



**UNIVERSIDAD  
TORCUATO DI TELLA**

Master in Management + Analytics

## **Marketing de Influencers en Instagram:**

Un framework para encontrar la mejor estrategia de marketing basado en algoritmos de clustering de influencers y similitud con el público objetivo

María Eugenia Irala  
Tutor: Elena Fumagalli

**Julio 2020**

## RESUMEN

A partir del nacimiento de las redes sociales, las compañías han lentamente abandonado las formas tradicionales de promocionar sus productos y servicios, para dar lugar a una nueva forma de hacer marketing: el llamado Marketing de Influencers. Este se basa en utilizar a los *influencers* de las redes sociales (nuevas “celebridades” en Internet) para que estos publiciten los productos de las empresas o marcas a través de sus publicaciones visitadas por millones de usuarios. No obstante, esta reciente estrategia de marketing online plantea una nueva problemática para las empresas, relacionada con la selección de los mejores influencers para promocionar los productos de manera exitosa, generando ganancias a las firmas. Dado que la literatura sobre esta temática es diversa y no sugiere una estrategia definitiva a seguir por las compañías, el presente estudio propone un conjunto de herramientas analíticas que puedan servirle a estas al momento de enfrentarse al problema principal del Marketing de Influencers. Particularmente, se sugiere la utilización de algoritmos de clustering sobre los datos de los influencers actuales para encontrar patrones que los caractericen y el posterior uso de estos resultados para llevar a cabo experimentos controlados que ayuden a comprender cómo influyen distintas características de los influencers sobre el comportamiento de los consumidores. De esta forma, gracias a la base de datos de influencers de Instagram provista por la compañía Upfluence, corrimos el algoritmo de *k-prototypes* para formar grupos de influencers e identificar sus características representativas. A partir de estos resultados, creamos perfiles de influencers ficticios de Instagram con el objetivo de realizar un experimento (N=213) que analice el impacto de la similitud entre el influencer y el consumidor sobre la disposición a comprar y pagar por el producto y la disposición a recomendar y a republicar el contenido del influencer. De la etapa de clustering, concluimos que la cantidad de seguidores (sinónimo de popularidad), es una de las variables más interesantes al momento de identificar distintos grupos de influencers, mientras que de la etapa experimental, concluimos que la similitud entre el influencer y el seguidor (medida a partir de la similitud ideológica o de postura ante la legalización del aborto en Argentina) parece tener un impacto significativo sobre la disposición a comprar del seguidor, cuando este está en contra de la legalización del aborto. En conclusión, esta investigación no solo puede considerarse como el primer intento de combinar un análisis de clustering (técnica de Machine Learning) con experimentos (método de investigación en Marketing), sino que también propone que las compañías tengan en cuenta la similitud que existe entre su público objetivo y los influencers que pretenden seleccionar para publicitar sus productos.

## Influencer Marketing on Instagram: A framework to find the best marketing strategy using clustering algorithms on influencer data and similarity with the target audience

### ABSTRACT

Since the emergence of social networks, companies have slowly abandoned traditional strategies for promoting their products and services, in favor of a new way of doing marketing: the so-called Influencer Marketing. It consists in advertising products through social media influencers (considered “online celebrities” on the Internet) and their posts viewed by millions of followers. However, along with this recent online marketing strategy a new business problem also arises: which influencers should companies and brands choose in order to maximize their sales? Given that the literature on this topic is diverse and does not suggest a definitive strategy to be followed by companies, this study proposes a set of analytical tools that can be useful for firms when facing the main problem of Influencer Marketing. Particularly, we suggest running clustering algorithms on current influencers data in order to find patterns that characterize them and using these results to carry out controlled experiments that help to understand how different influencers characteristics affect consumer behavior. Therefore, using the database of Instagram influencers provided by the company Upfluence, we ran a *k-prototypes* algorithm to identify groups of influencers. With these results, we created fictitious profiles of Instagram influencers in order to run an experiment (N = 213) that studies the impact of the similarity between the influencer and their followers on the willingness to buy and pay for the product and the willingness to recommend and republish the content of the influencer. From the clustering step, we conclude that the number of followers (measure of popularity) is one of the most interesting variables when identifying different groups of influencers, while from the experimental step, we conclude that the similarity between the influencer and their followers (which was measured using the ideological similarity or position regarding the legalization of abortion in Argentina) seems to have a significant effect on the willingness to buy (the product) of the follower when they are against abortion. In conclusion, this research can not only be considered as a first attempt to combine clustering analysis (machine learning technique) and experiments (marketing research method), but also proposes that companies should take into account the similarity that exists between their target audience and the influencers they select to advertise their products.

# Índice

<b>1. Introducción</b>	<b>3</b>
1.1. Motivación	3
1.2. Influencers	5
1.3. Instagram	8
1.4. Marketing por Redes Sociales	10
1.5. El Problema	12
1.6. Objetivo	13
1.7. Metodología	16
<b>2. Los Datos</b>	<b>18</b>
<b>3. Análisis Exploratorio</b>	<b>21</b>
3.1. Análisis dataset instagram	21
3.2. Análisis dataset instagram_posts	30
<b>4. Clustering</b>	<b>41</b>
4.1. Ingeniería de Datos, Transformaciones y Nuevas Variables	43
4.2. Implementación de <i>k-prototypes</i>	45
4.3. Conclusiones de Clustering	51
<b>5. Etapa Experimental</b>	<b>53</b>
5.1. Similitud entre el influencer y sus seguidores	54
5.2. Selección de la categoría "fashion influencers"	54
5.3. Promoción de productos: nivel de autoexpresión	55
5.4. Objetivo de realizar un Pretest	56
5.5. Pretest: Etapa Pre-Experimental	57
5.6. Resultados del Pretest	59
5.7. Experimento Principal	62
5.8. Diseño Experimental	63
5.9. Preprocesamiento y Análisis de las Respuestas	64
5.10. Análisis de los Resultados	72
<b>6. Conclusión</b>	<b>84</b>
<b>7. Bibliografía</b>	<b>88</b>
<b>8. Apéndice</b>	<b>91</b>
8.1. Variables de los Datasets	91
8.2. Transformaciones de las Variables	92
8.3. Gráficos Adicionales (Análisis Exploratorio)	94
8.4. Listado Final de Variables utilizadas para hacer Clustering	97

8.5.	<i>Influencer Personas</i> : Influencers Representativos a partir de la primera versión del algoritmo de Clustering.....	97
8.6.	Pretest: <i>Influencer Personas</i> y sus respectivas publicaciones en Instagram.....	101
8.7.	Pretest: Preguntas de la Etapa Pre-Experimental.....	106
8.8.	Experimento Principal: <i>Influencer Personas</i> y sus respectivas publicaciones en Instagram.....	107
8.9.	Experimento Principal: Preguntas del Experimento Final.....	108
8.10.	Experimento Principal: Resultados de Regresiones de la Disposición a Pagar, Recomendar y Republicar.....	113
8.11.	Experimento Principal: Cálculo de Potencia $(1 - \beta)$ post hoc para todas las Regresiones.....	120

# 1. Introducción

## 1.1 Motivación

El surgimiento de las plataformas de redes sociales en la última década ha cambiado completamente nuestra forma de interactuar y nuestra percepción de la vida de otras personas. Desde que plataformas como Facebook, Instagram, YouTube y Twitter aparecieron, las personas han comenzado paulatinamente a integrar estas innovaciones tecnológicas a su vida diaria, no solo utilizándolas como nuevos canales de comunicación, sino también como plataformas desde las cuáles leer noticias online, compartir fotos, experiencias e incluso comprar productos.

Por supuesto, las compañías y marcas eventualmente se percataron de que las redes sociales se convertían en una poderosa oportunidad para hacer marketing online. De hecho, esta creciente necesidad de las empresas por tener una presencia online fue contemporánea al surgimiento de los *influencers* en las redes sociales (en inglés SMIs: *social media influencers*), quienes pueden ser definidos como “celebridades online”, que son capaces de influenciar las actitudes y opiniones de las audiencias al exhibir sus vidas privadas a través de las redes. De esta forma, el *Influencer Marketing* o Marketing de Influencers surge como una combinación entre dos factores: el comportamiento de los consumidores estaba siendo altamente influenciado por las redes sociales y las compañías comenzaron a tener un interés en estos influencers digitales para utilizarlos como herramienta de comunicación y de marketing.

Es así como, en la actualidad, las empresas están abandonando cada vez más las formas tradicionales de marketing (TV, radio, diarios, etc.) y los *sponsors* tradicionales (actrices y actores, celebridades, modelos, etc.) para adentrarse más en el mundo del marketing por redes sociales o marketing digital, recurriendo a personalidades famosas en estas plataformas, como *vloggers*, influencers, personalidades famosas en Instagram, para que promocionen sus marcas o productos (Marwick 2015). Esto requiere un gran cambio de perspectiva, ya que mientras que con la manera tradicional sólo bastaba conseguir un *sponsor* muy famoso que publicite el producto/servicio en algún canal de TV que se adecúe al público objetivo, con el marketing online existe la posibilidad de ser extremadamente específicos con el público objetivo y la manera de llegar a él, por lo cual las estrategias de marketing deberán ser más detalladas y elaboradas para cumplir con el objetivo propuesto.

No obstante, a partir del surgimiento del Marketing de Influencers, aparece un nuevo problema de negocios: ¿Qué influencers deberían elegir las compañías de manera tal que se maximicen sus beneficios de venta? Lo cierto es que las empresas parecen no haber cambiado sus estrategias lo suficiente como para implementar óptimamente el marketing online. De hecho, según la literatura sobre esta temática, las firmas tradicionalmente usan el número de seguidores o el

llamado *engagement rate* para decidir cuál influencer contratar para promocionar sus productos y servicios.

Anil Narassiguin y Selina Sargent en su artículo llamado *Data Science for Influencer Marketing: feature processing and quantitative analysis*, publicado en 2019, afirman que el Marketing de Influencers todavía se basa en métricas cuantitativas simples como el tamaño de la comunidad (cantidad de seguidores) o el *engagement rate*<sup>1</sup> para estimar el valor de un influencer. Por lo cual, las compañías siguen centrándose en encontrar al influencer más “famoso” para publicitar sus productos, replicando los métodos tradicionales utilizados en marketing que involucraban buscar a las personalidades más conocidas y famosas. Sin embargo, haciendo esto, se están desperdiciando muchas de las oportunidades nuevas que brinda el marketing por redes sociales.

En concordancia con esto, estudios más recientes cuestionan esta metodología y proponen nuevas formas de encontrar el "influencer óptimo" para construir la estrategia de marketing online más rentable. Junto con estos enfoques disruptivos, identificamos algunos inconvenientes de la metodología establecida:

- Es bien sabido que, hoy en día, las empresas tienen acceso a una gran cantidad de datos. Por lo tanto, el hecho de que solo se utilicen unas pocas variables numéricas (número de seguidores, cantidad de *likes*, comentarios) es reductivo debido a que no se están teniendo en cuenta las características de la audiencia.
- Existe una suposición implícita de que cuánto más grande/famoso es el influencer (en términos de cantidad de seguidores), mayor es el retorno para las compañías. Sin embargo, según Disha Dinesh en *Why Micro-Influencers are a Social Media Marketing Imperative for 2017*, los consumidores tienden a confiar más en los “micro influencers” o influencers con menos seguidores debido a que la aversión a las publicidades está aumentando.

En base a lo dicho anteriormente, puede decirse que existe una brecha entre lo que actualmente están haciendo las compañías en términos de estrategias de marketing online y lo que sería mejor hacer para maximizar sus beneficios. Esta brecha debería abordarse mediante la

---

<sup>1</sup> El llamado *engagement rate* se define como la tasa de participación del influencer en las redes sociales. Formalmente, puede definirse como una métrica que indica cuán visible es el influencer en las redes sociales y se calcula dividiendo la cantidad de *engagements* (cantidad de comentarios, *likes*) por la cantidad de impresiones (o sea, la cantidad de veces que el contenido del influencer fue visto). Como esta última variable es difícil de obtener, se divide la cantidad *engagements* promedio (cantidad de comentarios, *likes* por publicación) por la cantidad de seguidores. De esta forma, se obtiene un *proxy* del porcentaje de seguidores que ven el contenido del influencer (Narassiguin y Sargent 2019).

investigación de qué características de los influencers se relacionan con las características de los consumidores (por ejemplo, similitud ideológica, física, etc.) y de cómo cambia el comportamiento del consumidor ante distintas características de los influencers. En consecuencia, nuestro objetivo principal es desarrollar un nuevo enfoque que aborde los dos inconvenientes mencionados anteriormente y ayude a las empresas a adoptar una perspectiva más amplia en el proceso de elegir el influencer óptimo para promocionar sus productos. Para hacer esto, estamos interesadas en investigar qué características de los influencers tienen un efecto en la disposición a comprar el producto de los consumidores, en su disposición a comprar, en su disposición a recomendar y, por último, en su disposición a republicar el contenido del influencer.

Particularmente, este trabajo consta de dos partes. Una primera parte involucra ahondar sobre las características de los influencers actuales (particularmente de Instagram), a partir de una base de datos provista por Upfluence, una empresa relacionada directamente con el Marketing de Influencers. Luego, en una segunda etapa, realizaremos un experimento que nos ayudará a entender más profundamente cómo las características de los influencers pueden afectar el comportamiento del consumidor y, por ende, tener un efecto sobre variables de negocio relevantes, como la disposición a comprar el producto publicitado o pagarlo.

## 1.2 Influencers

Hemos hecho referencia al surgimiento de los llamados *influencers* en las redes sociales, o también llamados SMIs, pero ¿quienes son realmente los influencers y a qué se dedican?

En particular, los influencers de las redes sociales pueden ser descriptos como “celebridades virtuales”, y representan un nuevo tipo de agente independiente que, a partir de compartir su vida privada, sus opiniones e incluso los productos que utiliza, da forma a las actitudes de la audiencia (usuarios de las redes sociales) a través de blogs, tweets y el uso de otras redes (Freberg et al. 2011). Aunque algunos estudios consideran a estos nuevos agentes como negativos para las compañías, por la posibilidad de que estos perjudiquen la reputación de las marcas o generen rumores falsos que se propaguen por Internet (Gorry y Westbrook 2009), otros consideran que la aparición de los influencers ofrece nuevas posibilidades para las marcas y organizaciones, que pueden potencialmente forjar “alianzas” con estos agentes para promocionar sus productos y servicios (Gaines-Ross 2003).

La particularidad de estos nuevos “jugadores” o “famosos” en el entorno online o en Internet, es que se los ha estudiado desde diversos ángulos. Desde un punto de vista sociológico, se ha estudiado a los influencers como comunicadores y fuentes de información nuevas en el mundo. De hecho, se los ha asociado con posibles “modelos a seguir” por parte de la comunidad



adolescente o como símbolos o actores relevantes al promover el debate de ciertas problemáticas sociales. En particular, Aran-Ramspott, Fedele y Tarrago (2018) en *Youtubers' social functions and their influence on pre-adolescence* demuestran que los pre-adolescentes consideran a los influencers como referentes del entretenimiento y los sienten cercanos a una cultura digital adolescente. Por otra parte, Crystal Abidin (2019) en *Yes Homo: Gay influencers, homonormativity, and queerbaiting on YouTube* plantea el hecho de que los influencers que se manifiestan como parte de la comunidad LGBT+<sup>2</sup> utilizan su voz para promover el debate y el activismo discursivo, generando así mayor cercanía con sus seguidores y fortaleciendo, al mismo tiempo, su carrera online como celebridades.

Estos estudios nos dejan en claro que los influencers tienen un rol de comunicadores en estas nuevas plataformas online. Por otra parte, algunas investigaciones se han enfocado en estudiar las características específicas de estos influencers, como por ejemplo, la cantidad de *hashtags*<sup>3</sup> o páginas web que mencionan en promedio en sus publicaciones en Twitter (Lahuerta-Otero y Cordero-Gutiérrez 2016), mientras que otras se han enfocado en estudiar la lógica de difusión de los mensajes de los influencers a través de Internet mediante el *retweeting* (republicar contenido en Twitter) o, en términos más genéricos, *rebroadcasting* (difusión o republicación del contenido del influencer). Por ejemplo, Zhan, Moe y & Schweidel (2016) en *Modeling the role of message content and influencers in social media rebroadcasting* concluyen que la republicación de contenido depende no solo del mensaje, sino también de la compatibilidad o *fit* entre el usuario que republica y el mensaje.

Sin embargo, la mayoría de los estudios que tienen como sujeto de estudio a los influencers, están orientados a entender qué rol juegan estos en el **marketing online**. En particular, este tipo de marketing busca promocionar productos y servicios a través de Internet, por lo cual, siendo los influencers los nuevos actores relevantes en las redes, se convirtió en una oportunidad tentadora para las empresas intentar utilizarlos para difundir con más efectividad las publicidades de sus marcas.

Es en este punto donde los trabajos de investigación comienzan a intentar entender cuál debería ser la estrategia a seguir por las empresas para promocionarse mejor. La cantidad de seguidores

---

<sup>2</sup> Lesbianas, Gays, Bisexuales, Transexuales, Intersexuales, Queer y Asexuales (Abidin 2019)

<sup>3</sup> Se define como *hashtags* a las palabras que son utilizadas con el símbolo # por delante. En general, estas palabras ayudan a que ciertos temas se difundan a través de las redes y que sea más fácil encontrar información referida a estos tópicos. De esta forma, los *hashtags* ayudan a organizar la información de las redes sociales (Lahuerta-Otero y Cordero-Gutiérrez 2016). Por ejemplo, *#JustDoIt* expresa el lema de la marca *Nike* por lo que todos los comentarios que utilicen este *hashtag* podrán ser fácilmente encontrados por la compañía, simplemente filtrando los comentarios que contengan *#JustDoIt*.

de estas celebridades online pasó a ser una característica muy estudiada, ya que se considera un sinónimo de popularidad. Es así como De Veirman, Cauberghe y Hudders (2017) en *Marketing through Instagram influencers: the impact of number of followers and product divergence on brand attitude* plantean que aquellos influencers en Instagram con mayor cantidad de seguidores son percibidos como más populares y por ende los usuarios en las redes los ponen en el rol de "líderes de opinión". No obstante, descubren que el hecho de que el influencer siga a pocas cuentas genera un efecto negativo en la audiencia, haciendo que esta se vea reacia a comprar productos promocionados por este tipo de influencers. Luego, investigaciones como la expuesta en *Influencer Marketing: How Message Value and Credibility Affect Consumer Trust of Branded Content on Social Media* llevada a cabo por Lou y Yuan (2019) plantean cómo el valor informativo del contenido generado por influencers, la confiabilidad, el atractivo y la similitud de estos con sus seguidores afecta positivamente la confianza de los seguidores en las publicaciones de marca de los influencers, que posteriormente influyen en el conocimiento de la marca y las intenciones de compra.

De esta manera, varios autores han intentado comprender más en profundidad el Marketing de Influencers, y han intentando dar respuesta a la pregunta de qué influencers debería usar una compañía al momento de promocionar sus productos y servicios online. Surgen así conceptos como credibilidad del influencer, nivel de similitud con los seguidores y popularidad, que efectivamente van a tener un impacto sobre la disponibilidad a pagar o comprar el producto de la audiencia.

Con respecto a la popularidad del influencer, Zietek (2016) ha propuesto categorías para clasificar a los influencers según la cantidad de seguidores, distinguiendo entre:

- Micro-Influencers: Aquellos que tienen menos de 15,000 seguidores.
- Regular-Influencers: Aquellos que tienen entre 15,000 seguidores y 50,000 seguidores.
- Rising-Influencers: Aquellos que tienen entre 50,000 seguidores y 100,000 seguidores.
- Mid-Influencers: Aquellos que tienen entre 100,000 seguidores y 500,000 seguidores.
- Macro-Influencers: Aquellos que tienen entre 500,000 seguidores y 1,000,000 seguidores.
- Mega-Influencers: Aquellos que tienen más de 1,000,000 seguidores.

Por supuesto, esta categorización no es la única que existe para clasificar a los influencers. A su vez, pueden clasificarse según el contenido que difunden en las redes sociales. Por ejemplo, existen influencers de moda, de belleza o maquillaje, influencers que promocionan principalmente sus viajes o actividades deportivas y muchos más.

Dado esto, podemos concluir que los influencers tienen varias características que las empresas deben tener en cuenta al momento de seleccionar aquellos que sean idóneos para poder publicitar sus productos. Sin embargo, la mayoría de los estudios realizados sobre Marketing de Influencers se han realizado en pequeños subconjuntos de datos (cientos a miles de influencers) para redes sociales específicas y en su mayoría datos estructurados (número de seguidores, *engagement rate*, ROI, etc.), por lo que las conclusiones suelen ser similares y relacionadas directamente con asociar éxito de la promoción del producto con la popularidad del influencer (gran cantidad de seguidores y alto *engagement rate*). Pocos estudios abordan cómo otras características de los influencers pueden afectar el comportamiento del consumidor.

En este trabajo de investigación, nos proponemos estudiar más en profundidad las características de estos nuevos agentes, haciendo un análisis exhaustivo con una mayor cantidad de datos y aplicando técnicas de Analytics y Data Science que aún no se han aplicado en este campo de estudio.

### 1.3 Instagram

Hemos hablado del surgimiento de las plataformas de redes sociales y hemos mencionado las más conocidas (Facebook, Instagram, YouTube, Twitter). No obstante, en el presente trabajo nos enfocaremos en Instagram.

Esta red social, propiedad de Facebook, nació en Estados Unidos en 2010. A diferencia de Facebook, que se enfoca en que aquellos que tengan una cuenta puedan crearse un perfil para compartir fotos, experiencias, opiniones, crear grupos para compartir determinado contenido, etc., Instagram se enfoca en que las personas que lo utilicen compartan principalmente contenido visual, como fotografías y videos muy cortos. Dejando de lado la lógica propia de Facebook, en la cual ser “amigo” de un usuario generaba que ambos se siguieran mutuamente, en Instagram, un usuario puede seguir a otro aunque este último no lo siga. La cantidad de seguidores de un perfil en específico se contabilizan como “followers”, y tienen la capacidad de dar *like*, compartir y/o comentar el contenido de los perfiles a los cuales siguen. Cuantos más seguidores tiene un perfil, mayor cantidad de interacciones totales recibirá. Los perfiles también tienen la posibilidad de subir videos o fotos denominadas “Historias”, que desaparecen a las 24 horas de haberlas subido.

Si bien los influencers suelen tener una presencia fuerte en todas las redes sociales que se utilizan actualmente, Instagram es una de las redes preferidas de estas celebridades online, debido a la gran oportunidad que ofrece de compartir contenido plenamente visual, permitiendo así mostrar la vida privada y experiencias del influencer.

A su vez, Instagram es una plataforma que se basa en la estética visual y las imágenes filtradas, lo que la convierte en un ecosistema adecuado para promocionar productos de belleza, popularizar ciertas imágenes corporales y abogando por estilos de vida lujosos y marcas de lujo prominentes (Djafarova y Rushworth, 2017). Instagram también es una aplicación de redes sociales que permite a los usuarios reunir seguidores, conectarse con diferentes marcas y entidades, y facilitan las interacciones sociales entre los consumidores (Blight et al., 2017).

Con respecto a su importancia con respecto a otras redes sociales, según el portal Statista, para septiembre de 2019, Instagram era la segunda red social más popular en los Estados Unidos según usuarios mensuales activos (121.23 millones). En primer lugar, se encontraba Facebook con 169.76 millones de usuarios mensuales activos, en tercer lugar Facebook Messenger (106.4 millones de usuarios) y en cuarto lugar Twitter (81.47 millones de usuarios). Para la misma fecha, Instagram también resultó la segunda aplicación más popular en los Estados Unidos según porcentaje de usuarios de teléfonos móviles con la aplicación descargada, siendo que el 64.92% de los portadores de celulares poseían la aplicación.

En el mundo profesional del marketing online, Instagram también toma un papel muy relevante. Según Statista, para enero 2019, Instagram mundialmente era la segunda red social más utilizada por profesionales de marketing, con un 73% de los expertos en el área usándola. El ranking lo lidera Facebook con un 94% de los profesionales usándola y en tercer lugar se encuentra Twitter con un 59%.

Lo cierto es que muchas empresas eligen Instagram como target para su presupuesto de marketing. Esto puede observarse claramente en números ya que esta red social está teniendo un crecimiento periódico cada año en cuanto a ganancias por publicidad. Mientras que en 2018 Instagram generó 6.18 billones de dólares en conceptos de ganancia por publicidad, en 2019 ese monto subió un 53% a 9.45 billones de dólares. Para 2020, se estima un crecimiento parecido, y se espera que las ganancias por publicidad alcancen los 13.86 billones de dólares. Si bien Facebook sigue siendo la red social que mayores ganancias por publicidad genera (cerca de 69.9 billones de dólares en 2019), el tipo de publicidad que exhibe es diferente a la que se puede encontrar en Instagram. Mientras que en Facebook se pueden encontrar en su mayoría publicidades mucho más parecidas al estilo de la televisión (las empresas pagan para que reproduzcan su publicidad pre-grabada sobre algún producto determinada cantidad de veces al día por un período de tiempo), en Instagram la existencia y popularidad de los *insta-influencers* abrieron el paso a que, además del tipo de publicidad anteriormente descrito, se expanda la publicidad a través de sponsors. En 2017, 2.58 millones de publicaciones en Instagram estaban patrocinadas, creciendo esa cantidad a 3.7 millones en 2018 y a 4.95 millones en 2019.

Teniendo en cuenta estas características únicas de Instagram, este estudio investigará las cualidades de los influencers en esta plataforma. Por otra parte, la predominancia del contenido fotográfico y visual nos será de utilidad al crear perfiles de influencers ficticios durante la etapa de experimentación de manera fácil y realista.

## 1.4 Marketing por Redes Sociales

Los SMIs atraen millones de seguidores al compartir contenido en su perfil de redes sociales como YouTube, Instagram, Twitter, etc. Normalmente, cada uno se desenvuelve alrededor de un dominio en particular como, por ejemplo, belleza, moda, videojuegos, entre otros. Si bien las celebridades tradicionales también lograron introducirse en las redes sociales, lo que las diferencia de los influencers es que estos últimos construyeron sus carreras en línea mientras que los primeros ya eran conocidos por otros motivos antes del surgimiento de las redes sociales (Schouten et al. 2020).

En la actualidad, las empresas invierten grandes cantidades de dinero en marketing online y sponsors en redes sociales. Para encontrar soluciones óptimas al problema de asignación de presupuesto en marketing online, se han realizado una diversidad de estudios académicos en cuanto al marketing online a través de influencers, para investigar qué rol juegan en las intenciones de compra y de pago de los consumidores. Por ejemplo, Colliander y Dahlén (2011) en su artículo *Following the fashionable friend: the power of social media: weighing publicity effectiveness of blogs versus online magazines*, encontraron que una publicación sobre una marca de ropa de moda resultaba en una mejor actitud hacia la marca y una mayor intención de compra comparado con un artículo en una revista. La explicación de este resultado fue que los compradores se sentían más cercanos al influencer. Por otro lado, Lee y Watkins (2016), en *YouTube vloggers' influence on consumer luxury Brand perceptions and intentions*, demostraron que los influencers afectan positivamente las intenciones de compra para marcas de lujo.

También se ha sugerido en varios estudios, entrevistas y artículos que los influencers son percibidos como más creíbles y "memorables" que sus pares tradicionales (actores, modelos, cantantes, etc.). A su vez, las reseñas sobre productos que hacen los SMIs tienen un impacto significativo mayor en el comportamiento de los consumidores con respecto a las publicidades hechas por celebridades tradicionales, debido a que los influencers son percibidos como más confiables, similares y cercanos a sus audiencias (Schouten et al. 2020).

Teniendo en cuenta la creciente participación de los SMIs en las estrategias de marketing online de las empresas, puede decirse que el funcionamiento y efectividad del marketing a través de

redes sociales puede resumirse en dos procesos mayores: la **identificación** con el influencer sponsor y la **credibilidad** percibida del influencer sponsor.

El proceso de identificación funciona de manera tal que si los consumidores creen que comparten intereses, valores y/o características con el influencer, será más probable que adopten sus creencias y actitudes (Cialdini 1993; Kelman 2006). La identificación, según analizan Schouten, Janssen y Verspaget (2020) puede ser tanto **percibida** como **deseada**. En el caso de celebridades tradicionales, la identificación surgía principalmente de la identificación deseada, mientras que con los influencers en redes sociales la identificación es percibida. Esto es así ya que mientras que las celebridades se veían como personas inalcanzables, los influencers son vistos más como “amigos” o pares a su público. Los SMI se esfuerzan a diario para ser cercanos a sus seguidores, por lo que es común responder comentarios, republicar contenido de sus seguidores, dialogar o realizar actividades para unir a la comunidad. Como resultado de la identificación a través de las similitudes percibidas con los influencers, surge que éstos se vean como fuentes más auténticas y confiables en comparación a las celebridades tradicionales. Ante esto, es posible concluir que la publicidad llevada a cabo por influencers será más efectiva dado que existe una mayor probabilidad de que los consumidores acepten productos de sponsors con los cuales se identifiquen y en los que puedan confiar.

La credibilidad, por otro lado, funciona de modo que un sponsor creíble tiene mayor poder de persuasión y genera mayor disposición a recomendar y republicar. Los influencers son percibidos como fuentes más creíbles que las celebridades tradicionales ya que son más afines a compartir experiencias personales y a mostrar el producto en contexto de uso. Aunque un buen porcentaje del contenido de los influencers sea sponsoreado y con el propósito de elevar las ventas de un producto, los influencers tienden a reflejar opiniones honestas y el público generalmente lo percibe como recomendaciones auténticas (Evans et al. 2017). Varios estudios han encontrado que los SMIs son fuentes creíbles ya que demuestran haber probado el producto ellos mismos y realmente tener una opinión formada sobre el mismo (Uzunoglu and Kip 2014).

Por estos motivos, el marketing online a través de redes sociales tomó gran importancia en la actualidad, llegando incluso a reemplazar en buena parte al marketing a través de celebridades tradicionales. En los últimos años, los profesionales del marketing han ido reduciendo sus presupuestos en publicidades tradicionales a la par de aumentar sus gastos en marketing digital. Según la compañía de manejo de las redes sociales Lyfe Marketing, en 2018 se destinaron en Estados Unidos 114.84 billones de dólares a campañas de marketing tradicional y 108.64 billones de dólares a marketing digital. Esta tendencia coincide con lo señalado por Chen Lou y Shupeí Yuan (2019) en *Influencer Marketing: How Message Value and Credibility Affect Consumer Trust of Branded Content on Social Media*: El Marketing de Influencers ha mostrado un crecimiento

exponencial en los últimos años, observándose ya por 2018 que el 39% de los especialistas en marketing planeaba aumentar su presupuesto para el marketing de influencers y el 19% de estos especialistas tenían la intención de gastar más de 100.000 dólares por campaña.

En 2019, según Lyfe Marketing, el marketing digital pasó a encabezar los presupuestos de publicidades, siendo que se destinaron 129.34 billones de dólares al marketing digital y solo 109,48 millones de dólares al marketing tradicional. Se estima que esta tendencia continuará así, con los recursos destinados a las nuevas formas de marketing en la suba y los de marketing tradicional estancados o en la baja. Para 2020, esta empresa estima que se destinarán 151.29 billones de dólares al marketing virtual y 107.13 billones de dólares al tradicional. Para 2023, la suma destinada al e-marketing fue estimada en 201.83 billones de dólares mientras que para el tradicional se estiman solo 100.48 billones.

Debido al creciente papel del marketing digital y de influencers en las estrategias de promoción de las empresas actuales, es de vital importancia dedicar recursos a la investigación de la forma más óptima de explotar el marketing por redes sociales.

## 1.5 El Problema

Si bien se han llevado a cabo varios estudios al respecto, hay todavía mucha confusión e incertidumbre sobre cuáles son los factores y variables que impulsan la propensión de pago y de compra a través de marketing por redes sociales. Existe una diversidad de estudios que demuestran que el marketing online está resultando ser más efectivo que el tradicional, pero nadie ha sido capaz de investigar las diferencias entre distintos conjuntos de influencers y cómo elegir al influencer que resultará más efectivo según el producto que se publicite. Sospechamos que debe existir algún tipo de conexión o *insight* entre el tipo de producto, de influencer y de consumidor que, si se llega a analizar y estructurar, se pueda aprovechar al máximo y de esa manera elevar el nivel de ventas y de éxito de las campañas publicitarias en redes sociales.

Según Karen Freberg, Graham y McGaughey (2011) en *Who are the social media influencers? A study of public perceptions of personality*, entender la personalidad percibida de los influencers provee herramientas para optimizar el "capital de influencers" que utiliza una empresa. Es decir, si podemos encontrar el driver en la personalidad del influencer (o del producto) que impulsa las ventas, podremos obtener mejoras en los beneficios de las empresas al mejorar el "stock" de influencers con los que se trabaja. El uso óptimo de redes sociales se desencadena en mayores ventas y mayores *profits* (Galeotti y Goyal 2009), por lo que dominando esta área se pueden obtener mejoras empresariales significativas.

Ya que el problema de descubrir el nodo más persuasivo en una red social se comprobó que es NP hard (More y Lingam 2019), lo que intentaremos al realizar este estudio es brindar una serie de herramientas analíticas que ayuden a las empresas a encontrar a decidir con qué tipo de influencers debe trabajar dado su contexto de producto y público objetivo.

## 1.6 Objetivo

Según lo desarrollado anteriormente, sabemos que actualmente no existe una estrategia definitiva a seguir por las compañías al momento de seleccionar influencers para publicitar productos y marcas en las nuevas plataformas sociales en Internet. Se han elaborado diversos estudios enfocados en describir a los influencers y el efecto de su popularidad, sus ventajas con respecto a las celebridades tradicionales, y varias investigaciones siguen haciendo hincapié en el número de seguidores y el *engagement rate* como métrica principal a observar para seleccionar SMIs.

Con el presente trabajo, buscamos seguir la línea de estas investigaciones, pero enfocándonos en generar un conjunto de herramientas analíticas que pueden servirle a las empresas al momento de enfrentarse al problema principal del Marketing de Influencers. En particular, proponemos la utilización de algoritmos de clustering utilizando los datos reales de influencers en redes sociales para así encontrar patrones característicos de estos. De esta manera, estamos buscando tener en cuenta todas las características posibles de los influencers para luego formar grupos de influencers o influencers representativos.

La literatura relacionada con la investigación del Marketing de Influencers ha hecho varios intentos de segmentar a los influencers en distintos grupos según sus características.

En *Who are the social media influencers? A study of public perceptions of personality*, Karen Freberg, Graham y McGaughey (2011) se propusieron detectar atributos clave de los influencers a partir de analizar las percepciones de alumnos de una gran universidad pública, enfocándose en cuatro SMIs de muestra. A partir de las respuestas de los estudiantes, los autores generaron prototipos de perfiles para cada uno de los cuatro influencers seleccionados. Una particularidad de los resultados obtenidos es que los SMIs, a diferencia de la percepción de los prototipos de CEOs en una compañía, fueron percibidos como agentes que disfrutaban de dar consejos a su público. Luego, Del Fresco García et al. (2016), a partir de un Análisis de Redes Sociales (en inglés: *Social Network Analysis*) basado en la construcción de grafos para visualizar las relaciones entre los usuarios (nodos), identifican a los influencers como aquellos usuarios con una posición de alto rango en la red a partir de su cantidad de retweets en Twitter, menciones y respuestas. Al analizar los resultados, los autores encontraron que los influencers pueden ser categorizados en



tres tipos según su rol en la red social: diseminadores, relacionales y líderes. Con un enfoque similar y utilizando también *Social Network Analysis*, Himelboim y Golan (2019) identifican tres tipos de influencers: primarios (“centros” o “ejes”), contextuales (“puentes”) y de baja influencia (“aislados”). A diferencia de Del Fresco García et al., estos autores analizaron los tweets que específicamente hacían mención de un video de YouTube con una campaña de publicidad de Heineken, por lo cual logran enmarcar cada rol de influencer dentro de una publicidad online viral.

A su vez, varios artículos se enfocan no sólo en descubrir características que identifiquen a los influencers, sino también desarrollar estrategias para detectarlos con mayor probabilidad en las redes sociales. Esta tendencia puede identificarse en *Looking for the perfect tweet. The use of data mining techniques to find influencers on twitter* de Lahuerta-Otero y Cordero-Gutiérrez (2016), cuyo objetivo de investigación es encontrar aquellas características que identifican los tweets de influencers, como por ejemplo, cantidad de *hashtags* utilizados o cantidad de caracteres promedio por tweet. No obstante, este tipo de estrategia busca facilitar (para los expertos en marketing en las empresas) el hallazgo de influencers a partir del análisis exhaustivo de lo que publican en Twitter. Por otro lado, en *Influence factorization for identifying authorities in twitter*, Alp y Öğüdücü (2019) recolectan datos de Twitter y, a través de un algoritmo, identifican distintos tópicos y a los influencers más relevantes dentro de cada tópico (utilizando técnicas algorítmicas y modelos de Machine Learning). More y Lingam (2019), por su parte, en su artículo llamado *A SI model for social media influencer maximization*, proponen utilizar el modelo *Suspected-Infected* (SI), usado para analizar el proceso de difusión en pandemias, para identificar a los influencers centrales (nodos más “persuasivos”) dentro de una red social.

Teniendo en cuenta estos estudios, nos pareció particularmente interesante el enfoque de Narassiguin y Sargent (2019), el cual se caracteriza no sólo por analizar dos redes sociales además de Twitter (red preferida dentro de los artículos encontrados sobre esta temática), como Facebook e Instagram, sino también incorporar variables demográficas propias del influencer (idioma, edad, género) con el objetivo de plantear una metodología de análisis descriptivo de los SMIs. Por otro lado, los autores aclaran que su metodología deja abierta la puerta a utilizar métodos de clustering para caracterizar grupos de influencers presentes en las redes sociales. Con esto en mente y a diferencia de los trabajos realizados con el objetivo de segmentar SMIs, proponemos utilizar algoritmos de clustering no sólo incorporando variables relacionadas directamente con la actividad en redes sociales (publicaciones, *likes*, comentarios), sino también variables demográficas o descriptivas del influencer como individuo (idioma, edad, género, categoría, si tiene otras redes sociales, etc.).

Por otro lado, a partir de esta estrategia que busca entender qué influencers existen en las redes sociales actualmente, quisimos analizar, mediante un experimento, cómo influyen ciertas características del influencer en el comportamiento de los consumidores.

Específicamente, queremos investigar si la similitud entre el influencer y sus seguidores tiene un impacto en los resultados de negocio de las empresas. Esta pregunta de investigación se ha analizado ampliamente en la bibliografía de esta temática, aunque nunca utilizando resultados propios de un análisis de datos profundo. En la mayoría de ellos se han utilizado encuestas con influencers reales y se le ha preguntado a los participantes del experimento si se identificaban o se sentían similares a estos. Esto puede encontrarse en *Celebrity vs. Influencer endorsements in advertising: the role of identification, credibility, and Product-Endorser fit* de Schouten et al. (2020), en donde se realiza una comparación entre los influencers y las celebridades tradicionales con respecto a su efectividad en actividades de promoción o publicidad. Estos autores hacen este estudio seleccionando las categorías *fitness* y *beauty*, y haciendo una encuesta de pretest para obtener, para cada categoría, el influencer y la celebridad de mayor público conocimiento. A partir de esto, diseñan una encuesta final donde se le presenta a los participantes un influencer y celebridad junto con una publicidad de un producto de cada categoría. En particular, los autores buscan medir el nivel de similitud percibida por los participantes con respecto a los influencers y las celebridades, simplemente a través de afirmaciones como “esta persona piensa como yo” o “esta persona se comporta como yo” a la cual los participantes debían responder en una escala de 1 (en total desacuerdo) a 7 (totalmente de acuerdo). Esto no generó complicaciones debido a que entre 50% y 60% de los encuestados (N=131) conocían a los influencers y celebridades elegidos. Schouten et al. pudieron concluir con este estudio que los participantes se sintieron más identificados con los influencers, se percibieron con mayor similitud a los influencers que a las celebridades y expresaron una mayor confianza en los influencers. Una de las implicancias de esto, plantean los autores, es que los influencers parecen ser más efectivos que las celebridades tradicionales en campañas de marketing o publicidad.

Ki y Kim (2019), por su parte, se propusieron investigar el mecanismo por el cual los influencers en las redes sociales persuaden a los consumidores de comprar o adoptar ciertas marcas. Plantean que un mediador entre la actividad del influencer y la acción del consumidor es el *deseo de imitar* por parte de los consumidores, y lo probaron a partir de una encuesta a 395 individuos cuyas respuestas alimentaron un modelo estadístico Probit. Si bien esto no se enmarca dentro del estudio de la similitud percibida (que es la que nos interesa analizar), podría formar parte del concepto de similitud deseada. Luego, en *Social media influencers as endorsers to promote travel destinations: an application of self-congruence theory to the Chinese Generation Y*, Xu y Pratt (2018) realizan un análisis similar a Schouten et al. (2020) y Ki y Kim

(2019) pero aplicado directamente en la categoría de promoción de viajes en China. A través de una encuesta y del uso de ANOVAs, estos autores buscan analizar el impacto de la *congruencia* entre el influencer de viajes y los consumidores sobre la posterior intención de visita al destino de los consumidores. Seleccionaron dos influencers muy conocidos por los jóvenes en China a partir de un pretest, y definieron la presencia de congruencia con el participante a partir de construir un índice de congruencia (surgido de las respuestas del participante con respecto al influencer y a él mismo) que expresara cuán cercana estaba la imagen ideal del consumidor de la imagen del influencer. Como puede notarse, esta investigación también hace uso del concepto de similitud deseada. Por último, Lou y Shupeí (2018) proponen un modelo teórico, utilizando Regresiones de Mínimos Cuadrados Parciales (*Partial Least Squares*) para entender los efectos del marketing de influencers, incorporando el valor del contenido del influencer, su credibilidad, su atractivo y similitud con sus seguidores. Específicamente, estos autores quieren ver qué efectos tienen todas estas variables en la intención de compra y el *awareness* de las marcas promocionadas. Haciendo uso de encuestas, Lou y Shupeí llegan a la conclusión de que los usuarios de las redes sociales tienen confianza en las publicidades de marcas hechas por SMIs, por lo cual sugieren a las compañías no sólo enfocarse en la cantidad de seguidores y niveles de *engagement*, sino también en cuán confiable/attractivo perciben los usuarios al influencer, el nivel de similitud entre ambos, entre otros. Este estudio fue hecho a partir de hacerles preguntas a los participantes sobre los influencers que actualmente estaban siguiendo, sin pre-seleccionar influencers mediante algún pretest.

