STATISTICAL ANALYSIS OF SALES OF ELECTRIC VEHICLES

INTRODUCTION

The electric vehicle is on the verge of experiencing rapid growth in vehicle markets. The broad-scale adoption of the electric vehicle could bring significant changes for society in terms of not only the technologies we use for personal transportation, but also taking a major step to a cleaner and safer environment and moving our economies away from petroleum. In this project we have collected data on ‘Electric vehicle sales and petrol price from September 2020 to September 2021’ and on ‘Electric vehicle sales and number of charging stations in each state present till October 2021’, and we tried to find how petrol prices and number of charging stations affect the sales of electric vehicles.

We used a linear regression model and correlation (Pearson method) in R to conduct our analysis. And plotted graphs to represent linear regression models for both datasets.

CONTENTS

Modelling of the Problem

In this analysis we have taken the available datasets from previously done surveys by respected institutions. We have tried to implement on those datasets what we have learned in our statistics course this semester.

DATASET 1

We created a dataset which includes monthly sales of electric vehicles, petrol price and diesel price from September 2020 to September 2021.

And imported it in R using the following **command**:

```
1  library(readxl)
2  EV<- read_excel("EV_analysis.xlsx")
3  View(EV)
4
```

Output:

Filter

	Month	EV_Sales	Petrol_Price(₹)	Diesel_Price(₹)
1	2020-09-01	10637	82.08	73.56
2	2020-10-01	10897	81.06	70.63
3	2020-11-01	12858	81.06	70.46
4	2020-12-01	14978	82.34	72.42
5	2021-01-01	16212	83.71	73.87
6	2021-02-01	19111	86.30	76.48
7	2021-03-01	26200	91.17	81.47
8	2021-04-01	14174	90.56	80.87
9	2021-05-01	3309	90.40	80.73
10	2021-06-01	11149	94.49	85.38
11	2021-07-01	26127	98.81	89.18
12	2021-08-01	28906	101.84	89.87
13	2021-09-01	34316	101.34	88.77

PROBLEM 1: Plot the graph: Electric Vehicles sale Vs Petrol Price

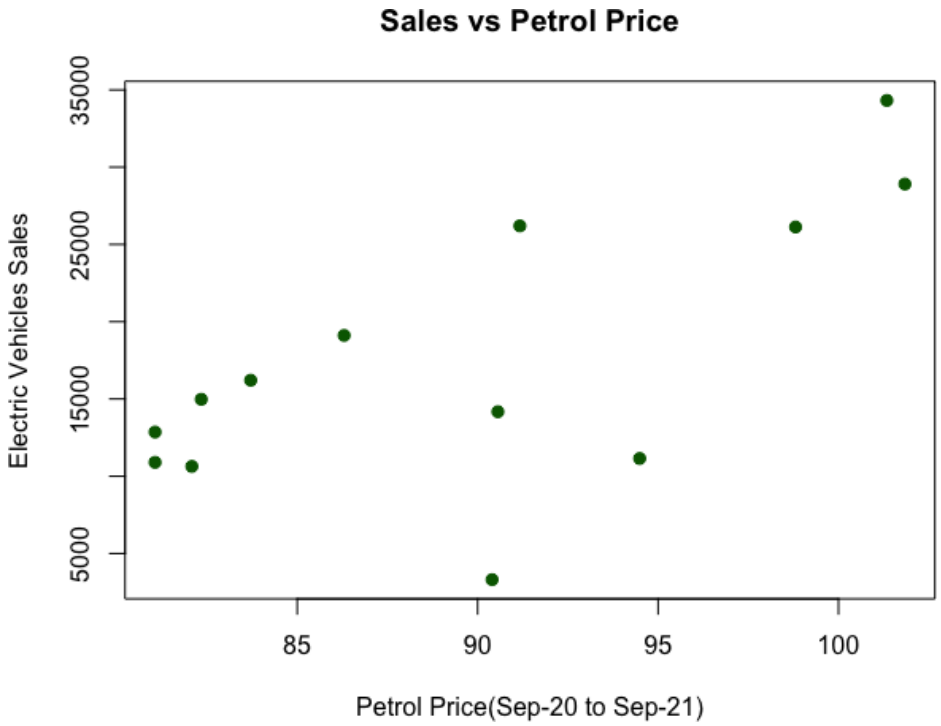
Command:

```
4
5  y=EV$EV_Sales
6
7  x=EV$`Petrol_Price(₹)`
8
9  plot(x,y, main="Sales vs Petrol Price",
10      xlab="Petrol Price(Sep-20 to Sep-21)",
11      ylab = "Electric Vehicles Sales",
12      pch=19,
13      col="dark green")
14
```

From the EV data frame we stored EV_sales in y and Petrol_Price in x.

