Imagen que contiene Círculo

Descripción generada automáticamenteLogotipo

Descripción generada automáticamente**UNIVERSIDAD TECNOLÓGICA DE PANAMÁ**

**FACULTAD DE INGENIERÍA DE SISTEMAS COMPUTACIONALES**

**LIC. EN INGENIERÍA DE SOFTWARE**

**ESTADISTICA CON APOYO INFORMATICO**

**SEMESTRAL**

**TEMA:**

**REPORTE ESTADÍSTICO DE 10 BASE DE DATOS**

**FACILITADOR: JUAN MARCOS CASTILLO phd**

**INTEGRANTE:**

|  |  |
| --- | --- |
| **PEREZ, OSCAR** | **8-993-1395** |

**GRUPO: 1SF-131**

**Índice**

[**Introducción** 3](#_Toc141228890)

[**Reporte Estadístico de 10 bases de datos** 4](#_Toc141228891)

[**Análisis de las bases de datos** 4](#_Toc141228892)

[**1.** **Malls más grandes del Mundo** 4](#_Toc141228893)

[**2.** **Producto Interno Bruto (PIB) anual por provincia en la República de Panamá** 8](#_Toc141228894)

[**3.** **Indicadores de Fecundidad en la República de Panamá** 12](#_Toc141228895)

[**4.** **Exportaciones por zona francas** 16](#_Toc141228896)

[**5.** **Muertes por cáncer en los estados unidos** 26](#_Toc141228897)

[**6.** **Nacimientos Vivos en la República de Panamá** 33](#_Toc141228898)

[**7.** **Cantidad de calorías en comidas rápida** 39](#_Toc141228899)

[**8.** **Autos** 44](#_Toc141228900)

[**9.** **Exportaciones principales mercaderías** 51](#_Toc141228901)

[**10.** **Precios de Laptops** 0](#_Toc141228908)

[Conclusión 7](#_Toc141228909)

## **Introducción**

Este informe en formato presenta un análisis estadístico basado en 10 bases de datos seleccionadas de diversas fuentes, como INEC y Kaggle. El análisis incluye una variedad de técnicas visuales, como gráficas de barra y pastel, así como histogramas y paretogramas para segmentar la muestra. Se describen las variables, su naturaleza numérica o no numérica, y el formato de los datos. Se investiga la correlación entre variables y se realiza un análisis de regresión lineal, evaluando la relación entre las variables numéricas, utilizando herramientas de Excel para obtener resultados significativos. Se presentan las pruebas estadísticas, gráficas de probabilidad normal, intervalos de confianza y evaluaciones para cada modelo. Este proyecto tiene como objetivo comprender y aplicar diversas técnicas estadísticas para analizar los datos recopilados y obtener conocimientos significativos que puedan ser útiles en la toma de decisiones y la resolución de problemas.

# 

# **Reporte Estadístico de 10 bases de datos**

Para la realización de este reporte se escogieron 10 bases de datos, de las cuales 6 fueron escogidas del Instituto Nacional de estadística y censo (INEC) y las otras 5 provenían de Kaggle. El objetivo de este reporte es llevar a cabo el análisis estadístico de estas 10 bases de datos tomando en consideración distintos aspectos que serán presentados en el siguiente reporte.

**Bases de datos escogidas para el análisis**

|  |  |
| --- | --- |
| INEC | Kaggle |
| PIB por provincia | Malls más grandes |
| Exportaciones por zona francas | Autos |
| Nacimientos | Laptops |
| Fecundidad | Cáncer |
| Exportaciones principales | Calorías en comida rápida |

Para esto se utilizaron herramientas como: Excel y Python junto con un conjunto de librerías.

# **Análisis de las bases de datos**

## **Malls más grandes del Mundo**

En la primera base de datos acerca de los malls más grandes del mundo, se segregó la muestra y solo se escogieron aquellos malls que se habían construido desde el año 2010 en adelante. La selección llevada a cabo tuvo sus bases en el interés y la curiosidad por conocer estos datos.

Fuente: <https://www.kaggle.com/datasets/drahulsingh/largest-malls>

Composición de la base de datos a analizar:

|  |  |  |
| --- | --- | --- |
| **Variable** | **Tipo de Dato** | |
| País | Categórico |
| Ciudad | Categórico |
| Año de Apertura | Numérico |
| Superficie Bruta Alquilable | Numérico |
| Tiendas | Numérico |

La pregunta en la cual se basó el análisis de esta base de datos fue:

* ¿Cuál es la cantidad de tiendas en los países con los malls más grande?

Gráficos Obtenidos a partir del análisis estadístico:

Gráfica 1. Cantidad total de tiendas vs Países

Como china es el país con más tiendas entre los malls más grandes del mundo, se segmentó la muestra. En base a este dato, se hizo el siguiente análisis numérico:

|  |  |
| --- | --- |
| Promedio de tiendas | 1807.5 |
| Desviación Estándar | 2353.913155 |
| Varianza | 863723.8095 |

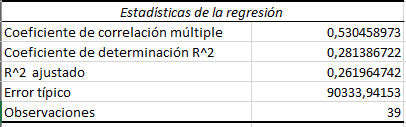
Una vez se observó que China es el país con mayor cantidad de tiendas, se decidió analizar cómo es la dispersión entre las áreas de terreno de cada mall vs la cantidad de tiendas que tiene.

Para este análisis se tomó como base la siguiente pregunta:

* ¿Cómo es la dispersión de los datos entre el terreno vs las tiendas que hay?

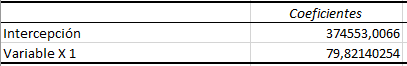
Gráfico 3. Cantidad de tiendas ocupadas en los malls en China vs Terreno disponible

Con estos datos veremos la relación entre el área y las tiendas.



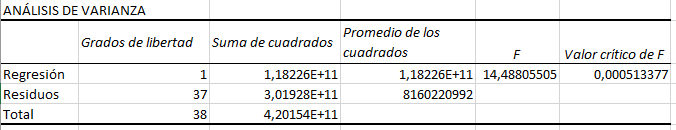
Los resultado nos muestran que no existe correlación entre el terreno de

A continuación, los intervalos de confianza



La tabla de Anova

Análisis de Varianza



A continuación, las gráficas de la probabilidad normal.

Gráfico 3. Suma de la superficie Bruta por País

Gráfico, Gráfico de barras

Descripción generada automáticamente

Se observó que Filipinas es quien más espacio tiene, pero no quien más tiendas tiene. Podríamos decir que China aprovecha mejor el espacio colocando más tiendas.  
Con esto se puede concluir que China es el país con mayor cantidad de tiendas en los malls y quien mejor aprovecha sus terrenos a la hora de colocar tiendas dentro de estos.

## **Producto Interno Bruto (PIB) anual por provincia en la República de Panamá**

El PIB es una medida clave para evaluar el crecimiento económico de un país. Al conocer cómo contribuyen diferentes regiones al PIB total, es posible realizar un seguimiento más preciso del crecimiento económico a nivel local y nacional. Por lo tanto, el objetivo de este análisis busca conocer qué áreas contribuyen más al PIB para así identificar las disparidades económicas y la concentración de la riqueza en ciertas zonas.

Fuente: <https://www.inec.gob.pa/DASHBOARDS/PIB/PIB_ANUAL_POR_PROVINCIA>

Composición de la base de datos a analizar:

|  |  |
| --- | --- |
| **Variable** | **Tipo de Dato** |
| Código Provincia | Numérico |
| Provincia | Categórico |
| Año | Numérico |
| Composición Constante | Numérico |
| Composición Corriente | Numérico |
| Per Cápita Constante | Numérico |
| Per Cápita Corriente | Numérico |
| Valor Constante | Numérico |
| Valor Corriente | Numérico |
| Valor | Numérico |
| Variación Porcentual Constante | Numérico |
| Variación porcentual Corriente | Numérico |

La pregunta en la cual se basó el siguiente análisis fue:

* ¿Qué provincias aparte de la capital, aportan más PIB?

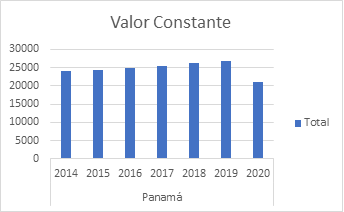
Como sabemos que Panamá es la provincia que más aporta se decidió descartarla y llevar a cabo la confección de un gráfico para descubrir que otras provincias aportan.

Gráfico 4. Suma de PIB vs Provincia

A partir del gráfico anterior se pudo concluir que la provincia de Colón aporta en gran cantidad al valor PIB.

1. ¿En qué año Panamá tuvo su pico de Valor Constante?

Una pregunta que surge, al ver más a fondo el valor Constante de Panamá es ¿Cuando alcanzó su punto más alto a través de los años?, la respuesta se puede apreciar en la siguiente gráfica.



Podemos observar que el punto más alto del valor constante fue en 2019, sin embargo, algo que también resulta llamativo es que en 2020 fue su punto más bajo, probablemente debido a los efectos de la pandemia del COVID-19.

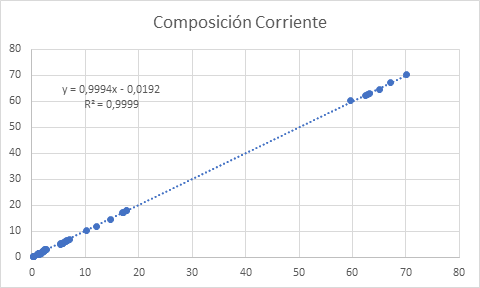
1. ¿En qué escenario, Panamá se encuentra más igualado con otras provincias?

Revisando las distintas gráficas, surgió la curiosidad de ver en qué caso, Panamá se encontraba con valores más cercanos a otras provincias.

Aquí podemos notar que en el valor Per Cápita por provincia, la provincia de Panamá y Colón, se encuentran bastante igualadas entre sí, pero aún, muy alejadas de las demás provincias.

1. ¿Cuál es la relación entre composición corriente y composición constante?

A continuación, se muestra una gráfica para representar la regresión lineal realizada entre la composición corriente y la composición constante.



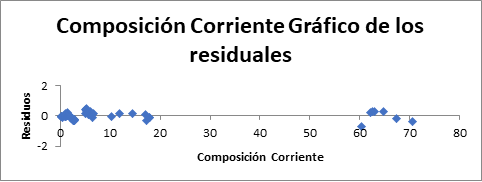
|  |  |
| --- | --- |
| Estadísticas de la regresión | |
| Coeficiente de correlación múltiple | 0.99993539 |
| Coeficiente de determinación R^2 | 0.99987078 |
| R^2 ajustado | 0.99986888 |
| Error típico | 0.21505224 |
| Observaciones | 70 |

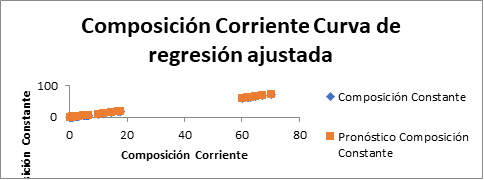
Estos resultados indican que la regresión lineal entre la composición corriente y la composición constante es altamente significativa y que el modelo de regresión explica prácticamente toda la variabilidad en la composición corriente. Es probable que la composición constante sea un predictor muy fuerte de la composición corriente, y el modelo tiene un alto poder predictivo en función de los valores ajustados y ajustados del coeficiente de determinación (R^2).

