******

***Lima norte***

***Ingeniería de Sistemas e informática***

**TRABAJO DE INVESTIGACIÓN PARA OPTAR POR EL GRADO ACADÉMICO DE BACHILLER EN INGENIERÍA DE SISTEMAS E INFORMÁTICA**

**Aplicación de Data Science para la toma de decisiones en el área de ventas en RMV VILLAR, 2023**

**Autores**

*Carrión Valverde, Paolo junior (U18304402)*

*Silva Bruno, Ever Javier (U18310958)*

**Docentes**

*VARGAS VARGAS, GUATAMA CLODOMIRO*

*MARIN VERASTEGUI, WILSON RICARDO*

***Lima, Perú\_***

***2023***

Índice

1. Resumen 3
2. Problema de investigation 4
3. Problemas 4
   1. 2.1. Problema General 4
   2. 2.2. Problemas Específicos 4
4. 3.Objetivos. 5
   1. 3.1. Objetivo General 5
   2. 3.2. Objetivo Específico 5
5. Justificación. 4
6. Estado del Arte 5
7. Marco teórico 7
8. Hipotesis 13
9. Metodologia 13
   1. Alcance 13
   2. Enfoque 14
   3. Diseño 14
   4. Población 15
   5. Tipo de muestra 15
   6. Técnica 16
   7. Instrumento 16
10. CRONOGRAMA DE TRABAJO 17
11. PRESUPUESTO 19
12. BIBLIOGRAFÍA 19
13. ANEXOS 23

**1. Resumen**

En el presente proyecto de investigación se constituye mediante el análisis de la problemática que se ha localizado en la empresa RMV VILLAR, de lo cual se ha detectado un estancamiento en las ventas en este 2023.

En relación al desarrollo de la investigación en mención se ha utilizado la metodología de Data Science, de lo cual se procederá utilizar técnicas e instrumentos para la recopilación y prepararlos para el análisis, interpretación y visualización de datos, con ello se espera obtener resultados propicios y así poder responder la pregunta de investigación. Asimismo, el objetivo principal del proyecto de investigación es medir el impacto de la aplicación de Data Science para la toma de decisiones en el área de ventas en RMV VILLAR. Por otro lado, el tipo de investigación que se emplea es explicativa y cuyo enfoque es cuantitativo de diseño no experimental, por ello se utilizaran las técnicas encuestas, técnicas de modelamiento matemático y análisis estadístico, visualización y exploración de datos, con ello permitirá brindar alternativas posibles para un plan de acción y mejorar la toma de decisiones, las cuales se puedan implementar en el área de ventas.

En conclusión, mediante el presente proyecto de investigación se quiere dar a conocer el impacto del uso de la metodología de Data Science, y así poder responder la pregunta de investigación, con ello mejorar e incrementar las ventas en la empresa.

**Palabras claves**: Data science, Python, Metodología, base de datos, algoritmo, paradigma

**2. Problema de Investigación\_**

La empresa **RMV VILLAR** no está obteniendo las ventas deseadas que se habían planeado alcanzar para este periodo del año, debido a que los clientes ya no están comprando los equipos electrónicos como anterioriores trimestres, como consecuencia se está produciendo un estancamiento en sus ventas y no se están logrando obtener los resultados deseados, por tal motivo no les permite seguir creciendo y posicionarse en el mercado tecnológico. Por esa razón, se busca tomar mejores decisiones para resolver de la mejor manera el problema identificado.

Hoy en día, con la aparición de nuevas tecnologías, entre ellas la ciencia de datos que se está empezando aplicar en todas las ramas empresariales y tecnológicas. Es por ello que la estrategia que se está planteando es implementar la metodología de Data Science para poder estudiar los comportamientos de consumo de los clientes. Con ello se extraerá, procesará y se hará la visualización de la información de sus clientes para que se tomen decisiones mediante las herramientas usadas en el procesamiento de los datos, con esto se busca entender a los clientes y así lograr obtener mayores ventas. Por todo esto se tiene la siguiente interrogante: **¿Cuál es el impacto de la aplicación de Data Science para la toma de decisiones en el área de ventas en RMV VILLAR, 2023?**

**3. Problemas**

**3.1. Problema General**

¿Cuál es el impacto de la aplicación de Data Science para la toma de decisiones en el área de ventas en RMV VILLAR, 2023?

**3.2. Problemas Específicos**

PE1: ¿Cuál es el impacto de la aplicación de Data Science en la calidad del servicio de venta para la toma de decisiones en el área de ventas en RMV VILLAR, 2023?

PE2: ¿Cuál es el impacto de la aplicación de Data Science en la mejora de la eficiencia para la toma de decisiones en el área de ventas en RMV VILLAR, 2023?

PE3: ¿Cuál es el impacto de la aplicación de Data Science en la optimización de procesos para la toma de decisiones en el área de ventas en RMV VILLAR, 2023?

**4.Objetivos**

**4.1. Objetivo General**

Medir el impacto de la aplicación de Data Science para la toma de decisiones en el área de ventas en RMV VILLAR, 2023.

**4.2. Objetivo Específico**

● Medir el impacto de la aplicación de Data Science en la calidad del servicio de venta para la toma de decisiones en el área de ventas en RMV VILLAR, 2023.

● Medir el impacto de la aplicación de Data Science en la mejora de la eficiencia para la toma de decisiones en el área de ventas en RMV VILLAR, 2023.

● Medir el impacto de la aplicación de Data Science en la optimización de procesos para la toma de decisiones en el área de ventas en RMV VILLAR, 2023.

**5. Estado del arte**

La investigación propuesta se abarca en el proceso de ventas de la empresa RMV VILLAR cuyo fin es de plantear una propuesta de desarrollo que nos permita conocer de la mejor manera a los clientes y sus comportamientos de consumo para ofrecerle mejores productos o servicios de acuerdo a sus necesidades, para ello apoyándose de la arquitectura de data science donde se utilizará diferentes herramientas para lograr su aplicación entre ellas tenemos Python, machine learning, base de datos SQL. Las herramientas tecnológicas que se usarán en la investigación ya mencionadas ayudan a la recolección de información, encontrar los indicadores de la empresa y deficiencias en el proceso, con todo ello se debe realizar un análisis profundo para que se pueda tener una mejor toma de decisiones y poder ofrecer mejores productos a sus clientes, obteniendo como resultado aumentar las ventas mediante el uso de DATA SCIENCE. Con el conocimiento proporcionado permite generar un mejor panorama y un mejor desempeño [1]. Por eso, se plantea desarrollar en la empresa la arquitectura de data science y a su vez apoyado en Python para desarrollar mejor el proceso de ventas en el año 2023, con ello se busca generar mejores ganancias y más clientes satisfechos.

Para iniciar esta investigación del tema propuesto debemos tener en cuenta que ¡Hasta el 70% e incluso más de los trabajos analíticos realizados por corporaciones han fallado! Incluso luego de que estos corporativos hayan efectuado grandes inversiones, que implican tiempo, el talento y sobre todo mucho dinero, en desarrollar lo que ellos pensaban que serían buenos software de datos y análisis. Debido a que los empresarios tuvieron la difícil tarea de la toma de decisiones y el equipo de análisis no aplicaron las dos cosas fundamentales: En qué consiste este "aspecto más importante" y cómo hacer que este "aspecto más importante", con ello se obtendrá un programa de análisis exitoso [6]. como nos menciona

Está tipo de investigación está hecha con el objetivo de ayudar en la mejora del proceso de ventas de una empresa o negocio, apoyado en DATA SCIENCE para poder realizar un análisis completo de la información y utilizando herramientas tecnológicas como Python que hoy en día se ha vuelto de mucha utilidad para lograr una buena toma de decisiones y como consecuencia aumentar las ventas y los ingresos. Para ello, las aplicaciones con más inteligencia están realizando un mejor uso de los conocimientos adquiridos de los datos, teniendo un mayor impacto en cada industria. En medio de toda está revolución se hallan las herramientas y métodos que lo impulsan, desde el procesamiento de las pilas masivas de datos generados todos los días hasta el aprendizaje que ayuda a tomar acción útil. También las redes neuronales profundas, junto con avances en la maquinaria clásica para el aprendizaje y la informática escalable de la unidad de procesamiento de gráficos (GPU) de propósito general, se han vuelto componentes críticos de la inteligencia artificial, ayudando a muchos de estos asombrosos avances y también para reducir la barrera a la adopción de estas nuevas tecnologías [7].

Hay muchas compañías que cuentan con millones de datos que no saben utilizar de la mejor manera toda esa información, pero existen muchos programas para poder estudiar, analizar los datos y poder aplicarlos de manera útil en la empresa, en donde una de las más importantes es la inteligencia artificial, la cual desarrolla un aprendizaje automático en donde no se requiere una persona que esté resolviendo dichos problemas, sino que el sistema automáticamente lo aprende y lo resuelve.

La ciencia de datos se puede aplicar en distintas ramas y no solo en la que tienen que ver con tecnología. En México durante la pandemia se implementó un algoritmo para analizar los conglomerados de mortalidad por COVID 19, centrado principalmente en un algoritmo híbrido basado en la metodología Batch FMDS para el aprendizaje automático en un marco de Data Science. Como resultados se encontraron dos indicadores claves que estaban relacionados a la mortalidad por COVID 19 en México. Uno de ellos era la densidad poblacional y el otro es el porcentaje de población en pobreza, uno de sus principales resultados fue que los conglomerados con altos valores de tasa de mortalidad tenían en común altos valores de densidad poblacional y bajos niveles de pobreza. [8]. Así de profunda y útil es la ciencia de datos acompañado con un algoritmo híbrido que permite el cálculo mediante el uso de fórmulas matemáticas y estadísticas, de lo cual da como resultados indicadores útiles que ayuden en una correcta toma de decisiones.

La ciencia de datos es un área multidisciplinaria que se enfoca principalmente en el estudio de patrones subyacentes en los datos (estos no tienen que ser obligatoriamente masivos), combinando también métodos matemáticos, estadísticos, computacionales y de negocios. La ciencia de datos en el ámbito organizativo tiene premisas muy claras como “lo que no se puede medir es imposible de controlarlo y si hay algo que no se puede controlar nunca se podrá mejorar”. La ciencia de datos también se puede aplicar con una metodología de multitarea automatizada que contempla alcances estadísticos y articula métodos de análisis para estudiar indicadores organizativos con la finalidad de brindar respuesta a interrogantes pragmáticas [1].

En la actualidad se sigue estudiando y analizando cuáles serían las mejores alternativas para realizar DATA SCIENCE y con ello poder tener un panorama más claro de los resultados que se pueden obtener y la utilidad que se le puede dar [9]. Por ende, las empresas están implementando esta nueva tecnología debido a los grandes resultados que están obteniendo a través del buen uso de los datos almacenados en sus compañías para mejorar y agilizar sus procesos de producción, de salud, comerciales, etc.