And created a scatter plot of y and x using the function plot().

Plot:



PROBLEM 2: Plot the graph: Electric Vehicles sale Vs Diesel Price

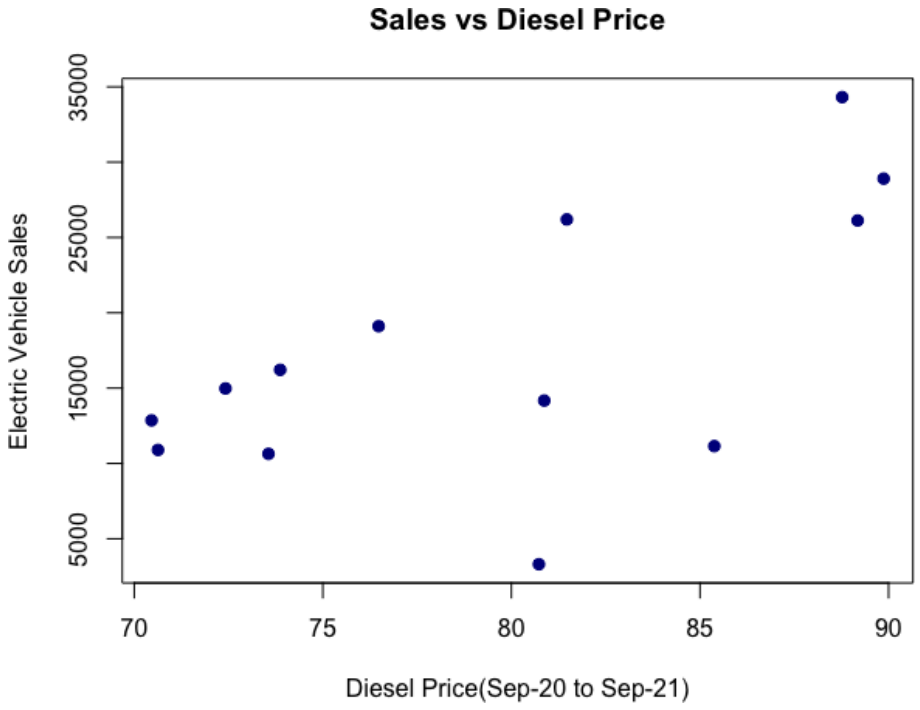
Command:

```
#.....Sales vs Diesel.............#

y=EV$EV_Sales

z=EV$`Diesel_Price(₹)`

plot(z,y, main="Sales vs Diesel Price",
     xlab="Diesel Price(Sep-20 to Sep-21)",
     ylab = "Electric Vehicle Sales",
     pch=19,
     col="dark blue")
```

From the EV dataframe we stored EV_sales in y and Diesell_Price in z.

And created a scatter plot of y and z using the function plot().

Plot:



PROBLEM 3: Create a Linear Regression Model and find relation between electric vehicles sale and petrol price

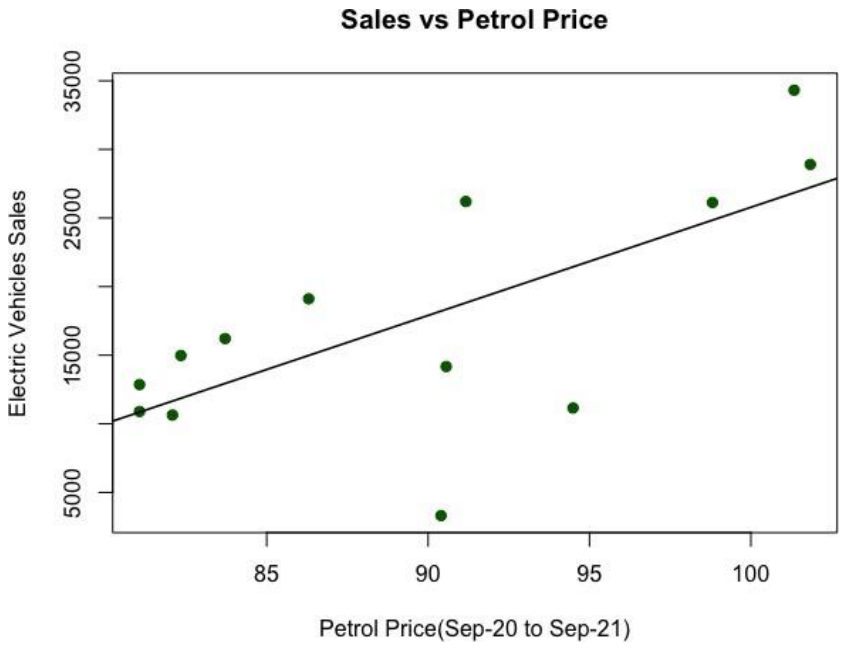
Command:

```
#.....linear_Regression.....#

mod=lm(y~x)
mod
abline(mod, col="black",lwd=1.5)

```

Output:

```
> mod=lm(y~x)
> mod

Call:
lm(formula = y ~ x)

Coefficients:
(Intercept)            x
   -53053.8         788.4

> abline(mod, col="black",lwd=1.5)
```

Plot:



PROBLEM 4: Find correlation.

