

**ESPECIALIZACIÓN EN PRICING ANALYTICS**

“Fijación y optimización de precios”

**Integrantes Grupo 3**:

*Cornelio Portella, Giovana Edeliza*

*Delgado Ramirez, Emma Gabriela*

*Dominguez Avila Gino Francisco*

**10 de enero del 2023**

**INDICE**

Tabla de contenido

1. **Introducción3**

1.1 Datos Generales2

1.2. Breve Reseña Histórica3

1.3. Misión y Visión3

1.4. Productos4

1.5. Alcance del Proyecto5

1. **La Propuesta8**
   1. Detalle de la Propuesta8
   2. Objetivo Principal y Específicos9
   3. Diseño de la Solución9
   4. Flujo de Datos 9
   5. Entregables14
   6. Valor Esperado 14

**PROPUESTA TÉCNICA PARA PROYECTO DE PRICING ANALYTICS EN LA COMERCIALIZACIÓN DE HUEVOS DE LA EMPRESA “XYZ” EN UNA CADENA DE SUPERMERCADO**

1. **Introducción**
   1. **Datos Generales**

A modo de ejemplo, se va a considerar a la empresa XYZ cuya su principal función es de producir la principal productora y distribuir huevos a nivel nacional.

* **RUC:** xxxxxxxxxx
* **Razón Social:** XYZ
* **Giro:** productora y comercializadora de huevos a nivel nacional.
* **Tamaño de la empresa:** empresa grande con más de 3000 trabajadores.

* 1. **Breve Reseña histórica**

La empresa XYZ inició sus actividades al inicio con solo 1000 gallinas ponedoras. Despues de algunos años establece una relación con los supermercados.

Comercializan huevos enriquecidos con Omega 3, así como huevos de corral, doble yema y blancos.

En los días actuales, cuenta con más de 6 millones de aves, brindando trabajo a más de 3000 familias en sus 4.5 mil Ha que posee alrededor de Latam.

* 1. **Misión y Visión**

**Misión:**

“Fomentar el desarrollo, con entusiasmo y crecimiento, amando la tierra y sus productos avícolas.”

**Visión:**

“Contribuir ofreciendo productos nutritivos y saludables de manera eficiente y sostenible”

* 1. **Productos:**

La empresa XYZ gracias a su amplia variedad de huevos:

***A) Huevos Pardos***

* Fuente de proteínas, vitaminas y minerales.
* Código de trazabilidad (permiten conocer el histórico, la ubicación y la trayectoria del huevo).
* Empaque ecológico, seguro y resistente. Fácil de almacenar y apilar.

***B) Huevos Pardos Jumbo***

* 10 % más grandes.
* Fuente de proteínas, vitaminas y minerales.
* Código de trazabilidad (permiten conocer el histórico, la ubicación y la trayectoria del huevo).
* Empaque seguro y resistente. Fácil de almacenar y apilar.

***C) Huevos Super Jumbo***

* 20 % más grandes.
* Fuente de proteínas, vitaminas y minerales.
* Código de trazabilidad (permiten conocer el histórico, la ubicación y la trayectoria del huevo).
* Empaque seguro y resistente. Fácil de almacenar y apilar.

***D) Huevos Blancos***

* Fuente de proteínas, vitaminas y minerales.
* Fecha de vencimiento y Código de trazabilidad (permiten conocer el histórico, la ubicación y la trayectoria del huevo).
* Empaque ecológico, seguro y resistente. Fácil de almacenar y apilar.

***E) Huevos de Corral***

* Yema de color más intenso
* Gallinas criadas libres de jaulas (Tiene sello de Certificación de Bienestar Animal)
* Fuente de vitaminas, proteínas y minerales.
* Código de trazabilidad (permiten conocer el histórico, la ubicación y la trayectoria del huevo).
* Empaque seguro y resistente. Fácil de almacenar y apilar.

***F) Huevos Orgánicos***

* Gallinas criadas libres de jaulas, su alimentación es orgánica (Tiene sello de Certificación de RTPO: Reglamento técnico para Productos Orgánicos)
* Fuente de vitaminas, proteínas y minerales.
* Código de trazabilidad (permiten conocer el histórico, la ubicación y la trayectoria del huevo).
* Empaque seguro y resistente. Fácil de almacenar y apilar.

***G) Huevos de Codorniz***

* Contiene Omega 3
* Fuente de vitaminas, proteínas y minerales.
* Empaque seguro y resistente. Fácil de almacenar y apilar.

* 1. **Alcance**

El proyecto de pricing de la empresa XYZ se desarrollará del producto huevos Pardos pack de 30 unidades de una cadena de supermercados.

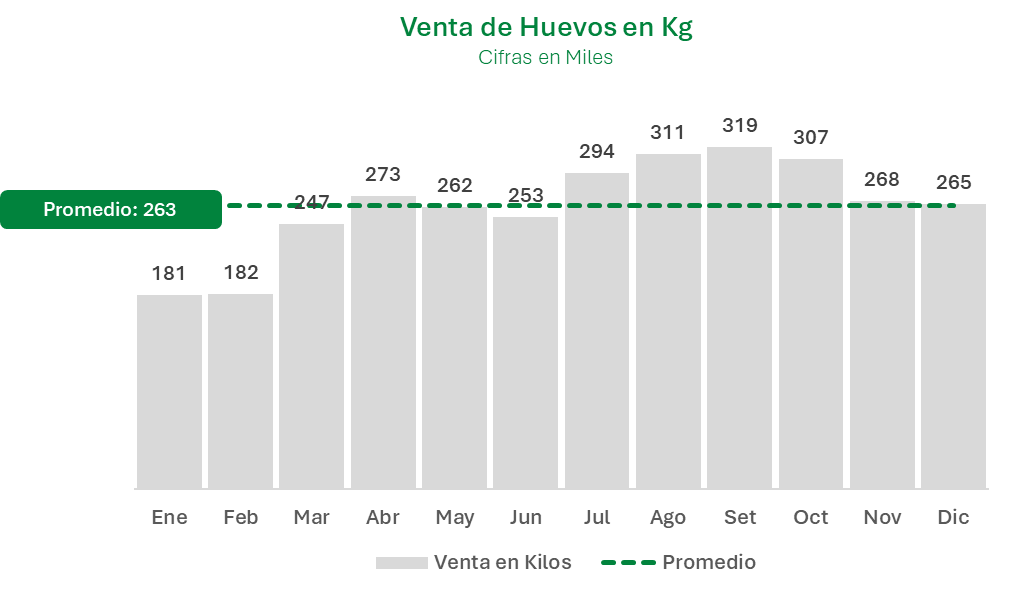
El modelo de pricing deberá determinar el precio optimo del producto Huevos Pardos XYZ y del Huevo Pardo de la Marca Blanca, obteniendo el mayor margen para cada producto y para cada zona.

|  |  |
| --- | --- |
| **N° Registros\*** | 12,122 registros |
| **Fecha de la Información** | Enero – diciembre 2023 |
| **N° regiones de venta** | Capital y 17 regiones de provincia |
| **Precio de Venta** | 7.24 - 13.68 |
| **Ventas Anuales en soles** | 30.9 MM |
| **Ventas Anuales en KG** | 3.2 MM |

\*Datos transformados, se desea evidenciar el proceso seguido para fines académicos.

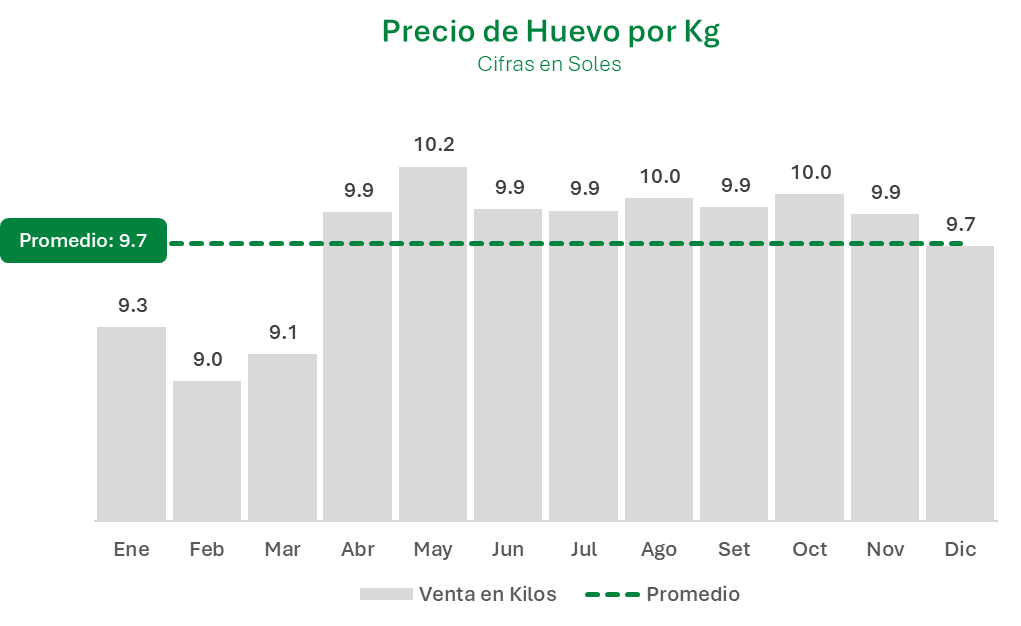
Las ventas promedio por mes es de un aproximado de 263 mil kilogramos, los cuales presentan una estacionalidad marcada respecto al ciclo escolar, siendo los meses de enero y febrero, periodo de vacaciones donde las ventas se reducen un 31% respecto al promedio anual.

De igual manera, los meses de julio a octubre son los meses de mayor consumo que se registra durante el año con un consumo promedio de 308mil kilogramos por mes (17% más que el promedio anual)



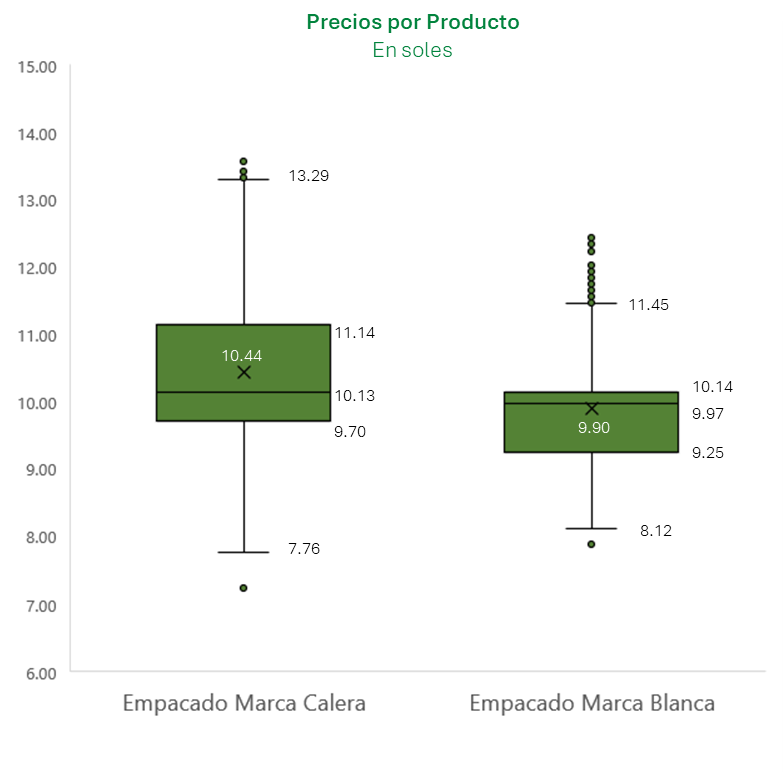
*Fuente: Elaboración propia*

El precio promedio de venta del año 2023 fue de 9.7soles por kilogramo. El precio también muestra una estacionalidad de menor precio durante el primer trimestre debido a una mayor producción de las gallinas ponedoras por la temporada de verano, lo cual hace que el precio se reduzca.



