Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:

1) Arbaaz Malik : **Email:** malikarbaaz267@gmail.com

- Data Cleaning
  a) Dropping duplicates
  b) Handling null and missing values
  c) Handling Outliers

- EDA (Univariate and Bivariate Analysis)
  a) Quantity
  b) Unitprice
  c) Top 10 repeatedly sold items
  Top 10 countries who's buying maximum products
  d) e) Number of shopping monthly, per day, hourly and whether it is morning, afternoon or evening.

- Correlation analysis between all the variables.
- Data Transformation (we have applied Log Transformation which basically shrink the data from large to small
  a) Log Transform on Quantity
  b) Log Transform on Unitprice
  c) Log Transform on TotalAmount

- Machine Learning Clustering algorithms :
  a) K-Means with Silhouette score
  b) K-Means with Elbow method
  c) Hierarchical clustering

2) Huzaifa Khan : **Email:** huzaifakhan2974@gmail.com

- Data Cleaning
  a) Dropping duplicates
  b) Handling Null/nan values
  c) Handling Outliers

- Univariate and Bivariate Analysis
  a) Top 10 repeatedly sold items
  b) Top 10 countries buying maximum products
  c) Number of shopping monthly, per day, hourly and whether it is morning, afternoon or evening.

- Data Transformation
  a) Log Transform on Quantity
  b) Log Transform on Unitprice
  c) Log Transform on TotalAmount

- Correlation analysis between all the variables
- ML Clustering algorithms used:
  a) K-Means Clustering
  b) Hierarchical clustering

Please paste the GitHub Repo link.

Github Link:- https://github.com/Malikarbaaz/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)

Business all over the world are growing everyday. With the help of technology, they have access to a wider market and hence, a large customer base.

Customer segmentation refers to categorizing into different groups with similar characteristics. Customer segmentation can help business focus on each customer group in a different way, in order to maximize benefits for customers as well as the business.

This project mainly deals in segmenting customers of an online business storfe in the UK.

We have done in this project:

1. **Data Cleaning**
2. **EDA**
3. **Data Transformation**
4. **ML Clustering algorithm**
5. **Clustering Profiling**

<u>Conclusion:</u>

- **Throughout the analysis we went through various steps 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.**
- **we saw how we can segment our customer depending on our business requirements. We perform RFM for our entire customer base**

- **RFM analysis can help in answering many questions with respect to their customers and this can help companies to make marketing strategies for their customers, retaining their slipping customers and providing recommendations to their customer based on their interest.**
- **Using cluster profiling the average of recency, frequency and monetary values for each customer segment was identified.**
- **We used the K-means algorithm to segment our customer in various clusters having similar similarity. K-means did a pretty good job here, Also we remember that the more the number of cluster we take the better the result we get (seperation of multiple cluster).**