Command:

```
23  #..............correlation.............#
24
25  cor(x,y,method = 'pearson')
26
```

Output:

```
> cor(x,y,method = 'pearson')
[1] 0.6814023
>
```

DATASET 2

We collected data which includes State Wise sales of electric vehicles in October 2021. And the number of charging stations present in those states as a representation of the infrastructure present in those states. **We know infrastructure is a factor in EV sales in India.**

And imported it in R using the following **command**:

```
1  library(readxl)
2  sales <- read_excel("SalesVsChargingStations.xlsx")
3  View(sales)
4
```

Output:

MTE_Project.R | EV | MTE_proj_pb2.R | sales

Filter

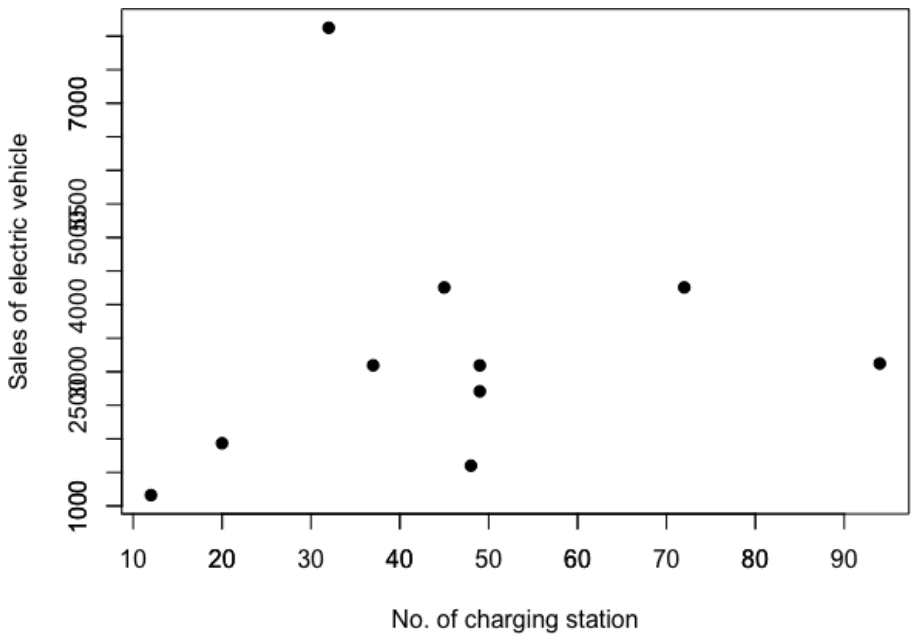
	STATE/UTs	NO._OF_ CHARGING_STATIONS(Units)	SLAES_OF_EV_IN_OCTOBER2021(Units)
1	DELHI	94	3122
2	UTTAR PRADESH	32	8124
3	BIHAR	37	3095
4	KARNATAKA	45	4255
5	TAMIL NADU	49	2708
6	MAHARASHTRA	72	4256
7	ASSAM	20	1934
8	RAJASTHAN	49	3094
9	HARYANA	12	1161
10	GUJARAT	48	1600

PROBLEM 5: Plot the graph: Electric Vehicles sale Vs Number of Charging Stations.

Command:

```
y=sales$`SLAES_OF_EV_IN_OCTOBER2021(Units)`
x=sales$`NO._OF_ CHARGING_STATIONS(Units)`

bins=seq(1000,8000,500)
bins2=seq(0,90,10)

plot(x,y,
     xlab="No. of charging station",
     ylab="Sales of electric vehicle",
     pch=19)
axis(1,at=bins2,labels = bins2)
axis(2,at=bins,labels = bins)

```

Plot:



PROBLEM 6: Create a Linear Regression Model and find relation between electric vehicles sale and number of charging stations.

