**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| **Shubhashis Singh ([shubhashis4866@gmail.com](mailto:shubhashis4866@gmail.com))** 1.Getting insight from data 2. Data Wrangling 3. Preprocessing the dataset 4. Data Cleaning 5.Feature Engineering 6. Exploratory Data Analysis 7. Analyze of Numerical variables 8. Model Training 9.RFM Model Analysis 10.K Means Clustering 11.Recency and Monetary 12.DBSCAN Algorithm on Recency and Monetary 13.KMeans Clustering 14.Applying Elbow Method on Frequency and Monetary 15.DBSCAN Algorithm on Recency,Frequency and Monetary 16.Hierarchical Clustering 17.Summary And Conclusion |
| **Please paste the GitHub Repo link.** |
| Github Link:- https://shubh4866.github.io/Online\_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)** |
| Customer segmentation is a way to split customers into groups (customer segments) based on certain shared characteristics. All customers share the common need for your product or service; beyond that, they possess distinct demographic differences (i.e., age, gender) and tend to have additional sociology-economic, lifestyle, or other behavioral differences that offer companies useful marketing insights that they can use to deepen customer relationships.  Customer segmentation is important because it allows brands to understand their target audience more clearly, and create customization. Targeting an entire mass audience is not an effective marketing strategy, because buyers interact and purchase differently.  Our data contains all the transactions occurring between 01/12/2010 and 09/12/2011 for a UK-based and registered non-store online retail.The company mainly sells unique all-occasion gifts. Many customers of the company are wholesalers.  We have 8 columns and 541909 rows.we can see that our data have some null values so we drop them as it can impact our prediction and also we drop some values where InvoiceNo starts with C means it's a cancellation.we perform Feature Engineering for enhance predictive models as it involves isolating key information, highlighting patterns and bringing in someone with domain expertise and perform Exploratory Data Analysis to understand the data more clearly.  We perform below ML Models:-  1.RFM Model Analysis  2.K Means Clustering  3.Recency and Monetary  4.DBSCAN Algorithm on Recency and Monetary  5.Hierarchical Clustering Firstly we did clustering based on RFM analysis. We had 4 clusters/Segmentation of customers based on RFM score.  * Platinum customers=1263 ( less recency but high frequency and heavy spendings) * Gold customers=1324 (good recency,frequency and monetary) * Silver customers=981(high recency, low frequency and low spendings) * Bronze customers=770 (very high recency but very less frequency and spendings)   we can conclude that :-  \* Above clustering is done with recency,frequency and monetary data(Kmeans Clustering) as all 3 together will provide more information.  \* Cluster 0 has a high recency rate but very low frequency and monetary. Cluster 0 conatins 2414 customers.  \* Cluster 1 has low recency rate but they are frequent buyers and spends very high money than other customers as mean monetary value is very high.Thus generates more revenue to the retail business |