



UNIVERSIDAD DE TALCA
FACULTAD DE ECONOMÍA Y NEGOCIOS
ESCUELA DE INGENIERÍA INFORMÁTICA EMPRESARIAL

DESARROLLO DE MODELO PREDICTIVO PARA LA ESTIMACIÓN DE DEMANDA DE *PELLETS* DE USO INDUSTRIAL Y DOMICILIARIA PARA LA EMPRESA FORESTAL RIO CLARO.

Autor: Jorge Arturo Meza Carrasco

Profesor Guía: Luis Eduardo Canales

Proyecto de memoria para optar al título de INGENIERO INFORMÁTICO EMPRESARIAL

TALCA – CHILE
2022

CONSTANCIA

La Dirección del Sistema de Bibliotecas a través de su unidad de procesos técnicos certifica que el autor del siguiente trabajo de titulación ha firmado su autorización para la reproducción en forma total o parcial e ilimitada del mismo.



Talca, 2023

ÍNDICE DE CONTENIDOS

| | | |
|-------|--|----|
| 1 | CAPÍTULO: INTRODUCCIÓN..... | 8 |
| 1.1 | Objetivo general | 10 |
| 1.2 | Objetivos específicos..... | 10 |
| 2 | CAPÍTULO: MARCO TEÓRICO..... | 11 |
| 2.1 | ENERGIAS RENOVABLES NO CONVENCIONALES (ERNK)..... | 11 |
| 2.1.1 | Biomasa..... | 13 |
| 2.2 | MODELOS Y HERRAMIENTAS DE ANÁLISIS PREDICTIVOS..... | 17 |
| 2.2.1 | Tipos de modelos predictivos..... | 17 |
| 2.2.2 | Técnicas de modelado..... | 19 |
| 3 | CAPÍTULO: METODOLOGÍA | 24 |
| 4 | CAPÍTULO: DESARROLLO DE METODOLOGÍA Y ANÁLISIS DE RESULTADOS 27 | |
| 4.1 | Descripción de la empresa | 27 |
| 4.2 | Comprensión del negocio..... | 28 |
| 4.2.1 | Objetivos del negocio..... | 28 |
| 4.2.2 | Valoración de la solución..... | 29 |
| 4.2.3 | Objetivos del análisis de datos..... | 30 |
| 4.2.4 | Realizar secuencia de proyecto..... | 30 |
| 4.3 | Comprensión de los datos..... | 31 |
| 4.3.1 | Recolección datos iniciales | 31 |
| 4.3.2 | Descripción de los datos..... | 32 |
| 4.3.3 | Exploración de los datos..... | 36 |
| 4.3.4 | Verificar la calidad de los datos..... | 37 |
| 4.4 | Preparación de los datos | 38 |

| | | |
|-------|---|----|
| 4.4.1 | Selección de los datos | 38 |
| 4.4.2 | Limpieza de los datos..... | 40 |
| 4.5 | Estructuración y formateo de los datos | 41 |
| 4.6 | Modelado..... | 44 |
| 4.6.1 | Selección de técnicas de modelado: | 46 |
| 4.6.2 | Generación de un plan de pruebas..... | 47 |
| 4.6.3 | DESARROLLO DE MODELOS PREDICTIVOS..... | 49 |
| 4.6.4 | Promedio móvil simple (<i>simple moving average SMA</i>) | 52 |
| 4.6.5 | Suavizado exponencial simple (<i>simple exponential smoothing SES</i>)..... | 54 |
| 4.6.6 | Suavizado exponencial doble o Método de Holt (<i>Double exponential Smoothing</i>)..... | 57 |
| 4.6.7 | Método de Holt-Winters (<i>triple exponential Smoothing</i>)..... | 60 |
| 4.6.8 | ARIMA Y AUTOARIMA (<i>autoregressive integrated moving average</i>). | 64 |
| 4.7 | MODELOS DEMANDA INDUSTRIAL. | 69 |
| 4.7.1 | Media Movil Industrial (SMA)..... | 71 |
| 4.7.2 | Suavizado exponencial simple industrial (SES)..... | 72 |
| 4.7.3 | ARIMA industrial. | 73 |
| 4.8 | EVALUACIÓN..... | 74 |
| 4.8.1 | Evaluación de los resultados | 74 |
| 4.8.2 | Revisión | 77 |
| 4.8.3 | Determinación próximos pasos | 80 |
| 4.9 | PLAN DE IMPLEMENTACIÓN | 81 |
| 5 | Conclusiones..... | 84 |
| 6 | REFERENCIAS..... | 86 |
| 7 | ANEXOS | 89 |

INDICE DE FIGURAS

| | |
|---|----|
| Figura 1: Procesos de transformación de biomasa. Fuente: Patiño 2014..... | 14 |
| Figura 2: Diagrama de modelo CRISP DM (elaboración propia)..... | 26 |
| Figura 3: Grafico de comportamiento ventas 2021 | 36 |
| Figura 4: Grafico comportamiento de ventas 2022 | 37 |
| Figura 5: Resultados de la funcion summary | 40 |
| Figura 6: Comportamiento de ventas mensuales..... | 52 |
| Figura 7: Proyección 1 mes con SMA | 53 |
| Figura 8: Proyección 5 meses con SMA..... | 54 |
| Figura 9: Proyección 1 mes con SES..... | 56 |
| Figura 10: Proyección 5 meses con SES..... | 57 |
| Figura 11: Proyección 1 mes con Holt..... | 59 |
| Figura 12: Proyección 5 meses con Holt | 60 |
| Figura 13:Comportamiento de datos al replicar un periodo..... | 63 |
| Figura 14: Proyecciones con datos replicados usando HW | 64 |
| Figura 15: Proyecciones ARIMA con co-variable..... | 69 |
| Figura 16: Comportamiento de ventas industrial | 70 |
| Figura 17: Proyecciones ventas industriales usando SMA..... | 71 |
| Figura 18: Proyección de ventas industriales usando SES | 72 |
| Figura 19: Proyecciones de ventas SES para 5 meses | 73 |

INDICE DE TABLAS

| | |
|---|----|
| Tabla 1 Plantas en regiones (fuente INFOR) | 9 |
| Tabla 2 Comparación de herramientas para análisis de datos (Fuente: Aqlan et al., 2020) | 23 |
| Tabla 3 Secuencia de proyecto genérico para la creación del modelo final | 31 |
| Tabla 4 Títulos de datos entregados por la empresa | 32 |
| Tabla 5 Contenido resumen ventas de pellets..... | 33 |
| Tabla 6 Tabla de ventas netas..... | 33 |
| Tabla 7 Resumen de la cantidad de ventas en general | 34 |
| Tabla 8 Cantidad productos venta nacional | 34 |
| Tabla 9 Detalle de las ventas realizadas | 34 |
| Tabla 10 Columnas eliminadas de la base de datos | 39 |
| Tabla 11 Columnas que se dejaron en la base de datos | 39 |
| Tabla 13 Resumen de base de datos..... | 44 |

RESUMEN

El análisis de datos de series temporales como un recurso que permita obtener ventajas se ha vuelto un gran aliado e importante para las empresas. Forestal Rio Claro es una empresa que siente la necesidad de aprovechar este recurso para mejorar la incertidumbre de stock en la demanda de uno de sus productos como lo es el *pellet*. El presente proyecto de tesis se encuentra enfocado en la realización de algunos modelos de análisis de datos que permitan poder predecir la creciente demanda de *pellets* en la que se ha visto enfrentada la empresa Forestal Rio Claro, la cual ha facilitado los registros de sus ventas para asegurar el cumplimiento de los objetivos del proyecto, donde se acude al uso de principalmente 2 softwares para trabajar con datos, MS Excel y R Studio, los que se destacan por su baja curva de aprendizaje y bajo costo de utilización, así también como la gran capacidad y funcionalidades para trabajar con datos. Los resultados mostraron que hay modelos más efectivos que otros y también que permiten estimar mayor cantidad de periodos futuros. Sin embargo, todos requieren de una gran cantidad de datos para un mejor rendimiento. Se puede concluir que un buen modelo de estimación de demanda, junto con la cantidad de datos necesarios puede ser un gran aporte a la hora de prevenir la falta de stock y apoyar a un mejor funcionamiento del área productiva. Se espera que este proyecto pueda ser ejecutado en el futuro con el surgimiento de nueva información y una mayor cantidad de datos.

Palabras claves: demanda, modelos predictivos, series temporales.

SUMMARY

The analysis of time series data as a resource to obtain advantages has become a great and important ally for companies. Forestal Rio Claro is a company that feels the need to take advantage of this resource to improve the uncertainty of stock in the demand of one of its products such as *pellets*. This thesis project is focused on the realization of some data analysis models that allow to predict the growing demand for *pellets* in which the company Forestal Rio Claro has been facing, which has provided the records of its sales to ensure compliance with the objectives of the project, where the use of mainly 2 softwares to work with data, MS Excel and R Studio, which stand out for their low learning curve and low cost of use, as well as the great capacity and functionality to work with data. The results showed that some models are more effective than others and also allow estimating a greater number of future periods. However, all of them require a large amount of data for better performance. It can be concluded that a good demand estimation model, together with the necessary amount of data, can be a great contribution to prevent stock-outs and support a better performance of the production area. It is hoped that this project can be implemented in the future with the emergence of new information and a greater amount of data.

Keywords: demand, predictive models, time series.

1 CAPÍTULO: INTRODUCCIÓN

El aumento del precio de los combustibles tradicionales (fósiles), la contaminación del planeta, la incesante búsqueda por encontrar nuevas fuentes de energías renovables, entre otros factores que han llevado a que las empresas comiencen a ver la biomasa como un recurso de valor considerable para el futuro (Pinilla, Luengo, Navarrete y Navarrete, 2020), todo esto entrega a las energías renovables un gran protagonismo debido a las nuevas exigencias gubernamentales motivadas por el calentamiento global, además de otros incentivos generados por los gobiernos como el de Chile con el programa de recambio de calefactores en los hogares, entregando una estufa a *pellets* a cambio de una tradicional. La empresa forestal Rio Claro se encuentra ubicada en la región del Maule, Chile y durante los últimos años ha trabajado arduamente produciendo con éxito esta valiosa fuente de energía extraída de desechos de madera (aserrín en su mayoría). Estos desechos de madera se obtienen de los diferentes procesos productivos con los que cuenta la empresa. Por otro lado, la satisfacción del cliente es de gran importancia, por lo que dentro de los factores de mayor relevancia para poder cumplir con este requisito es lograr anticiparse a la demanda del producto (Herrera, García, Izquierdo y Pérez, 2009). Para entregar la mayor satisfacción a sus clientes la empresa busca encontrar la manera de producir constantemente los *pellets* necesarios que ayuden a cubrir la demanda que sus clientes requieren.

Motivado por entregar la mejor calidad de servicio a sus clientes, es donde aparece la necesidad del gerente de operaciones de la empresa (Forestal Rio Claro) de poder anticiparse a la demanda que se irá generando mes a mes sobre el producto, debido a que actualmente no existe un método definido para poder estimar esta demanda del *pellet*, lo que le genera incertidumbre con respecto al proceso productivo. Basado en los datos almacenados, como por ejemplo datos de ventas y cotizaciones se busca encontrar un modelo predictivo de demanda que permita ayudar en los procesos productivos y anticiparse a la demanda con la que contará la empresa, tanto la demanda que se va a requerir para *pellets* de uso domiciliario, como también lo que se va a requerir para uso industrial. Según estadísticas del instituto forestal (INFOR),

aproximadamente 82% de la producción de *pellets* es para uso domiciliario, mientras que solo un 18% es para uso industrial e institucional, siendo la región del Bío Bío la que cuenta con un 70% de producción aproximado (Suarez et al., 2020). La producción de *pellets* se lleva a cabo principalmente en 3 regiones, Bío Bío, Maule y Araucanía.

Tabla 1 Plantas en regiones (fuente INFOR)

| REGIONES | plantas productivas | |
|------------------|---------------------|---------------------------------|
| | Numero de plantas | Rango de años de funcionamiento |
| BÍO BÍO | 9 | 1-13 |
| MAULE | 5 | 2-4 |
| ARAUCANIA | 5 | 3-5 |

Podemos notar que, dentro de las regiones con mayor producción de *pellets*, de las 3 principales, la del Maule ha sido una de las últimas en incorporarse al proceso, pero en los últimos años ha tenido un notable crecimiento. Empresas como Forestal Rio Claro son las que ayudan al crecimiento de la región en la generación de energías renovables. Se espera que, con la incorporación de un modelo predictivo para la demanda, ayude a la empresa a reducir los riesgos de stock y optimizar sus operaciones, entregando mayor seguridad en la toma de decisiones.

Para este proyecto se definieron los siguientes productos entregables para dar solución a las necesidades de la empresa en el proceso de la estimación de la demanda:

Los productos esperados para este proyecto son los siguientes:

- Modelos de estimación de demanda de *pellets* para uso domiciliario.
- Modelos de estimación de demanda de *pellets* para uso industrial.
- Propuesta de proceso para estimar la demanda utilizando los modelos antes mencionados.

Los objetivos establecidos para este proyecto, tanto general como específicos son los siguientes.

1.1 Objetivo general

- Desarrollar modelos predictivos para la estimación de demanda de biomasa (*pellets*) para la empresa Forestal Rio Claro, tanto la demanda que se requiere de *pellets* para uso de las industrias y para uso domiciliario.

1.2 Objetivos específicos

- OE1: Identificar las variables que influyen en la empresa para la demanda de *pellets*.
- OE2: Establecer criterios de validez para dichas variables.
- OE3: Estimar la demanda de *pellets* mediante modelos de análisis de datos.
- OE4: Entregar modelo con propuesta de proceso para el uso de este en el software R para la estimación de la demanda.

2 CAPÍTULO: MARCO TEÓRICO

El desarrollo de este marco teórico está enfocado desde lo macro a lo micro, buscando entregar el conocimiento necesario para comprender el estado actual de las energías renovables no convencionales (ERNC) al que pertenecen los *pellets* (biomasa), además dar a conocer por que la biomasa y la familia las ERNC son importantes de desarrollar, así también como es que se ha fomentado la utilización y las dificultades que se deben enfrentar. Luego de entregar un contexto general ya se especifica más en el proyecto donde se hace una revisión sobre la biomasa, los modelos predictivos para estimar la demanda de la empresa y las herramientas que se pueden utilizar para llevar a cabo dichos modelos.

2.1 ENERGIAS RENOVABLES NO CONVENCIONALES (ERNC)

Basado en las políticas económicas para el año 2050, se espera crear mayores incentivos al uso de las energías renovables no convencionales (ERNC), siendo estas de vital importancia para el ambiente económico de los países (Pinilla et al., 2020). Con estas medidas se espera lograr que un 70% de la energía utilizada en el 2050, provenga de ERNC, donde Chile es un referente tanto en inversión nacional como en extranjera (Romero, 2017). Es entonces que, si en Latino América esperan un mayor crecimiento en energías renovables, tendrán que adecuarse a una gobernanza que les permita manejar los múltiples desafíos que se enfrentan al implementar nuevas fuentes de energía, ya sean fiscales, regulatorios, inversión pública entre otros (Altomonte, 2017).

Chile lleva años trabajando en el desarrollo de ERNC, habiendo tenido gran avance entre los años 2012-2016, siendo la energía eólica y solar las que lograron un gran avance significativo, donde se aprovechan del potencial de recursos naturales del país (Romero, 2017). Además de que en el gobierno de 2013 se propusieron metas a la generación de energía eléctrica entre 2014 y 2025 donde la energía del país debiera

provenir de energías renovables no convencionales (Rudnick, 2014). En dichas metas se estableció en la ley 20/25 que para el año 2025 se debía alcanzar 20% de participación en el mercado de energías renovables no convencionales (Instituto Libertad, 2018). Si bien se cree que vamos bien encaminados, Messer (2013) cree que para poder aumentar el desarrollo de las ERNC con mayor rapidez se necesitan 2 factores que trabajen en conjunto, la disponibilidad tecnológica y las políticas de fomento, aunque Romero (2017), plantea que existen más factores que son importantes a la hora de ayudar al desarrollo de las ERNC, tales como; voluntad política, incrementos relevantes de demanda, acceso financiero bancario y cumplimiento de la ley 20.698.

La ley 20.698 se creó en el año 2013 como una forma de incentivar u obligar a que las empresas generadoras de energía incorporen ERNC en su entrega final a los clientes, de no cumplirse se establecen sanciones económicas (Rebolledo, 2014). Como parte de las ERNC en Chile se identifican fuentes de energía como la eólica, solar, pequeña hidroeléctrica, biogás, geotermia, energía de los mares y la biomasa (Instituto libertad, 2018). La biomasa fue una de las energías renovables con más rápido crecimiento en participación de mercado, alcanzando rápidamente un 45% de participación en el año 2014 posterior a la ley, seguido por la hidráulica (23%), eólica (22%) y solar con 9,9% (Rudnick, 2014). Un destacado medio internacional en el año 2017 destacó a Chile por su avance en energías renovables, esto debido a las gestiones realizadas por el primer gobierno de Sebastián Piñera donde un 74% de proyectos aprobados en materia energética eran energías renovables (Instituto Libertad, 2017). Por otro lado, Altomonte (2017) postula que los avances de Chile en materia de energías renovables se deben principalmente a las mejoras de las tecnologías involucradas, además de que las propuestas del gobierno no ayudaron a expandir las ERNC, siendo el propio mercado el que se ha encargado de asignar más recursos a ese sector.

La inserción de las energías renovables en Chile no ha estado absuelta de dificultades y desafíos que se deben enfrentar, en materia de energía eléctrica existe desde un principio una complejidad en la viabilidad para conectarse a la red eléctrica (Rudnick, 2014). También dentro de la lucha por ganar licitaciones para proyectos de

energías renovables existen empresas que quedan fuera y estas suscriben contratos con clientes que demandan pequeñas cantidades, lo que hace reducir la energía que se le demandará a las empresas adjudicadas (Romero 2017). En cuanto al funcionamiento se cree que algunas fuentes de energías renovables pueden ser inestables, así también como los precios pueden ser variables dejando de ser un negocio interesante para las empresas de energía (Instituto libertad, 2018).

2.1.1 Biomasa

García (2002) define la biomasa como toda la materia orgánica que se origina de procesos biológicos, donde el componente energético proviene principalmente del sol, además menciona la existencia de 3 tipos de biomasa;

- Biomasa vegetal: aquella que proviene de las plantas, a partir de la fotosíntesis. Existen cultivos energéticos que tienen la función de producir biomasa para transformarla en combustible.
- Biomasa animal: producida por animales que consumen vegetales, aunque su uso para fines energéticos es poco considerable.
- Biomasa residual: se produce de los desechos orgánicos que se generan en los trabajos realizados por las personas donde se involucran plantas o animales. Algunas subdivisiones de la biomasa residual son: la biomasa Industrial, biomasa urbana y biomasa agraria. A esta última pertenecen los residuos forestales.

Dentro de las transformaciones que suceden con la biomasa es posible obtenerla en 3 estados de la materia, en forma sólida, líquida y gaseosa (Pradhan, Mahanjani y Arora, 2018). Los biocombustibles sólidos provienen principalmente de los residuos y los líquidos provenientes de azúcares y aceites, donde el transporte del sólido es más complejo por lo que es recomendable producirlo en la misma zona donde se convertirán en energía (Velázquez, 2018). Existen procesos con usos de tecnologías definidos que se deben utilizar para la producción de energía mediante la biomasa, Patiño (2014) señala; la combustión, gasificación, pirolisis y procesos biológicos. Además, que para la

conversión de la biomasa hay 2 caminos conocidos; el bioquímico que utiliza enzimas y microorganismos para convertir la biomasa en líquidos y termoquímico que convierte la biomasa en formas de energía mediante uso de calor y control de oxígeno (Pradhan et al., 2018). Los procesos y tecnologías más conocidos y utilizados de transformación de biomasa los describe Patiño (2014) de la siguiente manera:

- **Combustión directa:** Proceso que no requiere tratamiento previo de la biomasa, donde se busca generar calor, ya sea en los hogares o en las industrias para la generación de vapor a alta temperatura. Para este proceso se debe introducir la biomasa en una caldera a alta temperatura.
- **Co-combustión:** Uso de caldera que se alimenta de carbón, teniendo un comportamiento más estable para procesar la biomasa.
- **Gasificación:** proceso donde se puede obtener principalmente dióxido de carbono y amoniaco. Es un proceso termoquímico donde se transforma un sólido en gaseoso. Este método aún está en investigación, se ha desarrollado electricidad en algunas pruebas.
- **Pirolisis:** proceso que convierte la biomasa en combustibles útiles mediante degradación térmica sin oxígeno. Su descomposición se realiza utilizando calor.
- **Procesos Biológicos:** proceso mediante el que se obtiene el bioetanol y biogás, se generan por fermentación y digestión.

