Bike showroom Data analysis

Mid term submission

Submitted by

Chaudhary Umangkumar Harshadbhai

21f3001035

IITM Online BS Degree Program,

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 60003

Table of Contents:

No	Content	Page No.
1	Executive Summary and Title	2
2	Proof of originality of the Data	3
3	Metadata	5
4	Descriptive Statistics	6
5	Explanation of Analysis Process/Method	7
6	Results and Findings	8

Executive Summary (Bike Showroom)

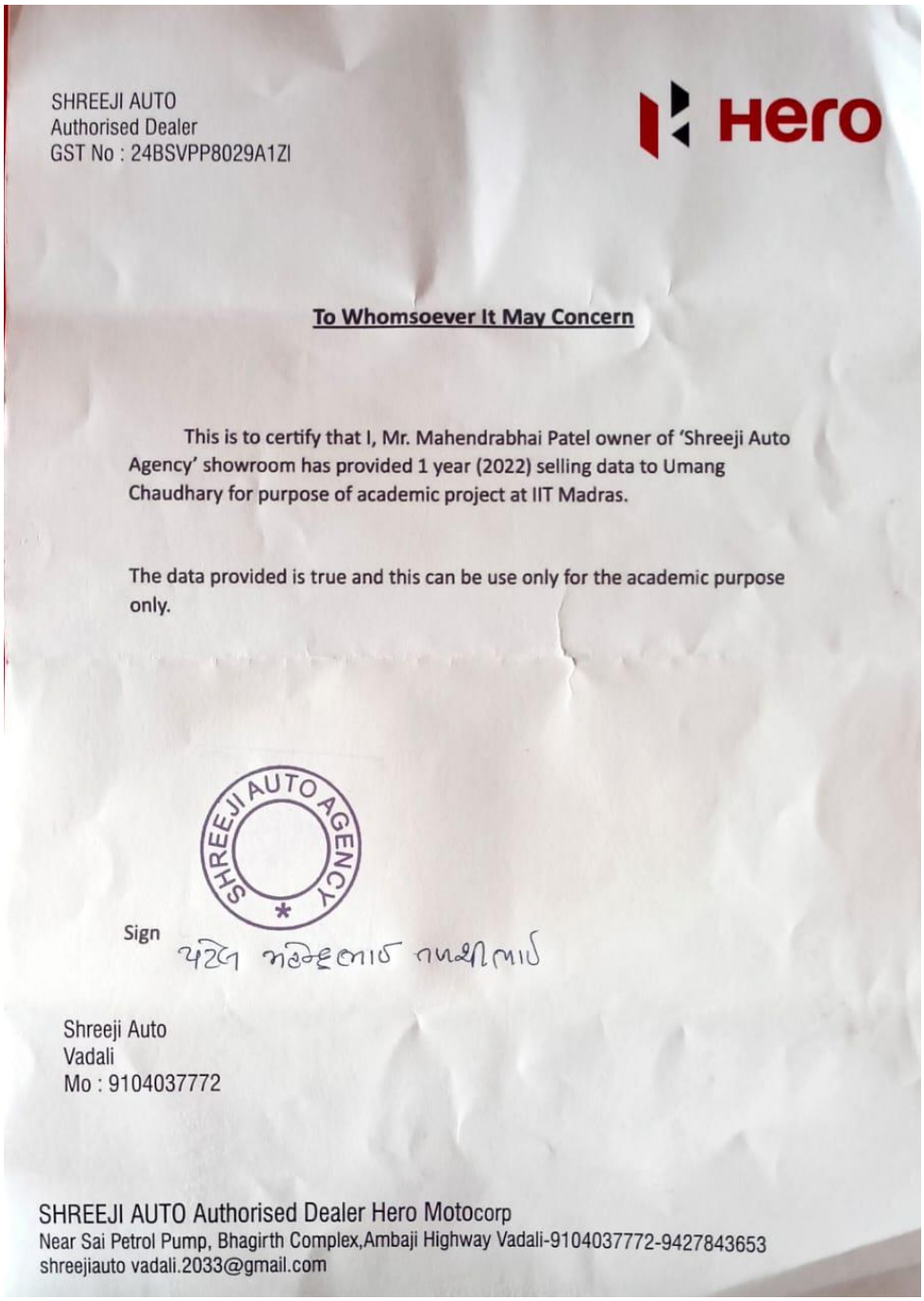
"Shreeji Auto, Agency" is located in 'Vadali', Dis-Sabarkantha, Gujarat. (Location link). It's been 15 years since this showroom opened. 'Shreeji Auto Agency' sells Hero Bikes and scooters. The bike dealership has a large selection of bikes from several reputable manufacturers to suit different riding demands and preferences. They sell scooters like the enjoyment, Maestro, and Destini as well as motorcycles in a variety of models like Splendour Plus, Passion Pro, HF Deluxe, and others. As they sell bikes to customers, bike showrooms are B2C businesses. The bike showroom is partnered by two people named Mahendrabhai and Rajubhai. Additionally, there are a variety of employees, including technicians, salespeople, service managers, financial specialists, etc. They offer certain free bike services to the consumer in order to compete with other bike dealerships and boost sales. Showroom remains open 6 days of the week (closed on Monday) but open in festival days as some customer believes to buy bike on festival day. Showroom's working hours are from morning 9:00 am to evening 6 pm. They keep track of sales data in their own programme, and I receive the data in excel format. As a business analyst, I will comprehend the data and do data cleaning to remove extraneous data from the file. After that, I'll analyze data that has been cleaned up so that I may present insights into the issues the showroom is encountering through various visual graphs. This will result in more sales, which will improve income and profit for the company.

