Assignment2

Manas Kore

06/16/2024

The dataset being analyzed contains information on Battery Electric Vehicles (BEVs) and Plug-in Hybrid Electric Vehicles (PHEVs) currently registered through the Washington State Department of Licensing (DOL). The dataset includes details such as the vehicle identification number (VIN), county, city, state, postal code, model year, make, model, electric vehicle type, eligibility for Clean Alternative Fuel Vehicle (CAFV), electric range, base MSRP, legislative district, DOL vehicle ID, vehicle location, electric utility, and 2020 census tract.

This data was collected to monitor the distribution and characteristics of electric vehicles in Washington State. Researchers might use this data to answer questions about the geographic distribution of electric vehicles, the popularity of different models, and trends in electric vehicle adoption.

The dataset is saved in a CSV (Comma-Separated Values) file format, which is a type of delimited text file. The delimiter used is a comma, and it can be opened with any text editor, spreadsheet software like Microsoft Excel, or data analysis programs like R.

# Reading the data into R

library(readr) # loading the readr package

# Use read\_csv() function to read the data

ev\_data <- read\_csv(“/Users/mkore/Downloads/Electric\_Vehicle\_Population\_Data.csv”)

# Cleaning the data

library(dplyr) # loading the dplyr package

# Renaming columns for easier reference

ev\_data <- ev\_data %>% rename( VIN = VIN (1-10), County = County, City = City, State = State, Postal\_Code = Postal Code, Model\_Year = Model Year, Make = Make, Model = Model, EV\_Type = Electric Vehicle Type, CAFV\_Eligibility = Clean Alternative Fuel Vehicle (CAFV) Eligibility, Electric\_Range = Electric Range, Base\_MSRP = Base MSRP, Legislative\_District = Legislative District, DOL\_Vehicle\_ID = DOL Vehicle ID, Vehicle\_Location = Vehicle Location, Electric\_Utility = Electric Utility, Census\_Tract = 2020 Census Tract )

# Getting the number of rows and columns

num\_rows <- nrow(ev\_data) num\_cols <- ncol(ev\_data)  
  
This dataframe has 186880 rows and 17 columns.

# Create a data frame for column descriptions

column\_descriptions <- data.frame( Column\_Name = colnames(ev\_data), Description = c( “Vehicle Identification Number (first 10 characters)”, “County of vehicle registration”, “City of vehicle registration”, “State of vehicle registration”, “Postal code of vehicle registration”, “Year of the vehicle model”, “Manufacturer of the vehicle”, “Model of the vehicle”, “Type of electric vehicle (BEV or PHEV)”, “Eligibility for Clean Alternative Fuel Vehicle”, “Electric range of the vehicle in miles”, “Manufacturer’s Suggested Retail Price”, “Legislative district of vehicle registration”, “Unique vehicle ID assigned by the Department of Licensing”, “Location of the vehicle in geographical coordinates”, “Electric utility provider”, “2020 Census tract code” ) )

# Display the table using kable

library(knitr) kable(column\_descriptions, col.names = c(“Column Name”, “Description”))

# Selecting three columns for summary statistics

selected\_columns <- ev\_data %>% select(Electric\_Range, Base\_MSRP, Model\_Year)

# Calculating summary statistics

summary\_stats <- selected\_columns %>% summarise( Min\_Range = min(Electric\_Range, na.rm = TRUE), Max\_Range = max(Electric\_Range, na.rm = TRUE), Mean\_Range = mean(Electric\_Range, na.rm = TRUE), Missing\_Range = sum(is.na(Electric\_Range)),

Min\_MSRP = min(Base\_MSRP, na.rm = TRUE),  
Max\_MSRP = max(Base\_MSRP, na.rm = TRUE),  
Mean\_MSRP = mean(Base\_MSRP, na.rm = TRUE),  
Missing\_MSRP = sum(is.na(Base\_MSRP)),  
  
Min\_Year = min(Model\_Year, na.rm = TRUE),  
Max\_Year = max(Model\_Year, na.rm = TRUE),  
Mean\_Year = mean(Model\_Year, na.rm = TRUE),  
Missing\_Year = sum(is.na(Model\_Year))

)

summary\_stats

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

summary(cars)

## speed dist   
## Min. : 4.0 Min. : 2.00   
## 1st Qu.:12.0 1st Qu.: 26.00   
## Median :15.0 Median : 36.00   
## Mean :15.4 Mean : 42.98   
## 3rd Qu.:19.0 3rd Qu.: 56.00   
## Max. :25.0 Max. :120.00