Capstone Project Submission

Team Member's Name, Email and Contribution:
Contributor's Role: **Kiran Ahire:** **Email id : kiranahire73@gmail.com** 1. Descriptive Analysis 2. Data Wrangling 3. Correlation between data and outlier detection 4. EDA – Analysis of Categorical & Numerical Data 5. Feature Engineering 6. RFM model 7. K means clustering 8. Silhoutte and Elbow method
Please paste the GitHub Repo link.
Github Link:- https://github.com/kiranahire03/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 the practice of dividing customers into groups that reflects similarity among customers in each group. The goal of segmenting customers is to decide how to relate to customers in each segment in order to maximize the value of each customer is to the business. The contents of the dataset had features such as invoiceno , stockcode , description, quantity , unitprice , customerid and country. The problem statement was to build an unsupervised machine learning algorithm to perform customer segmentation. We started with data wrangling in which we tried to handle null values, duplicates and performed feature modifications.

Next, we did some exploratory data analysis and tried to draw observations from the features we had in the dataset. Next, we formulated some quantitative factors such as recency, frequency and monetary known as RFM model for each of the customers. We implemented Kmeans clustering algorithm on these features. We also performed silhouette and elbow method analysis to determine the optimal no. of clusters which was 2. We saw customers having high recency and low frequency and monetary values were part of one cluster and customers having low recency and high frequency, monetary values were part of another cluster**.** We also implemented shap techniques to understand what is going on inside our model. We saw higher values frequency , Monetary and high values of Recency deciding other class. However , there can be more modification on this analysis. One may Choose to cluster into more cluster depending on company objectives and preferences.

The labelled feature after clustering can be fed into classification supervised machine learning algorithms that could predict the classes for new set of observations. The clustering can also be performed on new set of features such as type of products each customer prefer to buy often, finding out customer lifetime value (clv), segmenting on the basis of time period they visit and much more.