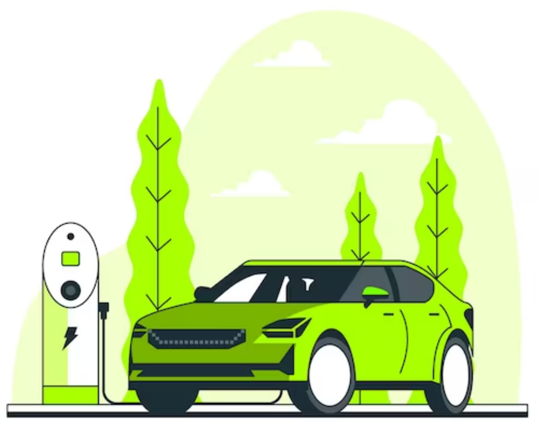
Northeastern University
Mechanical and Industrial Engineering Department
IE 6200: Engineering Probability and Statistics
Prof. Rehab Ali
Fall 2023

Shift of Fuel Choices towards EV: A Comparative Analysis of Conventional, Alternative, and Electric Vehicles in the U.S.

Final Report



Group #4

Aditya Ayyappan
Shashank Bharadwaj Ramachandra
Melanie Edmund
Aaditya Gupta
Esha Joshi
Rakshith Patil

Table Of Contents

Introduction

In the ever-evolving landscape of the automotive industry, the emergence of electric vehicles (EVs) has shaken things up, challenging conventional ideas about how we typically perceive and fuel vehicles. Initially met with skepticism, the tide has gradually turned as the first commercially available electric vehicles hit the market, coupled with the relentless surge in gas prices and growing environmental consciousness. As this transformative wave continues to reshape the automotive terrain, it becomes imperative for consumers to navigate this dynamic landscape with foresight. This project undertakes an analysis, delving into the nuanced aspects of conventional and alternative fuel vehicles versus electric vehicles in the United States.

This project aims to conduct a comprehensive comparative analysis, with a primary focus on scrutinizing the sales trends of conventional and alternative fuel type vehicles versus electric vehicles (EVs) in the context of the total car population in the USA. One of the key objectives of the project is to analyze whether EV trends are rising enough to overcome conventional fuel usages. Leveraging previously obtained data, and in-depth market analysis. This project seeks to provide valuable insights into the automotive industry's transition. However, the project consciously avoids delving into intricate technical details, legislative analyses, or exhaustive financial assessments, focusing instead on delivering a holistic understanding of the industry's transformative journey.

Data Collection

The data collection process for this project involved a thorough exploration of various online sources and datasets related to the automotive industry in the United States. Extensive research was conducted on different websites, and diverse data sets were considered before selecting the most suitable ones to meet the specific requirements of the project. The chosen datasets primarily originated from reputable government websites, ensuring reliability and accuracy in the information presented. The primary sources referenced in the project include [1] Vehicle Sales Data [2] Fuel Economy Data [3] Data acquired from government websites. These datasets were compiled and organized to form a comprehensive repository of information for the analysis.

The selected datasets were not created through conducting surveys or independent data collection but were rather curated from existing sources, namely the Alternative Fuels Data Centre and the Fuel Economy US website, both affiliated with the US Department of Energy. The variables, or columns, within the dataset encompass crucial information related to the automotive landscape in the United States. These variables include state-specific data, such as the total number of electric vehicles (EVs), plug-in hybrid electric vehicles (PHEVs), hybrid electric vehicles (HEVs), biodiesel users, ethanol/flex (E85) users, compressed natural gas (CNG) users, gasoline vehicles, diesel vehicles, and various combinations representing different fuel types. Assumptions made during the analysis include considering registration data across states as an accurate representation of vehicle sales.

The major variables for modelling focus on yearly sales trends of electric, alternative, and conventional vehicles from 2016 to 2022, along with the associated costs and government incentives, providing a comprehensive foundation for the project's analytical endeavors.

