



**UNIVERSIDAD
TORCUATO DI TELLA**

Master in Management & Analytics

**Predicción del cambio en el
comportamiento de los consumidores
de bebidas post pandemia de COVID-19
en Chile**

María del Rosario de la Torre

Licenciada en Administración de Empresas (U.B.A)

Tutor: Guido de Caso

Resumen

Desde la aparición del COVID-19 y la nominación del mismo como pandemia, el mundo entero se ha visto obligado a modificar gran parte de su rutina diaria que había tenido lugar hasta entonces.

Dentro de esas modificaciones, una de las que más ha tenido relevancia, es la del cambio de los hábitos: de consumo, de planificación de sus finanzas, de apoyo psicológico, administración de sus hogares, etc.

Los consumidores, en sí, ya estaban cambiando sus hábitos o costumbres, pero ahora lo están haciendo a una velocidad antes impensada. En este contexto de cambio drástico, las empresas deben estar preparadas para afrontarlo. Y esto requiere que se anticipen al tipo de consumidor que está surgiendo, desarrollando las capacidades necesarias para adaptarse lo más rápido posible.

Dado que aún no existen diversos artículos e investigaciones de este tema, el presente estudio propone un conjunto de herramientas analíticas que puedan servirle a las organizaciones al momento de enfrentarse al problema de adaptación de cambio de los hábitos de consumo “obligatorio”, en el sentido de que los consumidores se vean imposibilitados a conservar sus conductas cotidianas.

Abstract

Since the appearance of COVID-19 and its denomination as a pandemic, the whole world has been forced to modify much of its daily routine that had taken place.

Among these modifications, one that has had the most relevance is the change in habits: consumption, planning their finances, psychological support, managing their homes, etc.

Consumers were already changing their habits, but now they are doing so, at an unthinkable speed. In this context of drastic change, companies must be prepared to face it. To do so, they need to anticipate the type of consumer that is emerging, developing the necessary capacities to adapt and react as quickly as possible.

Given that, there is still not enough articles or research on this subject. This study propose a set of analytical tools that can be useful to organizations when facing the problem of this adapting to “mandatory” change, in consumption habits, which are unable to maintain in their daily behaviors.

Índice

Introducción.....	6
Motivación.....	6
¿Qué es el hábito de consumo?	7
La Pandemia del COVID-19.....	10
Pandemia en Chile	12
El problema.....	13
Objetivo	13
Metodología.....	17
Los Datos/ Dataset	18
Análisis Exploratorio.....	22
Venta en Hectolitros por Grupo de Productos	22
Venta en Hectolitros por Marca y Calibre	25
Venta en Hectolitros por Canal	34
Matriz de Correlación.....	38
Clustering.....	42
Métrica de Similaridad	42
Elección de cantidad de Clusters	43
Grupo de Productos	45
Marcas de Cervezas	46
Conclusión Clustering.....	47
Etapa Predictiva	47
Grupo de Productos	50
Canal	59
Predicción de ventas totales	70
Lineamientos Estratégicos y Plan de Acción: recomendaciones	72
Etapa Prescriptiva: Conclusión	81
Referencias bibliográficas.....	85
Anexo	87

Variables del Dataset	87
Listado final de las variables utilizadas	89
Gráficos adicionales	89
Análisis Exploratorio de Datos Regiones.....	93
Predicción Pernod Ricard	94
Tabla Valor Observado – Predicción por día.....	95
Información adicional Cámara de Comercio de Santiago de Chile.....	98

Introducción

Motivación

La pandemia del coronavirus ha llegado al mundo para poner en jaque diversos aspectos de la vida diaria de las personas. En muchos países, la gente tuvo que cumplir con el confinamiento impuesto o sugerido por los entes públicos a falta de una cura definitiva e inminente.

Este aislamiento, en algunos casos obligatorio y en otros opcional, forzó a las personas a adaptarse a un nuevo ritmo y forma de vida. Una de las cosas que se ha modificado sin duda en la vida de las personas es el hábito de consumo. Se entiende a este último, como el comportamiento que tienen los individuos al realizar una compra dependiendo del lugar donde la efectúan, la frecuencia con la cual realizan la misma, el tipo de producto que consumen y los momentos de compra, entre otros.

Es ya sabido y evidente que las compras Online han tenido un rol muy importante desde que existe el confinamiento. El temor al contagio por aglomeración de gente o contacto con superficies, ignorando el grado de desinfección de éstas, ha causado que mucha gente se vuelque a las compras por Internet. Muchos países y empresas estaban ya preparados para este incremento exponencial y contaban con un canal desarrollado, pero muchos otros no, por lo que tuvieron que adecuarse ágilmente para hacer frente a la gran demanda inesperada.

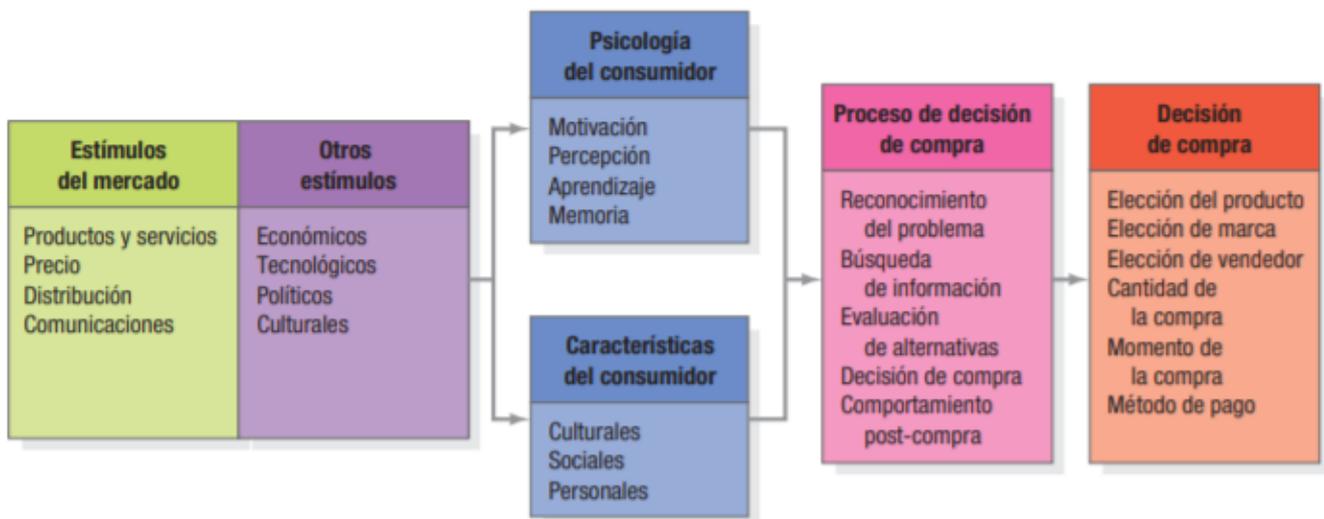
Para lo que nadie está preparado y ninguna persona puede saber de antemano, es cómo será esta nueva forma de hacer las compras. Para muchas personas, cambiará el hábito de consumo. Aquí surgen varios interrogantes como, por ejemplo: ¿Se seguirán realizando las compras a través del canal online o se volverán a hacer compras de forma personal?, ¿las compras consistirán en aprovisionamientos de magnitud para satisfacer las necesidades de varios días o semanas o para saciar deseos del momento?, ¿Los empaques elegidos serán los grandes, pesados o se preferirán artículos de fácil manipulación?, ¿La visibilidad en las góndolas en supermercados o almacenes seguirán siendo tan importante como antes para las empresas de consumo masivo o predominará más bien una estrategia digital para la visibilidad de sus productos en los nuevos días canales?.

Estas y muchas otras cuestiones tienen en vilo a las empresas productoras, ya que de esta forma se cambian las reglas del juego: la forma de producción, de empaquetado, logística de los productos, y mucho más.

¿Qué es el hábito de consumo?

Philip Kotler y Kevin L. Keller en su libro “Marketing Management” analizan el comportamiento de los consumidores e indican que el mismo consiste en una acción realizada debido a algún tipo de estímulo percibido. Partiendo de la idea del “conductismo”, asumen que una compra es el resultado de estímulos externos: estímulos de Marketing y del entorno. Estos estímulos, se cuelan en la conciencia del consumidor generando la decisión de compra a través de la combinación de procesos psicológicos y las características propias de cada individuo.

Kotler y Keller plantean el siguiente modelo de comportamiento de los consumidores:



Describen, además, que existen cuatro procesos psicológicos que influyen en las respuestas del consumidor:

1. Motivación:

Analizan cómo la motivación se convierte en una necesidad a través de 3 teorías y cómo varían las estrategias de marketing de acuerdo a cada una de ellas.

- a. Sigmund Freud: supone que las personas no entienden por completo sus motivaciones debido a que las fuerzas psicológicas que modelan el comportamiento de las mismas son, en su mayoría, inconscientes.
- b. Maslow: parte de la idea de la jerarquización de las necesidades humanas. La conocida Pirámide de Maslow indica que las necesidades más importantes para un individuo son las Fisiológicas (alimento, agua, vivienda), luego las necesidades de seguridad (seguridad, protección), seguido por las necesidades de sociales (sentimiento de pertenencia, amor), necesidades de estima (autoestima, reconocimiento, estatus) y por último las necesidades de autorealización (desarrollo y realización personal). Maslow sugiere que las personas intentarán primero satisfacer sus necesidades más básicas y luego las que sigan en orden de relevancia.
- c. Herzberg: en su teoría, distingue los factores “desmotivadores” (que provocan insatisfacción) y los “motivadores” (que provocan satisfacción). Para que se produzca una compra es necesario que existan factores motivadores y una ausencia de factores desmotivadores.

2. Percepción:

La percepción es un proceso en el que una persona elige, organiza e interpreta la información para formar una imagen del mundo que lo rodea de acuerdo a Berelson y Steiner en su libro *Human Behavior: An Inventory of Scientific Findings*.

Cuando una persona realiza una acción (se encuentra motivada para realizar la misma o dicho acto voluntario), se ve influida por su percepción de la situación. Existen tres procesos que interfieren en el desarrollo de diferentes percepciones ante un mismo objeto:

- a. Atención Selectiva: existe la atención voluntaria e involuntaria. El desafío es explorar y entender qué estímulos notarán las personas.
- b. Distorsión Selectiva: se trata de la predisposición de las personas a interpretar la información para que se ajuste a sus percepciones.
- c. Retención Selectiva: las personas solamente retienen la información que confirma sus actitudes y creencias.

3. Aprendizaje:

Los cambios en el comportamiento son impulsados por el aprendizaje a través de la experiencia. La teoría del aprendizaje indica que el mismo, se produce por la interrelación de impulsos, estímulos, señales, respuestas y refuerzos. Un impulso es un fuerte estímulo que lleva a concretar una acción cualquiera.

Los expertos en marketing toman la teoría del aprendizaje como fundamento para crear necesidades en los individuos e incentivar la demanda de ciertos productos a través de los impulsos. Este objetivo lo logran utilizando señales motivadoras y procurando un refuerzo positivo.

4. Memoria:

Se reconocen dos tipos de memoria: a Corto Plazo (MCP) y a Largo Plazo (MLP). La primera, se caracteriza por ser un reservorio de información temporal y limitado, mientras que la segunda, es principalmente ilimitado y permanente.

Una de las teorías de la MLP indica que la estructura de la misma supone que las personas forman modelos asociativos.

Las empresas utilizan estos tipos de asociaciones vinculadas para motivar el conocimiento de sus marcas. Para esto, se deben asegurar de generar las experiencias necesarias de los individuos para crear las estructuras de conocimiento y así las mantengan en su MLP.

La Pandemia del COVID-19

¿Qué es una Pandemia?

De acuerdo a la Real Academia Española, el término pandemia hace referencia a una enfermedad epidémica que se extiende a muchos países o que ataca a casi todos los individuos de una localidad o región.

Se diferencia de una epidemia o un brote epidémico, ya que la propagación de estos últimos es más acotada (a un país o región en particular) y por tiempo determinado.

¿Qué es un coronavirus?

Los coronavirus son una extensa familia de virus causantes de enfermedades que van desde el resfriado común al síndrome respiratorio agudo severo (SRAS).

Las recomendaciones habituales para no propagar la infección son: la buena higiene de manos y cubrirse la boca y la nariz al toser y estornudar. Asimismo, se debe evitar el contacto estrecho con cualquier persona que presente síntomas, como tos o estornudos.

¿Qué es la COVID-19?

De acuerdo a la Organización Mundial de la Salud (OMS), la COVID-19 es la enfermedad causada por el nuevo coronavirus SARS-CoV-2.

El día 11 de marzo de 2020, la Organización Mundial de la Salud (OMS) anunció que, debido a la rápida propagación y el nivel de gravedad del brote, la COVID-19 era caracterizada como una Pandemia.

Luego de este anuncio, varios países asumieron medidas de prevención, siguiendo las recomendaciones de la OMS:

- Distanciamiento social
- Higienización de manos, ropa y superficies

- Utilización de máscara médica
- Aislamiento de pacientes con síntomas
- Testeos masivos

¿Cómo se propaga la COVID-19?

De acuerdo a lo indicado por la Organización Panamericana de la Salud, el virus se propaga de persona a persona a través de las gotículas que emanan de la nariz o la boca de una persona infectada al toser, estornudar o al hablar. Una persona se puede contagiar si inhala estas gotículas procedentes de la persona infectada por el virus.

Debido a la forma en la que se propaga el virus, las medidas para contener la transmisión masiva, según coinciden entidades académicas y prestigiosos profesionales, es distanciamiento de más de un metro y medio, utilización de mascarilla o barbijo e higienización permanente de las manos.

Es posible que algunas personas infectadas sean asintomáticas, por lo que en cualquier momento conviene que se mantenga distancia física con todas las personas en el caso que nos encontremos en una zona donde puede circular el virus.

¿Cómo evitar la propagación de la COVID-19?

Como se indica en el párrafo anterior, es muy importante que se mantenga distancia, se utilice mascarilla y se higienice bien las manos en todo momento en el que estamos afuera o en contacto con personas que podrían estar infectadas.

Debido a la fácil transmisión que se da con la COVID-19, algunas de las medidas más utilizadas por todo el mundo fue el distanciamiento, aislamiento y cuarentena preventiva y obligatoria en algunos casos.

El distanciamiento físico significa guardar distancia física de las personas. La OMS recomienda que la distancia debe ser de por lo menos un metro y sugiere que se tome esta medida incluso si la persona no presenta síntomas ni ha tenido exposición conocida a la COVID-19.

El aislamiento supone separar a las personas que están infectadas o que presentan síntomas de COVID-19 para evitar que propaguen la enfermedad.

La cuarentena es la restricción de las actividades o el apartamiento de las personas que no están necesariamente enfermas pero que pudieron haber contraído el virus. El objetivo de la cuarentena es prevenir la propagación de la enfermedad en el momento en el que las personas comienzan a tener síntomas o consideran que pudieron ser infectados.

Pandemia en Chile

Como en la mayoría de los países de Latinoamérica, las noticias por COVID-19 comenzaron a resonar cada vez con mayor fuerza a fines del primer trimestre del 2020.

La OMS aún no había decretado la dispersión del virus como pandemia, cuando el 3 de marzo de 2020 arribó a la capital de Chile un médico que había estado de viaje por el sudeste asiático. Poco y nada se sabía de este virus y en el nuevo continente la opción de “importación” del virus parecía casi imposible. Aún así, el presidente Sebastián Piñera, anuncia un Plan de Acción para enfrentar el brote. Este plan, incluye la adquisición de nuevo equipamiento, refuerzo de la red de salud, controles migratorios y el despliegue de un operativo de detección y tratamiento.

Desde que se confirmó ese primer caso, el ritmo de propagación del virus no dio tregua y comenzó a expandirse por las 16 regiones del país. El día 18 de marzo, con ya 104 casos confirmados, el presidente decretó el estado de catástrofe en todo el territorio chileno por 90 días. El mismo se extendió luego por 180, 270 y 360 días en las diferentes regiones.

El país trasandino atravesó el peor momento de la pandemia en el mes de junio del 2020, con un pico de 6.938 casos confirmados y 195 fallecidos diarios.

En noviembre del 2020 se comenzó a percibir los resultados de las medidas tomadas y los casos seguían decreciendo, por lo que las medidas de confinamiento fueron reducidas en diciembre para el período vacacional. El 24 de diciembre se lanza el plan nacional de vacunación y llegan al país el

primer cargamento de vacunas del laboratorio Pfizer para dar inicio al plan de vacunación para el personal de salud y esenciales.

El 3 de febrero de 2021, se comienza el plan de vacunación masiva al mismo tiempo que los casos comenzaban a crecer nuevamente y vislumbraban una segunda ola de contagios. Ante este panorama, el país decide cerrar las fronteras nuevamente a partir del primero de abril y decretar un nuevo confinamiento más estricto rigiendo sobre un 83% de la población.

El plan de vacunación de Chile se ve en todo el mundo como un caso de éxito. Este se debe a principalmente tres factores:

- Posibilidad financiera para la compra de vacunas
- Capacidad logística para la distribución de las mismas
- Capacidad institucional y estructura para implementar un plan de vacunación efectivo

El objetivo principal del gobierno chileno es vacunar al 80% de su población para julio del 2021 y generar así inmunidad de rebaño.

El problema

Los cambios en los hábitos de consumo se dan permanentemente por avances de la tecnología, cambios de modas y muchos otros asuntos que hoy en día no le damos tanta relevancia.

Los cambios en las costumbres de los consumidores son difíciles de predecir cuando se dan en respuesta a nuevas tendencias, desarrollos o descubrimientos ya que, en la mayoría de los casos, no contamos con información histórica que replique las mismas condiciones ante un nuevo evento. En cambio, ante una catástrofe natural (terremoto, tsunami, alud), demográfica (guerras) o epidemiológica, para lamentación nuestro, sí se puede predecir o pronosticar aproximadamente cuál será el comportamiento de un grupo de personas o una sociedad entera, dependiendo del lugar y momento donde se dé el siniestro.

Objetivo

En base a lo expuesto anteriormente, las empresas necesitan saber y conocer con anticipación cómo cambiarán los hábitos de los consumidores para así adaptarse a la nueva demanda. Este entendimiento debe ser muy amplio, en el sentido de abarcar todas las variantes respecto al cliente objetivo al cual se dirigen.

En consecuencia, el principal objetivo de este trabajo es desarrollar un nuevo enfoque que ayude a las empresas a la detección temprana de cambios de consumo de sus clientes, derivados de una modificación en sus estilos de vida y alteración de prioridades.

Se propone un conjunto de herramientas que ayudará a las empresas a prever estos cambios. Concretamente, se crearon algoritmos de clustering basados en la venta histórica de los años 2018, 2019 y 2020 de una de las cerveceras más grandes de Chile. De esta manera se pudieron detectar patrones característicos consumo (o cambios del mismo) a principios del 2020, cuando comenzó a propagarse el virus. Con esto, se pueden analizar todos los comportamientos de los compradores para luego formar grupos representativos que nos ayudará a aplicar una herramienta de predicción.

Debido a la poca información que se tiene de los efectos de la pandemia de la COVID-19, existen pocos análisis que revelen los cambios en los hábitos de consumo durante la misma. Uno de los más interesantes que pudimos encontrar es de la consultora EY *“Future Consumer Index: cómo COVID-19 está cambiando los comportamientos del consumidor”* por Kristina Rogers, que se redactó al mes de haberse decretado la pandemia mundial. En el mismo, se reafirma que el COVID-19 aceleró un proceso que ya se venía desarrollando y estaba evolucionando.

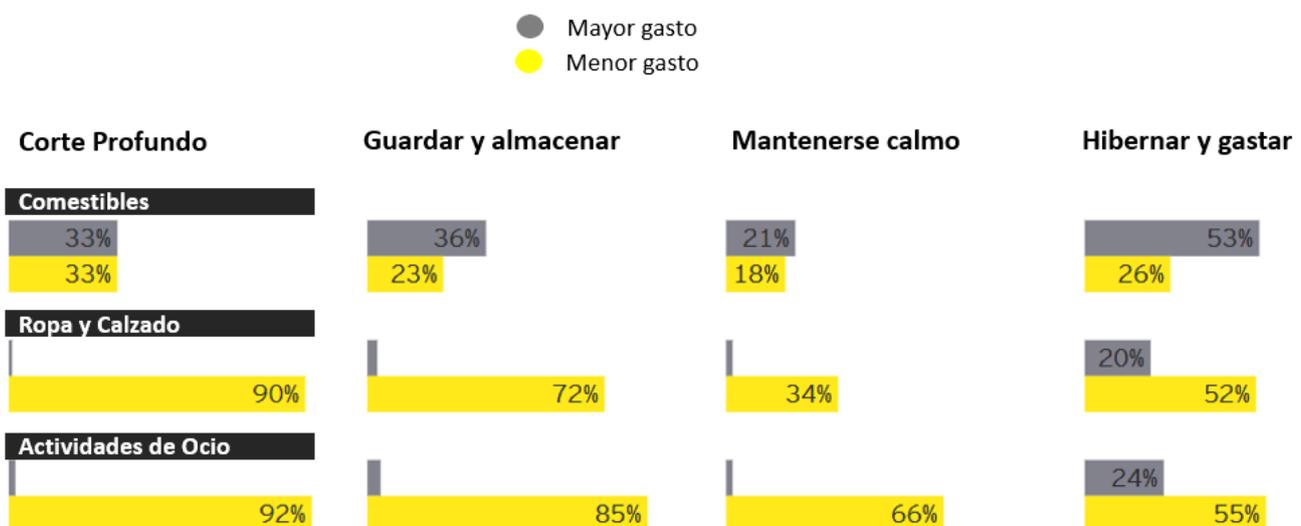
En este análisis, se identificaron cuatro segmentos de acuerdo al comportamiento de los consumidores, que emergieron en los primeros meses de la pandemia:

1. Guardar y almacenar (35%): consumidores que no estaban tan preocupados por la pandemia, pero sí por sus familias. Éstos son pesimistas cuando se les consulta acerca de los efectos a largo plazo.
2. Corte profundo (27%): es el grupo más golpeado por la pandemia. Muy pesimistas sobre los efectos a largo plazo. Disminuyeron los gastos en todas las categorías. La mayor parte de este grupo son consumidores de más de 45 años.

3. Mantenerse calmo (26%): consumidores que no modificaron sus hábitos de gastos. No se vieron muy afectados por la pandemia, sino más bien preocupados por el comportamiento de acopio de gran parte de la sociedad.
4. Hibernar y gastar (11%): preocupados por la pandemia, pero mejor posicionados para atravesarla. Son optimistas de cara al futuro. Gastan más en todos los ámbitos. La edad de la mayoría de este segmento es de entre 18 y 44 años.

Una vez reveladas estas categorías, le hicieron la siguiente pregunta para entender un poco más el comportamiento de los consumidores:

- ¿Cómo se vio afectada la manera en que gastan el dinero en las categorías de comestibles, ropa y actividades de ocio? Las respuestas arrojaron los siguientes datos:



Un 40% de los consumidores del segmento “Hibernar y gastar” indican que compran con menos frecuencia. El 42% dice que los productos que compran han cambiado y el 46% que las marcas ahora son más importantes que antes.

El segmento que ha mostrado más cambios en las distintas categorías es el de “Corte profundo”. Se ve una clara tendencia al menor gasto, sobre todo en las categorías menos esenciales como Ropa y calzado y Actividades de ocio. De acuerdo al reporte, este comportamiento se debe a que casi un cuarto de las personas que componen este segmento ha tenido suspendidos sus empleos, de forma temporal y permanente. El 78% de este

conjunto, está comprando con menos frecuencia y el 64% sólo compra lo esencial. Además, el 33% de estos consumidores, sienten que las marcas son menos importantes que antes.

Proyectando un poco cómo seguirán modificándose los hábitos de consumo, también se les ha consultado a los “categorizados” si esperan volver a sus antiguos comportamientos en un futuro próximo. Muchos de los mismos respondieron que es poco probable.

En línea con la última pregunta, notaron que post COVID-19, podrían surgir 5 nuevos segmentos a los cuales migrarán cada uno de los comentados anteriormente:

1. Vuelta a la normalidad (31%): gastos sin cambios. La vida cotidiana no se ha visto realmente afectada. Menos preocupados por la pandemia en general. La mayor parte de este segmento pertenecía al de “Mantenerse calmo”
2. Cautelosamente extravagante (25%): ingresos medios a altos. Centrado en la salud, pero optimista, a pesar de la gran recesión mundial que se vaticina. Mayores gastos en categorías más importantes para ellos. La mayor parte de este segmento pertenecía al de “Guardar y almacenar”
3. Mantenerse mensurado (22%): menores gastos y recortes profundos. Intentando establecerse nuevamente. Son más pesimistas sobre el futuro. La mayor parte de este segmento pertenecía al de “Guardar y almacenar”
4. Corte Continuo (13%): menos educación y trabajo en este segmento. Grandes cortes de gastos, cambio en las formas y tipos de compras. La pandemia es una preocupación. La mayor parte de este segmento pertenecía al de “Corte profundo”
5. Vuelta explosiva (9%): jóvenes y con trabajo. Mayores gastos en todas las categorías. Su vida cotidiana es más disruptiva. Son los más optimistas. La mayor parte de este segmento pertenecía al de “Hibernar y gastar”

Por último, se analizan en detalle los comportamientos del grupo 1 y 2 “Vuelta a la normalidad” y “Cautelosamente extravagante” y se consulta acerca de las categorías en las que se espera gastar más dinero una vez que

termine la pandemia. El grupo número 2 tiene una mayor predisposición a gastar mucho más que el grupo número 1, sobre todo en las primeras 4 categorías: Actividades de ocio, Vacaciones, Comidas en restaurantes y Compra de Ropa y calzado. En lo que respecta al interés de este trabajo, se ve que hay una mayor propensión del segmento “Cautelosamente extravagante” (un 21%) a gastar más en bebidas alcohólicas, en comparación con el 1% del segmento “Vuelta a la normalidad”.

De una forma u otra, las empresas deberán investigar y reconocer estos cambios ya que las ayudará a contar con una ventaja y a planificar el futuro con mayor certeza. Así, no sólo deberán satisfacer las expectativas de los consumidores actuales, sino también anticipar sus necesidades futuras.

Metodología

Para desarrollar el presente trabajo y ofrecer una herramienta útil a las empresas, en particular de consumo masivo, la metodología implementada constó de cuatro etapas:

- *Preparación de los datos:* los dataset provistos contienen información real de ventas de una compañía de consumo masivo del rubro bebidas. Los mismos tuvieron que ser trabajados, en un principio, de forma manual ya que cada archivo (con la información de ventas de cada año analizado <2018, 2019 y 2020>) contenía más de 3 millones de líneas cada uno y 33 columnas. El trabajo manual consistió en separar mes por mes el .CSV en diferentes solapas para que, al leerlo con el código, se puedan visualizar todas las líneas sin perder ningún dato. Luego de esto, se procedió a la carga y la concatenación de toda la información por año.
- *EDA (Análisis Exploratorio de Datos):* para poder comprender los datos con los que se trabajará, en primer lugar, se deben explorar, describir, analizar y visualizar la naturaleza de los mismos. Una vez concatenados los archivos con la información de los 3 años, se fueron seleccionando algunas columnas de interés para aplicar técnicas simples de resumen de datos y gráficos que nos llevó a observar algunos comportamientos generales. Luego, se graficaron dos

matrices de correlación para entender la relación entre diferentes *features*. Por último, se utilizaron algoritmos de clustering en series de tiempo para obtener otros insights acerca de las preferencias de consumo utilizando la métrica Dynamic Time Warping (DTW) y Suma de Cuadrados residuales o errores (SSE) y Coeficiente de Shilouette para obtener la cantidad de clusters a aplicar.

- *Etapa Predictiva:* con la información obtenida del Análisis Exploratorio de Datos, se trabajó para poder anticipar el comportamiento de los consumidores ante un evento fortuito que haga cambiar sus hábitos forzosamente (como sucedió con la pandemia de la COVID-19). En esta etapa se utilizó Prophet, un modelo de series de tiempo desarrollado por Facebook, óptimo para predecir futuros valores de la serie analizada.
- *Etapa Prescriptiva:* en esta última etapa, se realiza un análisis de los resultados provistos en las etapas previas para poder realizar las recomendaciones pertinentes a la empresa.

Los Datos/ Dataset

Los datos utilizados en este trabajo fueron provistos por una de principales compañías de bebidas más grandes de Latinoamérica. Se trata de Compañía Industrial Cervecera (CCU), líder en todas las categorías en las que participa incluyendo cervezas, gaseosas, aguas minerales, jugos y vinos entre otros.

Los dataset provistos, contienen información de todas las ventas de la compañía de los años 2018, 2019 y 2020 de la operación en Chile. La información completa de cada año se estructura en un archivo .CSV con 33 columnas y más de 3 millones de filas cada uno. Los datos se presentan en una tabla dinámica, la cual hay que desglosar para poder conocer toda la información que contiene.

Al querer visualizar la información de cada año en el comando, se observa que no se muestra el total de las líneas, por lo que se resuelve reformar el

archivo manualmente dividiendo la información de cada mes en una solapa distinta. Una vez hecho esto con los 3 archivos de cada año, se ubica en un Dropbox y luego se leen desde allí, concatenando todos los años para formar un solo dataset.

Se analiza el tipo de información que tiene cada columna y se explica en el anexo qué es cada una:

```
Index(['ANIO', 'MES', 'TIPO_VENTA', 'OFICINA', 'DESCR_OFICINA', 'DISTRITO',
      'ZONA', 'SKU', 'GRUPO_PROD', 'TIPO', 'LICENCIA', 'CATEGORIA',
      'SUBCATEGORIA', 'MARCAVENTAPAIS', 'AGR_MARCA', 'DES_MARCA',
      'AGR_EMPAQUE', 'DES_EMPAQUE', 'RETORNABILIDAD', 'TAMANIO', 'OCASION',
      'CANAL', 'NIVEL', 'RELEVANCIA', 'FLAG_SKU_PPM', 'VENTA_HL', 'VENTA_CJ',
      'VTA_BRUTA', 'DCTO', 'FLETE', 'COSTO', 'IMP_G', 'IMP_E'],
      dtype='object')
```

```
--- -----
0 ANIO int64
1 MES int64
2 ID_EMPRESA int64
3 TIPO_VENTA object
4 OFICINA int64
5 DESCR_OFICINA object
6 DISTRITO object
7 ZONA object
8 SKU int64
9 GRUPO_PROD object
10 TIPO object
11 LICENCIA object
12 CATEGORIA object
13 SUBCATEGORIA object
14 MARCAVENTAPAIS object
15 AGR_MARCA object
16 DES_MARCA object
17 AGR_EMPAQUE object
18 DES_EMPAQUE object
19 RETORNABILIDAD object
20 TAMANIO object
21 OCASION object
22 CANAL object
23 NIVEL object
24 RELEVANCIA float64
25 FLAG_SKU_PPM int64
26 VENTA_HL float64
27 VENTA_CJ float64
28 VTA_BRUTA float64
29 DCTO float64
30 FLETE float64
31 COSTO float64
32 IMP_G float64
33 IMP_E float64
dtypes: float64(9), int64(6), object(19)
```

Luego de analizar todos los datos, se seleccionan solamente las columnas con las que vamos a trabajar y que nos sirve para hacer los análisis previstos comentados en el apartado anterior. Se tomó esta decisión ya que la cantidad de líneas y columnas tornaban al dataset un cúmulo de información difícil de manipular por el espacio en memoria.

A continuación, se detalla algunos de los datos más relevantes que incluye el dataset y que se utilizaron principalmente para hacer el Análisis Exploratorio de Datos:

- ANIO/MES: indica el año y mes en el que se realizó la venta

- **DISTRITO:** contiene la información de todas las regiones del país y además detalla algunos canales que se toman como “zona”
- **GRUPO_PROD:** cuenta con la información del tipo de producto
 - Analcohólicas
 - Cervezas
 - Confites
 - Jugos
 - Licores
 - Otros
 - Pernod Ricard (CCU realiza la distribución y venta de estos productos)
 - Vinos
- **AGR_MARCA:** cuenta con la marca de cada uno de los productos.
- **AGR_EMPAQUE:** cuenta con la información del calibre de cada uno de los productos. Este dependerá también del grupo de productos y de la marca. CCU produce y comercializa más 15 tipos de empaques.
- **CANAL:** detalla el medio o sector en el que se realiza la venta
 - Consumo: refiere a todas las ocasiones en las que se consume el producto en el local que lo administra y ofrece, como por ejemplo: restaurantes, discoteques, bares, etc.
 - Compra: hace referencia a las ventas que se realizan en almacenes, licorerías, etc. Donde el cliente compra y se lo lleva a su hogar
 - Supermercado
 - Mayorista
 - E-Commerce: venta online
- **VENTA_HL:** total en hectolitros por línea. En la industria de bebidas la venta en hectolitros es el índice principal que se observa, con el cual realizan previsiones y planean los objetivos a mediano y largo plazo.

