**CONSUMER BEHAVIOUR TOWARDS BUYING AN ELECTRIC VEHICLE**

**A Case Study**



**University Of Calcutta**

**UNIVERSITY ROLL NO.:** 213026-11-00-57

**UNIVERSITY REGISTRATION NO.:**  026-1211-0130-21

**PAPER CODE:** DSE B2

**ABSTRACT**

With the current depletion of fossil fuels and its price hike, there is a need for another energy resource to run the vehicle. The automobile sector is considering Electric Vehicles as a solution to the industry and environment in India. However, the current market penetration of EV is relatively low in spite of governments implementing Electric vehicle policies. Through this project, awareness level about Electric vehicles and factors influencing consumers to consider purchasing Electric vehicles will be studied and consumer response for the same will be analyzed. The research will be based on primary data. In this project a survey was conducted among **448** individuals. The data was collected through a detailed questionnaire. Since the youth are the future of the nation, therefore their opinion matters and hence the target for the data collection would be youth of the age group 18-30 years old.

1. ***Objective of the study***:

* To understand the buying behavior of consumers towards electric vehicles.
* To recognize the important parameters affecting consumers’ vehicle purchasing decisions.
* To understand the growth of electric vehicles industry in India.
* To recognize consumers’ perception of internal combustion engines.

1. ***Problems :***

The growth of electric vehicles is very slow in India, and not many electric vehicle industries are coming to India for setup. Consumers who all are already having vehicles are not willing to switch to an electric vehicle immediately. The major factors that deters consumers from buying an EV is long recharging time , mileage of vehicle , lack of charging infrastructure , replacement of battery is expensive.

***Keywords :*** Electric vehicle , depletion of fossil fuels , environment , government policies , market penetration , awareness level , consumers’ perception , purchasing decisions , buying behavior.

**CONTENTS**

1.Introduction……………………………………………………………………………………………. 7- 14

1.1 Rationale of the Study

1.2 Introduction to Industry

1.3 Introduction to the companies

2.The Dataset……………………………………………………………………………………………… 15 - 22

3.Methodology…………………………………………………………………………………………… 23 - 57

3.1 Statement of the problem

3.2 Objective of the study

3.3 Area of the Survey

3.4 Sampling techniques

3.5 Data Collection

3.6 Software and programming. language used

3.7. Explanatory Data Analysis

3.7.1. Diagrammatic representation

3.8. Chi square test of association

3.8.1. Test statistic

3.8.2. Interpretation of our data based on test

3.8.3.Inference

3.9. Analysis of variance(ANOVA)

3.9.1.Definition

3.9.2. Use of one way ANOVA

3.9.3. Assumptions of ANOVA

3.9.4. Test based on one way ANOVA

3.9.5. Fisher Individual Test for difference of means

3.10. Multiple Logistic Regression

3.10.1. Assumption

3.10.2. Logit transformation

3.10.3. General model

3.10.4. Removing outliers, influential and leverages

3.10.5.Interpretation

3.11. ROC-AUC curve

3.12. Principal Component Analysis

3.12.1. Basic terminologies of PCA in ML

3.12.2. How does PCA work ?

3.12.3. Types of PCA plot

3.12.4 Interpretation

4. Conclusion …………………………………………………………………………………………… 58

5.References ……………………………………………………………………………………………. 59

6.Acknowledgement ………………………………………………………………………………… 60

**List of tables**

**Table-1:** First six data points

**Table -2:** Contingency table of Gender with respect to preference of type of vehicle

**Table-3:** Contingency table of occupation with respect to gender

**Table-4:** Contingency table of Battery replacement of EV is expensive with next vehicle will be EV

**Table-5:** Contingency table of cost to buy an EV is same as CNG/Fuel Vehicle with next vehicle will be EV

**Table-6:** Contingency table of occupation with EV is expensive

**Table-7:** Contingency table of age with next vehicle will be EV

**Table-8:** Contingency table of influenced by gov. incentives with next vehicle will be EV

**Table-9:** Contingency table of influenced by gov. incentives with switch to EV

**Table-10:** Contingency table of residential area with switch to EV

**Table-11:** Anova table

**Table-12:** Anova: Single Factor (summary)

**Table-13:** Anova table for Single Factor

**Table-14:** Table for Fisher Individual Tests for Differences of Means

**Table-15:** Eigen analysis of the Correlation Matrix

**Table-16:** Principal Component Loadings Table

**List of figures**

Figure 1: Picture of Electric Car

Figure 2: Picture of different types of Electric Vehicle

Figure 3: Pie Chart for gender

Figure 4: Bar plot for Age Vs. Gender

Figure 5: Pie Chart for residential area

Figure 6: Bar plot for Occupation

Figure 7: Pie Chart for Family’s Annual Income

Figure 8: Bar chart of taxes

Figure 9: Bar chart of Type of vehicles respondents have/had

Figure 10: Pie Chart for Educational Qualification

Figure 11: Bar chart for preference of vehicles

Figure 12: Barplot of Gender vs. Preference of vehicle

Figure 13: Pie Chart for number of people own an EV

Figure 14: Bar Chart for knowledge about EV

Figure 15: Bar chart for opinion about reduction of Global warming

Figure 16: Bar Chart for opinion about EV better than regular Cars

Figure 17: Bar chart for opinion about EV can save a lot of money

Figure 18: Bar Chart for opinion about EVs are expensive

Figure 19: Bar Chart for opinion about EV costs aboutthe same to buy a Fuel/CNG vehicles

Figure 20: Bar chart for government incentives influenced decision to buy EV

Figure 21: Pie Chart for Next vehicle will be EV

Figure 22: Pie Chart for switch to EV

Figure 23: ROC curve for model 1

Figure 24: ROC curve for model 2

Figure 25 : Scree Plot

Figure 26 : Score Plot

Figure 27 : Loading Plot

**1: INTRODUCTION**

* 1. ***Rationale of the Study: -***

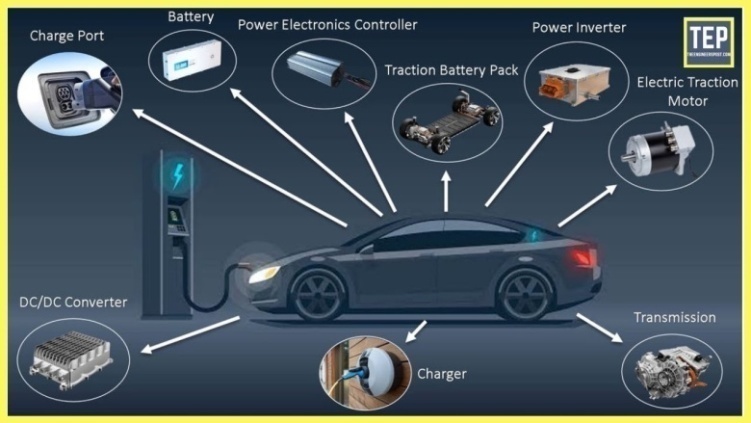
**“The time is right for electric cars – in fact the time is critical.”**

**---- Carlos Ghosn**

Although the rapid progress of the global economy and technology has advance human civilization, it has also caused tremendous damage to the global ecological environment. Therefore, humans are thinking seriously about the environment and its sustainable development. One of the solutions to these environmental problems is new energy vehicles.

From an energy perspective, more abundant energy sources for vehicles will improve the reliability and balance of energy consumption. Coupled with the intelligent development of electric vehicles, traffic status and road usage will be significantly improved. The IEA (2017a) has indicated that, based on vehicle fuel cycle calculations, electric passenger vehicles in Europe in 2015 emitted 50%less carbon dioxide than gasoline vehicles and 40% less carbon dioxide than diesel vehicles. When emissions related to vehicle manufacturing are considered, carbon dioxide emissions are reduced.

However, Ellingsen et al, clearly stated that, considering the full life cycle of vehicles (manufacture, usage, and scrap), Under the current European electricity production structure, pure electric vehicles can reduce greenhouse gas emissions by about 30% compared to internal combustion engine vehicles. For countries with carbon-intensive power production structures (such as India), with the full life cycle of vehicles considered, the reduction may even be greater. With the dual pressure of resource reduction and environmental changes, electric vehicles will become the mainstream development trend of the future automotive industry. Therefore, it is a critical issue to develop low-carbon, energy-saving, and intelligent electric vehicles to reduce environmental impact. This paper studies consumers’ opinions of electric vehicles in an uncertain environment, and analyzes the factors influencing consumers’ acceptance of electric vehicles, in order to improve the penetration of electric vehicles into the market and to provide reference suggestions for future researchers.

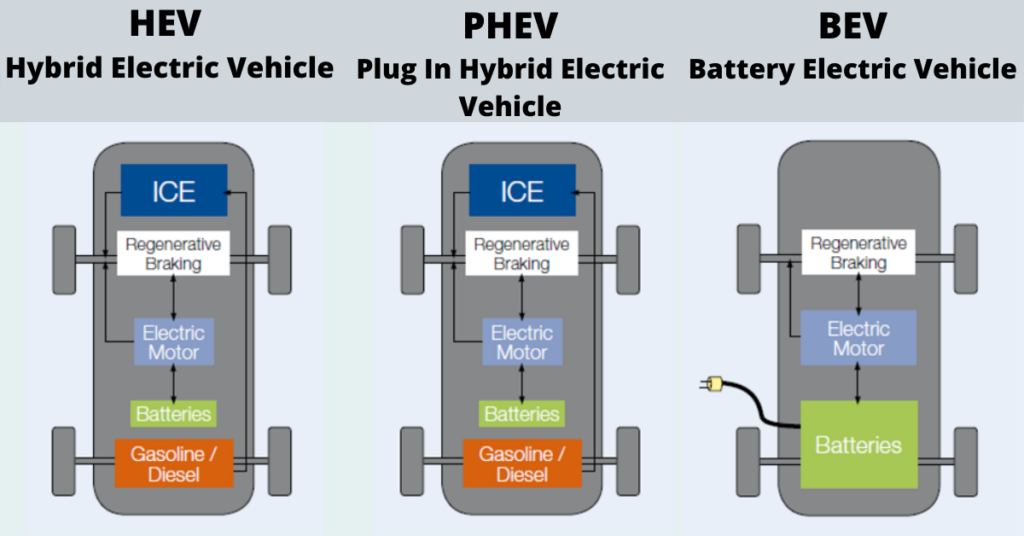


**Fig. -1 : Picture of Electric Car**

* ***Types of Electric Vehicle:---***

When we talk about electric vehicles, we usually refer to three types:

1. Hybrid electric vehicles (HEV)
2. Plug-in hybrid electric vehicles (PHEV) and
3. Battery electric vehicles (BEV)



***Fig. 2: Picture of different types of Electric Vehicle***

* ***Hybrid electric vehicles (HEV) :-***

The hybrid electric vehicle has both a conventional internal combustion engine (ICE) and an electric-powered propulsion system. HEVs use both petrol and electricity as power, HEVs startup using electric motor then the petrol engine takes over as the speed rises.

In this the batteries are charged by its regenerative braking system, this system is a mechanism that recovers the energy by slowing down a moving vehicle, which converts kinetic energy into electrical energy, which is stored in batteries or super-capacitors, or the energy is directly used by the motor in some vehicles.

Both the Internal combustion engine and electric motor are controlled by computer software which ensures the best fuel economy and leads to fewer tailpipe emissions as compared to traditional vehicles with an internal combustion engine.

* ***Plug-in hybrid electric vehicles (PHEV) :-***

We can say Plug-in Hybrid Electric Vehicle is a combination of both the Hybrid electric vehicle and battery electric vehicle, it combines an internal combustion engine with a battery-powered motor like a hybrid vehicle whose batteries can be charged by plugging into an external source of electric power.

The plug-in hybrid vehicle also uses the regenerative braking system for recharging batteries with the plugging in an external source of electricity. Because of the plug-in rechargeable system, the operating cost of PHEVs reduces as charging from the grid might cost less than charging of batteries from an onboard generator.

PHEVs can run longer only on electricity when charged full than HEVEs which extend the range of the vehicle which ultimately leads to fuel economy and less emission by a vehicle.

* ***Battery electric vehicles (BEV):-***

BEVs battery electric vehicles are also called all-electric vehicles or only EVs or pure electric vehicle. The electric vehicle has onboard batteries whose power is used to operate the electric motor with no other secondary fuel engine (internal combustion engine).

The vehicle store the electricity in chemical form in high-capacity batteries or supercapacitors present onboard thus not required any fuel tank or fuel cell. Unlike the traditional internal combustion engine, BEVs runs with **zero-emission** and no other harm to the environment.

The range of an all-electric vehicle depends on the vehicle but the average range of vehicle available in the [market](https://en.wikipedia.org/wiki/Market_(economics)) is 200 km and few luxuries and performance vehicle has a range of 450 km.

BEVs completely powered by batteries onboard, which are charged from external sources with the help of special chargers, and the charging speed is dependent on the type of charger. The chargers are classified by the speed of charging.

Generally, there are three types of **EV chargers**available are, level 1, level 2, level 3, or DC fast charger. 1) Level one takes almost 8 to 10 hours to fully charge an EV, and it uses standard 120volts outlet to charge. 2) Level 2 chargers require special EV charging stations that can provide 240volts, which can charge an EV within 4 to 6 hours. 3) Level 3 or DC fast charger also requires a special charging station, and this fast charger can charge an EV up to 80% to 90% in about 4 hours.

* 1. **Introduction to the Industry *:-***

Of all the oil consumed in the World, 20 percent is used for transportation. Further, passenger vehicles use 70 percent of transportation oil. By 2050,there may be as manyas 1.5 billion carson the road, compared to 750 million in 2010.This type of demand represents both a challenge and an opportunity to capitalize on new vehicle technologies, and in the process, reap substantial economic development benefits. In a world where oil is a limited resource, an alternate source of transportation fuel –electricity –is not only a smart investment, but as some would say, it is an inevitable one. Further, the switch to electric vehicles will generate demand for existing jobs and create new jobs as well. As study after study confirms, job growth in electric vehicle industries will outweigh any reduction of jobs in traditional fuel industries, resulting in net job growth. Electric vehicles create additional economic development opportunities by improving quality of life, reducing energy spending, and decreasing reliance on foreign oil.

Now let’s take a step back to know where this electric vehicle comes from, Electric vehicles technology came into existence in 1837 in Scotland by Robert Anderson. He made the first electric crude vehicle which works on both electric motor and crude engine. After that many scientist performed different research and invented different electric vehicle. Each vehicle would have more improvement compared to previous electric vehicle. Scientist made improvements in battery technology, electric motor which improved efficiency and speed of vehicle day by day. Abstract Electric mobility revolution is considered as biggest revolution for automobile industrial sector. This revolution came into existence due to depletion of crude oil, fossil fuels, natural gas etc. To overcome this crisis many technologies have been developed by different nations in our world. In India electric vehicles was introduced in the year 2011 by Mahindra electric mobility Ltd. And Revai electric company car i.e. Revai electric car. The global automotive industry is undergoing a paradigm shift at present in trying to switch to alternative/less energy intensive options. India, too, is investing in this electric mobility shift. The burden of oil imports, rising pollution, and as well as international commitments to combat global climate change are among key factors motivating India’s recent policies to speed up the transition to e-mobility.

The Indian automotive industry is the fifth largest in the world and is slated to be the third largest by 2030. Catering to a vast domestic market, reliance on the conventional modes of fuel intensive mobility will not be sustainable. In an effort to address this, federal policymakers are developing a mobility option that is “Shared, Connected, and Electric” and have projected an ambitious target of achieving 100 percent electrification by 2030.By making the shift towards electric vehicles (EVs), India stands to benefit on many fronts: it has a relative abundance of renewable energy resources and availability of skilled manpower in the technology and manufacturing sectors.

