



Universidad del Desarrollo

Facultad De Ingeniería

**ANÁLISIS Y MODELAMIENTO DE LOS MÁRGENES POR NEGOCIOS DE LA
EMPRESA DERCO MEDIANTE SERIES TEMPORALES Y REDES
NEURONALES RECURRENTE LSTM**

POR:

PATRICIO SOTO SANDOVAL & JONATHAN ÁLVAREZ VALLEJOS

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

PROFESOR GUÍA:

SR. ALONSO ASTROZA & SR. CRISTIAN CANDIA

DICIEMBRE 2022



SANTIAGO

AGRADECIMIENTOS

Agradezco a Daniela mi esposa y Maximiliano mi hijo, por darme el espacio, tiempo y motivación para estudiar largas jornadas y llevar a buen término cada uno de los ramos. Igualmente agradezco a Hermógenes, por no molestarse en las ocasiones que dedicaba todo el día a completar algún trabajo. Y no puedo dejar de lado a la gran comunidad de personas que comparten su conocimiento, de forma muchas veces anónima, y sin esperar ninguna retribución, aprendí muchísimo con ustedes.

Jonathan J. Álvarez Vallejos

Agradezco a Dios y mi familia, Elizabeth y Dominga, por la oportunidad de seguir creciendo profesional y personalmente, por darme el tiempo y la motivación necesaria para estar horas en el computador aprendiendo y desarrollándome, en lo técnico y en lo humano.

Patricio A. Soto Sandoval



TABLA DE CONTENIDO

LISTADO DE FIGURAS	1
LISTADO DE TABLAS	3
RESUMEN	4
INTRODUCCIÓN	5
TRABAJOS RELACIONADOS	6
Sobre el crecimiento del parque automotriz en Chile:	6
Sobre el concepto de postventa:	8
Sobre técnicas, métodos y modelos de machine learning:	8
HIPÓTESIS Y OBJETIVOS	15
Hipótesis:	15
Objetivos generales:	15
Objetivos específicos:	15
DATOS Y METODOLOGÍA	16
DATOS	16
Patentes:	16
Ventas Derco:	18
METODOLOGÍA	26
RESULTADOS	32
Resultados con Prophet:	32
Resultados con Modelos LSTM:	39
CONCLUSIONES	44
BIBLIOGRAFÍA	45



LISTADO DE FIGURAS

- **Figura 1:** Evolución del parque vehicular motorizado en comunas capitales regionales y otras comunas.
- **Figura 2:** Módulo de repetición en una RNN estándar (contiene una sola capa).
- **Figura 3:** Módulo de repetición en un LSTM (contiene cuatro capas que interactúan).
- **Figura 4:** Evolución del Parque Automotriz 2016 - 2021.
- **Figura 5:** Evolución del Parque Automotriz 2016 - 2021. Marcas DERCO v/s Otras Marcas.
- **Figura 6:** Patentes por año de permiso y año del vehículo.
- **Figura 7:** Evolución diaria del margen por negocio.
- **Figura 8:** Boxplots de márgenes por negocio.
- **Figura 9:** Boxplots y gráficos de densidad de márgenes por negocio (post limpieza de datos).
- **Figura 10:** Evolución diaria del margen por negocio (post limpieza de datos).
- **Figura 11:** Evolución anual del margen por negocio.
- **Figura 12:** Evolución mensual del margen por negocio.
- **Figura 13:** Evolución semanal del margen por negocio.
- **Figura 14:** Correlaciones entre márgenes por negocio.
- **Figura 15:** Chord Plot Frecuencia de compra entre negocios.
- **Figura 16:** Función de pérdida (mse) de entrenamiento modelo LSTM Nuevos.
- **Figura 17:** Gráfico sin regresores adicionales (Prophet - nuevos).
- **Figura 18:** Gráfico en Margen Retail con regresor adicional de márgenes de Autoplanet (Prophet - nuevos).
- **Figura 19:** Gráfico sin regresores adicionales (Prophet - sstt).
- **Figura 20:** Gráfico en Margen Retail con Margen Retail de Nuevos como regresor adicional (Prophet - sstt).
- **Figura 21:** Gráfico en Margen Retail con Margen Retail de Nuevos, Amicar y Mesón como regresores adicionales (Prophet - sstt).



- **Figura 22:** *Gráfico sin regresores adicionales (Prophet - meson).*
- **Figura 23:** *Gráfico Margen Retail con Margen Retail de Nuevos como regresor adicional (Prophet - meson).*
- **Figura 24:** *Gráfico Margen Retail con Margen Retail de Nuevos y Amicar como regresores adicionales (Prophet - meson).*
- **Figura 25:** *Gráfico Margen Retail con Margen Retail de Nuevos, Amicar y Servicio Técnico como regresores adicionales (Prophet - meson).*
- **Figura 26:** *Gráfico sin regresores adicionales (Prophet - autoplanet).*
- **Figura 27:** *Gráfico Margen Retail con Margen Retail de nuevos como Regresor Adicional (Prophet - autoplanet).*
- **Figura 28:** *Gráfico sin regresores adicionales (Prophet - amicar).*
- **Figura 29:** *Gráfico Margen Retail con Margen Retail de nuevos como regresor adicional (Prophet - amicar).*
- **Figura 30:** *Entrenamiento modelo LSTM margen nuevos en base logarítmica.*
- **Figura 31:** *Gráfico modelo LSTM solo nuevos.*
- **Figura 32:** *Gráfico modelo LSTM multivariable nuevos (sstt como variable paralela).*
- **Figura 33:** *Gráfico modelo LSTM solo sstt.*
- **Figura 34:** *Gráfico modelo LSTM multivariable sstt (nuevos como variable paralela).*
- **Figura 35:** *Gráfico modelo LSTM solo meson.*
- **Figura 36:** *Gráfico modelo LSTM multivariable meson (nuevos como variable paralela).*
- **Figura 37:** *Gráfico modelo LSTM solo autoplanet.*
- **Figura 38:** *Gráfico modelo LSTM multivariable autoplanet (nuevos como variable paralela).*
- **Figura 39:** *Gráfico modelo LSTM solo amicar.*
- **Figura 40:** *Gráfico modelo LSTM multivariable amicar (nuevos como variable paralela).*



LISTADO DE TABLAS

- **Tabla 1:** *Fechas de corte para normalización series temporales por negocio.*
- **Tabla 2:** *Estacionaridad por negocio. Método “Dickey-Fuller”.*
- **Tabla 3:** *Resultados análisis de causalidad de “Granger” desde lag_1 a lag_4 .*
- **Tabla 4:** *Revisión general de resultados de análisis de causalidad de “Granger”.*
- **Tabla 5:** *Evaluación resultados MAPE (mean absolute percentage error).*
- **Tabla 6:** *Resultados hiper parámetros por validación cruzada.*
- **Tabla 7:** *Resultados de Métricas de la Predicción (Prophet - nuevos).*
- **Tabla 8:** *Resultados de Métricas de la Predicción (Prophet - sstt).*
- **Tabla 9:** *Resultados de Métricas de la Predicción (Prophet - meson).*
- **Tabla 10:** *Resultados de Métricas de la Predicción (Prophet - autoplanet).*
- **Tabla 11:** *Resultados de Métricas de la Predicción (Prophet - amicar).*
- **Tabla 12:** *Resultados de Métricas modelos LSTM.*



RESUMEN

El objetivo de este proyecto es analizar y modelar los márgenes por negocio de la empresa DERCO mediante series temporales y redes neuronales recurrentes. Bajo la hipótesis a comprobar respecto a la relación en el aumento del margen de vehículos nuevos sobre el margen de ganancia de las unidades de negocio restantes. Se exploraron los registros de venta de los últimos 12 años, logrando plantear correlaciones e identificar patrones para el posterior modelamiento. Los modelos de las series fueron realizados por medio de Prophet y RNN LSTM. Se obtuvo como principal resultado que los modelos elaborados con Prophet, presentan un mejor desempeño para el caso en estudio, considerando las limitaciones de tiempo, recursos e información disponible. De ello podemos concluir que para este caso particular, la influencia del margen de venta de los vehículos nuevos es muy relevante para el margen de las otras unidades de negocio. Como mejora al análisis y modelamiento se propone explorar el detalle de venta (productos) por cliente y efectuar nuevos modelos de redes neuronales probando otras arquitecturas.



INTRODUCCIÓN

El presente proyecto de grado tiene por objetivo el análisis y modelamiento de los márgenes por negocios de la empresa DERCO mediante series temporales clásicas y redes neuronales LSTM. Supeditado a la hipótesis de comprobar relaciones de los distintos negocios en cuanto a su margen de ganancia.

DERCO es una empresa chilena que opera en cuatro países de latinoamérica, principalmente en los rubros automotriz, maquinaria, repuestos y accesorios y la compra-venta de automóviles usados. Los negocios en estudio corresponden a “amicar (crédito automotriz), autoplanet (repuestos y accesorios), meson (repuestos y accesorios), nuevos (venta de vehículos nuevos) y sstt (servicios técnicos)”.

Las series temporales se analizarán con Prophet, herramienta desarrollada por Facebook que se caracteriza por ser un método potente y optimizado para trabajar con series empresariales por su buen manejo de las tendencias y estacionalidades. Su algoritmo subyacente es una generalización de un modelo de regresión aditiva. Por otro lado, contrastaremos los resultados con un modelo de machine learning basado en redes neuronales recurrentes (RNN), en particular con una red Long Short Term Memory (LSTM), dada su probada capacidad para aprender secuencias a largo plazo.

El interés específico del estudio surge debido a la exploración inicial de los datos y la confirmación de algunas correlaciones previamente establecidas de acuerdo a la naturaleza del negocio. Dado el contexto empresarial, como equipo, estimamos que el estudio y análisis de los márgenes de las distintas unidades negocio será de utilidad y provecho para la empresa DERCO. Por lo tanto, la elaboración de modelos predictivos fundamentados en series de tiempo, surge como una alternativa atractiva.

Para establecer un marco de investigación se comienza con la revisión de trabajos relacionados al tema en estudio, para luego dar paso a las hipótesis y objetivos del proyecto. Posteriormente se presenta la exploración detallada de los datos, orientada a continuar con la metodología que incluye la propuesta de los modelos a contrastar. Finalmente se exponen los resultados y conclusiones, sin dejar de lado las limitaciones y sugerencias de trabajos futuros.



TRABAJOS RELACIONADOS

A continuación se exponen extractos relevantes sobre trabajos relacionados al tema en estudio. Como también, se resume parte de la documentación revisada para conformar el marco teórico del presente proyecto de título.

Sobre el crecimiento del parque automotriz en Chile:

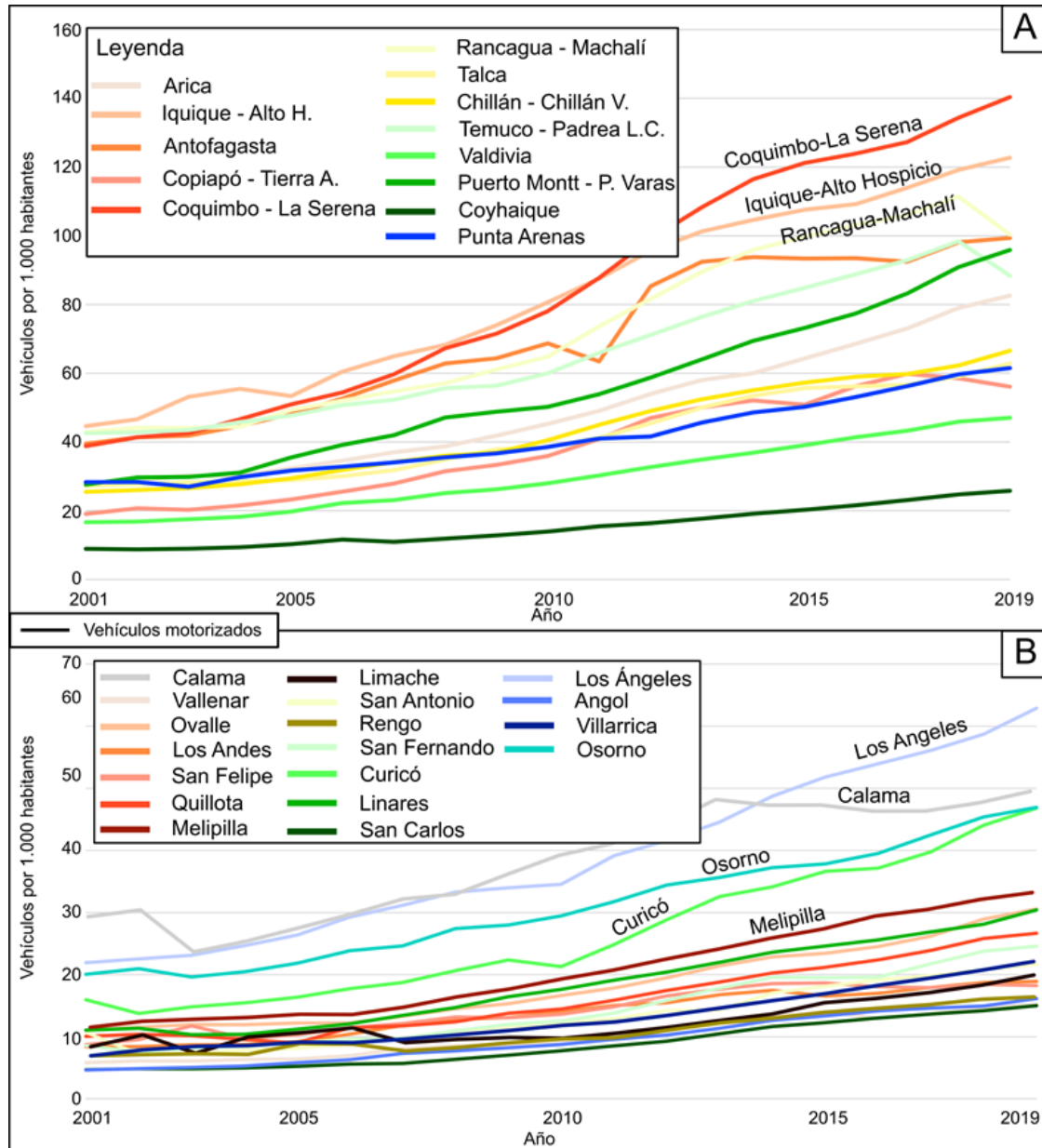
Desde la crisis financiera internacional de 2008 - 2009 la industria global se concentra en tres regiones: América del Norte, la Unión Europea y Asia. Destaca que los segmentos de mayor apropiación de valor agregado se concentran en cuatro países: EU, Alemania, Japón y Corea. China aparece como un quinto actor en importancia, pues desde 2009 figura como la principal productora de vehículos en el mundo (en 2016 produjo más de 28 millones de unidades o casi el 30% de la producción mundial). Destaca que desde 2009 se encuentran entre las 50 principales automotrices mundiales, 21 empresas chinas: Changan, BAIC, Dongfeng Motor, FAW, Chery, BYD, SAIC, Geely, etc... Lo anterior es relevante pues en 2000 las empresas chinas no figuraban en el top 50. (Marcial Flores & Ortiz Velázquez, 2018)

Chile ha sido parte del proceso de globalización en cuanto al movimiento de capitales, tratados de libre comercio y caída de aranceles, e incremento en tecnología, factores que han impactado el mercado del automóvil, el cual ha experimentado un aumento considerable en los últimos años e incluso se estima que con la pandemia por covid-19 tal proceso se haya acelerado aún más. El explosivo aumento del parque vehicular en los últimos años, queda plasmado en los números del Instituto Nacional de Estadística de Chile (INE). En efecto, según esta institución, el parque vehicular motorizado se cuadruplicó en 19 años, pasando de 1.314.129 vehículos motorizados en el año 2000 a 5.599.733 en 2019. De acuerdo con la Cámara de Comercio Automotriz de Chile, en el 2019, en el país existen prácticamente 50 marcas operando, lo que habla de una amplia gama y cuya concentración de mercado se desarrolla en torno a 20 de ellas aproximadamente. (Maturana et al., 2022)



Figura 1: Evolución del parque vehicular motorizado en comunas capitales regionales (A) y otras comunas (B). Fuente: Ciudades intermedias y parque automotriz en Chile.

