



**UNIVERSIDAD
TORCUATO DI TELLA**

MASTER IN MANAGEMENT + ANALYTICS

PRONOSTICANDO LAS DINÁMICAS
COMPETITIVAS EN LA INDUSTRIA CERVECERA
ARGENTINA

TESIS

Matías Ariel Breiman

Mayo 2022

Tutora: Magdalena Cornejo

Resumen

La cerveza es una de las bebidas alcohólicas más consumidas en el mundo. Particularmente en Argentina, es la de mayor consumo junto con el vino. En una industria que creció más de un 10% en 2021 y en la que, cada vez más, proliferan nuevas variedades y sabores que buscan adaptarse a las necesidades de los consumidores, poder contar con proyecciones precisas respecto a volumen de venta se vuelve un desafío complejo para los principales jugadores del mercado.

En Cervecería y Maltería Quilmes (CMQ), la compañía cervecera líder en el mercado argentino, la proyección de la demanda es un pilar fundamental en la operatoria y planificación de corto, mediano y largo plazo.

Con un portfolio de más de 100 SKUs activos que se distribuyen a lo largo y ancho del territorio nacional, contar con un pronóstico lo más preciso posible para todos ellos es determinante para los resultados de la empresa, puesto que dichas proyecciones constituyen el *input* que se usa para realizar el plan de producción. Si las proyecciones son demasiado optimistas, resultando en una venta menor a la esperada, los productos quedan varados en las plantas y depósitos, lo cual deriva no solo en costos de almacenamiento de la mercadería, sino eventualmente también en costos de derrame por el vencimiento de los líquidos. Por otro lado, si las proyecciones son muy pesimistas, podría incurrirse en costos de *stock-out*. Por estas razones, la precisión del pronóstico debe ser alta no solo a nivel producto, sino también a nivel geográfico.

A pesar de contar con varias herramientas y *drivers* internos para la construcción del *forecast*, tales como la venta histórica y diversos equipos regionales que proveen un análisis focalizado en sus respectivas plazas, existe la oportunidad de sumar de manera pertinente el volumen histórico de los demás jugadores del mercado, tal de poder incorporar las dinámicas competitivas a las proyecciones.

El objetivo de este trabajo consiste en explorar la posibilidad de refinar el *forecast* de demanda actual de la compañía proyectando el volumen de todos los competidores, construyendo así un modelo de industria y participación de mercado que permita mejorar la toma de decisiones, no solo en el campo de la proyección de la demanda, sino también en el de la planificación estratégica.

Se puede pensar en la industria como la conjunción de dos macro-variables: el segmento y el calibre de venta. El análisis de ambas puede darnos una idea de lo que el consumidor de cerveza está buscando a la hora de adquirir un producto y, por ende, donde debiera estar el foco de las empresas proveedoras de este.

Teniendo en cuenta estas dos dimensiones competitivas del mercado, se construirá un ranking de prioridades a atacar, que estará dado por la detección de aquellas combinaciones segmento-calibre donde se encuentren las mayores oportunidades de captura de participación de mercado. La recomendación de negocio estará basada en la priorización de la asignación de recursos a aquellas combinaciones que se encuentren en la cima del ranking.

Abstract

Beer is one of the most consumed alcoholic beverages in the world. Particularly in Argentina, it is the most consumed along with wine. In an industry that grew more than 10% in 2021 and in which, increasingly, new varieties and flavors proliferate that seek to adapt to the needs of consumers, being able to have accurate projections regarding sales volume becomes a complex challenge for the main market players.

At Cervecería y Maltería Quilmes (CMQ), Argentina's leading brewing company, demand forecasting is a fundamental pillar in short, medium, and long-term operations and planning.

With a portfolio of more than 100 active SKUs that are distributed throughout the national territory, having the most accurate forecast possible for all of them is decisive for the company's results since these projections constitute the input that is used to make the production plan. If the projections are too optimistic, resulting in sales being lower than expected, the products will remain stranded in the production plants and warehouses, which derives not only in storage costs of the merchandise but eventually also in spill costs due to the expiration of the liquids. On the other hand, if the projections are too pessimistic, they could incur shortage costs. For these reasons, forecast accuracy must be high not only at the product level, but also at the geographic level.

Despite having several tools and internal drivers for forecast construction, such as historical sales and various regional teams that provide focused analysis on their respective markets, there is an opportunity to add in a relevant way the historical volume of the other players in the market, to be able to incorporate competitive dynamics into the projections.

The objective of this work is to explore the possibility of refining the demand forecast by projecting the volume for all competitors, thus building an industry and market share model that allows an improvement regarding decision-making, not only in the field of demand forecasting, but also in strategic planning.

The industry can be thought of as the conjunction of two macro-variables: the sales segment and caliber. Analysing both of them can give us an idea of what the beer consumer is looking for when purchasing a product and, therefore, where the focus of the suppliers should be.

Taking these two competitive dimensions into account, a ranking of priorities to attack will be constructed, which will be given by the detection of those segment-caliber combinations where the greatest opportunities to capture market share are found. The business recommendation

will be based on prioritizing the resource allocation to those combinations that are at the top of the ranking.

Índice

1. Introducción	9
1.1. Contexto	9
1.2. Problema	10
1.3. Objetivo	12
2. Datos	13
2.1. Fuente, variables y clasificaciones	13
2.2. Análisis exploratorio.....	17
3. Metodología	25
3.1. Consideraciones generales.....	25
3.2. Tratamiento univariado.....	26
3.3. Tratamiento multivariado	40
4. Resultados	48
4.1. Modelos univariados	48
4.2. Modeos multivariados	52
5. Mapa competitivo	59
6. Conclusiones.....	64
Referencias Bibliográficas	67

Índice de Tablas

Tabla 1. Variables disponibles	13
Tabla 2. Clasificación de las marcas de la industria en los distintos segmentos.....	16
Tabla 3. Comparación del R^2 obtenido bajo los modelos Holt-Winters.....	36
Tabla 4. MAPE obtenido bajo el modelo de Holt-Winters	48
Tabla 5. Mejores modelos área-segmento univariados de <i>PyCaret</i>	51
Tabla 6. Mejores modelos área-calibre univariados de <i>PyCaret</i>	52
Tabla 7. Ranking de combinaciones prioritarias según índice TO para 2022.....	65

Índice de Figuras

Figura 1. Interacción de equipos de CMQ en el armado del <i>forecast</i> de demanda	10
Figura 2. División del mercado cervecero argentino en regiones y áreas geográficas	14
Figura 3. Evolución de volumen de venta de la industria cervecera argentina	17
Figura 4. Evolución de <i>market share</i> de los principales jugadores de la industria	18
Figura 5. Evolución de <i>market share</i> argentino, por fabricante y calibre	19
Figura 6. Evolución de <i>market share</i> argentino, por fabricante y segmento.....	19
Figura 7. Evolución de volumen de venta por área geográfica.....	21
Figura 8. Mapa competitivo para Argentina, correspondiente a 2021.....	22
Figura 9. Mapa competitivo por área geográfica de Argentina, correspondiente a 2021.....	23
Figura 10. Head del <i>dataset</i> con nivel de apertura área-calibre.....	28
Figura 11. Head del <i>dataset</i> con nivel de apertura área-segmento	28
Figura 12. Evolución de la venta en hectolitros, a nivel área-calibre.....	28
Figura 13. Evolución de la venta en hectolitros, a nivel área-segmento	29
Figura 14. Descomposición de las series temporales área-segmento	32
Figura 15. Descomposición de las series temporales área-calibre	33
Figura 16. Fit de los modelos de Holt-Winters en las series área-segmento.....	34
Figura 17. Fit de los modelos de Holt-Winters en las series área-calibre	35
Figura 18. Set de modelos disponibles en el módulo Regression de <i>PyCaret</i>	38
Figura 19. Set de modelos disponibles en el módulo <i>time_series</i> de <i>PyCaret</i>	39
Figura 20. Mapa de correlaciones entre las series área-segmento	40
Figura 21. Mapa de correlaciones entre las series área-segmento	41
Figura 22. Evolución de precios y mapas de correlación, por segmento.....	46
Figura 23. Evolución de precios y mapas de correlación, por calibre	47
Figura 24. Mejores modelos área-segmento univariados del módulo <i>Regression</i> de <i>PyCaret</i> ..	49
Figura 25. Mejores modelos área-segmento univariados del módulo <i>Time Series</i> de <i>PyCaret</i> ..	49
Figura 26. Mejores modelos área-calibre univariados del módulo <i>Regression</i> de <i>PyCaret</i>	50
Figura 27. Mejores modelos área-calibre univariados del módulo <i>Time Series</i> de <i>PyCaret</i>	50
Figura 28. Output de modelos VARIMA área-segmento.....	53
Figura 29. Output de modelos VARIMA área-calibre	53
Figura 30. Output de modelos VARIMA con covariables de precios, área-segmento	54
Figura 31. Output de modelos VARIMA con covariables de precios, área-calibre	55

Figura 32. <i>Benchmark</i> de modelos entrenados y selección de modelos finales.....	56
Figura 33. Proyecciones finales, según mejores modelos área-segmento	58
Figura 34. Proyecciones finales, según mejores modelos área-calibre	58
Figura 35. Proyección del mapa competitivo por área geográfica, correspondiente a 2022	60
Figura 36. Proyección del mapa competitivo nacional, correspondiente a 2022.....	61

1. Introducción

1.1. Contexto

A la hora de realizar un pronóstico de cualquier índole, uno siempre está sujeto a convivir con cierto margen de error, lo cual es lógico dado que no se conoce de antemano el output futuro. Por supuesto, el margen de error que se debiera considerar aceptable varía según el uso y las implicancias que tenga nuestra proyección (Morlidge, 2015).

En el caso de un *forecast* de demanda de cerveza para CMQ, se debe estimar la venta de cada uno de los productos del portfolio a nivel mensual, en cada región del país, con un horizonte anual (12 meses).

El contar con una precisión elevada a nivel producto trae consigo importantes ventajas. Al tomarse estas proyecciones que realiza el equipo de *Demand Planning* como premisa para elaborar el plan de producción mensual de la compañía, si la venta se comporta conforme al *forecast*, se logra minimizar la posibilidad de incurrir en costos de derrame de producto o *stock-out*. Esto es así porque se planifica la producción y los envasados de los productos para poder disponibilizarlos en la cantidad y región en donde se requieren. Además, se planifica el stock de arranque para cada mes en todos los casos, lo cual por supuesto depende de la venta mensual. Para aquellos productos donde la venta supere en demasía a las expectativas iniciales, el stock de arranque del mes siguiente se verá negativamente afectado y, en el caso en que este comportamiento se repita por varios meses, se corre el riesgo de tener quiebres, no pudiendo abastecer el mercado cuando probablemente siga habiendo demanda de dichos SKUs. Por el contrario, si la venta es considerablemente menor a la esperada y no se logra corregir el comportamiento con el correr de los meses, los productos se vencerán y, por ende, no podrán salir al mercado. A todo esto, hay que agregar los costos de tráfico entre plantas y depósitos cuando se requiere paliar algún desbalanceo de venta entre regiones. Dicho de otro modo, se puede tener una buena precisión a nivel producto agregado, pero la operativa diaria y los costos asociados a los movimientos recién mencionados acabaran afectando de forma negativa a la venta si no se logra un pronóstico sólido, también a nivel geográfico.

1.2. Problema

Como está visto, el *forecast* de demanda es entonces una herramienta que contiene información primordial, la cual impacta de manera directa en la operativa diaria de CMQ. No solo direcciona el manejo táctico de los envasados, producciones y despliegue de productos en el mes inmediato, sino que también aporta visibilidad sobre la evolución de la estrategia de la compañía respecto al crecimiento de las distintas marcas, segmentos y calibres en cada una de las regiones del país. Todo esto queda plasmado en el *forecast* mensualizado del próximo año, lo cual guía la toma de decisiones ligadas a cuestiones tales como la compra de mayores parques de envases retornables, instalación de nuevas líneas productivas en las plantas y optimización en el uso de depósitos y centros de distribución.

Cabe destacar que, generalmente, cuanto más volátil sea la demanda que se quiera proyectar, encontraremos mayores obstáculos para arribar a pronósticos robustos (Tashman, Sglavo, & Gilliland, 2015).

En CMQ, el proceso de *forecasting* se repite de forma iterativa todos los meses, actualizando todas las proyecciones. Se contempla así la venta histórica cerrada hasta el último mes. A su vez, se toma contacto con los distintos equipos que aportan información referida a su campo de *expertise*.

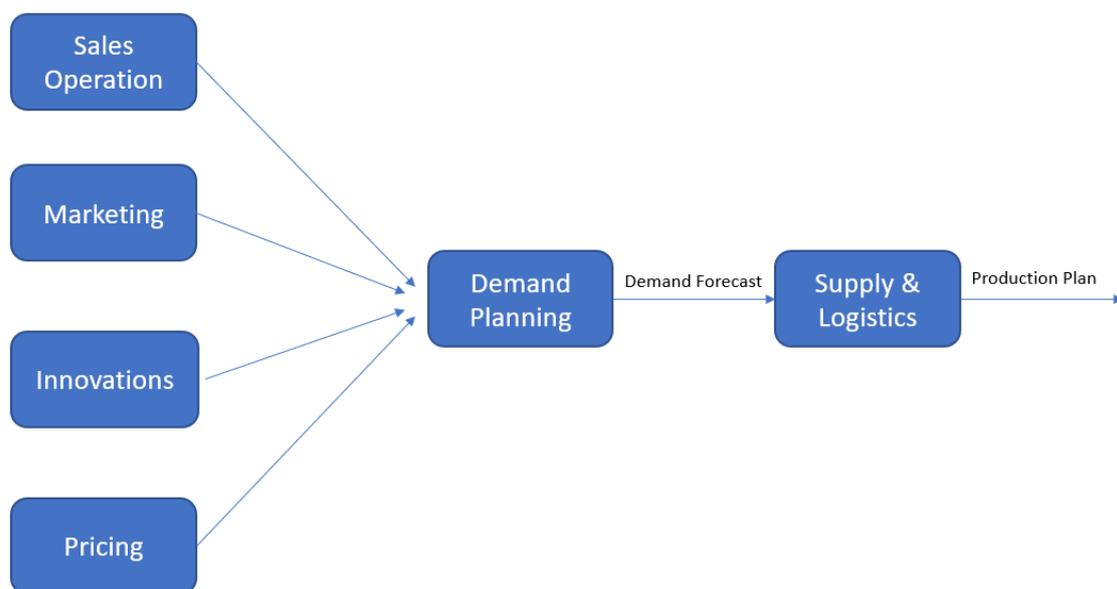


Figura 1. Interacción de los distintos equipos de CMQ en el armado del forecast de demanda.

El equipo de *Sales Operations* está compuesto por los equipos regionales de venta de todo el país. Son aquellos que se encuentran más cercanos a los clientes y trabajan codo a codo junto a

ellos para desarrollar y fortalecer la performance del portfolio de CMQ en sus respectivas plazas. Su aporte al *forecast* de demanda es fundamental, dado que conocen de primera mano las necesidades, preocupaciones y oportunidades que se presentan en los distintos distribuidores, cadenas de supermercados y demás canales de venta.

Por otro lado, los equipos de Márketing de cada una de las marcas facilitan al equipo de *Demand Planning* información referida a campañas, publicidades, activaciones comerciales en los puntos de venta y demás acciones de comunicación que pudieran constituir un *driver* de venta en sí mismas.

Otro actor que presta su ayuda a la construcción del *forecast* es el equipo de Innovaciones. El mercado cervecero muta y se expande, cada vez con mayor rapidez, hacia nuevas tendencias. Un ejemplo es la proliferación de cervezas sin alcohol, un nicho que ha crecido mucho en el último tiempo (Lopez, 2020).

A su vez, la comunicación con el equipo de *Pricing* es muy importante para la correcta lectura del sendero de precios que seguirán los productos en las distintas regiones. Los momentos de aumentos generalizados suelen ser altamente volátiles en cuanto a venta, pudiendo observarse una aceleración en los despachos los días previos, seguidos por una pronunciada baja en las salidas diarias, una vez que los nuevos precios entran en vigencia.

En resumen, la coordinación con numerosos equipos tiene lugar permanentemente, para contemplar todos los factores que pudieran incidir en la demanda y así tener proyecciones de venta robustas. El equipo de *Demand Planning* actúa como nexo entre todos ellos y *Planning Industrial*, quien se encarga de desarrollar el plan de producción que asegure poder entregar la venta plasmada en el *forecast*, como se mencionó anteriormente.

El proceso parece ser lo suficientemente sólido como para trasladar en forma oportuna y precisa los objetivos y proyecciones, alimentando al plan de producción y despachando los productos según lo previsto. Sin embargo, en ocasiones la venta no se comporta siguiendo todas las directrices del *forecast*. Esto lleva a pensar que hay factores que pudieran no estar siendo tenidos en cuenta, al menos en la medida en que lo requieren.

Si bien es cierto que CMQ conserva la posición de líder en el mercado argentino, la competencia ha ganado terreno en la industria en los últimos años, principalmente a manos de la empresa de origen chileno CCU. Este escenario invita a pensar en la necesidad de entender en profundidad cómo, dónde y por qué se produjo este fenómeno. Pero más importante aún, en el potencial que tendría poder predecir el volumen de la competencia de la misma forma en que se proyecta

el volumen propio, de modo de poder tener un *forecast* a nivel industria que capture las tendencias y patrones de la dinámica competitiva en cada una de las regiones del país.

1.3. Objetivo

Teniendo en consideración la situación competitiva actual, surge la pregunta de si sería posible refinar el *forecast* de demanda proyectando también el volumen de la competencia, construyendo así un modelo de industria que permita mejorar la toma de decisiones, no solo en el campo de la proyección de la demanda, sino también en la planificación estratégica. Tener visibilidad hacia adelante de las tendencias de venta de la competencia permitiría determinar en dónde la misma se intensificará y dónde se estará debilitando, en qué marcas o productos sucederá, lo cual puede disparar distintas iniciativas para proteger el *market share*, que pueden ir desde ajustes en la estructura de precios hasta el lanzamiento de nuevos productos.

Por lo tanto, el objetivo de esta tesis es elaborar un modelo de predicción de volumen y *market share* de la industria cervecera en Argentina, con distintos niveles de apertura, que permita identificar patrones de evolución en cada caso. El *output* debe ser la proyección del mapa competitivo en los próximos meses, que se podrá ir actualizando mensualmente conforme se obtengan los datos cerrados de cada periodo. Una vez obtenidos los pronósticos de volumen, se construirá un simulador de participación de mercado a distintos niveles. Se identificarán qué segmentos y calibres son los que más crecerán en los próximos meses, de modo de establecer un orden de prioridades a atacar. Para ello, también se determinará si dicho crecimiento vendrá en mayor medida dado por CMQ o la competencia. Esto es fundamental para entender el tipo de acción a tomar.

2. Datos

2.1 Fuente, variables y clasificaciones

Para la formulación de los modelos predictivos, se utilizarán datos que son provistos a la compañía a través de la consultora Scentia (<https://scentiaconsulting.com.ar/>). Esta firma trabaja con un sistema de escaneo de tickets de compra en el punto de venta, lo cual le permite recabar una gran cantidad de información referida a los productos transaccionados, siendo la misma una muestra representativa del volumen total industria. Esto se pudo comprobar para los productos pertenecientes a CMQ dado que, por información interna, se conoce la venta poblacional de los mismos.

Por acuerdo contractual entre ambas empresas, el primer reporte se elaboró en enero de 2018, con datos correspondientes a dicho mes. Por ello, la disponibilidad de información está restringida a esa fecha. Desde entonces, Scentia envía mensualmente una actualización del informe en cuestión. La base de datos que se utilizó en este trabajo contiene registros (filas) mensuales, desde enero de 2018 a diciembre de 2021, para cada producto transaccionado en cada región del país, dentro de la industria. Las variables (columnas) que conforman la base son las siguientes:

Variable	Descripción
Periodo	Mes y año del registro.
CANAL	Canal de venta del registro.
REGION	Región del país donde se produjo la transacción.
AREA	Área del país donde se produjo la transacción.
EAN	Código EAN asociado a cada producto.
Desc_2	Descripción del producto asociada al código EAN.
Manuf	Fabricante del producto.
Brand	Marca asociada al producto.
Segmento	Segmento de venta del producto, según arquitectura de precios.
TIPO_ENV	Calibre asociado al producto.
VolumenLts	Volumen de venta en litros.
Importe_Item	Valor monetario de la transacción en ARS (pesos argentinos).

Tabla 2. Nombres de las variables en la base de datos y su descripción correspondiente.

Para dar máxima claridad de los datos con los que contamos, a continuación se explica el significado de cada una de las variables que componen la base, así como los valores que pueden tomar.

La variable *Periodo* indica en qué mes y año fue realizada la transacción, y va desde enero 2018 a diciembre 2021.

La variable *Canal* aporta información sobre el canal de venta de la transacción en cuestión. Esta clasificación se divide en K+T (Kioscos y Tradicionales), *Self* Cadenas (supermercados) y *Self* Independientes (Autoservicios).