En la figura 1 se resume lo planteado por Patiño (2014) respecto a las tecnologías y procesos asociados a la transformación de biomasa.

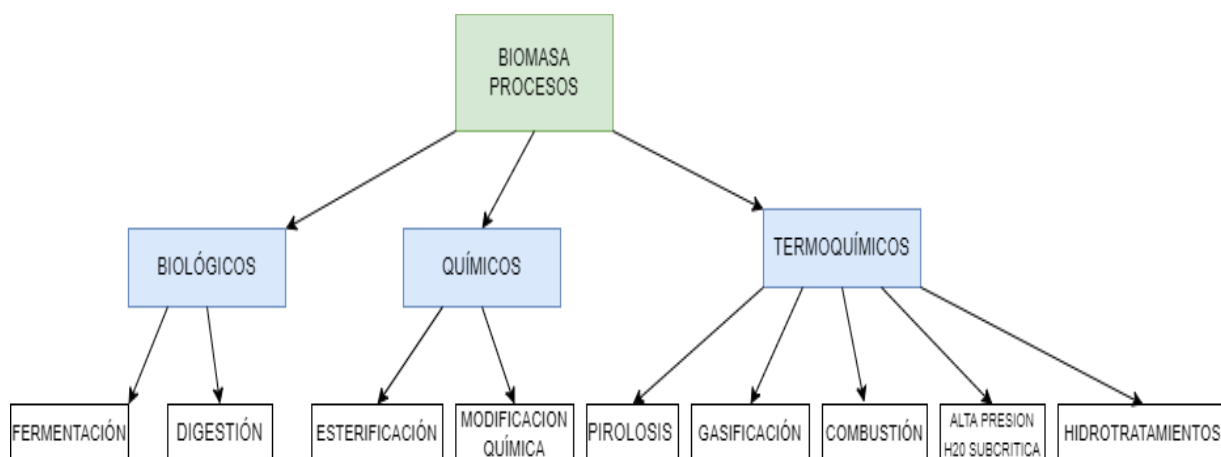


Figura 1: Procesos de transformación de biomasa. Fuente: Patiño 2014

En los países desarrollados la forma más popular de biomasa es la utilización de *pellets* para la calefacción, siendo los residuos de la industria maderera la materia prima más común para el proceso de peletización. En Chile la industria de la madera es el principal sector de suministro de biomasa, con un 87% y solo 10% entre subproductos agrícolas, animales y cultivos energéticos (Pinilla et al., 2020). Pero no todos los residuos forestales sirven para aplicaciones de energía, ya que también se pueden crear subproductos de la madera en base a estos; siendo los serrines, virutas, corteza y polvo de lijado los principales residuos para uso energético (Velázquez, 2018).

El *pellet* por su parte es un biocombustible sólido de 6 milímetros de diámetro y entre 10-30 milímetros de largo, se obtiene prensando los residuos madereros (aserrín). En Chile 97% es producido a partir de pino radiata (Pinilla et al., 2020). Caracterizado por su alta densidad y baja humedad permite una eficiencia de conversión energética de alrededor de 75%, además su forma procesada permite generar menos emisiones de partículas contaminantes (Pradhan et al., 2018). Estos desechos deben ser triturados, molidos y compactados de modo que se obtenga una alta densidad, para luego obtener una forma cilíndrica que se corta en trozos de 3 a 6 cm (Velázquez, 2018). Para su fabricación es necesario contar con la maquinaria adecuada y seguir el proceso establecido, el que Pinilla et al., (2020) lo resume en los siguientes pasos:

- **Abastecimiento:** la empresa puede trabajar con abastecimiento propio o externo, es decir utilizando desechos propios o abastecerse de manera externa comprando la biomasa a un proveedor para producir el *pellet*.
- **Secado:** en este proceso se debe quitar la humedad de la biomasa, hasta obtener un grado menor a 10%, se debe evitar un sobrecalentamiento por lo que este proceso debe ser cuidadoso.
- **Humectación:** luego para lograr la compactación se debe humedecer levemente hasta encontrar el punto de pegado de fibras.
- **Prensado:** para lograr un *pellet* con superficie lisa y durable se debe prensar con una alta presión y temperatura, donde la compactación debe ser de alta

calidad sino el *pellet* podría romperse y generar alto porcentaje de cenizas en la combustión.

- **Tamizado:** Mediante el uso de cintas tamizadoras se clasifican y seleccionan los *pellets*, sacando los defectuosos y los residuos que podrían generar complicaciones en la combustión (humo y cenizas).
- **Embolsado:** Luego de tener los *pellets* listos se deben embolsar en los diferentes tipos de formatos, ya sea sacos, bolsas, granel entre otros. Siempre cuidando no exponerlo a la humedad.
- **Almacenamiento:** se debe disponer de bodegas para este proceso, dicho lugar debe ser seco.

Otro punto importante para considerar es la calidad del *pellet*, el conjunto de normas europeas que establece el comité técnico CEN/TC 335 se encarga de establecer una estandarización para la producción en serie (Velázquez, 2018). Para lograr un producto de buena calidad se deben tener algunas consideraciones que Pinilla et al., (2020) describe como las siguientes:

- **Brillo:** apariencia lisa y brillante, semejante a un barnizado.
- **Color:** debe ser de color uniforme, si bien depende de la materia prima si es más oscuro o claro debe mantener la tendencia de los demás productos.
- **Estructura:** debe ser compacto y no tener grietas
- **Densidad:** su densidad debe ser mayor a la del agua, al sumergirlo debe hundirse.
- **Tiempo de inmersión:** si pasado 5 minutos aún está hundido y compacto es posible que tenga aglutinantes que podrían provocar humo en exceso.

Actualmente usar *pellets* es común en los países europeos y cada vez más en el resto del mundo, si bien la mayor parte de estos *pellets* provienen de la madera, se espera que se puedan estudiar y desarrollar más *pellets* provenientes de otros sectores (Pradhan et al., 2018).

2.2 MODELOS Y HERRAMIENTAS DE ANÁLISIS PREDICTIVOS.

En la actualidad el uso de modelos predictivos se ha vuelto parte de las estrategias empresariales (empresas grandes, medianas y pequeñas). Recolectar datos que te permitan generar un modelo predictivo que logre detectar patrones de comportamiento es fundamental para el éxito de estos modelos (Espino, 2017). Poder contar con el stock necesario para satisfacer la demanda es parte fundamental de una buena administración, así como la dinámica para adaptarse al mercado cambiante (Cedeño y Carpio, 2022). Para Espino (2017) Identificar variables pasadas, relacionarlas para poder predecir eventos futuros es uno de los fundamentos principales del análisis predictivo, siendo indispensable contar con una buena cantidad de datos si se quiere una predicción más precisa. El uso correcto de los datos históricos ayuda a las empresas a poder optimizar los procesos y crear estrategias que se adapte a las necesidades que se van generando con el tiempo (Chávez y Saucedo, 2016).

2.2.1 Tipos de modelos predictivos

El modelo predictivo es aquel que busca predecir el comportamiento, utilizando características individuales como entrada y obtiene una calificación predictiva en la salida; existen algunos tipos de modelos aplicables dentro de un análisis predictivo, Espino (2017) menciona:

- **Modelo predictivo:** “modelo de relación entre el rendimiento específico de unidad en una muestra y uno o más atributos o características conocidos de la unidad” (p.13). Se debe trabajar con un conjunto de entrenamiento, los cuales son datos con atributo y comportamiento conocido dentro de una muestra y un grupo de prueba que es una muestra con atributos parecidos, pero con comportamiento desconocido.
- **Modelos descriptivos:** utilizados para clasificar clientes o grupos de estos, generalmente buscan identificar relaciones entre los clientes y los productos sin buscar la probabilidad de que actúe de cierta manera. Algunos modelos

de este tipo son los modelos de simulación, teoría de colas y técnicas de previsión.

- **Modelos de decisión:** Estos modelos describen la relación que existe entre los datos conocidos, el pronóstico y resultado de una decisión donde se involucran muchas variables. Se utiliza para modelar decisiones que se toman solo una vez o que se repetirá en más ocasiones. Ejemplos son: arboles de decisión, análisis de Pareto, análisis Swot o matriz de decisiones.
- **Modelos de conjunto:** también conocido como modelos de ensamble, los que consisten en la combinación de dos o más modelos para obtener resultados sintetizados y más precisos. Un ejemplo es el modelado de bosque aleatorio donde se combinan árboles de decisión para analizar diferentes datos de una muestra obteniendo un promedio de los resultados.
- **Modelo de elevación:** este modelo busca predecir la influencia que se puede alcanzar en el comportamiento de una persona. En resumen, este modelo predice el incremento a la respuesta de una acción de marketing, es decir identificar los clientes influenciables y centrar los esfuerzos en llegar a ellos.

Uno de los aspectos más importantes para tener en cuenta cuando se lleva a cabo alguno de estos modelos es poder validarlo, debido a que se debe comprobar su funcionalidad de manera correcta (Espino, 2017). Además, para realizar un modelo predictivo se debe seguir una metodología buscando alcanzar acciones proactivas y correctivas que mejoren rendimiento y mitiguen riesgos (Márquez Resinas y Ruiz, 2018). Existen metodologías genéricas para realizar modelos de análisis predictivo; CRISP-DM (*Cross industry Standard process for data mining*) es una metodología genérica que propone IBM para realizar modelos de análisis predictivos (Chávez et al., 2016). Si bien la mayoría de las personas utiliza sus propias metodologías (Eckerson, 2007) concluyó que, para realizar análisis predictivo, muchos analistas se basan en CRISP-DM por el enfoque de sentido común que cuenta, incluyendo generalmente 6 pasos principales:

1. Definición del proyecto: Aquí se definen los objetivos y resultados que se esperan obtener mediante el análisis predictivo.
2. Exploración: Se deben observar los datos para determinar cuál es el modelo más adecuado para el proyecto.
3. Preparación de datos: se seleccionan, extraen y transforman los datos para la construcción del modelo.
4. Construcción del modelo: se construye, prueba y valida el modelo evaluando si se cumplen las métricas y objetivos del proyecto.
5. Implementación: Se aplican los resultados obtenidos en el modelado a los procesos y decisiones del negocio.
6. Administración de modelos: en este proceso se busca la mejora del modelo, rendimiento, precisión, estandarización de herramientas y reutilización de este.

2.2.2 Técnicas de modelado

Para poder llevar a cabo los análisis de datos, se debe recurrir a diferentes técnicas de modelado, manuales o de aprendizaje automático que permitan predecir los resultados futuros (Cedeño et al., 2022), de forma general, Espino (2017) considera que estas técnicas pueden dividirse en técnicas de regresión y de aprendizaje computacional.

3.2.1 Técnicas de regresión

Estas técnicas consisten en establecer una un modelo de ecuación matemática que permita representar las interacciones entre diferentes variables a estudiar. Dentro de las técnicas de regresión mencionadas por Espino (2017) podemos mencionar:

3.2.2 Modelo de regresión lineal: El modelo de regresión lineal busca analizar la relación que existe entre una variable dependiente y una independiente, teniendo como objetivo seleccionar los parámetros que minimicen la suma de los errores al cuadrado (Espino 2017). El modelo de regresión lineal puede ser simple o múltiple dependiendo de si es una variable independiente o un conjunto, se usa comúnmente

cuando la variable dependiente es una variable continua y si la variable es dicotómica (2 valores) ya se debe utilizar un modelo de regresión logística (Peláez, 2016)

3.2.3 Modelo de árboles de clasificación y regresión: básicamente son una técnica no paramétrica de árboles de decisión conocida como CART, donde si la variable es categórica produce arboles de clasificación y si es numérica arboles de decisión.

3.2.3 Curvas de regresión adaptativa multivariable: técnica no paramétrica que ajusta regresiones lineales para construir modelos flexibles, el nudo es de gran importancia dentro de este modelo debido a que es donde se intersecan las curvas para que un modelo de regresión local de paso a otro.

Otros modelos de regresión a considerar son análisis de supervivencia, regresión logística, series temporales y elección discreta (Espino, 2017).

3.3 Modelos de aprendizaje computacional (*Machine learning*)

Son modelos que tienen la capacidad de aprender en base a los datos, encontrar patrones y regularidades en los datos permitiendo que pueda lograr un desempeño óptimo al observar datos que no se han conocido previamente (González 2015). Algunos ejemplos que Espino (2017) menciona de estos modelos son los siguientes:

3.3.1 Redes neuronales: son utilizadas cuando no hay un conocimiento exacto de la naturaleza de los valores de entrada y salida, donde estas redes deben aprender la relación que puede existir entre estos valores, son capaces de modelar funciones complejas existiendo tres tipos de aprendizajes, por esfuerzo, supervisado y no supervisado. Las redes neuronales además procesan información en tiempo real, en paralelo, pudiendo reconocer y clasificar patrones en sistemas complejos, se utilizan mucho para predecir datos económicos y financieros (Cedeño et al., 2022).

3.3.2 Máquinas de vectores de soporte: máquinas de aprendizaje utilizadas para clasificaciones binarias y estimaciones de regresión. Detectan y explotan patrones complejos de datos ordenando y clasificando datos. Existen las máquinas de vectores lineales, polinomiales, sigmoide entre otros.

3.3.3 Clasificador bayesiano ingenuo: se utiliza cuando existe una cantidad de predicciones elevada y está basado en la regla de probabilidad condicional de bayes utilizado para la clasificación de los datos.

3.4.4 K-vecinos más cercanos: Modelo de reconocimiento de patrones que utiliza valores positivos y negativos, observando el punto que se desea predecir genera una nueva muestra basándose en las distancias del vecino más cercano dentro del conjunto de datos.

Por otro lado, Cedeño et al., (2022) también menciona un grupo de modelos predictivos que sirven básicamente para ayudar a predecir inventario y mantener la oferta de productos; dentro de los que podemos destacar:

- **Modelo Wilson:** cuando existe una demanda de un producto constante uniforme y conocida.
- **Modelo Holt Winter:** necesita al menos 2 periodos anteriores para poder pronosticar, acepta variables de tiempo y es adaptable a nueva información.
- **ETS:** Utiliza suavizado para la tendencia y estacionalidad, donde ambos componentes deben estar presente.
- **BATS:** solo modela periodos enteros de duración, es un método de suavizado exponencial que no es tan efectivo en naturaleza compleja.

En resumen, existen variadas técnicas para análisis predictivo, siendo algunas más exclusivas y efectivas que otras, donde todas tienen fortalezas y debilidades. Pero para (Espino, 2017) no solo la técnica utilizada importa sino también la herramienta (software) elegida para el análisis, existiendo algunas más populares y de uso gratuito. Herramientas como SAS y Tableau pueden tener una curva de aprendizaje elevada, además de un costo alto, por lo que optar por herramientas como MS Excel, R, Weka de uso libre es una gran alternativa para realizar modelos predictivos (Aqlan, Nwokeji y Shamsan, 2020). MS Excel es una muy buena herramienta para realizar modelos basados en series de tiempos (Martínez, 2017), donde se puede decir que “una serie de tiempo es un conjunto de observaciones de una variable medida en puntos sucesivos en el tiempo a lo largo de periodos sucesivos” (Villarreal, 2016, p.5). Para realizar modelos predictivos de estimación de demanda (lo que se busca en este proyecto),

(Martínez, 2017) describe 5 métodos de series de tiempo para utilizar en la herramienta MS Excel, r u otros;

- **Promedio Móvil Simple:** Método de suavizado que utiliza información histórica para predecir el futuro de la variable. En este método se promedian k valores con los datos más recientes para utilizar dicho promedio como el pronóstico del periodo siguiente (Villarreal, 2016).
- **Promedio Móvil Ponderado:** En este método se asigna un valor mayor a los datos más recientes obteniendo un promedio basado en asignar un porcentaje de importancia mayor a los datos más nuevos. Se cree que lo reciente es más cercano a la tendencia futura (Villarreal, 2016).
- **Suavización exponencial simple:** Se utiliza cuando los patrones de demanda son aleatorios, se usa cuando se necesita eliminar los impactos generados por elementos irregulares históricos. Se usa un coeficiente de suavización (Martínez, 2017).
- **Método de Holt:** Este método considera la tendencia o estacionalidad, no solo los valores históricos de los datos, mitigando los cambios que ocurren ocasionalmente en el tiempo.
- **Método Winters:** Método que considera tendencia, nivel y estacionalidad de las series de tiempo a trabajar, por lo que se adapta fácilmente al surgimiento de nuevos datos.

Además de MS Excel que es una herramienta de las más conocidas, existen muchas otras que sirven para llevar a cabo análisis de datos y modelos de estimación, algunas se pueden utilizar de manera libre y gratuita mientras que otras necesitan de licencia para su uso (Espino, 2017). Basado en una comparativa de costo, curva de aprendizaje, dificultad de uso y dificultad en el proceso de instalación podemos realizar una tabla comparativa de las herramientas más utilizadas para el aprendizaje de Análisis de datos (Aqlan et al., 2020).

En la siguiente tabla se resume una comparativa de los softwares de análisis de datos.

Tabla 2 Comparación de herramientas para análisis de datos (Fuente: Aqlan et al., 2020)

| SOFTWARE/FACTORES | Costo | Curva de aprendizaje | Dificultad de uso | Dif. proceso de instalación |
|-------------------|--------|----------------------|-------------------|-----------------------------|
| R | Gratis | Media | Media | Moderado |
| MS Excel | Bajo | Baja | Baja | Fácil |
| SAS | Alto | Muy alta | Alta | Moderado |
| Tableau | Alto | Muy alta | Alta | Moderado |
| JASP | Gratis | Alta | Alta | Fácil |
| Python | Gratis | Media | Media | Moderado |
| JAMOVI | Gratis | Alta | Alta | Fácil |
| NLTK | Gratis | Alta | Alta | Moderado |

El análisis de datos ya no es solo de uso exclusivo de grandes corporaciones, con las nuevas tecnologías cualquier persona tiene acceso a herramientas gratuitas que le permitan hacer análisis para su negocio, solo se necesitan datos para trabajar de manera más efectiva (Espino, 2017). En conclusión, los datos son de gran importancia para la estrategia de la empresa, si se utilizan sabiamente pueden entregar una gran ventaja competitiva, el valor que tienen es tan alto que algunos los consideran el petróleo del siglo 21, por lo que resguardarlos es de vital importancia. Poder realizar modelos para estimar y predecir comportamientos futuros basados en datos puede ayudar a los gerentes a tomar mejores decisiones sobre el rumbo de una empresa, motivo por el que los modelos de análisis tienen que ser lo más preciso posible.

3 CAPÍTULO: METODOLOGÍA

El desarrollo del proyecto se lleva a cabo mediante la adaptación y utilización de la metodología CRISP-DM (Cross - Industry Standard Process for Data Mining), que es muy utilizada en minería de datos, si bien este proyecto no aborda específicamente ese tipo de análisis, esta metodología sirve para seguir el mismo proceso en la obtención de resultados específicos como estimar la demanda de un producto para una empresa. Esta metodología tiene la característica de ser flexible y mostrar el ciclo vital de un proyecto de análisis de datos, donde siguiendo los 6 pasos principales de la metodología, se puede llegar a obtener los resultados de un proyecto. Los pasos por seguir en la metodología CRISP-DM se describen a continuación:

1.- Comprender el negocio (situación actual de la empresa)

La primera etapa de esta metodología busca ayudar a comprender el contexto y actualidad de la empresa en la que se lleva a cabo el proyecto. Se incluyen etapas como:

- Determinar objetivos del negocio.
- Valorar la solución entregada.
- Determinar los objetivos del análisis de datos.
- Realizar un plan para el proyecto.

2.- Comprensión de los datos:

En esta etapa se tiene un primer acercamiento con los datos, donde se interactúa y familiariza con estos. Cuenta con las siguientes etapas:

- Recolectar datos.
- Describir datos.
- Explorar datos.
- Verificar calidad de los datos.

