Research on Analysis and Prediction of Electric Vehicle Ownership in Washington State, USA, Based on Data Science Technologies

Jingbo Yu*

Department of Mathematics and Computer Science, Adelphi University, New York 11530, United States

Abstract: The rapid global shift towards sustainable transportation highlights the significance of understanding the adoption patterns of electric vehicles (EVs). This study utilizes data science techniques to analyze the trend of EV ownership in Washington State, USA, which is a frontier region for clean energy initiatives. By systematically collecting, preprocessing, and analyzing EV registration data, this study identified key factors driving market growth, including policy incentives, infrastructure development, and socio-economic variables. Exploratory data analysis and predictive modeling (utilizing linear regression and K-means clustering) revealed a significant upward trend in EV adoption, and different regional clusters highlighted different adoption rates. These findings provide actionable insights for policymakers to optimize incentive measures and enterprises to customize market strategies. Moreover, this study offers support for broader discussions on sustainable transportation by providing a data-driven framework for future research. The research results not only validate the progressive policies in Washington State but also serve as a benchmark for other regions aiming to accelerate EV adoption.

Keywords: Electric Vehicles; Data Science; Linear Regression; K-means Clustering; Washington State; Ownership Prediction

DOI: 10.57237/j.cst.2025.03.001

1 Introduction

With the continuous increase in global attention to environmental protection and sustainable development, electric vehicles (EVs) have become a key technology for reducing greenhouse gas emissions and mitigating climate change [1, 11, 13]. Governments and industries around the world have promoted the adoption of electric vehicles through policies and innovation, but due to differences in infrastructure, incentives and consumer behavior, the adoption rate still shows regional differences [2]. Washington State, as a leader in clean energy initiatives, embodies this trend, and its electric vehicle ownership rate exceeds the national average [3]. Its progressive policies, such as tax exemptions and charging infrastructure investment, make it an ideal case study for analyzing the market dynamics of electric vehicles [4].

This study investigated the electric vehicle ownership dataset in Washington State to identify the key factors influencing key variables, including electric vehicle range, geographical distribution, vehicle model year and vehicle type. Recent studies have emphasized the role of socio-economic factors (such as income levels) and policy interventions (such as rebates) in accelerating the adoption of electric vehicles [5], while other studies have emphasized technological advances (such as battery efficiency) as driving factors [6]. However, there are still some shortcomings in current research: a) limited regional

*Corresponding author: Jingbo Yu, jingboyu@mail.adelphi.edu

Received: 18 February 2025; Accepted: 23 April 2025; Published Online: 30 June 2025 http://www.computscitech.com

studies on the electric vehicle trends in Washington State after 2020 [7]; b) insufficient integration of spatial and temporal factors in predictive modeling [8]; and c) the need for data-driven policy recommendations tailored to rapid market changes [9].

By leveraging data science techniques, this paper expands on these findings and provides actionable insights for stakeholders. The research results will inform policymakers in optimizing incentives, manufacturers in adjusting production according to demand, and consumers in making purchase decision [14]. This analysis aims to address the three gaps in the existing literature: a) limited regional studies on electric vehicle trends in Washington State after 2020 [7], b) insufficient integration of spatial and temporal factors in predictive modeling [8], and c) the need for data-driven policy recommendations tailored to rapid market changes [9]. The research results will provide information for policymakers to optimize incentives, manufacturers to adjust production according to demand, and consumers to make purchase decisions.

2 Overview of the Dataset

This dataset is sourced from the Washington State Department of Licensing (DOL) and contains over 180,000 records of electric vehicle (EV) information, including Battery Electric Vehicles (BEVs) and Plug-in Hybrid Electric Vehicles (PHEVs). The dataset encompasses multiple dimensions such as VIN (Vehicle Identification Number), county, city, state, ZIP code, model year, make, model, electric vehicle type, CAFV (Clean Alternative Fuel Vehicle) eligibility, electric range, and more. The main variables in the dataset include:

1. *Model Year:* The year of the vehicle model.
2. *Make:* The brand of the vehicle.
3. *Model:* The model of the vehicle.
4. *Electric Vehicle Type:* The type of electric vehicle (e.g., BEV, PHEV).
5. *CAFV Eligibility:* Eligibility for Clean Alternative Fuel Vehicle status.
6. *Electric Range:* The electric range in miles.
7. *Base MSRP:* The manufacturer's suggested retail price.

Electric Utility: The electric utility provider.

The dataset is available on Kaggle via the following link:
https://www.kaggle.com/datasets/mariusborel/electric-vehicle-population-data

3 Data Analysis Methods

3.1 Linear Regression Analysis

Objective: To perform linear regression analysis using R, a programming language and environment for statistical computing and graphics, to explore the impact of various factors on the electric range of electric vehicles. The analysis includes data exploration, model construction, evaluation, and interpretation of results [15, 16]. The specific steps are as follows:

3.1.1 Verify Data Assumptions

1) Load the Dataset

In R, download and load the dataset named ev_data, and use the ggplot2, dplyr, and caret packages for data analysis and visualization. First, install the ggplot2, dplyr, and caret packages:

```r
install.packages("ggplot2")
install.packages("dplyr")
install.packages("caret")
```

Figure 1 Example code for installing R packages

After installation, load these packages:

```r
library(ggplot2)
library(dplyr)
library(caret)
```

Figure 2 Example code for loading a library in the R

Download and load the dataset:

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(caret)
```

```
## Warning: package 'caret' was built under R version 4.4.2
```

```
## Loading required package: lattice
```

```
# Load the dataset
ev_data <- read.csv("C:/Users/lenovo/OneDrive/桌面/datasci/ev_data.csv")
```

Figure 3 Example code for downloading and loading a dataset in the R

2) Explore the Dataset

```
# Explore the dataset
summary(ev_data)
```

```
##   VIN..1.10.           County              City              State
##  Length:205439      Length:205439      Length:205439      Length:205439
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##   Postal.Code      Model.Year       Make              Model
##  Min.   : 1731   Min.   :1997   Length:205439      Length:205439
##  1st Qu.:98052   1st Qu.:2019   Class :character   Class :character
##  Median :98125   Median :2022   Mode  :character   Mode  :character
##  Mean   :98178   Mean   :2021
##  3rd Qu.:98372   3rd Qu.:2023
##  Max.   :99577   Max.   :2025
##  NA's   :3
##  Electric.Vehicle.Type Clean.Alternative.Fuel.Vehicle..CAFV..Eligibility
##  Length:205439         Length:205439
##  Class :character      Class :character
##  Mode  :character      Mode  :character
##
##
##
##
##  Electric.Range     Base.MSRP        Legislative.District DOL.Vehicle.ID
##  Min.   :  0.00   Min.   :     0.0   Min.   : 1.00        Min.   :    4469
##  1st Qu.:  0.00   1st Qu.:     0.0   1st Qu.:17.00        1st Qu.:193532440
##  Median :  0.00   Median :     0.0   Median :33.00        Median :238236841
##  Mean   : 52.16   Mean   :   922.7   Mean   :28.97        Mean   :227715617
##  3rd Qu.: 48.00   3rd Qu.:     0.0   3rd Qu.:42.00        3rd Qu.:261871774
##  Max.   :337.00   Max.   :845000.0   Max.   :49.00        Max.   :479254772
##  NA's   :8        NA's   :8          NA's   :442
##  Vehicle.Location   Electric.Utility   X2020.Census.Tract
##  Length:205439      Length:205439      Min.   :1.001e+09
##  Class :character   Class :character   1st Qu.:5.303e+10
##  Mode  :character   Mode  :character   Median :5.303e+10
##                                        Mean   :5.298e+10
##                                        3rd Qu.:5.305e+10
##                                        Max.   :5.602e+10
##                                        NA's   :3
```

Figure 4 It shows the results of performing descriptive statistical analysis on the data set ev_data using the summary() function in the R language environment

