



CARTA DE AUTORIZACIÓN

CÓDIGO

AP-BIB-FO-06

VERSIÓN

1

VIGENCIA

2014

PÁGINA

1 de 2

Neiva, 31 de agosto del 2020

Señores

CENTRO DE INFORMACIÓN Y DOCUMENTACIÓN

UNIVERSIDAD SURCOLOMBIANA

Ciudad

El (Los) suscrito(s):

PEDRO ANDRÉS SÁNCHEZ CHARRY, con C.C. No. 1.075.279.194,
_____, con C.C. No. _____,
_____, con C.C. No. _____,
_____, con C.C. No. _____,

Autor(es) de la tesis y/o trabajo de grado o _____

Titulado Aplicación del Machine Learning para el análisis predictivo de las ventas en la taquilla Modulo
enterario del grupo empresarial COOMOTOR

presentado y aprobado en el año 2020 como requisito para optar al título de

MATEMÁTICO

Autorizo (amos) al CENTRO DE INFORMACIÓN Y DOCUMENTACIÓN de la Universidad Surcolombiana para que, con fines académicos, muestre al país y el exterior la producción intelectual de la Universidad Surcolombiana, a través de la visibilidad de su contenido de la siguiente manera:

- Los usuarios puedan consultar el contenido de este trabajo de grado en los sitios web que administra la Universidad, en bases de datos, repositorio digital, catálogos y en otros sitios web, redes y sistemas de información nacionales e internacionales "open access" y en las redes de información con las cuales tenga convenio la Institución.
- Permita la consulta, la reproducción y préstamo a los usuarios interesados en el contenido de este trabajo, para todos los usos que tengan finalidad académica, ya sea en formato Cd-Rom o digital desde internet, intranet, etc., y en general para cualquier formato conocido o por conocer, dentro de los términos establecidos en la Ley 23 de 1982, Ley 44 de 1993, Decisión Andina 351 de 1993, Decreto 460 de 1995 y demás normas generales sobre la materia.
- Continúo conservando los correspondientes derechos sin modificación o restricción alguna; puesto que, de acuerdo con la legislación colombiana aplicable, el presente es un acuerdo jurídico que en ningún caso conlleva la enajenación del derecho de autor y sus conexos.

Vigilada Mineducación



CARTA DE AUTORIZACIÓN

CÓDIGO

AP-BIB-FO-06

VERSIÓN

1

VIGENCIA

2014

PÁGINA

2 de 2

De conformidad con lo establecido en el artículo 30 de la Ley 23 de 1982 y el artículo 11 de la Decisión Andina 351 de 1993, “Los derechos morales sobre el trabajo son propiedad de los autores”, los cuales son irrenunciables, imprescriptibles, inembargables e inalienables.

EL AUTOR/ESTUDIANTE:

Firma: _____

EL AUTOR/ESTUDIANTE:

Firma: _____

EL AUTOR/ESTUDIANTE:

Firma: _____

EL AUTOR/ESTUDIANTE:

Firma: _____



TÍTULO COMPLETO DEL TRABAJO: Aplicación del Machine Learning para el análisis predictivo de las ventas en la taquilla Modulo Centenario del grupo empresarial COOMOTOR

AUTOR O AUTORES:

| Primero y Segundo Apellido | Primero y Segundo Nombre |
|----------------------------|--------------------------|
| SÁNCHEZ CHARRY | PEDRO ANDRÉS |

DIRECTOR Y CODIRECTOR TESIS:

| Primero y Segundo Apellido | Primero y Segundo Nombre |
|----------------------------|--------------------------|
| MEDINA ARCE | YINETH |

ASESOR (ES):

| Primero y Segundo Apellido | Primero y Segundo Nombre |
|----------------------------|--------------------------|
| MEDINA ARCE | YINETH |

PARA OPTAR AL TÍTULO DE: MATEMÁTICO

FACULTAD: CIENCIAS EXACTAS Y NATURALES

PROGRAMA O POSGRADO: MATEMÁTICA APLICADA

CIUDAD: NEIVA

AÑO DE PRESENTACIÓN: 2020

NÚMERO DE PÁGINAS: 63

TIPO DE ILUSTRACIONES (Marcar con una X):

Diagramas X Fotografías ___ Grabaciones en discos ___ Ilustraciones en general X Grabados ___
Láminas ___ Litografías ___ Mapas ___ Música impresa ___ Planos ___ Retratos ___ Sin ilustraciones ___ Tablas
o Cuadros X



SOFTWARE requerido y/o especializado para la lectura del documento: EXCEL, R-STUDIO. PYTHON, OVERLEAF (LATEX).

MATERIAL ANEXO: Tablas de datos suministrados por el grupo empresarial COOMOTOR.

PREMIO O DISTINCIÓN (*En caso de ser LAUREADAS o Meritoria*):

PALABRAS CLAVES EN ESPAÑOL E INGLÉS:

| <u>Español</u> | <u>Inglés</u> | <u>Español</u> | <u>Inglés</u> |
|----------------|---------------|------------------------|---------------------|
| 1. Ventas | Sales | 6. Conjunto de datos | Datasets |
| 2. Aprendizaje | Learning | 7. Minería de datos | Data mining |
| 3. Maquina | Machine | 8. predicción | Prediction |
| 4. Regresión | Regression_ | 9. Almacén de datos | Data warehouse |
| 5. Polinomial | Polynomial | 10. Regresión múltiple | Multiple regression |

RESUMEN DEL CONTENIDO: (Máximo 250 palabras)

El proyecto tiene como propósito aplicar los Modelos de Predicción de forma dinámica, sobre el método Data Scientist, para estimar las metas de ventas de la taquilla Modulo Centenario, ubicado en el terminal de transporte del grupo empresarial COOMOTOR; teniendo en cuenta factores que afectan el número de pasajes vendidos, como son los turnos, las temporadas, el puesto de la taquilla, entre otros. El proceso de desarrollo comprendió cinco etapas:

Etapa 1: Delimitación del problema taquilla "modulo centenario" del terminal de transporte e identificación de los posibles factores que afectan las ventas en la taquilla.

Etapa 2: En esta etapa se realizó, la extracción de datos de las ventas realizadas durante el periodo comprendido desde el año 2011 hasta mediados del 2019, la cual se obtuvo de los archivos del software FICS, Se continuo con la exportación de datos a la hoja de cálculo de EXCEL; para la limpieza de datos se utilizó el lenguaje de programación PYTHON, y finalmente se creó los dataset.

Etapa 3: Se exportó el dataset a R-Studio, igualmente se instalaron los paquetes necesarios para gráficos, se activó la librería y la función plots (Diagrama de dispersión). Se emplearon algunos parámetros lo que permitió visualizar el comportamiento de los datos y la correlación entre ellos.

Etapa 4: Modelación y evaluación de datos. Al dataset de entrenamiento se le hizo el análisis de correlación y mediante modelos de regresión se obtuvo el modelo matemático, luego se hizo la validación y simulación del mismo.

Etapa 5: Mediante un informe escrito se comunicó los resultados de interés y puestas en producción a la gerencia de la empresa.



ABSTRACT: (Máximo 250 palabras)

The purpose of the project is to apply the Prediction Models dynamically, on the Data Scientist method, to estimate the sales goals of the Modulo Centenario box office, located in the transport terminal of the COOMOTOR business group; taking into account factors that affect the number of tickets sold, such as shifts, seasons, ticket office, among others. The development process comprised five stages: Stage 1: Delimitation of the "centennial module" box office problem of the transport terminal and identification of the possible factors that affect sales at the box office. Stage 2: In this stage, the extraction of data from sales made during the period from 2011 to mid-2019 was carried out, which was obtained from the files of the FICS software, it continued with the export of data to the EXCEL spreadsheet; The PYTHON programming language was used to clean the data, and finally the datasets were created. Stage 3: The dataset was exported to R-Studio, the necessary packages for graphics were also installed, the library and the plots function (Scatter diagram) were activated. Some parameters were used which allowed visualizing the behavior of the data and the correlation between them. Stage 4: Data modeling and evaluation. Correlation analysis was performed on the training dataset and the mathematical model was obtained through regression models, then validated and simulated. Stage 5: By means of a written report, the results of interest and putting into production were communicated to the management of the company.

APROBACION DE LA TESIS

Nombre Presidente Jurado: MAURO MONTEALEGRE CARDENAS

Firma:

Nombre Jurado: MAURO MONTEALEGRE CARDENAS

Firma:

Nombre Jurado: EDINSON OSWALDO DELGADO RIVAS

Firma:



UNIVERSIDAD
SURCOLOMBIANA

Facultad de Ciencias Exactas y Naturales
Programa Matemáticas Aplicada

Aplicación del Machine Learning para el
análisis predictivo de las ventas en la taquilla
Modulo Centenario del grupo empresarial
COOMOTOR

Presentado por:

Pedro Andrés Sánchez Charry

Neiva - Huila

2020

Índice general

| | |
|---|-----------|
| 1. Introducción | 5 |
| 2. Descripción del problema | 7 |
| 2.1. Pregunta del problema | 7 |
| 2.2. Antecedentes | 7 |
| 2.2.1. Modelo de regresión lineal múltiple para el pronóstico de ventas de bolsas ecológicas para la empresa | 8 |
| 2.2.2. IA y "machine learning" para optimizar el transporte de mercancías por carretera | 8 |
| 2.2.3. SmartBus: Big Data y Data Science en Transporte Urbano | 8 |
| 2.2.4. Aplicación del análisis de regresión lineal simple para la estimación de los precios de las acciones de Facebook, Inc. | 8 |
| 3. Justificación | 9 |
| 4. Marco teórico | 11 |
| 4.1. Data mining | 11 |
| 4.1.1. Aspectos básicos del Data mining | 11 |
| 4.2. KDD (KNOWLEDGE DISCOVERY IN DATABASES) | 12 |
| 4.2.1. Data Warehouse o almacén de datos | 12 |
| 4.2.2. Procesos del KDD | 12 |
| 4.3. El ámbito de los modelos predictivos | 13 |
| 4.4. Algoritmos estadísticos | 13 |
| 4.4.1. Modelo de regresión lineal simple | 13 |
| 4.4.2. Ecuación de regresión lineal simple | 13 |
| 4.4.3. Ecuación de regresión estimada simple | 14 |
| 4.4.4. Método de mínimos cuadrados | 15 |
| 4.4.5. Modelo de regresión múltiple | 16 |
| 4.4.6. Ecuación de regresión múltiple | 16 |
| 4.4.7. Ecuación de regresión múltiple estimada | 16 |
| 4.4.8. Transformaciones de modelos de regresión no lineales | 17 |
| 4.4.9. Transformaciones de modelos cuadráticos | 18 |
| 4.4.10. Variables ficticias (o dummies) | 18 |
| 4.5. Machine learning | 18 |
| 4.6. Relevancias significativas del data science | 19 |
| 5. Objetivos | 21 |
| 5.1. Objetivo general | 21 |
| 5.2. Objetivos específicos | 21 |
| 6. Metodología | 23 |
| 6.1. Diseño de la investigación | 23 |
| 6.2. Enfoque de la investigación | 23 |
| 6.3. Población | 23 |
| 6.4. Herramienta metodológica | 24 |

| | |
|---|-----------|
| 7. Análisis y discusión de resultados | 25 |
| 7.1. Etapa 1: Enmarcar el problema | 25 |
| 7.1.1. ¿Cuánto venden aproximadamente en temporada alta? | 25 |
| 7.1.2. ¿Cuánto venden aproximadamente en temporada baja? | 27 |
| 7.1.3. ¿Las ventas dependen del día en la semana? | 28 |
| 7.1.4. ¿Las ventas de un fin de semana con día festivo son la misma sin un día festivo? | 29 |
| 7.1.5. Conclusión | 29 |
| 7.2. Etapa 2: Adquirir y preparar los datos | 30 |
| 7.3. Etapa 3: Explorar los datos | 31 |
| 7.4. Etapa 4: Modelizar y evaluar los datos | 33 |
| 7.5. Etapa 5: Comunicar los resultados | 37 |
| | |
| 8. Conclusiones | 45 |
| | |
| 9. Anexos | 47 |
| 9.1. Anexo 1: Base de datos sin filtrar | 47 |
| 9.2. Anexo 2: Base de datos filtrada | 51 |
| 9.3. Anexo 3: Base de datos con variables dummies | 55 |
| 9.4. Anexo 4: Tabulación de los resultados del modelo predictivos | 57 |

Capítulo 1

Introducción

El proyecto tiene como propósito aplicar los Modelos de Predicción de forma dinámica, sobre el método Data Scientist, para estimar las metas de ventas de la taquilla Modulo Centenario, ubicado en el terminal de transporte del grupo empresarial COOMOTOR; teniendo en cuenta factores que afectan el número de pasajes vendidos , como son los turnos, las temporadas, el puesto de la taquilla, entre otros. El proceso de desarrollo comprendió cinco etapas:

Etapas 1: Delimitación del problema taquilla "modulo centenario"del terminal de transporte e identificación de los posibles factores que afectan las ventas en la taquilla.

Etapas 2: En esta etapa se realizó, la extracción de datos de las ventas realizadas durante el periodo comprendido desde el año 2011 hasta mediados del 2019, la cual se obtuvo de los archivos del software FICS, Se continuo con la exportación de datos a la hoja de cálculo de EXCEL; para la limpieza de datos se utilizó el lenguaje de programación PYTHON, y finalmente se creó los dataset.

Etapas 3: Se exportó el dataset a R-Studio, igualmente se instalaron los paquetes necesarios para gráficos, se activo la libreria y la función plots (Diagrama de dispersión). Se emplearon algunos parámetros lo que permitió visualizar el comportamiento de los datos y la correlación entre ellos.

Etapas 4: Modelación y evaluación de datos. Al dataset de entrenamiento se le hizo el análisis de correlación y mediante modelos de regresión se obtuvo el modelo matemático, luego se hizo la validación y simulación del mismo.

Etapas 5: Mediante un informe escrito se comunicó los resultados de interés y puestas en producción a la gerencia de la empresa.

Capítulo 2

Descripción del problema

Hoy en día todas las empresas desean saber cuanto venderán en un futuro ya sea próximo o lejano, esto con el fin de crear metas o incluso hacer inversiones para la empresa, ya que al saber la cantidad de ingresos que se esperan, los administradores podrán tomar decisiones con mayor facilidad y tener un control sobre la misma. El grupo empresarial COOMOTOR hizo una inversión de 60 mil millones a finales del año 2018, esta inversión ha sido para comprar la mayor cantidad de acciones de las terminales de Cali, Melgar y Armenia. Por otro lado, también se invirtió en estaciones de gasolina y CDA's, esto deja a la empresa con una deuda muy grande y nace la necesidad de optimizar las ventas de las taquillas de todo el país.

2.1. Pregunta del problema

En este trabajo de grado se desea saber **¿Como aportar al grupo empresarial COOMOTOR un modelo matemático para recuperar lo invertido por medio de ventas de pasajes en la taquilla "Modulo Centenario" ?** Para resolver esta pregunta, fue necesario utilizar la base de datos de las ventas en esta taquilla, donde la empresa la facilito y así lograr su debido análisis.

Normalmente para las predicciones de ventas se utilizan técnicas como: series de tiempo, redes neuronales artificiales, inteligencia artificial, entre otros. Para este trabajo se realizó la investigación por medio de una regresión polinomial multivariada, ya que es una técnica del Data Science y además está al alcance de lo visto durante el ciclo académico del pregrado de matemática aplicada. Por otro lado, también consideré que era muy importante hacerlo por este medio ya que si la empresa desea tener un progreso y recuperar lo invertido en el menor tiempo posible, es importante superar la línea de la función resultante en la regresión. para poder encontrar dicha ecuación de predicción, se toma la base de datos y por medio de una simulación por PYTHON se obtendrán un modelo matemático donde se verá los resultados sobre las ventas esperadas por los días del año. El alcance que tiene este modelo, es a mediano plazo, sin embargo, se puede prolongar de tal manera que solo se debe actualizar los datos hasta la fecha donde se aplique nuevamente la simulación para adquirir el modelo.

Es importante resaltar que esta propuesta de investigación se realizó antes de la pandemia por el **covid-19** y que se está viviendo a la fecha de hoy, por lo tanto en los resultados no se tuvo en cuenta esta situación.

2.2. Antecedentes

Desde hace unos años el estudio del análisis predictivo de los datos, a lo que hoy conocemos data science o machine learning, ha despertado el interés de muchos investigadores y empresas de todos los sectores a nivel mundial. A partir de ese momento, se han realizado trabajos de investigación relacionados con el machine learning.

La gestión de los investigadores actualmente está en constante cambio. Por consecuencia los cuestionarios de modelos predictivos son valiosos para conocer diversas opiniones y así ayudar a fortalecer las

futuras investigaciones.

A continuación, conocerán algunos trabajos de investigación con respecto a la regresión lineal, regresión lineal múltiple, machine learning y big data aplicado al sector de la economía y del transporte.

2.2.1. Modelo de regresión lineal múltiple para el pronóstico de ventas de bolsas ecológicas para la empresa

En el año 2020, Geraldine Forero Gómez y Jonathan Alexander Martínez Lozano, realizaron una investigación sobre "Modelo de regresión lineal múltiple para el pronóstico de ventas de bolsas ecológicas para la empresa". La información que buscan con la empresa Boleco SA respecto a la venta de bolsas ecológicas es a que segmento de mercado se venderá y en que patrones de tiempo es necesaria la planeación de todas las áreas pertenecientes al proceso productivo de la empresa; en los pronósticos de ventas se deben clasificar estas necesidades a corto y a largo plazo.

Al crear un modelo mediante el método de regresión lineal múltiple para un pronóstico de ventas de bolsas ecológicas permitió conocer de forma más exacta y precisa las características con las cuales se debe interactuar, y de esta forma definir las variables necesarias que intervienen directamente en el proceso de venta para lograr el objetivo deseado.

2.2.2. IA y "machine learning" para optimizar el transporte de mercancías por carretera

En el año 2019, la empresa Ontruck operador español de transporte regional de mercancías por carretera, desarrolla algoritmos que con ayuda de inteligencia artificial (IA), machine learning y business intelligence (BI) poder hacer un cálculo dinámico de los precios de cada servicio, el en rutado de las mercancías teniendo en cuenta la situación en tiempo real de las carreteras y la asignación de cargas según el tipo de carga y el perfil de los transportistas.

A través de estas tecnologías, Ontruck quiere ofrecer, tanto a sus clientes como a los transportistas, " un servicio más personalizado", así como una gestión de cargas inteligente, con el objetivo de mejorar los índices de satisfacción de ambas partes. Para ello, además de sus desarrollos propios utiliza otras tecnologías como servicios GIS para geolocalización, geocoding y el tiempo estimado de llegada (ETA).

2.2.3. SmartBus: Big Data y Data Science en Transporte Urbano

En el año 2017, Carlos Rosado Mora presenta un trabajo de investigación que trata sobre la realización de un estudio y análisis de ciertos tipos de datos procedentes de la red de autobuses de Madrid (EMT). El sistema se basa en una plataforma Big Data mediante la cual se almacenarán grandes cantidades de datos procedentes de diversas fuentes, tanto internas a la EMT como externas, con el fin de aplicar técnicas de aprendizaje automático para segmentar las líneas de autobús y posteriormente realizar una predicción de la demanda diaria en una de ellas. Se realizará un análisis de los resultados obtenidos mediante el cual se pueda mejorar la eficiencia de la red de autobuses de la EMT por medio del análisis de técnicas de regresión y clustering en grandes cantidades de datos. Para finalizar, se implementará un DashBoard mediante el cual el usuario final pueda ver los resultados de nuestro estudio y pueda tomar decisiones en base a un criterio basado en grandes cantidades de datos.

2.2.4. Aplicación del análisis de regresión lineal simple para la estimación de los precios de las acciones de Facebook, Inc.

En el año 2017, Brenes González, realizó una investigación sobre " Aplicación del análisis de regresión lineal simple para la estimación de los precios de las acciones de Facebook, Inc.". Donde el análisis de regresión lineal, es una herramienta sumamente importante en el mundo de las Finanzas, debido a que permite realizar proyecciones y pronósticos de una variable dependiente explicada por una o más variables independientes. El objetivo de este trabajo, fue determinar una ecuación que permitiera estimar el precio promedio mensual de las acciones de la empresa Facebook, Inc., a través de un modelo de regresión lineal simple. Los coeficientes betas estimados para el modelo fueron significativos tanto para la constante como para la pendiente, medidos a través de la prueba estadística t. Así mismo, se realizó la prueba global de significancia de los coeficientes betas, determinada a través de la prueba estadística F, esta resultó ser sumamente significativa, lo cual, permitió la validación del modelo.

Capítulo 3

Justificación

Actualmente Data Scientist es una de las profesiones mas solicitadas en el mercado, como en la economia, seguridad, salud o educación. Los profesionales Expertos en éste método científico además de utilizar las técnicas las estadísticas para explorar y analizar grandes cantidades de datos de múltiples fuentes y hasta en diferentes formatos, tienen una fuerte visión de negocio para ser capaz de extraer y transmitir recomendaciones a los responsables de negocio de su empresa.

El grupo empresarial COOMOTOR, es orgullo para el departamento del Huila, siendo la mejor de la región en el tema de transporte terrestre. Este trabajo de grado, mediante el uso de las técnicas de análisis de datos, data science , tiene como objetivo extraer y transmitir predicciones sobre las ventas a futuro, que lleven a la empresa, no solo a ser la más taquillera y solicitada en el terminal de transporte de Neiva, sino que ella misma trate de superar sus propias expectativas de ventas lo mejor posible. También la gerencia lo puede utilizar como referente de apoyo en la toma de decisiones en temas de inversiones para el mejoramiento en al área comercial de la empresa , estrategias de venta, en la estructuración de planes de contingencias en caso de dificultades debidas por ejemplo, a desastres naturales, paros nacionales, o problemas de salud publica entre otros. Para lograr lo propuesto se contó con software (FICS) dotado de la base de datos en ventas otorgada por el área de sistemas, el software libre como PYTHON para la modelación, ejecución y validación del modelo matemático propuesto.

Capítulo 4

Marco teórico

4.1. Data mining

Según (SAS THE POWER TO KNOW, 2019) Son distintos procesos y métodos cualitativos en donde trabajamos con grandes cantidades de datos haciendo uso y ayudándonos de herramientas como las bases de datos, los procesos estocásticos, estadísticos y también algoritmos computacionales para generar información oculta entre los datos. La información al ser analizada y estudiada va generando conocimiento.

" El data mining nace a raíz de la cuarta revolución industrial o más conocida como la industria 4.0 La inteligencia artificial es señalada como elemento central de esta transformación, íntimamente relacionada la acumulación creciente de grandes cantidades de datos (big data), el uso de algoritmos para procesarlos y la interconexión masiva de sistemas y dispositivos digitales" (Andrea Talero Molano, 2019). Hoy en día las personas en vez de extraer minerales, extraen grandes cantidades de datos para las empresas.

El proceso del data mining tiene 5 etapas principales para el análisis, estas son:

1. Enmarcar el problema.
2. Adquirir y preparar los datos.
3. Explorar los datos.
4. Modelar y evaluar los datos.
5. Comunicar los resultados y puesta en producción.

4.1.1. Aspectos básicos del Data mining

" En el mundo actual tenemos información por todos los medios, esta información surge de gran variedad de datos" (Juan Gabriel Gomila, 2018); sin embargo, tener datos e información no implica tener conocimiento. Cuando tenemos muchos datos, es importante saber la diferencia entre datos, información y conocimiento.

