Tahfiq Williams, Aaron Almeida

Professor Fonkam

DS 220

7th December, 2023

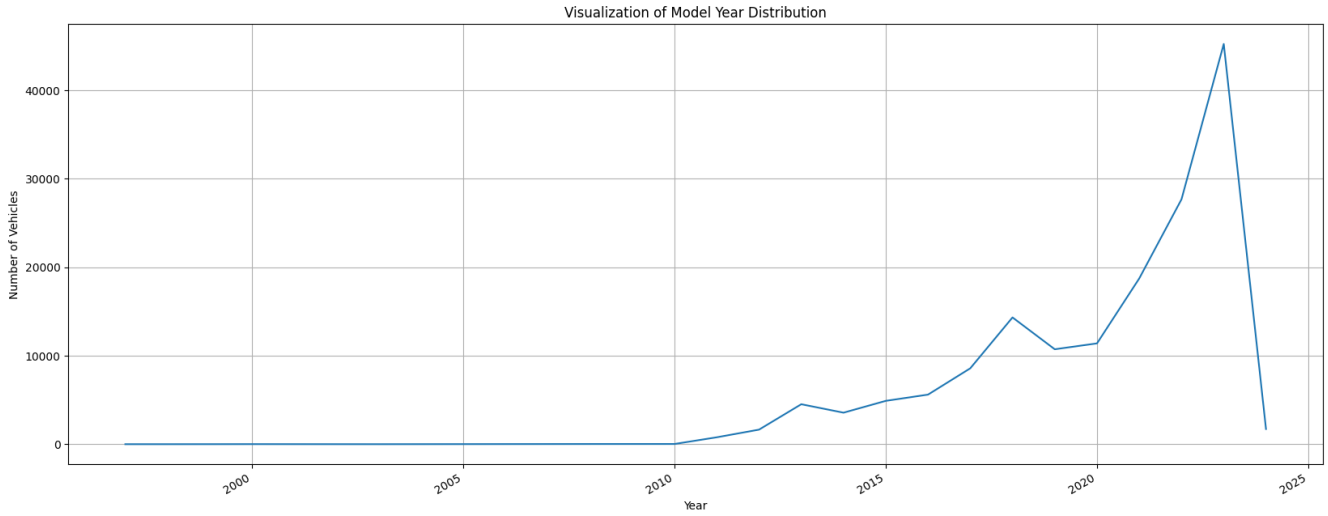
Final Project - EVs in Washington State

Introduction

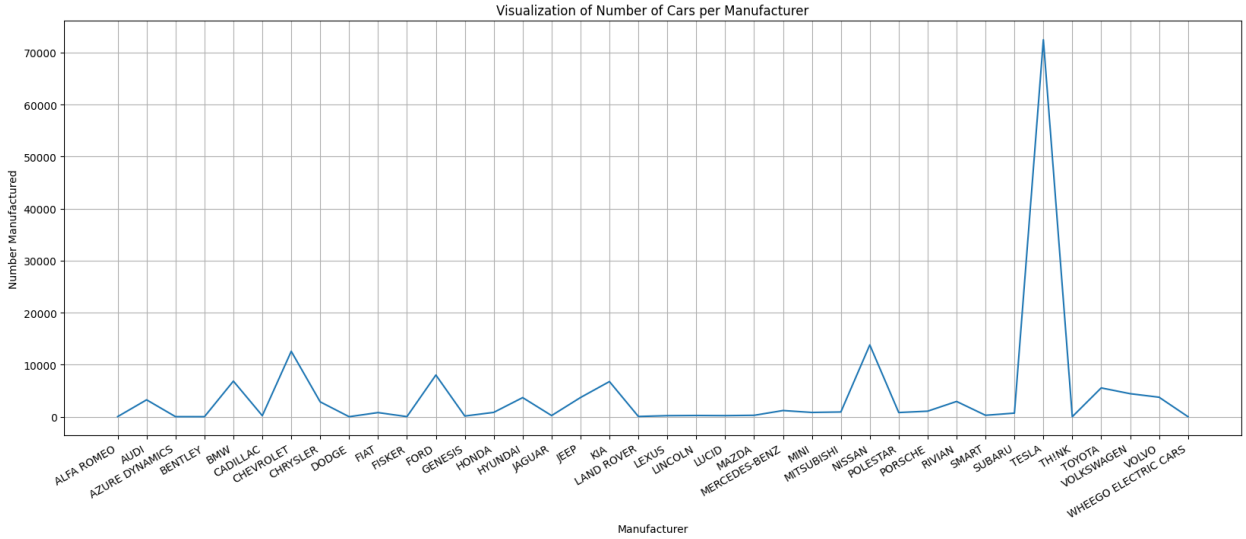
The dataset that we chose is the Electric Vehicle population in the state of Washington. It represents the various types of EVs ranging from ones that have a limited hybrid capacity to pure electric vehicles that provide hundreds of miles of range on a single charge. The data is quite substantial and informative as it goes into the depths with various parameters that allow us to better understand the region and its situations. This can also lead to us gathering more information on the infrastructure available for electric vehicles in certain counties of the state and any underlying causes that may be relevant to the same.

