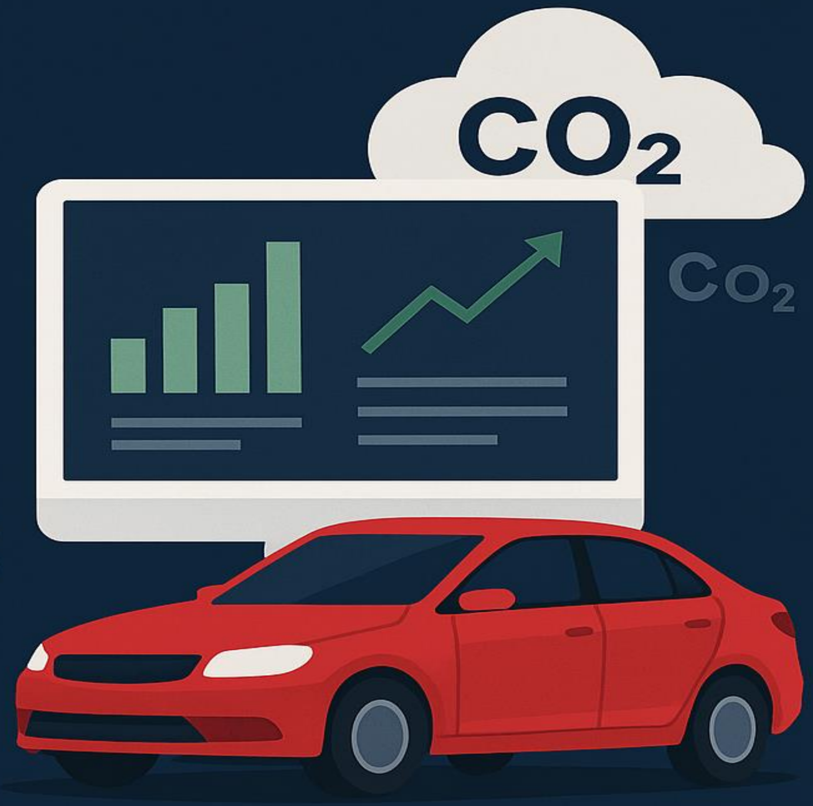
OPTIMIZING VEHICLE CHOICES: A DATA-DRIVEN APPROACH TO FUEL CONSUMPTION AND CO_2 EMISSIONS IN CANADA



By:
Jani Begam Zahir Hussain

INTRODUCTION

Problem Statement

Today's consumers rely on static fuel efficiency labels like EnerGuide, which provide only generic, lab-tested ratings. These labels don't consider:

- Real-Long-term fuel costs
- Environmental penalties like carbon tax
- Personalized vehicle usage behavior

This gap leaves consumers under-informed, policymakers under-prepared, and manufacturers out of tune with emissions realities.

Objective

To develop a machine learning-powered web dashboard that dynamically predicts:

- CO_2 emissions (g/km)
- Annual fuel cost (CAD)
- Eco and fuel efficiency scores
- Estimated carbon tax impact

based on user-selected vehicle configurations and up-to-date fuel pricing.

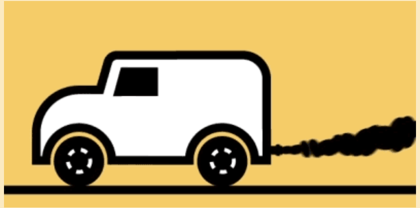
Significance

This solution transforms how vehicle choices are made in Canada by offering:

