

PROFESSIONAL TRAINING REPORT

AT

**Sathyabama Institute of Science and Technology
(Deemed to be University)**

Submitted in partial fulfillment of the requirements for the award
Of Bachelor of Technology Degree in Information Technology

By

SANTHOSH J

40120081



**DEPARTMENT OF INFORMATION TECHNOLOGY
SCHOOL OF COMPUTING**

SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)**

Accredited with Grade "A" by NAAC

**JEPPIAAR NAGAR, RAJIV GANDHI SALAI,
CHENNAI - 600 119.TAMILNADU**

APRIL-2023



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE

www.sathyabama.ac.in

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the Bonafide work of **SANTHOSH J ,40120081** who carried out the project entitled " **BIKE SALES PREDICTION**" under my supervision from January 2023 to April 2023.

INTERNAL GUIDE

Dr.C.Geetha M.E., Ph.D.,

HEAD OF THE DEPARTMENT

Dr. R. Subhashini M.E., Ph.D.,

Submitted for Viva voce Examination held on 29-04-2023

Internal Examiner

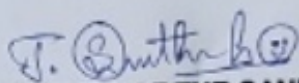
External Examiner

DECLARATION

I santhosh j hereby declare that the Project Report entitled "**BIKE SALES PREDICTION**" in **power bi** done by me under the guidance of **Ms.C.Geetha** At Sathyabama Institute of Science and Technology (Deemed to be University), Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai – 600119 is submitted in partial fulfillment of the requirements for the award of Bachelor of Technology degree in Information Technology

DATE: 29/04/23

PLACE: Chennai


SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **Board of Management of SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr.T.SASIKALA M.E.,Ph.D, Dean**, School of Computing **Dr.R.Subhashini** Head of the Department of Information Technology for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **MS.C.GEETHA** for his valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Information Technology** who were helpful in many ways for the completion of the project

TRAINING CERTIFICATE

No.: 34354

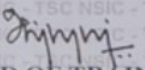

एन एस आई सी
N S I C

राष्ट्रीय लघु उद्योग निगम लिमिटेड
NATIONAL SMALL INDUSTRIES CORPORATION LIMITED
(A Govt. of India Enterprise)
NSIC - TECHNICAL SERVICES CENTRE
Sector B-24, Guindy Industrial Estate, Ekkaduthangal, Chennai - 600 032.

CERTIFICATE

This is to certify that Mr. **SANTHOSH J.** S/o.
Shri. **V. K. JAYARAMAN**, student of III year
B.Tech. (IT), Sathyabama Institute of Science and
Technology has undergone "**SKILL DEVELOPMENT
PROGRAM ON DATA ANALYTICS**" conducted by
us for a period of one week (30 Hours) from
09.03.2023 to 14.03.2023.




HEAD OF TRAINING


HEAD OF CENTRE

ABSTRACT

This data contains bike sales in India and the relationship between leading profits with age groups of people and according to males and females and needs that can be purchased when buying a bike and we can predict future prices using machine learning once the product is developed so the machine will make the decision in setting the price and this data was analyzed by reviewing bicycle data and extracting insights into sales in each of their countries. And calculate the percentage of profits in each of the countries. Also, who buys bikes more, males or females, and what are the most expensive types of bikes, who the most years in which bikes were sold? The results were presented in graphic images that easily simulate this information.

Key words—Auto-regressive integrated moving average, forecasting, motorcycle parts, time series.

TABLE OF CONTENT

CHAPTER	TITLE	PAGE NO
	ABSTRACT	
	LIST OF FIGURES	
1	INTRODUCTION	10
2	AIM AND SCOPE	11
2.1	AIM	11
2.2	SCOPE	12
3	SYSTEM REQUIREMENTS	13
4	METHODOLOGY	14
4.1	Data collection	15
4.2	Data mining	16
4.3	Data analysis	17
4.4	Types of analysis	18
4.4.1	Descriptive analysis	18
4.4.2	Diagnostic analysis	19
4.4.3	Predictive analysis	20
4.4.4	Prescriptive analysis	21
4.5	Power bi	22
4.5.1	Power bi uses	22
4.5.2	Power bi features	23
4.6	Architecture	24
4.7	Project implementation	25
4.7.1	Getting the data	25

4.7.2	Data preprocessing	25
4.7.3	Forest random classifier	25
5.	Indian two wheeler market analysis	27
5.1	Road to owning a two-wheeler	29
5.1.2	Key factors contributing to the purchase decision	29
5.2	Impact by the surrounding culture	30
5.2.3	Case study analysis	30
6	Tvs motor performance study report	31
6.1	Bajaj auto pic up speed	33
6.2	Honda twowheeler market assessment report	35
7	Interaction among visuals in project	37
7.1	Pie chart	37
7.2	Count of bike model by owner	38
7.3	slicer	39
7.4	Overall representation of data in visuals	40
8	CONCLUSION	41
9	REFERENCE	42

1.INTRODUCTION

The motorcycle is a vehicle that is most in demand by the people of Indonesia. Apart from the relatively low price, the ease of maintenance is also a separate consideration. The Central Statistics Agency (BPS) shows the number of motorbikes in INDIA in 2016 has reached more than 105 million units. Since

2012 - 2016, the number of bikes has increased by 8.32 percent per year [1]. These statistics show that every year, motorcycle users are growing. Along with the

growth of motorcycle users, the need for repair services or so-called workshops and sales of motorcycle parts is also increasing. In the field of selling motorcycle parts, stock availability is critical. In addition to increasing revenue, stock management

is also essential to maintain consumer confidence. The sellers must ensure the availability of stock so as not to suffer losses. Auto parts must be available when the consumer needs the spare parts. Too little amount of stock makes it possible to run out of stock, which results in not maximizing store revenue. Conversely, the amount of too much stock allows the capital turnover to be wrong. These happen because capital settles into stock for a long time. In addition to the related capital turnover, excessive amounts of stock can cause a buildup of storage warehouses.

The bicycle market is projected to register a CAGR of 4.09% over the upcoming five years.

The sales of bicycles initially witnessed a slump due to the closure of offline stores worldwide due to the COVID-19 pandemic-induced restrictions. However, after the pandemic, the demand for bicycles started picking up recently and almost came back to normal, with significant growth in consumers trying to avoid public transportation due to a fear of contracting the virus. Moreover, with the growing health consciousness among consumers, bicycles are now seen as an increasingly dependable and healthy mobility option. For instance, in September 2021, the Dutch e-bike brand VanMoof raised over USD 128 million from venture capitalists as sales of battery-powered bikes (e-bikes) tripled during the pandemic.

Furthermore, the number of cycling events increased in the middle of 2021. Over the forecast period, favorable government and employer initiatives are expected to drive the market studied. In addition, various governing bodies in Europe, Asia-Pacific, and Africa, particularly South Africa, promote road racing, touring and leisure cycling, track racing, off-road racing, and mountain biking events.

2. AIM AND SCOPE

2.1 AIM

The global bicycle market size was valued at USD 78.33 billion in 2021 and is projected to grow from USD 82.50 billion in 2022 to USD 127.83 billion by 2029, exhibiting a CAGR of 6.5% during the forecast period. The global COVID-19 pandemic has been unprecedented and staggering, with bicycles experiencing higher-than-anticipated demand across all regions compared to pre-pandemic levels. Based on our analysis, the global market exhibited a growth of 48.6% in 2020 as compared to 2019. As the global number of the health-conscious populace is increasing, so is the demand for bikes. Bikes are highly utilized by the populace for activities such as commuting, exercise, sport, hiking, and others. Therefore, manufacturers offer various types or categories specially designed for these applications to cover the larger customer base. Some of the categories offered by manufacturers include mountain bikes, road bikes, hybrid bikes, cruiser bikes, and others. Rising traffic congestion or urbanization and growing environmental concerns among the populace are driving the global market growth. For instance, according to the United Nations Conference on Trade and Development (UNCTAD) Handbook of Statistics, the global urban population percentage grew from 51.6% in 2010 to 56.2% in 2020. The share of the urban population has increased rapidly in the previous decade. It is generally higher in developed economies (around 79.2% in 2020) than in developing economies (approximately 51.6% in 2020). The issue of bike stations that is when a customer goes to bike station and finds that there is no place to repay the bike can be answered with the forecasts. If it known how many bikes will be needed each day then the bike redistribution can be planned accordingly.

2.2 SCOPE

In this project, the proposed methods for prediction of bike buyers using four machine learning techniques that were used include SVM, Random Forest, Decision tree, KNN. The system is implemented by using all the models and evaluates their performance. Performance evaluation is completely based on different performance metrics. Random Forest Classifier resulted highest accuracy of 0.86. From the above results Random Forest plays a key role in shaping improved classification accuracy of a dataset. To extend the project as to predict bikes using IOT, in that all bike stations can be connected over the internet using IOT. The sensors can be employed which will collect the information about the current bike stations. This is the future scope of project. Suppose, if we have given training to our machine learning model by a dataset and we test it by a completely different dataset. Then, it will create difficulties for our model to understand the correlations between the models. If we train our model very well and its training accuracy is also very high, but we provide a new dataset to it, then it will decrease the performance. So we always try to make a machine learning model which performs well with the training set and also with the test dataset. The dependent variable can be categorized into levels, so this problem falls under Classification algorithms in machine learning.

