

## Reto: Caso Camaro

Entregable final del reto.  
Documento editable para el participante.

**Nombre del participante:**

**Indicaciones:**

- Guarda este archivo de la siguiente manera: **NombreApellido\_reto\_C1SC1**
- Una vez terminado el reto deberás subirlo a la plataforma.

**Instrucciones:**

Elabora un resumen ejecutivo con medidas de tendencia central, de dispersión y de forma, que describan las variables involucradas en el precio del auto Camaro.

Para lograrlo, lee cada uno de los pasos y realiza lo solicitado.

1. Clasifica cada una de las variables que definen la compra de los carros Camaro, según su naturaleza cuantitativa o categórica.

Dada la lista de variables, clasifícalas en cuantitativas o categóricas y muestra su escala (nominal, ordinal, intervalo o razón).

Cuantitativas	Categóricas
Precio de Venta (Razón), Antigüedad (Razón) y Millaje (Razón)	Condición (Ordinal), Vendedor (Nominal)

2. Analiza las variables cuantitativas y responde las preguntas.

Variable cuantitativa	¿Presenta datos atípicos?	¿Cuántos?	¿Cómo se observa?	¿Qué valores tienen?
Precio de Venta	Si	2	Utilizando graficas boxplots	Mayores de 16K
Antigüedad	No	0		
Millaje	Si	2	Utilizando graficas boxplots	Mayores de 96

Variable cuantitativa	¿Presenta sesgo?	¿Cómo se manifiesta?
Precio de Venta	Si – Positivo	La media es mayor que la mediana, y el parámetro sesgo tiene un valor dentro del rango (.5, +Inf)
Antigüedad	Si - Negativo	Presenta sesgo dado que la media es menor que la mediana. No obstante, el parámetro sesgo está dentro de [-.5, .5]
Millaje	No	La media, la moda y la mediana tienen el mismo valor. Además, el parámetro sesgo está dentro de [-.5, .5]

Variable cuantitativa	¿Sigue una distribución normal?	Justificación
Precio de Venta	Si	Se acepta la Ho usando Anderson-Darling, con un p-value de .01
Antigüedad	Si	Se acepta la Ho usando Anderson-Darling, con un p-value de .10
Millaje	Si	Se acepta la Ho usando Anderson-Darling, con un p-value de .15

Variable cuantitativa	¿Qué medida de tendencia central describe mejor a la variable?	Justificación
Precio de Venta	Mediana	Dado el sesgo positivo que presenta la distribución de la variable.
Antigüedad	Media	Dado que su distribución no presenta sesgo, que es un valor continuo y que es la métrica que mejor representa la edad del conjunto de Camaros a analizar.
Millaje	Media	La media, la moda y la mediana tienen el mismo valor.

Variable cuantitativa	¿Qué medida de variabilidad describe mejor a la variable?	Justificación
Precio de Venta	Rango Intercuartil	La presencia de outliers y el marcado sesgo hacia valores por debajo de la mediana tienen como consecuencia que no se pueda usar la desviación estándar para interpretar correctamente la variabilidad de los datos.
Antigüedad	Desviación Estándar	Dado que no se presentan outliers, la desviación estándar es un buen parámetro para medir la dispersión de los datos relacionados con la edad de los Camaros a analizar.
Millaje	Rango Intercuartil	La presencia de outliers afecta la interpretación real de la desviación estándar. Por lo tanto, es preferible usar el rango intercuartil para entender la variabilidad de los datos.

Utilizando Chebyshev calcula el intervalo a dos desviaciones estándar de distancia de la media e interpreta su significado.

Resultados de los cálculos de intervalos de  $n$  desviaciones estándares de distancia desde la media:

Case	Mean	STDs	Sigmas	Lo_Limit	Hi_Limit	n_Elements	Percentage
Price	8053.45	4380.27	1	3673.18	12433.72	23	0.74
Price	8053.45	4380.27	2	-707.09	16813.99	29	0.94
Price	8053.45	4380.27	3	-5087.36	21194.26	31	1.00
Age	4.84	2.00	1	2.84	6.84	21	0.68
Age	4.84	2.00	2	0.84	8.84	29	0.94
Age	4.84	2.00	3	-1.17	10.84	31	1.00
Mileage	54.90	25.48	1	29.42	80.38	21	0.68
Mileage	54.90	25.48	2	3.95	105.86	29	0.94
Mileage	54.90	25.48	3	-21.53	131.34	31	1.00

Interpretación:

Al observar la tabla anterior, se puede concluir que las distribuciones para las variables Antigüedad y Millaje son normales, dado que cumplen la regla empírica para conjuntos de datos en forma de campana, tabla siguiente.