Una vez que contamos con todos los datos de los años (2018, 2019 y 2020) y las columnas seleccionadas para analizar, visualizamos el dataset completo para confirmar la cantidad de datos contenidos en el mismo:

Tabla 1: GRUPO_PROD

	ANIO	MES	GRUPO_PROD	VENTA_HL
0	2018	1	ANALCOHOLICAS	6.84
1	2018	1	CONFITES	1.80
2	2018	1	ANALCOHOLICAS	0.00
3	2018	1	ANALCOHOLICAS	0.36
4	2018	1	ANALCOHOLICAS	0.00
...
20693609	2020	12	VINOS	0.00
20693610	2020	12	JUGOS	1.00
20693611	2020	12	VINOS	0.00
20693612	2020	12	CERVEZAS	0.00
20693613	2020	12	CERVEZAS	0.00

20693614 rows × 4 columns

Tabla 2: AGR MARCA y AGR EMPAQUE

	ANIO	MES	GRUPO_PROD	AGR_EMPAQUE	AGR_MARCA	VENTA_HL
0	2018	1	ANALCOHOLICAS	PRB 2LT	PAP	6.84
1	2018	1	CONFITES	ESTUCHE 600 GR	GALLETON	1.80
2	2018	1	ANALCOHOLICAS	PET 500	GATORADE	0.00
3	2018	1	ANALCOHOLICAS	PET 1,5LT	PEPSI ZERO	0.36
4	2018	1	ANALCOHOLICAS	PET 250	PEPSI	0.00
...
20693609	2020	12	VINOS	GNB 750	CASTILLO DE MOLINA	0.00
20693610	2020	12	JUGOS	1 LITRO	VIVO	1.00
20693611	2020	12	VINOS	CARTON	GATO	0.00
20693612	2020	12	CERVEZAS	LN 355	ESCUDO	0.00
20693613	2020	12	CERVEZAS	LN 355	AUSTRAL	0.00

20693614 rows × 6 columns

Tabla 3: CANAL

	ANIO	MES	CANAL	VENTA_HL
0	2018	1	COMPRA	6.84
1	2018	1	COMPRA	1.80
2	2018	1	CONSUMO	0.00
3	2018	1	SUPERMERCADO	0.36
4	2018	1	MAYORISTA	0.00
...
20693609	2020	12	SUPERMERCADO	0.00
20693610	2020	12	SUPERMERCADO	1.00
20693611	2020	12	MAYORISTA	0.00
20693612	2020	12	CONSUMO	0.00
20693613	2020	12	COMPRA	0.00

20693614 rows × 4 columns

Tabla 4: DISTRITO

	ANIO	MES	DISTRITO	VENTA_HL
0	2018	1	ANTOFAGASTA/CALAMA	6.84
1	2018	1	ANTOFAGASTA/CALAMA	1.80
2	2018	1	ANTOFAGASTA/CALAMA	0.00
3	2018	1	ANTOFAGASTA/CALAMA	0.36
4	2018	1	ANTOFAGASTA/CALAMA	0.00
...
20693609	2020	12	OSOR/PTO MONTT/CASTR/COYH	0.00
20693610	2020	12	OSOR/PTO MONTT/CASTR/COYH	1.00
20693611	2020	12	OSOR/PTO MONTT/CASTR/COYH	0.00
20693612	2020	12	OSOR/PTO MONTT/CASTR/COYH	0.00
20693613	2020	12	OSOR/PTO MONTT/CASTR/COYH	0.00

20693614 rows x 4 columns

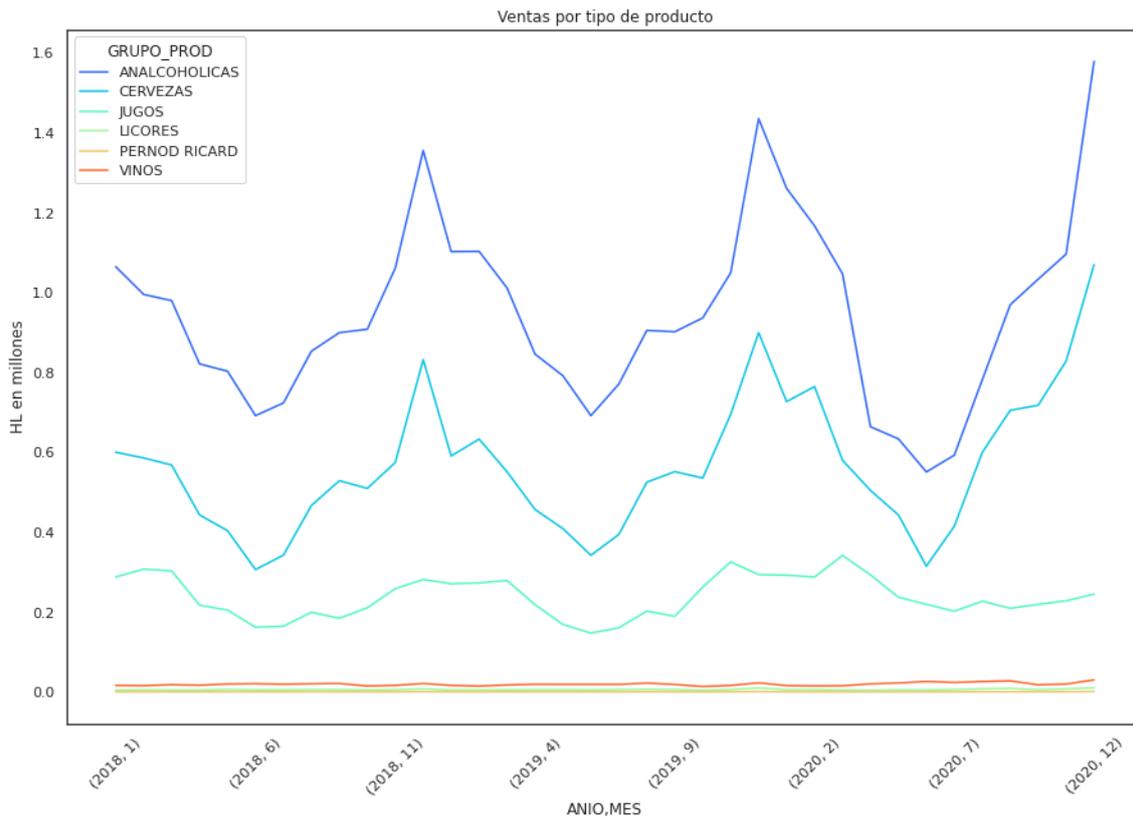
Análisis Exploratorio

El objetivo de este Análisis es explorar, describir, resumir y visualizar la naturaleza de los datos mediante el empleo de técnicas de resumen de datos y métodos gráficos.

Además de los análisis esenciales, se utilizaron también algoritmos de clustering y se aplicaron matrices de correlación entre distintos features, para poder obtener otra perspectiva de la información.

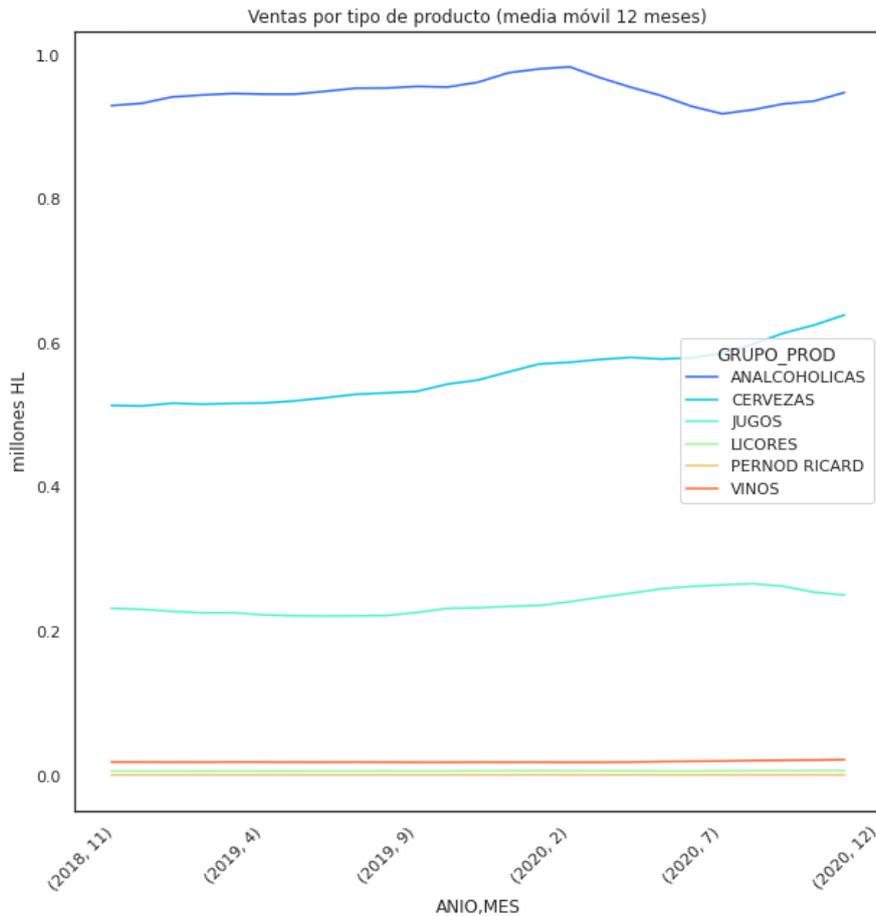
Venta en Hectolitros por Grupo de Productos

Se eliminó la variable “Confites” y “Otros” ya que ambas tenían poca información y no aportaban valor a nuestro análisis. De hecho, se ve que la venta de “Confites” se discontinúa en a fines del 2018. Una vez eliminados esos grupos de producto, realizamos un gráfico para poder observar mejor el comportamiento de esta variable:



A priori, en este gráfico se puede ver que la categoría con mayor venta es la de *analcohólicos*, seguido por *cerveza* y luego *jugos*. Se observa una tendencia a la baja en el segundo trimestre del 2020, cuando hubo un pico de casos de COVID-19 y desde el gobierno Nacional se reforzaron las medidas de confinamiento en la mayoría de las regiones. Esta caída se ve más pronunciada en la categoría de cervezas, con una mejor recuperación que las analcohólicas y jugos luego de agosto del 2020. Respecto a la categoría *jugos*, vemos que esta se mantiene bastante estable en comparación a las oscilaciones y bruscos ascensos y descensos de las otras categorías más vendidas. Se asocia este comportamiento a que jugos no tiene una curva de estacionalidad tan marcada como las cervezas, y además el coeficiente de correlación de las tres primeras categorías varía considerablemente, como veremos más adelante.

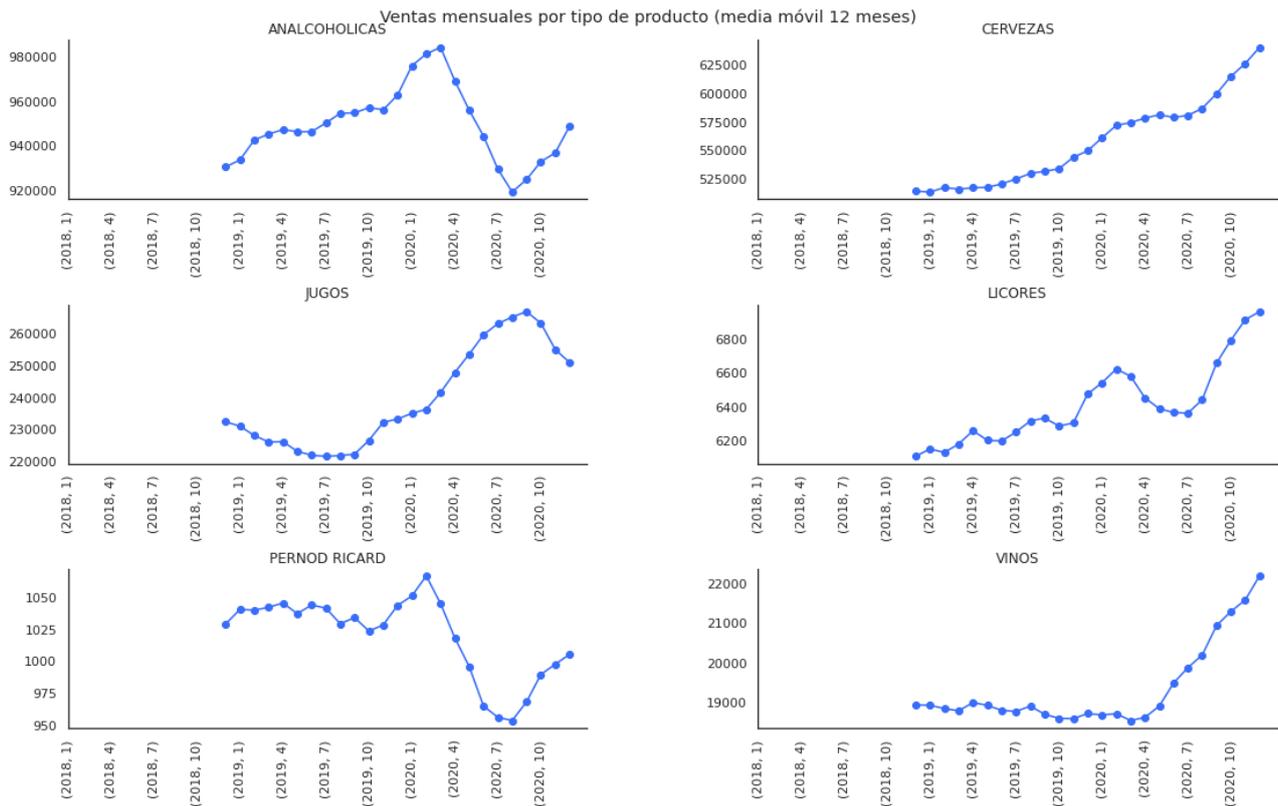
Teniendo en cuenta que en esta industria la estacionalidad de cada producto o categoría juega un rol importante, creemos que lo mejor es visualizar y analizar las medias móviles (12 meses) para suavizar estas fluctuaciones y remover esos efectos, aun incurriendo en el costo de perder las primeras 11 observaciones:



Con este detalle, puede observarse la tendencia creciente en la venta de cervezas a lo largo del 2020 con un mayor impulso a partir de agosto de ese año. A diferencia de la venta de bebidas analcoholicas, que presentan una caída muy marcada en el tercer trimestre del 2020 y luego una lenta recuperación hacia fines de año.

El comportamiento mencionado arriba puede explicarse principalmente por la baja de casos de COVID-19 y la relajación de medidas. La gente, agotada del encierro y con la llegada de un mejor clima, también se distendió y se volcó al consumo de cerveza con la excusa de “celebración”.

Asimismo, para poder validar la anterior conclusión, resulta conveniente visualizar cada serie por separado, dado que la escala de las bebidas analcohólicas y las cervezas (que tienen el mayor porcentaje de ventas) minimizan la dinámica de las otras categorías (jugos, vinos, licores y pernod ricard):



Observando las otras categorías con menor volumen de venta, se ratifica el aumento y tendencia creciente de las bebidas alcohólicas (licores, vinos y pernod ricard), al mismo tiempo que decaen las ventas en los jugos en el último trimestre del 2020.

Venta en Hectolitros por Marca y Calibre

Del mismo modo en que se examinó la información del dataset por grupo de productos, llevaremos a cabo un análisis del resto de las variables, pero concentrándonos en la categoría de Cervezas ya que fue la que mayor variación tuvo en toda la serie de tiempo.

Marca

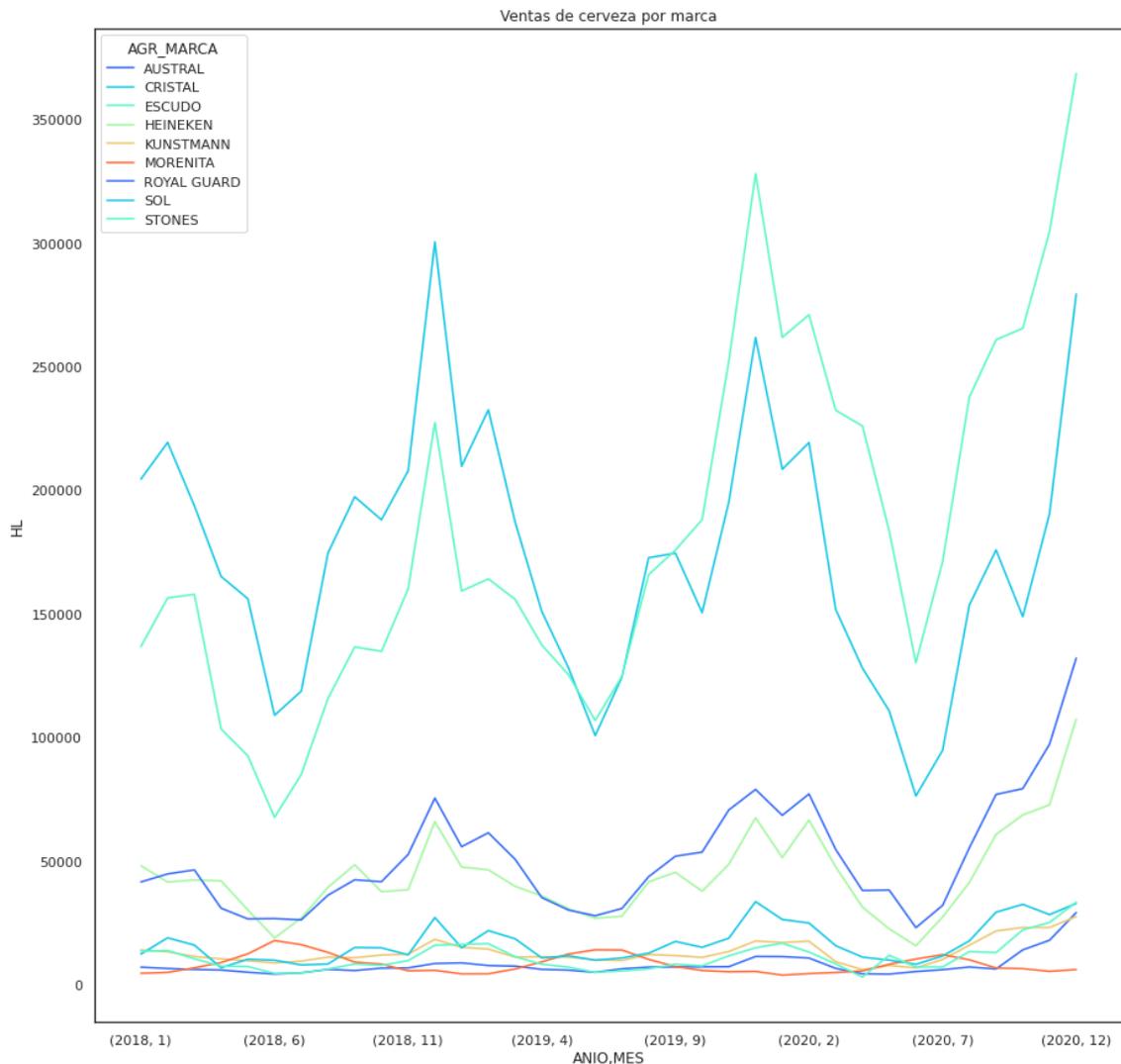
En primer lugar, se visualiza y se hace un conteo de toda la información contenida en el feature “AGR_MARCA” para detectar si existen datos que no sean informativos o que no aporten valor al análisis:

CRISTAL	658299
KUNSTMANN	531573
ESCUDO	474076
HEINEKEN	353739
ROYAL GUARD	333363
AUSTRAL	298488
STONES	175984
MORENITA	133682
SOL	130009
COORS	72118
DOLBEK	62728
DORADA	55775
GUAYACAN	29910
ANDES	21434
IMPERIAL	20609
SZOT	14892
TECATE	8495
BLUE MOON	7975
BAVARIA	1226
BUDWEISER	23
BIRRA MORETTI	4
EDELWEISS	1

Name: AGR_MARCA, dtype: int64

Se identifican las marcas con menos de 100.000 observaciones y se eliminan del análisis para que el dataset esté más “limpio” y podamos detectar patrones interesantes.

Asimismo, para poder analizar mejor la información, nos enfocamos en las 9 marcas más relevantes y luego se procede a graficar la evolución de las mismas para entender cómo se comportan:

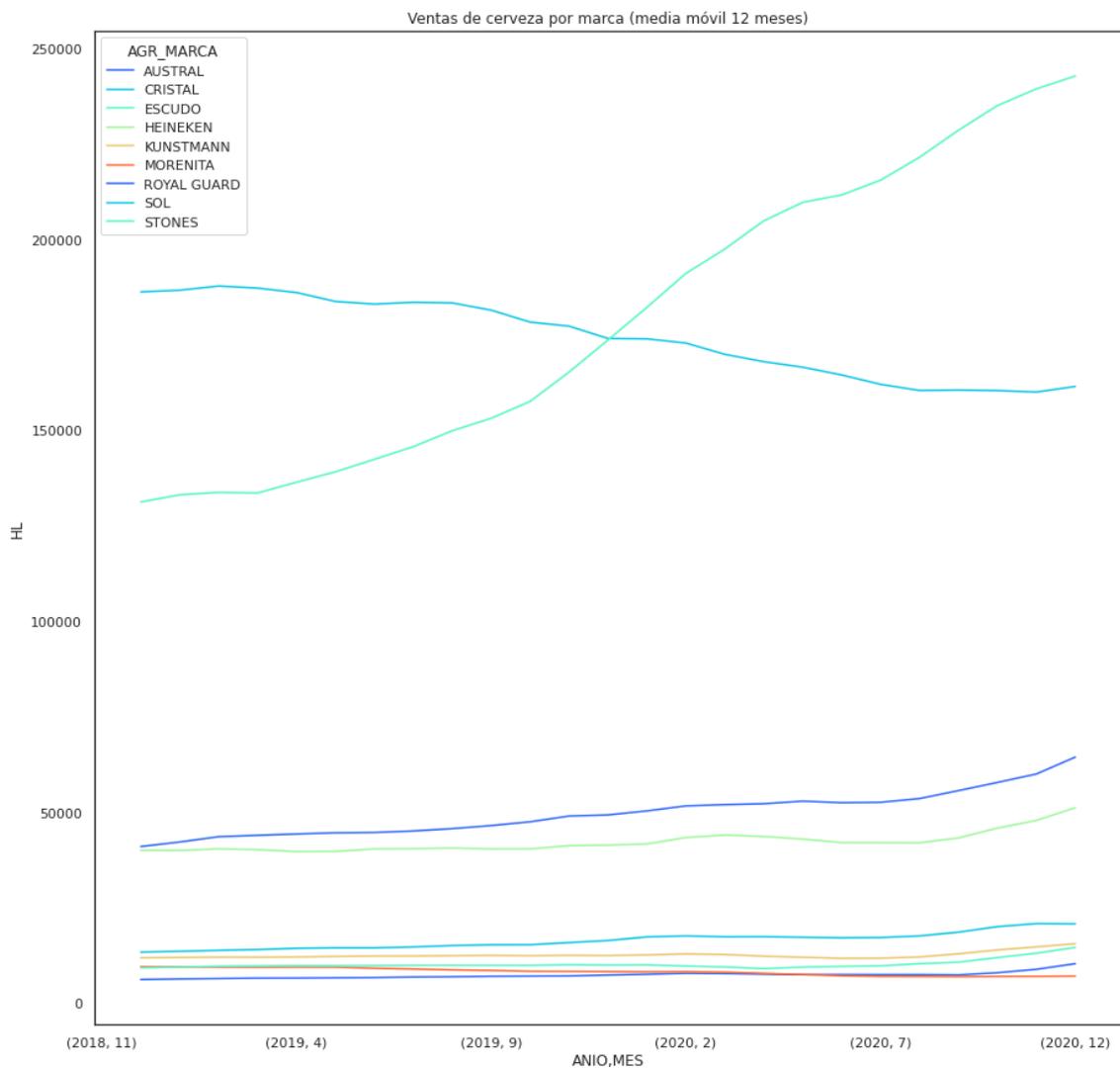


Se observa que las marcas que más se consumen son principalmente Cristal y Escudo, ambas mainstream de CCU Chile, con un liderazgo muy disputado e históricamente, Cristal siempre a la cabeza. De hecho, hasta mediados del 2019 la marca líder era Cristal hasta que Escudo la desplazó y tomó el primer puesto con la mayor cantidad de ventas, despegando y diferenciándose del resto de las marcas para fines del 2020. Este cambio de preferencia, se dio por un nuevo lanzamiento de la marca Escudo: Escudo “Silver”, buscando impactar y atraer a un consumidor más joven, con un perfil de cerveza más liviano, en contraposición con la Escudo “Original” que es más fuerte y robusta. Asimismo, la estrategia de precio de la nueva Escudo era más agresiva, buscando competir directamente con la marca mainstream de la competencia (Becker de AB InBev). CCU Chile armó una gran campaña alrededor de este lanzamiento, con el mayor éxito en el canal Compra, con una activación de “2 Escudo Silver x <luca>”: \$1.000 pesos chilenos que es un billete o solo dos monedas.

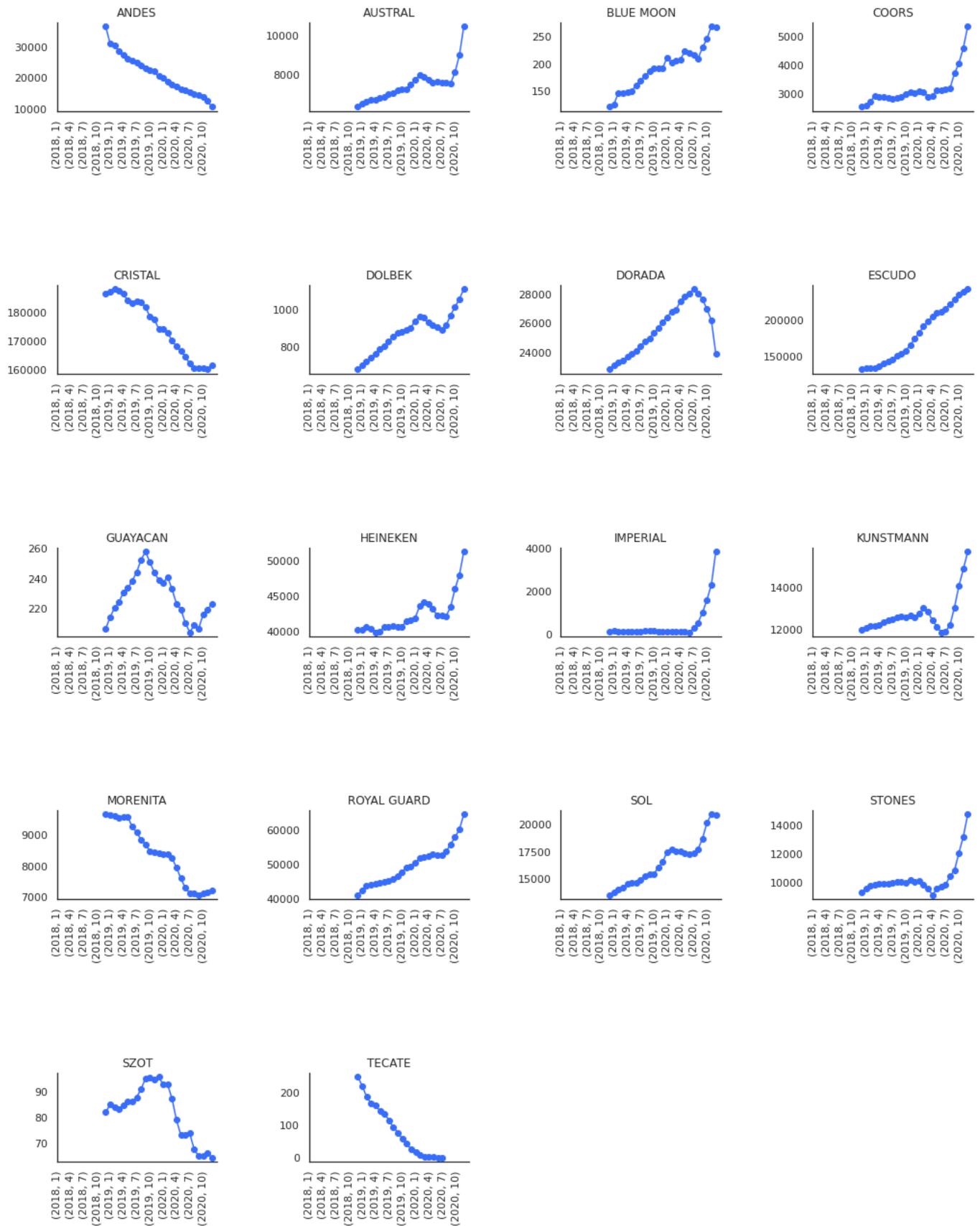
Del mismo modo, pero en una escala más abajo, vemos a las marcas Heineken y Royal Guard disputándose el liderazgo de ese rango. Para fines del 2020 el tercer puesto se lo llevaba Royal Guard.

De esta forma, notamos una mayor preferencia y favoritismo por las marcas Nacionales antes de las internacionales. Veremos más en detalle la relación entre las marcas con la matriz de correlación.

Análogamente como se hizo con el Tipo de Producto, aplicaremos la media móvil al análisis de las marcas para observar las tendencias sin los efectos de la estacionalidad y luego visualizaremos cada serie por separado.



Ventas mensuales por marca de cerveza (media móvil 12 meses)



Tal como se conjeturaba, la marca Escudo se convertía en líder del mercado chileno de cervezas para fines del 2019 con su nuevo lanzamiento, seguida por Cristal, histórica líder en ventas, y Royal Guard, otra cerveza nacional muy consumida por el público chileno.

