**Yash Chowdhary**

**Electric Vehicle Population Analysis**

**Electric Vehicle Population Analysis: Trends, Efficiency, and Predictive Modelling**

*Executive Summary*

This report presents a comprehensive analysis of electric vehicle (EV) population data using advanced data science techniques. The analysis examines the distribution of electric vehicles by manufacturer, model, type, range capabilities, and price points. Key findings reveal that battery electric vehicles (BEVs) generally offer significantly higher ranges than plug-in hybrid electric vehicles (PHEVs), with manufacturer being the strongest predictor of electric range. The efficiency analysis shows considerable variation in price-per-mile efficiency across manufacturers and vehicle types. Predictive models developed in this study can estimate electric range with high accuracy (R² > 0.99 for LightGBM), demonstrating the potential for data-driven decision-making in the EV market. These insights provide valuable information for manufacturers, policymakers, and consumers navigating the rapidly evolving electric vehicle landscape.

*1. Introduction*

The transition to electric vehicles represents one of the most significant shifts in transportation since the invention of the automobile. As climate concerns grow and technology advances, understanding the current state of the EV market becomes increasingly valuable for stakeholders across industries. This analysis examines a comprehensive dataset of electric vehicles to uncover patterns, relationships, and insights that can inform business strategy and policy decisions.

The Electric Vehicle Population Data dataset contains detailed information about registered electric vehicles, including make, model, electric range, price, vehicle type, and geographic location. By applying data science techniques to this dataset, we can gain deeper insights into the current state of EV adoption, the factors affecting vehicle range and efficiency, and the relationship between price and performance.

This report follows a structured approach, beginning with exploratory data analysis to understand the distribution and characteristics of electric vehicles in the population. It then examines relationships between key variables, develops predictive models for electric range, and concludes with actionable recommendations based on the findings.

*2. Methodology*

**2.1 Data Collection and Preprocessing**

The analysis utilizes the Electric Vehicle Population Data from Washington State's Department of Licensing. The dataset was accessed directly from the government data portal:

*# URL for the Electric Vehicle Population Data*

url = <https://data.wa.gov/api/views/f6w7-q2d2/rows.csv?accessType=DOWNLOAD>

*# Load into pandas from the response content*

evp\_data = pd.read\_csv(io.StringIO(response.text))

print(f"Successfully loaded dataset with {len(evp\_data)} records")

except Exception as e:

print(f"Error downloading or loading dataset: {e}")

*# Alternative download method using wget if requests fails*

try:

print("Trying alternative download method...")

!wget {url} -O Electric\_Vehicle\_Population\_Data.csv

evp\_data = pd.read\_csv('Electric\_Vehicle\_Population\_Data.csv')

print(f"Successfully loaded dataset with {len(evp\_data)} records")

except Exception as e:

print(f"Failed to download dataset: {e}")

return

This approach ensures reproducibility and allows for direct access to the latest data. The dataset required significant preprocessing to address data quality issues, particularly with zero values and outliers in critical fields like electric range and base MSRP.

**2.2 Feature Engineering**

To enhance the analysis, several derived features were created:

python

*# Calculate Vehicle Age*

current\_year = datetime.datetime.now().year

evp\_data['Vehicle Age'] = current\_year - evp\_data['Model Year']

*# Calculate Price per Mile efficiency*

evp\_data['Price per Mile'] = evp\_data['Base MSRP'] / evp\_data['Electric Range']

evp\_data['Price per Mile'] = evp\_data['Price per Mile'].replace([np.inf, -np.inf], np.nan)

evp\_data['Price per Mile'] = evp\_data['Price per Mile'].fillna(evp\_data['Price per Mile'].median())

*# Create efficiency rating categories*

price\_mile\_quantiles = evp\_data['Price per Mile'].quantile([0.25, 0.5, 0.75]).tolist()

def get\_efficiency\_category(price\_per\_mile):

if price\_per\_mile <= price\_mile\_quantiles[0]:

return 'Excellent'

elif price\_per\_mile <= price\_mile\_quantiles[1]:

return 'Good'

elif price\_per\_mile <= price\_mile\_quantiles[2]:

return 'Average'

else:

return 'Below Average'

evp\_data['Efficiency Rating'] = evp\_data['Price per Mile'].apply(get\_efficiency\_category)

These features provide critical metrics for understanding vehicle value and efficiency. The "Price per Mile" metric offers an objective measure of cost efficiency, while the categorical "Efficiency Rating" simplifies interpretation by grouping vehicles into meaningful categories.