3 SYSTEM REQUIREMENTS

Windows

- Microsoft® Windows® 7/8/10 64-bit
- 4 GB RAM minimum, 8 GB RAM recommended
- 2 GB of available disk space minimum,
- 4 GB Recommended 500 MB for IDE + 1.5 GB for Android SDK and emulator system image
- 1280 x 800 minimum screen resolution

Mac

- Mac® OS X® 10.10 Yosemite or higher, up to 10.14 macOS Mojave
- 4 GB RAM minimum, 8 GB RAM recommended
- 2 GB of available disk space minimum,
- 4 GB Recommended 500 MB for IDE + 1.5 GB for Android SDK and emulator system image
- 1280 x 800 minimum screen resolution

Linux

- GNOME or KDE desktop - Tested on gLinux based on Debian.
- 64-bit distribution capable of running 32-bit applications
- GNU C Library glibc 2.19 or later
- 4 GB RAM minimum, 8 GB RAM recommended
- 2 GB of available disk space minimum

4 METHODOLOGY

The different machine learning techniques in this project, predicting whether a bike buyer buys a bike or not. This system will increase the possibilities of buying bikes based on accuracy. Considered a dataset which is categorized so that dataset follows under classification model. On the same dataset by applying all classification algorithms and will compare accuracy values from all algorithms and find best suitable algorithm for predicting bike buyers which gives best performance.

Some of the classification algorithms used are:

- Random Forest Classifier
- Support Vector Machine
- Decision trees

Another major branding myth that needs to be eradicated is the thinking that if some strategy worked for a company, it's most likely to work for another company too. Each brand is unique in its own way and its connection that it forms with the customer, and each of these bonds need to be made stronger with a different algorithm. Using what worked on another brand onto your company has been one of the major causes of failures in a multitude of brands, and has demolished the existing bonds with the customers.



FIGURE 4.1

4.1 DATA COLLECTION

Two wheeler has an entirely different branding strategy for each of its products, as well as individual segments. They strategies based on their target customers, products and pricing. The premium segments have an even more elaborate and differing battle strategy for branding. Some examples of the premium segment include Harley Davidson, Triumph, KTM and Indian. Branding, as said earlier involves a great deal of balance, and multi tasking in the multifaceted life cycle of the product, and is not only tested by the strength of advertisements. The two wheeler branding also has some interesting news on the gossip front. TVS is known for making use of cricket celebrities, the Scooty and Scooter segment are endorsed by Bollywood celebrities and at other times, Scooty is advertised by new faces, predominantly women. Also frowned upon, is relaunching and reverting back products with the same name. This raises many questions about the brand as such. For example, TVS relaunched their TVS Victor in 2007 after it was discontinued in 2001. Initially, TVS launched their highly successful Scooty product series. Later, they broadened it to launch various products named Scooty, Scooty zest 110, Scooty Pep and Scooty Pep Plus. A variant, namely Jupiter has made its entry into the market is proving to be very successful in taking the place of Honda Activa. An added advantage to the company is having Big B as an ambassador that has entrusted several customers in its web of trust and legacy. Although TVS is a name that is known almost everywhere, it still has miles to go and milestones to reach before it can be satisfied. They have strongly influenced the mass with well-known faced. Now has come the time to move the mass with its technology and design and take major strides in that area too. Till date, it has achieved this only with its Apache.

Similarly, Bajaj's signature product has always been the Pulsar. It has extended this product multiple times with changes made in each version of the bike.

There also exist companies like Yamaha which are not entirely keen on relaunching any of their products, though they do have variants of a product in the market. Yamaha have made a bold move as the Brand Ambassador with bikes and mopeds. Furthermore, their advertisements feature John Abraham and Deepika Padukone which have gained them a permanent place in the hearts of speed lovers and respective fan clubs.

4.2 DATA MINING

In this section, there is an overview of the process of data mining and data modeling, collecting the data, through the data preparation and finally the data modeling. Data cleaning and formatting can be considered as the most critical part of the whole project according to several data scientists. Figure shows how the process of data mining works to extract knowledge using algorithms.

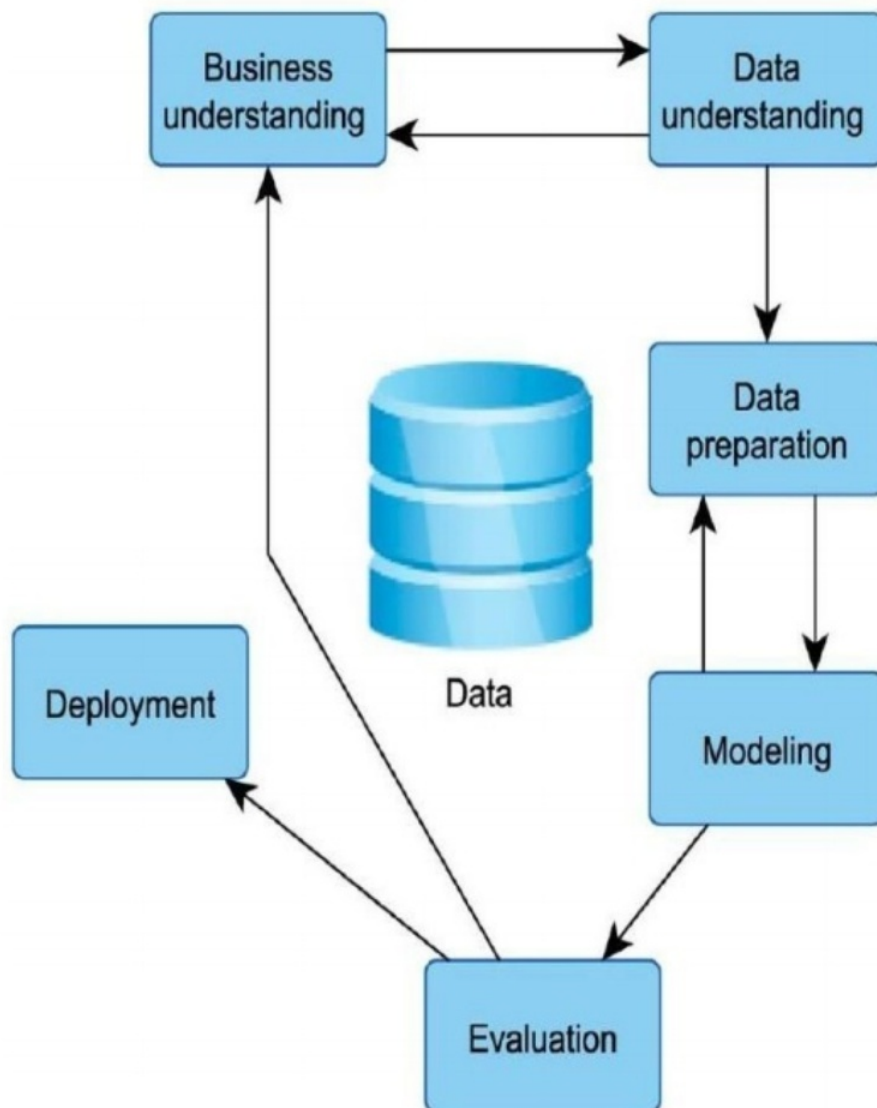


FIGURE 4.2 : DATA MINING

4.3 DATA ANALYSIS

The process of data analysis, or alternately, data analysis steps, involves gathering all the information, processing it, exploring the data, and using it to find patterns and other insights. The process of data analysis consists of: Data Requirement Gathering: Ask yourself why you're doing this analysis, what type of data you want to use, and what data you plan to analyze.

Data Collection: Guided by your identified requirements, it's time to collect the data from your sources. Sources include case studies, surveys, interviews, questionnaires, direct observation, and focus groups. Make sure to organize the collected data for analysis. Data Cleaning: Not all the data you collect will be useful, so it's time to clean it up. This process is where you remove white spaces, duplicate records, and basic errors. Data cleaning is mandatory before sending the information on for analysis.

Data Analysis: Here is where you use data analysis software and other tools to help you interpret and understand the data and arrive at conclusions. Data analysis tools include Excel, Python, R, Looker, Rapid Miner, Charito, Meta base, Rehash, and Microsoft Power BI. Data Interpretation: Now that you have your results, you need to interpret them and come up with the best courses of action based on your findings.

Data Visualization: Data visualization is a fancy way of saying, "graphically show your information in a way that people can read and understand it." You can use charts, graphs, maps, bullet points, or a host of other methods. Visualization helps you derive valuable insights by helping you compare datasets and observe relationships.