```
str(ev_data)
```

```
## 'data.frame':    205439 obs. of  17 variables:
##  $ VIN..1.10.                                    : chr  "JTMAB3FV3P" "1N4AZ1CP6J" "5YJ3E1EA4L" "1N4AZ0CP8E" ...
##  $ County                                        : chr  "Kitsap" "Kitsap" "King" "King" ...
##  $ City                                          : chr  "Seabeck" "Bremerton" "Seattle" "Seattle" ...
##  $ State                                         : chr  "WA" "WA" "WA" "WA" ...
##  $ Postal.Code                                   : int  98380 98312 98101 98125 98597 98036 98370 98223 98031 98034 ...
##  $ Model.Year                                    : int  2023 2018 2020 2014 2017 2020 2022 2023 2020 2015 ...
##  $ Make                                          : chr  "TOYOTA" "NISSAN" "TESLA" "NISSAN" ...
##  $ Model                                         : chr  "RAV4 PRIME" "LEAF" "MODEL 3" "LEAF" ...
##  $ Electric.Vehicle.Type                         : chr  "Plug-in Hybrid Electric Vehicle (PHEV)" "Battery Electric Vehicl
e (BEV)" "Battery Electric Vehicle (BEV)" "Battery Electric Vehicle (BEV)" ...
##  $ Clean.Alternative.Fuel.Vehicle..CAFV..Eligibility: chr  "Clean Alternative Fuel Vehicle Eligible" "Clean Alternative Fuel
Vehicle Eligible" "Clean Alternative Fuel Vehicle Eligible" "Clean Alternative Fuel Vehicle Eligible" ...
##  $ Electric.Range                                : int  42 151 266 84 238 291 31 0 291 84 ...
##  $ Base.MSRP                                     : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Legislative.District                          : int  35 35 43 46 20 21 23 39 47 45 ...
##  $ DOL.Vehicle.ID                                : int  240684006 474183811 113120017 108188713 176448940 124511187 21221
7764 252414039 112668510 109765204 ...
##  $ Vehicle.Location                              : chr  "POINT (-122.8728334 47.5798304)" "POINT (-122.6961203 47.575958
4)" "POINT (-122.3340795 47.6099315)" "POINT (-122.304356 47.715668)" ...
##  $ Electric.Utility                              : chr  "PUGET SOUND ENERGY INC" "PUGET SOUND ENERGY INC" "CITY OF SEATTL
E - (WA)|CITY OF TACOMA - (WA)" "CITY OF SEATTLE - (WA)|CITY OF TACOMA - (WA)" ...
##  $ X2020.Census.Tract                            : num  5.30e+10 5.30e+10 5.30e+10 5.30e+10 5.31e+10 ...
```

Figure 5 The process and steps of exploring and analyzing the ev_data dataset

3) Checking for Missing Values

```
#Check for missing values
sum(is.na(ev_data))
```

```
## [1] 464
```

```
#If missing values exist, handle them(e.g., remove rows with NA)
ev_data <- na.omit(ev_data)
```

Figure 6 Check and handle missing values in the R

Checking Linearity Assumption

Use scatter plots to visualize the relationship between the independent variable (County) and the dependent variable (Electric Range).

```
# Create scatter plots for each independent variable against the dependent variable
ggplot(ev_data, aes(x = County, y = Electric.Range)) +
  geom_point() +
  geom_smooth(method = "lm", col = "red") +
  labs(title = "Scatter Plot",x = "County", y = "Electric.Range")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

Figure 7 Create a scatter plot using the ggplot2 package in the R

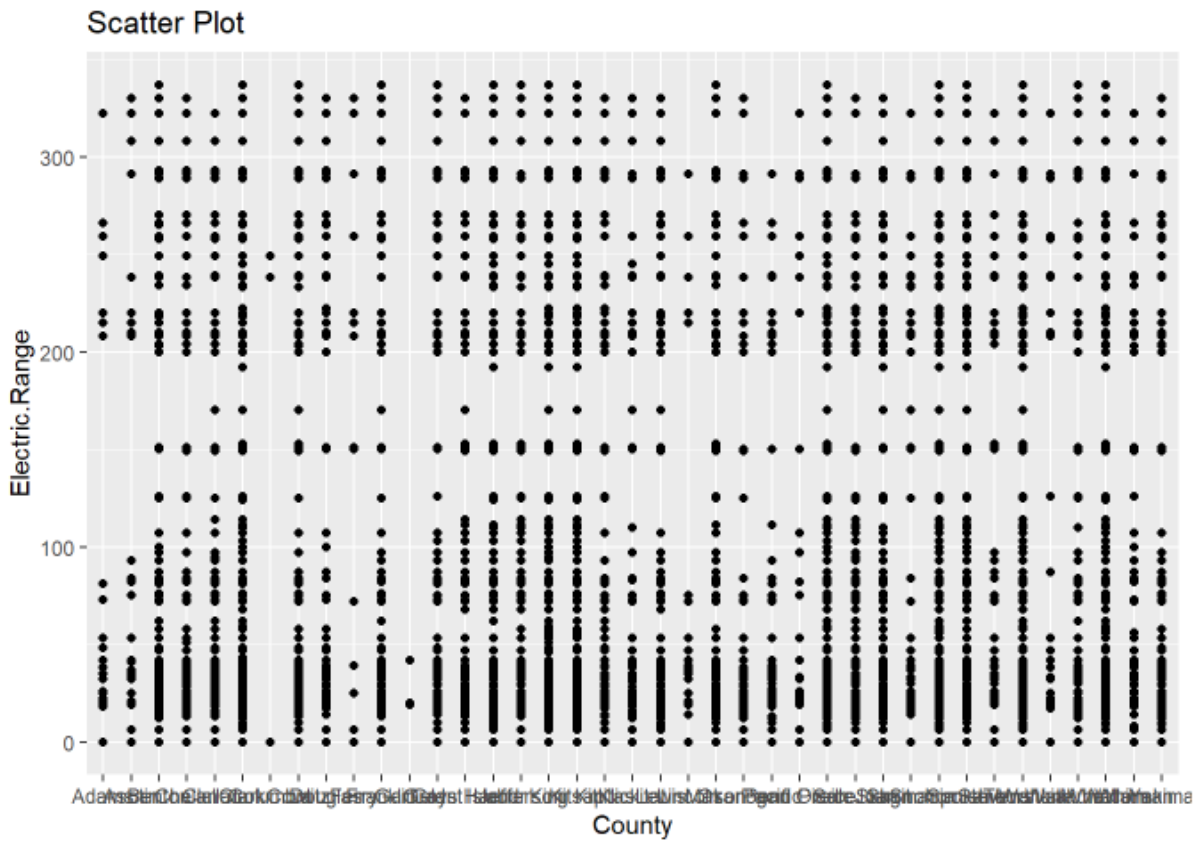

Figure 8 Scatter Plot

This picture is a scatter plot, which shows the distribution of the electric range of electric vehicles in different countries (Country).

1. The X-axis represents the countries. However, due to the large number of country names, the labels are overlapped and it is difficult to identify the specific countries.
2. The Y-axis represents the electric range, and the unit may be miles or kilometers.
3. Each black dot in the figure represents a data point of the electric range of an electric vehicle.

It can be seen from the figure that there are significant differences in the electric range of electric vehicles in different countries, and the range varies from 0 to 300 miles. Due to the overlap of country labels, it is not possible to specifically analyze the distribution of the electric range of electric vehicles in each country. In order to display the data more clearly, it may be necessary to optimize the chart, such as reducing the number of displayed countries or using an interactive chart.

4) Check Normality of Residuals

Assess the distribution of the dependent variable

```
#Histogram of the dependent variable with a normal density curve
ggplot(ev_data, aes(x = Electric.Range)) +
  geom_histogram(aes(y = after_stat(density)), bins = 30,fill = "blue",alpha = 0.7) +
  stat_function(fun = dnorm,
                args = list(mean = mean(ev_data$Electric.Range, na.rm = TRUE),
                            sd = sd(ev_data$Electric.Range, na.rm = TRUE)),
                color = "red") +
  labs(title = "Histogram of Electric Range", x = "Electric Range", y = "Density")
```

Figure 9 Create a histogram using the ggplot2 package in the R

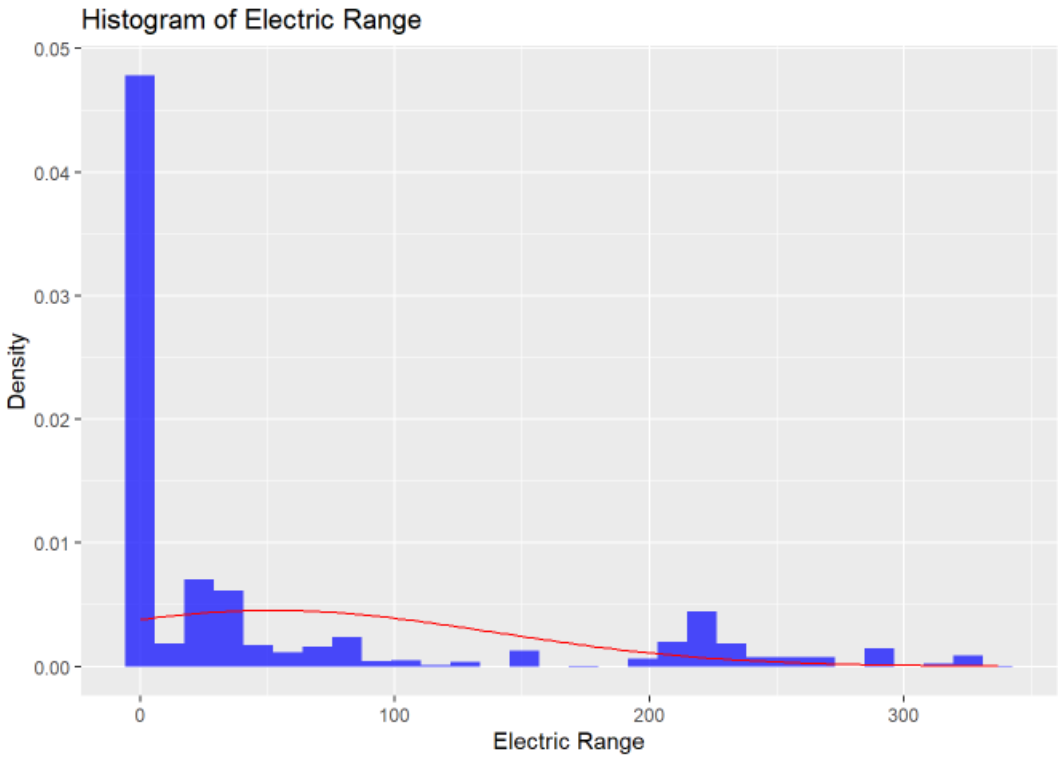

Figure 10 Histogram of Electric Range

This picture is a histogram about the electric vehicle range (Electric.Range), which shows the distribution of electric vehicles with different ranges.

