

Industry Standard Documentation

Project Plan:

- **Tasks:**

Data Collection: Gather customer dataset given, loyalty programs, and other sources.

Data Cleaning: Cleanse and preprocess the collected data to ensure accuracy and consistency.

Exploratory Data Analysis (EDA): Perform EDA to understand the distribution, relationships, and basic statistics of the data.

Clustering Analysis: Apply clustering algorithms (e.g., K-means) to segment customers based on behavior patterns.

Visualization: Create visualizations (e.g., charts, graphs, heatmaps) to present insights from the data analysis.

Documentation: Document findings, methodologies, and insights for future reference and reporting. Data collection, data cleaning, EDA, clustering, visualization, documentation

- **Timeline:**

For this then timeline required with the mildstone

- **day1:**
 - Data Collection and Cleaning
- **day 2:**
 - Exploratory Data Analysis (EDA)
- **day 3:**
 - Clustering Analysis
- **day 4:**
 - Visualization and Documentation

- **Resources:**

Here are the two different resources we have to need i.e human resource and the technical resources.

Human Resources:

- Data Scientists
- Data Analyst
- Project Manager

Technical Resources:

- Data storage (Cloud storage or on-premise servers)
- Data analysis tools (Python, R)
- Visualization tools (Tableau, Power BI)
- Cluster computing (if needed for large-scale data processing)

- **Risks:**

- **Data Quality Issues:** Incomplete or inaccurate data from POS systems and loyalty programs.
- **Mitigation:** Implement data validation checks and cleansing routines during data cleaning phase.

Algorithm Performance: Poor clustering results due to inappropriate algorithm selection or parameter tuning.

- **Mitigation:** Conduct thorough testing and validation of clustering algorithms on sample datasets before full-scale implementation.

Visualization Limitations: Inability to effectively communicate insights due to limitations of chosen visualization tools.

- **Mitigation:** Evaluate multiple visualization tools and techniques to find the most suitable for presenting complex data.
- Data quality issues, algorithm performance, visualization limitations