



Universidad del Desarrollo
Facultad de Ingeniería

FORECAST DE VENTAS EN ESCENARIO DE PANDEMIA Y RETIROS PREVISIONALES

Caso aplicado a ventas de importante empresa del rubro automotriz

POR: MATÍAS FELIPE BULL RUIZ

Capstone project presentado a la Facultad de Ingeniería de la Universidad del Desarrollo para optar al grado académico de Magíster en Data Science

PROFESORES GUÍA:

Dr. Cristian Candia-Castro Vallejos

Sr. Alonso Astroza Tagle

Diciembre 2022

SANTIAGO, CHILE

TABLA DE CONTENIDO

RESUMEN	1
1. INTRODUCCIÓN	2
2. TRABAJOS RELACIONADOS	3
3. HIPÓTESIS Y OBJETIVOS	4
3.1. HIPÓTESIS	4
3.2. OBJETIVOS ESPECÍFICOS	4
4. DATOS Y METODOLOGÍA	4
4.1. DATOS	4
4.2. METODOLOGÍA.....	4
5. RESULTADOS	7
5.1. ANTES DE PANDEMIA	11
5.2. ENCIERRO	12
5.3. RETIROS	13
5.4. REGRESIÓN LINEAL UNIVARIADA	15
5.5. ESTACIONARIEDAD.....	15
5.6. MODELO ARIMA	16
5.7. FORECAST – ANTES DE PANDEMIA	17
5.8. FORECAST – ENCIERRO	19
5.9. FORECAST – RETIROS	20
5.10. REGRESIÓN LINEAL MULTIVARIADA.....	23
5.11. MODELO GRADIENT BOOSTING REGRESSOR.....	24
5.12. ELECCIÓN DEL MEJOR MODELO.....	26
6. CONCLUSIONES	27
BIBLIOGRAFÍA	29

Resumen

El siguiente trabajo es un intento por entender el nivel de impacto que tuvo la pandemia COVID-19 junto al posterior retiro de ahorros previsionales en las ventas de una importante empresa del rubro automotriz hasta marzo de 2022. A pesar de contar con información desde 2013 en adelante se trabajará con datos de enero de 2018 en adelante, utilizando como punto de quiebre el mes de marzo de 2020 donde se declara la pandemia por parte de la Organización Mundial de la Salud y los países comienzan a restringir el movimiento de las personas.

Este trabajo se divide en dos partes. En la primera parte de Datos y Metodología se describe los datos y forma de trabajo y en una segunda parte, más extensa de Resultados, se procede a detallar los modelos de Serie de tiempo explorados, como también dos alternativas más para medir el impacto que no resultan satisfactorias porque sus resultados no tienen un contexto de negocio que las acompañe. A pesar de eso es posible determinar un modelo que permite predecir las ventas sin pandemia COVID-19 y luego medir el impacto de los Retiros, suponiendo que la pandemia COVID-19 no hubiese existido. Finalmente es posible estimar un impacto en ventas que cae abruptamente durante la época de encierro, entre marzo y julio de 2020, con una estimación de casi 26 mil millones mensuales en pérdidas respecto a este escenario predicho de ventas sin COVID-19. Luego se observa un freno en la pérdida debido a la disponibilidad de liquidez por retiro de ahorros previsionales hasta que el liberar un tercer retiro comienza la recuperación logrando quedar sin pérdidas por menor ventas en febrero de 2022.

Se puede estimar que la pandemia COVID-19 tuvo un impacto negativo directo en las ventas de la empresa del rubro automotriz y los retiros tuvieron un impacto positivo que solo se vio claramente al llegar el tercer retiro, llevando a recuperar las ventas de Importante empresa del rubro automotriz en febrero de 2022.

1. Introducción

A inicio del 2020, se vió el surgimiento de una variante del virus respiratorio sincial que produjo un contagio rápido y mortal. La prensa reportaba como rápidamente iban creciendo los casos de contagio por sars-cov-2 en China y luego en distintas partes del mundo. Rápidamente se empezó a hablar de lo complejo de la situación y la palabra Pandemia comenzó a tener más importancia en el día a día, palabra que en la generación post 1960, estaba más presente en los libros que en el día a día. Los contagios y complicaciones respiratorias eran tales que fue necesario cerrar fronteras de países, colegios, empresas. El contagio a nivel mundial obligó a cerrar empresas para evitar el contagio entre sus trabajadores, el mundo enteró tuvo un giro abrupto a la tele presencialidad.

El temor por salir y encontrarse con alguien en el pasillo del edificio, sin saber si estaba contagiado o no; evitar a las personas en la calle, en el bus, en el metro; los aviones completamente vacíos. Algunos dejaron las grandes ciudades para vivir en lugares con menor densidad de población, otros tan solo se quedaban en sus casas y dejaban que solo uno de los integrantes saliera para hacer las compras o buscar medicamentos. Quienes pudieron, gracias a su poder adquisitivo, adquirieron autos para desplazarse alejados del resto de las personas. Le llamaron distancia social en un momento, pero era distancia física.

Las empresas que dependían directamente de la venta de vehículos comenzaron a ver rápidamente los efectos del cierre. La necesidad de algunas personas por evitar el contacto físico llevó a que aumentara la demanda por vehículos; ante la escasez de oferta de vehículos nuevos, subieron los precios; La demanda de vehículos fue tal que los precios de autos usados también vieron un aumento en sus precios.

Uno de los efectos directos del virus sars-cov-2 fue la restricción de movilidad, a nivel nacional esto derivó en el cierre de lugares de trabajo y con ello la pérdida de puestos trabajo que en consecuencia mermaron los ingresos de las personas, llegando a afectar sus ahorros y su capacidad de compra. Si bien en un inicio, hubo ciertas medidas que buscaron minimizar estos efectos, ninguno fue tan llamativo como la propuesta de retiro del 10% de los fondos destinados a las pensiones. En total fueron 3 retiros, cada uno con sus condiciones especiales de retiro, produciendo un aumento inmediato de liquidez en las personas.

A nivel mundial, el virus afectó desde la creación de chips hasta las importaciones y exportaciones, dos procesos que afectan directamente al mercado automotriz y que a partir de la escasez de chips redujo considerablemente la producción y el stock de vehículos a nivel mundial, situación que trajo consigo el aumento de los precios de vehículos nuevos y usados. A pesar de esto, según datos de la Asociación Nacional Automotriz de Chile, ANAC, entre enero y diciembre de 2021 se vendieron 415.581 vehículos livianos, un 60,6% más que el mismo periodo del año 2020. ¿Es posible asegurar que parte de este aumento en la venta de autos responde a un aumento en la liquidez de las personas?.

Este trabajo nace con la idea original de estudiar si los retiros de fondos de pensiones tienen una relación causal con las ventas de autos en una importante empresa del rubro automotriz. Estudio que puede tomar más del tiempo disponible para la realización del capstone. En base a eso es que se decide dividir el trabajo, donde en esta primera parte se estudiará las relaciones entre variables propias del conjunto de datos entregados, como también la incorporación de nuevas variables que permitan entender las externalidades que pueden afectar la venta de vehículos en la empresa del rubro automotriz y una posterior medida de impacto entre las ventas reales y las ventas esperadas, que serán producto de la predicción de ventas que se logre realizar.

2. Trabajos Relacionados

Una serie de industrias vieron impactadas sus ventas durante la época de pandemia, problemas de producción, distribución y abastecimiento. La industria automotriz abarca una serie de complejas conexiones con otras industrias y no solo debe ser visto desde el punto de vista doméstico. La variedad de materias primas es enorme y en constante evolución que desafía a esta industria a ser eficiente y aumentar sus ingresos (Buruzs, A., & Torma, A., 2018). Desde que la Organización Mundial de la Salud decretó la Pandemia COVID-19 en marzo de 2020, los países comenzaron a restringir las actividades no esenciales y una serie de industrias y países comenzaron a cerrar sus fábricas, tanto por el aumento explosivo de casos y la gravedad de sus síntomas como también para proteger a los grupos más susceptibles de contagio. Decisión que trajo consecuencias como la falta de materias primas y retraso en envíos de mercadería, afectando directamente al mercado automotor. (Lewicki, 2020). Volvo calculó que la caída en ventas en marzo de 2020 fue de un 41% respecto a 2019 (Dooley, J., 2020, April 21) y con una potencial de caída de 60%, 80%. Diseñó un impacto de negocio para 2020 basado en la recuperación y el mercado chino, consistente en 3 etapas: de marzo a abril una etapa de contagio y fuera de control, de mayo a junio ganando control y a partir de julio a diciembre una nueva normalidad.

Katrakazas, C. (2021), realiza un trabajo para identificar el impacto de la pandemia COVID-19 en la conducta de manejo usando datos reales de movilidad y pronóstico con series de tiempo, donde se utilizó un modelo XGBoost para identificar los indicadores más influyentes del COVID-19 y un modelo SARIMA para obtener pronósticos de conducta de manejo.

3. Hipótesis y Objetivos

3.1. Hipótesis

El aumento de liquidez y las condiciones de encierro de la pandemia se asocia positivamente en las ventas de Importante empresa del rubro automotriz.

3.2. Objetivos específicos

- Agregar las estadísticas diarias del agregado monetario M1, compuesto de billetes, monedas, cuentas vistas y cuentas corrientes.
- Agregar datos de casos covid y restricción de movilidad.
- Agregar variables de tipo de cambio (peso chileno por dólar) e inflación.
- Agregar tasa del banco central como proxy de las tasas para créditos.
- Relacionar las ventas de La empresa del rubro automotriz con las fechas de los 3 retiros de fondos de pensiones.
- Análisis exploratorio de datos
- Entrenar modelos de regresión que permitan predecir las ventas.
- Determinar modelo que permita realizar un pronóstico de ventas.
- Medir el impacto de cambios de liquidez sobre sobre la venta de autos.

4. Datos y Metodología

4.1. Datos

Los datos en gran parte son propiedad de La empresa del rubro automotriz y representan las ventas diarias de la compañía entre 2013 y marzo de 2022. Dada las características ya descritas anteriormente hay una serie de situaciones ocurriendo al mismo tiempo que pueden estar representadas en variables externas.

4.2. Metodología

En términos de negocio es tentador poder contar con una herramienta que permita pronosticar las ventas futuras para determinar aquellas metas de negocio que habilitarán el cumplimiento de los objetivos del directorio y que este último pueda disponibilizar los recursos para su desarrollo; el ejercicio en términos generales puede abarcar varios años hacia adelante, aunque lo común es ver esto año a año con un nivel de detalle que permita disponer de los recursos necesarios. Los resultados se revisan mes a mes y se mide su cumplimiento, su desvío de la meta y las oportunidades de mejora. Pero así también es importante conocer

el impacto que han tenido ciertos eventos para dimensionar los efectos en el futuro y poder adelantarse a los efectos y tomar decisiones a tiempo. En este último punto es que se quiere profundizar y así permitir tener herramientas que en un futuro ayuden a dar respuesta a tiempo al negocio para minimizar pérdidas o aprovechar una oportunidad de negocio.

