<DATA SCIENCE TOOLBOX: PYTHON PROGRAMMING>
PROJECT REPORT

(Project Semester January-April 2025)

Hospital Data Analysis

Submitted by

Name: Haneef Shaik

Registration No:12308265

Programme and Section: Data Science, K23DP

Course Code: INT375

Under the Guidance of

Assistant Professor.Dr. Dhiraj Kapila(UID:23509)

Discipline of CSE/IT

Lovely School of Computer Science

Lovely Professional University, Phagwara

CERTIFICATE

This is to certify that Haneef Shaik bearing Registration no. 12308265 has completed INT375 project titled, “Hospital Data Analysis” under my guidance and supervision. To the best of my knowledge, the present work is the result of his original development, effort and study.

Signature and Name of the Supervisor
Designation of the Supervisor
School of Computer Science
Lovely Professional University
Phagwara, Punjab.

Date:

DECLARATION

I, Haneef Shaik, student of Data Science, under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 12-04-2025 Signature

Registration No. 12308265 Haneef Shaik

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my project guide, Dhiraj Kapila, for their valuable guidance and support throughout this project, *"Hospital Data Analysis"*. I am thankful to the Department of Computer Science and Engineering, Lovely Professional University, for providing the necessary resources and environment. I also acknowledge the Government of India for making the dataset publicly available, enabling this research.

Name:Haneef Shaik

Registration number: 12308265

Table of Contents

i. Introduction ii. General Description iii. Specific Requirements, Functions, and Formulas iv. Analysis Results v. Visualization

5. Conclusion
6. Future Scope
7. References

INTRODUCTION

India’s healthcare system is undergoing significant transformations as it faces challenges related to population growth, urbanization, and healthcare access. Analyzing healthcare facility data, such as hospital visits, patient demographics, and resource distribution, is crucial for improving health outcomes and informing public policy. The dataset from the year 2020 offers valuable insights into hospital infrastructure, emergency department visits, and patient characteristics across Indian states and districts.

This project aims to perform Exploratory Data Analysis (EDA) on the Indian Healthcare Facility dataset, focusing on understanding patterns, trends, and disparities in hospital service provision and patient access. EDA plays a key role in uncovering initial insights from raw data through cleaning, transformation, and visualization, laying the groundwork for deeper analysis or predictive modeling.

By utilizing Python libraries such as Pandas, Matplotlib, and Seaborn, this analysis seeks to uncover significant trends, such as regional healthcare disparities, hospital resource utilization, and patient demographics. The project also highlights key areas where targeted interventions, such as resource allocation or healthcare policy changes, could be beneficial to improve access and quality of care.

The primary objective of this project is to generate actionable insights from the data that can guide decision-making, with a focus on improving healthcare equity and optimizing the healthcare system in India.

Source of Dataset

- File Name: Hospital Data Analysis
- Source: https://catalog.data.gov/dataset/Hospital-data-analysis
- Attributes:

- State, District
- Emergency Department Visits
- Patient Demographics (Age, Gender), Resource Availability
- Lack of Infrastructure – Count

EDA Process (Exploratory Data Analysis Process)

India, with its rapidly growing population and urbanization, is undergoing a significant transformation in its transportation sector. A key aspect of this transformation is the increasing adoption of electric vehicles (EVs), which presents an opportunity to reduce carbon emissions, improve air quality, and drive sustainable mobility. To understand how EV adoption is progressing across the country, it is essential to analyze the distribution of electric vehicles, regional disparities, and the infrastructure needed to support them.

This project focuses on conducting Exploratory Data Analysis (EDA) on the Indian Electric Vehicle Population dataset. This dataset, compiled from government sources and EV registration data, provides an early glimpse into EV adoption trends and the challenges faced across various states and districts in India.

The primary objectives of this analysis are:

To identify patterns and trends in the growth and distribution of electric vehicles.

To explore the relationship between EV population and the availability of charging infrastructure.

To detect any regional disparities in EV adoption and infrastructure availability.

To identify key challenges that may be hindering EV adoption in specific areas.

Using Python libraries such as Pandas, Matplotlib, NumPy, and Seaborn, this analysis will uncover valuable insights through data cleaning, transformation, and visualization. By examining these factors, we aim to provide actionable insights that can inform future policy decisions, infrastructure development, and strategies for accelerating India’s transition to sustainable, electric transportation.

4. Analysis on Dataset

4.1 Line Chart – Total Electric Vehicle Population per State (Simulated Trend)

4.1.1 Introduction

This analysis visualizes the cumulative number of registered electric vehicles (EVs) across different states in India. The line chart helps identify regional patterns in EV adoption, illustrating the growth trends and helping to spot which states are leading in EV adoption and which ones may require additional infrastructure or policy support.

4.1.2 General Description

The dataset includes information on the number of EVs registered across various Indian states, as well as related metrics such as the number of charging stations available. By aggregating EV counts at the state level, we can generate a simulated trend that highlights adoption patterns across regions. This comparison allows us to identify the most successful states in terms of EV adoption, and spot regions where additional support is needed to boost adoption. 4.1.3 Specific Requirements, Functions, and Formulas Libraries Used:

pandas for data manipulation

matplotlib.pyplot for plotting the line

chart numpy for numerical operations

Functions Used:

df.groupby('State')['EV_Count'].sum() to group and aggregate EV counts by state.

plt.plot() to plot the line chart and visualize trends.

