Project Summary
Online Retail Customer Prediction

Customer segmentation is one of the key aspects of business decision support system. In order to grow the business intelligently in competitive market, identification of potential customer should be done timely. This proposes an integrated approach for determining target customers using predictive model and also discover their associative buying patterns using K-Means algorithm and other methods. After identification of targeted customers and their associative buying pattern, the business managers take the strategic profitable decisions accordingly.

Online retail business is growing rapidly now-a-days, customers and wholesalers usually buy products of their interest because it's very easy to select and view more catalogues. It is very Important for companies to provide the products that their customers need and when they need and how much they have the capability to purchase and to deal with these problems most of the retail companies use customer segmentation technique

First step was to import the dataset through pandas & then did data wrangling and feature engineering. Large NULL values were present in CustomerID and Description column. As customer IDs are uniquely assigned to customers so we cannot use it with other values. Therefore, we dropped the NULL values.

EDA (exploratory data analysis) was carried out, in which visualization of different features has taken into account with bar plot, distplot, heatmap and boxplot. With the help of barplot, got the insights of top customers, top countries, top months of sales, top days of sale and top hours.

Then carried out data transformation using RFM (Recency, Frequency and Monetary) table to get more deep insights of particular customers. We assigned labels on each customer ID according to their purchasing power. Some outliers were present in the RFM table, to remove the outliers IQR (inter quantile range) was applied and the data got transformed into standard scale.

For Cluster modeling, implemented the K-Means Clustering Algorithm and later for Optimal Clusters Modeling implemented the K-Means with Silhouette Analysis and the K-Means with Elbow Method.

Then got the following results:

- K-Means = Optimal Clusters (3)
- K-Means with Silhouette Analysis= Optimal Clusters (2)
- K-Means with Elbow Method = Optimal Clusters (4)

Contributor Roles:

1 **Sharath Diwakar** (*mailmesharathd@gmail.com*)

1) Data Cleaning
2) Data Wrangling
3) Feature Engineering:
   - Introducing new variables with Date Time
   - Introducing total amount with Quantity and Unit price
4) EDA (Exploratory Data Analysis):
   - Bar plot, Box plot, Dist plot
   - Correlation Map
5) Data Transformation:
   - RFM Analysis
6) Modeling:
   - K-Means Clustering
   - K-Means with Silhouette Analysis
   - K-Means with Elbow Method
7) Technical Documentation
8) Project Summary
9) Project PPT

GitHub Link: *https://github.com/Sharath2021/Online-Retail-Customer-Segmentation*