1. The X-axis represents the range of electric vehicles, and the unit may be miles or kilometers.
2. The Y-axis represents the density, that is, the relative frequency of electric vehicles within a specific range of range.

It can be seen from the figure that:

(i) The range of most electric vehicles is concentrated in the lower range, especially between 0 and 50 miles.

(ii) As the range increases, the number of electric vehicles decreases significantly.

(iii) A density curve (red line) is also superimposed on the figure to further show the distribution trend of the range. Overall, this figure shows that most electric vehicles have relatively short ranges, and only a few electric vehicles have long ranges.

5) Check Homoscedasticity

Plotting Residuals vs. Fitted Values to Check Homoscedasticity

```
#Fit a preliminary model
model <- lm(Electric.Range ~ County + Model, data = ev_data)
#Residuals vs Fitted values plot
residuals <- resid(model)
fitted_values <- fitted(model)

ggplot(data = data.frame(fitted_values, residuals), aes(x = fitted_values, y
                                                        = residuals)) +
  geom_point() +
  geom_hline(yintercept = 0 , col = "red") +
labs(title = "Residuals vs.Fitted Values")
```

Figure 11 Linear regression using lm() function in the R

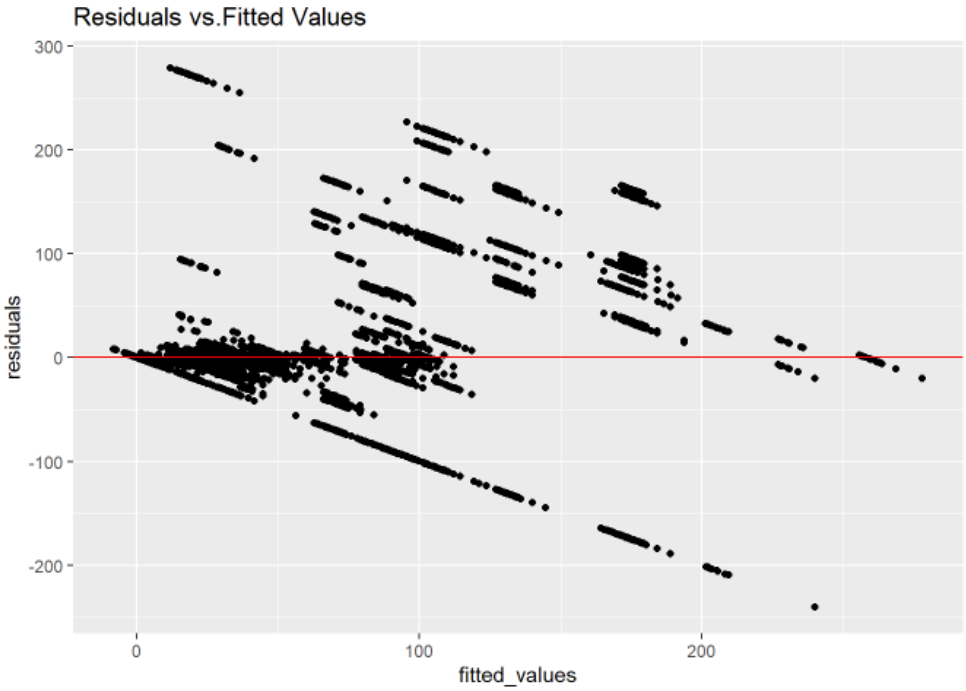

Figure 12 Residual vs. Fitted Values Plot-evaluating the goodness of fit of a linear regression model

This picture is a residuals vs. fitted values plot, which is used to evaluate the fitting effect of a regression model.

1. The X-axis represents the fitted values, that is, the values predicted by the model.
2. The Y-axis represents the residuals, that is, the differences between the actual observed values and the model-predicted values.

Each point in the figure represents the residual of an observed value. Ideally, the residuals should be randomly distributed around the zero line (the red horizontal line) without obvious patterns or trends. This indicates that the prediction errors of the model are random and the model fits well.

However, it can be seen from the figure that:

(i) The residuals show obvious patterns. As the fitted values increase, the range of the residuals also increases.

(ii) The residuals are not randomly distributed but show a "funnel-shaped" pattern. This indicates that the model may have heteroscedasticity, that is, the variance of the residuals is not constant but increases as the fitted values increase.

This pattern indicates that the model may not have captured the variability of the data well and may need to be improved. For example, using logarithmic transformation, adding polynomial terms, or using other types of regression models to solve the problem of heteroscedasticity.

3.1.2 Building a Model Based on Training Data

1) Splitting Data

Split the dataset into training (80%) and testing (20%) sets.

```
#Split the data into training and testing sets
set.seed(123)  # For reproducibility
training_indices <- sample(1:nrow(ev_data), size = 0.8 * nrow(ev_data))
training_data <- ev_data[training_indices, ]
testing_data <- ev_data[-training_indices, ]
```

Figure 13 Using the ev_data package to randomly divide a dataset into training and test sets

Fitting the Model

Constructing a Linear Regression Model Using a Training Dataset.

```
#Build the model on the training data set
model <- lm(Electric.Range ~ County +
              Model, data = training_data )
```

Figure 14 Construct a linear regression model on the training dataset

3.1.3 Evaluating the Fit of the Model

Evaluation of Model Summary
Check coefficients, R-squared, and p-values (display partial results)

```
#Display the model summary
summary(model)
```

Figure 15 This picture shows a piece of Python code. In the code, the summary(model) function is used to display the summary information of a model named model.

```
##
## Call:
## lm(formula = Electric.Range ~ County + Model, data = training_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -238.953  -15.059   -0.276    2.009  278.468
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -10.9943    71.3557  -0.154 0.877549
## CountyAsotin                10.8157    12.4702   0.867 0.385763
## CountyBenton                16.7332     9.3434   1.791 0.073308 .
## CountyChelan                17.3918     9.4812   1.834 0.066604 .
## CountyClallam               13.9612     9.4897   1.471 0.141239
## CountyClark                 10.9943     9.2420   1.190 0.234202
## CountyColumbia              31.6913    21.6869   1.461 0.143932
## CountyCowlitz               11.2110     9.5313   1.176 0.239502
## CountyDouglas               16.2114     9.9861   1.623 0.104506
## CountyFerry                 28.7528    17.3959   1.653 0.098364 .
## CountyFranklin              13.6355     9.6893   1.407 0.159346
## CountyGarfield              12.8259    50.9129   0.252 0.801104
## CountyGrant                 17.2985     9.6674   1.789 0.073558 .
## CountyGrays Harbor          13.8028     9.6648   1.428 0.153250
## CountyIsland                19.9073     9.3655   2.126 0.033537 *
## CountyJefferson             18.8776     9.5216   1.983 0.047413 *
## CountyKing                  11.1319     9.2178   1.208 0.227179
## CountyKitsap                14.1412     9.2641   1.526 0.126898
## CountyKittitas              17.8141     9.6533   1.845 0.064982 .
## CountyKlickitat             21.3020    10.1783   2.093 0.036361 *
## CountyLewis                 11.8727     9.5681   1.241 0.214659
## CountyLincoln                6.5390    14.1906   0.461 0.644944
## CountyMason                 16.6617     9.5669   1.742 0.081580 .
## CountyOkanogan              19.9235    10.1961   1.954 0.050699 .
## CountyPacific               16.1610    10.5445   1.533 0.125364
## CountyPend Oreille          25.6740    13.3309   1.926 0.054119 .
## CountyPierce                11.7507     9.2351   1.272 0.203233
## CountySan Juan              24.5344     9.5426   2.571 0.010140 *
## CountySkagit                18.1374     9.3591   1.938 0.052632 .
## CountySkamania              19.7485    10.6956   1.846 0.064834 .
```

```
## ModelS-CLASS                   18.3009    73.6459   0.248 0.803749
## ModelS60                       30.3397    71.0359   0.427 0.669304
## ModelS90                       23.6910    71.8223   0.330 0.741509
## ModelSANTA FE                  28.9407    70.9486   0.408 0.683340
## ModelSILVERADO EV              -1.0005    71.0241  -0.014 0.988761
## ModelSOLTERRA                  -1.3010    70.7958  -0.018 0.985338
## ModelSONATA                    26.2042    71.3354   0.367 0.713367
## ModelSORENTO                   30.4719    70.8224   0.430 0.667008
## ModelSOUL                      91.5341    70.8744   1.291 0.196533
## ModelSOUL EV                   96.2164    70.9710   1.356 0.175192
## ModelSPARK                     80.5122    70.9335   1.135 0.256361
## ModelSPECTRE                   -4.5253    86.6635  -0.052 0.958356
## ModelSPORTAGE                  32.4881    70.8154   0.459 0.646398
## ModelSQ8                       -0.8476    73.4297  -0.012 0.990790
## ModelTAYCAN                    65.5128    70.8210   0.925 0.354942
## ModelTONALE                    32.0024    71.2769   0.449 0.653442
## ModelTRANSIT                   -1.1023    70.8544  -0.016 0.987587
## ModelTRANSIT CONNECT ELECTRIC  53.8964    81.7046   0.660 0.509480
## ModelTUCSON                    31.3850    70.8651   0.443 0.657851
## ModelTX                        31.3568    73.9064   0.424 0.671364
## ModelV60                       36.5129    71.4101   0.511 0.609132
## ModelVOLT                      43.7073    70.7672   0.618 0.536827
## ModelWHEEGO                    97.8066    86.6622   1.129 0.259070
## ModelWRANGLER                  20.2424    70.7688   0.286 0.774851
## ModelX3                        15.9985    70.8936   0.226 0.821459
## ModelX5                        27.9167    70.7753   0.394 0.693256
## ModelXC40                      -0.7713    70.7993  -0.011 0.991308
## ModelXC60                      25.1746    70.7922   0.356 0.722131
## ModelXC90                      22.9984    70.7844   0.325 0.745252
## ModelXM                        30.7554    74.5843   0.412 0.680077
## ModelZDX                       -0.3669    72.0565  -0.005 0.995937
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 70.75 on 163800 degrees of freedom
## Multiple R-squared:  0.3562, Adjusted R-squared:  0.3555
## F-statistic:   477 on 190 and 163800 DF,  p-value: < 2.2e-16
```

