Project Plan:

1. **Tasks:**

   1.1 Data Collection
   - Collect the mall customers dataset.
   - Verify the completeness and accuracy of the dataset.

   1.2 Data Cleaning
   - Handle missing values.
   - Remove duplicates.
   - Normalize data.

   1.3 Exploratory Data Analysis (EDA)
   - Perform descriptive statistics.
   - Create visualizations to understand data distributions and relationships.

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

   1.5 Visualization
   - Create visualizations to represent customer segments.
   - Develop interactive dashboards in Power BI.

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

2. **Timeline:**

   Project Timeline with Milestones

Task	Start Date	End Date	Milestones
Data Collection	13 July 2024	13 July 2024	Data collection completed
Data Cleaning	13 July 2024	14 July 2024	Cleaned dataset ready
EDA	14 July 2024	14 July 2024	EDA insights generated
Clustering	14 July 2024	14 July 2024	Clustering model trained
Visualization	15 July 2024	16 July 2024	Visualizations and dashboards created
Documentation	15 July 2024	16 July 2024	Documentation completed
Project Completion		16 July 2024	Final delivery

Resources:

1.1 Human Resources

- **Me**: Responsible for all tasks including data collection, data cleaning, EDA, clustering, visualization, and documentation.

1.2 Technical Resources

- **Software**:
  - **Python**: For data analysis, clustering, and visualization.
  - **Jupyter Notebook**: For interactive data analysis 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 (recommended 16GB or higher), adequate storage.

1.3 Other Resources

- **Training Materials**: For self-study and skill enhancement.

- **Access to Data Sources**: Ensure access to the mall customers data set and any additional required data.

3. **Risks:**

3.1 Data Quality Issues

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

3.2 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.

3.3 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.