**Executive Summary**

**Project Title: Market Segmentation Using Machine Learning**

**Objective:**

The primary goal of this project is to perform customer segmentation using unsupervised machine learning techniques. The intent is to identify distinct customer profiles based on transactional behavior, enabling targeted marketing strategies and data-driven business decisions.

**Data Overview:**

* **Dataset**: Customer data consisting of 8,950 entries with 18 attributes (including customer ID, balance, frequency of purchases, credit limit, cash advances, and other card usage metrics).
* **Source**: CSV file titled Customer Data.csv.
* **Key Attributes**: BALANCE, PURCHASES, ONEOFF\_PURCHASES, INSTALLMENTS\_PURCHASES, CASH\_ADVANCE, CREDIT\_LIMIT, PAYMENTS, MINIMUM\_PAYMENTS, and transaction frequencies.

**Data Preparation**

1. **Missing Value Treatment**:
   * Missing values in MINIMUM\_PAYMENTS and CREDIT\_LIMIT were imputed using the mean.
2. **Feature Reduction**:
   * CUST\_ID was dropped as it is a unique identifier and not useful for clustering.
3. **Outlier Detection and Handling**:
   * Used IQR-based technique to identify outliers.
   * Applied **Winsorization** (capping method) to reduce the influence of extreme values.
4. **Feature Scaling**:
   * Standardization was applied using **StandardScaler** to normalize data distribution before clustering.

**Exploratory Data Analysis (EDA):**

* **Visualization Tools**: Used seaborn and matplotlib for KDE plots, boxplots, violin plots, pairplots, and correlation heatmaps.
* **Insights**:
  + Strong correlation observed between PURCHASES and ONEOFF\_PURCHASES.
  + Cash advance features were negatively correlated with purchase behaviors.

**Dimensionality Reduction:**

* **PCA (Principal Component Analysis)**:
  + Reduced high-dimensional data to 2 principal components for effective visualization of clusters.

**Clustering Technique:**

* **Model Used**: **K-Means Clustering**
* **Optimal Clusters**: Identified as **4** using the **Elbow Method**.
* **Cluster Characteristics**:
  1. **Cluster 0**: High purchases with low cash advance—likely active and reliable users.
  2. **Cluster 1**: High balance and high cash advance—potentially risky or credit-seeking customers.
  3. **Cluster 2**: Very high purchases and high credit limits—premium or high-value customers.
  4. **Cluster 3**: Low balance, low purchases—possibly dormant or low-engagement customers.
* **Visualized Using**: Scatter plots of PCA-reduced data colored by cluster.

**Cluster Profiling:**

Each cluster was evaluated and labeled based on transaction behavior, which can be used for tailored marketing strategies:

* **Cluster 0**: Consistent Installment Shoppers
* **Cluster 1**: High Cash Advance Users
* **Cluster 2**: Affluent Frequent Spenders
* **Cluster 3**: Low Activity/Inactive Users

**Model Evaluation & Deployment:**

1. **Model Export**:
   * K-Means model saved using joblib (kmeans\_model.pkl).
2. **Supervised Learning for Prediction**:
   * **Decision Tree Classifier** trained to classify customers into clusters.
   * **Accuracy**: Achieved **~92%** on test data.
   * Confusion matrix and classification report confirm high precision, recall, and F1-scores across all clusters.
3. **Deployment Ready**:
   * Final model exported as final\_model.sav using pickle.
   * Cluster-labeled data saved as Clustered\_Customer\_Data.csv for future use.

**Key Business Value:**

* **Marketing Personalization**: Enables targeted campaigns based on behavioral clusters.
* **Risk Mitigation**: Identifies risky profiles for credit risk evaluation.
* **Customer Retention**: Insights into inactive users can guide re-engagement strategies.
* **Product Development**: Helps align offerings with segment-specific needs.

**Conclusion:**

This project successfully demonstrates how unsupervised learning (K-Means) and dimensionality reduction (PCA) can uncover meaningful customer segments. The accuracy of the supervised Decision Tree model further supports its integration into real-world customer analytics pipelines for banking, finance, and retail sectors.