A Statistical Analysis of Electric Vehicles Population Data

IN-CLASS PRESENTATION-2

AA-5221- APPLIED ANALYTICS AND METHOD- 1

TEAM TECH ALLIANCE

Introduction

01

The shift towards sustainability has driven humans to adopt Electric vehicles.

02

We analysed electric vehicle registrations in Washington State to understand EV adoption using various factors.

03

BEVs (Battery Electric Vehicles) run solely on electricity, while PHEVs (Plug-in Hybrid Electric Vehicles) combine a battery with a gasoline engine.

04

The insights derived can help electric vehicle producers' customers make informed decisions on EV incentives.

Background

Our team has chosen this topic considering the 3 major points

Analysing electric vehicles is fascinating

As of 2024, there are nearly 12.5 million alternate fuel vehicles in use in the United States

Technological transformations making EVs more efficient

Our research focused majorly on

Comparison of EV types, Models, Make

Trends in Electric Range

CAFV Eligibility

Dataset Details

Columns Present in Dataset	
• VIN (1-10) • County • City • State • Postal Code • Model year • Make • Model • Electric Vehicle Type	• Clean Alternative Fuel Vehicle(CAFV) Eligibility • Base MSRP • Legislative District • DOL Vehicle ID • Vehicle Location • Electric Utility • 2020 Census Tract

The dataset was obtained from the official Washington State open data portal website.

We have used the “Electric Vehicles Population Data.” for our analysis.

The dataset consists of information on Battery Electric vehicles(BEVs) and Plug-in Hybrid Electric Vehicles(PHEVs) that are currently registered through the Washington State Department of Licensing(DOL)

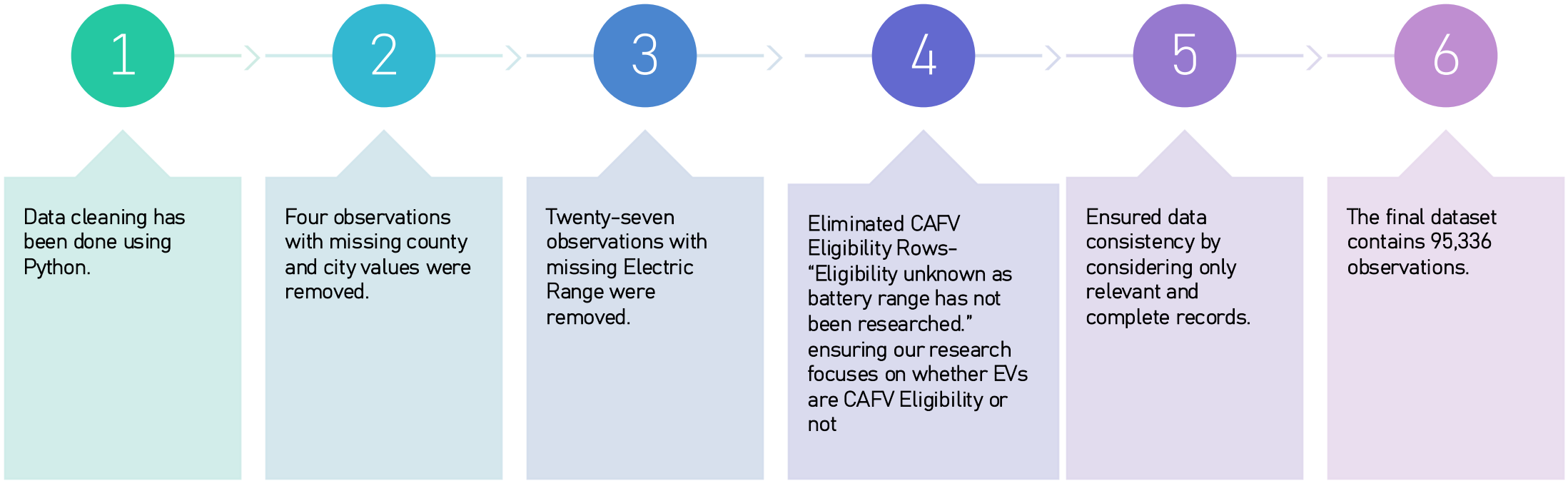
The dataset consists of 232k Rows with 17 columns

Variables in the Dataset

Variable Name	Variable Type	Description/Values
Model	Nominal	Specific Model of the Electric Vehicle
City	Nominal	Name of the city where the vehicle is registered.
State	Nominal	The state of vehicle registration (mostly Washington).
Model Year	Continuous	Year the vehicle model was manufactured (e.g., 2015, 2020).
Make	Nominal	Brand of the vehicle (e.g., Tesla, Nissan, Chevrolet).
Electric vehicle type	Nominal	Type of EV: Battery Electric Vehicle (BEV) or Plug-in Hybrid Electric Vehicle (PHEV).
CAFV Eligibility	Nominal	Indicates if the vehicle is eligible for Clean Alternative Fuel Vehicle incentives or not
Electric Range	Continuous	Maximum distance (in miles) the EV can travel on a full charge.