Por otro lado, las variables *Región* y *Área* indican en qué ubicación del territorio argentino fue llevada a cabo la transacción del registro. Un área agrupa a varias regiones, como se muestra en la siguiente figura:

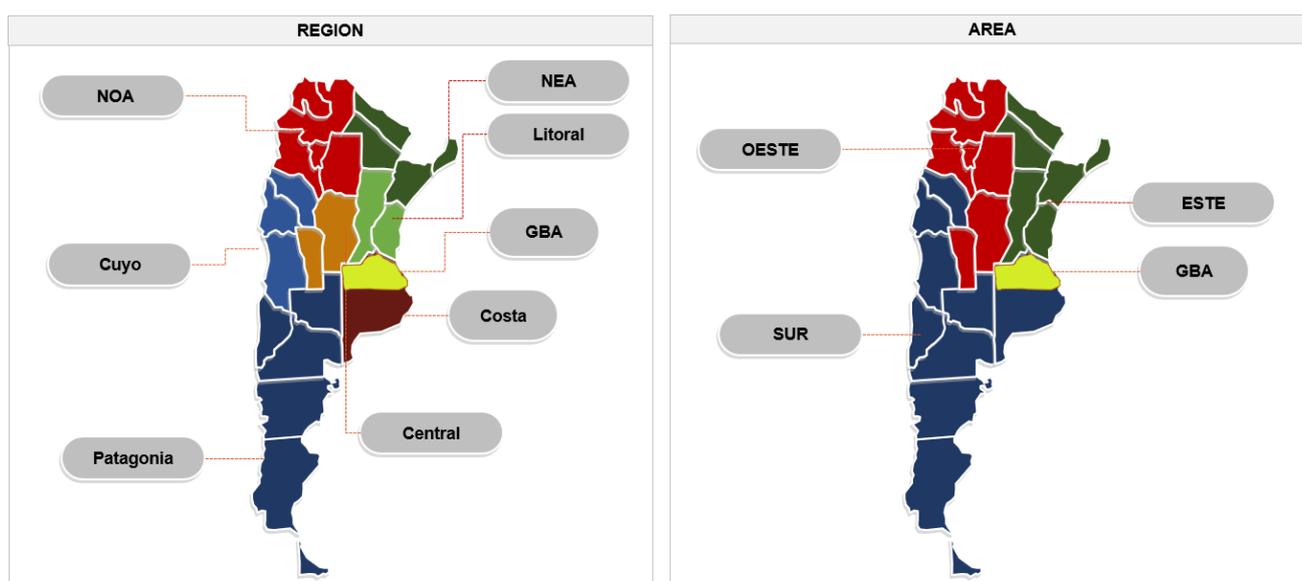


Figura 2. División del mercado cervecero argentino en regiones y áreas geográficas.

De esta forma, el área GBA es idéntica a la región GBA, el área Este se compone de las regiones Litoral y NEA, el área Oeste está conformada por las regiones Central y NOA y el área Sur abarca las regiones de Patagonia, Cuyo y Costa.

Continuando con la descripción de las variables, se encuentran el código EAN y la descripción del producto asociada a ese código. Indican cual fue el SKU que formó parte de la transacción en cuestión.

La variable *Manuf* informa acerca del fabricante o comercializador que realizó la venta. Los jugadores del mercado incluidos en la base de Scentia son: ABI (CMQ), CCU, Rabieta, Peñón del Águila, Antares, Schofferhofer, Cepas Argentinas, ANAPA S.A, Marcas Propias y Otros Fabricantes.

Por otro lado, la variable *Brand* indica la marca correspondiente al producto transaccionado, mientras que la variable *Segmento* clasifica las marcas en distintos grupos, según su precio por litro. En general, los productos pertenecientes a segmentos con mayor precio por litro son aquellos que ostentan un margen de rentabilidad mayor para el fabricante. Existen 4 segmentos en la industria cervecera Argentina. En orden ascendente según el precio, el primer lugar encontramos el segmento *Value*, que está compuesto por las marcas más económicas del mercado. En segundo lugar, encontramos el segmento *Core*, que nuclea a las marcas más volumétricas del mercado en la actualidad. Las mismas poseen un precio moderado. En tercer lugar, se encuentra el segmento *Core Plus*, en el cual encontramos marcas con un posicionamiento de precio mayor a las del segmento *Core*, apuntadas a un consumidor más refinado e informado. Por último, el segmento *Premium* está representado por aquellas marcas que apuntan a los consumidores más exigentes, quienes están dispuestos a pagar el elevado precio que tienen sus productos, relativo a los demás segmentos. En su mayoría, son marcas de renombre no solo local, sino global.

VALUE	CORE	CORE PLUS	PREMIUM	
1890	ANDES	27 EAZY	1906	OTRAS MARCAS
BAJO CERO	BRAHMA BEATS	ANDES ORIGEN	AMSTEL	OTRO MUNDO
BALTICA	BRAHMA CHOPP	ANDES PORTER	ANTARES	OTTWEILER
BIECKERT	BRAHMA DORADA	ANDES RED	BARBAROJA	PATAGONIA
HANOVER	BRAHMA DUPLO	BUDWEISER	BEAGLE	PAULANER
IGUANA	BRAHMA LIME	IGUANA SUMMER	BECKS	PEÑON DEL AGUILA
ISENBECK	BRAHMA MALZBIER	IMPERIAL	BLUE MOON	PILSEN
ISENBECK DARK	CERVEZA CORDOBA	ISENBECK BLACK IPA	CARLSBERG	PILSENER
OTRAS MARCAS	MARCAS PROPIAS	ISENBECK HONEY	CASSARO	RABIETA
PALERMO	NORTE	MILLER	CORONA	SCHOFFERHOFER
ST WENDELER	NORTE PORTER	MILLER LITE	CUSQUEÑA	SOL
	OTRAS MARCAS	SALTA CAUTIVA	DAB	SPATEN
	QUILMES CLASICA	SKOL	ESTRELLA GALICIA	STELLA ARTOIS
	QUILMES VARIEDADES		FLENSBURGER	TEMPLE
	SALTA		FRANZISCANER	TSINGTAO
	SALTA VARIEDADES		GOOSE ISLAND	WARSTEINER
	SANTA FE		GROLSCH	ZILLERTAL
	SCHNEIDER		GUINNESS	
	SCHNEIDER VARIEDADES		HEINEKEN	
			HERTOG JAN	
			HOEGAARDEN	
			KUNSTMAN	
			LEFFE	
			LOWENBRAU	
			MALTIN	
			ME ECHO LA BURRA	
			MICHELOB	
			NEGRA MODELO	

Tabla 2. Clasificación de las marcas de la industria en los distintos segmentos.

Además, se cuenta con información acerca del calibre al que pertenece cada producto, representado en la variable *TIPO_ENV*. La clasificación de la misma abarca 3 categorías: Botella retornable, Lata y OW (*One-Way*/descartable) sin Lata.

Finalmente, las variables *VolumenLts* e *Importe_Item* reflejan los montos correspondientes a cada transacción. La primera indica la cantidad de litros vendida, mientras que la segunda informa el importe monetario que se pagó por dicho volumen, en pesos argentinos.

2.2 Análisis exploratorio

En esta sección, se explora en profundidad el *dataset*, con el objetivo de encontrar *insights* que permitan describir el comportamiento de la industria cervecera argentina en los últimos 4 años.

En primer lugar, al graficar la evolución del volumen de venta nacional sin ningún nivel de apertura, se observa que el mercado se comporta con una estacionalidad mensual marcada a lo largo de los años. Mientras que los picos de venta se producen en los meses de diciembre y enero, notamos puntos mínimos en los meses de junio o julio. Jean-Louis Bertrand y Mila Parnaudeau demuestran en su estudio *The impact of climate variability on the private sector* (2015) que el consumo de cerveza se acelera cuando la temperatura supera los 23 °C. Por ello, no debería sorprender que las ventas sean relativamente altas en los meses de verano, en contraste con el invierno.



Figura 3. Evolución de volumen de venta de la industria cervecera argentina (en hectolitros).

Teniendo una noción de cómo se comporta el volumen de la industria en su conjunto, resulta interesante observar cómo se distribuye la venta entre los principales jugadores del mercado. Como se ha mencionado anteriormente, la industria está altamente concentrada entre 2 jugadores; ABI/CMQ y CCU. No obstante, esta última viene ganando participación de mercado en los últimos años a costa de la primera. A principios de 2018, ABI contaba con un *Share Of Market (SOM)* de 70%, el cual se fue deteriorando con el paso de los meses, hasta llegar a niveles de 65% durante 2021. Para dar visibilidad de lo que esto representa, en una industria de 20

millones de hectolitros anuales equivaldría a una pérdida de volumen de 1 millón de hectolitros, o 100 millones de litros.

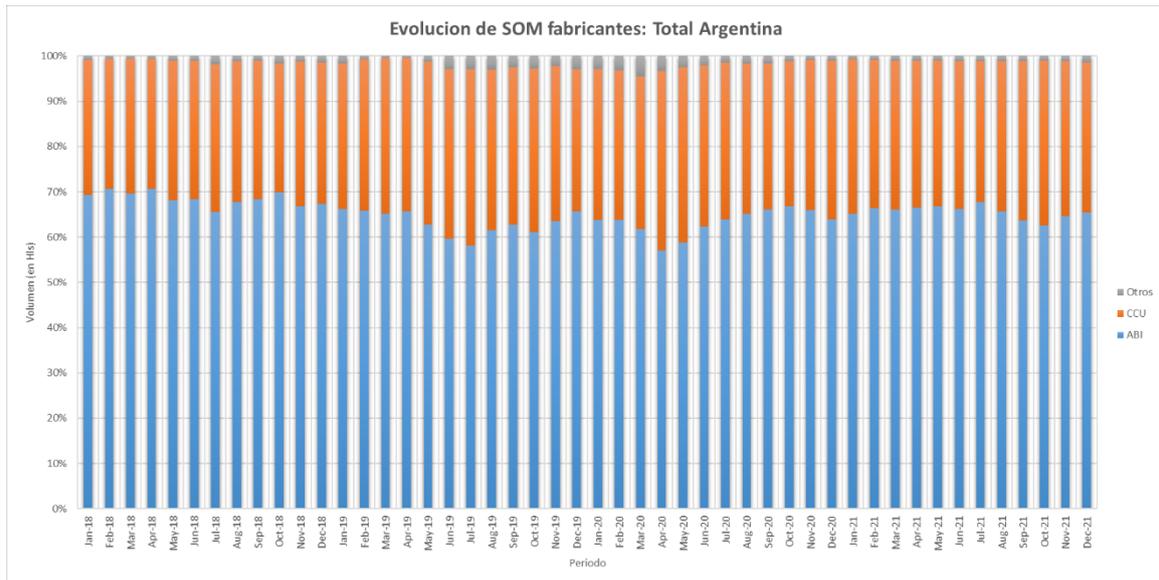


Figura 4. Evolución de market share de los principales jugadores de la industria cervecera argentina.

Sin dudas, el avance de su principal competidor sitúa a CMQ en la necesidad de un entendimiento de la situación y sus causas. Para ello, resulta interesante observar el grafico anterior, pero agregando las dimensiones de Calibre y Segmento. Para ello, se consideran únicamente ABI y CCU, dado que representan más del 95% del mercado y esto permitirá una lectura gráfica más sencilla.

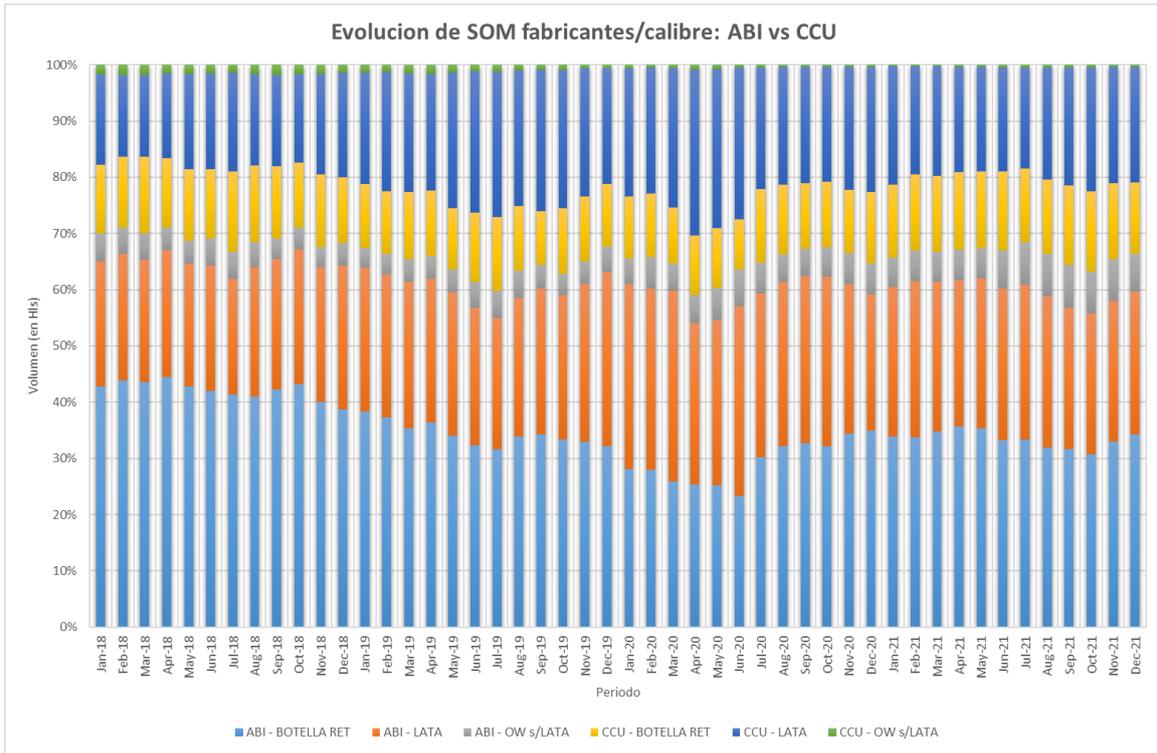


Figura 5. Evolución de market share argentino, por fabricante y calibre.

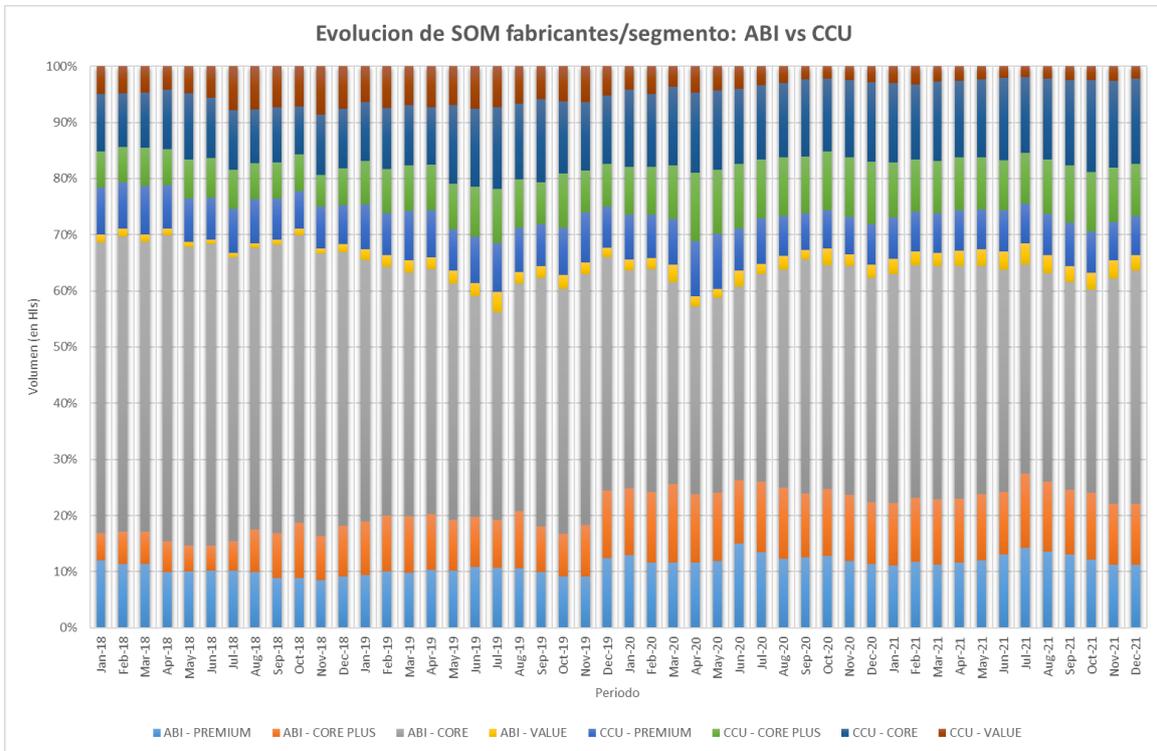


Figura 6. Evolución de market share argentino, por fabricante y segmento.

A nivel calibre, lo que se nota gráficamente es que, para ambas compañías, crece la participación de la lata en la venta total. No obstante, a diferencia de ABI, CCU mantiene la participación de su botella retornable. En el líder del mercado, la misma cae considerablemente, lo cual resulta una señal de alarma si se considera que, históricamente, es su calibre más volumétrico. De aquí se desprende una primera explicación para entender la pérdida de *SOM* de ABI. El mercado está virando hacia un mayor consumo de latas, lo cual se explica por diversas razones tales como el cierre del gap de precios vs el litro retornable, la optimización del empaque respecto tanto al almacenaje del mismo como al consumo individual y la eliminación de la incomodidad que pueda generar en el consumidor el tener que llevar los envases al almacén cada vez que quiera realizar una compra. Por otro lado, el enfriamiento de la lata se da en menor tiempo que el de la botella retornable, lo cual amplía las ocasiones de consumo disponibles.

Es que CCU fue pionera en el entendimiento de la capacidad que tenía este calibre para aumentar sus ventas. Fue así como, en 2016, anunció un plan de inversiones por \$2.250 millones, con fuerte foco en la ampliación de la capacidad de envasado de latas en su planta ubicada en Luján. En 2020, formalizó un nuevo plan de desembolsos por \$4.500 millones, destinados a continuar aumentando la capacidad de envasado de dicha planta (Compte, 2020).

Por otro lado, al mirar la evolución de venta de los segmentos, se nota una disminución progresiva del *Core* en ABI, mientras que CCU aumenta su participación. No obstante, CMQ crece con mayor fuerza en los segmentos *Premium* y *Core Plus*. Si bien esto podría ser positivo para esta última en términos de facturación y EBITDA, no lo es en términos de participación de mercado, ya que su crecimiento de volumen *Above Core* no llega a compensar caída de participación que sufre en sus marcas *mainstream*. Un punto a tener en cuenta es que los segmentos *Premium* y *Core Plus* poseen mixes de lata mayores a los segmentos *Core* y *Value*, por lo cual la batalla por prevalecer y continuar fortaleciéndolos está enfocada en continuar desarrollando el calibre que más crecimiento tuvo en la industria en los últimos años.

Otro *insight* que pudiera resultar importante para la planificación estratégica futura, es que la venta se ha federalizado en términos relativos. Si bien el área GBA conserva la posición dominante en volumen de venta durante todos los meses de la base de datos, se puede observar claramente como las áreas del interior del país ganan participación. De hecho, GBA muestra una tendencia ligeramente negativa, mientras el resto de las áreas crecen año a año, tal como se muestra en la figura 7.

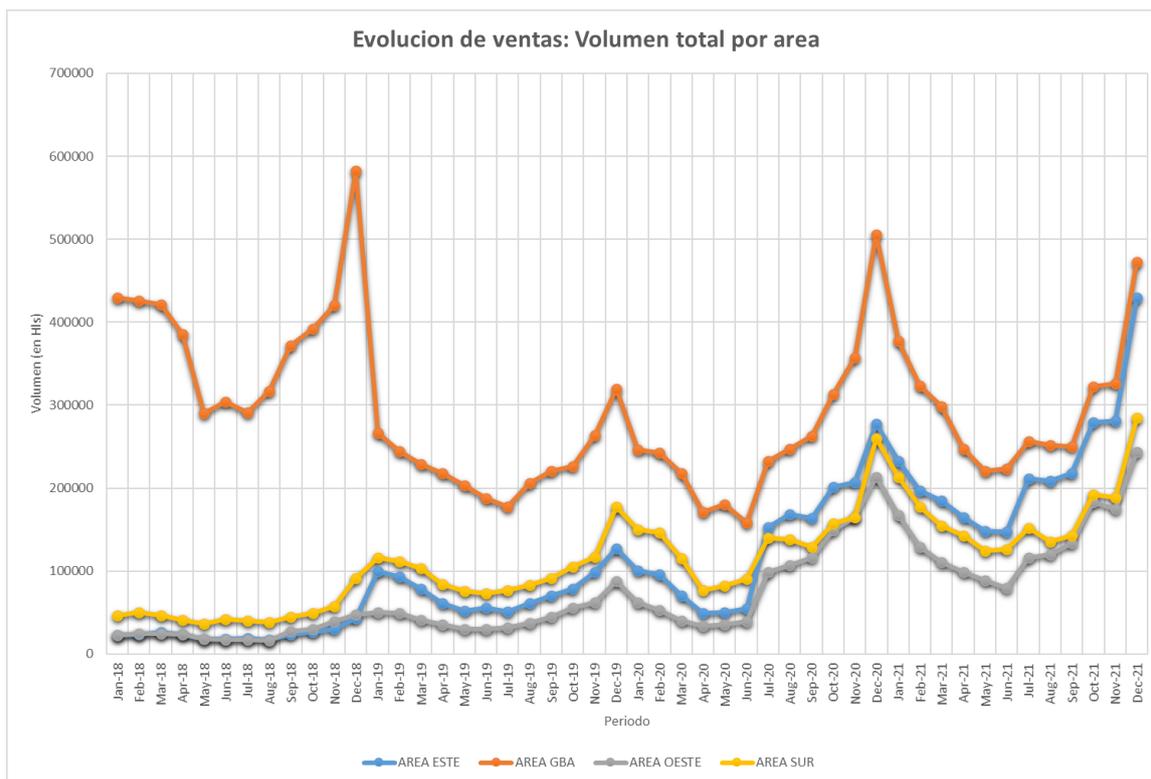


Figura 7. Evolución de volumen de venta por área geográfica.

Para ilustrar mejor el objetivo de este trabajo, se puede pensar en la industria como la conjunción de dos macro-variables: el segmento y el calibre de venta. El análisis de ambas puede dar una idea de lo que el consumidor de cerveza está buscando a la hora de adquirir un producto y, por ende, donde debieran hacer foco las empresas proveedoras de este.

Mientras la evolución de los segmentos da noción del nivel de calidad de los productos buscados por los consumidores, así como su disposición a pagar un valor extra por aquellos SKUs *Premium*, el calibre brinda información acerca de qué tipos de empaque prefieren.

Teniendo esto en cuenta, es posible construir un mapa competitivo que contemple la participación de mercado de CMQ/ABI y la de sus competidores, para entender en donde se encuentran las mayores oportunidades para crecer en *market share*.

A continuación, se muestra cómo se desarrolló el mercado para todo el año 2021.