Por otro lado, el caso de la variable Precio de Ventas, presenta un sesgo positivo que tiene como consecuencia el esperar que la variable en mención no cumpla con la regla empírica antes mencionada. Sin embargo, al evaluar la regla, así como el teorema de Chebyshev, tabla siguiente, se puede concluir que la variable cumple con ambos criterios; esto debido a que la variable presenta una distribución leptocúrtica, que tiene como consecuencia el que la mayor cantidad de precios de ventas a analizar se encuentren en el intervalo  $[\mu - 1*\sigma, \mu + 1*\sigma]$

Intervalo	Desigualdad de Chebyshev Para cualquier conjunto de datos	Regla empírica Para conjuntos de datos forma campana
$\mu \pm 1\sigma$		$\approx 68\%$
$\mu \pm 2\sigma$	$\geq 75\%$	$\approx 95\%$
$\mu \pm 3\sigma$	$\geq 89\%$	$\approx 100\%$

- Analiza las variables categóricas. Elige las medidas resumen más adecuadas y describe cada una de ellas.

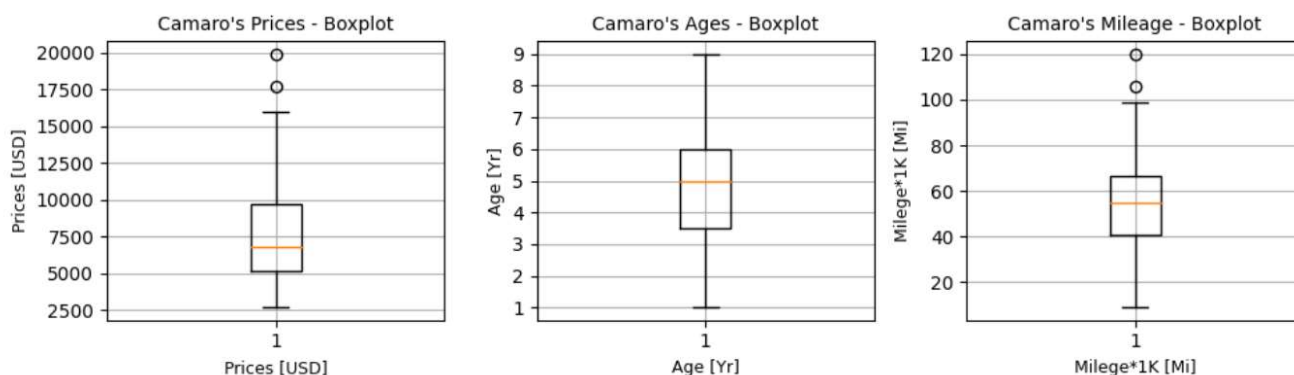
Variable cualitativa	Medidas de resumen más adecuadas	Descripción
Condición	Moda	Medida que permite ver el valor que se repite mas veces dentro de la variable categórica a analizar. También se usa la frecuencia de aparición de los valores, para si relacionarlos con parámetros como Millaje y Antigüedad, ya que la condición, como es de esperar, debe de ir de la mano de las variables antes mencionadas.
Vendedor	Moda	La medida de resumen se selecciono para determinar si existe algún sesgo, en los datos proporcionados, hacia alguna de las clases representadas dentro de la variable en cuestión, ya que al ser una variable binaria se puede presentar el caso.

- Construye un reporte estadístico. Resalta los hallazgos más importantes del análisis que realizaste y que serían factores clave en la elección de un carro Camaro.