- **Datos:** Son elementos primarios de información que por sí solos son irrelevantes como apoyo a la toma de decisiones.
- **Información:** La información se puede definir como un conjunto de datos procesados y que tienen un significado (relevancia, propósito y contexto), y que por lo tanto son de utilidad para quien debe tomar decisiones, al disminuir su incertidumbre.
- **Conocimiento:** El conocimiento es una mezcla de experiencia, valores e información que sirve como marco para la incorporación de nuevas experiencias e información, y es útil para la acción.

4.2. KDD (KNOWLEDGE DISCOVERY IN DATABASES)

" Es un proceso en el cual se identifican patrones útiles en los datos para obtener posteriormente el conocimiento útil y valido (relevante)" (Andrea Talero Molano, 2019). El conocimiento que resulta debe ser presentado de manera clara y comprensible ya que de esta manera se puede decir que el trabajo fue un éxito (El trabajo de generar conocimiento).

4.2.1. Data Warehouse o almacén de datos

Es una colección de datos de varias fuentes que están almacenadas bajo un esquema unitario, usualmente este esquema unitario recibe en un único lugar de esta manera:



Figura 4.1: Data Warehouse.

lo que hace es tomar esas tres bases de datos que pueden estar en sitios distintos, almacenarlos datos relativos al estudio que se va a realizar en un solo almacén.

Al tener un almacén de datos, beneficia en que, almacenarlo ahorraría bastante tiempo y además bastante recursos que también es muy importante para la empresa. Ya que, al tener distintas bases de datos separadas, va a tomar tiempo en organizar y consumiría recursos.

4.2.2. Procesos del KDD

los procesos del KDD se utilizan en la segunda etapa del análisis de datos, estos son:

1. Se comienza con los datos (Esta es la fuente de análisis del KDD).
2. Posteriormente de estos datos se hacen una selección que es almacenada en el data warehouse.
3. Luego los datos del datawarehouse son procesados.
4. Luego son transformados, estos datos transformados son los que van a servir de entrada al proceso de minería de datos y de salida se convertirá en los patrones.

Esto se resume en el siguiente diagrama:

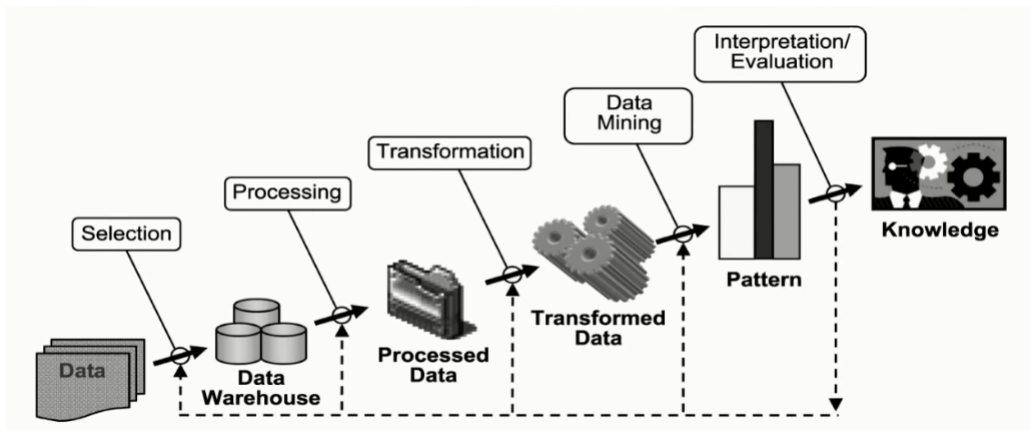


Figura 4.2: Proceso del KDD.

En conclusión, el KDD lo que hace es procesar grandes cantidades de datos, para identificar patrones relevantes en estos datos y para generar al final, resultados que serán convertidos en conocimiento.

4.3. El ámbito de los modelos predictivos

Según (uan Gabriel Gomila, 2018) La modelización predictiva se define como un conjunto de algoritmos de la estadística, que no se puede hacer a mano si no con un software, que cuando se aplican a datos es capaz de votar una función matemática donde va ser útil para la empresa.

4.4. Algoritmos estadísticos

Son la plantilla del modelo, para llevar a cabo una receta, se responsabiliza de crear ecuaciones matemáticas basados en datos históricos, analiza los datos, cuantifica las relaciones entre las variables, la convierte en una ecuación matemática y esto nos servirá para predecir datos.

4.4.1. Modelo de regresión lineal simple

Según (Pat Fernandez, Martínez Menchaca, Pat Fernández, y Martínez Luis, 2013) La ecuación con que se describe la relación entre y con x y en la que se da un término para el error, se le llama modelo de regresión. El modelo que se emplea en la regresión lineal simple es el siguiente.

$$y = \beta_0 + \beta_1 x + \epsilon \quad (4.1)$$

β_0 y β_1 se conocen como los parámetros del modelo, y ϵ (la letra griega épsilon) es una variable aleatoria que se conoce como término del error. El término del error da cuenta de la variabilidad de y que no puede ser explicada por la relación lineal entre x y y .

4.4.2. Ecuación de regresión lineal simple

la ecuación que describe la relación entre el valor esperado de y , que se denota $E(x)$, y x se le llama ecuación de regresión. La ecuación de regresión para la regresión lineal simple es la siguiente.

$$E(x) = \beta_0 + \beta_1 x \quad (4.2)$$

La gráfica de la ecuación de regresión lineal simple es una línea recta; β_0 es la intersección de la recta de regresión con el eje y , β_1 es la pendiente y $E(x)$ es la media o valor esperado de y para un valor dado de x .

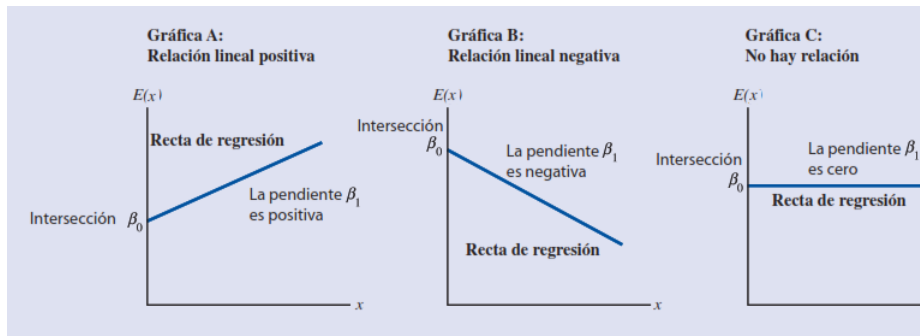


Figura 4.3: Gráfica de la regresión lineal.

La recta de regresión de la gráfica A indica que el valor medio de y está relacionado positivamente con x . La recta de regresión de la gráfica B indica que el valor medio de y está relacionado negativamente con x , valores menores de $E(x)$ corresponden a valores mayores de x . La recta de regresión de la gráfica C muestra el caso en el que el valor medio de y no está relacionado con x ; es decir, el valor medio de y es el mismo para todos los valores de x .

4.4.3. Ecuación de regresión estimada simple

Si se conocieran los valores de los parámetros poblacionales β_0 y β_1 se podría emplear la ecuación 4.2 para calcular el valor medio de y para un valor dado de x . Sin embargo, en la práctica no se conocen los valores de estos parámetros y es necesario estimarlos usando datos muestrales. Se calculan estadísticos muestrales (que se denotan b_0 y b_1) como estimaciones de los parámetros poblacionales β_0 y β_1 . Sustituyendo en la ecuación de regresión b_0 y b_1 por los valores de los estadísticos muestrales β_0 y β_1 se obtiene **la ecuación de regresión estimada**.

La ecuación de regresión estimada de la regresión lineal simple se da a continuación.

$$\hat{y} = b_0 + b_1x \quad (4.3)$$

Como el valor de \hat{y} proporciona tanto una estimación puntual de $E(x)$ para un valor dado de x como una estimación puntual de un solo valor de y para un valor dado de x , a \hat{y} se le llamará simplemente valor estimado de y .

En la siguiente figura se presenta en forma resumida el proceso de estimación en la regresión lineal simple.

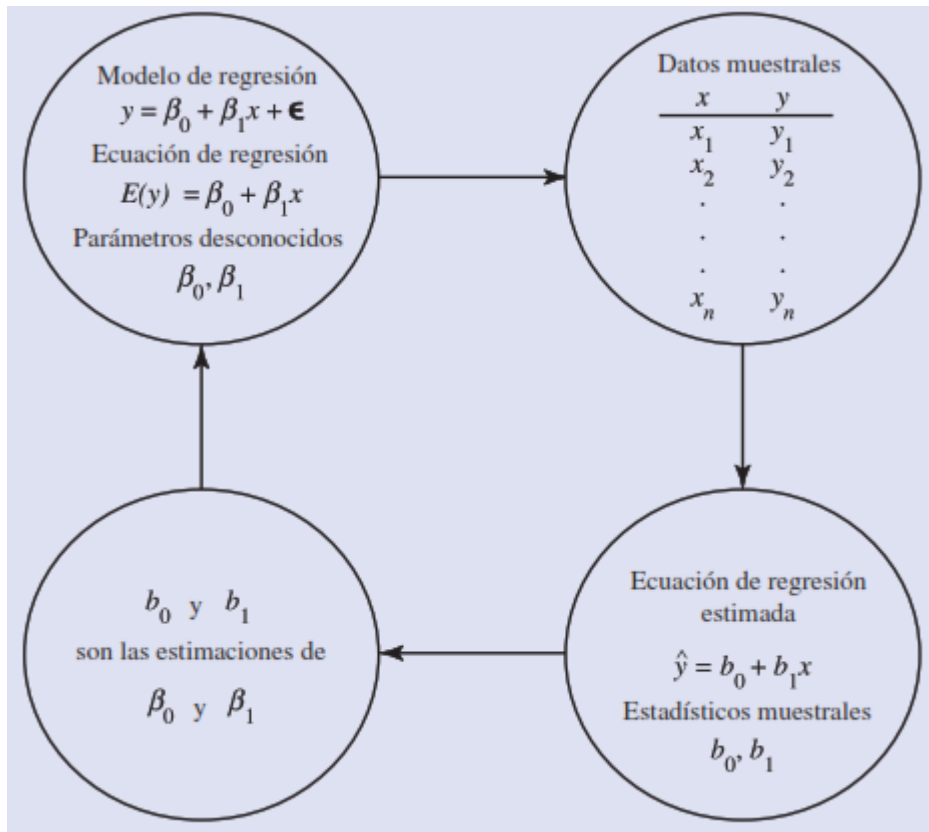


Figura 4.4: Procesos de estimación en la regresión lineal.

4.4.4. Método de mínimos cuadrados

Según (Anderson D., Sweeney D. y Williams T., 2001) En el método de mínimos cuadrados se usan los datos muestrales para obtener los valores de b_0 y b_1 que minimicen la suma de los cuadrados de las desviaciones (diferencias) entre los valores observados de la variable dependiente y_i y los valores estimados de la variable dependiente. El criterio que se emplea en el método de mínimos cuadrados es el de la expresión 4.4.

$$\min \sum (y_i - \hat{y}_i)^2 \quad (4.4)$$

Donde

y_i = valor observado de la variable dependiente en la observación i .

\hat{y}_i = valor estimado de la variable dependiente en la observación i .

Se puede usar cálculos diferenciales para demostrar que los valores de b_0 y b_1 que minimiza la expresión 4.4 se pueden encontrar usando las ecuaciones 4.5 y 4.6.

$$b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \quad (4.5)$$

$$b_0 = \bar{y} - b_1 \bar{x} \quad (4.6)$$

Donde

x_i = Valor de la variable independiente en la observación i .

y_i = Valor de la variable dependiente en la observación i .

\bar{x} = Media de la variable independiente.

\bar{y} = Media de la variable dependiente.

4.4.5. Modelo de regresión múltiple

Según (Mario F. Triola 2018) los conceptos de modelo de regresión y ecuación de regresión lineal simple, son aplicables en el caso de la regresión múltiple. A la ecuación que describe cómo está relacionada la variable dependiente y con las variables independiente x_1, x_2, \dots, x_p se le conoce como **modelo de regresión múltiple**.

Se supone que el modelo de regresión múltiple toma la forma siguiente

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \epsilon \quad (4.7)$$

En el modelo de regresión múltiple $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ son parámetros y el término del error ϵ es una variable aleatoria. Examinando con atención este modelo se ve que y es una función lineal de x_1, x_2, \dots, x_p más el término del error ϵ . El término del error corresponde a la variabilidad en y que no puede atribuirse o explicarse al efecto lineal de las p variables independientes.

4.4.6. Ecuación de regresión múltiple

Se discutirán los supuestos para el modelo de regresión múltiple y para ϵ . Uno de los supuestos es que la media o valor esperado de ϵ es cero. Una consecuencia de este supuesto es que la media o valor esperado de y , que se denota $E(y)$, es igual a $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$. A la ecuación que describe cómo está relacionada la media de y con x_1, x_2, \dots, x_p se le conoce como **ecuación de regresión múltiple**.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad (4.8)$$

4.4.7. Ecuación de regresión múltiple estimada

Si se conocieran los valores de $\beta_0, \beta_1, \dots, \beta_p$ se podría usar la ecuación 4.11 para calcular la media de las y para valores dados de x_1, x_2, \dots, x_p . Desafortunadamente, los valores de estos parámetros no suelen conocerse, es necesario estimarlos a partir de datos muestrales. Para calcular los valores de los estadísticos muestrales b_1, b_2, \dots, b_p , que se usan como estimadores puntuales de los parámetros $\beta_0, \beta_1, \dots, \beta_p$ se emplea una muestra aleatoria simple. Con los estadísticos muestrales se obtiene la siguiente **ecuación de regresión múltiple estimada**.

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_p x_p \quad (4.9)$$

Donde

$$b_0, b_1, b_2, \dots, b_p \text{ son las estimaciones de } \beta_0, \beta_1, \beta_2, \dots, \beta_p$$

$$\hat{y} = \text{Valor estimado de la variable dependiente}$$

Este proceso de estimación en la regresión múltiple se muestra en la siguiente figura

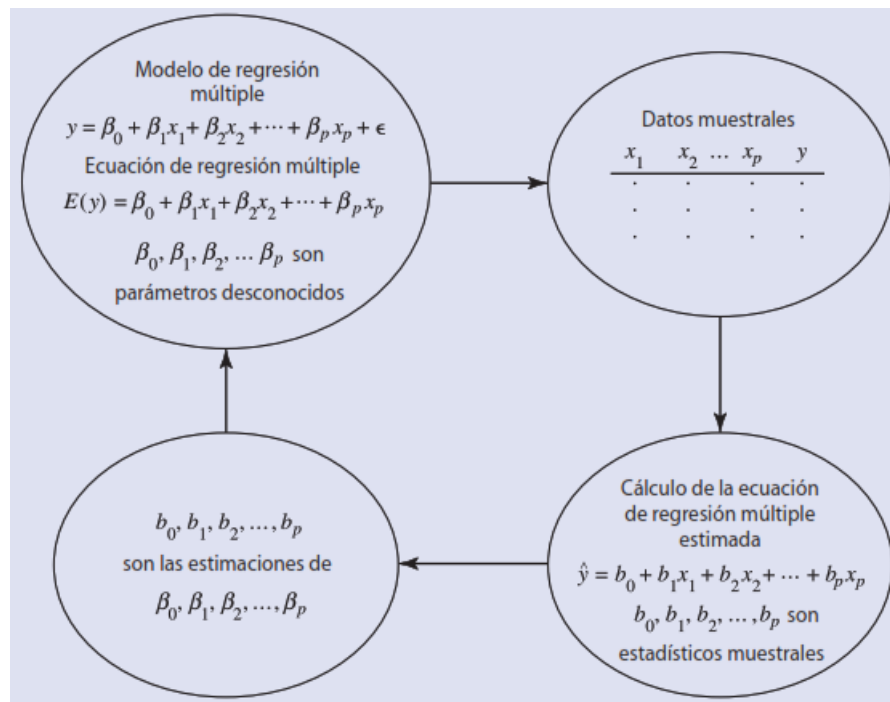


Figura 4.5: proceso de estimación en la regresión múltiple.

4.4.8. Transformaciones de modelos de regresión no lineales

El análisis de regresión puede utilizarse para estimar relaciones lineales que predicen una variable dependiente en función de una o más variables independientes. Estas aplicaciones son muy importantes, sin embargo algunas relaciones económicas y empresariales no son estrictamente lineales.

Examinando el algoritmo de mínimos cuadrados, vemos que manipulando con cuidado los modelos no lineales, es posible utilizar los mínimos cuadrados en un conjunto más amplio de problemas aplicados. Los supuestos sobre las variables independientes en la regresión múltiple no son muy restrictivos. Las variables independientes definen puntos en los que medimos una variable aleatoria Y . Suponemos que hay una relación lineal entre los niveles de las variables independientes X_j donde $j = 1, \dots, k$ y el valor esperado de la variable dependiente Y . Podemos aprovechar esta libertad para ampliar el conjunto de modelos que pueden estimarse. Por lo tanto, podemos ir más allá de los modelos lineales en nuestras aplicaciones del análisis de regresión múltiple. En las siguientes imágenes se ven 3 ejemplos:



Figura 4.6: Gráfica de un modelo no lineal.

- (a) Las funciones de oferta pueden no ser lineales.
- (b) El aumento de la producción total con un aumento del número de trabajadores puede ser cada vez menor a medida que se añaden más trabajadores.
- (c) El coste medio por unidad producida a menudo se minimiza en un nivel de producción intermedio.

4.4.9. Transformaciones de modelos cuadráticos

La función cuadrática

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \epsilon \quad (4.10)$$

puede transformarse en un modelo lineal de regresión múltiple definiendo nuevas variables:

$$\begin{aligned} Z_1 &= X_1 \\ Z_2 &= X_1^2 \end{aligned}$$

y después especificando el modelo

$$y_i = \beta_0 + \beta_1 z_{1i} + \beta_2 z_{2i} + \epsilon_i \quad (4.11)$$

que es lineal en las variables transformadas. Las variables cuadráticas transformadas pueden combinarse con otras variables en un modelo de regresión múltiple. Por lo tanto, podemos ajustar una regresión cuadrática múltiple utilizando variables transformadas. El objetivo es encontrar modelos que sean lineales en otras formas matemáticas de una variable.

4.4.10. Variables ficticias (o dummies)

En el análisis de la regresión múltiple, hemos supuesto hasta ahora que las variables independientes, x_i , existían en un rango y contenían muchos valores diferentes. Sin embargo, en los supuestos de la regresión múltiple la única restricción a la que están sujetas las variables independientes es que son valores fijos. Por lo tanto, podríamos tener una variable independiente que tomara solamente dos valores: $x_i = 0$ y $x_i = 1$. Esta estructura se denomina normalmente variable ficticia o variable dummies, y veremos que constituye un valioso instrumento para aplicar la regresión múltiple a situaciones en las que hay variables categóricas. Un importante ejemplo es una función lineal que varía en respuesta a alguna influencia. consideremos primero una ecuación de regresión simple:

$$Y = \beta_0 + \beta_1 X_1$$

Supongamos ahora que introducimos una variable ficticia, X_2 , que toma los valores 0 y 1 y que la ecuación resultante es:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

Cuando $X_2 = 0$ en esta ecuación, la constante es β_0 , pero cuando $X_2 = 1$, la constante es $\beta_0 + \beta_2$. Vemos, pues, que la variable ficticia desplaza la relación lineal entre Y y X_1 en el valor del coeficiente β_2 . De esta forma, podemos representar el efecto de los desplazamientos en nuestra ecuación de regresión.

4.5. Machine learning

Según (Christopher M. Bishop, 2006) es una rama de la inteligencia artificial (IA) que se encarga de generar algoritmos que tienen la capacidad de aprender. Esto sirve para que no estemos programando de forma explícita (es decir, que no necesitamos programar horas y horas a pensar en todos los escenarios posibles).

Simplemente alimentamos un algoritmo con grandes volúmenes de datos y que el algoritmo aprenda y sepa que hacer en cada una de estas excepciones

En el machine learning existen dos tipos de aprendizajes:

- **Algoritmos supervisados:** Es cuando entrenamos un algoritmo de machine learning dándole las preguntas (características) y respuestas (etiquetas). Es decir, son una serie de algoritmo donde los datos históricos tienen un valor de salida además de los datos de partida, el modelo lo que hace es hacer uso de las variables de salida desde los datos históricos aparte de los datos de entrada. **Ejemplo:** Regresión línea, regresión logística, arboles de decisiones y demás
- **Algoritmos no supervisados:** En este caso lo único que alimentamos de este algoritmo es de características. Nunca vamos a darle etiquetas. Es decir, no requiere de datos históricos para crear mi propio modelo. **Ejemplo:** Clustering.

Nota: Para este trabajo de grado, se realizara el aprendizaje por algoritmo supervisado

4.6. Relevancias significativas del data science

Según la revista HARVARD BUSINESS REVIEW, los datos son el nuevo petróleo del siglo XXI, además que los analistas de datos realizan aportes muy significativos a la empresa, para ellos hacen uso de las fuentes de datos mas grandes del mundo, estas son:

- El Facebook comparten 2'500.000 de contenidos ya sean fotos, videos, publicaciones, etc. (por minuto)
- En Youtube se sube 72 horas de video por (min)
- Se crean 2.5 exabyte de información (1 exabyte = 1018 bytes), se requiere 1 millón de DD de 1 TB para guardarlos

Grandes empresas multinacionales les compran datasets a estas fuentes de datos. Por ejemplo, cada vez que alguien compra en Amazon, recibe pago en paypal. Esto se debe a que Amazon se encarga de contratar analistas de datos, para crear modelos predictivos, donde la empresa ya sabe dónde o en que parte del mundo una persona va realizar su siguiente compra en dicha página.

Capítulo 5

Objetivos

5.1. Objetivo general

Pronosticar las ventas de los pasajes por día en la taquilla "Modulo Centenario" para el grupo empresarial COOMOTOR, basado en la aplicación del machine learning por medio de un modelo de regresión polinomial múltiple, que garantice una mejora en el proceso productivo de la empresa ubicada en la ciudad de Neiva.

5.2. Objetivos específicos

1. Identificar los factores que influyen en las ventas ya sea que aumenten o disminuyen para tomar las variables que aportarán en la creación del modelo matemático utilizando la base de datos del grupo empresarial COOMOTOR sobre las ventas en la taquillas Modulo Centenario perteneciente al terminal de Neiva.
2. Estructurar un modelo de regresión polinomial multivariado donde va predecir los ingresos en ventas de la taquilla modulo centenario determinando las variables predictoras o influyentes para su creación.
3. Evaluar el modelo de regresión polinomial multivariado mediante una simulación por PYTHON haciendo uso de paquetes como PANDAS para el manejo de la base de datos, SKLEARN para la creación del modelo y NUMPY para la relación que tienen las variables predictoras.

Capítulo 6

Metodología

6.1. Diseño de la investigación

El presente estudio tendrá un diseño experimental de tipo cuasiexperimental, pues es el que mejor se adapta a las necesidades del estudio.

El diseño experimental busca medir probabilísticamente la relación causal entre las variables influyentes del modelo matemático predictivo, y estar en posibilidad de confirmar o rechazar las hipótesis sometidas a prueba.

Por su parte, el diseño de tipo cuasiexperimental "son una derivación de los estudios experimentales, en los cuales la asignación de las variables no es aleatoria aunque el factor de exposición es manipulado por el investigador".

