EVON Final Report

Predicting Electric Vehicle Registrations in Ontario

Team Members: Geoff Krause, Hariharasudhan Giridharan, Ye Wang, Xuelian Xi
Faculty of Computer Science - Dalhousie University - CSCI 6409, Winter 2024

1. Problem Details

Statistics Canada

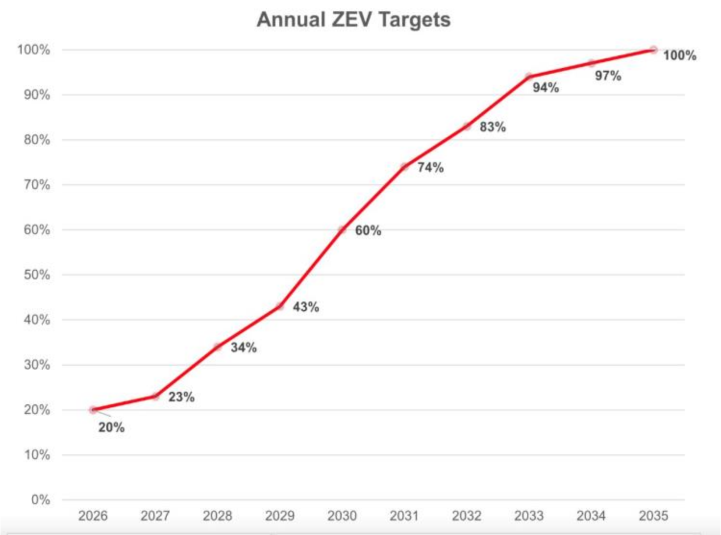
Year 2023 Q1-Q3 :
1,286,951
Registered Vehicles
100,644
BEV (Battery Electric)
7.81%
Proportion

↓

Year 2026 :
20%
Proportion

↓

Year 2035 :
100%
Proportion



01 Background Information

The Electric Vehicle (EV) market in Canada is experiencing rapid growth, with 132,783 EVs out of 1,286,951 total vehicles registered by September 2023.
Canada's climate goals target phasing out gas vehicles by 2035 and achieving a 20% EV market share by 2026。

02 Problem Statement

The adoption of EVs varies across different demographics and geographic locations, influenced by factors like income, employment, housing, and commuting patterns. **Understanding how these factors affect EV adoption is crucial to predict future growth patterns and infrastructure needs.**

03 Problem Significance

Anticipating the relationship between population characteristics and EV adoption is essential for both governmental policy-making and business strategies.

2. Project Details

Project Objectives

- ✓ **Develop a model to forecast EV adoption trends** in Ontario based on demographic and population statistics.
- ✓ Use insights to aid **organizations in strategy development and infrastructure planning.**

Empower strategic decisions

Project Scope

- ✓ Focus on Ontario due to its **large population, diverse geography, and detailed EV registration data.**
- ✓ **Utilize 2023 EV registration data and 2021 Census data,** segmented by Forward Sortation Area (FSA), to analyze EV adoption rates.

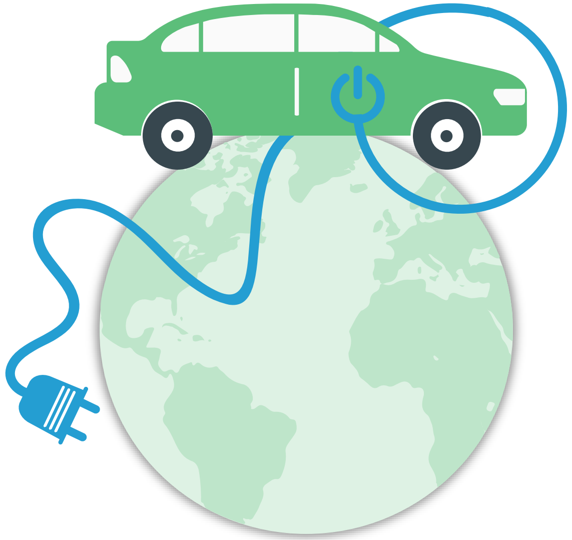
Navigate the complexities

Milestones

- ✓ **Data Collection:** Gather and preprocess 2023 EV registration data and 2021 Census data from Statistics Canada.
- ✓ **Model Development:** Test various classification models to analyze demographic influence on EV adoption rates.
- ✓ **Optimization & Validation:** Optimize model performance for accurate prediction of EV adoption categories (Low, Medium, High).
- ✓ **Insight Generation:** Provide actionable insights for policy-making, infrastructure investment, and marketing strategies .

