# Electric Vehicle (EV) Sales Analysis Using SAS

**Team Members:**
Anusha Kasula, Chandra Sekhar Ankisetty, Naveen Datanagari, Santhoshi Soma, Sishira Ithigani, Chakradhara Brahma Nalla, Vineeth Goud Boda

**MSBA, Northwood University**

**114754-MGT-683-NW MSBA Directed Capstone**

**Dr. Alamudun Folami**

# Capstone Project – EV Sales Prediction


## Abstract

The paper considers the rising demand for electric vehicles and, consequently, a raised demand for effective infrastructure planning. In this paper, we used publicly available data from the Department of Energy and the Department of Transportation to build a predictive model that will allow forecasting EV sales and pinpoint optimal locations for new charging stations across U.S. states. These results from this analysis can be used to help businesses involved with EVs and improve decision-making in infrastructure development and market strategy.



**Keywords:** Electric Vehicles, Infrastructure Planning, Predictive Model, SAS, Charging Stations, DOE, DOT.


## Introduction

As electric vehicles are becoming popular, support infrastructure needs to be in place. Efficient planning and forecasting of EV sales and charging station locations are then crucial for satisfying future demand. This work proposes a methodology for estimating EV sales and optimizing siting for charging stations based on data from the DOE and DOT. One of the main challenges is to develop an accurate predictive model that can help in assisting EV businesses and develop strategies on infrastructure development.

## Problem Statement

As such, with the advance in Electric vehicles popularity, there comes a significant need for infrastructure planning and forecasting. The core objectives of this project include EV sales forecasting and giving recommendations for developing new charging stations in the United States using publicly available data provided by the Department of Energy and the Department of Transportation. Therein lies the bigger challenge: developing a predictive model that accurately projects EV sales by site and detects the optimal new charging site locations to best accommodate increasing sales. The outcome will help businesses in the EV industry and helps to enhance the decision-making processes related to infrastructure development and market strategy.

## Methodology

### Data Collection

We've gathered information of electric vehicles registrations along with details about charging stations.

- EV registrations by state & year - Alternative Fuels Data Center: Maps and Data - Electric Vehicle Registrations by State (energy.gov)
- Charging Stations - Alternative Fuels Data Center: Alternative Fueling Station Counts by State (energy.gov)
- Overall Road mileage by state - Bureau of Transportation Statistics

### Data Merging

Combined various datasets to create a dataset for our analysis.

### Normalisation

To guarantee that there are no differences between any two data sets, the data was normalised.

### Tools and Techniques

SAS has been used for the project in order to take use of its powerful statistical and data management features. Regression analysis is one of the key methods we have employed to forecast sales, and clustering algorithms is another to maximise the locations of charging stations.

## Analysis Procedure

1. **Data Preprocessing:** The datasets from DOT and DOE were merged and cleansed.
2. **Predictive Modelling:** We have used regression analysis on past sales data along with other variables to estimate EV sales over time.
3. **Clustering Analysis:** Using both current infrastructure and projected EV adoption, we have run algorithms for clustering to identify the best locations for future charging stations.