Por lo cual, dado el enfoque de la literatura actual, proponemos una estrategia novedosa para investigar el impacto de la similitud (entre influencer y seguidor) sobre el comportamiento de los seguidores: utilizar una temática altamente polémica en la Argentina como lo es la legalización del aborto para poder analizar qué efecto tiene la similitud ideológica entre el influencer y el potencial seguidor sobre la disposición a pagar, comprar, recomendar y republicar. De esta forma, estamos combinando un análisis de negocio con conceptos ideológicos presentes en la sociedad. Al mismo tiempo, planteamos un estudio más profundo de la similitud percibida por los seguidores utilizando una temática específica y haciendo uso de influencers ficticios (creados a partir de la primera etapa de clustering) a diferencia de los trabajos de investigación ya existentes que usan influencers conocidos por la sociedad.

## 1.7 Metodología

Con el objetivo de investigar las características de los influencers en las redes sociales y cómo estas características pueden tener un efecto en el comportamiento del consumidor y variables de negocio, realizaremos un análisis compuesto de dos etapas:

**Parte 1 - Etapa Descriptiva:** Realizar un análisis exploratorio de los datos proporcionados y correr un algoritmo de clustering sobre estos para obtener diferentes perfiles de influencers. A partir de los resultados de esta etapa, construiremos “Personas” que representen las características representativas de los grupos identificados con el algoritmo. Es decir, utilizando los resultados del algoritmo, podremos construir influencers representativos de los patrones o tendencias que existen actualmente entre los influencers en Instagram.

**Parte 2 - Etapa Experimental:** Teniendo en cuenta los resultados de la etapa anterior, realizaremos un experimento utilizando perfiles ficticios de influencers en Instagram, contruidos a partir de las “Personas” que surgieron del algoritmo de clustering. La idea es manipular algunas variables que obtuvimos del paso descriptivo y evaluar cómo los diferentes perfiles o comportamientos del influencer (por ejemplo: un influencer afectan variables relevantes de negocio como la disposición a pagar, disposición a comprar, a recomendar o a republicar. Después de la recopilación de los datos, estos se analizarán mediante regresiones para verificar si la variable manipulada tiene un efecto significativo en nuestras variables dependientes seleccionadas.

## 2. Los Datos

Los datos que utilizamos en este estudio fueron provistos por Upfluence<sup>4</sup>, una compañía que recolecta y vende información de influencers a diferentes empresas interesadas en publicitar sus productos a través de los influencers en las redes sociales. En particular, los datos que recibimos fueron extraídos de Instagram.

Contamos con dos bases de datos:

- **instagram:** este dataset cuenta con información de 99,877 influencers. Cada fila corresponde a un influencer y cada columna corresponde a una característica o *feature* del influencer. En total, la base cuenta con 20 columnas: `influencer_id`, `email`, `name`, `language`, `country`, `address`, `gender`, `age_bracket`, `instagram_id`, `username`, `full_name`, `bio`, `website`, `followers`, `engagement_rate`, `total_posts`, `total_engagements`, `total_likes`, `total_comments`, `category`.
- **instagram\_posts:** este dataset tiene información de 438,548 posts. Cada fila corresponde a un posts de un influencer mientras que cada columna corresponde a una característica o *feature* del post. En total, la base cuenta con 12 columnas: `id`, `instagram_id`, `text`, `type`, `location_name`, `timestamp`, `likes`, `comments`, `post_id`, `views`, `media_type`, `thumbnail_url`.

Estas dos bases pueden unirse mediante la clave primaria **instagram\_id**, que es una combinación de números que identifican de manera única a la cuenta de instagram del influencer.

Con respecto al método de muestreo, Upfluence tomó datos desde el 4 de septiembre de 2019 hasta el 4 de diciembre de 2019. Por ende, los posts que se encuentran en la base *instagram\_posts* son los pertenecientes a los influencers de la base *instagram* pero que a la vez ocurrieron entre las dos fechas mencionadas.

A continuación se puede ver una muestra de los datos de **instagram**:

---

<sup>4</sup> <https://www.upfluence.com/>

Tabla 1: Muestra de influencers de la base de datos "instagram"  
Primeras 10 columnas

influencer_id	email	name	lang	country	address	gender	age_bracket	instagram_id	username
1515915	NA	pumora	de	DE	Roman Northern Gate	Gender_FEMALE	25-34	277756	pumora_anne
1503762	crystal@thedappergirl.com	thedappergirl	en	US	Texas,USA	Gender_FEMALE	25-34	194935	thedappergirl
2840663	babaluccia@hotmail.it	babaluccia	en	IT	Italy	Gender_FEMALE	35-54	256174	babaluccia
1775083	theincogneatist@gmail.com	theincogneatist	en	US	NewYork,NY,USA	Gender_FEMALE	18-24	282341	theincogneatist
1670661	oana@walkcatwalk.com	Oana	en	RO	Bucharest, Romania	Gender_FEMALE	0-17	260497	_walkcatwalk_
1724053	damsel@londondamsel.co	londondamsel	en	GB	London,UK	Gender_FEMALE	25-34	291001	damselinlondon
1483087	beyondthenail@gmail.com	spilledpolish	en	CA	Mississauga, ON, Canada	Gender_FEMALE	18-24	187991	spilledpolish
1617706	hello@goldfieldsgirl.com	goldfields_girl	en	AU	Golden Triangle Motel	Gender_FEMALE	25-34	289808	goldfieldsgirl
101274	klemen@ujusansa.si	Klemen	en	CN	NenjiangRiver,China	Gender_FEMALE	0-17	85644	mowgliii
134954	anthony@manvsclock.com	Anthony Middleton	en	MX	Mexico City, Federal	Gender_MALE	25-34	249804	manvsclock
1610710	hana@nirvanacakery.com	nirvanacakery	en	GB	London,UK	Gender_FEMALE	25-34	284863	nirvanacakery
1641762	megan@houseofgrays.com	houseofgrays	en	US	Laguna Beach, CA, USA	Gender_FEMALE	25-34	269710	houseofgraysco
1501980	mary@marymurnane.com	marymurnane	en	US	Chicago,IL,USA	Gender_FEMALE	35-54	286566	mary_murnane
1589215	realsimplegood@gmail.com	realsimplegood	en	US	Portland,OR,USA	Gender_MALE	18-24	275618	realsimplegood
1489138	NA	Kercia Jane	en	US	Minnesota,USA	Gender_FEMALE	18-24	269983	kerciajane
1492587	nicki@rebelinanevdress.co	by Nicki Nowicki	en	DE	Munich,Germany	Gender_FEMALE	25-34	249968	rebelinanevdress
1616116	NA	thestylecat	en	CA	Toronto,ON,Canada	Gender_FEMALE	25-34	287588	thestylecat
1590296	rachel@simplystoked.co	simplystoked__	en	AU	Australia	Gender_FEMALE	25-34	288987	simplystoked__
769893	medge@myvoguishdiaries	myvoguishdiary	en	CA	Toronto,ON,Canada	Gender_FEMALE	25-34	284996	myvoguishdiaries

### Siguientes 10 columnas

full_name	bio	website	followers	engagement	total_posts	total_engagement	total_likes	total_comments	category
Hand Embroidery by Anna	» Hand embroidery	http://pumora.de	17002	0.83673	722	109411	106156	3255	home
crystal dallas blogger	style + beauty #blogger	http://thedappergirl.com	12547	0.75888	349	52075	46311	5764	beauty
babaluccia	Blogger influence	https://www.babaluccia.com	18649	0.51515	560	119915	100580	19335	fashion
by Melissa Frusco	Curly Hair • Houston	https://www.theincogneatist.com	32788	0.43222	455	336189	310545	25644	beauty
Oana	Travel & Lifestyle	https://walkcatwalk.com	36418	0.36975	611	358916	325204	33712	travel
London Damsel	Londoner. Traveller	http://londondamsel.co.uk	7709	0.33272	641	80084	73082	7002	family
Maddy	Lifestyle/Beauty Blogger	https://amzn.to/2G8vZ8j	6935	0.31265	107	30308	28912	1396	beauty
Deneale - Goldfields Girl	• Lipstick wielder	https://linktr.ee/goldfieldsgirl	8085	0.31002	305	37457	34503	2954	family
Klemen	SURF LIFE PHOTOS	http://bit.ly/Sardis	12231	0.29347	57	31997	31125	872	travel
Traveller Anthony Middleton	♂ Former loser	https://manvsclock.com	19769	0.28867	235	78538	75735	2803	food
Hana Nirvana Cakery	Author of plant-based	https://nirvanacakery.com	10065	0.15618	214	120895	112514	8381	food
Megan Gray	What I love most about	http://houseofgrays.com	11374	0.15616	250	56139	54564	1575	home
Mary Murnane	Fashion   Photography	https://marymurnane.com	13240	0.15494	334	117756	103536	14220	fashion
The Real Simple Good Life	Justin + Erica Winick	https://tap.bio/@realsimplegood	59020	0.15424	553	177743	149437	28306	food
Kercia Jane Design	Designer + Artist	https://linktr.ee/kerciajane	5872	0.15416	158	39446	35491	3955	home
by Nicki Nowicki	♥ Fashion & Anti	https://wp.me/p6qf1p	11344	0.15410	231	129429	123698	5731	fashion
Jess Cheng	Senior creative @	https://youtu.be/7j8v	8736	0.15377	370	129273	116759	12514	beauty
Rachel Stokes	wife + mama raising	http://www.simplystoked.com	5554	0.15336	239	58768	50567	8201	family
Medge Beauvoir - Yellow	Fashion   Beauty	http://www.myvoguishdiaries.com	14644	0.15261	352	220373	205913	14460	beauty

En el Apéndice se describe en profundidad las variables y de qué tipo son (numéricas, *string* o de texto, *factor*, etc.). No obstante, es útil mencionar que las variables **age\_bracket** y **category** son de tipo factor (variables categóricas) cuyas categorías son las siguientes:

- **age\_bracket** cuenta con las categorías 0-17, 18-24, 25-34, 35-54.
- **category** cuenta con las categorías *beauty*, *family*, *fashion*, *food*, *gaming*, *home*, *lifestyle*, *sport*, *technology* y *travel*.

Estas variables nos servirán luego para hacer un análisis descriptivo de los datos. Particularmente, el método de muestreo de Upfluence, incluyó muestrear 10,000 influencers

para cada valor de *category* utilizando un mecanismo de búsqueda que detecta palabras clave específicas que pueden asociarse con cada categoría.

A su vez, el **engagement rate** fue calculado por Upfluence en base a la siguiente fórmula:

$$\text{engagement rate} = \frac{\text{average engagement}}{\text{followers}}$$

$$\text{engagement rate} = \frac{\text{total engagements/total posts}}{\text{followers}}$$

Donde la variable *total engagements* simplemente es igual a la suma del total de comentarios y el total de *likes*. De esta forma, el **engagement rate** es, efectivamente, un indicador de la performance del influencer con respecto a su visibilidad en las redes sociales.

Con respecto a la base de datos **instagram\_posts**, en la siguiente tabla puede observarse una pequeña muestra.

Tabla 2: Muestra de posts de influencers de la base de datos "instagram posts"

id	instagram_id	text	type	location_name	timestamp	likes	comments	post_id	views	media_type	thumbnail_url
457132837	547505	Everything But The Bag	picture	Arlington, Vir	1574115498	156	28	B5BfSEFgVjG	0	image	https://scontent-iad3-1.c
457132836	547505	Quick trip to NYC, my f	picture	Eataly NYC Fla	1573933198	36	2	B48DkmgwSG	0	image	https://scontent-iad3-1.c
436519812	547505	Tbt to a favorite fall ou	picture	Arlington, Vir	1571671920	88	35	B34qhydANH	0	image	https://scontent-iad3-1.c
436519811	547505	I lucked out in my teen	picture	Arlington, Vir	1571360023	67	7	B3vXoZcAxb9	0	image	https://scontent-iad3-1.c
432960387	547505	When I think of baseba	picture	Nationals Parl	1571272566	85	15	B3sw0gGA8eS	0	image	https://scontent-iad3-1.c
432960389	547505	Hiiiiii there. It's been a	picture	Arlington, Vir	1571182906	155	20	B3qFzsNABoX	0	image	https://scontent-dfw5-2.
432960388	547505	Majorrrrr Target boot s	picture	Arlington, Vir	1571098854	98	15	B3nlfalgeuX	0	image	https://scontent-iad3-1.c
407830306	547505	Team cozy hoodies and	picture	Arlington, Vir	1568164062	93	14	B2QH0aqgWT	0	image	https://scontent-iad3-1.c
407830311	547505	Left or right, which wou	picture	Arlington, Vir	1568071988	146	38	B2NYM7NgsB	0	image	https://scontent-iad3-1.c
407830307	547505	Look what's baaaaack!	picture	Arlington, Vir	1567988234	295	37	B2K4c_3A7dK	0	image	https://scontent-iad3-1.c
465971968	1340199	Tellement beau ce cha	picture	Picomtal	1575103119	2588	2	B5e7BS-l7QW	0	image	https://scontent-iad3-1.c
465971967	1340199	Wherever you go beca	picture	Picomtal	1575102580	1617	0	B5e5_d6o_RZ	0	image	https://scontent-iad3-1.c
465971970	1340199	If it's doesn't open it'	picture	Provence-Alpe	1575102097	1678	0	B5e5Ekjl-Cu	0	image	https://scontent-iad3-1.c
465971969	1340199	Happy week end ! 🥰	picture	Provence-Alpe	1575101612	2462	0	B5e4JVvodZI	0	image	https://scontent-iad3-1.c

Puede observarse a simple vista que la variable **timestamp**, que nos indicaría la fecha y horario preciso del post, no está en un formato legible. Esta es una de las cuestiones que abordaremos al momento de hacer un análisis descriptivo sobre los posts de los influencers.

Otra variable relevante de este dataset es **type**. Esta variable es categórica y nos indica si el post en cuestión es una imagen o un video.

Para una explicación más detallada de las variables, ir al Apéndice al final de este trabajo.

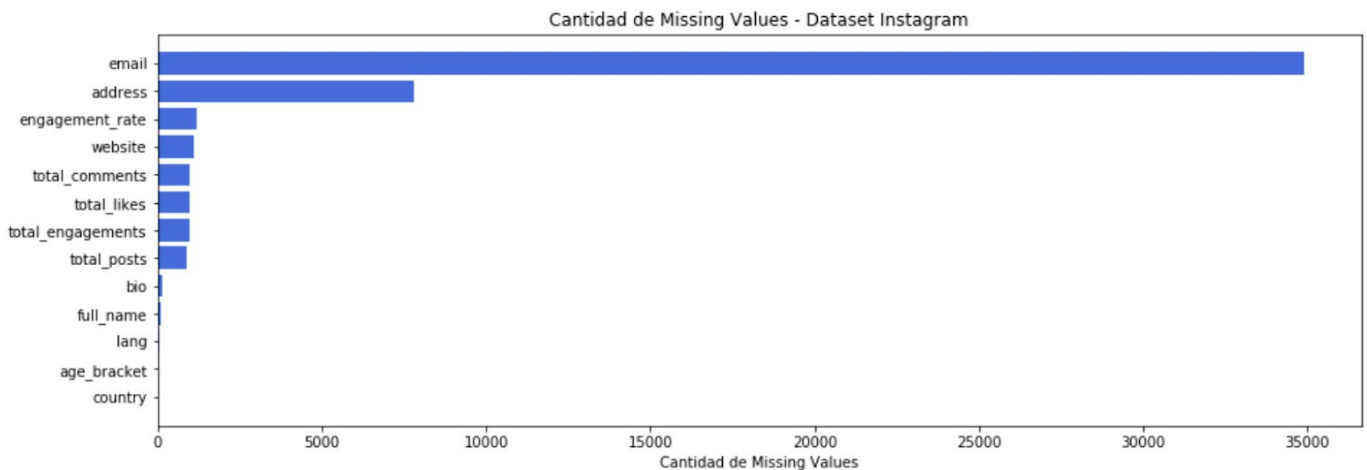
### 3. Análisis Exploratorio

En esta sección haremos un análisis descriptivo y exploratorio de cada dataset para poder obtener los primeros *insights* sobre las características de los influencers y sus posts. Parte de los resultados de este primer análisis nos llevará a realizar transformaciones en las variables para luego utilizarlas en la etapa de *clustering*.

#### 3.1 Análisis dataset *instagram*

Uno de los primeros pasos al momento de analizar datos, es hacer un conteo de los llamados valores faltantes o *missing values*. En el gráfico a continuación, puede observarse la cantidad de missing values que posee cada variable o feature de este dataset.

Figura 1: Cantidad de valores faltantes para cada variable en el dataset "instagram"



Salta a la vista que las variables 'email' y 'address' son las que exhiben una mayor cantidad de missing values. Por ende, decidimos eliminar estas variables del dataset y no incluirlas en el análisis.

Por otra parte, también identificamos algunas variables que creíamos que no serían informativas para el análisis exploratorio ni para el posterior algoritmo de clustering. Estas son 'website', 'full\_name', 'name' y 'country'. En primer lugar, las variables 'full\_name' y 'name' corresponden a nombres propios del influencer, por lo cual tomarán un valor diferente para cada uno de ellos, siendo muy poco probable que se detecte un patrón interesante y relacionado con las características que definen a los influencers actuales. Lo mismo ocurrirá con la variable 'website' la cual corresponde a una dirección de página web o URL, la cual no aporta información en sí misma sobre las características del influencer. No obstante, más adelante sí incluimos una variable que informa si el influencer tiene un sitio web junto con otras redes sociales.



En segundo lugar, la variable 'country' puede ser en principio interesante, debido a que informa sobre la nacionalidad de los influencers. Sin embargo, al hacer un análisis profundo de esta variable, detectamos que no estaba reportada de manera uniforme, contando con 538 valores únicos en total, siendo varios de estos referidos a un mismo país pero de diferentes formas. Por ejemplo, 'country' puede tomar el valor de 'United States', 'US', 'UnitedStates', 'Dallas, Texas, US', 'Florida Keys, FL, USA', etc. combinando no sólo el nombre del país sino también nombres de Estados dentro de Estados Unidos; esto ocurre también para otros países en la base de datos, como Australia, Canadá o Alemania. Lo cierto es que actualmente las redes sociales permiten un nivel de globalización que genera que no importe el lugar geográfico en el cual se ubique un influencer, pudiendo igualmente transmitir información constantemente a sus seguidores en todo el mundo. De esta forma, termina tomando mayor importancia el idioma en el cual se comunica el influencer a través de las redes, definiendo así un conjunto determinado de seguidores que puedan comprender sus publicaciones (por esto es que conservamos la variable 'lang' de idioma). Es así como decidimos no incurrir en un esfuerzo extra de ingeniería sobre 'country' para hacerla más uniforme y clara, dado que el país de origen actualmente no parece ser una variable relevante al momento de identificar características representativas y patrones de los SMIs. Por supuesto, esta elección fue hecha teniendo en cuenta que queremos realizar un análisis lo más general posible para que en un futuro las organizaciones o empresas puedan hacer uso de estas herramientas analíticas como gusten, incluyendo el país de origen o no, según sea relevante para la promoción de sus productos.

De esta manera, procedimos a eliminar 'website', 'full\_name', 'name' y 'country' de la base de datos.

A partir de este primer filtrado de variables, comenzamos a analizar aquellas que sí nos parecieron interesantes para poder entender las características de los influencers.

### 3.1.1 Análisis de variable 'lang'

Esta variable nos indica el idioma utilizado por el influencer en sus post. A partir de esta información, quisimos saber cuál era el idioma más utilizado por los influencers por lo cual construimos la siguiente tabla:

Tabla 3: Cantidad y Porcentaje de Influencers según el valor de la variable 'lang' (lenguaje)

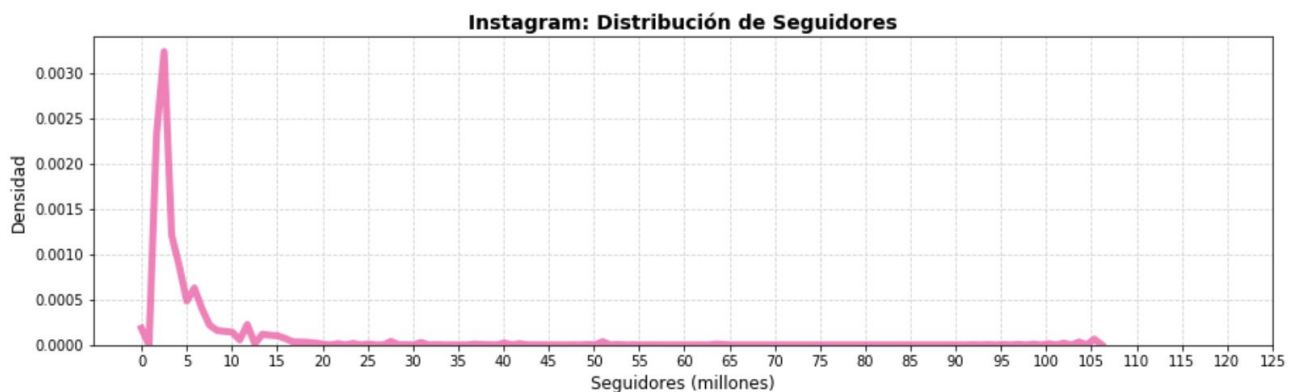
lang	en	pt	es	fr	de	it
Cantidad	90788.0	1240.00	1222.00	1159.00	983.00	831.00
(%) en Dataset	90.9	1.24	1.22	1.16	0.98	0.83

Puede notarse a simple vista que la primera fila reporta la cantidad de influencers que utilizan el idioma en cuestión y la segunda fila reporta el porcentaje que representa esta cantidad de influencers en el dataset total. Con esto en mente, podemos concluir que el 90.9% de los influencers de nuestra base de datos utiliza el idioma inglés, mientras que el restante 9.1% utiliza el portugués, español, italiano, francés, alemán, etc. Debido a que la variable 'lang' puede tomar 49 valores distintos (es una variable del tipo factor), esta tabla fue reducida y solo exponemos el resultado de los 6 idiomas más usados.

### 3.1.2 Distribución de seguidores

Cómo ya hemos mencionado, el dataset *instagram* posee variables como la cantidad de seguidores que tiene el influencer, la cantidad de *likes*, la cantidad de comentarios, etc. Consideramos interesante graficar la distribución de la cantidad de seguidores, para poder detectar la forma de esa distribución.

Figura 2: Distribución de Seguidores (en Millones)



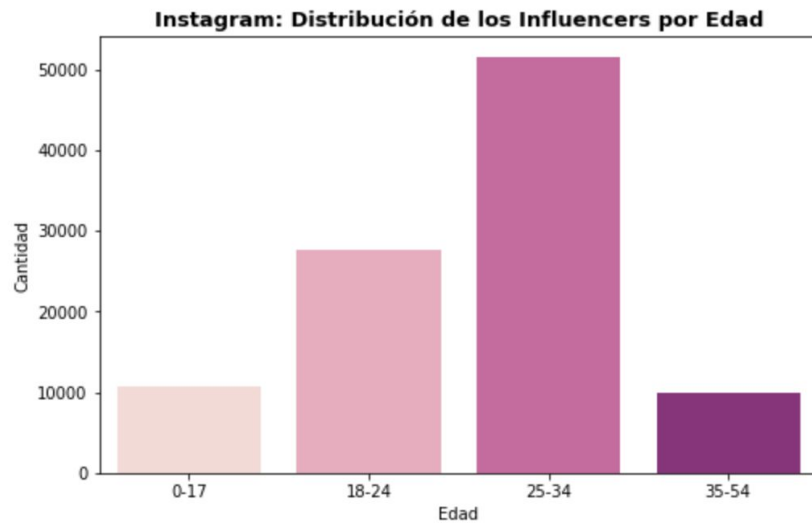
Es claro que esta distribución es asimétrica, es decir, que la mayoría de los influencers tienen entre 0 y 2 millones de seguidores, mientras que unos pocos pueden lograr tener más de 5, 10 o 15 millones. Esto nos está diciendo que contamos con un dataset que cuenta con pocos Mega-Influencers, que son aquellos que cuentan con más de 1 millón de seguidores.

Por otro lado, la asimetría de esta distribución nos da el indicio de que seguramente debemos transformar esta variable al momento de correr el algoritmo de clustering ya que, al basarse este en medidas de distancia, es necesario contar con variables simétricas y con la misma unidad de medida.

### 3.1.3 Distribución de edad

Con el objetivo de entender cuáles son los rangos de edad más “populares”, decidimos graficar la distribución del dataset (cantidad de observaciones/influencers) en función de los rangos de edad, así podíamos detectar la edad y generación de los influencers.

Figura 3: Histograma de los Influencers por rango etario

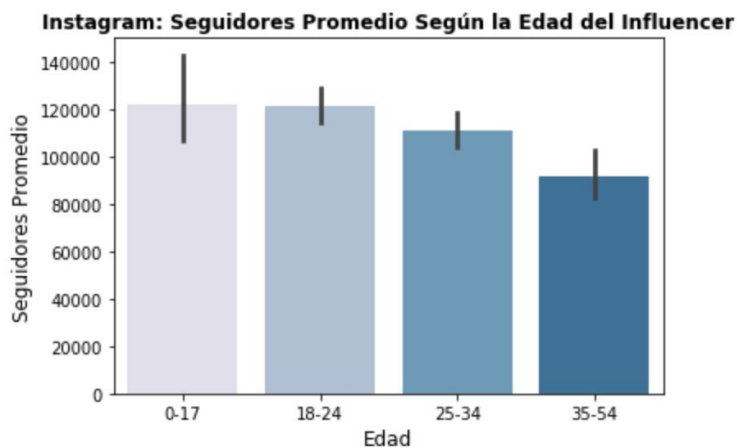


Es intuitivo haber encontrado que la mayoría de los influencers en el dataset tienen entre 25 y 34 años. Es ampliamente sabido que la red social Instagram fue creada recientemente (hace 10 años), e involucra una relación constante con la última tecnología (celulares con pantalla táctil y alta resolución de cámara). Por ende, es natural que individuos jóvenes sean más hábiles al momento de utilizar la plataforma y convertirse en influencers. En términos generacionales, tenemos en su mayoría influencers que son *millennials* o *centennials*.

### 3.1.4 Seguidores promedio por edad

Si analizamos la cantidad promedio de seguidores según la edad de los influencers, obtenemos el siguiente gráfico:

Figura 4: Cantidad de Seguidores Promedio por rango etario

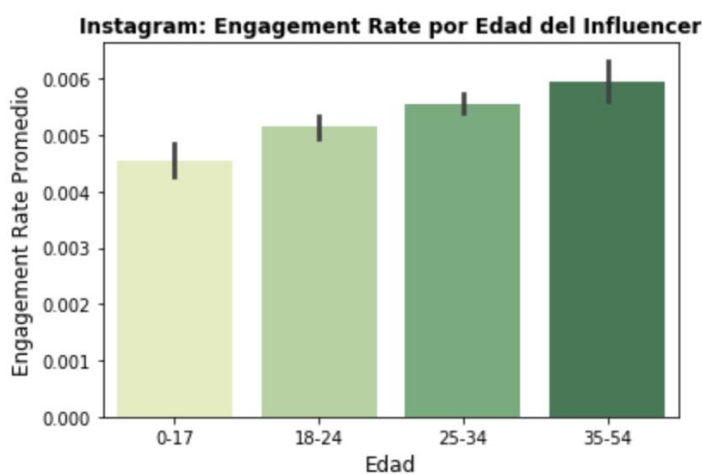


Es muy interesante e intuitiva la conclusión detrás: a mayor edad del influencer, menor es la cantidad promedio de seguidores. Esto es predecible dado lo encontrado anteriormente: los influencers en su mayoría son *centennials* o *millennials*, por lo cual publicarán contenido acorde con su edad y sus intereses. Esto implica que atraerán seguidores que probablemente sean de la misma generación, la cual, en este caso, coincide con ser la generación que más facilidad tiene al manejar la tecnología y las redes sociales. No obstante, vemos que los influencers de hasta 17 años tienen bastante varianza en la cantidad de seguidores: hay algunos que tienen muchos seguidores y otros que tienen muy pocos, dando una señal de que hay niveles muy distintos de popularidad para este rango etario.

### 3.1.5 Engagement rate por edad

Siguiendo con el análisis por rango etario, a continuación exhibimos un gráfico que nos muestra el 'engagement rate' promedio por edad del influencer:

Figura 5: Engagement Rate Promedio por rango etario



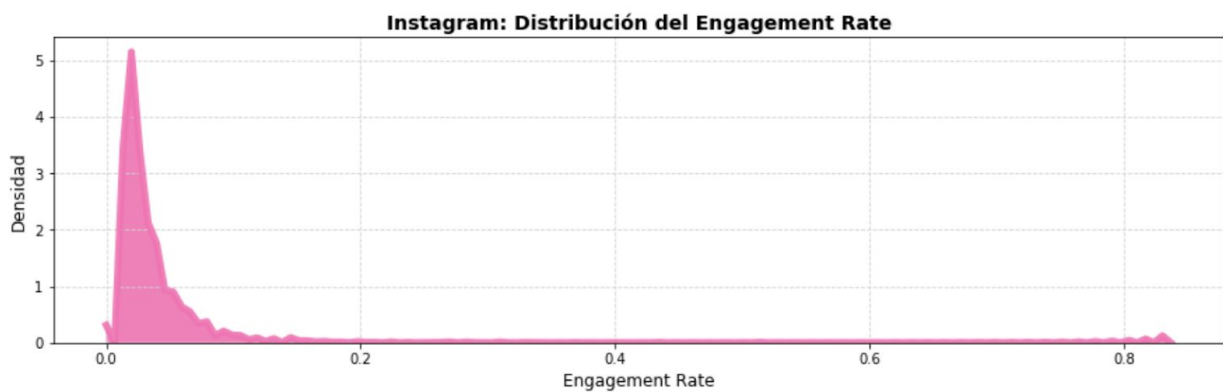
Lo que observamos es que el mayor 'engagement rate' promedio se da en el rango etario 35-54, lo cual nos está diciendo que los influencers que se encuentran dentro de ese rango tienen una mejor medida de performance en promedio con respecto a su visibilidad en Instagram. En algún punto, la conclusión podría estar relacionada con la 'fidelidad' de los seguidores: los influencers de mayor edad tienen menos seguidores (probablemente de la misma generación) pero estos suelen ser más 'fieles' dando siempre un *like* a cada publicación o comentando siempre en cada post. Por ende, el mayor 'engagement rate' puede deberse a una menor cantidad de seguidores que son más fieles y dejan sus *likes* y comentarios en todas las publicaciones (que seguro son menos que las de influencers más jóvenes).

### 3.1.6 Distribución del Engagement rate

Ahora veremos cómo se distribuye la variable 'engagement rate'. Recordemos que esta variable se construyó a partir de los valores de las variables 'total likes', 'total comments', 'total posts' y 'followers', por lo cual es de esperar que todas estas variables tengan una distribución similar a la de 'engagement rate'. De hecho, ya vimos que la distribución de la variable 'followers' era bastante asimétrica.

La forma de la distribución es, efectivamente, similar a la distribución de los seguidores. En particular, la asimetría nos está diciendo que la mayoría de los influencers tienen un bajo 'engagement rate'.

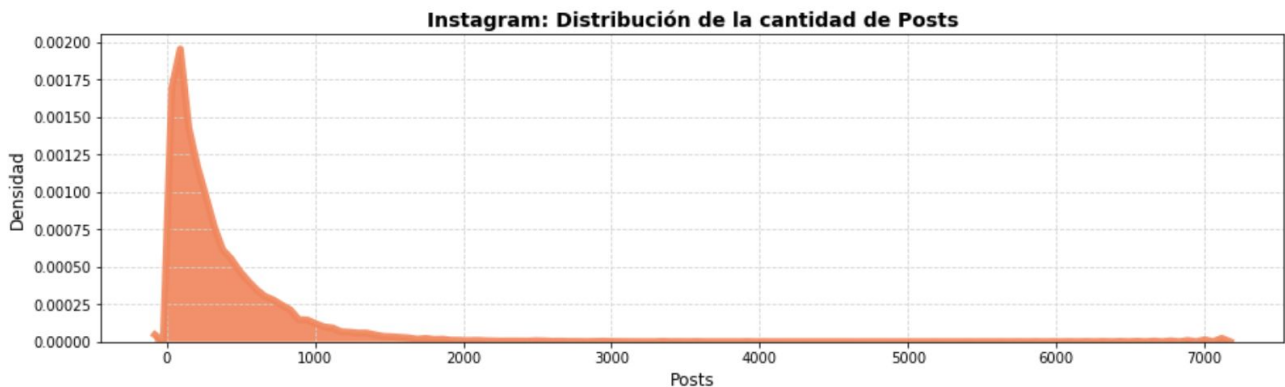
Figura 6: Distribución del Engagement Rate



### Distribución de la cantidad de publicaciones

Al graficar la distribución de la cantidad de posts, podemos ver que la mayoría de los influencers tienen entre 0 y 1000 publicaciones, con un pico o máximo en 100 publicaciones.

Figura 7: Distribución de la cantidad de publicaciones de los influencers



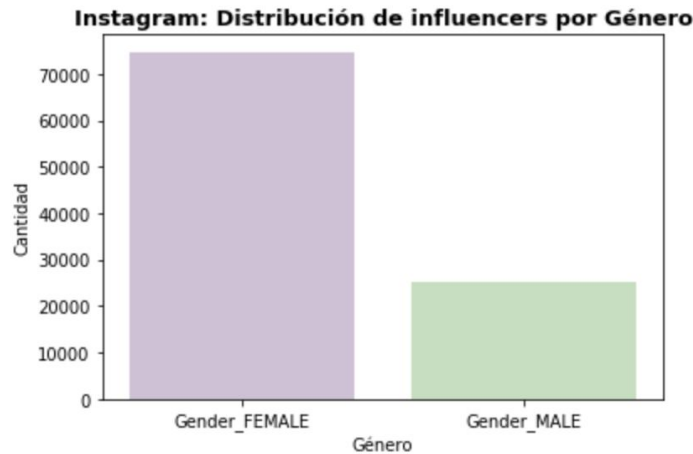
Es importante aclarar que la cantidad de publicaciones de cada influencer es representativa del período para el cual tenemos datos, es decir, comprenden la cantidad de posts realizados en tres meses. Por lo cual, para obtener el número total de publicaciones del influencer, debería multiplicarse la cantidad de publicaciones por alguna medida de antigüedad del influencer en

Instagram. A su vez vemos que se confirma lo que dijimos antes: la distribución es asimétrica. Es de esperarse que con la cantidad de *likes* y comentarios, ocurra lo mismo.

### 3.1.7 Cantidad de influencers por género

Otra variable importante a analizar es 'gender'. En particular, podemos analizar cuántas observaciones en el dataset corresponden a mujeres y cuántas a hombres.

Figura 8: Distribución de las observaciones por Género

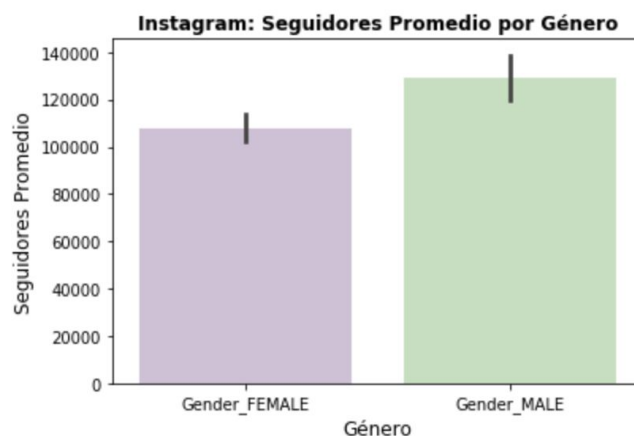


Es claro que el dataset presenta una muestra desbalanceada, en donde tenemos mayoría de influencers siendo de género femenino. Esto en parte es lógico, debido a que algunas de las categorías ('fashion', 'beauty') son casi exclusivamente de mujeres y, por otro lado, se ha registrado en otros trabajos de investigación (Schouten et al. 2020) que la mayoría de los contenidos relacionados con categorías como 'fashion', 'beauty', 'food' y 'fitness' son generados por mujeres.

### 3.1.8 Cantidad de seguidores promedio por género

Cuando nos concentramos en la cantidad de seguidores promedio por género, encontramos que los influencers masculinos tienen un promedio más alto con respecto a las mujeres.

Figura 9: Cantidad de Seguidores Promedio por Género

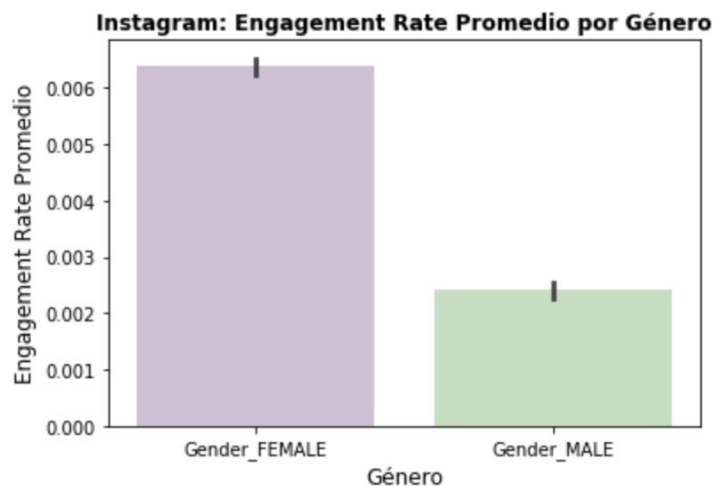


Una explicación posible a este resultado es que los influencers masculinos suelen hacerse primero 'famosos' en otras plataformas como YouTube o Twitch, o en otro medios (por ejemplo, pueden ser futbolistas, periodistas, modelos, etc.), y luego trasladan toda esa masa de seguidores a Instagram. Por ende, al llegar a esta plataforma, ya tienen una masa de seguidores establecida.

### 3.1.9 Engagement rate promedio por género

A pesar del resultado anterior, al analizar el 'engagement rate' por género, observamos que las influencers mujeres tienen un mayor 'engagement rate' promedio con respecto a los hombres. En parte, esto puede estar diciéndonos que las mujeres tienen seguidores más fieles que están mucho más atentos a las publicaciones y que reaccionan ante estas comentando o poniendo un *like*. Es decir que el porcentaje de seguidores que en promedio siempre reaccionan a las publicaciones de estas influencers es mayor en comparación con los influencers hombres.

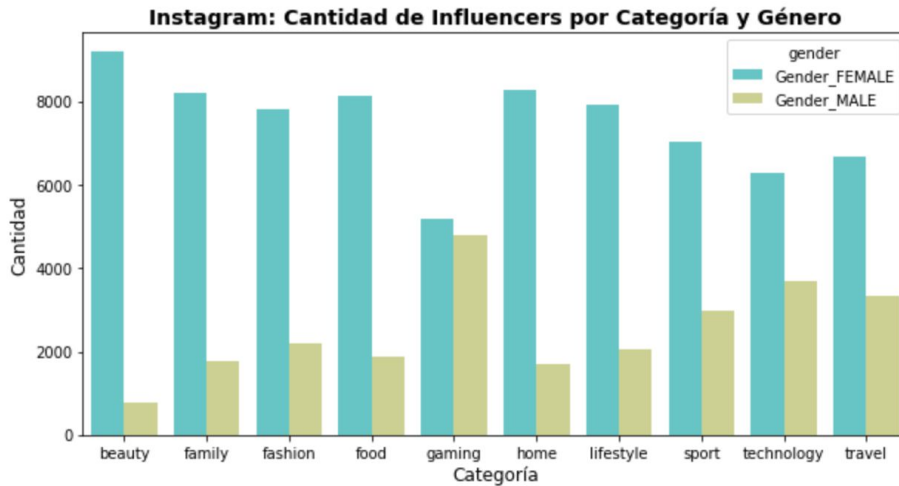
Figura 10: Engagement Rate Promedio por Género



### 3.1.10 Distribución de influencers por género y categoría

Por último, podemos profundizar la distribución por género al dividirla por categoría:

Figura 11: Cantidad de observaciones por Categoría y Género

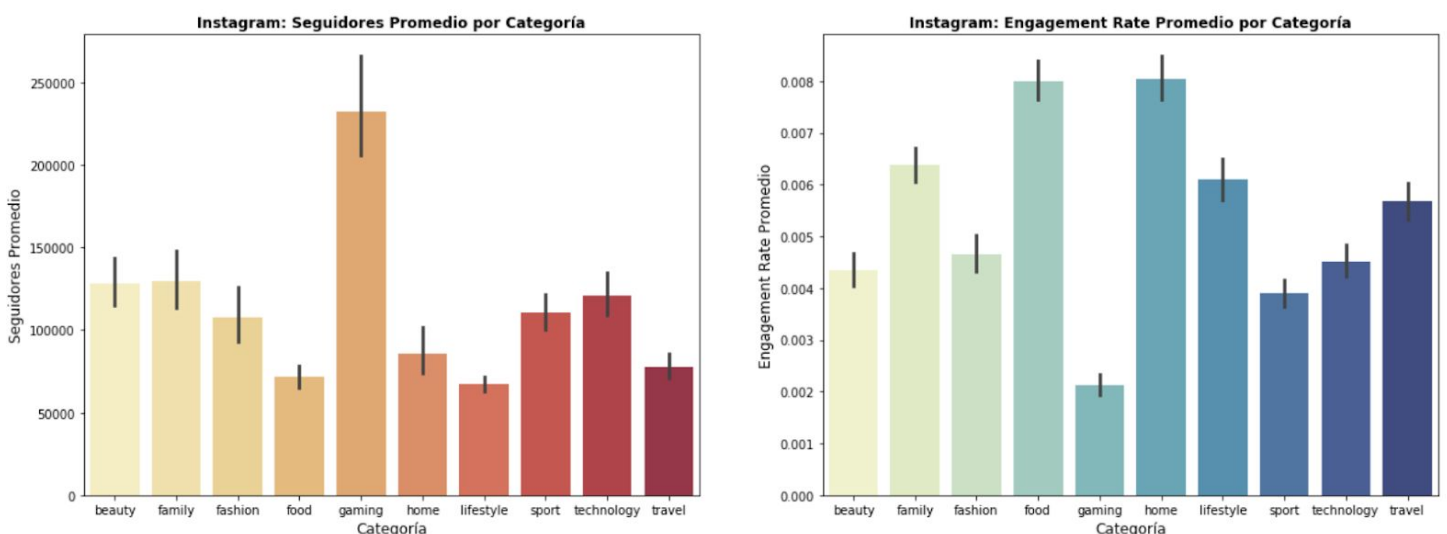


Aquí volvemos a observar la mayoría de mujeres influencers en el dataset. No obstante, vemos que la diferencia entre mujeres y hombres es mucho mayor para categorías como 'beauty', 'family', 'home', 'lifestyle', 'food' y 'fashion'. Luego, para las categorías 'travel', 'technology', 'sport' la diferencia es un poco menor, siendo 'gaming' la categoría más equilibrada de todas. Esto tiene sentido, ya que 'gaming' es una categoría que aún hoy suele ser dominada por hombres *gamers*.

