Capstone Presentation

Mahindra First Choice Services (MFCS) Project

Prepared by : HARSH SHAH (DSMR 16th March batch)

Problem Statement

- MFCS aims to expand country wide network of their workshops and improve their overall business
- Leverage the data and address the issues they have.

Data Science objectives:

i.   **Ownership pattern of cars by analysis**:

ii.   **Type of order each state receives:**

iii.   **Customer Lifetime value prediction:**

Why solve this problem?

- To help MFCS better understand the spending patterns of customers

- Handle seasonal demands across nation in a better way

- To Conduct **targeted** marketing campaigns and sales promotions.

- Improve business by Customer Behavior Analytics

Data

Current data points in the dataset:

Invoice_datetime | Jobcard_datetime | Order_type| Plant_Name | Location | Total_Amt | Make Customer_type | Part_Name

Additional Datapoints that boost better data analysis:

Servicing Time can help to analyse the total time required to complete the service comparatively the associated revenue generated by it.

Age and Gender of Customer can really help in identifying customer segments.
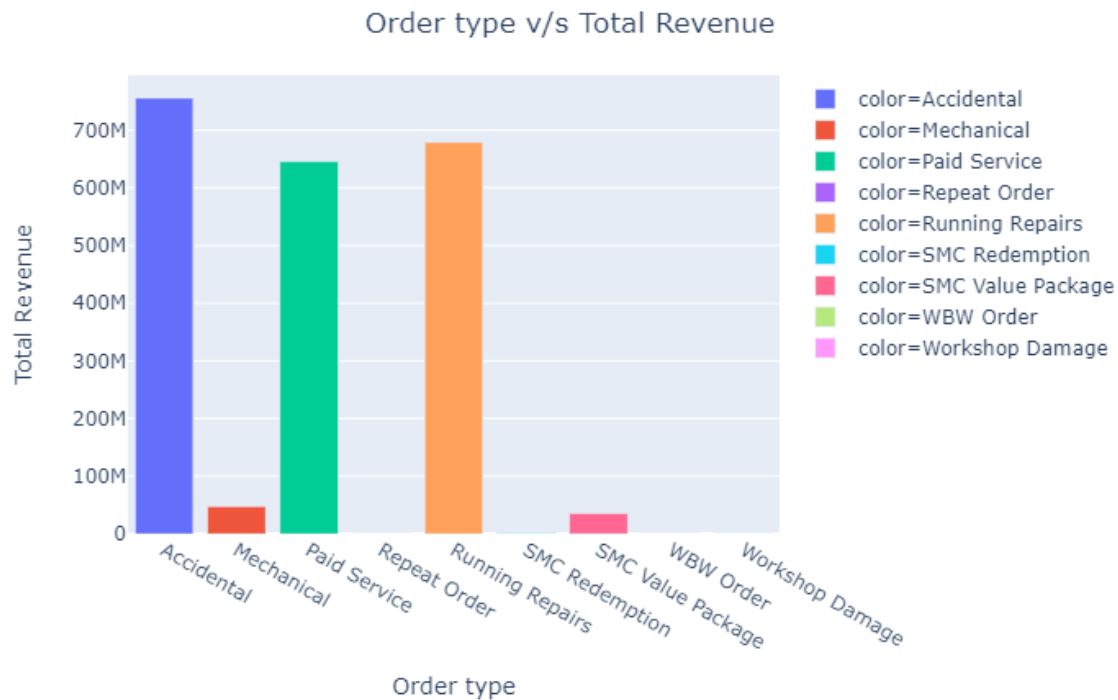
Technician related data points to identify his as well as particular garage's efficiency collectively.