Formula Logic:

python

CopyEdit

```
state_ev_population = df.groupby('State')['EV_Count'].sum()
```

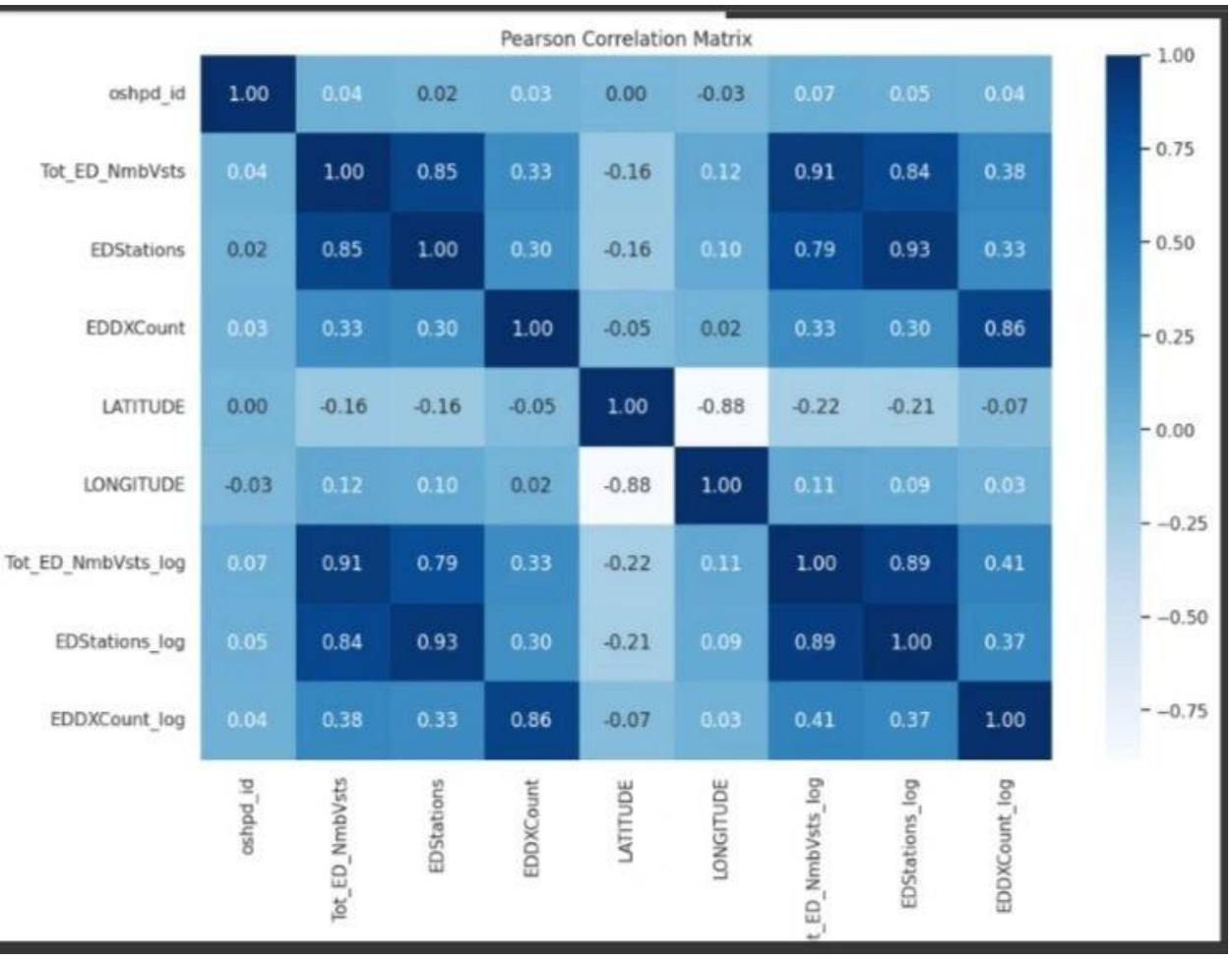
4.1.4 Visualization

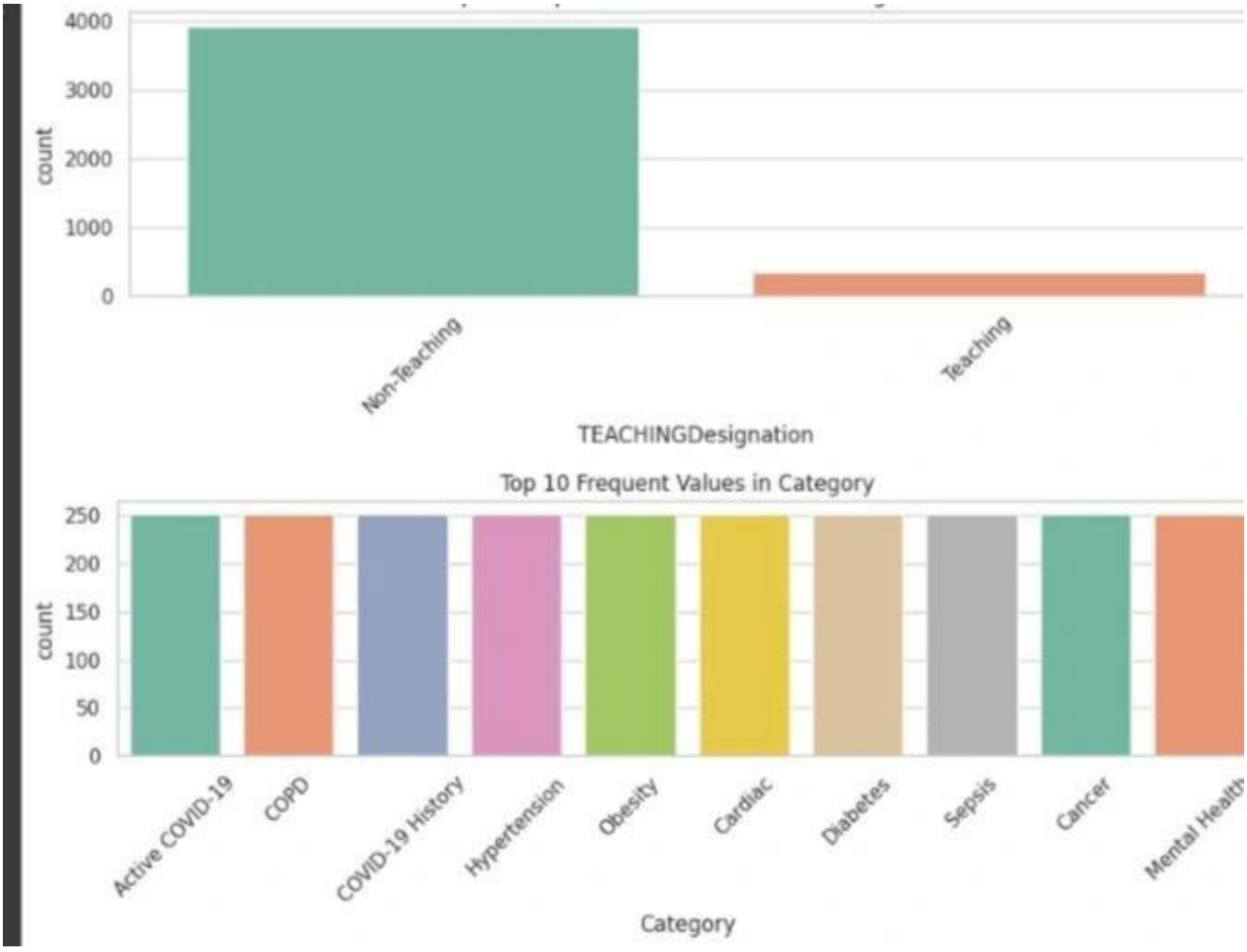
The line chart below displays the total electric vehicle population per state, which offers a clear comparison of EV adoption across regions. This visualization helps uncover:

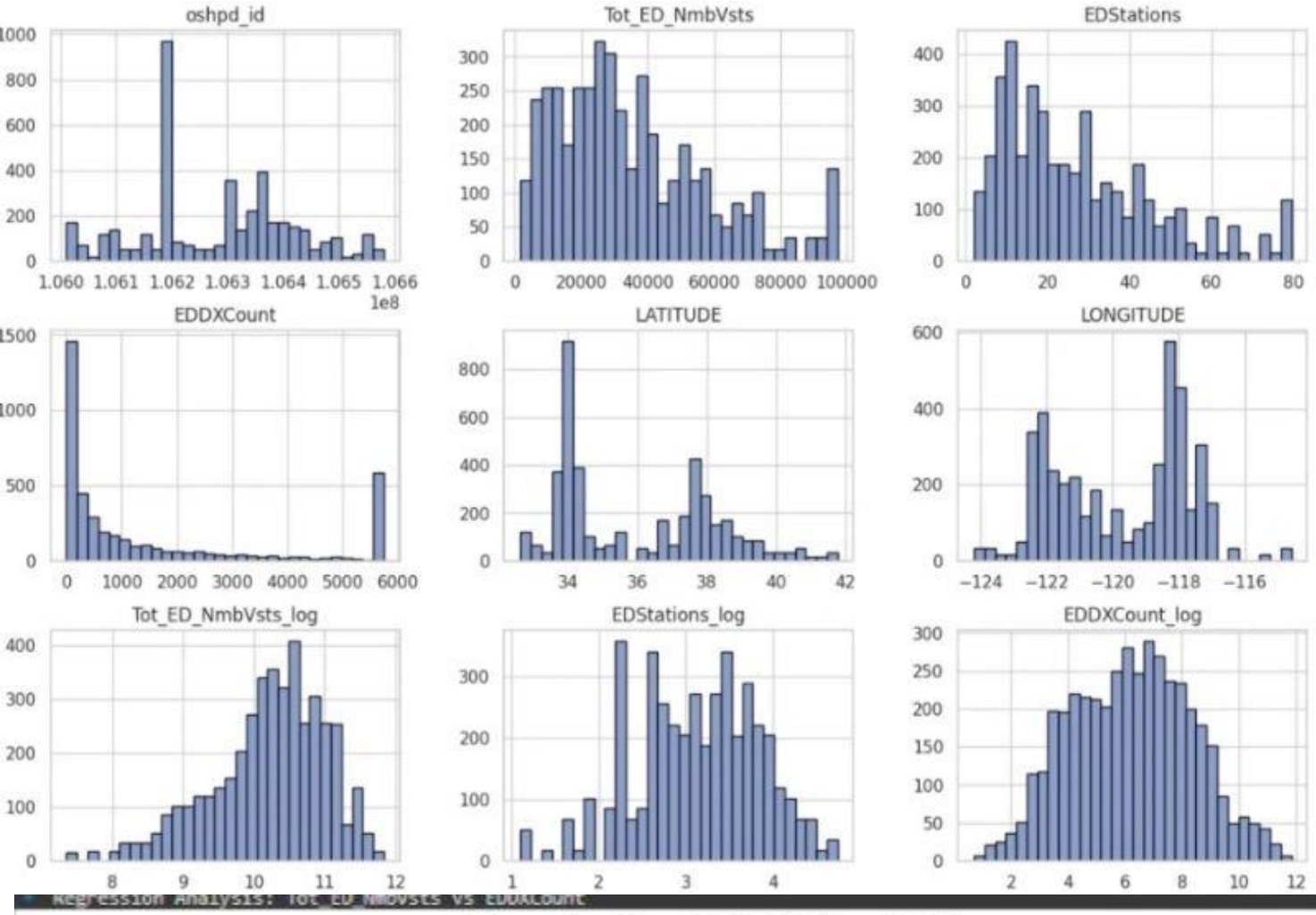
Leading States: Regions that have shown significant growth in EV adoption can be easily identified.