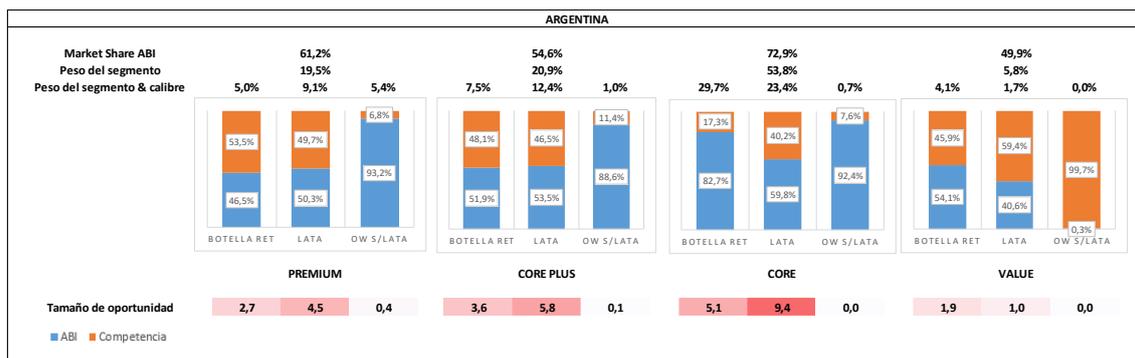


Figura 8. Mapa competitivo para Argentina, correspondiente a 2021.

En primer lugar, se calcula el *Market Share ABI* como el porcentaje de volumen vendido por CMQ sobre el total de la industria, en todo 2021, para cada segmento. También se muestra el peso cada segmento sobre el total de la venta. Por otro lado, el peso del segmento & calibre se calcula distribuyendo el peso del segmento entre los distintos calibres que lo conforman, según su venta dentro de dicho segmento. Notar que la suma de los pesos segmento & calibre es igual al peso del segmento en todos los casos. Debajo, se encuentra graficada la participación de mercado de ABI y su competencia, para cada una de las combinaciones segmento & calibre. Por último, pero fundamental, se encuentra el tamaño de oportunidad (TO). Este indicador se calcula para cada combinación segmento & calibre, y consiste en la multiplicación entre el peso del segmento & calibre y la participación de mercado de la competencia en la combinación en cuestión.

$$TO = \text{Peso del segmento \& calibre} * \text{Market Share competencia} * 100$$

El tamaño de oportunidad es de mucha utilidad, ya que pondera las combinaciones donde la competencia ostenta una participación de mercado considerable que podría potencialmente ser capturada o disminuida, con el peso relativo de dichas combinaciones en el mercado. De poco serviría invertir recursos adicionales de manera intensiva en una combinación donde la competencia tenga participación mayoritaria, si dicha combinación no tiene mayor relevancia en la industria. Un claro ejemplo sería observar OW s/Lata, dentro de lo que es el segmento *Value*, para lo que es Argentina. La competencia tiene prácticamente todo el mercado en su poder, por lo cual a priori pareciera ser una buena iniciativa lanzar nuevas marcas y productos que correspondan a esta combinación, o bien invertir recursos para fortalecer aquellos que ya existen. Sin embargo, el peso de esta combinación es prácticamente nulo, resultando en un TO muy bajo. Esto nos da la pauta de que, al menos, no se debiera priorizar la inversión incremental allí.

El mayor TO de 2021 a nivel nacional lo obtuvo el calibre Lata, dentro del segmento Core. Esto es consistente con lo mencionado anteriormente acerca del fuerte crecimiento que viene teniendo el calibre en los últimos años, haciendo que su peso en el mercado se fortalezca. Además, es el calibre del segmento Core (el más volumétrico de la industria) en donde la competencia tiene el mayor *market share*.

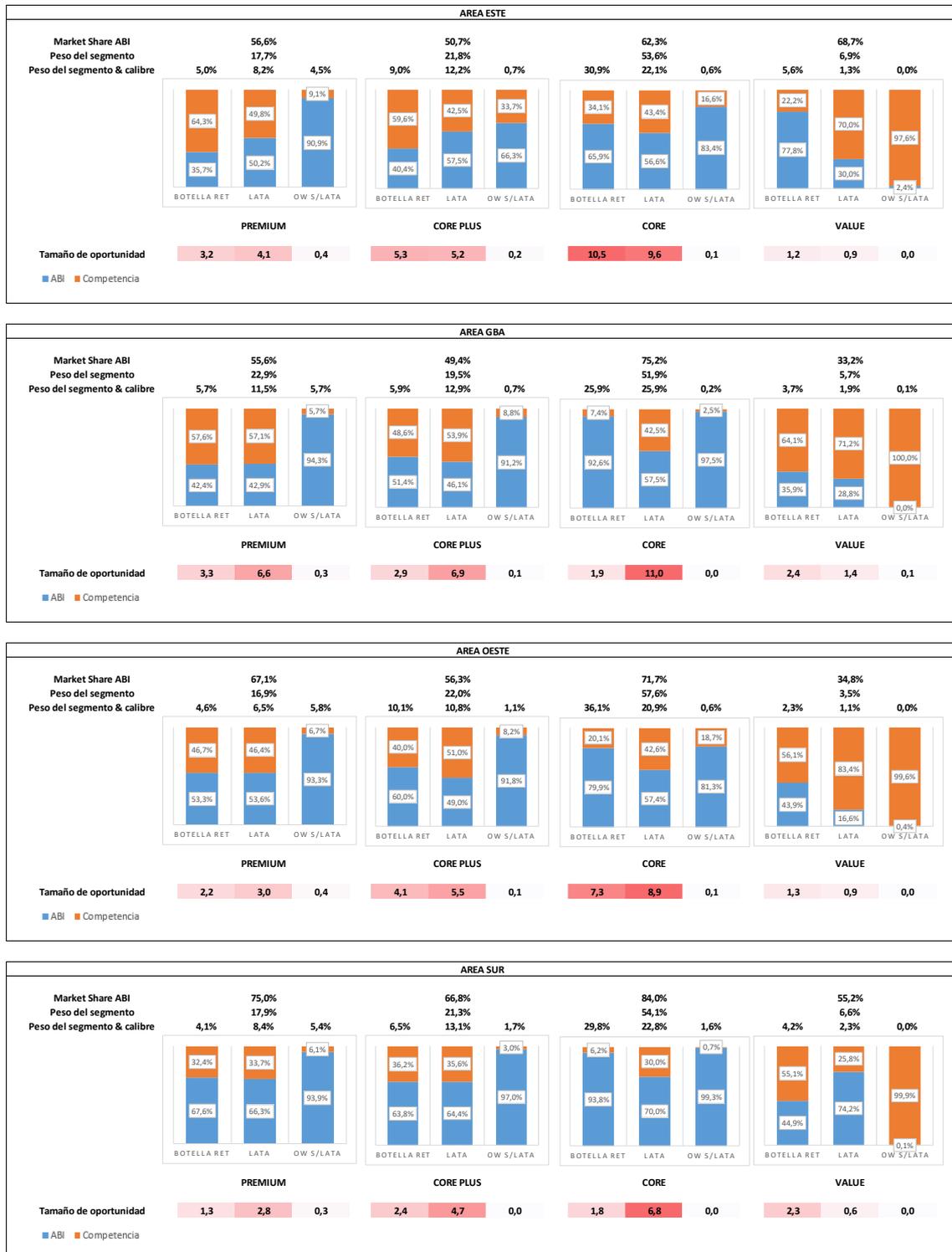


Figura 9. Mapa competitivo por área geográfica de Argentina, correspondiente a 2021.

Resulta igualmente interesante observar este mapa a nivel área, representado en la figura 9. Cómo se podría intuir, no todas las regiones del país se comportan de igual manera en términos de venta.

Es útil entender cómo el TO puede priorizar la inversión a nivel área, independientemente de si las combinaciones señaladas resultan equivalentes en dos o más casos. Incluso encontramos que, en área Este, la mayor oportunidad se encuentra en el litro retornable. Este ejemplo ilustra la intuición del valor que puede tener ejecutar este tipo de análisis a nivel geográfico.

Por lo tanto se propone, a través de las proyecciones que puedan otorgarnos distintos modelos estadísticos de series temporales, construir este mapa competitivo de la industria cervecera para el año 2022, identificando aquellas combinaciones segmento & calibre que proyecten el mayor TO y contrastando los resultados contra los datos reales de 2021. De este modo, podremos construir un ranking de combinaciones, que servirá para la priorización respecto a la asignación de recursos que se tengan a disposición para capturar participación de mercado.

3. Metodología

3.1 Consideraciones básicas

En términos generales, la manera de analizar y proponer una solución para el problema en cuestión será a través de la construcción de modelos de *forecasting*, en donde el volumen de venta será nuestra variable a predecir. Es posible pensar a la base de datos disponible como un conjunto de series temporales, en donde cada serie corresponde al volumen vendido de un determinado producto, en determinada región, para cada mes, desde enero de 2018 a diciembre de 2021.

Teniendo en consideración que se cuenta con información sobre más de 1.800 productos para cada región del país, estamos hablando de aproximadamente unas 15.000 series temporales susceptibles de ser analizadas y modeladas. Sin dudas, la complejidad de trabajar la información en el mayor nivel de granularidad posible sería muy ambicioso y exigente, tanto teóricamente como computacionalmente. Por supuesto, el poder arribar a modelos con una precisión robusta en ese escenario daría una visibilidad del mapa competitivo a futuro que mejoraría al máximo la capacidad y anticipación de la toma de decisiones referidas a cuestiones tales como posicionamiento de precios, optimización de la inversión en marketing y publicidad, desarrollo de nuevos productos, marcas y calibres, entre otras.

Sin embargo, por lo expuesto anteriormente, este trabajo propone abordar esta problemática desde un nivel de agregación de la información que equilibre la practicidad en la construcción y entrenamiento de los modelos predictivos con la relevancia de los *insights* que puedan obtenerse del output de estos.

Cómo se ha introducido en el análisis exploratorio, dos dimensiones importantes a la hora de mapear el escenario competitivo de la industria cervecera argentina son el calibre y el segmento de venta. Por lo tanto, para una primera aproximación al modelo de industria que se plantea construir, se agrupará la información de venta mensual de cada área geográfica por cada una de estas variables, de la siguiente forma:

$$V_{jkl} = \sum_{i=1}^I X_{ijkl} \quad \forall j, k, l \quad (1)$$

$$V_{hkl} = \sum_{i=1}^I X_{ihkl} \quad \forall h, k, l \quad (2)$$

Donde V_{jkl} (V_{hkl}) representa la venta del j-ésimo segmento (h-ésimo calibre), en el k-ésimo área y el l-ésimo mes. X_{ijkl} (X_{ihkl}) representa la venta del i-ésimo producto, en el j-ésimo segmento (h-ésimo calibre) y el k-ésimo área, correspondiente al l-ésimo mes.

De esta forma, se construyeron dos nuevos conjuntos de datos; uno agrupando la venta a nivel área-segmento y otro agrupando a nivel área-calibre. No hay pérdida de volumen en el proceso, aunque se sacrifica granularidad de la información en pos de una reducción muy significativa en la dimensionalidad del problema.

Existe gran cantidad de literatura acerca de análisis y *forecasting* de series temporales. En este caso, para construir el modelo de industria, es necesario predecir múltiples series, que dependerán del tipo de apertura que se analice. En el caso de trabajar a nivel área-segmento, el desafío consiste en predecir 16 series simultáneamente (4 áreas x 4 segmentos). Si el análisis se realiza con apertura área-calibre, la cantidad de series se reduce a 12, ya que existen 3 clasificaciones posibles de esta variable en la base.

3.2 Tratamiento univariado

En primer lugar, es interesante analizar cada combinación área-segmento o área-calibre como si fueran independientes. Bajo este supuesto, se dan las condiciones para correr modelos univariados correspondientes a cada una de ellas, en donde el volumen futuro dependerá exclusivamente de los rezagos propios.

El objetivo es obtener proyecciones de venta para los próximos 12 meses, comenzando en enero de 2022. Esta cantidad de *steps* hacia adelante permitirá extrapolar los patrones inmersos en cada serie, contemplando la estacionalidad mensual.

El hecho de trabajar con series univariadas habilita la posibilidad de entrenar una variedad de modelos que no podrían ser considerados dentro de las opciones viables de otro modo.

Aun así, el desafío no deja de ser poder pronosticar todas las series que componen el mercado, cualquiera sea el nivel de apertura elegido. Por supuesto, encontrar un único modelo que sea funcional a todas las series temporales es poco probable, por lo cual se intentará, en esta

instancia, entrenar un conjunto de modelos sobre las series, evaluar su desempeño y finalmente tomar como modelo final aquel que arroje el menor error.

La métrica clave de performance que se utilizará para medir el *accuracy* de los modelos será el MAPE (*Mean Square Percentage Error* o Error Porcentual Absoluto Medio). Este indicador es particularmente útil en este caso, ya que trabajar con una medida de desvío en términos porcentuales permitirá analizar y contrastar modelos para distintas áreas, segmentos y calibres que, si bien se encuentran en la misma unidad de medida, en muchos casos difieren notablemente en su magnitud.

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \frac{|A_t - F_t|}{|A_t|} \quad (3)$$

Donde A_t y F_t representan el valor real de venta y el Forecast para el período t , respectivamente. n será igual a 12, ya que se propone predecir la venta del próximo año.

Para este fin, se utilizará *PyCaret*, una librería de Python que conforma una herramienta integral de gestión de modelos de aprendizaje automático. Sus funcionalidades van desde el análisis exploratorio y visualización de datos, hasta el entrenamiento simultaneo de varios modelos. La misma se nutre de elementos de otras librerías, tales como *Scikit-Learn*, y es capaz de trabajar con *dataframes* de *Pandas*, librería que se utilizó para poder cargar y manipular la base de datos. A su vez, también se utilizó *Matplotlib*, para poder visualizar gráficamente los datos, así como el *forecast* obtenido en cada caso.

Cómo plataforma para ejecutar código, se utilizó Google Colab, donde cómo primer paso se importó la base de datos y se construyeron 2 *dataframes*. Uno de ellos, agrupando el volumen de venta a nivel área-segmento. El otro, agrupado con apertura área-calibre. De esta forma, se obtuvieron observaciones sin *missing values*, para cada uno de los meses entre enero de 2018 y diciembre de 2021, siendo las columnas de cada set de datos, las correspondientes a cada combinación área-segmento o área-calibre, según corresponda, tal como se observa en las figuras 10 y 11. Recordar que, por cuestiones de disponibilidad, no se cuenta con datos anteriores a enero 2018.

Periodo	AREA ESTE_BOTELLA RET	AREA ESTE_LATA	AREA ESTE_OW s-LATA	AREA GBA_BOTELLA RET	AREA GBA_LATA	AREA GBA_OW s-LATA	AREA OESTE_BOTELLA RET	AREA OESTE_LATA	AREA OESTE_OW s-LATA	AREA SUR_BOTELLA RET	AREA SUR_LATA	AREA SUR_OW s-LATA
0 2018-01-01	12706.209037	8380.234301	1291.388425	235246.110569	167267.488803	27342.340388	12483.058403	8217.450182	2001.860297	24169.787027	18327.136325	4519.643967
1 2018-02-01	14666.361929	7711.960018	1350.242886	236150.755314	163438.122069	26258.075817	14762.941815	8061.792893	2062.900453	28420.806569	17130.188733	4737.881831
2 2018-03-01	15828.406071	9099.240310	1463.462048	238879.466521	155831.958897	26984.428259	14227.890106	8517.863833	2246.484430	26575.043446	15979.856988	4257.667933
3 2018-04-01	13186.872500	8917.726455	1134.469239	216619.195458	147686.264758	20864.953800	14382.642151	7380.096948	1713.178328	22633.201475	14868.128908	3522.721299
4 2018-05-01	9682.108306	7606.151274	955.544702	163351.968615	110963.266077	15920.059244	8820.006682	7691.723498	1278.014404	17376.269689	16030.581346	3164.538389

Figura 10. Head del dataset con nivel de apertura área-calibre.

Periodo	AREA ESTE_CORE	AREA ESTE_CORE PLUS	AREA ESTE_PREMIUM	AREA ESTE_VALUE	AREA GBA_CORE	AREA GBA_CORE PLUS	AREA GBA_PREMIUM	AREA GBA_VALUE	AREA OESTE_CORE	AREA OESTE_CORE PLUS	AREA OESTE_PREMIUM	AREA OESTE_VALUE	AREA SUR_CORE	AREA SUR_CORE PLUS	AREA SUR_PREMIUM	AREA SUR_VALUE
0 2018-01-01	14484.858900	2513.060618	3801.856522	1578.055722	263104.020048	48933.054922	87460.169956	29356.694834	14902.008706	2126.372644	4737.245930	936.741602	28292.226532	4369.601560	11163.225686	3191.513642
1 2018-02-01	14573.449005	3049.980478	4147.540343	1957.595008	261294.178237	52524.344154	84994.460408	27033.970401	17507.835231	1880.357756	4481.889122	1017.573051	30875.563733	5094.000038	11145.539381	3173.773981
2 2018-03-01	16016.688046	3075.578918	4503.228548	1996.094116	254573.980000	54065.442754	86488.506015	25967.324900	16414.251446	2421.977331	4710.721103	1445.288489	30446.623649	4153.987304	9670.766120	2540.991294
3 2018-04-01	14766.234176	3074.140993	3795.439169	1603.253837	249142.344575	47203.791417	68154.126760	20670.151344	16648.995773	2223.113175	3659.928507	943.887972	26262.232554	3292.277796	9150.259466	2319.281876
4 2018-05-01	11830.148763	2098.548555	2815.016558	1500.090406	188602.107170	33639.282985	52317.076642	17678.827139	12649.383852	1273.092265	2953.451880	913.816587	23199.789302	3935.209084	7320.840848	2115.568190

Figura 11. Head del dataset con nivel de apertura área-segmento.

Una vez organizados los datos de esta forma, se pueden graficar las series temporales que deben entrenarse para poder predecir sus valores futuros.

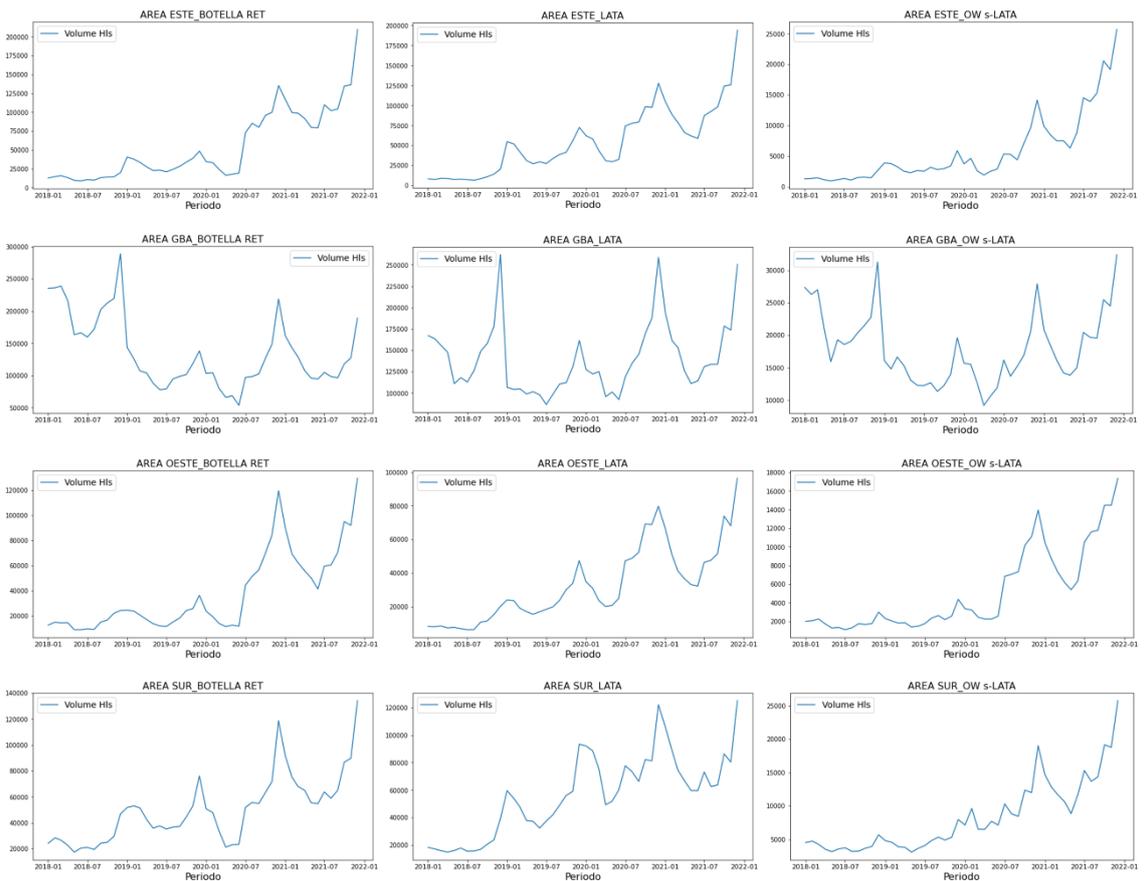


Figura 12. Evolución de la venta en hectolitros, a nivel área-calibre.

Esta visualización es importante porque da claridad sobre el hecho de que existe cierta diversidad entre el comportamiento del mercado como un todo (tal como se describe en la sección 2) y la evolución que se observa cuando tomamos una fracción de este.

No sólo el volumen que maneja cada región del país es distinto, sino que las tendencias que se pueden visualizar difieren entre las áreas. Por ejemplo, en GBA se ve una tendencia negativa en la venta de su botella retornable, así como también del segmento *Core*. Este comportamiento no se replica en las regiones del interior.

