**Exploring Electric Vehicle Adoption Through Hadoop-Driven Urban Analytics**

Authors: Lee Hoon Jae, Daniel Demetrios

Instructor: Jongwook Woo

Department of Information Systems, California State University

Los Angeles

Tel. 323-343-2916, Fax. 323-343--5209

e-mail: jwoo5@caltstatela.edu

**Abstract:** This research contributes to the growing knowledge of sustainable urban transportation systems by leveraging Hadoop for large-scale data analysis. This project examines the adoption of electric vehicles (EVs) and the challenges of integrating them into urban transportation systems, focusing on California and New York. We analyzed EV sales, charging infrastructure, and adoption trends using big data tools like Hadoop and Hive. While California leads in EV adoption, expanding costly Level 3 fast chargers) remains critical, especially in urban areas like New York City, where private charging is limited. New York needs to catch up in infrastructure development, which is compounded by the slow transition of traditional automakers to affordable EV production.

This study highlights the cost disparity between Level 2 (~$4,500) and Level 3 (~$50,000) chargers, emphasizing the need for targeted public investments. Visualizations created with Tableau reveal adoption patterns and infrastructure gaps, offering actionable insights for policymakers and planners. This research contributes to strategies promoting EV adoption, reducing emissions, and building sustainable urban transportation systems.

1. **Introduction**

Given the urgent need to address climate change and promote sustainable transportation solutions, this study explores extensive datasets on EV registration counts, charging station locations, and consumer adoption rates, primarily in California and New York City. This project uses Hadoop and HiveQL to process and analyze datasets to obtain valuable insights into the evolving landscape of EV adoption, a critical component in addressing climate change and reducing urban carbon footprints. California, a leader in EV adoption, offers a model of an urban area with comprehensive charging networks and supportive policies. New York City, however, offers a contrasting model and highlights the challenges typical of densely populated urban areas, such as limited charging infrastructure, high vehicle costs, and varying levels of consumer awareness [8][9]. This project identifies key trends and barriers to EV adoption and offers data-driven insights that contribute to understanding sustainable urban transportation systems.

1. **Related Work**

NASA highlights the urgency of addressing rising CO₂ levels and escalating climate challenges [1] while increasing global temperatures underscore the need for actionable mitigation strategies [2].

While electric vehicles (EVs) significantly reduce CO₂ emissions, comprehensive studies analyzing the interplay of EV adoption, charging infrastructure, and market dynamics remain scarce. Research often focuses on automakers’ transitions to EVs, evaluating their compliance with emission reduction goals and evolving consumer demands [3].

Furthermore, insights from the U.S. Department of Transportation and Atlas EV Hub reveal geographic disparities in EV adoption and infrastructure availability, with high-adoption states like California contrasting sharply with lagging regions [5][6][7]. California exemplifies successful EV promotion through its robust charging infrastructure and supportive policies [4], which boost consumer confidence in zero-emission vehicle (ZEV) initiatives.

**3. Background/Specifications**

Our project utilizes EV registration and alternative fueling station datasets collected from multiple states. These datasets collectively amount to approximately 2.55 GB, ideal for scalability testing. Although not extremely large, the data volume provides a realistic modelling scope that could be extended to larger datasets. The datasets are processed using HiveQL, with data cleansing performed in Python before loading.

*Table 1 Data Specification*

|  |  |
| --- | --- |
| Dataset |  |
| ca\_ev\_registrations\_public.csv | 267MB |
| fl\_ev\_registrations\_public.csv | 27.3MB |
| wi\_ev\_registrations\_public.csv | 266MB |
| CO\_EV\_Registrations.csv | 2.05MB |
| CT\_EV\_Registrations.csv | 357MB |
| ME\_EV\_Registrations.csv | 15.9MB |
| MN\_EV\_Registrations.csv | 12.8MB |
| MT\_EV\_Registrations.csv | 60.7MB |
| NC\_EV\_Registrations.csv | 1.75MB |
| NJ\_EV\_Registrations.csv | 42.9MB |
| NM\_EV\_Registrations.csv | 13MB |
| NY\_EV\_Registrations.csv | 776MB |
| OR\_EV\_Registrations.csv | 14.2MB |
| TN\_EV\_Registrations.csv | 11.8MB |
| TX\_EV\_Registrations.csv | 345MB |
| VA\_EV\_Registrations.csv | 21.7MB |
| VT\_EV\_Registrations.csv | 5.99MB |
| Alternative\_Fueling\_Stations.csv | 39.1MB |
| New\_ZEV\_Sales\_Last\_updated\_08-06-2024\_ada.csv | 2.7 MB |
| Vehicle\_Population\_Last\_updated\_04-30-2024\_ada.csv | 6.5 MB |

This project utilizes a Hadoop-based architecture with external tables created in Hive, leveraging HDFS for storage and data processing. The below table shows the Hadoop specification for our project.

