Data Analysis on Electric Vehicle Specifications and Prices

Introduction

With heightened environmental concerns and a series of technological advancements, the global transportation landscape is undergoing a transformative shift as electric vehicles (EVs) are becoming more mainstream. This is especially true in Germany, wherein political parties such as The Greens, a party focused on environmental issues (Sloat, 2020), have managed to influence government policy on EVs. Various policy measures, such as tax incentives and subsidies, along with increased consumer awareness, have led to a massive increase in demand for these vehicles (EIA, 2020).

Unlike conventional vehicles that are reliant on gas or diesel-powered engines, EVs use an electric motor powered by batteries or a fuel cell (USCS, 2015). EVs are often more efficient and known to reduce air pollution compared to their traditional counterparts (USCS, 2015). Currently, global sales of EVs comprise one-fifth of the overall car market and more than 10 million vehicles were sold around the world in 2022 (IEA, 2023). As EV sales are expected to increase even more in the upcoming years and we are faced with the challenges of energy security, it is vital we understand EV specifications and their pricing. Through a linear regression analysis of data scraped from the Electric Vehicle Database (lhan, 2023), this study aims to further understand which factors drive the prices of EVs. In doing so, we hope to find correlations between EV specifications and their pricing, while also informing stakeholders on the various aspects that affect the EV market.

Research Question

Which specifications for electric vehicles influence their price in Euros in Germany, what is the size of their effects, and how statistically significant is each?

Description of Data

The dataset we chose is titled, “Electric Vehicle Specifications and Prices” and is provided by the EV Database. This dataset contains various EV information including price, efficiency ratings, battery capacity, and driving range. With 360 observations and 9 variables in total (7 of which are numeric), the dataset covers a diverse range of information that will enhance the understanding of consumers, manufacturers, and policymakers in the EV landscape understanding of the EV landscape.

Data Dictionary

- Battery: The capacity of the vehicle's battery in kilowatt-hours (kWh).
- Car name: The model name of the electric vehicle.

- Car name link: A direct link to the corresponding page on EV Database for more in-depth information.
- Efficiency: The energy efficiency rating of the vehicle in watt-hours per kilometer (Wh/km).
- Fast charge: The fast charging capability of the vehicle in minutes for a certain charging percentage.
- Price: The price of the electric vehicle in Euros (€) in Germany.
- Range: The driving range of the vehicle on a single charge in kilometers.
- Top speed: The maximum speed the vehicle can achieve in kilometers per hour.
- Acceleration 0-100: The acceleration time from 0 to 100 kilometers per hour.

Created Variables:

- Brand: The make of the EV.
- Average Price: The average price of all EVs of the same make in the dataset.

Data Exploration

Describing the Data - Summary Statistics:

In this section, we will perform an exploratory data analysis (EDA) on the Electric Vehicles specifications and prices. The goal of EDA is to develop a better understanding of the dataset through the identification of patterns, relationships, and anomalies. Through this analysis, we can make informed decisions on how to preprocess and model the data.

To get a simple overview of the numeric variables to explore what sort of ranges the data fell into, we used function summary() in R. From this we see that there is a wide range between min and max values for each category. We seek to understand how these ranges impact price and assume that there will be a positive correlation between higher values and higher prices.

```
> summary(ev_df[,c(-2, -3)])
    Battery         Efficiency      Fast_charge      Price.DE.          Range          Top_speed      acceleration..0.100.
 Min.   : 21.30   Min.   :137.0   Min.   : 170   Min.   : 22550   Min.   :135.0   Min.   :125.0   Min.   : 2.100
 1st Qu.: 57.50   1st Qu.:171.0   1st Qu.: 360   1st Qu.: 45690   1st Qu.:295.0   1st Qu.:155.8   1st Qu.: 4.900
 Median : 71.00   Median :188.0   Median : 520   Median : 56942   Median :380.0   Median :180.0   Median : 6.750
 Mean   : 71.19   Mean   :195.2   Mean   : 553   Mean   : 67264   Mean   :369.7   Mean   :180.9   Mean   : 7.289
 3rd Qu.: 85.00   3rd Qu.:208.2   3rd Qu.: 680   3rd Qu.: 73100   3rd Qu.:446.2   3rd Qu.:200.0   3rd Qu.: 9.000
 Max.   :123.00   Max.   :295.0   Max.   :1290   Max.   :218000   Max.   :685.0   Max.   :320.0   Max.   :19.100
```

