EV Data Analysis Project

# Introduction

This project utilizes a recent addition to the automotive industry, Electric vehicles (EVs). They are the future of transportation on the road for a more sustainable society. Yet this adoption is not as simple as putting vehicles on the road, there are multiple variables of influence that affect the decision-making process of attaining an EV. This dataset explores those factors, which can be analyzed to show how they interact with EV adoption in the United States.

# Business Task

Using ChatGPT (CGPT) to generate some business tasks, I decided to combine two of them, the two being, Business Task 1: Identify Key Drivers of EV Adoption by State, and Business Task 2: Infrastructure Planning Recommendations.

|  |  |  |
| --- | --- | --- |
| Business Task | Identify Key Drivers of EV Adoption by State | Infrastructure Planning Recommendations |
| Objective | Determine which socioeconomic, policy, and infrastructure factors are most strongly associated with higher EV Share (%) across states. | Identify which U.S. states are underinvested in charging infrastructure relative to their current and projected EV adoption rates. |
| Deliverable(s) | A report with key insights and visualizations (e.g., heatmap, regression summary) identifying top predictors of EV adoption. | A dashboard or briefing slide deck for infrastructure planners with:   * A list of top 5 priority states for charger deployment * Suggested number/type of outlets needed to meet future demand |

I will assume of being a Junior Data Analyst at traditional vehicle manufacturing company that wants to explore the venture of manufacturing and selling EVs. These business tasks will be used to assist their decision-making process on the usage side ( if people will buy their EV offering in the US).

# Summary of data sources

## Where is it from?

This data source called “EV adoption USA” was published by on Kaggle. Link: <https://www.kaggle.com/datasets/surajshivakumar/ev-adoption-usa/data>

## Where was the data collected from?

The information to create this was extracted from:

* EV Registrations: National Renewable Energy Laboratory (NREL)
* Socioeconomic Indicators: US Census Bureau (population, income, education, labor force, unemployment)
* Charging Infrastructure & Incentives: Alternative Fuels Data Center (AFDC)
* Fuel Economy & Vehicle Registrations: Bureau of Transportation Statistics
* Gasoline Prices: American Automobile Association (AAA)
* Electricity Prices: Energy Information Administration (EIA)
* CO2 Emissions: Bureau of Transportation Statistics

## How was the data collected?

“Data was downloaded via CSV or API and combined on Year and State”

It was licenced from MIT, Copyright granted for free use by Mark Otto (2013) and Matthew Fong (2017)

# Meta data on column information

| **Variable** | **Description** |
| --- | --- |
| state | US state |
| year | Year of observation |
| EV Registrations | Number of Electric Vehicles registered |
| Total Vehicles | Total number of all vehicle registrations in the state |
| EV Share (%) | Percentage of total vehicles that are electric vehicles |
| Stations | Number of public EV charging stations |
| Total Charging Outlets | Total number of individual charging plugs available at public stations |
| Level 1 | Number of Level 1 charging outlets |
| Level 2 | Number of Level 2 charging outlets |
| DC Fast | Number of DC Fast charging outlets |
| fuel\_economy | Average fuel economy of all vehicles in the state (e.g., MPG) |
| Incentives | Presence and/or details of state-level EV incentives |
| Number of Metro Organizing Committees | Number of metropolitan planning organizations in the state |
| Population\_20\_64 | Working-age population (ages 20-64) |
| Education\_Bachelor | Number of people with a Bachelor's degree or higher |
| Labour\_Force\_Participation\_Rate | Percentage of the working-age population in the labor force |
| Unemployment\_Rate | Percentage of the labor force that is unemployed |
| Bachelor\_Attainment | Percentage of the total population with a Bachelor's degree or higher |
| Per\_Cap\_Income | Average income per person in the state |
| affectweather | A measure of concern or belief about climate change impacts |
| devharm | A measure of concern about potential harm from development |
| discuss | A measure of how often individuals discuss environmental issues |
| exp | A measure of environmental experience or exposure |
| localofficials | A measure of trust or engagement with local environmental officials |
| personal | A measure of personal responsibility towards the environment |
| reducetax | Support for reducing taxes related to environmental policies |
| regulate | Support for government regulation on environmental issues |
| worried | A measure of worry or concern about environmental problems |
| price\_cents\_per\_kwh | Average price of electricity per kilowatt-hour (in cents) in the state |
| gasoline\_price\_per\_gallon | Average price of gasoline per gallon in the state |
| Total | Total number of all registered vehicles (redundant with 'Total Vehicles') |
| Trucks | Number of registered trucks in the state |
| Trucks\_Share | Percentage of total vehicles that are trucks |
| Party | Predominant political party affiliation in the state |

# Analysis plan

## Part I

* Open the csv file on excel to get a sense of the structure; read through all of it with the meta data table to understand the data inputted.
* Clean data on RStudio.
* Write down idea of metrics to answer the business tasks.
* Manipulate data to curate those answers.
* Create a static report using R Markdown.

