

ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL



Facultad de Ingeniería en Electricidad y Computación

PROYECTO DE TITULACIÓN

Previo a la obtención del Título de:

Magister en Ciencia de Datos

TEMA:

PRONÓSTICOS DE VENTAS MEDIANTE TÉCNICAS CONVENCIONALES DE SERIES TEMPORALES PARA UNA EMPRESA DISTRIBUIDORA DE PRODUCTOS AL POR MAYOR Y POR MENOR.

Presentado por:

RICARDO ANDRÉS GARCÍA SEGOVIA

GUAYAQUIL – ECUADOR

Año: 2023

RESUMEN

El uso de los pronósticos en las empresas se vuelve más recurrente en los años más recientes. Esto como resultado de nuevas tecnologías de información, del procesamiento incremental de las computadoras y del desarrollo de software especializados que generan pronósticos. Dicho esto, las empresas cuentan con más facilidades para el uso de pronósticos en sus operaciones.

La empresa “AAA” cuyo giro de negocio se basa en la distribución mayorista y minorista de productos y utiliza los pronósticos para los procesos de planeación comercial. De este proceso de planeación se derivan las asignaciones de presupuesto. Por ende, el uso de los pronósticos con lleva un gran impacto en la organización en todos los niveles. Con estos pronósticos, se realiza la compra de inventarios, se contrata personal, se adquieren nuevos activos y servicios.

En este proyecto realizado para la empresa en mención, se ha generado pronósticos para cada una de las unidades de negocios y para cada uno de sus productos mediante el uso de técnicas convencionales de pronósticos para series temporales. A estos modelos, se han incorporado variables que han aportado para significativamente para obtener mejores resultados en las métricas de evaluación RMSE y MAPE.

Finalmente, se ha realizado un modelo de datos en una herramienta de inteligencia de negocios que integra los pronósticos con los datos de venta real, así como también con todas las dimensiones asociadas tal como: unidad de negocios, ciudad y producto. Con ello, el usuario final podrá utilizar la herramienta de forma interactiva.

ABSTRACT

The use of forecasts in companies has become more recurrent in recent years. This is due to the development of new information technologies, the processing capacity of computers and the development of specialized software that generate forecasts. That said, companies have more facilities for the use of forecasts in their operations.

The company "AAA" whose line of business is based on the wholesale and retail distribution of products uses the forecasts for the commercial planning processes. Budget allocations are derived from this planning process. Therefore, the use of forecasts has a great impact on the organization at all levels. With these forecasts, inventories are purchased, personnel are hired, and new assets and services are acquired.

In this project, carried out for the company in question, forecasts have been generated for each of the business units and for each of their products through the use of conventional forecasting techniques for time series. Variables have been incorporated into these models that have significantly contributed to obtain better results in the RMSE and MAPE evaluation metrics.

Finally, a data model has been made in a business intelligence tool that integrates the forecasts with the actual sales data as well as with all the associated dimensions such as: business unit, city and product. With this, the end user will be able to use the tool interactively.

DEDICATORIA

Dedico esta trabajo de titulación a mi familia. en especial a Roxana quien es mi inspiración para seguir adelante día a día.

AGRADECIMIENTOS

A mi madre y a mi familia que me han apoyado para culminar esta maestría aun cuando las circunstancias estuvieron cuesta arriba.

De igual forma agradezco a mis profesores de la maestría quienes me han dado todas las herramientas para poder desarrollar esta tesis. Así como también a mi compañero Alejandro Obando quien ha sido un gran apoyo a lo largo de este camino.

Ricardo Garcia

DECLARACION EXPRESA

“Los derechos de titularidad y explotación, me corresponde conforme al reglamento de propiedad intelectual de la institución; Yo Ricardo Andrés García Segovia doy consentimiento para que la ESPOL realice la comunicación pública de la obra por cualquier medio con el fin de promover la consulta, difusión y uso público de la producción intelectual”

Ricardo Andrés García Segovia

Autor

COMITÉ EVALUADOR

Dr. Sergio Bauz

PROFESOR TUTOR

Mgtr. Eduardo Cruz

PROFESOR EVALUADOR

ABREVIATURAS

ÍNDICE GENERAL

CAPÍTULO 1	8
1. PLANTEAMIENTO DE LA PROBLEMÁTICA	8
1.1 Descripción del Problema	8
1.2 Justificación	9
1.3 Objetivos	10
1.3.1 Objetivo General	10
1.3.2 Objetivos Específicos	11
1.4 Metodología:	11
1.5 Resultados Esperados:	11
1.6 Dataset:	12
CAPÍTULO 2	13
2. MARCO TEÓRICO Y ESTADO DEL ARTE	13
2.1 Introducción a los pronósticos	13
2.2 Clasificación de métodos y tipos de pronósticos.	14
2.3 Las series temporales	16
2.3.1 Definición	16
2.3.2 Características de series de tiempo	16
2.3.2 Clasificación de series de tiempo	17
2.3.3 Autocorrelación	17
2.3.4 Ruido Blanco	19
2.3.5 Componentes de las series de tiempo	20
2.3.6 Métodos de descomposición	21
2.4 Modelos convencionales de pronósticos en series de tiempo.	22
2.4.2 Modelos ARIMA	29

2.5 Modelos de Regresión en series temporales.....	33
2.6 Métricas de evaluación de pronósticos	34
CAPÍTULO 3	37
3. DISEÑO E IMPLEMENTACIÓN	37
3,1 introducción.....	37
3.2 El modelo de arquitectura tecnológica	39
3.3 Almacenamiento de los datos.....	40
3.4 Obtención de datos.....	41
3.5 Preprocesamiento y validación de datos.....	41
3.6 Generación de pronósticos y selección del mejor modelo.....	42
3.7 Diseño de Dashboard en Power BI para análisis de resultados.....	43
CAPÍTULO 4	45
4. ANÁLISIS DE RESULTADOS	45
4.1 Antecedentes	45
4.2 El dataset	45
4.3 Análisis Exploratorio de datos	46
4.3.1 UEN #1.....	48
4.3.2 UEN #2.....	56
4.3.3 UEN #3.....	61
4.3.4 UEN #4.....	64
4.4 Preprocesamiento de datos.....	67
4.2.1 Preprocesamiento: UEN #2.....	68
4.2.2 Preprocesamiento: UEN #3.....	68
4.5 Selección del mejor modelo.	69
4.4.1 Selección del modelo: UEN #1.....	70
4.4.2 Selección del modelo: UEN #2.....	73

4.4.3 Selección del modelo: UEN #3.....	76
4.4.4 Selección del modelo: UEN #4.....	79
4.6 Generación de pronósticos.....	80
4.7 Prototipo.....	81
5. CONCLUSIONES Y RECOMENDACIONES.....	86
5.1 CONCLUSIONES.....	86
5.2 RECOMENDACIONES.....	88
6. REFERENCIAS.....	89
7. ANEXOS.....	91

ÍNDICE DE FIGURAS

Figura 1	<i>Correlograma de autocorrelación: estacionalidad</i>	18
Figura 2	<i>Correlograma de autocorrelación: tendencia y estacionalidad</i>	19
Figura 3	<i>Correlograma de autocorrelación: ruido blanco</i>	19
Figura 4	<i>Forma de comportamiento del componente de tendencia</i>	20
Figura 5	<i>Forma de comportamiento del componente de estacionalidad</i>	21
Figura 6	<i>Forma de comportamiento del componente de residuos</i>	21
Figura 7	<i>Modelos de errores aditivos</i>	28
Figura 8	<i>Modelos de error multiplicativos</i>	29
Figura 9	<i>Esquema del diseño del proyecto de implementación</i>	38
Figura 10	<i>Modelo de arquitectura tecnológica</i>	39
Figura 11	<i>Modelo de Entidad Relación</i>	40
Figura 12	<i>Modelo de Entidad Relación pronóstico prototipo</i>	44
Figura 13	<i>Participación de ventas por Unidad de negocios</i>	46
Figura 14	<i>Detalle de cobertura y portafolio de productos por Unidad de negocios</i>	47
Figura 15	<i>Evolutivo por ciudad UEN #1</i>	48
Figura 16	<i>Evolutivo semanal ciudad UEN #1</i>	49
Figura 17	<i>Correlación de ventas entre ciudades UEN #1</i>	49
Figura 18	<i>Evolutivo por productos UEN #1</i>	50
Figura 19	<i>Correlación de ventas entre productos UEN #1</i>	51
Figura 20	<i>Evolutivo por semana producto UEN #1</i>	51
Figura 21	<i>Estacionalidad semanal: Unidad de negocios #1</i>	52
Figura 22	<i>Gráfico de autocorelación UEN #1</i>	53
Figura 23	<i>Descomposición temporal producto #1</i>	54
Figura 24	<i>Descomposición temporal producto #2</i>	54

Figura 25 <i>Descomposición temporal producto #3</i>	55
Figura 26 <i>Descomposición temporal producto #4</i>	56
Figura 27 <i>Evolutivo mes por productos – correlación UEN #2</i>	57
Figura 28 <i>Evolutivo semanal por producto UEN #2</i>	57
Figura 29 <i>Estacionalidad Semanal: UEN #2</i>	58
Figura 30 <i>Gráfico de autocorrelación UEN #2</i>	59
Figura 31 <i>Descomposición temporal producto # 5</i>	60
Figura 32 <i>Descomposición temporal productos # 6 y 7</i>	60
Figura 33 <i>Evolutivo mes por productos UEN #3</i>	61
Figura 34 <i>Correlación día y mensual productos UEN#3</i>	62
Figura 35 <i>Evolutivo semanal productos UEN# 3</i>	62
Figura 36 <i>Estacionalidad semanal: UEN #3</i>	63
Figura 37 <i>Descomposición productos UEN #3</i>	64
Figura 38 <i>Evolutivo mensual UEN #4</i>	64
Figura 39 <i>Evolutivo semanal UEN #4</i>	65
Figura 40 <i>Estacionalidad semanal UEN #4</i>	66
Figura 41 <i>Gráfico de autocorrelación UEN #4</i>	66
Figura 42 <i>Descomposición unidad de negocios #4</i>	67
Figura 43 <i>Imputación de valores atípicos UEN #2</i>	68
Figura 44 <i>Imputacion de valores atipicos UEN #3</i>	68
Figura 45 <i>Pronósticos generados por TSLM para productos #1 y #4</i>	72
Figura 46 <i>Pronósticos generados por ARIMA para productos #2 y #3</i>	73
Figura 47 <i>Pronósticos generados por TSLM para productos #5 y #6</i>	76
Figura 48 <i>Pronósticos generados por ETS para productos #7</i>	76
Figura 49 <i>Pronósticos generados por ARIMA CON REGRESORES productos #8 y #9</i>	78

Figura 50 <i>Pronósticos generados por ETS producto #10</i>	80
Figura 51 <i>Pantalla de identificación del usuario</i>	81
Figura 52 <i>Pantalla de error de contraseña</i>	82
Figura 53 <i>Pantalla de Índice</i>	82
Figura 54 <i>Pantalla de PRONÓSTICOS</i>	83
Figura 55 <i>Pantalla de PRONÓSTICOS filtrado por mes en curso</i>	85

ÍNDICE DE TABLAS

Tabla 1 <i>Criterios de selección de modelos de series temporales y aprendizaje supervisado</i>	15
Tabla 2 <i>Modelos combinados tendencia y estacionalidad</i>	27
Tabla 3 <i>Tabla de variables con tipo de datos</i>	45
Tabla 4 <i>Evaluación RMSE día UEN #1</i>	70
Tabla 5 <i>Evaluación MAPE día UEN #1</i>	71
Tabla 6 <i>Evaluación MAPE mes UEN #1</i>	72
Tabla 7 <i>Evaluación RMSE día UEN #2</i>	74
Tabla 8 <i>Evaluación MAPE día UEN #2</i>	74
Tabla 9 <i>Evaluación MAPE mes UEN #2</i>	75
Tabla 10 <i>Evaluación RMSE día UEN #3</i>	77
Tabla 11 <i>Evaluación MAPE día UEN #3</i>	77
Tabla 12 <i>Evaluación MAPE mes UEN #2</i>	78
Tabla 13 <i>Evaluación RMSE día UEN #4</i>	79
Tabla 14 <i>Evaluación MAPE día UEN #4</i>	79
Tabla 15 <i>Evaluación MAPE mes UEN #4</i>	80

CAPÍTULO 1

1. PLANTEAMIENTO DE LA PROBLEMÁTICA

1.1 Descripción del Problema

A partir de la era digital, gran cantidad de empresas tiene acceso a grandes volúmenes de datos que crecen a un ritmo exponencial. No obstante, contar con esta cantidad de datos no necesariamente significa que estas empresas tengan la capacidad de extraer información valiosa que les permita tomar decisiones acertadas en el futuro (Xindong Wu, 2014).

En consecuencia, las empresas han detectado la necesidad de poder explotar sus datos a fin de evaluar posibles implicaciones futuras y como prepararse oportunamente. El que podría suceder y como prepararse son cuestionamiento que pueden ser resueltas con el uso de analítica predictiva, específicamente con pronósticos (CONSTANTIN, 2016).

Los pronósticos de ventas es una actividad empresarial en la cual las empresas regularmente fallan. La precisión en estos pronósticos permite a las empresas cumplir con las necesidades de todas las partes involucradas dentro de la cadena de valor tanto proveedores como clientes. Por un lado, se puede prever situaciones de quiebre de stock, así como también anticipar que los clientes sean atendidos por la competencia. (Bienstock, 1998)

En la empresa “AAA” los pronósticos, a lo largo del tiempo, no han sido estimados mediante algún modelo matemático confiable. Mas bien, las predicciones de ventas han sido en base al conocimiento empírico del área comercial. Esto ha generado estimaciones no confiables que generan implicaciones en los procesos y operación de la empresa tales como:

- Asignaciones irracionales de presupuesto generando alta rotación de personal de fuerza de ventas.

- Sobre abastecimiento de productos afectando al flujo de efectivo de la compañía.
- Pérdida de clientes por no contar con el suficiente volumen en productos estratégicos especialmente a distribuidores.

Bajo estas circunstancias la empresa aspira a implementar alguna solución tecnológica de bajo coste que permita no solo la generación de pronósticos sino también les permita entender ciertos patrones con los cuales estas proyecciones han sido construidas para que, de esta forma, sea el primer paso de su proceso de planeación.

1.2 Justificación

Un pronóstico de venta consiste en predecir la cantidad de producto que un consumidor va a comprar dado las características propias del producto y las condiciones actuales de venta. Los pronósticos son vitales para una eficiente operación y ayuda a los miembros de la organización a tomar decisiones tales como: construir una nueva planta, la cantidad de inventario en las bodegas, el número de trabajadores que contratar, los salarios a pagar a la fuerza de ventas, etc. (Barker, 1999)

Si bien es cierto, los pronósticos nunca van a ser exactos con la realidad estos nos proveen un escenario substancial y objetivo de lo que puede ocurrir. Por medio de estos, la empresa puede simular escenarios de futuros cambios en el ambiente empresarial para así poder tomar acciones oportunas. (KURZAK, 2012)

Las gerencias, por ejemplo, necesitan constantemente pronosticar del tamaño del mercado o de alguna categoría de producto en específicas. Y, cuando suscitan problemas ellos necesitan los pronósticos para poder alinear las estrategias de cara a los competidores, proveedores, clientes, instituciones y toda parte interesada. (Armstrong, 2012)

Actualmente se encuentran disponibles en el mercado muchas herramientas de pronósticos. Sin embargo, estas herramientas son módulos anexos a grandes ERP que por su cuantía financiera no los hace accesible a la gran mayoría de las empresas.

Habría que decir también que, estos módulos generan pronósticos que responden a modelos matemáticos preestablecidos que se aplican de forma no especializada. Con lo cual, no garantiza que el pronóstico generado sea el mejor posible.

La implementación del presente proyecto tiene como objetivo la creación de una solución tecnológica, por medio de herramientas de código abierto, que permita la generación de pronósticos por medio de algoritmos convencionales para series temporales. Con ello, la empresa no estará condicionada el requerimiento de capital que este tipo de soluciones conllevan.

En consecuencia, el alcance del presente proyecto incluye la creación de un tablero iterativo elaborado en un aplicativo de inteligencia empresarial que permitirá al usuario no solo obtener pronósticos sino también poder analizar la información con el objetivo de la toma acertada de decisiones.

1.3 Objetivos

1.3.1 Objetivo General

Diseñar una herramienta de pronósticos de ventas usando modelos de series Temporales para una empresa de consumo masivo.

1.3.2 Objetivos Específicos

- Estimar pronósticos de ventas generados a partir de técnicas de series temporales convencionales.
- Establecer criterios de selección del mejor modelo de pronóstico de ventas.
- Desarrollar visualizaciones interactivas de los pronósticos de ventas mediante la implementación de un dashboard que ayude a la toma de decisiones oportunas.

1.4 Metodología:

Para el desarrollo del presente proyecto se llevará a cabo la siguiente metodología:

- Se utilizarán una base de datos estructurados de ventas a nivel día de una empresa comercializadora de productos del sector de consumo masivo.
- Posteriormente se realizarán conexiones a la base de datos para poder acceder a los mismos y realizar los procesamientos y análisis iniciales como: Transformaciones, anonimizarían, limpieza, imputaciones, etc., estos pasos serán realizados haciendo uso de programas estadísticos y de código abierto.
- Después, haciendo uso de los mismos programas mencionados en el párrafo anterior se generarán los pronósticos de series de tiempo para cada uno de los productos que hayan sido vendidos por la empresa, para la generación de estos pronósticos se utilizarán las siguientes familias de técnicas convencionales: Métodos simples, Suavizamiento Exponencial y Promedios Móviles Integrados Autorregresivos (ARIMA) y modelos lineales.
- Para finalizar, aquel modelo que genere mayor precisión estadística será utilizado para la generación de pronósticos. Estos pronósticos estarán disponibles en un tablero de inteligencia empresarial que será interactivo para el usuario.

1.5 Resultados Esperados:

Al finalizar el proyecto, la empresa “AAA” tendrá a su disposición un tablero de pronósticos interactivo que estará alojado en el SharePoint de la empresa para el consumo de los usuarios interesados. Este tablero contará con visualizaciones de predicciones de venta que permitirá al usuario no solo a entender los pronósticos sino también el comportamiento histórico de los

productos en cada unidad de negocio. Adicional, este tablero contara con los indicadores de comparación clave de la empresa de tal forma que al usuario podrá fácilmente realizar sus analisis.

Se espera que los pronósticos generados por la herramienta sirvan como base para los procesos comerciales y posteriormente para los procesos operacionales relacionadas al giro del negocio

1.6 Dataset:

Para el presente proyecto se tiene acceso a una base de datos estructurados en la cual se encuentra registrada una muestra de las transacciones de ventas desde enero 2022 hasta marzo 2023, cabe mencionar que estos datos han sido anonimizados a través del uso de etiquetas descriptivas ficticias para los campos con datos categóricos y con el uso de factores de conversión para los campos con datos numéricos, esto con la finalidad de proteger la confidencialidad de la información.

CAPÍTULO 2

2. MARCO TEÓRICO Y ESTADO DEL ARTE

2.1 Introducción a los pronósticos

Los pronósticos sirven como una herramienta que nos indica un estimado de la probabilidad de ocurrencia de eventos futuros (Juárez, 2016). La finalidad de elaborar pronósticos es disminuir la incertidumbre de tal manera que se permita tomar decisiones con menos riesgo.

La teoría de pronosticar está basada en la premisa del que el conocimiento actual y el pasado puede ser utilizado para realizar predicciones en el futuro. De manera particular para las series temporales, existe la creencia que esto es posible dado la identificación de patrones en la data histórica. No obstante, no se espera la precisión exacta en las predicciones de valores futuros más bien, lo que se espera es un valor esperado dentro de intervalos de confianza (Petropoulos, 2022).

Con el desarrollo y la proliferación de los computadores y de su software asociado, la elaboración de pronósticos ha captado cada vez más atención. Por ende, cualquier administrador o usuario está al alcance de utilizar técnicas muy sofisticadas para el análisis de datos asociados a los pronósticos. Habiendo dicho esto, es de gran importancia poder entenderlas ya que, un uso inadecuado de estas da como resultado decisiones erradas (Hawke, 2005).

El pronóstico de ventas, por ejemplo, es la piedra angular de las actividades de planeación (Haberleitner, 2010). El beneficio obtenido de estos pronósticos es obtener un flujo optimizado de productos y/o servicios dentro de la cadena de suministro. En esa misma dirección, (Rabi, 2012) explica como los pronósticos deficientes: reduce la disponibilidad de productos, cambia las preferencias de los clientes hacia los competidores e impacta de forma significativa a las necesidades de capital de la empresa.

Por otro lado, los pronósticos con buena precisión guía a las empresas a ofrecer mejor atención al cliente y a la optimización del inventario. Es por ello que las compañías están en la obligación de cada vez mejorar la precisión de los pronósticos con la finalidad de poder satisfacer asertivamente las necesidades de sus clientes. (Kumar, 2019)

2.2 Clasificación de métodos y tipos de pronósticos.

Dentro del campo de los pronósticos en series temporales, existen dos métodos disponibles los cualitativos y los cuantitativos.