Finalmente, como se menciona hay muchas empresas que no saben utilizar sus datos de manera correcta para tomar mejores decisiones, a esto se le suma también la variedad de software que hay en el mercado para aplicar Data Science. Por esta razón la aplicación de Data Science en el proceso de ventas con la ayuda de la implementación Python, es una gran alternativa debido a que es una de las mejores herramientas hoy en día ya que se puede trabajar con una gran facilidad de manejo debido a que tiene bibliotecas que facilitan su uso, lo cual logra buenos resultados en el desarrollo de DATA SCIENCE para las empresas.

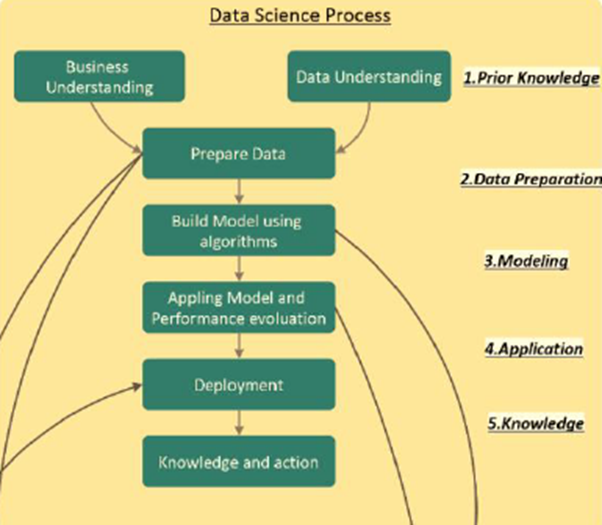
**6. Marco teórico**

En esta presente investigación se va a realizar la aplicación de Data Science para medir su influencia en el proceso de ventas. Data Science puede ser nombrada como un campo interdisciplinario que usa esquemas científicos y algoritmos muy útiles para poder conseguir juicios críticos y conocimientos útiles tras un análisis de los datos, por ello hoy en día Data Science está siendo aceptada como un nuevo paradigma en múltiples campos de la ciencia empírica como son las finanzas, economía o las operaciones. Se considera como un nuevo paradigma del progreso científico [9]. Su objetivo principal es obtener información de los datos y aplicarlo de mejor manera en la toma de decisiones. Data Science casi siempre se asocia con la programación para manejar grupos de datos masivos. Un punto muy importante de data Science que utilizan técnicas estadísticas avanzada, matemáticas y computacionales para poder hallar temas concretos y poder utilizarlo en un conjunto de datos. [9]

Data Science se pueden dividir en dos partes en modelos de aprendizaje supervisado y no supervisado, por ello en la parte supervisada o también llamada guía interna predicen la importancia de las variables de salida en función a las variables de entrada para que se pueda desarrollar un modelo a partir un conjunto de datos de prueba, en cambio la no supervisadas o también llamadas no dirigidas presenta patrones ocultos en datos no marcados, por ello no hay elementos de previsibilidad [14].

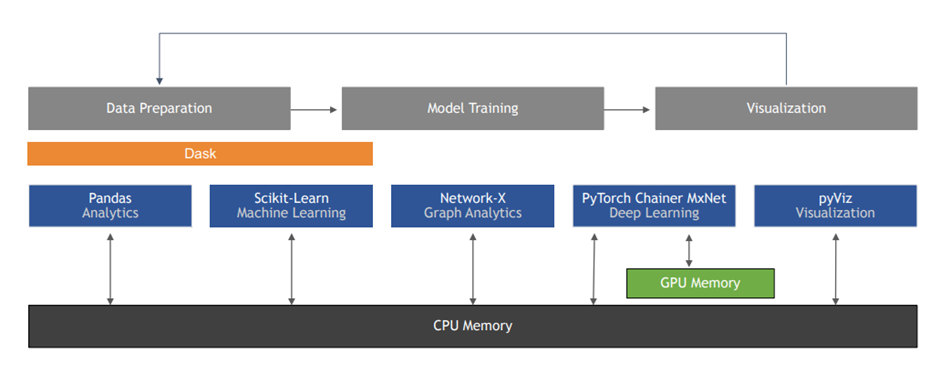
El algoritmo se ha usado para resolver un problema mediante un proceso racional paso a paso, por ello un profesional de ciencia de datos lo primero debe determinar qué algoritmo o algoritmos va utilizar, en los últimos años se han realizado muchos algoritmos para poder facilitar el uso de la ciencia de datos [14].

En el proceso de data Science implica la progresión secuencial del análisis realizado y los descubrimientos hallados, de lo cual este proceso se divide en 5 partes de las cuales son: se identificará el de desafío, se recopila todas las muestras de datos que se tienen, se diseñará el modelo que se va usar, se probará el modelo ya hecho para ver cómo funciona realmente en la vida real, por último, se implementará el modelo y se darán los resultados. [23]



**figura 1 – los procesos de data Science fuente** [14]

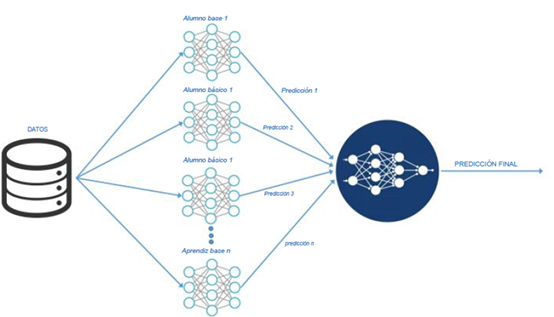
Para lograr la aplicación de la ciencia de datos se debe hacer mediante el uso de software especializado que nos brinde herramientas para el manejo de los datos. Python es uno de los lenguajes de programación muy populares hoy en día, debido a que es fácil de aprender y con sus bibliotecas ayuda a desarrollar trabajos web, ciencias computacionales, aplicaciones científicas y numéricas, por ello Python es ideal para ser utilizado en el análisis de datos y también el aprendizaje automático [9]. Python es simple, claro e intuitivo, es por ello que hay muchos científicos e ingenieros que lo eligen para realizar diversas aplicaciones científicas y numéricas. Por su nivel de facilidad para aprender permite ingresar de manera rápida a los proyectos logrando obtener información valiosa en poco tiempo y ahorrando recursos [7].

****

**Figura 2**. El ecosistema estándar de Python para el aprendizaje automático, la ciencia de datos y la computación científica [7].

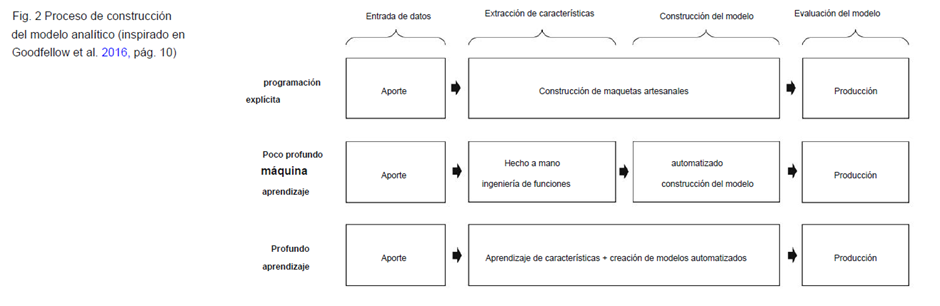
Para el análisis de los datos y poder luego tomar mejores decisiones se necesita la aplicación de algoritmos de aprendizaje automático. Machine learning surgió como un subconjunto de la Inteligencia Artificial (IA), ofreciendo una nueva dirección para construir una comprensión conceptual de cómo trabaja el cerebro humano. Es continuamente usado en el campo científico, de lo cual se concentra en el diseño de los modelos informáticos y algoritmos que desarrollará tareas determinadas, que mayormente consisten en el reconocimiento de patrones, que no tienen necesidad de programarlas de manera explícita [7].

El machine learning comprende el “aprender de los datos”, con la aplicación de algoritmos de aprendizaje automático, posteriormente la computadora reconoce todos estos patrones y utiliza “lo que aprendió” en los nuevos datos [10]. Mediante la aplicación de machine learning se podrá realizar un aprendizaje automático, con ello poder tomar mejores decisiones.

****

**figura 3 - Apilamiento de modelos fuente** [10]

El Proceso para la construcción de modelos analíticos basados en machine learning se divide en 4 partes. En primer lugar, tenemos a la entrada de datos, es este caso los mercados informáticos tienen varios puntos para encontrar la información como son aplicaciones, sitios web, redes sociales, etc. Estos datos pueden ser utilizados para construir modelos analíticos. En segundo lugar, está la extracción de características, una característica se obtiene de la entrada de datos que no han sido procesados con la finalidad de proporcionar una adecuada representación. El objetivo principal es preservar la información y factores que sean relevantes para una buena tarea de aprendizaje. [16] En tercer lugar, tenemos la construcción del modelo, en este punto poder decir que la entrada es usada para un algoritmo de aprendizaje que pueda identificar patrones y también relaciones que son relevantes en la tarea del aprendizaje automático. Por último, tenemos a la evaluación del modelo, aquí se toma en cuenta el rendimiento, recursos empleados y la interpretación que ha adquirido. El objetivo es que el modelo pueda satisfacer la tarea del aprendizaje automático como se buscaba. [10]



**figura 4 – Proceso para la construcción de modelo analíticos** **Machine Learning fuente**[16]

En este punto de la investigación debemos tener claro el significado de un proceso de ventas para en un futuro poder mejorarlo, el proceso de ventas está compuesto por un conjunto de pasos por los cuales se atraviesa hasta efectuar la venta de un producto o servicio.

El proceso inicia desde la identificación del cliente, luego pasa a la demostración del producto, continúa con la argumentación de la venta hasta llegar a la ansiada negociación y por último cerrar la venta, todos estos pasos son parte del proceso de ventas.

Para lograr conseguir una venta con éxito se debe mostrar confianza y credibilidad, con ello se logra que cliente se sienta más seguro de que está realizando la compra de los productos.

Luego de comprender el proceso de ventas, tenemos que tomar en cuenta la toma de decisiones, pues ambos van a ir de la mano para lograr el éxito de esta presente investigación. Hoy en día en el mundo entero es de suma importancia la toma de decisiones, pues gracias a ello se puede definir el buen futuro de un negocio o una organización, también en la vida diaria se ha vuelto en un componente fundamental, de igual manera en el ámbito profesional, social, etc. Por ello, en pocas palabras la toma de decisiones se trata de un proceso en donde se tiene varias opciones y se tiene que elegir a una de ellas. Para tener una mejor toma de decisiones se debe respaldar en un análisis previo de estas opciones para tomar la decisión correcta. [15]

Para una buena toma de decisiones, previamente se debe realizar un correcto análisis de los datos. De la misma manera que la ciencia de datos, también el análisis de datos tiene como finalidad obtener información de los datos y aplicarlo de mejor manera en las tomas de decisiones. En algunas ocasiones el análisis de datos se asocia con el uso de Microsoft Excel y algunas otras herramientas para extraer datos y encontrar patrones [9].

El propósito principal del análisis de datos es recopilar, procesar y analizar los datos para que los encargados puedan tomar mejores decisiones [11], con ello se podrá conocer la realidad de lo que se está estudiando con la obtención de información.