A continuación, los intervalos de confianza

|  |  |
| --- | --- |
|  | Coeficientes |
| Intercepción | 0.020456863 |
| Composición Corriente | 1.00048838 |
|  |  |

Gráficos de Regresión Lineal





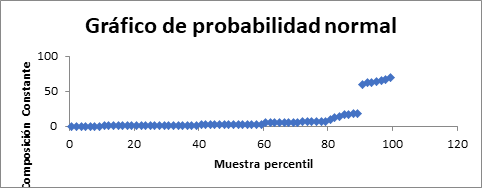


Tabla Anova

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *Grados de libertad* | *Suma de cuadrados* | *Promedio de los cuadrados* | *F* | *Valor crítico de F* |
| *Regresión* | 1 | 24334.52228 | 24334.52228 | 526180.6793 | 5.8739E-134 |
| *Residuos* | 68 | 3.144827585 | 0.046247464 |  |  |
| *Total* | 69 | 24337.6671 |  |  |  |

El valor F calculado es 526180.6793 y el valor crítico de F es 5.8739E-134. Dado que el valor F calculado (526180.6793) es significativamente mayor que el valor crítico de F (5.8739E-134), podemos rechazar la hipótesis nula y afirmar que la regresión es altamente significativa.

En otras palabras, hay evidencia estadística suficiente para concluir que existe una relación significativa entre la composición corriente y la composición constante en base a este análisis de varianza (ANOVA). La regresión entre ambas variables es estadísticamente significativa y, por lo tanto, el modelo de regresión es adecuado para explicar la relación entre ellas.

## **Indicadores de Fecundidad en la República de Panamá**

La base de datos contiene información sobre indicadores demográficos relacionados con la fecundidad en distintas ubicaciones y años. Incluye variables numéricas como la edad media de la fecundidad, tasas de natalidad y fecundidad. El análisis de estos datos permite identificar tendencias y patrones en la dinámica reproductiva de la población, brindando información valiosa para la formulación de políticas y la planificación familiar. Se decidió evaluar y comparar la fecundidad de una de las comarcas del país vs una provincia y compararlas a través de los años 2010-2019. El objetivo de este análisis se basó en ver cómo influye la taza de la fecundidad en el crecimiento de la población.

Fuente:<https://www.inec.gob.pa/DASHBOARDS/Vitales/EV_INDICADORES_FECUNDIDAD>

|  |  |
| --- | --- |
| **Variable** | **Tipo de Dato** |
| Ubicación | Categórico |
| Año | Numérico |
| Edad Media de la fecundidad | Numérico |
| Tasa Bruta de Natalidad | Numérico |
| Tasa Bruta de reproducción | Numérico |
| Tasa de Fecundidad General | Numérico |
| Tasa global de fecundidad | Numérico |

Composición de la base de datos a analizar:

La pregunta en la cual se basó el siguiente análisis fue:

* ¿Cómo varía la tasa de fecundidad en áreas de fácil acceso vs áreas de difícil acceso a través de los años?

A partir de los datos presentados se obtuvo el siguiente resultado:

Se le extrajo a estos datos la varianza, la desviación y promedio

|  |  |
| --- | --- |
| Promedio de la tasa global de fecundidad | 2,563030303 |
| Desviación Estándar | 72,10448485 |
| Varianza | 0,439661493 |

Se puede atribuir la diferencia presentada a la falta de educación sexual en las áreas de difícil acceso.

¿A qué se deberá el aumento en la tasa global de fecundidad en Panamá durante el año 2019?

Además, se observó un valor atípico en el año 2019 en la provincia de Panamá donde su tasa global de fecundidad aumentó abruptamente en comparación con los años anteriores.

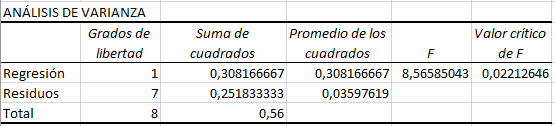
Se utilizarán los datos de los años y la tasa de fecundidad de panamá para hacer la tabla ANOVA y sus gráficos de regresión lineal.

|  |  |
| --- | --- |
| *Estadísticas de la regresión* | |
| Coeficiente de correlación múltiple | 0,741820476 |
| Coeficiente de determinación R^2 | 0,550297619 |
| R^2 ajustado | 0,486054422 |
| Error típico | 0,189673906 |
| Observaciones | 9 |

Vemos que el coeficiente de correlación está en un rengo aceptable, podemos decir que sí hay correlación entre los años y la tasa de global fecundidad.

Gráficos de Regresión Lineal

Tabla ANOVA



## **Exportaciones por zona francas**

La aproximación del análisis estadístico de esta base de datos se fundamenta en un enfoque guiado por la curiosidad acerca del comercio internacional de Panamá, específicamente dirigido a estudiar aspectos relevantes sobre las zonas francas y su impacto en el panorama exportador del país.

Composición de la base de datos a analizar:

|  |  |  |
| --- | --- | --- |
| **Variable** | **Tipo de Dato** | |
| Año | Categórico |
| Continente | Categórico |
| Vía | Categórico |
| Zonas francas | Categórico |
| Puerto de desembarque | Categórico |
| Mes | Categórico |
| Arancel | Numérico categórico |
| Valor FOB | Numérico |
| Peso Neto | Numérico |

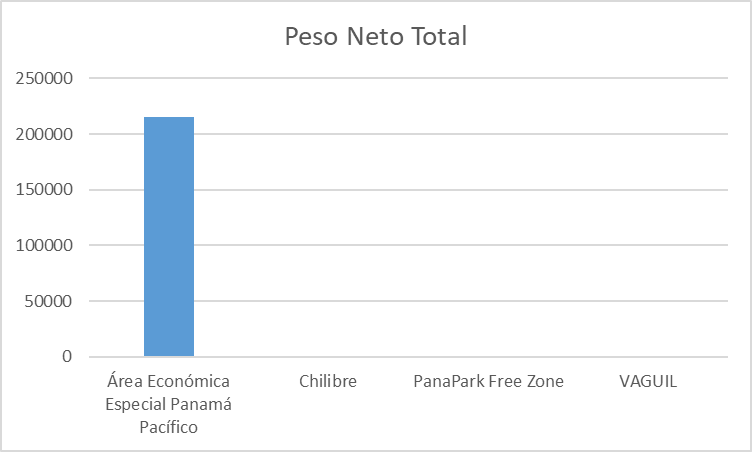
*Análisis de las Bases de Datos*:

Estos análisis se hicieron en Excel.

1. ¿Qué zona franca exporta la mayor cantidad de peso neto?

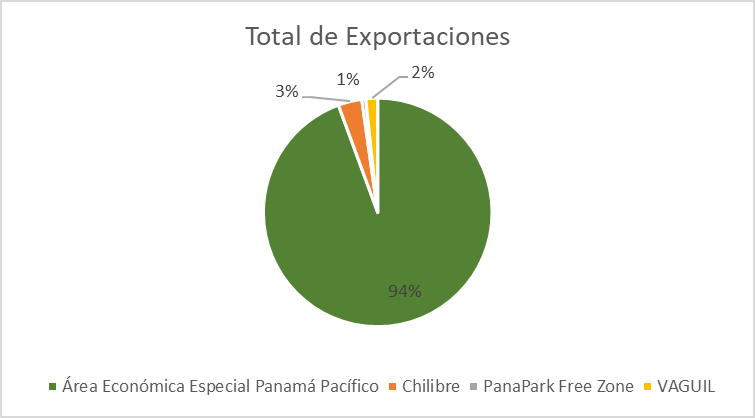
En el gráfico adjunto se observa que la Zona Franca Panamá Pacífico presenta la mayor cantidad de peso neto en las exportaciones del arancel analizado en los últimos años, lo cual destaca como una diferencia significativa en comparación con las demás zonas francas.

|  |  |
| --- | --- |
| **Etiquetas de fila** | **Suma de Peso neto** |
| Área Económica Especial Panamá Pacífico | 215237 |
| Chilibre | 1039 |
| PanaPark Free Zone | 25 |
| VAGUIL | 6 |
| **Total general** | **216307** |

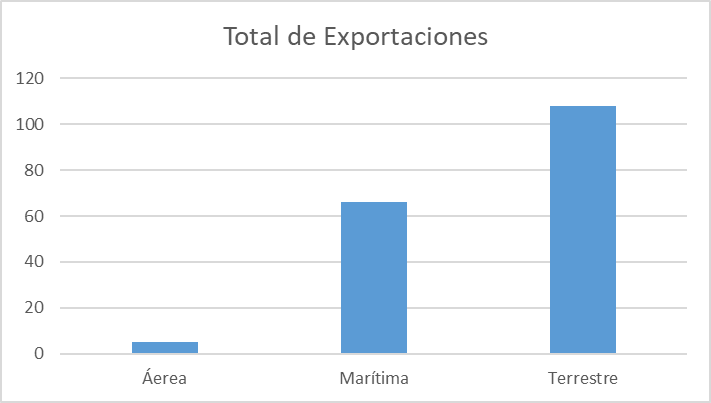


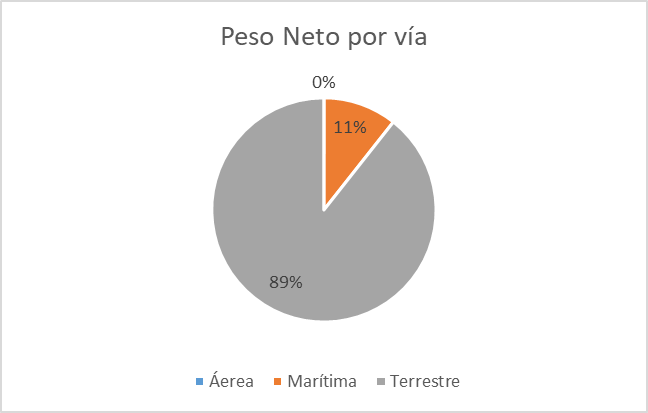
El siguiente gráfico muestra que las exportaciones del arancel 7117.19.00.00 en la Zona Franca de Panamá Pacífico representan el 94% del total de exportaciones de dicho arancel durante los años 2017-2021. Debido a este hecho, el análisis a continuación se centrará exclusivamente en esta Zona Franca, sin considerar las demás.

|  |  |
| --- | --- |
| **Etiquetas de fila** | **Recuento de Zonas Francas** |
| Área Económica Especial Panamá Pacífico | 169 |
| Chilibre | 6 |
| PanaPark Free Zone | 1 |
| VAGUIL | 3 |
| **Total general** | **179** |



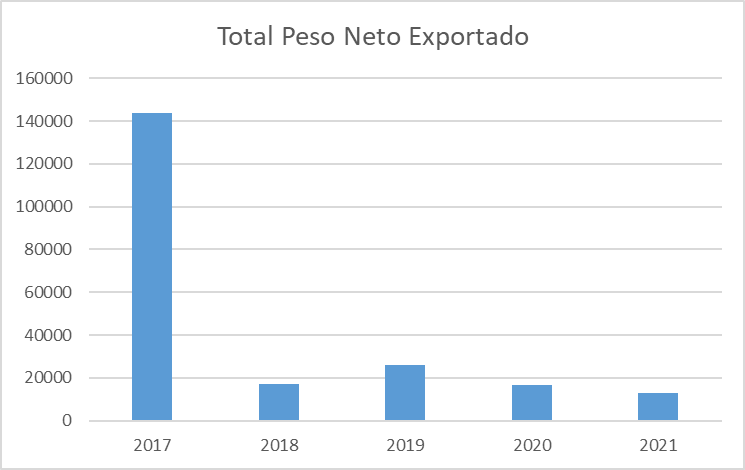
1. ¿Por cuál vía de transporte se suele exportar este arancel?