Problem objectives:

1. Increase sales and revenue of the showroom
2. Problem occurs due to different customer behavior.
3. Sales based on different trends, patterns and seasonality
4. Location wise sales are lower compare to other location.

Proof of originality of the Data

Letter head:



SHREEJI AUTO
Authorised Dealer
GST No : 24BSVPP8029A1ZI

Hero

To Whomsoever It May Concern

This is to certify that I, Mr. Mahendrabhai Patel owner of 'Shreeji Auto Agency' showroom has provided 1 year (2022) selling data to Umang Chaudhary for purpose of academic project at IIT Madras.

The data provided is true and this can be use only for the academic purpose only.

SHREEJI AUTO AGENCY

Sign

Shreeji Auto
Vadali
Mo : 9104037772

SHREEJI AUTO Authorised Dealer Hero Motocorp
Near Sai Petrol Pump, Bhagirth Complex,Ambaji Highway Vadali-9104037772-9427843653
shreejiauto vadali.2033@gmail.com

Images:

Recorded Video with owner: (Accessible from IITM email only)

https://drive.google.com/file/d/1Fg501neMqLlyqFTOaZmA0XugZxzxdpgi/view?usp=sharing

Metadata (Bike showroom data analysis)

Data provided by the showroom is given in the below link:

https://1drv.ms/x/s!ArIId7fuekEDlWmTIsm6KP-rUYm6

Name of the above file is “old data.xlsx”. Data was entered in showroom’s private software by showroom. It is downloaded in excel file. This data is provided by financial specialist of the showroom for education purpose only and it is confidential.

This data contains 345 rows and 15 columns.

```
RangeIndex: 345 entries, 0 to 344
Data columns (total 15 columns):
 #   Column           Non-Null Count  Dtype 
---  ------           --------------  ----- 
 0   DATE             345 non-null    object
 1   SKU              345 non-null    object
 2   SKU Description  345 non-null    object
 3   NAME             345 non-null    object
 4   CITY/Village     345 non-null    object
 5   contact ID       237 non-null    object
 6   MO               345 non-null    int64 
 7   HP               345 non-null    object
 8   BILL NO          341 non-null    object
 9   REG NO           329 non-null    object
 10  STATE            333 non-null    object
 11  Dealer           345 non-null    object
 12  Vehicle class    342 non-null    object
 13  plant            345 non-null    object
 14  PRICE            345 non-null    object
dtypes: int64(1), object(14)
memory usage: 40.6+ KB
```

There is different type of columns like string, float and integer in which only 1 column is integer, others are of string datatype. From the above image, names of 15 columns are given. The data is collected of year 2022. Size of excel data file is approximate 41 kb.