¿En jaque la sustentabilidad territorial? ÍCONOS Revista de Ciencias Sociales.





Sobre el concepto de postventa:

La postventa es un conjunto de actividades relacionadas entre sí, orientadas hacia un objetivo único, que es crear un estado de ánimo positivo en el cliente por haber excedido sus expectativas, un estado de ánimo que derive en la fidelización del mismo. El objetivo final para la organización es lograr mantener cada vez un mayor número de clientes fidelizados para complementar las estrategias de captación de clientes que se utilicen. Desde este enfoque el cliente es el centro de actividades concernientes a la garantía, formación técnica, índices de satisfacción, gestión de inventarios, instalaciones y venta de accesorios. (Parra Caballero, 2016)

Así, los concesionarios deberían concentrar sus esfuerzos en satisfacer a sus clientes con el objetivo de guiarlos a que recomienden los servicios recibidos a posibles y potenciales consumidores y a que repitan su compra. Esto se lograría con el mejoramiento de los servicios de postventa. Las actividades claves en el servicio postventa se pueden resumir en tres: la atención y relación con el cliente, la asistencia técnica en el taller de servicios y, por último, el adecuado suministro de repuestos. (Erika et al., 2013)

Sobre técnicas, métodos y modelos de machine learning:

En el pasado, las pocas empresas que sustentaban sus decisiones de negocios en datos, generalmente utilizaban modelos predictivos estadísticos estáticos. Habitualmente utilizando data histórica, con cero capacidad de aprender con una nueva observación empírica. Hoy los equipos comerciales utilizan el aprendizaje automático para predecir la propensión de compra en todos los canales, hacer recomendaciones personalizadas a clientes, predecir la lealtad a largo plazo y anticipar los posibles riesgos de los proveedores y compradores.

El beneficio de machine learning para escalar a través del amplio espectro de gestión de contratos, servicio al cliente, finanzas, legal, ventas, calidad, precios y desafíos de producción... Es atribuible a su capacidad de aprender y mejorar continuamente. Los algoritmos de aprendizaje automático son de naturaleza iterativa, aprenden



constantemente y buscan optimizar los resultados. Cada vez que se realiza un error de cálculo, los algoritmos de aprendizaje corrigen el error y comienzan otra iteración del análisis de datos. Estos cálculos ocurren en milisegundos, lo que hace que el machine learning sea excepcionalmente eficiente para optimizar las decisiones y predecir los resultados. El proceso de complementar la econometría con el aprendizaje automático, en América Latina, viene bastante lento. Las empresas aún no se han dado cuenta que el proceso de acceder a este tipo de factor diferencial, determinante de ventaja competitiva, es totalmente viable, y hasta cierto punto, crítico para el éxito en un mundo cada vez más determinado por la data. (Vallarino, 2016)

El uso de métodos de modelado no lineal (a diferencia de los modelos lineales generalizados habituales derivados del modelado estadístico) y la extracción de conocimiento a partir de datos, se basa en la idea fundamental de que se pueden derivar modelos a partir de datos con la ayuda de algoritmos y que este proceso de modelado puede ejecutarse automáticamente en su mayor parte, porque el algoritmo "hace el trabajo". (Hofmann et al., 2017)

En el nivel más general, los algoritmos de aprendizaje automático (ML) se pueden subdividir en dos categorías: supervisados y no supervisados, dependiendo de si el algoritmo respectivo requiere o no que se especifique una variable de destino.

Aparte de las variables de entrada (predictores), los algoritmos de aprendizaje supervisado también requieren los valores objetivo conocidos (etiquetas). Las etiquetas, es decir, las designaciones correctas, para los datos normalmente se asignan manualmente. Este conjunto correcto de las variables de entrada y su correcta clasificación constituyen un conjunto de datos de entrenamiento. El aprendizaje supervisado se usa principalmente para predecir valores numéricos (regresión) y con fines de clasificación (predecir la clase apropiada).

Los algoritmos de aprendizaje no supervisados no se centran en variables de destino individuales, sino que tienen el objetivo de caracterizar un conjunto de datos en general. Los algoritmos de ML no supervisados se utilizan a menudo para agrupar conjuntos de datos, es decir, para identificar relaciones entre puntos de datos individuales (que pueden



constar de cualquier número de atributos) y agruparlos en grupos. Los algoritmos, a su vez, pueden utilizarse como entrada para métodos supervisados. Los ejemplos de aprendizaje no supervisado incluyen la formación de grupos de clientes en función de su comportamiento de compra o datos demográficos, o la agrupación de series temporales para agrupar millones de series que antes no eran obvias. (Michalewicz et al., 1997)

El Deep learning ha demostrado ser de gran ayuda en diferentes ámbitos, como el reconocimiento de objetos, la clasificación de imágenes, el reconocimiento del lenguaje y diferentes sectores financieros (detección de fraude, fijación de precios, creación de carteras, gestión de riesgos...). Los datos de serie cronológica desempeñan un papel importante en muchos de estos problemas, lo que demuestra la capacidad que tienen las redes neuronales de tratar este tipo de información. En particular, a las redes Long Short Term Memory (LSTM) se les da muy bien aprender de secuencias, lo que las hace muy atractivas en este tipo de contexto. Las LSTM son un tipo especial de redes recurrentes. La característica principal de las redes recurrentes es que la información puede persistir introduciendo bucles en el diagrama de la red, por lo que, básicamente, pueden «recordar» estados previos y utilizar esta información para decidir cuál será el siguiente. Esta característica las hace muy adecuadas para manejar series cronológicas. Mientras las redes recurrentes estándar pueden modelar dependencias a corto plazo (es decir, relaciones cercanas en la serie cronológica), las LSTM pueden aprender dependencias largas, por lo que se podría decir que tienen una «memoria» a más largo plazo. (Garzón, 2018)

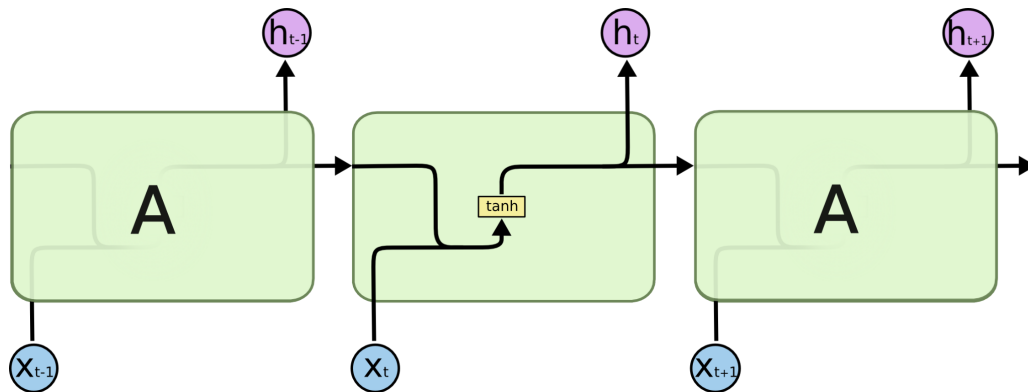
Una Red Recurrente convencional tiene una memoria de corto plazo, resultado de las múltiples transformaciones que sufre el estado oculto a su paso por la red. Las Redes LSTM resuelven este problema usando el concepto de la celda de estado (que es la memoria de la red). En estas Redes LSTM es posible agregar o eliminar información a esta celda de estado usando una serie de compuertas (forget, update y output) que permiten discriminar entre la información relevante e irrelevante. Esto hace que las Redes LSTM sean mucho más robustas que las redes recurrentes convencionales, pues como su nombre lo indica poseen una memoria tanto de largo (long) como de corto (short) plazo. Todo esto explica el auge que actualmente tienen las Redes LSTM para el



desarrollo de diferentes aplicaciones, y son realmente la arquitectura a usar cuando hablamos de Redes Neuronales Recurrentes. (Sotaquirá, 2019)

Las redes de memoria a largo plazo (normalmente llamadas "LSTM") son un tipo especial de RNN, capaces de aprender dependencias a largo plazo. Fueron introducidas por Hochreiter y Schmidhuber en 1997, y fueron refinadas y popularizadas por muchas personas en trabajos posteriores. Funcionan extremadamente bien en una gran variedad de problemas, y ahora son ampliamente utilizadas. Las LSTM están diseñadas explícitamente para evitar el problema de la dependencia a largo plazo. Recordar información durante largos periodos de tiempo es prácticamente su comportamiento por defecto. Todas las redes neuronales recurrentes tienen la forma de una cadena de módulos de red neuronal que se repiten. En las RNN estándar, este módulo de repetición tendrá una estructura muy simple, como una sola capa \tanh .

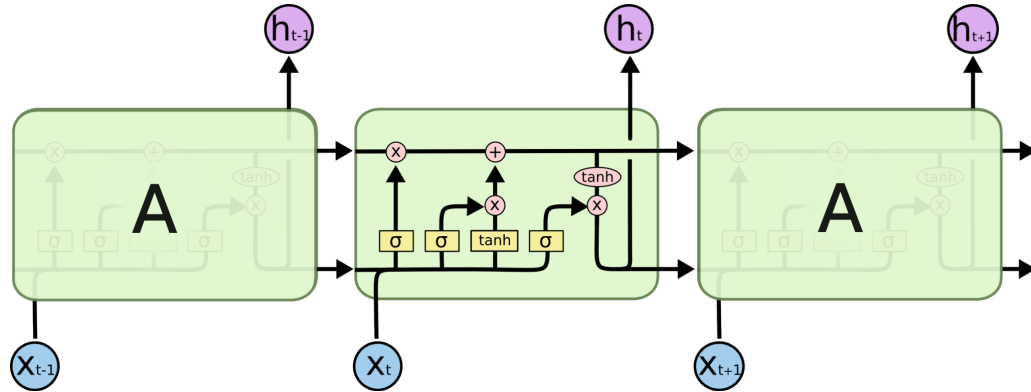
Figura 2: Módulo de repetición en una RNN estándar (contiene una sola capa). Fuente: *Understanding LSTM Networks*



Las LSTM también tienen esta estructura en cadena, pero el módulo de repetición tiene una estructura diferente. En lugar de tener una sola capa de red neuronal, hay cuatro, que interactúan de una manera muy especial.



Figura 3: Módulo de repetición en un LSTM (contiene cuatro capas que interactúan).
Fuente: *Understanding LSTM Networks*



En el diagrama anterior, cada línea lleva un vector completo, desde la salida de un nodo hasta las entradas de otros. Los círculos rosas representan operaciones puntuales, como la suma de vectores, mientras que los recuadros amarillos son capas de la red neuronal aprendidas. Las líneas que se fusionan denotan la concatenación, mientras que una línea que se bifurca denota que su contenido se copia y las copias van a lugares diferentes. (Graves, 2015)

Prophet, una herramienta desarrollada por Facebook con el objetivo de hacer lo más fácil posible el análisis de series temporales. Permite encontrar automáticamente la estacionalidad, tendencia y ofrece parámetros fáciles de entender sin necesidad de tener un amplio conocimiento y experiencia con series temporales.

Prophet presenta un modelo aditivo con las siguientes componentes:

$$y(t) = g(t) + s(t) + h(t) + \epsilon$$

Siendo:

- $g(t)$: Que modeliza la tendencia del modelo, la cual describe el comportamiento a largo plazo.
- $s(t)$: Que modeliza la estacionalidad del modelo mediante series de Fourier.



- $h(t)$: Que modeliza el efecto de las vacaciones o eventos que tienen impacto en nuestra serie.
- ϵ : Que representa el término del error irreducible.

La versatilidad de Prophet permite pronosticar un gran número y una variedad de problemas de series temporales, por lo que se ha convertido en una herramienta muy a tener en cuenta. (Ángel RC, 2020)

El objetivo de este paquete es dar a los usuarios una herramienta potente y de fácil uso para predecir resultados de negocio. Está muy bien documentado en su web, por lo que es fácil aprender a usarlo y ajustarlo. El algoritmo subyacente es una generalización de un modelo de regresión aditivo, que se descompone en tres componentes principales (tendencia, estacionalidad y festivos), más los regresores adicionales. La tendencia y la estacionalidad son dos componentes de las series temporales muy importantes pero difíciles de cuantificar. Prophet hace un gran trabajo capturando ambos. Además, la relación precisión – velocidad de este algoritmo es muy alta. Es un modelo que tiene fácil descomposición, por lo que la interpretabilidad es alta y es fácil extraer los coeficientes de cada una de las componentes del modelo. Así se puede estudiar el impacto de la estacionalidad, tendencia, festivos y otras variables regresoras. Prophet alcanza sus mejores resultados con series estacionarias o series con una tendencia y estacionalidad marcadas. Suele ser peor cuando aumenta la aleatoriedad de la muestra.

Como se ha explicado con anterioridad, el modelo de Prophet se basa en la suma o producto de distintas componentes: la tendencia, las componentes estacionales, la componente de festivos y las componentes del resto de regresores.

La tendencia puede ser:

- Lineal (“linear”): una función continua formada por trozos de rectas que crecen o decrecen entre cada punto de cambio de tendencia (“changepoint”).
- Constante (“flat”): una componente constante que forma una recta horizontal. No cambia la tendencia en ningún changepoint.
- Logística (“logistic”): una función logística que no aplica en nuestros modelos.



El efecto del resto de variables, ya sean estacionales o del resto de regresores, puede ser aditivo o multiplicativo. Esto se traduce en que la componente se suma o se multiplica a la de tendencia. (Alonso, 2022)



HIPÓTESIS Y OBJETIVOS

Hipótesis:

Para efecto de una mejor comprensión de la problemática en estudio, estructuramos la hipótesis a modo de la siguiente pregunta:

¿Si aumenta la venta de vehículos nuevos, entonces, el margen de ganancia de los negocios relacionados a la postventa y venta de accesorios, tendrán un crecimiento positivo?

Objetivos generales:

- Identificar correlaciones entre unidades de negocios de la empresa DERCO, y por medio de estas, elaborar un modelo para determinar la variación en el margen de un negocio en base a otro.
- Determinar cómo se relacionan las distintas unidades de negocio de la misma compañía con respecto a sus respectivos márgenes de ganancia.
- Descubrir patrones en la trayectoria de compra y la trazabilidad del cliente.

Objetivos específicos:

- Comprobar si es que al comprar un vehículo nuevo, el cliente compra accesorios adicionales, lo que por efecto del volumen de ventas, provocaría un crecimiento en el margen del negocio.
- Constatar si es que al subir la venta de vehículos nuevos, aumenta el número de mantenciones programadas y con ello mejorar el margen de postventa.
- Verificar si es que el crecimiento del parque automotriz, aumenta la demanda de repuestos de marcas importadas por DERCO, mejorando con ello la venta y margen de esta unidad de negocio.