## Part II

* Open the unedited csv file on Tableau, if needed use the edited ( on R ) file.
* Create visualizations for the business task.
* Create a dynamic dashboard.
* Create a storyline for presentations.

# Change log

Removed the redundant column A, made column B, A. Renamed this new column as “ID Number”

Renames “state” to “State” and “year” to “Year.”

Imported data into RStudio.

Install and loaded relevant packages ( dyplr, readr, ggplot2, tidyverse, etc)

Created “complete\_states” table.

Created these charts using r; EV Registrations over time in the US, EV Stations in the U.S over time, Charging stations over time, Total incentives granted over time in the U.S., Total Vehicles and EVs in the U.S from 2016 to 2023, Median EV % share from 2016 to 2023, Distribution of EV share (%) in 2016 compared to 2023.

Created a R Markdown HTML to illustrate the graphs and code used to make them.

Created a visualizations, dashboards, and Storylines on Tableau. This was done using the “complete\_states” dataset made on R.

# Observations about data

During the data inspection process, I found there was some null data. For political parties in parties between 2016 and 2017, there was no data input. fuel\_economy, incentives, number of metro organizing committees, affect weather, devharm, discuss, exp, local officials, personal, reduce tax, gasoline\_price\_per\_gallon, total\_trucks, trucks\_share, party.

This will significantly affect reporting of this project, so I will create a “clean” dataset that only shows states with complete data. “complete\_states” was made.

Only 50 states have complete data in the years 2018, 2020, 2021, 2022, and 2023.

**I confirmed that there was no completed data in 2019, as to why is a mystery.**

EVs are slowly becoming a choice for Americans as there is a steady recognition and concern for the environment. Regardless of political differences, it is becoming more apparent to residents that climate change is influential.

Trucks are just a segment of the larger competition of Internal Combustion Engine (ICE) vehicles. Truck share and purchases (in respect to average income) show that it is increasing steadily in demand.

# Limitations of the data

The metadata descriptions of social measurements, such as worry, were poorly defined. How it was calculated was also poorly defined. This made it difficult to interpret and report findings.

Social measurements, sources were not given explicitly.

Incomplete data was a minor issue, as it was hard to create consistent trends.

When a complete dataset was made, it created a void of data in 2019, which is explicable due to COVID-19.

The data stops at 2023, and this report is being made in 2025. So, relevance can be challenged.

# **Report**

## **Identifying Key Drivers of EV Adoption by U.S. State**

### ***Key Findings***

Based on the data analyzed through R and visualized in the accompanying Tableau storyline, several key drivers of electric vehicle (EV) adoption have emerged across U.S. states:

* **Income levels**: States with higher per capita income tend to have greater EV adoption.
* **Government incentives**: Both federal and state-level EV incentives positively influence adoption.
* **Charging infrastructure**: The availability of public EV charging stations correlates with EV registration trends.
* **Environmental concern**: States with higher levels of environmental awareness and engagement show stronger EV uptake.

There is a gradual but steady cultural shift among Americans toward environmental responsibility. Despite political and regional differences, climate change is increasingly recognized as a tangible issue, contributing to growing EV adoption.

In contrast, **internal combustion engine (ICE) trucks** remain dominant in many regions, particularly in the South. Data shows rising truck purchases—especially in relation to average income—indicating continued demand in that segment.

### ***Strategic Recommendation***

A **slow and strategic entry** into the EV market is recommended.

With increasing public and governmental momentum toward clean energy, your company is well-positioned to enter a pivotal moment in transportation history. However, this entry should be **phased and targeted**, beginning in states with supportive conditions.

**Recommended target states**:

* **California**
* **Texas**
* **Florida**

These states have:

* High EV share
* Strong environmental awareness
* Pre-existing charging infrastructure
* Supportive government policies

### **Infrastructure Planning Insight**

The data reveals a **feedback loop** between EV adoption and charging station deployment:

* Higher EV registrations tend to drive the development of charging infrastructure.
* However, the availability of charging stations also encourages further EV purchases.

#### **Strategic Recommendation**

Start in states where **charging infrastructure already exists** and EV share is relatively high. This approach minimizes initial infrastructure costs, allowing the company to **integrate with existing systems** rather than build from scratch.

For example, launching in **California** would leverage:

* A well-developed public and private charging network
* Strong consumer demand
* State-level incentives
* Available skilled labor and suppliers

The major cost in such regions will be building out **maintenance and support infrastructure**, such as:

* First party and third-party repair networks
* Training programs for EV-specific mechanics

### **Conclusion**

Through **deliberate and regionally targeted implementation**, the company can minimize risk while entering the EV market. Focusing on high-readiness states like **California, Texas, Florida, and New York** will offer the best return on early investments.

Additionally, the **EV truck market** presents a long-term opportunity:

* Truck ownership remains popular in southern states.
* EV truck offerings are still limited, with few companies achieving strong market share.
* A successful entry will depend on aligning EV technology with the **cultural expectations of truck users**.

**Final note**: A well-executed **marketing strategy** that speaks to identity, reliability, and lifestyle—not just sustainability—could set the company apart in the competitive EV truck landscape.