### 3.1.11 Distribución de seguidores y Engagement rate promedio por categoría

Si ahora analizamos cómo se comportan el promedio de seguidores y el 'engagement rate' por categoría, notamos que la categoría 'gaming' es la que tiene mayor promedio de seguidores, pero con alta varianza, mientras que las categorías 'food' y 'home' son las que reportan mayor 'engagement rate'.

Figura 12: Engagement Rate y Seguidores Promedio por Categoría





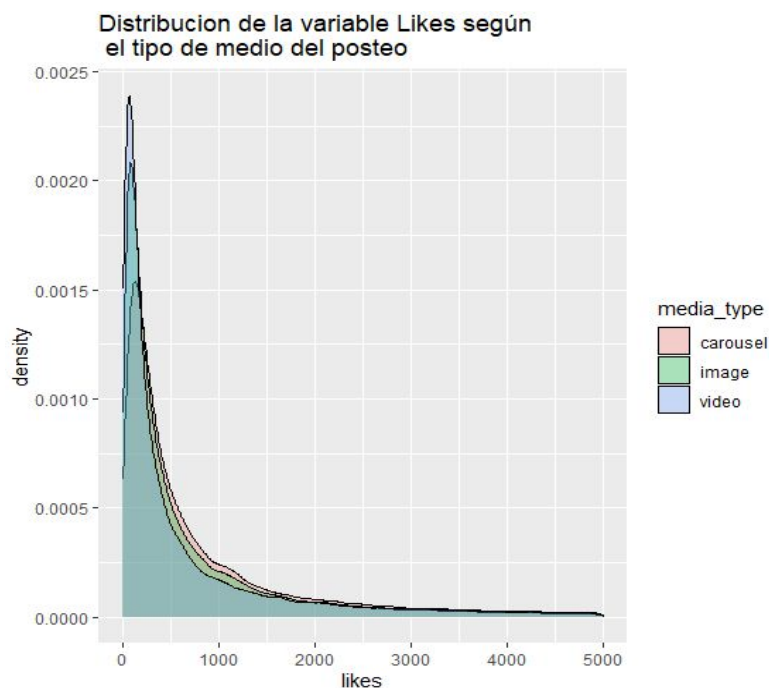
## 3.2 Análisis dataset *instagram\_posts*

Este dataset contenía información sobre cada post, cuyas variables más relevantes concluimos que eran 'comments' (comentarios) y 'likes'. También se tenía información sobre el día del posteo de la imagen y el tipo de publicación.

### 3.2.1 Distribución de la variable *likes* según el tipo de publicación

Pensamos que la distribución de *likes* entre imágenes, videos y carousel podría ser diferente ya que las personas no reaccionan de la misma manera a los distintos tipos de publicación. Por esto es que hicimos este gráfico que separa la distribución de *likes* entre los tres tipos de publicación.

Figura 13: Distribución de *likes* según el tipo de publicación (imagen, video, carousel)

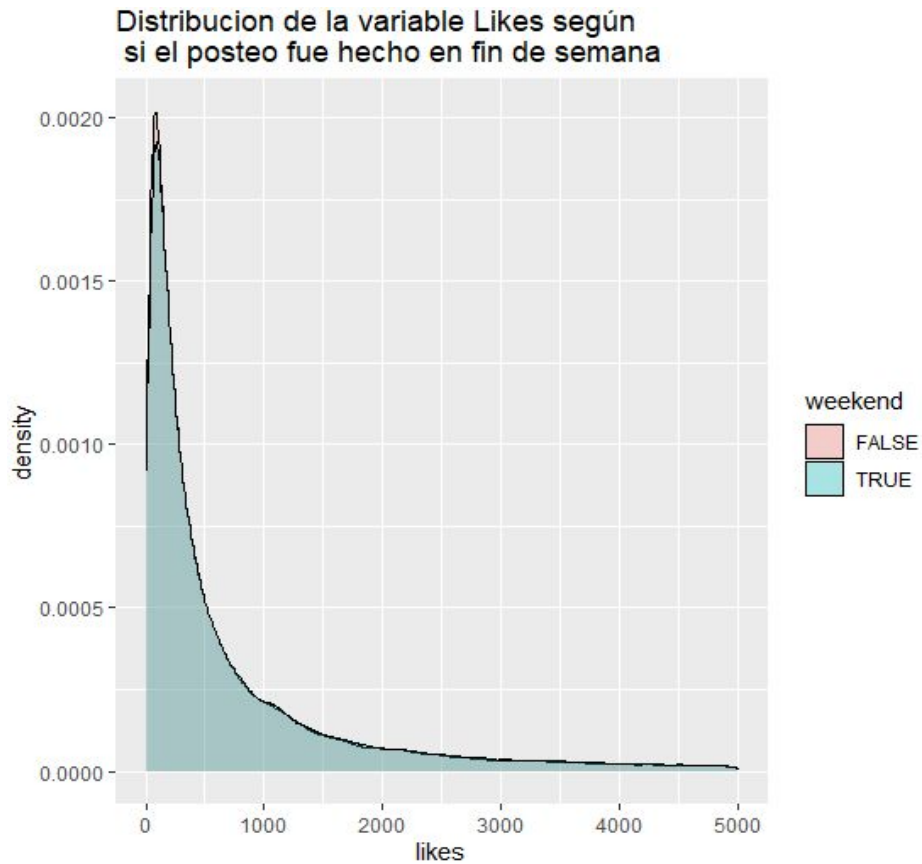


Para este gráfico se utilizaron las variables 'likes' y 'media\_type' (tipo de publicación). Podemos ver que los videos tienen la menor varianza de cantidad de *likes*, los carousel (posteos con un conjunto de imágenes para ver) tienen la mayor varianza y las imágenes están en el medio. Es decir, los *likes* están más concentrados en su media para los videos y menos concentrados para los carousel. Los promedios de *likes* en imágenes y videos parecen ser muy parecidos, mientras que el promedio para carousel parece estar ligeramente más a la derecha. Esto puede ser así ya que es común que la gente no vea todas las imágenes incluidas en un carousel, y al reducir la atención de la persona hacia la publicación es menos probable que le ponga *like* o interactúe con esta ya que no se siente inmerso, logrando que la distribución de *likes* hacia carouseles sea más variable. Por otro lado, tanto videos como imágenes al ser un solo elemento al cual prestarle atención, la persona sí le presta mayor atención y llega a interactuar.

### 3.2.2 Distribución de la variable likes según si se publicó un fin de semana o no

También nos pareció interesante averiguar si la distribución de *likes* era distinta según el tipo del día en cuál fue hecha la publicación (fin de semana o día de semana).

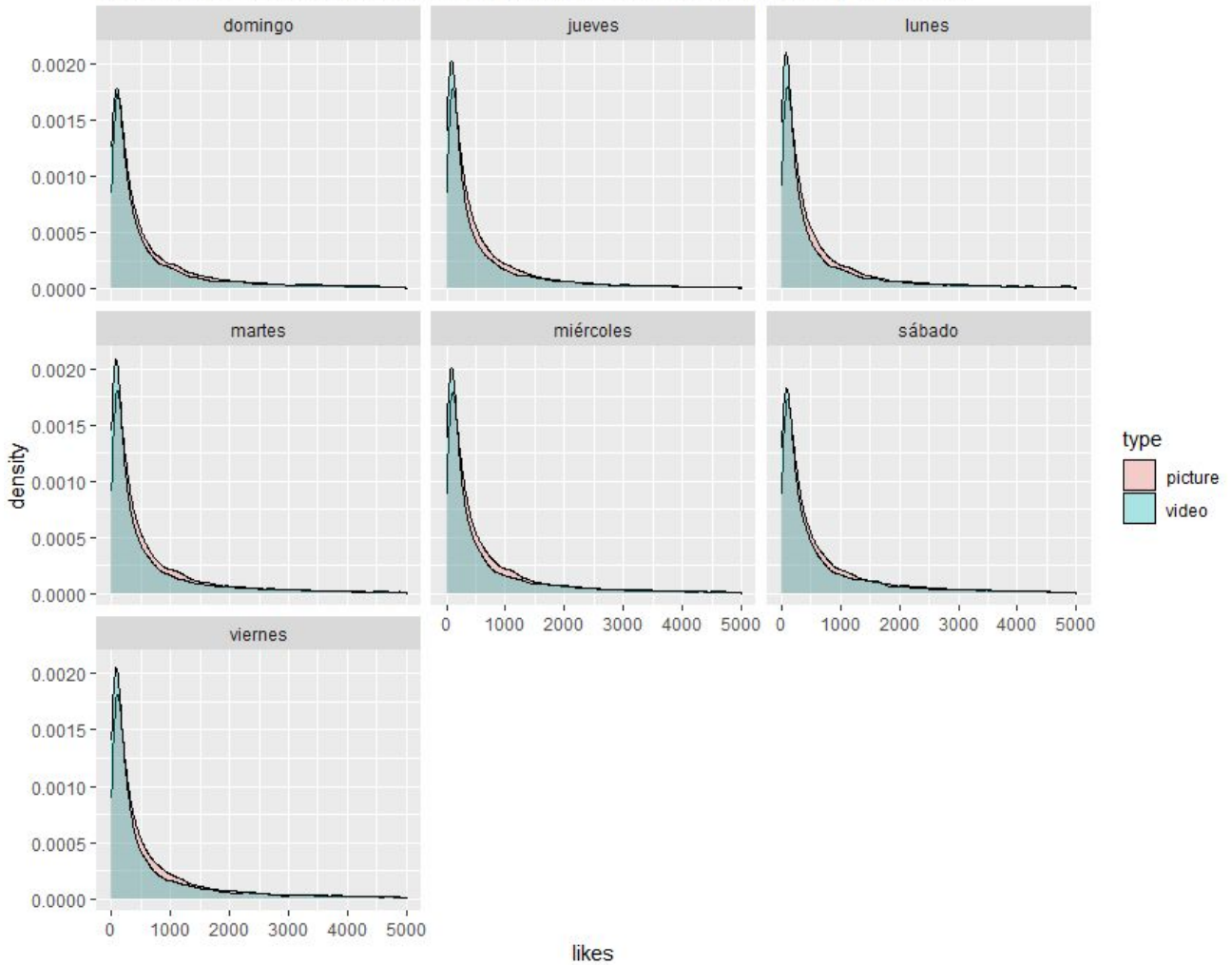
Figura 14: Distribución de *Likes* según si la publicación fue hecha durante el fin de semana



Para este gráfico se utilizaron las variables 'likes' y 'weekend' (variable creada a partir del dataset que es igual a TRUE si el posteo fue hecho en un fin de semana y FALSE si fue en un día de semana). No hay una diferencia apreciable entre la distribución de los *likes* a posts publicados durante el fin de semana y en la semana. Esto puede ser porque, si bien los usuarios tienen más tiempo para ver publicaciones durante el fin de semana, suelen dar *like* a las publicaciones en cuanto las ven, por más que hayan sido publicadas durante la semana. Es decir, cuando llegue el fin de semana y los usuarios tengan más tiempo libre aprovecharán para ver todas las publicaciones que se hayan perdido durante la semana e interactuarán con las que crean necesario/deseen, resultando en que el día de publicación no le afecta en nada.

### 3.2.3 Distribución de la variable likes según el tipo de publicación separado por día

Figura 15: Distribución de la variable *Likes* por día de la semana y tipo de publicación

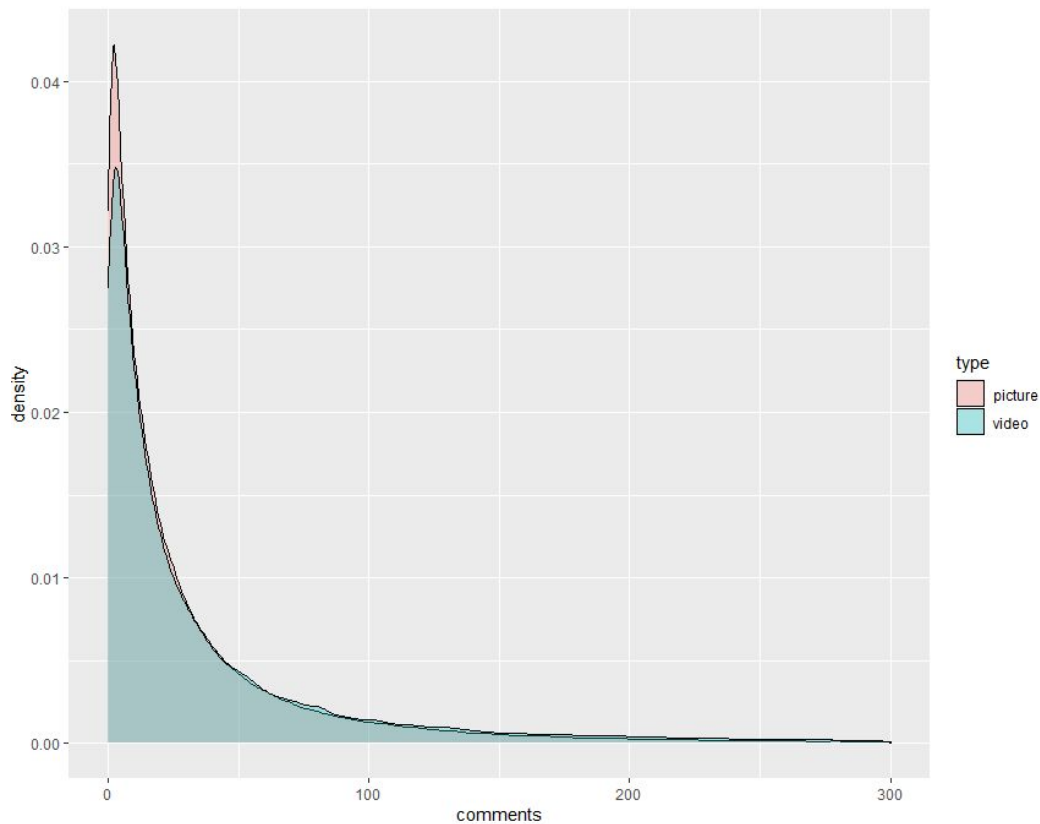


Para este gráfico se utilizaron las variables 'likes', 'weekday' (día de la semana en que se hizo la publicación) y 'type' (tipo de publicación: imagen o video). Se puede ver en estos gráficos que las distribuciones de *likes* de publicaciones hechas los sábados y los domingos es igual para videos y para fotos, mientras que para el resto de la semana la distribución de *likes* en los videos tiene menor varianza que la de las fotos. Las medias de *likes* tanto para videos como para fotos parece ser la misma para todos los días. La distribución de *likes* para fotos parece ser la misma para todos los días, y esto puede ser resultado del insight descrito en el gráfico anterior. En cambio, los *likes* parecen acumularse más alrededor de su media para los videos publicados en un día de semana.

### 3.2.4 Distribución de los comentarios según el tipo de publicación

Otra variable que era necesario analizar era comentarios.

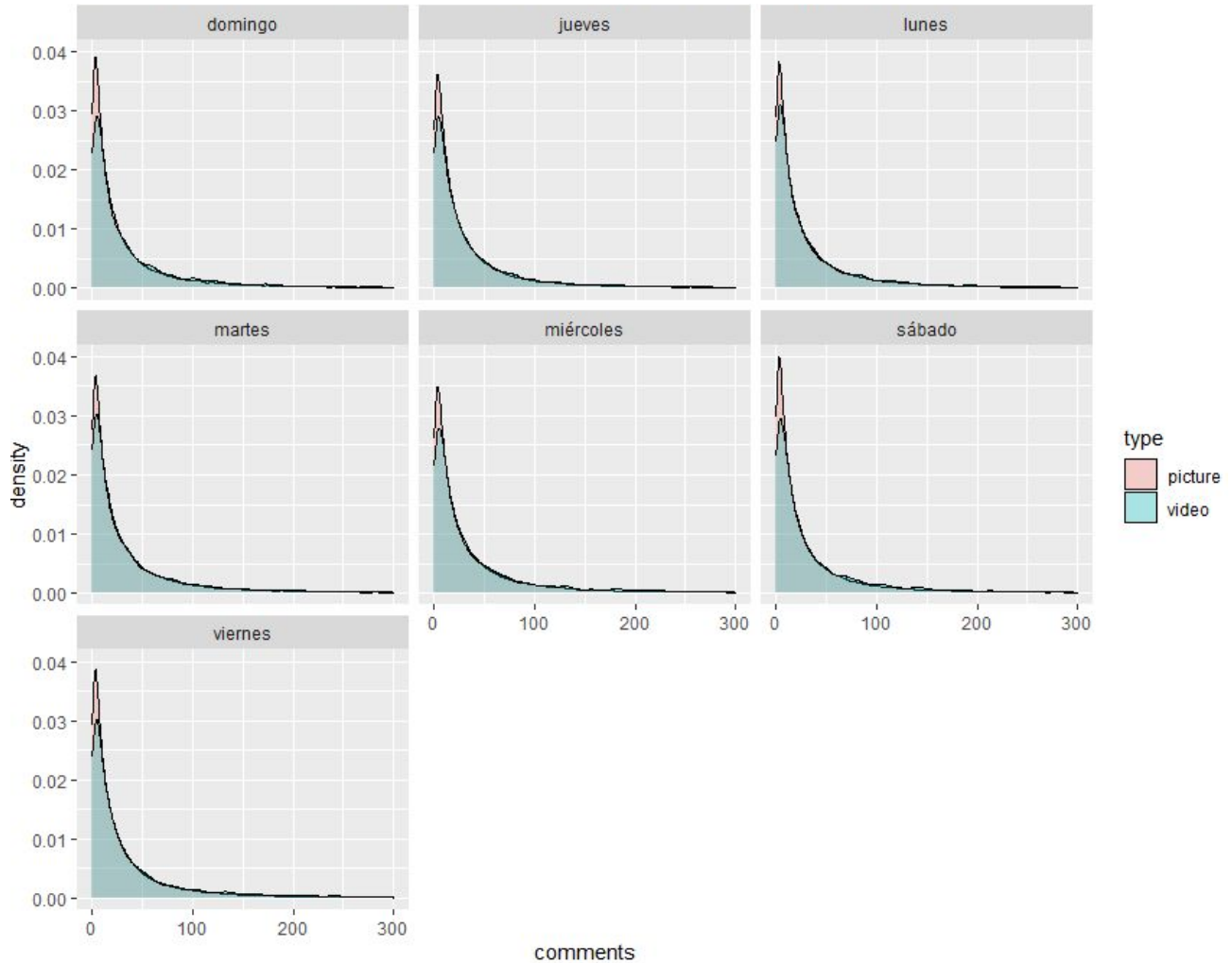
Figura 16: Distribución de comentarios por tipo de publicación (imagen o video)



Para este gráfico se utilizaron las variables 'comments' (comentarios) y 'type' (tipo de publicación). Si hacemos el gráfico de la distribución de comentarios discriminado entre fotos y videos, podemos observar que hay mayor concentración en la media en el caso de imágenes. Esto es así ya que es más probable que la gente comente a una foto y que haya una cantidad menos variable de seguidores que siempre comentan. En cambio, hay menor proporción de los seguidores de un influencer que ven los videos ya que muchos los pasan por alto.

### 3.2.5 Distribución de la variable comentarios según el tipo de publicación separado por día

Figura 17: Distribución de comentarios según tipo de publicación y día de la semana

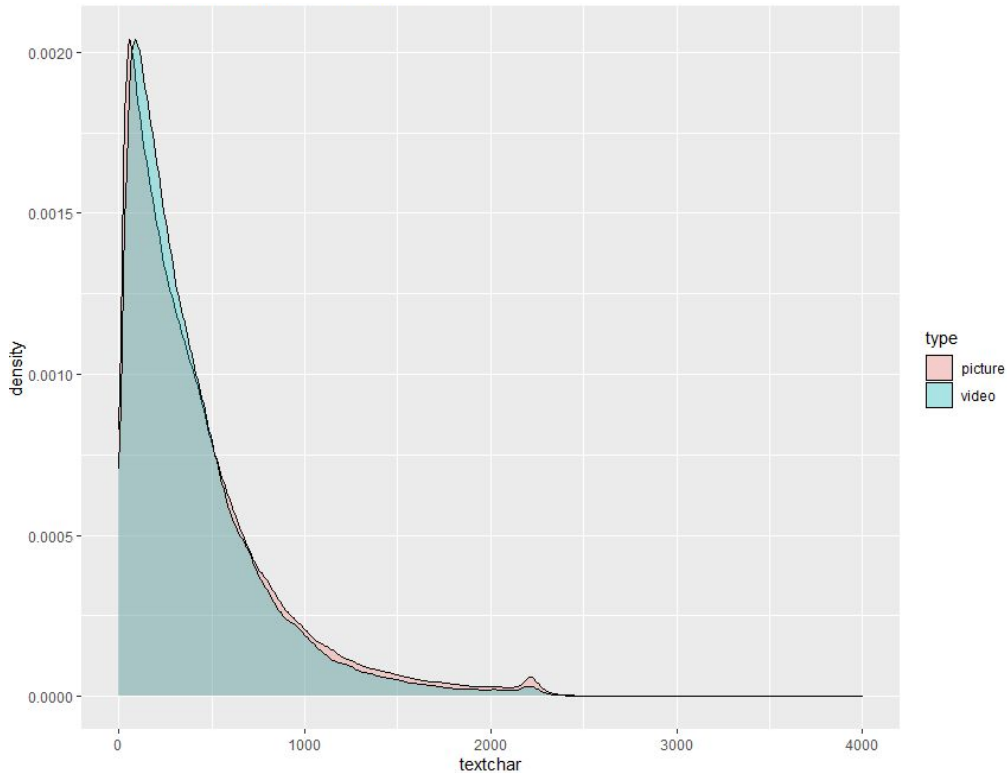


En este gráfico fueron utilizadas las variables 'comments' (comentarios), 'weekday' (día de la semana) y 'type' (tipo de publicación). Los resultados que se pueden apreciar son que para imágenes, la distribución de comentarios se concentra más alrededor de su media en los días viernes, sábado y domingo. Esto puede resultar de esta manera ya que, al contrario de los *likes* que se dan indiscriminadamente a las publicaciones de interés en cuanto se ven (de manera instantánea), los comentarios requieren de más tiempo, por lo que es razonable que solo se comente en publicaciones hechas en el mismo día en que se vieron.

### 3.2.6 Distribución de la variable descripción según tipo de publicación

Cada publicación contiene una descripción hecha por el influencer. La variable descripción fue convertida a cantidad de caracteres para poder estudiar si hay alguna diferencia entre la distribución de caracteres entre distintos tipos de publicaciones, para poder analizar el comportamiento del influencer.

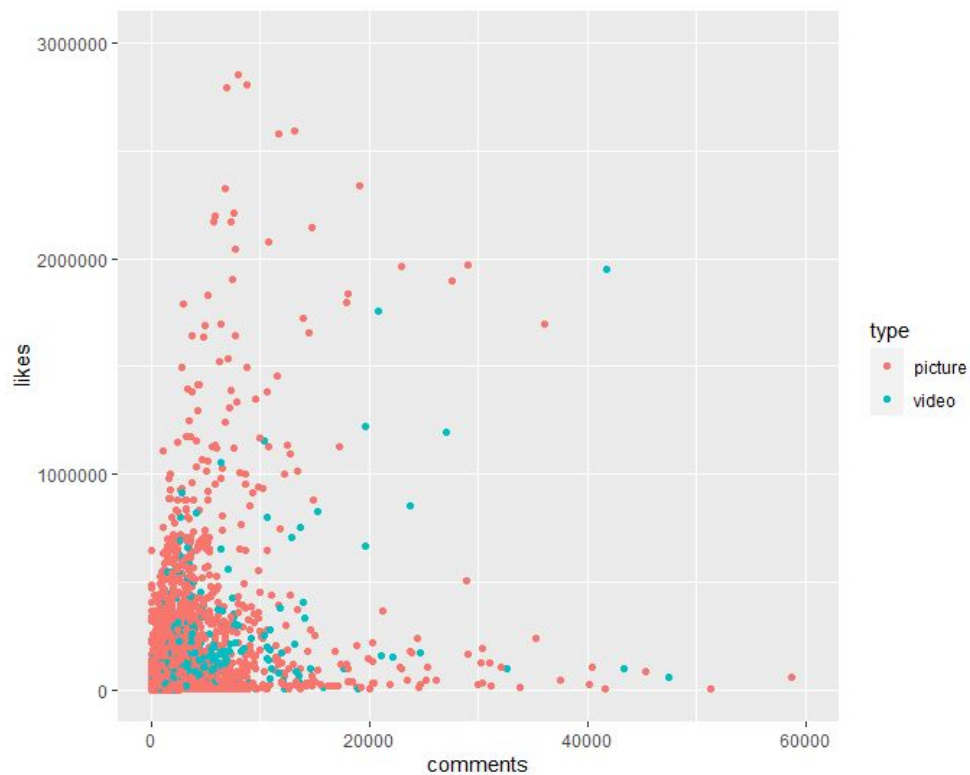
Figura 18: Distribución de la cantidad de caracteres en la descripción de la publicación por tipo de publicación (imagen o video)



Para este gráfico fueron utilizadas las variables 'textchar' (cantidad de caracteres incluidos en la descripción de la publicación - variable 'text') y 'type' (tipo de publicación). La media de la cantidad de caracteres usados en las publicaciones del tipo video es ligeramente mayor que en imágenes. Esto puede ser así ya que los videos que suben los influencers normalmente son tutoriales, "challenges" (eventos o comportamientos grabados para que el público los imite), campañas, etc, que requieren de mayor explicación para que el público entienda el contexto. En cambio, la mayoría de las imágenes/fotos no requieren de una descripción muy extensa porque son más "autocontenidas" y no requieren de tanto contexto para que el público la aprecie.

### 3.2.6 Relación entre comentarios y likes

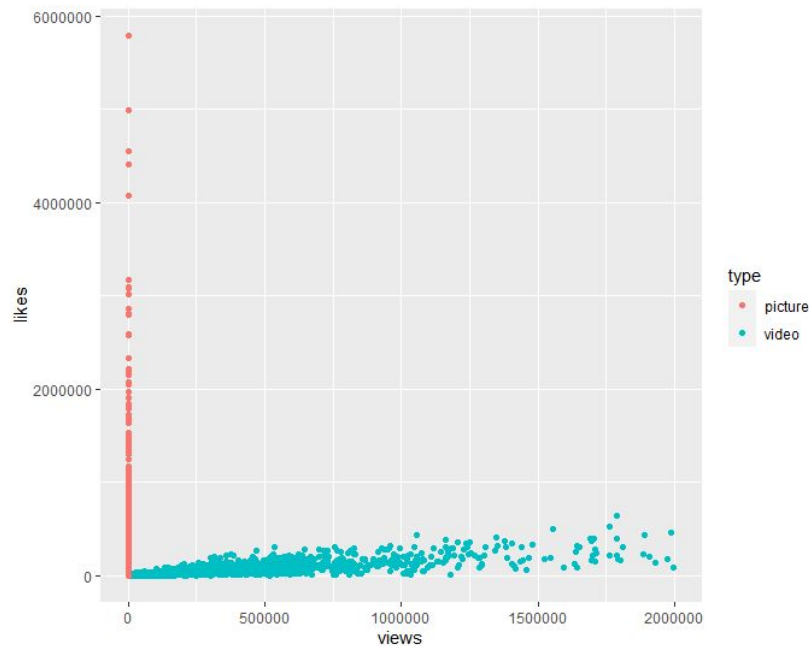
Figura 19: Scatter Plot entre los comentarios y los *likes*



Para hacer este gráfico se utilizaron las variables 'comments' (comentarios), 'likes' y 'type' (tipo de publicación). Se hubiera esperado que haya una clara correlación positiva entre comentarios y *likes*, ya que ambas variables están impulsadas por tener mayor cantidad de seguidores. En cambio, si bien hay cierta tendencia positiva, no es tan marcada como esperábamos, concluyendo en que estas dos variables son más independientes entre sí de lo que pensamos en un principio. Es decir, habrá influencers con un público que pone *likes* a sus publicaciones pero no comenta, otros que solo comentan sus fotos pero sin poner *like* y una combinación de las dos.

### 3.2.7 Relación entre visitas y likes

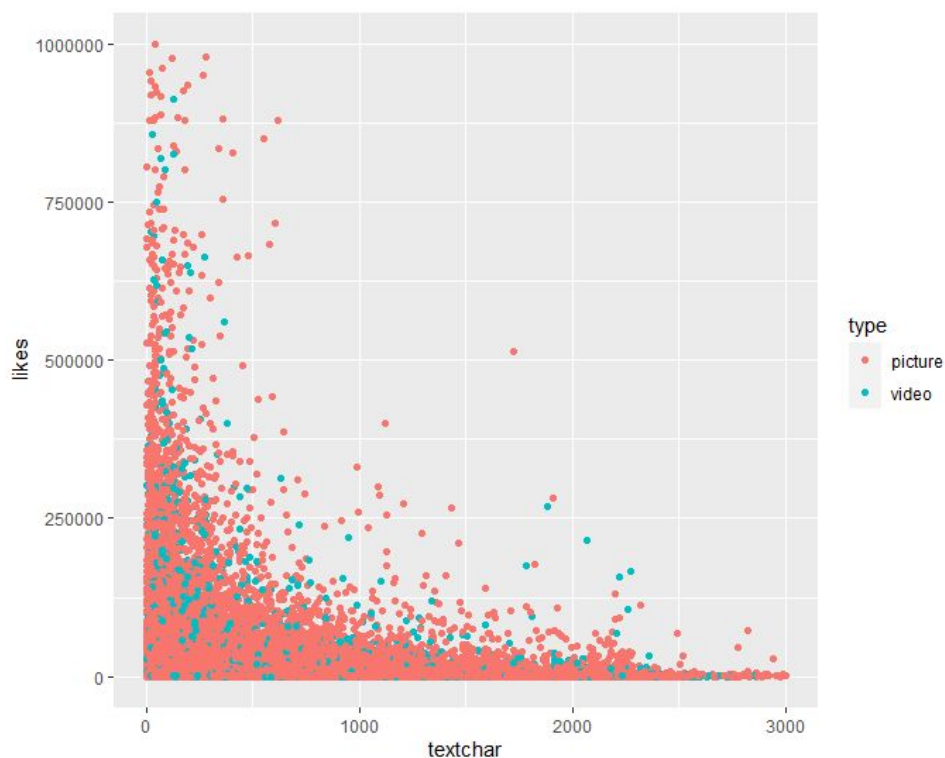
Figura 20: Scatter Plot entre visitas y *likes*



Para este gráfico se utilizaron las variables 'views' (visitas), 'likes' y 'type' (tipo de publicación). Podemos ver una clara correlación positiva entre 'views' y 'likes' para los videos, lo cual tiene sentido ya que para poner *like* en un video es sensato haberlo visto antes. En cambio, vemos que la variable 'views' es siempre 0 para imágenes, lo que nos indica que este tipo de publicación no acumula 'views', siendo esta una variable exclusiva para publicaciones del tipo video.

### 3.2.8 Relación entre cantidad de caracteres en la descripción y likes

Figura 21: Scatter Plot entre la cantidad de caracteres en la descripción de la publicación y *likes*

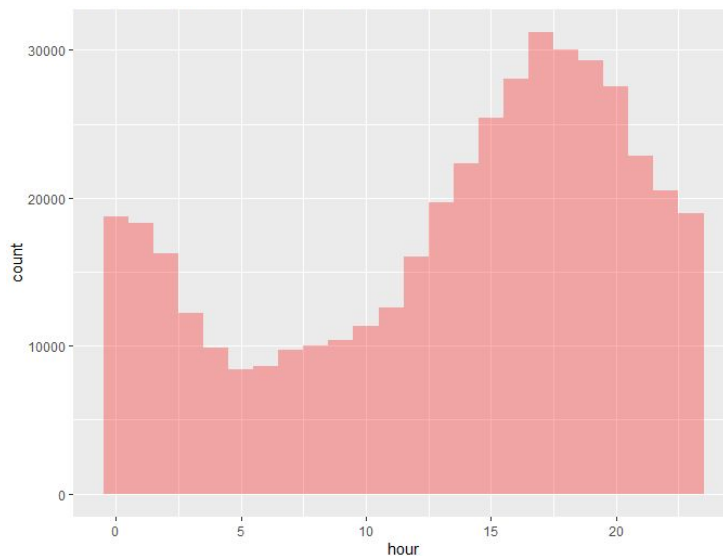




Nos pareció interesante averiguar si había alguna relación entre la cantidad de caracteres incluidos en la descripción de la publicación y los *likes* que esta recibía. Sin embargo, tras hacer el gráfico de dispersión entre esas dos variables, discriminado por tipo de publicación, no encontramos ninguna correlación significativa. Esto quiere decir que manipular la longitud de la descripción de una publicación no genera un cambio apreciable en la cantidad de *likes* que recibirá.

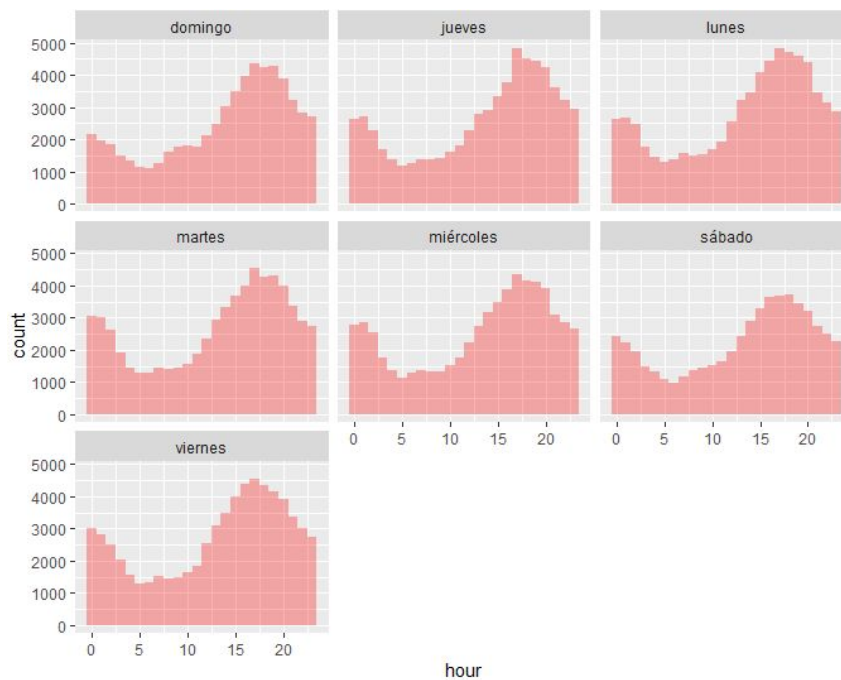
### 3.2.9 Histograma de cantidad de publicaciones por hora del día

Figura 22: Distribución de la cantidad de publicaciones por hora del día



Este histograma nos muestra que la hora más popular para publicar es a las 17 hs. Las horas con mayor cantidad de publicaciones son entre las 17 hs y las 19 hs, y aquellas con menor cantidad de publicaciones son entre las 4 hs y 11 hs. Las horas más populares coinciden con el período del día donde la mayoría de las personas terminan su jornada laboral o de clases, teniendo más tiempo de ver redes sociales. Por lo cual los influencers intentan programar sus publicaciones para que caigan dentro de este rango horario dado que es más fácil viralizarse y conseguir más visitas y *likes*.

Figura 23: Distribución de la cantidad de publicaciones por hora del día y por día de la semana



A su vez, las horas populares no parecen cambiar significativamente según el día de la semana.

### 3.2.10 Hashtags y Arrobas más utilizados por categoría

Para poder analizar cuáles hashtags y arrobas (forma de etiquetar una cuenta en Instagram) eran utilizados por cada categoría, simplemente agregamos la categoría de cada influencer como columna en el dataset *instagram\_posts* haciendo uso de la clave primaria en común entre los dos datasets ('instagram\_id'). De esta forma, realizamos un histograma basado en los hashtags y arrobas, de manera que contamos la cantidad de posts que mencionan cada uno de ellos. A continuación, debido a que son 10 categorías, exponemos los gráficos correspondientes a 'fashion' y 'beauty'.

Figura 24: Frecuencia de *Hashtags* para las categorías "Beauty" y "Fashion"

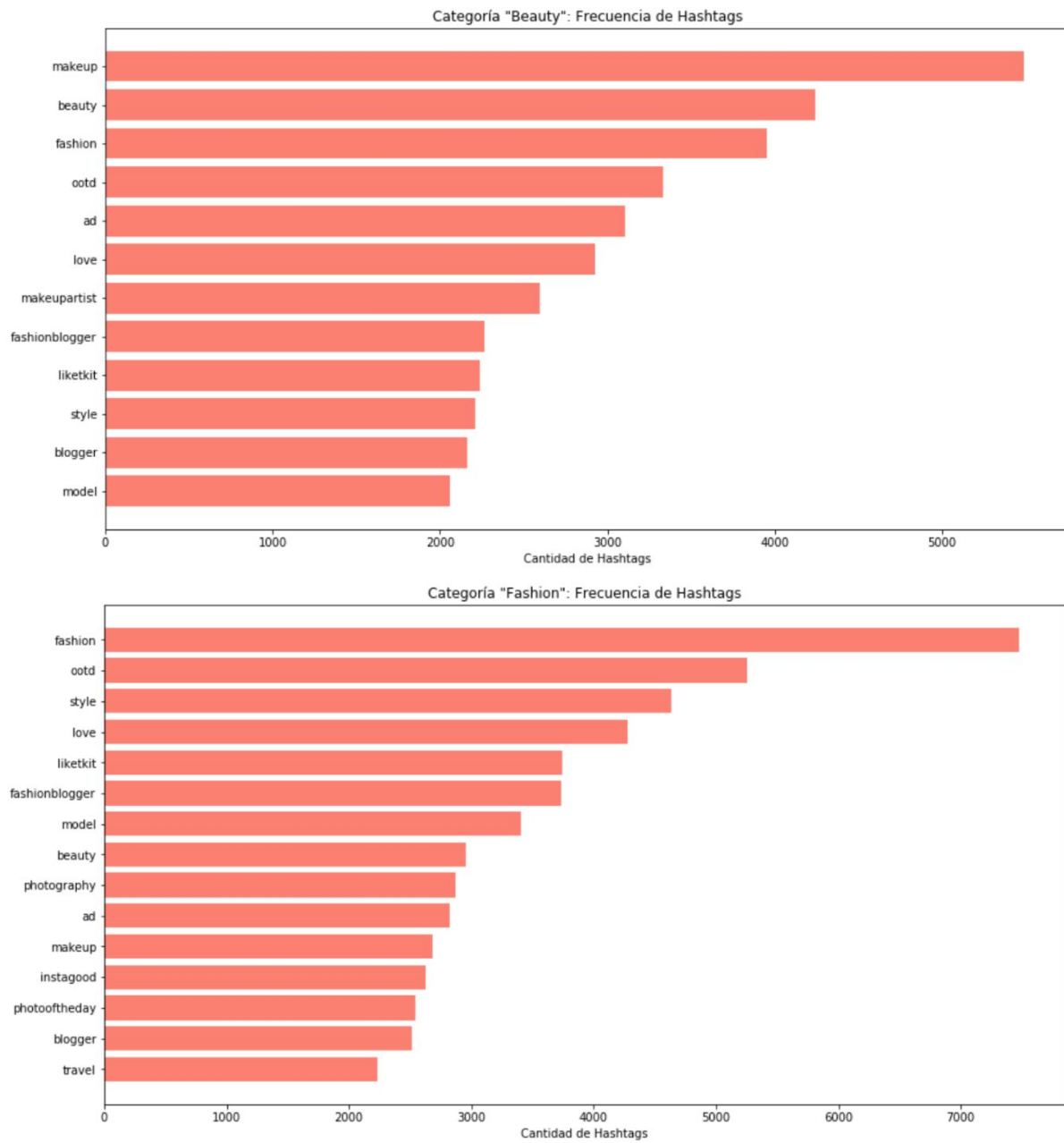
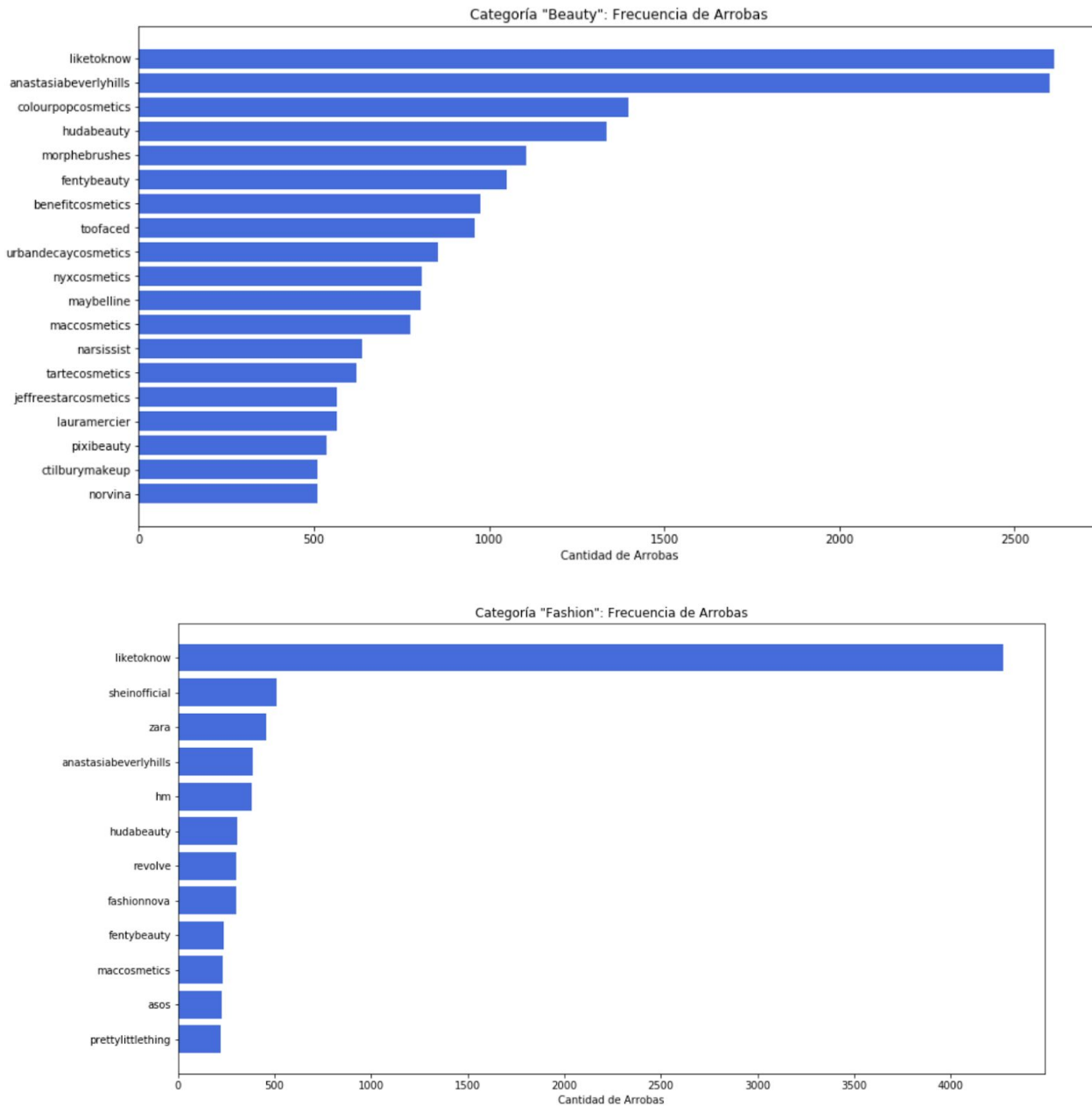


Figura 25: Frecuencia de Arrobas para las categorías "Beauty" y "Fashion"



## 4. Clustering

Uno de los objetivos de este trabajo es hacer uso de las bases de datos con las que contamos para correr algoritmos de clustering. Estos algoritmos tienen la cualidad de detectar patrones en los datos de forma tal que las observaciones puedan dividirse en distintos grupos o *clusters*. Aplicándolo al caso particular de este estudio, correr un algoritmo de clustering sobre los dataset que poseemos nos va a ayudar a detectar distintos tipos de influencers que existen hoy en las redes sociales, junto con sus características.