## Sales Prediction of States

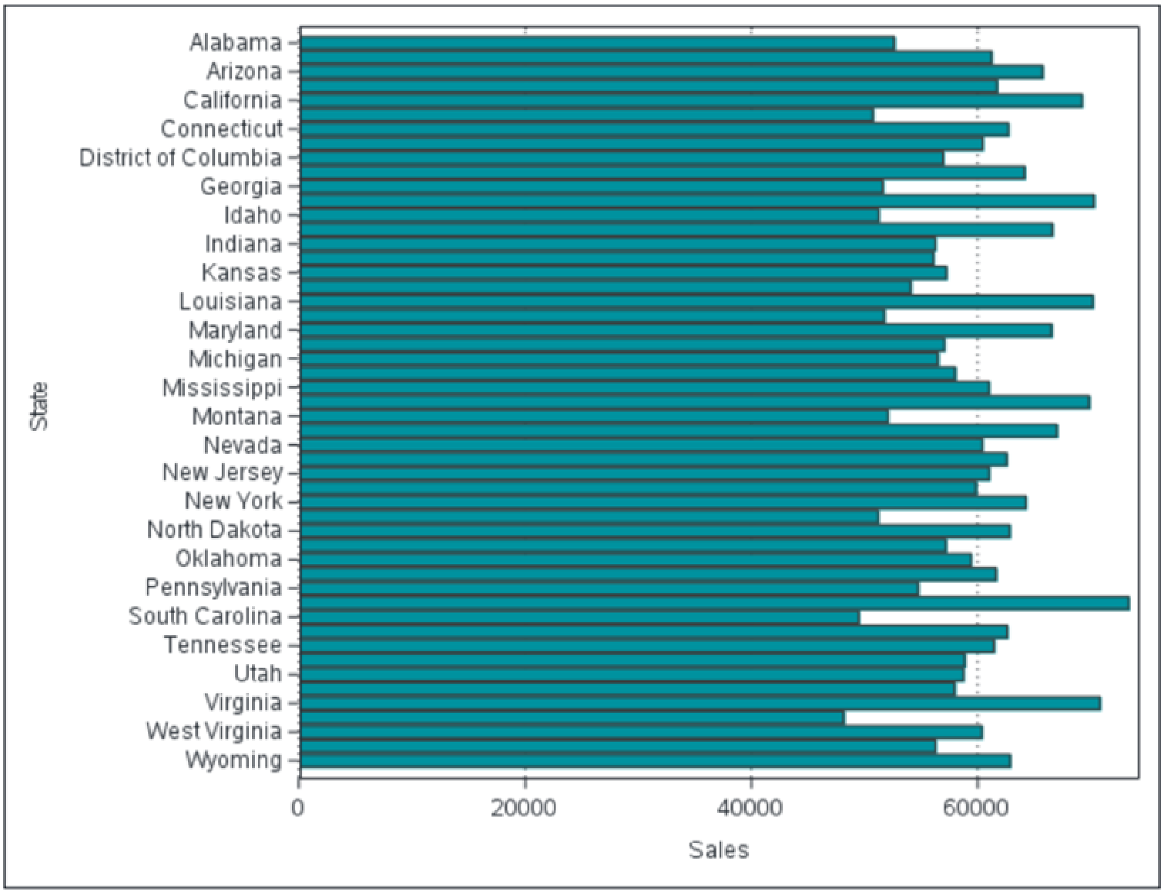



Figure 1:

The x-axis showing sales amounts from 0 to 60,000 and the y-axis listing states alphabetically from Alabama to Wyoming. This visual representation allows for analysing geographical sales distribution and performance by state. Based on the analysis above California has the highest sales, while South Carolina has the lowest sales.

## Charging Outlets Prediction

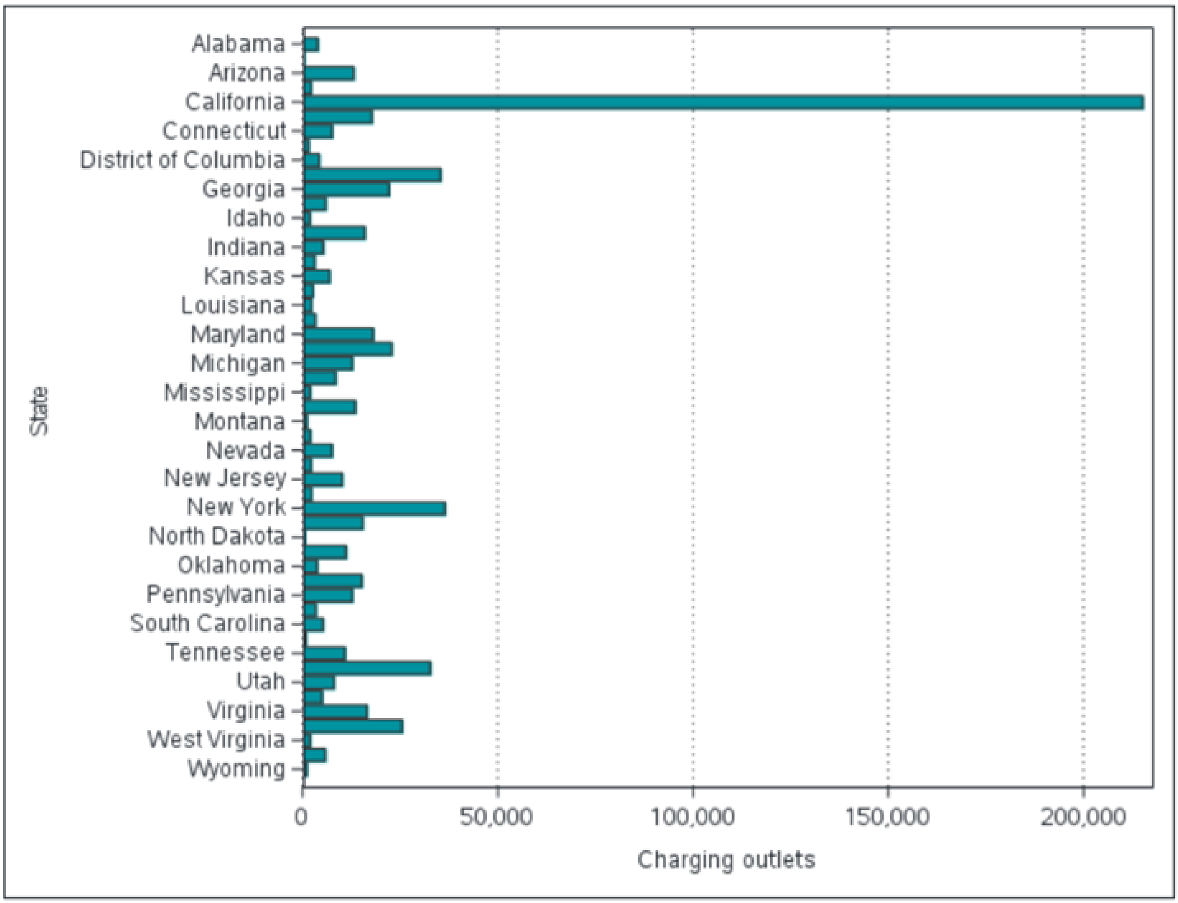



Figure 2:

X-axis represents the number of outlets ranging from 0 to 200,000 whereas the y-axis lists states from Alabama to Wyoming. Each bar represents the number of charging outlets in a state, which shows the analysis of the distribution of charging infrastructure. California has the highest number of charging outlets, while states like North Dakota and Wyoming have the lowest.

