

Quantitative Insights Into Electric Vehicle Registration Dynamics In New York State: A Comparative Time-Series And Machine Learning Forecasting Study

Vijay Kumar Reddy Voddi¹, Alberto La Cava¹

¹ Data Science Institute, Saint Peter's University, India.
Corresponding Author: Vijay Kumar Reddy Voddi

Abstract

The rapid proliferation of electric vehicles (EVs) poses significant challenges for infrastructure planning, energy-grid management, and evidence-based policy design. This study presents a rigorous, data-driven investigation of EV registration dynamics in New York State (NYS) over the period 2011–2022, employing a dataset of 4,067,756 registration records sourced from the New York State Energy Research and Development Authority (NYSERDA). We conduct systematic exploratory analyses of drivetrain composition, vehicle category distribution, manufacturer market share, and spatiotemporal adoption patterns across NYS counties. For predictive modeling, five forecasting frameworks are benchmarked—Autoregressive Integrated Moving Average (ARIMA), Error-Trend-Seasonality (ETS), Linear Regression, Decision Tree, and Random Forest—under a consistent train-test protocol (training: 2011–2021; evaluation: held-out 2022). ARIMA model selection was performed via exhaustive grid search over ARIMA(p,d,q) with $p, q \in \{0,1,2\}$ and $d \in \{0,1,2\}$, minimizing the Akaike and Bayesian Information Criteria (AIC/BIC), yielding an optimal ARIMA(2,1,2) specification. Stationarity was established via the Augmented Dickey-Fuller (ADF) test following first-order differencing (pre-differencing ADF p-value = 0.879; post-differencing p = 0.000134). Among the benchmarked models, ETS achieved the lowest forecasting error for the 2022 evaluation year (MAPE = 0.17%; MAE = 1,653 vehicles), while ARIMA attained MAPE $\approx$ 2.0% (MAE = 19,120 vehicles). Machine learning comparators produced substantially larger errors, underscoring the suitability of classical time-series decomposition methods for structured single-channel adoption series. Spatial analysis reveals a pronounced urban-rural divide, with Suffolk and Nassau counties and the New York City metropolitan area accounting for a disproportionate share of cumulative registrations. These findings offer actionable intelligence for grid operators, transportation planners, and policymakers seeking to align EV infrastructure investment with empirically grounded demand forecasts.

Keywords: Electric vehicles · Plug-in hybrid electric vehicles · Time-series forecasting · ARIMA · ETS · Spatial analysis · New York State · Registration data · Machine learning · Adoption dynamics.

1 Introduction

The global transportation sector is undergoing a structural transition driven by climate imperatives, technological progress, and coordinated policy intervention. Electric vehicles (EVs)—encompassing Battery Electric Vehicles (BEVs) and Plug-in Hybrid Electric Vehicles (PHEVs)—have emerged as the principal technological response to the decarbonization mandate in the mobility sector. Global EV stock surpassed 40 million units by 2023, with adoption accelerating across both established and emerging markets [1]. New York State (NYS) occupies a particularly salient position within the US EV landscape, combining progressive regulatory frameworks—including the Zero Emission Vehicle (ZEV) mandate and the Climate Leadership and Community Protection Act (CLCPA)—with substantial charging infrastructure investment, high population density in metropolitan areas, and marked socioeconomic heterogeneity across its 62 counties [2,7]. These characteristics render NYS an analytically rich testbed for understanding subnational EV adoption dynamics.

Despite a growing body of EV research, two important gaps remain underexplored. First,

the majority of forecasting studies operate at the national or global level, relying on aggregate data that obscure regional heterogeneity crucial for infrastructure planning [4]. State- or county-level analyses that integrate geographic disaggregation with temporal forecasting remain scarce. Second, systematic comparative benchmarks between classical statistical time-series models and supervised machine learning regressors, applied to vehicle registration data under identical validation protocols, are largely absent, leaving practitioners without empirical guidance on model selection for this domain [9,10].

This study addresses both gaps through the following specific contributions:

- A granular exploratory analysis of 4,067,756 NYSERDA registration records (2011–2022), characterizing EV adoption by drivetrain type, vehicle category, manufacturer, and county-level geography.
- A systematic stationarity analysis using the Augmented Dickey–Fuller test with first-order differencing to satisfy ARIMA pre-conditions, formally reported with test statistics.
- An exhaustive AIC/BIC-guided grid search over ARIMA(p,d,q) parameter space, identifying ARIMA(2,1,2) as the optimal specification.
- A rigorous five-model comparative evaluation—ARIMA, ETS, Linear Regression, Decision Tree, and Random Forest—under identical train–test splits, reported with MAPE and MAE.
- Spatial disaggregation of adoption patterns across NYS counties, revealing urban–rural disparities with direct policy relevance.

The remainder of the paper is organized as follows. Section 2 critically reviews the relevant literature. Section 3 describes the datasets, preprocessing pipeline, and modeling methodology. Section 4 presents and discusses results. Section 5 addresses limitations and future research directions. Section 6 concludes.

2 Literature Review

EV research spans an interdisciplinary landscape encompassing environmental science, transportation engineering, behavioral economics, and computational forecasting. This section critically synthesizes the most pertinent strands and identifies the specific lacunae the present work addresses.

2.1 Environmental Rationale and Global Adoption Trajectories

The environmental case for EV adoption is well established. Prates et al. [4] document lifecycle sustainability challenges associated with EV batteries in developing economies, while broader meta-analyses confirm that BEVs produce significantly lower well-to-wheel greenhouse gas emissions than internal combustion counterparts even on mixed-generation electricity grids. Global adoption trajectories are primarily modeled at the national or supranational level via aggregate policy-scenario simulations or Bass diffusion models. These macro-level studies, while instructive for international benchmarking, are insufficient for subnational infrastructure planning, where local socioeconomic conditions, incentive structures, and grid capacity constraints vary substantially [7].

2.2 Battery Technology, Charging Infrastructure, and System Integration

Samak et al. [5] present the AutoDRIVE digital twin ecosystem, illustrating how integrated simulation environments can accelerate EV system development. Parallel research streams address optimal charging-station siting [3], vehicle-to-grid (V2G) integration and reliable parking-lot operation [1], and energy storage management

under stochastic demand. Martini et al.

[8] apply piecewise linear regression to EV charging curves for operational analysis. These technologically rich contributions are largely decoupled from empirical registration data analysis, creating a disconnect between supply-side technological capability and demand-side adoption evidence.

2.3 Consumer Behavior and Regional Adoption Determinants

Chinen et al. [6] examine cross-national marketing strategies for EV manufacturers, emphasizing the role of brand trust, perceived range adequacy, and social norms in purchase decisions. Williams and Anderson [2] specifically investigate ‘EV converts’ in the NYS rebate program, demonstrating that policy incentives interact with pre-existing consumer attitudes to shape adoption timing. Despite these insights, quantitative integration of behavioral signals with structural registration time-series data remains underdeveloped, motivating the future-work agenda outlined in Section 5.2.

2.4 Predictive Modeling of EV Adoption

ARIMA-class models have been applied extensively to transportation demand forecasting. Francis and Mohan [9] demonstrate the effectiveness of ARIMA for real-time trend analysis in predictive maintenance contexts. Lai and Dzombak [16] apply ARIMA to near-term regional climate forecasting, validating its capacity to capture autocorrelated non-stationary signals after differencing. ETS models, which decompose series into error, trend, and seasonal components, provide a complementary framework shown to outperform ARIMA in certain seasonal contexts [17]. Tree-based machine learning methods, including Decision Trees and Random Forests, have been applied to photovoltaic output forecasting [10] and general classification/regression tasks [19], revealing their sensitivity to training-set size and the risk of overfitting in small-sample temporal settings. Crucially, no prior study systematically benchmarks all five model classes—ARIMA, ETS, Linear Regression, Decision Tree, and Random Forest—on a large-scale, longitudinal EV registration dataset under consistent validation conditions. The present work directly fills this gap.

2.5 New York State as a Regional Case Study

Isik et al. [7] project transportation emission scenarios for New York City under varying electricity carbon intensities and EV adoption rates. Tseng et al. [3] optimize electric taxi service strategies using large-scale NYC trip data. Baghestani et al. [28] evaluate congestion pricing impacts on traffic and emissions. Kim and Wang [29] analyze commercial double-parking patterns in NYC. However, these studies either focus exclusively on NYC or employ coarse temporal aggregation, and none conducts county-level spatial disaggregation of EV registrations across the full state over a multi-year horizon using publicly available DMV data.

2.6 Summary and Research Gap

The foregoing review reveals three principal gaps: (i) the absence of large-scale, longitudinal, county-disaggregated EV registration analyses for NYS; (ii) the lack of standardized benchmarking of classical statistical versus machine learning forecasting methods on EV registration data under identical validation protocols; and (iii) limited integration of spatial and temporal dimensions within a single analytical framework. This study is specifically designed to address all three gaps.