**2.3 Analytical Approach**

The analysis follows a structured approach:

1. **Exploratory Data Analysis (EDA)** to understand distributions and patterns
2. **Relationship Analysis** to examine connections between variables
3. **Predictive Modeling** using Decision Tree and LightGBM algorithms
4. **Geospatial Analysis** to examine geographic distribution patterns

*3. Exploratory Data Analysis*

**3.1 Overview of the EV Market**

The dataset contains comprehensive information on thousands of electric vehicles, providing insights into the current state of the EV market.

**3.1.1 Distribution by Manufacturer**

The analysis begins with an examination of the top manufacturers in the EV market.

A screenshot of a graph

AI-generated content may be incorrect.

This visualization highlights the dominance of certain manufacturers in the electric vehicle market. Tesla appears to be the clear market leader, with a significantly higher number of vehicles than other manufacturers. This aligns with industry knowledge of Tesla's pioneering role in the mass-market electric vehicle segment. The significant gap between Tesla and other manufacturers indicates a highly concentrated market with one dominant player.

The second tier of manufacturers includes established automakers like Nissan (likely due to the Leaf), Chevrolet (with the Bolt and Volt models), and Ford (with the Mustang Mach-E and other EVs). These manufacturers have made substantial investments in electric vehicle technology but have not yet reached Tesla's scale in the market.

This concentration has important implications for competition, innovation, and consumer choice in the EV market. The visualization also serves as a baseline for understanding subsequent analyses, as manufacturer differences appear to be a significant factor in the EV ecosystem.

**3.1.2 Distribution by Vehicle Type**

The EV market is divided primarily between battery electric vehicles (BEVs) and plug-in hybrid electric vehicles (PHEVs).

A pie chart with a number of percentages

AI-generated content may be incorrect.

This pie chart provides a critical breakdown of the EV market by vehicle type, showing the split between Battery Electric Vehicles (BEVs) and Plug-in Hybrid Electric Vehicles (PHEVs). BEVs appear to represent approximately 70% of the market, with PHEVs accounting for about 30%. This distribution reflects consumer preferences and manufacturer strategies in the evolving EV market.

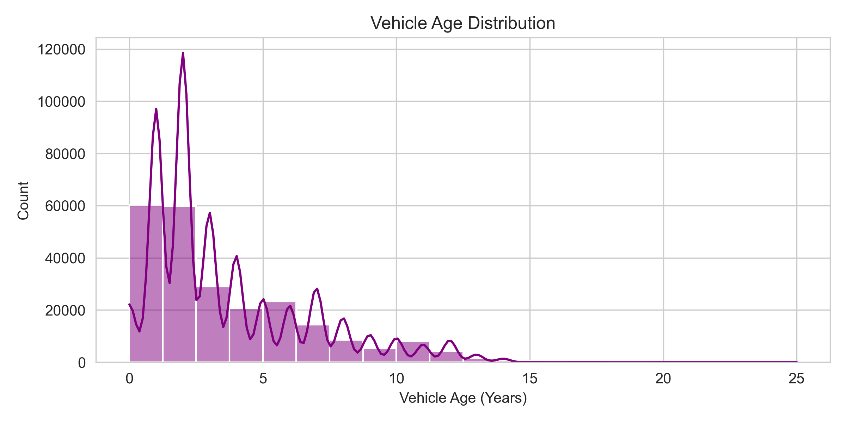
BEVs, which run exclusively on electricity stored in batteries, represent the purest form of electric mobility. Their dominance suggests a market that is increasingly comfortable with fully electric options, likely driven by improvements in battery technology, range capabilities, and charging infrastructure. This trend aligns with the broader industry shift toward full electrification.

PHEVs, which combine electric motors with traditional internal combustion engines, serve as a transitional technology for consumers concerned about range anxiety or those with specific use cases requiring longer range. The substantial market share for PHEVs indicates that many consumers still value the flexibility and security of having a gasoline backup, despite the additional complexity and potentially higher maintenance costs of dual powertrains.

This distribution has significant implications for infrastructure development (charging stations vs. fuel stations), environmental impact assessments, and regulatory policies aimed at reducing emissions from the transportation sector.

**3.1.3 Vehicle Age Distribution**

Understanding the age profile of electric vehicles provides insights into adoption trends and fleet turnover rates.