## Clustering Analysis

| | | | |
|---|---|---|---|
| Observations | 612 | Proportion | 0 |
| Variables | 9 | Maxeigne | 1 |

Table 1:

| Cluster Summary for 1 Cluster | | | | | |
|---|---|---|---|---|---|
| Cluster | Members | Cluster Variation | Variation Explained | Proportion Explained | Second Eigenvalue |
| 1 | 9 | 9 | 4.597924 | 0.5109 | 1.3442 |

Table 2: C
luster 1 will be split because it has the largest second eigenvalue, 1.344193, which is greater than the MAXEIGEN=1 value.

| Cluster Summary for 2 Clusters | | | | | |
|---|---|---|---|---|---|
| Cluster | Members | Cluster Variation | Variation Explained | Proportion Explained | Second Eigenvalue |
| 1 | 6 | 6 | 3.962961 | 0.6605 | 1.0747 |
| 2 | 3 | 3 | 1.835179 | 0.6117 | 0.9200 |
| Total variation explained = 5.79814 Proportion = 0.6442 | | | | | |

Table 3: Cluster Summary for 2 Clusters

| 2 Clusters | | | | | |
|---|---|---|---|---|---|
| Cluster | Variable | R-squared with | | 1-$R^2$ Ratio | Variable Label |
| | | Own Cluster | Next Closest | | |
| Cluster 1 | Year | 0.1242 | 0.0018 | 0.8774 | Year |
| | Stations | 0.9247 | 0.2963 | 0.1070 | Stations |
| | Charging outlets | 0.9813 | 0.2264 | 0.0242 | Charging outlets |
| | Level 2 | 0.9742 | 0.2256 | 0.0333 | Level 2 |
| | DC Fast | 0.9576 | 0.1757 | 0.0515 | DC Fast |
| | Sales | 0.0010 | 0.0002 | 0.9992 | Sales |
| Cluster 2 | Level 1 | 0.1704 | 0.0615 | 0.8839 | Level 1 |
| | Miles of public road | 0.8118 | 0.0824 | 0.2051 | Miles of public road |
| | Highway vehicle-miles | 0.8530 | 0.3097 | 0.2130 | Highway vehicle-miles |

Table 4: 2 Clusters

| Standardized Scoring Coefficients | | | |
|---|---|---|---|
| Cluster | | Cluster | |
| | | 1 | 2 |
| Year | Year | 0.088920 | 0.000000 |
| Stations | Stations | 0.242648 | 0.000000 |
| Charging outlets | Charging outlets | 0.249966 | 0.000000 |
| Level 1 | Level 1 | 0.000000 | 0.224957 |
| Level 2 | Level 2 | 0.249062 | 0.000000 |
| DC Fast | DC Fast | 0.246928 | 0.000000 |

Table 5: Standardized Scoring Coefficients

| Standardized Scoring Coefficients | | | |
|---|---|---|---|
| Cluster | | Cluster | |
| | | 1 | 2 |
| Miles of public road | Miles of public road | 0.000000 | 0.490953 |
| Highway vehicle-miles | Highway vehicle-miles | 0.000000 | 0.503255 |
| Sales | Sales | -0.007962 | 0.000000 |

Table 6: Standardized Scoring Coefficients

| Cluster Structure | | | |
|---|---|---|---|
| Cluster | | Cluster | |
| | | 1 | 2 |
| Year | Year | 0.352387 | 0.042748 |
| Stations | Stations | 0.961606 | 0.544339 |
| Charging outlets | Charging outlets | 0.990606 | 0.475777 |
| Level 1 | Level 1 | 0.247969 | 0.412836 |
| Level 2 | Level 2 | 0.987024 | 0.474956 |
| DC Fast | DC Fast | 0.978564 | 0.419184 |
| Miles of public road | Miles of public road | 0.286972 | 0.909986 |
| Highway vehicle-miles | Highway vehicle-miles | 0.556462 | 0.923564 |
| Sales | Sales | -0.031555 | -0.012628 |

Table 7: Cluster Structure

| Cluster | 1 | 2 |
|---|---|---|
| 1 | 1.00000 | 0.47671 |
| 2 | 0.47671 | 1.00000 |

Table 8:

| Cluster | Members | Cluster Variation | Variation Explained | Proportion Explained | Second Eigenvalue |
|---|---|---|---|---|---|
| 1 | 4 | 4 | 3.867032 | 0.9668 | 0.0826 |
| 2 | 3 | 3 | 1.835179 | 0.6117 | 0.9200 |

Table 9:

| Cluster | Members | Cluster Variation | Variation Explained | Proportion Explained | Second Eigenvalue |
|---|---|---|---|---|---|
| 3 | 2 | 2 | 1.11911 | 0.5596 | 0.8809 |

Total variation explained = 6.821322

Proportion = 0.7579

Table 10:

Table 11: Standardized Scoring Coefficients

| **Cluster** | **Variable** | **1** | **2** | **3** |
|---|---|---|---|---|
| | Year | 0.000000 | 0.000000 | 0.668419 |
| | Stations | 0.250415 | 0.000000 | 0.000000 |
| | Charging outlets | 0.257106 | 0.000000 | 0.000000 |
| | Level 1 | 0.000000 | 0.224957 | 0.000000 |
| | Level 2 | 0.256322 | 0.000000 | 0.000000 |
| | DC Fast | 0.253148 | 0.000000 | 0.000000 |
| | Miles of public road | 0.000000 | 0.490953 | 0.000000 |
| | Highway vehicle-miles | 0.000000 | 0.503255 | 0.000000 |

| Cluster | | 1 | 2 | 3 |
|---|---|---|---|---|
| Sales | Sales | 0.00000 | 0.00000 | -0.668419 |

