



Universidad Torcuato di Tella

Master in Management + Analytics

**Marketing de Influencers en
Instagram: un framework para
encontrar la mejor estrategia de
marketing basado en clustering,
influencers de distinto tamaño y
encaje entre producto-sponsor**

Teresita Inés Cancelo
Tutora: Elena Fumagalli

RESUMEN

A partir del surgimiento 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 “sponsoreen” los productos de las empresas o marcas a través de sus publicaciones, que son 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 sus productos de manera que se generen mayores ganancias para 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 éstas 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 popularidad del influencer y el encaje que tenga con el producto sobre la disposición a comprar, 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 influyentes al momento de identificar distintos grupos de influencers, mientras que de la etapa experimental, concluimos que la cantidad de seguidores del influencer, en conjunto con el encaje con el producto patrocinado, parecen tener un impacto muy significativo sobre la disposición a comprar, pagar y republicar del seguidor. 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 popularidad del influencer y su relación con el producto que se quiera publicitar para tomar la mejor decisión de marketing online.

Influencer Marketing on Instagram: A framework for finding the best marketing strategy based on clustering algorithms, different sizes of influencers and sponsor-product fit.

ABSTRACT

Since the emergence of social networks, companies have slowly abandoned traditional strategies of promoting their products and services, in favor of a new way of marketing their products: the so-called Influencer Marketing. It consists in advertising products through social media influencers (considered “online celebrities” on the Internet) and their posts, which are viewed by millions of followers. However, along with this recent shift towards online marketing strategies, a new business problem also arised, related with selecting the best influencers to promote their products so that more profit is generated. 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 Instagram influencers profiles in order to run an experiment (N = 213) to study the impact of the influencer’s popularity, together with the influencer-product fit, 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 (a way to measure popularity) is one of the most influential variables when identifying different groups of influencers, while from the experimental step, we conclude that the influencer’s amount of followers, in combination with its fit with the product being sponsored, seem to have a significant effect on the willingness to buy, to pay and to republish of the follower. 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 popularity of the influencer and its relation with the product to take an optimal online marketing decision.

1. Introducción.....	4
1.1 Motivación.....	4
1.2 Influencers.....	6
1.3 Instagram.....	10
1.4 Marketing por Redes Sociales.....	12
1.5 El Problema.....	14
1.6 Objetivo.....	15
1.7 Metodología.....	18
2. Los Datos.....	19
3. Análisis Exploratorio.....	22
3.1 Análisis del dataset instagram.....	22
3.2 Análisis dataset instagram_posts.....	32
4. Clustering.....	44
4.1 Ingeniería de Datos, Transformaciones y Nuevas Variables.....	45
4.2 Implementación de k-prototypes.....	48
4.3 Conclusiones de Clustering.....	54
5. Etapa Experimental.....	55
5.1 Popularidad del influencer y relación con su personalidad percibida.....	55
5.2 Diferenciación entre productos con y sin ajuste al influencer.....	56
5.3 Categoría de belleza dentro de los “insta-influencers”.....	56
5.4 Objetivo de realizar un pretest.....	57
5.5 Pretest: Etapa Pre-Experimental.....	59
5.6 Resultados del Pretest.....	60
5.7 Experimento.....	63
5.8 Diseño del experimento.....	64
5.9 Preprocesamiento de los datos del experimento.....	65
5.10 Resultados del Experimento.....	71
6. Conclusión.....	85
7. Apéndice.....	90
7.1 Variables de los Datasets.....	90

7.2 Transformaciones de las Variables.....	91
7.3 Gráficos Adicionales (Análisis Exploratorio).....	93
7.4 Listado Final de Variables utilizadas para hacer Clustering.....	96
7.5 Influencer Personas: Influencers Representativos a partir de la primera versión del algoritmo de Clustering.....	96
7.6 Pretest: Influencer Personas y sus respectivas publicaciones en Instagram.....	100
7.7 Pretest: Preguntas de la Etapa Pre-Experimental.....	105
7.8 Experimento Principal: Influencer Personas y sus respectivas publicaciones en Instagram.....	106
7.9 Experimento Principal: Preguntas del Experimento Final.....	108
8. Bibliografía.....	113

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*¹ 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 ya que no se están teniendo en cuenta las características de la audiencia.

¹ 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).

- 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 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 re-publicar 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+² 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*³ o páginas web que mencionan en promedio en sus publicaciones en

² Lesbianas, Gays, Bisexuales, Transexuales, Intersexuales, Queer y Asexuales (Abidin 2019)

³ 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

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* (re-publicar contenido en Twitter) o, en términos más genéricos, *rebroadcasting* (difusión o re-publicación 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 re-publicación de contenido depende no solo del mensaje, sino también de la compatibilidad o *fit* entre el usuario que re-publica 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 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

la información de las redes sociales (Lahuerta-Otero y Cordero-Gutiérrez 2016). Por ejemplo, *#JustDolt* 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 *#JustDolt*.

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). 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.

Instagram resulta una red social muy popular, sobretodo entre las generaciones más nuevas. Según la página 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 FB 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 la poseían la aplicación. En el primer lugar del ránking estaba Facebook con un 90,9% de los usuarios de móviles con tenencia de la app, en tercer lugar FB messenger con 56,97% y en cuarto lugar twitter con 43,62%. En el mundo profesional del marketing online, instagram toma mucha importancia. 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 ránking lo lidera Facebook con un 94% de los profesionales usándola y en tercer lugar se encuentra twitter con un 59%. Muchas empresas eligen esta red social como target para su presupuesto de marketing. Esto se ve en números ya que Instagram 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 generan (cerca de 69,9 billones de dólares en 2019), el tipo de publicidad que se puede encontrar es diferente al de instagram. Mientras que en Facebook se pueden encontrar en su mayoría publicidades mucho más parecidas al estilo de la televisión (las empresas pagan por 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 abrió el paso para que, además de el 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.

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 Shouten, 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, re-publicar 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 re-publicar. 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 patrocinado 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).

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. En 2019, 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.

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. Debido a esto es que 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 de manera de utilizar 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 alguna combinación entre cantidad de seguidores del influencer y nivel de relación con el producto publicitado 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 ni combinando ambas variables en un mismo experimento. En el paper *When less is more: the impact of macro and micro social media influencers' disclosure*, por Samantha Kay, Rory Mulcahy & Joy Parkinson, se llevó a cabo un experimento para evaluar las reacciones y efectividad de publicaciones patrocinadas en macro influencers y micro influencers. Sus resultados mostraron que los consumidores expuestos a micro influencers reportaron tener mayor nivel de conocimiento de producto y sentir al producto más atractivo. Además, su experimento reveló que cuando un consumidor es expuesto a micro influencers, tiene mayor propensión de compra que al exponerse a macro influencers. Por otro lado, en el paper *Celebrity vs. Influencer endorsements in advertising: the role of identification, credibility, and Product-Endorser fit*, por Alexander P. Schouten, Loes Janssen & Maegan Verspaget, se experimentó sobre la efectividad de los sponsorship en celebridades tradicionales e influencers, mediados por la variable de encaje entre sponsor y producto. Como resultado, encontraron que el encaje entre producto y sponsor afectaba directamente a la actitud hacia la publicidad, siendo que los sponsorships con alto encaje con el sponsor generaron una actitud más positiva y mayor intención de compra que los que no tenían encaje. También descubrieron que no

había una relación significativa entre el tipo de sponsor (celebridad o influencer) y encaje entre sponsor y producto.

Es por esto que este trabajo quiere innovar combinando ambas variables, tamaño del influencer con encaje entre producto y sponsor. De esta forma, estamos combinando un análisis de negocio con un análisis sobre cómo percibe el público los mensajes de los influencers según sus características y el tipo de publicaciones que sponsorea. Al mismo tiempo, nuestro trabajo también se diferenciará de la literatura anteriormente nombrada al hacer uso de influencers ficticios (creados a partir de la primera etapa de clustering) y no de influencers conocidos por la sociedad, ya que puede llegar a introducir cierto sesgo en el análisis si los participantes del experimento conocen o ya siguen al influencer en cuestión.

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 re-publicar. 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 Upfluence⁴, 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**:

⁴ <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_bracke	instagram_id	username
1515915	NA	pumora	de	DE	Roman Northern Gat	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, Can	Gender_FEMALE	18-24	187991	spilledpolish
1617706	hello@goldfieldsgirl.com	goldfieldsgirl	en	AU	Golden Triangle Mote	Gender_FEMALE	25-34	289808	goldfieldsgirl
101274	klemen@ujusansa.si	Klemen	en	CN	NenjiangRiver,China	Gender_FEMALE	0-17	85644	mowgliiii
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, US	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@rebelinawdress.co	by Nicki Nowicki	en	DE	Munich,Germany	Gender_FEMALE	25-34	249968	rebelinawdress
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@myvogueishdiaries	myvogueishdiary	en	CA	Toronto,ON,Canada	Gender_FEMALE	25-34	284996	myvogueishdiaries

Siguientes 10 columnas

full_name	bio	website	followers	engagement	total_posts	total_engag	total_likes	total_comn	category
Hand Embroidery by Ann	» Hand embroide	http://pumora.de	17002	0.83673	722	109411	106156	3255	home
c r y s t a l ☆ dallas blogg	style + beauty #b	http://thedapperg	12547	0.75888	349	52075	46311	5764	beauty
babaluccia	Blogger influence	https://www.baba	18649	0.51515	560	119915	100580	19335	fashion
by Melissa Frusco	Curly Hair • Hous	https://www.thein	32788	0.43222	455	336189	310545	25644	beauty
Oana	Travel & Lifestyle	https://walkcatwa	36418	0.36975	611	358916	325204	33712	travel
London Damsel	Londoner. Travell	http://londondam	7709	0.33272	641	80084	73082	7002	family
Maddy	Lifestyle/Beauty B	https://amzn.to/2	6935	0.31265	107	30308	28912	1396	beauty
Deneale – Goldfields Girl	👄 • Lipstick wield	https://linktr.ee/g	8085	0.31002	305	37457	34503	2954	family
Klemen	SURF LIFE ∞ PHO	http://bit.ly/Sardi	12231	0.29347	57	31997	31125	872	travel
Traveller 🌍 Anthony Mic	👨♂️ Former lose	https://manvscloc	19769	0.28867	235	78538	75735	2803	food
Hana 🌿 Nirvana Cakery	Author of plant-b	https://nirvanaca	10065	0.15618	214	120895	112514	8381	food
Megan Gray	What I love most a	http://houseofgra	11374	0.15616	250	56139	54564	1575	home
Mary Murnane	Fashion Photogr	https://marymurr	13240	0.15494	334	117756	103536	14220	fashion
The Real Simple Good Life	Justin + Erica Win	https://tap.bio/@	59020	0.15424	553	177743	149437	28306	food
Kercia Jane Design	👉 Designer + Art	https://linktr.ee/k	5872	0.15416	158	39446	35491	3955	home
by Nicki Nowicki	👄 Fashion & Anti	https://wp.me/p6	11344	0.15410	231	129429	123698	5731	fashion
Jess Cheng	Senior creative @A	https://youtu.be/	8736	0.15377	370	129273	116759	12514	beauty
Rachel Stokes	wife + mama raisi	http://www.simply	5554	0.15336	239	58768	50567	8201	family
Medge Beauvoir–Yellowle	Fashion Beauty	http://www.myvod	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	B5Bf5EFgVjG	0	image	https://scontent-iad3-1.c
457132836	547505	Quick trip to NYC, my f	picture	Eataly NYC Fla	1573933198	36	2	B48DkmwgSG	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 Park	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.c
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	B2QH0aqqWT	0	image	https://scontent-iad3-1.c
407830311	547505	Left or right, which wo	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 chá	picture	Picomtal	1575103119	2588	2	B5e7BS-l7QW	0	image	https://scontent-iad3-1.c
465971967	1340199	Wherever you go bec	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-Alp	1575102097	1678	0	B5e5Ekj-Cu	0	image	https://scontent-iad3-1.c
465971969	1340199	Happy week end ! 🥰	picture	Provence-Alp	1575101612	2462	0	B5e4JVodZI	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

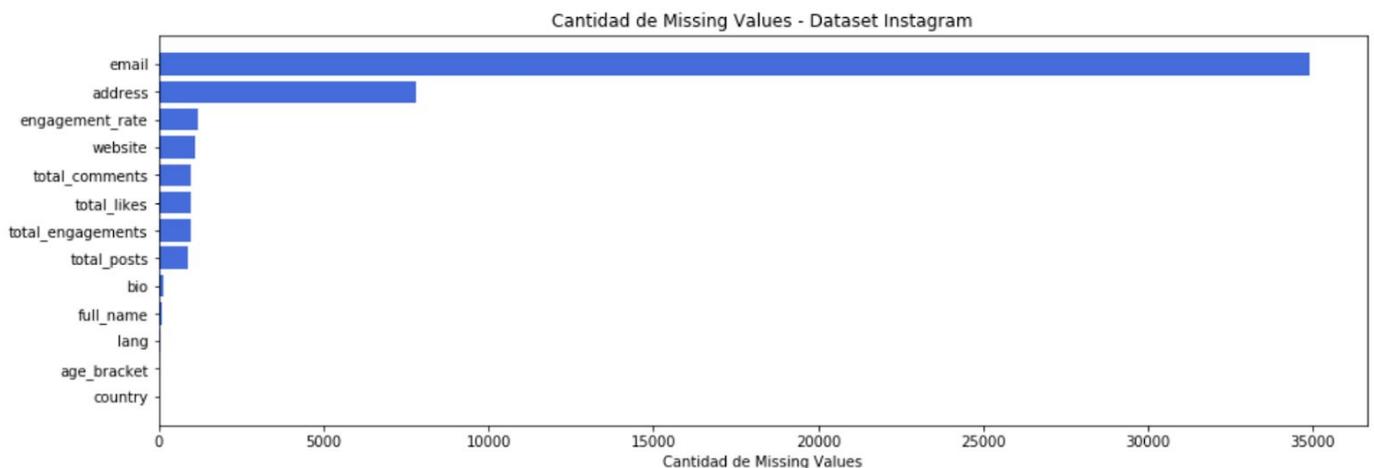
En esta sección haremos un análisis descriptivo y exploratorio de cada dataset para poder obtener los primeros *insights* sobre las características de los influencers y sus posts. Parte de los resultados de este primer análisis nos llevará a realizar transformaciones en las variables para luego utilizarlas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

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.

Figura 1: Cantidad de missing values por variable en la base de datos "instagram"



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'. Procedimos a eliminarlas del dataset. 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)

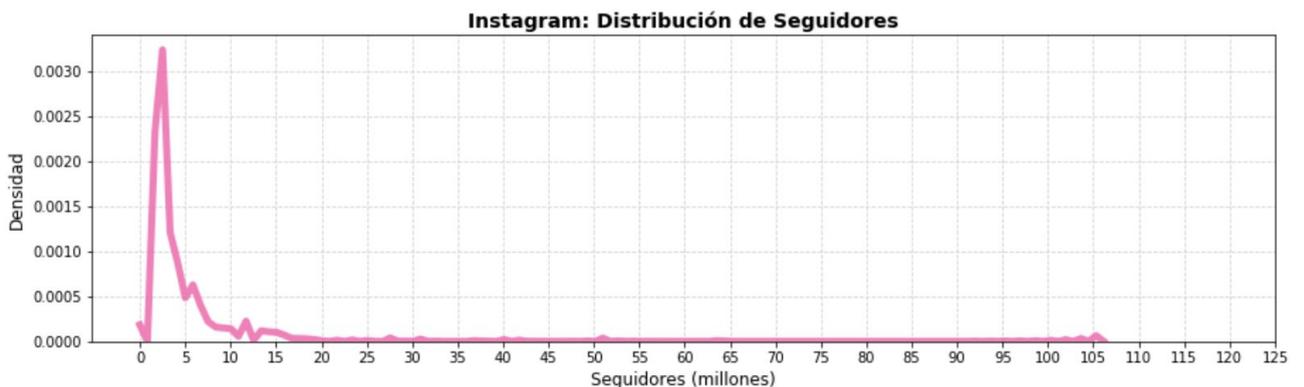
lang	en	pt	es	fr	de	it
Cantidad	90788.0	1240.00	1222.00	1159.00	983.00	831.00
(%) en Dataset	90.9	1.24	1.22	1.16	0.98	0.83

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)



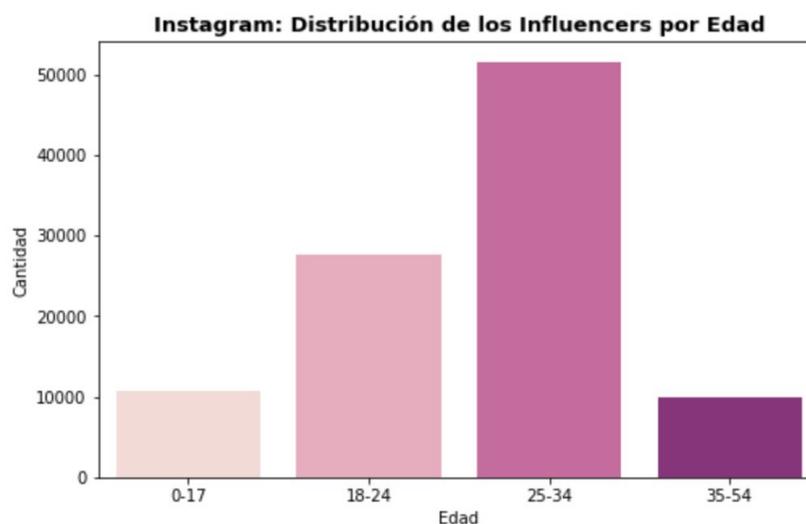
Es claro que esta distribución es asimétrica, es decir, que la mayoría de los influencers tienen entre 0 y 2 millones de seguidores, mientras que unos pocos pueden lograr tener más de 5, 10 o 15 millones. Esto nos está diciendo que contamos con un dataset que cuenta con pocos Mega-Influencers, que son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

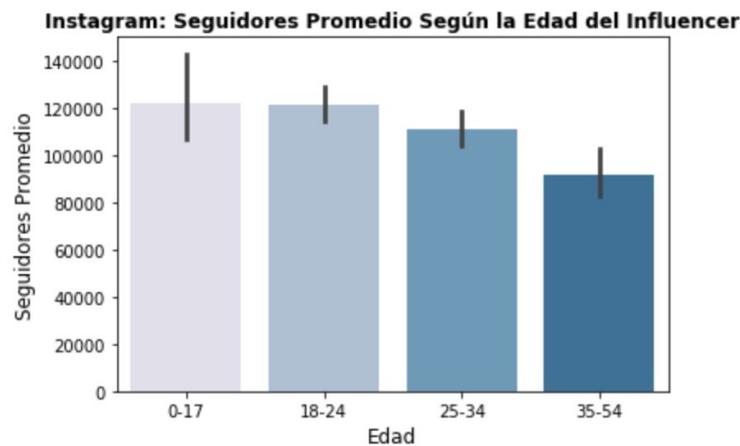


Es intuitivo haber encontrado que la mayoría de los influencers en el dataset tienen entre 25 y 34 años. Es ampliamente sabido que la red social Instagram fue creada recientemente (hace 10 años), e involucra una relación constante con la última tecnología (celulares con pantalla táctil y alta resolución 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

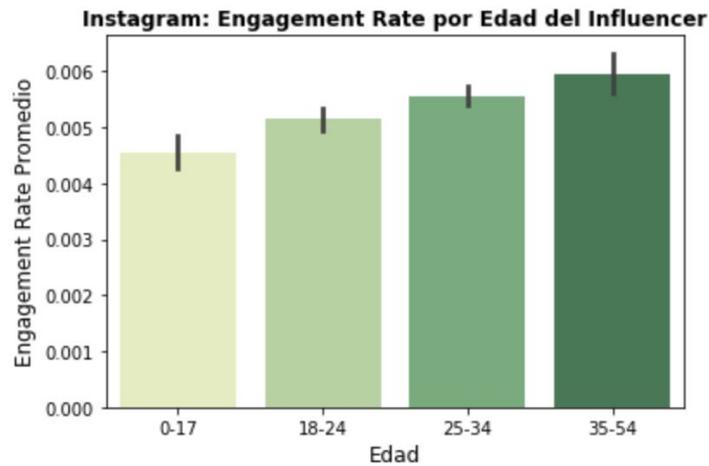


Es muy interesante e intuitiva la conclusión detrás: a mayor edad del influencer, menor es la cantidad promedio de seguidores. Esto es predecible dado lo encontrado anteriormente: los influencers en su mayoría son *centennials* o *millennials*, por lo cual publicarán contenido acorde con su edad y sus 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



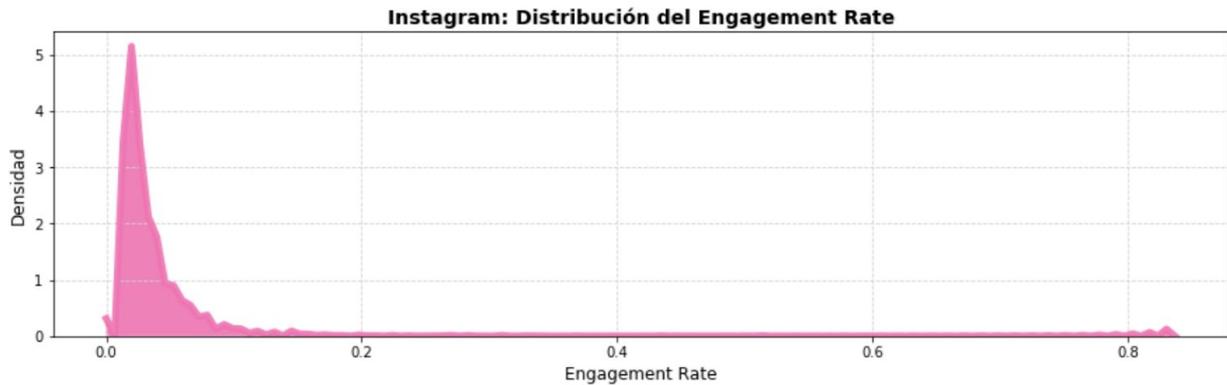
Lo que observamos es que el mayor 'engagement rate' promedio se da en el rango etario 35-54, lo cual nos está diciendo que los influencers que se encuentran dentro de ese rango tienen una mejor medida de performance en promedio con respecto a su visibilidad en Instagram. En algún punto, la conclusión 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'.

