Assignment_09_PothineniKalyan

PothineniKalyan

2023-06-02

Introduction

As electric vehicles become more predominant, it is essential to understand the characteristics and trends of the electric vehicle market. The sample data provides valuable insights into consumer/customer behavior, electric vehicle adoption rates, and potential policy interventions in Washington state.

We will perform exploratory data analysis, cleansing, and modeling techniques to uncover patterns and trends in the electric vehicle market. It involves identifying car models and makes of popular electric vehicle types. Also, examine the relationship between electric range and consumer demand. Machine learning algorithms can be applied to predict future electric vehicle adoption and help policymaker's interventions to promote electric vehicle use in Washington state.

For example, policymakers, electric vehicle manufacturers, and energy companies would be interested in this research as stakeholders. Electric vehicle manufacturers can use this information to design and market electric vehicles that meet the needs of consumers.

Analysis of electric vehicle data is a significant data science problem with practical implications for various stakeholders. It requires the application of various data science techniques to uncover insights and trends in the market (electric vehicle market)

Research questions

- How does the distribution of electric vehicles vary across different regions in Washington state, such as by county, city, or zip code?
- What factors influence electric vehicle adoption in Washington state, such as incentives, availability of charging infrastructure, or demographic characteristics?
- What is Washington State's current electric vehicle adoption rate, and how has it changed?
- How does the availability and accessibility of electric vehicle charging infrastructure differ across different regions in Washington state
- How does the adoption rate and usage pattern of electric vehicles in Washington state compare to other states with similar demographic and economic characteristics?
- How do Washington State's electric vehicle policies and programs compare to other states?
- What is the potential impact of increasing the electric vehicle adoption rate in Washington state on reducing greenhouse gas emissions and improving air quality?

Approach

Plan to use various data science techniques to understand the characteristics and trends of the electric vehicle market. Specifically, we will first clean data, identify outliers and missing values, and create plots/visualizations to understand the distribution of different variables. Preprocess the data to ensure its quality and consistency. Then, we will analyze exploratory data to uncover patterns and trends related to adoption rates and consumer behavior in the electric vehicle market. Analysis can involve identifying which models and makes are most popular, understanding the distribution of electric vehicle types across different regions, and examining the relationship.

Next, we will use machine learning algorithms to predict future electric vehicle adoption rates and help stakeholders design targeted interventions. Additionally, we will analyze the environmental and economic impacts of electric vehicle adoption, such as reductions in greenhouse gas emissions and potential cost savings. We will also examine the differences in electric vehicle adoption rates and usage patterns between rural and urban areas and investigate the potential impact of increasing electric vehicle adoption in Washington.

How the approach addresses (fully or partially) the problem.

The approach will address the problem statement partially. It might only address some aspects of the problem statement but will provide valuable insights and potential environmental and economic impacts.

- Identifying factors that influence electric vehicle adoption, such as demographic characteristics, incentives, and availability of charging infrastructure
- Comparing electric vehicle adoption rates and usage patterns between different areas of Washington state
- Conducting a comparative analysis of electric vehicle policies and programs in other states to identify best practices and lessons that we can learn from their experiences

The proposed approach will provide partial insights and recommendations to policymakers, vehicle manufacturers, and energy companies on promoting electric vehicle adoption, reducing greenhouse gas emissions, and creating economic and social benefits in Washington State.

Data (Minimum of 3 Datasets - but no requirement on the number of fields or rows)

Below datasets are for Washington State

- Electric vehicle population data: https://catalog.data.gov/dataset/electric-vehicle-population-data
- Electric Vehicle Title and Registration Activity: https://catalog.data.gov/dataset/electric-vehicle-title-and-registration-activity
- Electric Vehicle Population Size History: https://catalog.data.gov/dataset/electric-vehicle-population-size-history
- Electric Vehicle Population Size History By County: https://catalog.data.gov/dataset/electric-vehicle-population-size-history-by-county
- WA Tax Exemptions - Potential Eligibility by Make/Model Excluding Vehicle Price Criteria: https://catalog.data.gov/dataset/wa-tax-exemptions-potential-eligibility-by-make-model-excluding-vehicle-price-criteria

Required Packages

- tidyverse: for data cleaning, wrangling, and visualization
- readxl: for reading Excel files
- writexl: for writing data frames to Excel file
- stats: for a Wide range of stats functions (linear regression, hypothesis testing)
- dplyr: for data manipulation and summarization
- purrr: functional data manipulation and to analyze multiple data frames
- ppcor: to commute correlation, helpful in examining the relationship between variables

- car: For regression analysis and model diagnostics
- ggplot2: for creating high-quality and customizable visualizations
- tidyr: for reshaping and tidying data
- lubridate: for working with dates and times
- maptools: for working with spatial data and creating maps

Of course, other packages may be helpful depending on the specific analyses and modeling techniques used.