Cabe aclarar que este diseño no cumple con los requisitos de un "preexperimento" ni de experimento "puro", ya que no hay aleatorización de los sujetos a los grupos de tratamiento y control, además los grupos ya estaban formados antes de la investigación.

6.2. Enfoque de la investigación

El presente trabajo será diseñado bajo el planteamiento metodológico del enfoque cuantitativo, puesto que éste es el que mejor se adapta a las características y necesidades de la investigación.

El enfoque cuantitativo utiliza la recolección y el análisis de datos para contestar preguntas de investigación y probar hipótesis establecida previamente, y confía en la "medición numérica, el conteo y frecuentemente en el uso de la estadística para establecer con exactitud patrones de comportamientos en una población".

Del enfoque cuantitativo se tomará una base de datos sobre las ventas de la taquilla del modulo centenario ubicado en el terminal de transporte de Neiva. A pesar de que algunas de las variables declaradas son categóricas tales como el día, mes, entre otras, se va hacer uso de las variables "dummies", donde se convertirá dichas variables, en cuantitativas dándole valores binarios (0 y 1).

6.3. Población

La población se define como " un conjunto de todos los elementos que estamos estudiando, acerca de los cuales intentamos sacar conclusiones".

La población de estudio de esta trabajo de investigación estará conformado por todos los clientes que compran sus pasajes en la taquilla "Modulo Centenario" ubicado en el terminal de transporte de Neiva durante el año 2011 hasta mediados del año 2019.

6.4. Herramienta metodológica

Los datos de las ventas de pasajes en la taquilla Modulo Centenario se obtendrán por medio de la base de datos que utiliza la empresa, en este caso el software FICS, luego se tabularán en una hoja de calculo de EXCELL y serán procesados para la obtención del modelo matemático predictivo por medio del software PYTHON.

Capítulo 7

Análisis y discusión de resultados

7.1. Etapa 1: Enmarcar el problema

Antes de comenzar con la investigación, fue importante preguntar sobre el comportamiento de las ventas, esto con el fin de definir los parámetros a evaluar y poder crear un modelo que nos diga con la mayor certeza posible la cantidad de ingresos por día, es importante recalcar que los turnos del día van desde las 6 a.m. hasta la misma del otro día. Para comenzar a definir los parámetros, consultamos directamente con el área de comercio y resolvimos cada una de las siguientes preguntas

7.1.1. ¿Cuánto venden aproximadamente en temporada alta?

Para resolver esta pregunta, fue importante investigar sobre cuales son las temporadas altas. Luego, para saber cuanto eran los ingresos, nos facilitaron los datos del año 2018 de la taquilla "modulo centenario", los resultados fueron los siguientes:

- Temporada de fin de año (primera semana de diciembre hasta segunda semana de enero).

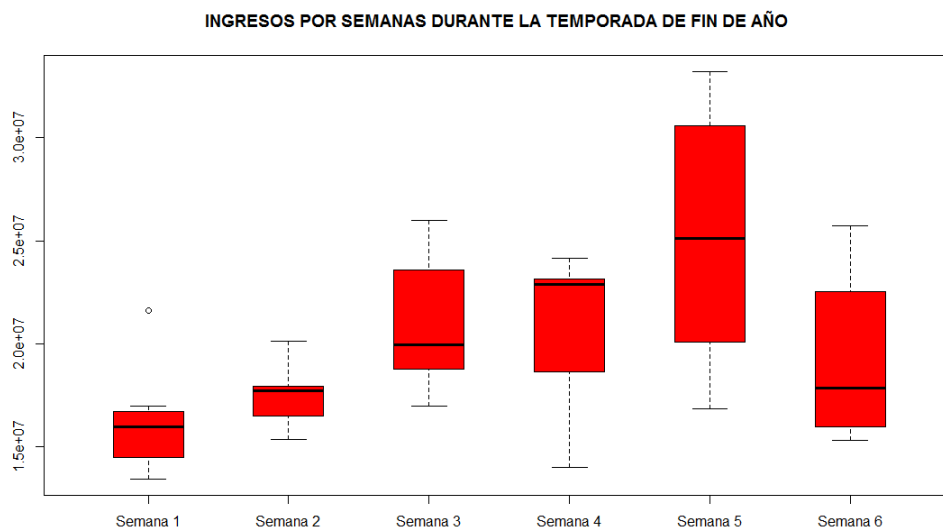


Figura 7.1: Boxplot sobre ingresos del año 2018 en temporada de fin de año

| Semana 1 | Semana 2 | Semana 3 | Semana 4 |
|-------------------|-------------------|-------------------|-------------------|
| Min. :13468100 | Min. :15372300 | Min. :17005500 | Min. :14035900 |
| 1st Qu. :14496200 | 1st Qu. :16495400 | 1st Qu. :18801200 | 1st Qu. :18663850 |
| Median :15993100 | Median :17745100 | Median :19940200 | Median :22869600 |
| Mean :16219571 | Mean :17453729 | Mean :21101843 | Mean :20674629 |
| 3rd Qu. :16729200 | 3rd Qu. :17969000 | 3rd Qu. :23587550 | 3rd Qu. :23156850 |
| Max. :21625000 | Max. :20129900 | Max. :25989700 | Max. :24175500 |

| Semana 5 | Semana 6 |
|-------------------|-------------------|
| Min. :16844000 | Min. :15314600 |
| 1st Qu. :20085400 | 1st Qu. :16000300 |
| Median :25111400 | Median :17866900 |
| Mean :25206971 | Mean :19424700 |
| 3rd Qu. :30563950 | 3rd Qu. :22540600 |
| Max. :33194700 | Max. :25709600 |

- Semana santa.

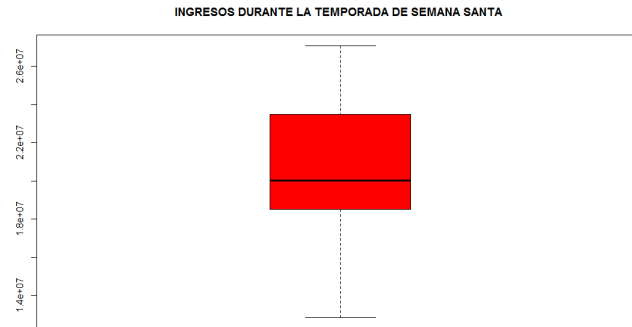


Figura 7.2: Boxplot sobre ingresos del año 2018 en semana santa

| Semana santa |
|-------------------|
| Min. :12859600 |
| 1st Qu. :18688550 |
| Median :20025500 |
| Mean :20507700 |
| 3rd Qu. :22801200 |
| Max. :27072200 |

- Temporada de San Pedro en el Huila (primera semana de junio hasta la primera semana de julio).

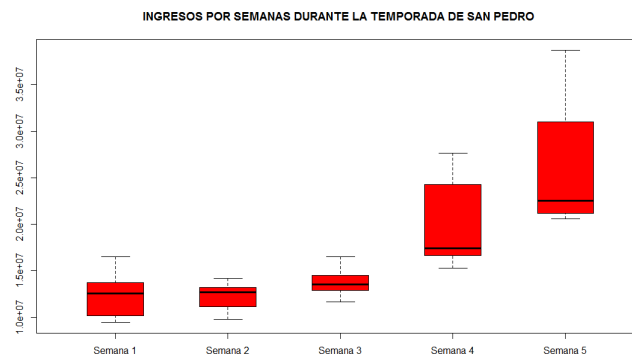


Figura 7.3: Boxplot sobre ingresos del año 2018 en temporada de San Pedro

| Semana 1 | Semana 2 | Semana 3 | Semana 4 | Semana 5 |
|------------------|------------------|------------------|------------------|------------------|
| Min. : 9489000 | Min. : 9813200 | Min. :11640200 | Min. :15302400 | Min. :20603600 |
| 1st Qu.:10143300 | 1st Qu.:11162000 | 1st Qu.:12871250 | 1st Qu.:16664750 | 1st Qu.:21200950 |
| Median :12564700 | Median :12721600 | Median :13543500 | Median :17449900 | Median :22528700 |
| Mean :12327757 | Mean :12216714 | Mean :13779414 | Mean :20334314 | Mean :26607400 |
| 3rd Qu.:13721700 | 3rd Qu.:13242700 | 3rd Qu.:14519550 | 3rd Qu.:24294300 | 3rd Qu.:31011750 |
| Max. :16510600 | Max. :14172800 | Max. :16490600 | Max. :27669800 | Max. :38694100 |

- Semana de receso escolar (segunda semana de octubre)



Figura 7.4: Boxplot sobre ingresos del año 2018 en la semana de receso escolar

7.1.2. ¿Cuánto venden aproximadamente en temporada baja?

Analizamos las ventas de dos meses (febrero y abril), ya que son meses donde habita poca presencia de pasajeros.

- Mes de febrero

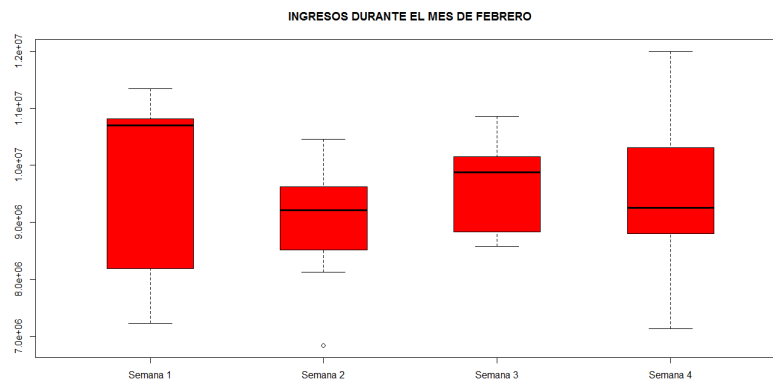


Figura 7.5: Boxplot sobre ingresos en febrero del 2018

| Semana 1 | Semana 2 | Semana 3 | Semana 4 |
|------------------|------------------|------------------|------------------|
| Min. : 7225800 | Min. : 6837600 | Min. : 8579500 | Min. : 7128900 |
| 1st Qu.: 8192250 | 1st Qu.: 8519050 | 1st Qu.: 8835700 | 1st Qu.: 8805500 |
| Median :10702800 | Median : 9209000 | Median : 9873900 | Median : 9256000 |
| Mean : 9613700 | Mean : 8968357 | Mean : 9610971 | Mean : 9516029 |
| 3rd Qu.:10818000 | 3rd Qu.: 9620300 | 3rd Qu.:10147100 | 3rd Qu.:10308250 |
| Max. :11346800 | Max. :10453200 | Max. :10857800 | Max. :11999800 |

- Mes de Abril

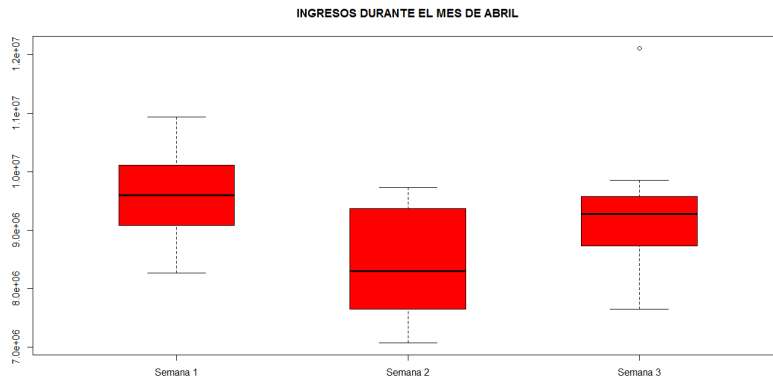


Figura 7.6: Boxplot sobre ingresos en abril del 2018

| Semana 1 | Semana 2 | Semana 3 |
|-------------------|------------------|------------------|
| Min. : 8273400 | Min. : 7074200 | Min. : 7652200 |
| 1st Qu.: 9078000 | 1st Qu.: 7651200 | 1st Qu.: 8735000 |
| Median : 9594200 | Median : 8298500 | Median : 9280000 |
| Mean : 9596186 | Mean : 8449529 | Mean : 9381386 |
| 3rd Qu.: 10106650 | 3rd Qu.: 9372900 | 3rd Qu.: 9578950 |
| Max. : 10936400 | Max. : 9725800 | Max. : 12109600 |

7.1.3. ¿Las ventas dependen del día en la semana?

Se toma los datos de los meses (agosto y septiembre) del 2018 para hacer los comparativos, los resultados fueron:

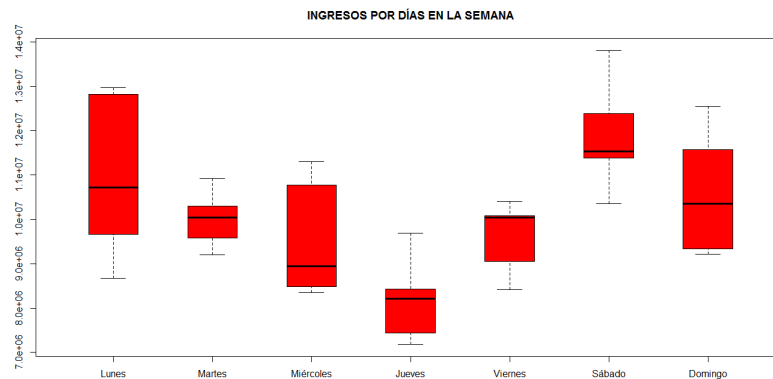


Figura 7.7: Boxplot sobre ingresos entre agosto y septiembre del 2018

| Lunes | Martes | Miércoles | Jueves |
|-------------------|-------------------|-------------------|------------------|
| Min. : 8669000 | Min. : 9200800 | Min. : 8353000 | Min. : 7182700 |
| 1st Qu.: 9662500 | 1st Qu.: 9581600 | 1st Qu.: 8477800 | 1st Qu.: 7443000 |
| Median : 10713100 | Median : 10036000 | Median : 8940800 | Median : 8209000 |
| Mean : 10968340 | Mean : 10006680 | Mean : 9570000 | Mean : 8190300 |
| 3rd Qu.: 12823700 | 3rd Qu.: 10294000 | 3rd Qu.: 10771600 | 3rd Qu.: 8432800 |
| Max. : 12973400 | Max. : 10921000 | Max. : 11306800 | Max. : 9684000 |

| Viernes | Sábado | Domingo |
|-------------------|-------------------|-------------------|
| Min. : 8421600 | Min. : 10350200 | Min. : 9213000 |
| 1st Qu.: 9047600 | 1st Qu.: 11378200 | 1st Qu.: 9341700 |
| Median : 10038800 | Median : 11528600 | Median : 10357400 |
| Mean : 9597280 | Mean : 11891700 | Mean : 10607100 |
| 3rd Qu.: 10077200 | 3rd Qu.: 12391200 | 3rd Qu.: 11571600 |
| Max. : 10401200 | Max. : 13810300 | Max. : 12551800 |

Luego, los días de la semana influyen para las ventas de taquilla, donde los días lunes, sábado y domingo presenta mayor ingreso.

7.1.4. ¿Las ventas de un fin de semana con día festivo son la misma sin un día festivo?

Tomamos un fin de semana con festivos del mes de agosto y otro sin festivos del mismo mes, los resultados fueron los siguientes:

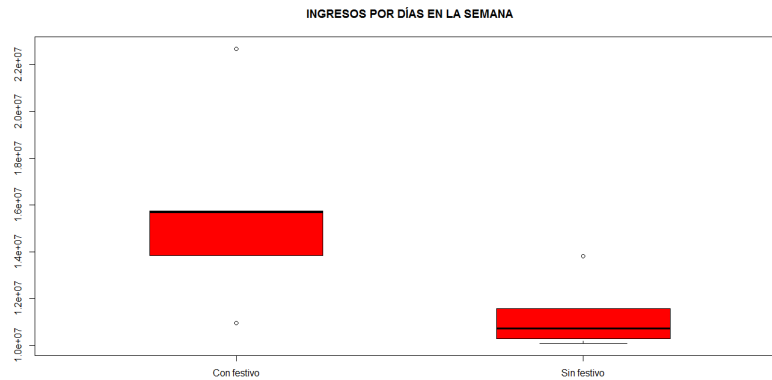


Figura 7.8: Boxplot sobre ingresos con días festivos en agosto del 2018

| Con festivo | Sin festivo |
|------------------|------------------|
| Min. :10952200 | Min. :10077200 |
| 1st Qu.:13850600 | 1st Qu.:10294000 |
| Median :15687200 | Median :10713100 |
| Mean :15785820 | Mean :11293240 |
| 3rd Qu.:15756000 | 3rd Qu.:11571600 |
| Max. :22683100 | Max. :13810300 |

7.1.5. Conclusión

Observando las gráficas podemos definir las variables influyentes que estarán presentes durante la investigación, estas variables son:

- Año
- Temporada de fin de año
- Temporada de San Pedro
- Semana santa
- Días antes de semana santa
- Semana de receso escolar
- Días antes de la semana de receso escolar
- Festivos
- Días antes de festivos
- Lunes
- Martes
- Miércoles
- Jueves
- Viernes
- Sábado
- Domingo

La mayoría de las variables son categóricas, es decir cualitativas, como su valor numérico no se pueden establecer, para ello la importancia de convertir dichas variables en **dummies**.

7.2. Etapa 2: Adquirir y preparar los datos

Obtener los datos que necesitamos es uno de los mayores inconvenientes que se tiene a la hora de practicar alguna técnica del data science. Muchas de las empresas son celosas, puesto que los datos, así como ayuda a conseguir modelos matemáticos para el análisis y toma de decisiones, también puede mostrar puntos débiles de la empresa.

En este caso, el reconocido empresario del departamento del Huila, el doctor ARMANDO CUELLAR ARTEAGA nos facilitó adquirir los datos sobre las ventas en la taquilla del terminal de transporte de Neiva, para ello fue necesario comunicarse con el área de sistemas.

La adquisición de los datos fue un proceso demorado, puesto que al área de sistemas no se encontraba disponible para recolectar datos de venta del año 2011 hasta 2019, para solucionar este problema, la ingeniera ROCIO nos facilitó una cuenta del software llamado **FICS**, la empresa lo utiliza para acceder a la base de datos, su proceso fue demasiado engorroso, ya que teníamos que exportar los datos de forma manual, es decir, abrimos la relación de ventas por día, se copiaba y luego se pegaba en una hoja de cálculo (EXCELL), este proceso se repetía por cada día del año hasta llegar al último día que registraba el programa en el año 2019. la recolección de 5 años duró un mes y medio, por fortuna, llegó el correo de la ingeniera con todos los datos que estábamos necesitando ya en formato de EXCELL (véase anexo 1).

A partir de la base de datos correspondiente al año 2017, venía unificado con todas las taquillas ubicadas en el terminal de transporte de Neiva, por lo tanto se separó la base de datos correspondiente a la taquilla "modulo centenario z en el mismo archivo de EXCELL se crearon hojas donde representa las ventas por meses (para ello fue importante crear una tabla dinámica), el dataset quedó como muestra las imágenes (véase anexo 1, fig. 9.7 y fig. 9.8).

Una vez que tengamos los datos separados por meses, se unifican todas las bases de datos separadas por años y luego se categorizan los datos, para ello se utilizan las tablas dinámicas de EXCELL, teniendo en cuenta las variables que se declararon en la etapa anterior, donde cada variable la identificamos de la siguiente manera:

- Año.
- Temporada de fin de año (**T.D**).
- Temporada de San Pedro (**T.SP**).
- Semana santa (**S.S**).
- Días antes de semana santa (**D.a.S.S**).
- Semana de receso escolar (**R.E**).
- Días antes de la semana de receso escolar. (**D.a.R.E**).
- Festivos.
- Días antes de festivos (**D.a.f**).
- Lunes.
- Martes.
- Miércoles.
- Jueves.
- Viernes.
- Sábado.
- Domingo.

En las siguientes imágenes se observa como quedaron las bases de datos después de su limpieza (véase anexo 2).

Finalmente, se realiza un código en PYTHON para terminar de crear las variables **dummies** o más conocidas como las variables categóricas o ficticias, con respecto a la columna **Dia** y **Mes**.

- Creación de la función para variables dummies

```
[ ] data.columns.values
↳ array(['Año', 'Ingresos', 'Dia', 'Festivo', 'D.a.f', 'T. D.', 'T.SP',
        'R.E', 'D.a.R.E', 'S.S.', 'D.a.S.S.'], dtype=object)

[ ] def dummies(x,y):
    df = pd.get_dummies(x[y])
    x = x.drop([y], axis = 1)
    x = pd.concat([x,df], axis = 1)
    return x
```

Figura 7.9: Código en PYTHON para variables dummies

- Creación de variables dummies por días

```
[ ] data = dummies(data, "Dia")
data.head()
```

| | Año | Ingresos | Festivo | D.a.f | T. D. | T.SP | R.E | D.a.R.E | S.S. | D.a.S.S. | Domingo | Jueves | Lunes | Martes | Miércoles | Sabado | Sábado | Viernes |
|---|------|----------|---------|-------|-------|------|-----|---------|------|----------|---------|--------|-------|--------|-----------|--------|--------|---------|
| 0 | 2011 | 4634000 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 1 | 2011 | 6642000 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 2011 | 8649000 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 3 | 2011 | 8437000 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 4 | 2011 | 9306000 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |

Figura 7.10: Variables dummies por días.