Las etapas de este capstone contemplan:

- Etapa 1: Recolección y limpieza de datos
- Etapa 2: Análisis exploratorio de datos y selección de variables
- Etapa 3: Modelamiento de datos
- Modelos a revisar:
 - Modelo base: Serie de tiempo univariada de ventas semanales, total nacional.
 - Vector auto regresivo: Serie de tiempo multivariada de ventas semanales, total nacional.
 - Gradient Boosting Regressor con ventas semanales, total nacional.
- Etapa 5: Selección del modelo

El proceso descrito es más bien no lineal y lleno de iteraciones que llevan a retroceder a las etapas anteriores. Naturalmente nacen nuevas preguntas que deben ser respondidas, por lo que es necesario volver a una etapa anterior para buscar aquel análisis que permita mejorar nuestras respuestas y se acerque más a lo buscado.

Durante la etapa de recolección y limpieza de datos se obtuvo la información de La empresa del rubro automotriz y patentes, pero adicionalmente la idea era identificar aquellas variables del entorno que podrían afectar la venta diaria de La empresa del rubro automotriz. Para representar la dificultad de compra de autos por parte de La empresa del rubro automotriz, se agrega el dólar observado. Por el lado del cliente, se busca representar la dificultad o facilidad para adquirir un auto y para eso se utiliza la Tasa de Política Monetaria (TPM), que definirá que tan costoso o no es adquirir un crédito, lo que se relaciona a su vez con la UF que busca representar el costo de la vida en el tiempo. Los agregados monetarios M1 y M2 buscan ser una aproximación al nivel de liquidez de las personas, especialmente en los meses donde hubo disponibilidad para retirar el 10% de los ahorros previsionales. Adicionalmente se agregan los rangos de fechas, donde era posible realizar los 3 retiros previsionales del 10%. Los casos covid por comuna buscan representar la realidad de casos covid de su comuna, razón por la cual alguien podría decidir si comprar un auto o no porque su comuna tiene muchos casos covid y quiere evitar el contacto físico. Se asume que las variables de dólar observado, UF, TPM y agregados monetarios afectan a cada comuna y región tal como vienen por lo que no es necesario aperturar o distribuir el indicador por población u otra apertura particular. No así para los casos covid por comuna que se agregan al conjunto de datos según corresponda al día y comuna.

De acuerdo a la naturaleza de los datos, se hizo un tratamiento ad-hoc para aquellos valores faltantes. Es así como para los datos de origen económico, donde es común tener información para los días hábiles entre lunes y viernes, es que se utilizó el dato del último día hábil para completar la información de los días sábado y domingos, a diferencia de los casos covid por ejemplo que se interpoló linealmente la información de los días faltantes. Para el caso de los datos provenientes de las ventas de La empresa del rubro automotriz, era común encontrar días sin venta por comuna. En este caso se decide completar con cero ventas, dado que representa de mejor manera la información faltante y mantiene la periodicidad que se requiere para los datos en serie de tiempo.

Los datos atípicos se detectaron mediante la agrupación de información a nivel de día de venta y comuna, bajo este criterio se logró extraer un total de 4.938 registros.

En total se cuenta con 1.143.447 filas, separadas en 30 columnas. Información a nivel de venta diaria desde el 2 de enero de 2013 al 28 de marzo de 2022, para 16 regiones y 339 comunas. Ya contando con estos datos ordenados es posible empezar a trabajar en la serie de tiempo, dado que se cuenta con una estructura periódica básica.

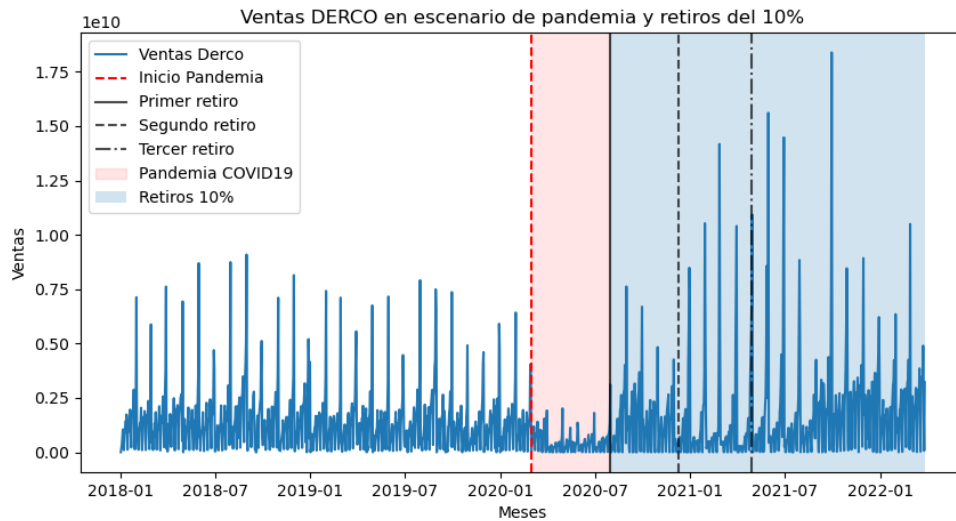
Para la parte descriptiva de los datos se trabajará con un detalle a nivel de comuna y los resultados presentados a nivel regional, mientras que los modelos se trabajarán de manera agregada nacional y semanal. Esto para evitar ciertos problemas de datos con valor 0 en las series de tiempo y así poder facilitar el análisis y exploración de los datos.

Los datos se dividirán en tres grupos, Prepandemia, Encierro y Retiros, de tal manera que se pueda separar claramente los efectos a medir a partir de las fechas donde comenzó la pandemia y cuando se realizaron los retiros.

Para medir los efectos de Encierro y Retiros, se procederá a realizar una predicción de los datos según las fechas que se quiera revisar y compara la predicción con las ventas reales del período en cuestión. Adicionalmente se revisará si las medias de las ventas son distintas a las ventas prepandemia para conocer si estamos por debajo o sobre las ventas en un escenario normal sin COVID-19 y sin retiros.

5. Resultados

Agrupando la información a nivel diario nacional es posible observar un crecimiento en ventas a posterior a los retiros, lo que permite elevar la pregunta si efectivamente estas ventas son una respuesta al aumento de liquidez por parte de las personas.



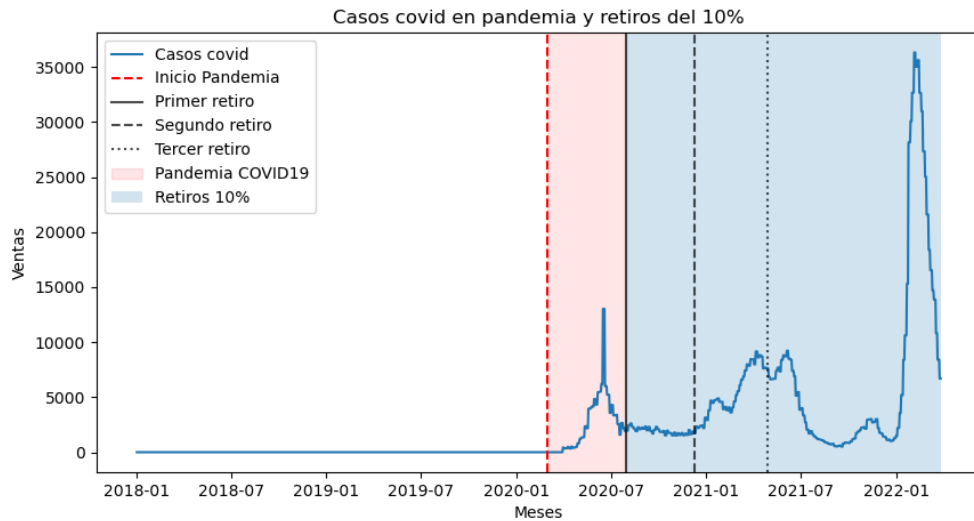
	<i>Pre pandemia</i>	<i>Pandemia</i>	<i>Retiro 1</i>	<i>Retiro 1 y 2</i>	<i>Retiro 1, 2 y 3</i>
<i>count</i>	791	151	133	272	607
<i>mean</i>	1.368.608.430	421.348.878	1.338.261.259	1.270.063.828	1.623.625.322
<i>std</i>	1.329.442.013	395.244.942	1.292.295.328	1.818.893.305	2.081.491.291
<i>min</i>	0	0	0	0	0
<i>25%</i>	553.642.240	147.581.356	262.164.782	164.897.171	323.739.356
<i>50%</i>	1.126.012.399	330.120.220	1.196.454.540	698.135.093	1.074.112.268
<i>75%</i>	1.754.671.400	560.643.710	1.971.051.504	1.654.942.492	2.170.986.666
<i>max</i>	9.088.940.012	2.015.431.334	7.619.229.256	14.173.230.415	18.384.779.998

Si bien se observa una diferencia en las ventas, los únicos promedios de venta significativamente diferentes de la época pre pandemia son las ventas de Pandemia y los retiros desde el 1 al 3.

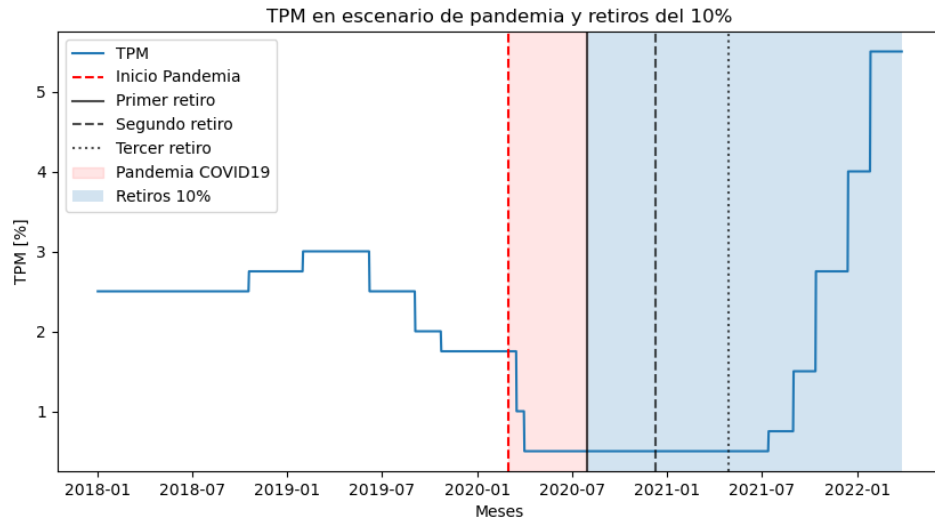
<i>Vs Promedio venta Pre Pandemia</i>	<i>p-value</i>
<i>Pandemia</i>	1,7337E-17
<i>Retiro 1</i>	0,8060
<i>Retiro 1 y 2</i>	0,3404
<i>Retiro 1, 2, y 3</i>	0,0054

En términos generales podemos decir que, en el acumulado, los retiros de fondos previsionales provocaron un aumento de la venta promedio que es significativamente diferente a la venta promedio pre pandemia.

