**MARKET SEGMENTATION ANALYSIS OF ELECTRIC VECHICLES**

**MARKET**

**NAME: MANISH C**

**DATE: 16/04/2023**

**OVERVIEW OF ELECTRIC MARKET**

An electric vehicle (EV) is a vehicle that uses one or more electric motors for propulsion. It can be powered by a collector system, with electricity from extravehicular sources, or it can be powered autonomously by a battery (sometimes charged by solar panels, or by converting fuel to electricity using fuel cells or a generator). EVs include, but are not limited to, road and rail vehicles, surface and underwater vessels, electric aircraft and electric spacecraft.

EVs first came into existence in the mid-19th century, when electricity was among the preferred methods for motor vehicle propulsion, providing a level of comfort and ease of operation that could not be achieved by the gasoline cars of the time. Internal combustion engines were the dominant propulsion method for cars and trucks for about 100 years, but electric power remained commonplace in other vehicle types, such as trains and smaller vehicles of all types.

In the 21st century, EVs have seen a resurgence due to technological developments, and an increased focus on renewable energy and the potential reduction of transportation's impact on climate change and other environmental issues. Project Drawdown describes electric vehicles as one of the 100 best contemporary solutions for addressing climate change.

Government incentives to increase adoption were first introduced in the late 2000s, including in the United States and the European Union, leading to a growing market for the vehicles in the 2010s. Increasing public interest and awareness and structural incentives, such as those being built into the green recovery from the COVID-19 pandemic, is expected to greatly increase the electric vehicle market. The International Energy Agency said in 2021 that governments should do more to meet climate goals, including policies for heavy electric vehicles. Electric vehicle sales may increase from 2% of global share in 2016 to 30% by 2030. Much of this growth is expected in markets like North America, Europe and China; a 2020 literature review suggested that growth in use of electric 4-wheeled vehicles appears economically unlikely in developing economies, but that electric 2-wheeler growth is likely. There are more 2 and 3 wheel EVs than any other type.

**PROBLEM STATEMENT (E V MARKET)**

Task is to analyze the Electric Vehicles Market using *Segmentation* analysis and come up with a feasible strategy to enter the market, targeting the segments most likely to use their product in terms of Geographic, Demographic, Psychographic, and Behavioral.

The three most feasible segmentation based on the availability of data are:

1. Geographic segmentation.
2. Vehicle feature segmentation
3. Customer behavioural segmentation

**This Report and the python code project focused on Vechicle feature segmentation**

Vehicle feature segmentation is the process of dividing vehicles into distinct groups or segments based on their features or characteristics. This segmentation can be based on factors such as body type, performance, fuel efficiency, safety features, luxury features, entertainment systems, ADAS, and intended use. Vehicle feature segmentation helps manufacturers and buyers to categorize and compare different vehicles based on their unique features and characteristics, enabling buyers to find a vehicle that matches their specific needs and preferences and helps manufacturers to target specific markets and customer segments.

**DATA COLLECTION**

We will analyze some data related to various popular electric vehicles available in the Market.

Electric Vehicle Data

1. Vehicle Battery capacity (in KWH)
2. Acceleration (0-100) in Seconds
3. Top Speed in Km/hr
4. Range of Vehicle in km
5. The efficiency of Vehicle in Wh/km
6. Fast charge speed in km/hr
7. Drive Configuration
8. Number of seats
9. Price in Euro

Data Source on Kaggle; <https://www.kaggle.com/datasets/geoffnel/evs-one-electric-vehicle-dataset>

The data consists of 104 vehicles and there are some missing values as well. We can conduct various data analytics visualizations to understand the data and information. This gives us an idea about the market as a whole and overall data distribution. Often, this type of data gives insight into the market and lets businesses conduct market research. The electric vehicle market is growing at a fast pace, with proper investment and research, the field can be improved and a higher level of efficiency can be achieved.

