**Pregunta 5**

* **Si se creara 4 categorías para la variable tipo de producto, ¿cuál categoría   es la más frecuente?**

Crear las 4 categorías y comenzar a distribuir los datos dentro de las categorías creadas, para posteriormente representar estos datos en un **diagrama de pastel o torta** para que se facilite la visualización de cuál es la más frecuente.

diagrama de pastel

Lo primero que vamos a hacer es crear 4 categorías que nos permitan asignar cada tipo de producto dentro de estás nuevas categorías ordenadas según su naturaleza, las categorías elegidas son:

1. Alimentos y Bebidas
2. Cuidado Personal y Cosméticos
3. Ropa y Accesorios
4. Materiales y Productos del Hogar

Luego agrupamos cada variable en su categoría:

**1. Alimentos y Bebidas**

Snacks

Cárnicos

Cereales

Frutas

Alimento infantil

Bebida

Verduras

**2. Cuidado Personal y Cosméticos**

Cosméticos

Cuidado personal

**3. Ropa y Accesorios**

Ropa

**4. Materiales y Productos del Hogar**

Material de oficina

Doméstico











Análisis: -------------------------------------------------------------

Interpretación: -------------------------------------------------------------

**Descripción:**

El gráfico de pastel (o gráfico de torta) muestra la distribución porcentual de cuatro categorías de productos de venta:

1. Alimentos y bebidas
2. Cuidado personal y cosméticos
3. Ropa y accesorios
4. Materiales y productos de hogar

Cada sector del gráfico representa el porcentaje correspondiente a cada categoría. Tomar en cuenta que dentro de cada categoría hay un tipo específico de producto cuya frecuencia se ha sumado al total.

**Interpretación:**

Se observa lo siguiente:

Alimentos y bebidas: Representa el 60% del total. Esta es la categoría más grande y, por lo tanto, la más frecuente.

Cuidado personal y cosméticos: Representa el 15% del total.

Ropa y accesorios: Representa el 8% del total.

Materiales y productos de hogar: Representa el 17% del total.

De acuerdo con el gráfico, la categoría más frecuente de productos de venta es Alimentos y bebidas, que constituye el 60% del total. Esta categoría domina claramente en comparación con las otras, que representan porcentajes significativamente menores.

1. **¿Cómo se distribuye la variable Importe coste total?**

Para responder a esta pregunta procedemos a ordenar los datos de menor a mayor, calcular los cuartiles: el primer cuartil (Q1) marca el 25% de los datos, la mediana (Q2) marca el 50%, y el tercer cuartil (Q3) marca el 75%.

Determinar el rango intercuartil, que es la diferencia entre Q1 y Q3, calcular los límites de los bigotes, Identificar los valores atípicos, que son los datos fuera de estos límites.

**Dibujar el diagrama:** la caja se extiende desde Q1 hasta Q3, con una línea en la mediana (Q2), los bigotes llegan hasta los límites calculados, y los valores atípicos se marcan como puntos separados.



**Descripción**

El siguiente diagrama de cajas y bigotes muestra la distribución de la variable "importe coste total" en un registro de ventas.

**Interpretación**

La mediana, representada por la banda dentro de la caja azul, es de aproximadamente €459,283.44. Esto indica que la mitad de los datos están por encima de este valor y la otra mitad por debajo.

La media, señalada por la marca 'x', es alrededor de €907,076.52.

El primer cuartil (Q1) se sitúa en aproximadamente €169,841.47, lo que significa que el 25% de los datos están por debajo de este valor.

El tercer cuartil (Q3) es de cerca de €1,178,942.29, indicando que el 75% de los datos están por debajo de este valor.

El mínimo se encuentra alrededor de €31.79, y el máximo es €26,68,371.68

·        El rango intercuartílico (IQR): La diferencia entre el tercer cuartil y el primer cuartil (Q3 - Q1) es 1,178,942.29 - 169,841.47 = 1,009,100.82.

·        El límite inferior (Q1 - 1.5IQR) es 169,841.47 - 1.5(1,009,100.82) = -334,708.94.

·        El límite superior (Q3 + 1.5IQR) es 1,178,942.29 + 1.5(1,009,100.82) = 2,692,593.52.

·        Se observan varios puntos individuales por encima del bigote superior, los cuales representan valores atípicos. El valor atípico más alto es de aproximadamente €5,068,488.89.



·        El bigote superior (la línea que se extiende desde la caja hacia arriba) es notablemente más largo que el bigote inferior. Esto sugiere que hay una mayor dispersión de los datos en la cola derecha de la distribución.

. La mediana (la línea dentro de la caja) está más cerca del primer cuartil (Q1) que del tercer cuartil (Q3), esto indica asimetría a la derecha. En otras palabras, la mediana se encuentra desplazada hacia la parte inferior de la caja y se presenta una asimetría a la derecha.

1. **¿Cómo se distribuye el número de unidades que se venden?**

**Diagrama de cajas:**  este tipo de gráfico es útil para visualizar la distribución de un conjunto de datos, así como para identificar valores atípicos y resumir algunas características clave de la distribución, como la mediana, los cuartiles y los valores mínimos y máximos.



Descripción

El siguiente diagrama de cajas y bigotes muestra la distribución de la variable "unidades de venta" en un registro de ventas.

Interpretación

* Del gráfico podemos observar que la mediana de unidades de venta es 5125.5, lo que significa que la mitad de las ventas están por debajo de este valor y la otra mitad por encima.
* La media o promedio de los datos es 5024.481 unidades de venta.
* La media es significativa porque es muy similar a la mediana, indicando una distribución equilibrada sin sesgo significativo.
* El primer cuartil (Q1) es 2657.5. Esto significa que el 25% de las ventas están por debajo de este valor.
* El tercer cuartil (Q3) es 7475.25. Esto significa que el 75% de las ventas están por debajo de este valor y el 25% restante está por encima.
* El rango intercuartílico (IQR): La diferencia entre el tercer cuartil y el primer cuartil (Q3 - Q1) es 7475.25 - 2657.5 = 4817.75. Este es el rango donde se encuentra el 50% central de los datos.
* El límite inferior (Q1 - 1.5IQR) es 2657.5 - 1.54817.75 = -4569.125.
* El límite superior (Q3 + 1.5IQR) es 7475.25 + 1.54817.75 = 14701.875.
* Dado que ningún dato se encuentra fuera de este rango y como se observa en la gráfica, no hay presencia de datos atípicos.
* También podemos notar que la forma de la distribución es casi simétrica, puesto que la mediana está centrada en la caja y muy cercana a la media. Además, los bigotes son aproximadamente de la misma longitud en ambos lados.

Nota sobre los límites superior e inferior

Aunque los cálculos para los límites inferior y superior (utilizando 1.5 veces el IQR) pueden dar valores negativos o muy altos, estos valores son teóricos para identificar posibles outliers. En la práctica, los datos reales no pueden ser negativos (para unidades de venta) y cualquier valor fuera de los bigotes del diagrama se consideraría un outlier. En este caso, el diagrama no muestra ningún outlier, lo que respalda la simetría y la normalidad de la distribución.