2016 Light-Duty Vehicle Registration Counts by State and Fuel Type

State	Electric (EV)	Plug-In Hybrid Electric (PHEV)	Hybrid Electric (HEV)	Biodiesel	Ethanol/ Flex (E85)	Compressed Natural Gas (CNG)	Propane	Hydrogen	Methanol	Gasoline	Diesel	Unknown Fuel	Electric	Alternative	Conventional
Alabama	500	900	29,100	0	4,28,300	20,100	0	0	0	37,77,300	1,26,500	53,900	30,500	4,48,400	39,03,800
Alaska	200	200	5,000	0	55,700	4,900	0	0	0	5,25,900	44,800	19,400	5,400	60,600	5,70,700
Arizona	4,700	4,400	89,600	0	4,27,300	17,500	0	0	100	48,05,000	1,79,500	1,12,800	98,700	4,44,800	49,84,500
Arkansas	200	500	19,100	0	3,20,500	12,600	0	0	0	20,97,800	96,800	22,200	19,800	3,33,100	21,94,600
California	1,41,500	1,16,700	9,66,700	0	13,22,600	80,600	0	1,300	400	2,72,41,000	7,10,400	1,15,500	12,24,900	14,03,200	2,79,51,400
Colorado	5,300	3,800	74,000	0	3,55,500	15,900	0	0	0	41,80,500	2,26,800	89,900	83,100	3,71,400	44,07,300
Connecticut	2,000	2,800	43,200	0	1,60,800	3,200	0	0	0	26,90,900	59,700	30,200	48,000	1,64,000	27,50,600
Delaware	300	700	11,200	0	68,300	1,400	0	0	0	7,49,800	17,000	7,200	12,200	69,700	7,66,800
District of Columbia	600	500	11,800	0	17,600	300	0	0	0	2,77,800	3,200	1,300	12,900	17,900	2,81,000
Florida	11,600	10,100	2,07,100	0	11,68,900	18,000	0	0	0	1,39,29,200	3,53,300	1,33,600	2,28,800	11,86,900	1,42,82,500
Georgia	18,000	4,000	86,400	0	7,10,700	26,900	0	0	0	73,77,700	2,06,700	1,14,800	1,08,400	7,37,600	75,84,400
Hawaii	4,200	1,200	22,600	0	43,600	1,500	0	0	0	9,53,300	17,800	7,400	28,000	45,100	9,71,100
Idaho	400	600	16,300	0	1,07,100	12,800	0	0	0	13,54,300	1,23,000	56,200	17,300	1,19,900	14,77,300
Illinois	5,800	6,000	1,46,500	0	9,68,600	18,500	0	0	0	87,48,600	2,00,900	94,500	1,58,300	9,87,100	89,49,500
Indiana	1,300	2,400	54,500	0	5,69,300	19,900	0	0	0	48,55,200	1,56,800	82,500	58,200	5,89,200	50,12,000
Iowa	400	1,200	26,900	0	3,67,400	13,400	0	0	0	24,18,600	1,07,600	80,600	28,500	3,80,800	25,26,200
Kansas	600	1,000	26,200	0	2,50,600	14,200	0	0	0	20,95,000	90,000	42,400	27,800	2,64,800	21,85,000
Kentucky	500	900	30,900	0	3,36,000	15,600	0	0	0	32,85,300	1,15,200	47,500	32,300	3,51,600	34,00,500
Louisiana	400	500	17,700	0	4,69,700	11,900	0	0	0	30,47,300	1,52,400	36,600	18,600	4,81,600	31,99,700
Maine	300	1,000	17,900	0	1,04,700	1,700	0	0	0	10,11,700	31,000	12,900	19,200	1,06,400	10,42,700
Maryland	3,200	5,000	85,900	0	3,27,800	6,100	0	0	0	40,39,100	1,05,200	44,500	94,100	3,33,900	41,44,300
Massachusetts	3,600	5,200	98,300	0	3,15,600	3,400	0	0	0	46,23,900	79,000	38,800	1,07,100	3,19,000	47,02,900
Michigan	1,600	11,600	64,900	0	11,86,000	19,800	0	0	0	69,02,900	1,90,800	81,100	78,100	12,05,800	70,93,700

Fig 1: Sample of data set

Data Visualization

Data visualization is a powerful tool that transforms complex datasets into comprehensible and meaningful insights through graphical representation. By utilizing charts, graphs, and other visual elements, data visualization facilitates a more intuitive understanding of trends, and relationships within information. The data here is represented in the form of scatter plots, line graphs and histograms.

Trend Lines

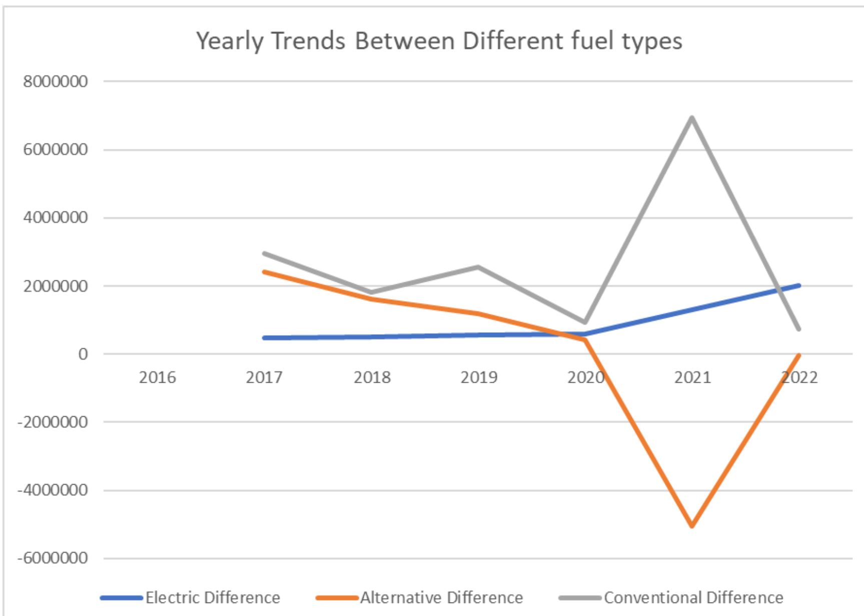


Fig 2. Graph showing yearly trends between different fuel types.