- We analysed Electric Vehicle type, Electric Range, Model year, Make, Model, and CAFV Eligibility Columns.

Dataset Cleaning



Research Questions Explored

01

Is there a significant difference in the electric range between Battery Electric Vehicles (BEVs) and Plug-in Hybrid Electric Vehicles (PHEVs)?

02

Is there a relationship between the model year and the electric range of electric vehicles?

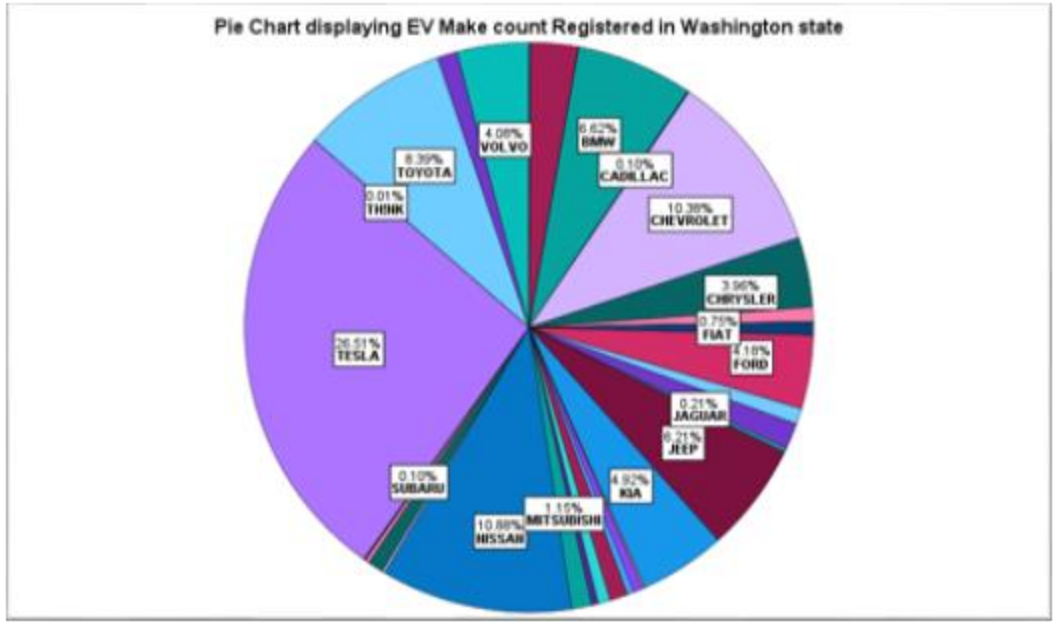
03

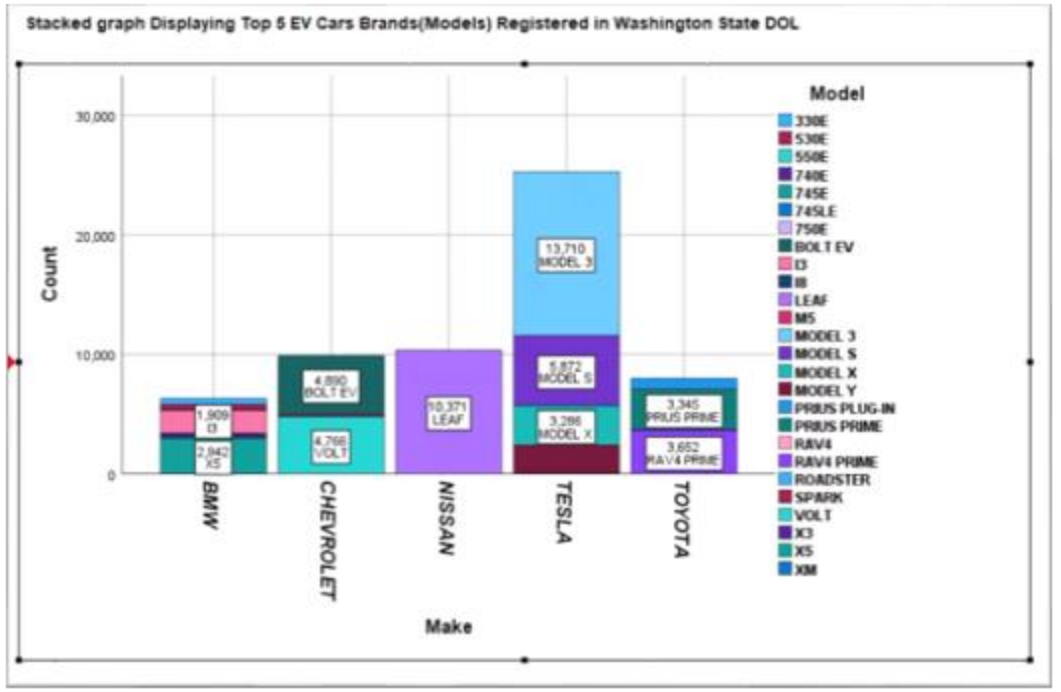
Are Battery Electric Vehicles more likely to be CAFV eligible than Plug-in Hybrid Electric Vehicles?