This initial exploration displays the top performing cars for several variables. The Maserati GranTurismo Folgore, as well as the Tesla Model S Plaid and Lucid Air Dream Edition P, were some of the top performers in speed and acceleration. It is interesting to note that the Porche Taycan had the quickest acceleration rate, but was not considered one of the top 5 in terms of speed. Mercedes had the top 5 most efficient cars with energy efficiency ratings ranging from 290 to 295 Wh/km. Lastly, Lucid Motors was one of the top performers in speed and battery capacity.

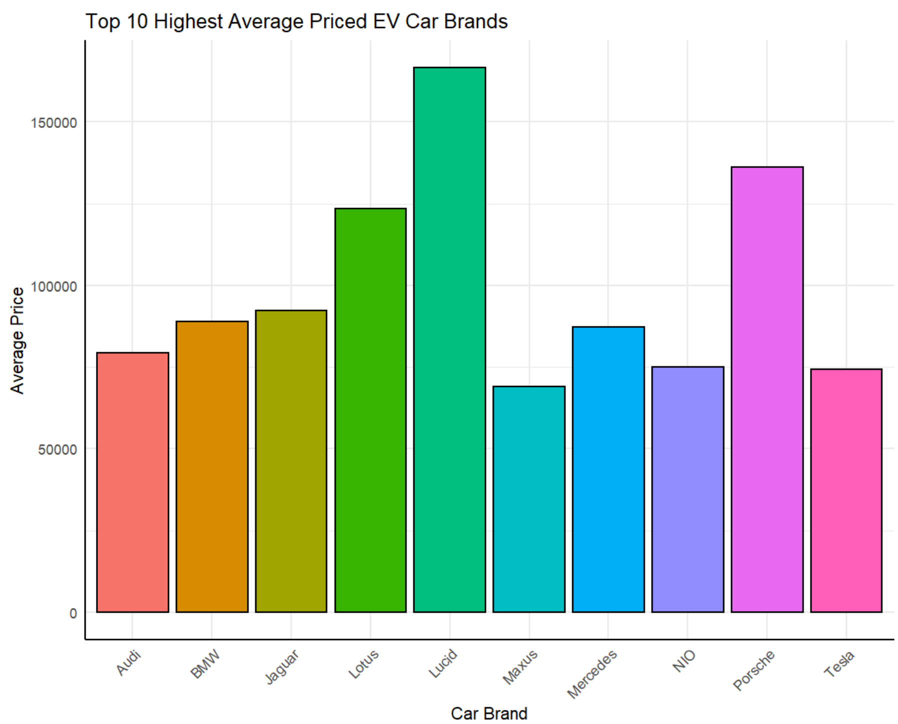
	Car_name	Top_speed
1	Maserati GranTurismo Folgore	320
2	Tesla Model S Plaid	282
3	Lucid Air Dream Edition R	270
4	Lucid Air Grand Touring	270
5	Lucid Air Dream Edition P	270

	Car_name	acceleration..0.100.
1	Porsche Taycan Turbo S Sport Turismo	2.8
2	Maserati GranTurismo Folgore	2.7
3	Lucid Air Dream Edition P	2.7
4	Tesla Model X Plaid	2.6
5	Tesla Model S Plaid	2.1

	Car_name	Efficiency
1	Mercedes eVito Tourer Extra-Long 90 kWh	295
2	Mercedes eVito Tourer Extra-Long 60 kWh	293
3	Mercedes EQV 300 Long	290
4	Mercedes eVito Tourer Long 90 kWh	290
5	Mercedes EQV 300 Extra-Long	290