- Creación de variables dummies por meses

```
[ ] data = dummies(data, "Mes")
data.head()
```

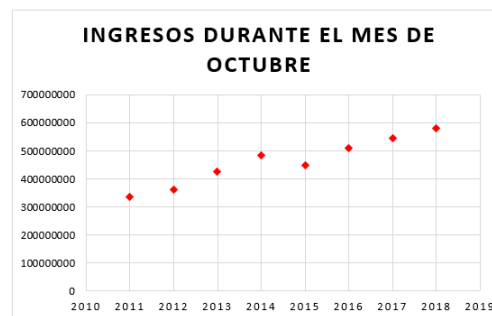
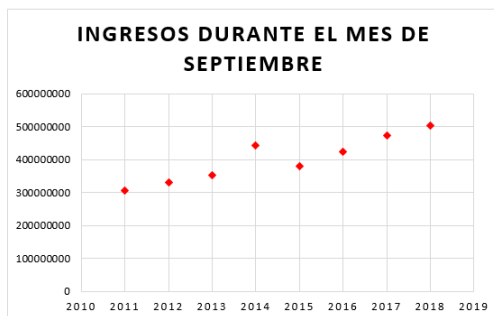
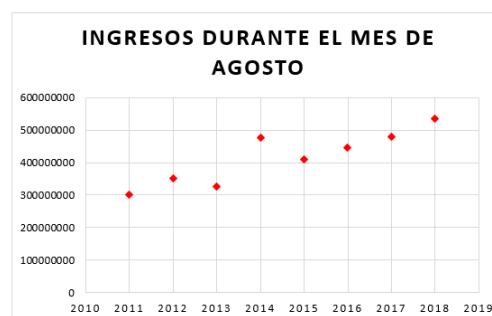
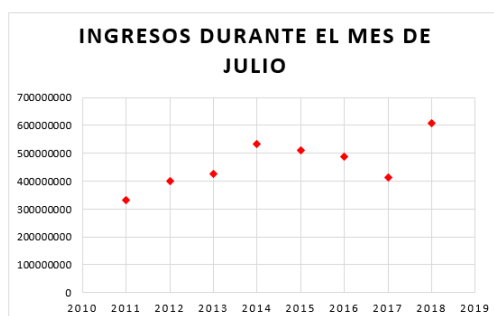
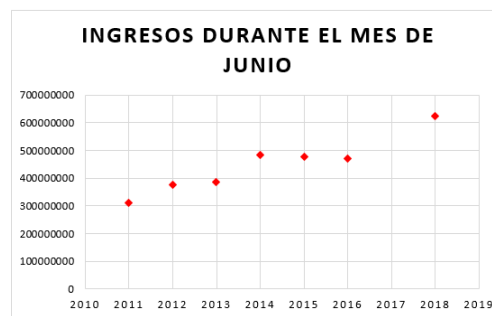
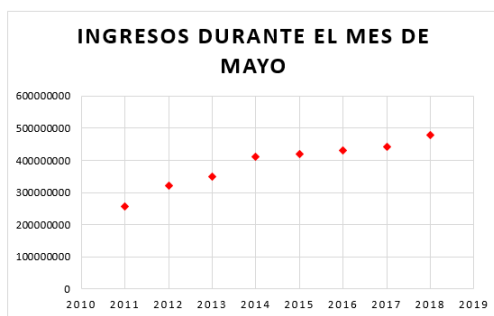
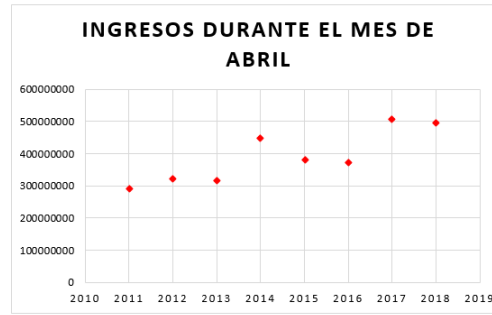
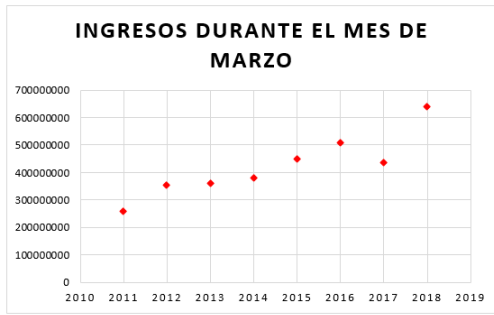
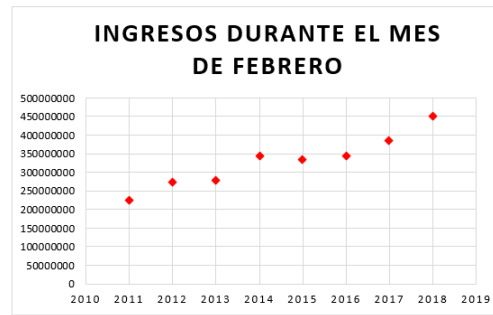
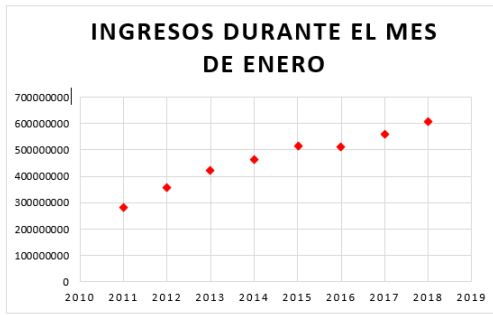
| | .S. | Domingo | Jueves | Lunes | Martes | Miércoles | Sábado | Viernes | Abril | Agosto | Diciembre | Enero | Febrero | Julio | Junio | Marzo | Mayo | Noviembre | Octubre | Septiembre |
|---|-----|---------|--------|-------|--------|-----------|--------|---------|-------|--------|-----------|-------|---------|-------|-------|-------|------|-----------|---------|------------|
| 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

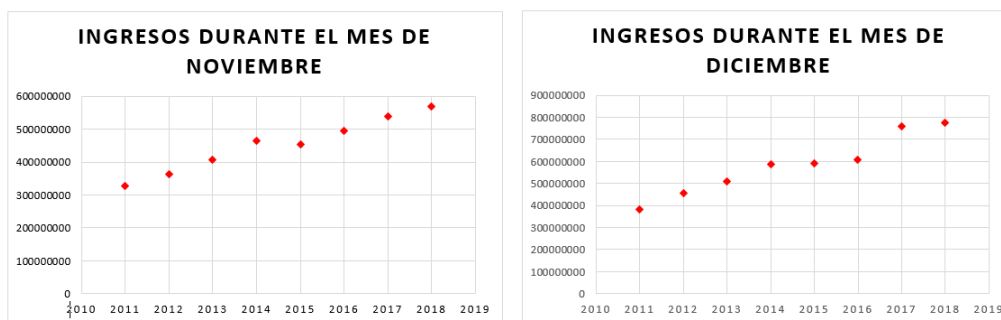
Figura 7.11: Variables dummies por meses.

La base de datos con las variables dummies quedaron de la siguiente manera (véase anexo 3)

7.3. Etapa 3: Explorar los datos

Es necesario observar de forma gráfica el comportamiento de los datos. Para este trabajo de investigación, su modelo estadístico es una regresión polinomial multivariado, en este caso nuestro modelo tiene 16 variables, luego su gráfico no es posible realizarlo ya que solo se puede visualizar graficas hasta 3 dimensiones. Sin embargo, se realizará gráficas de regresión de forma intuitivas, donde se relacionará los ingresos mensuales con respecto al año, para ello basta con observar dicha relación para la taquilla modulo centenario.





En las gráficas podemos observar que al trascurrir los años, los ingresos mensuales van aumentando de una forma lineal, este comportamiento lo confirmamos preguntando en el área de auditoría interna por el aumento de las tarifas de los pasajes al año, donde la respuesta fue que van aumentando entre los 500 y 1000 pesos, por lo que se considera que los ingresos incrementa.

7.4. Etapa 4: Modelizar y evaluar los datos

Una vez que tengamos los datasets preparados. se hará la creación del modelo predictivo, en este caso se realizará un modelo de regresión polinomial multivariado.

la simulación fue creada mediante el programa PYTHON, donde su ejecución se realizó por medio de GOOGLE COLABORATORY, utilizando la GPU virtual que ofrece GOOGLE desde Drive. Para llegar al modelo, se realizaron los siguientes procedimientos:

1. Se instalaron los paquetes que se requieren para la creación del modelo:

```
import pandas as pd
import numpy as np
import io
from google.colab import files
from sklearn.linear_model import LinearRegression
```

Cada paquete permite realizar lo siguiente:

- **pandas:** Nos permitirá manipular los datasets y realizar operaciones entre ellos.
 - **numpy:** Se utiliza para realizar cálculos estadísticos, tales como media aritmética, moda, mediana, entre otros.
 - **io y google.colab:** Nos permitirá abrir cualquier archivo, en este caso un archivo .csv (valores separados por comas), lo cual es un formato que vamos a utilizar para manipulación de los datasets.
 - **sklearn:** Se utiliza para la creación del modelo predictivo.
2. Se abre el archivo en .csv

```
d = files.upload()
Elegir archivos | enero.csv
• enero.csv(application/vnd.ms-excel) - 11245 bytes, last modified: 18/7/2019 - 100% done
Saving enero.csv to enero.csv
```

Figura 7.12: Subiendo base de datos a PYTHON.

3. Se crea el datasets al que deseamos crear el modelo, para ello se define la variable "data".

```
data = pd.read_csv(io.BytesIO(d["enero.csv"]), sep = ";", encoding = "ISO-8859-1")
data.head()
```

| | Año | Ingresos | Festivo | D.a.f | T. D. | T.SP | R.E | D.a.R.E | S.S. | D.a.S.S. | Domingo | Jueves | Lunes | Martes | Miércoles | Sábado | Viernes |
|---|------|----------|---------|-------|-------|------|-----|---------|------|----------|---------|--------|-------|--------|-----------|--------|---------|
| 0 | 2011 | 4634000 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 1 | 2011 | 6642000 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 2011 | 8649000 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 3 | 2011 | 8437000 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 4 | 2011 | 9306000 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |

Figura 7.13: Base de datos en PYTHON para el modelo predictivo

4. Nos aseguramos que todos los datos estén en formato " entero".

```
data.dtypes
```

| | |
|-----------|--------|
| Año | int64 |
| Ingresos | int64 |
| Festivo | int64 |
| D.a.f | int64 |
| T. D. | int64 |
| T.SP | int64 |
| R.E | int64 |
| D.a.R.E | int64 |
| S.S. | int64 |
| D.a.S.S. | int64 |
| Domingo | int64 |
| Jueves | int64 |
| Lunes | int64 |
| Martes | int64 |
| Miércoles | int64 |
| Sábado | int64 |
| Viernes | int64 |
| dtype: | object |

Figura 7.14: Variables de la base de datos.

5. Se define las variable a predecir y las variables predictivas, donde sabemos que se va predecir los ingresos en la taquilla para cualquier año y las variables influyentes (predictivas) son las 16 que ya hemos mencionado en las etapas anteriores.

```
[5] data.columns.values
```

```
array(['Año', 'Ingresos', 'Festivo', 'D.a.f', 'T. D.', 'T.SP', 'R.E', 'D.a.R.E', 'S.S.', 'D.a.S.S.', 'Domingo', 'Jueves', 'Lunes', 'Martes', 'Miércoles', 'Sábado', 'Viernes'], dtype=object)
```

```
[6] feature_cols = ['Año', 'Festivo', 'D.a.f', 'T. D.', 'T.SP', 'R.E', 'D.a.R.E', 'S.S.', 'D.a.S.S.', 'Domingo', 'Jueves', 'Lunes', 'Martes', 'Miércoles', 'Sábado', 'Viernes']
```

```
[7] X = data[feature_cols]
     Y = data["Ingresos"]
```

Figura 7.15: Declaración de variables predictoras y predecibles

6. A continuación se creará y a la vez se evaluará 2 tipos de regresión (lineal y polinomial):

- primero se creará el modelo de la regresión lineal:

```
[8] lm = LinearRegression()
     lm.fit(X,Y)
```

```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)
```

Figura 7.16: Código de la regresión lineal en PYTHON.

Luego se observan los resultados del modelo, para ello se observa la nueva columna llamada "prediction".

```
[10] data["prediction"] = lm.predict(pd.DataFrame(data[feature_cols]))
data.head()
```

| | Año | Ingresos | Festivo | D.a.f | T. D. | T.SP | R.E | D.a.R.E | S.S. | D.a.S.S. | Domingo | Jueves | Lunes | Martes | Miércoles | Sábado | Viernes | prediction |
|---|------|----------|---------|-------|-------|------|-----|---------|------|----------|---------|--------|-------|--------|-----------|--------|---------|--------------|
| 0 | 2011 | 4634000 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 6.580766e+06 |
| 1 | 2011 | 6642000 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1.009414e+07 |
| 2 | 2011 | 8649000 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1.300833e+07 |
| 3 | 2011 | 8437000 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1.320159e+07 |
| 4 | 2011 | 9306000 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1.261612e+07 |

Figura 7.17: Resultados de la regresión lineal.

Continuamos con la evaluación del modelo, para ello observamos el R^2 ya ajustado

```
[9] lm.score(X,Y)
0.7221139023388516
```

Figura 7.18: R^2 ajustado

observamos que el R^2 es muy bajo, luego su coeficiente de correlación no es recomendable para realizar un modelo de regresión lineal, no obstante, probaremos el porcentaje de error, para ello hacemos la suma de las diferencias al cuadrado y el error estándar residual.

```
[19] SSD = np.sum((data["Ingresos"] - data["prediction"])**2)
[20] RSE = np.sqrt(SSD/(len(data)-len(feature_cols)-1))
RSE
2348835.132866323
```

Figura 7.19: Error estandarizado.

Finalmente se calcula el porcentaje de error

```
[21] mean_ingresos = np.mean(data["Ingresos"])
[22] error = RSE/mean_ingresos *100
error
15.701369367169654
```

Figura 7.20: Porcentaje del error.

- por ultimo se creará el modelo de la regresión polinomial (grado 4):


```
[26] from sklearn.preprocessing import PolynomialFeatures
      from sklearn import linear_model

[27] poly = PolynomialFeatures(degree = 4)

[28] X_poly = poly.fit_transform(X)

[29] pm = linear_model.LinearRegression()
      pm.fit(X_poly,Y)

↳ LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)
```

Figura 7.21: Código para una regresión polinomial.

Luego se observan los resultados del modelo que están ubicados en la nueva columna llamada "prediction".

```
[31] data["prediction"] = pm.predict(X_poly)
      data.head()
```

| | Año | Ingresos | Festivo | D.a.f | T. D. | T.SP | R.E | D.a.R.E | S.S. | D.a.S.S. | Domingo | Jueves | Lunes | Martes | Miércoles | Sábado | Viernes | prediction |
|---|------|----------|---------|-------|-------|------|-----|---------|------|----------|---------|--------|-------|--------|-----------|--------|---------|--------------|
| 0 | 2011 | 4634000 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 4.638841e+06 |
| 1 | 2011 | 6642000 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 8.923816e+06 |
| 2 | 2011 | 8649000 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 8.389963e+06 |
| 3 | 2011 | 8437000 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 8.708137e+06 |
| 4 | 2011 | 9306000 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1.018349e+07 |

Figura 7.22: Resultado de una regresión polinomial.

Continuamos con la evaluación del modelo, para ello observamos el R^2 ya ajustado

```
[30] pm.score(X_poly,Y)

↳ 0.8655353099390275
```

Figura 7.23: R^2 ajustado

observamos que el R^2 es muy alto, luego su coeficiente de correlación es el adecuado para realizar un modelo de regresión lineal, no obstante, probaremos el porcentaje de error, para ello hacemos la suma de las diferencias al cuadrado y el error estándar residual.

```
[32] data["prediction"] = np.array(data["prediction"]).astype('int')

[33] SSD = np.sum((data["Ingresos"]-data["prediction"])**2)

[34] RSE = np.sqrt(SSD/(len(data) - len(feature_cols) -1))
      RSE

↳ 1633891.6266550485
```

Figura 7.24: Error estandarizado.

Finalmente se calcula el porcentaje de error

```

mean_ingreso = np.mean(data["Ingresos"])

[36] error = RSE/mean_ingreso *100
      error
10.922152677753145

```

Figura 7.25: Porcentaje del error.

Evaluando los dos modelos de regresión, podemos observar que los datos se ajustan mejor a la regresión polinomial multivariado de grado 4. Lastimosamente no se pudo evaluar polinomios de grado mayores o iguales a 5 ya que la maquina virtual de GOOGLE se reiniciaba por sus innumerables calculos que le tocaba hacer.

7.5. Etapa 5: Comunicar los resultados

Ya teniendo los datasets preparados por meses con sus respectivos parámetros, se aplica el modelo matemático de regresión polinomial de grado 4. a continuación, se observa los resultados sobre los ingresos que se esperan para el año 2020.

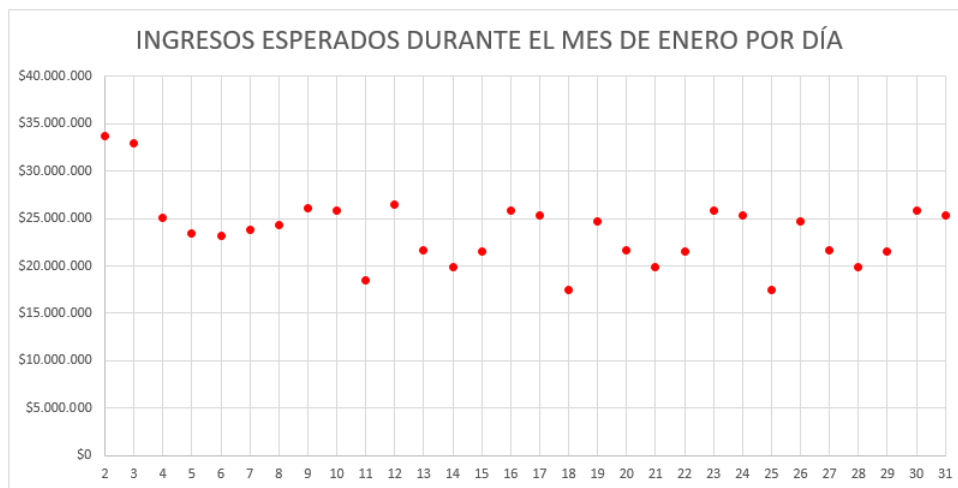


Figura 7.26: Predicción de ingresos en enero

Durante los primeros días del mes de enero, se puede observar que las ventas tienen un comportamiento dinámico debido a la temporada de fin de año, a partir del día 15 de enero comienza a tener una curva de tal forma que se repite cada 7 días (días de la semana)

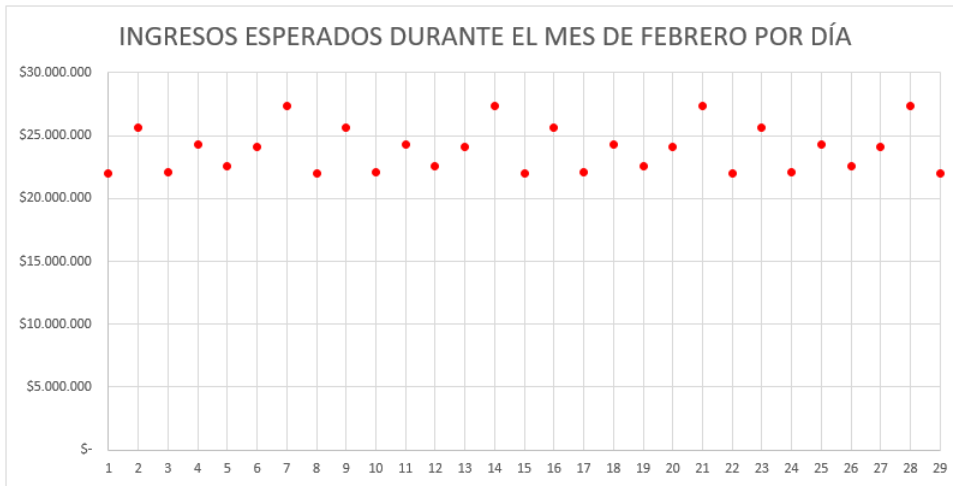


Figura 7.27: Predicción de ingresos en febrero

Durante el mes de febrero, se puede observar que los datos repiten un mismo patrón cada 7 días (días de la semana), lo que significa que las ventas en febrero no se ven afectadas por alguna temporada (variable) definidas en esta investigación.

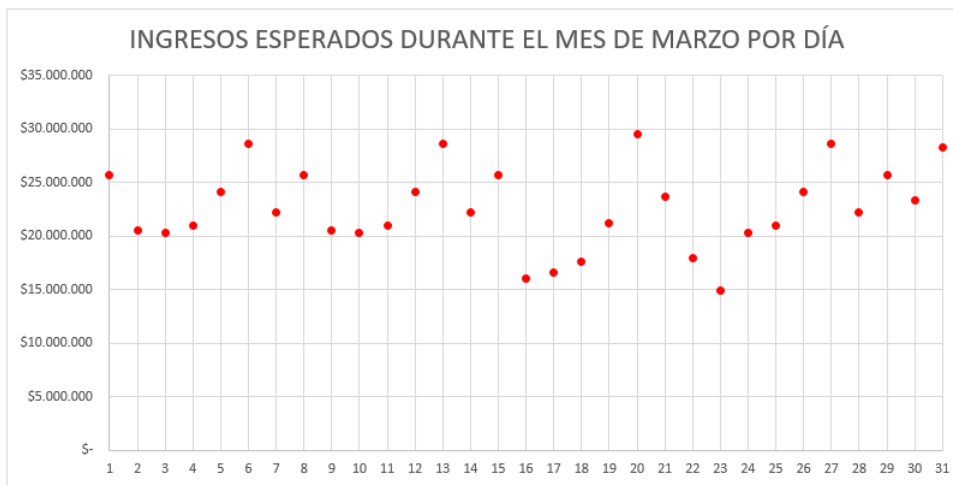


Figura 7.28: Predicción de ingresos en marzo

Durante el mes de marzo, se puede observar que los datos repiten un mismo patrón cada 7 días (días de la semana) hasta llegar al día 16, a partir de este día las ventas comienzan a cambiar ya que se aproxima un lunes festivo que le corresponde al día 23 y también se acerca la temporada de semana santa,

Durante el mes de abril, se puede observar que los primeros días las ventas tienen un comportamiento dinámico debido a la temporada de semana santa, adicional a esto el modelo no pudo predecir el fin de semana de esta temporada ya que los datos históricos varían entre el mes de marzo y abril.

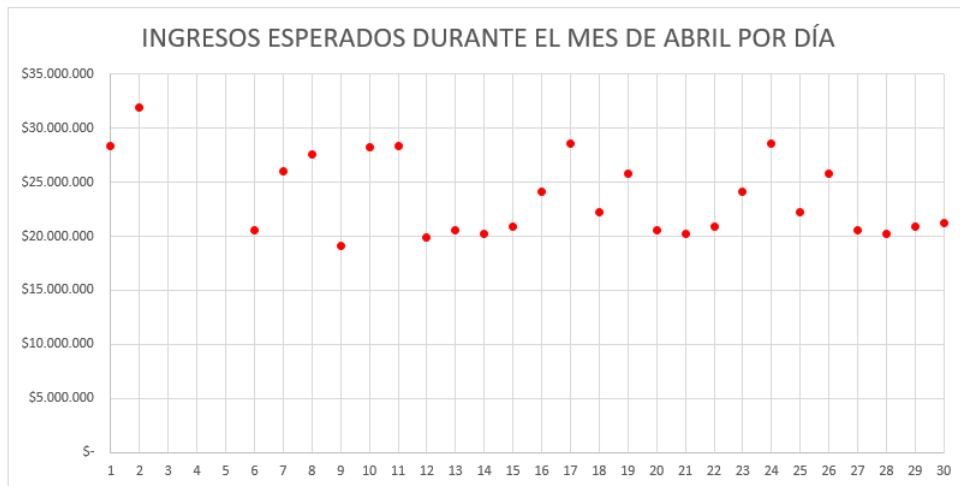


Figura 7.29: Predicción de ingresos en abril

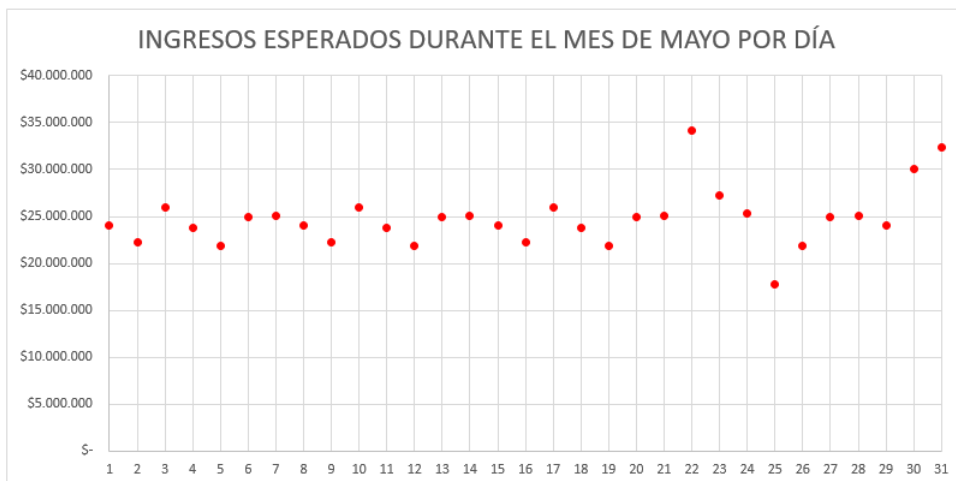


Figura 7.30: Predicción de ingresos en mayo

Durante el mes de mayo, se puede observar que los datos repiten un mismo patrón cada 7 días (días de la semana) hasta llegar al día 22, a partir de este día las ventas comienzan a cambiar ya que se aproxima la temporada de "san pedro en el huila".