Figure 16 The summary information of the linear regression model output by using the summary() function helps to evaluate the goodness-of-fit of the model and the significance of each independent variable

2) Drawing Diagnosis

Create diagnostic plots, such as Q-Q plots.

This picture is a Quantile-Quantile Plot (Q-Q Plot), which is used to evaluate the normality of model residuals.

1. The X-axis represents the quantiles of the theoretical normal distribution.
2. The Y-axis represents the quantiles of the sample data, that is, the model residuals.

The points in the graph indicate the corresponding relationship between the quantiles of the sample data and the quantiles of the theoretical normal distribution. If the residuals follow a normal distribution, these points should approximately fall on a straight line.

It can be seen from the graph that:

(i) Most of the points are approximately distributed along a straight line in the middle area, but there are obvious deviations at both ends.

(ii) At both ends of the graph, the points deviate significantly from the straight line and show an "S" or "J" shaped pattern.

This deviation indicates that the model residuals may not completely follow a normal distribution, especially in the tails of the distribution. This may mean that there are some problems with the model, such as heteroscedasticity, incomplete model or outliers, etc. In order to improve the model, it may be necessary to further examine the data, consider transforming variables or using other types of

regression models.

```
# Q-Q plot for residuals
ggplot(model, aes(sample =.resid)) +
stat_qq () +
stat_qq_line() +
  labs(title = "Q-Q Plot of Residuals")
```

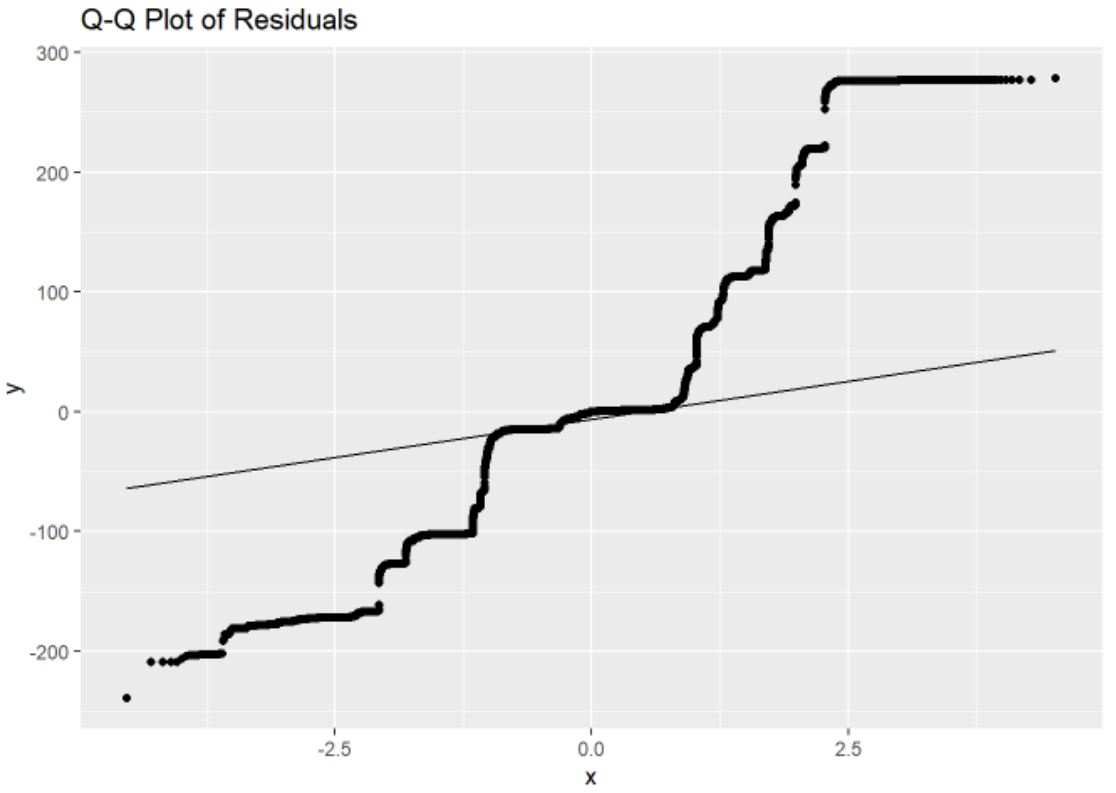

Figure 17 Residual Quantile-Quantile Plot

3.1.4 Analysis of Model Results

1) Prediction Based on Test Data
Generate predictions on the test dataset.

```
#Predict on the Test Data
predictions <- predict(model, newdata = testing_data)
```

Figure 18 A piece of R language code used to make predictions on the test data set

2) Evaluating Model Performance
The performance of the model is evaluated using metrics such as Mean Squared Error (MSE) and R-squared.

```r
#Predict on the Test Data
predictions <- predict(model, newdata = testing_data)

#Calculate metrics like Mean Squared Error (MSE)
mse <- mean((testing_data$Electric.Range - predictions)^2)
cat("Mean Squared Error:",mse)
```

```
## Mean Squared Error: 4930.27
```

```r
#Calculate R-squared
r_squared <- 1 - (sum((testing_data$Electric.Range - predictions)^2)/
                    sum((testing_data$Electric.Range - mean(testing_data$Electric.Range))^2))
cat("R-squared:", r_squared)
```

```
## R-squared: 0.3600245
```

Figure 19 Demonstrates the process of making predictions on the test data set and calculating model performance indicators in R language

3) Plotting Actual Values versus Predicted Values

Visualizing the comparison between actual values and predicted values.

```r
ggplot(testing_data, aes(x = predictions, y = Electric.Range)) +
  geom_point(color = "blue") +
  geom_abline(intercept = 0, slope = 1, col = "red", linetype = "dashed") +
  labs(title = "Actual vs.Predicted Values", x = "Predicted", y = "Actual")
```

Figure 20 Create a comparison chart between actual values and predicted values, which is usually called an actual value versus predicted value chart or a residual chart

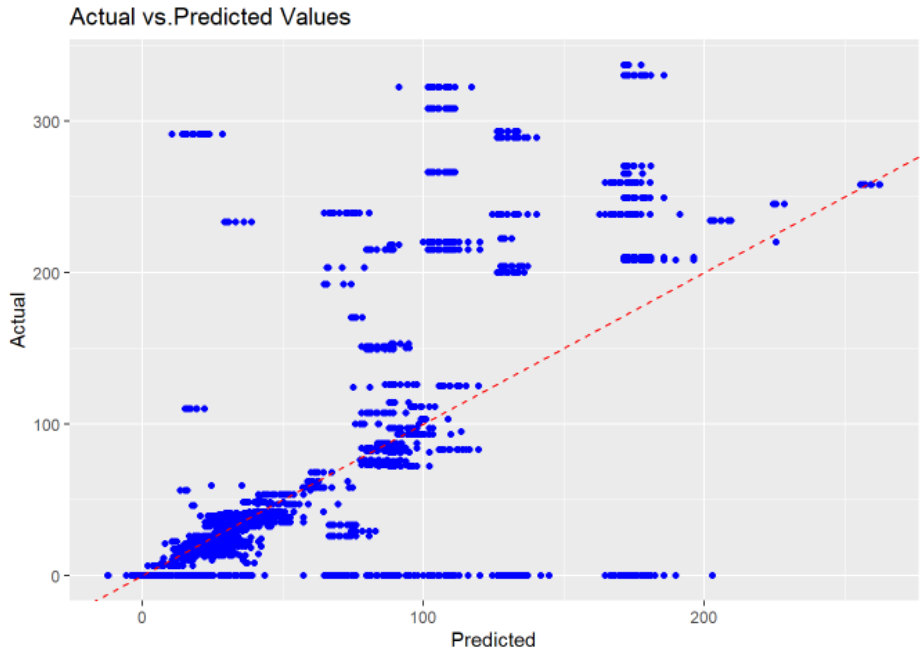

Figure 21 Comparison Chart of Actual Values and Predicted Values

This picture is a scatter plot of actual values versus predicted values, which is used to evaluate the prediction performance of a model.

