Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:

Name : GOPAL JI GUPTA **Email: gopaljigupta45@gmail.com**

Role :-

1) Data Cleaning :-
   - Dealing with null values, duplicate data and outliers present in our data.
2) Exploratory Data Analysis :-
   - Performing univariate and bivariate analysis of our variables and plot them to visualize the relationships between them.
   - Checking and visualizing the correlation between our dependent and independent variables.
   - Trying to discover trends, patterns in our data with the help of statistical summary and graphical representations.
3) Data Preprocessing & Feature Engineering :-
   - Creating several new variables like recency, frequency and monetary value from given data to better achieve our objective of customer segmentation.
   - Performing RFM score analysis.
   - Log transforming our data.
   - Scaling our data prior to feeding it to ML models.
4) Model Implementation :-
   - Fitting various models on our data to segment our customers into several clusters.
   - The Models implemented are :-
     1. K Means clustering with Silhouette score analysis.
     2. K Means clustering with Elbow method
     3. Agglomerative Hierarchical Clustering.
     4. DBSCAN
5) Data Visualization :-
   - Using several kinds of charts like pie chart, joint plot, heatmap, pair plot, barplot, boxplot etc to better visualize data and understand correlation and trends.
6) Conclusion :-
   - Drawing some insights from the data and the results of our various clustering models on our input data.

Please paste the GitHub Repo link.

Github Link:- https://github.com/gopaljigupta45/Online_Retail_Customer_Segmentation

Drive folder link :-

https://drive.google.com/drive/folders/18dO8ge9WPz38AbUJ9VmFUGchIq0IN7Q3?usp=sharing

Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

With advancements in technology, the world around us has been transformed and continues to evolve still at a rather fast pace. This is especially true for Businesses because technological advancements have made it possible to know a lot about one's customers and understand their behavior and interests and strategize around it. This is where customer segmentation comes into play.

Customer segmentation is the process of separating customers into groups on the basis of their shared behavior or other attributes. The groups should be homogeneous within them and should also be heterogeneous to each other. The overall aim of this process is to identify high-value customer base i.e. customers that have the highest growth potential or are the most profitable and most frequent buyers. It can also help us identify customers who we are at the risk of losing.

RFM analysis:-

We have used RFM (Recency, Frequency & Monetary) Model to segments customers into different groups. **Recency** signifies the how recently order was placed, **Frequency** signifies the number of times a customer ordered and **Monetary** signifies the total amount spent on the orders. Our objective is to rank our customers into segments indicating best to worst based on these factors. That way, we can formulate better relationships with our best customers to keep them happily engaged with our business, we can make efforts to retain the customers that we are at risk of losing and use several other strategies for different segments in order to maximize our customer base and sales revenue.

In this exercise, our goal is to do customer segmentation on a dataset containing all the transactions between 01/12/2010 and 09/12/2011 for a UK based and registered non-store online retail company.

Conclusions Drawn :-

- We saw that most of the customers belong to the United Kingdom which makes sense as it is the country where our business is based out of.
- We saw that there are around 8872 instances where an order was canceled.
- We saw that most cancelled products according to our data are REGENCY CAKESTAND 3 TIER, Manual, POSTAGE, JAM MAKING SET WITH JARS etc.
- We also saw that most number of cancellations were made by residents of the United Kingdom which makes sense as UK residents have made the most orders as well.
- We saw that the winter months have the most sales with November, October and December having higher number of orders placed compared to the rest of the year.
- We also saw that most orders were placed on Thursdays and no orders were placed on Saturdays.
- We saw how we can segment our customers depending on our business requirements. We performed Recency, Frequency and Monetary value analysis for our entire customer base and used it to rank our customers.
- RFM analysis can help in answering many questions with respect to our customers and this can help companies to make marketing strategies for their customers, retaining their at risk of leaving customers and providing recommendations to their customers based on their interest.
- We have used Agglomerative Hierarchical Clustering, elbow method and silhouette score to find optimal no. of clusters.
- **The optimal no. of clusters in Agglomerative Hierarchical clustering is 2.**
- **The optimal no. of clusters in K-means with silhouette score and Elbow method is 2.**
- **The optimal no. of clusters in Agglomerative Clustering with threshold value 40 is 3.**
- **The optimal no. of clusters in DBSCAN clustering is 3.**

……..