Transforming insights into action

3. Dataset Details



Data Collection

EV Data: Obtained from Statistics Canada.

Census Data: 2021 Census Profile data from Statistics Canada, providing demographic details.

Data Stats

EV Dataset Variables: FSA (area identifier), PHEV (Plug-in Hybrid EVs), BEV (Battery EVs), TotalEV (total EVs).
550 instances for distinct geographic areas.

Census Dataset Details: Includes demographics for 1646 FSAs across Canada, with 521 in Ontario.
Selected 143 characteristics related to population demographics, resulting in 1,370,751 total rows.

Data Preparation

Normalization: Population and dwelling counts normalized; EV registrations adjusted to per 10,000 people metric.

Feature Selection: 26 features with the highest correlation to normalized TotalEV selected for analysis.

Categorical Variable Creation: TotalEV converted to a categorical variable (Low, Medium, High adoption) for classification purposes.

Quality Checks: High-quality data with minimal cleaning needed, except for normalization and removal of 4 FSAs with small or no populations.

Table 1: Variables in *Electric Vehicles in Ontario – By Forward Sortation Area* dataset

Variable Name	Type	Description
FSA	Categorical/Identifier (text)	Identifies the individual areas (instances)
PHEV	Continuous (numeric)	Hybrid EVs registered in the area
BEV	Continuous (numeric)	Battery-powered EVs registered in the area
TotalEV	Continuous (numeric)	Total EVs registered in the area

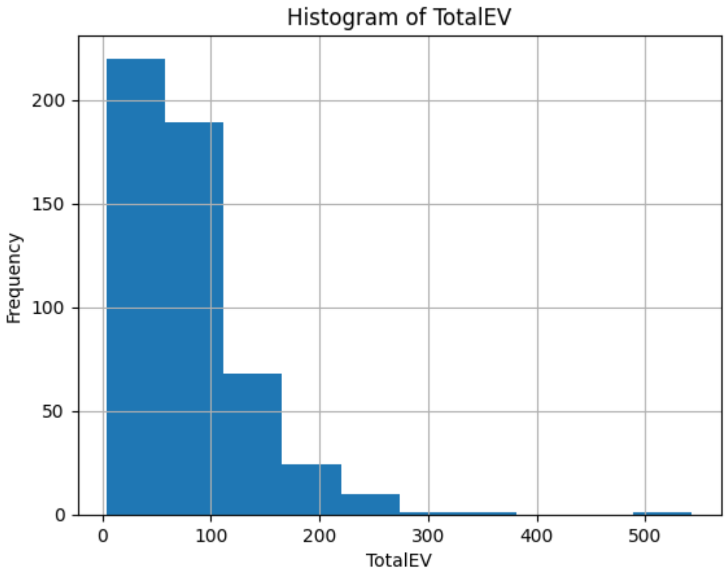
Table 2: Columns in *2021 Census Profile* dataset

Variable Name	Type	Description
GEO_NAME	Categorical/Identifier (text)	Equivalent to FSA
CHARACTERISTIC_ID	--	Identifies the variable/characteristic
CHARACTERISTIC_NAME	--	Identifies the variable/characteristic
C1_COUNT_TOTAL	Continuous (numeric)	Provides the raw data for most features
C2_COUNT_MEN+	Continuous (numeric)	Provides the raw data for the "num_males" feature

4. Dataset Insights

Key Findings:

- **EV adoption is significantly influenced by income, employment in high-paying sectors, higher education levels, and flexible commuting options.**
- **"High" adoption areas show greater variable dispersal.**
- **Some personal demographics exhibit patterns with EV adoption.**



Scatter Plot Insights (Figure 2):

