

# Capstone Project Submission

## Instructions:

- i) Please fill in all the required information.
- ii) Avoid grammatical errors.

### **Team Member's Name, Email and Contribution:**

Hritik Sharma

Email-[hritik.2.sharma@gmail.com](mailto:hritik.2.sharma@gmail.com)

Contribution-

1. Data cleaning
2. Data preprocessing
3. Feature engineering
4. Contributed to strategizing the exploration process
5. Creating the RFM model
6. Applying K-means clustering
7. Applying silhouette score method and elbow method

Sudhanshu Chouhan

Email – [Sudhanshurocks1998@gmail.com](mailto:Sudhanshurocks1998@gmail.com)

Contribution-

1. Data Wrangling
2. Strategizing Roles for the analysis
3. Contributed to strategizing the exploration process
4. Contributed to strategizing the model creation process
5. Completed the presentation of the final report

**Please paste the GitHub Repo link.**

Github Link:-

<https://github.com/Link/to/Repohttps://github.com/hritiksharma11/Capstone-Project-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)**

## **Problem Statement:**

Online Retail Customer Segmentation Online-Retail-Customer-Segmentation In this project, your task is to identify major customer segments on a transnational data set which 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.

## **Recency-Frequency-Monetary (RFM) model to determine customer value:**

The RFM model is quite useful model in retail customer segmentation where only the data of customer transaction is available. RFM stands for the three dimensions:

Recency – How recently did the customer purchase? Frequency – How often do they purchase? Monetary Value – How much do they spend? A combination of these three attributes can be defined to assign a quantitative value to customers. e.g. A customer who recently bought high value products and transacts regularly is a high value customer.

## **Segmentation with K-means clustering:**

Initially, the data is subject to important stages in an analytics pipeline: exploratory analysis, preprocessing, feature engineering and standardization. Then, the unsupervised classification technique, K-means clustering algorithm, is used to determine the ideal segments of customers. Silhouette analysis and related cluster visualizations are leveraged to deduce the optimum value of "K" (number of clusters) in the algorithm. The observations from the results are elaborately discussed before reaching the conclusion from the business perspective.

## **Conclusions**

- K-Means Clustering with Silhouette gives the highest score of 0.39 for number of clusters 2.
- Sales has been increased from 2010 to 2011.

- The business can focus on these different clusters and provide to customers of each sector in a different way, which would not only benefit the customer but also the business at large.