If you want to make informed decisions, you need data, but there's more to it. The data in question must be accurate.

4.4.TYPES OF DATA ANALYSIS

The four types of data analysis are:

- ☐ Descriptive Analysis
- ☐ Diagnostic Analysis
- ☐ Predictive Analysis
- ☐ Prescriptive Analysis

4.4.1DESCRIPTIVE ANALYSIS

Descriptive analytics answer the question, "What happened?". This type of analytics is by far the most commonly used by customers, providing reporting and analysis centered on past events. It helps companies understand things such as:

- How much did we sell as a company?
- What was our overall productivity?
- How many customers churned in the last quarter?

Descriptive analytics is used to understand the overall performance at an aggregate level and is by far the easiest place for a company to start as data tends to be readily available to build reports and applications. It's extremely important to build core competencies first in descriptive analytics before attempting to advance upward in the data analytics maturity model. Core competencies include things such as data modeling fundamentals, communicating data with right visualization and basic dashboard skills.

It's likely you've adopted some form of descriptive analytics internally, whether that be static P&L statements, PDF reports, or reporting within an analytics tool. For a true descriptive analytics program to be implemented, the concepts of repeatability and automation of tasks must be top of mind.

4.4.2.DIAGNOSTIC ANALYSIS

Diagnostic analytics, just like descriptive analytics, uses historical data to answer a question. But instead of focusing on “what”, diagnostic analytics addresses the critical question of why an occurrence or anomaly occurred within your data. Diagnostic analytics also happens to be the most overlooked and skipped step within the analytics maturity model. Anecdotally, I see most customers attempting to go from “what happened” to “what will happen” without ever taking the time to address the “why did it happen” step. This type of analytics helps companies answer questions such as:

- Why did our company sales decrease in the previous quarter?
- Why are we seeing an increase in customer churn?
- Why is a specific basket of products vastly outperforming their prior year sales figures?

Diagnostic analytics tends to be more accessible and fit a wider range of use cases than machine learning/predictive analytics. You might even find that it solves some business problems you earmarked for predictive analytics use cases. Being at the diagnostic analytics phase likely means you’ve adopted a modern analytics tool. Most modern analytics tools contain a variety of search-based, or lightweight artificial intelligence capabilities.

These features allow for detailed insights a layer deeper (for example: the Key Drivers visualization in Power BI, or Qlik’s search-based insight functionality). To be clear, these are an effective lightweight means to address diagnostic analytics use cases but are not a means to a full-scale implementation.

4.4.3.PREDICTIVE ANALYSIS

Predictive analytics is a form of advanced analytics that determines what is likely to happen based on historical data using machine learning. Historical data that comprises the bulk of descriptive and diagnostic analytics is used as the basis of building predictive analytics models. Predictive analytics helps companies address use cases such as:

- Predicting maintenance issues and part breakdown in machines.
- Determining credit risk and identifying potential fraud.
- Predict and avoid customer churn by identifying signs of customer dissatisfaction.

At the outset of any predictive analytics build, three core elements need to be established:

- Identify a problem to solve,
- Define what is you want to predict, and
- State what you will achieve by doing so.

To start you should collect existing data, organize data in a useful way to allow for data modeling, cleanse your data and review overall quality, and finally determine your modeling objective.

While modeling takes up the spotlight in predictive analytics, data prep is a crucial step that needs to happen first. This is why organizations with a rock-solid foundation in descriptive and diagnostic analytics are better equipped to handle predictive analytics. Simply put, the time and effort to prep, transform, and ensure data quality for retrospective reporting has already taken place. The groundwork should be relatively well laid to quickly identify and leverage data for the modeling phase

4.4.4.PRESCRIPTIVE ANALYSIS

Prescriptive analytics is the fourth and final pillar of modern analytics. Prescriptive analytics pertains to true guided analytics where your analytics is prescribing or guiding you toward a specific action to take. It is effectively the merging of descriptive, diagnostic, and predictive analytics to drive decision making. Existing scenarios or conditions (think your current fleet of freight trains) and the ramifications of a decision or occurrence (parts breakdown on the freight trains) are applied to create a guided decision or action for the user to take (proactively buy more parts for preventative maintenance).

Prescriptive analytics requires strong competencies in descriptive, diagnostic, and predictive analytics which is why it tends to be found in highly specialized industries (oil and gas, clinical healthcare, finance, and insurance to name a few) where use cases are well defined. Prescriptive analytics help to address use cases such as:

- Automatic adjustment of product pricing based on anticipated customer demand and external factors.
- Flagging select employees for additional training based on incident reports in the field.

Prescriptive analytics' primary aim is to take the educated guess or assessment out of data analytics and streamline the decision-making process. Prescriptive analytics is commonly considered the merging of descriptive, diagnostic, and predictive analytics. Getting started isn't so much a step-by-step list but rather the time and effort up front to build your competencies within the analytics maturity curve.

4.5.POWER BI

Microsoft Power BI is a business intelligence (BI) platform that provides nontechnical business users with tools for aggregating, analyzing, visualizing and sharing data. Power BI's user interface is fairly intuitive for users familiar with Excel, and its deep integration with other Microsoft products makes it a versatile self-service tool that requires little upfront training.

Users can download an application for Windows 10, called Power BI Desktop, and native mobile apps for Windows, Android and iOS devices. There is also Power BI Report Server for companies that must maintain their data and reports on premises. That version of Power BI requires a special version of the desktop app -- aptly called Power BI Desktop for Power BI Report Server.

4.5.1.USES OF POWER BI

Microsoft Power BI is used to find insights within an organization's data. PowerBI can help connect disparate data sets, transform and clean the data into a data model and create charts or graphs to provide visuals of the data. All of this can be shared with other Power BI users within the organization.

The data models created from Power BI can be used in several ways for organizations, including the following:

- telling stories through charts and data visualizations;
- examining "what if" scenarios within the data; and
- creating reports that can answer questions in real time and help with forecasting to make sure departments meet.

4.5.2.FEATURES OF POWER BI

Microsoft has added a number of data analytics features to Power BI since its inception and continues to do so. Some of the most important features are the following:

Artificial intelligence. Users can access image recognition and text analytics in PowerBI, create machine learning models using automated ML capabilities and integrate with Azure Machine Learning.

Hybrid deployment support. This feature provides built-in connectors that allow PowerBI tools to connect with a number of different data sources from Microsoft, Salesforce and other vendors.

Quick Insights. This feature allows users to create subsets of data and automatically apply analytics to that information.

Common data model support. Power BI's support for the common data model allows the use of a standardized and extensible collection of data schemas (entities, attributes and relationships).

Cortana integration. This feature, which is especially popular on mobile devices, allows users to verbally query data using natural language and access results using Cortana, Microsoft's digital assistant.

Customization. This feature allows developers to change the appearance of default visualization and reporting tools and import new tools into the platform.

APIs for integration. This feature provides developers with sample code and application program interfaces (APIs) for embedding the Power BI dashboard in other software products.

Self-service data prep. Using Power Query, business analysts can ingest, transform, integrate and enrich big data into the Power BI web service.