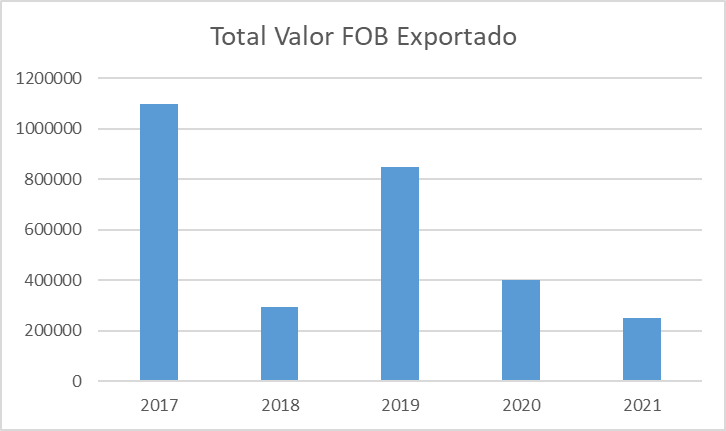




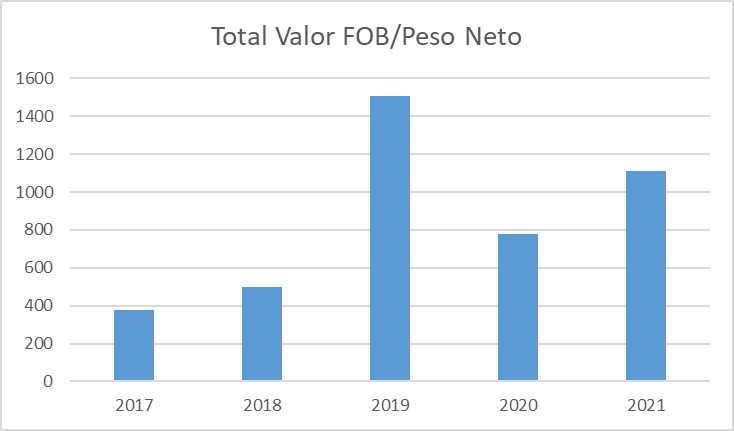
Los gráficos revelan que las vías más utilizadas para la exportación de este arancel son, en primer lugar, la vía terrestre, seguida de la vía marítima. Además, el Peso Neto de las exportaciones se concentra prácticamente al 100% en estas dos vías, lo cual indica una clara predominancia. Por lo tanto, el análisis que se presenta a continuación no tomará en cuenta la vía aérea.

1. ¿Cómo se ha comportado la exportación de este arancel? ¿Cómo se comporta el Peso Neto y el valor FOB de este arancel?





Al analizar el comportamiento de las exportaciones del arancel basándonos en los gráficos, durante los años estudiados se observa una notable disminución tanto en el Peso Neto exportado como en el Valor FOB a partir del año 2018. Sin embargo, el Valor FOB experimenta un aumento en el año 2019, pero vuelve a caer en el año siguiente. Esta fluctuación posiblemente se deba a los efectos de la pandemia, no obstante, también se tendrían que estudiar otros factores como cambios en la demanda del mercado, fluctuaciones económicas, entre otros, pues también pueden haber influido en estas tendencias observadas en los datos.



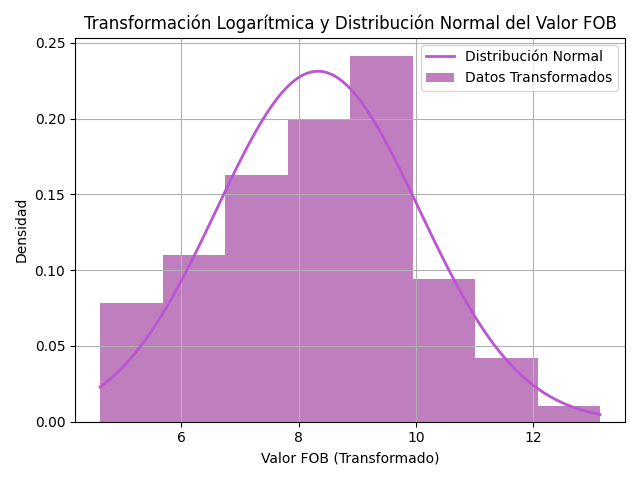
Dado al pico en 2019 en el valor FOB, también se realizó un análisis en el comportamiento del valor FOB entre peso neto y podemos ver en el grafico previo que presenta una tendencia creciente, podíamos decir que esta tendencia se vio afectada en el 2020, esto podría ser producto de la pandemia, pero de todas maneras se mantiene la tendencia de que el valor FOB por peso neto va en aumento de este arancel (bisuterías de metal) cada año.

1. ¿Promedio de valor FOB pagado en el periodo estudiado?

A continuación, se presenta un análisis de los datos numéricos del valor FOB en la siguiente tabla.

|  |  |  |  |
| --- | --- | --- | --- |
| Media | Mediana | Varianza | Desviación Estándar |
| 16364.49 | 4982 | 2084508073 | 45656.41 |

Distribución normal de la columna Valor FOB.



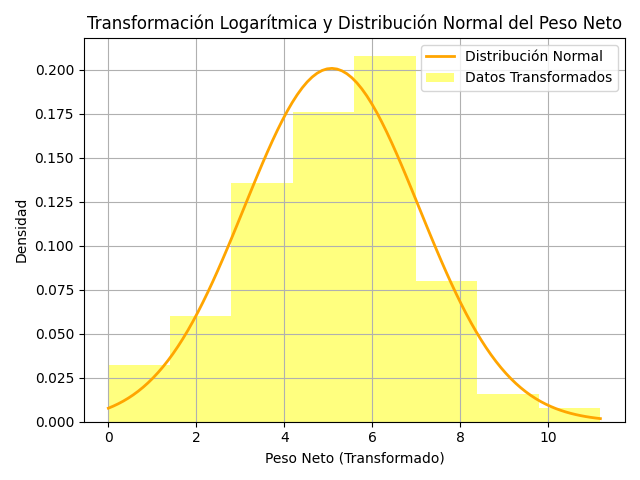
Al examinar los datos del valor FOB en el estudio, se observa que siguen una distribución normal. Sin embargo, es importante señalar que la mayoría de los datos se concentran hacia el lado izquierdo del pico de la distribución normal, lo que sugiere una asimetría hacia la derecha, indicando un sesgo positivo. En otras palabras, hay una extensión de la cola de la distribución hacia los valores más altos, lo que implica que algunas observaciones tienen valores considerablemente más elevados, ejerciendo influencia en esta asimetría en los datos.

1. ¿Promedio de Peso Neto exportado en el periodo estudiado?

A continuación, se presenta un análisis de los datos numéricos del peso neto en la siguiente tabla.

|  |  |  |  |
| --- | --- | --- | --- |
| Media | Mediana | Varianza | Desviación Estándar |
| 1208.42 | 180 | 35336621 | 5944.46 |

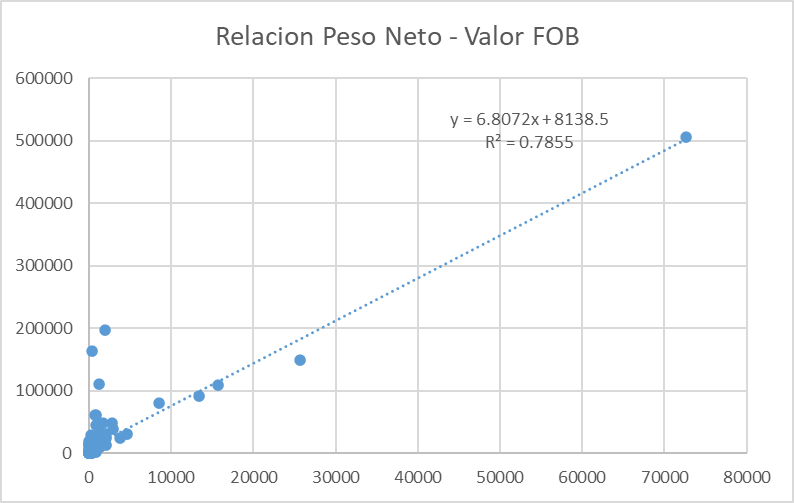
Distribución normal de la columna Peso Neto.



Al examinar los datos del Peso Neto en el estudio, se observa que el histograma muestra una forma similar a la distribución típica en forma de campana que caracteriza a las distribuciones normales. Esto sugiere la posibilidad de que los datos sigan una distribución normal. Sin embargo, para asegurar y confirmar esta suposición, se recomienda realizar otros análisis estadísticos que validen de manera más rigurosa si la distribución se ajusta realmente a una distribución normal. Aunque la distribución es bastante simétrica, todavía presenta un ligero sesgo positivo.

1. ¿Cuál es la relación entre el Peso Neto y valor FOB?

En el gráfico presentado a continuación se muestra la representación gráfica de la regresión lineal que relaciona el Peso Neto y el Valor FOB.



|  |  |
| --- | --- |
| *Estadísticas de la regresión* | |
| Coeficiente de correlación múltiple | 0.88629878 |
| Coeficiente de determinación R^2 | 0.78552553 |
| R^2 ajustado | 0.78431381 |
| Error típico | 21203.764 |
| Observaciones | 179 |

En este caso, se obtuvo un coeficiente de correlación de 0.886 lo que indica que hay una correlación fuerte y positiva entre el Peso Neto y el Valor FOB. Además, un valor de R^2 de 0.785 indica una correlación significativa y fuerte entre el Peso Neto y el Valor FOB. Esto implica lo que significa que aproximadamente el 78.55% de la variabilidad en el Valor FOB puede ser explicada por el Peso Neto y por lo tanto el Peso Neto es un buen indicador o predictor del Valor FOB dentro del marco de la regresión lineal utilizada.

A continuación, los intervalos de confianza para la intersección y la variable x1.

