Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:
The team members *SIBANI CHOUDHURY and Sangamesh Chandankera* ***Email:-*** ***choudhurysibani120@gmail.com*** ***sangameshchandan@gmail.com*** ***Contributor roles :-*** ***1. Sibani Choudhury*** ***1. Preview data*** ***2. Check total number of rows and column types.*** ***3. Check the null values and duplicate values.*** ***4. Exploratory Data Analysis.*** ***5. Analysing the data by column wise*** ***6. customer segmentation with RFM analysis*** ***7. Using Elbow and Silloute method*** ***8. Unsupervised machine learning*** ***K means clustering*** ***Hierachical clustering*** ***9. conclusion*** ***2. Sangamesh Chandankera*** ***1. Loading Data.*** ***2. Check total number of rows and column types.*** ***3. Check the null values and duplicate values.*** ***4. Exploratory Data Analysis.*** ***5. Data Visualization: Analyzing the Relationship by datetime variable.*** ***6. Customer RFM (Recency, Frequency, Monetary) analysis.*** ***7. Building and evaluating the model*** ***K Means clustering*** ***Principal component analysis*** ***8. Conclusion.***
Please paste the GitHub Repo link.
https://github.com/choudhurysibani120/customer-segmentation/blob/main/Online_Retail_Customer_Segmentation_sibani_choudhury_Capstone_Project.ipynb
Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

In this project, our task was 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.
There are 541909 rows and 8 columns in our dataset. There were more than 120000 null values present in CustomerID Column it main column as other column was filled with zero and drop all values. Various plots are visualized to see Outliers and Applied Inter Quartile Range method. Data was used different units so its scaled using Standard Scaler and normalise data.
Customer Segmentation is the process of dividing customers into groups based on common characteristics so companies can market to each group effectively and appropriately. Customers are segmented according to their similarities in behavior and habits.To find Number Clusters we applied Elbow Method and silhouette score the Selected Cluster Size with Visualized Graph. K-Means Clustering was applied. Along that RFM analysis was made. Dendrogram Linkage and Hierarchical Agglomerative Clustering Models are also applied. Given Data for Customer Segmentation most of them are irrelevant like StockCode, Description.etc and there is no relation. After Applying Elbow and Silhouette score are more at cluster size =3 or 2 Same results applied with Dendrogram results of Kmeans Clusters Centers in plots appears better than Hierarchical Agglomerative Clustering.Cluster_0 CustomerID's take more time gap between Each Oder they Placed (Rarely). Cluster_1 CustomerID's Makes always a Bulk Purchases which leads high Spending's (Retailers).Cluster_2 CustomerID's Has Highest Orders Placed (Small Shops with less inventory).