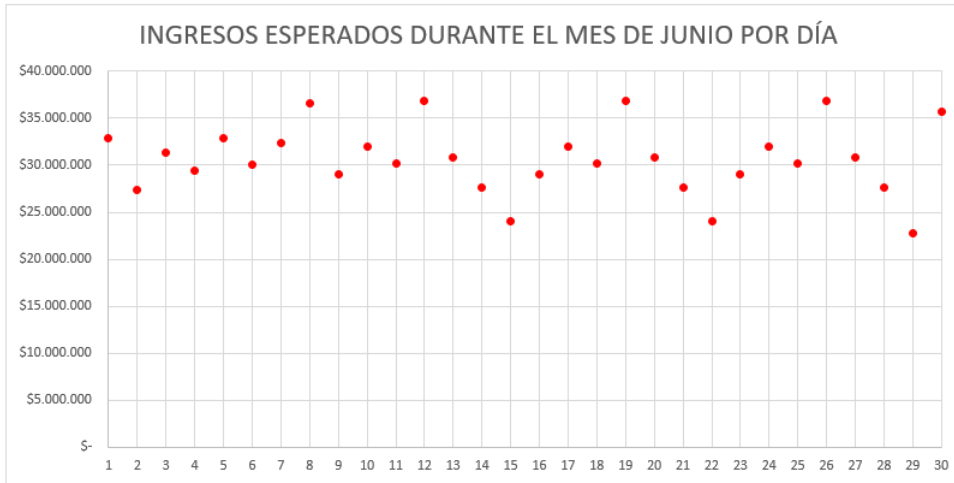


Figura 7.31: Predicción de ingresos en junio

Durante el mes de junio, se puede observar que los datos tienen un comportamiento muy dinámico debido a que en esta temporada, a parte de que las ventas aumentan considerablemente, también se encuentra con varios lunes festivos, lo que hace que las ventas tengan este comportamiento.

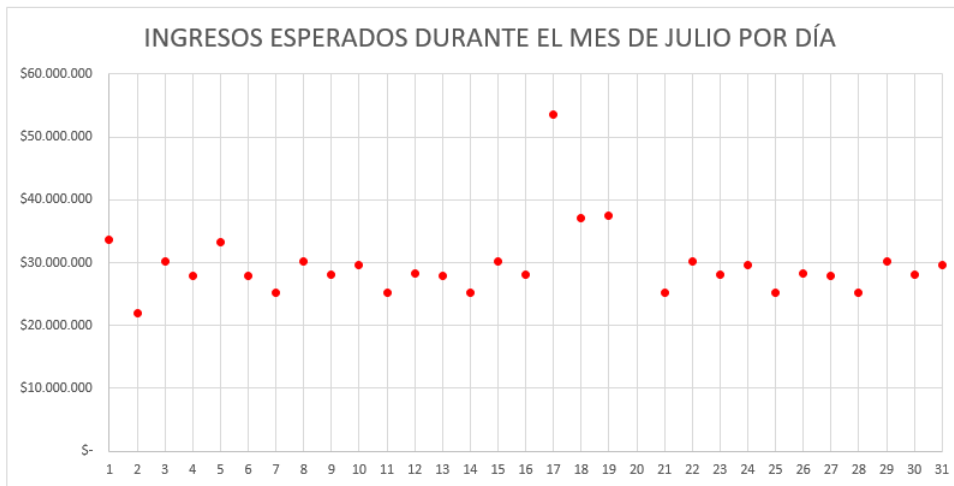


Figura 7.32: Predicción de ingresos en julio

Durante los primeros días del mes de julio, se puede observar que las ventas tienen un comportamiento dinámico debido a la temporada de "San Pedro". Adicional a esto, el modelo no pudo predecir las ventas del día 20 de julio ya que es un día patrio en Colombia, por lo tanto todos los años cambia de día y esto hace que los datos históricos no puedan aportar al modelo.

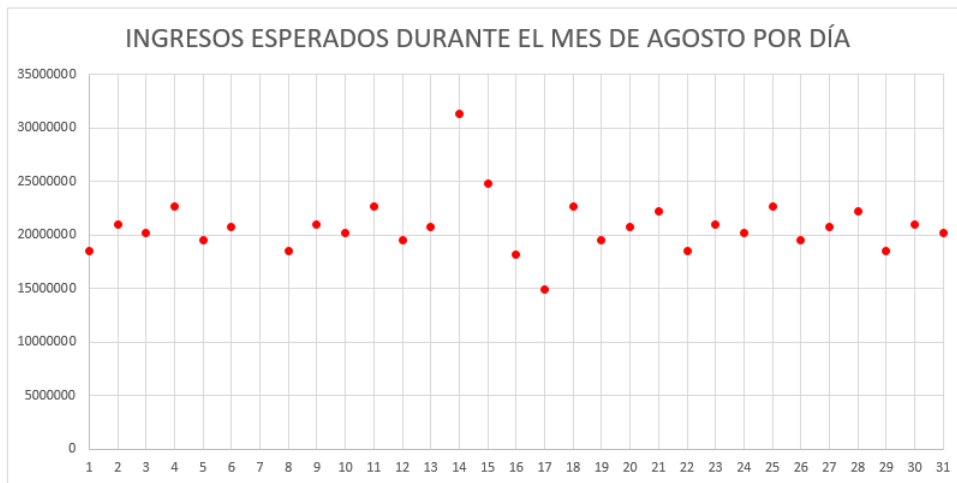


Figura 7.33: Predicción de ingresos en agosto

Durante el mes de agosto, se puede observar que los datos repiten un mismo patrón cada 7 días (días de la semana) hasta llegar al día 13, a partir de este día las ventas comienzan a cambiar ya que se aproxima un lunes festivo que le corresponde al día 17. Adicional a esto, el modelo no pudo predecir las ventas del día 7 de agosto ya que es un día patrio en Colombia, por lo tanto todos los años cambia de día y esto hace que los datos históricos no puedan aportar al modelo.

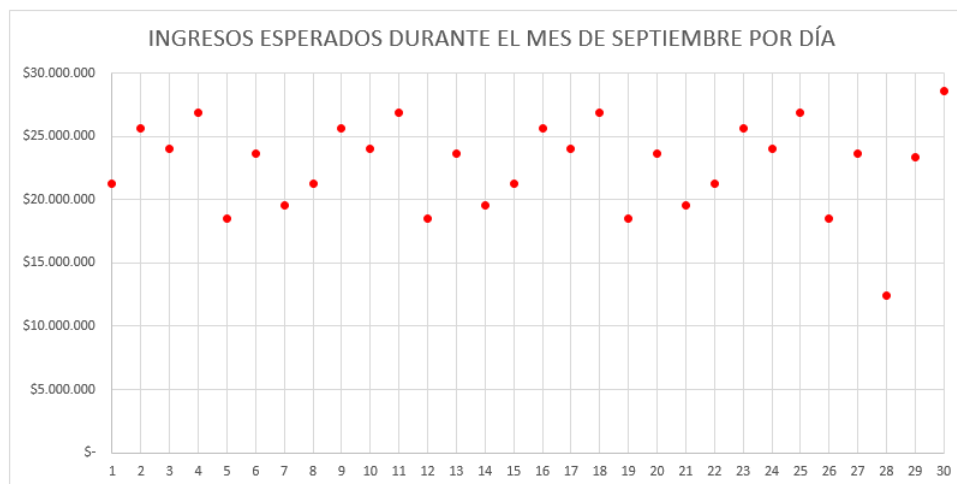


Figura 7.34: Predicción de ingresos en septiembre

Durante el mes de septiembre, se puede observar que los datos repiten un mismo patrón cada 7 días (días de la semana) hasta llegar al día 16, a partir de este día las ventas comienzan a cambiar ya que se acerca la temporada de "semana de receso escolar".

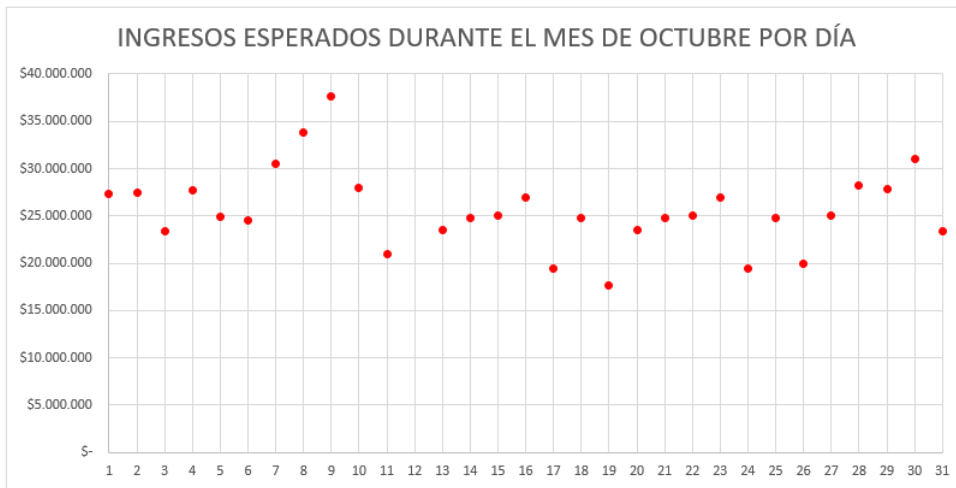


Figura 7.35: Predicción de ingresos en octubre

Durante el mes de octubre, se puede observar que los primeros días las ventas tienen un comportamiento dinámico debido a la temporada de "semana de receso escolar", luego comienza a repetirse un mismo patrón cada 7 días (días de la semana), al finalizar el mes, los datos vuelven a tener un comportamiento dinámico ya que se aproxima un lunes festivo.

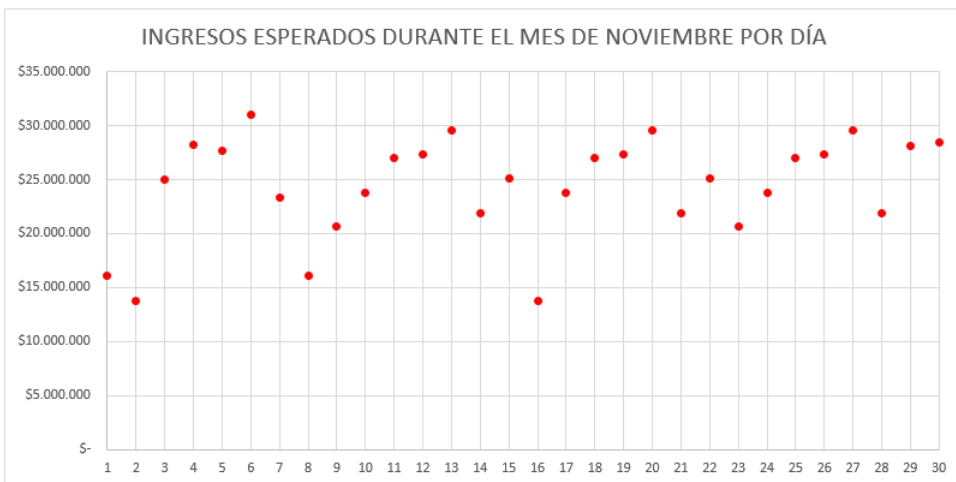


Figura 7.36: Predicción de ingresos en noviembre

Durante el mes de Noviembre, se puede observar que los datos repiten un mismo patrón cada 7 días (días de la semana) hasta llegar al día 21, a partir de este día las ventas comienzan a tener un comportamiento dinámico ya que se acerca la temporada de "fin de año".

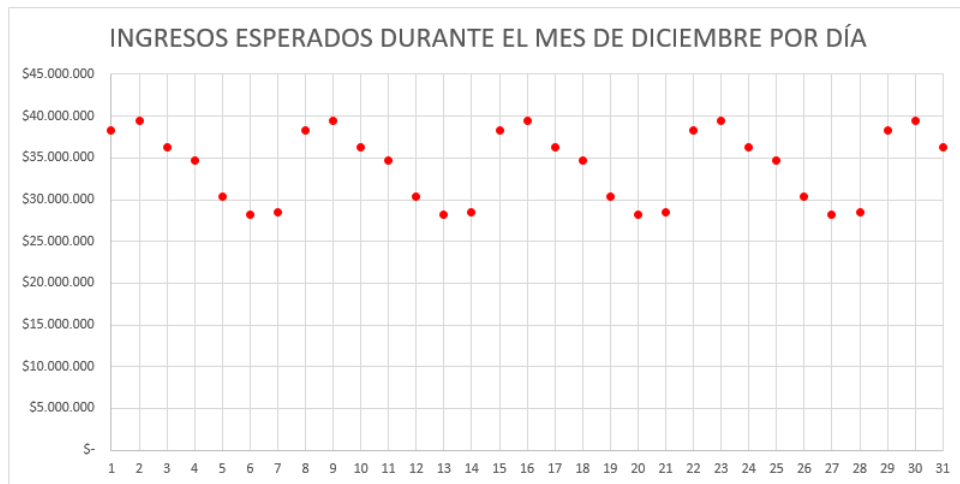


Figura 7.37: Predicción de ingresos en diciembre

Durante el mes de diciembre, se puede observar que los datos repiten un mismo patrón cada 7 días (días de la semana) ya que las ventas en diciembre mantienen constantemente altas.

Capítulo 8

Conclusiones

En este trabajo de grado se aprende a obtener una ecuación de regresión polinomial multivariado que describe el comportamiento entre varias variables, lo que permite pronosticar valores a futuro con cierto grado de certeza, lo cual constituye una herramienta muy útil para las clases relacionadas con estadística o matemática computacional en el programa de matemática aplicada de la universidad surcolombiana, ya que permite al estudiante manipular datos reales y aplicados en una de las mejores empresas que tiene nuestra región y así fomentar la investigación en la ciencia de los datos.

Para el desarrollo de este trabajo de grado se ha propuesto un modelo estadístico usando técnicas del data science para poder crear unas metas de ventas en las taquillas del grupo empresarial COOMOTOR del terminal de transporte de Neiva, en particular se tomó la taquilla conocida como "Modulo Centenario", donde se ha permitido predecir de cuanto será el ingreso esperado por día durante el año 2020, como se ha podido observar, estos ingresos presentan un incremento con respecto a los de años anteriores.

Los resultados han mostrado que cuando se trata de predecir días festivos fijos como lo son el 8 y 25 de diciembre, 1 de enero y días patrios como es el caso del 20 de julio (día de la independencia de Colombia) y 7 de agosto (día de la batalla de Boyacá), el modelo muestra algunas fallas debido a que el incremento de las ventas varían bastante dependiendo del día de la semana que le corresponda, es decir, si un 20 de julio le corresponde un lunes, las ventas no van a ser igual cuando corresponden a un miércoles o jueves. El modelo al tratar de supervisar los datos históricos, notará que sus ventas son muy dispersas, por lo tanto se tiene complicaciones al tratar de predecir los ingresos futuros, no obstante, para los demás días, el modelo predice los ingresos que se tendrán con un mínimo error que se encuentra entre el 10 y 20 por ciento por mes

Con estos resultados, la administración podrá establecer metas de venta para que la empresa pueda superarse así misma año tras año, ya que este modelo también se puede aplicar a futuro, donde se tiene que actualizarse los datos y así obtener predicciones para el año que lo requiera.

Para trabajos a futuro, se espera que esta investigación pueda ser ampliada utilizando conjuntos de datos actualizados ya sea de esta empresa u otra relacionada con el transporte público de personas, esto con el fin de involucrar la investigación en los estudiantes o incluso profesionales con respecto a la ciencia de los datos y así obtener mejores resultados.

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|----|---|--------------|---------------------|-------------------|-----------------------|---------------------|-----------------------|---------------|----------------------|--------------------|----------------|---------------|--------------|---|
| 2 | | | | | | | | | | | | | | |
| 3 | | Caja | Usuario | Comp/Vouch | Fecha/Hora Ope | Medio de pag | Tipo de boleto | Pasaje | Origen-Destit | Fecha Viaje | Importe | % | Comit | |
| 4 | | 13284 | HEYDY PATRICIA PAEZ | | 1.01.2014 6:06 | Efectivo | EFFECTIVO -NAV- | MCE-1256510 | NEIV - BOGO | 1.01.2014 7:20 | 45.000,00 | 0 | 0 | |
| 5 | | 13284 | HEYDY PATRICIA PAEZ | | 1.01.2014 7:10 | Efectivo | EFFECTIVO -NAV- | MCE-1256511 | NEIV - BOGO | 2.01.2014 0:30 | 45.000,00 | 0 | 0 | |
| 6 | | 13284 | HEYDY PATRICIA PAEZ | | 1.01.2014 8:24 | Efectivo | EFFECTIVO -NAV- | MCE-1256524 | NEIV - BOGO | 1.01.2014 8:30 | 45.000,00 | 0 | 0 | |
| 7 | | 13284 | HEYDY PATRICIA PAEZ | 16433 | 1.01.2014 8:47 | Tarjeta Debito | TARJETA DEBIT | MCE-1256525 | NEIV - BOGO | 1.01.2014 18:02 | 45.000,00 | 0 | 0 | |
| 8 | | 13284 | HEYDY PATRICIA PAEZ | 16433 | 1.01.2014 8:48 | Tarjeta Debito | TARJETA DEBIT | MCE-1256526 | NEIV - BOGO | 6.01.2014 12:00 | 45.000,00 | 0 | 0 | |
| 9 | | 13284 | HEYDY PATRICIA PAEZ | | 1.01.2014 8:53 | Efectivo | EFFECTIVO -NAV- | MCE-1256527 | NEIV - BOGO | 2.01.2014 7:20 | 45.000,00 | 0 | 0 | |
| 10 | | 13284 | HEYDY PATRICIA PAEZ | | 1.01.2014 8:58 | Efectivo | EFFECTIVO -CFV- | MCE-1256528 | MEDE - NEIV | 1.01.2014 15:30 | 72.000,00 | 0 | 0 | |
| 11 | | 13284 | HEYDY PATRICIA PAEZ | 16434 | 1.01.2014 9:14 | Tarjeta Debito | TARJETA DEBIT | MCE-1256529 | NEIV - IBAG | 1.01.2014 19:01 | 30.000,00 | 0 | 0 | |
| 12 | | 13284 | HEYDY PATRICIA PAEZ | 16434 | 1.01.2014 9:14 | Tarjeta Debito | TARJETA DEBIT | MCE-1256530 | NEIV - IBAG | 1.01.2014 19:01 | 30.000,00 | 0 | 0 | |
| 13 | | 13284 | HEYDY PATRICIA PAEZ | 16434 | 1.01.2014 9:14 | Tarjeta Debito | TARJETA DEBIT | MCE-1256531 | NEIV - IBAG | 1.01.2014 19:01 | 30.000,00 | 0 | 0 | |
| 14 | | 13284 | HEYDY PATRICIA PAEZ | 16435 | 1.01.2014 9:17 | Tarjeta Debito | TARJETA DEBIT | MCE-1256532 | NEIV - BOGO | 1.01.2014 8:30 | 45.000,00 | 0 | 0 | |
| 15 | | 13284 | HEYDY PATRICIA PAEZ | 16435 | 1.01.2014 9:17 | Tarjeta Debito | TARJETA DEBIT | MCE-1256533 | NEIV - BOGO | 1.01.2014 8:30 | 45.000,00 | 0 | 0 | |
| 16 | | 13284 | HEYDY PATRICIA PAEZ | | 1.01.2014 9:41 | Efectivo | EFFECTIVO -ZIR- | MCE-1256536 | NEIV - PERE | 1.01.2014 14:31 | 44.000,00 | 0 | 0 | |
| 17 | | 13284 | HEYDY PATRICIA PAEZ | | 1.01.2014 9:41 | Efectivo | EFFECTIVO -ZIR- | MCE-1256537 | NEIV - PERE | 1.01.2014 14:31 | 44.000,00 | 0 | 0 | |
| 18 | | 13284 | HEYDY PATRICIA PAEZ | | 1.01.2014 9:41 | Efectivo | EFFECTIVO -ZIR- | MCE-1256538 | NEIV - PERE | 1.01.2014 14:31 | 44.000,00 | 0 | 0 | |
| 19 | | 13284 | HEYDY PATRICIA PAEZ | | 1.01.2014 9:51 | Efectivo | EFFECTIVO -NAV- | MCE-1256539 | NEIV - BOGO | 1.01.2014 23:15 | 45.000,00 | 0 | 0 | |
| 20 | | 13284 | HEYDY PATRICIA PAEZ | | 1.01.2014 9:57 | Efectivo | EFFECTIVO -NAV- | MCE-1256540 | NEIV - BOGO | 1.01.2014 23:15 | 45.000,00 | 0 | 0 | |
| 21 | | 13284 | HEYDY PATRICIA PAEZ | | 1.01.2014 9:59 | Efectivo | EFFECTIVO -NAV- | MCE-1256541 | NEIV - BOGO | 1.01.2014 10:10 | 45.000,00 | 0 | 0 | |
| 22 | | 13284 | HEYDY PATRICIA PAEZ | | 1.01.2014 10:16 | Efectivo | EFFECTIVO -NAV- | MCE-1256542 | NEIV - BOGO | 3.01.2014 4:00 | 45.000,00 | 0 | 0 | |
| 23 | | 13284 | HEYDY PATRICIA PAEZ | | 1.01.2014 10:30 | Efectivo | EFFECTIVO -HOR- | MCE-1256545 | NEIV - BOGO | 1.01.2014 8:30 | 45.000,00 | 0 | 0 | |
| 24 | | 13284 | HEYDY PATRICIA PAEZ | | 1.01.2014 10:30 | Efectivo | EFFECTIVO -HOR- | MCE-1256546 | NEIV - BOGO | 1.01.2014 8:30 | 45.000,00 | 0 | 0 | |
| | | Enero | Febrero | Marzo | Abril | Mayo | Junio | Julio | Agosto | Septiembre | Octubre | Noviembre ... | | |