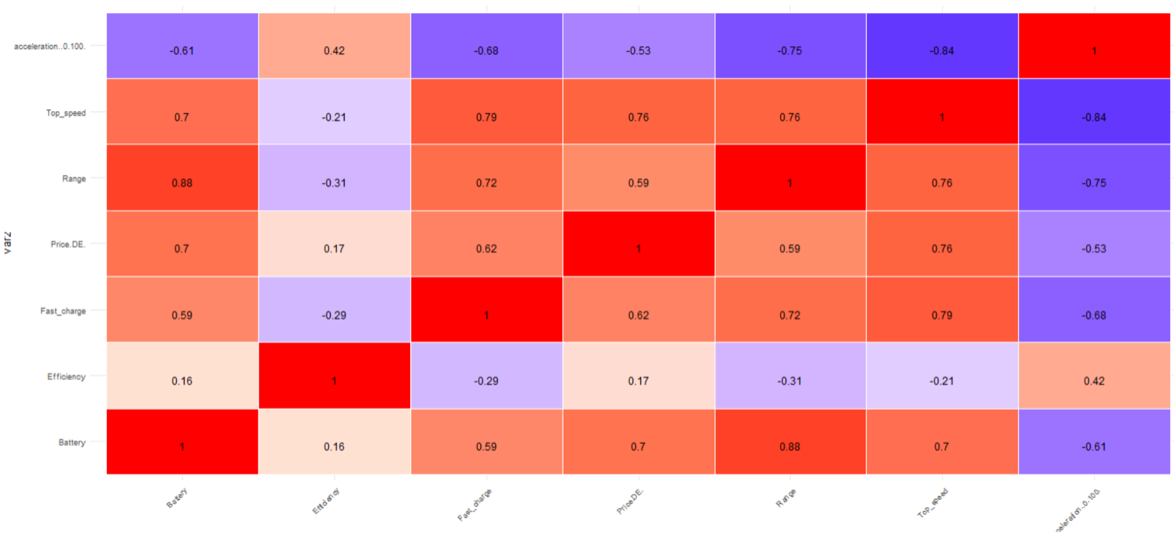
	Car_name	Battery
1	VinFast VF 9 Extended Range	123
2	Lucid Air Dream Edition R	118
3	Lucid Air Dream Edition P	118
4	Lucid Air Grand Touring	112
5	Lotus Eletre R	109

The following bar chart visualization depicts the top 10 highest priced EV car brands. To create this graph we created two new variable columns: brand and average price per brand. Out of the top 10, the three most expensive brands of EV were Lucid, Lotus, and Porsche.



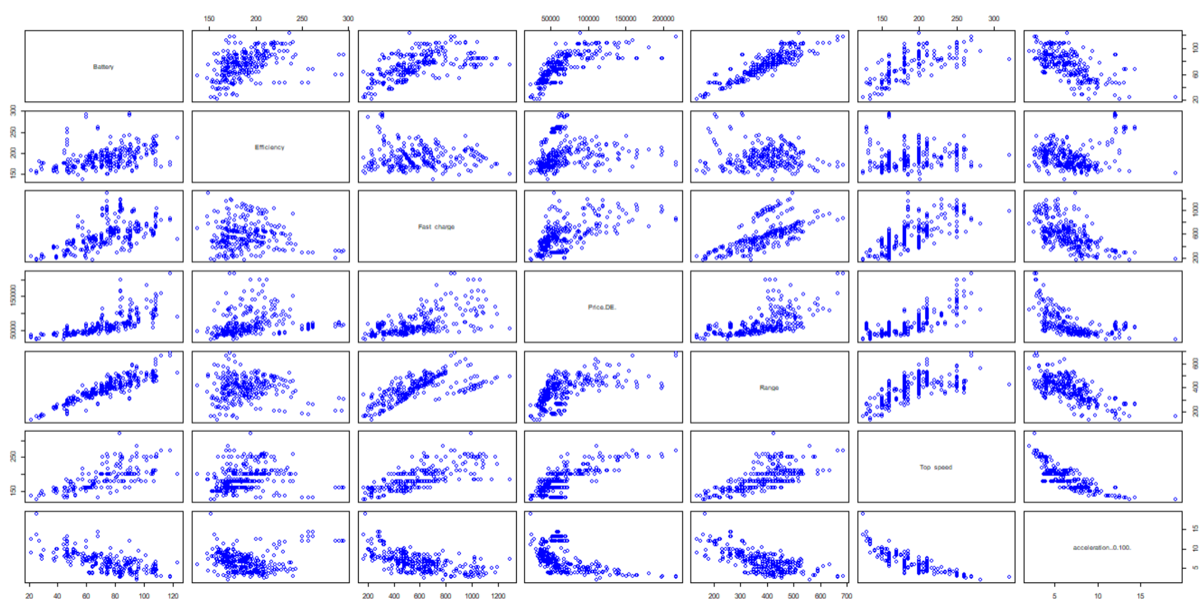
Correlation between Variables:

