Graphical user interface, text, application

Description automatically generated

import seaborn as sns

sns.displot(df, x='price\_usd', hue='manufacturer\_name')

df['price\_usd'].std()

rango = df['price\_usd'].max() - df['price\_usd'].min()

rango

median = df['price\_usd'].median()

Q1 = df['price\_usd'].quantile(q=0.25)

Q3 = df['price\_usd'].quantile(q=0.75)

min\_val = df['price\_usd'].quantile(q=0)

max\_val = df['price\_usd'].quantile(q=1.0)

print(minval, Q1, median, Q3, max\_val)

Detectar outliers

IQR = Q3 - Q1

minlimit = Q1 - 1.5 \* IQR

maxlimit = Q3 + 1.5 \* IQR

print(minlimit, maxlimit)

ESTADISTICA DESCRIPTIVA

* BLOQUE ANALITICO: MEDIDAS DE TENDENCIA CENTRAL, DISPERSION
* VISUAL: CON COMPONENTES VISUALES,

|  |  |
| --- | --- |
| Chart, histogram  Description automatically generated | Chart  Description automatically generated |

Implica que mas de 3,000 es outliner y debe eliminarse del analisis

|  |  |
| --- | --- |
|  |  |
| Chart, box and whisker chart  Description automatically generated | Esto nos ayuda a ver los outliers por cada categoria  Vemos que los electricos son menos, y son mas caros  De modo que asi eliminamos comparar peras y manzanas  Porque si los electricos son mas caros, son un tema que debe verse separados |