Visualization of Data

The data set has almost 160,000 entries, representing 38 car manufacturers that produce some form of electric-powered/supported vehicles. The data goes back to vehicles produced in the year 1997 up until 2023 with brand new electric vehicles up to late 2023. This allows us to visualize a strong upward trend as the year moves on while peaking with over 28% of the total electric vehicle population being vehicles of the 2023 Model Year. The 2024 Model Year cars are already showing a strong trend and we can expect to see it even surpass the record-breaking 2023 performance as seen in the graph below.



The Electric Vehicle sphere seems to have a clear leader in the Market, with Tesla being responsible for over 45% of the market share. This can indicate various things and can prompt us to question how and why they have managed to maintain such a strong hand with so many other competitors. One thing for sure is that they have heavily invested in their ecosystem which makes owning a Tesla extremely convenient.



Query List:

1.) Which cars present in the dataset have an electrical range of 0-100? Which county has the most of these cars?

2.) Out of all of the Tesla's present within Washington, how many are specifically Model 3's?

3.) What city has mostly newer (cars between 2020 and present) model cars? (or we can just do how many newer model cars are in the dataset)

4.) What car make is the most common in the entire dataset?

5.) What postal code has the newest cars?

6.) What Tesla model is most popular in Washington?

Query Results

1) The county that has cars with an electric range between 0 to 100 miles is King County. We can take away from this observation that either the electric charging infrastructure is not as developed as surrounding counties or a large majority of EV owners do not have long mileage as a requirement with just short distance commutes.

This query filters through the dataset and includes only the rows that have their 'Electric Range' values set as less than or equal to 100. This query provides a chart-based visual representation of the information it retrieves.

