**Project Charter**

**Project Title:** Customer Segmentation for a Retail Store

**Project Manager:** Nikhil Kumar Singh

**Start Date:** 13/07/2024

**End Date:** 17/07/2024

**Objectives:** To segment customers into distinct groups based on their purchasing behavior.

**Scope:** Data cleaning, EDA, customer segmentation using K-Means, visualization using Matplotlib and Power BI.

**Deliverables:** Insights, conclusions, and recommendations.

**Project Plan**

**Tasks:** Data collection, data cleaning, EDA, clustering, visualization, documentation

**Data Collection**

Gather data from sources such as transactions, customer surveys, stores. Ensure the data includes relevant features like purchase history, demographics, and customer preferences.

**Data Cleaning**

Process the raw data to handle missing values, correct inconsistencies, and remove duplicates. Standardize formats, normalize numerical data, and encode categorical variables to prepare the dataset for analysis.

**Exploratory Data Analysis (EDA)**

Analyze the cleaned data to understand its structure and underlying patterns. Use summary statistics, visualizations (like histograms, box plots, and scatter plots), and correlation matrices to identify trends, anomalies, and relationships between features.

**Clustering**

Apply clustering algorithms (such as K-means, or hierarchical clustering) to group customers based on similarities in their behaviors and characteristics.

**Visualization** Create visual representations through matplotlib and create dashboard using PowerBI of the clustered data to convey insights effectively. Use charts, graphs, and plots (e.g., bar charts, histogram,boxplots and scatter plots) to highlight customer segments, their behaviors, and key differentiators.

**Resources:**

* **Data Resources**
* **Software and tools**

**Data Cleaning Tools:**

* Python libraries (e.g., Pandas, NumPy)

**Exploratory Data Analysis Tools:**

* Data visualization libraries (e.g., Matplotlib, Seaborn)

**Clustering Tools:**

* Machine learning libraries (e.g., Scikit-learn)
* Clustering algorithms (e.g., K-means)

**Visualization Tools:**

* Visualization libraries (e.g., matplotlib,Power BI)

**Additional Resources:**

**Learning Materials:**

Online tutorials on data science and machine learning

**Risks:**

Data quality issues, algorithm performance, visualization limitations

**Data Quality Issues**

* Incomplete, inaccurate, or inconsistent data can skew segmentation results, leading to unreliable customer profiles.

**Algorithm Performance**

* **Incorrect algorithm selection or parameter tuning** may result in clusters that do not effectively capture customer similarities and differences.

**Visualization Limitations**

* **Difficulty in representing complex data** can obscure insights, potentially leading to misinterpretation of customer segments.