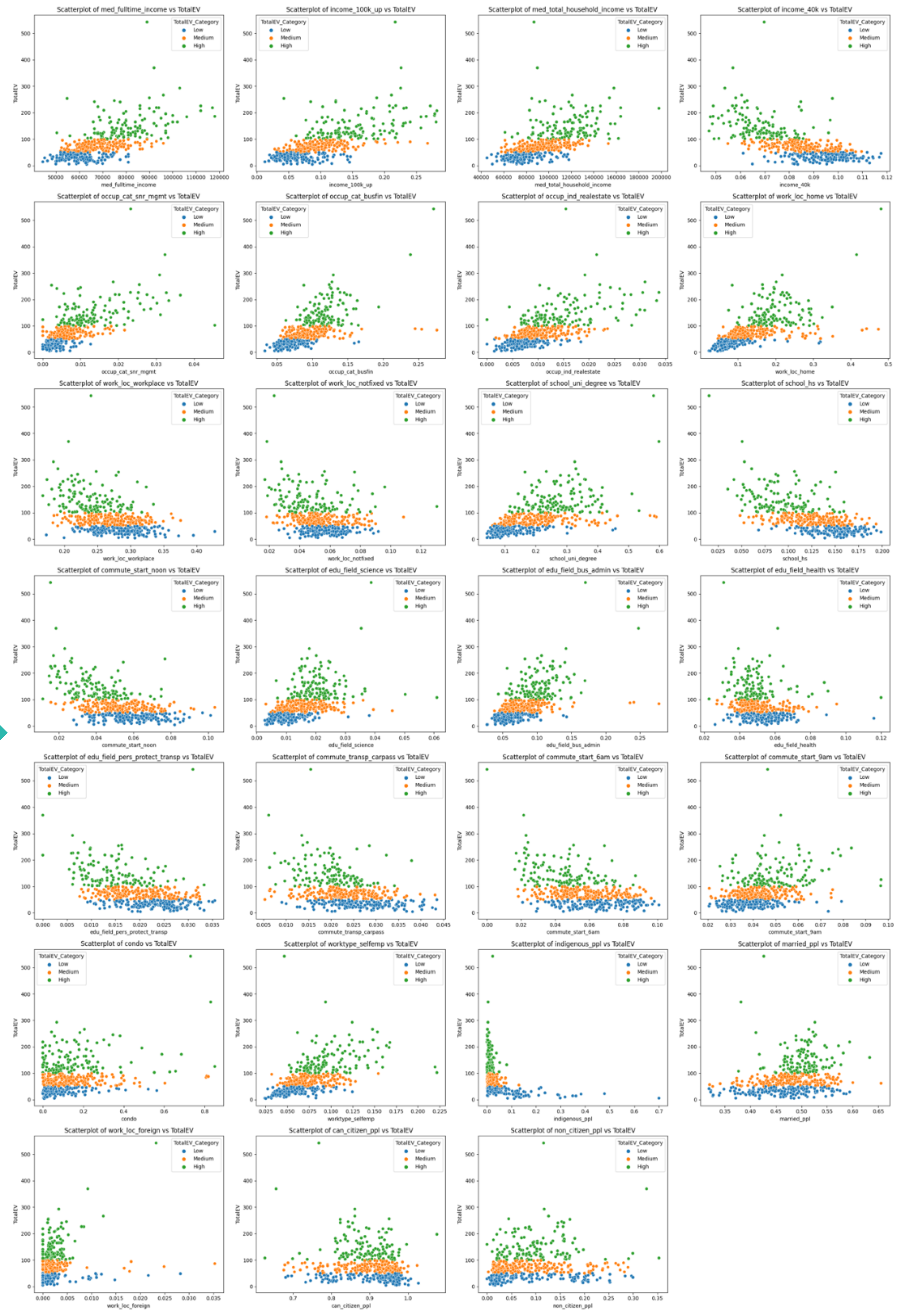
Income & Education: Strong correlation in "High" category; higher income and education levels indicate greater EV adoption.

Employment & Commuting Patterns: Higher work-from-home rates correlate with higher EV ownership. Less adoption in areas requiring early morning commutes.