Los métodos cualitativos son útiles cuando no se posee con suficientes datos disponibles. En este tipo de casos, se puede utilizar ya sea una analogía histórica o un consenso de panel. Por un lado, la analogía histórica busca aprovechar la experiencia de los participantes sobre el mercado para incursionar en uno nuevo. Por otro lado, el consenso de panel reúne un grupo de expertos en el tema para poder llegar a una conclusión. (Gallegos, 2014)

Los métodos cuantitativos se pueden realizar cuando existe suficiente información sobre el pasado y cuando se piensa que ciertos patrones del pasado van a seguir sucediendo en el futuro. (Rob J Hyndman, 2021).

Dentro de los métodos cuantitativos podemos detallar dos grandes ramas los métodos convencionales y los métodos de aprendizaje supervisado. Dentro de los métodos convencionales, podemos encontrar los modelos de univariados de series temporales y los modelos de regresión o explicativos.

Por otro lado, los métodos de aprendizaje supervisado nos encontramos con técnicas referentes a la inteligencia artificial tal es el caso del aprendizaje automático y del aprendizaje profundo. La utilización de cualquiera estos modelos van a depender del objetivo del pronosticador, así como también de la disponibilidad de datos

En la Tabla 1 se muestra las pautas de selección del modelo de series temporales y de aprendizaje supervisado. Dada la naturaleza de estos últimos, se recomienda utilizarlos, cuando existe gran cantidad de datos teniendo en cuenta que, el tiempo de procesamiento es menor que los métodos estadísticos tradicionales (Peixeiro, 2022).

Tabla 1

Criterios de selección de modelos de series temporales y aprendizaje supervisado

Criterios de seleccion	Métodos estadísticos Tradicionales	Métodos de aprendizaje supervisado
Numero de observaciones	$n < 10.000$	$N \geq 1000$
Granularidad de la serie	Anual, mensual, trimetral	Diaria – Horas
Relacion entre variables	Lineal	No lineal

Además, se debe agregar que, las ecuaciones que componen los modelos estadísticos tradicionales asumen linealidad entre las variables y normalidad en sus residuos. Si no se cumplen los supuestos mencionados, quiere decir que el método tradicional no se ajusta a la data. En dichos escenarios, se recomienda el uso de los métodos de aprendizaje supervisado **(Peixeiro, 2022)**.

En contraste con lo mencionado, los métodos estadísticos tradicionales cuentan con gran rendimiento si la granularidad de la serie es mensual, trimestral o anual. También, cuentan con buen rendimiento sobre data día siempre que el tamaño de la muestra no sea tan grande. Dentro de estos métodos, los modelos más utilizados son los de suavizado exponencial y ARIMA. Ambos modelos cuentan con variaciones que toman en consideración los diferentes componentes de la serie de tiempo. Lo cual, hace de estos modelos precisos. (Rob J Hyndman, 2021)

2.3 Las series temporales

2.3.1 Definición

Se denomina series de tiempo a la secuencia de observaciones en determinados momentos del tiempo, ordenados secuencial y cronológicamente. Las observaciones de la serie temporal están denotadas por Y_1, Y_2, \dots, Y_t donde Y_t es el valor de la serie en el instante t . (Hurtado, 2008)

2.3.2 Características de series de tiempo

Las series de tiempo muestran ciertas características o patrones a lo largo del tiempo que, es necesario identificar previo a la selección de algún modelo ya que, habiendo detectado estos patrones, será mucho más fácil escoger un modelo que modele estos comportamientos (Koehler, 2008). Estos patrones son: Tendencia, estacionalidad y ciclo.

- Patrón tendencial: La tendencia existe cuando hay un crecimiento o decrecimiento prologado de la serie de tiempo. en muchas ocasiones se refiere como tendencia al cambio de dirección de la serie sin embargo estos cambios están dentro del comportamiento tendencial.
- Patrón Estacional: la estacionalidad ocurre cuando factores estacionales influyen en nuestra serie. La estacionalidad regularmente es fija y se encuentra dentro de un periodo conocido. Por ejemplo, esta podría ser mensual o los días de la semana.

- Patrón cíclico: El patrón cíclico ocurre cuando la serie de tiempo presenta un crecimiento o decrecimiento que no responde a una frecuencia fija. Este patrón es usualmente relacionado con la estacionalidad, pero son totalmente diferentes. El patrón cíclico está relacionado las condiciones económicas y el ciclo del negocio y su duración es aproximadamente 2 años.

2.3.2 Clasificación de series de tiempo

En cuanto a la clasificación de series temporales estas, dependiendo de las características propias de su media, covarianza y varianza a través del tiempo estas se denomina estacionarias o no estacionarias.

- **Las series no estacionas** son aquellas cuya media, varianza y covarianza cambian a través del tiempo. Los cambios en la media originan una tendencia crecente o decreciente con lo cual no la vuelve constante.
- **Las series estacionarias** son aquellas cuya media y varianza oscilan constantemente a través del tiempo.

2.3.3 Autocorrelación

Prosiguiendo nuestro análisis, es importante identificar dentro de las series de tiempo, como esta se encuentra relacionada entre si con sus periodos pasados (Fulton, 2017). Esta distinción nos permitirá ajustar nuestro modelo con la finalidad de obtener resultados con mejor precisión.

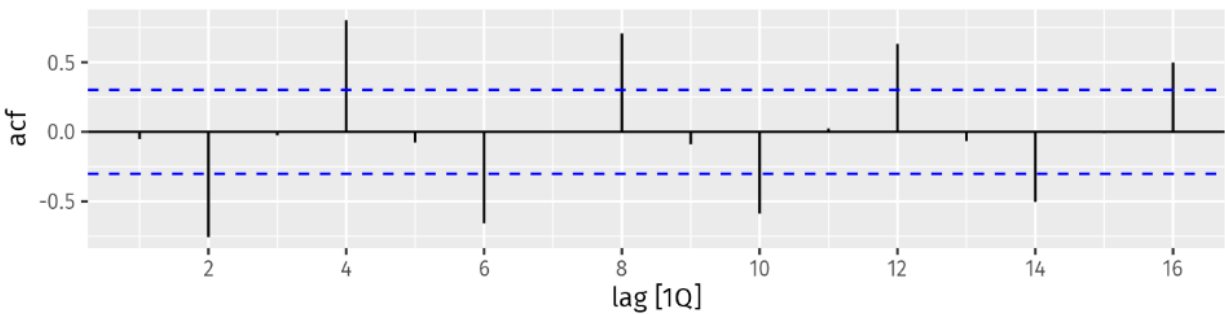
El concepto de auto correlación hace referencia a la relación lineal que existe entre el valor de la serie y sus periodos previos (Rob J Hyndman, 2021). El grado de autocorrelación es representado en la siguiente ecuación:

$$r_k = \frac{\sum_{t=k+1}^t (y_t - \hat{y})(y_{t-k} - \hat{y})}{\sum_{t=1}^t (y_t - \hat{y})^2} \quad (1)$$

Gráficamente, se puede detectar la autocorrelación mediante un *correlograma* en donde cada barra representa el valor de r_k de sus valores rezagados. En el ejemplo de la figura 1 se muestra una serie de corte trimestral la cual muestra una estacionalidad anual.

Figura 1

Correlograma de autocorrelación: estacionalidad

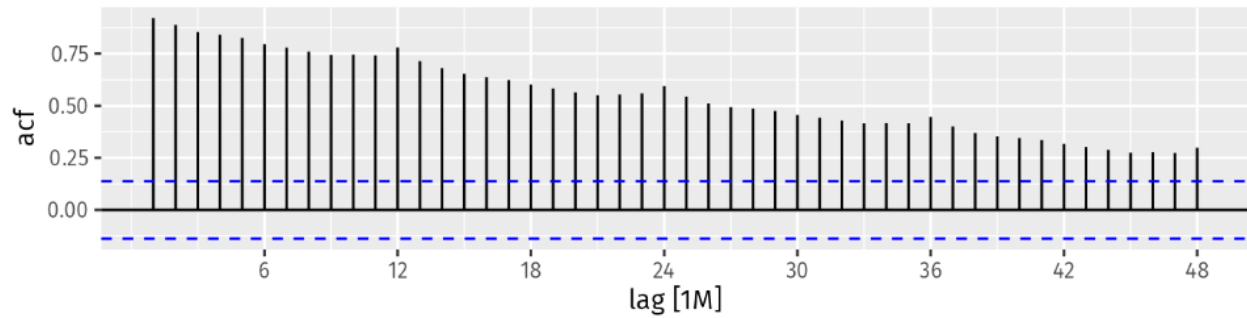


Nota: Tomada de (Rob J Hyndman, 2021)

Cuando la serie presenta tendencia, la autocorrelación es más alta en sus periodos más próximos la cual decae para sus periodos más antiguos. Así mismo, cuando la serie presenta estacionalidad, la autocorrelación es alta en periodos múltiplos de su estacionalidad. Estos conceptos se pueden graficar en la figura. Donde, se presenta una serie mensual con tendencia y estacionalidad cada año. Para este ejemplo, la estacionalidad se encuentra en los periodos múltiplos del 12.

Figura 2

Correlograma de autocorrelación: tendencia y estacionalidad



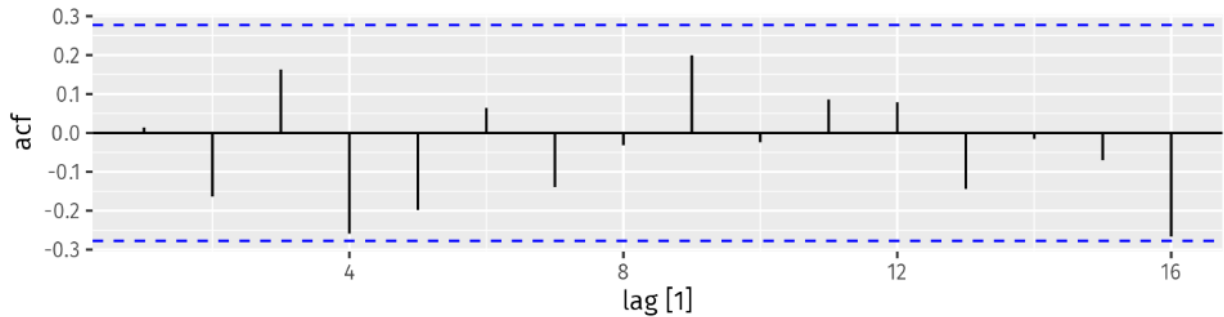
Nota: Tomada de (Rob J Hyndman, 2021)

2.3.4 Ruido Blanco

Por otro lado, cuando la serie no presenta autocorrelación, se la denomina “Ruido blanco” o “Estacionarias”. Estas series no presentan ni tendencia ni estacionalidad. Estas series que no presentan ninguno de los dos. En la figura 3 se puede observar un correlograma de una serie de ruido blanco en la cual la correlación existente entre sus periodos rezagados se encuentra cercanos a 0.

Figura 3

Correlograma de autocorrelación: ruido blanco



Nota: Tomada de (Rob J Hyndman, 2021)

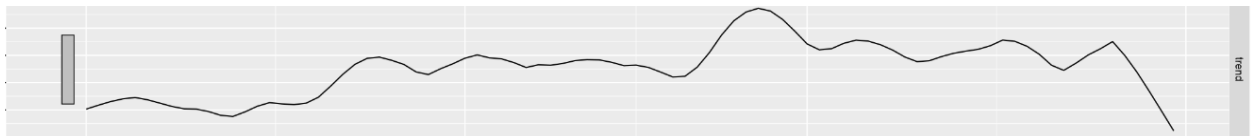
2.3.5 Componentes de las series de tiempo

Los datos de series de tiempo pueden mostrar una variedad de patrones como los mencionados en la sección 2.4. No obstante, cuando descomponemos una serie, usualmente tendencia esta combinada con el patrón cíclico lo cual se lo conoce como el componente tendencial – cíclico (Tsay, 2005). Dicho eso, una serie de tiempo está compuesta por un componente: tendencial – cíclico, estacional y residuos.

Tendencia – ciclo: El componente de tendencial cíclico es la parte de la serie de tiempo que detalla el cambio de la media a largo plazo. Este componente tendencial no es afectado por las fluctuaciones de corto plazo ni por estacionalidades. Una creciente o disminución de este componente se puede explicar por cambios en la economía, apertura o cierre de puntos de venta, lanzamiento de nuevos productos, etc

Figura 4

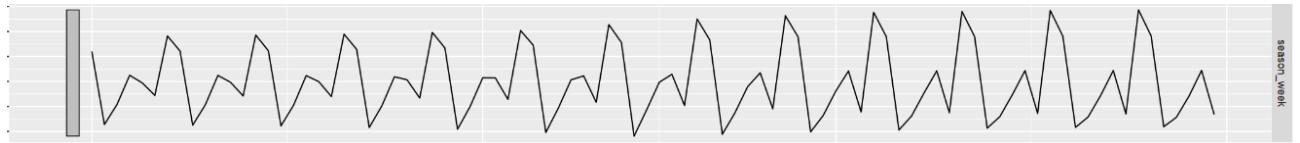
Forma de comportamiento del componente de tendencia



Estacionalidad: Este componente de la serie hace referencia a las fluctuaciones similares continuas dentro de periodos iguales o menores a un año. Las series temporales pueden contar con múltiples estacionalidades principalmente si la data es de granularidad diaria u horaria. Por ejemplo, el consumo de luz puede tener una estacionalidad en las horas pico dentro del día, puede también tener una estacionalidad un día de la semana, así como también dentro de una época del año. la figura 5 se muestra un ejemplo del componente estacional

Figura 5

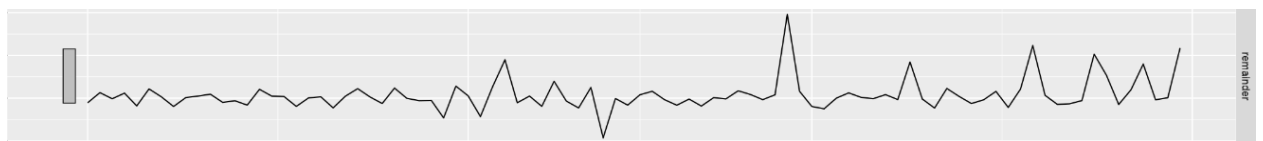
Forma de comportamiento del componente de estacionalidad



Residuos: Los residuos es el componente de la serie de tiempo que no puede ser explicada por el componente tendencial ni por el componente cíclico. Estos residuos serán de gran importancia en la selección de modelos ya que si estos muestran autocorrelación el modelo no será el adecuado. En la figura 6 se puede visualizar el componente residual

Figura 6

Forma de comportamiento del componente de residuos



2.3.6 Métodos de descomposición

Examinemos brevemente ahora los métodos de descomposición usualmente implementados. Dependiendo a los patrones que muestre la serie estos métodos pueden ser aditivo o multiplicativo (Rob J Hyndman, 2021).

El método de descomposición aditiva es aplicable cuando los patrones de la serie de tiempo se muestran constante a través del tiempo y es simplemente la sustracción los componentes tendenciales cíclico, estacional y los residuos dada la siguiente expresión:

$$y = T + S + R \quad (2)$$

Donde

T: Tendencia.

S: Estacional.

R: Residuos.

Por otro lado, el método de descomposición multiplicativa es aplicable cuando la serie temporal presenta patrones no lineales y cambios abruptos en el tiempo. En estos casos la descomposición de la serie temporal se la obtiene dividiendo los componentes sobre la serie original dada la siguiente expresión:

$$y = T x S x R \quad (3)$$

2.4 Modelos convencionales de pronósticos en series de tiempo.

2.4.1 Suavizado Exponencial

Los modelos de suavizado exponencial fueron propuesto a finales de los años 50 (**Brown, 1959; Holt, 1957; Winter, 1960**). Los pronósticos generados por estos métodos son, básicamente, promedios ponderados de observaciones pasadas, donde estos pesos caen exponencialmente cada vez que las observaciones son más antiguas. Lo importante en la aplicación de estos métodos es el reconocimiento de los componentes de la serie (tendencia y estacionalidad). En el siguiente punto abordaremos todas las variantes de este método.

En relación con lo mencionado en el punto que antecede, las variantes de los modelos de suavizado exponencial dependerá de los componentes de la serie. En suma, podemos dividir los modelos de suavizado exponencial en tres grupos: métodos simples, métodos de tendencia y métodos estacionales. Cada uno de ellos, cuentan con diferentes variantes.

2.4.1.1 Método simple.

Este método es ideal para pronosticar series de tiempo donde no se distinga claramente el patrón de tendencia o estacionalidad. La ecuación 1 muestra la forma de cálculo del método simple:

$$\hat{y}_{t+1/t} = \alpha y_t + \alpha(1 - \alpha)y_{t-1} + \alpha(1 - \alpha)^2 y_{t-2} + \dots \quad (4)$$

Se puede observar que los periodos pasados se multiplica con un parámetro de suavizado (α). Donde $0 \leq \alpha \leq 1$. Si el valor del parámetro es cercano a 1 se da mayor importancia a los periodos recientes. Si el valor del parámetro es cercano a 0 se da menos importancia a periodos recientes. Además, por cada nivel o periodo pasado el parámetro se va elevando n veces con lo cual, los periodos iniciales pierden relevancia para el pronóstico.

Para la generación de pronósticos este método cuenta con solo un componente y este es el nivel (l_t). Donde el nivel es la ecuación suavizada.

$$\text{Ecuación para el pronóstico: } \hat{y}_{t+h/t} = l_t \quad (5)$$

$$\text{Ecuación de suavizamiento: } l_t = \alpha y_t + (1 - \alpha)l_{t-1} \quad (6)$$

2.4.1.2 Métodos de tendencia (Holt).

En 1957, Holt amplió el método simple de tal forma que permitió incorporar la tendencia en los pronósticos. La fórmula para el cálculo de los valores ajustados es:

$$\text{Ecuación para el pronóstico: } \hat{y}_{t+h/t} = l_t + hb_t \quad (7)$$

$$\text{Ecuación de Nivel: } l_t = \alpha y_t + (1 - \alpha) (l_{t-1} + b_{t-1}) \quad (8)$$

$$\text{Ecuación de Tendencia: } b_t = \beta * (l_t - l_{t-1}) + (1 - \beta) b_{t-1} \quad (9)$$

Donde β^* es el parámetro de suavizado para la tendencia y toma valores entre 0 y 1. Se puede observar que la ecuación para la generación de pronósticos ya no es aplanada, sino que incorpora un componente de tendencia.

2.4.1.3 Método de tendencia amortiguada.

En relación con el punto anterior, se puede observar que la formula incorpora un parámetro β^* lineal para tendencia. Esto genera un problema para los pronósticos de largo plazo ya que tiende a sobre estimarlo. Una alternativa para este problema es amortiguar dicha tendencia incluyendo un parámetro (ϕ) que toma valores entre 0 y 1 y que se multiplica con el parámetro β^* .

$$\hat{y}_{t+h/t} = l_t + (\phi + \phi^2 + \dots + \phi^h)b_t \quad (10)$$

$$l_t = \alpha y_t + (1 - \alpha)(l_{t-1} + \phi b_{t-1}) \quad (11)$$

$$b_t = \beta(l_t - l_{t-1}) + (1 - \beta)\phi b_{t-1} \quad (12)$$

2.4.1.4 Métodos de estacionalidad (Holt- Winters).

Holt (1957) y Winters (1960) extiende el método de tendencia con el objetivo de tomar en consideración la estacionalidad. Con ello, para la ecuación anterior se añade un nuevo componente (st). Este componente incluye nivel y tendencia que se multiplica con un parámetro (γ).

En esta notación se utiliza (m) para referenciar la frecuencia de la estacional. Para series trimestrales $m = 4$, mensuales $m = 12$, etc.

Este método tiene dos variaciones: aditivo y multiplicativo. Se utiliza el aditivo cuando las variaciones estacionales son constantes a través de del tiempo. Por otro lado, el multiplicativo es útil cuando las variaciones estacionales crecen proporcionalmente a través del tiempo.

2.4.1.4.1 Método aditivo:

$$\hat{y}_{t+h/t} = l_t + hb_t + S_{t+h-m(k+1)} \quad (13)$$

$$l_t = \alpha(y_t - S_{t-m}) + (1 - \alpha)(l_{t-1} + b_{t-1}) \quad (14)$$

$$b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1} \quad (15)$$

$$s_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m} \quad (16)$$

2.4.1.4.2 Método multiplicativo:

$$\hat{y}_{t+h/t} = (l_t + hb_t)S_{t+h-m(k+1)} \quad (17)$$

$$l_t = \alpha \frac{y_t}{s_{t-m}} + (1 - \alpha)(l_{t-1} + b_{t-1}) \quad (18)$$

$$b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1} \quad (19)$$

$$s_t = \gamma \frac{y_t}{(l_{t-1} + b_{t-1})} + (1 - \gamma)s_{t-m} \quad (20)$$

2.4.1.5 Innovaciones para el método de suavizado exponencial.

En las secciones anteriores, se ha mencionado diferentes tipos de variantes para el modelo de suavizado exponencial. Desde modelos que solo incluyen nivel hasta aquellos que incorporan tendencia y estacionalidad. En total, se podría resumir estos modelos en 9 diferentes combinaciones.