Ya hemos mencionado el uso de Python para esta presente investigación, pero debe estar acompañado de un software de base de datos que lo complemente y que trabajen de la mano, en este caso SQL server es un lenguaje computacional para trabajar con datos, este lenguaje está diseñado para permitir consultar, actualizar, obtener y calcular información de bases de datos. Es uno de los principales sistemas para gestionar bases de datos relacionales. También tiene una interfaz que es el Management Studio, en donde se podrá trabajar haciendo consultas y ver los resultados [12]. La herramienta SQL que ya vamos utilizando varios años trabajando, debido a su facilidad y su panel con todas las herramientas necesarias para hacer una base de datos o múltiples consultas.

Microsoft Excel es una herramienta muy útil para extraer información de gran valor en grandes cantidades de datos. También es muy eficaz para obtener cálculos sencillos cómo realizar sumas, restas, ordenar la información, aplicar filtros y muchas otras utilidades que se pueden explotar y que son de mucha ayuda. Su estructura está compuesta por celdas donde se escribe la información y estas mismas están agrupadas por filas y columnas [13].

**7. Hipótesis**

● **Hipótesis General**

La aplicación de Data Science mejora significativamente la toma de decisiones en el área de ventas en RMV VILLAR, 2023.

● **Hipótesis Específicas**

HE1: La aplicación de Data Science mejora significativamente la calidad de servicio de venta para la toma de decisiones en el área de ventas en RMV VILLAR, 2023.

HE2: La aplicación de Data Science mejora significativamente la eficiencia para la toma de decisiones en el área de ventas en RMV VILLAR, 2023.

HE3. La aplicación de Data Science mejora significativamente la optimización de procesos para la toma de decisiones en el área de ventas en RMV VILLAR, 2023.

**8. Resultados y Discusión**

**9. Metodología**

**A) Alcance**

Está investigación tiene un alcance de tipo explicativo, porque se va a responder las causas y comprender los fenómenos del estudio. Todo ello mediante la implementación de esta metodología se pretende brindar respuestas de manera eficiente y eficaz al momento de tomar decisiones así examinando las relaciones causales de diversas variables. Por ello, se procederá a examinar el comportamiento humano y su toma de decisiones de los clientes, de esta manera gracias a este estudio visto de manera explicativo llegaremos comprender los fenómenos del estudio, al conocer mejor sus elementos cruciales.

**B) Enfoque**

La propuesta del trabajo de investigación se centra en el enfoque cuantitativo porque se van a medir variables numéricas en el proceso de ventas del presente proyecto de investigación. Con ello, se aplicará el enfoque cuantitativo para analizar las ventas realizadas por la empresa RMV VILLAR, con el principal propósito de encontrar resultados que nos puedan ayudar en la búsqueda para mejorar las ventas futuras.

**C)Diseño**

La presente investigación tiene como diseño metodológico de tipo no experimental de corte transversal porque no hay intervención de la variable independiente y las herramientas serán implementadas en un determinado tiempo. Por ello, se procederá a utilizar en está presente investigación el diseño metodológico de tipo no experimental y a su vez de corte transversal que permitirá analizar el problema que se presenta en las ventas obtenidas y encontrar respuestas para mejorar este proceso.

**D)Población**

Se ha tomado como población a todos los registros de ventas de la empresa RMV VILLAR. Para está presente investigación hemos determinado que la población a estudiar son todos los registros de ventas de los últimos dos años de la Empresa RMV VILLAR, que son un aproximadamente de 962 registros de ventas.

**E) Tipo de muestra.**

Para esta presente investigación se ha tomado un tipo de muestreo probabilístico que a su vez es de tipo aleatorio estratificado porque se van a usar extractos o grupos de los registros de ventas para facilitar el estudio de la investigación, está muestra conformada por el 50% de los registros de ventas de los últimos dos años en la empresa RMV VILLAR, dando como resultado 481 registros de ventas.

|  |  |  |  |
| --- | --- | --- | --- |
|  | POBLACION | MUESTRA | % |
| REGISTROS DE VENTAS | 962 | 481 | 50 |

Fuente propia

N =962

α=5% P=50%

**F) Técnica**

Para C. Rojas, R. Ignacio [20] una técnica es un procedimiento propio, de lo cual es validado con la práctica y está orientado a obtener y convertir datos para poder solucionar problemas, para ello la técnica va a acompañada de un instrumento que son muy útiles. En este trabajo se utilizarán estas principales técnicas para poder recoger datos que sean pertinentes sobre las variables involucradas en esta investigación:

● Encuestas

● Técnicas de modelamiento matemático y análisis estadístico

● Visualización y exploración de datos

**G) Instrumento**

Para M. Useche, W. Artigas, B. Queipo y É. Perozo [19] nos menciona que los instrumentos son herramientas muy necesarias para los investigadores, de lo cual se utilizan para la extracción, registro y almacenamiento de información ya sea en papel o digital, con ello poder medir el comportamiento de las variables que se están estudiando. En esta investigación se van a utilizar los principales instrumentos que son:

● Formato de Cuestionario

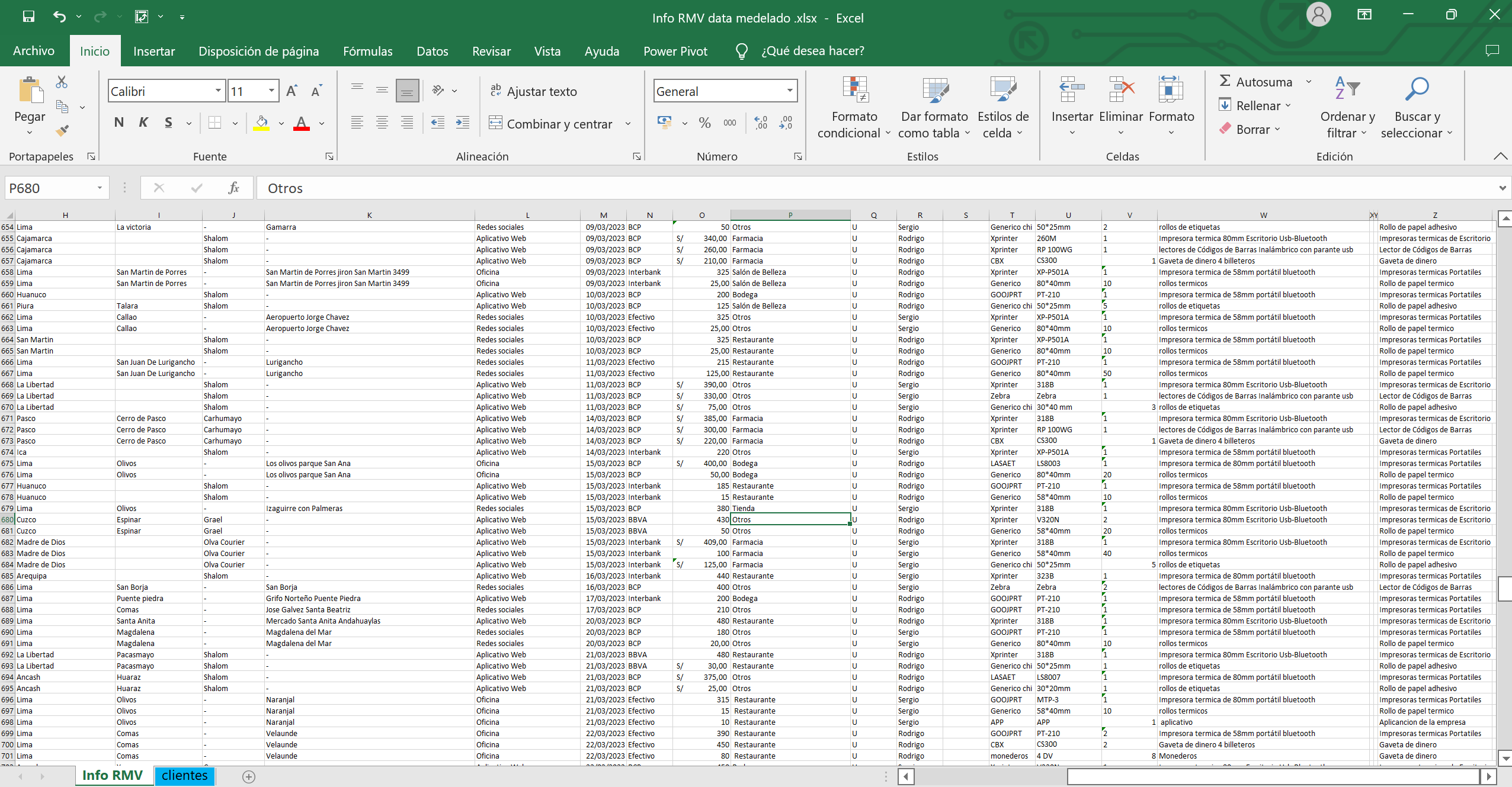
● Guía de Análisis Documental del proceso de venta

● Python (para uso estadística y análisis de datos)

● Power BI, Excel

**Implementation del Proyecto**

* Recolección de información

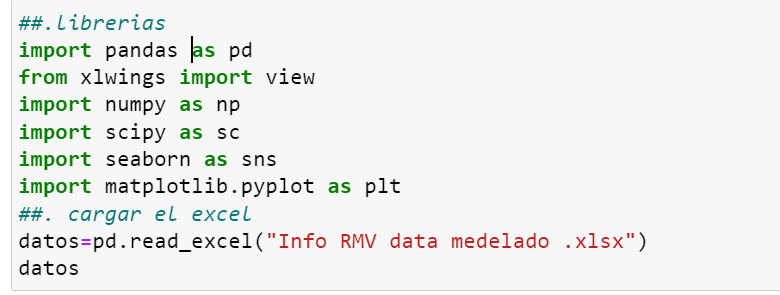


Fuente: Base de datos de empresa RMV

En la imagen nos muestra la base de datos de las ventas que se han realizado en los años 2022 y 2023, en donde se detalla toda la información necesaria para el análisis de los datos, en donde se empleara la aplicación de data Science.

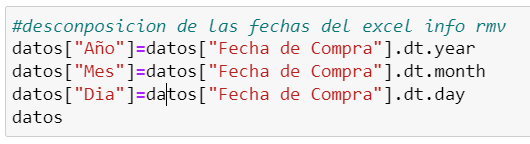
La herramienta que se utilizo es el programa Jupyter notebook y con el lenguaje de programación que se adecua mejor para la aplicación de data science es Python, debido a que tiene facilidad y varias librerías útiles para el desarrollo del proyecto de investigación.

