



**FACULTAD DE POSTGRADO
TESIS DE POSTGRADO**

**IMPLEMENTACIÓN DE MODELOS PREDICTIVOS
PARA INCREMENTAR LAS VENTAS DE JUGOS
EN LEYDE S.A.**

SUSTENTADO POR:

SERGIO MOLINA ANDERSON

**PREVIA INVESTIDURA AL TÍTULO DE
MÁSTER EN
ANALÍTICA DE NEGOCIOS**

SAN PEDRO SULA, CORTÉS

HONDURAS, C.A.

ENERO, 2025

**UNIVERSIDAD TECNOLÓGICA CENTROAMERICANA
UNITEC**

FACULTAD DE POSTGRADO

AUTORIDADES UNIVERSITARIAS

**PRESIDENTE EJECUTIVO / RECTORA
ROSALPINA RODRÍGUEZ**

**SECRETARIO GENERAL /
PRORRECTOR
ROGER MARTÍNEZ MIRALDA**

**VICERRECTOR ACADÉMICO NACIONAL
JAVIER ABRAHAM SALGADO LEZAMA**

**DECANA DE POSTGRADO
ANA DEL CARMEN RETTALLY VARGAS
IMPLEMENTACIÓN DE MODELOS PREDICTIVOS
PARA INCREMENTAR LAS VENTAS DE JUGOS
EN LEYDE S.A.**

**TRABAJO PRESENTADO EN CUMPLIMIENTO DE LOS
REQUISITOS EXIGIDOS PARA OPTAR AL TÍTULO DE**

**MÁSTER EN
ANALÍTICA DE NEGOCIOS**

ASESOR METODOLÓGICO:

HENRY ANTONIO OSORTO RUIZ

ASESOR TEMÁTICO:

MARVIN ROBERTO MENDOZA VALENCIA

MIEMBROS DE LA TERNA:

GABRIELA GARAY

DANIEL LUNA

RICARDO RODRIGUEZ

DERECHOS DE AUTOR

© Copyright 2024
Sergio Molina Anderson

Todos los derechos son reservados



**FACULTAD DE POSTGRADO IMPLEMENTACIÓN
DE MODELOS PREDICTIVOS PARA
INCREMENTAR LAS VENTAS DE JUGOS EN
LEYDE S.A.**

**AUTOR:
SERGIO MOLINA ANDERSON**

Resumen

En esta investigación se presenta la implementación de un modelo basado en técnicas de análisis de datos para revisar el comportamiento de ventas de una de las empresas más reconocidas de Honduras, LEYDE SA, la cual cuenta con más de 50 años en el mercado nacional, especializándose en productos lácteos y jugos procesados. La empresa ha experimentado un declive en su marca de jugos a nivel nacional durante los últimos cuatro años. Por lo tanto, para revisar tendencias, estacionalidad, ciclos y ruido gris, utilizaremos un análisis de series de tiempo. Adicionalmente, a través de clusterización, buscamos identificar patrones de consumo en las diferentes sucursales y concluiremos pronosticando las ventas utilizando el algoritmo Random Forest. El objetivo principal es ayudar al equipo de ventas y marketing a mejorar sus KPI o indicadores y con ello devolver las marcas de jugos a sus niveles pico de ventas, ya que alguna vez fueron productos estrella a nivel nacional, especialmente el jugo de naranja. Esto ayudará a mejorar las estrategias de ventas y marketing dirigidas a los diversos tipos de clientes para retener a los clientes leales que han apoyado la promoción y venta de estos productos durante años. Los datos de ventas de los últimos 10 años se utilizarán para alimentar los modelos a través de algoritmos de aprendizaje automático, herramientas que hoy en día cobran cada vez más valor para las empresas.

Palabras claves: Aprendizaje Automático, Clusterización, Patrones de Consumo, Random Forest y Series de Tiempos.



GRADUATE SCHOOL
IMPLEMENTACIÓN DE MODELOS PREDICTIVOS
PARA INCREMENTAR LAS VENTAS DE JUGOS
EN LEYDE S.A.

AUTOR:
SERGIO MOLINA ANDERSON

Abstract

This research presents the implementation of a model based on data analysis techniques to review the sales behavior of one of the most recognized companies in Honduras, LEYDE S.A., which has been in the national market for over 50 years, specializing in dairy products and processed juices. The company has experienced a decline in its juice brand nationwide over the last four years. Therefore, to review trends, seasonality, cycles, and gray noise, we will use a time series analysis. Additionally, through clustering, we aim to identify consumption patterns across different branches, and we will conclude by forecasting sales using the Random Forest algorithm. The main objective is to assist the sales and marketing team in improving their KPIs or indicators and thereby restore the juice brands to their peak sales levels, as they were once star products nationwide, especially the orange juice. This will help improve sales and marketing strategies targeted at various customer types to retain loyal customers who have supported the promotion and sale of these products for years. Sales data from the last 10 years will be used to feed the models through machine learning algorithms, tools that are becoming increasingly valuable for companies today.

Palabras claves: Aprendizaje Automático, Clusterización, Patrones de Consumo, Random Forest y Series de Tiempos.

DEDICATORIA

Dedico A Dios, fuente inagotable de amor, sabiduría y fortaleza. Agradezco por ser mi guía en cada paso de este camino, por darme el discernimiento necesario en los momentos de duda y la paz en los tiempos de adversidad. Tu presencia ha sido el pilar sobre el cual construyó este logro, y cada logro mío es para Tu gloria.

A mi querida familia, que ha sido mi refugio y apoyo constante. Gracias por su amor incondicional, por las palabras de ánimo en los momentos difíciles, y por su fe en mí cuando más lo necesitaba. Este esfuerzo no sería posible sin ustedes, quienes me han inspirado a seguir adelante ya nunca rendirme. A ustedes les dedico este logro, como un pequeño tributo a su enorme sacrificio y amor.

“El mejor modo de predecir el futuro es inventándolo”

Alan Key

Sergio Molina Anderson

AGRADECIMIENTOS

Quiero expresar mi más profundo primeramente a Dios por darme el tiempo necesario para aprender de una maestría, segundo un agradecimiento a la Universidad Tecnológica Centroamericana (UNITEC) por brindarme la oportunidad de formarme académicamente y proporcionarme las herramientas necesarias para desarrollar este trabajo de Analítica de Datos. A los catedráticos, cuyo conocimiento, paciencia y dedicación fueron esenciales para mi crecimiento académico y profesional, mi más sincera gratitud; sus enseñanzas han sido fundamentales en cada paso de este camino. Finalmente, a mi familia, que con su apoyo incondicional, comprensión y amor, me motivaron a seguir adelante incluso en los momentos más difíciles. Sin ustedes, nada de esto hubiera sido posible.

Sergio Molina Anderson

ÍNDICE DE CONTENIDO

DEDICATORIA	vii
AGRADECIMIENTOS.....	viii
ÍNDICE DE CONTENIDO.....	ix
ÍNDICE DE FIGURAS.....	xiii
ÍNDICE DE TABLAS.....	xvi
CAPÍTULO I. PLANTEAMIENTO DE LA INVESTIGACIÓN.....	17
1.1 INTRODUCCIÓN	17
1.2 ANTECEDENTES DEL PROBLEMA.....	17
1.3 DEFINICIÓN DEL PROBLEMA.....	19
1.3.1 ENUNCIADO DEL PROBLEMA	19
1.3.2 FORMULACIÓN DEL PROBLEMA	21
1.3.3 PREGUNTAS DE INVESTIGACIÓN	22
1.4 OBJETIVOS DEL PROYECTO.....	23
1.4.1 OBJETIVO GENERAL	23
1.5 JUSTIFICACIÓN.....	23
CAPITULO II. MARCO TEÓRICO.....	24
2.1 ANÁLISIS DE LA SITUACIÓN ACTUAL.....	24
2.1.1 ANÁLISIS DEL MACROENTORNO.....	24
2.1.2 ANÁLISIS DEL MICROENTORNO	28
2.2 CONCEPTUALIZACIÓN	30
2.2.1 SERIES DE TIEMPOS	30
2.2.2 COSTOS DE PRODUCCION	30
2.2.3 ANÁLISIS EDA	31
2.2.4 ANÁLISIS RFM.....	32
2.2.5 BIG DATA.....	32
2.2.6 CLUSTERIZACIÓN.....	33

2.2.7 CRISP-DM	34
2.2.8 CRM	35
2.2.9 DATA MINING O MINERÍA DE DATOS	36
2.2.10 DIFERENCIAS ENTRE CLUSTERIZACIÓN Y CLASIFICACIÓN	37
2.2.11 ETL	38
2.2.12 INTELIGENCIA DE NEGOCIOS	39
2.2.13 K-MEANS	39
2.2.14 KNIME	41
2.2.15 MACHING LEARNING.....	42
2.2.16 PERFILAMIENTO DE CLIENTES	43
2.2.17 SEGMENTACIÓN DE CLIENTES.....	43
2.2.18 BOSQUES ALEATORIOS o RANDOM FOREST	44
2.3 TEORÍAS DE SUSTENTO	44
2.3.1 BASES TEÓRICAS	44
2.3.2 METODOLOGÍAS DESARROLLADAS.....	47
2.4 MARCO LEGAL	49
2.4.1 LA PROTECCIÓN DE DATOS EN HONDURAS	49
2.4.2 POLÍTICA DE CIBERSEGURIDAD EN LA EMPRESA LEYDE	49
CAPÍTULO III. METODOLOGIA DE LA INVESTIGACIÓN	51
3.1 CONGRUENCIA METODOLOGICA	51
3.1.1 MATRIZ METODOLÓGICA	51
3.1.2 ESQUEMA DE VARIABLES DE ESTUDIO	55
3.1.3. MATRIZ DE OPERALIZACION DE LAS VARIABLES	60
3.1.4 HIPÓTESIS DE LA INVESTIGACION.....	63
3.2 ENFOQUE Y METODOS	64
3.2.1. ESTADÍSTICA DESCRIPTIVA	65
3.2.2. ESTADÍSTICA INFERENCIAL.....	65
3.2.3 ESTADÍSTICA APLICADA A MODELOS PREDICTIVOS Y SERIES DE TIEMPO	66
3.2.4. EVALUACIÓN DE MODELOS (ESTADÍSTICA INFERENCIAL).....	66
3.2.5. VISUALIZACIÓN DE DATOS	67
3.3 DISEÑO DE LA INVESTIGACIÓN.....	67

3.3.1 POBLACIÓN	67
3.3.2 MUESTRA	68
3.3.3 TÉCNICAS DE MUESTREO	68
3.4 TÉCNICAS, INSTRUMENTOS Y PROCEDIMIENTOS APLICADOS	68
3.4.1 PROCEDIMIENTOS APLICADOS	69
3.5 FUENTES DE INFORMACIÓN	69
CAPÍTULO IV. RESULTADOS Y ANÁLISIS.....	72
4.1 EXPLORACION DE DATOS.....	72
4.2 RESULTADOS Y ANALISS DE LAS TECNICAS APLICADAS.....	76
4.2.1. APLICABILIDAD DE LA SEGMENTACIÓN POR CLÚSTER	76
4.2.2. APLICABILIDAD DEL ANÁLISIS DE LA SERIE DE TIEMPO Y RANDOM FOREST	77
4.2.3. APLICACIÓN DE ESTRATEGIAS APLICADAS POR CLÚSTER.....	77
4.2.4. APLICACIÓN DEL REBRANDING Y REPOSICIONAMIENTO.....	77
4.2.5. APLICACIÓN DE ESTRATEGIAS DE EXPANSIÓN O RESTRUCTURACIÓN	78
4.2.6. APLICACIÓN DE CONTROL DE INVENTARIOS	78
4.2.7. APLICACIÓN DE ESTRATEGIAS DIGITALES.....	78
4.2.8. APLICACIÓN EN LA PLANIFICACIÓN FINANCIERA.....	78
4.2.9. USO DE INDICADORES DE DESEMPEÑO BASADO EN CLUSTERES.....	79
4.2.10. APLICACIÓN DE ESTRATEGIAS REGIONALES Y ECONÓMICAS	79
4.3 RESULTADOS CUANTITATIVOS	81
4.3.1. ALGORITMO DE CLUSTERIZACION.....	83
4.3.2 ALGORITMO DE SERIES DE TIEMPOS	90
4.3.3 ALGORITMO DE RANDOM FOREST	96
4.4 RESULTADOS DESCRIPTIVOS DE LOS MODELOS	97
CAPÍTULO V. CONCLUSIONES Y RECOMENDACIONES	113
5.1 CONCLUSIONES	113

5.2 RECOMENDACIONES	115
CAPÍTULO VI. APLICABILIDAD	117
6.1 NOMBRE DE LA PROPUESTA.....	117
6.2 JUSTIFICACIÓN DE LA PROPUESTA.....	117
6.3 ALCANCE DE LA PROPUESTA	117
6.3.1 Objetivo General	117
6.3.2 Objetivos Específicos	117
6.4 DESCRIPCIÓN Y DESARROLLO A DETALLE DE LA PROPUESTA	118
6.4.1 DESCRIPCIÓN CLARA DEL QUE Y COMO SE HARÁN	118
6.4.2 DESARROLLO DE TODOS LOS ELEMENTOS NECESARIOS (HERRAMIENTAS, INSTRUMENTOS, PROCESOS)	119
6.5 MEDIDAS DE CONTROL (INDICADORES, MEDICIONES)	119
6.6 CRONOGRAMA DE IMPLEMENTACIÓN Y PRESUPUESTO	119
6.7 CRONOGRAMA DE IMPLEMENTACIÓN Y PRESUPUESTO	120
REFERENCIAS	122
GLOSARIO.....	124
ANEXOS.....	127

ÍNDICE DE FIGURAS

Figura 1 Diagrama de Modelo CRISP-DM	34
Figura 2. Se observa una tendencia aditiva en las series de tiempos y luego una caída de la tendencia a partir del 2021	69
Figura 3. Datos de las ventas en unidades de Jugos por parte de la empresa LEYDE S.A. de Honduras desde el 2014 a la fecha de Septiembre del 2024	70
Figura 4 Observamos el comportamiento de las 3 sucursales de mayor venta en el país de las 11 que se tienen por parte de la empresa LEYDE S.A.	71
Figura 5. Observamos el comportamiento de las 3 sucursales de mayor venta en el país de las 12 que se tienen por parte de la empresa LEYDE S.A. en miles de millones en un gráfico de Pastel.	71
Figura 6. Logotipo Power BI.	77
Figura 7. Logotipo Knime.	78
Figura 8. Workflow de Transformación y Preparación realizado en knime.	79
Figura 9. Uso de nodo de agrupamiento de datos previo al gráfico de Scatter Plot y el de Line Plot.	80
Figura 10. Revisión de los datos en un nodo de Scatter Plot y en un nodo de Line Plot.	80
Figura 11. Revisión de nodos que elaboran el agrupamiento por kMeans.	81
Figura 12. Revisión de nodos para revisar la parte descriptiva del algoritmo K-means.	83
Figura 13. Algoritmo de los Nodos para descomponer las Series de Tiempos de las Ventas.	86
Figura 14. Configuración del Nodo de Time Series Classical Descomposition.	88
Figura 15. Configuración del Nodo de Joiner.	90
Figura 16. Workflow para nuestro Bosque Aleatorio.	92
Figura 17. Diagrama de Knime para identificar patrones de ventas por medio de	

Kmedia o clusterización por sucursales en Honduras de la empresa LEYDE S.A. ...	93
Figura 18. Diagrama de pastel en el software Knime usando la herramienta Kmedia para segmentar en 4 clúster las sucursales de acuerdo a su Patrón de tendencia en las ventas.	94
Figura 19. Hoja de Excel donde graficamos los datos usando el Nodo de SimpleKmeans 3.7 de Knime buscando la SSE (Suma de los Cuadrados) desde el Clúster 2 al 10 para saber dónde se forma el codo y determinar las cantidades de clúster ideal para usar la herramienta.	95
Figura 20. Gráfico de Área bajo la curva para revisar la data de ventas de LEYDE en unidades de jugo de Naranja.	96
Figura 21. Diagrama de Series de tiempos de la herramienta Knime para evaluar la estacionalidad de las ventas en la empresa LEYDE S.A.	97
Figura 22. Gráfico de Polynomial Curve Fitting Trend	99
Figura 23. Gráficos que generamos usando el Nodo de Knime Time Series Classical Decomposition	100
Figura 24. Gráfico de Nuestra Serie Ajustada con un rezago o Fitted	101
Figura 25. Datos que nos genera el algoritmo de Knime para Nuestra Serie de Tiempo	101
Figura 26. Figura del grafico de Scatter Plot donde podemos ver si hay o no autocorrelación de los datos de las ventas en unidades de la empresa LEYDE S.A. en Honduras de Jugos Procesados desde el 2014 al 2024.	102
Figura 27. Observamos las diferentes Sucursales de LEYDE por Colores como referencia de nuestros Gráficos	103
Figura 28. Figura del gráfico de Box Plot o diagrama de caja donde podemos evaluar rango (sin datos atípicos), datos atípicos, rango Inter cuartil (también conocido como RIC), cuartiles (Q1, Q2 y Q3), mediana (Q2) y mínimo y máximo de los datos de las ventas.	103
Figura 29. Figura del Gráfico de cajas condicional (JavaScript) Un gráfico de cajas muestra parámetros estadísticos robustos: mínimo, cuartil inferior, mediana, cuartil	

superior y máximo.	104
Figura 30. Gráfico de un workflow de Knime para nuestra data de ventas de series de tiempo de la empresa LEYDE S.A. para revisar de forma descriptiva los datos y así poder interpretarlos.	105
Figura 31. Gráfico de un workflow de Knime para nuestra data de ventas de series de tiempo de la empresa LEYDE S.A. para revisar de forma descriptiva los datos y así poder interpretarlos.	106
Figura 32. Gráfico de la Media de los Factores estacionales y muestra cómo varía la media en función de los periodos estacionales de forma mensual donde las ventas tienden a ser más altas o más bajas en promedio.	106
Figura 33. Algoritmo de Random Forest para predecir en nuestra Data la Venta Futura con un 96% de Accuracy	107
Figura 34. Resultados del Predictor de Random Forest.	107
Figura 35. Fotografía de la Presentación del Jugo Coctel de Frutas de la Empresa LEYDE S.A.	121
Figura 36. Fotografía de la Presentación de los Jugo de diferentes Frutas de la Empresa LEYDE S.A.	122
Figura 37, Fotografía del Famoso Jugo de Naranja que más venta tiene	122
Figura 38. Fotografía del Jugo de Piña de la Empresa LEYDE S.A.	123
Figura 39. Nuevas presentaciones de Jugos como Naranja Agria y Jamaica de la Empresa LEYDE	123
Figura 40. FOTOGRAFIA DEL PLANTEL DE LEYDE SPS (SUCURSAL)	127

ÍNDICE DE TABLAS

Tabla 1. Resumen Corporativo -----	
37	
Tabla 2. Esquema Resumido de las Variables de Estudio -----	57
Tabla 3. Fases de la Investigación -----	64
Tabla 4. Plan de desarrollo basados en la metodología CRISP-DM -----	75
Tabla 5. Ejemplo de la Data de Ventas Históricas de LEYDE del producto filtrado Jugo de Naranja -----	
78	
Tabla 6. Cronograma de Implementación y Presupuesto -----	
114	
Tabla 7. Presupuesto Anual del Coordinador -----	
114	
Tabla 8. Matriz de Concordancia -----	115
Tabla 9. Benchmarking o Estudio de la Competencia -----	
124	

CAPÍTULO I. PLANTEAMIENTO DE LA INVESTIGACIÓN

1.1 INTRODUCCIÓN

La empresa LEYDE S.A., fundada en Honduras, es una de las principales productoras y distribuidoras de productos lácteos en el país, destacándose por su compromiso con la calidad y el desarrollo del sector agro industrial. Desde su creación, LEYDE ha evolucionado como una empresa líder en el mercado, expandiendo su portafolio de productos y consolidando una red de distribución que abarca gran parte del territorio hondureño. Su historia está marcada por la constante innovación en la producción y procesamiento de lácteos y jugos envasados adoptando tecnologías de punta que le han permitido mantenerse competitiva en un entorno cambiante.

En el presente planteamiento del problema, vamos a revisar la Data Histórica de ventas de los últimos diez años de la Empresa LEYDE con una trayectoria en Honduras por más de 50 años desde sus inicios en el Litoral Atlántico en la ciudad de Ceiba con capital hondureño. Hace 5 años todavía la familia de Jugos en LEYDE era una de las pioneras en el mercado; era una de los productos estrella de la compañía, pero ha venido decreciendo y perdiendo mercado por tanto este enfoque del problema está basado en esa premisa de mejorar con herramientas de machine learning, pronósticos y causales que nos puedan dar el horizonte para mejorar estos indicadores o Kpi's de ventas. Este análisis resulta fundamental en un contexto empresarial donde la toma de decisiones basada en datos se ha vuelto un pilar para el éxito y la sostenibilidad a largo plazo.

1.2 ANTECEDENTES DEL PROBLEMA

Como antecedente del problema central que tiene LEYDE S.A. con las bajas de las ventas en su familia de Jugos Procesados quiero a continuación hablar de los problemas que hemos tenido como contexto en una empresa familiar que fundamos en el 2009 donde decidimos a raíz que teníamos ganadería de leche y vendíamos lo que producíamos a un Centro Recolector de Leche ubicado en el Municipio de Victoria en Yoro; darle valor agregado a la misma y para lo cual industrializamos el proceso de Pasteurización, Descremado, Cuajado y Prensado para Elaborar Quesos Duros, Semi Duros y Blandos; la empresa se llama Lácteos Victoria S. de R.L. de C.V. (LACTOVI), lo que nos llevó a producir, buscar la logística o canales de distribución de nuestros productos, diversificar nuestra cartera de productos con la finalidad de encontrar el punto de equilibrio de que productos eran más rentable dentro de la demanda, bajar costos operativos de forma constante, luchar contra cuentas por cobrar con supermercados que iban en quiebra como en

el caso de Las Antorchas que nos debían más de 10 millones de Lempiras en Cuentas x Cobrar y a todo esto analizo que sin las herramientas adecuadas para análisis de datos no podremos en la actualidad seguir tomando buenas decisiones como eliminar del mercado un producto que no se logra vender, además canalizar las cantidades correctas por centros de distribución, a que clientes si entregarles el pedido y a que clientes no conforme ellos solicitan por el tema de las devoluciones constantes, así concluyo que esta Maestría de Analítica de Datos nos da esas herramientas de pronóstico y de algoritmos con inteligencia artificial para ayudarse y evitar cometer errores que al final es dinero en pérdidas y buscar la mayor rentabilidad en cada industria que se aplique.

En el contexto de la empresa LEYDE que tiene una trayectoria de 50 años produciendo y comercializando productos lácteos y jugos de sabores envasados con una amplia variedad, surge la necesidad de analizar y mejorar las estrategias de ventas para maximizar la rentabilidad y la satisfacción del cliente. Usando las herramientas de análisis de datos vamos a revisar la situación actual de su cartera de productos de jugos envasados, la empresa enfrenta desafíos en cuanto a la optimización de sus ventas debido a la falta de comprensión de las tendencias del mercado y las preferencias de los consumidores. Este problema se manifiesta en varios aspectos clave:

- I. Variedad de Productos y Dificultad para Prever la Demanda: La empresa ofrece una amplia gama de jugos con diferentes sabores, lo que dificulta prever con precisión la demanda de cada producto. Esta falta de visibilidad sobre las tendencias de consumo conduce a problemas de inventario o de venta perdida, falta de materia primas como el concentrado de jugos debido a que son frutas de temporada, desperdicio de producto y pérdida de oportunidades de venta en todo el país de Honduras.
- II. Competencia en un Mercado Saturado: El mercado de bebidas, incluidos los jugos, es altamente competitivo ya que se han venido desarrollado a través de emprendedores tanto de empresas solidadas en el mercado formal e informal con numerosas marcas y opciones disponibles para los consumidores. La empresa necesita identificar estrategias efectivas para diferenciarse de la competencia y destacar sus productos en un entorno saturado apoyándose de una estadística que nos oriente el costo de Marketing versus las Ventas por año.
- III. Cambios en las Preferencias del Consumidor: Las preferencias y hábitos de consumo de los clientes están en constante evolución ya que las personas anhelan mejor salud; lo que hace que sea crucial para la empresa mantenerse al tanto de las tendencias

emergentes y adaptar su oferta de productos en consecuencia. La falta de comprensión sobre qué sabores son más populares en diferentes segmentos de mercado dificulta la capacidad de la empresa para satisfacer las demandas cambiantes de los consumidores como jugos de sabores como naranja, toronja, piña, limón, frutas varias, zanahoria. Para ellos se pueden observar sabores que actualmente no se desarrollan como sábila, naranja agria, mango, fresa, sandía, entre otros.

- IV. Eficiencia en la Cadena de Suministro y Distribución: La falta de datos precisos sobre las tendencias de ventas también afecta la eficiencia de la cadena de suministro y distribución para lo cual se mandan a 11 centros de distribución en todo el país la producción del día a día. La empresa necesita optimizar sus procesos logísticos para garantizar la disponibilidad oportuna de productos basado en datos de proyección de demanda en los centros de distribución adecuados y evitar costos innecesarios asociados con el almacenamiento y transporte excesivos.
- V. Rentabilidad y Sostenibilidad del Negocio: Mejorar las ventas no solo se trata de aumentar los ingresos, sino también de garantizar la rentabilidad a largo plazo y la sostenibilidad del negocio ya que se debe revisar constantemente el costo de la materia prima de los concentrados de jugos ya que son cambiantes en todo el año debido a la escases de la misma en ciertos meses. La empresa debe identificar oportunidades para impulsar las ventas de productos con márgenes de beneficio más altos y minimizar el impacto ambiental de su operación, especialmente en términos de empaque y desperdicio de producto.

El problema que enfrenta la empresa se centra en la necesidad de comprender y anticipar las tendencias del mercado de jugos para mejorar la eficiencia operativa, la rentabilidad y la satisfacción del cliente. Identificar los sabores más populares, diferenciarse de la competencia, adaptarse a los cambios en las preferencias del consumidor y optimizar la cadena de suministro son aspectos clave que deben abordarse para alcanzar el éxito en este mercado altamente competitivo.

1.3 DEFINICIÓN DEL PROBLEMA

1.3.1 ENUNCIADO DEL PROBLEMA

En el contexto la industria de bebidas en Honduras ha tenido un mercado potencial a través de los años con jugos de consumo de diferentes sabores, se puede definir como un sector altamente dinámico y competitivo. Con el crecimiento del interés por un estilo de

vida saludable y la creciente conciencia sobre los beneficios de consumir productos naturales para la salud, los jugos han ganado popularidad entre los consumidores de todas las edades. Sin embargo, esta popularidad también ha llevado a un aumento en la competencia, con numerosas marcas y variedades de jugos disponibles en el mercado como se puede observar en los Anexos de marcas que compiten con diferentes sabores, presentaciones, precios, etc. En este contexto, una empresa que se especializa en la venta de jugos de diferentes sabores enfrenta desafíos específicos relacionados con la optimización de sus ventas y para lo cual la empresa LEYDE maneja una Data de más de diez años donde se ha podido observar el comportamiento de las ventas por mes, temporadas y en los distintos 18 departamentos del país. La amplia gama de sabores que ofrece puede ser tanto una ventaja como un obstáculo; por los momentos se manejan como producto estrella la venta de Jugo de Naranja, pero se han venido mostrando tendencias más marcadas en otros sabores. Por un lado, proporciona opciones para satisfacer una variedad de gustos y preferencias de los consumidores que eso implica un reto en la toma de decisiones y en el área de desarrollo de nuevos productos. Por otro lado, la gestión eficaz de esta diversidad de productos puede resultar complicada, especialmente al tratar de prever la demanda y evitar el desperdicio de inventario o la venta perdida por no tener inventarios en esas temporadas pico como el verano seco y marcado de los meses de marzo y abril para conseguir zumos de frutas que se escasea en Honduras.

Además, el mercado de jugos está en constante evolución, con cambios en las preferencias del consumidor y el surgimiento de nuevas tendencias. Por ejemplo, mientras que ciertos sabores pueden ser populares en un momento dado, las preferencias del consumidor pueden cambiar rápidamente debido a factores como la temporada como se mencionó, las tendencias de salud de consumir menos azúcar para evitar enfermedades como la Diabetes o las campañas de marketing de la competencia. Esta volatilidad en las preferencias del consumidor hace que sea crucial para la empresa mantenerse ágil y adaptable para mantener su posición en el mercado.

La eficiencia en la cadena de suministro y distribución también es fundamental en este contexto. La capacidad de la empresa para asegurar una disponibilidad constante de productos en los puntos de venta adecuados, mientras se minimizan los costos asociados con el almacenamiento y el transporte, es esencial para garantizar la satisfacción del cliente y la rentabilidad del negocio. Actualmente se cuentan con 11 centros de distribución con las rutas de detalle, supermercados e institucional y se maneja una preventa para evitar andar producto en rutas sin la venta asegurada.

En resumen, la empresa se encuentra en un entorno desafiante donde debe equilibrar la diversidad de su oferta de productos con la necesidad de prever y adaptarse a las cambiantes tendencias del mercado para lo cual debemos revisar la Data y sacar indicadores clave para mejorar las ventas de esa Familia en específico dentro de la empresa LEYDE; mejorar las ventas de sus jugos implica entender y anticipar estas tendencias, diferenciarse de la competencia, optimizar la cadena de suministro y garantizar una experiencia satisfactoria para el cliente.

Revisión de Variables de interés:

- I. Proyección de ventas para el siguiente año por Sabores.
- II. Revisión de Costos generales que afectan el precio a través del año de producción de cada fruta.
- III. De forma Demográfica revisar las tendencias y gustos por localidades a través de los 11 centros de Distribución en el país.
- IV. Identificar el crecimiento de la Competencia a través de datos específicos como preferencias de encuestas en puntos de ventas como Supermercados.
- V. Revisar las tendencias de sabores para readecuar la capacidad instalada de la empresa para sabores y volúmenes.
- VI. Revisar precios de Mercados en bases de Datos como la FHIA/SIMPAH para ver comportamiento de precios versus a la producción local de frutas de temporada. actual de la industria del consumo masivo, la cual se caracteriza por una intensa competencia, las empresas enfrentan el desafío de personalizar sus servicios y ofertas para satisfacer las necesidades de un cliente cada vez más informado y exigente. La capacidad de una empresa para identificar y comprender las diversas necesidades y preferencias de sus clientes es fundamental para diseñar estrategias de marketing efectivas y mejorar la retención de clientes.

1.3.2 FORMULACIÓN DEL PROBLEMA

El principal problema que enfrenta LEYDE es el decrecimiento de las ventas en su línea de jugos durante los últimos 4 años. Este problema ha afectado no solo los ingresos de la empresa, sino también su capacidad para competir en el mercado de jugos y bebidas. A pesar de los esfuerzos tradicionales de marketing y ventas, no se ha logrado un cambio

positivo, lo que evidencia la necesidad de implementar enfoques más innovadores y basados en datos. La naturaleza competitiva del mercado actual y la gran cantidad de datos generados diariamente en la empresa sugieren que las estrategias tradicionales pueden no ser suficientes para abordar este problema. El uso de técnicas de Machine Learning permite aprovechar grandes volúmenes de datos históricos para encontrar patrones, realizar predicciones y optimizar la toma de decisiones estratégicas en áreas como ventas, marketing y fidelización de clientes.

LEYDE posee una vasta cantidad de datos históricos de ventas, segmentación de clientes, preferencias de consumo y campañas de marketing. La explotación de esta información mediante algoritmos de aprendizaje automático puede proporcionar “insights” valiosos sobre los factores que afectan las ventas, así como mejorar la segmentación de clientes y predecir comportamientos de compra futuros. En este contexto, el uso de Machine Learning no solo tiene el potencial de revertir la tendencia negativa en las ventas, sino también de diseñar estrategias más eficientes y personalizadas para los diferentes tipos de consumidores.

1.3.3 PREGUNTAS DE INVESTIGACIÓN

Pregunta general

¿Puede un modelo de series de tiempos y clusterización ayudar al departamento de ventas de LEYDE a pronosticar, identificar y segmentar eficientemente los perfiles de sus clientes para mejorar los indicadores de la familia de jugos envasados específicamente dentro de su cartera de productos?

