THE UNIVERSITY OF HONG KONG
FACULTY OF EDUCATION

Bachelor of Arts and Sciences in Social Data Science [BASc(SDS)]
Assignment Cover Sheet

Year of Study: ~~3rd~~ / 5th (please delete as appropriate)

Course Code:	BSDS4999	Student Name:	Chan Wang Tik
Course Title:	Project	Student No:	3036101769
Course Teacher:	Dr. Man Fung, Lo	Student's email:	u3610176@connect.hku.hk
Due Date:	18/5/2025		

Assignment Topic: Investigating the Social and Economic Impacts of Electric Vehicle Adoption in Support of Sustainable Development Goals

---

Student Declaration:

This assignment is entirely my own work except where I have duly acknowledged other sources in the text and listed those sources at the end of the assignment; I have not previously submitted this work to this University or any other institution for a degree, diploma or other qualification; I understand that I may be orally examined on my submission.

I have read the information on the website "What is Plagiarism" (https://tl.hku.hk/plagiarism/), which gives details of plagiarism, and I have observed all the requirements set out on the website. I also acknowledge that HKU has a strict policy on student plagiarism, and I have read and understood the Policy on Student Plagiarism in Undergraduate and Taught Postgraduate Curricula.

In addition (please tick all that apply):

- [x] I have submitted this assignment to *Turnitin*, have reviewed the Originality Report, and revised my assignment as necessary to ensure that my work is free of plagiarism.

- [ ] This assignment includes data from my classroom, such as video and audio recordings of my lessons and copies of student work. The data is used solely for the purpose of completing this assignment. The school and the students are not identified in this assignment. Where necessary I have used pseudonyms. I will destroy the data upon formal assessment of my assignment and my course grade is endorsed by the Faculty of Education.

- [x] **I have used Generative Artificial Intelligence (GenAI) to complete this assignment.** I acknowledge that I have done so in alignment with the assignment requirements set by the course instructor. I have provided proper attribution to the GenAI tool(s) used in the assignment. I understand that the use of GenAI is subject to the terms and conditions of the platform and that it is my responsibility to ensure that I use it appropriately and ethically, while maintaining Academic Integrity.
    - I also acknowledge that I have received support from GenAI for a portion of this assignment. Specifically, the following portion of my work was completed with the support of GenAI: Final Project Report
    - I attribute the use of GenAI in my work. GenAI Tool(s) Used: Grok
    - I have thoroughly reviewed and edited the contribution of GenAI Tool(s) to ensure accuracy, authenticity, and integrity of the information included in the assignment.
- [ ] **I have *not* used GenAI to complete this assignment.** I declare that I have completed this assignment without the use of any GenAI tools or assistance. The work submitted is entirely my own except where I have duly acknowledged other sources in the text.

Note. If students have any questions about academic integrity or the use or citation of GenAI tools in coursework tasks or assignments, they should speak with their course instructor.

Signed: Date: 21/5/2025

Title: Investigating the Social and Economic Impacts of Electric Vehicle Adoption in Support of Sustainable Development Goals

Introduction

As part of the Bachelor of Arts and Sciences in Social Data Science (BASc(SDS)) program at The University of Hong Kong, the final year project (BSDS4999) requires students to apply social data science methods to address research questions aligned with the United Nations Sustainable Development Goals (SDGs). This project investigates global electric vehicle (EV) adoption patterns using three complementary approaches: descriptive statistics to highlight disparities, k-means clustering to identify regional archetypes, and multiple linear regression (MLR) to predict EV sales share. The analysis aligns with SDG 7 (Affordable and Clean Energy) by promoting sustainable transport and SDG 9 (Industry, Innovation, and Infrastructure) by examining industrial capacity and infrastructure's role in EV adoption. The report including background, literature review, methodology, findings, discussion, and limitations.

Background

The transition to electric vehicles (EVs) is critical for reducing greenhouse gas emissions and advancing clean energy (SDG 7) while fostering innovation in transportation infrastructure and industrial capacity (SDG 9). Global EV adoption varies significantly due to economic, social, infrastructural, and industrial factors. For example, Austria's 26% EV sales share in 2023 contrasts sharply with Brazil's 3%, reflecting disparities in wealth, infrastructure, and policy support. Understanding these patterns and their drivers is essential for designing targeted interventions to accelerate sustainable development.

This study employs:

1. Descriptive Statistics: Summarizes EV sales share, stock share, sales volumes, and growth rates across regions and vehicle types, highlighting disparities.

2. K-Means Clustering: Groups regions based on EV adoption metrics (sales share, stock), socioeconomic factors (GDP per capita, urban population), and industrial

capacity (manufacturing employment).

3. Multiple Linear Regression: Predicts EV sales share using the same features to quantify their impact.

The analysis uses data from the IEA Global EV Data 2024, World Bank GDP per Capita, World Bank Urban Population, and SDG Indicator 9.2.2 (Manufacturing Employment) datasets. By combining descriptive insights, archetypes, and predictive modeling, the study informs policies for sustainable transport and industrialization.

Literature Review

EV adoption is driven by multiple factors, as evidenced by prior research:

1. Economic Factors: Higher GDP per capita enables EV adoption through purchasing power and incentives (Hardman et al., 2017). Norway's 93% EV sales share in 2023 reflects strong economic support (IEA, 2024).

2. Urbanization: Urban areas with dense populations and infrastructure facilitate EV adoption (Li et al., 2019). Cities like Amsterdam (Netherlands, 35% sales share) lead in charging networks.

3. Industrial Capacity: Strong manufacturing sectors, as in China (38% sales share), reduce EV costs, aligning with SDG 9 (BloombergNEF, 2023).

4. Infrastructure Availability: Charging points are critical, but data gaps in emerging economies like Brazil (3% sales share) pose challenges (Hardman et al., 2018).