En primer lugar, se proceder a cargar las librerías a usar y la base de datos que es el Excel, por ello se procedió a guardarlo en la variable datos



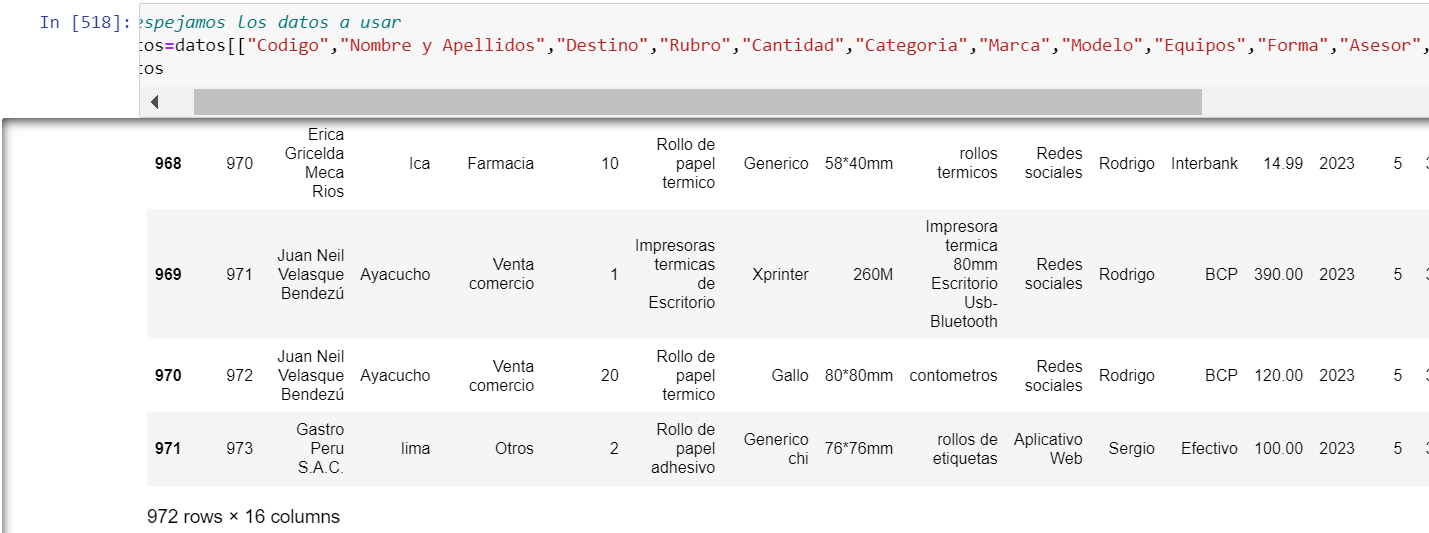
Fuente: Elaboración propia

Antes de realizar la limpieza de los datos se procedió a descomponer la fecha en tres partes, para poder hacer las comparaciones con facilidad ,como se muestra en la imagen



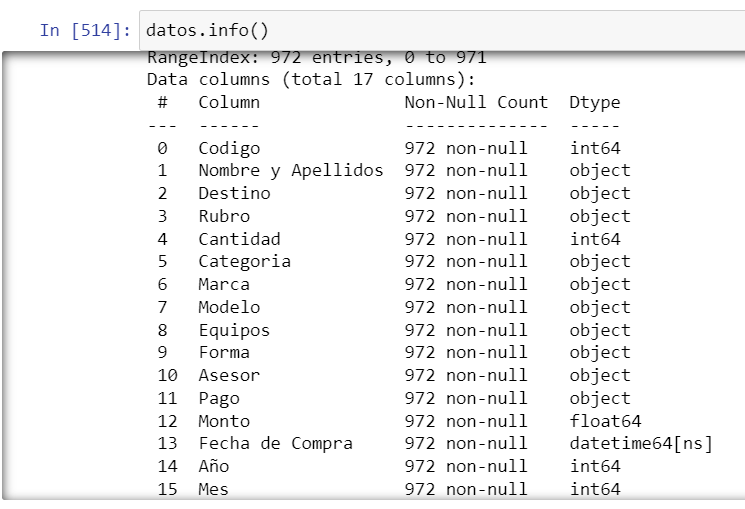
Fuente: Elaboración propia

El tercer paso en la limpieza de datos se tiene que ordenar y seleccionar las columnas que vamos a trabajar



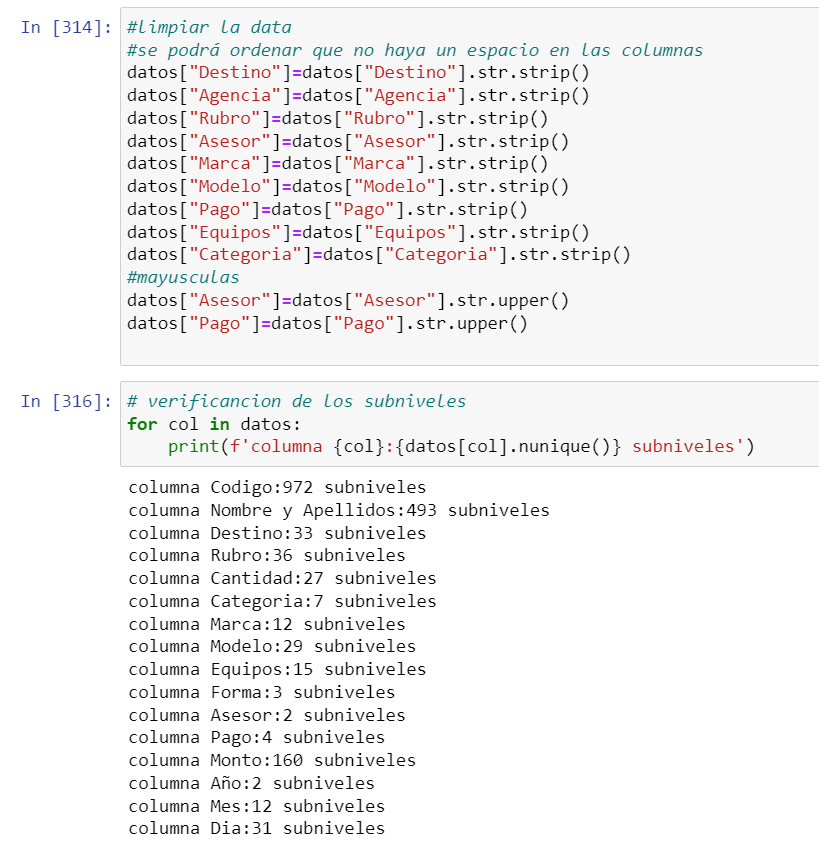
Fuente: Elaboración propia

En el cuarto paso se ejecuta dato.info() ,de lo cual nos muestra los nombres de cada columna ,también cuantos datos hay en cada columna y por ultimo nos muestra el tipo de dato es cada columna como se observa.



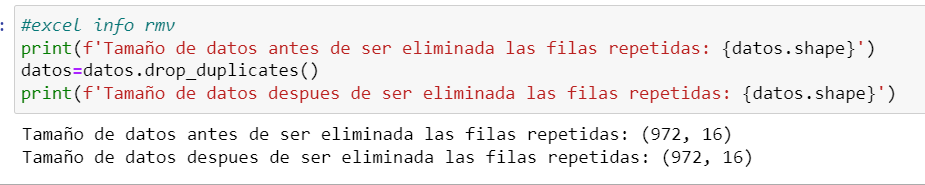
Fuente: Elaboración propia

En este quinto paso pondremos que las columnas con hay datos que tienen dos variables iguales de lo cual se procederá a quitar un espacio que pueda afectar a la contabilidad de los productos, pasar todo a string los que son letras, por ultimo se cambiara de mayúscula a minúscula para no encontrar datos que solo se diferencia por letra mayuscula