*Table 2 Hadoop Specifications*

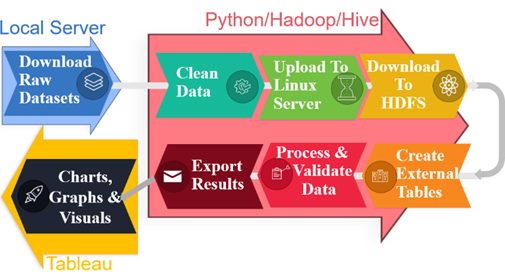
|  |  |
| --- | --- |
| Number of nodes | 5 (2 Master + 3 Work) |
| OCPUs | 10 |
| CPU Speed | 2195.084 MHz |
| Memory | 150 GB |
| Storage | 678 GB |
| HDFS Capacity | 147 GB |
| Hadoop Version | 3.2.1 |
| Hive Version | 3.1.2 |

1. **Implementation and Workflow**

Our workflow began with acquiring and preparing multiple state-specific EV registration datasets and alternative fueling station data. These datasets were downloaded in CSV format from credible open data sources such as state government and NASA websites to ensure reliability and comprehensiveness. The California dataset required Python preprocessing to resolve data quality issues, including line breaks within fields that could introduce errors during analysis. After preprocessing, all data files were securely uploaded to a Linux server and transferred to the Hadoop Distributed File System (HDFS), ensuring scalable storage and efficient data management.

To facilitate structured querying and analysis, we organized the datasets into three separate directories based on state groupings and created external tables within the Hive environment. Perform data engineering with HiveQL to create tables such as ev1\_sum and fl\_sum generated by extracting key fields, including the vehicle make, model, and registration\_valid\_date, and grouping the data by year. These tables were then unified into a comprehensive dataset, all\_vehicles\_sum, using the UNION ALL operation, which consolidated EV registration records across all states. This consolidated table served as the primary dataset for further analysis.

After data processing, the final cleaned and structured dataset, all\_veh.csv, was exported for visualization and advanced analysis. Tableau was utilized to integrate this dataset with the Alternative\_Fueling\_Stations dataset, enabling the creation of detailed visualizations. These visualizations highlighted state-wise EV distribution, growth trends over time, and geographical patterns in vehicle make and model preferences. Through this systematic workflow, the project uncovered valuable insights into EV adoption patterns and infrastructure distribution, providing a robust foundation for understanding the factors influencing the adoption of sustainable transportation systems.



*Figure 1 – Workflow Implementation*

1. **Data Cleaning/Preparation**

We implemented a multi-step data cleaning process to ensure data quality and consistency across the EV registration datasets. The California dataset presented a unique challenge due to line breaks within fields, specifically in the Vehicle ID field, which introduced potential NULL values. To address this, we developed a Python script to preprocess the data. The script identified fragmented lines by detecting entries starting with "CA-" and combined them into a single, coherent record. The cleaned dataset was then saved as cleaned\_ca\_ev\_registrations\_public.csv, ensuring all records were correctly formatted for subsequent analysis in the Hive environment.

Once preprocessing was completed, the datasets were uploaded to a Linux server and transferred to the Hadoop Distributed File System (HDFS). Directories were created in Hadoop, and datasets were categorized into tables based on states: ev1 for California, fl for Florida, wa for Washington, wi for Wisconsin, and ev3 for the remaining states (CO, CT, ME, MN, etc.). These tables were structured to allow consistent querying of vehicle details across states. The schema for these tables was configured with standard settings, such as OpenCSVSerde, to handle CSV parsing with quotes and TBLPROPERTIES ("skip.header.line.count"="1") to exclude headers during data ingestion.