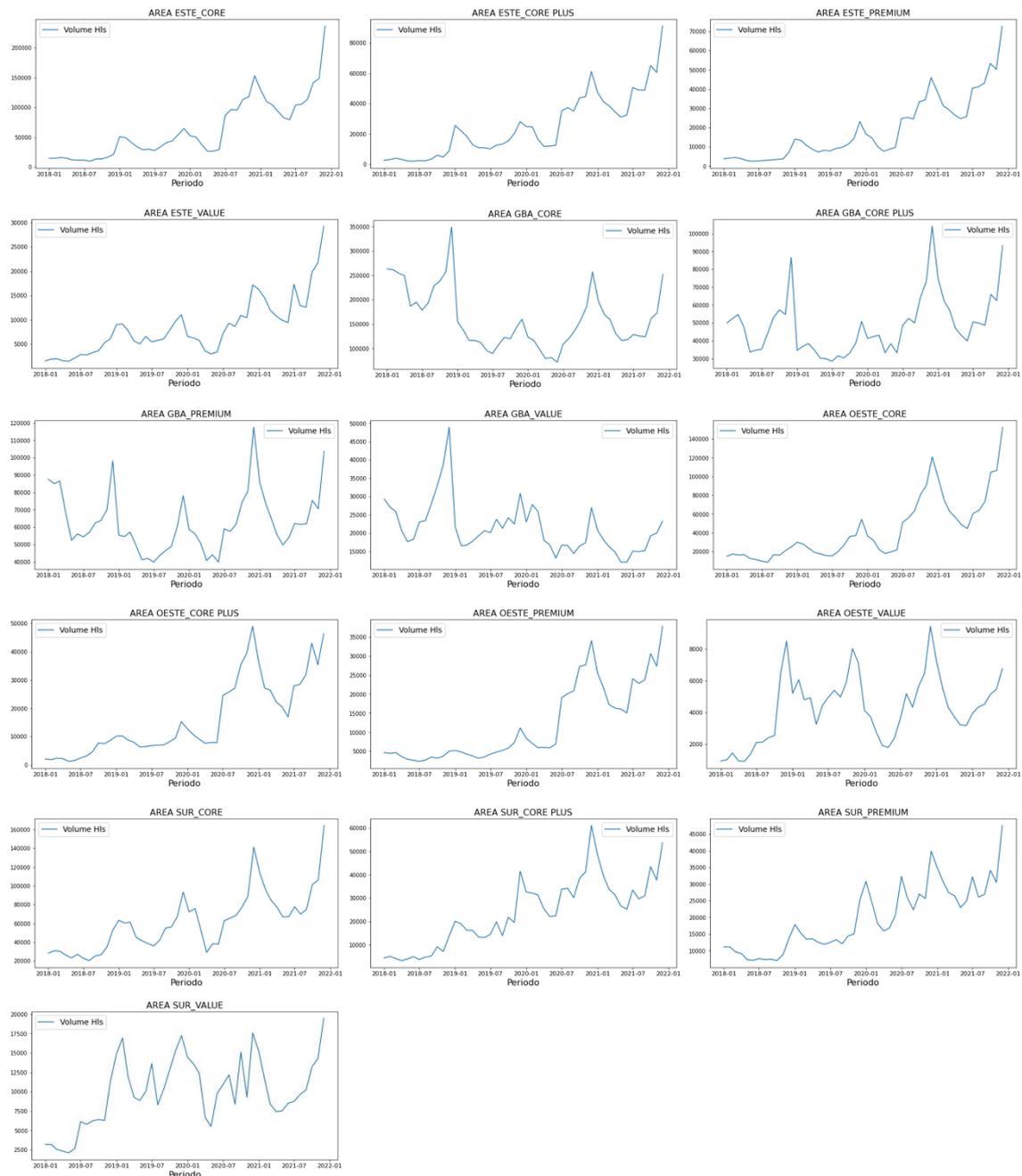


Figura 13. Evolución de la venta en hectolitros, a nivel área-segmento.

Entonces, se cuenta con 16 series para entrenar a nivel área-segmento y 12 a nivel área-calibre, con 48 observaciones y periodicidad mensual en todos los casos.

A nivel gráfico, dado que se observa cierta tendencia y estacionalidad en todas las series que componen nuestro conjunto de datos, resulta interesante analizar el método de Holt-Winters cómo una primera aproximación para poder predecir la venta futura.

Fue publicado inicialmente por Charles C. Holt en 1957 y actualmente goza de gran popularidad en lo que respecta a modelado y *forecasting* de series temporales. Dicho método, también conocido como Suavizado Exponencial Triple, es de naturaleza recursiva y busca modelar 3 aspectos fundamentales de una serie temporal; un valor típico o promedio, una pendiente o tendencia a lo largo del tiempo y un patrón de recurrencia cíclica o estacionalidad. Este último factor fue propuesto en 1960 por Peter Winters, un estudiante de Holt en la universidad de Texas en Austin.

El modelo en cuestión consiste en tres ecuaciones de suavizado fundamentales para realizar el fit de los datos, más una cuarta utilizada para computar las predicciones (Trull et al., 2020). Las proyecciones derivadas de este método están basadas en medias ponderadas de las observaciones pasadas, con el peso de estas decreciendo de forma exponencial a medida que nos alejamos del presente.

Un aspecto a tener en cuenta es que este modelo posee dos variantes; aditiva y multiplicativa, que esencialmente difieren en el modo en que es abordada la variabilidad correspondiente al factor estacional de la serie. Mientras que el método aditivo es recomendable cuando la variabilidad estacional es aproximadamente constante a lo largo del tiempo, el método multiplicativo suele funcionar mejor cuando nos encontramos con una variabilidad estacional que cambia de modo proporcional con el nivel de la serie (Hyndman y Athanasopoulos, 2018).

La variante multiplicativa presenta las siguientes ecuaciones para los factores de nivel, tendencia y estacionalidad, sumadas a la ecuación correspondiente para calcular los pronósticos.

$$\text{Nivel: } N_t = \alpha \left(\frac{X_t}{E_{t-m}} \right) + (1-\alpha)(N_{t-1} + T_{t-1})$$

$$\text{Tendencia: } T_t = \beta(N_t + N_{t-1}) + (1 - \beta)T_{t-1}$$

$$\text{Estacionalidad: } E_t = \gamma \left(\frac{X_t}{N_t} \right) + (1 - \gamma)E_{t-m}$$

$$\text{Forecast: } X_t(k) = (N_t + k T_t) + (1 - \beta)E_{t-m+k}$$

Donde α , β y γ son parámetros de suavizado de cada ecuación/factor. Todos ellos pueden tomar valores entre 0 y 1, e indican el nivel de ponderación que tendrán las observaciones pasadas cuando se estimen valores actuales y/o futuros. Por otro lado, m indica la cantidad de meses que conforman un ciclo estacional. Por ejemplo, si consideramos una estacionalidad mensual, su valor debe ser 12.

La variante aditiva del modelo de Holt-Winters contempla el siguiente sistema de ecuaciones:

$$\text{Nivel: } N_t = \alpha (X_t - E_{t-m}) + (1-\alpha)(N_{t-1} + T_{t-1})$$

$$\text{Tendencia: } T_t = \beta(N_t + N_{t-1}) + (1 - \beta)T_{t-1}$$

$$\text{Estacionalidad: } E_t = \gamma(X_t - N_{t-1} - T_{t-1}) + (1 - \gamma)E_{t-m}$$

$$\text{Forecast: } X_t(k) = N_t + k T_t + E_{t-m+k}$$

Previo a entrenar este tipo de modelos en nuestras series temporales, utilizando el paquete *statsmodels* de Python, se extrajo la tendencia y estacionalidad de cada una de las combinaciones área-segmento y área-calibre.

Lo que se observa en prácticamente todas las series es un ciclo estacional que se repite por año calendario, con picos en el mes de diciembre y valores mínimos en junio/julio.

Esto da indicios favorables para imponer $m=12$ y trabajar con estacionalidad mensual a la hora de entrenar los modelos de Holt-Winters.

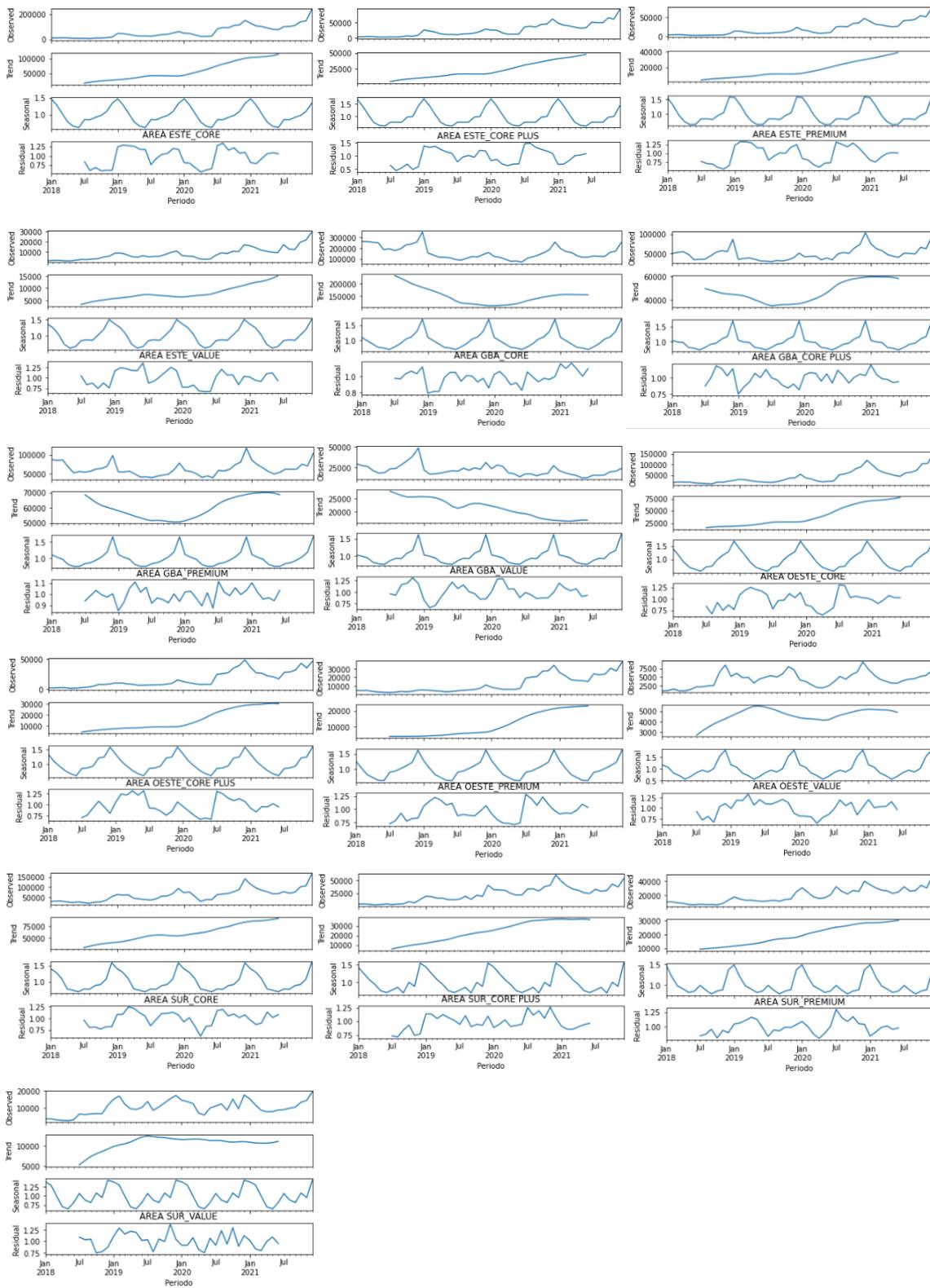


Figura 14. Descomposición de las series temporales área-segmento.

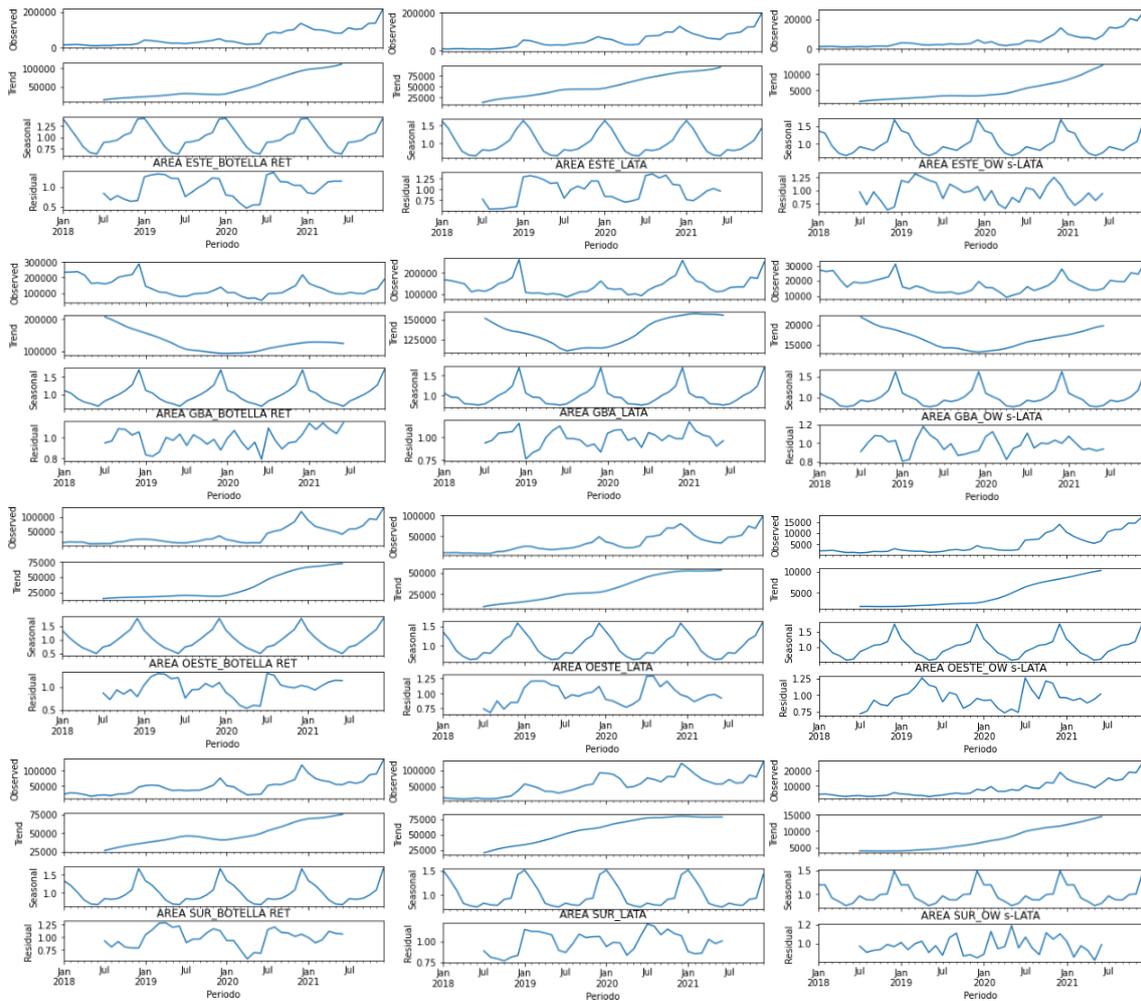


Figura 15. Descomposición de las series temporales área-calibre.

Se procedió a realizar el fit del modelo Holt-Winters en todas las series, tanto es su versión con estacionalidad aditiva como multiplicativa, y se computó el R^2 para determinar cuál es la variante que mejor aplica en estos casos.

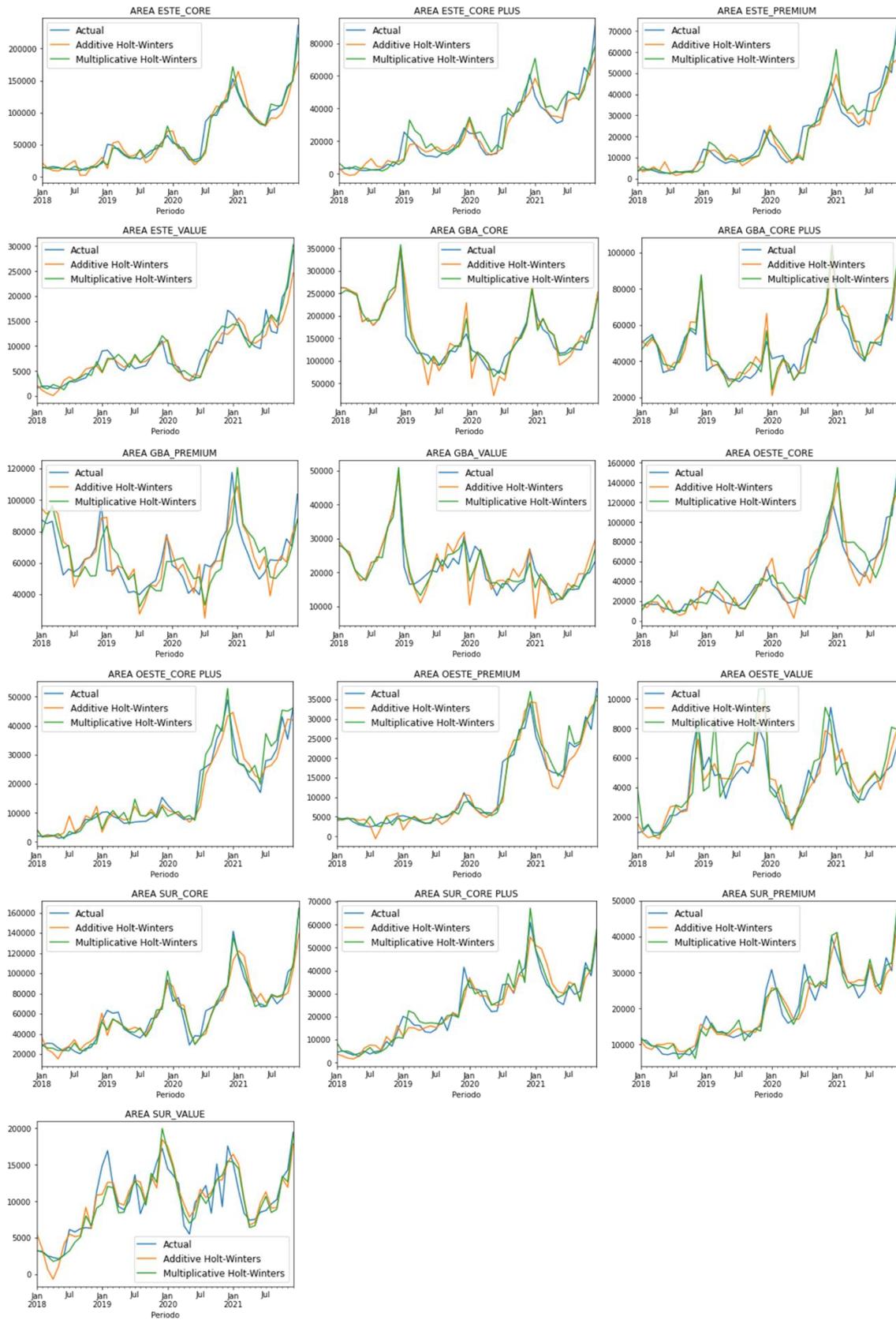


Figura 16. Fit de los modelos aditivo y multiplicativo de Holt-Winters en las series área-segmento.

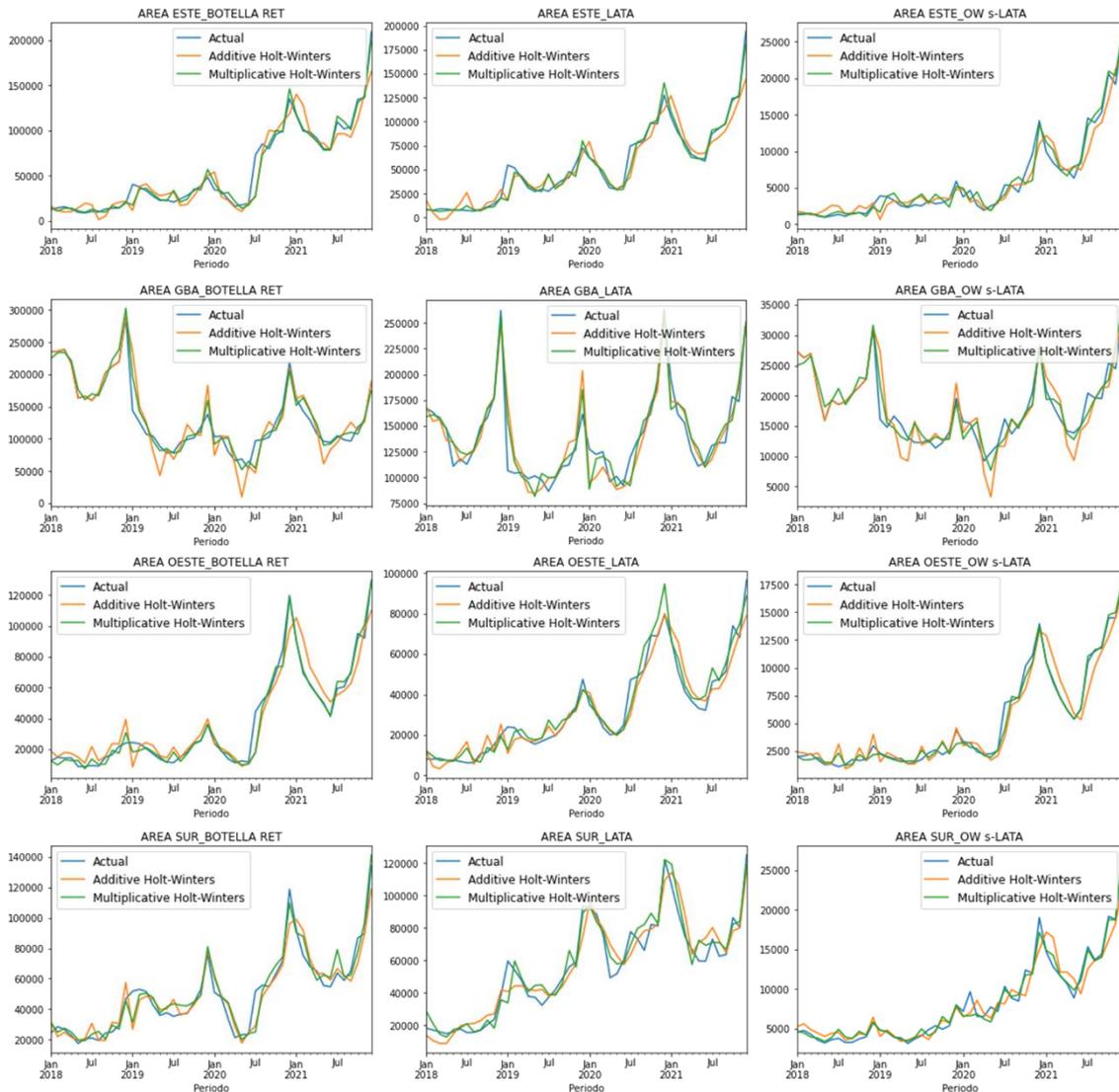


Figura 17. Fit de los modelos aditivo y multiplicativo de Holt-Winters en las series área-calibre.

