

## TASK -2

### DATASET: SUPER MARKET SALES

#### STEP-1:

Display basic information about the dataset

##### 1. Dataset Structure

- **Total Rows:** 1000
- **Columns:** 9
- **Missing Values:** No missing data detected.
- **Data Types:**
  - Categorical: Gender, Invoice ID, Branch, City, Customer type, Product line
  - Numerical: Unit price, Quantity, Tax 5%

##### 2. Summary Statistics

- Unit price ranges from **10.08** to **99.96**, mean is **55.67**
- Quantity ranges from **1** to **10**, with an average of **5.51**
- Tax 5% ranges from **1.07** to **985.00** (needs investigation—possibly an outlier)

##### 3. Unique Values

- 6 product lines, 3 branches, 2 customer types, and 2 genders
- `Invoice ID` is unique (likely the identifier)

#### STEP-2

1. Visualize missing data heatmap
2. Distribution: Histograms
3. Boxplots for outlier detection
4. Countplot for categorical variables
5. Pairplot for relationship overview

### Insights from Visual EDA

#### 1. Missing Data:

- **No missing values:** Confirmed with the `sns.heatmap()` — dataset is clean.

#### 2. Distribution & Outliers:

- **Unit Price:** Roughly normal, some high-value outliers.
- **Quantity:** Slight right skew, ranges 1–10, looks fine.

- **Tax 5%:** Right-skewed, a few **large outliers** — worth investigating.

### 3. Boxplots:

- Helped detect:
  - **Outliers** in `Tax 5%` and `Unit price`.
  - **Reasonable spread** in `Quantity`.

### 4. Category Counts:

- **Gender** and **Customer Type** are evenly split.
- **Product Line** distribution varies:
  - **Fashion accessories** and **Food and beverages** most common.
- **City** and **Branch** are balanced — likely 1 branch per city.

### 5. Pairplot:

- Visual correlations:
  - `Tax 5%` grows linearly with both `Unit price` and `Quantity` (as expected).
  - No unexpected anomalies.

### STEP-3:

- **Correlation Heatmap**
- **Grouping and Aggregation: Category-level Patterns**

### Correlation Heatmap (with Total Sales)

Feature 1	Feature 2	Correlation
Unit price	Tax 5%	<b>0.99</b>
Quantity	Total Sales	<b>0.58</b>
Unit price	Total Sales	0.60
Quantity	Tax 5%	0.58

## Groupby + Aggregation Analysis

### 1.By Product Line:

Product Line	Total Sales (Sum)	Total Sales (Avg)	Quantity (Sum)	Quantity (Avg)
Food and beverages	₹53,471	₹307	952	5.47
Sports and travel	₹52,498	₹316	920	5.54
Home and lifestyle	₹51,297	₹321	911	5.69
Fashion accessories	₹51,720	₹291	902	5.07
Electronic accessories	₹51,750	₹304	971	5.71
Health and beauty	₹46,851	₹308	854	5.62

#### Insights:

- **Top sales (sum):** Food & beverages
- **Highest per-sale average:** Home & lifestyle
- **Highest quantity sold:** Electronic accessories

### 2 By Branch:

Branch	Total Sales (Sum)	Avg per Sale
A	₹101,143	₹297
B	₹101,141	₹305
C	₹105,304	₹321

**Insight:** Branch C outperforms others both in total and average sale value.

### 3 By Customer Type:

Customer Type	Total Sales (Sum)	Avg per Sale
Member	₹156,403	₹312
Normal	₹151,184	₹303

**Insight:** Members contribute slightly more in both sales volume and sale value.