Fuels are categorized as conventional, alternative, and electric. We can see from the blue line that electric fuel utilization has gradually increased from 2017 to 2020 and a steeper rise from 2020 to 2022, the overall increase from 2016 to 2022 is 350.14%. The orange line represents alternative fuels, the use of these fuels has sharply declined in 2020 but further increases from 2021 to 2022. The grey line shows the trend for conventional fuels. There is an overall decrease from 2017 to 2022 with a steep increase from 2020 to 2021.

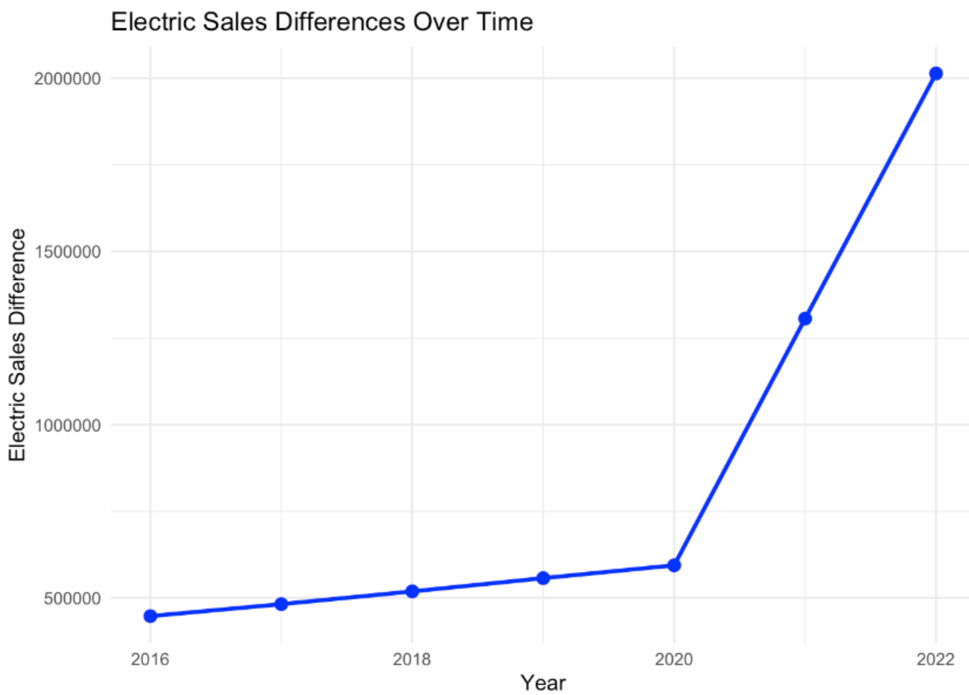


Fig 3. Graph showing yearly trends for electric vehicle sales from 2016 to 2022

From Fig 3 we can understand that the overall trend for electric vehicle sales is increasing over time. The sales are gradually increasing from 2016 to 2022 and then there is a sharp incline from 2020 to 2022.

Share of Sales

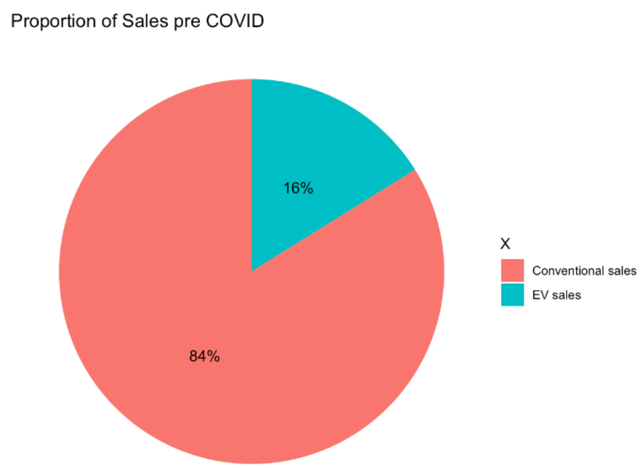


Fig 4. Pre-Covid sales

From Fig 4 we can see that there are 16% EV sales before covid and to 84% of vehicle sales for conventional fuels.