Preguntas específicas

- I. ¿Cuáles son los principales factores que han influido en el decrecimiento de las ventas de la línea de jugos de LEYDE S.A. en los últimos 4 años?
- II. ¿Cómo pueden las técnicas de series de tiempo ayudar a prever las ventas futuras y anticipar cambios en la demanda de los productos de jugos?
- III. ¿Qué patrones de comportamiento de los clientes pueden ser identificados mediante técnicas de clustering y cómo se pueden utilizar para segmentar el mercado?
- IV. ¿Qué estrategias de marketing y ventas pueden ser implementadas con base en los resultados obtenidos a partir de los modelos de Machine Learning para aumentar la fidelización y mejorar los indicadores de ventas?

1.4 OBJETIVOS DEL PROYECTO

1.4.1 OBJETIVO GENERAL

Aplicar técnicas de Machine Learning para analizar y predecir las ventas de la línea de jugos de LEYDE S.A., con el fin de optimizar los indicadores de ventas y desarrollar estrategias de marketing y fidelización más efectivas.

1.4.2 OBJETIVOS ESPECÍFICOS

1. Analizar los datos históricos de ventas de la línea de jugos de los últimos 10 años para identificar patrones, tendencias y factores asociados al decrecimiento de ventas.
2. Utilizar técnicas de series de tiempo para pronosticar las ventas futuras de la línea de jugos y anticipar posibles fluctuaciones o tendencias.
3. Implementar algoritmos de clasificación o clusterización mediante K-Means, para segmentar a los clientes según su comportamiento de compra.
4. Desarrollar recomendaciones para optimizar las estrategias de marketing y ventas, basadas en los resultados del análisis de datos y las predicciones generadas por los modelos de Machine Learning.

1.5 JUSTIFICACIÓN

Este proyecto o tesis añadirá una evaluación de la situación actual de la empresa LEYDE y ayudará a la toma de decisiones buscando incrementar los márgenes de ganancias, volviendo a poner la marca en la familia de Jugos con mejores indicadores de ventas, así como una mayor participación en el mercado de Honduras. Usaremos en este proyecto de investigación la implementación de un modelo basándose en algoritmos de pronósticos de series de tiempos y de clusterización para la segmentación de la cartera actual de clientes y buscar la fidelidad de los clientes actuales.

La implementación de los modelos de series de tiempos y de clusterización proporcionará a la empresa “insights” analíticos valiosos que facilitarán la toma de decisiones, adoptando técnicas avanzadas de segmentación que pueden ayudarle a posicionarse como líderes innovadores en su industria.

CAPITULO II. MARCO TEÓRICO

2.1 ANÁLISIS DE LA SITUACIÓN ACTUAL

2.1.1 ANÁLISIS DEL MACROENTORNO

¿Para hablar del Macroentorno debemos primero definir Que es Big Data?; y nos referimos al conjunto de datos o combinaciones de conjuntos de datos cuyo tamaño (volumen), complejidad (variabilidad) y velocidad de crecimiento (velocidad) dificultan su captura, gestión, procesamiento o análisis mediante tecnologías y herramientas convencionales, tales como bases de datos relacionales y estadísticas convencionales o paquetes de visualización, dentro del tiempo necesario para que sean útiles. Aunque el tamaño utilizado para determinar si un conjunto de datos determinado se considera Big Data no está firmemente definido y sigue cambiando con el tiempo, la mayoría de los analistas y profesionales actualmente se refieren a conjuntos de datos que van desde 30-50 Terabytes a varios Petabytes.

“La naturaleza compleja del Big Data se debe principalmente a la naturaleza no estructurada de gran parte de los datos generados por las tecnologías modernas” (Brynjolfsson, 2014, p. 87). Como ejemplos están los web logs, la identificación por radiofrecuencia (RFID), los sensores incorporados en dispositivos, la maquinaria, los vehículos, las búsquedas en Internet, las redes sociales como Facebook, computadoras portátiles, teléfonos inteligentes y otros teléfonos móviles, dispositivos GPS y registros de centros de llamadas.

Lo que hace que Big Data sea tan útil para muchas empresas es el hecho de que proporciona respuestas a muchas preguntas que las empresas ni siquiera sabían que tenían. En otras palabras, proporciona un punto de referencia. Con una cantidad tan grande de información, los datos pueden ser moldeados o probados de cualquier manera que la

empresa considere adecuada. Al hacerlo, las organizaciones son capaces de identificar los problemas de una forma más comprensible.

La recopilación de grandes cantidades de datos y la búsqueda de tendencias dentro de los datos permiten que las empresas se muevan mucho más rápidamente, sin problemas y de manera eficiente. También les permite eliminar las áreas problemáticas antes de que los problemas acaben con sus beneficios o su reputación.

El análisis de Big Data ayuda a las organizaciones a aprovechar sus datos y utilizarlos para identificar nuevas oportunidades. Eso, a su vez, conduce a movimientos de negocios más inteligentes, operaciones más eficientes, mayores ganancias y clientes más felices. Las empresas con más éxito con Big Data consiguen valor de las siguientes formas:

1. Reducción de coste: Las grandes tecnologías de datos, como Hadoop que es un software bajo licencia libre y el análisis basado en la nube, aportan importantes ventajas en términos de costes cuando se trata de almacenar grandes cantidades de datos, además de identificar maneras más eficientes de hacer negocios. Con la velocidad de Hadoop y la analítica en memoria, combinada con la capacidad de analizar nuevas fuentes de datos, las empresas pueden analizar la información inmediatamente y tomar decisiones basadas en lo que han aprendido.
2. Nuevos productos y servicios: Con la capacidad de medir las necesidades de los clientes y la satisfacción a través de análisis viene el poder de dar a los clientes lo que quieren. Con la analítica de Big Data, más empresas están creando nuevos productos para satisfacer las necesidades de los clientes. En la mayoría de los casos, con el fin de utilizar eficazmente el Big Data, debe combinarse con datos estructurados (normalmente de una base de datos relacional) de una aplicación comercial más convencional, como un ERP (Enterprise Resource Planning) o un CRM (Customer Relationship Management).

“El macroentorno del Big Data abarca una serie de factores globales y externos que influyen en su desarrollo y expansión. Estos factores incluyen aspectos políticos, económicos, sociales, tecnológicos, ecológicos y legales (PESTEL)”. (Brynjolfsson, 2014, p. 56)

2.1.1.1 EJEMPLOS EMPRESAS APLICANDO CLUSTERIZACIÓN

A continuación, daremos Ejemplos de 15 grandes empresas que han usado

Clusterización y que son empresas que han cruzado fronteras en muchos países:

i. Amazon (Optimización de Inventarios)

“Amazon usa técnicas avanzadas de clustering en su plataforma de comercio electrónico para gestionar inventarios en múltiples centros de distribución. Los algoritmos de clustering agrupan productos por características de demanda, rotación y ubicación geográfica.” (Hu, 2020, p. 7)

ii. Toyota (Control de Calidad)

“Toyota utiliza el clustering para el análisis de defectos en su línea de producción de automóviles. Los datos de defectos se agrupan en categorías similares, lo que permite identificar patrones y causas raíz más rápidamente, mejorando el control de calidad”. (Mahfoud, 2019, p. 57)

iii. Walmart (Análisis de Producción y Logística)

“Walmart utiliza clustering para agrupar datos sobre patrones de ventas y optimizar la distribución de productos a sus tiendas. Los algoritmos permiten determinar las necesidades de inventario en función de las tendencias de compra en diferentes regiones”. (Wan, 2021, p. 176)

iv. General Electric (Mantenimiento Predictivo)

“GE utiliza técnicas de clustering en el mantenimiento predictivo de maquinaria pesada, agrupando máquinas con patrones de desgaste y fallos similares. Esto ayuda a predecir cuándo es necesario el mantenimiento antes de que ocurra una falla crítica”. (Yan, 2020, p. 16)

v. Nestlé (Optimización de la Cadena de Suministro)

“Nestlé aplica algoritmos de clustering para agrupar plantas de producción y centros de distribución según la demanda regional, mejorando la eficiencia en la cadena de suministro y reduciendo costos logísticos”. (Garcia, 2022, p. 15)

vi. Ford (Optimización de Producción y Proveedores)

“Ford usa clustering para agrupar proveedores según la calidad, tiempos de entrega y costos, optimizando la selección de proveedores en sus procesos de fabricación. También agrupan las diferentes plantas según eficiencia y rendimiento, mejorando la gestión de recursos y capacidades”. (Rakesh, 2018, p. 29)

vii. Zara (Gestión de Inventarios y Producción Just-in-Time)

La empresa de moda Zara utiliza el clustering de datos para optimizar el inventario en sus tiendas físicas y centros de distribución, basado en las preferencias locales de los clientes y la rotación de productos. También agrupa las tendencias de moda en diferentes regiones para producir en función de la demanda. (Silva, 2019, p. 29)

viii. Coca-Cola (Análisis de Preferencias del Consumidor y Segmentación de Mercado)

“Coca-Cola utiliza técnicas de clustering para segmentar a sus consumidores en diferentes grupos basados en preferencias de productos, comportamientos de compra y ubicaciones geográficas. Esta segmentación ayuda en campañas de marketing más efectivas y en la personalización de productos”. (Chaudhary, 2021, p. 19)

ix. Airbnb (Segmentación de Clientes y Análisis de Precios)

“Airbnb emplea algoritmos de clustering para agrupar propiedades y usuarios en función de patrones de comportamiento, preferencias de ubicación y niveles de gasto. Esto ayuda a personalizar la oferta de alojamientos y optimizar precios dinámicos”. (Nicolau, 2019, p. 32)

x. IKEA (Optimización del Inventario y Logística)

“IKEA utiliza clustering para gestionar el inventario en sus almacenes y tiendas, agrupando productos por similitudes en demanda, volumen de ventas y características logísticas. Esto ayuda a optimizar el almacenamiento y mejorar la eficiencia en el transporte de mercancías”. (Lindberg & Hedman, 2020, p. 44)

xi. Apple (Segmentación de Clientes para Desarrollo de Producto)

“Apple emplea clustering para identificar patrones de comportamiento entre diferentes grupos de usuarios de sus productos, lo que le permite desarrollar nuevos productos o mejorar los existentes en función de las preferencias del cliente”. (Johnson &

Chen, 2022, p. 39)

xii. Pfizer (Investigación y Desarrollo Farmacéutico)

“Pfizer utiliza clustering en su área de I+D para agrupar datos genómicos y farmacológicos, facilitando la identificación de patrones y correlaciones entre pacientes con características similares y el efecto de los medicamentos”. (Tang & Wang, 2020, p. 15)

xiii. Nike (Análisis de Tendencias del Mercado y Producción Personalizada)

“Nike utiliza clustering para agrupar a sus clientes en segmentos basados en patrones de compra, preferencias deportivas y comportamiento digital. Esto le permite adaptar su oferta de productos y marketing a las tendencias locales”. (Gomez & Lee, 2021, p. 49)

xiv. Tesla (Control de Calidad y Análisis de Rendimiento de Baterías)

“Tesla aplica clustering para agrupar datos de rendimiento de baterías en sus vehículos eléctricos, detectando patrones de fallas o anomalías para mejorar el control de calidad y la eficiencia en el diseño de nuevos modelos”. (Smith & Martinez, 2021, p. 58)

xv. Unilever (Análisis de la Cadena de Suministro y Demanda Regional)

“Unilever utiliza clustering para agrupar sus centros de distribución y fábricas según patrones de demanda, ayudando a optimizar el abastecimiento de materias primas y la producción en función de las necesidades regionales”. (Patel & Fernandes, 2020, p. 27)

Con estos ejemplos podemos observar la importancia de aplicar los algoritmos de Clustering y Series de Tiempo en la data de la empresa y estaremos a la vanguardia de nuestro análisis de datos.

2.1.2 ANÁLISIS DEL MICROENTORNO

El microentorno de los datos se refiere a los factores internos y cercanos que influyen directamente en cómo una empresa gestiona y utiliza sus datos. A continuación, daremos ejemplos de 5 empresas que aprovechan los elementos de su microentorno para optimizar la toma de decisiones y mejorar su competitividad. (Emprendedores, 2023)

a. Amazon

Amazon es un ejemplo destacado de cómo las empresas utilizan su microentorno de datos para ofrecer servicios personalizados. El gigante del comercio electrónico aprovecha los datos generados por sus clientes, como el historial de compras, búsquedas y preferencias, para ofrecer recomendaciones personalizadas, optimizar la logística y mejorar la experiencia del cliente. Además, Amazon utiliza los datos para gestionar su cadena de suministro, predecir la demanda de productos y ajustar su inventario en tiempo real, logrando una mayor eficiencia operativa.

b. Netflix

Netflix utiliza los datos generados por los usuarios (microentorno) para perfeccionar su sistema de recomendaciones. Analiza el historial de visualización, las calificaciones y el comportamiento de sus suscriptores para ofrecer contenido que les interese. Esta personalización aumenta la retención de usuarios y mejora la experiencia del cliente. Además, Netflix utiliza estos datos para tomar decisiones estratégicas sobre la producción de contenido original, como identificar géneros o tendencias.

c. Zara (Inditex)

Zara, perteneciente al grupo Inditex, es otro gran ejemplo de cómo las empresas pueden utilizar el microentorno de datos para optimizar la gestión de su negocio. Zara recopila datos de ventas diarias, comentarios de los clientes y tendencias emergentes en sus tiendas y plataformas digitales para ajustar rápidamente su oferta de productos. Gracias a estos datos, la empresa puede lanzar nuevas colecciones y ajustar su inventario de manera ágil, logrando reducir los tiempos de producción.

d. Google

Google se basa en una vasta cantidad de datos generados por los usuarios para optimizar sus servicios, como las búsquedas, los anuncios personalizados y Google Maps. A través del análisis de datos del comportamiento de los usuarios, Google puede ofrecer publicidad extremadamente dirigida y relevante. Además, la empresa utiliza el feedback constante de sus usuarios para mejorar sus algoritmos de búsqueda y desarrollar nuevos productos.

e. Uber

Uber utiliza su microentorno de datos, que incluye datos en tiempo real sobre la ubicación de los conductores y pasajeros, para optimizar sus operaciones. A través del análisis de estos datos, Uber ajusta los precios dinámicamente, calcula el tiempo estimado de llegada y mejora la eficiencia de su red. También aprovecha estos datos para entender patrones de demanda.

2.2 CONCEPTUALIZACIÓN

2.2.1 SERIES DE TIEMPOS

“Las series de tiempo son una herramienta gráfica ampliamente utilizada y de gran utilidad en el análisis de datos en el ámbito empresarial”. (Chatfield, 2003, p. 2) Se representan mediante una gráfica de líneas donde el eje vertical muestra la escala del métrico de interés, como defectos, costos, dimensiones de piezas, quejas de clientes, entre otros. Por otro lado, el eje horizontal representa el tiempo o intervalos en los que se recopilaron los datos. A través de esta representación gráfica, se pueden identificar diversas características y patrones en los datos.

Las series de tiempo nos permiten identificar tendencias, ciclos, estacionalidad y cambios de comportamiento en nuestros datos y procesos. Por ejemplo, al observar la distribución de quejas de clientes en una empresa durante los meses de enero a diciembre, podemos notar que febrero es el mes con más quejas, mientras que abril tiene menos. Esto nos lleva a cuestionarnos qué sucedió en febrero que lo diferencia de abril y cuál podría ser la causa de este comportamiento.

2.2.2 COSTOS DE PRODUCCION

Los costos de producción (también llamados costos de operación) son los gastos necesarios para mantener un proyecto, línea de procesamiento o un equipo en funcionamiento. En una compañía estándar, la diferencia entre el ingreso (por ventas y otras entradas) y el costo de producción indica el beneficio bruto. (Drury, 2018)

Esto significa que el destino económico de una empresa como LEYDE está asociado con: el ingreso (por ej., los bienes vendidos en el mercado y el precio obtenido) y el costo de producción de los bienes vendidos. Mientras que el ingreso, particularmente el ingreso

por ventas, está asociado al sector de comercialización de la empresa, el costo de producción está estrechamente relacionado con el sector tecnológico; en consecuencia, es esencial que la administración conozca de costos de producción.

El costo de producción tiene dos características opuestas, que algunas veces no están bien entendidas en los países en vías de desarrollo. La primera es que para producir bienes uno debe gastar; esto significa generar un costo. La segunda característica es que los costos deberían ser mantenidos tan bajos como sea posible y eliminados los innecesarios.

Esto no significa el corte o la eliminación de los costos indiscriminadamente. Cuando se analiza la importancia dada al costo de producción en los países en vías de desarrollo, otro aspecto que debería ser examinado respecto a una determinada estructura de costos, es que una variación en el precio de venta tendrá un impacto inmediato sobre el beneficio bruto porque éste último es el balance entre el ingreso (principalmente por ventas) y el costo de producción.

“En consecuencia, los incrementos o las variaciones en el precio de venta, con frecuencia son percibidos como la variable más importante (junto con el costo de la materia prima), particularmente cuando existen amplias variaciones del precio”. (LA FAO org, n.d.)

2.2.3 ANÁLISIS EDA

Los científicos de datos utilizan el “análisis de datos exploratorios” (EDA) para analizar e investigar conjuntos de datos y resumir sus características principales, a menudo empleando métodos de visualización de datos. EDA ayuda a determinar la mejor manera de manipular las fuentes de datos para obtener las respuestas que necesita, lo que facilita a los científicos de datos descubrir patrones, detectar anomalías, poner a prueba una hipótesis o verificar suposiciones.

“EDA se utiliza principalmente para ver qué datos se pueden revelar más allá de la tarea formal de modelado o de la prueba de hipótesis y proporciona una mejor comprensión de las variables del conjunto de datos y las relaciones entre ellas”. (Tukey, 1977, p. 43)

También puede ayudar a determinar si las técnicas estadísticas que usted está considerando para el análisis de datos son adecuadas. Desarrolladas originalmente por el matemático estadounidense John Tukey en la década de 1970, las técnicas EDA siguen siendo un método ampliamente utilizado en el proceso de descubrimiento de datos en la actualidad.

2.2.4 ANÁLISIS RFM

“Es mucho más sencillo y barato fidelizar un cliente que captar uno nuevo”. (Kotler & Keller, 2016, p. 21). Gracias al análisis RFM, el ecommerce ve aumentar sus ingresos, utilizando los datos de los clientes para ofrecerles lo que necesitan en cada momento. Será muy complicado que esos usuarios abandonen la marca, cuando se les está tratando de manera personal.

El modelo RFM, cuyas siglas corresponden a los términos ingleses Recency, Frequency, Monetary, es una herramienta empleada en marketing que, bien usada, permite incrementar el número de ventas en ecommerce.

Se trata de una técnica basada en la segmentación de clientes, analizando su comportamiento: ¿Cuándo fue la última vez que realizó una compra? ¿Con qué frecuencia lo hace? ¿Cuánto se gasta? Con esa información es más sencillo elaborar una estrategia de venta, llegando incluso a personalizar el mensaje hacia esos clientes.

“El RFM permite entender mejor cómo funciona la base de clientes, conociendo su comportamiento para nuestra empresa. Cuanto más conozcamos a nuestros clientes, más eficaz será nuestro mensaje, mejores resultados obtendremos en la venta”. (Hugues, 1994)

2.2.5 BIG DATA

Cuando hablamos de Big Data nos referimos a conjuntos de datos o combinaciones de conjuntos de datos cuyo tamaño (volumen), complejidad (variabilidad) y velocidad de crecimiento (velocidad) dificultan su captura, gestión, procesamiento o análisis mediante tecnologías y herramientas convencionales, tales como bases de datos relacionales y estadísticas convencionales o paquetes de visualización, dentro del tiempo necesario para que sean útiles.

Aunque el tamaño utilizado para determinar si un conjunto de datos determinado se considera Big Data no está firmemente definido y sigue cambiando con el tiempo, la mayoría de los analistas y profesionales actualmente se refieren a conjuntos de datos que van desde 30-50 Terabytes a varios Petabytes.

“La naturaleza compleja del Big Data se debe principalmente a la naturaleza no estructurada de gran parte de los datos generados por las tecnologías modernas” (García, 2020, p. 20). Como ejemplo tenemos los web logs, la identificación por radiofrecuencia (RFID), los sensores incorporados en dispositivos, la maquinaria, los vehículos, las

búsquedas en Internet, las redes sociales como Facebook, computadoras portátiles, teléfonos inteligentes y otros teléfonos móviles, dispositivos GPS y registros de centros de llamadas.

En la mayoría de los casos, con el fin de utilizar eficazmente el Big Data, debe combinarse con datos estructurados (normalmente de una base de datos relacional) de una aplicación comercial más convencional, como un ERP (Enterprise Resource Planning) o un CRM (Customer Relationship Management).

2.2.6 CLUSTERIZACIÓN

El clustering es una técnica de análisis de datos que se utiliza para identificar patrones y estructuras ocultas dentro de un conjunto de datos. Su objetivo principal es agrupar objetos o elementos similares, es decir, su objetivo es el de crear “clúster”, de manera que los elementos dentro de estas agrupaciones sean más similares entre sí que con respecto a los que encontramos en otros clústeres o grupos. (Perez, 2019)

Usos comunes de clusterización de datos:

- i. Segmentación de Mercado: En marketing, la clusterización se utiliza para dividir a los clientes en grupos llamados «segmentos» o «clúster» en función de sus comportamientos de compra, preferencias de producto, ubicación geográfica y otros atributos. Esto permite a las empresas personalizar sus estrategias de marketing, mensajes publicitarios y ofertas promocionales para satisfacer las necesidades específicas de cada segmento.
- ii. Recomendación de Contenido: Plataformas como Netflix, Amazon y Spotify utilizan esta técnica de agrupación de datos para mejorar la experiencia del usuario al recomendar contenido relevante. Cuando identifican patrones de visualización o escucha similares entre usuarios, pueden sugerir películas, productos o canciones que otros usuarios con perfiles similares también disfrutaron.
- iii. Detección de Fraude y/o anomalías: En el sector financiero, la clusterización se utiliza para identificar patrones sospechosos de actividad fraudulenta. Por ejemplo, los bancos pueden agrupar transacciones financieras en base a criterios como la ubicación geográfica, la frecuencia y el monto para detectar actividades anómalas, como el uso no autorizado de tarjetas de crédito.

- iv. Agrupación de Documentos: ¿Crear Clústeres para organizar grandes colecciones de documentos de texto, como correos electrónicos, informes y artículos? Pues claro, ya que esto facilita la búsqueda y la recuperación de información relevante. Los motores de búsqueda, por ejemplo, pueden utilizar este método para agrupar páginas web relacionadas por temas.
- v. Biología y Genómica: En la investigación biológica, la clusterización se utiliza para agrupar genes, proteínas o secuencias de ADN con patrones de expresión similares.
- vi. Optimización de Cadenas de Suministro: Las empresas lo utilizan para agrupar productos o clientes en grupos con características similares en la cadena de suministro.
- vii. Análisis de Redes Sociales: En este ámbito, el clustering es realmente interesante y muy usado por todas las plataformas agrupando usuarios con intereses y comportamientos similares. De esta forma facilita la identificación de comunidades y la recomendación de conexiones con otros usuarios y/o contenido relevante. Además, las empresas también pueden utilizar esta información para enfocar sus anuncios en grupos específicos
- viii. Exploración de Datos Geoespaciales: En sistemas de información geográfica (SIG), la clusterización se utiliza para identificar áreas geográficas con características similares, como densidad de población, ingresos, uso de suelo o amenazas naturales.
- ix. Segmentación de Imágenes y Videos: En visión por computadora también es común usar Machine Learning y Clusterización de datos para agrupar objetos o escenas similares en imágenes o videos.

En cada uno de estos casos, la clusterización de datos ayuda a identificar patrones en los conjuntos de datos, lo que permite tomar mejores decisiones, optimizar la eficiencia y personalizar las experiencias de los usuarios.

2.2.7 CRISP-DM

CRISP-DM, que son las siglas de Cross-Industry Standard Process for Data Mining, es un método probado para orientar sus trabajos de minería de datos. Como metodología, incluye descripciones de las fases normales de un proyecto, las tareas necesarias en cada fase y una explicación de las relaciones entre las tareas.

Como modelo de proceso, CRISP-DM ofrece un resumen del ciclo vital de minería de datos.

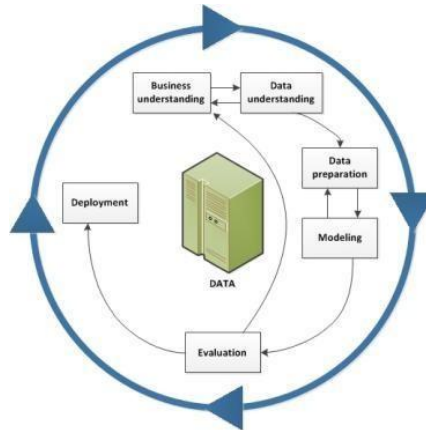


Figura 1 Diagrama de Modelo CRISP-DM

Fuente: (ibm.com)

El ciclo vital del modelo contiene seis fases con flechas que indican las dependencias más importantes y frecuentes entre fases. La secuencia de las fases no es estricta. De hecho, la mayoría de los proyectos avanzan y retroceden entre fases si es necesario.

El modelo de CRISP-DM es flexible y se pueden personalizar fácilmente. Por ejemplo, si la organización intenta detectar actividades de blanqueo de dinero, es probable que necesite realizar una criba de grandes cantidades de datos sin un objetivo de modelado específico. En lugar de realizar el modelado, su trabajo se centrará en explorar y visualizar datos para descubrir patrones sospechosos en datos financieros. CRISP-DM permite crear un modelo de minería de datos que se adapte a sus necesidades concretas.

En tal situación, las fases de modelado, evaluación y despliegue pueden ser menos relevantes que las fases de preparación y comprensión de datos. Sin embargo, es muy importante considerar algunas cuestiones que surgen durante fases posteriores para la planificación a largo plazo y objetivos futuros de minería de datos.

2.2.8 CRM

CRM es la sigla utilizada para "Customer Relationship Management" (Gestión de Relación con los Clientes). Sin embargo, la definición completa de CRM va más allá: se trata de una gestión integrada de ventas, marketing, atención al cliente y todos los puntos de contacto. (Martinez, 2018)

Estamos en la era del cliente, de la transformación digital y de las nuevas tecnologías. En este escenario, las relaciones también evolucionan y este fenómeno ha dado lugar al surgimiento de un concepto conocido como experiencia del cliente. Por eso, la experiencia que proporciona un CRM implica estar centrado en el cliente. Se trata de un concepto mucho más amplio que el de software o plataforma: es una estrategia, un proceso, una herramienta y una tecnología.

2.2.9 DATA MINING O MINERÍA DE DATOS

El proceso de escarbar en los datos para descubrir conexiones ocultas y predecir tendencias futuras tiene una larga historia. A veces denominado "knowledge discovery in databases," el término "data mining" no se acuñó hasta la década de 1990. Pero sus cimientos comprenden tres disciplinas científicas entrelazadas: la estadística (el estudio numérico de las relaciones entre datos), la inteligencia artificial (inteligencia similar a la humana desplegada por programas informáticos y/o máquinas) y el aprendizaje automático (algoritmos que pueden aprender de los datos para hacer predicciones). Lo que era viejo vuelve a ser nuevo, ya que la tecnología de extracción de datos sigue evolucionando para adaptarse al potencial ilimitado de los macrodatos y a una potencia informática asequible.

En la última década, los avances en el poder y la velocidad del tratamiento (datos) nos han permitido ir más allá de las prácticas manuales, tediosas y que consumen mucho tiempo al análisis de datos rápido, fácil y automatizado. Cuanto más complejo es el conjunto de datos recopilados, mayor es el potencial que hay para descubrir insights relevantes. (Garcia J. , 2019)

Retailers, empresas alimentos como LEYDE, bancos, manufactureros, proveedores de telecomunicaciones y aseguradoras, entre otros, utilizan la minería de datos para descubrir relaciones entre todo, desde optimización de precios, promociones y demografía hasta cómo afectan la economía, el riesgo, la competencia y las redes sociales a su modelo de negocio, ingresos, operación y relaciones con clientes.

Como concepto el Data Mining es el proceso de extraer información útil de una acumulación de datos, a menudo de un almacén de datos o de la recopilación de conjuntos de datos vinculados. Las herramientas de minería de datos incluyen potentes capacidades estadísticas, matemáticas y analíticas cuyo propósito principal

es examinar grandes conjuntos de datos para identificar tendencias, patrones y relaciones para respaldar la toma de decisiones y la planificación fundamentadas. (Torres, 2020)

2.2.10 DIFERENCIAS ENTRE CLUSTERIZACIÓN Y CLASIFICACIÓN

Clusterización y clasificación son dos técnicas fundamentales en el análisis de datos, particularmente en el ámbito de Machine Learning. A pesar de que ambas buscan segmentar conjuntos de datos, se diferencian en varios aspectos clave.

1. Clusterización (Clustering)

“La clusterización es una técnica de aprendizaje no supervisado en la que los datos no están etiquetados” (Fernández, 2021, p. 15), es decir, no se tiene información previa sobre las categorías o clases a las que pertenecen. El objetivo de la clusterización es agrupar datos similares en un mismo conjunto, creando “clusters” o grupos en función de características compartidas. Estos grupos son formados automáticamente por el algoritmo sin intervención humana. Ejemplos comunes de algoritmos de clusterización incluyen K-Means, DBSCAN y hierarchical clustering.

La clusterización se usa principalmente para descubrir patrones ocultos en los datos o para segmentar mercados en estudios de marketing, donde los grupos o segmentos de clientes no están claramente definidos de antemano.

Ejemplo de clusterización: Agrupar clientes en base a su comportamiento de compra, identificando segmentos como “compradores frecuentes”, “compradores ocasionales” o “compradores impulsivos”, sin tener previamente definida esa clasificación.

2. Clasificación

Por otro lado, “la clasificación es una técnica de aprendizaje supervisado, lo que significa que se trabaja con datos previamente etiquetados” (López, 2021, pág. 23). En la clasificación, el objetivo es asignar nuevos datos a una de varias clases predefinidas, basándose en la experiencia previa de un conjunto de entrenamiento. Los algoritmos de

clasificación aprenden a partir de ejemplos para luego predecir a qué clase pertenece un nuevo dato. Ejemplos comunes de algoritmos de clasificación incluyen Support Vector Machines (SVM), árboles de decisión, y redes neuronales.

La clasificación se utiliza para tareas donde se desea asignar una etiqueta o clase específica a cada entrada, como en el reconocimiento de imágenes o en la detección de fraudes.

Ejemplo de clasificación: Asignar a un cliente en una tienda en línea a una categoría como “cliente potencial”, “cliente regular” o “cliente VIP”, en función de su historial de compras.

Tabla 1. Resumen Corporativo

Aspecto	Clusterización	Clasificación
Tipo de aprendizaje	No supervisado	Supervisado
No supervisado	No se utilizan etiquetas predefinidas	Utiliza etiquetas de clase predefinidas
Objetivo	Descubrir grupos o patrones ocultos en los datos	Asignar una clase específica a nuevos datos
Algoritmos comunes	K-Means, DBSCAN, clustering jerárquico	SVM, árboles de decisión, redes neuronales
Aplicaciones	Segmentación de mercado, análisis exploratorio	Clasificación de correos (spam/no spam), diagnósticos médicos

Fuente: Elaboración Propia

2.2.11 ETL

Extracción, transformación y carga (ETL) es el proceso consistente en combinar datos de diferentes orígenes un gran repositorio central llamado

almacenamiento de datos. ETL utiliza un conjunto de reglas comerciales para limpiar y organizar datos en bruto y prepararlos para el almacenamiento, el análisis de datos y el machine learning (ML). (Martinez, 2021, p. 19)

Puede abordar necesidades de inteligencia empresarial específicas mediante análisis de datos (como la predicción del resultado de decisiones empresariales, la generación de informes y paneles, la reducción de la ineficacia operativa y más).

2.2.12 INTELIGENCIA DE NEGOCIOS

El business intelligence (BI) o inteligencia de negocios combina análisis de negocios, minería de datos, visualización de datos, herramientas e infraestructura de datos, y las prácticas recomendadas para ayudar a las organizaciones a tomar decisiones más basadas en los datos. En la práctica, sabes que tienes una inteligencia de negocios moderna cuando tienes una visión integral de los datos de tu organización y los utilizas para impulsar el cambio, eliminar las ineficiencias y adaptarte rápidamente a los cambios del mercado o del suministro.