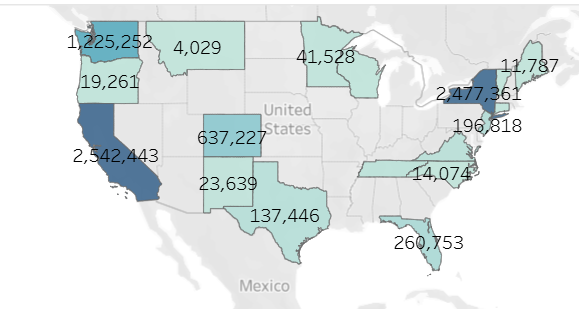
We utilized regular expressions and substring functions in Hive to standardise fields across all tables. For instance, the vehicle make and model data was extracted from the vehicle\_name field using the REGEXP\_EXTRACT function, while the year was extracted from the registration\_valid\_date field using the SUBSTRING function, similar to the techniques demonstrated in our coursework. After extracting these fields, data was grouped by state, year, make, and model to produce state-specific summary tables that counted vehicle registrations. These summary tables streamlined the data and simplified subsequent analysis.

Finally, the individual summary tables were consolidated into a single comprehensive table, all\_vehicles\_sum, using the UNION ALL operation. The UPPER function was applied to standardize the make and model fields to uppercase to maintain consistency and avoid case-sensitive discrepancies during analysis. The cleaned and consolidated dataset was then exported as all\_veh.csv and integrated with the Alternative\_Fueling\_Stations.csv dataset for visualization in Tableau. These visualizations highlighted state-wise EV distribution, vehicle trends by make and model, and the geographical spread of charging infrastructure, providing valuable insights into the evolving EV landscape.

1. **State-Wide EV Adoption Trends**

After the cleaned and consolidated data was loaded into Tableau, various visualization techniques were employed, including bar charts, heat maps, geo-spatial visualizations, and line graphs, to analyze and communicate trends in EV adoption effectively. The first visualization, a geo-spatial heat map (Figure II), illustrates the cumulative EV sales across states. In this map, states with higher cumulative sales are represented in darker shades, signifying greater levels of EV adoption. This high-level visualization provides an overview of the states that have been most active in embracing EV technology.

The heat map reveals that California leads cumulative EV registrations, given its strong policy support and extensive charging infrastructure. New York also demonstrates substantial EV adoption, with cumulative registrations comparable to California in some regions. However, it is essential to clarify that these counts include a mix of EV types, such as plug-in hybrid electric vehicles (PHEVs) and hybrid electric vehicles (HEVs), alongside fully electric vehicles (BEVs). As such, while New York’s cumulative registration numbers are significant, there may be a stronger preference for hybrids than BEVs in this state. This observation underscores the need for further analysis to differentiate consumer preferences between hybrid and fully electric models, which could provide more nuanced insights into regional adoption patterns.



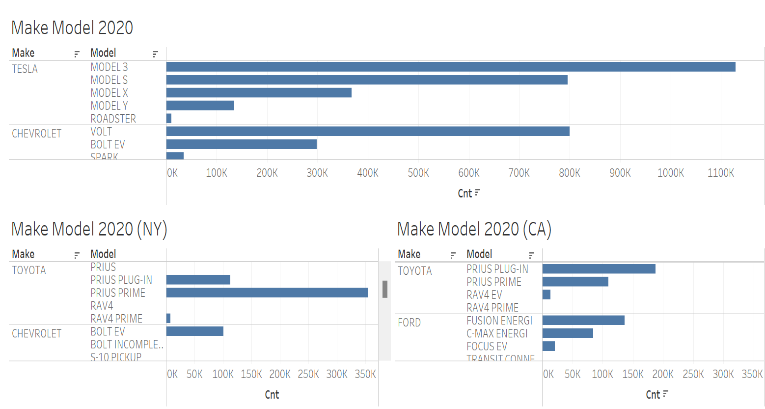
*Figure II – Cumulative Registrations per State up to 2020*

1. **Consumer Preferences for Ev Models Through 2020**

In Figure III, cumulative EV sales by model for 2020 were analyzed, explicitly focusing on consumer preferences in California and New York. The analysis revealed that the Tesla Model 3 was the best-selling model overall, with 1.13 million units sold, followed by the Model S (795,667), Chevrolet Volt (799,846), and Nissan Leaf (771,411). Toyota’s Prius Prime, a plug-in hybrid electric vehicle (PHEV), also recorded substantial sales, with 546,337 units.