**Data Pre-processing:**

Libraries used in python :

* Pandas
* Numpy
* Matplotlib
* Seaborn
* from sklearn. Model selection import train test split
* from sklearn. linear model import Linear Regression
* from sklearn. metrics import mean squared error, r2\_score

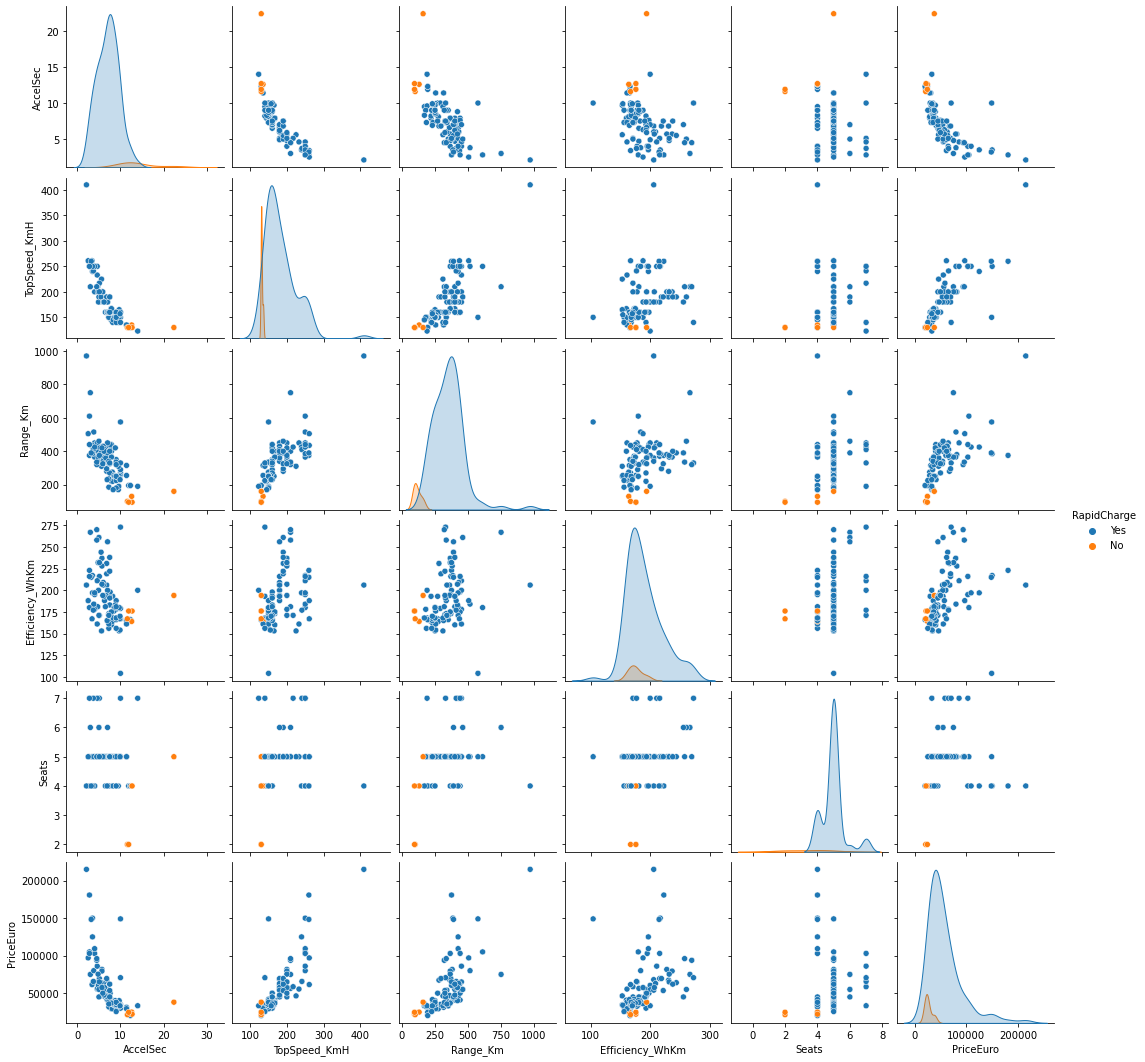
Steps involved:

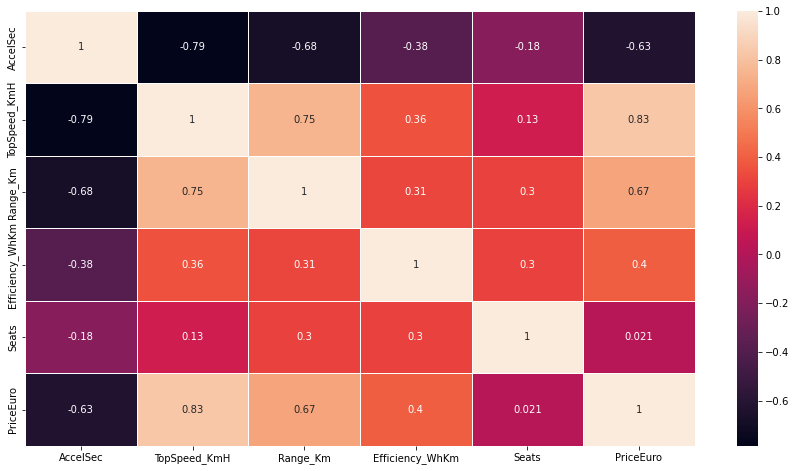
* Data manipulation
* Custom functions for plots
* Drop irrelevant columns

Data manipulation (for narrowing down to get relevant fields in the dataset):

This step involves filtering and selecting only the relevant fields or columns from the dataset. This is usually done to reduce the size of the dataset and to focus only on the information that is needed for the analysis.

Let us understand the correlation better using a heat map.

****

****

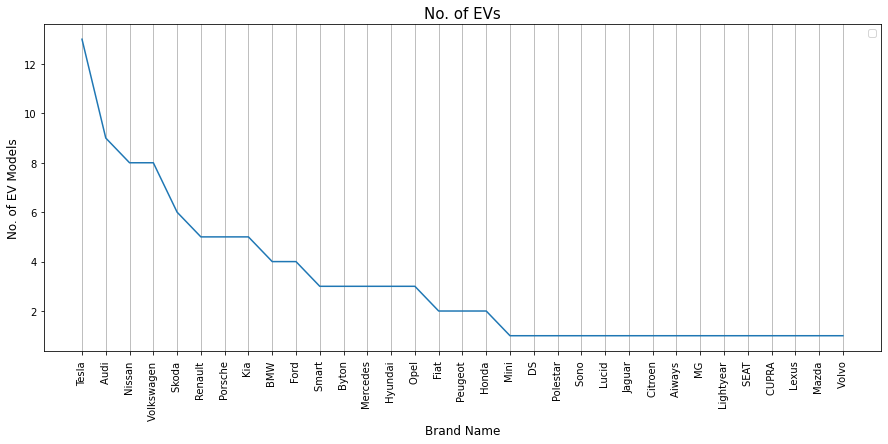
The lighter the colour, the more the correlation between the two data points

