

Capstone Problem Statement



Introduction	3
Objective	3
Geolocation Based Customer Analysis:	3
Market Segmentation:	4
Approach	4
Data Cleaning	5
Data Processing	5
Further Recommended Problem Statements	5

Introduction

Mahindra First Choice Services (MFCS) is a company of Mahindra Group and is India's leading chain of multi-brand car workshops with over 335+ workshops present in 267+ towns & 24 states. It has serviced over 10,50,000 cars. The company aims to establish countrywide network of over 400 workshops by March 2018.

Mahindra would now like to leverage the data that they have and address the key issues they have. Read along to know how you can help them improve their business.

The dataset consist of three aspects:

1. **Customer data:** where the details of the customer like the car owned, state and place of residence, order type, etc are present. Data dimension is of **534000 Customer entries**
2. **Invoice data:** where information related to customer visits and transactions are recorded, whether a customer as insurance claims, bifurcation of the amount paid, for what type of service did the customer came for, etc...
3. **Material Inventory:** where information related to what kind of service did the customer took and what kind of material was used to service, Labor information and the cost for the service, Plant and plant name where the customer took the service.

Objective

Geolocation Based Customer Analysis:

The idea is to explore how various factors like car make & model, time and type of service etc. vary with location. Since the servicing industry is local in nature, this kind of an analysis could possibly render some really interesting business insights.

Furthermore, this analysis will enable us to formulate more concrete machine learning problems.

From the data at hand it is possible to extract insights about customer behaviour especially the following questions can be addressed

- Problem Statement-1: Identifying the ownership pattern of cars throughout the country. This also captures the problem wherein information regarding the spending patterns can be identified
- Expected Business Outcome: Mahindra First Choice Services will be benefited in multiple ways. Knowing the ownership pattern targeted marketing campaigns could be carried out. Knowing the spending patterns services could be suited to the particular spending pattern.

- Problem Statement-2: Identify the type of order each state receives and present it as an interactive visualization.
- Expected Business Outcome: This could potentially give information about how Mahindra First Choice needs to be prepared to tackle various seasonal cases

Market Segmentation:

Market segmentation is the process of dividing a market of potential customers into internally homogeneous and mutually heterogeneous groups or segments, based on different characteristics captured in the data. Groups created through such a segmentation exercise many times reveal behavioral patterns which are different from generally accepted segments by the business. The exercise is broadly known as “clustering” and is aimed at finding the consumers who will respond similarly to various stimuli by detecting underlying behavior patterns.

Though clustering falls under a Machine Learning problem category called unsupervised learning, which requires extensive efforts, it is possible to carry out a visual analysis in a relatively short timespan.

- Problem Statement: Customer Lifetime value prediction - Based on Customer segments, predict the revenue that can be extracted from each segment over a life of the car -Regression/Time Series.
- Expected Business Outcome: This would be beneficial to Mahindra First Choice Services to identify the various segments in the market. Also, these segmentations would allow for targeted marketing activities and sales promotions.

Approach

You should perform the following activities

1. Cleaning the data
2. Processing and preparing the data for further analysis
3. Analyzing data through various visual tools
4. Building Predictive Models

Data Cleaning

In this process, you can

- Come up with effective measures to handle large volume data.
- Impute the missing values.
- Encode the categorical variables.

Data Processing

- Preparing the data by tagging geolocation with positions
- Deriving Relevant features from multiple tables.
- Aggregating information for each state for countrywide analysis eg. number of Maruti Cars in each state etc.

Further Recommended Problem Statements

1. Inventory Management and Recommendation

- a. We can find different relations on what inventory is used at what scale in how much amount in different states, and their relations with car model and what type of customer it was. With years of data, we can see how the trend changes, what type of demand is increasing, and this correlation with different car companies and models might give a good insight and prediction on requirements for the future.

2. Marketing Recommendation

- a. Based on how much revenue is generated from each marketing source, and how has it varied one can find what type of customers use which marketing source, average income per marketing source and how much does it cost. Thus with a time trend, one can predict where to invest your marketing resources and where to cut down. This can be combined with other data like type of vehicles and type of repair done for more engaging insights.

3. Customer Prediction

- a. Based on last year's data can you come up with numbers for this years customers, what will be the type of cars, issues, and inventory. This will solely focus on time series prediction unlike earlier where time series predictions are used for further insights. This will thus be more technical oriented using which other tools can be made. This will go deeper into the analysis of different techniques and coming up with the best technique for time series prediction.