A continuación se presenta la evolución de los casos covid a nivel nacional, identificando las fechas de los retiros previsionales solo con fines ilustrativos.

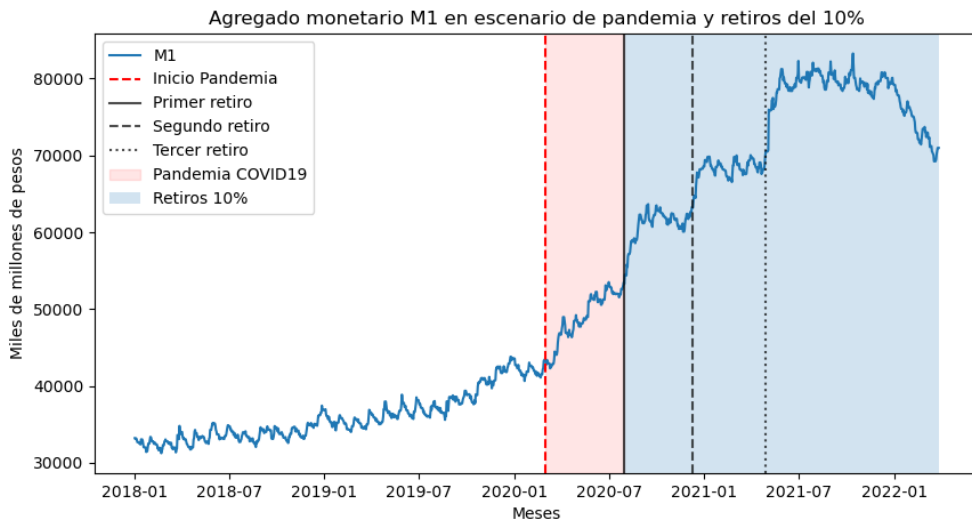
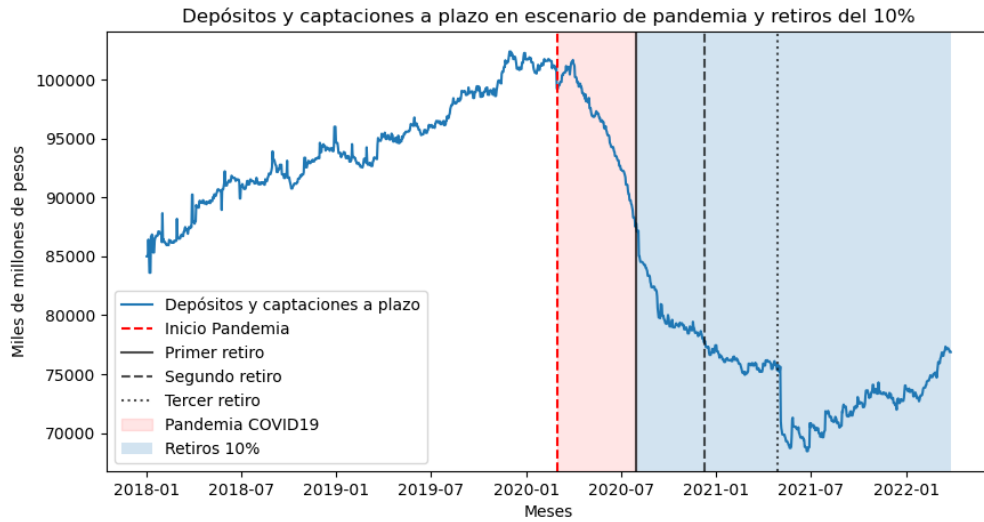


La Tasa de Política Monetaria es una tasa que actúa en el corto plazo y puede representar la facilidad de acceso a un crédito de consumo. En este caso se observa que durante los primeros meses de pandemia y primeros retiros previsionales se tuvieron las menos tasas de los últimos años, lo que presenta una gran oportunidad para adquirir un auto.



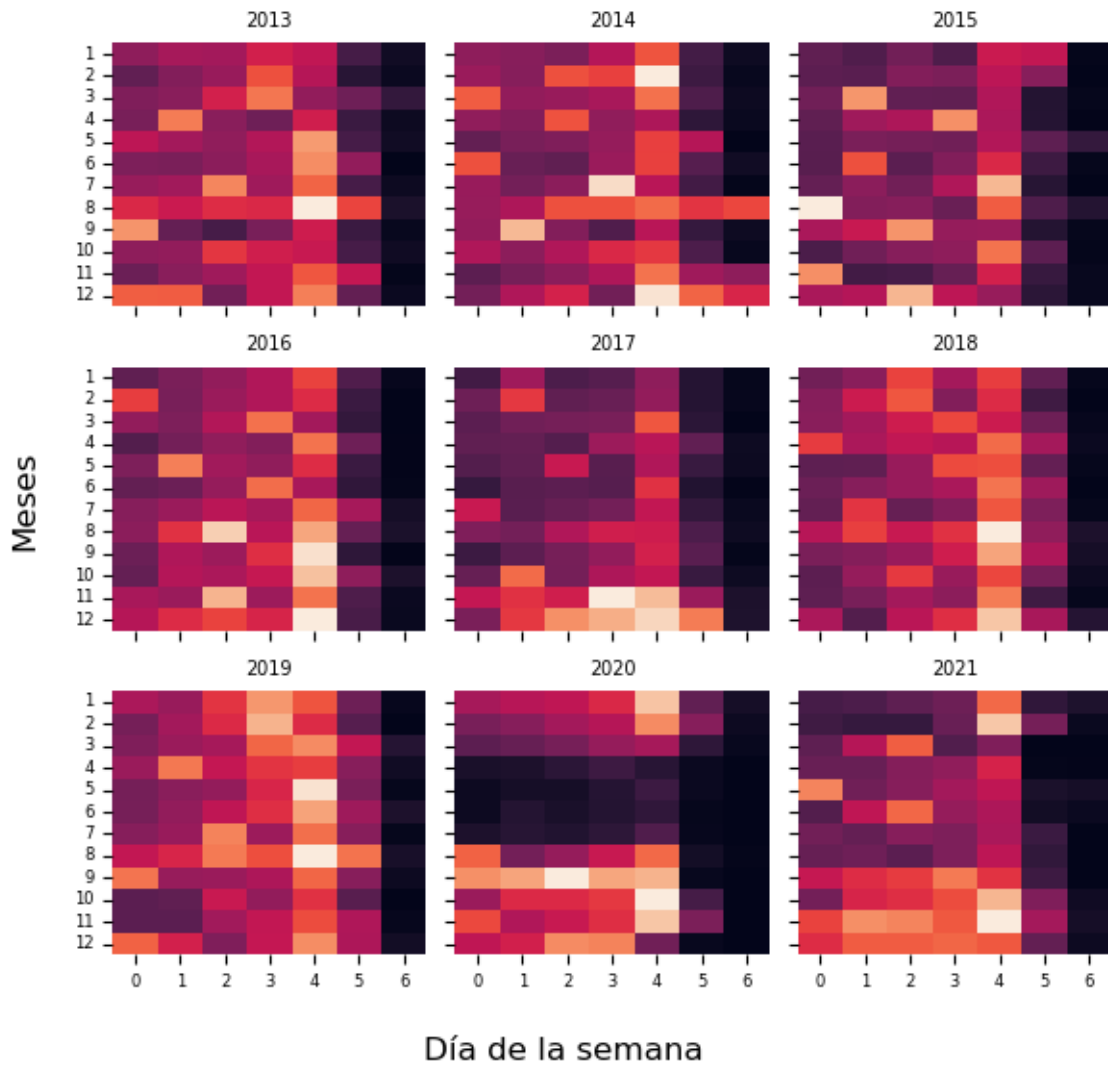
A partir de la pandemia se observa una caída abrupta de los depósitos a largo plazo, mientras el agregado M1 sube, en menor medida. Es posible pensar que las personas deciden dejar de depositar dinero en el banco para aumentar la liquidez en situación de incertidumbre. Esto también podría tener un estímulo arrastrado por motivo del estallido social ocurrido durante el último trimestre del 2019.

Se observa claramente como los retiros afectan la disminución de ahorros de largo plazo para aumentar las cuentas de mayor liquidez agrupadas en M1, considerando que los retiros previsionales comenzaron el 30 de julio de 2020.



Una vez entendidas las series de tiempo anteriores, es relevante entender como se distribuye la compra de autos por mes y día de semana para todos los años. Donde se aprecia un patrón claro durante la semana y que se ve interrumpido entre marzo y julio de 2020.

Concentración de compras por mes y día de semana

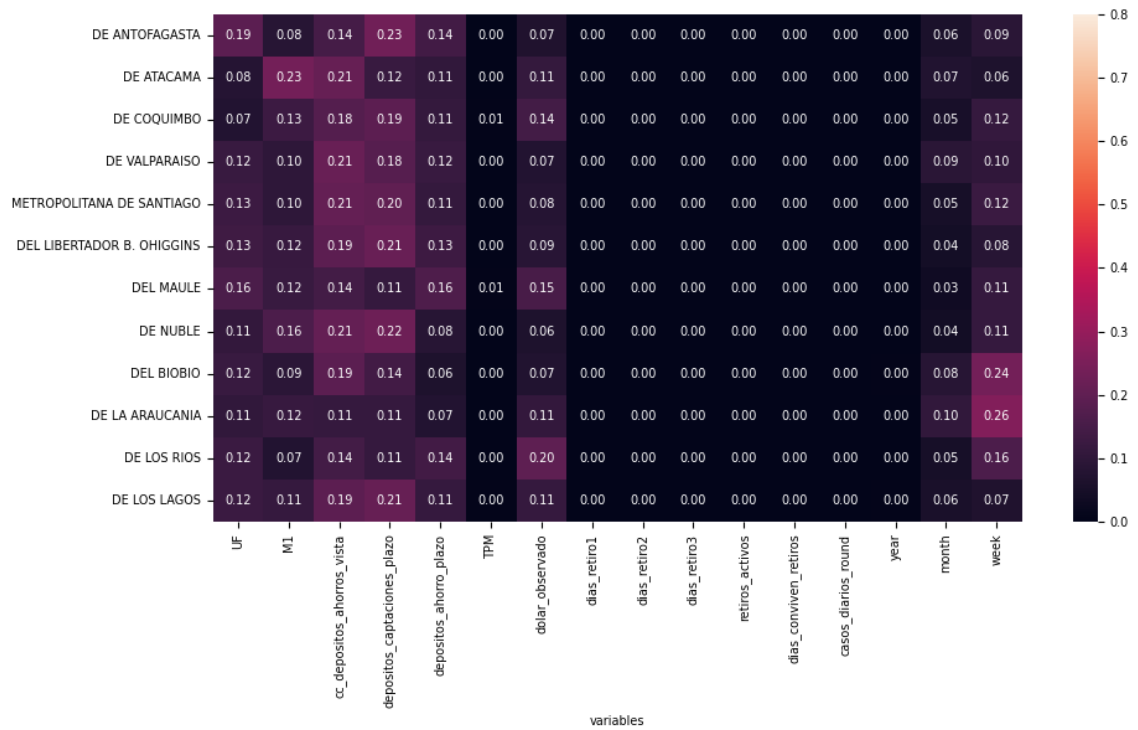


Una vez entendido el comportamiento de las series de tiempo con las fechas relevantes dentro del periodo revisado, resulta interesante entender los patrones y relaciones entre los datos. Entendiendo que el precio de venta puede estar influenciado por otros factores, se decide utilizar un modelo de Gradient Boosting Classifier para entender el impacto de las variables sobre la variable objetivo precio_de_lista.