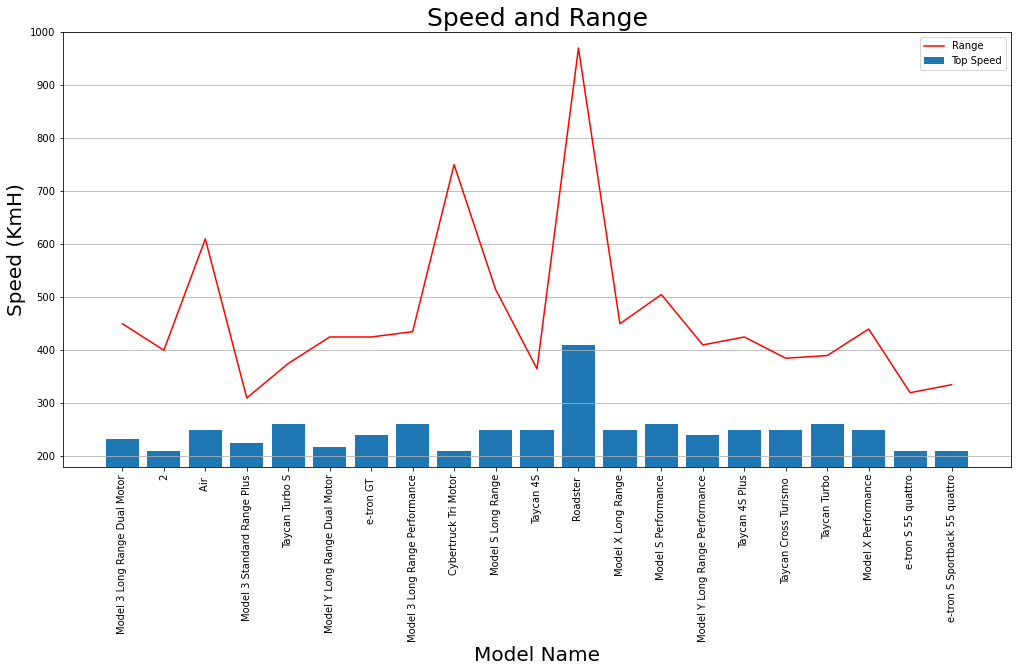
For example, Top Speed has a high correlation to price in Germany and price in the UK. So, it makes sense. A car that has a higher top speed will have a higher price.  Similarly, vehicle range and KWH has a high correlation of 0.89, it is obvious as the battery capacity, more will be the range of the vehicle. The correlation heatmap is thus, a great way to understand which data columns are related to each other.

Other data visualisations and various types of plots can be made to visualise the data distribution and understand the whole data. Doing proper data exploration can help in the overall data-driven decision-making process.