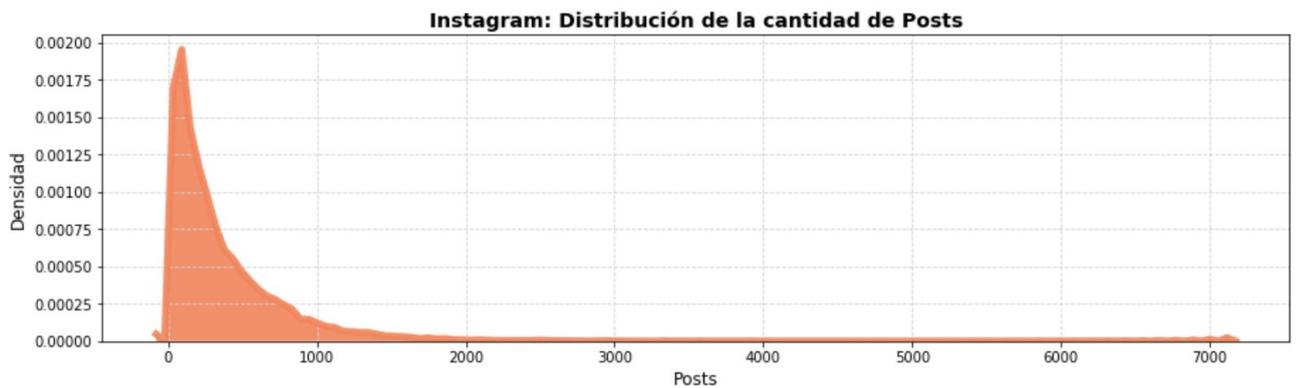
Figura 6: Distribución del Engagement Rate



3.1.7 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

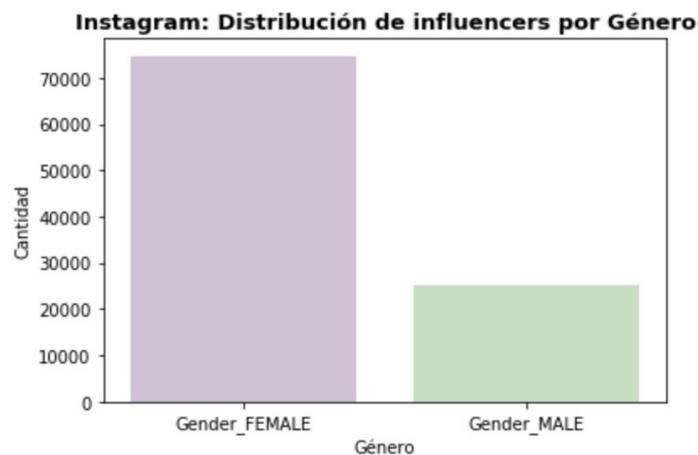


Es importante aclarar que la cantidad de publicaciones de cada influencer es representativa del período para el cual tenemos datos, es decir, comprenden la cantidad de posts realizados en tres meses. Por lo cual, para obtener el número total de publicaciones del influencer, debería multiplicarse la cantidad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

3.1.8 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

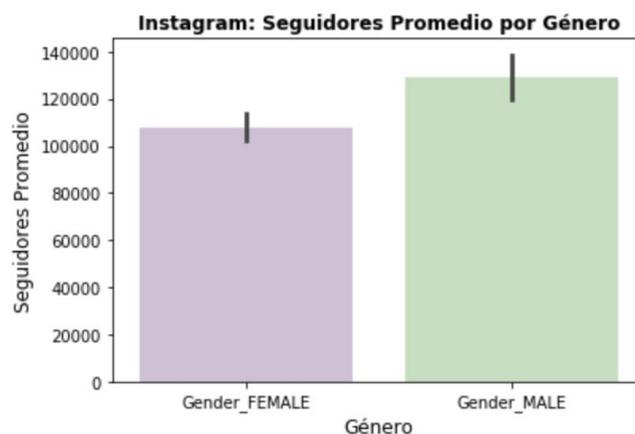


Es claro que el dataset presenta una muestra desbalanceada, en donde tenemos mayoría de influencers siendo de género femenino. Esto en parte es lógico, debido a que algunas de las categorías ('fashion', 'beauty') son casi exclusivamente de mujeres y, por otro lado, se ha registrado en otros trabajos 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.

3.1.9 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

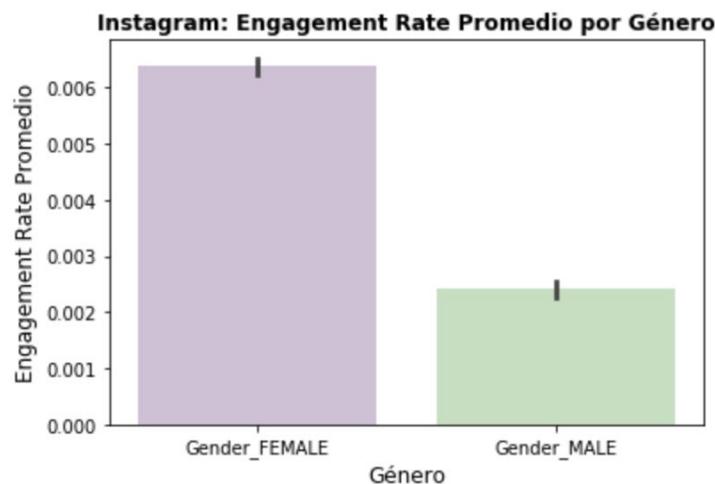


Una explicación posible a este resultado es que los influencers masculinos suelen hacerse primero ‘famosos’ en otras plataformas como YouTube o Twitch, o en otro medios (por ejemplo, pueden ser futbolistas, periodistas, modelos, etc.), y luego trasladan toda esa masa de seguidores a Instagram. Por ende,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

3.1.10 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.

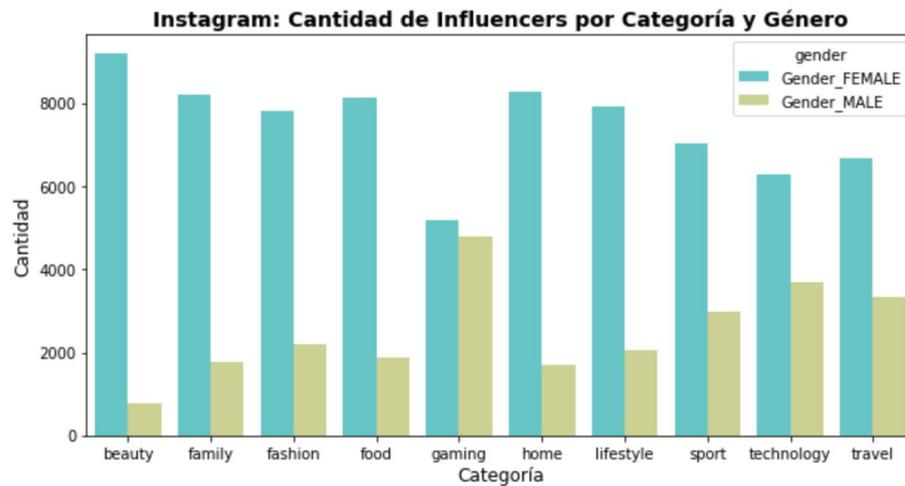
Figura 10: Engagement Rate Promedio por Género



3.1.11 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

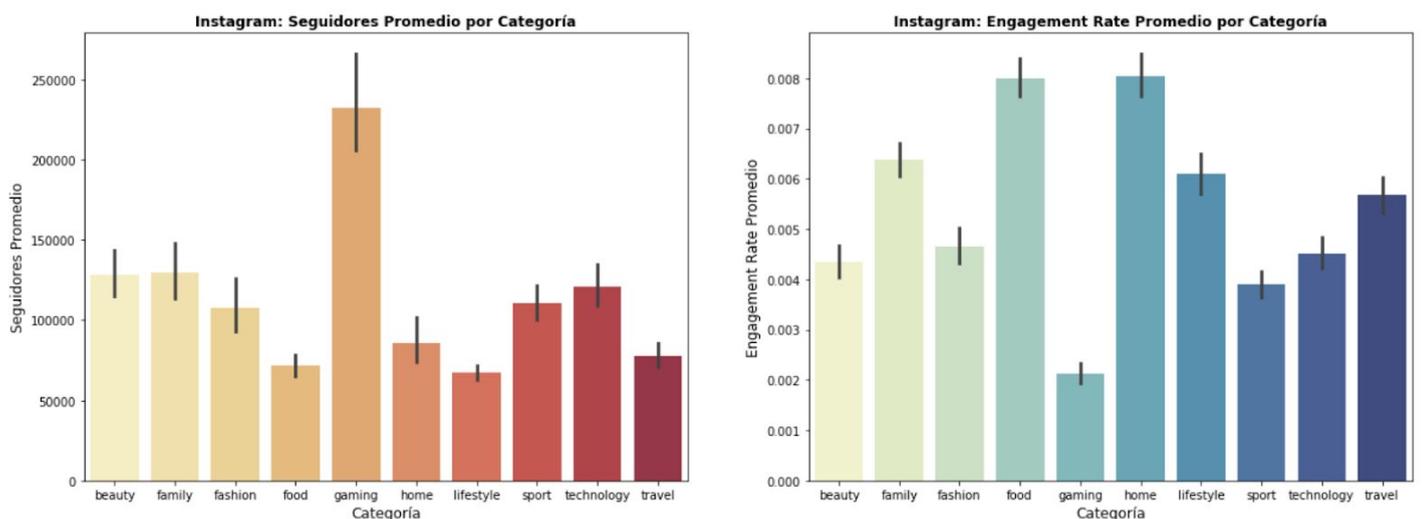


Aquí volvemos a observar la mayoría de mujeres influencers en el dataset. No obstante, vemos que la diferencia entre mujeres y hombres es mucho mayor para categorías como 'beauty', 'family', 'home', 'lifestyle', 'food' y 'fashion'. Luego, para las categorías 'travel', 'technology', 'sport' la diferencia 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*.

3.1.12 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



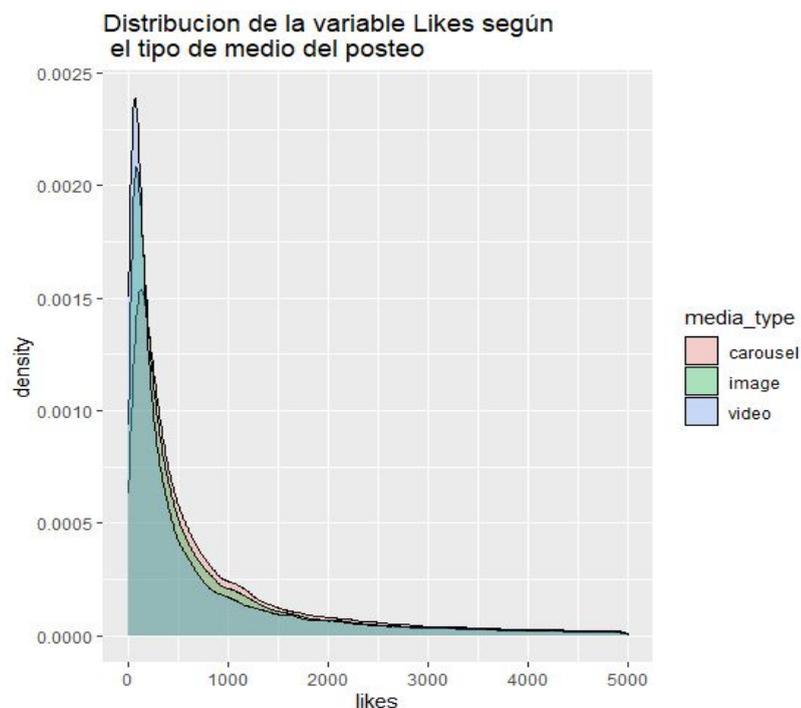
3.2 Análisis dataset *instagram_posts*

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)



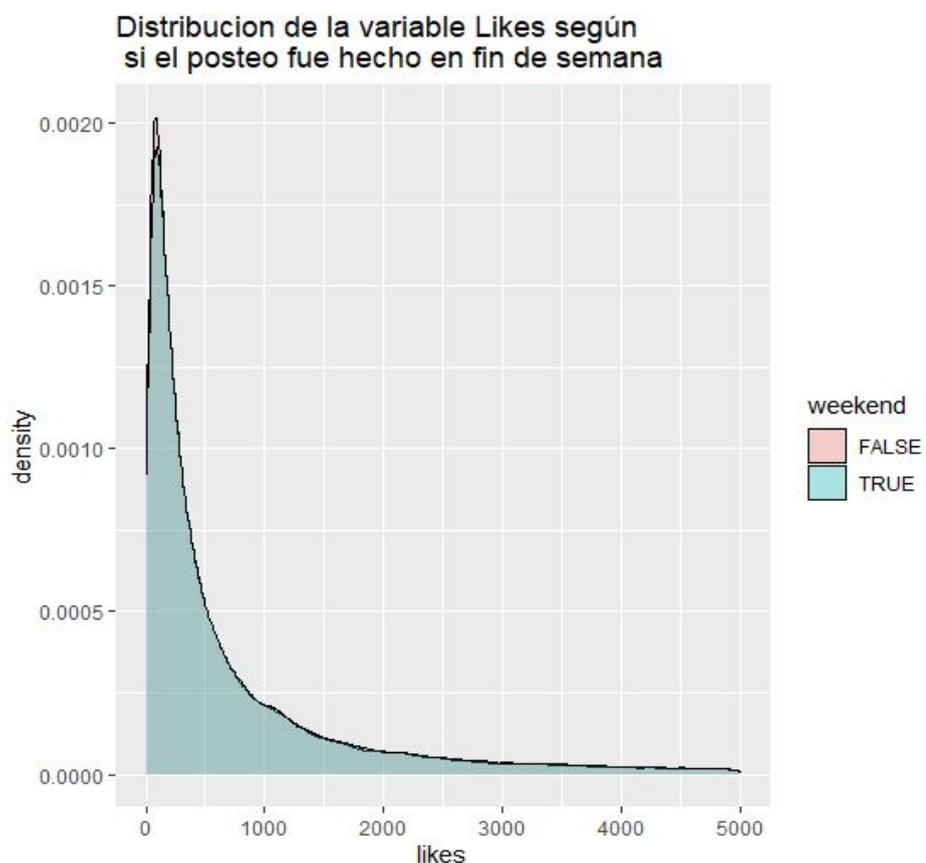
Para este gráfico se utilizaron las variables ‘likes’ y ‘media_type’ (tipo de publicación). Podemos ver que los videos tienen la menor varianza de cantidad de *likes*, los carousel (posteos con un conjunto de imágenes para ver) tienen la mayor varianza y las imágenes están en el medio. Es decir, los *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 parecidas, 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 carruseles 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

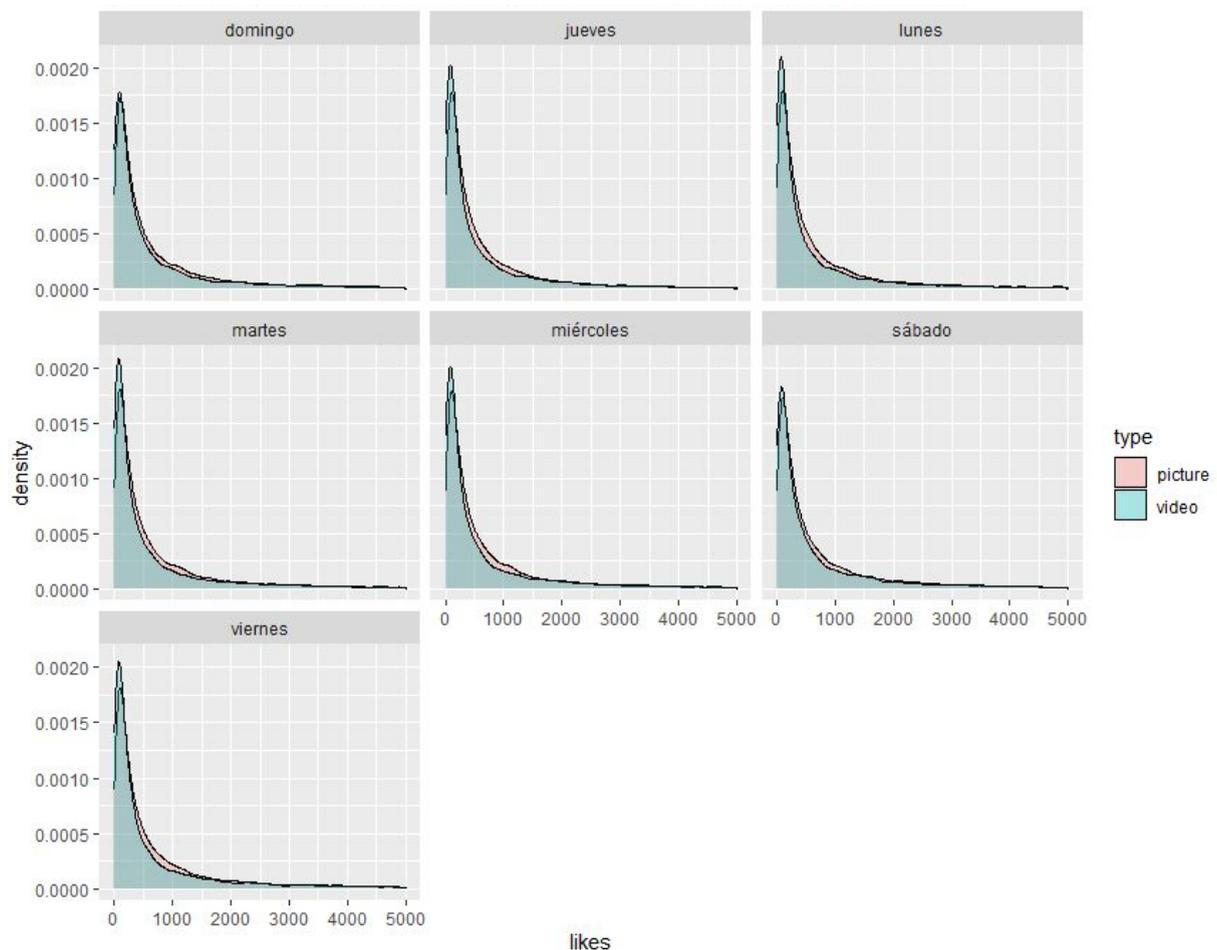


Para este gráfico se utilizaron las variables 'likes' y 'weekend' (variable creada a partir del dataset que es igual a TRUE si el posteo fue hecho en un fin de semana y FALSE si fue en un día de semana). No hay una diferencia apreciable entre la distribución de los *likes* a posts publicados durante 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



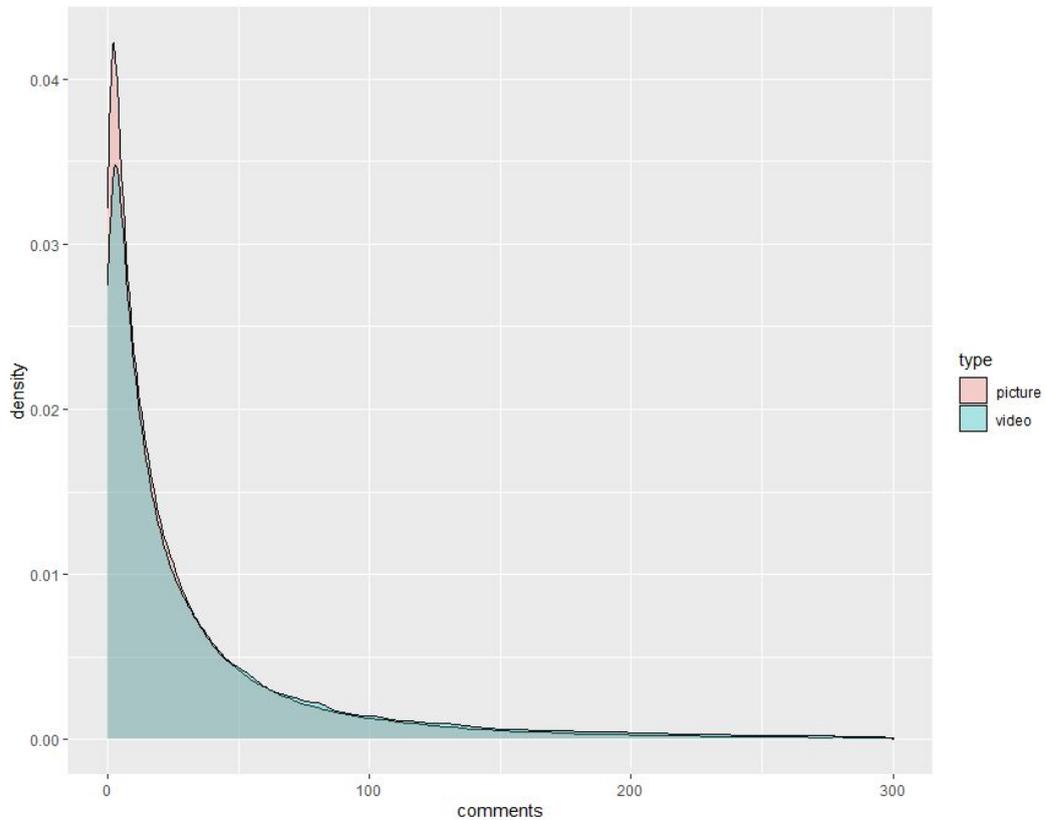
Para este gráfico se utilizaron las variables 'likes', 'weekday' (día de la semana en que se hizo la publicación) y 'type' (tipo de publicación: imagen o video). Se puede ver en estos gráficos que las distribuciones de *likes* de publicaciones hechas los sábados y los domingos es igual para videos y 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.