4.6.ARCHITECTURE

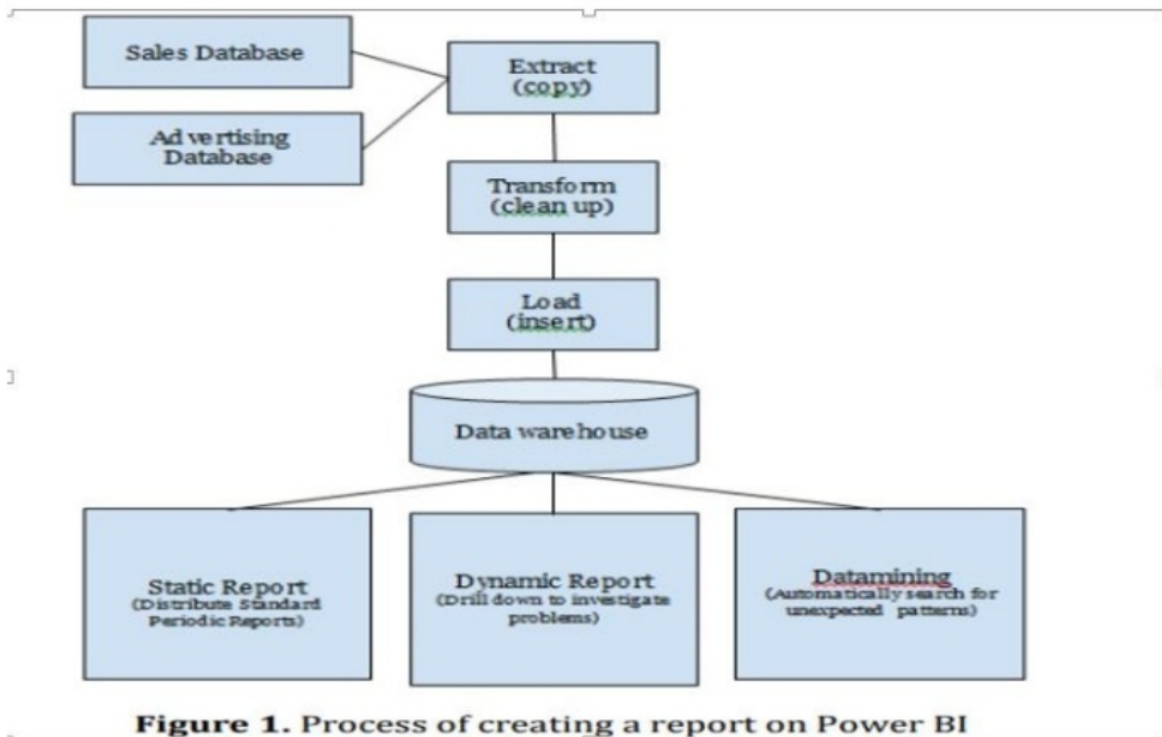


Fig 4.2: Architecture

4.7.PROJECT IMPLEMENTATION

Bike sharing systems use analytics to a great extent successfully. From the above algorithms, we consider Decision Tree Algorithm because this project depends on human decision that is whether a person able to buy a bike or not. The system architecture is as follows:
The dataset includes;

4.7.1 GETTING THE DATA

To create a machine learning model, the first thing we required is a dataset as a machine learning model completely works on data. However, sometimes, we may also need to use an HTML or xlsx file. The Bike Buyer Dataset comprised of 13 different attributes of 6995 customers. The bike buyer were described as either „Yes“ or „No“ on the basis of bike buyer or not. The class variable or attribute is Bike Buyer. Dataset consists of both categorical and numerical attributes. There are few missing values also.

4.7.2 DATA PREPROCESSING

When a data is incomplete or inconsistent or noisy then data preprocessing is needed. There are so many routes to deal with unprocessed data: i)Data Cleaning: To overfill the missing values in data by using mean or median or mode or any formulae. ii)Data Reduction: In this step, the dataset is altered so the results processed by the any model will be the identical but not needed values in dataset are removed. iii)Data Integration: In this step, data is combined from various origin if required and redundancies are removed too. iv)Data Transformation: In this step, operations like normalization is performed. Considered Random Forest Classifier because this project depends on human decisions and consider 13 attributes which may split into group of decision trees. This algorithm deals with decision trees that are to find whether a customer will able to buy a bike or not.

4.7.3 FOREST RANDOM CLASSIFIER

Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset. Decision Tree is a pictorial representation for getting all the possible emulsions to a issue based on given constrains which is known as decision tree

The following diagram shows the datasets taken

	A	B	C	D	E	F	G	H
	name	selling_price	year	seller_type	owner	km_driven	ex_showroom_price	
1	name							
2	Royal Enfield Classic 350	175000	2019	Individual	1st owner	350		
3	Honda Dio	45000	2017	Individual	1st owner	5650		
4	Royal Enfield Classic Gunmetal Grey	150000	2018	Individual	1st owner	12000	148114	
5	Yamaha Fazer FI V 2.0 [2016-2018]	65000	2015	Individual	1st owner	23000	89643	
6	Yamaha SZ [2013-2014]	20000	2011	Individual	2nd owner	21000		
7	Honda CB Twister	18000	2010	Individual	1st owner	60000	53857	
8	Honda CB Hornet 160R	78500	2018	Individual	1st owner	17000	87719	
9	Royal Enfield Bullet 350 [2007-2011]	180000	2008	Individual	2nd owner	39000		
10	Hero Honda CB2 extreme	30000	2010	Individual	1st owner	32000		
11	Bajaj Discover 125	50000	2016	Individual	1st owner	42000	60122	
12	Yamaha FZ16	35000	2015	Individual	1st owner	32000	78712	
13	Honda Navi	28000	2016	Individual	2nd owner	10000	47255	
14	Bajaj Avenger Street 220	80000	2018	Individual	1st owner	21178	95955	
15	Yamaha YZF R3	365000	2019	Individual	1st owner	1127	351680	
16	Jawa 42	185000	2020	Individual	1st owner	1700		
17	Suzuki Access 125 [2007-2016]	25000	2012	Individual	1st owner	55000	58314	
18	Hero Honda Glamour	25000	2006	Individual	1st owner	27000		
19	Yamaha YZF R15 S	40000	2010	Individual	2nd owner	45000	117926	
20	Royal Enfield Classic Gunmetal Grey	150000	2018	Individual	1st owner	23000	148114	
21	Yamaha FZ25	120000	2018	Individual	1st owner	39000	132680	
22	Hero Passion Pro 110	15000	2008	Individual	1st owner	60000		
23	Honda Navi [2016-2017]	26000	2016	Individual	1st owner	17450	44389	
24	Honda Activa I	32000	2013	Individual	2nd owner	20696	53900	
25	Jawa Standard	180000	2019	Individual	1st owner	2000		
26	Royal Enfield Thunderbird 350	110000	2016	Individual	1st owner	20000		
27	Honda Dream Yuga	25000	2012	Individual	1st owner	35000	56147	
28	TVS Apache RTR 160 4V	80000	2018	Individual	1st owner	15210		
29	Honda Navi [2016-2017]	42000	2017	Individual	1st owner	24000	44389	
30	Yamaha Fazer [2009-2016]	40000	2013	Individual	3rd owner	35000	84751	

FIGURE 4.3 : DATASET 1

	A	B	C	D	E	F	G	
	name	selling_price	year	seller_type	owner	km_driven	ex_showroom_price	
34	Hero Xtreme Sports	43000	2016	Individual	2nd owner	50000	78350	
35	Honda X-Blade	65000	2018	Individual	1st owner	27000		
36	Honda CB Shine SP	62000	2019	Individual	1st owner	10000	71049	
37	Honda Activa 5G	60000	2018	Individual	1st owner	6502	57557	
38	Hero Glamour 125	45000	2018	Individual	1st owner	21000	61600	
39	Honda CBR-250R	120000	2016	Individual	1st owner	16000	202310	
40	Hero Honda Passion PRO [2012]	30000	2011	Individual	1st owner	42000		
41	Royal Enfield Thunderbird 350	90000	2015	Individual	3rd owner	13000		
42	Bajaj Dominar 400	125000	2017	Individual	1st owner	12800		
43	Hero Glamour FI	50000	2018	Individual	1st owner	20000	71150	
44	KTM 390 Duke	175000	2017	Individual	1st owner	4600		
45	Hero Passion XPro	28000	2012	Individual	1st owner	81000	56750	
46	Yamaha FZ S V 2.0	85000	2018	Individual	1st owner	8000	84042	
47	Hero Achiever 150	29900	2016	Individual	1st owner	55000	69750	
48	Bajaj Discover 125	25000	2011	Individual	1st owner	57000	60122	
49	Yamaha FZ16	30000	2009	Individual	2nd owner	26000	78712	
50	Yamaha Saluto	60000	2018	Individual	1st owner	13000	59844	
51	Yamaha FZ16	35000	2012	Individual	2nd owner	29000	78712	
52	Bajaj Discover 100	25000	2010	Individual	1st owner	43000	42859	
53	Honda CB Trigger	30000	2015	Individual	1st owner	40000	73413	
54	Royal Enfield Electra 5 S	90000	2010	Individual	2nd owner	40000		
55	Hero Splendor PRO	35000	2015	Individual	2nd owner	24000	49900	
56	Honda Activa 5G	70000	2019	Individual	1st owner	3000	59422	
57	Hero Honda Passion Plus	25000	2006	Individual	1st owner	75000		
58	Bajaj Pulsar 150	25000	2010	Individual	1st owner	42000		
59	Royal Enfield Bullet 350	90000	2014	Individual	1st owner	13000		
60	Bajaj Discover 125	25000	2013	Individual	1st owner	55000	57549	
61	Bajaj Pulsar 150 [2001-2011]	42000	2011	Individual	1st owner	23000		
62	Hero Splendor PRO	35000	2015	Individual	2nd owner	24000	51994	
63	Hero Xtreme Sports	50000	2014	Individual	1st owner	80000	80350	

FIGURE 4.4 : DATASET 2

5.INDIAN TWO WHEELER MARKET ANALYSIS

[Indian Two Wheeler](#) Industry Analysis 2022 is the combination of the [Electric Two Wheeler](#) and ICE Two Wheeler Segment. **The Indian Two Wheeler unit in 2021 stands at about 13.7 million units which is equivalent to the sales volume of 2013 and is expected to register a 5 percent CAGR over the next 5 years.** The year 2021 wasn't really a cakewalk for the Indian two-wheeler industry as it had to navigate across various hurdles to keep up the pace with its sales volume. However, the growth was a meager 3.5% up from the year CY 2020 which proved to be fatal for the two-wheeler industry. The two-wheeler industry has always had a significant foothold across the nation and the cumulative market value stood at 1321.8 billion INR during this year.