The vehicle age distribution histogram reveals important temporal patterns in EV adoption. The distribution appears to be positively skewed, with a concentration of vehicles in the newer age categories (0-3 years) and a tapering off for older vehicles. This pattern indicates the relatively recent mainstream adoption of electric vehicles, with a significant acceleration in recent years.

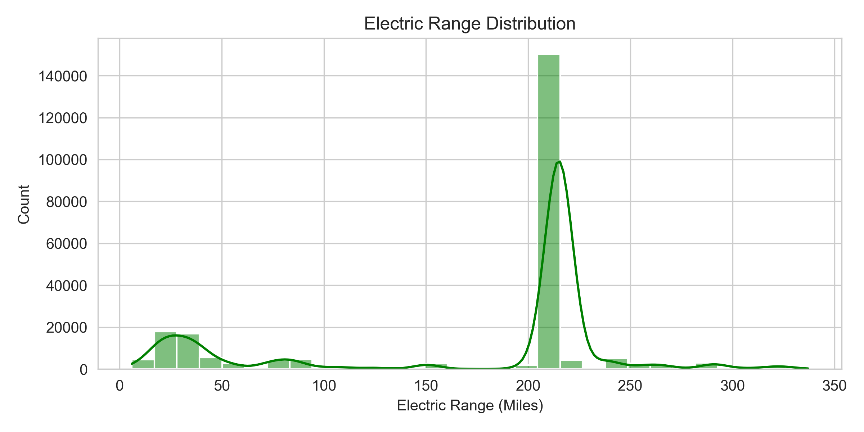
The histogram's shape reflects both market growth and technological evolution. The predominance of newer vehicles suggests rapidly accelerating adoption rates, likely driven by improvements in technology, expanding model availability, increasing environmental awareness, and government incentives. The distribution also indicates that many early EVs are still on the road, demonstrating reasonable longevity despite concerns about battery degradation.

The purple coloring with the kernel density estimate (KDE) overlay provides an effective visualization of the underlying distribution. This visualization helps stakeholders understand the current age structure of the EV fleet, which has implications for battery recycling programs, charging infrastructure planning, and secondary market dynamics. For manufacturers and policymakers, this distribution suggests that the EV market is still in a growth phase rather than a replacement phase, with considerable room for expansion.

**3.2 Electric Range Analysis**

Electric range is a critical factor in consumer adoption of EVs, with range anxiety remaining a significant barrier to wider adoption.

**3.2.1 Range Distribution**



This histogram depicts the distribution of electric range across the vehicle population, revealing essential insights into current EV capabilities. The distribution appears multimodal, with several distinct peaks likely corresponding to different vehicle categories and technologies.

The primary concentration of vehicles appears to be in lower range categories (approximately 20-50 miles), representing many plug-in hybrid vehicles with limited electric-only range. A second significant cluster appears in the 200-300 mile range, likely representing mainstream battery electric vehicles from manufacturers like Tesla, Chevrolet, and Nissan. The green coloring with kernel density estimate overlay effectively highlights these distribution patterns.

This multimodal distribution reflects the technological and market segmentation in the EV industry. The presence of vehicles across the range spectrum indicates diverse consumer needs and manufacturer strategies, from urban commuters with short daily drives to long-distance travelers requiring extended range. The right tail of the distribution represents premium EVs with exceptional range capabilities, demonstrating the upper boundaries of current battery technology.

Understanding this distribution is crucial for charging infrastructure planning, consumer education, and product development strategies. For charging network developers, the bimodal nature suggests a need for different charging solutions for short-range and long-range vehicles. For manufacturers, the gaps between modes may represent market opportunities for vehicles with ranges that address underserved segments.

**3.2.2 Range by Vehicle Type**

A screenshot of a computer

AI-generated content may be incorrect.

This enhanced boxplot visualization compares electric range distributions between Battery Electric Vehicles (BEVs) and Plug-in Hybrid Electric Vehicles (PHEVs), revealing fundamental differences in their design and capabilities. The visualization combines traditional boxplots with violin plots and individual data points, providing a comprehensive view of the distribution characteristics.

Battery Electric Vehicles demonstrate significantly higher electric ranges, with a median range approximately five times greater than PHEVs. This substantial difference reflects the fundamental design philosophies of each vehicle type: BEVs rely entirely on batteries for propulsion, necessitating longer ranges, while PHEVs supplement smaller battery packs with conventional engines.