Se pudo comprobar que, en la mayoría de los casos, el modelo con factores multiplicativos logra explicar una mayor proporción de la variabilidad de la evolución de la venta, tanto de los distintos segmentos como de los calibres. Esto queda evidenciado en la comparación del R^2 para cada caso, que queda plasmada en la tabla 3.

Área-Segmento			Área-Calibre		
Serie	R2 add	R2 mul	Serie	R2 add	R2 mul
AREA ESTE_CORE	0,89	0,96	AREA ESTE_BOTELLA RET	0,91	0,96
AREA ESTE_CORE PLUS	0,90	0,86	AREA ESTE_LATA	0,89	0,96
AREA ESTE_PREMIUM	0,92	0,92	AREA ESTE_OW s-LATA	0,92	0,96
AREA ESTE_VALUE	0,86	0,89	AREA GBA_BOTELLA RET	0,81	0,93
AREA GBA_CORE	0,79	0,92	AREA GBA_LATA	0,84	0,87
AREA GBA_CORE PLUS	0,84	0,89	AREA GBA_OW s-LATA	0,71	0,87
AREA GBA_PREMIUM	0,73	0,73	AREA OESTE_BOTELLA RET	0,91	0,97
AREA GBA_VALUE	0,60	0,84	AREA OESTE_LATA	0,92	0,94
AREA OESTE_CORE	0,91	0,91	AREA OESTE_OW s-LATA	0,94	0,98
AREA OESTE_CORE PLUS	0,91	0,92	AREA SUR_BOTELLA RET	0,88	0,91
AREA OESTE_PREMIUM	0,92	0,96	AREA SUR_LATA	0,91	0,93
AREA OESTE_VALUE	0,81	0,48	AREA SUR_OW s-LATA	0,91	0,98
AREA SUR_CORE	0,88	0,95			
AREA SUR_CORE PLUS	0,91	0,93			
AREA SUR_PREMIUM	0,90	0,94			
AREA SUR_VALUE	0,78	0,80			

Tabla 3. Comparación del R^2 obtenido bajo los modelos aditivo y multiplicativo de Holt-Winters.

Se decidió, entonces, avanzar con el entrenamiento del modelo multiplicativo de Holt-Winters.

Se procedió a separar todas las series en un conjunto de entrenamiento y otro de testeo. Para el primero, se tomaron los primeros 36 meses de la base datos, es decir, desde enero de 2018 a diciembre de 2020. El año 2021 fue separado para formar parte del segundo conjunto.

Luego, se entrenó el modelo y se computaron los valores de las proyecciones para el año 2021, las cuales finalmente fueron contrastadas con los valores reales de dicho año en cada caso, obteniendo el MAPE para cada serie.

Dada la heterogeneidad entre series que parece observarse gráficamente en las figuras 12 y 13, y la variedad de resultados obtenidos luego de correr el modelo de Holt-Winters en todas ellas, es deseable poder testear y entrenar una amplia variedad de modelos en cada caso. Esto es así ya que es posible que, según el comportamiento de la venta en cada segmento y/o calibre, tengan que considerarse distintos aspectos que no puedan ser capturados por un único modelo.

PyCaret tiene 2 módulos que son capaces de llevar a cabo estas tareas de aprendizaje supervisado; El módulo *Regression* y el módulo *Time Series*. Cabe destacar que este último está en fase Beta, por lo cual se utilizaron ambos módulos para el análisis. Además, cada uno de ellos cuenta con distintos sets de modelos para entrenar.

El módulo *Regression* contiene los siguientes algoritmos, que fueron testeados para cada serie de los dataset construidos, siempre que fuera posible:

ID	Model
lr	Linear Regression
lasso	Lasso Regression
ridge	Ridge Regression
en	Elastic Net
lar	Least Angle Regression
llar	Lasso Least Angle Regression
omp	Orthogonal Matching Pursuit
br	Bayesian Ridge
ard	Automatic Relevance Determination
par	Passive Aggressive Regressor
ransac	Random Sample Consensus
tr	TheilSen Regressor
huber	Huber Regressor
kr	Kernel Ridge
svm	Support Vector Machine
knn	K Neighbors Regressor
dt	Decision Tree
rf	Random Forest
et	Extra Trees Regressor
ada	AdaBoost Regressor
gbr	Gradient Boosting Regressor
mlp	Multi Level Perceptron
xgboost	Extreme Gradient Boosting
lightgbm	Light Gradient Boosting Machine
catboost	CatBoost Regressor

Figura 18. Set de modelos disponibles en el módulo *Regression* de *PyCaret*.

En todos los casos, se dividió cada serie en un conjunto de entrenamiento y un conjunto de *testing*, para evaluar la performance de cada modelo. Cabe destacar que, como se está trabajando con series temporales, no sería correcto elegir el set de *testing* de manera aleatoria como se suele hacer en otros tipos de problemas de *machine learning*, donde las observaciones son independientes entre sí. En este caso, esa dependencia existe y, por ello, se deben tomar ciertos períodos (observaciones) para testear la performance de nuestros modelos. Particularmente, la parte final del set de datos, tal de que el entrenamiento y posterior evaluación de los modelos se haga de manera correcta (Radečić, 2021).

Se procedió a entrenar todos los modelos del módulo *Regression* de *PyCaret* y se construyó un ranking de performance para cada serie, ordenado por MAPE.

Una vez obtenidos los mejores modelos para cada serie, se procedió a realizar las predicciones para el año 2022, agregando los 12 meses de ese año a la venta histórica. Con esta información, estaríamos en condiciones de construir el mapa competitivo proyectado.

También, se corrió el proceso análogo con el set de modelos incluidos en el módulo *Time_Series* de *PyCaret*.

ID	Model
naive	Naive Forecaster
grand_means	Grand Means Forecaster
polytrend	Polynomial Trend Forecaster
arima	ARIMA
auto_arima	Auto ARIMA
exp_smooth	Exponential Smoothing
croston	Croston
ets	ETS
theta	Theta Forecaster
tbats	TBATS
bats	BATS
lr_cds_dt	Linear w/ Cond. Deseasonalize & Detrending
en_cds_dt	Elastic Net w/ Cond. Deseasonalize & Detrending
ridge_cds_dt	Ridge w/ Cond. Deseasonalize & Detrending
lasso_cds_dt	Lasso w/ Cond. Deseasonalize & Detrending
lar_cds_dt	Least Angular Regressor w/ Cond. Deseasonalize & Detrending
llar_cds_dt	Lasso Least Angular Regressor w/ Cond. Deseasonalize & Detrending
br_cds_dt	Bayesian Ridge w/ Cond. Deseasonalize & Detrending
huber_cds_dt	Huber w/ Cond. Deseasonalize & Detrending
par_cds_dt	Passive Aggressive w/ Cond. Deseasonalize & Detrending
omp_cds_dt	Orthogonal Matching Pursuit w/ Cond. Deseasonalize & Detrending
knn_cds_dt	K Neighbors w/ Cond. Deseasonalize & Detrending
dt_cds_dt	Decision Tree w/ Cond. Deseasonalize & Detrending
rf_cds_dt	Random Forest w/ Cond. Deseasonalize & Detrending
et_cds_dt	Extra Trees w/ Cond. Deseasonalize & Detrending
gbr_cds_dt	Gradient Boosting w/ Cond. Deseasonalize & Detrending
ada_cds_dt	AdaBoost w/ Cond. Deseasonalize & Detrending
lightgbm_cds_dt	Light Gradient Boosting w/ Cond. Deseasonalize & Detrending

Figura 19. Set de modelos disponibles en el módulo *Time_Series* de *PyCaret*.

Cómo se observa, se entrenaron una gran variedad de modelos, desde procesos lineales simples cómo ARIMA y auto-ARIMA hasta modelos de árboles de decisión tales como *Random Forest* o *Light Gradient Boosting Machines*, con variantes para tratar la estacionalidad y la tendencia de la serie a entrenar.

3.3 Tratamiento multivariado

Una vez que se obtuvieron los modelos univariados de mejor performance según MAPE, los mismos constituyen un *benchmark* a la hora de intentar refinar los resultados. El supuesto de que cada serie en nuestros *dataset* es independiente y puede ser entrenada de manera aislada es poco robusto.

Naturalmente, uno esperaría que exista algún tipo de correlación entre series que componen el mercado. Esto puede ayudar a mejorar la precisión de los modelos, si se logra identificarlas correctamente.

Para este fin, utilizando la librería *Seaborn*, se construyó un mapa de calor con las correlaciones de las distintas series, tanto para el caso área-segmento como para el caso área-calibre.



Figura 20. Mapa de calor representando las correlaciones entre las series área-segmento.

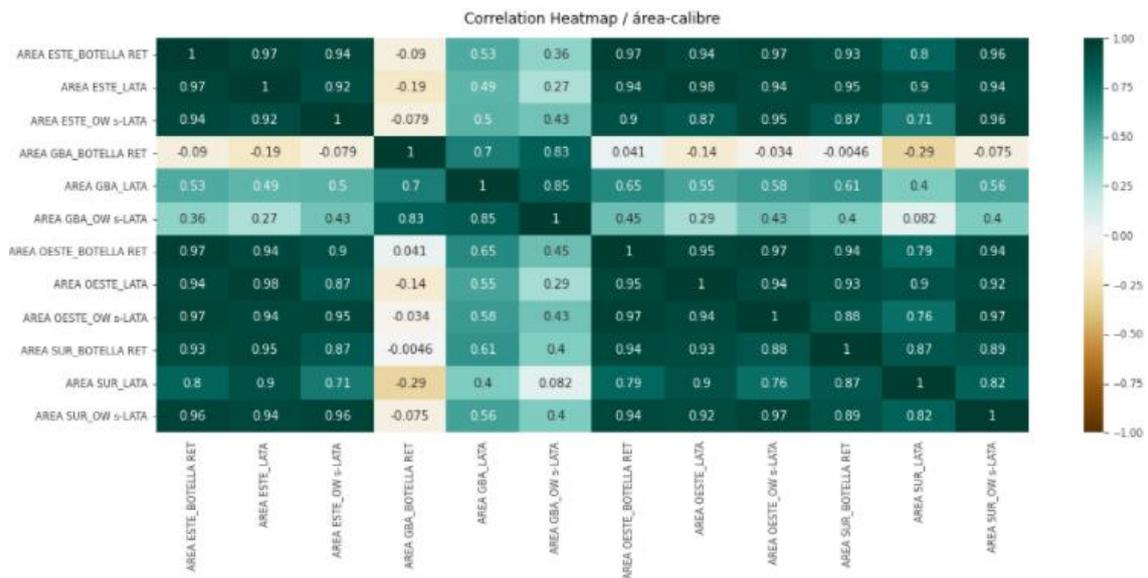


Figura 21. Mapa de calor representando las correlaciones entre las series área-calibre.

Un hallazgo interesante al observar las correlaciones es el hecho de que, a nivel área, tanto los distintos segmentos como los calibres tienen un valor cercano a 1. Uno podría haber intuido que, dentro de una región geográfica, si el mercado fuera estable, el crecimiento de algún segmento o calibre tendería a derivar en el decrecimiento de algún otro, por lo que deberían observarse correlaciones negativas. Esto no es así ya que el mercado argentino, como se ha mencionado anteriormente, tiene una tendencia creciente en los últimos años. Lo que sí puede suceder, es que cambien los pesos de los segmentos o calibres en determinada área, o mismo que se modifique la distribución de la venta entre áreas, hipótesis que podrá comprobarse una vez que se obtenga el output final para los niveles de apertura elegidos y se construya la proyección de la industria en 2022.

Por lo tanto, en esta instancia, se decidió partir los set de datos que se construyeron inicialmente en 4; uno por cada área geográfica, para llevar a cabo el análisis y entrenamiento de modelos multivariados.

Para llevar a cabo esta tarea, se utilizó extensivamente la librería *Darts*, desarrollada por la compañía suiza *Unit8* en 2020. Además de modelos estadísticos clásicos de *forecasting* de series temporales, esta librería es capaz de entrenar modelos de regresión y de redes neuronales. Además, cuenta con soporte para modelos multivariados y entrenamiento de varias series simultáneamente.

Se dividieron nuevamente las series en un set de entrenamiento y un set de testeo. El primero se conformó con las observaciones comprendidas entre los años 2018 y 2020, mientras que el año 2021 fue separado para usarse a la hora de validar los modelos.

Lo que se buscó es entonces, sacar provecho de las correlaciones de las distintas series de venta de segmentos y calibres dentro de cada área, entrenando modelos multivariados.

Los modelos VAR son la generalización, a múltiples variables, de los modelos univariados autoregresivos (AR). La representación de los valores para cada período temporal t es en forma de un vector de N variables, donde cada posición en dicho vector es el valor que toma cada serie (en este caso, el volumen de venta) en el período en cuestión.

Sin pérdida de generalidad, en vectores autorregresivos, se intenta relacionar el valor presente de cada variable con el pasado del resto de las variables incluidas en el modelo (Hyndman y Athanasopoulos, 2018).

Llevado a nuestro caso, un modelo de este tipo a nivel área es representado de la siguiente forma:

$$V_{i,t} = C_i + \sum_{p=1}^P \phi_{ii,p} V_{i,t-p} + \sum_{p=1}^P \sum_{j=1}^J \phi_{ij,p} V_{j,t-p} + e_{i,t} \quad (4)$$

Donde $j \neq i$ y $V_{i,t}$ representa el volumen de venta del i -ésimo segmento o calibre, en el período t . $\phi_{ii,p}$ es un coeficiente que representa el impacto del p -ésimo rezago de la variable V_i sobre sí misma, mientras que $\phi_{ij,p}$ representa el poder de influencia del p -ésimo rezago del j -ésimo segmento/calibre sobre $V_{i,t}$. $e_{i,t}$ representa el término de error.

Agregando el componente correspondiente a un modelo de medias móviles, es posible construir un modelo VARMA o VARIMA, si incluimos dentro del mismo el orden de integración de las series en cada caso.

$$V_{i,t} = C_i + \sum_{p=1}^P \phi_{ii,p} V_{i,t-p} + \sum_{p=1}^P \sum_{j=1}^J \phi_{ij,p} V_{j,t-p} + e_{i,t} + \sum_{q=1}^Q \theta_{i,q} e_{t-q} + \sum_{q=1}^Q \sum_{j=1}^J \theta_{j,q} e_{t-q} \quad (5)$$

Por otro lado, es posible agregar covariables a un modelo VARIMA, que no sean estrictamente el volumen de venta de algún segmento o calibre, pero que sigan la misma estructura de serie temporal que estos últimos. En la base de datos utilizada en este trabajo, como se observa en la tabla 1, se cuenta con información del importe monetario asociado a la venta de cada producto, en cada período disponible. Para que la información de precios sea relevante, se decidió transformarla a una misma unidad de medida en cuanto a importe y volumen. Esto quiere decir que no solamente importará que la unidad monetaria sea equivalente en todos los casos. Naturalmente, al contar con envases de aproximadamente medio litro, el precio absoluto de las latas es menor al de los productos de botella retornable, donde mayoritariamente el contenido es de 1 litro. No obstante, el precio promedio por litro de la lata es mayor al del litro, por lo que el consumidor obtiene una peor relación contenido/importe en la lata a pesar de tener esta última, en promedio, un precio nominal menor. Este ejemplo ilustra el problema que se tendría al comparar precios de distintas unidades de volumen. Se procedió entonces a calcular el precio por litro, que se puede obtener a través de la división entre las variables *Importe_Item* y *VolumenLts* en cada caso.

Cómo se decidió trabajar la información a nivel segmento y calibre, se necesita el precio promedio por litro de cada segmento y calibre, en cada área, para cada período disponible. Es posible obtenerlo utilizando las ecuaciones 1 y 2, dividiendo la sumatoria de importes monetarios asociados a cada producto del segmento/calibre en cuestión sobre el volumen correspondiente de dichos productos. Esto queda plasmado en las ecuaciones 6 y 7.

$$P_{jkl} = \frac{\sum_{i=1}^I I_{ijkl}}{\sum_{i=1}^I X_{ijkl}} \quad \forall j, k, l \quad (6)$$

$$P_{hkl} = \frac{\sum_{i=1}^I I_{ihkl}}{\sum_{i=1}^I X_{ihkl}} \quad \forall h, k, l \quad (7)$$

Donde P_{jkl} (P_{hkl}) representa el precio promedio por litro del j-ésimo segmento (h-ésimo calibre) en el k-ésimo área, correspondiente al l-ésimo período o mes e I_{ijkl} (I_{ihkl}) representa el importe total del i-ésimo producto, asociado al j-ésimo segmento (h-ésimo calibre), en el k-ésimo área y correspondiente al l-ésimo mes. Las definiciones de X_{ijkl} e X_{ihkl} son análogas a las brindadas en la introducción de las ecuaciones 1 y 2.

De esta forma, las series de precio construidas están en términos nominales, lo cual no termina de resultar útil para estudiar cómo afecta el precio de la cerveza al volumen de la industria. Sobre todo en un contexto de alta inflación como el que existe en el país, dichas series nominadas en pesos argentinos serán estrictamente crecientes en el tiempo. Una opción para contrarrestar esta situación es utilizar precios reales, ajustándolos por inflación mediante el Índice de Precios al Consumidor. Otra alternativa, con la cual se decidió avanzar, es calcular el ratio entre el precio por litro de determinado segmento o calibre y el precio por litro promedio de la industria en su conjunto, en un periodo determinado. Se prefiere esta opción dado que, en los modelos multivariados, se intenta explotar las correlaciones entre el comportamiento de la venta de los distintos segmentos y calibres dentro de un área. Realizando la transformación recién mencionada, será posible entender en cierto modo el nivel de esfuerzo que requiere adquirir, en promedio, un producto de un segmento o calibre en particular, contrastado con lo que se desembolsa en promedio en toda la industria.

Se plasma la construcción de estas nuevas variables en las ecuaciones 8 y 9:

$$Px_{jkl} = \frac{P_{jkl}}{\frac{\sum_{j=1}^J \sum_{i=1}^I I_{ijkl}}{\sum_{j=1}^J \sum_{i=1}^I X_{ijkl}}} \quad \forall j, k, l \quad (8)$$

$$Px_{hkl} = \frac{P_{hkl}}{\frac{\sum_{h=1}^H \sum_{i=1}^I I_{ihkl}}{\sum_{h=1}^H \sum_{i=1}^I X_{ihkl}}} \quad \forall h, k, l \quad (9)$$

Donde Px_{jkl} (Px_{hkl}) es el ratio entre el precio promedio por litro del j-ésimo segmento (h-ésimo calibre) en el k-ésimo área y el l-ésimo mes, sobre el precio promedio por litro de la totalidad de los segmentos (calibres) en dicho área y mes en cuestión.

En las figuras 22 y 23 se grafican las series de las variables Px , junto con el mapa de correlación entre ellas y las series de volumen. Se observa que existen varios pares de correlación negativa. Tendría sentido pensar que, a mayor nivel de Px , es decir, a mayor precio que tenga un determinado segmento o calibre respecto al promedio de la industria, su propio volumen tienda a caer. De todas formas, no estamos en condiciones de afirmar nada semejante, puesto que la correlación no implica causalidad (O'Neil y Schutt, 2013).

A su vez, se evidencia la jerarquía de los distintos segmentos en cada una de las áreas. En orden ascendente respecto a Px , el *Value* encabeza la lista, pasando luego por el *Core* y *Core Plus*, para llegar por último al nivel de precios del *Premium*. No obstante, se ve una tendencia general a la baja en el ratio de este último segmento en todos los casos.

En cuanto a la evolución de Px en los distintos calibres, se observa que OW s-lata es el de mayor índice, aunque también es notoria su tendencia a la baja en todas las áreas. Lo sigue la lata que, a partir de mediados de 2020, comienza a encarecerse relativamente. Por último, aparece la botella retornable, que ofrece el precio por litro más bajo al consumidor.

Se decidió entrenar entonces, además del modelo VARIMA descrito anteriormente, uno del mismo tipo, pero que incorpore estas covariables de precio.

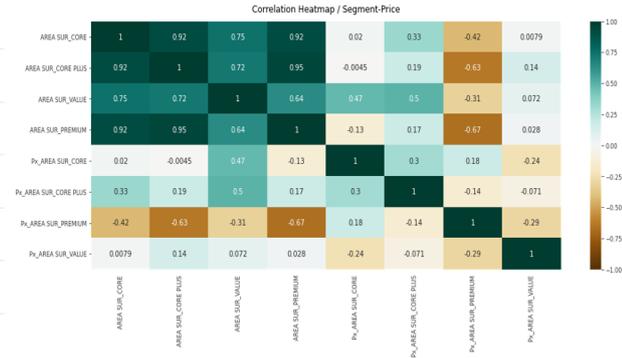
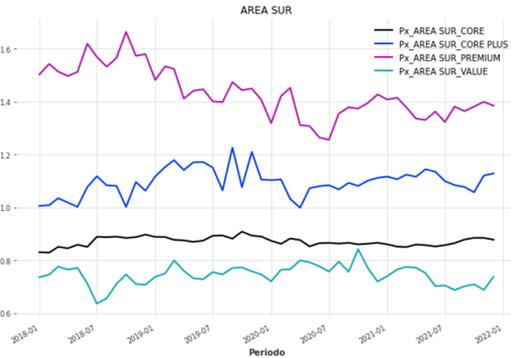
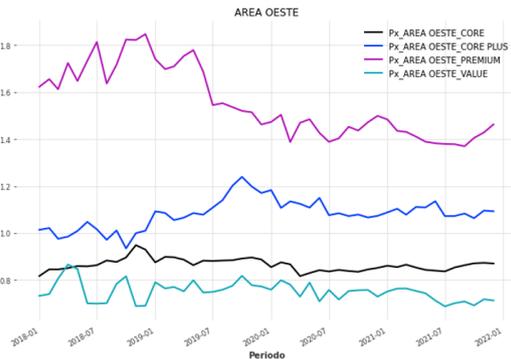
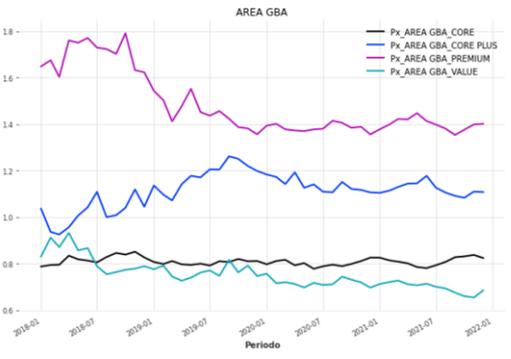


Figura 22. Evolución de precios y mapas de correlación con el volumen de venta, por segmento.

