# Exploratory Data Analysis (EDA) – Task 2

- **Project Title:** Exploratory Data Analysis – Task 2
- **Your Name:** Soniya Sourath

## 1. Objective / Introduction

The objective of this analysis is to perform detailed exploratory data analysis (EDA) on the given dataset to uncover meaningful business insights, detect anomalies, analyze variable relationships, and identify patterns.

## 2. Dataset Overview

- **Number of rows:** 2000
- **Number of columns:** 11
- **Column types:** float64(5), int64(2), object(4)
- **Missing values summary:** The dataset contains some missing values in a few columns. Most columns have complete data except for age, quantity, price_per_unit, cogs, and total_sale.

## 3. EDA Analysis & Visualizations

## 3.1 Univariate Analysis

**Numerical Columns:** - Histogram: - Boxplot: - Skewness:
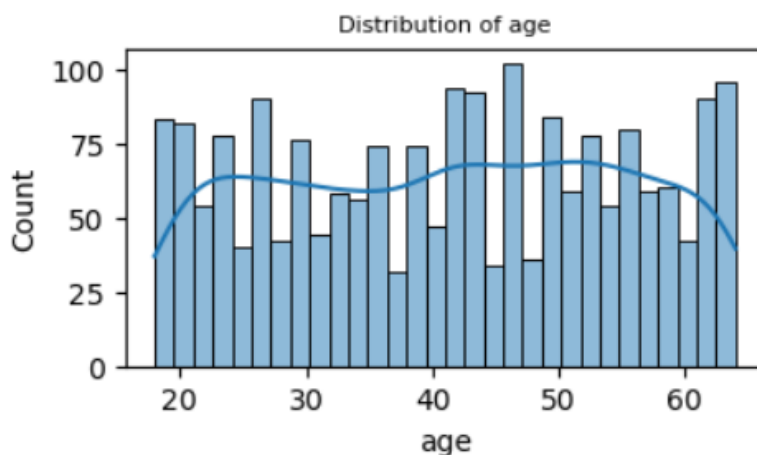
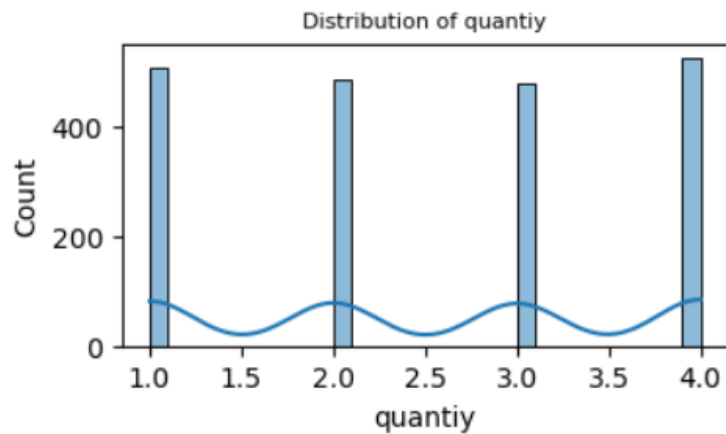**Figure 1.1: Histogram of Distribution of age** >



Distribution of age

**Figure 1.2: Histogram of Distribution of Quantity** >


Distribution of quantiy

**Figure 1.3: Histogram of Distribution of price_per_unit** >


Distribution of price_per_unit
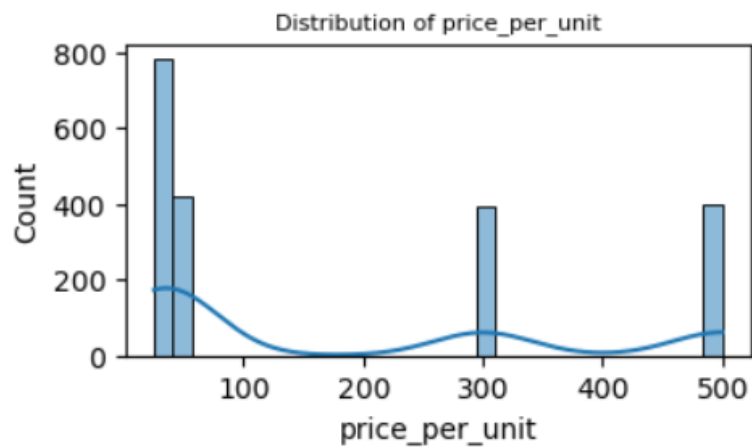
**Figure 1.4: Histogram of Distribution of Cogs** >