The FAME India Scheme is an incentive scheme for promotion of electric and hybrid vehicles. It aims to promote electric mobility and gives financial incentives for enhancing EV production and the creation of electric transportation infrastructure. In 2015 the Ministry of Heavy Industries and Public Enterprises launched FAME to incentivize the production and promotion of eco-friendly vehicles including EV and hybrid vehicles. The scheme is proposed for establishing charging infrastructure) . The National Electric Mobility Mission Plan (NEMMP) 2020, a National Mission document providing the vision and therefore the roadmap for the faster adoption of EVs and its manufacturing. This plan has been designed to boost national fuel security, to supply affordable and environmentally friendly transportation, and to enable the Indian automotive industry to attain global manufacturing leadership.

* 1. **Introduction to the Companies :-**

Regardless of the country’s ambitious targets, India’s EV space is at a nascent stage. It is less than 1% of the total vehicle sales however has the potential to grow to more than 5% in a few years. At present there are more than 5 lac electric twowheelers and few thousand electric cars on Indian roads. The industry volumes have been fluctuating, mostly depending on the incentives offered by the government. However, looking at it differently – India offers the world’s largest untapped market, especially in the two-wheeler segment. 100 percent foreign direct investment is allowed in this sector under the automatic route.

Responding to this opportunity, leading players like OLA Electric Mobility Pvt, Ather Energy, and Mahindra Electrics are rapidly growing their market presence. Moreover, certain states like Karnataka and Tamil Nadu are rolling out innovative and timely investor-friendly policies besides building necessary infrastructure. Recently, the American electric vehicle and clean energy company Tesla Inc. marked its entry into India by incorporating its subsidiary, Tesla India Motors and Energy Pvt Ltd, in Bangalore.

In February 2021, Ather Energy, India’s first intelligence EV manufacturer moved its US$86.5 million factory from Bangalore (Karnataka) to Hosur (Tamil Nadu). Ather Energy’s factory is said to have an annual production capacity of 0.11 million two-wheelers.

This week, Ola Electric, the subsidiary of the unicorn Indian ride-hailing start-up, also announced that it would be setting up the world’s largest electric scooter plant in Hosur (which is a two and a half-hour drive from Bangalore over the next 12 weeks, at a cost of US$330 million, and aiming to produce 2 million units a year. By 2022, Ola Electric wants to scale up production to pump out 10 million vehicles annually or 15 percent of the world’s e-scooters.

There have also been positive developments in the expansion of charging infrastructure across the country – states like Andhra Pradesh, Uttar Pradesh, Bihar, and Telangana are setting impressive targets for the deployment of public charging infrastructure to increase uptake of electric vehicles in the country.

Recently, Sterling and Wilson Pvt. Ltd (SWPL), India’s leading engineering, procurement, and Construction Company announced its entry into the electric mobility segment in India. It has signed a 50-50 joint venture with Enel X, to be incorporated on April 1, 2021, to launch and create innovative charging infrastructure in India.

**2. The DATASET**

A simple still reasonable connection is comprehended between environment and consumer behaviour towards buying an EV among people of age below 18-30 above in our dataset. Here the dataset has been tactfully divided between two genders - Female and Male. The questionnaire was formed keeping in mind the likes and dislikes form the potential customer of electronic vehicle and their willing to change to electronic vehicle.

**The questionnaire was divided into three parts**

•**Section 1:** The personal information of the respondent.

•**Section 2 :** Awareness , usage patterns and perception towards EV.

•**Section 3 :** Weather the respondent is willing to switch to an electronic vehicle or not and the reason.

The first few data points of our data set are tabulated below:

**Table-1: First six data points**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Name :** | **What's your**  **age?** | **What's your gender ?** | **What's your occupation?** | **Highest educational qualification?** |
| Ankana Ghosh | 18-22 | Female | Student | Graduate |
| Joydeep Pan | 23-30 | Male | Others | Diploma |
| Riddhima | 18-22 | Female | Student | Graduate |
| Saptarsi Bhattacharjee | 23-30 | Male | Others | Post Graduate |
| Poushomi Chatterjee | 18-22 | Female | Student | Graduate |
| Tanupriyo Das | 18-22 | Male | Student | Higher Secondary |
| **What is your family's residential locality?** | **What's your yearly family income (in Lakh)?** | **Do you pay taxes?** | **Which type of vehicle do you have/had?** | **What do you prefer?** |
| Rural | 9Lakh-15Lakh | Yes | Electric Vehicle, CNG / Fuel Vehicle, Pedal - driven Vehicle | Electric Vehicle |
| Rural | 9Lakh-15Lakh | Yes | CNG / Fuel Vehicle | Electric Vehicle |
| Urban | 2Lakh-8Lakh | Yes | CNG / Fuel Vehicle | CNG / Fuel Vehicle |
| Rural | 9Lakh-15Lakh | Yes | Electric Vehicle, CNG / Fuel Vehicle, Pedal - driven Vehicle | Electric Vehicle |
| Urban | 2Lakh-8Lakh | Yes | CNG / Fuel Vehicle | CNG / Fuel Vehicle |
| Rural | Less than 2 Lakh | No | CNG / Fuel Vehicle, Pedal - driven Vehicle | Electric Vehicle |
| **How would you rate your knowledge about Electric Vehicle?** | **Do you own an electric vehicle or have you ever used it?** | **How long have you owned or used an electric vehicle?** | **Which type of electric vehicle do you have ?** | **Which type of electric vehicles would you like to buy if you don't have an electic vehicles?** |
| Good | Yes | Less than 5 years | Electric Scooty |  |
| Good | Yes | Less than 5 years |  | E - Bikes, Electric Scooty |
| Terrible | No | I didn't use / own an electric vehicle |  |  |
| Good | Yes | Less than 5 years | Electric Scooty | Electric Scooty |
| Good | No | I didn't use / own an electric vehicle |  | Electric Scooty |
| Average | No | I didn't use / own an electric vehicle |  | E - Bikes |
| **From which source do you get the most of your knowledge about electric vehicles?** | **What do you think about the following statements about electric vehicles? [Can reduce global warming]** | **What do you think about the following statements about electric vehicles? [Better than regular cars in terms of satisfying consumer needs]** | **What do you think about the following statements about electric vehicles? [Can save a lot of money to the owner]** | **What do you think about the following statements about electric vehicles? [Are very expensive]** |
| Newspaper, Television advertisement, Social Media ads, Friends, Hoardings and poster | Strongly Agree | Agree | Strongly Agree | Strongly Disagree |
| Newspaper, Television advertisement, Social Media ads, Friends, Hoardings and poster | Strongly Agree | Agree | Strongly Agree | Disagree |
| Social Media ads | Agree | I don't know | Agree | Strongly Agree |
| Newspaper, Friends | Strongly Agree | Agree | Strongly Agree | Disagree |
| Newspaper, Social Media ads, Friends | Strongly Agree | I don't know | Agree | I don't know |
| Television advertisement, Social Media ads | Strongly Agree | I don't know | Agree | Agree |
| **What do you think about the following statements about electric vehicles? [Maintenance and infrastructure is well developed]** | **What do you think about the following statements about electric vehicles? [The cost to charge an Electric Vehicle is much less than the fuel costs for a petrol or diesel vehicle]** | **What do you think about the following statements about electric vehicles? [Costs about the same to buy a petrol or diesel vehicles]** | **How to increase the mileage of an electric vehicles?** | **Is battery replacement of Electric Vehicle expensive?** |
| Strongly Agree | Strongly Agree | Strongly Disagree | Increasing power of battery, Reducing the weight of the vehicle, Not applying the breaks very frequently, Electric cars are most efficient when driven at around 50-60 mph | Yes |
| Agree | Strongly Agree | Strongly Agree | Increasing power of battery, Reducing the weight of the vehicle, Not applying the breaks very frequently, Electric cars are most efficient when driven at around 50-60 mph | Yes |
| Agree | I don't know | I don't know | Reducing the weight of the vehicle | Yes |
| Agree | Agree | Disagree | Increasing power of battery, Reducing the weight of the vehicle, Not applying the breaks very frequently | Yes |
| Agree | Agree | I don't know | Increasing power of battery | Maybe |
| Agree | Agree | I don't know | Increasing power of battery | Maybe |
| **Is the performance of an Electric Vehicles same as Internal Combustion Engine (ICE)?** | **Does electric vehicle needs PUC (Pollution Under Control) certificate ?** | **Have government incentives influenced your decision to purchase an electric vehicle?** | **How likely that your next vehicle will be an electric vehicle?** | **What factor encourage you to consider electric vehicles?** |
| Maybe | No | Yes | I want to buy an electric vehicle as soon as possible | Price, Positive environmental effect, New trends, Beneficial financial or insurance option, Low noise level, Promotion, References, Petrol price hikes |
| Yes | No | Yes | I want to buy an electric vehicle during next 5 years | Price, Positive environmental effect, New trends, Beneficial financial or insurance option, Low noise level, Promotion, Petrol price hikes |
| Maybe | Yes | No | I don't want to buy an electric vehicle | New trends, Low noise level, Petrol price hikes |
| No | No | No | I want to buy an electric vehicle during next 5 years | Price, Positive environmental effect, Low noise level |
| Maybe | Yes | Maybe | I want to buy an electric vehicle during next 10 years | Price, Low noise level, Petrol price hikes |
| Yes | No | No | I want to buy an electric vehicle during next 5 years | Positive environmental effect, Low noise level, Petrol price hikes |
| **What factor discourage you to consider electric vehicles?** | **If you already have an electric vehicle, it is of which company?** | **If you want to buy an electric vehicle, then which company will you prefer?** | **Which company's electric vehicle will you suggest to others?** | **Will you switch to an electric vehicle if the company offers an exchange value on your owned vehicle to buy an electric vehicle?** |
| Long recharging time, Lack of charging infrastructure, Mileage of vehicle | Hero | Tata, Hyundai, Mahindra, Hero, TVS | Tata, Hyundai, Mahindra, Hero, TVS | Yes |
| Long recharging time, Mileage of vehicle |  | Tata, Hyundai, Mahindra, Hero, TVS | Tata, Hyundai, Mahindra, Hero, TVS | Yes |
| Long recharging time, Price, Lack of trust to new technologies, Price of electricity, Mileage of vehicle |  |  | Tata, Hyundai, Hero | Maybe |
| Lack of charging infrastructure, Mileage of vehicle | Hero | Tata, Hero, TVS | Tata, Hero, TVS | Maybe |
| Long recharging time, Lack of consumer choice, Lack of trust to new technologies, Mileage of vehicle |  |  |  | Maybe |
| Price |  | Tata | Revolt | Yes |
| **On a scale of 1 to 5, how would you rate your experience with my survey?** |  |  |  |  |
| 5 |  |  |  |  |
| 5 |  |  |  |  |
| 4 |  |  |  |  |
| 5 |  |  |  |  |
| 4 |  |  |  |  |
| 5 |  |  |  |  |

Since our dataset is basically Primary Data, a realistic assumption can be inferred very easily. out of **total 448 data points**; **182 are female** variables , **266 are male** variables.

**A small-scale idea about the variables are given below:**

1. **What is your Age? :** Age of a person can also influence their vehicle preference.
2. **What is your Gender? :** To understand more acutely about the choice of vehicles , we divide the data into male and female.
3. **What's your occupation? :** Buying a vehicle is expensive so in that case occupation gives us an idea about the choice of vehicle of the consumerbased on their occupation
4. **Highest educational qualification? :** It plays an important role regarding the knowledge about EV and environmental pollution .
5. **What is your family's residential locality? :** It plays an important role regarding the facility of buying an EV or charging station of EV,etc.
6. **What's your yearly family income (in Lakh)? :** Family income is important for buying a vehicle. This gives us an idea about the choice of vehicle of the consumer.
7. **Do you pay taxes? :** It is a yes/ No type question**.**We know that if the annual income is over Rs. 2,50,000 in a financial year then he/she have to pay taxes . So , this is a cross question for checking their annual income is whether above or below the 2,50,000 /- or not.
8. **Which type of vehicle do you have/had? :** I get an idea about which type of vehicle they have/ had.
9. **What do you prefer? :** Gives an idea about the preferrence of vehicle.
10. **How would you rate your knowledge about Electric Vehicle? :** This gives an idea about the consumers’ knowledge about Ev .
11. **Do you own an electric vehicle or have you ever used it? :** It’s a yes/no type question. This also gives us some data about the consumption of EV.
12. **How long have you owned or used an electric vehicle? :** This gives us the information about how long a consumer used an EV .
13. **Which type of electric vehicle do you have ? :** This gives an idea about the type of EV the consumers’ have.
14. **Which type of electric vehicles would you like to buy if you don't have an electic vehicles? :** This gives an idea about the choice of EV of a consumer that they wants to buy if they don’t have an EV and this may also give us a relation between how annual income is related to buying an EV.
15. **From which source do you get the most of your knowledge about electric vehicles? :** This question helps us to know about the source of popularity of EV.
16. **What do you think about the following statements about electric vehicles? [Can reduce global warming] :** It gives us consumers’ opinion about reduction of global warming
17. **What do you think about the following statements about electric vehicles? [Better than regular cars in terms of satisfying consumer needs]:** It gives us consumers’ opinion whether EV is better than regular cars or not in terms of satisfying consumer needs.
18. **What do you think about the following statements about electric vehicles? [Can save a lot of money to the owner] :** It gives us consumers’ opinion whether an EV can save a lot of money to the owner or not.
19. **What do you think about the following statements about electric vehicles? [Are very expensive] :** It gives us consumers’ opinion about EV is expensive or not.
20. **What do you think about the following statements about electric vehicles? [Maintenance and infrastructure is well developed] :** It gives us consumers’ opinion about maintenance and infrastructure of an EV.
21. **What do you think about the following statements about electric vehicles? [The cost to charge an Electric Vehicle is much less than the fuel costs for a petrol or diesel vehicle] :** It gives us consumers’ opinion about the cost to charge an EV is less or not with respect to fuel costs.
22. **What do you think about the following statements about electric vehicles? [Costs about the same to buy a petrol or diesel vehicles] :** It gives us consumers’ opinion whether cost for buying an EV is same as fuel/CNG vehicle or not.
23. **How to increase the mileage of an electric vehicles? :** This question is for checking the knowledge of a consumer regarding an EV.
24. **Is battery replacement of Electric Vehicle expensive? :** It gives us consumers’ opinion about replacing battery of an EV is expensive or not. It’s a Yes/No type question.
25. **Is the performance of an Electric Vehicles same as Internal Combustion Engine (ICE)? :** It’s a yes/no type question. This question is for checking the knowledge of a consumer regarding an EV.
26. **Does electric vehicle needs PUC (Pollution Under Control) certificate ? : :** It’s a yes/no type question. This question is for checking the knowledge of a consumer regarding an EV.
27. **Have government incentives influenced your decision to purchase an electric vehicle? : :** It’s a yes/no type question. This question is to understand whether a consumer is influenced by government incentives or not.
28. **How likely that your next vehicle will be an electric vehicle? :** It gives us an idea about the consumption of EV by the consumer .
29. **What factor encourage you to consider electric vehicles? :** This questions helps us to understand the factors which encourages a consumer to buy EV
30. **What factor discourage you to consider electric vehicles? :** This questions helps us to understand the factors which discourages a consumer to buy EV.
31. **If you already have an electric vehicle, it is of which company? :** This question helps us to analyze which company’s EV is mostly consumed by a consumer.
32. **If you want to buy an electric vehicle, then which company will you prefer? :** This question helps us to analyze which company’s EV is mostly preferred by a consumer.
33. **Which company's electric vehicle will you suggest to others? :** This question helps us to analyze which company’s EV is mostly suggested by a consumer to others.
34. **Will you switch to an electric vehicle if the company offers an exchange value on your owned vehicle to buy an electric vehicle? :** It’s a yes/no type question
35. **On a scale of 1 to 5, how would you rate your experience with my survey? :** This question is just to put a rating of the experience with my survey.

**3. METHODOLOGY**

**3.1. Statement of the problem:**

The topic of the project is ‘a study on the consumer behavior towards buying electric vehicles: with special reference to youngsters. The goal of the study is to examine the current state of electric car potential, as well as customer attitudes and preferences regarding electric vehicles, particularly among young people. The research would also aid in concentrating on the numerous elements that may have influenced the use of electric vehicle.