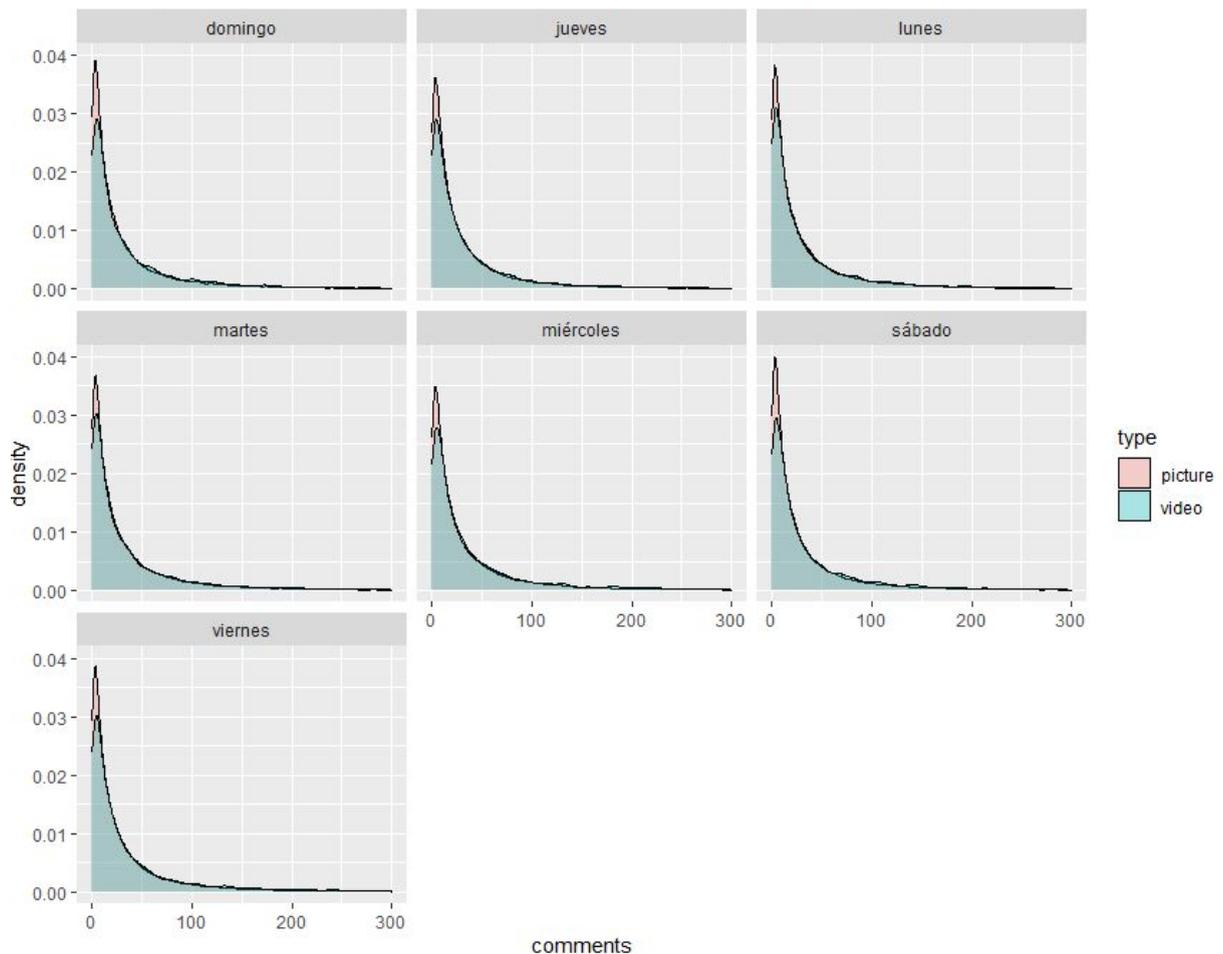
Figura 16: Distribución de comentarios por tipo de publicación (imagen o video)



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

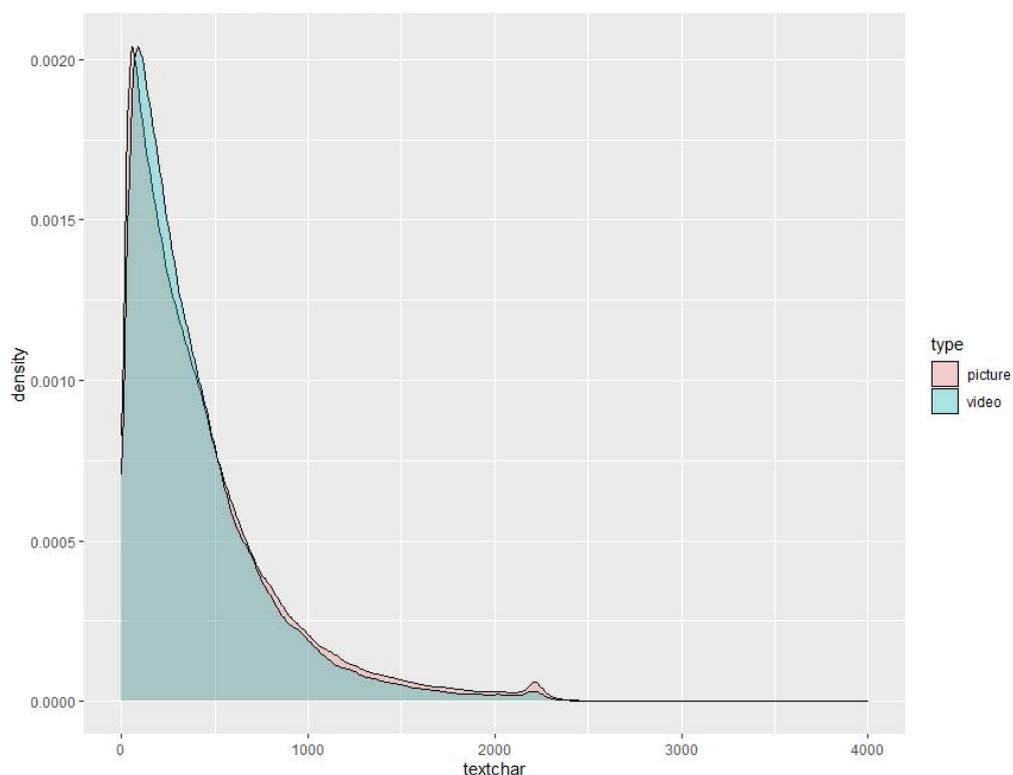


En este gráfico fueron utilizadas las variables 'comments' (comentarios), 'weekday' (día de la semana) y 'type' (tipo de publicación). Los resultados que se pueden apreciar son que para imágenes, la distribución de comentarios se concentra más alrededor de su media en los días viernes, sábado y domingo. 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.

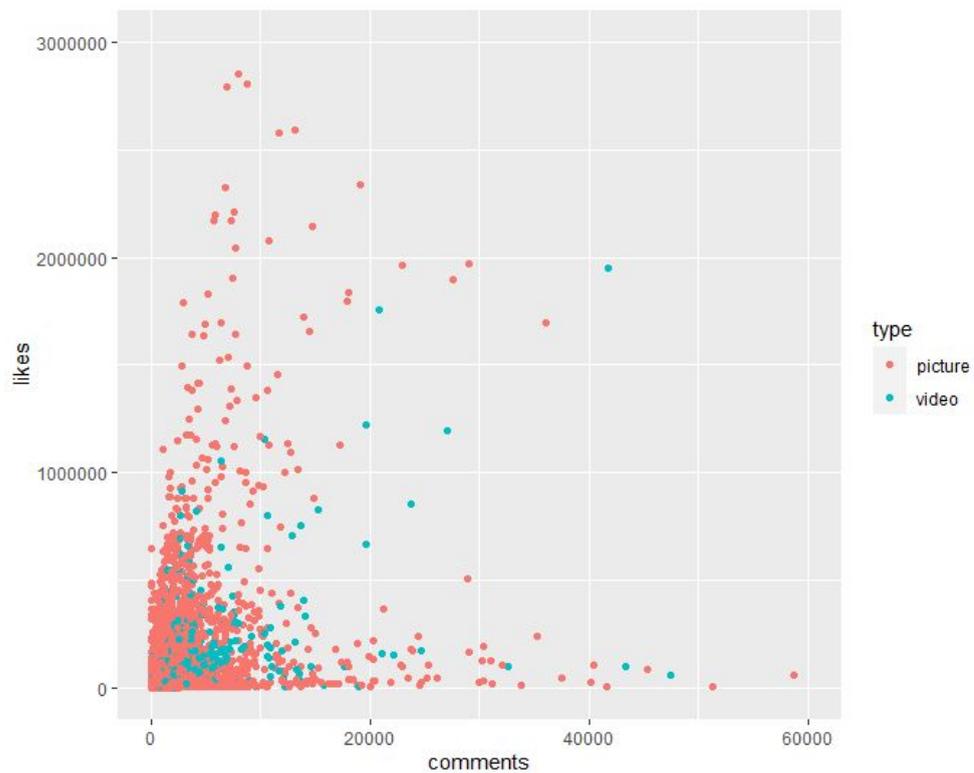
Figura 18: Distribución de la cantidad de caracteres en la descripción de la publicación por tipo de publicación (imagen o video)



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.

3.2.7 Relación entre comentarios y likes

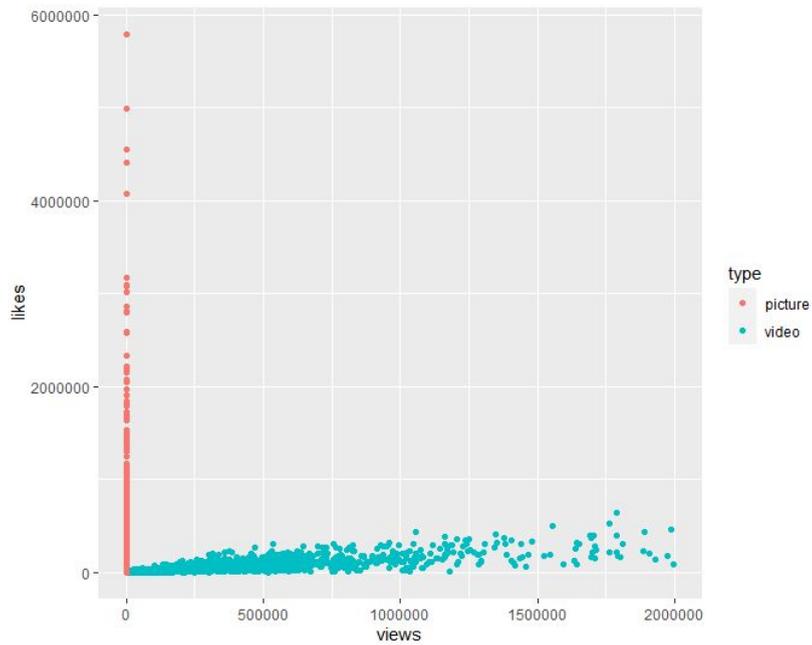
Figura 19: Scatter Plot entre los comentarios y los *likes*



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 creíamos. 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.

3.2.8 Relación entre visitas y likes

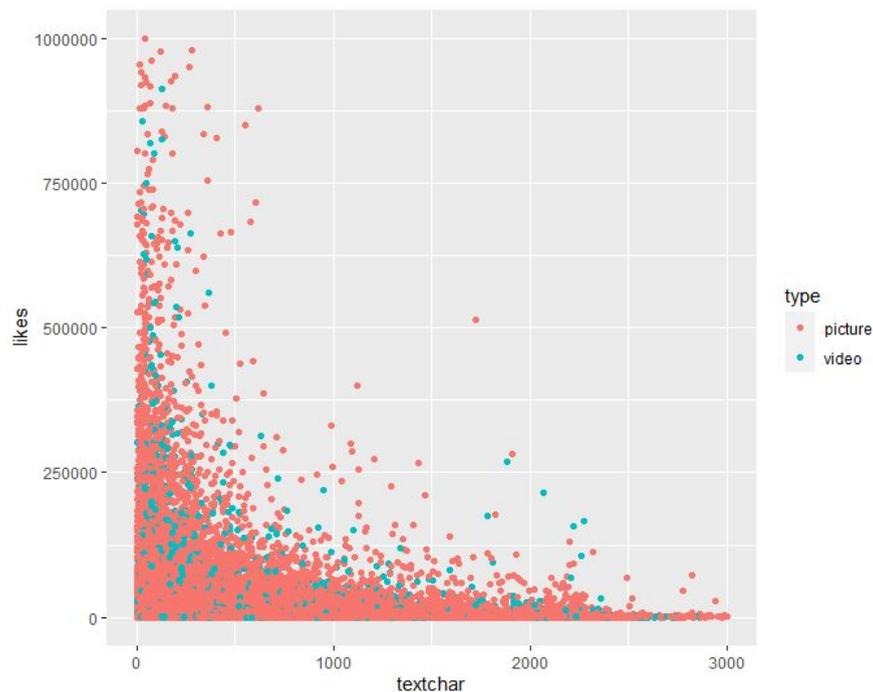
Figura 20: Scatter Plot entre visitas y *likes*



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.

3.2.9 Relación entre cantidad de caracteres en la descripción y likes

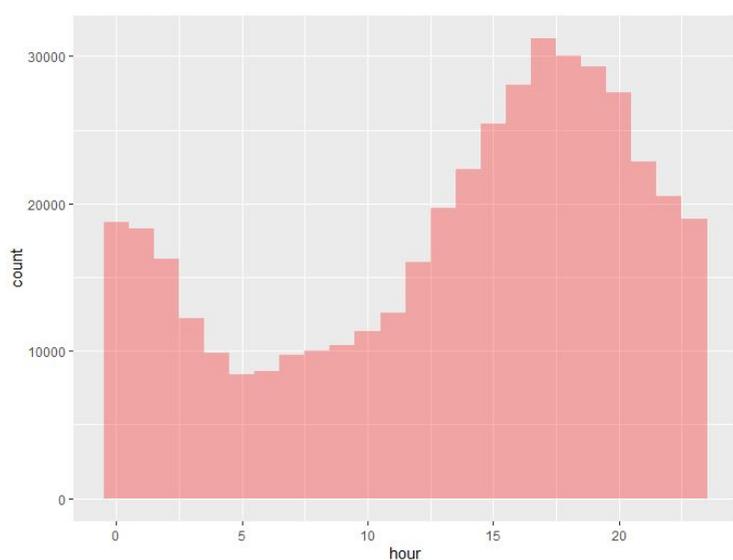
Figura 21: Scatter Plot entre la cantidad de caracteres en la descripción de la publicación y *likes*



Nos pareció interesante averiguar si había alguna relación entre la cantidad de caracteres incluidos en la descripción de la publicación y los *likes* que ésta 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.

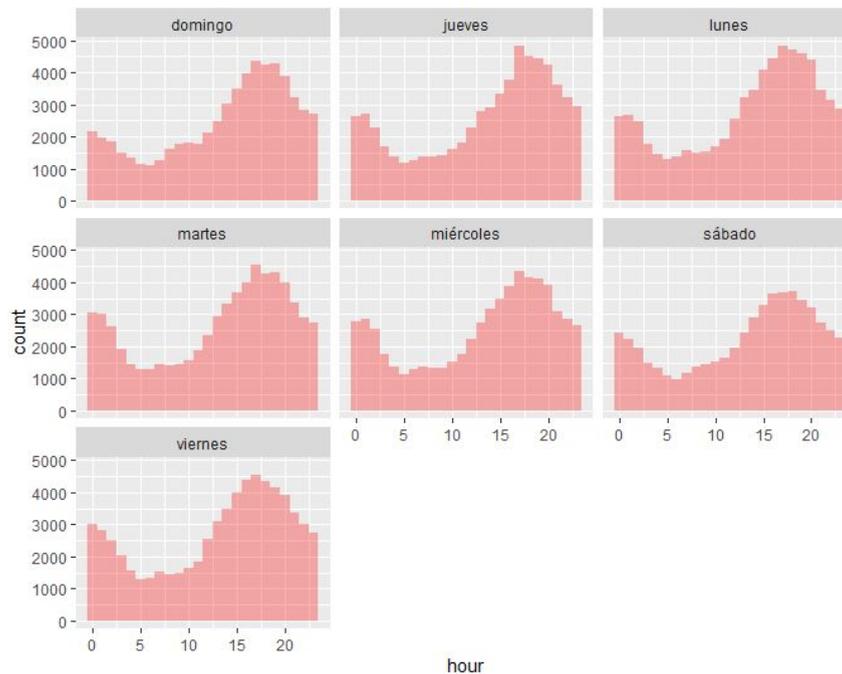
3.2.10 Histograma de cantidad de publicaciones por hora del día

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.

3.2.11 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”

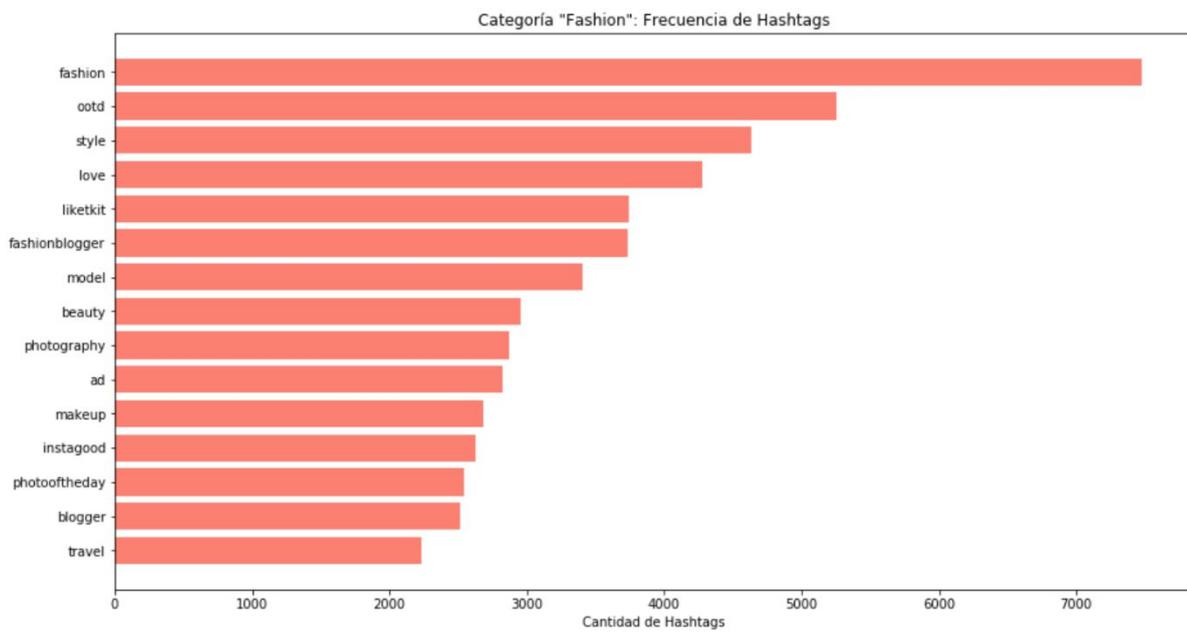
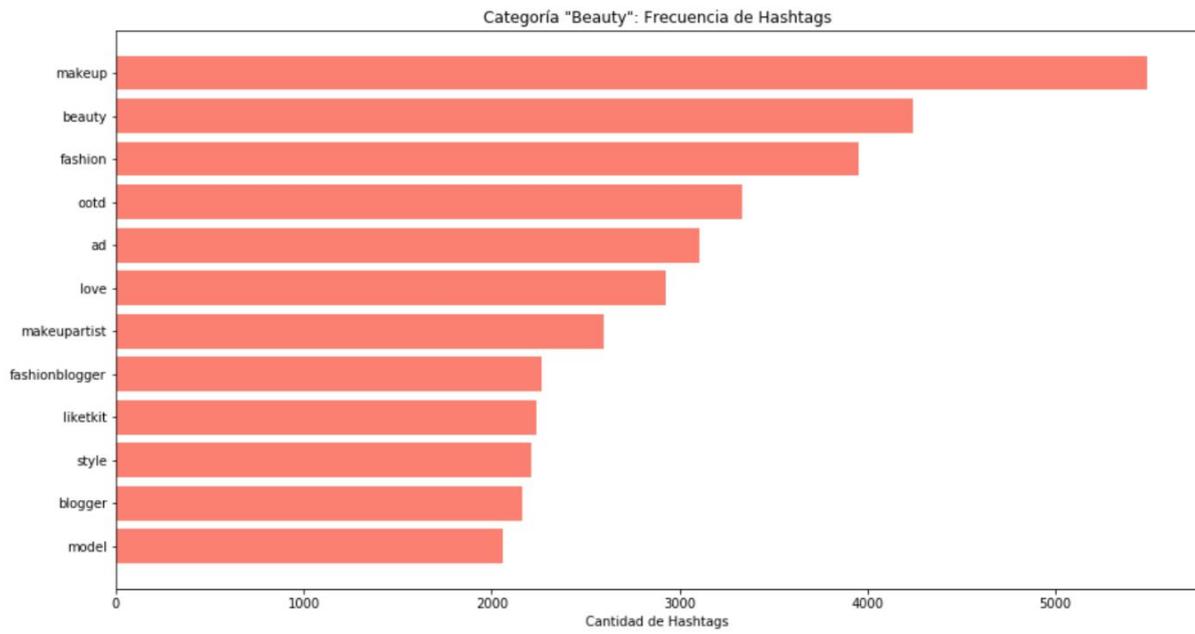
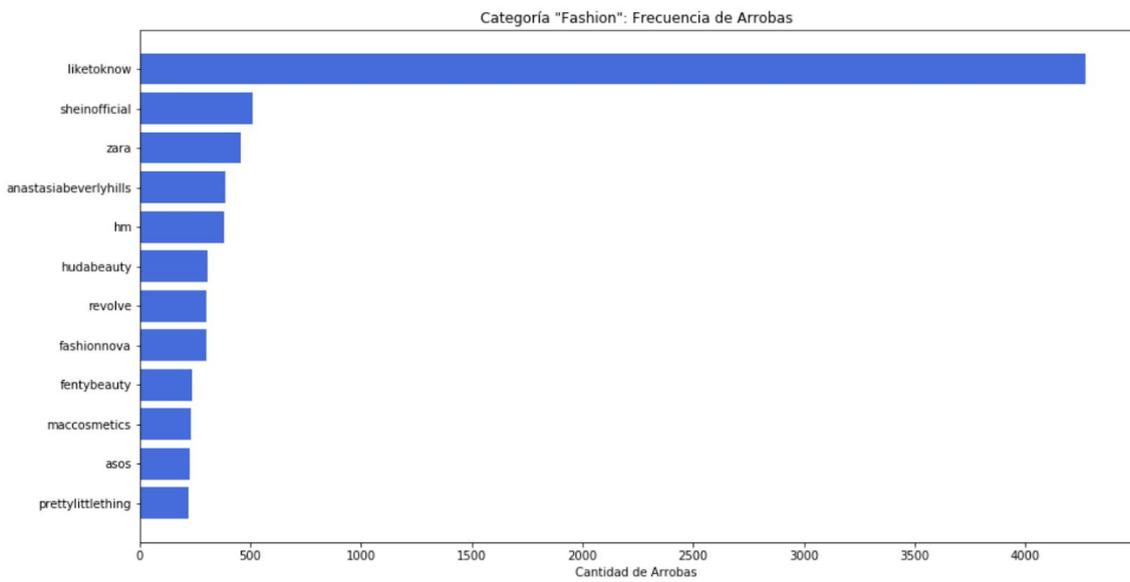
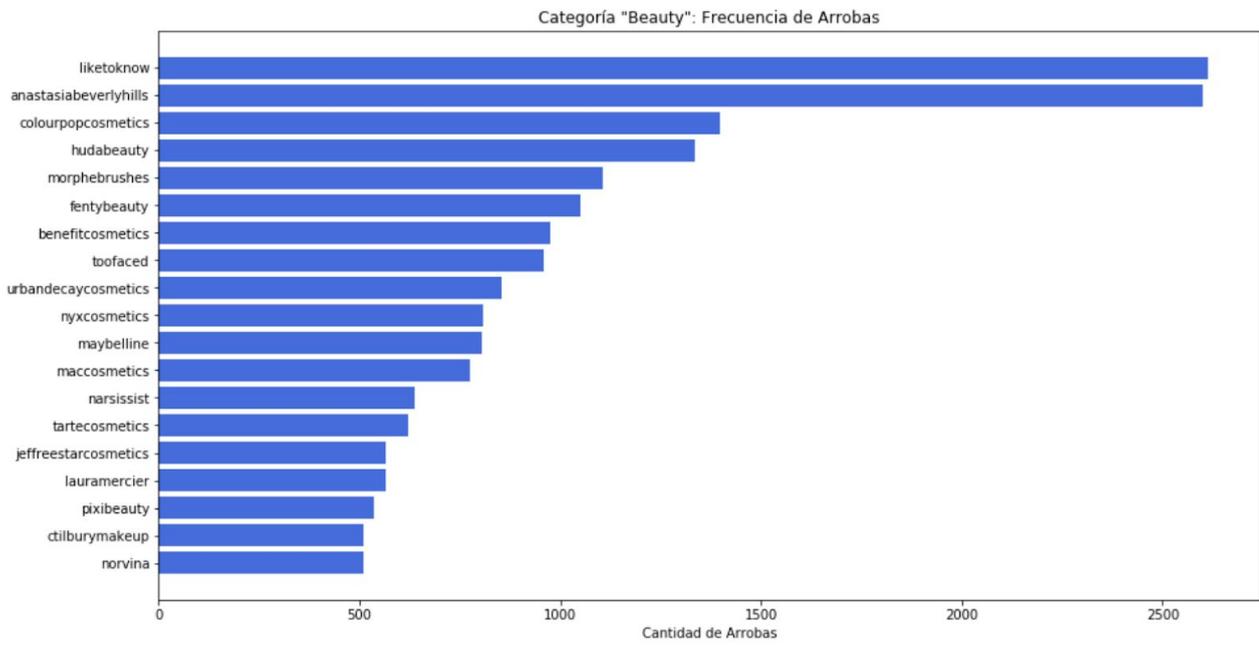


Figura 25: Frecuencia de Arrobas para las categorías "Beauty" y "Fashion"



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 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 ya 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 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.