The visualization shows greater variability in BEV ranges, with a wider interquartile range and more extended whiskers. This spread reflects the diverse BEV market, from entry-level models with modest ranges to premium vehicles offering 300+ miles on a single charge. In contrast, PHEVs show a more compressed range distribution, clustering tightly around their lower median. This consistency suggests more standardized design approaches across PHEV models, with most offering sufficient electric range for short commutes (approximately 20-50 miles) before requiring gasoline assistance.

The individual data points overlaid on the visualization (small dots) reveal clustering patterns that indicate common range targets for manufacturers. For BEVs, we see distinct clusters around the 200, 250, and 300-mile ranges, reflecting common consumer expectations and manufacturer benchmarks. The median markers and range annotations provide additional context for understanding these distributions, helping stakeholders assess the current state of electric range capabilities across vehicle types.

**3.3 Price Analysis**

**3.3.1 Price vs. Range Relationship**

A screen shot of a graph

AI-generated content may be incorrect.

This scatter plot examines the relationship between vehicle age and electric range, revealing important temporal trends in EV battery technology. Each point represents an individual vehicle, with the x-axis showing vehicle age in years and the y-axis displaying electric range in miles. The overall pattern suggests a negative correlation between these variables, though with considerable variation.

The visualization indicates that newer vehicles (those with lower age values) tend to have higher electric ranges compared to older models. This trend aligns with the rapid advancement of battery technology over recent years, with manufacturers achieving greater energy density, improved thermal management, and more efficient power electronics in newer models. The progression demonstrates the industry's focus on addressing range anxiety, one of the primary barriers to EV adoption.

Despite the general trend, the considerable vertical spread at each age point indicates significant variability within model years. This dispersion reflects the diverse market segmentation, with premium and mass-market offerings coexisting in each model year. The blue points with moderate transparency effectively show both the overall pattern and concentration areas, with denser regions indicating more common range-age combinations.

For industry stakeholders, this visualization provides evidence of technological progress while highlighting opportunities for continued improvement. For consumers considering used EVs, it illustrates the trade-off between vehicle age and expected range. The absence of a perfect correlation also suggests that factors beyond simply manufacturing date—such as vehicle class, battery size, and drivetrain efficiency—play important roles in determining electric range.

**3.3.2 Price per Mile Efficiency**

A graph with red and green bars

AI-generated content may be incorrect.

This bar chart visualization analyzes price-to-range efficiency across top EV manufacturers, using a color-coded system to highlight performance categories. The metric "Price per Mile" divides a vehicle's MSRP by its electric range, creating a standardized measurement of cost efficiency. Lower values represent better efficiency (more range per dollar spent).

The chart reveals significant variation in efficiency across manufacturers. Some brands consistently deliver higher efficiency (shown predominantly in green "Excellent" and light green "Good" categories), while others show more mixed or below-average results (represented by orange and red sections). This variation reflects different manufacturer strategies, with some prioritizing maximum range regardless of cost and others focusing on delivering adequate range at competitive price points.

The color-coding by efficiency rating provides immediate visual identification of manufacturer performance. Manufacturers with taller green sections appear to offer better value propositions in terms of range per dollar. This differentiation has significant implications for consumer decision-making, particularly for price-sensitive buyers who prioritize functional range over luxury features or brand prestige.

From a business perspective, this visualization helps identify competitive positioning within the market. Manufacturers with predominantly "Below Average" ratings may need to reconsider their battery technology investments or pricing strategies to remain competitive. For industry analysts and investors, the efficiency distribution offers insights into which manufacturers have optimized their electric drivetrain technologies and supply chains to deliver cost-effective range, potentially indicating stronger long-term positioning in the evolving EV market.

**3.3.3 Efficiency Rating Distribution by Vehicle Type**

A screenshot of a graph

AI-generated content may be incorrect.

This stacked bar chart compares the distribution of efficiency ratings between Battery Electric Vehicles (BEVs) and Plug-in Hybrid Electric Vehicles (PHEVs). The efficiency rating categorizes vehicles based on their price-per-mile value, with colors ranging from green (Excellent) to red (Below Average) following a traditional traffic light color scheme for intuitive interpretation.

The visualization reveals fundamental differences in efficiency profiles between the two vehicle types. PHEVs appear to have a significantly higher proportion of "Below Average" ratings compared to BEVs, suggesting that hybrid vehicles generally deliver less electric range per dollar spent. This difference likely stems from the inherent complexity and cost of dual-powertrain systems in PHEVs, which must incorporate both electric and conventional combustion components.