Distribution of cogs
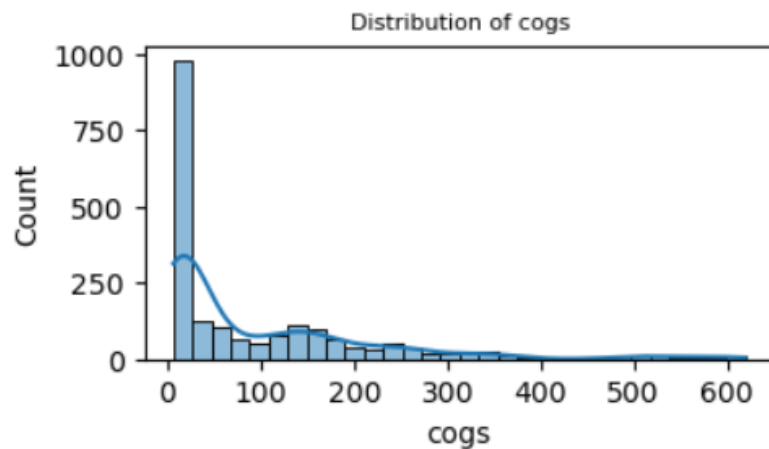
**Figure 1.5: Histogram of Distribution of Total_sale >**


Distribution of total_sale
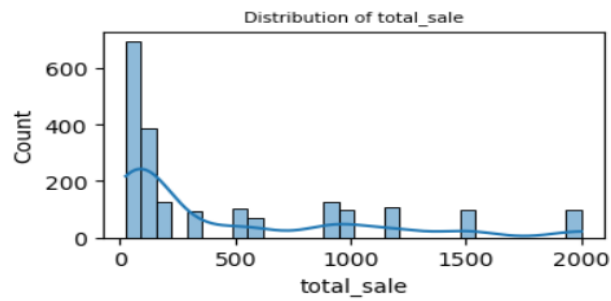
**Figure 2.1: Boxplot of Age>**
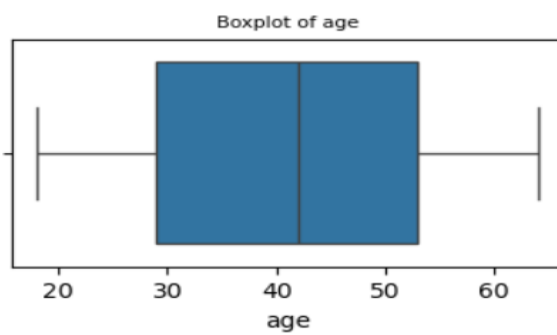

Boxplot of age

**Figure 2.2: Boxplot of Quantity>**


Boxplot of quantiy
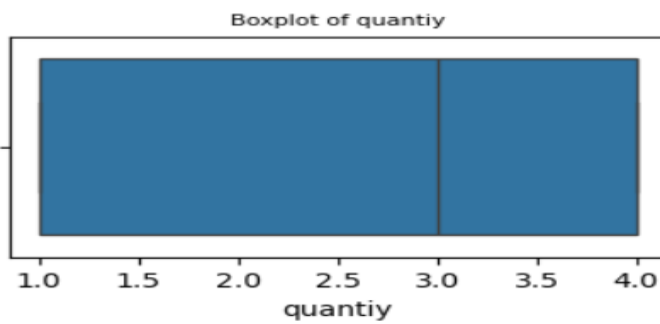
**Figure 2.3: Boxplot of price_per_unit>**