Es importante tener en cuenta que esta es una definición muy moderna de inteligencia de negocios y que BI también ha tenido fama de ser una palabra de moda. La inteligencia de negocios tradicional “surgió originalmente en la década de 1960 como un sistema de intercambio de información entre organizaciones” (Romero, 2022, p. 45). Se desarrolló aún más en la década de 1980 junto con modelos informáticos para la toma de decisiones y convertir los datos en información antes de convertirse en una oferta específica de parte de equipos de BI con soluciones de servicio dependientes de TI. Las soluciones de BI modernas priorizan el análisis flexible de autoservicio, los datos gobernados en plataformas confiables, los usuarios corporativos empoderados y la velocidad para obtener información.

2.2.13 K-MEANS

“El algoritmo k-means es un método de agrupamiento que divide un conjunto de datos en k grupos o clusters” (Pérez, 2021, pág. 56). Los datos se agrupan de tal manera que los puntos en el mismo clúster sean más similares entre sí que los puntos en otros clusters.

Del universo de algoritmos de aprendizaje no supervisado, K-means sigue siendo uno de los algoritmos más conocidos para el aprendizaje no supervisado, aunque alternativas más avanzadas como DBSCAN o algoritmos basados en clustering espectral han ganado popularidad en ciertos escenarios debido a su capacidad para manejar conjuntos de datos más complejos y de mayor dimensión. La razón por la que existe este método es porque en 2024, la cantidad total de datos creados, capturados, copiados y consumidos globalmente ha superado los 200 Zettabytes, impulsada principalmente por el crecimiento de dispositivos IoT, la inteligencia artificial generativa y las redes 5G. Con el algoritmo k-means es posible recopilar grandes cantidades de información similar en un mismo lugar, hecho que ayuda a encontrar patrones y hacer predicciones en grandes conjuntos de datos. Para utilizar el algoritmo K-means, primero se especifica el número de clusters deseados (k). Por ejemplo, al establecer «k» igual a 2, su conjunto de datos se agrupará en 2 grupos, mientras que si establece «k» igual a 4, agrupará los datos en 4 grupos.

Cada grupo está representado por su centro o centroide, que corresponde a la media aritmética de los puntos de datos asignados al grupo. De esta manera, el algoritmo funciona a través de un proceso iterativo hasta que cada punto de datos está más cerca del centroide de su propio grupo que de los centroides de otros grupos, minimizando la distancia dentro del grupo en cada paso.

A continuación, detallamos cómo funciona paso a paso el algoritmo K-means:

- A. Especificar el número de clústers deseados (k): El primer paso es especificar cuántos clústeres queremos dividir el conjunto de datos. Este número se denomina k.
- B. Seleccionar k puntos al azar del conjunto de datos como los centroides iniciales de cada clúster: Luego, se eligen k puntos al azar del conjunto de datos para servir como los centroides iniciales de cada clúster. Estos centroides son el punto central o el promedio de cada clúster.
- C. Asignar cada punto del conjunto de datos al clúster cuyo centroide esté más cerca: A continuación, el algoritmo asigna cada punto del conjunto de datos al clúster cuyo centroide esté más cerca. Para hacer esto, se calcula la distancia entre cada punto y cada centroide y se asigna el punto al clúster cuyo centroide tenga la menor distancia.

- D. Recalcular los centroides de cada clúster como la media de todos los puntos del clúster: Una vez que todos los puntos han sido asignados a un clúster, se recalculan los centroides de cada clúster como la media de todos los puntos del clúster. Esto significa que se actualiza la posición del centroide para reflejar la nueva agrupación.
- E. Repetir los pasos 3 y 4 hasta que los centroides de los clusters ya no cambien o hasta que se alcance el número máximo de iteraciones.

2.2.14 KNIME

“KNIME (Konstanz Information Miner) pertenece a una nueva generación de herramientas dominadas como Plataformas de Data Science y Machine Learning por Gartner” (López M., 2022, pág. 36). Estas herramientas permiten a científicos de datos expertos, analistas o usuarios de negocio interactuar con sus datos y crear, desplegar y gestionar sus modelos de analítica avanzada. Las herramientas integran las funcionalidades principales para realizar proyectos de minería de datos: importación de datos, preparación de datos, exploración de datos, modelado, evaluación y despliegue. Dentro de estas herramientas KNIME se encuentra en el grupo de líderes del último diagrama de Gartner conocido también como la “navaja suiza” del mercado.

Como puntos fuertes podemos destacar tenemos la Facilidad de uso; a través de interfaz visual la programación de aplicaciones en KNIME es altamente intuitiva. Conectando visualmente nodos que encapsulan distintas funciones e integrando módulos automatizados de Machine Learning y modelos preprogramados, facilita la analítica avanzada a los usuarios de negocio sin experiencia en este ámbito. Al mismo tiempo brinda un ecosistema óptimo para desarrolladores avanzados con la posibilidad de integrar programación en Python o R.

Extensas funcionalidades: KNIME ofrece el ciclo completo de Data Mining. Con la posibilidad de conectarse a múltiples y heterogéneas fuentes de datos podremos unificar datos provenientes de distintas BBDD, archivos y servicios web diversos como Azure, etc. Con muy poco esfuerzo. Con una gran variedad de nodos para el preprocesamiento la herramienta ofrece las condiciones óptimas para la generación de procesos de ETL automatizada. Finalmente ofrece los principales algoritmos y métodos de evaluación para la generación de modelos potentes. Adicionalmente dispone de múltiples extensiones (Text

Processing, Big Data con Spark y Hadoop, Deep Learning con TensorFlow y Keras y muchas más) que empoderan la herramienta aún más. (Data Solutions, n.d.)

El Bajo costo de adquisición e implantación: KNIME cuenta con una versión gratuita “KNIME Analytics Platform” para el uso personal, así como una versión de pago “KNIME Server” para el uso en organizaciones que quieren llevar sus actividades de data mining a un nuevo nivel.

2.2.15 MACHING LEARNING

“La inteligencia como la capacidad de predecir el futuro, por ejemplo, el peso de un vaso que vamos a levantar o la reacción de los demás a nuestros actos, en base a los patrones almacenados en la memoria (el marco memoria-predicción)” (Hawking, 2004, p. 28). Ese mismo principio está detrás del Machine Learning (ML), también conocido como aprendizaje automático.

El Machine Learning es una disciplina del campo de la Inteligencia Artificial que, a través de algoritmos, dota a los ordenadores de la capacidad de identificar patrones en datos masivos y elaborar predicciones (análisis predictivo). Este aprendizaje permite a los computadores realizar tareas específicas de forma autónoma, es decir, sin necesidad de ser programados. El término se utilizó por primera vez en 1959. Sin embargo, ha ganado relevancia en los últimos años debido al aumento de la capacidad de computación y al boom de los datos. Las técnicas de aprendizaje automático son, de hecho, una parte fundamental del Big Data.

Distintos algoritmos de Machine Learning:

Los algoritmos de Machine Learning se dividen en tres categorías, siendo las dos primeras las más comunes:

- 1) Aprendizaje supervisado: estos algoritmos cuentan con un aprendizaje previo basado en un sistema de etiquetas asociadas a unos datos que les permiten tomar decisiones o hacer predicciones. Un ejemplo es un detector de spam que etiqueta un e-mail como spam o no dependiendo de los patrones que ha aprendido del histórico de correos (remitente, relación texto/imágenes, palabras clave en el asunto, etc.).

- 2) Aprendizaje no supervisado: estos algoritmos no cuentan con un conocimiento previo. Se enfrentan al caos de datos con el objetivo de encontrar patrones que permitan organizarlos de alguna manera. Por ejemplo, en el campo del marketing se utilizan para extraer patrones de datos masivos provenientes de las redes sociales y crear campañas de publicidad altamente segmentadas.
- 3) Aprendizaje por refuerzo: su objetivo es que un algoritmo aprenda a partir de la propia experiencia. Esto es, que sea capaz de tomar la mejor decisión ante diferentes situaciones de acuerdo a un proceso de prueba y error en el que se recompensan las decisiones correctas. En la actualidad se está utilizando para posibilitar el reconocimiento facial, hacer diagnósticos médicos o clasificar secuencias de ADN.

2.2.16 PERFILAMIENTO DE CLIENTES

Representa la percepción que una empresa tiene sobre el cliente óptimo al que aspira dirigir sus iniciativas de marketing, con el propósito de dinamizar las transacciones. En esencia, se trata de una delineación meticulosa de aquellos posibles consumidores a quienes se pretende captar y persuadir (Malaga, 2023, p. 45).

“El perfilamiento de clientes implica detallar las particularidades de un individuo o un conjunto de consumidores, abarcando aspectos demográficos, ubicación geográfica y rasgos psicológicos, así como sus hábitos de compra, solvencia financiera y registros previos de compras” (Moreno, 2022, p. 32).

Es, esencialmente, un proceso de identificación de los atributos de aquellos individuos que son más propensos a realizar la compra de un producto o servicio y beneficiarse de este. Al precisar estos atributos de tu público meta, es posible dividir tu cartera de clientes en distintos segmentos basados en estos perfiles.

2.2.17 SEGMENTACIÓN DE CLIENTES

La segmentación de clientes, también llamada segmentación de mercado, permite a las empresas dividir a sus clientes en subconjuntos homogéneos dentro de un mercado diverso. Esta técnica ayuda a comprender mejor a los clientes ya orientar decisiones de marketing hacia grupos claramente definidos, enfocándose en

satisfacer necesidades específicas. Además, permite identificar los grupos de clientes más rentables, facilitando el enfoque en su retención (Analitycs, 2019).

Mediante la información recabada a través de la segmentación de la clientela, es posible discernir los atributos esenciales del consumidor prototípico en la empresa LEYDE, permitiendo así determinar las categorías de individuos que conformarán la audiencia meta a ser posteriormente analizada y categorizada.

2.2.18 BOSQUES ALEATORIOS o RANDOM FOREST

Los Random Forest son un conjunto de algoritmos de aprendizaje automático que pertenecen a la categoría de los métodos de conjunto, específicamente a los "bosques aleatorios". Este método consiste en la creación de múltiples árboles de decisión independientes que se combinan para producir una predicción más robusta y precisa. Cada árbol toma una muestra aleatoria del conjunto de datos y realiza predicciones. Luego, los resultados de todos los árboles se promedian (en regresión) o se toman como votación mayoritaria (en clasificación) para obtener un resultado final, lo cual reduce el riesgo de sobreajuste y mejora la precisión del modelo.

"Random Forest es una técnica de aprendizaje automático de ensamblaje que crea múltiples árboles de decisión y combina sus predicciones para mejorar la precisión y evitar problemas como el sobreajuste" (Breiman, 2001, p. 45)

2.3 TEORÍAS DE SUSTENTO

2.3.1 BASES TEÓRICAS

Se revisó la literatura encontrando algunas teorías basadas en la segmentación de clientes, entre las cuales se puede mencionar:

2.3.1.1 TEORIA DE SEGMENTACION DE MERCADOS

La segmentación de mercado es un enfoque clave en marketing y ventas que respalda el uso de técnicas de clusterización. Al agrupar a los clientes o productos en segmentos basados en patrones de comportamiento de compra, la organización puede

adaptar sus estrategias de ventas y marketing a cada grupo, aumentando la efectividad de sus esfuerzos. Esta teoría sostiene que no todos los clientes tienen los mismos intereses, valores o comportamientos de compra, y que un análisis profundo de estos segmentos permite una mejor gestión de recursos y estrategias.

Wedel y Kamakura (2000) presentan un enfoque profundo sobre la segmentación de mercado y las técnicas de clusterización, argumentando que estas herramientas facilitan la identificación de patrones de comportamiento de compra, lo cual es esencial para el desarrollo de estrategias de marketing efectivas.

2.3.1.2 TEORIA DEL CICLO DE VIDA DEL PRODUCTO

La teoría del ciclo de vida del producto establece que cada producto pasa por distintas fases (introducción, crecimiento, madurez y declive) en su permanencia en el mercado. Al aplicar análisis de series de tiempo, se pueden identificar tendencias que indiquen en qué fase se encuentra un producto en específico. Esta teoría proporciona una base para interpretar los patrones temporales en las ventas y anticipar decisiones sobre promoción, rediseño, y discontinuación de productos.

Kotler y Keller (2016) explican detalladamente las fases del ciclo de vida del producto y cómo cada etapa requiere diferentes estrategias de marketing. Señalan que las series de tiempo ayudan a observar cambios en las ventas ya adaptar las estrategias en consecuencia para maximizar el rendimiento en cada fase del ciclo de vida.

2.3.1.3 TEORÍA DEL COMPORTAMIENTO DEL CONSUMIDOR

Basada en la psicología y sociología, esta teoría explica cómo y por qué los consumidores toman decisiones de compra. Utilizando técnicas de clusterización, podemos identificar patrones específicos en los datos históricos que reflejan distintos tipos de comportamiento del consumidor. Este enfoque permite a las empresas crear perfiles de clientes según su comportamiento de compra y personalizar estrategias de marketing, ventas y fidelización.

Solomon (2018) analiza cómo los principios de psicología y sociología son fundamentales para comprender los patrones de comportamiento de compra. Este libro detalla cómo la segmentación y la clusterización permiten identificar grupos de

consumidores con motivaciones similares, lo que es clave para crear estrategias de marketing personalizadas.

2.3.1.4 TEORÍA DE LOS PROCESOS ESTOCÁSTICOS

Esta teoría es crucial en el análisis de series de tiempo. Un proceso estocástico asume que el comportamiento futuro de una serie de tiempo tiene una probabilidad basada en observaciones pasadas. Esto respalda el uso de modelos como ARIMA o Modelos de Suavizado Exponencial en la predicción de ventas futuras. La teoría asume que el comportamiento de las series de tiempo puede descomponerse en componentes como tendencia, estacionalidad y ruido, permitiendo una modelización más precisa.

Box, Jenkins y Reinsel (2015) presentan una base sólida para el uso de modelos de series de tiempo como ARIMA, explicando cómo los procesos estocásticos permiten que el comportamiento futuro de una serie dependa de sus observaciones pasadas. Su teoría descompone la serie en componentes de tendencia, estacionalidad y ruido, lo que mejora la precisión de las predicciones.

2.3.1.5 TEORÍA DE REDES Y PROPAGACION DE INFORMACION

En el contexto de ventas, la teoría de redes puede ser útil para analizar cómo se propaga la información sobre productos o servicios a través de redes sociales y boca a boca. Utilizando la clusterización, se pueden identificar grupos de usuarios influyentes o con patrones de compra similares. En un análisis de series de tiempo, esta teoría también ayuda a entender cómo la información en redes sociales afecta el comportamiento de ventas de forma dinámica.

La Revista de la Universidad de Manouba en Tunez (2024) afirma que La teoría de las redes sociales es esencial para entender cómo la información sobre productos o servicios se propaga a través de plataformas como redes sociales y el boca a boca. Al analizar redes, podemos identificar nodos claves o usuarios influyentes que juegan un papel crucial en la difusión de información. Además, las técnicas de clusterización permiten agrupar usuarios con comportamientos similares, lo cual facilita la creación de estrategias de marketing personalizadas. En cuanto a la integración con series de tiempo, los modelos de difusión de información, como los basados en cadenas de Markov o difusión aleatoria, pueden predecir cómo los cambios en las redes sociales afectan las ventas a lo largo del tiempo

2.3.1.6 TEORÍA DE SISTEMAS DINAMICOS

La teoría de sistemas dinámicos explica cómo los sistemas complejos cambian con el tiempo y cómo sus componentes interactúan entre sí. En análisis de series de tiempo, los datos de ventas se consideran como un sistema dinámico influido por diversas variables (marketing, estacionalidad, etc.). Esta teoría respalda el uso de modelos dinámicos y técnicas de simulación para entender el comportamiento de las ventas en el tiempo y las posibles respuestas a cambios en condiciones de mercado.

2.3.1.7 TEORÍA DE LA COMPETENCIA Y ESTRATEGIA DE PORTER

Porter sostiene que las empresas pueden ganar ventaja competitiva mediante estrategias de liderazgo en costo, diferenciación o enfoque. La clusterización de ventas puede mostrar cómo ciertos productos o regiones responden a estas estrategias, ayudando a diseñar estrategias competitivas en cada segmento. Esto permite identificar oportunidades de mercado y ajustar la estrategia de ventas según los patrones revelados por los datos.

Estas teorías proveen una base sólida para interpretar los resultados y justificar el uso de análisis de clusterización y series de tiempo en la toma de decisiones empresariales, ya que permiten entender mejor la dinámica del mercado y la demanda.

2.3.2 METODOLOGÍAS DESARROLLADAS

Basados en las metodologías desarrolladas en dos empresas como Amazon y Walmart que han aplicado técnicas de clusterización y series de tiempo a su histórico de ventas para mejorar sus estrategias de negocio.

1. Amazon:

Amazon aplica la clusterización para segmentar a sus clientes y personalizar las recomendaciones de productos según sus preferencias de compra. Además, usa análisis de series de tiempo para predecir la demanda de productos y gestionar de manera óptima sus inventarios en tiempo real. Además, Amazon segmenta a sus clientes con base en el historial

de navegación, compras y preferencias, utilizando algoritmos como el K-Means y el clustering jerárquico. “Esta segmentación le permite ofrecer recomendaciones más precisas y personalizadas a cada usuario” (Smith K. , 2019, p. 45).

Amazon emplea series de tiempo para gestionar su cadena de suministro, anticipando picos de demanda y optimizando la reposición de inventarios en diferentes centros de distribución, sobre todo en fechas de alta demanda como Black Friday y Prime Day.

2. Walmart:

Walmart utiliza técnicas de clusterización y análisis de series de tiempo en sus datos históricos de ventas para optimizar su cadena de suministro y personalizar la distribución de productos a nivel local. Estas técnicas le permiten entender patrones de demanda específicos por región y estación del año.

Walmart utiliza el clustering para segmentar tiendas según características de ventas y comportamiento de clientes en diferentes zonas geográficas, ajustando sus estrategias de inventario y publicidad en función de estos segmentos. “La empresa emplea análisis de series de tiempo para pronosticar la demanda y anticipar inventarios en diferentes regiones” (Doe & Perez, 2020, p. 145), especialmente en épocas festivas y durante eventos meteorológicos (como huracanes) que afectan el comportamiento de compra.

Estas referencias presentan un enfoque teórico y práctico de cómo empresas líderes como Amazon y Walmart aplican la clusterización y las series de tiempo para mejorar su desempeño en ventas y logística.

2.4 MARCO LEGAL

2.4.1 LA PROTECCIÓN DE DATOS EN HONDURAS

Honduras enfrenta un gran desafío en la promoción y protección efectiva de los derechos relacionados con la protección de datos. Aunque estos derechos están reconocidos formalmente, el país carece de las herramientas, mecanismos y una cultura social lo suficientemente sólidos para garantizar su implementación de manera práctica y efectiva, que es el propósito principal de cualquier derecho formalizado.

La Constitución de la República, en su artículo 76, garantiza el “derecho al honor, a la intimidad personal, familiar y a la propia imagen”, y en el artículo 100 establece que “toda persona tiene derecho a la inviolabilidad y al secreto de las comunicaciones... salvo resolución judicial... los documentos personales únicamente estarán sujetos a inspección o fiscalización de la autoridad competente.” Estos artículos proporcionan una base para el reconocimiento de los derechos a la protección de datos” (Zelaya, 2021).

Desde abril de 2018, la Asamblea Legislativa de Honduras ha aplazado el debate final necesario para la aprobación de la normativa sobre Protección de Datos Personales. Este proyecto es fundamental, ya que actualmente no existe un marco legal en Honduras que regule la protección de datos personales, lo que hace de esta propuesta un avance legislativo importante. Sin embargo, algunos conceptos propuestos en la legislación son susceptibles a interpretaciones ambiguas o incorrectas.

2.4.2 POLÍTICA DE CIBERSEGURIDAD EN LA EMPRESA LEYDE

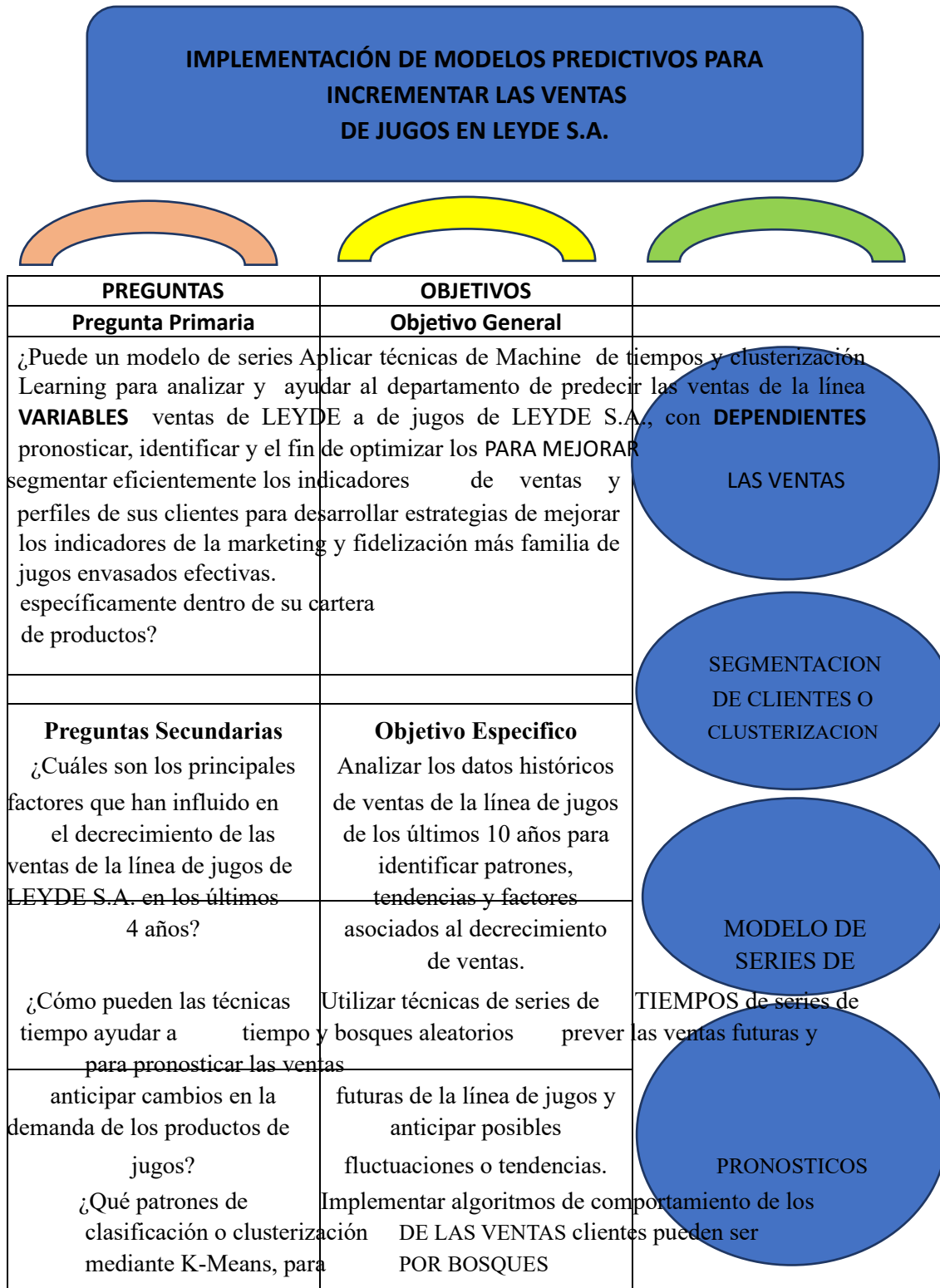
LEYDE cuenta con un departamento de TI asignado para todas las políticas de seguridad que deben enfocarse en garantizar la protección tanto de sus empleados, controles de inventarios, ventas, asegurando un ambiente de trabajo seguro y el cumplimiento de normativas locales e internacionales. Estas políticas incluyen protocolos estrictos de higiene en las instalaciones, especialmente en áreas de procesamiento de alimentos, para evitar cualquier tipo de contaminación. Asimismo, se mantienen capacitaciones periódicas sobre seguridad industrial, el manejo adecuado de maquinaria y la respuesta ante emergencias. La empresa también cuenta con sistemas de monitoreo y control de acceso a las instalaciones, así como medidas de ciberseguridad que protejan los datos sensibles y operativos, asegurando la integridad de la información y la continuidad de las operaciones.

De igual manera, existen Normativas de Uso de Equipos de Cómputo y Normativas de Gestión Documental y Archivos. La primera busca establecer el uso correcto de equipos como laptops, computadoras de escritorio, monitores, teclados, ratones, servidores, entre otros. La segunda tiene como finalidad definir directrices para la recepción y custodia tanto física como digital de expedientes, garantizando la confidencialidad y el resguardo adecuado de la documentación.

CAPÍTULO III. METODOLOGIA DE LA INVESTIGACIÓN

3.1 CONGRUENCIA METODOLOGICA

3.1.1 MATRIZ METODOLÓGICA



identificados mediante técnicas de clustering y análisis RFM, y cómo se segmentar el mercado?

segmentar a los clientes según su comportamiento de compra. pueden utilizar para

ALEATORIOS

	<p>Desarrollar recomendaciones para optimizar las estrategias de marketing y ventas, basadas en los resultados del análisis de datos y las predicciones generadas por los modelos de Machine Learning.</p>	
--	--	--

Aquí presentamos una matriz metodológica para la empresa LEYDE S.A que procesa y vende jugos de sabores con el objetivo de incrementar las ventas a mediano plazo. Esta matriz se enfoca en diferentes áreas que la empresa puede considerar para mejorar su desempeño y aumentar sus ventas. Donde además haremos énfasis en los siguientes aspectos que se deberá hacer un estudio concluyente por parte de mercadeo y ventas en los siguientes ítems.

1. Análisis del mercado y competencia:

- Investigación de mercado para identificar tendencias de consumo de jugos y preferencias de sabor.
- Análisis de la competencia para comprender sus estrategias de ventas, precios y productos.
- Identificación de nichos de mercado no explotados o áreas geográficas con potencial de crecimiento.

2. Desarrollo de productos:

- Investigación y desarrollo de nuevos sabores de jugos que puedan atraer a diferentes segmentos de clientes.
- Pruebas de sabor y aceptabilidad del producto con grupos de muestra.
- Evaluación de la viabilidad de productos orgánicos, sin azúcar añadido u otras tendencias de salud.

3. Estrategias de marketing y promoción:

- Desarrollo de una estrategia de marca sólida que resalte los valores de la empresa y la calidad de sus productos.
- Implementación de campañas publicitarias en medios tradicionales y digitales.
- Uso de redes sociales y marketing de influencia para llegar a un público más amplio.
- Ofrecer degustaciones gratuitas en puntos de venta estratégicos para aumentar la visibilidad y generar interés en el producto.

4. Distribución y canales de venta:

- Evaluación de los canales de distribución existentes y búsqueda de nuevas oportunidades. Actualmente existen 3 canales de Ventas como ser: Ruteo, Institucionales y Supermercados.
- Negociación con supermercados, tiendas de conveniencia y otros minoristas para ampliar la presencia del producto.
- Exploración de la venta online a través de plataformas de comercio electrónico o la propia página web de la empresa.

5. Fijación de precios y estrategias de ventas:

- Análisis de costos para establecer precios competitivos pero rentables.
- Ofrecer descuentos y promociones temporales para incentivar la compra. -
Implementar programas de fidelización para mantener a los clientes recurrentes y aumentar el valor de vida del cliente.

6. Retroalimentación y mejora continua:

- Recopilación de comentarios de los clientes sobre los productos y servicios.
- Análisis de datos de ventas para identificar patrones y áreas de mejora. -
Realización de encuestas periódicas de satisfacción del cliente para evaluar la percepción de la marca y la calidad del producto.

- Ajuste de estrategias según los resultados obtenidos para optimizar el rendimiento y aumentar las ventas.

Esta matriz proporciona un marco de trabajo integral que la empresa LEYDE S.A. puede seguir para mejorar sus ventas de jugos de sabores. Es importante adaptar estas estrategias a las necesidades específicas y recursos disponibles de la empresa, así como realizar un seguimiento constante para medir el progreso y realizar ajustes según sea necesario.

3.1.2 ESQUEMA DE VARIABLES DE ESTUDIO

El esquema de las variables de estudio en una investigación que aplica técnicas de Machine Learning para mejorar las ventas de la familia de jugos de LEYDE S.A. se puede dividir en dos grandes grupos:

Las variables dependientes (resultado que queremos predecir) y variables independientes (factores o predictores que influyen en el resultado). Estas variables provienen de los datos históricos de ventas, las características de los clientes, y otros factores contextuales.

1. Variables Dependientes (VD)

Estas son las variables que se desean predecir o explicar mediante el análisis y los modelos de Machine Learning. En este caso, el enfoque está en las ventas:

- VD1: Ventas de la Familia de Jugos (en unidades o valor monetario)
- Descripción: Representa la cantidad de unidades vendidas o el ingreso generado por la venta de jugos en un período determinado (por ejemplo, mensual o trimestral). Esta es la variable principal que se quiere optimizar.
- Tipo: Continua.

2. Variables Independientes (VI)

Estas son las variables que influyen en el comportamiento de las ventas y se utilizan para entrenar los modelos de predicción o segmentación. Pueden ser categóricas o continuas, y abarcan factores tanto internos como externos.

2.1. Variables de Características de los Producto

□ VI1: Tipo de Producto

Descripción: Indica las diferentes variantes de jugos que produce LEYDE (sabores, tamaños, empaques, etc.).

- Tipo: Categórica (por ejemplo: "jugo de naranja", "jugo de piña", "jugo de mango").

□ VI2: Precio del Producto

- Descripción: Precio de venta de cada tipo de jugo.

- Tipo: Continua (en moneda local).

□ VI3: Promociones o Descuentos Aplicados

- Descripción: Indica si los productos han tenido descuentos o promociones en un período determinado.

- Tipo: Categórica (0 = sin descuento, 1 = con descuento).

2.2. Variables de Comportamiento del Cliente

Estas variables están relacionadas con el comportamiento de compra de los clientes y se usarán principalmente en la segmentación de clientes mediante el análisis RFM (Recencia, Frecuencia, Valor Monetario).

□ VI4: Recencia (R)

- Descripción: Tiempo transcurrido desde la última compra realizada por el cliente.

- Tipo: Continua (en días).

□ VI5: Frecuencia de Compra (F)

- Descripción: Número de veces que un cliente ha comprado jugos en un período de tiempo determinado.

- Tipo: Continua.

□ VI6: Valor Monetario (M)

- Descripción: Cantidad de dinero gastada por un cliente en la compra de jugos durante un período.

Tipo: Continua.

2.3. Variables Temporales y Estacionales

Este grupo de variables está relacionado con el factor temporal y su influencia en la demanda.

□ VI7: Mes o Estación del Año

- Descripción: Indica el mes o la estación del año en que se realiza la compra.

Puede ser relevante para identificar patrones estacionales.

- Tipo: Categórica (enero, febrero, marzo... / verano, invierno, etc.).

□ VI8: Días Festivos o Eventos Especiales

- Descripción: Marca si la venta coincide con un evento especial o festivo (como Navidad, Semana Santa, Día de la Madre, etc.), lo que podría influir en la demanda.

- Tipo: Categórica (0 = día normal, 1 = día festivo o especial).

2.4. Variables de Marketing y Publicidad

-
Estas variables se relacionan con las acciones de marketing de la empresa y cómo estas pueden influir en las ventas.

□ VI9: Gasto en Publicidad

- Descripción: Inversión en publicidad y campañas promocionales enfocadas en la familia de jugos.

- Tipo: Continua (en moneda local).

□ VI10: Canales de Distribución Utilizados

- Descripción: Canales a través de los cuales se distribuyen los productos (supermercados, tiendas locales, ventas en línea).

- Tipo: Categórica (ej.: "supermercado", "venta en línea", "distribuidores locales").

2.5. Variables Externas

Estas son variables exógenas que no dependen directamente de la empresa pero que podrían influir en las ventas.