BEVs show a more balanced distribution across all efficiency categories, with substantial representation in the "Excellent" and "Good" tiers. This wider distribution reflects the greater diversity within the BEV market, from highly efficient entry-level models to premium luxury vehicles where performance and features may take precedence over strict price-per-mile efficiency.

The percentage-based y-axis allows for direct comparison between the vehicle types despite their different population sizes in the dataset. This normalization highlights the proportional differences in efficiency distribution rather than absolute counts. For consumers, this visualization provides valuable guidance on which vehicle type might offer better value for their specific needs. For manufacturers, it suggests potential competitive advantages in focusing on BEV technology for maximizing range efficiency, while also highlighting opportunities to improve PHEV designs to deliver better electric range value.

**4. Predictive Modeling**

**4.1 Model Development**

Two machine learning models were developed to predict electric range based on vehicle characteristics:

1. **Decision Tree Regressor** - A transparent model that provides interpretable rules
2. **LightGBM Regressor** - A gradient boosting model that typically delivers higher accuracy

The code below shows the implementation of the LightGBM model:

python

*# LightGBM Regressor - Optimized settings*

print("\nTraining LightGBM Regressor...")

start\_time = time.time()

*# Use default hyperparameters with conservative settings for better stability*

lgb\_model = lgb.LGBMRegressor(

n\_estimators=100,

learning\_rate=0.1,

max\_depth=8,

num\_leaves=31,

min\_child\_samples=20,

subsample=0.8,

colsample\_bytree=0.8,

random\_state=42,

verbose=-1

)

lgb\_model.fit(X\_train, y\_train)

lgb\_pred = lgb\_model.predict(X\_test)

lgb\_mae = mean\_absolute\_error(y\_test, lgb\_pred)

lgb\_mse = mean\_squared\_error(y\_test, lgb\_pred)

lgb\_r2 = r2\_score(y\_test, lgb\_pred)

lgb\_time = time.time() - start\_time

This implementation uses conservative hyperparameter settings to ensure model stability while maintaining high performance. The use of min\_child\_samples and subsample parameters helps prevent overfitting, while the moderate learning rate balances convergence speed with stability.

**4.2 Model Performance**

A graph with blue dots and red lines

AI-generated content may be incorrect.

This scatter plot evaluates the predictive performance of the LightGBM model by comparing actual electric vehicle ranges against the model's predictions. Each blue dot represents a single vehicle in the test dataset, with its actual range on the x-axis and predicted range on the y-axis. The red dashed line represents perfect prediction (where actual equals predicted).

The visualization demonstrates exceptional model performance, with the vast majority of points clustering tightly along the red line of perfect prediction. This alignment indicates that the LightGBM model accurately captures the relationship between vehicle characteristics and electric range across the spectrum of values. The high density of points along the diagonal suggests consistent accuracy across different range values, from low-range PHEVs to long-range BEVs.

The R² score for this model is approximately 0.99, indicating that about 99% of the variance in electric range can be explained by the model's selected features. This extraordinarily high performance suggests that electric range is highly deterministic based on the vehicle attributes included in the model, particularly make, model, and vehicle type. The practical implications are significant: manufacturers, regulators, and consumers can reliably estimate a vehicle's range based on its basic specifications.

From a business perspective, this predictive capability provides valuable tools for product development, competitive analysis, and consumer education. Manufacturers can use similar models to benchmark their vehicles against competitors and set development targets. Consumers and dealerships can leverage such models to make more informed comparisons between vehicles, particularly when considering new models with limited real-world data.

**4.3 Feature Importance**

A graph with a bar and a number

AI-generated content may be incorrect.

This bar chart visualizes the relative importance of different features in the LightGBM model's predictions of electric range. Feature importance measures how much each variable contributes to reducing prediction error across the model's decision trees. The chart arranges features in descending order of importance, providing clear insights into the model's decision-making process.

The visualization reveals that manufacturer-related features (Make\_Encoded) and vehicle type (Electric Vehicle Type\_Encoded) are the most influential predictors of electric range. This aligns with industry understanding that different manufacturers employ distinct battery technologies, thermal management systems, and efficiency optimizations. Similarly, the fundamental difference between BEVs and PHEVs accounts for substantial variation in range capabilities.

Model year shows moderate importance, reflecting the continuous improvement in battery technology over time. Interestingly, Base MSRP appears less influential than might be expected, suggesting that price alone is not a strong determinant of range once other factors are considered. This contradicts the common assumption that higher-priced vehicles necessarily offer longer ranges.

