Part 1. Data Analysis and Insight

1. Data Overview

The dataset titled “Electric Vehicles” contains an overview of the distribution of electric vehicles in several cities in the state of Washington, USA. Although the exact source of the data is not specified in the file, it was likely obtained from a vehicle registration database with 202 records and 6 fields.

- Vehicle Identification Number: Unique identifier of the vehicle
- City: The city where the vehicle is registered
- Postal code: The postal code associated with the vehicle's registration address
- Model year (2011-2025): The year the vehicle model was manufactured
- Make: Manufacturer’s name
- Model: The specific model of the vehicle

This data can be used to analyze popular electric vehicle models or the market share of manufacturers in different regions.

	A	B	C	D	E	F
1	Vehicle Identification Number	City	Postal Code	Model Year	Make	Model
2	JTMAB3FV3P	Seabeck	98380	2023	TOYOTA	RAV4 PRIME
3	1N4AZ1CP6J	Bremerton	98312	2018	NISSAN	LEAF
4	5YJ3E1EA4L	Seattle	98101	2020	TESLA	MODEL 3
5	1N4AZ0CP8E	Seattle	98125	2014	NISSAN	LEAF
6	5YJYGDEE5L	Lynnwood	98036	2020	TESLA	MODEL Y
7	1G1FX6S00H	Yelm	98597	2017	CHEVROLET	BOLT EV
8	5YJYGDEE5L	Lynnwood	98036	2020	TESLA	MODEL Y
9	KM8S6DA23N	Poulsbo	98370	2022	HYUNDAI	SANTA FE
10	7FCTGAAA1P	Arlington	98223	2023	RIVIAN	R1T
11	5YJYGDEE9L	Kent	98031	2020	TESLA	MODEL Y
12	1N4AZ0CP2F	Kirkland	98034	2015	NISSAN	LEAF
13	1G1FX6S0XH	Bremerton	98337	2017	CHEVROLET	BOLT EV

2. Data Cleaning

Using the filter function to detect 11 missing data. Firstly, in column A (Vehicle Identification Number) have 4 missing data is in rows 70, 74, 79, and 93. Since this value is the unique identifier of the vehicle, it cannot be left blank and cannot be inferred or filled in manually, so I decided to delete these rows, as they are missing important information. Since 4 rows have been deleted now the total number of rows is 198

	A	B	C	D	E	F
1	Vehicle Identification Number	City	Postal Cod	Model Yea	Make	Model
70		Oak Harbor	98277	2019	TESLA	MODEL 3
74		Olympia	98512	2015	NISSAN	LEAF
79		Seattle	98112	2023	BMW	X5
93		Seattle	98117	2015	NISSAN	LEAF

Secondly, in column B (City), there are 2 missing values in rows 24 and 31.

	A	B	C	D	E	F
1	Vehicle Identification Number	City	Postal Cod	Model Yea	Make	Model
24	1G1RD6E46C		98506	2012	CHEVROLET	VOLT
31	5YJ3E1EB7J		98012	2018	TESLA	MODEL 3

These can be imputed by comparing with the Postal Code. The postal codes 98012 and 98506 correspond to the cities of Bothell and Olympia, respectively.

	A	B	C	D	E	F
1	Vehicle Identification Number	City	Postal Cod	Model Yea	Make	Model
24	1G1RD6E46C	Olympia	98506	2012	CHEVROLET	VOLT
31	5YJ3E1EB7J	Bothell	98012	2018	TESLA	MODEL 3

Thirdly, in column E (Make) there are two missing values in rows 16 and 97.

	A	B	C	D	E	F
1	Vehicle Identification Number	City	Postal Cod	Model Yea	Make	Model
16	1G1RB6S50J	Centralia	98531	2018		VOLT
97	5YJ3E1EB5J	Poulsbo	98370	2018		MODEL 3

They can be attributed by comparing with the Model. In which, the VOLT model is mainly produced by CHEVROLET and the MODEL 3 model is mainly produced by TESLA.