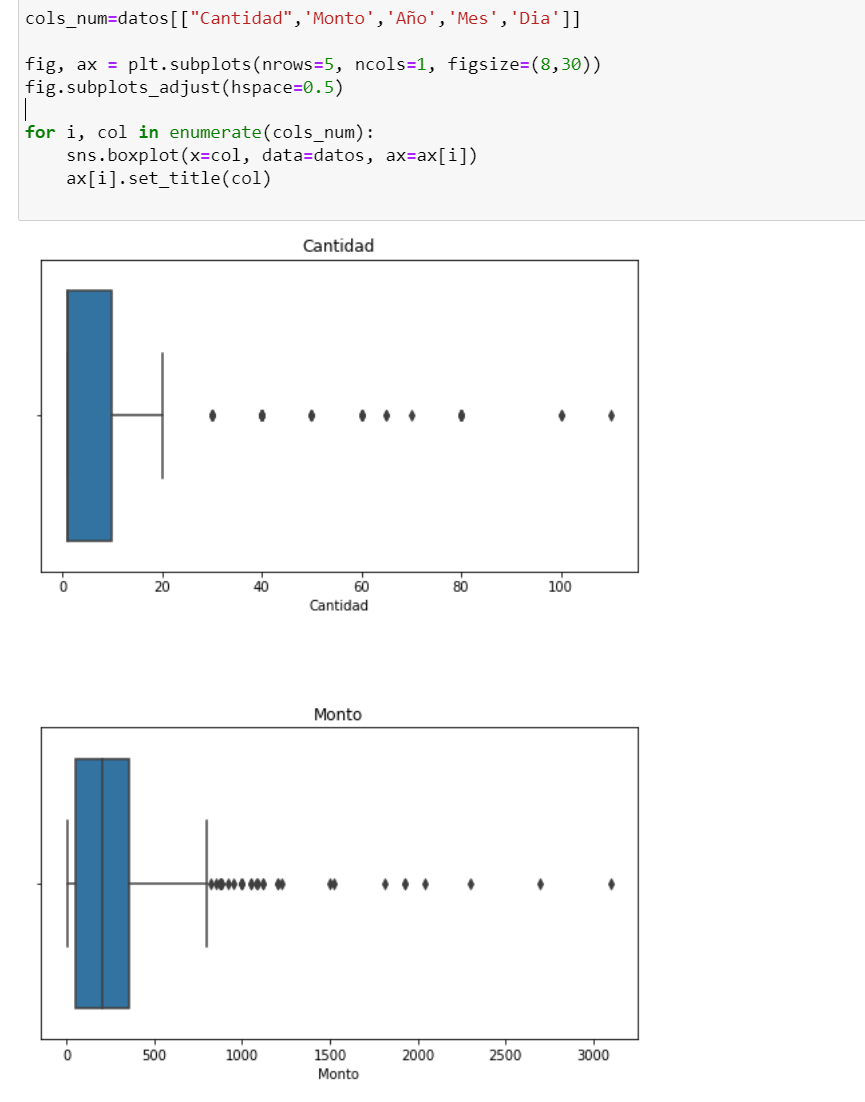


Fuente: Elaboración propia

En el sexto paso

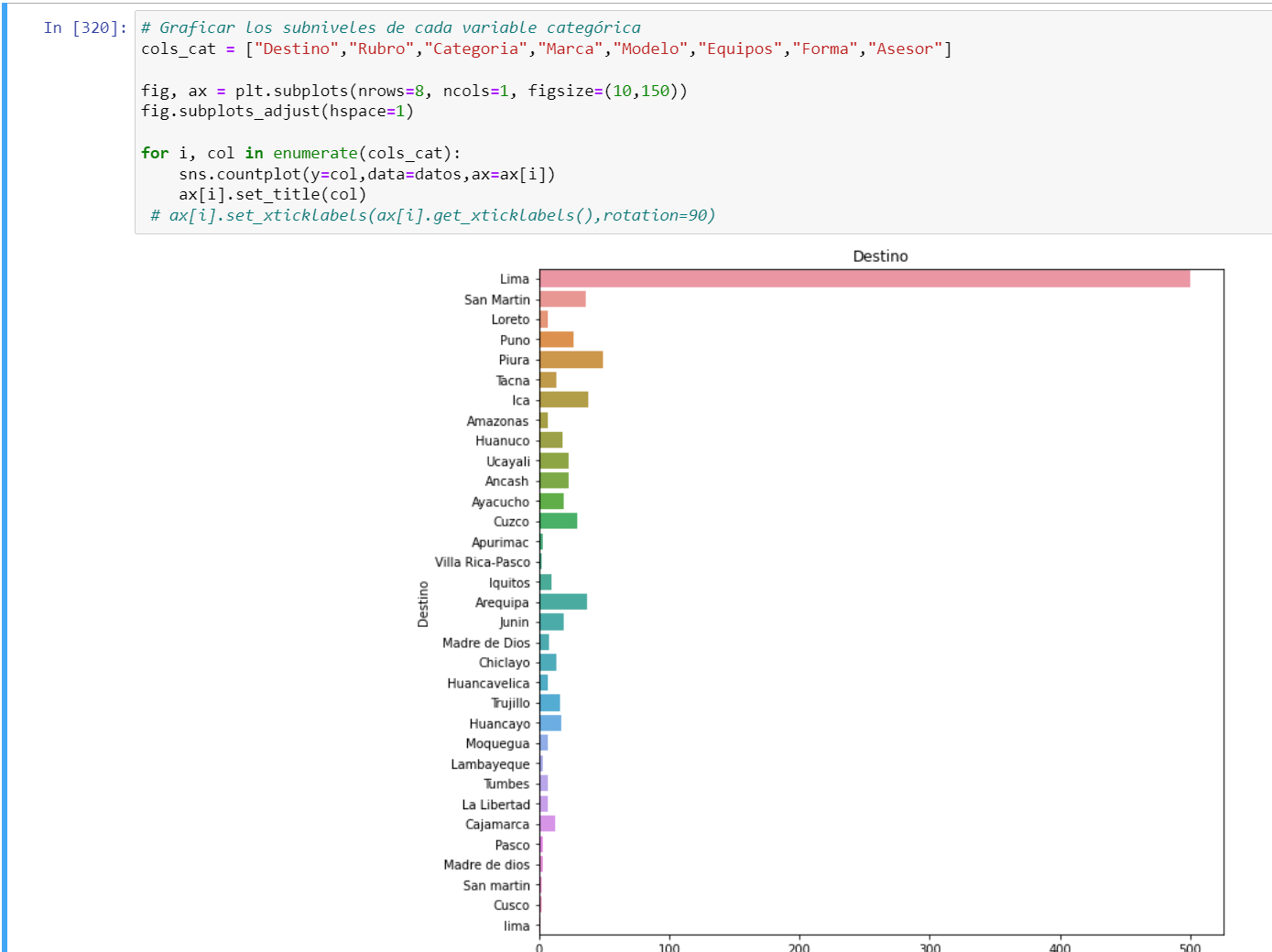


Fuente: Elaboración propia



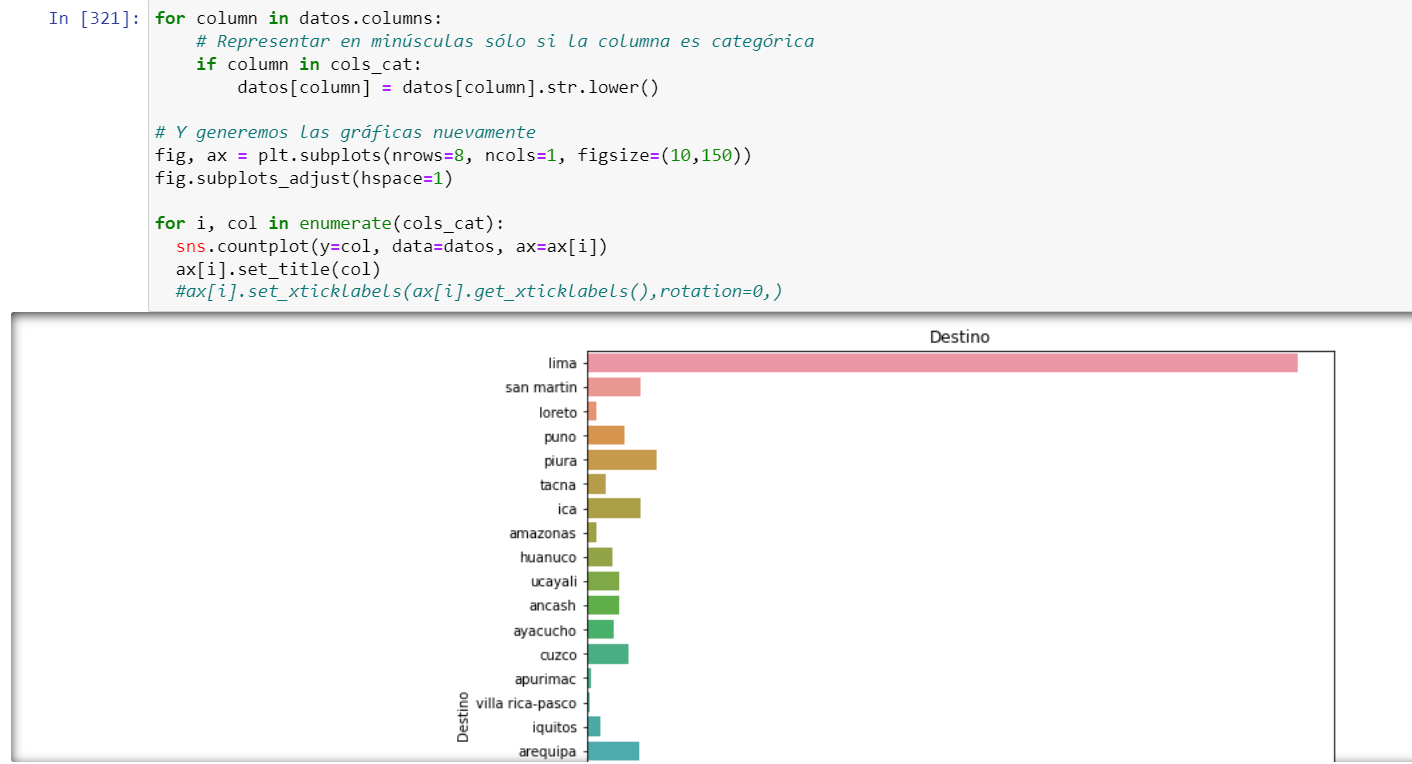
Fuente: Elaboración propia

En este paso se proceder hacer las gráficas para poder ver si hay erros o similitudes de las variables que están en string ,debido a que hay datos con mal digitación,etc.