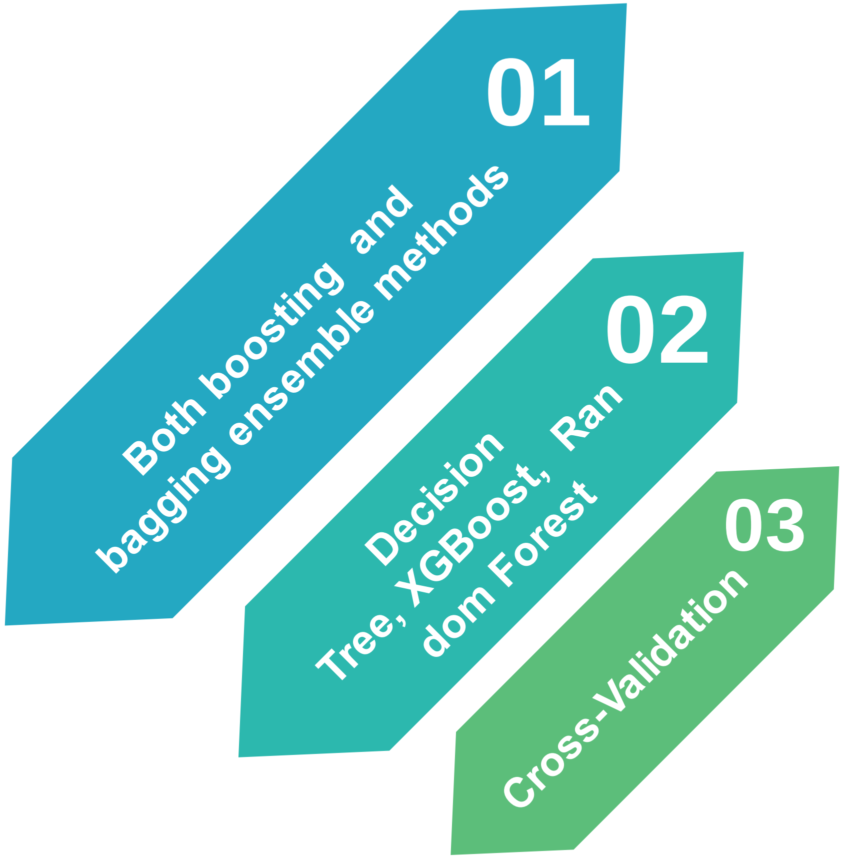
Personal Demographics: Less clear correlations with EV ownership, but notable patterns in citizenship, indigenous status, and marriage rates.

Histogram Overview (Figure 1):

- Majority of FSAs have ≤100 EVs registered per 10,000 people.
- "High" adoption threshold set at 100 EVs/10,000 people.
- "Medium" adoption threshold set at 50 EVs/10,000 people.



5. Modelling



Models Motivation

Decision Tree, XGBoost, and Random Forest were chosen for their ability to manage continuous features and a categorical target (High, Medium, Low) effectively. This diverse model selection adheres to the "No Free Lunch" theorem, **incorporating both boosting (XGBoost) and bagging (Random Forest) ensemble methods alongside a simpler, computationally efficient model (Decision Tree).**

Training and Validation Details

- **Decision Tree:** Emphasized for its simplicity and no need for feature scaling. Hyperparameters like 'max_depth', 'min_samples_split', and 'min_samples_leaf' were fine-tuned using GridSearchCV to enhance performance and prevent overfitting.
- **XGBoost:** Chosen for its proficiency in handling complex interactions through gradient boosting. Parameters including learning rate and 'n_estimators' optimized to surpass Decision Tree performance.
- **Random Forest:** Selected for its robustness and high-dimensional data handling, with hyperparameters like 'max_depth' and 'n_estimators' optimized for competitive performance.

Evaluation Metrics

- **Validation Accuracy Results:** Decision Tree (0.7020), XGBoost (0.7938), Random Forest (0.84), showcasing Random Forest's superior performance.
- **Metrics Used:** Accuracy, precision, recall, F1-score, and confusion matrices to assess model performance across different classification accuracies and error rates.
- **Cross-Validation:** Employed k-fold cross-validation to mitigate overfitting and validate model reliability, with average performance across all folds determining final evaluation.

	Hyper-parameter1	Hyper-parameter	Hyper-parameter3	Hyper-parameter4	Validation Accuracy
Decision Tree	max_depth 3	min_samples_leaf 4	min_samples_split 2		0.7020
XG Boost	learning_rate 0.1	max_depth 4	n_estimators 100		0.7938
Random Forest	max_depth 10	min_samples_leaf 1	min_samples_split 10	n_estimator 50	0.84

Table M: Best Hyperparameters of Decision Tree, XG Boost and Random Forest models by Grid Search Crossvalidation

6. Results

	Accuracy	Precision (MacroAverage)	Recall (MacroAverage)	F1-Score (MacroAverage)
Decision Tree	0.7290	0.74	0.71	0.72
XGBoost	0.8323	0.83	0.83	0.83
Random Forest	0.8387	0.83	0.83	0.83

Table X: Performance evaluation of Decision Tree, XG Boost and Random Forest models based on macro average.