The significance of study is to analyze customer’s attitudes toward electric vehicles as well as projected demand in the future. As a result, statistical study of people’s desire for electric vehicles would be possible. The research would also aid in deducing the experiences and perspectives of various electric vehicle customers.

**3.2. Objective of the study:**

• To know the awareness level of Electric vehicle among consumers.

• To examine the factors affecting the decision to purchase electric vehicle.

**3.3 Area of the survey:**

According to the aim of the survey the area for sample collection should be broader, so I have collected data from Urban as well as Rural area (i.e. in my village).

**3.4 Sampling techniques:**

A group of units or elements which have well defined characteristics under study, called population. The population may be finite or infinite, a finite population in which unit of population is finite and an infinite population is one in which member of population is infinite. A sample is a finite subset of statistical individuals in a population and a number of individuals/units in a sample is called a sample size. On the basis of sample, we can estimate about the population parameter in which we are interested. The sample was selected by using random sampling as sampling technique & Sample size collected for survey is 448.

**3.5 Data collection:**

The data has been collected through random sampling, from the sample respondents with the help of the questionnaire which was distributed online (i.e. by Google Form).

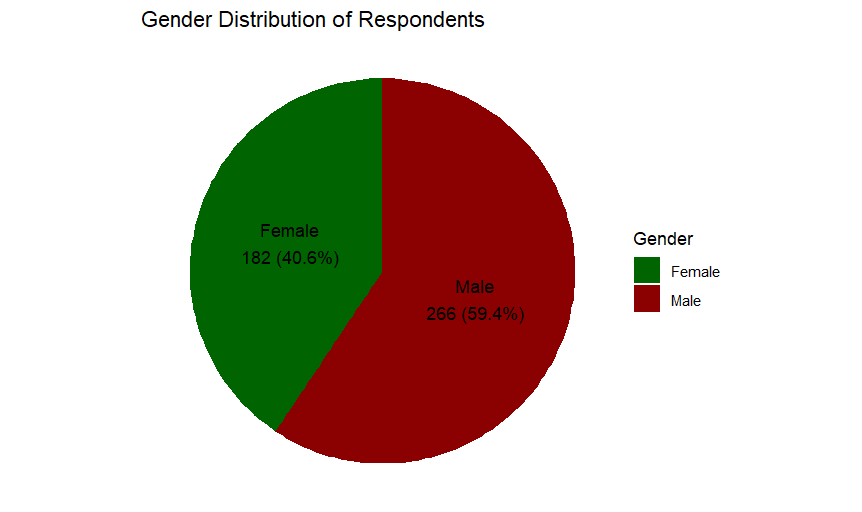
**3.6 Software and Programming Language Used:**

Data analysis involves summarizing the Raw data and interpreting their meaning which provides clear answer to questions in which we are interested. For this purpose, I have used software named as MS-EXCEL & MS- WORD. Then I have done the analysis and interpret the data using statistical tools (bar chart and pie chart) by writing code in R programming language and also used Chi-square test. I have also done logistic fitting and graphs of PCA.

**3.7. Explanatory Data Analysis**

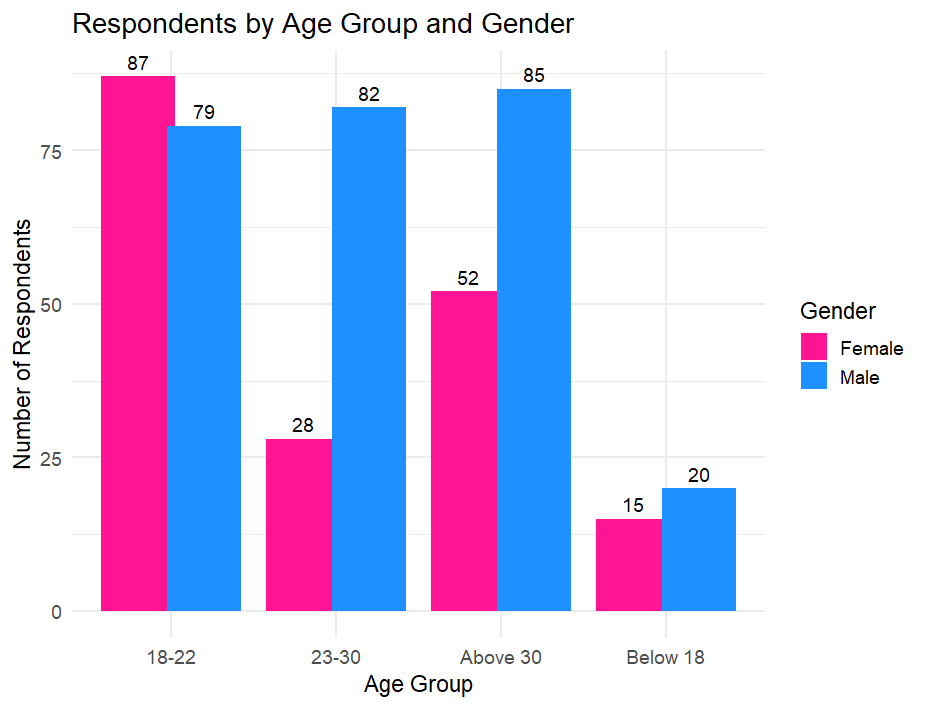
**3.7.1 Diagrammatic representation**

Diagrammatic Presentation of Data gives an immediate understanding of the real situation to be defined. Using the “ggplot” command, some divided bar diagrams along with some pie charts have been presented and just at a glance, an inference about Male, Female and finally the whole dataset can be drawn.



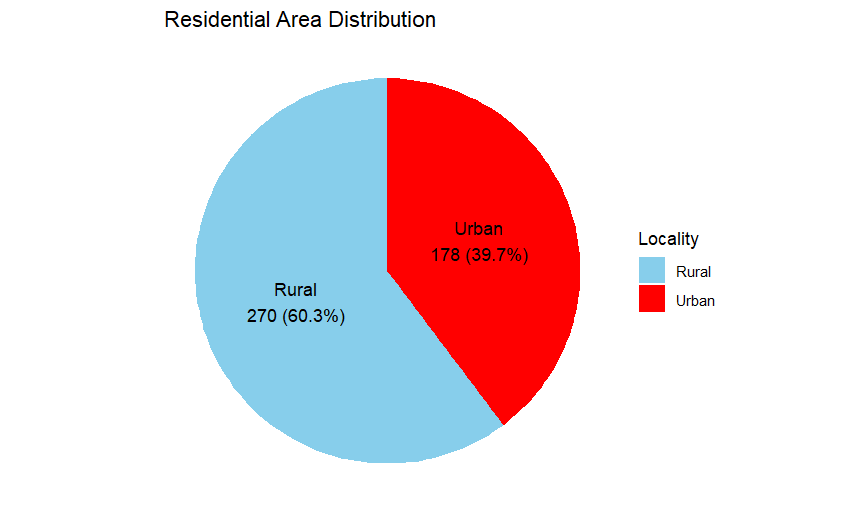
**Figure 3:** Pie Chart for gender

**Interpretation:** The data we get from our survey is that there are 266 male respondents and 182 female respondents.



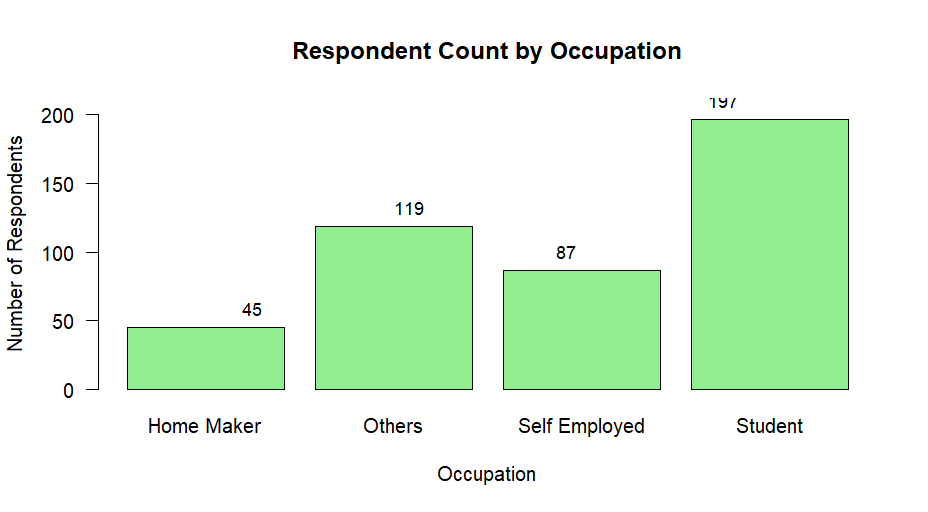
**Figure 4:** Bar plot for Age Vs. Gender

**Interpretation:** This graph represents Age Vs. Gender. We can see that in between age group of 18-22 years 87 respondents are female and 79 respondents are male. In between age group of 23 – 30 years total number of female respondents are 28 and total number of male respondents are 82. Above 30 years of old there are 52 female respondents and 85 male respondents. And, also below 18 years of age 15 respondents are female and 20 respondents are male.



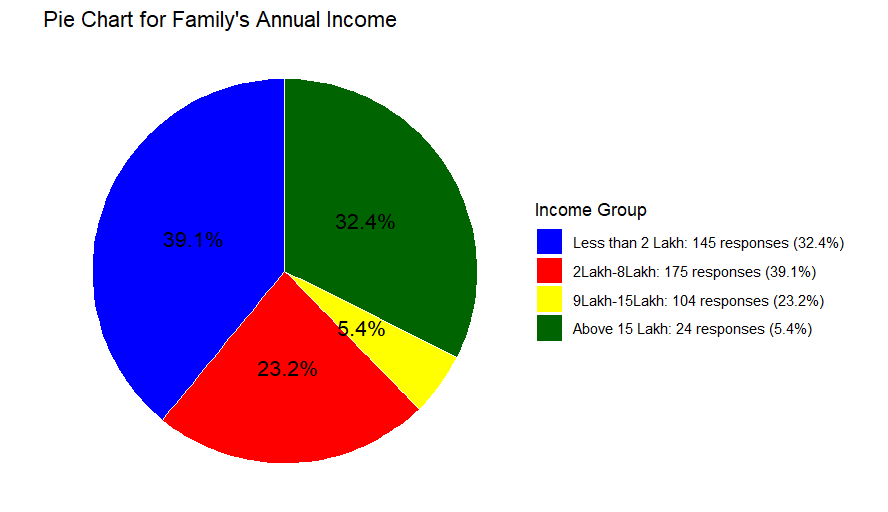
**Figure 5:** Pie Chart for residential area

**Interpretation:** Out of 448 respondents, 178 reside in urban areas and 270 reside in rural areas.



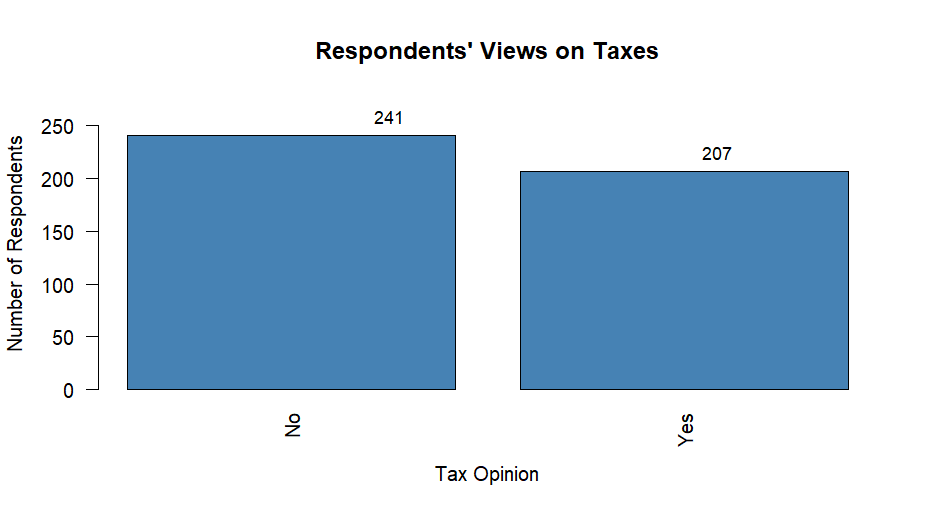
**Figure 6:** Bar plot for Occupation

**Interpretation:** The graph represents that out of 448 respondents , the occupation of 45 respondents are home maker , 87 are self employed , 197 are students and 119 have other ways of occupation.



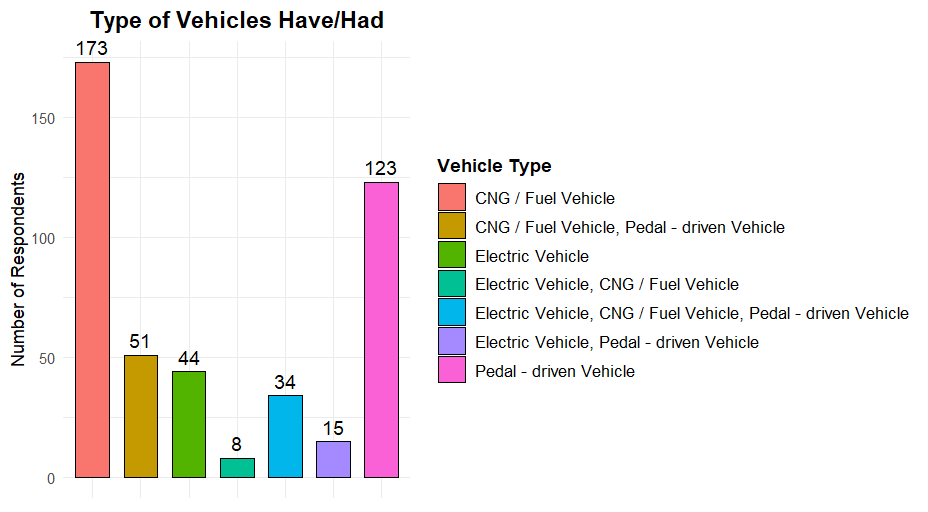
**Figure 7:** Pie Chart for Family’s Annual Income

**Interpretation:** The pie chart represents that 145 respondents have family income less than 2 Lakh , 175 respondents have family income in between 2 Lakh – 8 Lakh . In between 9-15 Lakh there are 104 respondents and above 15 Lakh there are 24 respondents .



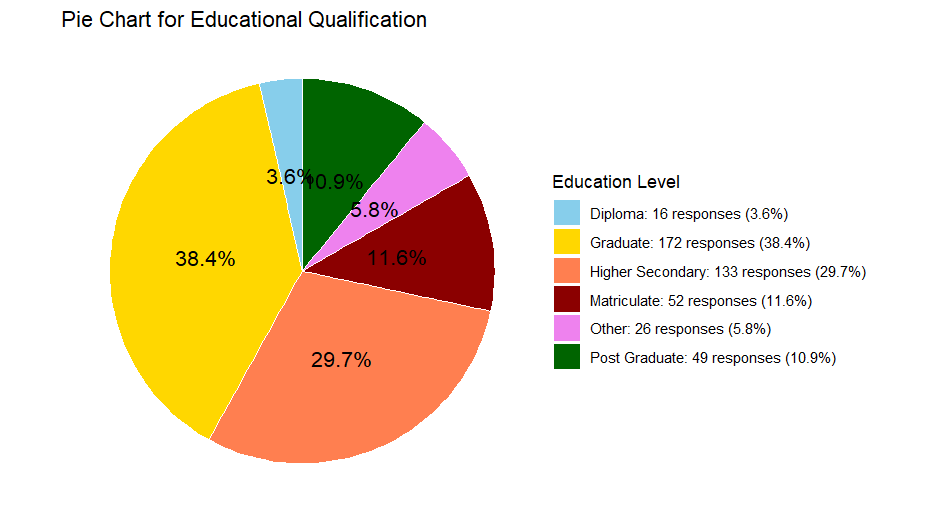
**Figure 8:** Bar chart of taxes

**Interpretation:** The bar chart represents that 241 respondents doesn’t pay any taxes and 207 people pay taxes. From this we can also conclude that 241 people have annual family income less than 3.5Lakh.