Debido a que existen varios modelos de clustering, la primera decisión que debimos tomar fue la de qué algoritmo elegir. No obstante, dado que para otros trabajos habíamos usado datos muy similares y habíamos utilizado el algoritmo *k-means* para hacer clustering con muy buenos resultados, decidimos que *k-means* era el modelo ideal para empezar a experimentar.

Sin embargo, un desafío al que debimos enfrentarnos y que no nos había ocurrido en trabajos anteriores, es la presencia de variables categóricas relevantes en el dataset. Como ya hemos mencionado al describir las variables, contamos con algunas categóricas como 'category', 'gender', 'age\_bracket', 'lang', etc. que son importantes al momento de describir al influencer.

Ahora bien, ¿por qué esto representó un desafío? Porque *k-means* es un algoritmo que funciona con variables únicamente numéricas debido a que computa medidas de distancia. En particular, *k-means* divide al conjunto de datos en  $K$  subconjuntos distintos sin solapamiento y asigna cada observación al cluster cuyo centroide es el más cercano según la distancia Euclídea (entendido centroide como la media de los valores de todas las observaciones pertenecientes a ese cluster).

Por ende, para poder realizar esta partición, el algoritmo se inicializa con determinada cantidad de *clusters* cuyos centros o centroides (valor medio del *cluster*) son aleatorios, y luego calcula una medida de distancia (distancia euclidiana) entre cada observación del dataset con respecto a cada centroide. De esta forma, *k-means* asigna cada observación a su centroide más cercano (lo asigna a un *cluster*). Luego el algoritmo recalcula los centroides promediando todas las observaciones de los *clusters* y vuelve a realizarse la asignación. Finalmente, cuando la asignación ya no cambie luego de varias iteraciones, se considera que *k-means* ya encontró los *clusters* que caracterizan a ese conjunto de datos.

Investigando sobre qué otros algoritmos de clustering existían descubrimos el algoritmo de clustering *k-modes*. Según *Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values*, de Zhexue Huang, "el algoritmo de *k-modes* reemplaza los promedios de los clusters (el centroide) con modas, y utiliza un método basado en la frecuencia para actualizar las modas en el proceso de clustering para minimizar la función de costo". Es decir, realiza el mismo proceso de *k-means* anteriormente descrito, pero en vez de utilizar el promedio de los valores de las observaciones para actualizar clusters y calcular los centroides, utiliza la moda de los valores categóricos. El problema que encontramos con este algoritmo es inmediatamente opuesto al que le encontrábamos a *k-means* y es que no estaríamos teniendo en cuenta los valores de las variables numéricas, los cuales son igual de valiosos que las categóricas para describir al influencer.

Finalmente, pudimos encontrar *k-prototypes*, un algoritmo que combina *k-means* (para variables numéricas) con *k-modes* (para variables categóricas). Este algoritmo fue nombrado y descrito

por primera vez en el artículo *An improved k-prototypes clustering algorithm for mixed numeric and categorical data* (Jinchao Ji et al. 2013). Fue creado con el fin de hacer clustering en bases de datos que mezclan tanto variables numéricas como categóricas. En vez de tratar de encontrar la similitud entre observaciones, este algoritmo trata de encontrar la desigualdad entre observaciones y agrupa los puntos con menos disimilitud entre sí en un cluster. La disimilitud total entre dos observaciones es la suma entre la disimilitud para atributos numéricos (la distancia Euclídea al cuadrado) y la disimilitud para atributos categóricos (número de atributos categóricos no coincidentes). El proceso de asignación es parecido al de los otros dos algoritmos. Se inicia con centroides aleatorios y se asigna cada observación al centroide del cual es menos disímil. El centroide luego se actualiza con el promedio de los atributos numéricos y la moda de los categóricos, y se iteran los pasos de reasignación de observaciones y actualización de centroides hasta que los centroides no varíen mucho al actualizarlos.

## 4.1 Ingeniería de Datos, Transformaciones y Nuevas Variables

Antes de aplicar *k-prototypes* al dataset, tuvimos que realizar una serie de transformaciones a las variables. Además, creamos nuevas variables que agregamos al dataset *instagram* utilizando la información de los posts de la base de datos *instagram\_posts*.

En primer lugar, tal como detectamos en el análisis descriptivo, las variables **total\_comments**, **total\_likes**, **total\_engagements** (suma de las dos primeras), **total\_posts**, **engagement\_rate** y **followers** exhiben una distribución asimétrica. Dado que *k-prototypes* también calcula medidas de distancia con las variables numéricas, vamos a querer que todas exhiban una distribución lo más simétrica posible de manera que las variables no tengan valores que presenten grandes diferencias entre sí o valores extremos. En estos casos, aplicar **logaritmo natural** nos permite suavizar la distribución de estas variables y hacerlas más simétricas. Por lo tanto, eso fue lo que hicimos.

Luego, generamos nuevas variables tanto en el dataset *instagram* como en *instagram\_post*.

En el caso de *instagram* transformamos primero la variable **username** en una que reporte la cantidad de caracteres del nombre de usuario. De esta forma, transformamos una variable de texto en una variable numérica para poder incluirla en el análisis. Al graficar su distribución, esta se mostró bastante simétrica. Luego, generamos la variable **redes\_soc**, la cual nos dice la cantidad de redes sociales que el influencer menciona en su **bio**. Esta variable fue construida a partir de buscar las palabras "gmail", "youtube", "hotmail", "twitch", etc. en el texto de la variable bio de cada influencer. El histograma de esta variable por categoría puede encontrarse en el Apéndice.

Con respecto *instagram\_posts*, usamos los datos para poder generar variables agregadas que luego pudieran ser fácilmente unidas al dataset *instagram*. Recordemos que *instagram\_posts* posee datos de los posts de los influencers que se encuentran en *instagram*, por lo cual, nos pareció interesante analizar la frecuencia con que cada influencer publica en Instagram, el porcentaje de publicaciones que son fotos o videos, la cantidad de *hashtags* promedio que usa cada influencer en sus posts y la cantidad de *arrobas* (forma de etiquetar a otras cuentas en Instagram) promedio que usa cada influencer en sus posts. Por ende, creamos las siguientes variables:

- **frecuencia\_post**: para crear esta variable, primero se transformó la variable **timestamp** (que está en formato horas desde las 00:00 del 01-ene-1970) para que exhiba la fecha y horario exactos de la publicación. Luego se ordenaron los posts de menos reciente a más reciente para cada influencer, y se tomó la diferencia en horas entre cada post. Así, calculamos para cada influencer el *promedio* de la diferencia en horas entre cada post y lo agregamos como variable en el dataset *instagram*.
- **picture\_share**: Del total de publicaciones de cada influencer, se calculó cuántas de estas eran imágenes y cuántas eran videos (a partir de la variable **type**). Luego se calculó el porcentaje que representa cada tipo de publicación sobre el total de posts para cada influencer, y se añadió el porcentaje de imágenes como **picture\_share** al dataset *instagram*.
- **video\_share**: porcentaje de posts que son videos sobre el total de publicaciones, utilizando la variable **type**.
- **hashtagspromedio**: A partir de la variable **text** (texto que aparece en la publicación), se contó la cantidad de *hashtags* que aparecían en este. Luego, se promedió la cantidad de *hashtags* utilizados en los posts para cada influencer, y se agregó como **hashtagspromedio** al dataset *instagram*.
- **arrobaspromedio**: Creada con metodología similar a **hashtagspromedio** pero contando los arrobas. Esta variable nos va a indicar cuántas cuentas de instagram etiqueta en promedio cada influencer en sus posts.

Todas de estas nuevas variables también mostraron tener distribuciones asimétricas (ver Apéndice), por lo cual se les aplicó el logaritmo natural.

A su vez, había cierta cantidad de observaciones que tenían uno o varios valores NaN entre sus columnas. Como los algoritmos de clustering requieren exclusivamente de valores **numéricos** para *k-means*, **categoricos** para *k-modes* y **ambos** para *k-prototypes* para funcionar, necesitamos eliminar todas las filas que tengan al menos un NaN. Hicimos esto con la función `na.omit()` en R. Decidimos eliminar los valores faltantes debido a que al ejecutar la función

na.omit() , solo se perdían un total de 2667 observaciones, cerca de 3% de la cantidad de observaciones totales. Al no perderse demasiada información, optamos por la opción de eliminarlos en vez de estimarlos con riesgo a introducir sesgo en las medias o varianza de las variables. De esta forma, optamos por seguir una metodología conservadora con el objetivo de producir resultados lo más insesgados posibles.

También nos dimos cuenta de que había varias observaciones que, si bien no tenían valores NaN, tenían el valor 0 tanto en la variable **total\_posts** (cantidad de posts) y **engagement\_rate** (tasa de participación). Esto entorpecería posteriormente nuestro análisis de clustering ya que modificaría las medias de los centros de cada cluster dando un panorama no tan verídico o incluso podría llegar a crear un cluster nuevo solo con las observaciones con valores 0, por lo cual decidimos eliminar estas observaciones. Dada la importancia de la variable 'engagement\_rate' y 'total\_posts' para definir la actividad diaria de los influencers en Instagram, y dado que las observaciones cuyo valor de 'engagement\_rate' y 'total\_posts' eran iguales a cero también contaban con las variables 'total\_comments', 'total\_likes' y 'total\_engagements' en cero, consideramos a estos casos también como missing values. Si bien se iban a perder observaciones e información con este paso, realmente no se podían estimar los valores faltantes sin hacer supuestos demasiado grandes y probablemente irreales. Aplicando esta estrategia, eliminamos solo un 15% de las observaciones totales, dejándonos aún un gran volumen de datos para hacer un análisis descriptivo pertinente.

Luego de todas estas modificaciones y eliminaciones, terminamos con un dataset resultante de 84.637 observaciones y 17 variables, que combinaba variables del dataset *instagram* original y variables nuevas creadas a partir del dataset *instagram\_posts*. A todas las variables numéricas, antes de correr el algoritmo de clustering, les aplicamos la función `scaling()` en R para re-escalarlas y expresarlas en una misma unidad. En particular, la función `scaling()` le resta a cada observación de cada variable el promedio de esa variable y luego divide por el desvío estándar de la misma. En el Apéndice se encontrará el listado de las variables finales que utilizamos para hacer el clustering.

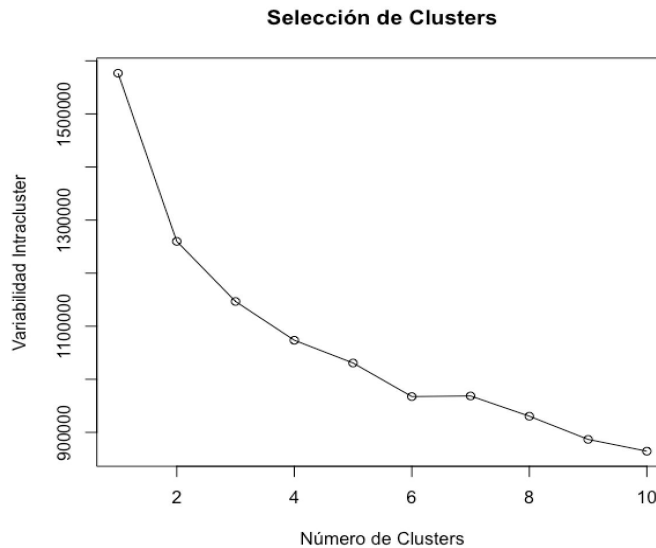
## 4.2 Implementación de *k-prototypes*

El primer modelo de clustering que corrimos fue con la totalidad del dataset (luego de los cambios del inciso anterior). Como el algoritmo de *k-prototypes* requiere que la cantidad de clusters a obtener sea especificada de antemano, corrimos el algoritmo para todas las cantidades de clusters entre 1 y 10 y guardamos el "withinness" (una métrica que representa la suma de las distancias de cada observación de su centroide) para cada iteración. Con esta lista de valores,



realizamos el *Elbow Graph*, un gráfico de líneas que muestra cómo varía la “withinness” al cambiar la cantidad de clusters especificados en el algoritmo. El gráfico de elbow que se obtuvo es el siguiente:

Figura 26: *Elbow Graph* para seleccionar la cantidad de clusters



Al aumentar la cantidad de clusters, el “withinness” baja debido a que cada vez más observaciones estarán más cerca de su centroide, al punto de que si la cantidad de clusters es igual a la cantidad de observaciones, el “withinness” será 0 ya que los centroides resultantes serán iguales a cada observación del dataset. Cada vez que se aumenta en 1 la cantidad de clusters, bajará menos el “withinness”. La idea de este gráfico es seleccionar la cantidad de clusters para los cuales no vale la pena seguir aumentando el número dado que de ahí en más la disminución del “withinness” no compensará la disminución en utilidad de la cantidad de clusters.

Decidimos utilizar 4 clusters porque es donde consideramos que está el “codo” del gráfico. Al recuperar los valores de cada uno de los 4 centroides obtuvimos la siguiente tabla:

Tabla 4: Valores de los centroides para cada variable según el cluster

Variables	Cluster			
	1	2	3	4
lang	en	en	en	en
gender	female	female	female	female
age_bracket	25-34	25-34	25-34	25-34
category	fashion	beauty	gaming	food
username	12.74	12.12	12.43	13.33
followers	24,839	483,458	83,582	28,986
engagement_rate	0.10%	0.60%	0.16%	1.20%
total_posts	65	608	197	416
total_engagements	25,869	4,487,799	215,777	171,413
total_likes	24,876	4,414,707	209,543	163,031
total_comments	993	73,092	6,234	8,382
redes_soc	1.13	1.24	1.23	1.09
hashtagspromedio	7.93	3.6	7.67	7.78
frecuencia_post	132 hs	59 hs	101 hs	87 hs
arrobaspromedio	1.16	1.21	1.67	1.16
picture_share	94%	89%	51%	96%
video_share	6%	11%	49%	4%

Por si acaso, también realizamos esta tabla para un número de clusters igual a 5 y 6 para observar cómo se comportaban las variables y qué resultado hacía más sentido de negocio. Sin embargo, el resultado al hacer 4 clusters fue el más interesante que obtuvimos. Tras analizar estos resultados, creamos 4 “Personas” que podrían representar a cada cluster que se pueden ver en el Apéndice. Podemos notar que todos los clusters resultantes tienen la variable ‘gender’ en femenino y tienen los mismos rangos de edad. Esto coincide con lo encontrado en la etapa descriptiva, donde veíamos que los influencers eran en su mayoría mujeres y que el rango de edad más frecuente era el de 25-34 años. Lo mismo ocurre con el idioma, el cual es el mayoritario también (‘en’ - inglés). La creación de las *Influencer Personas*, es decir, los influencers representativos de cada cluster obtenido, nos sirvió esencialmente para tener más claro cuáles eran los patrones generales presentes en los datos y las variables clave para distinguir entre distintos tipos de influencers.

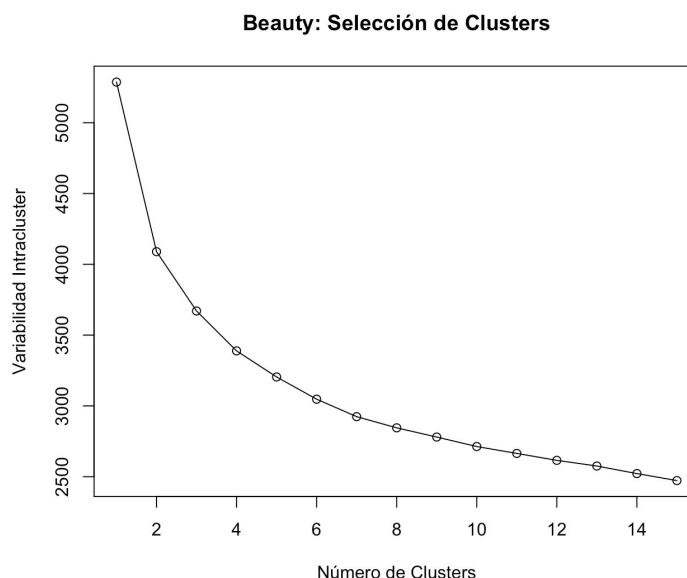
Como ya mencionamos, corrimos el modelo de clustering también para un número de clusters igual a 5 y 6, para observar cómo iban cambiando los clusters en términos de los centroides, y cuáles eran las variables que nos resultaban más útiles para interpretar la división de los clusters. De esta forma, descubrimos que el número de seguidores (variable ‘followers’) parecía ser una de las variables que más cambiaba entre los clusters resultantes. Es decir, en cada corrida del algoritmo, siempre obteníamos algunos clusters con pocos seguidores (entre 20K y 30K seguidores), otros con muchos seguidores (entre 200K y 600K seguidores) y otros con un

número de seguidores “en el medio” (entre 60K y 150K seguidores). De hecho, en la tabla mostrada anteriormente, podemos observar que obtuvimos dos clusters cuyos centroides para la variable seguidores toman el valor de 24,839 y 28,986, otro con valor 83,582 y otro con el mayor número de seguidores llegando a 483,458. Por ende, concluimos que probablemente siempre íbamos a obtener una clusterización que nos de como resultado clusters con distintos niveles de seguidores, quizás relacionado con la presencia de influencers Micro, Regulares, ‘Rising’, Mid y Macro.

Por otro lado, también podemos notar que los 4 clusters resultantes son de **distintas categorías**. Al correr el algoritmo con más clusters, se mantiene el hecho de siempre obtener clusters con diferentes categorías, por lo que también nos pareció una variable relevante al momento de interpretar y darle un sentido a los clusters. Decidimos entonces que sería interesante si eligiéramos unas cuantas categorías que nos sean las más atractivas y ejecutar el algoritmo de clustering sólo en las observaciones pertenecientes a esas categorías. De esta manera, podríamos ver cómo se conforman los grupos dentro de cada categoría, ya que tal vez cada una de ellas tiene una composición diferente de grupos diferenciados.

Entonces, realizamos el mismo proceso de ejecutar el algoritmo para 1 a 10 clusters y crear el *Elbow Graph* para cada una de las categorías. En particular, mostraremos los gráficos resultantes para las categorías ‘beauty’ y ‘fashion’, que nos resultaron particularmente interesantes y que posteriormente decidimos utilizar para la fase de experimentación.

Figura 27: *Elbow Graph* para seleccionar la cantidad de clusters para la categoría “Beauty”



Decidimos quedarnos con 4 clusters ya que es donde concluimos que se encontraba el “codo”. Con respecto a los centroides obtenidos a partir de este análisis, el cuadro a continuación los exhibe:

Tabla 5: Valores de los centroides para cada variable según el cluster (categoría “Beauty”)

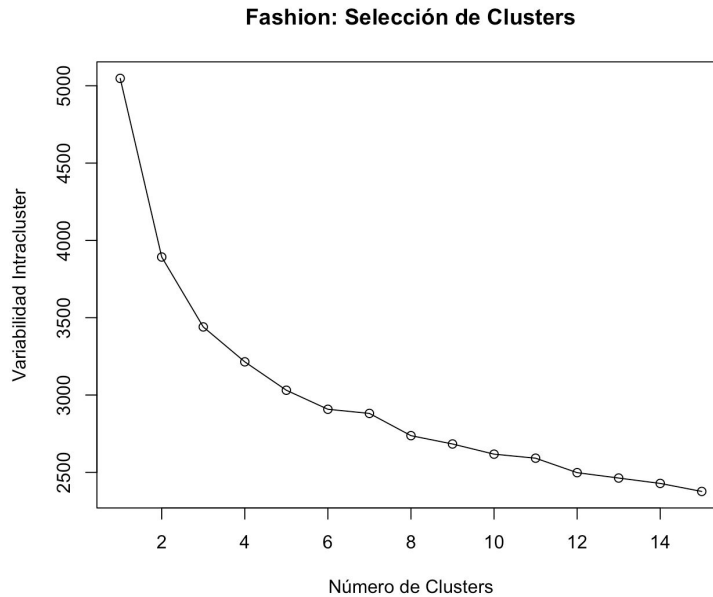
Variables	Cluster			
	1	2	3	4
lang	en	en	en	en
gender	female	female	female	female
age_bracket	25-34	25-34	18-24	25-34
username	12.49	12.77	11.87	13.38
followers	46,176	29,967	410,376	89,131
engagement_rate	0.31%	0.17%	0.80%	0.69%
total_posts	132	69	533	419
total_engagements	89,432	27,302	4,442,597	554,150
total_likes	86,551	26,150	4,356,102	537,827
total_comments	2,882	1,153	86,494	16,323
redes_soc	1.18	1.21	1.28	1.20
hashtagspromedio	0.65	12.10	1.47	11.95
frecuencia_post	120 hs	125 hs	64 hs	77 hs
arrobaspromedio	1.21	1.75	1.35	2.41
picture_share	91%	88%	88%	88%
video_share	9%	12%	12%	12%

Se puede observar que al igual que en el caso más general de clustering en dónde no dividimos por categoría, obtuvimos aproximadamente 3 niveles de clusters basados en el número de seguidores. Primero, tenemos el cluster menos popular con ~29K seguidores, el más popular con ~410K seguidores y en el medio tenemos dos clusters: uno de ~46K seguidores y otro de ~89K seguidores. De esta forma, vemos que el clustering parece siempre encontrar este patrón tanto a nivel general como a nivel categoría, distinguiendo entre influencers más Micro o Regulares, Rising o Mid e influencers Macro.

Otro *insight* interesante que surge de la tabla anterior es que el cluster con mayor cantidad de seguidores es aquel que publica con más frecuencia contenido (publican cada 64 horas en promedio, es decir, cada 2 días y medio) y que a su vez tiene el ‘engagement rate’ más alto, contando con la mayor cantidad de comentarios y de *likes*. Por supuesto, tiene la mayor cantidad de posts en el período analizado (del 4 de septiembre al 4 de diciembre) y un promedio de edad más bajo que el de otros clusters (rango 18-24). Esto último es muy interesante dado el diagnóstico que habíamos hecho de la variable ‘age bracket’ y de ‘followers’, finalmente concluyendo que los influencers más jóvenes suelen tener más seguidores dado su constante relación con la tecnología y publicaciones de contenido frecuentes.

Con respecto al uso de *hashtags* es remarcable que el cluster con menos seguidores es aquel que tiene en promedio mayor cantidad de *hashtags* por post. Esto tiene mucho sentido dado que los influencers en este grupo están en pleno crecimiento y quieren hacerse conocer.

Figura 28: *Elbow Graph* para seleccionar la cantidad de clusters para la categoría "Fashion"



Al igual que en el caso de la categoría 'beauty', el *Elbow Graph* nos está indicando que la variabilidad dentro de cada cluster comienza a disminuir mucho más "lento" a partir de 4 clusters. Podría discutirse que esto ocurre en realidad a partir de 6 clusters, ya que no es del todo claro cuál es el lugar justo en el cual se forma un "codo". No obstante, luego de correr el algoritmo con 4, 5 y 6 clusters, concluimos que utilizar 4 clusters tenía sentido tanto a nivel comparativo (para comparar con el clustering más general) pero también a nivel de simplificación en el análisis e intuición de negocios. Por ende, también analizamos los resultados en función de 4 clusters.

Tabla 6: Valores de los centroides para cada variable según el cluster (categoría "Fashion")

Variables	Cluster			
	1	2	3	4
lang	en	en	en	en
gender	female	female	female	female
age_bracket	25-34	25-34	25-34	25-34
username	13.40	12.80	12.60	12.30
followers	68,231	26,239	47,910	345,384
engagement_rate	0.80%	0.10%	0.40%	0.90%
total_posts	426	58	124	560
total_engagements	417,808	25,832	78,946	3,526,589
total_likes	403,863	25,001	76,303	3,472,408
total_comments	13,945	830	2,643	54,181
redes_soc	1.10	1.10	1.10	1.20
hashtagspromedio	13.05	12.88	0.61	1.53
frecuencia_post	75	133	122	60
arrobaspromedio	1.79	1.57	1.10	1.41
picture_share	91%	89%	93%	92%
video_share	9%	11%	7%	8%

Al igual que en la categoría 'beauty', el cluster con mayor cantidad de seguidores en promedio tiene el mayor 'engagement rate', la mayor frecuencia de publicaciones y la mayor cantidad de publicaciones en el período analizado. Comparando con los clusters que expusimos para 'beauty', en este caso podemos ver que el porcentaje de contenido que son videos es menor para todos los clusters, lo cual es intuitivo dado que los influencers que se dedican a la categoría 'beauty' suelen hacer más videos mostrando cómo maquillarse, técnicas y otros consejos. Por otro lado, los *hashtags* promedio también son más altos nuevamente para el cluster con menos seguidores.

Con respecto a la cantidad de seguidores para todos los clusters, el patrón es similar al encontrado en las otras corridas del algoritmo: contamos con un cluster con pocos seguidores (~26K), otro con muchos seguidores (~345K) y otros dos con un nivel intermedio (~47K y ~68K).

### 4.3 Conclusiones de Clustering

Del primer clustering que ejecutamos podemos concluir que las dos variables que nos resultaron más útiles al momento de interpretar y darle sentido a los clusters fueron la variable de categoría y la cantidad de seguidores. Sin embargo, tras ejecutar el algoritmo dentro de cada categoría, pudimos ver que la variable más interesante pasa a ser principalmente la cantidad de seguidores. Si hay algo en común que vimos en todos los algoritmos de clustering por categoría corridos, es que la cantidad óptima de clusters siempre ronda alrededor de cuatro clusters y siempre se componen de un cluster con relativamente pocos seguidores representando al

Micro-Influencer, un cluster mediano cercano al Micro-Influencer, un cluster de influencer mediano más cercano al Macro-Influencer y, por último, un Macro-Influencer con relativamente muchos seguidores.

Con estos *insights* obtenidos, vamos a proceder a la fase de experimentación. En esta fase, vamos a basarnos en los resultados obtenidos en la etapa de clustering para poder construir modelos de influencer o “Personas” influencers que representen de alguna manera los clusters armados, y luego manipularemos distintas variables relacionadas a estos influencers para hacer un estudio del comportamiento de los consumidores. En particular, queremos evaluar cómo reaccionan los individuos en Instagram ante un influencer que vende un producto, y si el deseo de compra de este producto cambia según las características del influencer.

## 5. Etapa Experimental

A partir de los resultados que obtuvimos de la etapa de clustering, procederemos con la etapa experimental, que se caracterizará por utilizar como base los resultados obtenidos anteriormente y luego manipular ciertas características de estos resultados para así evaluar cómo influyen estas características sobre el comportamiento del consumidor.

Como ya hemos explicado, los influencers se han convertido, al igual que las celebridades, en figuras públicas que son capaces de influenciar el comportamiento de los individuos o audiencias que los siguen a través de un medio relativamente reciente: las redes sociales. Esto ha llamado la atención de las compañías y empresas actuales, interesadas en vender sus productos a través de este nuevo medio. De esta forma, los influencers pasaron a ser vistos por las empresas como posibles intermediarios entre la compañía en sí y los consumidores, y como posibles promotores de los productos que estas venden.

Dado este nuevo contexto, lo que nos interesaba evaluar es cómo cambia el comportamiento del consumidor (la disposición a comprar, disposición a pagar, disposición a difundir, etc.) de acuerdo a las características del producto y del influencer que lo promociona. De esta forma, las empresas podrán tener mayor información al momento de decidir el/los influencers óptimos para publicitar sus productos y darle una mejor utilización a sus recursos.

En particular, nuestra pregunta de interés es la siguiente: *La similitud entre el influencer y sus seguidores ¿influye en los resultados de negocio?* Es intuitivo pensar que cuanto más identificados se sienten los seguidores con el influencer, más están dispuestos a pagar por el producto anunciado por el influencer. Sin embargo, lo más interesante es que este proceso de identificación también podría tener un efecto sobre la voluntad de sugerir el producto a otros, la disposición a pagar y voluntad de compra, entre otros.

Con el objetivo de medir estas relaciones, realizaremos una serie de experimentos con influencers ficticios. En primer lugar, realizaremos un pretest que nos ayudará a definir qué perfiles de influencers deberíamos utilizar y en segundo lugar, realizaremos un experimento final que buscará medir los efectos de una mayor similitud sobre la disposición a pagar, voluntad de compra, de sugerir el producto a otros, etc.



## 5.1 Similitud entre el influencer y sus seguidores

Para poder responder a nuestra pregunta de investigación, primero debíamos decidir qué significaba que un seguidor sienta que tiene similitud con el influencer. Particularmente, pensamos que la similitud con estas figuras de las redes sociales podría darse teniendo en cuenta similitudes físicas o ideológicas. Dado que pensamos que la similitud ideológica puede llegar a generar resultados más interesantes al momento de evaluar cómo esta influye en la disponibilidad a pagar y otros resultados de negocio, decidimos enfocarnos en este tipo de similitud para realizar nuestro análisis.

Es así como comenzamos a pensar cuáles eran los factores sobre los cuales enfocarnos. Si bien actualmente existen varios debates ideológicos que atraviesan a la sociedad, en la Argentina, uno de los debates que más polaridad produce es el correspondiente a **la legalización del aborto**. En los últimos años, el debate entre los que apoyan la legalización y los que no la apoyan se ha hecho muy presente, generando marchas multitudinarias de ambos grupos. De hecho, durante 2019 se debatió en el Congreso de la Nación Argentina un proyecto de ley para legalizar el aborto, el cual finalizó con la decisión conjunta de la Cámara de Diputados de no legalizarlo. No obstante, este debate probablemente se retome debido a la gran polémica que genera y el constante pedido de debate por parte de los que están a favor de la legalización.

De esta forma, decidimos que podríamos experimentar generando perfiles de influencers que sean similares pero difieran en su ideología. Esta diferencia en opinión se expresaría a través de *hashtags*, *emojis*, etc. que actualmente son utilizados por aquellos que están a favor o en contra de la legalización del aborto.

## 5.2 Selección de categoría “fashion influencers”

Por otra parte, también debimos seleccionar a qué **categoría** pertenecerían estos influencers ficticios, que luego mostraríamos al público para que responda una serie de preguntas. De las diez categorías presentes en el dataset *instagram*, la que nos pareció más precisa utilizar es **fashion** (o **moda**, si lo queremos expresar en español), dado que es una categoría que está altamente relacionada con la autoexpresión e identidad de las personas. De esta manera, tanto la categoría elegida como la dimensión sobre la que evaluaremos la similitud, estarán muy relacionadas con elementos que identifican a una persona: ideología, opiniones, principios, autoexpresión, estilo, identidad, entre otros.

### 5.3 Promoción de productos: nivel de autoexpresión

Por supuesto, los influencers que construyamos para poder experimentar y evaluar cómo influye la similitud sobre los resultados de negocio, deben estar promocionando o vendiendo un producto. Es aquí donde quisimos agregar una dimensión adicional: **el nivel de autoexpresión del influencer reflejado en el producto**. Esta dimensión busca capturar cuán relacionado está el producto con el influencer en términos de su autoexpresión, por lo cual, evaluaremos tres tipos de productos que el influencer puede vender:

1. Producto de su propia marca: el influencer puede estar altamente involucrado con el producto dado que este es de su propia marca. Esto transmitirá una idea de mayor autoexpresión por parte del influencer, debido a que su propia marca estará relacionada con la propia identidad del influencer, sus gustos, su estilo, etc.
2. Producto de otra marca: el influencer no está relacionado con el producto al mismo nivel que en el caso anterior. Simplemente, el influencer publicita el producto indicando la marca a la que pertenece. El nivel de autoexpresión del influencer en este caso es más bajo debido a que si bien promociona el producto indicado y su marca (señal que indica que claramente le gusta esa marca en particular), no es de su autoría el producto en sí (como en el caso anterior).
3. Producto de marca indefinida: el influencer en este caso publicita un producto de marca indefinida. Este es el nivel más bajo de autoexpresión por parte del influencer, ya que este ni siquiera menciona la marca en cuestión, solo muestra el producto indicado que es de su gusto.

Por supuesto, en el caso particular de este experimento, los productos serán siempre piezas de ropa o vestimenta, dado que utilizaremos la categoría *fashion* o moda.

Esta dimensión agregada se convertirá en otra variable independiente en nuestro análisis. La intuición que nos guía es la siguiente: si un individuo tiene un nivel de similitud bajo con el influencer, cuanto más relacionado esté el producto con el influencer en sí mismo, menor va a ser la probabilidad de que el seguidor compre efectivamente el producto (o desee hacerlo); no obstante, a mayor “neutralidad” del producto (producto de otra marca o sin marca), esa probabilidad no será tan baja, dada la relación menos estrecha entre el producto y el influencer. Llamaremos “ajuste producto-influencer basado en la autoexpresión” a esta dimensión, que en particular se medirá a partir de preguntarle a los individuos si piensan que el producto está altamente relacionado con el influencer y si ellos piensan que van a ser asociados con las ideas del influencer si compran el producto.

## 5.4 Objetivo de realizar un Pretest

Una vez definidas todas las dimensiones que tendremos en cuenta, nos queda definir el experimento que queremos realizar. Principalmente, el objetivo será crear influencers ficticios con distintas ideologías (un influencer a favor de la legalización del aborto vs. un influencer en contra) que a su vez vendan un producto de su propia marca o un producto “neutral” (de otra marca/sin marca). Luego, se les hará preguntas a los individuos que participen del experimento para evaluar cómo influyen el nivel de similitud y de autoexpresión de los seguidores con el producto sobre la disposición a pagar por el producto, la voluntad de sugerir el producto a otros y la intención de recompra del seguidor.

Para poder hacer el análisis con perfiles lo suficientemente creíbles, decidimos realizar un pretest para poder definir cuáles eran los mejores perfiles a utilizar en el experimento principal. Por lo cual, creamos en total cuatro perfiles de influencers pertenecientes a la categoría moda, con tres productos asociados a cada uno.

La creación de estos perfiles ficticios se hizo en base a lo obtenido a partir de aplicar técnicas de clustering. En particular, seleccionamos el cluster que tuviera mayor cantidad de seguidores ya que nos interesaba crear influencers que tuvieran un alto nivel de popularidad, para hacer el efecto de la opinión e ideología mucho más fuerte. Es natural pensar que si un influencer pequeño, con baja cantidad de seguidores, está a favor o en contra del aborto, no es algo que a los seguidores les va a importar demasiado. En cambio, si el influencer es altamente conocido y opina de cierta forma, es más probable que sus dichos tengan repercusiones en las redes sociales.

Por otro lado, este cluster también exhibe el nivel más alto de engagement rate, lo cual también es algo que buscamos y queremos que posean los influencers ficticios. Entre otros datos interesantes, también puede notarse que este cluster menciona, en promedio, una red social adicional en su bio (que en general es la dirección de mail y si menciona otra red social, menciona YouTube), publica en promedio cada dos días y publica en su mayoría imágenes o fotos (concordante con el hecho de que hablamos de influencers de moda). Notar que la cantidad de seguidores expuestos en todos los perfiles de Instagram ficticios es aproximadamente el mismo que el obtenido en el cluster con mayor engagement rate de la categoría moda o *fashion*, y la edad expuesta en los perfiles está dentro del rango 25-34 años. La cantidad de publicaciones totales surgen de suponer que los influencers ficticios tienen una antigüedad de aproximadamente tres años utilizando la plataforma. Por último, todos los perfiles

son de género femenino, lo cual coincide con el hecho de que el género del cluster es femenino y que la mayoría de los influencers que publican contenido de moda son mujeres.

De esta forma, utilizando información obtenida de la etapa de clustering, creamos dos perfiles de moda que estén a favor de la legalización del aborto y otros dos que estén en contra. Los cuatro perfiles tendrán en común la cantidad de seguidores, la mención de una o dos redes sociales en la bio, la cantidad de publicaciones en el perfil y el uso de varios *hashtags* (entre dos o tres). Por ende, la única diferencia entre los dos perfiles a favor de la legalización o entre los dos perfiles en contra de la legalización, es cómo está redactada la bio, la imagen del influencer, los *emojis* utilizados y los *hashtags*.

Luego, a cada perfil están asociadas tres publicaciones del mismo producto, pero con la diferencia de que una publicación será de la propia marca del influencer, otra hará referencia a otra marca y la última publicación será de una marca indefinida.

Todos estos perfiles y sus publicaciones asociadas pueden verse en el Apéndice. El objetivo es que a partir del pretest, podamos elegir **un perfil a favor de la legalización del aborto, uno en contra**. Los elegiremos en base a correr ciertos test estadísticos que nos ayuden a identificar cuáles de los perfiles es más creíble. A su vez, **elegiremos entre la publicación referida a una marca en particular y la de marca indefinida**. Queremos averiguar cuál de las dos es la más “neutral” con respecto a la publicación de la marca propia del influencer.

Luego, con esta información que surja del pretest, armaremos los perfiles definitivos que utilizaremos en el experimento final. Este experimento tiene como objetivo mostrarles a las personas que participen del experimento un influencer al azar (este puede estar a favor de la legalización o en contra) con una publicación al azar (publicación de marca propia vs. de otra marca/marca indefinida -definiremos cuál utilizar a partir del pretest-) y que estos participantes respondan ciertas preguntas. Haremos una mayor explicación de esta etapa más adelante. A continuación, describiremos el pretest más en profundidad.

## 5.5 Pretest: Etapa Pre-Experimental

Realizamos un pretest ( $n = 20$ , edad promedio = 28.4 desvío standard de la edad = 6.47) para elegir los perfiles de influencers y publicaciones que nos ayudarían a manipular exitosamente: 1) los valores y puntos de vista percibidos del influencer (a favor de la legalización del aborto vs. en contra de la legalización del aborto) y 2) ajuste producto-influencer basado en la autoexpresión (marca del influencer vs. otra marca/sin marca). Con el objetivo de evaluar si nuestro diseño de

perfiles y publicaciones funcionaban como previsto, le preguntamos a los participantes que evalúen cuatro perfiles de influencers ficticios pero realistas y, para cada influencer, evaluar dos de sus publicaciones. Todas los perfiles creados que utilizamos y las preguntas hechas a los participantes pueden encontrarse en el Apéndice.