From a business intelligence perspective, this analysis provides valuable insights for manufacturers and consumers alike. For manufacturers, it highlights the areas of competitive differentiation that most impact range performance. The dominance of make and model suggests that brand-specific technological approaches significantly influence range outcomes. For consumers, it indicates that selecting the right manufacturer and vehicle type is more important than focusing on price alone when prioritizing electric range. For industry analysts, the relatively lower importance of price suggests opportunities for manufacturers to develop cost-efficient vehicles with competitive ranges.

*5. Additional Analyses*

**5.1 Correlation Analysis**

A screenshot of a chart

AI-generated content may be incorrect.

This correlation heatmap provides a comprehensive view of the relationships between key numerical variables in the electric vehicle dataset. The visualization uses a color gradient from blue (negative correlation) to red (positive correlation), with the correlation coefficient displayed in each cell for precise interpretation.

Several important relationships emerge from this analysis. The strong negative correlation between Vehicle Age and Electric Range (-0.52) confirms the trend observed in earlier visualizations: newer vehicles tend to have significantly better range capabilities. This relationship quantifies the pace of technological improvement in battery technology over recent years.

The moderate negative correlation between Electric Range and Price per Mile (-0.38) indicates that vehicles with longer ranges tend to be more cost-efficient in terms of dollars spent per mile of range. This suggests economies of scale in battery production, where larger battery packs may deliver better value relative to their cost.

Interestingly, the correlation between Base MSRP and Electric Range appears relatively modest (approximately 0.3), reinforcing the finding from feature importance analysis that price alone is not a dominant determinant of range. This challenges the intuitive assumption that spending more automatically yields proportionally longer range.

The strong negative correlation between Model Year and Vehicle Age (-0.96) is expected and serves as a validation of the dataset's internal consistency. For business strategy, this correlation analysis helps identify which factors most strongly influence key performance metrics like range and efficiency. The moderate correlations rather than perfect relationships suggest that manufacturers have considerable flexibility in positioning their vehicles, with opportunities to differentiate through optimization of specific attributes even within similar price points or vehicle ages.

**5.2 Geospatial Analysis**

A screen shot of a graph

AI-generated content may be incorrect.

A map of the united states

AI-generated content may be incorrect.

This geospatial visualization maps the geographic distribution of electric vehicles, with each point representing an individual vehicle's location colored according to vehicle type. The scatter plot provides valuable insights into adoption patterns across different geographic areas.

The visualization reveals distinct clustering patterns that highlight urban and suburban concentrations of electric vehicles. Dense clusters likely correspond to major metropolitan areas, showing significantly higher EV adoption compared to rural regions. This pattern aligns with expectations based on factors including charging infrastructure availability, commuting distances, and demographic characteristics of early technology adopters.

The color-coding by vehicle type (represented by the color bar) adds an additional dimension of analysis, allowing identification of potential geographic preferences for BEVs versus PHEVs. Some regions appear to have higher concentrations of certain vehicle types, potentially reflecting differences in driving requirements, charging infrastructure, or regional incentive programs.

From a business and policy perspective, this distribution offers critical insights for infrastructure planning and market development. For charging network operators, the visualization highlights high-density areas that may require expanded charging options. For manufacturers and dealerships, it identifies regions with strong EV penetration for targeted marketing and those with lower adoption that may represent growth opportunities. For policymakers, understanding the spatial distribution of EVs helps in designing geographically appropriate incentives and addressing potential "charging deserts" in underserved areas.

**5.3 EV Adoption Trends**

A graph with a line going up

AI-generated content may be incorrect.

This line chart tracks electric vehicle adoption trends over time, with model years on the x-axis and vehicle counts on the y-axis. The visualization uses a connected line with markers at each data point to show year-over-year changes in the number of registered electric vehicles.

The chart reveals a dramatic upward trajectory in EV adoption, particularly in recent years. The curve shows modest growth in earlier years, followed by an inflection point where adoption accelerates substantially. This hockey-stick pattern is characteristic of technologies moving from early adoption to early majority phases in the technology adoption lifecycle.

Several factors likely contribute to this accelerating adoption curve. Technological improvements in battery capacity, energy density, and charging speed have made EVs increasingly practical for everyday use. Expanding model availability across different vehicle classes has broadened consumer options beyond early compact EVs. Government incentives and environmental regulations have provided financial motivation and regulatory pressure for electrification. Finally, growing charging infrastructure has reduced range anxiety concerns.

