

Capstone Mini Project: Customer Segmentation

Objective

Cluster mall customers into different segments based on purchasing behavior and demographics. Evaluate and visualize the clusters.

Step 1: Load and Understand the Data

Use pandas to load the Mall Customer Segmentation dataset. Conduct basic exploratory data analysis (EDA) to understand the dataset structure.

Step 2: Data Preprocessing & Feature Engineering

- Select numerical features (e.g., Age, Income, Spending Score).
- Standardize data using StandardScaler.
- Handle categorical features using one-hot encoding if needed.

Step 3: K-Means Clustering

- Use the Elbow Method to identify the optimal number of clusters.
- Apply KMeans clustering and obtain cluster labels.

Step 4: Dimensionality Reduction with PCA

- Use PCA to reduce features to 2 components for visualization.
- Plot the customer segments in a 2D space with colors indicating clusters.

Step 5: Evaluate and Tune with Grid Search

- Create a Scikit-learn pipeline with scaling and KMeans.
- Use GridSearchCV to find the optimal number of clusters.

Expected Outcomes

- Cluster assignments for each customer.
- 2D PCA visualization of clusters.
- Reusable ML pipeline for clustering tasks.

Optional Extensions

- Try DBSCAN or Hierarchical Clustering.
- Add Silhouette Score for cluster evaluation.
- Build a Plotly dashboard for interactive visualization.