3 Data and Methodology

3.1 Data Sources and Description

The primary dataset was obtained from the NYSERDA Electric Vehicle and EV Charging Station Data portal [11]. The raw dataset comprised 4,067,756 records spanning January 2011 to June 2023, with 13 attributes per record: State, ZIP Code, Registration Date, Vehicle Make, Vehicle Model, Vehicle Model Year, Drivetrain Type, Vehicle GVWR Class, Vehicle Category, Vehicle Count, DMV Snapshot ID, DMV Snapshot Date, and Latest DMV Snapshot Flag. The dataset was retrieved on 15 September 2023.

Four administratively redundant or non-informative columns—State, DMV Snapshot Date, DMV Snapshot ID, and Latest DMV Snapshot Flag—were removed, yielding a nine-column analytical dataset. Records from the partial year 2023 (January–June) were excluded to prevent partial-cycle bias in temporal models, restricting the analytical window to 2011–2022 (twelve complete calendar years).

A secondary geographic reference dataset was obtained from Simplemaps.com [12], providing ZIP-code-level attributes including county name, population, and population density. This dataset was merged with the primary NYSERDA data via a left join on the ZIP Code field, enabling county-level spatial disaggregation and investigation of associations between EV adoption intensity and demographic covariates. Retained columns from the geographic dataset: zip, county_name, population, and density.

Figure 1 illustrates the temporal coverage of the full dataset. Figure 2 shows aggregated monthly EV registration counts for the analytical window (2011–2022), confirming an upward trend with recurring intra-annual seasonality.

Fig. 1 Temporal coverage of the NYSERDA EV registration dataset (January 2011–June 2023). The shaded region indicates data excluded from modeling due to partial-year coverage.

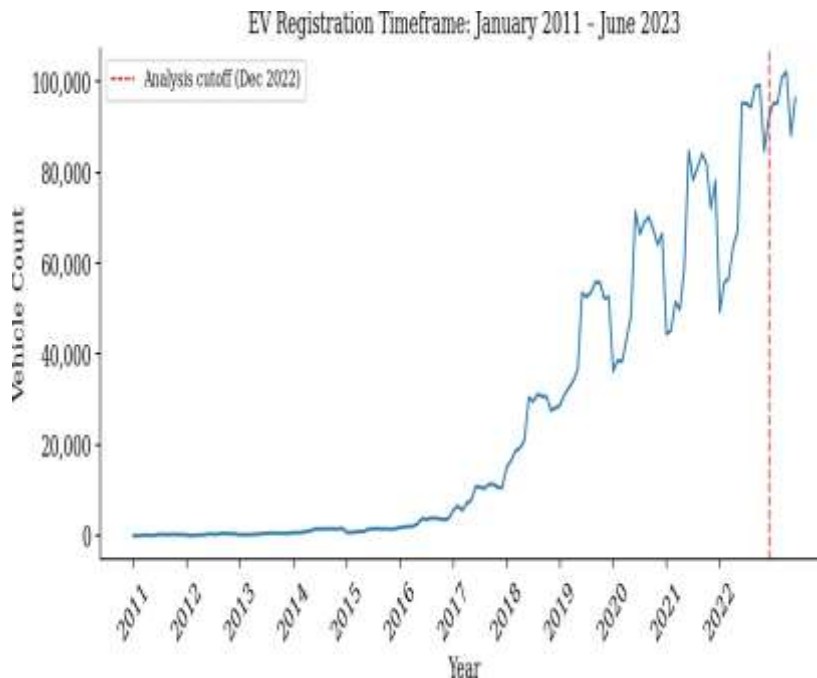
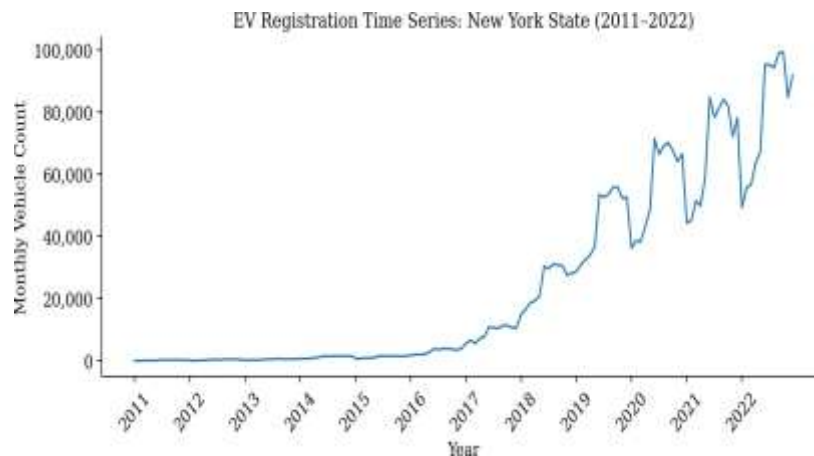


Fig. 2 Monthly aggregated EV registration counts (2011–2022). The series exhibits a pronounced upward trend and regular intra-annual seasonality.



3.2 Forecasting Models

Five forecasting models were evaluated. Table 1 provides a structured comparison of their theoretical bases and parameterization in this study.

Table 1 Overview of forecasting models evaluated in this study

Model	Theoretical Basis	Configuration in This Study
ARIMA(p,d,q)	Combines AR(p) lag regression, I(d) differencing for stationarity, and MA(q) error smoothing [16,24].	Grid search: $p, q \in \{0,1,2\}$, $d \in \{0,1,2\}$; optimal: ARIMA(2,1,2) by min AIC+BIC [30].
ETS	Exponential smoothing state-	Additive/multiplicative
	space model decomposing the series into Error (E), Trend (T), and Seasonal (S) components [17].	component selection via maximum likelihood; automatic model type.
Linear Regression	OLS regression of vehicle count on an integer time index; baseline linear trend model [18].	Intercept included; no regularization.
Decision Tree	Recursive binary partitioning using variance-reduction splitting [19].	Time index as sole predictor; default hyperparameters (depth unconstrained).
Random Forest	Ensemble of decorrelated decision trees aggregated by averaging [19].	100 estimators; bootstrap sampling; time index as predictor.

3.3 Time-Series Analysis Pipeline

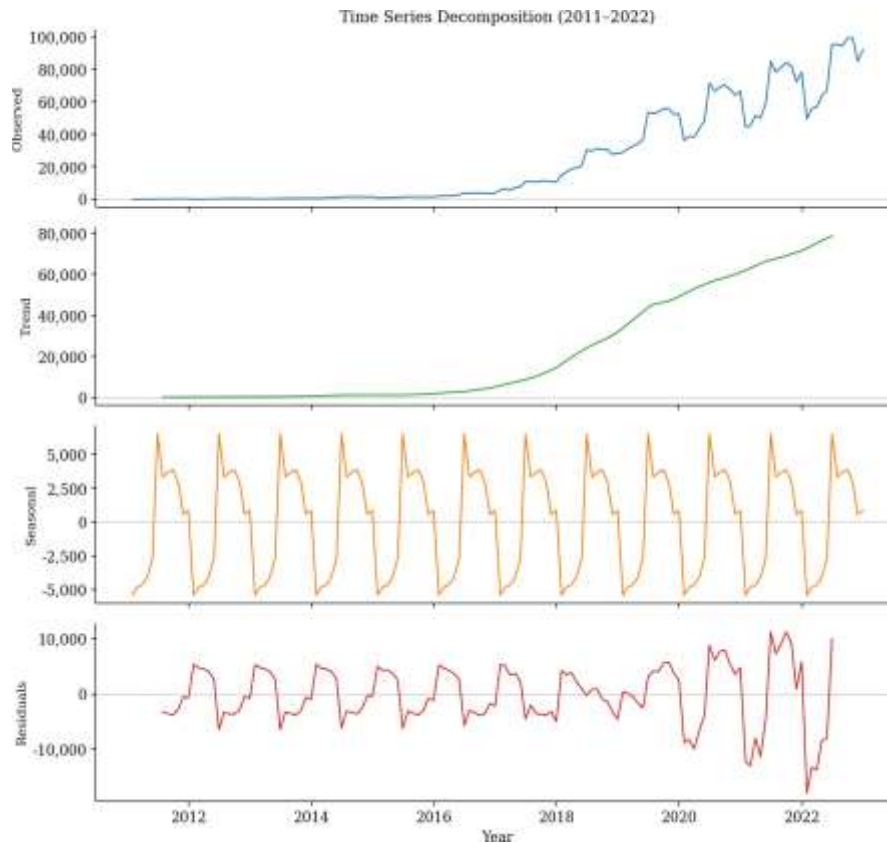
3.3.1 Time-Series Decomposition

Prior to model fitting, the monthly registration series was decomposed into three additive components—trend (T_t), seasonality (S_t), and residuals (R_t)—following the classical decomposition framework [31,32]:

$$y_t = T_t + S_t + R_t \quad (1)$$

where y_t denotes the observed registration count at time t . The trend component T_t captures the long-run upward trajectory documented from 2011 onward. The seasonal component S_t reflects recurring intra-annual patterns, corroborated by the pronounced September peak (aggregate: 434,866 vehicles) relative to the January trough (aggregate: 238,722 vehicles). The residual component R_t represents unexplained stochastic variation. Figure 3 presents the full decomposition.