Los valores y puntos de vista del influencer fueron manipulados agregando *hashtags*, símbolos, colores y frases que están típicamente asociadas con estar a favor y en contra de la legalización del aborto. A su vez, el ajuste producto-influencer basado en la autoexpresión fue manipulado a partir de tener una influencer de **moda** (*fashion influencer*) promocionando ropa de su propia marca o una marca indefinida u otra marca. En general, a los participantes se les mostraron al azar cuatro perfiles de influencers de moda: dos perfiles a favor del aborto y dos perfiles en contra del aborto. Luego de haber visto cada perfil, a los participantes se les mostraron dos publicaciones de cada influencer promocionando un producto. Una de las publicaciones promocionaba ropa de la propia marca del influencer, mientras que la otra publicación promocionaba ropa de otra marca o de marca indefinida.

En primer lugar, a los participantes se les mostró la página de perfil de Instagram del Influencer. Se les pidió que respondieran una serie de preguntas sobre la persona del perfil que acababan de ver. Inicialmente, se les pidió que respondan si pensaban que el influencer tenía su propia marca, si le pagaban por promocionar productos de otras marcas o si promocionaba su propia marca, en una escala de Likert de 7-puntos (1 = extremadamente improbable, 7 = extremadamente probable). Adicionalmente, se les pidió que respondieran qué tan probable les parecía que este influencer esté a favor de la legalización del aborto o en contra, usando una escala de Likert de 7-puntos (1 = definitivamente en contra de la legalización del aborto; 7 = definitivamente a favor de la legalización del aborto). Finalmente, se les preguntó a qué categoría pensaban que pertenecía el influencer (1 = definitivamente un influencer de moda, 7 = definitivamente un influencer de belleza).

En segundo lugar, a los participantes se les mostró dos de las publicaciones de estos influencers. Para medir el ajuste producto-influencer basado en la autoexpresión, se les pidió a los participantes que expresaran qué tan probable les parecía que el producto promocionado esté asociado con la persona promocionándolo, así como también qué tan probable les parecía que se les asociaría con la persona y que la respaldan/aprueban si compraran el producto publicitado (1= para nada, 7=extremadamente).

## 5.6 Resultados del Pretest

### 5.6.1 Valores y puntos de vista del influencer

Con el objetivo de determinar qué perfil fue el más efectivo en transmitir la imagen de una persona a favor de la legalización del aborto o en contra, comparamos las puntuaciones obtenidas de las respuestas de los participantes a la pregunta “esta persona está definitivamente en contra de la legalización del aborto (1), definitivamente a favor de la legalización del aborto (7)” para todos los perfiles que le mostramos a los participantes. Ejecutamos un ANOVA de medidas repetidas unifactorial para comparar el efecto del tipo de influencer sobre las puntuaciones de los valores percibidos del influencer. Como era de esperar, los resultados mostraron que el tipo de influencer tiene un efecto significativo sobre si el influencer es considerado a favor o en contra del aborto ( $F(7,133) = 90.03, p < .001, \eta^2 = .82$ ). Comparaciones post hoc fueron analizadas utilizando t-tests dependientes con una corrección de Bonferroni (Bonferroni 1936)<sup>5</sup>. Los dos perfiles a favor del aborto resultaron tener puntuaciones significativamente mayores que los perfiles en contra del aborto (Tabla 7). La Figura 29 muestra las puntuaciones promedio junto a otros intervalos de confianza. En ella hacemos referencia a los dos perfiles ficticios a favor de la legalización del aborto como “pro1” y “pro2”, mientras que hacemos referencia a los dos perfiles ficticios en contra de la legalización como “contra1” y “contra2”.

Tabla 7: Promedios (M) y Desvíos Estándar (SD)

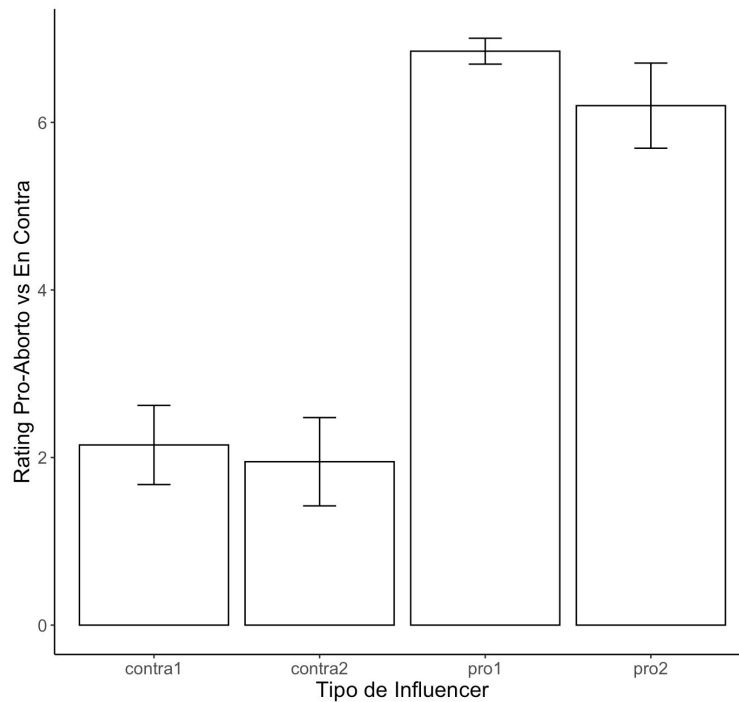
	M	SD
Pro-Aborto 1 <sup>a</sup>	2.15	1.48
Pro-Aborto 2 <sup>a</sup>	1.95	1.65
Contra-Aborto 1 <sup>a</sup>	6.85	.48
Contra-Aborto 2 <sup>a</sup>	6.20	1.59

Nota: Los promedios con el mismo superíndice no son significativamente distintos entre sí.

---

<sup>5</sup> La corrección de Bonferroni es un método estadístico que se utiliza para contrarrestar el problema de las comparaciones múltiples. La importancia de este método estadístico radica en que una prueba de hipótesis se basa en rechazar una hipótesis nula si ésta tiene poca probabilidad de ocurrencia, pero si se prueban múltiples hipótesis, incrementa la probabilidad de un evento "extraño" y por lo tanto aumenta la probabilidad de que se rechace incorrectamente una hipótesis nula (error tipo I). Es por esto que la corrección de Bonferroni se utiliza para compensar ese aumento ya que prueba cada hipótesis de manera individual con un nivel significativo de  $\alpha/m$  (nivel de alfa general deseado sobre  $m$  número de hipótesis). Fue nombrado por primera vez en el paper *Teoria statistica delle classi e calcolo delle probabilita*, por Carlo Emilio Bonferroni.

Figura 29: Puntuaciones Promedio de la pregunta relacionada a la ideología del Influencer (1= en contra de la legalización, 7= a favor de la legalización)



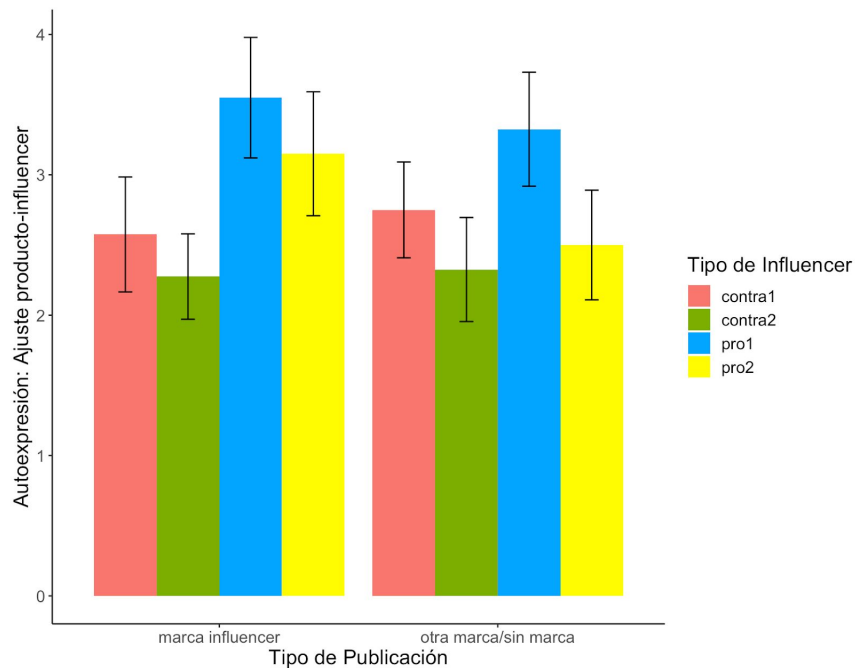
### 5.6.2 Ajuste producto-influencer basado en la autoexpresión

En primer lugar, corrimos un análisis de fiabilidad sobre las tres preguntas/afirmaciones que buscaban medir el ajuste producto-influencer basado en la autoexpresión y encontramos que dejando de lado una de ellas, nuestra medida de fiabilidad mejoraba sustancialmente (de  $\alpha = .60$  a  $\alpha = .90$ ) por lo que decidimos crear una medida compuesta con dos de las tres preguntas. Corrimos un ANOVA de medidas repetidas con dos factores para comparar el efecto del tipo de influencers y el tipo de publicación (producto con marca del influencer vs. de otra marca/sin marca) sobre esta medida compuesta, la cual tiene como objetivo representar el ajuste producto-influencer basado en la autoexpresión. Lo que intuitivamente buscamos encontrar es que las publicaciones con productos de marca propia del influencer generen una respuesta positiva (más alta en el rango del 1 al 7) con respecto a que si se comprara el producto, los seguidores claramente se sentirían identificados con el influencer. No obstante, esta respuesta debería ser más bien neutral si el producto no pertenece a ninguna marca o es de otra marca.

Los resultados mostraron un efecto significativo del tipo de influencer ( $F(7,133) = 6.38, p < .001, \eta^2 = .14$ ), un efecto significativo del tipo de publicación ( $F(1, 19) = 43.58, p < .001, \eta^2 = .12$ ), y una interacción significativa de los dos ( $F(7,133) = 14.47, p < .001, \eta^2 = .15$ ). Comparaciones post hoc fueron analizadas usando t-test dependientes con una corrección de Bonferroni. Desafortunadamente no encontramos diferencias entre la publicación con marca propia del influencer y la publicación de otra marca/sin marca dado un mismo tipo de influencer (a favor del aborto o en contra). Como ya hemos mencionado, esperábamos que la publicación de otra

marca/sin marca obtuviera una puntuación más baja en la autoexpresión percibida del producto-influencer, independientemente del tipo de influencer. La Figura 30 muestra las puntuaciones promedio junto con los intervalos de confianza.

Figura 30: Puntuaciones Promedio de las Preguntas relacionadas a la autoexpresión del Influencer (por tipo de publicación y tipo de Influencer)

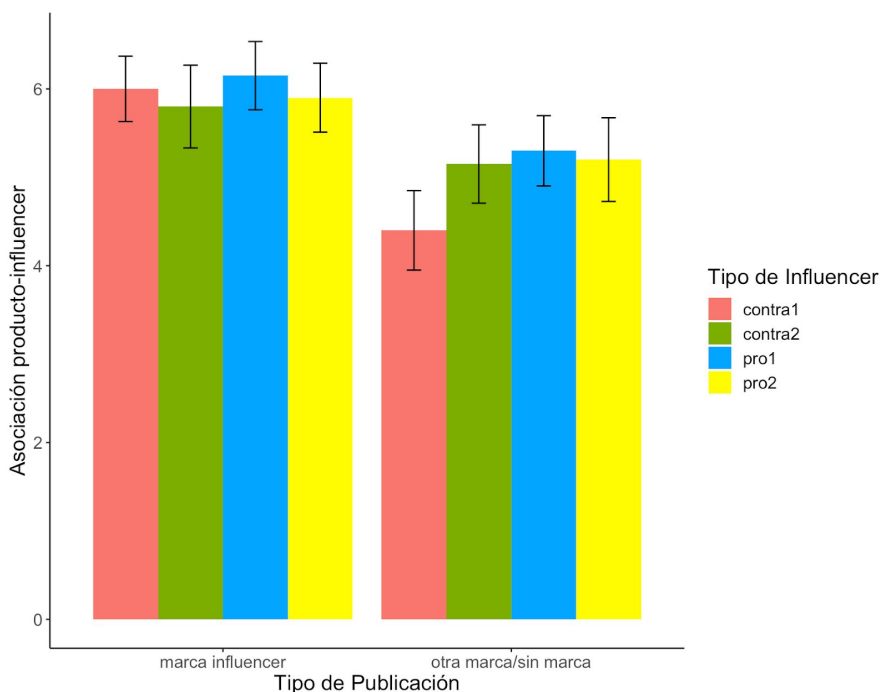


Dado que no utilizamos la primera pregunta/afirmación (“El producto publicitado está asociado fuertemente con la persona que lo publicita”) en la medida compuesta, decidimos correr el mismo análisis para ver si el tipo de influencer y la publicación tenían un efecto sobre la puntuación. Corrimos un ANOVA de medidas repetidas con dos factores para comparar el efecto del tipo de influencers y el tipo de publicación utilizando sólo las puntuaciones de la primera pregunta y encontramos un efecto significativo del tipo de influencer ( $F(7,133) = 31.14, p < .001, \eta^2 = .41$ ), un efecto significativo del tipo de publicación ( $F(1, 19) = 67.76, p < .001, \eta^2 = .14$ ), y una interacción significativa de los dos ( $F(7,133) = 10.61, p < .001, \eta^2 = .09$ ). Comparaciones post hoc fueron analizadas usando t-test dependientes con una corrección de Bonferroni. En este caso, sí encontramos diferencias significativas entre la publicación con marca propia del influencer y la publicación de otra marca/sin marca dado un mismo tipo de influencer. Particularmente, para el influencer “pro1” (a favor de la legalización del aborto -perfil 1-), la diferencia entre las puntuaciones promedio de la publicación con marca propia del influencer (“post1”) y la publicación sin marca/otra marca (“post2”) resultó significativa ( $M_{post1} = 6.15, SD = 1.73$  vs.  $M_{post2} = 5.30, SD = 1.78, p = .013$ ), para el influencer “pro2” (a favor de la legalización del aborto -perfil 2-) la diferencia entre las puntuaciones promedio resultó marginalmente significativa ( $M_{post1} = 5.90, SD = 1.74$  vs.  $M_{post2} = 5.20, SD = 2.12, p = .079$ ) y para los dos



influencers en contra de la legalización del aborto, “contra1” y “contra2”, la diferencia resultó significativa también (contra1:  $M_{post1} = 6.00$ ,  $SD = 1.65$  vs.  $M_{post2} = 4.40$   $SD = 2.01$ ,  $p < .001$  y contra2:  $M_{post1} = 5.80$ ,  $SD = 2.09$  vs.  $M_{post2} = 5.15$   $SD = 2.98$ ,  $p = .006$ ). La Figura 31 muestra las puntuaciones promedio junto con los intervalos de confianza.

Figura 31: Puntuaciones Promedio de las Preguntas relacionadas a la asociación entre el Influencer y el producto publicitado (por tipo de publicación y tipo de Influencer)



## 5.7 Experimento Principal

Teniendo en cuenta los resultados anteriores, pudimos concluir que los perfiles que maximizan las diferencias entre un influencer a favor de la legalización del aborto y un influencer en contra de la legalización y uno con marca propia vs. otra marca/sin marca son “pro1” (Mery Style @merystyle\_ok) y “contra1” (Camila Chen Martinez @camichenfashion). En particular, en la Figura 29 observamos que las puntuaciones con respecto a la ideología del influencer resultaron compatibles con la ideología que efectivamente queríamos que tenga cada influencer. Luego, en la Figura 31 notamos que las diferencias en puntuación entre la publicación de un producto con marca propia del influencer y otra con otra marca/sin marca resultaron, en su mayoría, significativas para todos los tipos de influencer. No obstante, decidimos quedarnos con “pro1” dado que la diferencia resultó significativa con un *p-valor* más pequeño ( $p = .013$ ), y con “contra1”, por la misma razón ( $p < .001$ ).

Esta selección de influencers nos servirá para construir nuevos perfiles para el experimento principal. Utilizaremos las descripciones de cada uno de los influencers elegidos junto con los *emojis* y *hashtags* que transmitan la ideología propia de cada perfil. No obstante, ambos perfiles tendrán la misma foto y nombre de usuario (en particular, elegimos el nombre y foto de Camila Chen Martinez @camichenfashion) debido a que no queremos que la imagen y nombre utilizados puedan sesgar nuestros resultados finales. A su vez, para la publicación en donde se exhiba un producto que no sea de la marca del influencer, es decir, que no esté relacionado con su persona, utilizaremos el producto de otra marca. En otras palabras, para evaluar el nivel de autoexpresión del influencer con el producto y cómo esto influye en la disposición a comprar, pagar, recomendar y republicar del individuo, usaremos una publicación con un producto de la marca del influencer y otra publicación con un producto de otra marca. Esta decisión fue tomada ya que la marca indefinida no nos parecía interesante para nuestros experimentos ni tampoco realista de utilizar, dada la actividad de promoción que realizan los influencers actuales.

Las imágenes finales de los influencers para el experimento final pueden verse en el Apéndice. Es así como, en total, contamos con **cuatro** configuraciones posibles: tenemos dos perfiles de influencers con diferentes ideologías, y cada uno de ellos cuenta con dos publicaciones, una de la marca propia del influencer y otra de otra marca.

## 5.8 Diseño Experimental

El objetivo principal de este experimento final es poder evaluar el efecto de la similitud en términos ideológicos entre influencer y su seguidor sobre la disposición a comprar, pagar, recomendar y republicar de este último. Para hacer esto, le mostraremos aleatoriamente a los participantes del experimento uno de los cuatro perfiles previamente seleccionados y contruidos a partir de los resultados del pretest. Por lo cual, cada participante verá el perfil del influencer a favor de la legalización del aborto o el perfil del influencer en contra, junto con alguna de las dos publicaciones posibles (marca propia vs. otra marca).

Luego, para obtener información sobre la reacción de los participantes al ver los perfiles, les realizaremos una serie de preguntas. Estas pueden encontrarse detalladas en el Apéndice. A partir de estas, contaremos con una medida de disposición a comprar, de disposición a recomendar y de disposición a republicar para cada participante del experimento una escala de Likert de 7-puntos. Con respecto a la disposición a pagar, simplemente le preguntaremos a cada participante cuánto está dispuesto a pagar por el producto publicitado por el influencer en pesos argentinos (ARS).

A su vez, haremos otra serie de preguntas orientadas a entender la personalidad del consumidor. Esta parte del cuestionario está orientada a entender la identidad moral del participante (si este se siente una persona con ciertos valores - *trabajadora, afectuosa, amigable, servicial*, etc. - o si quiere serlo, y cómo lo simboliza). Con las respuestas que se obtengan de esta parte de la encuesta, construiremos una “Escala de Identidad Moral” o, en inglés, *Moral Identity Scale* (Aquino y Reed 2002).

Por último, agregaremos preguntas demográficas para saber la edad de los participantes, su género, su ideología con respecto a la legalización del aborto (1= definitivamente en contra de la legalización del aborto, 7=definitivamente pro legalización del aborto), su interés en las políticas públicas del país (1=para nada interesado o al tanto de políticas públicas, 7=extremadamente interesado o al tanto de políticas públicas) y su uso de la plataforma Instagram.

Es importante aclarar que el universo de interés para este experimento son mujeres argentinas. La selección de este universo no solo se debe al hecho de que la legalización del aborto afectaría a las mujeres directamente y que por lo tanto estas suelen tener una postura definida sobre este hecho, sino que también hemos descubierto en los datos que la mayoría de los influencers en Instagram son mujeres, teniendo en cuenta que estas se centran en producir contenido de categorías como “fashion” y “beauty”.

Una vez finalizada la construcción de la encuesta con las preguntas descriptas antes, decidimos difundirla a través de LinkedIn, Instagram, Twitter y nuestros conocidos. Esta metodología de muestreo es conocida como *snowball sampling*, ya que incentivamos a aquellos que respondían la encuesta a difundirla también a sus conocidos.

## 5.9 Preprocesamiento y Análisis de las Respuestas

El experimento fue realizado por un total de 229 personas. Nuestro target principal, como hemos establecido anteriormente, son mujeres que utilicen instagram. Ya que el experimento se pudo haber filtrado a individuos fuera de nuestro target, es necesario una limpieza de las observaciones.

### 5.9.1 División de las observaciones por género

En primer lugar, fue necesario filtrar las observaciones que corresponden a hombres. Para ello, analizamos cuántos participantes se habían identificado como hombres y cuántos como mujeres.

Tabla 8: Distribución de los participantes del experimento por género

	Género de los participantes		
	Masculino	Femenino	Otro
Cantidad de participantes	9	219	1

A partir de esto, eliminamos las observaciones del género masculino, quedándonos en total con 220 observaciones.

### **5.9.2 División de las observaciones por uso de Instagram**

En segundo lugar, también consideramos necesario filtrar las observaciones de aquellos participantes que manifestaron que no utilizaban Instagram (respondieron “No” ante la pregunta número # - “¿Tienes una cuenta de instagram?”). A continuación, puede observarse cómo se distribuyen los individuos según su uso de Instagram.

Tabla 9: Respuestas a la pregunta “¿Tienes una cuenta de instagram?”

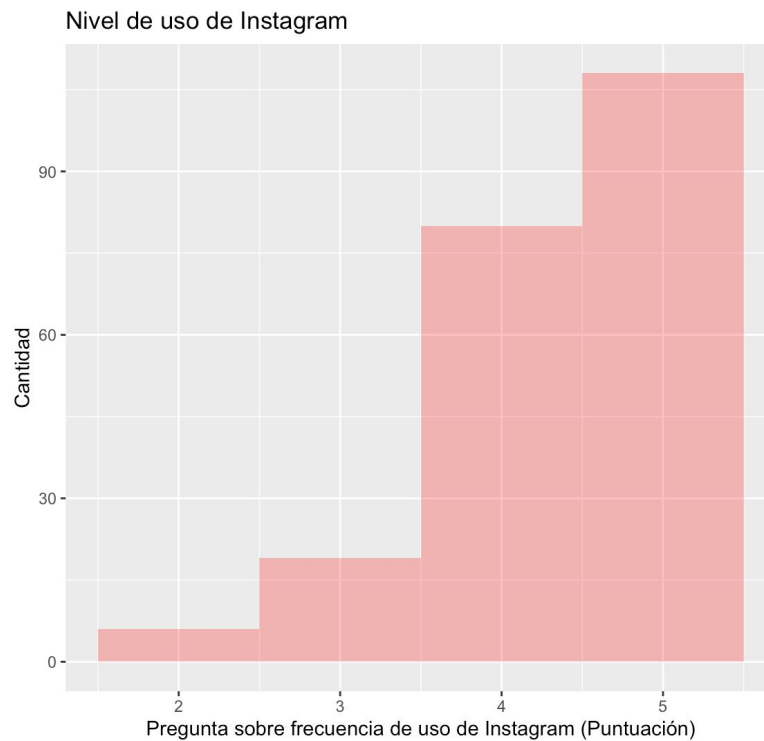
Sí (Tengo cuenta de instagram)	No (No tengo cuenta de instagram)
213	7

Del total de las observaciones luego de filtrar por género, un total de 213 individuos declararon sí ser propietarios de una cuenta en Instagram mientras que solo 7 dijeron no poseer una. Es así como decidimos eliminar estas últimas de la base de datos, y nos quedamos con las 213 respuestas de mujeres con cuenta de Instagram.

### **5.9.3 Frecuencia de uso de Instagram**

Los participantes que sí poseen una cuenta en Instagram, debían también reportar el nivel de uso que le dan a la red social. A continuación, puede verse la distribución de las respuestas según lo respondido por los participantes en la pregunta “¿Qué tan seguido utilizas tu cuenta de instagram?” (1=Nunca, 2=Casi Nunca, 3=A veces, 4=Bastante, 5=Demasiado).

Figura 32 - Respuestas ante la pregunta “¿Qué tan seguido utilizas tu cuenta de instagram?”



Al observar el histograma, podemos ver que una pequeña cantidad de observaciones respondió que “Casi Nunca” utiliza su cuenta de Instagram, pero que la mayoría de los individuos que participaron del experimento utilizan Instagram “Bastante” o “Demasiado”. Por lo cual, consideramos que casi toda la muestra en cuestión conoce cómo utilizar la plataforma, su dinámica y formas de marketing online.

#### 5.9.4 Disposición a pagar en una misma escala

Como ya hemos mencionado, en este experimento algunos participantes tuvieron que responder cuál era su disposición a pagar (en pesos argentinos) por un producto de la marca del influencer o por un producto de otra marca. En particular, la “otra marca” seleccionada fue 47 Street, una marca muy conocida por las mujeres y que varios influencers promocionan actualmente en las redes sociales.

Un potencial problema de estas dos publicaciones es que puede que la disposición a pagar por el producto de otra marca sea mayor a la del producto de la marca propia del influencer, por el simple hecho de que 47 Street es conocida por múltiples personas (tanto por aquellas que tienen Instagram como por aquellas que no) y sus precios son conocidos y públicos, mientras que la marca propia del influencer (que es ficticio) es desconocida y por ende, los individuos no tienen un punto de referencia claro a partir del cual emitir su propia disposición a pagar. De esta manera, sabemos de antemano que puede haber cierto sesgo en la diferencia de disposición a

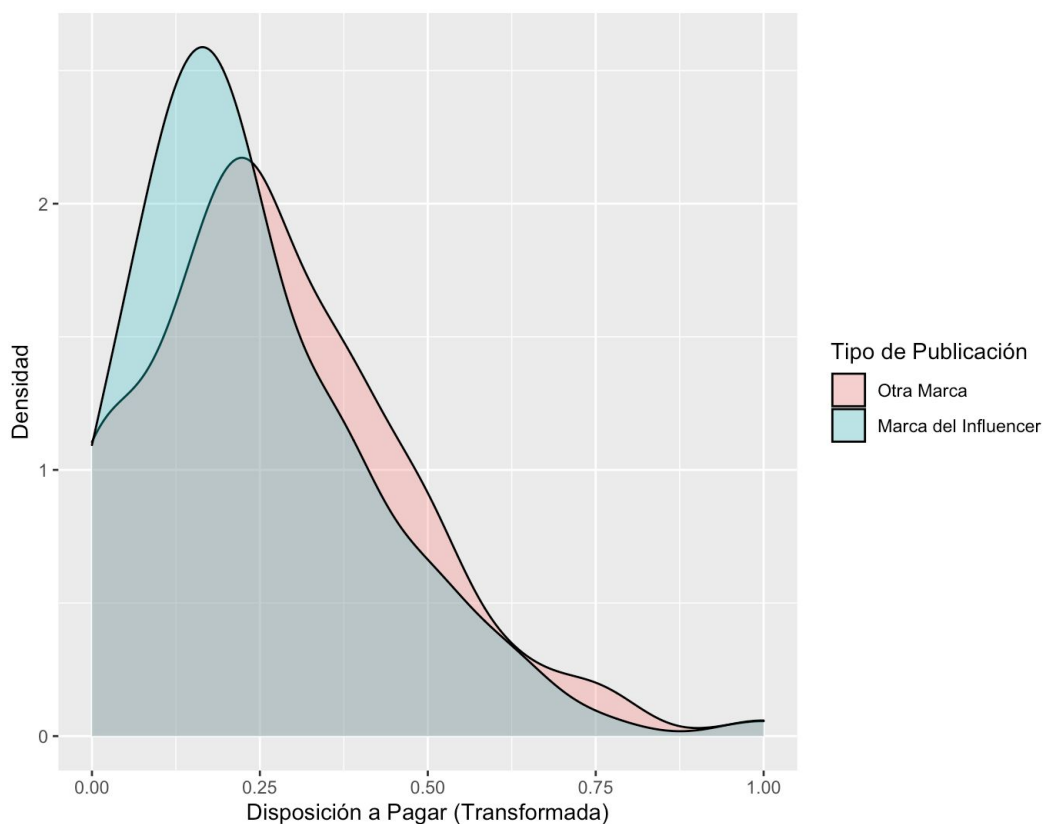
pagar entre la publicación con un producto de la marca del influencer vs. un producto de otra marca.

Es por esta razón que decidimos que, antes de analizar esta variable, debíamos re-escalarla para que puedan ser comparadas las propensiones de pago de una publicación y la otra. No obstante, se deberá utilizar una forma de re-escalado que no modifique la media de las distribuciones para ambos tipos de publicación ya que esto podría perjudicar nuestros análisis posteriores de las medias. De esta manera, re-escalamos la variable de disposición a pagar ( $X$ ) con la siguiente metodología:

$$X_{new} = \frac{X_i - \min(X)}{\max(X) - \min(X)}$$

Donde  $X_i$  representa el valor de la disposición a pagar del participante  $i$ ,  $\min(X)$  representa la disposición a pagar mínima de todas las observaciones y  $\max(X)$  es la máxima disposición a pagar de todas las observaciones. Con este re-escalamiento, ahora las propensiones de pago de cada tipo publicación tomarán valores entre 0 y 1. Las distribuciones de la disposición a pagar, luego de esta transformación, pueden observarse en el siguiente gráfico:

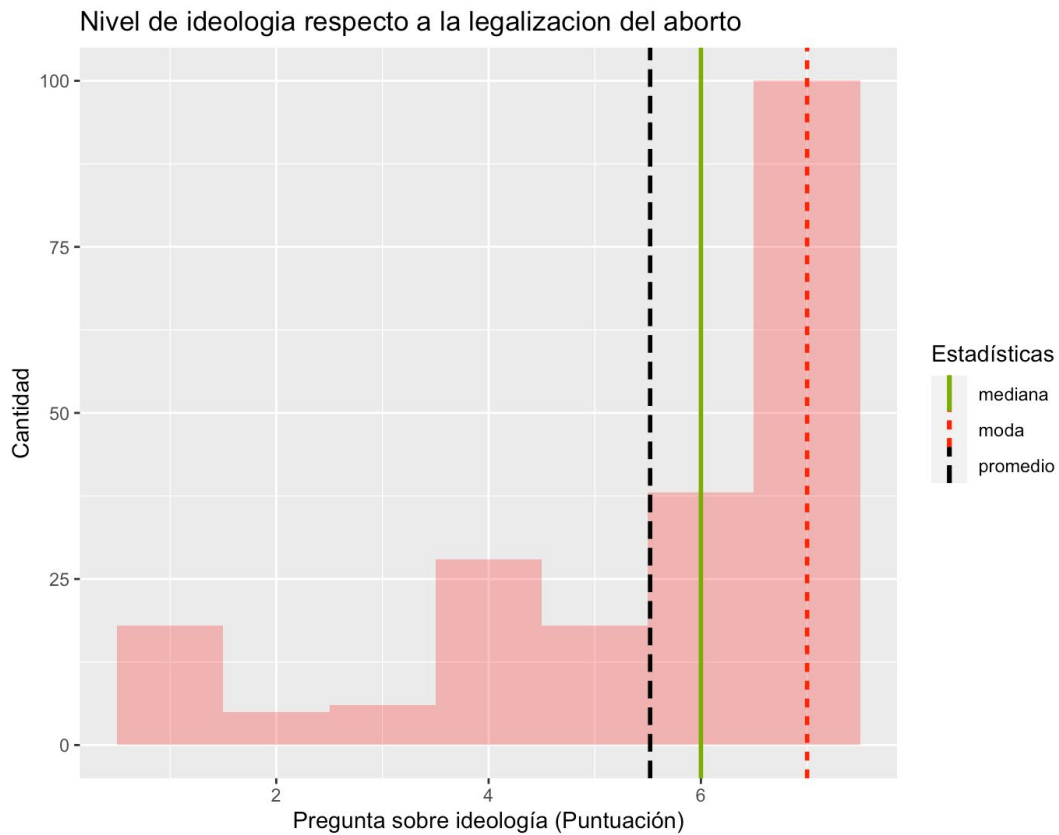
**Figura 33: Distribución de la disposición a pagar re-escalada**



### 5.9.5 Análisis de la ideología de los participantes

Los participantes del experimento tuvieron que contestar si se consideraban una persona a favor de la legalización del aborto o en contra (1= definitivamente en contra de la legalización del aborto, 7= definitivamente pro legalización del aborto). Podemos analizar la distribución de las respuestas obtenidas para caracterizar nuestra muestra, a través del siguiente gráfico:

**Figura 34: Distribución de las respuestas ante la pregunta sobre ideología referida al aborto**



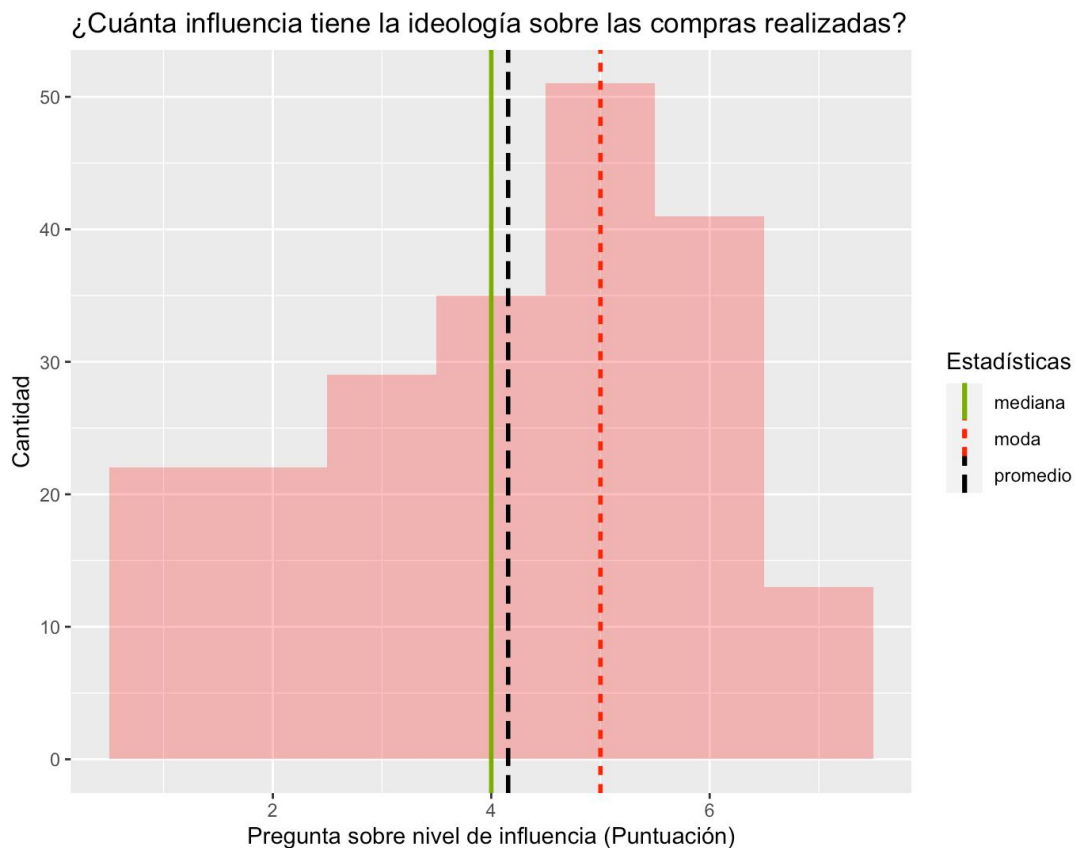
Podemos notar que contamos con una muestra que se encuentra ideológicamente más cerca de estar a favor de la legalización del aborto. De hecho, contamos con aproximadamente 100 individuos que contestaron que estaban definitivamente a favor de la legalización del aborto. La mediana de la distribución se encuentra en 6, la moda en 7 y el promedio en 5.52. Esto podría considerarse como una limitación al momento de realizar los test y análisis posteriores, debido a que contamos con menos muestras de individuos que estén en contra de la legalización del aborto. No obstante, no contábamos con los recursos para realizar una muestra estratificada, y por ende, intentamos recaudar la mayor cantidad de respuestas posibles de manera de obtener la mayor cantidad de individuos en ambos grupos (a favor y en contra).

### 5.9.6 Nivel de influencia de la ideología sobre las decisiones de compra

Una de las preguntas demográficas que le hicimos a los participantes tenía como propósito indagar sobre cuánta influencia tienen las ideas de los individuos al momento de realizar una

compra. En particular, la pregunta era la siguiente: “¿A menudo se ve influenciado por sus principios e ideologías al momento de realizar una compra?” (1= para nada, 7= extremadamente). Al igual que en los casos anteriores, podemos graficar la distribución de las respuestas a esta pregunta.

Figura 35: Distribución de las respuestas a la pregunta “¿A menudo se ve influenciado por sus principios e ideologías al momento de realizar una compra?”



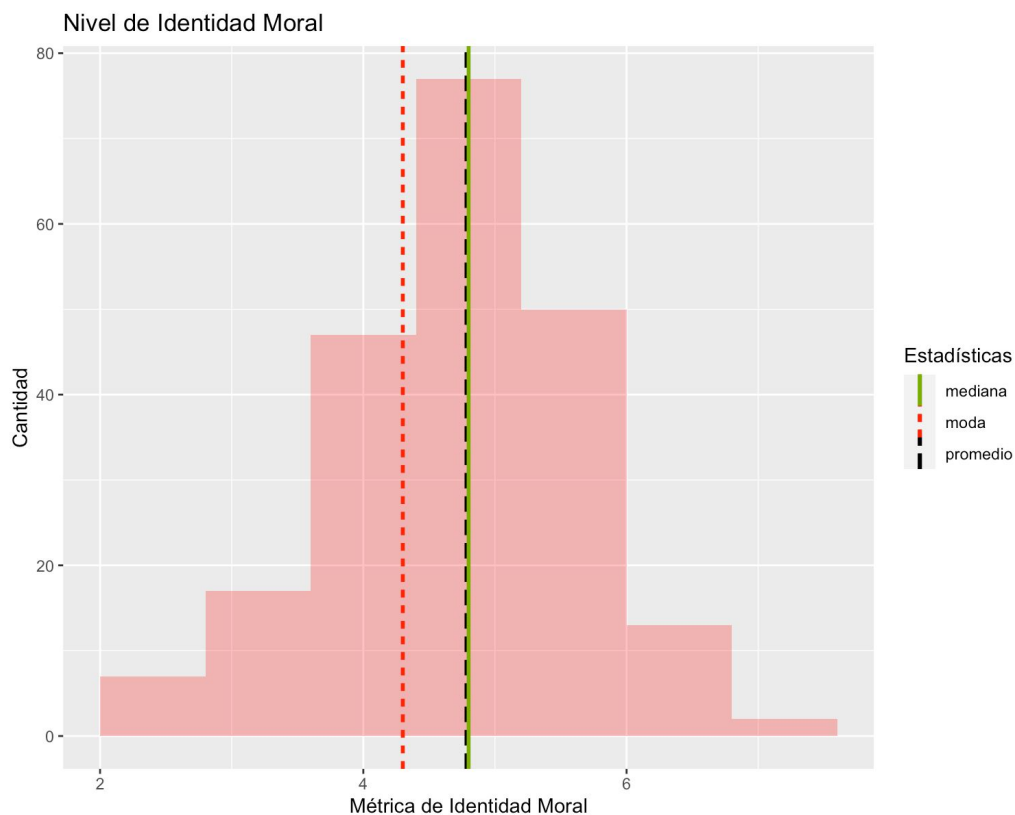
La mediana de las respuestas se encuentra en 4, el promedio en 4.15 y la moda en 5. Por lo cual, observamos que en promedio, los individuos no se perciben como personas influenciadas por sus ideas al momento de tomar decisiones de consumo. Mirando únicamente este gráfico, podríamos concluir que la similitud ideológica entre un influencer y su seguidor no debería influenciar significativamente la disposición a comprar de este último. No obstante, debemos analizar más en profundidad las demás variables para poder obtener una mejor respuesta a nuestra pregunta de investigación. A su vez, hay que tener en cuenta que estas son las percepciones de los participantes. Puede que los resultados finales sí muestren cierto nivel de influencia de la ideología sobre las compras.



### 5.9.7 Nivel de Identidad Moral de los participantes

A partir de las respuestas obtenidas de las preguntas de la sección “Escala de Identidad Moral” (ver Apéndice), realizamos un promedio de todas ellas para formar una nueva variable agregada que buscará medir cuán deseosos están los participantes por llegar a ser un “modelo” de persona moral (cuyas características son *afectuoso, compasivo, justo, amigable, generoso, servicial, trabajador, honesto y amable*). De esta manera, cuanto mayor sea el valor de esta variable promedio, a la que llamaremos Nivel de Identidad Moral, más deseoso estará el participante por ser el modelo de persona moral y más querrá mostrarlo. Decidimos realizar un gráfico de la distribución de esta variable, para poder tener una visión general del Nivel de Identidad Moral de que exhibieron los participantes.

Figura 36: Distribución del Nivel de Identidad Moral



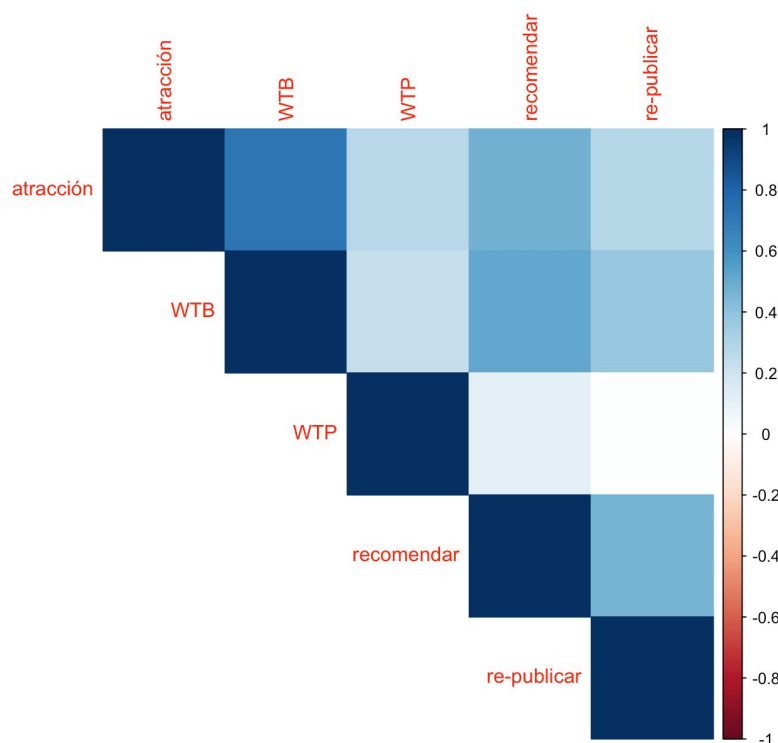
La distribución parece ser bastante simétrica, centrándose en el promedio de Nivel de Identidad Moral igual a 4.78. La mediana se ubica en 4.80, muy similar al promedio, mientras que la moda se ubica en 4.3. Podemos concluir que los participantes no parecen tener una tendencia fuerte a desear ser este modelo de persona moral descrito anteriormente, y que tampoco buscar mostrarlo demasiado. Parece que los participantes, en promedio, son neutrales con respecto a querer ser una persona con las características morales enumeradas.