This correlation matrix showed us that *Range* and *Battery* were very highly correlated with a value of 0.88. To avoid multicollinearity, we decided to remove *Range* from our linear regression model.



Correlations Scatterplot Matrix:

Mostly all variables have a positive relationship except Acceleration 0-100, which has a negative relationship to all the other numeric variables. Efficiency has nearly no correlation with the other variables.



Hypothesis Testing:

Null Hypothesis: H_0: $\beta1=\beta2=\beta3=\beta4=\beta5=0$

Alternative Hypothesis: H_1: $\beta1$ or $\beta2$ or $\beta3$ or $\beta4$ or $\beta5 \neq 0$

Regression Model:

After running our regression analysis, we uncovered that Efficiency and Top Speed were the most statistically significant in impacting EV prices. Based on the multiple R squared values, the variables substantially explain the variation in our dependent variable and explain the model approximately 72%.

```
Call:
lm(formula = carsnorange$Price.DE. ~ ., data = carsnorange)

Residuals:
   Min     1Q Median     3Q    Max 
-53627 -11716   -143   8242  84828 

Coefficients:
                      Estimate  Std. Error t value             Pr(>|t|)    
(Intercept)        -158634.947   14227.588 -11.150 < 0.0000000000000002 ***
Battery                267.626      87.583   3.056              0.00245 ** 
Efficiency             291.038      45.659   6.374        0.00000000069 ***
Fast_charge             18.097       7.589   2.384              0.01772 *  
Top_speed              700.965      70.710   9.913 < 0.0000000000000002 ***
acceleration..0.100.  1780.757     760.505   2.342              0.01986 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 18490 on 301 degrees of freedom
  (53 observations deleted due to missingness)
Multiple R-squared:  0.7167,	Adjusted R-squared:  0.712 
F-statistic: 152.3 on 5 and 301 DF,  p-value: < 0.00000000000000022
```

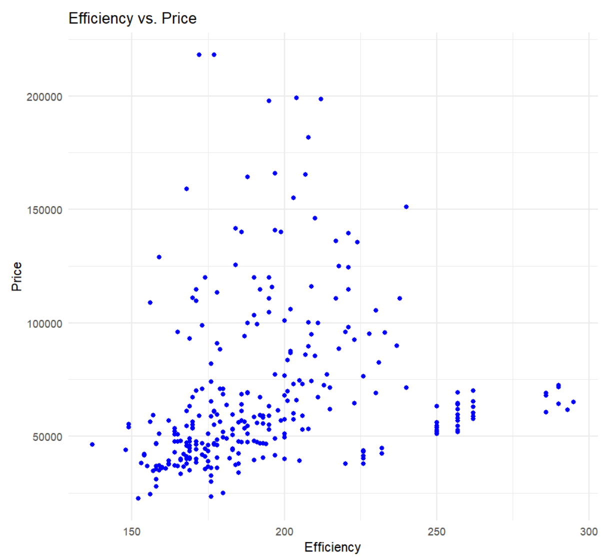
Combined Significance of Efficiency and Top Speed