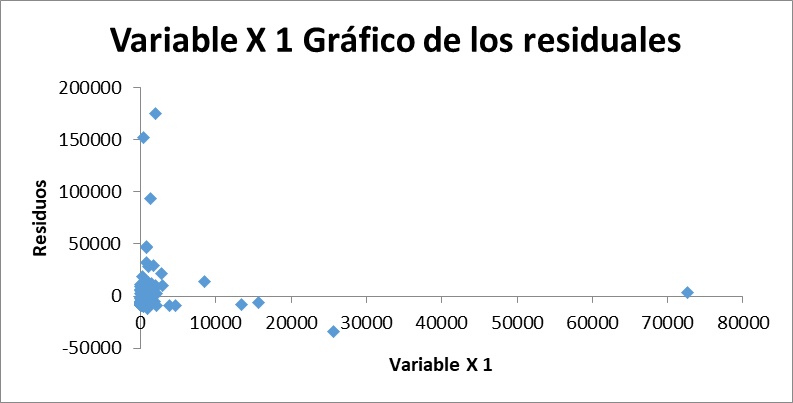
|  |  |
| --- | --- |
|  | *Coeficientes* |
| Intercepción | 8138.51877 |
| Variable X 1 | 6.80721447 |

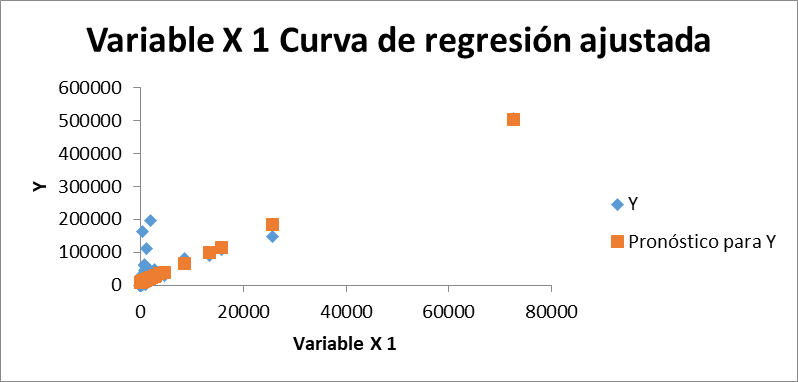
|  |  |  |
| --- | --- | --- |
| Intervalo de confianza B0 | | |
| 3178.38099 | <B0 | 13098.6566 |

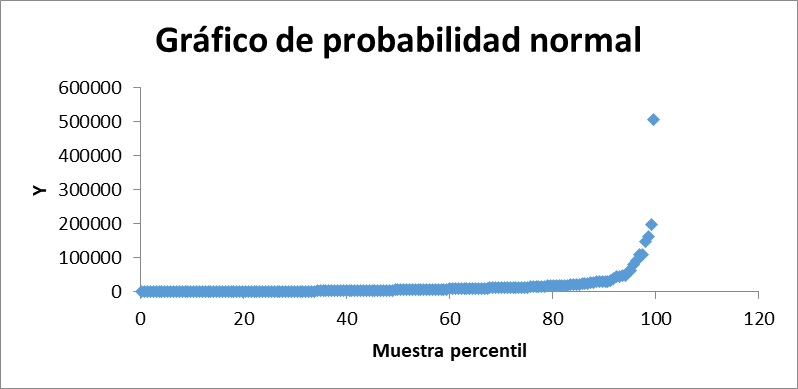
|  |  |  |
| --- | --- | --- |
| Intervalo de confianza B1 | | |
| 6.16643898 | <B1 | 7.44798996 |

El intervalo de confianza al 95% para la intercepción (β0) y el coeficiente de x1 (β1) nos proporciona una estimación que sugiere que, en caso de tomar numerosas muestras distintas y calcular el intervalo de confianza para cada una de ellas, aproximadamente el 95% de dichos intervalos abarcarían los valores reales de la población correspondientes a la intercepción y el coeficiente de x1.

También obtenemos el grafico de residuales y el grafico de probabilidad normal para analizar mejor la regresión.







ANOVA

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *Grados de libertad* | *Suma de cuadrados* | *Promedio de los cuadrados* | *F* | *Valor crítico de F* |
| Regresión | 1 | 2.9146E+11 | 2.9146E+11 | 648.273047 | 4.5273E-61 |
| Residuos | 177 | 7.9579E+10 | 449599606 |  |  |
| Total | 178 | 3.7104E+11 |  |  |  |

*Hipótesis utilizadas*:

* Hipótesis nula (H0): No hay una relación significativa entre el Peso Neto y el Valor FOB en las importaciones realizadas en las zonas francas.
* Hipótesis alternativa (H1): Existe una relación significativa entre el Peso Neto y el Valor FOB en las importaciones realizadas en las zonas francas.

*Resultados de las hipótesis según el análisis ANOVA*:

El análisis de varianza realizado en el modelo de regresión reveló una fuerte correlación positiva entre el Peso Neto y el Valor FOB, indicando que a medida que el Peso Neto aumenta, también tiende a aumentar el Valor FOB de las importaciones.

Dado que el valor obtenido de la estadística F es mayor que el valor crítico de F, se rechaza la hipótesis nula, mostrando que hay suficiente evidencia estadística para afirmar que existe una relación significativa entre el Valor FOB y el Peso Neto, y por tanto el modelo tiene una capacidad significativa para predecir el Valor FOB (variable dependiente) en función del Peso Neto (variable independiente).

Estos hallazgos respaldan la relevancia del peso neto como factor influyente en el valor FOB de las exportaciones realizadas en las zonas francas.

## **Muertes por cáncer en los estados unidos**

Las muertes por cáncer constituyen un importante problema de salud a nivel mundial, siendo una de las principales causas de mortalidad. Factores como el tabaquismo, la dieta poco saludable y la predisposición genética pueden aumentar el riesgo de padecerlo. El cáncer abarca distintos tipos, como el de pulmón, mama, próstata y colon, entre otros. La detección temprana es clave para mejorar el pronóstico y la supervivencia, y los avances en investigación han llevado a tratamientos más efectivos. La promoción de hábitos saludables y la investigación continua son fundamentales para enfrentar esta enfermedad y reducir su impacto. Tomando esto en consideración, se decidió analizar esta base de datos con el fin de detectar que variables influyen de gran manera en las muestras.

Fuente: <https://data.world/nrippner/cancer-linear-regression-modeltutorial/workspace/file?filename=death+.csv>

Composición del documento:

|  |  |
| --- | --- |
| **Variable** | **Tipo de Dato** |
| Country | Categórico |
| FIPS | Categórico |
| Met Objective of 45.5 | Categórico |
| Age-Adjusted Death Rate | Numérico |
| Lower 95% Confidence interval for Death Rate | Numérico |
| Upper 95% Confidence Interval for Death Rate | Numérico |
| Average Deaths per Year | Numérico |
| Recent Trend (2) | Numérico |
| Recent 5-Year Trend (2) in Death Rates | Numérico |
| Lower 95% Confidence Interval for Trend | Numérico |
| Upper 95% Confidence Interval for Trend | Numérico |

|  |  |
| --- | --- |
| Media de Age-Adjusted Death Rate | 53.18514042 |
| Desviación Estándar de Age-Adjusted Death Rate | 16,43237564 |
| Varianza de Age-Adjusted Death Rate | 14.03150255 |

Histogramas de frecuencia

1. ¿Cuál es la tendencia general entre todos los estados?

Para hallar solución a esta pregunta, realizamos un conteo de la categoría de tendencia y obtenemos el siguiente gráfico:

Gráfico de la suma de las distintas tendencias

Podemos resaltar que “stable” es la tendencia que más se repite en los distintos estados, por una gran mayoría.

1. ¿La mayoría de los estados lograron llegar a la meta de bajar la tasa de mortalidad al 45.5?

Para responder esta pregunta, realizamos algo similar al punto anterior, contando todas las ocasiones en las que se cumplió la meta.

Nos percatamos de que la mayoría de los estados no lograron llegar a la meta de tener una tasa de mortalidad de 45.5.

1. ¿Cuál es la relación entre Age-Adjusted Death Rate y Upper 95% Confidence Interval for Death Rate?

A continuación, se muestra una gráfica para representar la regresión lineal realizada entre dichas variables.

|  |  |
| --- | --- |
| *Estadísticas de la regresión* | |
| Coeficiente de correlación múltiple | 0.91501503 |
| Coeficiente de determinación R^2 | 0.8372525 |
| R^2 ajustado | 0.8371946 |
| Error típico | 7.51207879 |
| Observaciones | 2813 |

El valor del coeficiente de correlación múltiple es 0.91501503. Esto indica una relación positiva bastante fuerte entre el "Age-Adjusted Death Rate" y el "Upper 95% Confidence Interval for Death Rate". Un valor cercano a 1 sugiere que ambas variables están estrechamente relacionadas y que la tasa de mortalidad ajustada por edad tiende a aumentar a medida que el intervalo de confianza superior de la tasa de mortalidad también aumenta.

A continuación, los intervalos de confianza

|  |  |
| --- | --- |
|  | *Coeficientes* |
| Intercepción | 1.70959342 |
| Age-Adjusted Death Rate | 1.21387111 |

Gráficos de Regresión Lineal

Tabla ANOVA

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Grados de libertad | Suma de cuadrados | Promedio de los cuadrados | F | Valor crítico de F |
| Regresión | 1 | 24334.52228 | 24334.52228 | 526180.6793 | 5.8739E-134 |
| Residuos | 68 | 3.144827585 | 0.046247464 |  |  |
| Total | 69 | 24337.6671 |  |  |  |

El análisis de varianza (ANOVA) indica que existe una relación estadísticamente significativa entre la tasa de mortalidad ajustada por edad y el intervalo de confianza superior de la tasa de mortalidad. El modelo de regresión es adecuado y explica una gran parte de la variabilidad observada en la tasa de mortalidad ajustada por edad a través de su relación con el intervalo de confianza superior.

## **Nacimientos Vivos en la República de Panamá**

Fuente: <https://www.inec.gob.pa/DASHBOARDS/Vitales/EV_NACIMIENTOS>

Analizar una base de datos de Nacimientos Vivos en la República de Panamá podría ser una tarea relevante y valiosa por diversas razones. Algunas de las posibles razones para llevar a cabo dicho análisis podrían incluir:

* Salud Pública: El análisis de datos de nacimientos vivos podría proporcionar información vital sobre la salud materna e infantil en Panamá.
* Políticas de Bienestar Social: Los datos de nacimientos vivos pueden ser útiles para formular y evaluar políticas públicas relacionadas con el bienestar social, como programas de apoyo a las madres y recién nacidos, incentivos para la planificación familiar y servicios de atención médica prenatal.
* Estudios Demográficos: Los datos de nacimientos vivos son esenciales para llevar a cabo estudios demográficos y proyecciones poblacionales. Esto es importante para comprender cómo la población de Panamá está cambiando con el tiempo y para preparar planes futuros.
* Investigación Científica: Los datos de nacimientos vivos pueden ser una fuente valiosa para investigaciones científicas en áreas como genética, epidemiología y salud pública. También pueden proporcionar información para estudios sobre factores de riesgo durante el embarazo y el impacto de ciertas intervenciones médicas.

La composición de la base de datos a utilizar:

|  |  |
| --- | --- |
| **Variable** | **Tipo de Dato** |
| Año | Categórico |
| Edad de las madres | Categórico |
| Sexo del bebé | Categórico |
| Provincia | Categórico |
| Cantidad de Nacidos | Numérico |

El objetivo de este análisis es determinar la cantidad de nacimientos en cada provincia para identificar la tasa de crecimiento y determinar cuál provincia presenta la mayor tasa de natalidad.

En el año 2020, se observa que la provincia de Panamá registra la tasa de natalidad más alta, evidenciando una diferencia significativa en comparación con otras provincias, tal como se muestra en el histograma. Por esta razón, se decide focalizar el análisis únicamente en la provincia de Panamá para estudiar el aumento de la tasa de natalidad y determinar los rangos de edad en los que las mujeres son más propensas a tener hijos.

Se puede apreciar que el grupo de edad con mayor número de embarazos en la población panameña se encuentra entre los 20 y 34 años, que son los rangos de edad en los que se registra una mayor fecundidad.

La gráfica muestra una tendencia decreciente en los nacimientos en la provincia de Panamá a lo largo del tiempo. Es interesante destacar que la provincia de Panamá representa aproximadamente el 37,17% de todos los nacimientos en el país cada año, con una media de 24,365 nacimientos.

1. **¿Qué provincia posee la mayor cantidad de Nacimientos vivos en la república de Panamá?**

En la siguiente tabla podemos ver los valores de nacimientos vivos por cada provincia en la república de Panamá.

|  |  |
| --- | --- |
| Provincias | Nacimientos vivos |
| 1 | 13387 |
| 2 | 12303 |
| 3 | 15727 |
| 4 | 24714 |
| 5 | 3244 |
| 6 | 4744 |
| 7 | 3374 |
| 8 | 73104 |
| 9 | 12540 |
| 10 | 31208 |

En el siguiente gráfico podemos observar que la provincia con la mayor cantidad de nacimientos vivos en la república de Panamá en la provincia #8 “Panamá”, su capital.

