Exploratory Data Analysis (EDA) by Deepika Bhukya

1. Introduction:

In this section, the goal of the Exploratory Data Analysis (EDA) was to understand the structure, patterns, and potential insights of the dataset. The analysis involves identifying important relationships between different features, understanding the distribution of numerical features, and preparing the data for further analysis or modeling.

2. Data Overview:

The dataset contains customer information, product details, and transaction data. Key features include:

- **CustomerID**: Unique identifier for each customer.
- **ProductID**: Unique identifier for each product.
- **Quantity**: Number of items purchased.
- **Price**: Price of the product.
- **TotalValue**: Total value of the transaction.

3. Data Cleaning:

During EDA, the data was checked for missing values, duplicated rows, and outliers. Any inconsistencies were handled by:

- Removing rows with missing or invalid data.
- Correcting errors in column formats (e.g., date formatting).

4. Univariate Analysis:

Univariate analysis was performed on individual features. Key findings:

- **Distribution of Product Prices**: A histogram was plotted to visualize the price distribution of products.
- **Total Value of Transactions**: Box plots and histograms were used to identify the spread and potential outliers in the total value spent by customers.

(Insert histograms/plots here)

5. Bivariate Analysis:

In the bivariate analysis, relationships between pairs of features were analyzed:

- **Price vs. Quantity**: A scatter plot was used to check the relationship between the price of products and the quantity purchased.
- **TotalValue vs. Quantity**: A correlation heatmap showed that there is a positive relationship between the total value spent and the quantity purchased.

(Insert scatter plot/correlation heatmap here)

6. Insights:

- **Top Spending Products**: Identifying the products with the highest total value and the most frequently purchased items.
- **Customer Segments**: Customers who spend more tend to purchase higher quantities of products.

7. Conclusion:

Through EDA, we were able to uncover several trends in the data, including the purchasing behavior of customers, the most popular products, and the relationships between various features. This analysis sets the stage for deeper analysis or modeling, such as clustering and predictive modeling.