Capstone Project Submission

Name:- Ajit Kumar Toppo
Email: ajitkumartoppo96@gmail.com

Github Link:- customer_segmentation

Drive link:- https://drive.google.com/drive/folders/1ImJgySX8Wvu3QlvKfFEQwPaNUgmfHP5V?usp=share_link

Online Retail Customer Segmentation

Context:

Customer Segments (or Market Segmentation) allow the companies to be able to utilize their resources (time, finance) to serve their goals: increasing sales, increasing profits, retaining important customers as well as implementing marketing campaigns more effectively.

Problem Statement:

For this project we have transaction records of online purchases done by customers from a UK based online retail. Our objective will be to segment the customers based on RFM analysis and K-means clustering. 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.

Approach:

First I began with the understanding of our data present, imported the required libraries. Found out the missing values present and treated accordingly. Did analysis to find how many unique products,customers are there? From which place do most of the customers belong to? Performed RFM analysis on the data and categorized it into four quartiles so that we can find our best customers, loyal customers, big spenders, almost lost, lost customers, lost cheap customers based on the RFM score. Checked the distribution of data based on recency, frequency, and monetary and found that the data has skewed distribution so applied log transformation to transform the skewed data to approximately conform to normality. Then used the elbow method and silhouette score to find out the optimal number of clusters and applied the K-means clustering algorithm to obtain customer segments.

Conclusion:

From all the graphs we see the k-means has created two clusters for CustomerID 0 and 1. Customers with relatively high frequency of buying, high monetary value who purchased recently are grouped in cluster 0. Customers with relatively low frequency of buying, low monetary value some of them purchased recently while for some its relatively long time are placed in cluster 1. We observed clusters from RFM analysis also by which can be used by business to understand customer behavior, plan business strategies, marketing campaigns, etc. to target, incentivise and attract customer base.