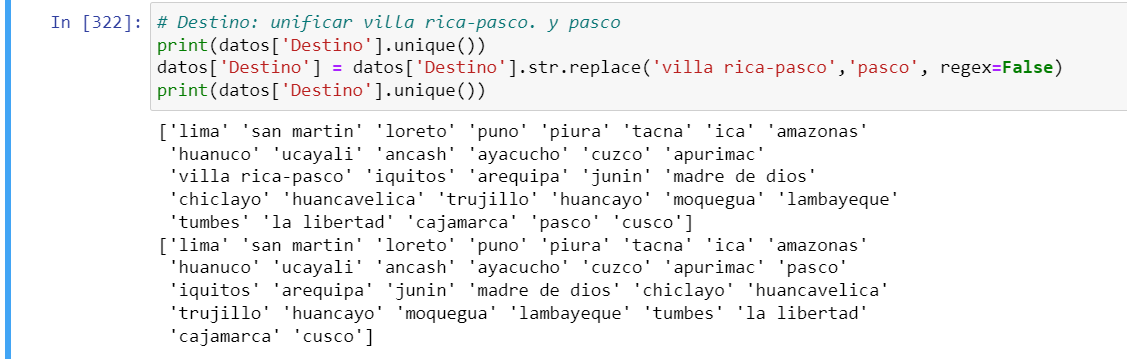


Fuente: Elaboración propia

En este paso se observa que se



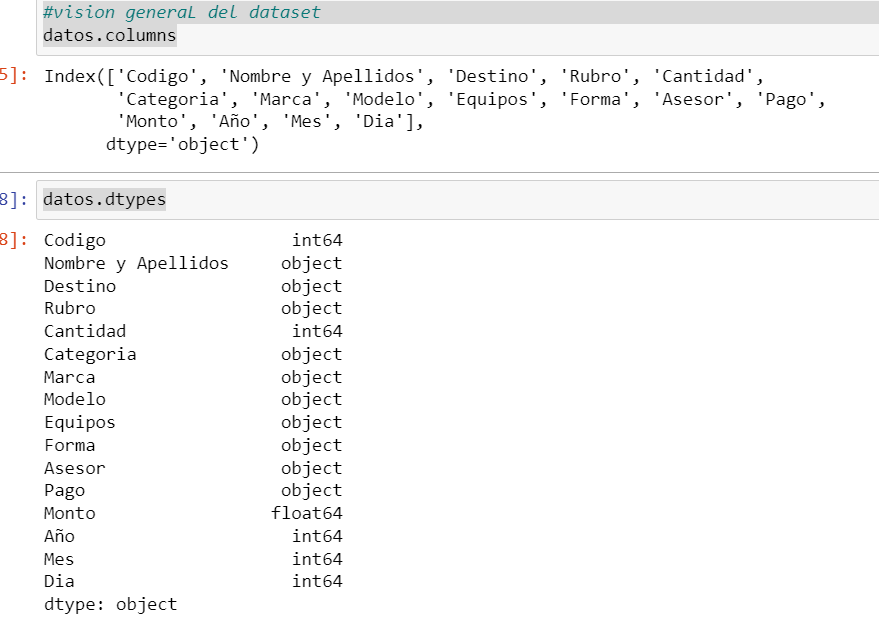
Fuente: Elaboración propia





Fuente: Elaboración propia

exploración de datos

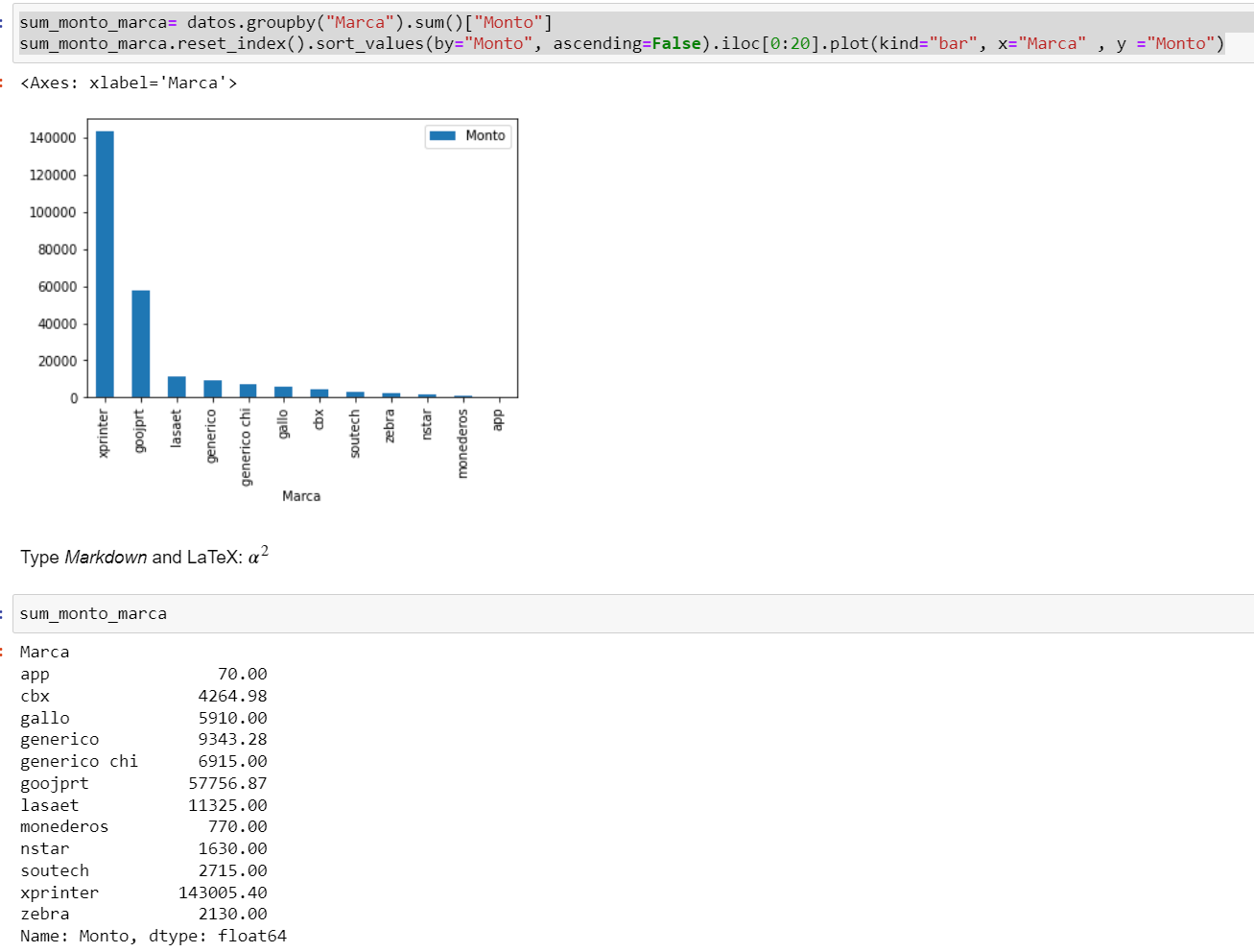


Fuente: Elaboración propia

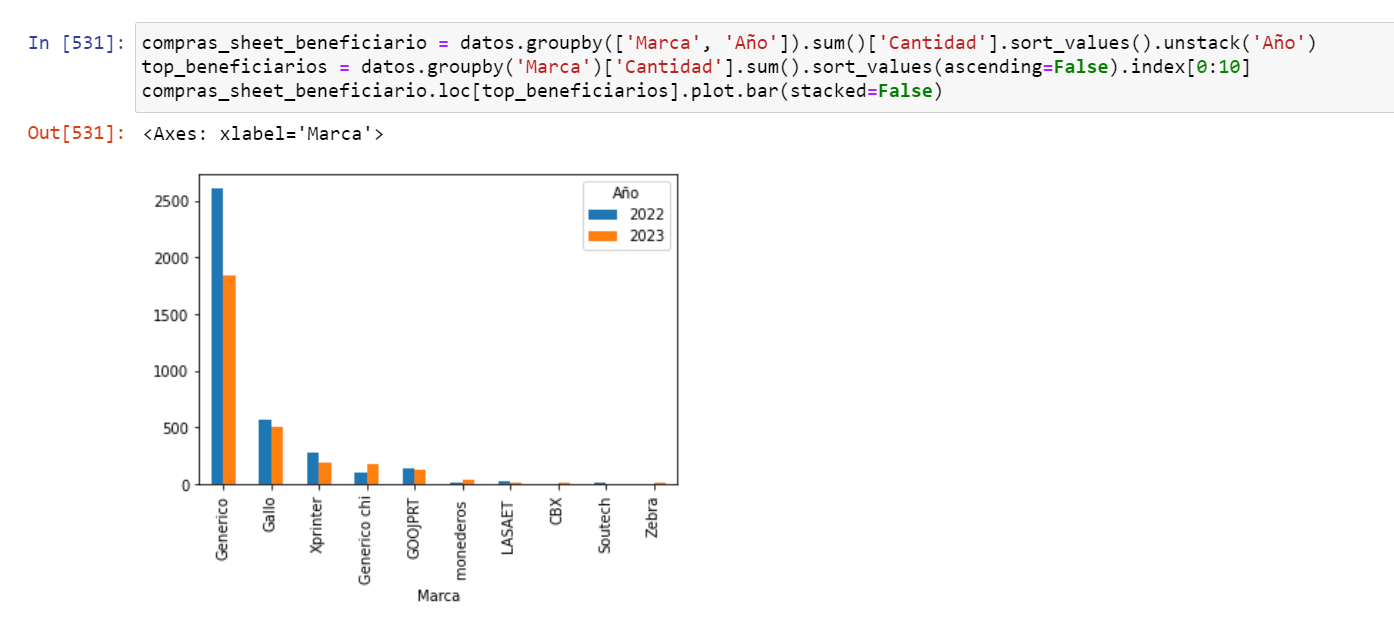


Fuente: Elaboración propia

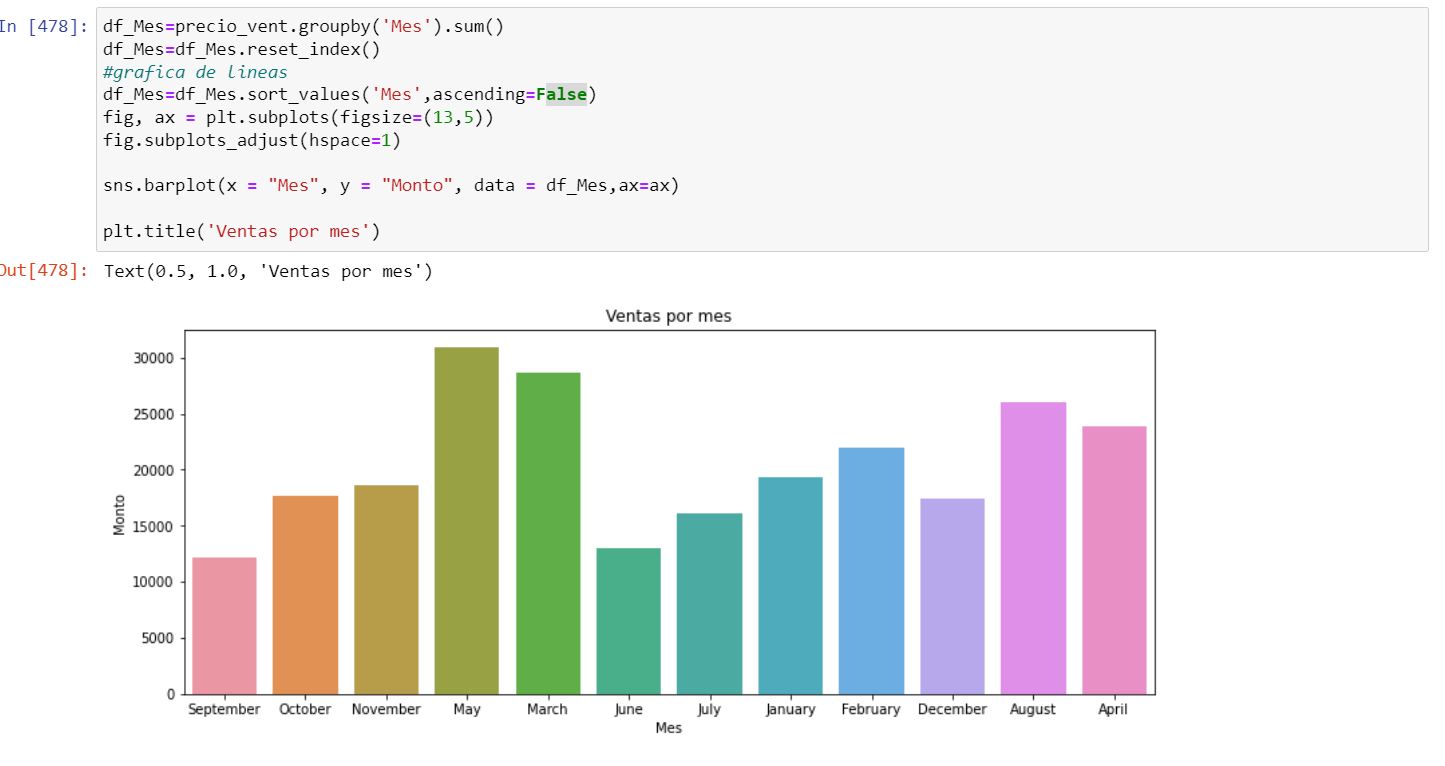
En este grafico de barras se obtuvo el la conclusión de que las ventas realizadas, con ello él mayor producto vendido es la marca Xprinter llegando a superar los 14000 soles y en segundo lugar la marca GOOJPRT



