



**Evidencia 1. Análisis estadístico de la base de datos**

Gabriel Eduardo Melendez Zavala | A01638293 | IDM

Profesores:

Juan Carlos Lavariega Jarquín

Rafael Benítez Medina

Patricia Domínguez Alavez

**Análisis de datos para la toma de decisiones Gpo 570**

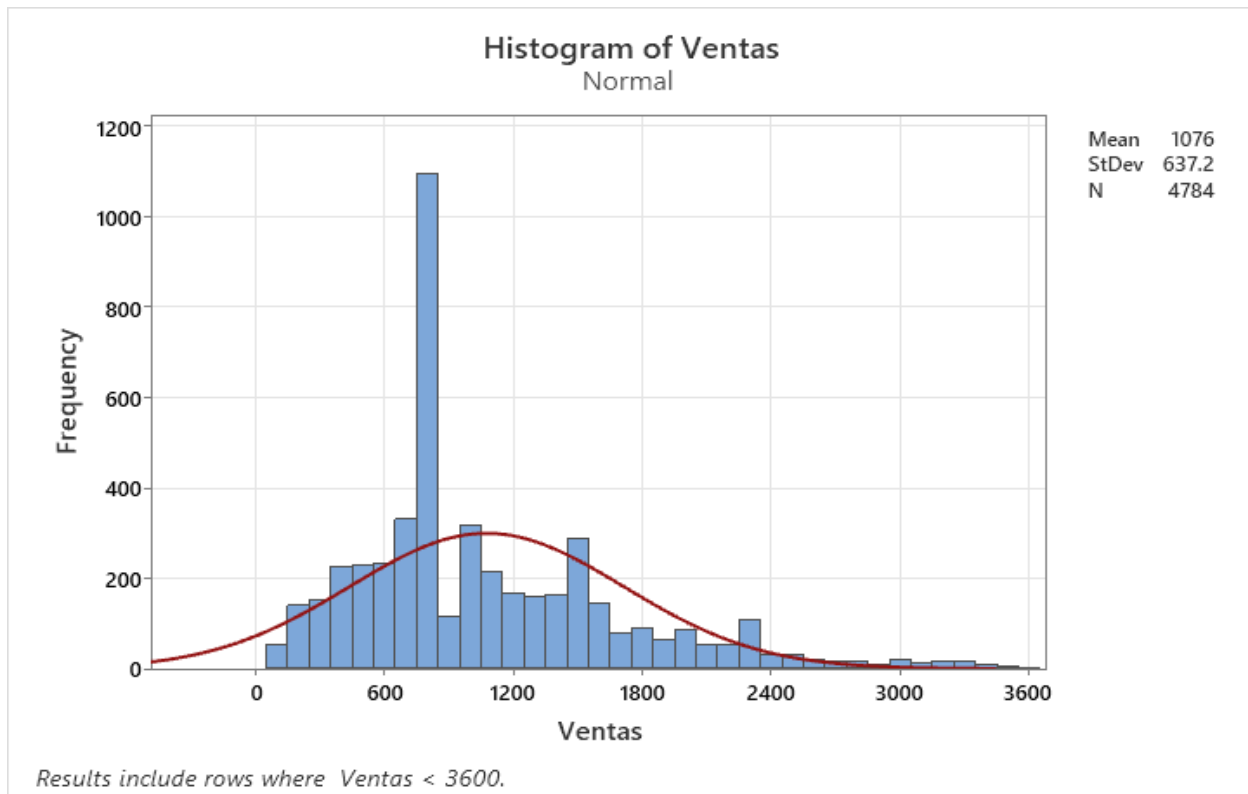
## Evidencia 1. Análisis estadístico de la base de datos

Cualitativas:

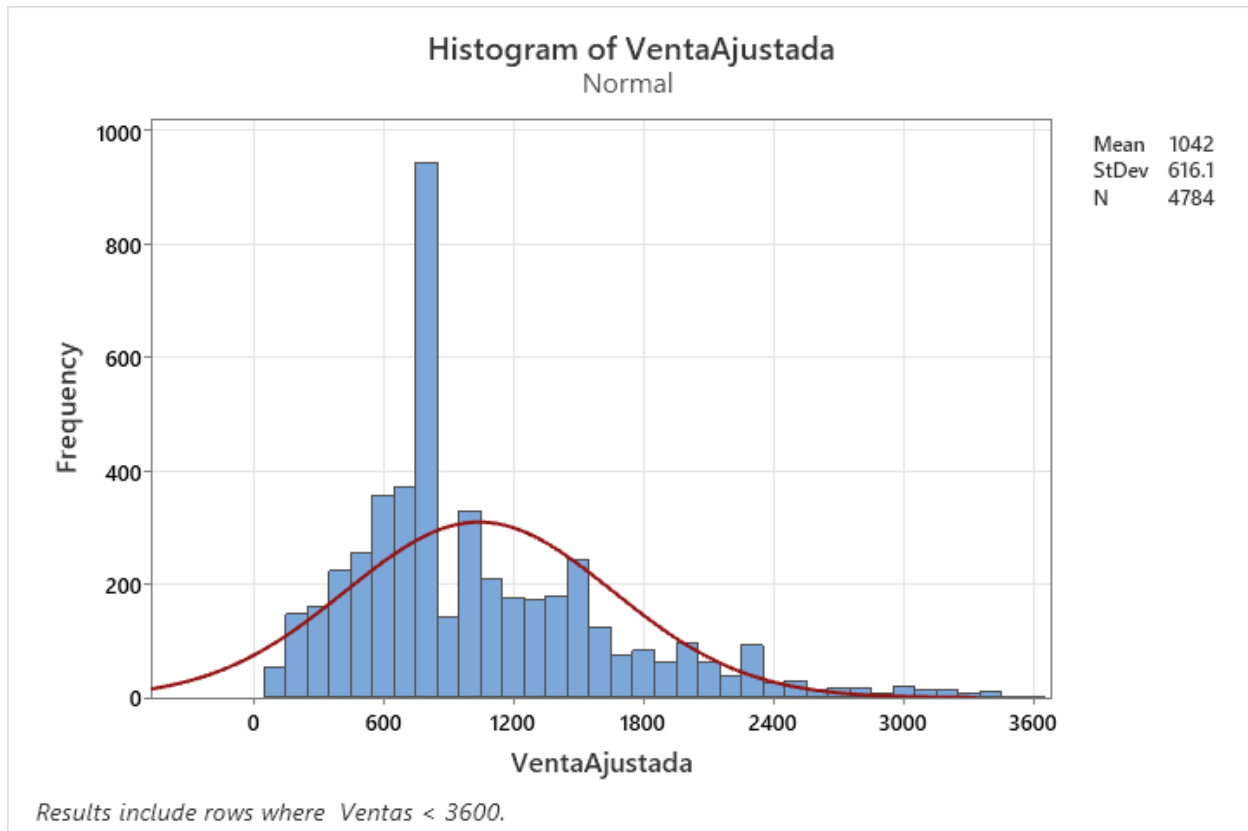
2 variables cualitativas que están relacionadas son el idHora y grupoHora ya que son politómicas, dicotómicas y el idHora es ordinal que se relaciona con el grupoHora que es nominal. Estas variables son relevantes en la situación analizada ya que nos permiten conocer a que grupo horario las ordenes son realizadas.

Cuantitativas:

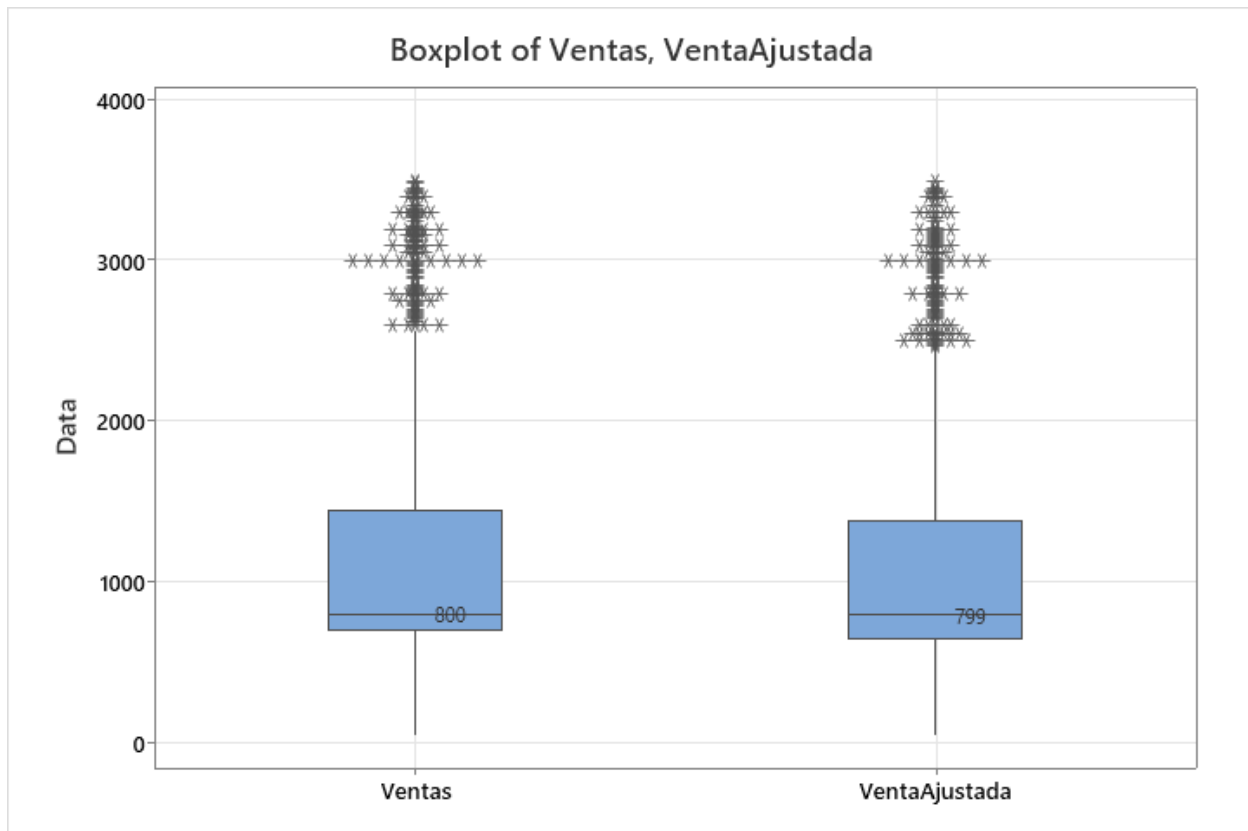
2 variables cuantitativas que se relacionan son las ventas y las ventas Ajustadas, ambas son continuas y están relacionadas dado a que las ventas Ajustadas son la resta de las ventas contra el descuento que fue aplicado. Estas variables son relevantes para el análisis ya que describen el principal factor de desempeño para cualquier negocio, el número de ventas



Este histograma representa la frecuencia de los datos “ventas”. Se puede observar que el promedio de las ventas que son menores a 3,600 es de 1076 son una  $s$  de 637.2.



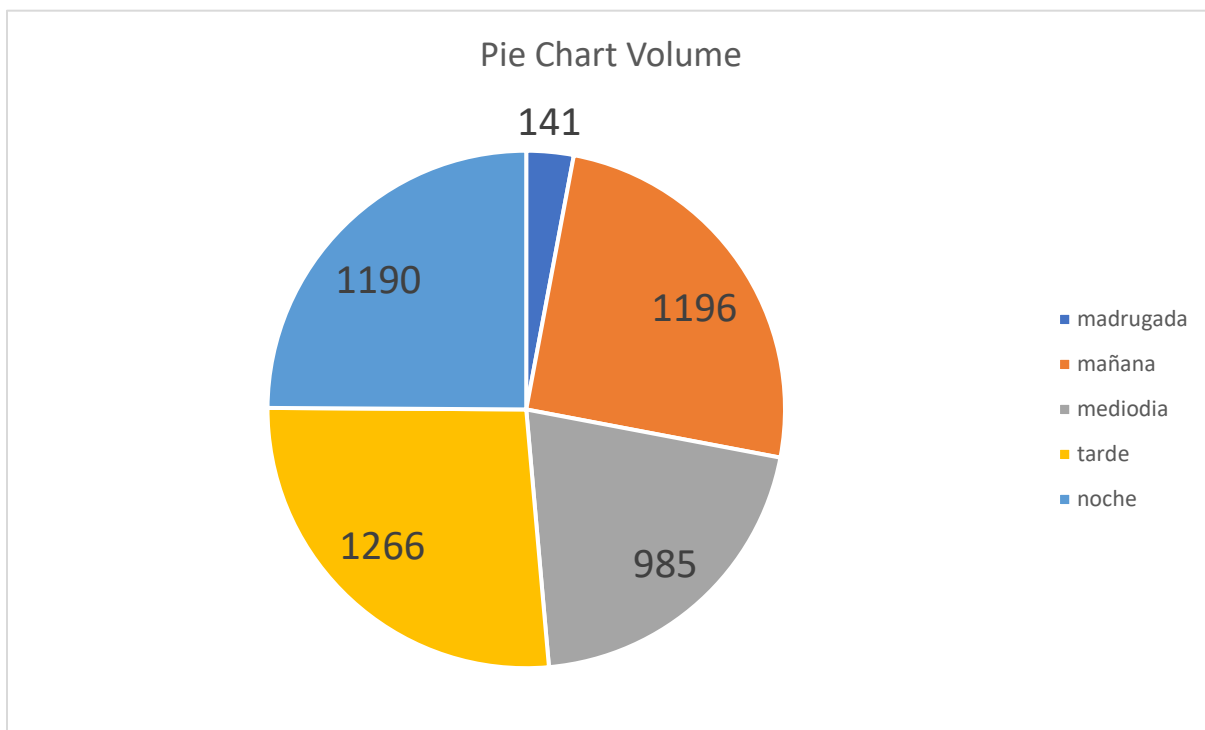
En contraste al histograma de VentaAjustada se puede observar una pequeña diferencia en el promedio de los datos; 1042, así como la  $s$  de 616.1. Esto es el resultado de ajustar las ventas totales con su respectivo descuento para obtener como resultado la “VentaAjustada”.