Plots and Table Needs

Here are some possible plots and tables that could help illustrate the findings for each research question:

> How does the distribution of electric vehicles vary across different regions in Washington state, such as by county, city, or zip code?

- Plot: Map showing the number or percentage of electric vehicles registered in each county or city in Washington state, with different colors representing different ranges of values.
- Table: Summary table of the number or percentage of electric vehicles registered in each county or city in Washington state.

> What factors influence electric vehicle adoption in Washington state, such as incentives, availability of charging infrastructure, or demographic characteristics?

- Plot: Scatter plot or bar chart showing the relationship between electric vehicle adoption rates and potential explanatory variables, such as incentives, availability of charging infrastructure, or demographic characteristics.
- Table: Summary table of the correlation coefficients between electric vehicle adoption rates and potential explanatory variables, along with p-values and 95% confidence intervals.

> What is Washington State's current electric vehicle adoption rate, and how has it changed?

- Plot: Line plot showing the number or percentage of electric vehicles registered in Washington state by year, with separate lines for different vehicle types or regions.
- Table: Summary table of the number or percentage of electric vehicles registered in Washington state by year, broken down by vehicle types or regions.

> How do the availability and accessibility of electric vehicle charging infrastructure differ across different regions in Washington state?

- Plot: Map showing the number or density of electric vehicle charging stations in each county or city in Washington state, with different colors representing different ranges of values.
- Table: Summary table of the number or density of electric vehicle charging stations in each county or city in Washington state.

> How does the adoption rate and usage pattern of electric vehicles in Washington state compare to other states with similar demographic and economic characteristics?

- Plot: Bar chart or scatter plot showing the electric vehicle adoption rates or usage patterns in Washington state compared to other states with similar demographic and economic characteristics.

- Table: Summary table of the electric vehicle adoption rates or usage patterns in Washington and other similar states, along with statistical tests for significant differences.

  How do Washington State's electric vehicle policies and programs compare to other states?

- Plot: Bar chart showing the key features or outcomes of Washington State's electric vehicle policies and programs compared to other states.
- Table: Summary table of the critical features or outcomes of Washington State's electric vehicle policies and programs, along with those of other states.

  What is the potential impact of increasing the electric vehicle adoption rate in Washington state on reducing greenhouse gas emissions and improving air quality?

- Plot: Line plot showing the trend in greenhouse gas emissions or air quality in Washington state under different scenarios of increasing electric vehicle adoption rates.
- Table: Summary table of the percentage reduction in greenhouse gas emissions or improvement in air quality in Washington state due to increasing electric vehicle adoption rates, broken down by different scenarios or time horizons.

Questions for future steps

To answer the research questions, you may need to learn the following:

- How to scrape data from various sources and merge datasets
- How to perform spatial analysis and create maps to visualize the distribution of electric vehicles and charging infrastructure across different regions in Washington state
- How to perform a regression analysis to identify factors that influence electric vehicle adoption rates
- How to perform cluster analysis to group regions with similar electric vehicle adoption patterns
- How to perform comparative analysis to compare Washington state's electric vehicle policies and programs to other states
- How to estimate the potential impact of increasing electric vehicle adoption rates on reducing greenhouse gas emissions and improving air quality.

Week-2

How to import and clean the datasets

To prepare and clean the data for analysis, the following steps can be taken:

- Data Import: Use the appropriate R packages, such as readxl, to import the electric vehicle population data from Excel or CSV files into R.

- Data Inspection: Explore the structure of the dataset by examining the dimensions, variable names, and data types. Use functions like dim(), names(), and str() to gain an understanding of the data.

- Handling Missing Values: Identify missing values in the dataset using functions like is.na() and assess the extent of missing data. Decide on an appropriate strategy to handle missing values, such as imputation or deletion.

- Data Cleaning: Perform necessary data cleaning steps to ensure data integrity and consistency. This may involve removing duplicate records, correcting inconsistencies in variable names or categories, standardizing formats

- Variable Transformation: Convert variables to their appropriate data types. For example, convert categorical variables to factors or character strings, and convert date or time variables to the appropriate date/time format
- Filtering and Subset Selection: Subset the dataset to include only the relevant variables needed for analysis. Remove any variables that are not applicable to the research questions
- Creating Derived Variables: Create new variables only if necessary, by performing calculations or transformations on existing variables. For example, calculating the age of the electric vehicles based on the model year or creating indicators for specific geographic regions.
- Data Integration: If required, merge the electric vehicle population data with other relevant datasets, such as demographic data or charging infrastructure data, using common identifiers like county, city, or zip code.
- Data accuracy Validation: Validate the cleaned dataset to ensure the accuracy and consistency of the data. Check for any inconsistencies or discrepancies that may have occurred during the cleaning process.