04

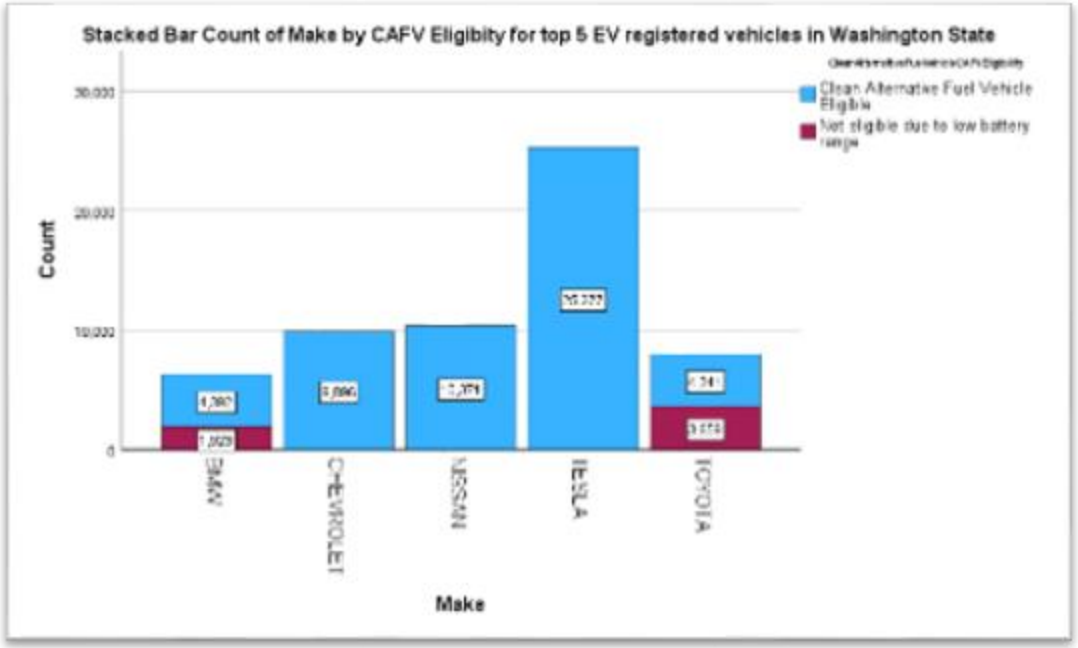
Is there a significant relationship between the model year and electric range for Tesla electric vehicles registered in Washington State?