*Las respuestas previamente presentadas se basan en el siguiente reporte:*

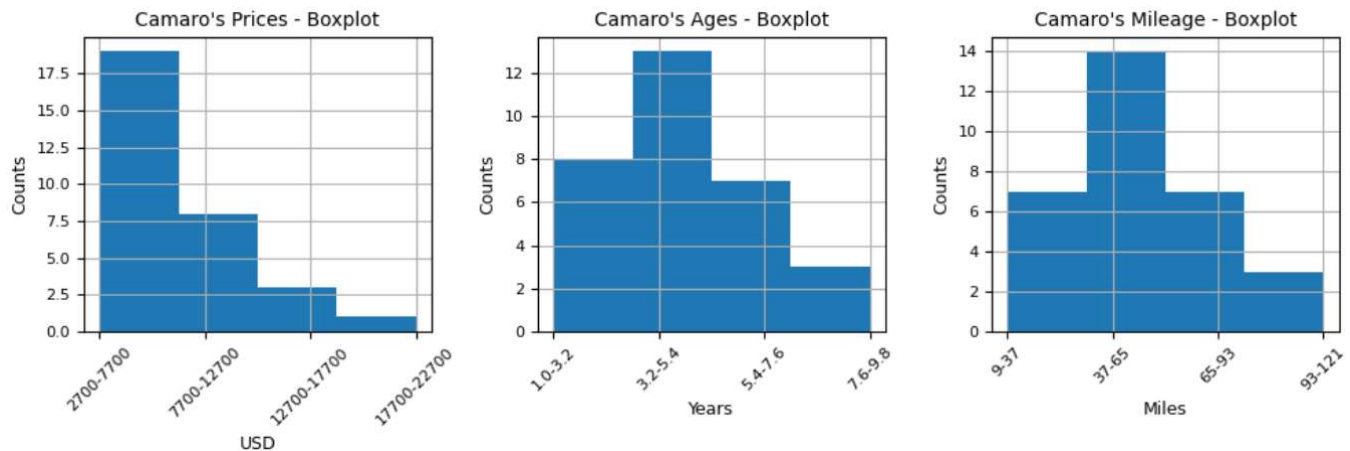
**Análisis de comportamiento estadístico de las variables numéricas:**

- **Boxplots**



*Se puede observar en las graficas presentadas que existes anomalías en las distribuciones relacionadas con las variables Precios de Venta y Millaje. Además, se puede observar ciertos sesgos en el tamaño de los bigotes de los boxplots así como en la mediana de la distribución intercuartil.*

## - Histograms



Una vez agrupados los elementos de las variables numéricas a analizar, se puede observar que la distribución de dichos elementos parece Gaussiana. No obstante, un análisis más detallado necesita efectuarse para determinar la validez de las distribuciones Normales.

## - Medidas Estadísticas

	count	mean	std	min	25%	50%	75%	max	median	mode	kurtosis	skew	cv
<b>Price</b>	31.00	8053.45	4380.27	2700.00	5150.00	6800.00	9685.00	19888.00	6800.00	6800.00	0.90	1.10	54.39
<b>Age</b>	31.00	4.84	2.00	1.00	3.50	5.00	6.00	9.00	5.00	4.00	-0.28	0.37	41.37
<b>Mileage</b>	31.00	54.90	25.48	9.00	40.50	55.00	66.50	120.00	55.00	55.00	0.54	0.54	46.41

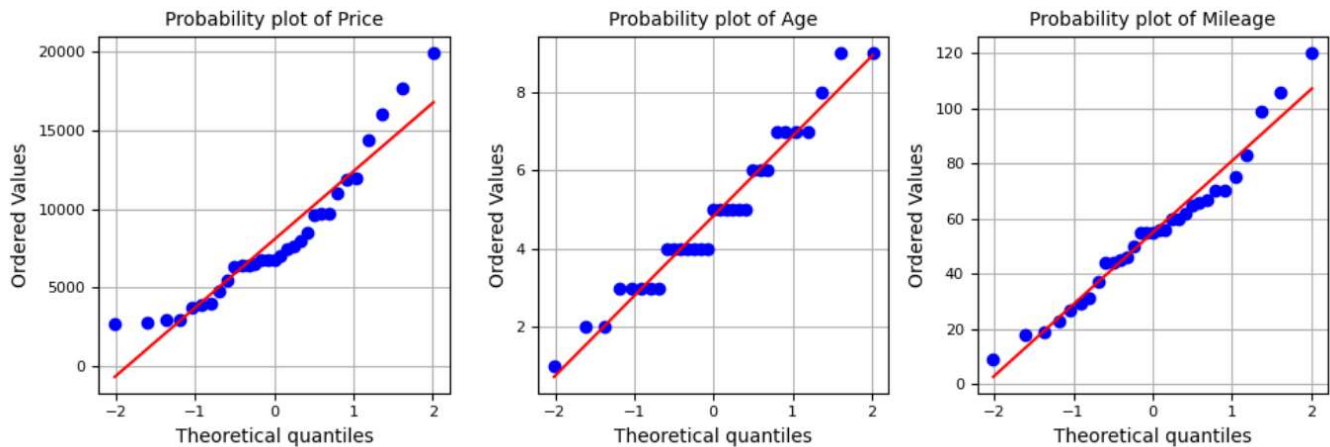
La tabla presentada ayuda a entender mejor el comportamiento de las variables numéricas a analizar. Se puede observar que existe un sesgo positivo en la variable Precio de Ventas y un pequeño sesgo negativo en la variable Antigüedad. También se puede observar que, aunque tienen un comportamiento Normal, existen características dentro de ellas, como la curtosis de Fisher y el Skew, que necesitan ser analizadas a mayor detalle para validar su comportamiento.