3.- Preparación de los datos:

En esta etapa se preparan los datos de tal forma que queden listos para ser procesados por el software que se utilizará en la etapa de modelado. Se pueden encontrar las siguientes actividades en esta etapa:

- Seleccionar datos.
- Limpiar datos.
- Estructurar datos.
- Formatear datos.

4.- Modelado:

Esta etapa es la más importante y fundamental para el desarrollo del proyecto, acá se deben elegir y probar los modelos que se van a implementar para lograr obtener los diferentes resultados que mejor se acomoden a la empresa. Se realizan modelos más simples y otros complejos que permitan comparar las diferencias que se obtienen en cada uno. Además, en algunos modelos se consideran factores como estacionalidad y tendencia. En esta etapa se pueden identificar las siguientes actividades:

- Seleccionar la técnica de modelado.
- Generar plan para probar los modelos.
- Construir modelos de estimación.

5.- Evaluar:

En esta etapa es donde se realiza una evaluación de los modelos construidos, es decir, si estos cumplen con los requerimientos del negocio. Se identifica el modelo que mejor se adapta y si los resultados varían lo suficiente entre modelos. Se consideran actividades como:

- Evaluar resultados.
- Revisar proceso.

- Determinar pasos a seguir.

6.- Implementación:

En este punto se define cómo se va a entregar el modelo a la empresa y de qué forma se debe implementar. Es una etapa de finalización del proyecto que permite hacer una interacción final con la empresa en la que se llevó a cabo el modelado de datos. Las siguientes actividades forman parte de esta etapa:

- Plan de proceso para implementación.
- Plan para monitorear y mantener.
- Revisar el proyecto.

La figura 2 que se muestra a continuación representa gráficamente el proceso de la metodología y las relaciones existentes entre cada uno de los pasos que se deben llevar a cabo para el cumplimiento del proyecto.

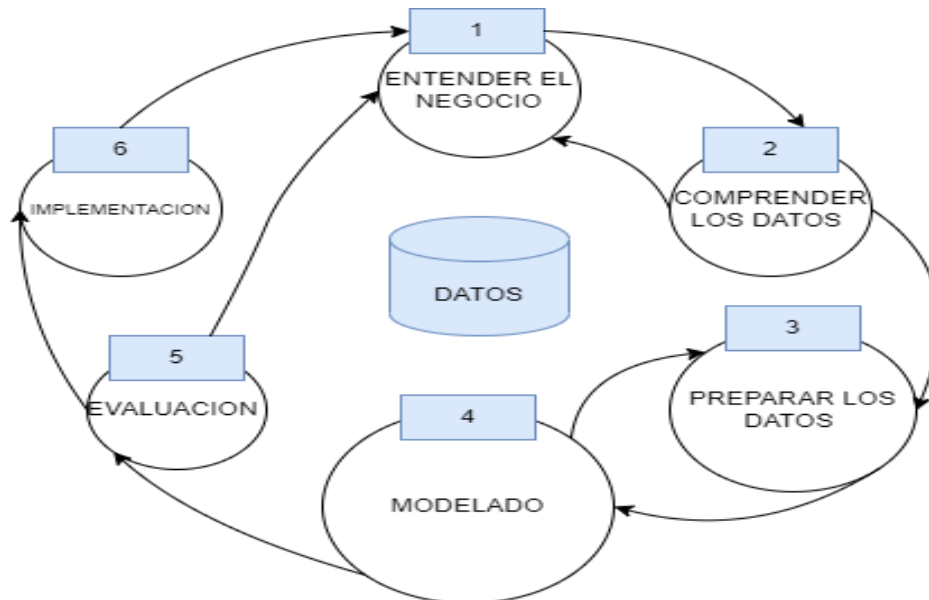


Figura 2: Diagrama de modelo CRISP DM (elaboración propia)

4 CAPÍTULO: DESARROLLO DE METODOLOGÍA Y ANÁLISIS DE RESULTADOS

4.1 Descripción de la empresa

Forestal Rio Claro (FRC) es una empresa chilena dedicada principalmente a fabricar, proveer y producir madera impregnada y *pellets* para el uso industrial y domiciliario. Su casa matriz se encuentra ubicada en Lote 2 - división predio Flor María, Maule, Talca km 266 Ruta 5 sur. También cuenta con otras sucursales en las comunas de Colina y Peñaflo. Esta empresa cuenta con más de 25 años de experiencia en el sector trabajando bajo un estándar de calidad internacional como EUROFINS y la certificación chilena Nch 819 OF.2012.

La visión de la empresa es ser reconocida como líder en el mercado nacional, latinoamericano y referente internacional de madera impregnada a través de procesos de excelencia e innovando en función de las expectativas de las personas que confían en el trabajo realizado por la empresa.

La misión es garantizar un servicio adaptado a las necesidades que demandan los clientes, colaborando con sus requerimientos y solicitudes para garantizar el éxito de los proyectos. Para ello se genera valor a través de la responsabilidad, la excelencia y dedicación en todos los aspectos del día a día.

Para cumplir con estos compromisos, la empresa cuenta con una línea de productos que se orientan al sector agrícola, eléctrico, construcción y hogar. Se encarga de proveer *pellets* a lo largo de todo Chile, tanto para venta a los domicilios como para las empresas que lo requieran para uso en la industria. Trabaja con países de todo el mundo, principalmente de Sudamérica y África a los que les provee materiales de madera impregnada para su utilización. Cuenta principalmente con 5 pilares de procesos a seguir como organización, los que les ayudan a mantener un alto estándar de calidad:

Stock permanente: se lleva a cabo con el objetivo de entregar una solución rápida y oportuna a los clientes, acompañándolos y resolviendo dudas. La empresa debe mantener constante stock de madera y *pellets*.

Calidad: Se busca constantemente garantizar la entrega de productos confiables, por lo que el uso de las certificaciones de calidad adquiridas (EUROFINS, Nch) son una constante exigencia en el proceso productivo.

Entrega oportuna: se asegura la entrega a tiempo de los productos requeridos por los clientes, por lo que se cuenta con un proceso gestionado de entregas.

Capacidad logística: se cuenta con una capacidad logística que permite llegar a cualquier parte de Chile y otros países que lo requieran.

Flexibilidad: se entienden los requerimientos de cada cliente, por lo que se adaptan a las necesidades de cada uno en caso de ser requerido.

Si bien, Forestal Rio claro tiene una amplia línea de productos a la venta, el producto específico que se aborda en este proyecto es el saco de *pellet* de 18 kg y el *pellet* a granel, los cuales están hechos 100% de pino radiata; una especie muy común de madera en la región. Para este producto se cuenta con una amplia red de distribución que alcanza varias regiones de Chile, encontrándose en diferentes puntos de venta en cada una de ellas, lo que le permite tener un mayor alcance al público.

4.2 Comprensión del negocio

4.2.1 Objetivos del negocio

En busca de la constante mejora y enfoque en el cliente, Forestal Rio Claro se encuentra en búsqueda de un modelo que le permita estimar la demanda que tiene de sus productos. En este proyecto se aborda la específicamente la demanda del *pellet*, un producto que va de la mano con la conciencia ambiental que tiene la empresa, buscando reducir las emisiones de CO2 a la atmósfera. La empresa cuenta con

recopilación de datos históricos de ventas que serán analizados y permitirán realizar modelos de estimación de demanda futura, considerando las variaciones históricas que se puedan apreciar en los datos. Los objetivos del negocio que se pueden encontrar en la empresa van ligados directamente con el cumplimiento de uno de los cinco pilares de proceso mencionados anteriormente. En este caso, mediante el uso de modelos predictivos, se busca mantener el stock permanente para así cumplir con el objetivo de entregar una solución rápida y oportuna a sus clientes. Se deben probar una serie de modelos predictivos donde en algunos se consideran tendencia y estacionalidad, esto para observar si difieren completamente de los que no lo consideran, junto con establecer criterios para la aceptación del modelo que mejor se adapte a la realidad que tiene la empresa actualmente en cuanto a la demanda del producto.

4.2.2 Valoración de la solución

Es de gran importancia para la empresa utilizar todos los recursos disponibles para obtener ventajas que le permitan posicionar sus productos por sobre la competencia. Es por esto que, utilizar un recurso como los datos históricos con los que cuenta la empresa, permite facilitar el proceso del área productiva, siendo de gran importancia para esta producir la cantidad necesaria de *pellets*. Se puede generar un gran impacto en la toma de decisiones de la empresa utilizando los datos, que al ser llevados a un software de análisis que permita crear modelos predictivos, facilitaría la tarea de los encargados del área de producción.

Económicamente, este proyecto tendrá un impacto positivo para la empresa, por lo que la propuesta de implementar un modelo predictivo para estimar la demanda de una manera precisa es de vital importancia para generar mayores beneficios. Es por esto por lo que, la solución entregada en este proyecto tiene gran valor en cuanto al apoyo a las decisiones que se van a tomar en el futuro dentro del área productiva.

4.2.3 Objetivos del análisis de datos.

En la propuesta del proyecto se plantean objetivos que se deben alcanzar mediante el análisis de datos, para esto se deben realizar una serie de actividades que permitan llegar a dar cumplimiento a cada uno de ellos. Las actividades que permitirán alcanzar dichos objetivos del proyecto se mencionan a continuación:

- **Definir el ecosistema donde se van a procesar los datos;** en el ecosistema o software elegido para procesar datos se llevarán a cabo los modelos y procesamiento de los datos que finalmente darán el cumplimiento de los objetivos que se plantearon.
- **Definir técnicas de modelado a utilizar;** se definirán una serie de tipos de modelos que se van a llevar a cabo para así dar con la mejor solución o la que se adapte mejor a las necesidades de la empresa.
- **Identificar fórmulas;** estas son las que van a ayudar a realizar cada uno de los modelos y que permitirán entregar la mejor precisión a cada uno de ellos.
- **Identificar las variables;** establecer cuáles son las que afectan en la demanda del producto y cómo estas variables podrían hacer variar el modelo si se consideran o no dentro del análisis de los datos, por ejemplo; cómo la estacionalidad puede o no afectar para que el modelo de estimación de demanda se vea diferente a otro modelo que no la considera.

4.2.4 Realizar secuencia de proyecto

Se realiza un plan o secuencia de proyecto que contemple las tareas que se abordarán en la etapa de modelado. Mediante este plan de trabajo se puede apreciar un resumen genérico de los pasos que se van a seguir en la construcción de los modelos. En la Tabla 3 se genera una secuencia de las etapas que se abordarán en el proceso de modelado.

Tabla 3 Secuencia de proyecto genérico para la creación del modelo final

| | |
|------------------------------------|--|
| Recopilado de datos | <ul style="list-style-type: none"> • Se seleccionan los datos a utilizar. • Se cargan en una tabla dentro de una base de datos. |
| Elección de herramienta a utilizar | <ul style="list-style-type: none"> • Se elige(n) la(s) herramienta(s) a utilizar para el procesado de los datos. |
| Elección de modelos | <ul style="list-style-type: none"> • Se seleccionan los modelos de estimación de demanda con los que se va a trabajar. |
| Creación de modelos | <ul style="list-style-type: none"> • Se llevan a cabo los modelos de predicción de demanda, realizando un reporte de cada uno, así como las fórmulas asociadas a estos. |
| Generación de resultados finales | <ul style="list-style-type: none"> • Se elige el resultado final que será el modelo para la utilización en la empresa. |

4.3 Comprensión de los datos

Se recolectan los datos a utilizar para la construcción de los modelos. Estos datos son proveídos directamente por la empresa, obtenidos de la base de datos del software que utilizan para llevar a cabo sus ventas.

4.3.1 Recolección datos iniciales

Los datos que han recolectado de la empresa para este proyecto se encuentran en diferentes tablas dinámicas del software MS Excel, cada una de las cuales cuenta con información con respecto a las ventas que ha realizado la empresa a partir del año 2021 hasta junio de 2022. Es decir, información de un año y medio de antigüedad para

realizar los diferentes modelos de estimación. Para mejorar la comprensión del contenido se procedió a cambiar el nombre de las hojas del MS Excel correspondiente. Esto para que existiera una mayor relación con respecto a los datos contenidos en las tablas.

A continuación, en la Tabla 4 se muestra la modificación de los nombres iniciales:

Tabla 4 Títulos de datos entregados por la empresa

| Nombre hoja de tabla original | Nombre modificado para comprensión |
|--------------------------------------|---|
| Tablas Marzo unidades | Resumen ventas toneladas <i>pellets</i> |
| \$\$\$ | Ventas netas junio 2021 |
| Tablas Marzo | Resumen cantidad general |
| Tabla verificación | Cantidad ventas maderas. |
| Vtas marzo depurado | Resumen detallado |
| Exportación m3 | Metros cúbicos exportados |
| Listado | Listado |
| Maestra código m3 | Código ventas m3 |
| Maestra listada (2) | Productos a la venta |

Ya con la revisión de las hojas de tablas dinámicas y su renombramiento, se puede trabajar de manera más sencilla con los datos que posee cada una. Se dejaron algunas hojas con el mismo nombre debido a que se considera que no interfieren con los datos que se necesitan para el proyecto.

4.3.2 Descripción de los datos

Se procede a realizar una descripción de los datos que se encuentran contenidos en las tablas dinámicas mencionadas anteriormente en la Tabla 4. Se destaca el contenido de datos de la Tabla 9 **resumen detallado** descrita en el listado como la más relevante para llevar a cabo el proyecto.

Resumen ventas toneladas *pellets*

Tabla 5 Contenido resumen ventas de *pellets*

| ATRIBUTO | TIPO | DESCRIPCIÓN |
|---------------|----------|--|
| Mes | Numérico | Realiza un filtro de las ventas de toneladas de <i>pellets</i> mensuales que se han realizado. |
| Año | Numérico | Realiza un filtro anual de las toneladas de <i>pellets</i> vendidos. |
| Pellet | Numérico | Permite ver el número de toneladas de <i>pellets</i> vendidos. |

Ventas netas junio 2021 (\$)

Tabla 6 Tabla de ventas netas

| ATRIBUTO | TIPO | DESCRIPCIÓN |
|---------------------------|------------|---|
| MERCADO | Caracteres | Realiza una diferencia de las ventas en los diferentes mercados tanto nacional como de exportación. |
| Nombre del cliente | caracteres | Se realiza un filtro de todas las ventas realizadas a cada uno de los clientes con los que cuenta la empresa. |
| GE | Caracteres | Permite elegir el resumen de ventas de diferentes categorías de productos a elegir. |

Resumen cantidad general

Tabla 7 Resumen de la cantidad de ventas en general

| ATRIBUTO | TIPO | DESCRIPCIÓN |
|-----------------|-------------|---|
| MERCADO | Caracteres | Permite filtrar entre mercado nacional y de exportación. |
| GE | Caracteres | Permite elegir el o las categorías de productos en que se desean ver las cantidades vendidas. |

Cantidad ventas maderas

Tabla 8 Cantidad productos venta nacional

| ATRIBUTO | TIPO | DESCRIPCIÓN |
|---------------------------|-------------|--|
| Mercado | caracteres | Permite elegir solo del mercado nacional. |
| Nombre del cliente | caracteres | Permite ver toda la lista de clientes y las cantidades que ha comprado cada uno. |
| GE | caracteres | Permite elegir las categorías de los productos. |
| Desc. Producto | caracteres | Permite ver todos los tipos de productos de cada categoría. |

Resumen detallado

Tabla 9 Detalle de las ventas realizadas

| ATRIBUTO | TIPO | DESCRIPCIÓN |
|-----------------|-------------|---|
| Id | Numérico | Permite ver el identificador de la venta. |
| Día | Numérico | Permite elegir el día |

| | | |
|---------------------------|------------|--|
| | | específico de las ventas. |
| Mes | Numérico | Permite ver las ventas realizadas por mes. |
| Libro vtas | Numérico | Permite ver el libro de ventas mensuales. |
| Cantidad | Numérico | Permite filtrar por cantidad de productos vendidos |
| Desc.producto | Carácter | Se puede ver el detalle del producto que se vendió, su nombre y peso. |
| Precio unitario | Numérico | Se puede ver el precio de venta por unidad de producto vendido. |
| Total neto línea | Numérico | Se puede ver el total neto de cada compra realizada. |
| Rut del cliente | Numérico | Permite ver los Rut de los clientes o empresas. |
| Nombre del cliente | Numérico | Permite ver todos los nombres de los clientes que han comprado. |
| Cod.vendedor | caracteres | Se puede filtrar por las ventas que han realizado los diferentes vendedores. |
| Tipo de doc | caracteres | Permite ver los tipos de comprobantes que permiten realizar las ventas |
| Nombre doc | Caracteres | Permite filtrar por nombre del comprobante de venta, boleta, factura etc. |
| Cod.producto | Numérico | Se puede ver el código del producto que se ha vendido. |
| GE2 | Carácter | Permite filtrar por nombre del producto. |

Si bien existen otros atributos para esta tabla, no se considera que contengan la información necesaria ni el impacto suficiente para ser considerada dentro de la descripción, motivo por el que se decidió no mencionar dichos atributos.

La Tabla 9 resumen detallado mostrada anteriormente, es la más completa para extraer la información necesaria para realizar el proyecto, debido a que permite filtrar la

información por diferentes atributos que pueden ser de utilidad a la hora de realizar el modelo de análisis predictivo, motivo por el que se decidió trabajar directamente con la información que esta proporciona y si es necesario considerar alguna información proveniente de alguna de las otras tablas.

Las últimas 4 tablas de información mencionadas al principio, no serán resumidas con sus atributos porque contienen información que menciona la otra línea de productos de la empresa o contienen información incompleta.

4.3.3 Exploración de los datos

Para comprender el contenido de los datos se realiza una visualización gráfica de la Tabla 5 “resumen ventas toneladas de *pellets*”

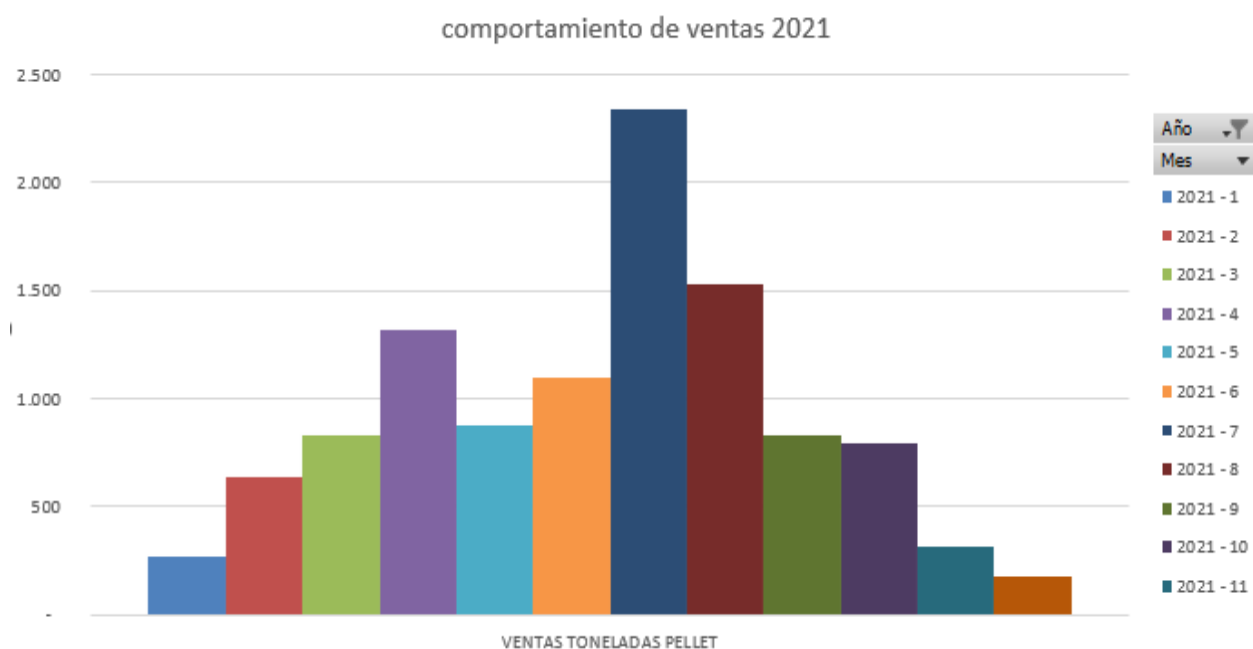


Figura 3: Grafico de comportamiento ventas 2021

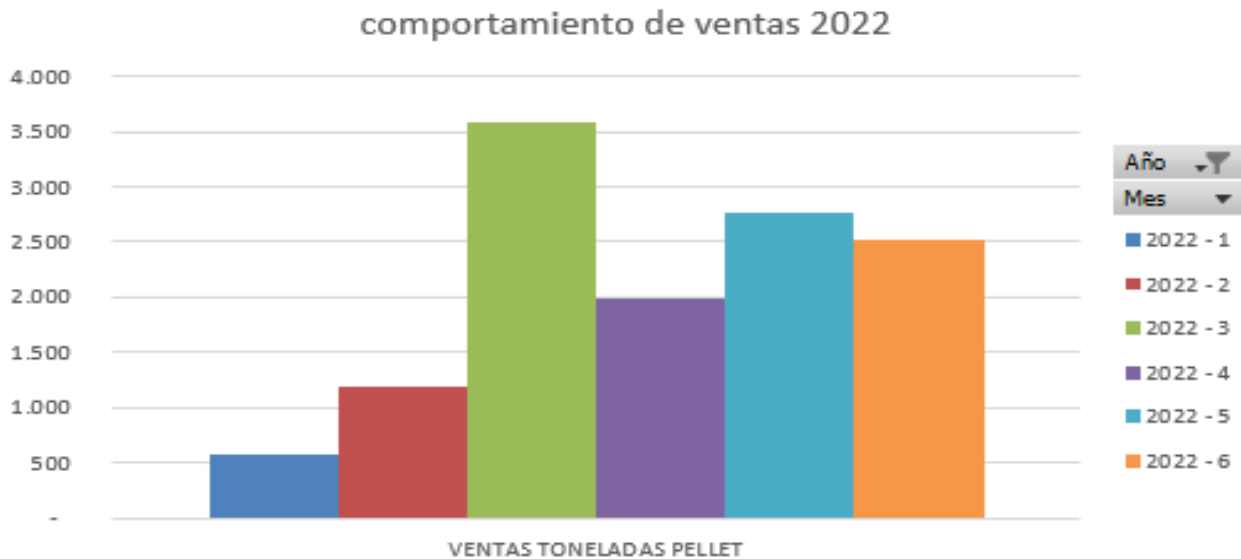


Figura 4: Grafico comportamiento de ventas 2022

Mediante este gráfico resumido del año 2021 ya se pueden apreciar algún patrón de la demanda como lo es la influencia de la estacionalidad en ella, es decir que dentro de los meses de mayor temperatura dicha demanda generalmente es baja, mientras que en los meses de mayor frío la demanda crece por el uso de las estufas. También se puede apreciar una anomalía en el aumento repentino de las ventas en el mes de abril.