Data collection timeline: [6th June]

After understanding of the data, I came to knew that some columns are not useful so I dropped some columns and change the datatype of some columns for better data analysis.

Here is the link given below of cleaned data:

https://1drv.ms/x/s!ArIId7fuekEDlWZsHPK914LZhWeB

Descriptive Statistics

Here is the null values of all columns of cleaned dataset.

```
DATE              0
Month             0
SKU               0
SKU Description   0
NAME              0
CITY/Village      0
Tehsil            0
HP                0
PRICE             0
bike colour       0
vehicle type      0
dtype: int64
```

Above image is of null values containing rows. ‘contact id’ column contains 108 null values, ‘REG NO’ column contains 16 null values, ‘BILL NO’ column contains 4 null values, ‘Vehicle class’ column contain 3 null values, ’STATE’ column contains 12 null values of old data.

Statistics of ‘PRIZE’ column:

```
count       346.000000
mean      62853.032601
std       10578.372626
min       37942.000000
25%       58425.840000
50%       64099.840000
75%       67179.840000
max      118027.840000
Name: PRICE, dtype: float64
```

In the data there is only one column ‘PRIZE’ is numerical so statistics of this column is given above. There is total 346 rows and mean of the prize is 62853.032601. Also, standard deviation, minimum value, maximum value, 25%, median (50%), 75% of the column is given in the image. And Mode of the prize columns is 67179.84 Rs.

Explanation of Analysis Process/Method

The major programme I utilised for this project was Microsoft Excel, and I also used Python for some data statistics. Therefore, the owner's data was unstructured and the dataset had certain undesirable columns. After eliminating those rows, I created some more rows from the remaining rows. When the data was small enough, I utilised Ms. Excel to create pivot tables for visualisation.

Data Preprocessing:

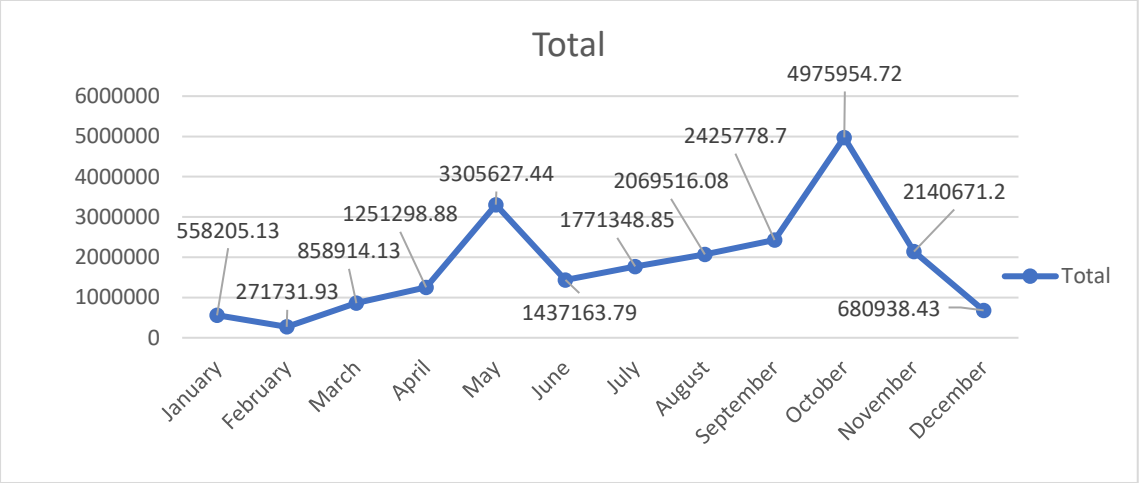
- Month column is derived from the DATE column using excel formula =TEXT(Cell,"mmmm").
- Contact ID, BILL NO, REG NO, STATE, DEALER, vehicle class, plant, these are columns are removed as these are unnecessary columns.
- Bike colour is derived from SKU description with the help of showroom's owner.
- I have manually created column 'Tehsil' with the help of showroom's owner.