Now we understand about the Total number of models from respected brand and

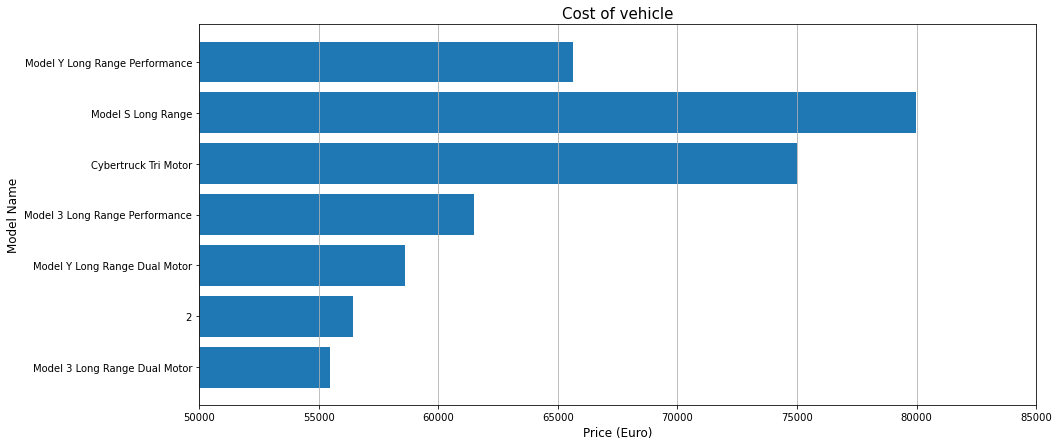
Top Speed and Range filtration





There are much of Models speed above 200kmH and subsequently with the range

Filteration according to Price Range

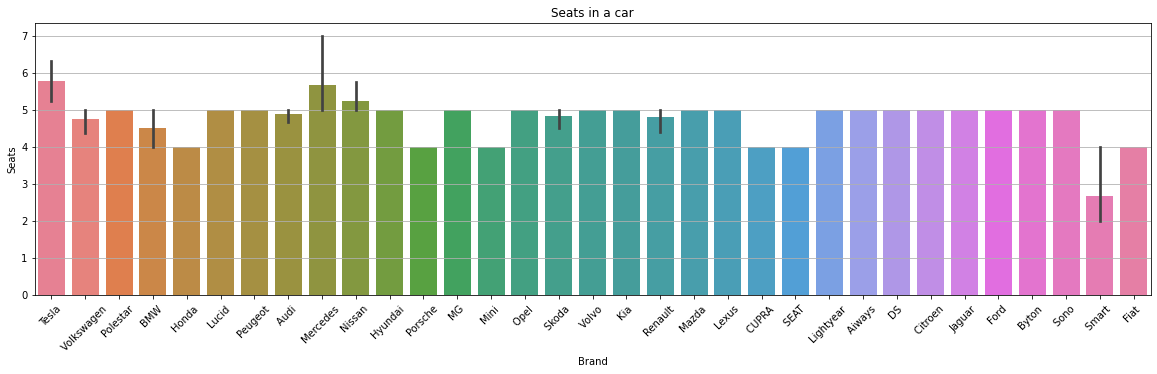


This shows the top speed of around more than 400KmH is the Roadster which is from tesla and other models that are quiet good in top speed

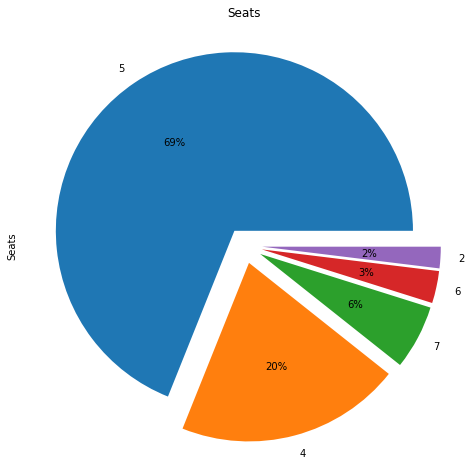
This is the list of vehicle with specifications that satisfies the condition of

eg. A person wnted to buy a vehicle which must be an EV with the top speed more than 200 KmH and range more than 400 Km. They are 4 person family and working so budget is around 50000 euros to 80000 euros (this is the budget more than 60% people look for)

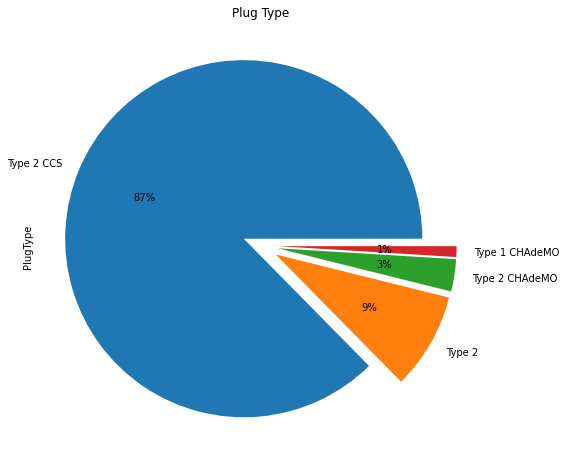
Now lets understand the number of seats in each car and there correlation and assume which car has the highest and lowest number of seats