Este método de clustering fue nombrado y descrito por primera vez en 2013 en el paper “An improved *k-prototypes* clustering algorithm for mixed numeric and categorical data”, por Jinchao Ji et al.

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 ende, 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 y correr el 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

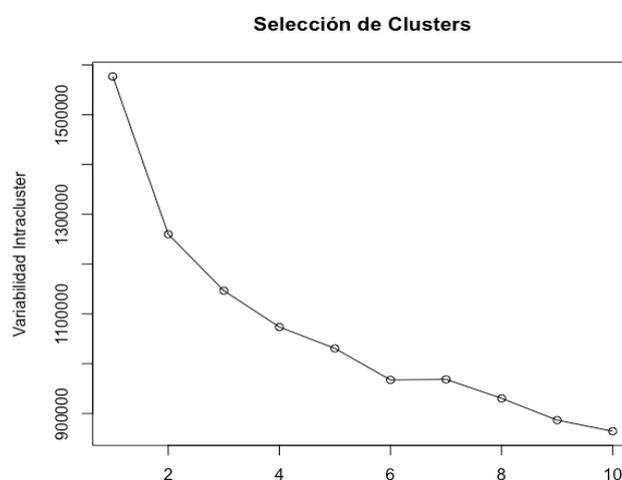
importancia de la variable 'engagement_rate' y 'total_posts' para definir la actividad diaria de los influencers en Instagram, y dado que las observaciones cuyo valor de 'engagement_rate' y 'total_posts' eran iguales a cero también contaban con las variables 'total_comments', 'total_likes' y 'total_engagements' 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%

Enriqueciendo el tema, 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

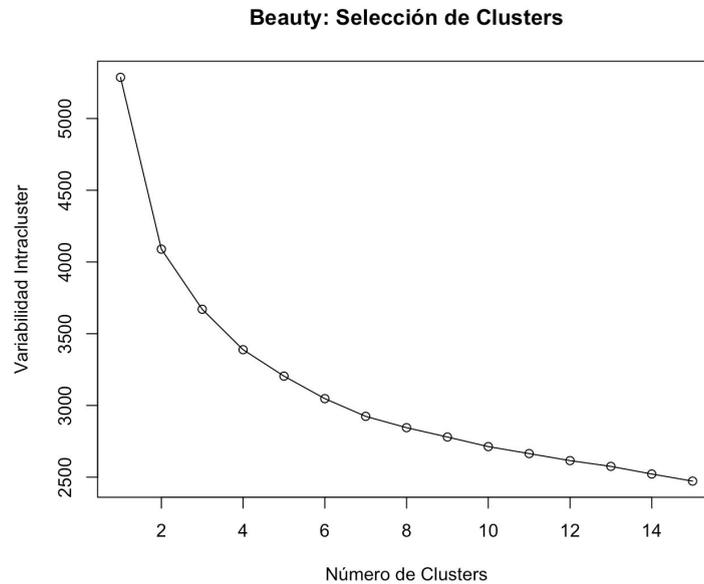
eran en su mayoría mujeres y que el rango de edad más frecuente era el de 25-34 años. Lo mismo ocurre con el idioma, el cual es el mayoritario también ('en' - inglés). La creación de las *Influencer Personas*, es decir, los influencers representativos de cada cluster obtenido, nos sirvió esencialmente 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”)

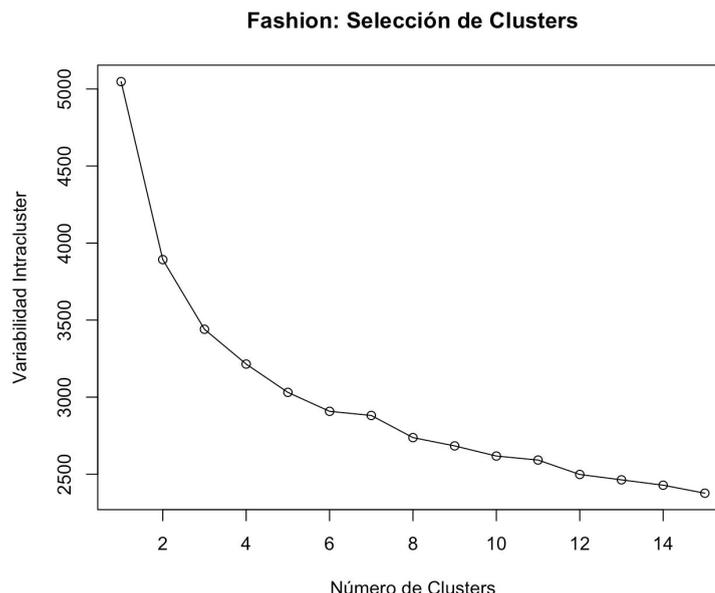
Variables	Cluster			
	1	2	3	4
lang	en	en	en	en
gender	female	female	female	female
age_bracket	25-34	25-34	18-24	25-34
username	12.49	12.77	11.87	13.38
followers	46,176	29,967	410,376	89,131
engagement_rate	0.31%	0.17%	0.80%	0.69%
total_posts	132	69	533	419
total_engagements	89,432	27,302	4,442,597	554,150
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 pareciera que el clustering siempre encuentra 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: Hasta qué punto el fit entre producto e influencer afecta el resultado de las ventas? El sentido común nos sugeriría que cualquier producto podría ser publicitado utilizando cualquier influencer popular (popularidad medida en cantidad de seguidores) y generar buenos resultados empresariales. Sin embargo, lo que queremos investigar es si diferentes productos requieren, o no, influencers con características diferentes para elevar las ventas. Determinados productos podrían requerir que el influencer sea percibido como confiable para que los consumidores elijan comprar lo que publicita, mientras que otros productos podrían requerir que el influencer se vea más profesional. Por este motivo, queremos explorar si la personalidad percibida del influencer debe coincidir con la “personalidad” del producto para elevar la intención de compra y/o propensión de pago del consumidor.

5.1 Popularidad del influencer y relación con su personalidad percibida

En primer lugar, la primer variable independiente elegida fue la popularidad del influencer. Cuantos más seguidores tenga un influencer, más conocido es y más alcance tienen sus publicaciones. Es por esto que las empresas deciden frecuentemente publicitar sus productos con algún influencer con muchos seguidores sin hacer mayor investigación. Sin embargo, un influencer “macro” tendrá una personalidad diferente y su público le exigirá o esperará diferentes actitudes en comparación a un influencer “micro”. Mientras que los macroinfluencers se ven más distantes a uno, más profesionales y con mayor producción, los microinfluencers son percibidos más confiables, más honestos y mayor parecidos a un amigo que a una celebridad. Estos diferentes rasgos de personalidad entre bloggers más populares y menos populares es lo que puede estar generando que unos funcionen bien en combinación con ciertos productos mientras que los otros con productos diferentes.

5.2 Diferenciación entre productos con y sin ajuste al influencer

En segundo lugar, la segunda variable independiente elegida fue el ajuste entre producto e influencer. Es interesante hacer una distinción entre productos que están dentro del dominio o categoría de un influencer en específico y los productos que no lo están. En la práctica, se puede ver que influencers sponsorean productos que “coinciden” con la temática de su perfil de instagram, lo cual tiene sentido ya que combina bien con su contenido normal, pero también productos fuera de su categoría (ejemplo, un blogger de moda sponsoreando productos comestibles y alimenticios). Si bien esta segunda situación genera cierto desconcierto a veces entre los consumidores y se ve como “fuera de lugar”, en ciertos influencers genera un desagrado mayor que en otros. Es por esto que decidimos trabajar con estos dos tipos de productos, los que se ajustan al dominio y habilidad influencer que los sponsorea y los que no, e intentaremos averiguar si los rasgos de personalidad derivados de la popularidad de cada influencer tiene que ver con el éxito o no de las diferentes combinaciones de influencer-producto.

5.3 Categoría de belleza dentro de los “insta-influencers”

En último lugar, elegimos centrarnos en la categoría específica de belleza, la cual incluye bloggers de maquillaje y skincare principalmente. Esto es así ya que es una categoría que surgió hace relativamente poco y está siendo muy popular entre jóvenes y adultos. Además, hay una gran dicotomía entre los bloggers populares de esta categoría y los menos populares. Normalmente, por lo visto en la práctica, los grandes influencers de

belleza tienen sus propias marcas de maquillaje o skincare, tienen canales de youtube y son populares en esa red social al mismo tiempo que en instagram, tienen equipos profesionales (cámara, maquillaje, zona de photoshoots, etc) para mejorar la calidad de su contenido, entre otras, mientras que los influencers chicos son más “amateurs” en lo que hacen y los materiales que poseen, pero son más propensos a interactuar directamente con sus seguidores. Estas diferencias marcadas nos pueden modelar mejor las variables que queremos estudiar y sus relaciones entre sí, de manera de estudiar mejor los efectos que causan en las propensiones de pago e intención de compra.

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 distintos niveles de popularidad y profesionalismo (un influencer con +1.1M de seguidores con experiencia en su área vs. uno -20mil seguidores mucho más amateur) que a su vez vendan un producto que coincida con sus categorías de dominio o uno fuera de esta. Luego, se les hará preguntas a los individuos que participen del experimento para evaluar cómo influyen el nivel de popularidad, profesionalismo y ajuste entre influencer-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 este motivo, creamos en total cuatro perfiles de influencers pertenecientes a la categoría belleza, con dos 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 principio no quisimos basar en el cluster que tuviera mayor cantidad de seguidores y el que tuviera la menor ya que nos interesaba maximizar la diferencia de seguidores entre los dos influencers, para hacer el efecto de la popularidad y profesionalismo mucho más fuerte. El problema que encontramos con esto es que la diferencia de seguidores entre el macroinfluencer y el microinfluencer no era lo suficientemente grande (410.376 seguidores vs 29.966 seguidores), por lo cual decidimos ampliar la diferencia haciendo que el influencer menos popular tenga 16.9mil seguidores y el más popular 1.1M seguidores. Es natural pensar que un influencer pequeño con baja cantidad de seguidores será mucho menos profesional, tendrá menos años de

construcción de su carrera y contará con recursos mucho más “amateurs” que uno con más cantidad de seguidores, mayor cantidad de años en el rubro y material profesional con el cual trabajar.

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 en promedio este cluster menciona una red social adicional en su bio, publica en promedio cada dos días y publica en su mayoría imágenes o fotos (concordante con el hecho de que hablamos de influencers de belleza).

De esta forma, utilizando información obtenida de la etapa de clustering, creamos dos perfiles de belleza que sean macroinfluencers y otros dos que sean microinfluencers. Las diferencias entre los dos perfiles más populares y entre los dos menos populares será el contenido de la bio, diferencia en profesionalismo de las imágenes publicadas, cantidad de seguidores y cantidad de publicaciones. Las diferencias entre los dos perfiles micro (y a su vez entre los dos perfiles macro) será el tono de la biografía (una mucho más “amigable” con emojis y más cálida y otra más seria) y las fotos publicadas (un perfil más variable en cuanto a sus fotos y otro más apegado a subir solo publicaciones relacionadas a la belleza). Además, entre los dos macroinfluencers se agregará la diferencia de que uno tendrá el “tick” que demuestra que es un perfil comprobado por instagram y el otro que no lo tendrá; esto se hará para ver si causa alguna diferencia de cómo se percibe el influencer entre los seguidores.

Luego, a cada perfil están asociadas dos publicaciones de productos diferentes. Los cuatro productos elegidos para testear serán una paleta de maquillaje (alto encaje entre influencer y producto), una suscripción mensual a una aplicación de audiolibros, un producto de tecnología (la consola de videojuegos Nintendo Switch para los microinfluencers y unos auriculares para los macroinfluencers) y una barrita de cereal protéica.

Todos estos perfiles y sus publicaciones asociadas pueden verse en el Apéndice. El objetivo es que a partir del pretest, podamos elegir **un perfil de macroinfluencer y uno de microinfluencer**. 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 el producto sin**

encaje producto-influencer que cause mayor sensación de estar fuera de la categoría de belleza.

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 ser un microinfluencer o un macroinfluencer) con una publicación al azar (publicación que entra dentro de la categoría belleza o uno que no) 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) el profesionalismo percibido del influencer (profesional vs. amateur) y tamaño percibido (micro vs. macro); y 2) ajuste producto-influencer basado en la habilidad (ajuste producto-influencer alto vs. ajuste producto-influencer bajo). 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.

El tamaño y profesionalismo de un influencer fue manipulado al cambiar el número de followers y el profesionalismo de las imágenes publicadas en la parte visible del perfil (mayormente relacionado con la habilidad del influencer, en este caso de maquillaje). El ajuste entre producto-influencer basado en la habilidad fue manipulado al hacer que un influencer de maquillaje promocióne un producto relacionado al maquillaje o uno no relacionado al maquillaje (comida, tecnología). En general, a los participantes se les mostraron al azar cuatro perfiles de influencers de maquillaje: dos microinfluencers y dos macroinfluencers. Tras ver el perfil de cada uno, a los participantes se les mostraron dos de las publicaciones promocionando un producto de estos influencers. Los participantes vieron una publicación sobre un producto de maquillaje y una publicación sobre otro producto (comida o tecnología).

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 era profesional, amateur y si tenía muchos seguidores en una escala de Likert de 7-puntos (1 = extremadamente improbable, 7 = extremadamente probable). 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 le mostró dos de las publicaciones de estos influencers. Para medir el ajuste producto-influencer basado en la habilidad, se les preguntó a los participantes que expresaran si el producto publicitado caía dentro del dominio de la persona que lo publicita, y si la persona publicitando el producto tiene la credibilidad para hacerlo (1= para nada, 7= extremadamente).

5.6 Resultados del Pretest

5.6.1 Tamaño percibido del influencer

Para determinar cuál perfil resultaba más efectivo para representar la imagen de un macro influencer, comparamos las evaluaciones de los participantes en el ítem “Esta persona tiene muchos seguidores” en todos los perfiles mostrados. Ejecutamos un test ANOVA de medidas repetidas unidireccionales para comparar el efecto del tipo de influencer en el tamaño percibido del influencer. Como era de esperar, los resultados mostraron que el tipo de influencer tenía un efecto significativo sobre si el influencer se percibió como “macro” ($F(7,133) = 37.87, p < .001, \eta^2 = .51$). Comparaciones post hoc se analizaron usando t-tests dependientes con una corrección de Bonferroni⁵. Los dos perfiles macro resultaron tener puntuaciones significativamente mayores que los perfiles micro (tabla 7). La Figura 29 muestra las puntuaciones promedio junto a otros intervalos de confianza.

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

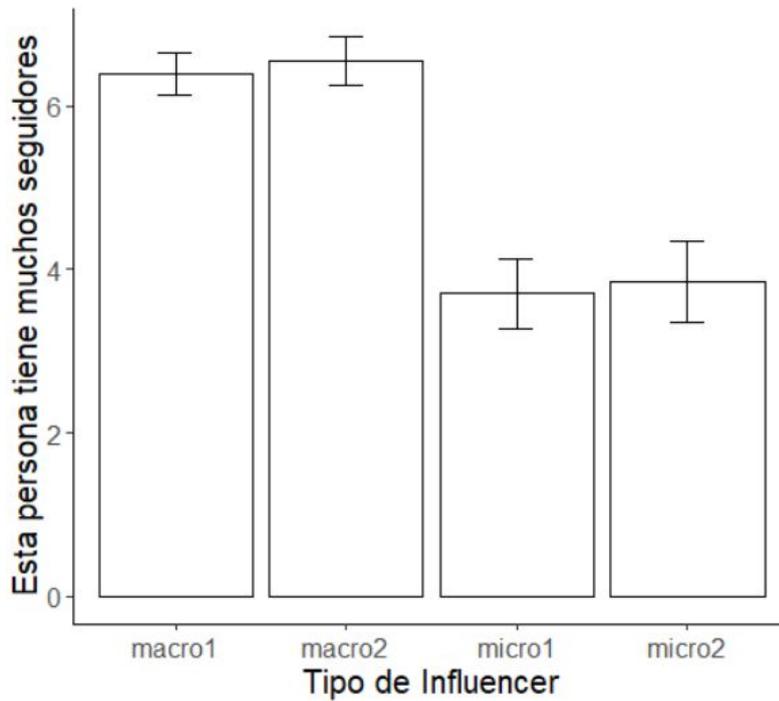
	M	SD
Micro 1 ^d	3.70	1.32

⁵ 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 α/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.

Micro 2 ^d	3.85	1.54
Macro 1 ^e	6.40	0.81
Macro 2 ^e	6.55	0.93

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

Figura 29: Esta persona tiene muchos seguidores



5.6.2 Profesionalismo percibido del influencer

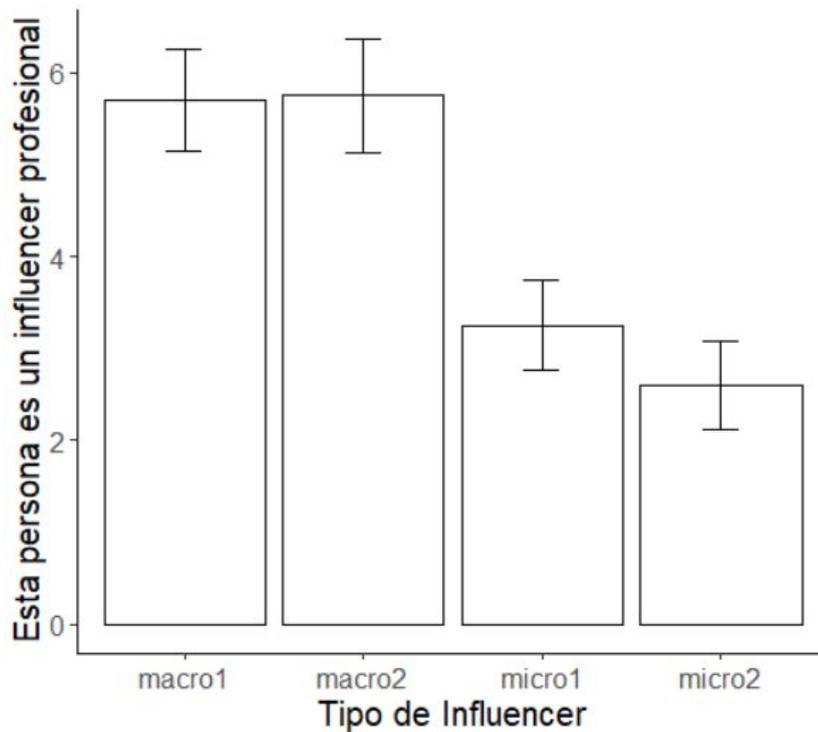
Para determinar cuál perfil era más efectivo al retratar la imagen de un influencer profesional, comparamos las puntuaciones de los participantes en el ítem “Esta persona es un influencer profesional” en todos los perfiles mostrados. Corrimos un test ANOVA de medidas repetidas unidireccionales para comparar los efectos del tipo de influencer en el profesionalismo percibido del influencer. Como era de esperar, los resultados mostraron que el tipo de influencer tiene un efecto significativo sobre si el influencer era percibido como profesional ($F(7,133) = 30.22, p < .001, \eta^2 = .49$). Las comparaciones post hoc se analizaron utilizando t-tests dependientes con una corrección de Bonferroni. Se encontró que los dos perfiles macro tenían puntajes significativamente mayores que los perfiles micro (tabla 8). La figura 30 muestra los puntajes promedio junto a sus intervalos de confianza.