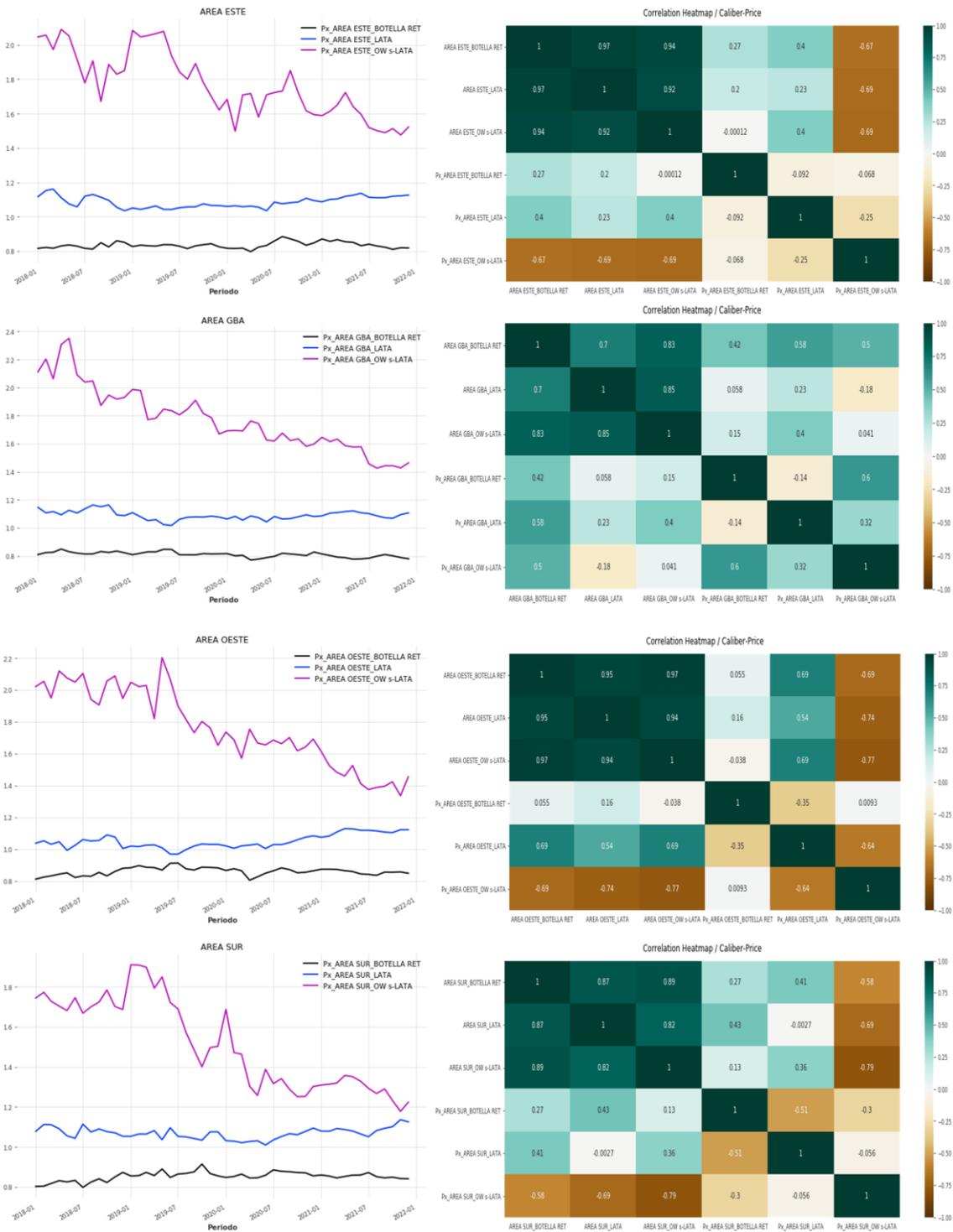


Figura 23. Evolución de precios y mapas de correlación con el volumen de venta, por calibre.

4. Resultados

4.1 Modelos univariados

Cómo se introdujo en la sección 3, una primera aproximación para entrenar modelos univariados fue el método Holt-Winters. Se comprobó que, en este caso, el modelo multiplicativo explica una mayor proporción de la variabilidad de las series, por lo cual se utilizó esa variante para entrenar los datos correspondientes a los años 2018, 2019 y 2020. Luego, se realizaron las predicciones para el año 2021, computando el MAPE en cada caso.

Área-segmento		Área-calibre	
Serie	MAPE	Serie	MAPE
AREA ESTE_CORE	31%	AREA ESTE_BOTELLA RET	25%
AREA ESTE_CORE PLUS	23%	AREA ESTE_LATA	19%
AREA ESTE_PREMIUM	25%	AREA ESTE_OW s-LATA	18%
AREA ESTE_VALUE	19%	AREA GBA_BOTELLA RET	28%
AREA GBA_CORE	15%	AREA GBA_LATA	10%
AREA GBA_CORE PLUS	27%	AREA GBA_OW s-LATA	22%
AREA GBA_PREMIUM	39%	AREA OESTE_BOTELLA RET	83%
AREA GBA_VALUE	22%	AREA OESTE_LATA	70%
AREA OESTE_CORE	31%	AREA OESTE_OW s-LATA	64%
AREA OESTE_CORE PLUS	78%	AREA SUR_BOTELLA RET	12%
AREA OESTE_PREMIUM	86%	AREA SUR_LATA	37%
AREA OESTE_VALUE	48%	AREA SUR_OW s-LATA	51%
AREA SUR_CORE	18%		
AREA SUR_CORE PLUS	73%		
AREA SUR_PREMIUM	29%		
AREA SUR_VALUE	21%		

Tabla 4. MAPE obtenido bajo el modelo de Holt-Winters.

Cómo se observa, los valores de MAPE obtenidos son muy variados. El modelo de Holt-Winters tuvo buen rendimiento en AREA ESTE_VALUE, AREA GBA_CORE y AREA SUR_CORE por el lado de la venta por segmento, mientras que AREA ESTE_LATA, AREA ESTE_OW s-LATA, AREA GBA_LATA y AREA SUR_BOTELLA RET fueron las series con mejor desempeño en la venta por calibre. En todos estos casos, el valor de MAPE se situó debajo de 20%.

Aun así, la amplia variedad en los resultados refuerza la idea de que, probablemente, sea difícil encontrar un único modelo que minimice el MAPE en todos los casos. De aquí la motivación para

experimentar con las herramientas que nos provee *PyCaret* para poder entrenar múltiples modelos en cada una de las series.

Los resultados de los modelos obtenidos utilizando la librería *PyCaret* se muestran a continuación.

A partir del entrenamiento de los modelos detallados en las figuras 18 y 19, se seleccionaron los que mejor rendimiento (menor MAPE) tuvieron para cada serie área-segmento y área calibre.

	Model	MAPE	time_series
knn	K Neighbors Regressor	0.4028	AREA ESTE_CORE
en	Elastic Net	0.3758	AREA ESTE_CORE PLUS
ridge	Ridge Regression	0.3717	AREA ESTE_PREMIUM
en	Elastic Net	0.3598	AREA ESTE_VALUE
gbr	Gradient Boosting Regressor	0.2705	AREA GBA_CORE
par	Passive Aggressive Regressor	0.1940	AREA GBA_CORE PLUS
gbr	Gradient Boosting Regressor	0.1666	AREA GBA_PREMIUM
ridge	Ridge Regression	0.2011	AREA GBA_VALUE
ridge	Ridge Regression	0.3857	AREA OESTE_CORE
ridge	Ridge Regression	0.3625	AREA OESTE_CORE PLUS
omp	Orthogonal Matching Pursuit	0.3953	AREA OESTE_PREMIUM
omp	Orthogonal Matching Pursuit	0.3139	AREA OESTE_VALUE
ridge	Ridge Regression	0.2464	AREA SUR_CORE
ridge	Ridge Regression	0.2041	AREA SUR_CORE PLUS
lasso	Lasso Regression	0.1300	AREA SUR_PREMIUM
knn	K Neighbors Regressor	0.2504	AREA SUR_VALUE

Figura 24. Mejores modelos área-segmento univariados del módulo *Regression* de *PyCaret*.

	Model	MAPE	time_series
polytrend	Polynomial Trend Forecaster	0.3545	AREA ESTE_CORE
lightgbm_cds_dt	Light Gradient Boosting w/ Cond. Deseasonalize...	0.3197	AREA ESTE_CORE PLUS
polytrend	Polynomial Trend Forecaster	0.3254	AREA ESTE_PREMIUM
lightgbm_cds_dt	Light Gradient Boosting w/ Cond. Deseasonalize...	0.1999	AREA ESTE_VALUE
croston	Croston	0.2018	AREA GBA_CORE
grand_means	Grand Means Forecaster	0.2477	AREA GBA_CORE PLUS
grand_means	Grand Means Forecaster	0.1722	AREA GBA_PREMIUM
br_cds_dt	Bayesian Ridge w/ Cond. Deseasonalize & Detren...	0.2092	AREA GBA_VALUE
lightgbm_cds_dt	Light Gradient Boosting w/ Cond. Deseasonalize...	0.4162	AREA OESTE_CORE
lightgbm_cds_dt	Light Gradient Boosting w/ Cond. Deseasonalize...	0.4606	AREA OESTE_CORE PLUS
croston	Croston	0.4537	AREA OESTE_PREMIUM
grand_means	Grand Means Forecaster	0.2673	AREA OESTE_VALUE
lightgbm_cds_dt	Light Gradient Boosting w/ Cond. Deseasonalize...	0.1834	AREA SUR_CORE
dt_cds_dt	Decision Tree w/ Cond. Deseasonalize & Detrending	0.2455	AREA SUR_CORE PLUS
br_cds_dt	Bayesian Ridge w/ Cond. Deseasonalize & Detren...	0.1795	AREA SUR_PREMIUM
grand_means	Grand Means Forecaster	0.2321	AREA SUR_VALUE

Figura 25. Mejores modelos área-segmento univariados del módulo *Time Series* de *PyCaret*.

	Model	MAPE	time_series
en	Elastic Net	0.4600	AREA ESTE_BOTELLA RET
ridge	Ridge Regression	0.3047	AREA ESTE_LATA
ridge	Ridge Regression	0.3566	AREA ESTE_OW s-LATA
gbr	Gradient Boosting Regressor	0.3171	AREA GBA_BOTELLA RET
rf	Random Forest Regressor	0.1617	AREA GBA_LATA
gbr	Gradient Boosting Regressor	0.2233	AREA GBA_OW s-LATA
huber	Huber Regressor	0.4770	AREA OESTE_BOTELLA RET
ridge	Ridge Regression	0.2391	AREA OESTE_LATA
ridge	Ridge Regression	0.4161	AREA OESTE_OW s-LATA
en	Elastic Net	0.3170	AREA SUR_BOTELLA RET
rf	Random Forest Regressor	0.1949	AREA SUR_LATA
omp	Orthogonal Matching Pursuit	0.2964	AREA SUR_OW s-LATA

Figura 26. Mejores modelos área-calibre univariados del módulo *Regression* de *PyCaret*.

	Model	MAPE	time_series
polytrend	Polynomial Trend Forecaster	0.3973	AREA ESTE_BOTELLA RET
lightgbm_cds_dt	Light Gradient Boosting w/ Cond. Deseasonalize...	0.3352	AREA ESTE_LATA
polytrend	Polynomial Trend Forecaster	0.2696	AREA ESTE_OW s-LATA
croston	Croston	0.1963	AREA GBA_BOTELLA RET
grand_means	Grand Means Forecaster	0.1986	AREA GBA_LATA
croston	Croston	0.1592	AREA GBA_OW s-LATA
par_cds_dt	Passive Aggressive w/ Cond. Deseasonalize & De...	0.4333	AREA OESTE_BOTELLA RET
croston	Croston	0.4077	AREA OESTE_LATA
lightgbm_cds_dt	Light Gradient Boosting w/ Cond. Deseasonalize...	0.4564	AREA OESTE_OW s-LATA
lightgbm_cds_dt	Light Gradient Boosting w/ Cond. Deseasonalize...	0.2203	AREA SUR_BOTELLA RET
dt_cds_dt	Decision Tree w/ Cond. Deseasonalize & Detrending	0.1854	AREA SUR_LATA
ada_cds_dt	AdaBoost w/ Cond. Deseasonalize & Detrending	0.1993	AREA SUR_OW s-LATA

Figura 27. Mejores modelos área-calibre univariados del módulo *Time Series* de *PyCaret*.

Como se observa, se obtuvieron resultados variados según la serie temporal, módulo de *PyCaret* y modelo seleccionado. Dado que ambos módulos se utilizaron para el mismo fin, se pueden combinar los resultados y tomar finalmente el modelo de menor MAPE por serie, según si este haya sido obtenido mediante el mejor modelo del módulo *Regression* o del módulo *Time Series*.

Serie	Modelo seleccionado	MAPE
AREA ESTE_CORE	polytrend	35.45%
AREA ESTE_CORE PLUS	lightgbm_cds_dt	31.97%
AREA ESTE_PREMIUM	polytrend	32.54%
AREA ESTE_VALUE	lightgbm_cds_dt	19.99%
AREA GBA_CORE	croston	20.18%
AREA GBA_CORE PLUS	par	19.40%
AREA GBA_PREMIUM	gbr	16.66%
AREA GBA_VALUE	ridge	20.11%
AREA OESTE_CORE	ridge	38.57%
AREA OESTE_CORE PLUS	ridge	36.25%
AREA OESTE_PREMIUM	omp	39.53%
AREA OESTE_VALUE	grand_means	26.73%
AREA SUR_CORE	lightgbm_cds_dt	18.34%
AREA SUR_CORE PLUS	ridge	20.41%
AREA SUR_PREMIUM	lasso	13.00%
AREA SUR_VALUE	grand_means	23.21%

Tabla 5. *Mejores modelos área-segmento univariados de ambos módulos de PyCaret.*

Tanto en la dimensión de calibre como de segmento de venta, se encuentra una mejor performance de los modelos univariados en áreas GBA y Sur, mientras que en Este y Oeste la predicción no resulta ser demasiado precisa, con valores de MAPE superando el 30% en la mayoría de los casos.

Esta fue una fuerte motivación para explorar modelos multivariados que pudieran de algún modo responder la pregunta de si la correlación entre distintos calibres y segmentos dentro de cada región geográfica es relevante a la hora de predecir el volumen de venta.

Serie	Modelo seleccionado	MAPE
AREA ESTE_BOTELLA RET	polytrend	39.73%
AREA ESTE_LATA	ridge	30.47%
AREA ESTE_OW s-LATA	polytrend	26.96%
AREA GBA_BOTELLA RET	croston	19.63%
AREA GBA_LATA	rf	16.17%
AREA GBA_OW s-LATA	croston	15.92%
AREA OESTE_BOTELLA RET	par_cds_dt	43.33%
AREA OESTE_LATA	ridge	23.91%
AREA OESTE_OW s-LATA	ridge	41.61%
AREA SUR_BOTELLA RET	lightgbm_cds_dt	22.03%
AREA SUR_LATA	dt_cds_dt	18.54%
AREA SUR_OW s-LATA	ada_cds_dt	19.93%

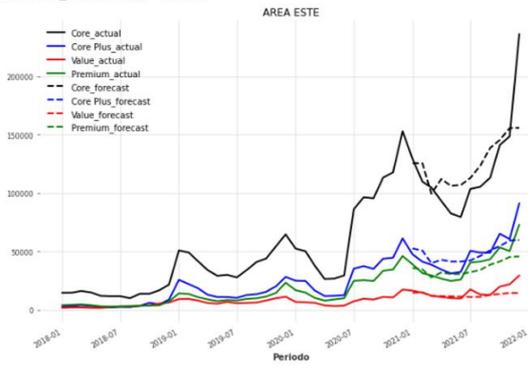
Tabla 6. Mejores modelos área-calibre univariados de ambos módulos de PyCaret.

4.2 Modelos multivariados

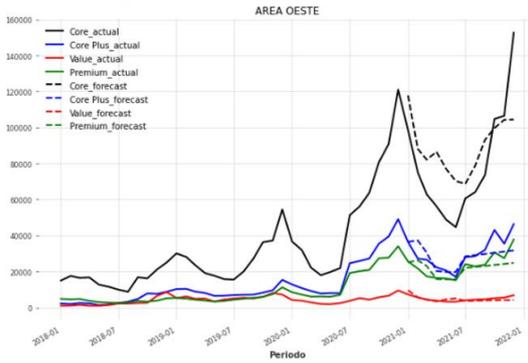
Cómo se explicó en la sección 3, se utilizó la librería *Darts* para poder entrenar los modelos VARMA, tomando cada área como unidad de análisis por separado. Para la elección de los parámetros p (orden o número de rezagos del modelo autorregresivo) y q (tamaño de la ventana de media móvil) se utilizó la técnica de *gridsearch* (Fuentes, 2018). Dada la cantidad de observaciones con las que contamos en la base de datos, se estableció un límite de 6, tanto para p como para q . Notar que, en términos estadísticos, el límite teórico en donde los modelos dejarían de funcionar sería tener una cantidad de parámetros a estimar superior a la cantidad de observaciones (Hyndman y Athanasopoulos, 2018).

Los resultados obtenidos por área fueron los siguientes:

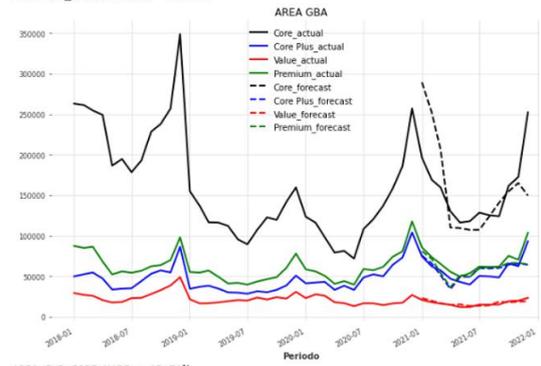
AREA ESTE_CORE MAPE = 16.36%
 AREA ESTE_CORE PLUS MAPE = 16.71%
 AREA ESTE_VALUE MAPE = 18.91%
 AREA ESTE_PREMIUM MAPE = 17.12%



AREA OESTE_CORE MAPE = 28.17%
 AREA OESTE_CORE PLUS MAPE = 13.80%
 AREA OESTE_VALUE MAPE = 22.33%
 AREA OESTE_PREMIUM MAPE = 12.50%



AREA GBA_CORE MAPE = 19.69%
 AREA GBA_CORE PLUS MAPE = 15.30%
 AREA GBA_VALUE MAPE = 10.96%
 AREA GBA_PREMIUM MAPE = 11.02%



AREA SUR_CORE MAPE = 13.71%
 AREA SUR_CORE PLUS MAPE = 20.71%
 AREA SUR_VALUE MAPE = 19.38%
 AREA SUR_PREMIUM MAPE = 12.94%

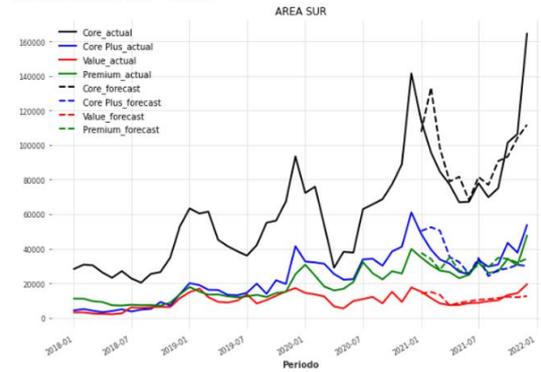
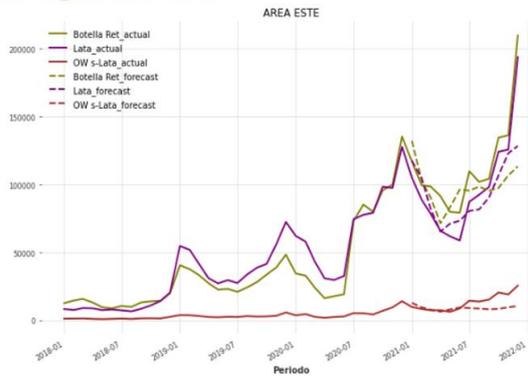
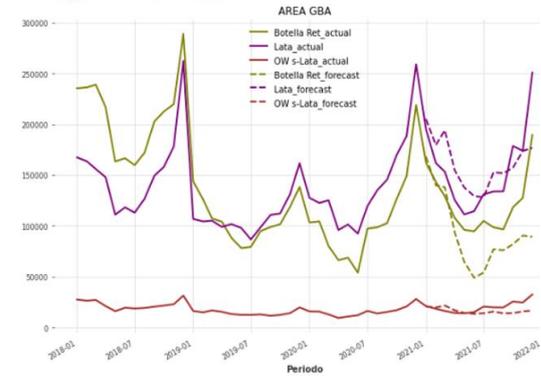


Figura 28. Output de modelos VARIMA área-segmento.