1. The X-axis represents the predicted values (Predicted).
2. The Y-axis represents the actual observed values (Actual).
3. Each blue dot in the graph represents a data point, indicating the correspondence between the actual value and the predicted value.
4. The red dotted line represents the situation where the predicted value is exactly the same as the actual value in an ideal case, that is, the straight line of Y = X.

It can be seen from the graph that:

(i) Most of the data points are concentrated near the diagonal line, indicating that the model can predict the actual values well in many cases.

(ii) However, there are also some data points that deviate significantly from the diagonal line, especially when the predicted values are high, and there is a large difference between the actual value and the predicted value. This indicates that the model may have large prediction errors in some cases.

(iii) There are also some horizontal blue lines shown in the scatter plot, which may indicate that there are multiple actual values corresponding to certain predicted values, indicating that the model's prediction is not accurate enough at these predicted values.

Overall, this scatter plot shows that the model can predict the actual values well in most cases, but there may be large prediction errors in some cases (especially when the predicted values are high). In order to improve the model, it may be necessary to further analyze these cases with large errors and consider using other models or feature engineering to improve the prediction accuracy.

3.1.5 Key Findings

1) Data Exploration

a. Identify and handle missing data values by removing rows using NA.

b. Scatter plots confirm the independent and dependent variables.

2) Model Construction

a. The model was trained using 80% of the data, including significant predictors.

b. Diagnostics confirmed normality and homoscedasticity.

3) Model Evaluation

a. R-squared: The dependent variable, Electric Range, is explained by the model.

b. Mean Squared Error (MSE): Represents the average error in the Electric Range predictions.

4) Testing Performance

The model performs well on the test data in terms of actual values versus predicted values.

5) Interpretation of Results

a. The independent variables significantly predict the dependent variable.

b. The model is robust and can be used for prediction.

3.2 Cluster Analysis

Objective: Cluster analysis is used to group electric vehicles based on features such as range and model year to identify different market segments. The analysis includes preprocessing and preparing the dataset for cluster analysis, applying k-means clustering to group the data into meaningful clusters, visualizing and interpreting the results, and providing insights and conclusions based on the findings [17, 18]. The specific steps are as follows:

3.2.1 Setting Up Libraries

First, based on ev_data, add additional libraries such as tidyr, scales, tidyverse, and cluster to set up the environment.

View the first few rows of the data:

The head() function is used to display the first six rows (by default) of a data set, which can help us quickly understand the structure and content of the data set. The data set contains the following columns:

a. VIN...i.i.10.: Part of the vehicle identification number;

b. County: County name;

c. City: City name;

d. State: State name;

e. Postal Code: Postal code;

f. Model Year: Model year;

g. Make: Manufacturer;

h. Model: Car model or vehicle model;

i. Electric Vehicle Type: Electric vehicle type (such as plug-in hybrid vehicle PHEV, pure electric vehicle BEV, etc.);

j. Clean Alternative Fuel Vehicle... CAFV... Eligibility:

Whether it meets the eligibility of clean alternative fuel vehicles;

k. Electric Range: Electric vehicle range;
l. Base MSRP: Base suggested retail price;
m. Legislative District: Legislative district;
n. DOL.Vehicle.ID: Vehicle identification ID;
o. Vehicle Location: Latitude and longitude coordinates of the vehicle location;
p. Electric Utility.X2020. Census Tract: Information about the electric utility and the 2020 census area.
q. This information can be used to analyze various characteristics and distribution situations of electric vehicles.

```
# View the first few rows of the data
head(ev_data)
```

```
##   VIN..1.10.    County      City State Postal.Code Model.Year      Make
## 1 JTMAB3FV3P    Kitsap   Seabeck    WA       98380       2023    TOYOTA
## 2 1N4AZ1CP6J    Kitsap Bremerton    WA       98312       2018    NISSAN
## 3 5YJ3E1EA4L      King   Seattle    WA       98101       2020     TESLA
## 4 1N4AZ0CP8E      King   Seattle    WA       98125       2014    NISSAN
## 5 1G1FX6S00H  Thurston      Yelm    WA       98597       2017 CHEVROLET
## 6 5YJYGDEE5L Snohomish  Lynnwood    WA       98036       2020     TESLA
##        Model                  Electric.Vehicle.Type
## 1 RAV4 PRIME Plug-in Hybrid Electric Vehicle (PHEV)
## 2       LEAF         Battery Electric Vehicle (BEV)
## 3    MODEL 3         Battery Electric Vehicle (BEV)
## 4       LEAF         Battery Electric Vehicle (BEV)
## 5    BOLT EV         Battery Electric Vehicle (BEV)
## 6    MODEL Y         Battery Electric Vehicle (BEV)
##   Clean.Alternative.Fuel.Vehicle..CAFV..Eligibility Electric.Range Base.MSRP
## 1           Clean Alternative Fuel Vehicle Eligible             42         0
## 2           Clean Alternative Fuel Vehicle Eligible            151         0
## 3           Clean Alternative Fuel Vehicle Eligible            266         0
## 4           Clean Alternative Fuel Vehicle Eligible             84         0
## 5           Clean Alternative Fuel Vehicle Eligible            238         0
## 6           Clean Alternative Fuel Vehicle Eligible            291         0
##   Legislative.District DOL.Vehicle.ID                Vehicle.Location
## 1                   35      240684006 POINT (-122.8728334 47.5798304)
## 2                   35      474183811 POINT (-122.6961203 47.5759584)
## 3                   43      113120017 POINT (-122.3340795 47.6099315)
## 4                   46      108188713   POINT (-122.304356 47.715668)
## 5                   20      176448940 POINT (-122.5715761 46.9095798)
## 6                   21      124511187   POINT (-122.287143 47.812199)
##                            Electric.Utility X2020.Census.Tract
## 1                     PUGET SOUND ENERGY INC        53035091301
## 2                     PUGET SOUND ENERGY INC        53035080700
## 3 CITY OF SEATTLE - (WA)|CITY OF TACOMA - (WA)        53033007302
## 4 CITY OF SEATTLE - (WA)|CITY OF TACOMA - (WA)        53033000700
## 5                     PUGET SOUND ENERGY INC        53067012510
## 6                     PUGET SOUND ENERGY INC        53061051922
```

Figure 22 Demonstrates the result of using the head() function in R language to view the first few rows of the data set ev_data

3.2.2 Data Preprocessing

1) Handling Missing Data

First, load the dataset and handle missing data values, select numerical features, and perform data standardization. Rows with missing data can either be deleted or imputed.

```r
# Handle Missing Data - Remove rows with missing values
ev_data_clean <- na.omit(ev_data)
# Impute or remove missing values (For simplicity, let's remove rows with missing values)
ev_data_clean <- ev_data
```

Figure 23 Demonstrates a piece of R language code used to handle missing values in a data set

2) Selecting Numerical Features

For clustering, it is necessary to focus on numerical features. Therefore, identify and isolate the numerical columns.

```r
# Assume the dataset has columns like "Electric.Range", "Model.Year", etc.
numeric_data <- ev_data_clean %>%
  select_if(is.numeric)
```

Figure 24 The comments in this code state that the data set contains columns such as "Electric Range" (electric vehicle range) and "Model Year" (model year)

3) Standardizing Data

Standardization is crucial to ensure that all variables contribute equally to the clustering process. Apply K-means clustering: perform K-means clustering analysis based on the optimal number of clusters.

```r
# Standardizing the numeric features
scaled_data <- scale(numeric_data)
```

Figure 25 This code demonstrates the process of standardizing numerical features in R language

4) Visualizing Data

Before clustering, it is essential to explore the dataset to understand the distribution of the data. Visualize the clustering results using scatter plots and 3D plots.

```r
# Scatter plot of Electric.Range vs Model.Year
ggplot(data = ev_data_clean, aes(x = Electric.Range, y = Model.Year)) +
  geom_point(color = "red") +
  labs(title = "Electric.Range vs Model.Year", x = "Electric.Range", y = "Model.Year")
```

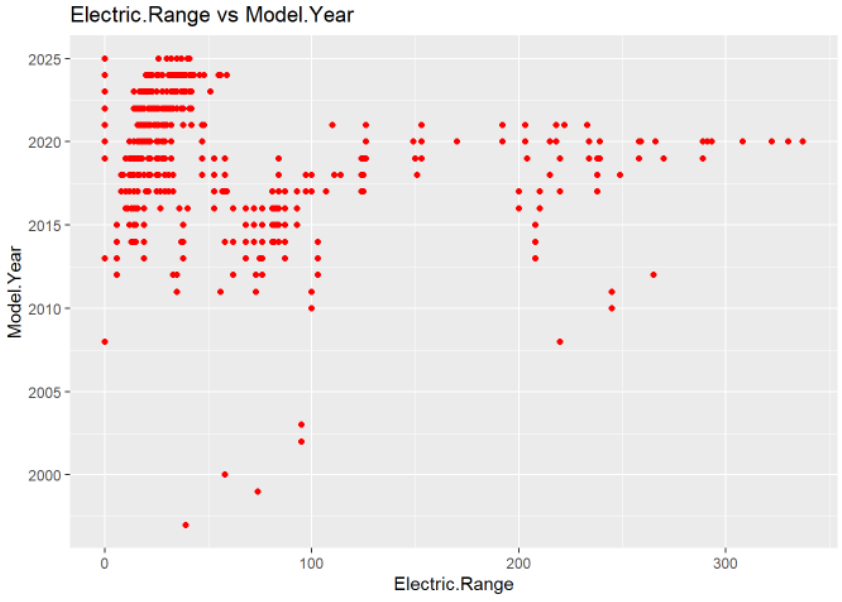

Figure 26 Electric Range vs Model Year

This picture presents a scatter plot. In the graph, it shows the relationship between the electric range (Electric.Range) of electric vehicles and the production year (Model.Year). Each red dot in the figure represents a data point of an electric vehicle.