Tabla 8: Promedios (M) y Desvíos Estándar (SD)

	M	SD
Micro 1 ^c	3.25	1.53
Micro 2 ^c	2.60	1.48
Macro 1 ^d	5.70	1.76
Macro 2 ^d	5.75	1.92

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

Figura 30: Esta persona es un influencer profesional

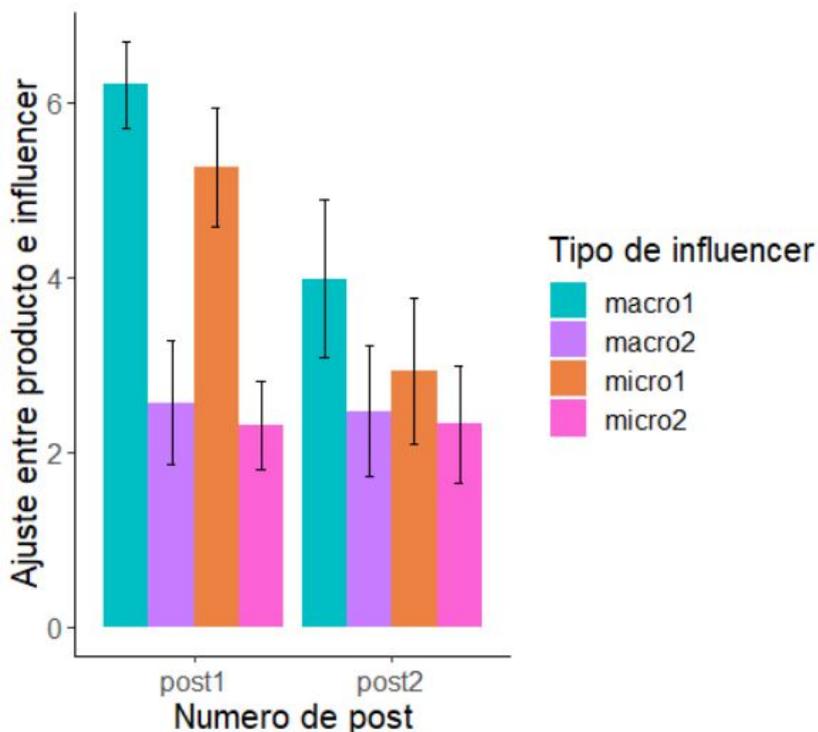


5.6.3 Ajuste producto-influencer basado en la habilidad

Primero, creamos una medida compuesta al promediar los puntajes de los tres ítems que median el ajuste producto-influencer basado en la habilidad ($\alpha = 0.9$). Ejecutamos un test ANOVA de medidas repetidas bidireccional para comparar el efecto del tipo de influencer y la tipología del post en el ajuste producto-influencer percibido. Los resultados mostraron un efecto principal significativo del tipo de influencer ($F(7,133) = 31.76, p < .001, \eta^2 = .39$), un efecto significativo central de la tipología de la publicación ($F(1, 19) = 43.46, p < .001, \eta^2 = .06$) y una interacción significativa de las dos ($F(7,133) = 12.49, p < .001, \eta^2 = .08$) sobre el

ajuste percibido entre producto e influencer. Comparaciones post hoc fueron analizadas utilizando t-tests dependientes con una corrección de Bonferroni. Los resultados revelaron que nuestra manipulación fue efectiva para el micro influencer 1 (Mfitpost = 5.25, SD = 1.45 vs. Mnofitpost = 2.93 SD = 1.78, $p < .001$) y para el macro influencer 2 (Mfitpost = 6.20, SD = 1.51 vs. Mnofitpost = 3.98 SD = 1.59, $p < .001$). La figura 31 muestra los puntajes promedio en conjunto con sus intervalos de confianza.

Figura 31: Promedio del ajuste percibido entre producto e influencer



5.7 Experimento

En base a los resultados del pretest, podemos concluir que los perfiles que maximizan las diferencias entre un influencer popular y profesional en su área y uno menos popular y menos profesional son macro1 (Laura Blakonian @LauraBlushh) y micro1 (Sofi Melandi @WorldOfSofi). A su vez, los productos elegidos para maximizar la diferencia entre un producto con ajuste producto-influencer y uno sin ajuste son la paleta de sombras de ojos (maquillaje) y el audiolibro (app a base de suscripción mensual para escuchar audiolibros). Dado esto, estas serán las combinaciones de productos-influencers que se utilizarán en el siguiente experimento.

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 popularidades, y cada uno de ellos cuenta con dos publicaciones, una de un producto con encaje respecto a la categoría y el otro sin.

5.8 Diseño del experimento

El objetivo principal de este experimento final es poder evaluar el efecto del encaje producto-influencer en conjunto con la popularidad del mismo sobre la disposición a comprar, pagar, recomendar y re-publicar de los consumidores. Para hacer esto, le mostraremos aleatoriamente a los participantes del experimento uno de los dos perfiles previamente seleccionados y contruidos a partir de los resultados del pretest. Por lo cual, cada participante verá el perfil del influencer más popular o del influencer menos popular, junto con alguna de las dos publicaciones posibles (paleta de maquillaje vs. suscripción mensual a una aplicación de audiolibros).

Luego, para obtener información sobre la reacción de los participantes al ver los perfiles y sus publicaciones, 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 re-publicar 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) para el caso de la paleta de maquillaje, y cuánto está dispuesto a pagar por una suscripción mensual en pesos argentinos (ARS) para el caso de la aplicación de los audiolibros.

A su vez, haremos otra serie de preguntas orientadas a entender la personalidad del consumidor. Esta parte del cuestionario está orientado a entender la actitud respecto a la consistencia del participante (si este se siente una preferencia por la consistencia en sí mismo y en otros o si es neutral ante esta). Esta escala se denomina preferencia por la consistencia y está relacionada a la escala de necesidad de consistencia.

Por último, agregaremos preguntas demográficas para saber la edad de los participantes, su género, su uso de la plataforma Instagram y si sigue influencers dentro de la categoría de belleza.

Es importante aclarar que el universo de interés para este experimento son mujeres. La selección de este universo se debe al hecho de que 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”, y que en la práctica sus seguidoras son en mayoría mujeres.

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

5.9 Preprocesamiento de los datos del experimento

El experimento fue llevado a cabo en 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 9: Distribución de los participantes del experimento por género

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

Del total de 229 observaciones, solo 9 fueron hombres y 1 fue otro. 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 que habían repuesto que no utilizaban instagram (responder “No” ante la pregunta número # - ¿Tienes una cuenta de instagram?) o que lo utilizaban poco (responder “Casi Nunca” ante la pregunta número # - ¿Qué tan seguido utilizas tu cuenta de instagram?). A continuación, puede observarse cómo se distribuyen los individuos según su uso de Instagram.

Tabla 10: Respuestas a la pregunta “¿Tienes una cuenta de instagram?”

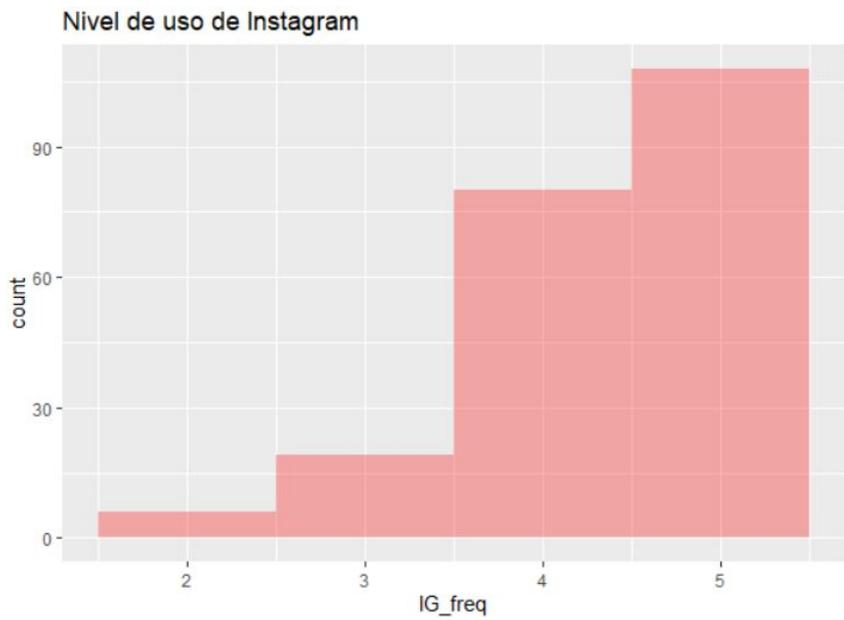
Sí (Tengo cuenta de instagram)	No (No tengo cuenta de instagram)
213	7

Del total de las observaciones luego de filtrar por género, un total de 213 individuos declararon sí ser propietarios de una cuenta de 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?”



Respuestas ante la pregunta ¿Qué tan seguido utilizas tu cuenta de instagram?, siendo 1 - “Nunca” y 5 - “Demasiado”

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.

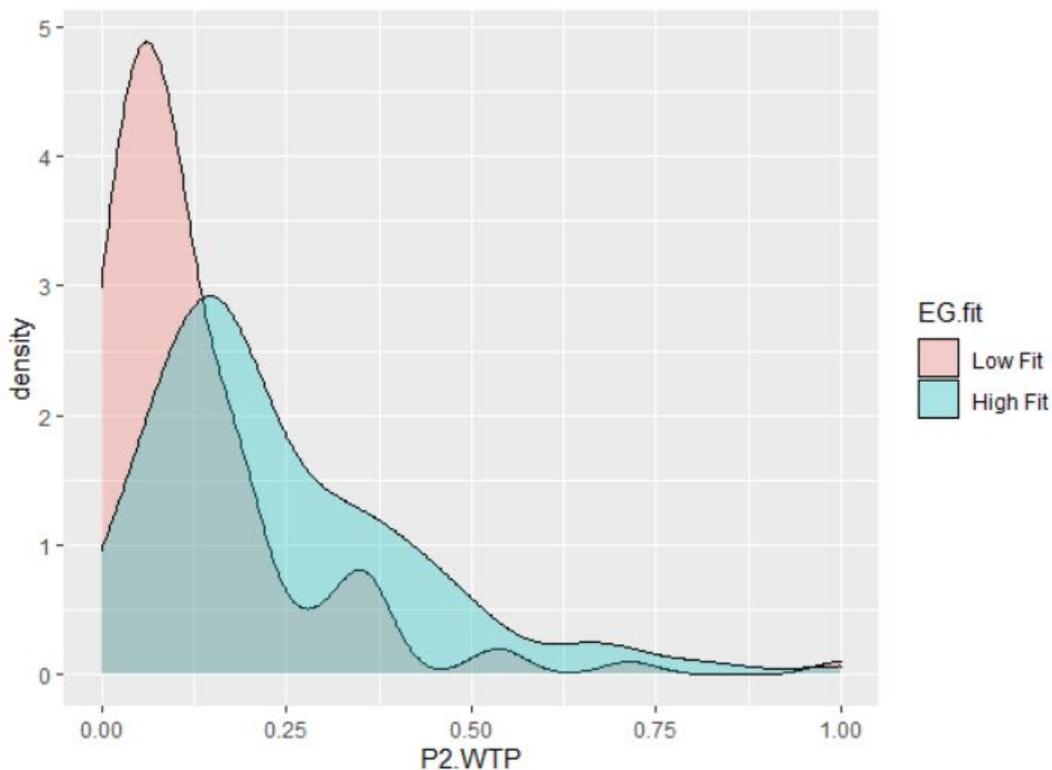
5.9.4 Disposición a pagar en una misma escala

Como trabajamos con dos productos diferentes (una paleta de sombras de ojos y una suscripción mensual a una aplicación de audiolibros), las propensiones a pago estarán en rangos diferentes debido a la naturaleza diferente de los productos. Es por esto que, antes de analizar esta variable, debemos re-escalarla para que puedan ser comparadas las propensiones de pago de un producto con el otro. De todas maneras, se deberá utilizar una forma de re-scaling que no modifique la media de las distribuciones para ambos productos ya que esto podría perjudicar nuestros análisis posteriores de las medias. Es por esto que decidimos utilizar min-max scaling en las propensiones de pago de un producto y del otro, para lograr que se encuentren en la misma escala.

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

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



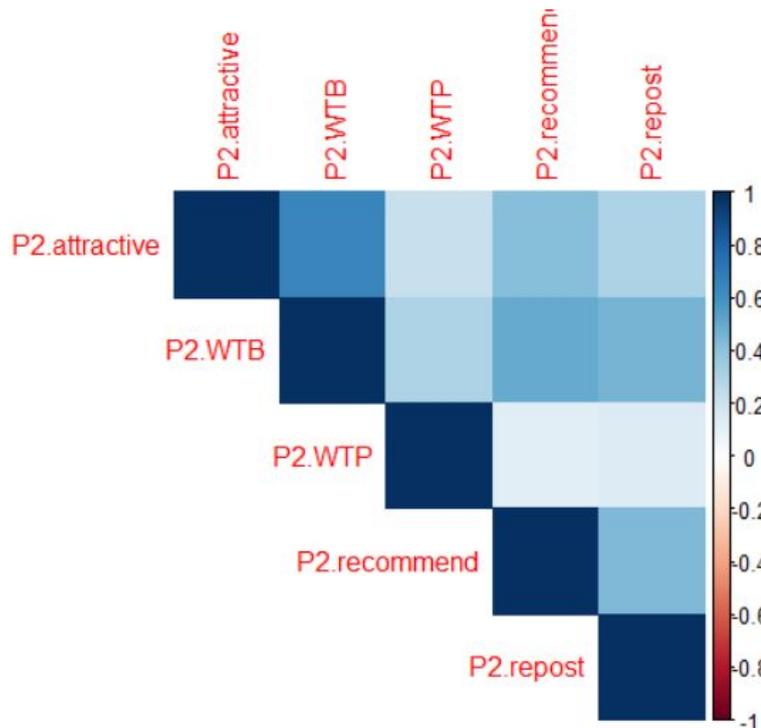
Siendo Low Fit el producto sin ajuste producto-influencer (la suscripción a la app de audiolibros) y High Fit el producto con ajuste (la paleta de maquillaje).

5.9.5 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 re-publicar. A continuación, nos pareció interesante observar las correlaciones entre las respuestas a estas preguntas.

Figura 34: Matriz de Correlación entre las respuestas referidas a la publicación



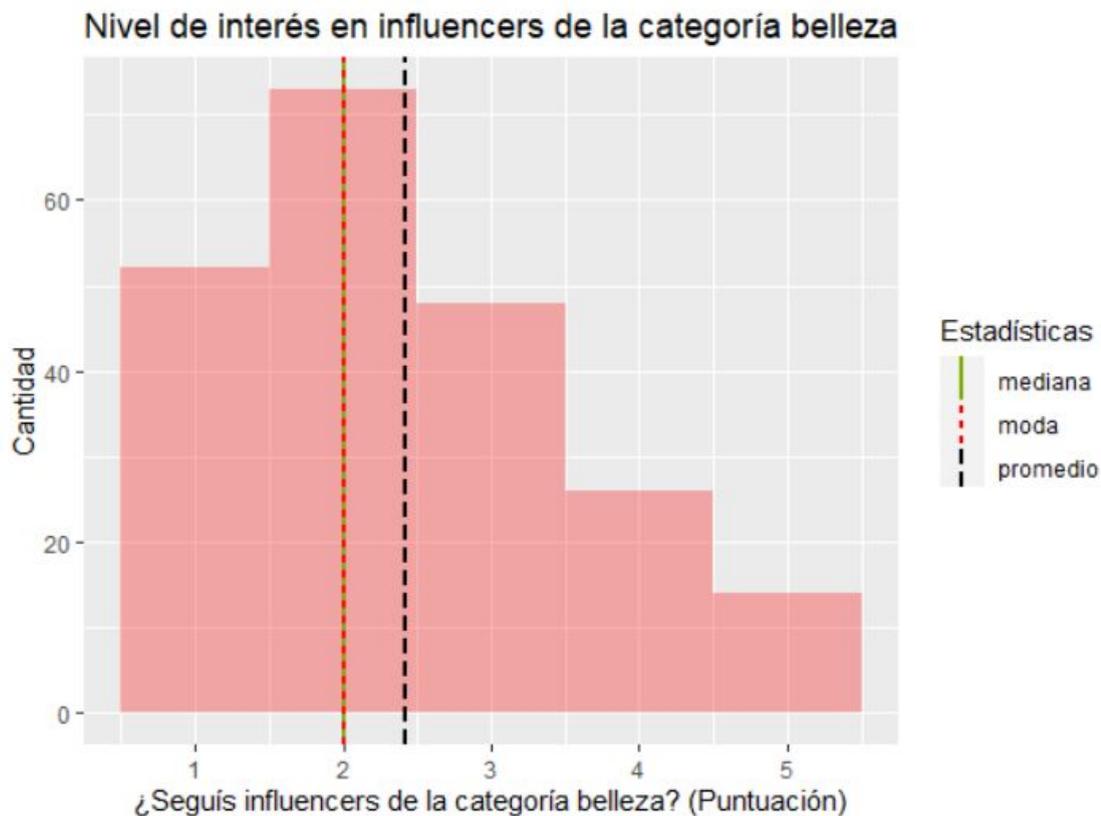
Las etiquetas que aparecen en la matriz hacen referencia a las preguntas relacionadas a la publicación. La etiqueta “P2.attractive” hace referencia a la respuesta a la pregunta “¿Cuán atractivo te parece el producto que se muestra en la publicación?” (1= nada atractivo, 7= muy atractivo), “P2.WTB” hace referencia a la disposición a comprar (*Willingness To Buy*), “P2.WTP” refiere a la disposición a pagar (*Willingness To Pay*), “P2.recommend” refiere a la disposición a recomendar y “P2.repost” 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 re-publicar, mientras que la disposición a comprar también correlaciona con la disposición a recomendar y re-publicar.

5.9.6 División según categoría de influencers seguidos

Nos interesaba saber la distribución del interés por influencers de belleza en las observaciones recolectadas. Si vemos el siguiente histograma, podemos apreciar que dos de los tres estadísticos principales (moda y mediana) caen dentro del nivel 2 - Poco, mientras que la media cae más cercano a la mitad del rango (3 - Algo).

Figura 35: Histograma sobre nivel de interés por influencers de belleza



5.9.7 Manipulación de las preguntas correspondientes al nivel de preferencia por consistencia

Contábamos con 18 variables correspondientes a las preguntas para crear el índice de preferencia por consistencia. Cada pregunta se podía responder con una escala de 1 - Totalmente en desacuerdo a 7 - Totalmente de acuerdo. Para todas las preguntas menos una (la número 18), responder 1 - Totalmente en desacuerdo era equivalente a tener aversión por la consistencia y responder 7 - Totalmente de acuerdo significaba tener preferencia por la consistencia. En el caso de la pregunta número 18, era inmediatamente al revés, por lo que se transformaron los datos de esa variable invirtiéndolos ($8 - X_{var}$) para que sean consistentes con los de las demás preguntas.

Para evaluar la fiabilidad de la escala que estamos construyendo sobre preferencia por la consistencia, se calculó el Alfa de Cronbach⁶ sobre las 18 variables anteriormente nombradas.

Figura 36: Test de fiabilidad calculando el Alfa de Cronbach sobre las preguntas de preferencia por la consistencia

```
Reliability analysis
Call: psych::alpha(x = mainstudy$PFC.items, check.keys = TRUE)

raw_alpha std.alpha G6(smc) average_r S/N ase mean sd median_r
0.88      0.88      0.9      0.28 7.1 0.012 4.3 0.96 0.29

Lower alpha upper      95% confidence boundaries
0.85 0.88 0.9
```

Al ser este coeficiente 0.88 (mayor a 0.8), se puede afirmar que los items son consistentes entre sí y están correlacionados entre sí, por lo cual podemos crear una escala confiable.

Tras hacer esto, se tomó el promedio de las respuestas a las 18 preguntas para crear la variable índice de preferencia por consistencia (PFC).

5.10 Resultados del Experimento

5.10.1 Análisis de la propensión de pago

En primer lugar, analizaremos cómo influyen las características percibidas del influencer en la propensión de pago, sin controlar y controlando por la preferencia por la consistencia del consumidor.

⁶ 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 demuestre 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”.

Sin controlar por el índice de preferencia por la consistencia

Se llevó a cabo la prueba de Levene⁷ para la homogeneidad de las varianzas a través de los 4 grupos de observaciones (Micro Influencer - Low Fit, Micro Influencer - High Fit, Macro Influencer - Low Fit, Macro Influencer - High Fit).

Tabla 11- Test de Levene para la Homogeneidad de la Varianza (centro = mediana)

Df	Valor estadístico F	P-value (>F)
3	1.8543	0.1385

Códigos de significancia: 0 '****' 0.001 '***' 0.01 '**' 0.05 '.' 0.1 ' ' 1

Al ser el p-valor mayor a 0.05, no se puede rechazar la hipótesis nula del test de que las varianzas poblacionales son iguales (es decir, hay homogeneidad de varianza entre algunos grupos). Por esto es que luego al realizar el test ANOVA se deberá implementar la corrección con la prueba de White.

Tabla 12- Test ANOVA para la diferencia de medias entre grupos de la propensión de pago

	Df	F	P-test (>F)
(Intercept)	1	34.1649	0.00000001936***
EG.size	1	0.3954	0.5301579
EG.fit	1	11.9562	0.0006604***
EG.size:EG.fit	1	0.0186	0.8916527
Residuos	208		

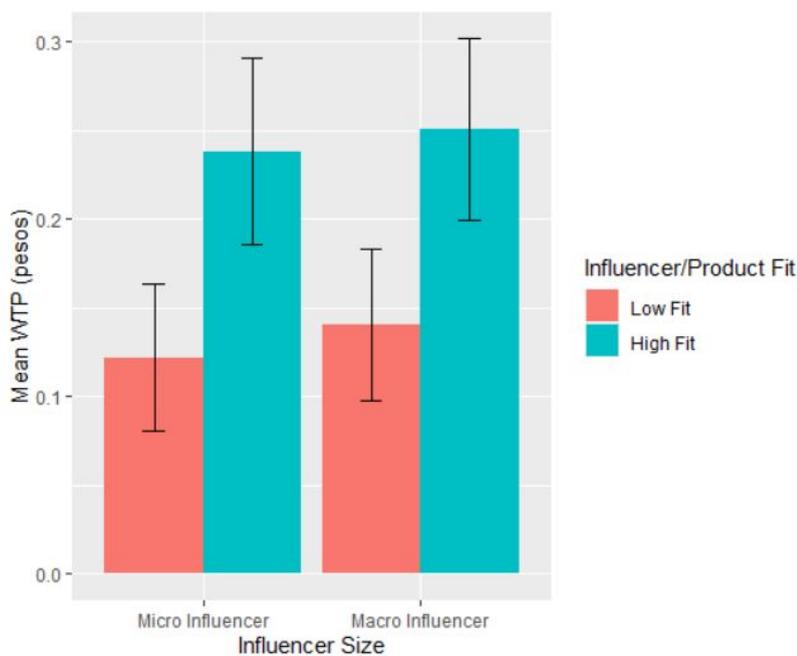
Códigos de significancia: 0 '****' 0.001 '***' 0.01 '**' 0.05 '.' 0.1 ' ' 1

Las etiquetas de las variables representan "EG.size" al tamaño del influencer (Microinfluencer o Macroinfluencer), "EG.fit" al encaje entre el producto y el influencer (Low Fit o High Fit) y EG.size:EG.fit a la interacción entre ambas variables.