Boxplot of price_per_unit

**Figure 2.4: Boxplot of Cogs>**
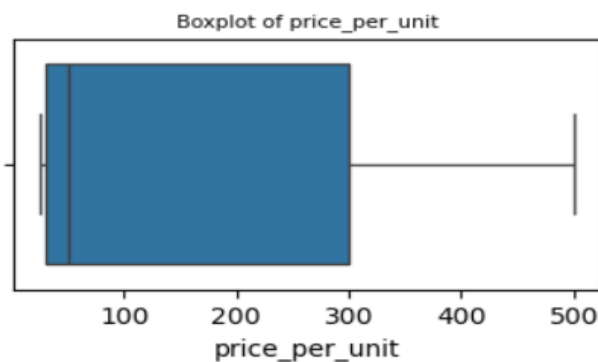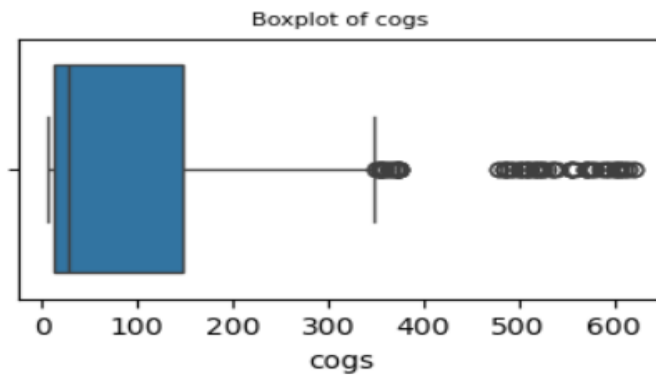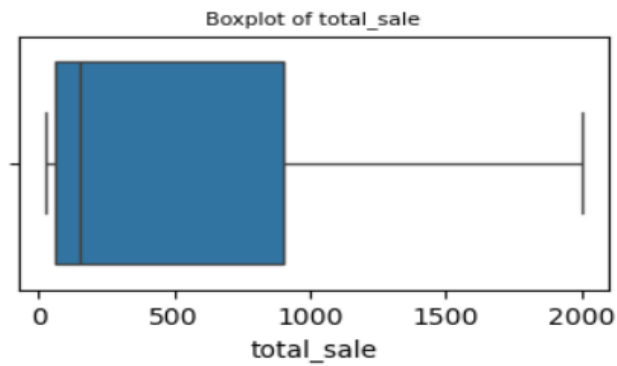