What does the final datasets look like?

Electric Vehicle Population Data

```r
ev_population_df <- read.csv("data/week-9/Electric_Vehicle_Population_Data.csv")

#Remove unnecessary columns
ev_population_df <- subset(ev_population_df, select = -c(VIN..1.10.,
                                                          DOL.Vehicle.ID,
                                                          Vehicle.Location
                                                        ))
#rename columns
colnames(ev_population_df) <- c("County", "City", "State", "Postal_Code",
                                "Model_Year", "Make", "Model", "EV_Type",
                                "CAFV_Eligibility", "Electric_Range",
                                "Base_MSRP", "Legislative_District",
                                "Electric_Utility", "Census_Tract")


#Handle Missing values
ev_population_df <- ev_population_df[complete.cases(ev_population_df), ]

head(ev_population_df, n=5)
```

```
##     County    City State Postal_Code Model_Year   Make   Model
## 1   Yakima  Yakima    WA       98908       2018  TESLA MODEL X
## 2   Kitsap Poulsbo    WA       98370       2021  HONDA CLARITY
## 3     King Seattle    WA       98199       2019  TESLA MODEL 3
## 4     King Seattle    WA       98119       2013 NISSAN    LEAF
## 5 Thurston   Lacey    WA       98516       2017  TESLA MODEL S
##                                  EV_Type
## 1         Battery Electric Vehicle (BEV)
## 2 Plug-in Hybrid Electric Vehicle (PHEV)
## 3         Battery Electric Vehicle (BEV)
## 4         Battery Electric Vehicle (BEV)
```

```
## 5         Battery Electric Vehicle (BEV)
##                          CAFV_Eligibility Electric_Range Base_MSRP
## 1 Clean Alternative Fuel Vehicle Eligible            238         0
## 2 Clean Alternative Fuel Vehicle Eligible             47         0
## 3 Clean Alternative Fuel Vehicle Eligible            220         0
## 4 Clean Alternative Fuel Vehicle Eligible             75         0
## 5 Clean Alternative Fuel Vehicle Eligible            210         0
##   Legislative_District                             Electric_Utility
## 1                   14                                   PACIFICORP
## 2                   23                       PUGET SOUND ENERGY INC
## 3                   36 CITY OF SEATTLE - (WA)|CITY OF TACOMA - (WA)
## 4                   36 CITY OF SEATTLE - (WA)|CITY OF TACOMA - (WA)
## 5                   22                       PUGET SOUND ENERGY INC
##   Census_Tract
## 1  53077001100
## 2  53035090400
## 3  53033005803
## 4  53033005804
## 5  53067012226
```

Electric Vehicle Population size history

```r
ev_population_history <- read.csv("data/week-9/Electric_Vehicle_Population_Size_History.csv")

#Convert Date column to the appropriate date format
ev_population_history$Date <- as.Date(ev_population_history$Date,
                                      format = "%B %d %Y")

#rename columns
colnames(ev_population_history) <- c("Date", "PHEV_Count", "BEV_Count",
                                     "EV_Total")


#Checking for missing values
sum(is.na(ev_population_history))
```

```
## [1] 0
```

```r
head(ev_population_history, n=5)
```

```
##         Date PHEV_Count BEV_Count EV_Total
## 1 2017-01-31       7688     14737    22425
## 2 2017-02-28       7889     15205    23094
## 3 2017-03-31       8093     15595    23688
## 4 2017-04-30       8287     16103    24390
## 5 2017-05-31       8498     16568    25066
```