5.1. Antes de pandemia

Antes de pandemia, corresponde al periodo comprendido desde el 1 de enero de 2018 al día anterior que se decretara la pandemia en Chile.

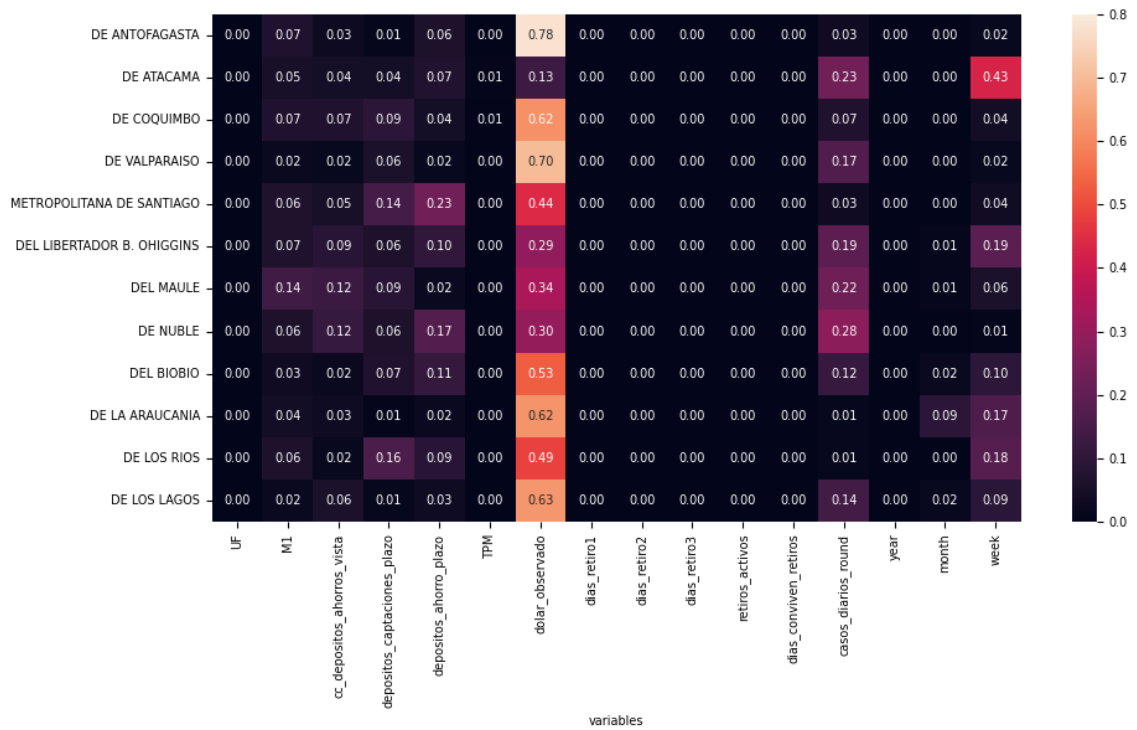
Se observa que el precio de lista es influido por el comportamiento de la UF, el nivel de liquidez M1 y el dinero disponible en las cuentas vista de las personas, algunos depósitos a plazo y el dólar observado; con una leve importancia de la semana del año. Las personas pagan de acuerdo a lo que tienen disponible de en sus cuentas vistas, eventualmente el pie y recurren a ciertos ahorros.



5.2. Encierro

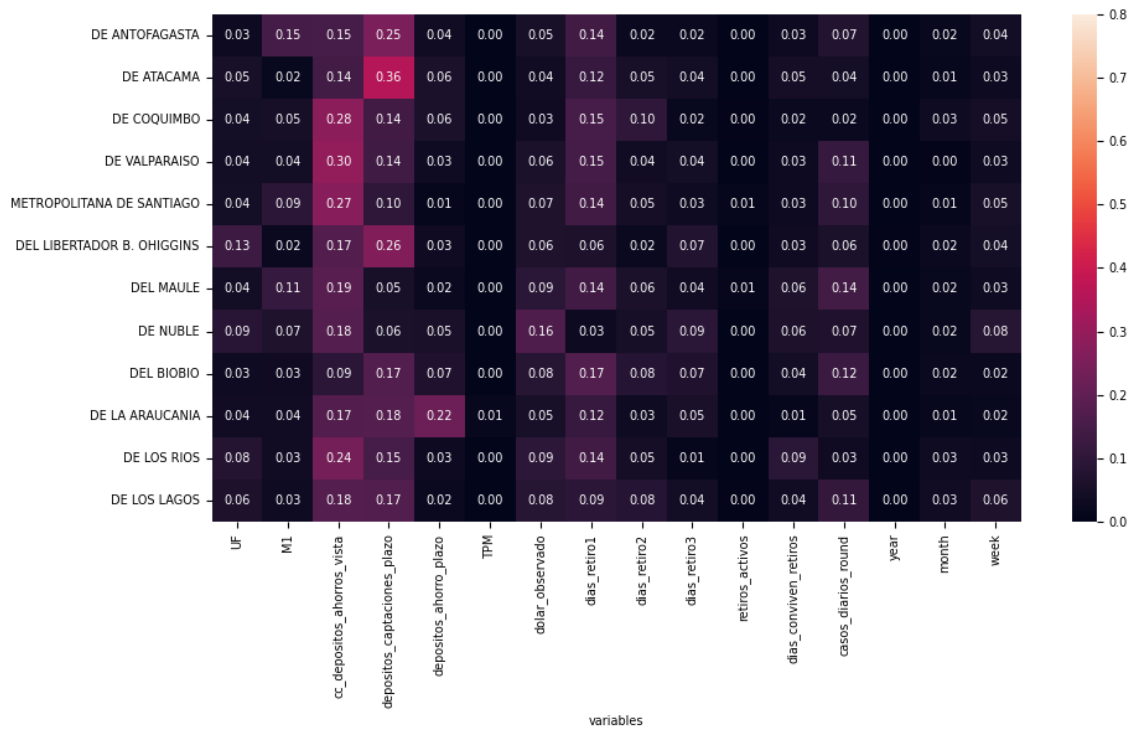
Encierro, corresponde al período desde el inicio de la pandemia, el 1 de marzo de 2020 hasta el primer retiro previsual el 30 de julio de 2020.

Solo se observa que las ventas dependen fuertemente del dólar, seguido por los casos covid de la región. En menor medida los ahorros de largo plazo impactaron en las ventas.



5.3. Retiros

Retiros, corresponde al periodo comprendido entre el 30 de julio de 2020 en adelante, hasta marzo de 2022. Aquí nuevamente se ve un impacto importante de los ahorros, pero ahora se hace presente la importancia de los días que han corrido desde el primer, segundo y tercer retiro, donde el primer retiro tiene un mayor impacto que el resto. De igual manera se mantiene el impacto de los casos covid de la región.

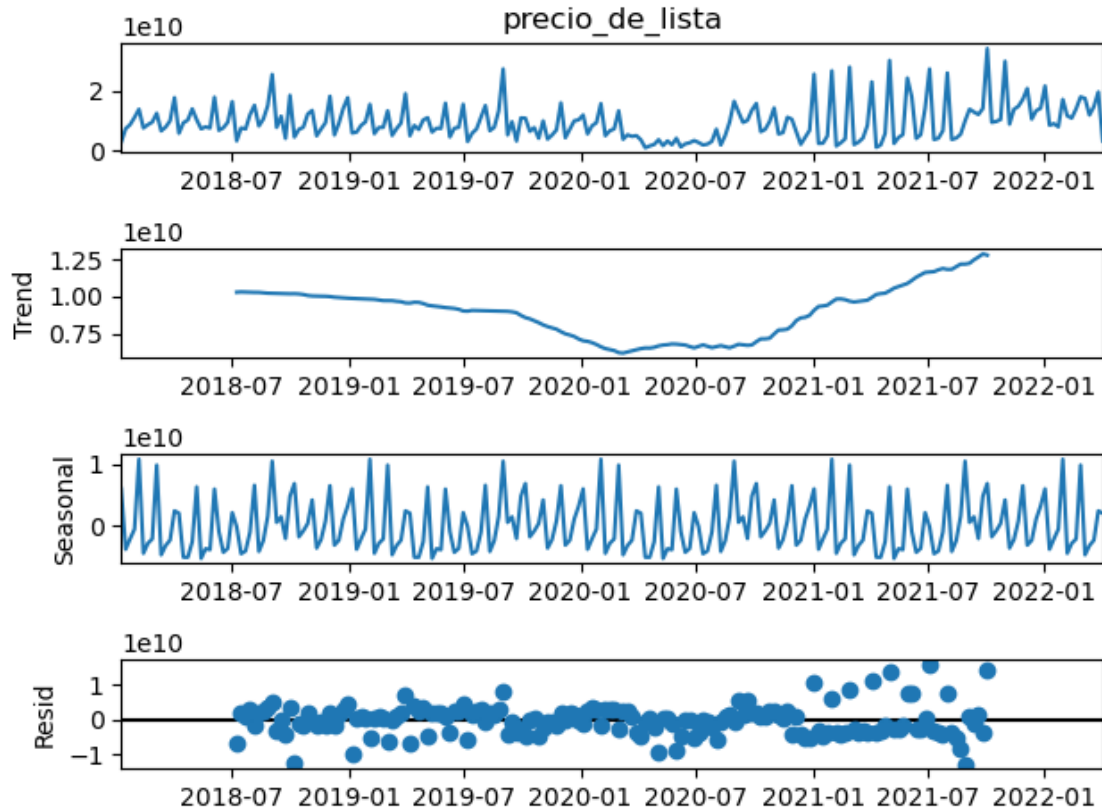


Los principales hallazgos tienen relación con el cambio en la importancia de los datos sobre la venta diaria. En particular permite explicar que durante este período si hay información de venta que se explica por una mayor disponibilidad de dinero en las cuentas corrientes y cuentas vistas y en una menor medida los casos diarios de covid por región pudieron impactar en la decisión de comprar un auto.