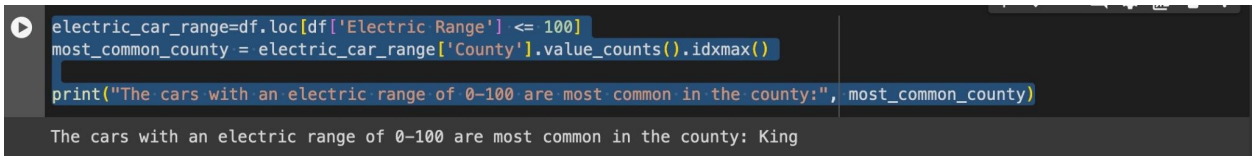
```
electric_car_range_df = df[df['Electric Range'] <= 100]

print("Cars with an electric range of 0-100:")
print(electric_car_range_df)
```

```
Cars with an electric range of 0-100:
        VIN (1-10)    County           City State  Postal Code  ...  \
0       2C4RC1N71H    Kitsap      Bremerton    WA      98311.0  ...
1       2C4RC1N7XL   Stevens       Colville    WA      99114.0  ...
2       KNDC3DLCXN    Yakima         Yakima    WA      98908.0  ...
6       WMZYU7C51K    Yakima         Yakima    WA      98902.0  ...
7       7SAYGDEF8N    Yakima         Yakima    WA      98901.0  ...
...            ...       ...            ...   ...          ...  ...
159461  7SAYGDEF4P  Thurston        Olympia    WA      98516.0  ...
159462  KM8JBDA2XP  Skamania      Underwood    WA      98651.0  ...
159463  1G1FZ6S02M    Skagit            Bow    WA      98232.0  ...
159464  YV4H60CX2P      King      Sammamish    WA      98029.0  ...
159466  7SAYGDEF6N    Island  Camano Island    WA      98282.0  ...

        Legislative District DOL Vehicle ID                Vehicle Location  \
0                       23.0      349437882  POINT (-122.6466274 47.6341188)
1                        7.0      154690532     POINT (-117.90431 48.547075)
2                       14.0      219969144  POINT (-120.6027202 46.5965625)
6                       14.0      146830148   POINT (-120.524012 46.5973939)
7                       15.0      207786505      POINT (-120.500225 46.6043)
...                      ...            ...                              ...
159461                  22.0      254707820  POINT (-122.7474291 47.0821119)
159462                  14.0      235949514  POINT (-121.5312858 45.7348285)
159463                  40.0      148544168   POINT (-122.440636 48.5613885)
159464                   5.0      240200754  POINT (-121.9993659 47.5484866)
159466                  10.0      208285619  POINT (-122.5310901 48.2192797)

                                    Electric Utility 2020 Census Tract
```

This query filters through the dataset and includes only the rows that have their 'Electric Range' values set as less than or equal to 100. This query provides a chart-based visual representation of the information it retrieves.

This query checks for the values in the 'Electric Range' column of the dataset that are less than or equal to 100. Once the query checks for that and collects that information, it stores it

within the electric_car_range variable. From there it uses the value_counts() method to go through the amount of unique values within the ‘County’ column. Then the idmax() method finds the most common county within the filtered DataFrame. This information is stored in the most_common_county variable.



2) The Tesla Model 3 is by far the most popular car. The count of 28,589 suggests that there is a very significant presence of this particular model, Model 3, within the dataset, which gives significant insight into the popularity of the Model with consumers. This also allows us to begin to conclude why this popularity exists. This dataset is taken in Washington, and with Washington being the state with the 2nd highest percentage of electric vehicles in the United States according to sources like U.S News, this high number of Tesla Model 3’s makes sense. With Tesla being one of the biggest innovators in clean energy and electric vehicles and being the biggest name in the industry, it's no surprise that its models would be the most popular.

This query filters through the dataset and includes only the rows where the value for the ‘Make’ column is ‘TESLA’ and where the value for the ‘Model’ column is ‘MODEL 3’. The ‘&’ operator is used for the logic operator ‘AND’ and allows the query to filter the data based on both conditions. After filtering through the dataset and retrieving the information it has queried for, the code stores this information in the variable tesla_count_model3. The final line prints out the information retrieved from the query, outputting a message that makes it easier to deduce what this query is doing, as well as the count of rows where the 'Make' is 'TESLA' and the 'Model' is 'MODEL 3' in the dataset. The overall result/ output is the number of Tesla Model 3

cars in the dataset. This particular query, while pretty simple, provides a good amount of information and real-world insights.

```
tesla_count_model3=len(df[(df['Make'] == 'TESLA') & (df['Model'] == 'MODEL 3')])
print("The number of Tesla Model 3 cars is:", tesla_count_model3)
```

```
The number of Tesla Model 3 cars is: 28589
```