In California, the Tesla Model 3 and Model S dominated EV registrations. Notably, New York recorded even higher Model 3 registrations (331,271) compared to California (306,928) but significantly fewer Model S registrations (213,296) compared to California (301,845). A particularly striking observation was the strong preference for the Toyota Prius Prime in New York, with 354,583 units registered, compared to only 109,239 in California. This trend suggests that New York consumers prefer hybrid models, reflecting a potential hesitancy to transition fully to zero-emission vehicles. This hesitancy is likely influenced by concerns over battery electric vehicle (BEV) limitations, such as range and charging infrastructure availability, which remain critical factors in consumer decision-making.

It is important to note that this analysis was limited to data up to 2020, as the most recent available California EV dataset ends that year. This provides a snapshot of EV adoption trends through 2020, while later data is addressed in subsequent sections.



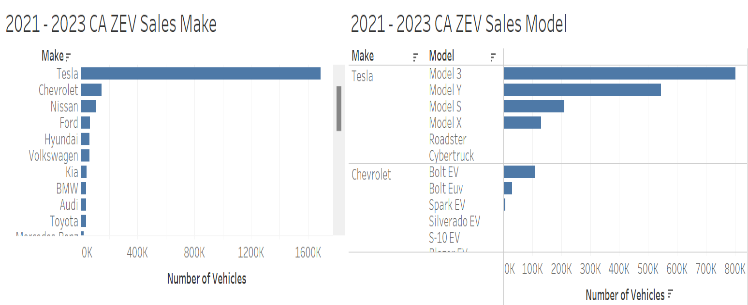
*Figure III Make Model Registrations for 2020*

1. **ZEV Market Leaders in CA (2021-2023)**

In Figure IV, California's Zero Emission Vehicle (ZEV) market from 2021 to 2023 is analyzed, emphasizing the dominance of battery electric vehicles (BEVs). As expected, Tesla leads the market with 1,689,663 registrations during this period. Among Tesla models, the Model 3 is the most popular, accounting for 801,700 registrations, followed by the Model Y with 544,890 registrations.

Beyond Tesla, the Chevrolet Bolt EV is the next most registered ZEV, with 109,244 units, reflecting its appeal as a more affordable electric vehicle option. In contrast, the Subaru Solterra recorded the fewest registrations, with only 2,566 units, highlighting its limited penetration in the competitive ZEV market.

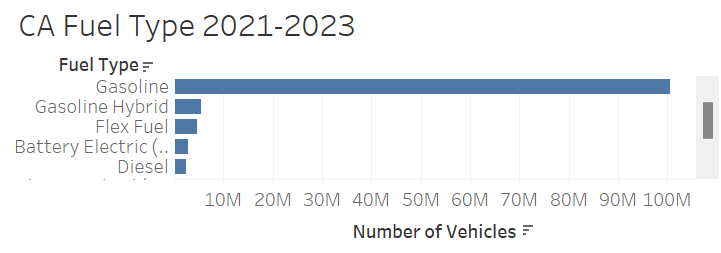
This strong preference for Tesla models underscores the brand's dominance in California's ZEV market, attributed to its established reputation, extensive Supercharger network, and consumer trust in its range and performance. The substantial gap between Tesla and other manufacturers highlights the significant challenges competing brands face in capturing market share, even as the overall demand for ZEVs continues to rise. This disparity suggests that while the market for ZEVs is expanding, it remains concentrated among a few dominant players, posing barriers to broader competition.



*Figure IV ZEV Make and Model in CA 2021-2023*

1. **Fuel Type Dominance and Regional ZEV Trend In CA**

Figure V analyzes vehicle fuel types in California from 2021 to 2023, emphasizing the continued dominance of traditional gasoline-powered vehicles. According to the bar chart, gasoline remains the most prevalent fuel type, with an overwhelming total of 100,614,050 vehicles. Despite their growing popularity, battery electric vehicles (BEVs) are a distant second, with 2,772,064 registered vehicles. This stark disparity highlights the persistent reliance on gasoline-powered vehicles in California, even as the state positions itself as a leader in Zero Emission Vehicle (ZEV) adoption.



