**STAT 401 – Final Project - Cassandra Davies**

**Section 1 - Introduction:**

Research Question – Which factors affect the price of electric vehicles?

Hypothesis – Certain factors do affect the price of electric vehicles more than others.

The data set for this project contains information relevant to the research question and hypothesis. Specifically, 360 samples with 9 different descriptors.

**Section 2 – Exploratory Data Analysis:**

The original data set for this project shows the prices for 360 electric vehicles. The original 9 descriptors are: car name, car name link, price, battery, efficiency, fast charge, range, top speed, and acceleration. Here’s a screenshot of the first 20 rows of the excel file, for reference:

Tabla

Descripción generada automáticamente

As you can see in row 18, not all of the data is present for every sample.

R Studio did a good job of automatically removing those from the models I built, as shown in Section 3.

**Section 3 – Regression Model Building:**

I built 5 different models. First, I set up the price variable as the response on the y-axis. Then, I added the following variables to share the x-axis: efficiency, fast charge, range, top speed, acceleration, battery. The summary for the first model can be found in the first image below. The second model is first model minus three specific data points that I chose to remove because they seemed to have high leverage. The summary for the second model (second image below) shows a lowered residual standard error and an increased adjusted R-squared value. The third model is the second model minus the “range” predictor variable. This did not improve the model’s residual standard error or the adjusted R-squared values, so I reinstated the “range” predictor variable in the fourth model. Also in the fourth model, I added an interaction term, “batRange”, which is the interaction between the “range” and “battery” predictor variables. This interaction term improved the residual standard error and the adjusted R-squared values. The fifth and final model is the fourth model minus the “acceleration” predictor variable. Out of the five models created in this project, the fifth model shows the lowest residual standard error, the highest adjusted R-squared, and the best significance codes for its variables, as shown in the figures below.

1 Texto, Tabla

Descripción generada automáticamente 2 Tabla

Descripción generada automáticamente

3 Texto, Carta

Descripción generada automáticamente

4 Texto

Descripción generada automáticamente 5 Texto

Descripción generada automáticamente

**Section 4 – Results And Conclusion:**

Gráfico, Gráfico de líneas

Descripción generada automáticamente Gráfico, Gráfico de dispersión

Descripción generada automáticamente Gráfico, Gráfico de dispersión

Descripción generada automáticamente Gráfico, Gráfico de dispersión

Descripción generada automáticamente I removed the three data points (330, 148, 60) to see if their exclusion improved the model, and I do think that it did. The model summary comparison can be found in Section 3.

**My conclusion is that the predictor variables that showed the strongest relationship to price were top speed, followed by efficiency and fast charge.**

Here are the final plots for the fifth model:

Gráfico, Gráfico de dispersión

Descripción generada automáticamente Gráfico, Gráfico de dispersión

Descripción generada automáticamente

Gráfico, Gráfico de dispersión

Descripción generada automáticamente Interfaz de usuario gráfica, Gráfico, Gráfico de líneas

Descripción generada automáticamente

**Section 5 – Further Questions:**

I do wonder if there are other factors that contribute to the price of electric vehicles, perhaps other factors that were not measured when this data set was made. It would be interesting to collect newer data and take in a larger number of predictor variables.

**Appendix:**

Here is the code script I used in R Studio to build these data models and analyze them.

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

The R file can also be found on my github repository <https://github.com/CEDavies>

This is my own unaided work.

* Cassandra Davies