5.4. Serie de tiempo univariada

Para trabajar con una serie de tiempo univariada de las ventas se procede a realizar un remuestreo de los datos para llevarlos desde una periodicidad diaria a semanal. De esta manera se elimina los días con venta 0 y se disminuye el ruido provocado por contar con información de mayor detalle. Con esto es posible contar con una modelo base donde es posible trabajar como punto de partida y comparación para posibles mejoras.

Descomponiendo la serie no es posible señalar si es estacionaria o no y para el caso de la estacionalidad se observa cierto patrón pero no es claro decir si es solo para 1 semana, 5 o más. Por lo tanto se decide trabajar directamente con un modelo ARIMA.



5.5. Estacionariedad

Una vez hecho el remuestreo de los datos por semana, se procede a revisar si la serie es estacionaria o no, utilizando el Test de Dickey-Fuller Aumentado, donde se obtiene un p-value de 0,3168, que es mayor que 0,05 y por lo tanto no se puede rechazar la hipótesis nula, resultando en una serie no estacionaria. Dado este resultado es necesario realizar una primera diferenciación sobre la cual se realiza nuevamente el test de

Dickey-Fuller Aumentado y se obtiene un p-value de $1,9903e-9$, que permite rechazar la hipótesis nula y es posible indicar que la serie diferenciada es estacionaria.

De esta manera es necesario comenzar con las primeras pruebas para determinar qué modelos utilizar. A primera vista este podría ser un modelo ARIMA o SARIMA, pero la estacionalidad (s) no se ve tan claramente.

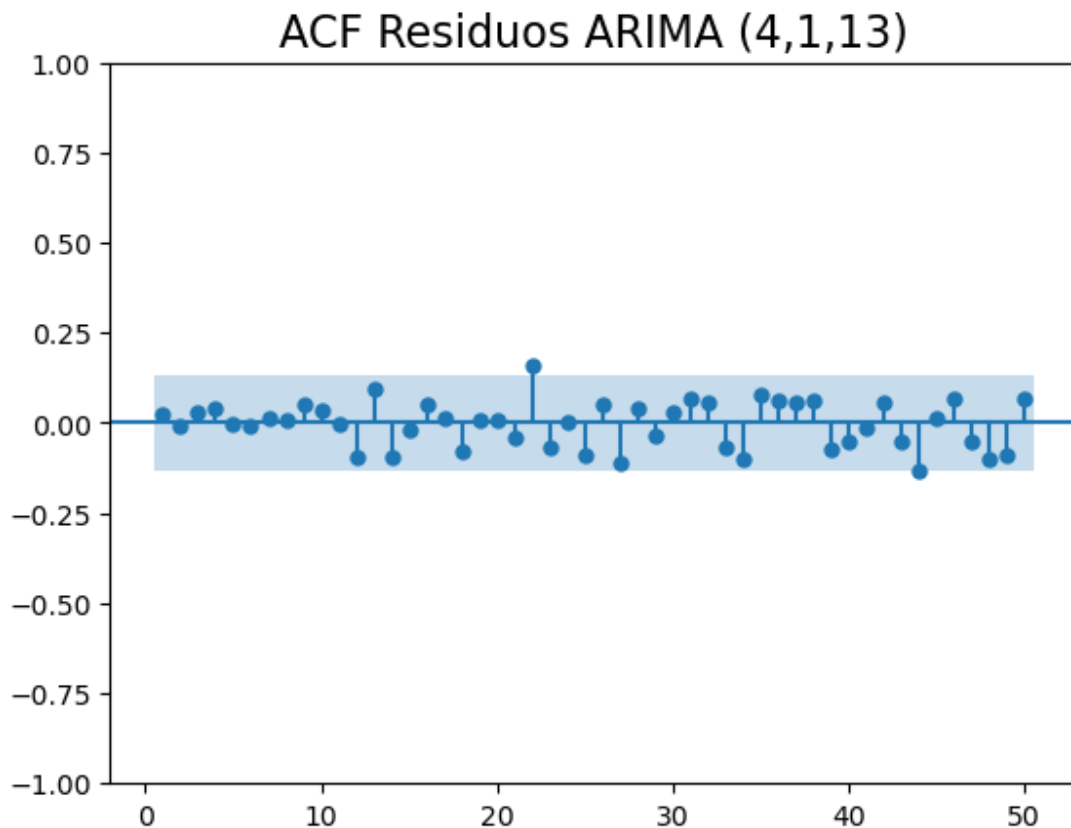
5.6. Modelo ARIMA

Analizando la función de autocorrelación parcial y función de autocorrelación es posible determinar los autoregresivos 4, 9 y 13 como posible soluciones mientras que para la media móvil también se puede revisar las alternativas de 4, 9 y 13. Al evaluar los posibles modelos se obtienen los siguientes resultados:

<i>Modelo</i>	<i>AIC</i>	<i>p-value</i>
<i>ARIMA (4,1,0)*</i>	10493,5117	ar.L4 = 0,000
<i>ARIMA (9,1,0)</i>	10430,2287	ar.L9 = 0,262
<i>ARIMA (13,1,0)</i>	10422,1008	ar.L13 = 0,799
<i>ARIMA (4,1,4)*</i>	10458,8670	ar.L4 = 0,000 ma.L4 = 0,000
<i>ARIMA (4,1,9)</i>	10432,8357	ar.L4 = 0,000 ma.L9 = 0,469
<i>ARIMA (4,1,13)*</i>	10433,9915	ar.L4 = 0,002 ma.L13 = 0,033
<i>ARIMA (9,1,4)</i>	10421,6886	ar.L9 = 0,457 ma.L4 = 0,000
<i>ARIMA (9,1,9)</i>	10427,9411	ar.L9 = 0,943 ma.L9 = 0,930
<i>ARIMA (9,1,13)</i>	10425,4970	ar.L9 = 0,999 ma.L13 = 0,860
<i>ARIMA (13,1,4)</i>	10425,6411	ar.L13 = 0,890 ma.L4 = 0,928
<i>ARIMA (13,1,9)</i>	10431,7913	ar.L13 = 0,757 ma.L9 = 0,934
<i>ARIMA (13,1,13)</i>	10429,4194	ar.L13 = 0,184 ma.L13 = 0,383

Bajo el criterio del menor AIC, el mejor modelo es ARIMA (9,1,4), pero el noveno lag autoregresivo no es significativo (p-value = 0,457) y en consecuencia el coeficiente de ese noveno lag no es estadísticamente distinto de 0. Siguiendo con este criterio, las posibles opciones son 3: ARIMA (4,1,0), ARIMA (4,1,9), ARIMA (4,1,13).

Si comparamos el Log – Likelihood ratio entre los modelos antes mencionados, el modelo ARIMA (4,1,13) obtiene un p-value = 0,000 al compararlo tanto contra ARIMA (4,1,0) y ARIMA (4,1,4) de esta manera agregar más restricciones al modelo permite explicar mejor el comportamiento de la venta en La empresa del rubro automotriz.

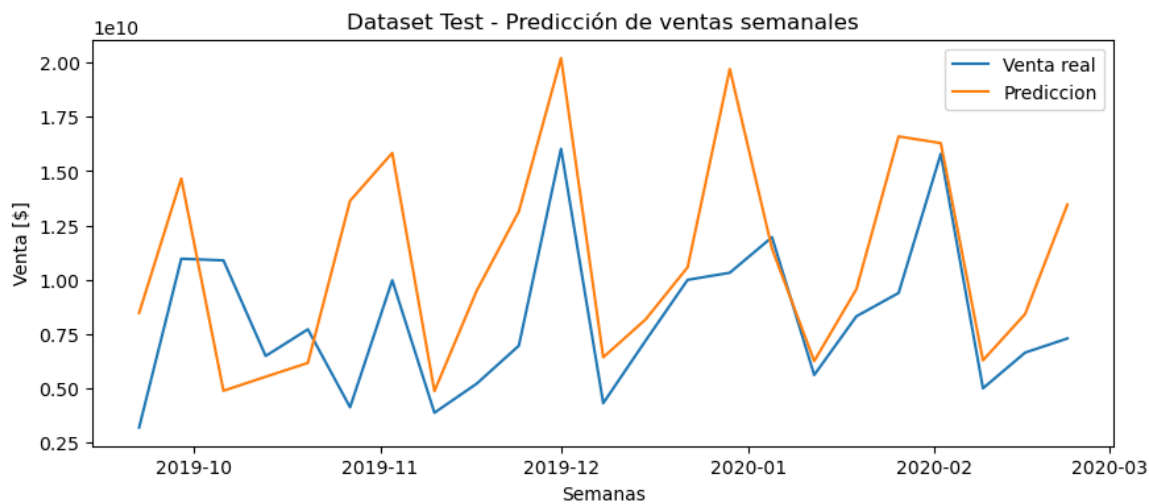
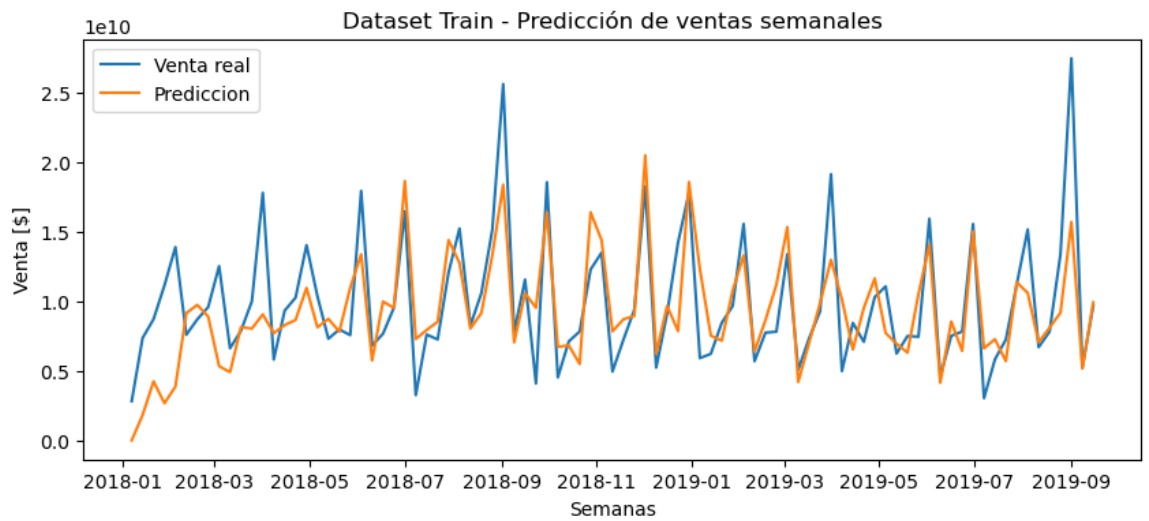


Adicionalmente revisando la función de autocorrelación de los residuos se observa que no existe autocorrelación de los residuos por lo tanto se valida nuevamente el modelo.