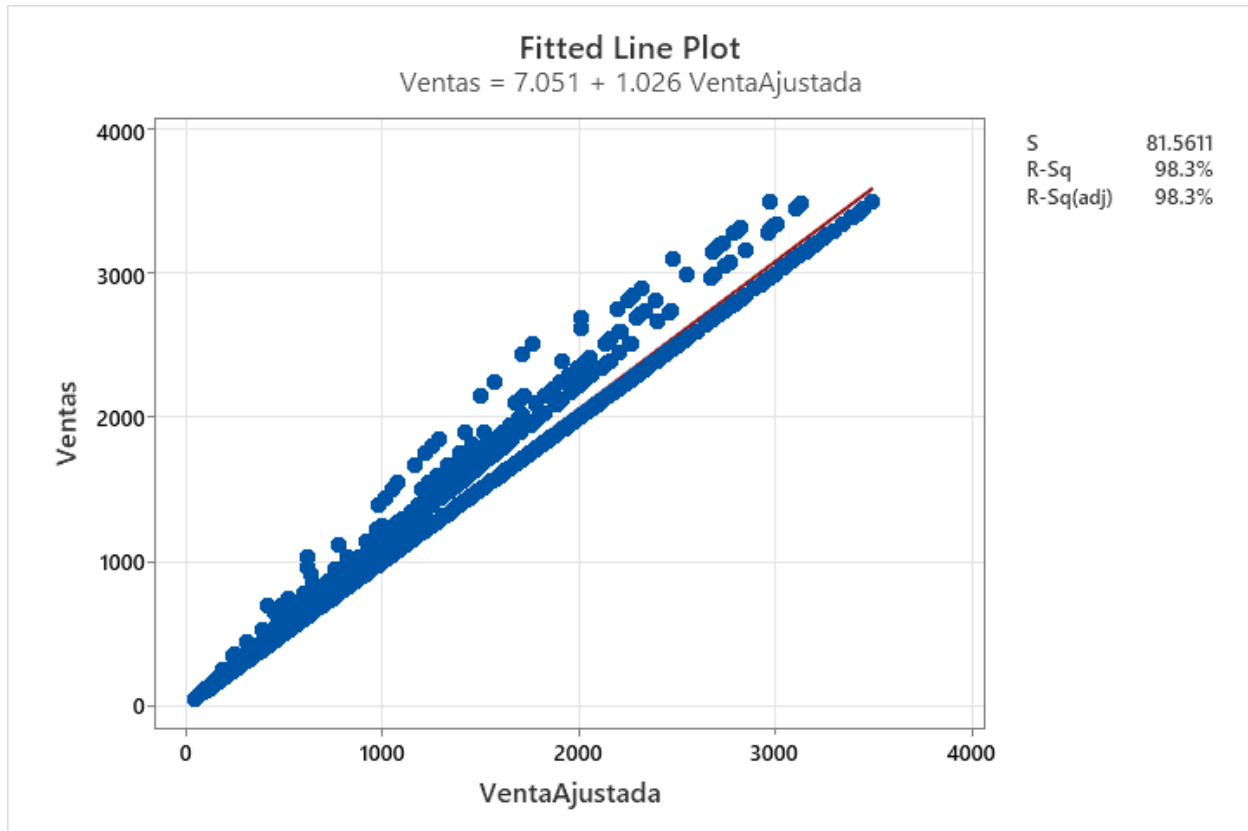


En el boxplot de ventas, ventaAjustada podemos observar que la mediana de los datos se encuentra en 800,  $Q_3 = 1449$  y  $Q_1 = 699$  para Ventas a diferencia de una mediana de 799  $Q_3 = 1377$  y  $Q_1 = 649$  para ventas Ajustadas. Los bigotes de ambos gráficos de caja no varían demasiado con un rango de 50 a 2558 para ventas y de 50 a 2469 para ventas ajustadas. También se puede observar los valores denotados como valores atípicos denotados por el "\*", estos valores representan todos aquellos datos que están por fuera de los bigotes de la caja y comúnmente son limpiados del análisis. Sin embargo, mantuve ciertos valores atípicos para que se pueden ver representados en el grafico de caja y bigote.

madrugada	mañana	mediodia	tarde	noche
141	1196	985	1266	1190

La tabla de frecuencia muestra el numero de ordenes realizadas por grupo, se puede obsevar que con esta infromacion se creo el siguiente diagrama de pie. Es relevante analizar la hora debido a que la tecstore puede organizar sus horarios de trabajo con respecto a su personal sobre las horas del dia donde se esperan más ordenes. Todos los grupos a excepcion de la madrugada se mantienen relativamente estabels con el mismo numero de ordenes. La tarde es el numero con mayores ordenes seguido por la mañana y la noche .





La ecuación de regresión obtenida del gráfico de ventasAjustadas contra ventas indica que las ventas siguen a las ventas ajustadas con la siguiente fórmula:  $y(x) = 7.051 + 1.026x$  esto significa que por cada incremento en ventas ajustadas las ventas incrementan en conjunto por un factor de 1.026.

La correlación entre estas variables o en otras palabras la  $r^2$  es de 0.983 esto significa cuanta proporción de la variable dependiente es explicada por la variable independiente o que tan bien el modelo de regresión se ajusta a los datos.

Prueba de hipótesis para probar la significancia de la regresión,

$H_0: \beta_1 = 0$

$H_1: \beta \neq 0$

## Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	1873397893	1873397893	281620.32	0.000