□ VI11: Factores Económicos (Inflación, PIB, etc.)

- Descripción: Condiciones económicas que pueden influir en el poder adquisitivo de los consumidores.

- Tipo: Continua (porcentaje o valor de referencia).

□ VI12: Competencia (Lanzamientos o Promociones de Competidores)

- Descripción: Información sobre la actividad de los competidores en términos de lanzamientos de productos o campañas de marketing.

- Tipo: Categórica (0 = sin actividad competitiva relevante, 1 = con actividad relevante).

Justificación de las Variables

- Variables de ventas (VD1): Es el principal indicador que se busca mejorar, ya que refleja el desempeño de la familia de jugos.

- Variables independientes (VI1-VI12): Estas se seleccionaron por su potencial influencia en el comportamiento de las ventas. Las variables relacionadas con los productos y clientes permitirán analizar patrones de consumo, mientras que las variables estacionales y de marketing ayudarán a ajustar las estrategias comerciales de la empresa.

Este esquema proporciona una estructura clara de las variables que se utilizarán en los modelos de Machine Learning para optimizar las ventas y entender los comportamientos del mercado de LEYDE S.A.

3.1.3. MATRIZ DE OPERALIZACION DE LAS VARIABLES

Tabla 2. Esquema Resumido de las Variables de Estudio

Título de In la Investigación	Objetivo General	Tipo	Código	Variable	Tipo de Dato	Dimensiones	Items
		Dependent e (VD)	VD1	Ventas de la familia de jugos	Continua	Volu men de venta s Venta s por categ oría de jugos	Total, de ventas mensuales, variación mensual Ventas por tipo de jugo, ventas en diferentes regiones
		Independientes (VI)	VI1	Tipo de producto	Catagórica	Categoría de jugos	Sabor, tamaño, presentación
			VI2	Precio del producto	Continua	Precios aplicados	Precio medio, precio promocional.
			VI3	Promocione s o descuentos aplicados	Catagórica	Tipos de promoción	Descu entos, combo s, cupon es

. IMPLEMENTACIÓN DE MODELOS PREDICTIVOS PARA MEJORAR LAS VENTAS DE LA FAMILIA DE JUGOS EN LEYDE S.A.	Aplicar técnicas de Machine Learning para analizar y predecir las ventas de la línea de jugos de LEYDE S.A		VI4		Continua		
				Recencia (R)		Tiempo desde	Días desde última

						última compra	compra
		VI5	Frecuencia de compra (F)	Continua	Frecuencia de compra		Número de compras por mes, frecuencia por cliente

		VI6	Valor monetario (M)	Continua	Gasto total del cliente	Promedio de gasto, gasto máximo
		VI7	Mes o estación del año	Catagórica	Periodicidad	Meses, estaciones (verano, invierno)
		VI8	Días festivos o eventos especiales	Catagórica	Eventos de impacto en ventas	Festividades nacionales, eventos promocionales
		VI9	Gasto en publicidad	Continua	Inversión en publicidad	Gasto mensual en medios (digital, TV, radio)
		VI10	Canales de distribución utilizados	Catagórica	Canales de venta	Supermercados, tiendas de conveniencia
						a, en línea
		VI11	Factores económicos (Inflación, PIB, etc.)	Continua	Indicadores económicos	Inflación, tasa de interés, PIB

			VI12	Competencia (actividad de competidores)	Categoría	Actividad de competidores	Lanzamiento de nuevos productos, promociones de competidores

Fuente: Elaboración Propia

3.1.4 HIPÓTESIS DE LA INVESTIGACION

La hipótesis de esta investigación se formula en base al objetivo principal de mejorar los indicadores de ventas de la familia de jugos de la empresa LEYDE utilizando técnicas de Machine Learning. Se plantea que los factores internos (como el comportamiento de los clientes) y externos (como estacionalidad o condiciones del mercado) influyen en el desempeño de las ventas y que la aplicación de modelos predictivos y técnicas de segmentación permitirá a la empresa optimizar sus estrategias.

H1: Hipótesis General (HG)

La implementación de técnicas de Machine Learning para el análisis de series de tiempo y la segmentación de clientes permitirá mejorar significativamente las ventas de la familia de jugos de LEYDE, optimizando las estrategias de marketing y ventas.

Hipótesis Específicas (HE)

- Los modelos predictivos de series de tiempo permitirán anticipar con precisión la demanda futura de la familia de jugos, identificando patrones de estacionalidad, tendencias y ciclos de ventas, lo que contribuirá a una mejor planificación de producción y distribución.
- La segmentación de clientes o clusterización, permitirá identificar grupos de clientes con comportamientos de compra similares, facilitando el diseño de campañas de marketing personalizadas que incrementarán la fidelización y la frecuencia de compra.

- El algoritmo de Bosques Aleatorios nos dará una proyección de las ventas para el próximo año en las distintas sucursales. La modelización de estas variables permitirá ajustar las estrategias comerciales y maximizar el rendimiento de ventas en periodos clave.
- La aplicación de modelos de Machine Learning para la predicción de ventas reducirá el margen de error en la estimación de la demanda de jugos, en comparación con los métodos tradicionales de pronóstico utilizados por LEYDE, mejorando la eficiencia operativa y la rentabilidad.

Como Justificación de las Hipótesis general se basa en la premisa de que las técnicas de Machine Learning permiten aprovechar grandes volúmenes de datos para tomar decisiones informadas, optimizando la estrategia comercial de LEYDE S.A. Y las hipótesis específicas están diseñadas para validar el impacto de cada técnica utilizada (series de tiempo, clusterización y bosques aleatorios) en la mejora de los indicadores de ventas.

Estas hipótesis serán probadas a través de la aplicación de modelos predictivos y técnicas de análisis de datos sobre el historial de ventas de la empresa, con el objetivo de proponer estrategias efectivas para revertir la tendencia decreciente en la familia de jugos.

H0: Hipótesis Nula

Es lo contrario o que estas técnicas aplicadas no ayuden a mejorar los niveles de ventas del siguiente año.

3.2 ENFOQUE Y METODOS

Para mejorar las ventas de la empresa LEYDE, que procesa jugos de sabores, es crucial realizar un análisis exhaustivo de la data utilizando un enfoque sistemático y métodos analíticos adecuados.

La presente investigación sigue un enfoque cuantitativo y descriptivo-exploratorio, basado en el análisis de datos históricos de ventas de la empresa LEYDE con el objetivo de utilizar técnicas de Machine Learning para optimizar los indicadores de ventas de la familia de productos de jugos. A través del uso de algoritmos de aprendizaje automático, se pretende identificar patrones de consumo, predecir la demanda futura y segmentar a los clientes, permitiendo así la toma de decisiones estratégicas más informadas.

Para poder interpretar los datos históricos de ventas de los últimos diez años en LEYDE y poder emplear algoritmos de machine learning como clustering, bosques aleatorios y análisis de series de tiempo, estamos utilizando principalmente estadística descriptiva, estadística inferencial y algunas técnicas más avanzadas de estadística aplicada a modelos predictivos.

A continuación, haremos un desglose de los tipos de estadística que se están utilizando en diferentes partes del trabajo:

3.2.1. ESTADÍSTICA DESCRIPTIVA

Antes de aplicar los algoritmos de machine learning, es fundamental hacer un análisis exploratorio de los datos. La estadística descriptiva permite resumir y visualizar las características esenciales de los datos.

- Medidas de tendencia central: Como la media, la mediana y la moda, para obtener una idea del comportamiento promedio de las ventas en LEYDE
- Medidas de dispersión: Como la desviación estándar, varianza, percentiles, y rango intercuartil, que ayudarán a entender la variabilidad de las ventas.
- Visualizaciones: Histogramas, gráficos de barras, diagramas de dispersión (scatterplots) y diagramas de caja (boxplots) para identificar patrones, posibles outliers y distribuciones.

3.2.2. ESTADÍSTICA INFERENCIAL

Aunque nuestro enfoque principal es el uso de machine learning, en algunos casos vamos a emplear métodos de estadística inferencial para hacer suposiciones o generalizaciones sobre las ventas futuras o para hacer pruebas de hipótesis.

- Intervalos de confianza: Para estimar el rango dentro del cual podrían estar las ventas futuras con un cierto nivel de confianza.
- Pruebas de hipótesis: Vamos a verificar si hay diferencias significativas en las ventas entre diferentes períodos de tiempo (antes y después de la estacionalidad).
- Modelos de regresión: Vamos a evaluar una regresión múltiple como parte de los algoritmos de predicción o como comparativo a los modelos más avanzados.

3.2.3 ESTADÍSTICA APLICADA A MODELOS PREDICTIVOS Y SERIES DE TIEMPO

Aquí entra el uso más avanzado de la estadística aplicada en los modelos de machine learning:

- Análisis de series de tiempo: En este caso después de revisar los ciclos, tendencias y estacionalidad de las ventas históricas, se aplicara un modelos como random forest para modelar y prever las ventas futuras.

- Descomposición de series temporales: Este es un paso importante donde se descomponen los datos de ventas en tendencia de LEYDE como su estacionalidad y ruido.
- Autocorrelación y autocorrelación parcial (ACF y PACF): Estas funciones son fundamentales para entender la correlación en diferentes lags (desplazamientos) en las ventas.

- Clusterización:

- Técnicas como K-means, DBSCAN, o jerárquicas nos ayudaran a identificar patrones de ventas agrupando productos, clientes o periodos de tiempo con comportamientos similares.
- Distancias estadísticas: Las distancias euclidianas o las métricas de similitud/correlación entre diferentes conjuntos de datos de ventas se utiliza la estadística en el proceso de clustering.

3.2.4. EVALUACIÓN DE MODELOS (ESTADÍSTICA INFERENCIAL)

Una parte clave de aplicar modelos de machine learning en la empresa LEYDE es la evaluación de su desempeño utilizando métricas estadísticas:

- Errores en series temporales: Puedes usar métricas como MAE (Error Absoluto Medio), RMSE (Raíz del Error Cuadrático Medio), MAPE (Error Porcentual Absoluto Medio) para medir la precisión de los modelos de series de tiempo.

- Silhouette score: En el clustering, puedes usar medidas como el coeficiente de silhouette para evaluar la calidad de los grupos obtenidos.
- Cross-validation: El uso de validación cruzada (k-fold) para validar la capacidad predictiva de los modelos.

3.2.5. VISUALIZACIÓN DE DATOS

Aunque esto no es estadística pura, la visualización de datos es una parte importante del análisis estadístico y los modelos predictivos, ya que facilita la interpretación y comunicación de los resultados en LEYDE.

- Gráficos de series temporales: Para visualizar las tendencias y patrones en las ventas.
- Mapas de calor y gráficos de dispersión en clusterización: Para entender la segmentación obtenida por los algoritmos.

3.3 DISEÑO DE LA INVESTIGACIÓN

Este estudio es de naturaleza aplicada porque busca ofrecer soluciones prácticas a un problema específico dentro de la empresa. Además, combina aspectos de investigación correlacional para analizar las relaciones entre las variables, y de investigación predictiva a través de modelos que anticipan comportamientos futuros en base a datos históricos.

3.3.1 POBLACIÓN

Los datos utilizados en esta investigación provendrán principalmente de la base de datos interna de ventas de LEYDE, que incluye información detallada sobre:

- Ventas históricas de la línea de jugos (últimos 10 años).
- 11 sucursales en todo el país donde la planta de procesamiento en la ciudad de Ceiba despacha de forma diaria o semanal producto para la venta.
- Factores externos como la estacionalidad y ciclos de demanda en el mercado.
- Estos datos serán complementados con información del sector de bebidas y estudios de mercado para entender mejor las tendencias de consumo en Honduras.

3.3.2 MUESTRA

La población incluye las ventas de la familia de los Jugos en LEYDE de los últimos 10 años históricos o 117 meses desde Enero del 2014 a Septiembre del 2024. La muestra se define según Hernández-Sampieri (Hernández Sampieri et al., 2014) como “un subconjunto de elementos que pertenecen a un conjunto definido en sus características al que llamamos población”. Para efectos de esta investigación la muestra está compuesta por 1404 registros de ventas de las distintas sucursales en los 117 meses. En el proceso de transformación y limpieza de los datos se suprimieron los “Missing Values” o valores nulos. Representa una muestra no probabilística.

3.3.3 TÉCNICAS DE MUESTREO

Para este estudio la técnica de muestreo que se utilizó fue por selección intencionada o muestreo de conveniencia, el cual consiste en la selección por un procedimiento no aleatorio de una muestra cuyas características de la población sean muy similares o parecidas a la de la población objetivo (Casal & Mateu, 2003).

3.4 TÉCNICAS, INSTRUMENTOS Y PROCEDIMIENTOS APLICADOS

La investigación se desarrollará en las siguientes fases mostradas en la siguiente tabla:
Tabla de las fases de la Investigación

Tabla 3. Fases de la Investigación

Fase	Actividad	Duración estimada
Fase 1: Recolección de datos	Obtención y limpieza de datos de ventas	1 semana
Fase 2: Análisis Exploratorio	Análisis inicial y visualización de datos	1 semana
Fase 3: Modelos predictivos	Implementación de modelos de series de tiempo	1 semana
Fase 4: Segmentación	Clusterización de clientes y análisis de patrones	1 semana
Fase 5: Validación	Evaluación y ajuste de los modelos	1 semana
Fase 6: Propuesta estratégica	Desarrollo de estrategias basadas en los resultados	1 semana

Fuente: Elaboración Propia

3.4.1 PROCEDIMIENTOS APLICADOS

Desarrollo y Validación del Modelo:

- i. Se dividirán los datos en conjuntos de entrenamiento (70%) y validación (30%) para asegurar que los modelos sean precisos y no estén sobre ajustados.
- ii. La técnica de validación cruzada será utilizada para evaluar el rendimiento y la precisión de los modelos predictivos.
- iii. Los resultados de la segmentación de clientes y las predicciones de ventas serán comparados con las tendencias observadas para verificar la precisión de los algoritmos.

Propuesta de Estrategias de Optimización:

A partir de los resultados obtenidos, se desarrollarán estrategias de marketing y ventas para mejorar los indicadores de la familia de jugos. Estas estrategias podrían incluir:

- A. Campañas de marketing personalizadas: dirigidas a los distintos segmentos de clientes identificados mediante la clusterización.
- B. Ajustes en la planificación de producción: basados en las predicciones de demanda futura generadas por los modelos de series de tiempo.
- C. Programas de fidelización de clientes: orientados a retener a los clientes de mayor valor y mejorar la frecuencia de compra de los clientes menos activos.

3.5 FUENTES DE INFORMACIÓN

Un análisis de tendencias y pronósticos de mercado para una empresa de jugos podría implicar varios pasos y enfoques:

1. Recopilación de datos: La empresa recopila datos relevantes de diversas fuentes, como ventas pasadas, información demográfica, tendencias de consumo, datos de redes sociales, informes de la industria, etc. Estos datos pueden incluir información sobre preferencias de sabor, hábitos de compra, cambios en la demanda, innovaciones de productos, entre otros.

2. **Análisis exploratorio de datos:** Se lleva a cabo un análisis inicial de los datos para identificar patrones, tendencias y correlaciones. Esto puede incluir visualización de datos, análisis estadístico descriptivo y técnicas de minería de datos para descubrir información relevante y insights preliminares.

3. **Identificación de tendencias emergentes:** Utilizando herramientas de análisis de datos, se identifican tendencias emergentes en la industria de los jugos. Esto podría incluir el aumento en la demanda de ciertos sabores o ingredientes, cambios en las preferencias de los consumidores hacia productos saludables o naturales, tendencias de estilo de vida que afectan las decisiones de compra, etc.

4. **Modelado de pronósticos:** Se desarrollan modelos predictivos utilizando técnicas como series temporales, análisis de regresión u algoritmos de machine learning para predecir la demanda futura de los productos de jugo. Estos modelos pueden tener en cuenta variables como estacionalidad, eventos especiales, campañas de marketing, cambios en la economía, entre otros.

5. **Evaluación de escenarios:** Se realizan análisis de sensibilidad y simulaciones para evaluar cómo diferentes escenarios podrían afectar la demanda de los productos de jugo. Por ejemplo, cómo un aumento en los precios de los ingredientes podría impactar en los márgenes de ganancia, o cómo cambios en la estrategia de marketing podrían influir en las ventas.

6. **Desarrollo de estrategias proactivas:** Con base en los hallazgos del análisis de datos y los pronósticos realizados, la empresa desarrolla estrategias proactivas para mantenerse competitiva en el mercado. Esto podría implicar el lanzamiento de nuevos productos o sabores en línea con las tendencias emergentes, ajustes en la cadena de suministro para responder rápidamente a cambios en la demanda, campañas de marketing dirigidas a segmentos específicos de consumidores, entre otras acciones.

7. **Monitoreo y ajuste continuo:** La empresa monitorea continuamente los resultados de sus estrategias y realiza ajustes según sea necesario en función de la evolución del mercado y los nuevos datos disponibles. Este ciclo de retroalimentación continua ayuda a mantener la competitividad y la relevancia en un entorno empresarial en constante cambio.

Al implementar la metodología propuesta en el análisis de ventas en la empresa de jugos de sabores LEYDE se ofrecerá una serie de beneficios claros. Primero, nos permitirá comprender a fondo las preferencias y necesidades de nuestros clientes, lo que nos permite adaptar nuestro producto y estrategias de marketing de manera más efectiva. Además, al analizar la cadena de valor y la cadena de suministro, podemos identificar áreas de optimización que podrían mejorar la eficiencia operativa y reducir costos.

Mediante el análisis de segmentación y correlación, podemos identificar segmentos de mercado clave y factores que influyen en las ventas, lo que nos permite dirigir estrategias específicas para impulsar el crecimiento. La evaluación de precios y estrategias de marketing nos ayudará a maximizar la rentabilidad y la visibilidad de nuestra marca. En resumen, al aplicar esta metodología de análisis de datos de manera sistemática, podemos tomar decisiones más informadas y estratégicas que conduzcan a un aumento sostenido en las ventas y la competitividad de la empresa en el mercado de jugos de sabores.

Se espera que esta investigación proporcione:

- a) Modelos predictivos precisos para anticipar la demanda de jugos, lo que permitirá a LEYDE mejorar su planificación de producción y estrategias de ventas.
- b) Una segmentación efectiva de clientes, que permitirá diseñar campañas de marketing personalizadas y mejorar la fidelización de los clientes.
- c) Estrategias basadas en datos que optimicen los esfuerzos comerciales y ayuden a revertir la tendencia decreciente de las ventas en la familia de jugos.

De esta manera, la empresa podrá tomar decisiones más informadas y basadas en evidencia, mejorando su competitividad en el mercado.

CAPÍTULO IV. RESULTADOS Y ANÁLISIS

En un mundo empresarial cada vez más competitivo, la capacidad de tomar decisiones estratégicas basadas en datos es fundamental para garantizar el éxito de una organización. La empresa LEYDE, dedicada al comercio de productos lácteos y otros derivados, se enfrenta a una serie de desafíos relacionados con la gestión eficiente de sus recursos y la maximización de sus ventas. Con el auge de la ciencia de datos y las herramientas de análisis avanzado, se abre una oportunidad única para implementar metodologías que permitan optimizar las decisiones comerciales a partir de datos históricos.

El propósito de este capítulo es presentar los resultados de aplicar técnicas de análisis de series de tiempo y clusterización para mejorar las ventas de LEYDE. El análisis de series de tiempo permitirá identificar patrones y tendencias a lo largo del tiempo, brindando información valiosa sobre la estacionalidad, ciclos de demanda y otras variables clave que afectan el rendimiento de la empresa. Por otro lado, la clusterización facilitará la segmentación de clientes y productos, lo que permitirá diseñar estrategias de ventas personalizadas y más efectivas.

Mediante el uso de herramientas avanzadas de análisis de datos, este trabajo tiene como objetivo proporcionar a LEYDE una visión más profunda de su operación, mejorando la toma de decisiones e impulsando sus resultados comerciales. En las siguientes secciones, se describirán los datos utilizados, las técnicas de análisis implementadas y los resultados obtenidos, con el fin de proporcionar recomendaciones concretas que contribuyan al crecimiento sostenible de la empresa.

4.1 EXPLORACION DE DATOS

En el entorno empresarial actual, la capacidad de analizar y aprovechar grandes volúmenes de datos se ha convertido en una ventaja competitiva clave para la toma de decisiones estratégicas. LEYDE, una empresa líder en la producción y comercialización que tiene una baja de su marca de jugos desde los últimos 4 años y se enfrenta al reto de optimizar sus estrategias de ventas para mantenerse competitiva en un mercado cada vez más dinámico y exigente.

Usando técnicas avanzadas de análisis de datos, específicamente el Análisis Exploratorio de Datos (EDA) y el apoyándose del uso de series de tiempo y métodos de clusterización, para identificar patrones de comportamiento de ventas y segmentar adecuadamente el mercado. El enfoque en series de tiempo permitirá modelar las tendencias y estacionalidades que influyen en la demanda de los jugos, mientras que la clusterización proporcionará una mejor comprensión de los distintos segmentos de clientes o puntos de venta, facilitando así la personalización de estrategias comerciales.

Mediante este enfoque, se busca no solo mejorar la toma de decisiones en la planificación de ventas y marketing, sino también contribuir a una asignación más eficiente de los recursos y al desarrollo de estrategias dirigidas a maximizar las oportunidades de crecimiento y rentabilidad en el mercado hondureño.

Los Insight destacados tenemos un total de 117 meses desde Enero del 2014 a Septiembre del 2024 con 1404 registros de unidades vendidas en las 12 sucursales que tiene la empresa LEYDE en todo el país. Su año de mayor venta es el año 2021 con 17,944,903 unidades vendidas de jugos en sus diferentes presentaciones y sabores.

Además de las 12 sucursales que existen, 3 de ellas son las que más venden siendo las ciudades principales de San Pedro Sula, Tegucigalpa y Ceiba donde actualmente está la planta de procesamiento. A continuación, mostramos dos gráficos preliminares para evaluar los picos y tendencias que muestra la Data de las ventas:



Figura 2. Se observa una tendencia aditiva en las series de tiempos y luego una caída de la tendencia a partir del 2021

Fuente: Elaboración Propia

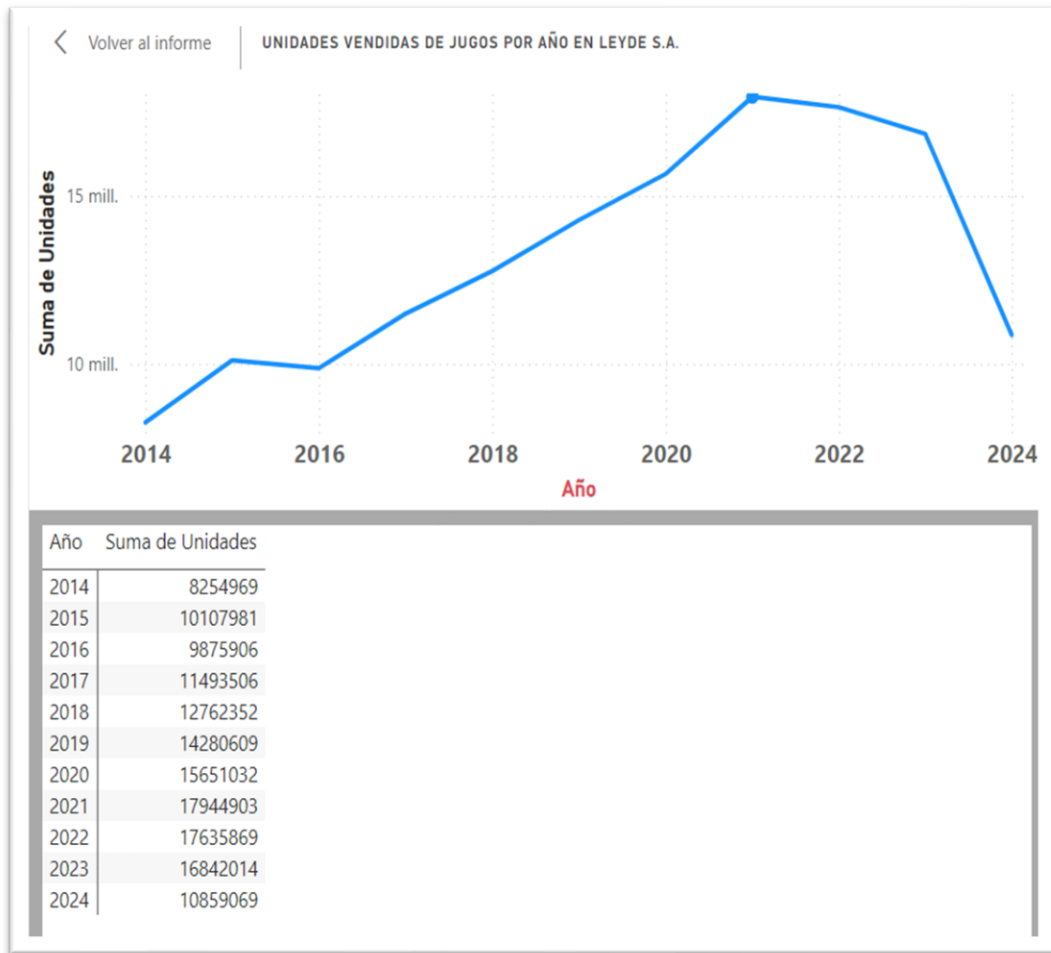


Figura 3. Datos de las ventas en unidades de Jugos por parte de la empresa LEYDE S.A. de Honduras desde el 2014 a la fecha de Septiembre del 2024

Fuente: Elaboración Propia

Dentro de las 11 sucursales que se entrega productos desde la planta de procesamiento en Ceiba tenemos las siguientes:

1. San Pedro sula
2. Tegucigalpa
3. La Ceiba
4. Tocoa Colon
5. Santa Rosa de Copan
6. Roatán
7. Comayagua
8. Siguatepeque

9. Santa Barbara

10. Choluteca

11. Olancho

A continuación, evaluaremos el gráfico de las 3 sucursales que venden el 84% de las ventas totales:

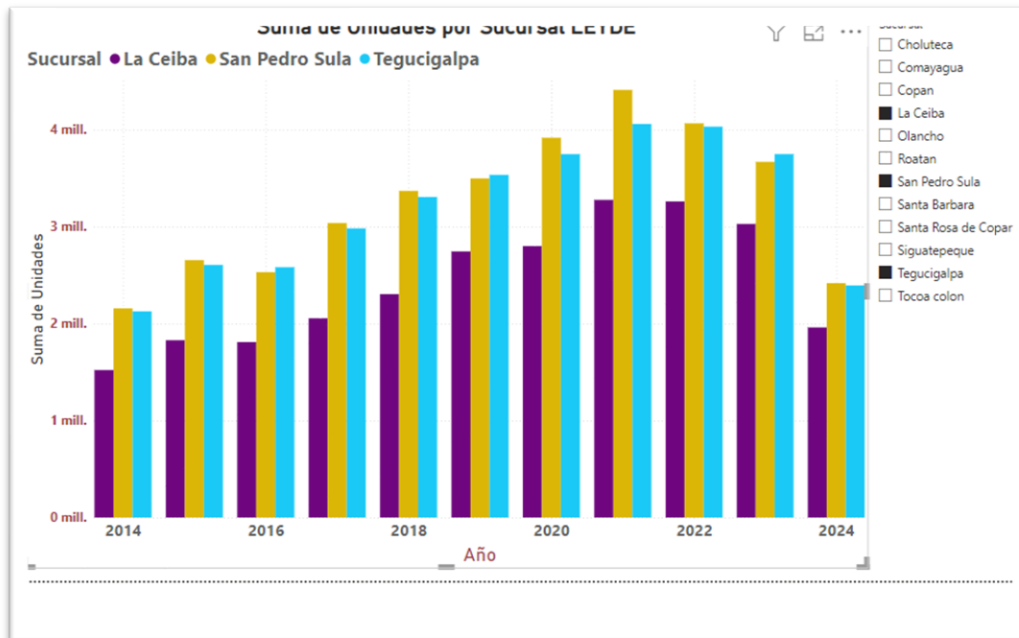


Figura 4 Observamos el comportamiento de las 3 sucursales de mayor venta en el país de las 11 que se tienen por parte de la empresa LEYDE S.A.

Fuente: Elaboración Propia

Luego observamos una caída de las ventas desde el año 2021 al presente de una forma preocupante de un 30% aproximadamente.

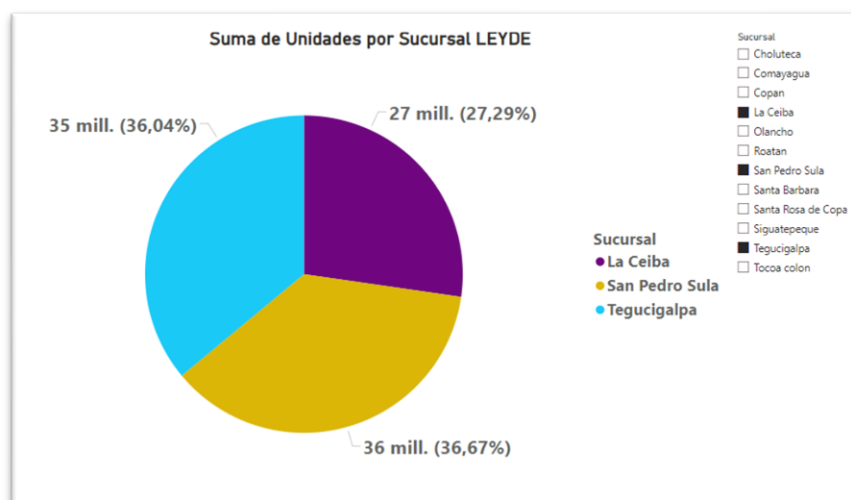


Figura 5. Observamos el comportamiento de las 3 sucursales de mayor venta en el país de las 12 que se tienen por parte de la empresa LEYDE S.A. en miles de millones en un gráfico de Pastel.

Fuente: Elaboración Propia

4.2 RESULTADOS Y ANALISS DE LAS TECNICAS APLICADAS

Aquí se realizó una evaluación exhaustiva de los datos históricos de clientes de los últimos 10 años con el fin de analizar la disminución de las ventas en los últimos años. El objetivo principal de esta investigación fue identificar estrategias para mejorar los KPI's relacionados con las ventas de jugos procesados de la empresa LEYDE. A partir del análisis de estos datos, se definieron objetivos claros que guiarán la propuesta presentada, y se estableció un plan para su desarrollo. En términos de aplicabilidad, se sugieren diversas acciones encaminadas a optimizar el rendimiento de las ventas, con un enfoque basado en la clusterización de clientes y la proyección de series de tiempo, lo que permitirá una toma de decisiones más informada y estratégica.

Dentro del objetivo general que se planteó de Aplicar técnicas de Machine Learning para analizar y predecir las ventas de la línea de jugos de LEYDE, con el fin de optimizar los indicadores de ventas y desarrollar estrategias de marketing y fidelización más efectivas

El Análisis de la Data basado en el desarrollo de los algoritmos que se implementaron en las ventas por diez años de la empresa LEYDE en su producto o familia de jugos procesados dará los siguientes alcances para el departamento de Ventas y Marketing:

4.2.1. APLICABILIDAD DE LA SEGMENTACIÓN POR CLÚSTER

El uso del análisis de clusterización que se desarrollo es extremadamente útil para mejorar la gestión de las sucursales de manera diferenciada. La empresa puede aplicar esta metodología para la:

- Personalización de estrategias de marketing: Adaptar campañas publicitarias y promociones específicas para los diferentes clústeres, logrando mayor efectividad al enfocar los esfuerzos en las características y comportamientos de compra de cada grupo de sucursales.
- Optimización de recursos: Asignar recursos, como personal o inventario, de manera más eficiente, priorizando las sucursales con mayor potencial de ventas o crecimiento.
- Mejora en la toma de decisiones: Facilitar decisiones sobre que inventario de jugos debemos tener en las sucursales basadas en el rendimiento específico de cada clúster.