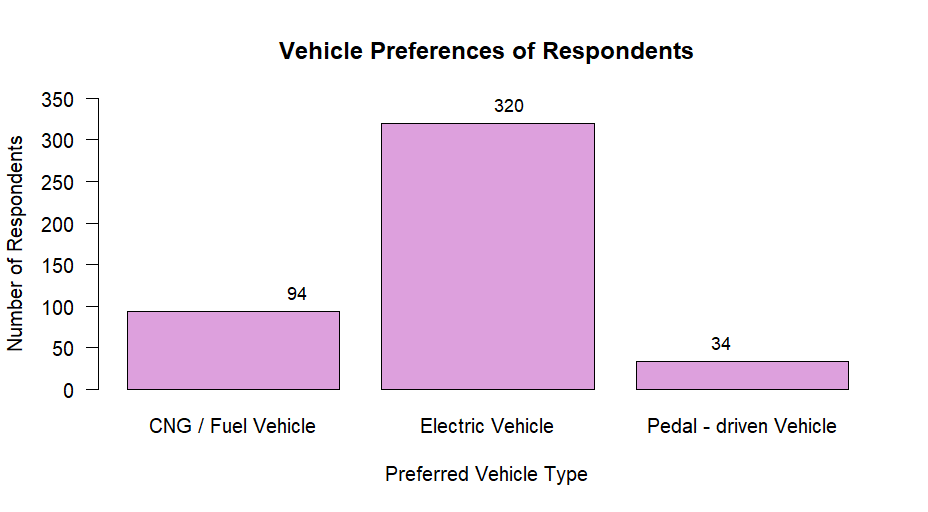
**Figure 9:** Bar chart of Type of vehicles respondents have/had

**Interpretation:** These bar chart gives us the details of vehicles respondents have/had. Out of 448 respondents 173 respondents have only CNG/ Fuel vehicle , 51 respondents have CNG/Fuel vehicle and pedal – driven vehicle , 45 respondents have only EV , 8 respondents have EV and CNG/Fuel vehicle , 34 respondents have EV , CNG/Fuel vehicle as well as pedal driven vehicle and 122 respondents have only pedal - driven vehicle .



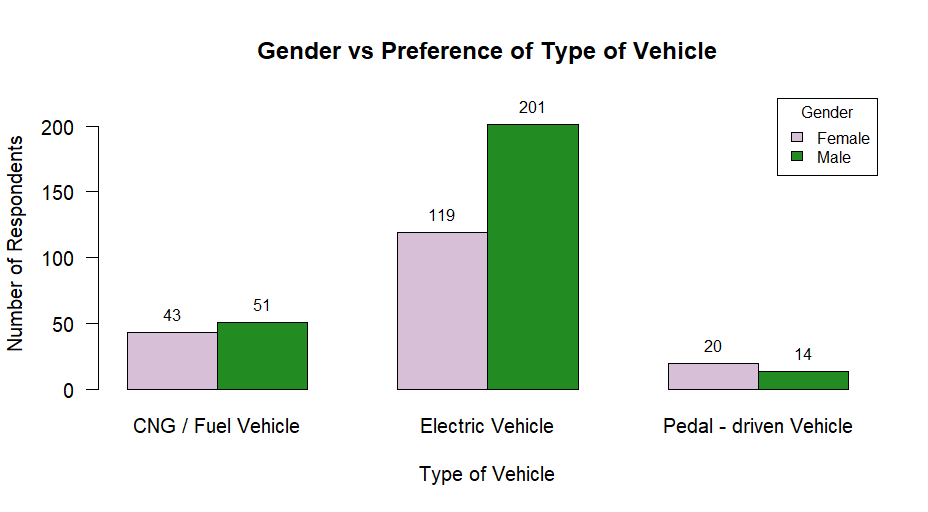
**Figure 10:** Pie Chart for Educational Qualification

**Interpretation:** From this pie chart we can see that 16 respondents have educational qualification of Diploma, 52 respondents have educational qualification upto matriculate , 133 respondents have educational qualification upto higher secondary , 172 respondents have educational qualification upto graduation , 49 respondents have educational qualification upto post diploma and 29 respondents have any other educational qualification.



**Figure 11:** Bar chart for preference of vehicles

**Interpretation:** Among 448 respondents 94 prefers CNG/Fuel vehicle, 320 prefers EV, 34 prefers pedal – driven vehicle.

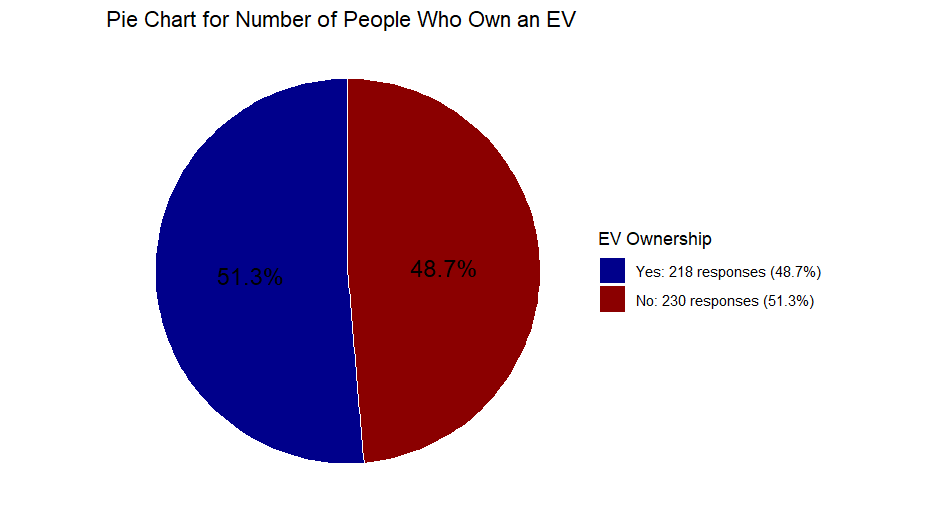


**Figure 12:** Barplot of Gender vs. Preference of vehicle

**Interpretation:** This bar plot gives us an idea of gender vs. preference of vehicles.

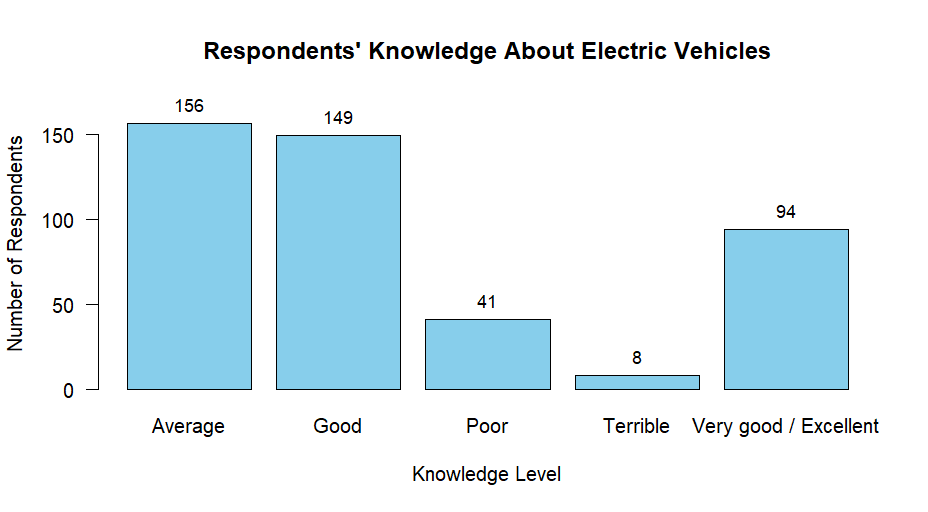
In case of females, 43 respondents pefers CNG/Fuel vehicle, 119 prefers Electric Vehicle and 20 prefers Pedal-driven vehicle.

In case of male respondents, 51 prefers CNG/Fuel vehicle, 201 prefers EV and 14 prefers Pedal-driven Vehicle.



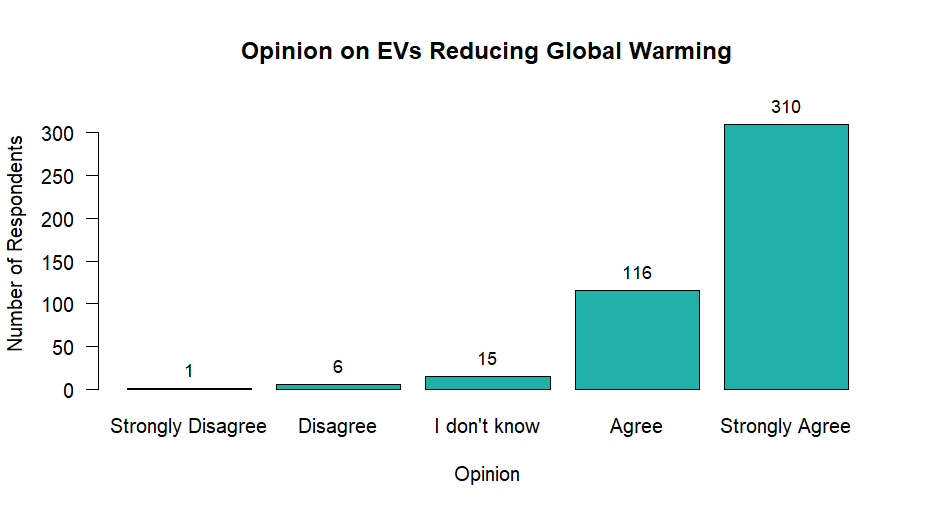
**Figure 13:** Pie Chart for number of people own an EV

**Interpretation:** Among 448 respondents 218 already have an EV and 230 respondents doesn’t have an EV.



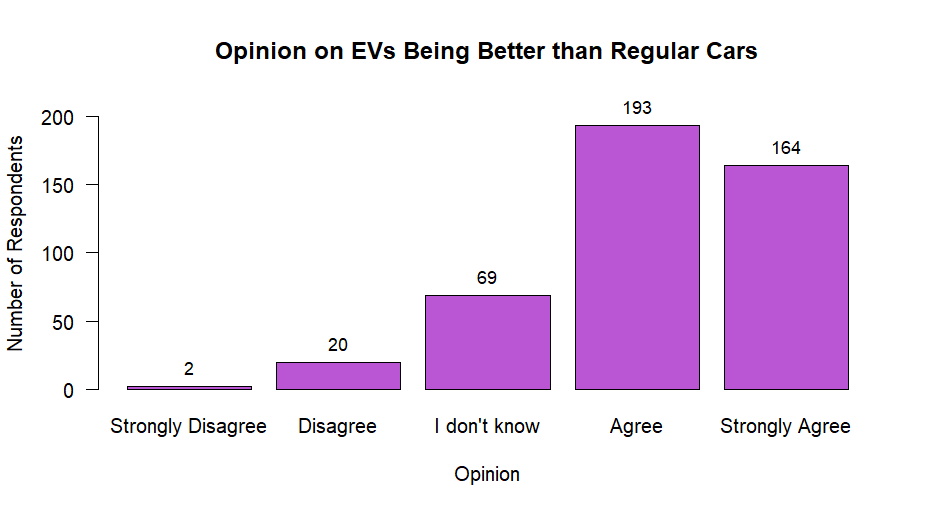
**Figure 14:** Bar Chart for knowledge about EV

**Interpretation:** This bar plot represents that 94 respondents have very good/excellent knowledge about EV, 156 respondents have average knowledge about EV, 149 have good knowledge and 41 have poor knowledge and 8 have terrible knowledge about EV.



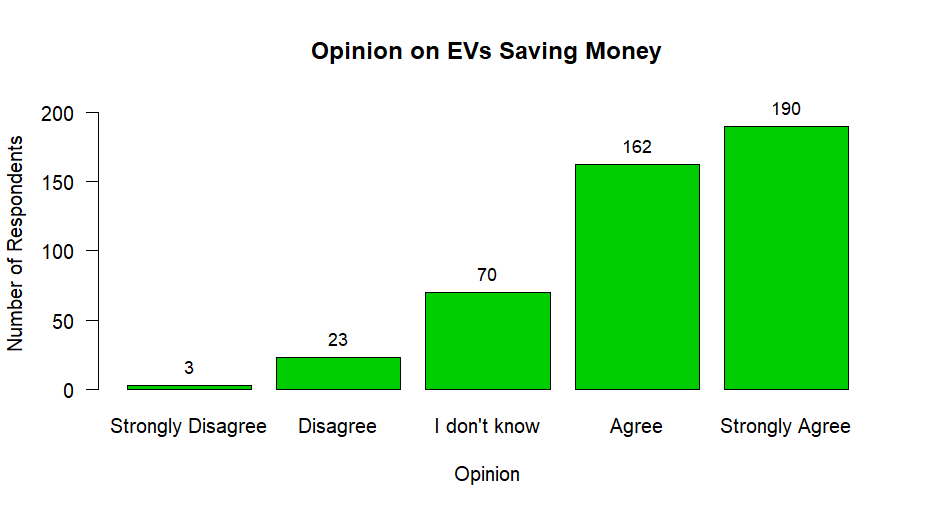
**Figure 15:** Bar chart for opinion about reduction of Global warming

**Interpretation:** From this bar plot we can conclude that 310 respondents strongly agrees with the fact that EV reduces global warming. 116 respondents agrees that EV reduces global warming. 15 respondents doesn’t know about this topic and 6 respondents disagrees and 1 respondent disagreed that EV can reduce global warming. So, from this data we can conclude that 426 people are aware of the fact that EV can reduce global warming.



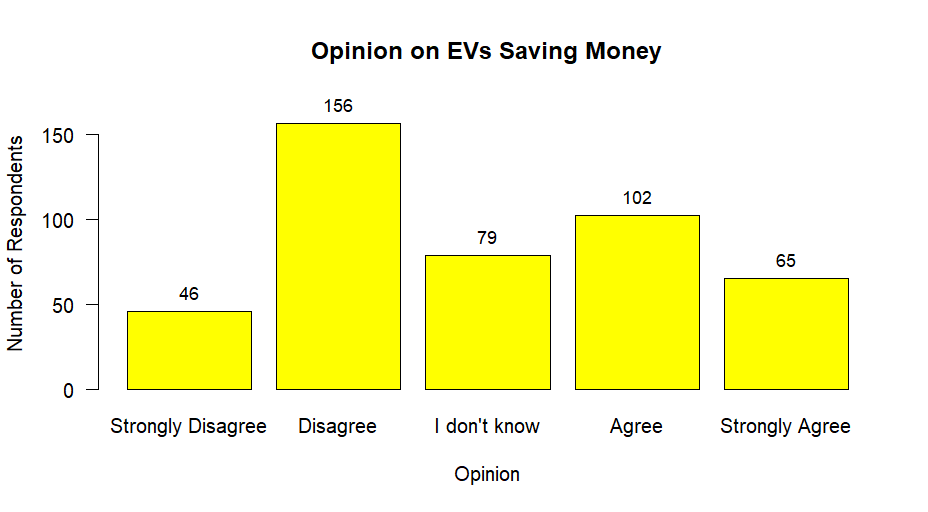
**Figure 16:** Bar Chart for opinion about EV better than regular Cars

**Interpretation:** According to this bar chart 2 respondents strongly disagreed and 20 respondents disagreed with the fact that EV is better than regular cars. 193 respondents agrees that EV is better than regular cars. 69 respondents responded that they didn’t know about this fact. And 164 respondents strongly disagreed with the fact that EV serves better than regular car.