Exploratory Data Analysis
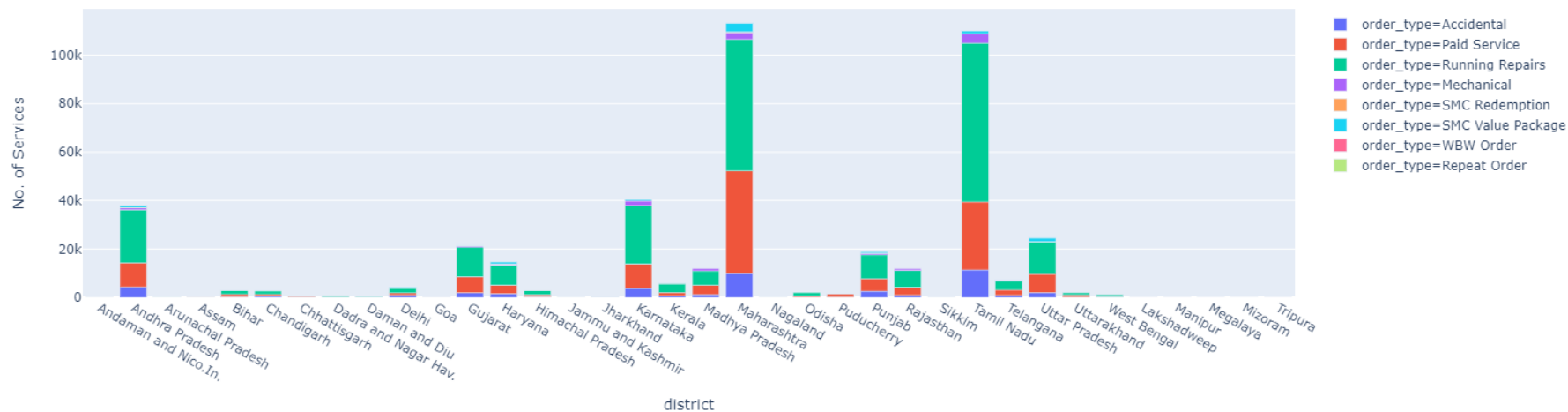→ Order each state receives
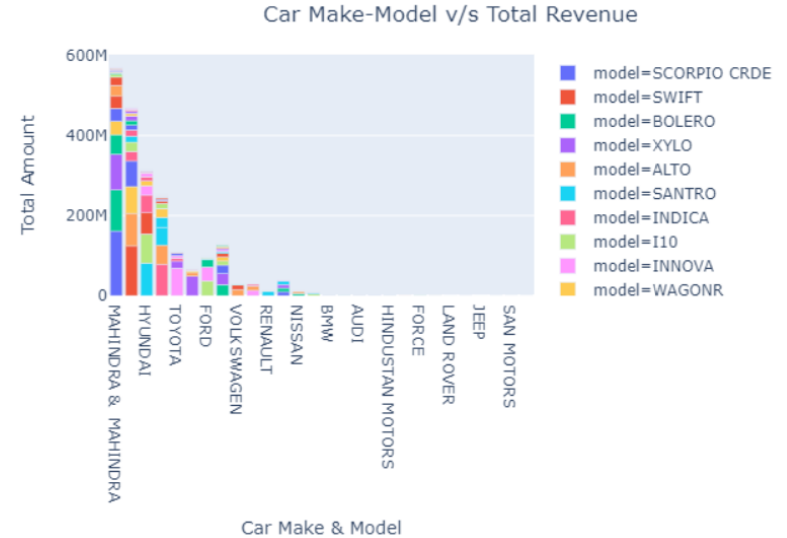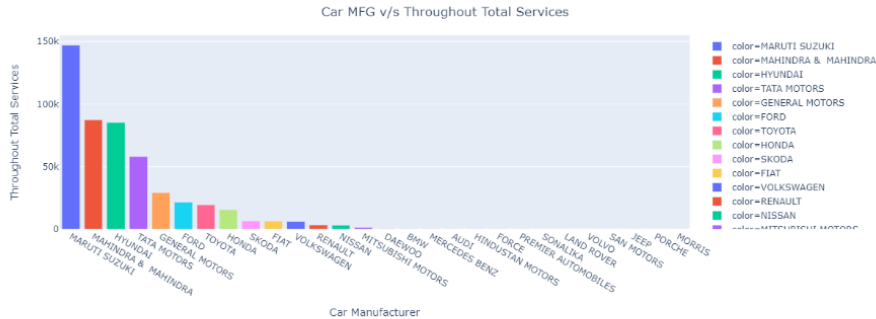
Order each state receives - EDA