Para confirmar que la variable Nivel de Identidad Moral, construida a partir de los promedios de las respuestas de cada participante en la sección sobre Identidad Moral de la encuesta, es una escala confiable, corrimos un un análisis de fiabilidad sobre todas las respuestas de la sección. Encontramos que el Alfa de Cronbach (Cronbach 1951)<sup>6</sup> es un valor suficientemente alto ( $\alpha = .78$ ), por lo cual tenemos evidencia suficiente para afirmar que la escala construida es confiable.

### 5.9.8 Correlación entre las respuestas referidas a la publicación del influencer

Los participantes observaron aleatoriamente uno de los perfiles ficticios junto con un tipo de publicación (marca propia del influencer vs. otra marca), y luego contestaron preguntas específicas sobre cuán atractivo les parecía el producto mostrado, cuánto estarían dispuestos a pagar por el producto, el nivel de disposición a comprar, recomendar y republicar. A continuación, nos pareció interesante observar las correlaciones entre las respuestas a estas preguntas.

Figura 37: Matriz de Correlación entre las respuestas referidas a la publicación



<sup>6</sup> El Alfa de Cronbach es un coeficiente que se utiliza para evaluar la fiabilidad de una escala de medida. Fue ideada por Lee Cronbach en 1951. Este coeficiente apareció por primera vez en el paper *Coefficient Alpha and the internal structure of tests*, por Lee J. Cronbach, y establece que su importancia radica en que “cualquier investigación basada en alguna medida debe tener exactitud o dependencia o, como lo llamamos comúnmente, fiabilidad de medida. Un coeficiente de fiabilidad demuestra si el diseñador del experimento estuvo en lo correcto al esperar que cierta colección de items den por resultado declaraciones interpretables acerca de las diferencias individuales”.

Las etiquetas que aparecen en la matriz hacen referencia a las preguntas relacionadas a la publicación. La palabra “atracción” hace referencia a la respuesta a la pregunta “¿Cuán atractivo te parece el producto que se muestra en la publicación?” (1= nada atractivo, 7= muy atractivo), “WTB” hace referencia a la disposición a comprar (*Willingness To Buy*), “WTP” refiere a la disposición a pagar (*Willingness To Pay*), “recomendar” refiere a la disposición a recomendar y “republicar” hace referencia a la disposición a republicar.

Al poner atención en las correlaciones, vemos que la disposición a comprar está altamente relacionado con cuán atractivo es el producto para el individuo. A su vez, cuán atractivo se ve el producto correlaciona con la disposición a recomendar. Esta última se relaciona positivamente con la disposición a republicar, mientras que la disposición a comprar también correlaciona con la disposición a recomendar y republicar.

## 5.10 Análisis de los Resultados

Una vez que procesamos los datos obtenidos del experimento, recurrimos a hacer una serie de estudios de regresión para poder determinar cómo influyen las variables independientes de interés sobre las que consideramos variables dependientes.

Las variables independientes (o, en algunos casos, de control) dentro del dataset de respuestas son las siguientes:

- *Brand*: variable binaria que toma el valor 0 cuando la publicación del influencer que vió el participante es de otra marca, y toma el valor 1 cuando la publicación es de la marca propia del influencer. Esta variable fue transformada al tipo factor, de manera tal que se identifica como variable categórica siendo 0 la categoría “Otra Marca” y 1 la categoría “Marca del Influencer”.
- *Value*: variable binaria que toma el valor 0 si el influencer que vió el participante está en contra de la legalización del aborto, y toma el valor 1 cuando el influencer está a favor de la legalización del aborto. Esta variable fue transformada al tipo factor, de manera tal que se identifica como variable categórica siendo 0 la categoría “Influencer Pro-Vida”<sup>7</sup> y 1 la categoría “Influencer Pro-Aborto”.

---

<sup>7</sup> El término Pro-Vida (o en inglés *pro-life*) hace referencia a aquellas personas cuya posición moral es tal que están en contra de la legalización del aborto y su práctica, debido a que atenta contra la vida de un ser humano en formación. De esta forma, estas personas afirman que la vida humana comienza al momento de ocurrir la fecundación.

- *Abort*: variable que puede tomar valores enteros entre 1 y 7. En particular, es la variable que refleja las respuestas de los participantes a la pregunta sobre su ideología con respecto a la legalización del aborto (1= definitivamente en contra de la legalización del aborto, 7= definitivamente pro legalización del aborto). Dado esto, los participantes que hayan elegido responder con los números 1 o 7, representan las posturas extremas o polos opuestos de la muestra.
- *Moral Identity*: variable que surge de tomar el promedio de todas las respuestas que dio un participante a las preguntas de la sección sección “Escala de Identidad Moral” (ver Apéndice). De esta forma, cuanto mayor sea el valor de esta variable, mayor será el deseo del participante de ser una persona con ciertos valores morales (ya descriptos antes) y de querer mostrarlo.
- *Influence*: variable que puede tomar valores enteros entre 1 y 7. También refleja las respuestas de los participantes, pero esta vez a la pregunta sobre cuánta influencia tienen sus ideas al momento de realizar una compra (1= para nada, 7= extremadamente) Por lo cual, aquellos que hayan respondido con valores muy cercanos a 1 son aquellos que se consideran muy independientes de sus ideas y principios al momento de realizar una compra y, por el contrario, los que hayan respondido con valores muy cercanos a 7, son los participantes que se sienten altamente influenciados por sus ideas al momento de comprar.

Por otro lado, las variables dependientes son:

- *Willingness To Pay (WTP)*: variable que mide la disposición a pagar de los participantes del experimento. Anteriormente aclaramos que esta variable estaba medida en pesos argentinos, pero luego fue sometida a un re-escalamiento para que la disposición a pagar por productos de la marca del influencer sea comparable con la disposición a pagar por productos de otra marca. En consecuencia, la disposición a pagar ahora tomará valores entre 0 y 1.
- *Willingness To Buy (WTB)*: variable que mide la disposición a comprar (el producto expuesto por el influencer) de los participantes del experimento. Surge de la respuesta a la pregunta “¿Cuán probable es que compres el producto que se muestra en la publicación?” (1 = poco probable, 7 = muy probable). Por ende, la disposición a comprar tomará valores enteros entre 1 y 7.
- *Willingness To Recommend (recommend)*: variable que mide la disposición a recomendar el influencer a otras personas. Al igual que la disposición a comprar, tomará valores enteros entre 1 y 7, donde el número 1 representará una muy baja probabilidad de recomendar al influencer mientras que el número 7 representa una probabilidad alta de recomendarlo.

- *Willingness To Repost (repost)*: variable que mide la disposición a republicar el contenido publicado por el influencer que vio el participante. Surge de la respuesta a la pregunta “¿Qué tan probable es que reposteas contenido publicado por este influencer?” (1 = poco probable, 7 = muy probable), por lo que la disposición a republicar también tomará valores enteros entre 1 y 7.

A partir de las variables con las que contamos, quisimos medir el efecto de las variables independientes sobre las variables dependientes. En otras palabras, queremos medir el efecto diferencial que tienen aquellos participantes que son similares a los influencers que vieron sobre las variables de interés, como la disposición a comprar, pagar, recomendar o republicar. Nuestra intuición es que individuos que tienen una mayor similitud ideológica con los influencers van a tener una mayor disposición a pagar, comprar, recomendar y republicar publicaciones de marcas propias del influencer que publicaciones de otra marca, ya que al existir similitud, el consumidor preferirá aquella publicación con la que pueda identificarse o acercarse aún más al influencer. De manera contraria, aquellos individuos que no se sientan similares ideológicamente a los influencers, van a preferir el producto publicitado de otra marca, de manera que puedan hacer la compra pero sin estar altamente relacionados con el influencer y su ideología.

De esta forma, podemos expresar nuestra hipótesis principal al llevar a cabo este estudio:

*Hipótesis: Cuanto más identificado o similar se sienta el individuo con respecto al influencer, mayor será su disposición a comprar y su disposición a pagar por productos publicitados de la marca propia del influencer. A su vez, mayor será su disposición a recomendar al influencer y republicar su contenido.*

Para evaluar si nuestra hipótesis es correcta, procederemos a hacer una serie de regresiones con las variables descritas anteriormente.

### **5.10.1 Análisis de Regresión: Disposición a Comprar (WTB)**

Como ya hemos mencionado, nuestra intención es analizar si la similitud entre el influencer y el consumidor tienen un efecto sobre la disposición a comprar el producto publicitado. Teniendo en cuenta todas las variables que surgieron a partir del desarrollo del experimento, decidimos realizar la siguiente regresión:

$$(1) \quad WTB = \beta_0 + \beta_1 Brand + \beta_2 Abort + \beta_3 Value + \beta_4 (Brand \times Abort) + \beta_5 (Value \times Abort) + \beta_6 (Value \times Brand) + \beta_7 (Value \times Brand \times Abort)$$

Notemos que tanto *Brand* como *Value* son variables binarias, por lo cual, esta regresión está determinando cuatro rectas posibles según los valores que pueden tomar *Brand* y *Value*. A

continuación, mostraremos las rectas que se forman a partir de los distintos valores que pueden tomar las dos variables binarias.

Tabla 10: Modelos de Regresión según los valores de las variables binarias

	<b>Influencer Pro-Vida</b> ( <i>Value=0</i> )	<b>Influencer Pro-Aborto</b> ( <i>Value=1</i> )
<b>Otra Marca</b> ( <i>Brand=0</i> )	$WTB = \beta_0 + \beta_2 Abort$	$WTB = \beta_0 + \beta_3 + (\beta_2 + \beta_5) Abort$
<b>Marca del Influencer</b> ( <i>Brand=1</i> )	$WTB = \beta_0 + \beta_1 + (\beta_2 + \beta_4) Abort$	$WTB = \beta_0 + \beta_1 + \beta_3 + \beta_6 + (\beta_2 + \beta_4 + \beta_5 + \beta_7) Abort$

Puede observarse que ante las cuatro combinaciones posibles de perfil de influencer y tipo de publicación, tenemos diferentes rectas de regresión, con diferentes ordenadas al origen y diferente pendiente. Nuestro interés está el efecto marginal total de la variable *Abort*, es decir, que deseamos saber cuál es el efecto de la ideología del consumidor, dada la ideología del influencer sobre la disposición a comprar. Como vemos, si los coeficientes son significativos, podremos comparar las pendientes y evaluar si son positivas o negativas. Es importante aclarar que cuanto mayor sea la variable *Abort*, más a favor de la legalización del aborto estará el individuo, y por ende será más similar al Influencer Pro-Aborto; por el contrario, cuando menor sea la variable *Abort*, el individuo estará más en contra de la legalización del aborto y por ende será más similar al Influencer Pro-Vida.

Utilizando R, corrimos la regresión (1). En la tabla siguiente pueden observarse los coeficientes estimados, errores estándar, t-estadísticos y p-valor para cada coeficiente.

Tabla 11: Efecto de la Similitud entre Influencer y Consumidor sobre la Disposición a Comprar

Variables	Coefficientes Estimados	Errores Estándar (SE)	<i>t</i>	<i>p</i>
<i>Brand</i>	2.448	0.9779	2.503	0.01308 **
<i>Abort</i>	0.2902	0.1177	2.466	0.01447 **
<i>Value</i>	1.6176	0.9519	1.699	0.09078 *
<i>Brand x Abort</i>	-0.4654	0.1695	-2.746	0.00657 ***
<i>Value x Abort</i>	-0.2529	0.1624	-1.557	0.12091
<i>Value x Brand</i>	-2.5940	1.3419	-1.933	0.05461 *
<i>Value x Brand x Abort</i>	0.4938	0.2302	2.145	0.03309 **
Constante	0.5088	0.6726	0.756	0.45023

Nota: \*\*\*, \*\*, \* indican significatividad al 1%, 5% y 10% respectivamente. *N* = 213.  
Potencia *post hoc* = 0.655

Los resultados obtenidos indican que gran parte de los coeficientes estimados son significativos, lo cual nos afirma que efectivamente existe un efecto significativo de la similitud sobre la disposición a comprar. No obstante, recordemos que lo que nos interesa son los efectos marginales totales. Esto puede calcularse en R fácilmente, por lo que mostraremos a continuación los efectos marginales totales para cada caso expuesto en la Tabla 10.

Tabla 12: Efecto Marginal Total de la variable *Abort*

Modelo	Efecto Marginal Total de <i>Abort</i>	Errores Estándar (SE)	<i>t</i>	<i>p</i>
Otra Marca   Influencer Pro-Vida ( <i>Brand</i> =0, <i>Value</i> =0)	0.29019	0.11766	2.466	0.0145 **
Otra Marca   Influencer Pro-Aborto ( <i>Brand</i> =0, <i>Value</i> =1)	0.03725	0.11195	0.333	0.7397
Marca del Influencer   Influencer Pro-Vida ( <i>Brand</i> =1, <i>Value</i> =0)	-0.17519	0.12196	-1.437	0.1524
Marca del Influencer   Influencer Pro-Aborto ( <i>Brand</i> =1, <i>Value</i> =1)	0.06565	0.10826	0.606	0.5449

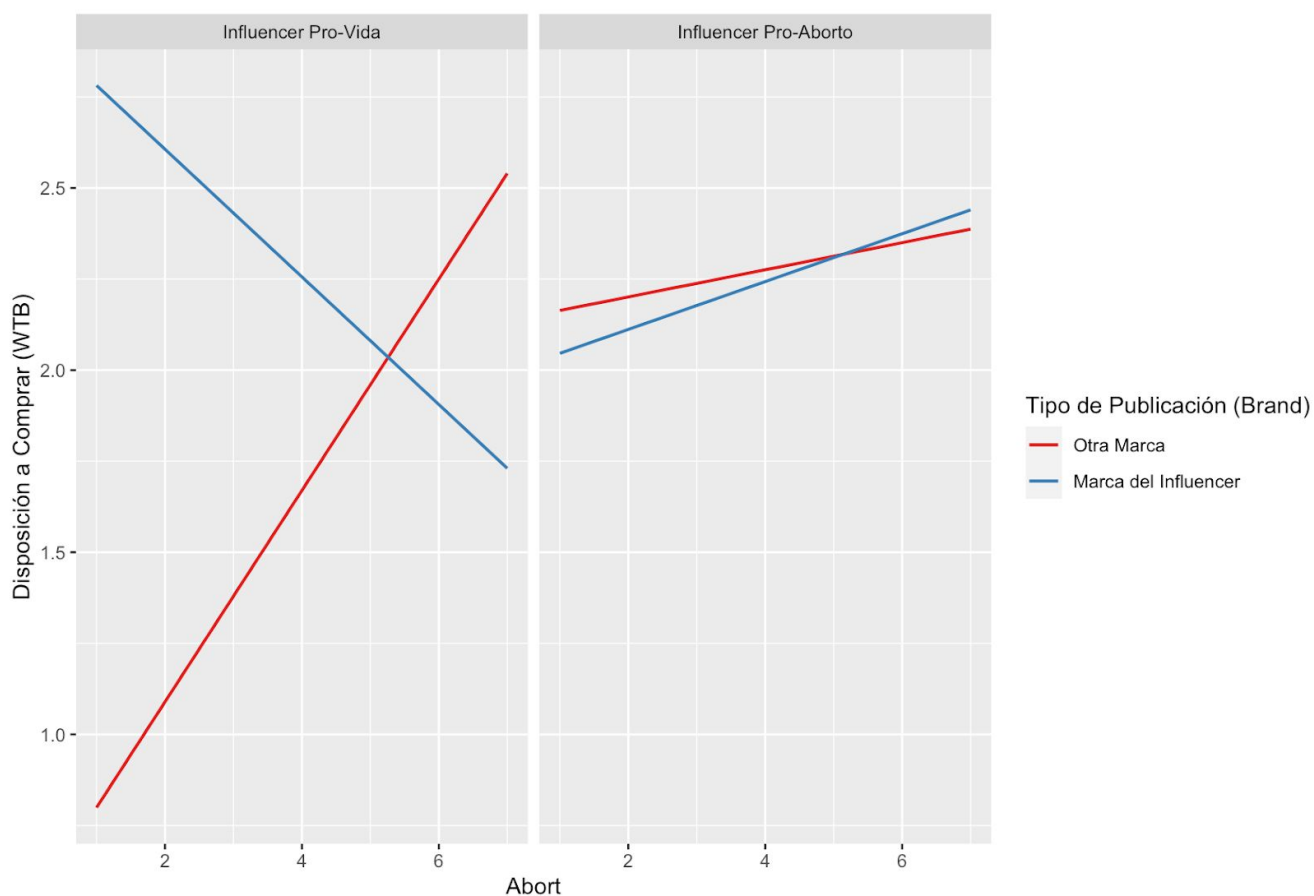
Nota: \*\*\*, \*\*, \* indican significatividad al 1%, 5% y 10% respectivamente. *N* = 213.

Los efectos marginales totales que se muestran arriba, corresponden a las estimaciones de las pendientes de las rectas que describimos en la Tabla 10. De esta forma, por ejemplo, el efecto marginal de la variable *Abort* para el modelo “Marca del Influencer | Influencer Pro-Aborto” es igual a la suma de los coeficientes  $\beta_2$ ,  $\beta_4$ ,  $\beta_5$ ,  $\beta_7$  estimados en la Tabla 11. Lo que puede

deducirse a partir de la información de la Tabla 12 es que el único efecto marginal total significativo es el correspondiente a "Otra Marca | Influencer Pro-Vida". Pero ¿cómo podemos interpretar estos resultados?

Para poder hacer un análisis más intuitivo de los resultados que obtuvimos, procederemos a graficar las cuatro rectas correspondientes a los cuatro modelos de regresión que componen a la regresión (1). De esta manera, podremos visualizar más claramente lo expuesto hasta ahora.

Figura 38: Modelos de Regresión para medir el Efecto de la Similitud sobre la Disposición a Comprar de los Consumidores



Observando atentamente el gráfico anterior, podemos notar que, **para el caso en el cual los participantes del experimento vieron al Influencer Pro-Vida**, la disposición a comprar el producto publicitado de Otra Marca disminuye a medida que la variable *Abort* va disminuyendo. En otras palabras, cuanto más en contra de la legalización del aborto esté el individuo, más disposición a comprar tendrá por la Marca del Influencer Pro Vida y menos disposición a comprar tendrá por la Otra Marca. Sin embargo, cuando el individuo vio al Influencer Pro-Aborto, no parecen haber efectos demasiado significativos.

El hecho de que el único efecto marginal total significativo de *Abort* fue el del modelo "Otra Marca | Influencer Pro-Vida", nos está diciendo que la pendiente de la recta de color rojo que se



encuentra a la izquierda es significativamente distinta de cero. Si bien el modelo “Marca del Influencer | Pro Vida” no tiene una pendiente significativa (es decir, el efecto marginal total de *Abort*), su ordenada al origen ( $\beta_0 + \beta_1$ ) sí es significativa considerando que  $\beta_1$  es significativa. Esto puede comprobarse agregando los intervalos con un 90% de confianza a las estimaciones de las rectas (Ver este gráfico en el Apéndice).

De esta forma, puede concluirse que para aquellos participantes en contra de la legalización del aborto que vieron una publicación de un influencer Pro-Vida, la disposición a comprar es mayor para la publicación de la Marca del Influencer que para la de Otra Marca, de manera que en este caso, la similitud sí genera una tendencia a querer comprar el producto propio del Influencer.

No obstante, parece que esta relación no aparece para el caso de similitudes entre el Influencer Pro-Aborto y los participantes a favor de la legalización. Una posible explicación de esto es que ambos grupos (Pro-Vida y Pro-Aborto) sean distintos entre sí en términos de sus características y su reacción ante una persona que piensan diferente. Los resultados obtenidos parecen mostrar que los individuos que se manifestaron como Pro-Aborto no manifiestan cambios en su disposición a comprar al momento de ver un Influencer Pro-Vida. Por ende, estos consumidores no parecen darle tanta importancia a la similitud.

Si controlamos por las variables *Moral Identity* o *Influence*, los resultados se mantienen. En el primer caso, simplemente agregamos a la regresión (1) la variable *Moral Identity* como covariable o variable de control. De esta forma, estaríamos aislando el efecto propio de tener el deseo de parecerse a un modelo de persona moral y querer mostrarlo a través de la vestimenta o de actividades, sobre la disposición a comprar. Así, al controlar por esta variable, obtendremos una estimación aún más precisa del efecto marginal total de la ideología del individuo (*Abort*) sobre la disposición a comprar el producto. A continuación mostramos las estimaciones de los coeficientes y los efectos marginales totales.

Tabla 13: Efecto de la Similitud entre Influencer y Consumidor sobre la Disposición a Comprar controlando por Identidad Moral

Variables	Coefficientes Estimados	Errores Estándar (SE)	<i>t</i>	<i>p</i>
<i>Brand</i>	2.5242	0.9711	2.599	0.01002 **
<i>Abort</i>	0.3026	0.1169	2.588	0.01034 **
<i>Value</i>	1.7121	0.9458	1.810	0.07173 *
<i>Moral Identity</i>	0.2516	0.1231	2.043	0.04234 *
<i>Brand x Abort</i>	-0.4663	0.1682	-2.773	0.00607 ***
<i>Value x Abort</i>	-0.2675	0.1613	-1.658	0.09879 *
<i>Value x Brand</i>	-2.8014	1.3355	-2.098	0.03717 **
<i>Value x Brand x Abort</i>	0.5281	0.229	2.306	0.02211 **
Constante	-0.7998	0.9251	-0.865	0.38827

Nota: \*\*\*, \*\*, \* indican significatividad al 1%, 5% y 10% respectivamente. *N* = 213.  
Potencia *post hoc* = 0.825

Tabla 14: Efecto Marginal Total de la variable *Abort* al controlar por Identidad Moral

Modelo	Efecto Marginal Total de <i>Abort</i>	Errores Estándar (SE)	<i>t</i>	<i>p</i>
Otra Marca   Influencer Pro-Vida ( <i>Brand</i> =0, <i>Value</i> =0)	0.30263	0.11692	2.588	0.0103 **
Otra Marca   Influencer Pro-Aborto ( <i>Brand</i> =0, <i>Value</i> =1)	0.03510	0.11109	0.3160	0.7523
Marca del Influencer   Influencer Pro-Vida ( <i>Brand</i> =1, <i>Value</i> =0)	-0.16366	0.12115	-1.351	0.1782
Marca del Influencer   Influencer Pro-Aborto ( <i>Brand</i> =1, <i>Value</i> =1)	0.09694	0.10852	0.893	0.3727

Nota: \*\*\*, \*\*, \* indican significatividad al 1%, 5% y 10% respectivamente. *N* = 213.

Al agregar *Influence* como variable de control, los resultados también se mantuvieron, aunque el coeficiente correspondiente a la variable de control no parece ser significativo. Esto nos estaría diciendo que el hecho que los participantes se consideren personas que se ven influenciadas por sus ideas y principios al momento de realizar una compra, no parece realmente tener un efecto sobre la disposición a comprar.

Tabla 15: Efecto de la Similitud entre Influencer y Consumidor sobre la Disposición a Comprar controlando por Nivel de Influencia

Variables	Coefficientes Estimados	Errores Estándar (SE)	<i>t</i>	<i>p</i>
<i>Brand</i>	2.48316	0.98484	2.521	0.01245 **
<i>Abort</i>	0.29050	0.11792	2.464	0.01458 **
<i>Value</i>	1.60418	0.9547	1.680	0.09443 *
<i>Influence</i>	-0.02302	0.06418	-0.359	0.72017
<i>Brand x Abort</i>	-0.46881	0.17009	-2.756	0.00638 ***
<i>Value x Abort</i>	-0.25003	0.16296	-1.534	0.12649
<i>Value x Brand</i>	-2.6201	1.34674	-1.946	0.05309 *
<i>Value x Brand x Abort</i>	0.49782	0.23092	2.156	0.03227 **
Constante	0.59432	0.71493	0.831	0.40678

Nota: \*\*\*, \*\*, \* indican significatividad al 1%, 5% y 10% respectivamente. *N* = 213.  
Potencia *post hoc* = 0.659

Tabla 16: Efecto Marginal Total de la variable *Abort* al controlar por Nivel de Influencia

Modelo	Efecto Marginal Total de <i>Abort</i>	Errores Estándar (SE)	<i>t</i>	<i>p</i>
Otra Marca   Influencer Pro-Vida ( <i>Brand</i> =0, <i>Value</i> =0)	0.2905	0.11792	2.464	0.0146 **
Otra Marca   Influencer Pro-Aborto ( <i>Brand</i> =0, <i>Value</i> =1)	0.04047	0.11254	0.3600	0.7196
Marca del Influencer   Influencer Pro-Vida ( <i>Brand</i> =1, <i>Value</i> =0)	-0.17831	0.12252	-1.455	0.1471
Marca del Influencer   Influencer Pro-Aborto ( <i>Brand</i> =1, <i>Value</i> =1)	0.06948	0.10901	0.637	0.5246

Nota: \*\*\*, \*\*, \* indican significatividad al 1%, 5% y 10% respectivamente. *N* = 213.

Como último comentario sobre el análisis anterior, es importante resaltar el valor de la potencia para cada una de las regresiones llevadas a cabo. La potencia estadística se define como la habilidad de un test de hipótesis de detectar un efecto que existe en la población, es decir, puede describirse como la probabilidad de detectar un efecto “real” o “verdadero” cuando este existe, dado un número determinado de muestra<sup>8</sup>. En la práctica, el análisis de potencia se utiliza para

<sup>8</sup> Para una probabilidad de error de tipo II de  $\beta$ , la potencia estadística correspondiente es  $1 - \beta$ . Por ejemplo, si el experimento E tiene una potencia estadística de 0.7 y el experimento F tiene una potencia estadística de 0.95, entonces hay una mayor probabilidad de que el experimento E tenga un error de tipo II

determinar el tamaño de la muestra para un trabajo de investigación, de manera que ese tamaño sea lo suficientemente significativo para encontrar resultados. En general, un valor de potencia de 0.80 es considerado suficiente para afirmar que el tamaño de muestra es tal que los resultados encontrados son significativos y pueden aplicar a la población analizada. En este caso particular, hicimos un análisis de potencia post hoc<sup>9</sup> para todas las regresiones planteadas. Puede observarse tanto en la Tabla 11, 13 y 15 que la potencia toma los valores 0.655, 0.825 y 0.659, pudiéndose concluir que el incluir la variable de identidad moral MID resulta en un análisis con una mayor probabilidad de que los resultados sean detectables en la población. No consideramos que los valores 0.655 y 0.659 minimicen los resultados encontrados, pero sí nos indican que el tamaño de muestra debería aumentarse en futuras investigaciones. Este análisis post hoc se llevó a cabo utilizando la función `pwr.f2.test()` en R.

### 5.10.2 Análisis de Regresión: Disposición a Pagar (WTP), Disposición a Recomendar y Republicar

Siguiendo la misma metodología, hicimos los análisis de regresión correspondientes a la disposición a pagar, recomendar y republicar.

$$(2) \quad WTP = \beta_0 + \beta_1 Brand + \beta_2 Abort + \beta_3 Value + \beta_4 (Brand \times Abort) + \beta_5 (Value \times Abort) \\ + \beta_6 (Value \times Brand) + \beta_7 (Value \times Brand \times Abort)$$

---

que el experimento F. Esto reduce la sensibilidad del experimento E para detectar efectos significativos. Sin embargo, el experimento E es, en consecuencia, más confiable que el experimento F debido a su menor probabilidad de un error de tipo I.

<sup>9</sup> El análisis de potencia se puede realizar antes (a priori) o después (post hoc) de la recopilación de datos. El análisis de potencia a priori se lleva a cabo antes del estudio de investigación y, por lo general, se utiliza para estimar tamaños de muestra suficientes para lograr la potencia adecuada. El análisis post-hoc de la "potencia observada" se realiza después de que se ha completado un estudio y utiliza el tamaño de la muestra y el tamaño del efecto obtenidos para determinar cuál fue la potencia en el estudio, asumiendo que el tamaño del efecto en la muestra es igual al tamaño del efecto en la población. Mientras que la utilidad del análisis de poder prospectivo en el diseño experimental es universalmente aceptada, el análisis de poder post hoc es fundamentalmente defectuoso (Hoenig & Heisey 2001). Caer en la tentación de utilizar el análisis estadístico de los datos recopilados para estimar el poder resultará en valores poco informativos y engañosos. En particular, se ha demostrado que el "poder observado" post hoc es una función uno a uno del valor p alcanzado. Esto se ha ampliado para mostrar que todos los análisis de poder post hoc sufren de lo que se llama la "paradoja del enfoque de poder" (PAP), en la que se cree que un estudio con un resultado nulo muestra más evidencia de que la hipótesis nula es realmente cierta cuando el valor p es menor, ya que el poder aparente para detectar un efecto real sería mayor. De hecho, en general se entiende que un valor p más pequeño hace que sea relativamente menos probable que la hipótesis nula sea cierta.

$$(3) \text{ recommend} = \beta_0 + \beta_1 \text{Brand} + \beta_2 \text{Abort} + \beta_3 \text{Value} + \beta_4 (\text{Brand} \times \text{Abort}) + \beta_5 (\text{Value} \times \text{Abort}) \\ + \beta_6 (\text{Value} \times \text{Brand}) + \beta_7 (\text{Value} \times \text{Brand} \times \text{Abort})$$

$$(4) \text{ repost} = \beta_0 + \beta_1 \text{Brand} + \beta_2 \text{Abort} + \beta_3 \text{Value} + \beta_4 (\text{Brand} \times \text{Abort}) + \beta_5 (\text{Value} \times \text{Abort}) \\ + \beta_6 (\text{Value} \times \text{Brand}) + \beta_7 (\text{Value} \times \text{Brand} \times \text{Abort})$$

No obstante, en todos estos casos, ningún coeficiente resultó significativo. Al parecer, la similitud solo tuvo un efecto en la disposición a comprar pero no en la disposición a pagar por el producto, ni en la disposición a recomendar al influencer o republicar su contenido. Pueden verse todos las tablas de regresión y gráficos asociados a las regresiones (2), (3) y (4) en el Apéndice. Los valores de potencia post hoc calculados para estas regresiones dieron muy bajos para la regresión (2) y (3), pero considerablemente alto (en el orden de 0.656) para la regresión (4), lo cual nos indica que en un futuro debería hacerse este análisis con un tamaño de muestra mucho mayor que 213 para poder detectar si efectivamente la similitud entre el influencer y el consumidor tiene un impacto sobre la disposición a pagar, recomendar y republicar<sup>10</sup>.

### 5.10.3 Discusión

De esta forma, nuestro estudio experimental principal nos sirvió para consolidar una metodología de trabajo que nos permitiera una mayor comprensión de la relación entre los Influencers y los consumidores. En este caso, dadas las respuestas que obtuvimos por parte de los participantes, podemos concluir que la similitud sí parece tener un efecto significativo en la disposición a comprar para aquellos individuos en contra de la legalización del aborto que se vieron expuestos al Influencer Pro-Vida. Sin embargo, no se puede afirmar lo mismo para la disposición a pagar, recomendar y republicar.

Por supuesto, existen muchos aspectos de este experimento que podrían mejorarse. En primer lugar, de las 213 observaciones con las que nos quedamos luego de eliminar a aquellos individuos que eran de género masculino y sin Instagram, solo unos pocos pueden considerarse en contra de la legalización del aborto. La mayor parte de las respuestas de estos individuos a la pregunta sobre su ideología tienden a ser más cercanas al número 7, es decir, al número que hace referencia a la ideología a favor de la legalización del aborto. De hecho, solo 23 individuos de los 213 respondieron con un 1 o un 2. Esto nos da una señal de que, si se repitiera este

---

<sup>10</sup> Al correr estas regresiones incluyendo la variable MID, los valores de potencia aumentaron, al igual que en el caso de la regresión de disposición a comprar. En particular, para la regresión (2) el valor de potencia pasó de 0.415 a 0.520, para la regresión (3) pasó de 0.317 a 0.679 y para la regresión (4) pasó de 0.656 a 0.763. Los resultados de las regresiones con las variables de control incluidas tampoco dieron significativos para la disposición a pagar, recomendar y republicar.

experimento, posiblemente debería seguirse una metodología de muestreo estratificado, para tener una cantidad balanceada de individuos de cada grupo.

Por otro lado, la publicación de Otra Marca, como ya hemos discutido, puede haber generado un sesgo al momento de reportar la disposición a pagar por parte del individuo. Debido a que, en general, las marcas propias de los influencers no tienen un nivel de precios similar a las marcas establecidas en el mercado, no es del todo correcto hacer comparaciones entre la disposición a pagar de la publicación de Otra Marca con la Marca del Influencer. Es por esto que hemos hecho un re-escalamiento de esta variable para reducir el sesgo, aunque posiblemente en un futuro deba modificarse esta forma de hacer referencia a otra marca que no sea la del influencer.

Por último, el análisis de potencia post hoc realizado resultó de utilidad para verificar cuán significativa fue nuestra muestra para detectar efectos significativos. Estos valores, si bien no estuvieron por encima del valor 0.80 (un umbral generalizado en este tipo de análisis), fueron relativamente cercanos, entre un 0.60 y 0.70, centrándonos en las regresiones realizadas para medir el efecto de la similitud sobre la disposición a comprar. En particular, al incluir la variable MID, generó un aumento en la potencia post hoc. Con respecto a los resultados que no fueron significativos (disposición a pagar, recomendar y republicar), la valor de la potencia dio relativamente bajo, entre valores de 0.30 a 0.60. No obstante, al hacer las regresiones incluyendo MID e *Influence* respectivamente, observamos un aumento considerable de la potencia post hoc, dándonos indicios de que el tamaño de muestra de nuestra etapa experimental se acerca más a un valor idóneo para detectar efectos significativos cuando incluimos estas variables de control. En el Apéndice se incluye el cálculo de las potencias post hoc para todos los modelos y el tamaño de muestra necesario para que la potencia alcance un nivel de 0.80.

## 6. Conclusión

La presente investigación se ha centrado en lograr una mayor comprensión del llamado Marketing de Influencers, a partir de desplegar una serie de técnicas analíticas que pueden ser de utilidad a las empresas al momento de seleccionar qué influencers elegir para publicitar sus productos en las redes sociales.

Inicialmente, a partir de la gran cantidad de datos obtenidos por parte de la compañía Upfluence, hicimos un análisis exploratorio profundo. Esto implicó no sólo analizar las variables a través de varios gráficos que nos permitiesen visualizar la información, sino que también decidimos correr un algoritmo de clustering sobre los datos, de manera que podamos identificar patrones que caractericen a los influencers actuales. A partir de correr el algoritmo de *k-prototypes* sobre toda la base de datos, pudimos identificar cuatro grupos de influencers con sus propias características, siendo la cantidad de seguidores y la categoría del influencer las variables más interesantes al momento de interpretar los cuatro clusters obtenidos. Utilizando estos resultados, construimos cuatro *Influencer Personas* a modo de ejemplo, para poder visualizar cómo el algoritmo pudo agrupar los datos que describen a los influencers en las redes sociales. Esta metodología puede ser usada por las compañías para poder dar forma a los resultados que surgen de aplicar algoritmos complejos, interpretarlos y comunicarlos.

Luego de aplicar esta técnica algorítmica sobre toda la base de datos, decidimos aplicarlo a cada categoría por separado. Nos centramos en exhibir los resultados de la categoría *fashion* y *beauty*, categorías que se han mencionado frecuentemente en la bibliografía de SMIs. Lo que descubrimos es que los clusters obtenidos también se diferenciaban principalmente en la cantidad de seguidores. De esta forma, podemos concluir que los resultados del clustering están alineados con lo expuesto por la literatura sobre el Marketing de Influencers: la popularidad del influencer es una de las variables clave para poder distinguir entre distintos tipos de SMIs.

En segundo lugar, decidimos utilizar los resultados del clustering para poder investigar cómo influyen las características del influencer sobre el comportamiento del consumidor. Esta etapa busca ir más allá de los resultados obtenidos luego de ejecutar técnicas propias de disciplinas modernas como *Analytics* y *Data Science*, a partir de la realización de experimentos que nos permitan entender con mayor profundidad qué variables le importan al consumidor al momento de realizar una compra (además de la cantidad de seguidores). De esta manera, decidimos realizar un experimento con perfiles de influencers ficticios (de la categoría *fashion*), utilizando información del clustering.

El objetivo del experimento llevado a cabo fue entender cuál es el efecto de la similitud entre el influencer y el consumidor sobre diversas variables de interés para las compañías que quieren

promocionar sus productos online: disposición a pagar, a comprar el producto y disposición a recomendar y republicar contenido del influencer. Para definir similitud, utilizamos un perfil ficticio que estuvieran a favor de la legalización del aborto ("Pro-Aborto"), otro perfil ficticio que estuviera en contra de la legalización del aborto ("Pro-Vida") y preguntamos a los participantes del experimento cuál era su ideología con respecto a la legalización del aborto en la Argentina. De esta forma, en base a hacer uso de una de las temáticas más polémicas en la Argentina actual, pudimos concluir que aquellos participantes que se consideran en contra del aborto, tuvieron una mayor disposición a comprar el producto de la marca propia del influencer cuando este era **similar**, es decir, cuando el influencer era Pro-Vida. No obstante, los resultados no fueron significativos para el caso de la disposición a pagar, recomendar y republicar, lo cual nos indica que este experimento puede mejorarse para obtener mejores resultados. Hemos mencionado dos posibles mejoras del experimento llevado a cabo. En primer lugar, se debería aumentar el tamaño de muestra, para asegurar un valor de potencia estadística mayor, principalmente para aquellos casos en los que se quiere detectar un efecto de la similitud (entre el influencer y el seguidor) sobre la disposición a pagar, recomendar y republicar. En segundo lugar, la toma de esta muestra debería hacerse de manera aleatoria y estratificada para asegurar que los grupos que nos proponemos estudiar (en este caso, individuos Pro-Vida y Pro-Aborto) tengan un tamaño similar.

Es importante mencionar que nuestra forma de medir la similitud entre influencer y seguidor es innovadora con respecto a la literatura analizada. Recordemos que los trabajos de investigación que se han propuesto analizar el nivel de similitud, siempre lo han hecho utilizando influencers conocidos mayoritariamente por la sociedad y cuyas ideas u opiniones son conocidas por su público. De esta manera, medir la similitud sólo requería preguntar a los participantes si se sentían identificados con el SMI seleccionado, o si pensaban similar. Nuestro enfoque complejiza esta metodología, no sólo haciendo uso de influencers creados a partir de un análisis profundo de segmentación, sino también proponiendo centrarse en una similitud percibida e ideológica. Todos los autores que han sido mencionados y han estudiado la similitud entre influencer y seguidor, han encontrado que esta tiene un efecto positivo sobre la intención de compra, confianza hacia el influencer y *awareness* de la marca promocionada. Nuestros resultados siguen esta misma tendencia, afirmando que la similitud efectivamente tiene un impacto en la disposición a comprar para los individuos Pro-Vida del experimento.

Con respecto a las implicancias de nuestra investigación, el presente estudio tiene varias aplicaciones interesantes cuando se trata de la toma de decisiones de negocios.

En primer lugar, nuestro enfoque puede eventualmente ayudar a las compañías y marcas a seleccionar una mejor estrategia de Marketing de Influencers: en lugar de seleccionar al



influencer más popular, pueden aplicarse primero técnicas analíticas sobre las bases de datos disponibles y realizar diversos experimentos para poder comprender más profundamente el impacto de la selección de cierto influencer sobre el comportamiento del consumidor. Por lo cual, las empresas podrían utilizar nuestra metodología para ejecutar su propio análisis de clustering y obtener una solución personalizada. Utilizando sus propios perfiles de influencers, podrían realizar experimentos o pruebas A / B con su base de clientes y sus productos.

Por otra parte, nuestros resultados pueden ser útiles para compañías intermediarias como Upfluence (nuestro proveedor de datos) al momento de tener que dar un consejo a sus clientes (otras empresas o marcas) sobre qué influencer seleccionar para promocionar productos. Las firmas como Upfluence que operan como intermediarias entre empresas e influencers, podrán explicar con mayor claridad qué tipo de influencers están presentes actualmente en las redes sociales y podrán sugerir experimentos posibles de acuerdo al perfil de la compañía. Estos consejos basados en datos ciertamente ayudarán a las empresas a tomar decisiones de marketing online.

Nuestro hallazgo relacionado al estudio experimental, también transmite un mensaje a las compañías al momento de elegir a los influencers que publicitarán sus productos. Si bien nuestros resultados fueron parciales y aplicaron a una parte de la muestra (aquellos individuos en contra de la legalización del aborto), si nos abstraemos, podemos concluir que las compañías deben seleccionar influencers cuidadosamente, teniendo en cuenta la audiencia a la que quieren llegar. Dado que la similitud ideológica, de valores y pensamientos parece ser un factor importante al momento de comprar un producto por parte de un influencer en las redes sociales, las empresas deben intentar seleccionar aquellos SMIs que sean similares a la audiencia a la que quieren atraer. De esta manera, puede concluirse que una posible estrategia de marketing online que surge a partir de nuestro hallazgo, es elegir influencers que sean parecidos a los consumidores representativos que la empresa o marca en cuestión desea convencer de comprar su producto. Por supuesto, nuestro estudio fue hecho tomando a las mujeres argentinas como universo de interés y hemos elegido la categoría *fashion* para simplificar la etapa experimental. Sin embargo, esta temática debe seguir siendo investigada y analizada con mayor profundidad, incluyendo más categorías, otras formas de plantear la similitud, y con otros universos (hombres, distintos rangos etarios, estudiantes, etc.).