Indian Two-Wheeler Market overview

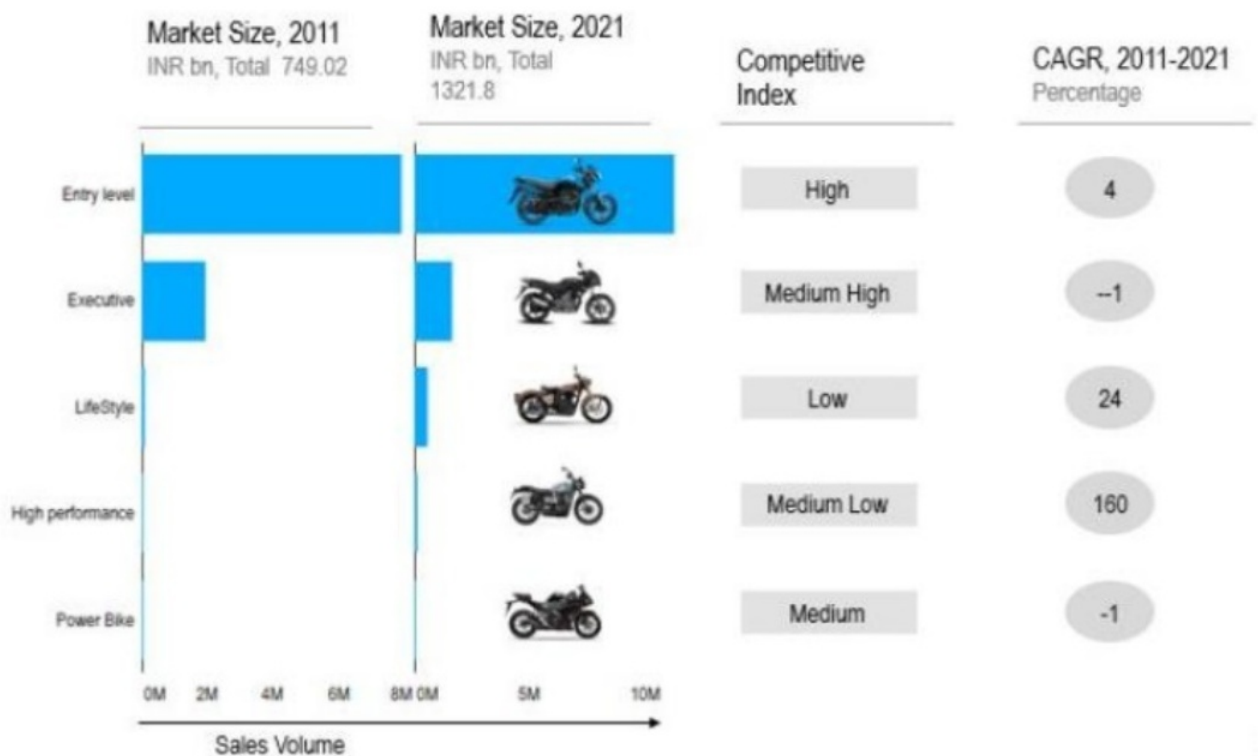


FIGURE:5.2

ACG has put forth various reasons that led to the disastrous outing for the two-wheeler industry, and the most prominent of them is the rapid and incessant shot up in customer preference to buy Electric Two Wheeler, fuel prices, massive inflation, and also the exorbitant interest rates. All these factors continued to trouble the already plagued two-wheeler industry and led to its downfall.

When we take a look at the market dynamics post-2018, we notice an astonishing fact that the scooter and the motorcycle segment are mostly in tandem with the growth curve and the overall performance in the market.

The two-wheeler segment in India has enjoyed a flawless run in the market and thanks to the wide variety of product choices it offers to its customer base.

The two-wheeler segment can be broadly classified under 5 categories i.e Economy, Executive, High performance, Lifestyle, and Power Bike. Out of all the categories listed above, the economy and the executive category have established a firm hold in the market by amassing a record number of sales in the past. Their market presence stands at a whopping 93% as far as the year 2021 is considered. While the other categories have not been able to get big in the market but have decently performed as they all witnessed good growth in the yesteryear.

Taking a look at the growth curve of the various bike categories reveals an interesting insight that the overall growth rate has been encouraging.

Competitive Landscape – Market level

OEM	Entry level (Not Price Sensitive)	Executive	LifeStyle	Power Bike	High performance
Hero					
Bajaj Auto		48%		45%	
Honda				41%	
TVS Motors		22%			
Yamaha					98%
Kawasaki					
Royal Enfield			89%		
Suzuki					
KTM					
LML					
Mahindra					
Piaggio					
Triumph					
HD					

ACG
Automotive Consulting Group

FIGURE:5.3

5.1.Road to owning a two-wheeler:

Procuring a two-wheeler for facilitating commuting needs has been the top priority for the Indian community at the moment. The buyers take into consideration various factors before zeroing in on their purchase call. The key areas of consideration are price range, ergonomics and style, economy, status and prestige symbol, emotions, ease of comfort and convenience, and most importantly the maintenance cost and safety. The weightage of these factors depends on the type of segment. Ultimately, the customer makes the final call in regards to procuring a two-wheeler. The hierarchy of the factors that influence the purchase decision will vary from person to person, and the key areas like age and income of the buyer, gender, and most critically the geographical base play a critical role in influencing the buyer. The report formulated from our end depicts the fine relation between the purchase decision of the buyer and the influence it has on the OEM's market dynamics. The report also highlights the corrective action that needs to be inculcated for being able to build on the trust of the potential buyers.

5.1.2. Key factors contributing to the Purchase decision:

The marketing and the advertisement channels do constitute a major part in getting the traction from the customer base and thereby enhance the sales volume. The brands have started focusing on unraveling the various standout features of their product in a visual mode and this in turn has enabled them to reach a wide range of audiences. For instance, the two-wheeler products like Pulsar and Apache have been periodically making the headlines by coming up with an advertisement to showcase the product's best in class power ability, seamless ride experience, and the confidence it instills in the rider. Altogether, these advertisements have emerged as a strong medium to directly engage with the potential customer base. Taking a look at the ad conceptualized by Pulsar and Apache reveals that the brands are trying their best to attract the customers to get their hands on the product and we will soon get to know whether the brands would be able to reap the best with this tactical Move. The recently concluded study has revealed a scintillating insight that the Word of Mouth(WOM) stands tall in influencing the mindset of the buyers to make purchase-related decisions, it is then closely followed by the product visibility attributes which also attracts the customers to own a brand. Both these factors cumulatively sum up to almost 75% of the influencing medium, out of which 40% is by the WOM and the rest is from the product attributes.

5.2.Impact by the surrounding culture:

Let's take a look at the most luxurious two-wheeler brand i.e the Harley Davidson in the Indian context. We get to know that Harley Davidson has evolved to be one of the key prestige and status indication in the Indian scenario. And the buyers usually get their hands on this product to enhance its popularity and traction in society than using it as a mainstream commuting option.

Taking a further keen look in regards to how the competition has turned out to be in the two-wheeler industry for the last few years reveals an astounding point that the dominance of the two-wheeler segment is segment specific, and even the segment leaders have been able to make the most of the market by maintaining a popular product chunk of not more than 2 to 3 two-wheeler models.

5.2.1.Sentiment Analysis:

It is one of the prime factors that have a good impact on the buyer, this analysis helps the brand to formulate a product strategy and a sales plan. Over the last half a decade the TVS brand has done commendably well in hitting the positive sentiment well.

5.2.2. Impact induced by the culture:

The profile of the person making the purchase decision: the two-wheeler industry has always been customer-driven. The industry has been able to quickly adapt itself to deliver to the needs of the customers, which has been a prominent reason for its flawless run in the market. Getting to understand the customer's mindset is one of the keys to success.

5.2.3. Case study analysis:

1. Ola Impact, Market share, Customer experience, and its impact on Electric Two Wheeler Industry and Overall Indian Two Wheeler segment
2. The study conducted on a customer who previously owned a Bajaj Pulsar and went on to purchase a TVS Apache has revealed the various factors that influenced the owner to make his buying preference.

6.TVS Motor Performance Study Report:

TVS is one of the oldest two-wheeler manufacturing brand in our country. TVS has earned the fame of being one of the most trusted, reputed and respected brands based in south India. The very famous and popular TVS 50 mopeds and TVS XL still continue to exist in the Indian market. But on the other hand, the then competitors of TVs such as LUNA, Hero Majestic and Suvega vanished from the Indian Moped industry a long time back which clearly proves the strong foothold of TVs in the Indian two-wheeler market. The most important reason for this success of TVS 50 is mainly because of its very quick initial pick up, good capacity to carry the load and a wide range of service networks. TVS 50 has a very good brand image in the automobile market of our country especially in south India which is evident from the fact that almost every house has a TVS 50 moped.