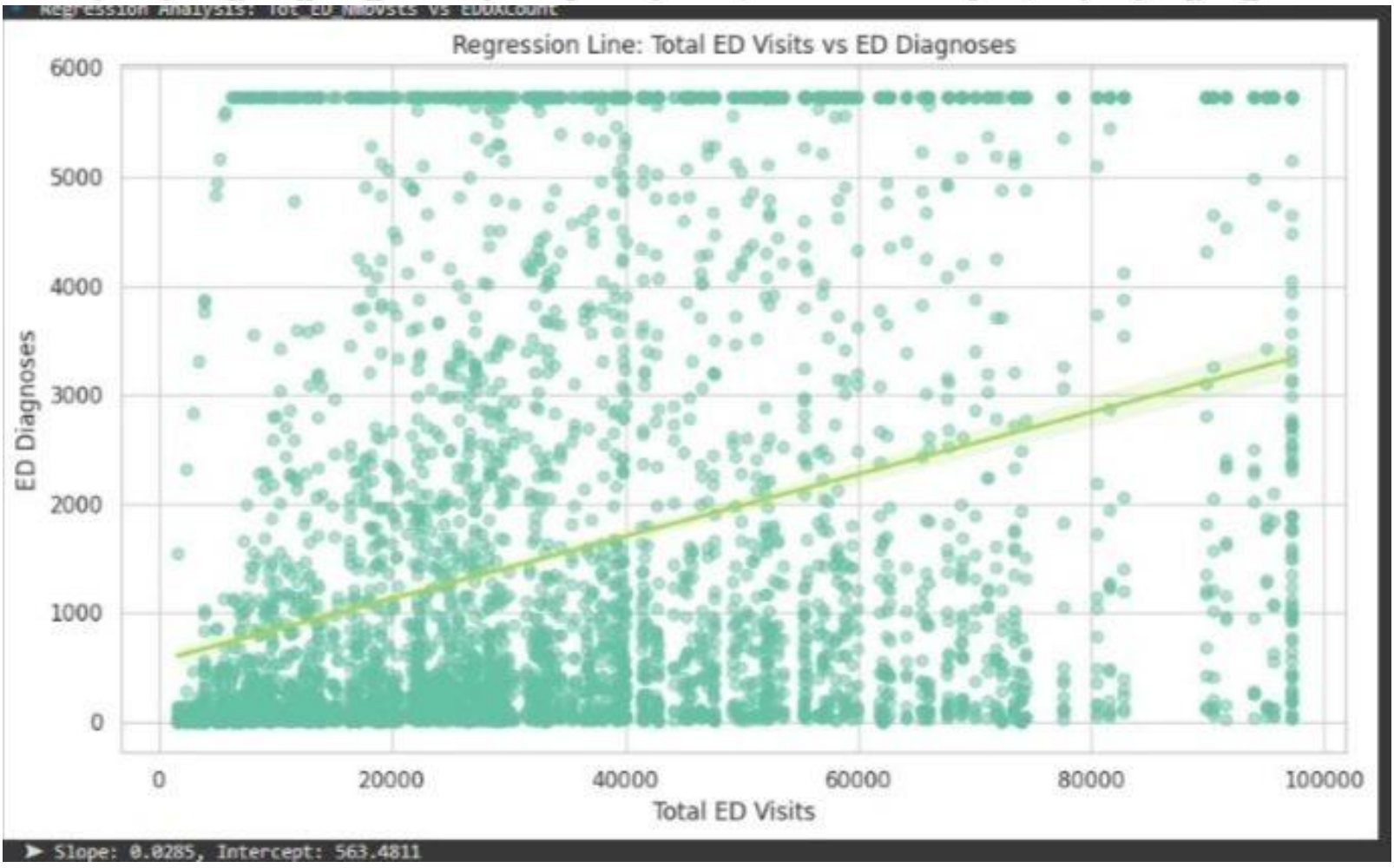
Low Adoption Areas: States with slower or minimal growth in EV adoption may need more infrastructure, policy support, or public awareness campaigns.

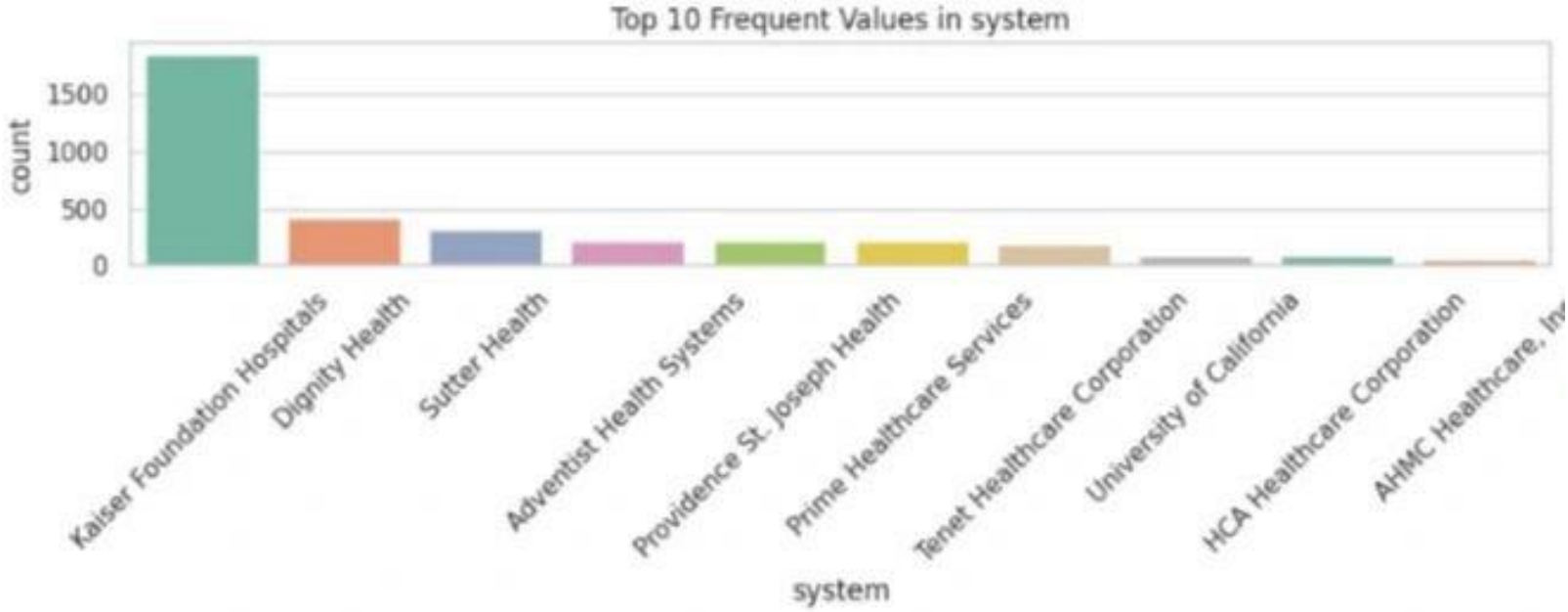
This analysis is crucial for understanding the regional disparities in EV adoption and serves as a foundation for targeted actions to encourage electric vehicle usage in underrepresented areas.