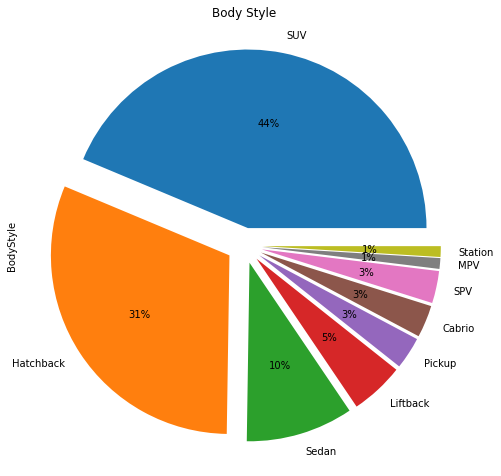


Mercedes, Tesla and Nissan have the highest number of seats and Smart the lowest



According to the above pie plot we can assume that the cars type which has 5 number of seats segments are have the most upto 69% from the observed data

Now lets understand about the type of plug used for charging, Cars and their body type from the given set of data

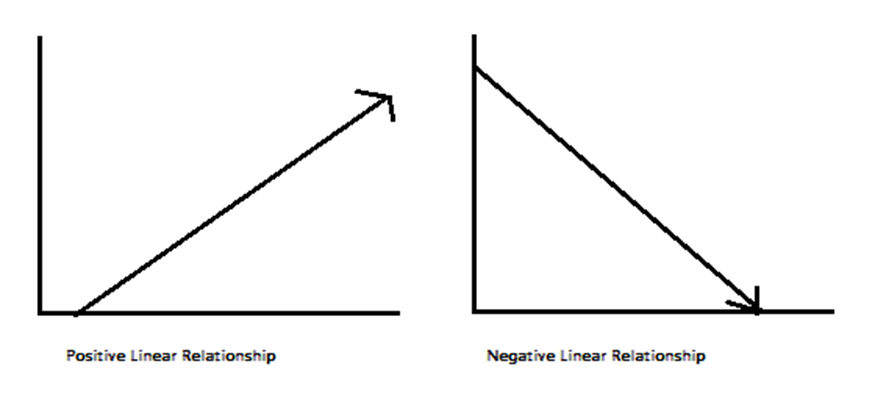


from the above plots we can assume that most of the cars brand uses Type 2css upto 87% and most cars brand produces SUV and hatchback of 44% and 31% respectively.

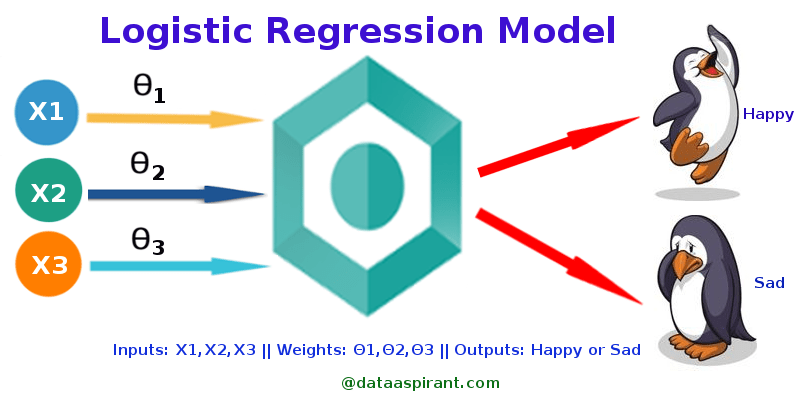
**SEGMENTATION ANALYSIS:**

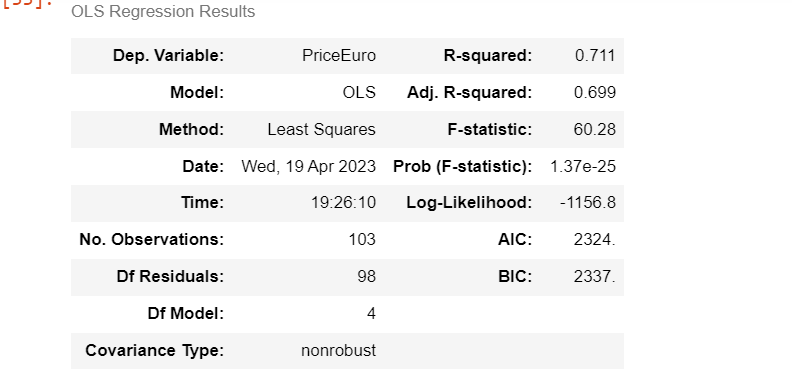
The above datasets is segmented by Linear Regression, Logistic Regression and by confusion matrix