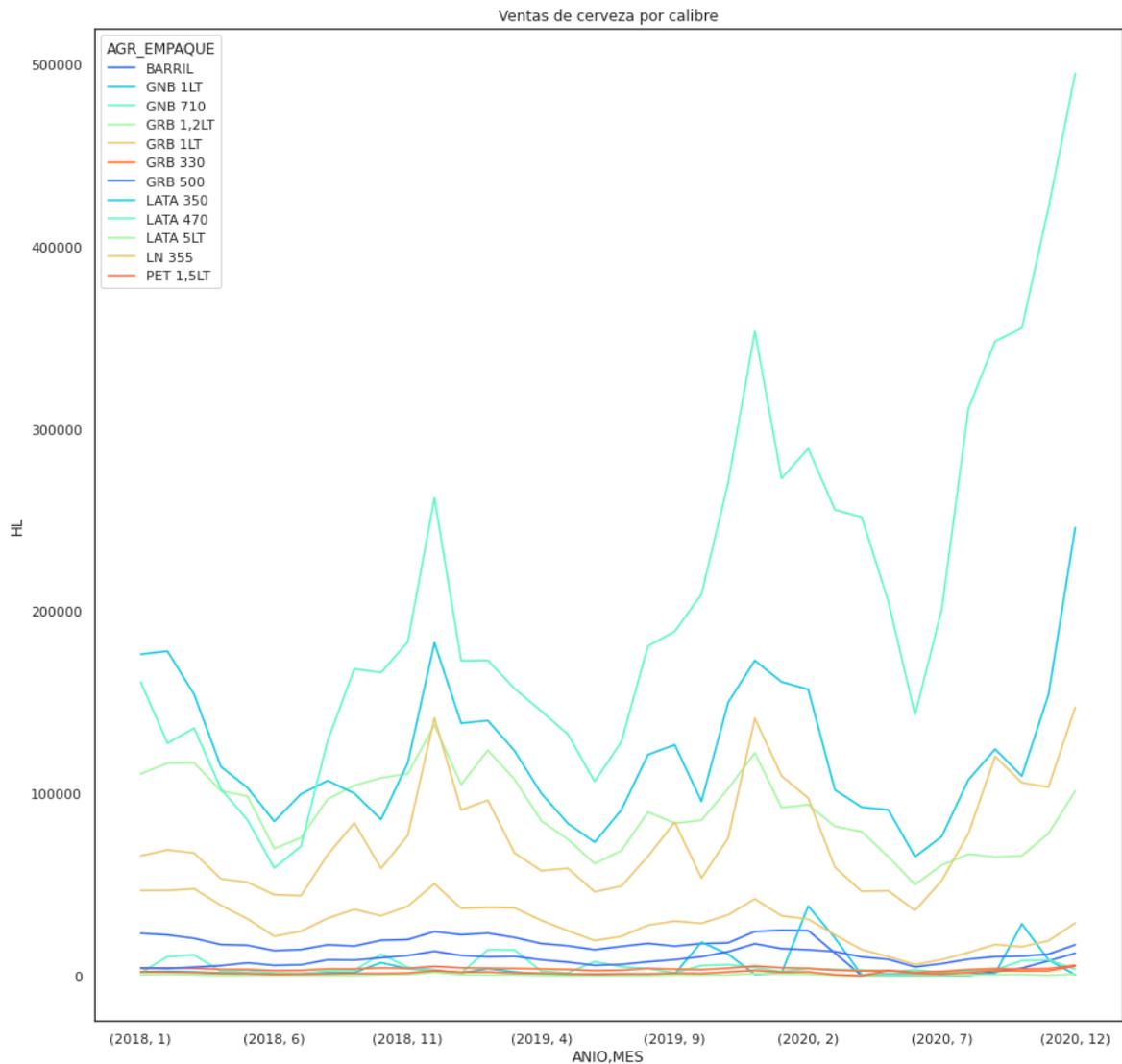
Calibre

Analizamos la información, ahora, por calibre concentrándonos sólo en la categoría Cervezas.

```
LN 355      1502804
LATA 350    650435
LATA 470    333284
GRB 1LT     236365
GRB 1,2LT  226625
BARRIL     167922
GRB 330     77210
GNB 1LT     73995
GNB 710     36652
GRB 500     32686
PET 1,5LT  29056
LATA 5LT    13053
LATA 269    3874
GNB 700     362
LATA 355     56
GNB 750     24
Name: AGR_EMPAQUE, dtype: int64
```

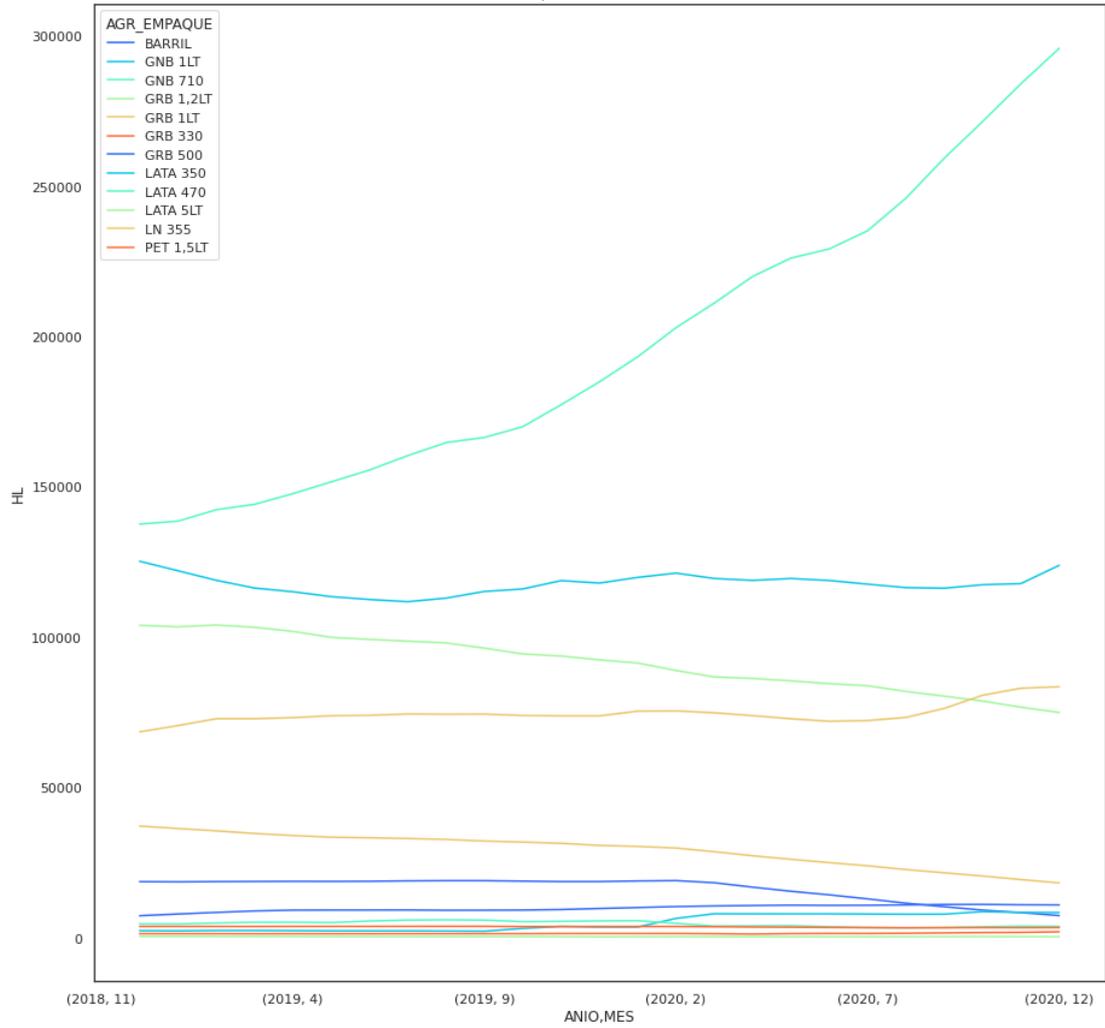
Eliminamos los calibres que tienen menos de 10.000 observaciones para poder realizar un análisis más representativo.

Del mismo modo en que se analizaron la:
gráficos de la data de calibres:

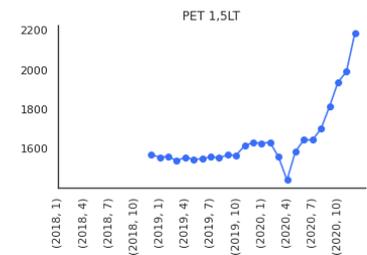
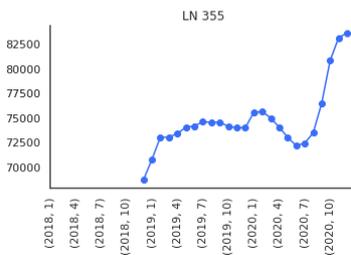
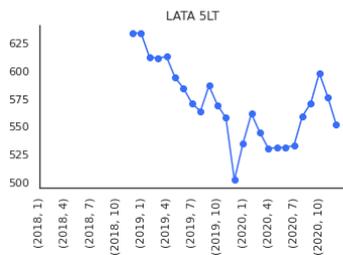
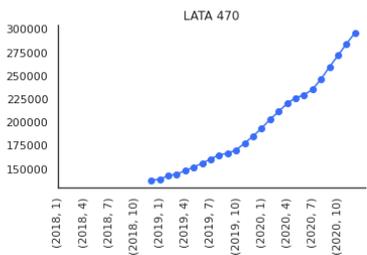
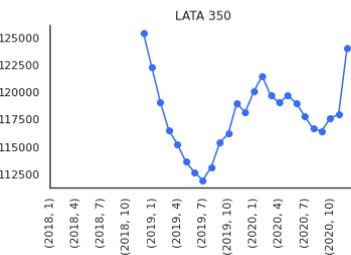
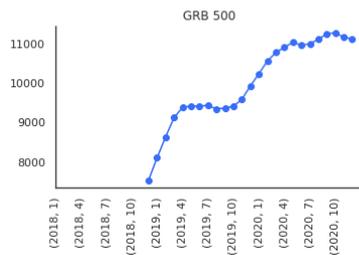
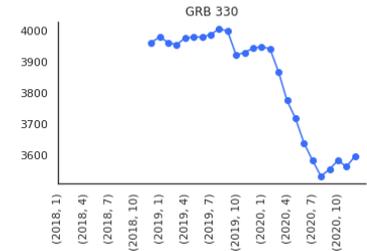
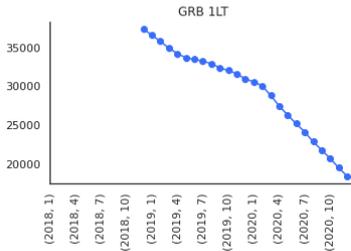
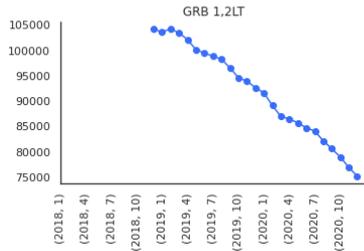
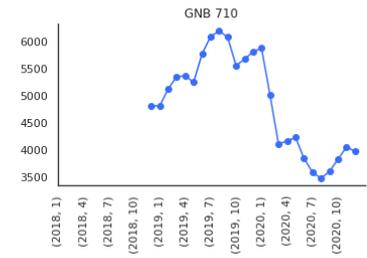
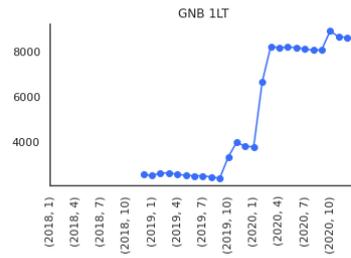
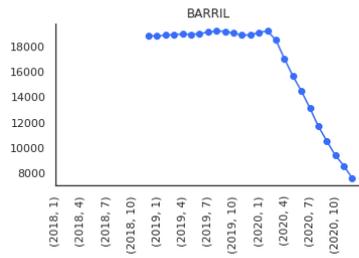


A priori, se puede ver un gran incremento en el calibre lata 470cc, siendo el empaque líder cuando de cerveza se trata. Se puede observar que, para mediados del 2018, la lata 470cc supera al calibre lata 350cc. Lo analizaremos con más detalle con el gráfico de medias móviles, para quitar el efecto de estacionalidad, y los gráficos individuales de cada calibre.

Ventas de cerveza por marca (media móvil 12 meses)



Ventas mensuales por calibre de cerveza (media móvil 12 meses)



Como primer punto importante y validando la conclusión prematura previa, se puede detectar un aumento en la demanda del calibre 470cc formato lata, estableciéndose como el calibre líder en las ventas por sobre la lata 350cc que fue históricamente el formato más vendido en Chile. A partir del 2018 la lata 470cc gana terreno y esto se consolida durante la pandemia. El aumento en la demanda de este calibre se da principalmente por un cambio de estrategia a nivel compañía. La misma, al igual que la competencia,

comenzaron a impulsar este formato más grande por una cuestión de conveniencia de precio por litro, ya que el precio de la lata está atada al valor del aluminio y el mismo es muy volátil. A partir del 2019 todas las marcas de CCU Chile están ejecutando el calibre 470cc y desincentivando el consumo de la lata 355cc. Asimismo, las marcas importadas también han perdido relevancia debido a alzas en el tipo de cambio que no les permiten llegar a puntos bajos de precio.

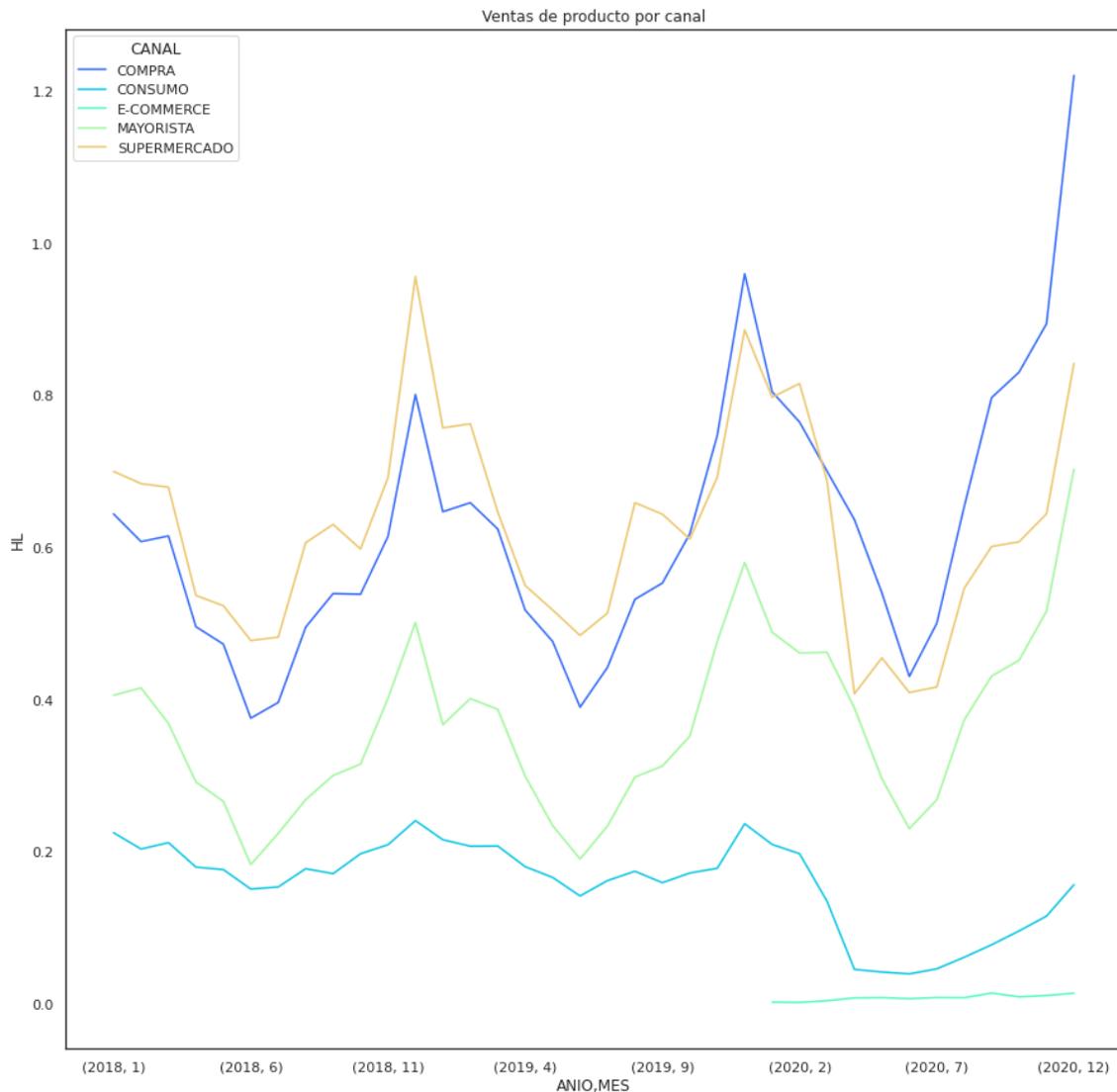
Se advierte, asimismo, una caída importante en los empaques con mayor volumen de cerveza como el barril y las botellas (GNB y GRB) de 1,5 o 1 litro. Esto se debe a que aquellos calibres son utilizados principalmente en el canal “Consumo” que, como veremos en el próximo apartado, corresponde a los sitios en donde la bebida se consume en el lugar, como por ejemplo: restaurantes, bares, discoteques, etc. Desde que comenzó la pandemia, estos lugares se cerraron y no fue, hasta fines del 2020, que comenzaron a abrir de a poco sus puertas con varias restricciones.

Venta en Hectolitros por Canal

Una parte muy importante de este estudio radica en el análisis de los comportamientos de compra en los distintos canales. Llamamos canal a todos los lugares o medios por los que los consumidores pueden realizar sus compras de producto. Como se mencionó algunos apartados más arriba, en el dataset provisto la información de canales se divide en:

- *Consumo*: refiere a todas las ocasiones en las que se consume el producto en el local que lo administra y ofrece como, por ejemplo: restaurantes, discoteques, bares, etc.
- *Compra*: hace referencia a las ventas que se realizan en almacenes, licorerías, etc. Donde el cliente compra y se lo lleva a su hogar
- *Supermercado*
- *Mayorista*
- *E-Commerce*: venta online

Aplicamos la misma serie de gráficos para poder analizar desde otra perspectiva la información:



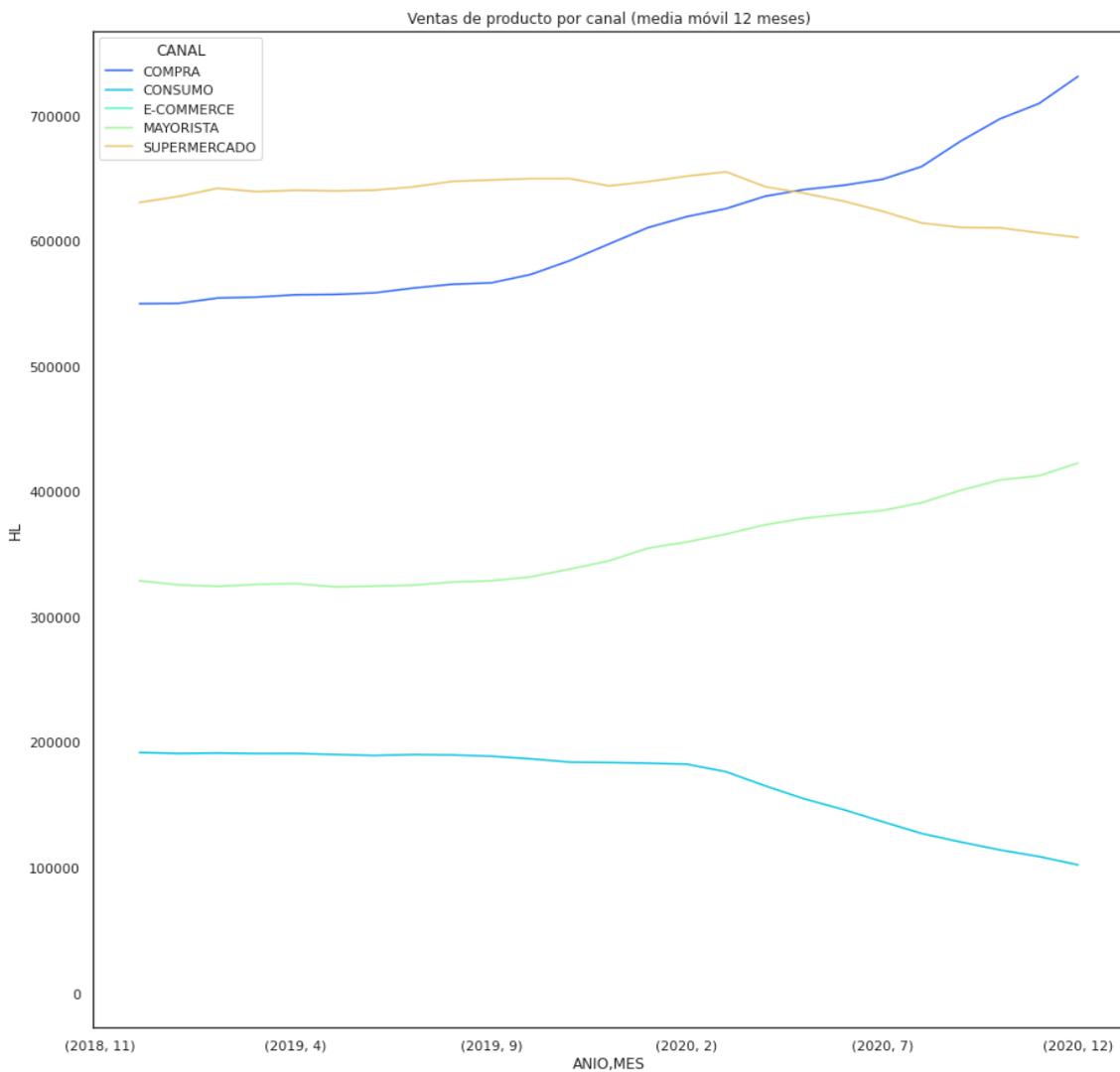
A priori, en este grafico vemos que durante los años 2018 y 2019 el mayor volumen de venta correspondía a los canales Supermercado y Compra con una magnitud y comportamiento bastante similar, muy por arriba de los canales Mayorista y Consumo que se llevan una menor cuota del conjunto. En los tres primeros canales, se pueden observar algunos picos de demanda muy parecidos por la estacionalidad. El canal e-Commerce comienza a registrarse a partir de fines del 2019 y se ve una tendencia creciente durante el 2020.

Hacia fines del 2019 se puede visualizar un aumento de ventas en el canal Compra, superando al canal Supermercado. Esto se debe a un ajuste en la política de precios en el canal “tradicional” (compra) que impulsó la compañía, como un plan para la captura del market share tanto para las

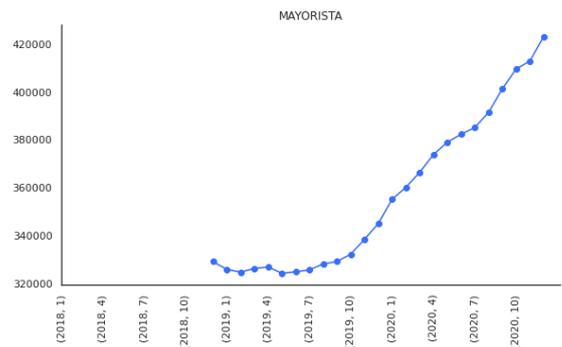
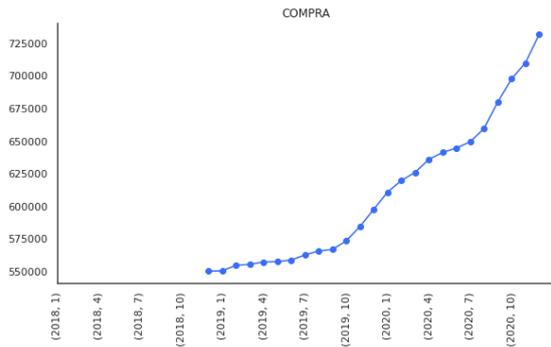
marcas mainstream como premium y evitar que la competencia le gane esa porción del mercado.

A partir del 2020, con la llegada de la pandemia, vemos cómo caen las ventas en general, pero con una mayor tendencia a la baja en canal Supermercado y “Consumo”, potenciando aún más el cambio que se venía gestando en el 2019, con un aumento en el canal Compra.

Lo veremos con más detalle en el gráfico de las medias móviles y con los gráficos individuales de cada canal:



Ventas mensuales por canal (media móvil 12 meses)



Con el detalle de estos gráficos, se corrobora la fuerte caída en la venta en el canal Supermercado en el segundo trimestre de 2020, cuando los casos de COVID-19 positivo aumentan de forma acelerada y la gente comienza a explorar cierto temor al contagio en los lugares más concurridos, por lo que se vuelca a la compra de bebidas alcohólicas en lugares más pequeños con poco tráfico de gente. De esta forma, el canal “Compra” supera al Supermercado.

Asimismo, se ve un desplome en el canal “Consumo” dado por las restricciones de circulación y cierre de lugares como restaurantes, bares, etc.

Matriz de Correlación

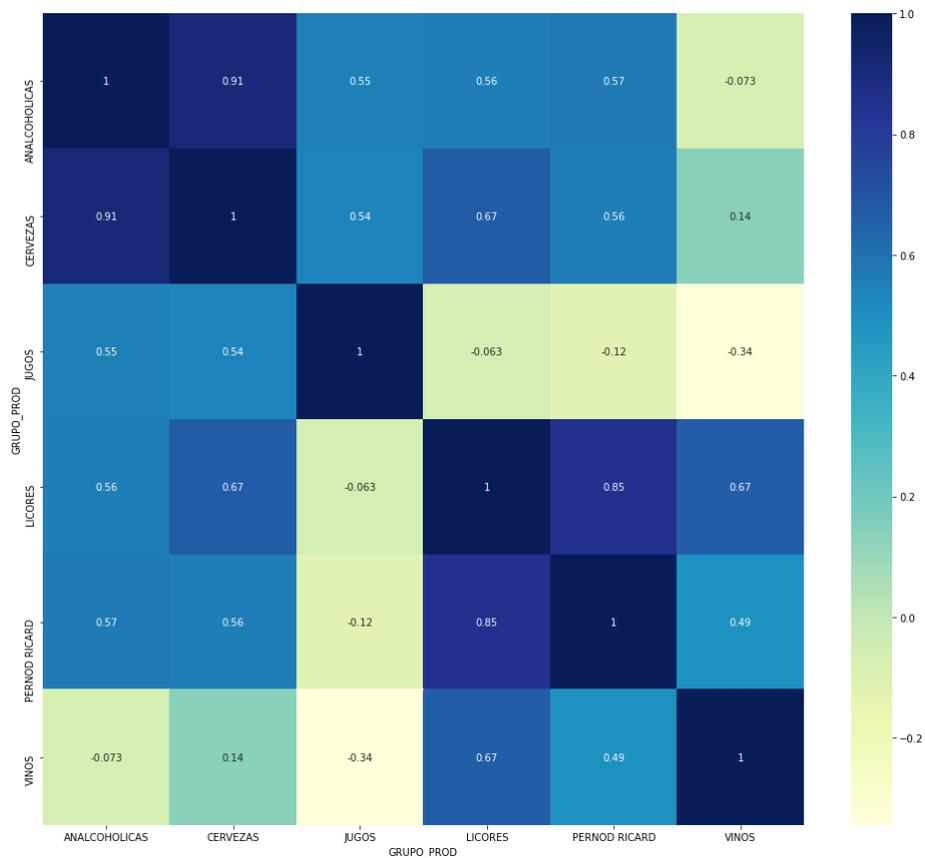
Para entender la relación entre dos elementos o variables y ahondar un poco más en lo que se vio en el análisis exploratorio de datos, visualizamos a continuación dos matrices de correlación que nos ayudará a entender el comportamiento de los consumidores a lo largo del tiempo.

Una matriz de correlación exhibe los valores de correlación entre variables. En este caso, utilizamos el coeficiente de Pearson, que mide el grado de relación lineal entre cada par de observaciones:

$$r = \frac{\sum z_x z_y}{N - 1}$$

Tal como lo desarrolla Aleksander Dietrichson, en su trabajo “Métodos Cuantitativos, Coeficientes de Correlación”, este coeficiente puede tomar valores que varían entre -1 y 1, indicando una relación positiva si el índice toma el valor de 1, relación negativa con -1 y si no hay asociación entre las variables el coeficiente es igual a 0. Una relación positiva (1), advierte que a medida que aumenta el valor de una variable, también lo hace el valor de la otra. Una relación negativa, en cambio, muestra que a medida que aumenta el valor de una variable, la otra disminuye.

Grupo de Productos



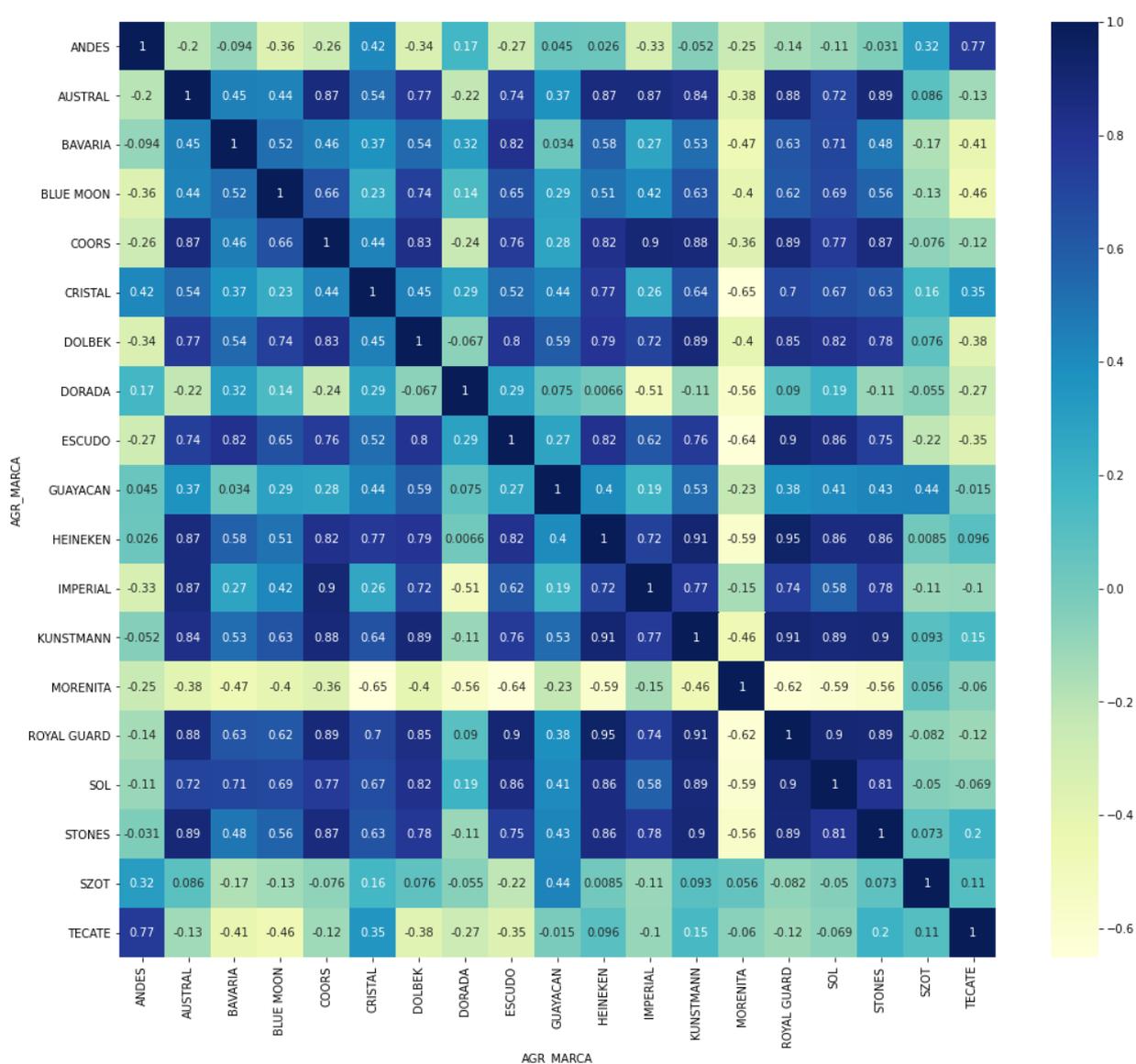
En este gráfico, observamos una alta correlación positiva entre la categoría Analcohólicas y Cervezas que valida lo indicado más arriba cuando se analizó la evolución de ventas de acuerdo a los tipos de producto. Esto puede explicarse por el tipo de bebida alcohólica que es la cerveza: suave, refrescante y liviana (en la mayoría de los casos) y puede reemplazar o compensar muy bien a una bebida gaseosa o saborizada.

No sucede lo mismo con la categoría Analcohólicas y Vinos, que son productos totalmente distintos en el sentido del sabor y ocasiones de consumo. Se puede ver una correlación negativa también con los jugos y el grupo de bebidas alcohólicas más fuertes como los licores, los productos de Pernod Ricard y los vinos.

Respecto a los jugos, que, a diferencia de las analcohólicas y cervezas tiene un comportamiento más estable en cuanto a las ventas, vemos que el índice de correlación con las dos categorías más vendidas es cercano al 0.5. Esto significa que la relación es positiva pero no necesariamente al variar una varía la otra en la misma medida.

Marcas de Cervezas

CCU Chile cuenta con casi el 75% del mercado de bebidas en el país y lo mismo sucede con las cervezas, produciendo y comercializando más de 20 marcas. En éste análisis nos concentraremos sólo en las más relevantes que pueden darnos algunos insights acerca del comportamiento de los consumidores en esta categoría.