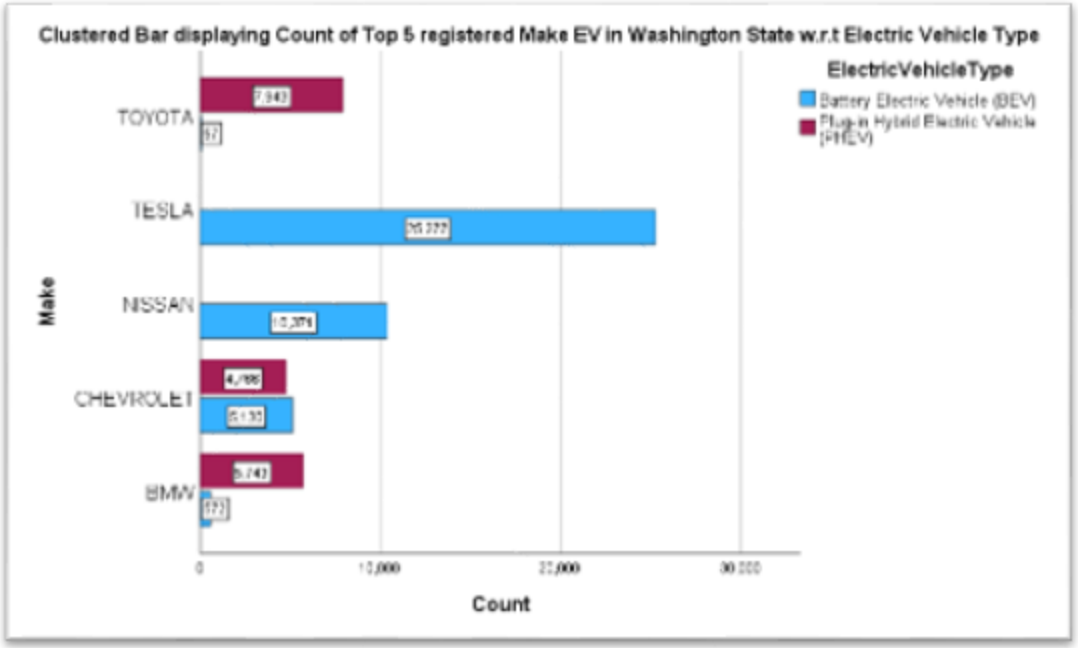
Descriptive Statistics



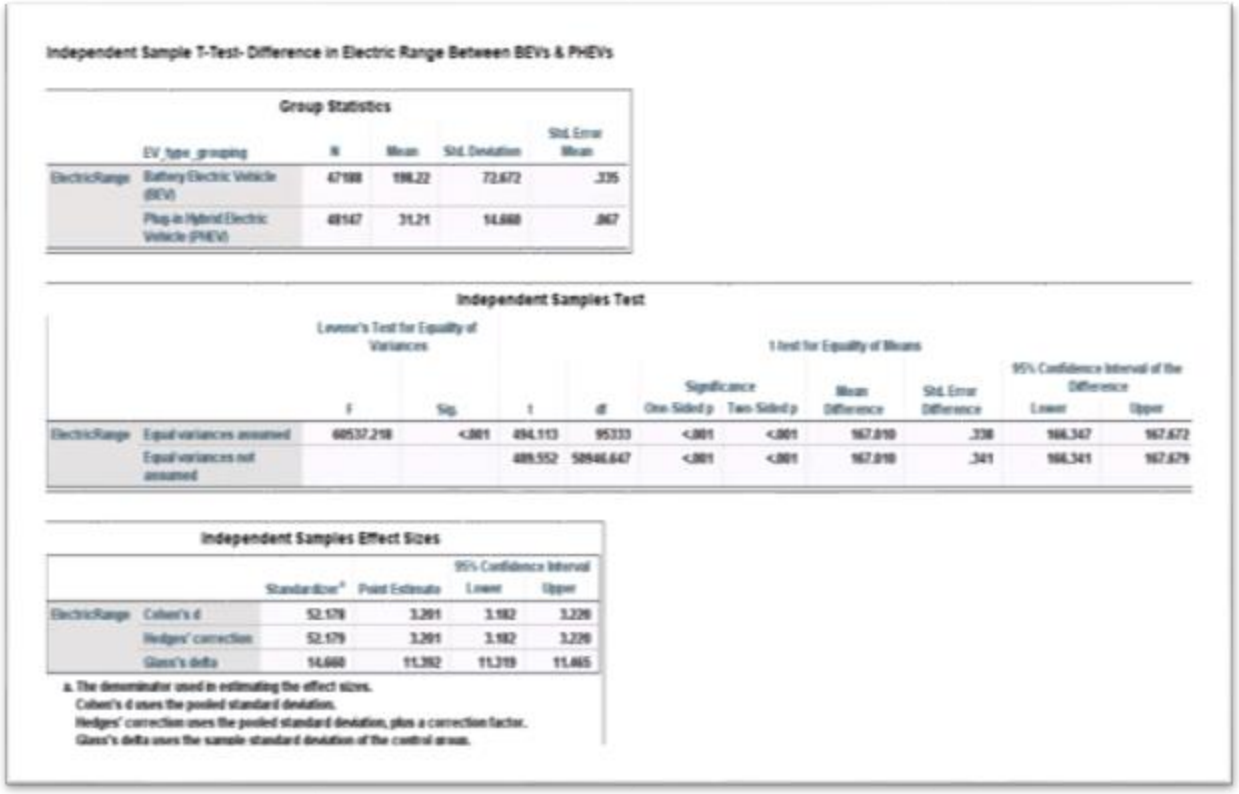


Descriptive Statistics





Difference in Electric Range Between BEVs & PHEVs



Independent Sample T-Test- Difference in Electric Range Between BEVs & PHEVs

Group Statistics

	EV_type_grouping	N	Mean	Std. Deviation	Std. Error Mean
ElectricRange	Battery Electric Vehicle (BEV)	67188	198.22	72.672	.335
	Plug-in Hybrid Electric Vehicle (PHEV)	48147	31.21	14.668	.067

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance				95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference	Lower	Upper
ElectricRange	Equal variances assumed	60537.218	<.001	494.113	95333	<.001	<.001	167.010	.338	166.347	167.672
	Equal variances not assumed			489.552	58946.647	<.001	<.001	167.010	.341	166.341	167.679

Independent Samples Effect Sizes

		Standardizer[a]	Point Estimate	95% Confidence Interval Lower	Upper
ElectricRange	Cohen's d	52.179	3.201	3.182	3.220
	Hedges' correction	52.179	3.201	3.182	3.220
	Glass's delta	14.668	11.392	11.319	11.465

a. The denominator used in estimating the effect sizes.
Cohen's d uses the pooled standard deviation.
Hedges' correction uses the pooled standard deviation, plus a correction factor.
Glass's delta uses the sample standard deviation of the control group.

Interpretation

