



EVIDENCIA 1

Análisis estadístico de la base de datos

Análisis de datos para la toma de decisiones (Gpo 572)

Profesores:

Gibran Sayeg Sánchez, Lorena Guadalupe Gómez Martínez, Juan Carlos Lavariega Jarquín, Pablo Joaquín López Granados, M.C. Patricia Domínguez Alavez, Juan Ángel Martínez Arzola y María de los Ángeles Constantino González

Annette Pamela Ruiz Abreu

A01423595

Índice

Introducción	2
Parte 1: Selección de variables	3
Parte 2: Estadística Descriptiva	5
Parte 3: Correlación y Regresión Lineal	9
Parte 4: Conclusión	11
Bibliografía	12

Introducción

TECstore nace con el objetivo de fomentar el orgullo, sentido de pertenencia y promover el espíritu Tec a través de productos oficiales, así como la creación de nuevas experiencias que conecten a la comunidad del Tecnológico de Monterrey con sus marcas. Actualmente cuenta con 11 tiendas físicas, 1 kiosko en Fashion Drive Monterrey, la tienda en línea y tiendas oficiales en Mercado Libre, Amazon y Rappi.

Uno de los grandes problemas de la marca es su lento crecimiento y adquisición de clientes. Esta tienda, como la mayoría de los negocios del siglo XXI, cuenta con una gran cantidad de datos recopilados a lo largo del tiempo que no se han estado aprovechando para poder entender y solucionar el problema de visibilidad.

El objetivo de este proyecto es analizar los datos de TECstore para ayudar a los ejecutivos, los cuales son nuestra audiencia, a determinar cuáles son sus áreas de oportunidad y qué pueden hacer para mejorar su crecimiento.

Al ser la responsable del análisis de clientes, mi responsabilidad es conocer al mercado de TECstore mediante la analítica descriptiva para posteriormente proponer acciones para atraer al mercado faltante.

(ITESM, 2022)

Parte 1: Selección de variables

Determina y justifica qué variables cuantitativas y cualitativas utilizarás en tu análisis. Mínimo 2 variables cuantitativas que consideres están relacionadas, y 2 cualitativas, explica porque dichas variables son relevantes en la situación analizada.

Variables Cualitativas

Género: Al tener que analizar los clientes, el género es una de las dimensiones más importantes para poder diseñar estrategias enfocadas al género de clientes que menos compre en la TECstore.

Edad: La edad nos permitirá dividir en clases a los clientes para posteriormente identificar qué grupos de edad son los que más o menos compran en la TECstore. Además, es muy importante tener esta división de edad porque se tendrán que diseñar propuestas diferentes para cada grupo (adolescente, joven, adulto, etc.); ya que hay grandes brechas entre generaciones y se requiere de una estrategia diferente para cada tipo de audiencia.

Dominio: El dominio será útil para enfocarnos a un solo segmento del mercado y poner de lado las ventas departamentales; ya que estas no nos interesan y pueden afectar nuestro análisis.

En general, estas variables cualitativas son necesarias para el análisis de clientes; ya que es necesario conocer a nuestro mercado para poder diseñar propuestas efectivas e identificar tendencias o áreas de oportunidad.

Variables Cuantitativas

Subtotal: El subtotal se usará para poder observar las ventas totales de los clientes. Obviamente esta es la métrica más importante para un negocio; ya que lo que le interesa a la TECstore es incrementar sus ventas. Esto será posible identificando los clientes que menos gastan y diseñando estrategias para aumentar las ventas en ese grupo de la población.

Campos Calculados

CantidadClientes (COUNTD(idClientes)): Este campo calculado nos permitirá contar la cantidad de clientes que hay. Podremos estudiar esta métrica usando diferentes dimensiones como género, dominio y edad.

TicketPromedio (SUM(Subtotal)/COUNTD(idOrder)): Este campo calculado nos permitirá observar el ticket promedio de los clientes; es decir, podremos ver cuánto gastan en promedio los clientes de cierto género, dominio y/o edad. Además, nos servirá para ver las regiones en donde en promedio gasten los suficiente para obtener un envío gratis.