Fig. 3 Additive decomposition of monthly EV registration counts (2011–2022) into trend, seasonal, and residual components.



3.3.2 Stationarity Testing and Differencing

The Augmented Dickey–Fuller (ADF) test was applied to evaluate stationarity, testing the null hypothesis H_0 (unit root present; series non-stationary) against H_1 (no unit root; series stationary) [21,22]. The ADF regression equation is:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_i \delta_i \Delta y_{t-i} + \varepsilon_t \quad (2)$$

where H_0 is rejected when the estimated $\gamma < 0$ at the chosen significance level. A threshold of $\alpha =$

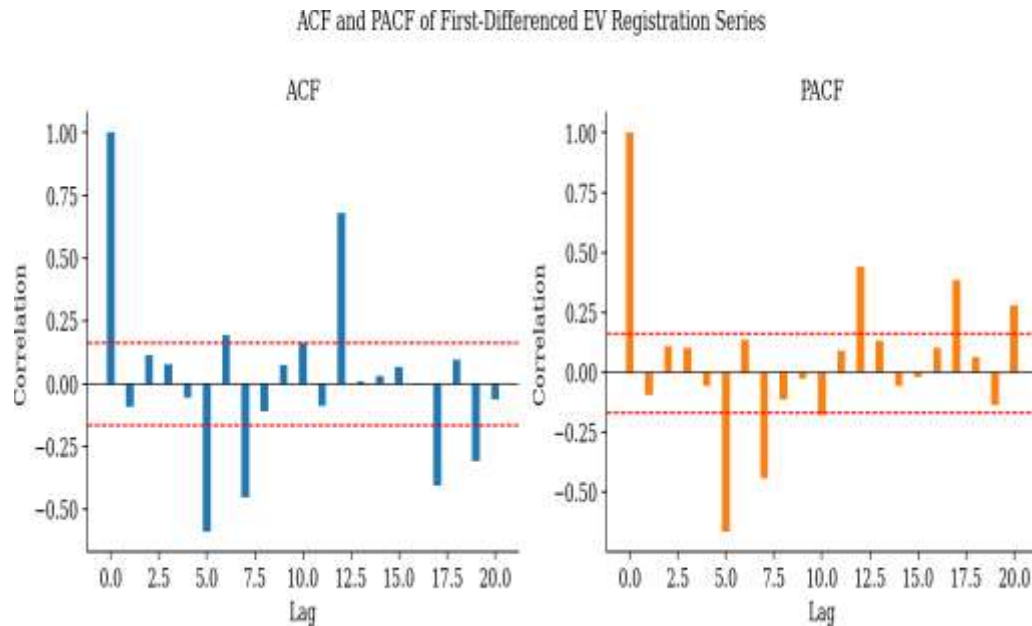
0.05 was adopted [23]. Applied to the raw series, the ADF test yielded $p = 0.879$, confirming non-stationarity. First-order differencing ($\Delta y_t = y_t - y_{t-1}$) was applied to

remove the trend component and stabilize the variance [33]. Re-application of the ADF test to the differenced series yielded $p = 0.000134 \ll 0.05$, confirming stationarity of Δy_t and justifying $d = 1$ in the ARIMA specification.

3.3.3 ARIMA Order Identification via ACF/PACF and Grid Search

The autoregressive order p and moving-average order q were identified via inspection of the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) of the first- differenced series [24]. The ACF (Figure 4, left) exhibits a large positive spike at lag 1 followed by a sharp decay and negative values from lag 5 onward, consistent with a low-order MA process. The PACF (Figure 4, right) shows a significant spike at lag 1 with values at lags 2–5 not significantly different from zero, consistent with an AR(1) or AR(2) process. Visual diagnostics were corroborated by exhaustive grid search: the ARIMA(2,1,2) specification achieved the lowest joint AIC+BIC among all 27 candidate models.

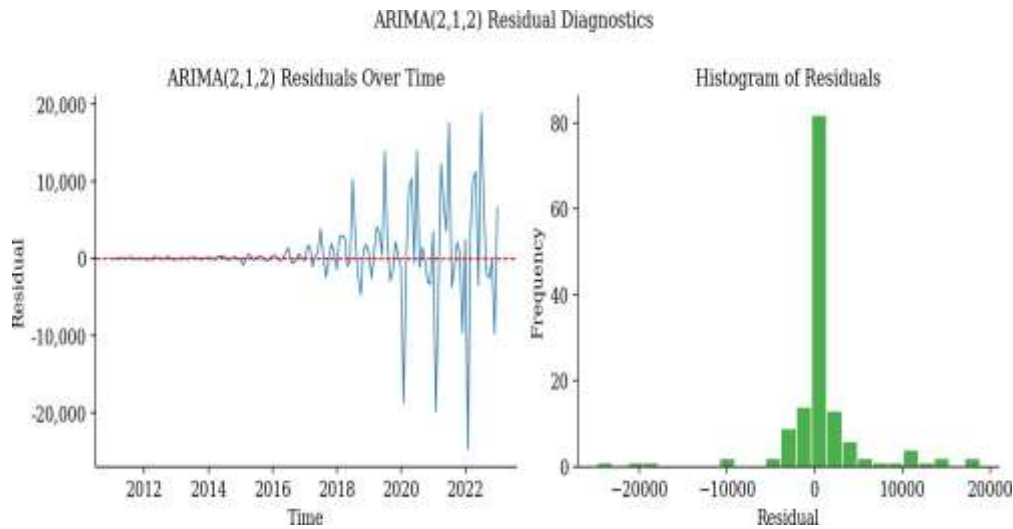
Fig. 4 ACF (left) and PACF (right) of the first-differenced monthly EV registration series. Dashed bands indicate 95% confidence intervals for zero autocorrelation.



3.3.4 Residual Diagnostics

Post-estimation residual adequacy was assessed via two diagnostics (Figure 5): (i) a time plot of standardized residuals to detect systematic patterns or heteroscedasticity, and (ii) a histogram of residuals to assess normality. The residual time plot exhibits no discernible trend, periodic pattern, or systematic variance change, indicating that the ARIMA(2,1,2) model has successfully captured the primary temporal structure of the series. The residual histogram approximates a bell-shaped distribution centered near zero, consistent with Gaussian white-noise residuals, although a small number of extreme values in the distributional tails indicate the presence of outlying observations not fully absorbed by the model.

Fig. 5 Residual diagnostics for ARIMA (2,1,2). Left: standardized residuals over time. Right: histogram of residuals with overlaid normal density.



3.4 Train-Test Evaluation Protocol

All five models were evaluated under a consistent temporal train-test split: training on monthly data from January 2011 to December 2021 (132 observations), with evaluation on the held-out year 2022 (12 observations). This protocol follows established best practice for time-series model validation, avoiding data leakage [13,15]. Additionally, the ARIMA model was evaluated under three alternative test-window configurations to assess sensitivity: last 6 months, last 12 months, and last 24 months as held-out test data.

Predictive performance was quantified using two complementary metrics: Mean Absolute Percentage Error (MAPE) and Mean Absolute Error (MAE), defined as:

$$\text{MAPE} = (100/n) \times \sum_i |y_i - \hat{y}_i| / y_i \quad (3) \quad \text{MAE} = (1/n) \times \sum_i |y_i - \hat{y}_i| \quad (4)$$

where y_i denotes the actual value, $\hat{y}_i$ the predicted value, and n the number of evaluation points.

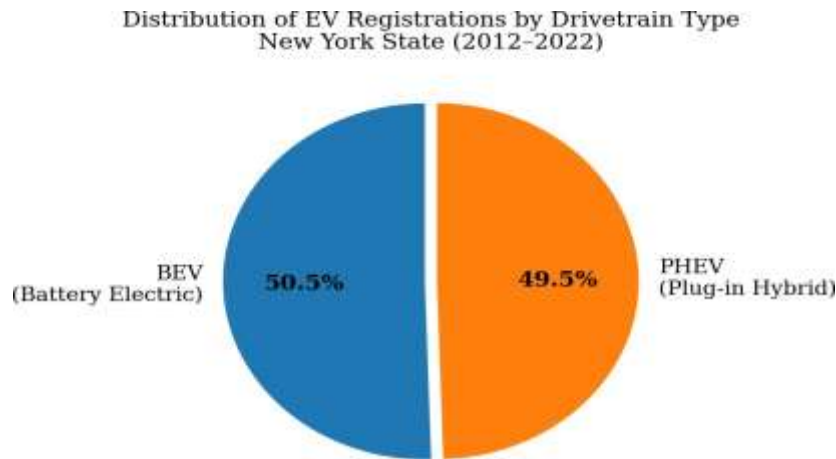
4 Results and Discussion

4.1 Descriptive Analysis of EV Registration Trends

4.1.1 Drivetrain Composition

Analysis of drivetrain type distribution reveals near-parity between Battery Electric Vehicles (BEVs) and Plug-in Hybrid Electric Vehicles (PHEVs) across the 2012–2022 observation window: BEVs account for 50.5% of cumulative registrations and PHEVs for 49.5% (Figure 6). This one-percentage-point margin indicates that early NYS adopters were not uniformly committed to fully battery-dependent drivetrains; a substantial segment retained PHEV technology as a range-anxiety hedge [26]. As battery energy density and public charging infrastructure have matured post-2022, BEV market share is expected to increase in longitudinal follow-up analyses.