Figura 9.4: Base de datos sin filtrar del año 2014

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|----|---|--------------|------------------------------|-------------------|-----------------------|---------------------|-----------------------|---------------|----------------------|--------------------|----------------|----------|--------------|---|
| 1 | | | | | | | | | | | | | | |
| 2 | | | | | | | | | | | | | | |
| 3 | | Caja | Usuario | Comp/Vouch | Fecha/Hora Ope | Medio de pag | Tipo de boleto | Pasaje | Origen-Destit | Fecha Viaje | Importe | % | Comit | |
| 4 | | 16552 | FRANCY ELENA ALVAREZ MORALES | | 1.01.2015 6:28 | Efectivo | EFFECTIVO -NAV- | MCE-184152 | NEIV - BOGO | 1.01.2015 7:20 | 46.000,00 | 0 | 0 | |
| 5 | | 16552 | FRANCY ELENA ALVAREZ MORALES | | 1.01.2015 6:28 | Efectivo | EFFECTIVO -NAV- | MCE-184153 | NEIV - BOGO | 1.01.2015 7:20 | 46.000,00 | 0 | 0 | |
| 6 | | 16552 | FRANCY ELENA ALVAREZ MORALES | | 1.01.2015 6:28 | Efectivo | EFFECTIVO -NAV- | MCE-184154 | NEIV - BOGO | 1.01.2015 7:20 | 46.000,00 | 0 | 0 | |
| 7 | | 16552 | FRANCY ELENA ALVAREZ MORALES | | 1.01.2015 6:57 | Efectivo | EFFECTIVO -NAV- | MCE-184155 | NEIV - BOGO | 1.01.2015 7:20 | 46.000,00 | 0 | 0 | |
| 8 | | 16552 | FRANCY ELENA ALVAREZ MORALES | | 1.01.2015 6:57 | Efectivo | EFFECTIVO -NAV- | MCE-184156 | NEIV - BOGO | 1.01.2015 7:20 | 46.000,00 | 0 | 0 | |
| 9 | | 16552 | FRANCY ELENA ALVAREZ MORALES | | 1.01.2015 7:17 | Efectivo | EFFECTIVO -NAV- | MCE-1361738 | NEIV - BOGO | 1.01.2015 7:20 | 46.000,00 | 0 | 0 | |
| 10 | | 16552 | FRANCY ELENA ALVAREZ MORALES | | 1.01.2015 7:21 | Efectivo | EFFECTIVO -CFV- | MCE-1361739 | BOGO - NEIV | 4.01.2015 19:00 | 46.000,00 | 0 | 0 | |
| 11 | | 16552 | FRANCY ELENA ALVAREZ MORALES | | 1.01.2015 7:21 | Efectivo | EFFECTIVO -CFV- | MCE-1361740 | BOGO - NEIV | 4.01.2015 19:00 | 46.000,00 | 0 | 0 | |
| 12 | | 16552 | FRANCY ELENA ALVAREZ MORALES | | 1.01.2015 7:48 | Efectivo | EFFECTIVO -NAV- | MCE-1361741 | NEIV - BOGO | 1.01.2015 8:30 | 46.000,00 | 0 | 0 | |
| 13 | | 16552 | FRANCY ELENA ALVAREZ MORALES | | 1.01.2015 8:03 | Efectivo | EFFECTIVO -ZIR- | MCE-1361742 | NEIV - CALLI | 3.01.2015 19:25 | 57.000,00 | 0 | 0 | |
| 14 | | 16552 | FRANCY ELENA ALVAREZ MORALES | | 1.01.2015 8:05 | Efectivo | EFFECTIVO -NAV- | MCE-1361743 | NEIV - BOGO | 1.01.2015 8:30 | 46.000,00 | 0 | 0 | |
| 15 | | 16552 | FRANCY ELENA ALVAREZ MORALES | | 1.01.2015 8:13 | Efectivo | EFFECTIVO -NAV- | MCE-1361744 | NEIV - BOGO | 1.01.2015 8:30 | 46.000,00 | 0 | 0 | |
| 16 | | 16552 | FRANCY ELENA ALVAREZ MORALES | | 1.01.2015 8:13 | Efectivo | EFFECTIVO -NAV- | MCE-1361745 | NEIV - BOGO | 1.01.2015 8:30 | 46.000,00 | 0 | 0 | |
| 17 | | 16552 | FRANCY ELENA ALVAREZ MORALES | | 1.01.2015 8:13 | Efectivo | EFFECTIVO -NAV- | MCE-1361746 | NEIV - BOGO | 1.01.2015 8:30 | 46.000,00 | 0 | 0 | |
| 18 | | 16552 | FRANCY ELENA ALVAREZ MORALES | | 1.01.2015 8:20 | Efectivo | EFFECTIVO -NAV- | MCE-1361749 | NEIV - BOGO | 1.01.2015 8:30 | 46.000,00 | 0 | 0 | |
| 19 | | 16552 | FRANCY ELENA ALVAREZ MORALES | | 1.01.2015 8:21 | Efectivo | EFFECTIVO -NAV- | MCE-1361750 | NEIV - BOGO | 1.01.2015 8:30 | 46.000,00 | 0 | 0 | |
| 20 | | 16552 | FRANCY ELENA ALVAREZ MORALES | | 1.01.2015 8:23 | Efectivo | EFFECTIVO -NAV- | MCE-1361752 | NEIV - BOGO | 1.01.2015 10:10 | 46.000,00 | 0 | 0 | |
| 21 | | 16552 | FRANCY ELENA ALVAREZ MORALES | | 1.01.2015 8:23 | Efectivo | EFFECTIVO -NAV- | MCE-1361753 | NEIV - BOGO | 1.01.2015 10:10 | 46.000,00 | 0 | 0 | |
| 22 | | 16552 | FRANCY ELENA ALVAREZ MORALES | | 1.01.2015 8:39 | Efectivo | CLIENTE PREFEJ | MCE-1361754 | NEIV - BOGO | 1.01.2015 11:15 | 46.000,00 | 0 | 0 | |
| 23 | | 16552 | FRANCY ELENA ALVAREZ MORALES | | 1.01.2015 9:39 | Efectivo | EFFECTIVO -NAV- | MCE-1361756 | NEIV - BOGO | 1.01.2015 10:10 | 46.000,00 | 0 | 0 | |
| | | Enero | Febrero | Marzo | Abril | Mayo | Junio | Julio | Agosto | Septiembre | Octubre ... | | | |

Figura 9.5: Base de datos sin filtrar del año 2015

9.2. Anexo 2: Base de datos filtrada

Las siguientes imágenes tienen como objetivo presentar los datos después de un procedimiento de limpieza.

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|----|---|-------|---------------|-----------|---------|-------|-------|-------|--------|------------|---------|------|----------|---|
| 1 | | | | | | | | | | | | | | |
| 2 | | Año | Ingresos | Día | Festivo | D.a.f | Mes | T.D. | T.SP | R.E | D.a.R.E | S.S. | D.a.S.S. | |
| 3 | | 2011 | \$ 4.634.000 | Sabado | 1 | 0 | Enero | 1 | 0 | 0 | 0 | 0 | 0 | |
| 4 | | 2011 | \$ 6.642.000 | Domingo | 0 | 0 | Enero | 1 | 0 | 0 | 0 | 0 | 0 | |
| 5 | | 2011 | \$ 8.649.000 | Lunes | 0 | 1 | Enero | 1 | 0 | 0 | 0 | 0 | 0 | |
| 6 | | 2011 | \$ 8.437.000 | Martes | 0 | 1 | Enero | 1 | 0 | 0 | 0 | 0 | 0 | |
| 7 | | 2011 | \$ 9.306.000 | Miércoles | 0 | 1 | Enero | 1 | 0 | 0 | 0 | 0 | 0 | |
| 8 | | 2011 | \$ 10.119.000 | Jueves | 0 | 1 | Enero | 1 | 0 | 0 | 0 | 0 | 0 | |
| 9 | | 2011 | \$ 12.409.000 | Viernes | 0 | 1 | Enero | 1 | 0 | 0 | 0 | 0 | 0 | |
| 10 | | 2011 | \$ 11.542.000 | Sábado | 0 | 1 | Enero | 1 | 0 | 0 | 0 | 0 | 0 | |
| 11 | | 2011 | \$ 7.712.000 | Domingo | 0 | 1 | Enero | 1 | 0 | 0 | 0 | 0 | 0 | |
| 12 | | 2011 | \$ 7.630.000 | Lunes | 1 | 0 | Enero | 1 | 0 | 0 | 0 | 0 | 0 | |
| 13 | | 2011 | \$ 9.264.000 | Martes | 0 | 0 | Enero | 1 | 0 | 0 | 0 | 0 | 0 | |
| 14 | | 2011 | \$ 8.048.000 | Miércoles | 0 | 0 | Enero | 1 | 0 | 0 | 0 | 0 | 0 | |
| 15 | | 2011 | \$ 8.393.000 | Jueves | 0 | 0 | Enero | 1 | 0 | 0 | 0 | 0 | 0 | |
| 16 | | 2011 | \$ 10.720.000 | Viernes | 0 | 0 | Enero | 1 | 0 | 0 | 0 | 0 | 0 | |
| 17 | | 2011 | \$ 9.325.000 | Sábado | 0 | 0 | Enero | 1 | 0 | 0 | 0 | 0 | 0 | |
| 18 | | 2011 | \$ 9.790.000 | Domingo | 0 | 0 | Enero | 1 | 0 | 0 | 0 | 0 | 0 | |
| 19 | | 2011 | \$ 9.423.000 | Lunes | 0 | 0 | Enero | 1 | 0 | 0 | 0 | 0 | 0 | |
| 20 | | 2011 | \$ 8.996.000 | Martes | 0 | 0 | Enero | 1 | 0 | 0 | 0 | 0 | 0 | |
| 21 | | 2011 | \$ 9.031.000 | Miércoles | 0 | 0 | Enero | 1 | 0 | 0 | 0 | 0 | 0 | |
| 22 | | 2011 | \$ 11.459.000 | Jueves | 0 | 0 | Enero | 1 | 0 | 0 | 0 | 0 | 0 | |
| 23 | | 2011 | \$ 10.663.000 | Viernes | 0 | 0 | Enero | 1 | 0 | 0 | 0 | 0 | 0 | |
| | | Enero | Febrero | Marzo | Abril | Mayo | Junio | Julio | Agosto | Septiembre | Octubre | | | |

Figura 9.9: Base de datos filtrado del mes de enero desde el año 2011 hasta el año 2019

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|----|---|-------|---------------|-----------|---------|-------|---------|-------|--------|------------|---------|------|----------|---|
| 1 | | | | | | | | | | | | | | |
| 2 | | Año | Ingresos | Día | Festivo | D.a.f | Mes | T.D. | T.SP | R.E | D.a.R.E | S.S. | D.a.S.S. | |
| 3 | | 2011 | \$ 8.451.000 | Martes | 0 | 0 | Febrero | 0 | 0 | 0 | 0 | 0 | 0 | |
| 4 | | 2011 | \$ 7.330.000 | Miércoles | 0 | 0 | Febrero | 0 | 0 | 0 | 0 | 0 | 0 | |
| 5 | | 2011 | \$ 9.008.000 | Jueves | 0 | 0 | Febrero | 0 | 0 | 0 | 0 | 0 | 0 | |
| 6 | | 2011 | \$ 9.421.000 | Viernes | 0 | 0 | Febrero | 0 | 0 | 0 | 0 | 0 | 0 | |
| 7 | | 2011 | \$ 6.767.000 | Sábado | 0 | 0 | Febrero | 0 | 0 | 0 | 0 | 0 | 0 | |
| 8 | | 2011 | \$ 9.007.000 | Domingo | 0 | 0 | Febrero | 0 | 0 | 0 | 0 | 0 | 0 | |
| 9 | | 2011 | \$ 8.042.000 | Lunes | 0 | 0 | Febrero | 0 | 0 | 0 | 0 | 0 | 0 | |
| 10 | | 2011 | \$ 7.357.000 | Martes | 0 | 0 | Febrero | 0 | 0 | 0 | 0 | 0 | 0 | |
| 11 | | 2011 | \$ 8.079.000 | Miércoles | 0 | 0 | Febrero | 0 | 0 | 0 | 0 | 0 | 0 | |
| 12 | | 2011 | \$ 8.150.000 | Jueves | 0 | 0 | Febrero | 0 | 0 | 0 | 0 | 0 | 0 | |
| 13 | | 2011 | \$ 11.419.000 | Viernes | 0 | 0 | Febrero | 0 | 0 | 0 | 0 | 0 | 0 | |
| 14 | | 2011 | \$ 6.762.000 | Sábado | 0 | 0 | Febrero | 0 | 0 | 0 | 0 | 0 | 0 | |
| 15 | | 2011 | \$ 9.443.000 | Domingo | 0 | 0 | Febrero | 0 | 0 | 0 | 0 | 0 | 0 | |
| 16 | | 2011 | \$ 6.619.000 | Lunes | 0 | 0 | Febrero | 0 | 0 | 0 | 0 | 0 | 0 | |
| 17 | | 2011 | \$ 7.003.000 | Martes | 0 | 0 | Febrero | 0 | 0 | 0 | 0 | 0 | 0 | |
| 18 | | 2011 | \$ 6.295.000 | Miércoles | 0 | 0 | Febrero | 0 | 0 | 0 | 0 | 0 | 0 | |
| 19 | | 2011 | \$ 6.519.000 | Jueves | 0 | 0 | Febrero | 0 | 0 | 0 | 0 | 0 | 0 | |
| 20 | | 2011 | \$ 9.664.000 | Viernes | 0 | 0 | Febrero | 0 | 0 | 0 | 0 | 0 | 0 | |
| 21 | | 2011 | \$ 5.877.000 | Sábado | 0 | 0 | Febrero | 0 | 0 | 0 | 0 | 0 | 0 | |
| 22 | | 2011 | \$ 10.056.000 | Domingo | 0 | 0 | Febrero | 0 | 0 | 0 | 0 | 0 | 0 | |
| 23 | | 2011 | \$ 8.710.000 | Lunes | 0 | 0 | Febrero | 0 | 0 | 0 | 0 | 0 | 0 | |
| | | Enero | Febrero | Marzo | Abril | Mayo | Junio | Julio | Agosto | Septiembre | Octubre | | | |

Figura 9.10: Base de datos filtrado del mes de febrero desde el año 2011 hasta el año 2019

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|----|---|------------|-----------------|--------------|----------------|--------------|------------|--------------|-------------|------------|----------------|-------------|-----------------|---|
| 1 | | | | | | | | | | | | | | |
| 2 | | Año | Ingresos | Día | Festivo | D.a.f | Mes | T. D. | T.SP | R.E | D.a.R.E | S.S. | D.a.S.S. | |
| 3 | | 2011 | \$ 6.970.000 | Martes | 0 | 0 | Marzo | 0 | 0 | 0 | 0 | 0 | 0 | |
| 4 | | 2011 | \$ 6.674.000 | Miércoles | 0 | 0 | Marzo | 0 | 0 | 0 | 0 | 0 | 0 | |
| 5 | | 2011 | \$ 10.047.000 | Jueves | 0 | 0 | Marzo | 0 | 0 | 0 | 0 | 0 | 0 | |
| 6 | | 2011 | \$ 10.307.000 | Viernes | 0 | 0 | Marzo | 0 | 0 | 0 | 0 | 0 | 0 | |
| 7 | | 2011 | \$ 7.068.000 | Sábado | 0 | 0 | Marzo | 0 | 0 | 0 | 0 | 0 | 0 | |
| 8 | | 2011 | \$ 10.547.000 | Domingo | 0 | 0 | Marzo | 0 | 0 | 0 | 0 | 0 | 0 | |
| 9 | | 2011 | \$ 7.056.000 | Lunes | 0 | 0 | Marzo | 0 | 0 | 0 | 0 | 0 | 0 | |
| 10 | | 2011 | \$ 7.784.000 | Martes | 0 | 0 | Marzo | 0 | 0 | 0 | 0 | 0 | 0 | |
| 11 | | 2011 | \$ 6.821.000 | Miércoles | 0 | 0 | Marzo | 0 | 0 | 0 | 0 | 0 | 0 | |
| 12 | | 2011 | \$ 7.997.000 | Jueves | 0 | 0 | Marzo | 0 | 0 | 0 | 0 | 0 | 0 | |
| 13 | | 2011 | \$ 9.470.000 | Viernes | 0 | 0 | Marzo | 0 | 0 | 0 | 0 | 0 | 0 | |
| 14 | | 2011 | \$ 5.256.000 | Sábado | 0 | 0 | Marzo | 0 | 0 | 0 | 0 | 0 | 0 | |
| 15 | | 2011 | \$ 8.534.000 | Domingo | 0 | 0 | Marzo | 0 | 0 | 0 | 0 | 0 | 0 | |
| 16 | | 2011 | \$ 6.812.000 | Lunes | 0 | 1 | Marzo | 0 | 0 | 0 | 0 | 0 | 0 | |
| 17 | | 2011 | \$ 6.879.000 | Martes | 0 | 1 | Marzo | 0 | 0 | 0 | 0 | 0 | 0 | |
| 18 | | 2011 | \$ 8.696.000 | Miércoles | 0 | 1 | Marzo | 0 | 0 | 0 | 0 | 0 | 0 | |
| 19 | | 2011 | \$ 10.980.000 | Jueves | 0 | 1 | Marzo | 0 | 0 | 0 | 0 | 0 | 0 | |
| 20 | | 2011 | \$ 14.114.000 | Viernes | 0 | 1 | Marzo | 0 | 0 | 0 | 0 | 0 | 0 | |
| 21 | | 2011 | \$ 11.227.000 | Sábado | 0 | 1 | Marzo | 0 | 0 | 0 | 0 | 0 | 0 | |
| 22 | | 2011 | \$ 9.014.000 | Domingo | 0 | 1 | Marzo | 0 | 0 | 0 | 0 | 0 | 0 | |
| 23 | | 2011 | \$ 6.552.000 | Lunes | 1 | 0 | Marzo | 0 | 0 | 0 | 0 | 0 | 0 | |
| | | | | Marzo | | | | | | | | | | |
| | | Enero | Febrero | Marzo | Abril | Mayo | Junio | Julio | Agosto | Septiembre | Octubre | | | |

Figura 9.11: Base de datos filtrado del mes de marzo desde el año 2011 hasta el año 2019

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|----|---|------------|-----------------|--------------|----------------|--------------|------------|--------------|-------------|------------|----------------|-------------|-----------------|---|
| 1 | | | | | | | | | | | | | | |
| 2 | | Año | Ingresos | Día | Festivo | D.a.f | Mes | T. D. | T.SP | R.E | D.a.R.E | S.S. | D.a.S.S. | |
| 3 | | 2011 | \$ 9.379.000 | Viernes | 0 | 0 | Abril | 0 | 0 | 0 | 0 | 0 | 0 | |
| 4 | | 2011 | \$ 6.815.000 | Sábado | 0 | 0 | Abril | 0 | 0 | 0 | 0 | 0 | 0 | |
| 5 | | 2011 | \$ 9.917.000 | Domingo | 0 | 0 | Abril | 0 | 0 | 0 | 0 | 0 | 0 | |
| 6 | | 2011 | \$ 8.696.000 | Lunes | 0 | 0 | Abril | 0 | 0 | 0 | 0 | 0 | 0 | |
| 7 | | 2011 | \$ 7.188.000 | Martes | 0 | 0 | Abril | 0 | 0 | 0 | 0 | 0 | 0 | |
| 8 | | 2011 | \$ 8.738.000 | Miércoles | 0 | 0 | Abril | 0 | 0 | 0 | 0 | 0 | 0 | |
| 9 | | 2011 | \$ 8.142.000 | Jueves | 0 | 0 | Abril | 0 | 0 | 0 | 0 | 0 | 0 | |
| 10 | | 2011 | \$ 10.264.000 | Viernes | 0 | 0 | Abril | 0 | 0 | 0 | 0 | 0 | 0 | |
| 11 | | 2011 | \$ 6.053.000 | Sábado | 0 | 0 | Abril | 0 | 0 | 0 | 0 | 0 | 0 | |
| 12 | | 2011 | \$ 10.534.000 | Domingo | 0 | 0 | Abril | 0 | 0 | 0 | 0 | 0 | 0 | |
| 13 | | 2011 | \$ 8.436.000 | Lunes | 0 | 0 | Abril | 0 | 0 | 0 | 0 | 0 | 1 | |
| 14 | | 2011 | \$ 10.187.000 | Martes | 0 | 0 | Abril | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 15 | | 2011 | \$ 13.340.000 | Miércoles | 0 | 0 | Abril | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 16 | | 2011 | \$ 12.082.000 | Jueves | 0 | 0 | Abril | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 17 | | 2011 | \$ 18.086.000 | Viernes | 0 | 0 | Abril | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 18 | | 2011 | \$ 14.246.000 | Sábado | 0 | 0 | Abril | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 19 | | 2011 | \$ 12.330.000 | Domingo | 0 | 0 | Abril | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 20 | | 2011 | \$ 16.271.000 | Lunes | 0 | 0 | Abril | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 21 | | 2011 | \$ 14.018.000 | Martes | 0 | 0 | Abril | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 22 | | 2011 | \$ 12.901.000 | Miércoles | 0 | 0 | Abril | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 23 | | 2011 | \$ 8.231.000 | Jueves | 0 | 0 | Abril | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| | | | | Abril | | | | | | | | | | |
| | | Enero | Febrero | Marzo | Abril | Mayo | Junio | Julio | Agosto | Septiembre | Octubre | | | |

Figura 9.12: Base de datos filtrado del mes de abril desde el año 2011 hasta el año 2019

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O |
|----|---|------------|-----------------|-------------|----------------|--------------|------------|--------------|-------------|------------|----------------|-------------|-----------------|---|---|
| 1 | | | | | | | | | | | | | | | |
| 2 | | Año | Ingresos | Día | Festivo | D.a.f | Mes | T. D. | T.SP | R.E | D.a.R.E | S.S. | D.a.S.S. | | |
| 3 | | 2011 | \$ 8.577.000 | Domingo | 0 | 0 | Mayo | 0 | 0 | 0 | 0 | 0 | 0 | | |
| 4 | | 2011 | \$ 7.241.000 | Lunes | 0 | 0 | Mayo | 0 | 0 | 0 | 0 | 0 | 0 | | |
| 5 | | 2011 | \$ 7.430.000 | Martes | 0 | 0 | Mayo | 0 | 0 | 0 | 0 | 0 | 0 | | |
| 6 | | 2011 | \$ 7.188.000 | Miércoles | 0 | 0 | Mayo | 0 | 0 | 0 | 0 | 0 | 0 | | |
| 7 | | 2011 | \$ 9.913.000 | Jueves | 0 | 0 | Mayo | 0 | 0 | 0 | 0 | 0 | 0 | | |
| 8 | | 2011 | \$ 12.250.000 | Viernes | 0 | 0 | Mayo | 0 | 0 | 0 | 0 | 0 | 0 | | |
| 9 | | 2011 | \$ 7.601.000 | Sábado | 0 | 0 | Mayo | 0 | 0 | 0 | 0 | 0 | 0 | | |
| 10 | | 2011 | \$ 10.412.000 | Domingo | 0 | 0 | Mayo | 0 | 0 | 0 | 0 | 0 | 0 | | |
| 11 | | 2011 | \$ 7.932.000 | Lunes | 0 | 0 | Mayo | 0 | 0 | 0 | 0 | 0 | 0 | | |
| 12 | | 2011 | \$ 7.448.000 | Martes | 0 | 0 | Mayo | 0 | 0 | 0 | 0 | 0 | 0 | | |
| 13 | | 2011 | \$ 7.466.000 | Miércoles | 0 | 0 | Mayo | 0 | 0 | 0 | 0 | 0 | 0 | | |
| 14 | | 2011 | \$ 10.509.000 | Jueves | 0 | 0 | Mayo | 0 | 0 | 0 | 0 | 0 | 0 | | |
| 15 | | 2011 | \$ 11.362.000 | Viernes | 0 | 0 | Mayo | 0 | 0 | 0 | 0 | 0 | 0 | | |
| 16 | | 2011 | \$ 6.277.000 | Sábado | 0 | 0 | Mayo | 0 | 0 | 0 | 0 | 0 | 0 | | |
| 17 | | 2011 | \$ 9.225.000 | Domingo | 0 | 0 | Mayo | 0 | 0 | 0 | 0 | 0 | 0 | | |
| 18 | | 2011 | \$ 7.803.000 | Lunes | 0 | 0 | Mayo | 0 | 0 | 0 | 0 | 0 | 0 | | |
| 19 | | 2011 | \$ 6.366.000 | Martes | 0 | 0 | Mayo | 0 | 0 | 0 | 0 | 0 | 0 | | |
| 20 | | 2011 | \$ 6.418.000 | Miércoles | 0 | 0 | Mayo | 0 | 0 | 0 | 0 | 0 | 0 | | |
| 21 | | 2011 | \$ 9.717.000 | Jueves | 0 | 0 | Mayo | 0 | 0 | 0 | 0 | 0 | 0 | | |
| 22 | | 2011 | \$ 11.245.000 | Viernes | 0 | 0 | Mayo | 0 | 0 | 0 | 0 | 0 | 0 | | |
| 23 | | 2011 | \$ 7.114.000 | Sábado | 0 | 0 | Mayo | 0 | 0 | 0 | 0 | 0 | 0 | | |
| | | | | Mayo | | | | | | | | | | | |
| | | Enero | Febrero | Marzo | Abril | Mayo | Junio | Julio | Agosto | Septiembre | Octubre | | | | |