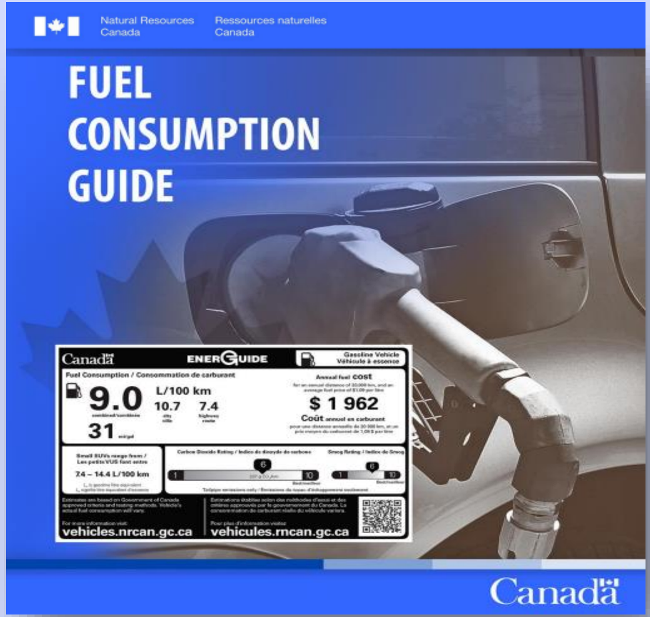
- **Consumers:** Personalized insights that go beyond marketing and labels—enabling informed, cost-effective, and environmentally conscious decisions.
- **Policymakers:** Accurate emissions forecasts and cost models to guide carbon tax policy, rebates, and eco-incentives.
- **Automakers:** Data-driven signals for design improvements and compliance with sustainability targets.

It's more than just picking a car. It's empowering a nation to choose sustainability with confidence.

"What if choosing your next car wasn't just about style or speed—but about saving money and the planet, all powered by data?"



DATA OVERVIEW



Data Source:

Natural Resources Canada (NRCan) – *Fuel Consumption Ratings Dataset*, provided by the Government of Canada.

A trusted and publicly available resource under Canada's Open Data initiative.

Important Factors:

1. Identification of the Vehicle:

- Make
- Model
- Model Year
- Vehicle Class

2. Performance Details:

- Fuel Type (e.g., Gasoline, Diesel, E85, Electric)
- Transmission Type (e.g., Automatic, Manual, CVT, Electric Drive)
- Engine Size (in Litres)

3. Measures of Efficiency:

- City, Highway, and Combined Fuel Consumption (L/100 km)
- CO_2 Emissions (grams/km)
- Estimated Annual Fuel Cost (in CAD)

Type of Data:

- Structured tabular dataset, Regression Dataset
- Designed for descriptive and exploratory analysis
- Suitable for regression modeling and predictive insights

Dataset Size:

- 8,296 vehicle records with 15 features.
- Covers model years from 2015 to 2025
- Clean and ready for analysis with no missing values

Exploratory Data Analysis (EDA)

Pandas Profiling:
- High correlations between CO_2 emissions and engine size, cylinders, and fuel consumption.
- Fuel type patterns vary notably across brands.
- Some fields like Model need additional cleaning due to inconsistency.
- Dataset is clean: no missing values detected.

Histogram Analysis
- **Engine Size**: Right-skewed; most engines fall between 1–3L.
- **Cylinders**: Peaks at 4 & 6 cylinders; higher counts are rare.
- **Fuel Consumption**: Most vehicles consume 5–15 L/100 km.
- **CO_2 Emissions**: Majority emit 100–300 g/km, with high outliers >500 g/km.
- **MPG**: Left-skewed; most cars are moderately efficient.

Action: Outliers flagged for removal; skewed features normalized.

Shapiro-Wilk Normality Test
- None of the numerical features follow a normal distribution ($p < 0.05$).
- Fuel consumption and emissions are highly skewed due to outliers.
- CO_2 rating and smog rating show non-normal behavior from categorical-like distribution.
- Suggests use of non-parametric methods or transformation for modeling.

Action: Applied RobustScaler for normalization to mitigate the effect of skewed distributions and enhance model performance.

Q-Q Plot Observations
- Most features deviate significantly from the diagonal line, confirming non-normality.
- Model Year, Cylinders, Ratings: Step-like patterns indicate discrete values.
- Fuel Consumption & Emissions: Skewed tails highlight presence of extreme values/outliers.
- CO_2 Emissions: Slight skew at high values, mostly aligns with normal line.