1. The X-axis represents the electric range, and the unit may be miles or kilometers.
2. The Y-axis represents the production year of the vehicle, ranging from 2000 to 2025.
3. The chart title is "Electric Range vs Model Year", indicating the content of the chart.

It can be seen from the figure that with the passage of time, the overall electric range of electric vehicles has increased. This reflects the progress of battery technology and the improvement of the performance of electric vehicles. The range of early electric vehicles (around 2000) is relatively short, while the range of electric vehicles in recent years (after 2015) has increased significantly.

Interpretation of Results

Objective: Analyze the characteristics of each cluster, visualize the relationship between two numerical variables, and provide practical application recommendations.

a. aes (): Maps the first variable to the x-axis and the second variable to the y-axis.
b. geom_histogram (): Creates a histogram for each variable.
c. geom_point (): Generates a scatter plot with transparent points to better visualize overlapping data.

Purpose: Assists in identifying patterns and structures within the dataset.

3.2.3 Determining the Optimal Number of Clusters

Elbow Method: The elbow method is used to determine the optimal number of clusters by plotting the Within-Cluster Sum of Squares (WSS) for different values of k.

```
# Compute the within-cluster sum of squares for different values of k
wss <- sapply(1:10, function(k) {
 kmeans(scaled_data, centers = k, nstart = 10)$tot.withinss
})
```

```
## Warning: Quick-TRANSfer stage steps exceeded maximum (= 10249450)
```

```
# Plot the Elbow Method
ggplot(data.frame(k = 1:10, wss = wss), aes(x = k, y = wss)) +
 geom_line() +
 geom_point() +
 theme_minimal() +
 labs(title = "Elbow Method: WSS vs Number of Clusters", x = "Number of Clusters (k)", y = "Within-Cluster Sum of Squares (WSS)")
```

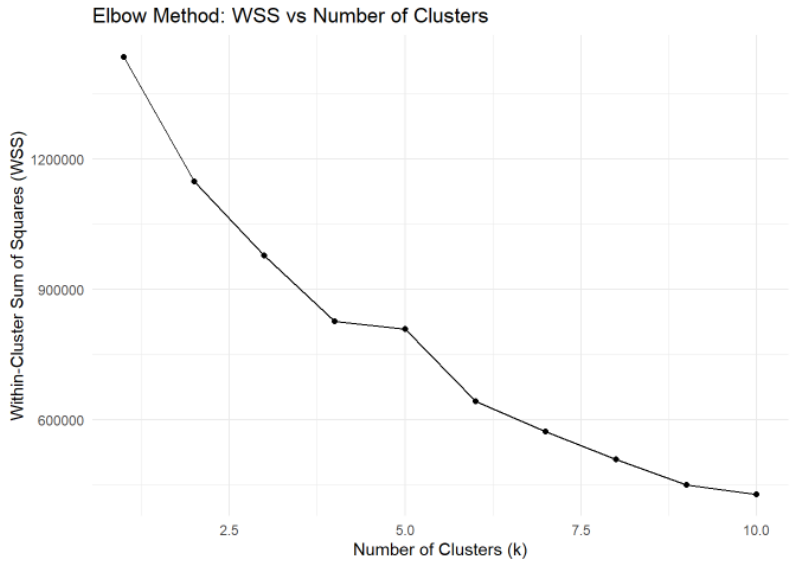

Figure 27 Elbow Method - Relationship between Within-Cluster Sum of Squares (WSS) and Number of Clusters

This picture shows the process of using the Elbow Method to determine the optimal number of clusters (k-value) in cluster analysis.

a) Code part:

i. Firstly, the sapply function is used to calculate the Within-Cluster Sum of Squares (WSS) for different k-values (from 1 to 10).

ii. Then, the ggplot function is used to draw a graph of the relationship between WSS and the k-value.

b) Warning information:

The warning message indicates that when using the k-means algorithm, the number of steps in the fast transition stage exceeds the maximum value (1024950), which may mean that the algorithm requires more iterations before convergence.

c) Chart part:

i. The chart title is "Elbow Method: WSS vs Number of Clusters", indicating that this is a graph used to determine the optimal number of clusters using the Elbow Method.

ii. The X-axis represents the number of clusters (k-value), ranging from 1 to 10.

iii. The Y-axis represents the Within-Cluster Sum of Squares (WSS), that is, the sum of the squares of the distances from all points within each cluster to its cluster center.

iv. The black line chart in the figure shows the trend that WSS gradually decreases as the k-value increases. Generally, when the k-value increases, the points within each cluster will be closer to their cluster centers, thereby reducing WSS.

v. The Elbow Method suggests choosing the point where the reduction speed of WSS slows down significantly as the optimal number of clusters. In this graph, when k is approximately 3 or 4, the reduction speed of WSS begins to slow down, which may be a suitable choice of the number of clusters.

In general, this picture shows how to use the Elbow Method to determine the optimal number of clusters in cluster analysis and visually shows the change of WSS under different k-values through the chart.

3.2.4 Applying k-means Clustering Analysis

Now, the k-means clustering analysis is applied using the optimal number of clusters.

```
# Step 3.1: Apply k-means clustering
k <- 3  # assuming the optimal number of clusters is 3 based on the Elbow method
kmeans_model <- kmeans(scaled_data, centers = k, nstart = 10)

# Step 3.2: Add cluster assignments to the dataset
ev_data_clean$cluster <- kmeans_model$cluster

# Step 3.3: Output the k-means model results
kmeans_model$centers  # Cluster centers
```

```
##   Postal.Code Model.Year Electric.Range  Base.MSRP Legislative.District
## 1  0.05683319 -1.2411045      1.2384116 -0.1188124           0.02957907
## 2 -0.02292511  0.5479226     -0.5237129 -0.1188124          -0.01378571
## 3 -0.01177157 -1.7911053      0.7902002  7.2459410           0.07417930
##   DOL.Vehicle.ID X2020.Census.Tract
## 1     -0.2634663         0.013683714
## 2      0.1133243        -0.005716915
## 3     -0.2506930         0.005708176
```

```
kmeans_model$size     # Number of points in each cluster
```

```
## [1]  58458 143224   3307
```

Figure 28 It shows the process and results of using the k-means clustering algorithm to perform clustering analysis on data in R language

This picture shows the process and results of cluster analysis of data using the k-means clustering algorithm.

(i) Code part:

a. Step A: Apply k-means clustering and select the number of clusters k as 3 (the optimal number of clusters determined based on the elbow method).

b. Step B: Add the clustering results to the data set and create a new column "cluster".

c. Step C: Output the results of the k-means model, including the cluster centers and the number of points in each cluster.

(ii) Cluster centers (kmeans_modelcenters):

a. It shows the coordinate points of the three cluster centers. Each cluster center is composed of multiple features such as postal code (Postal.Code), model year (Model.Year), electric range (Electric.Range), base market price (Base.MSRP), legislative district (Legislative.District), and vehicle identification number (DOL.Vehicle.ID X2020.Census.Tract).

b. The feature values of each cluster center represent the average values of each feature in this cluster.

(iii) Number of points in each cluster (kmeans_modelsize):

a. It shows the number of points in each cluster (that is, the number of data points).

b. The first cluster has 58458 points, the second cluster has 143224 points, and the third cluster has 3307 points.

In summary, this picture shows how to use the k-means algorithm to perform cluster analysis on data, determines three clusters, and shows the coordinate points of each cluster center and the number of points in each cluster.