- **Independent Samples t-test ($p < 0.001$):** Significant difference in electric range between BEVs (198.22 miles) and PHEVs (31.21 miles).
- **Levene's Test for Equality of Variances ($p < 0.001$):** We use the "Equal variances not assumed" results as unequal variances observed
- **t-value (489.552) and p-value (< 0.001):** Strong evidence of a significant difference in electric range among EV types
- **Effect Size (Cohen's d = 52.179):** Extremely large effect size, indicating a very strong difference in electric range between the two vehicle types.
- **Confidence Interval:** The range is between 166.34 and 167.68 miles, further confirming the large difference in the electric range.
- BEVs have a significantly higher electric range than PHEVs, with a very strong effect size.

Relationship Between Model Year & Electric Range

Pearson Correlations

Correlations

		ModelYear	ElectricRange
ModelYear	Pearson Correlation	1	-.185**
	Sig. (2-tailed)		<.001
	N	95335	95335
ElectricRange	Pearson Correlation	-.185**	1
	Sig. (2-tailed)	<.001	
	N	95335	95335

**. Correlation is significant at the 0.01 level (2-tailed).

Confidence Intervals

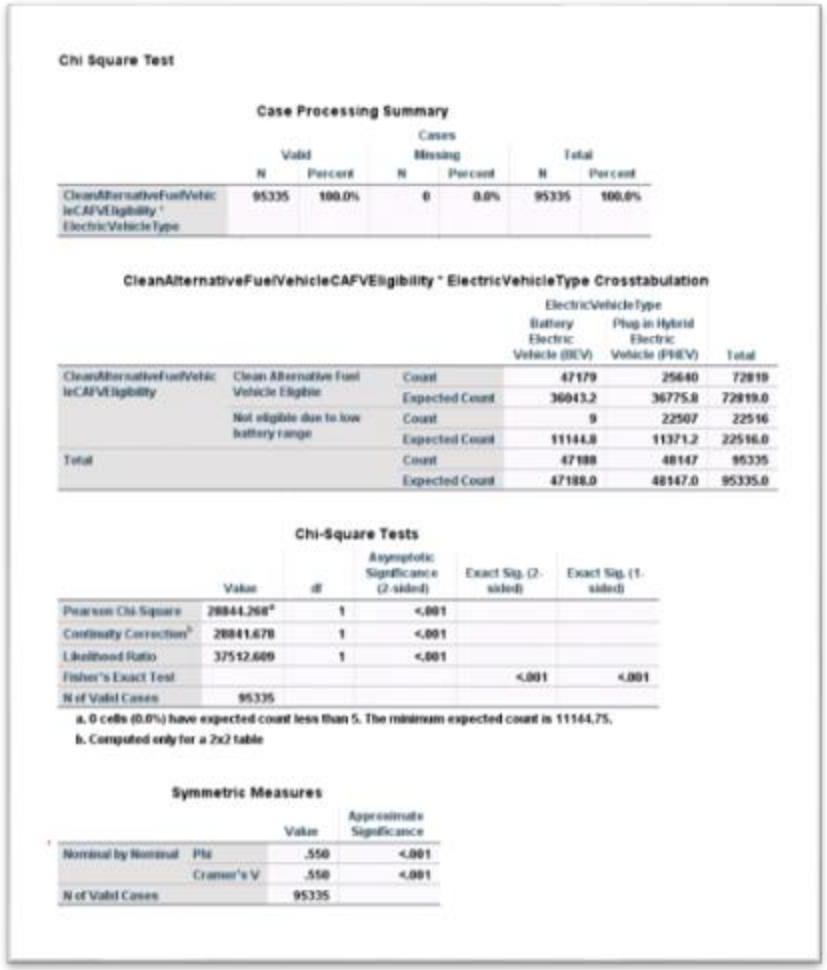
			95% Confidence Intervals (2-tailed)[a]	
	Pearson Correlation	Sig. (2-tailed)	Lower	Upper
ModelYear - ElectricRange	-.185	<.001	-.191	-.179

