OMKAR SAJJANSINGH CHAUHAN | Machine Learning | Week 2 Task

Task Description

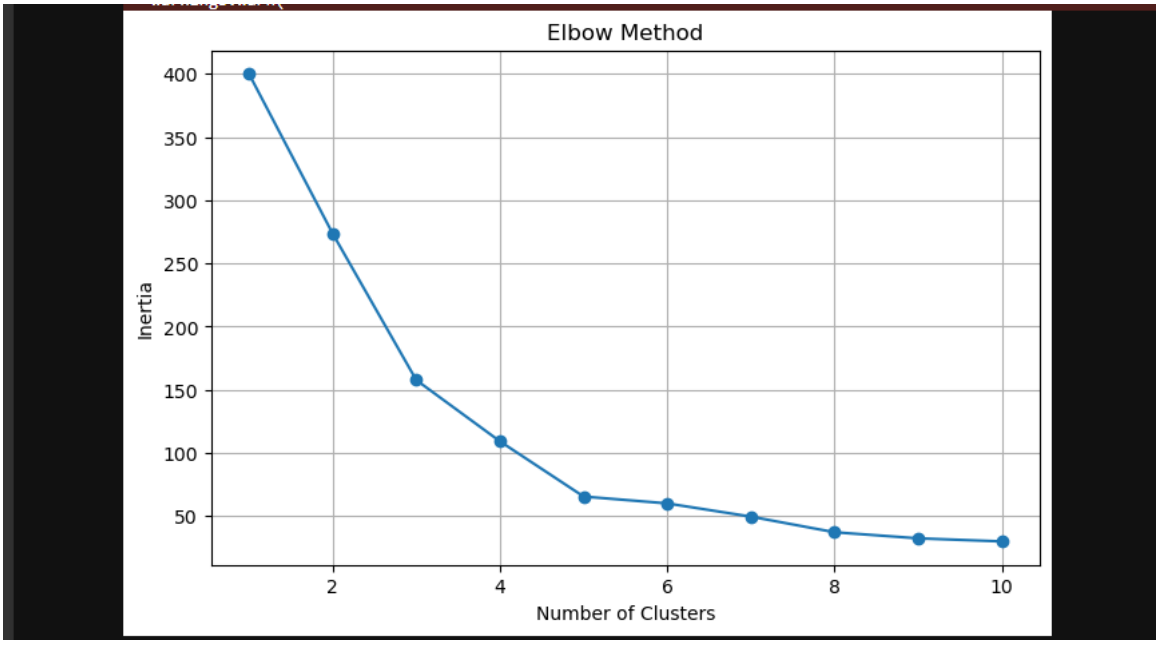
In this task, I implemented a K-Means clustering algorithm to perform customer segmentation using a dataset from Kaggle. The objective was to group customers based on purchase patterns and gain business insights through unsupervised learning. I performed data preprocessing, used the Elbow Method to determine the optimal number of clusters, and visualized the results.

Tools & Platforms Used

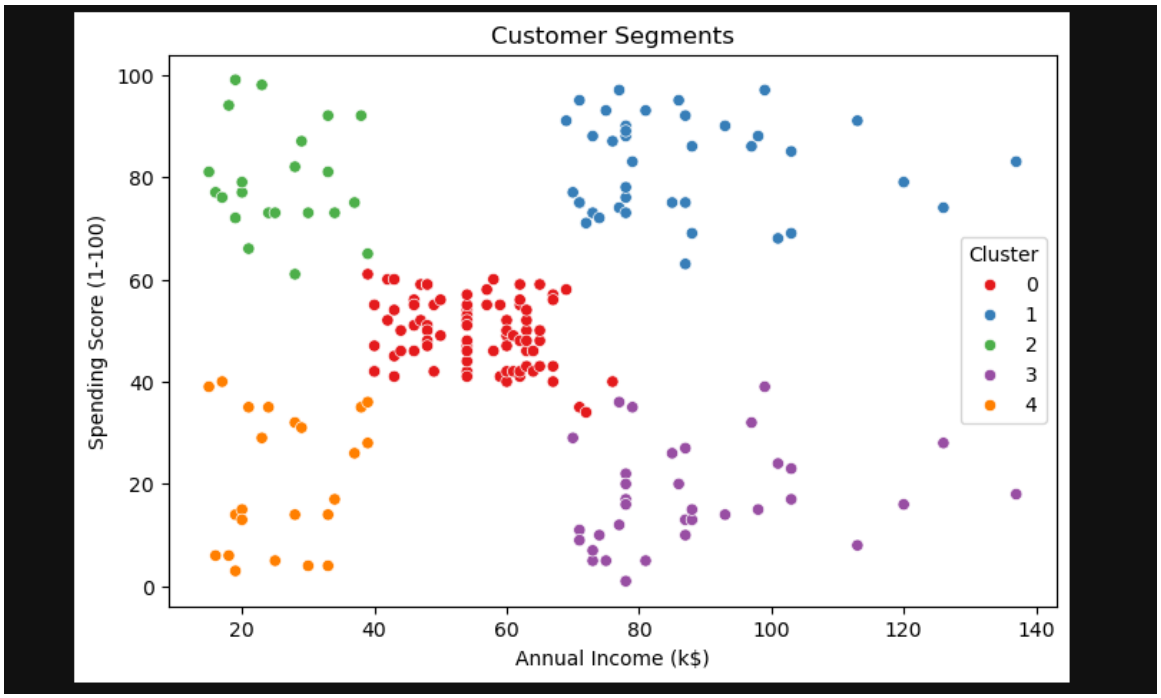
- Python
- Jupyter Notebook
- Scikit-learn
- Pandas & NumPy
- Matplotlib & Seaborn
- Plotly
- GitHub

Screenshots / Links

GitHub Repo : https://github.com/BhaveshThakur-57/Credora-Internship-Machine-Learning-/tree/main/Task%202



Elbow Method to Determine Optimal Clusters



Visualize the Clusters

Challenges Faced & How I Solved Them

1. Determining the Optimal K: Initially uncertain how to choose the number of clusters. Solved using the Elbow Method and Silhouette Score for better clarity.

2. Data Normalization: Features with different scales affected clustering. Resolved using `StandardScaler` to normalize data before clustering.

3. Cluster Interpretation: After clustering, interpreting what each cluster represented in business terms was tricky. I analyzed average values in each cluster to understand behavior patterns like high spenders or frequent shopper

4. Visualization: Used Seaborn and optionally Plotly for plotting clusters. Added centroid markers and color-coding for clarity.