Error	4776	31770962	6652		
Total	4777	1905168855			

Hay suficiente evidencia estadística para rechazar la hipótesis nula lo que significa que la variable independiente (ventas ajustadas) se relaciona con las ventas con un nivel de significancia del 95%. Cabe mencionar que el p-value es menor que el nivel de significancia .05.

En conclusión, el análisis estadístico de la base de datos de TecStore proporciona información valiosa sobre las relaciones entre las variables cualitativas y cuantitativas. Se ha identificado que las variables idHora y grupoHora están relacionadas, lo que permite conocer a qué grupo horario pertenecen las órdenes. Además, las variables de ventas y ventas ajustadas están estrechamente relacionadas, sin embargo se está realizando un sobreajuste dado que las ventas ajustadas son el resultado de las ventas aplicando el descuento pero dado a que la base de datos carecía de variables cuantitativas para realizar un análisis de regresión adecuado. Asimismo, los histogramas y boxplots revelan diferencias en la distribución de las ventas y las ventas ajustadas, con ligeras variaciones en promedios y rangos. Los valores atípicos se destacan en ambos conjuntos de datos, lo que puede ser importante para futuros análisis.

La tabla de frecuencia y el diagrama de pastel muestran la distribución de las órdenes por grupo horario, lo que podría ayudar a la Tecstore a planificar sus recursos y personal en momentos de mayor demanda. De igual manera, el análisis de regresión revela una fuerte relación entre las ventas y las ventas ajustadas, con un coeficiente de correlación  $r^2$  de 0.983, lo que sugiere que

las ventas siguen de cerca las ventas ajustadas. La prueba de hipótesis respalda la significancia de esta relación, lo que significa que las ventas ajustadas son un predictor significativo de las ventas totales.

En resumen, estos hallazgos brindan una comprensión más profunda de los datos y pueden servir como base para la toma de decisiones estratégicas en la Tecstore, especialmente en lo que respecta a la gestión de horarios y la relación entre ventas y descuentos.