From a business forecasting perspective, this curve provides valuable signals about market momentum. The steep recent increase suggests the EV market has achieved escape velocity from niche status and is moving toward mainstream adoption. For manufacturers, this trend validates investments in electric platforms and suggests the potential for continued growth. For suppliers and supporting industries (charging networks, battery manufacturers, etc.), it indicates expanding market opportunities. For policymakers, it shows that adoption incentives appear to be working, though continued support may be needed to maintain momentum through the full market transition.

**More Visualizations**

A graph of electric cars

AI-generated content may be incorrect.

A graph with different colored lines

AI-generated content may be incorrect.A graph of different colored lines

AI-generated content may be incorrect.

A graph with blue dots and black text

AI-generated content may be incorrect.

A diagram of a type of type of type

AI-generated content may be incorrect.

A graph of blue and orange bars

AI-generated content may be incorrect.

A graph of a bar chart

AI-generated content may be incorrect.

A screenshot of a graph

AI-generated content may be incorrect.

A graph showing a number of vehicles

AI-generated content may be incorrect.

A graph of a graph showing a long line

AI-generated content may be incorrect.

A graph of blue rectangular bars

AI-generated content may be incorrect.

A diagram of a network

AI-generated content may be incorrect.

A screenshot of a graph

AI-generated content may be incorrect.

A graph with blue dots and red lines

AI-generated content may be incorrect.

A graph of a vehicle age

AI-generated content may be incorrect.

A graph with blue bars

AI-generated content may be incorrect.

A green rectangular object with black text

AI-generated content may be incorrect.

A graph with a line going up

AI-generated content may be incorrect.

A screenshot of a graph

AI-generated content may be incorrect.

A graph with blue lines

AI-generated content may be incorrect.

A diagram of a graph

AI-generated content may be incorrect.

A graph with a line

AI-generated content may be incorrect.

*6. Conclusions and Recommendations*

**6.1 Key Findings**

1. **Market Concentration**: The electric vehicle market shows significant concentration, with Tesla maintaining a dominant position among manufacturers.
2. **Vehicle Type Distribution**: Battery Electric Vehicles (BEVs) comprise approximately 70% of the electric vehicle population, with Plug-in Hybrid Electric Vehicles (PHEVs) making up the remainder.
3. **Range Capabilities**: BEVs offer substantially higher electric ranges than PHEVs, with median ranges approximately five times greater.
4. **Efficiency Metrics**: Price-per-mile efficiency varies significantly across manufacturers and vehicle types, with PHEVs generally showing lower efficiency ratings than BEVs.
5. **Range Prediction**: Vehicle characteristics, particularly make and model, can predict electric range with exceptional accuracy (R² > 0.99 with LightGBM).
6. **Technological Improvement**: Newer vehicles consistently demonstrate better range capabilities, reflecting rapid advancement in battery technology.
7. **Geographic Distribution**: EV adoption shows distinct clustering in urban and suburban areas, with significant regional variation.
8. **Adoption Acceleration**: EV registrations have increased dramatically in recent years, indicating market transition from early adoption to early majority phases.

**6.2 Recommendations**

**For Manufacturers:**

1. **Focus on Range Efficiency**: Prioritize improvements in range-per-dollar metrics, as this appears to be a key differentiator for consumers.
2. **PHEV Optimization**: Manufacturers producing PHEVs should address their lower efficiency ratings by improving electric-only range capabilities or reducing costs.
3. **Model Diversification**: Continue expanding model offerings across different vehicle categories to capture diverse consumer needs.
4. **Targeted Geographic Expansion**: Use geospatial insights to identify underserved markets with growth potential.

**For Policymakers:**

1. **Infrastructure Development**: Address charging infrastructure gaps, particularly in regions showing lower EV adoption.
2. **Tailored Incentives**: Consider vehicle-type-specific incentives that address the different adoption barriers for BEVs versus PHEVs.
3. **Consumer Education**: Develop programs to increase awareness of electric range capabilities and efficiency metrics to reduce range anxiety.
4. **Long-term Planning**: Prepare for accelerating EV adoption by ensuring electrical grid capacity and charging infrastructure can meet projected demand.

**For Consumers:**

1. **Value Assessment**: Consider price-per-mile efficiency when evaluating vehicles to identify models offering the best range for the investment.
2. **Vehicle Type Selection**: Choose between BEV and PHEV based on personal driving patterns, with PHEVs better suited for those with occasional long trips and limited charging access.
3. **Technological Timing**: Recognize the rapid improvement in range capabilities when deciding between purchasing now or waiting for future models.