4.2.2. APLICABILIDAD DEL ANÁLISIS DE LA SERIE DE TIEMPO Y RANDOM FOREST

El uso de series de tiempo para el pronóstico de ventas proporciona herramientas precisas para:

- Planificación de la demanda: La empresa puede planificar con mayor precisión la producción y distribución de jugos, reduciendo costos por exceso o falta de inventario, ajustando la oferta a la demanda prevista.
- Optimización de campañas estacionales: Utilizar los patrones estacionales identificados para lanzar campañas promocionales en los meses de mayor demanda, maximizando el retorno sobre la inversión.
- Ajuste de precios y ofertas: Durante las temporadas de baja demanda, la empresa puede ajustar los precios o lanzar promociones específicas para estabilizar las ventas y evitar caídas drásticas.

4.2.3. APLICACIÓN DE ESTRATEGIAS APLICADAS POR CLÚSTER

- Adaptación de portafolio de productos: La empresa puede ajustar el portafolio de productos en las diferentes sucursales según las preferencias regionales identificadas en cada clúster. Por ejemplo, introducir o retirar ciertos productos dependiendo de la zona o preferencia del consumidor.
- Fidelización y retención de clientes: Implementar programas de fidelización diferenciados por clúster para mantener a los clientes actuales y convertir los picos temporales de ventas en una tendencia sostenida.

4.2.4. APLICACIÓN DEL REBRANDING Y REPOSICIONAMIENTO

En respuesta al deterioro de la imagen de la marca y la caída en las ventas, la empresa puede aplicar:

- Redefinición del mensaje de la marca: Un relanzamiento de la marca con una estrategia de marketing que refleje los valores actuales del consumidor (por ejemplo, productos más saludables, sostenibles) puede mejorar la percepción del producto.
- Ajustes en el empaque o diseño de productos: Considerar mejoras en el empaque para

atraer a un segmento de consumidores que esté más enfocado en la estética o en productos ecológicos.

4.2.5. APLICACIÓN DE ESTRATEGIAS DE EXPANSIÓN O RESTRUCTURACIÓN

- Cierre o reestructuración de sucursales: Basándose en el análisis de clusterización, la empresa puede cerrar sucursales no rentables o reasignar recursos a aquellas que muestran mayor potencial.
- Identificación de nuevas áreas de expansión: Usar los datos de ventas y clusterización para identificar posibles áreas geográficas con potencial de apertura de nuevas sucursales o expansión de la presencia de la marca.

4.2.6. APLICACIÓN DE CONTROL DE INVENTARIOS

- Gestión eficiente de inventarios: Con el uso de pronósticos basados en series de tiempo, la empresa puede ajustar sus niveles de inventario a la demanda proyectada para cada clúster y temporada. Esto evitará sobreproducción o desabastecimiento, optimizando los costos de almacenamiento y logística.

4.2.7. APLICACIÓN DE ESTRATEGIAS DIGITALES

- Expansión de canales de venta online: Aprovechar las tendencias del comercio electrónico implementando estrategias digitales (por ejemplo, ventas en línea, publicidad en redes sociales) para diversificar los canales de distribución y captar a consumidores que prefieren hacer compras en línea.
- Marketing digital segmentado: Desarrollar campañas de marketing digital dirigidas específicamente a los consumidores de cada clúster, utilizando publicidad en redes sociales, correos electrónicos personalizados y estrategias de remarketing basadas en el comportamiento de compra.

4.2.8. APLICACIÓN EN LA PLANIFICACIÓN FINANCIERA

- Presupuestación y asignación de recursos: Basándose en las proyecciones de ventas y el rendimiento de cada clúster, la empresa puede planificar sus presupuestos de

manera más eficiente, asignando mayores recursos a las áreas o sucursales con mejor proyección y ajustando costos en zonas con bajo rendimiento.

4.2.9. USO DE INDICADORES DE DESEMPEÑO BASADO EN CLUSTERES

- KPI's por clúster: Definir indicadores de rendimiento específicos para cada clúster de sucursales permitirá medir con mayor precisión la efectividad de las estrategias implementadas y ajustar decisiones en tiempo real.

4.2.10. APLICACIÓN DE ESTRATEGIAS REGIONALES Y ECONÓMICAS

- Ajuste a condiciones económicas locales: La empresa puede adaptar sus estrategias a las condiciones económicas y preferencias de cada región, implementando precios diferenciados o promociones que respondan a las particularidades económicas de los clústeres identificados.

La aplicabilidad de los métodos de clusterización, series de tiempo y las estrategias derivadas del análisis en la empresa LEYDE está basado en la metodología CRISP-DM que es crucial para que la empresa tome decisiones informadas, optimice sus operaciones y recupere su posicionamiento en el mercado de jugos y se muestra en una tabla a continuación. Esto implica personalizar estrategias a nivel regional, ajustar el portafolio de productos, mejorar la planificación financiera y logística, y adaptar la marca a las nuevas exigencias del consumidor. La correcta implementación de estas recomendaciones ayudará a contrarrestar la tendencia negativa en ventas, mejorar la eficiencia operativa y fortalecer el crecimiento de la empresa.

Tabla 4. Plan de desarrollo basados en la metodología CRISP-DM

Fases	Tareas	Entradas	Salidas
Comprensión del Negocio	<ul style="list-style-type: none"> Evaluación de la situación. 	Información sobre la empresa	Análisis de la información del negocio.

	<ul style="list-style-type: none"> Objetivos del negocio. 		
	<ul style="list-style-type: none"> Plan del proyecto 		

Compresión de los datos	<ul style="list-style-type: none"> Recopilación de los datos. 	Recolección inicial de los datos de clientes y transacciones de la empresa.	Obtención de la base de datos con las que se va a trabajar.
	<ul style="list-style-type: none"> Descripción de los datos. 		
	<ul style="list-style-type: none"> Exploración de los datos. 		
	<ul style="list-style-type: none"> Verificación de la calidad de los datos 		
Preparación de los datos	<ul style="list-style-type: none"> Selección de datos a analizar (muestreo). 	Preprocesamiento de los datos proporcionados	Datos seleccionados y preparados para aplicar la Minería de Datos.
	<ul style="list-style-type: none"> Limpieza de los datos. 	por la empresa.	
	<ul style="list-style-type: none"> Integración y formato de los datos. 		
Modelado	<ul style="list-style-type: none"> Seleccionar las técnicas de modelado. 	Desarrollo de técnica Clustering y Series de Tiempos.	Generación del modelo de clusterización y series de tiempos.
	<ul style="list-style-type: none"> Diseñar un modelo de comprobación. 		
	<ul style="list-style-type: none"> Construir el modelo. 		

	<ul style="list-style-type: none"> • Evaluar el modelo. 		
Evaluación	<ul style="list-style-type: none"> • Evaluar los resultados obtenidos. 	Modelado de Minería de Datos.	Evaluación e interpretación de los resultados obtenidos.

Fuente: Elaboración Propia

4.3 RESULTADOS CUANTITATIVOS

A continuación, se hará una descripción de todas las herramientas, instrumentos y procesos utilizados para el desarrollo de la propuesta, así como los algoritmos desarrollados para su análisis específico:

Las Herramientas utilizadas en este proyecto son de Open Source por lo tanto no tienen costo:

POWER BI: Es una plataforma empresarial integral y expansible en el ámbito de la inteligencia de negocios (BI), caracterizada por atributos de autoservicio que resultan idóneos para empresas de gran envergadura. Establece una conexión con los datos, proporciona capacidades visuales y facilita la integración fluida de elementos visuales en las aplicaciones de uso cotidiano.



Figura 6. Logotipo Power BI.

Fuente: (fuente específica no válida)

Knime: representa una vanguardia en la categoría emergente de herramientas identificadas por Gartner como Plataformas de Ciencia de Datos y Aprendizaje Automático.

Estas herramientas facultan a profesionales de la ciencia de datos, analistas y usuarios empresariales para interactuar con sus datos, dando lugar a la creación, implementación y gestión de modelos de análisis avanzado.

En cuanto a sus destacadas características, se puede resaltar su facilidad de uso. La programación de aplicaciones en KNIME se torna altamente intuitiva gracias a una interfaz visual. La conexión visual de nodos que encapsulan diversas funciones, junto con la integración de módulos automatizados de Aprendizaje Automático y modelos predefinidos, simplifica la analítica avanzada para usuarios empresariales que carecen de experiencia en este ámbito.

Simultáneamente, proporciona un entorno óptimo para desarrolladores avanzados al permitir la integración de programación en Python o R.



Figura 7. Logotipo Knime.

Fuente: (fuente específica no válida)

La recolección de los datos las ventas históricas en las 11 sucursales desde el año 2014 a Septiembre del 2014 fue proporcionada por el personal de Planning de la empresa LEYDE S.A. de Honduras.

Tabla 5. Ejemplo de la Data de Ventas Históricas de LEYDE del producto filtrado Jugo de Naranja

Mes	Sucursal	Producto	Unidades

Fuente: Elaboración Propia

Hacemos una descripción de los datos de la tabla a continuación:

- Sucursal: el cual comprende el nombre del cliente para la planta procesadora que sería la sucursal en el mapa de Honduras, el nombre de la sucursal es el mismo nombre de la ciudad donde radica, como municipio de Honduras importante o departamento.
- Producto: en este caso la data esta filtrada por Jugos de Naranja que es la familia de productos que nos interesa.
- Unidades: son las unidades que se venden por cada mes del año.
- Mes: es el Año y mes que se vendió y va en un rango desde enero del 2014 a septiembre de 2024.

4.3.1. ALGORITMO DE CLUSTERIZACION

Haremos una descripción del uso del Algoritmo en el software o herramienta que se desarrolló en el programa de Knime usando el nodo de K-media para la Data y así agrupar las sucursales de acuerdo a su venta histórica de diez años; para la integración y formato de los datos se hizo una serie de nodos y pasos que se detallan a continuación:

Para ello haremos de primera instancia una transformación y Preparación de los Datos:

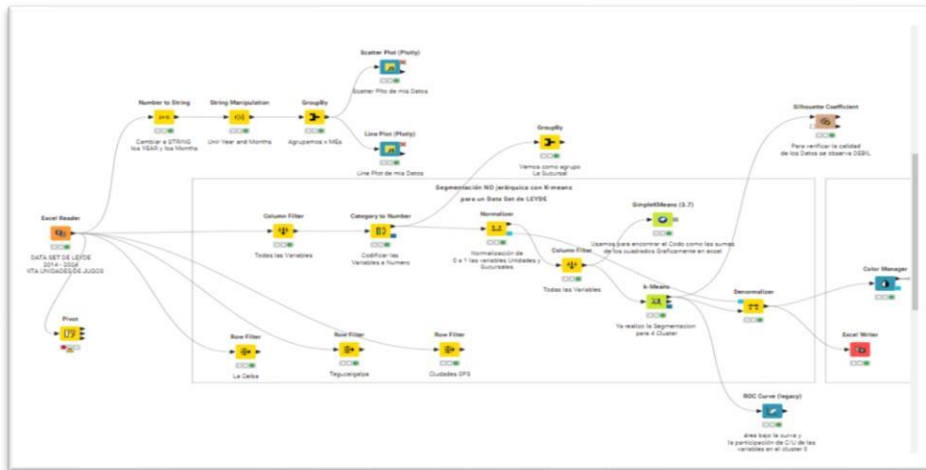


Figura 8. Workflow de Transformación y Preparación realizado en knime.

Fuente: Elaboración Propia

En este paso se usaron varios nodos para revisar de forma exploratoria siendo el más influyente el Nodo de GroupBy que se muestra a continuación para revisar la tendencia de la gráfica y revisión de la data histórica en un nodo Scatter Plot y el de Line Plot.



Primeramente usamos el Nodo Agrupar Por (GroupBy) :

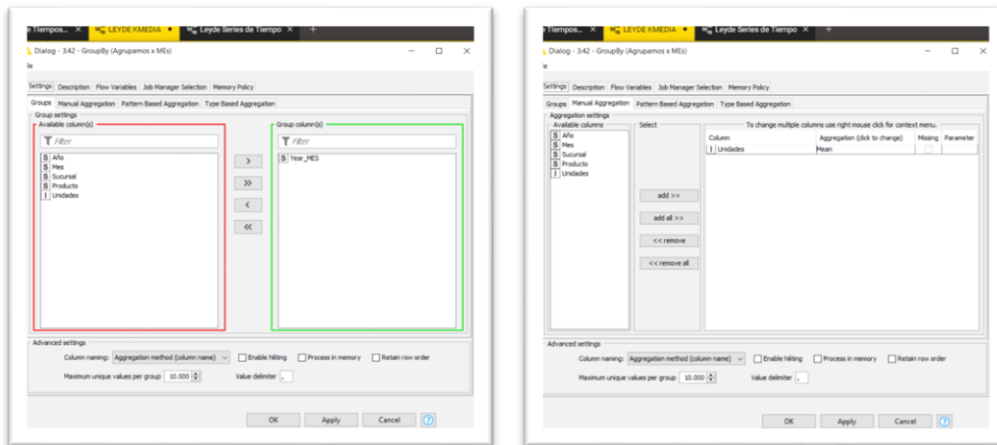


Figura 9. Uso de nodo de agrupamiento de datos previo al gráfico de Scatter Plot y el de Line Plot.

Fuente: Elaboración Propia

Luego observamos en la exploración de los datos usando nuestros nodos de graficos un diagrama de dispersión y un diagrama de línea donde vemos el histórico de mes a mes de los últimos diez años de la venta de Jugos procesados con sus picos y caídas y haciendo referencia a una serie de tiempo con tendencia y aditiva.

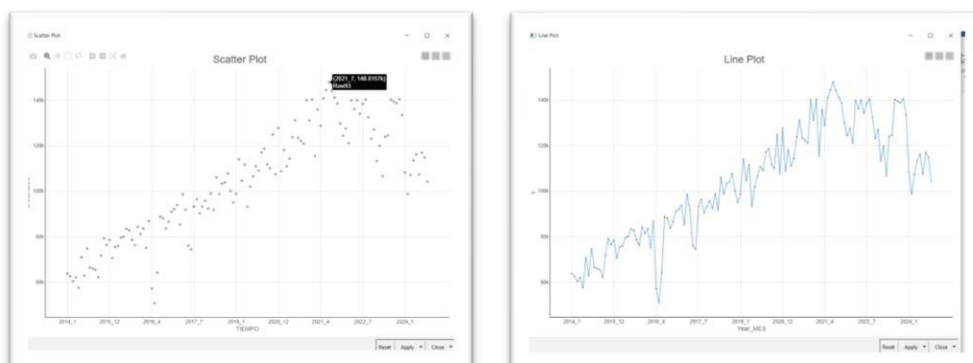


Figura 10. Revisión de los datos en un nodo de Scatter Plot y en un nodo de Line Plot.

Fuente: Elaboración Propia

Aquí se puede observar a priori que hay una tendencia o una dispersión de datos desde el 2014 hasta el 2021 de forma creciente y luego decrece con picos de subidas y bajadas hasta el 2024.

Para prepararnos para el uso del Nodo K Means debemos usar una serie de nodos de preparación y verificación que detallaremos a continuación:

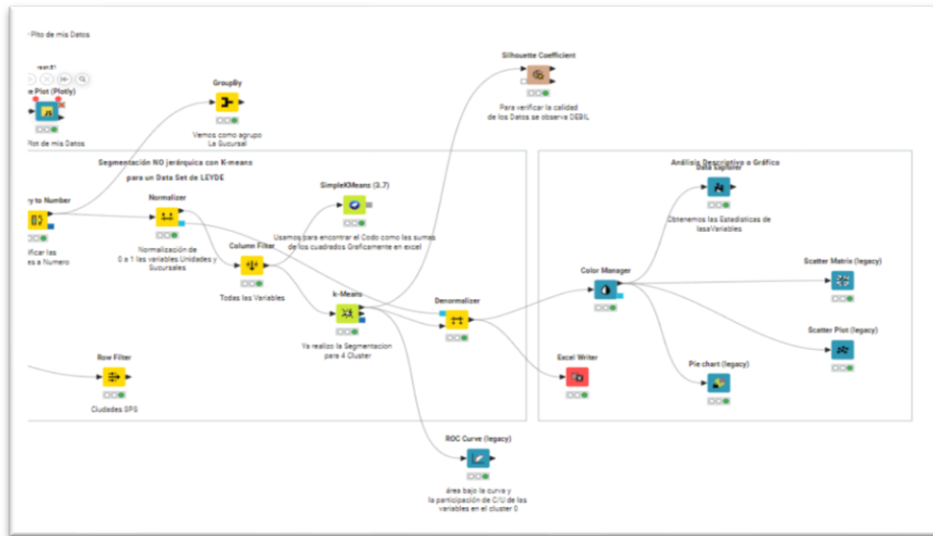


Figura 11. Revisión de nodos que elaboran el agrupamiento por kMeans.

Fuente: Elaboración Propia

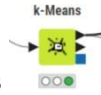
A continuación, detallamos los Nodos relevantes:

1. **Nodo Normalizer**  :

Este nodo se utiliza para normalizar o escalar los datos, ajustando las características a un rango común, generalmente entre 0 y 1. Esto es especialmente útil para algoritmos de aprendizaje automático que dependen de la magnitud de las variables, como K-Means, ya que garantiza que cada característica tenga una contribución equitativa en los cálculos de distancia.

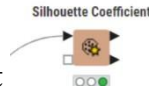
2. **Nodo Simple K-Means 3.7**  :

Implementa el algoritmo de agrupamiento K-Means, que asigna puntos de datos a los grupos o "clústeres" basados en la similitud (medida habitualmente con la distancia euclidiana). La versión Simple K-Means 3.7 es una versión específica de este algoritmo, optimizada para KNIME.



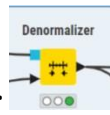
3. **Nodo K-Means** :

Similar al nodo anterior, pero puede ofrecer configuraciones adicionales o una implementación más avanzada del algoritmo de K-Means. Permite segmentar los datos en varios clusters de acuerdo con patrones o características comunes.



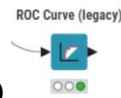
4. **Nodo Silhouette Coefficient** :

Este nodo calcula el coeficiente de silueta, una métrica que mide qué tan bien se encuentran los datos agrupados en sus respectivos clusters. Los valores cercanos a 1 indican que los puntos están bien agrupados, mientras que los valores negativos sugieren mala agrupación o asignación incorrecta de puntos.



5. **Nodo Denormalizer** :

Revertir la normalización aplicada previamente. Este nodo es útil para devolver los datos a su escala original después de haber aplicado el análisis o procesamiento, como los resultados de un algoritmo de clustering.



6. **Nodo ROC Curve (legacy)** :

Se utiliza para evaluar el rendimiento de un modelo de clasificación binaria. La curva ROC (Receiver Operating Characteristic) traza la tasa de verdaderos positivos frente a la tasa de falsos positivos. Este nodo es una versión "legacy", lo que significa que puede ser una versión funcional antigua, pero aún funcional, para analizar modelos predictivos en KNIME.

Estos nodos son esenciales para el preprocesamiento, análisis y evaluación de datos en flujos de trabajo de clustering. (Knime Web Site, n.d.)

A continuación, revisamos la parte descriptiva de la información y hacemos una salida de una data de Excel ya categorizada en 4 clúster para orientar a la fuerza de ventas y mercadeo a implementar decisiones acordes a sus características y patrones de relevancia.

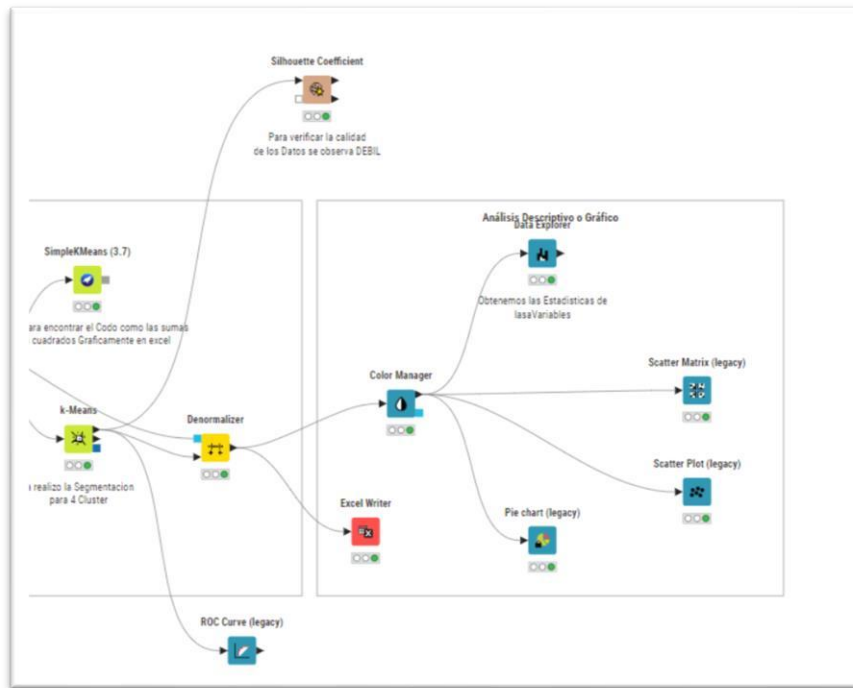
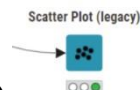


Figura 12. Revisión de nodos para revisar la parte descriptiva del algoritmo Kmeans.
Fuente: Elaboración Propia

Para revisar estos nodos descriptivos del algoritmo de K-Means vamos a presentar lo siguiente:

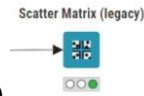


1. El **nodo Scatter Plot (Legacy)** en KNIME es una versión más antigua del gráfico de dispersión. Este nodo se utiliza para visualizar la relación entre dos variables numéricas, representando los puntos de datos como coordenadas en un plano cartesiano.

El gráfico de dispersión es útil para:

- **Detectar relaciones entre variables:** Permite ver si existe una correlación positiva, negativa o nula entre dos características.
- **Identificar patrones:** Como agrupamientos, outliers o tendencias lineales/no lineales.
- **Explorar distribuciones:** Puedes observar cómo se distribuyen los datos en el espacio en función de las dos variables seleccionadas.

Aunque el nodo es "legacy" (antiguo), sigue siendo útil si prefieres una versión más básica o si estás trabajando con una versión antigua de KNIME que aún lo soporte. Sin embargo, puede carecer de algunas de las funciones avanzadas o mejoras gráficas de nodos más recientes, como personalización o interactividad mejorada.

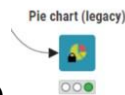


2. El **nodo Scatter Matrix (Legacy)** en KNIME permite visualizar una **matriz de gráficos de dispersión** para múltiples variables. A diferencia de un gráfico de dispersión simple que compara solo dos variables, una matriz de dispersión muestra todas las combinaciones posibles de pares de variables dentro de un conjunto de datos.

Este nodo es útil para:

- **Analizar relaciones entre varias variables simultáneamente:** Puedes observar todas las posibles correlaciones entre las variables en un solo gráfico.
- **Identificar patrones globales:** Como correlaciones lineales o no lineales, tendencias, o agrupamientos entre más de dos variables.
- **Explorar interacciones:** Entre las variables y cómo estas pueden estar relacionadas en pares. Cada celda de la matriz muestra un gráfico de dispersión para un par de variables, lo que facilita la comparación entre todas las variables en el dataset.

Al ser "legacy", puede ser una versión anterior del nodo actual de Scatter Matrix en KNIME, por lo que puede no tener algunas características avanzadas como la interactividad o personalización gráfica más moderna, pero sigue siendo funcional para exploraciones visuales básicas

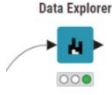


3. El **nodo Pie Chart (Legacy)** en KNIME genera un gráfico de torta o pastel (pie chart) que se utiliza para mostrar la proporción o porcentaje de cada categoría dentro de un conjunto de datos. Cada "rebanada" del gráfico representa una categoría, y el tamaño de la rebanada es proporcional a la magnitud o frecuencia de esa categoría en relación con el total.

Este nodo es útil para:

- **Visualizar distribuciones categóricas:** Permite ver cómo se distribuyen las distintas categorías dentro de un grupo de datos, facilitando la comparación entre ellas.
- **Mostrar proporciones:** Es ideal para mostrar las participaciones relativas de cada categoría respecto al total.
- **Comprender datos simples:** Es una herramienta visual intuitiva para representar datos categóricos.

Si bien este nodo es una versión "legacy", lo que significa que es antiguo, sigue funcionando bien para crear gráficos de torta simples. Sin embargo, puede carecer de características más avanzadas presentes en versiones modernas, como personalización detallada o interactividad.

4. El **nodo Data Explorer**  en KNIME se utiliza para realizar un análisis exploratorio y descriptivo de un conjunto de datos, generando una vista gráfica y resumida de las características de las variables. Este nodo es útil para obtener una comprensión rápida de la distribución y las propiedades de los datos antes de realizar cualquier procesamiento adicional o modelado.

Las principales características del **nodo Data Explorer** incluyen:

- **Estadísticas descriptivas:** Muestra métricas como el promedio, mediana, desviación estándar, valores mínimos y máximos, y otros resúmenes estadísticos para variables numéricas.
- **Distribución de datos:** Genera histogramas automáticos para cada variable numérica, lo que permite visualizar la distribución de los datos y detectar sesgos o patrones inusuales.
- **Manejo de valores faltantes:** Indica la cantidad de valores faltantes en cada variable, lo cual es útil para identificar problemas de calidad de datos.
- **Gráficos de barras:** Para variables categóricas, el nodo muestra gráficos de barras que ilustran la frecuencia de las diferentes categorías.
- **Visualización conjunta:** Proporciona una vista completa de todas las variables del dataset en un solo informe, facilitando la comparación entre variables y permitiendo identificar outliers o patrones anómalos.

Este nodo es muy valioso en las etapas iniciales del análisis de datos, ya que permite explorar y conocer el comportamiento de las variables de manera rápida y efectiva.

4.3.2 ALGORITMO DE SERIES DE TIEMPOS

A Continuación, haremos una descripción del Algoritmo que se desarrolló en el programa de Knime usado para la Data y así realizar los pronósticos de Series de Tiempo:

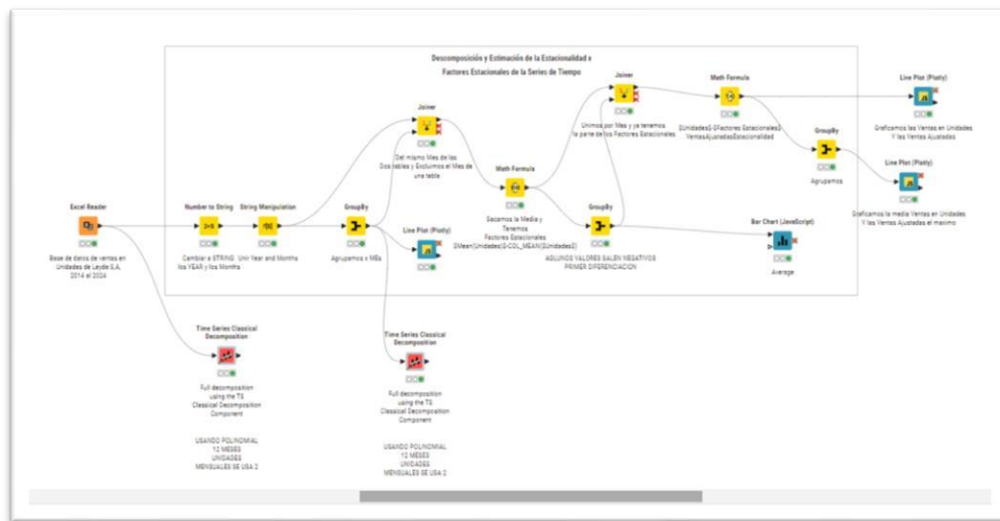



Figura 13. Algoritmo de los Nodos para descomponer las Series de Tiempos de las Ventas.

Fuente: Elaboración Propia

Aquí utilizamos varios nodos en las series de tiempo que explicaremos a continuación:

- El nodo **Time Series Classical Decomposition**  en KNIME se utiliza para descomponer una serie temporal en sus componentes fundamentales: tendencia, estacionalidad y componente residual o aleatorio. Esta técnica es muy útil cuando se trabaja con datos que muestran patrones de estacionalidad o una tendencia subyacente.

Componentes del modelo de reparación clásica:

1. Tendencia (Trend): Representa la dirección general de los datos a lo largo del tiempo (creciente, decreciente o estable). Este componente se obtiene eliminando las fluctuaciones estacionales de la serie.
2. Estacionalidad (Seasonality): Este componente captura los patrones repetitivos en la serie temporal que ocurren en intervalos regulares (diarios, semanales, mensuales, etc.).
3. Residual (Random/Noise): Es el componente aleatorio que no está explicado ni por la tendencia ni por la estacionalidad. Representa las variaciones no explicadas en los datos.

Funcionamiento del nodo en KNIME:

Configuración del nodo:

1. Selección de columnas (Selección de columnas): Se especifica la columna que contiene la serie temporal a descomponer y en nuestro caso tenemos el atributo de ventas de unidades por mes en LEYDE.
2. Season length (Duración de la estacionalidad): Define el número de periodos por ciclo de estacionalidad (sería 12 para datos mensuales con un ciclo anual en el caso de la data de la empresa LEYDE).
3. Tipo de modelo (Tipo de modelo): Puedes elegir entre el modelo aditivo o multiplicativo (el modelo de la serie de tiempo que presenta la data de LEYDE es aditivo).
4. Método de descomposición (Método de análisis): Permite seleccionar si se desea extraer solo la tendencia, solo la estacionalidad, o ambas por medio de la Media Móvil o la Polinomial.

Resultados del nodo:

El nodo nos genera columnas adicionales en la tabla de salida que contienen cada uno de los componentes descompuestos: tendencia, estacionalidad y residuos. Estos resultados nos permiten analizar más a fondo el comportamiento de la serie, identificar patrones subyacentes.

Este nodo es especialmente útil para descomponer una serie temporal antes de aplicar otros métodos, como predicción con ARIMA, suavizado exponencial, o técnicas de aprendizaje automático.

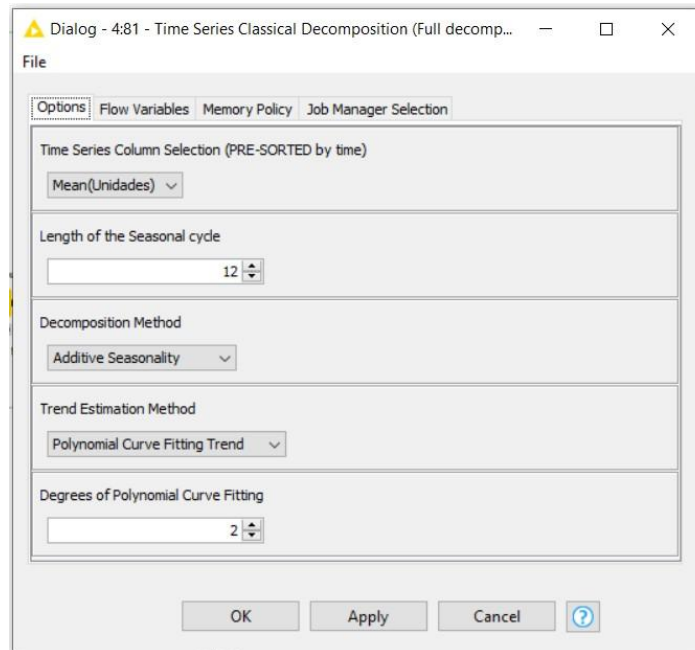
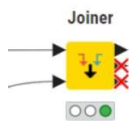


Figura 14. Configuración del Nodo de Time Series Classical Decomposition.

Fuente: Elaboración Propia



- El nodo Joiner en KNIME se utiliza para combinar dos tablas de datos con base en una o más columnas comunes (similares a lo que hace una unión o "join" en bases de datos SQL). Es muy útil cuando necesitas combinar información de diferentes fuentes en un solo conjunto de datos.