DATOS Y METODOLOGÍA

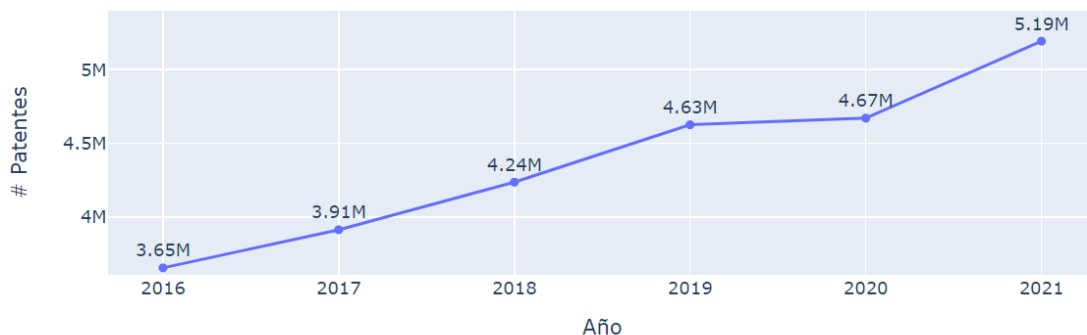
DATOS

Patentes:

Para identificar la venta de vehículos nuevos se analizó el dataset correspondiente al crecimiento del parque automotriz, el cual se encuentra en el archivo “patentes_historico_02.parquet”. El motivo de este análisis se debe a que no se cuenta con la base de datos de venta de otros concesionarios, como para contrastar la información, y a la dificultad de obtener dichos registros.

Los datos analizados muestran una propensión alcista del parque automotriz desde el año 2016 hasta el año 2021, donde se logra apreciar que la tendencia se estanca el año 2020 producto de la pandemia.

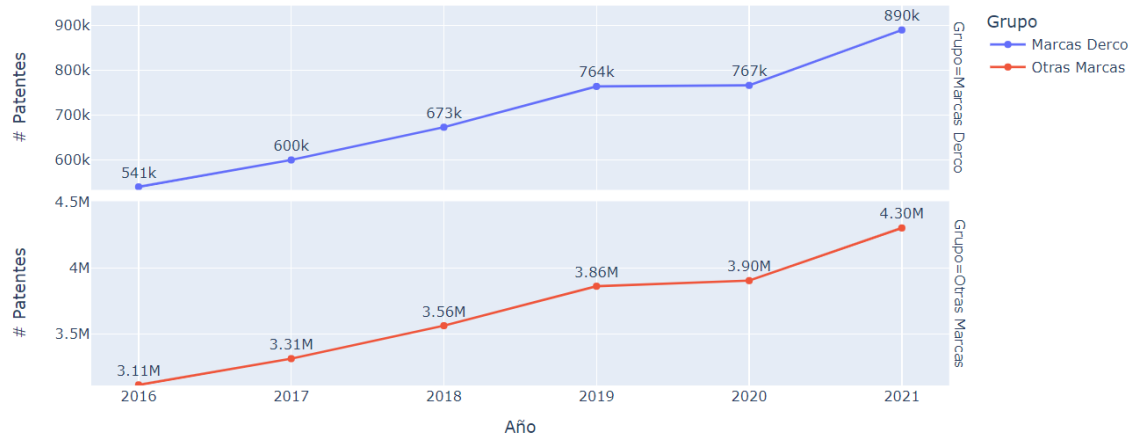
Figura 4: Evolución del Parque Automotriz 2016 - 2021.



Para entender mejor el perímetro a abordar, además, se separó el parque automotriz por vehículos DERCO y otras marcas, con el fin de comparar el comportamiento de estos 2 grupos. De esta forma, mediante una aproximación gráfica, comprobamos que su comportamiento evolutivo es similar. Por lo tanto podríamos suponer que el aumento general de la venta de vehículos, no alteró la tendencia individual de las marcas DERCO.



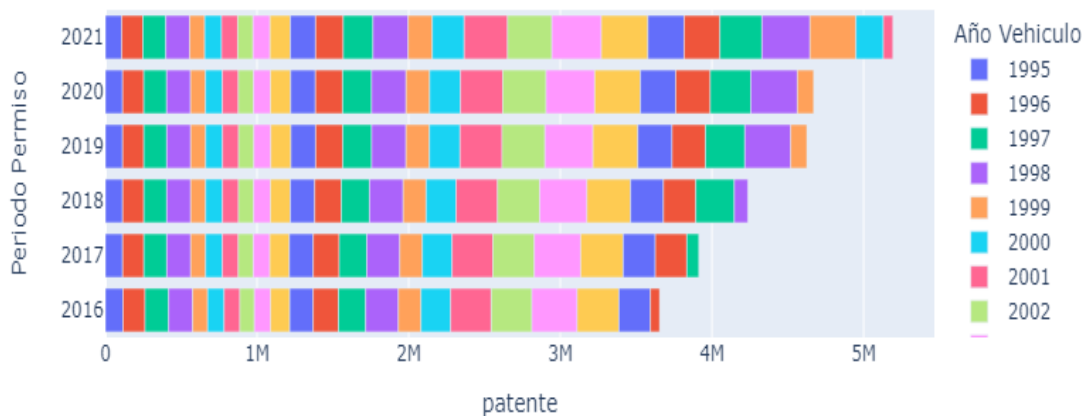
Figura 5: Evolución del Parque Automotriz 2016 - 2021. Marcas DERCO v/s Otras Marcas.



Marca	Suzuki	Mazda	Renault	Great Wall	JAC	Changan	Haval
Cantidad de Patentes	414.922	225.961	121.932	62.577	43.821	39.291	5.373

Adicionalmente se analizó el parque automotriz por año de fabricación y marca de cada vehículo, tanto para DERCO como para el parque total.

Figura 6: Patentes por año de permiso y año del vehículo.





Ventas Derco:

Por otro lado, se analizó el dataset “tabla_journey_javier.csv” compuesto por los registros históricos de los márgenes por venta de la empresa Derco separado por cada negocio, incluyendo fecha de operación y el detalle de clientes por Rut encriptado. Luego de explorar el dataset, y eliminar registros sin información, se conservan casi 12 millones de entradas. Posteriormente, después de una exploración inicial y creación de tablas auxiliares, se mantienen las siguientes columnas:

```
[ 'fecha_transaccion', 'year_month_t', 'negocio', 'marca', 'detalle', 'retail', 'precio_de_lista', 'margen_retail', 'year_venta', 'comuna', 'persona', 'encrypt_rut' ]
```

Para la exploración del margen por negocio se prepara el dataset de tal forma que se pueda observar el detalle del margen retail por cada uno de los negocios en cada fecha de venta registrada. Obteniendo la siguiente estructura:

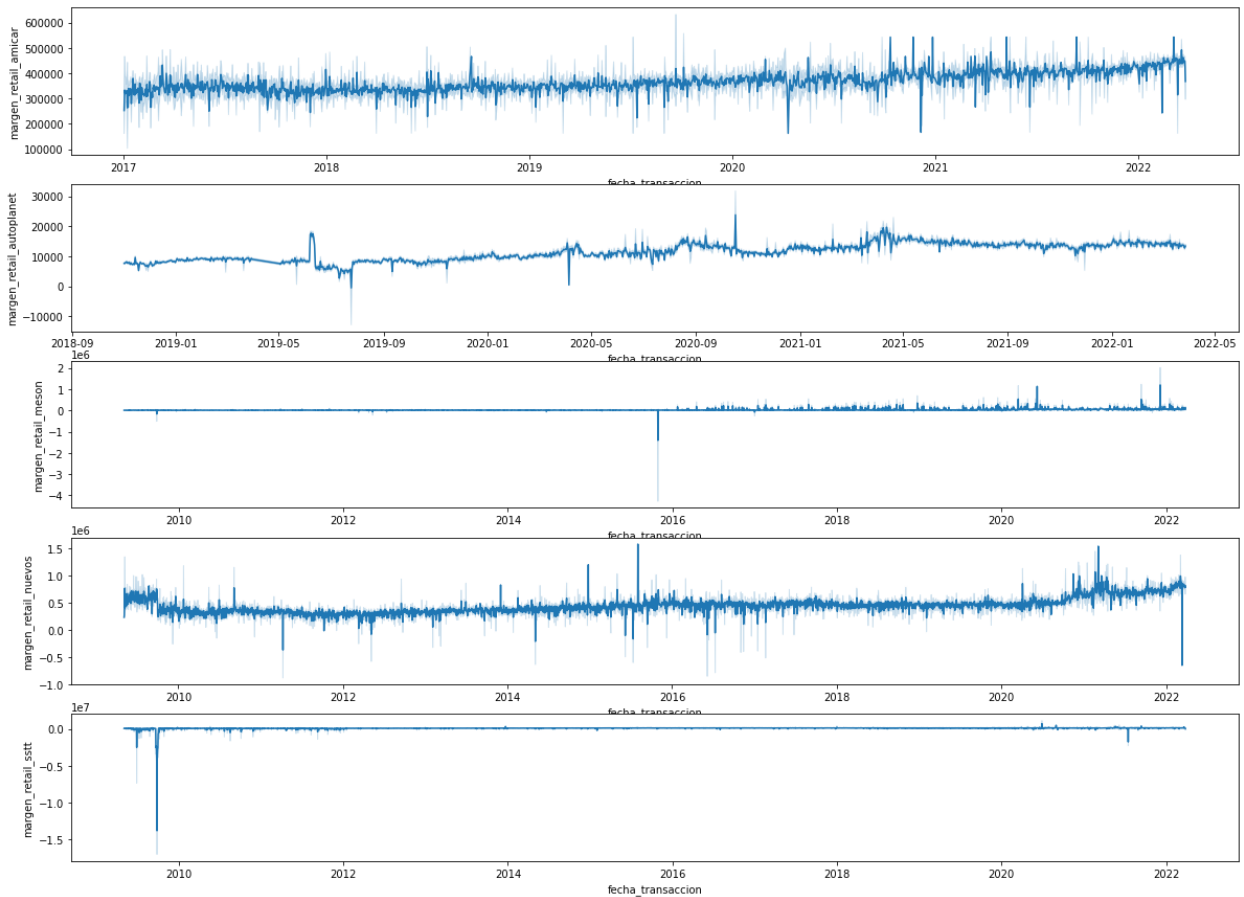
- *fecha_transaccion* 8605185 *non-null datetime64[ns, UTC]*
- *encrypt_rut* 8605185 *non-null object*
- *persona* 8605185 *non-null object*
- *marca* 8605185 *non-null object*
- *margen_retail_amicar* 194371 *non-null float64*
- *margen_retail_autoplanet* 4844085 *non-null float64*
- *margen_retail_meson* 1864695 *non-null float64*
- *margen_retail_nuevos* 596163 *non-null float64*
- *margen_retail_sstt* 1106219 *non-null float64*

Se comienza por una exploración visual, consiguiendo una representación del margen por día de cada negocio (en orden desde arriba: amicar, autoplanet, meson, nuevos y sstt). El margen amicar presenta una evolución, si bien dispersa, contenida entre márgenes razonables, al igual que el margen por venta de vehículos nuevos, pero



por otro lado, en el margen de autoplanet, meson y sstt se observan puntos (mínimos y máximos) exageradamente separados de la media.

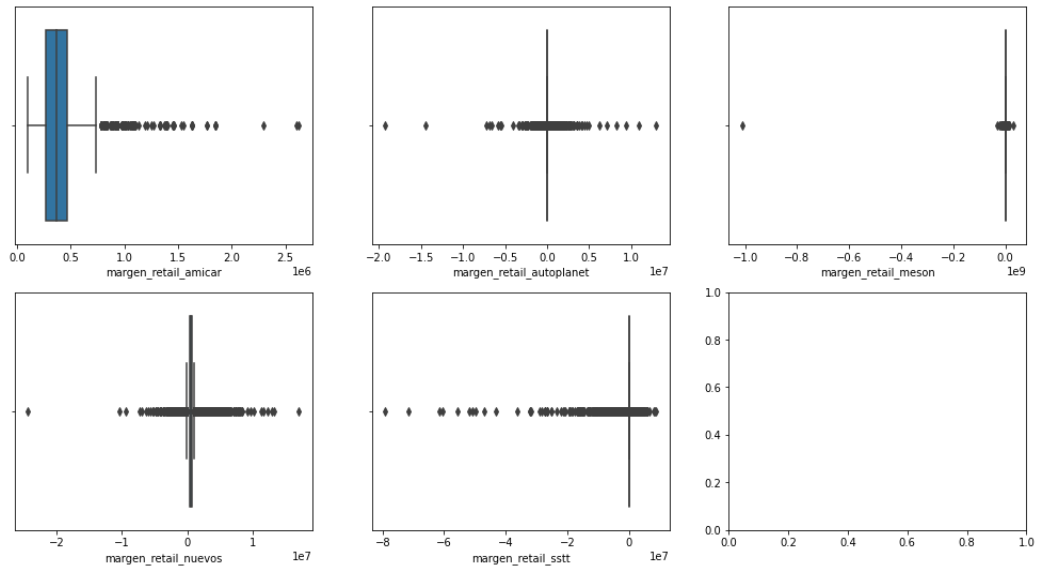
Figura 7: Evolución diaria del margen por negocio.



Esto se confirma con los gráficos de caja (boxplot), donde solamente en amicar se logra apreciar la mediana, cuartiles y bigotes (whisker). En los gráficos restantes, los datos atípicos hacen perder completamente la perspectiva de la distribución, y solo se observa una línea y puntos fuera de los cuartiles. Incluso el margen de nuevos presenta este comportamiento, el que no es tan obvio en el gráfico de línea.

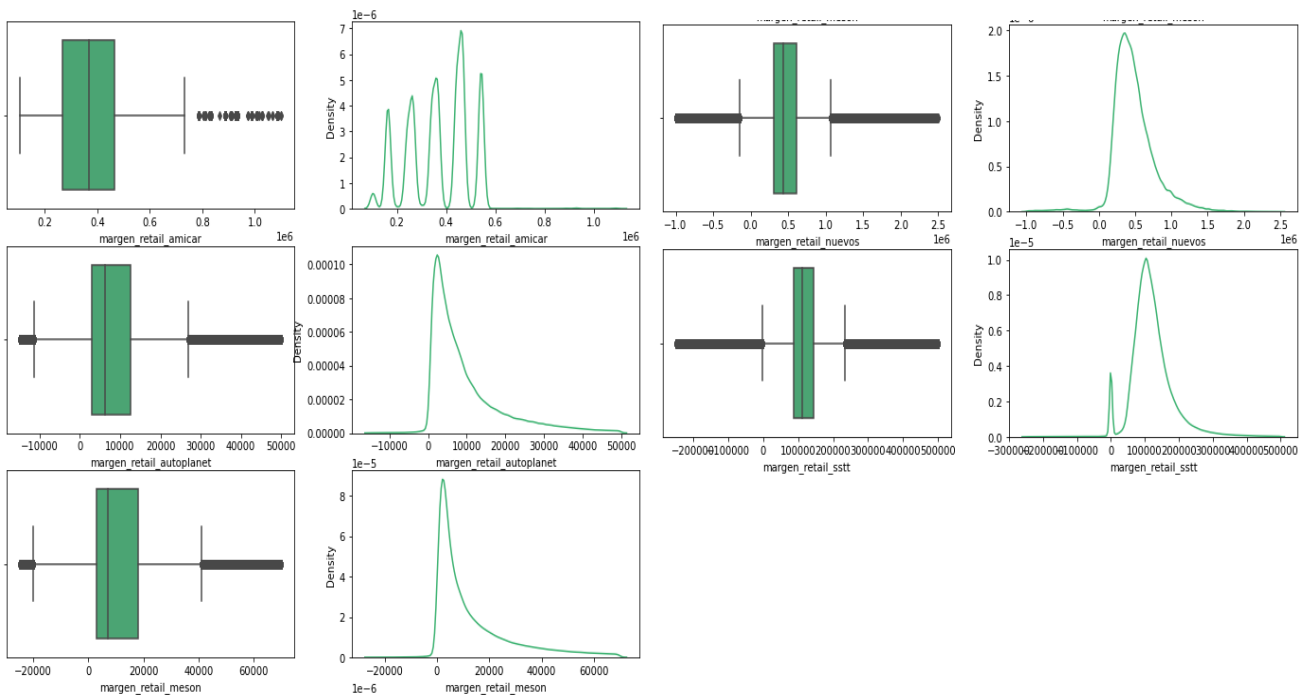


Figura 8: Boxplots de márgenes por negocio.