3) The first query filters through the dataset and includes only the values where the 'Model Year' is greater than or equal to 2020. We set the condition within our question that a 'newer car' is a car that has a model year between 2020 and the present year. So as we query, we should filter through and only collect rows where the 'Model Year' is either 2020 or greater (like 2021, 2022, or 2023). The first query returns a chart-based representation of the information it collects.

```
df.loc[df['Model Year'] >= 2020]
```

	VIN (1-10)	County	City	State	Postal Code	Model Year	Make	Model	Electric Vehicle Type	Clean Alternative Fuel Vehicle (CAFV) Eligibility	Electric Range	Base MSRP	Legislative District	DOL Vehicle ID	Vehicle Location	El
1	2C4RC1N7XL	Stevens	Colville	WA	99114.0	2020	CHRYSLER	PACIFICA	Plug-in Hybrid Electric Vehicle (PHEV)	Clean Alternative Fuel Vehicle Eligible	32	0	7.0	154690532	POINT (-117.90431 48.547075)	
2	KNDC3DLCXN	Yakima	Yakima	WA	98908.0	2022	KIA	EV6	Battery Electric Vehicle (BEV)	Eligibility unknown as battery range has not b...	0	0	14.0	219969144	POINT (-120.6027202 46.5965625)	
7	7SAYGDEF8N	Yakima	Yakima	WA	98901.0	2022	TESLA	MODEL Y	Battery Electric Vehicle (BEV)	Eligibility unknown as battery range has not b...	0	0	15.0	207786505	POINT (-120.500225 46.6043)	
9	5YJ3E1EB1M	Kitsap	Poulsbo	WA	98370.0	2021	TESLA	MODEL 3	Battery Electric Vehicle (BEV)	Eligibility unknown as battery range has not b...	0	0	23.0	211699309	POINT (-122.64177 47.737525)	
15	YV4BR0CL3N	Snohomish	Bothell	WA	98021.0	2022	VOLVO	XC90	Plug-in Hybrid Electric Vehicle	Not eligible due to low battery range	18	0	1.0	194401833	POINT (-122.179458 47.802589)	

The second query does a very similar thing as the first, as it filters through the dataset and collects where the 'Model Year' is greater than or equal to 2020. However, it is different from the first query because it does not provide a chart-based representation of the data it collects, instead returning just one numerical value. This is because of the len() method. This method calculates and returns the number of rows in the dataset, effectively returning the count of rows

where ‘Model Year’ is greater than or equal to 2020. Also, unlike the first query, the second query stores this information within the new_car_count variable. The last line prints out the information retrieved from the query, outputting a message that makes it easier to deduce what this query is doing, as well as the number of new cars (cars having a model year between 2020 and the present year) within the whole dataset.

This query could provide us insight into how many new electric vehicles are being purchased by consumers within Washington. If compared with how many electric vehicles there are in Washington that aren't considered new, we could gain insight into just how often people are buying electric vehicles.

4) This query goes through the dataset and accesses the ‘Make’ column. The query uses the value_counts() method on the ‘Make’ column once it's accessed to count the occurrences of unique values within the dataset. The idmax() method is then applied to find the car make with the maximum count- in terms of our original question, this finds the car make that appears the most and therefore is the most common in the dataset. Once the information for this query is retrieved, it is stored within the most_common_make variable. The final line of this query works to print out the retrieved information, as well as a message to give insight into what the entire query is returning.

