Project Plan 1. Tasks

1.1 Data Collection

- Gather the mall customers dataset.
- Verify the dataset's completeness and accuracy.

1.2 Data Cleaning

- Address any missing values.
- Remove duplicate entries.
- Normalize the data for consistency.
- Scaling data using Min-Max Scaling.

1.3 Exploratory Data Analysis (EDA) ·

Perform descriptive statistical analysis.

- Create visualizations to understand data distributions and relationships.

1.4 Clustering

- Select and apply appropriate clustering algorithms (e.g., K-means).
- Train and evaluate the clustering model.

1.5 Visualization

- Develop visualizations to represent customer segments.
- Create interactive dashboards using Power BI.

1.6 Documentation

- Document the data analysis and clustering process.
- Prepare user guides and technical documentation.

2. Timeline

Task	Start Date	End Date	Milestones
Data Collection	15 July 2024	15 July 2024	Data collection
Data cleaning	15 July 2024	16 July 2024	Cleaned dataset ready
EDA	16 July 2024	16 July 2024	EDA insights generated
Clustering	17 July 2024	16 July 2024	Clustering model trained
Visualization	17 July 2024	16 July 2024	Visualizations and dashboard created
Documentation	16 July 2024	17 July 2024	Documentation completed
Project Completion		17 July 2024	Finaly Delivery

3. Resources

3.1 Human Resources

- Me: Responsible for data collection, cleaning, EDA, Min-Max Scaling, clustering, visualization, and documentation.

3.2 Technical Resources

Software

- **Python**: For data analysis, clustering, and visualization.
- **Matplotlib**: For creating visualizations.
- **Seaborn**: For statistical data visualization.
- **Scikit-learn**: For clustering algorithms.
- **Power BI**: For creating interactive dashboards.

Hardware

- Computer: Modern multi-core processor, minimum 8GB RAM (16GB or higher recommended), adequate storage.

3.3 Other Resources

- Training materials: For self-study and skill enhancement.
- Access to data sources: Ensure access to the mall customers dataset and any additional required data.

4. Risks

4.1 Data Quality Issues

- **Risk**: Incomplete or inaccurate data.
- **Mitigation**: Implement data validation and cleaning processes. Regularly update and verify the dataset.

4.2 Algorithm Performance

- **Risk**: Clustering algorithm may not perform well on the dataset.
- **Mitigation**: Evaluate multiple algorithms and select the best-performing one. Perform hyperparameter tuning.

4.3 Visualization Limitations

- **Risk**: Visualizations may not effectively convey insights.
- **Mitigation**: Use best practices for data visualization. Seek feedback from peers or mentors to ensure visualizations meet the intended goals.