Se observa una gran correlación positiva entre las siguientes marcas:

- Austral y Royal Guard (0,88)
- Austral y Stones (0,89)
- Coors e Imperial (0,9)
- Coors y Royal Guard (0,89)
- Doblek y Kustmann (0,89)
- Escudo y Royal Guard (0,9)
- Heineken y Royal Guard (0,95)
- Heineken y Kunstmann (0,91)
- Kusntmann y Sol (0,89)
- Royal Guard y Stones (0,89)
- Sol y Royal Guard (0,9)
- Stones y Kusntmann (0,9)

Correlación negativa:

- Dorada e Imperial (-0,51)
- Morenita con:
 - Cristal (-0,65)
 - Dorada (-0,56)
 - Escudo (-0,64)
 - Heineken (-0,59)
 - Royal Guard (-0,62)
 - Sol (-0,59)
 - Stones (-0,56)

Tanto la correlación positiva como negativa entre las marcas de cerveza, puede explicarse por tres variables que afectan la decisión de los consumidores a la hora de hacer su compra:

- Precio: se pueden distinguir tres niveles de precio:
 - Alto: cervezas importadas o internacionales y de nicho
 - Medio: cervezas con algún diferencial o con cierta popularidad en el mercado
 - Bajo: cervezas de marcas masivas o populares
- Status: se refiere al status o el prestigio que le da al consumidor comprar cierta marca de cerveza. Esto depende también el precio y del origen de la cerveza.
- Tipo de cerveza:
 - Fuertes: cervezas con cuerpo, robustas y densas.

- Suaves: cervezas fáciles de tomar y refrescantes.

La correlación positiva se da, en particular, con marcas de cerveza que cumplen las siguientes condiciones:

- Precio Medio – Status Alto
- Precio Alto – Status Alto
- Precio Medio – Tipo de cerveza Suave
- Precio Bajo – Tipo de cerveza Fuerte

Por otro lado, en la correlación positiva, se puede visualizar claramente la siguiente relación:

- Precio Bajo – Tipo de cerveza Suave – Status Medio/Bajo

Clustering

Para realizar este análisis trabajamos con un Clustering en series de tiempo. Este método, divide una colección de serie temporal en función de la similitud en las características entre éstas. Las series temporales se pueden agrupar de acuerdo a tres reglas:

- Contener valores similares a lo largo del tiempo (que se está analizando)
- Tender a aumentar o a disminuir al mismo tiempo
- Tener patrones repetidos similares

En este caso, nos centraremos en la evolución de venta de, por un lado, sólo cervezas y por otro lado todos los grupos de producto. Lo que nos va a interesar va a ser fraccionar las distintas series de tiempo (dadas por las ventas en hectolitros de cada marca), basándonos en una métrica de similaridad y distancia.

Métrica de Similaridad

Una de las métricas de similaridad más populares y utilizadas es la de distancia euclídea junto con el algoritmo k-medias, pero para series de

tiempo como el que queremos analizar, esta métrica no es la mejor opción ya que la misma es invariante a los cambios en el tiempo e ignora la dimensión temporal de los datos.

Por este motivo, se indagó sobre otras métricas que pudieran servir para este trabajo. En esta búsqueda, se encontró la Dynamic Time Warping (DTW) que se utiliza generalmente para medir la similaridad entre dos secuencias temporales que pueden no estar alineadas en velocidad, longitud o en el propio tiempo. Esta métrica se ajusta perfectamente para aplicar a los datos con los que contamos.

Elección de cantidad de Clusters

Al trabajar con el algoritmo k-medias, debemos indicarle previamente la cantidad de clusters que debe buscar. Como no tenemos la “verdadera” etiqueta de los clusters, utilizaremos algunas métricas de performance de clusterización que nos ayudarán a encontrar la cantidad de clusters como un hiperparámetro que maximice esta métrica.

SSE

La Suma de Cuadrados residuales o errores (SSE) es una técnica estadística que se utiliza para medir la cantidad de varianza en un conjunto de datos que no se explica por un modelo de regresión. Es una medida de coherencia interna de los clusters. Viene dada por la siguiente formula:

$$\sum_{i=0}^n \min_{\mu_j \in c} (\|x_i - \mu_j\|)^2$$

μ = Media

c = valores observados

Es la métrica “método de codo” que consiste en graficar la inercia o SSE correspondientes a la aplicación de clusterización con diferente numero

predefinido de clusters. Se decide la cantidad óptima de clusters cuando hay un punto de inflexión (codo) en el gráfico.

Coefficiente de Silhouette

Este coeficiente mide cómo de similares son los objetos de un mismo cluster comparado con otros clusters. Se calcula usando la distancia intra-cluster media (a) y la distancia media al cluster más cercano (b) para cada observación. El coeficiente de Silhouette para una muestra es:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

El mayor valor es 1 y el menor es -1. Los valores cercanos a cero (0) indican que los clusters se están solapando, mientras que los valores negativos indican que una observación ha sido asignada al cluster equivocado.

Dado que k-medias es sensible a los centroides aleatorios del inicio, iteraremos el proceso varias veces para asegurar la robustez de los resultados. Se definieron 10 estados aleatorios diferentes. Una vez aplicadas ambas métricas para los dos datasets definidos, vemos que para los grupos de p productos, el número de clusters se visualiza de manera más evidente en el Coeficiente de Silhouette y para las marcas de cervezas en SSE, y la misma es $k = 3$.

Gráfico 1: Coeficiente de Silhouette y SSE por Grupo de Productos

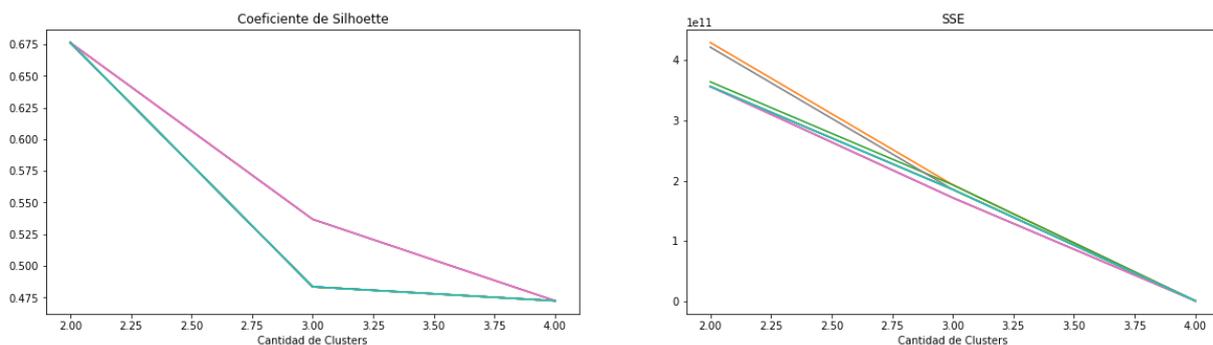
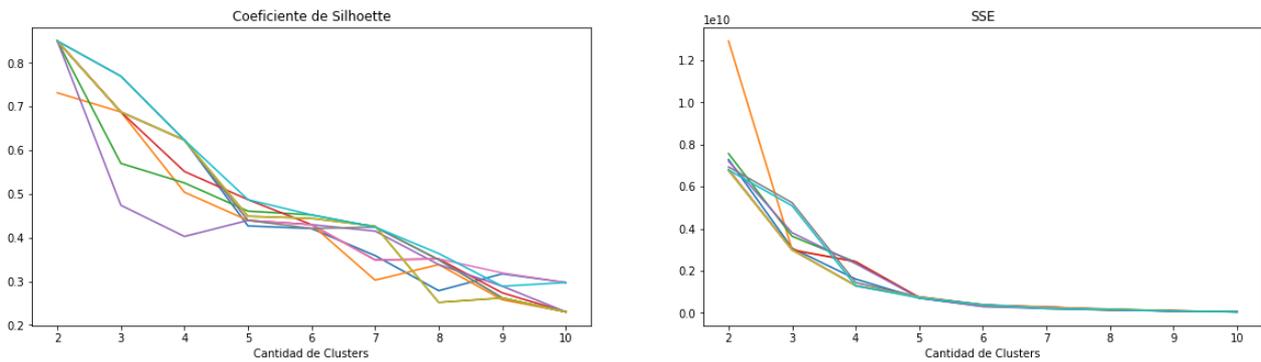
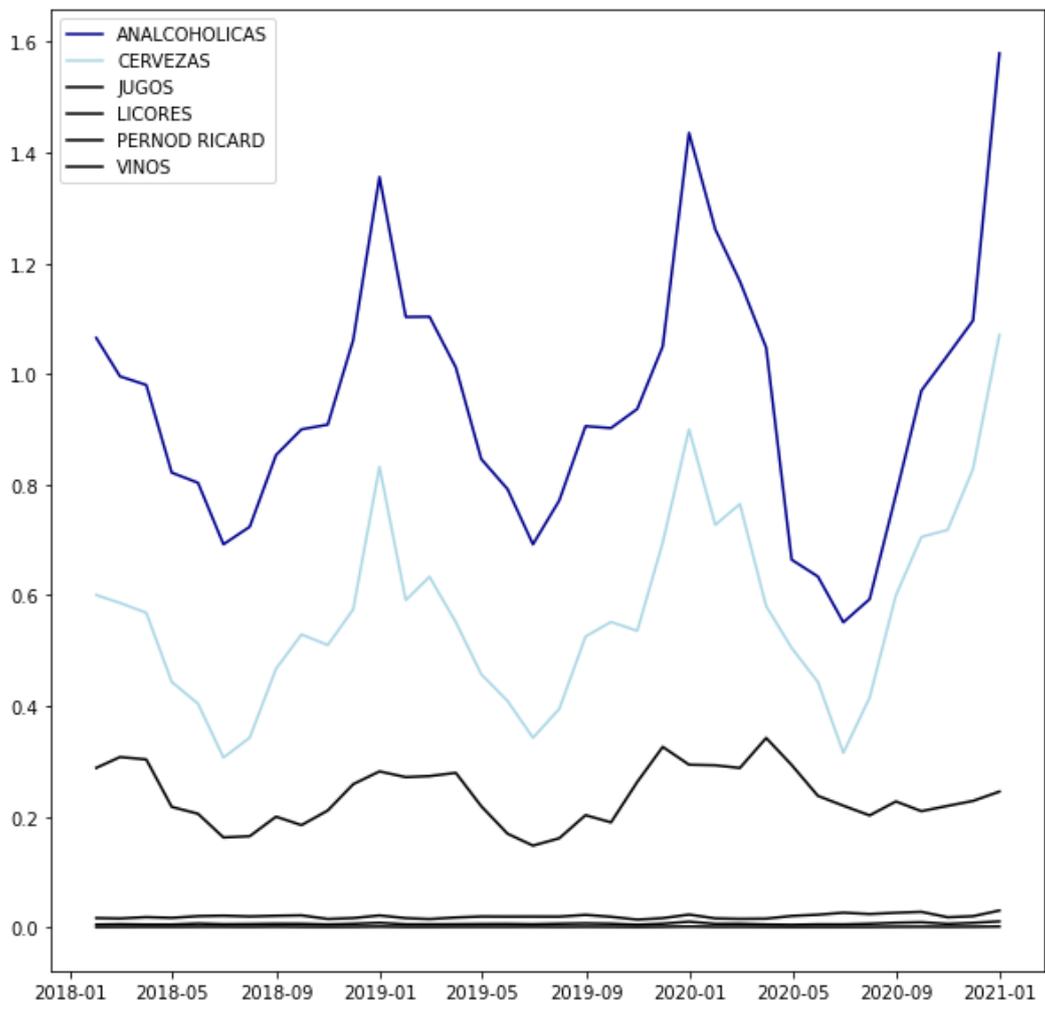


Gráfico 2: Coeficiente de Silhouette y SSE para marcas de Cervezas



Una vez que tenemos el número k y la métrica DTW con la cual trabajar, avanzamos con el análisis de agrupamiento:

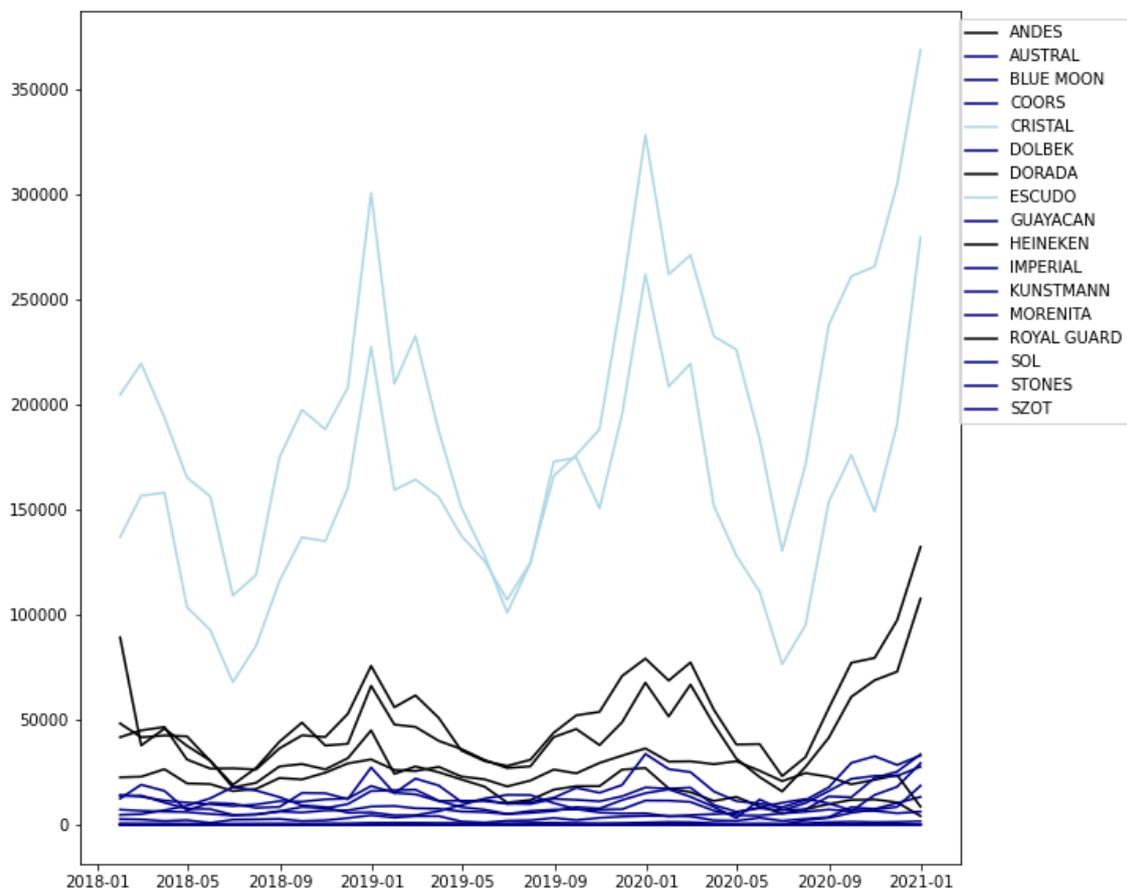
Grupo de Productos



Observando el grafico, podemos separar los Grupos de Productos en 3 clusters que se pueden interpretar de la siguiente forma:

- Ventas Altas: bebidas de consumo cotidiano como las bebidas Analcohólicas.
- Ventas Medias: bebidas de consumo social como la Cerveza
- Ventas Bajas: bebidas de consumo ocasional, como Jugos, Licores, Vinos y Bebidas Blancas.

Marcas de Cervezas



De la misma manera en que se consideró para el gráfico de los grupos de productos, podemos separar las marcas de cervezas en 3 clusters que se pueden interpretar, análogamente, de la siguiente forma:

- Ventas Altas: con Escudo y Cristal
- Ventas Medias: Andes, Dorada, Heineken y Royal Guard

- Ventas Bajas: todo el resto de las marcas, Austral, Blue Moon, Coors, Doblek, Guayacan, Imperial, Kustnmann, Morenita, Sol, Stones y Szots.

Conclusión Clustering

Sin dudas, el número de clusters preferible, es tres. Es decir, que existen tres grandes grupos en los que los datos de las series que lo componen son similares entre sí y distintos a las series de los otros grupos.

- Ventas Altas: dentro de este grupo, se encuentran las bebidas de consumo cotidiano como las Bebidas Analcohólicas y en cuanto a las marcas de cerveza, las que tienen mayor volumen de consumo y son más masivas como Escudo y Cristal.
- Ventas Medias: aquí se encuentran las bebidas de consumo social con alcohol: las cervezas. Respecto a las marcas, vemos que aparecen algunas marcas nacionales y otras internacionales que son más populares.
- Ventas Bajas: se pueden detectar todas las bebidas de consumo ocasional como los jugos o bebidas con mayor graduación alcohólica como los Licores, Vinos y Bebidas Blancas. Dentro de las cervezas encontramos algunas internacionales con un precio alto o nacionales con precio muy bajo.

Etapa Predictiva

Para realizar el análisis predictivo se utilizó Prophet, software de código abierto desarrollado por el equipo Core Data Science de Facebook.

Prophet se utiliza principalmente para análisis de series de tiempo. Es un procedimiento para pronosticar datos de series de tiempo basado en un modelo aditivo, en el que las tendencias no lineales se ajustan a las estacionalidad anual, semanal, diaria y otros efectos.

Se eligió este modelo para hacer las predicciones porque, de acuerdo a la descripción realizada por el equipo de Facebook, funciona mejor con series

que tienen fuertes efectos estacionales y varias temporadas de datos históricos. Además, se indica que es robusto ante los datos faltantes y los cambios en la tendencia, y normalmente maneja bien los valores atípicos.

Para la predicción, se podrían haber utilizado otros modelos, como, por ejemplo, el Modelo Autorregresivo de Media Móvil (ARMA) que, de acuerdo a José A. Mauricio, en su trabajo “Introducción al Análisis de Series Temporales”, tienen un buen desempeño predictivo en el corto plazo, pero pierde la capacidad de pronóstico en horizontes más amplios. Además, el modelo ARMA, no tiene en cuenta la estacionalidad en la serie de tiempo, y en el dataset de bebidas este aspecto es fundamental.

El dataset provisto, cumple con todas las condiciones mencionadas arriba por tratarse de las ventas históricas de bebidas que tienen una marcada estacionalidad (en particular las bebidas alcohólicas) y además tienen el efecto de la pandemia durante todo el 2020.

Se realiza un análisis predictivo por Grupo de Productos y otro por Canal. Se eligieron sólo estos dos features, porque se cree que son los que demostrarán de forma mucho más clara un cambio en los hábitos de consumo en el largo plazo.

Para cada uno de los subgrupos dentro de “Grupo de Producto” y “Canal”, se realiza un breve diagnóstico de predicción, utilizando la función *cross validation* y, en consecuencia, métricas de performance como *Error Cuadrático Medio (RMSE)* y *Error Absoluto Medio (MAE)*:

La función *Cross Validation* permite realizar pronósticos aplicando un esquema de *cross validation* temporal, que forma un tipo de enfoque de entrenamiento de “ventana deslizante” utilizando datos históricos. Esto se hace seleccionando puntos de corte en el historial, y para cada uno de ellos ajustando el modelo usando datos sólo hasta ese punto de corte. Luego se pueden comparar los valores pronosticados con los valores reales. El primer parámetro que definimos es el horizonte de predicción (*horizon*), luego el “*training data*” (*initial*) indicando cuánto tiempo deben entrenarse los datos antes de comenzar las pruebas, y, por último, se define el período (*period*) que marca con qué frecuencia detenerse y realizar una predicción.

El *Error Cuadrático Medio (RMSE)* realiza una diferencia entre los valores estimados y los valores reales, elevando estas diferencias al cuadrado para

luego calcular el promedio de todas ellas. El RMSE mide la magnitud de error. Su ecuación es la siguiente:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} = \sqrt{MSE}$$

y = valor observado

\hat{y} = valor predicho

N = número de valores analizados

El *Error Absoluto Medio (MAE)* muestra qué tan cercana es la predicción hecha al resultado real. Se calcula la diferencia entre el valor obtenido y el valor real en valor absoluto y luego se calcula el promedio. Su ecuación es la siguiente:

$$MAE = \sqrt{\frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|}$$

y = valor observado

\hat{y} = valor predicho

N = número de valores analizados

El RMSE en conjunto con el MAE se utilizan para determinar la variación del error en el conjunto de mediciones. A mayor diferencia entre MAE y RMSE mayor varianza entre los errores individuales de la muestra.

Luego de cada análisis individual, se muestra un pronóstico conjunto para arribar a la predicción de cantidades totales de hectolitros vendidas.

Grupo de Productos

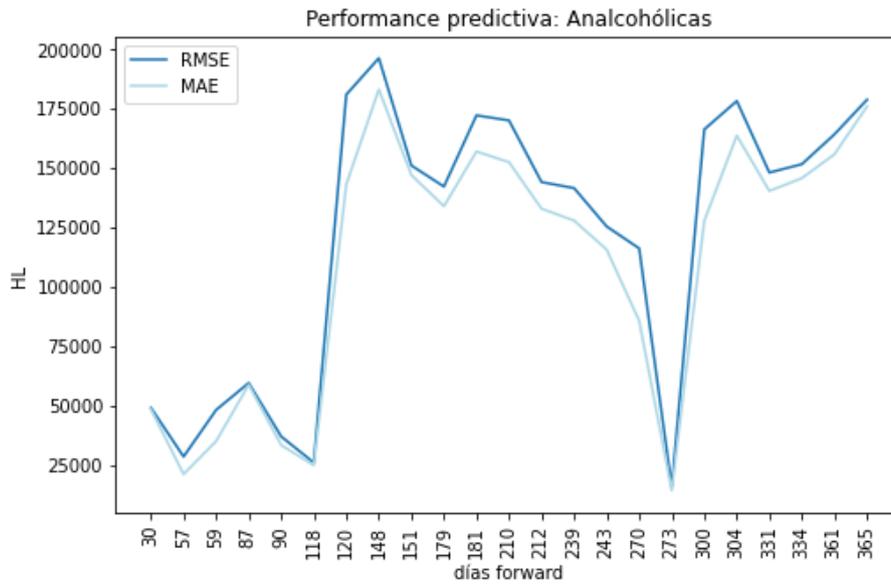
Para poder trabajar este dataset sobre el modelo de Prophet, primero debemos acondicionarlo y agregarle el índice con la fecha:

GRUPO_PROD	ANALCOHOLICAS	CERVEZAS	JUGOS	LICORES	PERNOD RICARD	VINOS	mes
2018-01-31	1.065042e+06	6.005771e+05	288260.592001	5083.311049	746.668500	16893.05832	2018-01-31
2018-02-28	9.956238e+05	5.862512e+05	307836.939131	5804.840150	888.583000	16150.50355	2018-02-28
2018-03-31	9.800959e+05	5.685548e+05	303483.654372	5192.854250	1030.045000	18654.56792	2018-03-31
2018-04-30	8.216989e+05	4.435962e+05	217797.247103	5201.284850	982.439000	17277.84774	2018-04-30
2018-05-31	8.033501e+05	4.041813e+05	205372.654973	6982.419200	1126.694500	20379.33606	2018-05-31
2018-06-30	6.921829e+05	3.068207e+05	162554.941308	5758.117900	1106.831000	21174.72158	2018-06-30
2018-07-31	7.239278e+05	3.429592e+05	164837.653251	6021.827250	973.732000	19837.20608	2018-07-31
2018-08-31	8.535037e+05	4.672627e+05	200132.049673	6614.242220	1266.010000	21013.71577	2018-08-31
2018-09-30	8.999423e+05	5.293142e+05	184781.086844	6517.715200	968.647000	21760.07985	2018-09-30
2018-10-31	9.084184e+05	5.100292e+05	211230.620784	5301.640310	889.052000	15309.20337	2018-10-31
2018-11-30	1.061377e+06	5.746869e+05	258790.279624	6500.672450	993.425500	16897.14339	2018-11-30
2018-12-31	1.355860e+06	8.322310e+05	281885.557400	8284.496460	1373.579500	21643.30640	2018-12-31
2019-01-31	1.102994e+06	5.911894e+05	271523.543510	5600.934490	883.679000	16790.76107	2019-01-31
2019-02-28	1.103397e+06	6.334866e+05	273425.373970	5556.154060	882.462498	15158.33669	2019-02-28
2019-03-31	1.011789e+06	5.516090e+05	279282.661160	5790.953450	1055.765498	17974.96592	2019-03-31
2019-04-30	8.462552e+05	4.574013e+05	219001.873330	6135.444190	1024.569000	19762.71340	2019-04-30
2019-05-31	7.919072e+05	4.091925e+05	169333.089146	6310.710800	1024.932500	19556.72440	2019-05-31
2019-06-30	6.921281e+05	3.425862e+05	147797.511951	5708.208530	1189.214501	19636.10129	2019-06-30
2019-07-31	7.713691e+05	3.947663e+05	161047.209110	6651.831400	943.221500	19507.18178	2019-07-31
2019-08-31	9.056208e+05	5.255257e+05	202858.715100	7397.915800	1119.526000	22667.21524	2019-08-31
2019-09-30	9.022054e+05	5.520138e+05	189777.630180	6718.995250	1026.303000	19227.21798	2019-09-30

Realizado este paso, se comienza a trabajar con cada subgrupo.

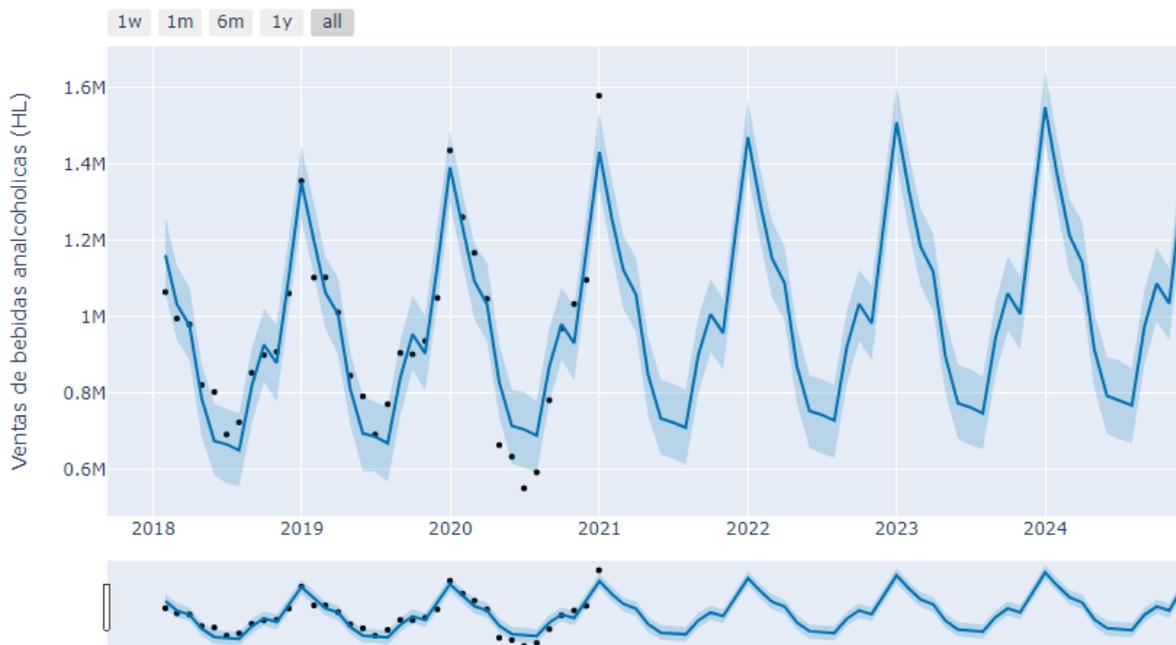
Bebidas Analcohólicas

Diagnóstico del modelo



En el eje x se muestran “días forward”, que corresponden a los días de horizonte establecidos para realizar el diagnóstico de predicción del modelo. Se observa que para las bebidas analcohólicas, el modelo tiene una performance relativamente buena dentro de los 4 meses vista y luego empeora terminando con un error al alza.

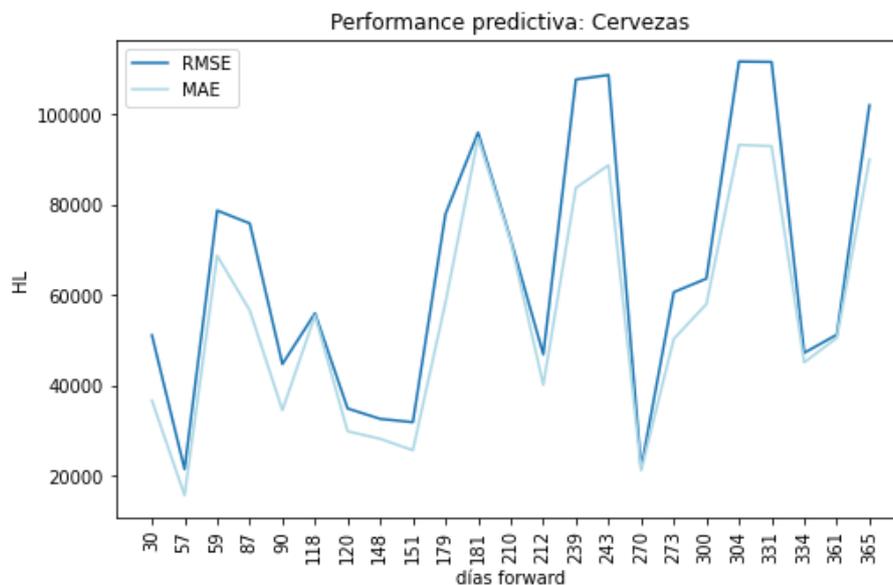
Predicción



En el gráfico de predicción para las bebidas analcohólicas realizado para los próximos 4 años, se puede ver un leve aumento en las ventas año tras año, con la estacionalidad característica de la categoría. En este gráfico se advierte claramente lo comentado con anterioridad, la categoría de analcohólicos no se mostró tan afectada por la pandemia con una caída del 23% vs AA, más bien se vio beneficiada ya que en diciembre 2020 se puede observar un pico de ventas que se pronostica irá aumentando a un ritmo del 2% año tras año.

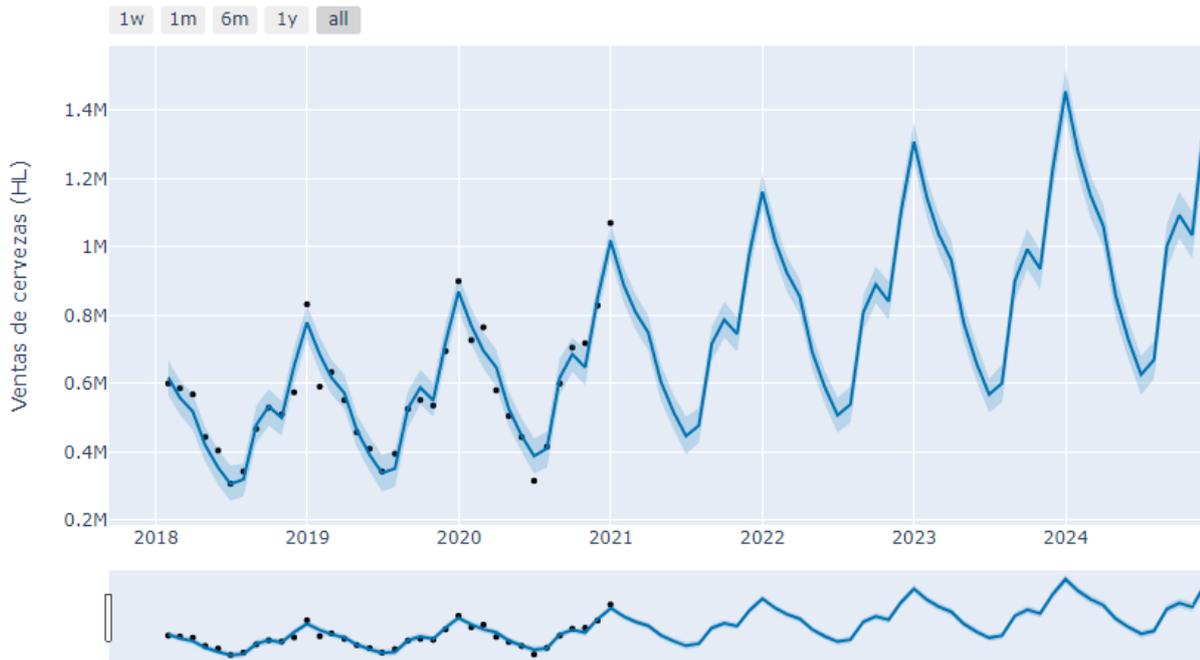
Cervezas

Diagnóstico del modelo



En el eje x se muestran “días forward”, que corresponden a los días de horizonte establecidos para realizar el diagnóstico de predicción del modelo. Se puede percibir que este modelo en particular para cervezas, presenta una performance muy buena en la predicción a dos meses vista. Luego varía bastante con algunos picos.

