

SPRINT 8.01

NIVEL 1. EJERCICIO 1

Se ha realizado un gráfico de histograma que muestra la frecuencia de gasto en transacciones por rango de gasto.

Se han establecido unos rangos de gasto de 20 €.

El gráfico muestra que el rango de gasto más frecuente es el de entre 20 y 40 €. Los siguientes rangos más frecuentes están distribuidos por todo el rango de gasto.

El rango de gasto menos frecuente es el menor de 20€ lo que indica que los productos más baratos no se encuentran entre los más solicitados.

El segundo rango menos frecuente es el de entre 200 y 220€. Es necesario hacer una comprobación más exhaustiva de los productos vendidos en las transacciones de este rango para ver si hay alguna relación entre determinados productos y ese rango de ventas o es una anomalía estadística teniendo en cuenta que el rango siguiente de venta de 220-240€ se encuentra entre los más frecuentes.

NIVEL 1. EJERCICIO 2

Se ha realizado un gráfico scatterplot para mostrar la relación entre el peso y el precio de los productos.

No existe relación entre el peso del producto y su precio salvo en los productos de más peso entre los cuales no hay productos con un precio inferior a 25€ en los productos de 3kg o inferiores a 75€ en los superiores a 3 kg.

Es de subrayar la falta de productos en la franja de 1.5 a 2 kg.

NIVEL 1. EJERCICIO 3

Se ha realizado un gráfico de barras horizontales que muestra el número de usuarios por país.

Lo primero que se puede observar es que todos los usuarios corresponden solo a 3 países todos ellos de habla inglesa. Para aumentar el número de usuarios sería necesario aumentar el número de países en los que se venden nuestros productos. Por ejemplo. Haciendo campañas de publicidad en países de habla hispana.

La diferencia de usuarios entre países puede ser debida a la diferencia de población y por tanto de posibles clientes.

NIVEL 1. EJERCICIO 4

Se ha realizado un gráfico boxplot que muestra la distribución del gasto por transacción y país.

Se puede observar que la media de gasto en USA es superior a la de los otros países y que tiene una distribución mas sesgada ya que la mediana es superior a la media.

NIVEL 1. EJERCICIO 5

Se ha realizado un gráfico de barras que muestra el número de transacciones aceptadas y rechazadas por país.

Se puede observar que Canada es el país con más transacciones y también el único con transacciones rechazadas.

NIVEL 1. EJERCICIO 6

Se ha realizado un gráfico swarmplot para comprobar la distribución de las transacciones rechazadas por país.

Como muestra el gráfico anterior en este también se observa que todas las transacciones declinadas se han producido en Canada pero no se observa ningún patrón distintivo en su distribución por lo que parece ser un problema aleatorio posiblemente debido a algún problema con el sistema de pago en ese país. No obstante, es conveniente hacer un seguimiento de este problema por si fuese algún intento de uso fraudulento del sistema de pago.

NIVEL 1. EJERCICIO 7

Se realiza un pairplot para comprobar las relaciones entre distintos pares de variables numéricas por país.

No se observa ninguna relación entre las distintas variables.

El pairplot también muestra un gráfico de densidad(kde) de las distintas variables. El gráfico de densidad no muestra datos exactos sino que muestra tendencias con curvas suavizadas.

NIVEL 2. EJERCICIO 1

Se crea una tabla de correlación entre las variables numéricas de amount, price y weight.

La escala de valores es de 0 a 1.

La única relación significativa que se puede observar es entre las variables de price y weight. Esta relación es de un 19%.

Se realiza un gráfico heatmap para visualizar la tabla de correlación.

NIVEL 2. EJERCICIO 2

Se ha creado un gráfico jointplot que muestra la relación entre las variables numéricas de price y weight.

Este gráfico también muestra un histograma con curva kde de la distribución de cada variable.

El histograma de weight muestra un menor número de productos en la franja de 1.5 a 2.5 kg.

El histograma de price muestra un rango de precio mas uniforme.

La linea de regresión muestra visualmente la correlación existente entre las dos variables. En este caso es del 19%.

la banda de color mas claro alrededor de la linea muestra la incertidumbre de la variación.

NIVEL 3. EJERCICIO 1

Se ha creado un gráfico violinplot conjuntamente con un boxplot para visualizar la distribución del gasto por transacción y por país.

Aunque el violinplot incluye un boxplot en si mismo, se ha decidido mostrar un gráfico boxplot externo para una mejor visualización de los datos.

En el gráfico boxplot se puede observar que la media de gasto en USA es superior a la de los otros países y que tiene una distribución mas sesgada ya que la mediana es superior a la media.

En el gráfico violinplot se puede observar que la densidad de amount en USA está más localizada en los rangos superiores de amount, mientras que en UK lo está mas en la zona de la media. En Canada densidad está más distribuida a lo largo de todo el rango de amount.

NIVEL 3. EJERCICIO 2

Se ha realizado un gráfico FacetGrid para visualizar la relación entre amount y price por pais.

La menor densidad de puntos en el gráfico de UK muestra que en este pais se realizan un menor número de transacciones y la linea de regresión muestra una mayor relación entre las variables que en los demás paises, aunque el margen de variación también es mayor.

La linea de regresión correspondiente a USA muestra una tendencia casi plana mostrando una relación prácticamente nula entre las variables.

La densidad del scatter en el gráfico de Canada muestra que es el país en el que se realizan mas transacciones y la linea de regresión muestra una relación ligeramente inversa entre las variables