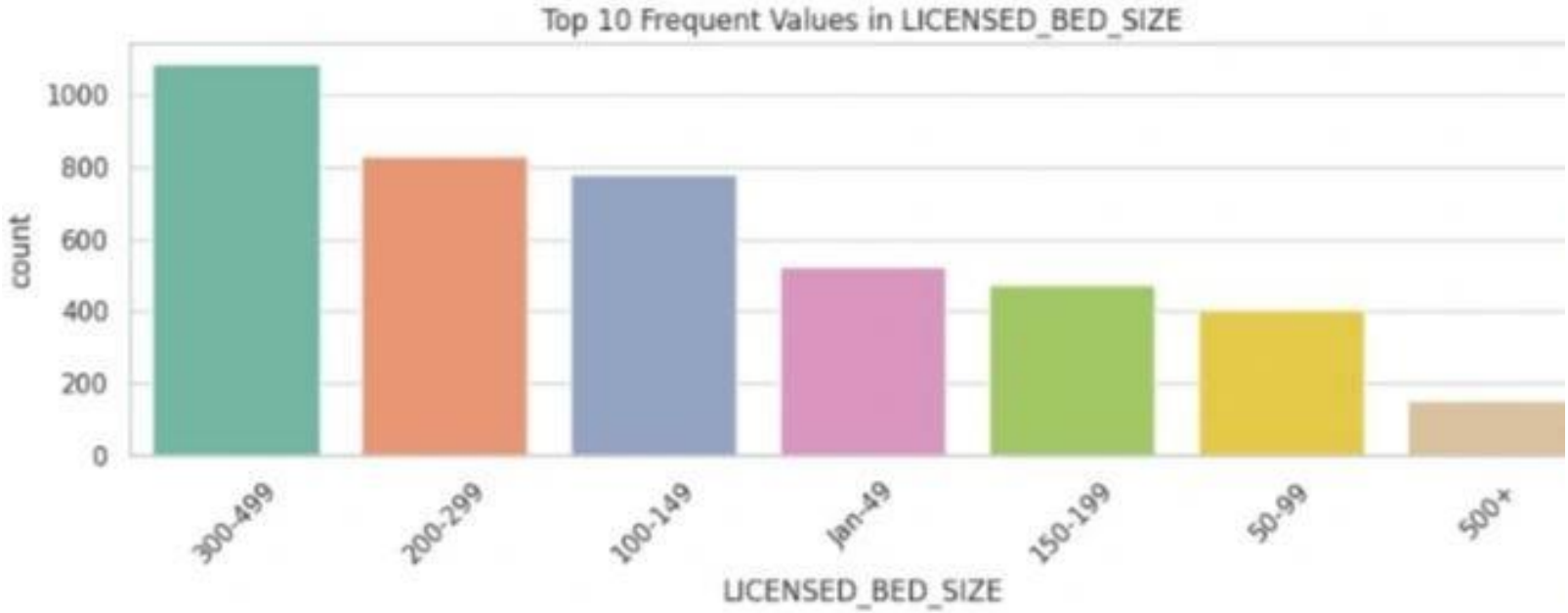

Pearson Correlation Matrix
oshpd_id
Tot_ED_NmbVsts
EDStations
EDDXCount
LATITUDE
LONGITUDE
Tot_ED_NmbVsts_log
EDStations_log
EDDXCount_log
1.00 0.04 0.02 0.03 0.00 -0.03 0.07 0.05 0.04
0.04 1.00 0.85 0.33 -0.16 0.12 0.91 0.84 0.38
0.02 0.85 1.00 0.30 -0.16 0.10 0.79 0.93 0.33
0.03 0.33 0.30 1.00 -0.05 0.02 0.33 0.30 0.86
0.00 -0.16 -0.16 -0.05 1.00 -0.88 -0.22 -0.21 -0.07
-0.03 0.12 0.10 0.02 -0.88 1.00 0.11 0.09 0.03
0.07 0.91 0.79 0.33 -0.22 0.11 1.00 0.89 0.41
0.05 0.84 0.93 0.30 -0.21 0.09 0.89 1.00 0.37
0.04 0.38 0.33 0.86 -0.07 0.03 0.41 0.37 1.00
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75