Predicción

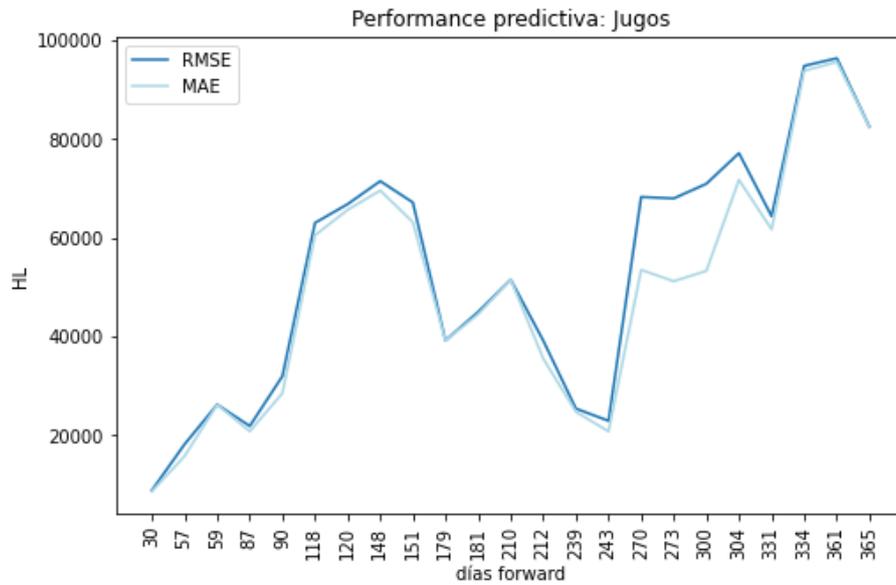


En la predicción de esta categoría se puede visualizar un incremento año tras año de un 9% en promedio, tomando en cuenta el año entero con su estacionalidad.

También se ve claramente, como esta categoría se vio “beneficiada” con la pandemia, creciendo un 5% vs AA.

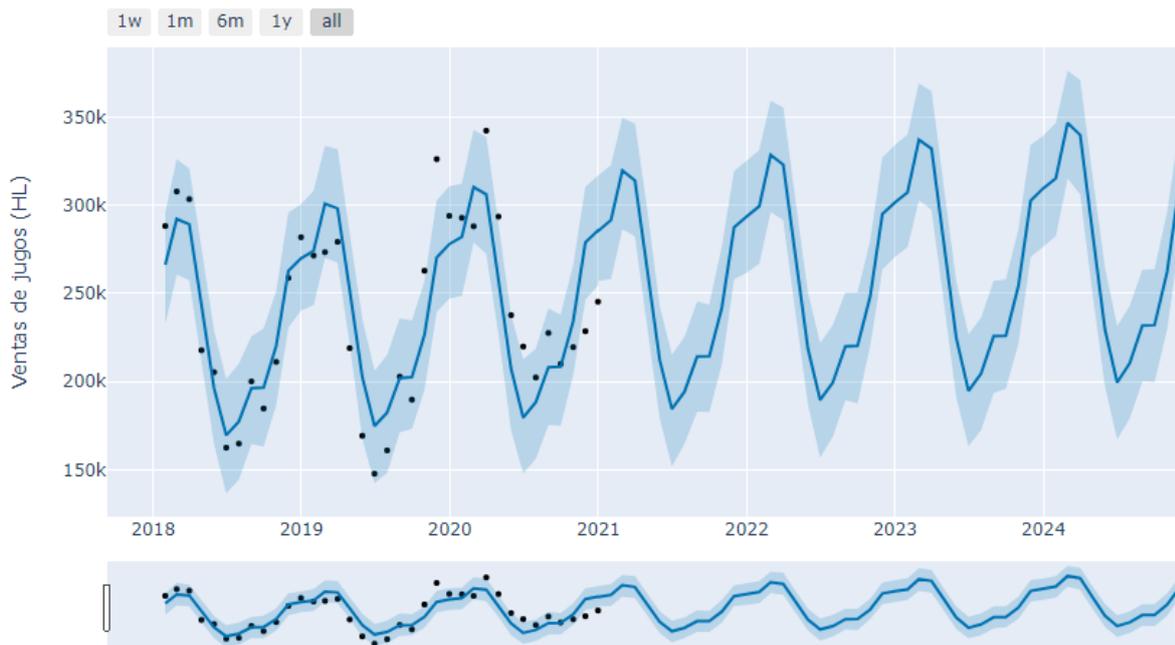
Jugos

Diagnóstico del modelo



En el eje x se muestran “días forward”, que corresponden a los días de horizonte establecidos para realizar el diagnóstico de predicción del modelo. La performance del modelo en la categoría jugos se observa muy bien hasta el tercer mes, al igual que la categoría analcohólicos.

Predicción

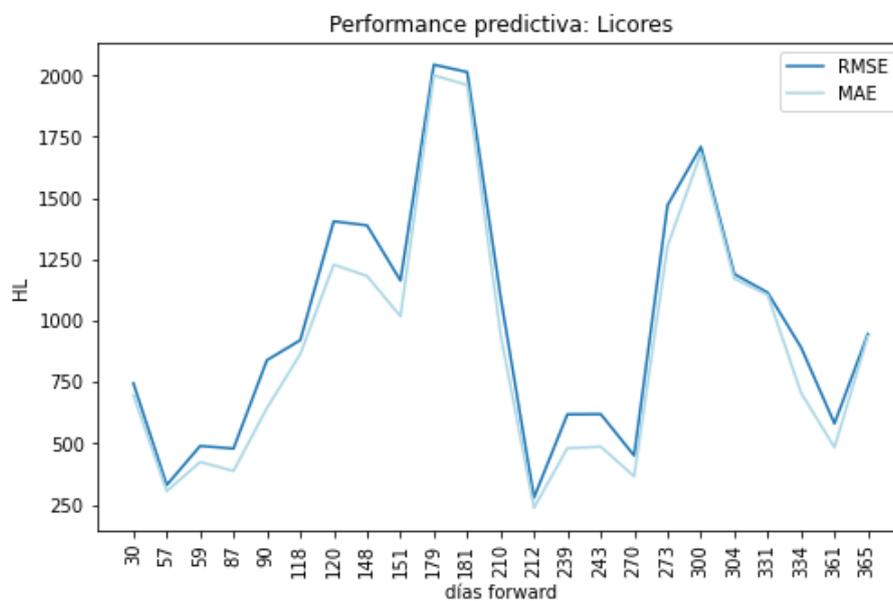


La categoría de Jugos es la que más favorecida resultó por la pandemia en el primer semestre. El nivel de consumo aumentó durante la cuarentena más estricta en Chile, pero para fin de año, este aumento se neutraliza con una caída del 17% vs AA. Asimismo, como se ve en el gráfico de arriba, esta categoría sufre (al igual que las demás) las alteraciones propias de la estacionalidad y de los productos sustitutos.

Para esta categoría, la tercera con más volumen de la compañía, se puede predecir una caída considerable para el 2021, pero luego se normaliza con un aumento estable para los próximos años.

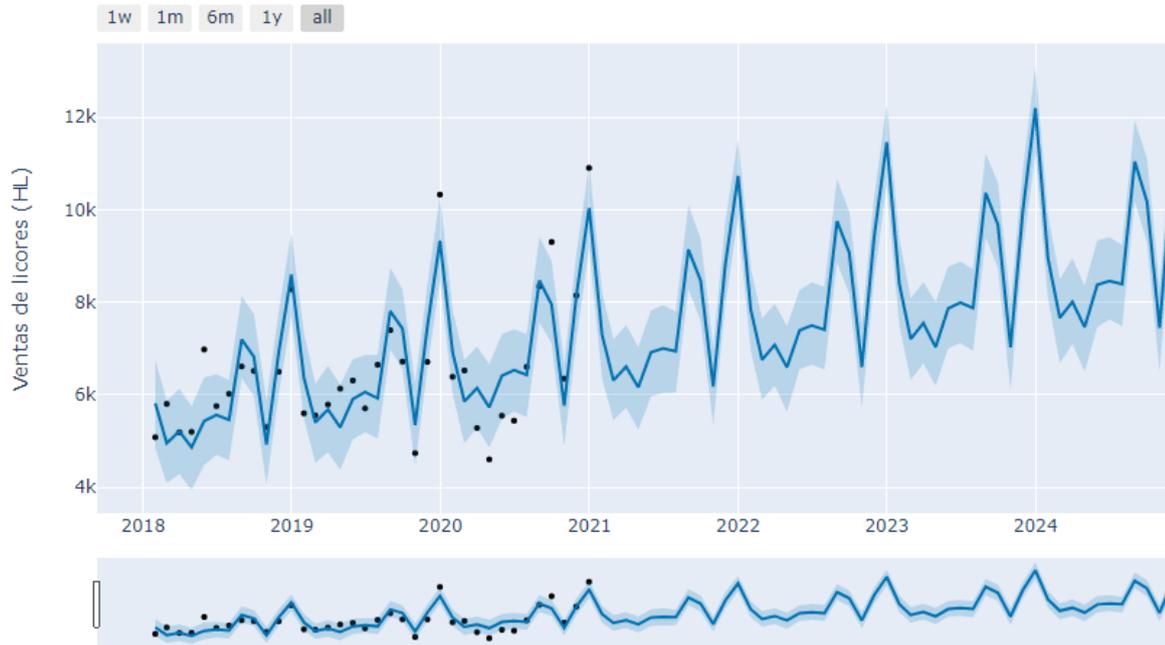
Licores

Diagnóstico del modelo



En el eje x se muestran “días forward”, que corresponden a los días de horizonte establecidos para realizar el diagnóstico de predicción del modelo. Al igual que en la categoría Jugos, en Licores, el modelo se comporta muy bien hasta el tercer mes y luego el desempeño empieza a empeorar.

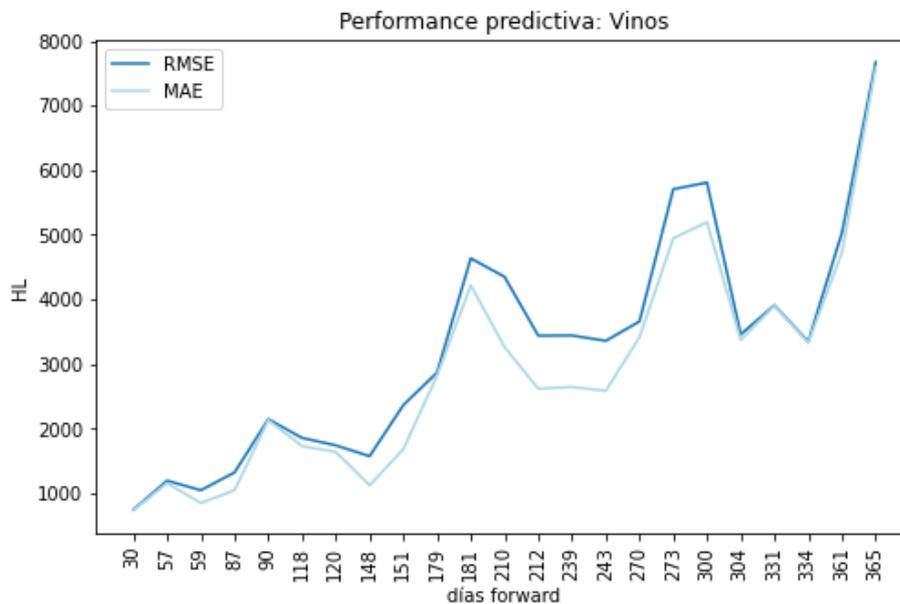
Predicción



En esta categoría se predice una leve caída para fines del 2021 (vs AA), pero un crecimiento promedio del 6% por año hasta el 2024.

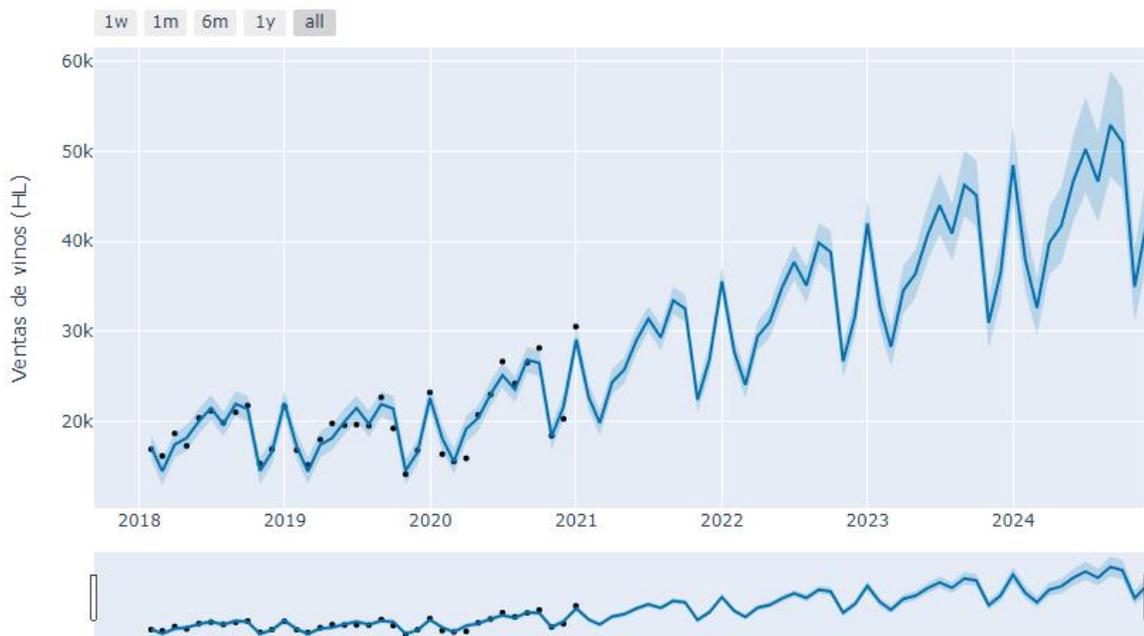
Vinos

Diagnóstico del modelo



En el eje x se muestran “días forward”, que corresponden a los días de horizonte establecidos para realizar el diagnóstico de predicción del modelo. Puede observarse que este modelo para la categoría vinos, tiene una muy buena performance en el corto plazo, pero presenta un sesgo a la sobreestimación conforme avanza el horizonte de predicción.

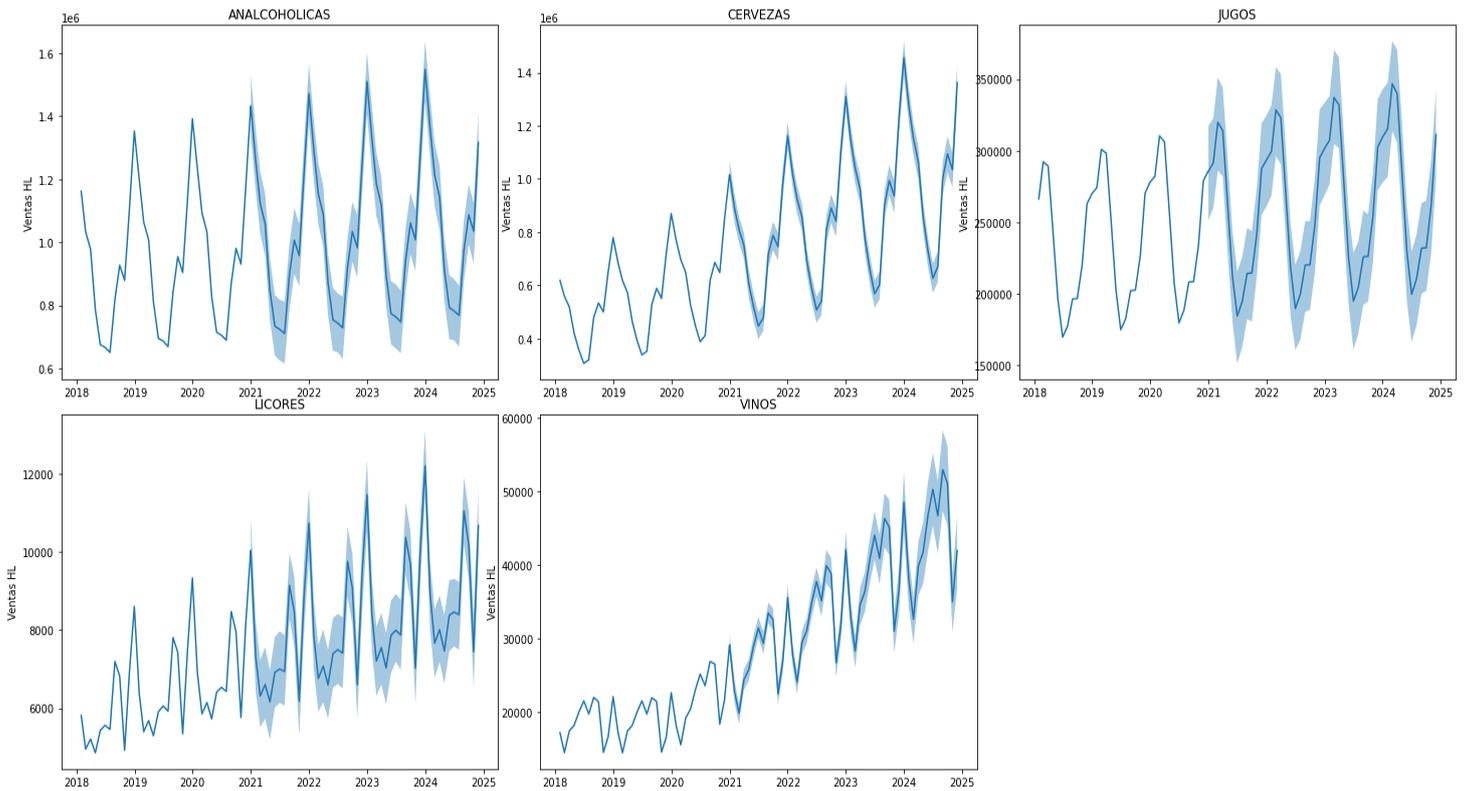
Predicción



De acuerdo a este gráfico, podemos predecir que Vinos será la categoría con mayor aumento de volumen con un crecimiento exponencial a partir del 2021. Pero a la luz del diagnóstico realizado con anterioridad, conviene tomar con cautela esta tendencia creciente proyectada a futuro.

Predicción de ventas totales por Grupo de Producto

Para poder visualizar mejor la predicción para cada categoría y poder estimar el total de las ventas para del período 2021-2024, se analizan todas las categorías a continuación:



Indudablemente, el mercado de bebidas crece año tras año y CCU como compañía, se apropia de la mayor proporción de ese aumento, al ser la empresa líder en el mercado. A pesar de la performance de los modelos en la estimación de los errores, en general se puede asegurar que la predicción prevista de tendencia creciente para la mayoría de las categorías es un hecho concreto.

En el primer año de pandemia vemos que las preferencias de consumo entre las bebidas cambiaron notablemente: en el momento más estricto y complejo de la pandemia en 2020, se percibe un aumento considerable en la categoría de jugos que venía creciendo a muy bajas tasas años anteriores, incluso en el 2019 el consumo cayó versus año anterior. En contraposición, las bebidas analcoholicas que venían mostrando un aumento de volumen año tras año sufre una caída muy fuerte en el primer semestre del 2020, recuperándose hacia fines del año sin llegar al volumen logrado pre-pandemia. Las categorías de Cervezas y Vinos fueron las únicas categorías que no se vieron afectadas por la pandemia. De hecho, ambas presentaron un incremento casi tres veces mayor para finales del 2020 en comparación al año anterior, por lo que creemos que hay una gran oportunidad de seguir desarrollando estas dos categorías, sobre todo la de Cervezas.

Canal

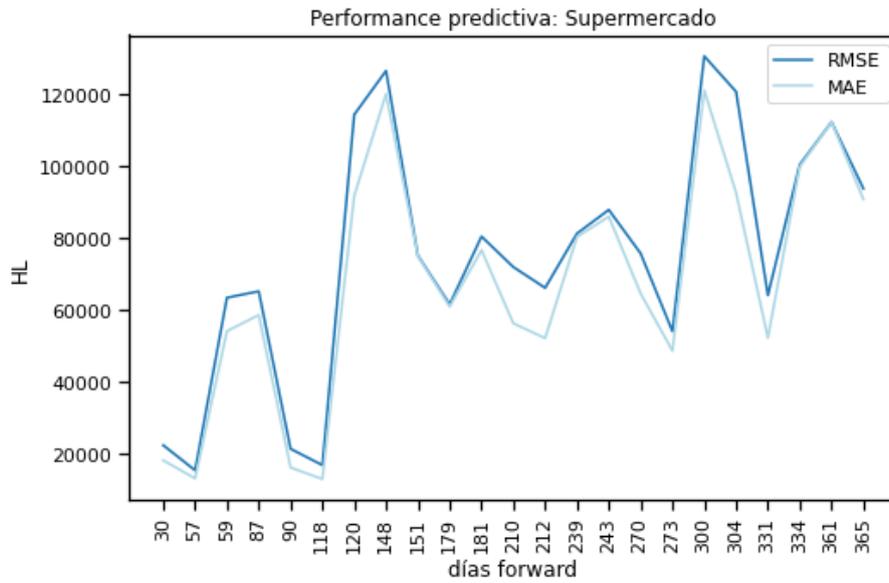
Del mismo modo que se hizo con “Grupo de Productos” vamos a acondicionar la data para poder trabajar la predicción:

CANAL	COMPRA	CONSUMO	E-COMMERCE	MAYORISTA	SUPERMERCADO	mes
2018-01-31	6.646537e+05	240012.852630	NaN	410412.36266	701352.807530	2018-01-31
2018-02-28	6.299403e+05	217699.896240	NaN	420556.82882	685322.534911	2018-02-28
2018-03-31	6.478873e+05	232479.526990	NaN	377320.46216	680752.783872	2018-03-31
2018-04-30	5.274216e+05	200222.627997	NaN	300482.29224	538557.362483	2018-04-30
2018-05-31	5.099512e+05	197223.103650	NaN	278281.92218	525149.525283	2018-05-31
2018-06-30	4.115149e+05	171017.002290	NaN	193190.36622	479216.305658	2018-06-30
2018-07-31	4.316257e+05	171713.797290	NaN	233384.18325	483608.115841	2018-07-31
2018-08-31	5.315215e+05	199139.801850	NaN	281236.49742	607781.509073	2018-08-31
2018-09-30	5.697232e+05	188560.939180	NaN	310314.88778	631707.474204	2018-09-30
2018-10-31	5.763976e+05	219323.627410	NaN	328238.55514	599466.075774	2018-10-31
2018-11-30	6.425465e+05	226282.115100	NaN	413835.70687	693232.653274	2018-11-30
2018-12-31	8.248987e+05	256725.769680	NaN	510984.44549	957407.567110	2018-12-31
2019-01-31	6.473991e+05	216460.611180	NaN	367600.23965	757522.818140	2019-01-31
2019-02-28	6.591939e+05	207923.120520	NaN	401840.11750	762949.042910	2019-02-28
2019-03-31	6.246738e+05	208177.982990	NaN	387731.85446	646918.531050	2019-03-31
2019-04-30	5.181281e+05	181034.947490	NaN	299935.51714	550482.511680	2019-04-30
2019-05-31	4.770302e+05	166923.838870	NaN	235088.98146	518282.079036	2019-05-31

Se puede ver que el canal e-Commerce no tiene ventas durante el 2018 y la mayor parte del 2019, ya que el mismo comienza su operación a finales del 2019. Teniendo esto se muestra a continuación cada canal por separado:

Supermercado

Diagnóstico del modelo



En el eje x se muestran “días forward”, que corresponden a los días de horizonte establecidos para realizar el diagnóstico de predicción del modelo. El desempeño de este modelo para el canal supermercado, es bastante inestable y variable, pero con tendencia decreciente a medida que avanza el horizonte de predicción.

Predicción

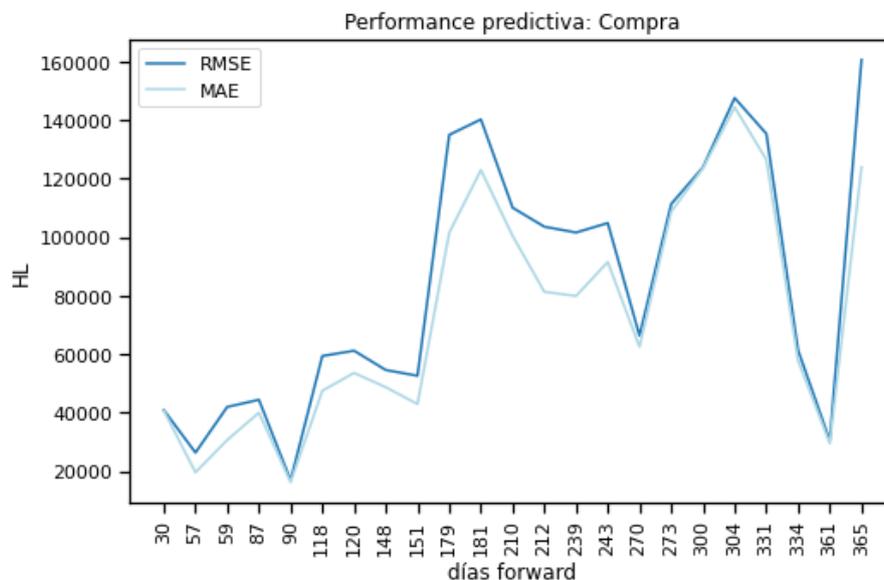


El pronóstico del canal supermercado muestra una tendencia a la baja para los próximos 5 años, con las marcadas curvas de la estacionalidad.

En los últimos 2 años de data real, se visualiza un aumento de ventas en temporada baja (junio y julio) y una disminución en los meses fuertes (diciembre y enero), pero un crecimiento en valores absolutos entre el 2018 y 2019. A partir del 2020, con la aparición de la pandemia, notamos una caída muy fuerte en los momentos más críticos de la pandemia (junio y julio 2020) y un tímido incremento del volumen de compra para fines de ese año. Si bien los supermercados fueron uno de los pocos establecimientos que quedaron abiertos durante todo el año 2020, no fue suficiente para igualar la venta de los años anteriores.

Compra

Diagnóstico del modelo



En el eje x se muestran “días forward”, que corresponden a los días de horizonte establecidos para realizar el diagnóstico de predicción del modelo. La performance de este modelo es muy buena para los primeros cuatro meses, pero luego va disminuyendo su rendimiento a medida que se avanza en el horizonte de estudio.

Predicción

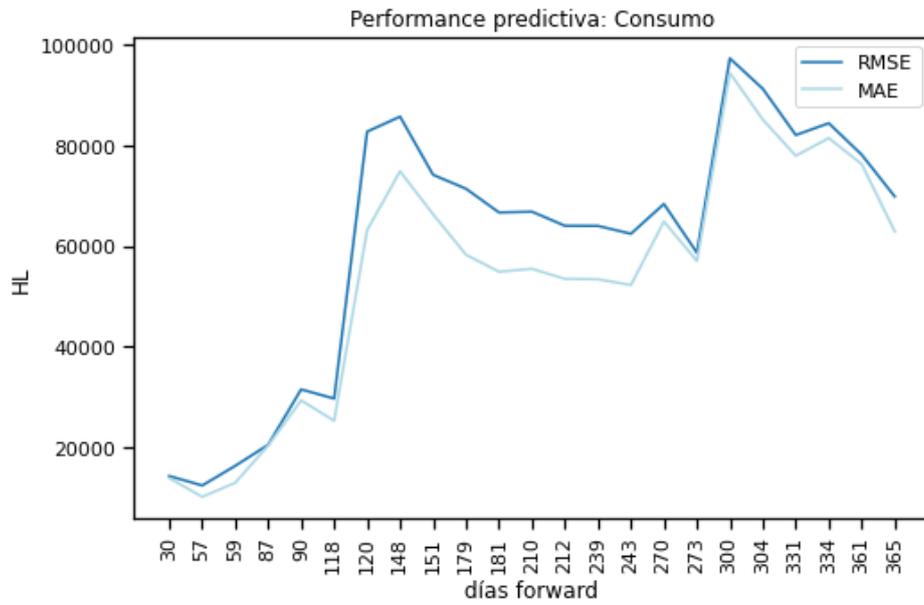


Como sucede con la categoría de Vinos, hay que ser prudentes a la hora de analizar este gráfico de predicción ya que el error del modelo, crece a medida que se extiende el período de observado.

De todas maneras, esta predicción tiene sentido, ya que las ventas en el canal supermercado caen y es posible que las mismas sean sustituidas por compras en otros puntos o comercios más pequeños y con menor circulación de gente. Además de la pandemia, el cambio de estrategia que tuvo la compañía a fines del 2019, influye también en la disminución del volumen volcado a este canal.

Consumo

Diagnóstico del modelo



En el eje x se muestran “días forward”, que corresponden a los días de horizonte establecidos para realizar el diagnóstico de predicción del modelo. Para el canal consumo, se ve que el modelo tiene una buena performance para los primeros 90 días y luego empeora.

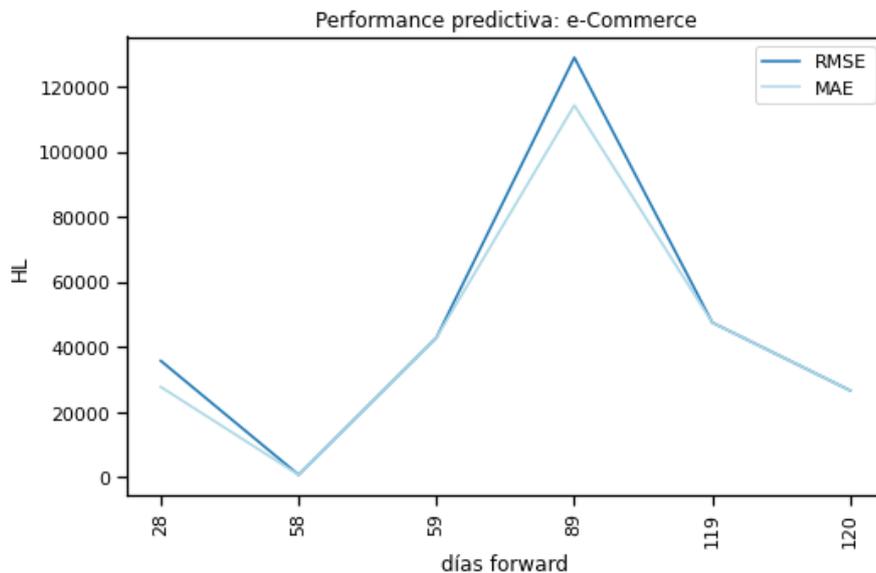
Predicción



Este canal, Consumo, es uno de los más afectados por la pandemia y las predicciones no son muy optimistas. A pesar de la performance del diagnóstico del modelo, creemos que el pronóstico observado para este canal es acertado. Si bien, podemos ver algún repunte en los momentos más laxos de la pandemia, la sociedad en general se volvió más consciente en muchos aspectos sobre todo respecto a la higiene, al distanciamiento y a los gastos económicos.

e-Commerce

Diagnóstico del modelo



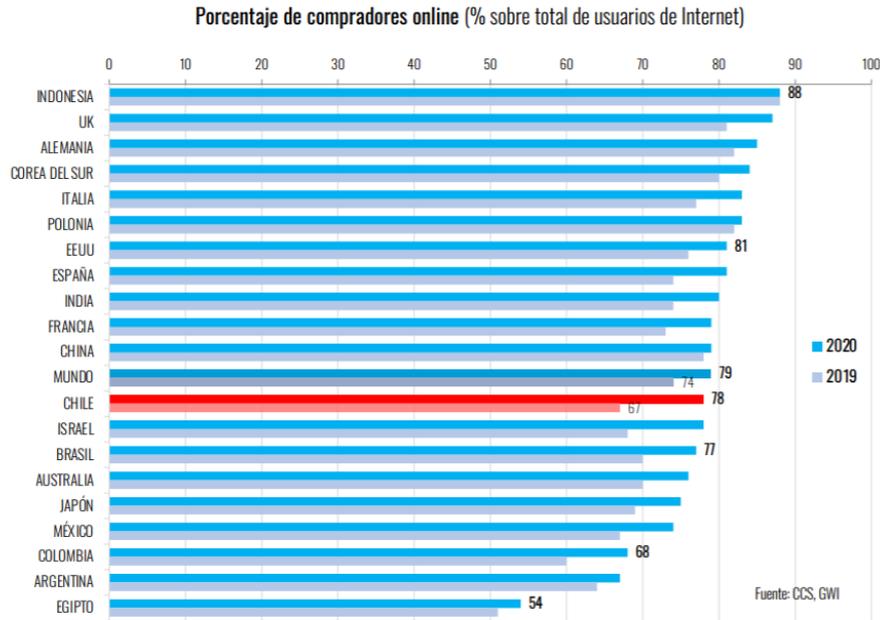
En el eje x se muestran “días forward”, que corresponden a los días de horizonte establecidos para realizar el diagnóstico de predicción del modelo. La performance de este modelo es buena los primeros dos meses, luego, el error aumenta significativamente, pero vuelve a caer cuando se calcula para el horizonte definido. Debido a que la compañía comenzó a registrar ventas en el canal e-Commerce a finales del 2019, el período inicial que se tomó para el diagnóstico del modelo fue más corto que el resto, pero aseguramos que el análisis es representativo.