Analysis:

- Data is of year 2022. Data has 11 columns after preprocessing of the data.
- I have created pivot table using 'month' as rows and 'prize' as values, based on that made visualization graph. This graph describes month wise sales of year 2022.
- Location wise sales can be identified with the help of column 'Tehsil'.
- Using the column 'vehicle type', it can be decided that which type of vehicle should show room keep whether it is bike or scooter.
- 'bike colour' is used to decide that customer choose which colour most.
- Made pivot table of all SKU's and if count of any SKU is less than or equal to 5, I replace it to other SKU and Made a new pivot table.

Results and Findings

Revenue generated in 2022:

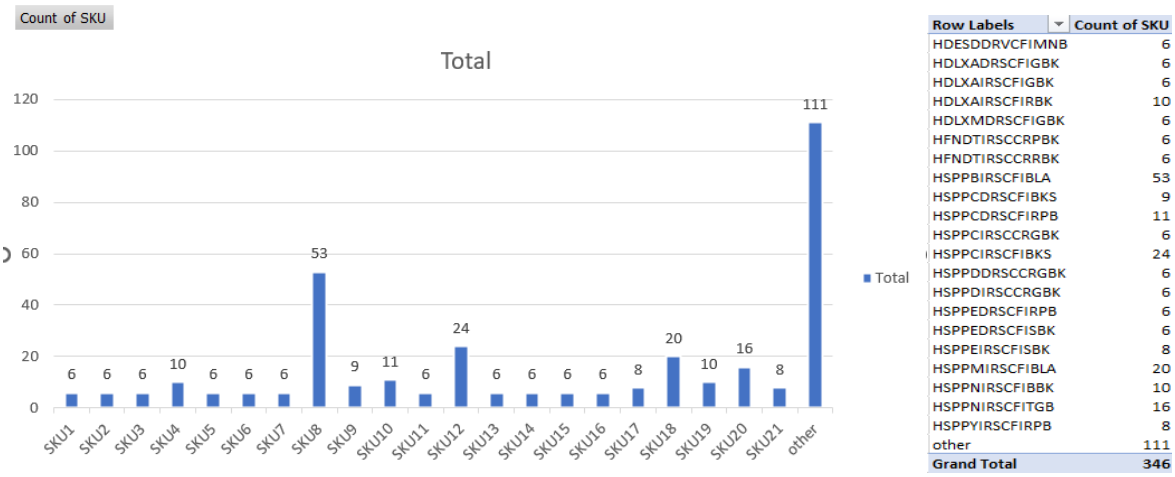


The graph above shows the revenue for each month in 2022. The amount of money created in October is higher than in any other month. Revenue for the first five months is ranked higher, then the graph starts to decline in June. Once more, revenue rose, but only dramatically in the months of October and again in the months of November and December. Thus, October 2022 is the year's busiest sales month. Sales are higher during the 'Navratri' and 'Diwali' festivals, which explains why October is a better month for sales.

