Project Plan

1. Tasks

The project will be divided into the following key tasks:

Data Collection

- Identify and gather the Mall Customers dataset.
- Ensure data is in a suitable format for analysis.

Data Cleaning

- Handle missing values (imputation or removal).
- Remove duplicates.
- Transform categorical variables into numerical format.
- Normalize or standardize numerical features

Exploratory Data Analysis (EDA)

- Analyze customer demographics and purchasing behavior.
- Generate visualizations to identify trends and patterns.
- Summarize findings from the EDA.

Clustering

- Implement the K-Means clustering algorithm.
- Determine the optimal number of clusters using methods like the Elbow Method.
- Validate and assess the quality of the clusters.

Visualization

- Create visualizations using Matplotlib and Seaborn to represent customer segments.
- Develop interactive dashboards in Power BI for stakeholder presentation.

Documentation

- Document the methodology, findings, and insights.
- Prepare a final report and presentation for stakeholders.

2. Timeline

The project timeline is structured over a 3-month period with key milestones:

Task	Start Date	End Date	Milestone
Data Collection	July 14, 2024	July 16, 2024	Dataset acquired
Data Cleaning	July 16, 2024	July 18, 2024	Data cleaned
Exploratory Data Analysis	July 18, 2024	July 20, 2024	EDA completed
Clustering	July 20, 2024	july22, 2024	Clustering completed
Visualization	july22, 2024	July 23, 2024	Visualizations created
Documentation	July 23,2024	July 24,2024	Final report drafted
Review and Presentation	July 24,2024	July 25,2024	Presentation delivered

3. Resources

The following resources will be needed to successfully complete the project:

Human Resources:

- Data Analyst/Scientist: Responsible for data analysis, cleaning, and modeling.
- Business Analyst: To assist with requirements gathering and stakeholder communication.
- IT Support: To ensure the necessary software and hardware are available.

Software/Tools:

- Python (with libraries: Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn)
- Jupyter Notebook for development and analysis.
- Power BI for visualization and reporting.

Hardware:

- A computer with sufficient processing power and memory to handle data analysis tasks.

4. Risks

Identifying potential risks is crucial for project success. Here are some risks associated with the project:

Data Quality Issues:

- **Risk**: The dataset may contain missing values, outliers, or inconsistencies that could affect analysis.
- **Mitigation**: Implement thorough data cleaning and validation processes. Use data profiling techniques to identify and address quality issues early.

Algorithm Performance:

- **Risk**: The K-Means algorithm may not perform well if the data is not well-prepared or if the optimal number of clusters is not correctly identified.
- **Mitigation**: Experiment with different preprocessing techniques and validate clustering results using metrics such as Silhouette Score. Consider alternative clustering algorithms if necessary.

Visualization Limitations:

- **Risk**: The visualizations may not effectively communicate insights or may be limited by the capabilities of the chosen tools.
- **Mitigation**: Plan visualizations in advance, ensuring they align with the insights to be communicated. Seek feedback from stakeholders on visualizations to ensure clarity and effectiveness.

Conclusion

This project plan outlines the key tasks, timeline, resources, and risks associated with your customer segmentation project. By following this plan, you can ensure a structured approach to achieving your project objectives and delivering valuable insights to the retail store.