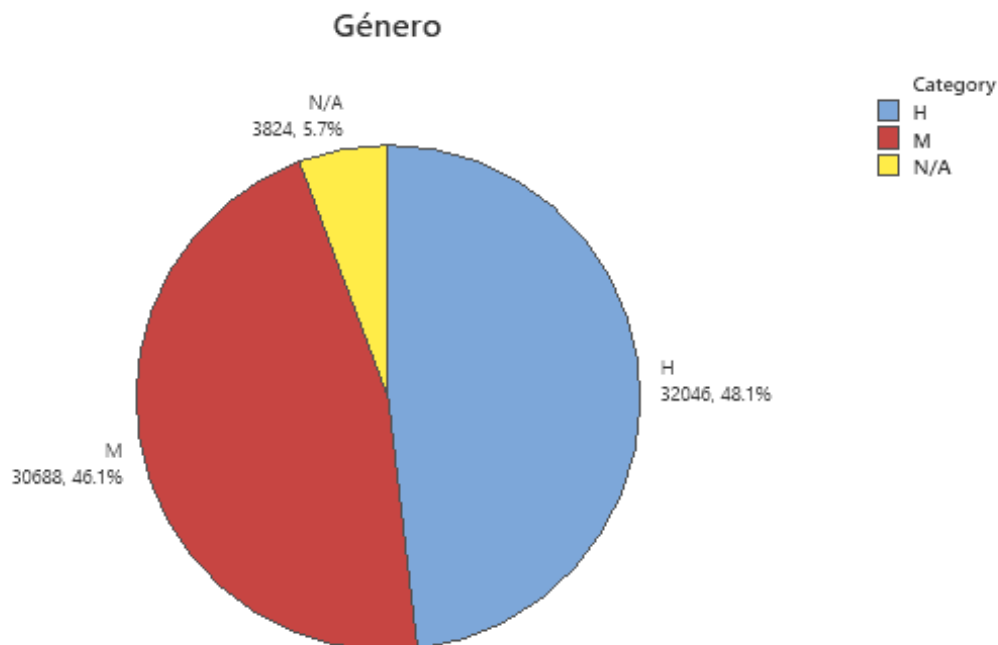
Parte 2: Estadística Descriptiva

Describe las variables cuantitativas seleccionadas por medio de medidas descriptivas de tendencia central, dispersión y de ubicación, selecciona la representación gráfica más adecuada para cada una de ellas. Describe las variables cualitativas con base en tablas de frecuencia y representación gráfica. Argumenta por qué la relevancia de analizar estas variables con dichas medidas descriptivas y representaciones tabulares o gráficas.

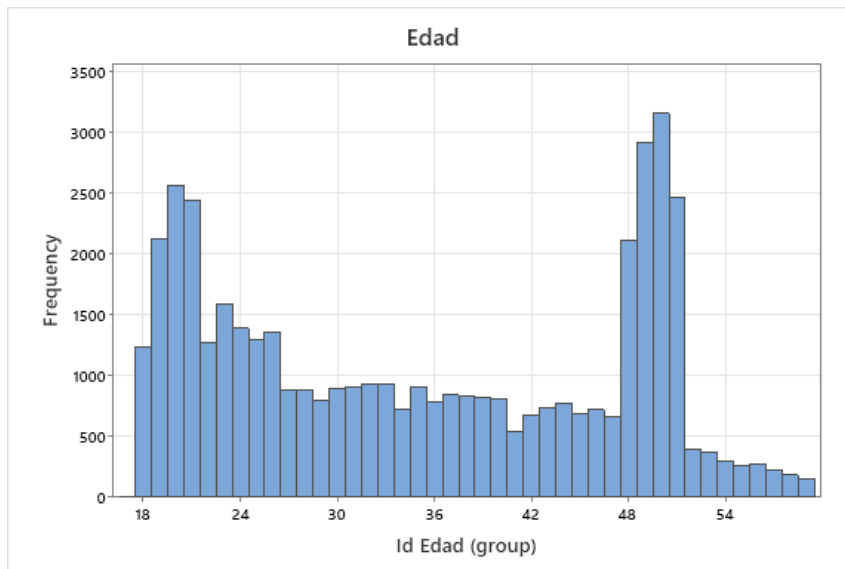
Variables Cualitativas

Es importante mencionar que para un mejor análisis, todas las gráficas se deberían modificar dependiendo de los filtros o especificaciones que se requieran.