	Accuracy	Precision (WeightedAverage)	Recall (WeightedAverage)	F1-Score (WeightedAverage)
Decision Tree	0.7290	0.74	0.73	0.73
XGBoost	0.8323	083	0.83	0.83
Random Forest	0.8387	0.84	0.84	0.84

Table Y: Performance evaluation of Decision Tree, XG Boost and Random Forest models based on weighted average.

-	Precision	Recall	F1-Score
Decision Tree - High	0.74	0.59	0.66
Decision Tree - Low	0.86	0.80	0.83
Decision Tree - Medium	0.63	0.75	0.69
XGBoost - High	0.79	0.79	0.79
XGBoost - Low	0.85	0.96	0.91
XGBoost - Medium	0.83	0.74	0.78
Random Forest - High	0.81	0.77	0.79
Random Forest - Low	0.88	0.96	0.92
Random Forest - Medium	0.81	0.77	0.79

Table Z: Performance evaluation on target classes of Decision Tree, XG Boost and Random Forest models.

Model Performance Summary

- **Accuracy Highlights**:
  - **Decision Tree**: Moderate (0.7290).
  - **XGBoost**: Strong (0.8323).
  - **Random Forest**: Strongest (0.8387).
- **F1-Score for "Low" Class**:
  - Exceptionally high in XGBoost (0.91) & Random Forest (0.92).

Model Comparison Insights

Optimal Model: Random Forest, due to its superior accuracy and stability.

- **Performance Trends**:
- **Decision Tree**: Good generalization.
- **XGBoost**: High training scores but overfitting signs.
- **Random Forest**: Best generalization and stable performance.
- **Key Takeaways**:
- Random Forest and XGBoost excel in predicting low EV adoption scenarios.
- Challenges remain in accurately predicting "High" and "Medium" categories.

7. Key Findings

01 Interpretation of Results

Random Forest stands out with the highest accuracy (0.8387) and F1-scores, especially for the "Low" class (F1-score: 92%).

02 Implications

Ensemble methods like Random Forest and XGBoost overcome limitations of basic algorithms, enhancing predictions in small datasets.