Feature Selection
- Employed **Boruta algorithm** to identify **top influential features** for CO_2 emissions.
- Cross-validated with **VIF analysis** and correlation matrix.

Action: Retained only statistically and practically relevant predictors, ensuring low redundancy and high model accuracy.

Box Plots
- **Model Year:** Left-skewed; most vehicles are from 2020–2025.
- **Engine Size & Cylinders:** Right-skewed; 4–6 cylinders are most common.
- **Fuel Consumption (City, Highway, Combined):** Right-skewed; a few heavy-duty outliers.
- **MPG (Miles Per Gallon):** Left-skewed; most vehicles cluster around 25–30 mpg.
- **CO_2 Emissions:** Right-skewed with significant outliers.
- **Smog Rating:** More symmetric; centered around 5.

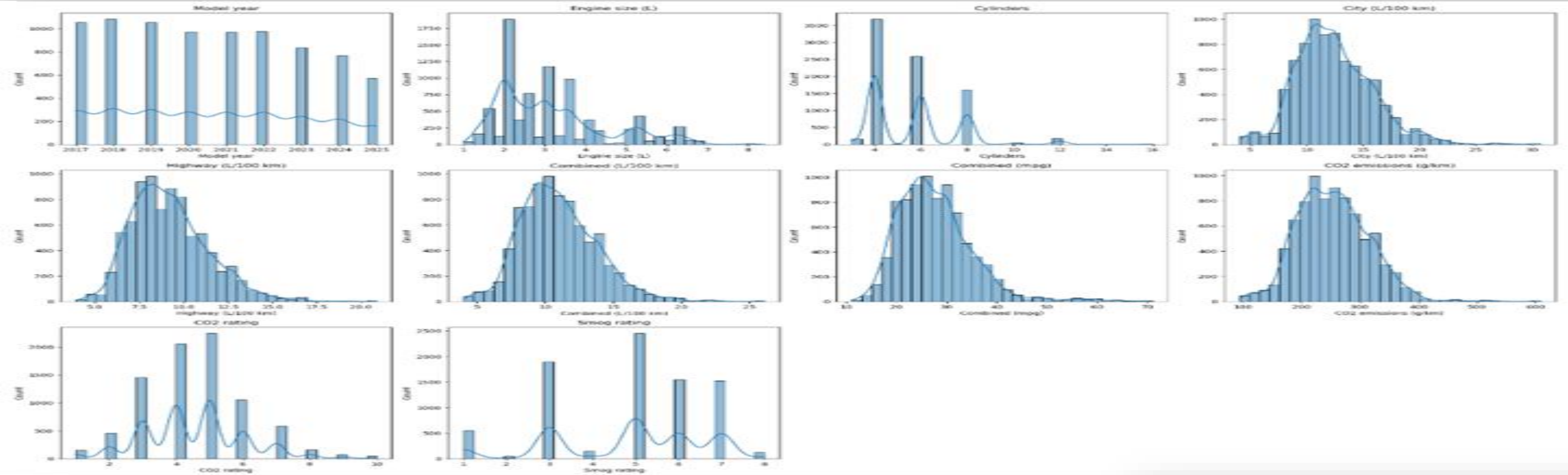
Conclusion: Most fuel and emission-related features show skewness and outliers, justifying normalization and outlier treatment.