The query ultimately returns the car make that is the most popular in the entire dataset, which is Tesla. This query provides very similar insight and information that the query for

question 2 provided. This query, however, is just more general than the query for our second question. This query just tells us that Tesla as a brand is the most common car maker of the electric cars in Washington, while the query for question two tells us what specific model of Tesla is the most popular. Thus, this query provides us with very similar insight, as again it shows that Tesla is the most popular brand for consumers of electric cars in Washington. This is most likely due to the fact that Washington is one of the leading states for electric cars and Tesla is one of the biggest, most trusted, and most-known names in that particular industry. It makes sense that in one of the biggest states for electric cars, the leading brand for electric cars would have the most popular make owned by consumers.

```
[32] most_common_make = df['Make'].value_counts().idxmax()

     print("The most common car make is:", most_common_make)

     The most common car make is: TESLA
```

5) This query first filters through the dataset and includes only the rows and values where ‘Model Year’ is greater than or equal to 2020. As previously mentioned, we set our standard for a new car to have a model year between 2020 and the present year (2023), so we first have to filter through the data to collect the values within that range. The query then stores this retrieved information in the newer_cars variable to be used further. The query then counts the amount of unique occurrences of postal codes in the ‘Postal Code’ column using the value_counts() method, while the idmax() method then returns the postal code with the maximum count. The query then works to find the postal code with the most occurrences within the filtered, retrieved information within the newer_cars variable. This information is then stored within the in_postal_code variable, which is what we’re looking to ultimately return. The final line of the query returns the result of the query, which is the postal code that has the most number of new cars. This provides an answer to our question, telling us that postal code 98052 is the postal code

with the most new cars owned. This query provides various insights into this specific postal code.

One insight it could provide is into the affluence and economic standing of this specific postal code. A high concentration of newer cars- especially electric cars, which can be more expensive than regular cars- in a neighborhood suggests that its residents are more affluent and have a higher economic standing than others. This could also give insight into the values and demographics of the residents within this postal code. The residents here may place a higher value on their environmental impact, which may explain why the concentration of electric cars is higher here. This could also reveal certain demographics, like age in particular, as electric cars are more common amongst younger consumers.

```
newer_cars=df.loc[df['Model Year'] >= 2020]
in_postal_code = newer_cars['Postal Code'].value_counts().idxmax()
print("The postal code with the newest cars:", in_postal_code)
```

```
The postal code with the newest cars: 98052.0
```

6) The first part of this query filters through the dataset to include only rows where the ‘Make’ value is ‘TESLA’. As our original question is asking about what model from Tesla specifically is the most popular, it's important for us to first filter through the data to only include information about Tesla and its models. This information is then stored in the variable tesla_wa. The query then uses the value_counts() method to count the unique occurrences within the ‘Model’ column. The idmax() method is then used to find the Tesla Model with the maximum count in the dataset, in the context of our question, this gives the most popular Tesla model in the dataset. This information is then stored in the variable most_popular_tesla_model. The query then finally prints out the information we retrieved, returning the most popular model present in the dataset. This answers our original question, telling us that the most popular Tesla model in Washington is the Model Y. The insights that this query and its result provide us with are mostly

regarding consumer preferences. It tells us that there is a high concentration of Tesla Model Y's in the dataset, meaning that there is a strong preference for this model over all others among electric car consumers.

Understanding this preference could help dealerships in advertising, marketing, and which products they should supply for their customers. This could also provide various other insights that aren't just limited to consumer preferences. For example, this could provide insight to Tesla about their most purchased model so that they may adapt their other models to be more similar to it. On the other hand, it could also be useful to Tesla competitors to have this information as well, as they could adapt their cars to be more like Tesla's Model Y in order to better compete.

```
[36] tesla_wa= df[(df['Make'] == 'TESLA')]
     most_popular_tesla_model = tesla_wa['Model'].value_counts().idxmax()

     print("The most popular Tesla model in Washington:", most_popular_tesla_model)

     The most popular Tesla model in Washington: MODEL Y
```

Conclusion

There are many resources available online that can provide a better understanding of Electric Vehicles and the major impact that will impose on the future of the automotive industry. This article provides the reader with a clear idea of what they can expect to see in the coming decades within the industry and the significance it will have on the consumer. Overall, electric vehicles are the future and we will definitely see ICE (Internal Combustion Engine) vehicles phasing out slowly over time.