Predicción

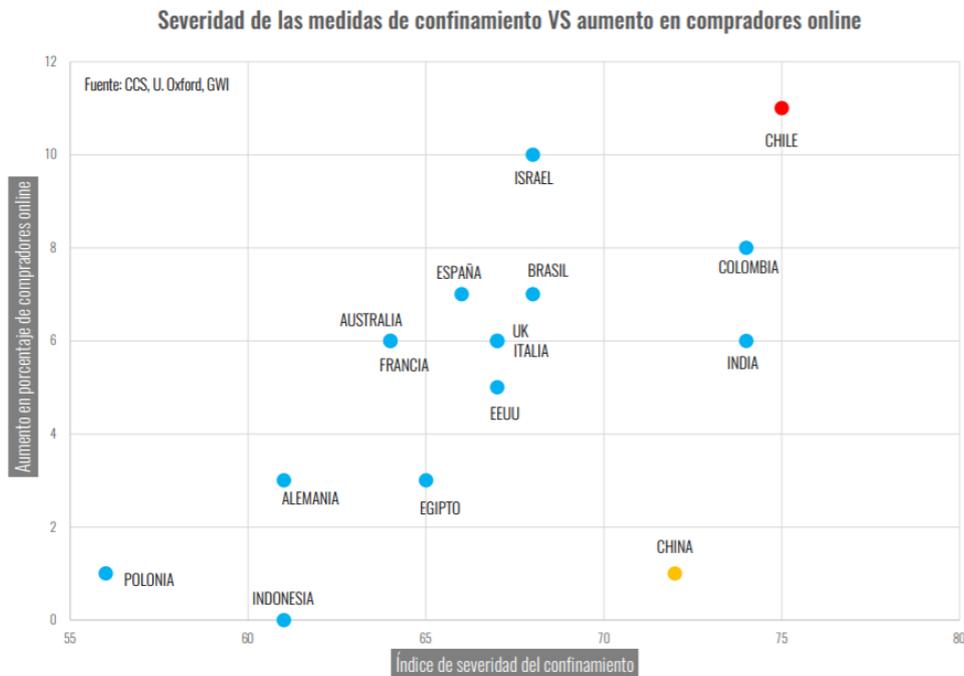


Asociado con la caída en el volumen en el canal supermercado, está el incremento en este canal. El e-Commerce o comercio electrónico, ha sido una salvación para muchos negocios que se vieron perjudicados con la caída en sus ventas presenciales por las medidas de distanciamiento social, además de una alternativa para que muchos consumidores puedan seguir realizando sus compras habituales de forma segura.

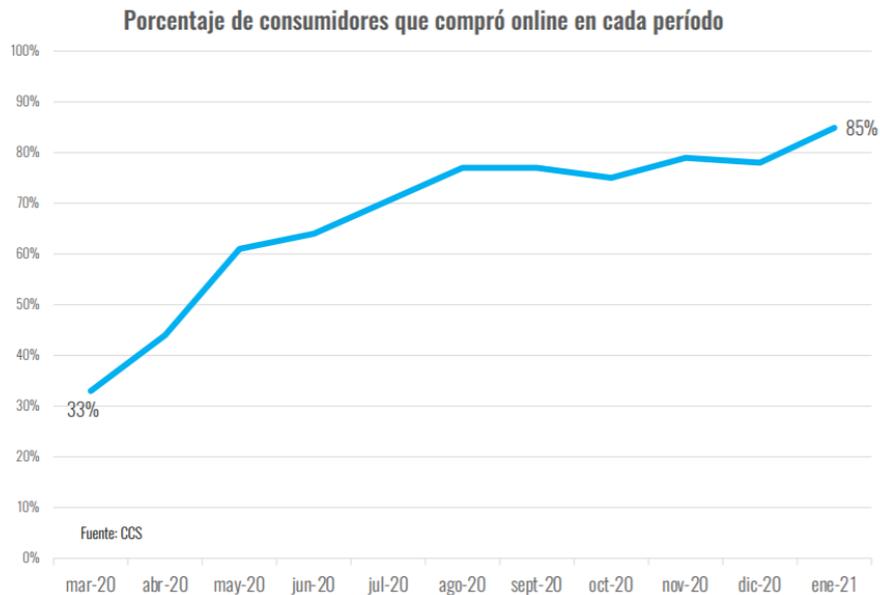
De acuerdo a un estudio realizado por la Cámara de Comercio de Santiago, el porcentaje de compradores online aumentó 11% en el 2020, uno de los mayores incrementos mundiales, seguido por Israel, Colombia, Brasil y España.



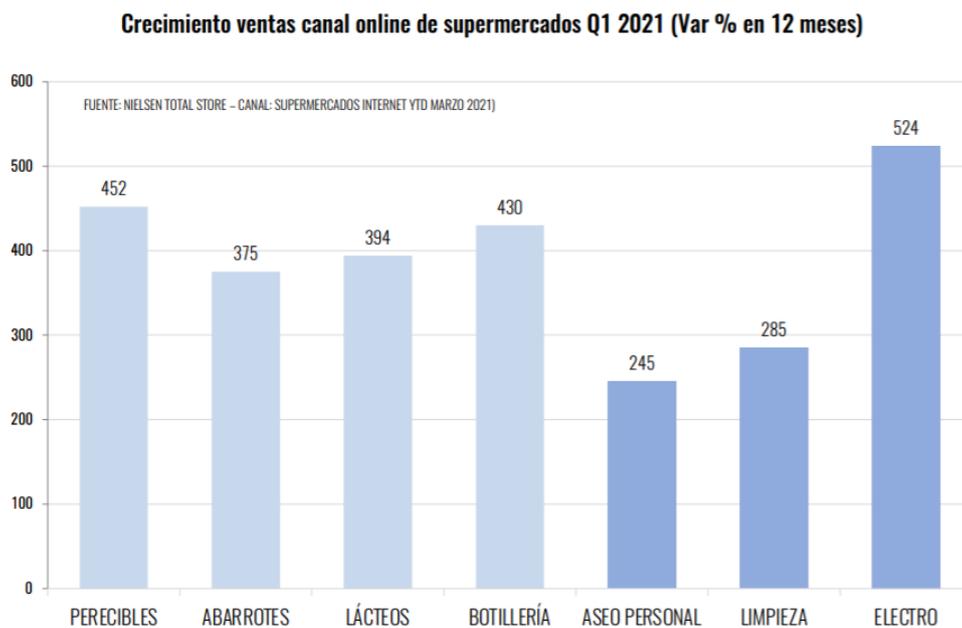
Si bien Chile fue uno de los países que impuso medidas más estrictas de aislamiento y controles más rigurosos, fue también el que más creció en compradores online si se lo compara con otros países con medidas de confinamientos severos durante todo el 2020.



Análogamente como se mostró en el apartado del análisis exploratorio de datos, en este estudio, también se puede ver el crecimiento que tuvo el canal a lo largo de todo el 2020 con un incremento sostenido en el período más complejo y estricto de la pandemia, que fue de mayo a agosto.

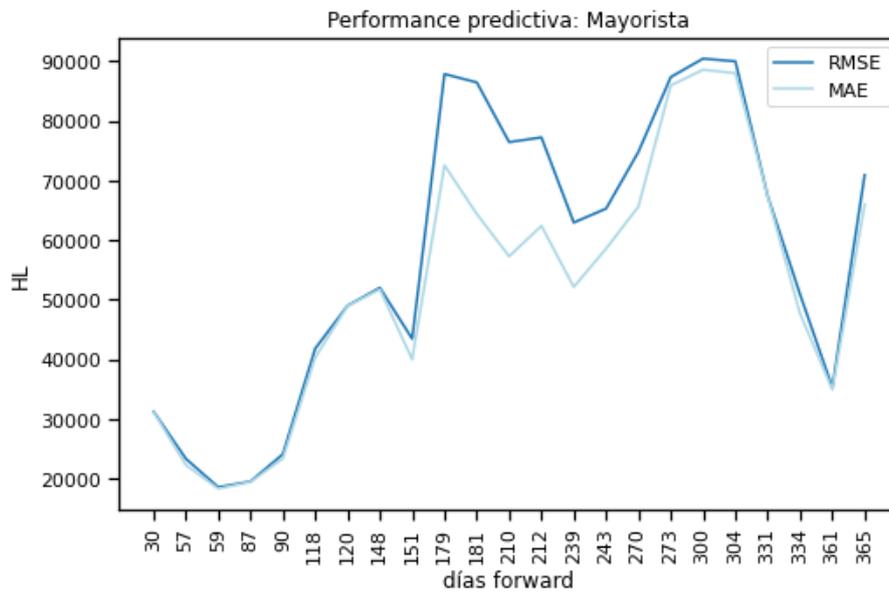


Se observa, también que la categoría de bebidas alcohólicas “Botillería” es una de las que más crecimiento tuvo en las ventas del canal online de supermercados, luego de la categoría de “Electro” y “Perecibles”



Mayorista

Diagnóstico del modelo



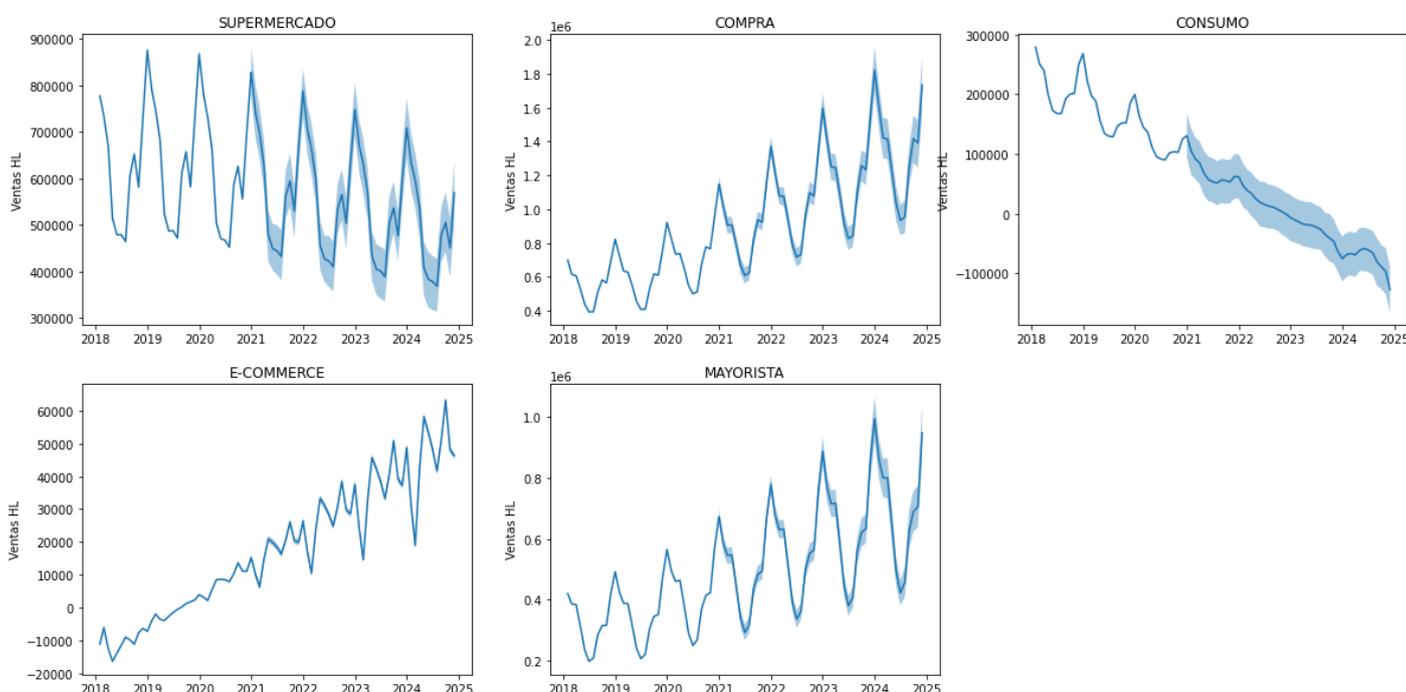
En el eje x se muestran “días forward”, que corresponden a los días de horizonte establecidos para realizar el diagnóstico de predicción del modelo. La performance de este modelo para el canal mayorista, es muy buena y aceptable para los primeros 3 meses, luego empeora.

Predicción



El canal mayorista fue el menos impactado por la pandemia. El mismo venía creciendo a un ritmo del 10%, 15% y el mismo se mantuvo durante todo el 2020. De acuerdo a la predicción arrojada por el modelo Prophet, se observa que el año 2021 es el de menor crecimiento para este canal, pero luego a partir del 2022 recupera y se estabilizan las tasas de crecimiento que acompañan a la industria en general.

Predicción de ventas totales por Canal



Indudablemente, existen dos grandes “perdedores” y dos grandes “ganadores” cuando se habla de los canales de venta o consumo. Dentro del primer grupo, se encuentran los canales Supermercado y Consumo (bares, restaurantes, etc) que sin duda fueron los más golpeados por las medidas restrictivas, la sensación de vulnerabilidad de la gente y la pandemia en sí. La caída del volumen en los mismos era previsible si se observaba el comportamiento de los países más “avanzados” (en cuestión de transición y manejo del COVID-19), donde los espacios en el que el tráfico de gente era considerado masivo o los lugares donde la gente se

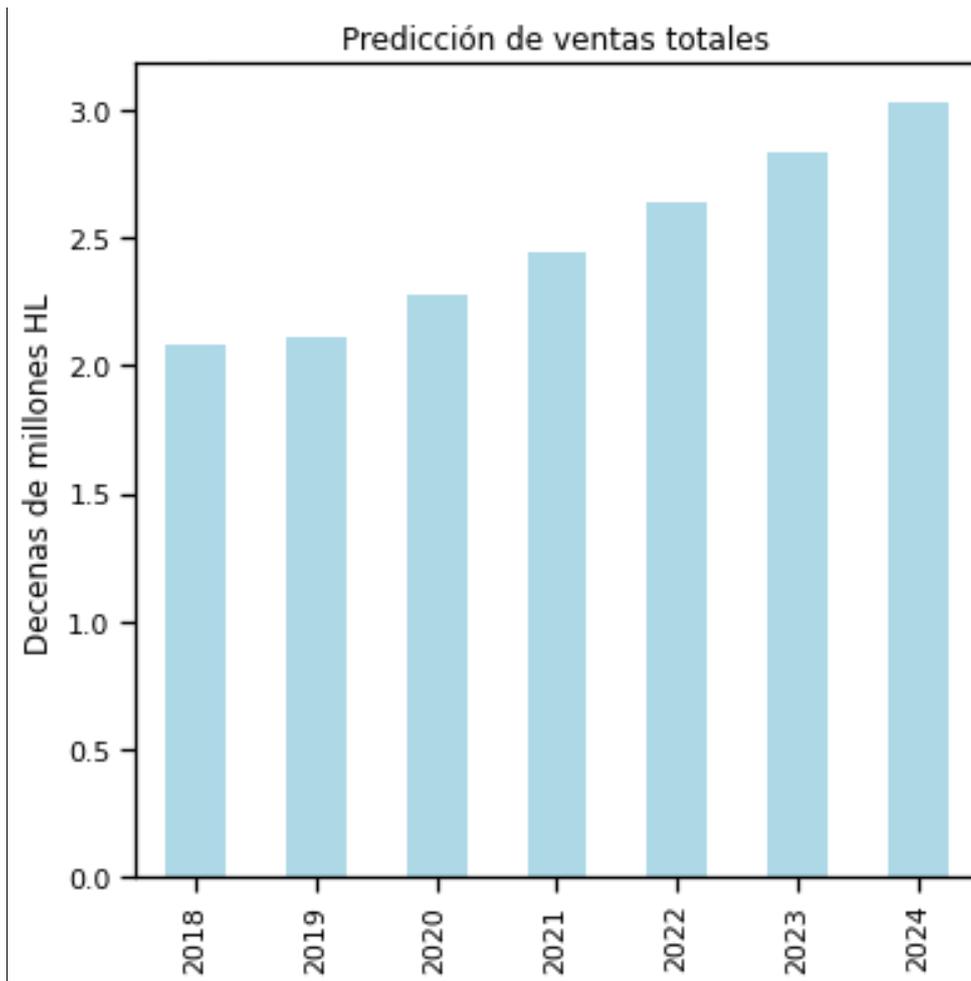
sentía expuesta al contagio, fueron cerrados o vieron una disminución desmedida de la circulación.

Del otro lado, se encuentran los “ganadores” de la pandemia, que se convirtieron en una especie de “lugar seguro”, ya sea por la baja circulación de gente o por la comodidad para realizar la operación de compra. Estos son el canal “Compra” o “Tradicional” que incluye a los almacenes o botillerías y “e-Commerce” que contempla todas las transacciones que se realizan de manera online. Más allá del cambio de estrategia que tuvo la compañía en el canal tradicional a fines del 2019 y del incremento y evolución del canal online de la mano de la tecnología, la pandemia aceleró y ayudó a establecer anticipadamente, un proceso de cambio de hábito que ya venía experimentando el consumidor.

Estas conclusiones, están en línea con el artículo “Consumo de hogares chilenos empieza a estabilizarse” realizado por la consultora *Kantar*. En el mismo, aseguran que el atributo “cercanía”, es uno de los más relevantes para los chilenos a la hora de elegir dónde hacer sus compras. En canal tradicional, se observa un aumento del 25% en el consumo para el tercer trimestre del año, representando un 32% del gasto de la canasta, en comparación con el 28% en el mismo periodo del año anterior. En cuanto al canal online, indican que continúa adquiriendo relevancia, duplicando su participación versus año anterior pasando del 1% al 2% y con una penetración del 10%

Predicción de ventas totales

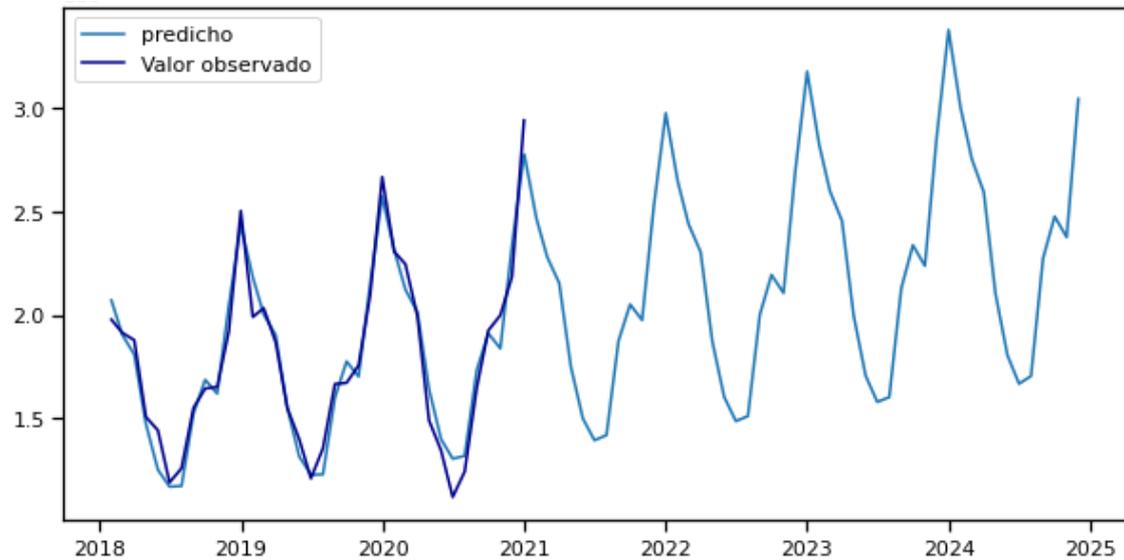
Luego de realizar cada una de las predicciones por Grupo de Productos y por Canales, unimos toda la información para obtener la venta total final para el año 2021, 2022, 2023 y 2024. A continuación, se muestran los resultados:



Más allá de los cambios de hábito de los consumidores, podemos observar que la tendencia que se venía viendo creciente en el volumen de ventas entre los años 2018 y 2019, continúa al alza, pudiendo concluir que la baja del consumo en los canales Supermercado y Consumo son compensadas por el crecimiento que tiene y seguirá teniendo el canal online y tradicional.

Asimismo, ya vimos que el consumo de casi todas las bebidas alcohólicas aumenta considerablemente, mientras que las bebidas analcohólicas se mantienen en sus valores históricos con un modesto aumento de volumen.

Se puede decir que la performance del modelo es óptima y confiable ya que la predicción para los años 2018, 2019 y 2020 es muy aproximada a los Valores Observados del dataset provisto:



Lineamientos Estratégicos y Plan de Acción: recomendaciones

Una vez analizado el dataset y las predicciones, podemos llegar a la conclusión descrita en el apartado anterior. Pero ahora, ¿cómo pueden las empresas aprovechar esta herramienta y toda esta información para tomar decisiones?

El objetivo de esta última sección, es sugerir algunos lineamientos estratégicos, basados en la industria de bebidas de Chile, para que puedan explotar al máximo este trabajo.

Comenzamos con la definición de Estrategia. De acuerdo a *Alfred D. Chandler Jr*, en su libro *“Strategy and Structure”*, la misma, consiste en la determinación de los objetivos y metas a largo plazo de una empresa, la adopción de cursos de acción, y la asignación de los recursos necesarios para llevar a cabo esas metas establecidas.

La planificación estratégica es un proceso que se debe dividir en etapas para poder alcanzar los objetivos, comenzando con el análisis de la situación actual, siguiendo por la definición de los lineamientos estratégicos y, por último, desarrollando un plan de acción. En cada etapa o fase, nos centraremos en la evaluación de distintas cuestiones clave.

Para poder dar un marco óptimo a lo que vamos a analizar siempre es importante poder determinar el universo sobre el cuál vamos a hacer el

análisis. Vamos a tomar a Chile como el territorio sobre el cuál plantearemos esta estrategia, y analizaremos el consumo masivo de bebidas, como sub segmento. Haremos foco por supuesto en los mercados actuales sobre los que se paran estas empresas, pero teniendo alguna perspectiva sobre posibles innovaciones, oportunidades y cambios, que venían dándose en esta industria, y como el estudio anterior puede afectar en esto.

Análisis situación actual

En primer lugar, se debe realizar una evaluación del entorno externo. Para esto, nos apoyaremos en el modelo de las Fuerzas de Porter, desarrollado en su libro “Ser Competitivo”, ya que nos permitirá analizar la intensidad competitiva de la industria. Detallamos a continuación cada fuerza:

- Potenciales entrantes: si bien en los últimos años han aparecido diversos productores de cerveza artesanal o bodegas “boutique”, estos no representan una amenaza fuerte, ya que para ingresar a esta categoría de bebidas es necesario, en primer lugar, tener el capital necesario para invertir, en segundo lugar, lograr economías de escala y por último poseer un fácil acceso a los canales de distribución. Pero lo que sí puede presentar una potencial amenaza, es la fusión o compra del principal competidor (AB InBev) con una cervecera o bodega más chica y la expansión de la misma en el canal *online*. Como se describe previamente ese es análisis del mercado actual, donde se encuentra el otro gran jugador mundial dentro del territorio Chileno. Pero no hay que perder de perspectiva la posibilidad que ofrece un país como Chile, que fomenta la competencia de una manera más abierta que otros países de la región. La posible apertura a otros jugadores mundiales, a través de importaciones sin grandes sobrecostos, es una amenaza latente, teniendo en cuenta las reglamentaciones puestas por el gobierno chileno.
- Clientes: hoy los usuarios cuentan con mayor poder porque tienen más información de manera inmediata, por lo que cualquier desacierto de parte de CCU, puede costar la pérdida de clientes a costa de la competencia. Es ya sabido que conseguir un nuevo cliente es mucho más caro que retener a uno existente.

- Proveedores: en este sentido, la compañía cuenta con un gran poder de negociación con los proveedores ya que los insumos utilizados son, en el mayor de los casos, *commodities*.
- Productos Sustitutos: a pesar de la participación de mercado con la que cuenta la empresa, la intensidad competitiva en esta materia es considerable, ya que existen muchos productos similares por los que pueden ser reemplazados las bebidas que comercializa CCU.
- Rivalidad competitiva: en este punto, CCU, cuenta con una gran ventaja competitiva, ya que posee más de un 70% del market share de las bebidas, en particular de las cervezas. De todas formas, la amenaza del competidor está latente.

En segundo lugar, se debe realizar un análisis interno de la compañía, para entender en qué parte de la cadena de valor encontramos las mayores fuentes de diferenciación versus la competencia. En este punto, creemos oportuno desarrollar un análisis FODA para poder entender la situación interna de la empresa teniendo en cuenta, también, la situación externa que comentamos en detalle anteriormente:

- Fortalezas: CCU es el principal productor y distribuidor de diversas categorías de bebidas en Chile, con más de un 70% del *market share*. Asimismo, cuenta la fuerza de distribución con la que cuenta es propia, permitiéndole llegar a todo el país con economías de escala. Es también “dueño” del portfolio con las mejores marcas a nivel internacional a lo largo de todas sus categorías.
- Oportunidades: establecerse como el jugador principal en el canal *online*, tomando el impulso proporcionado por la pandemia y la cuarentena obligatoria. Además de esto, es un buen momento para nuevos lanzamientos o innovaciones con el objetivo de no perder a los nuevos clientes que son cada vez más exigentes y versátiles. De esta forma, consolida y ratifica su liderazgo en el mercado.
- Debilidad: ausencia de gestión del área “Business Inteligente”, que se dedique realizar investigaciones y análisis de datos, tanto de la compañía como de la competencia, para la toma de decisiones. Este departamento, es sumamente relevante y necesario para poder alinear de los procesos de Marketing, ventas y atención al cliente con las necesidades actuales y futuras de los clientes, de manera de

aumentar la participación del mercado, el *share of wallet* y la retención de clientes.

- Amenaza: su principal competidor, AB InBev, es la cervecera más fuerte en la región y cuentan con una gran soporte y análisis de buenas prácticas que implementaron en el país exitosamente. Esto se puede observar, en particular, en el canal *online* donde tienen una gran ventaja competitiva en cuanto a estrategias digitales.

Las Competencias Clave son aquellas capacidades que confieren a la organización una fortaleza competitiva única. Son las capacidades distinguidas de una organización, que confieren un impacto competitivo único y durable. Por este motivo, creemos útil también, detallar las competencias clave de la compañía:

- Líder del mercado de bebidas en Chile
- Logística propia
- Producción de envases retornables y no retornables
- Liderazgo en costos
- Amplio portafolio de marcas

Lineamientos estratégicos

Una vez que tenemos el análisis externo e interno, podemos avanzar con la definición de los lineamientos estratégicos. Para esto, es fundamental evaluar las oportunidades que tiene la organización, con el siguiente proceso:

- Entender el mercado, la propia organización y lo que buscan los clientes
- Generar oportunidades de alto nivel
- Evaluación a alto nivel
- Evaluación de alternativas estratégicas
- Evaluación en más detalle
- Priorización de las oportunidades

Para buscar las oportunidades que tiene la organización nos podemos basar en las siguientes herramientas:

Matriz Producto/Mercado

Esta matriz, nos servirá para la evaluación de las diferentes oportunidades.

	Productos Actuales	Nuevos Productos
Mercados Actuales	Cervezas sin alcohol Bebidas sin TACC Cervezas artesanales	Cervezas cannábicas Cervezas con baja graduación alcohólica Bebidas <i>ready to drink</i>
Nuevos Mercados/ Segmentos	Cervezas con maltas proteicas sin alcohol Cervezas bajas calorías Radler sin alcohol	Bebidas innovadoras (en todo su proceso de producción y distribución)

En el cuadrante “Productos Actuales – Nuevos Mercados/Segmentos” colocamos las cervezas con malta proteica sin alcohol, bajas calorías y Radler sin alcohol, que si bien ya existen algunos productos con estas especificaciones, creemos que aún hay mucho por desarrollar. Es un nuevo mercado en crecimiento y con cada vez más necesidades, orientado a los consumidores deportistas o que buscan el cuidado de la salud.

En cuanto al cuadrante de “Nuevos Productos – Mercados Actuales”, encontramos que existe una gran oportunidad para desarrollar nuevos productos como las cervezas cannábicas, con baja graduación alcohólica y las bebidas *ready to drink* (listas para tomar). Las bebidas cannábicas están tomando cada vez mayor relevancia y son una gran promesa para la industria. De acuerdo a un informe de Fortune Business Insights, se proyecta que la industria crecerá más de un 1.000% para el año 2027. Esto dependerá mucho del contexto del país en relación a la regulación en el consumo de marihuana, pero creemos que es una gran oportunidad de desarrollar un nuevo producto.

Por último, en “Nuevos Productos – Nuevos Mercados/Segmentos” se encuentran las Bebidas innovadoras como por ejemplo bebidas como el Vodka producido con ethanol derivado de las emisiones de dióxido de carbono, vino en lata o bebidas realizadas con una técnica cero desechos.

A medida que la pandemia transforma nuestros hábitos de compra y alimentación, es muy importante para las empresas mantenerse a la vanguardia y estar al tanto del cambiante y sofisticado gusto de los consumidores. Al mismo momento en que la pandemia estimuló el consumo de alcohol, en contraste también ha reforzado el enfoque en la salud y en el cuidado del medio ambiente. Así lo confirma una investigación de la agencia *FMCG Gurus*, en la que han demostrado que el 60% de los consumidores ahora está más atento al impacto que sus alimentos y bebidas tienen en el medio ambiente. Además, el 56% indica que presta más atención a los ingredientes naturales que contiene el producto al realizar sus compras.

De acuerdo a lo expuesto anteriormente, las empresas deberán buscar la forma en que sus nuevos productos sean lo más transparentes posibles a lo largo de toda su cadena de producción y distribución.

Matriz BCG

Una vez que tenemos evaluadas y definidas las prioridades de las oportunidades, las analizamos con esta matriz de gestión de portafolios. En la misma se observan 4 caminos distintos a tomar en base a la participación del mercado y tasa de crecimiento en el mismo de los productos a analizar.



Con todo lo dicho anteriormente, podemos posicionar las oportunidades de la siguiente forma:

- Desarrollar (alta Tasa de crecimiento del mercado y alta Participación del mercado): esta situación es la más beneficiosa para las compañías. Colocamos en esta sección a las cervezas bajas calorías y cervezas artesanales. Creemos que hay mucho camino por recorrer en ambas subcategorías.
- Ordeñar (baja Tasa de crecimiento del mercado y alta Participación del mercado): en este cuadrante podemos situar a las cervezas tradicionales Premium que la compañía ya comercializa, como por ejemplo la marca Heineken. Las mismas llevan una gran porción del mercado, pero las proyecciones de crecimiento de mercado no son al alza.
- Invertir (alta Tasa de crecimiento del mercado y baja Participación del mercado): en este cuadrante entran todas las oportunidades que tienen gran potencial como por ejemplo las cervezas cannábicas y las bebidas innovadoras.
- Salirse (baja Tasa de crecimiento del mercado y baja Participación del mercado): acá situamos a las cervezas de bajo precio. Si bien la recomendación no es salirse del mercado, sí creemos que no es necesario una inversión para tener más desarrollo en este segmento.

Estrategias competitivas: Plan de acción

El objetivo de esta última sección, es poder detallar cómo sería un principio de plan de acción tomando todos los lineamientos previamente analizados. Vamos a seguir la estructura de definición de segmentos objetivos y detectar ventajas competitivas sostenibles para poder sustentar esta estrategia.

Al estar analizando el sector de consumo masivo, lo evidente, es sugerir abarcar la totalidad de los segmentos del mercado con los productos que ya existen y con los que se debe innovar. Por este motivo, a pesar de que la recomendación gira en torno a desarrollarse en todos los segmentos del mercado para llegar a todos los consumidores posibles, hay algunos en los que se recomienda hacer un mayor foco.