Esta tabla contiene la mitad del año, pero al igual que la del año anterior tiene un comportamiento similar, se puede apreciar una baja demanda en los primeros meses, sin embargo, también tiene una anomalía en la demanda del mes de marzo específicamente.

4.3.4 Verificar la calidad de los datos

Mediante la revisión realizada de los datos se pueden establecer algunas conclusiones con respecto a la calidad estos; los datos obtenidos son de un periodo

muy cercano por lo que con la realización de varios modelos diferentes se espera poder obtener resultados cercanos a la futura demanda de la empresa. En la Tabla 9 elegida (Resumen detallado) los datos son claros, pero existen muchos datos de productos que no son relevantes para este proyecto; como aquellos que contienen valores nulos, otros con valores negativos e incompletos. Los datos para realizar los modelos predictivos para demanda domiciliar se pueden considerar suficientes para el proyecto, considerando que están todos aquellos registros dentro del tiempo de venta del producto. Por otro lado, los datos para el uso industrial (*pellet* a granel) son de menor calidad ya que son menos y solo se registran ventas a una empresa en meses diferidos.

4.4 Preparación de los datos

Se procede a realizar la preparación de los datos en formato MS Excel para ser analizados por el software R Studio. Con esto se generan las tablas con la base de datos que se necesitará para realizar los modelos de estimación de demanda.

4.4.1 Selección de los datos

Para efectos del proyecto se observó que había tablas en hojas de datos en MS Excel que no contenían los datos necesarios para el proyecto (datos relacionados con las ventas de *pellets*), por lo que se hizo una separación en una hoja MS Excel diferente de la tabla elegida a la que se le debe realizar la limpieza. Esta tabla es resumen detallado (tabla 9).

De la tabla resumen detallado se eliminan las columnas que no serán consideradas en ningún análisis para el proyecto:

Las columnas eliminadas fueron;

Tabla 10 Columnas eliminadas de la base de datos

| COLUMNA | Motivo |
|-----------------------------|--|
| ID | Solo contiene el código de la venta realizada. |
| Libro Vtas. | Contiene la fecha y mes en un formato diferente. |
| Rut del cliente | El Rut del comprador no se considera. |
| Cod.vendedor | El registro del vendedor no será considerado. |
| Tipo Doc | Contiene el nombre resumido del tipo de recibo que se entregó en la venta, pero se utilizará el nombre completo del tipo de documento. |
| Número del documento | No se considera influyente. |
| Cod.producto | No es necesario. |
| Mercado | No existe venta a otro mercado de este producto. |
| Export | No se exporta este producto. |
| Cantidad2 | Esta columna está repetida. |

Estas columnas se eliminaron de la tabla porque no se consideran para el proyecto, en cambio se dejaron para el análisis las siguientes columnas:

Tabla 11 Columnas que se dejaron en la base de datos

| | |
|-------------------------|---|
| Día | El día en que se realizó la compra. |
| Mes | El mes que se realizó la compra. |
| Año | El año en que se realizó la compra. En este caso es 2021 o 2022. |
| Cantidad | La cantidad del producto comprado por el cliente. |
| Desc. Producto | Si se lleva pack múltiplo de 12 bolsas de <i>pellets</i> (18kg) o una cantidad elegida de bolsas. |
| Precio unitario | El precio de la unidad, en este caso la bolsa de <i>pellets</i> . |
| Total, neto | El total neto de la venta. |
| Nombre cliente | El cliente que realizó la compra. |
| Nombre documento | El tipo de documento con el que se hizo el comprobante, este puede ser boleta, factura, nota de crédito, nota débito etc. |

Estas columnas contienen información relevante para el modelado, si bien no todo será necesario en la predicción, se pueden extraer algunos análisis diferentes que permitan dar mayor consistencia al proyecto. En total ha quedado un registro de 3.550 datos por columna, con un total de 31.950 datos con los que se va a trabajar para la base de datos principal que se utilizará.

4.4.2 Limpieza de los datos

Ya que las columnas que no se consideran para el análisis fueron eliminadas de la base de datos en MS Excel, se procede a exportar a R Studio para realizar la limpieza. En la limpieza se buscan *outliers* (Valores atípicos), valores negativos y NA (datos perdidos) si se identifican.

Luego de haber exportado la base de datos a R con el nombre de *Detalle_de_ventas* se procedió a realizar una revisión general con la función **Summary** de R (Figura 5:), donde se obtuvieron algunos resultados preliminares:

```
> summary(Detalle_de_ventas)
  Día      Mes      Año      Cantidad      Desc. Producto
Min.   : 1.00   Min.   : 1.000   Min.   :2021   Min.   : -28000.0   Length:3550
1st Qu.: 9.00   1st Qu.: 4.000   1st Qu.:2021   1st Qu.:  1.0     Class :character
Median :17.00   Median : 6.000   Median :2022   Median :  10.0     Mode  :character
Mean   :16.35   Mean   : 5.624   Mean   :2022   Mean   :  460.5
3rd Qu.:24.00   3rd Qu.: 7.000   3rd Qu.:2022   3rd Qu.:  288.0
Max.   :31.00   Max.   :12.000   Max.   :2022   Max.   :140000.0

Precio Unitario  Total Neto Línea  Nombre del Cliente  Nombre Doc
Min.   : -268084   Min.   : -4946832   Length:3550         Length:3550
1st Qu.:  2687    1st Qu.:  40235    Class :character    Class :character
Median :  3437    Median :  84504    Mode  :character    Mode  :character
Mean   : 24873    Mean   :  973039
3rd Qu.: 40235    3rd Qu.:  820228
Max.   :268086    Max.   :23918400
```

Figura 5: Resultados de la función *summary*

- Se detectaron valores perdidos (NA) en la base de datos. Esto se comprobó utilizando la función ***is.na(Detalle_de_ventas)*** en R Studio, luego se aplica la función ***sum(is.na(Detalle_de_ventas))*** , la que arrojó como resultado 14 datos perdidos. Luego para eliminar los datos perdidos se utiliza la función ***na.omit(Detalle_de_ventas)***.

```
> sum(is.na(Detalle_de_ventas))  
[1] 14
```

- Se pudo apreciar que, en cantidad, precio unitario y total neto existen números negativos los cuales serán limpiados de la base de datos.

El total de números negativos y 0 encontrados fueron

| | |
|-----------------|------------------------|
| Cantidad | 88 negativos y 1 cero. |
| Precio unitario | 86 negativos y 1 cero. |
| Total neto | 92 negativos. |

Estos resultados se obtuvieron utilizando la función sum, donde se suma la cantidad de números que sean menores que 1.

sum(Detalle_de_ventas\$Cantidad<1)

sum(Detalle_de_ventas\$`Precio Unitario`<1)

sum(Detalle_de_ventas\$`Total Neto Linea`<1)

Los números negativos que serán eliminados de la base de datos están directamente relacionados por lo que se eliminarán las columnas que contienen los negativos y NA, con todas sus filas.

Para eliminar las columnas negativas se decidió exportar la base de datos sin NA a MS Excel nuevamente para realizar este proceso.

4.5 Estructuración y formateo de los datos

Se procede a realizar una revisión de la estructura y dar un formato que permita trabajar con los datos. Para esto se realiza un guardado de los cambios realizados en la base de datos utilizando la plataforma R Studio, creando una nueva base de datos, para guardar la nueva, se ha utilizado el siguiente comando en R que permite guardar los cambios que se han llevado a cabo.

El comando **setwd** permite elegir el directorio donde se va a realizar el guardado de los cambios de la base de datos; quedando estructurado de la siguiente manera:

```
setwd("C:/Users/... ")
```

Luego con el comando **write.table** se procede a guardar la nueva base de datos, donde el comando completo queda de la siguiente manera.

```
write.table(Detalle_de_ventas, file ="Detalle_ventas.csv", sep = ";", row.names = F)
```

La nueva base de datos queda con el nombre de "Detalle_ventas.csv". Para poder trabajar de manera cómoda con estos datos se procedió a importar la base de datos en un nuevo archivo de MS Excel dando la instrucción que proviene en formato CSV para poder pasarla a formato XLSX.

En el nuevo archivo "Detalle_ventas.xlsx" se pudo observar que la columna "cantidad" necesitaba ser reestructurada debido a que no era posible trabajar con los datos en el formato en que se encontraba. Esta columna es la que se considera la más importante para realizar los modelos, debido a que muestra las cantidades vendidas del producto, pero no estaba mostrando en algún formato de peso (kilogramos, toneladas, etc.). Las cantidades que indicaba esta columna hacía referencia a las bolsas de 18kg y los paquetes de bolsas de *pellets* que vende la empresa. Para mejor comprensión se adjuntan las imágenes.

| Cantidad | Desc. Producto |
|----------|--------------------------|
| 1680 | BOLSA PELLETS (18 KILOS) |
| 840 | BOLSA PELLETS (18 KILOS) |
| 840 | BOLSA PELLETS (18 KILOS) |
| 1680 | BOLSA PELLETS (18 KILOS) |
| 1680 | BOLSA PELLETS (18 KILOS) |
| 1680 | BOLSA PELLETS (18 KILOS) |
| 1680 | BOLSA PELLETS (18 KILOS) |

| Cantidad | Desc. Producto |
|----------|----------------------------------|
| 2 | PACK 12 BOLSAS PELLET (18 KILOS) |
| 2 | PACK 12 BOLSAS PELLET (18 KILOS) |
| 12 | PACK 12 BOLSAS PELLET (18 KILOS) |
| 2 | PACK 12 BOLSAS PELLET (18 KILOS) |
| 2 | PACK 12 BOLSAS PELLET (18 KILOS) |
| 2 | PACK 12 BOLSAS PELLET (18 KILOS) |
| 2 | PACK 78 BOLSAS PELLET (18 KILOS) |
| 2 | PACK 78 BOLSAS PELLET (18 KILOS) |
| 2 | PACK 60 BOLSAS PELLET (18 KILOS) |

En la base de datos había las siguientes descripciones de productos a la venta:

- BOLSA PELLETS (18 KILOS)

- PACK 12 BOLSAS PELLETS (18 KILOS)
- PACK 48 BOLSAS PELLETS (18 KILOS)
- PACK 60 BOLSAS PELLETS (18 KILOS)
- PACK 78 BOLSAS PELLETS (18 KILOS)
- PELLET A GRANEL

Como la columna cantidad hacía referencia a estos productos, se debió dar el formato en Kilogramos que corresponde para realizar el modelado, para esto se utilizó en MS Excel la siguiente función que permitió realizar la transformación.

```
=SI.CONJUNTO(
[@[Desc. Producto]]="BOLSA PELLETS (18 KILOS)";[@Cantidad]*18;
[@[Desc. Producto]]="PACK 12 BOLSAS PELLET (18 KILOS)";[@Cantidad]*12*18;
[@[Desc. Producto]]="PACK 24 BOLSAS PELLET (18 KILOS)";[@Cantidad]*24*18;
[@[Desc. Producto]]="PACK 36 BOLSAS PELLET (18 KILOS)";[@Cantidad]*36*18;
[@[Desc. Producto]]="PACK 48 BOLSAS PELLET (18 KILOS)";[@Cantidad]*48*18;
[@[Desc. Producto]]="PACK 60 BOLSAS PELLET (18 KILOS)";[@Cantidad]*60*18;
[@[Desc. Producto]]="PACK 78 BOLSAS PELLET (18 KILOS)";[@Cantidad]*78*18;
[@[Desc. Producto]]="PELLET A GRANEL";[@Cantidad]
)
```



Esta función permitió crear una nueva columna con los kilogramos vendidos de *pellets* durante el año 2021 y 2022. Por último, se hizo la separación de los *pellets* que son para uso industrial y los que son de uso domiciliario. La descripción del producto que corresponde a uso Industrial es la llamada “*PELLET A GRANEL*” por ende, se realiza un filtro en dicha descripción dejando solamente los datos que corresponden a *pellet* a granel. Luego se extrajeron esos datos y se exportaron a una nueva base de datos que se le llamó “Ventas industrial”.

En resumen, se obtienen 2 bases de datos diferentes, donde una guarda la información que corresponde a ventas a clientes que compraron *pellets* para uso domiciliario Y la otra base de datos corresponde a las ventas de producto a clientes que compraron para uso industrial.

Tabla 12 Resumen de base de datos

| Nombre BBDD | Columnas | Cantidad de datos por columna | Datos totales |
|---|----------|-------------------------------|---------------|
| Detalle_ventas (uso domiciliario) | 9 | 3445 | 31005 |
| Ventas_industrial (uso industrial) | 9 | 10 | 100 |

En el siguiente proceso corresponde llevar estas bases de datos al software R Studio para realizar los diferentes modelos elegidos.

| Data | | |
|-------------------|--------------------------|---|
| Detalle_ventas | 3445 obs. of 9 variables |  |
| ventas_industrial | 10 obs. of 10 variables |  |

Con estos nombres se trabajarán las bases de datos en el software de modelado elegido para este proyecto.

4.6 Modelado

Comprender los datos es crucial para poder establecer qué tipo de modelos se van a llevar a cabo en el análisis, además de establecer las métricas que permitirán medir la precisión del modelo y las series de tiempo a trabajar. Este apartado de modelado está basado en el libro *forecasting principles and practice de* Hyndman y Athanasopoulos (2018), donde se extraen interpretaciones y explicaciones de los modelos.

Para el análisis de este proyecto se consideraron las series de tiempo y los elementos básicos que las componen debido a que se considera que los datos se ven alterados por la existencia de estos componentes, aun así, se trabajaron algunos modelos de predicción que no tienen considerado los elementos de las series de tiempo para el producto, esto con la finalidad de evaluar las diferencias que conlleva la inclusión de estos elementos en el modelo.

Series de tiempo

Una serie de tiempo nos permite observar el comportamiento de una variable durante el transcurso del tiempo. Se utiliza dicha variable para poder predecir el comportamiento futuro.

Existen patrones de series de tiempo que se tendrán en consideración para el desarrollo de este proyecto.

Los patrones de serie de tiempo que se tendrán en consideración son;

- **Tendencia:**

Se considera que existe tendencia cuando a largo plazo los datos presentan aumento o disminución, dicha tendencia puede ser lineal o no lineal lo que sería más bien un cambio de orientación de la variable a estudiar, ya sea de creciente a decreciente o viceversa.

- **Estacionalidad**

Este componente se puede encontrar en una serie de tiempo cuando existen factores estacionales que la afectan, puede ser un día, semana, mes etc. La estacionalidad se caracteriza por tener una frecuencia conocida y fija. La determinación de la existencia de dicha estacionalidad se comprueba con analizar si existe autocorrelación en los valores. Dicha autocorrelación mide la relación lineal entre valores rezagados de la serie de tiempo, siendo el periodo con mayor autocorrelación la frecuencia de estacionalidad.

Otro patrón de series de tiempo que existe muy a menudo y que se debe tener en claro para no caer en confusiones es el patrón cíclico.

- **Cíclico**

En este patrón existe una subida y bajada de datos sin una frecuencia, por lo general se produce por el efecto de condiciones económicas. Dura alrededor de 2 años y se suele confundir con el patrón de comportamiento estacional en algunos casos. Para poder diferenciarlos se debe mirar las fluctuaciones, si la frecuencia no es fija se debe a que es cíclico, en cambio si la frecuencia es fija y se puede asociar con el calendario, el patrón es estacional. Además, hay que señalar que las magnitudes del patrón cíclico son más variables que el estacional.

En muchos casos las series de tiempo que incorporan estos patrones deben estar identificados en los datos para poder elegir de manera correcta el tipo de modelo que se va a implementar en el análisis, ya que este modelo debe ser capaz de capturar dichos patrones.

4.6.1 Selección de técnicas de modelado:

- **Promedio móvil simple:** Esta es una técnica sencilla para estimar demanda, es efectiva cuando las series de tiempo se comportan de manera estable. Si bien esta técnica puede no ser la más efectiva en este caso, ya que no considera tendencia, estacionalidad ni ciclo, será de utilidad para compararlo entre ellos y con los demás modelos, también ver si existe diferencia que se pueda establecer.
- **Suavizado exponencial simple:** Este método, al igual que el anterior tampoco considera los patrones de series de tiempo, aunque otorga mayor relevancia a los datos dependiendo su antigüedad. Se considera este modelo porque se ha visto una creciente evolución en las ventas de *pellets* en el último periodo. Por lo

que las predicciones pueden variar más al compararlas con el promedio móvil simple.

- **Método Holt - Winters (tendencia, estacionalidad):** Este método extiende el suavizado exponencial simple permitiendo estimar datos que tienen tendencia. Utiliza 3 ecuaciones que permiten llevar a cabo el modelo, una para predecir, otra para establecer un nivel de los datos y finalmente otra para tendencia. Para estacionalidad se considera lo mismo que lo anterior, pero cambia la ecuación de tendencia por una que considera estacionalidad, la que puede ser aditiva o multiplicativa.
- **Modelos ARIMA:** Estos modelos proporcionan un enfoque diferente a la hora de realizar pronósticos de series de tiempo ya que trabajan con un modelo autorregresivo y media móvil. Son unos de los más utilizados en los pronósticos de series temporales. Estos tienen como objetivo la descripción de la autocorrelación de los datos.

4.6.2 Generación de un plan de pruebas

Luego de construidos los modelos se debe validar y probar la calidad de estos. Existen métricas que permiten validar la calidad de los modelos. Basado en *Measuring forecast accuracy* (Hyndman, 2014) se consideran las siguientes métricas para este proyecto:

1. **Error absoluto medio (MAE):** Es una medida muy utilizada en las series de tiempo para medir el error de pronóstico. Con esta medida se puede ver qué tan certera es la predicción comparada con los valores reales.

$$MAE = \frac{1}{n} \sum_{i=1}^n |f_i - Y_i|$$

Donde n = es el número de datos no faltantes.

f_i = son las observaciones actuales.