Order type v/s Total Revenue
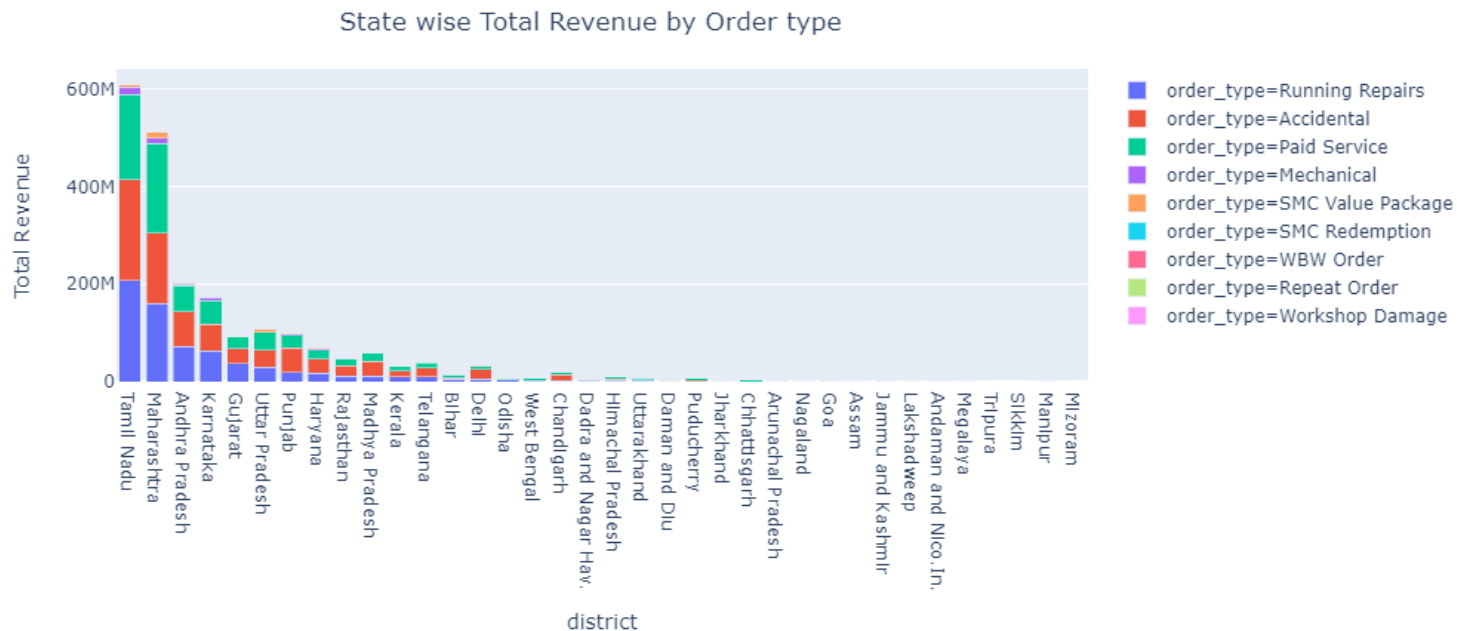
Order each state receives - EDA