⁷ La prueba de Levene es un test estadístico que se utiliza para evaluar si la varianza de dos o más grupos son iguales. La prueba toma la hipótesis nula de que las varianzas poblacionales de los grupos son iguales (homogeneidad de la varianza). Fue ideada por Howard Levene en 1960, apareciendo por primera vez en "Contributions to probability and statistics. Essays in honor of Harold Hotelling" bajo el título "Robust Tests for Equality of Variances".

Se puede ver que el efecto principal, significativo al 99% de confianza, que genera una diferencia en la propensión de pago es el encaje entre el influencer y el producto. Además, se calculó el poder estadístico (1 - probabilidad de cometer un error tipo II) de esta prueba utilizando una significancia del 1%, la cantidad de observaciones y las medias por grupo y dio como resultado 0,9011, es decir que hay un 90,11% de chances que la hipótesis nula sea rechazada si la hipótesis alternativa es verdadera por lo tanto la prueba tiene mucha potencia. Generamos el gráfico para observar la diferencia de medias separados por el encaje entre producto influencer.

Figura 37: Diferencia de medias para la propensión de pago



Intervalos de confianza al 95%

Podemos observar que, tanto para microinfluencers como para macroinfluencers, la propensión de pago del consumidor se eleva si el producto está relacionado con la categoría del influencer. Para un producto con alto ajuste con la categoría del influencer, no hay una diferencia significativa en la propensión de pago entre los micro y macro influencers.

Controlando por el índice de preferencia por la consistencia

Se realizó un test de ANOVA sobre la propensión de pago controlando por el índice de preferencia por la consistencia.

Tabla 13: Test ANOVA para la diferencia de medias entre grupos de la propensión de pago

	Df	F	P-value (>F)
(Intercept)	1	0.1029	0.7487
PFC	1	1.2152	0.2716
EG.size	1	0.8899	0.3466
EG.fit	1	2.6126	0.1076
PFC:EG.size	1	0.9547	0.3297
PFC:EG.fit	1	0.3021	0.5831
EG.size:EG.fit	1	1.4791	0.2253
PFC:EG.size:EG.fit	1	1.5212	0.2189
Residuals	204		

Se puede ver que no hay ninguna variable significativa que influya en la propensión a pagar de los consumidores.

5.10.2 Análisis de la propensión a comprar

En segundo lugar, analizaremos cómo influyen las características percibidas del influencer en la propensión de compra, sin controlar y controlando por la preferencia por la consistencia del consumidor.

Sin controlar por el índice de preferencia por la consistencia

Se llevó a cabo la prueba de Levene para la homogeneidad de las varianzas a través de los 4 grupos de observaciones (Micro Influencer - Low Fit, Micro Influencer - High Fit, Macro Influencer - Low Fit, Macro Influencer - High Fit).

Tabla 14: Test de Levene para la Homogeneidad de la Varianza (centro = mediana)

Df	Valor estadístico F	P-value (>F)
3	2.2292	0.08586 .

Códigos de significancia: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Al ser el p-valor mayor a 0.05 pero menor a 0.1, se puede rechazar la hipótesis nula del test de que las varianzas poblacionales son iguales con una confianza del 90% (es decir, no hay homogeneidad de varianza entre algunos grupos).

Tabla 15: Test ANOVA para la diferencia de medias entre grupos de la propensión de compra

	Df	F	P-test (>F)
(Intercept)	1	123.1734	< 0.0000000000000000 22 ***
EG.size	1	8.5886	0.003759 **
EG.fit	1	5.1704	0.023991 *
EG.size:EG.fit	1	2.7554	0.098424 .
Residuos	209		

Códigos de significancia: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

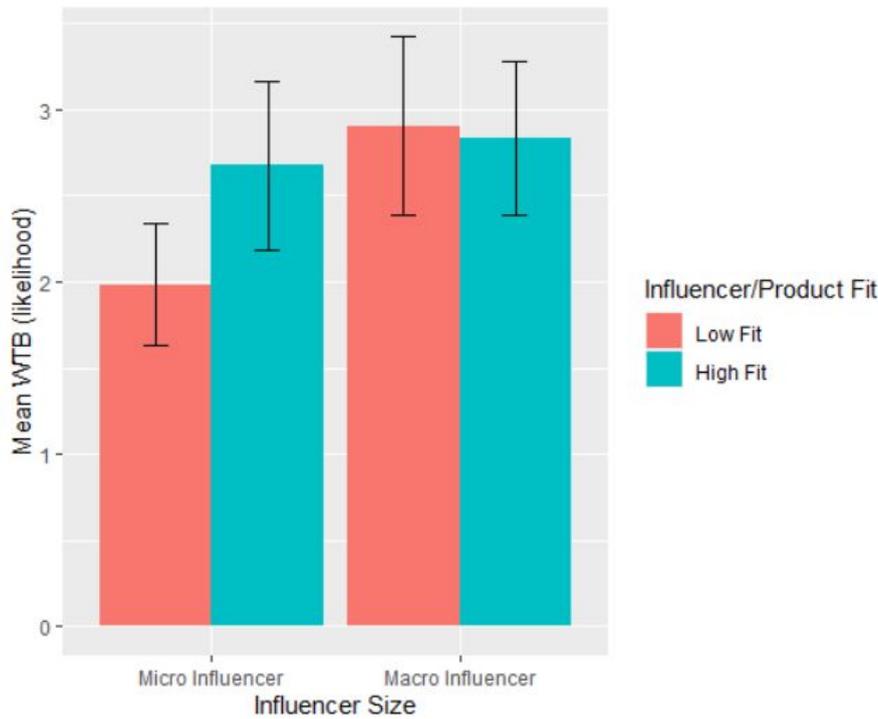
El tamaño del influencer es una característica que genera una diferencia en la media de propensión de compra significativa al 99%. Además, se calculó el poder estadístico (1 - probabilidad de cometer un error tipo II) de esta prueba y dio como resultado 0,7593, es decir que hay un 75,93%⁸ de chances que la hipótesis nula sea rechazada si la hipótesis alternativa es verdadera. Por otro lado, el encaje entre el producto y el influencer generan una diferencia en la media significativa al 95%, con una potencia del 41.51%⁹. Finalmente, la interacción entre el tamaño del influencer y el encaje entre producto e influencer genera una diferencia en la media significativa al 0,90%, es decir es marginalmente significativa, con una potencia del 86,88%.

Generamos el gráfico para observar la diferencia de medias separados por el encaje entre producto influencer.

⁸ Se requerirían 260 observaciones para llegar a una potencia del 80%

⁹ Se requerirían 644 observaciones para llegar a una potencia del 80%

Figura 38: Diferencia de medias para la propensión de compra



Intervalos de confianza al 95%

Este gráfico muestra que cuando el influencer es micro, el encaje entre producto e influencer es un determinante significativo para la propensión de compra, mientras que si es un macroinfluencer la diferencia en la media de propensión de compra no es significativa.

Controlando por el índice de preferencia por la consistencia

Se realizó un test de ANOVA sobre la propensión de pago controlando por el índice de preferencia por la consistencia.

Tabla 16: Test ANOVA para la diferencia de medias entre grupos de la propensión de compra

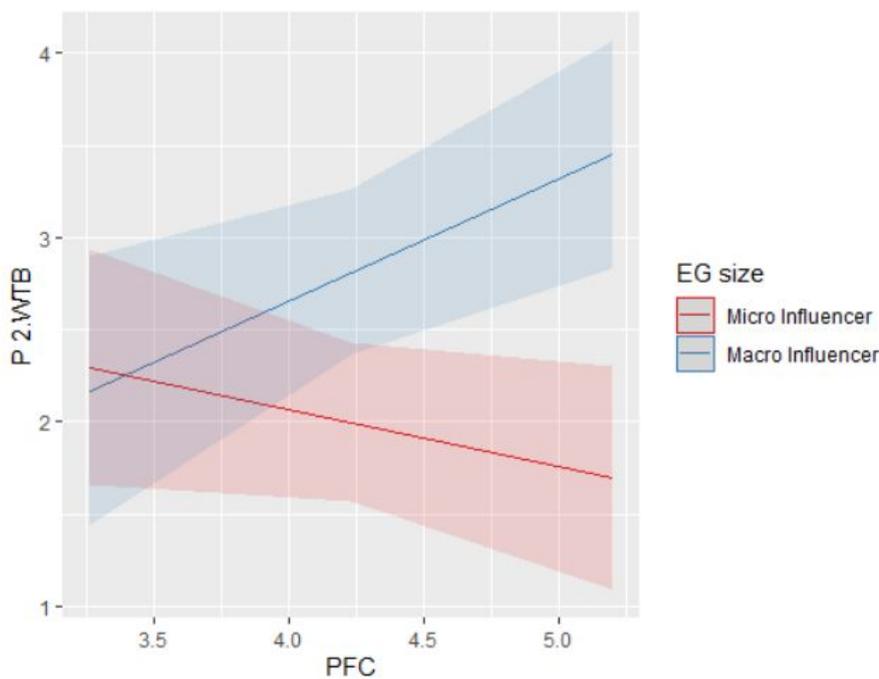
	Df	F	P-value (>F)
(Intercept)	1	16.002	0.00008834 ***
PFC	1	2.8268	0.09423 .
EG.size	1	2.4196	0.12137
EG.fit	1	0.0093	0.92340

PFC:EG.size	1	4.2282	0.04102 *
PFC:EG.fit	1	0.1652	0.68488
EG.size:EG.fit	1	0.6073	0.43669
PFC:EG.size:EG.fit	1	1.1517	0.28445
Residuals	205		

Códigos de significancia: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Se puede observar que la interacción entre el índice de preferencia por la consistencia y el tamaño del influencer es significativa al 95% para determinar la propensión de compra, con una potencia del 72%¹⁰.

Figura 39: Diferencia de medias para la propensión a comprar



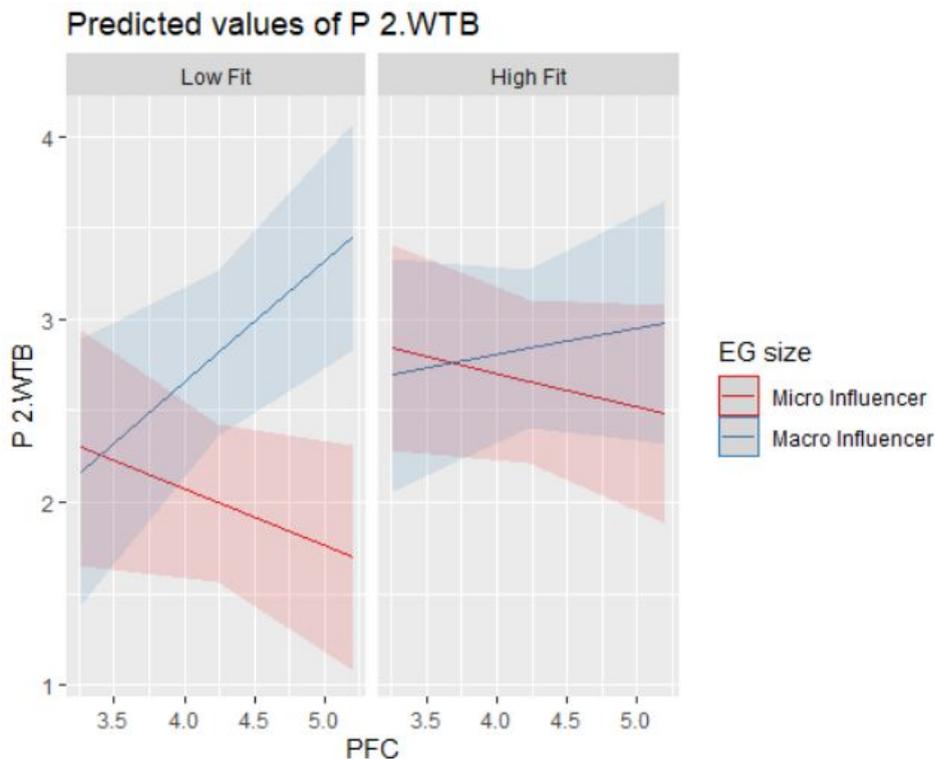
Intervalos de confianza al 95%

Tras explicitar gráficamente la relación descubierta entre la interacción del tamaño y preferencia por la consistencia sobre la propensión de compra, podemos ver cómo para niveles bajos de preferencia por la consistencia del consumidor la propensión a la compra de los consumidores no es estadísticamente diferente para microinfluencers o para macroinfluencers. En cambio, para niveles altos de preferencia por consistencia, con un

¹⁰ Se necesitarían 260 observaciones para llegar a una potencia del 80%

95% de confianza, la propensión de compra es más alta para influencers con grandes cantidades de seguidores que para influencers con poca cantidad de seguidores. A medida que el consumidor posee una mayor preferencia por la consistencia, tendrá una mayor propensión a la compra si se trata de un macro influencer, pero una menor propensión de compra si se trata de un microinfluencer.

Figura 40: Diferencia de medias para la propensión a comprar



Intervalos de confianza al 95%

Si realizamos el mismo gráfico anterior pero además dividido entre bajo encaje entre producto-influencer y alto encaje, podemos ver que la afirmación realizada anteriormente se cumple para productos con bajo ajuste con el influencer, pero para productos con alto ajuste la propensión de compra no es estadísticamente diferente para micro o macro influencers, para ningún nivel de preferencia por la consistencia. Una PFC alta demuestra, además, ser perjudicial para el éxito de las publicidades de los microinfluencers, ya que a medida que la PFC sube, la propensión de pago baja tanto para productos sin encaje como para con encaje. A su vez, el efecto contrario ocurre con los macroinfluencers, donde a mayor PFC mayor propensión de pago generan, independientemente del producto. Esto tal vez es causa de que el público está acostumbrado a que los macroinfluencers publiciten una gran variedad de productos o es más esperable que mayor cantidad de empresas los

contraten, mientras que de los microinfluencers se espera que hagan un esfuerzo extra por mantenerse consistentes en sus publicaciones y contenido.

5.10.3 Análisis de la propensión a recomendar

En tercer lugar, analizaremos cómo influyen las características percibidas del influencer en la propensión a recomendar, sin controlar y controlando por la preferencia por la consistencia del consumidor.

Sin controlar por el índice de preferencia por la consistencia

Se llevó a cabo la prueba de Levene para la homogeneidad de las varianzas a través de los 4 grupos de observaciones (Micro Influencer - Low Fit, Micro Influencer - High Fit, Macro Influencer - Low Fit, Macro Influencer - High Fit).

Tabla 17: Test de Levene para la Homogeneidad de la Varianza (centro = mediana)

Df	Valor estadístico F	P-value (>F)
3	2.5806	0.05456 .

Códigos de significancia: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Al ser el p-valor mayor a 0.5 y menor a 0.1, se puede rechazar la hipótesis nula del test de que las varianzas poblacionales son iguales con una confianza del 90% (es decir, hay homogeneidad de varianza entre algunos grupos).

Tabla 18: Test ANOVA para la diferencia de medias entre grupos de la propensión a recomendar

	Df	F	P-test (>F)
(Intercept)	1	168.7123	<0.0000000000000002 ***
EG.size	1	0.6834	0.4093
EG.fit	1	0.8944	0.3454
EG.size:EG.fit	1	0.3386	0.5613
Residuos	209		

Códigos de significancia: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Se puede ver que no hay ninguna variable significativa que influya en la propensión a recomendar de los consumidores.

Controlando por el índice de preferencia por la consistencia

Se realizó un test de ANOVA sobre la propensión de pago controlando por el índice de preferencia por la consistencia.

Tabla 19: Test ANOVA para la diferencia de medias entre grupos de la propensión a recomendar

	Df	F	P-value (>F)
(Intercept)	1	7.4645	0.006842 **
PFC	1	0.0000	0.998200
EG.size	1	1.1929	0.276024
EG.fit	1	0.3468	0.556600
PFC:EG.size	1	1.5323	0.217185
PFC:EG.fit	1	0.1729	0.678003
EG.size:EG.fit	1	0.1225	0.726675
PFC:EG.size:EG.fit	1	0.2070	0.649623
Residuals	205		

Códigos de significancia: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

El análisis no cambia respecto a no controlar por el índice de PFC ya que no hay ninguna variable significativa que influya en la propensión a recomendar de los consumidores.

5.10.4 Análisis de la propensión a repostear

En último lugar, analizaremos cómo influyen las características percibidas del influencer en la propensión a repostear contenido, sin controlar y controlando por la preferencia por la consistencia del consumidor.

Sin controlar por el índice de preferencia por la consistencia

Se llevó a cabo la prueba de Levene para la homogeneidad de las varianzas a través de los 4 grupos de observaciones (Micro Influencer - Low Fit, Micro Influencer - High Fit, Macro Influencer - Low Fit, Macro Influencer - High Fit).

Tabla 20: Test de Levene para la Homogeneidad de la Varianza (centro = mediana)

Df	Valor estadístico F	P-value (>F)
3	1.9275	0.1262

Códigos de significancia: 0 '****' 0.001 '***' 0.01 '**' 0.05 '.' 0.1 '.' 1

Al ser el p-valor mayor a 0.1, no se puede rechazar la hipótesis nula del test de que las varianzas poblacionales son iguales (es decir, hay homogeneidad de varianza entre algunos grupos). Por esto es que luego al realizar el test ANOVA se deberá implementar la corrección con la prueba de White.

Tabla 21: Test ANOVA para la diferencia de medias entre grupos de la propensión a repostear

	Df	F	P-test (>F)
(Intercept)	1	148.9209	< 0.0000000000000000 2 ***
EG.size	1	4.9708	0.02684 *
EG.fit	1	3.0796	0.08075 .
EG.size:EG.fit	1	4.0383	0.04576 *
Residuos	208		

Códigos de significancia: 0 '****' 0.001 '***' 0.01 '**' 0.05 '.' 0.1 '.' 1

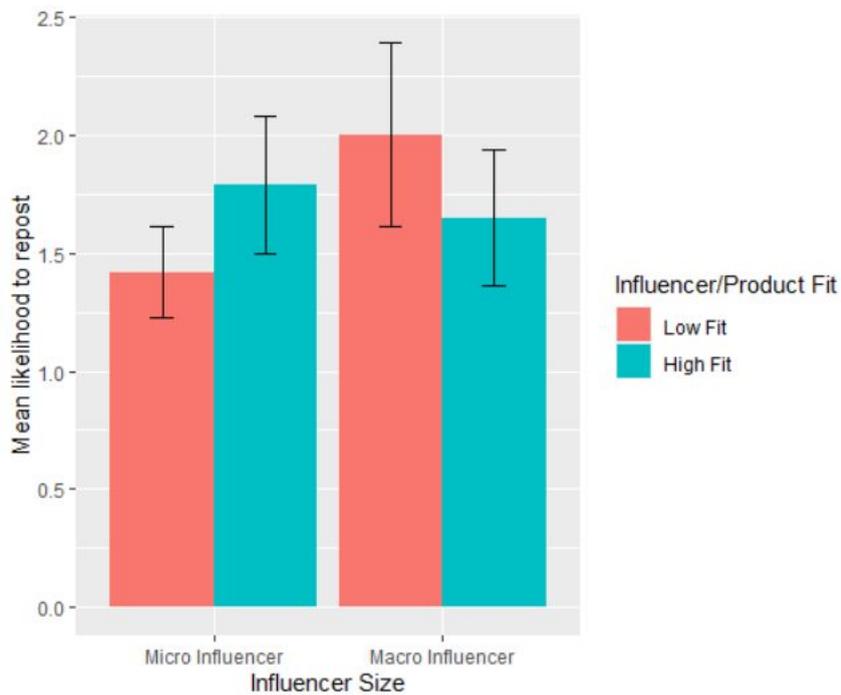
El tamaño del influencer es una característica que genera una diferencia en la media de propensión a repostear significativa al 95% con una potencia del 36,59% (se necesitarían 742 observaciones para obtener una potencia del 80%). Por otro lado, el encaje entre el producto y el influencer generan una diferencia en la media significativa al 90% con potencia del 10,14%¹¹. Finalmente, la interacción entre el tamaño del influencer y el encaje

¹¹ Se necesitarían 151.324 observaciones para obtener una potencia del 80%

entre producto e influencer genera una diferencia en la media significativa al 95% con una potencia del 95,58%.

Generamos el gráfico para observar la diferencia de medias separados por el encaje entre producto influencer.

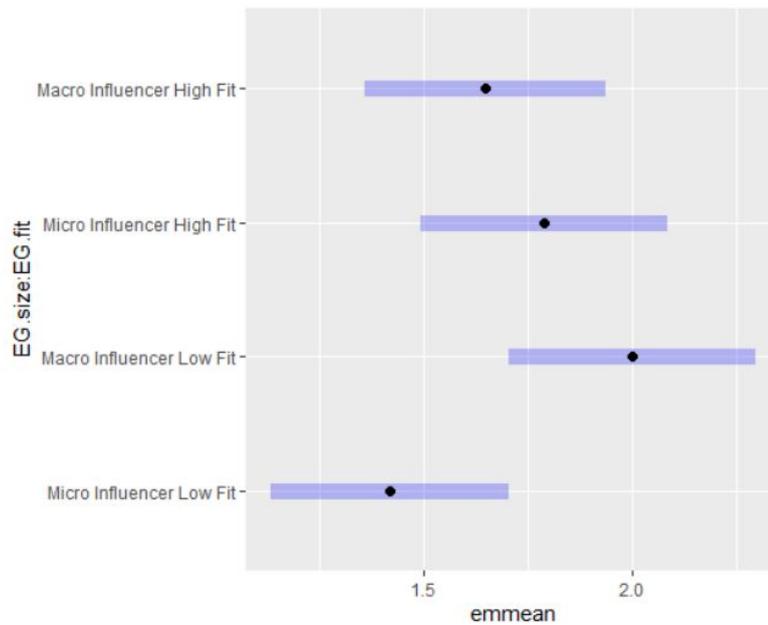
Figura 41: Diferencia de medias para la propensión a repostear



Intervalos de confianza al 95%

Podemos observar que, para un producto con poco encaje, un macroinfluencer elevará la propensión a repostear por sobre el microinfluencer.

