In this project, I performed **descriptive statistics** and **data cleaning** to prepare a retail dataset for analysis and dashboard creation in Power BI.

**Descriptive Statistics before Cleaning**

A screenshot of a table

AI-generated content may be incorrect. A table of numbers and symbols

AI-generated content may be incorrect.

**Quantity**

* High gap between Mean and median, that mean there are outliers.
* Based on median Half the orders are **3 items or fewer**
* Based on SD there are Extremely high variation (many unusually large orders)
* Skewness near to symmetric
* Based on high kurtosis value , there are extreme outlier

**Unit Price**

* Median = 2.08 then Half of items cost 2.08
* Based on SD there are very high variation in price
* Extremely **right-skewed** : most items are cheap, but a few are **extremely expensive**.
* Based on high kurtosis value ,there are strong outlier

**Data Cleaning Steps**

**1. Filtered:** 1,454 rows with null Description, Unit Price = 0, and null Customer ID By Removed this columns

A screenshot of a computer

AI-generated content may be incorrect.

**2.Removed duplicates**: from 540,455 To 535,187 rows.

**3. Replaced null** in Customer ID with "Unknown Customer" , 24% of customers are unknown.

**4. Removed**: 2 rows with negative Unit Price (treated as data entry errors).

A green and black text

AI-generated content may be incorrect.

**5. Outliers**: Based on IQR, a large number of Unit Price outliers detected. Due to volume, they were **kept** but flagged for separate analysis.



A table with text on it

AI-generated content may be incorrect.

**New Featured Created**

A screenshot of a computer screen

AI-generated content may be incorrect.

1-If quantity is negative , then return, else sales

2-If Customer is un Known, then Guest else Registred

**Dashboard**

**A screenshot of a computer

AI-generated content may be incorrect.**