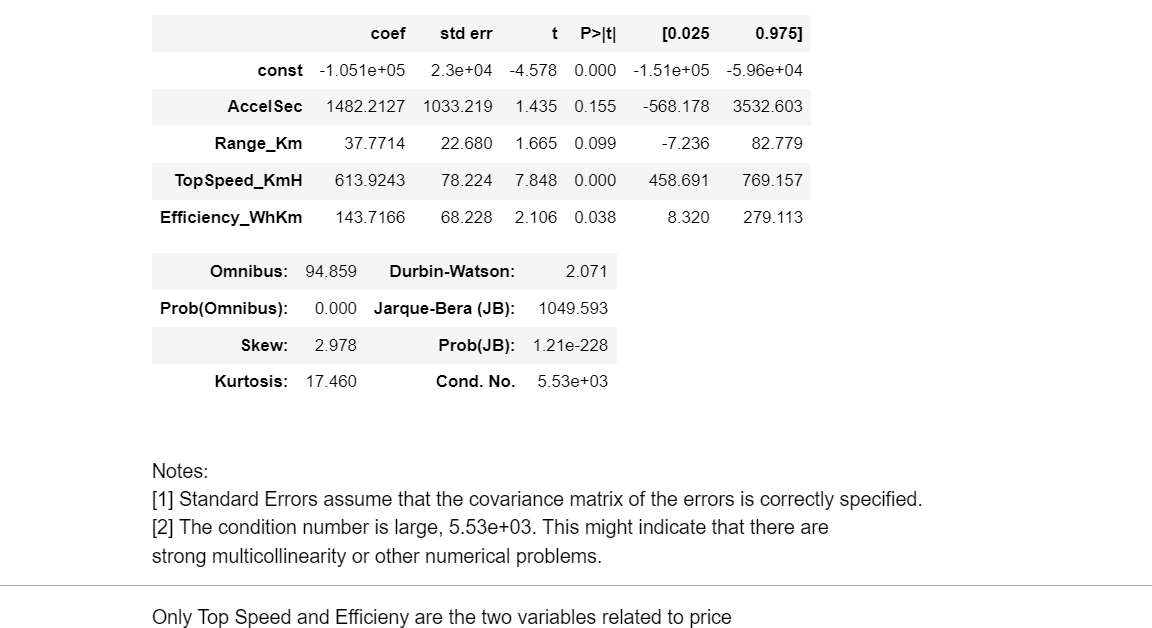
* **Linear Regression:** [Linear regression](https://en.wikipedia.org/wiki/Linear_regression) is a statistical model that examines the linear relationship between two (Simple Linear Regression ) or more (Multiple Linear Regression) variables — a dependent variable and independent variable(s). Linear relationship basically means that when one (or more) independent variables increases (or decreases), the dependent variable increases (or decreases) too:



* **Logistic Regression** : [**Logistic Regression**](https://en.wikipedia.org/wiki/Logistic_regression) is a Machine Learning classification algorithm that is used to predict the probability of a categorical dependent variable. In logistic regression, the dependent variable is a binary variable that contains data coded as 1 (yes, success, etc.) or 0 (no, failure, etc.). In other words, the logistic regression model predicts P(Y=1) as a function of X.







Above figure shows the fitting the model and summarizing.

* The R-squared value of this electric car dataset is 78.35225979903606
* Which means Around 78% of the dependant variable has been explained by the independent variables
* The Accuracy score of this electric car dataset is 95.23809523809523
* Which means the data is accurate upto 95%