Electric_Vehicle_Population_Size_History_By_County

```r
ev_population_by_county <- read.csv("data/week-9/Electric_Vehicle_Population_Size_History_By_County.csv

#Convert Date column to the appropriate date format
```

```r
ev_population_by_county$Date <- as.Date(ev_population_by_county$Date,
                                        format = "%B %d %Y")

#rename columns
colnames(ev_population_by_county) <- c("Date", "County", "State",
                                       "Vehicle_Primary_Use", "BEVs",
                                       "PHEVs", "EV_Total", "Non_EV_Total",
                                       "Total_Vehicles", "Percent_EV")

#Checking for missing values
sum(is.na(ev_population_by_county))
```

```
## [1] 0
```

```r
head(ev_population_by_county, n=5)
```

```
##         Date    County State Vehicle_Primary_Use BEVs PHEVs EV_Total
## 1 2020-06-30    Juneau    AK           Passenger    0     2        2
## 2 2020-06-30       Lee    FL           Passenger    1     0        1
## 3 2020-06-30      Maui    HI           Passenger    1     0        1
## 4 2020-06-30 Middlesex    MA           Passenger    1     1        2
## 5 2020-05-31     Jones    GA           Passenger    1     0        1
##   Non_EV_Total Total_Vehicles Percent_EV
## 1           28             30       6.67
## 2           45             46       2.17
## 3           31             32       3.12
## 4           99            101       1.98
## 5            1              2      50.00
```

Electric_Vehicle_Title_and_Registration_Activity

```r
ev_registration <- read.csv("data/week-9/Electric_Vehicle_Title_and_Registration_Activity.csv")

# Remove unnecessary columns
ev_registration <- ev_registration[, c("Clean.Alternative.Fuel.Vehicle.Type",
                                       "VIN..1.10.", "DOL.Vehicle.ID",
                                       "Model.Year", "Make", "Model",
                                       "Vehicle.Primary.Use", "Electric.Range",
                                       "Odometer.Reading",
                                       "Odometer.Code",
                                       "New.or.Used.Vehicle", "Sale.Price",
                                       "Sale.Date", "Base.MSRP", "Transaction.Type",
                                       "DOL.Transaction.Date", "Transaction.Year",
                                       "County", "City", "State.of.Residence",
                                       "Postal.Code")]


# Handle missing values
ev_registration[ev_registration == ""] <- NA

# Get the current column names
old_colnames <- colnames(ev_registration)
```

```r
# Replace periods with underscores in column names
new_colnames <- gsub("\\.", "_", old_colnames)
# Assign the new column names to the DataFrame
colnames(ev_registration) <- new_colnames

# Convert data types
ev_registration$Model_Year <- as.integer(ev_registration$Model_Year)
ev_registration$Electric_Range <- as.numeric(ev_registration$Electric_Range)

# Standardize date formats
ev_registration$Sale_Date <- as.Date(ev_registration$Sale_Date
                                     , format = "%B %d %Y")
ev_registration$DOL_Transaction_Date <- as.Date(ev_registration$DOL_Transaction_Date
                                                , format = "%B %d %Y")

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

```
## [1] 505002
```

```r
# Remove rows with missing sales date
ev_registration <- ev_registration[complete.cases(ev_registration$DOL_Transaction_Date), ]

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

```
## [1] 505002
```

```r
head(ev_registration, n=5)
```

```
##      Clean_Alternative_Fuel_Vehicle_Type VIN__1_10_ DOL_Vehicle_ID Model_Year
## 1         Battery Electric Vehicle (BEV) 7SAYGDEF6N      197458138       2022
## 2         Battery Electric Vehicle (BEV) KNDCC3LG2N      177586440       2022
## 3         Battery Electric Vehicle (BEV) KNDCC3LG2N      177586440       2022
## 4 Plug-in Hybrid Electric Vehicle (PHEV) 1G1RD6E43E      205937539       2014
## 5 Plug-in Hybrid Electric Vehicle (PHEV) 1G1RD6E43E      205937539       2014
##        Make   Model Vehicle_Primary_Use Electric_Range Odometer_Reading
## 1     TESLA Model Y           Passenger              0                0
## 2       KIA    Niro           Passenger              0               20
## 3       KIA    Niro           Passenger              0                0
## 4 CHEVROLET    Volt           Passenger             38            62742
## 5 CHEVROLET    Volt           Passenger             38                0
##                                        Odometer_Code New_or_Used_Vehicle
## 1 Odometer reading is not collected at time of renewal                 New
## 2                                       Actual Mileage                 New
## 3 Odometer reading is not collected at time of renewal                 New
## 4                                       Actual Mileage                Used
## 5 Odometer reading is not collected at time of renewal                Used
##   Sale_Price  Sale_Date Base_MSRP      Transaction_Type DOL_Transaction_Date
## 1          0       <NA>         0 Original Registration           2022-05-04
## 2      45419 2021-08-23         0        Original Title           2021-09-15
## 3          0       <NA>         0 Original Registration           2021-09-15
```

```
## 4          0 2019-04-09         0        Original Title           2022-06-10
## 5          0       <NA>         0 Original Registration           2022-06-10
##   Transaction_Year   County    City State_of_Residence Postal_Code
## 1             2022 Thurston OLYMPIA                 WA       98502
## 2             2021 Thurston OLYMPIA                 WA       98502
## 3             2021 Thurston OLYMPIA                 WA       98502
## 4             2022 Thurston OLYMPIA                 WA       98502
## 5             2022 Thurston OLYMPIA                 WA       98502
```