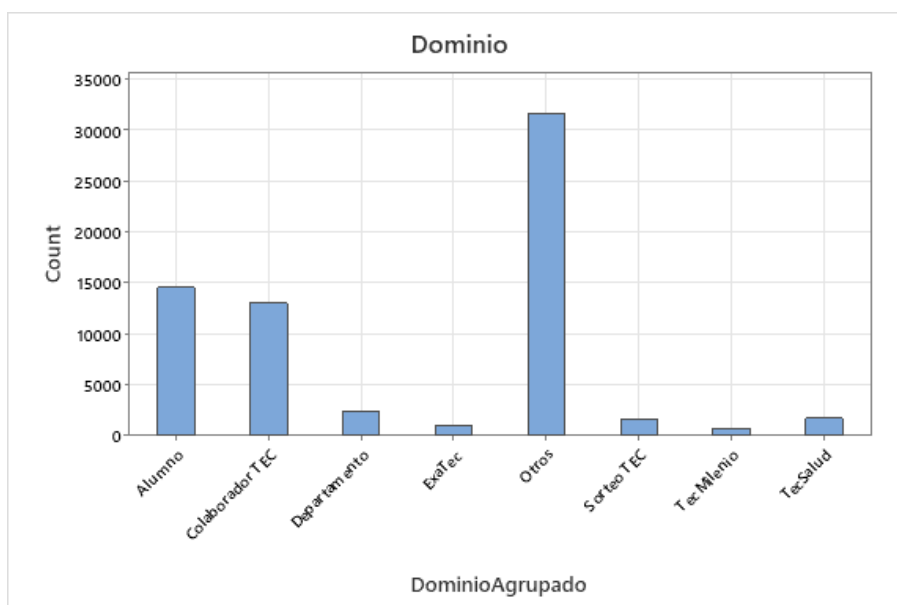
El género se puede representar en una gráfica circular porque son pocas categorías/clases (3) y esto nos permite ver con facilidad el género que predomina en la totalidad de los pedidos.



La edad se puede representar en un histograma; ya que podemos dividir las edades en clases podemos ver el centro, la extensión y la forma del conjunto de datos. También podemos usar el histograma como una herramienta visual para comprobar la normalidad. Además, las variables numéricas y las cualitativas ordinales usualmente se presentan en histogramas en vez de gráficos de barras.



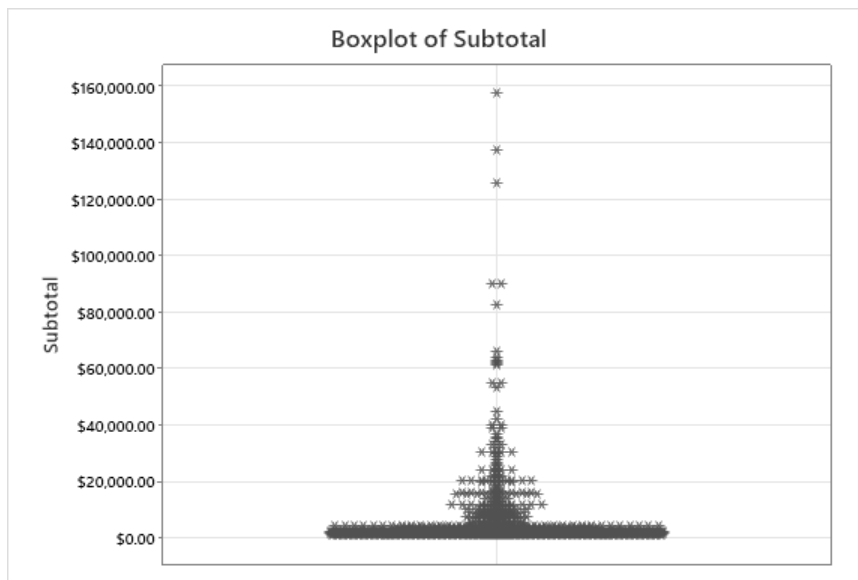
El dominio puede ser representado en un gráfico de barras ya que es una variable cualitativa nominal; es decir, las clases no tienen un orden ni una es más importante que otra.