*Fuente: Elaboración propia*

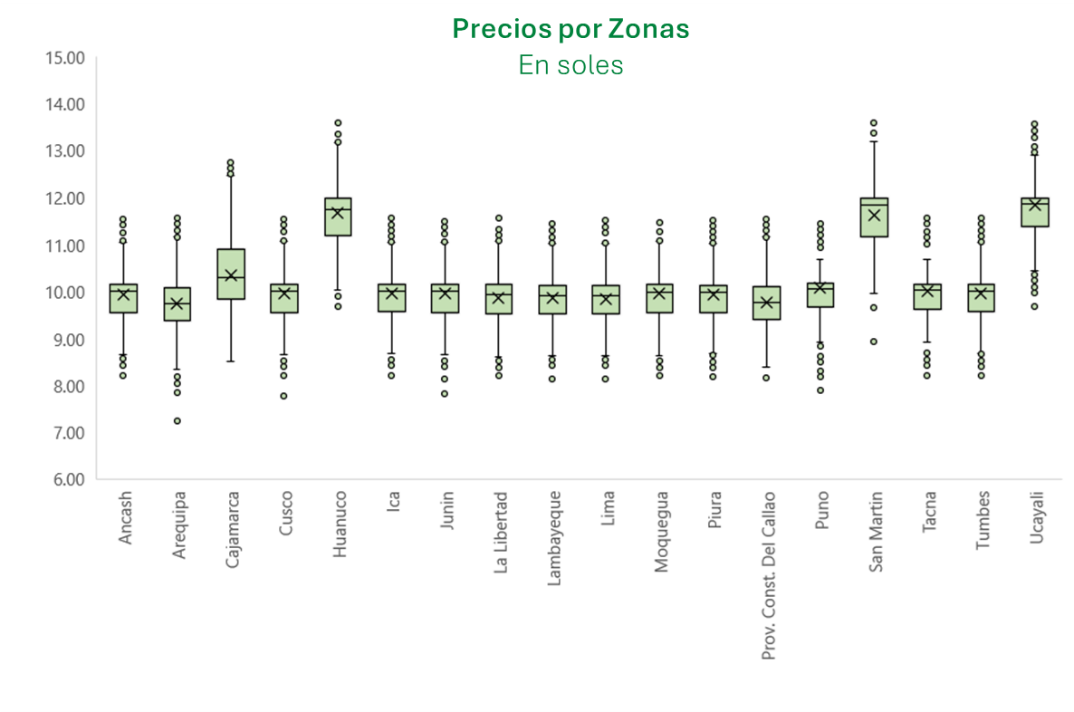
En caso del precio por producto, en caso del huevo pardo de Marca Blanca el promedio es de 9.90 y sus rangos de precios es más corto, oscilan durante el año entre 8.12 y 11.45 soles. En caso del huevo pardo de XYZ el promedio es de 10.44 (5% adicional) y su rango de precios es más amplio (oscila entre 7.76 a 13.29 soles)



XYZ Marca Blanca

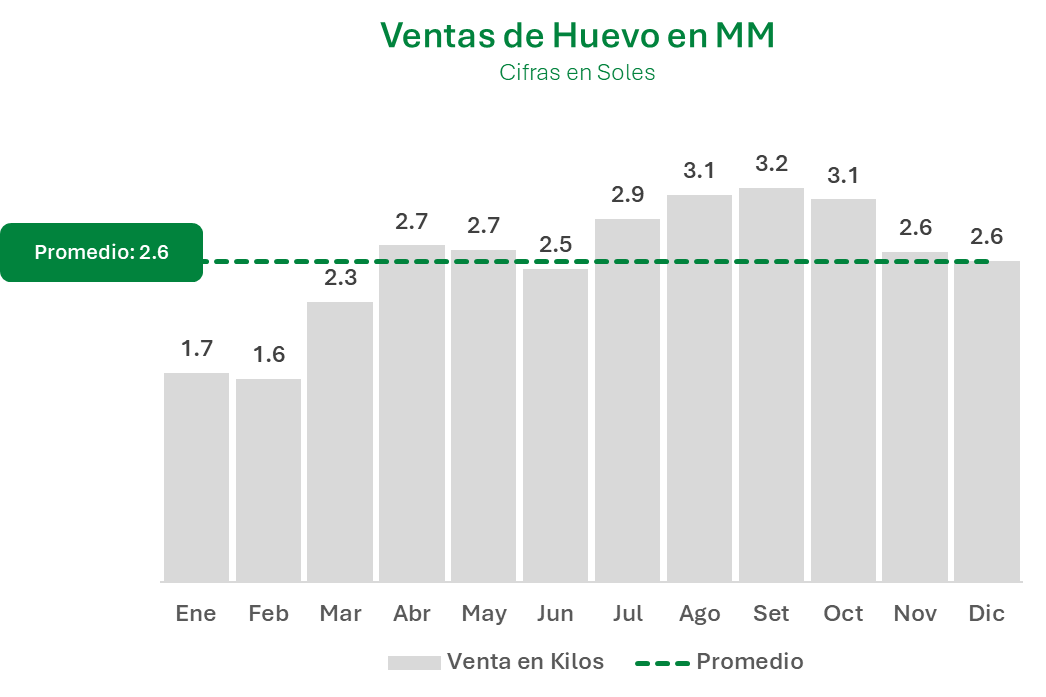
*Fuente: Elaboración propia*

Asimismo, el precio muestra un diferencial es 4 regiones: Huanuco, San Martín y Ucayali y Cajamarca donde el precio es superior a las otras 14 regiones. Esto principalmente se debe al incremento del costo para poder abastecer del producto a zonas de poco acceso logístico.

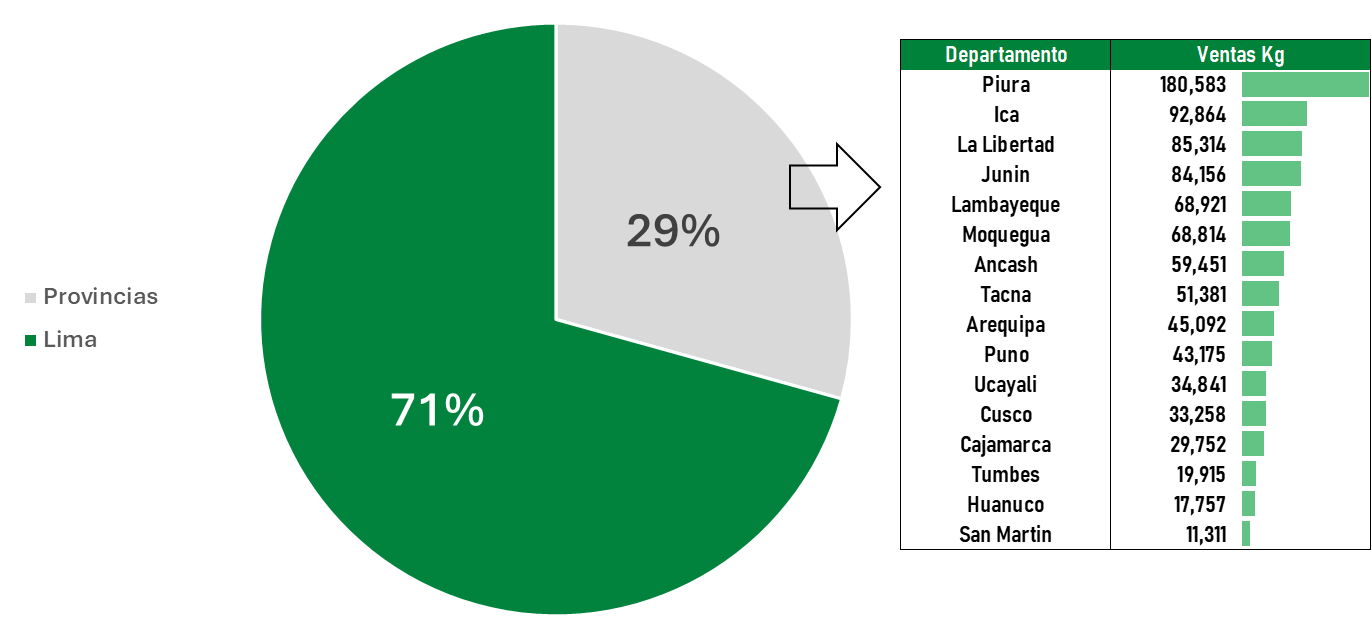


*Fuente: Elaboración propia*

Los ingresos por ventas de la línea de supermercado registran un promedio de 2.6 millones de soles promedio por mes en el año 2023, y con un ingreso anual por ventas de 30.9MM de soles.



En cuanto al canal de supermercados, el 71% de las ventas se realizan en Lima y el 33% en las 17 otras regiones del país donde las principales son: Piura (6%), Ica (3%), La Libertad (3%) y Junín (3%)



1. **Propuesta:** 
   1. **Detalle de la propuesta**

En el presente trabajo se va a obtener mediante el uso de un modelo de regresión (elasticidad constante) para obtener una función de la demanda de los productos de huevos de la marca XYZ vs Línea Blanca en las presentaciones de pack x 30 unid de un Supermercado en diferentes zonas de Lima y provincias y en función al precio, fecha de venta, costo, descuentos entre otras variables .

Nuestra propuesta técnica para el proyecto de Pricing Analytics en la venta de huevos de la marca “XYZ” tiene como objetivo principal la fijación y optimización de precios para maximizar los ingresos y/o rentabilidad en la comercialización de los huevos en pack de 30 unidades. Para lograr esto, utilizaremos técnicas como Clustering, Regresión (Estimación de la función de demanda) y Optimización de precios usando Python.

En primer lugar, realizaremos un análisis exploratorio para conocer la distribución de los datos, detectar valores nulos y outliers, y posible correlaciones entre las variables; posteriormente se realizará un análisis de Clustering para segmentar las ventas del Supermercado (K-means) , según su ubicación, comportamiento de compra, días de la semana( posibles variantes) y preferencias(entre marca blanca o Calera ). Esto nos permitirá identificar patrones y características específicas de cada grupo, lo que nos ayudará a definir estrategias de precio más efectivas para cada segmento.

Luego, se estimará la función de demanda utilizando el modelo de regresión de elasticidad constante. A través de este proceso, seremos capaces de entender cómo las variables clave: el precio, la promoción, producto línea blanca y la disponibilidad, afectan la demanda de los huevos.

Una vez que tengamos la estimación de la función de demanda, utilizaremos técnicas de “Optimización de Precios” para determinar los precios óptimos para cada segmento de territorios y determinar la frontera eficiente de cada producto. Esto implica considerar los costos de producción, la elasticidad de la demanda y otros factores relevantes para optimizar los precios.

En cuanto al diseño de la solución, implementaremos un modelo de Pricing Analytics que recopilará y analizará los datos relevantes de ventas, precios, promociones, entre otros. Este modelo nos permitirá generar informes y recomendaciones automáticas sobre los precios óptimos para cada producto y segmento de clientes.

* 1. **Objetivo General y Específicos**

**Objetivos General:**

Establecer un modelo de pricing de “XYZ” para una cadena de Supermercado, de productos de huevos pardo.

**Objetivos específicos:**

* Optimizar la estrategia de precios en la venta de huevos pardo en packs de 30 unidades.
* Estimar la función de demanda mediante técnicas de regresión para comprender la relación entre el precio y la cantidad demandada.
* Segmentar por Zonas mediante técnicas de clustering para adaptar estrategias de precios a diferentes perfiles por zona.
* Implementar un modelo de optimización de precios para maximizar ingresos y/o rentabilidad de los productos en estudio.
  1. **Diseño de la Solución:**

1. **ETL y Data Cleaning:** es la obtención y limpieza de los datos.
2. **Análisis Exploratorio de los datos:** Obtención de las principales estadísticas descriptivas, tablas uni y bivariadas, y gráficos de las variables cualitativas y cuantitativas.
3. **Segmentación de Clientes:** Utilización de algoritmos de clustering para identificar grupos homogéneos de regiones basados en comportamientos de compra, preferencias y sensibilidad al precio.
4. **Regresión para Estimación de la Función de Demanda:** Aplicación de modelos de regresión lineal y logística para entender la relación cuantitativa entre el precio de los huevos y la demanda, considerando variables como producción, temporada, eventos especiales, etc.
5. **Optimización de Precios:** Desarrollo de un modelo de optimización de precios que tenga en cuenta los resultados de la regresión y las segmentaciones de clientes para determinar los precios óptimos.
   1. **Flujo de Datos:**



* Recopilación de datos de ventas históricas, precios, y variables contextuales (estacionalidad de la demanda, días hábiles y feriados, etc.)
* Preprocesamiento de datos para limpiar y organizar la información.
* Aplicación de técnicas de clustering
* Estimación de la función de la demanda mediante regresión.
* Integración de resultados en el modelo de optimización de precios.
* Recomendaciones por producto.
  1. **Resumen de las técnicas utilizadas**

A continuación, se mencionan los conceptos y técnicas utilizadas en este estudio - Marco Teórico:

* + 1. **Elasticidad precio:**

La elasticidad precio de la demanda revela cuánto varia la cantidad demandada de un bien o servicio frente a los cambios verificados en su precio. (ECONOMIPEDIA, 2024)

En el mercado existen productos que son sensibles a los cambios del precio, es decir que un cambio en el precio y se disminuye la cantidad demandada de dicho producto , de igual manera una disminución en el precio causará una subida en la cantidad demandada.

Se puede clasificar la elasticidad en 3 tipos principales, demanda elástica unitaria, demanda inelástica y demanda elástica (INTUIT MAILCHIMP, 2024)

**Demanda elástica unitaria:**

La demanda elástica unitaria se produce cuando la elasticidad precio de la demanda es igual a 1. La elasticidad unitaria se produce cuando el cambio de precio es equivalente a la cantidad demandada. En última instancia, el porcentaje de cambio es el mismo en ambos.

**Demanda inelástica:**

La demanda inelástica se produce cuando la elasticidad precio de la demanda es inferior a 1.

Los productos inelásticos suelen ser aquellos que resultan imprescindibles para nuestra vida, como el combustible, los alimentos, el papel higiénico o los medicamentos. En el caso de estos productos, aunque los precios aumenten sustancialmente, la gente seguirá comprándolos. Sin embargo, si el precio cae, la demanda podría aumentar y la oferta general se podría reducir.

**Demanda elástica:**

La demanda elástica es cualquier valor de elasticidad mayor que 1. En los productos elásticos, un pequeño cambio en el precio provoca grandes cambios en la demanda. Normalmente se les considera bienes no esenciales, como ropa, coches, ciertos dispositivos, entretenimiento y artículos de lujo.