Finalmente, teniendo en cuenta la literatura existente sobre el Marketing de Influencers y SMIs, nuestro trabajo puede considerarse como el primer intento de combinar un análisis de clustering (técnica de Machine Learning) con experimentos (método de investigación en Marketing). Este método nos permite obtener información e *insights* que no podrían haberse obtenido únicamente con la etapa descriptiva o relacionada a analizar grandes volúmenes de datos (*big*

*data*). Creemos que esta combinación de metodologías podría agregarse a la “caja de herramientas” del científico de datos y permitir que los gerentes o agentes de toma de decisiones puedan obtener información fructífera y práctica. Sin embargo, nuestro desarrollo deja la puerta abierta a que se utilicen otras técnicas dentro de la disciplina de *Data Science* para combinar con el uso de clustering, enriqueciendo la metodología “híbrida” propuesta. Por ejemplo, podría llevarse a cabo un análisis de *Natural Language Processing* (NLP) sobre los textos de las publicaciones de los influencers para poder detectar patrones de comportamiento, personalidad y tono en que se comunica con sus seguidores.

## 7. Bibliografía

- Abidin, C. (2019). Yes Homo: Gay influencers, homonormativity, and queerbaiting on YouTube. *Continuum*, 33(5), 614-629.
- Alp, Z. Z., & Öğüdücü, Ş. G. (2019). Influence factorization for identifying authorities in twitter. *Knowledge-Based Systems*, 163, 944-954.
- Aran Ramspott, S., Fedele, M., & Tarragó, A. (2018). Funciones sociales de los youtubers y su influencia en la preadolescencia. *Comunicar*, 2018, 1 octubre, Vol. XXVI, nº 57, 4º trimestre.
- Aquino, K., & Reed II, A. (2002). The self-importance of moral identity. *Journal of personality and social psychology*, 83(6), 1423.
- Blight, M. G., Ruppel, E. K., & Schoenbauer, K. V. (2017). Sense of community on Twitter and Instagram: Exploring the roles of motives and parasocial relationships. *Cyberpsychology, Behavior, and Social Networking*, 20(5), 314-319.
- Bonferroni, C. (1936). Teoria statistica delle classi e calcolo delle probabilita. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze*, 8, 3-62.
- Cialdini, R. (1993). *The psychology of influence*. New York: William Morrow & Co.
- Colliander, J., & Dahlén, M. (2011). Following the fashionable friend: The power of social media: Weighing publicity effectiveness of blogs versus online magazines. *Journal of advertising research*, 51(1), 313-320.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *psychometrika*, 16(3), 297-334.
- Del Fresno García, M., Daly, A. J., & Segado Sanchez-Cabezudo, S. (2016). Identifying the new Influences in the Internet Era: Social Media and Social Network Analysis. *Revista Española de Investigaciones Sociológicas*, (153).
- De Veirman, M., Cauberghe, V., & Hudders, L. (2017). Marketing through Instagram influencers: the impact of number of followers and product divergence on brand attitude. *International Journal of Advertising*, 36(5), 798-828.
- Dinesh, D. (2017), Why Micro-Influencers are a Social Media Marketing Imperative for 2017, *EContent Magazine*
- Djafarova, E., & Rushworth, C. (2017). Exploring the credibility of online celebrities' Instagram profiles in influencing the purchase decisions of young female users. *Computers in Human Behavior*, 68, 1-7.
- Evans, N. J., Phua, J., Lim, J., & Jun, H. (2017). Disclosing Instagram influencer advertising: The effects of disclosure language on advertising recognition, attitudes, and behavioral intent. *Journal of Interactive Advertising*, 17(2), 138-149.

- Freberg, K., et al. (2010), Who are the social media influencers? A study of public perceptions of personality, *Public Relations Review*, doi: 10.1016/j.pubrev.2010.11.001
- Gaines-Ross, L. (2003). CEO capital. *Executive Excellence*, 20(2), 20-20.
- Galeotti, A., & Goyal, S. (2009). Influencing the influencers: a theory of strategic diffusion. *The RAND Journal of Economics*, 40(3), 509-532.
- Gorry, G. A., & Westbrook R. A. (2009), Academic research: Winning the Internet confidence game, *Corporate Reputation Review*, 12(3), 195–203. <https://doi.org/10.1057/crr.2009.16>
- Himmelboim, I., & Golan, G. J. (2019). A social networks approach to viral advertising: The role of primary, contextual, and low influencers. *Social Media+ Society*, 5(3), 2056305119847516.
- Hoening, J. M., & Heisey, D. M. (2001). The abuse of power: the pervasive fallacy of power calculations for data analysis. *The American Statistician*, 55(1), 19-24.
- Huang, Z. (1998). Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data mining and knowledge discovery*, 2(3), 283-304.
- Ji, J., Bai, T., Zhou, C., Ma, C., & Wang, Z. (2013). An improved k-prototypes clustering algorithm for mixed numeric and categorical data. *Neurocomputing*, 120, 590-596.
- Kelman, H. C. (2006). Interests, relationships, identities: Three central issues for individuals and groups in negotiating their social environment. *Annu. Rev. Psychol.*, 57, 1-26.
- Ki, C. W. C., & Kim, Y. K. (2019). The mechanism by which social media influencers persuade consumers: The role of consumers' desire to mimic. *Psychology & Marketing*, 36(10), 905-922.
- Lahuerta-Otero, E., & Cordero-Gutiérrez, R. (2016). Looking for the perfect tweet. The use of data mining techniques to find influencers on twitter. *Computers in Human Behavior*, 64, 575-583.
- Lee, J. E., & Watkins, B. (2016). YouTube vloggers' influence on consumer luxury brand perceptions and intentions. *Journal of Business Research*, 69(12), 5753-5760.
- Lou, C., & Yuan, S. (2019). Influencer marketing: how message value and credibility affect consumer trust of branded content on social media. *Journal of Interactive Advertising*, 19(1), 58-73.
- More, J. S., & Lingam, C. (2019). A SI model for social media influencer maximization. *Applied Computing and Informatics*, 15(2), 102-108.
- Narassiguin, A., Sargent, S. (2019), *Data Science for Influencer Marketing: feature processing and quantitative analysis*, ffhal-02120859f

Schouten, A. P., Janssen, L. & Verspaget, M. (2020), Celebrity vs. Influencer endorsements in advertising: the role of identification, credibility, and Product-Endorser fit, *International Journal of Advertising*, 39:2, 258-281, DOI: 10.1080/02650487.2019.1634898

Uzunoglu, E., & Kip, S. M. (2014). Brand communication through digital influencers: Leveraging blogger engagement. *International Journal of Information Management*, 34(5), 592-602.

Xu, X., & Pratt, S. (2018). Social media influencers as endorsers to promote travel destinations: an application of self-congruence theory to the Chinese Generation Y. *Journal of travel & tourism marketing*, 35(7), 958-972.

Zhang, Y., Moe, W. W., & Schweidel, D. A. (2017). Modeling the role of message content and influencers in social media rebroadcasting. *International Journal of Research in Marketing*, 34(1), 100-119.

Zietek, N. (2016). Influencer Marketing: the characteristics and components of fashion influencer marketing.

## 8. Apéndice

### 8.1 Variables de los Datasets

#### 8.1.1 Dataset 'instagram' (instagram.tsv)

**influencer\_id:** Variable del tipo integer que representa el número id asignado a cada influencer único.

**email:** Variable del tipo string que representa el email del influencer.

**name:** Variable del tipo string que representa el nombre completo del influencer.

**lang:** Variable del tipo factor que representa el idioma usado (en posts) en el perfil de instagram. Posee 49 niveles.

**country:** Variable del tipo string que representa el país de residencia del influencer.

**address:** Variable del tipo string que representa la dirección donde el influencer reside.

**gender:** Variable del tipo factor que representa el género del influencer. Posee 2 niveles: Gender\_FEMALE y Gender\_MALE.

**age\_bracket:** Variable del tipo factor que representa el rango de edad al cual pertenece el influencer. Posee 4 niveles: 0-17, 18-24, 25-34 y 35-54.

**instagram\_id:** Variable del tipo integer que representa el número de id asignado a cada instagram.

**username:** Variable del tipo string que representa el nombre de usuario del instagram.

**full\_name:** Variable del tipo string que representa el nombre completo del influencer.

**bio:** Variable del tipo string que representa la biografía establecida en el instagram.

**website:** Variable del tipo string que representa el sitio web del influencer.

**followers:** Variable del tipo entera que representa la cantidad de seguidores del perfil de instagram.

**engagement\_rate:** Variable del tipo numérica que representa la tasa de participación de los seguidores del perfil de instagram en las publicaciones. Se calcula  $(\text{likes} + \text{comentarios})/\text{followers}$ .

**total\_posts:** Variable del tipo entera que representa la cantidad total de publicaciones realizadas por el perfil de instagram,

**total\_engagements:** Variable del tipo entera que representa la suma total de interacciones realizadas en todas las publicaciones hechas por el perfil de instagram. Se calcula likes totales + comentarios totales.

**total\_likes:** Variable del tipo entera que representa la suma total de "likes" realizados en todas las publicaciones hechas por el perfil de instagram.

**total\_comments:** Variable del tipo entera que representa la suma total de comentarios realizados en todas las publicaciones hechas por el perfil de instagram.

**category:** Variable del tipo factor que representa la temática del instagram. Posee 10 niveles: beauty, family, fashion, food, gaming, home, lifestyle, sport, technology y travel.

### 8.1.2 Dataset 'instagram\_posts' (instagram\_posts.tsv)

**id:** Variable del tipo integer que distingue cada registro de manera única.

**instagram\_id:** Variable del tipo integer que representa el número de id asignado al instagram que realizó la publicación.

**text:** Variable del tipo string que representa la descripción que fue añadida a la publicación.

**type:** Variable del tipo factor que representa el tipo de la publicación. Posee 2 niveles: picture y video.

**location\_name:** Variable del tipo string que representa la localidad desde dónde fue hecha la publicación.

**timestamp:** Variable del tipo entera que representa la fecha y horario en el cual fue hecha la publicación, en formato horas desde las 00:00 del 01-ene-1970.

**likes:** Variable del tipo entera que representa la cantidad de "likes" que tuvo la publicación.

**comments:** Variable del tipo entera que representa la cantidad de comentarios que tuvo la publicación.

**post\_id:** Variable del tipo string que representa una sucesión alfanumérica asignada a cada publicación única, con en cual se puede encontrar la publicación.

**views:** Variable del tipo entera que representa la cantidad de visitas que tuvo la publicación.

**media\_type:** Variable del tipo factor que representa el tipo de herramienta multimedia usada en la publicación. Posee 3 niveles: carousel, image y video.

**thumbnail\_url:** Variable del tipo string que representa la url con la cual se puede encontrar a la publicación.

## 8.2 Transformaciones de las Variables

### 8.2.1 Dataset 'instagram'

**username:** Variable del tipo entera que representa la cantidad de caracteres que componen el nombre de usuario en instagram. Se obtuvo aplicando la función `nchar()` sobre la variable `username`.

**redes\_soc:** Variable del tipo entera que representa la cantidad de redes sociales diferentes a instagram en las cuales se puede encontrar al influencer. Se obtuvo buscando en la variable `bio` si estaban presentes los nombres de las distintas redes sociales y contandolas (YouTube, Gmail, Hotmail, Twitter, website y Twitch).

**picture\_share:** Variable del tipo numérica que representa la proporción de posts del tipo imagen en relación a los posts totales. Se obtuvo buscando el *instagram\_id* en el dataset *instagramPosts* para obtener las publicaciones de cada instagram, contando cuántas de esas publicaciones son del tipo imagen y dividiendo esa cantidad sobre el total de publicaciones del instagram encontradas en el dataset.

**video\_share:** Variable del tipo numérica que representa la proporción de posts del tipo video en relación a los posts totales. Se obtuvo buscando el *instagram\_id* en el dataset *instagramPosts* para obtener las publicaciones de cada instagram, contando cuántas de esas publicaciones son del tipo video y dividiendo esa cantidad sobre el total de publicaciones del instagram encontradas en el dataset.

**hashtagspromedio:** Variable del tipo numérica que representa la cantidad promedio de hashtags que el instagram usa en sus publicaciones. Para obtener esta variable, se creó en el dataset *instagramPosts* una variable llamada *hashtags* que extraiga los hashtags usados en cada publicación. Luego, se buscó cada *instagram\_id* del dataset *instagram* en el dataset *instagramPosts* para obtener las publicaciones de cada instagram, se contó la cantidad de hashtags en cada una de esas publicaciones y se sumó las cantidades, y se dividió esa suma entre la cantidad de publicaciones del instagram encontradas.

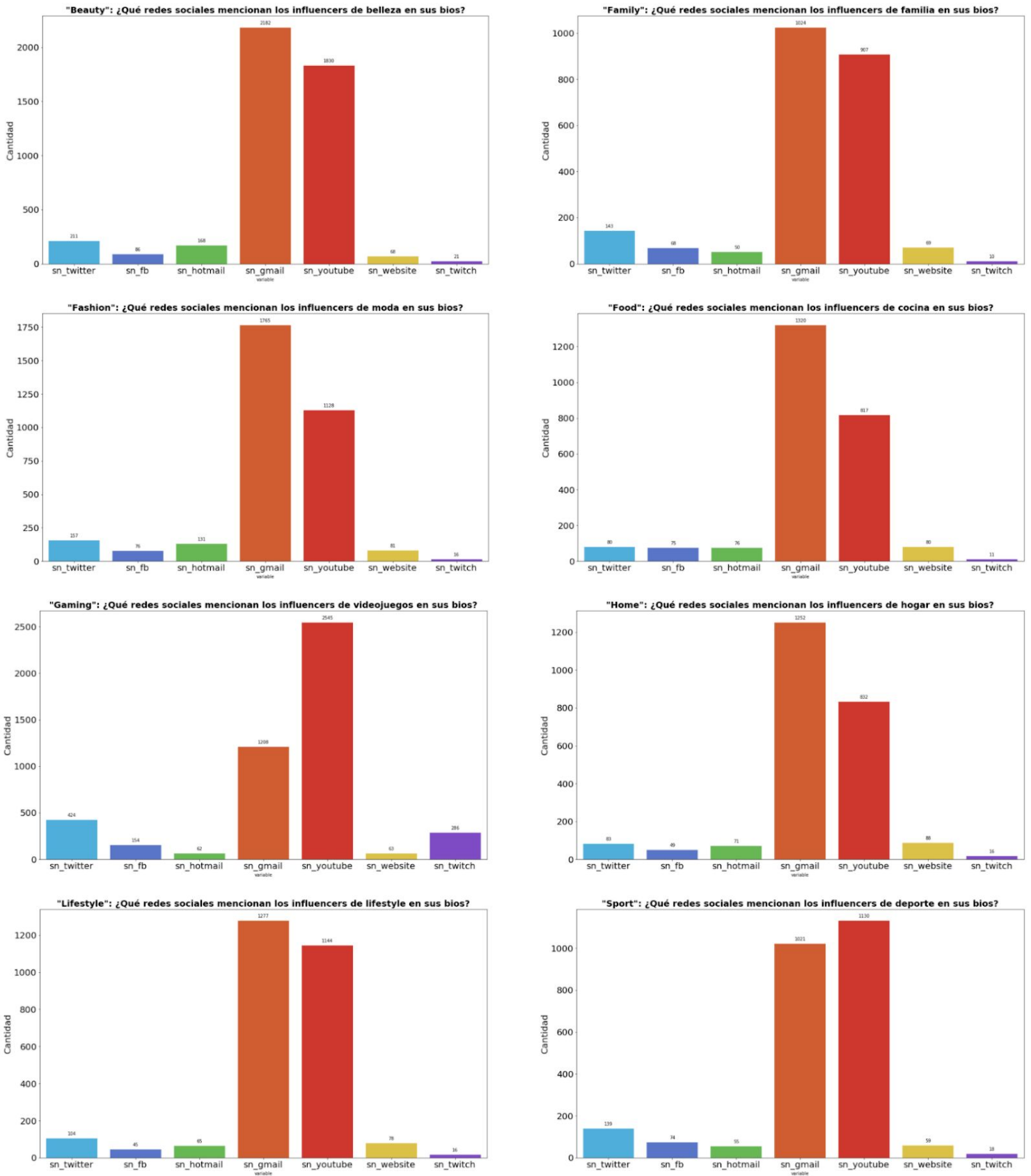
**arrobaspromedio:** Variable del tipo numérica que representa la cantidad promedio de arrobas (etiquetaciones) que el instagram agrega en sus publicaciones. Para obtener esta variable, se creó en el dataset *instagramPosts* una variable llamada *arrobaspromedio* que extraiga las arrobas usados en cada publicación. Luego, se buscó cada *instagram\_id* del dataset *instagram* en el dataset *instagramPosts* para obtener las publicaciones de cada instagram, se contó la cantidad de arrobas en cada una de esas publicaciones y se sumó las cantidades, y se dividió esa suma entre la cantidad de publicaciones del instagram encontradas.

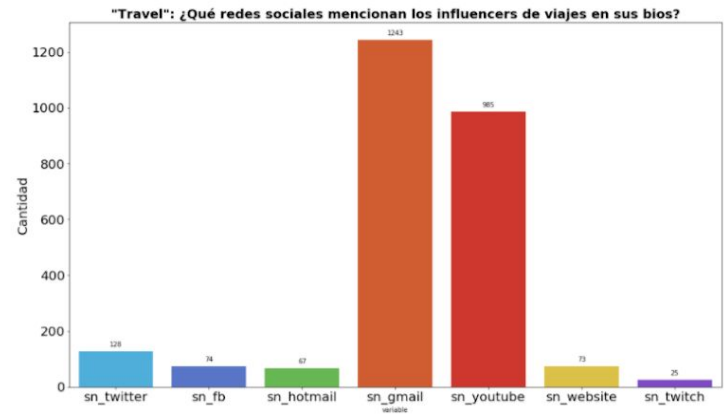
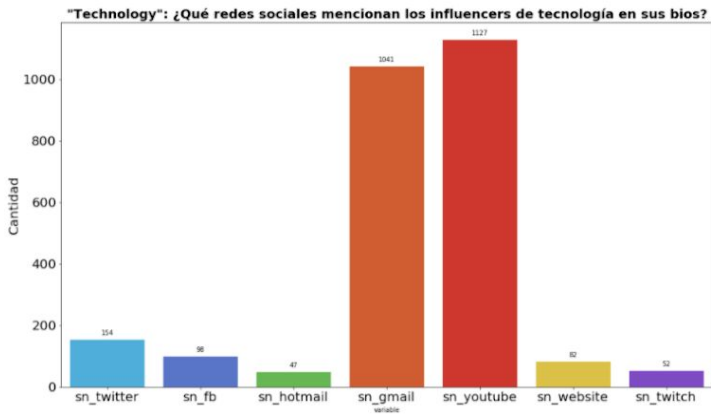
**frecuencia\_post:** Variable del tipo numérica que representa la cantidad de horas promedio entre cada publicación que hace el instagram. Para obtener esta variable, se buscó cada *instagram\_id* del dataset *instagram* en el dataset *instagramPosts* para obtener las publicaciones de cada instagram, se las ordenó cronológicamente las publicaciones según la variable *timestamp*, se restó cada publicación de la publicación inmediatamente anterior para obtener las horas entre cada par de publicaciones seguidas, y se sacó un promedio de todos los tiempos entre publicaciones obtenidos. Esta variable toma el valor NA en el caso de que del instagram no se hayan encontrado publicaciones en *instagramPosts* o que se haya encontrado una sola publicación en *instagramPosts* (ya que con una sola publicación no se puede calcular frecuencia entre dos posts seguidos).



### 8.3 Gráficos Adicionales (Análisis Exploratorio)

Figura 39: Histograma de redes sociales mencionadas en las bios de influencers, por categoría





Haciendo un rápido análisis de estos gráficos, observamos que la mayoría de los influencers, sin importar la categoría, mencionan sus cuentas de YouTube o su Gmail. Luego, algunos mencionan su website o blog, y para el caso particular de los influencers pertenecientes a la categoría "gaming", se menciona la cuenta de Twitch, plataforma en la cual se puede transmitir en vivo mientras se juega a diversos videojuegos.

Figura 40: Distribución del Tiempo entre Publicaciones (en Horas)



Figura 41: Distribución de la Participación de Videos en las Publicaciones



Figura 42: Distribución de la Participación de Imágenes en las Publicaciones



Figura 43: Distribución de la Cantidad de *Hashtags* Promedio utilizados por los influencers en sus publicaciones

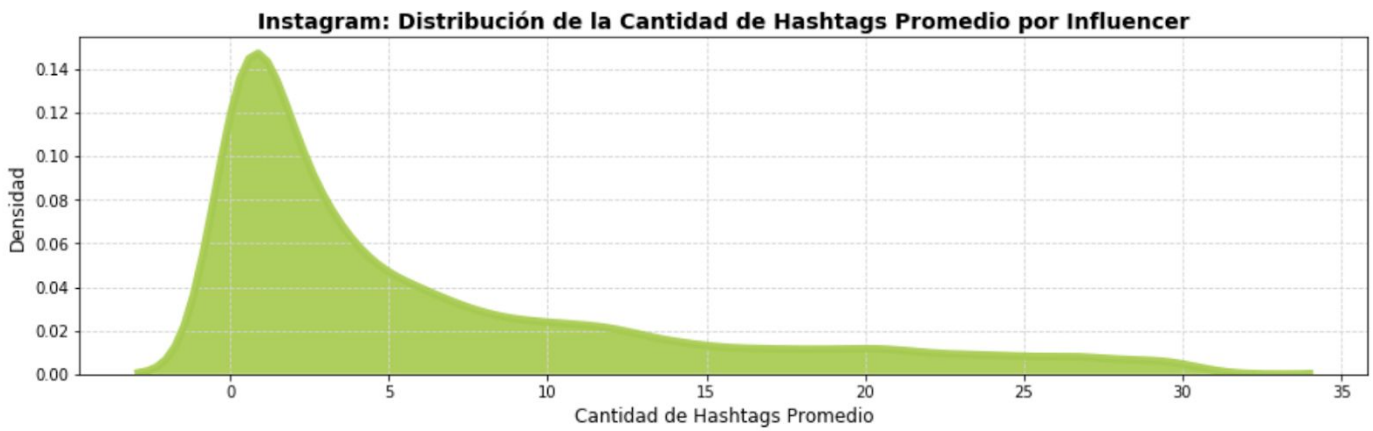
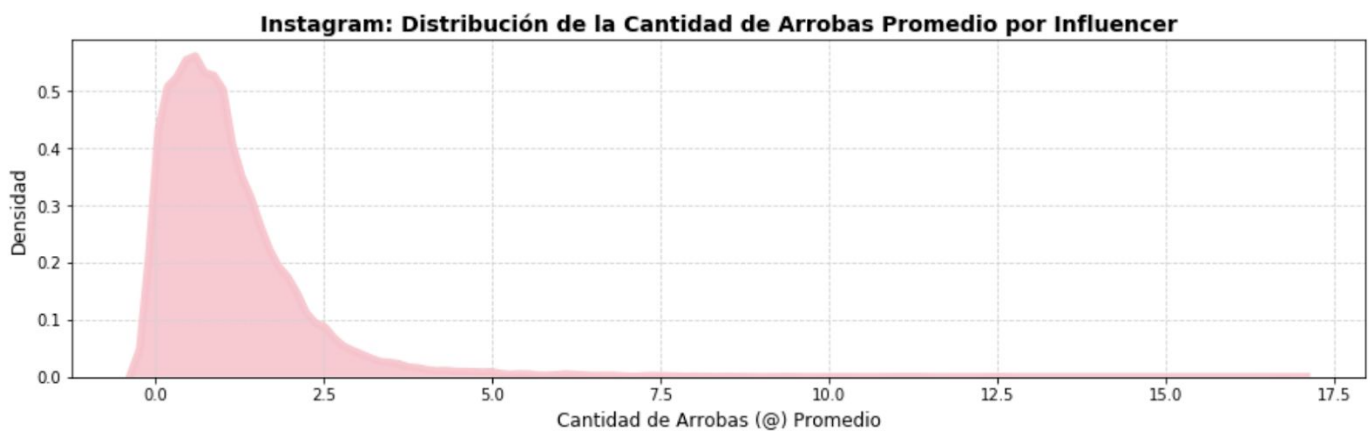


Figura 44: Distribución de la Cantidad de *Arrobas* (@) Promedio utilizados por los influencers en sus publicaciones



## 8.4 Listado Final de Variables utilizadas para hacer Clustering

A continuación enumeramos todas las variables presentes en el dataset final (que contiene 84.637 observaciones) para luego correr el algoritmo de clustering seleccionado (k-prototypes).





1. lang
2. gender
3. age\_bracket
4. username
5. followers
6. engagement\_rate
7. total\_posts
8. total\_engagements
9. total\_likes
10. total\_comments
11. category
12. redes\_soc
13. picture\_share
14. video\_share
15. hashtagspromedio
16. arrobaspromedio
17. frecuencia\_post

Recordemos que a todas aquellas variables que presentaron alguna asimetría las transformamos tomando el logaritmo natural. Finalmente, para que todas las variables queden expresadas en una misma unidad, las re-escalamos restandoles el promedio y dividiendo por el desvío estándar, a partir de la función `scaling()` en R.





## 8.5 *Influencer Personas*: Influencers Representativos a partir de la primera versión del algoritmo de Clustering

A partir de haber corrido por primera vez el algoritmo sobre los datos transformados, creamos un conjunto de Influencer Personas o influencers representativos de cada cluster. Esto nos sirvió para poder tener una primera idea de qué patrones estaban presentes en los datos, de manera tal que, al construir estas Personas, los datos fueran más visibles y entendibles. Por lo cual, utilizando las características de los clusters y el análisis exploratorio realizado, armamos las siguientes influencers representativos:

## Influencer Persona 1: Influencer de Moda *miranda.style*

<p>NOMBRE DE USUARIO EN INSTAGRAM</p> <p><b>miranda.style</b></p>	<p>MARKET SIZE</p> <p> <b>35 %</b></p>	<p>TYPE</p> <p><b>Fashion Influencer</b></p>	
	<h3>Publicaciones</h3> <ul style="list-style-type: none"><li>◦ Sube fotos de diferentes looks de ropa cada <b>5 o 6 días</b>.</li><li>◦ Usa muchos <i>hashtags</i> en sus posts para atraer a más público que busque ciertos estilos de ropa (<b>#fashion, #ootd</b> -outfit of the day- <b>#style, #love, #liketkit, #fashionblogger #model</b>, etc). Pueden incluir arrobas de sponsors o de medios que venden prendas de sus outfits (ej: <b>@LIKEtoKNOW.it</b>)</li><li>◦ Casi no sube videos, el <b>94% de sus publicaciones son fotos</b>.</li><li>◦ Se dedica a mostrar diferentes estilos y vestimentas, pero no utiliza más herramientas de Instagram (historias, <i>challenges</i>, sorteos) para generar conexiones más profundas con sus seguidores.</li><li>◦ Publica poco para no saturar a su público.</li></ul>		
<h3>Datos</h3> <p>Género: Femenino</p> <hr/> <p>Edad: 30 Años</p> <hr/> <p># Seguidores: ~25K</p> <hr/> <p>Engagement Rate: 0.095%</p> <hr/> <p># Publicaciones (3 meses): 70</p>	<h3>Seguidores</h3> <ul style="list-style-type: none"><li>◦ <b>Poca interacción directa con sus seguidores</b> → más contacto mediante <i>likes</i>, no tanto por comentarios.</li><li>◦ Los usuarios que la siguen lo hacen para <b>inspirarse en cuanto a looks de ropa</b>, pero no tienen conexión sentimental con ella.</li></ul>	<h3>Gain Points</h3> <p><b>Ayuda a los usuarios de Instagram a inspirarse y armar sus outfits</b>, ya sea porque son sus seguidores fieles o porque encuentran sus posts mediante los <i>hashtags</i>.</p>	<h3>Pain Points</h3> <p>Su escasa interacción (expresada en un bajo <i>Engagement Rate</i>) con su público puede generar un <b>incremento lento de su masa de seguidores</b>.</p>
<h3>Presencia en otras redes sociales</h3> <p> </p>			

## Influencer Persona 2: Influencer de Belleza *mykie.beauty*


<p>NOMBRE DE USUARIO EN INSTAGRAM</p> <p><b>mykie.beauty</b></p>	<p>MARKET SIZE</p> <p> <b>40 %</b></p>	<p>TYPE</p> <p><b>Beauty Influencer</b></p>
	<h3>Publicaciones</h3> <ul style="list-style-type: none"><li>• Sube fotos y videos de diferentes looks de maquillaje, técnicas de cuidados de piel, técnicas para maquillarse, etc.</li><li>• No usa muchos <i>hashtags</i> en sus posts porque ya tiene un público establecido (utiliza algunos como <b>#makeup, #beauty, #fashion</b>) Los arrobas de sus posts son informativos de marcas de cosméticos que utiliza (<b>@anastasiabeverlyhills, @colourpopcosmetics</b>, etc).</li><li>• Sube en su mayoría imágenes (90%), pero también <b>sube videos tutoriales (10%)</b>.</li><li>• <b>Publica cada 2 o 3 días</b>, es muy activa en su cuenta de Instagram.</li></ul>	
<h3>Datos</h3> <p>Género: Femenino</p> <hr/> <p>Edad: 27 Años</p> <hr/> <p># Seguidores: ~490K</p> <hr/> <p>Engagement Rate: 0.6%</p> <hr/> <p># Publicaciones (3 meses): 600</p> <hr/>	<h3>Seguidores</h3> <ul style="list-style-type: none"><li>• <b>Gran interacción directa con su público</b>, hace historias, <i>challenges</i> y contesta preguntas para tener contacto con sus seguidores.</li><li>• Tiene un alto <i>Engagement Rate</i>, lo cual implica que cada publicación recibe muchos <i>likes</i> y comentarios en promedio, en relación al total de sus seguidores.</li><li>• Los usuarios en Instagram la siguen para <b>inspirarse e informarse sobre tips de maquillaje y belleza</b>, y también por la <b>conexión emocional</b> que desarrollaron con ella.</li><li>• Le gusta crear confianza y cercanía con su público y <b>compartir momentos de su vida privada</b> (viajes, pareja y amigos, tutoriales con algún amigo o famoso/otro influencer).</li></ul>	
<h3>Gain Points</h3> <p>La gente en las redes la aprecia por la inspiración que genera y porque también <b>entretiene con su personalidad encantadora</b>.</p>	<h3>Pain Points</h3> <p>Dilema entre seguir subiendo <b>contenido sobre belleza y estética</b> (pero quedarse estancada en cantidad de seguidores) o centrarse más en subir <b>contenido entretenido y llegar a más usuarios</b>.</p>	
<h3>Presencia en otras redes sociales</h3> <p> </p>		

## Influencer Persona 3: Influencer de Videojuegos *harleyplays1*

NOMBRE DE USUARIO EN INSTAGRAM


**harleyplays1**

MARKET SIZE

 10 %

TYPE

**Gaming Influencer**



### Publicaciones

- Sube fotos casuales de ella y videos de sus momentos remarcables jugando **juegos virtuales**.
- Utiliza Instagram para generar una **conexión más cercana con su público** y que estos se enteren de eventos en su vida personal.
- Usa muchos *hashtags* en sus posts para que más personas la encuentren y comiencen a seguirla (**#ad, #art, #youtube, #cosplay, #youtuber**). A su vez, los *gamers* que se dedican a promocionar su actividad en torneos de videojuegos o *gameplays*, suelen tener sponsors (tecnología, equipos en torneos), por lo que utilizan *hashtags* que citen a estos (ej: suelen incluir **#ad**)
- **La mitad del contenido que sube en Instagram son videos** (que pueden estar relacionados con videos más largos en YouTube).
- **Publica cada 4 días**, ya que Instagram no es su plataforma principal de contacto con sus seguidores (utilización de YouTube y Twitch)

### Datos

Género: Femenino

Edad: 25 Años

# Seguidores: ~85K

Engagement Rate: 0.16%

# Publicaciones (3 meses): 200

### Seguidores

- Gran interacción con sus suscriptores/seguidores, pero no necesariamente en Instagram, sino más bien en plataformas de *streaming*.
- Bajo *Engagement Rate* porque su público interactúa con ella en otras redes sociales.
- Los usuarios de Instagram la siguen para **conocer más de su vida cotidiana y sentirse más cercanos a ella**, más allá de su performance como *gamer*.
- Le gusta crear confianza y cercanía con los que la siguen y hacerlos participar de juegos o *challenges*. Además, **está atenta a sus sugerencias para crear contenido nuevo (sugerencias de otros juegos por ejemplo)**.



### Gain Points

La gente en las redes la aprecia por su personalidad encantadora, sus comentarios sobre juegos de moda y sus *gameplays* entretenidos y graciosos.





### Pain Points

**Desearía aumentar su masa de seguidores en Instagram** (que es menor a la de sus otras redes) para así incorporar otras formas innovadoras de crear contenido orientado a los videojuegos y a su vida diaria.

### Presencia en otras redes sociales

## Influencer Persona 4: Influencer de Alimentos/Cocina *veganmel.eats*

NOMBRE DE USUARIO EN INSTAGRAM <b>veganmel.eats</b>		MARKET SIZE  15 %	TYPE <b>Food Influencer</b>
	<b>Publicaciones</b> <ul style="list-style-type: none"><li>• Sube fotos de recetas y de su dieta diaria.</li><li>• Usa muchos <i>hashtags</i> en sus posts para que las personas interesadas puedan encontrar su blog y sus recetas fácilmente (<b>#ad, #food, #feedfeed, #foodie, #f52grams, #foodporn, #foodphotography, #foodblogger</b>)</li><li>• El 96% de su contenido son imágenes (fotos de los platos que prepara/come), mientras que el resto son videos (recetas rápidas)</li><li>• Publica cada <b>3 o 4 días</b>, para no saturar a su público y para darle tiempo a crear contenido relevante.</li><li>• No hace demasiados posts sponsorados.</li></ul>		
	<b>Seguidores</b> <ul style="list-style-type: none"><li>• Buena interacción con sus seguidores, sobretodo en formato "preguntas y respuestas" para difundir información e ideas sobre estilos de dieta.</li><li>• Alto <i>Engagement Rate</i> ya que su público, aunque pequeño, es leal a ella.</li><li>• Los usuarios en Instagram que la siguen lo hacen por sus <b>consejos sobre alimentación variada</b> (vegana, vegetariana, etc) y para <b>adquirir ideas sobre nuevos platos o recetas</b>.</li><li>• Se esfuerza en generar confianza en sus seguidores y ser transparente con sus comidas y sus recetas porque sabe que la alimentación es algo muy delicado.</li></ul>		
<b>Datos</b> <ul style="list-style-type: none"><li>Género: Femenino</li><li>Edad: 34 Años</li><li># Seguidores: ~30K</li><li>Engagement Rate: 1.2%</li><li># Publicaciones (3 meses): 400</li></ul>	<b>Gain Points</b> <p>La gente en las redes la aprecia por la información e ideas que divulga, las cuales incluyen <b>tips para seguir una dieta balanceada y deliciosa</b> que puede <b>ajustarse a todo tipo de usuarios</b>.</p>	<b>Pain Points</b> <p>Tiene que esforzarse mucho en crear contenido innovador e interesante ya que <b>tiene riesgo de volverse repetitiva o no diferenciarse con respecto a otros influencers similares</b>.</p>	
<b>Presencia en otras redes sociales</b>  			

### 8.6 Pretest: *Influencer Personas* y sus respectivas publicaciones en Instagram

A partir del resultado del clustering utilizando la categoría "fashion", construimos *Influencer Personas* pero con un formato de perfiles de la plataforma Instagram, junto con publicaciones asociadas. Estos perfiles y publicaciones serán utilizados en el pretest.



## Perfil a favor de la legalización del aborto "Pro1": *merystyle\_ok*

### Perfil del Influencer



Instagram profile of *merystyle\_ok*. The profile shows a bio with the name "Mery Style", CEO of @meryscollection, and various hashtags including #AbortoLegalYa, #NiUnaMenos, and #QueSeaLey. The profile also displays statistics: 7692 posts, 348K followers, and 1244 following. There are icons for grid, gallery, and profile.

### Publicación de marca propia



Instagram post by *merystyle\_ok* featuring a grey hoodie with cutouts on the sleeves. The caption reads: "¿Tienen frío pero no tanto como para ponerse una campera? ¿No saben qué usar en #otoño? Obvio que @meryscollection viene a ayudarlas a todas ustedes a encontrar su propio estilo en estos climas raros. Diseñé este buzo muy cool. Link in bio para comprar!"

### Publicación de otra marca



Instagram post by *merystyle\_ok* featuring a grey hoodie with cutouts on the sleeves. The caption reads: "@cuesta\_blanca\_oficial me mandó este buzo tremeeendo para disfrutar del otoño y estoy ANSIOSA por usarlo!! El gris es un color básico que combina con todo, así que tienen que tener este buzo para agregarle rebeldía a sus días. Visiten la tienda online! #fashion #cuestablanca"

### Publicación de marca indefinida



Instagram post by *merystyle\_ok* featuring a grey hoodie with cutouts on the sleeves. The caption reads: "Este buzo con cortes locos me parece increíble para agregarle un toque de estilo a tu #outfit. Esos cortes le dan personalidad, soltura y rebeldía a tu imagen. Querés ser indomable? Clavate este buzo. Mirá mis historias para saber dónde conseguirlo."

## Perfil a favor de la legalización del aborto "Pro2": *agus.barrero*

### Perfil del Influencer



agus.barrero

7692 posts 348K followers 1244 following

Message

Agus Barrero

CEO @bqueen

26 años y experta en outfits #fashion #fashioninsta

- 📖 Fashion & books feminista y acuariana
- 💚 #abortolegal #quesealey
- ✉ Email agustina.barrero@hotmail.com
- 📖 Looks/tips en mi blog: agusbarrerolook.com

Followed by: lacialurivero, soythelmafardin and 38 others

Grid of fashion photos

### Publicación de marca propia



Instagram

agus.barrero



Para abrir la temporada de #otoño les traigo un nuevo modelo de buzo exclusivo de @bqueen 🥰!

Se me ocurrió este diseño porque me encantan las rosas 🌹 y tenía ganas de mostrarles una nueva prenda #fashion y #alamoda

Visiten la tienda online → [www.bqueen.com.ar](http://www.bqueen.com.ar)

### Publicación de otra marca



Instagram

agus.barrero



Los genios de @yagmour oficial me mandaron este buzo de la nueva temporada de otoño, con 50% OFF!! Me encanta la ropa de Yagmour, así que esto es un mimo al alma 🤍

Visiten la tienda online y comprenlo ya con descuento! → <https://www.yagmour.com.ar/>

### Publicación de marca indefinida



Instagram

agus.barrero



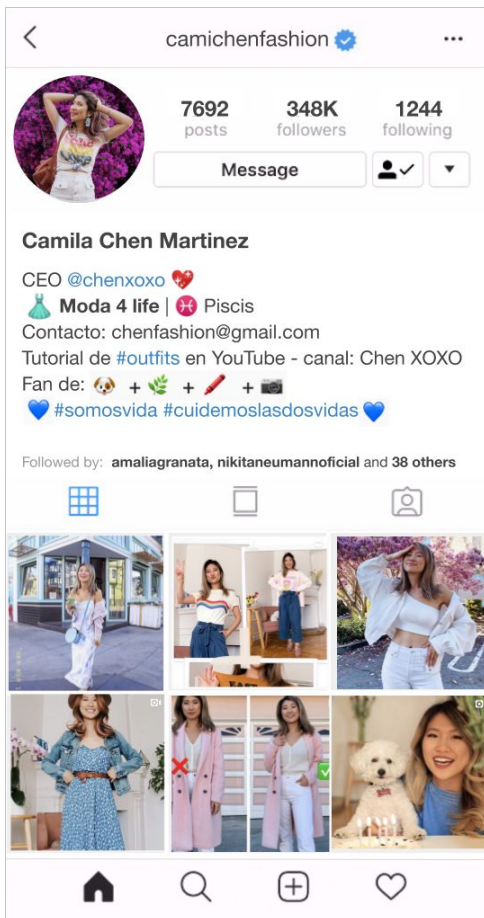
Amé profundamente este buzo para usar con unos jeans 🙌 en estas tardes con viento 🌬️

Cuando no tengo tiempo de elegir un buen #outfit, siempre me gusta tener estas prendas que van con casi todo ✨🌟 #fashionista

Link para comprarlo en mis historias! 🤍

## Perfil en contra de la legalización del aborto "Contra1": *camichenfashion*

### Perfil del Influencer



camichenfashion

7692 posts 348K followers 1244 following

Message

**Camila Chen Martinez**  
CEO @chenxoxo  
Moda 4 life | Piscis  
Contacto: chenfashion@gmail.com  
Tutorial de #outfits en YouTube - canal: Chen XOXO  
Fan de: #somosvida #cuidemoslasvidas

Followed by: amaliagrana, nikitaneumannoficial and 38 others

Grid of fashion photos and a dog.