1. **¿Cuál mes es el que contiene la mayor cantidad de nacimientos vivos en la república de Panamá?**

En la siguiente tabla podemos observar la similitud de la cantidad de nacimientos por mes en la república de Panamá. Se puede notar a su vez que el mes que contiene menos días es de igual forma el mes con menos nacimientos.

|  |  |
| --- | --- |
| Todas las Provincias | |
| Mes | Total |
| 1 | 16680 |
| 2 | 14463 |
| 3 | 15737 |
| 4 | 15480 |
| 5 | 15966 |
| 6 | 15395 |
| 7 | 16304 |
| 8 | 16652 |
| 9 | 17244 |
| 10 | 17473 |
| 11 | 16745 |
| 12 | 16206 |

En la siguiente gráfica de frecuencia nacimientos/meses se puede observar mejor el comportamiento de los nacimientos a lo largo de los 12 meses del año notando que su mayor pico de encuentra entre los meses 9 y 10 (septiembre y octubre)

En esta gráfica de frecuencia relativa encontramos los porcentajes de nacimientos en cada mes notan una diferencia de 0.0155 del mes con mayor tasa de natalidad contra el menor.

Por último, nuestro análisis numérico quedaría de la siguiente manera:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Media | Mediana | Varianza | Desviación Estándar | Coeficiente de Varianza |
| 16195.4167 | 16255 | 718682.99 | 847.7517 | 0.05235 |
|  |  |  |  |  |

1. **¿Cuál es la correlación que existe entre la edad de la madre y la cantidad de hijos?**

|  |  |
| --- | --- |
| *Estadísticas de la regresión* | |
| Coeficiente de correlación múltiple | 0.410009309 |
| Coeficiente de determinación R^2 | 0.168107634 |
| R^2 ajustado | 0.168071505 |
| Error típico | 1.422358027 |
| Observaciones | 23028 |

Como se puede observar en la siguiente tabla el análisis obtenido del coeficiente de correlación es muy bajo siendo este de 0.4100 por lo cual no podemos hacer una regresión lineal adecuada.

En esta gráfica podemos observar el comportamiento de los datos con respecto a la edad de la madre y la cantidad de hijos que tiene.

Puedo concluir que realmente la edad no es un factor determinante para la cantidad de hijos que tenga la madre. La cantidad hijos se mantiene bastante similar a lo largo de la vida de la madre.

## **Cantidad de calorías en comidas rápida**

En este estudio, se presenta una descripción detallada de las variables que conforman una base de datos relacionada con restaurantes y sus platillos. La base de datos contiene información relevante sobre una variedad de platillos ofrecidos por diferentes restaurantes. Entre las variables disponibles se encuentran: "restaurant," que representa el nombre del restaurante; "Item," que identifica el nombre del platillo; "calories," que indica la cantidad total de calorías presentes en el platillo; "cal\_fat," que señala las calorías provenientes de las grasas; "cholesterol," que cuantifica la cantidad de colesterol presente en el platillo en miligramos; "total\_carb," que refleja la cantidad de carbohidratos presentes en el platillo en gramos; "fiber," que mide la cantidad de fibra presente en el platillo en gramos; y "protein," que indica la cantidad de proteína presente en el platillo en gramos. Estos datos brindan una visión detallada de los aspectos nutricionales de los platillos ofrecidos, lo que puede ser de gran utilidad para los consumidores interesados en tomar decisiones informadas sobre sus elecciones alimenticias. El análisis de estas variables permitirá explorar patrones, comparar opciones y entender mejor el contenido nutricional de los platillos en diferentes restaurantes, contribuyendo así a una alimentación más consciente y saludable.

Fuente: <https://www.kaggle.com/datasets/ulrikthygepedersen/fastfood-nutrition>

Descripción de Variable:

|  |  |
| --- | --- |
| **Variable** | **Tipo de Dato** |
| restaurant | Categórico |
| Item | Categórico |
| calories | Numérico |
| cal\_fat | Numérico |
| cholesterol | Numérico |
| total\_carb | Numérico |
| fiber | Numérico |
| protein | Numérico |

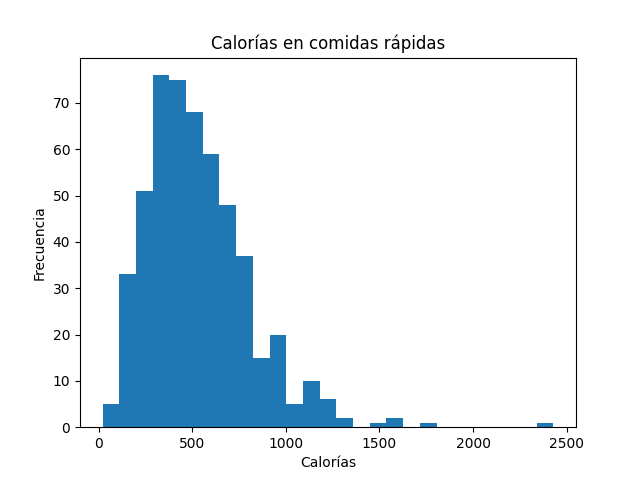
Descripción de la aproximación:

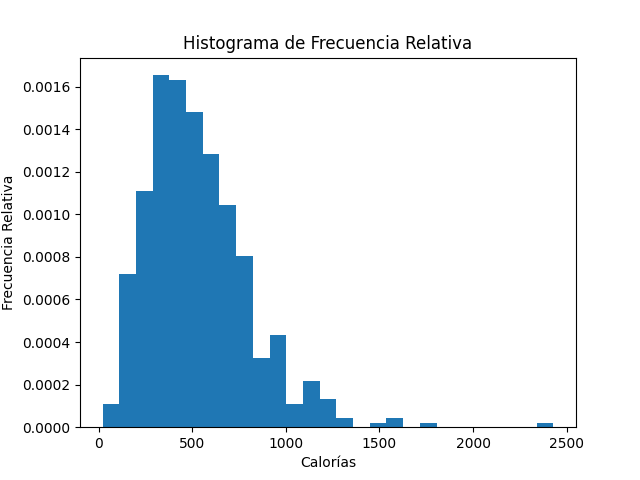
La aproximación del análisis estadístico de esta base de datos se fundamenta en un enfoque guiado por la curiosidad de conocer alternativas de comida rápida que no sean dañinas para nuestro cuerpo.

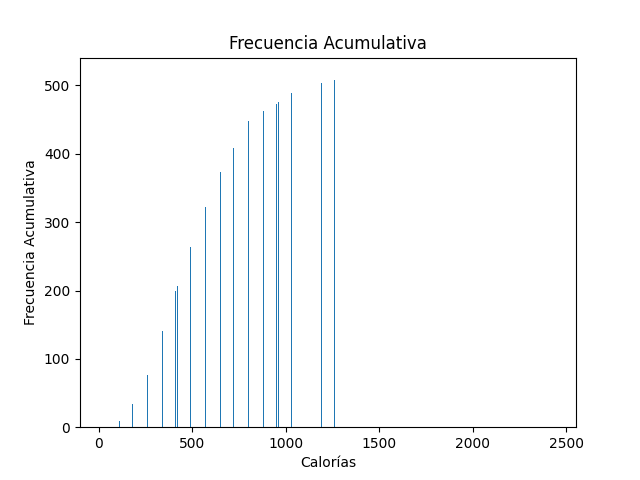
Herramienta utilizada:

-Python (Librerías: NumPy y Pandas)

Gráficas de Frecuencia:







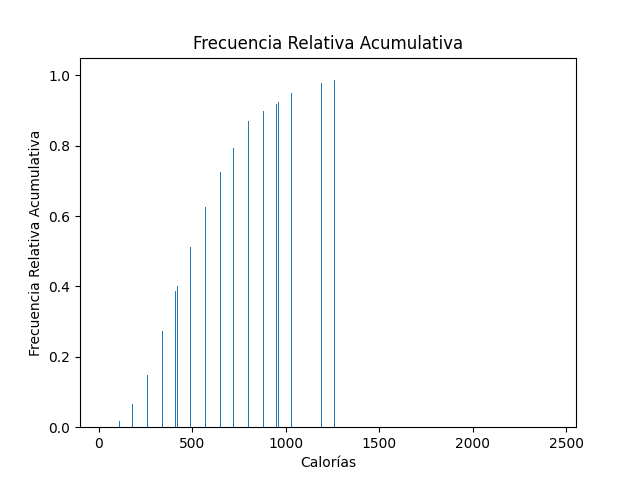
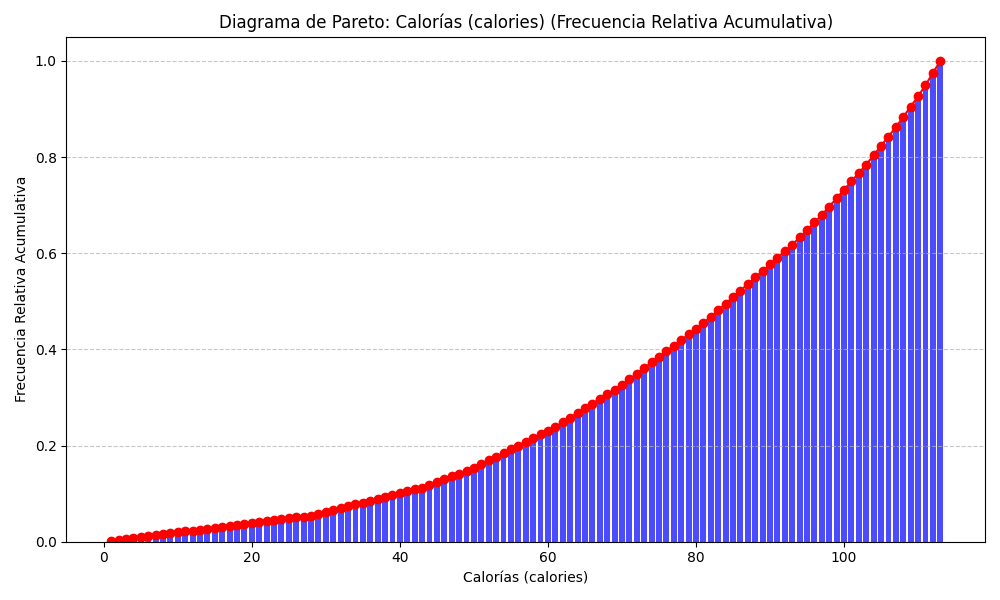


