**Customer Segmentation Analysis Report**

**Introduction**

**Objective:**

The objective of this analysis is to segment customers based on their purchasing behavior using clustering techniques. By identifying distinct customer segments, we aim to tailor marketing strategies and product offerings to improve customer satisfaction and increase sales.

**Data Source and Description:**

The dataset used for this analysis is a sales dataset containing various features such as:

* Order details
* Product information
* Sales figures
* Discounts
* Profits
* Shipping costs
* Customer information

**Data Preprocessing**

**Handling Missing Values and Outliers:**

* **Missing Values:**
  + Numeric columns with missing values were filled using the mean.
  + Categorical columns with missing values were filled using the mode.
* **Outliers:**
  + Outliers were not explicitly handled in this analysis but should be considered in further iterations for more accurate clustering.

**Feature Selection and Scaling:**

* **Feature Selection:**
  + The features selected for clustering include:
    - Ship Mode
    - Product Category
    - Sales
    - Quantity
    - Discount
    - Profit
    - Shipping Cost
    - Order Priority
    - Segment
* **Scaling:**
  + Numeric features were scaled using StandardScaler to normalize the data and ensure all features contribute equally to the clustering process.

**Clustering**

**Explanation of K-Means Clustering:**

K-Means is a partitioning clustering algorithm that aims to divide a set of observations into K clusters, where each observation belongs to the cluster with the nearest mean. It minimizes the within-cluster variance and assigns cluster centers iteratively.

**Determining the Optimal Number of Clusters (Elbow Method):**

The Elbow method was used to determine the optimal number of clusters by plotting the within-cluster sum of squares against the number of clusters and identifying the "elbow" point where the rate of decrease slows down.