	A	B	C	D	E	F
1	Vehicle Identification Number	City	Postal Cod	Model Yea	Make	Model
16	1G1RB6S50J	Centralia	98531	2018	CHEVROLET	VOLT
97	5YJ3E1EB5J	Poulsbo	98370	2018	TESLA	MODEL 3

Next, in column F (Model) there is 1 missing value in row 5.

	A	B	C	D	E	F
1	Vehicle Identification Number	City	Postal Cod	Model Yea	Make	Model
59	3C3CFFGE0F	Seattle	98133	2015	FIAT	

Compared with the Model Year, in 2015 FIAT manufacturers only produced 1 electric car model, which is the 500.

	A	B	C	D	E	F
1	Vehicle Identification Number	City	Postal Cod	Model Yea	Make	Model
59	3C3CFFGE0F	Seattle	98133	2015	FIAT	500
184	3C3CFFGE9D	Spokane	99202	2013	FIAT	500

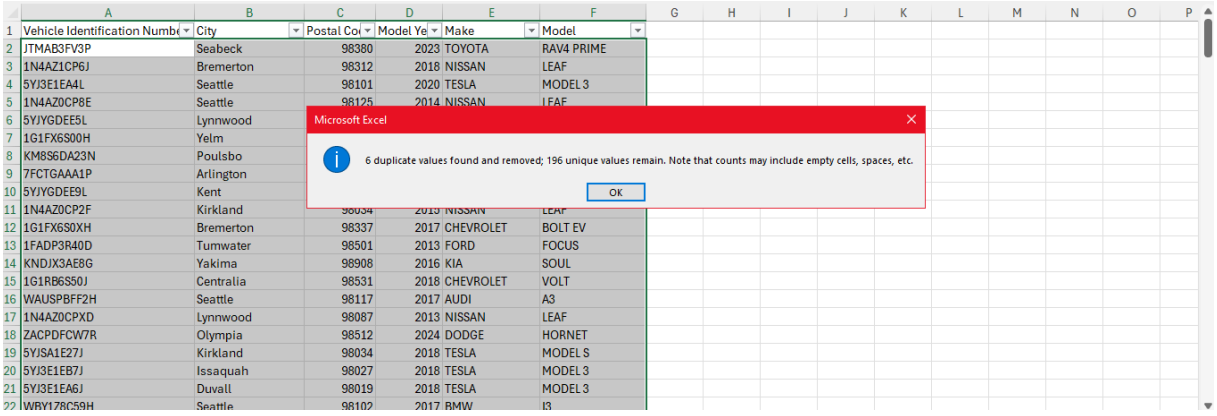
Lastly, in column D (Model Year) there are 2 missing values in rows 52 and 63

	A	B	C	D	E	F
1	Vehicle Identification Number	City	Postal Cod	Model Yea	Make	Model
52	KNDCC3LG1L	Seattle	98144		KIA	NIRO
63	1N4BZ0CP3H	Bremerton	98310		NISSAN	LEAF

I decided to fill in N/A instead of missing data because filling in Mode can artificially increase the number of cars registered in the year of high appearance, skewing the chart.