### Publicación de marca propia



Instagram

camichenfashion



♡ 💬 📌

Nuevamente material exclusivo de @chenxoxo !! Soy fan de la combinación celeste y rosa, se nota? Si aman este buzito tanto como yo, visiten mi tienda online en [www.chenxoxo.com.ar](http://www.chenxoxo.com.ar) y no se pierdan las **promos increíbles** de OTOÑO Les envío XOXOs! #fashion #xoxo #style #ootd

Home Search Add Heart

### Publicación de otra marca



Instagram

camichenfashion



♡ 💬 📌

Paseando por el shopping, decidí comprarme este buzo **DIVINO** de @47street me voló la cabeza No les parece genial? Si aman esta combinación de colores, visiten la tienda online en <https://www.47street.com.ar/> Los leo!! #fashion #fashionlover #style #47street

Home Search Add Heart

### Publicación de marca indefinida



Instagram

camichenfashion



♡ 💬 📌

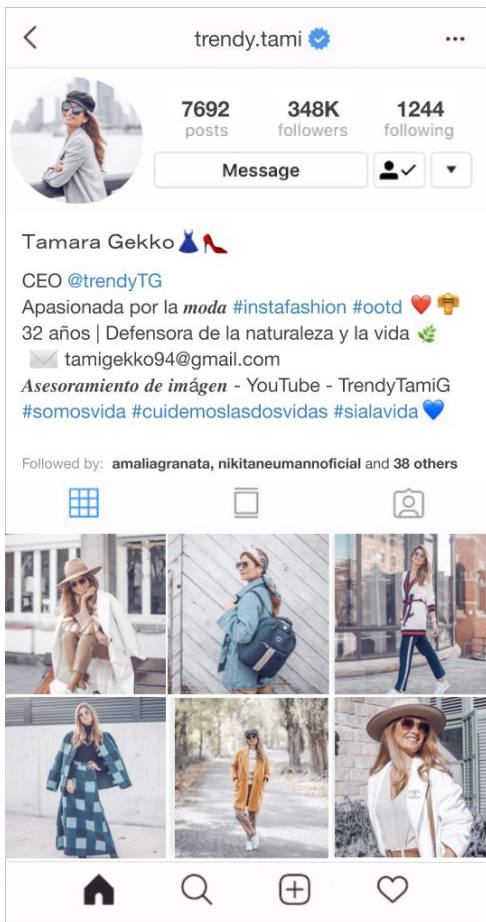
Les dije que soy fan de la combinación celeste y rosa? Hermoso este buzito para usar en estos días otoñales

Compralo YA en el link de mis historias #pink #otoño #fashion #fashionblogger #instagood

Home Search Add Heart

## Perfil en contra de la legalización del aborto "Contra2": *trendy.tami*

### Perfil del Influencer



Instagram profile of **trendy.tami** (verified). The profile shows 7692 posts, 348K followers, and 1244 following. The bio identifies her as Tamara Gekko, CEO of @trendyTG, and describes her as a fashion enthusiast and defender of life. It includes her email (tamigekko94@gmail.com) and mentions her YouTube channel (TrendyTamiG). A grid of six fashion photos is visible below the bio.

### Publicación de marca propia



Instagram post by **trendy.tami** (verified) featuring a grey ribbed sweater with ruffled details. The caption reads: "Nuevo sweater otoñal con mucha onda!! - de mi colección @trendyTG. Más productos y outfits increíbles en mis historias ✨. Seguí buscando tu estilo trendy 🥰". The post includes the hashtags #trendyTG, #trendystyle, #fashionblogger, and #fashion.

### Publicación de otra marca



Instagram post by **trendy.tami** (verified) featuring a grey ribbed sweater with ruffled details. The caption reads: "Hace poco me compré en @zara este sweater para los primeros días de frío 🍂❤️. Buenísimo el gris para esta temporada!! Combinalo con beige, verde, azul, lo que quieras! 🍂. Link de tienda online en mis historias #zara. #trendystyle #fashionblogger #fashion #otoñoivibes".

### Publicación de marca indefinida



Instagram post by **trendy.tami** (verified) featuring a grey ribbed sweater with ruffled details. The caption reads: "Amo los buzos grises para salir a caminar en otoño 🍂 ¿Ustedes qué opinan? Si quieren tener YA este buzo muy trendy, busquen el link en mis historias 🥰". The post includes the hashtags #trendystyle, #fashionblogger, #fashion, and #otoñoivibes.

## 8.7 Pretest: Preguntas de la Etapa Pre-Experimental

A continuación, se podrán ver las preguntas que debieron contestar los participantes del pretest. Todas las preguntas se responden utilizando una escala de Likert de 7-puntos, donde los valores 1 y 7 hacen referencia a respuestas extremas y opuestas.

### [Estudio de Influencers en Instagram]

En el presente estudio, estamos interesados en entender cómo los consumidores responden a las publicidades y perfiles de los influencers en las redes sociales. Te mostraremos varios perfiles de instagram de influencers y sus posts, y luego te pediremos que respondas algunas preguntas referidas a ellos.

>> *Se muestra aleatoriamente uno de los perfiles ficticios*

Luego de haber visto el **perfil** anterior, pienso que...

- Esta persona es una influencer **profesional**  
1 = poco probable; 7 = muy probable
- Esta persona es una influencer **amateur**  
1 = poco probable; 7 = muy probable
- Esta persona **tiene muchos followers**  
1 = muy en desacuerdo; 7 = muy de acuerdo
- Esta persona tiene su **propia marca**  
1 = poco probable; 7 = muy probable
- A esta persona **le pagan por publicitar** productos y servicios de varias marcas  
1 = poco probable; 7 = muy probable
- Esta persona publicita productos y servicios de **su propia marca**  
1 = poco probable; 7 = muy probable
- Esta persona está...  
1= definitivamente en contra de la legalización del aborto; 7=definitivamente pro legalización del aborto
- Esta persona es:  
1= definitivamente una influencer de moda; 7 = definitivamente una influencer de belleza

>> *Se muestran dos publicaciones (una de marca propia del influencer y la otra de otra marca o sin marca) y para cada una de ellas se debe contestar lo siguiente:*

Después de observar el post anterior pienso que...

- El producto publicitado está asociado fuertemente con la persona que lo publicita  
1 = para nada; 7 = extremadamente
- Si comprara este producto, sería claro que me identifico con esta persona  
1 = para nada; 7 = extremadamente
- Si comprara este producto, sería claro para los demás que apoyo a esta persona  
1 = para nada; 7 = extremadamente
- El producto publicitado se ajusta a las habilidades (al dominio?) de la persona que lo publicita  
1 = para nada; 7 = extremadamente
- La persona publicitando este producto/servicio tiene la competencia (está capacitado) para hacerlo  
1 = para nada; 7 = extremadamente
- La persona publicitando este producto/servicio tiene la credibilidad para hacerlo  
1 = para nada; 7 = extremadamente

## 8.8 Experimento Principal: Influencer Personas y sus respectivas publicaciones en Instagram

### Perfil a favor de la legalización del aborto: Versión Final

#### Perfil del Influencer

camichenfashion

7692 posts | 348K followers | 1244 following

Message

**Camila Chen Martinez**  
CEO @chenxoxo  
25 | Te tiro tips de moda | #FashionBlogger  
Argentina  
Feminista y viajera | Escorpiana  
#AbortoLegalYa #NiUnaMenos #QueSeaLey  
chenfashion@gmail.com

Followed by: lacialurivero, soythelmalfardin and 38 others

Grid of 8 fashion-related images.

#### Publicación de marca propia

Instagram

camichenfashion

Nuevamente material exclusivo de @chenxoxo !! Soy fan de esta combinación, se nota? Si aman este buzito tanto como yo, visiten mi tienda online en [www.chenxoxo.com.ar](http://www.chenxoxo.com.ar) y no se pierdan las **promos increíbles** de OTOÑO Les envío XOXOs! #fashion #xoxo #style #ootd

#### Publicación de otra marca

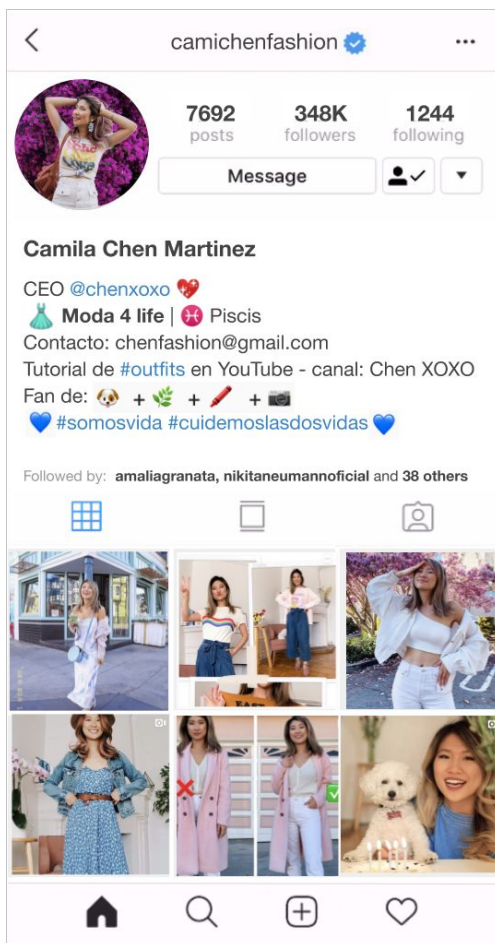
Instagram

camichenfashion

Paseando por el shopping, decidí comprarme este buzo **DIVINO** de @47street me voló la cabeza No les parece genial? Si aman esta combinación de colores, visiten la tienda online en <https://www.47street.com.ar/> Los leo!! #fashion #fashionlover #style #47street

## Perfil en contra de la legalización del aborto: Versión Final

### Perfil del Influencer



camichenfashion

7692 posts 348K followers 1244 following

Message

**Camila Chen Martinez**  
CEO @chenxoxo ❤️  
Moda 4 life | Piscis  
Contacto: chenfashion@gmail.com  
Tutorial de #outfits en YouTube - canal: Chen XOXO  
Fan de: 🍷 + 🌿 + 🍷 + 📧  
#somosvida #cuidemoslasdosvidas

Followed by: amaliagranaata, nikitaneumannoficial and 38 others

Grid of 8 fashion-related images

### Publicación de marca propia



Instagram

camichenfashion

Nuevamente material exclusivo de @chenxoxo ❤️ !!  
Soy fan de esta combinación, se nota? 🌸❤️ Si aman este buzito tanto como yo, visiten mi tienda online en [www.chenxoxo.com.ar](http://www.chenxoxo.com.ar) y no se pierdan las promos increíbles de OTOÑO 🍂❤️  
Les envío XOXOs! #fashion #xoxo #style #ootd

### Publicación de otra marca



Instagram

camichenfashion

Paseando por el shopping, decidí comprarme este buzo DIVINO de @47street 🍷 me voló la cabeza 🤪  
No les parece genial?  
Si aman esta combinación de colores 🌸❤️, visiten la tienda online en <https://www.47street.com.ar/>  
Los leo!! #fashion #fashionlover #style #47street

## 8.9 Experimento Principal: Preguntas del Experimento Final

La encuesta realizada en el experimento final cuenta con tres secciones. La primera sección tiene como objetivo estudiar las preferencias del consumidor con respecto a los perfiles ficticios construidos. Se le mostrará a cada participante de manera aleatoria alguno de los perfiles anteriores (a favor del aborto o en contra) junto con alguna de las dos publicaciones descritas (de la marca propia del influencer o de otra marca). A partir de esto, el participante deberá contestar cuál es su disposición a pagar por el producto, a comprarlo, recomendarlo o republicarlo. La segunda sección de preguntas buscan entender la personalidad del consumidor y la tercera sección corresponde a preguntas demográficas o personales, relacionadas con la opinión del participante sobre la legalización del aborto.

## **[Estudio sobre Preferencias del Consumidor]**

En este estudio, estamos interesadas en ciertas preferencias del consumidor. Te pediremos que te enfoques en el siguiente escenario corto e imagines que es real. Haz tu mejor esfuerzo por fingir que realmente estás en el escenario. Luego, responde cada pregunta como si estuvieras en el escenario en este preciso momento. Es decir, responde cómo te sentirías o comportarías en este momento, en ese escenario.

*[page break]*

### **Escenario seleccionado: Instagram Influencer**

Por favor, revisa cuidadosamente el siguiente perfil de instagram e imagina que estuviste siguiendo a esta persona por un buen tiempo. A menudo miras el perfil de esta persona en busca de inspiración y a veces compras lo que esta persona publicita.

*>> Se le muestra al usuario uno de los perfiles al azar (a favor o en contra de la legalización del aborto)*

*[page break]*

Ahora, imagina que esta persona que estás siguiendo acaba de publicar sobre el siguiente producto.

*>> Se le muestra al usuario una publicación al azar del Influencer (marca propia -autoexpresivo- u otra marca)*

- ¿Cuán atractivo te parece el producto que se muestra en la publicación?  
1 = nada atractivo; 7 = muy atractivo
- ¿Cuán probable es que compres el producto que se muestra en la publicación?  
1 = poco probable; 7 = muy probable
- ¿Cuál es el precio máximo (en pesos argentinos) que pagaría por el producto presentado en la publicación?  
\$ \_\_\_\_\_ (ARS)
- ¿Cuán probable es que recomiendes este influencer a otras personas (amigos, familia)?  
1 = poco probable; 7 = muy probable



- ¿Qué tan probable es que repostees contenido publicado por este influencer?  
1 = poco probable; 7 = muy probable

*[page break]*

### **[Estudio sobre la Personalidad del Consumidor]**

El siguiente estudio tiene como objetivo comprender la personalidad del consumidor. En particular, te mostraremos una serie de afirmaciones sobre ti y te pediremos que evalúes en qué medida estás de acuerdo o en desacuerdo con ellas.

*[page break]*

#### **>> [Escala de Identidad Moral]**

A continuación se enumeran algunas características que pueden describir a una persona:

*Afectuoso, compasivo, justo, amigable, generoso, servicial, trabajador, honesto y amable.*

La persona con estas características podrías ser tú, o podría ser alguien más. Por un momento, visualiza en tu mente el tipo de persona que tiene estas características. Imagina cómo esta persona piensa, siente y actúa. Cuando tengas una clara imagen de cómo sería esta persona, contesta las siguientes preguntas.

#### *>> Sub-Escala de Internalización*

1. Me sentiría bien si fuera una persona con estas características.  
1 = Totalmente en desacuerdo and 5 = Totalmente de acuerdo.
2. Ser una persona con estas características es una parte importante de quien soy.  
1 = Totalmente en desacuerdo and 5 = Totalmente de acuerdo.
3. Me sentiría avergonzada/o si fuera una persona con estas características. (R)  
1 = Totalmente en desacuerdo and 5 = Totalmente de acuerdo.
4. Tener estas características no es realmente importante para mí. (R)  
1 = Totalmente en desacuerdo and 5 = Totalmente de acuerdo.
5. Deseo profundamente tener estas características.  
1 = Totalmente en desacuerdo and 5 = Totalmente de acuerdo.

#### *>> Sub-Escala de Simbolización*

1. A menudo uso ropa que me identifica con estas características.  
1 = Totalmente en desacuerdo and 5 = Totalmente de acuerdo.
2. Las cosas que hago en mi tiempo libre (ej: hobbies) me identifican claramente como alguien que tiene estas características.  
1 = Totalmente en desacuerdo and 5 = Totalmente de acuerdo.
3. El tipo de libros y revistas que leo me identifican como una persona que tiene estas características.  
1 = Totalmente en desacuerdo and 5 = Totalmente de acuerdo.
4. El hecho de que tengo estas características es comunicado a otros por mi pertenencia a ciertas organizaciones.  
1 = Totalmente en desacuerdo and 5 = Totalmente de acuerdo.
5. Estoy involucrada/o en actividades que comunican a otros que tengo estas características.  
1 = Totalmente en desacuerdo and 5 = Totalmente de acuerdo.

[Nota: Todas las respuestas se responden utilizando una escala de Likert de 5-puntos. (R) indica aquellas respuestas que luego requerirán que se las exprese de manera inversa - en otras palabras, que el valor extremo 5 pase a ser el 1 y viceversa - .]

### **[Información Demográfica]**

Cuál es tu género?

- Femenino
- Masculino
- Otro\_\_ (por favor especifique)

¿Qué edad tienes?

\_\_\_\_\_

¿Tiene hijos?

- Si
- No

¿Tienes una cuenta en Instagram? Sí, No

(Si responde que Sí) →

→ ¿Qué tan seguido utilizas tu cuenta de Instagram? Nunca - Casi nunca - A veces - Bastante - Demasiado

¿Cuántos seguidores tienes? \_\_\_\_\_

¿A cuántas cuentas sigues? \_\_\_\_\_

¿Seguís influencers de la categoría moda? Si, No

**>> [Opinión sobre la Legalización del Aborto y otros valores]**

- Consideras que eres una persona que está...  
1= definitivamente en contra de la legalización del aborto; 7=definitivamente pro legalización del aborto
- ¿Te considerás una persona que está al tanto y le interesan las políticas públicas del país?  
1=para nada; 7=extremadamente
- ¿A menudo se ve influenciado por sus principios e ideologías al momento de realizar una compra?  
1=para nada; 7=extremadamente

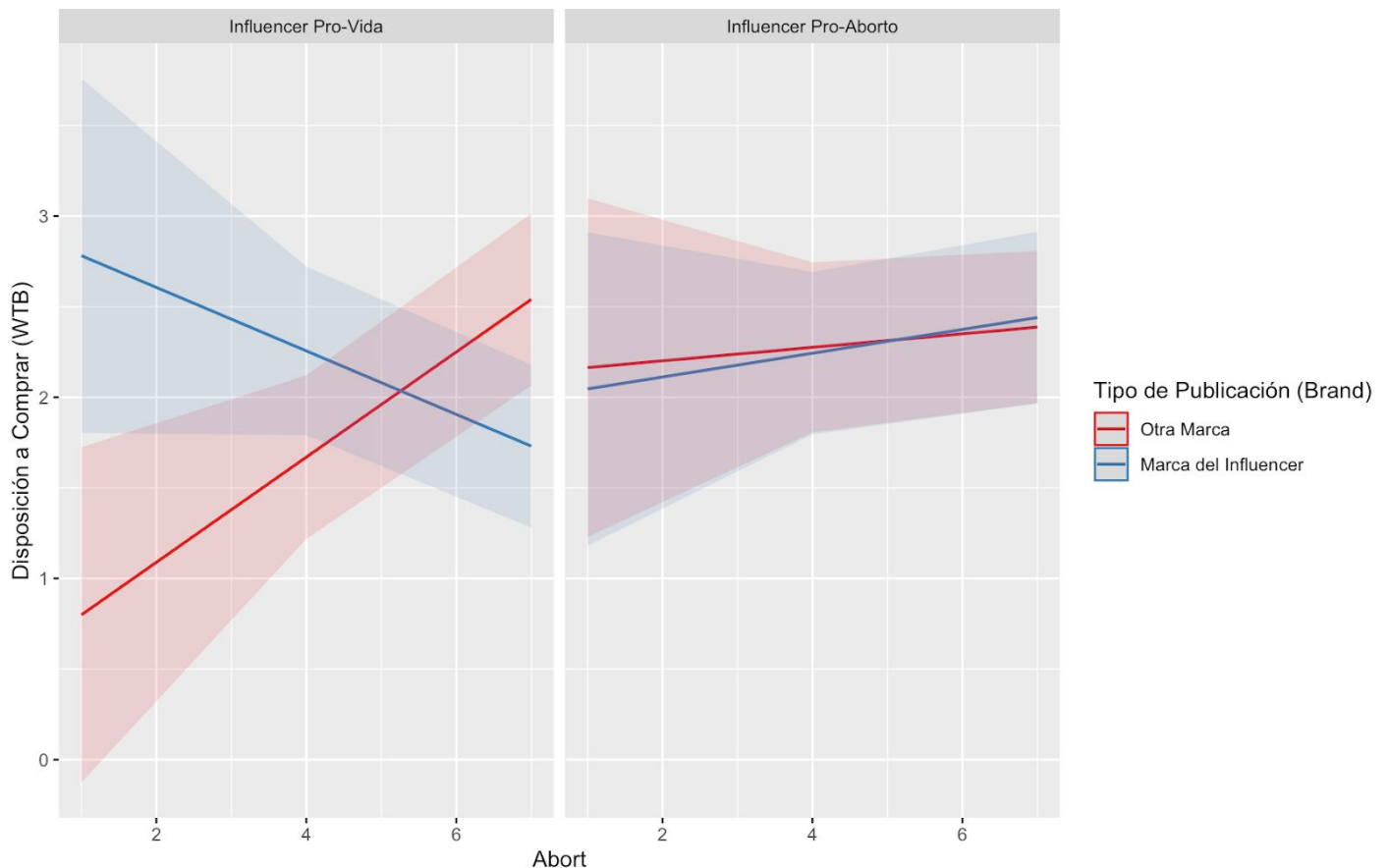
**>> [Chequeo de muestra: que los participantes no hayan participado del pretest]**

- ¿Ya has visto estos perfiles de instagram/preguntas con anterioridad? Si, No
- ¿Conocías los productos de las publicaciones presentadas antes de realizar este estudio?  
Sí/No Cuáles: \_\_\_\_
- ¿Estabas interesado en alguno de los productos de las publicaciones antes de realizar este estudio? Sí/No Cuáles: \_\_\_\_

## 8.10 Experimento Principal: Resultados de Regresiones de la Disposición a Pagar, Recomendar y Republicar

### 8.10.1 Disposición a Comprar (WTB)

Figura 45: Modelos de Regresión para medir el Efecto de la Similitud sobre la Disposición a Comprar de los Consumidores (con intervalos de confianza del 90%)



Al agregar los intervalos de confianza del 90% a las predicciones, puede verse que para aquellos que vieron al Influencer Pro-Vida, la disposición a comprar por la Marca del Influencer es significativamente diferente y mayor a la disposición a comprar por la Otra marca para  $Abort = 1$ . Por ende, los individuos que se identifican como Pro-Vida, al ser similares al influencer, están más dispuestos a comprar la Marca del Influencer. Esto implica que efectivamente existe un proceso de identificación y similitud percibida por parte de los consumidores.

### 8.10.2 Disposición a Pagar (WTP)

Recordemos que para este caso, corrimos la siguiente regresión:

$$WTP = \beta_0 + \beta_1 Brand + \beta_2 Abort + \beta_3 Value + \beta_4 (Brand \times Abort) + \beta_5 (Value \times Abort) + \beta_6 (Value \times Brand) + \beta_7 (Value \times Brand \times Abort)$$

A continuación exhibimos los resultados, que no fueron significativos.

Tabla 17: Efecto de la Similitud entre Influencer y Consumidor sobre la Disposición a Pagar

Variables	Coefficientes Estimados	Errores Estándar (SE)	t	p
<i>Brand</i>	0.0483	0.11778	0.41	0.68195
<i>Abort</i>	0.0089	0.01417	0.627	0.53106
<i>Value</i>	-0.0935	0.11466	-0.815	0.41598
<i>Brand x Abort</i>	-0.0123	0.02041	-0.603	0.54687
<i>Value x Abort</i>	0.0133	0.01956	0.680	0.49743
<i>Value x Brand</i>	-0.0177	0.16163	-0.11	0.91290
<i>Value x Brand x Abort</i>	-0.0008	0.02772	-0.029	0.97700
Constante	0.2323	0.08101	2.868	0.00457 ***

Nota: \*\*\*, \*\*, \* indican significatividad al 1%, 5% y 10% respectivamente. *N* = 213.

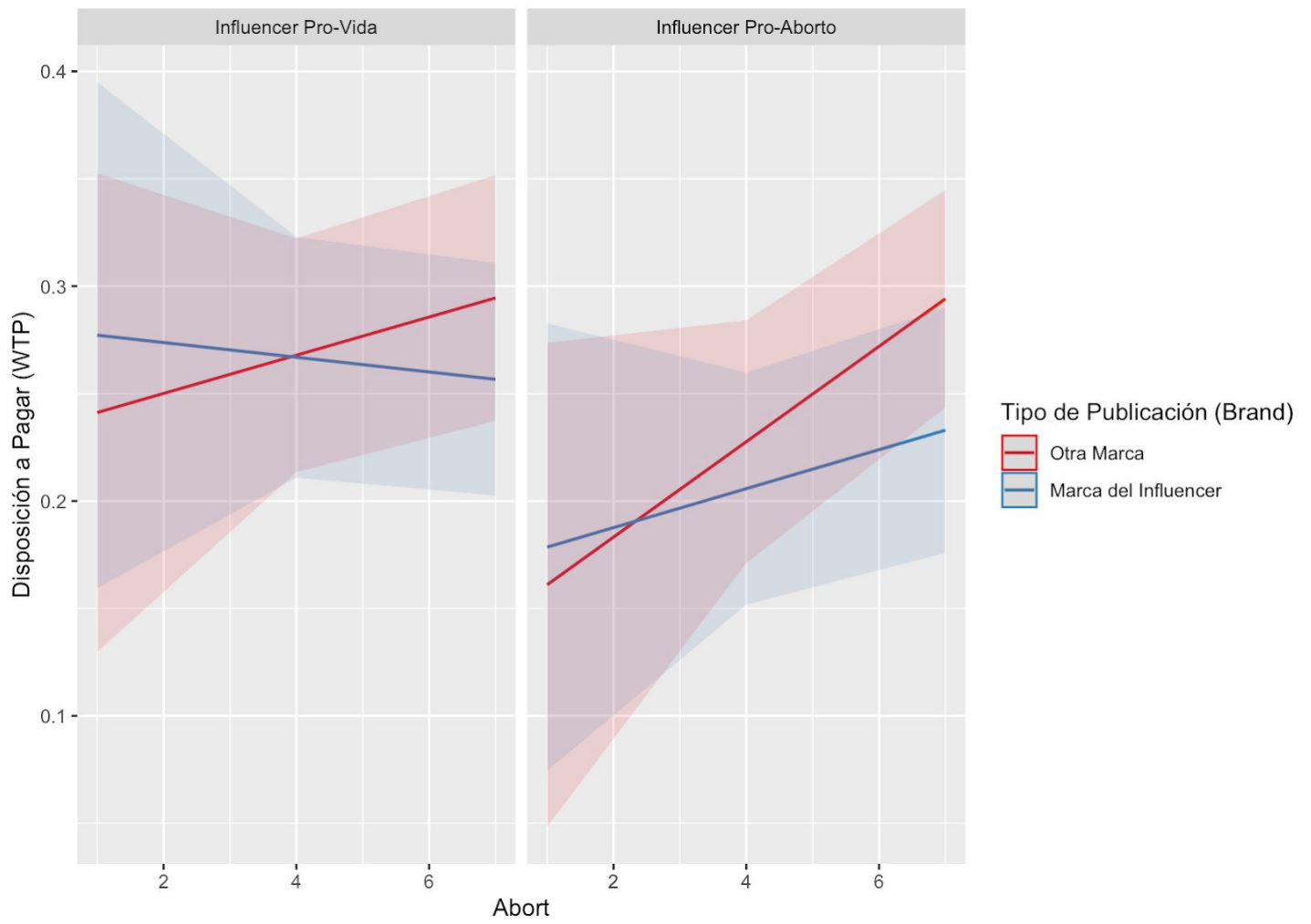
Potencia *post hoc* = 0.415

Tabla 18: Efecto Marginal Total de la variable *Abort*

Modelo	Efecto Marginal Total de <i>Abort</i>	Errores Estándar (SE)	t	p
Otra Marca   Influencer Pro-Vida ( <i>Brand</i> =0, <i>Value</i> =0)	0.008892	0.014172	0.627	0.53106
Otra Marca   Influencer Pro-Aborto ( <i>Brand</i> =0, <i>Value</i> =1)	0.022189	0.013483	1.646	0.10136
Marca del Influencer   Influencer Pro-Vida ( <i>Brand</i> =1, <i>Value</i> =0)	-0.003425	0.014689	-0.233	0.81586
Marca del Influencer   Influencer Pro-Aborto ( <i>Brand</i> =1, <i>Value</i> =1)	0.009072	0.013039	0.696	0.48739

Nota: \*\*\*, \*\*, \* indican significatividad al 1%, 5% y 10% respectivamente. *N* = 213.

Figura 46: Modelos de Regresión para medir el Efecto de la Similitud sobre la Disposición a Pagar de los Consumidores (con intervalos de confianza del 90%)



### 8.10.3 Disposición a Recomendar

Recordemos que para este caso, corrimos la siguiente regresión:

$$\begin{aligned}
 recommend = & \beta_0 + \beta_1 Brand + \beta_2 Abort + \beta_3 Value + \beta_4 (Brand \times Abort) + \beta_5 (Value \times Abort) \\
 & + \beta_6 (Value \times Brand) + \beta_7 (Value \times Brand \times Abort)
 \end{aligned}$$

A continuación exhibimos los resultados, que no fueron significativos.

Tabla 19: Efecto de la Similitud entre Influencer y Consumidor sobre la Disposición a Recomendar

Variables	Coefficientes Estimados	Errores Estándar (SE)	<i>t</i>	<i>p</i>
<i>Brand</i>	0.3509	1.05660	0.332	0.740
<i>Abort</i>	-0.0365	0.12714	-0.287	0.774
<i>Value</i>	-1.4411	1.02859	-1.401	0.163
<i>Brand x Abort</i>	-0.1280	0.18311	-0.699	0.485
<i>Value x Abort</i>	0.2013	0.17549	1.147	0.253
<i>Value x Brand</i>	0.4376	1.44995	0.302	0.763
<i>Value x Brand x Abort</i>	-0.0028	0.24868	-0.011	0.991
Constante	3.1782	0.72672	4.373	0.000 ***

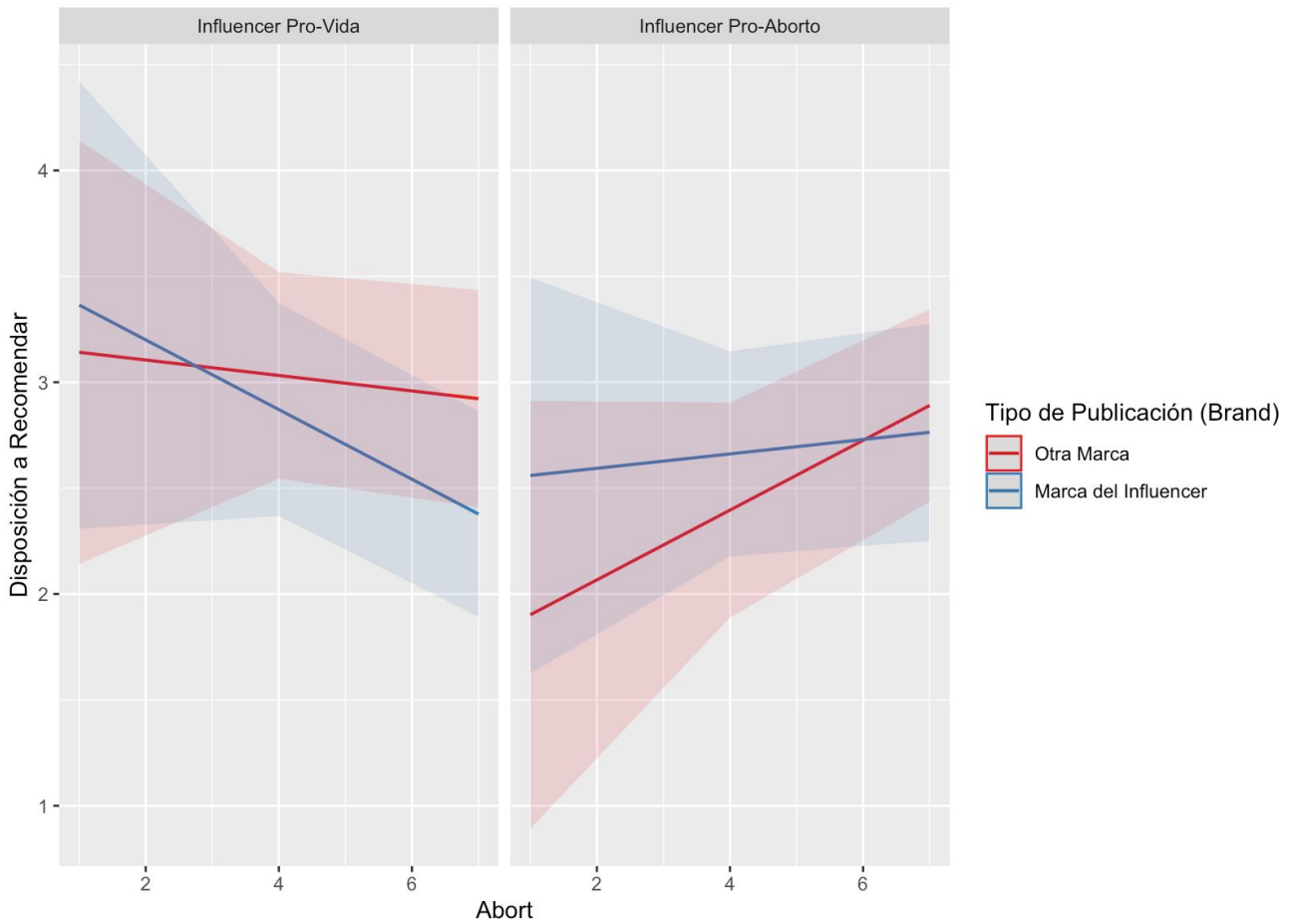
Nota: \*\*\*, \*\*, \* indican significatividad al 1%, 5% y 10% respectivamente. *N* = 213.  
Potencia *post hoc* = 0.317

Tabla 20: Efecto Marginal Total de la variable *Abort*

Modelo	Efecto Marginal Total de <i>Abort</i>	Errores Estándar (SE)	<i>t</i>	<i>p</i>
Otra Marca   Influencer Pro-Vida ( <i>Brand</i> =0, <i>Value</i> =0)	-0.0365	0.12714	-0.287	0.774
Otra Marca   Influencer Pro-Aborto ( <i>Brand</i> =0, <i>Value</i> =1)	0.1648	0.12096	1.362	0.175
Marca del Influencer   Influencer Pro-Vida ( <i>Brand</i> =1, <i>Value</i> =0)	-0.1645	0.13178	-1.249	0.213
Marca del Influencer   Influencer Pro-Aborto ( <i>Brand</i> =1, <i>Value</i> =1)	0.0340	0.11697	0.291	0.772

Nota: \*\*\*, \*\*, \* indican significatividad al 1%, 5% y 10% respectivamente. *N* = 213.

Figura 47: Modelos de Regresión para medir el Efecto de la Similitud sobre la Disposición a Recomendar de los Consumidores (con intervalos de confianza del 90%)



#### 8.10.4 Disposición a Republicar

Recordemos que para este caso, corrimos la siguiente regresión:

$$\begin{aligned}
 repost = & \beta_0 + \beta_1 Brand + \beta_2 Abort + \beta_3 Value + \beta_4 (Brand \times Abort) + \beta_5 (Value \times Abort) \\
 & + \beta_6 (Value \times Brand) + \beta_7 (Value \times Brand \times Abort)
 \end{aligned}$$

A continuación exhibimos los resultados, que no fueron significativos.



Tabla 21: Efecto de la Similitud entre Influencer y Consumidor sobre la Disposición a Republicar

Variables	Coefficientes Estimados	Errores Estándar (SE)	<i>t</i>	<i>p</i>
<i>Brand</i>	-0.2646	0.76827	-0.344	0.731
<i>Abort</i>	-0.0893	0.09244	-0.966	0.335
<i>Value</i>	0.5327	0.74790	0.712	0.477
<i>Brand x Abort</i>	0.0158	0.13314	0.118	0.906
<i>Value x Abort</i>	-0.0175	0.12760	-0.137	0.891
<i>Value x Brand</i>	-0.2425	1.05429	-0.23	0.818
<i>Value x Brand x Abort</i>	0.0054	0.18082	0.03	0.976
Constante	2.0593	0.52841	3.897	0.000 ***

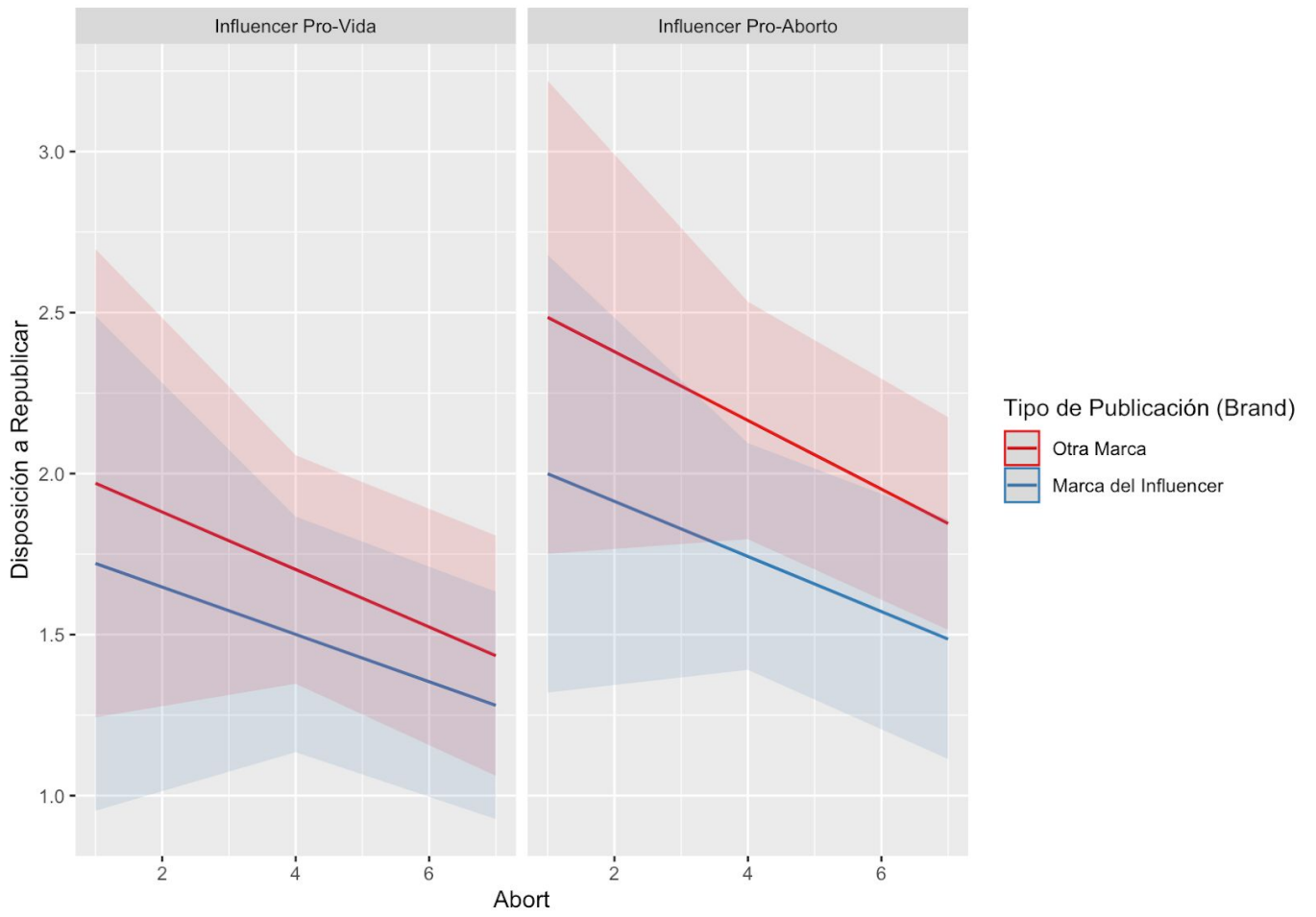
Nota: \*\*\*, \*\*, \* indican significatividad al 1%, 5% y 10% respectivamente. *N* = 213.  
Potencia *post hoc* = 0.657

Tabla 22: Efecto Marginal Total de la variable *Abort*

Modelo	Efecto Marginal Total de <i>Abort</i>	Errores Estándar (SE)	<i>t</i>	<i>p</i>
Otra Marca   Influencer Pro-Vida ( <i>Brand</i> =0, <i>Value</i> =0)	-0.0893	0.09244	-0.966	0.335
Otra Marca   Influencer Pro-Aborto ( <i>Brand</i> =0, <i>Value</i> =1)	-0.1067	0.08795	-1.214	0.226
Marca del Influencer   Influencer Pro-Vida ( <i>Brand</i> =1, <i>Value</i> =0)	-0.0735	0.09582	-0.767	0.444
Marca del Influencer   Influencer Pro-Aborto ( <i>Brand</i> =1, <i>Value</i> =1)	-0.0856	0.08505	-1.006	0.316

Nota: \*\*\*, \*\*, \* indican significatividad al 1%, 5% y 10% respectivamente. *N* = 213.

Figura 48: Modelos de Regresión para medir el Efecto de la Similitud sobre la Disposición a Republicar de los Consumidores (con intervalos de confianza del 90%)



## 8.11 Experimento Principal: Cálculo de Potencia $(1 - \beta)$ post hoc para todas las Regresiones

Con el objetivo de tener una mayor visibilidad de cuán significativo es el tamaño muestral ( $N=213$ ) de nuestro experimento principal para detectar diferencias (o no) entre los distintos tratamientos, calculamos la potencia  $(1 - \beta)$  post hoc para todas las regresiones llevadas a cabo en la investigación. A su vez, agregamos este cálculo para las regresiones cuya variable dependiente es la disposición a pagar, recomendar y republicar e incluyen variables de control (MID y nivel de influencia). Agregamos también el tamaño de muestra que hubiera sido necesario para obtener un nivel de potencia post hoc igual a 0.80.

Tabla 23: Potencia post hoc para todas las regresiones realizadas

Modelo	Resultado	Potencia $(1 - \beta)$ post hoc	$N$ post hoc
<i>Willingness To Pay (WTP)</i>	no significativo	0.4149428	456
<i>Willingness To Buy (WTB)</i>	<b>significativo</b>	0.6547901	281
<i>Willingness To Recommend</i>	no significativo	0.3167256	594
<i>Willingness To Repost</i>	no significativo	0.6560158	280
<i>Willingness To Pay (WTP)</i> incluyendo MID como control	no significativo	0.5200911	362
<i>Willingness To Buy (WTB)</i> incluyendo MID como control	<b>significativo</b>	0.8253536	202
<i>Willingness To Recommend</i> incluyendo MID como control	no significativo	0.6790249	269
<i>Willingness To Repost</i> incluyendo MID como control	no significativo	0.7626311	229
<i>Willingness To Pay (WTP)</i> incluyendo <i>Influence</i> como control	no significativo	0.5180056	363
<i>Willingness To Buy (WTB)</i> incluyendo <i>Influence</i> como control	<b>significativo</b>	0.6588713	278
<i>Willingness To Recommend</i> incluyendo <i>Influence</i> como control	no significativo	0.3432940	547
<i>Willingness To Repost</i> incluyendo <i>Influence</i> como control	no significativo	0.8188142	205