5.7. Forecast – Antes de pandemia

Aplicando el modelo ARIMA (4,1,13) a los datos comprendidos entre enero de 2018 y el 29 de febrero de 2020, agrupados por semana.

<i>Dataset</i> (semanal)	<i>Fecha inicio</i>	<i>Fecha fin</i>	<i>RMSE</i>	<i>Median Absolute Error</i>	<i>Mean absolute percentage error</i>
<i>Train</i>	2018-01-07	2019-09-15	3.374.153.678,0348	1.607.727.913,4520	0,2741
<i>Test</i>	2019-09-22	2020-02-23	4.529.677.719,7758	2.114.356.146,3791	0,5256



En términos generales se observa que durante la etapa de entrenamiento, las predicciones se ajustan a los datos reales con un error porcentual absoluto del 27%. Al hacer la predicción con las fechas del dataset de

test, el error porcentual sube a un 52%, donde se observa una sobre estimación de las semanas con mayor venta.

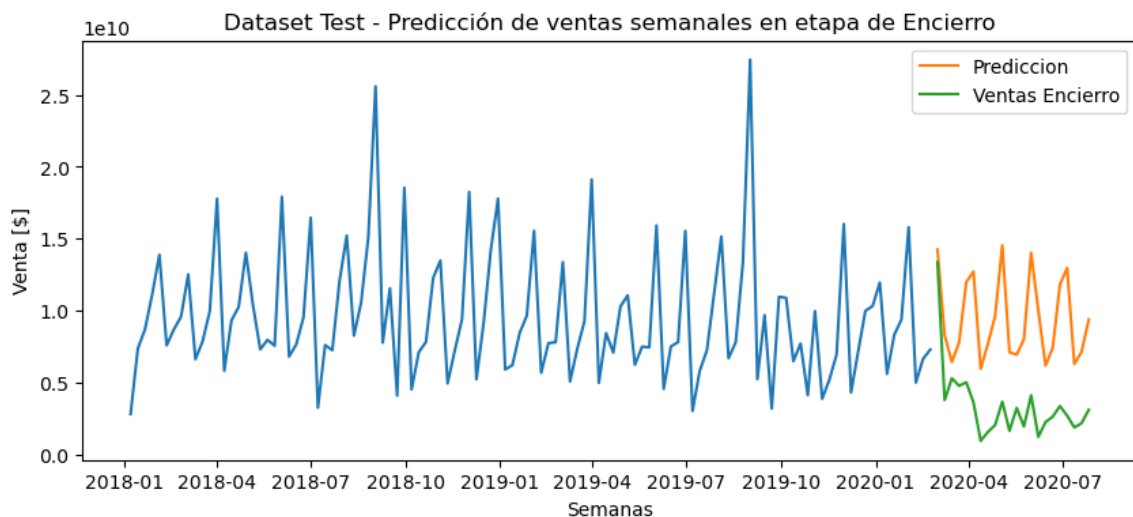
Las ventas reales representan, en promedio, el 82,0475% de las ventas predichas en el grupo de testeo.

5.8. Forecast – Encierro

Entrenando el modelo ARIMA (4,1,13) con información hasta el 29 de febrero de 2020, procedemos a hacer una predicción de como habrían sido las ventas si no hubiese existido el encierro entre el 1 de marzo de 2020 hasta el 30 de julio de 2020, cuando ocurre el primer retiro.

<i>Dataset</i> (semanal)	<i>Fecha inicio</i>	<i>Fecha fin</i>	<i>RMSE</i>	<i>Median Absolute Error</i>	<i>Mean absolute percentage error</i>
<i>Train</i>	2018-01-07	2020-02-29	3.278.964.669,3449	1.482.242.865,1590	0,2881
<i>Test</i>	2020-03-01	2020-07-30	6.589.436.284,5670	5.765.268.874,4591	2,5252

Las ventas durante la etapa de Encierro son menores a los mínimos históricos por semana. El efecto del encierro es superior a lo visto antes. Cómo se observaba anteriormente, como “si todos los días hubiesen sido domingos”.



Las ventas reales representan, en promedio, el 35,7621% de las ventas predichas.

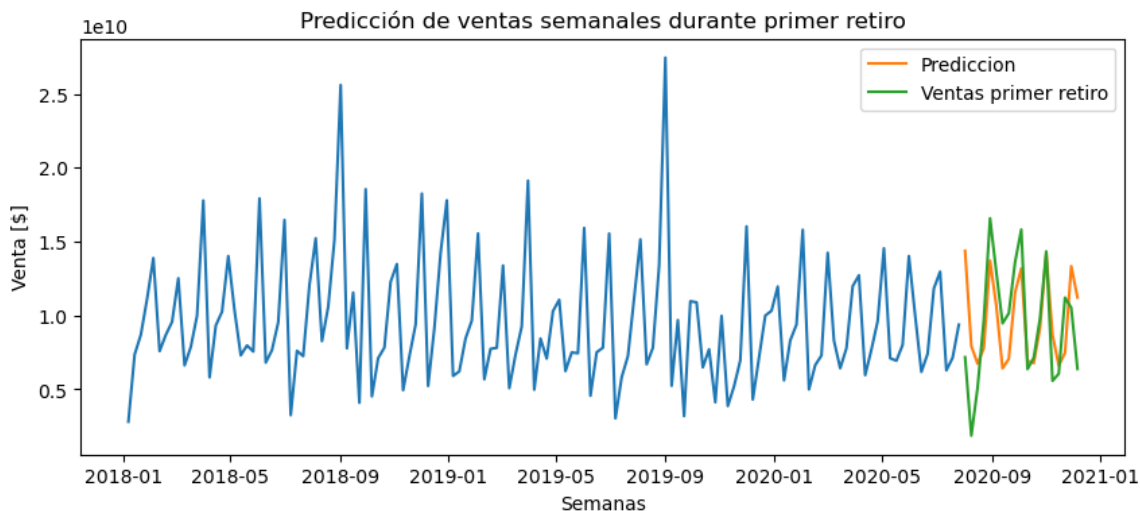
En promedio, mensualmente se habrían perdido unos 26 mil millones de pesos por efecto de las restricciones de movilidad.

<i>Mes</i>	<i>Venta real [\$]</i>	<i>Predicción [\$]</i>	<i>Diferencia</i>	<i>YoY real</i>	<i>YoY pred</i>
<i>Marzo 2020</i>	32.257.943.292	48.720.764.770	-16.462.821.478	-0,22	0,17
<i>Abril 2020</i>	8.163.866.595	35.989.265.846	-27.825.399.251	-0,80	-0,10
<i>Mayo 2020</i>	14.607.596.510	50.619.992.511	-36.012.396.001	-0,62	0,31
<i>Junio 2020</i>	9.475.336.177	35.424.788.079	-25.949.451.902	-0,73	-0,01
<i>Julio 2020</i>	9.857.900.344	35.751.365.418	-25.893.465.074	-0,76	-0,13

5.9. Forecast – Retiros

Para esta etapa se utilizará la predicción anterior, eliminando parcialmente el efecto encierro de la pandemia, y así comprender si hay diferencias importantes en cuanto al aumento de ventas provocadas por el primer retiro.

El periodo de análisis será predecir las ventas entre el primer retiro a partir del 30 de julio de 2020 y el segundo retiro el 10 de diciembre de 2020.



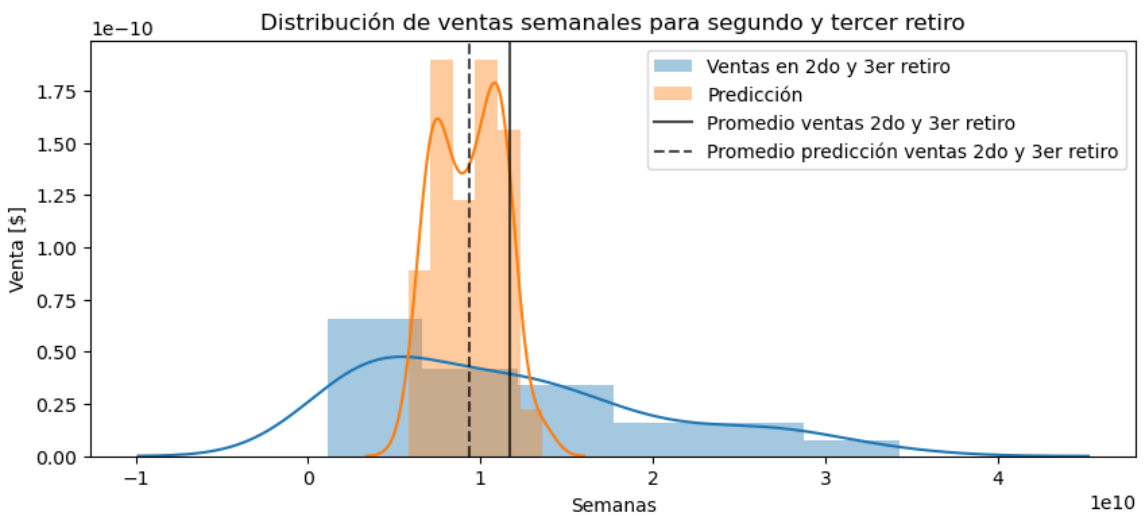
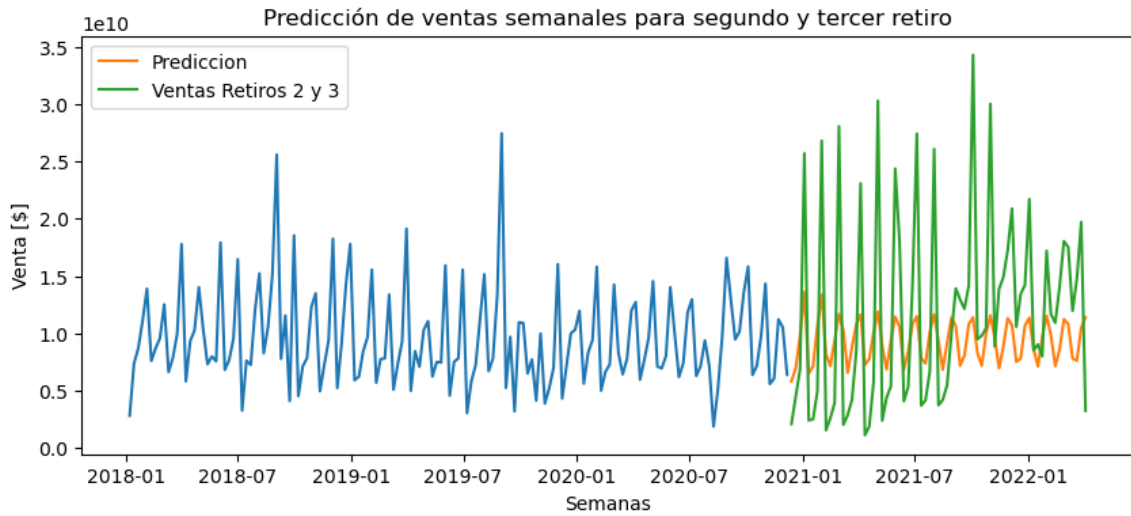
<i>Dataset</i>					<i>Mean</i>
<i>Retiro 1</i>	<i>Fecha inicio</i>	<i>Fecha fin</i>	<i>RMSE</i>	<i>Median Absolute</i>	<i>absolute</i>
<i>(semanal)</i>				<i>Error</i>	<i>percentage</i>
					<i>error</i>
<i>Train</i>	2018-01-07	2020-07-29	3.002.199.396.2193	1.060.181.500.1280	0,2404
<i>Test</i>	2020-07-30	2020-12-10	3.198.477.553.8850	2.605.782.265.0418	0,4360

Al realizar un test de t de student, entre los valores de la predicción y los valores reales del periodo comprendido entre el inicio del primer retiro y el comienzo del segundo retiro, el valor p resulta ser 0,8521 por lo tanto no se puede rechazar la hipótesis nula y la diferencia entre las medias de las predicciones de venta y las ventas reales es estadísticamente igual a 0. Al realizar el mismo test contra las ventas reales del periodo prepandemia, el valor p resulta en 0,9060. El efecto del primer retiro llevó las ventas a los niveles prepandémicos.