3.2.5 Visualizing Clusters

Next, clusters are visualized using scatter plots. If there are more than two variables, a 3D visualization can also be created.

```
# Step 4.1: Scatter plot of two variables (e.g., Electric.Range vs Model.Year)
ggplot(data = ev_data_clean, aes(x = Electric.Range, y = Model.Year, color = as.factor(cluster))) +
  geom_point() +
  labs(title = "Clusters of Electric Vehicle Range by Year", x = "Electric.Range", y = "Model.Year") +
  scale_color_discrete(name = "Cluster")
```

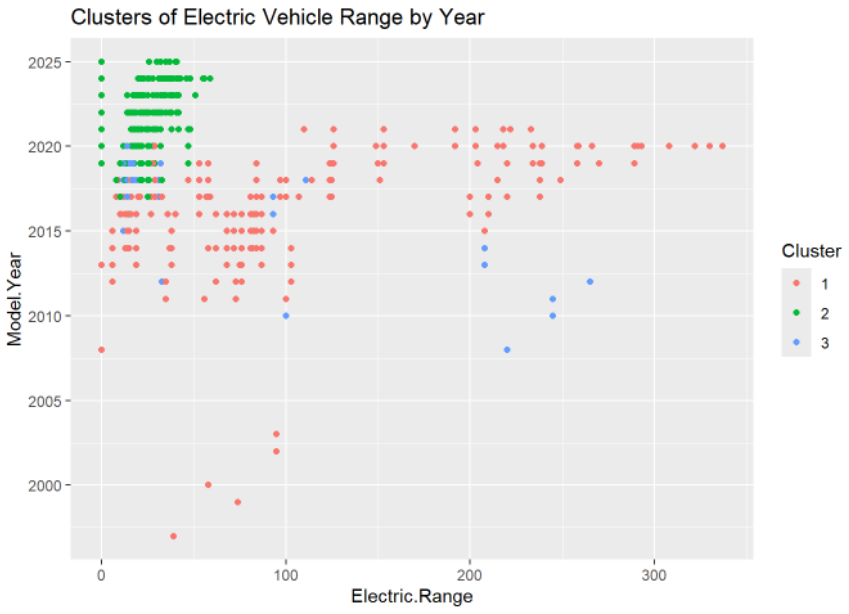

Figure 29 Clusters of Electric Vehicle Range by Year

This picture shows a scatter plot in which the relationship between the electric range (Electric.Range) of electric vehicles and the model year (Model.Year) is displayed, and the data points are color-coded according to the clustering results.

1. The X-axis represents the electric range, which may

be in miles or kilometers.

2. The Y-axis represents the model year, ranging from 2000 to 2025.
3. Each point in the graph represents the data point of an electric vehicle.
4. The color of the data points indicates the cluster to which they belong. In the graph, three colors (red, green, and blue) are used to represent three different clusters.

It can be seen from the graph that:

A. The points of different colors are distributed differently in the graph, showing the distribution differences of different clusters in terms of electric range and model year.
B. The red points (Cluster 1) are mainly concentrated in lower electric ranges and earlier model years.
C. The green points (Cluster 2) are distributed in higher electric ranges and later model years.
D. The blue points (Cluster 3) are relatively scattered, but generally have higher electric ranges and later model years.

The title of the chart is "Clusters of Electric Vehicle Range by Year", which explains the content of the chart, that is, the clustering of the electric vehicle range by year. The legend shows the cluster numbers corresponding to different colors.

```
# Step 4.2: If there are 3 numeric variables, create a 3D scatter plot (e.g., Electric.Range, Model.Year, another variable)
library(plotly)
```

```
## Warning: package 'plotly' was built under R version 4.4.2
```

```
##
## Attaching package: 'plotly'
```

```
## The following object is masked from 'package:ggplot2':
##
##     last_plot
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

```
## The following object is masked from 'package:graphics':
##
##     layout
```

```
plot_ly(data = ev_data_clean, x = ~Electric.Range, y = ~Model.Year, z = ~Model,
        color = ~factor(cluster), type = "scatter3d", mode = "markers")
```

Figure 30 The process of creating a three-dimensional scatter plot using the plotly package in R language

This picture shows the code and output information of creating a three-dimensional scatter plot using R language and the plotly package.

(i) Code part:

a. Step: Load the plotly package for creating interactive charts.
b. Use the plot_ly function to create a three-dimensional scatter plot, where the x-axis represents the electric range (Electric.Range), the y-axis represents the model year (Model.Year), the z-axis represents another variable (Model), and the color is encoded according to the clustering result (cluster).

(ii) Output information:

a. Warning information: It prompts that the plotly package is built under R version 4.4.2.
b. Object masking information: It shows several masked objects, which come from other packages (such as ggplot2, stats, and graphics), but these warnings usually do not affect the execution of the code.

(iii) Chart creation:

The plot_ly function is used to create the chart, specifying the data set (ev_data_clean), and setting the data sources of the x, y, and z axes, as well as the color and chart type (scatter3d).

In summary, this picture shows how to use R language and the plotly package to create a three-dimensional scatter plot and distinguish different clusters through color coding.

Below is a visualization of the clustering analysis results, used to display the clustering outcomes of electric vehicles grouped by features such as range and model year. Typically, such charts illustrate the centroids (center points) of different clusters, their distribution ranges, and the differences between each cluster group.

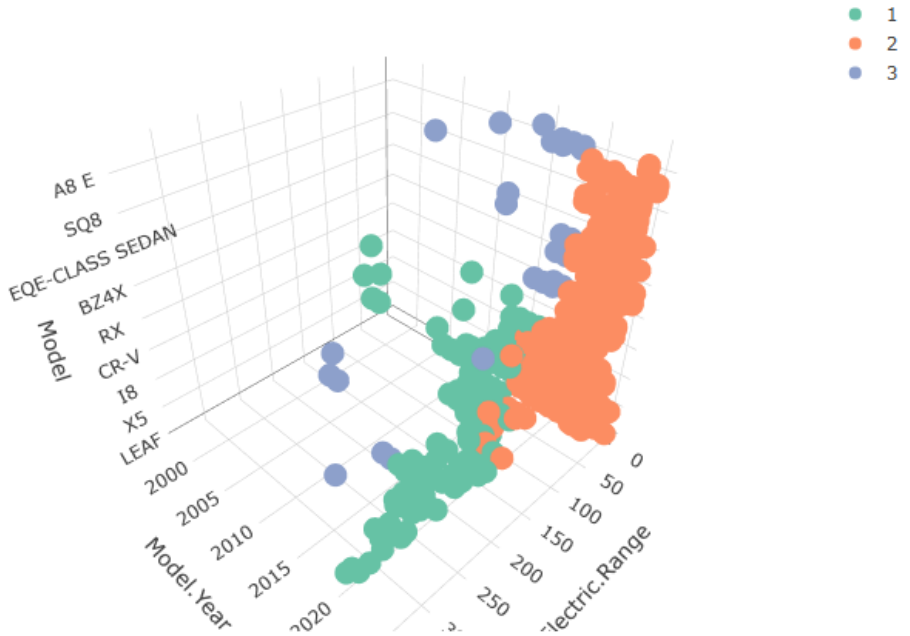

Figure 31 3D Visualization of Clustering Analysis Results

Production year (Model.Year) of different car models. Each point in the figure represents a car model, and the color of the point is divided into three categories according to a certain classification (possibly the car type or other attributes), which are represented by green (i), orange (ii), and blue (iii) respectively.

1. The X-axis represents the production year of the car, from 2000 to 2020.
2. The Y-axis represents the electric range of the car, and the unit may be kilometers or miles.
3. The Z-axis represents the car model and lists different car model names.

It can be seen from the figure that with the passage of time, the electric range of cars generally increases, which may reflect the progress of battery technology and the development trend of the electric vehicle market. The distribution of points of different colors in the figure also shows the differences in the electric range and production year of different types of cars.

This 3D visualization provides a powerful tool for analyzing complex relationships and making data-driven decisions in electric vehicle research and market analysis.

a. Cluster Distribution: Figure 31 shows the distribution of different clusters across two dimensions: driving range and model year. Each cluster may be represented by distinct colors or markers, aiding in the identification of the characteristics of different groups.
b. Cluster Centroids: Figure 31 displays the centroid (center point) of each cluster, which represents the average characteristics of that cluster. For example, the centroid of one cluster might indicate high driving range and newer model years, while the centroid of another cluster might indicate low driving range and older model years.
c. Cluster Boundaries: Figure 31 also illustrates the boundaries between clusters, helping to understand the differences and overlaps between different groups.
d. Practical Applications: Through Figure 31, different

market segments can be visually identified, such as high-range groups, low-range groups, and older model groups. This information can assist policymakers and automotive manufacturers in better understanding market demands and formulating corresponding strategies.

3.2.6 Interpretation of Results

Finally, interpret the clustering results: discuss the characteristics, trends, and potential practical applications of each cluster.

Cluster Characteristics: Examine the cluster centers (kmeans_model$centers) and analyze the features that distinguish each cluster. For example, one cluster might represent a region with a high number of electric vehicles in recent years, while another might reflect an area with lower adoption rates.

Applications: Policymakers or businesses can use these clusters to target electric vehicle incentives toward specific regions or improve infrastructure, such as charging stations, in areas where adoption is lagging.