****

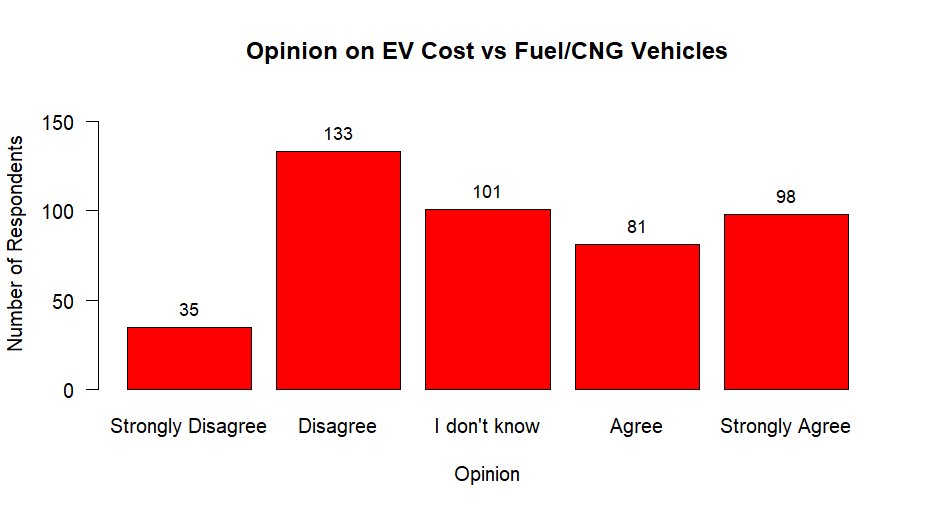
**Figure 17:** Bar chart for opinion about EV can save a lot of money

**Interpretation:** Out of 448 respondents, 164 strongly agreed and 193 agreed that EV saves a lot of money of the owner. 20 people disagreed and 2 people strongly disagreed with the same fact. And 69 people responded that they didn’t know about this fact.

****

**Figure 18:** Bar Chart for opinion about EVs are expensive

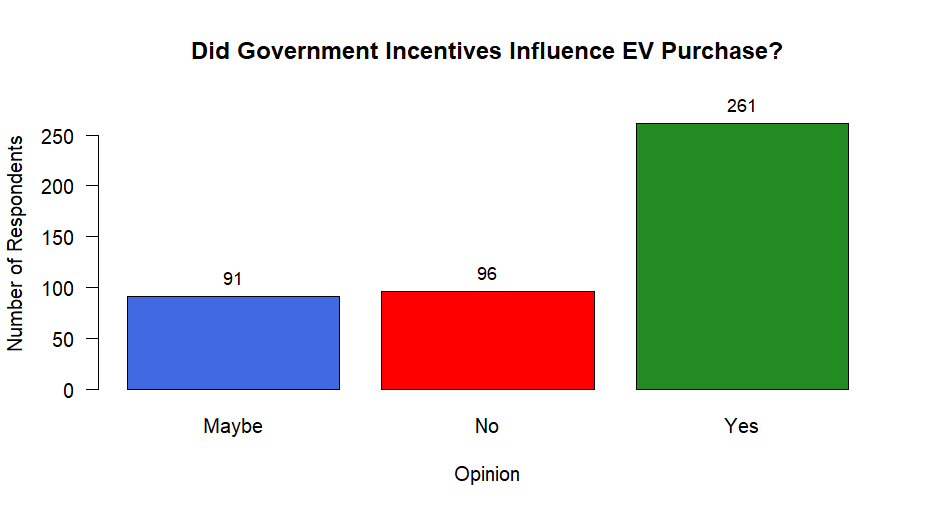
**Interpretation:** Among 448 respondents, 102 respondents responded that Ev are expensive to buy. 65 people strongly agreed that EVs are expensive. But, 156 respondents responded that Ev are not expensive, i.e., they disagreed to this fact and also 46 people strongly disagreed to this fact. 79 responded that they didn’t know about this fact.

****

**Figure 19:** Bar Chart for opinion about EV costs about

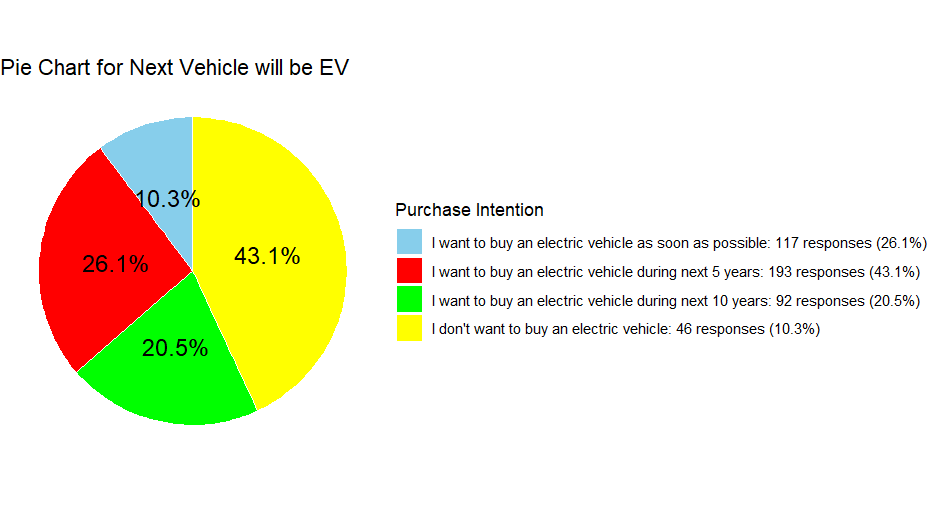
the same to buy a Fuel/CNG vehicles

**Interpretation:** Among 448 respondents, 81 agreed and 98 strongly agreed to the fact that EV costs about the same to buy a Fuel/ CNG vehicles. 35 strongly disagreed and 133 disagreed to the fact that EV costs about the same to buy a Fuel/CNG vehicle. And 101 responded that they didn’t know about this fact.

****

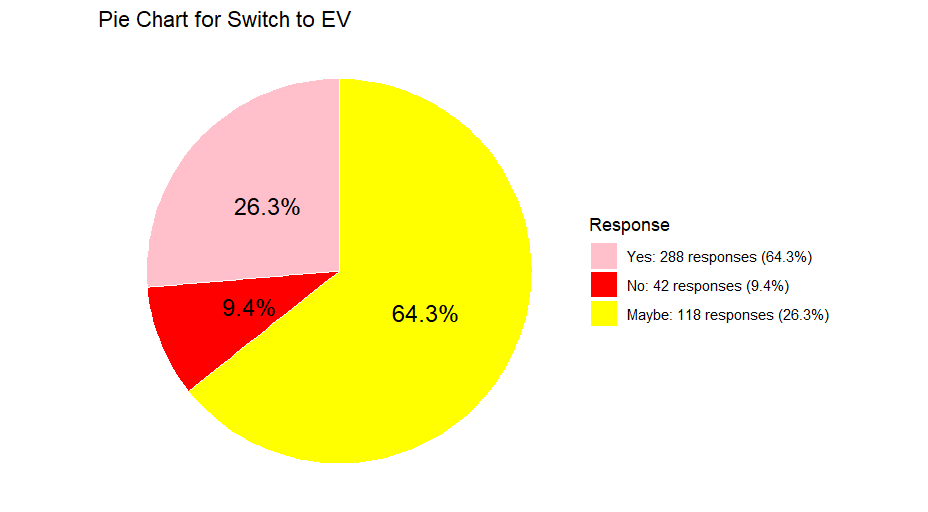
**Figure 20:** Bar chart for government incentives influenced decision to buy EV

**Interpretation:** The bar chart shows that government incentives have influenced the consumer decision to buy an EV. 261 respondents responded that they are influenced by government incentives. 96 responded that they aren’t influenced by government incentives. And lastly 91 responded that maybe they are influenced by government incentives.

****

**Figure 21:** Pie Chart for Next vehicle will be EV

**Interpretation:** The pie chart shows that 117 respondents wants to buy an EV as soon as possible. 193 responded that they wants to buy an EV during next 5 years. 92 responded that they wants to buy EV during next 10 years and 46 responded that they don’t want to buy an EV.

****

**Figure 22:** Pie Chart for switch to EV

**Interpretation:** The pie chart represents that among 448 respondents 288 respondents wants to switch to EV if the company offers an exchange value on their owned vehicle to buy an electric vehicle. 42 responded that they don’t want to switch to EV even if the company offers an exchange value on their owned vehicle to buy an electric vehicle. And 118 respondents responded that maybe they will switch to EV if the company offers an exchange value on their owned vehicle to buy an electric vehicle.

* 1. **Chi square test of association:**

The Chi-square Test provides a method for testing the association between the row and column variables in a two-way table (contingency table) consisting of categorical data. The **null hypothesis H0** assumes that there is no association between the variables (in other words, one variable does not vary according to the other variable), while the **alternative hypothesis H1** claims that some association does exist.

* + 1. ***Test statistic :***

This test is based on a test statistic that measures the divergence of the observed data from the values that would be expected under the null hypothesis of no association. It is also known as Pearsonian χ² statistic. It is given as,

χ2(cal.) = ~ χ2 (r-1)(s-1)

Where r and s are the no. of rows and number of columns of the contingency table and i=1, 2……r and j=1, 2…….. s.

Oij= Observed frequency of (i, j)th cell.

Eij= Expected frequency of (i, j)th cell.

Where Nio and Noj are marginal totals and N is the total number of observations.

The test statistic follows, under H0 a chi-square distribution with (r-1) (s- 1) degrees of freedom. The null hypothesis can be tested either at 5% or 1% level of significance.

If χ2 (cal.) < χ2(tab) then H0 may be accepted or if the p – value > 0.05 H0 may be accepted which shows that the two variables are independent of each other otherwise we may reject the H0 which shows that the two variables are not independent i.e., dependent of each other.

* + 1. **Interpretation Of Our Data Based On Test**

**Table : 2**

**Contingency table of Gender with respect to preference of type of vehicle**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Gender | Preference of type of vehicle | | | TOTAL |
| CNG/Fuel vehicle | Electric vehicle | Pedal -driven vehicle |
| Male | 51 | 201 | 14 | 266 |
| FEMALE | 43 | 119 | 20 | 182 |
| TOTAL | 94 | 320 | 34 | 448 |

Pearson's Chi-squared test

χ² = 7.2573

p value = 0.02655

df = 2

**Interpretation:** This table depicts how gender is associated with the preference of type of vehicle. Here, p-value < 0.05, thus the null hypothesis is rejected, i.e there is sufficient evidence of association between gender with preference of type of vehicle.

**Table :3**

**Contingency table of occupation with respect to gender**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Gender** | **OCCUPATION** | | | | **TOTAL** |
| **Home maker** | **Others** | **Self employed** | **Student** |
| **Male** | 44 | 17 | 12 | 109 | 182 |
| **FEMALE** | 1 | 102 | 75 | 88 | 266 |
| **TOTAL** | 45 | 119 | 87 | 197 | 448 |

**Pearson's Chi-squared test**

χ² = 138.792

p value < 2.2e-16

df = 3

**Interpretation:** This table depicts how occupation are associated with the gender. Here, p-value < 0.05, thus the null hypothesis is rejected, i.e there is sufficient evidence of association between occupation with respect to gender.

**Table : 4**

**Contingency table of Battery replacement of EV is expensive with next vehicle will be EV**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Battery replacement expensive** | **Next vehicle EV** | | | | |
| **I don’t want to buy an electric vehicle** | **I want to buy an electric vehicle as soon as possible** | **I want to buy an electric vehicle during next 10 years** | **I want to buy an electric vehicle during next 5 years** | **TOTAL** |
| **MAYBE** | 24 | 49 | 40 | 64 | 177 |
| **NO** | 2 | 16 | 4 | 7 | 29 |
| **YES** | 20 | 52 | 48 | 122 | 242 |
| **TOTAL** | 46 | 117 | 92 | 193 | 448 |

**Pearson's Chi-squared test**

χ² = 23.133

p value = 0.0007533

df = 6

**Interpretation:** This Table depicts how battery replacement of EV is expensive is associated with the next vehicle EV. Here, p-value < 0.05, thus the null hypothesis is rejected , i.e there is sufficient evidence of association between battery replacement of EV is expensive with next vehicle will be EV .

**Table: 5**

**Contingency table of cost to buy an EV is same as CNG/Fuel Vehicle with next vehicle will be EV**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Costs same as CNG/Fuel vehicle | Next vehicle would be EV | | | | TOTAL |
| I Don’t want to buy an EV | I want to buy an Ev as soon as possible | I want to buy an EV during the next 10 years | I want to buy an EV during the next 5 years |
| Agree | 10 | 23 | 21 | 27 | 81 |
| Disagree | 5 | 44 | 18 | 66 | 133 |
| I don’t know | 25 | 16 | 25 | 35 | 101 |
| Strongly agree | 5 | 28 | 11 | 54 | 98 |
| Strongly disagree | 1 | 6 | 17 | 11 | 35 |
| TOTAL | 46 | 117 | 92 | 193 | 448 |

**Pearson's Chi-squared test**

χ² = 70.798

p value = 2.271e-10

df = 12

**Interpretation:** This table depicts how cost to buy an EV is same as CNG/Fuel Vehicle

are associated with next vehicle will be EV. Here, p-value < 0.05, thus the null hypothesis is rejected, i.e., there is sufficient evidence of association between cost to buy an EV is same as CNG/Fuel Vehicle with next vehicle will be EV.

**Table : 6**

**Contingency table of occupation with EV is expensive**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Occupation**  **EV is**  **expensive** | **Home Maker** | **Others** | **Self Employed** | **Student** | **Total** |
| **Agree** | 10 | 14 | 16 | 62 | 102 |
| **Disagree** | 25 | 49 | 35 | 47 | 156 |
| **I don't know** | 4 | 17 | 10 | 48 | 79 |
| **Strongly Agree** | 3 | 11 | 22 | 29 | 65 |
| **Strongly Disagree** | 3 | 28 | 4 | 11 | 46 |
| **Total** | 45 | 119 | 87 | 197 | 448 |

**Pearson's Chi-squared test**

χ² = 76.756

p value = 0.2302

df = 12

**Interpretation:** This table depicts how occupation are associated with EV is expensive. Here, p-value > 0.05, thus the null hypothesis is accepted.

**Table : 7**

**Contingency table of age group with next vehicle will be EV**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Age group | Next vehicle would be EV | | | | TOTAL |
| I Don’t want to buy an EV | I want to buy an Ev as soon as possible | I want to buy an EV during the next 10 years | I want to buy an EV during the next 5 years |
| 18-22 | 35 | 26 | 48 | 57 | 166 |
| 23-30 | 6 | 33 | 12 | 59 | 110 |
| Above 30 | 3 | 43 | 23 | 68 | 137 |
| Below 18 | 2 | 15 | 9 | 9 | 35 |
| TOTAL | 46 | 117 | 92 | 193 | 448 |

**Pearson's Chi-squared test**

χ² = 65.246

p value = 9.087e-07

**Interpretation:** This table depicts age group are associated with next vehicle will be EV. Here, p-value < 0.05, thus the null hypothesis is rejected, i.e., there is sufficient evidence of association between age with next vehicle will be EV.

**Table : 8**

**Contingency table of influenced by gov. incentives with next vehicle will be EV**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Buy EV**    **Gov.**  **Incent**  **-ives**  **Influenced** | **I don't want to buy an electric** | **I want to buy an electric vehic** | **I want to buy an electric vehic** | **I want to buy an electric vehic** | **All** |
| Maybe | 18 | 15 | 24 | 34 | 91 |
| No | 21 | 16 | 29 | 30 | 96 |
| Yes | 7 | 86 | 39 | 129 | 261 |
| Total | 46 | 117 | 92 | 193 | 448 |

**Pearson's Chi-squared test**

χ² = 62.534

p value = 1.373e-11

df = 6

**Interpretation:** This table depicts how gov. incentives are associated with next vehicle will be EV. Here, p-value > 0.05, thus the null hypothesis is accepted.

