

Forecasting and Clustering EV Adoption in ON

Objectives

The goal of this project was to analyze and forecast electric vehicle (EV) adoption trends across Ontario using quarterly registration data organized by Forward Sortation Area (FSA).

Two specific objectives are:

1. Clustering Analysis: Identify groups of FSAs with similar EV adoption patterns to uncover geographic disparities, market maturity levels, and technology preferences.
2. Predictive Modelling: Forecast short-term EV adoption to guide infrastructure development, policy-making, and strategic planning.

Electric Vehicle (EV) adoption is central to Ontario's transition to sustainable transportation, and actionable insights are needed to support equitable infrastructure deployment and targeted incentives.

Data Preparation

The dataset, sourced from the Ontario Open Data

Catalogue(<https://data.ontario.ca/dataset/electric-vehicles-in-ontario-by-forward-sortation-area>), contained quarterly EV registration counts—including Battery Electric Vehicles (BEVs) and Plug-in Hybrid Electric Vehicles (PHEVs)—for Ontario FSAs from Q2 2023 to Q1 2025. A supplementary geographic dataset provided FSA coordinates, city names, and regional groupings. This data was downloaded locally as a CSV, which was used for preparation and modeling. The steps involved in the preparation of this data included filtering FSAs to Ontario using Postal Codes starting with K, L, M, N, and P. The data had missing points, so the missing geographic data were filled using public sources and centroid estimates.

Challenges

Some FSAs were missing from reliable geographic databases and were not officially recognized as postal regions by Canada Post or mapping authorities, creating notable gaps in geospatial plotting. Historical records were limited, which made it difficult to train stable forecasting models. As a result, LSTM models performed poorly. Overall, these issues point to the need for more extensive historical data, the inclusion of richer and more relevant features, the careful selection of simpler yet well-suited models, and the application of robust validation frameworks to improve forecast reliability.

Clustering Analysis

We applied K-Means clustering on three standardized metrics:

- Adoption Level: Total registered EVs in Q1 2025.
- Growth Rate: Percentage increase from Q2 2023 to Q1 2025.
- BEV Preference: Average BEV share over the period.

The model segmented Ontario FSAs into four clusters:

- High-Growth, BEV-Dominant Suburbs – Affluent suburban areas with high adoption and strong BEV preference (e.g., Oakville, Milton).
- Established Urban & Commuter Zones – Large urban centers with strong, steady growth (e.g., Toronto, Ottawa).
- Lagging Rural Regions – Low adoption and lower BEV preference, more reliance on PHEVs.
- New/Stagnant Micro-Markets – Very low adoption but high BEV preference among a small number of early adopters.

This segmentation confirms regional disparities and supports targeted infrastructure and policy strategies.

Forecasting and Clustering EV Adoption in ON

Predictive Modelling

- Linear Regression: Baseline model for simplicity and interpretability.
- Random Forest: Captures nonlinear relationships and interactions, robust for tabular datasets.

Additional models (ARIMA, LSTM, and an ensemble) were tested, but underperformed due to limited temporal data.

Predictive Modeling

The Random Forest model was tuned via GridSearchCV with 5-fold cross-validation, using negative RMSE as the scoring metric. Final parameters were n_estimators = 100, max_depth = None, and min_samples_split = 2.

The mean cross-validated R^2 was approximately 0.97, with minimal training-test score gaps, indicating balanced model fit without overfitting.

The Models were tested using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination.

- Random Forest: $R^2 \approx 0.98$, RMSE $\approx 2,964$, MAE ≈ 29 (best performer).
- Linear Regression: $R^2 \approx 0.77$, RMSE $\approx 29,211$, MAE ≈ 114 .
- ARIMA, LSTM, Ensemble: Negative R^2 values, confirming poor fit with limited data.

Random Forest outperformed all other approaches, providing accurate short-term forecasts.

Conclusions

The results from the predictive models indicate that Electric Vehicle adoption is growing across Ontario, but is highly uneven across regions. Location based data analysis shows affluent suburban FSAs in the Greater Golden Horseshoe lead this growth, with rural FSAs lagging. Also, location based, urban and suburban clusters prefer BEVs while rural areas lean toward PHEVs. The data only ranging between four quarters can show that short-term forecasting reliably identifies future high-growth FSAs, mainly in the Greater Toronto Area.

Next steps

Future exploration on the topic of Electric Vehicle adoption across Ontario can be improved and diversified by extending historical data to improve sequential forecasting models. There are also many other factors that impact this adoption rate. Taking into account these various factors, such as adding socioeconomic, demographic, and infrastructure variables, can enhance explanatory power and show more complex relationships between features.

Other than data collection related fixes, these models could be polished using rolling cross-validation for robust evaluation. This revision to the predictive models can then be used to potentially conduct qualitative research into consumer preferences to inform targeted Electric Vehicle adoption incentives.