* + 1. **ETL y Data Cleaning**

Los procesos ETL (Extracción, Transformación y Carga) son un término estándar que se utiliza para referirse al movimiento y transformación de datos. Se trata del proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos y cargarlos en otra base de datos (denominada data mart o data warehouse) con el objeto de analizarlos. También pueden ser enviados a otro sistema operacional para apoyar un proceso de negocio. (POWER DATA, 2024)

* + 1. **Data Cleaning**

***Antes de la extracción***

Aunque podría entenderse como una acción integrada en la fase de transformación de datos, en la actualidad la tendencia es considerar la limpieza de datos como una fase separada del proceso ETL.

Esta visión corresponde a una concepción más moderna y práctica del proceso. Para ahorrar tiempo y ganar en efectividad es conveniente unificar criterios, por ejemplo, introduciendo “av” en vez de “avenida” en todos los registros de una base de datos de direcciones postales, ANTES de empezar el proceso ETL propiamente dicho.

Tan importante es tener la información consolidada como que todos los datos sean correctos y con una visión única para todos los usuarios. Solo así se pueden lograr unos circuitos de trabajo y análisis de dichos datos realmente óptimos y efectivos. (AKVO, 2024)

***Después de la extracción***

Se debe chequear inconsistencias e inexactitudes en los datos, se puede utilizar la siguiente tabla como guía: acuracidad / exactitud, consistencia, integridad, puntualidad, unicidad y validez. Chequear por cada variable y revisar los valores extremos.

Las acciones más importantes son: tratar los valores perdidos o nulos; tratar los valores extremos y eliminar o corregir inconsistencias. (AKVO, 2024)

* + 1. **Análisis Exploratorio de Datos (EDA)**

El análisis exploratorio de datos (EDA por sus siglas en inglés) implica el uso de gráficos y visualizaciones para explorar y analizar un conjunto de datos. El objetivo es explorar, investigar y aprender, no confirmar hipótesis estadísticas.

El proceso de utilizar resúmenes numéricos y visualizaciones para explorar sus datos e identificar posibles relaciones entre variables se denomina análisis exploratorio de datos (EDA por sus siglas en inglés).

El análisis exploratorio de datos es un proceso de investigación en el que se usan estadísticas de resumen y herramientas gráficas para llegar a conocer los datos y comprender lo que se puede averiguar de ellos. (JMP, 2024)

Con el EDA, se pueden hallar anomalías en los datos, como valores atípicos u observaciones inusuales, revelar patrones, comprender posibles relaciones entre variables y generar preguntas o hipótesis interesantes que se pueden comprobar más adelante mediante métodos estadísticos más formales.

El análisis exploratorio de datos es como el trabajo detectivesco: se buscan claves y pistas que puedan conducir a la identificación de las posibles causas de origen del problema que se intenta resolver. Se exploran las variables de una en una, luego de dos en dos, y luego muchas variables a la vez.

Aunque el EDA abarca tablas de resúmenes estadísticos como la media y la desviación estándar, la mayor parte de las personas se centran en los gráficos. Se utiliza una variedad de gráficos y herramientas exploratorias, y se va allá donde se dirijan los datos. Si un gráfico o análisis no es informativo, mire los datos desde otra perspectiva.

Como el EDA implica explorar, es un proceso iterativo. Lo más probable es que se averigüen diferentes aspectos sobre los datos a partir de diferentes gráficos. Los objetivos típicos son comprender:

La distribución de variables en su conjunto de datos. Las relaciones entre variables.

Si sus datos tienen o no valores atípicos o puntos inusuales puede indicar problemas de calidad de los datos o conducir a descubrimientos interesantes. Si sus datos tienen o no patrones temporales. ( Price optimization DMC by Christian Taipe )

* + 1. **Técnicas para usar**

1. **Normalización / Estandarización**

La normalización de datos es un proceso fundamental en el preprocesamiento de datos, cuyo objetivo es ajustar la escala de las características numéricas para que sean comparables entre sí. A menudo, se refiere a la transformación de datos para que tengan un rango específico, como entre 0 y 1.

La estandarización y la normalización no son lo mismo, aunque a menudo se utilizan indistintamente y pueden generar cierta confusión. Ambas técnicas buscan ajustar y transformar los datos, pero tienen objetivos y enfoques ligeramente diferentes.

La estandarización se enfoca en la transformación de los datos para que tengan una media de cero y una desviación estándar de uno, mientras que la normalización busca ajustar los datos a un rango específico o a una escala definida. Ambas técnicas son útiles en diferentes situaciones y se seleccionan según los requisitos y objetivos del análisis de datos. (MEDIUM, 2024)

1. **Clustering**

La técnica de análisis cluster o análisis de conglomerados consiste en clasificar a los individuos en estudio formando grupos o conglomerados (cluster) de elementos, tales que los individuos dentro de cada conglomerado presenten cierto grado de homogeneidad en base a los valores adoptados sobre un conjunto de variables.

1. **Método K-means:**

El método consiste en dividir (en una clasificación inicial) los datos en un número k de clusters, especificado previamente. Un modo corriente de asignar los individuos a los k grupos antes de iniciar el proceso, es clasificándolos de acuerdo con alguna variable resumen, por ejemplo, suma de todas las variables implicadas o un procedimiento más refinado como puede ser la primera componente principal resultante de un análisis de componentes principales. La variable en cuestión se divide en k intervalos, codificados de 1 a k, y cada individuo será asignado al correspondiente código o grupo, según el intervalo al que pertenezca. A veces se aprovecha una clasificación previa obtenida mediante otro procedimiento como, por ejemplo, análisis jerárquico. También puede usar unas estimaciones previas de los centroides y con ellas una clasificación inicial.

1. **Regresión (Estimación de la función de demanda)**

La regresión lineal es una técnica de modelado estadístico que se emplea para describir una variable de respuesta continua como una función de una o varias variables predictoras. Puede ayudar a comprender y predecir el comportamiento de sistemas complejos o a analizar datos experimentales, financieros y biológicos. (MATH WORKS, 2024)

Las técnicas de regresión lineal permiten crear un modelo lineal. Este modelo describe la relación entre una variable dependiente y (también conocida como la respuesta) como una función de una o varias variables independientes Xi (denominadas predictores). La ecuación general correspondiente a un modelo de regresión lineal es:

Y=β0+∑ βk Xk +ϵi

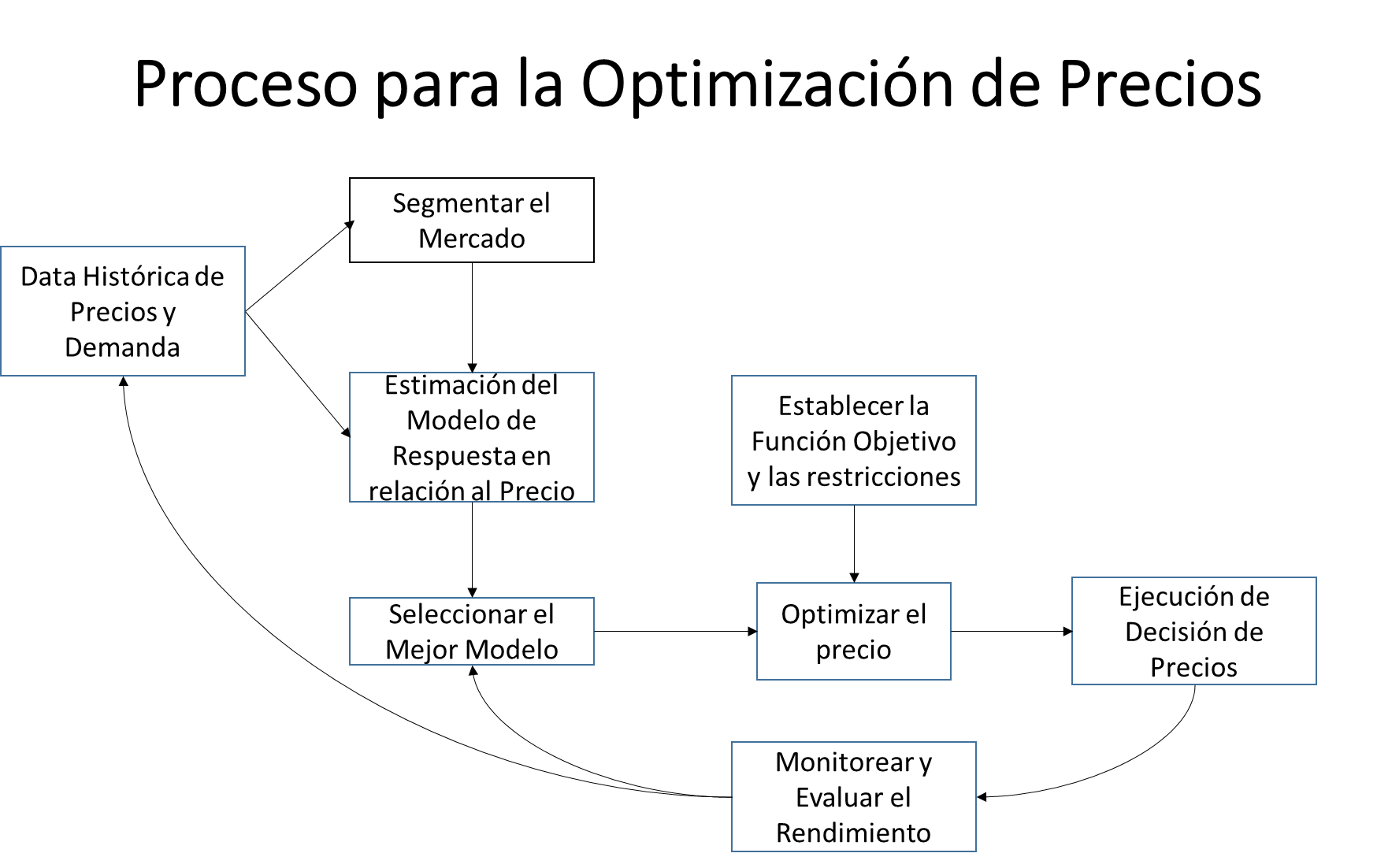
donde β representa las estimaciones de parámetros lineales que se deben calcular y

ϵ representa los términos de error.

1. **Optimización de precios**

El objetivo de Pricing Analytics es proporcionar el precio correcto para cada producto, para cada segmento de clientes, a través de cada canal. Se debe seguir el siguiente proceso:

**Proceso para la Optimización de precios**



Las empresas, a veces quieren perseguir objetivos diferentes o conflictivos como maximizar las ganancias y los ingresos minimizando los gastos. En estos casos podemos hacer uso de **las fronteras eficientes**.

La frontera eficiente es una valiosa herramienta de visualización porque permite a los tomadores de decisiones comprender cuán lejos están de la frontera eficiente y les permite elegir precios que satisfagan su compensación entre los dos objetivos (ingresos y rentabilidad).

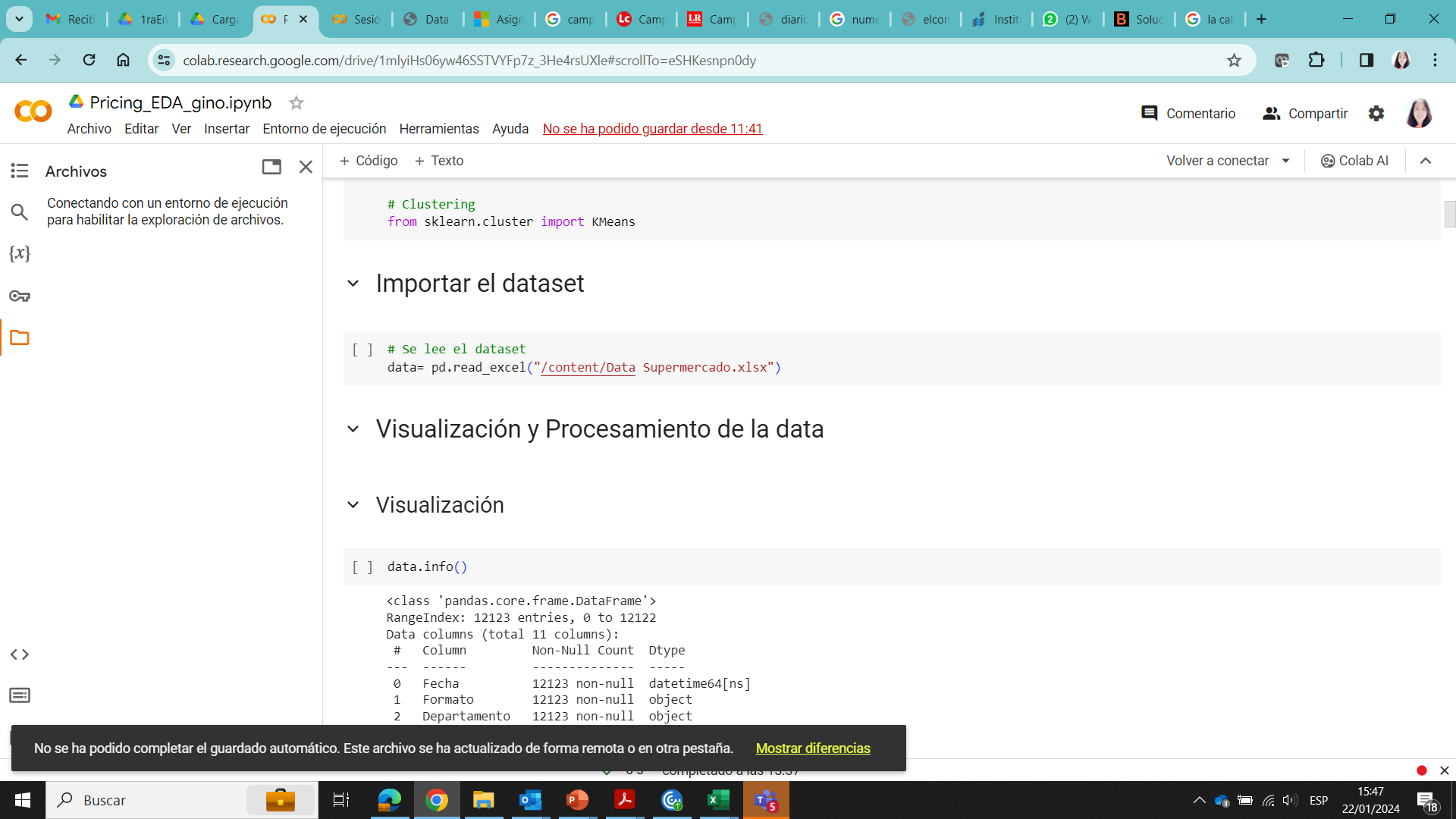
* 1. **Entregables:**

Los entregables que esperamos se obtengan de este proyecto incluyen cuadros detallados sobre los diferentes segmentos por zonas identificados(territorios), estimaciones de la función de demanda, recomendaciones de precios óptimos, frontera eficiente y una descripción completa del sistema de Pricing Analytics implementado que incluye lo siguiente:

* Segmentación de territorios
* Función de la demanda y las elasticidades de los productos.
* Rango de precios óptimos implementado por producto que maximicen la rentabilidad y/o ingreso.
* Informe detallado de la fijación de precios
* Código en Python
  1. **Valor Esperado:**

El valor esperado de esta propuesta técnica es maximizar los ingresos y/o rentabilidad de la venta de huevos pardo pack por 30 unidades en una cadena de supermercado, al establecer precios óptimos para cada segmento de zonas que permita maximizar las ventas de cada uno de los productos.

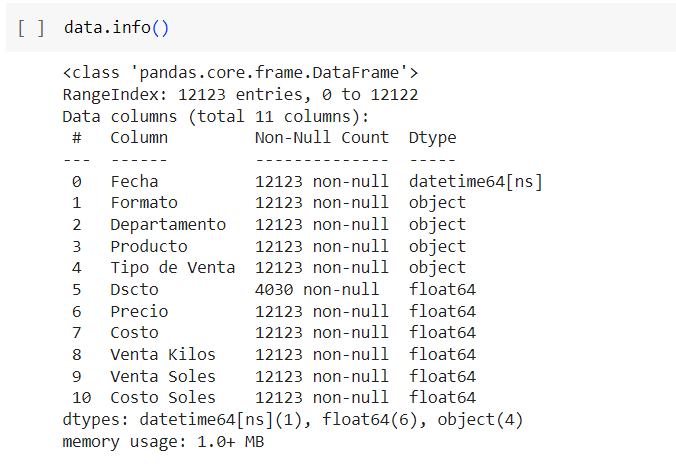
1. **Desarrollo del Proyecto:** 
   1. **Carga y visualización de objetos del dataset**



* 1. **Visualización de campos**

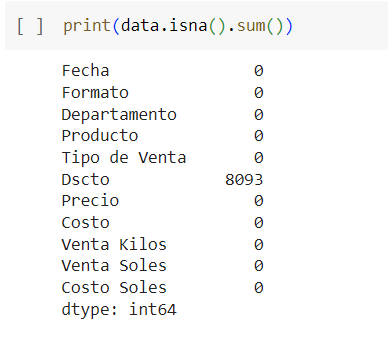


Tenemos 11 columnas, de las cuales podemos observar que la columna Dscto contiene registros NaN. También se compone por 12123 filas, pero como observamos en la siguiente imagen, Dscto solo tiene 4030 registros.



* 1. **Exploración de la información**

Hemos encontrado 8093 registros nulos en la columna Dscto.



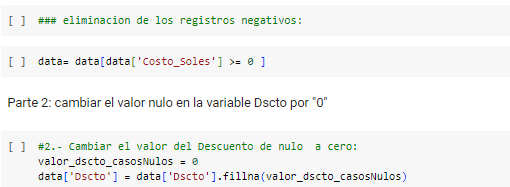
Hemos encontrado registros negativos en las columnas: Venta Kilos, Venta Soles y Costo Soles.



* 1. **Tratamiento de los valores atípicos:**



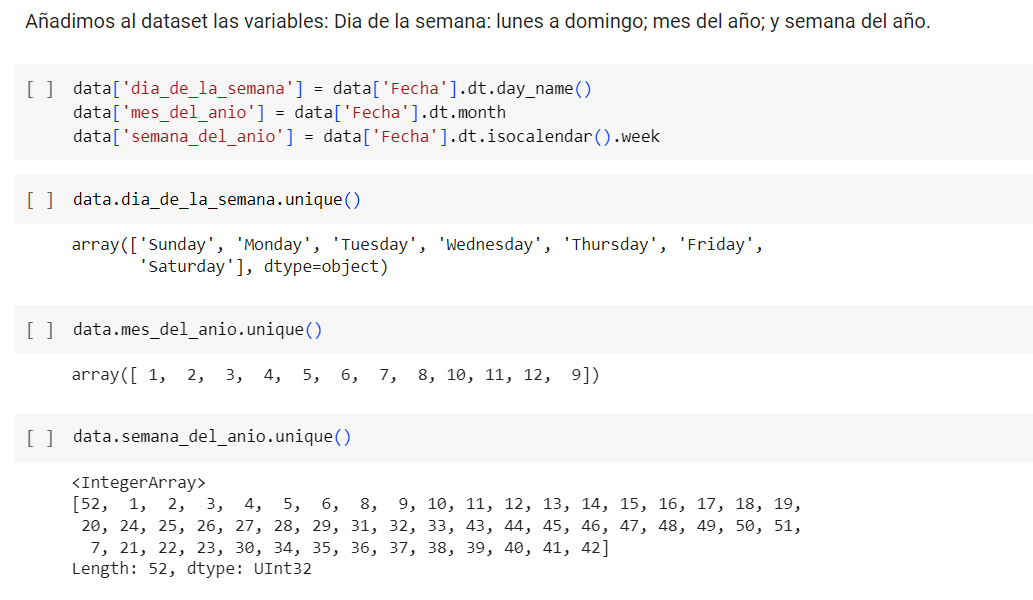
Hemos encontrado 3 registros de valores negativo que pasaremos a eliminar, debido a que se trata de casos de devoluciones de producto que corresponde al 0.02% de la información. Así mismo, cambiaremos los null por valores 0 en la columna Dscto.



* 1. **Creación de nuevas columnas:**

Al analizar el dataset se vio por conveniente crear variables de tiempo, debido a que la demanda difiere de los días de la semana, temporalidad escolar.

* Variables de tiempo



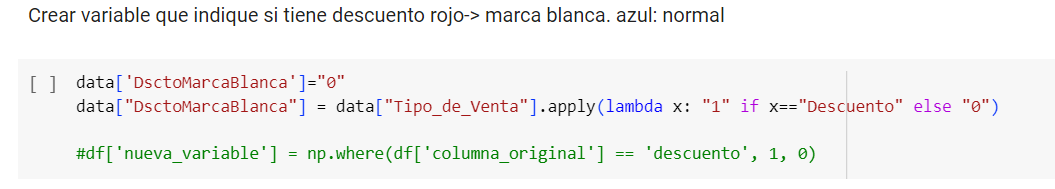
Asimismo para poder calcular el margen de retorno, se creó la columna de rentabilidad que resulta de la resta del precio menos el costo de los huevos.

* Columna de rentabilidad

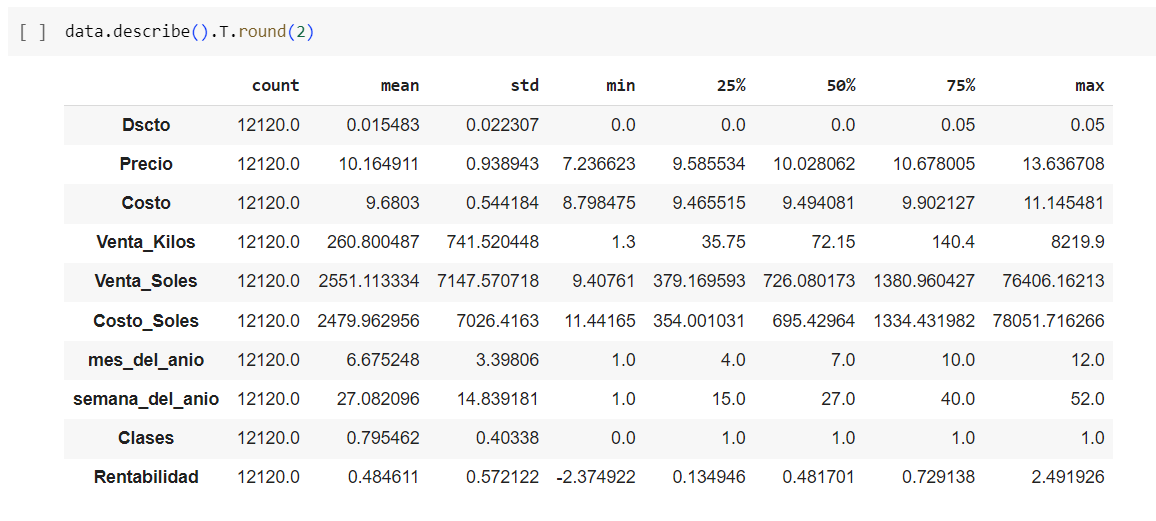


Y se creó una variable dummies donde especifica si la marca tuvo o no descuento a la fecha de la venta

* Columna que indica si tenemos descuentos en los productos



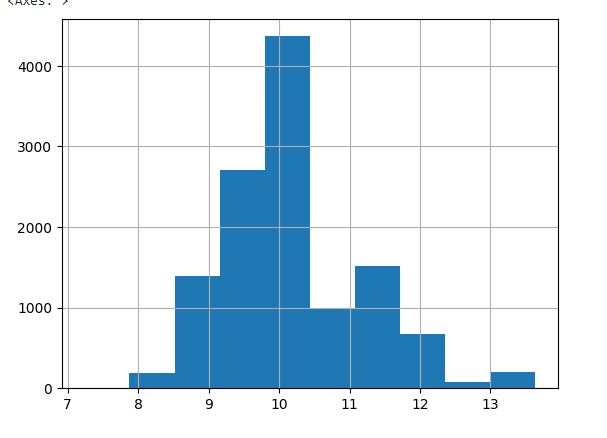
* 1. **Análisis Exploratorio**
     1. **Análisis descriptivo**
* Visualizamos que nuestros promedios en precio es 10.17, teniendo un mínimo de 7.24 y un máximo de 13.64 soles.
* En caso del costo, su promedio es 9.68, en ciertas temporadas llega a un mínimo de 8.80 y un máximo de 11.15 soles.
* En caso de la rentabilidad promedio es de 0.48 soles y una máxima que puede llegar a 2.49.
* Tenemos rentabilidad negativa para algunos registros donde la mínima es -2.37.
* Ya no tenemos valores negativos en la venta.



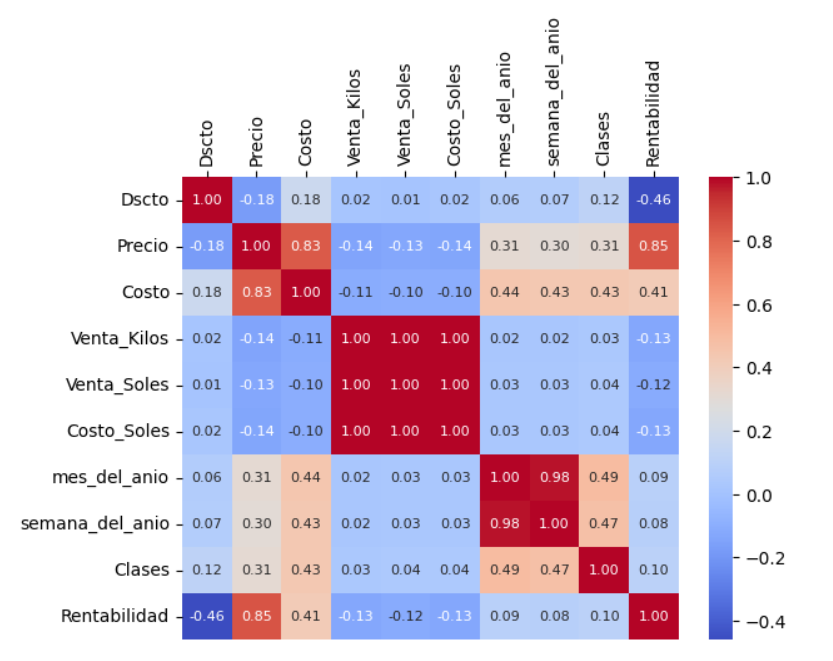
* + 1. **Precio**

En el siguiente histograma podemos ver que la mayor cantidad de registros se encuentran entre los precios 9.00 y 10.50 por kilo de huevo.

Los precios máximo es de 13.70 soles y el mínimo de 7.24 soles el cual varia de la zona geográfica y de la estacionalidad de la demanda y producción del huevo.