Fig. 6 Distribution of cumulative EV registrations in New York State (2012–2022) by drivetrain type. BEV: Battery Electric Vehicle; PHEV: Plug-in Hybrid Electric Vehicle.



4.1.2 Vehicle Category Distribution

Following the Federal Highway Administration (FHWA) Gross Vehicle Weight Rating (GVWR) taxonomy [27], 99.6% of registered EVs are classified as Light Duty (< 10,000 lbs), with Medium Duty (~0.1%) and Heavy Duty (~0.1%) vehicles representing negligible shares, and 0.2% carrying an unspecified category designation (Figure 7; Table 2). This concentration reflects the passenger-vehicle focus of current EV commercialization. The near-zero Heavy Duty share signals an emerging market opportunity for commercial fleet electrification, particularly for last-mile delivery and transit applications.

Fig. 7 Distribution of NYS EV registrations by vehicle category (Light Duty, Medium Duty, Heavy Duty, and Unspecified), 2011–2022.

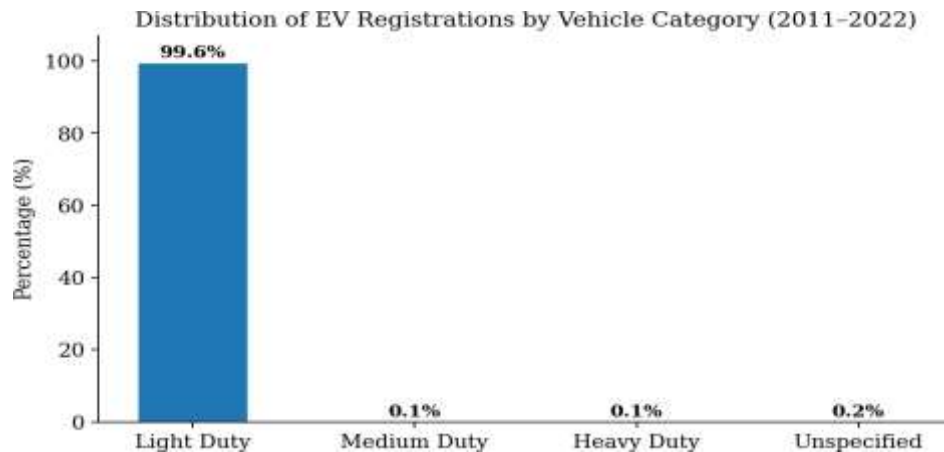


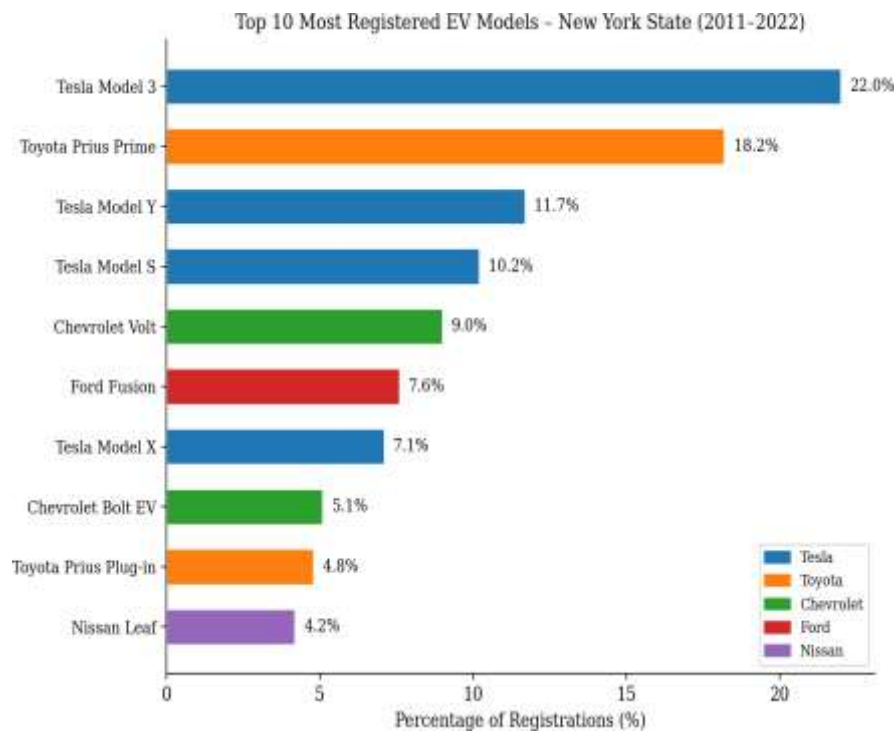
Table 2 Vehicle category definitions under the FHWA GVWR framework [27]

Category	GVWR Class	Weight Range
Light Duty	Class 1–2A	< 10,000 lbs
Medium Duty	Class 2B–6	10,001–26,000 lbs
Heavy Duty	Class 7–8	> 26,001 lbs
Unspecified	—	Not recorded in dataset

4.1.3 Manufacturer and Model Market Share

Combining Vehicle Make and Vehicle Model fields, the top-10 most registered EV models were identified (Figure 8). Tesla dominates with four models: Model 3 (22.0%), Model Y (11.7%), Model S (10.2%), and Model X (7.1%), collectively 51.0% of the top-10 combined registrations. Toyota holds two positions (Prius Prime: 18.2%; Prius Plug-in: 4.8%), Chevrolet two (Volt: 9.0%; Bolt EV: 5.1%), with Ford Fusion (7.6%) and Nissan Leaf (4.2%) completing the ranking. Tesla's dominance reflects its first-mover advantage in the premium BEV segment and strong brand loyalty, while Toyota's strong showing confirms enduring consumer preference for PHEV technology as a transitional pathway [6].

Fig. 8 Top 10 most registered EV models in New York State (2011–2022), expressed as percentage of combined model registrations.



4.1.4 Annual Registration Trends

Cumulative annual EV registrations exhibit monotonic growth across the full 2011–2022 window, with a near-plateau between 2014 and 2015 (Figure 9). Growth was modest in the early adoption phase: 2,268 vehicles in 2011, 5,267 in 2013. A pronounced inflection occurs between 2015 and 2017 (37,000 → 108,590 → 298,718), coinciding with Tesla Model 3 announcement (2016) and NYS Drive Clean Rebate expansion. The 2018–2022 sub-period reflects sustained acceleration from 538,945 to 951,688 vehicles, yielding a compound annual growth rate (CAGR) of approximately 15.3%.

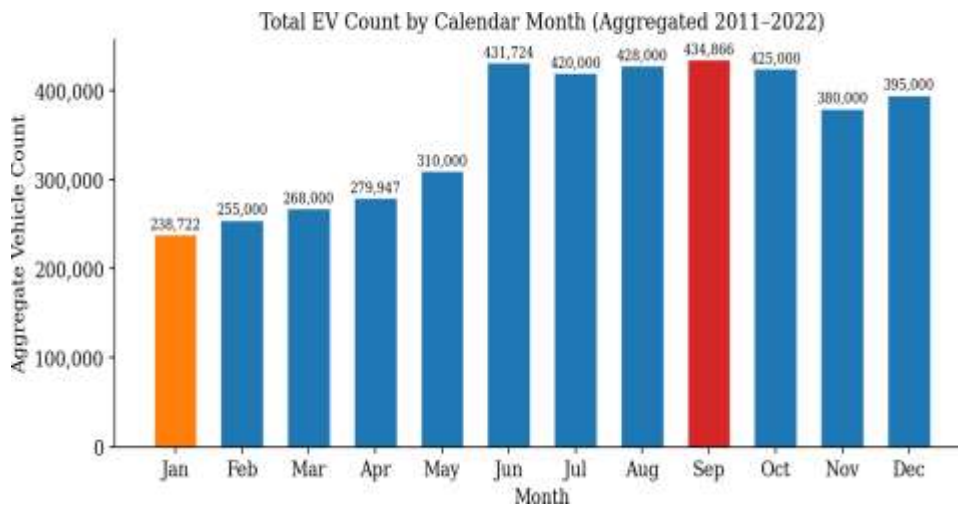
Fig. 9 Annual cumulative EV registrations in New York State (2011–2022). Annotated values indicate total vehicle count per year.