5. Analytical Approaches: Clustering segments regions by adoption patterns (Zhang et al., 2020), while regression quantifies drivers like income and infrastructure (Sierzchula et al., 2014). Descriptive statistics provide foundational insights into disparities (IEA, 2024).

This project addresses a gap in global studies by integrating descriptive statistics, clustering, and regression, aligning with SDGs 7 and 9. Descriptive analysis highlights disparities, clustering reveals structural patterns, and regression identifies actionable drivers.

Methodology

Data Sources

1. IEA Global EV Data 2024: EV sales share, stock share, sales volumes, and charging points (12,654 rows, 8 columns).

2. World Bank GDP per Capita: GDP per capita (USD) for 1960–2024, reshaped for 2023 (261 rows, 70 columns).

3. World Bank Urban Population: Urban population percentage for 1960–2024, reshaped for 2023 (261 rows, 70 columns).

4. SDG Indicator 9.2.2: Manufacturing employment percentage for 2023 (4,815 rows, 8 columns).

Data Preprocessing

- Region Mapping: Standardized names using region_map (e.g., 'United States' to 'USA').
- Filtering: Focused on 2023 historical data, selecting 30 iea_regions (26 after merging for clustering/regression).
- Features (Clustering/Regression):
    - ev_sales_share (%): Target for regression, predictor for clustering.
    - ev_stock (vehicles): Sum of BEV, PHEV, FCEV stock.
    - GDP per capita (USD): Economic indicator.
    - Urban population (% of total population): Urbanization level.
    - manufacturing_employment (%): Industrial capacity.
    - charging_points: Excluded due to missing data.
- Descriptive Features:
    - ev_sales_share (%) for cars, buses, trucks, vans.
    - ev_stock_share (%) for cars, buses, trucks, vans.
    - ev_sales (vehicles) by powertrain (BEV, PHEV, FCEV).
    - Average Annual Growth Rate (AAGR) for car sales share (2010–2023).
- Missing Data:

- charging_points: 100% missing for 2023, excluded.
- GDP per capita, Urban population, manufacturing_employment: 2 missing values each, imputed using median.
- Descriptive data: Missing values for non-car vehicle types (e.g., buses in Australia) noted but not imputed.

- Outlier Handling (Clustering): Winsorized ev_stock and GDP per capita at 5%.
- Scaling: Standardized features for clustering and regression.

Descriptive Analysis

- Metrics:
  - EV sales share (%) for cars and other vehicle types in 2023.
  - EV stock share (%) for cars and other vehicle types in 2023.
  - EV sales (vehicles) by region, year, and powertrain (sample).
  - AAGR for car sales share (2010–2023).
- Visualizations (using Python's Matplotlib/Seaborn):
  - Bar Chart: EV sales share for cars in 2023 by region, highlighting disparities (e.g., Norway 93% vs. Brazil 3%).
  - Heatmap: EV sales share across vehicle types (cars, buses, trucks, vans) for 2023, showing data availability.
  - Line Plot: EV sales growth for select regions (e.g., Australia, Austria) from 2011–2023.
  - Bar Chart: Top 10 regions by AAGR (2010–2023), emphasizing rapid growth (e.g., UAE 249.3%).

K-Means Clustering

- Algorithm: K-means with n_init=10.
- Optimal k: Selected k=6 via elbow and silhouette score plots.
- Implementation: Clustered 26 regions based on ev_sales_share, ev_stock, GDP per

capita, Urban population, manufacturing_employment.

Multiple Linear Regression

- Model: Linear regression predicting ev_sales_share.
- Features: ev_stock, GDP per capita, Urban population, manufacturing_employment.
- Data Split: 80% training, 20% testing (random_state=42).
- Evaluation: Mean Squared Error (MSE), R^2.
- Feature Importance: Regression coefficients.

Visualization

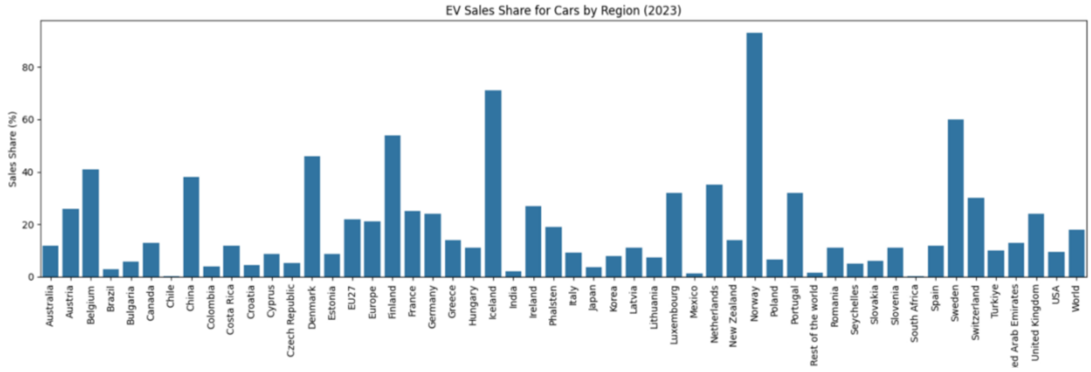
- Descriptive: Bar charts, heatmap, line plot for disparities and trends.
- Clustering: Elbow, silhouette, scatter (ev_sales_share vs. GDP per capita), pairplot.
- Regression: Actual vs. predicted scatter, feature importance bar plot, residuals histogram.

Results

Descriptive Analysis Results

1. EV Sales Share (%) for Cars in 2023:
   - Range: Norway led with 93%, followed by Iceland (71%), Sweden (60%), and Finland (54%). South Africa (0.29%), Chile (0.31%), and Mexico (1.3%) had the lowest shares.
   - Disparities: Austria's 26% vs. Brazil's 3% highlights economic and infrastructural gaps.