Por último, queda realizar la predicción de ventas para el segundo y tercer retiro. Tomando en cuenta que estamos reemplazando los datos que corresponden a las fechas de encierro, por las predicciones de ese mismo período y luego retomando los datos reales correspondientes al primer retiro. El periodo por predecir es desde el 13 de diciembre de 2020 al 4 de abril de 2022

<i>Dataset</i>					<i>Mean</i>
<i>Retiro 2 y 3</i>	<i>Fecha inicio</i>	<i>Fecha fin</i>	<i>RMSE</i>	<i>Median Absolute</i>	<i>absolute</i>
<i>(semanal)</i>				<i>Error</i>	<i>percentage</i>
					<i>error</i>
<i>Train</i>	2018-01-01	2020-12-12	2.997.181.508,8191	1.936.530.716.5304	0.2719
<i>Test</i>	2020-12-13	2022-04-03	8.849.949.845.9570	4.445.923.016.6907	0.8604

En términos generales se obtiene un resultado bastante dispar y aquí hay que tener en cuenta que la duración de los retiros fue de 1 año para cada uno por lo tanto aquí se también puede haber un efecto rezagado del primer retiro. Adicionalmente se observa mínimos en ventas semanales como los observados anteriormente en la etapa de encierro y esto sucede por que durante esa etapa aun existían restricciones de movilidad en algunas regiones del país. Ya en el último período de 2021 las restricciones de movilidad se suprimieron y las personas y empresas comenzaron una transición hacia una movilidad mucho más libre. Probablemente la llegada de nuevos autos y una preferencia por una movilidad más independiente, con una mayor disponibilidad de liquidez para pagar por un auto pueda mantener los precios de venta sin alcanzar los mínimos vistos en la etapa de encierro.



Al realizar la una prueba de t de student para las medias de las ventas y la predicción se obtiene un p-value de $5,3431e-05$ que por lo tanto es menor que 0.05 y en consecuencia se puede rechazar la hipótesis nula. Las medias de las predicciones son estadísticamente distintas de las medias de las ventas. Al realizar la misma comparación contra la época prepandemia el valor p es de 0,7316. Con este resultado no se puede rechazar la hipótesis nula y las medias no son estadísticamente distintas. De esta manera las ventas con retiros no se alejan demasiado de las ventas anteriores. Podríamos decir que “se está vendiendo lo mismo que antes”, o que se está regresando a esta nueva normalidad. Los retiros habrían acelerado esta vuelta a “la normalidad”.

5.10. Regresión lineal multivariada

Otra alternativa para realizar un pronóstico de las ventas es utilizar un modelo de Vector Auto Regresivo, que al igual que la regresión univariada, requiere que las series de tiempo sean estacionarias. Ahora bien las variables que serán ocupadas son, además de precio_de_lista, son UF, M1, cc_depositos_ahorros_vista, depositos_ahorro_plazo, depositos_captaciones_plazo, TPM y dolar_observado.

Es necesario diferenciar dos veces el dataset con las variables descritas, dado que en la primera diferenciación TPM no pasa la prueba aumentada de Dickey Fuller y obtiene un valor p de 0,4884, no pudiendo rechazar la hipótesis nula.

Bajo el criterio de información de Akaike (AIC), se selecciona el menor valor para el modelo, obteniendo el orden 11.

VAR Order Selection (* highlights the minimums)

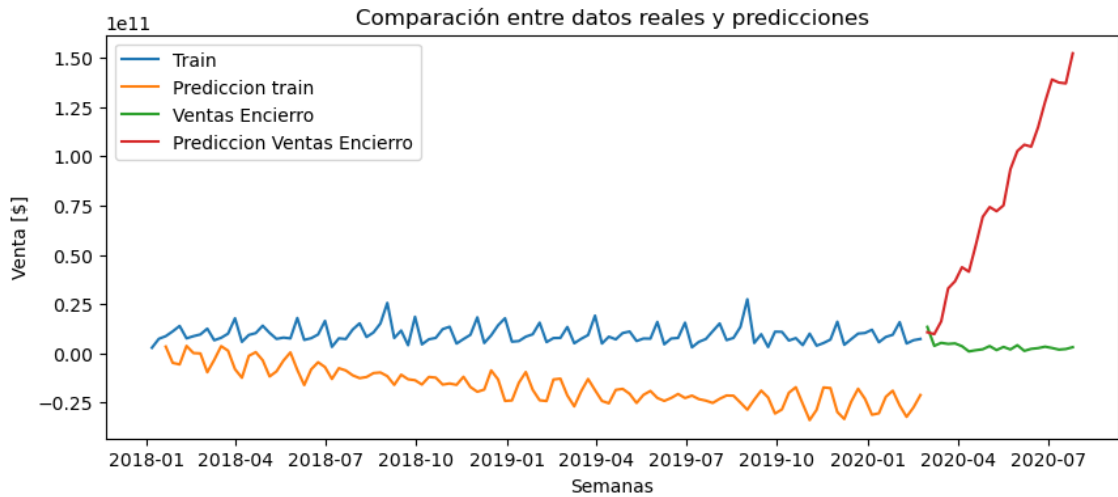
	AIC	BIC	FPE	HQIC
0	89.87	90.08	1.075e+39	89.96
1	87.10	88.99*	6.737e+37	87.86
2	85.78	89.34	1.835e+37	87.22
3	84.94	90.18	8.476e+36	87.06
4	84.43	91.35	5.740e+36	87.23
5	84.21	92.81	5.684e+36	87.69
6	83.82	94.10	5.441e+36	87.98
7	83.27	95.22	5.270e+36	88.11
8	82.01	95.64	3.313e+36	87.52
9	80.32	95.62	2.095e+36	86.51
10	78.04	95.03	1.618e+36	84.92
11	70.60*	89.26	4.047e+34*	78.15*

Una vez elegido el orden del modelo, se continúa con tu entrenamiento y posterior, cálculo de causalidad de Granger. Para ello, se elabora una matriz de causalidad de granger, desde donde se obtiene que, de manera individual, la UF, los depositos_captaciones_plazo y el dolar_observado no son granger causa del precio_de_lista, por lo que se sugiere quitar estas variables del set de datos. En cambio, al realizar el test de manera agrupada, UF, depositos_captaciones_plazo y el dolar_observado obtienen un p-value de 0,000 en todas las pruebas, por lo que sí son Granger causa de todas las otras variables.

	precio_de_lista_x	UF_x	M1_x	cc_depositos_ahorros_vista_x	depositos_captaciones_plazo_x	depositos_ahorro_plazo_x	TPM_x	dolar_observado_x
precio_de_lista_y	1.0000	0.4964	0.0000	0.0000	0.1963	0.0000	0.0060	0.4083
UF_y	0.0072	1.0000	0.1585	0.2276	0.0642	0.0018	0.1315	0.0639
M1_y	0.0000	0.0045	1.0000	0.0001	0.0195	0.0000	0.2874	0.3042
cc_depositos_ahorros_vista_y	0.0000	0.0076	0.0000	1.0000	0.0160	0.0002	0.3548	0.2896
depositos_captaciones_plazo_y	0.2378	0.4681	0.1798	0.1925	1.0000	0.0966	0.1187	0.0022
depositos_ahorro_plazo_y	0.0000	0.0000	0.0000	0.0000	0.0051	1.0000	0.0607	0.0078
TPM_y	0.0259	0.7075	0.0690	0.0874	0.0582	0.2447	1.0000	0.2008
dolar_observado_y	0.0189	0.2836	0.4589	0.3382	0.0917	0.0018	0.1047	1.0000

A continuación, se revisan las autocorrelaciones de los residuos de todas las variables de donde se obtiene que todos los rezagos son no significativos o se encuentran muy cerca de ser cero.

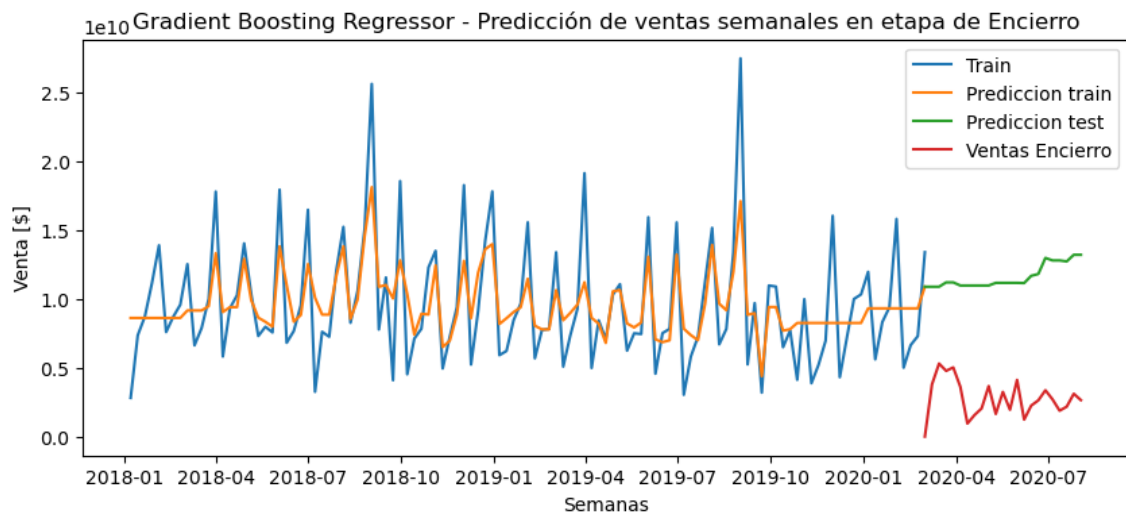
Al realizar las predicciones correspondientes a cada variable y graficando las ventas reales versus las ventas predichas se observa una línea recta creciente completamente alejada de la realidad. La interpretación de este resultado podría ser por la naturaleza creciente de todas las otras variables endógenas del modelo que están afectando a la variable precio_de_lista. Aun si fuera por esto es extraño el comportamiento que se observa al intentar predecir los datos de entrenamiento donde se obtienen hasta valores negativos de ventas semanales, lo cual no es posible. De esta manera se encuentra un resultado difícil de aplicar en la realidad por lo que se rechaza aplicar este modelo.