Figura 42: Diferencia de medias para la propensión a repostear



Intervalos de confianza al 90%

Este gráfico con intervalos de confianza al 90% nos muestra que para un gran encaje de producto-influencer, los influencers micro y macro son reposteados en igual medida. En cambio, si el encaje es bajo, se repostea menos a los microinfluencers que a los macroinfluencers.

Controlando por el índice de preferencia por la consistencia

Se realizó un test de ANOVA sobre la propensión de pago controlando por el índice de preferencia por la consistencia.

Tabla 22: Test ANOVA para la diferencia de medias entre grupos de la propensión a repostear

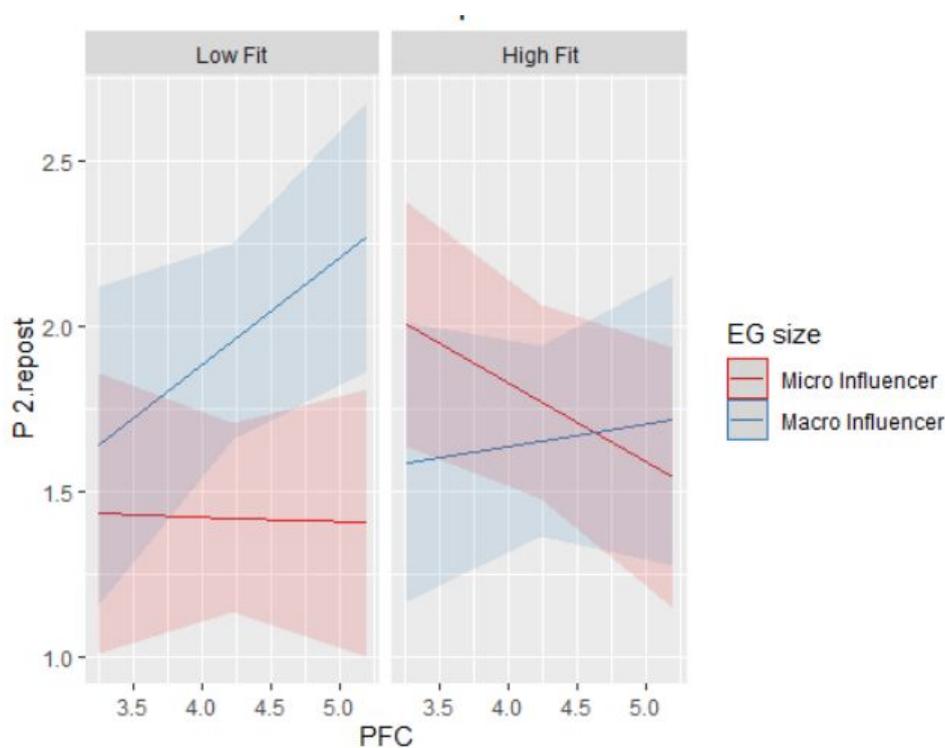
	Df	F	P-value (>F)
(Intercept)	1	12.0513	0.0006311 ***
PFC	1	0.0284	0.8663654
EG.size	1	0.3178	0.5735336
EG.fit	1	2.2706	0.1333910
PFC:EG.size	1	0.8062	0.3703028
PFC:EG.fit	1	1.3855	0.2405241

EG.size:EG.fit	1	0.0664	0.7969109
PFC:EG.size:EG.fit	1	0.0061	0.9376114
Residuals	205		

Códigos de significancia: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

No hay ninguna variable significativa que influya en la propensión a recomendar de los consumidores.

Figura 43: Diferencia de medias para la propensión a repostear



Intervalos de confianza al 90%

Se puede apreciar en el gráfico que, con una confianza del 90%, para productos con un bajo ajuste a la categoría del influencer, cuanto mayor sea la preferencia por la consistencia del consumidor, mayor será la propensión a repostear si el influencer es muy popular, mientras que esta propensión se mantendrá constante si se trata de un microinfluencer. Esto quiere decir que, para altos niveles de PFC, la propensión de repostear es mayor para macroinfluencers si el producto no encaja con la categoría del mismo. En cambio, para productos con alto ajuste con la categoría del influencer, las diferencias entre la propensión de reposteo no es estadísticamente diferente entre los dos tipos de influencers, para ningún nivel de PFC.

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 las 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 *belleza*), utilizando información del clustering.

El objetivo del experimento llevado a cabo fue entender cuál es el efecto de la popularidad y profesionalismo del influencer en conjunto con el encaje entre el influencer y el producto 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. La popularidad del influencer fue diferenciada según la cantidad de seguidores en los perfiles de instagram (pocos seguidores vs. muchos seguidores), y está ligada positivamente con el profesionalismo del influencer percibido por el público ya que un a mayor cantidad de seguidores el influencer tiende a tener más años de construcción de carrera, equipo más profesional y denotar más habilidad. Por otro lado, para diferenciar un producto que se ajuste al influencer contra uno que no, se eligieron un producto que esté dentro de la categoría y/o habilidad del perfil en cuestión y otro que no.

Los resultados del experimento demuestran varios puntos importantes:

- Si se publicita un producto con un influencer cuya categoría esté relacionada (en cuanto a habilidades requeridas, utilización, etc) al mismo, la propensión de pago del consumidor por el producto será estadísticamente igual si se trata de un micro influencer o un macro influencer. Si el producto no cae dentro de la categoría del influencer, de nuevo la propensión de pago de los compradores será igual sea el influencer popular o no, pero será menor en aproximadamente un 44% que si se tratara de un producto con alto encaje. Estos resultados no dependen de la preferencia por la consistencia que tenga el público del influencer sino que son independientes a este rasgo de personalidad.
- Si un producto es patrocinado por un influencer con el cual tenga encaje con su categoría, no hay diferencia estadística si el instagrammer era micro o macro. Si, por el contrario, el producto es patrocinado por alguien con el cual no encaja en su categoría y encima tiene pocos seguidores, la propensión a compra será menor en aproximadamente un 19%. Si el producto no encaja con el influencer pero este es macro, la propensión a compra de los consumidores será igual a que si el producto sí tuviera relación con la categoría del influencer. Este resultado sólo vale si el público tiene altos niveles de preferencia por la consistencia, ya que para bajos niveles de ésta la propensión a compra es igual para cualquier combinación entre popularidades de influencers y encaje del producto. Se puede ver con esto que lo

que generalmente hace que el público prefiera a influencers con más seguidores es la consistencia que demuestran en sus publicaciones y contenido.

- La propensión a recomendar no se ve alterada por ninguna combinación de popularidad de influencer y encaje del producto, y tampoco depende de la preferencia por la consistencia que tenga el público.
- Si un producto es publicitado por un influencer cuya categoría está relacionada al producto en cuestión, la propensión a repostear no varía si el influencer tiene muchos o pocos seguidores. Si el producto no está relacionado a la categoría del influencer, la propensión a repostear es mayor si se trata de un influencer muy popular. Además, la propensión de reposteo no es diferente estadísticamente para macro influencers si el producto tiene alto o bajo encaje con este. Estos resultados aplican para altos niveles de preferencia por la consistencia del público, mientras que para bajos niveles de este no hay diferencias significativas para ninguna combinación de influencer-producto.

Es interesante remarcar que, si bien literatura anteriormente citada sobre el tema encontró que el product-endorser fit aumenta la propensión de compra de igual manera en celebridades que en influencers (Alexander P. Schouten, Loes Janssen & Maegan Verspaget), nuestro estudio demostró que eso cambia si en vez de celebridades se utilizan dos influencers con distinto tamaño. Dentro del mundo de los influencers, la interacción entre tamaño del influencer y el encaje entre el producto y el sponsor sí son significativas sobre la propensión de compra que generan en el consumidor, ya que el encaje genera un efecto distinto sobre la propensión de compra dependiendo del tamaño del influencer. Otro punto importante es que mientras literatura anterior llegó a la conclusión que los micro influencers generaban mayor propensión a compra que los macro influencers (Samantha Kay, Rory Mulcahy & Joy Parkinson), nuestro estudio demostró que eso depende del contexto del producto en cuestión, es decir, el tamaño del influencer no es el único determinante a la hora de manipular la propensión de pago del comprador ya que se producen efectos diferentes sobre el consumidor si se trata de un producto relacionado al influencer o no.

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.

Tras analizar los resultados obtenidos, podemos concluir ciertas recomendaciones generales. En primer lugar, el rol que toma la preferencia por la consistencia del público objetivo en la elección del capital de influencers óptimos a utilizar para sponsorar un producto. Se debería estudiar y analizar este rasgo de personalidad en los que serían potenciales consumidores del producto a vender. Si el nivel de preferencia por la consistencia es bajo, no será significativo qué tipo de influencer promocione el producto. Por otro lado, si el nivel es alto, la recomendación general es utilizar un macroinfluencer que tenga ajuste con el producto para elevar la propensión de compra y de reposteo. En segundo lugar, la importancia de el encaje entre el influencer y el producto para elevar la propensión de pago. Se deberá hacer un análisis tanto del producto que se quiere vender como de los instagrammers disponibles en el mercado, y ajustar lo mejor posible la categoría a la que pertenece el producto con la del influencer. Por último, la ausencia de una característica *driver* para elevar la propensión a recomendar o eWOM (*electronic word-of-mouth*), es decir que es probable que este proceso sea afectado por otras características (como la compra y testeo del producto personalmente) y no por las percepciones que se tenga del influencer y su publicación promocionada.

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.

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. Por último, también se recomienda para posteriores investigaciones aumentar la cantidad de participantes del experimento para asegurar una potencia estadística mayor sobre las conclusiones.

7. Apéndice

7.1 Variables de los Datasets

7.1.1 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 $(likes + comentarios)/followers$.

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.

7.1.2 Dataset 'instagramPosts' (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.

7.2 Transformaciones de las Variables

7.2.1 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 posteos del tipo imagen en relación a los posteos 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 posteos del tipo video en relación a los posteos 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.

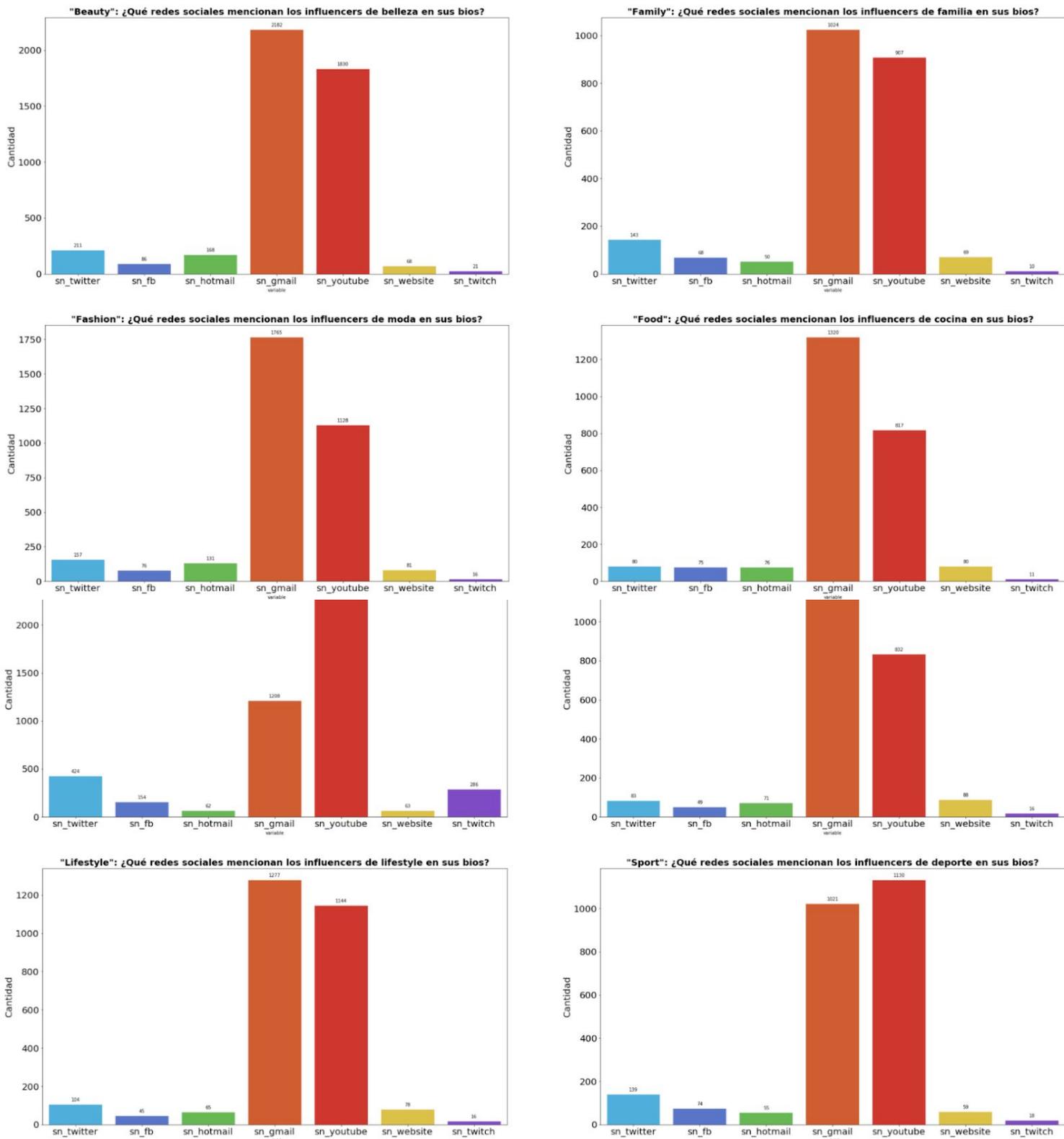
arrobaspromedio: Variable del tipo numérica que representa la cantidad promedio de arrobas (etiquetaciones) que el instagram agrega en sus publicaciones. Para obtener esta variable, se creó en el dataset *instagramPosts* una variable llamada *arrobaspromedio* que extraiga las arrobas usados en cada publicación. 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.

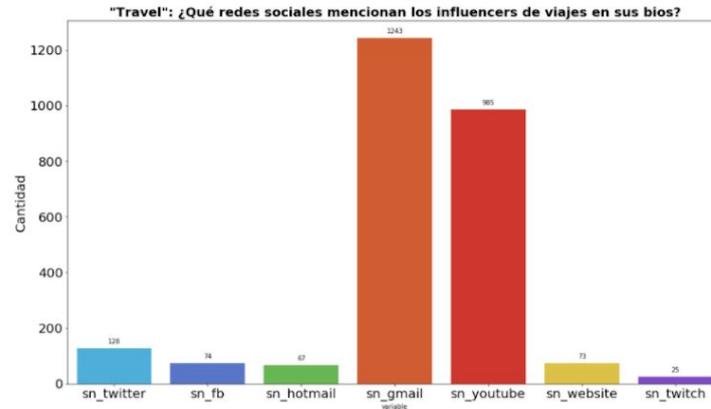
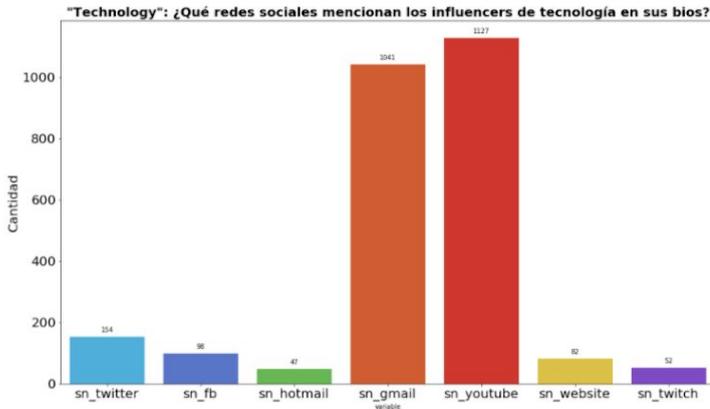
frecuencia_post: Variable del tipo numérica que representa la cantidad de horas promedio entre cada publicación que hace el instagram. Para obtener esta variable, se buscó cada *instagram_id* del dataset *instagram* en el dataset *instagramPosts* para obtener las publicaciones de cada instagram, se las 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 posteos seguidos).

7.3 Gráficos Adicionales (Análisis Exploratorio)

Figura 44: 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 45: Distribución del Tiempo entre Publicaciones (en Horas)



Figura 46: Distribución de la Participación de Videos en las Publicaciones



Figura 47: Distribución de la Participación de Imágenes en las Publicaciones

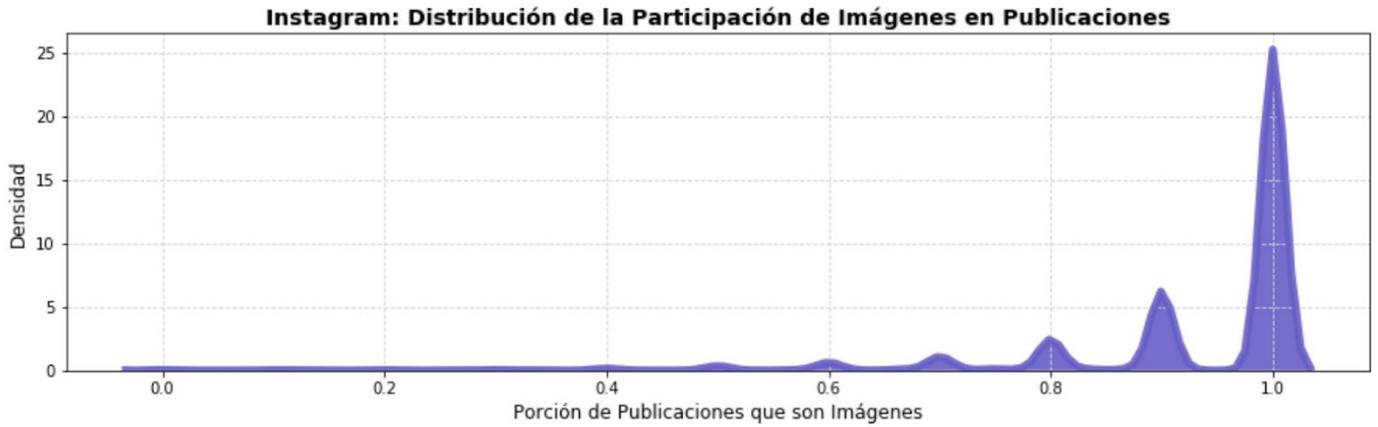


Figura 48: Distribución de la Cantidad de *Hashtags* Promedio utilizados por los influencers en sus publicaciones

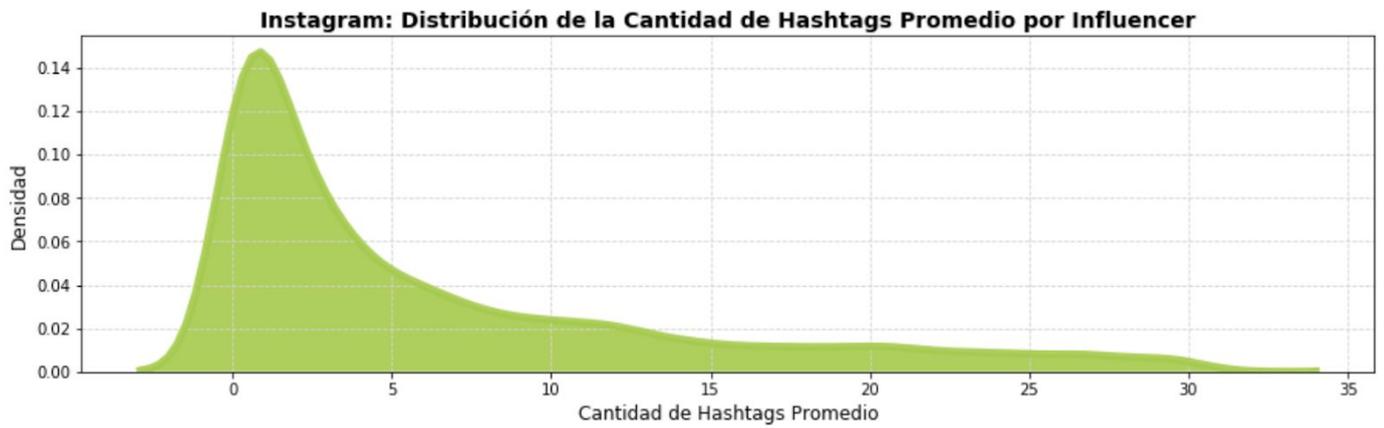
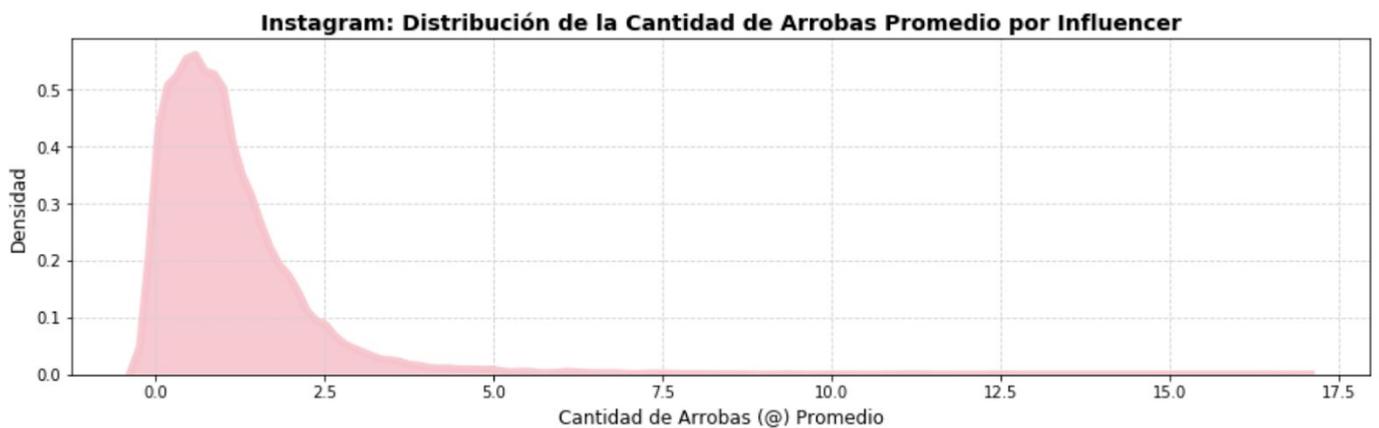


Figura 49: Distribución de la Cantidad de *Arrobos* (@) Promedio utilizados por los influencers en sus publicaciones



7.4 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.

