No hay relaciones lineales fuertes entre las variables, según los diagramas de dispersión.

Las variables parecen estar distribuidas de manera independiente.

Los datos muestran una dispersión uniforme a través de las variables, lo que sugiere que no hay valores atípicos significativos ni conglomerados.

Supongamos que tenemos un conjunto de datos que representa los salarios mensuales de los empleados de una empresa. Los datos son los siguientes:

[2000, 2500, 2800, 3000, 3200, 3500, 4000, 4200, 4500, 5000, 5500, 6000, 6500, 7000, 7500, 8000, 8500, 9000, 9500, 10000, 20000]

En este caso, podemos observar que el último valor, 20000, es mucho más alto que el resto de los salarios. Este valor es un ejemplo de un valor atípico, ya que se aleja significativamente del patrón general de los datos.

Este valor atípico puede influir en el análisis y la interpretación de los datos. Por ejemplo, si calculamos la media de los salarios, este valor atípico puede aumentar artificialmente el promedio, dando la impresión de que los salarios son más altos de lo que realmente son. Del mismo modo, si calculamos la desviación estándar, este valor atípico puede aumentar la dispersión de los datos.

Es importante identificar y manejar adecuadamente los valores atípicos. En algunos casos, puede ser apropiado eliminarlos del análisis si se consideran errores de medición o datos incorrectos. En otros casos, puede ser necesario aplicar técnicas de transformación de datos o utilizar métodos estadísticos robustos que sean menos sensibles a los valores atípicos.

En resumen, en el ejemplo anterior, el valor 20000 es un ejemplo de un valor atípico en el conjunto de datos de los salarios mensuales de los empleados.

Análisis grafico de cajas

La columna "Cantidad" representa la cantidad de documentos y tiene un promedio de aproximadamente 355.44 documentos. El valor mínimo es 101 y el valor máximo es 600. La dispersión de los datos se puede estimar utilizando la desviación estándar, que es de aproximadamente 116.40.

La columna "NoMes" representa el número del mes y tiene un promedio de aproximadamente 5.02. Los valores varían entre 1 y 9, lo que indica que los datos abarcan un período de 9 meses.

La columna "PesoGB" representa el peso en gigabytes y tiene un promedio de aproximadamente 20.51 GB. El valor mínimo es 10.04 GB y el valor máximo es 30.48 GB. La dispersión de los datos se puede estimar utilizando la desviación estándar, que es de aproximadamente 6.06.

Además, los cuartiles proporcionan información sobre la distribución de los datos. Por ejemplo, el 25% de los valores en la columna "Cantidad" son menores o iguales a 246, y el 75% de los valores son menores o iguales a 446. El 50% de los valores, que corresponde a la mediana, es igual a 377.