Fig.1 EV Sales Share for Cars by Region (2023)



High-income, urbanized regions (e.g., Norway, Austria) dominate, while emerging economies (e.g., Brazil, India) lag behind.

Fig.2 Global EV Sales for Cars by Powertrain (2010-2023)

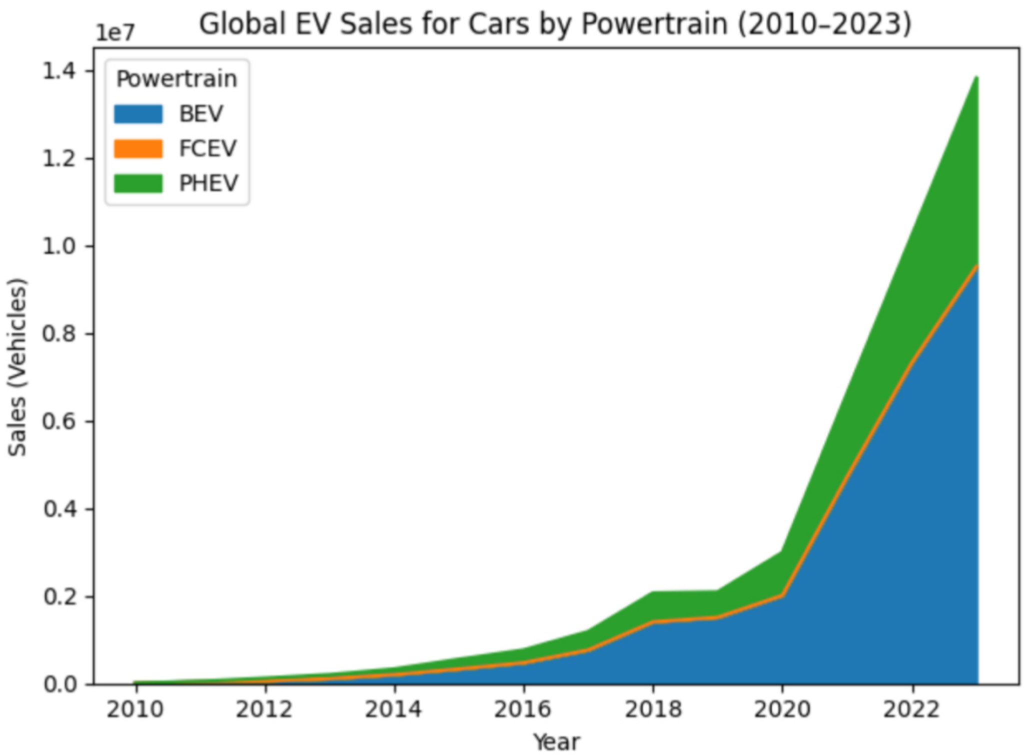


Fig.3 EV Sales Share Trends for Cars (Top 10 Regions, 2010-2023)

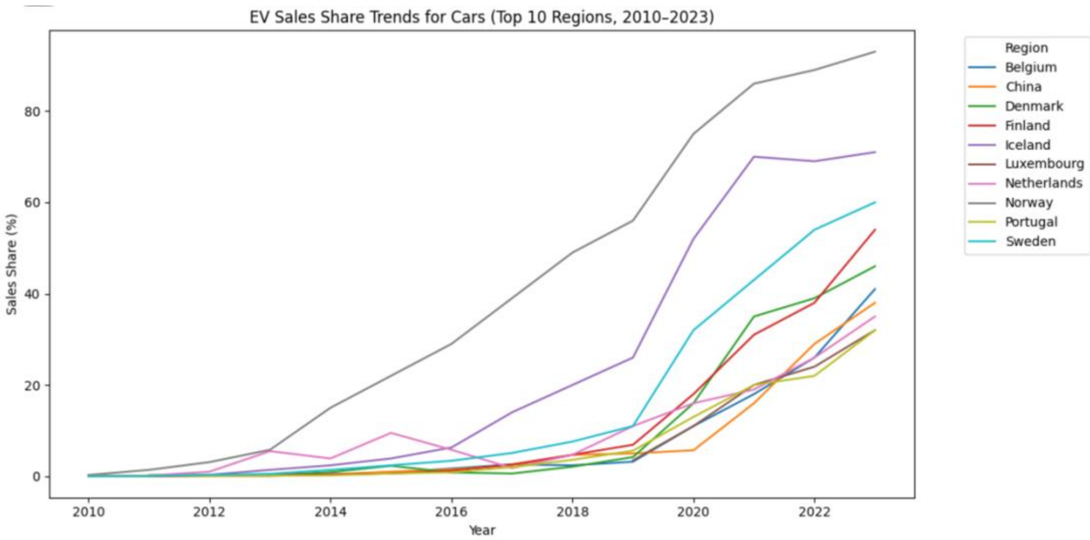
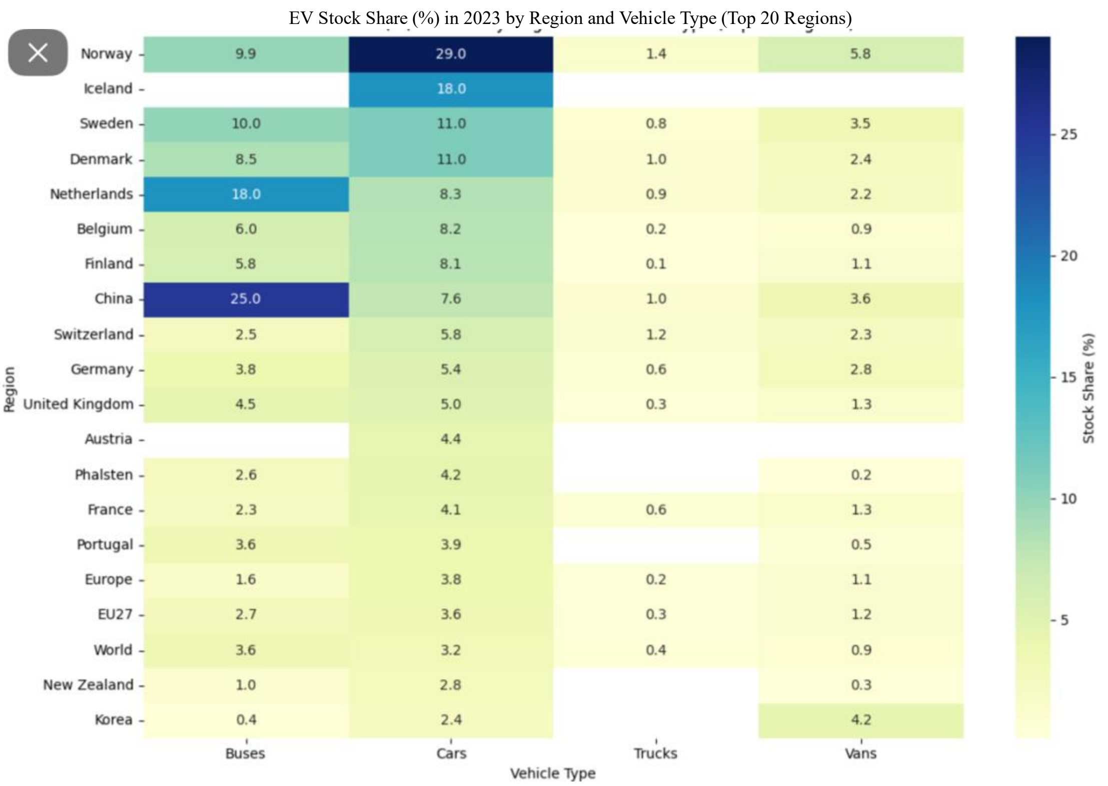


