Project Plan

Tasks:

- Data collection :

Gather necessary Mall Customers Data

- Data Cleaning :

Remove duplicates and handle missing values.

Handle outliers appropriately.

Normalize or scale numerical features if needed.

- Exploratory Data Analysis (EDA):

Perform descriptive statistics.

Visualize distributions of features (age, income, spending score).

Analyze correlations and patterns.

Identify key characteristics of different segments.

- Clustering :

Apply K-means clustering algorithm to cleaned data after scaling

Determine optimal number of clusters using methods like the Elbow method or Silhouette score.

Train clustering model and assign clusters to data points.

Interpret and name each cluster based on its characteristics.

- Visualization:

Create visualizations to represent clusters and their attributes.

- Documentation:

Document all steps and methodologies used.

Record insights and interpretations from EDA and clustering.

Prepare a final report or presentation summarizing findings and recommendations.

Timeline :

13/07/2024: Gather and clean data

14/07/2024: Perform EDA

15/07/2024: Clustering

16/07/2024: Model enhancement and visualization

17/07/2024: Documentation

Resources :

Data : Mall customer.csv

Software : jupyter notebook, Python libraries like pandas, numpy, scikit-learn, matplotlib and seaborn

Hardware : Computer system with analytical processing power

Personnel : Data analyst with expertise in preprocessing, eda, clustering and visualization.

Risks :

- Data quality issues :

Inaccurate and inconsistent data may sabotage the analysis.

Perform data cleaning to mitigate it.

- Algorithm performance :

K-means may not segment customers into good clusters.

To mitigate, select optimal number of clusters and define random state

- Visualization Limitations :

Visuals may not communicate insights effectively.

Use a variety of insights and improve based on feedback