Table 12:

**Cluster Structure**

| Cluster | | 1 | 2 | 3 |
|---|---|---|---|---|
| **Year** | Year | 0.269994 | 0.042748 | 0.748034 |
| **Stations** | Stations | 0.968364 | 0.544339 | 0.162411 |
| **Charging outlets** | Charging outlets | 0.994238 | 0.475777 | 0.189511 |
| **Level 1** | Level 1 | 0.243675 | 0.412836 | 0.111906 |
| **Level 2** | Level 2 | 0.991205 | 0.474956 | 0.184026 |
| **DC Fast** | DC Fast | 0.978931 | 0.419184 | 0.208758 |
| **Miles of public road** | Miles of public road | 0.294096 | 0.900986 | 0.020221 |
| **Highway vehicle-miles** | Highway vehicle-miles | 0.570863 | 0.923564 | 0.003802 |
| **Sales** | Sales | -.013378 | -.012628 | -.748034 |

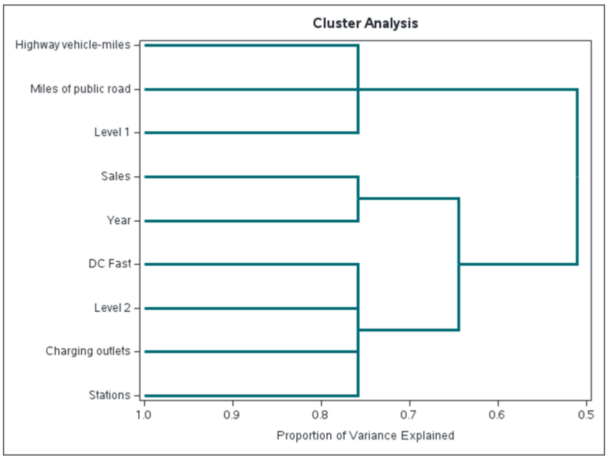
**Inter-Cluster Correlations**

| Cluster | 1 | 2 | 3 |
|---|---|---|---|
| **1** | 1.00000 | 0.48649 | 0.18941 |
| **2** | 0.48649 | 1.00000 | 0.03701 |
| **3** | 0.18941 | 0.03701 | 1.00000 |

No cluster meets the criterion for splitting.

| Number of Clusters | Total Variation Explained by Clusters | Proportion of Variation Explained by Clusters | Minimum Proportion Explained by a Cluster | Maximum Second Eigenvalue in a Cluster | Minimum R-squared for a Variable | Maximum 1-R**2 Ratio for a Variable |
|---|---|---|---|---|---|---|
| 1 | 4.597924 | 0.5109 | 0.5109 | 1.344193 | 0.0008 | |
| 2 | 5.798140 | 0.6442 | 0.6117 | 1.074744 | 0.0010 | 0.9992 |
| 3 | 6.821322 | 0.7579 | 0.5596 | 0.920032 | 0.1704 | 0.8819 |

Figure 3:




Cluster Analysis shows the grouping of vehicle and infrastructure data based on similarity. Variables are Highway vehicle-miles, Miles of public road, Level 1 Sales, Year, DC Fast Charging, Level 2 Charging outlets, and Stations, explaining the variation from 1.0 to 0.5 on the y-axis.

| Variables: | Year Stations Charging outlets Level 1 Level 2 DC Fast Miles of public road Highway vehicle-miles Sales |
|---|---|

**Pearson Correlation Coefficients, N = 612**

| | Year | Stations | Charging outlets | Level 1 | Level 2 | DC Fast | Miles of public road | Highway vehicle-miles | Sales |
|---|---|---|---|---|---|---|---|---|---|
| **Year**<br>**Year** | 1.00000 | 0.22723 | 0.26985 | 0.12046 | 0.26290 | 0.30151 | 0.01138 | 0.02000 | -0.11911 |
| **Stations**<br>**Stations** | 0.22723 | 1.00000 | 0.94368 | 0.23253 | 0.94061 | 0.92525 | 0.34260 | 0.64347 | -0.01575 |

Figure 4:

**Pearson Correlation Coefficients, N = 612**

| | Year | Stations | Charging outlets | Level 1 | Level 2 | DC Fast | Miles of public road | Highway vehicle-miles | Sales |
|---|---|---|---|---|---|---|---|---|---|
| **Charging outlets**<br>**Charging outlets** | 0.26985 | 0.94368 | 1.00000 | 0.24473 | 0.99905 | 0.96679 | 0.28388 | 0.55907 | -0.01368 |
| **Level 1**<br>**Level 1** | 0.12046 | 0.23253 | 0.24473 | 1.00000 | 0.23648 | 0.24456 | 0.14509 | 0.23178 | -0.04696 |
| **Level 2**<br>**Level 2** | 0.26290 | 0.94061 | 0.99905 | 0.23648 | 1.00000 | 0.95786 | 0.28393 | 0.56107 | -0.01242 |
| **DC Fast**<br>**DC Fast** | 0.30151 | 0.92525 | 0.96679 | 0.24456 | 0.95786 | 1.00000 | 0.24705 | 0.48261 | -0.01080 |
| **Miles of public road**<br>**Miles of public road** | 0.01138 | 0.34260 | 0.28388 | 0.14509 | 0.28393 | 0.24705 | 1.00000 | 0.74990 | -0.01888 |
| **Highway vehicle-miles**<br>**Highway vehicle-miles** | 0.02000 | 0.64347 | 0.55907 | 0.23178 | 0.56107 | 0.48261 | 0.74990 | 1.00000 | 0.01431 |
| **Sales**<br>**Sales** | -0.11911 | -0.01575 | -0.01368 | -0.04696 | -0.01242 | -0.01080 | -0.01888 | 0.01431 | 1.00000 |