Como criterio para eliminar los datos atípicos, se tomó el gráfico de densidad de los márgenes para buscar los puntos (mínimos y máximos) donde se asentara la distribución.

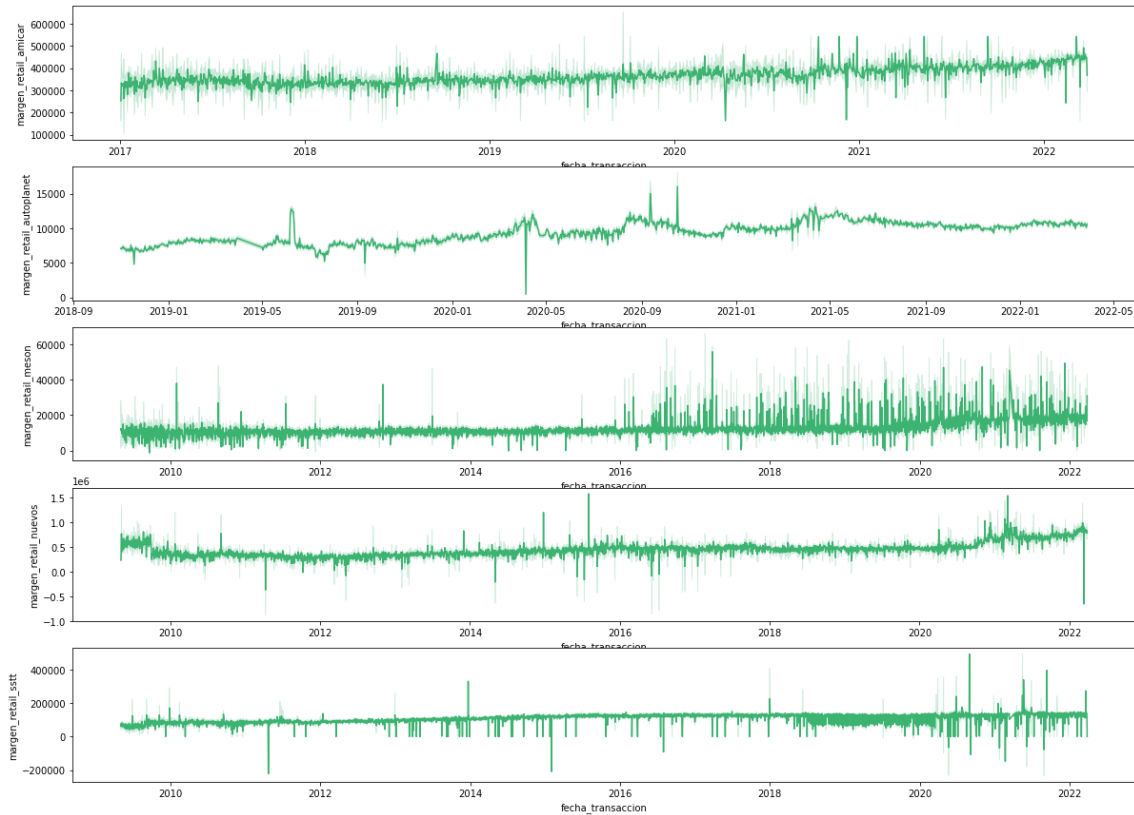
Figura 9: Boxplots y gráficos de densidad de márgenes por negocio (post limpieza de datos).





Con esto se logró disminuir la variabilidad excesiva de los datos, que no permitía observar tendencias, pero mantuvimos la consistencia de la información ya que conservamos 11.719.490 registros.

Figura 10: Evolución diaria del margen por negocio (post limpieza de datos).

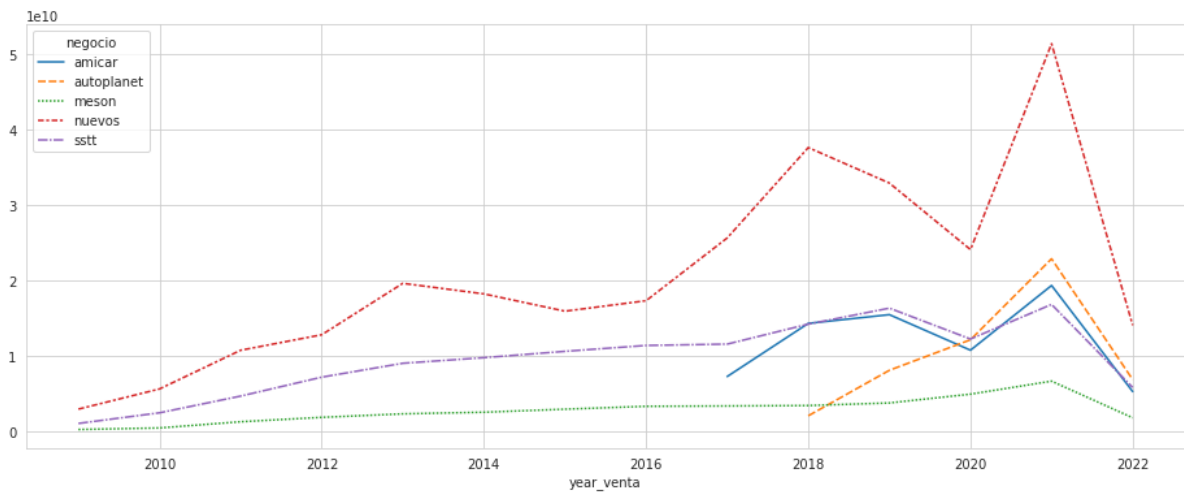


Una vez despejados los datos, se grafica los márgenes por año, mes y semana, con el objetivo de identificar posibles tendencias o estacionalidades que den indicios para profundizar en análisis posteriores (metodología).

En el gráfico por año, si bien se observan tendencias, no logramos apreciar estacionalidades. Por otro lado, notamos que al año 2020 el margen de autoplanet y meson crece a diferencia de las otras unidades de negocio. Luego, se observa una importante recuperación el año 2021 y finalmente un menor margen el año 2022, dado que no se cuenta con los datos del año completo.

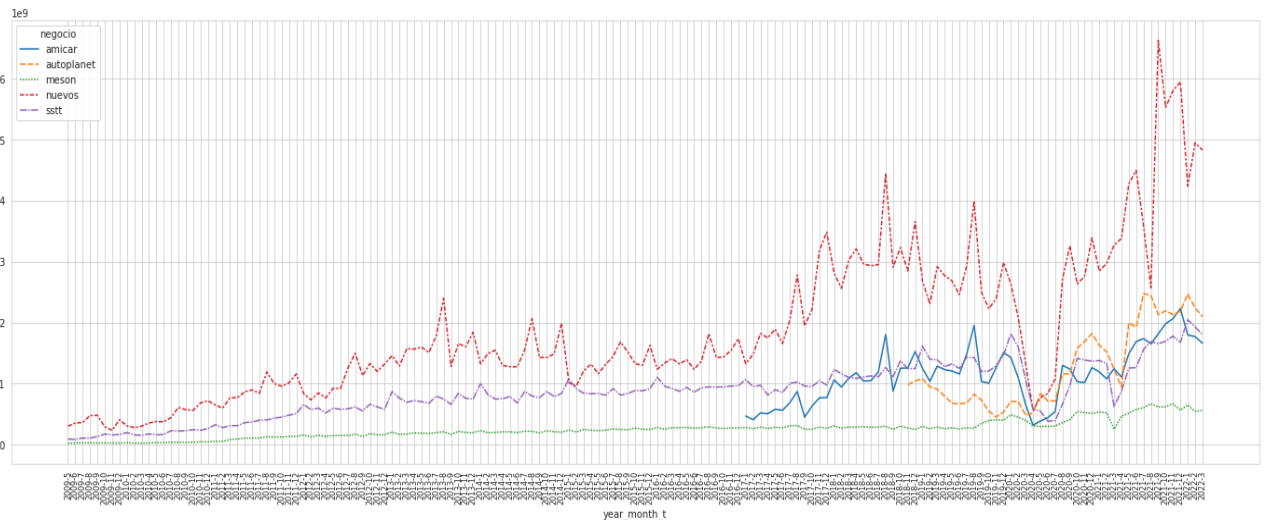


Figura 11: Evolución anual del margen por negocio.



En el gráfico mensual se observa cierta estacionalidad en los márgenes, y además, se percibe una posible relación entre las distintas unidades de negocio, pero considerando periodos o ventanas de tiempo entre un peak y otro (ej.: ventana de 4 meses entre un peak de nuevos y un peak de ssth). Si bien esto se explica por la naturaleza de los datos, es necesario explorarlo en mayor profundidad más adelante.

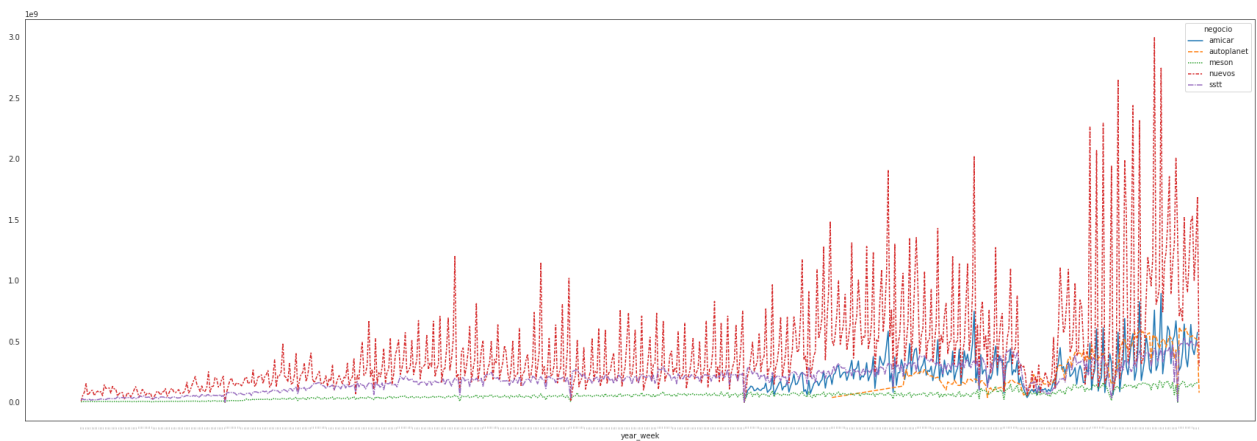
Figura 12: Evolución mensual del margen por negocio.





Para el gráfico del margen semanal, observamos bastante diferencia de una semana a otra (principalmente en nuevos) lo que nos hace suponer que las ventas se concentran en ciertas semanas del mes. A medida que avanzamos, la diferencia entre semanas es cada vez mayor, lo que genera abundante ruido para el análisis de estacionalidad. Cabe destacar, que en todos los casos es claramente evidente el efecto de la pandemia y posterior retiro de fondos previsionales.

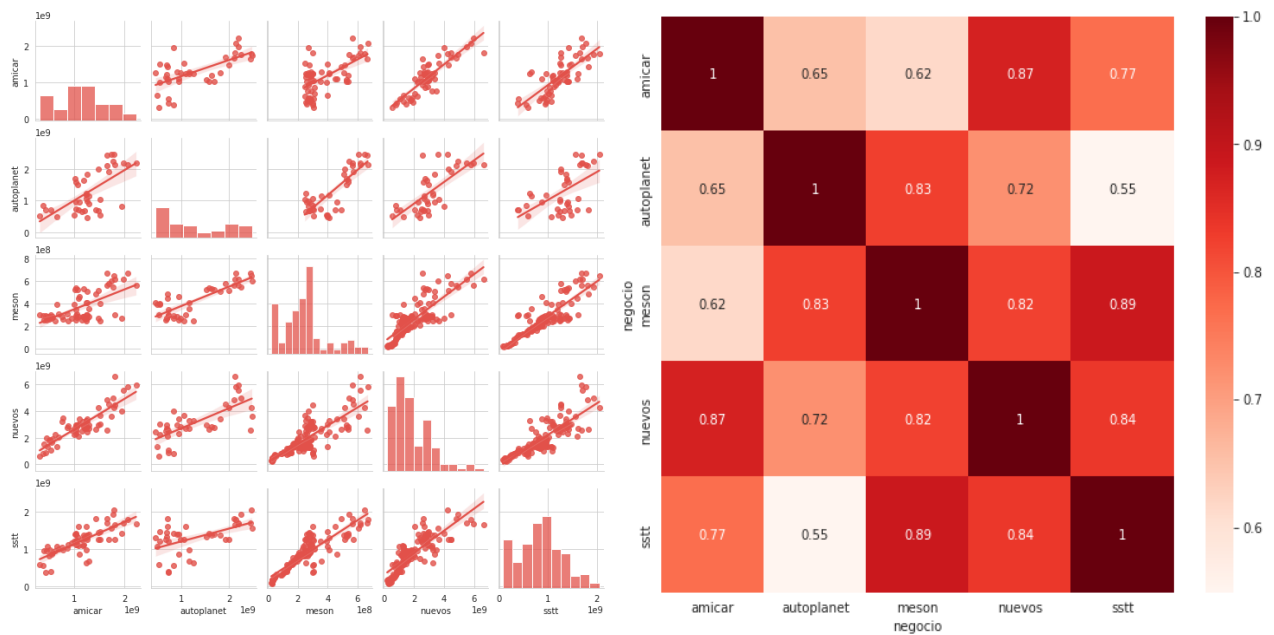
Figura 13: Evolución semanal del margen por negocio.



Por su parte, si analizamos la correlaciones entre márgenes por negocio (agrupados por mes-año para disminuir el ruido), tenemos una correlación fuerte entre nuevos y amicar (0,87); sstt y meson (0,89). Moderada entre meson y autoplanet (0,83); nuevos y meson (0,82); sstt y nuevos (0,84); nuevos y autoplanet (0,72); sstt y amicar (0,77). Y más débil entre autoplanet y amicar (0,65); meson y amicar (0,62); sstt y autoplanet (0,55). Las correlaciones están calculadas de forma directa, sin ajustar por ventanas de tiempo. Pero aun así, nos entregan información que confirma algunos supuestos del tipo de datos con el que se está trabajando. Por ejemplo, la baja correlación entre sstt y autoplanet es esperada, ya que si bien pertenecen a la misma empresa, los clientes podrían comprar los repuestos necesarios para la reparación directamente en meson y no en autoplanet. Al contrario, en otros casos, la correlación no implica causalidad, como lo es con el margen de meson y autoplanet (ambos son competidores).



