Capstone Project Submission

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

Team Member's Name, Email, and Contribution:

Team Member's Role:-

- **Sanjaya Kumar Khadanga**
  Email- skhadanga38@gmail.com
    - Data Understanding
    - Feature Analysis
    - Feature Engineering
    - Exploratory Data Analysis
    - RFM model
    - K-means

- **Bibhuti Bhusan Sahu**
  Email- sahubibhuti45@gmail.com
    - Data Understanding
    - Feature Analysis
    - Data Visualization
    - Multivariate Analysis
    - Silhouette
    - DBSCAN
    - Research Analytics
        - Technical documentation

- **Balaram panigrahy**
  Email- balarampanigrahy42@gmail.com
    - Data Understanding
    - Feature Analysis
    - Data Visualization
    - Multivariate Analysis
    - Elbow method
    - Research Analytics
        - Technical documentation

Please paste the GitHub Repo link.
Github Link:- https://github.com/Bibhuti-MLAI/Customer-Segmentation

Please write a summary of your Capstone project and its components. Describe the problem statement, your approaches, and your conclusions. (200-400 words)
The problem statement is about segmentanting the customers according to the behaviour of frequency of visit, products they are buying. The dataset has the data about a retail store in the United Kingdom. After analysing the dataset we have started our approach with each variable independently. We started with business understanding and then we moved into feature engineering and then we spent most of the time in Exploratory Data Analysis (EDA). In EDA we have gone through different plots which helps us to understand more about the data and we proceed for model building. EDA also helps us about the business decisions i.e. skewness and kurtosis, after which we also go for transformations for the highly skewed data and then we finalise the data for model building. In the model building part we have applied RFM model, Kmeans clustering and Hierarchical clustering to segment the customers. We have calculated the Silhouette score for each model and also the Elbow curve to get the optimum number of customers. For Hierarchical clustering we have drawn Dendrogram to get the optimum number of clusters. After all this we came to the conclusion that we have two segments of the customers based on their behaviour of visiting and the products they are buying.