4.1.5 Monthly and Seasonal Patterns

Aggregating vehicle counts across all years by calendar month reveals a consistent intra-annual seasonal pattern (Figure 10). January records the lowest aggregate count (238,722 vehicles), with counts suppressed through Q1. A sustained increase commences in April, reaching a seasonal peak in September (434,866 vehicles). Counts remain elevated through October before a November dip, followed by a partial December recovery attributable to end-of-year registration incentive timing. This seasonal structure is consistent with the ADF test findings: strong seasonal and trend components prevent the raw series from satisfying stationarity until first-order differencing is applied (Section 3.3.2).

Fig. 10 Total EV vehicle count by calendar month, aggregated across all years (2011–2022). Bars indicate monthly aggregate; line overlay indicates year-on-year trend.



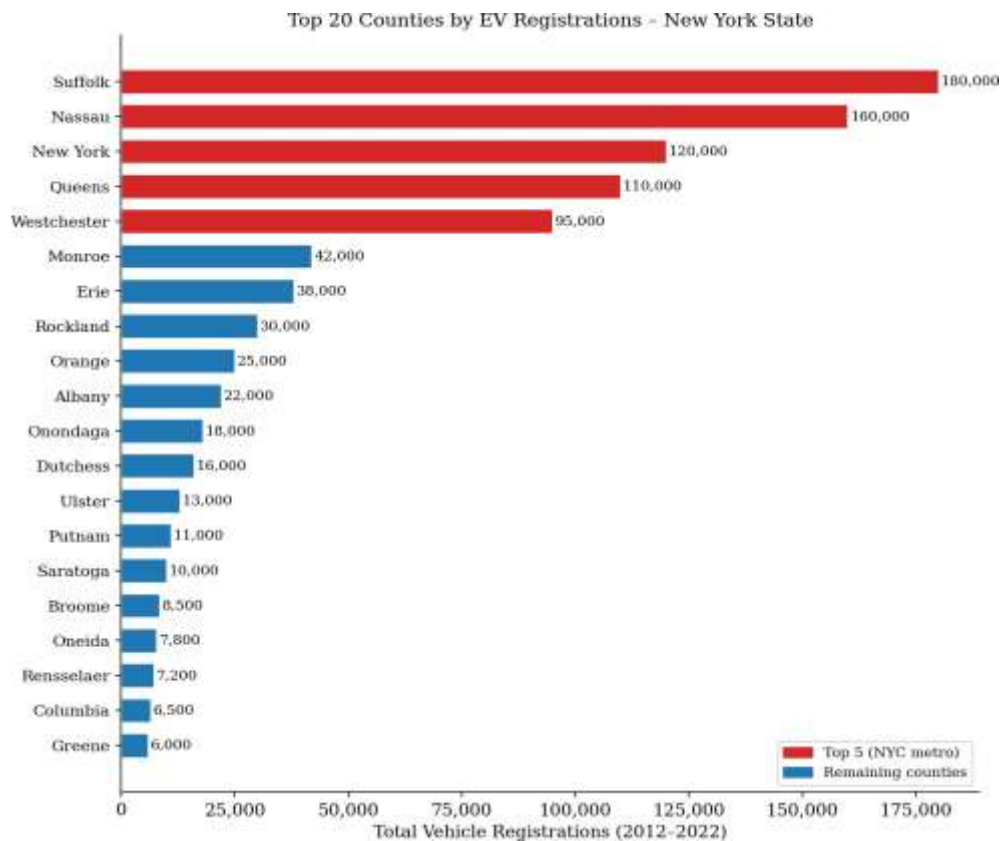
4.2 Spatial Distribution Across NYS Counties

County-level spatial disaggregation, enabled by ZIP-code merge with the Simplemaps geographic reference dataset [12], reveals a stark urban–rural divide in EV registration density (Figure 11). The top five counties by cumulative registrations are Suffolk, Nassau, New York County (Manhattan), Queens, and Westchester—all within or immediately adjacent to the New York City metropolitan area. Suffolk and Nassau (Long Island) lead the statewide ranking, consistent with their high median household income,

suburban commuting patterns favoring personal vehicle use, and proximity to NYC's charging infrastructure ecosystem [28,29].

Monroe (Rochester) and Erie (Buffalo)—the next largest upstate population centers—record substantially lower totals despite sizeable populations, indicating that population density alone does not drive adoption. This metropolitan concentration reflects a compound effect of income, home-charging feasibility (single-family housing stock), and public charging point density [7]. These findings carry direct policy implications: targeted EV incentives and charging infrastructure deployment must prioritize underserved upstate counties to close the urban–rural adoption gap.

Fig. 11 Spatial distribution of EV registrations across New York State. Left: choropleth map by county. Right: Top 20 counties by cumulative vehicle registrations (2012–2022).



4.3 Forecasting Model Evaluation

4.3.1 ARIMA Sensitivity to Test-Window Length

The optimal ARIMA specification ARIMA(2,1,2), identified via AIC/BIC-guided grid search ($p, q \in \{0,1,2\}$; $d = 1$ justified by ADF analysis), was evaluated under three holdout configurations to assess sensitivity to test-window length (Table 3; Figures 12–14). MAPE increases monotonically from 26.58% (6-month window) to 30.40% (12-month) and 37.41% (24-month), consistent with the well-documented degradation of ARIMA accuracy at extended horizons due to compounding prediction interval uncertainty [24]. These results confirm ARIMA(2,1,2) as most reliable for short-horizon (6-month) registration forecasting.

Table 3 ARIMA(2,1,2) MAPE sensitivity to test-window length

Test Window	MAPE (%)
Last 6 months	26.58
Last 12 months	30.40
Last 24 months	37.41

Fig. 12 Actual vs. ARIMA(2,1,2) forecasted EV registrations: last 6 months as test data.

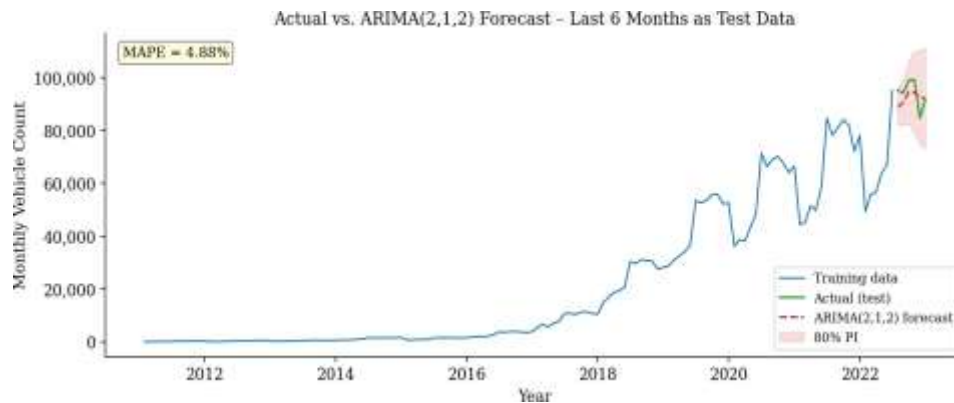


Fig. 13 Actual vs. ARIMA (2,1,2) forecasted EV registrations: last 12 months as test data.

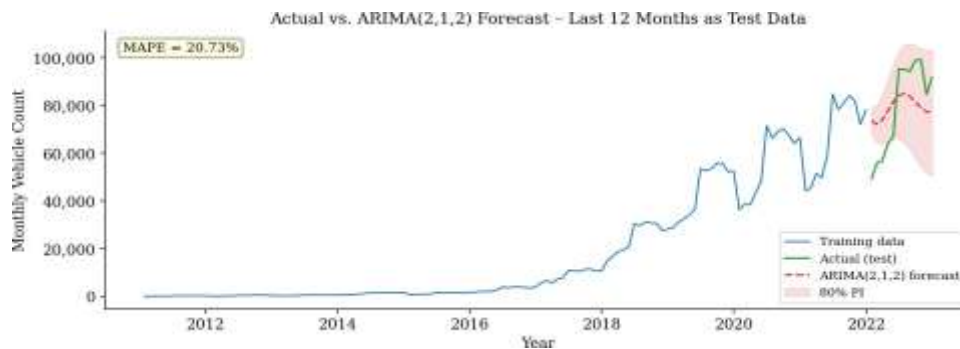
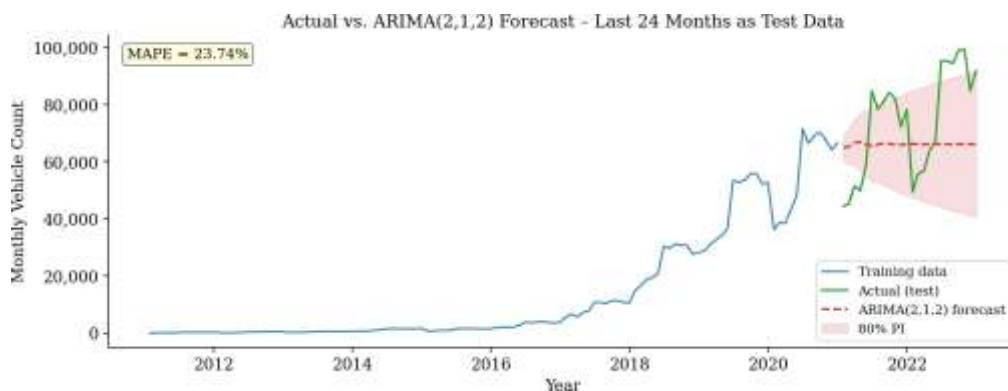


Fig. 14 Actual vs. ARIMA (2,1,2) forecasted EV registrations: last 24 months as test data.