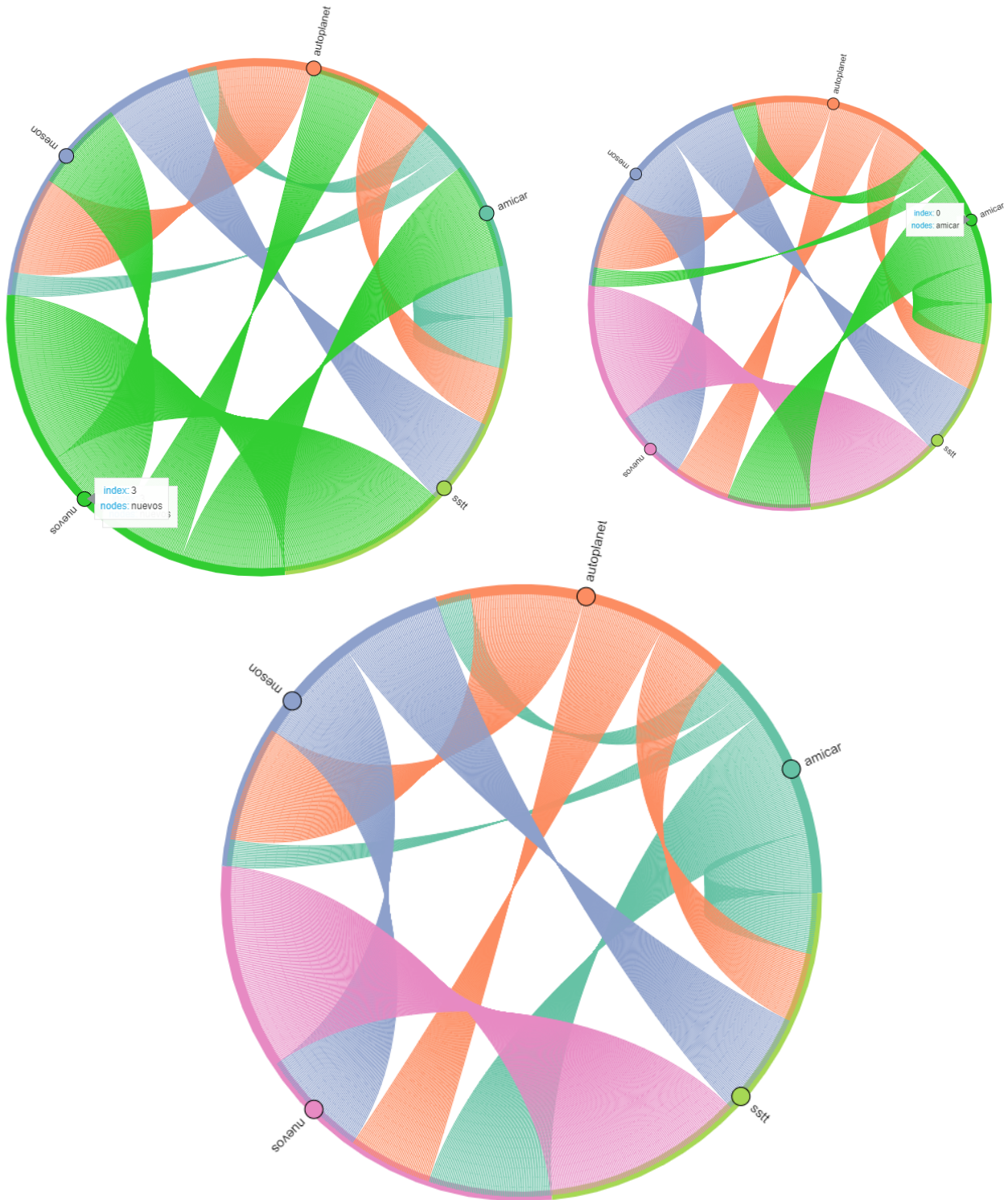
Figura 14: Correlaciones entre márgenes por negocio.



Finalmente se realizó un análisis de la frecuencia de compra entre negocios, a nivel de clientes (persona natural y jurídica). En algunos casos se registraron más de 120 mil transacciones, por lo que el máximo se ajustó a un total de 33 compras, abarcando con esto el 99% de las operaciones. Se consideró a los clientes que habían comprado en más de un negocio, independiente del número de transacciones registradas. El objetivo fue identificar cómo se distribuyen los clientes de un negocio a otro. En el resultado (observable en la siguiente figura) cada arco representa un negocio y las bandas que unen estos arcos, indican el número de clientes compartidos por negocio. En la figura podemos notar que casi la mitad de clientes que compran vehículos nuevos también registran transacciones en sstt, en un grado más bajo con amicar y meson, y en menor medida con autoplanet. Por otro lado, los clientes que tienen alguna transacción con amicar, presentan una baja relación con meson y autoplanet. Y los clientes de sstt, tienen una mayor relación con meson que con autoplanet. Esto último podría indicar que los clientes que acuden a sstt, prefieren comprar repuestos en el mismo concesionario.



Figura 15: Chord Plot Frecuencia de compra entre negocios.





METODOLOGÍA

Para lograr construir un modelo que pronostique los aumentos de márgenes desde algún otro margen del resto de negocios de DERCO se realizaron pruebas de causalidad de “Granger” (Granger, 1969). Para llevar a cabo estas pruebas se identificó previamente que las series temporales de cada negocio sean estacionarias, esto quiere decir que su media y varianza son constantes en el tiempo. Para este análisis se revisaron las fechas mínimas y máximas por serie de tiempo (negocio) y se normalizaron a la fecha máxima de la fecha mínima de “Fecha Min” (01-11-2018) y la fecha mínima de la fecha máxima de “Fecha Max”.

Tabla 1: Fechas de corte para normalización series temporales por negocio.

	Fecha Min	Fecha Max	Nueva Fecha Min	Nueva Fecha Max
amicar	02-01-2017	27-03-2022	01-11-2018	27-03-2022
autoplanet	01-11-2018	28-03-2022	01-11-2018	27-03-2022
meson	04-05-2009	27-03-2022	01-11-2018	27-03-2022
nuevos	05-05-2009	28-03-2022	01-11-2018	27-03-2022
sstt	04-05-2009	27-03-2022	01-11-2018	27-03-2022

Estos cortes son necesarios para comparar las series entre sí y poder obtener la matriz de causalidad.

Para cada una de estas series se revisó su estacionalidad utilizando el método “Dickey-Fuller” (Dickey & Fuller, 1981) con corte en el p-value de < 0.05 . Como resultado para cada serie nos encontramos que autoplanet, meson y sstt no son estacionarias. En este ejercicio se excluyeron los meses de pandemia por considerarlas fechas anormales en la venta de vehículos ya que la venta presencial no estaba disponible. Estas fechas son desde 01-04-2020 hasta el 31-08-2020.

Tabla 2: Estacionariedad por negocio. Método “Dickey-Fuller” .

	Estacionaria	No Estacionaria
amicar	X	
autoplanet		X
meson		X
nuevos	X	
sstt		X



Esto nos obliga a transformar las series en estacionarias para que el ejercicio de causalidad sea efectivo, obteniendo la diferencia de cada una de los registros de la serie con su registro anterior, según lo que se logra ver en el código.

```
#transformamos las series a estacionarias
list_negocios = ['meson', 'sstt', 'nuevos', 'amicar', 'autoplanet']
df_margen_tranformed = df_margen_filter_dates.pivot(index='fecha_transaccion',
                                                    columns='negocio', values='margen_retail')
df_margen_tranformed = df_margen_tranformed.diff().dropna()
df_margen_tranformed = pd.melt(df_margen_tranformed.reset_index(),
                              id_vars='fecha_transaccion',
                              value_vars=list_negocios).rename(columns={'value':'margen_retail'})
df_margen_tranformed
```

El ejercicio se realizó con temporalidad mensual, por lo que los lags que se identificaron corresponden a espacios de tiempo de 1 mes.

En las siguientes tablas se puede ver el resultado de los lags del 1 al 4, considerando las variables de negocio con término “_x” las series de tiempo regresoras y las terminadas en “_y” las posibles pronosticadas. La métrica considerada para la comparación es: “ssr_chi2test”

Tabla 3: Resultados análisis de causalidad de “Granger” desde lag_1 a lag_4 .

lag 1	amicar_x	autoplanet_x	meson_x	nuevos_x	sstt_x
amicar_y	1.00	0.32	0.00	0.02	0.00
autoplanet_y	0.33	1.00	0.89	0.22	0.80
meson_y	0.07	0.38	1.00	0.01	0.11
nuevos_y	0.14	0.61	0.21	1.00	0.10
sstt_y	0.02	0.52	0.04	0.00	1.00
lag 2	amicar_x	autoplanet_x	meson_x	nuevos_x	sstt_x
amicar_y	1.00	0.11	0.01	0.04	0.01
autoplanet_y	0.19	1.00	0.69	0.08	0.55
meson_y	0.28	0.18	1.00	0.12	0.09
nuevos_y	0.30	0.30	0.17	1.00	0.19
sstt_y	0.16	0.22	0.07	0.07	1.00



lag 3	amícar_x	autoplanet_x	meson_x	nuevos_x	ssft_x
amícar_y	1.00	0.07	0.02	0.04	0.01
autoplanet_y	0.46	1.00	0.81	0.28	0.79
meson_y	0.40	0.28	1.00	0.06	0.16
nuevos_y	0.15	0.04	0.10	1.00	0.13
ssft_y	0.46	0.26	0.18	0.09	1.00

lag 4	amícar_x	autoplanet_x	meson_x	nuevos_x	ssft_x
amícar_y	1.00	0.04	0.05	0.01	0.04
autoplanet_y	0.41	1.00	0.76	0.28	0.58
meson_y	0.48	0.30	1.00	0.16	0.20
nuevos_y	0.01	0.00	0.19	1.00	0.10
ssft_y	0.60	0.35	0.19	0.09	1.00

Como podemos observar en las tablas anteriores se analiza la causalidad por cada lag de tiempo (1 mes), lo que va mostrando variaciones según cada negocio, como ejemplo entre nuevo y meson en el primer mes hay una hipótesis nula pero en el segundo mes esa hipótesis nula se rechaza, por lo que se puede considerar esa causalidad. A continuación se agrupan los valores más bajos para un análisis general.

Tabla 4: Revisión general de resultados de análisis de causalidad de “Granger”.

	amícar_x	autoplanet_x	meson_x	nuevos_x	ssft_x
amícar_y	1.00	0.04	0.00	0.01	0.00
autoplanet_y	0.19	1.00	0.69	0.08	0.55
meson_y	0.07	0.18	1.00	0.01	0.09
nuevos_y	0.01	0.00	0.10	1.00	0.10
ssft_y	0.02	0.22	0.04	0.00	1.00

Donde podemos observar que dentro de un máximo de 4 lags se puede ver que la serie de tiempo del negocio “nuevos” se puede considerar causal de la mayoría de las demás series, con excepción de “autoplanet” la cual si bien su valor sigue siendo bajo no podemos decir que la hipótesis nula se rechaza para esa serie.

Recordar que lo que hace esta prueba es corroborar la hipótesis nula de la causalidad, por lo que los p-valores bajos nos indican que la nula causalidad no existe, pero no confirma la causalidad en sí.



Modelos con Prophet:

Se confeccionó un modelo de para cada negocio utilizando la herramienta para pronósticos de series temporales Prophet (Taylor & Letham, 2017) y se agregaron y compararon como regresoras las series de tiempo de los negocios con causalidad identificada para comparar entre sí.

- **Serie del margen retail sin regresores adicionales:** Esto solo considerando el margen retail de este negocio.
- **Serie del margen retail con regresores adicionales:** Para efectos de comparación con el anterior modelo y validar si mejora en algo incluyendo los márgenes de venta retail de los negocios con causalidad.

Para efectos del entrenamiento de los modelos se consideraron las fechas de la pandemia (01-04-2020 hasta el 31-08-2020) como feriados para poder darles una priorización distinta a las demás fechas. Además, se realizó una búsqueda de los mejores hiper paramétricos mediante la técnica de validación cruzada. Considerando ventanas de tiempo de 7 días dentro de un periodo de 250 días, esto nos dio 5 ventanas de tiempo dentro de cada uno de la grilla de parámetros a revisar dando un total de 48 combinaciones. Ejemplo de Grilla de Parámetros:

```
param_grid = {'changepoint_prior_scale': [0.01, 0.1, 0.3, 0.5],
              'seasonality_prior_scale': [0.01, 0.1, 5.0, 10.0],
              'holidays_prior_scale': [0.01, 0.1, 1],
              'seasonality_mode': ['additive']
            }
```

Los modelos se entrenaron con sus registros en base logarítmica y los resultados se midieron utilizando estos registros, luego para efectos de visualización se reconvirtieron a base exponencial.

Tabla 5: Evaluación resultados MAPE (mean absolute percentage error).

MAPE	DESCRIPCIÓN
<0.1	Alta Precisión en el Forecasting
0.1-0.2	Buen Forecasting
0.2-0.5	Forecasting Razonable
>0.5	Forecasting Impreciso



Las métricas utilizadas en cada modelo fueron MAE, MSE, RMSE y MAPE, guiandonos para la validación cruzada con RMSE y luego revisando el MAPE y el RMSE para la comparación y revisión de los regresores correspondientes a las series de otros negocios que tuvieran causalidad.

Modelos con LSTM (Redes Neuronales):

Para el modelamiento de series de tiempo con redes neuronales recurrentes LSTM, como punto de partida, se comenzó con la serie del margen de venta de vehículos nuevos. Pero, debido a la excesiva diferencia en la magnitud de los datos (min y max) se decidió, para todos los modelos, ingresar los datos en base logarítmica (del mismo modo que en el modelamiento con Prophet).

Con los datos ajustados y estructurados para el entrenamiento. Primero se definió un modelo secuencial, apilado y de un solo paso, para la predicción de las series de nuevos y sstt. Para ambos casos, se entrenó con el 85% de los datos, con una ventana de predicción de 21 días (3 semanas), ya que las pruebas efectuadas demostraron un buen comportamiento con dicho periodo.

La arquitectura de ambos modelos es la siguiente:

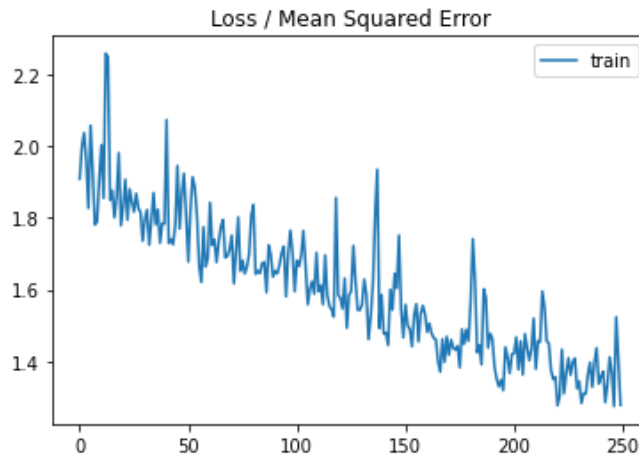
- Primera capa de entrada tridimensional (LSTM) / 100 neuronas / función de activación “relu” (Rectified Linear Unit). Genera una salida bidimensional, corregida para entrada de la segunda capa con el parámetro “return_sequences=True”.
- Segunda capa tridimensional (LSTM) / 100 neuronas / función de activación “relu”.
- Capa de salida “Dense”, caracterizada por estar profundamente conectada con su capa precedente.
- Compilado con optimizador “adam” y función de pérdida (loss) “mse” (error cuadrático medio).



```
# define model Stacked
model = Sequential()
model.add(LSTM(100, activation='relu', return_sequences=True,
              input_shape=(n_steps_in, n_features)))
model.add(LSTM(100, activation='relu'))
model.add(Dense(n_steps_out))
model.compile(optimizer='adam', loss='mse')
```

Los modelos se ajustaron con un batch size de 45 días y un epochs de 250 (estimado de acuerdo a su función de pérdida) .

