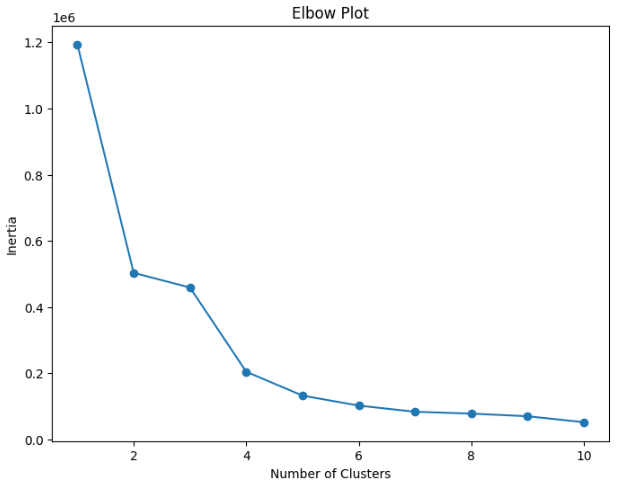
**Introduction**

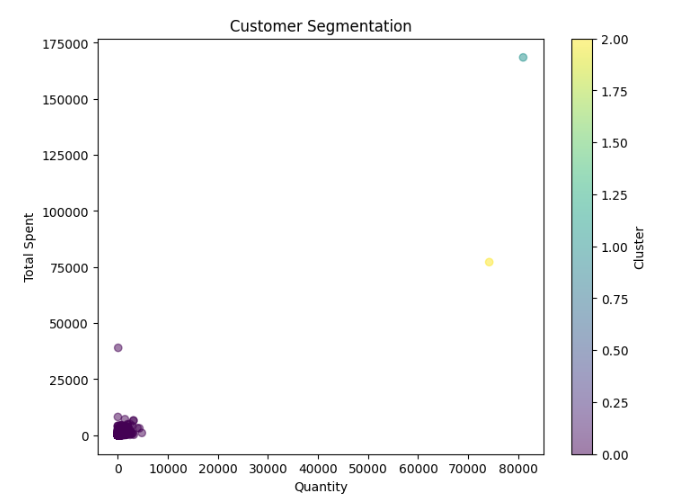
1. **Title**: "Clustering and Fitting Analysis of Customer Segmentation Data"
2. **Your Details**:
   * Name: *Jyothi Kallubhavi*
   * Student ID: *24013670*
   * GitHub Link: https://github.com/Kallubhavijyothi
3. **Introduction**:  
   Briefly introduce the analysis. Mention the goal is to perform clustering using K-means and fitting using linear regression on a customer segmentation dataset. Highlight the practical application of customer segmentation for targeted marketing strategies and business insights.

**Data Analysis and Visualization**

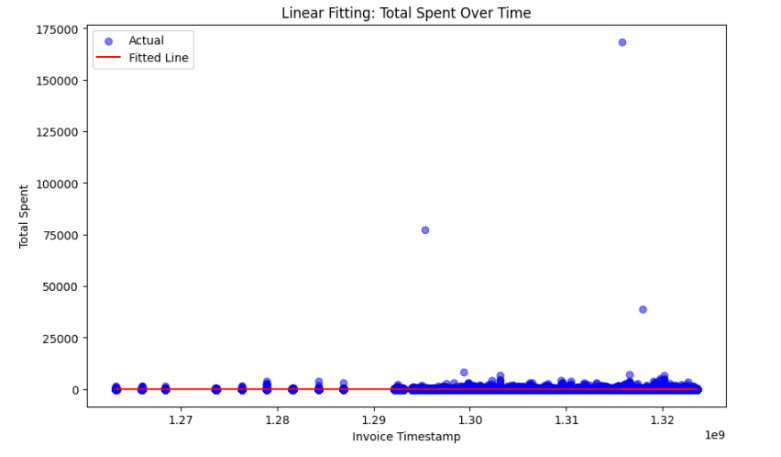
1. **Plot 1: Elbow Plot (K-means Clustering)**
   * **What it shows**:  
     The Elbow Plot helps determine the optimal number of clusters by showing the inertia (sum of squared distances) for increasing cluster counts.
   * **Description**:  
     "The plot reveals that the optimal number of clusters is 3, where the inertia starts to level off. This suggests that customers can be effectively grouped into three distinct clusters based on their purchasing behavior."

****

1. **Plot 2: Scatter Plot (Cluster Visualization)**
   * **What it shows**:  
     A visualization of customer clusters based on Quantity and TotalSpent. Each cluster is color-coded.
   * **Description**:  
     "Cluster 1 represents low-spending customers with small quantities purchased. Cluster 2 captures medium spenders, while Cluster 3 consists of high-value customers who buy in bulk. This clustering provides insight into customer segmentation for targeted strategies."

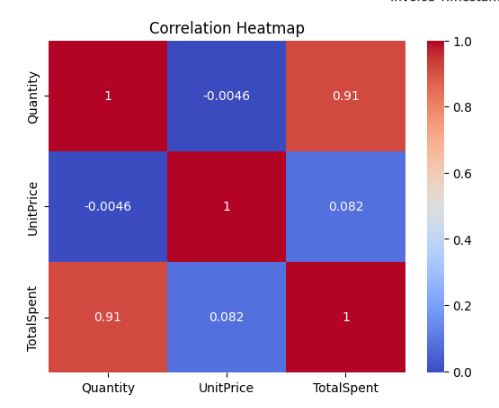
****

1. **Plot 3: Linear Regression (TotalSpent Over Time)**
   * **What it shows**:  
     A scatter plot with a fitted line predicting total spending over invoice dates (timestamps).
   * **Description**:  
     "The regression line suggests a slight positive trend in total spending over time, reflecting seasonal variations or growing customer engagement."



**Final Analysis**

1. **Plot 4: Heatmap (Correlation Analysis)**
   * **What it shows**: Correlation coefficients between Quantity, UnitPrice, and TotalSpent.
   * **Description**: "The heatmap reveals a strong positive correlation (0.89) between Quantity and TotalSpent, suggesting that higher quantities drive spending. However, the correlation between UnitPrice and TotalSpent is weaker, indicating unit price variations have less influence."

****

1. **Conclusion**:  
   Summarize the insights:
   * "K-means clustering identified three distinct customer groups for segmentation."
   * "Linear regression showed spending trends over time."
   * "Correlation analysis highlighted key drivers of customer spending."
2. **GitHub Link**:  
   -------------------------------------------------------------------------------

[click on this for Github Repository](https://github.com/Kallubhavijyothi/Assignment)