Diagrama de Pareto



Media: 530.9126

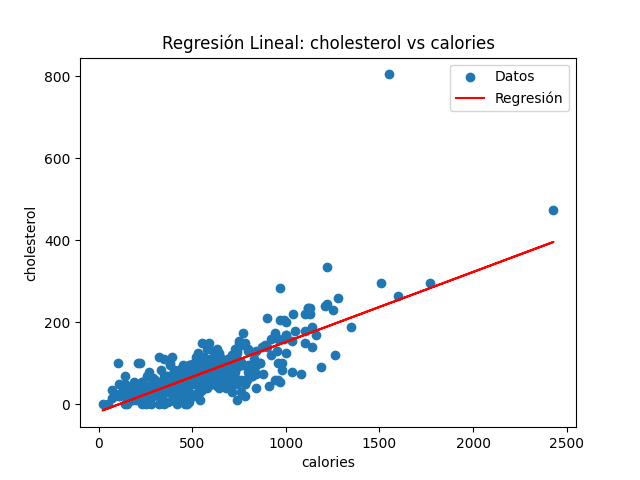
Varianza: 79770.1771

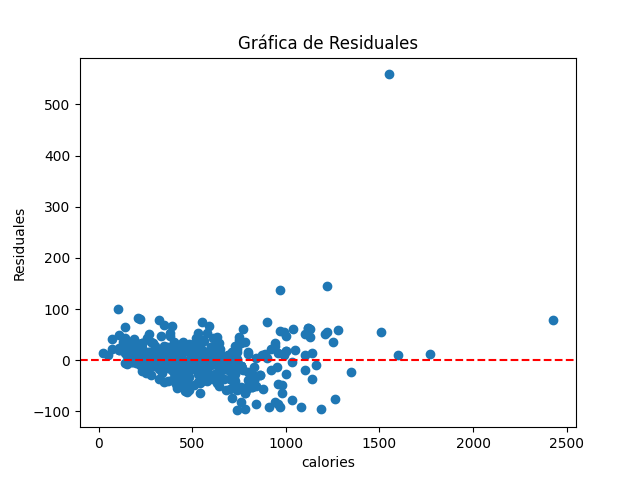
Desviación Estándar: 282.4361

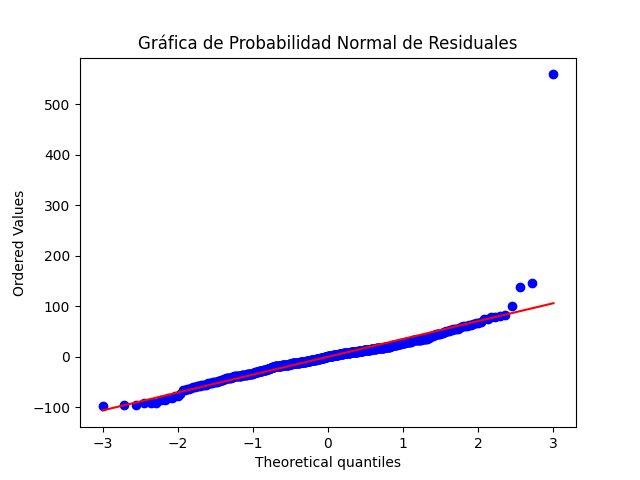
Coeficiente de correlación entre calories y cholesterol: 0.76246

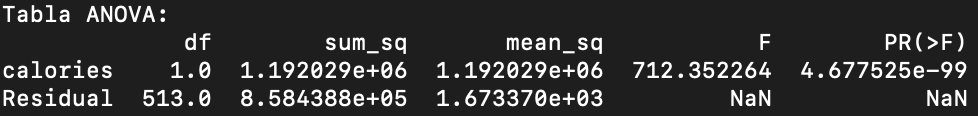
Análisis de Regresión Lineal

En la siguiente gráfica se puede observar el nivel de correlación que tienen los datos para determinar cuales variables hacen más dañino los platillos

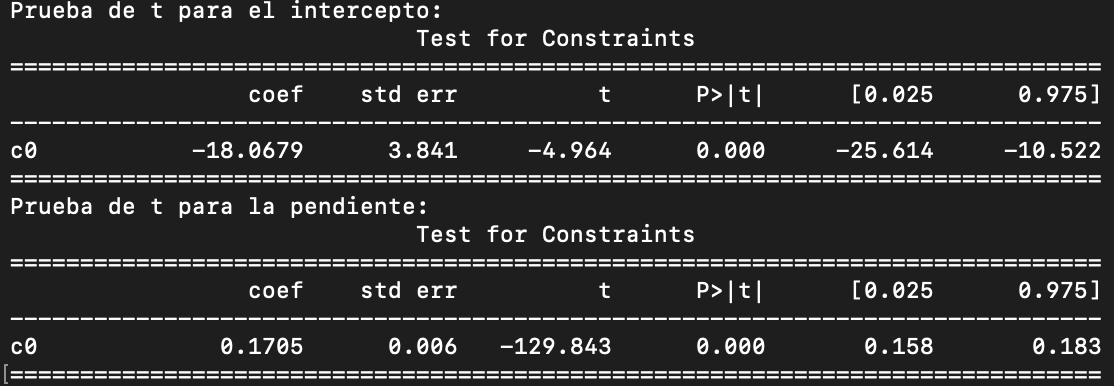


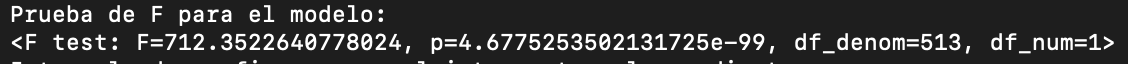






R cuadrado: 0.58134487931429





Los resultados del análisis de Regresión Lineal muestran que la variable independiente "calories" tiene un efecto significativo en la variable dependiente "cholesterol" en los platillos de los restaurantes de comida rápida en Estados Unidos. Según el coeficiente de determinación (R cuadrado), aproximadamente el 58.13% de la variabilidad en el contenido de colesterol puede explicarse por las variaciones en las calorías del platillo.

Además, al aplicar intervalos de confianza al 95%, se ha determinado que el valor del intercepto se encuentra con una probabilidad del 95% entre -25.613794 y -10.521950. Esto indica que existe un alto grado de confianza en que el verdadero valor del intercepto se encuentra dentro de este rango.

## **Autos**

El presente análisis estadístico se centra en una base de datos que contiene información detallada sobre diferentes modelos de automóviles. Esta colección de datos incluye variables clave que describen características importantes de cada vehículo. Entre las variables disponibles se encuentran: "name," que representa la marca y modelo del automóvil; "mpg," que indica la cantidad de millas recorridas por galón, brindando una perspectiva sobre la eficiencia del combustible; "cylinders," que muestra el número de cilindros presentes en el motor; "displacement," que proporciona el cilindraje total del motor en pulgadas cúbicas; "horsepower," que mide la potencia del automóvil en caballos de fuerza; "weight," que refleja el peso del vehículo en libras; "acceleration," que representa el tiempo que el automóvil tarda en alcanzar 100 km/h, ofreciendo una visión sobre su rendimiento en aceleración; "model\_year," que indica el año del modelo del automóvil; y "origin," que señala el país de procedencia del vehículo. Este análisis busca explorar patrones, tendencias y relaciones entre estas variables para obtener conocimientos significativos sobre la composición y el rendimiento de los automóviles representados en la base de datos. A través del empleo de técnicas estadísticas y visualizaciones, se pretende obtener una comprensión más profunda del conjunto de datos y, en última instancia, obtener información relevante para la toma de decisiones y futuras investigaciones en el ámbito automotriz.

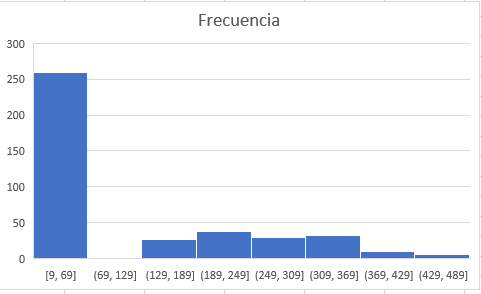
Fuente: <https://www.kaggle.com/datasets/tawfikelmetwally/automobile-dataset>

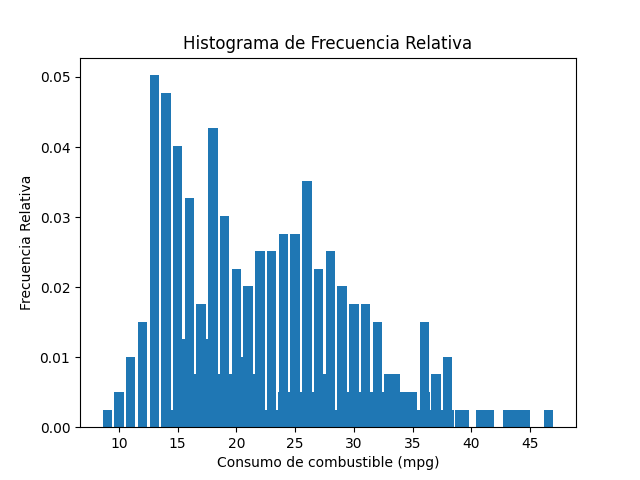
Para el análisis tuvimos las siguientes variables:

|  |  |
| --- | --- |
| **Variable** | **Tipo de Dato** |
| Nombre | Categórico |
| mpg | Numérico |
| Cilindros | Numérico |
| Desplazamiento | Numérico |
| Caballos de Fuerza | Numérico |
| Peso | Numérico |
| Aceleración | Numérico |
| Año del modelo | Numérico |
| Origen | Categórico |

Sacamos el promedio de los caballos de fuerza por su país de origen

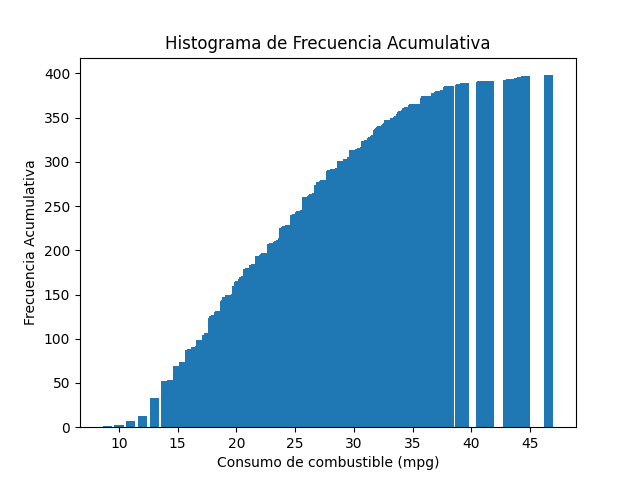
|  |  |  |
| --- | --- | --- |
| Origen | Promedio de caballos de Fuerza | |
| Europa | 80,55882353 |
| Japón | 79,83544304 |
| USA | 119,0489796 |
| Desviación Estándar | 38,4911599 | |
| Varianza | 1481,56939 | |

Histograma de frecuencia:



También me interesó el MPG que es las millas por galón

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  | | --- | |  | |  | |  | |  | |  | | | |  | | | |  |
|  |  | |  | |  | |
|  |  | |  | |  | |  | |
|  |  | |  | |  | |  | |
|  |  | |  | |  | |  | |
|  | |  | |  | |  | |  | |  | |  | |  | |
|  | |  | |  | |  | |  | |  | |  | |  | |
|  | |  | |  | |  | |  | |  | | |  | |  |  |
|  | |  | |  | |  | |  | |  | | |  | |  |  |