Figura 16: Función de pérdida (mse) de entrenamiento modelo LSTM Nuevos.



El desempeño y resultado del entrenamiento se detallan en el siguiente capítulo.

Por otro lado, y en base a las relaciones ya mencionadas en el análisis de causalidad de Granger. Se realizó un modelo de predicción multivariable con la misma arquitectura y ventana de 21 días. Pero que se caracteriza por tomar más de una variable y entrenar incorporando la relación estacional de las distintas series de tiempo. Este método ,“Multiple Parallel Series”, permite entrenar dos series de tiempo a la vez y lograr predicciones de forma independiente. Los datos deben ser reestructurados en lotes (de acuerdo a la ventana de predicción, cuántos datos del pasado tomará el modelo para predecir el valor siguiente) y se debe incorporar una tercera variable con la relación de las series.



RESULTADOS

La base para entrenamiento de cada modelo comienza desde el 01-11-2018 y terminan el 23-09-2021, para los registros de prueba estos comienzan el 24-09-2021 y terminan el 26-03-2022 correspondiente al 15% de la base total considerada.

Resultados con Prophet:

Para los siguientes modelos se usaron los siguientes hiper parámetros encontrados según proceso de validación cruzada, luego en cada gráfico y resultados se detallan los regresores utilizados, mostrando la diferencia entre cada regresores por modelo.

Tabla 6: Resultados hiper parámetros por validación cruzada.

modelo/negocio	daily_seasonality	changepoint_prior_scale	seasonality_prior_scale	holidays_prior_scale	seasonality_mode
nuevos	True	0.50	1.00	0.50	additive
amicar	True	0.50	1.00	1.00	additive
ssttt	True	10.00	0.01	0.01	additive
autoplanet	True	10.00	0.50	0.01	additive
meson	True	10.00	0.05	0.01	additive

Los gráficos muestran solo los resultados comparables con el dataset de pruebas, donde ‘y’ es el valor real y ‘yhat’ el valor predecido. No se consideran valores futuros debido a la naturaleza del ejercicio y que la base utilizada no está actualizada por lo que los valores futuros predecidos serían fechas ya pasadas que no tenemos cómo validar.

En las tablas de resultados las métricas consideraron los márgenes en base logarítmica, de igual manera que el entrenamiento. Para cada tabla de resultados la columna regresores con el registro en blanco “[]” se refiere al resultado del modelo individual, sin regresores adicionales.

Margen Retail de Vehículos Nuevos

El primer modelo que se construyó fue el margen retail donde se comparó el regresor adicional del margen de autoplanet con el propio margen de vehículos nuevos. Esto debido a los resultados del análisis de Causalidad de “Granger” de 0.04 para “autoplanet-nuevos”.



Figura 17: Gráfico sin regresores adicionales (Prophet - nuevos).

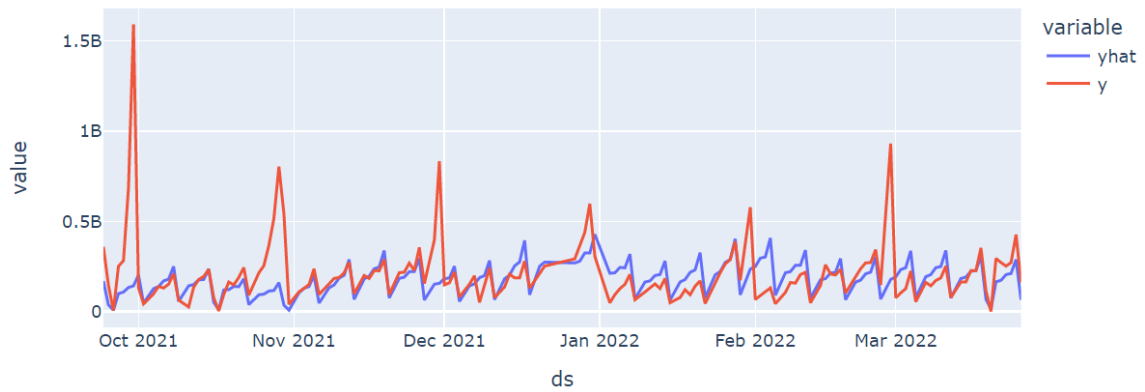


Figura 18: Gráfico en Margen Retail con regresor adicional de márgenes de Autoplanet (Prophet - nuevos).

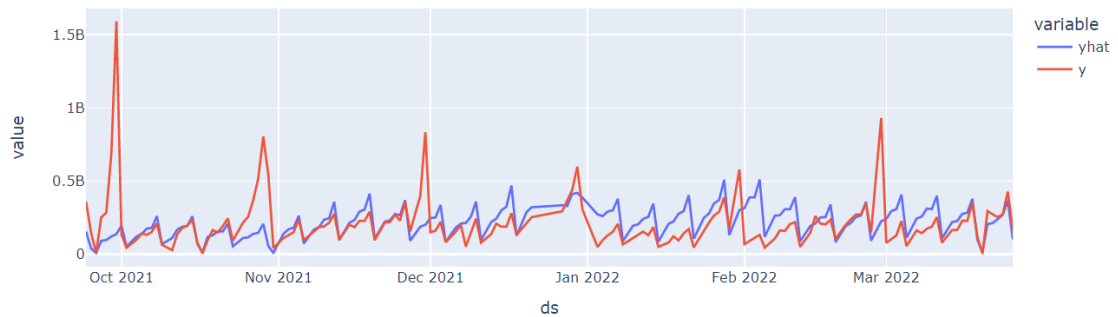


Tabla 7: Resultados de Métricas de la Predicción (Prophet - nuevos).

regresores	mae	mse	rmse	mape
[]	0.53	0.54	0.73	0.03
['autoplanet']	0.50	0.48	0.69	0.03

Podemos observar que no varía demasiado el resultado incluyendo la variable de margen retail de autoplanet, siendo que los resultados de las pruebas de causalidad reflejaban que podría ser una variable a considerar. Esto demuestra que la correlación no implica causalidad.



Margen Retail de Servicio Técnico.

En este modelo se consideraron para comparación los márgenes retail adicionales a los del negocio nuevos, amicar y meson, los cuales eran las posibles series con causalidad.

Figura 19: Gráfico sin regresores adicionales (Prophet - sstt).

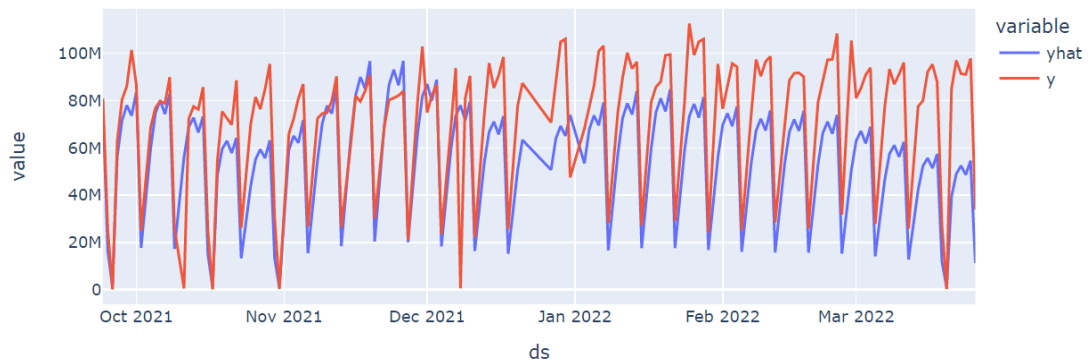


Figura 20: Gráfico en Margen Retail con Margen Retail de Nuevos como regresor adicional (Prophet - sstt).

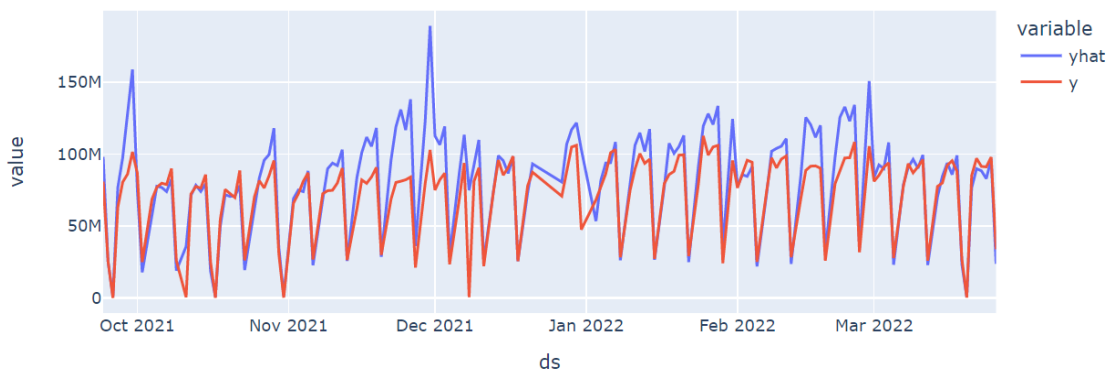


Figura 21: Gráfico en Margen Retail con Margen Retail de Nuevos, Amicar y Mesón como regresores adicionales (Prophet - sstt).

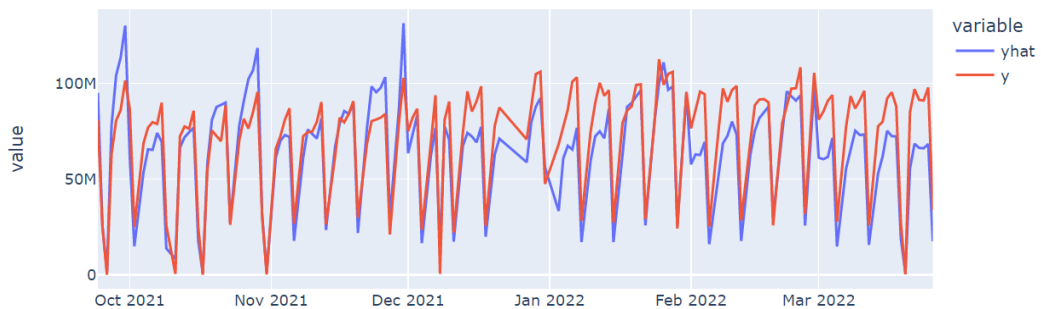




Tabla 8: Resultados de Métricas de la Predicción (Prophet - sstt).

regresores	mae	mse	rmse	mape
[]	0.35	0.54	0.74	0.02
['nuevos']	0.26	0.42	0.65	0.02
['nuevos', 'amicar']	0.31	0.39	0.62	0.02
['nuevos', 'amicar', 'meson']	0.29	0.30	0.55	0.02

Se puede observar claramente la mejora tanto en las gráficas como en el rmse, lo que confirma el uso de estas series como regresoras.

Margen Retail de Mesón.

Para el negocio de meson, el cual corresponde en su mayoría a la venta de repuestos de las marcas de vehículos que vende derco, se consideró según tabla del análisis de causalidad los márgenes de la venta de vehículos nuevos, amicar y servicio técnico como regresores adicionales.

Figura 22: Gráfico sin regresores adicionales (Prophet - meson).

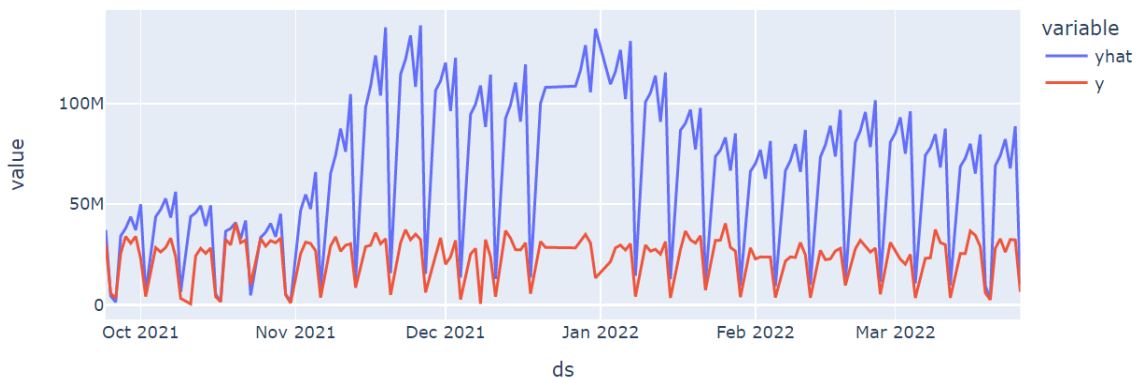




Figura 23: Gráfico Margen Retail con Margen Retail de Nuevos como regresor adicional (Prophet - meson).

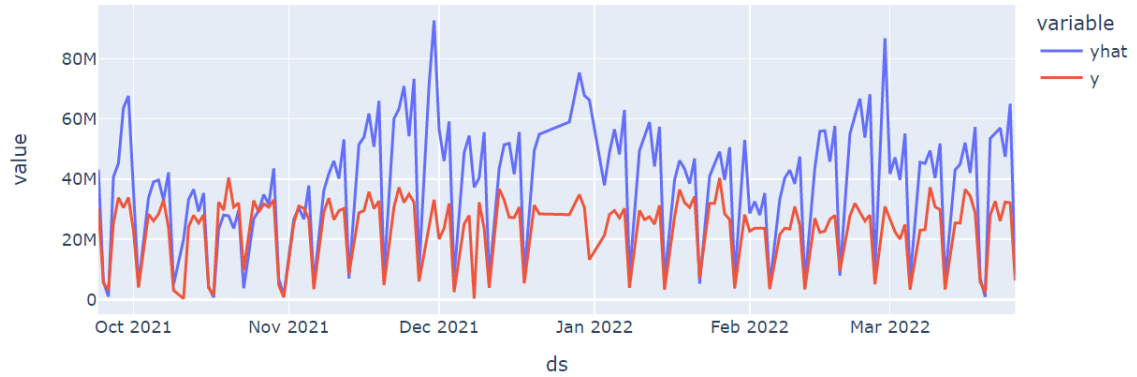


Figura 24: Gráfico Margen Retail con Margen Retail de Nuevos y Amicar como regresores adicionales (Prophet - meson).