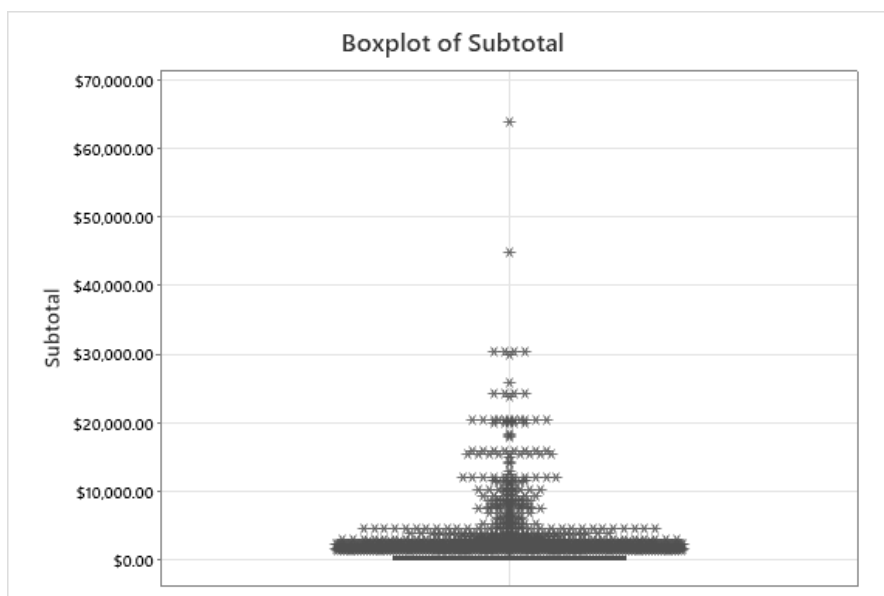
Variables Cuantitativas

Aunque el subtotal es una variable cuantitativa que en teoría podría representarse con un boxplot; sin embargo, el rango del subtotal es tan grande (aún eliminando las ventas departamentales) que no se alcanzan a apreciar los cuartiles y muchos de los datos salen como si fueran atípicos.

Boxplot con ventas departamentales:



Boxplot sin ventas departamentales:



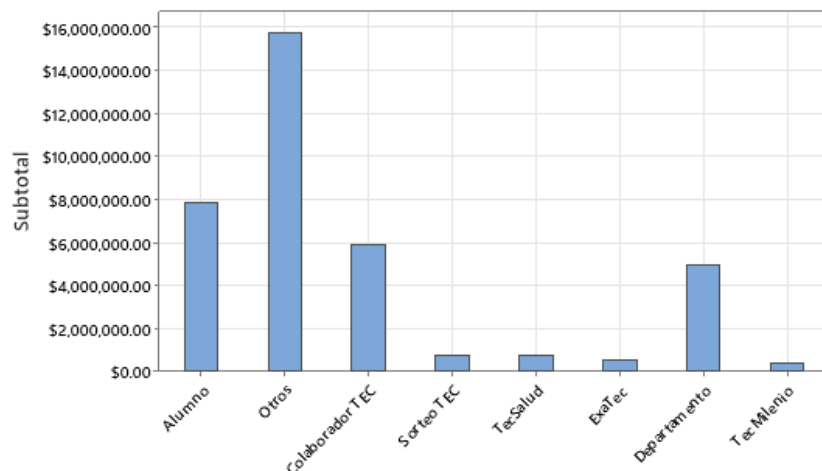
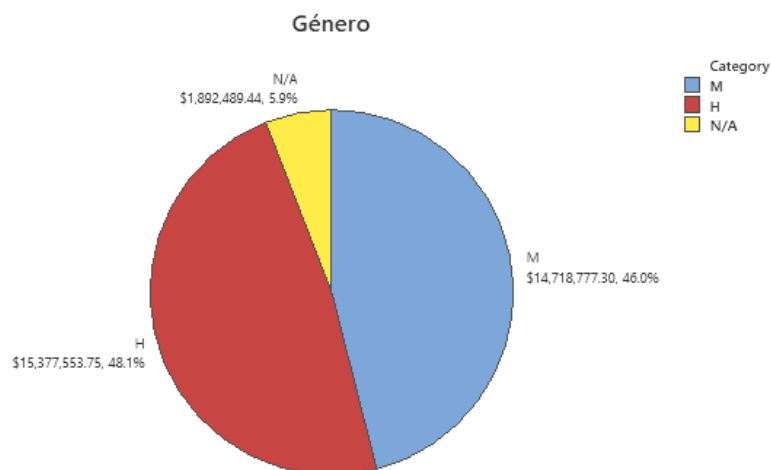
Por ende, la mejor opción para analizar el subtotal y el ticket promedio sería mostrar la analítica descriptiva y los gráficos utilizando las variables cualitativas; ya que las variables cuantitativas no nos dicen mucho sin analizar las categorías/variables cualitativas. Se excluyeron las ventas departamentales.