Tabla 2

Modelos combinados tendencia y estacionalidad

Componente de tendencia	Componente de estacionalidad		
	N	A	M
	(Ninguno)	(Aditivo)	(Multiplicativo)
N (Ninguno)	(N, N)	(N, A)	(N, M)
A (Aditivo)	(A, N)	(A, A)	(A, M)
Ad (Aditivo Atenuado)	(Ad, N)	(Ad, A)	(Ad, M)

Sin embargo, dentro de estos modelos no se ha incluido el componente del “error” o “residuo” que, como se menciona en la sección 2.3, es parte de la composición de una serie de tiempo. En esa misma dirección, incorporar el componente del error nos permitirá no solo obtener el pronóstico sino también intervalos de predicción. A estos modelos de suavizado exponencial con error incluido se los conoce como ETS. Para cada método existen dos modelos de error aditivos y multiplicativos.

Los residuos es la diferencia entre el valor real y el valor ajustado. Si el modelo, incorpora errores aditivos estos se suman en la ecuación. Por otra parte, si se incorporan errores multiplicativos estos se multiplican en la ecuación.

Dependiendo del método que se utilice ya se simple, con tendencia, estacional, tendencia y estacional, etc. La ecuación para el cálculo del error varia. Si, se utiliza un modelo simple entonces el error es $e_t = y_t - L_{t-1}$. Si, se utiliza un modelo con tendencia entonces el error es $e_t = y_t - L_{t-1} - b_{t-1}$.

A continuación, se detalla todas las posibles combinaciones de modelos ETS que en total suman 18.

Nueve combinaciones para errores aditivos y nueve combinaciones para errores multiplicativos. En la figura 7 se detallan los modelos con errores aditivos. Por otro lado, En la figura 8 se presentan los modelos con errores multiplicativos

Figura 7

Modelos de errores aditivos

	N	A	M
N	$y_t = \ell_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$	$y_t = \ell_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = \ell_{t-1}s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/\ell_{t-1}$
A	$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$	$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1})s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1} + \beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + b_{t-1})$
A _d	$y_t = \ell_{t-1} + \phi b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$	$y_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = \phi b_{t-1} + \beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + \phi b_{t-1})$

Nota: Tomada de (Rob J Hyndman, 2021)

Figura 8

Modelos de error multiplicativos

Trend	Seasonal		
	N	A	M
N	$y_t = \ell_{t-1}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$	$y_t = (\ell_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \alpha(\ell_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + s_{t-m})\varepsilon_t$	$y_t = \ell_{t-1}s_{t-m}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
A	$y_t = (\ell_{t-1} + b_{t-1})(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1})s_{t-m}(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
A _d	$y_t = (\ell_{t-1} + \phi b_{t-1})(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m}(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$

Nota: Tomada de (Rob J Hyndman, 2021)

2.4.2 Modelos ARIMA

Los modelos ARIMA y de suavizado exponencial son los dos enfoques más utilizados dentro de los métodos tradicionales. Mientras que el suavizado exponencial se enfoca en la describir y descomponer la serie en tendencia y estacionalidad. Los modelos ARIMA describen la auto correlación en la serie (Rob J Hyndman, 2021)

2.4.2.1 Supuestos de los modelos ARIMA

Para la aplicación de modelos ARIMA se deben cumplir tres condiciones: (1) la serie de tiempo debe ser univariada (2) deben ser estacionarios y (3) los residuos no deben estar correlacionados.

La estacionariedad es una propiedad de las series de tiempo en el cual su media y varianza son constantes a través del tiempo. Formalmente, para la aplicación de estos modelos la serie temporal no debe tener raíz unitaria. Uno de los contrastes más utilizados es la prueba de Dickey-Fuller en donde:

H_0 : La serie es no estacionaria, es decir, tiene raíz unitaria

H_1 : La serie es estacionaria, es decir, no tiene raíz unitaria

Si la serie no es estacionaria, entonces se debe diferenciar y/o transformar la serie. Por un lado, diferenciado la serie se obtiene que las medias sean igual. Este proceso se logra restando $y_t - y_{t-1}$.

Por otro lado, la serie se transforma con la finalidad de igualar las varianzas. Las transformaciones de Box y Cox son utilizadas para realizar este procedimiento.

$$Y_i^{(\lambda)} = \begin{cases} K_1(Y_i^\lambda - 1) & \text{si } \lambda \neq 0 \\ k_2 \ln(Y_i) & \text{si } \lambda = 0 \end{cases} \quad (21)$$

Por otro lado, los residuos no deben estar correlacionados. Estos deben ser independientes, aleatorios y normalmente distribuidos $(0, \sigma)$. Si esto no ocurre, significa que habría información en los residuos que el modelo es incapaz de captar. Lo cual implica que existe un sesgo en los pronósticos.

2.4.1.2 Variantes del modelo ARIMA

2.4.1.2.1 Modelo auto regresivo AR(p)

Un modelo auto-regresivo es una combinación lineal donde las variables predictoras vienen dados por los valores pasados de la serie de tiempo. Siendo así, un modelo auto regresivo de orden p puede ser descrito como:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t \quad (22)$$

Donde c es una constante, ϕ es el parámetro que multiplica a las variables rezagadas. Y (ε_t) es el termino de error o ruido blanco.

2.4.1.2.2 Modelo media móvil MA (q)

Por otro lado, una modelo media móvil es una combinación lineal donde las variables predictoras vienen dadas por los errores de los pronósticos de las variables rezagadas. La modelo media móvil de orden q puede ser descrito como:

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (23)$$

Donde c es una constante, θ es el parámetro que multiplica a los errores rezagadas. Y (ε_t) es el termino de error o ruido blanco.

2.4.1.2.3 Modelo ARMA (p, q)

Un modelo autor regresivo media móvil, es la combinación de los dos procesos descritos previamente. Donde (p) es la porción de autor regresiva y (q) es la porción media móvil. La ecuación matemática de un modelo ARMA (p, q) puede ser descrito como la combinatoria lineal de una constante, los valores rezagados de la serie Y_{t-p} , los errores pasados (ε_{t-q}) , y los errores actuales (ε_t) .

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (24)$$

2.4.1.2.4 Modelo ARIMA (p, d, q)

Los modelos descritos anteriormente se pueden utilizar en el supuesto que la serie presente un comportamiento estacionario. El parámetro d, indica el número de veces que la serie temporal debe diferenciarse para hacerla estacionaria. La expresión matemática de los modelos ARIMA es similar a los ARMA (p, q), con la diferencia que los valores rezagados no serán de la serie original sino de la serie diferenciada n veces. La serie diferenciada puede representarse como (y't), de esta forma, la ecuación sería:

(25)

$$y_t = c + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

2.4.1.2.5 Modelos SARIMA (p, d, q) (P, D, Q) m

Hasta ahora hemos descrito las técnicas de pronósticos para series estacionarias y no estacionarias. Sin embargo, las técnicas descritas asumen que la serie no cuenta con un componente estacional. El modelo SARIMA es una extensión de los modelos ARIMA que incorpora dentro de su ecuación parámetros y variables estacionales.

Los parámetros del patrón estacional vienen dados por las siglas (P, D, Q) m. Donde estos parámetros cumplen los mismos criterios descritos previamente con la diferencia que hacen referencia a los valores estacionales.

La ecuación del modelo:

$$(1 - \phi B)(1 - \phi B^s)(1 - B)(1 - B^s)yt = (1 - \phi B)(1 - \phi B)et \quad (26)$$

2.5 Modelos de Regresión en series temporales

Los modelos explicados en las secciones previas son ecuaciones lineales donde las variables explicativas son componentes de la propia serie. Estos modelos no aceptan variables exógenas que, en muchas ocasiones, son de mucha utilidad para poder explicar a la serie de tiempo (B.Kedem, 2005). Por ejemplo, si quisiéramos pronosticar el consumo de energía eléctrica podríamos considerar el nivel de temperatura como variable explicativa o el día de la semana. Los modelos de regresión en series de tiempo solucionan este problema. La ecuación de estos modelos viene dada de la siguiente forma:

$$y_t = B_0 + B_1X_1 + B_lX_l + e_t \quad (27)$$

Los coeficientes en la regresión son estimados mediante mínimos cuadrados. El principio de mínimos cuadrados se basa en escoger aquellos coeficientes que minimicen la suma cuadrática del error tal como se observa en la ecuación 28.

$$\sum_{t=1}^T \varepsilon_t^2 = \sum_{t=1}^T (y_t - \beta_1x_{1,t} - \dots - \beta_kx_{k,t})^2 \quad (28)$$

Si bien es cierto estos modelos llegan a ser de gran utilidad si se conocen las variables regresores. No obstante, en la práctica esto se vuelve limitado ya que para pronosticar se necesita conocer las variables regresores las cuales son desconocidas (Rob J Hyndman, 2021).

Dicho lo anterior, en los modelos de regresión de series de tiempo se puede incorporar variables 'dummy' que hagan referencia a patrones estacionales que ya se encuentran definidas en el calendario como, por ejemplo: incluir los días de la semana o festividades del año (Rob J Hyndman, 2021).

Como alternativa del punto anterior los modelos de regresión pueden incorporar series de Fourier (Vieira, 2023). Las series de Fourier son series de tiempo del seno y coseno de la unidad periódica. Estas series son utilizadas para considerar la estacionalidad de las horas del día, los días de la semana, el mes del año, etc. En la ecuación se muestra la aplicación de estas series como variables regresores

$$f(x) = a_0 + \sum_{n=1}^{\infty} \left(a_n \cos \frac{n\pi x}{L} + b_n \sin \frac{n\pi x}{L} \right) \quad (29)$$

2.6 Métricas de evaluación de pronósticos

Para concluir este capítulo, en esta sección haremos una breve revisión de las métricas de evaluación utilizado en los pronósticos. Estas métricas deben ser evaluadas en genuinos pronósticos sobre data que el modelo no ha visto. Con ello, el tamaño de los residuos de no es un indicador de que tan grande puede ser el error de predicción (**Rob J Hyndman, 2021**). Habiendo dicho eso, es una práctica común dividir los datos en dos porciones que se denominan datos de entrenamiento y de prueba.

Por un lado, los datos de entrenamiento sirven para que el modelo pueda estimar los parámetros del pronóstico y, los datos de prueba se lo utilizan para medir la precisión de los pronósticos. Estos conjuntos de datos tienen que dividirse de forma secuencial y no aleatoria. Donde el tamaño de la división va a depender de la longitud de la serie y del horizonte de tiempo a pronosticar.

Volviendo al tema que nos ocupa, las medidas de evaluación se las realiza a través de los errores de los pronósticos. Los errores se puede definir como la diferencia entre el valor real y el valor estimado o pronostico. De aquí, podríamos sumar los errores de distintas formas. Entre las medidas comúnmente usadas se encuentra: MAE (Error absoluto medio), RMSE (Raíz Cuadrática media del error), MAPE (Error absoluto medio porcentual). El uso de cualquiera de estas medidas va a depender de los objetivos del investigador.

El MAE es utilizado para comparar distintos modelos sobre una sola serie de tiempo o para comparar diferentes series de tiempo que cuenten con a misma unidad de medida. El modelo que menor MAE tenga es aquel que alinea los pronósticos a la mediana. Esta medida penaliza grandes errores y es utilizado cuando la data presenta datos atípicos. La fórmula del MAE es sencilla de interpretar y viene dado de la siguiente expresión.

$$MAE = \sum_{i=1}^N \frac{|y_i - \hat{y}_i|}{N} \quad (30)$$

El RMSE, a diferencia del MAE, es sensible a datos atípicos ya que penaliza mas pequeños errores y debido a que es mejor detectando estas pequeñas diferencias es una medida más precisa que el MAE. La fórmula del RMSE, es muy parecida al MAE y viene dada de la siguiente expresión

$$RMSE = \sqrt{\sum_{i=1}^N \frac{(y_i - \hat{y}_t)^2}{n}} \quad (31)$$

Para concluir esta revisión. El MAPE es una medida de precisión que es utilizada para cuantificar el error medio de las predicciones en forma de porcentaje. El porcentaje es calculado sobre el valor actual de la serie, por ende, si el valor actual de la serie es 0, la medida tiende al infinito. La fórmula del MAPE viene dado de la siguiente expresión.

$$MAPE = \sum_{i=1}^N \frac{|y_i - \hat{y}_i|}{y_t} \quad (32)$$

CAPÍTULO 3

3. DISEÑO E IMPLEMENTACIÓN

3,1 introducción

Los sistemas de información hoy en día juegan un rol fundamental en el desarrollo de las empresas. La inteligencia de negocios con sus plataformas juega un papel fundamental en la estrategia empresarial. Tal como menciona (Cano, 2007), el objetivo básico de estas herramientas es apoyar continuamente y de forma sostenible a las organizaciones con la finalidad de mejorar su competitividad esto lo realiza facilitando el acceso a información a las partes interesadas para la toma de decisiones

En esa misma dirección, (Silva Solano, 2017) define a la inteligencia de negocios como el conglomerado de metodologías, prácticas y capacidades enfocadas al uso de información que nos permite tomar decisiones más acertadas. La práctica de la inteligencia de negocios se logra a través de sistemas de tecnologías de la información y de un conocimiento del giro del negocio de la empresa.

. En ese sentido, según (Diaz, 2012) para poder implementar soluciones de inteligencia de negocios es importante poder contar con una estructura tecnológica que permita la extracción, transformación y carga de datos continuamente.

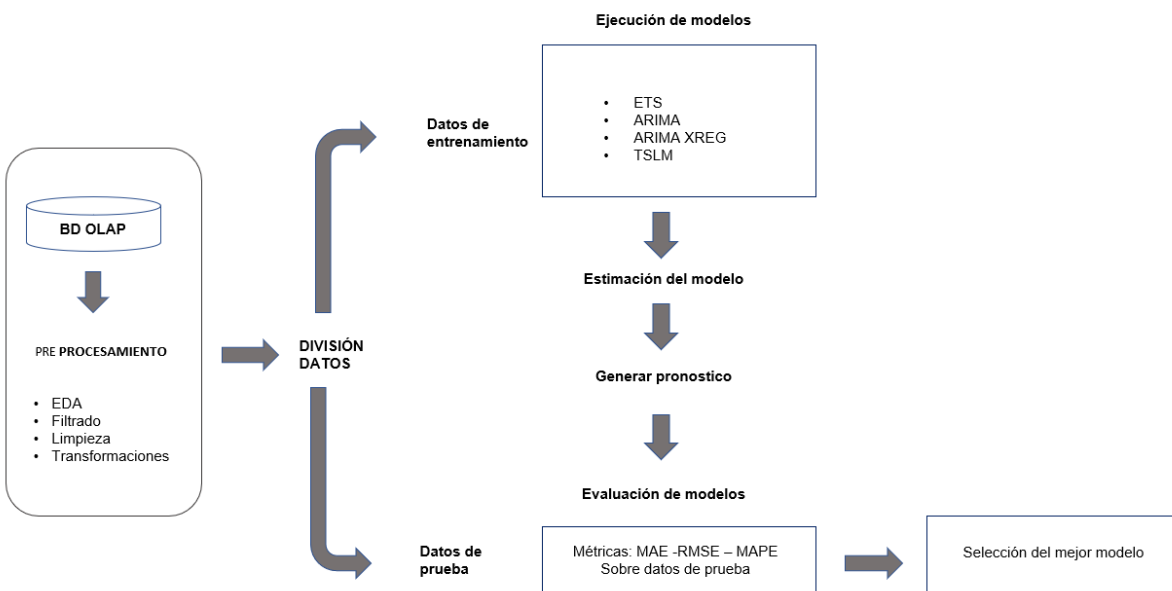
Habiendo dicho eso, en este capítulo se abordará al detalle el diseño de la arquitectura tecnológica a implementar en la empresa “AAA”. Desde el almacenamiento el procesamiento y la visualización de los pronósticos generados por los métodos aplicados además de los criterios de selección de modelos utilizados.

En la figura 9 se detalla el diseño del proyecto de implementación, se empieza por la obtención de datos mediante la consulta de una base de datos OLAP, después se selecciona los campos y se procede con la fase del procesamiento que empieza por un Análisis exploratorio. El análisis exploratorio se lo realiza con la finalidad de familiarizarse con la naturaleza de los datos que se va a estudiar, detectar posibles relaciones entre variables, detectar valores atípicos entre otros. (Rojo, 2005). Este análisis en series temporales nos permitirá también realizar transformaciones de Box – Cox esto, con la finalidad de corregir la varianza en la distribución de los datos.

Una vez finalizada dicha etapa se utilizará una librería especializada de código abierto que se encargará de dividir la data en entrenamiento y prueba. Esta misma librería aplicara los métodos convencionales de series de tiempo y evaluara los modelos mediante medidas de evaluación como son el RMSE y el MAPE. Esto, con el objetivo de seleccionar el mejor modelo.

Figura 9

Esquema del diseño del proyecto de implementación



3.2 El modelo de arquitectura tecnológica

La arquitectura que utilizaremos será cliente-servidor(C/S). en la cual como define (D. Polo, 2015), como un modelo de aplicaciones distribuidas en el que las tareas se reparte entre proveedores de recursos o servidores y los demandas, los clientes. Un cliente realiza la petición y el servidor responde.

Este modelo constara de tres bloques. El primero de ellos estará representado con el almacenamiento. Para este bloque utilizaremos una base de datos OLAP. Posteriormente haremos las peticiones por medio de la herramienta estadística denominada R en el cual, se desarrollará el análisis descriptivo y exploratorio, la limpieza, transformación de datos, imputación, estimación de modelos, generación de pronósticos y validación de modelos. Por último, utilizaremos una plataforma de inteligencia de negocios en la cual el usuario podrá utilizar los pronósticos de los modelos seleccionados. En la figura 10 se detalla el modelo de arquitectura propuesto

Figura 10

Modelo de arquitectura tecnológica



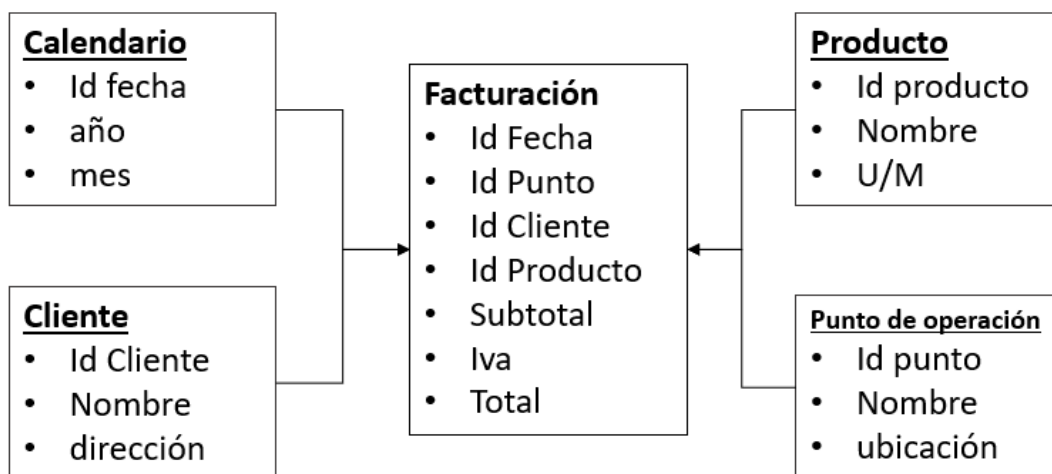
3.3 Almacenamiento de los datos

Para el caso de la empresa “AAA”, la herramienta que es utilizado para el almacenamiento es Microsoft SQL Server. SQL es un sistema de gestión de base de datos relacional cuyo lenguaje principal de consulta es el Transact- SQL.

Las consultas a esta base de datos se da gracias a los modelos lógicos basados en objetos. Entre estos modelos el más representativo es el modelo entidad relación en el cual se basa en la percepción de un mundo de objetos denominados Entidades y de las relaciones existente que se da entre cada uno de ellos. Una entidad es un objeto que puede distinguirse de otros dado que cada uno posee ciertos atributos que lo identifica. Una relación es una asociación entre varias entidades (Leon, 2008). En la figura 11 se puede ejemplificar un ejemplo del modelo entidad relación.

Figura 11

Modelo de Entidad Relación.



Esta herramienta tiene un rol fundamental en la arquitectura tecnológica ya que interviene activamente en todas las fases del proceso. Es la encargada de almacenar la información proveniente de las distintas fuentes de información de la empresa. También es la encargada de almacenar los pronósticos y la información histórica de ventas.

3.4 Obtención de datos

El consumo o la obtención de los datos se realizará mediante consultas al gestor de base de datos. Esta información surge del sistema transaccional de los puntos de venta. Al estar en un modelo de entidad relación nos permitirá extraer la información de nuestro interés como son fecha, unidad de negocios, región, productos y total de ventas. Por medio de la fecha, podremos crear nuevos campos como día de la semana, fines de semana, feriados, etc. que serán utilizadas como variables para la estimación y generación de pronósticos

3.5 Preprocesamiento y validación de datos

El objetivo principal del pre-procesamiento de datos es corregir las inconsistencias, ruido y valores faltantes de la información con la finalidad de no afectar los resultados de los pronósticos. Además, tal como se ha presentado en secciones anteriores, en las series de tiempo es una practica usual realizar transformaciones sobre la misma.