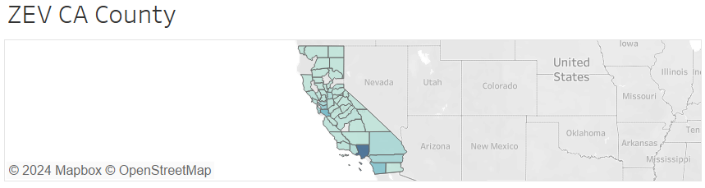
*Figure V CA Fuel Type 2021-2023*

1. **Regional ZEV Trend in CA**

To better understand ZEV adoption, a geospatial heat map visualizes ZEV registrations across California's counties. In this visualization, darker shades represent areas with higher concentrations of ZEVs, with Los Angeles County leading the state at 547,982 registrations. Following Los Angeles are Santa Clara County with 212,575 registrations and San Diego County with 165,063 registrations. These three counties alone account for a significant proportion of ZEVs in California, underscoring the impact of population density and infrastructure development in driving adoption.

This data reveals that though California has made remarkable progress in promoting ZEVs, particularly in major urban hubs with robust infrastructure and incentives, its overall market remains heavily dominated by gasoline-powered vehicles. The concentration of ZEV registrations in counties like Los Angeles and Santa Clara is likely influenced by more robust charging infrastructure, higher levels of consumer awareness, and state-level incentives tailored to densely populated regions. These factors create favourable conditions for ZEV adoption in urban centres but also reveal disparities in adoption rates across less populated or rural counties, where infrastructure is less developed.

Such findings highlight the ongoing challenges in achieving widespread ZEV adoption across California. The lack of equitable distribution of charging infrastructure and limited consumer confidence in less urbanized areas continue to hinder the state's efforts to transition fully to battery electric vehicles. To address these challenges, it will be essential for California to implement targeted strategies that expand charging networks, improve accessibility in underserved regions, and increase public awareness of ZEV benefits. By doing so, the state can move closer to achieving its ambitious clean energy and transportation goals, reducing its reliance on gasoline-powered vehicles and fostering a more sustainable future.

****

*Figure VI CA County ZEV Registrations 2021-2023*

1. **EV Charging Infrastructure Comparison: CA vs NY**

In Figure VII, a geospatial heat map illustrates the total number of EV charging stations across U.S. states as of 2024, focusing on comparing California and New York. Based on our analysis and the Alternative\_Fueling\_Stations dataset, this visualisation highlights the critical role infrastructure plays in supporting electric vehicle (EV) adoption, particularly for Zero Emission Vehicles (ZEVs).

The data reveals a significant disparity: California has 19,024 charging stations, while New York has only 5,155. This stark contrast underscores the importance of charging infrastructure availability in promoting ZEV adoption.

California’s extensive network of Level 2 and Level 3 fast chargers reflects the state’s commitment to reducing range anxiety and meeting the growing demand for battery electric vehicles (BEVs). These investments have positioned California as a leader in EV sales and ZEV registrations, fostering consumer confidence and encouraging widespread adoption. In contrast, New York’s relatively limited charging infrastructure presents a potential barrier to ZEV adoption despite the state’s high population density and strong interest in reducing emissions. The scarcity of charging stations, particularly outside major urban areas, may deter potential buyers who prioritize convenient access to charging facilities. Expanding New York’s charging network could be pivotal in addressing these challenges, reducing consumer hesitancy, and accelerating the transition to ZEVs, thereby better aligning the state with its clean energy objectives.

A map of the united states with numbers and a green line

Description automatically generated

*Figure VII Electric Charging Stations Across USA in 2024*

1. **Conclusion**

The present research study has explored the barriers to adopting electric vehicles (EVs) in urban environments like California and New York City. We identified key trends that influence consumer preferences and adoption patterns by analysing state-specific EV registration data and charging infrastructure. These insights clarify the factors that drive or hinder the transition to sustainable urban transportation.

The findings of this study reveal that California leads in both EV sales and charging infrastructure, supported by its extensive network of charging stations and progressive policies. However, California continues to face barriers in suburban and rural areas with limited charging infrastructure. Expanding access to Level 3 fast chargers is critical to overcoming range anxiety and encouraging further adoption of fully electric vehicles. Additionally, transitioning consumers from hybrid models to fully electric vehicles poses another hurdle, as preferences for hybrids persist in some regions.