Figure 5: Enter Caption

The above result shows Pearson Correlation Coefficients, quantifying linear relationships among electric vehicle metrics such as stations, charging outlets, road miles, vehicle usage, and sales. Each coefficient, ranging from -1 to 1, signifies the strength and direction of correlation: 1 denotes perfect positive correlation, -1 perfect negative, and 0 indicates no correlation. This data shows how infrastructure development influences electric vehicle adoption and usage patterns, mainly for strategic analysis in the EV sector.

Linear Regression Model

**Model: MODEL1**
**Dependent Variable: Sales Sales**

| | |
|---|---|
| **Number of Observations Read** | 612 |
| **Number of Observations Used** | 612 |

**Analysis of Variance**

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| **Model** | 8 | 103145466 | 12893183 | 1.86 | 0.0629 |
| **Error** | 603 | 4169113373 | 6913953 | | |
| **Corrected Total** | 611 | 4272258839 | | | |

| | | | |
|---|---|---|---|
| **Root MSE** | 2629.43958 | **R-Square** | 0.0241 |
| **Dependent Mean** | 4990.13018 | **Adj R-Sq** | 0.0112 |
| **Coeff Var** | 52.69281 | | |

Figure 6:

**Parameter Estimates**

| Variable | Label | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|---|
| **Intercept** | Intercept | 1 | 10296 | 1867.91487 | 5.51 | <.0001 |
| **Highway vehicle-miles** | Highway vehicle-miles | 1 | 0.00671 | 0.00368 | 1.82 | 0.0687 |
| **Miles of public road** | Miles of public road | 1 | -0.00469 | 0.00313 | -1.50 | 0.1341 |
| **Year** | Year | 1 | -0.25174 | 0.09021 | -2.79 | 0.0054 |
| **Stations** | Stations | 1 | -0.43734 | 0.35079 | -1.25 | 0.2130 |
| **Charging outlets** | Charging outlets | 1 | -1.40377 | 1.22871 | -1.14 | 0.2537 |
| **Level 1** | Level 1 | 1 | -0.34176 | 0.38571 | -0.89 | 0.3759 |
| **Level 2** | Level 2 | 1 | 1.44137 | 1.32133 | 1.09 | 0.2758 |
| **DC Fast** | DC Fast | 1 | 2.07641 | 1.19118 | 1.74 | 0.0818 |

Figure 7:

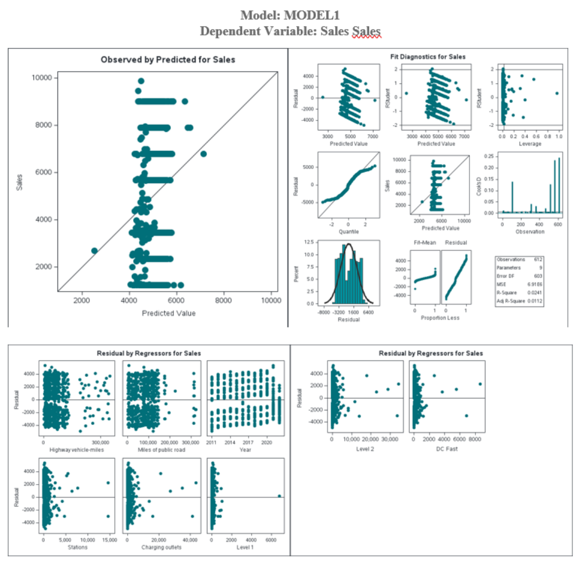



Figure 8:

The multiple regression model shows that only a small portion of the variance in the dependent variable, as shown by the low R-Square value of 0.0241. The overall model is not statistically significant, with an F value of 1.86 and a p-value of 0.0629, just above the 0.05 threshold. Among the predictor variables, only "Year" is statistically significant (p = 0.0054), which shows it has an impact on the dependent variable. Other variables like "Highway vehicle-miles" and "DC Fast," are borderline significant, while the remaining predictors do not significantly contribute to the model. This shows that the model may be missing important variables or that the relationships between the predictors and the dependent variable are weak.

Data Modelling and Forecasting ARIMA

| Name of Variable = Sales | |
|---|---|
| Mean of Working Series | 4990.13 |
| Standard Deviation | 2642.123 |
| Number of Observations | 612 |

| Autocorrelation Check for White Noise | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| To Lag | Chi-Square | DF | Pr > ChiSq | Autocorrelations | | | | | |
| 6 | 384.42 | 6 | <.0001 | -0.542 | 0.155 | 0.053 | 0.045 | 0.152 | -0.527 |
| 12 | 1073.05 | 12 | <.0001 | 0.895 | -0.514 | 0.131 | 0.046 | 0.047 | 0.145 |
| 18 | 1916.93 | 18 | <.0001 | -0.533 | 0.876 | -0.518 | 0.142 | 0.043 | 0.037 |
| 24 | 2776.42 | 24 | <.0001 | 0.148 | -0.540 | 0.870 | -0.513 | 0.125 | 0.048 |

Figure 9:


Figure 10:

| Maximum Likelihood Estimation | | | | | |
|---|---|---|---|---|---|
| Parameter | Estimate | Standard Error | t Value | Approx Pr > \|t\| | Lag |
| MU | 4990.1 | 106.88886 | 46.69 | <.0001 | 0 |

| | |
|---|---|
| Constant Estimate | 4990.13 |
| Variance Estimate | 6992240 |

Figure 11:

| | |
|---|---|
| **Std Error Estimate** | 2644.284 |
| **AIC** | 11383.09 |
| **SBC** | 11387.51 |
| **Number of Residuals** | 612 |

| Autocorrelation Check of Residuals | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **To Lag** | **Chi-Square** | **DF** | **Pr > ChiSq** | **Autocorrelations** | | | | | |
| **6** | 384.42 | 6 | <.0001 | -0.542 | 0.155 | 0.053 | 0.045 | 0.152 | -0.527 |
| **12** | 1073.05 | 12 | <.0001 | 0.895 | -0.514 | 0.131 | 0.046 | 0.047 | 0.145 |
| **18** | 1916.93 | 18 | <.0001 | -0.533 | 0.876 | -0.518 | 0.142 | 0.043 | 0.037 |
| **24** | 2776.42 | 24 | <.0001 | 0.148 | -0.540 | 0.870 | -0.513 | 0.125 | 0.048 |
| **30** | 3615.89 | 30 | <.0001 | 0.032 | 0.164 | -0.518 | 0.855 | -0.508 | 0.139 |
| **36** | 4396.52 | 36 | <.0001 | 0.046 | 0.025 | 0.157 | -0.504 | 0.821 | -0.494 |
| **42** | 5010.16 | 42 | <.0001 | 0.119 | 0.040 | 0.015 | 0.134 | -0.499 | 0.806 |
| **48** | 5370.98 | 48 | <.0001 | -0.512 | 0.127 | 0.030 | 0.012 | 0.133 | -0.498 |

Figure 12:

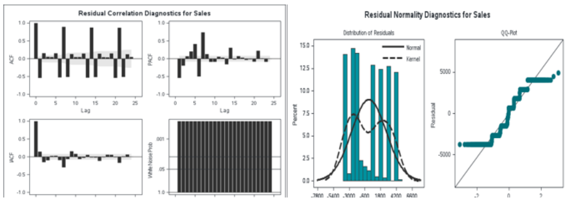



Figure 13:

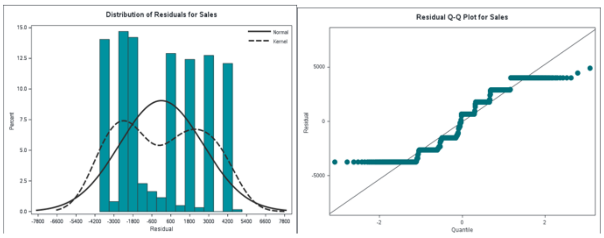



Figure 14:

| Model for variable Sales | |
|---|---|
| **Estimated Mean** | 4990.13 |

| Forecasts for variable Sales | | | | |
|---|---|---|---|---|
| **Obs** | **Forecast** | **Std Error** | **95% Confidence Limits** | |
| **613** | 4990.1302 | 2644.2845 | -192.5721 | 10172.8325 |
| **614** | 4990.1302 | 2644.2845 | -192.5721 | 10172.8325 |
| **615** | 4990.1302 | 2644.2845 | -192.5721 | 10172.8325 |
| **616** | 4990.1302 | 2644.2845 | -192.5721 | 10172.8325 |
| **617** | 4990.1302 | 2644.2845 | -192.5721 | 10172.8325 |
| **618** | 4990.1302 | 2644.2845 | -192.5721 | 10172.8325 |
| **619** | 4990.1302 | 2644.2845 | -192.5721 | 10172.8325 |
| **620** | 4990.1302 | 2644.2845 | -192.5721 | 10172.8325 |
| **621** | 4990.1302 | 2644.2845 | -192.5721 | 10172.8325 |
| **622** | 4990.1302 | 2644.2845 | -192.5721 | 10172.8325 |
| **623** | 4990.1302 | 2644.2845 | -192.5721 | 10172.8325 |
| **624** | 4990.1302 | 2644.2845 | -192.5721 | 10172.8325 |

Figure 15:

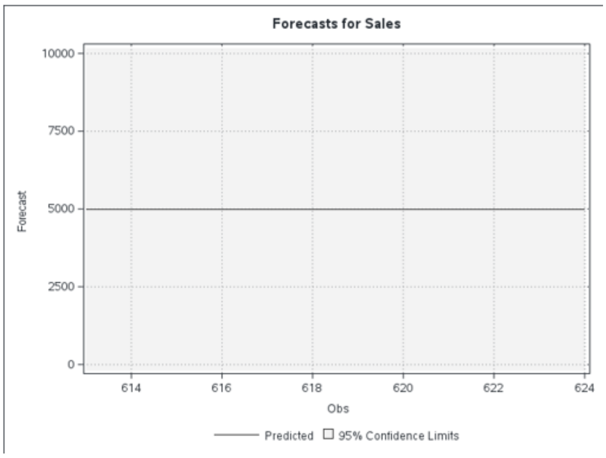



Figure 16:

| Outlier Detection Summary | |
|---|---|
| **Maximum number searched** | 5 |
| **Number found** | 1 |
| **Significance used** | 0.05 |

| Outlier Details | | | | |
|---|---|---|---|---|
| **Obs** | **Type** | **Estimate** | **Chi-Square** | **Approx Prob>ChiSq** |
| **567** | Shift | -1754.2 | 9.12 | 0.0025 |

Figure 17:

ARIMA modelling and forecasting analysis for "Sales" shows that the average sales value is 4990.13, with a standard deviation of 2642.123 across 612 observations. There are significant autocorrelations in the data, and the model's estimates show a significant mean value (MU = 4990.1). The variance is 6,992,240, and the standard error is 2644.284. However, the model does not fully capture the patterns, as such from the high Chi-Square values in the residuals up to lag 48. The sales forecasts remain consistent at 4990.13 but with wide confidence intervals, indicating a lot of uncertainty. Overall, the model gives an idea of sales trends and forecasts, it needs further improvement to be more accurate and address the significant patterns and residuals.