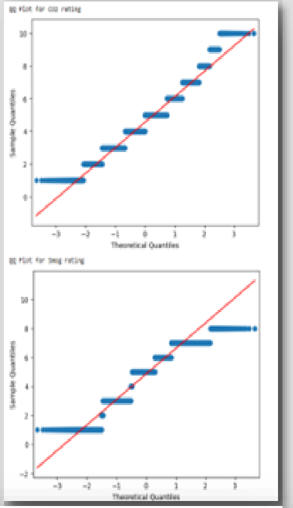
Spearman Correlation
- **22 strong correlations identified ($\rho > 0.8$),** confirming multicollinearity among key features.
- Fuel Efficiency Metrics (City, Highway, Combined): **Highly correlated ($\rho > 0.95$)**
- Combined vs. City Fuel Consumption → **0.99**
- Highway vs. Combined Fuel Consumption → **0.97**
- **Fuel Use vs. Emissions:**
- Combined Fuel Consumption ↔ CO_2 Emissions → **0.96**
- City Fuel Consumption ↔ CO_2 Emissions → **0.95**
- **Engine & Design Features:**
- Engine Size ↔ CO_2 Emissions → **0.84**
- Cylinders ↔ CO_2 Emissions → **0.83**
- **Environmental Ratings:**
- CO_2 Rating ↔ Emissions → **−0.97 (strong inverse)**
- Smog Rating ↔ Emissions → **−0.46 (moderate inverse)**
- Clear multicollinearity exists among fuel consumption metrics.

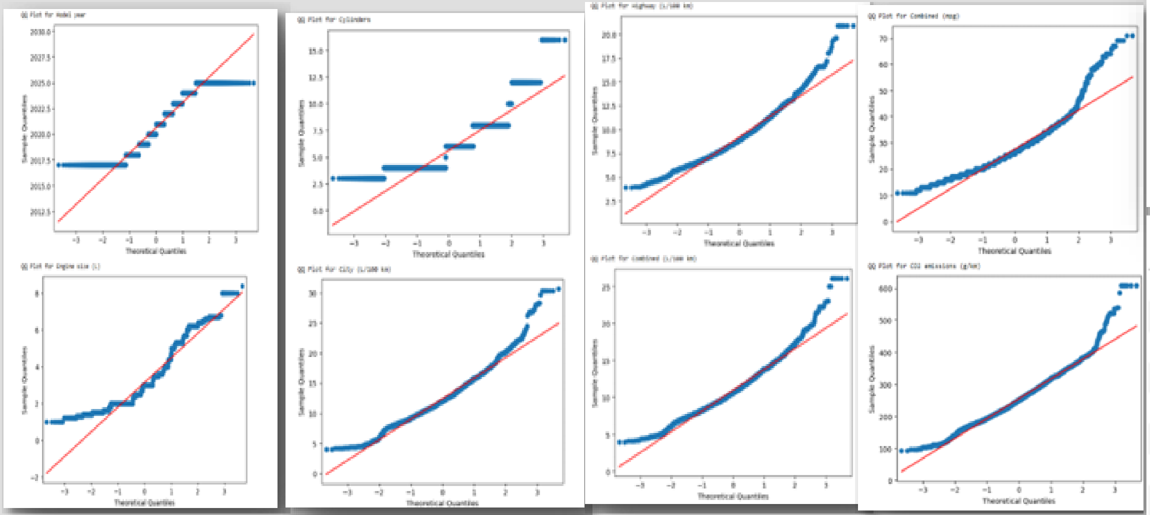
  Action: To address multicollinearity, dropped City and Highway fuel consumption features and retained Combined for model input and downstream calculations.

Outlier Detection & Handling
- Implemented Isolation Forest algorithm with 5% contamination.
- Removed ~300 outliers (e.g., cars with engine sizes >7L or >10 passengers).

Action: Improved model robustness and reduced noise; training dataset trimmed to 6,305 records for better generalization.