Figura 9.13: Base de datos filtrado del mes de mayo desde el año 2011 hasta el año 2019

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|----|---|------------|-----------------|------------|----------------|--------------|------------|--------------|-------------|------------|----------------|-------------|-----------------|---|
| 1 | | | | | | | | | | | | | | |
| 2 | | Año | Ingresos | Día | Festivo | D.a.f | Mes | T. D. | T.SP | R.E | D.a.R.E | S.S. | D.a.S.S. | |
| 3 | | 2011 | \$10.140.000 | Miércoles | 0 | 1 | Junio | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 4 | | 2011 | \$11.986.000 | Jueves | 0 | 1 | Junio | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 5 | | 2011 | \$14.048.000 | Viernes | 0 | 1 | Junio | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 6 | | 2011 | \$10.659.000 | Sábado | 0 | 1 | Junio | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 7 | | 2011 | \$7.474.000 | Domingo | 0 | 1 | Junio | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 8 | | 2011 | \$8.702.000 | Lunes | 1 | 0 | Junio | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 9 | | 2011 | \$8.738.000 | Martes | 0 | 0 | Junio | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 10 | | 2011 | \$8.403.000 | Miércoles | 0 | 0 | Junio | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 11 | | 2011 | \$7.920.000 | Jueves | 0 | 0 | Junio | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 12 | | 2011 | \$10.290.000 | Viernes | 0 | 0 | Junio | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 13 | | 2011 | \$7.557.000 | Sábado | 0 | 0 | Junio | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 14 | | 2011 | \$10.436.000 | Domingo | 0 | 0 | Junio | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 15 | | 2011 | \$8.920.000 | Lunes | 0 | 0 | Junio | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 16 | | 2011 | \$9.546.000 | Martes | 0 | 0 | Junio | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 17 | | 2011 | \$8.918.000 | Miércoles | 0 | 0 | Junio | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 18 | | 2011 | \$11.579.000 | Jueves | 0 | 0 | Junio | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 19 | | 2011 | \$11.642.000 | Viernes | 0 | 0 | Junio | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 20 | | 2011 | \$10.767.000 | Sábado | 0 | 0 | Junio | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 21 | | 2011 | \$12.531.000 | Domingo | 0 | 0 | Junio | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 22 | | 2011 | \$11.328.000 | Lunes | 0 | 1 | Junio | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 23 | | 2011 | \$10.950.000 | Martes | 0 | 1 | Junio | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| | | | | | | | Junio | | | | | | | |
| | | Enero | Febrero | Marzo | Abril | Mayo | Junio | Julio | Agosto | Septiembre | Octubre | | | |

Figura 9.14: Base de datos filtrado del mes de junio desde el año 2011 hasta el año 2019

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|----|---|------------|-----------------|------------|----------------|--------------|------------|--------------|-------------|------------|----------------|-------------|-----------------|---|
| 1 | | | | | | | | | | | | | | |
| 2 | | Año | Ingresos | Día | Festivo | D.a.f | Mes | T. D. | T.SP | R.E | D.a.R.E | S.S. | D.a.S.S. | |
| 3 | | 2011 | \$11.157.000 | Viernes | 0 | 1 | Julio | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 4 | | 2011 | \$8.804.000 | Sábado | 0 | 1 | Julio | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 5 | | 2011 | \$7.491.000 | Domingo | 0 | 1 | Julio | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 6 | | 2011 | \$8.712.000 | Lunes | 1 | 0 | Julio | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 7 | | 2011 | \$10.945.000 | Martes | 0 | 0 | Julio | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 8 | | 2011 | \$14.452.000 | Miércoles | 0 | 0 | Julio | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 9 | | 2011 | \$10.154.000 | Jueves | 0 | 0 | Julio | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 10 | | 2011 | \$12.462.000 | Viernes | 0 | 0 | Julio | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 11 | | 2011 | \$10.822.000 | Sábado | 0 | 0 | Julio | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 12 | | 2011 | \$12.696.000 | Domingo | 0 | 0 | Julio | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 13 | | 2011 | \$10.612.000 | Lunes | 0 | 0 | Julio | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 14 | | 2011 | \$8.811.000 | Martes | 0 | 0 | Julio | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 15 | | 2011 | \$10.809.000 | Miércoles | 0 | 0 | Julio | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 16 | | 2011 | \$10.908.000 | Jueves | 0 | 0 | Julio | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 17 | | 2011 | \$12.526.000 | Viernes | 0 | 0 | Julio | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 18 | | 2011 | \$10.809.000 | Sábado | 0 | 0 | Julio | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 19 | | 2011 | \$13.361.000 | Domingo | 0 | 0 | Julio | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 20 | | 2011 | \$10.060.000 | Lunes | 0 | 0 | Julio | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 21 | | 2011 | \$7.909.000 | Martes | 0 | 0 | Julio | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 22 | | 2011 | \$9.867.000 | Miércoles | 1 | 0 | Julio | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 23 | | 2011 | \$12.055.000 | Jueves | 0 | 0 | Julio | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | | | | | Julio | | | | | | | |
| | | Enero | Febrero | Marzo | Abril | Mayo | Junio | Julio | Agosto | Septiembre | Octubre | | | |

Figura 9.15: Base de datos filtrado del mes de julio desde el año 2011 hasta el año 2019

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|----|---|------------|-----------------|------------|----------------|--------------|------------|--------------|-------------|------------|----------------|-------------|-----------------|---|
| 1 | | | | | | | | | | | | | | |
| 2 | | Año | Ingresos | Día | Festivo | D.a.f | Mes | T. D. | T.SP | R.E | D.a.R.E | S.S. | D.a.S.S. | |
| 3 | | 2011 | \$9.383.000 | Lunes | 0 | 0 | Agosto | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | | 2011 | \$7.966.000 | Martes | 0 | 0 | Agosto | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | | 2011 | \$9.298.000 | Miércoles | 0 | 0 | Agosto | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | | 2011 | \$9.980.000 | Jueves | 0 | 0 | Agosto | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7 | | 2011 | \$11.805.000 | Viernes | 0 | 0 | Agosto | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 8 | | 2011 | \$8.146.000 | Sábado | 0 | 0 | Agosto | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 9 | | 2011 | \$10.928.000 | Domingo | 1 | 0 | Agosto | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 10 | | 2011 | \$8.508.000 | Lunes | 0 | 1 | Agosto | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 11 | | 2011 | \$10.534.000 | Martes | 0 | 1 | Agosto | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 12 | | 2011 | \$10.064.000 | Miércoles | 0 | 1 | Agosto | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 13 | | 2011 | \$13.217.000 | Jueves | 0 | 1 | Agosto | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 14 | | 2011 | \$15.667.000 | Viernes | 0 | 1 | Agosto | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15 | | 2011 | \$10.145.000 | Sábado | 0 | 1 | Agosto | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 16 | | 2011 | \$9.491.000 | Domingo | 0 | 1 | Agosto | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 17 | | 2011 | \$8.363.000 | Lunes | 1 | 0 | Agosto | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 18 | | 2011 | \$8.217.000 | Martes | 0 | 1 | Agosto | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 19 | | 2011 | \$8.860.000 | Miércoles | 0 | 0 | Agosto | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 20 | | 2011 | \$10.740.000 | Jueves | 0 | 0 | Agosto | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 21 | | 2011 | \$11.702.000 | Viernes | 0 | 0 | Agosto | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 22 | | 2011 | \$8.402.000 | Sábado | 0 | 0 | Agosto | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 23 | | 2011 | \$9.704.000 | Domingo | 0 | 0 | Agosto | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | | | | | Agosto | | | | | | | |
| | | Enero | Febrero | Marzo | Abril | Mayo | Junio | Julio | Agosto | Septiembre | Octubre | | | |

Figura 9.16: Base de datos filtrado del mes de agosto desde el año 2011 hasta el año 2019

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|----|---|------|---------------|-----------|---------|-------|------------|-------|------|-----|---------|------|----------|---|
| | | Año | Ingresos | Día | Festivo | D.a.f | Mes | T. D. | T.SP | R.E | D.a.R.E | S.S. | D.a.S.S. | |
| 3 | | 2011 | \$ 9.522.000 | Jueves | 0 | 0 | Septiembre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 4 | | 2011 | \$ 12.204.000 | Viernes | 0 | 0 | Septiembre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 5 | | 2011 | \$ 9.186.000 | Sábado | 0 | 0 | Septiembre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 6 | | 2011 | \$ 11.846.000 | Domingo | 0 | 0 | Septiembre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 7 | | 2011 | \$ 10.128.000 | Lunes | 0 | 0 | Septiembre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 8 | | 2011 | \$ 8.559.000 | Martes | 0 | 0 | Septiembre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 9 | | 2011 | \$ 8.688.000 | Miércoles | 0 | 0 | Septiembre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 10 | | 2011 | \$ 8.766.000 | Jueves | 0 | 0 | Septiembre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 11 | | 2011 | \$ 13.809.000 | Viernes | 0 | 0 | Septiembre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 12 | | 2011 | \$ 9.350.000 | Sábado | 0 | 0 | Septiembre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 13 | | 2011 | \$ 9.155.000 | Domingo | 0 | 0 | Septiembre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 14 | | 2011 | \$ 6.485.000 | Lunes | 0 | 0 | Septiembre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 15 | | 2011 | \$ 8.814.000 | Martes | 0 | 0 | Septiembre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 16 | | 2011 | \$ 8.526.000 | Miércoles | 0 | 0 | Septiembre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 17 | | 2011 | \$ 11.293.000 | Jueves | 0 | 0 | Septiembre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 18 | | 2011 | \$ 13.357.000 | Viernes | 0 | 0 | Septiembre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 19 | | 2011 | \$ 7.732.000 | Sábado | 0 | 0 | Septiembre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 20 | | 2011 | \$ 13.016.000 | Domingo | 0 | 0 | Septiembre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 21 | | 2011 | \$ 7.910.000 | Lunes | 0 | 0 | Septiembre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 22 | | 2011 | \$ 8.336.000 | Martes | 0 | 0 | Septiembre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 23 | | 2011 | \$ 9.678.000 | Miércoles | 0 | 0 | Septiembre | 0 | 0 | 0 | 0 | 0 | 0 | |

Figura 9.17: Base de datos filtrado del mes de septiembre desde el año 2011 hasta el año 2019

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|----|---|------|---------------|-----------|---------|-------|---------|-------|------|-----|---------|------|----------|---|
| | | Año | Ingresos | Día | Festivo | D.a.f | Mes | T. D. | T.SP | R.E | D.a.R.E | S.S. | D.a.S.S. | |
| 3 | | 2011 | \$ 8.734.000 | Sábado | 0 | 0 | Octubre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 4 | | 2011 | \$ 11.794.000 | Domingo | 0 | 0 | Octubre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 5 | | 2011 | \$ 9.269.000 | Lunes | 0 | 0 | Octubre | 0 | 0 | 0 | 1 | 0 | 0 | |
| 6 | | 2011 | \$ 8.706.000 | Martes | 0 | 0 | Octubre | 0 | 0 | 0 | 1 | 0 | 0 | |
| 7 | | 2011 | \$ 10.416.000 | Miércoles | 0 | 0 | Octubre | 0 | 0 | 0 | 1 | 0 | 0 | |
| 8 | | 2011 | \$ 11.105.000 | Jueves | 0 | 0 | Octubre | 0 | 0 | 0 | 1 | 0 | 0 | |
| 9 | | 2011 | \$ 14.865.000 | Viernes | 0 | 0 | Octubre | 0 | 0 | 0 | 1 | 0 | 0 | |
| 10 | | 2011 | \$ 10.826.000 | Sábado | 0 | 0 | Octubre | 0 | 0 | 0 | 1 | 0 | 0 | |
| 11 | | 2011 | \$ 12.963.000 | Domingo | 0 | 0 | Octubre | 0 | 0 | 0 | 1 | 0 | 0 | |
| 12 | | 2011 | \$ 11.928.000 | Lunes | 0 | 0 | Octubre | 0 | 0 | 1 | 0 | 0 | 0 | |
| 13 | | 2011 | \$ 13.070.000 | Martes | 0 | 0 | Octubre | 0 | 0 | 1 | 0 | 0 | 0 | |
| 14 | | 2011 | \$ 12.774.000 | Miércoles | 0 | 0 | Octubre | 0 | 0 | 1 | 0 | 0 | 0 | |
| 15 | | 2011 | \$ 15.225.000 | Jueves | 0 | 0 | Octubre | 0 | 0 | 1 | 0 | 0 | 0 | |
| 16 | | 2011 | \$ 17.928.000 | Viernes | 0 | 0 | Octubre | 0 | 0 | 1 | 0 | 0 | 0 | |
| 17 | | 2011 | \$ 14.252.000 | Sábado | 0 | 0 | Octubre | 0 | 0 | 1 | 0 | 0 | 0 | |
| 18 | | 2011 | \$ 10.637.000 | Domingo | 0 | 0 | Octubre | 0 | 0 | 1 | 0 | 0 | 0 | |
| 19 | | 2011 | \$ 10.113.000 | Lunes | 1 | 0 | Octubre | 0 | 0 | 1 | 0 | 0 | 0 | |
| 20 | | 2011 | \$ 9.174.000 | Martes | 0 | 0 | Octubre | 0 | 0 | 1 | 0 | 0 | 0 | |
| 21 | | 2011 | \$ 10.487.000 | Miércoles | 0 | 0 | Octubre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 22 | | 2011 | \$ 9.925.000 | Jueves | 0 | 0 | Octubre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 23 | | 2011 | \$ 11.916.000 | Viernes | 0 | 0 | Octubre | 0 | 0 | 0 | 0 | 0 | 0 | |

Figura 9.18: Base de datos filtrado del mes de octubre desde el año 2011 hasta el año 2019

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|----|---|------|---------------|-----------|---------|-------|-----------|-------|------|-----|---------|------|----------|---|
| | | Año | Ingresos | Día | Festivo | D.a.f | Mes | T. D. | T.SP | R.E | D.a.R.E | S.S. | D.a.S.S. | |
| 3 | | 2011 | \$ 9.193.000 | Martes | 0 | 1 | Noviembre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 4 | | 2011 | \$ 11.176.000 | Miércoles | 0 | 1 | Noviembre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 5 | | 2011 | \$ 13.451.000 | Jueves | 0 | 1 | Noviembre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 6 | | 2011 | \$ 19.151.000 | Viernes | 0 | 1 | Noviembre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 7 | | 2011 | \$ 9.950.000 | Sábado | 0 | 1 | Noviembre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 8 | | 2011 | \$ 9.897.000 | Domingo | 0 | 1 | Noviembre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 9 | | 2011 | \$ 9.831.000 | Lunes | 1 | 0 | Noviembre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 10 | | 2011 | \$ 8.188.000 | Martes | 0 | 1 | Noviembre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 11 | | 2011 | \$ 9.913.000 | Miércoles | 0 | 1 | Noviembre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 12 | | 2011 | \$ 14.820.000 | Jueves | 0 | 1 | Noviembre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 13 | | 2011 | \$ 17.583.000 | Viernes | 0 | 1 | Noviembre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 14 | | 2011 | \$ 13.059.000 | Sábado | 0 | 1 | Noviembre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 15 | | 2011 | \$ 7.591.000 | Domingo | 0 | 1 | Noviembre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 16 | | 2011 | \$ 9.463.000 | Lunes | 1 | 0 | Noviembre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 17 | | 2011 | \$ 8.424.000 | Martes | 0 | 0 | Noviembre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 18 | | 2011 | \$ 10.799.000 | Miércoles | 0 | 0 | Noviembre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 19 | | 2011 | \$ 10.639.000 | Jueves | 0 | 0 | Noviembre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 20 | | 2011 | \$ 11.719.000 | Viernes | 0 | 0 | Noviembre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 21 | | 2011 | \$ 8.404.000 | Sábado | 0 | 0 | Noviembre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 22 | | 2011 | \$ 8.637.000 | Domingo | 0 | 0 | Noviembre | 0 | 0 | 0 | 0 | 0 | 0 | |
| 23 | | 2011 | \$ 9.703.000 | Lunes | 0 | 0 | Noviembre | 0 | 0 | 0 | 0 | 0 | 0 | |

Figura 9.19: Base de datos filtrado del mes de noviembre desde el año 2011 hasta el año 2019

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|----|---|------|--------------|-----------|---------|-------|-----------|------|------|-----|---------|------|----------|---|
| 1 | | | | | | | | | | | | | | |
| 2 | | Año | Ingresos | Día | Festivo | D.a.f | Mes | T.D. | T.SP | R.E | D.a.R.E | S.S. | D.a.S.S. | |
| 3 | | 2011 | \$13.400.000 | Jueves | 0 | 0 | Diciembre | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | | 2011 | \$16.131.000 | Viernes | 0 | 0 | Diciembre | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | | 2011 | \$11.303.000 | Sábado | 0 | 0 | Diciembre | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | | 2011 | \$15.183.000 | Domingo | 0 | 0 | Diciembre | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7 | | 2011 | \$13.561.000 | Lunes | 0 | 1 | Diciembre | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 8 | | 2011 | \$13.270.000 | Martes | 0 | 1 | Diciembre | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 9 | | 2011 | \$15.171.000 | Miércoles | 0 | 1 | Diciembre | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 10 | | 2011 | \$11.466.000 | Jueves | 1 | 0 | Diciembre | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 11 | | 2011 | \$13.140.000 | Viernes | 0 | 1 | Diciembre | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 12 | | 2011 | \$12.925.000 | Sábado | 0 | 1 | Diciembre | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 13 | | 2011 | \$14.582.000 | Domingo | 0 | 1 | Diciembre | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 14 | | 2011 | \$12.524.000 | Lunes | 0 | 0 | Diciembre | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15 | | 2011 | \$14.130.000 | Martes | 0 | 0 | Diciembre | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 16 | | 2011 | \$14.880.000 | Miércoles | 0 | 0 | Diciembre | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 17 | | 2011 | \$13.502.000 | Jueves | 0 | 0 | Diciembre | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 18 | | 2011 | \$14.611.000 | Viernes | 0 | 0 | Diciembre | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 19 | | 2011 | \$13.154.000 | Sábado | 0 | 0 | Diciembre | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 20 | | 2011 | \$15.625.000 | Domingo | 0 | 0 | Diciembre | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 21 | | 2011 | \$13.212.000 | Lunes | 0 | 0 | Diciembre | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 22 | | 2011 | \$15.640.000 | Martes | 0 | 0 | Diciembre | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 23 | | 2011 | \$14.923.000 | Miércoles | 0 | 0 | Diciembre | 1 | 0 | 0 | 0 | 0 | 0 | 0 |

Figura 9.20: Base de datos filtrado del mes de diciembre desde el año 2011 hasta el año 2019

9.3. Anexo 3: Base de datos con variables dummies

Las siguientes imágenes tiene como objetivo presentar los datos después de procesarlo con el código de PYTHON y crear las variables dummies o variables categóricas.

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | |
|----|---|------|----------|---------|-------|------|------|-----|---------|------|----------|---------|--------|-------|--------|-----------|--------|---------|---|--|
| 1 | | | | | | | | | | | | | | | | | | | | |
| 2 | | Año | Ingresos | Festivo | D.a.f | T.D. | T.SP | R.E | D.a.R.E | S.S. | D.a.S.S. | Domingo | Jueves | Lunes | Martes | Miércoles | Sábado | Viernes | | |
| 3 | | 2011 | 4634000 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | |
| 4 | | 2011 | 6642000 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 5 | | 2011 | 8649000 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | |
| 6 | | 2011 | 8437000 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | |
| 7 | | 2011 | 9306000 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | |
| 8 | | 2011 | 10119000 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 9 | | 2011 | 12409000 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | |
| 10 | | 2011 | 11542000 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | |
| 11 | | 2011 | 7712000 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 12 | | 2011 | 7630000 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | |
| 13 | | 2011 | 9264000 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | |
| 14 | | 2011 | 8048000 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | |
| 15 | | 2011 | 8393000 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 16 | | 2011 | 10720000 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | |
| 17 | | 2011 | 9325000 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | |
| 18 | | 2011 | 9790000 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 19 | | 2011 | 9423000 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | |
| 20 | | 2011 | 8996000 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | |
| 21 | | 2011 | 9031000 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | |
| 22 | | 2011 | 11459000 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 23 | | 2011 | 10663000 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | |

Figura 9.21: Base de datos con variables dummies del mes de enero

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S |
|----|---|------|----------|---------|-------|------|------|-----|---------|------|----------|---------|--------|-------|--------|-----------|--------|---------|---|
| 1 | | | | | | | | | | | | | | | | | | | |
| 2 | | Año | Ingresos | Festivo | D.a.f | T.D. | T.SP | R.E | D.a.R.E | S.S. | D.a.S.S. | Domingo | Jueves | Lunes | Martes | Miércoles | Sábado | Viernes | |
| 3 | | 2011 | 8451000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 4 | | 2011 | 7330000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 5 | | 2011 | 9008000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | | 2011 | 9421000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 7 | | 2011 | 6767000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 8 | | 2011 | 9007000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 9 | | 2011 | 8042000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 10 | | 2011 | 7357000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 11 | | 2011 | 8079000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 12 | | 2011 | 8150000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 13 | | 2011 | 11419000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 14 | | 2011 | 6762000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 15 | | 2011 | 9443000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 16 | | 2011 | 6619000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 17 | | 2011 | 7003000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 18 | | 2011 | 6295000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 19 | | 2011 | 6519000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 20 | | 2011 | 9664000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 21 | | 2011 | 5877000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 22 | | 2011 | 10056000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 23 | | 2011 | 8710000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |

Figura 9.22: Base de datos con variables dummies del mes de febrero

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S |
|----|---|------|----------|---------|-------|------|------|-----|---------|------|----------|---------|--------|-------|--------|-----------|--------|---------|---|
| 1 | | | | | | | | | | | | | | | | | | | |
| 2 | | Año | Ingresos | Festivo | D.a.f | T.D. | T.SP | R.E | D.a.R.E | S.S. | D.a.S.S. | Domingo | Jueves | Lunes | Martes | Miércoles | Sábado | Viernes | |
| 3 | | 2011 | 6970000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 4 | | 2011 | 6674000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 5 | | 2011 | 10047000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | | 2011 | 10307000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 7 | | 2011 | 7068000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 8 | | 2011 | 10547000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 9 | | 2011 | 7056000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 10 | | 2011 | 7784000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 11 | | 2011 | 6821000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 12 | | 2011 | 7997000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 13 | | 2011 | 9470000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 14 | | 2011 | 5256000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 15 | | 2011 | 8534000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 16 | | 2011 | 6812000 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 17 | | 2011 | 6879000 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 18 | | 2011 | 8696000 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 19 | | 2011 | 10980000 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 20 | | 2011 | 14114000 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 21 | | 2011 | 11227000 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 22 | | 2011 | 9014000 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 23 | | 2011 | 6552000 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |

Figura 9.23: Base de datos con variables dummies de los meses marzo y abril

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V |
|----|---|------|----------|---------|-------|------|------|-----|---------|------|----------|---------|--------|-------|--------|-----------|--------|---------|-------|-------|------|---|
| 1 | | | | | | | | | | | | | | | | | | | | | | |
| 2 | | Año | Ingresos | Festivo | D.a.f | T.D. | T.SP | R.E | D.a.R.E | S.S. | D.a.S.S. | Domingo | Jueves | Lunes | Martes | Miércoles | Sábado | Viernes | Julio | Junio | Mayo | |
| 3 | | 2011 | 8577000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 4 | | 2011 | 7241000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 5 | | 2011 | 7430000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 6 | | 2011 | 7188000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 7 | | 2011 | 9913000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 8 | | 2011 | 12250000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 9 | | 2011 | 7601000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| 10 | | 2011 | 10412000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 11 | | 2011 | 7932000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 12 | | 2011 | 7448000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 13 | | 2011 | 7466000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 14 | | 2011 | 10509000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 15 | | 2011 | 11362000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 16 | | 2011 | 6277000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| 17 | | 2011 | 9225000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 18 | | 2011 | 7803000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 19 | | 2011 | 6366000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 20 | | 2011 | 6418000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 21 | | 2011 | 9717000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 22 | | 2011 | 11245000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 23 | | 2011 | 7114000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |

Figura 9.24: Base de datos con variables dummies de los meses mayo, junio y julio

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T |
|----|---|------|----------|---------|-------|-------|------|-----|---------|------|----------|---------|--------|-------|--------|-----------|--------|--------|---------|---|
| 1 | | | | | | | | | | | | | | | | | | | | |
| 2 | | Año | Ingresos | Festivo | D.a.f | T. D. | T.SP | R.E | D.a.R.E | S.S. | D.a.S.S. | Domingo | Jueves | Lunes | Martes | Miércoles | Sábado | Sábado | Viernes | |
| 3 | | 2011 | 9383000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | | 2011 | 7966000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 5 | | 2011 | 9298000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 6 | | 2011 | 9980000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7 | | 2011 | 11805000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 8 | | 2011 | 8146000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 9 | | 2011 | 10928000 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 10 | | 2011 | 8508000 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 11 | | 2011 | 10534000 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 12 | | 2011 | 10064000 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 13 | | 2011 | 13217000 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 14 | | 2011 | 15667000 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 15 | | 2011 | 10145000 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 16 | | 2011 | 9491000 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 17 | | 2011 | 8363000 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 18 | | 2011 | 8217000 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 19 | | 2011 | 8860000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 20 | | 2011 | 10740000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 21 | | 2011 | 11702000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 22 | | 2011 | 8402000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 23 | | 2011 | 9704000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Figura 9.25: Base de datos con variables dummies del mes de agosto

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W |
|----|---|------|----------|---------|-------|-------|------|-----|---------|------|----------|---------|--------|-------|--------|-----------|--------|---------|-----------|-----------|---------|------------|---|
| 1 | | | | | | | | | | | | | | | | | | | | | | | |
| 2 | | Año | Ingresos | Festivo | D.a.f | T. D. | T.SP | R.E | D.a.R.E | S.S. | D.a.S.S. | Domingo | Jueves | Lunes | Martes | Miércoles | Sábado | Viernes | Diciembre | Noviembre | Octubre | Septiembre | |
| 3 | | 2011 | 9522000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 4 | | 2011 | 12204000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| 5 | | 2011 | 9186000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 6 | | 2011 | 11846000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 7 | | 2011 | 10128000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 8 | | 2011 | 8599000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 9 | | 2011 | 8688000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 10 | | 2011 | 8766000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 11 | | 2011 | 13809000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| 12 | | 2011 | 9350000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 13 | | 2011 | 9155000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 14 | | 2011 | 6485000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 15 | | 2011 | 8814000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 16 | | 2011 | 8526000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 17 | | 2011 | 11293000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 18 | | 2011 | 13357000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| 19 | | 2011 | 7732000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 20 | | 2011 | 13016000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 21 | | 2011 | 7910000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 22 | | 2011 | 8336000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 23 | | 2011 | 9678000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |

Figura 9.26: Base de datos con variables dummies de los meses septiembre, octubre, noviembre y diciembre

9.4. Anexo 4: Tabulación de los resultados del modelo predictivos

Las siguientes imágenes tiene como objetivo presentar los resultados que se obtuvieron mediante una tabulación, luego de estructurar el modelo y evaluarlo.

| ENERO | |
|-------|---------------|
| Día | Ingresos |
| 1 | \$217.780.510 |
| 2 | \$33.684.131 |
| 3 | \$32.961.644 |
| 4 | \$25.033.764 |
| 5 | \$23.474.590 |
| 6 | \$23.165.150 |
| 7 | \$23.758.016 |
| 8 | \$24.258.270 |
| 9 | \$26.121.107 |
| 10 | \$25.810.220 |
| 11 | \$18.497.953 |
| 12 | \$26.422.942 |
| 13 | \$21.688.030 |
| 14 | \$19.886.962 |
| 15 | \$21.524.638 |
| 16 | \$25.890.105 |
| 17 | \$25.278.956 |
| 18 | \$17.486.844 |
| 19 | \$24.662.622 |
| 20 | \$21.688.030 |
| 21 | \$19.886.962 |
| 22 | \$21.524.638 |
| 23 | \$25.890.105 |
| 24 | \$25.278.956 |
| 25 | \$17.486.844 |
| 26 | \$24.662.622 |
| 27 | \$21.688.030 |
| 28 | \$19.886.962 |
| 29 | \$21.524.702 |
| 30 | \$25.890.112 |
| 31 | \$25.278.892 |

Figura 9.27: Tabla de ingresos en enero del 2020

FEBRERO

| Día | Ingresos | Día | Ingresos | Día | Ingresos |
|-----|---------------|-----|---------------|-----|---------------|
| 1 | \$ 22.022.101 | 11 | \$ 24.317.596 | 21 | \$ 27.374.870 |
| 2 | \$ 25.673.872 | 12 | \$ 22.554.896 | 22 | \$ 22.022.101 |
| 3 | \$ 22.079.120 | 13 | \$ 24.082.668 | 23 | \$ 25.673.872 |
| 4 | \$ 24.317.596 | 14 | \$ 27.374.886 | 24 | \$ 22.079.120 |
| 5 | \$ 22.554.896 | 15 | \$ 22.022.101 | 25 | \$ 24.317.596 |
| 6 | \$ 24.082.672 | 16 | \$ 25.673.872 | 26 | \$ 22.554.896 |
| 7 | \$ 27.374.870 | 17 | \$ 22.079.120 | 27 | \$ 24.082.672 |
| 8 | \$ 22.022.101 | 18 | \$ 24.317.596 | 28 | \$ 27.374.886 |
| 9 | \$ 25.673.872 | 19 | \$ 22.554.896 | 29 | \$ 22.022.098 |
| 10 | \$ 22.079.120 | 20 | \$ 24.082.672 | | |

Figura 9.28: Tabla de ingresos en febrero del 2020

MARZO

| Día | Ingresos | Día | Ingresos | Día | Ingresos |
|-----|---------------|-----|---------------|-----|---------------|
| 1 | \$ 25.749.288 | 11 | \$ 20.939.048 | 21 | \$ 23.654.099 |
| 2 | \$ 20.547.880 | 12 | \$ 24.094.690 | 22 | \$ 18.001.704 |
| 3 | \$ 20.272.418 | 13 | \$ 28.595.809 | 23 | \$ 14.864.680 |
| 4 | \$ 20.939.048 | 14 | \$ 22.246.497 | 24 | \$ 20.272.418 |
| 5 | \$ 24.094.690 | 15 | \$ 25.749.288 | 25 | \$ 20.939.048 |
| 6 | \$ 28.595.809 | 16 | \$ 15.998.760 | 26 | \$ 24.094.690 |
| 7 | \$ 22.246.497 | 17 | \$ 16.655.853 | 27 | \$ 28.595.809 |
| 8 | \$ 25.749.288 | 18 | \$ 17.605.416 | 28 | \$ 22.246.497 |
| 9 | \$ 20.547.880 | 19 | \$ 21.216.105 | 29 | \$ 25.748.776 |
| 10 | \$ 20.272.418 | 20 | \$ 29.490.785 | 30 | \$ 23.307.048 |
| | | | | 31 | \$ 28.310.399 |

Figura 9.29: Predicción de ingresos en enero del 2020

Figura 9.30: Tabla de ingresos en marzo del 2020

ABRIL

| Día | Ingresos | Día | Ingresos | Día | Ingresos |
|-----|----------------------|-----|---------------|-----|---------------|
| 1 | \$ 28.378.920 | 11 | \$ 28.389.774 | 21 | \$ 20.272.418 |
| 2 | \$ 31.923.816 | 12 | \$ 19.912.488 | 22 | \$ 20.939.048 |
| 3 | \$ 5.870.009.632.670 | 13 | \$ 20.547.880 | 23 | \$ 24.094.690 |
| 4 | \$ 5.870.015.980.958 | 14 | \$ 20.272.418 | 24 | \$ 28.595.809 |
| 5 | \$ 5.870.012.479.191 | 15 | \$ 20.939.048 | 25 | \$ 22.246.497 |
| 6 | \$ 20.617.000 | 16 | \$ 24.094.690 | 26 | \$ 25.749.288 |
| 7 | \$ 25.990.727 | 17 | \$ 28.595.809 | 27 | \$ 20.547.880 |
| 8 | \$ 27.614.504 | 18 | \$ 22.246.497 | 28 | \$ 20.272.418 |
| 9 | \$ 19.134.866 | 19 | \$ 25.749.288 | 29 | \$ 20.939.048 |
| 10 | \$ 28.301.921 | 20 | \$ 20.547.880 | 30 | \$ 21.216.437 |

Figura 9.31: Tabla de ingresos en abril del 2020

MAYO

| Día | Ingresos | Día | Ingresos | Día | Ingresos |
|-----|---------------|-----|---------------|-----|--------------|
| 1 | \$ 24.009.235 | 11 | \$ 23.827.435 | 21 | \$25.046.530 |
| 2 | \$ 22.206.171 | 12 | \$ 21.834.529 | 22 | \$34.139.555 |
| 3 | \$ 25.970.067 | 13 | \$ 24.893.667 | 23 | \$27.201.387 |
| 4 | \$ 23.827.435 | 14 | \$ 25.046.530 | 24 | \$25.271.107 |
| 5 | \$ 21.834.529 | 15 | \$ 24.009.235 | 25 | \$17.733.659 |
| 6 | \$ 24.893.667 | 16 | \$ 22.206.171 | 26 | \$21.834.529 |
| 7 | \$ 25.046.530 | 17 | \$ 25.970.067 | 27 | \$24.893.667 |
| 8 | \$ 24.009.235 | 18 | \$ 23.827.435 | 28 | \$25.046.530 |
| 9 | \$ 22.206.171 | 19 | \$ 21.834.529 | 29 | \$24.009.235 |
| 10 | \$ 25.970.067 | 20 | \$ 24.893.667 | 30 | \$30.061.667 |
| | | | | 31 | \$32.379.235 |

Figura 9.32: Tabla de ingresos en mayo del 2020

JUNIO

| Día | Ingresos | Día | Ingresos | Día | Ingresos |
|-----|---------------|-----|---------------|-----|--------------|
| 1 | \$ 32.834.659 | 11 | \$ 30.162.531 | 21 | \$27.587.171 |
| 2 | \$ 27.326.563 | 12 | \$ 36.801.379 | 22 | \$24.081.507 |
| 3 | \$ 31.380.579 | 13 | \$ 30.819.171 | 23 | \$29.019.491 |
| 4 | \$ 29.475.683 | 14 | \$ 27.586.659 | 24 | \$32.018.787 |
| 5 | \$ 32.848.739 | 15 | \$ 24.081.507 | 25 | \$30.162.531 |
| 6 | \$ 30.061.667 | 16 | \$ 29.019.491 | 26 | \$36.801.379 |
| 7 | \$ 32.379.235 | 17 | \$ 32.018.787 | 27 | \$30.819.171 |
| 8 | \$ 36.608.355 | 18 | \$ 30.162.531 | 28 | \$27.587.171 |
| 9 | \$ 29.019.491 | 19 | \$ 36.801.379 | 29 | \$22.714.683 |
| 10 | \$ 32.018.787 | 20 | \$ 30.819.171 | 30 | \$35.688.461 |

Figura 9.33: Tabla de ingresos en junio del 2020

JULIO

| Día | Ingresos | Día | Ingresos | Día | Ingresos |
|-----|---------------|-----|----------------|-----|--------------|
| 1 | \$ 33.686.947 | 11 | \$ 25.193.439 | 21 | \$25.306.046 |
| 2 | \$ 21.972.445 | 12 | \$ 28.213.611 | 22 | \$30.268.067 |
| 3 | \$ 30.268.531 | 13 | \$ 27.946.283 | 23 | \$28.151.383 |
| 4 | \$ 27.948.995 | 14 | \$ 25.306.046 | 24 | \$29.558.979 |
| 5 | \$ 33.254.083 | 15 | \$ 30.268.067 | 25 | \$25.193.439 |
| 6 | \$ 27.946.283 | 16 | \$ 28.151.383 | 26 | \$28.213.611 |
| 7 | \$ 25.306.046 | 17 | \$ 53.502.235 | 27 | \$27.946.283 |
| 8 | \$ 30.268.067 | 18 | \$ 37.126.766 | 28 | \$25.306.046 |
| 9 | \$ 28.151.383 | 19 | \$ 37.402.539 | 29 | \$30.268.067 |
| 10 | \$ 29.558.979 | 20 | -\$ 75.160.700 | 30 | \$28.151.262 |
| | | | | 31 | \$29.558.867 |

Figura 9.34: Tabla de ingresos en julio del 2020

AGOSTO

| Día | Ingresos | Día | Ingresos | Día | Ingresos |
|-----|----------------|-----|---------------|-----|--------------|
| 1 | \$ 18.569.702 | 11 | \$ 22.660.829 | 21 | \$22.244.070 |
| 2 | \$ 21.034.056 | 12 | \$ 19.536.968 | 22 | \$18.569.702 |
| 3 | \$ 20.220.488 | 13 | \$ 20.771.650 | 23 | \$21.034.056 |
| 4 | \$ 22.660.829 | 14 | \$ 31.342.566 | 24 | \$20.220.488 |
| 5 | \$ 19.536.968 | 15 | \$ 24.851.350 | 25 | \$22.660.829 |
| 6 | \$ 20.771.650 | 16 | \$ 18.192.712 | 26 | \$19.536.968 |
| 7 | \$ 263.760.153 | 17 | \$ 14.917.448 | 27 | \$20.771.650 |
| 8 | \$ 18.569.702 | 18 | \$ 22.660.829 | 28 | \$22.244.070 |
| 9 | \$ 21.034.056 | 19 | \$ 19.536.968 | 29 | \$18.569.757 |
| 10 | \$ 20.220.488 | 20 | \$ 20.771.650 | 30 | \$21.034.056 |
| | | | | 31 | \$20.220.232 |

Figura 9.35: Tabla de ingresos en agosto del 2020

SEPTIEMBRE

| Día | Ingresos | Día | Ingresos | Día | Ingresos |
|-----|---------------|-----|---------------|-----|--------------|
| 1 | \$ 21.267.243 | 11 | \$ 26.879.307 | 21 | \$19.554.123 |
| 2 | \$ 25.669.131 | 12 | \$ 18.548.995 | 22 | \$21.267.243 |
| 3 | \$ 24.015.875 | 13 | \$ 23.620.427 | 23 | \$25.669.131 |
| 4 | \$ 26.879.307 | 14 | \$ 19.554.123 | 24 | \$24.015.875 |
| 5 | \$ 18.548.995 | 15 | \$ 21.267.243 | 25 | \$26.879.307 |
| 6 | \$ 23.620.427 | 16 | \$ 25.669.131 | 26 | \$18.548.995 |
| 7 | \$ 19.554.123 | 17 | \$ 24.015.875 | 27 | \$23.620.427 |
| 8 | \$ 21.267.243 | 18 | \$ 26.879.307 | 28 | \$12.464.843 |
| 9 | \$ 25.669.131 | 19 | \$ 18.548.995 | 29 | \$23.381.527 |
| 10 | \$ 24.015.875 | 20 | \$ 23.620.427 | 30 | \$28.550.202 |

Figura 9.36: Tabla de ingresos en septiembre del 2020

OCTUBRE

| Día | Ingresos | Día | Ingresos | Día | Ingresos |
|-----|---------------|-----|----------------------|-----|--------------|
| 1 | \$ 27.303.796 | 11 | \$ 20.914.170 | 21 | \$24.835.002 |
| 2 | \$ 27.479.930 | 12 | \$ 3.139.041.521.605 | 22 | \$25.045.342 |
| 3 | \$ 23.379.235 | 13 | \$ 23.451.451 | 23 | \$26.972.666 |
| 4 | \$ 27.739.770 | 14 | \$ 24.835.002 | 24 | \$19.487.728 |
| 5 | \$ 24.889.594 | 15 | \$ 25.045.342 | 25 | \$24.755.706 |
| 6 | \$ 24.534.206 | 16 | \$ 26.972.666 | 26 | \$19.976.698 |
| 7 | \$ 30.539.962 | 17 | \$ 19.487.728 | 27 | \$25.011.478 |
| 8 | \$ 33.834.197 | 18 | \$ 24.755.706 | 28 | \$28.245.754 |
| 9 | \$ 37.576.762 | 19 | \$ 17.660.026 | 29 | \$27.778.938 |
| 10 | \$ 28.017.964 | 20 | \$ 23.451.451 | 30 | \$31.059.834 |
| | | | | 31 | \$23.379.978 |

Figura 9.37: Tabla de ingresos en octubre del 2020

NOVIEMBRE

| Día | Ingresos | Día | Ingresos | Día | Ingresos |
|-----|---------------|-----|---------------|-----|--------------|
| 1 | \$ 16.071.610 | 11 | \$ 27.026.810 | 21 | \$21.909.930 |
| 2 | \$ 13.729.530 | 12 | \$ 27.355.666 | 22 | \$25.084.410 |
| 3 | \$ 25.040.754 | 13 | \$ 29.552.122 | 23 | \$20.702.458 |
| 4 | \$ 28.280.698 | 14 | \$ 21.909.930 | 24 | \$23.809.626 |
| 5 | \$ 27.686.210 | 15 | \$ 25.084.410 | 25 | \$27.026.810 |
| 6 | \$ 31.059.834 | 16 | \$ 13.729.530 | 26 | \$27.355.666 |
| 7 | \$ 23.379.970 | 17 | \$ 23.809.626 | 27 | \$29.552.122 |
| 8 | \$ 16.071.610 | 18 | \$ 27.026.810 | 28 | \$21.909.930 |
| 9 | \$ 20.702.458 | 19 | \$ 27.355.666 | 29 | \$28.160.762 |
| 10 | \$ 23.809.626 | 20 | \$ 29.552.122 | 30 | \$28.425.722 |

Figura 9.38: Tabla de ingresos en noviembre del 2020

DICIEMBRE

| Día | Ingresos | Día | Ingresos | Día | Ingresos |
|-----|---------------|-----|---------------|-----|--------------|
| 1 | \$ 38.252.362 | 11 | \$ 34.698.362 | 21 | \$28.425.722 |
| 2 | \$ 39.398.266 | 12 | \$ 30.327.122 | 22 | \$38.252.362 |
| 3 | \$ 36.291.142 | 13 | \$ 28.160.762 | 23 | \$39.398.266 |
| 4 | \$ 34.698.362 | 14 | \$ 28.425.722 | 24 | \$36.291.142 |
| 5 | \$ 30.327.122 | 15 | \$ 38.252.362 | 25 | \$34.698.362 |
| 6 | \$ 28.160.762 | 16 | \$ 39.398.266 | 26 | \$30.327.122 |
| 7 | \$ 28.425.722 | 17 | \$ 36.291.142 | 27 | \$28.160.762 |
| 8 | \$ 38.252.362 | 18 | \$ 34.698.362 | 28 | \$28.425.722 |
| 9 | \$ 39.398.266 | 19 | \$ 30.327.122 | 29 | \$38.252.362 |
| 10 | \$ 36.291.142 | 20 | \$ 28.160.762 | 30 | \$39.398.266 |
| | | | | 31 | \$36.291.150 |

Figura 9.39: Tabla de ingresos en diciembre del 2020

Bibliografía

- [1] Jiawei Han, Micheline Kamber, Jian Pei (2012). DATA MINING. Concepts and Techniques. Morgan Kaufmann Publisher, USA.
- [2] Christopher M. Bishop (2006). Pattern Recognition and Machine Learning. Cambridge CB3 0FB, U.K.
- [3] Rafael Caballero Roldán, Enrique Martín Martín, Adrián Riesco Rodríguez (2006). BIG DATA CON PYTHON: Recolección, almacenamiento y proceso. Alfaomega Colombiana S.A. Bogotá.
- [4] SAS THE POWER TO KNOW. La Minería de Datos de la A a la Z (2019): Cómo Descubrir Conocimientos y Crear Mejores Oportunidades. SAS, México
- [5] Levin, R. y Rubin, D. (2004). Estadística para administración y economía. México: Pearson Educación.
- [6] Anderson, D., Sweeney, D. y Williams, T. (2001). Estadística para administración y economía (7 ed., Vol. II). México: Thomson.
- [7] Devore, J. L. (2005). Probabilidad y estadística para ingeniería y ciencias. (6a ed.). México: Thomson Learning.
- [8] Mario F. Triola (2018). Estadística (12 ed.). Pearson Educación de México, S.A.
- [9] Walpole, R. y Myers, R. (1999). Probabilidad y estadística para ingenieros. (6a ed.). Prentice Hall, México.
- [10] Juan Gabriel Gomila Salas (2018). Curso completo de Machine Learning: Data Science en Python. <https://www.udemy.com/course/machinelearningpython/>.
- [11] Juan Gabriel Gomila Salas (2018). Curso completo de Machine Learning: Data Science con RStudio. <https://www.udemy.com/course/r-data-science/>.
- [12] Vilar, J. (2006). Identificación de valores atípicos y observaciones influyentes. Universidad de La Coruña, La Coruña.
- [13] Anderson, David R., Dennis J. Sweeney y Thomas A. Williams (2008). Identificación de valores atípicos y observaciones influyentes. Cengage Learning Editores S.A., México D.C.
- [14] Jay L. Devore (2008). Probabilidad y Estadística para Ingeniería y Ciencias (7 ed.). California Polytechnic State University, San Luis Obispo.
- [15] Empresa Ontruck(2019). IA y ‘machine learning’ para optimizar el transporte de mercancías por carretera. <https://innovadores.larazon.es/es/ia-y-y-machine-learning-para-optimizar-el-transporte-de-mercancias-por-carretera/>.
- [16] Geraldine Forero Gómez, Jonathan Alexander Martínez Lozano (2020). Modelo de regresión lineal múltiple para el pronóstico de ventas de bolsas ecológicas para la empresa. Universidad Cooperativa de Colombia, Bogotá.
- [17] Carlos Rosado Moral (2017). SmartBus: Big Data y Data Science en Transporte Urbano. Universidad Autónoma de Madrid, España.