Fig.4 EV Stock Share (%) in 2023 by Region and Vehicle Type (Top 20 Regions)



K-Means Clustering

Fig.5 Cluster Summary (Mean Values) & Regions per Cluster

```
Cluster Summary (Mean Values):
   cluster  ev_sales_share      ev_stock  GDP per capita  \
0        0       16.750000  3.660200e+06    68556.319360
1        1       12.850000  1.959590e+05    28084.942763
2        2       38.000000  4.818000e+06    12614.061742
3        3       93.000000  9.002900e+05    87925.094419
4        4        9.110000  3.963029e+05    40706.856455
5        5       44.333333  4.235398e+05    65955.671413

  Urban population (% of total population)  manufacturing_employment
0                                 80.531500                 15.002250
1                                 62.928833                 16.462500
2                                 64.570000                 29.065000
3                                 83.995000                  7.046000
4                                 85.179600                 10.627150
5                                 88.096000                 10.977333

Regions per Cluster:
   cluster                                             region
0        0                                    [Germany, USA]
1        1  [Austria, India, Italy, Mexico, Poland, Portugal]
2        2                                           [China]
3        3                                          [Norway]
4        4  [Australia, Brazil, Canada, Chile, France, Jap...
5        5  [Belgium, Denmark, Finland, Netherlands, Swede...
```