a. Estimation is based on Fisher's r-to-z transformation.

Interpretation

- **Pearson Correlation (-0.185):** Weak negative correlation between Model Year and Electric Range.
- **Statistical Significance (p-value < 0.001):** The correlation is statistically significant at 0.01.
- **Confidence Interval (-0.191 to -0.179):** A 95% confidence interval confirms the negative correlation and statistical significance.
- Newer model years EVs tend to have slightly lower electric ranges with a weak relationship.

CAFV Eligibility Across EV Types



Chi Square Test

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
CleanAlternativeFuelVehicleCAFVEligibility * ElectricVehicleType	95335	100.0%	0	0.0%	95335	100.0%

CleanAlternativeFuelVehicleCAFVEligibility * ElectricVehicleType Crosstabulation

			ElectricVehicleType		
			Battery Electric Vehicle (BEV)	Plug in Hybrid Electric Vehicle (PHEV)	Total
CleanAlternativeFuelVehicleCAFVEligibility	Clean Alternative Fuel Vehicle Eligible	Count	47179	25640	72819
		Expected Count	36043.2	36775.8	72819.0
	Not eligible due to low battery range	Count	9	22507	22516
		Expected Count	11144.8	11371.2	22516.0
Total		Count	47188	48147	95335
		Expected Count	47188.0	48147.0	95335.0

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	28844.268[a]	1	<.001		
Continuity Correction[b]	28841.678	1	<.001		
Likelihood Ratio	37512.609	1	<.001		
Fisher's Exact Test				<.001	<.001
N of Valid Cases	95335				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 11144.75.

b. Computed only for a 2x2 table

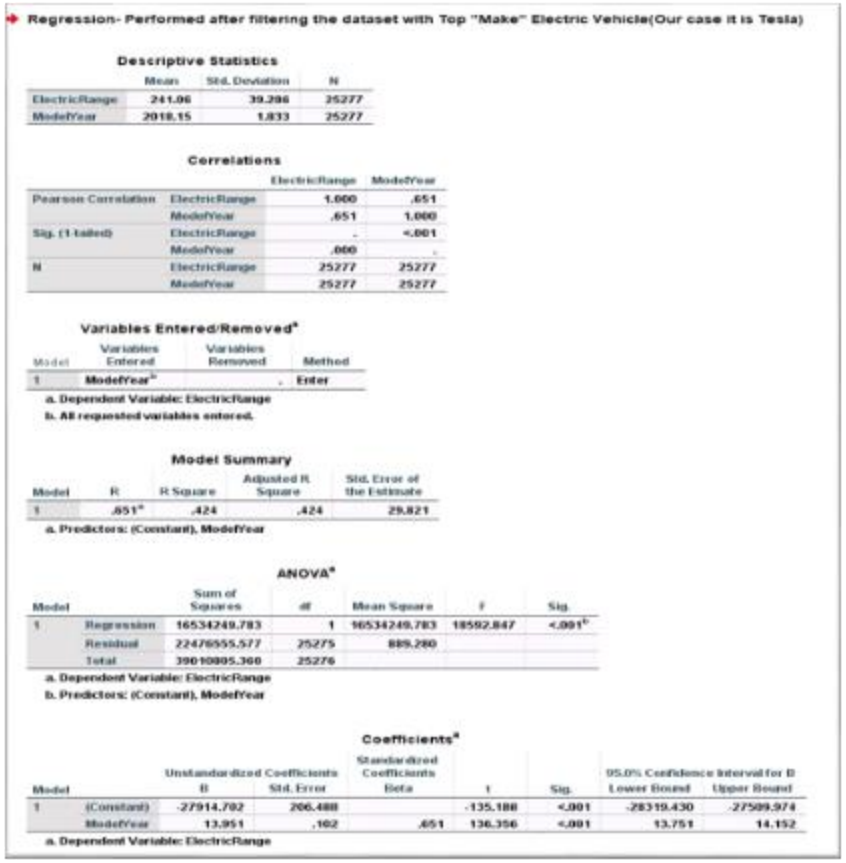
Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	.550	<.001
	Cramer's V	.550	<.001
N of Valid Cases		95335	

