

Mall Customers Segmentation Using Python

Project Title:

Mall Customers Segmentation

Objective:

The objective of this project is to segment mall customers into different groups based on their purchasing behavior and demographic data using clustering techniques. This helps businesses understand customer types, enabling targeted marketing and improved customer service.

Tools & Libraries Used:

- Python - Core programming language
- Pandas - Data manipulation
- NumPy - Numerical computations
- Matplotlib & Seaborn - Data visualization
- Scikit-learn - Machine learning (clustering algorithms)

Dataset Overview:

The dataset contains information about mall customers, including:

- CustomerID
- Gender
- Age
- Annual Income (k\$)
- Spending Score (1-100)

These attributes are used to understand patterns and segment the customers.

Steps Implemented:

1. Data Loading & Preprocessing:

- Loaded dataset using `pandas.read_csv()`.
- Basic checks using `info()`, `describe()` and `isnull().sum()` to confirm data integrity.
- Cleaned and encoded categorical variables (e.g., Gender using `LabelEncoder`).

2. Exploratory Data Analysis (EDA):

- Univariate analysis of Age, Income, and Spending Score using histograms and boxplots.
- Bivariate analysis using scatter plots and pairplots.
- Heatmaps used to observe correlations between numerical features.

3. Feature Selection for Clustering:

- Chose Annual Income and Spending Score as the key variables for segmentation (2D clustering).
- Applied `StandardScaler` for normalization before clustering.

4. K-Means Clustering:

- Used Elbow Method to determine the optimal number of clusters (k).
- Fitted the model using `KMeans(n_clusters=k)`.
- Visualized the clustered customer segments.

5. Cluster Analysis:

- Each cluster is analyzed based on mean income and spending score:
- High spenders with high income
- Low income low spenders
- Young moderate spenders, etc.

- Visualized with cluster labels on a scatter plot for interpretation.

Outcomes:

- Successfully segmented mall customers into distinct groups.
- Visual insights provide clear business strategies for targeting each group.
- Helped identify key customer profiles, such as:
 - High income, low spending (potential upsell targets) - Low income, high spending (value-focused customers)

Conclusion:

This project demonstrates a practical implementation of unsupervised learning via K-Means clustering for real-world business use. Mall customer data was analyzed and grouped effectively, offering valuable insights into consumer behavior and enabling better decision-making for marketing teams.

Future Scope:

- Add more features like product categories, visit frequency, etc.
- Implement hierarchical clustering or DBSCAN for comparison.
- Build an interactive dashboard using Streamlit or Power BI.
- Integrate with CRM systems for real-time personalization.