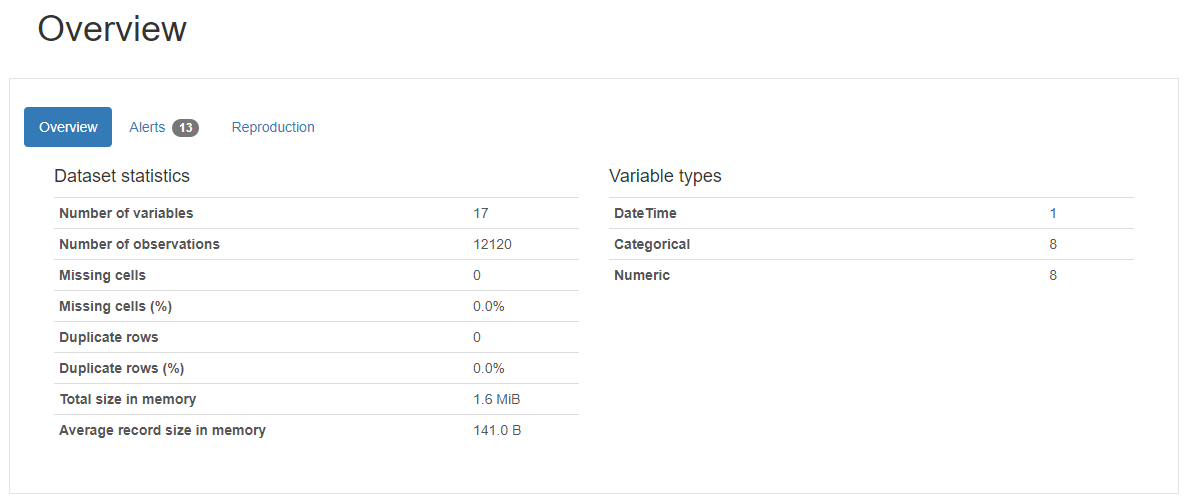
* + 1. **Elaboración de la gráfica de correlación**

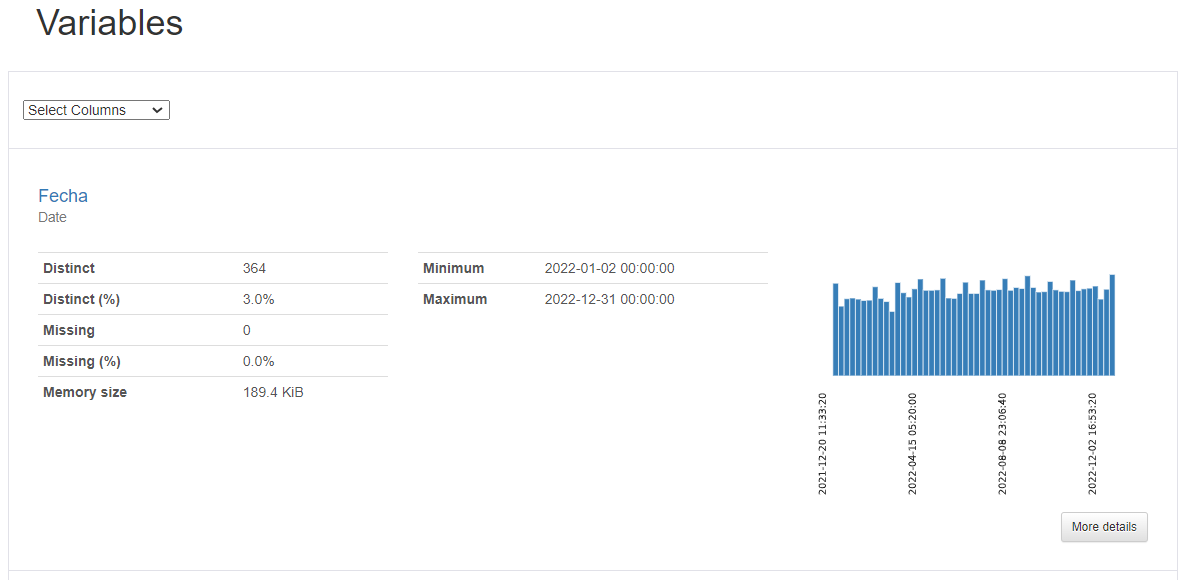


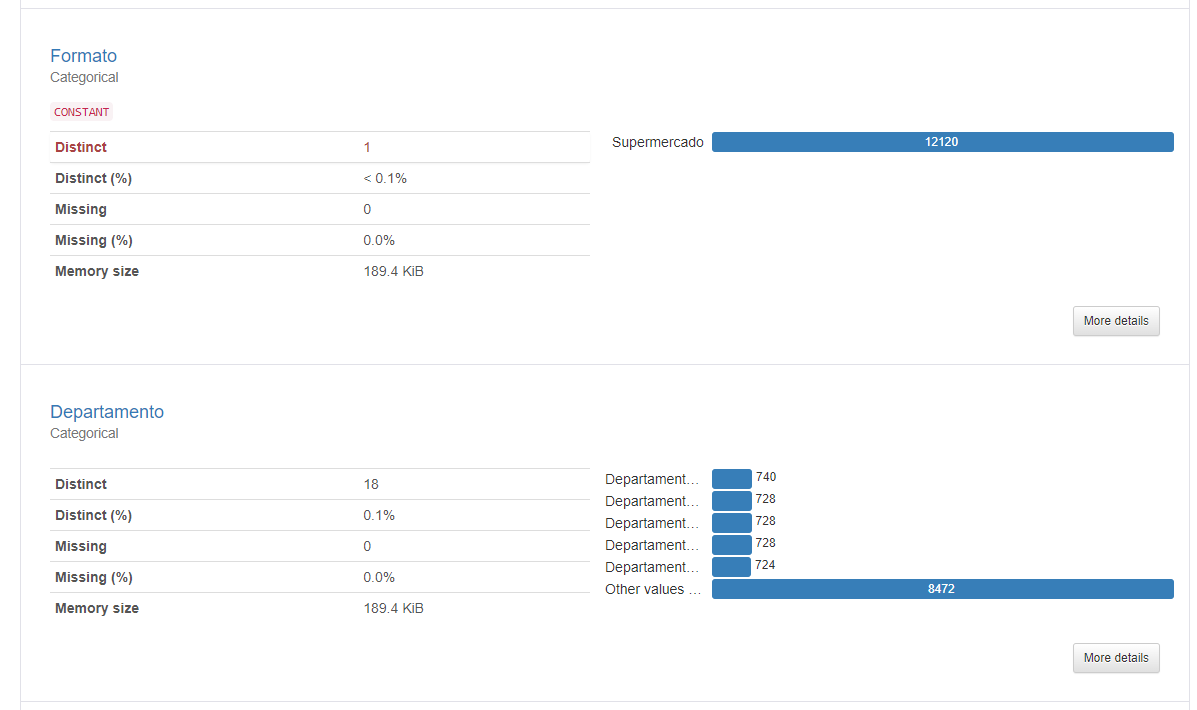
Podemos observar que tenemos una alta correlación entre precio y costo, y entre rentabilidad y precio. Pero observamos una correlación baja negativa entre precio volúmenes de venta, asimismo con la variable descuento.

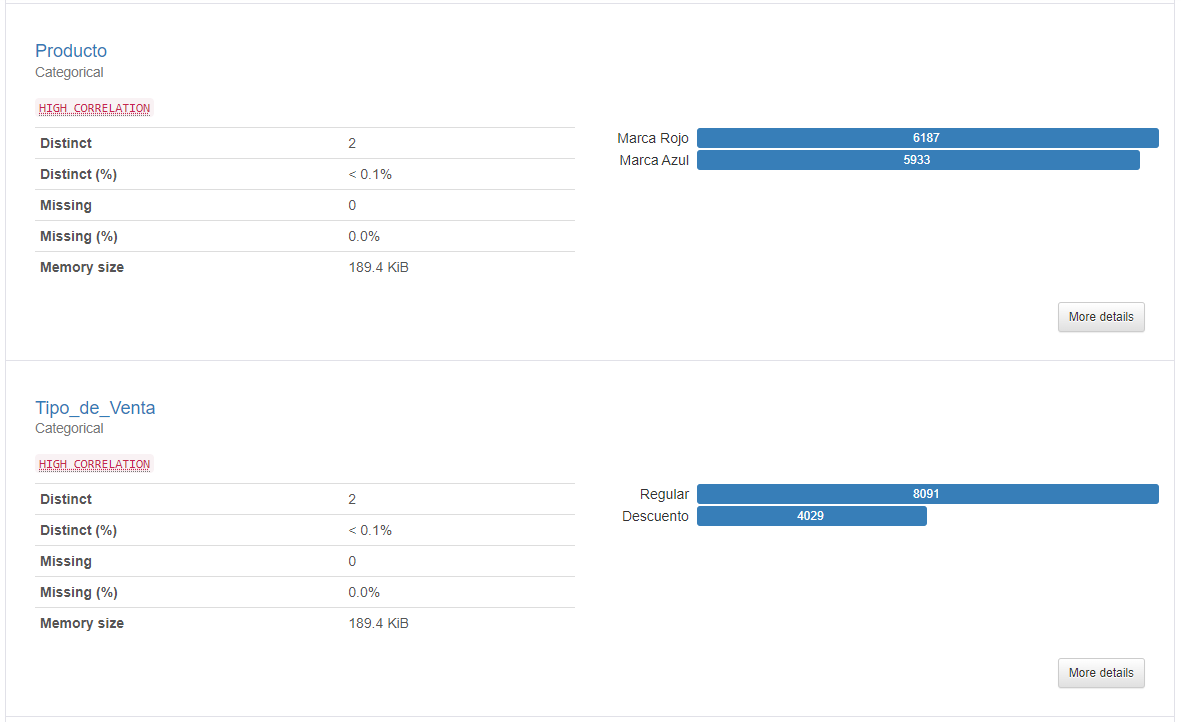
* + 1. **Análisis descriptivo total**

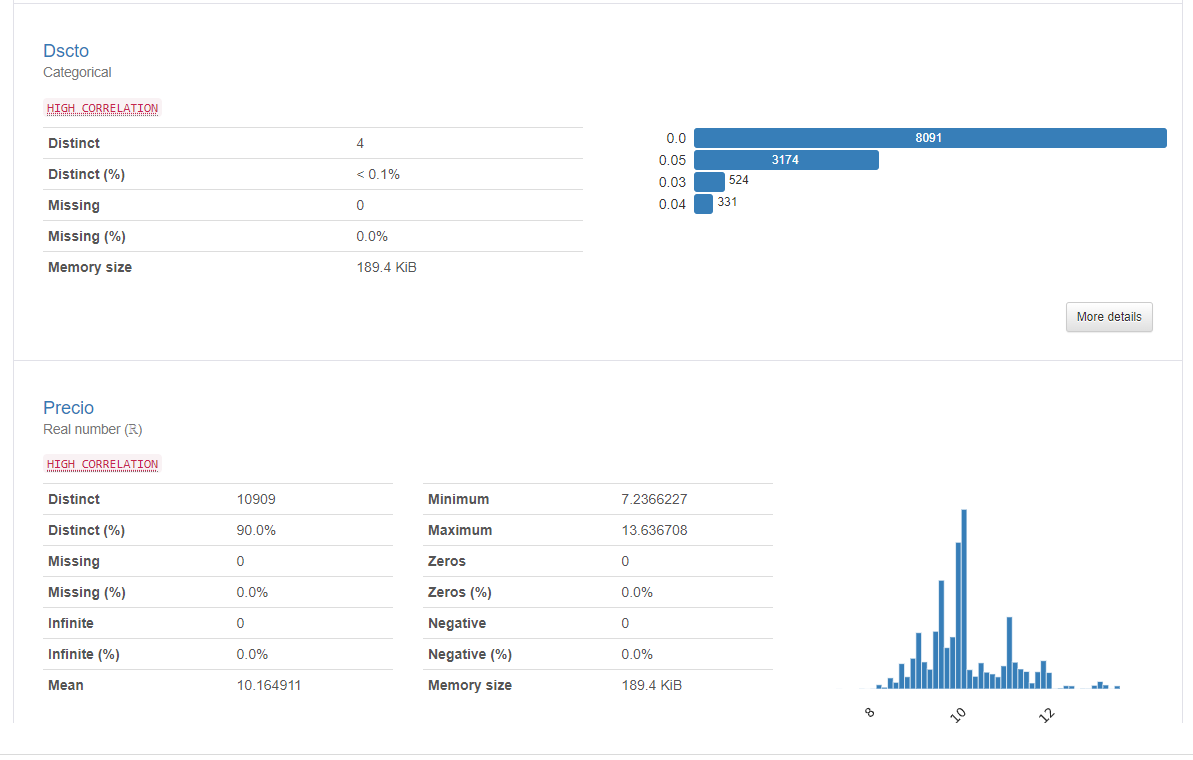
Se utilizo la función de profile report para generar un análisis descriptivo de todas las variables del dataset.