WA_Tax_Exemptions_-_Potential_Eligibility_by_Make_Model_Excluding_Vehicle_Price_Criteria

```r
wa_tax_exemptions <- read.csv("data/week-9/WA_Tax_Exemptions.csv")

# Handle missing values
wa_tax_exemptions[wa_tax_exemptions == ""] <- NA

# Get the current column names
old_colnames <- colnames(wa_tax_exemptions)
# Replace periods with underscores in column names
new_colnames <- gsub("\\.", "_", old_colnames)
# Assign the new column names to the DataFrame
colnames(wa_tax_exemptions) <- new_colnames

# Convert data types
wa_tax_exemptions$Model_Year <- as.integer(wa_tax_exemptions$Model_Year)

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

```
## [1] 0
```

```r
# Remove rows with missing sales date
ev_registration <- ev_registration[complete.cases(ev_registration$Sale_Date), ]

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

```
## [1] 246
```

```r
# Split the "Vehicle Model Description" column by space
split_description <- strsplit(wa_tax_exemptions$Vehicle_Model_Description, " ")

# Extract the first element after splitting
wa_tax_exemptions$Vehicle_Model_Description <- sapply(split_description, "[[", 1)

# Extract the rest of the data into a new variable
wa_tax_exemptions$Model_Extension <- sapply(split_description, function(x) ifelse(length(x) > 1, x[[2]]

head(wa_tax_exemptions, n=5)
```

```
##   Model_Year Vehicle_Make Vehicle_Model_Description Alternative_Fuel_Type
```

```
## 1       2023         Audi                     e-tron      Battery Electric
## 2       2023         Audi                     e-tron      Battery Electric
## 3       2023         Audi                     e-tron      Battery Electric
## 4       2023         Audi                     e-tron      Battery Electric
## 5       2023         Audi                     e-tron      Battery Electric
##   Model_Extension
## 1              GT
## 2         quattro
## 3               S
## 4               S
## 5       Sportback
```

Discuss how you plan to uncover new information in the data that is not self-evident

To uncover new information in the data that is not self-evident, I would employ various exploratory data analysis (EDA) techniques and statistical methods. Here are some approaches I would consider:

- Data Visualization: I would create visualizations such as histograms, box plots, scatter plots, and heatmaps to explore the relationships and patterns in the data. Visualizations can reveal insights that may not be apparent from raw data, helping identify trends, outliers, or clusters.
- Descriptive Statistics: Calculating summary statistics like mean, median, standard deviation, and correlation coefficients can provide a deeper understanding of the data. Examining the distribution of variables, identifying central tendencies, and measuring the strength of relationships between variables can uncover hidden patterns or associations.
- New Variables: I would derive new variables or features from the existing data that could potentially provide valuable insights. For example, combining multiple variables to create a new metric, calculating ratios or percentages, or grouping categorical variables into meaningful categories can reveal hidden relationships or patterns.
- Hypothesis Testing: Formulating hypotheses based on prior knowledge or assumptions, I would conduct statistical tests to determine if there is significant evidence to support or reject these hypotheses. This could involve t-tests, chi-square tests, ANOVA, or regression analysis, depending on the nature of the variables and research questions.

By combining these approaches and leveraging domain knowledge, I aim to uncover new insights and patterns in the data that are not immediately apparent. The iterative process of exploration, analysis, and interpretation will help reveal hidden information and contribute to a deeper understanding of the dataset.

What are different ways you could look at this data?

There are several different ways to explore and analyze the dataset.

- Descriptive Analysis
- Geographic Analysis
- Temporal Analysis
- Correlation and Regression Analysis
- Comparative Analysis
- Economic and Environmental Analysis
- Machine Learning and Predictive Modeling

How do you plan to slice and dice the data?

To slice and dice the data effectively, I will use various techniques and strategies based on above research questions

- Filtering: I will use filtering techniques to extract specific subsets of data based on criteria such as location, time period or vehicle characteristics
- Aggregation: Aggregating the data allows for summarizing information at different levels of granularity
- Grouping: Grouping the data by relevant variables such as vehicle make, model, or primary use enables comparative analysis and understanding of patterns within specific categories
- Time Series Analysis: I can slice the data into different time periods, such as months, quarters, or years, and analyze the trends, growth rates, and seasonality patterns in electric vehicle adoption over time.
- Cross-Tabulation: I can cross-tabulate variables like county and electric vehicle adoption to identify variations in adoption rates across different counties or cross-tabulate demographic characteristics with adoption rates to understand any correlations.
- Machine Learning Techniques: Using machine learning algorithms, I can apply techniques such as clustering or segmentation to group the data based on similar characteristics or patterns.