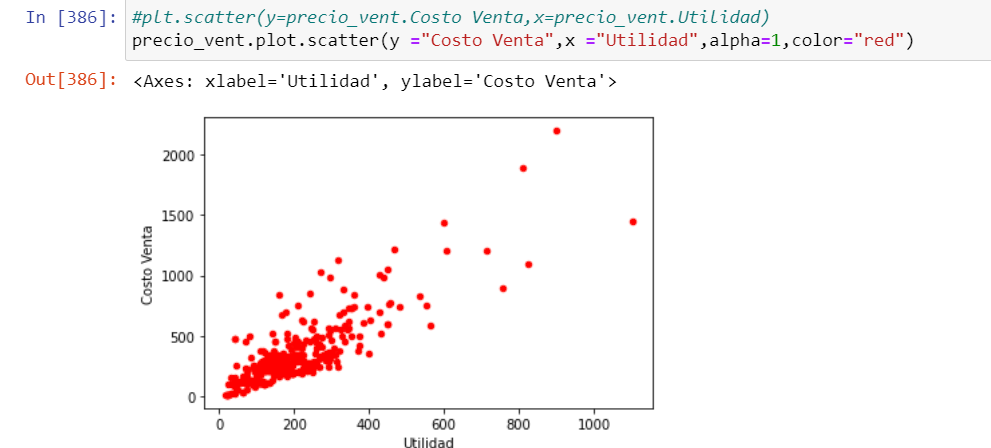
Fuente: Elaboración propia



Fuente: Elaboración propia



Fuente: Elaboración propia



Fuente: Elaboración propia

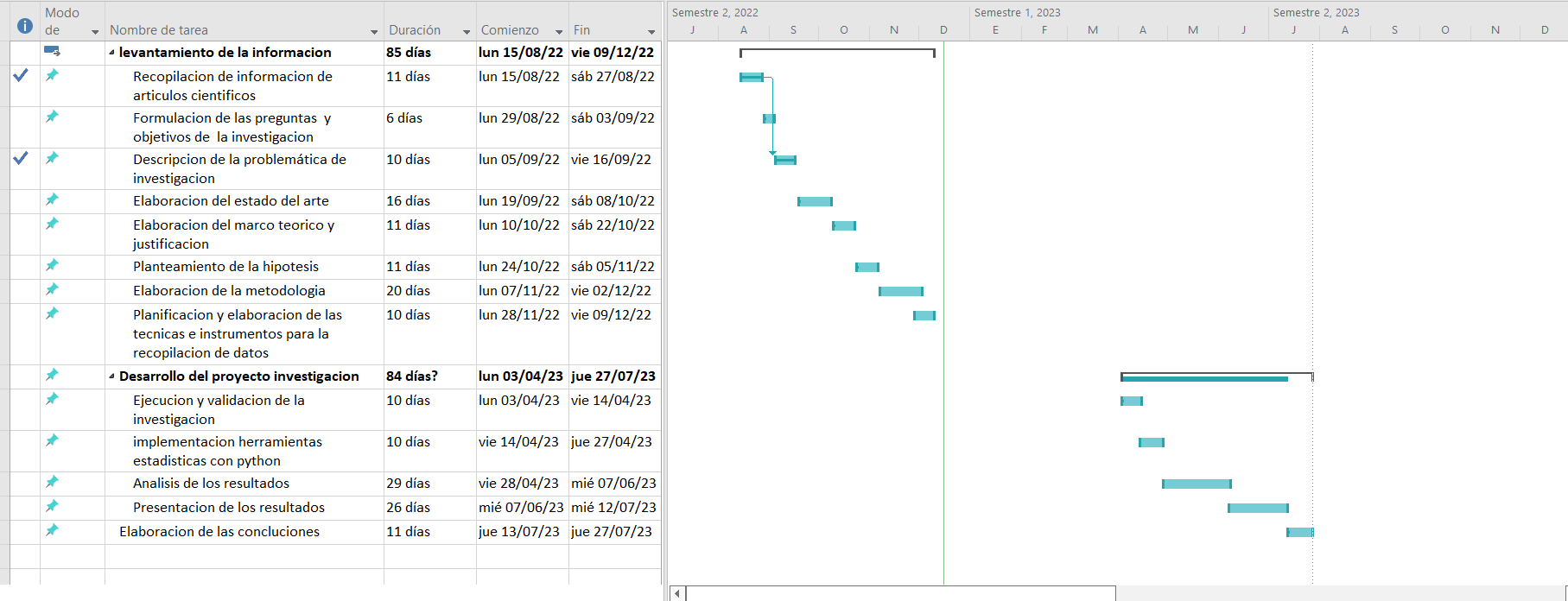
8. Resultados y Discusión

Data Science es una metodología que hoy en día su aplicación mejora la toma de decisiones en el área de ventas, según nuestros resultados. Ya que nos ayuda a obtener información de los datos y aplicarlo de la mejor manera en la toma de decisiones (Según P. Morgan, 2016 FUENTE 9). La información obtenida con la aplicación de Data Science nos da una vista de los productos que más se están vendiendo y con ello se puede tomar mejores decisiones en el ares de ventas.

Concluciones

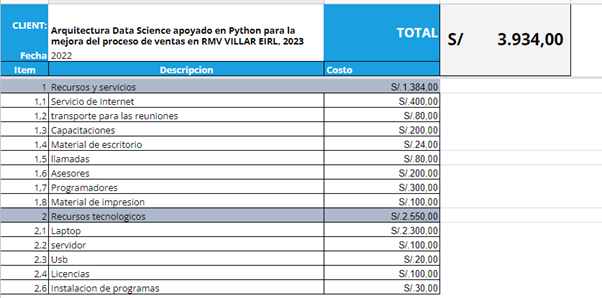
* Después de realizar la investigación sobre el impacto de Data Science para la toma de decisiones en el área de ventas, podemos concluir que la calidad de la información dentro de la toma de decisiones es muy importante. Ya que la calidad de la información nos ayuda a tener mejores resultados y como aplicarlos. Concluimos que existe una correlación significativa de la aplicación de Data Science y la calidad de la información para la toma de decisiones en el área de ventas de la empresa RMV VILLAR.
* Con los datos obtenidos con la parte estadística usando lupyter notebook con python dieron resultados que deberían de implementar acciones de marketing de acuerdo Alos equipos menos vendidos y reforzar Alos que fueron los más vendidos como xprinter que llego a recaudar un aproximado de 140000 soles
* Con la aplicación de data science se puedo llegar una idea clara de los producto que se deberían comprar para esta temporada ya que con el análisis de datos nos muestra que este mes baja las ventas por ende se debería de hacer una estrategia para que mejore las ventas de los productos mas vendidos

**9.Cronograma:**



**Figura 6: Elaboración propia link:https://drive.google.com/file/d/1ywY5rR5L2xTA37ziCZXMpAyUKIciKmY5/view?usp=sharing**

**10.** **Presupuesto**

****

**Cuadro 1: elaboración propia**

Mediante el cuadro 1 se mostrará lo que se está gastando para poder llevar a cabo la presente investigación, como reuniones, llamadas, etc.

**11. Bibliografía**

[1] J. P. Rave, G. J. Álvarez, and F. G. Echavarría, “Automated multitasking methodology for studying business indicators using data science,” *Comput. y Sist.*, vol. 24, no. 3, pp. 933–956, 2020, doi: 10.13053/CYS-24-3-3040.

[2] P. T. Nguyen, K. Shankar, E. A. Fedchenko, R. Novianty, and Rosdiana, “Data science and project engineering management techniques,” *Opcion*, vol. 35, no. Special Issue 19, pp. 2899–2921, 2019.

[3] J. Guaña-Moya, N. Salgado-Reyes, and C. Escobar-Terán, “Computing and data science using deep learning | La informática y la ciencia de datos utilizando Deep learning,” *RISTI - Rev. Iber. Sist. e Tecnol. Inf.*, vol. 2022, no. E49, pp. 116–127, 2022.

[4] A. Anton-Bravo and A. Serrano-Tellería, “Innovación en la docencia del periodismo a través de la ciencia de datos,” *Eur. Public Soc. Innov. Rev.*, vol. 6, no. 1, pp. 70–85, 2021, doi: 10.31637/epsir.21-1.6.

[5] G. R. Vivian and C. R. Cervi, “Utilizando Técnicas de Data Science para Definir o Perfil do Pesquisador Brasileiro da Área de Ciência da Computação,” *Proc. ERBD*, pp. 108–117, 2016.

[6] G. Burk, Scott;Sweenor, David E.;Miner, “It’s all analytics - Part II: Designing an integrated AI, analytics, and data science architecture for your organization,” *Book*, pp. 1–296, 2021, doi: 10.4324/9780429343957.

[7] S. Raschka, J. Patterson, and C. Nolet, “Machine learning in python: Main developments and technology trends in data science, machine learning, and artificial intelligence,” *Inf.*, vol. 11, no. 4, p. 193, Apr. 2020, doi: 10.3390/info11040193.

[8] J. Pérez-Ortega, N. N. Almanza-Ortega, K. Torres-Poveda, G. Martínez-González, J. C. Zavala-Díaz, and R. Pazos-Rangel, “Application of Data Science for Cluster Analysis of COVID-19 Mortality According to Sociodemographic Factors at Municipal Level in Mexico,” *Mathematics*, vol. 10, no. 13, 2022, doi: 10.3390/math10132167.

[9] P. Morgan, *DATA ANALYSIS FROM SCRATCH WITH PYTHON St ep By Step Guide*, vol. 110, no. 9. 2016.

[10] L. F.-R. Pérez and Á. R. Blasco, “A Data Science Approach to Cost Estimation Decision Making - Big Data and Machine Learning | Un enfoque de ciencia de datos para la toma de decisiones en la estimación de costes - Big Data y aprendizaje automático,” *Rev. Contab. Account. Rev.*, vol. 25, no. 1, pp. 45–57, 2022,doi:10.6018/RCSAR.401331.

[11] Decision Making - Big Data and Machine Learning | Un enfoque de ciencia de datos para la toma de decisiones en la estimación de costes - Big Data y aprendizaje automático,” *Rev. Contab. Account. Rev.*, vol. 25, no. 1, pp. 45–57, 2022.

[12] Oppel Andy and Sheldon Robert, *Fundamentos\_De\_Sql\_3Edi\_Oppel*, 3rd ed., vol. 1. México, 2009.

[13] M. López, C. Lagunes, and S. Herrera, “Excel como una herramienta asequible en la enseñanza de la Introducción,” *Teoría la Educ. Educ. y Cult. en la Soc. la Inf.*, vol. 7, no. 1, pp. 1–10, 2006, [Online]. Available:<http://www.redalyc.org/articulo.oa?id=201021084007>

[14] K. Jantakun, T. Jantakun, and T. Jantakoon, “The Architecture of System for Predicting Student Performance Based on Data Science Approaches (SPPS-DSA Architecture),” *Int. J. Inf. Educ. Technol.*, vol. 12, no. 8, pp. 778–785, 2022, doi: 10.18178/ijiet.2022.12.8.1684.

[15] M. Á. González and M. O. Soler, “El proceso de toma de decisiones profesionales a través del coaching,” *Electron. J. Res. Educ. Psychol.*, vol. 7, no. 2, pp. 877–900, 2009.

[16] C. Janiesch, P. Zschech, and K. Heinrich, “Machine learning and deep learning,” *Electron. Mark.*, vol. 31, no. 3, pp. 685–695, 2021, doi: 10.1007/s12525-021-00475-2.

[17] R. Hernández, C. Fernández, and P. Baptista, “Definición del alcance de la investigación que se realizará: exploratorio, descriptivo, correlacional o explicativo,” *Metodol. la Investig.*, vol. 6, pp. 88–101, 2014, [Online]. Available:<http://metabase.uaem.mx/bitstream/handle/123456789/2792/510_06_color.pdf?sequence=1>

[18] F. A. Sánchez Flores, “Fundamentos Epistémicos de la Investigación Cualitativa y Cuantitativa: Consensos y Disensos,” *Rev. Digit. Investig. en Docencia Univ.*, vol. 13, pp. 101–122, 2019, doi: 10.19083/ridu.2019.644.