Boxplot of cogs

**Figure 2.5: Boxplot of total_sale>**


Boxplot of total_sale

**Categorical Columns:** - Countplot:

**Figure 3.1: Countplot of Gender>**
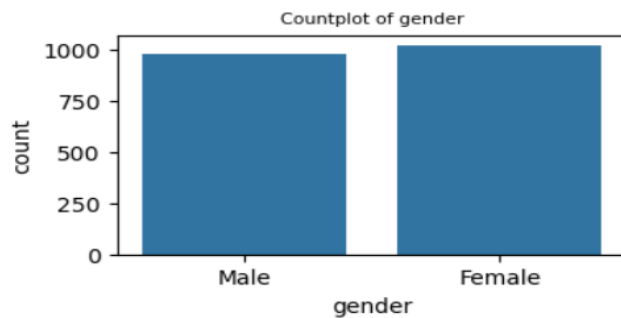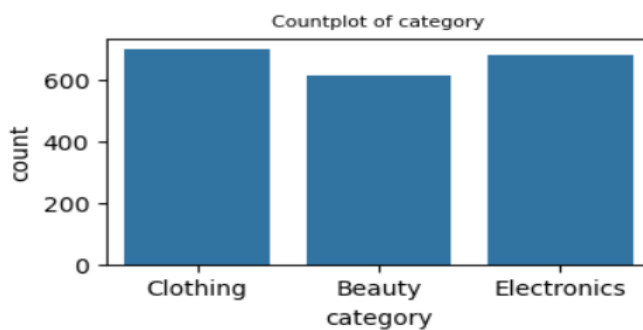

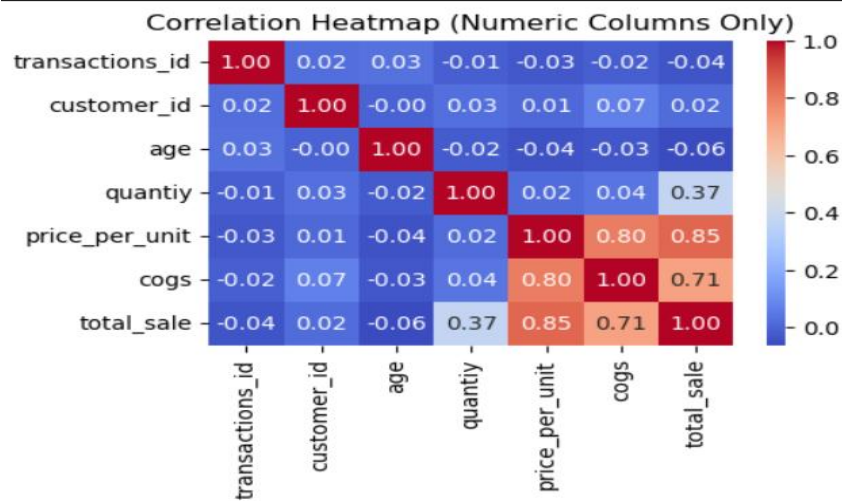Countplot of gender

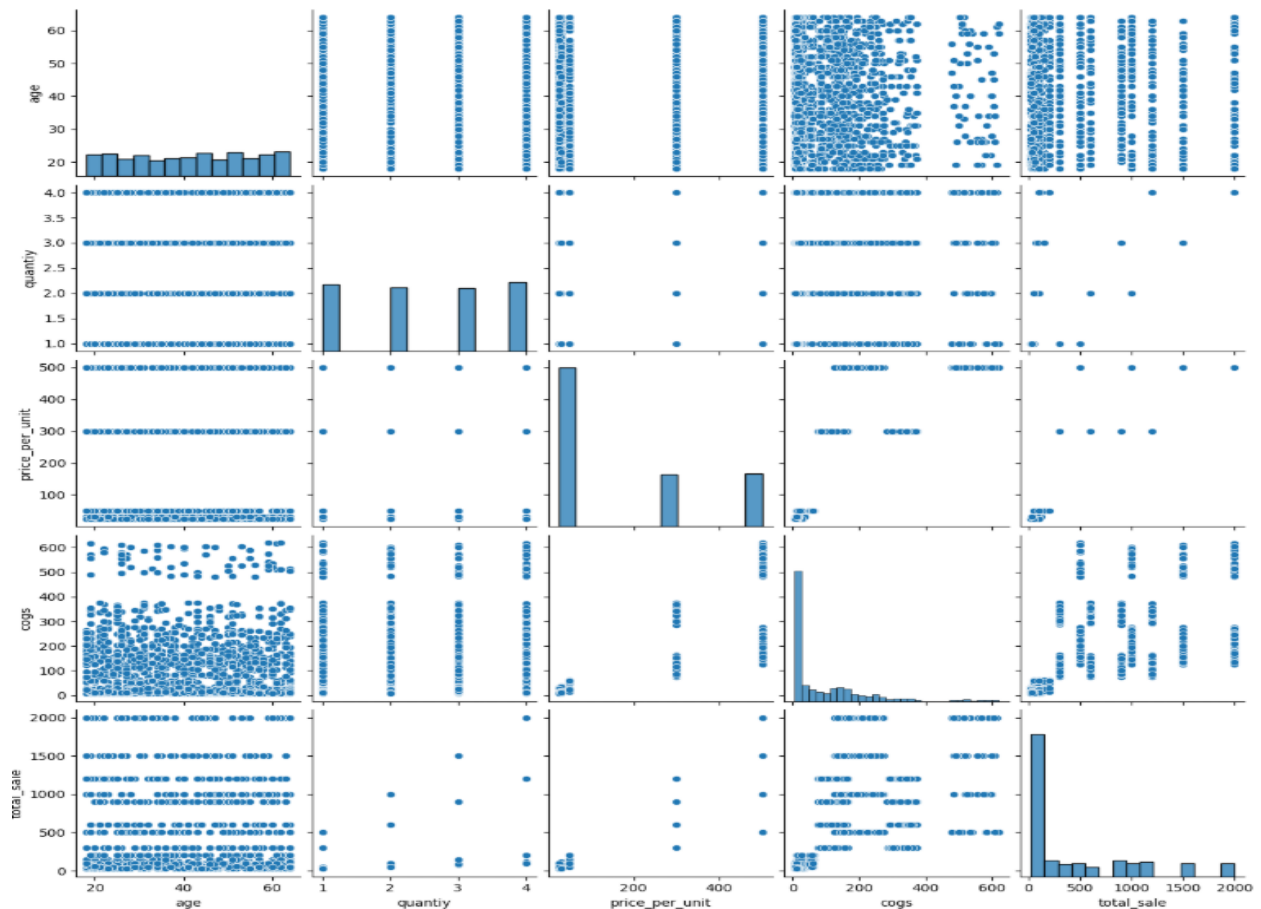**Figure 3.1: Countplot of category>**


Countplot of category

## 3.2 Bivariate / Multivariate Analysis

- Correlation Heatmap: Numerical columns only

**Figure 3: Correlation Heatmap** >



- **Pairplot**

- **Grouping Analysis:**

- **Figure 4.1: Total Sales by Product Category>**



Total Sales by Product Category
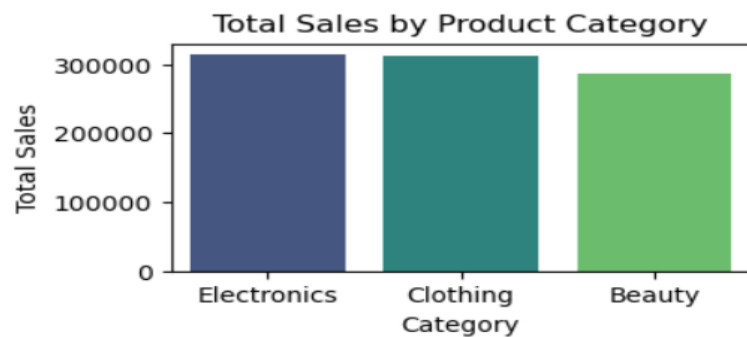
- **Figure 4.2: Total Sales by Gender >**



Total Sales by Gender
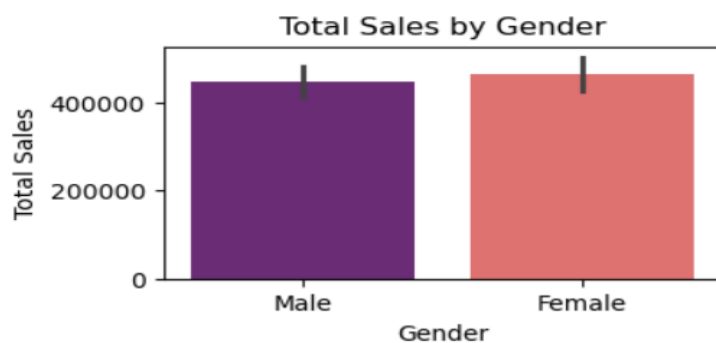
**Table 1: Average Sales per Category**

```
     category   total_sale
0      Beauty   468.692810
1    Clothing   443.751783
2 Electronics   458.786550
```

**Table 2: Total Sales per Category**

```
     category   total_sale
0      Beauty     286840.0
1    Clothing     311070.0
2 Electronics     313810.0
```

**Table 3: Average & Total Sales per Gender**

```
   gender  total_sale
0  Female  457.620452
1    Male  455.428571
   gender  total_sale
0  Female    465400.0
1    Male    446320.0
```

**Table 4: Total Sales by Category and Gender**

```
       category  gender  total_sale
0        Beauty  Female    149470.0
1        Beauty    Male    137370.0
2      Clothing  Female    162460.0
3      Clothing    Male    148610.0
4   Electronics  Female    153470.0
5   Electronics    Male    160340.0
```