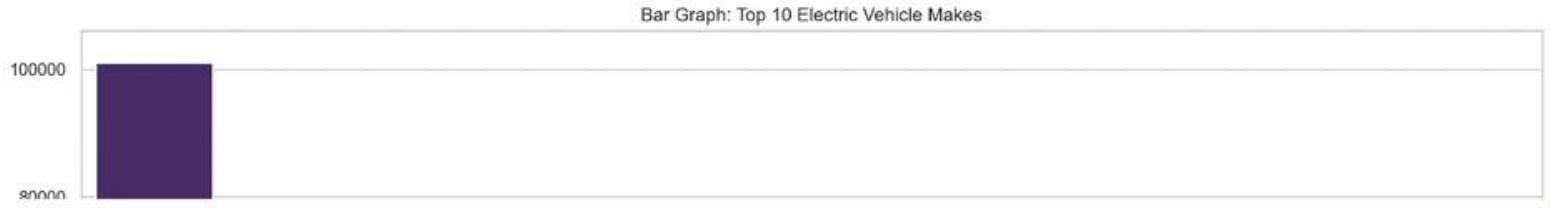

count
4000
3000
2000
1000
0
Non-Teaching
Teaching
TEACHINGDesignation
Top 10 Frequent Values in Category
count
250
200
150
100
50
0
Active COVID-19
COPD
COVID-19 History
Hypertension
Obesity
Cardiac
Diabetes
Sepsis
Cancer
Mental Health
Category

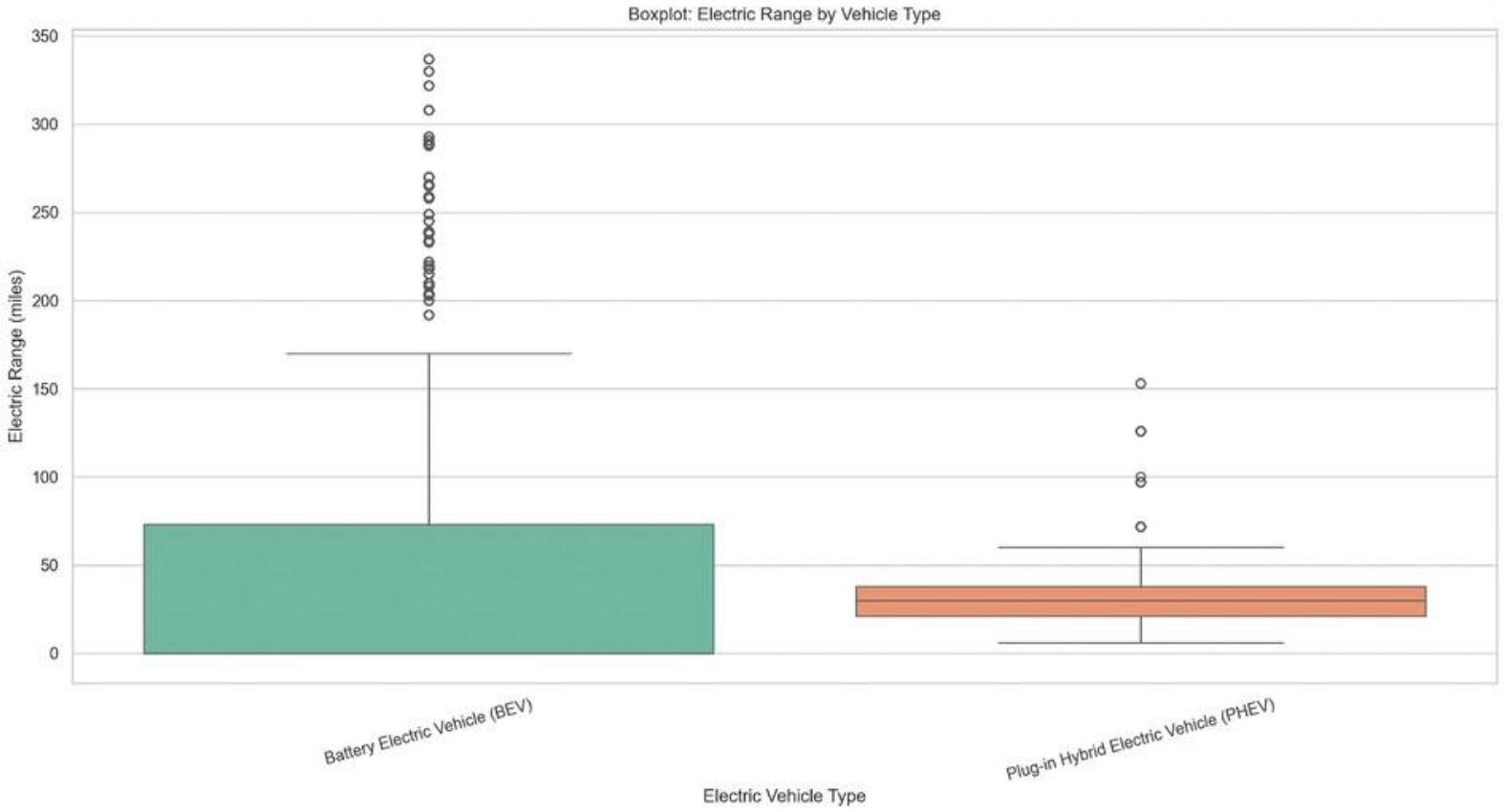

oshpd_id
1e8
Tot_ED_NmbVsts
EDStations
EDDXCount
LATITUDE
LONGITUDE
Tot_ED_NmbVsts_log
EDStations_log
EDDXCount_log