**Table: 9**

**Contingency table of influenced by gov. incentives with switch to EV**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Switch**  **to EV**  **Gov. incentives** | **Maybe** | **No** | **Yes** | **Total** |
| **Maybe** | 51 | 9 | 31 | 91 |
| **No** | 39 | 22 | 35 | 96 |
| **Yes** | 28 | 11 | 222 | 261 |
| **Total** | 118 | 42 | 288 | 448 |

**Pearson's Chi-squared test**

χ² = 130.299

p value = 0.216

df = 4

**Interpretation:** This table depicts how gov. incentives are associated with switching to EV. Here, p-value > 0.05, thus the null hypothesis is accepted.

**Table : 10**

**Contingency table of residential area with switch to EV**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Switch**  **to EV**  **Residen-**  **tial**  **Area** | **Maybe** | **No** | **Yes** | **Total** |
| Rural | 50 | 21 | 199 | 270 |
| Urban | 68 | 21 | 89 | 178 |
| Total | 118 | 42 | 288 | 448 |

**Pearson's Chi-squared test**

χ² = 27.006

p value = 0.325

df = 2

**Interpretation:** This table depicts how residential area are associated with switching to EV if the company offers an exchange value on your owned vehicle to buy an electric vehicle. Here, p-value > 0.05, thus the null hypothesis is accepted.

* + 1. **Inference**

A surprising part of the result is that most of the pairs of variables may be significantly associated in the total population, but independent in the gender based sub-populations and vice-versa.

In our study, this eventually introduces Simpson’s Paradox, which is a statistical phenomenon where an association between two variables in a population emerges, disappears or reverses when the population is divided into sub-populations.

**3.9 Analysis of variance**

**3.9.1 Definition**

ANOVA, which stands for Analysis of Variance, is a statistical test used to analyze the difference between the means of more than two groups.

A one-way ANOVA uses one independent variable, while a two-way ANOVA uses two independent variables.

**3.9.2 Use a one-way ANOVA**

When you have collected data about one categorical independent variable and one quantitative dependent variable, the independent variable should have at least three levels (i.e. at least three different groups or categories). ANOVA tells us if the dependent variable changes according to the level of the independent variable or not.

**3.9.3 Assumptions of ANOVA**

For the results of a one-way ANOVA to be valid, the following assumptions should be met:

1. **Normality –** Each sample was drawn from a normally distributed population.
2. **Equal Variances –** The variances of the populations that the samples come from are equal. You can use Bartlett’s Test to verify this assumption.

**3. Independence –** The observations in each group are independent of each other and the observations within groups were obtained by a random sample.

A one-way ANOVA uses the following null and alternative hypotheses:

**𝐻0 (null hypothesis):** μ1= μ2 = μ3 = … = μK (all the population means are equal)

**VS**

**𝐻1 (alternative hypothesis):** at least one population mean is different from the rest .

No matter which software you use, you will receive the following table as output:

**Table: 11**

**Anova Table**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **S.O.V.** | **df** | **SS** | **MS** | **F** | **p** |
| Treatment | 𝑑𝑓r | SSR | MSR | MSR/MSE | F dfr,dfe |
| Error | 𝑑𝑓e | SSE | MSE |
| Total | 𝑑𝑓t | SST | **\_\_\_\_\_** | **\_\_\_\_\_** | **\_\_\_\_\_** |

**k:** total number of groups

**n:** total observations

**SSR:** regression sum of squares

**SSE:** error sum of squares

**SST:** total sum of squares (SST = SSR + SSE)

**𝑑𝑓r :** regression degrees of freedom ( dfr = k-1)

**𝑑𝑓e:** error degrees of freedom ( dfe = n-k)

**𝑑𝑓t:** total degrees of freedom ( dft = n-1)

**MSR:** regression mean square (MSR = SSR/𝑑𝑓r )

**MSE:** error mean square (MSE = SSE/𝑑𝑓e )

**F:** The F test statistic (F = MSR/MSE)

**p:** The p-value that corresponds to F dfr,dfe

If the p-value is less than your chosen significance level (e.g. 0.05), then you can reject the null hypothesis and conclude that at least one of the population means is different from the others.

Note: If you reject the null hypothesis, this indicates that at least one of the population means is different from the others, but the ANOVA table doesn’t specify which population means are different. To determine this, you need to perform post hoc tests, also known as “multiple comparisons” tests.

**3.9.4 Tests based on one-way ANOVA**

**[Test-1]**

Denoting the “Age” column into four groups respectively:

● Group 1: Below 18

● Group 2: 18-22 years

● Group 3: 23-30 years

● Group 4: Above 30 years

Each group has entries of preference of vehicles for each individual. So, we want to test whether there is any effect of age grouping on preference of vehicle(𝐻𝑎) or not (𝐻0). The test report is as follows:

**Table: 12**

**Anova: Single Factor (summary)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Group** | **Count** | **Mean** | **Variance** |
| **Group 1** | |  | | --- | | 166 | | |  | | --- | | 1.7771 | | 0.42876304 |
| **Group 2** | |  | | --- | | 110 | | |  | | --- | | 1.8909 | |  | |  | | 0.24482704 |
| **Group 3** | |  | | --- | | 137 | |  | | |  | | --- | | 1.9343 | |  | | 0.10595025 |
| **Group 4** | |  | | --- | | 35 | | |  | | --- | | 1.9429 | | 0.17313921 |

**Table: 13**

**Anova table for Single Factor**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **S.O.V.** | **df** | **SS** | **MS** | **F** | **p** |
| **Age** | 3 | 2.226 | 0.7420 | 0.80 | 0.040 |
| **Error** | 444 | 117.738 | 0.2652 |
| **Total** | 447 | 119.964 | \_\_\_ | \_\_\_ | \_\_\_ |

**Interpretation: -** Based on the provided table, the p-value for the "Groups" variation is less than the significance level (typically 0.05). Therefore, we can reject the null hypothesis and conclude that there is sufficient evidence to suggest a significant difference between at least two of the age groups.

Now, to test which groups have significant differences between them, let us perform a Fisher Individual Tests for Differences of Means.

**3.9.5 Fisher Individual Tests for Differences of Means**

Use the individual confidence intervals to identify statistically significant differences between the group means, to determine likely ranges for the differences, and to determine whether the differences are practically significant. Fisher's individual tests table displays a set of confidence intervals for the difference between pairs of means.

The individual confidence level is the percentage of times that a single confidence interval includes the true difference between one pair of group means, if you repeat the study. Individual confidence intervals are available only for Fisher's method. All of the other comparison methods produce simultaneous confidence intervals.

Controlling the individual confidence level is uncommon because it does not control the simultaneous confidence level, which often increases to unacceptable levels. If you do not control the simultaneous confidence level, the chance that at least one confidence interval does not contain the true difference increases with the number of comparisons.

The confidence interval of the difference is composed of the following two parts:

* **Point estimate**

The point estimate is the difference between a pair of means and is calculated from the sample data. The confidence interval is centered on this value.

* **Margin of error**

The margin of error defines the width of the confidence interval and is determined by the observed variability in the sample and the confidence level. To calculate the upper limit of the confidence interval, the margin of error is added to the point estimate. To calculate the lower limit of the confidence interval, the margin of error is subtracted from the point estimate.

**Interpretation:**

Use the confidence intervals to assess the differences between group means. If the confidence intervals for any of the pairs of groups include zero, then it indicates that the differences between means are not statistically significant.

**Table: 14**

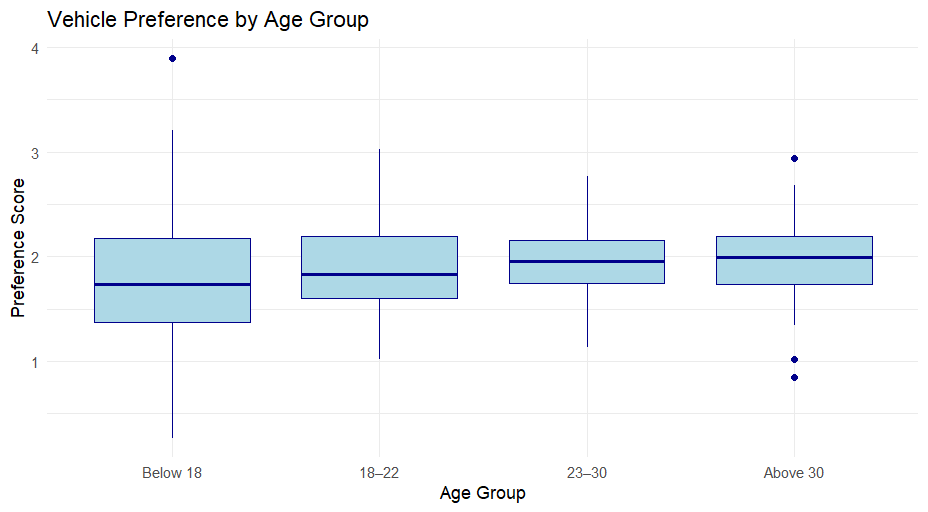
**Table for Fisher Individual Tests for Differences of Means**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Difference of Levels** | **Difference of Means** | **SE of Difference** | **95% CI** | **t-Value** | **Adjusted P-Value** |
| 23-30 - 18-22 | 0.1138 | 0.0633 | (-0.0106, 0.2382) | 1.80 | 0.073 |
| Above 30 - 18-22 | 0.1572 | 0.0594 | (0.0404, 0.2740) | 2.64 | 0.008 |
| Below 18 - 18-22 | 0.1657 | 0.0958 | (-0.0225, 0.3540) | 1.73 | 0.084 |
| Above 30 - 23-30 | 0.0434 | 0.0659 | (-0.0862, 0.1730) | 0.66 | 0.511 |
| Below 18 - 23-30 | 0.0519 | 0.0999 | (-0.1445, 0.2484) | 0.52 | 0.603 |
| Below 18 - Above 30 | 0.0086 | 0.0975 | (-0.1831, 0.2002) | 0.09 | 0.930 |

*We also get the* ***Simultaneous confidence level = 79.88%***

**The confidence intervals indicate the following:**

* The confidence interval for the difference between the means of 23-30 years and 18-22 years extends from -0.0106 to 0.2382. This range include zero, which indicates that the difference between these means is statistically insignificant.
* The confidence interval for the difference between the means of below 18 years and 18-22 years extends from -0.0225 to 0.3540. This range include zero, which indicates that the difference between these means is statistically insignificant.
* The confidence interval for the difference between the means of above 30 years and 23-30 years extends from -0.0862 to 0.1730. This range include zero, which indicates that the difference between these means is statistically insignificant.
* The confidence interval for the difference between the means of below 18 years and 23-30 years extends from -0.1445 to 0.2484. This range include zero, which indicates that the difference between these means is statistically insignificant.
* The confidence interval for the difference between the means of below 18 years and above years extends from -0.1831 to 0.2002. This range include zero, which indicates that the difference between these means is statistically insignificant.
* The confidence interval for the difference between the means above 30 years and 18-22 years extends from 0.0404 to 0.2740. This range doesn’t include zero, which indicates that the difference between these means is statistically significant.
* The 95% individual confidence level indicates that you can be 95% confident that each confidence interval contains the true difference for that specific comparison. However, the simultaneous confidence level indicates that you can be only 79.88% confident that all the intervals contain the true differences.



* There’s notable overlap among groups but subtle upward trends in median from "Below 18" to "Above 30".
* this boxplot reinforces the idea that **age might influence vehicle preference**, especially for older vs younger age brackets.

**3.10 Multiple logistic regression**

Multiple logistic regression is a statistical modeling technique used to predict the probability of a binary outcome (a dichotomous variable with two categories, often coded as 0 and 1) based on multiple predictor variables. It is an extension of simple logistic regression, which considers only one predictor variable.

In multiple logistic regression, the goal is to find the relationship between the binary response variable and two or more predictor variables by estimating the coefficients of the predictors. These coefficients represent the change in the log-odds of the outcome associated with a one-unit change in the predictor variable while holding other variables constant.

The model assumes that the relationship between the predictor variables and the log-odds of the binary outcome is linear on the logit scale. The logit function is the natural logarithm of the odds, and it maps the probability of success (the probability of the binary outcome being 1) to a continuous scale.

**3.10.1 Assumption**

Multiple logistic regression, like other regression models, relies on several assumptions to provide reliable and valid results. Here are the key assumptions of multiple logistic regression:

**Binary outcome:** The dependent variable (response variable) should be binary or dichotomous, representing two categories or classes. Logistic regression is specifically designed for modeling binary outcomes.

**Linearity of predictors and log-odds:** The relationship between the predictors and the log-odds of the outcome should be linear. This assumption implies that the relationship between each predictor variable and the log-odds of the outcome is a straight line when plotted on a logit scale.

**Independence of observations:** Observations in the dataset should be independent of each other. Independence assumption ensures that each observation provides unique information and that the model does not violate the assumption of independence in its estimation process.

**Absence of multicollinearity:** Predictor variables should be minimally correlated with each other. High correlations among predictors (multicollinearity) can lead to unstable estimates, make it difficult to interpret the individual effects of predictors, and affect the model's performance.

**Absence of influential outliers:** The presence of influential outliers can affect the estimated coefficients and undermine the validity of the model. It is important to check for and handle outliers appropriately.

**Adequate sample size:** Logistic regression typically requires a sufficient sample size to provide stable estimates and accurate inference. The rule of thumb is to have at least 10-15 observations per predictor variable to avoid overfitting and obtain reliable results.

**Assumption of no perfect separation:** Perfect separation occurs when the model can perfectly predict the outcome based on a combination of predictor variables, leading to infinite parameter estimates and unreliable inference. Techniques such as penalized regression or adding regularization terms can help address this issue.

It is essential to assess these assumptions when applying multiple logistic regression and consider potential violations that may impact the interpretation and validity of the results. Diagnostic tools, such as residual analysis and goodness-of-fit tests, can help evaluate the model's assumptions.

**3.10.2 Logit transformation**

In multiple logistic regression, the relationship between the predictor variables and the log-odds of the binary outcome is assumed to be linear on the logit scale. The logit transformation is used to convert the probability of success (the probability of the binary outcome being 1) to a continuous scale that spans from negative infinity to positive infinity.

The logit function is defined as the natural logarithm of the odds, which is the ratio of the probability of success (p) to the probability of failure (1 - p):

logit(p) = log(p / (1 - p))

In the context of multiple logistic regression, the model can be mathematically represented as:

logit(p) =β0 +∑ β𝑖𝑋i , i = 1,2….n

where:

logit(p) is the log-odds of the probability p of the binary outcome being 1,

β0 is the intercept or constant term,

βi ’s are the coefficients corresponding to the predictor variables

𝑋i ’s respectively where i=1,2,....,n

The logit transformation is useful in logistic regression because it allows us to model the linear relationship between the predictor variables and the log-odds of the binary outcome. This linear relationship simplifies the estimation process and facilitates the interpretation of the coefficients.

Once the logistic regression model is fitted, the predicted logit values can be converted back to probabilities using the inverse of the logit function, which is called the logistic function (also known as the sigmoid function):

p = 1 / (1 + exp(-logit(p)))

where:

p is the probability of success (the probability of the binary outcome being 1),

logit(p) is the logit value obtained from the logistic regression model.

The logistic function maps the logit values to probabilities that range between 0 and 1, making it suitable for representing probabilities of binary outcomes. These probabilities can then be used to make predictions on new data or evaluate the performance of the logistic regression model.

**3.10.3 General model**

The form of the model is,

log ( ) = β0 + β1 \* X1 +……+ βk \* Xk (3.3)

In this model, p is the probability that the dependent variable Y=1 and 𝑋i ’s are the independent variables (predictors). β0 is a constant and βi’s are known as the regression coefficients, which have to be 𝑖 estimated from the data. Logistic regression estimates the probability of a certain event occurring.

Logistic regression thus forms a predictor variable (log(p/(1-p)) which is a linear combination of the explanatory variables. The values of this predictor variable are then transformed into probabilities by a logistic function. Such a function has the shape of an S called “Sigmoid curve”.