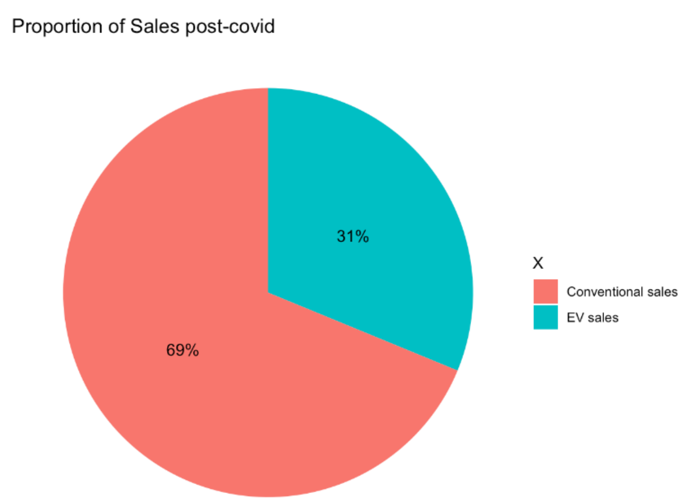


Fig 5. Post-Covid Sales

From Fig 5 we can see that 31% EV sales post covid and 69% of vehicle sales for conventional fuels. We can see there is an increase in EV sales from 16% pre-covid to 31% post-covid.

Scatter Plot

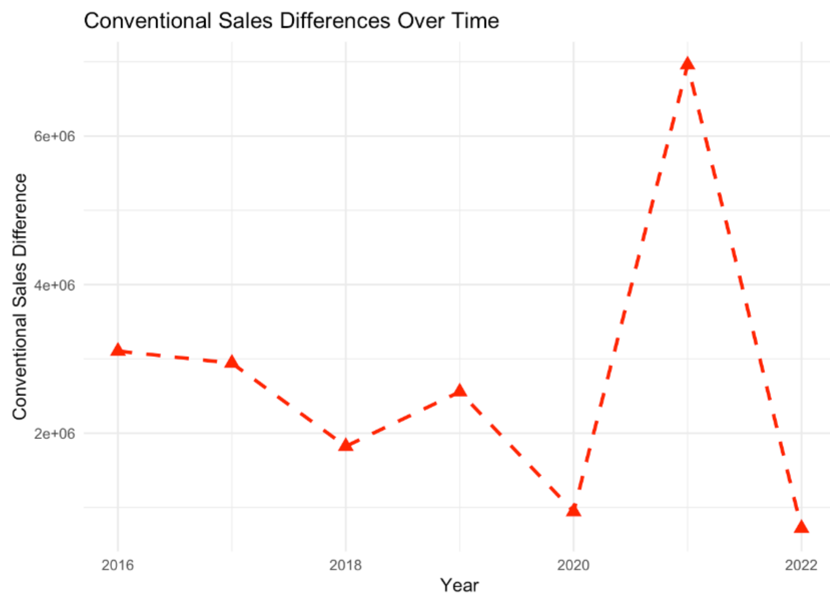


Fig 6. Line graph Scatter Plot for Conventional Vehicle Sales from 2016 to 2022

From Fig 6 we can understand that the overall trend for conventional vehicle sales is decreasing over time. The sales are gradually decreasing from 2016 to 2020 and then there is a sharp inclining and then declining from 2020 to 2022.

Statistical Analysis

Hypothesis Testing

In statistical hypothesis testing, the null hypothesis (H_0) posits no effect or difference, representing the default assumption, while the alternative hypothesis (H_1 or H_a) asserts a real effect or difference. Researchers use sample data to determine whether there is enough evidence to reject the null hypothesis in favor of the alternative hypothesis. The p-value, a measure of the probability of obtaining observed results if the null hypothesis is true, guides this decision. If the p-value is below a predetermined significance level (α), typically 0.05, the null hypothesis is rejected, suggesting support for the alternative hypothesis. If the p-value is higher than α, the null hypothesis is not rejected, indicating insufficient evidence for the alternative hypothesis. This process helps draw meaningful conclusions about population parameters based on observed sample data.

Hypothesis 1: Testing for Significance Difference in EV Sales Mean vs. Conventional Sales Mean (All Years)

Null Hypothesis (H_0): The mean of electric vehicle (EV) sales is equal to the mean of conventional vehicle sales.

Alternative Hypothesis (H_1): The mean of electric vehicle (EV) sales is not equal to the mean of conventional vehicle sales.

Inference: The one-sample t-test resulted in a p-value of 0.0001608, which is less than the 0.05 significance level. Therefore, there is sufficient evidence to reject the null hypothesis. The mean electric vehicle sales difference is significantly different from the population mean of conventional vehicle sales difference.

Hypothesis 2: Testing for Significance Difference in EV Sales Mean vs. Conventional Sales Mean (2016-2019)

Null Hypothesis (H_0): The mean of EV sales between 2016 and 2019 is equal to the mean of conventional sales between 2016 and 2019.