How could you summarize your data to answer key questions?

- Descriptive Statistics
- Tabular Summaries
- Visualization
- Geographic Mapping
- Comparative Analysis
- Correlation and Regression Analysis
- Machine Learning Predictions

I can effectively communicate the key findings and insights derived from the data, providing comprehensive answers to the key questions regarding electric vehicle adoption in Washington state.

Do you plan on incorporating any machine learning techniques to answer your research questions? Explain.

Yes, incorporating machine learning techniques can be valuable in answering research questions related to electric vehicle adoption. Machine learning algorithms can help uncover patterns, relationships, and predictive insights from the data.

- Predictive Modeling: Machine learning algorithms can be used to develop predictive models that estimate future electric vehicle adoption rates.
- Feature Selection: Machine learning algorithms can identify the most influential factors in electric vehicle adoption. By analyzing the feature importance, we can understand which variables have the most significant impact on adoption rates. This information can guide policy decisions and highlight the key areas of focus for promoting electric vehicle adoption.

- Clustering and Segmentation: Machine learning algorithms can group regions or areas based on similar characteristics to identify segments with distinct adoption patterns. This segmentation can help target specific regions or demographics for tailored policies, incentives, or marketing campaigns. By understanding the unique needs and challenges of different segments, policymakers can develop more effective strategies to promote electric vehicle adoption.

By incorporating machine learning techniques, we can leverage the power of advanced analytics to extract valuable insights, make predictions, and inform decision-making processes related to electric vehicle adoption

Questions for future steps

- How can we further refine and optimize the machine learning models used for predicting electric vehicle adoption rates? Are there additional variables or data sources that can improve the accuracy and reliability of the predictions?
- What are the specific demographic and socioeconomic factors that influence electric vehicle adoption in different regions of Washington state? Can we identify any unique characteristics or trends that may help tailor policies and interventions for specific populations?
- How can we assess the long-term environmental and economic impacts of increasing electric vehicle adoption in Washington state? Are there specific metrics or indicators that can provide a comprehensive evaluation of the benefits, such as reductions in greenhouse gas emissions, improvements in air quality, and potential cost savings?
- Can we conduct a comparative analysis of Washington state's electric vehicle policies and programs with those of other states to identify best practices, lessons learned, and potential areas for improvement? How can we ensure a robust and comprehensive comparison that considers various aspects, such as incentives, charging infrastructure, and regulatory frameworks?
- What are the potential barriers and challenges to increasing electric vehicle adoption in rural areas of Washington state? How can we address the unique needs and constraints of rural communities and develop targeted strategies to promote electric vehicle adoption in these regions?
- How can we engage key stakeholders, such as local governments, utilities, and community organizations, in the process of promoting electric vehicle adoption? What partnerships and collaborations can be fostered to support infrastructure development, awareness campaigns, and policy implementation?
- Can we explore innovative financing models and incentives that can make electric vehicles more affordable and accessible to a broader range of consumers? How can we address the upfront cost barrier and incentivize the transition to electric vehicles, particularly for low-income individuals and communities?
- How can we continuously monitor and track the progress of electric vehicle adoption in Washington state? What data collection and monitoring systems can be established to ensure ongoing evaluation, identification of emerging trends, and adaptation of strategies based on evolving market dynamics and technological advancements?

Week-3

Accelerating Electric Vehicle Adoption in Washington State: Data-Driven Insights and Recommendations

Problem Statement:

- The problem at hand is to understand the current state of electric vehicle (EV) adoption in Washington state and identify strategies to accelerate its growth.
- The goal is to promote sustainable transportation, reduce greenhouse gas emissions, and improve air quality. Key research questions include analyzing the distribution of EVs across regions, determining influential factors for adoption, assessing the availability of charging infrastructure, comparing policies with other states, and evaluating the environmental impact.