**3.10.4 Removing outliers, influential and leverages**

In the project few influential and leverage are detected and finally removed. In logistic regression, a set of observations whose values deviate from the expected range and produce extremely large residuals and may indicate a sample peculiarity is called outliers. These outliers can unduly influence the results of the analysis and lead to incorrect inferences. An observation is said to be influential if removing the observation substantially changes the estimate of coefficients. Influence can be thought of as the product of leverage and outliers. An observation with an extreme value on a predictor variable is called a point with high leverage. Leverage is a measure of how far an independent variable deviates from its mean. Detection of outliers and influential cases and corresponding treatment is very crucial task of any modeling exercise. A failure to detect outliers and hence influential cases can have severe distortion on the validity of the inferences drawn from such modeling exercise.

**3.10.5 Interpretation**

**3.10.5.1 Interpretation of the first model**

We wish to fit our Logistic Model, for which a binary response is required. Here our binary response is Gender (Y). Here we want to predict the consumption of EV patterns between genders. Here, our explanatory variables or covariates are the following:

➢ 𝑋1 = Age

➢ 𝑋2 = Occupation

➢ 𝑋3 = Highest educational qualification

➢ 𝑋4 = Family income

➢ 𝑋5 = Next vehicle would be EV

➢ 𝑋6 = Influenced by Government incentives

The regression model is given by,

ln(p/1-p ) = β0 +i Xi +e

The coefficient estimates, standard errors, z-values, and p-values are provided for each covariate. The p-values indicate the significance of each covariate in predicting the response variable. In this case, next vehicle would be EV is significant at 0.05 level of significance. The other covariates are not significant at 0.05 level of significance. The AIC (Akaike Information Criterion) for this model is 89.2658611732685. A lower AIC value indicates a better-fitting model.

**3.10.5.2 Interpretation of the second model**

We wish to fit our Logistic Model, for which a binary response is required. Here our binary response is residential area (Y). Here we want to predict the the level of concern regarding environmental pollution among the respondents. Here, our explanatory variables or covariates are the following:

➢ 𝑋1= Age

➢ 𝑋2 = Occupation

➢ 𝑋3 = Highest educational qualification

➢ 𝑋4 = Family income

➢ 𝑋5 = Reduces global warming

➢ 𝑋6 = Next vehicle would be EV

The regression model is given by,

ln(p/1-p ) = β0 +i Xi +e

The coefficient estimates, standard errors, z-values, and p-values are provided for each covariate. The p-values indicate the significance of each covariate in predicting the response variable. In this case, next vehicle would be EV is significant at 0.05 level of significance. The other covariates are not significant at 0.05 level of significance. The AIC (Akaike Information Criterion) for this model is 89.0815865038871. A lower AIC value indicates a better-fitting model.

**3.10.6 Confusion Matrix**

A confusion matrix is a popular performance evaluation tool used in machine learning and statistics to assess the performance of a classification model. It provides a summary of the model's predictions compared to the actual ground truth labels for a given dataset. The confusion matrix is especially useful when dealing with problems where the output can belong to multiple classes or categories. The confusion matrix is structured as a square matrix, with rows and columns representing the true classes and predicted classes, respectively. It consists of four main components: True Positives (TP): The number of samples that were correctly predicted as positive (correctly classified instances of the positive class). True Negatives (TN): The number of samples that were correctly predicted as negative (correctly classified instances of the negative class). False Positives (FP): The number of samples that were incorrectly predicted as positive (instances of the negative class wrongly classified as positive). False Negatives (FN): The number of samples that were incorrectly predicted as negative (instances of the positive class wrongly classified as negative). The confusion matrix allows us to understand the performance of a classification model in terms of these four components, which helps in calculating various evaluation metrics like accuracy, precision, recall (sensitivity), specificity, and F1-score.

* **True Positives (TP):** Model correctly predicted the positive class.
* **True Negatives (TN):** Model correctly predicted the negative class.
* **False Positives (FP):** Model incorrectly predicted positive when it was actually negative (Type I error).
* **False Negatives (FN):** Model incorrectly predicted negative when it was actually positive (Type II error).

The confusion matrix helps calculate key performance metrics:

* **Accuracy** = (TP + TN) / Total predictions
* **Precision** = TP / (TP + FP)
* **Recall (Sensitivity)** = TP / (TP + FN)
* **Specificity** = TN / (TN + FP)
* **F1 Score** = Harmonic mean of precision and recall

**3.10.7 Interpretation**

[**3.10.7.1**](https://3.4.7.1/) **Confusion matrix for first model**

|  |  |  |
| --- | --- | --- |
| **Predicted Labels** | **Actual Labels** | |
| 0 | 1 |
| 0 | 266 | 18 |
| 1 | 0 | 164 |

**True Positives (TP):** The number of samples correctly predicted as  
positive (actual 1, predicted 1) is **164**.  
**True Negatives (TN):** The number of samples correctly predicted as  
negative (actual 0, predicted 0) is **266**.  
**False Positives (FP):** The number of samples incorrectly predicted  
as positive (actual 0, predicted 1) is **0**.  
**False Negatives (FN):** The number of samples incorrectly predicted  
as negative (actual 1, predicted 0) is **18**.  
**Accuracy =** **0.96**  
**Precision =** **1**  
**Recall (Sensitivity) =** **0.901**  
**Specificity =** **1**  
**F1-score =** **0.948**

* **Accuracy:** Accuracy is the proportion of correctly classified  
  instances (both true positives and true negatives) out of all the  
  instances in the dataset. An accuracy of **0.96** means  
  that the model correctly predicts the outcome for approximately  
  **96%** of the data.
* **Precision:** Precision is the proportion of true positive predictions  
  (correctly predicted positive instances) out of all positive  
  predictions (both true positives and false positives). A precision of  
  **1** means that when the model predicts a positive outcome, it is  
  correct **100%** of the time.
* **Recall (Sensitivity):** Recall, also known as sensitivity, is the  
  proportion of true positive predictions out of all actual positive  
  instances in the dataset. A recall of **0.901** means  
  that the model can identify approximately **90.1%** of the actual  
  positive instances correctly.
* **Specificity:** Specificity is the proportion of true negative predictions  
  (correctly predicted negative instances) out of all actual negative  
  instances in the dataset. A specificity of **1**  
  indicates that the model can identify **100%** of the  
  actual negative instances correctly.
* **F1-score:** The F1-score is the harmonic mean of precision and  
  recall. It combines both metrics and provides a single score that  
  balances precision and recall. An F1-score of **0.948**  
  suggests that the model is performing **very well** in terms of both  
  precision and recall.
* The model has a high overall accuracy of **96%**, meaning that it  
  correctly predicts the outcome for approximately **96%** of the data  
  points.  
  The precision of **100%** suggests that when the model predicts a  
  positive outcome, it is correct every time. This  
  indicates that the model has **zero false positives**.  
  The recall (sensitivity) of **90.1%** indicates that the model can  
  identify the majority of actual positive instances correctly.  
  This means it has a **low rate of false negatives**, missing only a small  
  portion of the positive cases.  
  The high specificity of **100%** implies that the model has a **very low  
  rate of false positives**, correctly identifying all negative instances.  
  The high F1-score of **0.948** indicates an excellent balance  
  between precision and recall. This suggests that the model is  
  performing strongly in capturing the positive instances while  
  maintaining a perfect false positive rate.

[**3.10.7.**](https://3.4.7.1/)**2 Confusion matrix for second model**

|  |  |  |
| --- | --- | --- |
| **Predicted Labels** | **Actual Labels** | |
| 0 | 1 |
| 0 | 270 | 14 |
| 1 | 0 | 164 |

**True Positives (TP):** The number of samples correctly predicted as  
positive (actual 1, predicted 1) is **164.**  
**True Negatives (TN):** The number of samples correctly predicted as  
negative (actual 0, predicted 0) is **270.**  
**False Positives (FP):** The number of samples incorrectly predicted  
as positive (actual 0, predicted 1) is **0.**  
**False Negatives (FN):** The number of samples incorrectly predicted  
as negative (actual 1, predicted 0) is **14.**  
**Accuracy = 0.969**  
**Precision = 1**  
**Recall (Sensitivity) = 0.921  
Specificity = 1  
F1-score = 0.959**

**• Accuracy:** Accuracy is the proportion of correctly classified  
instances (both true positives and true negatives) out of all the  
instances in the dataset. An accuracy of 0.969 means  
that the model correctly predicts the outcome for approximately  
96.9% of the data.  
**• Precision:** Precision is the proportion of true positive predictions  
(correctly predicted positive instances) out of all positive  
predictions (both true positives and false positives). A precision of  
1 means that when the model predicts a positive outcome, it is  
correct 100% of the time.  
**•** **Recall (Sensitivity):** Recall, also known as sensitivity, is the  
proportion of true positive predictions out of all actual positive  
instances in the dataset. A recall of 0.921 means  
that the model can identify approximately 92.1% of the actual  
positive instances correctly.  
**• Specificity:** Specificity is the proportion of true negative predictions  
(correctly predicted negative instances) out of all actual negative  
instances in the dataset. A specificity of 1  
indicates that the model can identify 100% of the  
actual negative instances correctly.  
**• F1-score:** The F1-score is the harmonic mean of precision and  
recall. It combines both metrics and provides a single score that  
balances precision and recall. An F1-score of 0.959  
suggests that the model is performing very well in terms of both  
precision and recall.

**•** The model has a high overall accuracy of 96.9%, meaning that it  
correctly predicts the outcome for approximately 96.9% of the data  
points.  
The precision of 100% suggests that when the model predicts a  
positive outcome, it is correct every time. This  
indicates that the model has zero false positives.  
The recall (sensitivity) of 92.1% indicates that the model can  
identify the majority of actual positive instances correctly.  
This means it has a low rate of false negatives, missing only a small  
portion of the positive cases.  
The high specificity of 100% implies that the model has a very low  
rate of false positives, correctly identifying all negative instances.  
The high F1-score of 0.959 indicates an excellent balance  
between precision and recall. This suggests that the model is  
performing strongly in capturing the positive instances while  
maintaining a perfect false positive rate.

**3.11 ROC and AUC curve**

AUC - ROC curve is a performance measurement for the classification problems at various threshold settings. ROC stands for Receiver Operating Characteristic and AUC stands for Area under the curve. ROC is a probability curve and AUC represents the degree or measure of separability. It tells how much the model is capable of distinguishing between classes. Higher the AUC, the better the model is at predicting 0 classes as 0 and 1 classes as 1. By analogy, the Higher the AUC, the better the model is at distinguishing between patients with the disease and no disease. The ROC curve is plotted with TPR against the FPR where TPR is on the y-axis and FPR is on the x-axis.

* **True Positive Rate (TPR)**

The True Positive Rate (TPR), also known as Sensitivity or Recall, is the ratio of correctly predicted positive observations to the total actual positives. It is calculated as:

**TPR = [TP/(TP+FN)],** where TP is the number of true positives and FN is the number of false negatives.

It measures the proportion of actual positives that are correctly identified by the model.

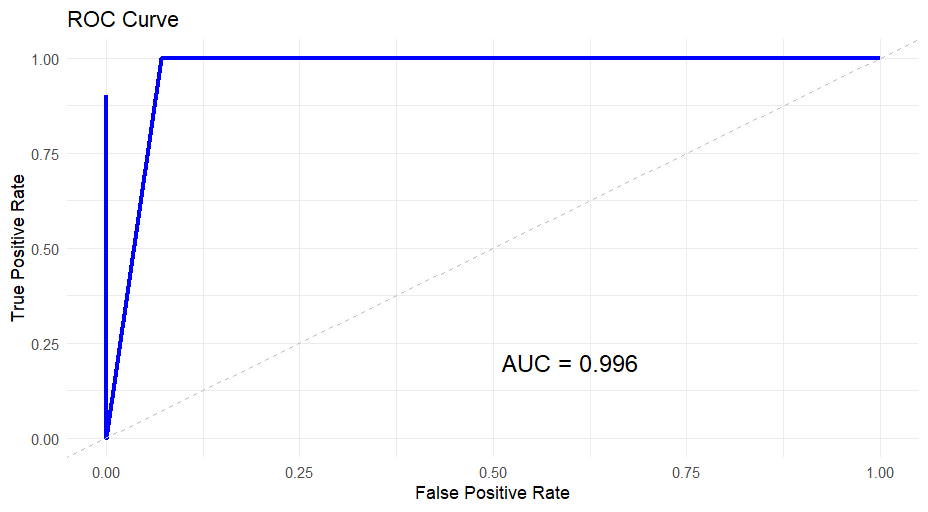
* **False Positive Rate (FPR)**

The False Positive Rate (FPR) is the ratio of incorrectly predicted positive observations to the total actual negatives. It is calculated as:

**FPR = [FP/(FP+TN)],** where FP is the number of false positives and TN is the number of true negatives.

FPR measures the proportion of actual negatives that are incorrectly identified as positive by the model.

**3.11.1 ROC Curve for model 1**



**Figure – 23:** ROC Curve for model

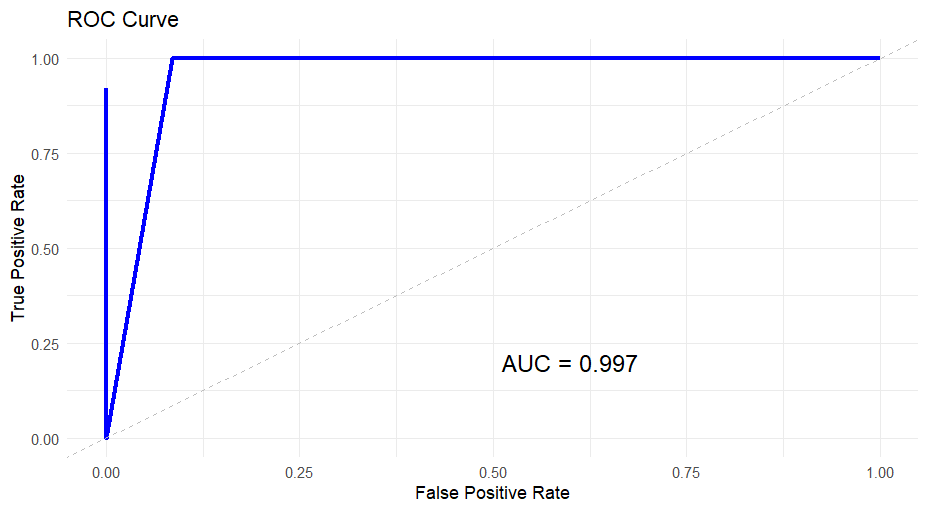
Here,

* AIC: 89.2658611732685
* AUC: 0.996

**Interpretation:** The ROC plot shows how well the logistic regression model distinguishes between male and female respondents based on predictors like age, occupation, education, income, EV preference, and government incentives. The curve's shape and the high AUC value (0.996) indicate excellent classification performance.

An AIC of **89.27** suggests the model fits the data well with a reasonable level of complexity. Lower AIC values are preferred when comparing models—so this score reflects a strong balance between accuracy and simplicity.

**3.11.3 ROC Curve for model 2**



**Figure – 24:** ROC Curve for model 2

Here,

* AIC: 89.0815865038871
* AUC: 0.997

**Interpretation:** The ROC plot shows how well the logistic regression model distinguishes between Urban and Rural respondents based on predictors like age, occupation, education, income, belief in EV impact, and EV purchase intention. The curve’s shape and the high AUC value (0.997) indicate excellent classification performance.

An AIC of 89.08 suggests the model fits the data well with a reasonable level of complexity. Lower AIC values are preferred when comparing models—so this score reflects a strong balance between accuracy and simplicity.

**3.12** **Principal Component Analysis**

Principal component analysis, or PCA, is a dimensionality reduction method that is often used to reduce the dimensionality of large data sets, by transforming a large set of variables into a smaller one that still contains most of the information in the large set.