Fig.6 Clusters: EV Sales Share vs GDP per Capita

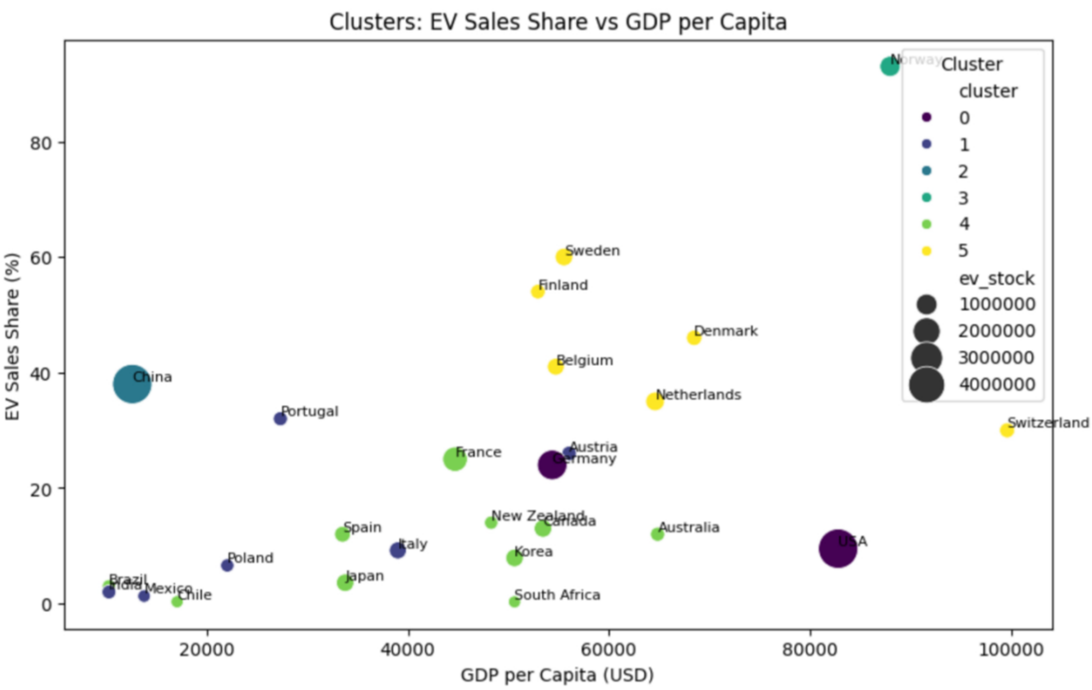


Fig7. Elbow Method for Optimal K

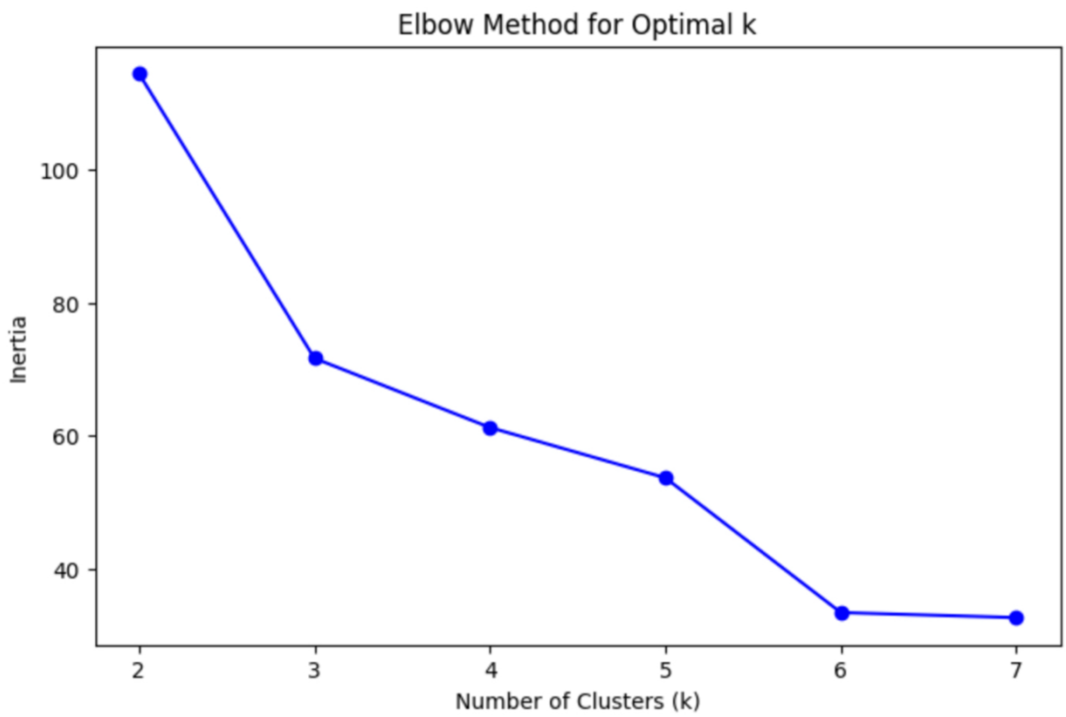