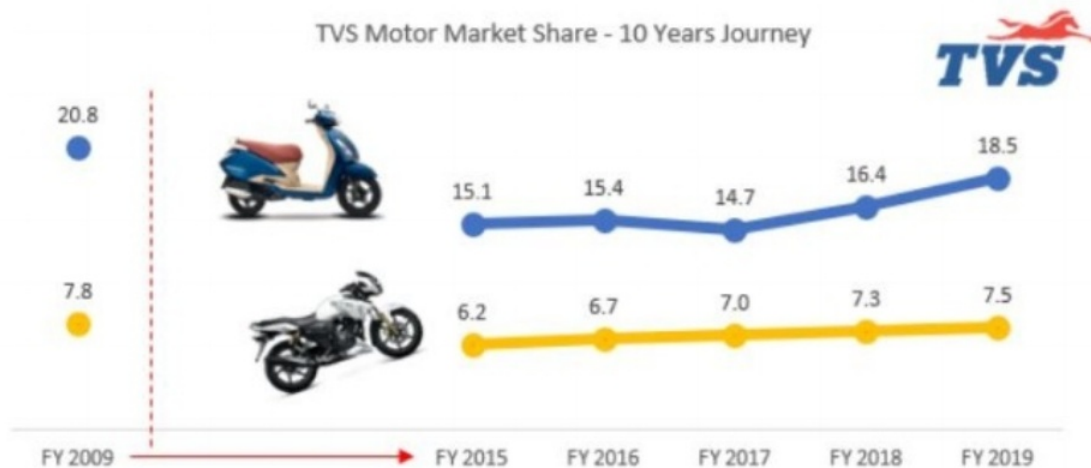


FIGURE 6.1.TVS Sales

The company has also continuously upgraded the product on a regular basis which has necessarily kept the competition at bay. In the process of constant up-gradation, the main modification that was done in the moped is the change from Dry clutch to Wet clutch along with the increase of the cubic capacity. At present, TVS has two of its products which are TVS XL Super and Jupiter. Both of these products from TVS lie within the top bestselling two-wheelers in India and is still going strong in the market.

TVS has a wide range of products starting from the economical segment to the premium ones, thereby catering to the need of almost all customers at its best. TVS lives up to the expectation of its customers by constantly upgrading its products by addressing the feedback from the customers. This is one of the major strengths of the company. The earlier model TVS 50 is currently upgraded to TVS XL Super which comes with a 4 stroke 100cc engine and kicks start but still, it continues with the centrifugal clutch wet type. All of the customers would have the feel of Moped that is non geared along with its superb initial pulling which is the main reason to why the customers do not feel the need of going for a geared Bike. There are many villages in India where the customers still find it difficult to handle the geared bikes and this is exactly where the TVS XL Super comes handy for those customers.

The Scooty is another of the very successful model from TVS which is quite popular among the female folks. This model has undergone several upgradations and finally, the Scooty Pep which has a 90cc engine is one of the largest selling models in this specific segment. After struggling a lot to get the desired volume with WEGO which is the auto gear 110cc 4 stroke scooter that TVS has brought in JUPITER by keeping the capacity of the engine same as WEGO. There was a very little overall change in looks of the model which is on par with Activa from Honda and it was very well received by the customers. When this model JUPITER has launched the model Activa from Honda was not quite easily available across which is why the customers had to wait for long. This is exactly where TVS has encashed the situation with the launch of JUPITER. Slowly and steadily, it gained a respectful share in the market in the specific segment. The overall performance of these models is satisfactory but the maintenance has always been a concern.

6.1.Bajaj Auto Picks Up Speed:

Bajaj Auto is one of the oldest automobile manufacturers of India which is more popularly known as “Hamara Bajaj”. Bajaj Auto has been ruling the scooter market of India for some decades with its household product “Bajaj Chetak”. The brand was particularly known for its trust, value as well as reliability. Basically ‘Hamara Bajaj’ summed up what Bajaj meant at that point of time. But, in the year 2001, the launch of Pulsar created a new category of common sports bikes which brought up the opportunity to deal with motorcycles. Now, the focus of the company is on motorcycle business and its well-known products that are CT, and Pulsar lies within the top 10 motorcycles which are being sold in India. The company worked in several areas for the transformation of brand perception from a scooter company to a motorcycle company and built some of the very powerful brands like Pulsar, Avenger, Dominar, V and Platina, each of them occupying its own space in the market.

The company is presently targeting to create a strong presence in the market of electric scooters or motorcycles which would certainly be an attractive **space to be in. Bajaj automobile is now looking forward to entering the** segment of an electric vehicle with new motorcycles and scooters in India. All these electric vehicle offerings by Bajaj Auto are expected to be launched in India by the year 2020.



FIGURE: 6.2.BAJAJ Sales Bar Graph

The company is presently targeting to create a strong presence in the market of electric scooters or motorcycles which would certainly be an attractive space to be in. Bajaj automobile is now looking forward to entering the segment of an electric vehicle with new motorcycles and scooters in India. All these electric vehicle offerings by Bajaj Auto are expected to be launched in India by the year 2020.

The company is aiming at about 20 to 25 percent market share as it has shown an improvement from the 15 to 18 percent market share in the year 2018. Bajaj is aiming for increasing its market share in the next 12 months to come. The 10 percent gain in the market share in the next two years would be something outstanding for Bajaj Auto.

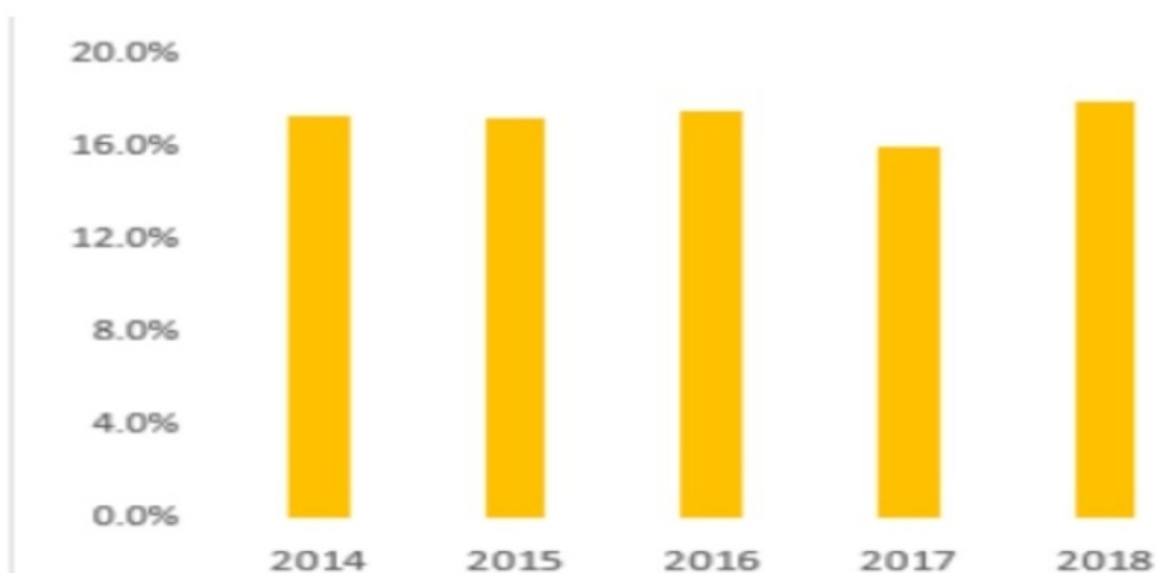


FIGURE:6.3.Profit

Over the last few years of time, the brand has shown tremendous growth and most importantly the growth is not limited to the Indian boundaries. At present, Bajaj is available in about 70 countries all over the world. But the markets keep on fluctuating thereby continuing to be uncertain at all the times. In spite of this, the company by the end of the year would have exceeded in exporting more than about 2 million units in a calendar year across to 70 countries over the world as per the export plan of the company. The exports which are talked about here would basically represent about 40 percent of the overall sales of the company.

6.2.Honda Two Wheeler Market Assessment Report:

Honda takes pride in being the second largest two-wheelers' manufacturers in our country. The company has an extremely good brand image in the Indian two-wheelers' market and hold of Honda in the country is quite strong. This is evident from the fact of the huge preference of Honda products by the Indians. Activa which is the First model Honda has gained a lot of popularity in the Indian two-wheelers' market. Something that is even better for the company is that Activa has become the household name in almost all the Indian cities as well as the villages. In the scooter segment, Activa is basically the largest selling model. Activa is very well known for its simplicity in design, trouble-free performance, great mileage and lastly a service friendly vehicle. These are the ultimate reasons as to why it is ruling the Indian market. "Unicorn" is Honda's first motorcycle in the country after its break up with "Hero". This is the first commuter bike with "Mono Suspension" in India was very well received by the consumers in India. Most importantly, "Unicorn" was the best-preferred bike in the 150cc category in the Indian market.