La idea que se plantea, en primer lugar, es segmentar el mercado cruzando parámetros de edad, nivel socioeconómico y tomar algunas variables actitudinales para luego saber cómo debemos comunicar a estos consumidores.

El mayor desafío que enfrenta la compañía, se encuentra en el segmento de los “consumidores innovadores”. Éstos son los que están abiertos a testear nuevos productos, que tienen una alta tolerancia hacia la prueba de nuevas cosas y no necesariamente son leales con una marca en particular. Este segmento, suele estar asociado con consumidores más jóvenes, de hasta 40 años, con un nivel socioeconómico AB y C1a (un poco más alto que la media), y que tienen mayor acceso a la información. Este segmento es el que genera los mayores desafíos, pero son quienes logran adoptar de una manera más rápida las tendencias, y en quienes se debe enfocar en esta primera instancia.

Asimismo, también se debe poner foco en un segmento según el tipo de comportamiento, que valoran lo saludable y son plenamente conscientes de los productos que consumen. Esta es una tendencia mundial en todas las categorías, y dada la oportunidad que existe en el futuro crecimiento de las “cervezas” no debe ser ajeno a esta conclusión. En este segmento, predomina un público más joven, con fácil acceso a dispositivos móviles, Internet y nuevas tecnologías. Con un gran interés por los productos saludables y el medio ambiente.

Un último segmento que también se vincula con los segmentos anteriores, es el público femenino, que históricamente tiene menor penetración en el mercado de las cervezas y los vinos, pero con una tendencia creciente, y con propuestas que puedan valorarse por el gran potencial de crecimiento.

El propósito la compañía debe ser enfocarse plenamente y apuntar a los tres segmentos descritos anteriormente con distintas propuestas de valor que permitan tener un diferencial de manera sostenible y a lo largo del tiempo. Una vez determinado el “hacia dónde” se deben enfocar, luego el objetivo es definir el “cómo”. Nuestra sugerencia, la desarrollamos en base a las ventajas competitivas, siempre alineadas con las capacidades diferenciales de la organización.

En lo que respecta al segmento de “*early adopters*” o “innovadores”; se debe tener una concepción de empresa que sea innovadora en todos los

aspectos: no sólo con respecto al consumidor, sino también en la concepción interna del este objetivo. Trabajar con flexibilidad, motivando la prueba y los posibles errores, con un equipo industrial dedicado y predispuesto, seguramente con un análisis de pérdida de eficiencia, pero con la posibilidad de desarrollar nuevos productos y seducir a este segmento. Es probable que la apuesta en este sentido, haga descender algunos KPI claves del sector industrial, pero debe ser aceptado para lograr los resultados planteados. De cara al consumidor, se debe priorizar las variedades de cervezas (de sabores, nuevas marcas, etc.), que generen la constante novedad en los distintos segmentos de precio.

En cuanto al segmento saludable y femenino, también hay un desafío alrededor de porqué los clientes elegirían a un jugador y no a otro. Un punto clave acá es ser “el primero” y tener una comunicación acorde a las necesidades de estos segmentos. Alineado con el objetivo de llegar al segmento “*early adopters*”, la compañía debe estar preparada para ofrecer una gama de productos innovadores (cerveza sin alcohol, mezcladas con fruta, maltas proteicas, etc.), que puedan responder a distintas ocasiones de consumo de estos consumidores. Se pueden pensar desde propuestas que puedan ayudar al rendimiento deportivo, como otras que permitan la relajación y el momento de esparcimiento con productos más saludables y de bajas calorías o baja graduación alcohólica. Las alternativas conocidas en el mercado de estos segmentos son pocas, es por eso que nuestra principal sugerencia es la de innovar. Y como se describe más arriba, el trabajo de innovar no es sólo una responsabilidad del área de “innovación”, sino también que deben estar involucrados otros sectores como por ejemplo industrial, planeamiento, finanzas y marketing.

En resumidas cuentas, para llegar a todos estos segmentos, la propuesta de valor debe estar enfocada en el concepto de “ventaja en diferenciación”, donde todos los integrantes de la compañía tienen un rol muy importante, deben estar comprometidos con dicho objetivo y creer profundamente en su trabajo.

Etapa Prescriptiva: Conclusión

El presente trabajo tiene por objetivo lograr una mayor comprensión del cambio en los hábitos de consumo ante algún evento cualquiera que modifique de manera forzosa la vida de las personas. Concretamente, se tomó el reciente acontecimiento que afectó y sigue afectando a toda la población mundial (pandemia de Covid-19), para poder anticipar estos cambios y que sea de utilidad a las empresas, en particular, de consumo masivo para poder actuar sobre planes de producción, objetivos estimados y demás.

A partir de un dataset facilitado por una de las compañías más grandes de bebidas de Chile, con los datos de ventas de los años 2018, 2019 y 2020, se realizó un análisis exploratorio de todas las variables contenidas en esos dataset. Esto nos dio visibilidad del comportamiento de las variables, a priori, más importantes para realizar este trabajo: Grupo de Productos (bebidas analcohólicas, cervezas, vinos, licores, etc), Marcas, Empaques (calibres), Canales, Regiones, y el feature por el que las empresas de este tipo miden los objetivos anuales: Venta por Hectolitro. Asimismo, graficamos algunas matrices de correlación para entender la conducta de los consumidores en los Grupos de Producto y las marcas de Cerveza que nos dieron algunos insights muy interesantes que tiene que ver sobre todo con el precio y estatus que da el consumo de cada categoría y marca.

Luego de analizar todas esas variables e interpretar el comportamiento de ventas a lo largo de los tres años, se aplicó un algoritmo de clustering sobre el dataset para poder identificar los patrones de comportamiento de compra. A partir de esto, se reconocen tres grandes grupos bien diferenciados:

- Ventas Altas: vemos que dentro de este cluster se encuentran las bebidas de consumo cotidiano como las Bebidas Analcohólicas y en cuanto a las marcas de cerveza, las que tienen mayor volumen de consumo y son más masivas como Escudo y Cristal.
- Ventas Medias: aquí se encuentran las bebidas de consumo social con alcohol: las cervezas. Respecto a las marcas, vemos que aparecen algunas marcas nacionales y otras internacionales que son más populares.

- Ventas Bajas: se pueden detectar todas las bebidas de consumo ocasional como los jugos o bebidas con mayor graduación alcohólica como los Licores, Vinos y Bebidas Blancas. Dentro de las cervezas encontramos algunas internacionales con un precio alto o nacionales con precio muy bajo que tienen poco volumen.

Ya con toda la información de ventas históricas desmenuzada y analizada, aplicamos un modelo de predicción de series de tiempo, creado por el equipo de Facebook. Se decidió elegir este modelo porque el mismo descompone las series de tiempo en tendencia, estacionalidad y vacaciones. Para la industria de bebidas la estacionalidad es el factor más importante ya que en temporada alta (verano) el volumen aumenta considerablemente y en temporada baja el mismo disminuye.

En los últimos años se ha observado una tendencia creciente en el consumo de bebidas sobre todo en las bebidas Analcohólicas, Cervezas y Licores, pero el último año analizado contamos con un suceso histórico que nos obligó a modificar nuestras conductas y logró un cambio en los hábitos y tendencias que modificaron y seguirán modificando la forma de consumir que tienen las personas, según podemos observar de las predicciones.

De los gráficos de predicción, podemos concluir que las bebidas alcohólicas comienzan a tener una mayor consideración en el momento de la compra, dejando atrás las bebidas sin alcohol como las gaseosas, aguas y jugos. Creemos que este comportamiento se da porque luego de estar más de un año aislados y con poco contacto directo con familiares, amigos, conocidos y pares, las personas sienten una necesidad imperiosa de sociabilizar, recuperar momentos perdidos, disfrutar y compartir la mayor parte de su tiempo. Para este tipo de acontecimientos o eventos, la bebida alcohólica como la Cerveza o el Vino es central porque invita a compartir y genera un ambiente de fiesta o celebración.

Tal como indica un estudio realizado por *Kantar* en el tercer trimestre del 2020, la pandemia y cuarentena en Chile, impactaron en el valor de la industria cervecera aumentando un 4% versus año anterior. Indican que las compras online y la necesidad de vida social, aunque sea a través de pantallas, fue un soporte para que el consumo no cayera. De acuerdo al

Barómetro COVID-19 (Ola 3 de Kantar), un 22% de los chilenos afirmaron estar comprando más alcohol que en el período pre-pandemia.

Del mismo modo, se concluye que la pandemia llegó para acelerar un proceso que se venía gestando y evolucionando a un ritmo muy lento, sobretodo en Latinoamérica. Este proceso, es el desarrollo del canal *online* o e-Commerce. Este canal es el que se llevará la mayor proporción del crecimiento del volumen en los próximos 3 años, por lo que las compañías deberán estar preparadas para enfrentarse a un consumidor más exigente en cuanto a inmediatez y experiencia de compra online.

En resumen, como se ha comentado anteriormente, el presente estudio tiene como objetivo proveer a las compañías una herramienta para tomar más y mejores decisiones de negocios.

En primer lugar, esta investigación puede eventualmente ayudar a las empresas a pensar en sus datos o información como lo que es: su activo máspreciado. En el siglo XXI, parece algo absurdo pensar que hay compañías que aún no utilizan su información para tomar decisiones, pero es más común de lo que parece. Como indican muchos gurús de la tecnología: “la información es poder”, y las compañías deberían utilizar y aprovechar la misma para su propio beneficio. Para poder poner en practica la técnica o herramienta expuesta en este trabajo, es fundamental que las empresas cuenten con toda su información histórica lo más detallada posible.

Por otra parte, los resultados de este trabajo también pueden ser útiles para la reformulación o el refuerzo de los planes para los próximos 3 años de las compañías de consumo masivo, en particular de bebidas. Las mismas deberán poner todos los esfuerzos en los canales Tradicionales y Online, al mismo tiempo que deberán pensar en desarrollar aún más la categoría de bebidas alcohólicas, pensadas para un público más joven. Deberán entender también, que todas las empresas están poniendo sus esfuerzos allí también, por lo que las opciones que tendrán los consumidores serán abundantes y deberán esforzarse para ser los mejores en su clase.

La nueva generación de consumidores es más exigente en sus experiencias digitales que en el plano físico, no son leales a una marca, les molesta perder el tiempo buscando un producto de una marca determinada o

directamente no encontrarlo. En el canal online, es muy sencillo encontrar fácilmente productos de la competencia y una mala experiencia que tenga cualquier consumidor, puede afectar negativamente el negocio. Por estos motivos, el “cambio de mentalidad” que deben hacer las compañías es evidente, necesario y urgente.

Finalmente, al momento de realizar este trabajo, no se encontraron demasiadas investigaciones que analicen las implicancias de la pandemia (en principio, porque en muchos países aún existen picos de contagios y muertes por COVID-19), por lo que puede considerarse como uno de los primeros intentos de desarrollo de herramientas para predecir cómo evolucionará y cambiarán los hábitos de los consumidores luego de vivir un evento fortuito y de tal magnitud, que obligó a modificar muchos aspectos de la vida cotidiana de la población mundial.

Limitaciones

Al analizar los dataset provistos, nos encontramos con mucha información no relevante para el desarrollo de este trabajo que hemos desechado. Por el contrario, se han incluido algunos features que, en principio, se pensó que podría darnos ciertos insights interesantes, pero luego se demostró lo contrario. Un ejemplo de esto, es el detalle de la venta por región. En primer lugar, detectamos que la información descripta allí incluía información de regiones y canales por igual. Esto se debe a cómo se organiza internamente la empresa, pero a los fines del análisis de este trabajo, no nos servía. En segundo lugar, al analizar la información por región vemos que el comportamiento de los consumidores entre ciudades o regiones era muy similar. Por un lado, se pueden obtener algunas conclusiones obvias, como por ejemplo que las regiones más grandes tienen la mayor concentración de ventas. Por otro lado, se ve claramente la tendencia a la baja del canal supermercado (ya que se toma como “región”) y un aumento en la venta del canal Mayorista.

Referencias bibliográficas

Kotler P. & Keller K. (2014) Marketing Management. Pearson. 160-165.

Bernard Berelson y Gary A. Steiner (1964). Human Behavior: An Inventory of Scientific Findings. Harcourt Brace Jovanovich. 90-95.

José A. Mauricio (2007). Introducción al Análisis de Series Temporales. Universidad Complutense de Madrid.

Alfred D. Chandler Jr. (1961). Strategy and Structure: Chapters in the History of the Industrial Enterprise. Beard.

Michael E. Porter. (2009). Ser Competitivo. Deusto.

Organización Panamericana de la Salud (2021). Nombre de la página: <https://www.paho.org/>.

Gobierno de Chile. (2021). Cifras Oficiales. Nombre de la página: <https://www.gob.cl/>.

Aleksander Dietrichson PhD. 22 de junio del 2019. Métodos Cuantitativos, Coeficientes de Correlación. Nombre de la página: <https://bookdown.org/dietrichson/metodos-cuantitativos/coeficientes-de-correlacion.html>

Kristina Rogers. 16 de abril del 2020. Future Consumer Index: cómo COVID-19 está cambiando los comportamientos del consumidor. Nombre de la página: https://www.ey.com/es_ar/consumer-products-retail/how-covid-19-could-change-consumer-behavior

Forecasting at scale. Nombre de la página:
<https://facebook.github.io/prophet/>

Comité de Comercio electrónico. Cámara de Comercio de Santiago (Chile).
Nombre de la página: <https://www.ecommerceccs.cl/wp-content/uploads/2021/04/Perspectivas-del-Comercio-Electr%C3%B3nico-George-Lever.pdf>

Kantar. 02 de agosto del 2020. Las Cervezas más valoradas y de gran crecimiento durante la pandemia. Nombre de la página:
<https://www.kantar.com/latin-america/inspiracion/coronavirus/las-cervezas-mas-valoradas>

Kantar. 03 de diciembre del 2020. Consumo de hogares chiles empieza a estabilizarse. Nombre de la página:
<https://www.kantarworldpanel.com/cl/Noticias/Consumo-de-hogares-chilenos-empieza-a-estabilizarse->

María José Ramírez. 14 de octubre del 2020. Así se clasifican los segmentos socioeconómicos en Chile. Nombre de la página:
<https://marketing4ecommerce.cl/como-se-clasifican-segmentos-socioeconomicos-en-chile/>

América Economía. 18 de enero del 2017. Guía: 12 pasos para importar mercancías a Chile. Nombre de la página:
[https://mba.americaeconomia.com/articulos/reportajes/guia-12-pasos-para-importar-mercancias-chile.](https://mba.americaeconomia.com/articulos/reportajes/guia-12-pasos-para-importar-mercancias-chile)

Fortune Business Insights. Cannabis Beverages Market Size. Nombre de la página: [https://www.fortunebusinessinsights.com/industry-reports/cannabis-beverages-market-100738.](https://www.fortunebusinessinsights.com/industry-reports/cannabis-beverages-market-100738)

Anexo

Variables del Dataset

- ANIO: variable del tipo integer que indica el año de la venta realizada
- MES: variable del tipo integer que indica el mes de la venta realizada
- ID_EMPRESA: variable del tipo integer que representa la UEN correspondiente a la venta realizada. Está relacionada a “GRUPO_PROD”
- TIPO_VENTA: variable del tipo object que representa la categoría de la transacción. Puede ser venta, devolución, entrega sin cargo, etc.
- OFICINA: variable del tipo integer que representa el número de oficina que concreta la venta
- DESCR_OFICINA: variable del tipo object que representa el nombre de la oficina, de acuerdo a la región.
- DISTRITO: variable del tipo object que representa la región donde se concreta la venta
- ZONA: variable del tipo object que representa las zonas del país de acuerdo a la región
- SKU: variable del tipo integer que representa el número de referencia único para cada producto
- GRUPO_PROD: variable del tipo object que representa las categorías comercializadas
- TIPO: variable del tipo object que indica la clase de producto
- LICENCIA: variable del tipo object que representa el nombre de la licencia de la marca
- CATEGORIA: variable del tipo object que representa la clase del producto
- SUBCATEGORIA: variable del tipo object que representa la clase del producto con mayor detalle
- MARCAVENTAPAIS: variable del tipo object que representa el nombre de la marca tal como se comercializa en el mercado
- AGR_MARCA: variable del tipo object que representa la marca del producto

- DES_MARCA: variable del tipo object que representa una descripción de la marca y variedad
- AGR_EMPAQUE: variable del tipo object que representa el calibre del producto
- DES_EMPAQUE: variable del tipo object que representa una descripción del calibre y cantidades del producto
- RETORNABILIDAD: variable del tipo object que representa la retornabilidad o no de un empaque
- TAMANIO: variable del tipo object que representa la presentación del producto
- OCASION: variable del tipo object que representa la ocasión de consumo de acuerdo al CANAL
- CANAL: variable del tipo object que representa el canal o medio por el cual es vendido el producto
- NIVEL: variable del tipo object que representa el nivel de imagen de los comercios que venden los productos
 - S: sofisticado (Premium)
 - E: estándar (normal)
 - B: básico
- RELEVANCIA variable del tipo float que representa la consideración de los comercios que venden los productos
- FLAG_SKU_PPM: variable del tipo integer que representa la activación o inactivación del SKU
- VENTA_HL: variable del tipo float que representa la cantidad de hectolitros vendidos
- VENTA_CJ: variable del tipo float que representa la cantidad de cajas de producto vendidas
- VTA_BRUTA: variable del tipo float que representa la venta antes de impuestos expresada en pesos chilenos
- DCTO: variable del tipo float que representa el descuento realizado por la venta expresado en pesos chilenos
- FLETE: variable del tipo float que representa el costo del flete expresado en pesos chilenos
- COSTO: variable del tipo float que representa el costo total de comercialización del producto expresado en pesos chilenos
- IMP_G: variable del tipo float que representa el impuesto sobre el producto

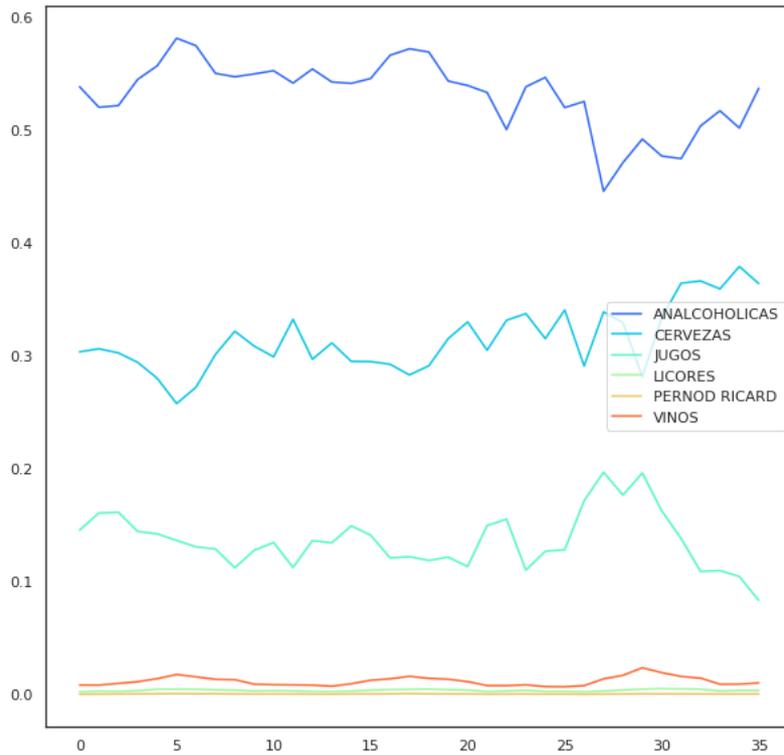
- IMP_E: variable del tipo float que representa el impuesto sobre el producto

Listado final de las variables utilizadas

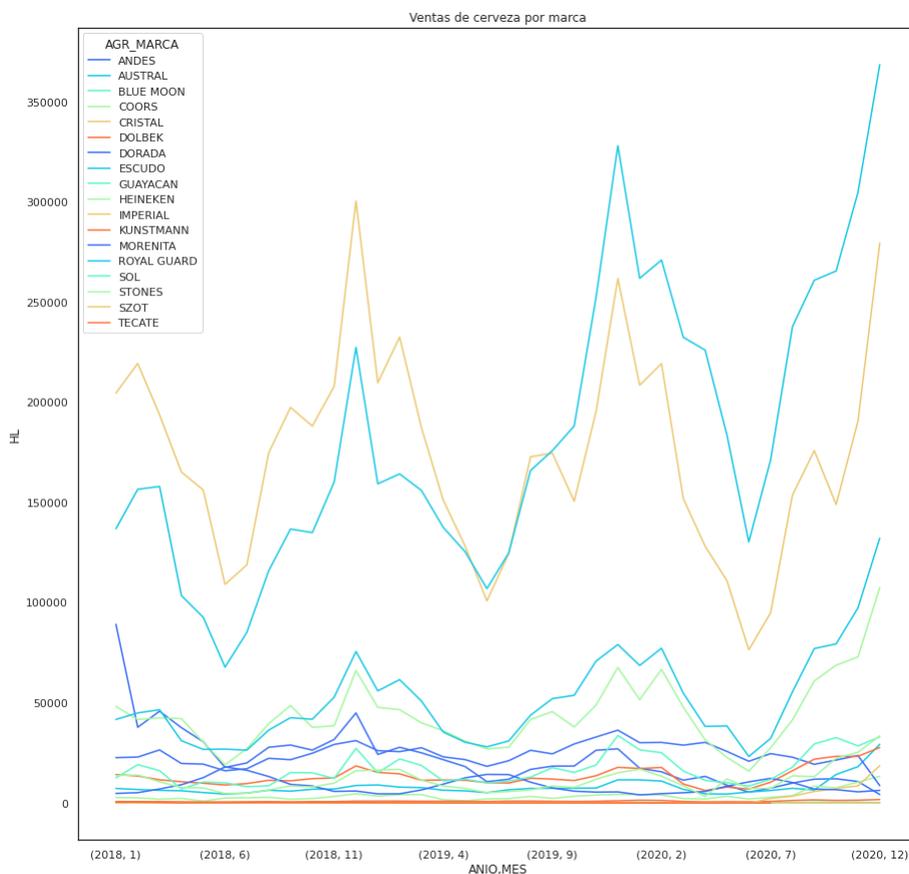
- ANIO: variable del tipo integer que indica el año de la venta realizada
- MES: variable del tipo integer que indica el mes de la venta realizada
- DISTRITO: variable del tipo object que representa la región donde se concreta la venta
- GRUPO_PROD: variable del tipo object que representa las categorías comercializadas
- AGR_MARCA: variable del tipo object que representa la marca del producto
- AGR_EMPAQUE: variable del tipo object que representa el calibre del producto
- CANAL: variable del tipo object que representa el canal o medio por el cual es vendido el producto
- VENTA_HL: variable del tipo float que representa la cantidad de hectolitros vendidos

Gráficos adicionales

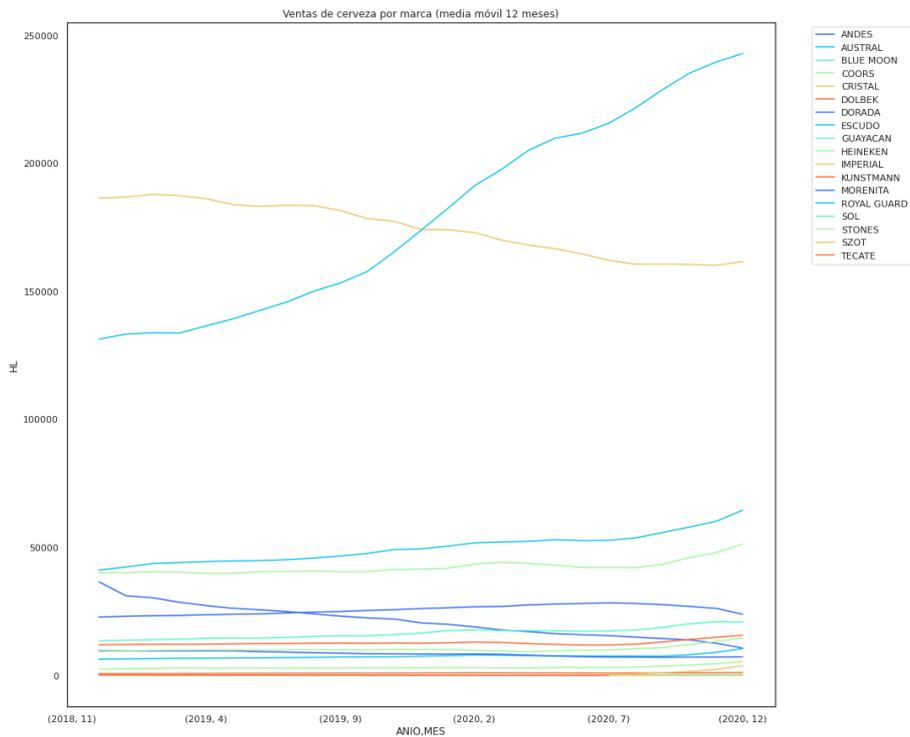
Análisis Exploratorio de Datos: Venta en Hectolitros por Grupo de Producto. Media móvil (12 meses) variable escalada



Análisis Exploratorio de Datos: Venta en Hectolitros de Marcas de Cervezas (todas)

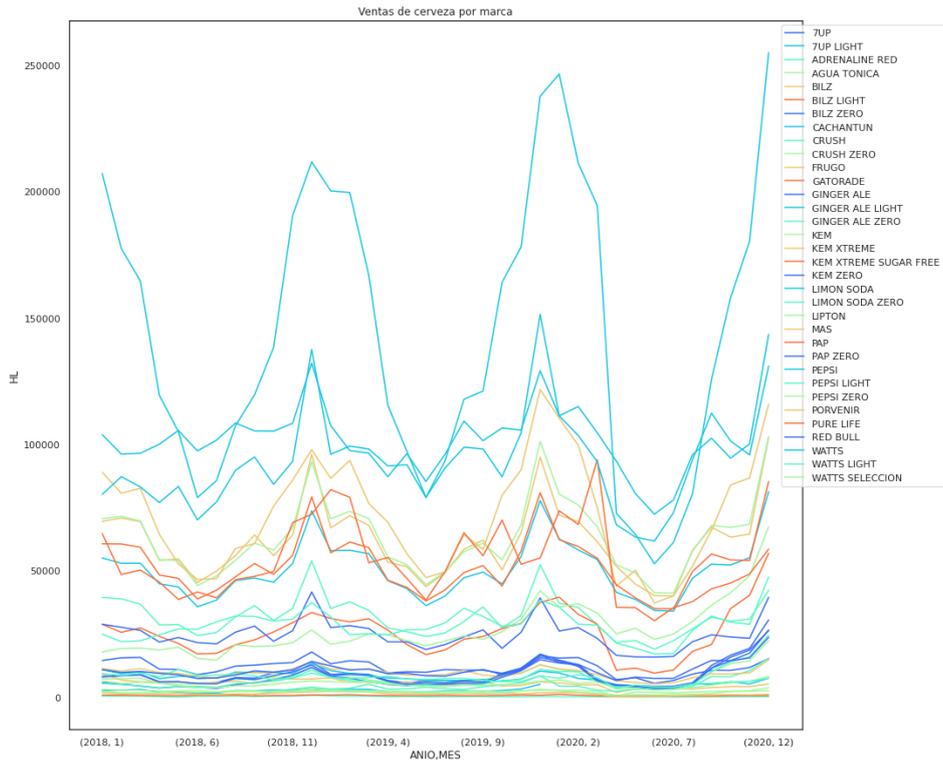


Análisis Exploratorio de Datos: Venta en Hectolitros de Marcas de Cervezas (todas). Media móvil (12 meses)

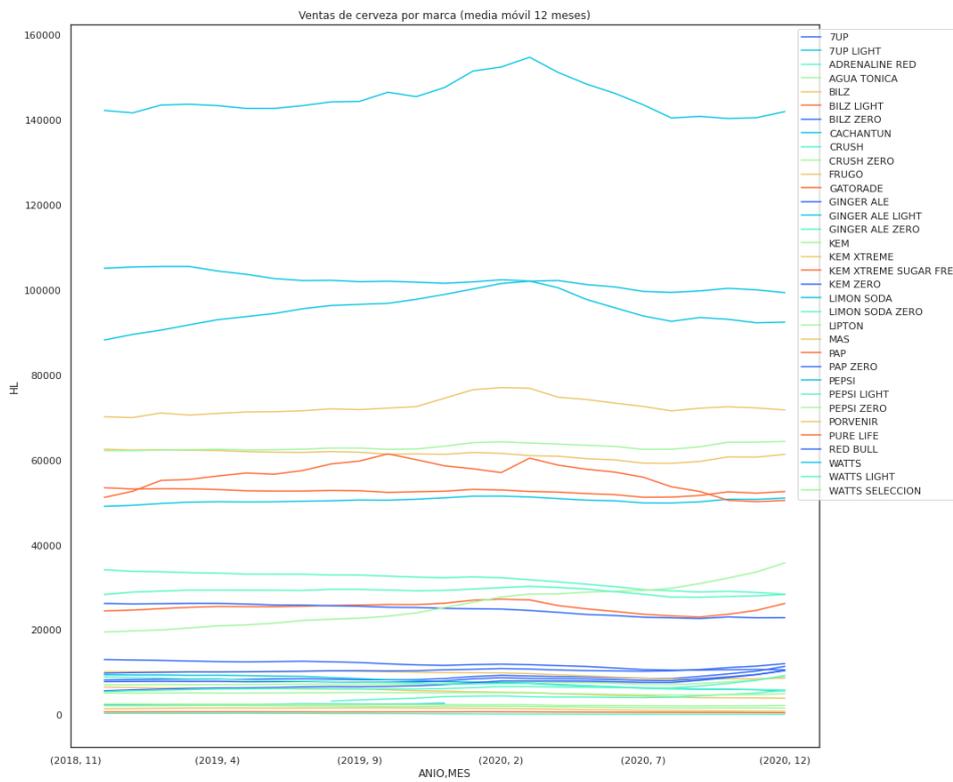


Análisis Exploratorio de Datos: Venta en Hectolitros de Bebidas Analcohólicas

WATTS	1533572
MIAS	1316234
CACHANTUN	784877
GATORADE	747182
PEPSI	746293
BILZ	711350
KEM	707539
PAP	676082
LIMON SODA	656749
CRUSH	595087
RED BULL	537326
GINGER ALE	494408
WATTS LIGHT	457715
PEPSI ZERO	368362
7UP	319671
LIPTON	294401
KEM XTREME	234369
WATTS SELECCION	223240
BILZ ZERO	203444
PAP ZERO	197378
GINGER ALE LIGHT	179982
CRUSH ZERO	177386
PURE LIFE	177315
PEPSI LIGHT	176945
LIMON SODA ZERO	144729
KEM ZERO	138361
FRUGO	120764
AGUA TONICA	112865
GINGER ALE ZERO	71126
PORVENIR	52900
KEM XTREME SUGAR FREE	40061
7UP LIGHT	39691
ADRENALINE RED	34755
7UP ZERO	18328
NOBIS	15236
AGUA TONICA ZERO	15172
PEPSI ZERO LIMON	14510
KEM MANGO DURAZNO	13105



Análisis Exploratorio de Datos: Venta en Hectolitros de Bebidas Analcohólicas. Media Móvil (12 meses)