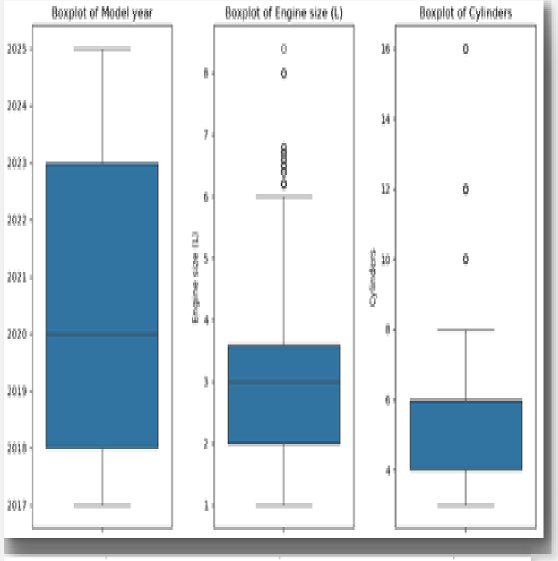


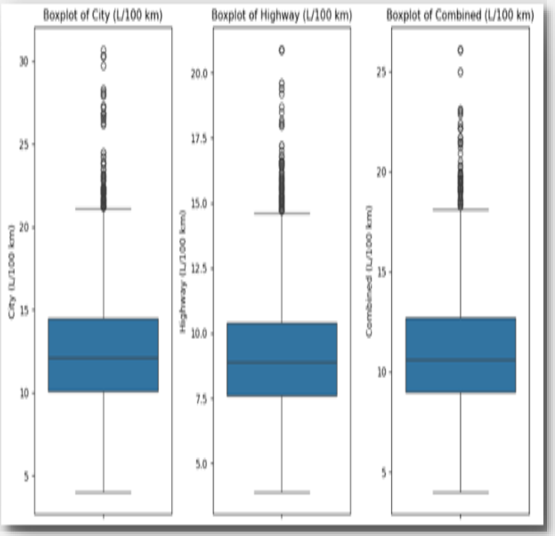


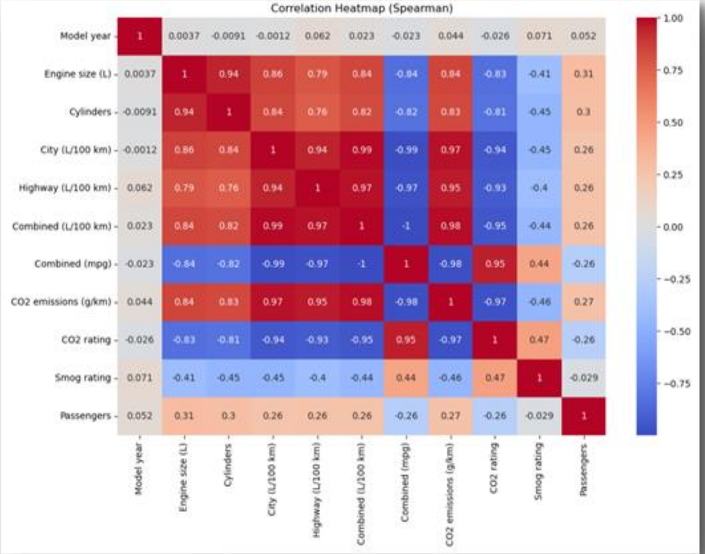


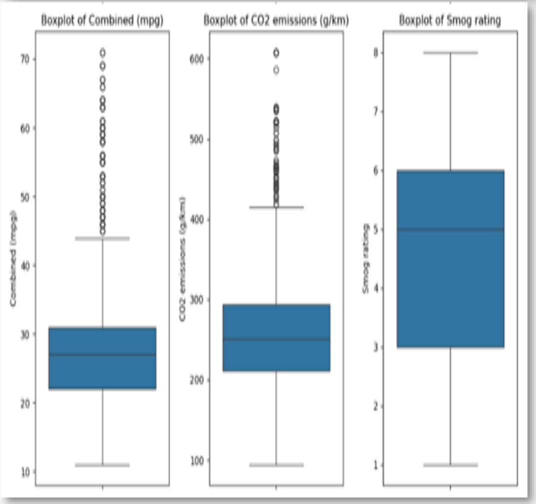

```
BorutaPy finished running.

Iteration: 	100 / 100
Confirmed: 	10
Tentative: 	2
Rejected: 	17

Final Selected Features:
['Model year', 'Engine size (L)', 'City (L/100 km)', 'Highway (L/100 km)', 'Combined (L/100 km)', 'Combined (mpg)', 'CO2 rating', 'Smog rating', 'Transmission_Automatic', 'Fuel type_E']
```