Para realizar la parte del preprocesamiento utilizaremos la herramienta de R con su interfaz de Rstudio. R, es un lenguaje y entorno de programación desarrollado inicialmente por Ross ihaka y Robert Gentleman a mediados de los años noventa que se especializa en análisis estadísticos y que cuenta con librerías especializadas para la generación de pronósticos

3.6 Generación de pronósticos y selección del mejor modelo

Una vez consultada la información de SQL, y después de realizar un análisis exploratorio de información, se realizará modelos de series temporales a nivel día: ETS, ARIMA, ARIMA CON REGRESORES Y TSLM (Time series lineal model)

Proceso ETS:

- Se dividirá la data en entrenamiento y prueba. Para la validación se obtendrá los datos días del último mes.
- Por medio de la función ETS, se obtendrá la combinación sugerida de tendencia, estacionalidad y error. Además, con ello, se obtendrá los parámetros de suavizamiento a utilizar según el caso.
- Con la finalidad de obtener mejores resultados, para los productos que apliquen, realizaremos transformaciones de variable y ajustaremos el parámetro 'estacional' en 1 semana para indicarle al modelo que la data cuenta con dicha frecuencia estacional.

Proceso ARIMA:

- Se dividirá la data en entrenamiento y prueba. Para la validación se obtendrá los datos días del último mes.
- Por medio de la función auto.arima, se obtendrá la combinación de parámetros autorregresivos y medias móviles(p,d,q) no estacionales y estacionales(P,D,Q) si es el caso.

Proceso ARIMA XREG:

- Se dividirá la data en entrenamiento y prueba. Para la validación se obtendrá los datos días del último mes.
- Por medio de la función auto.arima, se obtendrá la combinación de parámetros autorregresivos y medias móviles(p,d,q) no estacionales y estacionales(P,D,Q) si es el caso.
- Según el producto que corresponda, ingresaremos variables dummy que identifiquen patrones estacionales como el día de la semana, fines de semana, fines de mes, o eventos en particulares.

Proceso TSLM:

- Se dividirá la data en entrenamiento y prueba. Para la validación se obtendrá los datos días del último mes.
- En estos modelos, incluiremos los parámetros de tendencia, estacionalidad y series de Fourier

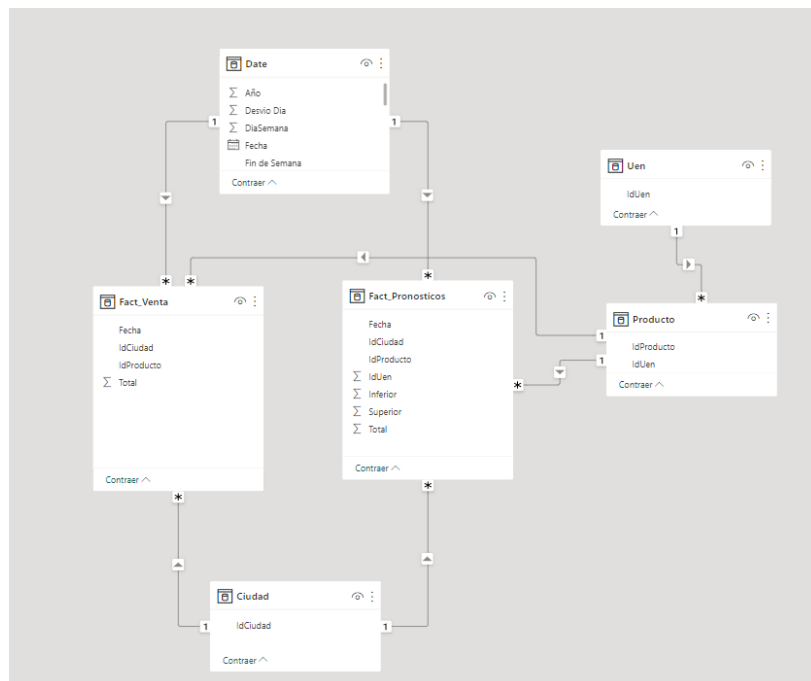
3.7 Diseño de Dashboard en Power BI para análisis de resultados

Power BI, como lo define Microsoft es una plataforma unificada y escalable de inteligencia empresarial (BI) con funciones de autoservicio apta para grandes empresas. Se ha escogido PowerBi como plataforma de inteligencia empresarial ya la empresa "AAA" cuenta con esta herramienta lo cual facilitara la curva de aprendizaje de cara a los usuarios.

Así mismo, por medio de esta herramienta se podrá un modelo de datos escalable que nos permita combinar información de distintas fuentes. Una de las fuentes de información vendrá directamente de la base de datos OLAP en un objeto denominado vista. Una vista es una tabla virtual definido por una consulta. Similar a una tabla, una vista consta de un conjunto de columnas y filas de datos con un nombre. Sin embargo, a menos que esté indizada, una vista no existe como conjunto de valores de datos almacenados en una base de datos (Microsoft, 2023). En estas vistas almacenaremos la venta diaria y las dimensiones de fecha, unidad de negocio, productos y las ciudades. Este modelo de datos se finalizará con la inclusión de los pronósticos que provendrá de un Excel que exportaremos desde R. En la figura se muestra el modelo de datos a utilizar

Figura 12

Modelo de Entidad Relación pronostico prototipo



CAPÍTULO 4

4. ANÁLISIS DE RESULTADOS

4.1 Antecedentes

En este capítulo, incursionaremos al detalle la solución de la herramienta. Iniciaremos por un EDA por cada unidad de negocios, esto con la finalidad de poder entender características pde los datos y esto a su vez, nos permita realizar un preprocesamiento adecuado.

Posteriormente, se explicará el análisis de resultados por medio de un aplicativo de visualización que permitirá al usuario poder hacer uso de los beneficios de los pronósticos.

4.2 El dataset

El dataset propuesto contiene 14,082 registros donde cada uno de estos representa la venta día de sus productos por cada ciudad y unidad de negocio. El intervalo de tiempo seleccionado para el análisis corresponde a 15 meses que va desde el 1 de enero del 2022 hasta el 31 de marzo del 2023. En la tabla 3, se detalla una breve descripción del dataset.

Tabla 3

Tabla de variables con tipo de datos

Variable	Tipo de datos
Fecha	Date
IdUen	Int
IdCiudad	Int
IdProducto	Int
Total	Double

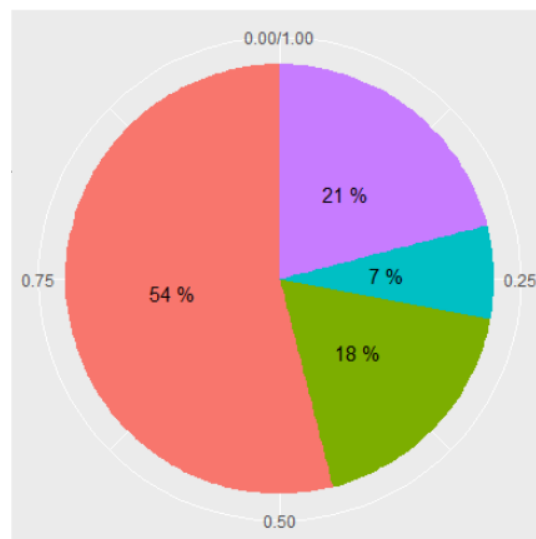
4.3 Análisis Exploratorio de datos

Los datos provienen de la empresa “AAA” cuyo giro de negocio se enfoca en la comercialización y distribución de artículos varios. Esta empresa cuenta con cuatro unidades de negocios cada una independiente entre sí. En la cual, cada unidad de negocio cuenta con su propio portafolio de producto y se distribuye de forma independiente a través de las 5 ciudades.

Conforme a lo que se menciona en el párrafo anterior, la empresa “AAA” cuenta con cuatro unidades de negocios. Donde la participación de venta con mayor peso recae en la unidad de negocio 1 con el 54% tal como detalla en la figura 13. Esto se debe principalmente a que esta unidad de negocios cuenta con una mayor cobertura a nivel de ciudades, un mayor portafolio de productos y, es la única unidad de negocios que llega al consumidor final por medio de distribuidores a los cuales se les ofrece descuentos o palancas comerciales lo cual genera un “push” de venta.

Figura 13

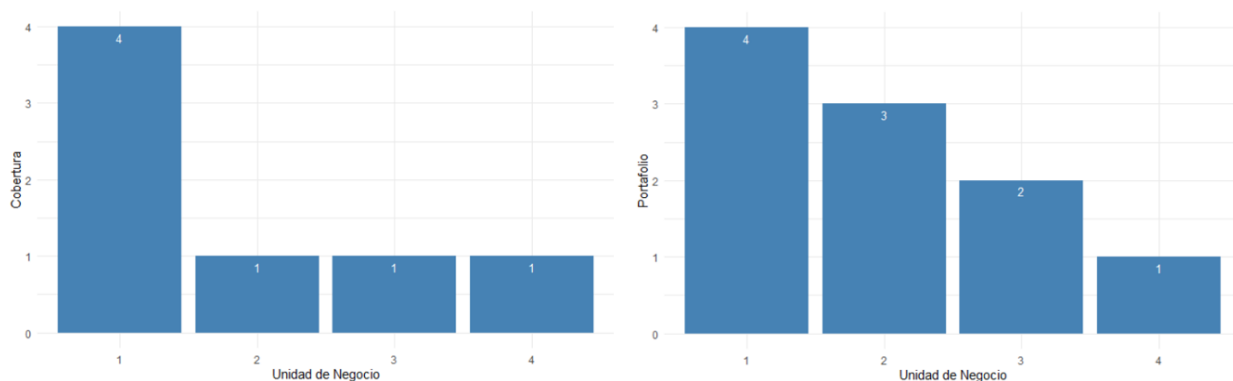
Participación de ventas por Unidad de negocios



En esa misma dirección, lo sigue la unidad de negocios 4 con el 21% de participación. Si bien, esta UEN solo cuenta con 1 producto y se distribuye en una sola ciudad, el precio de venta es superior a la de las otras unidades de negocio. Lo sigue la unidad de negocios #2 que cuenta con tres productos con el 18% de participación y por último la Unidad de negocio #3 con el 7% y un portafolio de dos productos, estas dos últimas unidades de negocios mencionadas cuentan con una ciudad de cobertura. En la figura 14 se muestra el detalle de cobertura y portafolio de productos

Figura 14

Detalle de cobertura y portafolio de productos por Unidad de negocios



A continuación, realizaremos un breve análisis por cada unidad de negocio, esto con la finalidad de entender el comportamiento y las distintas características de sus productos. Así mismo, con la finalidad de simplificar el análisis, de aquí en adelante a cada unidad de negocio las llamaremos "UEN".

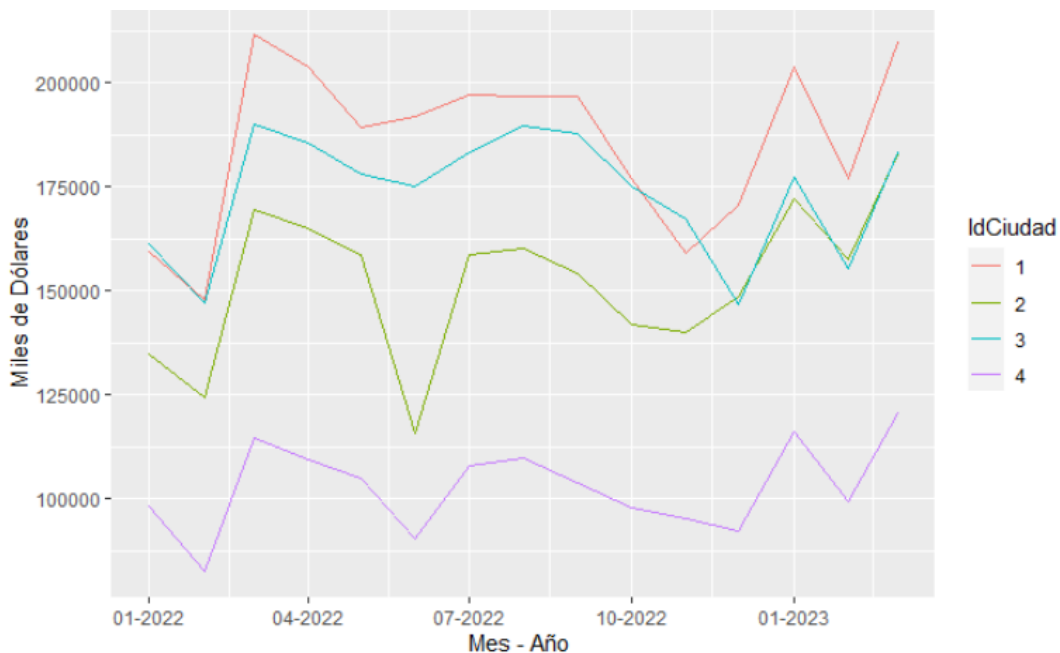
4.3.1 UEN #1

4.3.1.1 Análisis de cobertura.

Empezaremos esta unidad de negocios con un análisis de su cobertura ya que es la unidad que se distribuye en 4 ciudades distintas. Como se puede observar en la figura 15, las ventas entre las ciudades siguen un mismo comportamiento a través del tiempo. Esto es, que todas ellas suben o bajan en conjunto manteniendo su mismo nivel de venta.

Figura 15

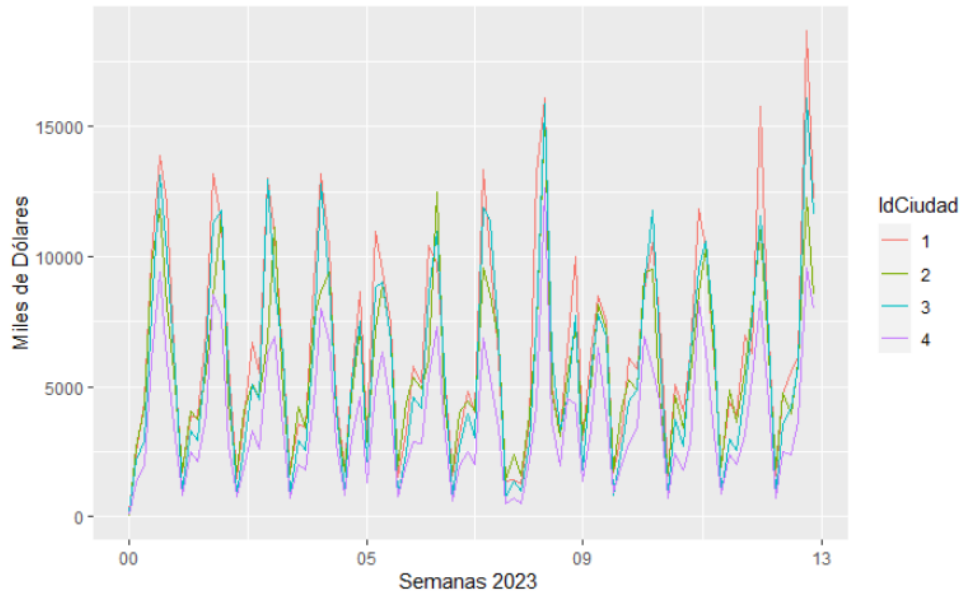
Evolutivo por ciudad UEN #1



Si hacemos un acercamiento a estas series y lo vemos desde un punto de vista semanal, en la figura 16 se puede observar que, a priori, la venta cuenta con picos que se repiten a través de las distintas semanas y ciudades. Es así que, podríamos aseverar que hay una homogeneidad entre las distintas ciudades.

Figura 16

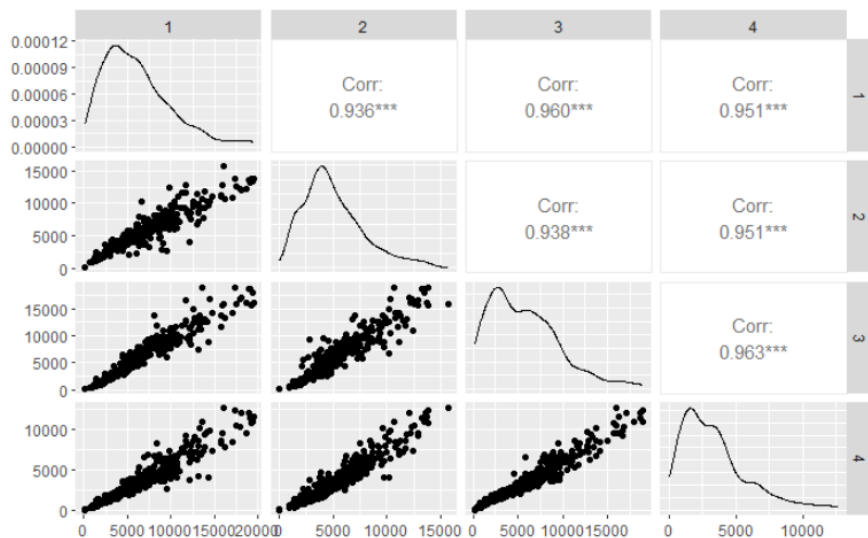
Evolutivo semanal ciudad UEN #1



Exploremos un poco la idea de que existe homogeneidad entre las ciudades correlacionando las ventas entre sí. En la figura 17, se detalla la correlación existente entre ciudades en la cual indica que hay una fuerte correlación entre todas las ciudades, todas ellas están por encima del 90%.

Figura 17

Correlación de ventas entre ciudades UEN #1



En definitiva, podríamos concluir que es posible utilizar un modelo único para todas estas ciudades. Es por ello que, para nuestro posterior análisis, realizaremos la evaluación de los modelos a nivel Uen – Producto.

4.3.1.2 Análisis de productos.

Si ampliamos nuestro análisis a nivel producto, en la figura 18 se puede observar que, a diferencia de las ciudades, si bien todos los productos cuentan con una alta correlación tal como lo detalla la figura 19, cada serie cuenta con un comportamiento o forma diferente. Por ejemplo, parecería que hay una leve tendencia positiva en el producto #2 desde enero del 2022, caso contrario con el producto #1 que al parecer hubo un incremento en sus niveles de venta por encima de los 250.000 mensuales y luego decreció para estabilizarse en los niveles de 200.000 dólares mes.