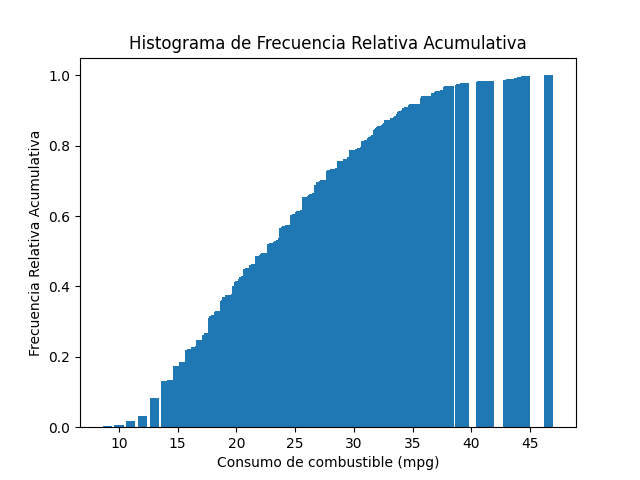
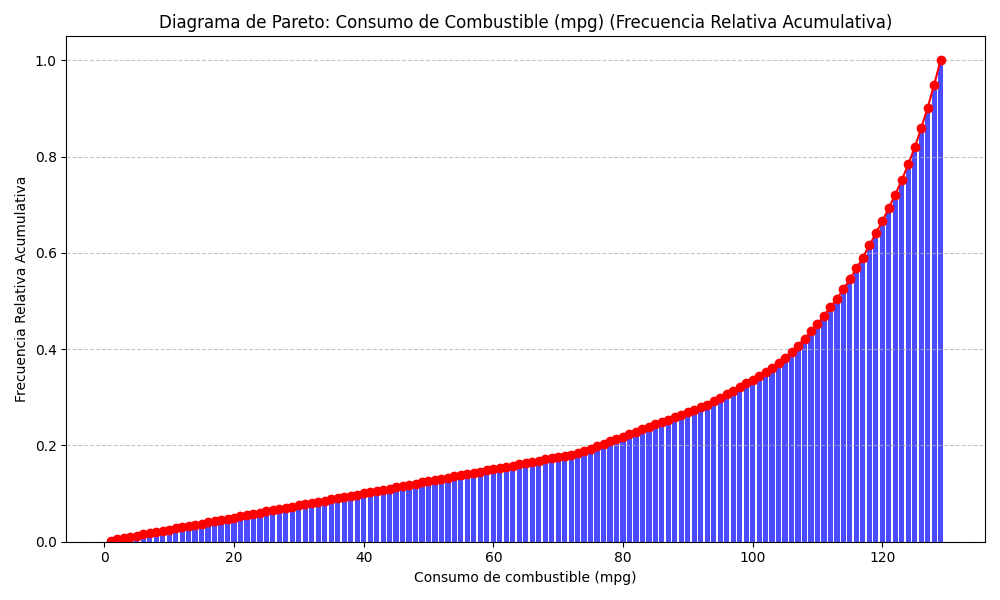


Diagrama de Pareto



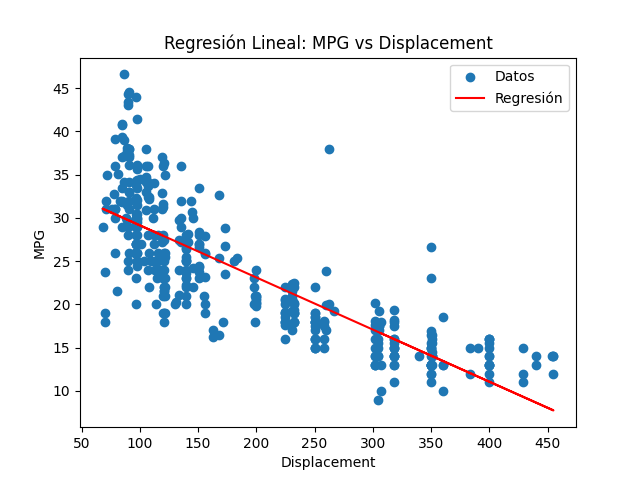
Media: 23.5145

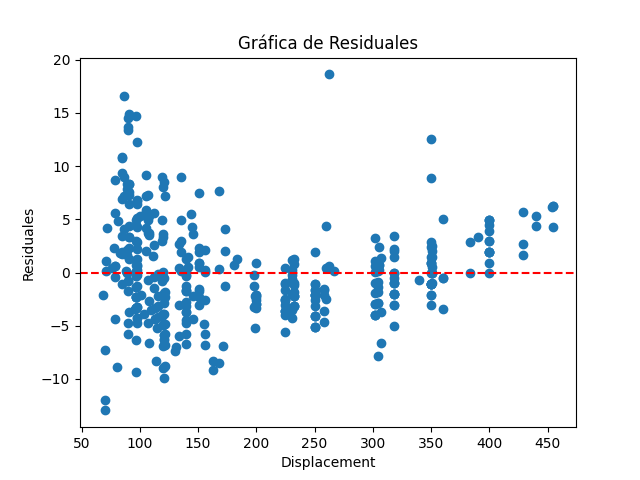
Varianza: 61.08961

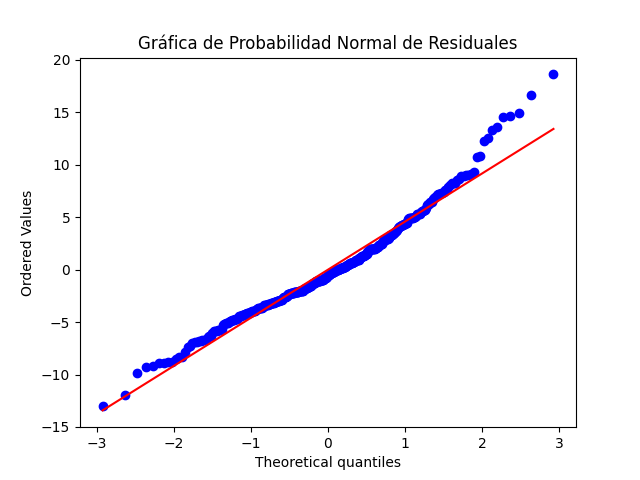
Desviación Estándar: 7.8156

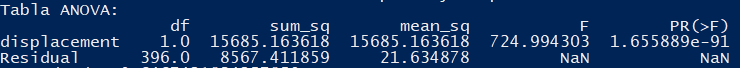
Coeficiente de correlación entre mpg y displacement: -0.80420

Análisis de Regresión Lineal

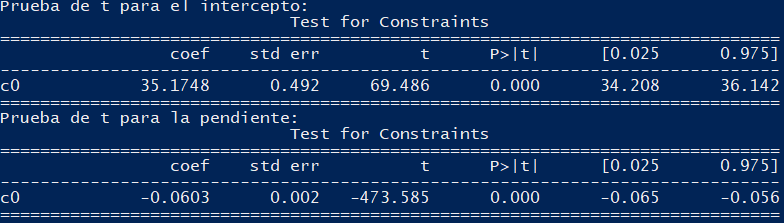




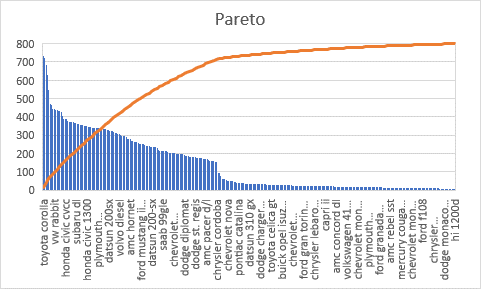
  
Tabla anova



R cuadrado: 0.6467421834257859





El término MPG representa la distancia en millas que un automóvil puede recorrer utilizando un galón de combustible. El objetivo es disminuir este valor para lograr un mayor ahorro económico en el consumo de combustible. Observamos que el coeficiente de correlación se encuentra dentro de un rango aceptable, lo que sugiere que existe una relación entre las millas por galón (mpg) y el cilindraje del automóvil.

## **Exportaciones principales mercaderías**

La aproximación del análisis estadístico de esta base de datos se fundamenta en un enfoque guiado por la curiosidad acerca del comercio internacional de Panamá, específicamente dirigido a estudiar aspectos relevantes sobre las principales mercaderías y su impacto en el panorama exportador del país.

Fuente:<https://www.inec.gob.pa/DASHBOARDS/Comercio/ExportacionesPrincipalesMercaderias>

Composición de la base de datos a analizar:

|  |  |
| --- | --- |
| **Variable** | **Tipo de Dato** |
| Grupo | Categórico |
| Suprimida | Categórico |
| Año | Categórico |
| Mes | Categórico |
| Arancel | Categórico |
| Código país | Categórico |
| Continente | Categórico |
| Peso Bruto | Numérico |
| Peso Neto | Numérico |
| Valor FOB | Numérico |

1. ¿Qué continente exporta la mayor cantidad de peso neto?

En el gráfico adjunto se aprecia una notable predominancia del peso neto en las exportaciones de la mercadería analizada en los últimos años en el continente africano, seguido de América en segundo lugar y Oceanía en tercer lugar.

|  |  |
| --- | --- |
| **Etiquetas de fila** | **Suma de Peso neto** |
| Africa | 102409444 |
| América | 32253078 |
| Asia | 4115655 |
| Oceania | 22028278 |
| **Total general** | **160806455** |

Peso Neto Total por continente

120000000

100000000

80000000

60000000

40000000

20000000

0

Africa

América

Asia

Oceania

A continuación, podemos observar que las exportaciones de la mercadería 1511.10 se dan con mayor frecuencia en América y representan más de la mitad del total de exportaciones de dicha mercadería durante los últimos años. Debido a este hecho, el análisis realizado a continuación se centrará exclusivamente en el continente americano, sin considerar a los demás continentes.

|  |  |
| --- | --- |
| **Etiquetas de fila** | **Recuento de Continente** |
| Africa | 75 |
| América | 140 |
| Asia | 4 |
| Oceania | 8 |
| **Total general** | **227** |

Frecuencia de Exportaciones por Continente

160

140

120

100

80

Total

60

40

20

0

Africa

América

Asia

Oceania

Frecuencia de Exportaciones por Continente

160

140

120

100

80

Total

60

40

20

0

Africa

América

Asia

Oceania

Total de Exportaciones

2% 4%

33%

62%

Africa América Asia Oceania

1. ¿Cómo se ha comportado la exportación de esta mercadería? ¿Cómo se comporta el Peso Neto y el valor FOB de este arancel?

Total Peso Neto Exportado

14000000

12000000

10000000

8000000

6000000

4000000

2000000

0

2017

2018

2019

2020

2021

Total Valor FOB Exportaciones

9000000

8000000

7000000

6000000

5000000

4000000

3000000

2000000

1000000

0

2017

2018

2019

2020

2021

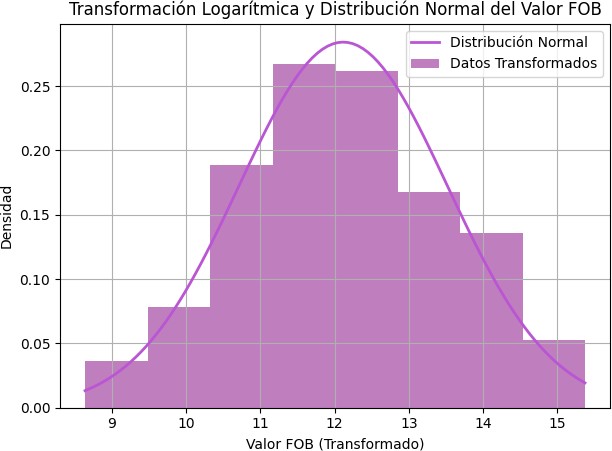
Al examinar los gráficos de las exportaciones de la mercadería 1511.10, se puede observar que tanto el Peso Neto exportado como el Valor FOB alcanzan sus niveles más altos en el año 2018. Sin embargo, a partir del año 2019, se observa una disminución en ambos valores. Esta tendencia descendente se acentúa en el año 2020, siendo este último año el punto más bajo tanto para el Peso Neto como para el Valor FOB. Estos datos sugieren una disminución en las exportaciones de la mercadería 1511.10 a partir del año 2019, con una notable caída en el año 2020.

1. ¿Promedio de valor FOB pagado en el periodo estudiado?

A continuación, se presenta un análisis de los datos numéricos del valor FOB en la siguiente tabla.

|  |  |  |  |
| --- | --- | --- | --- |
| Media | Mediana | Varianza | Desviación Estándar |
| 454436.537 | 174125 | 4.87872E+11 | 698478.13 |

Distribución normal de la columna Valor FOB.