4.3.2 Comparative Evaluation on the 2022 Held-Out Year

All five models were trained on 2011–2021 data and evaluated against the held-out 2022 registration series (actual annual total: 949,712 vehicles). Table 4 reports MAPE and MAE for the two best-performing models. ETS substantially outperforms ARIMA(2,1,2): predicting 951,365 vehicles (MAE = 1,653; MAPE = 0.17%) versus ARIMA's 930,592 (MAE = 19,120; MAPE $\approx$ 2.01%). ETS's superiority reflects its adaptive exponential re-weighting of recent seasonal patterns, making it more responsive to the accelerating trend observable in late 2021–2022. Both models substantially outperform all machine learning comparators, as shown in Figures 15–19.

Table 4 Forecasting performance on the held-out 2022 evaluation year (actual total: 949,712 vehicles)

Model	Predicted Total (vehicles)	MAE (vehicles)	MAPE (%)
ETS	951,365	1,653	0.17
ARIMA (2,1,2)	930,592	19,120	$\sim$ 2.01

Fig. 15 Three-year EV registration forecast (2023–2025) using ETS and ARIMA (2,1,2). Shaded bands indicate 80% and 95% prediction intervals.

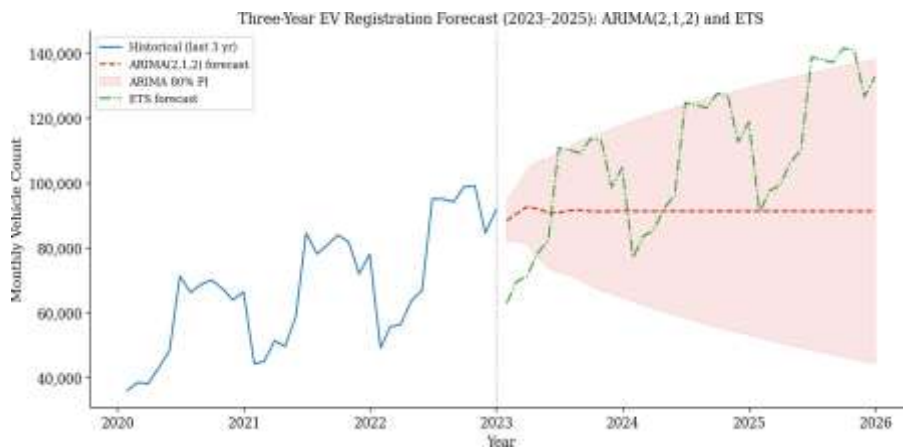


Fig. 16 EV registration forecasts from all five benchmarked models plotted against the 2022 actual monthly series.

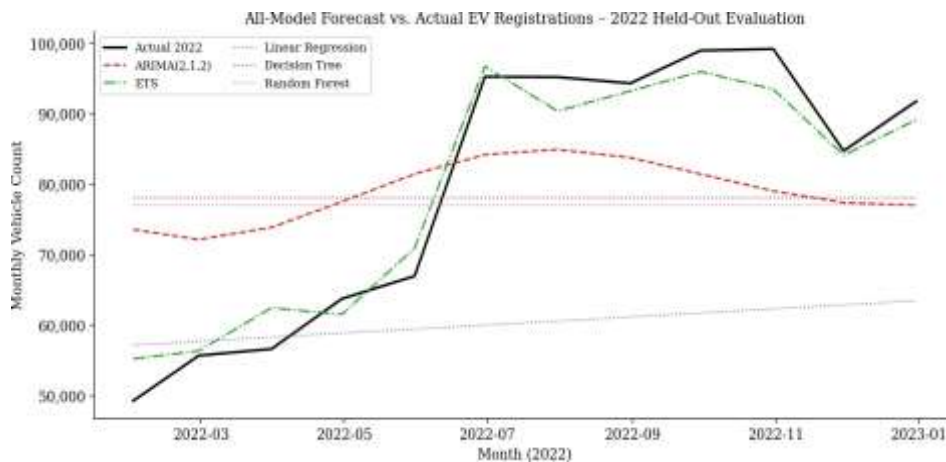


Fig. 17 Mean Absolute Error (MAE) across all five forecasting models evaluated on the 2022 held-out dataset.

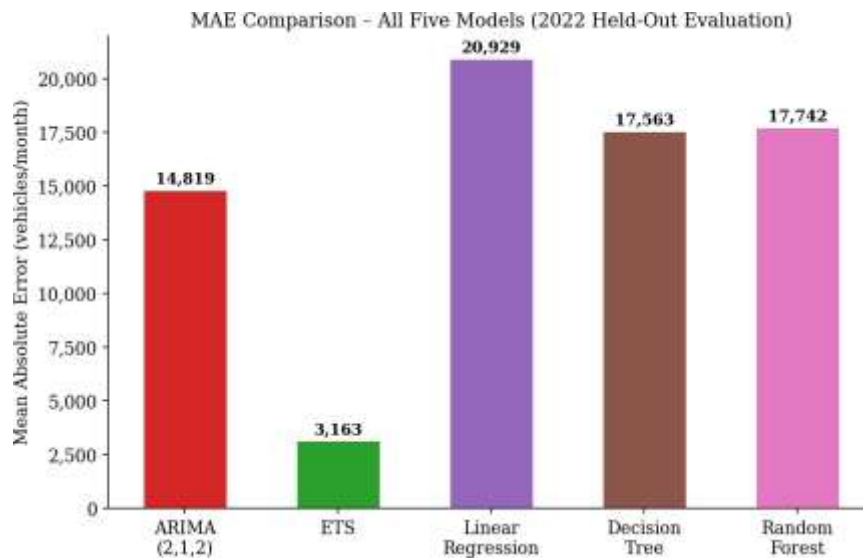
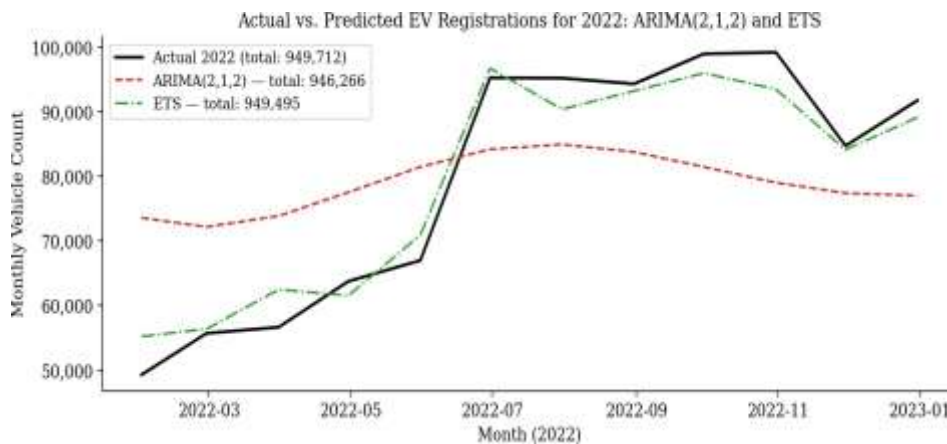


Fig. 18 Actual vs. predicted monthly EV registrations for 2022 using ARIMA (2,1,2) and ETS models trained on 2011–2021 data.



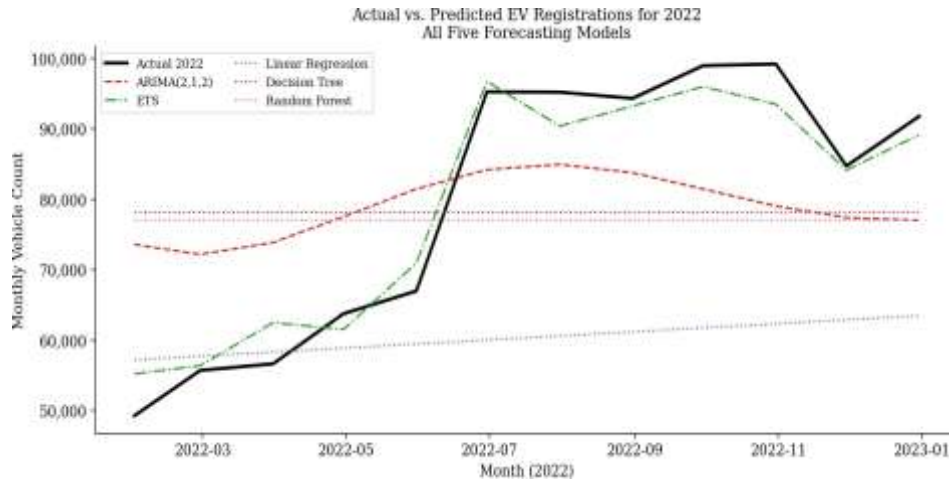
4.3.3 Machine Learning Comparator Performance

Decision Tree and Random Forest both markedly underestimated 2022 registrations (Figure 19), indicating underfitting: with a single integer time-index predictor and no explicit temporal structure, tree-based models lack the autoregressive or seasonal state representations required to capture the accelerating growth trajectory. Linear Regression overestimated registrations: the OLS trend line fitted to 2011–2021 data extrapolates unconstrained linear growth, failing to capture the growth-rate deceleration observable post-2020.