Alternative Hypothesis (H_1): The mean of EV sales difference between 2016 and 2019 is not equal to the mean of conventional sales between 2016 and 2019.

Inference: The Welch two-sample t-test resulted in a p-value of 0.0004527, which is less than the 0.05 significance level. Therefore, there is sufficient evidence to reject the null hypothesis. The mean electric vehicle sales difference is significantly different from the mean conventional vehicle sales difference between 2016 to 2019.

Hypothesis 3: Testing for Significance Difference in EV Sales Mean vs. Conventional Sales Mean (2020-2022)

Null Hypothesis (H_0): The mean of EV sales between 2020 and 2022 is equal to the mean of conventional sales between 2020 and 2022.

Alternative Hypothesis (H_1): The mean of EV sales between 2020 and 2022 is not equal to the mean of conventional sales between 2020 and 2022.

Inference: The Welch two-sample t-test resulted in a p-value of 0.1571, which is greater than the 0.05 significance level. Therefore, there is insufficient evidence to reject the null hypothesis. The mean electric vehicle sales difference is not significantly different from the mean conventional vehicle sales difference between 2020 and 2022.

Overall Inference: Between 2016 and 2019, there is a significant difference between EV sales and conventional vehicle sales. However, between 2020 and 2022, there seems to be insufficient evidence to accurately reject the null hypothesis. This suggests that buyer preference for EVs is climbing with a decline in conventional vehicle preferences, especially after the impact of COVID-19.

ANOVA Testing

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Factor	4	409849333	102462333	9.84	0.002
Error	10	104106667	10410667		
Total	14	513956000			

Fig 7. Pre COVID-ANOVA Analysis

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Factor	4	5960857333	1490214333	1.51	0.271
Error	10	9847606667	984760667		
Total	14	15808464000			

Fig 8. Post COVID ANOVA Analysis

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
3226.56	79.74%	71.64%	54.42%

Means

Factor	N	Mean	StDev	95% CI
Massachuttes	3	10767	929	(6616, 14917)
New york	3	16367	4070	(12216, 20517)
Georgia	3	6933	3101	(2783, 11084)
New jersey	3	13733	4554	(9583, 17884)
Flordia	3	22400	2066	(18249, 26551)

Pooled StDev = 3226.56

Fig. 9 Summary of 5 chosen states

An ANOVA analysis compares the observations of samples with different variables to determine differences in outcomes if any. Like Hypothesis testing, with an ANOVA analysis there is a null hypothesis that states something about the data sets being analyzed and if the calculated p-value is less than the level of significance we must reject the null hypothesis. In figures 11 and 12, the null hypothesis stated that the means in each sample were the same. With a level of significance of 0.05, figure 11 indicates that we must reject the null hypothesis and that there is a statistically significant difference between the means. However, figure 12 indicates that we must accept the null hypothesis and that there is no statistically significant difference between the means of the samples.