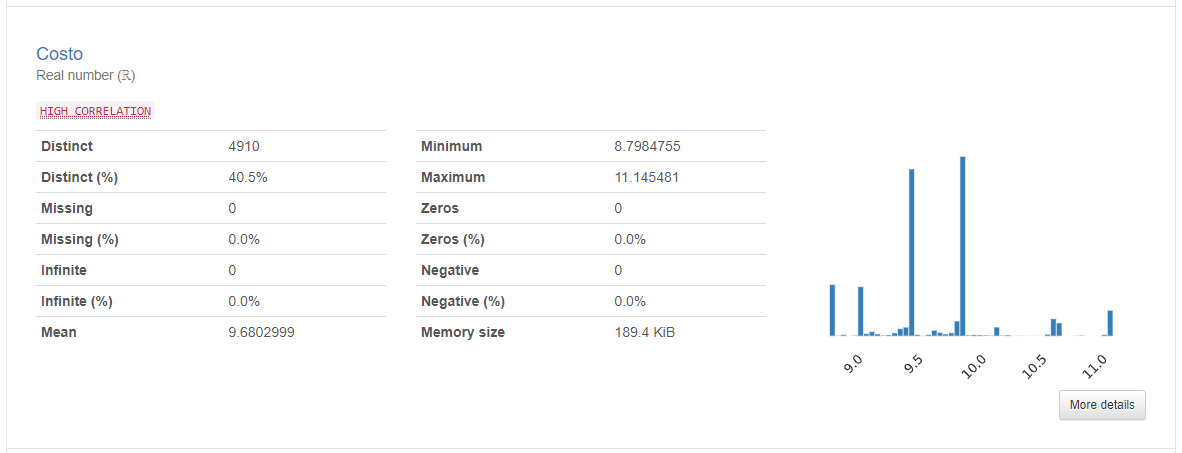


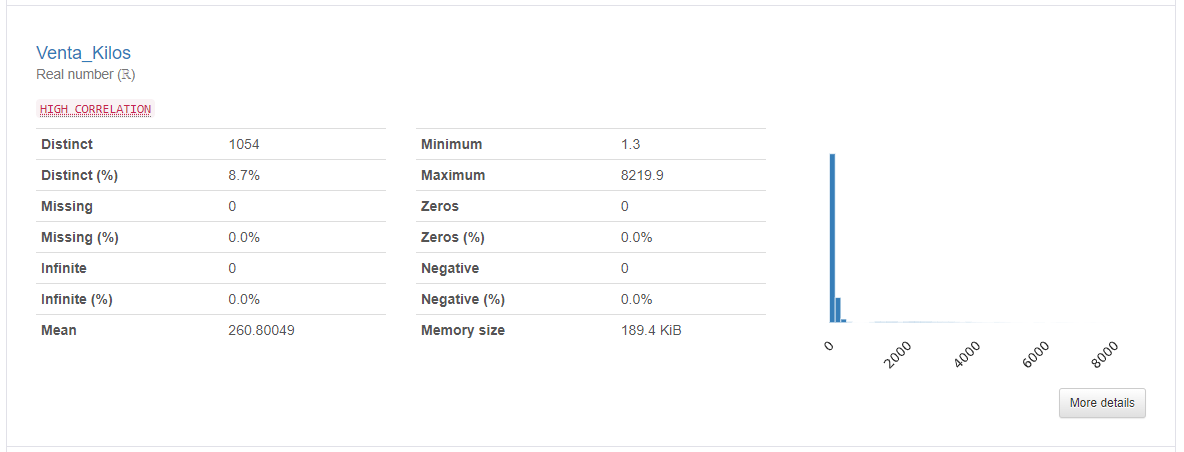


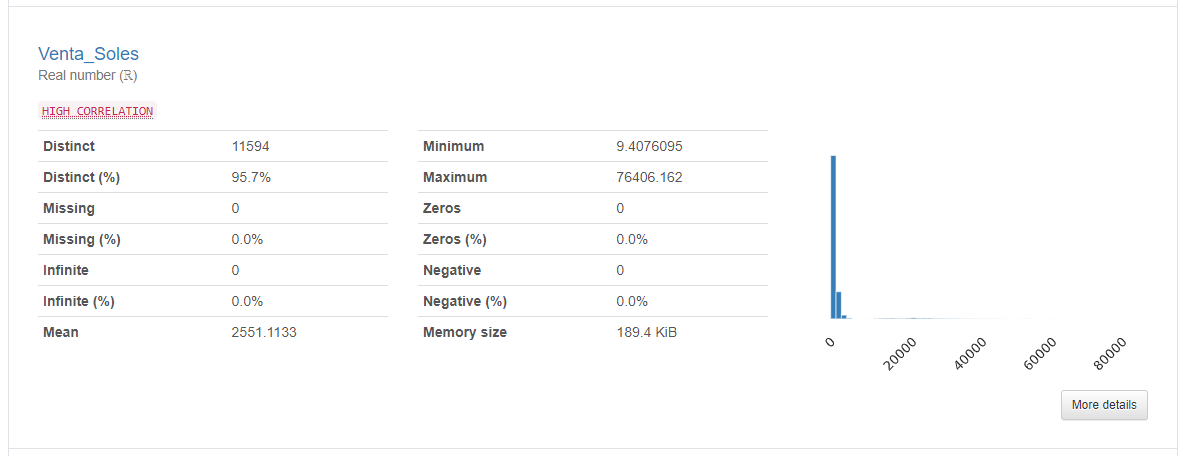


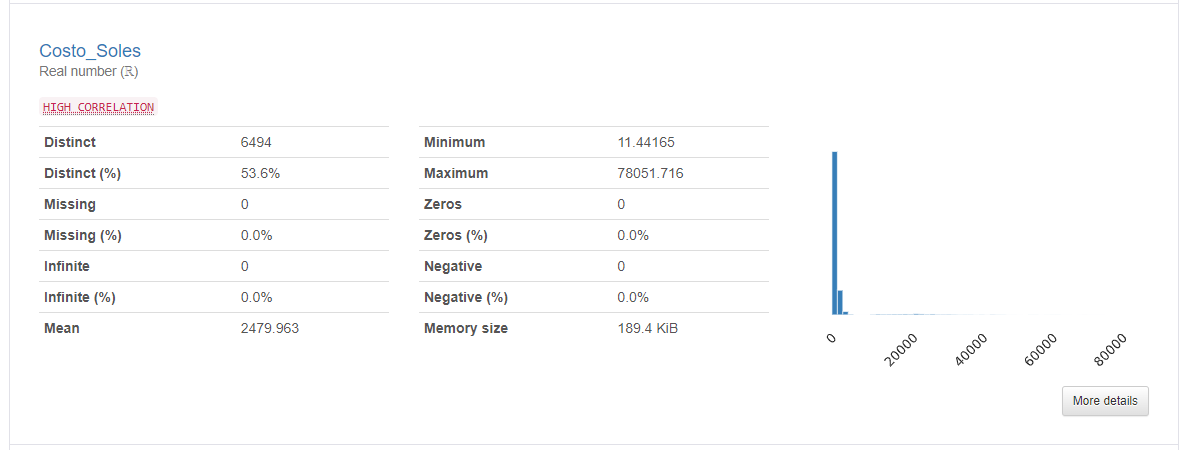


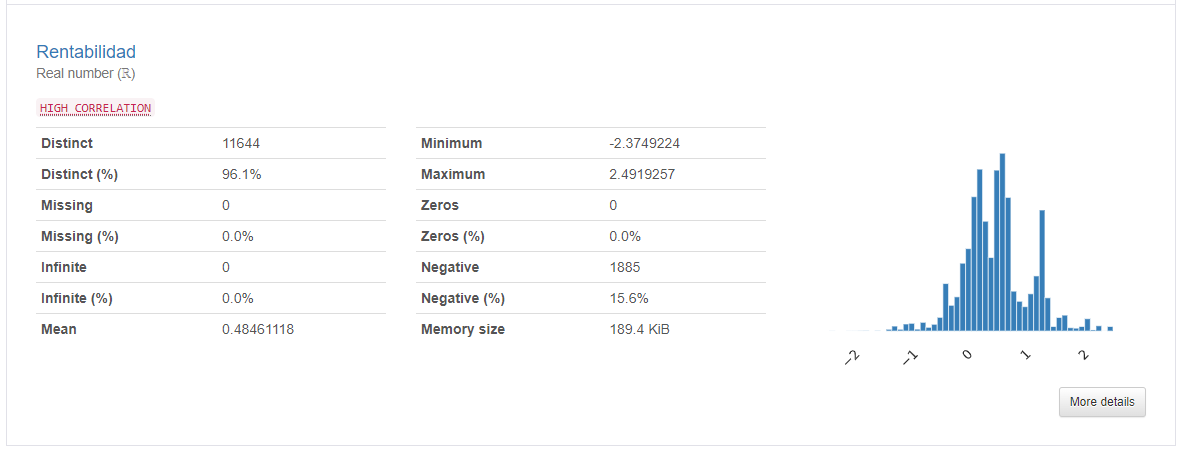




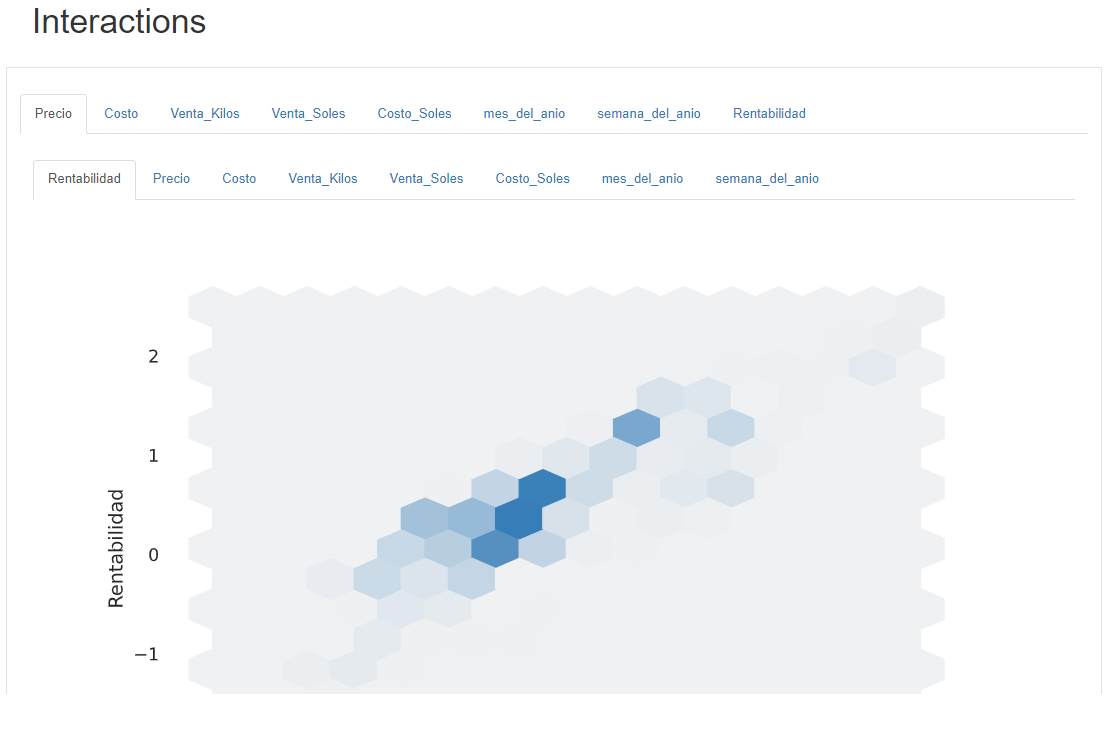




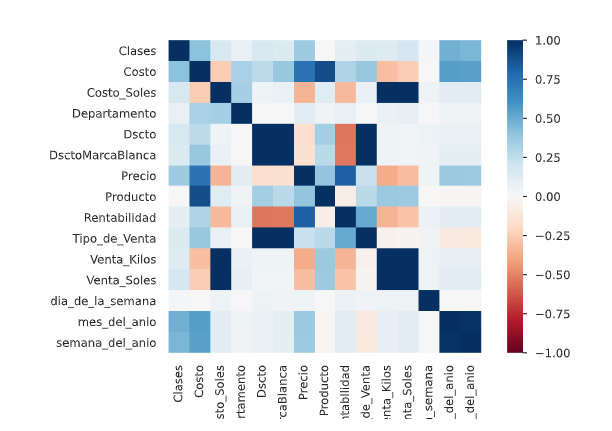




Asimismo la función te posibilita realizar interacciones entre las variables del dataset de manera bivariado.