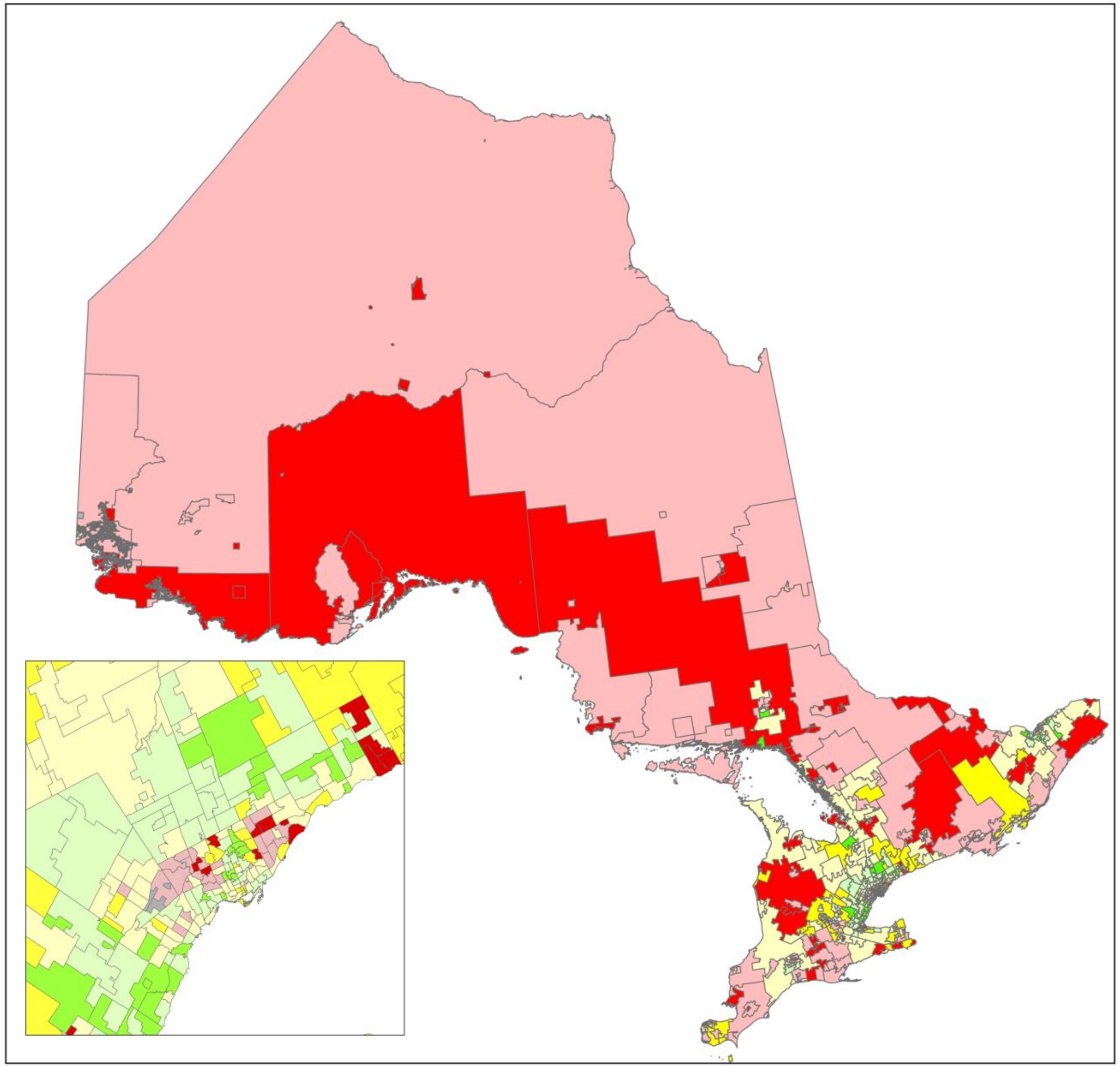
03 Limitations

Random Forest's performance dips for "High" and "Medium" EV classes.
Data limitations could impact prediction accuracy.

04 Future Strategies

Incorporate multi-year trends and detailed geographic data for dynamic and comprehensive regional EV adoption analysis.

8. Conclusion



Province of Ontario

Summary of Findings

Regions with higher income levels and population densities are more likely to register electric vehicles (EVs). This trend is reflecting a strong correlation between economic status and EV adoption.

Achievement of Objectives

- The project effectively achieved its objectives by developing a predictive model that can **accurately forecast EV adoption trends in Ontario based on key demographic and regional factors.**
- The model successfully identified **possible factors corresponding to EV adoption,** such as income, education, and workplace location.
- The research may be useful in **suggesting areas for governmental and corporate action.**

9. References

References

Data & Software

Emerging Technologies Office. (2024). *Electric Vehicles in Ontario – By Forward Sortation Area—Q4 2023* [CSV]. Ontario Data Catalogue. https://data.ontario.ca/dataset/electric-vehicles-in-ontario-by-forward-sortation-area/resource/dca5bef6-df38-4c45-be73-62e71b243d3d

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research, 12*(85), 2825–2830. https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html

Statistics Canada. (2022, February 9). *Census Profile Downloads, 2021*. https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/details/download-telecharger.cfm?Lang=E

Referenced Works

Blair, N. (2024, January). *Electric Vehicle Adoption Statistics in Canada*. Made in CA. https://madeinca.ca/electric-vehicle-adoption-statistics-canada/

Chen, C. F., de Rubens, G. Z., Noel, L., Kester, J., & Sovacool, B. K. (2020). Assessing the socio-demographic, technical, economic and behavioral factors of Nordic electric vehicle adoption and the influence of vehicle-to-grid preferences. *Renewable and Sustainable Energy Reviews, 121*, 109692. https://www.sciencedirect.com/science/article/abs/pii/S1364032119308974

Cousins, B. (2023, December). *What does 2024 have in store for the EV industry?* Bloomberg BNN. https://www.bnnbloomberg.ca/what-does-2024-have-in-store-for-the-ev-industry-1.2014925

Egbue, O., Long, S., & Samaranayake, V. A. (2017). Mass deployment of sustainable transportation: evaluation of factors that influence electric vehicle adoption. *Clean Technologies and Environmental Policy, 19*, 1927-1939. https://link.springer.com/article/10.1007/s10098-017-1375-4

Wang, Y., & Hewitt, E. L. (2019, October). Plug-in Electric Vehicle (PEV) Adoption in US Transport for Policy. In *2019 International Energy and Sustainability Conference (IESC)* (pp. 1-21). IEEE. https://ieeexplore.ieee.org/abstract/document/8976770

THANK YOU

Any questions?