Boxplot

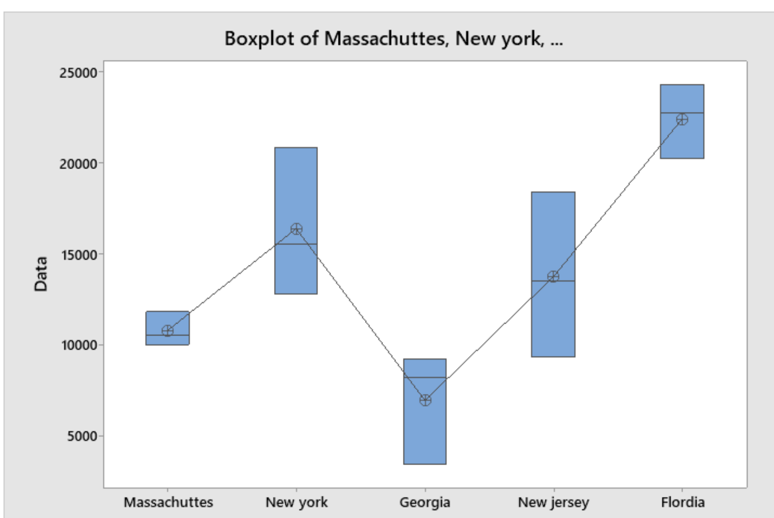


Fig 10. Boxplot of mean EV sales in 5 states prior to COVID

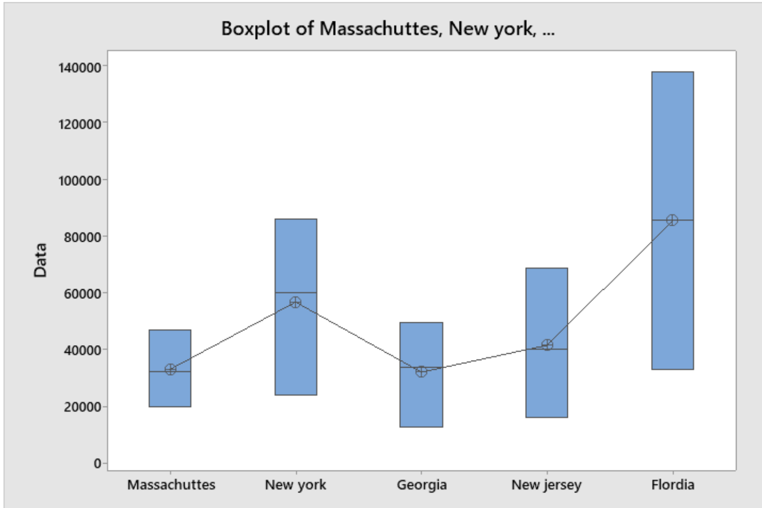


Fig 11. Boxplot of mean EV sales in 5 states after COVID

Figures 10 and 11 above are boxplots of the mean EV sales in 5 states prior to and after pandemic. Prior to the pandemic, the lack of overlap in the boxplot indicates almost no overlap in the EV purchasing habits between states. The means are fairly spread out and while Massachusetts and Florida have very small ranges, the others are quite spread out in comparison, which indicates a wider variety of purchasing numbers. In stark contrast, after covid, the median values all increased, are much closer together and the range of values are also more similarly distributed. This indicates that the states are now purchasing EVs at more similar rates than prior to COVID.

Confidence Interval of Means

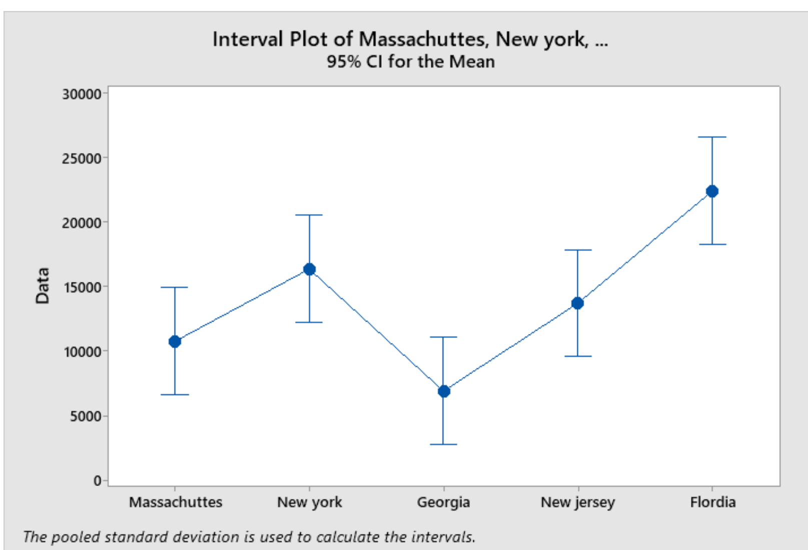


Fig 12. Plot of confidence interval of means for 5 states prior to COVID.

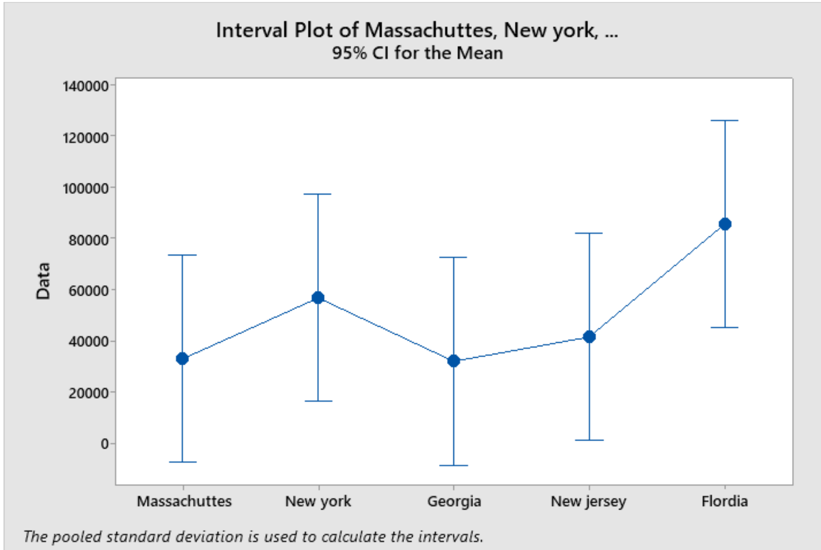


Fig 13. Plot of confidence interval of means after COVID.