**Figure 5: Stacked Bar Chart – Category-wise Sales by Gender** >
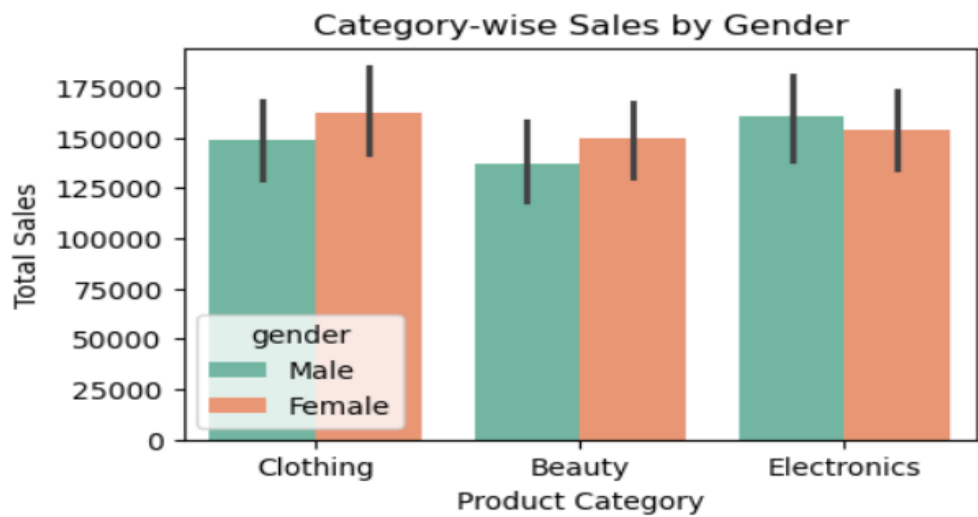
## 4. Outliers & Skewness

- Boxplots for outliers:



Boxplot for Outlier Detection
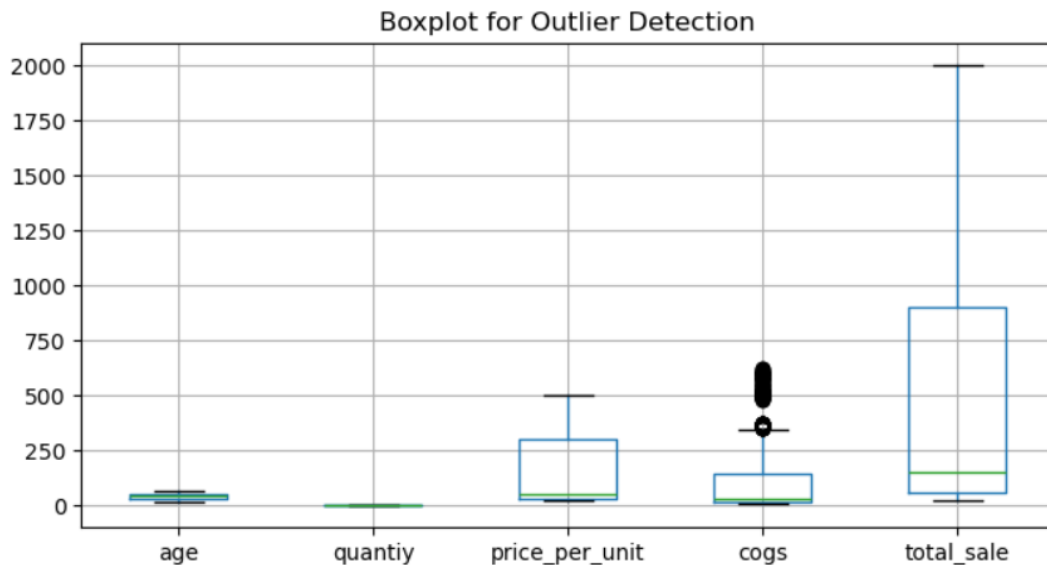
.

## 5. Key Insights / Observations

- **Category-wise Sales:** Electronics has the highest total sales, while Beauty products have the highest average sale per transaction.
- **Gender-wise Sales:** Female customers contributed slightly more to total and average sales compared to male customers.
- **Category + Gender:** Electronics sold slightly more to males, while Clothing and Beauty sold more to females.
- **Age Distribution:** Most customers fall into middle age groups, indicating core target customers for sales.
- **Outliers & Skewness:** Some numerical columns (e.g., Total Sale, Price per Unit) had outliers and slight skewness, which were handled appropriately to avoid distortion in analysis.

## 6. Conclusion / Recommendations

The exploratory data analysis provided meaningful insights into the retail sales dataset. The analysis revealed that Electronics has the highest total sales, while Beauty products have the highest average sale per transaction. Female customers contributed slightly more to total and average sales than male customers. Category-wise and

gender-wise grouping highlighted clear patterns in purchasing behavior. Outliers and skewed distributions were identified and handled appropriately. These findings can help the business make informed decisions regarding inventory management, marketing strategies, and customer targeting to optimize sales performance.

## 7. Tools / Libraries Used

- Python
- Pandas
- Matplotlib
- Seaborn
- Jupyter Notebook