Tipos de uniones:

El nodo Joiner permite realizar varios tipos de uniones (joins), que se basan en cómo se desea combinar las filas de ambas tablas según la coincidencia de las columnas clave:

1. Inner Join (Unión Interna):
 - Solo se incluyen las filas que tienen coincidencias en ambas tablas.
 - Las filas que no coinciden en ambas tablas son descartadas.
 - Fórmula matemática: $R=A \cap B$ Donde AAA es la primera tabla y BBB es la segunda tabla, y RRR es el resultado de la intersección de ambas tablas en las columnas clave.
2. Left Outer Join (Unión Externa Izquierda):
 - Incluye todas las filas de la primera tabla (izquierda), y agrega los valores coincidentes de la segunda tabla (derecha). ○ Si no hay coincidencia, los valores faltantes de la segunda tabla se rellenan con valores nulos.
 - Fórmula matemática: $R=A \cup (A \cap B)$ Se conservan todas las filas de AAA (izquierda) y se añaden coincidencias de BBB.
3. Right Outer Join (Unión Externa Derecha):
 - Incluye todas las filas de la segunda tabla (derecha), y agrega los valores coincidentes de la primera tabla (izquierda). ○ Si no hay coincidencia, los valores faltantes de la primera tabla se rellenan con valores nulos.
 - Fórmula matemática: $R=B \cup (A \cap B)$ Se conservan todas las filas de BBB (derecha) y se añaden coincidencias de AAA.
4. Full Outer Join (Unión Externa Completa):
 - Incluye todas las filas de ambas tablas.
 - Las filas que no coinciden en ambas tablas se completan con valores nulos para la tabla que no tiene la coincidencia.
 - Fórmula matemática: $R=A \cup B$ Esto conserva todas las filas de AAA y BBB, llenando los valores faltantes con nulos.

Configuración del nodo Joiner:

Al configurar el nodo Joiner, estos son los pasos más importantes:

1. Table Input (Entradas de tabla):

- El nodo tiene dos entradas: la tabla izquierda (la primera tabla) y la tabla derecha (la segunda tabla) que deseas unir.
2. Column Matching (Coincidencia de columnas):
 - Selecciona las columnas de ambas tablas que deben coincidir para realizar la unión. Estas columnas actúan como las claves en las uniones.
 - Puedes unir más de una columna.
 3. Join Mode (Modo de unión):
 - Elige el tipo de unión que quieres realizar: Inner Join, Left Outer Join, Right Outer Join, o Full Outer Join.
 4. Column Selection (Selección de columnas):
 - Después de la unión, puedes seleccionar qué columnas de cada tabla se incluirán en el conjunto de datos final.

El nodo Joiner genera una tabla combinada que contiene las columnas seleccionadas de ambas tablas según el tipo de unión especificado. Este nodo es extremadamente útil cuando se trabaja con datos que provienen de múltiples fuentes o para análisis que requieren la integración de diferentes conjuntos de datos.

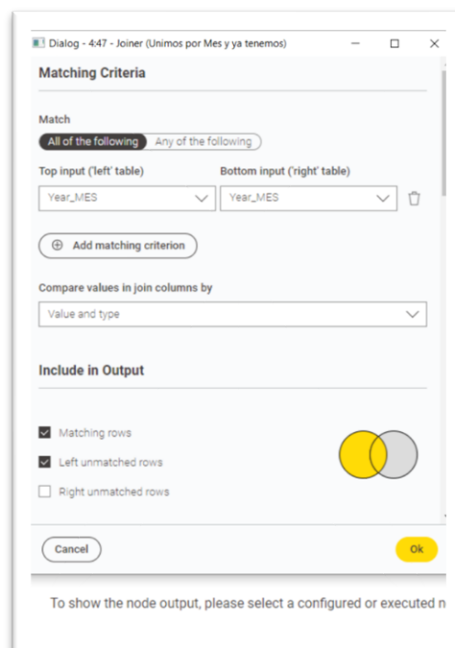


Figura 15. Configuración del Nodo de Joiner.

Fuente: Elaboración Propia



- El nodo Math Formula en KNIME es una herramienta que nos

permite realizar cálculos matemáticos y lógicos en una o más columnas de datos. Este nodo es muy versátil y se utiliza comúnmente para transformar, manipular o crear nuevas columnas a partir de expresiones matemáticas y funciones.

Funcionalidad principal del nodo:

- Podemos escribir fórmulas matemáticas usando operadores, funciones matemáticas, y condiciones lógicas para operar sobre las columnas existentes de tu tabla de datos.
- El resultado de la fórmula se almacena en una nueva columna, o puedes sobrescribir una columna existente.

Componentes y configuración del nodo:

1. Column Selection (Selección de columnas):

- Las columnas sobre las que vas a operar se especifican dentro de la fórmula.

Puedes referenciar cualquier columna del conjunto de datos por su nombre.

2. Math Expression (Expresión matemática):

- En esta sección defines la fórmula que deseas aplicar. Puedes utilizar una amplia gama de operaciones matemáticas, como:

- Operaciones aritméticas: +, -, *, /
- Funciones matemáticas: $\sin(x)$, $\cos(x)$, $\log(x)$, $\exp(x)$, $\text{pow}(x,y)$
- Condicionales: $\text{if}(\text{condition}, \text{true_value}, \text{false_value})$ (útil para aplicar condiciones lógicas).
- Funciones estadísticas: $\text{min}()$, $\text{max}()$, $\text{mean}()$, $\text{median}()$
- Constantes matemáticas: pi, e
- Funciones de redondeo: $\text{floor}(x)$, $\text{ceil}(x)$, $\text{round}(x, d)$ (redondea al número de decimales d).

3. Agregar una nueva columna o sobrescribir una existente:

- Puedes especificar el nombre de la nueva columna que contendrá el resultado de la fórmula.
- Si prefieres sobrescribir una columna

existente con los resultados de la fórmula, puedes seleccionarla para que los nuevos valores reemplacen los antiguos.

4.3.3 ALGORITMO DE RANDOM FOREST

En KNIME, los nodos para implementar el algoritmo de Random Forest se organizan en un flujo como se muestra en la siguiente figura 16 que incluye la preparación de datos, la configuración del modelo, el entrenamiento y la evaluación de los resultados. Aquí tienes una descripción de los nodos clave que puedes utilizar:

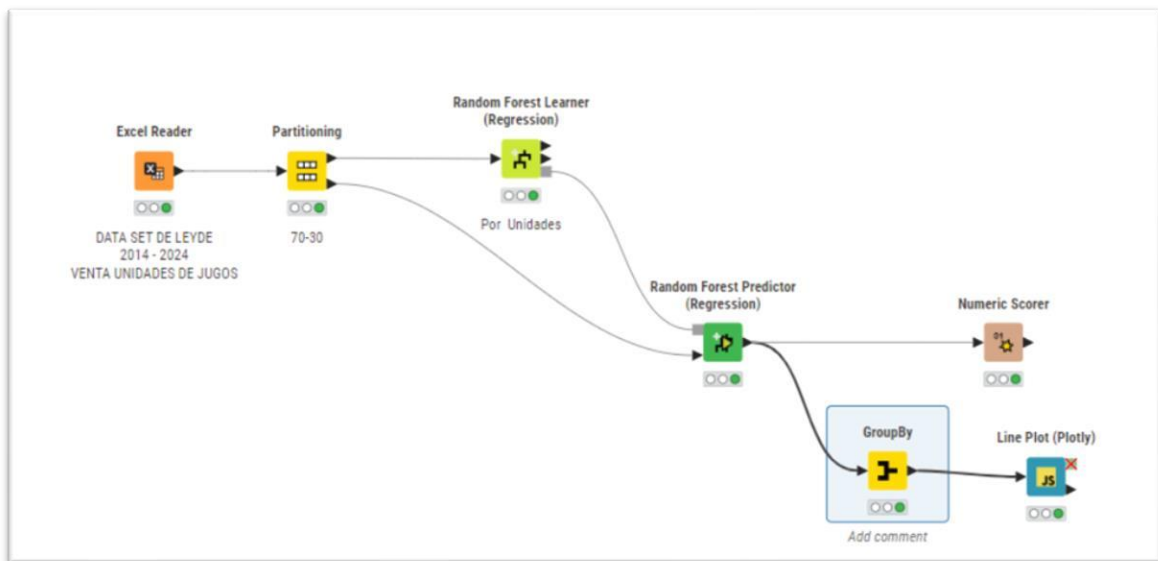


Figura 16. Workflow para nuestro Bosque Aleatorio.

Fuente: Elaboración Propia

1. File Reader o CSV Reader

- Para comenzar, es necesario importar los datos. Usa el nodo File Reader para cargar los datos de Excel de las ventas de LEYDE.

2. Partitioning (Particionado)

- Antes de entrenar el modelo, es importante dividir los datos en conjunto de entrenamiento y prueba para validar el rendimiento del modelo. El nodo Partitioning permite dividir los datos, comúnmente en un 70% para entrenamiento y 30% para prueba.

3. Random Forest Learner

- Este es el nodo principal para entrenar el modelo de Random Forest. En Random Forest Learner, donde podemos configurar varios parámetros como:
 - Número de árboles (Number of Trees): especifica cuántos árboles quieres en el bosque. Un número mayor puede mejorar la precisión, pero aumenta el tiempo de cálculo.
 - Número de atributos (Number of Attributes): el número de características a considerar para cada división de nodo.
 - Modo de impureza: KNIME utiliza criterios como el Gini index o entropía para dividir los nodos.

4. Random Forest Predictor

- Una vez entrenado el modelo, usamos el nodo Random Forest Predictor para predecir los valores del conjunto de prueba. Este nodo aplica el modelo entrenado a los datos de prueba y genera las predicciones.

5. Scorer

- Después de obtener las predicciones, usa el nodo Scorer para evaluar el rendimiento del modelo. Proporciona métricas como la precisión, sensibilidad, especificidad y matriz de confusión, que son útiles para analizar la calidad del modelo.

4.4 RESULTADOS DESCRIPTIVOS DE LOS MODELOS

A continuación, observaremos una serie de gráficos y figuras de todos los resultados descriptivos de la aplicación de los dos algoritmos en la Data de LEYDE:

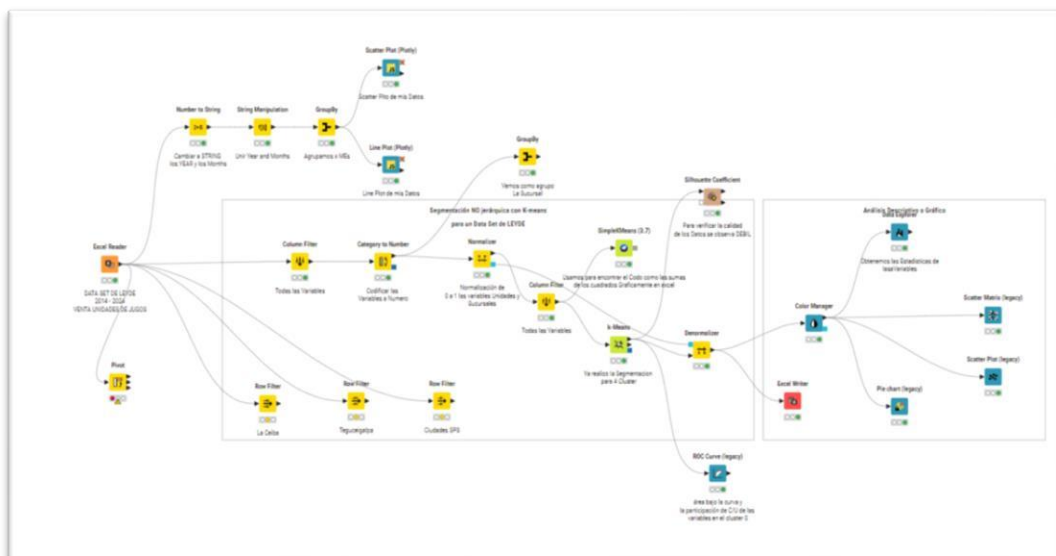


Figura 17. Diagrama de Knime para identificar patrones de ventas por medio de Kmedia o clusterización por sucursales en Honduras de la empresa LEYDE S.A.

Fuente: Elaboración Propia

Con el workflow o flujo de trabajo de la Figura 17 podemos observar toda la operación que realizamos con la base de datos de la empresa LEYDE para lo cual leemos los datos, los limpiamos, filtramos la información que ocupamos y entrenamos nuestro modelo de segmentación no supervisada y no jerarquizada con el nodo K-Means que nos ayuda a agrupar los objetos en K grupos basándonos en sus características. En este caso se logró determinar 4 agrupaciones para las 11 sucursales que son las encargadas de la fuerza de venta en todo el país basados en el histórico de unidades vendidas de jugos a través de 10 años (2014 al 2024). Al revisar el histórico de la data observamos en la Figura 18 que el 27.49% caen dentro del clúster “0”, el 16.74% dentro del clúster “1”, el 30.84 dentro del clúster “2” y el 24.93% dentro del clúster “3”. Estos datos ya normalizados son agrupados por la distancia euclidiana.

Aquí podemos determinar que dentro de las 11 sucursales hay de alto rendimiento en las ventas como las sucursales de San Pedro Sula, Tegucigalpa y Ceiba por la cantidad de rutas y el tamaño de la economía de las 3 ciudades principales del país.

Hay sucursales de bajo rendimiento porque son en lugares con ciudades en desarrollo como las de Santa Barbara y Colón; para lo cual debemos mejorar las estrategias de ventas y marketing en esa zona como hacer un análisis FODA de las instalaciones y la atención al cliente.

Hay sucursales con estacionalidad marcada ahí es donde nos vamos apoyar o beneficiar de una estrategia de inventarios o campañas adaptadas a su estacionalidad como Siguatepeque y Comayagua para no dejar desabastecidos con una mejor proyección de producción para las temporadas altas y evaluar como innovar en los aspectos del envase y el sabor específico de nuevas presentaciones.

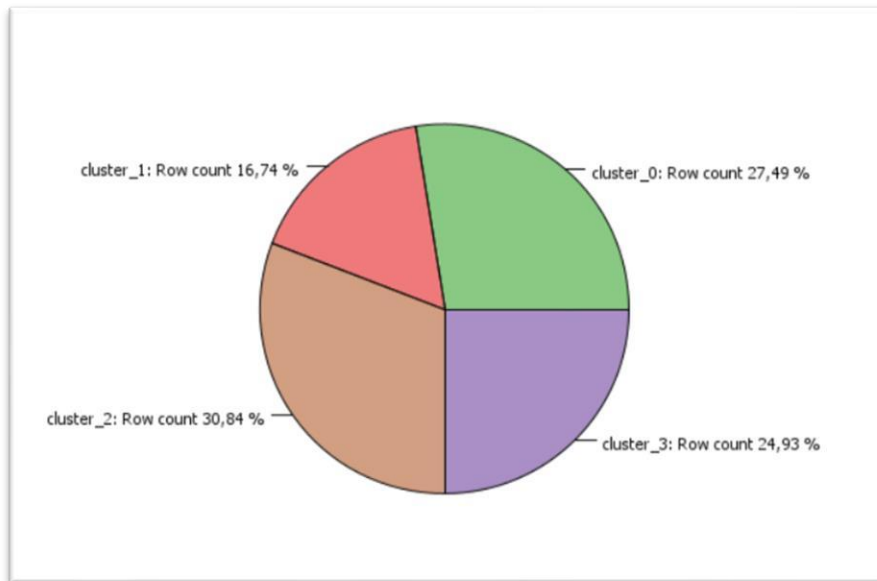


Figura 18. Diagrama de pastel en el software Knime usando la herramienta Kmedia para segmentar en 4 clúster las sucursales de acuerdo a su Patrón de tendencia en las ventas.

Fuente: Elaboración Propia

)

En la Figura 19 podemos observar que para poder determinar que nuestro algoritmo de K-Media optimiza con 4 clústeres revisamos de forma gráfica colocando desde $k=2$ a $k=10$ y obteniendo nuestro SSE o suma de cuadrados de los residuales o errores se forma un codo que denota que en ese específico quiebre esta nuestro k optimo. Para lo cual determinamos que es 4 usando el Nodo de Knime llamado SimpleKmeans (3.7).

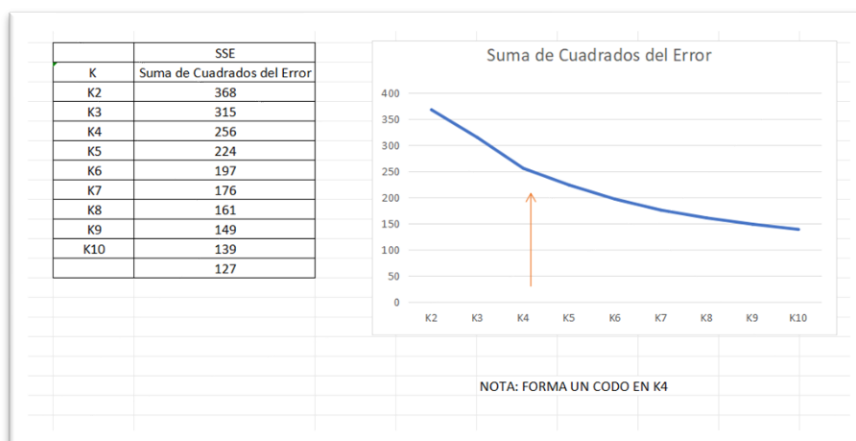


Figura 19. Hoja de Excel donde graficamos los datos usando el Nodo de SimpleKmeans 3.7 de Knime buscando la SSE (Suma de los Cuadrados) desde el Clúster 2 al 10 para saber dónde se forma el codo y determinar las cantidades de clúster ideal para usar la herramienta.

Fuente: Elaboración Propia

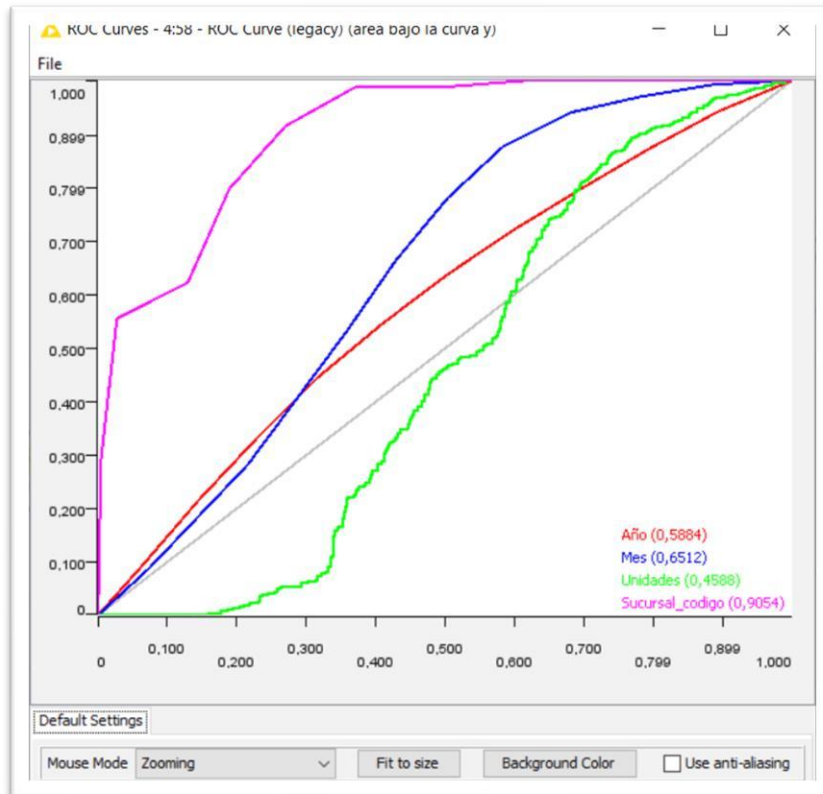


Figura 20. Gráfico de Área bajo la curva para revisar la data de ventas de LEYDE en unidades de jugo de Naranja.

Fuente: Elaboración Propia

Para apoyarnos con el grafico del área bajo la curva ROC (Receiver Operating Characteristic o Características Operativa del receptor) como herramienta estadística usamos el Nodo ROC CURVES y nos da la figura 20 y aquí medimos el acierto en la predicción de eventos que ocurren o que no ocurren. Para ello este indicador refleja la relación entre la tasa de verdaderos positivos como porcentaje en los 4 clústeres y la tasa de falsos positivos.

Tabla 6. Clasificación de las sucursales por clústeres con sus hallazgos

Clasificación de las Sucursales por Clúster					
Clasificación de acuerdo al comportamiento de las ventas en los últimos 10 años					
		HALLAZGOS			HALLAZGOS

CLUSTE R 0	SANTA BARBARA	REPRESENTA EL MAYOR NUMERO DE SUCURSALES Y MENOS VENTAS	CLUSTE R 1	SIGUATEPEQUE	REVISION DE INVENTARIOS POR VENTA PERDIDA
	ROATAN			SANTA ROSA DE COPAN	
	COMAYAGUA				
	CHULUTECA				
CLUSTE R 2	TOCOA COLON	OPORTUNIDAD DE MEJORAR O ASIGNACION DE RECURSOS	CLUSTE R 3	SPS	MAYOR VENTA A NIVEL PAIS
				LA CEIBA	
				TEGUCIGALPA	
				SAN PEDRO SULA	

Fuente: Elaboración Propia

En la Tabla anterior (numero 6) podemos observar los hallazgos a priori de acuerdo a la clasificación de k-media (algoritmo de knime) para lo cual ayudara a orientar a la fuerza de ventas y marketing en la toma de decisiones para mejorar las ventas.

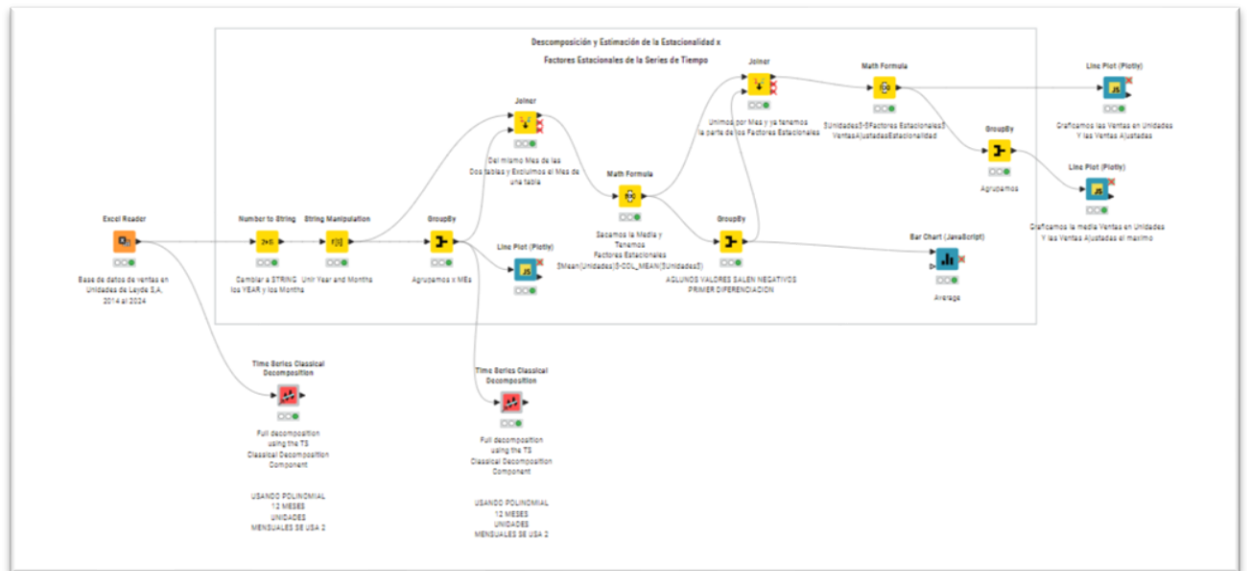


Figura 21. Diagrama de Series de tiempos de la herramienta Knime para evaluar la estacionalidad de las ventas en la empresa LEYDE S.A.

Fuente: Elaboración Propia

A continuación, realizamos como se muestra en la Figura 21 nuestro workflow o flujo de trabajo para la data de las ventas históricas de LEYDE al igual que el anterior para las Series de Tiempo.

Aquí nos interesa revisar lo siguiente usando series de tiempo en KNIME y sacar conclusiones útiles para la empresa:

1. Descomposición de la Serie de Tiempo

- **Tendencia:** Analizamos si hay una tendencia general de aumento, disminución o estabilidad en las ventas a lo largo del tiempo. Lo cual concluimos que hay una tendencia creciente por 6 años luego se observa que va de forma decreciente la venta en unidades de forma global a nivel país.

- **Estacionalidad:** Se observa patrones repetitivos dentro de cada año (12 meses) que indiquen variaciones estacionales. Esto es clave para entender cuándo las ventas tienden a ser más altas o bajas en cada año.

- **Ciclo:** Identificamos ciclos de ventas a largo plazo y podemos revisar periodos de crecimiento o contracción.

2. Análisis de Volatilidad

- Se Observa la estabilidad o volatilidad en las ventas para cada sucursal Donde predominan las ventas en las 3 ciudades más grandes del país como ser San Pedro Sula, Tegucigalpa y Ceiba.

3. Identificación de Anomalías

- Detectamos valores atípicos o eventos inesperados que podrían reflejar circunstancias inusuales, como promociones, crisis, o eventos económicos como la Pandemia **Covid** que inicio el 30 de enero de 2020 (4 años y 285 días) y finalizó el 5 de mayo de 2023.

4. Segmentación de Sucursales

- Consideramos agrupar sucursales con patrones de ventas similares en el proceso anterior y esto nos ayuda a encontrar grupos de sucursales que se comportan de manera parecida en cuanto a ventas, lo cual puede ser útil para decisiones de marketing y logística.

5. Modelos de Predicción

- Después de explorar la serie vamos a construir modelos predictivos usando un algoritmo de Bosques Aleatorios o Random Forest Learner y Predictor para generar predicciones para los datos futuros de ventas.

Al interpretar los resultados:

- Observamos los patrones de crecimiento para evaluar la salud general del negocio.
- Usamos las tendencias estacionales y la volatilidad para planificar inventario y estrategias de ventas.
- Usamos los grupos de sucursales para enfoques específicos de marketing y soporte.

El nodo que nos desarrolla todo el análisis es el Time Series Classical para ellos después de haber transformada y agrupada nuestra data de venta de 10 años en LEYDE podemos revisar los siguientes gráficos y obtener nuestras propias conclusiones de un pronóstico para el año 2025.

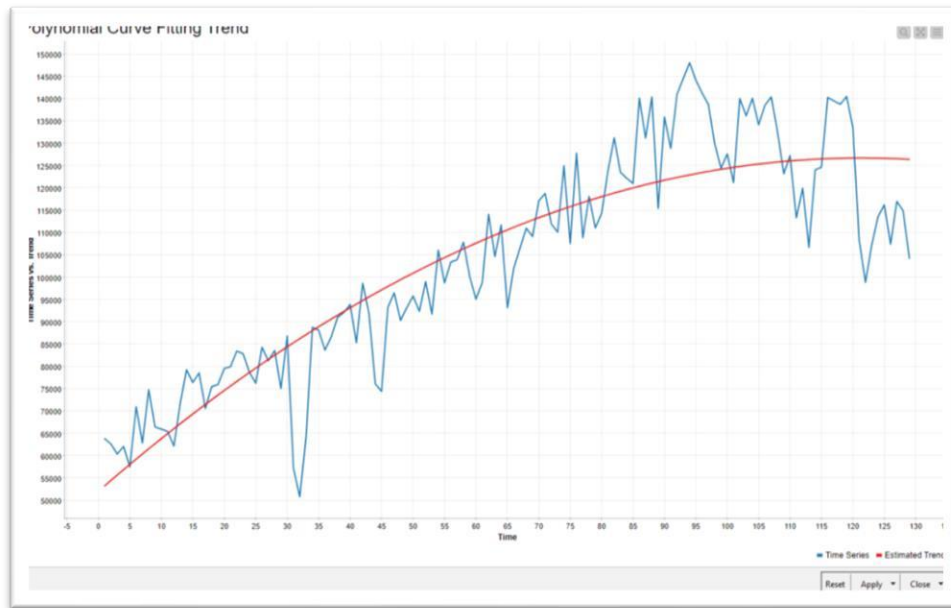


Figura 22. Gráfico de Polynomial Curve Fitting Trend

Fuente: Elaboración Propia

En la Figura 22 se observa nuestra serie de tiempo que tiene tendencia aditiva y se ve una estacionalidad marcada a través de los 10 años del histórico de ventas de la empresa LEYDE. La caída de las ventas está marcada desde el año 2021 dentro de las 11 sucursales a nivel país y para lo cual esa constante puede variar por muchos factores como los cambios en las preferencias de los clientes por comprar otras marcas o sabores de jugos que se ha incrementado esa oferta desde la Pandemia ya que muchas empresas determinaron que era un mercado seguro de hidratación y para la salud de las personas tomar bebidas naturales aunque dañinas por el tema de los azúcares como causa principal de la Diabetes.

Este gráfico de ajuste de curva polinomial nos ayuda a detectar tendencias subyacentes en los datos a lo largo del tiempo. Este método utiliza una función polinomial para ajustar una curva que representa una tendencia general en la serie, en lugar de los picos y valles específicos de cada punto de datos. Se observa una tendencia creciente que va a mantener un patrón de decreciente y notamos que debemos mejorar ese comportamiento y las métricas en cada dato individual.

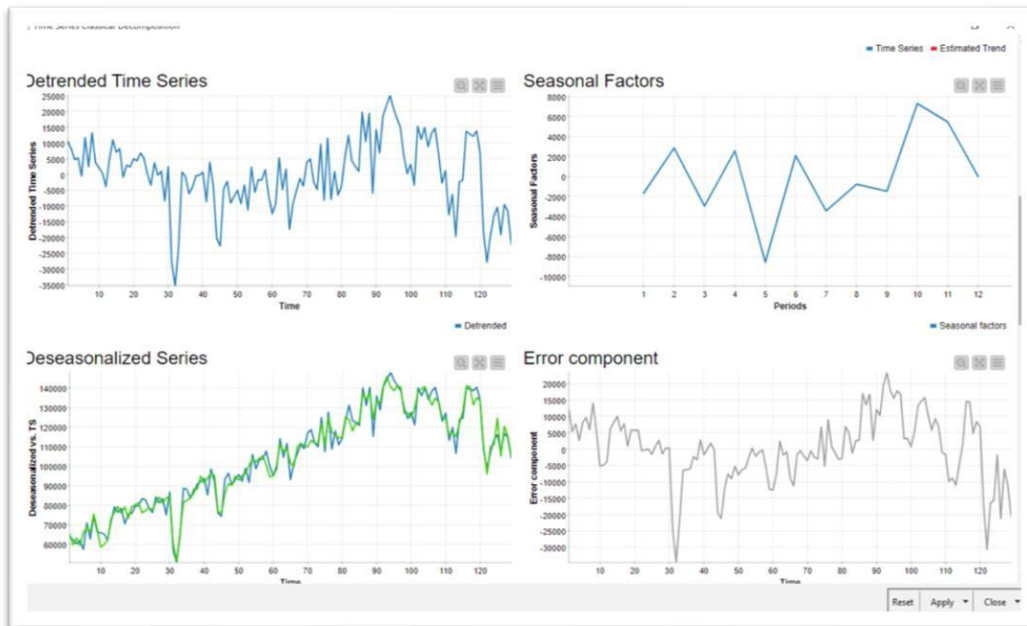


Figura 23. Gráficos que generamos usando el Nodo de Knime Time Series Classical Decomposition

Fuente: Elaboración Propia

Con los gráficos que se muestra en la figura 23 tenemos el primer componente de la serie de tiempo sin tendencia, factores estacionales, la serie desestacionalizada y el ruido gris. Para la Media Móvil hacemos el análisis obteniendo un rezago de la media de los datos, luego de la estacionalidad y de esa nueva media móvil hacemos dos rezagos hacia atrás.

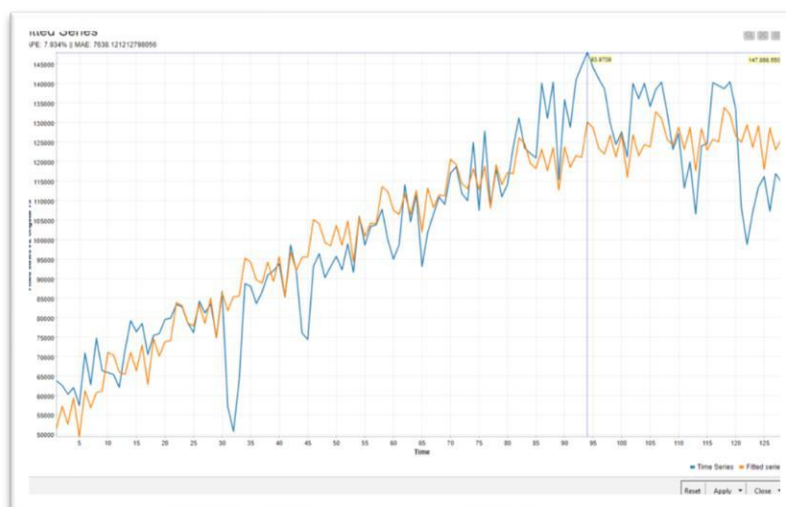


Figura 24. Gráfico de Nuestra Serie Ajustada con un rezago o Fitted

Fuente: Elaboración Propia

Este grafico de la Figura 24 representa la comparación entre los valores reales (observados) de nuestra serie temporal y los valores pronosticados (ajustados) aplicados por nuestro modelo en KNIME. Aquí evaluamos la precisión del modelo de forma visual y podemos observar las tendencias, patrones en los datos históricos.