Interpretation

- **Chi-Square Test ($p < 0.001$):** Significant association between CAFV eligibility and electric vehicle type (BEV vs. PHEV).
- **Cramer's V (0.550):** Strong association between the two variables.
- **Observed vs. Expected Counts:** The observed count of BEVs and PHEVs eligible for CAFV differ broadly, suggesting that these variables are independent.
- The electric vehicle type significantly affects eligibility for CAFV incentives

Electric Range Significance with Model Years for Tesla EV



Regression- Performed after filtering the dataset with Top "Make" Electric Vehicle(Our case it is Tesla)

Descriptive Statistics

	Mean	Std. Deviation	N
ElectricRange	241.06	39.286	25277
ModelYear	2018.15	1.833	25277

Correlations

		ElectricRange	ModelYear
Pearson Correlation	ElectricRange	1.000	.651
	ModelYear	.651	1.000
Sig. (1-tailed)	ElectricRange	.	<.001
	ModelYear	.000	.
N	ElectricRange	25277	25277
	ModelYear	25277	25277

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	ModelYear[b]	.	Enter

a. Dependent Variable: ElectricRange
b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.651[a]	.424	.424	29.821

a. Predictors: (Constant), ModelYear

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	16534249.783	1	16534249.783	18592.847	<.001[b]
	Residual	22476555.577	25275	889.280		
	Total	39010805.360	25276			

a. Dependent Variable: ElectricRange
b. Predictors: (Constant), ModelYear

Coefficients[a]

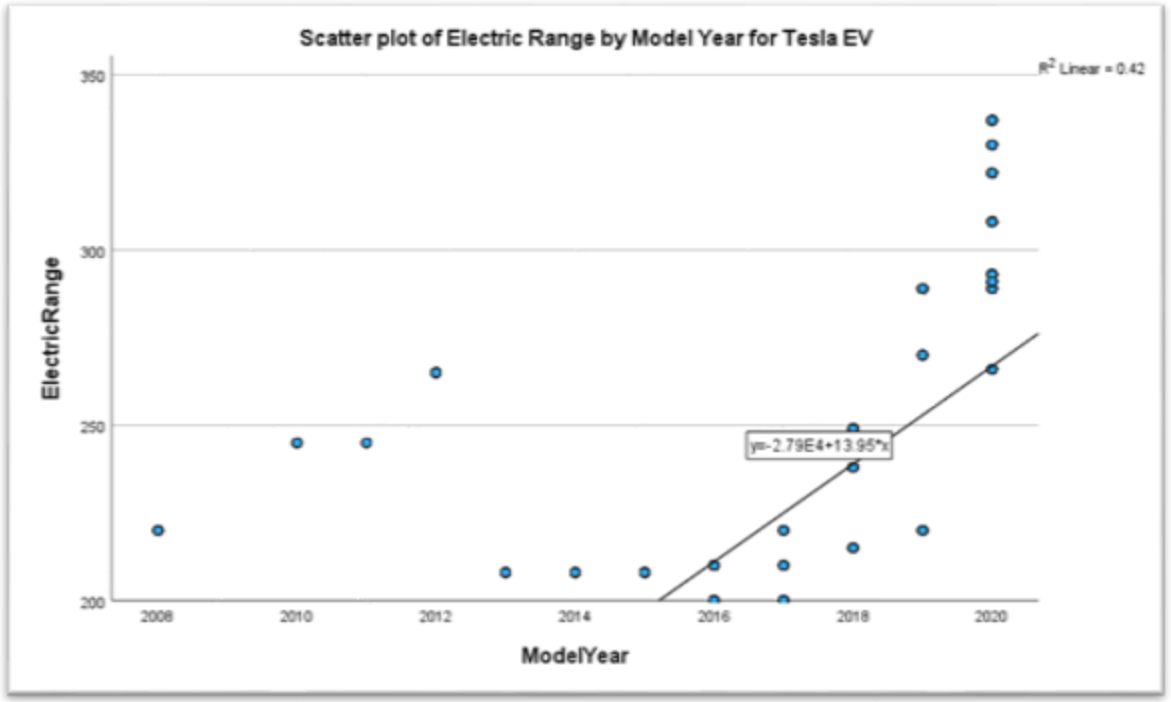
Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	-27914.702	206.488		-135.188	<.001	-28319.430	-27509.974
	ModelYear	13.951	.102	.651	136.356	<.001	13.751	14.152