Figura 18

Evolutivo por productos UEN #1

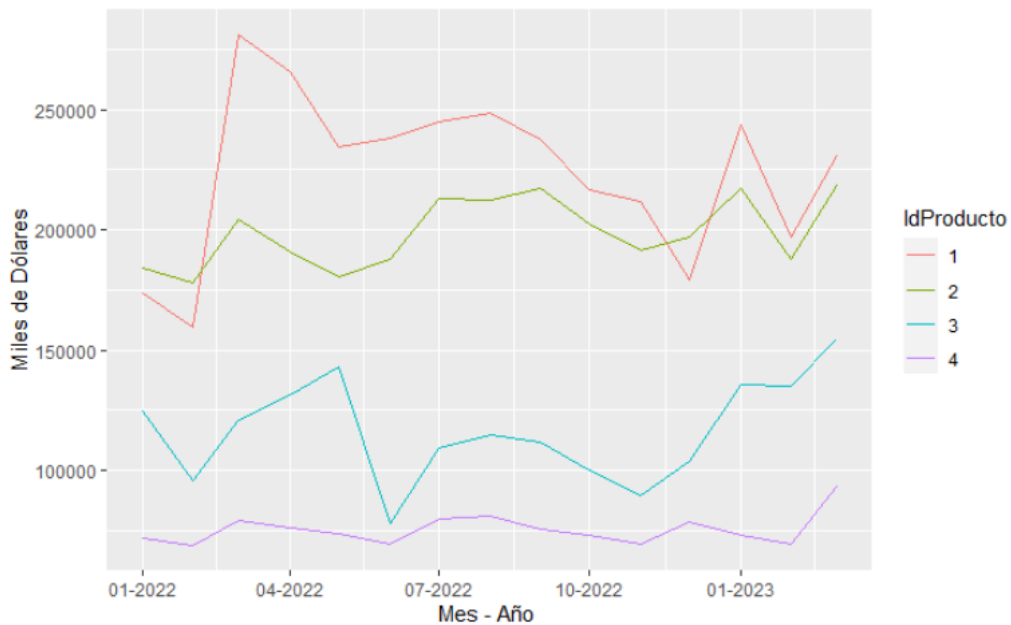
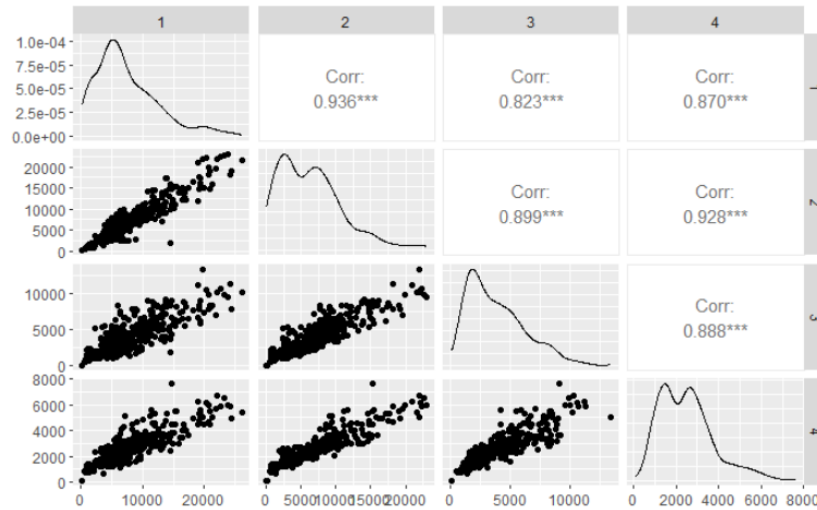


Figura 19

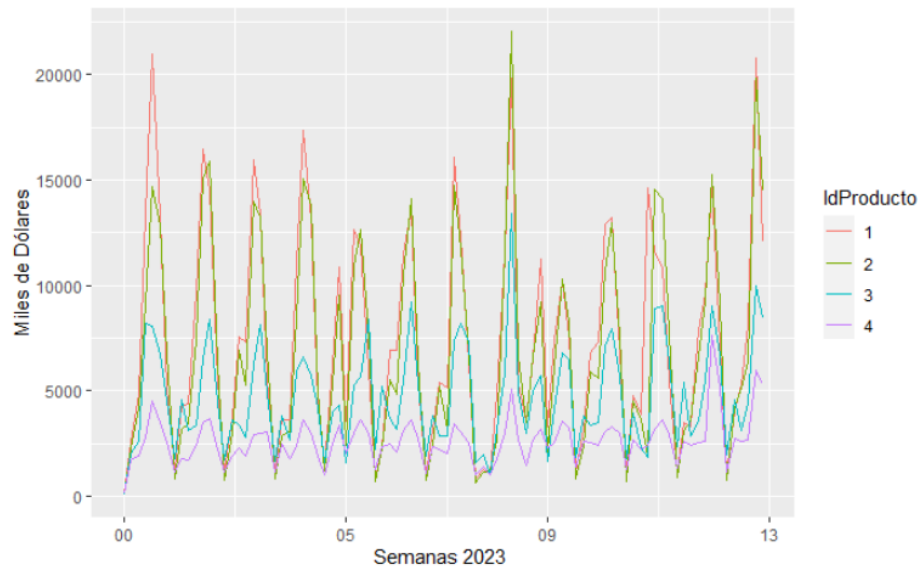
Correlación de ventas entre productos UEN #1



Si tomamos las últimas semanas del 2023, con el objetivo de identificar algún patrón estacional, en la figura 20 nos damos cuenta de que todos los productos presentan el mismo comportamiento, al parecer se incrementa a medidas que pasan los días de la semana y decrece los fines de semana.

Figura 20

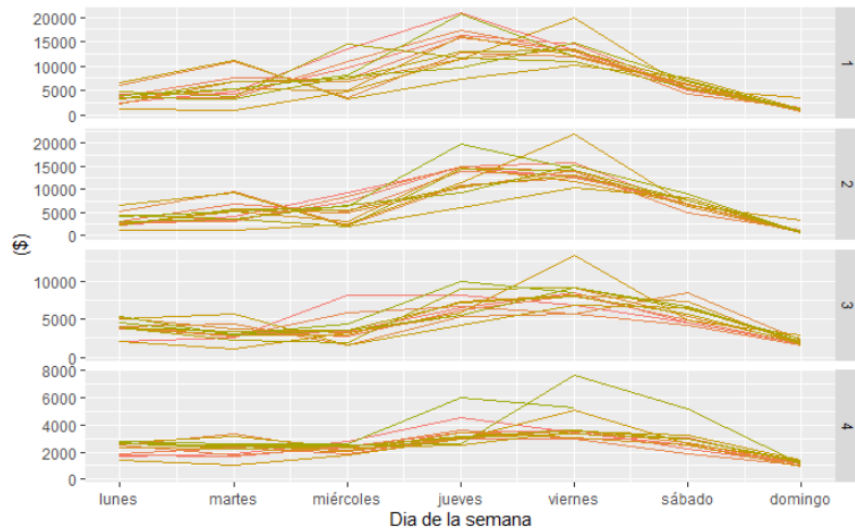
Evolutivo por semana producto UEN #1



Avanzando en nuestro razonamiento, realizando un gráfico estacional tal como se presenta en la figura 21, podemos afirmar que las ventas tienen un incremental a medidas que pasan los días y decrecen los fines de semanas. De modo que, en la fase del preprocesamiento, podríamos utilizar variables 'dummys' que hagan referencia los días de la semana.

Figura 21

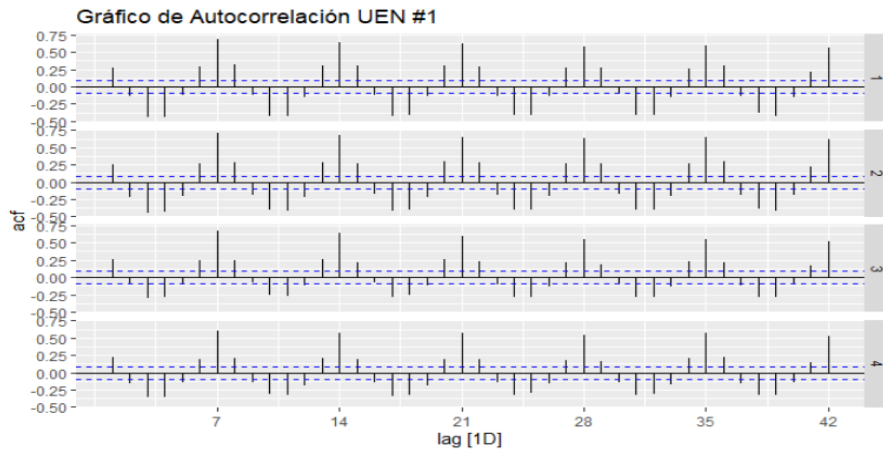
Estacionalidad semanal: Unidad de negocios #1



Llegados a este punto y, sabiendo que hay una presencia de estacionalidad semanal, realizando un gráfico de autocorrelación como detalla la figura 22 podremos observar que el valor actual se encuentra altamente relacionado con su valor séptimo rezagado y comienza a disminuir marginalmente en sus rezagos múltiplos del siete. Estos nos ayudan identificar que los modelos convencionales nos darían buenos resultados.

Figura 22

Gráfico de autocorrelación UEN #1



4.3.1.3 Descomposición de productos UEN #1.

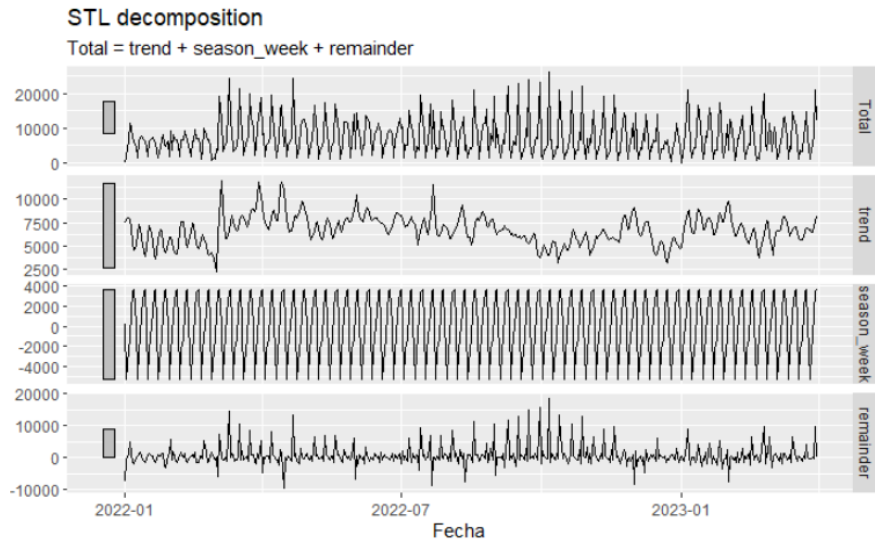
Prosigamos nuestro análisis descomponiendo cada una de las series temporales. En la sección que antecede, hicimos una breve descripción de los componentes de la serie mediante un gráfico evolutivo temporal.

En esta sección utilizaremos el modelo STL para poder identificar presencias de tendencia, estacionalidad y sus residuos. Para descomponer las series temporales configuraremos el parámetro de la tendencia con una media móvil de 7 y con una estacional periódica semanal de la función en R. A continuación, realicemos un breve análisis por cada uno de los productos.

El producto #1, como se presenta en la figura 23 no presenta una tendencia marcada a través del tiempo, podemos observar también extrayendo el componente estacional semanal da como resultado que los residuos se distribuyan normalmente y de forma aleatorio.

Figura 23

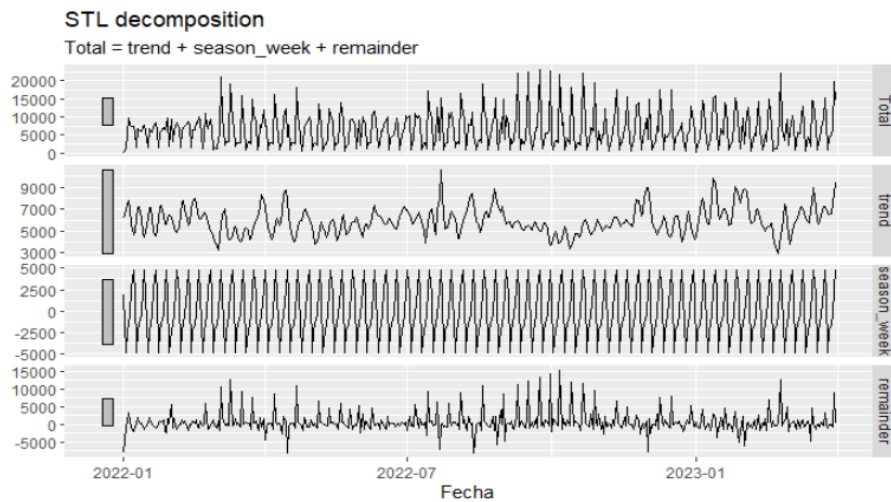
Descomposición temporal producto #1



Con respecto al producto #2, en la figura 24 se puede observar que cuenta con un comportamiento similar al producto #1. En donde no existe una tendencia marcada y los residuos se distribuyen aleatoriamente con una ligera autocorrelación entre los mismos periodos del producto #1

Figura 24

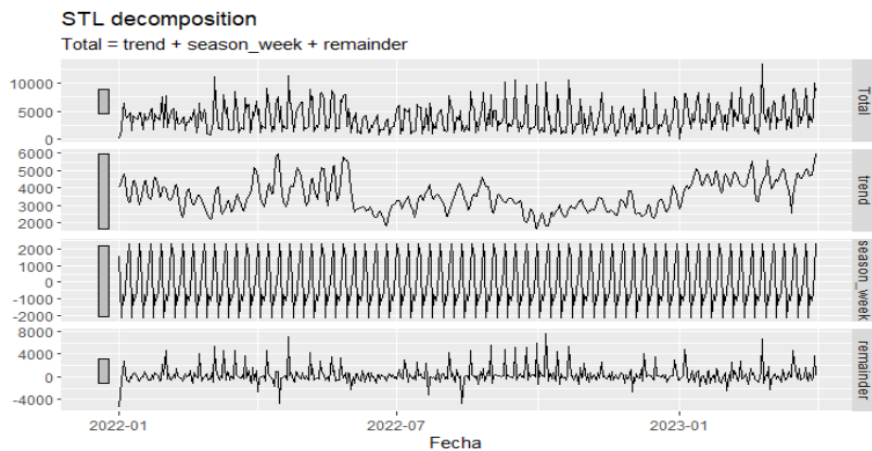
Descomposicion temporal producto #2



Siguiendo con el análisis, el producto #3, como se presenta en la figura 25 si presenta una ligera tendencia de ventas que inicia a finales del 2022, al igual que los dos productos previos presenta una autocorrelación en los residuos y un componente estacional semanal previamente descrito.

Figura 25

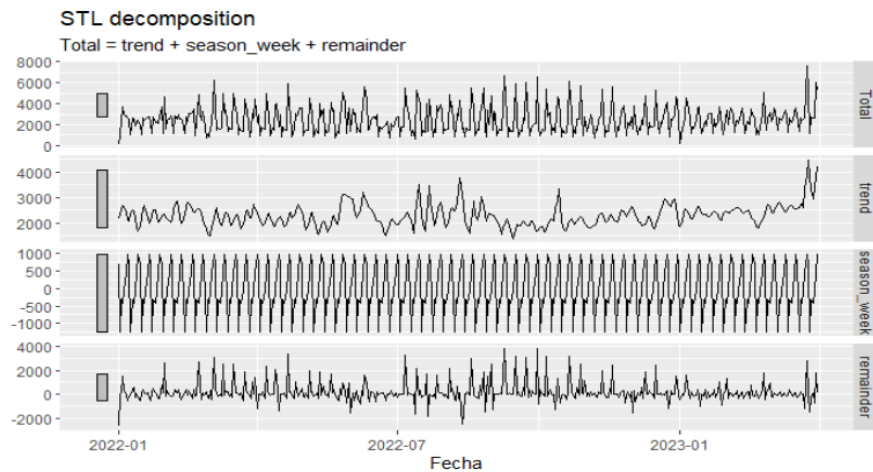
Descomposición temporal producto #3



Para concluir con esta unidad de negocio, la figura 26 se detalla la descomposición del producto #4. Este producto, al igual que el producto #3 cuenta con una ligera tendencia que inicia a finales del 2022 y así como los otros productos de esta unidad de negocio, presenta un patrón estacional y una ligera autocorrelación en los residuos.

Figura 26

Descomposición temporal producto #4



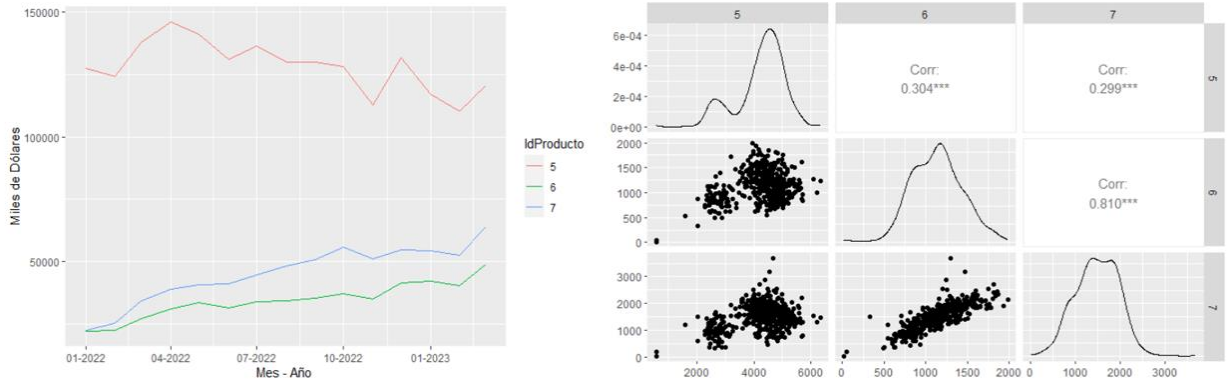
4.3.2 UEN #2

4.3.2.1 Análisis de productos

La UEN #2 se compone de tres productos y se distribuye en una sola ciudad. En la figura 27 se puede detallar un evolutivo de ventas por cada uno de sus productos y a correlación que se presenta entre cada uno de ellos. En ella se puede observar que el producto #5 es el más representativo a nivel de ventas y presenta una tendencia decreciente a lo largo del tiempo. Luego, los productos #6 y #7 si presentan una tendencia a lo largo del tiempo y presentan una alta correlación entre ambos productos 81%.

Figura 27

Evolutivo mes por productos – correlación UEN #2



Tomando como referencia las últimas semanas del 2023, podemos observar en la figura 28 que, la venta de todos los productos mantiene a un nivel promedio y una varianza constante. También, se puede observar que hay disminuciones de venta en un determinado día de la semana.

Figura 28

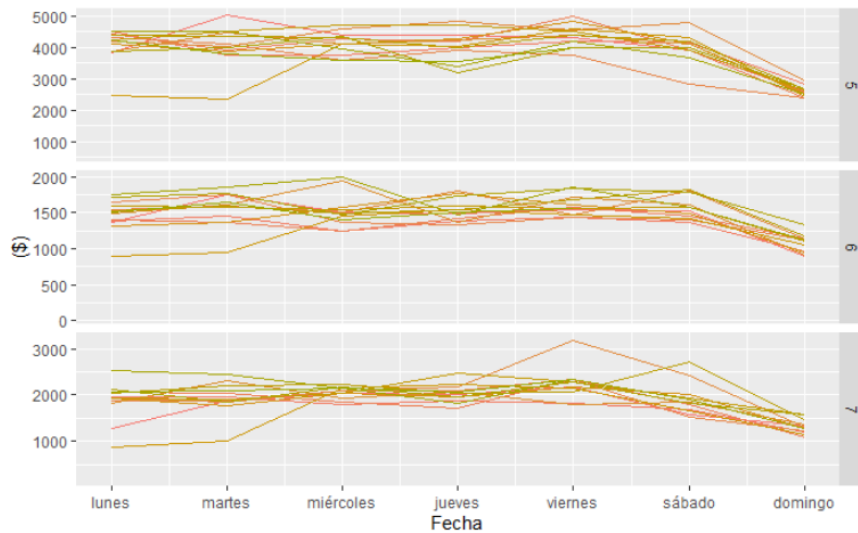
Evolutivo semanal por producto UEN #2



Realizando un gráfico estacional, podemos notar en la figura 29 que, en efecto las ventas los días de la semana se mantiene constante y hay una caída los fines de semana específicamente el domingo

Figura 29

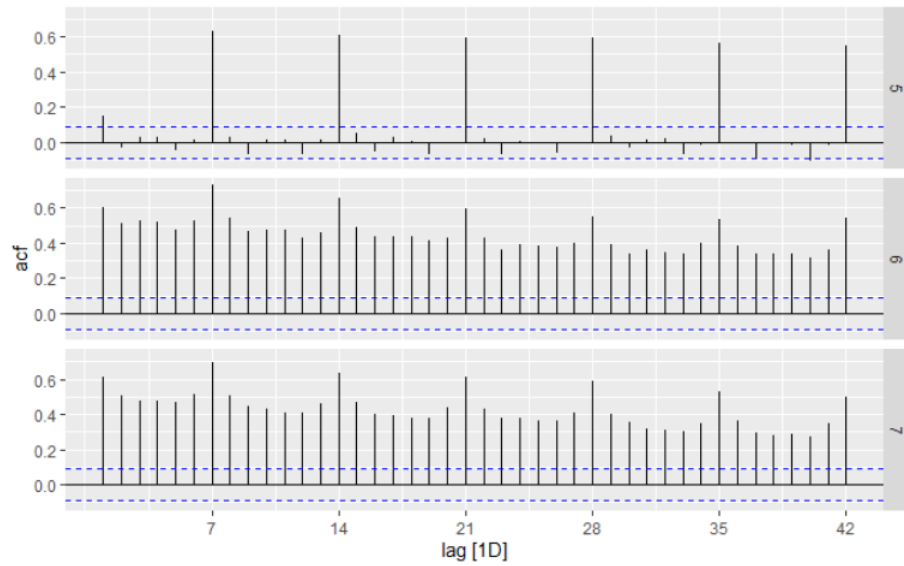
Estacionalidad Semanal: UEN #2



Así mismo, realizando un gráfico de autocorrelación tal como se presenta en la figura 30 podremos observar que el valor actual para el producto #5 se encuentra fuertemente relacionado con su valor rezagado en múltiplo del 7. Para los otros productos restantes, si bien hay una alta correlación con sus periodos rezagados múltiplo del 7 también hay una alta correlación con el periodo pasado lo que indica la presencia de un componente tendencial.

Figura 30

Gráfico de autocorelación UEN #2

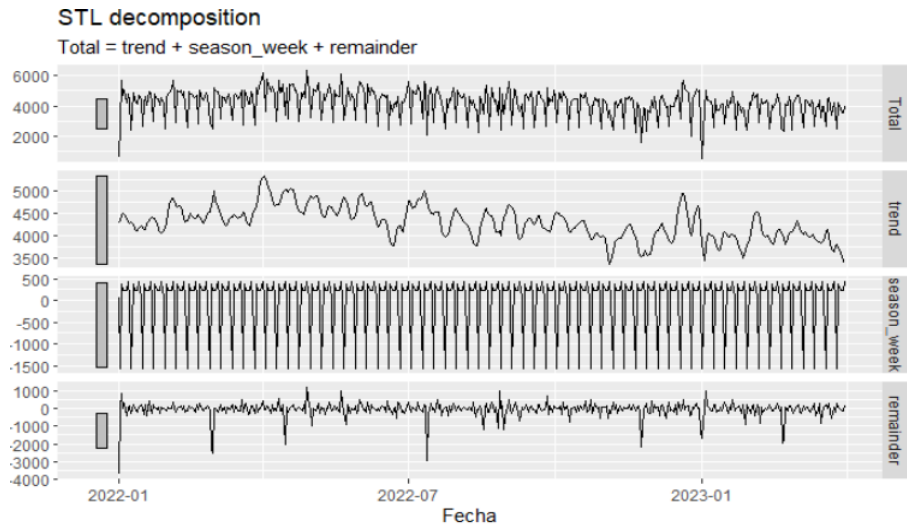


4.3.2.2 Descomposición series temporales UEN #2.

En las figuras 31 y 32, se detalla la descomposición de los productos de la unidad de negocios #2. En donde validamos las hipótesis previamente descritas. Con ellos, podemos destacar que todos los productos presentan un patrón estacional. En la figura 31 se puede observar los componentes del producto #5 el cual presenta una leve tendencia decreciente.