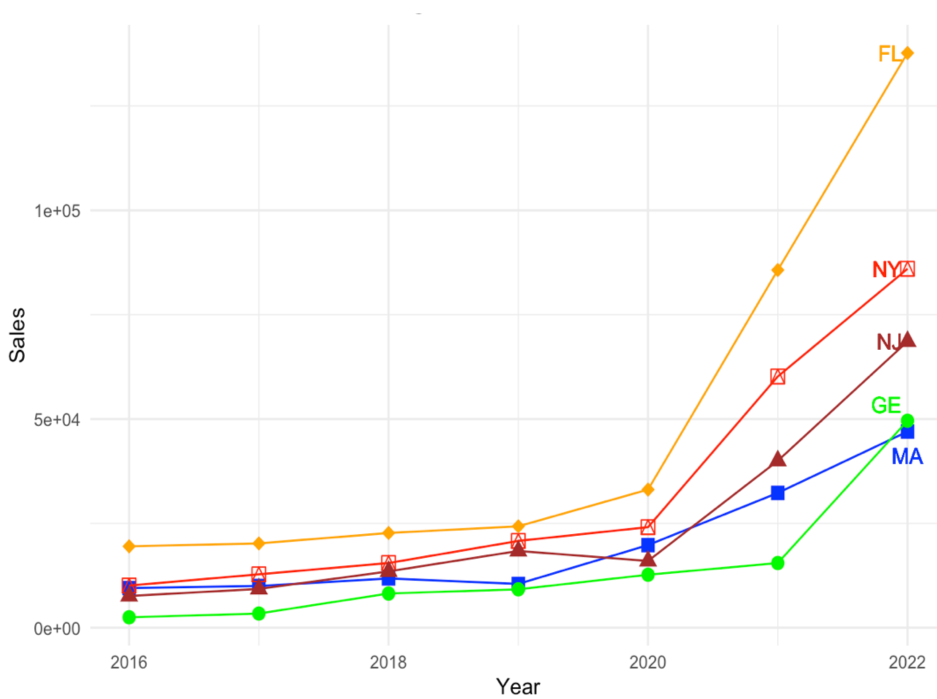


Fig.14 Trend of EV sales of 5 states used in ANOVA Analysis

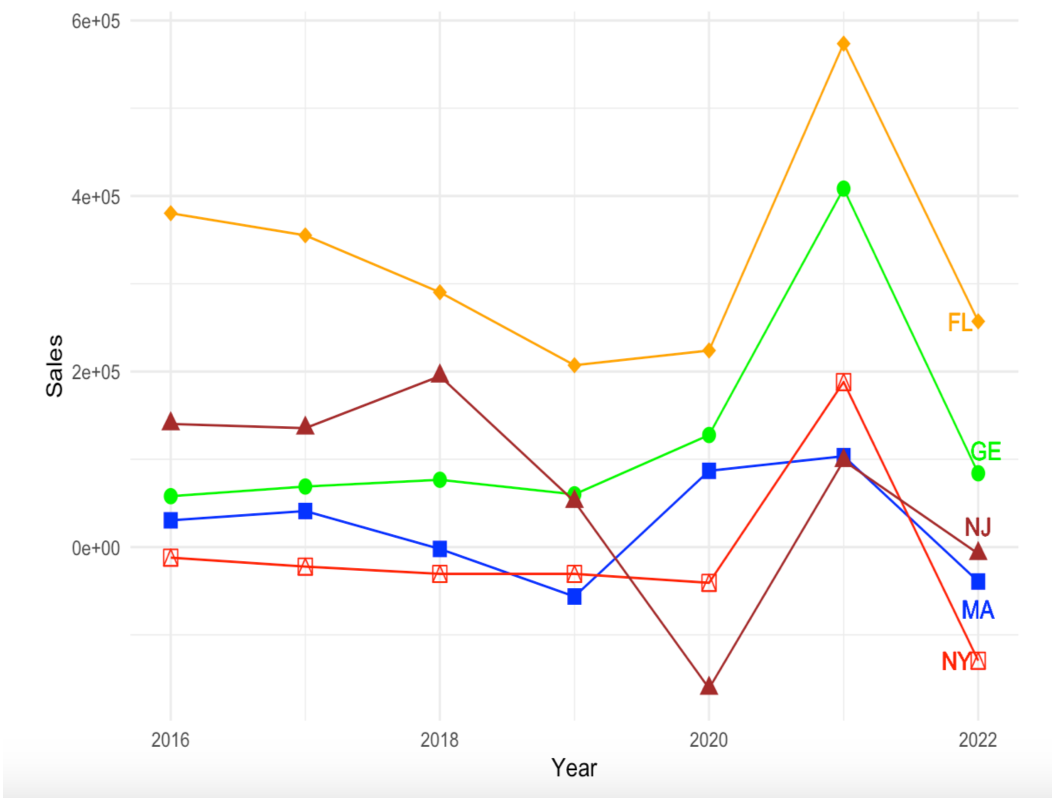


Fig.15 Trend of Non-EV sales of 5 states used in ANOVA Analysis

Results and Conclusion

All testing and analysis completed throughout the course of this project indicated an upwards trend in the purchasing of electric vehicles across the entire united states. The beginning of the increase was seen after the pandemic began in 2020 and the purchase of EVs rather than conventional vehicles has only continues to rise. Outside of another global pandemic or national crisis.