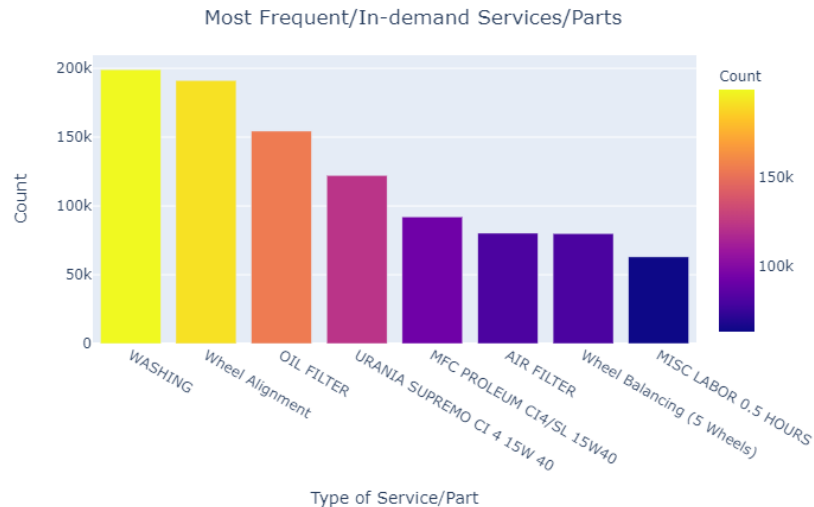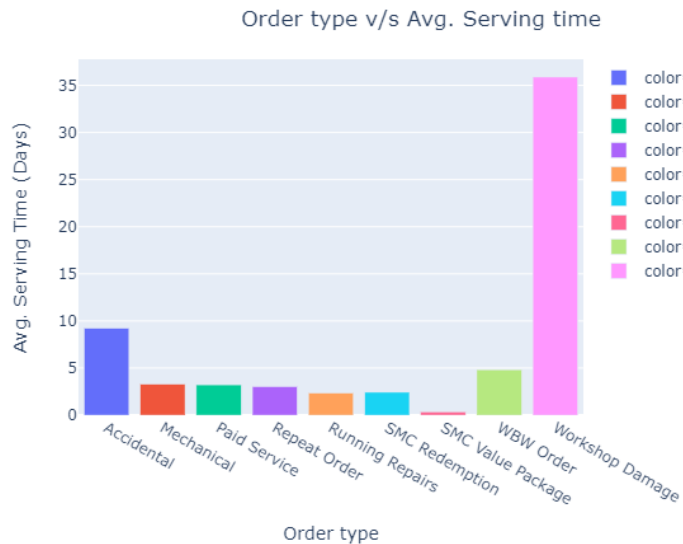

State wise Total Services

Exploratory Data Analysis

→Spending patterns of Cars

Exploratory Data Analysis

→ In demand parts and average service time by order type

Pipeline

DATA CLEANING

- Cleaned 78% of Pincode and City data points of Plant - Using external data(official government approved data by INDIA)
- There were more than 90% missing values for **Customer** (DOB, Marital Status, Occupation); Insurance (Policy No, Insurance company); so removed them.
- For few other data points, missing values for categorical variables were imputed as 'No Info' category.

Models and Approaches

→ For Customer Segmentation,
● Used K-Means Clustering..
● Used RFM Segmentation technique

● I decided to segment customers in 3 groups namely,
1. HIGHLY VALUED
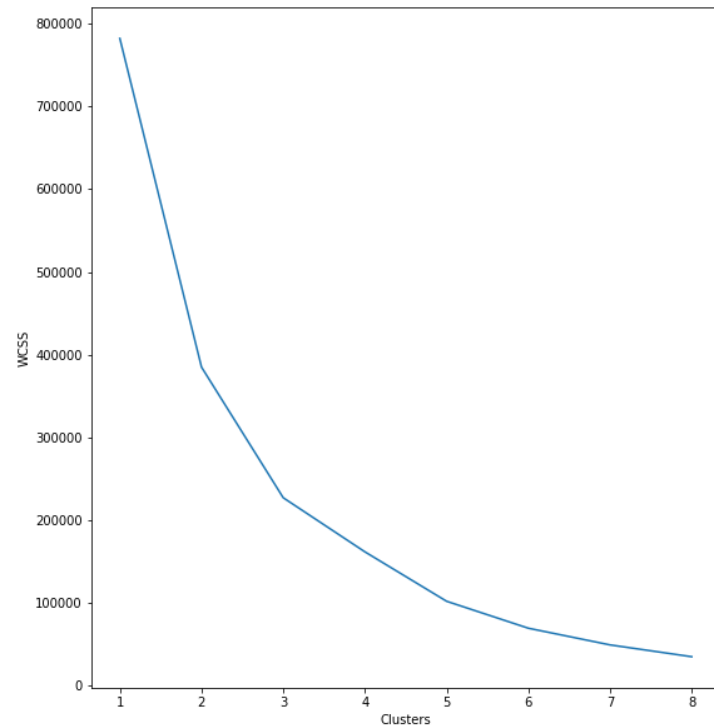2. MID VALUED and
3. LOW VALUED

Model Tuning

FEATURE ENGINEERING
- From jobcard_datetime and invoice_datetime, made new column **'service_time_days'** but major rows were seemed dirty data.

HYPERPARAMETER TUNING
- Hyperparameter : cluster = 6 (This value was better in results than others) - WCSS figure

WCSS

Using elbow method , it can be observed that k=6 seems to be ideal scenario.

Final Results

```
Low-Value      211542
Mid-Value       41784
High-Value        158
Name: Segment, dtype: int64
```

	customer_no	yearly_frequency	total_visits	monetary_value	recency	OverallScore	Segment	LTV	RecencyCluster	FrequencyCluster	RevenueCluster
0	E4981	533.600000	2668	2.371946e+06	48	15	High-Value	1.265670e+09	5	5	5
1	164806	23.000000	23	2.030874e+06	0	12	High-Value	4.671010e+07	5	2	5
2	174547	38.000000	38	6.006147e+04	72	9	Mid-Value	2.282336e+06	5	2	2
3	MFM15	31.400000	157	8.761328e+04	118	9	Mid-Value	2.751057e+06	5	2	2
4	189806	25.000000	25	5.025483e+04	61	9	Mid-Value	1.256371e+06	5	2	2

Insights & Decisions

Top Insights
- Karnataka has 100+ garages just like Maharashtra and Tamil Nadu but orders/garage ratio is almost half as compared to other two, so it can be improved.
- Running Repairs and Paid Services (Accidental is unpredictable) are top order types and hence can given more attention to generate major revenues.
- There seemed to be only 12 high values customers as per the modelling outcome, so better strategies can increase the count and also gender and age of customer data can help better customer segmentation apart from spending patterns.

THANK YOU