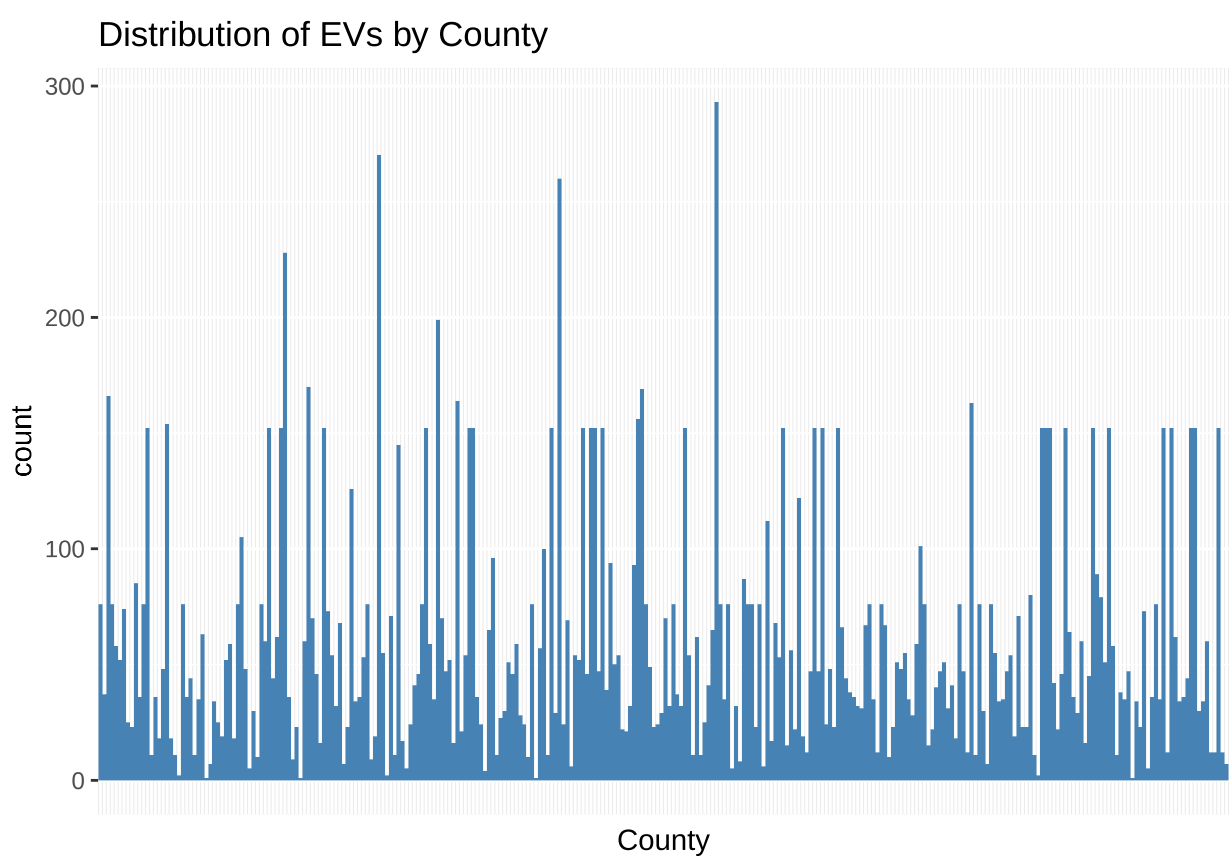
Approach and Methodology:

To address the problem statement, we employed a comprehensive data analysis approach. We collected and cleaned multiple datasets, including EV population data, EV registration activity, EV tax exemptions, and EV title and registration activity. We used R programming language and various packages such as dplyr and ggplot2 for data manipulation, visualization, and analysis.

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.2.3
```

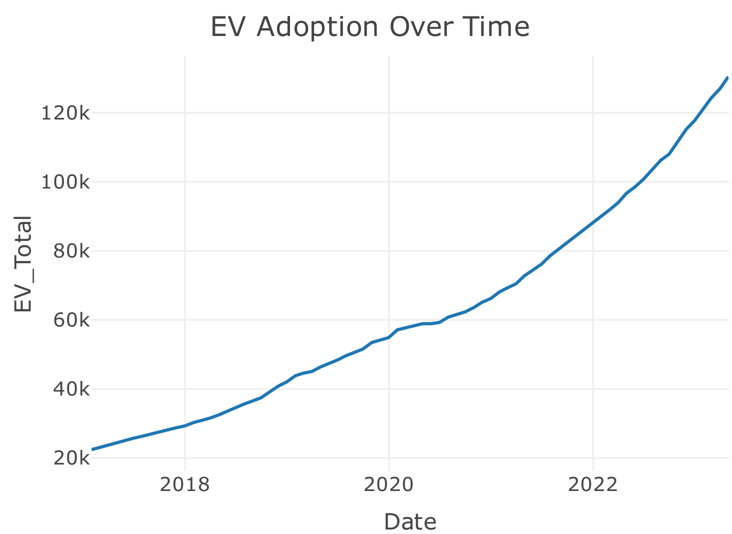
```
ggplot(ev_population_by_county, aes(x = County)) +
  geom_bar(fill = "steelblue") +
#  annotation_custom(grob = label_table, xmin = -Inf, xmax = Inf, ymin = -Inf, ymax = -Inf) +
  labs(title = "Distribution of EVs by County") +
  theme(axis.text.x = element_blank(), axis.ticks.x = element_blank())
```



