

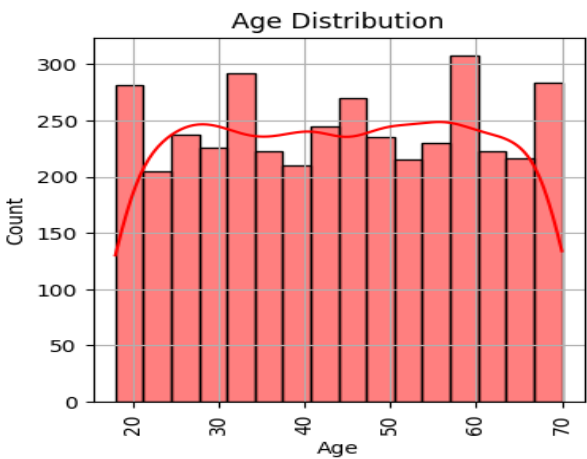
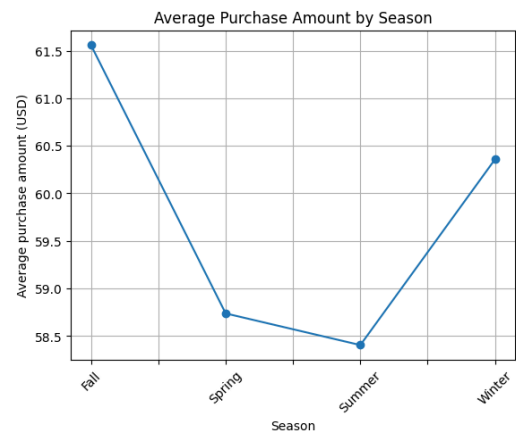
Introduction

This report analyzes a dataset containing information about customer demographics and purchasing behavior. The attributes include age, purchase amount (USD), and other relevant features. The analysis focuses on three key areas:

- 1. Identifying Relationships: Exploring relationships between customer characteristics and purchasing behavior.
- 2. Grouping Customers: Grouping customers with similar profiles using clustering techniques.
- 3. Predicting Purchase Amount: Predicting purchase amounts based on customer age using linear regression.

Relational Graph

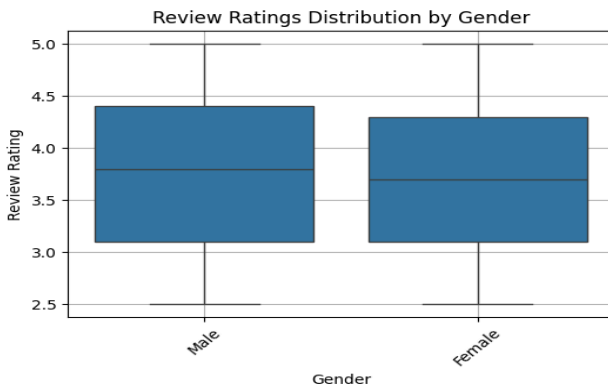
A line plot is generated to visualize the relationship between seasons and average purchase amounts. This analysis reveals seasonal trends in purchasing behavior, which could be useful for marketing and sales strategies.



Categorical Graph

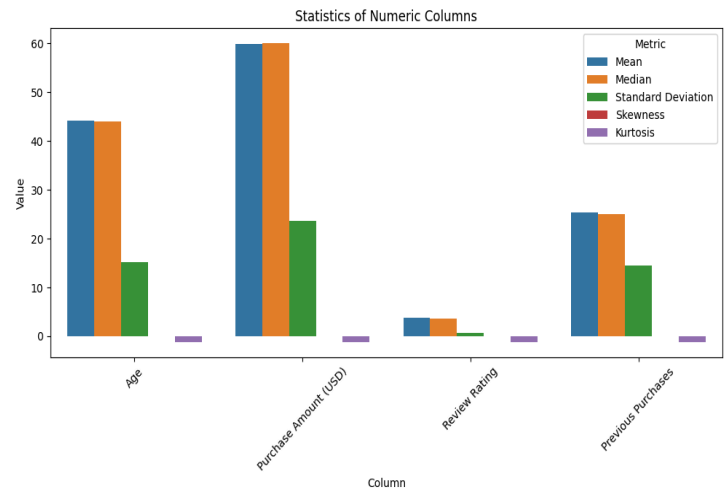
A histogram is created to explore the distribution of customer ages across the dataset. This helps identify whether age is concentrated in a particular range or spread evenly, which can inform targeted marketing efforts.

GITHUB REPO : <https://github.com/Osamasaaid11/clustering-and-fiting-submission>



A box plot is created to visualize the distribution of review ratings by gender. This helps in understanding the spread and central tendency of review ratings across different genders.