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

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

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

Influencer Persona 3: Influencer de Videojuegos *harleyplays1*

<p>NOMBRE DE USUARIO EN INSTAGRAM</p> <p>harleyplays1</p>	<p>MARKET SIZE</p> <p> 10 %</p>	<p>TYPE</p> <p>Gaming Influencer</p>
	<h3>Publicaciones</h3> <ul style="list-style-type: none">◦ 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 <i>hashtags</i> en sus posts para que más personas la encuentren y comiencen a seguirla (#ad, #art, #youtube, #cosplay, #youtuber). A su vez, los <i>gamers</i> que se dedican a promocionar su actividad en torneos de videojuegos o <i>gameplays</i>, suelen tener sponsors (tecnología, equipos en torneos), por lo que utilizan <i>hashtags</i> 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)	
<h3>Datos</h3> <p>Género: Femenino</p> <hr/> <p>Edad: 25 Años</p> <hr/> <p># Seguidores: ~85K</p> <hr/> <p>Engagement Rate: 0.16%</p> <hr/> <p># Publicaciones (3 meses): 200</p>	<h3>Seguidores</h3> <ul style="list-style-type: none">◦ Gran interacción con sus suscriptores/seguidores, pero no necesariamente en Instagram, sino más bien en plataformas de <i>streaming</i>.◦ Bajo <i>Engagement Rate</i> 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 <i>gamer</i>.◦ Le gusta crear confianza y cercanía con los que la siguen y hacerlos participar de juegos o <i>challenges</i>. Además, está atenta a sus sugerencias para crear contenido nuevo (sugerencias de otros juegos por ejemplo).	
<h3>Gain Points</h3> <p>La gente en las redes la aprecia por su personalidad encantadora, sus comentarios sobre juegos de moda y sus <i>gameplays</i> entretenidos y graciosos.</p>		<h3>Pain Points</h3> <p>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.</p>
<h3>Presencia en otras redes sociales</h3> <p> </p>		

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

<p>NOMBRE DE USUARIO EN INSTAGRAM</p> <p>veganmel.eats</p>	<p>MARKET SIZE</p> <p> 15 %</p>	<p>TYPE</p> <p>Food Influencer</p>
	<h3>Publicaciones</h3> <ul style="list-style-type: none">• Sube fotos de recetas y de su dieta diaria.• Usa muchos <i>hashtags</i> en sus posts para que las personas interesadas puedan encontrar su blog y sus recetas fácilmente (#ad, #food, #feedfeed, #foodie, #f52grams, #foodporn, #foodphotography, #foodblogger)• El 96% de su contenido son imágenes (fotos de los platos que prepara/come), mientras que el resto son videos (recetas rápidas)• Publica cada 3 o 4 días, para no saturar a su público y para darle tiempo a crear contenido relevante.• No hace demasiados posts sponsorados.	
<h3>Datos</h3> <p>Género: Femenino</p> <hr/> <p>Edad: 34 Años</p> <hr/> <p># Seguidores: ~30K</p> <hr/> <p>Engagement Rate: 1.2%</p> <hr/> <p># Publicaciones (3 meses): 400</p>	<h3>Seguidores</h3> <ul style="list-style-type: none">• Buena interacción con sus seguidores, sobre todo en formato "preguntas y respuestas" para difundir información e ideas sobre estilos de dieta.• Alto <i>Engagement Rate</i> ya que su público, aunque pequeño, es leal a ella.• Los usuarios en Instagram que la siguen lo hacen por sus consejos sobre alimentación variada (vegana, vegetariana, etc) y para adquirir ideas sobre nuevos platos o recetas.• 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.	
<h3>Gain Points</h3> <p>La gente en las redes la aprecia por la información e ideas que divulga, las cuales incluyen tips para seguir una dieta balanceada y deliciosa que puede ajustarse a todo tipo de usuarios.</p>		<h3>Pain Points</h3> <p>Tiene que esforzarse mucho en crear contenido innovador e interesante ya que tiene riesgo de volverse repetitiva o no diferenciarse con respecto a otros influencers similares.</p>
<h3>Presencia en otras redes sociales</h3> <p> </p>		

7.6 Pretest: *Influencer Personas* y sus respectivas publicaciones en Instagram

A partir del resultado del clustering utilizando la categoría “belleza”, 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 microinfluencer “micro1”: WorldOfSofi

Perfil del Influencer

WorldOfSofi

690 posts 16,9K followers 563 following

Message

Sofi Melandi

Makeup Blogger
23 años | 21/02/1996
Sol/Asc/Luna: ♋ ♋ ♋
Diseño Multimedia, UP
Buenos Aires, Argentina
DMs siempre abiertos para collabs

Grid of 6 images showing makeup looks and a person using a Nintendo Switch.

Publicación de un producto sin ajuste

WorldOfSofi

Un gato en mi regazo y una #Switch en mi mano ❤️
Es todo lo que necesito para hacer que este fin de semana sea perfecto! 🐱💖🌟 En @nintendo podés conseguir la tuya del color que quieras!
#nintendoswitch #switch #nintendo #ad #gamer #gamer #videogame #tecnologia #cat #catlover

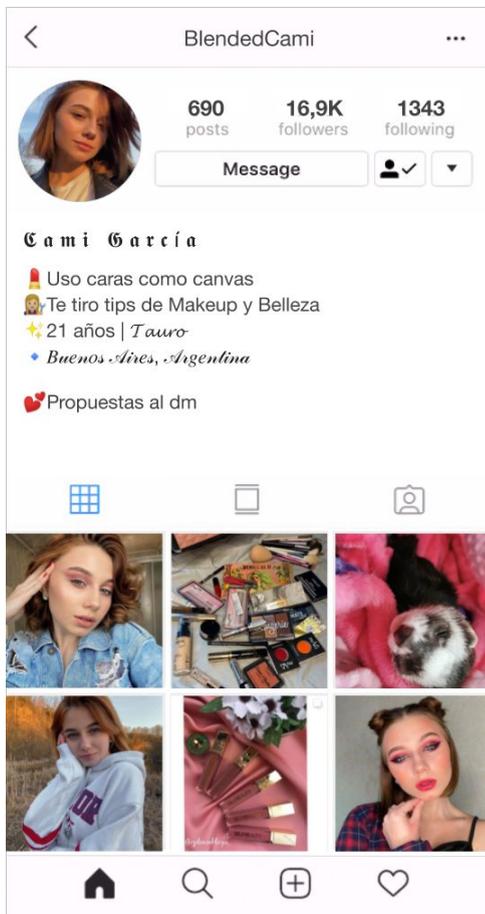
Publicación de un producto con ajuste

WorldOfSofi

Alguien más es fan de los colores neutros y marrones? 🐼 Esta paleta de @colourcandy está volviéndome LOCA. Son tonos súper cálidos que se pueden usar en cualquier situación, 10/10 🌟🌟
20% OFF con código WORLDSOFI 😊 #colourcandy #goingcoconutpalette #eyeshadow #sponsoredby

Perfil microinfluencer "micro2": BlendedCami

Perfil del Influencer



BlendedCami

690 posts 16,9K followers 1343 following

Message

Cami García

- 👉 Uso caras como canvas
- 👉 Te tiro tips de Makeup y Belleza
- 🌟 21 años | Tauro
- 📍 Buenos Aires, Argentina

Propuestas al dm

Publicación de un producto sin ajuste



Instagram

BlendedCami



Nunca me apasionó leer, pero desde que me recomendaron @audible no puedo dejar de escuchar #audiolibros 🎧 Puedo "escuchar" un libro mientras me maquillo y aprovechar al máximo mi tiempo ✨ Suscribite en www.audible.com/BlendedCami para 2 meses gratis! #audible #audiobooks #libro

Publicación de un producto sin ajuste



Instagram

BlendedCami



AMO estas barritas de granola @CerealNix. Cada vez que tengo hambre pero todavía no es la hora de comer, agarro una de estas y puedo seguir adelante con mucha fuerza 🙌 Visitá www.cerealnix.com y hacé tu pedido! 📱

#granola #barritas #cereal #cerealnix #sponsored

Perfil macroinfluencer “macro1”: LauraBlushh

Perfil del Influencer

The screenshot shows the Instagram profile of LauraBlushh. At the top, there is a back arrow, the username 'LauraBlushh', and a three-dot menu. Below this is a circular profile picture of Laura Blakonian. To the right of the profile picture, it displays '5320 posts', '1,1M followers', and '1183 following'. There is a 'Message' button and a checkmark icon. The bio reads: 'Laura Blakonian', 'Maquilladora profesional', 'CEO @BlakBeauty', and 'Descubre tu artista interior'. Below the bio, there is a link to a YouTube video: 'ÚLTIMO VIDEO: youtu.be/ZmHpfEzhZvU' and an email address: 'laurablak@gmail.com'. At the bottom of the profile, it says 'Followed by: plumandtore, geoffreystar, mykietutorials and 8 more'. The bottom navigation bar shows icons for home, search, post, and heart.

Publicación de un producto sin ajuste

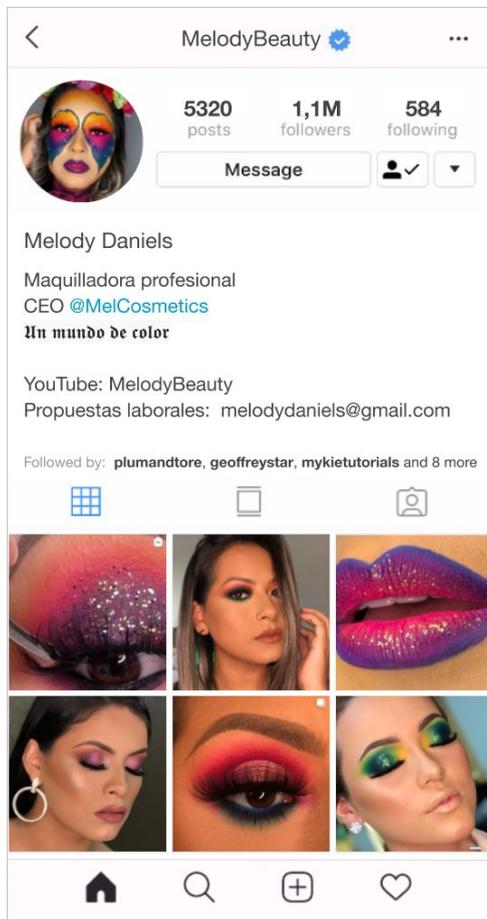
The screenshot shows an Instagram post by LauraBlushh. The post features a photo of a woman wearing large, pink and white Beats Solo Wireless headphones. The caption reads: 'Siempre me preguntan cómo escucho música 😊 Uso los audífonos #Beats Solo Wireless! La música te hace sentir libre! Los cables no deberían ser obstáculo. Con materiales duraderos y un ajuste cómodo, estos auris son la mejor opción para experimentar la libertad inalámbrica. #sponsored'. The bottom navigation bar shows icons for home, search, post, and heart.

Publicación de un producto con ajuste

The screenshot shows an Instagram post by LauraBlushh. The post features a photo of a ColourCandy makeup palette with various shades of eyeshadow and blush, along with a brush. The caption reads: 'Estoy tan contenta tener esta bella paleta de @colourcandy 😊 Es tan diversa que la puedo usar con maquillaje de día o de noche. La fórmula me encanta! No deja mucho residuo y tiene muy buena pigmentación. De mis favoritas ❤️ Usá mi código de descuento 😊: LAURABLUSH #sponsored'. The bottom navigation bar shows icons for home, search, post, and heart.

Perfil macroinfluencer "macro2": MelodyBeauty

Perfil del Influencer



Instagram profile of MelodyBeauty. The profile shows a grid of 6 posts, a bio with contact information, and a list of followed accounts. The bio includes: Melody Daniels, Maquilladora profesional, CEO @MelCosmetics, In mundo de color, YouTube: MelodyBeauty, Propuestas laborales: melodydaniels@gmail.com. The followed accounts are: plumantore, geoffreystar, mykietutorials and 8 more.

Publicación de un producto sin ajuste



Instagram post by MelodyBeauty featuring CerealNix bars. The post includes a photo of the bars and a caption: "El perfecto snack para la mediatarde y mediamañana. Estoy OBSESIONADA con estas barritas de @CerealNix, son livianas y tienen muchos nutrientes que hacen que tu piel y pelo se vean muy saludables 🍷 Usá mi código MBEAUTY y recibí tus primeras 10 barritas gratis 🌟🌟🌟 #ad".

Publicación de un producto sin ajuste



Instagram post by MelodyBeauty featuring Audible audiobooks. The post includes a photo of a stack of books and an Audible audiobook player. The caption reads: "Siempre me costó mucho encontrar tiempo para leer los libros que me interesaban 😭 Pero recientemente conocí los #audiobooks de @audible y no lo puedo creer, pude terminar 6 libros que tenía pendientes en solo dos semanas 😊 Los primeros 2 meses son gratis :) #sponsored".

7.7 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:*

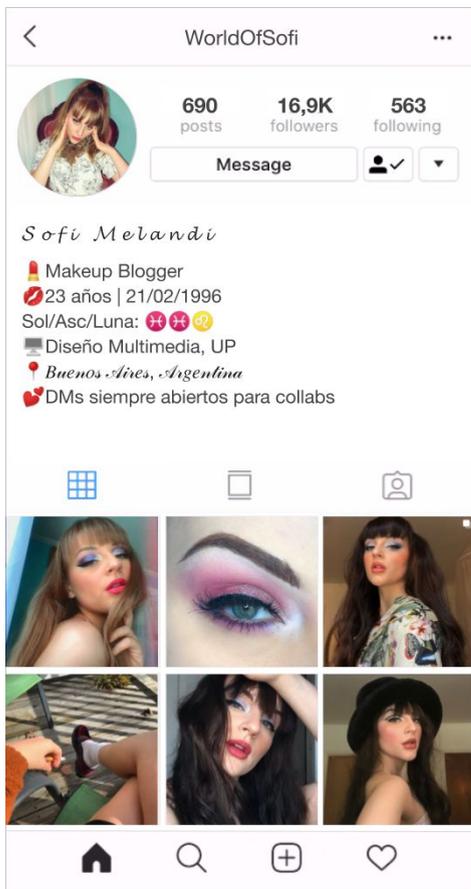
Después de observar el post anterior pienso que...

- 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 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

7.8 Experimento Principal: *Influencer Personas* y sus respectivas publicaciones en Instagram

Perfil microinfluencer: Versión Final

Perfil del Influencer



WorldOfSofi

690 posts | 16,9K followers | 563 following

Message

Sofi Melandi

- Makeup Blogger
- 23 años | 21/02/1996
- Sol/Asc/Luna: ♋♋♋
- Diseño Multimedia, UP
- Buenos Aires, Argentina
- DMs siempre abiertos para collabs

Grid of 6 images showing makeup looks and a person.

Publicación con ajuste



WorldOfSofi



Estoy tan contenta tener esta bella paleta de @colourcandy 🥰 Es tan diversa que la puedo usar con maquillaje de día o de noche. La fórmula me encanta! No deja mucho residuo y tiene muy buena pigmentación. De mis favoritas 🍷 Usá mi código de descuento 🙄: WORLDDFSOFI #sponsored

ajuste

Publicación sin



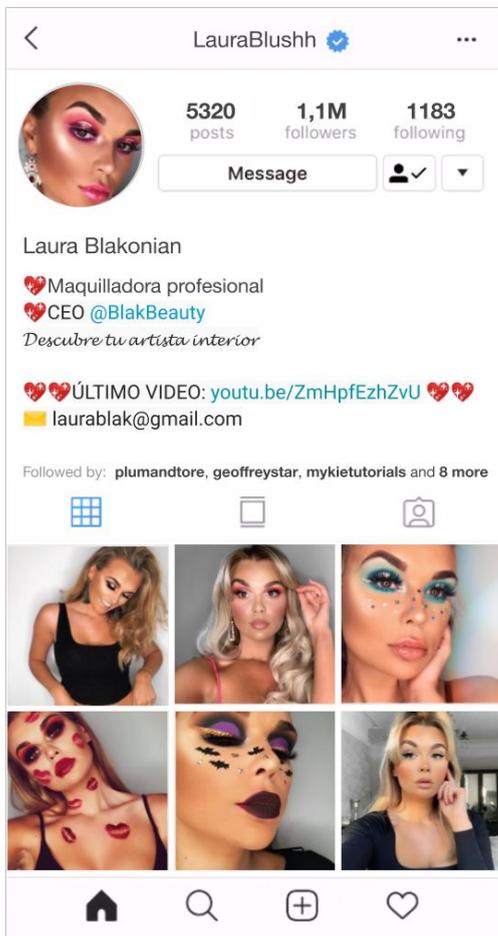
WorldOfSofi



Siempre me costó mucho encontrar tiempo para leer los libros que me interesaban 🙄 Pero recientemente conocí los #audiobooks de @audible y no lo puedo creer, pude terminar 6 libros que tenía pendientes en solo dos semanas 😊 Los primeros 2 meses son gratis ;) #sponsored

Perfil macroinfluencer: Versión Final

Perfil del Influencer



LauraBlushh

5320 posts 1,1M followers 1183 following

Message

Laura Blakonian

Maquilladora profesional
CEO @BlakBeauty
Descubre tu artista interior

ÚLTIMO VIDEO: youtu.be/ZmHpFzhZvU

laurablak@gmail.com

Followed by: plumandtore, geoffreystar, mykietutorials and 8 more

Publicación con ajuste



Instagram

LauraBlushh



Estoy tan contenta tener esta bella paleta de @colourcandy 🥰 Es tan diversa que la puedo usar con maquillaje de día o de noche. La fórmula me encanta! No deja mucho residuo y tiene muy buena pigmentación. De mis favoritas 🍷 Usá mi código de descuento 🙌: LAURABLUSH #sponsored

ajuste

Publicación sin



Instagram

LauraBlushh



Siempre me costó mucho encontrar tiempo para leer los libros que me interesaban 🙄 Pero recientemente conocí los #audiobooks de @audible y no lo puedo creer, pude terminar 6 libros que tenía pendientes en solo dos semanas 😂 Los primeros 2 meses son gratis :) #sponsored

7.9 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 (microinfluencer o macroinfluencer) junto con alguna de las dos publicaciones descritas (con ajuste al influencer o sin ajuste). A partir de esto, el participante deberá contestar cuál es su disposición a pagar por el producto, a comprarlo, recomendarlo o re-publicarlo. La segunda sección de preguntas buscan entender la personalidad del consumidor y la tercera sección corresponde a preguntas demográficas o personales.

[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 (micro o macroinfluencer)

[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 (con ajuste a la categoría del influencer o sin ajuste)

- ¿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 Preferencia por la Consistencia]

1. Prefiero estar cerca de personas cuyas reacciones puedo anticipar.
1 = Totalmente en desacuerdo and 5 = Totalmente de acuerdo.
2. Es importante para mí que mis acciones sean consistentes con mis creencias.
1 = Totalmente en desacuerdo and 5 = Totalmente de acuerdo.
3. Incluso si mis actitudes y acciones me parecieran consistentes entre sí, me molestaría que no se vean consistentes bajo la mirada de otros.
1 = Totalmente en desacuerdo and 5 = Totalmente de acuerdo.
4. Es importante para mí que los que me conocen puedan predecir qué haré.
1 = Totalmente en desacuerdo and 5 = Totalmente de acuerdo.
5. Quiero ser descrita/o por otros como una persona estable y predecible.
1 = Totalmente en desacuerdo and 5 = Totalmente de acuerdo.
6. Las personas admirables son consistentes y predecibles.
1 = Totalmente en desacuerdo and 5 = Totalmente de acuerdo.
7. Parecer consistente es una parte importante de la imagen que presento al mundo.
1 = Totalmente en desacuerdo and 5 = Totalmente de acuerdo.
8. Me molesta cuando alguien de quien dependo es impredecible.
1 = Totalmente en desacuerdo and 5 = Totalmente de acuerdo.

9. No me gusta dar la apariencia de ser inconsistente.
1 = Totalmente en desacuerdo and 5 = Totalmente de acuerdo.
10. Me incomoda si me doy cuenta que mi comportamiento contradice mis creencias.
1 = Totalmente en desacuerdo and 5 = Totalmente de acuerdo.
11. Un requerimiento importante para cualquier amigo mío es tener consistencia personal.
1 = Totalmente en desacuerdo and 5 = Totalmente de acuerdo.
12. Típicamente, prefiero hacer las cosas de la misma manera.
1 = Totalmente en desacuerdo and 5 = Totalmente de acuerdo.
13. Me desagradan las personas que están constantemente cambiando su opinión.
1 = Totalmente en desacuerdo and 5 = Totalmente de acuerdo.
14. Quiero que mis amigos cercanos sean predecibles.
1 = Totalmente en desacuerdo and 5 = Totalmente de acuerdo.
15. Es importante para mi que los otros me vean como una persona estable.
1 = Totalmente en desacuerdo and 5 = Totalmente de acuerdo.
16. Hago un esfuerzo para parecer consistente a otros.
1 = Totalmente en desacuerdo and 5 = Totalmente de acuerdo.
17. Me incomoda mantener dos creencias que son inconsistentes.
1 = Totalmente en desacuerdo and 5 = Totalmente de acuerdo.
18. No me molesta demasiado si mis acciones no son consistentes. (R)
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 belleza? Si, No

>> [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. Bibliografía

Abidin, C. (2019). Yes Homo: Gay influencers, homonormativity, and queerbaiting on YouTube. *Continuum*, 33(5), 614-629.

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.

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. *Publicazioni 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 Garcia, 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.

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.

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.

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.

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.

Levene, H. (1960). "Robust Tests for Equality of Variances," In: I. Olkin, et al., Eds., *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, Stanford University Press, Palo Alto, 1960, pp. 278-292.

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*, fhal-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

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

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*.