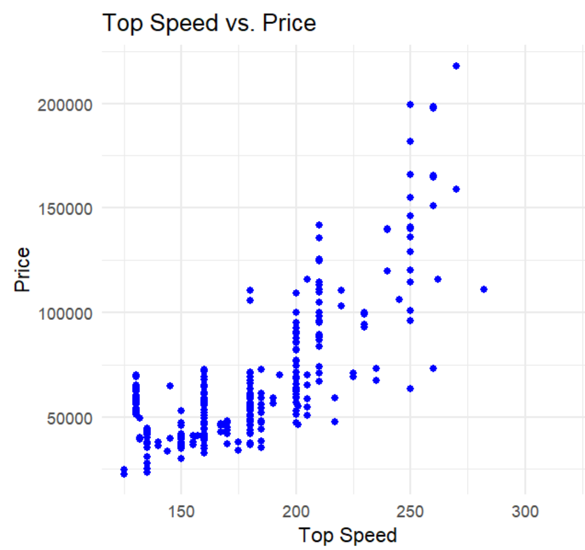
From the previous analysis, we saw that Efficiency and Top Speed were the most statistically significant individually, but we also wanted to evaluate if they have a combined significance. To check this, we ran a joint significance test and from running this test, we found they were in fact, very statistically significant when combined.

```
. test efficiency top_speed

 ( 1)  efficiency = 0
 ( 2)  top_speed = 0

       F(  2,   300) =   54.00
            Prob > F =    0.0000

. test efficiency top_speed fast_charge acceleration

 ( 1)  efficiency = 0
 ( 2)  top_speed = 0
 ( 3)  fast_charge = 0
 ( 4)  acceleration0100 = 0

       F(  4,   300) =   58.49
            Prob > F =    0.0000
```

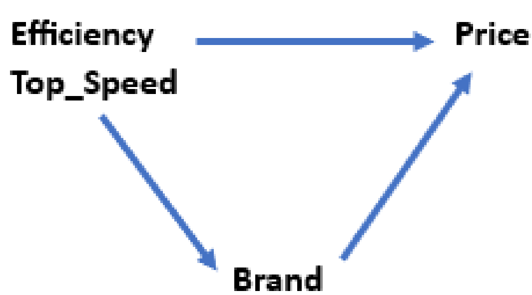
Impact of Brand on the pricing as a mediator:

When analyzing the impact of efficiency and top-speed in pricing, we saw that pricing moved almost in a similar pattern with efficiency and top-speed. However, there were certain instances, where we saw outliers with sudden spikes in prices in between. Looking at it deeper, we realized it was mostly dictated by certain specific brands.





So, we figured, brands also may have an impact on pricing and work as a mediator. In order to test out our hypothesis, we ran the analysis below taking average price per brand to evaluate the impact of brand on pricing.



Brand Impact Model

To explore the impact of brand on price we created the linear regression model below where y is price and the x variables are every brand of EV in the dataset. Lotus, Lucid, and Porche were found to be the most statistically significant when impacting price at the 0.001 level. These brands were followed by BMW and Mercedes at the 0.01 level of significance.

```
> carbrandmodel <- lm(cars$Price.DE. ~ cars$Brand, data = cars)
> summary(carbrandmodel)

Call:
lm(formula = cars$Price.DE. ~ cars$Brand, data = cars)

Residuals:
   Min     1Q Median     3Q    Max
-57600 -11654  -1440   9800  92880

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)           39490.0    15803.9   2.499 0.013074 *
cars$BrandAiways       4085.5    22350.1   0.183 0.855099
cars$BrandAudi        39770.7    16895.1   2.354 0.019310 *
cars$BrandBMW         49430.0    16824.6   2.938 0.003597 **
cars$BrandBYD          8859.5    17669.3   0.501 0.616505
cars$BrandCitroen      7901.3    16824.6   0.470 0.639008
cars$BrandCUPRA        2700.0    20402.8   0.132 0.894821
cars$BrandDacia      -15940.0    22350.1  -0.713 0.476357
cars$BrandDS           1050.0    27373.2   0.038 0.969431
cars$Brande.Go       -14500.0    27373.2  -0.530 0.596756
cars$BrandElaris      13600.0    27373.2   0.497 0.619720
cars$BrandFiat         5600.0    17312.3   0.323 0.746597
cars$BrandFisker      20125.0    19355.8   1.040 0.299415
cars$BrandFord        24885.0    19355.8   1.286 0.199692
cars$BrandGenesis     28710.0    19355.8   1.483 0.139198
cars$BrandHonda        4255.0    22350.1   0.190 0.849159
cars$BrandHyundai      8358.0    17312.3   0.483 0.629656
cars$BrandJaguar      52910.0    27373.2   1.933 0.054320 .
cars$BrandJeep        -2490.0    27373.2  -0.091 0.927590
cars$BrandKia         21837.8    17471.9   1.250 0.212454
```