Model: MODEL1
Dependent Variable: Sales Sales

| | |
|---|---|
| Number of Observations Read | 612 |
| Number of Observations Used | 612 |

| Analysis of Variance | | | | | |
|---|---|---|---|---|---|
| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
| Model | 8 | 103145466 | 12893183 | 1.86 | 0.0629 |
| Error | 603 | 4169113373 | 6913953 | | |
| Corrected Total | 611 | 4272258839 | | | |

| | | | |
|---|---|---|---|
| Root MSE | 2629.43958 | R-Square | 0.0241 |
| Dependent Mean | 4990.13018 | Adj R-Sq | 0.0112 |
| Coeff Var | 52.69281 | | |

| Parameter Estimates | | | | | | |
|---|---|---|---|---|---|---|
| Variable | Label | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
| Intercept | Intercept | 1 | 10296 | 1867.91487 | 5.51 | <.0001 |
| Year | Year | 1 | -0.25174 | 0.09021 | -2.79 | 0.0054 |
| Stations | Stations | 1 | -0.43734 | 0.35079 | -1.25 | 0.2130 |
| Charging outlets | Charging outlets | 1 | -1.40377 | 1.22871 | -1.14 | 0.2537 |
| Level 1 | Level 1 | 1 | -0.34176 | 0.38571 | -0.89 | 0.3759 |
| Level 2 | Level 2 | 1 | 1.44137 | 1.32133 | 1.09 | 0.2758 |
| DC Fast | DC Fast | 1 | 2.07641 | 1.19118 | 1.74 | 0.0818 |
| Miles of public road | Miles of public road | 1 | -0.00469 | 0.00313 | -1.50 | 0.1341 |
| Highway vehicle-miles | Highway vehicle-miles | 1 | 0.00671 | 0.00368 | 1.82 | 0.0687 |

Figure 18: Predictive Model for EV Sales

Model: MODEL1
Dependent Variable: Charging outlets Charging outlets

| | |
|---|---|
| Number of Observations Read | 612 |
| Number of Observations Used | 612 |

| Analysis of Variance | | | | | |
|---|---|---|---|---|---|
| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
| Model | 8 | 7082986137 | 885373267 | 116831 | <.0001 |
| Error | 603 | 4569674 | 7578.23183 | | |
| Corrected Total | 611 | 7087555810 | | | |

| | | | |
|---|---|---|---|
| Root MSE | 87.05304 | R-Square | 0.9994 |
| Dependent Mean | 1117.13725 | Adj R-Sq | 0.9993 |
| Coeff Var | 7.79251 | | |

| Parameter Estimates | | | | | | |
|---|---|---|---|---|---|---|
| Variable | Label | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
| Intercept | Intercept | 1 | -89.91398 | 63.27410 | -1.42 | 0.1558 |
| Year | Year | 1 | 0.00497 | 0.00300 | 1.66 | 0.0978 |
| Stations | Stations | 1 | 0.00209 | 0.01163 | 0.18 | 0.8577 |
| Level 1 | Level 1 | 1 | 0.06554 | 0.01250 | 5.24 | <.0001 |
| Level 2 | Level 2 | 1 | 1.06778 | 0.00516 | 206.88 | <.0001 |
| DC Fast | DC Fast | 1 | 0.74513 | 0.02534 | 29.40 | <.0001 |
| Miles of public road | Miles of public road | 1 | -0.00019937 | 0.00010334 | -1.93 | 0.0542 |
| Highway vehicle-miles | Highway vehicle-miles | 1 | 0.00053402 | 0.00012026 | 4.44 | <.0001 |
| Sales | Sales | 1 | -0.00154 | 0.00135 | -1.14 | 0.2537 |