Y_i = serie de tiempo estimada.

2. **Raíz del error cuadrático medio (RMSE):** Se encarga de medir el promedio de los errores de las predicciones. Al igual que el MAE, se encarga de comparar el valor observado con el valor del pronóstico donde las series de tiempo deben tener igual tamaño. Su fórmula está dada por.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2}$$

Donde:

n = número de datos no faltantes.

\hat{Y}_i = datos pronosticados.

Y_i = datos observados actuales.

3. **Error porcentual absoluto medio (MAPE):** A diferencia de las medidas anteriores que son dependientes de la escala de datos, esta medida pertenece a los errores porcentuales, ya que representa el error en porcentajes. El MAPE es uno de los más usados para medir el porcentaje de error en modelos de demanda y su fórmula está dada por.

$$MAPE = \frac{\sum_{t=1}^n \frac{|At - Ft|}{|At|}}{n}$$

Donde:

n : el número de periodos.

A_t : demanda real del producto.

F_t : demanda estimada.

Para comprender mejor los resultados que se puedan extraer de estas fórmulas, en la etapa de evaluación se realizaron las pruebas con los datos reales.

4.6.3 DESARROLLO DE MODELOS PREDICTIVOS

R Studio Software.

Para el modelado se utilizó el software R Studio, que es un software libre que trabaja con un lenguaje de programación interpretado, donde se pueden ir ejecutando las instrucciones directamente sin previa compilación del programa. Comúnmente es muy utilizado para realizar computación estadística y graficas debido a que cuenta con una variedad de técnicas que permiten realizar modelos como los de series temporales llevados a cabo en este proyecto. R es colaborativo y abierto donde los usuarios pueden publicar paquetes para hacerlo más robusto ayudando a mejorar sus funciones. R cuenta con un entorno en la nube conocido como R Studio Cloud que permite trabajar en una máquina virtual, la cual tiene la principal ventaja de no tener conflictos con las versiones de los paquetes a utilizar, motivo por el que se decidió utilizar esta versión para trabajar en el presente proyecto.

junto con las siguientes librerías;

- ***Library tidyverse***: se utiliza para facilitar la gráfica y manipulación de los datos mediante los paquetes ***ggplot2*** y ***dplyr***.

- **Library zoo:** se utiliza para proporcionar una infraestructura a los datos. Es muy utilizada en el pronóstico de series de tiempo. Funciones como **rollmean** se utilizan para calcular medias móviles.
- **Library smooth:** también se utiliza para realizar el promedio móvil simple.
- **Library fpp2:** se cargan paquetes para el análisis y visualización de modelos de series temporales.
- **Library forecast:** este se utiliza porque contiene herramientas y métodos para realizar cálculos de series de tiempo, es utilizado para modelos ARIMA y suavizados exponenciales.
- **Library tseries:** usado para medir estacionariedad en modelos de series temporales mediante la función `adf.test` (prueba de dickey-fuller), utilizada en modelo Arima.

Variables que influyen en la demanda.

- Se identificó influencia en la demanda de 4 variables que fueron útiles para la realización de los modelos;
 - Día, Mes, Año, Cantidad de kg vendidos.
- Se realizó la verificación de estos valores para que fueran numéricos de tal forma que no exista complicaciones a la hora de realizar los modelos.

La función ***is.numeric*** permite validar las variables para ver si son de tipo numérico.

```
> is.numeric(Detalle_Ventas$Dia)
[1] TRUE
> is.numeric(Detalle_Ventas$Mes)
[1] TRUE
> is.numeric(Detalle_Ventas$Anho)
[1] TRUE
> is.numeric(Detalle_Ventas$Cantidad_kg)
[1] TRUE
```

Primero se calculan los modelos de datos de ventas domiciliario.

Antes de realizar los modelos se hicieron algunos ajustes a los datos para poder trabajar de mejor manera en el software.

1.- Se renombraron las columnas para no tener problemas de comprensión, debido a que el software tiene conflictos con algunos caracteres de letras como la "ñ".

Mediante la función **colnames** quedaron de la siguiente manera.

```
colnames(Detalle_Ventas) <-  
  c("Dia", "Mes", "Anho", "Cantidad_kg", "Desc.Prod", "Precio.Unit",  
    "Total.Neto.Linea", "Nombre.Cliente", "Nombre.Doc")
```

2.- Se seleccionaron las columnas que se utilizarían para el modelado, siendo las primeras 4 columnas las elegidas.

```
Detalle_Ventas = Detalle_Ventas[ , 1:4]
```

3.- Se calcula la suma mensual total de la cantidad de kg de *pellets* vendidos y se agrupan por mes y año, creando una nueva base de datos llamada *Detalle_mensual* que será la que contenga la data ordenada para los modelos.

```
Detalle_mensual<-  
  Detalle_Ventas %>%  
  group_by(Anho, Mes) %>%  
  summarise( cant_mensual_kg = sum( Cantidad_kg ) )
```

4.- se crea una columna llamada fecha, que permitirá ver los datos gráficamente de mejor manera.

```
Detalle_mensual$fecha = paste( Detalle_mensual$Anho, Detalle_mensual$Mes,"1",sep = "-")  
Detalle_mensual$fecha <- as.Date( Detalle_mensual$fecha )
```

5.- se crea un gráfico en la figura 5 utilizando la función **ggplot** para observar el comportamiento de las ventas durante el tiempo (01/01/2021 – 01/06/ 2022).

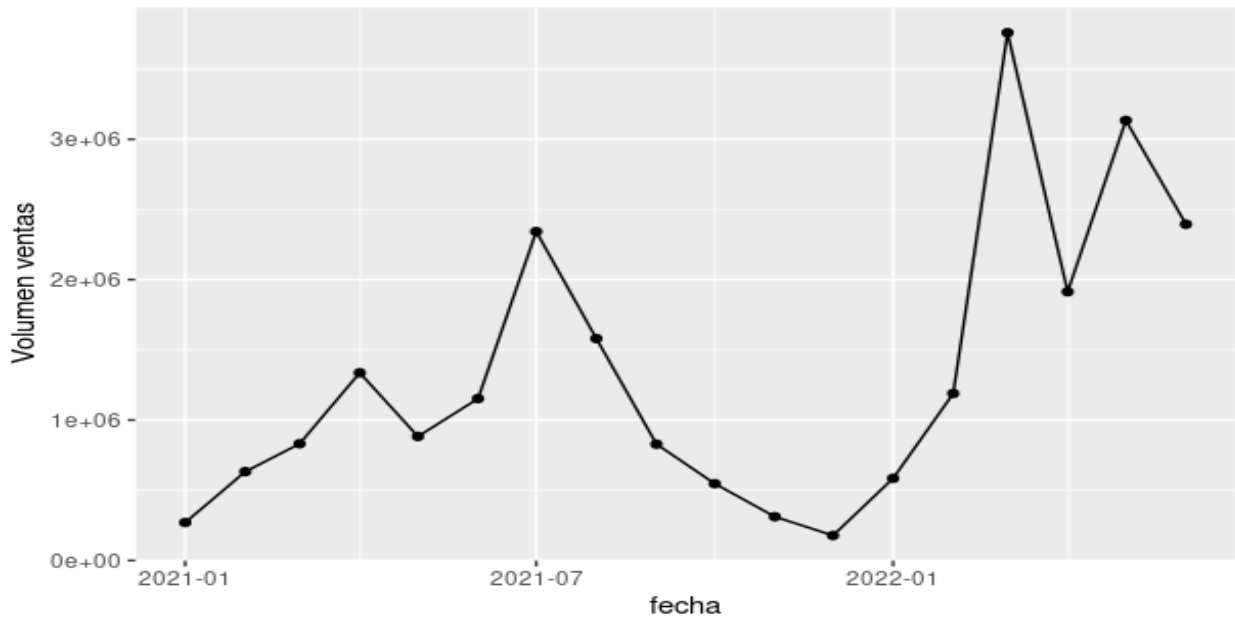


Figura 6: Comportamiento de ventas mensuales

4.6.4 Promedio móvil simple (*simple moving average SMA*)

Para el cálculo de la estimación de la demanda mediante el promedio móvil simple, se utiliza la librería ***smooth***.

El promedio móvil simple nos sirve para suavizar los datos históricos y además que se pueden hacer proyecciones hacia adelante mediante esta sencilla técnica. Su fórmula está dada por:

$$SMA = \frac{P1 + P2 + P3 + \dots + Pn}{n}$$

Donde.

P = es el promedio en periodo.

n = la cantidad de periodos.

Utilizando la librería **smooth** se generó la siguiente fórmula que permitió estimar la demanda mensual mediante el promedio móvil simple. Se debe dejar claro que; pese a que la función de pronóstico de **smooth** permite estimar varios meses en el futuro, el promedio móvil simple solo se utiliza para calcular un periodo debido a que depende de las observaciones reales anteriores. Para comenzar a observar algún patrón se hicieron pruebas con mayor cantidad de meses. Se hizo la prueba con 1 y 5 periodos para observar el comportamiento de la gráfica, el código para 1 periodo en adelante quedó de la siguiente manera.

```
media<-sma(ts(Detalle_mensual$cant_mensual_kg), order = 3)
pronostico<-forecast(media, h=1)
plot(forecast(pronostico))
```

Con *order = 3* se le está indicando que utilice 3 meses anteriores para realizar la estimación y con *h=1* se le está indicando que estime 1 mes hacia adelante, donde dicha estimación se ve reflejada con una x azul en el gráfico.

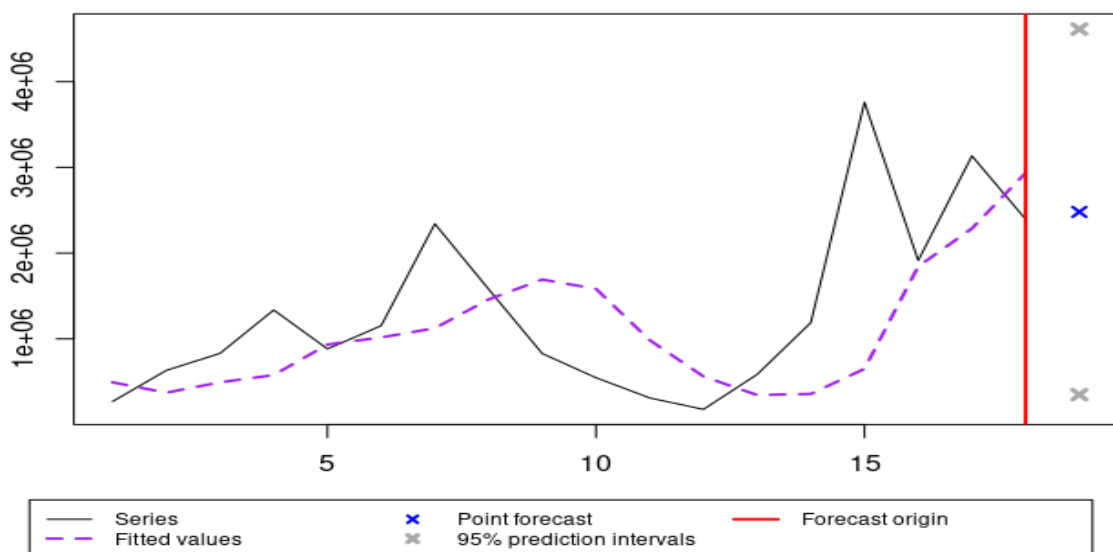


Figura 7: Proyección 1 mes con SMA

El gráfico para un periodo en el futuro muestra un comportamiento similar al último periodo salvo por una leve alza. Esto debido a que el punto de pronóstico (periodo 19) se ubica a la misma altura que dicho periodo que sería el 18.

Ahora observar cómo se comporta la gráfica si pronosticamos 5 meses en el futuro. Se modifica h en el código quedando de la siguiente manera.

```
media<-sma(ts(Detalle_mensual$cant_mensual_kg), order = 3)
pronostico<-forecast(media, h=5)
plot(forecast(pronostico))
```

Se obtiene la siguiente gráfica:

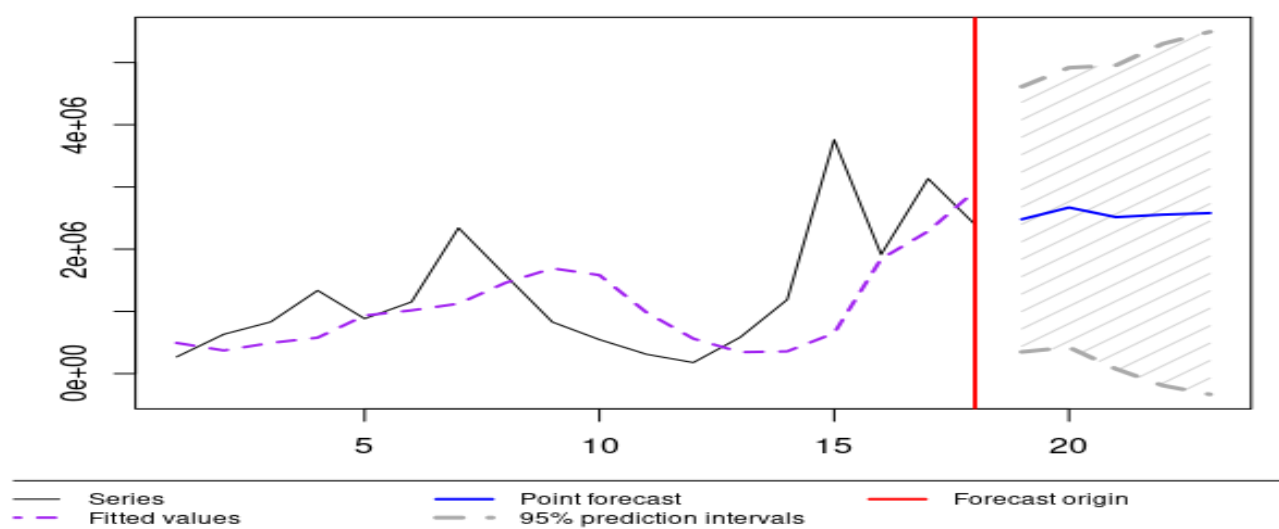


Figura 8: Proyección 5 meses con SMA

En esta gráfica se puede observar que el punto de pronóstico (línea azul) se mantiene relativamente constante, difiriendo mucho del comportamiento real de la serie de tiempo (línea gris) lo que lleva a concluir que este método no es efectivo para más de un periodo.

4.6.5 Suavizado exponencial simple (*simple exponential smoothing SES*)

Para realizar un suavizado exponencial simple se utiliza la librería **forecast**.

Este método se caracteriza por entregar mayor importancia a los datos más recientes de la serie de tiempo a diferencia del promedio móvil simple que entrega la misma importancia a cada uno de los datos, además de tener la ventaja de lograr óptimos resultados sin una gran cantidad de periodos pasados. La fórmula del SES está dada por:

$$F_t = F_{t-1} + \alpha(A_{t-1} - F_{t-1})$$

Donde:

F_t = nuevo pronóstico

F_{t-1} = pronóstico de ventas en unidades del periodo t-1.

A_{t-1} = demanda real en unidades del periodo t-1.

α = coeficiente de suavización (valor varía entre 0 y 1).

Utilizando las funciones de la librería **forecast** de R se puede realizar este cálculo de una manera rápida y efectiva. Al igual que el modelo anterior se hizo para 1 y 5 meses. El cálculo del SES en esta serie de tiempo quedó de la siguiente manera.

Primero se crea el SES con un numero h de periodos a calcular, en este caso es 1

```
fc <- ses(Detalle_mensual$cant_mensual_kg , h=1)
```

Con la función **summary(fc)** podemos ver todos los parámetros entre ellos la constante de suavizamiento, que en este caso se obtiene un valor de 0.6519

```
Smoothing parameters:  
alpha = 0.6519
```


Esta constante nos dice que mientras más el valor de Alpha se aproxima a 1 mayor es la importancia que le entrega a los datos de los últimos periodos. Por el contrario, si la constante se aproxima a 0 la importancia se la llevan los datos de los periodos anteriores.

El gráfico que se obtiene con $h = 1$ es el siguiente:

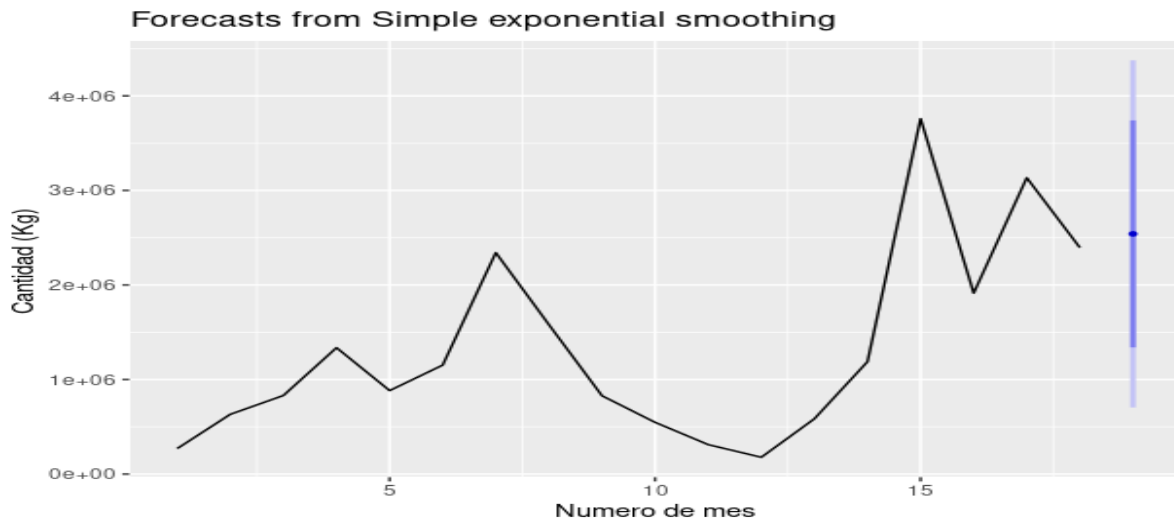


Figura 9: Proyección 1 mes con SES

El gráfico nos muestra un alza del punto de pronóstico con respecto al mes anterior, lo que significa que el método de suavizamiento exponencial proyecta un crecimiento en las ventas del periodo 19.

Ahora al probar con un $h=5$, es decir proyectando para 5 meses mediante este método, para esto cambiamos el valor de h .

```
fc <- ses(Detalle_mensual$cant_mensual_kg , h=5)
```

Los resultados que gráficamente que se obtienen son los siguientes.

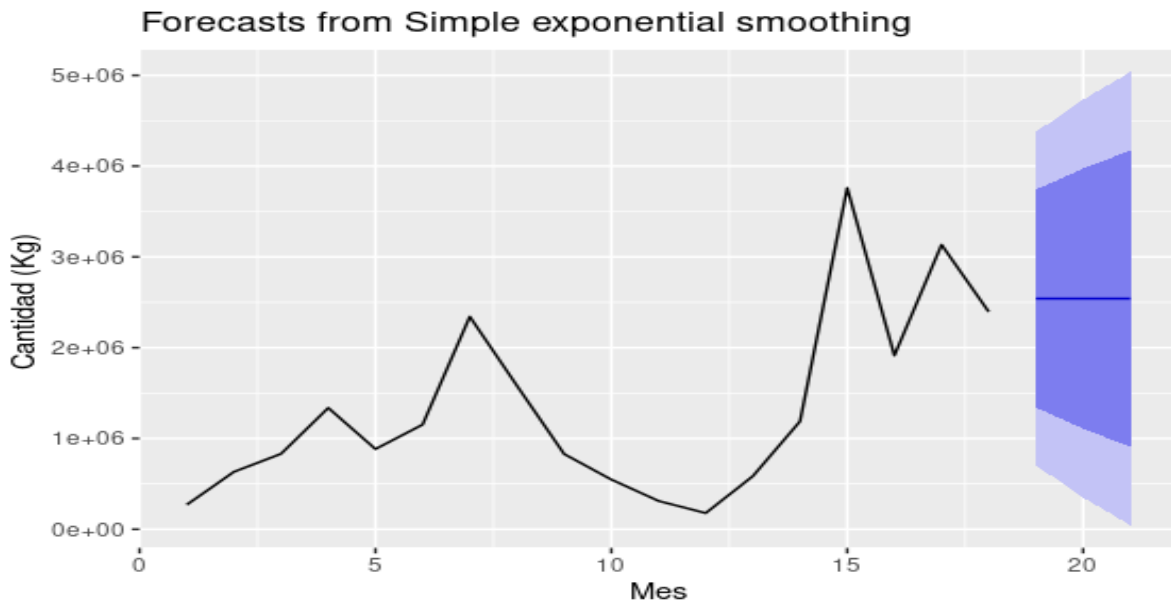


Figura 10: Proyección 5 meses con SES

Como se puede apreciar en el gráfico, para un periodo proyectado de 5 meses mediante este método se obtiene una constante, lo que significa que no es recomendable utilizar este modelo para un tiempo mayor a 1 periodo, ya que la demanda se repite indefinidamente.

