Final Project - CS1030

TJ Arbuckle

5/4/2023

Khan Academy analysis:

To start, section 1 is the Data Tools portion of Khan Academy's Data Analysis 101 course covers various tools and techniques used in data analysis. The course introduces students to spreadsheet software, such as Microsoft Excel and Google Sheets, which are commonly used for data analysis. To me this was personally old news. Having worked as a quality inspector for 3 years the concept of using spreadsheets and data to analyze large portions of data was not a new concept but for me it served as a very nice review.

The Next section covers the basics of Big Data, which refers to the processing and analysis of large and complex datasets that are too big to be managed by traditional data processing tools. I found this section very interesting as the applications of big data can have massive implications especially regarding big problems in our world today. Issues like insurance rates, social media, advertising, they all can be predicted and taken advantage of using big data.

The final section of the Data Analysis 101 course is biased in machine learning. Bias in machine learning is an interesting concept to me and the section explores it from a few key angles but the one I found the most interesting was the bias in machine learning. I personally hadn't considered how a machine learning algorithm could be biased because the people who train them are at some level biased in some way. By that understanding it's interesting to think that most of not all machine learning algorithms are biased and it made me start to consider the implications of this to a well-known machine learning being utilized such as Chat GPT.

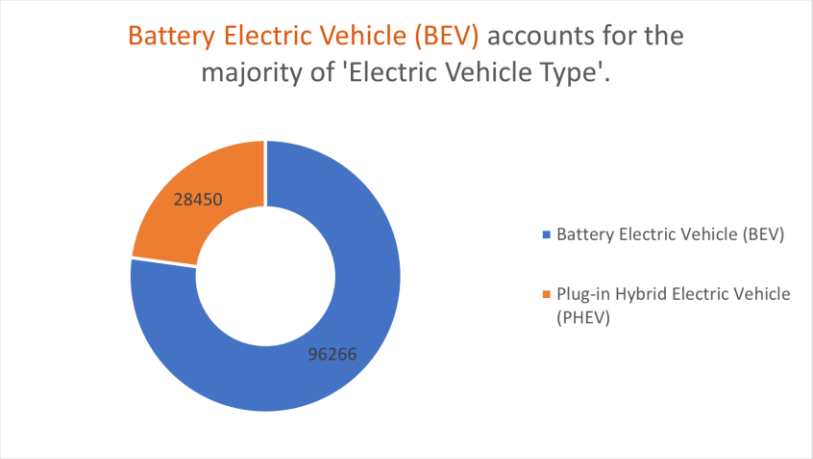
The final portion is the Unit Test. The unit test portion of the Data Analysis 101 course provides an opportunity for you to assess your understanding of the material and ensure that they you understand the concepts covered in each section. I got a 7/9 on the test so I'd say my understanding was good with some room for improvement.

For my Data analysis I wanted to explore a topic regarding the use of electric vehicles. While exploring data.gov I found a data set surrounding the total registered electric/plug in hybrid models. This immediately made me wonder if more people owned and operated hybrid vehicles or fully electric vehicles. I've personally always had an affinity for hybrid vehicles as they are the best for both worlds in my opinion, however, I can understand people not wanting to own them for reliability issues. So, I wanted to explore this. I immediately downloaded the data set and went to work in excel.

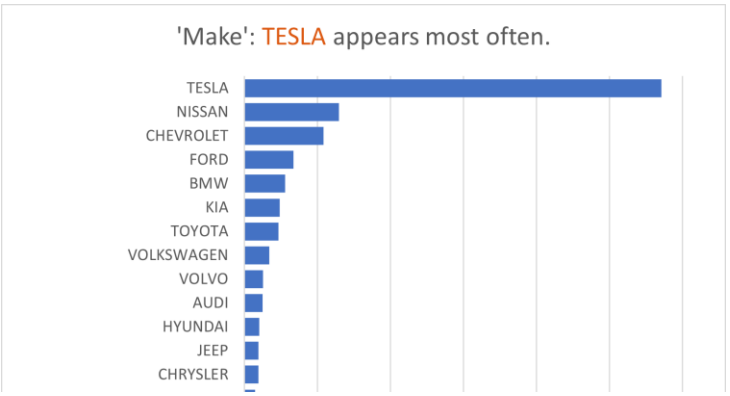
The first this I discovered was that my laptop was severely underpowered for processing this kind of data. I think it might be a limitation of excel on my laptop and its ability to properly utilize the hardware. This was probably due to the excel file having 124,716 rows of data indicating the first important piece of

data that we can gather from this data set. In the state of Washington in the year 2023 there are a total of 124,716 registered hybrid/electric vehicles. After consulting the federal highway administrations website, we know that there are a total of 2,578,732 registered vehicles in the state of Washington as of 2023 meaning when we do the math, we learn that electric/hybrid vehicles in the state of Washington make up 4.8%. This honestly shocked me. I knew that these cars would be slow to come on but some of the vehicles on this list are 2014 model years meaning that these cars have existed for almost 10 years and even after all that they don't even account for 5% of those registered cars. After crunching numbers and producing graphs this is the data set, I generated.