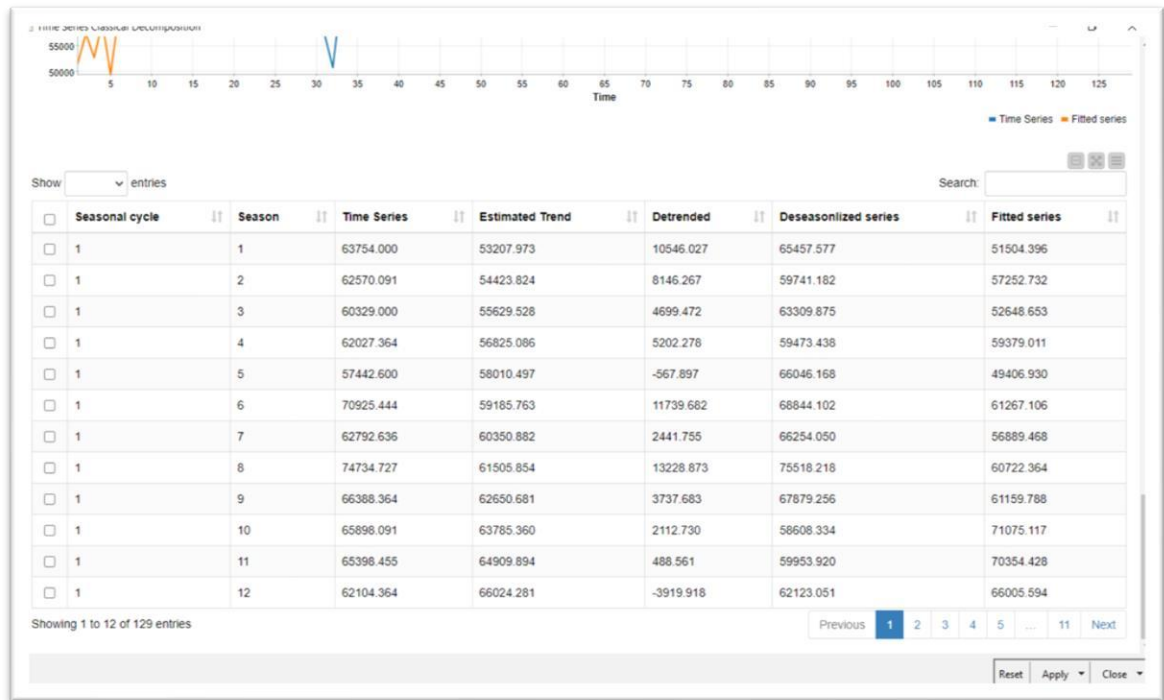


Figura 25. Datos que nos genera el algoritmo de Knime para Nuestra Serie de Tiempo

Fuente: Elaboración Propia

En nuestra Figura 25 podemos observar la data que nos da al realizar el algoritmo de series de tiempos donde podemos evaluar datos como la tendencia estimada, la series sin tendencia, la desestacionalidad de la serie y la serie ajustada para poder evaluar los datos como el comportamiento que vamos a tener en la venta de las distintas sucursales de la empresa LEYDE.

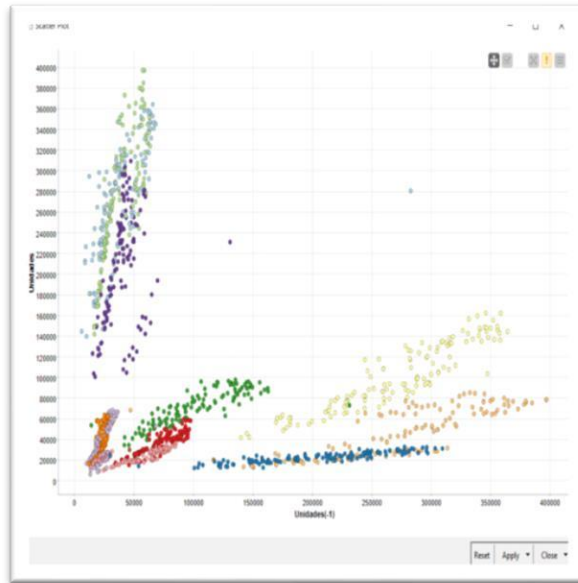


Figura 26. Figura del grafico de Scatter Plot donde podemos ver si hay o no autocorrelación de los datos de las ventas en unidades de la empresa LEYDE S.A. en Honduras de Jugos Procesados desde el 2014 al 2024.

Fuente: Elaboración Propia

Como se observa en la Figura 26 con este grafico de dispersión como se agrupan nuestros datos de ventas y podemos medir la relación entre los valores actuales y pasados de las series en diferentes retrasos (lags). Básicamente muestra cuanto depende los valores de la serie de sus valores en periodos anteriores. Aquí podemos ver una autocorrelación positiva que claramente nos da una tendencia o estacionalidad bien segmentada en nuestros datos.

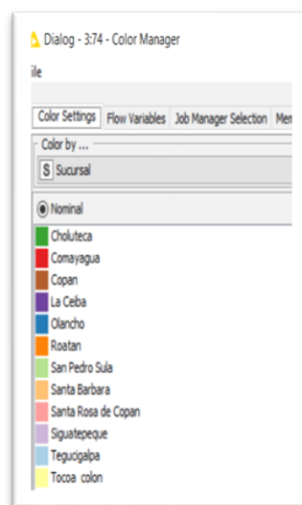


Figura 27. Observamos las diferentes Sucursales de LEYDE por Colores como referencia de nuestros Gráficos

Fuente: Elaboración Propia

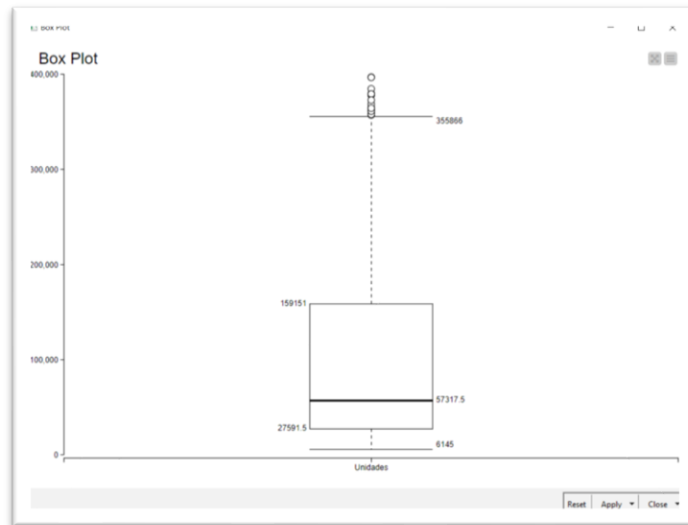


Figura 28. Figura del gráfico de Box Plot o diagrama de caja donde podemos evaluar rango (sin datos atípicos), datos atípicos, rango Inter cuartil (también conocido como RIC), cuartiles (Q1, Q2 y Q3), mediana (Q2) y mínimo y máximo de los datos de las ventas.

Fuente: Elaboración Propia

Para interpretar nuestro grafico de la figura 28 de caja y bigote revisamos los promedios de las unidades vendidas que van de 6145 en un mes determinado a 355866 unidades en otro pico de ventas y se puede observar que el primer cuartil tenemos 57317 unidades vendidas de jugos procesados en la empresa LEYDE.

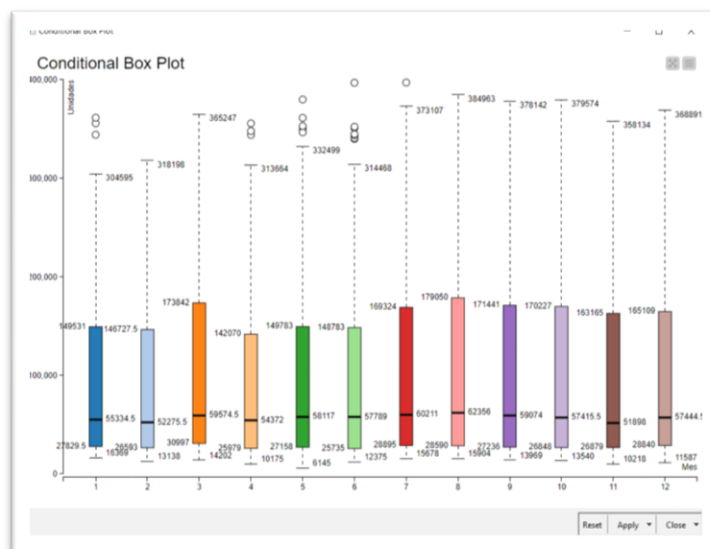


Figura 29. Figura del Gráfico de cajas condicional (JavaScript) Un gráfico de cajas muestra parámetros estadísticos robustos: mínimo, cuartil inferior, mediana, cuartil superior y máximo.

Fuente: Elaboración Propia

Aquí en la figura 29 es el mismo análisis, pero por medio de cada una de las diferentes sucursales que tiene la empresa LEYDE en todo el país.

Este gráfico de cajas muestra parámetros estadísticos robustos: mínimo, cuartil inferior, mediana, cuartil superior y máximo. Estos parámetros se consideran robustos porque no son sensibles a valores atípicos extremos.

El gráfico de cajas condicional divide los datos de una columna numérica en clases de acuerdo con otra columna nominal y crea un gráfico de cajas para cada una de las clases. Un gráfico de cajas para un atributo numérico se construye de la siguiente manera: La caja en sí va desde el cuartil inferior (Q1) hasta el cuartil superior (Q3). La mediana se representa como una barra horizontal dentro de la caja. La distancia entre Q1 y Q3 se denomina rango intercuartílico (IQR). Encima y debajo de la caja están los llamados "bigotes". Estos se dibujan en el valor mínimo y máximo como barras horizontales y se conectan a la caja mediante una línea punteada. Los bigotes nunca exceden $1.5 * IQR$. Esto significa que si hay algunos puntos de datos que superan $Q1 - (1.5 * IQR)$ o $Q3 + (1.5 * IQR)$, los bigotes se dibujan en el primer valor dentro de estos rangos y los puntos de datos se representan por separado como valores atípicos. Para los valores atípicos, se hace una distinción entre valores atípicos leves y extremos. Los valores atípicos leves son aquellos puntos de datos p para los cuales se cumple: $p < Q1 - (1.5 * IQR)$ Y $p > Q1 - (3 * IQR)$ o $p > Q3 + (1.5 * IQR)$ Y $p < Q3 + (3 * IQR)$. En otras palabras, los valores atípicos leves son aquellos puntos de datos que se encuentran entre $1.5 * IQR$ y $3 * IQR$. Los valores atípicos extremos son aquellos puntos de datos p para los cuales se cumple: $p < Q1 - (3 * IQR)$ o $p > Q3 + (3 * IQR)$. Por lo tanto, tres veces el ancho de la caja (IQR) marca el límite entre "valores atípicos leves" y "valores atípicos extremos". Los valores atípicos leves se representan como puntos, mientras que los valores atípicos extremos se muestran como cruces. Para identificar los valores atípicos, se pueden seleccionar y resaltar. Esto proporciona una visión rápida de las características extremas de un conjunto de datos de los datos de las ventas en unidades de la empresa LEYDE S.A. en Honduras de Jugos Procesados desde el 2014 al 2024.

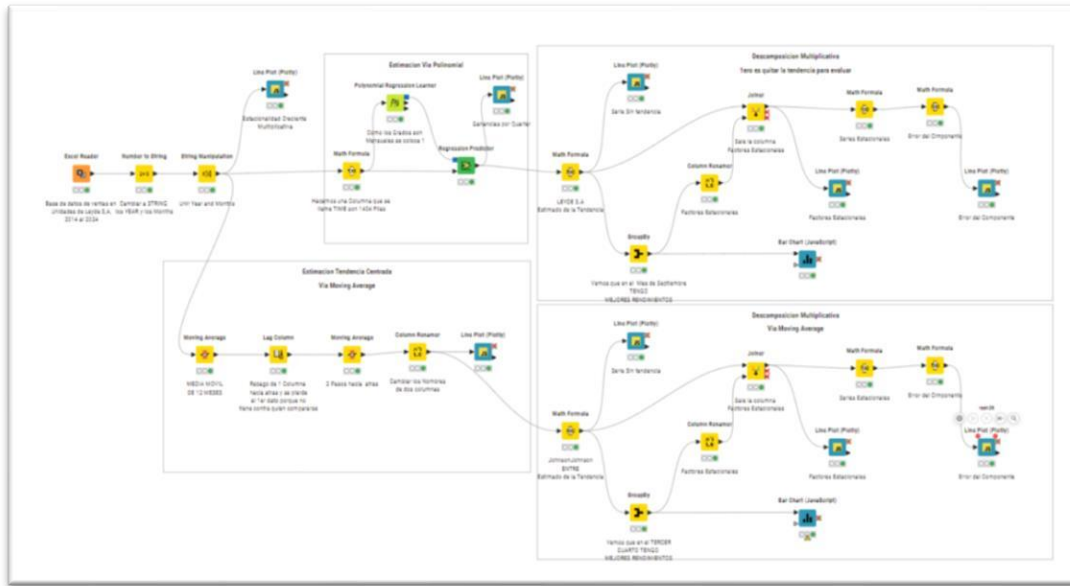


Figura 30. Gráfico de un workflow de Knime para nuestra data de ventas de series de tiempo de la empresa LEYDE S.A. para revisar de forma descriptiva los datos y así poder interpretarlos. Fuente: Elaboración Propia

Para nuestro flujo de trabajo se evaluaron varios caminos para revisar la data de LEYDE y poder determinar la forma óptima de buscar factores o tendencias de nuestra serie como se muestra en la figura 30.

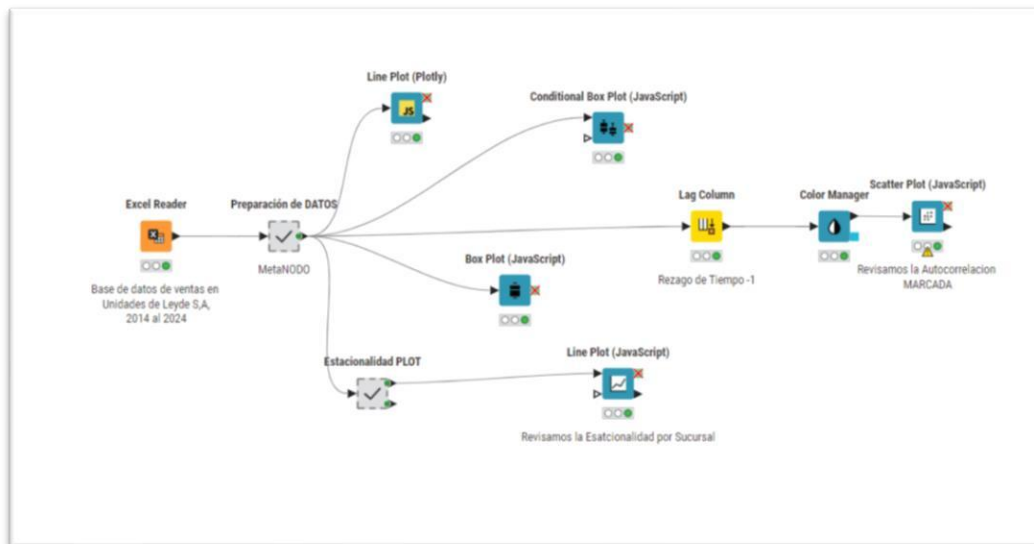


Figura 31. Gráfico de un workflow de Knime para nuestra data de ventas de series de tiempo de la empresa LEYDE S.A. para revisar de forma descriptiva los datos y así poder interpretarlos. Fuente: Elaboración Propia

Esta figura 31 nos muestra cómo hacemos el soporte de forma descriptiva de los datos para poder interpretarlos en las conclusiones y recomendaciones.

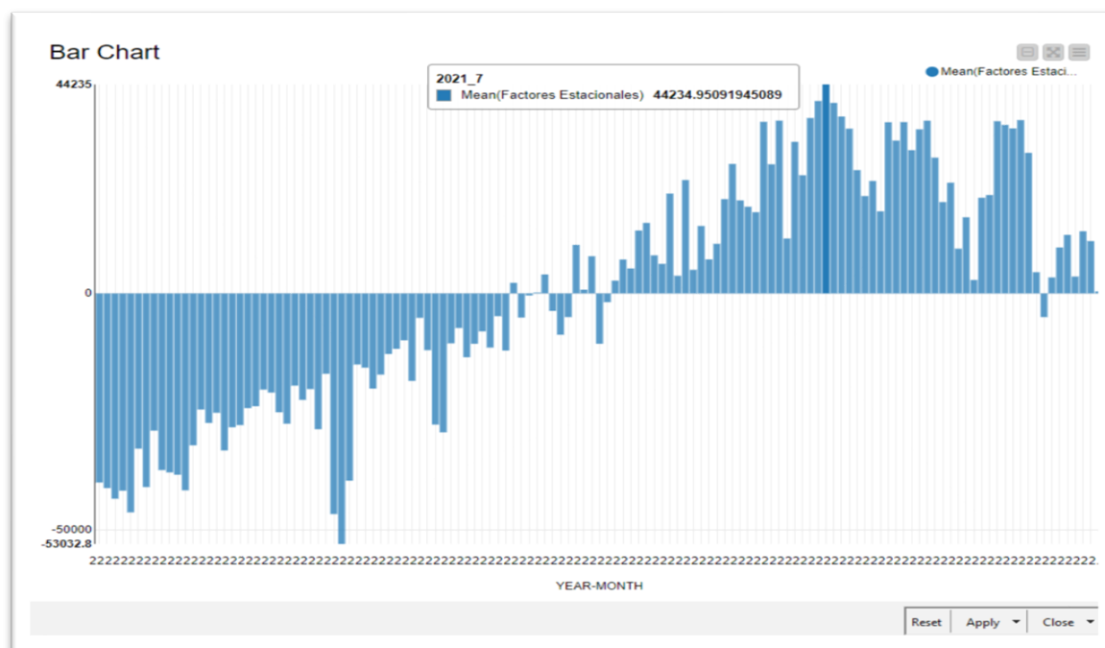


Figura 32. Gráfico de la Media de los Factores estacionales y muestra cómo varía la media en función de los periodos estacionales de forma mensual donde las ventas tienden a ser más altas o más bajas en promedio.

Fuente: Elaboración Propia

Todos estos resultados descriptivos nos servirán de soporte para dar nuestras conclusiones y recomendaciones para la empresa LEYDE.

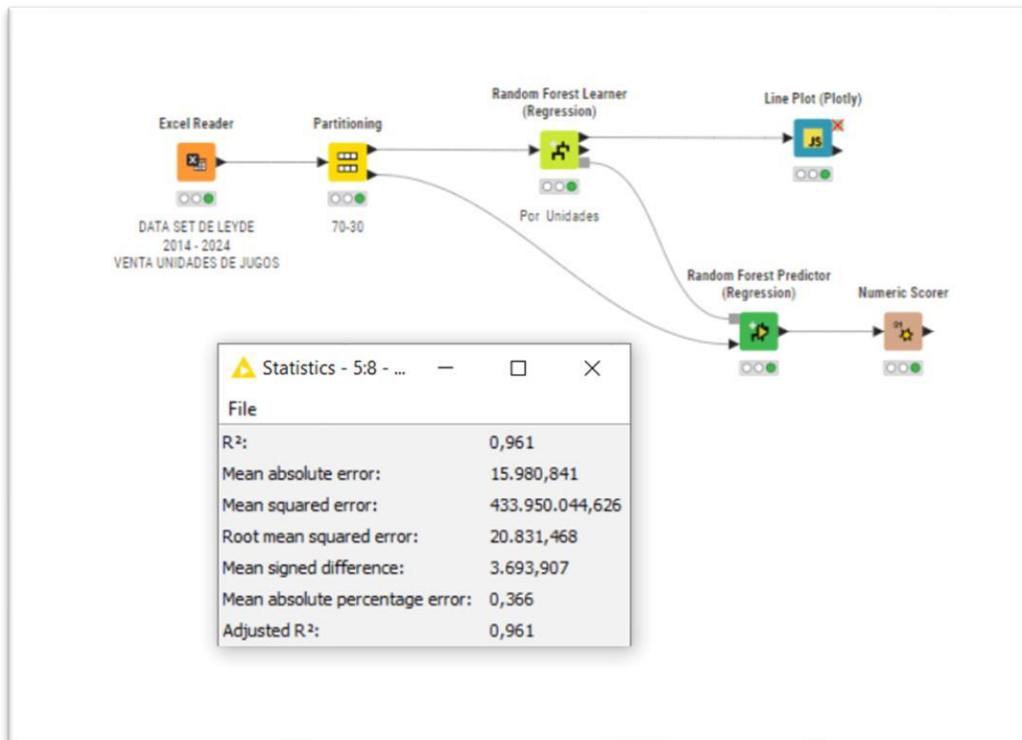


Figura 33. Algoritmo de Random Forest para predecir en nuestra Data la Venta Futura con un 96% de Accuracy

Fuente: Elaboración Propia

En la Figura 33 se observa cómo hacemos uso del algoritmo de **Random Forest** para predecir las ventas por sucursales como se muestra en la siguiente figura 31:

#	RowID	Sucursal	Mean(Unidades)
1	Row0	Cholulteca	67,994.842
2	Row1	Comayagua	35,458.941
3	Row2	La Ceiba	201,327.763
4	Row3	Olancho	23,184.385
5	Row4	Roatan	38,565.974
6	Row5	San Pedro Sula	295,643.487
7	Row6	Santa Barbara	53,318.179
8	Row7	Santa Rosa de Copan	24,454.154
9	Row8	Siguatepeque	40,807.641
10	Row9	Tegucigalpa	283,092.538
11	Row10	Tocoa colon	100,882.923

Figura 34. Resultados del Predictor de Random Forest. Fuente:

Elaboración Propia

CAPÍTULO V. CONCLUSIONES Y RECOMENDACIONES

A continuación, se presentan las conclusiones del proyecto de tesis y las recomendaciones dirigidas a la empresa LEYDE, enfocadas en la justificación de lo planteado en el capítulo 1, cuyo objetivo fue evaluar la situación actual de la organización. Este análisis está orientado a facilitar la toma de decisiones con el propósito de incrementar los márgenes de ganancia y reposicionar la marca de jugos con mejores indicadores de ventas, además de aumentar su participación en el mercado hondureño.

En el proyecto de investigación se implementó un modelo basado en algoritmos de pronósticos de series de tiempo y técnicas de clusterización para segmentar las 11 sucursales responsables de las ventas a nivel nacional. Estas sucursales, cuya estrategia de mercadeo y ventas está segmentada por zonas y clientes que se encargan de promover la fidelidad de los consumidores actuales.

La aplicación de los modelos de series de tiempo, los random forest y de clusterización proporcionó a la empresa valiosos "insights" basados en las preguntas planteadas al inicio del capítulo 1, así como en los objetivos generales y específicos establecidos.

5.1 CONCLUSIONES

Al observar la caída en las ventas de los últimos cuatro años de la familia de jugos en la empresa LEYDE S.A. y tras realizar un análisis de clusterización para agrupar las sucursales, además de emplear series de tiempo para el pronóstico de ventas para el año 2025, se identificaron diversas conclusiones.

1. Segmentación de sucursales y Demanda regional: El análisis de clusterización ha revelado patrones de comportamiento diferenciados entre las sucursales, lo que sugiere que las estrategias de marketing y ventas deberían adaptarse a las características específicas de cada grupo de sucursales, ya que no todas presentan los mismos niveles de ventas ni respuestas a las promociones. Basándonos en las 4 categorías antes mencionadas y entendiendo que las sucursales de San Pedro Sula, Tegucigalpa y Ceiba son las 3 más importantes con una fuerza de ventas de más del 80% del grupo de 11. Existen diferencias en la demanda regional de los productos de jugos LEYDE. Las sucursales ubicadas en áreas urbanas muestran una mayor estabilidad en las ventas, mientras que las rurales o menos pobladas experimentan mayores fluctuaciones a lo que la fuerza de marketing debe orientar la demanda de acuerdo a sabores por zonas.

2. Estacionalidad en las ventas y caída consistente en las ventas: Las series de tiempo utilizadas muestran patrones estacionales claros, lo que indica que hay periodos del año con mayor o menor demanda, lo cual debe ser considerado en la planificación de inventarios y campañas promocionales. La tendencia a la baja observada en los últimos cuatro años apunta a un posible cambio en las preferencias de los consumidores, la aparición de competidores o una percepción de menor valor del producto.

3. Precisión del modelo de pronóstico y concentración de las ventas: Las series de tiempo sugieren una ligera recuperación para 2025, pero no es suficiente para alcanzar los niveles de ventas de años anteriores, lo que alerta sobre la necesidad de intervenciones estratégicas más agresivas. El análisis ha identificado que un pequeño grupo de sucursales es responsable de la mayor parte de las ventas, mientras que otras sucursales contribuyen de manera muy marginal. Este fenómeno de concentración indica que algunas zonas podrían estar siendo desatendidas o mal gestionadas.

4. Impacto de promociones e impacto económico externo: Las promociones de ventas tienen efectos positivos, pero temporales. La ausencia de una estrategia de fidelización hace que las ventas caigan nuevamente después de las promociones. Los factores externos como la inflación o crisis económicas pueden haber afectado negativamente las ventas de productos no esenciales, como los jugos, y esto debe considerarse en el contexto del análisis de las caídas en las ventas.

5. Desviación del pronóstico y deterioro de la imagen de la marca: Si bien el pronóstico muestra una tendencia a la baja en la mayoría de las sucursales, algunas muestran signos de estabilización o incluso crecimiento, lo que indica que ciertos mercados pueden tener un potencial aún no explotado. La tendencia negativa en las ventas puede estar relacionada con un deterioro de la imagen de la marca, lo que afectaría la lealtad del cliente y justificaría la necesidad de un rebranding o mejoras en la percepción del producto.

Estas conclusiones son una parte integral de nuestro análisis, ayudando a explicar las razones detrás de la caída en ventas y proporcionando un plan de acción estratégico para revertir la tendencia.

5.2 RECOMENDACIONES

A continuación, damos una serie de Recomendaciones basado en nuestro análisis de clusterización y de series de tiempos para la empresa LEYDE:

1. Estrategias personalizadas para cada clúster de sucursales y campañas focalizadas por temporadas: Implementar estrategias de marketing diferenciadas para cada grupo de sucursales basado en el análisis de clusterización, adaptando promociones, precios y productos a las características específicas de cada mercado. Aprovechar las tendencias estacionales identificadas a través de las series de tiempo, lanzando campañas y promociones durante los periodos de mayor demanda para maximizar las ventas.

2. Innovación en productos y estrategias de fidelización: Diversificar el portafolio de productos introduciendo nuevas variedades de jugos o presentaciones que se ajusten mejor a las necesidades de los consumidores actuales. Implementar programas de fidelización o beneficios recurrentes que logren mantener al cliente comprometido más allá de los efectos de las promociones temporales.

3. Reestructuración de sucursales de bajo rendimiento y un Rebranding o reposicionamiento: Evaluar la posibilidad de reestructurar o cerrar sucursales con ventas persistentemente bajas y redistribuir los recursos en sucursales con mayor potencial o en áreas nuevas con mejor proyección de crecimiento. Considerar un relanzamiento de la marca para mejorar la percepción entre los consumidores y contrarrestar la posible caída en la imagen del producto.

4. Optimización de la logística e inventarios y una mejora del canal digital: Ajustar la cadena de suministro e inventarios para adaptarse mejor a las fluctuaciones estacionales y regionales detectadas, minimizando costos y desperdicios. Para la impulsión de sabores como la Limonada Rosa, Tamarindo y Jugo de Naranja Agria se debe comprar fruta en meses donde los precios son menores como el ejemplo del limón que en verano alcanza hasta L. 6.00 por unidad y en invierno baja a L. 0.50 centavos para lo cual se recomienda tercerizar el exprimido de la fruta y mantener en congeladores bajo cero el zumo aplicando estabilizadores y preservantes como el sorbato de potasio para usar esa materia prima en temporadas altas y mantener el punto de equilibrio de nuestros precios para la demanda.

Aprovechar la tendencia hacia el comercio electrónico y mejorar la presencia online de los productos de jugos LEYDE, ampliando las opciones de compra y mejorando la visibilidad de la marca.

5. Promociones regionalizadas y monitoreo constante del entorno económico: Lanzar promociones adaptadas a las necesidades y capacidades de compra de cada región o clúster de sucursales para maximizar el impacto de las campañas. Se deberá mantener una vigilancia constante de los factores económicos externos que puedan afectar el consumo de los jugos LEYDE y ajustar las estrategias de precios y promociones en consecuencia.

CAPÍTULO VI. APLICABILIDAD

6.1 NOMBRE DE LA PROPUESTA

Creación de la plaza de trabajo “Coordinador de Análisis de Datos de la Familia de Jugos-LEYDE” que será la persona encargada del análisis de los datos de las ventas reales versus a las metas establecidas para lograr incrementar los indicadores de ventas de las 11 sucursales del país.

6.2 JUSTIFICACIÓN DE LA PROPUESTA

El estudio de los datos desde el punto de vista de las ventas, presentado en esta investigación, revela la importancia de que las empresas cuenten con equipos especializados que se encarguen de extraer, depurar y transformar la información de los distintos procesos organizacionales, convirtiéndola en recursos. clave para la toma de decisiones.

6.3 ALCANCE DE LA PROPUESTA

Contar con la persona y las herramientas necesarias para brindar toda la información a la Gerencia General, la Gerencia de Ventas, Jefes de Ventas de cada sucursal para la toma de decisiones donde se implementarán reportes semanales en Power BI en la nube con la finalidad de revisar y mejorar los indicadores de ventas del próximo año 2025 para la familia de jugos de la empresa LEYDE.

6.3.1 Objetivo General

Implementar algoritmos de machine learning para establecer metas de ventas en las distintas sucursales basados en sus históricos tratando de lograr un incremento de las mismas mes a mes y así poder optimizar los indicadores de ventas de la familia de jugos de la empresa LEYDE de cara al año 2025.

6.3.2 Objetivos Específicos

1. Realizar un análisis semanalmente de los comportamientos de las sucursales según patrones de comportamiento en las ventas.

2. Revisar y comparar las metas versus el cumplimiento de las ventas del año 2025 con alta precisión en cada sucursal del país.
3. Utilizar algoritmos de machine learning para identificar las variables más influyentes en el desempeño de las ventas de la familia de jugos.
4. Generar recomendaciones estratégicas basadas en los resultados de los algoritmos implementados para mejorar la toma de decisiones comerciales.
5. Evaluar el impacto de las estrategias sugeridas en los márgenes de ganancia y participación de mercado de la empresa.

6.4 DESCRIPCIÓN Y DESARROLLO A DETALLE DE LA PROPUESTA

6.4.1 DESCRIPCIÓN CLARA DEL QUE Y COMO SE HARÁN

- Es indispensable contar con un profesional que tenga experiencia en análisis de datos y en la elaboración de modelos descriptivos y predictivos.
- Se requiere un profesional con experiencia en la interpretación y análisis de datos de áreas tanto administrativas como operativas.
- Es necesario solicitar la creación de una plaza de Coordinador de Análisis de Datos en la ciudad de la Ceiba, que cumpla con los requisitos académicos y profesionales.
- Una vez que se autorice la creación de la plaza, se debe proceder a la publicación de la oferta para recibir las hojas de vida de los candidatos.
- Solo serán considerados aquellos candidatos que cumplan con los requisitos del perfil establecido.
- Las entrevistas serán realizadas por el Gerente General y el Gerente de Planning.
- La persona seleccionada dependerá directamente de la Gerencia de Planning.
- La persona seleccionada tendrá la responsabilidad de realizar análisis descriptivos, y los paneles de control serán presentados semanalmente, quincenal y mensualmente a las distintas gerencias para facilitar la toma de decisiones.