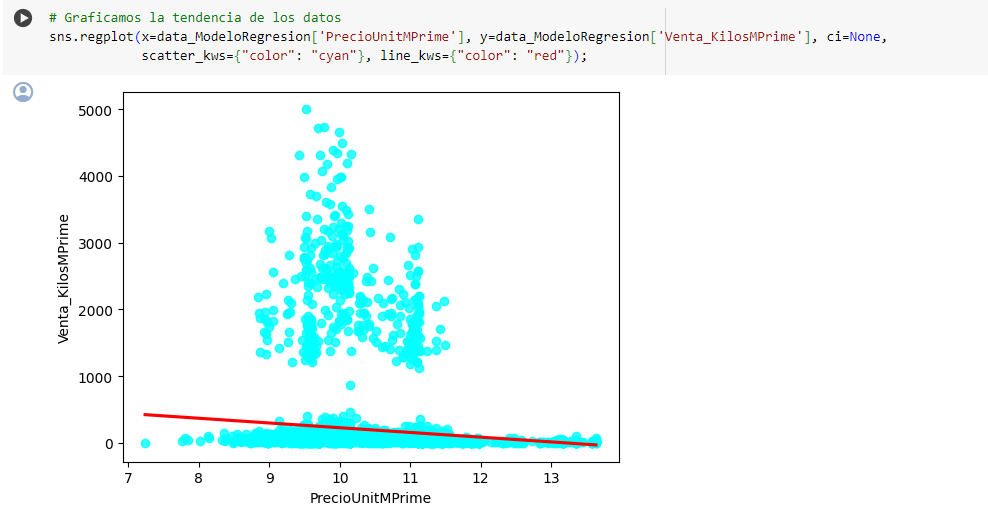


Y realiza el mapa de calor de las correlaciones:

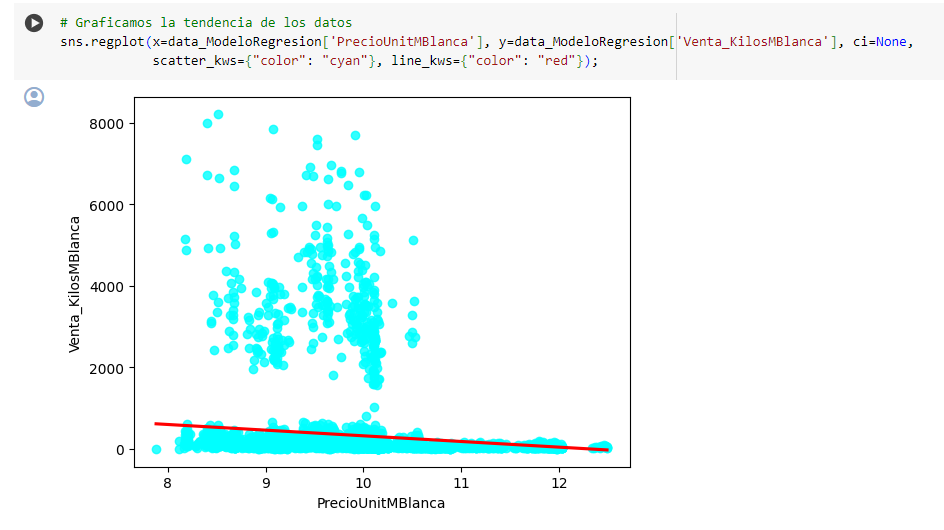


Donde se encuentran las correlaciones más altas en:

* Costos con venta en soles o venta en kilos
* Producto con costo
* Costo y precio
  1. **Análisis de Regresión**

****

En caso de las ventas de marca Prime(XYZ), se visualiza que a medida que aumenta el precio las ventas disminuyen y caso contrario el precio es menor, las ventas aumentan.



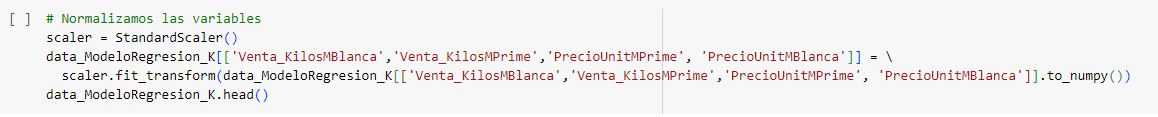
En caso de la marca blanca el comportamiento es similar, a medida que disminuye el precio de la marca blanca las ventas aumentan en mayor proporción a la marca prime y en caso de que se aumente el precio las ventas disminuyen.

* 1. **Clustering**

Primero agrupamos la data del dataset a nivel de departamento, de la cual haremos la segmentación de las zonas para luego utilizar estos segmentos en nuestra función de demanda.

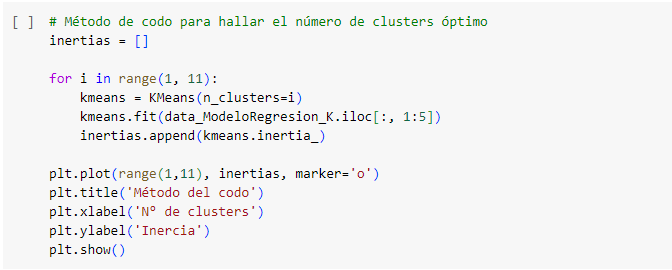
****

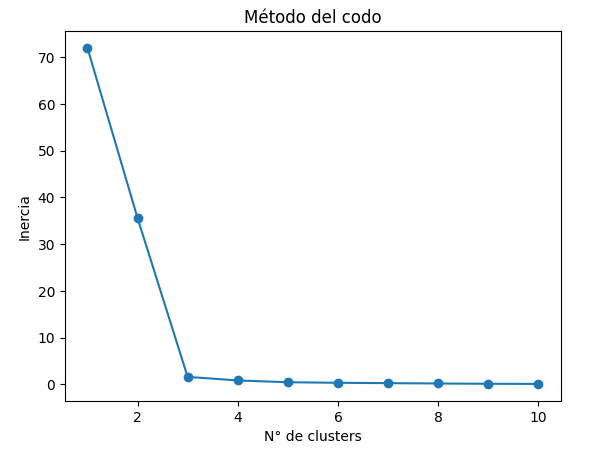
Posteriormente se normalizan las variables a utilizar en el modelo de cluster, como por ejemplo la venta de kilos de la marca blanca, la venta de kilos de la marca prime, y los precios unitarios de ambos productos.

****

****

Para determinar el nro. de cluster óptimos se utilizó el método del codo, obteniendo la métrica de 3 cluster para el modelo de segmentación.

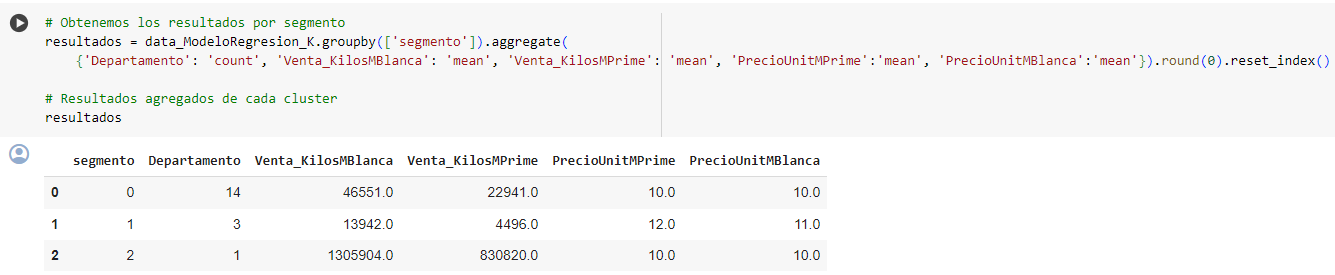


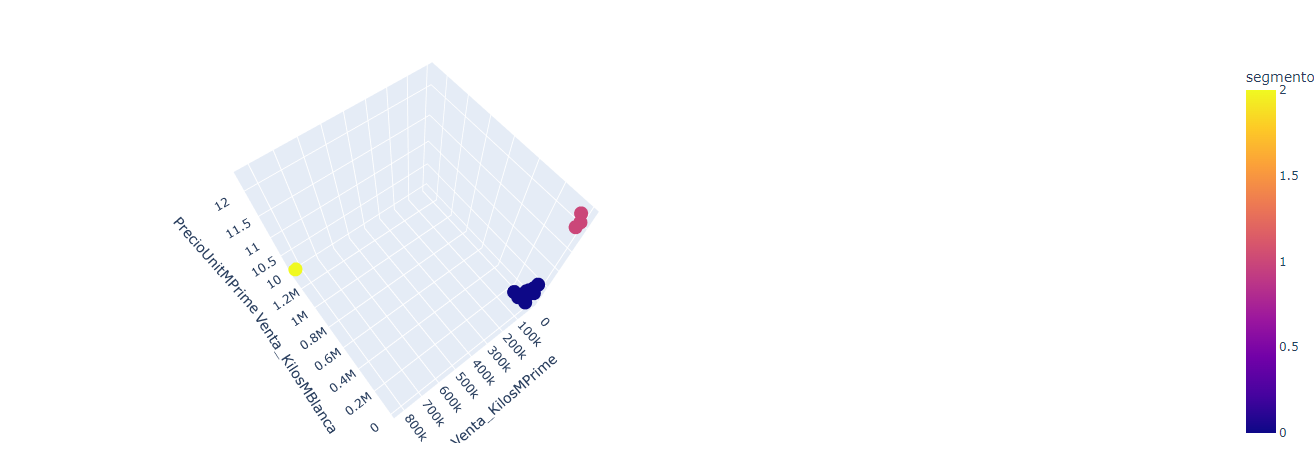


Se realizo el modelo de clusterización, obteniendo 3 segmentos geográficos diferenciados por el volumen de ventas y el precio del producto. Es decir de los 18 departamentos existentes en la data solo los quedamos con 3 segmentos donde:

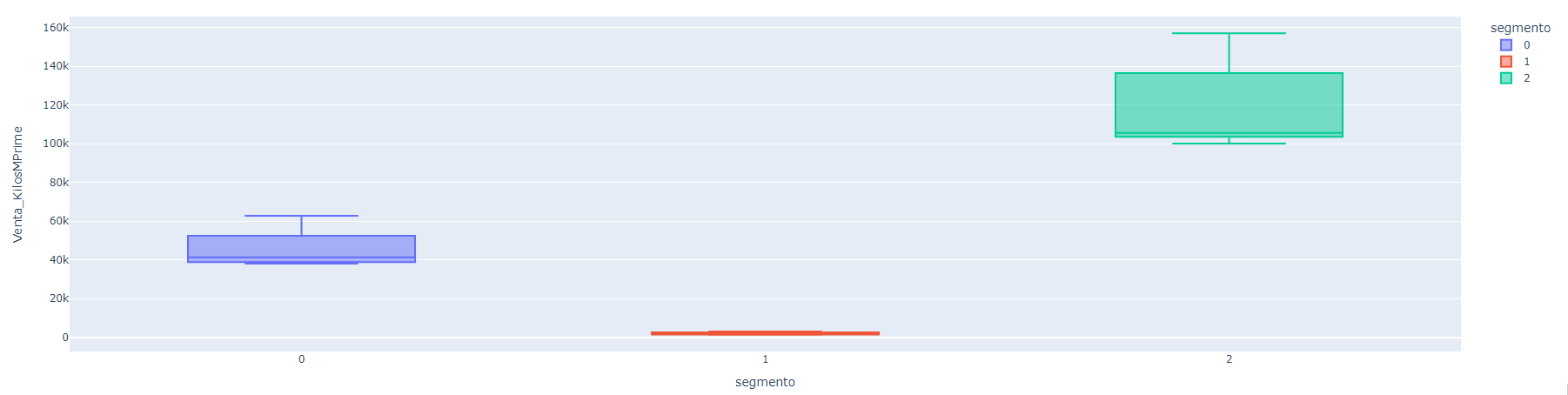
* 1 segmento: conformado por 14 departamentos, el precio promedio era 10 y las ventas bajas
* 2 segmento: conformado por 3 departamentos, donde el precio era alto y las ventas bajas
* 3 segmento: conformado por 1 departamento, el precio es bajo y las ventas muy altas





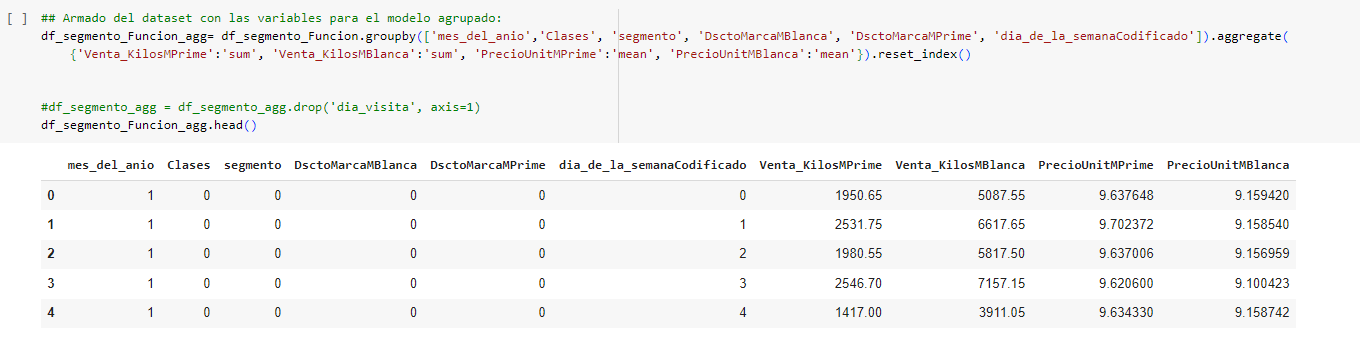


En el siguiente grafico se puede visualizar las ventas por cada segmento obtenido, donde el segmento 2 representa el mayor volumen de las ventas de la marca prime.