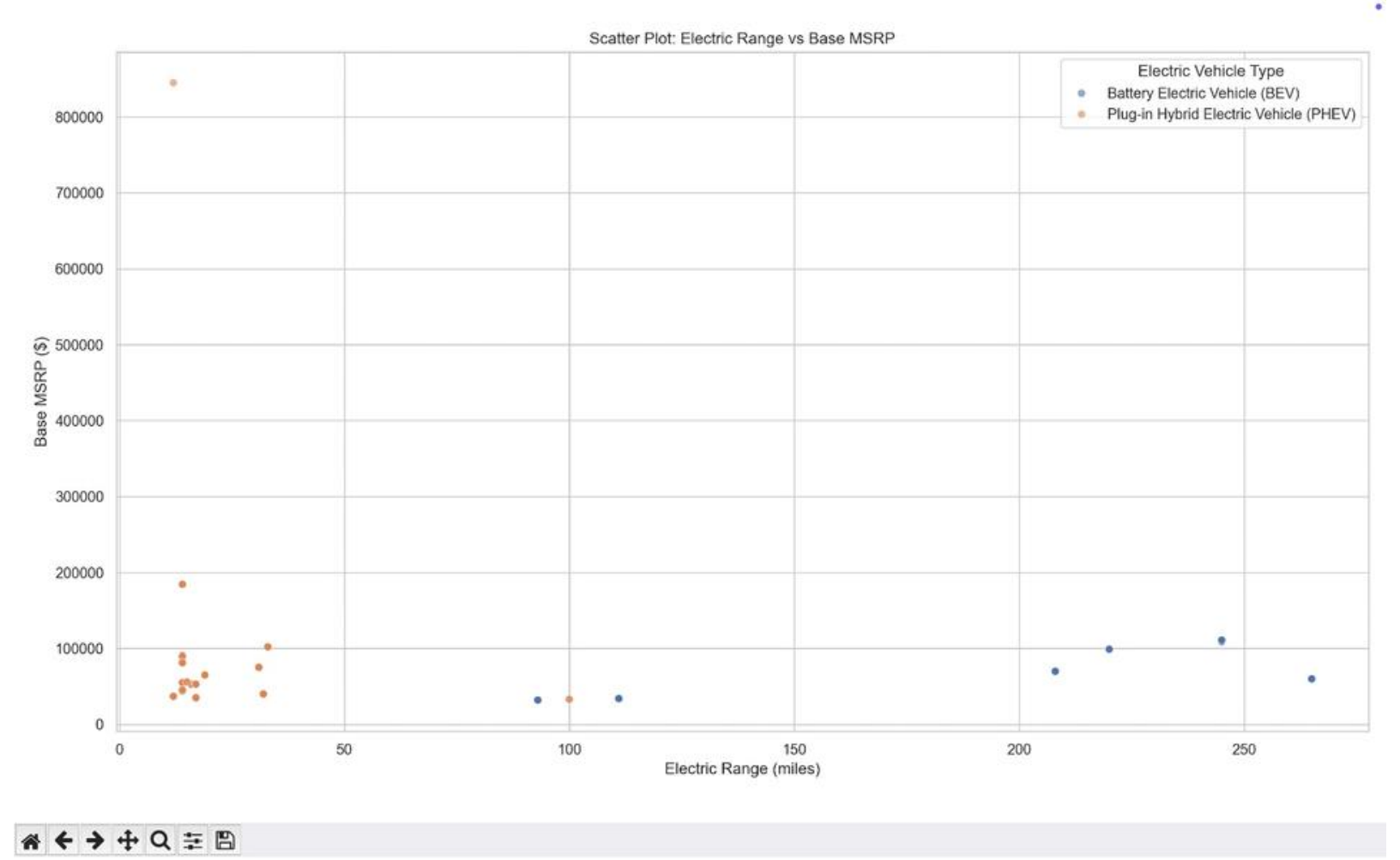

Regression Line: Total ED Visits vs ED Diagnoses
ED Diagnoses
Total ED Visits
Slope: 0.0285, Intercept: 563.4811