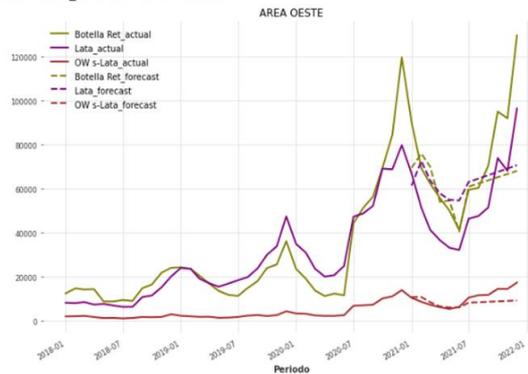
AREA ESTE_BOTELLA RET MAPE = 16.03%
 AREA ESTE_LATA MAPE = 12.68%
 AREA ESTE_OW s-LATA MAPE = 32.29%



AREA GBA_BOTELLA RET MAPE = 25.97%
 AREA GBA_LATA MAPE = 14.54%
 AREA GBA_OW s-LATA MAPE = 23.67%



AREA OESTE_BOTELLA RET MAPE = 15.00%
 AREA OESTE_LATA MAPE = 36.19%
 AREA OESTE_OW s-LATA MAPE = 22.08%



AREA SUR_BOTELLA RET MAPE = 14.80%
 AREA SUR_LATA MAPE = 21.15%
 AREA SUR_OW s-LATA MAPE = 18.33%

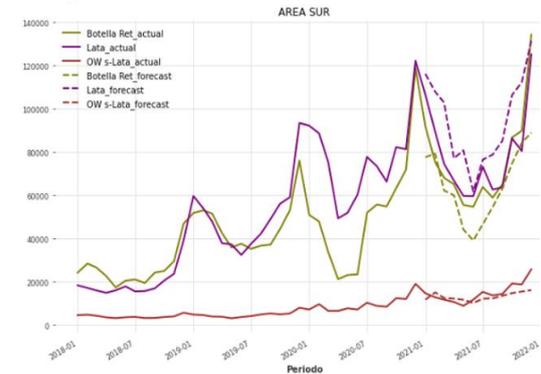


Figura 29. Output de modelos VARIMA área-calibre.

Comparando los resultados de los modelos VARIMA vs los modelos univariados, se observa una notable mejora en la performance de las predicciones a nivel general. Se logró explotar las interacciones que tiene el mercado entre los distintos segmentos y calibres que lo componen, tomando cada área geográfica como un mercado en sí mismo. Se lograron proyecciones robustas en el segmento *Core* de área Sur, en el *Core Plus* de área Oeste, en el *Premium* de áreas GBA, Oeste y Sur, y en el *Value* de área GBA. El MAPE de dichas series se situó debajo del 15%. Por el lado de las series de calibre, se obtuvo un desempeño similar en latas de áreas Este y GBA, sumado a botella retornable en área Sur.

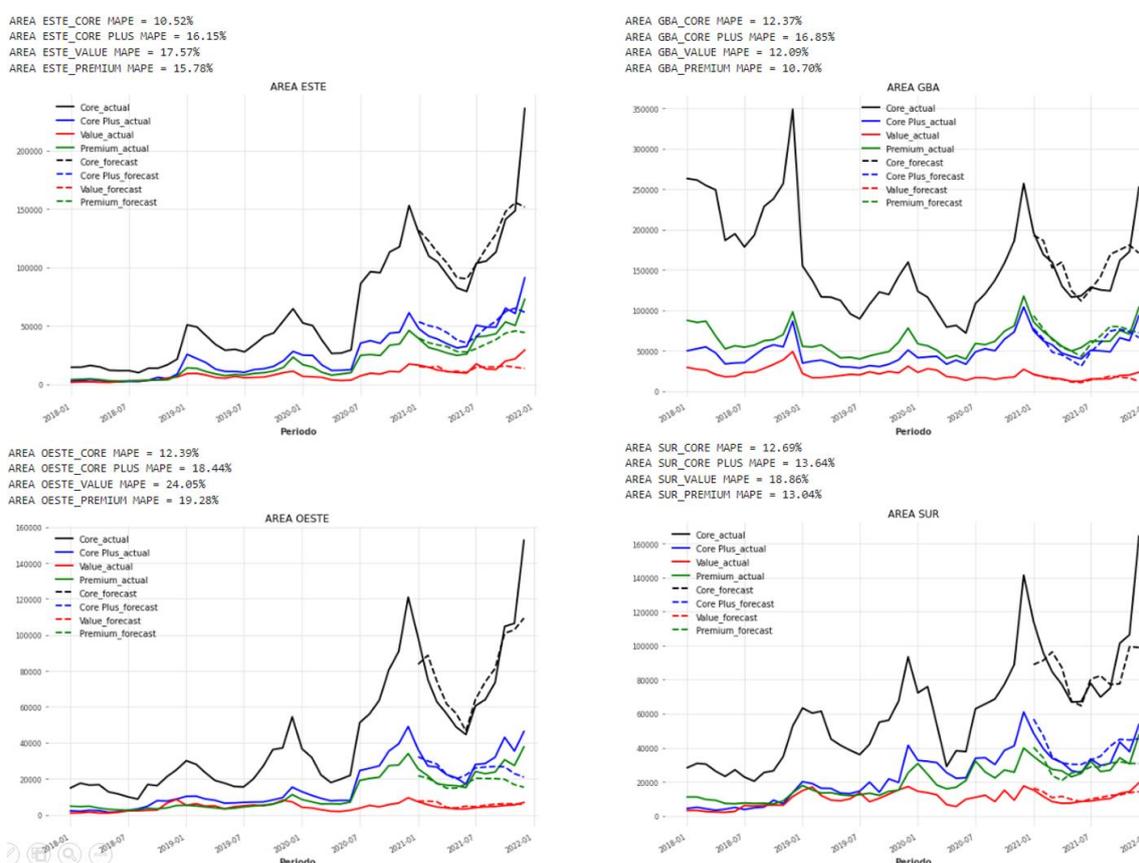


Figura 30. Output de modelos VARIMA con covariables de precios, área-segmento.

Finalmente, el modelo VARIMA con covariables de precio no solo logró aumentar aún más la precisión general, llevando a 16 series de las 28 totales (series área-segmento + series área-calibre) a valores de MAPE por debajo del 15%, sino que también logró que 27 de ellas tengan valores por debajo de 25%, hito que no se había logrado con ninguno de los modelos anteriores. Esto implica que introducir las relaciones de precio entre las distintas series de cada área es

relevante para lograr proyecciones más robustas, a la vez que reduce la volatilidad entre los errores que se obtuvieron entre todas las series, llevando el MAPE promedio al valor más pequeño posible.

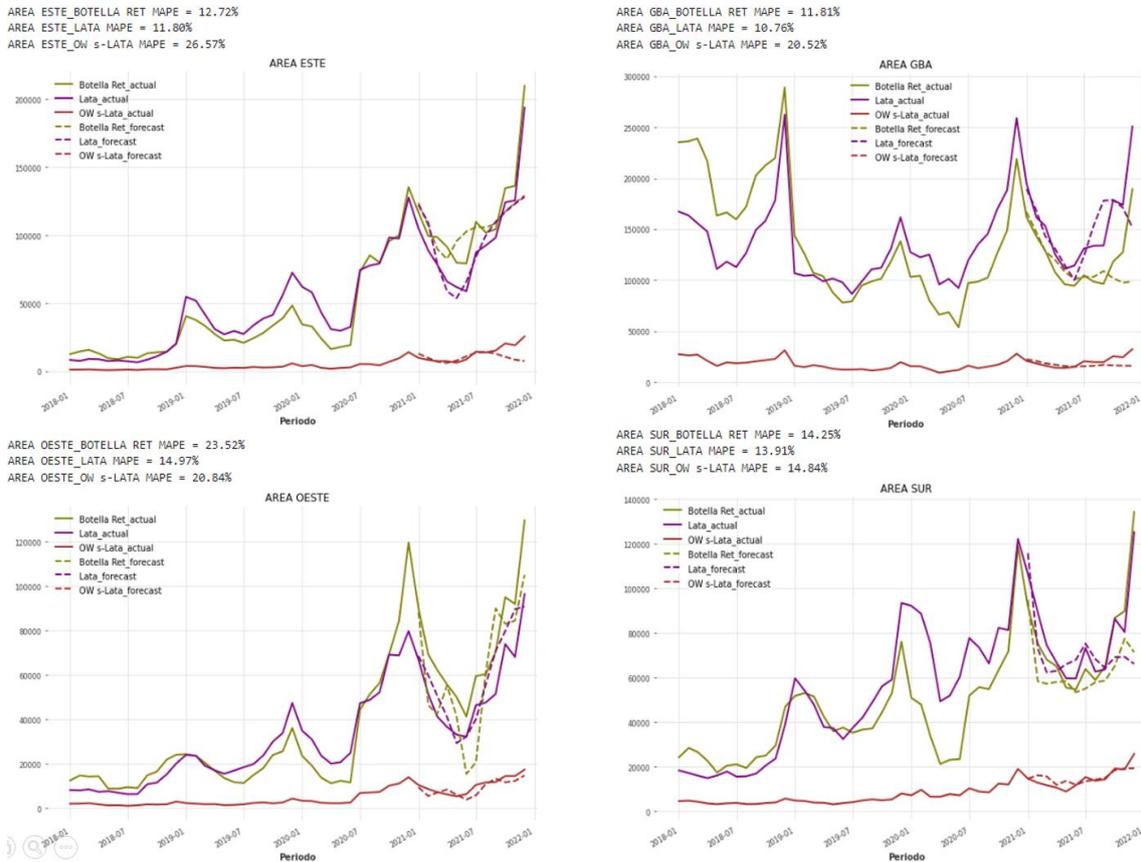


Figura 31. Output de modelos VARIMA con covariables de precios, área-calibre.

Computados los resultados de todos los modelos implementados, como último paso se realizó una comparación del MAPE obtenido bajo cada modelo, para cada una de las series entrenadas. Esto nos permite entender cuál fue el modelo de mayor precisión en cada caso, para luego seleccionarlo y realizar las proyecciones de la serie en cuestión.

Cómo se introdujo anteriormente a modo de hipótesis, no se encontró un modelo que sistemáticamente tenga un mejor desempeño que los demás en todas las series. Por ende, a los efectos de minimizar el MAPE, se construyeron las proyecciones finales tomando para su cómputo el mejor modelo para cada serie por separado. Dicho de otro modo, se realizó un ensamble de los modelos de mejor desempeño.

MAPE Benchmark						
Series/Model	Holt-Winters	PyCaret Best Model	VARIMA	VARIMA with Price Covariates	Best Model	MAPE Gain
AREA ESTE_CORE	31.0%	35.5%	16.4%	10.5%	10.5%	24.9%
AREA ESTE_CORE PLUS	23.0%	32.0%	16.7%	16.2%	16.2%	15.8%
AREA ESTE_PREMIUM	25.0%	32.5%	17.1%	15.8%	15.8%	16.8%
AREA ESTE_VALUE	19.0%	20.0%	18.9%	17.6%	17.6%	2.4%
AREA GBA_CORE	15.0%	20.2%	19.7%	12.4%	12.4%	7.8%
AREA GBA_CORE PLUS	27.0%	19.4%	15.3%	16.9%	15.3%	11.7%
AREA GBA_PREMIUM	39.0%	16.7%	11.0%	10.7%	10.7%	28.3%
AREA GBA_VALUE	22.0%	20.1%	11.0%	12.1%	11.0%	11.0%
AREA OESTE_CORE	31.0%	38.6%	28.2%	12.4%	12.4%	26.2%
AREA OESTE_CORE PLUS	78.0%	36.3%	13.8%	18.4%	13.8%	64.2%
AREA OESTE_PREMIUM	86.0%	39.5%	12.5%	19.3%	12.5%	73.5%
AREA OESTE_VALUE	48.0%	26.7%	22.3%	24.1%	22.3%	25.7%
AREA SUR_CORE	18.0%	18.3%	13.7%	12.7%	12.7%	5.7%
AREA SUR_CORE PLUS	73.0%	20.4%	20.7%	13.6%	13.6%	59.4%
AREA SUR_PREMIUM	29.0%	13.0%	12.9%	13.0%	12.9%	16.1%
AREA SUR_VALUE	21.0%	23.2%	19.4%	18.9%	18.9%	4.4%
AREA ESTE_Botella RET	25.0%	39.7%	16.0%	12.7%	12.7%	27.0%
AREA ESTE_LATA	19.0%	30.5%	12.7%	11.8%	11.8%	18.7%
AREA ESTE_OW s-LATA	18.0%	27.0%	32.3%	26.6%	18.0%	14.3%
AREA GBA_Botella RET	28.0%	19.6%	26.0%	11.8%	11.8%	16.2%
AREA GBA_LATA	10.0%	16.2%	14.5%	10.8%	10.0%	6.2%
AREA GBA_OW s-LATA	22.0%	15.9%	23.7%	20.5%	15.9%	7.8%
AREA OESTE_Botella RET	83.0%	43.3%	15.0%	23.5%	15.0%	68.0%
AREA OESTE_LATA	70.0%	23.9%	36.2%	15.0%	15.0%	55.0%
AREA OESTE_OW s-LATA	64.0%	41.6%	22.1%	20.8%	20.8%	43.2%
AREA SUR_Botella RET	12.0%	22.0%	14.8%	14.3%	12.0%	10.0%
AREA SUR_LATA	37.0%	18.5%	21.2%	13.9%	13.9%	23.1%
AREA SUR_OW s-LATA	51.0%	19.9%	18.3%	14.8%	14.8%	36.2%
Average	36.6%	26.1%	18.7%	15.7%	14.3%	25.7%

Holt-Winters			Best Model		
MAPE	# Series	# Series (%)	MAPE	# Series	# Series (%)
<15%	2	7%	<15%	18	64%
<20%	7	25%	<20%	26	93%
<25%	11	39%	<25%	28	100%
≥25%	17	61%	≥25%	0	0%

Figura 32. Resultados de MAPE obtenidos en los modelos entrenados y selección de mejor modelo.

El ensamble de los mejores modelos se detalla en la figura 32, donde se compara el MAPE obtenido para los distintos modelos entrenados y se selecciona como *Best Model* aquel modelo con menor MAPE en cada caso (pintado de verde).

Es interesante observar como el error promedio entre todas las series va disminuyendo a medida que agregamos mayor cantidad de variables explicativas relevantes al entrenamiento de los modelos. No obstante, el modelo VARIMA con covariables de precio, que es el que mayor cantidad de información incorpora, no resulta ser el de mejor desempeño en todos los casos, aunque se observa que lo es en 17 de las 28 series. Si existiera la restricción de tener que realizar

las proyecciones bajo un mismo modelo en todos los casos, la mejor decisión sería tomar dicho modelo. Al tener la posibilidad de ensamblar los mejores modelos en cada caso, se logra reducir el MAPE promedio de 15,7% a 14,3%.

Tomando el suavizado exponencial de Holt-Winters como el modelo base, se parte de un resultado donde tan solo 2 series obtuvieron un MAPE menor a 15% y 17 de ellas arrojaron un valor superior al 25%. Luego del ensamble de los mejores modelos, se logró que el 64% de las series obtengan un MAPE menor a 15%, un 93% de ellas queden por debajo de 20% y el 100% se sitúe debajo de 25%.

La ganancia de MAPE o *MAPE Gain* indica en cuantos puntos porcentuales se redujo el error en el modelo de mejor desempeño, partiendo desde el modelo de peor desempeño, el cual no necesariamente resulta ser el suavizado exponencial. De hecho, hubo 3 casos en donde el desempeño del modelo de Holt-Winters no pudo ser superado por modelos más sofisticados.

Una vez hecho el ensamble de los mejores modelos, se procedió a calcular las proyecciones finales mensuales de volumen para el año 2022, que se muestran en las figuras 33 y 34.

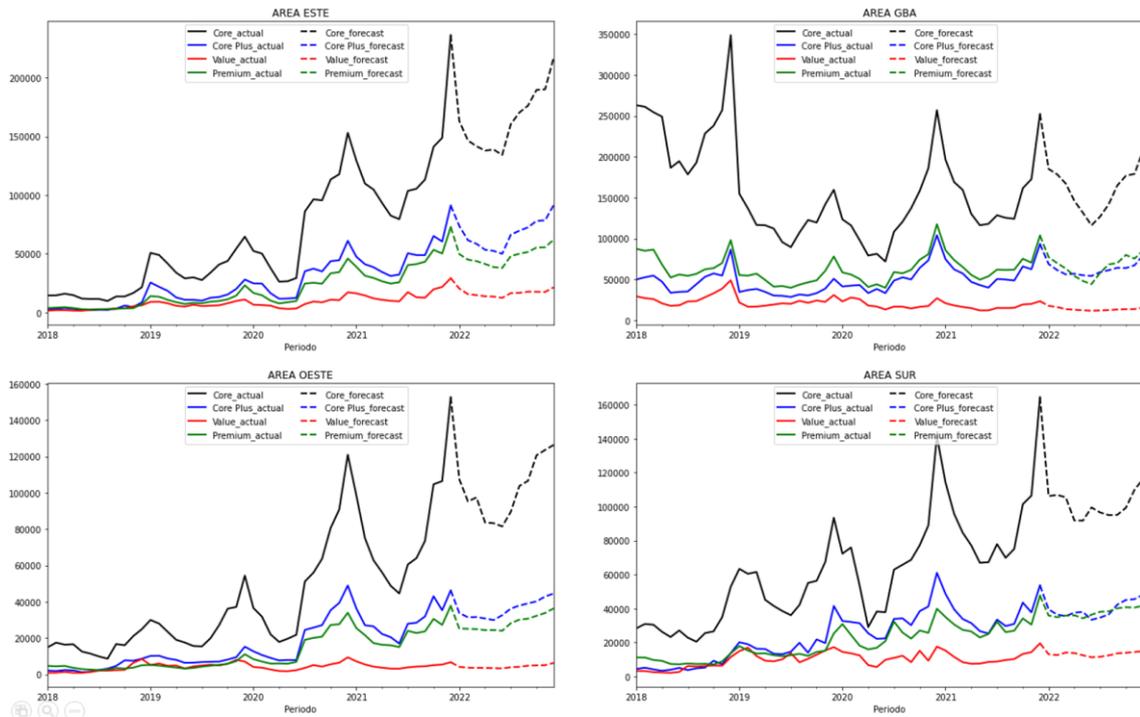


Figura 33. Proyecciones finales, según mejores modelos área-segmento.

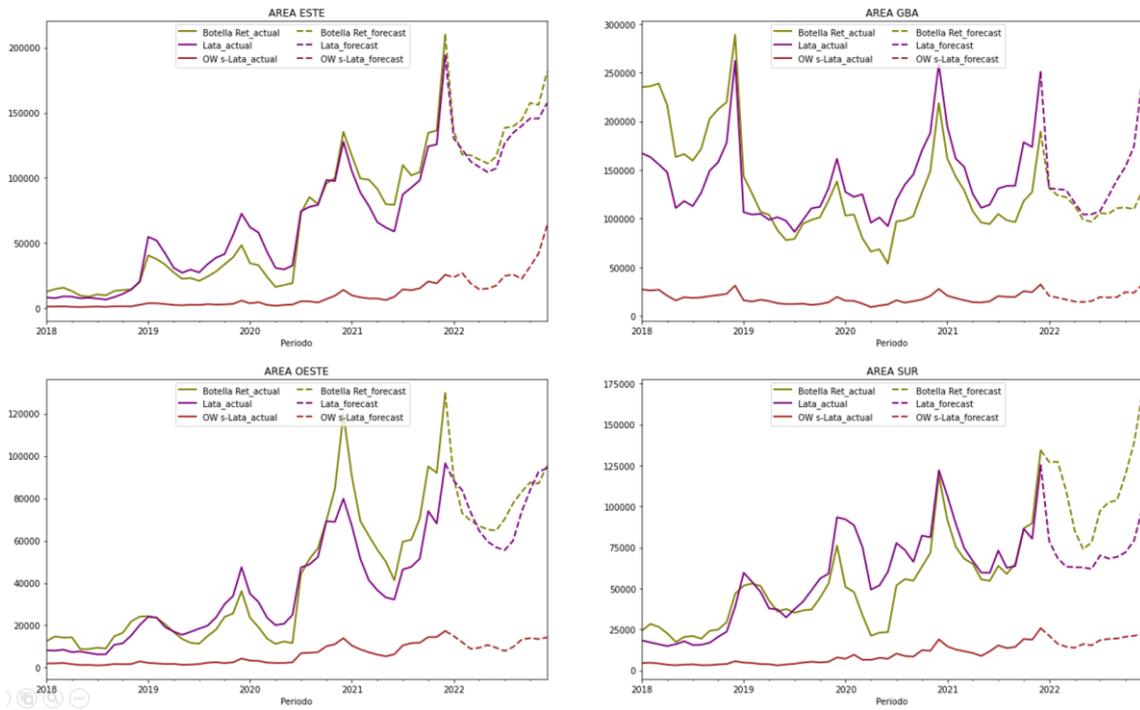


Figura 34. Proyecciones finales, según mejores modelos área-calibre.

5. Mapa competitivo

Con las proyecciones de volumen para el año 2022, calculadas a partir del ensamble de los mejores modelos para cada serie, se construyó el mapa competitivo proyectado para dicho período.

En la proyección del mercado, se tomaron dos supuestos no triviales, que vale la pena destacar:

- Se asumió una distribución de la participación de mercado idéntica a 2021 para cada combinación segmento-calibre, entre ABI y la competencia. Hay que tener en cuenta que el mapa competitivo real, hacia atrás, construye estas combinaciones a partir de la venta de cada producto existente en el mercado. Como los modelos entrenados a lo largo de este trabajo proyectan el mercado a nivel área-segmento y área-calibre, se sacrifica la partición del volumen entre los jugadores del mercado en pos de una mayor simplicidad. Notar que, para poder incorporar la dimensión compañía o jugador del mercado en el análisis, las series a proyectar se incrementarían al doble de las tratadas en este trabajo, siempre que haya disposición a tratar a la competencia de CMQ como un todo.
- Para la construcción del “peso del segmento & calibre” proyectado, se asumió que el volumen proyectado para cada calibre en cada área se distribuye entre los distintos segmentos conforme a dicha distribución en el año 2021. Si bien es cierto que se proyectaron las series área-segmento y área-calibre para cada zona, la suma de volumen no es equivalente en ambos casos, lo cual se explica en el hecho de que ambos modelos se ejecutaron de forma independiente. Nuevamente, el correr los modelos utilizando series tridimensionales de la forma área-segmento-calibre aumentaría la complejidad del problema, posiblemente afectando la flexibilidad de los modelos multivariados en cuanto a cantidad de rezagos que se puedan tomar, teniendo en cuenta la acotada longitud temporal de los datos disponibles.

En la proyección del mapa competitivo se incluyó la variación absoluta en el índice de tamaño de oportunidad, respecto a 2021. De esta forma, se ofrece visibilidad no solo de aquellas combinaciones segmento-calibre cuyas predicciones las sitúan como prioritarias sino que, también, se puede determinar cuáles son aquellas que proyectan un mayor movimiento, sea de crecimiento o declive.