4.6.6 Suavizado exponencial doble o Método de Holt (*Double exponential Smoothing*)

Para realizar este modelo se utiliza la librería **forecast**.

A diferencia del suavizado simple este modelo agrega una constante de suavizamiento llamada delta (δ) la que busca reducir el error entre demanda real y pronóstico. Mediante este modelo se busca prever la tendencia, incorporando al pronóstico uno de los patrones de series de tiempo. Para realizar el método de Holt está dado por la siguiente fórmula:

$$F_t = \alpha (\text{Demanda real del periodo anterior} + (1 - \alpha)(\text{Pronóstico del periodo anterior} + \text{tendencia estimada para el periodo anterior}))$$

$$T_t = \delta (\text{Pronostico de este periodo} - \text{Pronóstico del ultimo periodo} + (1 - \delta)(\text{Tendencia estimada para el último periodo}))$$

$$FIT_t = \text{Pronostico suavizado exponencialmente } (F_t) + \text{tendencia suavizada exponencialmente } (T_t)$$

Donde:

F_t = Pronóstico suavizado exponencialmente del periodo t.

T_t = Tendencia suavizada para el período t.

α = Constante de suavizamiento para el promedio.

δ = Constante de suavizamiento para la tendencia.

FIT_t = Pronóstico de demanda con tendencia.

Se realiza una prueba del modelo para 1 y 5 meses respectivamente, con la finalidad de ver el comportamiento al aumentar el número de meses de pronóstico. En R mediante la librería de **forecast** el cálculo se realiza de la siguiente manera:

```
fc_holt <- holt( Detalle_mensual$cant_mensual_kg , h=1)
```

Con $h=1$ proyectamos la demanda para el próximo periodo. En la siguiente gráfica se puede apreciar el comportamiento:

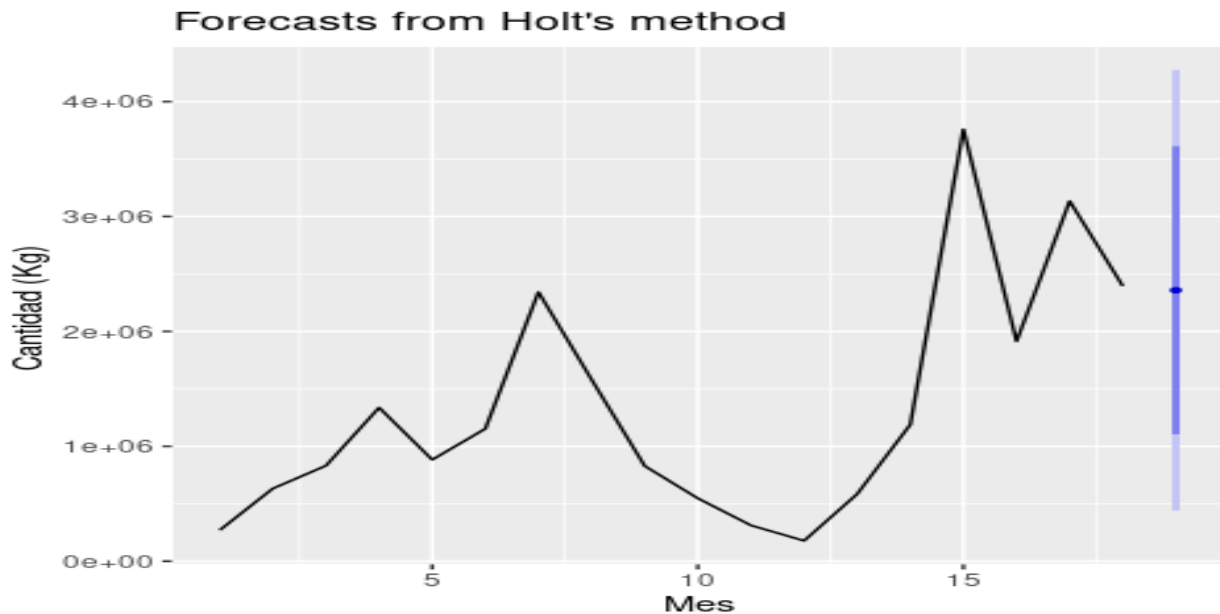


Figura 11: Proyección 1 mes con Holt

Para el siguiente período el modelo proyecta una leve baja en la demanda con respecto al periodo anterior, por lo que se realiza una proyección para 5 periodos con la finalidad de comprobar si esta tendencia será constante. Para esto modificamos el valor de h.

```
fc_holt <- holt( Detalle_mensual$cant_mensual_kg , h=5)
```

Realizamos la gráfica utilizando la función **autoplot**.

```
autoplot(fc_holt) + ylab("Cantidad (Kg) ") + xlab("Mes")
```

Se obtiene la siguiente gráfica proyectada para los próximos 5 meses:

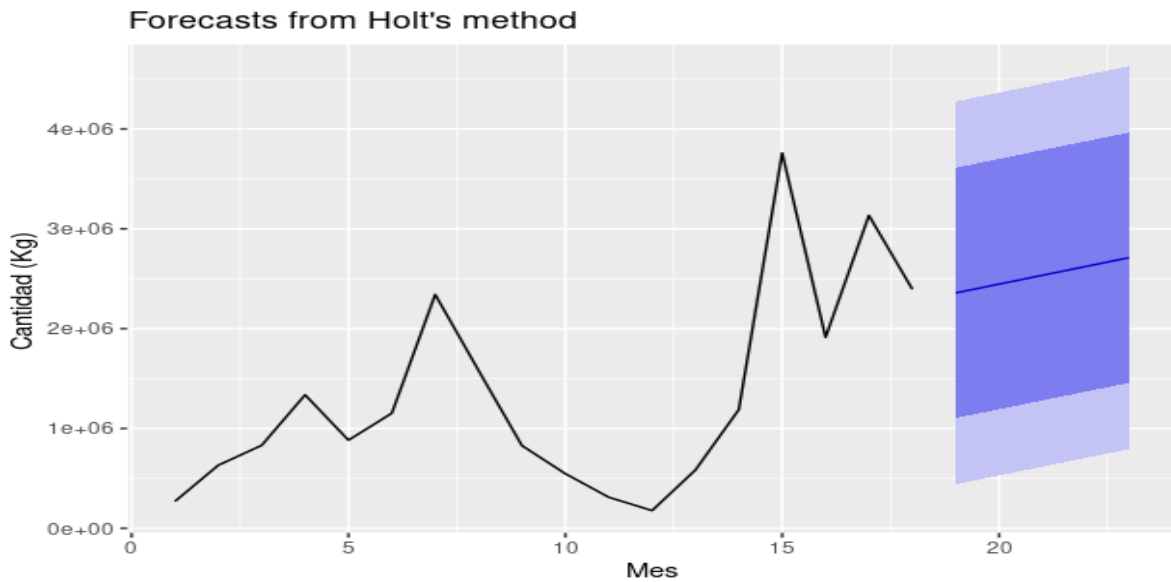


Figura 12: Proyección 5 meses con Holt

Basado en la cantidad de datos el modelo proyecta una tendencia creciente de la demanda futura (línea azul en la gráfica). Esto se puede explicar debido a que este modelo no captura la estacionalidad.

4.6.7 Método de Holt-Winters (*triple exponential Smoothing*)

Para realizar este modelo se utiliza la librería **fpp2**.

Este modelo a diferencia del método de Holt agrega otro componente de suavizado a la ecuación, el que corresponde a un componente de estacionalidad al modelo, otro patrón de series de tiempo, por lo que se tiene un suavizado para el nivel, otro para la tendencia y un tercero para estacionalidad. La fórmula para el cálculo de este modelo en su modo aditivo está dada por:

$$L_t = \alpha(Y_t - S_{t-p}) + (1 - \alpha)[L_{t-1} + T_{t-1}]$$

$$T_t = \gamma[L_t - L_{t-1}] + (1 - \gamma) T_{t-1}$$

$$S_t = \delta(Y_t - L_t) + (1 - \delta) S_{t-p}$$

$$\hat{Y}_t = L_{t-1} + T_{t-1} + S_{t-p}$$

Donde:

L_t = nivel en el tiempo t.

T_t = tendencia en el tiempo t.

γ = ponderación para la tendencia.

S_t = componente estacional en el tiempo t.

δ = ponderación para el componente estacional.

α = ponderación para el nivel

p = período estacional.

Y_t = valor de los datos en el tiempo t.

\hat{Y}_t = pronóstico de un período adelante, en el tiempo t.

En R se realiza utilizando la librería **forecast** del paquete **fpp2**.

La función para el cálculo en R quedo de la siguiente manera:

```
demand <- ts(Detalle_mensual$cant_mensual_kg, start = c(2021, 1), frequency = 12)
```

```
hw <- HoltWinters(demand)
```

La consola arrojó el siguiente mensaje de error;

```
> hw <- HoltWinters(demand)
Error in decompose(ts(x[1L:wind], start = start(x), frequency = f), seasonal) :
  time series has no or less than 2 periods
```

Esto ocurre debido que al ejecutar la base de datos que se ha utilizado para los demás modelos, no es posible realizarlo, debido a que no logra captar periodos repetitivos que le permitan capturar la estacionalidad porque la serie de tiempo tiene menos de 2 periodos completos.

Para lograr obtener alguna predicción de este modelo se ha decidido replicar hacia atrás los datos del primer año de la serie de tiempo, permitiendo obtener 2,5 periodos. Con esto se ha creado otra base de datos a la que se le llamó `Detalle_mensual1`. La fórmula modificada queda de la siguiente manera:

```
demand <- ts(Detalle_mensual1$cant_mensual_kg, start = c(2020, 1), frequency = 12)
```

```
hw <- HoltWinters(demand)
```

Se le realiza un *plot* a la variable *demand* para observar el comportamiento de los datos, se obtienen los siguientes valores de Alpha beta y gamma con la siguiente gráfica:

```
Call:
HoltWinters(x = demand)

Smoothing parameters:
alpha: 0.2349782
beta : 0.4319815
gamma: 0.1
```

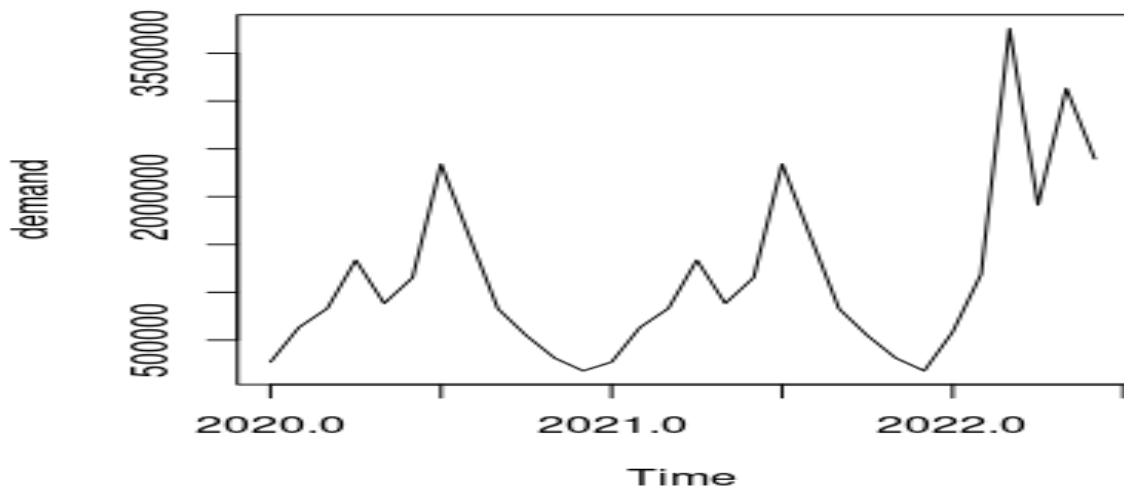


Figura 13:Comportamiento de datos al replicar un periodo

Se puede ver que los periodos 2020 y 2021 están repetidos los comportamientos por lo que se prosigue con la estimación del modelo HoltWinters.

Para esto se realiza una estimación a 12 meses con un intervalo de confianza de 0,95 mediante el siguiente comando:

```
forecast <- predict(hw, n.ahead = 12, prediction.interval = T, level = 0.95, seasonal= "Additive")
plot(hw, forecast)
```

Se realiza un plot a las variables (hw, forecast) creadas con el pronóstico y los datos observados ajustados. Se obtiene la siguiente gráfica:

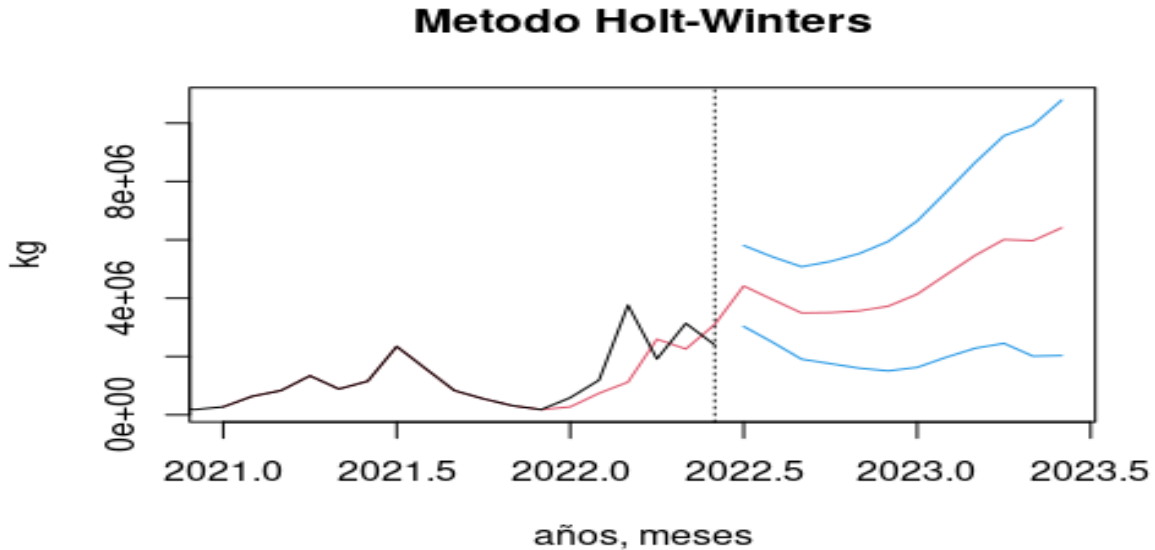


Figura 14: Proyecciones con datos replicados usando HW

Se puede apreciar mediante la línea roja que corresponde al pronóstico, que este modelo estima un crecimiento de la demanda para los periodos próximos, este ejercicio se realiza para observar una aproximación de cómo se comportaría el modelo si hipotéticamente se contara con una mayor cantidad de datos, debido a que los datos utilizados para realizarlo no se ajustan a la realidad.

4.6.8 ARIMA Y AUTOARIMA (*autoregressive integrated moving average*).

Este modelo de ARIMA entrega un enfoque diferente a los que se vieron anteriormente, es decir Holt y HoltWinters se enfocan en la tendencia y estacionalidad, mientras que ARIMA se basa en la autocorrelación de los datos para realizar predicciones. Durante esta sección se realiza una explicación resumida de los modelos ARIMA.

ARIMA es la combinación de un modelo autorregresivo y un modelo de media móvil, los que están dados por las siguientes fórmulas para el cálculo:

Modelo autorregresivo: Las observaciones son predecibles a partir de observaciones pasadas más un término de error.

$$y_t = c + \theta y_{t-1} + \varepsilon_t$$

Donde:

y_t = variable en un momento del tiempo.

c = constante que puede ser positiva, negativa o 0.

θ = coeficiente.

y_{t-1} = variable en el periodo anterior.

ε_t = Ruido blanco (*White noise*).

Modelo de media móvil de AR-I-MA: al contrario de utilizar valores pasados, este modelo (MA) utiliza errores de pronósticos pasados, no se debe confundir este modelo con la media móvil simple, este modelo (MA) funciona similar a una regresión. La fórmula de este modelo está dada por:

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

Donde:

Y_t = predicción de la variable en el periodo t.

ε_t = ruido blanco (*White noise*).

\mathcal{E}_{t-1} = ruido blanco (*White noise*) periodo anterior.

θ_1 = constante *tetha* en el periodo 1.

- **AUTOARIMA**: este modelo es una función de R que mediante variaciones en el algoritmo realiza un modelo automáticamente.

Se realiza este modelo en R Studio Cloud ejecutando los siguientes comandos:

```
modelo_autoarima<-auto.arima(Detalle_mensual$cant_mensual_kg)
summary(modelo_autoarima)
```

Con el *auto.arima* se genera el mejor modelo, el que se almacena en la variable *modelo_autoarima*. Se le aplica un *summary* a la variable y se obtuvieron los siguientes resultados.

```
> summary(modelo_autoarima)
Series: Detalle_mensual$cant_mensual_kg
ARIMA(0,1,0)

sigma^2 = 9.407e+11:  log likelihood = -258.47
AIC=518.93  AICc=519.2  BIC=519.77
```

Con estos resultados se puede extraer conclusiones con respecto a lo que nos podría aportar este modelo y es lo que se conoce como *Random Walk*; ARIMA (0,1,0). Esto representa (p, d, q), donde p es el orden de la parte autorregresiva, d es el grado de la primera diferenciación involucrada, y q corresponde al orden de la media móvil. Significa que los valores que puede predecir son aleatorios, es decir la predicción va a variar de forma aleatoria hacia arriba o hacia abajo con respecto al periodo anterior. Para lo que necesitamos calcular, que es la demanda mensual no es posible utilizar este modelo, ya que el área de producción dependería de un algoritmo aleatorio calcule cuanto producir.

MODELO ARIMA con xreg

Por lo tanto para buscar encontrar mejores resultados se aplica un modelo ARIMA con una co-variable, donde se elige el mes como co-variable, lo que permitirá al modelo poder estimar utilizando mas información de los datos.

Para esto primero se llama a la librería **tseries** la que nos permitirá llamar la prueba dickey fuller que mide la estacionariedad en los datos. No confundir estacionariedad con estacionalidad, la primera hace alusion a que las variaciones de la serie de tiempo son constantes, mientras la estacionalidad hace referencia al comportamiento que tiene dependiendo del periodo estacional del año. Para poder aplicar un modelo ARIMA la serie debe ser estacionaria, por lo que si la prueba dickey fuller muestra que la serie de tiempo no es estacionaria se debe hacer la conversión de esta.

Para medir la estacionariedad se realizó mediante el siguiente comando:

```
adf.test(Detalle_mensual$cant_mensual_kg)
```

Esta prueba arrojó los siguiente resultados

```
Augmented Dickey-Fuller Test  
data: Detalle_mensual$cant_mensual_kg  
Dickey-Fuller = -1.8243, Lag order = 2, p-value = 0.6393  
alternative hypothesis: stationary
```

Con un p-value > a 0,05 nos indica que la serie de tiempo no es estacionaria.

Debido al valor de p-value se utiliza la función *diff* para convertir los datos, y logaritmo para disminuir el tamaño de los valores de la serie de tiempo, en un proceso llamado transformación.

```
y = diff( log( Detalle_mensual$cant_mensual_kg))
```

Se crea la variable para contener el modelo ARIMA y se agrega la co-variable a la función `xreg`, que en este caso sera el Mes.

