Ashfaq Pathan: https://www.linkedin.com/in/ashfaq-pathan-a0b746128/

Exploratory Analysis

We have done EDA using Excel,Tablue,Python, you can fine all eda in eda folder give with this document.

Data Cleaning and Merging

Missing values: removed null value column where Null ration is more than 40%, we have dropped that columns  
After initial analysis of missing values, we had decided to drop below features having missing values more than 40%

Invoice Data Cleaning: - Dropped Columns: - with are not so impactful in data modelling

Customer Data Cleaning: - Dropped columns: - Title, Marital Status, Occupation, Date of birth, Death date

Jed Data Cleaning: - Dropped columns: with are not so impactful in data modelling

Data Merging: Shape of data frame (936200, 34)

Feature Engineering

Total: Invoice Date Time, Job Card Date Time using these two columns we have created a Column which contained total hr customer spend on car service

Clusters: We made two impact full clusters :( revenue with total hr spend) (Revenue with km\_ readings)

Recency:  How recently did the customer purchase?

Frequency: How often do they purchase?

Monetary Value: How much do they spend?

Feature Cleaning

City: Using pin code used to fined correct city

District: Using pin code used to fined correct state

Customer Segmentation

After all cleaning & data pre-processing tasks, A new data frame was created with the help of cleaned merged data frame (Clean.csv).

Approach

remove customers with revenue 0

Rule Based Clustering :3 clusters were identified based on total\_hr analysis on Average revenue. Low, Medium, High

K-means Clustering done a created 2 features

Customer Lifetime Value Calculation : total\_revenu / reccency

This calculation is based on a blog we find on website

<https://data-science-blog.com/blog/2020/03/19/how-important-is-customer-lifetime-value/>

Model dataset containing features

Created a Separate Data Frame of year2013

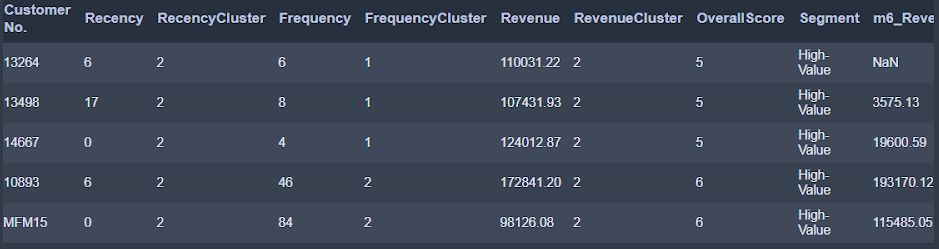
We created two different model

1 based on RFM

2 based on clusters with we made

Below we can see final dataset

RMF dataset\_0.85\_accru\_score



Cluster\_dataset\_0.94\_accru\_score