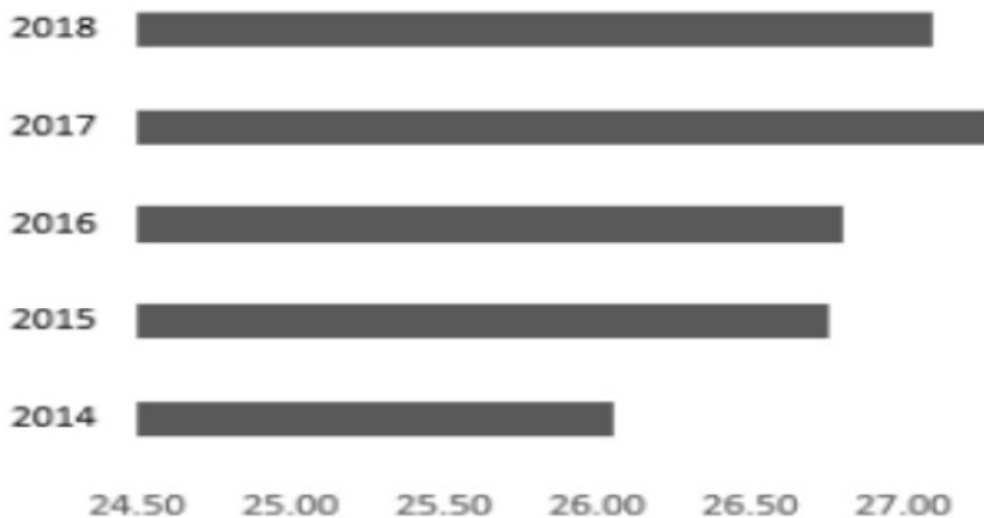


FIGURE.6.4.Honda Sales Bar Graph

Honda in its later stage has lost the strong grip on the motorcycle market of India and the only model that still lies within the top ten selling motorcycles

in the country is “CB Shine”. This model which ranked fourth in the largest selling bike in the Jan 2017 came down to the eighth position in the Dec 2018. This drop in the ranking was due to the volume of sales by about 17,543 units. However, CB shine sold more than 7 million vehicles by end of 2018. Honda has still not been able to re-grip its motorcycle business despite launching a number of other different models because of the fierce competition from Bajaj, Yamaha, and Royal Enfield. The new models Hornet and X-Blade was not able to fulfill all the expectations of the customers did not live up to the expectations of the customers. But on the other hand, other similar models with the same capacity by the other manufacturers such as Bajaj and Yamaha were able to increase their share in the market in this category. The customers had a clear idea in their mind that the Honda products are fuel efficient in nature and the ones who are in search of power, their choice would certainly be Bajaj or Yamaha. However, the ones who are looking forward to a Cruiser bike will be inclined to go for Royal Enfield.

The dealers of Honda are very well equipped in terms of infrastructure as well as manpower. Almost all the dealers are permitted to open their own outlet within the allotted territory for the purpose of extending the reach and each of the branches is running as a separate profit center. Adequately trained and vastly experienced manpower is available throughout the branches along with a centralized customer support team who are in the main office taking full control of all the grievances of the customers. This makes it quite clear that the Customer Satisfaction Index or the CSI process is implemented at its best and working in its full swing. Most importantly, the overall process under constant monitoring from the head office. Despite this, there still exists some unsatisfied customers of Honda. One of the main reasons behind this might be due to the fact of huge volumes where the reach is still not a hundred percent.

The relation between the company and dealership is completely in the business terms where everything is simply cash and carry. Honda expects all its dealers to completely adhere to the policies and guidelines of the company. However, in case of any kinds of deviation will result in the termination of the dealership. Now, due to a very high volume of sales, the dealers are at the receiving end thereby making themselves self-aligned with the policy which in turn gives highly successful results.

7.Interaction among visuals in project:

By default all visuals in power bi report page are interactive with each other that means, If you select any value in slicer it filters all other visuals in the report page, in similar way if you select any item in chart visuals that will affect the value of other visual too.

7.1.PIE CHART:

selling_price and First owner by bike model

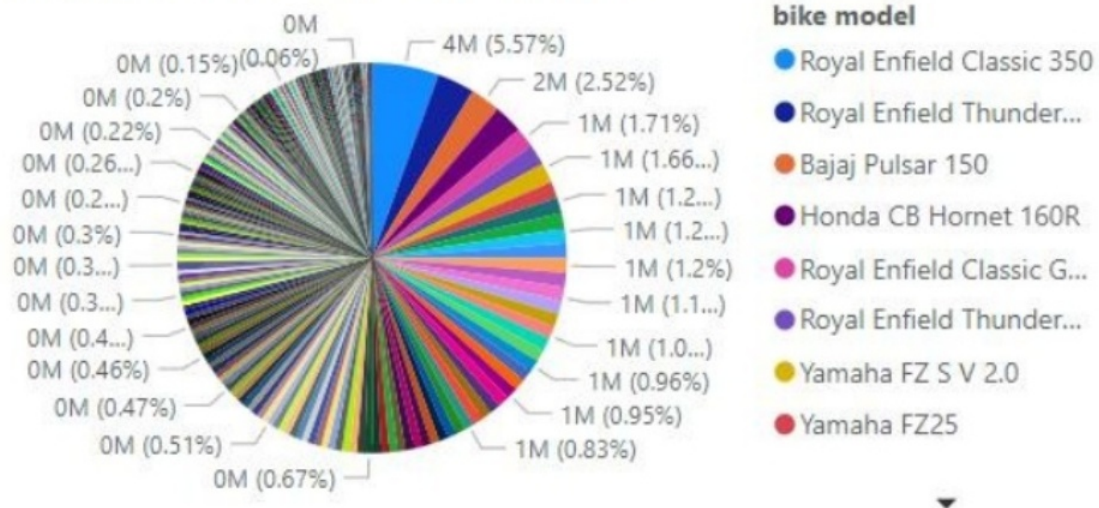


FIGURE.7.1.Sales in pie chart

Pie charts can be an excellent addition to your report. They will show a quick overview of your data, and act as an eye-catching visual accent. That being said, you need to know when to use them, and how to use them correctly for them to be a valuable part of your visual narrative, not a detriment.

7.2.Count of bike model by owner:

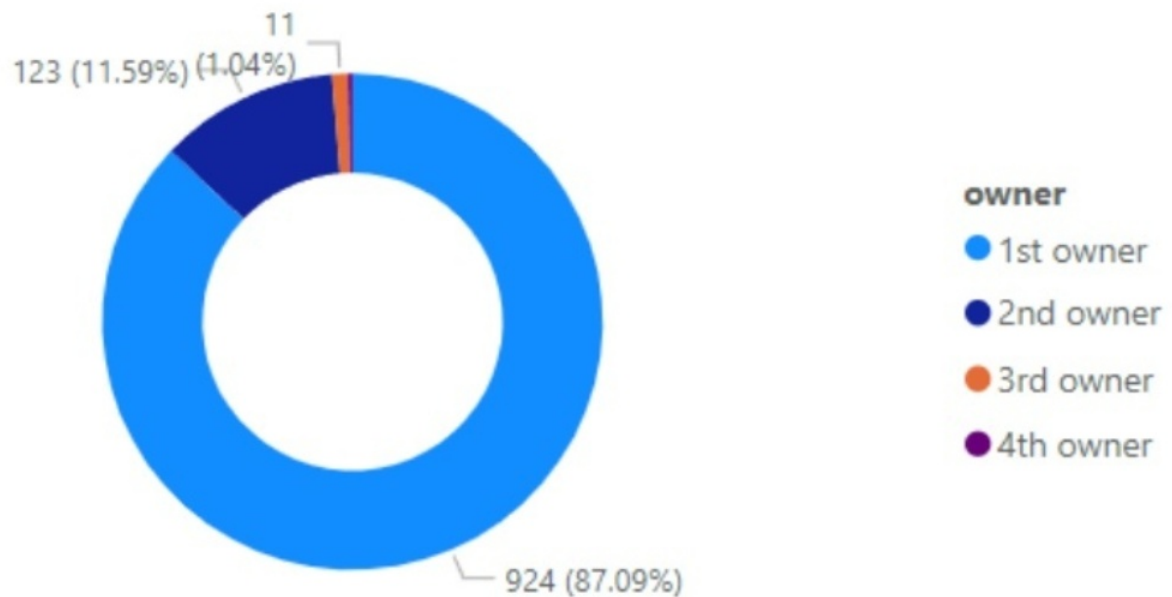


FIGURE.7.2.Representation of owners in donut chart.