Se realiza el pronóstico para los siguientes 5 meses utilizando el siguiente comando, donde la función `xreg` selecciona los meses a predecir;

```
fc_mod = forecast(mod_fc, h=10, xreg= c(7,8,9,10,11))
```

```
mod_fc = Arima( y , order = c(0,1,0) , xreg = Detalle_mensual$Mes[-1] )
```

S e aplica una transformacion inversa a los datos, exponencial para el logaritmo y suma acumulada para la para la diferencia (*diff.*), obteniendo los datos pronosticados en una variable.

```
fut_arima = exp( log( Detalle_mensual$cant_mensual_kg[18]) +  
                cumsum( fc_mod$mean ) )
```

Se crea una variable para las fechas futuras.

```
fecha_fut = seq.Date(from = as.Date("2022-7-1"),by="months", length.out = 5)
```

Se crean variables para para la demanda real (`df_new1`) y la demanda futura (`df_new2`).

```
df_new1 = Detalle_mensual[,3:4]  
df_new2 = data.frame( fecha=fecha_fut, cant_mensual_kg=fut_arima)
```

Se crea una variable para contener las las demandas

```
df_new1$group= "Real"
df_new2$group= "Futuro"
df_arima = rbind( df_new1, df_new2)
```

Finalmente se gráficán los resultados obtenidos.

```
ggplot(data=df_arima,aes(x=fecha, y= cant_mensual_kg, color=group)) +
  geom_point( ) + geom_line() +
  labs(x="Mes", y="Valor mensual", title="Modelo ARIMA con Xreg")
```

El modelo ARIMA arrojó el siguiente comportamiento de la demanda para los próximos 5 periodos:

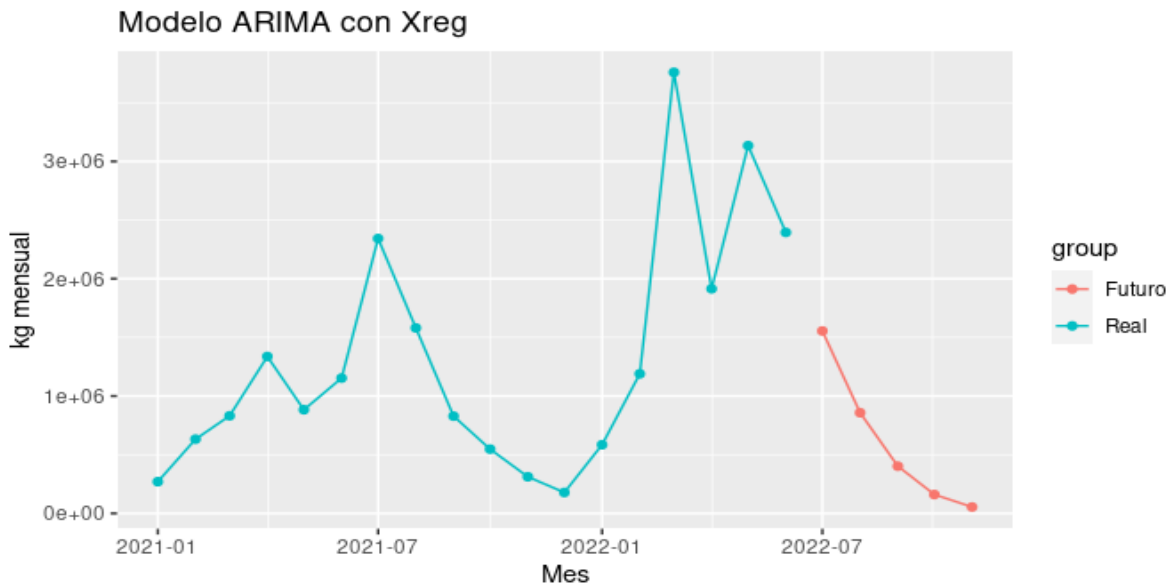


Figura 15: Proyecciones ARIMA con co-variable

4.7 MODELOS DEMANDA INDUSTRIAL.

Para la estimación de la demanda industrial se recurre a realizar una media móvil y un suavizado exponencial simple. Esto debido a la baja cantidad de datos, que si bien son grandes cantidades no se puede establecer una tendencia estacionalidad.

Se realizan los mismos pasos que los datos de ventas a domicilio donde se crea una variable llamada Detalle_industrial que almacena las sumas totales de ventas en kg por meses.

El comportamiento de las ventas se puede observar en el siguiente gráfico:

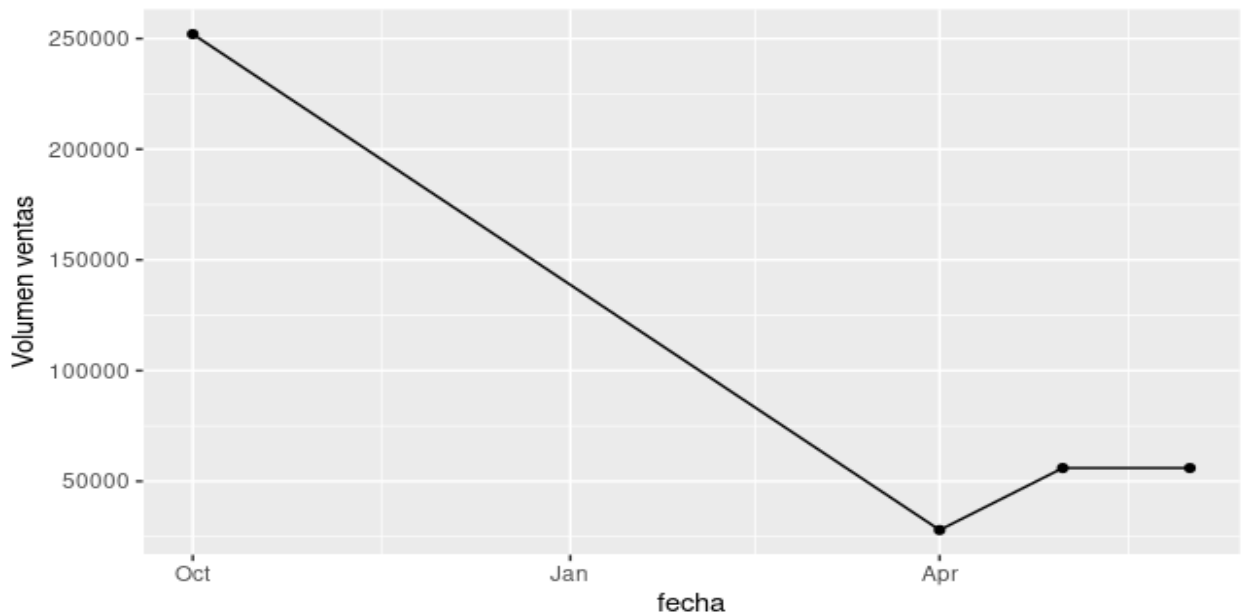


Figura 16: Comportamiento de ventas industrial

Las ventas para uso industrial se agruparon principalmente en 4 meses

Debido a la baja cantidad de periodos se procede a realizar la predicción solamente para 1 periodo en el futuro. Para esto se realiza un modelo de promedio móvil simple y un suavizado exponencial simple.

4.7.1 Media Movil Industrial (SMA)

Se realiza un cálculo en R para estimar un periodo de demanda en el futuro, para esto se utilizan los 4 meses anteriores. El cálculo en R queda de la siguiente manera:

```
media_industrial<-sma(ts(Detalle_industrial$cant_mensual_kg), order = 4)
pronostico1<-forecast(media_industrial, h=1)
plot(forecast(pronostico1))
```

Con la función `order = 4` le asignamos que se utilizaran los 4 meses anteriores para predecir la demanda del periodo siguiente y con `h = 1` se le asigna la cantidad de periodos a predecir. Luego se realiza un *plot* de la predicción, el que se puede observar en la siguiente gráfica:

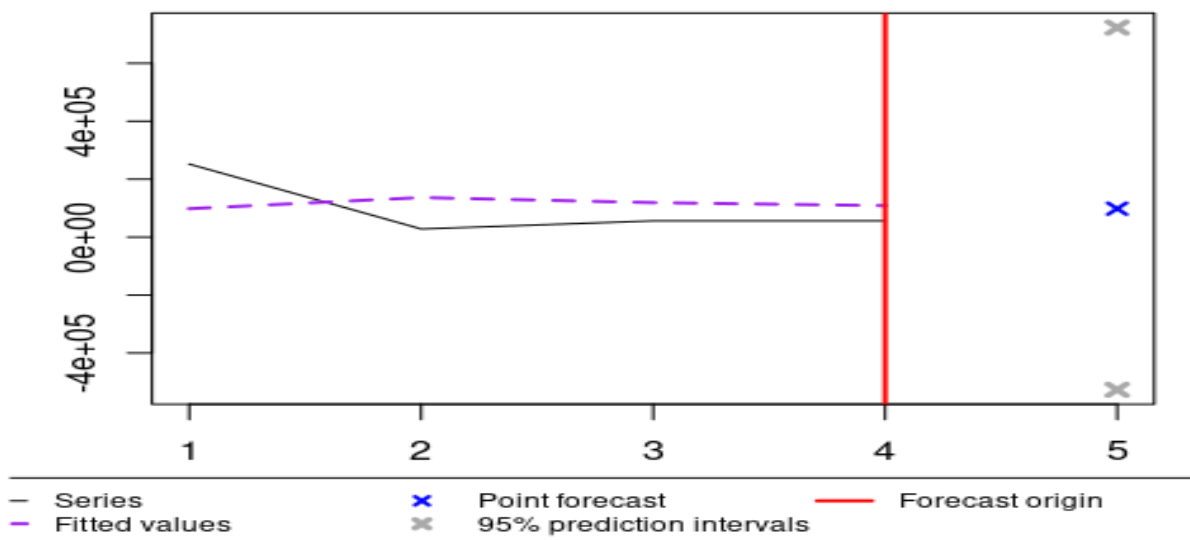


Figura 17: Proyecciones ventas industriales usando SMA

Se observa en el punto de pronóstico (x azul), un leve aumento en el valor de la demanda, esto motivado por el primer dato que se encontraba muy elevado en comparación a los siguientes tres. Pero en general con este modelo se espera que la

demanda se mantenga lineal durante el tiempo, aunque se debe evaluar si a medida que se agregan datos estos presentan saltos muy significativos. Lo que haría comenzar a evaluar la utilización de un modelo diferente.

4.7.2 Suavizado exponencial simple industrial (SES)

Se procede a realizar un suavizado exponencial para poder estimar el valor pronosticado para el siguiente periodo. En R utilizando la librería **forecast** se realiza utilizando las siguientes instrucciones:

```
fc_industrial <- ses(Detalle_industrial$cant_mensual_kg , h=1)
```

Se obtiene un alfa de 0,875 lo que indica que hay una mayor importancia a los datos del último periodo.

```
Smoothing parameters:  
alpha = 0.875
```

con $h=1$ para pronosticar solo un periodo se obtienen el siguiente resultado

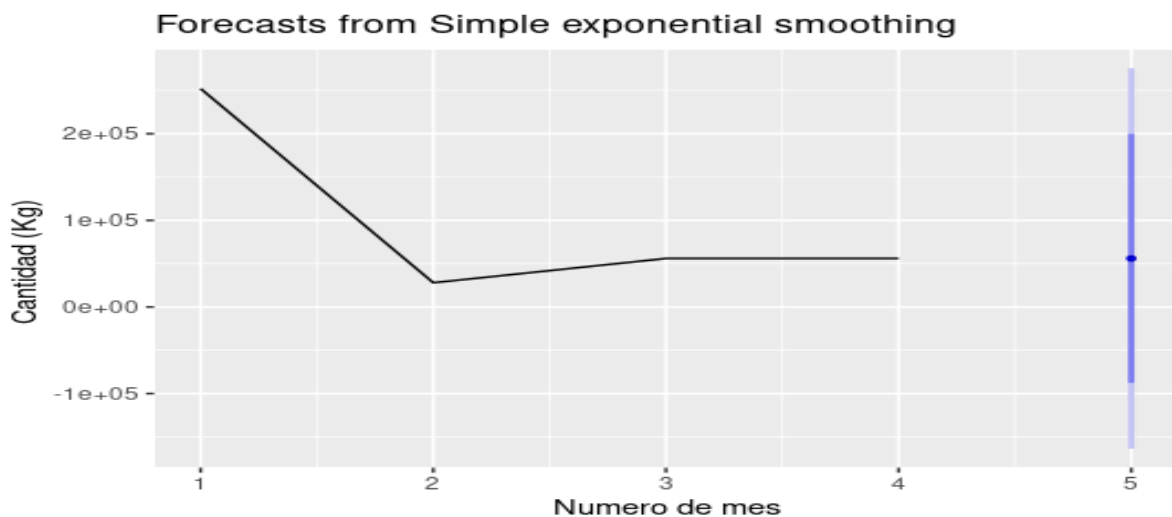


Figura 18: Proyección de ventas industriales usando SES

Se observa que la demanda estimada aproximadamente igual a la del último periodo y si se sigue aumentando la cantidad de meses estimados esta constante seguirá repitiéndose como se puede observar con un $h=5$ en la siguiente gráfica.

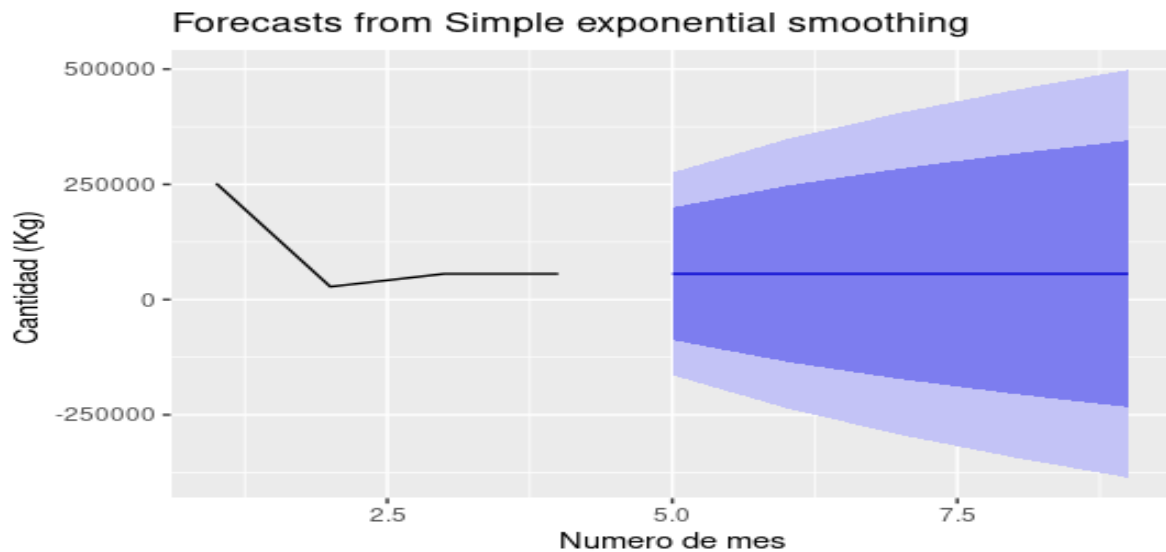


Figura 19: Proyecciones de ventas SES para 5 meses

A medida que se vayan sumando más datos es posible que el modelo vaya variando, dependiendo del comportamiento de estos, aunque también es posible que mantenga este comportamiento lineal si no se consiguen más clientes del sector industrial.

4.7.3 ARIMA industrial.

Se realiza un modelo de auto arima para observar el comportamiento de los predictores p,d,q .

```
modelo_autoarima1<-auto.arima(Detalle_industrial $cant_mensual_kg)
summary(modelo_autoarima1)
```

Se le aplica un *summary* a la variable creada `modelo_autoarima1` y se observan los siguientes resultados;

```
> summary(modelo_autoarima1)
Series: Detalle_industrial$cant_mensual_kg
ARIMA(0,0,0) with zero mean
```

Se puede observar que nos arroja un ARIMA(0,0,0) lo que significa que la serie de tiempo es *white noise*, quiere decir que las variables son independientes y se distribuyen idénticamente con media 0, es decir no existe una correlación entre ellas. Por lo que no es posible realizar un modelado en ARIMA con estos datos, debido a que tampoco cumple con la mínima cantidad de datos recomendados para realizar estos modelos.

4.8 EVALUACIÓN

En esta etapa se realiza la evaluación de los diferentes modelos realizados. Donde se aplican las métricas de error, lo que significa que podremos ver que tan cerca se encuentran los pronósticos con respecto a los datos reales.

4.8.1 Evaluación de los resultados

Con la finalidad de establecer cuál es el mejor modelo, o el que mejor se adapta a las necesidades de la empresa se realizó una evaluación mediante las métricas que se mencionaron en el plan de pruebas. Las métricas y su respectiva descripción en la interpretación general corresponden a las siguientes:

- **Error absoluto medio (MAE):** Nos entrega la media de la diferencia absoluta entre la predicción del modelo y el valor objetivo. Para realizar el cálculo se realiza mediante los siguientes pasos:

1.- se realiza una resta de la diferencia entre los valores que se han pronosticado y los valores reales en cada punto que se pronostica.

2.- se calcula el valor absoluto de lo que se obtuvo anteriormente.

3.- se suman los resultados obtenidos anteriormente.

4.- se realiza una división de la suma por la cantidad de puntos que se pronosticaron.

Interpretación: Se obtiene un promedio de los errores, donde si el valor obtenido es un valor demasiado elevado, significa que el error del pronóstico es alto, lo que nos indica que la predicción no es confiable, esto depende de la escala de los datos con los que se esté trabajando.

- **Raíz del error cuadrático medio (RMSE):** Esta medida de error se utiliza para medir cuánto es el error entre los dos conjuntos de datos es decir el valor pronosticado y el valor real, a diferencia del MAE le entrega mayor importancia a los *outliers* y se realiza mediante los siguientes pasos:

1.- restar las diferencias que existe entre los puntos pronosticados y los reales, y el resultado elevarlo al cuadrado.

2.- sumar los resultados obtenidos de los cuadrados de las diferencias.

3.- dividir el valor de la suma obtenida y dividirla por la cantidad de puntos que se han pronosticado.

4.- sacar la raíz cuadrada del resultado obtenido.

Interpretación: mientras mayor sea el resultado obtenido, mayor será el error de predicción y el modelo será de menor calidad o menos preciso.

- **Error porcentual absoluto medio (MAPE):** sirve para medir el comportamiento de un modelo en términos porcentuales, se calcula mediante los siguientes pasos:
 - 1.- restar valor pronosticado y valor real en los puntos pronosticados.
 - 2.- calcular el valor absoluto del resultado obtenido anteriormente.
 - 3.- se realiza una división del resultado obtenido por el valor real de los puntos de la serie temporal
 - 4.- sumar todos los valores que se obtuvieron en la división anterior y luego se divide por la cantidad de puntos que se pronosticaron.

Interpretación: esta medida tiene la característica de ser una de las más fácil de interpretar, es decir al entregar un porcentaje alto significa que el modelo contiene un alto error, si el porcentaje entregado es bajo significa que el modelo es más preciso y contiene menos error.

CALCULO DEL ERROR EN R STUDIO VENTAS DOMICILIARIA.

Suavizado exponencial simple:

```
> accuracy(fc)
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 179195.8 882750.3 644753.8 -14.91718 61.22641 0.8747985 -0.08685923
```

Metodo de Holt:

```
Error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set -192447.1 862459 678626.6 -108.7755 126.7733 0.920757 0.3733095
```

Autoarima

```
Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 118042 942573.4 696100 -5.491044 52.27 0.9444648 -0.4055726
```

ARIMA

Error measures:

| | ME | RMSE | MAE | MPE | MAPE | MASE | ACF1 |
|--------------|-------------|-----------|-----------|----------|----------|-----------|-----------|
| Training set | -0.02752952 | 0.7007583 | 0.5360831 | 80.66251 | 119.1316 | 0.7826969 | -0.627427 |

Promedio movil simple

Para esta métrica se realizó el calculo de forma manual en MS Excel utilizando las formulas (Anexo 1,2 y 3).

R nos entrega algunas métricas de medicion del error, dentro de las que se encuentran las que se plantean junto con otras que no seran consideradas.

4.8.2 Revisión

Para los resultados obtenidos de las metricas de error en cada uno de los modelos se realiza una revisión e intepretacion del numero obtenido de las tres metricas seleccionadas. Para poder guiar algunos resultados que entregan las metricas se realiza el calculo de de media de las cantidades de kg vendidos en total.

Para esto se utiliza la función mean con la base de datos y la columna a la cual se le calculará la media, queda de la siguiente manera:

```
mean(Detalle_mensual$cant_mensual_kg)
```

Ejecutando ese comando se obtuvo el siguiente resultado:

```
> mean(Detalle_mensual$cant_mensual_kg)
[1] 1326148
```

La media equivale a 1.326.148 kg de *pellets*, equivalente a 1326 toneladas aproximadamente.