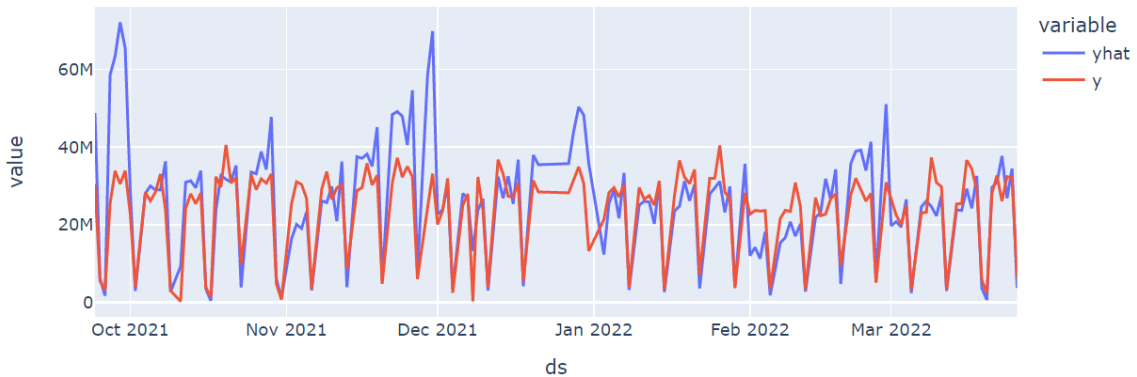


Figura 25: Gráfico Margen Retail con Margen Retail de Nuevos, Amicar y Servicio Técnico como regresores adicionales (Prophet - meson).

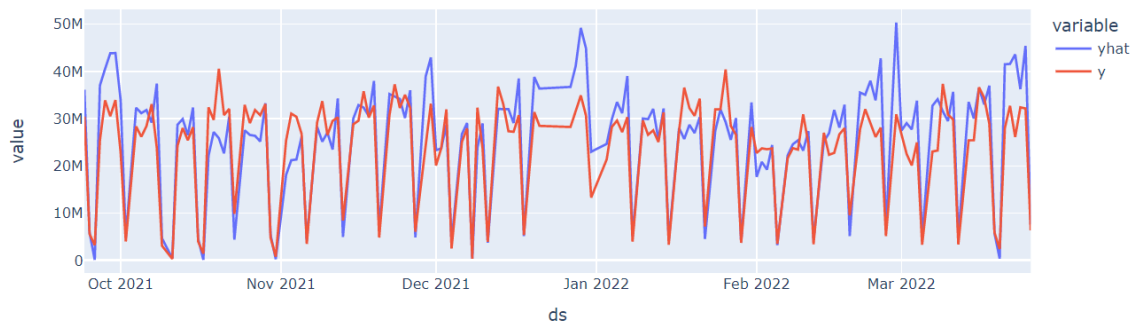




Tabla 9: Resultados de Métricas de la Predicción (Prophet - meson).

regresores	mae	mse	rmse	mape
[]	0.85	1.11	1.05	0.05
['nuevos']	0.60	0.63	0.79	0.04
['nuevos', 'amicar']	0.34	0.32	0.57	0.02
['nuevos', 'amicar', 'sstt']	0.25	0.19	0.44	0.02

Podemos observar cómo mejora el modelo principalmente con el margen de venta de los vehículos nuevos y amicar.

Margen Retail de Autoplanet.

Figura 26: Gráfico sin regresores adicionales (Prophet - autoplanet).

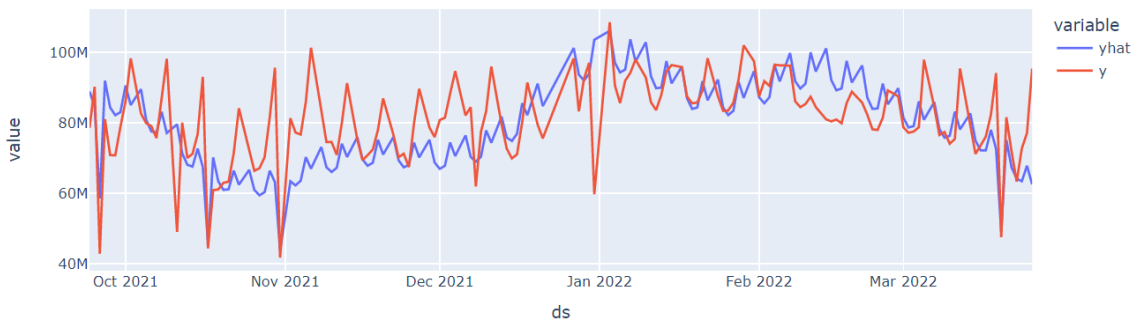


Figura 27: Gráfico Margen Retail con Margen Retail de nuevos como Regresor Adicional (Prophet - autoplanet).

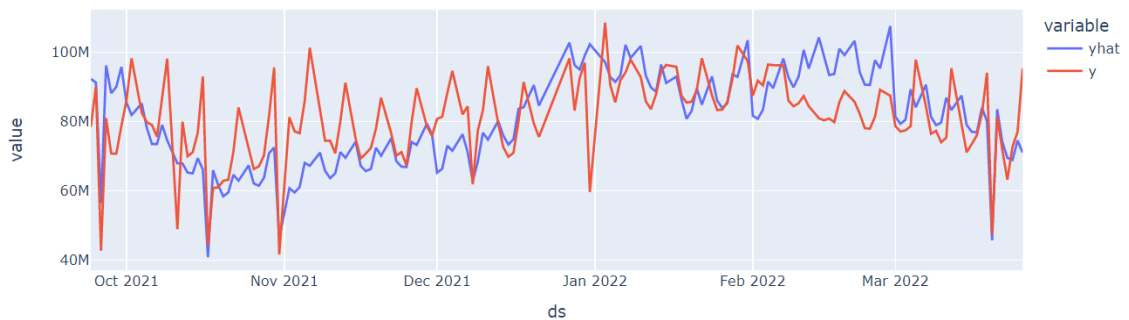




Tabla 10: Resultados de Métricas de la Predicción (Prophet - autoplanet)

regresores	mae	mse	rmse	mape
[]	0.10	0.02	0.14	0.01
['nuevos']	0.11	0.02	0.14	0.01

No se logra apreciar un mejor desempeño con la nueva variable regresora de los márgenes retail de venta de vehículos nuevos, esto va en línea con los resultados de las pruebas de causalidad anteriormente mencionadas, donde el resultado no estaba dentro de los parámetros del <0.05 pero igual manera no tenía un valor elevado para descartar la variable.

Margen Retail de Amicar.

Figura 28: Gráfico sin regresores adicionales (Prophet - amicar).

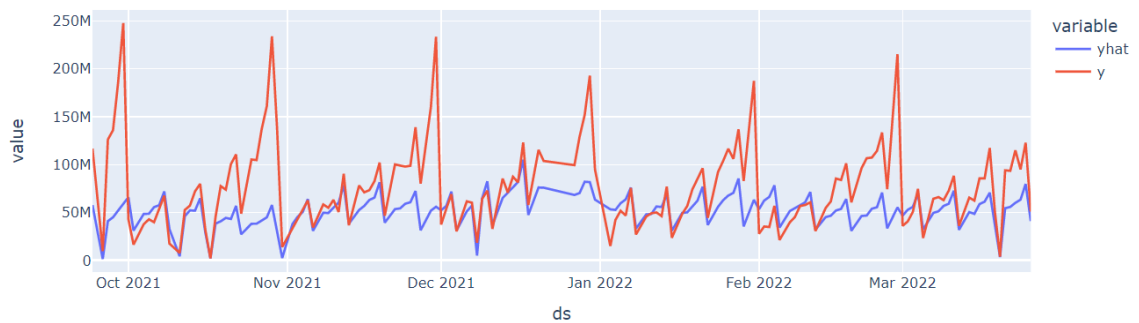


Figura 29: Gráfico Margen Retail con Margen Retail de nuevos como regresor adicional (Prophet - amicar).

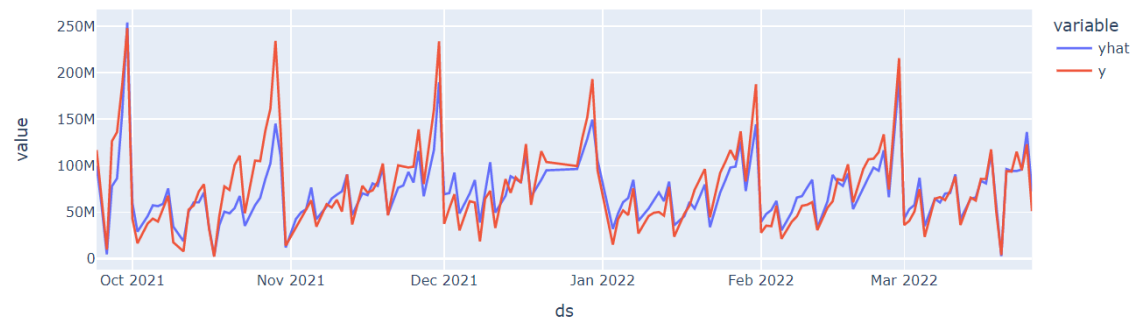




Tabla 11: Resultados de Métricas de la Predicción (Prophet - amicar)

regresores	mae	mse	rmse	mape
[]	0.44	0.36	0.60	0.02
['nuevos']	0.22	0.08	0.28	0.01
['sstt']	0.45	0.37	0.61	0.02
['nuevos', 'meson']	0.22	0.07	0.26	0.01
['meson', 'sstt']	0.43	0.34	0.58	0.02

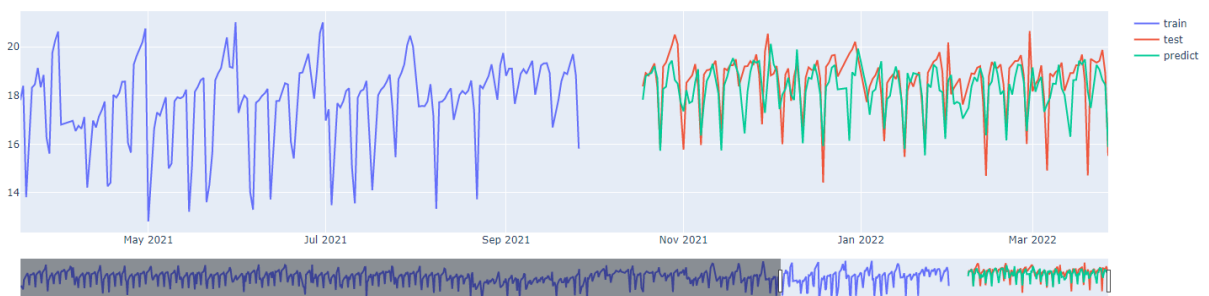
Logramos apreciar nuevamente la mejora que se logra con el margen retail de la venta de vehículos nuevos, las cuales desde un principio se sabía que estaban fuertemente relacionadas, no así con la variable de servicio técnico.

Resultados con Modelos LSTM:

A continuación se presentan los resultados de los modelos efectuados mediante redes neuronales LSTM. Se consideran las mismas métricas que en los modelos anteriormente expuestos (Prophet). Las predicciones se realizaron sobre la data de prueba, correspondiente al último 15% del total. Los modelos fueron entrenados con los datos en base logarítmica, por lo tanto, su representación gráfica fue transformada exponencialmente para comparar con los datos de prueba.

A modo de ejemplo se deja un muestra del modelo LSTM entrenado solo con datos de la serie nuevos en base logarítmica.

Figura 30: Entrenamiento modelo LSTM margen nuevos en base logarítmica.





Para una mejor visualización de los resultados, a continuación se exponen gráficamente los modelos LSTM considerando solo una variable (serie temporal del margen del negocio), versus los modelos multivariantes con variable paralela. Esto en relación al resultado de las pruebas de causalidad de “Granger”. Para todos los gráficos a continuación la línea en azul corresponde a los datos de prueba (test) y la línea en rojo a la predicción.

Figura 31: Gráfico modelo LSTM solo nuevos.

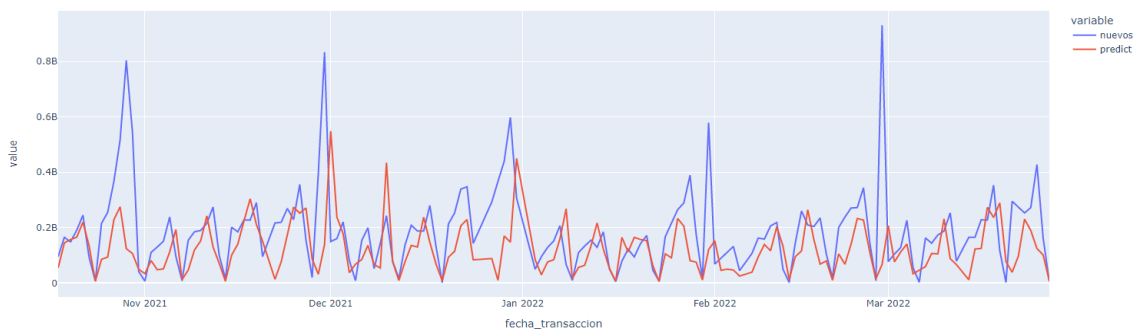


Figura 32: Gráfico modelo LSTM multivariable nuevos (sstt como variable paralela).

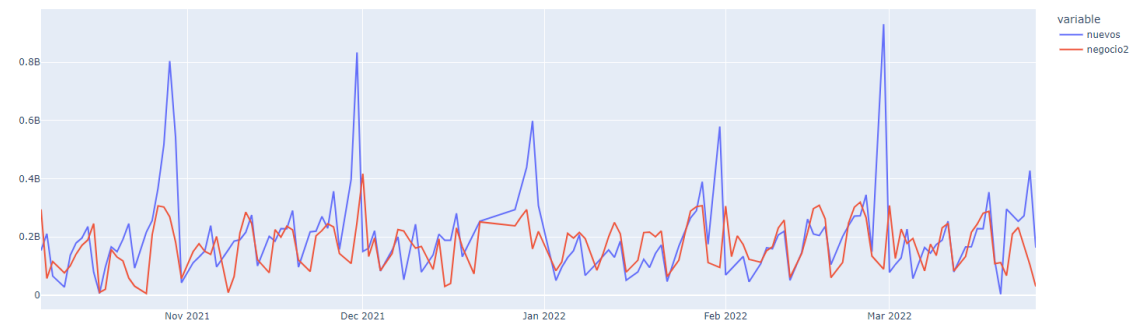


Figura 33: Gráfico modelo LSTM solo sstt.

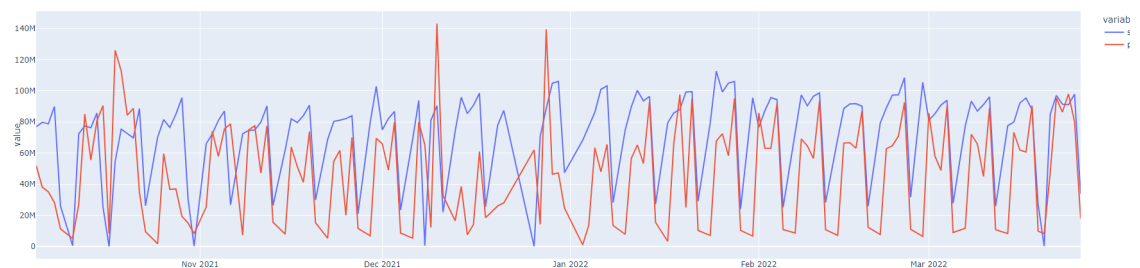




Figura 34: Gráfico modelo LSTM multivariable sstt (nuevos como variable paralela).