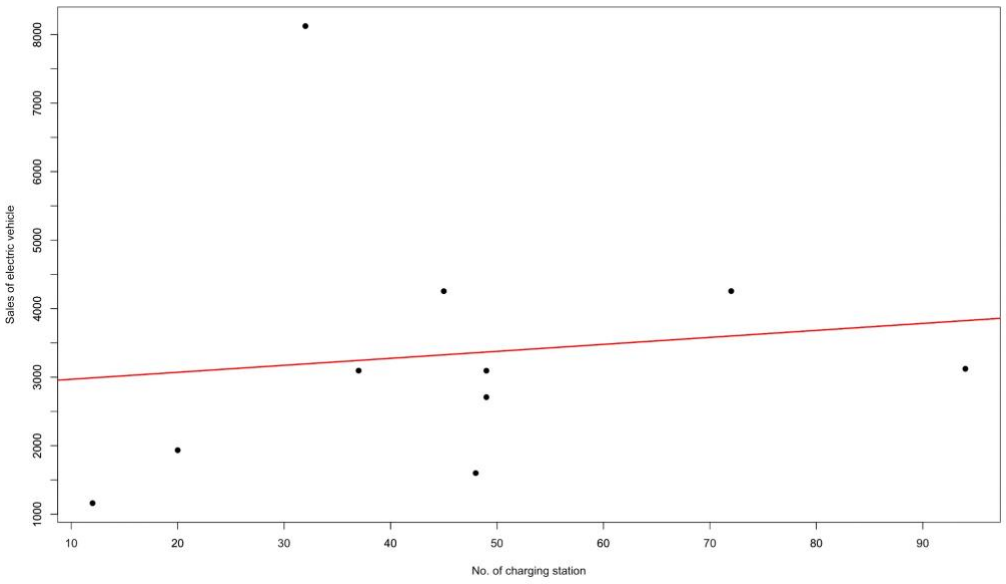
Command:

```
#.......linear_regression.............#

mod=lm(y~x)
mod
abline(mod,col="red", lwd=2)

```

Output:

```
> mod=lm(y~x)
> mod

Call:
lm(formula = y ~ x)

Coefficients:
(Intercept)            x
    2867.33        10.21

> abline(mod,col="red", lwd=2)
```

Plot:



PROBLEM 7: Find correlation.

Command:

```
23 #..............correlation.............#
24
25 cor(x,y,method = 'pearson')
26
```

Output:

```
> cor(y,x)
[1] 0.1235812
```

RESULT

DATASET 1

We plotted the data (in a simple scatterplot) and added the line we built with our linear model. R read the data first, with the read excel command, to create a data frame with the data, then create a linear regression with our data. The command plot takes two variables and plots them. In this case, it plots the Sales against the Petrol_Price of the Electric Vehicle. Then, add the line made by the linear regression with the command abline.

With the command summary(mod) we got the detailed information on the model's performance and coefficients.

- Intercept: -53053.8
- Coefficient of x: 788.4
  So, the line of regression is -> **y = -53053.8 + x (788.4)**

- p-value= 0.01033or 1.033% (which is less than 5%), thus petrol price is probably an excellent addition to our model.
- R-squared: 0.4643 (So the model can explain 46% of the total variability.)
- Correlation: 0.6814023

DATASET 2

We did the same operations on dataset 2 and got the following output on the model's performance and coefficient.

- Intercept: 2867.33
- Coefficient of x: 10.21
  So, the line of regression is -> **y = 2867.33 + x (10.21)**

- p-value= 0.7338 or 73.38% (which is very high), which means there's a 73% chance that this predictor is not meaningful for the regression.
- R-squared: 0.01527 (So the model can explain 1.5% of the total variability.)
- Correlation: 0.1235812

Thus from aur regression analysis, this predictor (number of charging stations present in each state) is not a meaningful predictor.

CONCLUSION

In this Statistical analysis project, we have tried to learn and understand about collection and process of data. The importance of type of data for any particular operation is significant. We have learnt that for analysing data and that too statistically, how careful you have to be before passing any observation.
While making this project we learned different functions of R and their implementation.
The institutions do a very thorough research and put in a lot of hard work to collect and process the data, and pass any observation or conclusion only after that. Otherwise all of this analytical process will lose its credibility. We have learnt that for making good observations you have to be specific about what you want to extract from your data.

REFERENCES

- https://www.petroldieselprice.com/petrol-price-today
- india-briefing.com/news/electric-vehicle-industry-in-india-why-foreign-investors-should-pay-attention-21872.html/
- https://www.datacamp.com/community/tutorials/linear-regression-R
- https://cea.nic.in/electric-vehicle-charging-reports/?lang=en
- https://auto.economictimes.indiatimes.com/news/two-wheelers/hero-electric-sells-over-6500-units-in-september-2021/86809740
- https://powermin.gov.in/sites/default/files/webform/notices/scan0016%20(1).pdf
- https://www.eqmagpro.com/electric-vehicle-charging-stations-across-india-state-wise-numbers/
- https://www.statisticshowto.com/probability-and-statistics/correlation-coefficient-formula/
- http://www.stat.yale.edu/Courses/1997-98/101/linreg.htm
- Class Notes