Figura 31

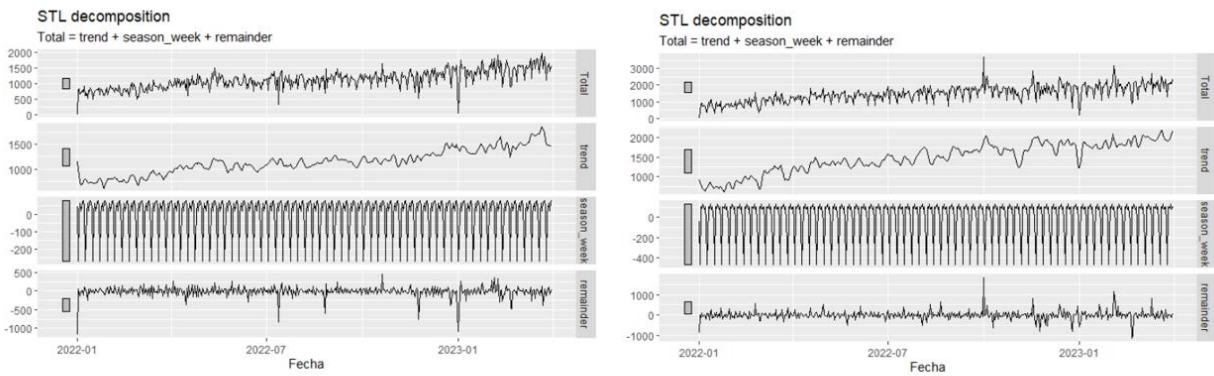
Descomposición temporal producto # 5



Los productos 6 y 7, como se detalla en la figura 32 presentan una marcada tendencia y para todos los productos los residuos se distribuyen aleatoriamente. Adicional, en los residuos podemos notar que todos estos cuentan con ciertos valores atípicos principalmente el producto #7.

Figura 32

Descomposición temporal productos # 6 y 7



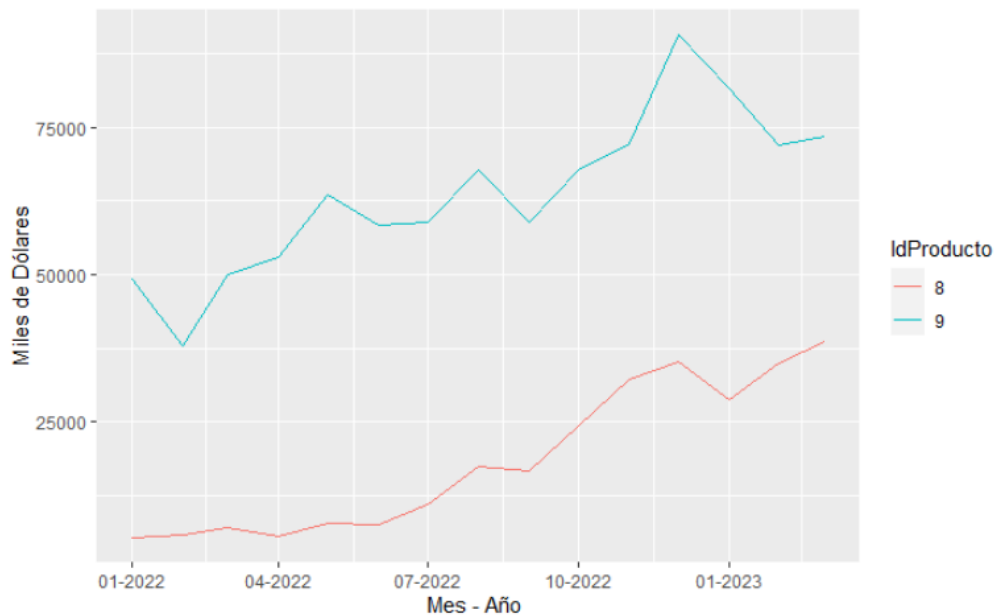
4.3.3 UEN #3

4.3.3.1 Análisis de productos

La UEN #3 se compone de dos productos y se distribuye en una sola ciudad. En la figura ,se puede detallar un evolutivo de ventas por cada uno de sus productos. En ella se puede observar que el producto #9 es el más representativo a nivel de ventas. Así mismo, ambos productos presentan con una marcada tendencia a lo largo del tiempo.

Figura 33

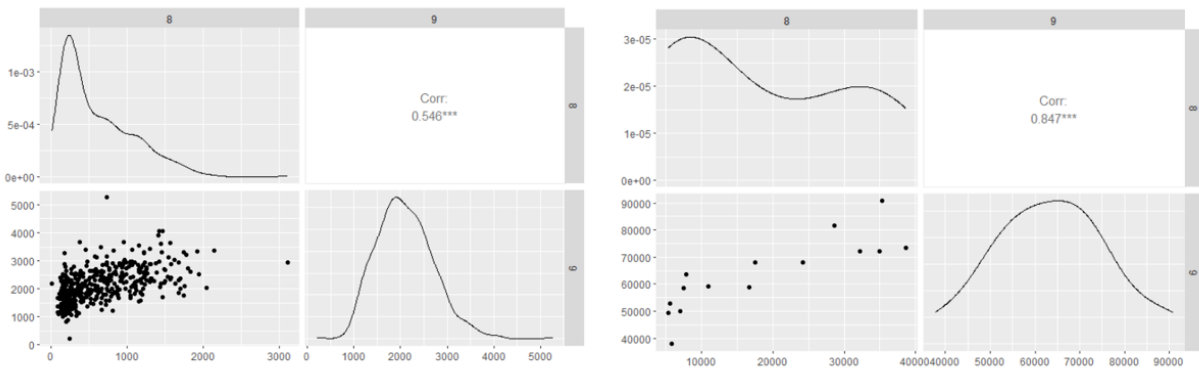
Evolutivo mes por productos UEN #3



En continuación con el análisis, podemos observar en la figura 34 aunque ambos productos tienen alta correlación a nivel mes, con un 84,7% a nivel día muestra una baja correlación de apenas un 54% lo cual indicaría que cada producto tiene diferentes niveles de venta a través de los días.

Figura 34

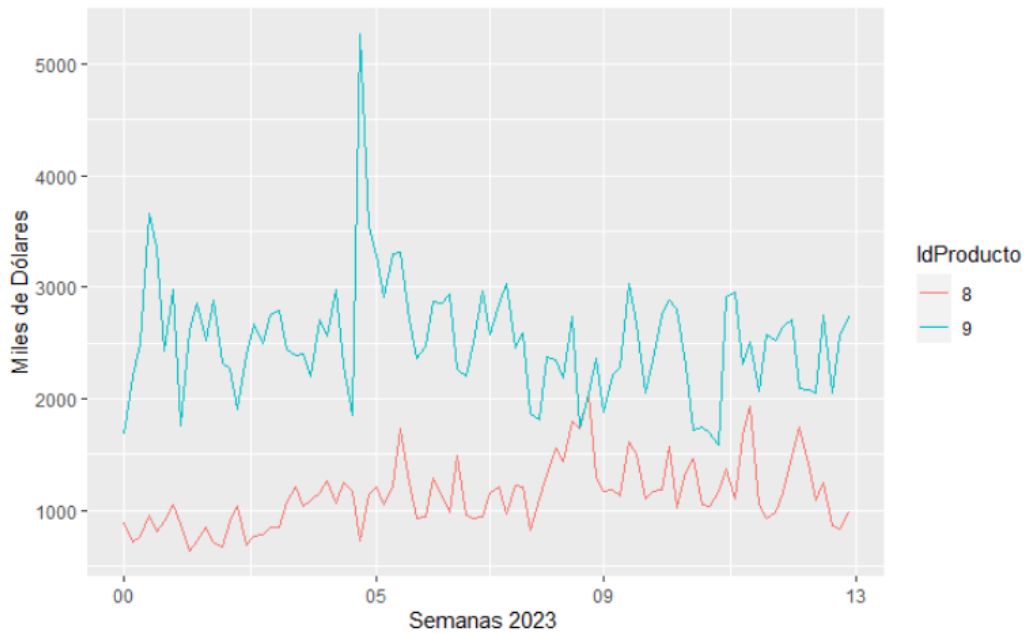
Correlación día y mensual productos UEN#3



Ampliando nuestro análisis, en la figura 35 se puede notar que no diferencia de las unidades negocios analizadas previamente, en esta UEN no se pueda detectar algún tipo de patrón estacional ambos productos se comportan de manera aleatoria.

Figura 35

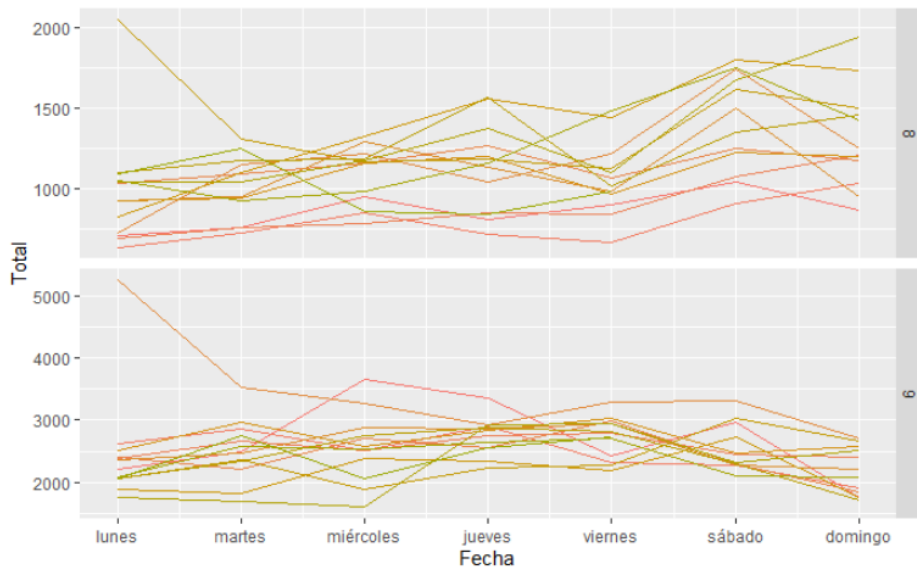
Evolutivo semanal productos UEN#3



Mediante un gráfico estacional como se talla en la figura 36 podemos validar que, en efecto, ambos productos no cuentan con algún tipo de estacional relacionado al día de la semana. En el producto #8, las ventas comienzan a incrementarse los fines de semana en contraste con el producto #9 que mantiene una venta constante todos los días de la semana.

Figura 36

Estacionalidad semanal: UEN #3

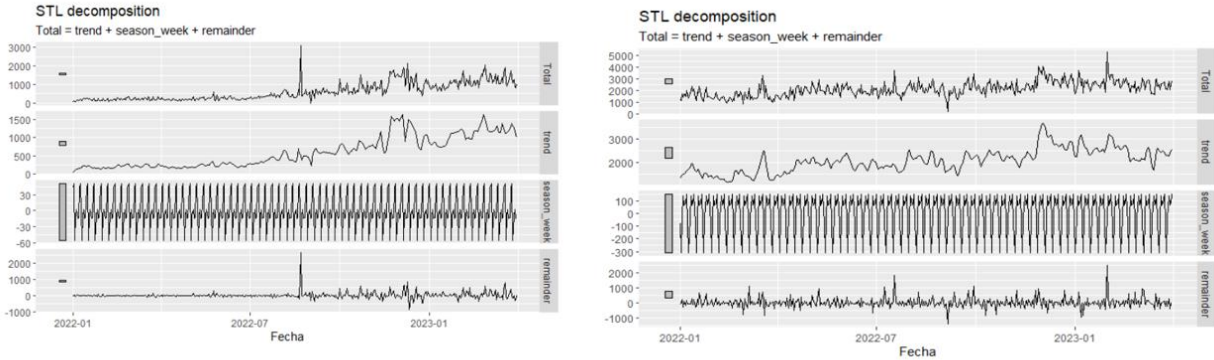


4.3.3.2 Descomposición series temporales UEN #3.

En la figura 37, se detalla la descomposición de los productos de la unidad de negocios #3. En donde validamos las hipótesis previamente descritas. Con ellos, podemos destacar que ambos productos crecen a través del tiempo y que presentan una estacionalidad a nivel semanal. Así también, observando sus residuos, ambos productos presentan valores atípicos y en su mayoría estos se distribuyen aleatoriamente.

Figura 37

Descomposición productos UEN #3

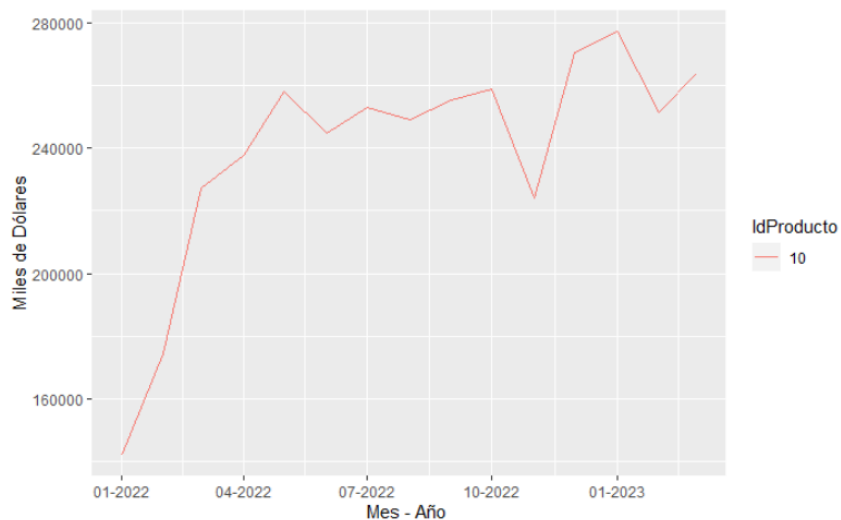


4.3.4 UEN #4

La unidad de negocios #4 está compuesta por un solo producto y es la segunda a nivel de importancia con respecto a los ingresos. Esta UEN, si bien presenta un incremento en sus niveles de venta hasta mayo del 2022, estas ventas se estabilizan en un promedio de 250.000 mes. El detalle se muestra en la figura 38

Figura 38

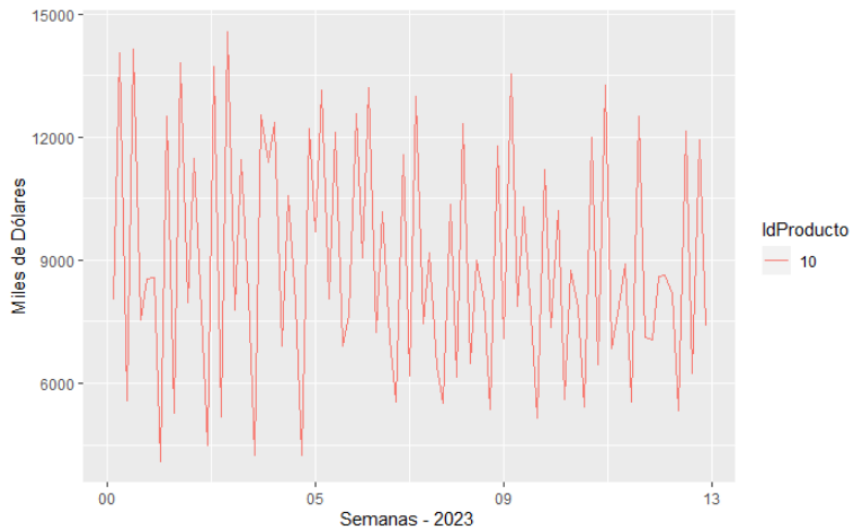
Evolutivo mensual UEN #4



A nivel diario, tomando las últimas semanas del 2023 como se puede observar en la figura 39, la serie presenta un comportamiento aleatorio en el cual no se distingue que exista un patrón estacional referente a los días de la semana.

Figura 39

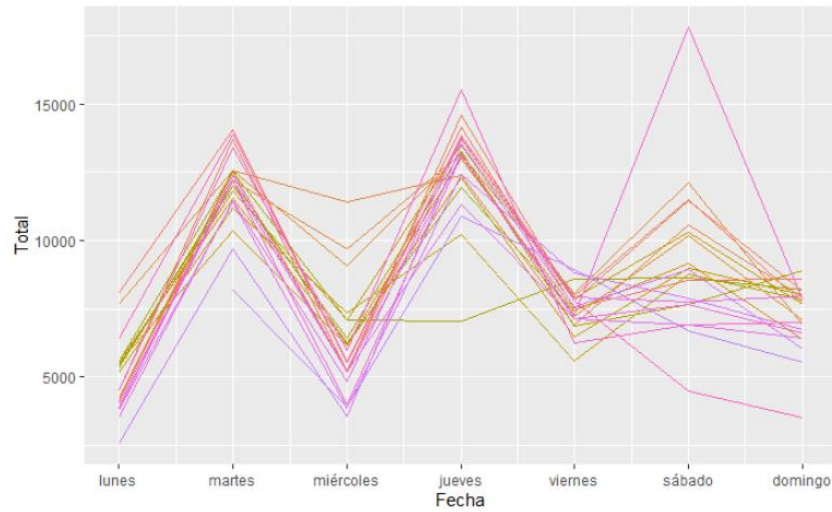
Evolutivo semanal UEN #4



Para contrastar nuestra hipótesis, en la figura 40 observa que la venta si responden al día de la semana. En la figura, se puede notar que la venta tiene un incremental los martes y jueves y cae los días lunes, miércoles y viernes

Figura 40

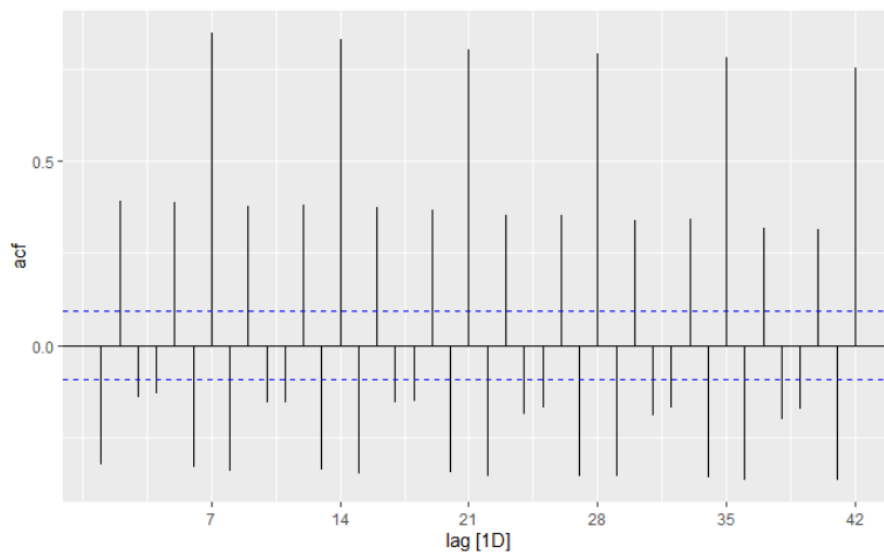
Estacionalidad semanal UEN #4



En esa misma dirección, en la figura 41 se puede observar la correlación existente con sus periodos rezagados múltiplos del 7. Así como también, muestra una correlación alta positiva y negativa pasando un rezago.