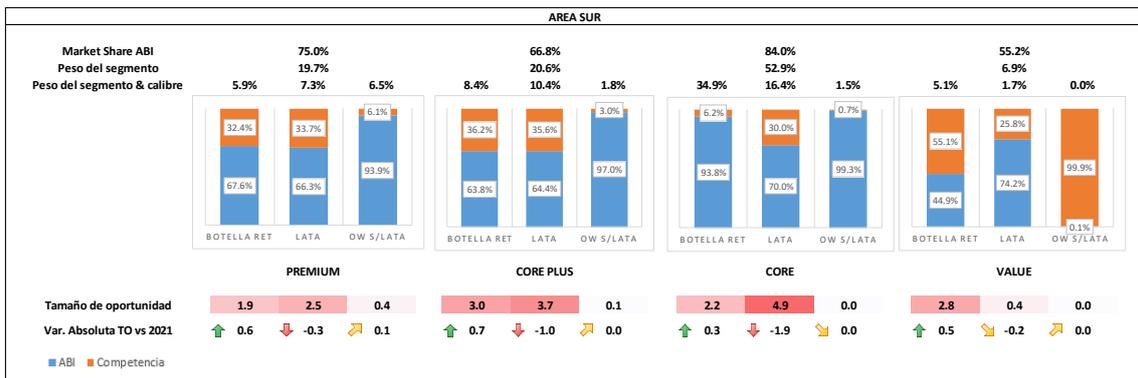
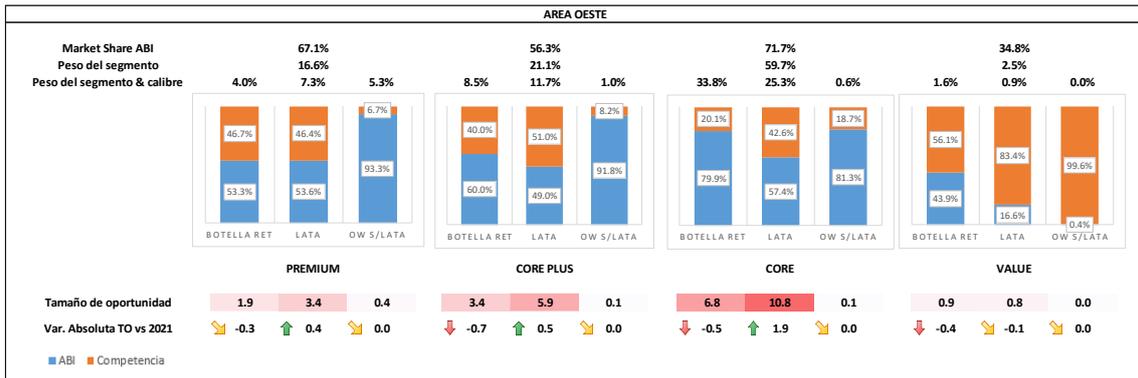
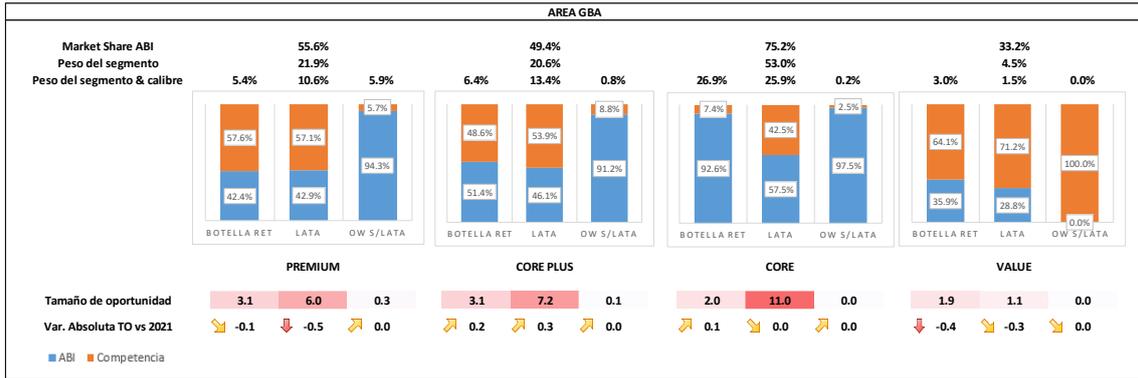
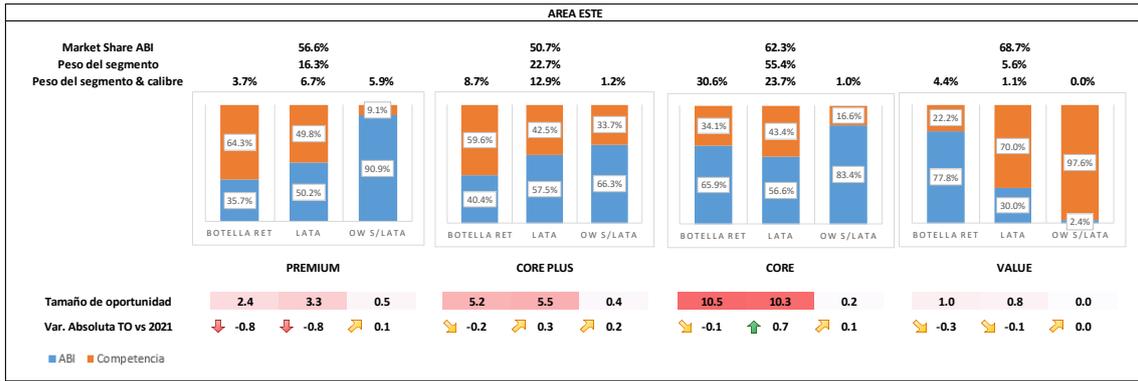


Figura 35. Proyección del mapa competitivo por área geográfica de Argentina, correspondiente a 2022.

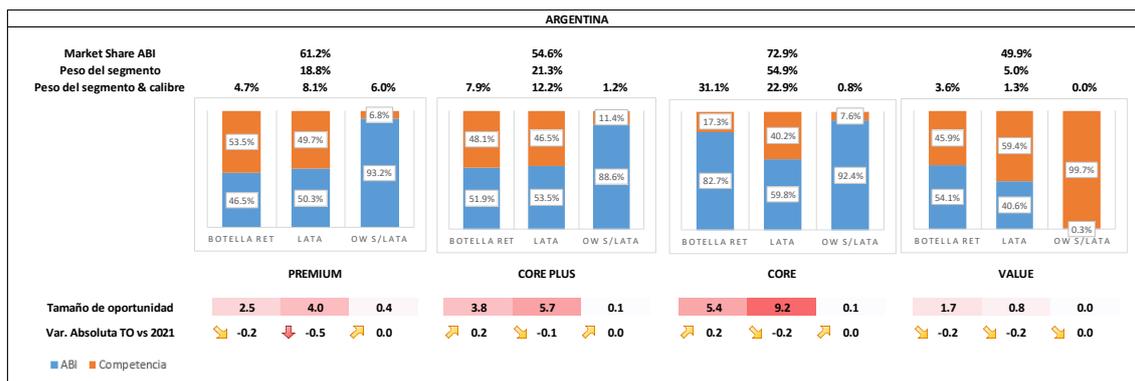


Figura 36. Proyección del mapa competitivo para Argentina, correspondiente a 2022.

A nivel nacional, lo que se observa en primera instancia es que el calibre lata, dentro del segmento *Core*, continuará teniendo un rol protagónico y supondrá la mayor oportunidad de capturar participación de mercado de los competidores de CMQ, al igual que en 2021. Sin embargo, su TO disminuye en 0,2, lo cual es compensado por una suba de igual magnitud en la botella retornable, dentro del mismo segmento. Este último ocupa el tercer lugar en las prioridades a atacar según el TO, a pesar de que la participación de mercado de ABI es superior al 80% en la combinación en cuestión. Esto tiene que ver con que, según nuestras proyecciones, seguirá siendo la más volumétrica del mercado argentino, con un peso del 31,1%.

El podio de combinaciones prioritarias lo completa la lata dentro del segmento *Core Plus*. Se proyecta una gran oportunidad para ABI en marcas como Budweiser y Andes Origen, de poder capturar *market share* de la competencia, que tiene a Imperial como marca insignia dentro del segmento.

Notar que, a excepción del *Value*, los pronósticos indican que las mayores oportunidades continuarán encontrándose en la lata. Esto es consistente con lo explicado en el apartado de análisis exploratorio, sobre el crecimiento de este calibre y sus ventajas frente a otros formatos. Dicha tendencia no es propia de algún segmento en particular.

En cuanto a movimientos en el tamaño de oportunidad respecto a 2021, los mismos son relativamente suaves, a excepción de la lata *Premium*, que cae medio punto en el índice de TO. Esto es debido a que, según las proyecciones, esta combinación perdería un punto porcentual sobre el total mercado durante 2022.

Al analizar el mapa competitivo a nivel área, pueden encontrarse distintos *insights* en cada caso.

En área Este, las proyecciones indican que la mayor oportunidad de captura de participación de mercado estará en la botella retornable *Core*, al igual que en 2021. Sin embargo, se pronostica un crecimiento de 0,7 en el TO de la lata, dentro del mismo segmento. Este movimiento dejaría los índices de ambas combinaciones prácticamente empatados, con lo cual el foco de inversión y desarrollo del mercado en esta región del país debería estar puesto en ambos calibres. Este *switching* también es pronosticado en el segmento *Core Plus*, a tal punto que la lata desplaza a la botella retornable como tercera combinación en términos de TO. Al igual que en el *Core*, los índices quedan emparejados. Esto indica que la prioridad de una combinación sobre otra es más débil que la que se puede observar a nivel nacional. Por otro lado, disminuye la oportunidad en el segmento *Premium*, donde se proyectan retrocesos de 0,8 en el índice TO, tanto en botella retornable como en la lata.

Con respecto a área GBA, se observa que la primera prioridad se encontrará claramente en la lata *Core*, según las proyecciones. Un dato destacable es que esta es la combinación que mayor TO obtuvo entre todas las combinaciones segmento-calibre de todas las áreas, con un valor de 11,0. Por lo tanto, a la hora de planificar una estrategia de captura de crecimiento en *market share* a nivel nacional, sin duda se debería poner especial atención en destinar la mayor cantidad de recursos a continuar desarrollando la lata *Core* en GBA. El podio de combinaciones prioritarias lo completan las latas *Core Plus* y *Premium*, a pesar de que esta última proyecta un retroceso en su TO de 0,5. La prevalencia de la lata dentro de las combinaciones con mayor oportunidad sigue la línea del mapa competitivo a nivel nacional.

Continuando con área Oeste, se observa nuevamente que mayor oportunidad se aloja en la lata *Core*, que crece 1,9 en el índice TO. Dicho movimiento constituye el mayor crecimiento proyectado para una combinación, tomando en cuenta la totalidad de las áreas geográficas. A su vez, se proyectan crecimientos en las latas *Core Plus* y *Premium*. Sin embargo, la botella retornable *Core* se sitúa como la segunda prioridad a atacar para CMQ. A pesar de que se proyecta un descenso en su TO, el peso de dicha combinación dentro del área la convierte en un blanco del cual será difícil correr el foco, ya que se proyecta que su volumen de venta será la tercera parte del total mercado en la región. A su vez, también se pronostica un debilitamiento en las oportunidades de las botellas retornables *Value*, *Premium* y *Core Plus*.

Finalmente, en área Sur se observa un comportamiento contrario respecto al de las demás zonas del país, si se considera la variación en el índice TO de las distintas combinaciones. La botella retornable proyecta crecimiento en todos los segmentos del mercado, mientras que la lata muestra un pronóstico opuesto. La lata *Core*, a pesar de mostrar una variación de -0,9 en su TO (la mayor caída proyectada para una combinación, tomando en cuenta la totalidad de las áreas), continuaría siendo la primera prioridad a atacar, seguida por lata *Core Plus* y litro retornable *Core Plus*, que desplazaría del tercer lugar a la lata *Premium*. Por otro lado, vale la pena nombrar como cuarta prioridad a la botella retornable *Value*. Este último es el segmento más pequeño del mercado en todas las áreas, pero por el *market share* de la competencia, combinado con el peso proyectado de esta combinación, se podría accionar tácticamente en área Sur, aprovechando por ejemplo el camino recorrido en los últimos años en área Este. Allí, de la mano de marcas como Bajo Cero y 1890, ABI logra dominar actualmente el mercado, con una participación de 77,8% en 2021.

Un punto importante a destacar es que todas estas recomendaciones que podemos extraer a partir del análisis del TO refieren exclusivamente a la priorización de la asignación de recursos adicionales o incrementales a aquellos que se encuentran alocados actualmente en cada caso. La intención es guiar la toma de decisiones referidas a dónde se debe fortalecer la inversión para crecer en participación de mercado, partiendo de la base que ya se tiene al momento del análisis. Dicho de otra forma, la recomendación es de carácter marginal. No engloba la asignación de la totalidad de los recursos disponibles, incluyendo aquellos que contribuyeron a otorgar la posición que CMQ ostenta hoy en día en cada combinación segmento & calibre. Para tomar decisiones en ese sentido, debería incluirse en el análisis algún factor que considere la pérdida potencial de *market share* que la compañía podría sufrir si se asignara la totalidad de los recursos en base al criterio del TO, en aquellas combinaciones que no resulten favorecidas en el ranking. Es decir, debería contemplarse de algún modo el *trade-off* entre el aumento de participación que se lograría por aumentar la cantidad de recursos en las combinaciones prioritarias y la disminución de *market share* que se observaría en aquellas combinaciones relegadas, debido a la quita de inversiones. Este es un tipo de análisis que no se contempla en este trabajo. Se asume como punto de partida la preservación de la participación de mercado actual, y se plantea el desafío de cómo optimizar la asignación de recursos adicionales o incrementales, cuando el objetivo primordial sea capturar el *market share* que está en poder de la competencia. En este sentido, el TO es una herramienta de gran utilidad.

6. Conclusiones

A lo largo de este trabajo, se ha buscado realizar una proyección de la industria cervecera argentina de cara al año 2022, junto con sus dinámicas competitivas. Se comenzó analizando la importancia e implicancias que tiene contar con un *forecast* robusto para los principales jugadores del mercado, particularmente CMQ. Se entendió que las proyecciones de cada compañía no debieran considerarse independientes del volumen de los competidores. De aquí es que surgió la necesidad de contar con un modelo de industria que permita anticipar futuros movimientos respecto al volumen del mercado y su composición. Analizando los datos disponibles, se comprendió que existen dos dimensiones competitivas principales en la industria; segmentos y calibres. Además, se detectó que su comportamiento es heterogéneo según la zona geográfica del país en donde nos encontremos. Teniendo eso en cuenta, esta tesis estuvo focalizada en trabajar al nivel de estas dimensiones principales, analizando cada área de Argentina como un mercado independiente.

En primera instancia, se entrenó el modelo de Holt-Winters a modo de *baseline* en cada una de las series temporales, obteniendo una amplia variedad de valores de MAPE. Esto resultó en un indicio de que sería complejo encontrar un único modelo que se adapte de la mejor manera para predecir todas las series temporales que componen el mercado. La situación era consistente con el hecho de que la evolución de la venta no se comportó de igual manera en el pasado para cada combinación área-segmento o área-calibre. De aquí surgió la motivación para decidir entrenar varios modelos en cada una de las series, tanto univariados como multivariados. Usando el MAPE como criterio de comparación, se eligió aquel modelo de mejor desempeño en cada caso. Por lo tanto, las proyecciones finales estuvieron dadas por el ensamble de los mejores modelos hallados para cada serie.

Una vez obtenidos los pronósticos, se construyeron los mapas competitivos proyectados para 2022, tanto a nivel área como nacional. En cada caso, utilizando el índice de Tamaño de Oportunidad o TO, se arribó a las recomendaciones referidas a qué combinaciones segmento-calibre se deberían priorizar en cuanto a inversión, desarrollo y asignación de recursos incrementales, en pos de capturar de la forma más eficiente, el *market share* que ostenta la competencia.

Rank	Área Este	Área GBA	Área Oeste	Área Sur
1	botella ret core	lata core	lata core	lata core
2	lata core	lata core plus	botella ret core	lata core plus
3	lata core plus	lata premium	lata core plus	botella ret core plus
4	botella ret core plus	botella ret premium	botella ret core plus	botella ret value
5	lata premium	botella ret core plus	lata premium	lata premium
6	botella ret premium	botella ret core	botella ret premium	botella ret core
7	botella ret value	botella ret value	botella ret value	botella ret premium
8	lata value	lata value	lata value	lata value
9	ow s/lata premium	ow s/lata premium	ow s/lata premium	ow s/lata premium
10	ow s/lata core plus	ow s/lata core plus	ow s/lata core	ow s/lata core plus
11	ow s/lata core	ow s/lata value	ow s/lata core plus	ow s/lata value
12	ow s/lata value	ow s/lata core	ow s/lata value	ow s/lata core

Tabla 7. Ranking de combinaciones prioritarias según índice TO a nivel área; Proyección 2022.

Notar que, para todas las zonas de país, el top 3 de combinaciones prioritarias está compuesto por al menos 2 combinaciones de calibre lata. Esto está relacionado con el desarrollo que tuvo dicho empaque en los últimos años, siendo CCU la empresa que pudo visualizar en primera instancia su potencial, tal cómo se explicó en el apartado de análisis exploratorio.

Con respecto a los próximos pasos y trabajo futuro, existen posibilidades de refinar el análisis y la calidad de las proyecciones. Por un lado, al contar con información de volumen de venta a nivel producto (pudiendo identificar a qué compañía pertenece dicho ítem), se puede estudiar la posibilidad de entrenar modelos más sofisticados con este nivel de granularidad. Notar que esto aumenta considerablemente la complejidad del problema y es posible que varios de los modelos entrenados en este trabajo no sean extrapolables a este fin, por la cantidad de parámetros que se tendrían que estimar. Principalmente, los modelos multivariados tendrían dificultades a la hora de trabajar con un número creciente de variables mientras la longitud temporal de las series siga siendo acotada. Recordar que, por disponibilidad de los datos, no se cuenta con información anterior a enero 2018. Esta situación no es favorable dada la importancia que tuvo la correlación entre los distintos segmentos y calibres dentro de cada área para lograr un menor error en las proyecciones de los modelos entrenados, como se evidencia en la figura 32. Aun así, es importante destacar el hecho de que poder lograr la máxima

granularidad en las proyecciones permitiría relajar los dos supuestos que se tomaron a la hora de construir los mapas competitivos.

Por otro lado, de igual manera que los precios de los distintos segmentos y calibres, así como sus interrelaciones, otorgaron la posibilidad de mejorar el desempeño de los modelos en varios casos, es natural pensar en que deben existir otras variables externas que tengan poder explicativo sobre el volumen de venta de cerveza en Argentina. Se podría plantear un set de variables referidas al contexto macroeconómico del país, tales como el Estimador Mensual de Actividad Económica (EMAE), el índice de precios al consumidor (IPC) o alguna variante del salario real (Índice de Variación de Salarios o Remuneración Imponible Promedio de los Trabajadores Estables) que puedan ayudar a explicar las fluctuaciones en el mercado, así como también tomar indicadores de precios de bienes sustitutos a la cerveza dentro de la industria de bebidas alcohólicas, como puede ser el vino.

Además, resultaría muy beneficiosa la colaboración que pueda existir entre *Demand Planning* y equipos como Márketing o Innovaciones, que puedan contribuir a desarrollar un análisis de carácter más cualitativo respecto a tendencias incipientes en el mercado. Así como CCU logró visualizar una oportunidad para desarrollar el calibre lata de forma prematura, explorar si existen combinaciones que tengan un gran potencial de crecimiento debido a conveniencia de empaques, cambios en los hábitos de consumo o algún otro factor *soft* puede enriquecer en gran medida la planificación estratégica. Este tipo de estudio debería ser complementario al expuesto en este trabajo, donde se utiliza esencialmente información histórica para realizar pronósticos futuros.

Por último, sería interesante llevar el análisis al campo de la monetización del *market share*. Como se ha visto anteriormente, existen diferencias entre los precios de los distintos segmentos y calibres que componen el mercado. Esto implica que, en términos de facturación, no resultaría equivalente capturar participación en cualquier combinación posible. Una decisión óptima desde el punto de vista del TO, índice construido en base al volumen, puede no serlo en términos monetarios. Por ello, incorporar esta dimensión de estudio permitiría enriquecer aún más las conclusiones.

Referencias Bibliográficas

- Bertrand, J. L. & Parnaudeau, M. (2015). *The Impact of Climate Variability on the Private Sector*.
- Cabras, I., Higgins, D. & Preece, D. (2016). *Brewing, Beer and Pubs: A Global Perspective (Ch. 8)*.
- Compte, J. M. (2020). *La chilena CCU invertirá \$ 4500 millones para duplicar su producción de cerveza en Luján*.
- Darts. <https://unit8co.github.io/darts>.
- Fuentes, A. (2018). *Mastering Predictive Analytics with scikit-learn and TensorFlow: Implement machine learning techniques to build advanced predictive models using Python (Ch. 2)*.
- Hyndman, R. & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice (2nd ed)*.
- Lazzeri, F. (2020). *Machine Learning for Time Series Forecasting with Python (Ch. 4)*.
- Lopez, M. A. (2020). *Estudios: Cerveza sin alcohol: la nueva tendencia saludable de los argentinos*.
- Morlidge, S. (2015). *How Good Is a "Good" Forecast?: Forecast Errors and Their Avoidability*.
- O'Neil, C. & Schutt, R. (2013). *Doing Data Science: Straight Talk from the Frontline (Ch. 11)*.
- PyCaret. <https://pycaret.org/>.
- Radečić, D. (2021). *Time Series From Scratch — Train/Test Splits and Evaluation Metrics*.

- Tashman, L., Sglavo, U. & Gilliland, M. (2015). *Business Forecasting: Practical Problems and Solutions*.
- Trull, O., García Díaz, J. & Troncoso, A. (2020). *Initialization Methods for Multiple Seasonal Holt–Winters Forecasting Models*.
- Tsihrintzis, G., Virvou, M & Lakhmi, J. (2021). *Advances in Machine Learning/Deep Learning-based Technologies (Ch. 8)*.