## - Verificación de la Normalidad (Anderson-Darling)

	Custom_AD_S	Custom_AD_Asqr	Scipy_AD_Asqr	Scipy_AD_Criticals	Scipy_AD_Significances	p_value	Ho	Hi
<b>Price</b>	-31.96	0.96	0.96	0.522, 0.595, 0.713, 0.832, 0.99	15.0, 10.0, 5.0, 2.5, 1.0	0.01	Accepted	Rejected
<b>Age</b>	-31.52	0.52	0.52	0.522, 0.595, 0.713, 0.832, 0.99	15.0, 10.0, 5.0, 2.5, 1.0	0.10	Accepted	Rejected
<b>Mileage</b>	-31.38	0.38	0.38	0.522, 0.595, 0.713, 0.832, 0.99	15.0, 10.0, 5.0, 2.5, 1.0	0.15	Accepted	Rejected

La normalidad de las distribuciones numéricas a analizar se verificó usando el método de Anderson-Darling. Esto, para aceptar o rechazar una  $H_0$  la cual consiste en que las distribuciones a analizar se comportan como distribuciones Gaussianas. En todos los casos analizados, la  $H_0$  se aceptó, ya que la estadística calculada del método, la  $A^2$ , se encuentra por debajo de los valores críticos que proporciona el método mencionado. En la tabla presentada se pueden observar los valores obtenidos al implementar el método, así como los P-values obtenidos con cada uno de los análisis.

### - Verificación de la Normalidad (QQ plots)



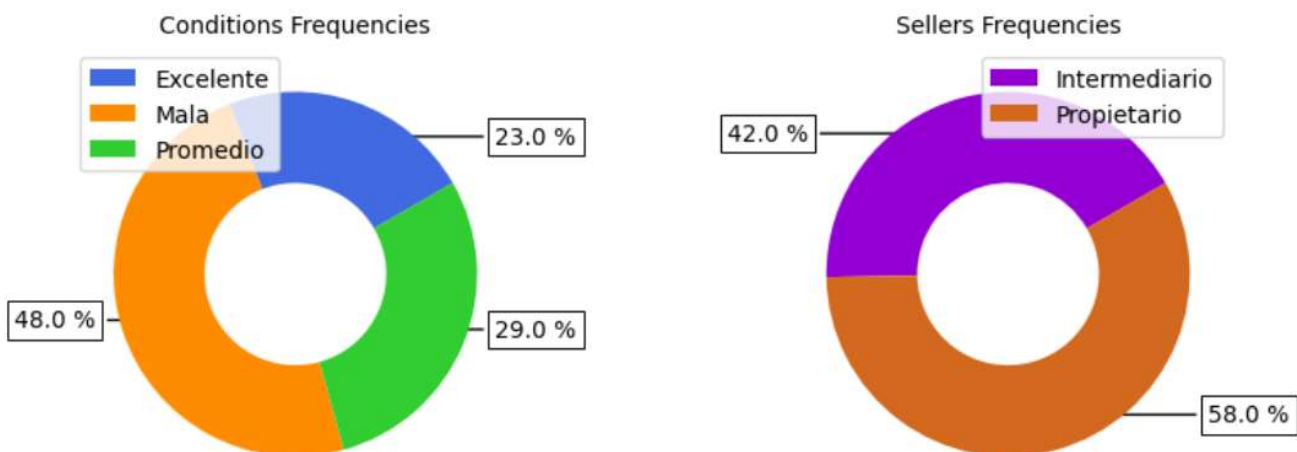
Las graficas QQ presentadas permiten la verificación de la normalidad de las distribuciones asociadas con las variables numéricas de interés. La verificación es posible al ver como se sobreponen los valores que contienen cada una de las variables mencionadas, puntos azules, sobre la línea ideal que representa una distribución normal, líneas rojas.