[19] M. C. Useche, W. Artigas, B. Queipo, and É. Perozo, *Técnicas e instrumentos de recolección de datos cuali-cuantitativos*, Primera ed. Colombia: Book, 2019. [Online]. Available: https://repositoryinst.uniguajira.edu.co/flip/index.jsp?pdf=/bitstream/handle/uniguajira/467/88. Tecnicas e instrumentos recolección de datos.pdf?sequence=1&isAllowed=y

[20] I. R. Rojas-Crotte, “Elementos para el diseño de técnicas de investigación: Una propuesta de definiciones y procedimientos de la investigación científica,” *Tiempo Educ.*, vol. 12, no. 24, pp. 277–297, 2011, [Online]. Available:<http://www.redalyc.org/pdf/311/31121089006.pdf>

[21] V. D. Sousa, M. Driessnack, I. Amélia, and C. Mendes, “Revisión de diseños de investigación resaltantes para enfermería. Parte 1: Diseños de investigación cuantitativa.,” *Rev Latino-am Enferm.* , vol. 15, no. 3, pp. 1–6, 2007, [Online]. Available: [www.eerp.usp.br/rlae](http://www.eerp.usp.br/rlae)

[22] M. Ato, J. J. López, and A. Benavente, “Un sistema de clasificación de los diseños de investigación en Psicología [A classification system of research designs in Psychology],” *An. Psicol.*, vol. 29, no. 3, pp. 1038–1059, 2013.

[23] Rollins John B., “Metodología Fundamental para la Ciencia de Datos,” *IBM Anal.*, pp. 1–6, 2015, [Online]. Available:<https://www.ibm.com/downloads/cas/6RZMKDN8>

[24] J. Ardila, N. Rodríguez, and F. Gil, “Población y muestreo,” *Epidemiol. clínica Investig. clínica*, pp. 129–139, 2004, [Online]. Available: http://www.medicapanamericana.com/Libros/Libro/3848/Epidemiologia-Clinica.html

12. **ANEXOS**

**MATRIZ DE CONSISTENCIA**

**Título: Aplicación de Data Science para la toma de decisiones en el área de ventas en RMV VILLAR, 2023**

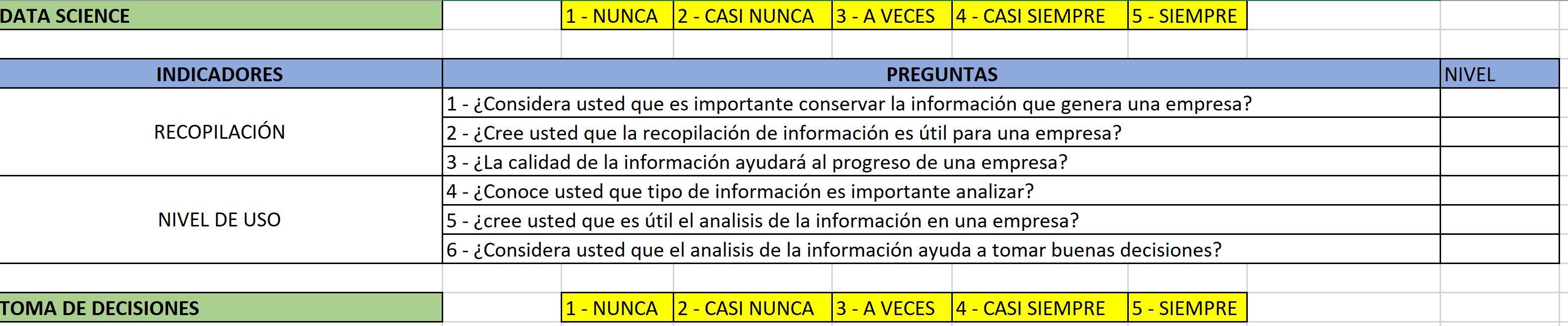
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **PROBLEMA** | **OBJETIVOS** | **HIPÓTESIS** | **VARIABLES** | **DIMENSIONES** | **INDICADORES** | **METODOLOGÍA** |
| **GENERAL:**  ¿Cuál es el impacto de la aplicación de Data Science para la toma de decisiones en el área de ventas en RMV VILLAR, 2023? | **GENERAL:**  Medir el impacto de la aplicación de Data Science para la toma de decisiones en el área de ventas en RMV VILLAR, 2023. | **GENERAL:**  La aplicación de Data Science mejora significativamente la toma de decisiones en el área de ventas en RMV VILLAR, 2023. | **VARIABLE 1:**  Data Science | * Calidad de la información * Análisis de información | * Recopilación * Disponibilidad * Nivel de uso * Evaluación * Filtrar los datos | **Enfoque:**  Cuantitativo  **Diseño:**  No Experimental  De corte Transversal  **Alcance:**  Explicativo  **Población:**  500 clientes  **Muestra:**  252 clientes |
| **PROBLEMAS ESPECÍFICOS:**  **PE1**: ¿Cuál es el impacto de la aplicación de Data Science en la calidad del servicio de venta para la toma de decisiones en el área de ventas en RMV VILLAR, 2023?  **PE**2: ¿Cuál es el impacto de la aplicación de Data Science en la mejora de la eficiencia para la toma de decisiones en el área de ventas en RMV VILLAR, 2023?  **PE**3: ¿Cuál es el impacto de la aplicación de Data Science en la optimización de procesos para la toma de decisiones en el área de ventas en RMV VILLAR, 2023? | **OBJETIVOS ESPECÍFICOS:**  **OE1**: Medir el impacto de la aplicación de Data Science en la calidad del servicio de venta para la toma de decisiones en el área de ventas en RMV VILLAR, 2023.  **OE2:** Medir el impacto de la aplicación de Data Science en la mejora de la eficiencia para la toma de decisiones en el área de ventas en RMV VILLAR, 2023.  **OE3:** Medir el impacto de la aplicación de Data Science en la optimización de procesos para la toma de decisiones en el área de ventas en RMV VILLAR, 2023. | **HIPÓTESIS ESPECÍFICAS:**  **HE1:** La aplicación de Data Science mejora significativamente la calidad de servicio de venta para la toma de decisiones en el área de ventas en RMV VILLAR, 2023.  **HE2:** La aplicación de Data Science mejora significativamente la eficiencia para la toma de decisiones en el área de ventas en RMV VILLAR, 2023.  **HE3:** La aplicación de Data Science mejora significativamente la optimización de procesos para la toma de decisiones en el área de ventas en RMV VILLAR, 2023. | **VARIABLE 2:**  Toma de decisiones | * Calidad del servicio de venta * Mejora de la eficiencia * Optimización de Procesos | * Tiempo de Demora de las ventas * Reducir errores * Número de Ventas realizados * Agilidad en las ventas |

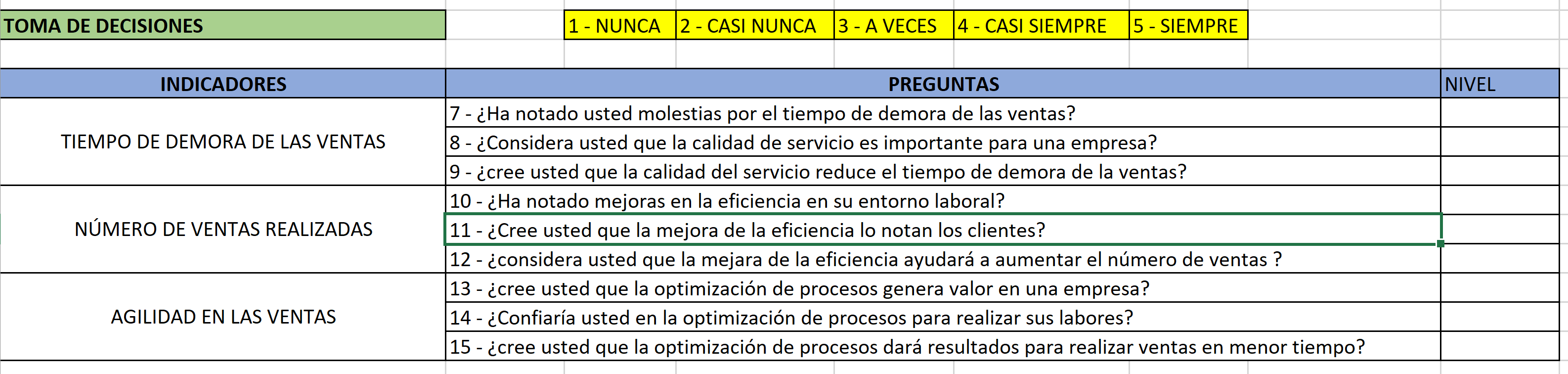
**MATRIZ DE OPERACIONALIZACIÓN DE LAS VARIABLES**

**Título: Aplicación de Data Science para la toma de decisiones en el área de ventas en RMV VILLAR, 2023**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **VARIABLES** | **DEFINICIÓN CONCEPTUAL** | **DEFINICIÓN OPERACIONAL** | **DIMENSIONES** | **INDICADORES** | **METODOLOGÍA** |
| Data Science | Data Science puede ser nombrada como un campo interdisciplinario que usa esquemas científicos y algoritmos muy útiles para poder conseguir juicios críticos y conocimientos útiles tras un análisis de los datos. En este proceso es importante la información obtenida, pues gracias a esta se va a realizar un análisis más profundo y obtener resultados más útiles. Hoy en día Data Science está siendo aceptada como un nuevo paradigma en múltiples campos de la ciencia empírica como son las finanzas, economía o las operaciones. Se considera como un nuevo paradigma del progreso científico [9]. | Se tomará para la información todos los datos de los clientes de la empresa RMV VILLAR durante el mes de noviembre del 2022, donde se propondrá la aplicación de Data Science para mejorar el área de ventas. | * Calidad de la información * Análisis de información | * Cantidad de datos valiosos del cliente * Nivel de uso * Evaluación * Filtrar los datos | **Enfoque:**  Cuantitativo  **Diseño:**  No Experimental  De corte Transversal  **Alcance:**  Explicativo  **Población:**  500 clientes  **Muestra:**  252 clientes |
| Toma de decisiones | Hoy en día en el mundo entero es de suma importancia la toma de decisiones, pues gracias a ello se puede definir el buen futuro de un negocio o una organización, también en la vida diaria se ha vuelto en un componente fundamental, de igual manera en el ámbito profesional, social, etc. La toma de decisiones ha escalado bastante en los negocios y principalmente en el proceso de ventas, pues una buena toma de decisiones podría mejorar este proceso y obtener mayores ventas para la empresa. En pocas palabras la toma de decisiones se trata de un proceso en donde se tiene varias opciones y se tiene que elegir a una de ellas. Para tener una mejor toma de decisiones se debe respaldar en un análisis previo de estas opciones para tomar la decisión correcta. [15] | * Calidad del servicio de venta * Mejora de la eficiencia * Optimización de Procesos | * Tiempo de Demora de las ventas * Número de Ventas realizados * Agilidad en las ventas |

**Cuestionario**





**Base Información**

