**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

| **Team Member’s Name, Email and Contribution:** |
| --- |
| **Suraj Kumar** ([surajkumar0892@gmail.com](mailto:surajkumar0892@gmail.com))   1. Data Understanding  * Data Wrangling * Data Analysis  1. Data Visualization: BarPlot, Line-Plot, Count-Plot 2. Feature Engineering: Data Preprocessing 3. RFM Segmentation and Analysis  * Calculating RFM Scores  1. Building Clustering Models  * K-Means clustering on RM data * K-Means clustering on FM data * K-Means clustering on RFM data * Comparing all the Models   **Shreya Ranjan** ([shreyasrivastav15@gmail.com](mailto:shreyasrivastav15@gmail.com))   1. Data Understanding  * Data Wrangling * Data Analysis  1. Data Visualization: BarPlot, Line-Plot, Count-Plot 2. Feature Engineering: Data Preprocessing 3. RFM Segmentation and Analysis  * Calculating RFM Scores  1. Building Clustering Models  * K-Means clustering on RM data * K-Means clustering on FM data * K-Means clustering on RFM data * Comparing all the Models |
| **GitHub Repo link:** |
| [**surajkumar089/Online\_Retail\_Customer\_Segmentation--Unsupervised-Machine-Learning-Capstone\_Project (github.com)**](https://github.com/surajkumar089/Online_Retail_Customer_Segmentation--Unsupervised-Machine-Learning-Capstone_Project)  [**Shreyaranjan16/Order\_Retail\_Customer\_Segmentation (github.com)**](https://github.com/Shreyaranjan16/Order_Retail_Customer_Segmentation) |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** |
| The goal of our research was to identify key client categories based on a transactional dataset and find out about or define them for online retail customers. By using customer segmentation, customers can be divided into groups according to their common characteristics by using customer segmentation. Marketing campaigns can be more effectively applied to specific audience subsets with the help of customer segmentation.  The dataset includes 54,1909 transactional data records with 8 characteristics. Prior to moving on, a significant number of duplicate records and missing values were removed. With the help of the data available, a few additional features that simplify and simplify data conversions while significantly enhancing model accuracy have been created.  In EDA, we learned a few things, including:   * which products are the most sold ones and the minimum sold. * The UK had the most number of consumers, while Saudi Arabia had the fewest. * The majority of customers made purchases on Thursday, while Friday saw the fewest transactions. * Most purchases were made between October and December during the holiday season, and between January and February, there were the fewest transactions. * The majority of purchases were made in the afternoon, with the evening seeing the fewest transactions.   We have decided to use the RFM approach for our transactional data, despite the fact that there are six different types of consumer segmentation models available. RFM is a technique that's usually utilized to identify customers based on the time since their most recent purchase, the total number of purchases they've made, and the total amount they've spent.  Organizations can determine how much of their revenue originates from new vs. repeat customers by using RFM analysis. Following the RFM study, K-Means clustering was used to identify various customer segments in our data.  We have used the silhouette score method and the elbow method to help us determine the best possible ‘k’ value, i.e., the optimal number of clusters.  We have additionally used K-Means clustering separately for Recency, Monetary and Frequency, Monetary along with Hierarchical clustering on RFM data to compare the clustering performances. |