Customer Behavior Analysis:

Sales of different bike models: -

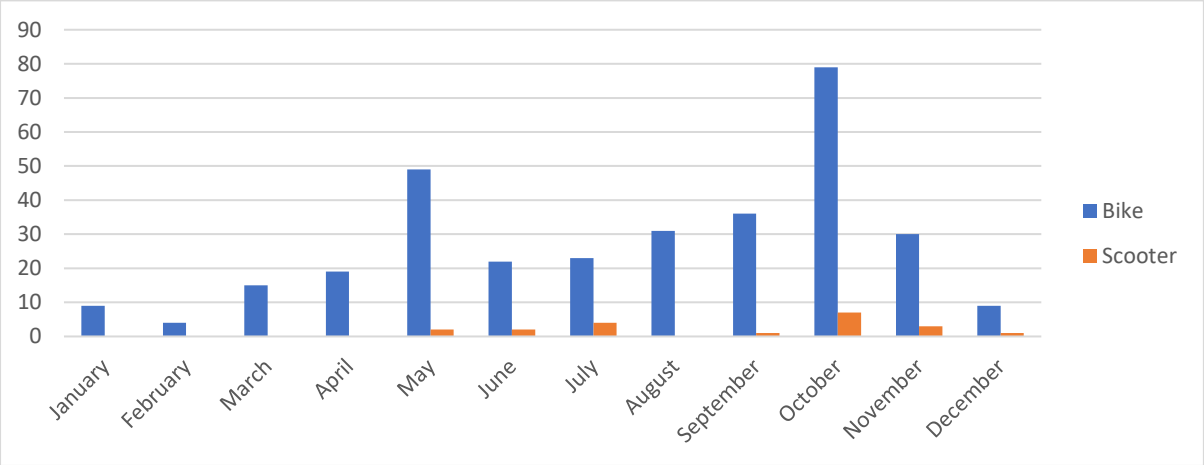
68 distinct bike models were available. I reached 21 SKU and another SKU, bringing the total to 22, and I gave each SKU a name, as seen in the table below.



Row Labels	Count of SKU
HDESDDRVCFIMNB	6
HDLXADRSCFIGBK	6
HDLXAIRSCFIGBK	6
HDLXAIRSCFIRBK	10
HDLXMDRSCFIGBK	6
HFNDTIRSCCRPBK	6
HFNDTIRSCCRRBK	6
HSPPBIRSCFIBLA	53
HSPPCDRSCFIBKS	9
HSPPCDRSCFIRPB	11
HSPPCIRSCCRGBK	6
HSPPCIRSCFIBKS	24
HSPPDDRSCCRGBK	6
HSPPDIRSCCRGBK	6
HSPPEDRSCFIRPB	6
HSPPEDRSCFISBK	6
HSPPEIRSCFISBK	8
HSPPMIRSCFIBLA	20
HSPPNIRSCFIBBK	10
HSPPNIRSCFITGB	16
HSPPYIRSCFIRPB	8
other	111
Grand Total	**346**

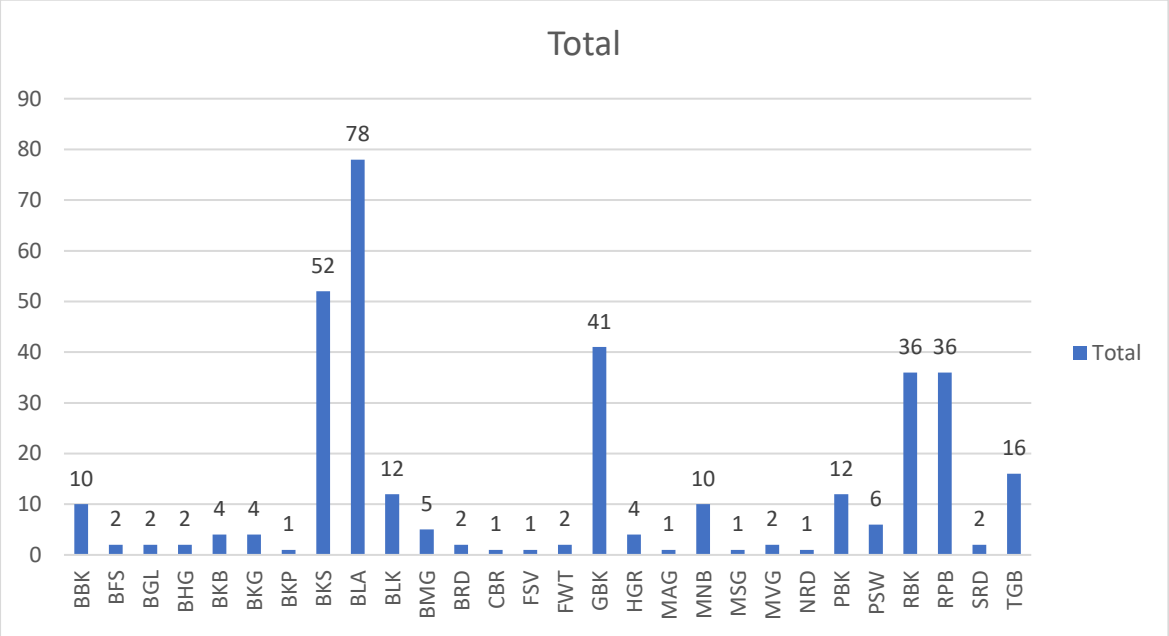
According to the graph, the bike with the largest sales is SKU8 (HSPPBIRSCFIBLA)(SPLENDOR + I3S BS6 FI DRUM SELF CAST BLACK AND ACCENT), whereas other categories have sales that are less than or equal to 5.

