Shailendra Kumar C S

Clustering-Based Analysis of Electric Vehicle Adoption in Washington State

Abstract

This research delves into the spread and nature of electric vehicles (EVs) in Washington State based on a database of more than 230,000 records. By preprocessing the data, statistical analysis, visualization, and clustering techniques, including K-Means and hierarchical clustering, the research discovers trends in EV adoption. Essential metrics like Base MSRP, Electric Range, and legislative district data were preprocessed, analyzed, and applied for clustering. The research indicates regional trends, car choices, and growth tendencies, assisting policymakers and companies to comprehend the forces of EV penetration.

Objectives

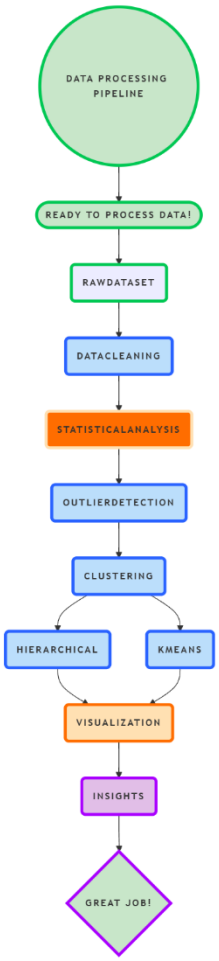
- Clean and preprocess the EV dataset for reliable analysis.
- Conduct exploratory data analysis to understand key features.
- Detect and address outliers and missing data.
- Apply clustering techniques to group vehicles based on meaningful attributes.
- Visualize and interpret trends in EV adoption across regions and time.

Proposed Methodology

1. Data Cleaning: Deleted missing or zero values in key columns (e.g., Electric Range, Base MSRP).
2. Imputation: Performed mean and median imputation for missing values.
3. Outlier Detection: Utilized IQR and boxplots for outlier detection.
4. Statistical Analysis: Conducted skewness, kurtosis, and correlation analysis.
5. Clustering: Performed both hierarchical and K-Means clustering to cluster EVs.
6. Visualization: Utilized heatmaps, histograms, line graphs, and pie charts to obtain insights.

Shailendra Kumar C S

Architecture (Flow Diagram)



Module Description (Clustering)

1. Hierarchical Clustering:
    - Sampled 1000 rows and utilized ward method for dendrogram visualization.
    - Result: Optimal number of clusters approximated to be 4.
2. K-Means Clustering:
    - Utilized on Base MSRP and Electric Range.
    - Elbow method proposed 4 optimal clusters.
    - Clustered data was plotted using a scatter plot displaying clear clusters.

Shailendra Kumar C S

Dataset

Name: Electric Vehicle Population Data
Source: Public dataset from Washington State
Size: ~235,000 rows × 17 columns

Important Fields:

- Model Year, Make, Model
- Electric Vehicle Type
- Electric Range
- Base MSRP
- County, City, Legislative District

Results

- Top EV Counties: King County dominates EV registrations.
- Popular Models: Tesla Model Y and Model 3 are the most registered.
- Growth Trend: EV growth has sharply increased post-2015, peaking in 2023.
- EV Type Distribution: Uniform distribution of BEV and PHEV across states.
- Mileage Leaders: Jaguar and Tesla lead in electric range.
- Cluster Analysis: Vehicles grouped into 4 clusters based on MSRP and range:
    - Low cost, low range
    - Mid-range vehicles
    - High-range, high-cost
    - Luxury outliers

Conclusion

The clustering method yielded actionable information on the EV market in Washington State. The increasing trend in EV registrations indicates a movement toward green transportation. Clustering assisted in segmenting vehicles for improved targeting by policymakers and manufacturers. The project illustrates how machine learning and EDA can inform decisions in the dynamic EV market.

GitHub Link

https://github.com/RadheKrishna-RadheRadhe/EV_Data-Analysis/tree/main

References

- https://data.wa.gov/Transportation/Electric-Vehicle-Population-Data/f6w7-q2d2/about_data
- https://catalog.data.gov/dataset/electric-vehicle-population-data
- https://sdgresources.relx.com/electric-vehicles-evs#:~:text=EVs%20also%20play%20a%20role,which%20has%20significant%20health%20impacts
- https://sdgresources.relx.com/goal-7-affordable-and-clean-energy
- https://sdgresources.relx.com/goal-11-sustainable-cities-and-communities
- https://sdgresources.relx.com/goal-13-climate-action