Revisión de las métricas:

1.- Suavizado exponencial simple

- **RMSE = 882.750**

Este valor nos indica que existe un error de esa cantidad en la predicción, esto se encuentra bajo la media de ventas de kilogramos de *pellets*, pero es una cantidad considerable no menor.

- **MAE = 644.753**

Este valor que se obtuvo en el cálculo del MAE es menor que la media, equivalente aproximadamente a la mitad, y también menor a la métrica de RMSE, donde si bien es más alentador que el del RMSE, sigue siendo elevado para una producción mensual de este producto.

- **MAPE: 61.22%**

En este caso esta medida porcentual es mayor a la mitad, lo que se considera bastante elevado para lo que se pudiera esperar en un modelo predictivo que ayude a estimar bien la demanda.

2.- Metodo de Holt

- **RMSE = 862.459**

Valor similar al obtenido en el SES, que si bien es mínimamente menor no es suficiente para creer que este modelo se comporta de mejor manera.

- **MAE = 678.626**

Este valor igual es similar al de SES, mínimamente mayor, lo que significa que contiene un error aun mayor que en el SES, ambos continúan siendo elevados.

- **MAPE = 126.7%**

Con esta métrica el resultado es mucho menos alentador que las demás, significa que existe una cantidad de error demasiado grande en la predicción con este modelo.

3.- Autoarima

- **RMSE = 942.573**

Error más elevado que los anteriores, y más cercano a la media

- **MAE = 696.100**

También es más elevado que los anteriores, lo que significa que con estas métricas es el modelo con mayor error hasta el momento.

- **MAPE = 52%**

Se podría pensar que es un modelo más exacto por tener un porcentaje de error más bajo comparado con los demás, pero se debe considerar que esto se genera debido a que este modelo dio como resultado una caminata aleatoria (*Random walk*).

4.- ARIMA

- **RMSE = 0,70**

El valor obtenido para esta métrica en este modelo es un valor más cercano a lo esperado, debido a que el ideal es 0 además considerando que la media se encuentra lejana es un resultado bueno y el modelo más preciso.

- **MAE = 0,53**

Al igual que la métrica anterior se encuentra muy lejana de la media y cercana a 0 lo que indica que es un resultado bueno y el modelo es más preciso basado en esta métrica.

- **MAPE = 119.13**

Al contrario de las métricas anteriores esta medida entrega un porcentaje muy elevado contradiciendo lo que entregaron las otras métricas.

5.- Promedio Móvil Simple.

- **RMSE= 581.390**

Este error al igual que los anteriores está muy elevado, considerando que se espera que en esta métrica esté más cercana a 0 para disminuir dicho error. Lo

que nos hace ver que este modelo podría ser un tanto impreciso para esa cantidad de datos.

- **MAE = 422,716**

Métrica que entrega un valor menor a los anteriores, sin embargo sigue siendo elevado para poder decir que es un modelo con buena precisión, lo que significa que no es confiable guiarse solo porque tiene un valor menor a los demás con esta métrica.

- **MAPE = 39%**

Esta métrica nos indica que el modelo contiene un margen de error del 39% es decir es bajo comparado con los obtenidos en los modelos anteriores utilizando solamente esta medida.

Cálculo del error en modelos de demanda industrial.

Si bien es posible obtener métricas de medición de los modelos de demanda industrial, se considera que el aporte que pueden hacer es irrelevante para el proyecto debido a la calidad de los datos. Esto porque se obtiene un comportamiento lineal debido a la baja cantidad de datos y además que dichos datos pertenecen a una sola empresa, lo que hace que el cálculo de la demanda no se pueda llevar a cabo de forma efectiva con modelos de series de tiempo.

4.8.3 Determinación próximos pasos

Luego de analizados los modelos y medidos los errores de cada uno se debe proceder con los siguientes pasos:

- 1.- Elección del mejor modelo para la empresa.
- 2.- Realizar manual de como implementar el modelo en R.
- 3.- Establecer el proceso para la implementación del modelo.
- 4.- Obtener conclusiones con respecto al proyecto realizado junto con los comentarios finales de los modelos.

4.9 PLAN DE IMPLEMENTACIÓN

Este plan de implementación contempla los pasos a seguir para la instalación del software R tanto local como en la nube, además de las instrucciones para la ejecución del modelo elegido en dicho software, así como el procedimiento a seguir en la carga de la base de datos tal cual fue entregada por la empresa para este proyecto.

Se espera que la base de datos que se usara para los modelamientos futuros contenga las mismas columnas que la utilizada en este proyecto, lo que facilitará el trabajo para la predicción y funcionamiento correcto del modelo.

Además de que la persona que se encargue de realizar el modelado en el futuro debe tener un grado de conocimiento de manejo de tablas de MS Excel y software relacionado con datos.

Para efectos de esta tesis solo se especifica el plan general, luego para entregar a la empresa se creará un manual de instrucciones más detallado.

El plan de implementación se divide en 3 etapas.

1. Preparación de la base de datos.
2. Instalación del software.
3. Ejecución del modelo elegido.

PREPARACIÓN DE LA BASE DE DATOS.

La base de datos de la empresa se encuentra en formato MS Excel con tablas divididas en hojas dentro del archivo por lo que es necesario realizar la separación.

- a. Separar tabla de datos a ocupar.
- b. Eliminar columnas innecesarias para el proyecto.
- c. Crear nueva columna con el nombre de “Cantidad en kg” la que contendrá la cantidad de *pellets* vendidos en kilogramos, para esto además se necesita ejecutar alguna fórmula en MS Excel como la función “Si.conjunto”.

- d. Renombrar el archivo a Detalle_Ventas, para facilitar la ejecución del código en R.

INSTALACIÓN DEL SOFTWARE R.

El software R se puede ejecutar tanto con instalación local como en la nube. Este proyecto se ejecutó principalmente en la versión de la nube de R Studio.

Para instalación local se deben seguir los siguientes pasos:

- a. Instalar el software de R, para eso se debe ingresar a la página oficial cloud.r-project.org donde podrá elegir para que sistema operativo descargar.
- b. Ejecutar el archivo descargado.
- c. Otorgar todos los permisos solicitados.
- d. Instalar el entorno de desarrollo integrado (IDE) R Studio.

Se instala este entorno de desarrollo R Studio porque permite trabajar de una manera más práctica, es decir permite facilitar tareas como administrar archivos, contar con herramientas para escribir y revisar el código, fluidez de trabajo, capacidades visuales, entre otras ventajas.

Instalación R Studio

- a. Ir al sitio oficial RStudio.com y buscar la pestaña download para descargar el archivo para ejecutar.
- b. Seleccionar la opción gratuita R Studio desktop y descargar.
- c. Ejecutar el instalador del archivo.
- d. Otorgar todos los permisos.

Con esto instalado solo faltarían los paquetes necesarios para ejecutar el modelo.

VERSIÓN EN LA NUBE (R STUDIO CLOUD)

Para ejecutar R Studio en la nube es más sencillo y eficiente que tenerlo instalado localmente, salvo por la desventaja de que la opción gratuita es limitada en cuanto a las horas y cantidad de CPU. Pero para este proyecto es más que suficiente.

Para obtener R Studio en la nube se deben seguir los siguientes pasos:

- a. Ingresar a la página rstudio.cloud.
- b. Crear una cuenta de usuario en la opción free.
- c. Crear proyecto para acceder a las vistas.

Con esta cuenta podrá acceder a sus proyectos cuando sea y donde sea desde cualquier dispositivo que lo permita, lo que se traduce en una gran ventaja con respecto a la versión local.

Ejecución del modelo.

Para ejecutar el modelo se deben seguir los siguientes pasos:

- a. Cargar la base de datos.
- b. Ordenar la base de datos.
- c. Instalar paquetes necesarios.
- d. Ejecutar el código del modelo.

5 Conclusiones

Durante el transcurso del proyecto se pudo observar diferentes comportamientos que ayudaron a extraer algunas conclusiones finales.

Lo primero que se puede concluir es que si se desea obtener resultados más efectivos y realistas con respecto a la demanda futura de una empresa es estrictamente necesario que exista una gran cantidad de datos de ventas, de modo que permitan a los modelos poder hacer análisis históricos para intentar predecir las ventas futuras, esta cantidad de datos se espera debe rondar alrededor de 4 años o unos 50 periodos aproximadamente, permitiendo mayor precisión en la predicción.

El uso de modelos de series temporales como los utilizados en este proyecto nos permitió trabajar con una cantidad de variables de datos limitadas, es decir se debe utilizar en su mayoría la variable de cantidad de productos vendidos, si se espera que influyan más variables en la predicción se debe recurrir a otro tipo de modelos más complejos.

Para el análisis de los datos se trabajó con el software de R Studio Cloud, si bien existen otros software y lenguajes que también permiten hacer análisis de series temporales se puede concluir que con R Studio cloud no hubo mayores limitantes, y sus ventajas de uso como la curva de aprendizaje, gratuidad de uso y poder trabajar en la nube en una máquina virtual lo hacen ser un excelente y recomendado software para realizar este y proyectos similares.

Dentro de los modelos que se realizaron en el software de R, hubo algunos de menor y mayor complejidad, es decir mientras más complejos se requirió de mayor tiempo poder realizarlos. Se observó que algunos como promedio móvil o suavizado exponencial simple pueden ser modelos confiables si lo que se desea predecir es un periodo adelante sin considerar tendencia ni estacionalidad en los datos. Por ende el mejor modelo y que mejor se adapta a los datos es el modelo ARIMA con co-variable, ya que este modelo permitió trabajar con la cantidad de ventas y mes, lo que hace que puede relacionar estas 2 variables y realizar una predicción que considere la

estacionalidad de los datos. Aunque si se requiere de una mayor cantidad de datos para asegurar una mejor calidad de la predicción.

Debido a la baja cantidad de datos para las ventas a industrias, la distancia temporal entre cada venta y que estas son realizadas a una sola empresa se consideró que no se necesita un modelo para predecir esa demanda, si en el futuro se agregan más empresas que requieran *pellet* a granel, se puede recurrir a un modelo predictivo que ayude a la empresa a estimar la demanda a las industrias, de momento no es un aporte predecir esta demanda industrial con un modelo de serie de tiempo.

En el futuro se espera que la demanda de los *pellets* vaya en aumento, y cada vez se esté agregando una mayor cantidad de clientes a la lista, lo que hará que la empresa necesite aumentar su producción y se requiera recurrir a modelos de análisis de datos para poder estimar la demanda, por el momento se necesita recaudar una mayor cantidad de estos datos para que los modelos tengan mayor efectividad y puedan predecir a mediano y largo plazo.

6 REFERENCIAS

- A. E. Márquez-Chamorro, M. Resinas and A. Ruiz-Cortés (2018). Predictive Monitoring of Business Processes: A Survey. *in IEEE Transactions on Services Computing*. vol. 11, no. 6, pp. 962-977. doi: 10.1109/TSC.2017.2772256.
- Altomonte, H. (2017). *Las energías renovables no convencionales en la matriz de generación eléctrica: tres estudios de caso*. Comisión económica para América Latina y el Caribe. Recuperado de: <http://hdl.handle.net/11362/40975>
- Cedeño Troya, F. & Carpio Torres, F. (2022). Modelos predictivos de sistemas de información en la gestión de abastecimientos del sector ferretero. *Revista Científica Ciencia Y Tecnología*, 22(34). doi:<https://doi.org/10.47189/rcct.v22i34.529>
- Chávez, J. y Saucedo, N. (2016). *Aplicación teórica de un modelo de análisis predictivo para desarrollar estrategias*. Revista Red Internacional de Investigadores en Competitividad. ISBN 978-607-96203-0-5.
- Eckerson, W. (2007). Predictive analytics. *Extending the Value of Your Data Warehousing Investment. TDWI Best Practices Report*, 1, 1-36. Recuperado de: <https://bit.ly/3leFcF8>
- Espino Timón, C. (2017). *Análisis predictivo: técnicas y modelos utilizados y aplicaciones del mismo - herramientas Open Source que permiten su uso*. Universitat Oberta de Catalunya. Recuperado de: <http://hdl.handle.net/10609/59565>
- F. Aqlan, J. C. Nwokeji and A. Shamsan (2020). Teaching an Introductory Data Analytics Course Using Microsoft Access and Excel. *IEEE Frontiers in Education Conference (FIE)*, pp. 1-10. doi: 10.1109/FIE44824.2020.9274247.
- García, J. G. (2002). Las Energías Renovables. *Revista de la Universidad del Azuay*, (29), 269-293. Recuperado de: <https://bit.ly/3s8lwp9>

- González, F. A. (2015). Modelos de aprendizaje computacional en reumatología. *Revista Colombiana de Reumatología*, 22(2), 77-78. doi: <https://doi.org/10.1016/j.rcreu.2015.06.001>
- Herrera, M., Díaz, J. C. G., García, R. P., & Izquierdo, J. *Métodos de remuestreo para la predicción de la demanda de agua*. Universitat Politècnica de València. Recuperado de: <https://bit.ly/37fzVJY>
- Instituto Libertad (2017). *Las energías renovables no convencionales en Chile*. Instituto Libertad Chile. Recuperado de: <https://bit.ly/3PtgUF5>
- Martínez, C. (Octubre 2017). *Excel Cartagena*. Obtenido de <https://excelcartagena.com/pronosticos-de-la-demanda/>
- Messer, N. (2013). *Políticas de fomento para las energías renovables no convencionales (ERNC)*. Pontificia Universidad Católica de Chile (Chile). Recuperado de: <https://bit.ly/3KFB0H3>
- Patiño Martínez, P. E. (2014). Biomasa residual vegetal: tecnologías de transformación y estado actual. *Innovaciencia Facultad De Ciencias Exactas Físicas Y Naturales*, 2(1), 45-52. doi: <https://doi.org/10.15649/2346075X.255>
- Peláez, I. M. (2016). Modelos de regresión: lineal simple y regresión logística. *Revista Seden*, 14, 195-214.
- Pinilla Suárez, J. C., Luengo Vergara, K., Navarrete, T., & Navarrete, F. (2020). *El pellet de madera en Chile su producción su uso y su mercado*. Instituto Forestal. Recuperado de: <https://bit.ly/3yd2RgX>
- Pradhan, P., Mahajani, S. M., & Arora, A. (2018). Production and utilization of fuel pellets from biomass: A review. *Fuel Processing Technology*, 181, 215-232. doi: <https://doi.org/10.1016/j.fuproc.2018.09.021>
- Rebolledo Canales, W. (2014). *Evaluación técnica-económica de la nueva Ley 20.698*. Universidad de Chile (Chile). Recuperado de: <https://bit.ly/3LLBoXD>

- Romero, J. C. (2017). *El mercado de energías renovables no convencionales (ERNC) en Chile: un sector de éxito, incertidumbres y futuros. Cuadernos de energía*, (52), 5-8. ISSN 1698-3009.
- Velázquez Martí, B. (2018). *Aprovechamiento de la biomasa para uso energético*. Editorial Universitat Politècnica de València. <http://hdl.handle.net/10251/113122>
- Villarreal, F. (2016). Introducción a los Modelos de Pronósticos. *Unión matemática Argentina-Universidad Nacional del Sur*, 1-121. Recuperado de: <https://bit.ly/3wCdr gb>
- Hyndman, R. J. (2014). *Measuring forecast accuracy*. Business forecasting: Practical problems and solutions, 177-183.
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*. (2nd ed.) OTexts. <https://otexts.org/fpp2/>

7 ANEXOS

Anexo 1(MAE promedio movil simple)

| fecha | x | y | x-y |
|------------|---------|---------|---------|
| 01-01-2021 | 270036 | | 270036 |
| 01-02-2021 | 632574 | | 632574 |
| 01-03-2021 | 831420 | 578010 | 253410 |
| 01-04-2021 | 1336788 | 933594 | 403194 |
| 01-05-2021 | 883224 | 1017144 | 133920 |
| 01-06-2021 | 1153062 | 1124358 | 28704 |
| 01-07-2021 | 2342160 | 1459482 | 882678 |
| 01-08-2021 | 1580562 | 1691928 | 111366 |
| 01-09-2021 | 828504 | 1583742 | 755238 |
| 01-10-2021 | 547200 | 985422 | 438222 |
| 01-11-2021 | 311580 | 562428 | 250848 |
| 01-12-2021 | 177822 | 345534 | 167712 |
| 01-01-2022 | 584964 | 358122 | 226842 |
| 01-02-2022 | 1189260 | 650682 | 538578 |
| 01-03-2022 | 3759102 | 1844442 | 1914660 |
| 01-04-2022 | 1913922 | 2287428 | 373506 |
| 01-05-2022 | 3133962 | 2935662 | 198300 |
| 01-06-2022 | 2394522 | 2480802 | 86280 |

| | |
|------------|---------|
| SAE | 6763458 |
| MAE | 422716 |

Anexo 2 (Rmse promedio movil simple)

| fecha | x | y | (X-Y)^2 |
|------------|---------|---------|-------------|
| 01-01-2021 | 270036 | | 72919441296 |
| 01-02-2021 | 632574 | | 4,0015E+11 |
| 01-03-2021 | 831420 | 578010 | 64216628100 |
| 01-04-2021 | 1336788 | 933594 | 1,62565E+11 |
| 01-05-2021 | 883224 | 1017144 | 17934566400 |
| 01-06-2021 | 1153062 | 1124358 | 823919616 |
| 01-07-2021 | 2342160 | 1459482 | 7,7912E+11 |
| 01-08-2021 | 1580562 | 1691928 | 12402385956 |
| 01-09-2021 | 828504 | 1583742 | 5,70384E+11 |
| 01-10-2021 | 547200 | 985422 | 1,92039E+11 |
| 01-11-2021 | 311580 | 562428 | 62924719104 |
| 01-12-2021 | 177822 | 345534 | 28127314944 |
| 01-01-2022 | 584964 | 358122 | 51457292964 |
| 01-02-2022 | 1189260 | 650682 | 2,90066E+11 |
| 01-03-2022 | 3759102 | 1844442 | 3,66592E+12 |
| 01-04-2022 | 1913922 | 2287428 | 1,39507E+11 |
| 01-05-2022 | 3133962 | 2935662 | 39322890000 |
| 01-06-2022 | 2394522 | 2480802 | 7444238400 |

defecto

RMSE

581390

normalizada

NRMSE

0,162341407

coeficiente de variación

CV(RMSE)

0,43840509

Anexo 3 (Mape promedio movil simple)

| fecha | x | y | $ x-y / y $ | $ y-x / x $ |
|------------|---------|---------|-------------|-------------|
| 01-01-2021 | 270036 | | | |
| 01-02-2021 | 632574 | | | |
| 01-03-2021 | 831420 | 578010 | 0,43841802 | 0,304791802 |
| 01-04-2021 | 1336788 | 933594 | 0,431872955 | 0,301614018 |
| 01-05-2021 | 883224 | 1017144 | 0,131662773 | 0,151626315 |
| 01-06-2021 | 1153062 | 1124358 | 0,025529235 | 0,024893718 |
| 01-07-2021 | 2342160 | 1459482 | 0,604788548 | 0,376864945 |
| 01-08-2021 | 1580562 | 1691928 | 0,06582195 | 0,070459748 |
| 01-09-2021 | 828504 | 1583742 | 0,476869339 | 0,911568321 |

| | | | | |
|------------|---------|---------|-------------|-------------|
| 01-10-2021 | 547200 | 985422 | 0,444704908 | 0,800844298 |
| 01-11-2021 | 311580 | 562428 | 0,446009089 | 0,805083767 |
| 01-12-2021 | 177822 | 345534 | 0,48537047 | 0,943145393 |
| 01-01-2022 | 584964 | 358122 | 0,633421013 | 0,387787966 |
| 01-02-2022 | 1189260 | 650682 | 0,827713076 | 0,45286817 |
| 01-03-2022 | 3759102 | 1844442 | 1,03807005 | 0,509339731 |
| 01-04-2022 | 1913922 | 2287428 | 0,163286451 | 0,195152154 |
| 01-05-2022 | 3133962 | 2935662 | 0,067548648 | 0,063274539 |
| 01-06-2022 | 2394522 | 2480802 | 0,034779075 | 0,036032244 |

| | | |
|------|-----|-----|
| MAPE | 39% | 40% |
|------|-----|-----|