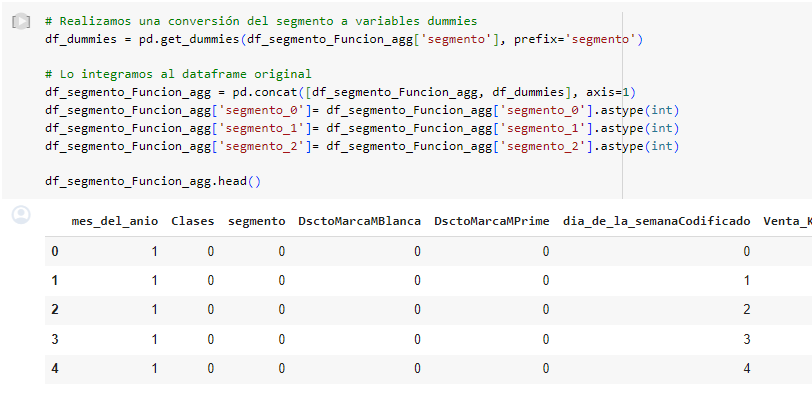


* 1. **Función de la demanda**

Armamos el dataset con la nueva variable del segmento para realizar el modelo agrupado

****

Realizamos la conversión de los segmentos en variables dummies para la inclusión al modelo

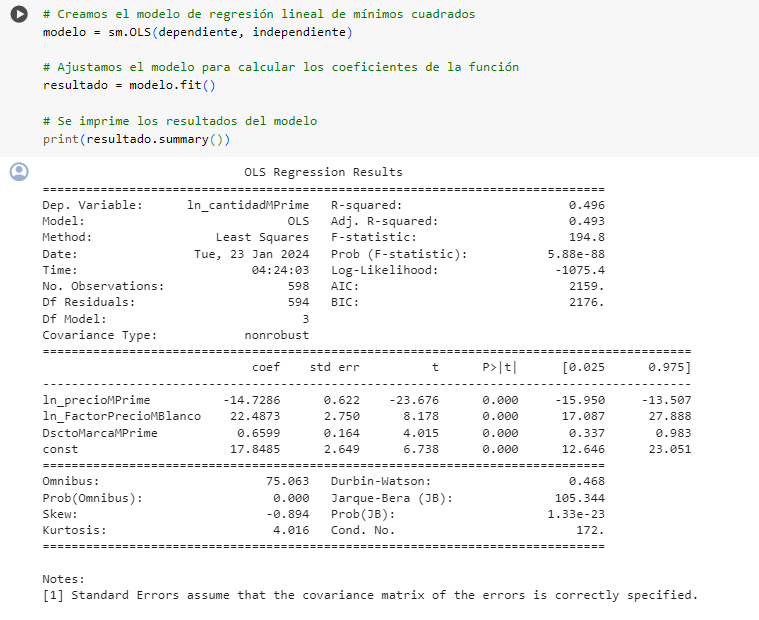
****

Generamos el modelo de regresión, para ello se creo una variable adicional del factor del precio entre la marca prime y la marca blanca, para determinar si hay un efecto de la reducción de precio en una marca genera la variación en la demanda del otro producto.

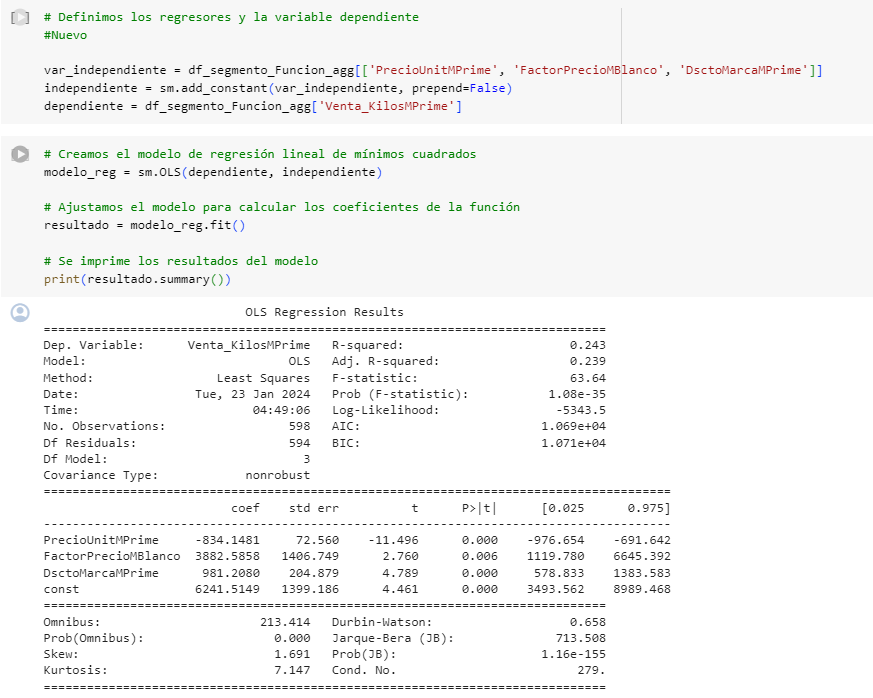
****

Creamos el modelo de mínimos cuadrados, donde nuestro r2 es casi un 50% , todos los coeficientes del modelo son aceptados yobtenemos una factor negativo en relación al aumento del precio genera una demanda negativa del producto prime como se visualizó

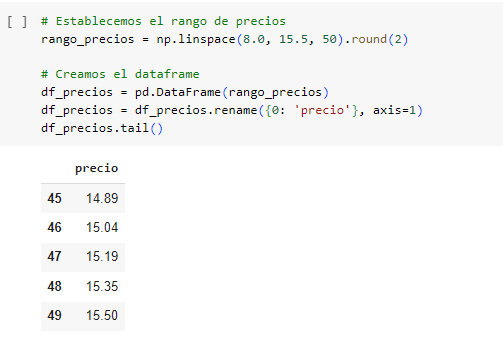
en la parte de regresión que se realizó anteriormente.

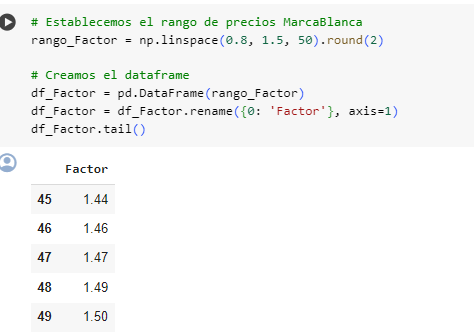
****

Generamos la estimación de la demanda, definimos la variable dependiente e independiente del modelo y finalmente generamos el modelo de regresión lineal de mínimos cuadrados.

****

Luego del modelo predecimos la demanda, para ello generamos valores para las variables de precio y factor

****

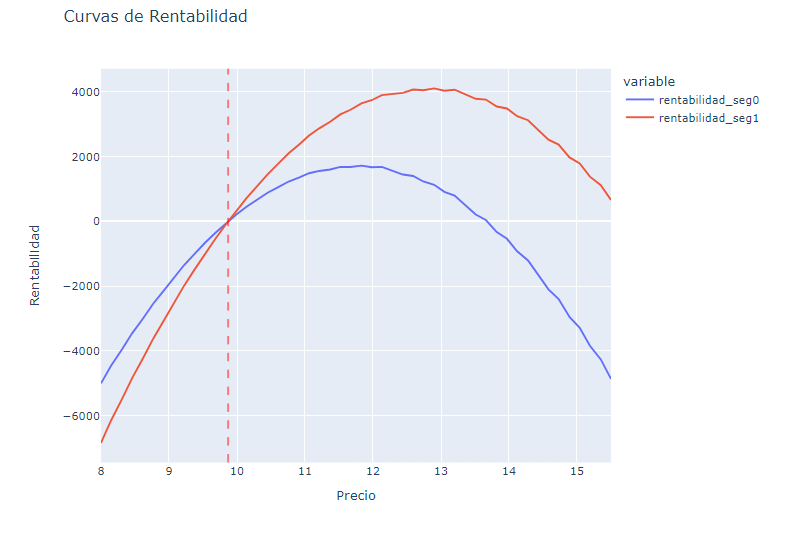
****

Luego obtenemos los siguientes resultados por cada segmento, donde se visualiza la que ambos segmento son más sensibles al precio. El segmento0 tiene una aceptación de precio menor con respecto al segmento 1

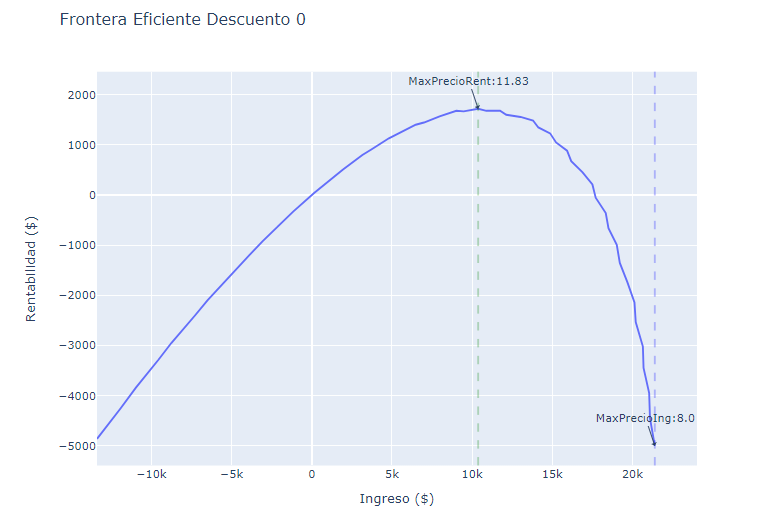
****

* 1. **Optimización de precios**

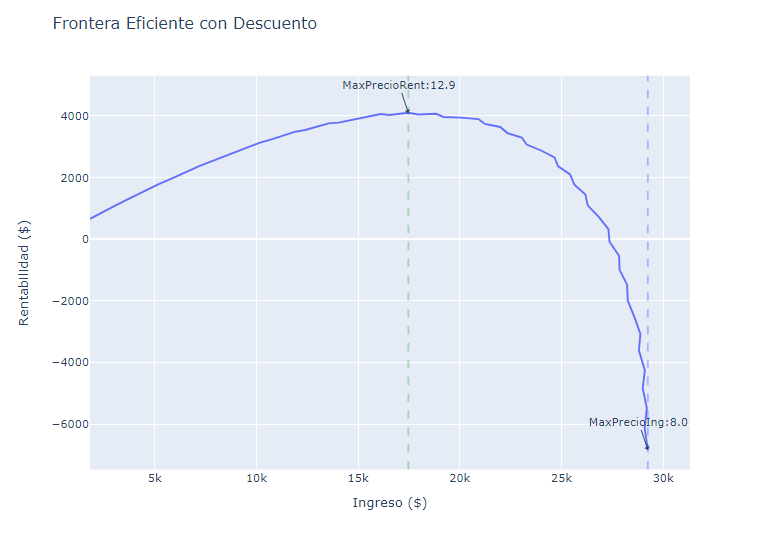
Cuando solo queremos maximizar los ingresos por la venta de los productos de huevo , se podría determinar erróneamente que el precio optimo es de 8 soles para todos los segmentos. Por ello es que se realiza una optimización del precio con las variables de ingreso y rentabilidad para generar un beneficio real a la compañía. A través de este se obtiene que el precio optimo es de 11.83 para obtener la mayor rentabilidad de ambos productos.



Que es validado con el modelo de frontera eficiente, que el precio optimo es 11.83 cuando no aplica descuento.

****

Y el precio optimo cuando hay descuento es de 12.90 soles

****

**Conclusiones**

* Se ha elaborado el modelo de precios para los productos de marca prime(XYZ) y marca blanca de una línea de producto y una cadena de supermercados.
* Se han realizado segmentos por las zonas geográficas que tienen diferente aceptación de precios.
* Se ha determinado el precio optimo para el producto prime cuando aplica o no descuento que es 11.83 y 12.90 respectivamente.

**Recomendaciones**

* Se recomienda ampliar la base de datos de mayor temporalidad para ver el efecto estacional del producto.
* Asimismo se puede aplicar el modelo a los otros productos y canales de la empresa.

# Bibliografía

*AKVO*. (08 de 01 de 2024). Obtenido de AKVO: https://datajourney.akvo.org/blog/how-to-clean-your-development-data?utm\_term=data%20cleaning&utm\_campaign=&utm\_source=adwords&utm\_medium=ppc&hsa\_acc=7028243667&hsa\_cam=2077694068&hsa\_grp=131836071793&hsa\_ad=565589428704&hsa\_src=g&hsa\_tgt=kwd-353628671407

*ECONOMIPEDIA*. (08 de 01 de 2024). Obtenido de httECONOMIPEDIA: https://economipedia.com/definiciones/elasticidad-precio-de-la-demanda.html

*INTUIT MAILCHIMP*. (08 de 01 de 2024). Obtenido de INTUIT MAILCHIMP: https://mailchimp.com/es/resources/price-elasticity-of-demand/#:~:text=La%20elasticidad%20precio%20de%20la,posible%20que%20consigas%20m%C3%A1s%20clientes

*JMP*. (08 de 01 de 2024). Obtenido de JMP STATISTICAL DISCOVERY: https://www.jmp.com/es\_co/statistics-knowledge-portal/exploratory-data-analysis.html

*MATH WORKS*. (08 de 01 de 2024). Obtenido de MATH WORKS: https://la.mathworks.com/discovery/linear-regression.html

*MEDIUM*. (08 de 01 de 2024). Obtenido de MEDIUM: https://medium.com/@nicolasurrego/transformando-datos-en-oro-c%C3%B3mo-la-estandarizaci%C3%B3n-y-normalizaci%C3%B3n-mejoran-tus-resultados-fbe0840d2b94

*POWER DATA*. (08 de 01 de 2024). Obtenido de POWER DATA: https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/bid/312584/procesos-etl-definici-n-caracter-sticas-beneficios-y-retos