6.4.2 DESARROLLO DE TODOS LOS ELEMENTOS NECESARIOS (HERRAMIENTAS, INSTRUMENTOS, PROCESOS)

- Apertura de plaza para Coordinador de Análisis de Datos
- Compra de computadora Laptop
- Compra de escritorio
- Compra de silla
- Compra de licencia Office

6.5 MEDIDAS DE CONTROL (INDICADORES, MEDICIONES)

- Pronósticos de las Metas de Mes a Mes
- Revisión de Cumplimiento de Metas
- Forecast mensuales y anuales.
- KPI áreas de ventas

6.6 CRONOGRAMA DE IMPLEMENTACIÓN Y PRESUPUESTO

Nos basamos en técnicas de implementación análoga para realizar basados en proyectos similares de apertura de puestos para realizar el siguiente cronograma de tiempos por semanas.

Tabla 7. Cronograma de Implementación y Presupuesto

DETALLE	SEMANAS									
	1	2	3	4	5	6	7	8	9	10
Solicitud de Plaza para el “Coordinador de Análisis de Datos de la Familia de Jugos-LEYDE”	■	■								
Recepción de Hoja de Vida de Candidatos			■							
Entrevistas y aplicación de pruebas de aptitud por parte de RRHH				■						
Entrevistas de las personas seleccionadas por la Gerencia General y Gerencia de Planning					■					
Compra de Computadora Laptop						■				
Compra de Escritorio							■			
Compra de Silla								■		
Compra de Licencia Office									■	
Contratación del Candidato										■
Inducción por parte de la Gerencia de Planning										■

Fuente: Elaboración Propia

Tabla 8. Presupuesto Anual del Coordinador

PRESUPUESTO ANUAL DEL COORDINADOR	
Sueldos	\$ 12.000,00
Plan Móvil	\$ 400,00
Depreciación vehículo	\$ 500,00
Pasivo Laboral	\$ 3.600,00
13 Avo- 14 Avo	\$ 1.000,00
Laptop	\$ 1.500,00
Silla y Escritorio	\$ 500,00
Licencias de Office	\$ 200,00
Total	\$ 19.700,00

Fuente: Elaboración Propia

6.7 CRONOGRAMA DE IMPLEMENTACIÓN Y PRESUPUESTO

Tabla 9. Matriz de Concordancia

TITULO DE LA INVESTIGACION						
IMPLEMENTACIÓN DE MODELOS PREDICTIVOS PARA MEJORAR LAS VENTAS DE LA FAMILIA DE JUGOS EN LEYDE						
Problema General	Objetivo General	Hipótesis	Teorías/Metodológicas de Sustento	Variables Independientes		
¿Puede un modelo de series de tiempos y clusterización ayudar al departamento de ventas de LEYDE a pronosticar, identificar y segmentar eficientemente los perfiles de sus clientes para mejorar los indicadores de la familia de jugos envasados específicamente dentro de su cartera de productos?	Aplicar técnicas de Machine Learning para analizar y predecir las ventas de la línea de jugos de LEYDE S.A., con el fin de optimizar los indicadores de ventas y desarrollar estrategias de marketing y fidelización más efectivas	<p>H1: Hipótesis General (HG)</p> <p>La implementación de técnicas de Machine Learning para el análisis de series de tiempo y la segmentación de clientes permitirá mejorar significativamente las ventas de la familia de jugos de LEYDE, optimizando las estrategias de marketing y ventas.</p> <p>H0: Un modelo de series de tiempo y de segmentación de clientes no es capaz de ayudar a mejorar las ventas de LEYDE</p>	<p>1. Teoría de Segmentación de Mercado</p> <p>2. Teoría del Ciclo de vida del Producto</p> <p>3. Teoría del Comportamiento del Consumidor</p> <p>4. Teoría de los Procesos Estocásticos</p> <p>5. Teoría de Redes y Propagación de Información</p> <p>6. Teoría de Sistemas Dinámicos</p> <p>7. Teoría de la Competencia y Estrategia de Porter</p>	Variables Independientes		
				"Modelo Clusterizacion"		
				Dimensiones	Indicadores	Unidad Medida
Problema Especifico	Objetivos Especificos			1. Indicadores de Clientes	Clústeres para dividir los clientes	Cantidad de venta de Productos por Sucursal por Mes
I. ¿Cuáles son los principales factores que han influido en	I. Analizar los datos históricos de ventas de la			2. Indicadores de Algoritmos		
			Metodologías			

<p>el decrecimiento de las ventas de la línea de jugos de LEYDE S.A. en los últimos 4 años? II. ¿Cómo pueden las técnicas de series de tiempo ayudar a prever las ventas futuras y anticipar cambios en la demanda de los productos de jugos? III. ¿Qué patrones de comportamiento de los clientes pueden ser identificados mediante técnicas de clustering y cómo se pueden utilizar para segmentar el mercado? IV. ¿Qué estrategias de marketing y ventas pueden ser implementadas con base en los resultados obtenidos a partir de los modelos de Machine Learning para aumentar la fidelización y mejorar los</p>	<p>línea de jugos de los últimos 10 años para identificar patrones, tendencias y factores asociados al decrecimiento de ventas. II. Utilizar técnicas de series de tiempo para pronosticar las ventas futuras de la línea de jugos y anticipar posibles fluctuaciones o tendencias. III. Implementar algoritmos de clasificación, como el análisis de clusterización mediante KMeans, para segmentar a los clientes según su comportamiento de compra. IV. Desarrollar recomendaciones para optimizar las estrategias de marketing y ventas, basadas en los resultados del análisis de datos y las</p>		<p>1. Algoritmo K-Means 2. Algoritmo de Series de Tiempos</p>			
<p>indicadores de ventas?</p>	<p>predicciones generadas por los modelos de Machine Learning.</p>					
<p>Tipo y Diseño</p>	<p>Población y Muestra</p>	<p>Técnicas y Procedimientos</p>	<p>Estadística Para Utilizar</p>			
<p>Estudio: Longitudinal</p>	<p>Población: Para efectos de esta investigación la muestra está compuesta por 1404 registros de ventas de las 11 sucursales en los 117 meses. En el proceso de transformación y limpieza de los datos se suprimieron los "Missing Values" o valores nulos. Representa una muestra no probabilística.</p>	<p>Técnica: Observación</p>	<p>Descriptivo</p>			
<p>Diseño: No experimental</p>		<p>Fase 1: Recolección de datos Obtención y limpieza de datos de ventas Fase 2: Análisis Exploratorio Análisis inicial y visualización de datos Fase 3: Modelos predictivos Implementación de modelos de series de tiempo Fase 4: Segmentación Clusterización de clientes y análisis de patrones Fase 5: Validación Evaluación y ajuste de</p>	<p>- Medidas de tendencia central: Como la media, la mediana y la moda, para obtener una idea del comportamiento promedio de las ventas en LEYDE - - Medidas de dispersión: Como la desviación estándar, varianza, percentiles, y rango Inter cuartil, que ayudarán a entender la variabilidad de las ventas. - Visualizaciones: Histogramas, gráficos de barras, diagramas de dispersión (scatterplots) y diagramas de caja (boxplots) para identificar patrones, posibles outliers y distribuciones.</p>			
<p>Alcance: Descriptivo</p>		<p>Muestra: Para efectos de esta</p>	<p>Inferencial</p>			

Enfoque: Inferencial	investigación la muestra está compuesta por 1404 registros de ventas de las 11 sucursales en los 117 meses. En el proceso de transformación y limpieza de los datos se suprimieron los “Missing Values” o valores nulos. Representa una muestra no probabilística.	los modelos Fase 6: Propuesta estratégica Desarrollo de estrategias basadas en los resultados	<ul style="list-style-type: none"> - Intervalos de confianza: Para estimar el rango dentro del cual podrían estar las ventas futuras con un cierto nivel de confianza. - Pruebas de hipótesis: Vamos a verificar si hay diferencias significativas en las ventas entre diferentes períodos de tiempo (antes y después de la estacionalidad). - Modelos de regresión: Vamos a evaluar una regresión múltiple como parte de los algoritmos de predicción o como comparativo a los modelos más avanzados.
-------------------------	--	---	---

Fuente: Elaboración Propia

REFERENCIAS

- Analitics. (2019). *Segmentación de clientes: Estrategias de mercado para una retención eficaz*. Analíticas.
- Box, G., Jenkins, J., & Reinsel, G. (2015). *Análisis de series temporales: pronóstico y control (5.ª ed.)*. Wiley.
- Breiman, L. (2001). *Bosques aleatorios. Aprendizaje automático*.
- Brynjolfsson. (2014). *Big Data*.
- Chatfield, C. (2003). *Importancia de las series de tiempo en el análisis empresarial*.
- Chaudhary, S. (2021). Market segmentation using clustering algorithms in the beverage industry. *International Journal of Marketing Research*.
- Data Solutions. (s.f.). Obtenido de <https://www.lisdatasolutions.com/es/que-es-knime>

- Doe, L., & Perez, M. (2020). Optimizing Retail with Clustering and Time Series Analysis: Walmart's Approach to Predictive Analytics. *Journal of Retail Management and Innovation*.
- Drury, C. (2018). *Contabilidad de gestión y costes*. Cengage Learning.
- Emprendedores. (2023). Obtenido de <https://emprendedores.es/ideas-de-negocio/ejemplosbig-data-en-empresas/>
- Fernández, L. (2021). *Aprendizaje Automático y Análisis de Datos*. Ciencias Computacionales.
- García. (2020).
- García. (2022). Clustering algorithms for supply chain optimization: A case study in the food industry. *Supply Chain Management Journal*.
- García, J. (2019). *El Impacto de la Tecnología en el Análisis de Datos Moderno*. Tecnológica.
- García, J. (2019). *El Impacto de la Tecnología en el Análisis de Datos Moderno*. Editorial. Tecnológica.
- Gomez, A., & Lee, H. (2021). Market trend analysis and customer clustering for product customization: A case study in Nike. *International Journal of Retail & Distribution Management*.
- Hawking, J. (2004). *Sobre la inteligencia*. Basic Book.
- Hu, W. (2020). *Inventory optimization with data clustering dor multi warehouse e-commerce*. Journal of Big Data.
- Hugues, A. (1994). Marketing estratégico de bases de datos: el plan maestro para iniciar y gestionar un programa de marketing rentable basado en el cliente. Probus Publishing.
- Johnson, L., & Chen, W. (2022). Clustering user preferences for product development in technology companies: The case of Apple. *Journal of Product*.
- Knime Web Site. (s.f.). Obtenido de <https://forum.knime.com/t/diccionario-glosario-de-nodos/>
- Kotler, P., & Keller, K. (2016). *Gestión de marketing (15.ª ed.)*. Pearson.
- Kotler, P., & Keller, K. (2016). *Gestión de marketing 15 Ed.* Pearson.
- LA FAO org. (s.f.). Obtenido de <https://www.fao.org/4/v8490s/v8490s06.htm>
- Lindberg, & Hedman. (2020). Data clustering in supply chain optimization: Case study of IKEA's inventory management. *Logistics and Transportation*.
- López, J. (2021). *Fundamentos de Aprendizaje Automático*. Tecnología.
- López, M. (2022). *Herramientas de Data Science: Una Guía Completa*. Tecnología.
- Mahfoud. (2019). Cluster analysis in manufacturing quality control: A case study in the automotive industry. *International Journal of Production Research*.
- Malaga, J. (2023). Definiendo el Cliente Óptimo en Marketing. *Revista de Marketing Estratégico*, vol. 5, no. 2.
- Manouba, U. d. (2024). Agrupamiento de datos de redes sociales para estrategias de marketing.
- Martinez, C. (2018). *Fundamentos del CRM y su Importancia en las Empresas Modernas*. Gestión Empresarial.
- Martinez, C. (2021). *Fundamentos de ETL y Almacenamiento de Datos: Tecnología de la Información*.
- Moreno, C. (2022). *Perfilamiento de Clientes en el Marketing Moderno*. Investigación de Mercado.
- Nicolau, W. (2019). Cluster analysis for price segmentation in the sharing economy: A case study of Airbnb. *Tourism Management*.
- Patel, K., & Fernandes, T. (2020). Supply chain clustering and optimization in the FMCG sector: A case study of Unilever. *Journal of Operations Management*.
- Pérez, A. (2021). *Algoritmos de Aprendizaje Automático: Una Guía Práctica*. Ciencia de Datos.
- Perez, J. (2019). *Análisis de Datos y Machine Learning*. Ciencias de la Computación.

- Rakesh, K. (2018). Supplier clustering using data mining techniques: A case study in the automotive industry. *Journal of Manufacturing Systems*, 29.
- Romero, A. (2022). *Evolución y Tendencias en Inteligencia de Negocios*. Análisis de Datos.
- Silva, M. (2019). Data clustering in fast fashion supply chains: A case of Zara. *Operations Research and Decisions*.
- Smith, D., & Martinez, J. (2021). Cluster-based analysis of electric vehicle battery performance: Case study at Tesla. *IEEE Transactions on Industrial Applications*.
- Smith, K. (2019). Data Analytics in Retail: How Amazon Uses Data to Drive Sales. *Journal of Business Analytics*.
- Solomon, M. (2018). *Comportamiento del consumidor: comprar, tener y ser (12.ª ed.)*. Pearson.
- Tang, X., & Wang, S. (2020). Data clustering techniques in pharmaceutical research: Case studies at Pfizer. *Pharmaceutical Research Journal*.
- Torres, F. (2020). *Data Mining: Fundamentos y Aplicaciones Prácticas*. Tecnologías de la Información.
- Tukey, J. (1977). Análisis exploratorio de datos.
- Wan, A. (2021). Clustering analysis for optimizing logistics networks in global retail. *Procedia Computer Science*.
- Wedel, M., & Kamakura, W. (2000). *Segmentación de mercado: fundamentos conceptuales y metodológicos*. Springer.
- Yan, X. (2020). Machine learning for predictive maintenance: A clustering-based approach. *IEEE Transactions on Industrial Informatics*.

GLOSARIO

A continuación, damos un glosario organizado alfabéticamente, que abarca términos clave relacionados con la aplicación de algoritmos de clusterización y series de tiempo para el análisis de ventas históricas en la empresa LEYDE:

Algoritmo: Conjunto de instrucciones o reglas definidas que se siguen para realizar cálculos, resolver problemas o analizar datos de manera sistemática.

Bosques Aleatorios: Es una técnica de aprendizaje automático de ensamblaje que crea múltiples árboles de decisión y combina sus predicciones para mejorar la precisión y evitar problemas como el sobreajuste

Clusterización (o Agrupamiento): Técnica de aprendizaje no supervisado utilizada para clasificar o agrupar datos en subconjuntos o "clusters" donde los objetos dentro de un grupo son más similares entre sí que aquellos en otros grupos.

Clúster: Subgrupo o conjunto de datos similares entre sí dentro de un conjunto más grande, obtenido mediante técnicas de clusterización.

Cohesión de Clúster: Medida de la similitud entre los elementos dentro de un mismo clúster. Un valor alto indica que los datos están bien agrupados.

Correlación: Relación entre dos variables en la que un cambio en una variable se asocia con un cambio en la otra. Puede ayudar a identificar patrones o tendencias.

Datos Históricos de Ventas: Conjunto de datos que refleja las ventas pasadas, usado para analizar patrones y realizar predicciones futuras.

Desviación Estándar: Medida de la variabilidad o dispersión de un conjunto de datos. Es útil para entender la consistencia de los datos en un clúster.

Estacionalidad: Comportamiento o patrón que se repite en los datos de ventas en intervalos específicos (mensuales, trimestrales, anuales). Es importante en análisis de series de tiempo.

Métricas de Similitud: Herramientas utilizadas para determinar la similitud o diferencia entre dos elementos, comúnmente aplicadas en técnicas de clusterización.

Normalización: Proceso de escalado de datos para que diferentes variables tengan un peso equivalente en el análisis. Se utiliza en algoritmos de clusterización para mejorar la precisión de los resultados.

Patrón de Comportamiento de Cliente: Análisis del comportamiento de los clientes en función de sus compras históricas, frecuencias y preferencias. Ayuda a identificar clústeres específicos de clientes.

Predicción: Estimación de valores futuros basada en modelos matemáticos y datos históricos. En ventas, se utiliza para prever la demanda futura y optimizar la planificación.

Series de Tiempo: Secuencia de datos recogidos en intervalos regulares a lo largo del tiempo. En ventas, se usa para identificar patrones de demanda y prever futuras tendencias.

Segmentación de Clientes: Técnica de agrupar a los clientes en diferentes segmentos según características comunes (edad, ubicación, frecuencia de compra, etc.). La clusterización es útil en este proceso.

Tendencia: Dirección general de los datos en una serie de tiempo, que puede ser creciente, decreciente o estable. Ayuda a entender el comportamiento a largo plazo de las ventas.

Visualización de Datos: Representación gráfica de datos para facilitar la interpretación de patrones, tendencias y agrupamientos obtenidos a partir de algoritmos de clusterización y series de tiempo.

ANEXOS

La Variedad de sabores en la familia de Jugos LEYDE S.A. obtenidas del portal de Facebook oficial de la empresa donde el producto estrella sigue siendo en de Naranja y los demás son frutos de temporada limitado al abastecimiento de los zumos de las materias primas y a los precios del mercado de los mismos como: Naranja Agria, Limonada Rosa, Ponche de frutas, tamarindo, Jamaica, te frio con limón.



Figura 35. Fotografía de la Presentación del Jugo Coctel de Frutas de la Empresa LEYDE S.A.

Fuente (Facebook-LEYDE)



Figura 36. Fotografía de la Presentación de los Jugo de diferentes Frutas de la Empresa LEYDE S.A.

Fuente (Facebook-LEYDE)



Figura 37, Fotografía del Famoso Jugo de Naranja que más venta tiene

Fuente (Facebook-LEYDE)



Figura 38. Fotografía del Jugo de Piña de la Empresa LEYDE S.A.
Fuente (Facebook-LEYDE)



Figura 39. Nuevas presentaciones de Jugos como Naranja Agria y Jamaica de la Empresa LEYDE
Fuente (Facebook-LEYDE)

En Honduras, varias marcas de empresas producen y comercializan jugos. Estas empresas brindan una amplia variedad de opciones de jugos y néctares en el mercado hondureño. Hay desde grandes compañías internacionales hasta marcas locales que se enfocan en ofrecer productos más frescos y naturales. Entre todas dominan el mercado de jugos en Honduras, con una amplia gama de productos que incluyen jugos 100% naturales, néctares y bebidas con sabores tropicales y populares en la región. Tabla del Benchmarking de la Competencia

Tabla 10. Benchmarking o Estudio de la Competencia

1. Tropicigas (Del Frutal)

- Jugos comercializados: Jugo Del Frutal (disponible en sabores como naranja, piña, mango, manzana, y más).
- Descripción: Del Frutal es una de las marcas más populares de jugos en el país, ofreciendo opciones de néctares y jugos 100% naturales.

2. SULA

- Jugos comercializados: Jugos SULA (naranja, piña, manzana, mango, entre otros).
- Descripción: La empresa Lactosa, propietaria de SULA, produce una amplia variedad de productos lácteos y jugos. SULA es conocida por ofrecer jugos con o sin pulpa y néctares.

3. Tropical

- Jugos comercializados: Jugos Tropical (naranja, piña, uva, y más).
- Descripción: Tropical es una marca local que produce bebidas de frutas y néctares con una gran variedad de sabores.

4. Big Cola

- Jugos comercializados: Jugos Big Fresh (piña, mango, naranja).
- Descripción: Aunque más conocida por sus refrescos, Big Cola también tiene una línea de jugos bajo la marca Big Fresh, disponible en sabores tropicales.

5. Coca-Cola (Jugos Del Valle)

- Jugos comercializados: Jugos Del Valle (naranja, manzana, uva, durazno).
- Descripción: Del Valle es una de las marcas más grandes de jugos en América Latina, y en Honduras se distribuyen bajo la licencia de Coca-Cola.

6. Zumo (Distribuidora de Bebidas ZUMO)

- Jugos comercializados: Jugos Zumo (naranja, piña, tropical).
- Descripción: Zumo es una empresa que se enfoca en ofrecer jugos naturales y tropicales en el mercado hondureño.

7. Florida Bebidas

- Jugos comercializados: Jugos y néctares de frutas como piña, mango, y naranja.
- Descripción: Aunque es una empresa costarricense, Florida Bebidas también tiene presencia en Honduras con su línea de jugos naturales.

8. Del Monte

- Jugos comercializados: Jugos Del Monte (piña, naranja, guayaba, durazno).
- Descripción: Del Monte es una marca internacional con una gran presencia en Honduras, ofreciendo jugos a base de frutas tropicales y sabores populares.

9. Jumex

- Jugos comercializados: Jumex (néctares y jugos de durazno, mango, manzana, guayaba, entre otros).
- Descripción: Jumex es una marca mexicana que también tiene distribución en Honduras, conocida por sus néctares y jugos.

10. La Famosa

- Jugos comercializados: Jugos de naranja, piña, manzana, y más.
- Descripción: La Famosa es una empresa centroamericana que tiene una presencia significativa en Honduras, produciendo jugos y néctares de frutas tropicales.

11. Dole

- Jugos comercializados: Jugos de piña y otras frutas tropicales.
- Descripción: Dole es conocida principalmente por la producción de frutas frescas, pero también comercializa jugos de alta calidad, especialmente jugos de piña.

12. Florida's Natural

- Jugos comercializados: Jugo de naranja 100% natural.
- Descripción: Aunque es una marca internacional, Florida's Natural está presente en Honduras y se especializa en jugos de naranja sin conservantes ni azúcar añadida.

13. Tampico Beverages

- Jugos comercializados: Jugos y néctares en sabores como mango, naranja, durazno, piña.
- Descripción: Tampico ofrece una variedad de bebidas de frutas con una mezcla de sabores tropicales y populares en América Latina.

14. Naturale

- Jugos comercializados: Jugos 100% naturales de naranja, piña, guayaba, y otros.
- Descripción: Naturale es una marca más pequeña que se enfoca en ofrecer jugos sin azúcar añadida ni conservantes, promoviendo opciones más saludables.

15. Kerns (Grupo Jumex)

- Jugos comercializados: Jugos de durazno, fresa, mango y otros sabores tropicales.
- Descripción: Kerns es otra marca de Jumex, reconocida por sus jugos y néctares enlatados que también se distribuyen en Honduras.

16. Natura's

- Jugos comercializados: Néctares de naranja, mango, durazno, guayaba, entre otros.
- Descripción: Natura's es una marca centroamericana con una fuerte presencia en el mercado hondureño, conocida por sus néctares envasados de frutas tropicales.

17. Sunny Delight (SunnyD)

- Jugos comercializados: Bebidas de naranja y otros sabores cítricos.
- Descripción: Sunny Delight, o SunnyD, es popular por sus bebidas cítricas que combinan jugo de naranja con otros ingredientes para crear una bebida dulce y refrescante.

18. Concentrados El Rey

- Jugos comercializados: Concentrados de frutas como naranja, piña, limón y más.
- Descripción: Aunque más enfocados en la producción de concentrados para bebidas, los productos de El Rey son muy utilizados en la industria de jugos en Honduras.

19. Banacol

- Jugos comercializados: Jugos de banano y otras frutas tropicales.
- Descripción: Banacol es una empresa centrada en productos de banano, pero también produce jugos y concentrados derivados de esta fruta.

20. Concentrados El Coyol

- Jugos comercializados: Jugos y concentrados de frutas tropicales.
- Descripción: El Coyol produce concentrados que son utilizados por empresas locales para la elaboración de jugos y néctares en Honduras.

21. FIFCO

- Jugos comercializados: Jugos tropicales y naturales.
- Descripción: FIFCO (Florida Ice & Farm Co.) es una empresa costarricense que ha ampliado su presencia en Honduras, ofreciendo bebidas naturales y jugos de frutas.

Fuente: Elaboración Propia



Figura 40. FOTOGRAFIA DEL PLANTEL DE LEYDE SPS (SUCURSAL)
Fuente (Elaboración Propia)

AUTORIZACIÓN DE LA EMPRESA LEYDE S.A.

Atención:

Ing. Virgilio Madrid; Encargado de Planning Corporativo

Estimado Ingeniero:

Yo, Sergio Molina Anderson de San Pedro Sula, autor del trabajo de postgrado titulado:

“IMPLEMENTACIÓN DE MODELOS PREDICTIVOS PARA MEJORAR LAS VENTAS DE LA FAMILIA

DE JUGOS EN LEYDE S.A.” presentado y aprobado en octubre 2024 como requisito previo para optar al título de máster en Analítica de Datos y reconociendo que la presentación del presente documento forma parte de los requerimientos establecidos del programa de maestrías de la Universidad Tecnológica Centroamericana (UNITEC), solicito por este medio el uso de la Data de Ventas Histórica de la familia de Jugos Procesados ya sea en Lempiras o Unidades como crea conveniente para:

- 1) Hacer un análisis con herramientas de Aprendizaje automático como es Clusterización para ayudar a identificar oportunidades de mejora en sus Kpi's de Ventas de esa familia para el 2025.
- 2) Hacer un análisis con herramientas de Aprendizaje automático como es Series de Tiempos para ayudar a identificar oportunidades de mejora en sus Kpi's de Ventas de esa familia para el 2025.

Asimismo, el autor cede de forma ilimitada y exclusiva a LEYDE y a UNITEC la titularidad de los derechos patrimoniales. Es entendido que cualquier copia o reproducción del presente documento con fines de lucro no está permitida sin previa autorización por escrito de parte de UNITEC.

En fe de lo cual se suscribe el presente documento en la ciudad de San Pedro Sula, a los 25 días del mes de del año octubre 2024.

Ing. Virgilio Madrid



*** La autorización firmada se encuentra adjunta a mí expediente**

CARTA DE AUTORIZACIÓN DE DERECHOS DE AUTOR

SAN PEDRO SULA, CORTES. ENERO, 2025

AUTORIDADES DE UNITEC

Por la presente, yo, Sergio Molina Anderson, con número de cuenta 12313295, autor de tesis titulada " **IMPLEMENTACIÓN DE MODELOS PREDICTIVOS PARA INCREMENTAR LAS VENTAS DE JUGOS EN LEYDE S.A.**", presentada como requisito parcial para optar al título de Maestría en Analítica de Negocios en la Universidad Tecnológica Centroamericana, declaro lo siguiente:

1. Originalidad del trabajo

Garantizo que la obra mencionada es producto de mi autoría y que cumple con los principios de originalidad y ética académica. Declaro que en el desarrollo de esta obra se han citado todas las fuentes consultadas y se han respetado los derechos de autor de terceros.

2. Cesión de derechos de reproducción

Autorizo a UNITECa incluir una copia digital o física de mi tesis en su repositorio académico, con fines de consulta, difusión y preservación, según las políticas establecidas por la institución.

3. Limitaciones de uso

Esta autorización se otorga con la condición de que mi obra sea utilizada únicamente para fines académicos y sin fines de lucro. Cualquier uso comercial deberá contar con mi autorización expresa.

4. Derechos patrimoniales y morales

Declaro que la cesión de derechos para la reproducción y difusión de mi obra no afecta mis derechos morales como autor, los cuales mantengo de manera irrenunciable e intransferible.

5. Responsabilidad legal

Asumo la responsabilidad legal sobre cualquier reclamo relacionado con el contenido de esta obra, exonerando a UNITEC de cualquier disputa derivada del uso del material presentado.

Sin otro particular, firmo esta carta en señal de conformidad.

Atentamente,



Sergio Molina Anderson

lactovi@gmail.com +504

95029553

**AUTORIZACIÓN DEL AUTOR(ES) PARA LA CONSULTA, REPRODUCCIÓN
PARCIAL O TOTAL Y PUBLICACIÓN FÍSICA Y ELECTRÓNICA DEL TEXTO
COMPLETO DEL TRABAJO FINAL DE GRADUACIÓN DE UNITEC Y CESIÓN DE
DERECHOS PATRIMONIALES**

Señores

CENTRO DE RECURSOS PARA

EL APRENDIZAJE Y LA INVESTIGACION (CRAI)

UNIVERSIDAD TECNOLÓGICA CENTROAMERICANA (UNITEC)

Tegucigalpa

Estimados Señores:

Yo, Sergio Molina Anderson, en San Pedro Sula, autor del trabajo de postgrado titulado: **IMPLEMENTACIÓN DE MODELOS PREDICTIVOS PARA INCREMENTAR LAS VENTAS DE JUGOS EN LEYDE S.A**, presentado y aprobado en Enero, 2025, como requisito previo para optar al título de Master en Analítica de Negocios en UNITEC(en lo sucesivo, el “Trabajo Final de Graduación”) y reconociendo que la presentación del presente documento forma parte de los requerimientos establecidos del programa de Postgrado de la Universidad Tecnológica Centroamericana (UNITEC) y del Centro Universitario Tecnológico (CEUTEC), por este medio **AUTORIZO/AUTORIZAMOS** a la Universidad Tecnológica Centroamericana (UNITEC) y el Centro Universitario Tecnológico (CEUTEC), para que:

- 1) A través de sus Centros Asociados y Bibliotecas de los “Centros de Recursos para el Aprendizaje y la Investigación (CRAI)”, para que con fines académicos, puedan libremente registrar, copiar o utilizar la información contenida en él, con fines educativos, investigativos o sociales. Asimismo, para que exponga mi trabajo como medio didáctico en los Centros de Recursos para el Aprendizaje y la Investigación (CRAI o Biblioteca), y con fines académicos permita a los usuarios de dichos centros su consulta y acceso mediante catálogos electrónicos, repositorios académicos nacionales o internacionales, página web institucional, así como medios electrónicos en general, internet, intranet, DVD, u otro formato conocido o por conocer, así como integrados en programas de cooperación bibliotecaria académicos, que permitan mostrar al mundo la producción académica de la Universidad a través de la visibilidad de su contenido.
- 2) De conformidad con lo establecido en la Ley de Derechos de Autor y de los Derechos Conexos de la República de Honduras, se autoriza para que permita

copiar, reproducir o transferir información del Proyecto de Graduación, conforme su uso educativo y debiendo citar en todo momento la fuente de información; esto permitirá ampliar los conocimientos a las personas que hagan uso del mismo, siempre y cuando resguarden la completa información textual o paráfrasis de esta.

Asimismo, en mi calidad de estudiante y/o autor del Trabajo Final de Graduación acepto que UNITEC/CEUTEC no se hace responsable del uso, reproducciones, venta y distribuciones de todo tipo de fotografías, imágenes, grabaciones, o cualquier otro tipo de presentación relacionado con el Trabajo Final de Graduación que el mismo autor distribuya antes y después de la entrega del documento a la Universidad.

Finalmente, declaro bajo fe de juramento, conociendo las consecuencias penales que conlleva el delito de perjurio: que soy autor(es) del presente Trabajo Final de Graduación, que el contenido de dicho trabajo es obra original de suscrito(s) y de la veracidad de los datos incluidos en el documento. Eximo a UNITEC/CEUTEC; así como el Tutor y Lector que han revisado el presente, por las manifestaciones y/o apreciaciones personales incluidas en el mismo, de cualquier responsabilidad por su autoría o cualquier situación de perjuicio que se pudiera presentar.

De conformidad con lo establecido en los artículos 9.2, 18, 19, 35 y 62 de la Ley de Derechos de Autor y de los Derechos Conexos; los derechos morales pertenecen al autor y son personalísimos, irrenunciables, imprescriptibles e inalienables, asimismo, por tratarse de una obra colectiva, cedo de forma ilimitada y exclusiva a la UNITEC/CEUTEC la titularidad de los derechos patrimoniales que surjan o se deriven del Trabajo Final de Graduación. Es entendido que cualquier copia o reproducción del presente documento con fines de lucro no está permitida sin previa autorización por escrito de parte de UNITEC/CEUTEC.

En fe de lo cual, se suscribe el presente documento en la ciudad de San Pedro Sula, Cortes a los 25 días del mes de Enero del 2025.



Sergio Molina Anderson

12313295