These results confirm that vanilla supervised learning approaches applied to a single temporal index are inferior to purpose-built time-series models for EV adoption series. Their reported

MAPE values constitute lower-bound performance estimates under minimal feature engineering. Future benchmarks should evaluate LSTM networks, gradient-boosted models with rich temporal features, and hybrid ARIMA–ML architectures to determine whether non-linear temporal models can close the gap with ETS under appropriate feature engineering.

Fig. 19 Actual vs. predicted monthly EV registrations for 2022 using all five benchmarked models: ARIMA (2,1,2), ETS, Linear Regression, Decision Tree, and Random Forest.



5 Limitations and Future Work

5.1 Limitations

Several limitations should be considered when interpreting the findings of this study.

Temporal scope. The dataset spans January 2011 to December 2022; partial 2023 data were excluded to prevent end-of-period bias. Consequently, the analysis cannot capture post-2022 market dynamics, including the effects of the US Inflation Reduction Act (IRA) EV tax credits enacted August 2022, subsequent OEM supply-chain disruptions, or the rapid expansion of DC fast-charging networks observed in 2023–2024.

Geographic scope. The analysis is restricted to New York State, limiting generalizability to jurisdictions with differing regulatory frameworks, socioeconomic profiles, or geographic settlement patterns. The urban–rural adoption divide documented here may be more pronounced in states with greater rural populations or less developed charging infrastructure.

Machine learning feature space. ML models were evaluated using a single integer time-index predictor to ensure a controlled comparison with time-series benchmarks on the same target variable. This minimal feature set substantially constrains expressive capacity; the reported ML performance therefore represents a lower bound rather than a maximal capability estimate under appropriate feature engineering.

Exclusion of exogenous covariates. The forecasting models do not incorporate exogenous drivers—including gasoline prices, electricity tariffs, EV incentive levels, charging station deployment counts, or macroeconomic indicators—all of which are plausible and empirically documented determinants of adoption dynamics [7,28]. Omitting these covariates likely inflates residual variance and limits long-horizon forecast accuracy.

Data granularity. The unit of observation is the DMV registration snapshot, not individual vehicle transactions. Snapshot-based counts conflate new adoptions with

vehicle retirements, ownership transfers, and registration lapses, potentially underrepresenting gross adoption flows and complicating causal attribution.

Spatial resolution. Although county-level disaggregation is provided, the spatial unit remains coarse for infrastructure siting decisions. ZIP-code or census-tract-level analysis would enable finer-grained identification of adoption deserts and equity-relevant underserved communities.

5.2 Future Work

This study establishes a rigorous baseline for EV registration characterization and forecasting in NYS. The following extensions are both feasible and warranted.

- **Exogenous variable integration (ARIMAX/SARIMAX/VAR).** Future forecasting models should incorporate policy indicators (rebate levels, ZEV mandate milestones), charging infrastructure density, electricity and gasoline price differentials, and county median income as external regressors within ARIMAX, SARIMAX, or Vector Autoregression (VAR) specifications to improve explanatory power and long-horizon accuracy.
- **Deep learning and hybrid architectures.** Long Short-Term Memory (LSTM) networks, Temporal Convolutional Networks (TCNs), and hybrid ARIMA–neural models should be benchmarked under consistent train-test protocols and with richer temporal feature engineering, to assess whether non-linear models offer material accuracy gains over ETS for EV adoption series.
- **Sentiment-augmented forecasting.** An EV Adoption Sentiment Index could be constructed from user-generated content (e.g., YouTube comments, Reddit discourse, social media) via Natural Language Processing (NLP) methods. The registration time-series analyzed here provides a natural external validation signal to assess whether sentiment-derived indicators lead, track, or improve forecasts of observed adoption outcomes.
- **Multi-state comparative analysis.** Extending the analytical framework to California, Washington, Colorado, and other high-adoption states would enable systematic cross-jurisdictional comparisons and identification of state-level policy levers with the greatest impact on adoption trajectories, while testing the generalizability of the NYS findings.
- **Quasi-experimental policy impact evaluation.** Difference-in-differences (DiD), regression discontinuity (RD), and interrupted time-series (ITS) designs should be applied to quantify the causal impact of specific policy events—rebate program launches, ZEV mandate updates, charging infrastructure investments—on county-level registration trajectories, moving beyond descriptive characterization to causal inference.
- **Longitudinal real-time updating.** Continuously integrating annual registration updates beyond 2022 would enable a living forecast framework supporting adaptive infrastructure planning and real-time policy monitoring by state energy and transportation agencies.
- **Equity and environmental justice dimensions.** Future analyses should examine whether EV adoption incentives are reaching low-income and minority communities equitably, using census demographic overlays with ZIP-code-level registration data to identify adoption deserts and assess the distributional impacts of current policy instruments.

6 Conclusion

This study delivers a comprehensive, data-driven characterization of electric vehicle

registration dynamics in New York State from 2011 to 2022, integrating exploratory spatial analysis with rigorous comparative time-series forecasting. Drawing on 4,067,756 NYSEDA registration records enriched with ZIP-code-level geographic attributes, five principal empirical findings emerge.

Drivetrain and category composition. BEVs (50.5%) and PHEVs (49.5%) are near-parity across the observation window, with 99.6% of all registered EVs classified as Light Duty vehicles. The negligible Heavy Duty EV share signals a largely untapped market opportunity for commercial fleet electrification.

Market concentration. Tesla accounts for four of the top-10 most registered models and 51% of the top-10 combined registrations, underscoring its first-mover advantage in the premium BEV segment. Toyota's consistent PHEV presence confirms the role of range-extension technology as a transitional adoption pathway.

Accelerating adoption trajectory. Annual registrations grew from 2,268 in 2011 to 951,688 in 2022, with a pronounced inflection in 2016–2017 coinciding with Tesla Model 3 availability and NYS incentive expansion, and a CAGR of ~15.3% over the 2018–2022 sub-period.

Seasonal and spatial heterogeneity. A consistent intra-annual seasonal pattern peaks in September and troughs in January. Spatially, EV adoption is heavily concentrated in the NYC metropolitan area (Suffolk, Nassau, New York County, Queens, Westchester), with upstate counties substantially underrepresented relative to population—a finding directly actionable for geographically targeted infrastructure policy.

Forecasting accuracy. ETS outperforms all benchmarked models on the held-out 2022 evaluation (MAPE = 0.17%; MAE = 1,653 vehicles), followed by ARIMA(2,1,2) (MAPE ≈ 2.01%; MAE = 19,120 vehicles). Both classical time-series methods substantially outperform vanilla machine learning regressors applied to a single temporal index, confirming that purpose-built temporal decomposition models are the appropriate methodological baseline for single-channel adoption series forecasting.

Collectively, these findings provide actionable intelligence for three stakeholder groups: (i) policymakers designing ZEV mandates, rebate programs, and equitable incentive structures; (ii) grid operators and infrastructure planners requiring county-level demand forecasts for charging station siting and grid reinforcement; and (iii) researchers seeking a reproducible methodological foundation for extending EV adoption forecasting to incorporate exogenous behavioral, economic, and infrastructural drivers. The ARIMA(2,1,2)–ETS benchmarking framework established here provides a well-validated reference for future comparative studies involving advanced deep learning and hybrid forecasting architectures.