5.11. Modelo gradient boosting regressor

Para este modelo se utilizó las variables mencionadas al inicio de este trabajo (UF, M1, cc_depositos_ahorros_vista, depositos_captaciones_plazo, depositos_ahorro_plazo, TPM, dolar_observado, días_retiro1, días_retiro2, días_retiro3, retiros_activos, días_conviven_retiros, year, month, week).

Al observar las series de tiempo de entrenamiento (Train) versus la Predicción de train se observa una falta de ajuste a los cambios semanales, donde además existe un cambio repentino y falta de ajuste a la curva original a partir de octubre de 2019. Conociendo de antemano lo sucedido en esas fechas, a partir del 18 de octubre en adelante, es que podemos interpretar que el modelo no logra captar lo sucedido con las variables entregadas y le falta información como para poder realizar una predicción cercana a esos datos. En particular esto afecta al resto de la predicción que ocurre en la etapa de Encierro. Afortunadamente se logra obtener un resultado de predicción bastante conservador para la etapa de Encierro, entre marzo y julio de 2020.



En promedio, mensualmente se habrían dejado de ganar uno 35 mil millones de pesos por efectos de las restricciones de movilidad.

Mes	Venta real [\$]	Predicción [\$]	Diferencia	YoY real	YoY pred
Marzo 2020	32.257.943.292	55.007.282.690	22.749.339.398	-0,22	0,32
Abril 2020	8.163.866.595	43.856.877.687	35.693.011.092	-0,80	0,09
Mayo 2020	14.607.596.510	55.590.074.255	40.982.477.745	-0,62	0,44
Junio 2020	9.475.336.177	47.575.875.446	38.100.539.269	-0,73	0,33
Julio 2020	9.857.900.344	51.476.135.400	41.618.235.056	-0,76	0,26

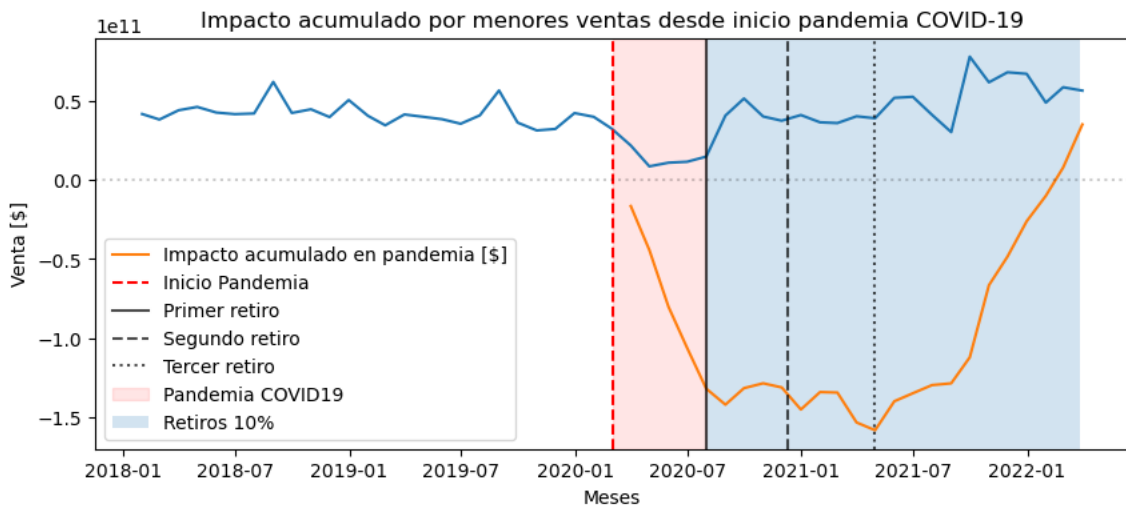
Las métricas de YoY pred calculadas se alejan completamente de los números calculados para meses anteriores. Por otra parte, el comportamiento obtenido con el modelo difiere mucho del comportamiento esperado para ventas semanales y además con la historia. Con una visión más de negocio, este modelo no convence como para continuar con un análisis.

El hallazgo en la predicción de los datos de entrenamiento a partir de octubre de 2019, abre la puerta a investigar los efectos y consecuencias de lo ocurrido a partir del 18 de octubre. Dado que esto afectó a muchos comercios durante esa época es que no se desea continuar y se deja como una oportunidad para explorar más adelante.

5.12. Elección del mejor modelo

En base a los modelos explorados, el modelo ARIMA (4,1,13) resultó ser el único que permitió obtener un proceso convincente y resultados conservadores como para descubrir el impacto de la pandemia COVID-19 y posteriores retiros provisionales del 10%. Por lo tanto, el impacto de la restricción de movilidad provocada por decretar Pandemia se puede estimar de unos 26 mil millones al mes, entre marzo y julio de 2020. Los retiros por su parte aceleraron la recuperación de La empresa del rubro automotriz, dando un respiro frente a los meses anteriores y logrando una recuperación a partir de febrero de 2021. Donde aquí juega un papel importante el vencimiento del primer y segundo retiro provisionales, sumado a un tercer retiro (sin condiciones), facilitando la compra de autos.

Considerando el período desde el inicio de la pandemia COVID-19 se puede medir el impacto acumulado por las menores ventas y es posible observar como el primer retiro provoca un freno en la caída de la curva, dando las primeras luces de recuperación. Con la llegada del segundo retiro este no alcanza a tener el mismo efecto que su antecesor, probablemente dado por las condiciones particulares de este retiro. La mayor pérdida acumulada se obtiene en abril de 2020 alcanzando los -158 mil millones o lo que serían unos 11.300 millones al mes comparando ventas reales contra predicción de ventas.



A partir del tercer retiro se ve un cambio en la tendencia, el primer retiro llegaba a su fin luego de cumplir un año, algunas personas no habían retirado sus ahorros hasta ahora y dependiendo del fondo donde se hayan

encontrado aun hacían crecer sus ahorros, pudiendo retirar un monto mayor que hace 1 año; para algunos, parecía la oportunidad para acceder a un auto. De ahí en adelante la curva comienza a recuperar la menor venta, logrando ganancias a partir de febrero de 2022.

6. Conclusiones

El modelo ARIMA resulta una buena herramienta para estimar el impacto de la pandemia y los retiros. Se puede observar que el modelo ARIMA se ajusta de buena manera a los cambios, situación que no se pudo observar con los otros modelos. Si bien el modelo ARIMA (4,1,13) no es la única solución para predecir las ventas del periodo analizado, si ofrece una buena forma de. De esta manera se puede concluir que el impacto de la pandemia fue importante y no tan solo por las ventas sino en como afecta a todo un ecosistema extracción de materias primas, fábricas de repuestos, importaciones y exportaciones, que sin duda generan un retraso en la recuperación de las empresas y países. En el caso de la empresa del rubro automotriz el impacto fue importante, alcanzando los casi 26 mil millones de pesos por mes, durante marzo y julio de 2020 para luego lograr una recuperación gracias a los retiros. Por lo tanto, la pandemia tiene una relación negativa con las ventas de la empresa del rubro automotriz y los retiros, tuvieron un efecto positivo pero rezagado, ya que su impacto se ve más fuerte al momento de llegar un tercer retiro, probablemente motivado por la caducidad del primer retiro previsional y por un ahorro provocado por esa espera.

A pesar de haber explorado algunas soluciones con otros modelos, estas no fueron satisfactorias y no ofrecían ventajas frente a una serie de tiempo univariada. Y es quizás en esta “simplicidad” de la serie de tiempo univariada que radica su riqueza, esto sin dejar de lado lo complejo que puede ser dar con un buen modelo de serie de tiempo. Siempre está la tentación de poder agregar nuevas variables para explicar un comportamiento pero esto trae consigo aumentar la complejidad del modelamiento y cálculo, pudiendo retrasar una entrega a tiempo en un trabajo o bien perdiéndose dentro de un modelo con el afán de dar con un resultado sin cuestionar si tiene sentido o no para el negocio.

Es posible utilizar Machine Learning, no solo como una herramienta para realizar predicciones a futuro, sino que también como una herramienta para entender comportamientos pasados, entender el impacto de ciertas variables sobre otras, aprovechando las ventajas que tiene para dar una explicación clara y directa a ciertos efectos y para entender comportamientos inesperados.

A futuro sería bueno profundizar en el análisis por región y de esa manera identificar si existe alguna región donde los impactos del encierro y los retiros fueron proporcionalmente mayores que a nivel nacional, como también continuar con análisis de modelos de redes neuronales que permitan explorar si existen mejores resultados de predicción. El hallazgo en la predicción del modelo de Gradient Boosting Regressor donde a partir de octubre de 2019 presenta dificultades para ajustarse a los datos y no deja de ser importante

haciendo posible que requiera un entendimiento más profundo, quizás otro conjunto de datos que permita entender de mejor manera el impacto que tuvieron los hechos delictivos en el dinamismo de la economía del país.

Bibliografía

Buruzs, A., & Torma, A. (2018). A Review on the Outlook of the Circular Economy in the automotive industry. *International Journal of Environmental and Ecological Engineering*, 11(6), 576–580. <http://waset.org/Publication/10008699>

Dooley, J. (2020, April 21). *How COVID-19 has changed the automotive business, and what Volvo is doing about it*. ClickZ. <https://www.clickz.com/how-covid-19-has-dramatically-changed-business-at-volvo-and-what-theyre-doing-about-it/261297/>

Ktrakazas, C., Michelaraki, E., Sekadakis, M., Ziakopoulos, A., Kontaxi, A., & Yannis, G. (2021). Identifying the impact of the COVID-19 pandemic on driving behavior using naturalistic driving data and time series forecasting. *Journal of Safety Research*, 78, 189–202. <https://doi.org/10.1016/j.jsr.2021.04.007>

Ramírez, J., Alarcón, J., Calzada, G., & Ponce, H. (2021). Mexican automotive industry sales behavior during the COVID-19 pandemic. In *Advances in Soft Computing* (pp. 265–276). Springer International Publishing.

The impact of the COVID-19 pandemic on a selected automotive industry – A case study of Poland. (2020, November 26). Ibima.org; International Business Information Management Association (IBIMA). <https://ibima.org/accepted-paper/the-impact-of-the-covid-19-pandemic-on-a-selected-automotive-industry-a-case-study-of-poland/>