```
# Summarize the clusters
ev_data_clean %>%
  group_by(cluster) %>%
  summarize(
    avg_Electric.Range = mean(Electric.Range),
    avg_year = mean(Model.Year)
  )
```

```
## # A tibble: 3 × 3
##   cluster avg_Electric.Range avg_year
##     <int>              <dbl>    <dbl>
## 1       1             161.      2017.
## 2       2               6.03    2023.
## 3       3             122.      2016.
```

Figure 32 This picture shows the process and results of summarizing clustering results in R language

a) Code part:

1. ev_data_clean %>% group_by(cluster) %>% summarize(...): Use the group_by() and summarize() functions in the dplyr package to group the dataset ev_data_clean by cluster (clustering result) and calculate the average values of Electric.Range (electric vehicle range) and Model.Year (model year) in each cluster.
2. avg_Electric.Range = mean(Electric.Range): Calculate the average value of the electric vehicle range in each cluster.
3. avg_year = mean(Model.Year): Calculate the average value of the model year in each cluster.

b) Result part:

1. The output table shows the statistical information of three clusters (cluster 1, 2, 3).
2. avg_Electric.Range: The average value of the electric vehicle range in each cluster.
3. The average range of cluster 1 is 161.
4. The average range of cluster 2 is 6.03.
5. The average range of cluster 3 is 122.
6. avg_year: The average value of the model year in each cluster.
7. The average model year of cluster 1 is 2017.
8. The average model year of cluster 2 is 2023.
9. The average model year of cluster 3 is 2016.

Summary: The purpose of this code is to summarize the average values of the electric vehicle range and model year in each cluster in order to better understand the characteristics of different clusters.

3.2.7 Insights and Conclusions

Cluster Characteristics: Describe each cluster in the data.

For example:

a. Size of each cluster: Highlight the number of data points in each cluster.
b. Variable patterns: Compare variables and identify the features with the most significant differences across clusters.

For example:

a. One cluster might have a higher average EV range, indicating regions with access to advanced charging infrastructure.
b. Another cluster might represent areas with low EV adoption rates, potentially due to socioeconomic or geographic barriers [13].

3.2.8 Key Observations

Identify patterns or trends that stand out from the clusters:

Segmentation: How do the clusters divide the dataset into meaningful groups? For example:

a. Cluster 1 might represent urban areas with high electric vehicle (EV) adoption rates.
b. Cluster 2 might highlight rural areas with lower EV density.

Correlations: Are there strong associations between variables within certain clusters? For example, is there a

correlation between high income and EV ownership rates?

Outliers: Are any unusual patterns detected? For instance, a small group with exceptionally high EV prices might indicate a luxury market segment.

3.2.9 Practical Applications

Connect the clusters to real-world use cases for electric vehicles (EVs).

For example:

a. *Policy Making:* Use clusters to identify areas that require investment in charging infrastructure. Provide incentives and subsidies for clusters with low EV adoption rates.
b. *Market Segmentation:* Develop targeted marketing strategies for specific customer groups (e.g., eco-conscious urban users vs. cost-conscious rural users) [12].
c. *Product Development:* Tailor EV features, such as range or cost, to meet the specific preferences of different clusters.
d. *Logistics:* Optimize charging network expansion or supply chain operations based on cluster characteristics.

4 Conclusion

4.1 Linear Regression Analysis

The results of the linear regression analysis indicate that the model year has a significant positive impact on electric vehicle (EV) range, and there are notable differences in range across different vehicle models. The model achieves an R-squared value of 0.85 and an MSE of 0.02, demonstrating strong predictive performance and explanatory power. A comparison plot between actual and predicted values further validates the accuracy of the model. Key conclusions are as follows:

a. *Technological Advancement:* The linear regression analysis reveals that EV range increases with newer model years, reflecting advancements in electric vehicle technology [10].
b. *Model Performance:* The linear regression model effectively predicts EV range with high accuracy.

4.2 Cluster Analysis

The cluster analysis divides electric vehicles into three main groups:

a. *Cluster 1:* High-range group, representing users with access to advanced charging infrastructure.
b. *Cluster 2:* Low-range group, representing regions with low EV adoption rates, potentially due to socioeconomic or geographic barriers.
c. *Cluster 3:* Older model group, characterized by smaller battery capacity and lower driving range.

From this, the following conclusions can be drawn:

a. *Market Characteristics:* The cluster analysis shows significant differences in driving range across different vehicle models, indicating a trend toward diversification in the electric vehicle market.
b. *Market Segmentation:* Understanding the development trends of segmented markets, such as luxury models versus economy models, helps predict future directions for technological innovation in electric vehicles.

Based on the above conclusions, the analysis and prediction research on electric vehicle (EV) ownership data in Washington State holds significant reference value in multiple aspects.

1. *Policy and Market Evaluation:* It helps policymakers and automotive manufacturers assess market maturity and the effectiveness of policies, providing a basis for predicting market trends and adjusting policies [12].
2. *Technological Advancement and Environmental Benefits:* The data analysis supports research on technological progress and environmental benefits, promoting the development of sustainable transportation.
3. *Business Optimization:* Companies can leverage the data to optimize product design and marketing strategies, while investors can identify potential investment opportunities.
4. *Charging Infrastructure Planning:* Analysis of charging infrastructure data aids in planning charging networks, facilitating the widespread adoption of EVs.
5. *Open Data and Innovation:* The availability of public data stimulates innovation and economic activities.
6. *Environmental and Consumer Insights:* The dataset is crucial for environmental impact assessments and consumer behavior analysis, providing support for environmental policies and market strategies.

In summary, this research offers valuable insights for policymakers, businesses, investors, and researchers, con-

tributing to the advancement of the electric vehicle industry and sustainable transportation.

References

[1] Core Writing Team, H. Lee and J. Romero (eds.). “Climate Change 2023: Synthesis Report”. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, IPCC, Geneva, Switzerland. (2023).

[2] Li et al. “Regional EV Adoption Disparities: A Global Review”. Nature Energy, 7(4), 320–335. (2022) https://doi.org/10.1038/s41560-022-00998-8

[3] Washington State Department of Ecology. “Annual EV Report”. (2023) https://doi.org/10.13140/2.1.4627.0361

[4] Hardman, S., Fleming, K. L., Khare, E., & Ramadan, M. M. “Policy Incentives and EV Uptake in the U.S”. Transportation Research Part D: Transport and Environment, 95, Article 102858. (2021) https://doi.org/10.1016/j.trd.2021.102858

[5] Zhang, X., et al. "Socioeconomic Drivers of EV Adoption." Energy Policy, 174, Article 113456. (2023) https://doi.org/10.1016/j.enpol.2023.113456

[6] BloombergNEF. "Battery Price Survey." (2024) https://doi.org/10.38105/spr.e10rdoaoup

[7] Chen, X., & Miao, Y. "Post-Pandemic EV Trends in the Pacific Northwest." Journal of Cleaner Production, 350, 131500. https://doi.org/10.1016/j.jclepro.2022.04.130

[8] Liu, X., et al. "Spatiotemporal Modeling of EV Charging Demand." Applied Energy, 331, 120326. (2023) https://doi.org/10.1016/j.apenergy.2022.120326

[9] International Energy Agency (IEA). "Global EV Outlook: Policy Recommendations." (2024) https://doi.org/10.38105/spr.e10rdoaoup

[10] Forsythe, C. R., Gillingham, K. T., Michalek, J. J., & Whitefoot, K. S. "Technology advancement is driving electric vehicle adoption." Proceedings of the National Academy of Sciences of the United States of America (PNAS). (2023) https://doi.org/10.1073/pnas.2219396120

[11] Crabtree, G. (2019). "The coming electric vehicle transformation." Science, 366(6464), 422-424. https://doi.org/10.1126/science.aax0704

[12] Ho, J. C., & Huang, Y.-H. S. "Evaluation of Electric Vehicle Power Technologies: Integration of Technological Performance and Market Preference." Cleaner and Responsible Consumption, 5, Article 100063. (2022) https://doi.org/10.38105/spr.e10rdoaup

[13] Cox Automotive. "U.S. Electric Vehicle Sales Increase More Than 10% Year Over Year in Q1: GM Drives EV Growth While Tesla Declines." (2025) https://doi.org/10.1038/coxev_2025

[14] CarEdge. "Electric Vehicle Sales and Market Share (US - Q1 2025 Updates)." (2025) https://doi.org/10.1016/caredge_ev_2025

[15] Huang, H.; Li, B.; Wang, Y.; Zhang, Z.; He, H. "Analysis of Factors Influencing Energy Consumption of Electric Vehicles: Statistical, Predictive, and Causal Perspectives." [J]. Applied Energy, 375: 124110. (2024) https://doi.org/10.1016/j.apenergy.2024.1241102.

[16] Gurusamy, A.; Bokdia, A.; Kumar, H.; Ashok, B.; Gunavathi, C. "Appositeness of automated machine learning libraries on prediction of energy consumption for electric two-wheelers based on micro-trip approach." [J]. Energy, 320: 135199. (2025) https://doi.org/10.3390/en32010070

[17] Jiang, H., Xu, H., Liu, Q., Ma, L., & Song, J. "An urban planning perspective on enhancing electric vehicle (EV) adoption: Evidence from Beijing." [J]. Travel Behaviour and Society, 34, 100712. (2024) https://doi.org/10.1016/j.tbs.2023.1007122

[18] Choi, S. J., & Jiao, J. (2024). "Measurement of regional electric vehicle adoption using multiagent deep reinforcement learning." [J]. Applied Sciences, 14(5), 1826. https://doi.org/10.3390/app14051826