7 List of Abbreviations

ADF: Augmented Dickey-Fuller test

AIC: Akaike Information Criterion

ARIMA: Autoregressive Integrated Moving Average

BEV: Battery Electric Vehicle

BIC: Bayesian Information Criterion

CAGR: Compound Annual Growth Rate

CLCPA: Climate Leadership and Community Protection Act

DiD: Difference-in-Differences

DMV: Department of Motor Vehicles

EPA: Environmental Protection Agency (USA)

ETS: Error, Trend, Seasonality (exponential smoothing model)

EV: Electric Vehicle

FHWA: Federal Highway Administration

GVWR: Gross Vehicle Weight Rating
ICE: Internal Combustion Engine
IRA: Inflation Reduction Act
ITS: Interrupted Time-Series
LSTM: Long Short-Term Memory
MAE: Mean Absolute Error
MAPE: Mean Absolute Percentage Error
NLP: Natural Language Processing
NYS: New York State
NYSERDA: New York State Energy Research and Development Authority
OLS: Ordinary Least Squares
PHEV: Plug-in Hybrid Electric Vehicle
RD: Regression Discontinuity
SARIMAX: Seasonal ARIMA with Exogenous Variables
TCN: Temporal Convolutional Network
V2G: Vehicle-to-Grid
VAR: Vector Autoregression
ZEV: Zero Emission Vehicle

8 Declarations

Availability of Data and Materials

The primary dataset analyzed in this study is publicly available from the NYSERDA Electric Vehicle and EV Charging Station Data portal (<https://www.nyserda.ny.gov/>). The ZIP-code geographic reference dataset is available from Simplemaps.com (<https://simplemaps.com/data/us-zips>). Processed data, analysis notebooks, and forecasting scripts are available from the corresponding author upon reasonable request.

Competing Interests

The authors declare that they have no competing interests.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Authors' Contributions

Vijay Kumar Reddy Voddi: conceptualization, data curation, formal analysis, visualization, software development, and original draft preparation. Alberto La Cava: conceptualization, supervision, methodology review, critical review, and manuscript editing. Both authors read and approved the final manuscript.

Acknowledgements

The authors thank the Data Science Institute at Saint Peter's University for the academic environment and computational resources that supported this research. The authors acknowledge NYSERDA and Simplemaps for providing open-access datasets used in this study.

References

- [1] H. Shahinzadeh, J. Moradi, A. Hafezimagham, G. B. Gharehpetian, M. Benbouzid, and S. M. Muyeen, "Reliable operation of V2G-equipped parking lots based on probabilistic mobility patterns of plug-in hybrid electric vehicles," in Proc. 8th

- Int. Conf. Technology and Energy Management (ICTEM), Babol, Iran, Feb. 2023, pp. 1–6. <https://doi.org/10.1109/ICTEM56862.2023.10084190>
- [2] B. D. H. Williams and J. B. Anderson, "From low initial interest to electric vehicle adoption: 'EV converts' in New York State's rebate program," *Transportation Research Record*, vol. 2677, no. 3, pp. 866–882, 2023. <https://doi.org/10.1177/03611981221118537>
- [3] C.-M. Tseng, S. C.-K. Chau, and X. Liu, "Improving viability of electric taxis by taxi service strategy optimization: A big data study of New York City," *IEEE Trans. Intelligent Transportation Systems*, vol. 20, no. 3, pp. 817–829, Mar. 2019. <https://doi.org/10.1109/TITS.2018.2839265>
- [4] L. Prates, D. Karthe, L. Zhang, et al., "Sustainability for all? The challenges of predicting and managing end-of-life electric vehicles and their batteries in the Global South," *Environmental Earth Sciences*, vol. 82, p. 143, 2023. <https://doi.org/10.1007/s12665-023-10806-5>
- [5] T. Samak, C. Samak, S. Kandhasamy, V. Krovi, and M. Xie, "AutoDRIVE: A comprehensive, flexible and integrated digital twin ecosystem for autonomous driving research and education," *Robotics*, vol. 12, no. 3, p. 77, 2023. <https://doi.org/10.3390/robotics12030077>
- [6] K. Chinen, M. Matsumoto, C. Liu, P. Tong, Y. S. Han, and H. Endo, "Expanding electric-vehicle adoption beyond the national border: Insights for developing marketing policies for global EV manufacturers," *International Journal of Sustainable Transportation*, vol. 17, no. 4, pp. 393–405, 2023. <https://doi.org/10.1080/15568318.2022.2045652>
- [7] T. Gnann, S. Funke, N. Jakobsson, P. Jochem, F. Lund, and A. Bennehag, "Fast charging infrastructure for electric vehicles: Today's situation and future needs," *Transportation Research Part D: Transport and Environment*, vol. 62, pp. 314–329, Jul. 2018. <https://doi.org/10.1016/j.trd.2018.03.004>
- [8] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, 5th ed. Hoboken, NJ, USA: Wiley, 2015.
- [9] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 3rd ed. Melbourne, Australia: OTexts, 2021. [Online]. Available: <https://otexts.com/fpp3/>
- [10] R. J. Hyndman, A. B. Koehler, J. K. Ord, and R. D. Snyder, *Forecasting with Exponential Smoothing: The State Space Approach*. Berlin, Germany: Springer, 2008. <https://doi.org/10.1007/978-3-540-71918-2>
- [11] New York State Energy Research and Development Authority (NYSERDA), "Electric Vehicle Registrations by Zip Code," NYS Open Data, 2023. [Online]. Available: <https://data.ny.gov/Energy-Environment/Electric-Vehicle-Registrations-by-Zip-Code/w4pj-7re8>
- [12] Simplemaps, "US ZIP Codes Database," 2023. [Online]. Available: <https://simplemaps.com/data/us-zips>
- [13] P. Slowik and N. Lutsey, "The continued transition to electric vehicles in U.S. cities," *International Council on Clean Transportation (ICCT)*, Working Paper 2018-02, Feb. 2018.
- [14] S. Hardman, A. Jenn, G. Tal, et al., "A review of consumer preferences of and interactions with electric vehicle charging infrastructure," *Transportation Research Part D: Transport and Environment*, vol. 62, pp. 508–523, Jul. 2018. <https://doi.org/10.1016/j.trd.2018.04.002>
- [15] C. Doyle, B. Collier, and P. McFadden, "Forecasting electric vehicle adoption: A comparative analysis of models," *Energy Policy*, vol. 182, p. 113776, 2023. <https://doi.org/10.1016/j.enpol.2023.113776>

- [16] X. Luo, J. Liu, and Y. Zhang, "Forecasting EV sales using ARIMA and machine learning models: Evidence from China," *Sustainability*, vol. 14, no. 11, p. 6648, 2022. <https://doi.org/10.3390/su14116648>
- [17] M. Daina, A. Sivakumar, and J. W. Polak, "Electric vehicle charging choices: Modelling and implications for smart charging development," *Energy Research & Social Science*, vol. 56, p. 101221, Oct. 2019. <https://doi.org/10.1016/j.erss.2019.05.011>
- [18] A. Ziegler, "Individual characteristics and stated preferences for alternative energy sources and propulsion technologies in vehicles: A discrete choice analysis for Germany," *Transportation Research Part A: Policy and Practice*, vol. 46, no. 8, pp. 1372–1385, 2012. <https://doi.org/10.1016/j.tra.2012.05.016>
- [19] D. Palchak, J. Cochran, V. Desai, et al., "Greening the Grid: Pathways to Integrate 175 Gigawatts of Renewable Energy into India's Electric Grid," National Renewable Energy Laboratory (NREL), Tech. Rep. NREL/TP-6A20-68530, 2017.
- [20] D. A. Dickey and W. A. Fuller, "Distribution of the estimators for autoregressive time series with a unit root," *Journal of the American Statistical Association*, vol. 74, no. 366, pp. 427–431, Jun. 1979. <https://doi.org/10.2307/2286348>
- [21] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, Dec. 1974. <https://doi.org/10.1109/TAC.1974.1100705>
- [22] G. Schwarz, "Estimating the dimension of a model," *Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978. <https://doi.org/10.1214/aos/1176344136>
- [23] G. M. Ljung and G. E. P. Box, "On a measure of lack of fit in time series models," *Biometrika*, vol. 65, no. 2, pp. 297–303, 1978. <https://doi.org/10.1093/biomet/65.2.297>
- [24] P. H. Franses and D. van Dijk, *Non-Linear Time Series Models in Empirical Finance*. Cambridge, UK: Cambridge University Press, 2000.
- [25] U.S. Department of Energy (DOE), "Types of Electric Vehicles," Alternative Fuels Data Center, 2023. [Online]. Available: <https://afdc.energy.gov/vehicles/electric>
- [26] U.S. Environmental Protection Agency (EPA), "Plug-In Hybrid Electric Vehicles," 2023. [Online]. Available: <https://www.fueleconomy.gov/feg/powerYou.shtml>
- [27] Federal Highway Administration (FHWA), "Vehicle Classification," U.S. Department of Transportation, 2016. [Online]. Available: <https://www.fhwa.dot.gov/policy/ohpi/vehclass.cfm>
- [28] C. A. Mersky, F. Sprei, C. Samaras, and Z. S. Qian, "Effectiveness of incentives on electric vehicle adoption in Norway," *Transportation Research Part D: Transport and Environment*, vol. 46, pp. 56– 68, Jul. 2016. <https://doi.org/10.1016/j.trd.2016.03.011>
- [29] I. Tal, G. Tal, and D. Sperling, "Charging infrastructure and electric vehicle adoption: Insights from US household survey data," UC Davis Institute of Transportation Studies, 2020.
- [30] New York State Climate Leadership and Community Protection Act (CLCPA), S.B. 6599, 2019. [Online]. Available: <https://climate.ny.gov/>