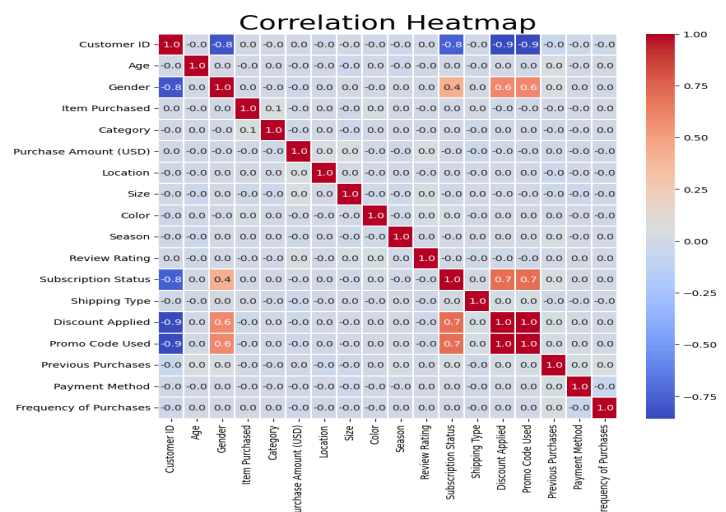
A **bar chart** provides a summary of statistical metrics for numerical columns in the dataset, including mean, median, and standard deviation. This visualization offers an overview of the distribution and variability of numeric features, aiding in understanding the dataset's characteristics.



A statistical analysis of the numerical features in the dataset reveals the following:

	Mean	Median	Standard Deviation	Skewness	Kurtosis
Age	44.06846153846154	44.0	15.207589127162382	-0.006377267719668403	-1.195093602981543
Purchase Amount (USD)	59.76435897435898	60.0	23.685392250875307	0.012696871813012472	-1.2365469509064881
Review Rating	3.7499487179487176	3.7	0.7162228139312417	0.004522856027108999	-1.179654568180405
Previous Purchases	25.35153846153846	25.0	14.447125170462309	0.003119954940061591	-1.1902001181924076

A correlation heatmap was generated to visualize relationships between numerical features. The heatmap highlights both positive and negative correlations, aiding in identifying key interdependencies among attributes. This analysis is crucial for understanding how

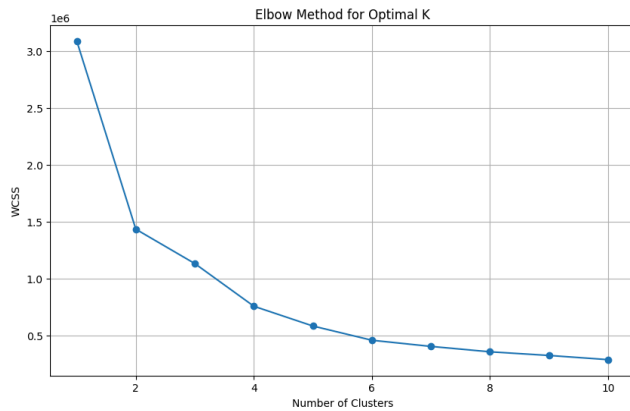


Student Name: Osama said

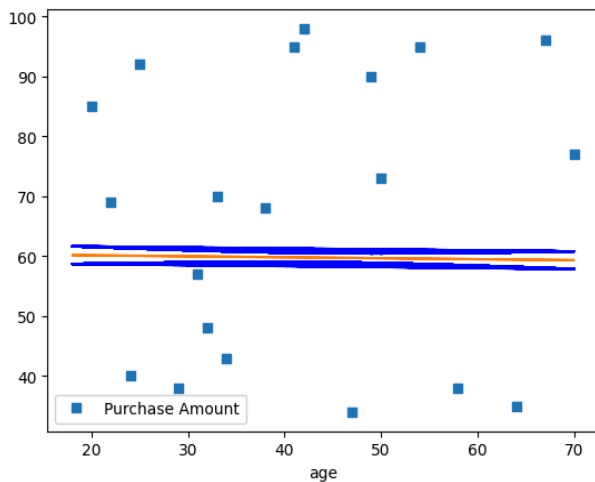
Student ID: 23031848

GITHUB REPO : <https://github.com/Osamasaaid11/clustering-and-fiting-submission>

different variables influence each other in the dataset.



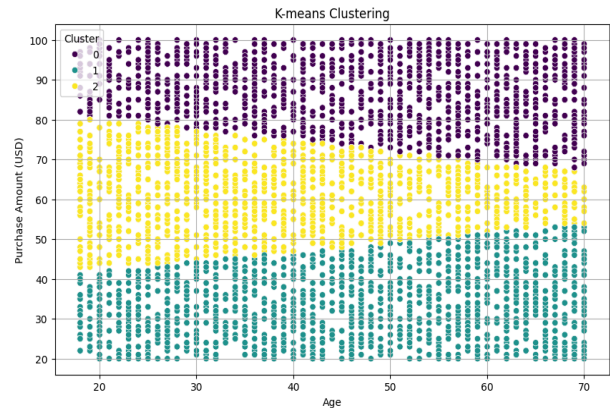
older customers with higher average purchase amounts, and customers with lower purchase amounts regardless of age. These insights can inform targeted marketing strategies to enhance customer engagement and drive sales growth.



- **High Prediction Error:** Mean squared error (560.793) indicates high prediction error, with an average deviation of 23.7 units.

Clustering Analysis:

The K-means clustering analysis grouped customers into three distinct clusters based on age and purchase amount. Using the elbow method, the optimal number of clusters was determined to be three. The clusters revealed distinct customer segments, including younger customers with varying purchase amounts,



Regression Analysis:

Linear regression was used to model the relationship between customer age and purchase amount (USD). The results reveal:

- **Negative Relationship:** Model coefficient (-0.016) indicates a slight negative association between age and purchase amount.
- **Model Intercept:** Intercept (60.480) represents the predicted purchase amount when age is zero.