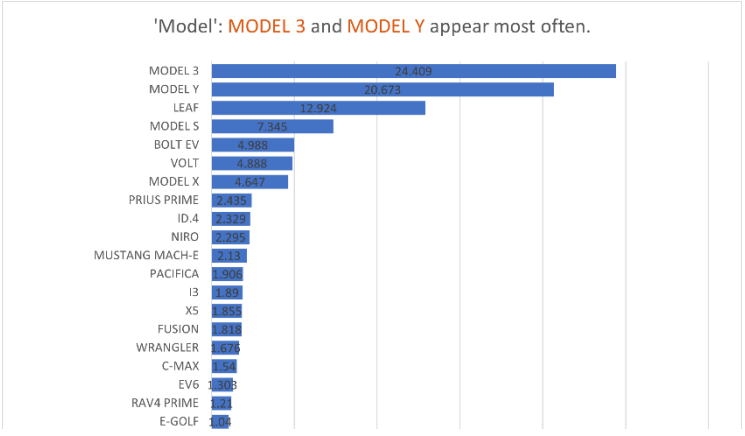
To start we look at the difference between my original hypostasis regarding the number of hybrids versus electric cars. As you can see by the chart, we can see that battery electric cars make up for the majority of the electrified vehicles on the road which I find fascinating. I would have assumed given the amount of time people spend on the road these days that hybrids would be more feasible but perhaps the distance of commute benefits the battery electric cars more.



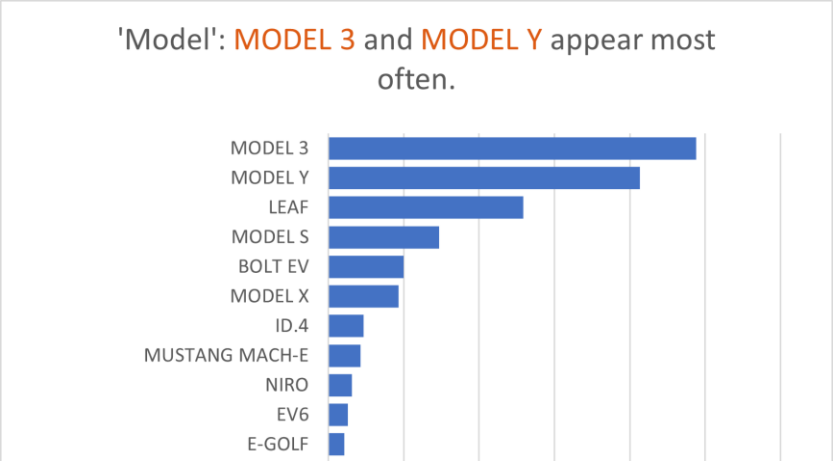
Next, I decided to explore the makes and models of the cars displayed in the overall data. To no one's surprise the Tesla is sitting on top with the Model 3 owning the majority of the vehicles out there. The next closest being Nissan however shocked me. I found it very interesting to see the progression of all the different models and makes of cars. The charts in this document are cropped to fit however the full versions and the Excel document are in the GitHub repository.



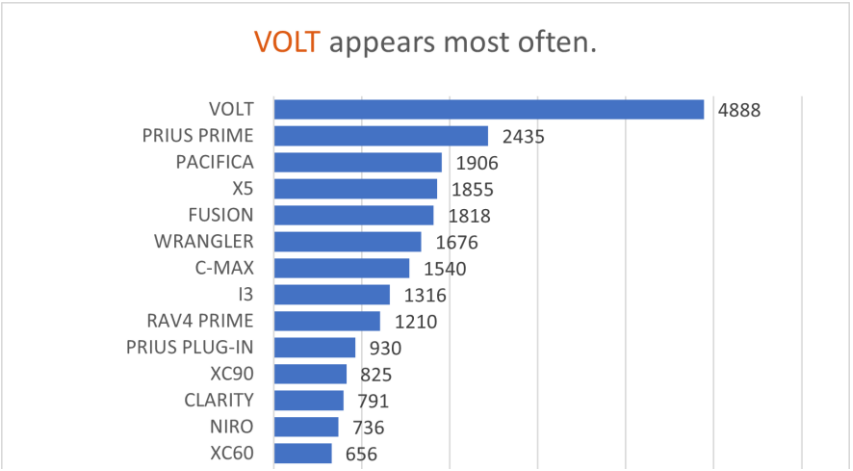
To go into some more interesting details, I split the data between the battery electric and hybrid vehicles to try and extract some more interesting conclusions.



Starting with the battery electric we once again see Tesla on top with their Model 3, Model Y, Model S and Model X only being broken up by the Nissan Leaf and the Chevrolet Bolt EV. This is not a surprise to me in the slightest although it was very interesting to see the GM model coming so high on the list given the experience, I've had with them thus far. This, however, is not where I found interesting data.



The most interesting data I found today was from the hybrid side of things. If we single out the hybrid vehicles and just explore the makes and models registered the results would be very interesting. The quintessential hybrid vehicle is the Toyota Prius. Given how many of them are on the road I figured they would be sitting at the top of the hybrid chart, but I was shocked when I finished putting together the data. The most popular model in the state of Washington is the Chevrolet Volt by a considerable margin. I never thought it would be the more popular model. It is worth noting that this could be a factor only because of the state this is based on. I imagine if this data set included all states (especially California) the data might favor Toyota heavier.



In conclusion, this data analysis on electric and hybrid vehicles in Washington in 2023 revealed that electric vehicles make up the majority of electrified vehicles on the road, with Tesla being the dominant brand. Surprisingly, the Chevrolet Volt emerged as the most popular hybrid model, surpassing the Toyota Prius. Although electric and hybrid vehicles still represent a small fraction (4.8%) of the total registered vehicles in the state, the study underscores the increasing adoption of electrified vehicles and highlights the potential influence of geographical factors on vehicle preferences. Further research could expand the scope of the analysis and investigate factors affecting consumer preferences and adoption rates.

Data Link: https://catalog.data.gov/dataset/electric-vehicle-population-data

All data and charts are listed on the Excel file in the GitHub repository. **This includes additional notes, charts and data not seen in this document.**