### - Comportamiento Variables Categóricas

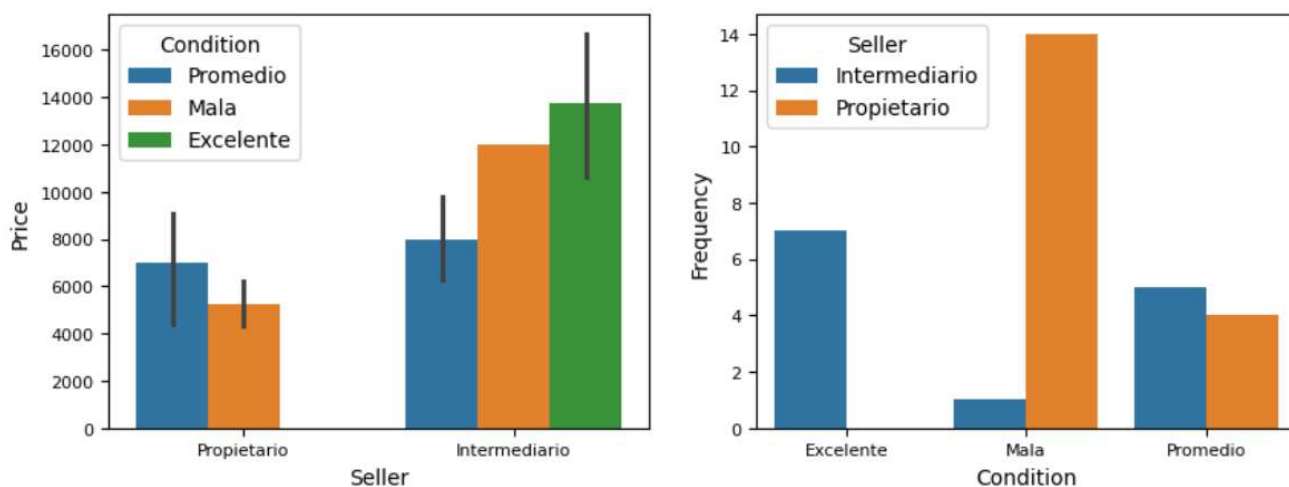
		Frequency
Seller	Condition	
	Excelente	7
	Mala	1
Intermediario	Promedio	5
	Mala	14
	Promedio	4
Propietario	Excelente	7
	Mala	1
	Promedio	5

Las variables categorías se analizaron mediante la frecuencia de los valores que las componen, como se presenta a modo de resumen en la tabla anterior. Es posible observar en los datos proporcionados, que la condición se encuentra estrechamente relacionada con la antigüedad del carro, así como con el Millaje. En conclusión, se puede determinar que un Intermediario posee carros de mejor calidad, donde posiblemente un carro en Mala condición se puede considerar como un valor anómalo. Por otro lado, en el caso de un vehículo vendido por un propietario, se puede concluir que la regla es que esté venda un carro en peor condición que un intermediario.





También se puede observar, mediante la grafica previa, que el mercado de Camaros usados se conforma por casi el 50% de Camaros en mal estado, de los cuales casi el total de ellos son vendidos por propietarios. Mientras que Camaros en una buena o excelente condición, son vendidos por intermediarios. Este ultimo punto puede observarse mejor en la siguiente figura, en su componente a la derecha.



La figura previa, el elemento a la izquierda, nos proporciona información clave para tomar la decisión de donde comprar un Camaro. Esta grafica nos muestra que, en general, Camaros vendidos por sus propietarios pueden ser adquiridos a precios menores que aquellos ofertados por Intermediarios. Sin embargo, existe una alta probabilidad de que el costo sea mayor en las reparaciones requeridas, ya que probablemente la condición de esos Camaros sea cuestionable, en comparación con aquellos vendidos por Intermediarios.

### Recomendación (Conclusión):

La conclusión de este documento es que, para poder hacer una compra eficiente en relación con el costo beneficio de un Camaro a comprar, se debe acudir a un Intermediario. Esto, dado que vehículos ofertados por propietarios tienden a tener una mala calidad, mientras que aquellos que se pueden adquirir mediante intermediarios son mas confiables. La confiabilidad de dichos Camaros, o Condición, va de la mano con la cantidad de Millaaje y



*Antigüedad de los mismos, es decir, aquellos con mejores condiciones serán mas nuevos y menos usados.*

*Es importante mencionar que dentro del conjunto de Camaros ofertados por Intermediarios, lo mas conveniente es adquirir camaros con una condición Promedio, ya que serán ellos los que generen un mayor costo-beneficio con relación a gastar menos y el poder adquirir un vehículo suficientemente confiable.*

*Afortunadamente, el encontrar un Camaro con las características mencionadas en el párrafo anterior es muy factible, ya que casi el 38% del mercado de Intermediarios se encuentra compuesto por los Camaros que deberían ser objetivos del comprador.*