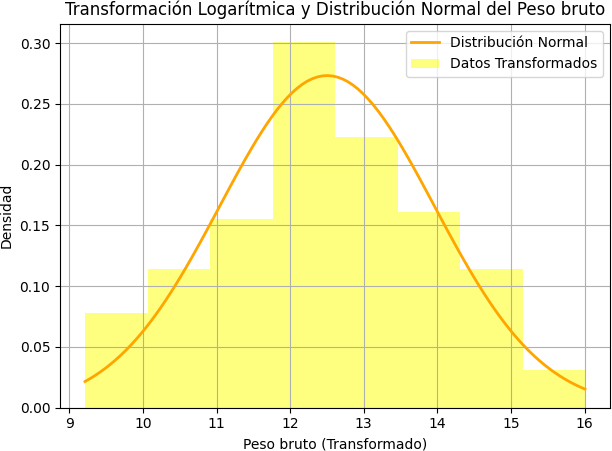


1. ¿Promedio de Peso Bruto importado en el periodo estudiado?

A continuación, se presenta un análisis de los datos numéricos del peso bruto en la siguiente tabla.

|  |  |  |  |
| --- | --- | --- | --- |
| Media | Mediana | Varianza | Desviación Estándar |
| 713163.242 | 250750 | 1.34314E+12 | 1158937.60 |

Distribución normal de la columna Peso Bruto.

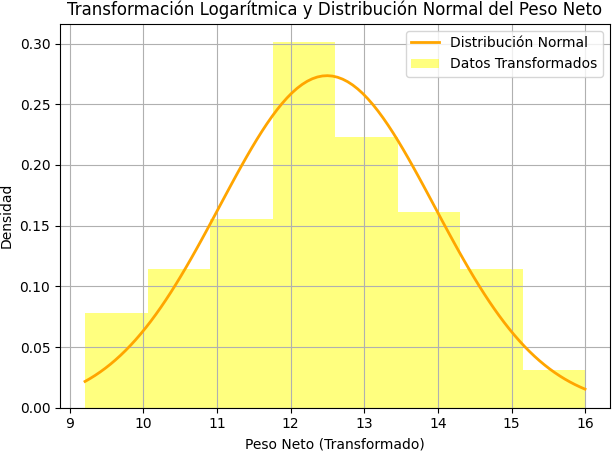


1. ¿Promedio de Peso Neto importado en el periodo estudiado?

A continuación, se presenta un análisis de los datos numéricos del peso neto en la siguiente tabla.

|  |  |  |  |
| --- | --- | --- | --- |
| Media | Mediana | Varianza | Desviación Estándar |
| 708398.48 | 250000 | 1.32922E+12 | 1152916.85 |

Distribución normal de la columna Peso Neto.



1. ¿Cuál es la relación entre el Peso Neto y valor FOB?

En el gráfico presentado a continuación se muestra la representación gráfica de la regresión lineal que relaciona el Peso Neto y el Valor FOB.



Relación Peso Neto - Valor FOB

6000000

5000000

4000000

3000000

2000000

1000000

0

0

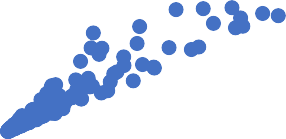
2000000

4000000

6000000

8000000

10000000



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | y | = 0.5848x + 401 | 49 |
|  |  |  | R² = 0.9318 |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

|  |  |
| --- | --- |
| *Estadísticas de la regresión* | |
| Coeficiente de correlación  múltiple | 0.96531609 |
| Coeficiente de determinación  R^2 | 0.93183515 |
| R^2 ajustado | 0.93153219 |
| Error típico | 182766.256 |
| Observaciones | 227 |

En este caso, se obtuvo un coeficiente de correlación de 0.965 lo que sugiere una correlación fuerte y positiva entre el Peso Neto y el Valor FOB. Además, el valor de R^2 de 0.932 es muy cercano a 1, lo que indica una correlación significativa entre el Peso Neto y el Valor FOB. Esto implica que alrededor del 93.2% de las fluctuaciones en el Valor FOB pueden ser atribuidas a las variaciones en el Peso Netos lo que significa que el Peso Neto es un muy buen indicador o predictor del Valor FOB dentro del marco de la regresión lineal utilizada.

A continuación, los intervalos de confianza para la intersección y la variable x1.

|  |  |
| --- | --- |
|  | *Coeficientes* |
| Intercepción | 40148.887 |
| Variable X 1 | 0.5848229 |

|  |  |  |
| --- | --- | --- |
| Intervalo de confianza B0 | | |
| 12226.4107 | <B0 | 68071.3632 | |

|  |  |  |
| --- | --- | --- |
| Intervalo de confianza B1 | | |
| 0.56415483 | <B1 | 0.60549096 |

ANOVA

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Grados de libertad* | | *Suma de cuadrados* | *Promedio de los cuadrados* | *F* | *Valor crítico de F* |
| Regresión | 1 | 1.0274E+14 | 1.0274E+14 | 3075.82143 | 3.282E-133 |
| Residuos | 225 | 7.5158E+12 | 3.3404E+10 |  |  |
| Total | 226 | 1.1026E+14 |  |  |  |

El análisis de varianza realizado en el modelo de regresión reveló una fuerte correlación positiva entre el Peso Neto y el Valor FOB, indicando que a medida que el Peso Neto aumenta, también tiende a aumentar el Valor FOB de las importaciones.



## **Precios de Laptops**

Fuente: <https://www.kaggle.com/datasets/muhammetvarl/laptop-price>

El presente análisis se enfoca en una base de datos que proporciona información detallada sobre diversas laptops disponibles en el mercado. Esta colección de datos incluye una serie de variables clave que describen las características y especificaciones técnicas de cada laptop. Entre las variables disponibles se encuentran: "Laptop," que representa el nombre completo del producto; "Status," que indica si la laptop es nueva o reconstruida; "Brand," que identifica la marca de la laptop; "Model," que especifica el modelo del dispositivo; "CPU," que muestra el nombre del procesador utilizado en la laptop; "RAM," que indica la cantidad de memoria RAM instalada, expresada en valor numérico; "Storage," que detalla la capacidad de almacenamiento de la laptop también en valor numérico; "GPU," que proporciona el modelo de la tarjeta gráfica integrada, y puede tener valor numérico nulo si no está presente; y "FinalPrice," que refleja el precio final de venta de la laptop.

Mediante el análisis de estas variables, se busca obtener una comprensión más profunda de la oferta de laptops en términos de marcas, modelos, especificaciones técnicas y precios. Esto permitirá a los consumidores y profesionales del sector tecnológico tomar decisiones informadas al seleccionar el equipo más adecuado para sus necesidades y presupuestos. A través de técnicas estadísticas y visualizaciones, se explorarán patrones, tendencias y posibles relaciones entre las variables, con el objetivo de obtener conocimientos significativos sobre la diversidad y características de las laptops disponibles en el mercado actual. Este análisis proporcionará una visión detallada de la industria de laptops y servirá como base para futuras investigaciones y toma de decisiones en el ámbito tecnológico.

Composición de la base de datos a analizar:

|  |  |
| --- | --- |
| **Variable** | **Tipo de Dato** |
| Laptop | Categórico |
| Status | Categórico |
| Model | Categórico |
| CPU | Categórico numérico |
| RAM | Numérico |
| Storage | Numérico |
| GPU | Numérico |
| FinalPrice | Numérico |

Descripción de la aproximación:

La aproximación del análisis estadístico de esta base de datos se fundamenta en un enfoque guiado por la curiosidad de conocer el comportamiento de los diferentes componentes que poseen las laptops lo cual influye en su precio final.

Análisis de la Base de Datos

Gráficas de Frecuencia:

Gráfico, Histograma

Descripción generada automáticamente

Gráfico, Histograma

Descripción generada automáticamente

Gráfico, Histograma

Descripción generada automáticamente

Gráfico, Histograma

Descripción generada automáticamente

Gráfico, Gráfico de líneas

Descripción generada automáticamente

Media: 15.41388888888889

Varianza: 97.37377643970974

Desviación Estándar: 9.86781518066232

Coeficiente de correlación entre RAM y Final Price: 0.7249463124103899

Análisis de Regresión Lineal

Gráfico

Descripción generada automáticamente

Gráfico

Descripción generada automáticamente

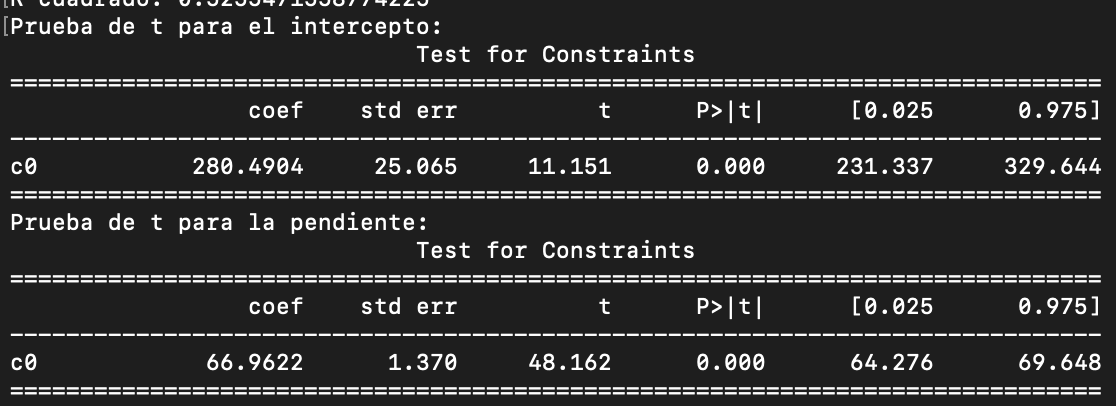
Gráfico, Gráfico de líneas

Descripción generada automáticamente

Texto

Descripción generada automáticamente

R cuadrado: 0.5255471558774225



Texto

Descripción generada automáticamente

Los resultados del análisis muestran que la variable "RAM" tiene un coeficiente de correlación positivo importante con la variable dependiente "FinalPrice". El coeficiente de correlación entre "RAM" y "FinalPrice" es de aproximadamente 0. 525, lo que indica una relación positiva entre la cantidad de RAM y el precio final.

Los intervalos de confianza al 95% para el intercepto están entre 231.337 y 329.644, y para la pendiente están entre 64.276 y 69.648. Esto significa que estamos 95% seguros de que el verdadero valor del intercepto y la pendiente se encuentran dentro de estos intervalos.

## Conclusión

En conclusión, el análisis estadístico realizado en este proyecto ha proporcionado una comprensión profunda de las bases de datos seleccionadas. Se utilizó una variedad de técnicas visuales, como gráficas de barra, pastel, histogramas y paretogramas, para explorar y segmentar los datos de manera efectiva. Se describieron detalladamente las variables y su naturaleza numérica o no numérica, lo que permitió entender la estructura y el formato de los datos. El análisis de conversión de distribución discreta a distribución continua se aplicó con criterio, mejorando la interpretación de los resultados. El estudio de correlación y regresión lineal reveló relaciones significativas entre variables numéricas y proporcionó herramientas valiosas para realizar predicciones y evaluaciones. En general, este proyecto ha demostrado la importancia de la estadística en el análisis de datos y cómo las herramientas de Excel pueden utilizarse para obtener resultados precisos y significativos. Los hallazgos obtenidos pueden ser fundamentales para la toma de decisiones y la resolución de problemas en diversos campos, destacando el valor de aplicar técnicas estadísticas rigurosas en el análisis de datos. También que fue la primera vez utilizando los Python para analizar estas pases de datos.