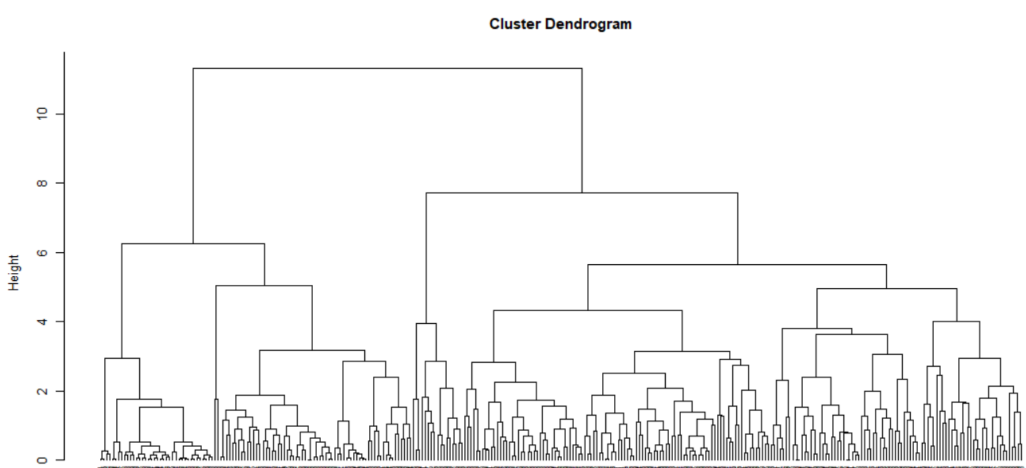
```
cars$BrandLexus       28510.0    27373.2   1.042 0.298586
cars$BrandLotus       84000.0    22350.1   3.758 0.000211 ***
cars$BrandLucid      127110.0    18699.5   6.798 7.11e-11 ***
cars$BrandMaxus       29500.0    27373.2   1.078 0.282157
cars$BrandMazda       -3500.0    27373.2  -0.128 0.898356
cars$BrandMercedes    47883.6    16225.4   2.951 0.003452 **
cars$BrandMG           1700.0    17312.3   0.098 0.921852
cars$BrandMini        -3190.0    22350.1  -0.143 0.886614
cars$BrandNIO         35470.0    17312.3   2.049 0.041471 *
cars$BrandNissan      11145.7    17920.0   0.622 0.534499
cars$BrandOpel        11897.3    16824.6   0.707 0.480108
cars$BrandORA          4833.3    20402.8   0.237 0.812922
cars$BrandPeugeot      9817.7    16824.6   0.584 0.560035
cars$BrandPolestar    29360.0    18248.8   1.609 0.108844
cars$BrandPorsche     96641.1    16658.8   5.801 1.89e-08 ***
cars$BrandRenault       164.3    17920.0   0.009 0.992692
cars$BrandSkoda       13817.1    17920.0   0.771 0.441371
cars$BrandSmart        5500.0    18248.8   0.301 0.763356
cars$BrandSsangYong    1000.0    27373.2   0.037 0.970886
cars$BrandSubaru      18000.0    27373.2   0.658 0.511385
cars$BrandTesla       34823.8    17312.3   2.012 0.045292 *
cars$BrandToyota      12634.2    18248.8   0.692 0.489342
cars$BrandVinFast     27733.3    20402.8   1.359 0.175218
cars$BrandVolkswagen  10863.6    17312.3   0.628 0.530872
cars$BrandVolvo       21134.5    17180.7   1.230 0.219746
cars$BrandZeekr       18580.0    18699.5   0.994 0.321325
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Residual standard error: 22350 on 263 degrees of freedom
  (51 observations deleted due to missingness)
Multiple R-squared:  0.6418,    Adjusted R-squared:  0.5805
F-statistic: 10.47 on 45 and 263 DF,  p-value: < 2.2e-16
```