Statistics

Variable	Mean	StDev	Sum	Minimum	Median	Maximum	Range	Mode
Subtotal	498.6	840.6	31988919.0	0.00	400.0	64000.0	64000.0	650
CORRECTOTicketPromedio	498.6	840.6	31988820.5	0.00	400.0	64000.0	64000.0	650

Variable	N for Mode
Subtotal	10012
CORRECTOTicketPromedio	10012

SUBTOTAL:



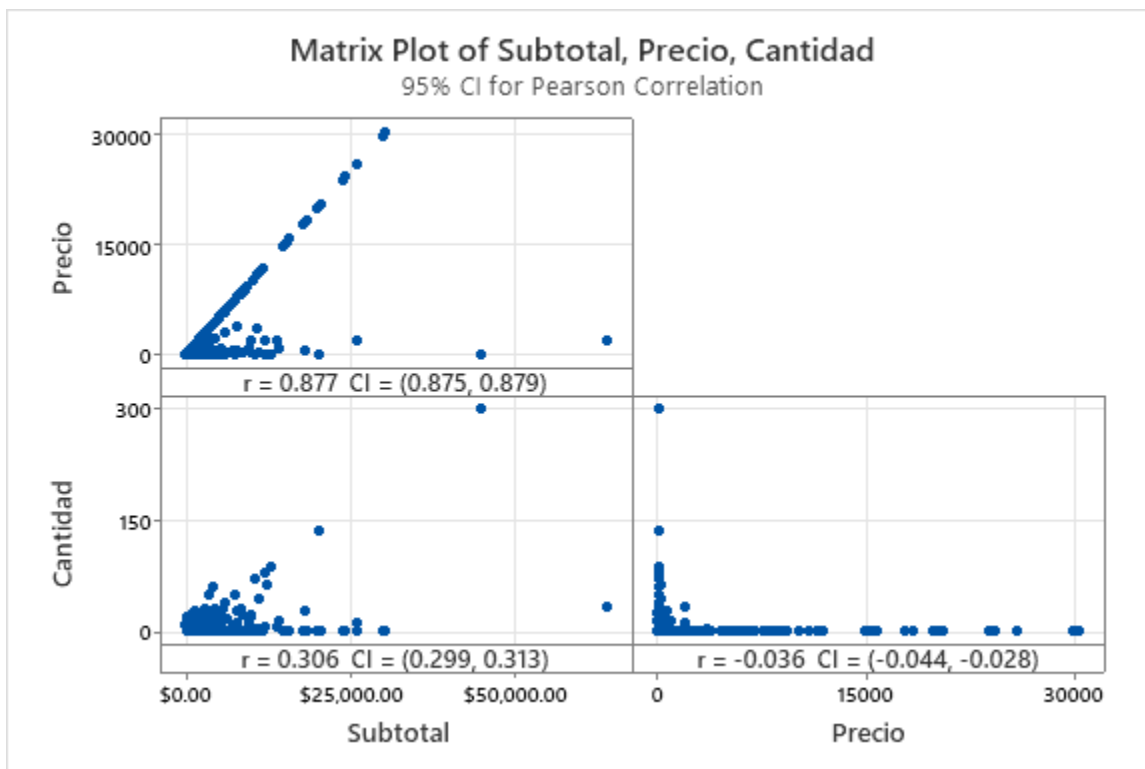
Parte 3: Correlación y Regresión Lineal

Utiliza al menos un análisis de correlación y/o de regresión lineal simple para explicar la relación entre dos variables cuantitativas. Interpreta la ecuación de regresión en términos del contexto del estudio.

Se hizo un análisis de correlación entre las tres variables cuantitativas que tenemos en la base de datos: subtotal, precio y cantidad. Se excluyeron las ventas departamentales.

Correlations

	Subtotal	Precio
Precio	0.877	
Cantidad	0.306	-0.036



Podemos observar que sí existe una gran correlación entre precio y subtotal, y una pequeña relación entre subtotal y cantidad. El subtotal es directamente proporcional a precio y cantidad.

Como sí existe una correlación entre las variables, probaremos hacer regresiones lineales simples y una múltiple para encontrar la variable respuesta: subtotal.

Subtotal y Precio

Regression Equation

$$\text{Subtotal} = 29.44 + 1.00016 \text{ Precio}$$

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
404.317	76.87%	76.87%	76.86%

Subtotal y Cantidad

Regression Equation

$$\text{Subtotal} = 327.14 + 147.68 \text{ Cantidad}$$

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
800.244	9.37%	9.37%	9.28%

Subtotal, Precio y Cantidad

Regression Equation

$$\text{Subtotal} = -166.47 + 1.01403 \text{ Precio} + 163.094 \text{ Cantidad}$$

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
287.734	88.28%	88.28%	88.11%

Con la R cuadrada (r-sq) podemos ver que el mejor modelo es el de regresión lineal múltiple.

Esta ecuación nos dice lo siguiente:

- Cuando precio y cantidad son iguales a cero, el subtotal es igual a -166.47.
- El subtotal aumenta 1.01403 unidades por cada unidad de precio.
- El subtotal aumenta 163.094 unidades por cada unidad de cantidad.

Esta interpretación tiene mucho sentido ya que en todas las compras el subtotal es el resultado de la suma de la multiplicación de cantidad por precio de cada artículo.

Parte 4: Conclusión

Argumenta o justifica como el análisis estadístico realizado en los pasos anteriores puede usarse en la toma de decisiones sobre la problemática del reto.

Como la responsable del análisis de clientes el análisis estadístico realizado anteriormente me permite conocer algunos de los datos más relevantes para mi análisis como género, dominio y edad. Además, he podido explorar las diferentes visualizaciones que podrían tener estos datos y seleccioné la mejor para poder explicar el comportamiento de los datos.

Las variables seleccionadas son de gran valor; ya que son las más importantes para poder conocer al mercado de TECstore y diseñar propuestas efectivas para atraer a nuevos clientes. Además, el boxplot de subtotal podría ser útil para identificar a los clientes que más compran para después analizarlos más a fondo para intentar descubrir qué factores afectan en que compren mucho en TECstore.

Finalmente, la regresión lineal y el análisis de correlaciones se podrá usar a la hora de diseñar las propuestas. Es decir, el objetivo de TECstore es tener el subtotal más alto posible, al utilizar la regresión lineal múltiple se podrá jugar con algunos valores para predecir la cantidad de artículos de cierto precio que se necesitan para llegar al subtotal deseado.

Bibliografía

Instituto Tecnológico y de Estudios Superiores de Monterrey. (2022). *Reto 2022*. Canvas.
Recuperado 3 de septiembre de 2022, de
<https://experiencia21.tec.mx/courses/312534/pages/reto-2022>