Capstone Project Submission

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| |  |  |  | | --- | --- | --- | | **NAMES** | **E-MAIL** | **CONTRIBUTION** | | Rahul Deshmukh | Deshmukhrahul1100@gmail.com | Entire Project work. | |
| **Please paste the GitHub Repo link.** |
|  |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** |

**Summary**

This study started with importing dataset, analysing dataset after this I have done pre-processing, I have checked for the null values as our dataset contains many null values in Customer id feature and we have to segment the customers, without customer id we are unable to segment customers therefore I have removed all the rows without Customer id.

After that I have done some exploratory data analysis (EDA) I came to know about top customers, Worst customers, periodical purchasing stats, most revenue generated weekday, purchase stats of country, top and lease purchasing country, top sold product, most revenue generated product, Customer stats, etc.

After that I have done some feature engineering to build RFM model (recency, frequency and monetary value) I have extracted and analyzed RFM score then I have created customer segment in 3 category bronze, silver and gold.

After that I have done data pre-processing for clustering with the help of log transformation, I have reduced Skewness of data then I have scaled data, after scaling I have extracted Silhouette Score Based on the inertia and silhouette score, I came to know that optimal number of clusters is 3.

Then I have implemented K-means clustering and plotted different graphs to visualize clusters. after that I have merger cluster column to data and used classification model for prediction. I have used Logistic Regression, Random Forest Classifier and XGBoost and done evaluation of it, my Optimal model was Random Forest as I got train Accuracy of 1.00 and test accuracy of 0.98 with it.