**EXPLAINATION**

Customer is the king I'm sure you might have heard this same in today's world. Customer is indeed the king and hence it's not only very important to retain the loyal customers but also to win back the ones who are on the verge of churning out. So, customer segmentation is one method if we want to know who are our most loyal customers and who are the ones who visit the platform or our e-commerce website infrequently that is they just visit the website for the sake of you know redeeming some coupons or just getting some discounts or we can say that they are only the one-time visitors. performing customer segmentation using machine learning techniques to create several clusters of customers according to their spending behaviour we would be able to create clusters of most loyal customers as well as the customers who are on the verge of churning. we will take a look at data set we are going to use as well as various data cleaning steps.

**what is customer segmentation?**

It's a methodology using which we can divide our customer base into group of individuals who are similar in terms of either gender spending behaviour frequency to visit age or other demographics customer segmentation allows companies to precisely target the customers who has a specific needs and desires. This way companies can design the targeted campaigns to the right group or audience. It may also be the key that during the creation of these individual groups or clusters company might identify new market segments on which company can focus more as it might be more lucrative, also using segmentation companies can identify groups that require extreme attention such that people in that group are on the verge of churning out there may be a group of customers or segments who has a highest potential value it can also help with creating targeted strategies that can capture your customers attention and create positive high-value experiences with your brand's so in a nutshell you can perform segmentation based on your business requirements. Here to perform customer segmentation we will first utilize RFM modelling to calculate the RFM scores for each customer and then we will apply unsupervised machine learning technique called k-means to group those customers into different groups based on RFM scores, RFM stands for

Recency, frequency and monetary.

**Recency** means how recently a customer purchased an item or product the more the recent purchases the better the score would be. **Frequency** represent show often a customer purchases an item or product the more frequent they purchase the better score would be for frequency and lastly **Monetary** represent show much the customer spends the more the mode the better monetary score would be now if you are wondering what are the different groups or segments which can be created based on our FM scoring and k-means clustering then let me take an example so according to business requirement a company might create three groups like high medium and low. A group or segment which is marked as high is the one which has customers with highest scores of our F and M the customer in this group tends to be more loyal to the company and spends a lot as compared to other groups they often visit to your website or platform. This is the group you don't want to lose. A medium group is one which visits the e-commerce website or platform but not as frequent as customers in high group they spend comparatively less money than the customers in high-growth and they might not have purchased product very recently. customers in this group has medium RFM scores then we have low group in which we have low scores of RFM values customers in this particular group are tend to churn out and there are those customers who has not bought from the platform ecommerce website from quite long in fact they've visited the website only once maybe to redeem some coupons or get some discount. Let calculate the RFM for each individual customer. First, I imported the necessary libraries like NumPy, pandas, Matplotlib in line to draw any matplotlib related graphs in Jupiter notebook. In the next cell I have imported data using read underscore csv method of pandas. In the next line I am validating the number of Records using the shape function so I am just validating the number of records and columns. Next, I have done some feature Engineering and Exploratory data analysis.

To calculate the recency frequency and monetary scores for each customer followed by creating the segments or clusters to group the customer with similar needs of behaviour, Recency purchases as well as how frequently they are buying, most loyal customers as well as the customers who are on the verge of churning out that is grouping them into various loyalty levels like Champions, Potential Customers, Need Attention. where champion group represents most loyal customers and Potential customers group represents the group who has not purchased anything from quite long. calculating the recency, frequency and monetary values for each customer from the data set I first import the date/time library and give it an alias name as DT in the next line I am setting up the latest date in the day time format to calculate Recency. Create RFM Modelling scores for each customer. The less the recency number the better it would be for our FM scoring. Lastly, we are just taking a look at subset of data using head function. Moving on in the next line we are performing some descriptive statistics on recency column and in order to do that we are making use of method called describe which gives that is a descriptive statistics like mean, standard deviation, maximum value, minimum value, median value etc. In the next cell I just wanted to see the recency distribution of data so I used a seaborne library to plot the histogram on recency column and I perform the same task for frequency and monetary column and seethe similar kind of observations for frequency and monetary values as well or monetary columns as well in the next cell we are creating the quartiles like 0.25,0.50,0.75so that we can subdivide the entire data set into four groups based on recency frequency and monetary values. In the next line we are just converting these quantiles into dictionary and you can see the data subdivision in the next cell here okay. Next we are creating two functions our scoring and F scoring here too in order to create segments which will be directed by values one two three and four one important thing to note here is that in our scoring function I am assigning value one to the lower value of recency because lower the value the better it is and which also means that customer is more engaged with a specific brand on the other hand F and M scoring function I am assigning value one to the highest value of frequency and monetary because higher the value off or frequency and monetary the better it is .

In the next cell I am assigning the F and M segment values & for each customer for recency column I am using our scoring function and providing that function as an argument to apply method along with our values as recency and quantiles similarly for frequency and monetary columns we are using F and M scoring functions and providing that function as an argument to apply method along with our values as frequency and quantize as well as monetary and quantile respectively. I wanted to see the output after adding these new columns called RF and M so I used the head method to print those first few records and you can see that new columns are F R M got added with individual segment values for each customer. So, basically our FM score gives the score to a customer's loyalty or engagement the lower the value of this our FM score is the lower and the customer would be as well as more engaged he would be with the brand. Based on our FM scores we are assigning the loyalty levels to each customer by creating three level Champion, Potential customer and Need Attention. So, if a person is in

Champion group then it means that it's our most valuable customer and is in fact a loyal customer as well and customer in Potential Customer category is the one who hasn't purchased from the brand from quite long and Need attention group is on the verge of churning out. Next we are calculating the score cards based on panda's q cut method here first argument it depicts the name of the data frame column which in our case is our FM score and second argument is Q cut comp which is kind of representing bins containing the same number of Records for each bin okay and the third argument is labels which in our case is loyalty level which we defined above so in the next line we are creating new column called RFM loyalty level or RFM underscore loyalty level containing these loyalty levels. Now we can see three loyalty levels in terms of Champion, Potential customer and Need attention. we can see that customers with highest monetary value are at the top.

**SUMMARY**

**so, based on this RFI modelling what should be the marketing strategy should be now?**

well customers with our FM 1 group of are the best customers and we can try to cross sell other products of our brand as well as we can encourage them to sign up for loyalty programs to enjoy some elite experiences like free same-day shipping priority access to newly launched products etc on the other hand if the customer is falling into our FM 2 group then company may try to offer some reward or coupon to trigger the spending from these group and if the customer is falling into our FM 3 group then company should give them discount so that they come and buy products.