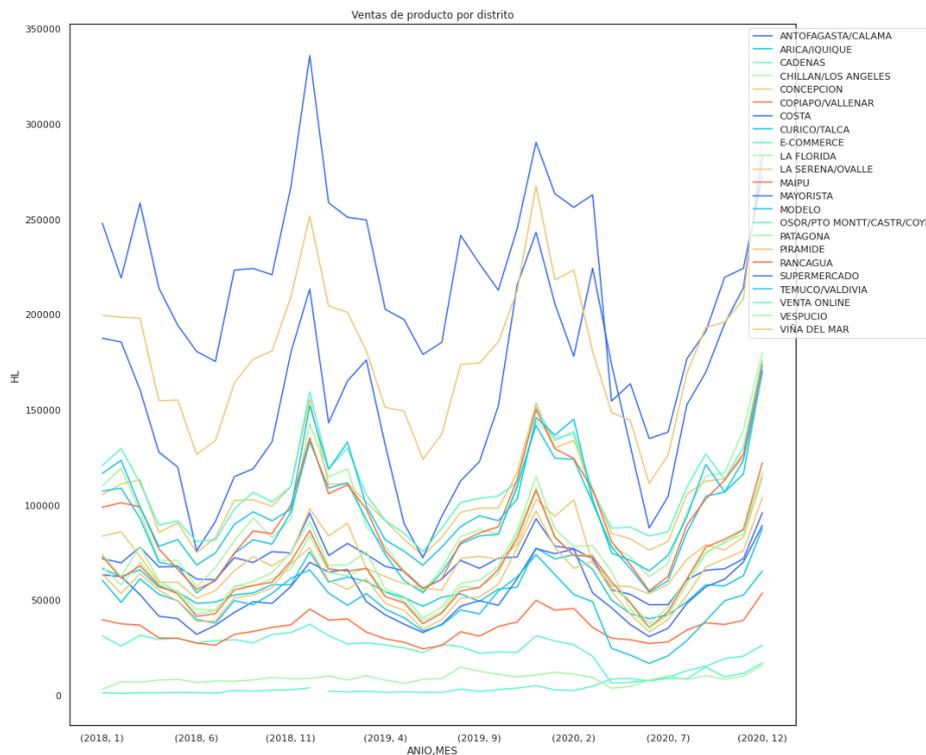
Análisis Exploratorio de Datos Regiones

Cantidad de observaciones por distrito/región

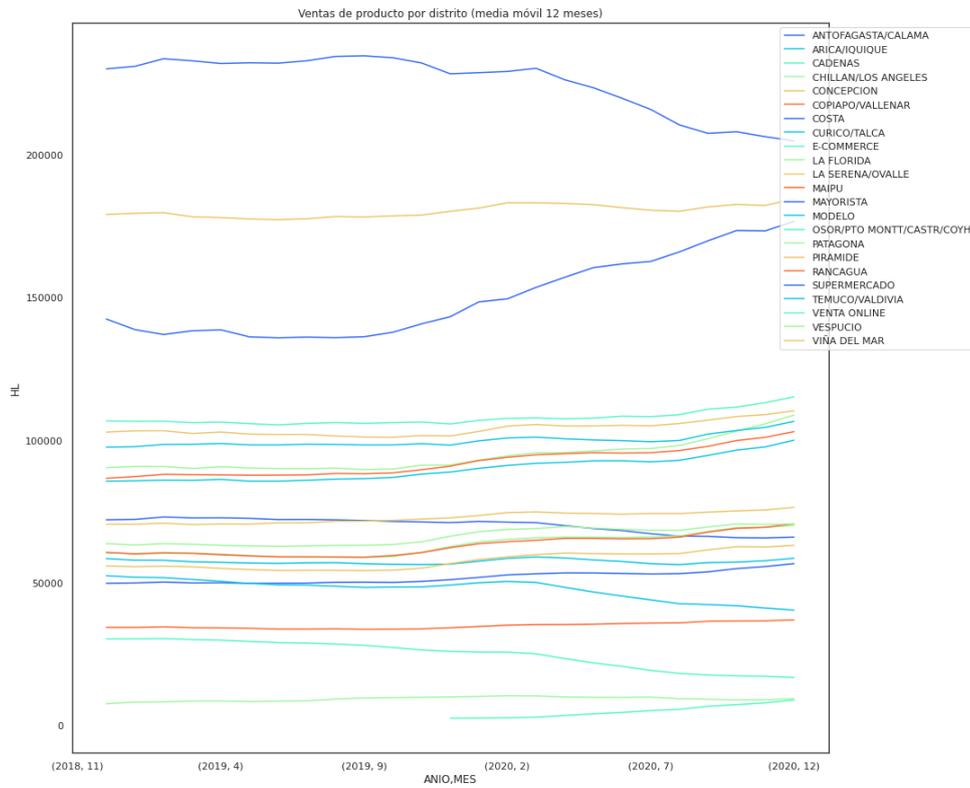
OSOR/PTO MONTT/CASTR/COYH	2416562
TEMUCO/VALDIVIA	1775309
CHILLAN/LOS ANGELES	1766134
CURICO/TALCA	1628494
VIÑA DEL MAR	1602762
LA SERENA/OVALLE	1196912
ANTOFAGASTA/CALAMA	1195620
RANCAGUA	1191773
ARICA/IQUIQUE	1130305
CONCEPCION	1053386
COSTA	926598
MODELO	778976
COPIAPO/VALLENAR	747174
VESPUICIO	581066
MAIPU	539660
PIRAMIDE	509035
LA FLORIDA	504548
SUPERMERCADO	242092
CADENAS	220447
MAJORISTA	182918
IMAGEN STGO	150920
PETROLERA	86313
IMAGEN	79101
IMAGEN VIÑA	73317
E-COMMERCE	37262
EVENTOS ESPECIALES	36636
VENTA ONLINE	12759
PATAGONIA	10825
ACTIVIDADES ESPECIALES	9890
VENDING	6820

Name: DISTRITO, dtype: int64

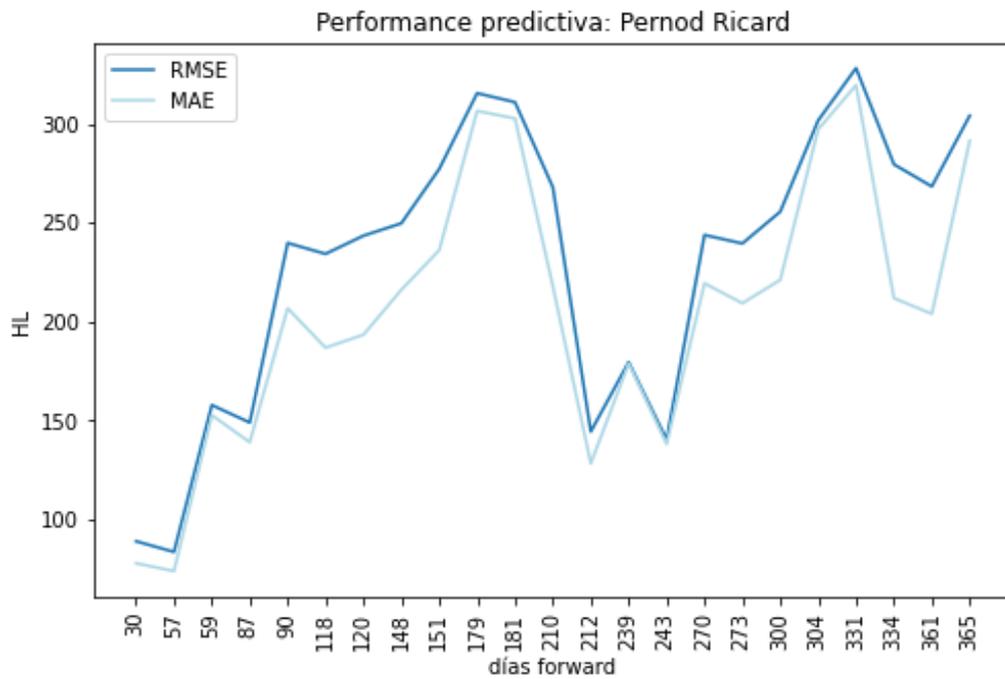
Venta en Hectolitros. Evolución



Venta en Hectolitros. Media móvil (12 meses)



Predicción Perno Ricard



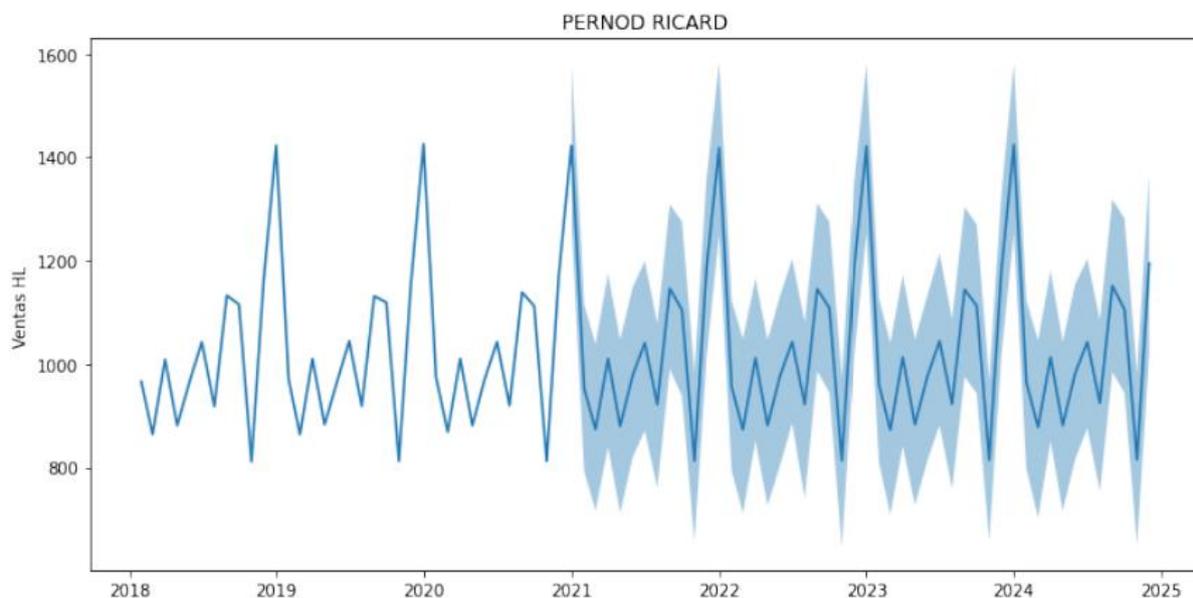


Tabla Valor Observado – Predicción por día

Fecha	Predicho	Valor Observado
31/1/2018	2.069.670	1.977.350
28/2/2018	1.903.620	1.913.440
31/3/2018	1.807.550	1.878.040
30/4/2018	1.473.210	1.507.540
31/5/2018	1.251.250	1.442.520
30/6/2018	1.169.510	1.190.710
31/7/2018	1.172.750	1.259.530
31/8/2018	1.523.280	1.551.060
30/9/2018	1.685.450	1.644.250
31/10/2018	1.618.870	1.652.070
30/11/2018	2.042.120	1.920.240
31/12/2018	2.432.590	2.502.650
31/1/2019	2.180.630	1.989.870
28/2/2019	2.002.420	2.032.790
31/3/2019	1.901.410	1.868.560
30/4/2019	1.550.640	1.550.610
31/5/2019	131.490	1.398.350
30/6/2019	12.270	1.210.230
31/7/2019	1.228.850	1.355.230

31/8/2019	159.960	1.666.310
30/9/2019	177.430	16.720
31/10/2019	170.160	1.755.940
30/11/2019	2.148.440	2.096.940
31/12/2019	2.571.160	2.666.160
31/1/2020	2.311.580	2.306.060
29/2/2020	2.121.530	2.244.720
31/3/2020	2.012.230	1.992.960
30/4/2020	1.639.680	1.489.840
31/5/2020	1.398.820	1.345.170
30/6/2020	1.304.550	1.120.290
31/7/2020	1.318.330	1.243.380
31/8/2020	1.731.350	164.660
30/9/2020	1.909.660	1.925.050
31/10/2020	1.837.050	199.840
30/11/2020	2.333.860	2.184.720
31/12/2020	2.772.390	2.939.650
1/1/2021	2.772.210	n/a
1/2/2021	2.465.910	n/a
1/3/2021	2.278.490	n/a
1/4/2021	2.152.990	n/a
1/5/2021	1.749.840	n/a
1/6/2021	1.498.640	n/a
1/7/2021	139.340	n/a
1/8/2021	1.418.280	n/a
1/9/2021	1.873.360	n/a
1/10/2021	2.049.360	n/a
1/11/2021	1.974.090	n/a
1/12/2021	2.522.050	n/a
1/1/2022	2.973.080	n/a
1/2/2022	2.643.620	n/a
1/3/2022	2.437.240	n/a
1/4/2022	2.302.980	n/a
1/5/2022	1.872.980	n/a
1/6/2022	1.601.980	n/a
1/7/2022	1.485.940	n/a
1/8/2022	1.510.150	n/a
1/9/2022	2.000.680	n/a
1/10/2022	2.192.410	n/a
1/11/2022	2.105.220	n/a
1/12/2022	2.686.460	n/a
1/1/2023	3.173.960	n/a
1/2/2023	2.821.860	n/a
1/3/2023	2.596.110	n/a
1/4/2023	2.453.290	n/a
1/5/2023	1.996.590	n/a

1/6/2023	170.550	n/a
1/7/2023	1.578.660	n/a
1/8/2023	1.601.670	n/a
1/9/2023	2.127.390	n/a
1/10/2023	2.335.640	n/a
1/11/2023	2.236.220	n/a
1/12/2023	2.850.000	n/a
1/1/2024	3.374.830	n/a
1/2/2024	3.000.630	n/a
1/3/2024	2.752.620	n/a
1/4/2024	2.592.210	n/a
1/5/2024	2.104.590	n/a
1/6/2024	1.804.530	n/a
1/7/2024	1.666.590	n/a
1/8/2024	1.703.650	n/a
1/9/2024	2.272.220	n/a
1/10/2024	2.474.230	n/a
1/11/2024	2.374.380	n/a
1/12/2024	3.042.470	n/a

Tabla Valores Observados y Predicciones por semestre. Variaciones y Promedio

Tabla por Canal en hectolitros

	REAL						PREDICCIÓN								
	jul-18	dic-18	jul-19	dic-19	jul-20	dic-20	jul-21	dic-21	jul-22	dic-22	jul-23	dic-23	jul-24	dic-24	
Supermercado															Crecimiento Promedio
Predicted															
Actual	4.836.081	9.574.076	5.142.426	8.863.732	4.171.216	8.421.387	4.438.813	6.683.846	4.223.106	6.331.253	4.007.050	5.980.359	3.779.880	5.688.084	Supermercado
Variación			6%	-7%	-19%	-5%	6%	-21%	-5%	-5%	-5%	-6%	-6%	-5%	-6%
Compras															Compras
Predicted															
Actual	4.316.257	8.248.987	4.429.256	9.600.700	5.007.260	12.205.940	6.078.895	11.526.010	7.169.442	13.409.980	8.261.659	15.285.290	9.330.486	17.313.310	13%
Variación			3%	16%	13%	27%	21%	-6%	18%	16%	15%	14%	13%	13%	
Consumo															Consumo
Predicted															
Actual	1.717.138	2.567.258	1.627.316	2.374.486	467.791	1.572.140	534.949	621.681	154.179	-903	-220				-74%
Variación			-5%	-8%	-71%	-34%	14%	-60%	-71%	-100%	-100%	-100%	-100%	#DIV/0!	
e-Commerce															e-Commerce
Predicted															
Actual				3.904.879	9.090.128	14.819.960	18.413.500	19.881.020	28.402.290	28.570.930	38.415.640	37.218.420	48.240.250	46.306.190	44%
Variación			#DIV/0!	#DIV/0!	#DIV/0!	280%	103%	34%	54%	44%	35%	30%	26%	24%	
Mayorista															Mayorista
Predicted															
Actual	2.333.842	5.109.844	2.344.832	5.807.076	2.688.290	7.031.285	2.934.276	6.661.808	3.370.432	7.568.531	3.807.344	8.471.592	4.232.379	9.466.242	10%
Variación			0%	14%	15%	21%	9%	-5%	15%	14%	13%	12%	11%	12%	

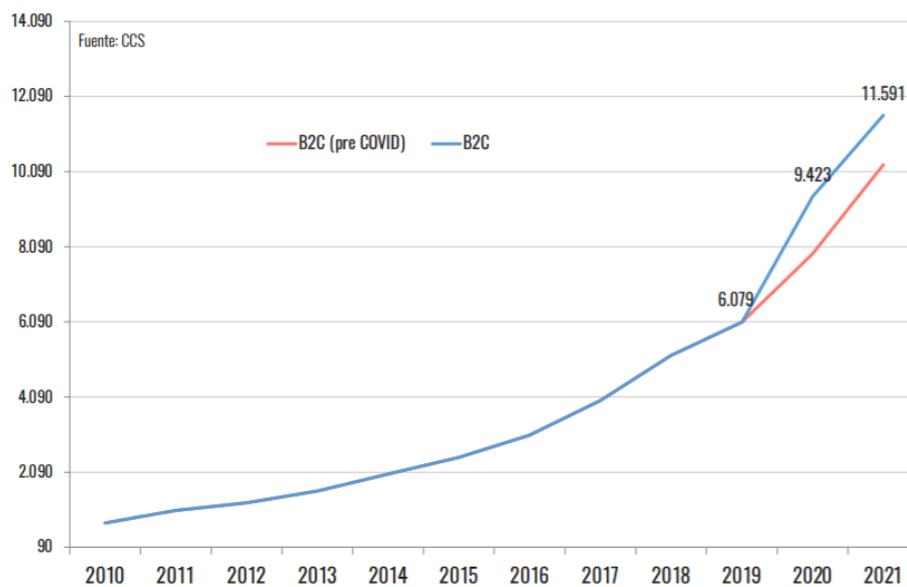
Tabla por Grupo de Producto en hectolitros

REAL							PREDICCIÓN							Crecimiento Promedio	
	jul-18	dic-18	jul-19	dic-19	jul-20	dic-20	jul-21	dic-21	jul-22	dic-22	jul-23	dic-23	jul-24	dic-24	
Analcohólicas															Analcohólicas
Predicted	723.927.800	1.355.860.000	771.369.100	1.435.682.000	593.201.400	1.579.199.000	723.816.500	1.216.049.000	743.626.900	1.245.145.000	763.450.800	1.274.074.000	781.912.600	1.316.587.000	2%
Actual															
Variación			7%	6%	-23%	10%	22%	-23%	3%	2%	3%	2%	2%	3%	
Cervezas															Cervezas
Predicted	342.959.200	832.231.000	394.766.300	899.827.000	415.325.600	1.070.393.000	446.737.200	1.162.160.000	507.494.700	1.105.144.000	568.416.100	1.226.827.000	626.377.600	1.362.070.000	9%
Actual															
Variación			15%	8%	5%	19%	8%	9%	14%	-5%	12%	11%	10%	11%	
Jugos															Jugos
Predicted	164.837.700	281.885.600	161.047.200	293.998.500	202.364.300	245.326.200	184.445.900	287.481.800	189.616.100	295.055.500	194.794.400	302.612.200	199.592.400	311.225.400	3%
Actual															
Variación			-2%	4%	26%	-17%	-9%	17%	3%	3%	3%	3%	2%	3%	
Licores															Licores
Predicted	6.021.827	8.284.496	6.651.831	10.330.390	6.605.504	10.906.060	7.006.174	8.840.640	7.499.938	9.403.038	7.994.511	9.960.950	8.461.402	10.673.360	3%
Actual															
Variación			10%	25%	-1%	6%	6%	-19%	7%	6%	7%	6%	6%	7%	
Vinos															Vinos
Predicted	1.983.721	2.164.331	1.950.718	2.320.325	2.124.212	3.052.776	3.139.663	2.686.468	3.769.836	3.171.187	4.400.496	3.652.394	5.024.635	4.191.360	17%
Actual															
Variación			-2%	7%	9%	32%	48%	-12%	20%	18%	17%	15%	14%	15%	

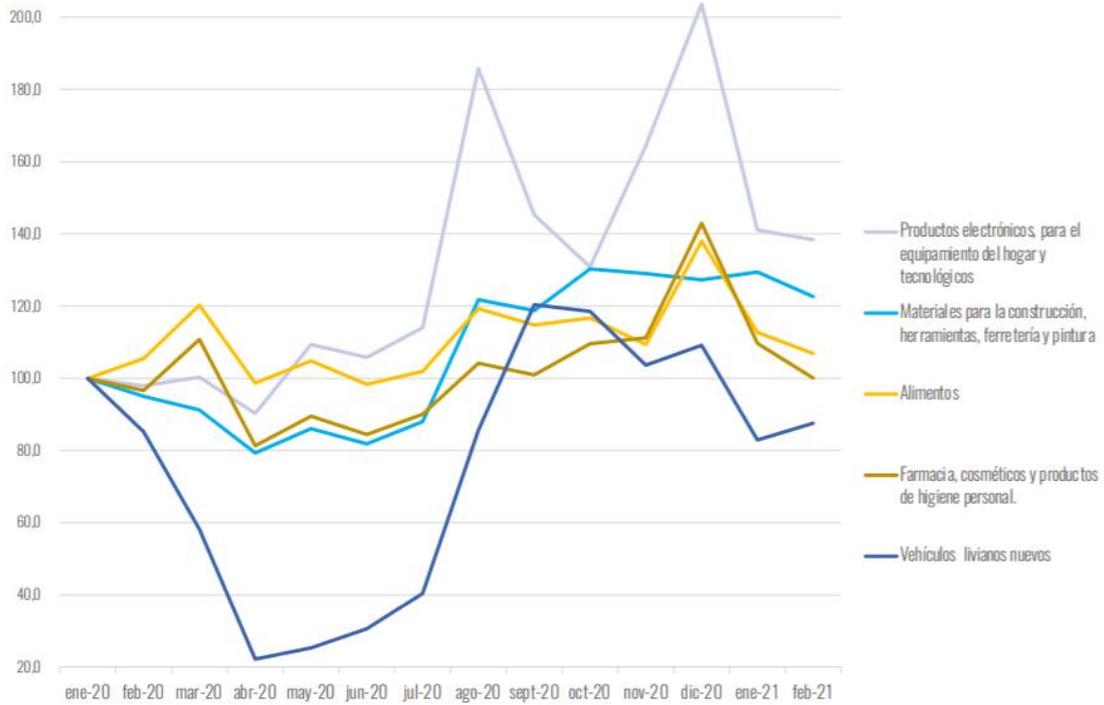
Información adicional Cámara de Comercio de Santiago de Chile

Gráficos soporte del artículo “Perspectivas-del-Comercio-Electrónico-George-Lever” de la Cámara de Comercio y Santiago

Proyección e-commerce, escenario antes y después de COVID-19

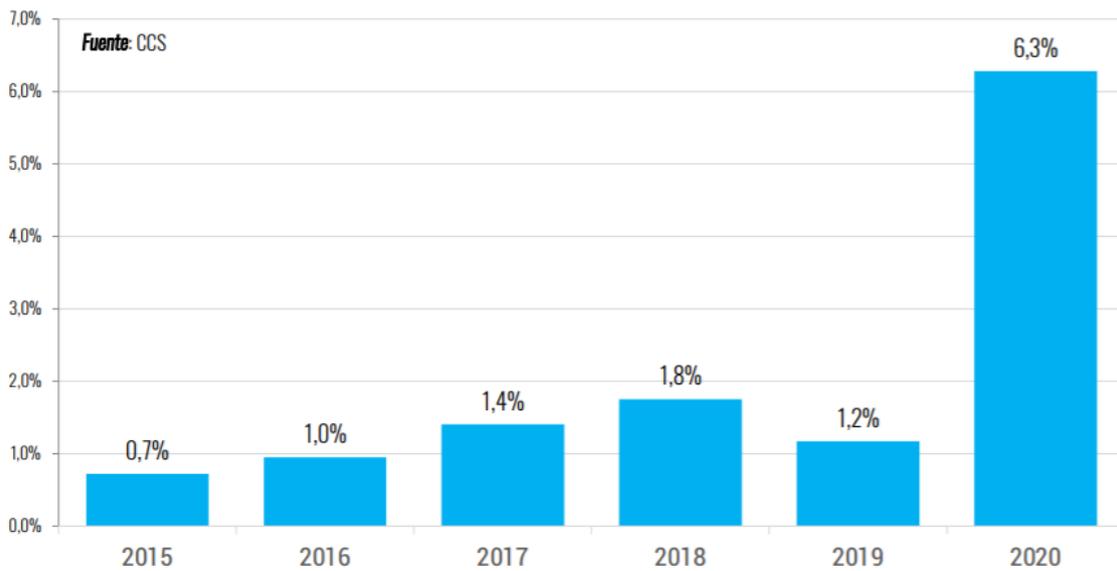


Evolución de las ventas del comercio (índice base enero 2020=100)

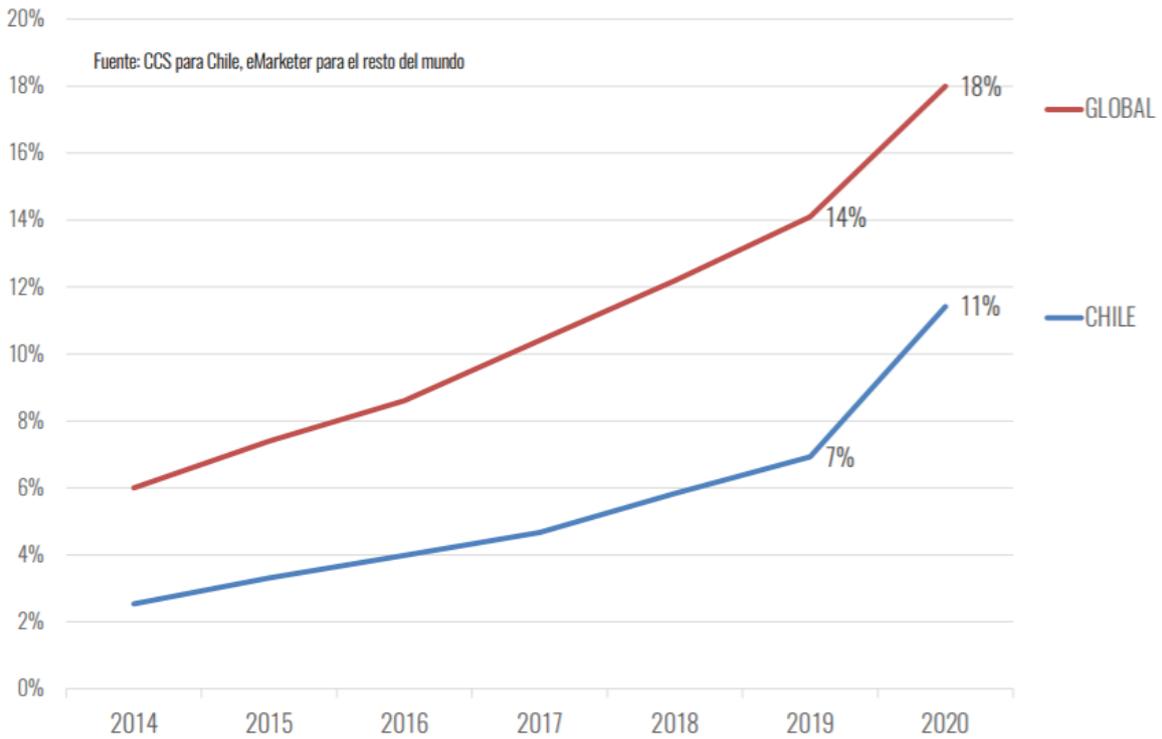


CHILE: Contribución del eCommerce al crecimiento del comercio

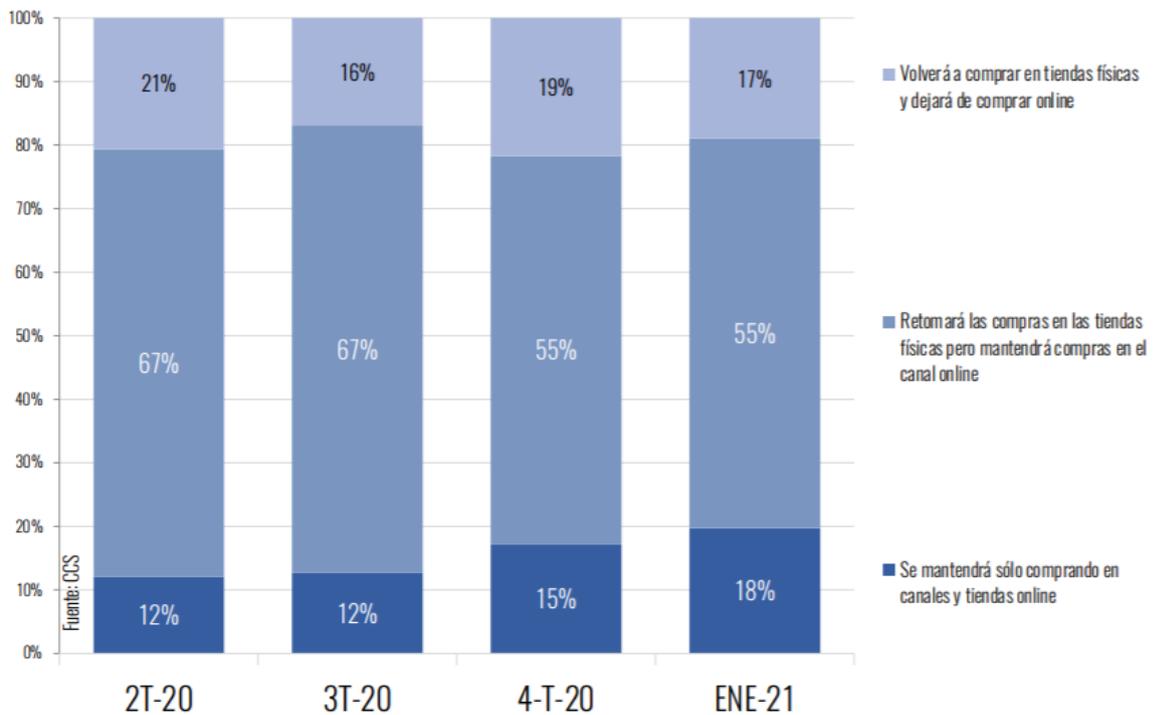
(Puntos porcentuales de crecimiento total del retail)



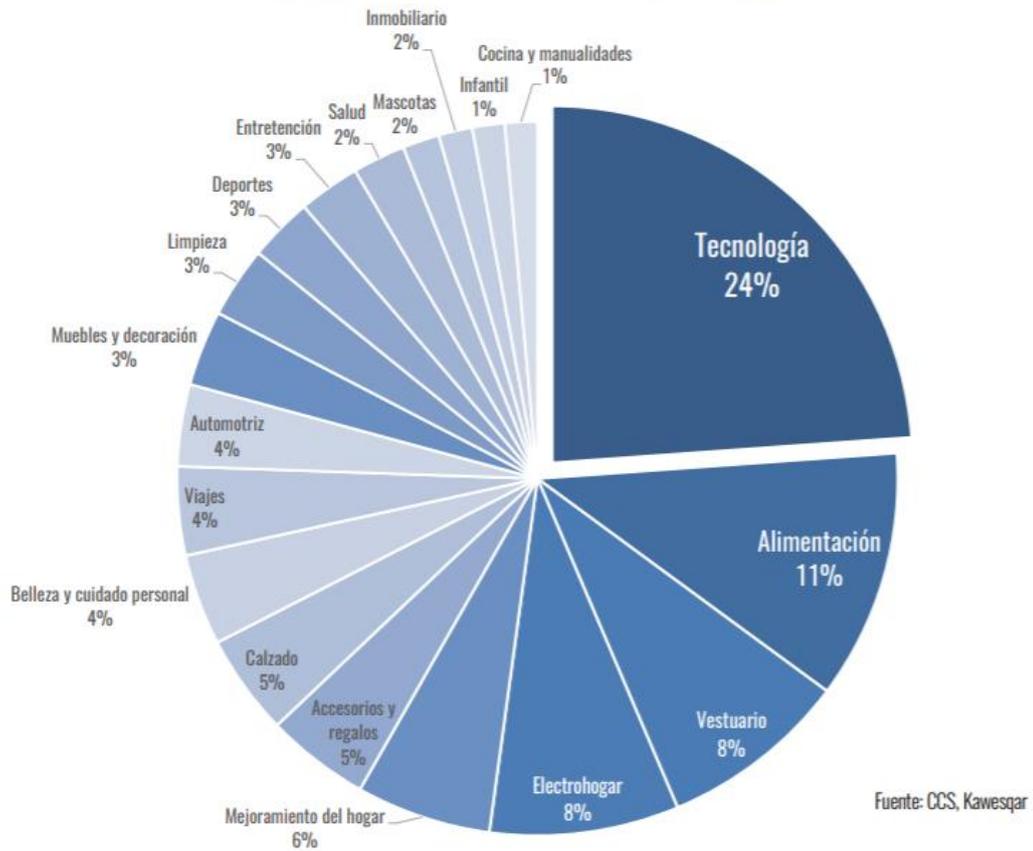
Penetración del eCommerce sobre ventas del comercio



Hábitos de compra post-pandemia



Composición de las compras online declaradas por consumidores 2020



Participación de algunas categorías dentro del gasto total online declarado por consumidores

