**Capstone Project Submission**

|  |
| --- |
| Name, Email, and Contribution: |
| Name: Tilak R  email id: arunnayak221@gmail.com  Contribution:  ● Feature Engineering  ● NLP  ● Model Building  ● Data Wrangling  ● Handling Missing and duplicate values  ● Exploratory Data Analysis  ● Model Building |
| GitHub Repo link. [Tilak46-R/Customer-segmentation-using-unsupervised-learning (github.com)](https://github.com/Tilak46-R/Customer-segmentation-using-unsupervised-learning) |
| GitHub Link:- [Tilak46-R (TILAK.R) (github.com)](https://github.com/Tilak46-R) |
| **Please write a summary of your Capstone project and its components. Describe the problem statement, your approaches, and your conclusions. (200-400 words)** |
| **In this case study, our task was to identify major customer segments on a transactional 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.**  **STEPS INVOLVED IN THE PROJECT.**   * **Handling missing values and Duplicates in the dataset** * **Removal of Cancelled Orders** * **Feature Engineering** * **Exploratory Data Analysis** * **RECENCY,FREQUENCY,MONETARY MODEL(RFM MODEL)** * **SCALING OUR DATA** * **Applying Elbow method on Recency, Frequency and Monetary**   **CONCLUSION:**   * **The Five most sold products are WHITE HANGING HEART T-LIGHT HOLDER And REGENCY CAKESTAND 3 TIER.** * **The least sold products are: Green with metal bag charm and White with metal bag charm and so on..** * **We can see that majority of the customers are from United kingdom followed by small portions of Germany, France etc..** * **We can see that there are only 4338 customers present and are responsible for all these transactions let us now understand the percentage share of the top 10 customers** * **From the above figure we can infer that the top ten customers out of 4338 have contributed nearly 8 to 9% of total share. We can infer them as big buyers or wholesalers.** * **From the above distribution plots are heavily right skewed. it is very hard to find out the distribution hence we will apply log Transformation function to it.** * **We can infer from the chart that people have purchased more items on Thursday followed by Wednesday and Tuesday and people have purchased least on Fridays.** * **We can infer that people have purchased more on week days rather than holidays or weekends.** * **People have purchased more on November followed by October and December may be due to Festivals and people have purchased least during Feb this may be due to Winter season** * **We can see that people have made more purchases in Autumn season that is during the months of September, October and November.** * **We can infer that more people have purchased during the afternoon and least people have purchased during Evening** * **From the elbow method we reach the conclusion that the optimal number of clusters is 3 for Recency, Frequency and Monetary values.** * **Cluster2 represents your 'Champions' or 'Loyal' customers.** |