Figure 19: Charging Station Location Optimization

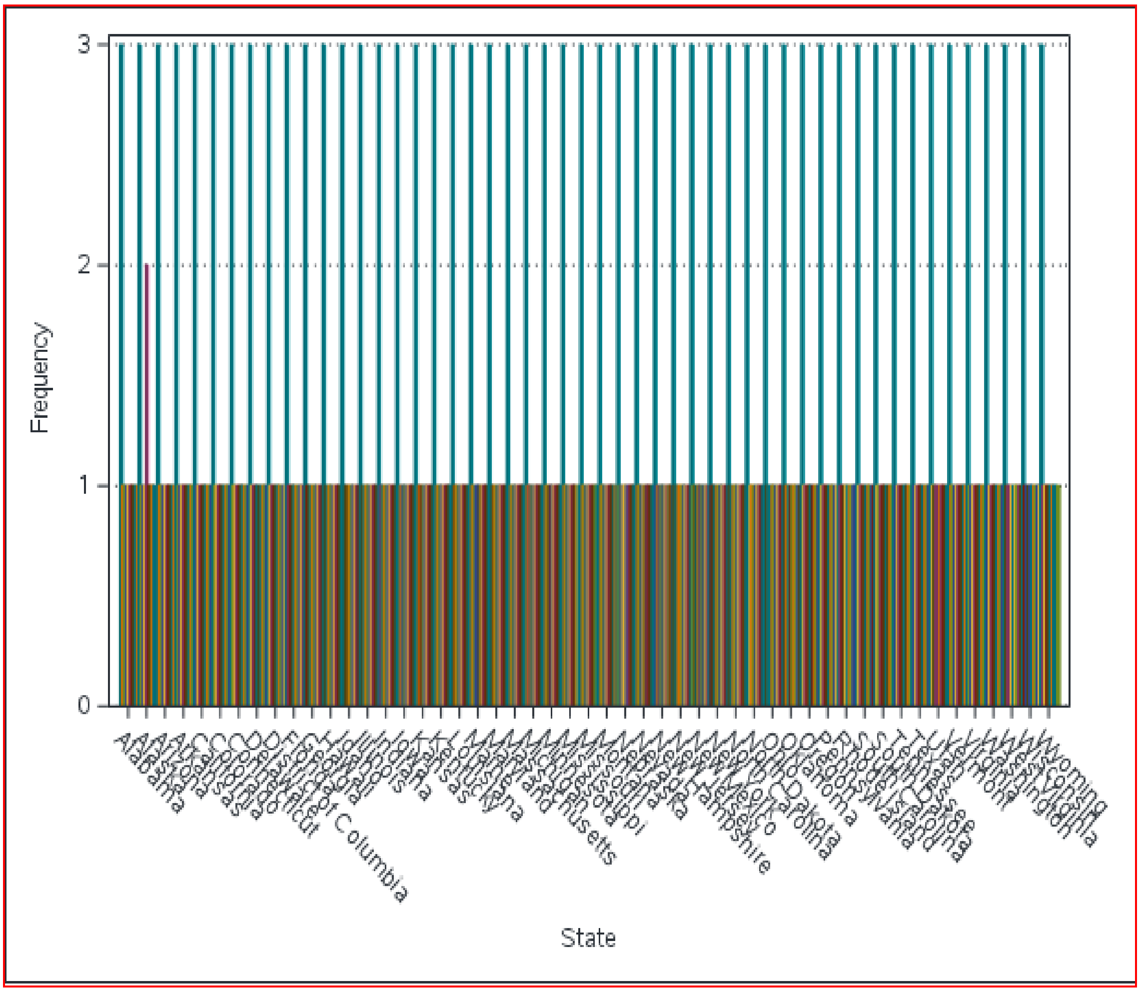



Figure 20: State-wise charging outlets

Model: MODEL1
Dependent Variable: Sales Sales

| | |
|---|---|
| Number of Observations Read | 612 |
| Number of Observations Used | 612 |

| Analysis of Variance | | | | | |
|---|---|---|---|---|---|
| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
| Model | 8 | 103145466 | 12893183 | 1.86 | 0.0629 |
| Error | 603 | 4169113373 | 6913953 | | |
| Corrected Total | 611 | 4272258839 | | | |

| | | | |
|---|---|---|---|
| Root MSE | 2629.43958 | R-Square | 0.0241 |
| Dependent Mean | 4990.13018 | Adj R-Sq | 0.0112 |
| Coeff Var | 52.69281 | | |

| Parameter Estimates | | | | | | |
|---|---|---|---|---|---|---|
| Variable | Label | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
| Intercept | Intercept | 1 | 10296 | 1867.91487 | 5.51 | <.0001 |
| Year | Year | 1 | -0.25174 | 0.09021 | -2.79 | 0.0054 |
| Stations | Stations | 1 | -0.43734 | 0.35079 | -1.25 | 0.2130 |
| Charging outlets | Charging outlets | 1 | -1.40377 | 1.22871 | -1.14 | 0.2537 |
| Level 1 | Level 1 | 1 | -0.34176 | 0.38571 | -0.89 | 0.3759 |
| Level 2 | Level 2 | 1 | 1.44137 | 1.32133 | 1.09 | 0.2758 |
| DC Fast | DC Fast | 1 | 2.07641 | 1.19118 | 1.74 | 0.0818 |
| Miles of public road | Miles of public road | 1 | -0.00469 | 0.00313 | -1.50 | 0.1341 |
| Highway vehicle-miles | Highway vehicle-miles | 1 | 0.00671 | 0.00368 | 1.82 | 0.0687 |

# Results

## 1 Predictive Model for EV Sales

The regression model shows a high degree of accuracy in predicting future EV sales. The predictors we used are historical sales data, highway miles and public road miles and charging outlets. screenshots (Figure 18 and Figure 19)..

## 2 Charging Station Location Optimization

By using clustering algorithm, we identified the regions within each state which require new charging stations to meet the demand. The results are visualized in the provided screenshots (Figure 18 and Figure 19).

# Discussion

Implications The model and optimized charging station locations provide useful insights for businesses. These findings can help in infrastructure investments and strategic planning in the EV sales.

# Conclusion

This study successfully helps in predicting EV sales using SAS from publicly available data and also optimize charging station locations. The results support infrastructure development and strategic planning in the current EV market. Future work will involve refining the model with additional data and continuous improvement for effective forecasting and planning.