New York also demonstrates impressive EV adoption rates but faces significant challenges. Not only is there a need for charging infrastructure in its suburban and rural areas, but it also faces the additional difficulty of its densely populated urban areas like New York City, where space to install charging infrastructure is limited. For this reason, the availability of Level 3 fast chargers becomes exceptionally crucial. However, the high installation cost of Level 3 chargers—approximately $50,000 per unit—poses a significant financial barrier compared to Level 2 chargers, which cost around $4,500 to install [8][9]. These costs underscore the importance of public investment and incentives to accelerate the development of fast-charging infrastructure in urban areas.

Moreover, the slow transition of traditional automotive manufacturers to electric vehicle production exacerbates the challenge. While some companies have announced ambitious goals, their progress in producing and marketing affordable and widely available EVs could be faster. More urgency is needed to ensure the adoption of EVs, especially in states like New York, where charging infrastructure gaps further deter potential buyers [9]. Addressing these barriers requires a coordinated approach involving governments, private companies, and local communities to accelerate the development of EV infrastructure and increase the availability of high-speed charging stations.

Tableau's visualisations offer a comprehensive view of EV distribution, sales trends, and infrastructure gaps across the analyzed states. These tools illustrate the progress made and pinpoint areas requiring further attention and investment. The results underscore the importance of targeted policies, such as expanding charging networks in underserved areas, increasing consumer awareness, and providing incentives to accelerate the adoption of battery electric vehicles (BEVs).

As the demand for EVs grows, this study contributes valuable, data-driven insights that can guide policymakers, city planners, and environmental advocates in making informed decisions. These findings emphasize the need for coordinated efforts to foster a smoother transition to sustainable transportation systems, reduce carbon emissions, and enhance the quality of urban living. By identifying actionable strategies and implementing practical solutions, this research aims to contribute to developing a low-emission future for cities worldwide.

**References**

[1] NASA. (2024, July). *Carbon Dioxide Concentration | NASA Global Climate Change*. Climate Change: Vital Signs of the Planet; NASA. <https://climate.nasa.gov/vital-signs/carbon-dioxide/?intent=121>

[2] NASA. (2023). *Global Surface Temperature | NASA Global Climate Change*. Climate Change: Vital Signs of the Planet; NASA. [Global Temperature | Vital Signs – Climate Change: Vital Signs of the Planet](https://climate.nasa.gov/vital-signs/global-temperature/?intent=121)

[3] *Global EV transition goals major car manufacturers*. (n.d.). EV Markets Reports. [Global EV transition goals major car manufacturers | EVMarketsReports.com](https://evmarketsreports.com/global-ev-transition-goals-major-car-manufacturers/)

[4] Commission, C. E. (n.d.). *ZEV and Infrastructure Stats Data*. California Energy Commission. [ZEV and Infrastructure Stats Data | California Energy Commission](https://www.energy.ca.gov/files/zev-and-infrastructure-stats-data)

[5] *Alternative Fueling Stations*. (2020). Bts.gov. [Alternative Fueling Stations | Geospatial at the Bureau of Transportation Statistics](https://geodata.bts.gov/datasets/usdot::alternative-fueling-stations/about)

[6] Atlas EV Hub. (2024). *State EV registration data*. [State EV Registration Data – Atlas EV Hub](https://www.atlasevhub.com/materials/state-ev-registration-data/)

[7] U.S. Department of Energy. (2022). *Alternative Fuels Data Center: Vehicle Registration Counts by State*. Afdc.energy.gov. [Alternative Fuels Data Center: Vehicle Registration Counts by State](https://afdc.energy.gov/vehicle-registration)

[8] DiNello, S. (2022, August 11). *What Does a Level 3 Charger Cost?* Future Energy. <https://futureenergy.com/ev-charging/what-does-a-level-3-charger-cost/>

[9] Daniel C. Vock (2024). *New York’s slow progress moving to EVs highlights obstacles cities face.* Route Fifty. <https://www.route-fifty.com/infrastructure/2024/05/new-yorks-slow-progress-moving-evs-highlights-obstacles-cities-face/396852/>