Fig.8 Silhouette Score vs. Number of Clusters

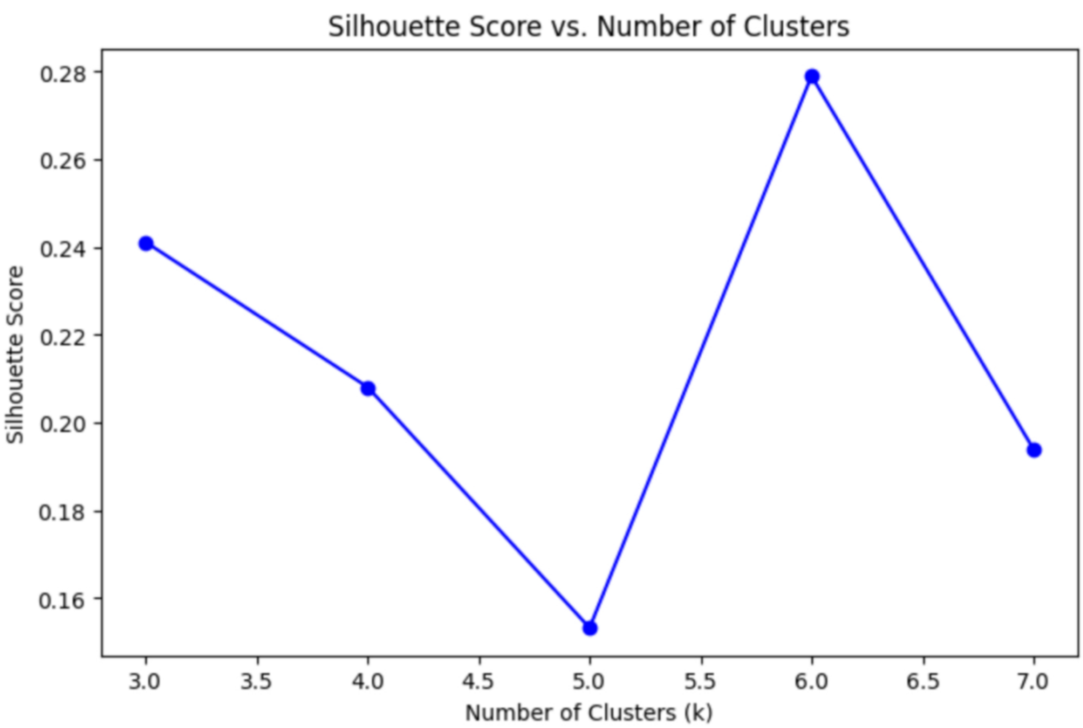
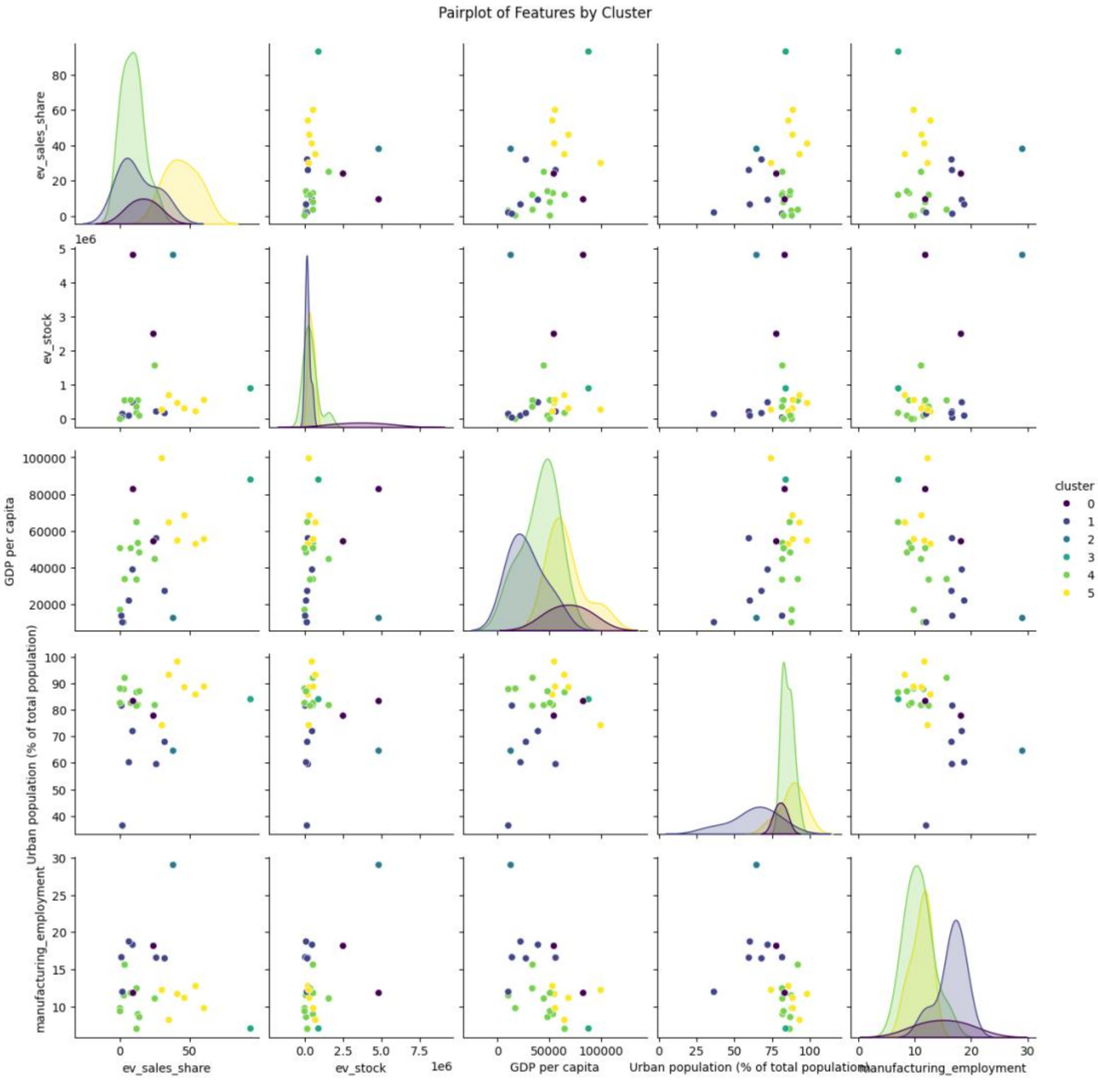