a. Dependent Variable: ElectricRange

Interpretation

- **Pearson Correlation:** A moderate positive relationship (0.651) between model year and electric range.
- **Significance:** The relationship is statistically significant ($p < 0.001$), confirming it's not by chance.
- **R-squared:** 42.4% of the variance in the electric range can be explained by model year.
- **Unstandardised coefficient(b):** Each additional model year increases the electric range by 13.95 miles.
- **Standardised coefficient(β):** For each standard deviation increase in the model year, the electric range increases by 0.651 standard deviations.
- The model year significantly predicts the electric range for Tesla vehicles in Washington State.

Scatterplot to display relationship of Model year and EV Range for Tesla Cars



Interpretation

- Tesla's electric range has increased significantly.
- Model year alone explains 42.4% of the range variation, indicating moderate predictive power
- Most gains in range occurred after 2016.
- The scatter plot displays the positive relationship between model year and electric range.

Limitations

1

Base MSRP(lowest Manufacturer-suggested retail price) column has not been considered for analysis due to the number of observations having 0

2

The analysis doesn't represent all the regions restricted with EVs registered in Washington state alone.

3

Findings apply only to vehicles with known CAFV status and may not reflect the EV population.

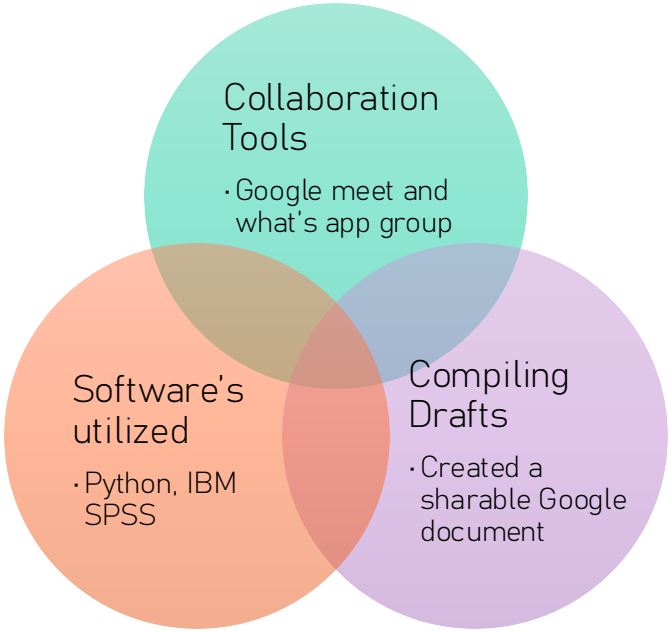
4

Other factors influencing EV adoption, such as charging infrastructure, price, etc, were not present in the dataset, limiting further analysis

RACI Chart- Teamwork

Assignment Tasks	Team Lead (Harika Aakula)	Research Strategist (Dinah Karia)	Data Analyst 1 (Mohammed Ateeq Ur Rahman)	Data Analyst 2 (Kranthi Kumar Duggirala)	Data Analyst 3 (Masthan Babu Padarthi)	Data Analyst 4 (Naresh Arepalli)	Content Curator (Venkat Reddy Meka)	Quality Specialist (Rajamouli Sanaka)
Research on EV Database & Selection	R	A	C	C	C	I	I	I
Dataset Cleaning	C	C	R	C	A	R	I	I
Dataset Analysis and Descriptive Statistics	R	I	A	R	R	R	I	I
Creating Research Questions	A	R	C	C	C	I	R	I
Addressing Research Questions and Interpretation	R	I	R	R	R	A	I	I
Draft Preparation and Compiling into Presentation	I	I	I	I	I	I	A	R
Feedback	R,A	C	C	A	C	C	R	C
Quality Check & Final Edits	C	I	I	I	I	I	R	A

R	Responsible
A	Accountable
C	Consulted
I	Informed



Conclusion/Key findings