Figura 41

Gráfico de autocorrelación UEN #4

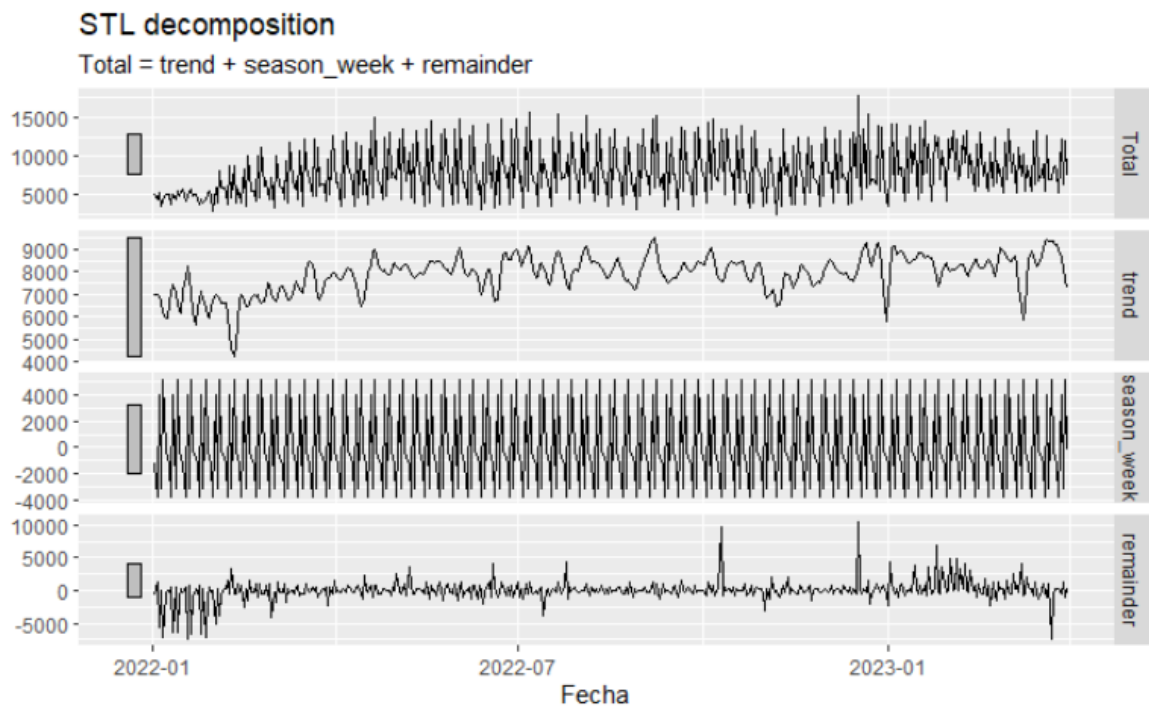


4.2.4.1 Descomposición series temporales UEN #4.

Tal como se observa en la figura 42, se puede destacar que no existe una marcada tendencia y que dicho componente se sitúa a nivel de 8000. Su componente estacional semanal fluctúa entre los 4000 y sus residuos se comportan de forma aleatorio entre 5000.

Figura 42

Descomposición unidad de negocios #4



4.4 Preprocesamiento de datos.

Debido a que ciertos productos tal como se presentó en la sección previa presentan ciertos valores atípicos es importante previo a generar cualquier modelo tratarlos de manera adecuada. Para ello utilizaremos la librería `imputeTS` de R que por medio de la función `tsouliers` nos ayudara a identificar y nos recomendará un valor para dicho punto.

Para ello lo realicemos en las unidades de negocios 2 y 3 ya que previamente mostraban ciertos valores atípicos.

4.2.1 Preprocesamiento: UEN #2

A continuación, en la figura 43 se muestra el detalle de los reemplazos realizado por la herramienta para cada producto de la unidad de negocios #2. El producto #5 presenta 3 reemplazos mientras que los productos 6 y 7 muestran 4 reemplazos respectivamente.

Figura 43

Imputación de valores atípicos UEN #2

Producto #5	
Index	reemplazos
1	2.771
328	3.772
366	3379

Producto #6	
Index	reemplazos
1	807
194	1992
328	1285
366	1124

Producto #7	
Index	reemplazos
274	1688
359	1617
366	1166
399	2286

4.2.2 Preprocesamiento: UEN #3

A continuación, en la figura 44 se muestra el detalle de los reemplazos realizado por la herramienta para cada producto de la unidad de negocios #3. El producto #8 presenta 17 reemplazos mientras que el producto #9 simplemente un reemplazo

Figura 44

Imputacion de valores atipicos UEN #3

Producto #8	
Index	reemplazos
234	397
244	405
274	804
295	1048
318	899
321	697
339	1417
340	1157
343	1301

Producto #8	
Index	reemplazos
345	1159
352	956
258	590
400	1235
421	1589
423	1520
443	1362
449	1454

Producto #9	
Index	reemplazos
395	2693

4.5 Selección del mejor modelo.

En esta sección, implementaremos varios modelos para cada producto, en la cual mediante la métrica de evaluación MAPE y RMSE seleccionaremos el mejor modelo con el menor indicador tanto en los datos de entrenamientos como los de prueba.

Para este ejercicio, evaluaremos los pronósticos generados para enero febrero y marzo del 2023. Con ello, para el pronóstico de enero los datos de entrenamiento serán hasta diciembre del 2022 y los datos de prueba serán los datos de venta de enero y posteriormente calcularemos las medidas de evaluación. Para febrero los datos de entrenamiento serán hasta enero 2023 y los datos de prueba serán los datos reales de venta de febrero. De igual forma, para marzo los datos de entrenamiento serán hasta febrero y los datos de prueba serán los datos reales de venta en marzo.

Dado que la empresa “AAA”, necesita medir la precisión de los pronósticos a nivel mes y nuestros pronósticos serán generados a nivel día se detallará ambas medias de error. La selección del modelo será aquel que menor error genere en los últimos 3 meses

Dado el análisis exploratorio, utilizaremos modelos de series temporales como ETS y ARIMA que incorporan componentes estacionales. Adicional a estos modelos utilizaremos métodos lineales que incorporan variables como TSLM (Time serie lineal model) y ARIMA CON REGRESORES. Las variables que incorporaremos a estos modelos serán relacionadas a los fines de semanas y a los días de la semana.

Para el caso de ARIMA CON REGRESORES, las variables que incorporaremos será una dummy que identifique si es fin semana (1 si es sábado o domingo) y dos variables periódicas como son el seno y el coseno.

Por otro lado, para el modelo de TSLM, incorporaremos variables de Fourier que son equivalente al seno y al coseno utilizados en el método anterior.

4.4.1 Selección del modelo: UEN #1.

En la tabla 4, se detalla los resultados obtenidos de la medida de evaluación RMSE, donde se puede observar que el modelo que mejor resultados obtiene para los datos a nivel día es TSLM donde su promedio de error día es de 1.995. A nivel individual este modelo genera los mejores resultados para el producto #1 y el producto #4. El modelo ARIMA por otro lado genera los mejores resultados para los productos #2 y #3.

Tabla 4

Evaluación RMSE día UEN #1

IdProducto	arima	arima_xreg	ets	tslm
1	3.514	3.308	3.516	2.878
ENE	3.937	4.041	4.024	3.185
FEB	3.352	3.254	3.445	2.872
MAR	3.254	2.629	3.078	2.576
2	2.428	2.572	2.592	2.445
ENE	2.565	2.351	2.589	2.296
FEB	2.684	2.938	2.643	2.672
MAR	2.036	2.426	2.544	2.366
3	1.619	1.661	1.631	1.955
ENE	1.764	1.606	1.651	1.956
FEB	1.885	1.864	1.771	2.232
MAR	1.209	1.513	1.470	1.676
4	763	717	757	704
ENE	764	563	748	562
FEB	515	536	523	517
MAR	1.008	1.053	1.001	1.033
Total	2.081	2.064	2.124	1.995

Así mismo, si evaluamos los modelos por medio del MAPE, podemos observar en la tabla 5 que el mejor modelo a nivel general es ARIMA con un 39,50%. A nivel individual, tal cual la medida anterior, este modelo genera los mejores resultados en los productos #2 y #3. El modelo TSLM sigue siendo el mejor para el producto #1 con un 45,49% y para el producto #4 el mejor modelo es ARIMA CON REGRESORES con un 23,50%.

Tabla 5*Evaluación MAPE día UEN #1*

IdProducto	arima	arima_xreg	ets	tslm
1	47,07 %	46,92 %	48,60 %	45,49 %
ENE	43,10 %	45,69 %	43,78 %	56,00 %
FEB	65,76 %	67,53 %	71,99 %	56,49 %
MAR	32,35 %	27,55 %	30,03 %	23,97 %
2	46,48 %	51,68 %	48,45 %	49,34 %
ENE	52,33 %	59,98 %	56,06 %	62,64 %
FEB	58,24 %	60,60 %	55,36 %	53,86 %
MAR	28,86 %	34,48 %	33,91 %	31,51 %
3	40,36 %	40,67 %	44,03 %	44,63 %
ENE	60,04 %	61,12 %	62,28 %	70,54 %
FEB	38,75 %	36,47 %	43,97 %	35,13 %
MAR	22,30 %	24,43 %	25,85 %	28,23 %
4	24,11 %	23,50 %	25,37 %	24,64 %
ENE	36,39 %	33,43 %	38,83 %	35,82 %
FEB	19,62 %	19,69 %	20,18 %	18,72 %
MAR	16,32 %	17,38 %	17,11 %	19,37 %
Total	39,50 %	40,70 %	41,61 %	41,02 %

En esta unidad de negocios los resultados del MAPE a nivel día generan resultado de error elevados debido a que las ventas día son influenciadas por palancas comerciales que las realizan principalmente por llegar a objetivos de venta. No obstante, si observamos este mismo indicador a nivel mes como se presenta en la tabla 6, los resultados son para todos los productos a excepción del # 3 se encuentra por debajo del 10%. De aquí podemos escoger el mejor modelo en donde el modelo TSLM es quien mejor resultados obtiene para el producto #1 y #4. Así mismo, aunque este modelo genere mejores resultados para el producto #2 se puede observar también que ARIMA genera resultados más homogéneos. Por último, para el producto #3 el mejor modelo es ETS. No obstante, en ARIMA los errores se van disminuyendo al pasar de los meses lo cual nos indica que podríamos escoger estos modelos.

Tabla 6

Evaluación MAPE mes UEN #1

IdProducto	arima	arima_xreg	ets	tslm
1	18,68 %	15,77 %	21,59 %	6,18 %
ENE	25,35 %	29,29 %	26,51 %	16,08 %
FEB	18,31 %	14,35 %	25,88 %	0,91 %
MAR	12,39 %	3,69 %	12,37 %	1,56 %
2	5,70 %	5,99 %	8,19 %	4,31 %
ENE	8,34 %	11,62 %	9,17 %	12,05 %
FEB	6,45 %	4,66 %	5,90 %	0,31 %
MAR	2,31 %	1,69 %	9,50 %	0,57 %
3	15,46 %	15,57 %	12,87 %	27,43 %
ENE	25,60 %	23,74 %	18,83 %	32,49 %
FEB	12,78 %	8,63 %	5,61 %	28,34 %
MAR	7,98 %	14,33 %	14,17 %	21,46 %
4	6,50 %	6,66 %	6,18 %	5,18 %
ENE	4,55 %	2,30 %	3,28 %	0,25 %
FEB	0,51 %	1,10 %	2,57 %	0,71 %
MAR	14,44 %	16,56 %	12,70 %	14,59 %

Dado los siguientes resultados, escogeremos los modelos TSLM para los productos #1 y #4 y ARIMA para los productos #2 y #3. En la figura 45 se muestra los pronósticos generados por TSLM para los productos #1 y #4. A su vez, la figura 46 se muestra los pronósticos generados por ARIMA para los productos restantes.

Figura 45

Pronósticos generados por TSLM para productos #1 y #4

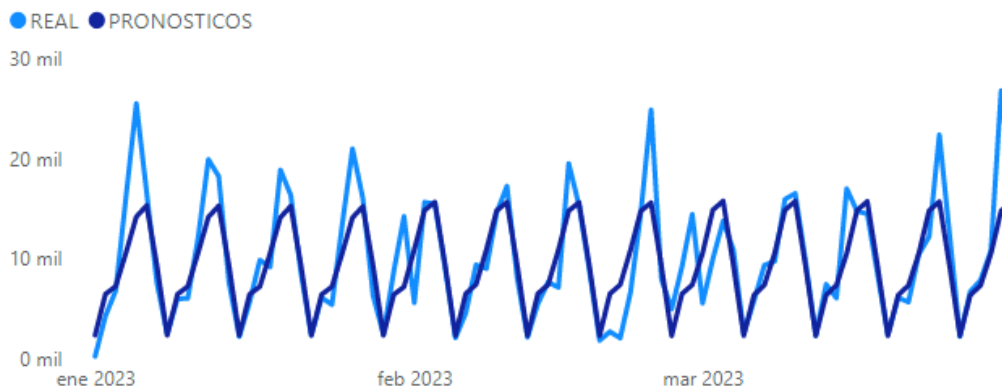
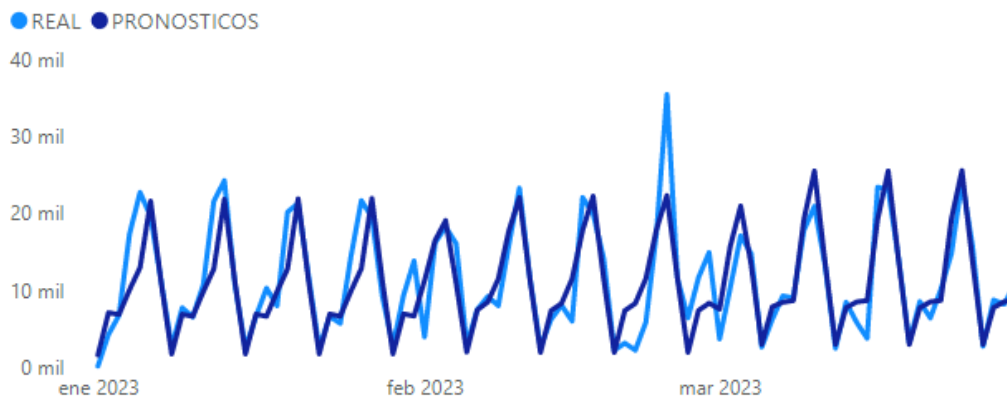


Figura 46

Pronósticos generados por ARIMA para productos #2 y #3



4.4.2 Selección del modelo: UEN #2.

En la tabla 7, se detalla los resultados obtenidos de la medida de evaluación RMSE para la UEN #2, donde se puede observar que el modelo TSLM genera mejores resultados para los productos #5 y #6. Para el producto restante, el mejor modelo es el ETS.

Así mismo, la tabla 8 se detalla los resultados obtenidos de la medida de evaluación MAPE, donde se puede observar que se repiten los mismos resultados. Donde el modelo TSLM genera un MAPE del 8.92% para el producto #5 y 9.11% para el producto #6. En esa misma dirección, el modelo ETS genera un MAPE 11.60% para el producto #7

Tabla 7*Evaluación RMSE día UEN #2*

IdProducto	arima	arima_xreg	ets	tslm
5	514	473	490	457
ENE	508	498	498	438
FEB	612	552	597	568
MAR	421	370	375	364
6	183	178	200	170
ENE	142	145	149	130
FEB	253	224	231	224
MAR	156	164	219	156
7	314	313	284	481
ENE	262	274	181	495
FEB	352	422	345	576
MAR	328	243	327	371
Total	337	321	325	369

Tabla 8*Evaluacion MAPE dia UEN #2*

IdProducto	arima	arima_xreg	ets	tslm
5	10,36 %	9,46 %	10,35 %	8,92 %
ENE	10,86 %	10,32 %	10,89 %	8,96 %
FEB	11,89 %	10,29 %	12,35 %	10,35 %
MAR	8,33 %	7,77 %	7,80 %	7,46 %
6	9,90 %	9,43 %	10,12 %	9,11 %
ENE	8,60 %	7,87 %	7,68 %	7,22 %
FEB	13,68 %	12,83 %	12,49 %	12,89 %
MAR	7,43 %	7,57 %	10,20 %	7,21 %
7	13,49 %	14,51 %	11,60 %	24,78 %
ENE	12,97 %	13,75 %	7,45 %	27,15 %
FEB	16,56 %	20,37 %	15,38 %	31,21 %
MAR	10,95 %	9,41 %	11,96 %	15,99 %
Total	11,25 %	11,13 %	10,69 %	14,27 %

A nivel mes, podemos observar en la tabla 9 que los modelos generan resultados similares entre productos. Para el producto #5 podríamos decidirnos por el modelo TSLM o ARIMA CON REGRESORES ya que ambos generan un MAPE promedio del 3%. Para el producto #6 el modelo ARIMA es el que mejores resultados genera. No obstante, los modelos TSLM y ARIMA CON REGRESORES generan resultados menores al 2.5%. Así mismo podríamos escoger entre los modelos ARIMA o ETS ya que ambos generan resultados del 4%. Dados los resultados obtenidos mediante las métricas de evaluación. Podríamos escoger el modelo TSLM para los productos #5 y #6 y escogeremos el modelo ETS para el producto #7.

Tabla 9

Evaluación MAPE mes UEN #2

IdProducto	arima	arima_xreg	ets	tslm
5	5,39 %	3,02 %	5,15 %	3,27 %
ENE	8,90 %	7,68 %	9,72 %	6,85 %
FEB	2,88 %	0,37 %	5,14 %	1,35 %
MAR	4,38 %	1,00 %	0,58 %	1,61 %
6	1,48 %	2,45 %	4,46 %	2,41 %
ENE	0,20 %	1,02 %	2,43 %	2,37 %
FEB	1,28 %	1,96 %	0,51 %	1,31 %
MAR	2,97 %	4,39 %	10,44 %	3,54 %
7	4,11 %	7,24 %	4,56 %	21,07 %
ENE	2,53 %	12,11 %	1,28 %	27,95 %
FEB	1,77 %	9,56 %	0,40 %	21,85 %
MAR	8,02 %	0,05 %	11,99 %	13,40 %

En la figura 47, se muestra los pronósticos generados por el modelo TSLM para los productos #5 y #6. Por otra parte, la figura 48, muestra los pronósticos generados por el modelo ETS para el producto #7

Figura 47

Pronósticos generados por TSLM para productos #5 y #6

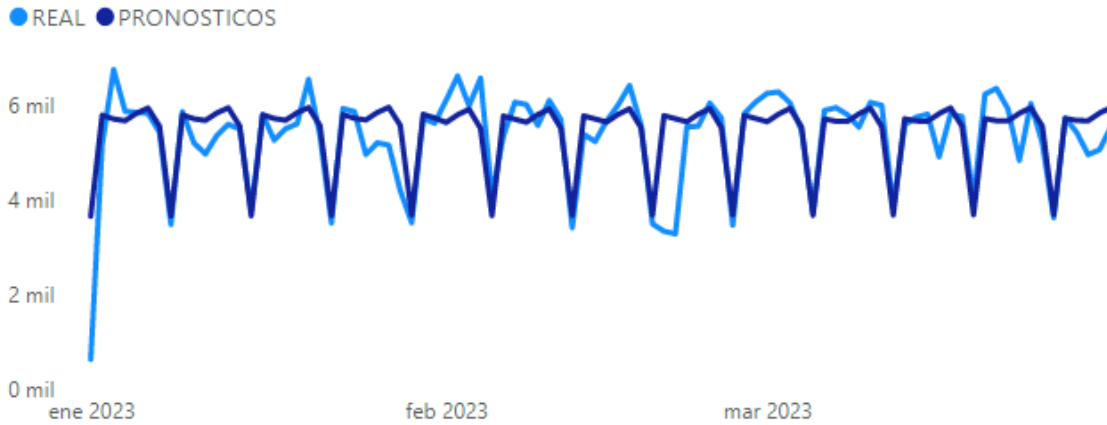
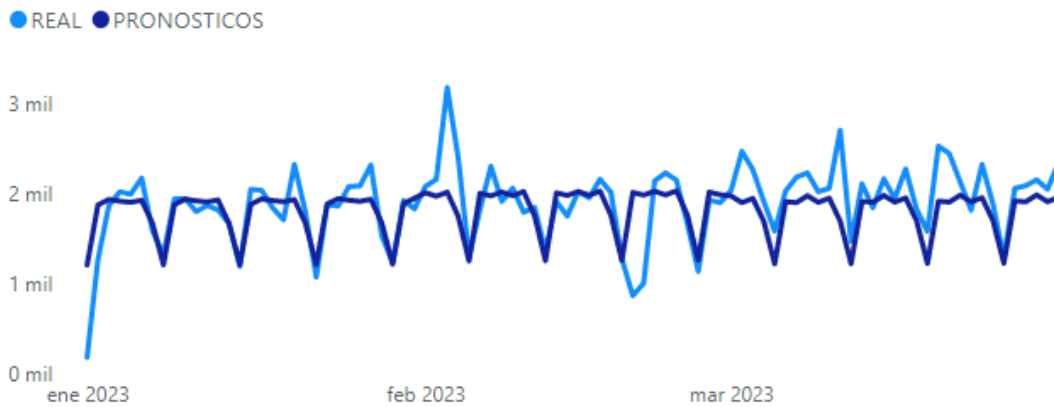


Figura 48

Pronósticos generados por ETS para productos #7



4.4.3 Selección del modelo: UEN #3.

Para la unidad de negocios #3 evaluando por la medida del RMSE. Podemos observar en la tabla 10 que el modelo de ARIMA CON REGRESORES es quien mejores resultados genera para el producto #8 con un RMSE promedio de 271. Así mismo, este modelo genera los mejores resultados para el producto #9. No obstante ARIMA y ETS generan resultados similares.