Sales difference between bikes and scooters: -



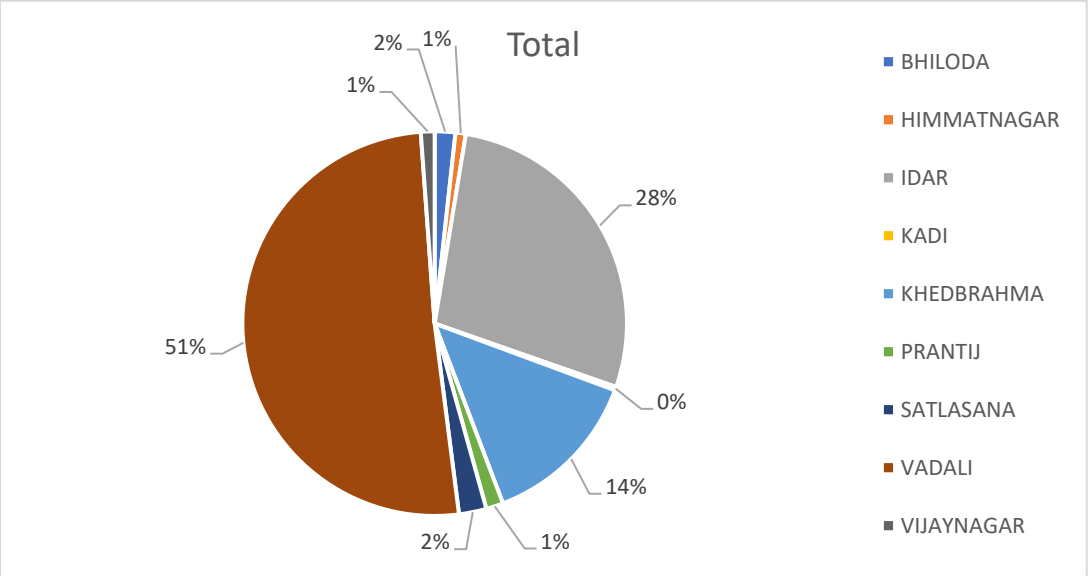
The graph plainly demonstrates that scooter sales are far lower than those of motorcycles. Scooter sales occur in several months, including October, July, and others.

Bike color analysis: -



The most popular bike colour is BLA (Black with Accent). Other popular colours are BKS, GBK, RBK, RPB, PBK, MNB, BLK, TGB, and BBK. Some bike colours, such as BRD and CBR, sell poorly.

Location Wise Analysis:



'VADALI' accounts for 51% of bike sales, followed by 'IDAR' at 28%. Selling in the KADI Tehsil is practically 0%.