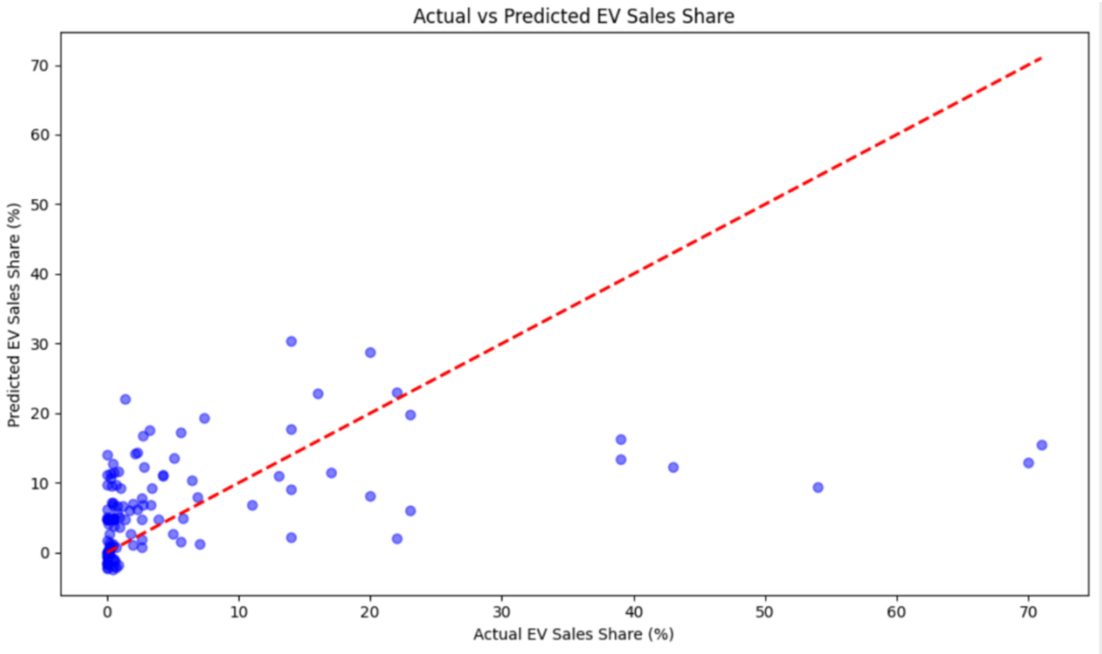
Fig.9 Pairplot of Features by Cluster

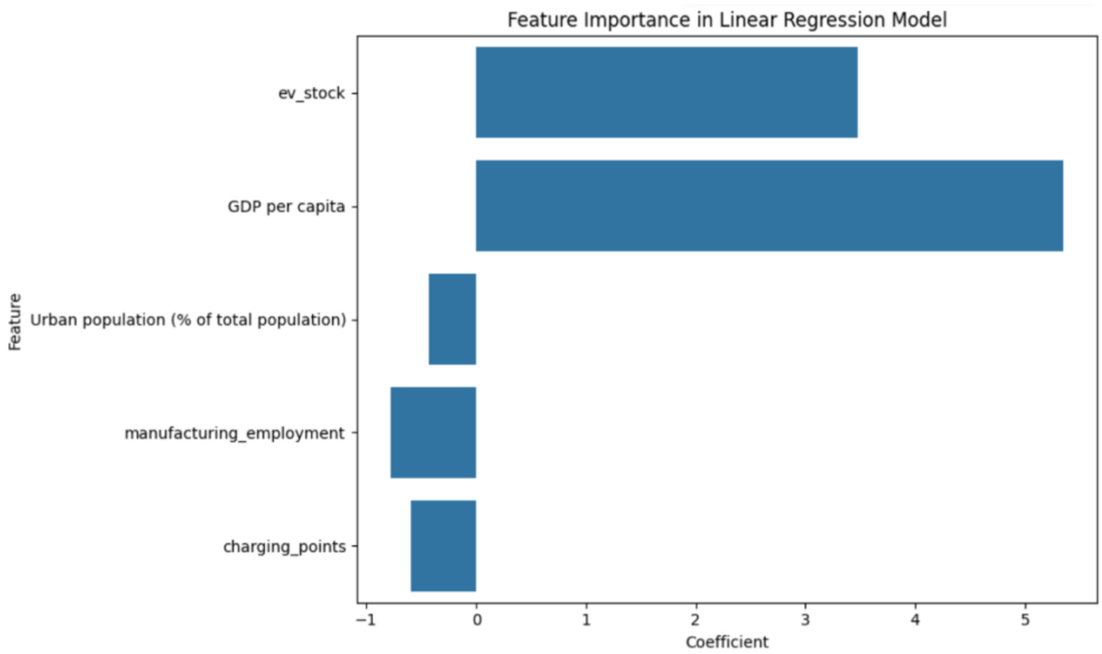


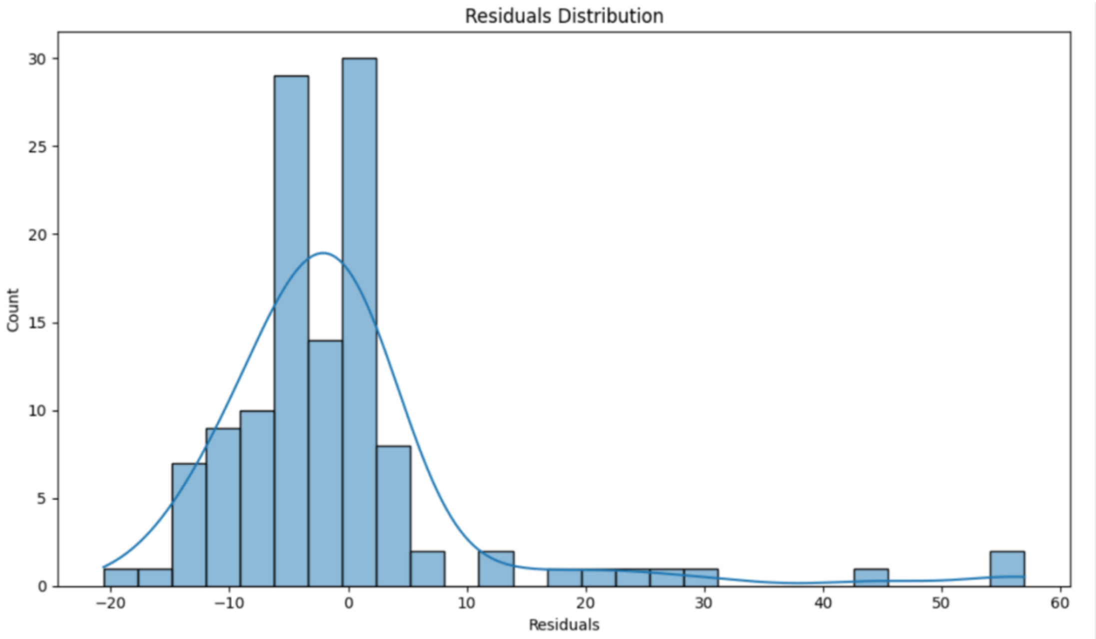
Multiple Linear Regression

```
Model Performance:
Train MSE: 112.5898
Test MSE: 129.9202
Train R²: 0.2356
Test R²: 0.1611
```

```
Feature Importance:
                                     Feature  Coefficient
0                                   ev_stock     3.473750
1                             GDP per capita     5.348962
2  Urban population (% of total population)    -0.426285
3                   manufacturing_employment    -0.777969
4                            charging_points    -0.589487
```


Actual vs Predicted EV Sales Share
Predicted EV Sales Share (%)
Actual EV Sales Share (%)
0
10
20
30
40
50
60
70


Feature Importance in Linear Regression Model
Feature
ev_stock
GDP per capita
Urban population (% of total population)
manufacturing_employment
charging_points
Coefficient
−1
0
1
2
3
4
5



Discussion

Descriptive Analysis Insights

The descriptive analysis reveals stark disparities in EV adoption:

- High Performers: Norway (93% sales share, 29% stock share), Iceland (71% sales share, 18% stock share), and Sweden (60% sales share, 11% stock share) lead, driven by wealth, policies, and infrastructure. Austria's 26% sales share and 4.4% stock share align with this group.
- Low Performers: Brazil (3% sales share, 0.21% stock share), South Africa (0.29% sales share, 0.044% stock share), and Mexico (1.3% sales share, 0.14% stock share) lag due to economic and infrastructural barriers.
- Vehicle Types: Cars dominate adoption, but buses (e.g., China 50%, Switzerland 65%) show progress in public transport, supporting SDG 7. Trucks and vans lag, indicating SDG 9 gaps in commercial vehicle innovation.
- Growth Trends: High AAGR in UAE (249.3%) and Costa Rica (143.8%) reflects emerging markets' potential, while Brazil's 100.7% growth suggests gradual progress.