```r
# Plot the trend of EV adoption over time
library(plotly)
```

```
## Warning: package 'plotly' was built under R version 4.2.3

##
## 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
```

```r
ev_population_history$Date <- as.Date(ev_population_history$Date, format = "%B %d %Y")


plot_ly(ev_population_history, x = ~Date, y = ~`EV_Total`,
        type = "scatter", mode = "lines") %>%
  layout(title = "EV Adoption Over Time", xaxis = list(title = "Date"), yaxis = list(title = "EV_Total")
```



```r
# Statistical Analysis
# Calculate correlation between variables
correlation <- cor(ev_population_df[, c("Electric_Range", "Base_MSRP")])

# Hypothesis testing
# Example: Compare the mean electric range between different electric vehicle types
# Assuming you have a categorical variable "Electric Vehicle Type" in your dataset
library(car)
```

```
## Warning: package 'car' was built under R version 4.2.3
```

```
## Loading required package: carData
```

```
## Warning: package 'carData' was built under R version 4.2.3
```

```r
lm_model <- lm(`Electric_Range` ~ `EV_Type`, data = ev_population_df)
Anova(lm_model)
```

```
## Warning in printHypothesis(L, rhs, names(b)): one or more coefficients in the hypothesis include
##      arithmetic operators in their names;
##   the printed representation of the hypothesis will be omitted
```

```
## Anova Table (Type II tests)
##
## Response: Electric_Range
##               Sum Sq     Df F value    Pr(>F)
## EV_Type     83298212      1    9011 < 2.2e-16 ***
## Residuals 1202985442 130136
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Our analysis involved exploratory data analysis, descriptive statistics, and correlation studies. We utilized data visualization techniques such as bar charts, and time series plots to uncover insights. While no specific modeling technique was required for this final phase, a recommendation for future implementation would be to develop predictive models for EV adoption rates based on the available datasets

Interesting Insights:

Our analysis revealed several interesting insights: **Distribution**: EVs were concentrated in urban areas such as Seattle, but there was growing adoption in other regions. **Influential Factors**: Factors such as incentives, charging infrastructure availability, and demographic characteristics played significant roles in EV adoption. **Charging Infrastructure**: Certain regions had better availability and accessibility of charging infrastructure than others. Policy Comparison: Washington state had implemented commendable EV policies, but there were opportunities to learn from other states. **Environmental Impact**: Increasing EV adoption could lead to substantial reductions in greenhouse gas emissions and improvements in air quality.

Implications to the Consumer:

The analysis provides valuable insights for consumers interested in EV adoption. It highlights the geographic distribution of EVs, allowing consumers to identify areas with higher EV concentration and potentially better charging infrastructure. The influential factors identified can help consumers understand the benefits, incentives, and support available for adopting EVs. Moreover, the analysis showcases the environmental benefits of EVs, encouraging consumers to make sustainable transportation choices.

Limitations and Future Improvements:

While our analysis provides valuable insights, it has some limitations.

- First, the analysis primarily relies on historical data, and future trends and developments may impact the findings. Incorporating real-time data and continuously monitoring EV adoption rates would provide more accurate insights.
- Second, the analysis does not account for individual consumer preferences and decision-making processes, which could be explored through surveys or interviews. - - Additionally, the analysis focuses on Washington state and may not capture nuances specific to other regions.

To improve and build upon this analysis, future researchers could:

- Include more granular data, such as charging station locations, to analyze charging infrastructure gaps in detail.
- Conduct predictive modeling to forecast EV adoption rates and evaluate the impact of different policy interventions.
- Conduct surveys or interviews to gather consumer preferences, barriers, and motivations for EV adoption.
- Expand the analysis to include economic assessments, such as total cost of ownership comparisons between EVs and internal combustion engine vehicles.

By addressing these limitations and building upon the insights gained, policymakers, stakeholders, and consumers can make informed decisions and develop targeted strategies to accelerate the adoption of electric vehicles, leading to a more sustainable and environmentally friendly transportation system in Washington state.