Tabla 10*Evaluación RMSE día UEN #3*

IdProducto	arima	arima_xreg	ets	tslm
8	413	271	388	330
ENE	450	238	275	334
FEB	221	192	202	244
MAR	569	383	688	413
9	454	425	446	542
ENE	515	415	433	414
FEB	441	424	511	522
MAR	407	437	394	691
Total	434	348	417	436

Siguiendo con la metodología planteada, en la tabla 11 se muestran los resultados obtenidos por la medida MAPE donde, al igual que para RMSE, el modelo ARIMA CON REGRESORES generan mejores resultados para el producto #8 con un MAPE DEL 23.45% y para el producto #9 un MAPE del 15.67%

Tabla 11*Evaluación MAPE día UEN #3*

IdProducto	arima	arima_xreg	ets	tslm
8	37,84 %	23,45 %	33,22 %	29,24 %
ENE	50,60 %	25,35 %	29,61 %	36,20 %
FEB	16,80 %	14,04 %	14,42 %	19,08 %
MAR	46,12 %	30,96 %	55,63 %	32,45 %
9	16,91 %	15,67 %	15,66 %	20,34 %
ENE	18,70 %	15,21 %	15,72 %	14,68 %
FEB	15,87 %	14,97 %	17,52 %	18,03 %
MAR	16,15 %	16,82 %	13,75 %	28,31 %
Total	27,37 %	19,56 %	24,44 %	24,79 %

A nivel mes, podemos observar en la tabla 12 que el modelo ARIMA CON REGRESORES es quien mejor desempeño cuenta para el producto #8 con un 13.84%. Para el producto #9 el mejor modelo es ARIMA con un 4.72%. No obstante, el modelo ETS y ARIMA CON REGRESORES dan resultados menores al 6%. De aquí, podemos elegir el modelo ARIMA CON REGRESORES para ambos productos. En la figura 49 se muestran los pronósticos generados con este modelo.

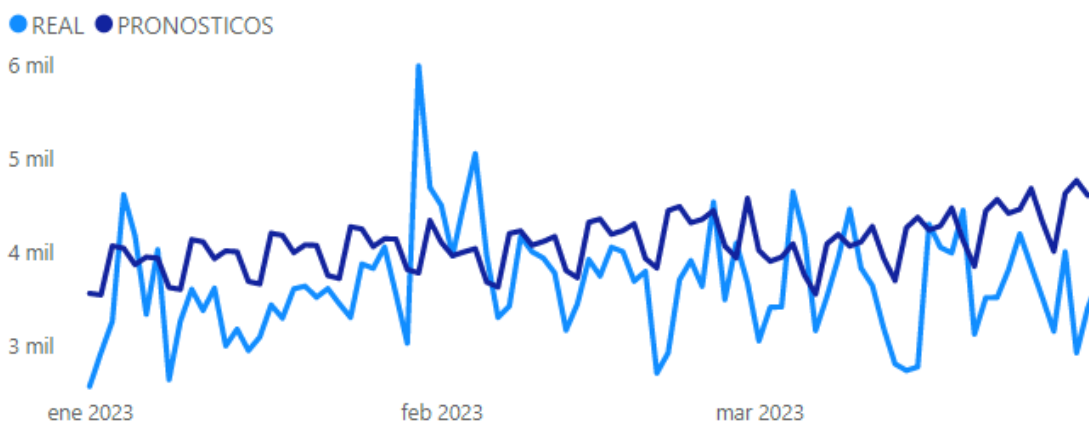
Tabla 12

Evaluación MAPE mes UEN #3

IdProducto	arima	arima_xreg	ets	tslm
8	29,25 %	13,84 %	25,60 %	22,56 %
ENE	45,58 %	18,12 %	25,54 %	33,26 %
FEB	4,73 %	0,83 %	3,45 %	8,82 %
MAR	37,42 %	22,56 %	47,79 %	25,61 %
9	4,72 %	5,92 %	5,29 %	12,84 %
ENE	7,14 %	5,17 %	5,79 %	4,39 %
FEB	4,86 %	5,91 %	10,00 %	10,60 %
MAR	2,17 %	6,69 %	0,08 %	23,52 %

Figura 49

Pronósticos generados por ARIMA CON REGRESORES productos #8 y #9



4.4.4 Selección del modelo: UEN #4.

En la tabla 13 se detalla el RMSE para la unidad de negocios #4. El modelo ETS es aquel que mejores resultados genera. No obstante, el modelo ARIMA genera resultados similares.

Desde el MAPE, en la tabla 14, el modelo ETS genera un MAPE DEL 13.20% y ARIMA con un 13.60%

Tabla 13

Evaluación RMSE día UEN #4

IdProducto	arima	arima_xreg	ets	tslm
10	1.688	1.880	1.632	2.061
ENE	1.933	2.521	1.942	1.904
FEB	1.646	1.723	1.521	1.971
MAR	1.485	1.395	1.433	2.307

Tabla 14

Evaluación MAPE día UEN #4

IdProducto	arima	arima_xreg	ets	tslm
10	13,60 %	18,40 %	13,20 %	18,04 %
ENE	14,83 %	26,04 %	14,89 %	17,34 %
FEB	14,95 %	16,15 %	13,79 %	18,06 %
MAR	11,01 %	13,00 %	10,92 %	18,71 %

Prosiguiendo nuestro análisis, a nivel mes podemos observar en la tabla 15 que el modelo que mejores resultados genera es ARIMA CON REGRESORES no obstante el modelo ETS genera resultados similares con un MAPE menor al 5%. Dado estos resultados, escogeremos el modelo ETS para la generación de pronósticos para la unidad de negocios #4. En la figura 50 se muestran los pronósticos generados con este modelo.

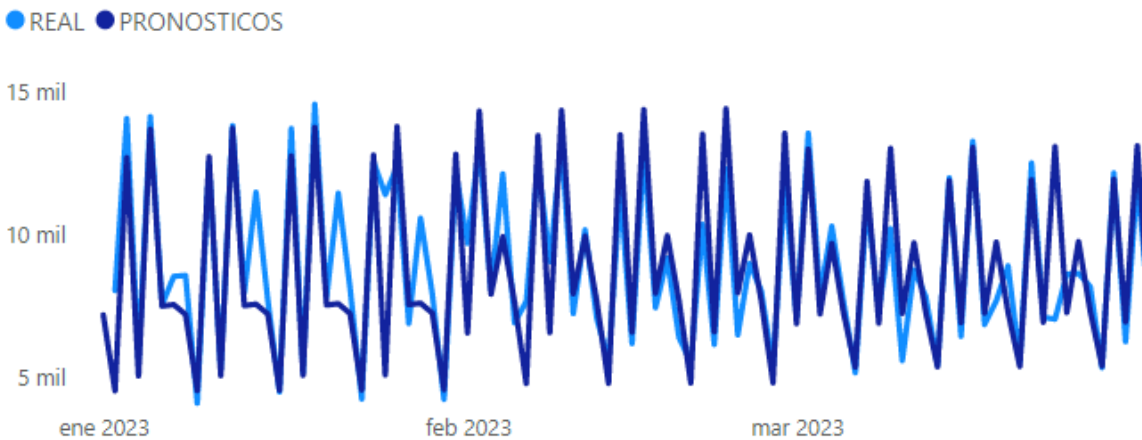
Tabla 15

Evaluación MAPE mes UEN #4

IdProducto	arima	arima_xreg	ets	tslm
10	5,41 %	4,19 %	4,51 %	6,48 %
ENE	6,00 %	6,46 %	6,77 %	1,44 %
FEB	5,80 %	4,02 %	3,21 %	4,42 %
MAR	4,43 %	2,11 %	3,57 %	13,57 %

Figura 50

Pronósticos generados por ETS producto #10



4.6 Generación de pronósticos.

En esta última sección, dado que hemos evaluado y seleccionado el mejor modelo para cada uno de los productos. Generaremos los pronósticos que servirán para la elaboración de la herramienta en PowerBI.

Los productos 1,4,5,6 serán estimados mediante pronósticos TSLM, por su parte, el producto 2 y 3 será estimado por medio de ARIMA. Los productos 8 y 9 serán estimados por medio de ARIMA CON REGRESORES y para los productos restantes 7 y 10 serán estimados por medio de los modelos ETS.

4.7 Prototipo.

Para finalizar ciclo, se procede a crear la herramienta que será de uso para la empresa “AAA”. Para ello se ha utilizado la herramienta de PowerBi. El diseño de la herramienta esta basado en la forma como la empresa visualiza los reportes.

El reporte cuenta con cinco pantallas que el usuario utilizara de forma interactiva. En la primera pantalla se muestra el LOG IN donde cada usuario que desee utilizar la herramienta deberá de colocar su usuario y contraseña. En la figura 51 se muestra la pantalla identificación del usuario

Figura 51

Pantalla de identificación del usuario



Si el usuario no ingresa la contraseña correcta le aparecerá un mensaje de error. De no ser así lo llevara a la pantalla de índice. La figura 52 muestra el mensaje de error y la figura 53 muestra la pantalla de índice donde el usuario podrá ingresar a la pantalla de pronósticos.

Figura 52

Pantalla de error de contraseña



Figura 53

Pantalla de Índice

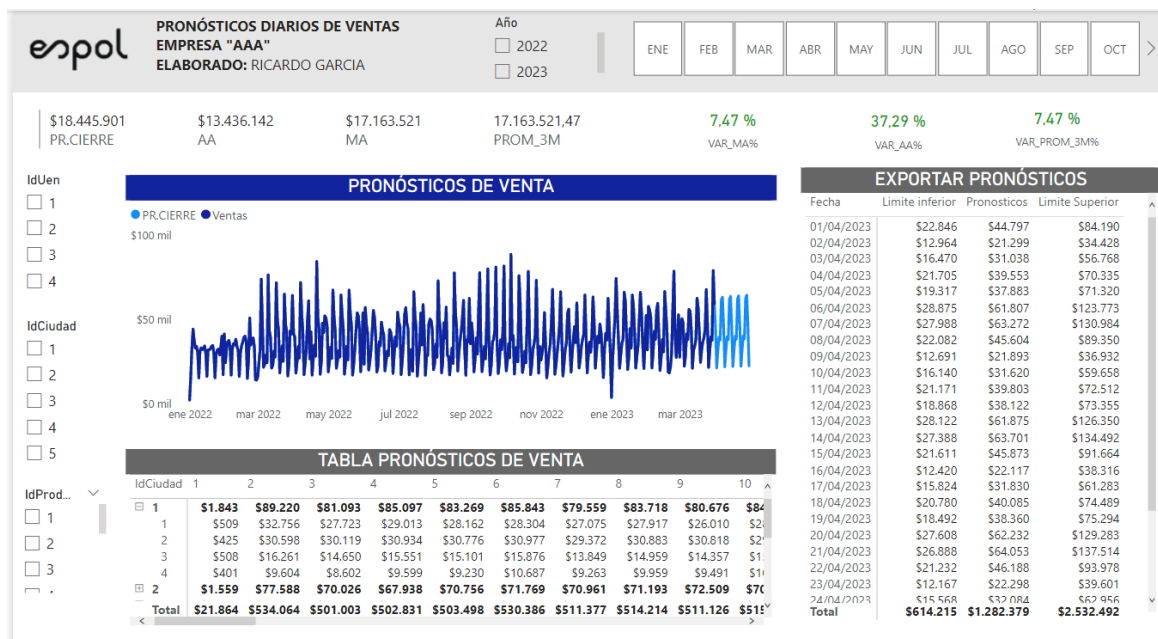


Si bien el prototipo cuenta con pantallas relacionado al análisis exploratorio y evaluación de modelos que hemos analizado a lo largo del presente capitulo. El usuario final ingresara directamente a la pantalla de pronósticos que detallaremos a continuación.

La figura 54 muestra el reporte de pronósticos. En la parte superior del reporte se encuentran los segmentadores correspondiente al tiempo (Año y Mes) con esto el usuario podrá visualizar el evolutivo de ventas y entender los pronósticos.

Figura 54

Pantalla de PRONÓSTICOS



Debajo de estos segmentadores se encuentra un tablero de indicadores que sirve como referencia para los pronósticos. Estos indicadores tienen coherencia cuando se selecciona un año y un mes en específico tal como se presenta en la figura 55. Los indicadores de referencias que utiliza esta empresa son: AA que hace referencia a la venta del año anterior, MA que hace referencia a la venta del mes pasado y PROM 3M que hace referencia a los últimos tres meses de venta. Este último indicador muestra como viene la tendencia. Si bien los pronósticos están a nivel diario, la empresa “AAA” siempre observa cual es la proyección de cierre por eso estos indicadores se miden a nivel mes.

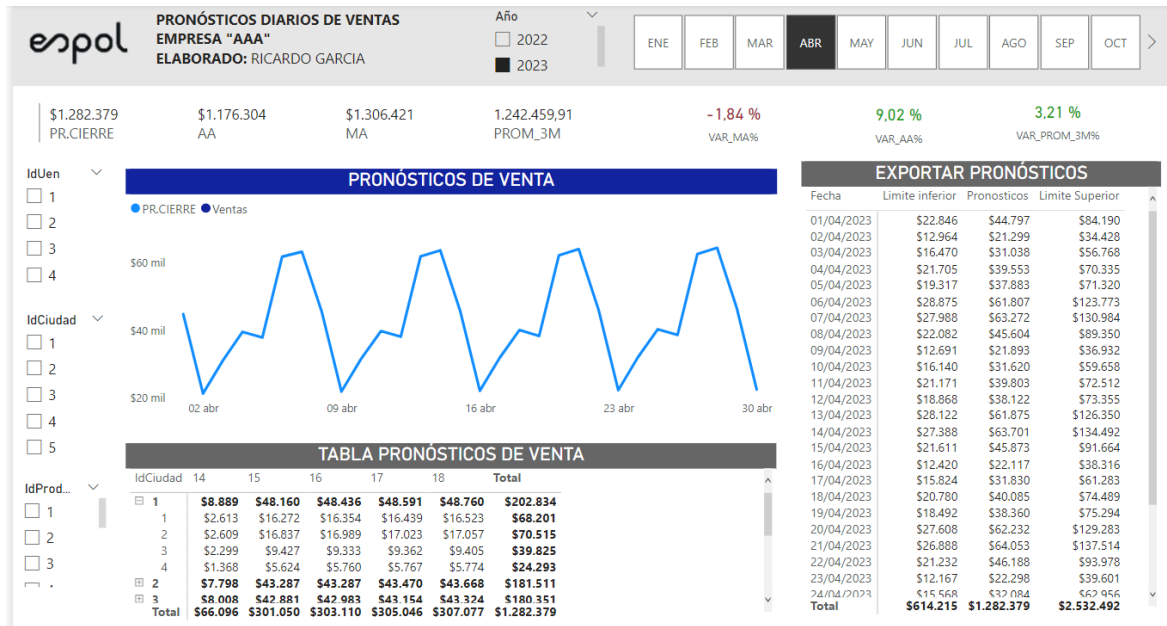
En la parte izquierda del reporte se encuentran los segmentadores correspondiente a los niveles de agrupación que utiliza la empresa esto son: Unidad de negocios, ciudad y producto, con ello el usuario podrá obtener pronósticos específicos de acuerdo con sus necesidades.

En la parte central del pronóstico se detalla un gráfico evolutivo y una tabla a nivel semana en la cual se tiene tantos datos reales y los proyectados.

Por último y no menos importante, en la parte derecha, se muestra una tabla en la cual se detalla tanto los pronósticos como sus intervalos al 95% con ello el usuario podrá tener flexibilidad de poder escoger el número que más le haga sentido. Esto es importante ya que la unidad de negocios #1 como se mencionó previamente, cuenta con palancas comerciales que la fuerza de ventas utiliza para poder llegar al objetivo. Con esto, si el departamento comercial decide implementar una palanca agresiva el usuario podrá utilizar el intervalo optimista. Así mismo, si se decide no implementarla ninguna palanca el usuario podría utilizar el intervalo pesimista.

Figura 55

Pantalla de PRONÓSTICOS filtrado por mes en curso



5. CONCLUSIONES Y RECOMENDACIONES

5.1 CONCLUSIONES

- Los modelos aplicados en esta investigación demuestran que no hay grandes diferencias entre los resultados obtenidos. En la mayoría de los productos, es posible haber escogido al menos un modelo más.
- Para los productos que cuentan con una marcada estacionalidad semanal la inclusión de series de Fourier no solo genera mejores resultados sino también mejor ajuste.
- El conocimiento del negocio es fundamental para la generación de los pronósticos ya que es posible incorporar al modelo ciertas variables que mejoran los modelos.
- Los datos previos al 2022 han sido afectados por la pandemia este es el caso para los años 2020 y 2021. Previo a estos años, el nivel de venta difiere considerablemente con el actual es por eso que se ha tomado los datos desde el 2022 año en el cual se comenzó a estabilizar el mercado. No obstante, debido a esto no se ha podido detectar si a nivel mes existe alguna estacionalidad.
- Dado lo mencionado en el punto anterior, estos modelos son útiles con una frecuencia diaria menor a dos años, fuera de este límite es recomendable la utilización de otros modelos de aprendizaje automático que permita incorporar varias estacionalidades.

- La herramienta de Power BI ha sido de gran utilidad en todas las etapas del proceso. En la fase exploratoria, ha permitido realizar visualizaciones de forma muy rápida y sencilla, además, permite incorporar código R para gráficos mas complejos. En esa misma dirección, esta herramienta nos permita comparar rápidamente los pronósticos de distintos modelos simplemente exportando los resultados y añadiéndolo al modelo de datos.

5.2 RECOMENDACIONES

- Para la unidad de negocios #1, la incorporación de alguna variable que haga referencia a los descuentos comerciales, gasto de marketing o crédito otorgado ayudaría a mejorar los resultados.
- Es importante dar constante seguimiento a los pronósticos generados por la herramienta. Para ello, la empresa “AAA” necesita contar con un profesional que pueda dar mantenimiento
- Se recomienda que los datos se actualicen a nivel semanal o quincenal con lo cual se tenga mejores resultados a corto plazo.
- La empresa debe dar prioridad a los proyectos que involucren análisis y ciencia de datos. No solo contando con profesionales de esta área sino también brindando las herramientas a los usuarios para que puedan generar soluciones en base a datos.

6. REFERENCIAS

- Armstrong, J. S. (2012). *Forecasting for Marketing*.
- B.Kedem. (2005). *Regression models for time series analysis*.
- Barker, M. J. (1999). SALES FORECASTING .
- Bienstock, M. a. (1998). Seven keys to better forecasting.
- Cano, J. L. (2007). *BUSINESS INTELLIGENCE: competir con informacion*.
- CONSTANTIN, C. (2016). The importance of Sales forecasting in establishing marketing strategies.
- D. Polo, L. P. (2015). "Comparison of Neural Network applied to prediction of times Series".
- Diaz, J. (2012). *Introduccion al Business Intelligence*.
- Fulton, C. (2017). *Time series anaysis and its application with R example*.
- Gallegos, M. (2014). Pronosticos paa negocios.
- Haberleitner, M. &. (2010). Implementation of a demand planning system using advance order information.
- Hawke, J. (2005). *Business Forecasting*. Pearson.
- Hurtado, C. (2008). *Series de Tiempo*. Universidad de Chile.
- Juárez, A. C. (2016). Análisis de series de tiempo en el pronóstico de la demanda.
- Koehler, A. B. (2008). *Forecasting with exponential smoothing*.
- Kumar, A. (2019). *Industrial Marketing Management*.
- KURZAK, L. (2012). IMPORTANCE OF FORECASTING IN ENTERPRISE MANAGEMENT.

- Leon, F. (2008). *Bases de datos relacionales: teoría y practica*.
- Microsoft. (2023). *learn.microsoft.com*. Obtenido de <https://learn.microsoft.com/es-es/sql/relational-databases/views/views?view=sql-server-ver16>
- Peixeiro, M. (2022). *Time series Forecasting in Python*.
- Petropoulos, F. (2022). *Forecasting: Theory and practice*.
- Rabi, F. &. (2012). Estimating benefits of Demand Sensing for consumer goods organisations.
- Rob J Hyndman, G. A. (2021). *Forecasting: Principles and Practice*.
- Rojo, J. (2005). *Analisis descriptivo y explotario de datos*.
- Ronald K. Klimberg, G. P. (2010). FORECASTING PERFORMANCE MEASURES – WHAT ARE THEIR PRACTICAL MEANING?
- Shokouhifar, M. M. (2022). Multivariate time-series blood donation/demand forecasting for resilient supply chain management during COVID-19 pandemic.
- Silva Solano, L. E. (2017). *Business Intelligence: un balance para su implementación*.
- Tsay, R. S. (2005). *Analysis of Financial Time Series*.
- Vieira, A. I.-N. (2023). Forecasting daily admissions to an emergency department considering single and multiple seasonal patterns.
- Xindong Wu, F. (2014). *Data Mining with Big Data*.

7. ANEXOS

Anexo 1

Detalle de Campos del Dataset

VARIABLES	TIPO	DESCRIPCION
Fecha	Date	Corresponde a la fecha en la que se realizó cada venta.
IdUen	int	Indica la identificación de la Unidad de negocio de la empresa
IdCiudad	int	Corresponde a la identificación de la ciudad donde se realiza la venta
IdProducto	int	Corresponde a la identificación del producto
Total	Float	Corresponde al ingreso generado por las unidades vendidas