Based on the multiple R squared values, brand explains the variation in our dependent variable approximately 64%.

Cluster Analysis:

We created a dendrogram to see the number of logical clusters that can be created from the given data and then based on the dendrogram, ran a k-means cluster analysis to see how the different variables feel within the various clusters versus prices.



Based on the dendrogram, we went with 4 predominant clusters and similar relationships were seen in clusters with efficiency and top_speed as our initial regressions analysis, reinforcing the correlation and even a causation associated between efficiency and top-speed to pricing.

```
Summary statistics: Mean
Group variable: group (Cluster ID)
```

group	battery	effici~y	fast_c~e	pricede	range	top_sp~d	acc~0100
1	97.03333	201.2778	872.7778	143548.5	486.9444	240.5556	3.783333
2	64.10987	194.2146	477.5107	51508.75	337.5107	167.103	8.176824
3	98.13333	194.6667	913.3333	202233.3	512.5	260	2.966667
4	92.852	200.04	745.4	98657.24	465.9	217.48	4.852
Total	71.38632	195.5863	552.8339	67529.88	370.6026	181.43	7.275896

Conclusion

In our study we conducted a linear regression model to determine which specifications affect the price of EVs. We then employed bar graphs and scatter plot matrices to explore and describe the data set. Through creating a regression model, we discovered that fast charging time and acceleration feed were statistically significant at the 0.05 level at predicting price. We also found that the battery was statistically significant at the 0.01 level at predicting price. Lastly, top speed and efficiency were statistically significant at the 0.001 level. The signs of the coefficients for the variables were positive, as we expected. Efficiency and Top speed were found to be the most statistically significant in impacting the price of EV. Brand was found to be a mediator. The brands in order of statistical significance with regard to their influence on EV price were Porsche, Lucid, and Lotus at the 0.001 level, BMW and Mercedes at the 0.01 level, and Audi, NIO, and Tesla at the 0.05 level.

Works Cited

lhan, Fatih. "Electric Vehicle Specifications and Prices." *Electric Vehicle Specifications and Prices*, Kaggle, 27 Nov. 2023, www.kaggle.com/datasets/fatihilhan/electric-vehicle-specifications-and-prices/data.

Sloat, Amanda. "Germany's New Centrists? The Evolution, Political Prospects, and Foreign Policy of Germany's Green Party." *Brookings*, The Brookings Institution, Oct. 2020, www.brookings.edu/articles/germanys-new-centrists-the-evolution-political-prospects-and-foreign-policy-of-germanys-green-party/.

"Demand for Electric Cars Is Booming, with Sales Expected to Leap 35% This Year after a Record-Breaking 2022 - News." *International Energy Agency*, IEA, 26 Apr. 2023, www.iea.org/news/demand-for-electric-cars-is-booming-with-sales-expected-to-leap-35-this-year-after-a-record-breaking-2022.

"U.S. Energy Information Administration - EIA - Independent Statistics and Analysis." *International - U.S. Energy Information Administration* , EIA, Dec. 2020, www.eia.gov/international/analysis/country/DEU.

"What Are Electric Cars?" *Union of Concerned Scientists*, UCS, 25 Feb. 2015, www.ucsusa.org/resources/what-are-electric-cars.