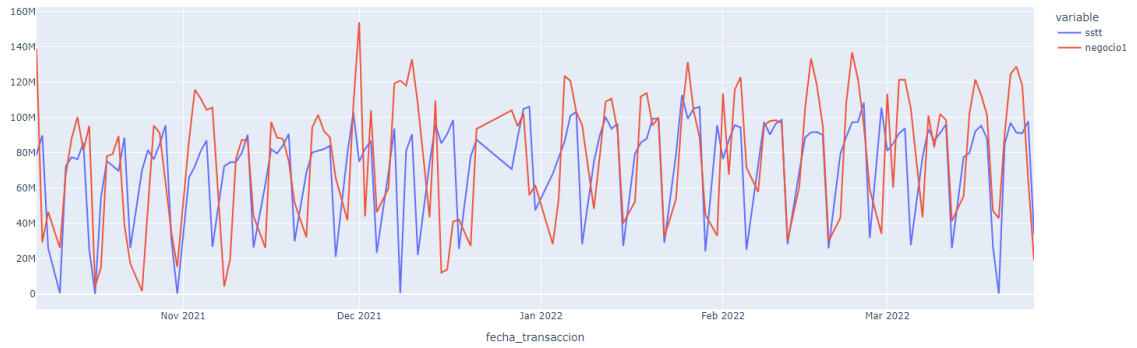


Figura 35: Gráfico modelo LSTM solo meson.

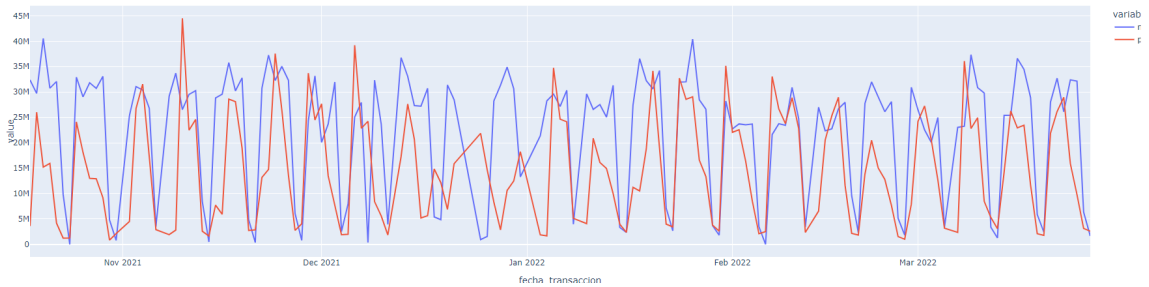


Figura 36: Gráfico modelo LSTM multivariable meson (nuevos como variable paralela).

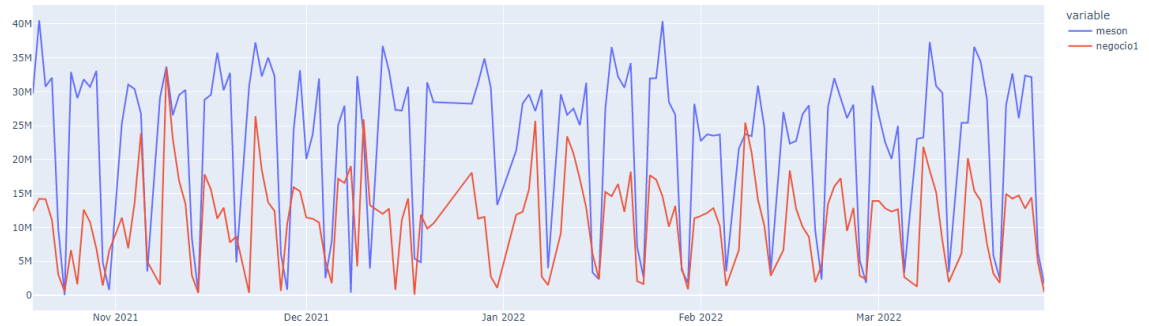


Figura 37: Gráfico modelo LSTM solo autoplanet.

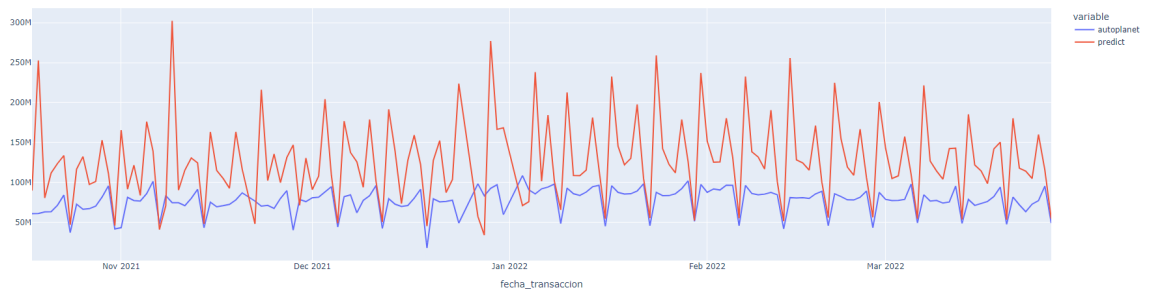




Figura 38: Gráfico modelo LSTM multivariable autoplanet (nuevos como variable paralela).

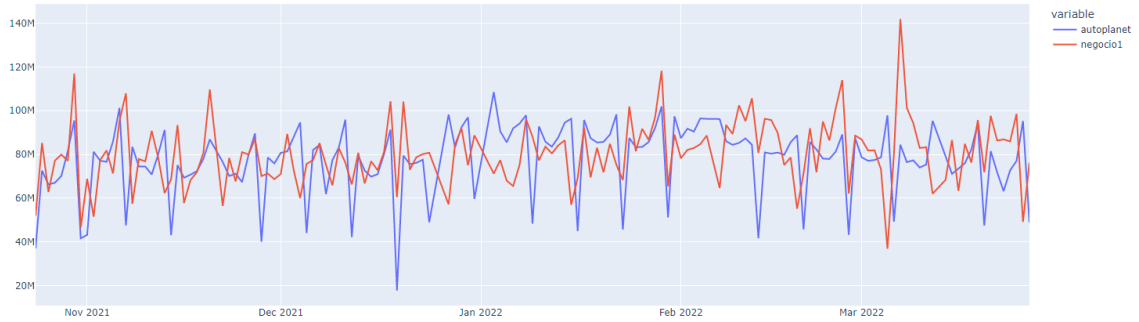


Figura 39: Gráfico modelo LSTM solo amicar.

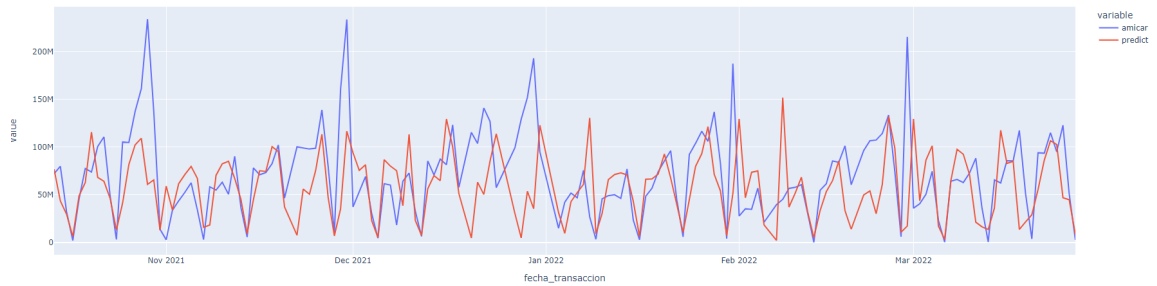


Figura 40: Gráfico modelo LSTM multivariable amicar (nuevos como variable paralela).

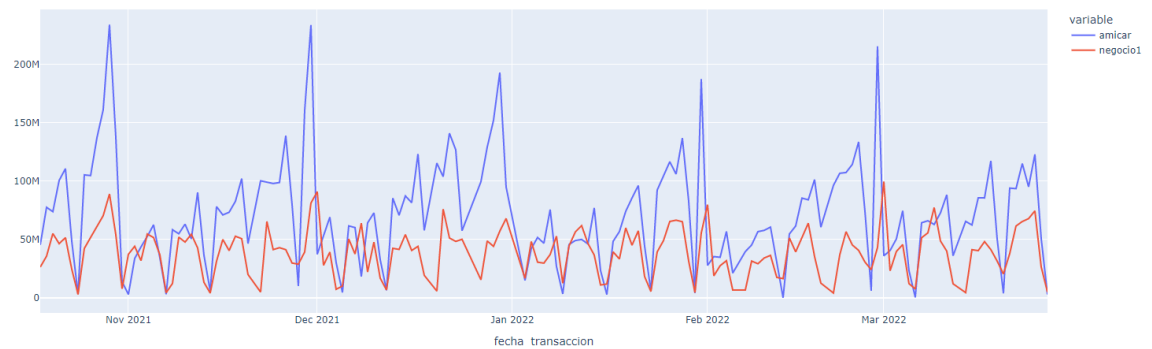




Tabla 12: Resultados de Métricas modelos LSTM

modelo	mae	mse	rmse	mape
solo_nuevos	0.73	0.95	0.97	0.039
solo_sstt	1.22	4.69	2.16	0.071
solo_meson	0.85	1.54	1.24	0.056
solo_autoplanet	0.48	0.33	0.57	0.026
solo_amicar	0.64	0.91	0.95	0.037
multivariable_nuevos (sstt)	0.54	0.72	0.84	0.029
mutivariable_sstt (nuevos)	0.57	1.14	1.06	0.035
multivariable_meson (nuevos)	0.99	1.70	1.30	0.062
multivariable_autoplanet (nuevos)	0.20	0.08	0.28	0.011
multivariable_amicar (nuevos)	0.78	1.01	1.00	0.044

En comparativa, los modelos que mejoran al sumar una variable paralela son: **nuevos**, **sstt** y **autoplanet**. Para las series **meson** y **amicar**, los modelos tienen un mejor desempeño sin sumar una variable paralela. Por otro lado, gráficamente, en los modelos de sstt y autoplanet (con nuevos como variable paralela), el ajuste de la predicción a los datos de prueba es bastante claro.



CONCLUSIONES

Según lo observado podemos decir que el margen retail de la venta de vehículos nuevos si puede ser usado como regresor en apoyo a la predicción del margen de los demás negocios, especialmente en los márgenes de amicar, sstt y meson. En relación a los resultados, se observa que los modelos de Prophet con regresores adicionales, presentan un mejor desempeño que los modelos simples (univariados) o los modelos de redes neuronales. Si bien, al comparar los modelos LSTM multivariantes con los modelos de múltiples regresores de Prophet, la diferencia no es significativa, se comprueba que la opción de agregar más de un regresor mejora el comportamiento de los modelos de las otras unidades de negocio (ej.: amicar y meson).

Como alternativa de mejora, es posible optimizar los parámetros e hiper parámetros de los modelos de redes neuronales, o incluso diseñar una arquitectura que aumente su capacidad predictiva. Sería interesante además probar con modelos de multipasos (predicción de varios periodos de una sola vez). Pero para efectos prácticos, el desempeño logrado con Prophet es el idóneo. Por otro lado, hubiera sido ideal contar con un mayor número de días en los registros, tanto para los años anteriores, por parte de los negocios de autoplanet y amicar, como también contar con fechas más cercanas a la actual para predecir resultados futuros con los modelos. De igual forma sería interesante poder hacer un análisis por cliente, si bien la base de datos tiene el registro de estos, el cruce por Rut no es suficiente. Ya que no se cuenta con el tipo de producto específico para los negocios que no son de venta de vehículos (como son los accesorios en autoplanet y los repuestos del automóvil) por lo que la compra del producto puede no estar asociada al vehículo nuevo adquirido. Es por esto que no se logró dar respuesta a los objetivos específicos listados al principio, pero si se abordan de manera indirecta con los datos de los márgenes de venta por cada negocio.

Sería además interesante poder analizar alguna variable externa a la base, como por ejemplo algún indicador económico del país (PIB o el IMACEC) para ver si afecta o no en los márgenes de cada negocio y su posible utilización en los modelos.



BIBLIOGRAFÍA

Alonso, M. (2022, April 26). *Análisis y Predicción de Series Temporales con FB Prophet Python*. Model Differently. Retrieved November 22, 2022, from https://www.modeldifferently.com/2022/04/analisis_prediccion_ts_prophet/

Ángel RC. (2020, April 24). *Forecasting con Prophet. Caso práctico prediciendo el consumo eléctrico en España*. Medium. Retrieved November 22, 2022, from <https://medium.com/@angel.r.chicote/forecasting-con-prophet-7fb36b25eb4b>

Dickey, D. A., & Fuller, W. A. (1981). Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root. *Econometrica*, 49(4), 1057-1072. <https://doi.org/10.2307/1912517>

Erika, E., Khulan, B., & Ting Ting, L. (2013). *After-Sales Services and Customer Relationship Marketing: A multiple case study within the Swedish heavy equipment machinery industry*. Linnaeus University, School of Business and Economics, Department of Marketing.

<http://lnu.diva-portal.org/smash/record.jsf?pid=diva2%3A635620&dswid=852>

Garzón, J. I. (2018, November 6). *Cómo usar redes neuronales (LSTM) en la predicción de averías en las máquinas*. GFT Blog English. Retrieved November 21, 2022, from <https://blog.gft.com/es/2018/11/06/como-usar-redes-neuronales-lstm-en-la-prediccion-de-averias-en-las-maquinas/>



- Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, 37(3), 424-438.
<https://doi.org/10.2307/1912791>
- Graves, A. (2015, August 27). *Understanding LSTM Networks -- colah's blog*. Colah's Blog. Retrieved November 21, 2022, from
<http://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- Hofmann, M., Neukart, F., & Bäck, T. (2017). *Artificial Intelligence and Data Science in the Automotive Industry*. <https://doi.org/10.48550/arXiv.1709.01989>
- Marcial Flores, A., & Ortiz Velázquez, S. (2018, mayo). Acumulación y financiarización en la industria automotriz de México. *Ola Financiera*, 11(30), 118 - 138.
<https://doi.org/10.22201/fe.18701442e.2018.30.65519>
- Maturana, F., Morales, M., Sepúlveda, U., & Maldonado, J. (2022, enero). Ciudades intermedias y parque automotriz en Chile. ¿En jaque la sustentabilidad territorial? *ÍCONOS Revista de Ciencias Sociales*, XXVI(72), 139 - 159.
<https://doi.org/10.17141/iconos.72.2022.4866>
- Michalewicz, Z., Baeck, T., & Fogel, D.B. (Eds.). (1997). *Handbook of Evolutionary Computation*. Taylor & Francis.
- Parra Caballero, C. (2016). *Relación entre el servicio postventa y la fidelidad del cliente en el sector automotriz, el caso de dos concesionarios de vehículos chinos en Lima*. Pontificia Universidad Católica del Perú. Facultad de Gestión y Alta Dirección. <http://hdl.handle.net/20.500.12404/12130>
- Sotaquirá, M. (2019, July 20). *¿Qué son las Redes LSTM?* Codificando Bits. Retrieved November 21, 2022, from <https://www.codificandobits.com/blog/redes-lstm/>



Taylor, S. J., & Letham, . (2017, September 27). Forecasting at scale. *Practical Data Science for Stats*. 10.7287/peerj.preprints.3190v2

Vallarino, D. (2016, diciembre). Econometría versus Machine Learning: La lucha por los datos. *Forbes México*.
<https://www.forbes.com.mx/econometria-versus-machine-learning/>