**6.3 Future Research Directions**

1. **Battery Degradation Analysis**: Expand the dataset to include historical range measurements to analyze how electric range changes over time and usage.
2. **Charging Behavior Patterns**: Combine this dataset with charging station usage data to better understand the relationship between range capabilities and charging behavior.
3. **Secondary Market Dynamics**: Investigate the relationship between vehicle age, range, and resale value to better understand EV depreciation patterns.
4. **Policy Impact Assessment**: Analyze how different regional incentives and regulations correlate with adoption rates and vehicle type preferences.

**7. References**

**Data Sources**

Washington State Department of Licensing. (2023). Electric Vehicle Population Data [Data file]. Retrieved from <https://data.wa.gov/Transportation/Electric-Vehicle-Population-Data/f6w7-q2d2>

**Python Libraries and Packages**

Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., ... & Oliphant, T.E. (2020). Array programming with NumPy. Nature, 585(7825), 357-362. <https://doi.org/10.1038/s41586-020-2649-2>

McKinney, W. (2010). Data Structures for Statistical Computing in Python. Proceedings of the 9th Python in Science Conference, 51-56. <https://conference.scipy.org/proceedings/scipy2010/mckinney.html>

Hunter, J.D. (2007). Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering, 9(3), 90-95. <https://doi.org/10.1109/MCSE.2007.55>

Waskom, M. (2021). Seaborn: Statistical Data Visualization. Journal of Open Source Software, 6(60), 3021. <https://doi.org/10.21105/joss.03021>

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825-2830. <https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T.Y. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. Advances in Neural Information Processing Systems, 30, 3146-3154. <https://proceedings.neurips.cc/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf>

Folium Development Team. (2023). Folium: Python Data, Leaflet.js Maps. GitHub repository. <https://github.com/python-visualization/folium>

**EV Market Research and Technical References**

International Energy Agency. (2023). Global EV Outlook 2023. IEA, Paris. <https://www.iea.org/reports/global-ev-outlook-2023>

BloombergNEF. (2023). Electric Vehicle Outlook 2023. Bloomberg Finance L.P. <https://about.bnef.com/electric-vehicle-outlook/>

Nykvist, B., & Nilsson, M. (2015). Rapidly falling costs of battery packs for electric vehicles. Nature Climate Change, 5(4), 329-332. <https://doi.org/10.1038/nclimate2564>

Hoekstra, A. (2019). The Underestimated Potential of Battery Electric Vehicles to Reduce Emissions. Joule, 3(6), 1412-1414. <https://doi.org/10.1016/j.joule.2019.06.002>

U.S. Department of Energy. (2023). Alternative Fuels Data Center - Electric Vehicle Benefits and Considerations. Retrieved from <https://afdc.energy.gov/fuels/electricity_benefits.html>

Lutsey, N., & Nicholas, M. (2019). Update on electric vehicle costs in the United States through 2030. The International Council on Clean Transportation (ICCT). <https://theicct.org/publication/update-on-electric-vehicle-costs-in-the-united-states-through-2030/>

National Renewable Energy Laboratory. (2022). Electric Vehicle Charging Infrastructure Trends from the Alternative Fueling Station Locator: First Quarter 2022. Retrieved from <https://www.nrel.gov/docs/fy22osti/83581.pdf>

**Machine Learning and Data Analysis References**

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An Introduction to Statistical Learning with Applications in R (2nd ed.). Springer. <https://www.statlearning.com/>

VanderPlas, J. (2016). Python Data Science Handbook: Essential Tools for Working with Data. O'Reilly Media. <https://jakevdp.github.io/PythonDataScienceHandbook/>

Brownlee, J. (2020). Feature Selection for Machine Learning with Python. Machine Learning Mastery. <https://machinelearningmastery.com/feature-selection-with-real-and-categorical-data/>

Kuhn, M., & Johnson, K. (2019). Feature Engineering and Selection: A Practical Approach for Predictive Models. Chapman and Hall/CRC. <https://doi.org/10.1201/9781315108230>

**8. Appendix: Methodology Details**

The analysis utilized Python with data science libraries including pandas, matplotlib, seaborn, scikit-learn, and LightGBM. Data cleaning addressed zero values and outliers in key fields, particularly electric range and base MSRP. Feature engineering created derived metrics including vehicle age, price per mile, and efficiency rating categories. Predictive modeling employed decision tree and gradient boosting approaches with cross-validation to ensure reliability.