A Donut chart demonstrates the relationship of parts to an entire. Also, a Donut chart can contain more than one information arrangement.

Donut Charts are best used to compare a particular section to the whole, rather than comparing individual sections with each other.

The **sum of the Donut chart values** must add up to 100%. **Like Pie charts, the formula of a Donut chart is = (given data/ total data)*360.**

7.3.SLICER:

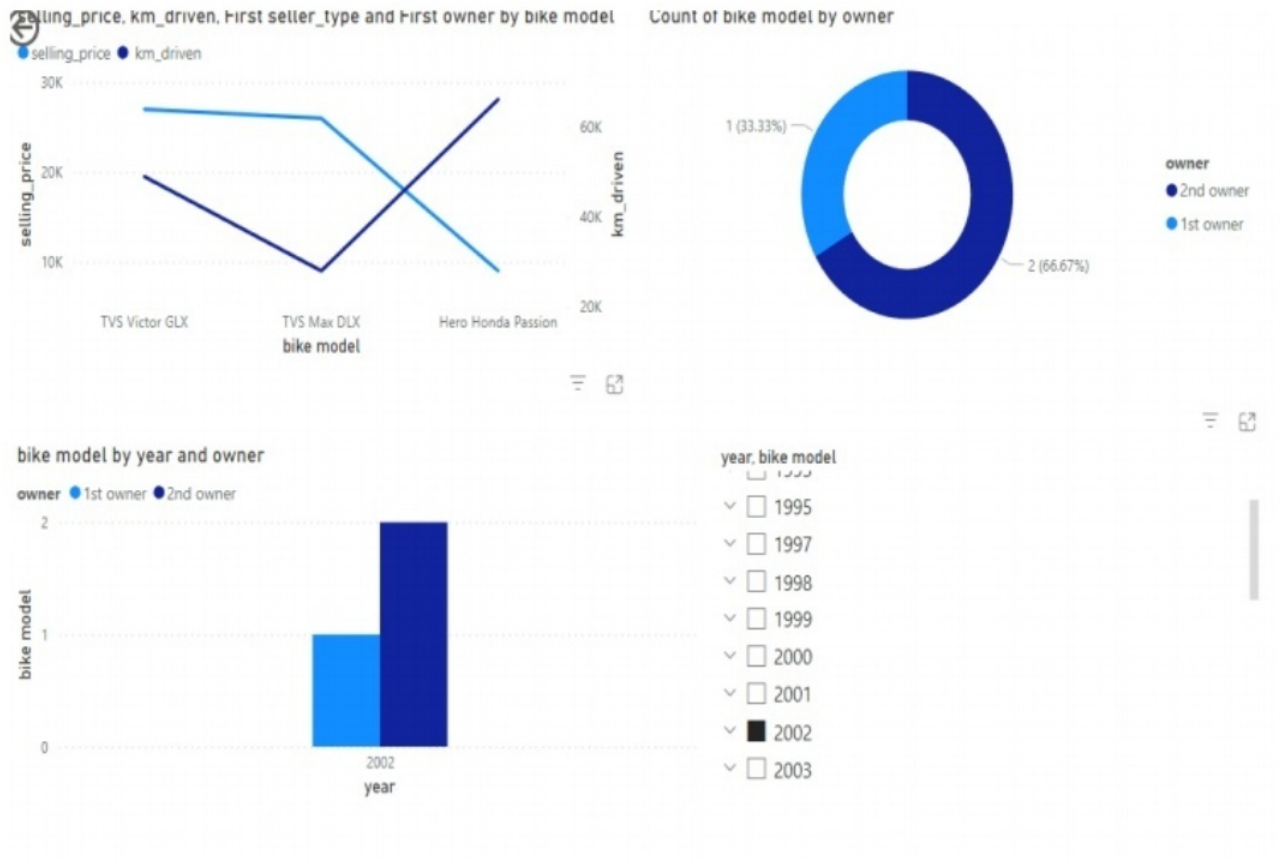


FIGURE.7.3.Data in Slicer

Slicers in Power BI are a type of on canvas visual filters. The slicers enable a user to sort and filter a packed report and view only the information they want. Unlike filters, the slicers are present as a visual on the report and let a user select values as they are analyzing the report. For instance, if you have a sales analysis report, you can make a slicer for years. From that slicer, you can select the year for which you want to see the sales metrics. The report visuals will automatically change showing the information for that year. We will understand slicers with more such examples in the sections to follow.

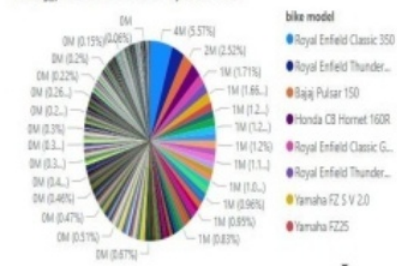
In the image of the report shown below, the visual indicated in the red box is a slicer. It contains the name of district managers. You can sort the information on district monthly sales by selecting the names of managers from the slicer.

7.4. Over all representation of data in visuals:

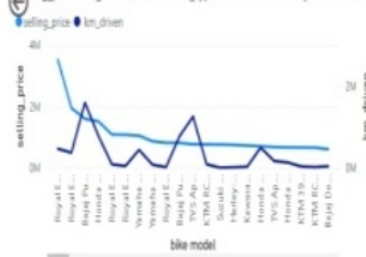


bike model	owner	seller_type
Activa 3g	1st owner	Individual
Activa 4g	1st owner	Individual
Aprilia SR 125	1st owner	Individual
Bajaj ct 100	1st owner	Individual
Bajaj Avenger [2015]	1st owner	Individual
Bajaj Avenger [2015]	2nd owner	Individual
Bajaj Avenger 150	1st owner	Individual
Bajaj Avenger 150 street	1st owner	Individual
Bajaj Avenger 220	1st owner	Individual
Bajaj Avenger 220 dtsi	1st owner	Individual
Bajaj Avenger Cruise 220	1st owner	Individual

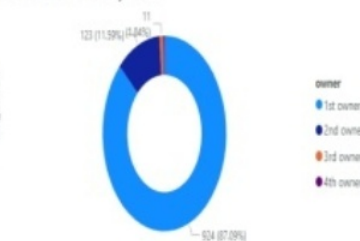
selling_price and First owner by bike model



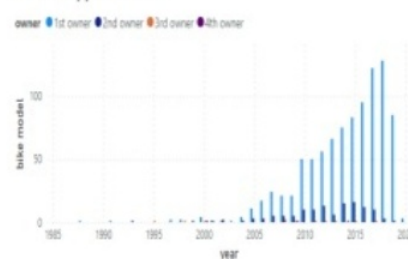
selling_price, km_driven, First seller_type and First owner by bike model



Count of bike model by owner



bike model by year and owner



year bike model



FIGURE.7.4.Sales represented in visuals

8.CONCLUSION:

In this project, the proposed methods for prediction of bike buyers using four machine learning techniques that were used include SVM, Random Forest, Decision tree, KNN. The system is implemented by using all the models and evaluates their performance. Performance evaluation is completely based on different performance metrics. Random Forest Classifier resulted highest accuracy of 0.86. From the above results Random Forest plays a key role in shaping improved classification accuracy of a dataset. To extend the project as to predict bikes using IOT, in that all bike stations can be connected over the internet using IOT. The sensors can be employed which will collect the information about the current bike stations. This is the future scope of project. I hope this implementation of LSTM to predict Capital Bike share demand was interesting to the readers. It helped me implement LSTM with Keras and strengthened my knowledge of Pandas.

Conclusion Sales forecasting is a critical part of the strategic planning process and allows a company to predict how their company will perform in the future. It allows them to not only plan for new opportunities, but also allows them to avert negative trends that appear in the forecast. A mission statement is important because it allows an organization to know exactly why they exist and serves as a guide for decisions. Both concepts are important to the success of the company and should not be overlooked throughout the strategic planning process.

9.Reference:

[1] K. Gebhert, R. Roland, "Impact of Weather on Capital Bike Share Trips," presented at the 92nd Annual Meeting of the Transportation Research Board 2013.

[2] J. Yoon, Fabio Pinelli, "Cityride: a predictive bike sharing journey advisor" 2012 IEEE 13th International Conference on Mobile Data Management.

[3] R. Giot, R. Cherrier "Predicting Bike Share Demand upto One hour ahead" 2013 IEEE 9th International Conference on Data Management, France.

[4] I. Frade, A. Ribbero, "Bicycle Sharing Systems Demand" unpublished.

[5] T. Rui, Lin Li Hua "Quantitative Research on Vehicle Exhausts Pollution in the City" Published in 2012.

[6] Y. Zhang, Z Huang "Performance evaluation of bike sharing system in Wuchang area of Wuhan, China", 6th China Planning Conference (IACP), 2012.

[7] Kaggle: Bike Share Demand,
"<https://www.kaggle.com/c/bike-sharing-demand>".