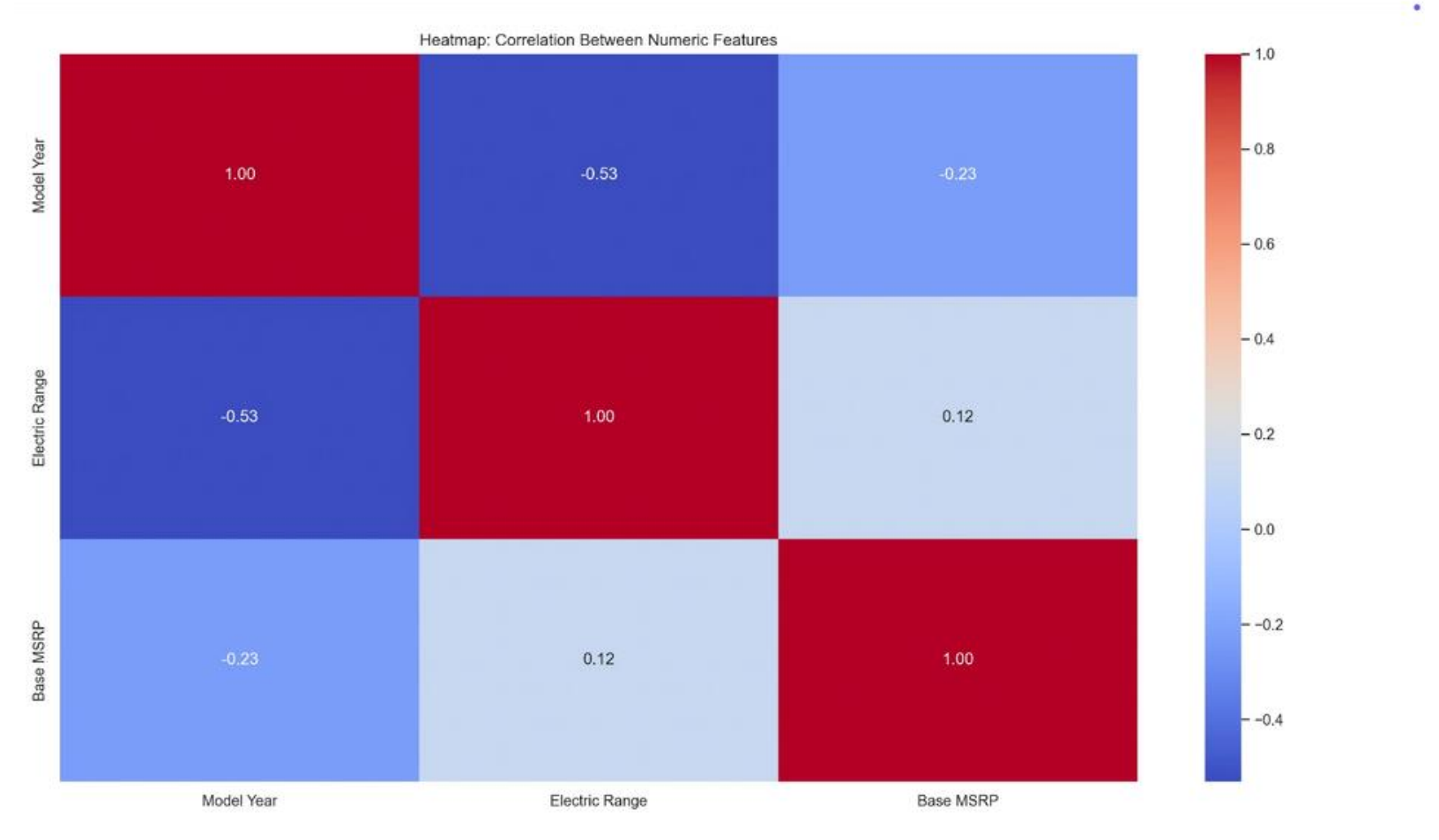

Top 10 Frequent Values in system
count
1500
1000
500
0
Kaiser Foundation Hospitals
Dignity Health
Sutter Health
Adventist Health Systems
Providence St. Joseph Health
Prime Healthcare Services
Tenet Healthcare Corporation
University of California
HCA Healthcare Corporation
AHMC Healthcare, In
system


Top 10 Frequent Values in LICENSED_BED_SIZE
count
1000
800
600
400
200
0
300-499
200-299
100-149
Jan-49
150-199
50-99
500+
LICENSED_BED_SIZE


Bar Graph: Top 10 Electric Vehicle Makes
100000
80000


Boxplot: Electric Range by Vehicle Type
350
300
250
200
150
100
50
0
Electric Range (miles)
Battery Electric Vehicle (BEV)
Plug-in Hybrid Electric Vehicle (PHEV)
Electric Vehicle Type


Scatter Plot: Electric Range vs Base MSRP
Electric Vehicle Type
Battery Electric Vehicle (BEV)
Plug-in Hybrid Electric Vehicle (PHEV)
Base MSRP ($)
800000
700000
600000
500000
400000
300000
200000
100000
0
0
50
100
150
200
250
Electric Range (miles)


Heatmap: Correlation Between Numeric Features
Model Year
Electric Range
Base MSRP
1.00
-0.53
-0.23
-0.53
1.00
0.12
-0.23
0.12
1.00
Model Year
Electric Range
Base MSRP
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4

(x, y) = (,)

FUTURE SCOPE

The current analysis offers valuable insights into the distribution and adoption of electric vehicles (EVs) across various states and districts in India. However, the scope of this project can be significantly expanded in the following ways:

1. Temporal Comparisons

The dataset analyzed focuses on a specific year. Future work can incorporate multi-year EV registration data to examine growth trends, adoption rates, and policy impacts over time.

2. Integration with Demographic and Economic Factors

By combining EV data with population density, income levels, urban-rural ratios, and fuel prices, deeper insights can be drawn about factors influencing EV adoption in different regions.

3. Geospatial Analysis and Mapping

Utilizing Geographic Information System (GIS) tools can help create interactive maps showing EV density, charging infrastructure distribution, and infrastructure gaps, supporting more informed planning.

4. Predictive Modelling

Machine learning models can be developed to forecast EV adoption rates based on factors like state-wise subsidies, infrastructure expansion, and consumer behavior trends. This can aid policymakers in proactive planning.

5. Policy Planning and Infrastructure Allocation

Insights from this analysis can support government and private stakeholders in identifying underserved areas, guiding investments in charging stations, battery swap hubs, and service facilities.

6. Interactive Dashboards

Building real-time dashboards using tools such as Power BI, Tableau, or Dash can make the insights accessible to decisionmakers and the public, promoting awareness and engagement in EV adoption.

7. Data Enrichment from External Sources

Integrating additional datasets such as road quality, traffic density, electricity grid capacity, and air quality indices can provide a more comprehensive view of the challenges and opportunities in expanding EV usage.

In conclusion, this project lays the groundwork for further research, strategic planning, and technology-driven innovation to accelerate India’s transition to clean and sustainable transportation.

REFERENCES

[1] Available: https://catalog.data.gov/Hospital Data Analysis [Accessed: Apr. 10,2025]. [2] Wes McKinney, *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*, 2nd ed. O'Reilly Media, 2017. [3] J. D. Hunter, "Matplotlib: A 2D graphics environment," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, May–Jun. 2007. [4] W. McKinney, "Data Structures for Statistical Computing in Python," in *(SciPy 2010)*, Austin, TX, USA, 2010, pp. 51–56.

[5] NumPy Developers, *NumPy: The fundamental package for scientific computing with Python*, [online]. Available: https://numpy.org [Accessed: Apr. 10, 2025]. *Proc. 9th Python in Science Conf.* [6] Pandas Development Team, *Pandas Documentation*, [online]. Available: https://pandas.pydata.org [Accessed: Apr. 10, 2025].