Limitations

Limited Exploration of Infrastructure Challenges:

The project acknowledges a limitation in delving into the intricate details of charging infrastructure challenges for electric vehicles. The success of widespread EV adoption is contingent upon a robust and accessible charging network. However, factors such as the current state of charging infrastructure, potential bottlenecks, and the timeline for expansion are not extensively explored in this study.

Ignoring these elements may limit the depth of understanding regarding the practical aspects that influence consumer decisions and the broader shift toward electric mobility.

External Economic Factor Not Fully Explored:

Economic conditions, including inflation rates, interest rates, and overall market stability, play a crucial role in shaping consumer purchasing behaviors. This project, while recognizing the impact of economic factors, does not comprehensively explore their intricacies. Fluctuations in economic conditions can significantly affect the affordability of vehicles, potentially influencing the transition from conventional to electric vehicles. The study acknowledges the existence of these external economic influences but falls short of providing an exhaustive examination of their specific impacts on the automotive industry's transformative journey.

Proposed Future Work

Consumer Demographics and Preferences: Understanding the intricate interplay between consumer demographics and preferences is important in deciphering the trajectory of electric vehicle (EV) adoption. By delving into the diverse fabric of consumer segments, including age, income levels, urban or rural residence, and environmental awareness, the project can discern the nuanced factors influencing vehicle choices. For instance, analyzing data to discern whether younger, environmentally conscious urban dwellers exhibit a higher propensity for EV adoption compared to older, suburban demographics with different priorities.

Market Dynamics Beyond Passenger Vehicles: Expanding the analysis to encompass a broader spectrum of the automotive market is essential for a comprehensive understanding of the transition to electric vehicles. While the focus often centers on passenger cars, evaluating market dynamics beyond this segment – encompassing electric trucks, buses, and two-wheelers will uncover distinct adoption patterns and challenges. Investigating these diverse market segments ensures a holistic approach to understanding the transformative journey of the automotive industry, offering valuable insights that extend beyond the realm of passenger vehicles.

Expanding Research Beyond USA: Different countries have varied regulatory frameworks, incentives, and policies influencing the automotive industry. Expanding the research to include multiple countries introduces the challenge of navigating diverse regulatory landscapes, making it complex to draw direct comparisons and generalize findings. Each region may have unique factors influencing the adoption of electric vehicles, such as government subsidies, emission standards, and import/export regulations.

Detailed Explanation of Charging Infrastructure: Future work should include a dedicated exploration of the challenges and opportunities associated with charging infrastructure for electric vehicles. This involves assessing the current state of charging networks, identifying potential obstacles to expansion, and forecasting the development timeline.

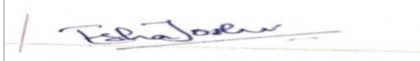
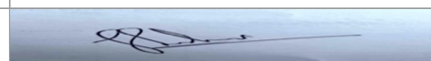
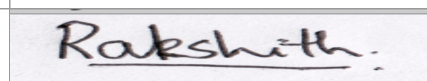
Bibliography

[1] "Alternative Fuels Data Center: Vehicle Registration Counts by State," *afdc.energy.gov*.
https://afdc.energy.gov/vehicle-registration?year=2016

[2] "Download Fuel Economy Data," fueleconomy.gov.
https://www.fueleconomy.gov/feg/download.shtml

TEAM MEMBER CONTRIBUTON REPORT

NAME	PERCENTAGE CONTRIBUTION	CONTRIBUTIONS
Esha Joshi	16%	Project goals, Data compilation, Data preparation, Formulation of hypothesis, Report compilation, Limitations and proposed work
Aaditya Gupta	17%	Problem statement, Project goals, Trends in R, Hypothesis testing, R markdown
Aditya Ayyappan	17%	Data Sorting in R, Hypothesis Testing in R, Programming in R, Hypothesis 1 formulation
Rakshit Patil	17%	Problem statement, Data preparation, Data Sorting in R, Hypothesis Testing, ANOVA Analysis, Hypothesis 2 formulation
Shashank Bharadwaj Ramachandra	16%	Data preparation, Identification of variables, Hypothesis Testing, Formatting, PPT compilation
Melanie Edmund	17%	Problem statement, Hypothesis question formulation, Results and Conclusion, Hypothesis Testing, Formatting,

Team Member	**Signatures**
Esha Joshi (Point of contact)	
Aaditya Gupta	
Aditya Ayyappan	
Shashank	
Rakshit Patil	
Melanie Edmund