K-Means Clustering insights

1. Cluster 0: Moderate-Adoption Industrial Leaders (Germany, USA):
    - Characteristics: Moderate ev_sales_share (16.75%), high ev_stock

(3,660,200), high GDP per capita ($68,556). Aligns with SDG 7 but lags leaders.

- Policy interventions recommendation: Expand charging networks, incentivize fleet transitions.

2. Cluster 1: Emerging Moderate-Adoption (Austria, India, Italy, Mexico, Poland, Portugal):

- Characteristics: Moderate ev_sales_share (12.85%), low ev_stock (195,959), moderate GDP per capita ($28,085). Mixed SDG 7/9 progress.

- Policy interventions recommendation: Subsidies, urban charging for Austria; manufacturing support for India.

3. Cluster 2: High-Production Emerging Leader (China):

- Characteristics: High ev_sales_share (38%), very high ev_stock (4,818,000), high manufacturing_employment (29.07%). Strong SDG 9, growing SDG 7.

- Policy interventions recommendation: Scale urban chargers, promote affordable EVs.

4. Cluster 3: Extreme-Adoption Leader (Norway):

- Characteristics: Extreme ev_sales_share (93%), high GDP per capita ($87,925), low manufacturing_employment (7.05). SDG 7/9 exemplar.

- Policy interventions recommendation: Model for others; phase out ICE vehicles.

5. Cluster 4: Low-to-Moderate Adoption Mixed (Australia, Brazil, Canada, etc.):

- Characteristics: Low ev_sales_share (9.11%), moderate ev_stock (396,303), high urbanization (85.18%). Lags in SDG 7/9.

- Policy interventions recommendation: Pilot programs (Brazil), infrastructure investment (Australia).

6. Cluster 5: High-Adoption Urban Leaders (Belgium, Denmark, Finland, Netherlands, Sweden):

- Characteristics: High ev_sales_share (44.33%), high GDP per capita

($65,956), high urbanization (88.10%). Strong SDG 7/9.

- Policy interventions recommendation: Expand fast chargers, mandate EV adoption.

Multiple Linear Regression Insights

The MLR confirms:

- Positive Drivers: ev_stock (3.473750) and GDP per capita (5.348962) drive ev_sales_share, supporting Cluster 3/5's high adoption.

- Unexpected Results: Negative Urban population (-0.426285) and manufacturing_employment (-0.777969) may reflect multicollinearity or small sample size, conflicting with Cluster 5's high urbanization.

- Performance: Moderate R^2 (0.1611 test) indicates predictive power but data limitations.

Societal Implications

The analyses support SDG 7 (clean energy) and SDG 9 (industrial innovation):

- Descriptive: Disparities (e.g., Austria 26% vs. Brazil 3%) highlight infrastructure needs.
- Clustering: Six archetypes guide tailored policies, from Norway's model to Brazil's pilots.
- Regression: Economic drivers emphasize wealth's role, but infrastructure gaps (missing charging data) remain critical.
- Policy Recommendations:
  - For regions in cluster 3 and 5 recommend Mandate EVs, expand chargers.
  - For regions in cluster 0 and 4 recommend Subsidies, infrastructure investment.
  - For regions in cluster 1 and 2 recommend Manufacturing support, urban charging.

Limitations

1. Missing Charging Points Data: Excluded due to 100% missing data for 2023, limiting infrastructure analysis.
2. Small Sample Size: 26 regions reduce clustering/regression robustness.
3. Missing Values: Imputed GDP per capita, Urban population, manufacturing_employment may introduce bias.
4. K-Means Assumptions: Spherical clusters may oversimplify; k=6 increases granularity but risks overfitting.
5. Regression Assumptions: Linear model misses non-linearities; negative coefficients suggest multicollinearity.
6. Descriptive Gaps: Sparse truck/van data limits commercial vehicle insights.
7. Static Analysis: 2023 focus misses temporal trends.

Conclusion

This study integrated descriptive statistics, k-means clustering (k=6), and multiple linear regression to analyze global EV adoption, aligning with SDGs 7 and 9. Descriptive analysis highlighted disparities (e.g., Norway 93% vs. Brazil 3%), clustering identified six archetypes (from Norway's extreme adoption to Brazil's low adoption), and regression confirmed ev_stock and GDP per capita as key drivers. The k=6 clustering provides nuanced insights, separating leaders like Norway and China from mixed regions like Brazil. Future work should incorporate charging data, expand regions, and explore non-linear models.

References

- BloombergNEF. (2023). Electric Vehicle Outlook 2023. BloombergNEF.
- Hardman, S., et al. (2017). A review of consumer preferences of and interactions with electric vehicle charging infrastructure. Transportation Research Part D: Transport and Environment, 54, 11–24.
- Hardman, S., et al. (2018). Driving the market for plug-in vehicles: Understanding consumer preferences. International Journal of Sustainable Transportation, 12(5), 343–356.
- IEA. (2024). Global EV Data 2024. International Energy Agency.

- Li, W., et al. (2019). A review of factors influencing consumer intentions to adopt battery electric vehicles. Renewable and Sustainable Energy Reviews, 78, 318–328.
- Sierzchula, W., et al. (2014). The influence of financial incentives and other socio-economic factors on electric vehicle adoption. Energy Policy, 68, 183–194.
- Zhang, X., et al. (2020). Clustering analysis of electric vehicle adoption in China. Energy Policy, 141, 111414.