* + 1. **Basic Terminologies of PCA in Machine Learning**

Before getting into PCA in machine learning, we need to understand some basic terminologies,

* **Variance**: For calculating the variation of data distributed across the dimensionality of the graph
* **Covariance**: Calculating dependencies and relationship between features
* **Standardizing data**: Scaling our dataset within a specific range for unbiased output
* **Covariance matrix**: Used for calculating interdependencies between the features or variables and also helps in reducing it to improve the performance
* **EigenValues and EigenVectors**: The eigenvectors aim to find the largest dataset variance to calculate the Principal Component. Eigenvalue means the magnitude of the Eigenvector. The eigenvalue indicates variance in a particular direction, whereas the eigenvector expands or contracts the X-Y (2D) graph without altering the direction.
* **Dimensionality Reduction**: Transpose of original data and multiply it by transposing of the derived feature vector. Reducing the features without losing information.

**3.12.2 How Does PCA Work?**

* The steps involved for PCA in ML are as follows-
* Original Data
* Normalize the original data (mean =0, variance =1)
* Calculating covariance matrix
* Calculating Eigen values, Eigen vectors, and normalized Eigenvectors
* Calculating Principal Component (PC)
* Plot the graph for orthogonality between PCs

**3.12.3 Types of PCA plot**

* **Loading plot -** The loading plot graphs the coefficients of each variable for the first component versus the coefficients for the second component.
* Scree Plot - The scree plot is used to determine the number of factors to retain in an exploratory factor analysis (FA) or principal components to keep in a principal component analysis (PCA). A scree plot always displays the eigenvalues in a downward curve, ordering the eigenvalues from largest to smallest. According to the scree test, the "elbow" of the graph where the eigenvalues seem to level off is found and factors or components to the left of this point should be retained as significant.
* Biplot - A biplot uses points to represent the scores of the observations on the principal components, and it uses vectors to represent the coefficients of the variables on the principal components.
* Score plot - The score plot project the observations onto a pair of PCs.

**3.12.4 Interpretation**

Here, I have used age, gender, occupation, educational qualification and family income as variables for plotting the PCA and finding out the correlation matrix.

* **Eigen analysis of the Correlation Matrix (Table: 15)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Eigen value | 1.1453 | 1.1172 | 0.9775 | 0.9547 | 0.8052 |
| Proportion | 0.229 | 0.223 | 0.196 | 0.191 | 0.161 |
| Cumulative | 0.229 | 0.453 | 0.648 | 0.839 | 1.000 |

The eigen values are related to the variances of the variables on which the correlation matrix is based. This table provides results from an eigen analysis of a correlation matrix, typically used in principal component analysis (PCA). Here's a breakdown of what each row represents:

* **Eigenvalue:** These are the eigenvalues of the correlation matrix. They indicate the amount of variance captured by each principal component. Higher eigenvalues mean that the component explains more variance in the data.
* **Proportion:** This is the proportion of the total variance explained by each principal component. It is calculated by dividing each eigenvalue by the sum of all eigenvalues.
* **Cumulative:** This shows the cumulative proportion of variance explained by the principal components up to that point. It helps in understanding how much of the total variance is explained as we include more principal components.
* **Interpreting the Table**

The first eigenvalue is **1.1453**, which explains **22.9%** of the total variance.  
The first two components together explain **45.3%** of the total variance.  
The first three components explain **64.8%** of the total variance.  
By the fourth component, **83.9%** of the total variance is explained.  
All five components together explain **100%** of the variance, as expected.

This table helps in deciding how many principal components to retain for dimensionality reduction or further analysis. Components with lower eigenvalues contribute less to explaining the variance and might be dropped to simplify the model without significant loss of information.

* **Principal Component Loadings Table (Table: 16)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **PC1** | **PC2** | **PC3** | **PC4** | **PC5** |
| Gender | -0.079 | -0.246 | 0.924 | -0.004 | -0.283 |
| Age | 0.151 | 0.648 | 0.243 | 0.683 | 0.177 |
| Occupation | -0.458 | 0.596 | -0.062 | -0.290 | -0.589 |
| Highest educational qualification | -0.715 | 0.061 | 0.158 | -0.130 | 0.666 |
| Family income | -0.500 | -0.400 | -0.243 | 0.657 | -0.314 |

This table represents the eigenvectors resulting from a Principal Component Analysis (PCA) on a dataset. Each row corresponds to a variable in the dataset, and each column (PC1, PC2, PC3, PC4, PC5) corresponds to a principal component. The values in the table indicate the loadings of each variable on the principal components.

Here's a brief explanation of each component:

* **PC1, PC2, PC3, PC4, PC5:** These are the principal components. Each principal component is a linear combination of the original variables, designed to capture the maximum amount of variance in the data.
* **Variable:** This lists the original variables that were analysed (e.g., gender, age, occupation, education, family information).
* **Values in the table:** These are the coefficients (loadings) for each variable on the respective principal components. They indicate the contribution of each variable to the principal

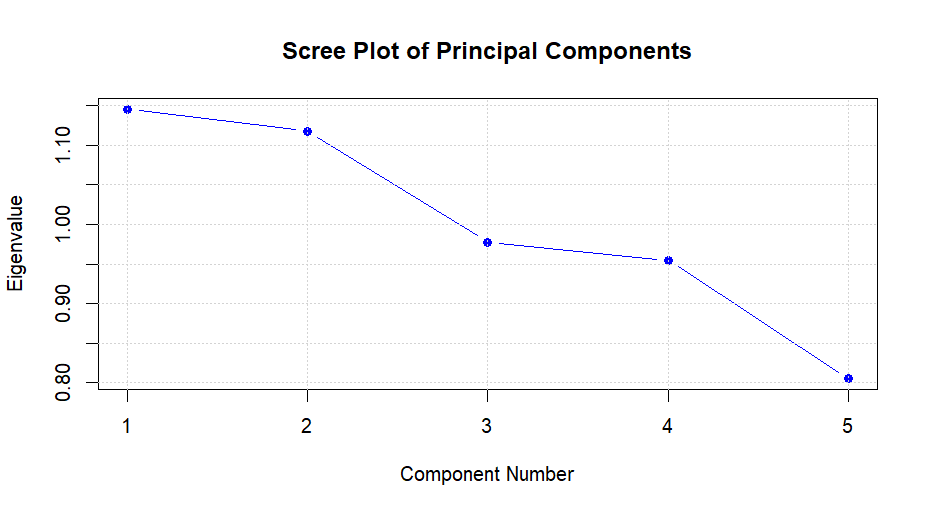
component. Higher absolute values suggest that the variable has a stronger influence on that component.

**In this context:**

* **Gender** has a strong loading on **PC3**.
* **Age** has strong loadings on **PC2** and **PC4**.
* **Occupation** has significant loadings on **PC1**, **PC2**, and **PC5**.
* **Education** has strong loadings on **PC1** and **PC5**.
* **Family income** has strong loadings on **PC1** and **PC4**.

These loadings help interpret the principal components and understand the underlying structure of the data.

* **Now we are plotting PCA graphs:**



**Figure – 25:** Scree Plot

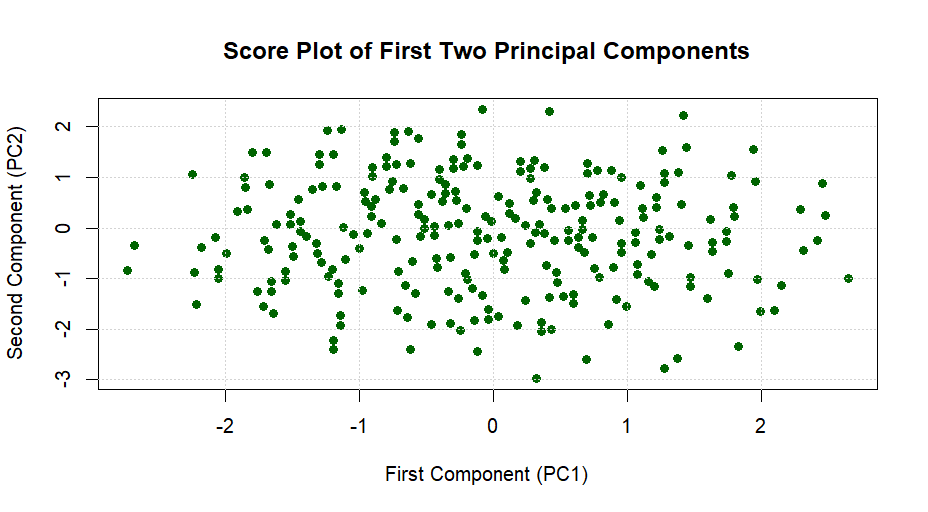
This graph is a scree plot, which is typically used in principal component analysis (PCA). It shows the eigenvalues associated with each principal component. Here's what each element of the graph represents:

- **Component Number (X-axis)**: This represents the principal components, ordered from the first to the last.

- **Eigen value (Y-axis)**: This represents the amount of variance explained by each principal component.

In a scree plot, the eigenvalues are plotted in descending order. The "elbow" or the point where the slope of the curve decreases sharply indicates the number of components that should be retained. The components to the left of this point account for most of the variance in the data, while the components to the right contribute less and may be considered for exclusion.

**Interpretation:** Here on the x-axis we’ve component numbers and on the y-axis we have eigenvalues. In the given scree plot, it looks like the eigenvalues start to level off after the **third component**, suggesting that the **first three components explain most of the variance** in the data.



**Figure – 26:** Score Plot

This graph is a score plot, which is also commonly used in principal component analysis (PCA). Here's what each element of the graph represents:

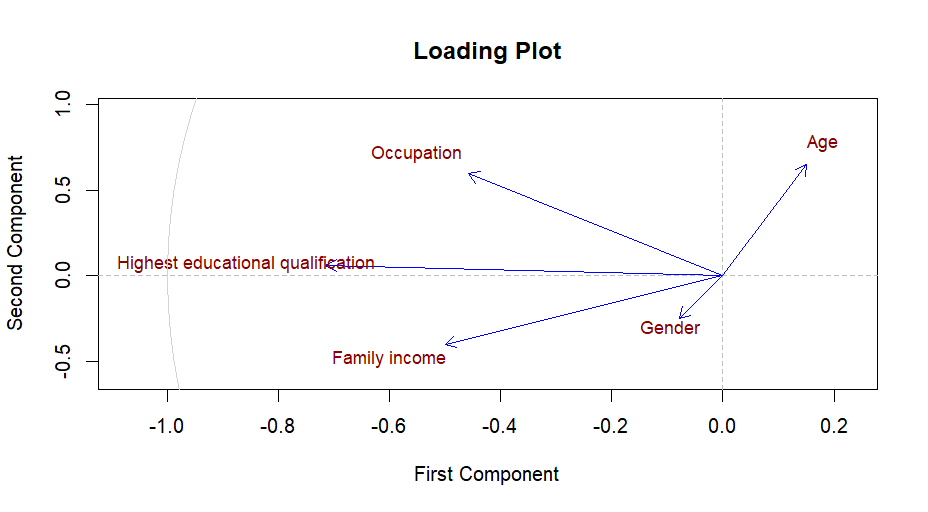
**- First Component (X-axis):** This represents the scores for the first principal component.

**- Second Component (Y-axis):** This represents the scores for the second principal component.

In a score plot, each point represents an observation (or a sample) in the new principal component space. The position of each point is determined by its scores on the first and second principal components.

The score plot is useful for visualizing the distribution and clustering of observations. It can help identify patterns, such as clusters or outliers, and can provide insights into the structure of the data. For example, points that are close together in the plot are similar in terms of the principal components, while points that are far apart are dissimilar.

**Interpretation:** Here on the x-axis we’ve first component and on the y-axis we have second component. In this plot, you can observe the spread and clustering of the data points along the first and second principal components. This can give you an idea of how the observations are distributed and how they relate to each other in the reduced dimensionality space



**Figure – 27:** Loading Plot

This graph is a loading plot from a Principal Component Analysis (PCA). Here's what it represents:

1. **Axes**: The axes represent the first and second principal components. These are linear combinations of the original variables that explain the most variance in the data.
2. **Vectors**: The vectors (lines) represent the loadings of each original variable on the first and second principal components. The direction and length of these vectors indicate the contribution of each variable to the principal components.

* **Gender**: Slightly positively correlated with the first component and slightly negatively correlated with the second component.
* **Age**: Positively correlated with the first component and slightly positively correlated with the second component.
* **Highest educational qualification**: Negatively correlated with the first component and slightly positively correlated with the second component.
* **Family Income**: Negatively correlated with the first component and slightly negatively correlated with the second component.
* **Occupation**: Negatively correlated with the first component and positively correlated with the second component.

**Interpretation:** Here on the x-axis we’ve first component and on the y-axis we have second component. The plot helps to understand the relationships between the variables and how they contribute to the principal components. Variables pointing in the same direction are positively correlated, while those pointing in opposite directions are negatively correlated. In summary, this loading plot shows how each variable contributes to the principal components and their correlations.

**Conclusion: -**

In India, there may be a need for energy transition in automobiles because of the depletion of fossil fuels assets and the consistent rise in gas fees. The government has taken steps to reduce pollution stages with the aid of promoting electric automobiles and providing buying subsidies .

The authorities has relaxed FDI regulations in order to promote output. EVs are being introduced in India by a number of new brands.

Governments and manufacturers should work together to construct the infrastructure and create positive climate for electric motors.

Government already have taken some incentives for the consumers such as

**1. Financial Incentives:** Governments often provide subsidies, tax credits, and rebates to reduce the purchase price of electric vehicles (EVs), making them more affordable for consumers.

**2.Infrastructure Development :** Investment in charging infrastructure, including public charging stations and home charging incentives, to support the widespread adoption of EVs.

**3. Regulatory Support**: Implementation of policies such as emission regulations, EV mandates, and benefits like reduced registration fees or access to carpool lanes to encourage the transition from traditional vehicles to EVs.

The respondents are aware of global climate conditions and are ready to change but their cost is an important factor while considering the purchase of EV. If sufficient infrastructure is available, respondents are inclined to simply accept EV’s as future buying option.

The cost of replacement of battery, Lack of charging station, Long recharging time and mileage are all major factors that limit consumer confidence.

If all these factors can be taken into consideration, then consumption of EV will increase which will reduce pollution and the problem of global warming will also reduce.

**REFERENCES: -**

* Gupta, S.C. and Kapoor, V.K.: Fundamentals of Mathematical Statistics, Sultan Chand and Sons publications.
* Gupta, S.C. and Kapoor, V.K.: Fundamentals of applied statistics, Sultan chand and sons publications.
* Principal Component Analysis in Machine Learning

(Link:- <https://www.analyticsvidhya.com/blog/2022/07/principal-component-analysis-beginner-friendly/>

and

<https://bioturing.medium.com/how-to-read-pca-biplots-and-scree-plots-186246aae063>)

* Fisher Individual Tests for Differences of Means

(Link:- <https://support.minitab.com/en-us/minitab/help-and-how-to/statistical-modeling/anova/how-to/one-way-anova/interpret-the-results/all-statistics-and-graphs/pairwise-comparisons/#:~:text=than%20Blend%202.-,Fisher%20Individual%20Tests%20for%20Differences%20of%20Means,the%20differences%20are%20practically%20significant>.)

* Regression Model for Ordinal Data , Peter Mccullagh , Journal of the Royal Statistical Society Series B,1980
* Categorical Data Analysis , Alan Agresti , 1990,John Wiley Sons
* Universal Business School, Mumbai Research paper ---- A study on Consumer Buying Behaviour towards Electric Vehicle (Link : <https://www.ijisrt.com/assets/upload/files/IJISRT22JUN196_(1).pdf>)

**ACKNOWLEDGEMENT**

I would like to express my sincere gratitude to all these individuals for mentoring and supporting me in completing this project ***“Consumer Behaviour Towards Buying An Electric Vehicle.”***

First and foremost, I want to thank my Assistant Professor of the department, **Prof. Anish Darnal**, for the continuous guidance, encouragement, and for giving invaluable suggestion at all stages for completion of this project. I am grateful for the knowledge and skills I have gained under his mentorship.

Again, I want to say my sincere thanks to my Associate professor and Head of Department, Prof**. Dr. Archana Das Roy**, for providing facilities existing in the department during this project work.

Last but not the least, I am very thankful to my parents for providing me their valuable time and financial help for completing this project work.