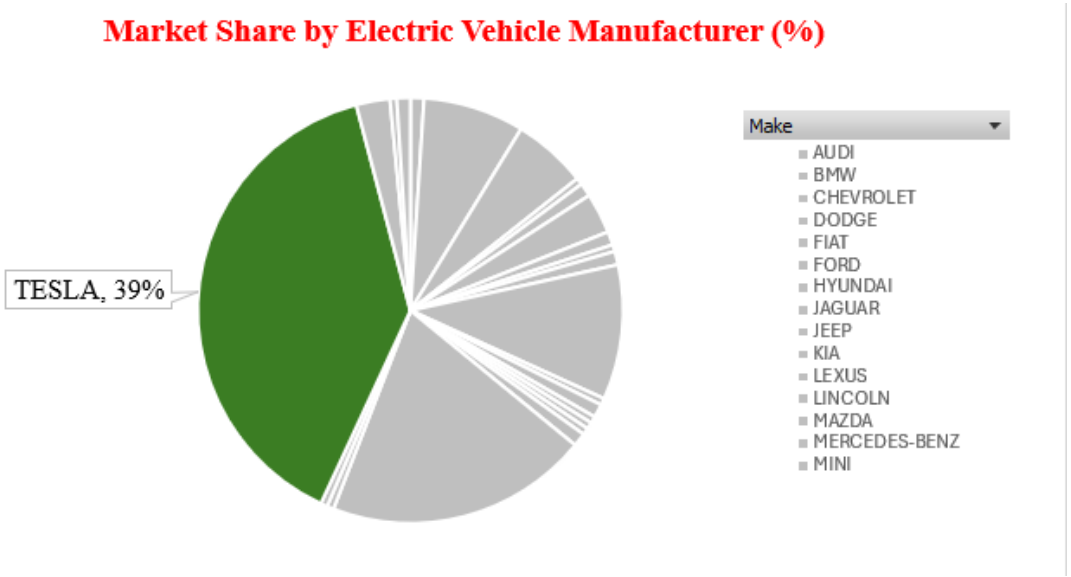
I chose to remove duplicate values in the Excel file because they could cause inaccuracies in the analysis. According to the notification, 6 duplicate rows were found and removed, leaving 196 unique entries.

Keeping duplicates would have affected the summary statistics, visualizations, and overall interpretation of the data. Each vehicle is only represented once, which improves the validity and clarity of the results.



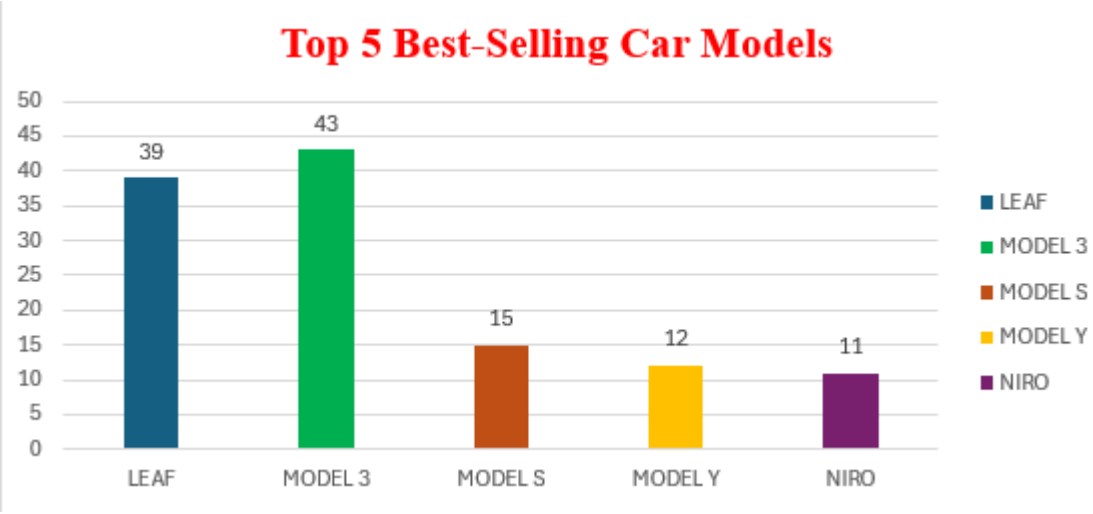
3. Descriptive Statistics

Insight 1: Tesla dominates electric vehicle market share



Tesla dominates the electric vehicle market in this dataset with 39% of all registrations, reaching 76 vehicles, almost double the number of its nearest rival, Nissan (39 vehicles). This leadership comes from the brand's strong presence on popular models, such as the Model 3 and the Model Y. Reflecting consumer preferences for advanced technology, its range and charging infrastructure in the region are improving.

Insight 2: Tesla dominates electric vehicle market share



The Tesla Model 3 was the best-selling electric vehicle in its class, with 43 registrations, far ahead of its sister model, the Model Y, which had 12. The Model 3's dominance is due to its winning combination of affordability, long range, and widespread charging support, making it the most popular vehicle among consumers in the region.