	Model year	Engine size (L)	Cylinders	City (L/100 km)	Highway (L/100 km)	Combined (L/100 km)	Combined (mpg)	CO2 emissions (g/km)	CO2 rating	Smog rating
count	8296.000000	8296.000000	8296.000000	8296.000000	8296.000000	8296.000000	8296.000000	8296.000000	8296.000000	8296.000000
mean	2020.599807	3.131666	5.610776	12.411656	9.193672	10.963175	27.520251	254.732763	4.550506	4.839320
std	2.477466	1.340042	1.903399	3.411809	2.187367	2.813596	7.512694	61.781550	1.556446	1.755687
min	2017.000000	1.000000	3.000000	4.000000	3.900000	4.000000	11.000000	94.000000	1.000000	1.000000
25%	2018.000000	2.000000	4.000000	10.100000	7.600000	9.000000	22.000000	211.000000	4.000000	3.000000
50%	2020.000000	3.000000	6.000000	12.100000	8.900000	10.600000	27.000000	251.000000	5.000000	5.000000
75%	2023.000000	3.600000	6.000000	14.500000	10.400000	12.700000	31.000000	294.000000	5.000000	6.000000
max	2025.000000	8.400000	16.000000	30.700000	20.900000	26.100000	71.000000	608.000000	10.000000	8.000000









(121, 3)
(55, 3)

	var1	var2	corr_value	corr_abs
58	Combined (L/100 km)	City (L/100 km)	0.991093	0.991093
49	Highway (L/100 km)	Combined (L/100 km)	0.967883	0.967883
62	Combined (L/100 km)	CO2 emissions (g/km)	0.956808	0.956808
40	City (L/100 km)	CO2 emissions (g/km)	0.951233	0.951233
95	CO2 rating	CO2 emissions (g/km)	-0.946069	0.946069
74	Combined (mpg)	CO2 rating	0.933674	0.933674
47	Highway (L/100 km)	City (L/100 km)	0.926290	0.926290
13	Engine size (L)	Cylinders	0.922445	0.922445
51	Highway (L/100 km)	CO2 emissions (g/km)	0.920661	0.920661
71	Combined (mpg)	Combined (L/100 km)	-0.920455	0.920455

```
Model year                          int64
Engine size (L)                   float64
Cylinders                           int64
City (L/100 km)                   float64
Highway (L/100 km)                float64
Combined (L/100 km)               float64
Combined (mpg)                      int64
CO2 emissions (g/km)                int64
CO2 rating                          int64
Smog rating                         int64
Passengers                        float64
Make_Ferrari                        int32
Make_Ford Motor Company             int32
Make_General Motors                 int32
Make_Honda                          int32
Make_Hyundai Motor Group            int32
Make_Mercedes-Benz Group            int32
Make_Nissan-Renault Alliance        int32
Make_Stellantis                     int32
Make_Tata                           int32
Make_Toyota                         int32
Make_Volkswagen Group               int32
Make_Volvo                          int32
Transmission_Automatic              int32
Transmission_CVT                    int32
Transmission_Dual-Clutch            int32
Transmission_Manual                 int32
Fuel type_E                         int32
Fuel type_X                         int32
Fuel type_Z                         int32
dtype: object
```

Variance Inflation Factor (VIF) for Features:

	VIF Factor	Feature
5	69401.615822	Combined (L/100 km)
3	27221.943271	City (L/100 km)
4	10011.775051	Highway (L/100 km)
0	641.844970	Model year
6	196.627944	Combined (mpg)
7	139.816374	CO2 rating
2	106.188523	Cylinders
1	69.781035	Engine size (L)
28	29.392850	Fuel type_Z
27	27.805361	Fuel type_X
9	19.708926	Passengers
8	14.488072	Smog rating
12	4.426812	Make_General Motors
22	4.217121	Transmission_Automatic
26	4.032055	Fuel type_E
17	3.438523	Make_Stellantis
19	3.176253	Make_Toyota
11	3.002253	Make_Ford Motor Company
20	2.862468	Make_Volkswagen Group
15	2.409914	Make_Mercedes-Benz Group
14	2.172274	Make_Hyundai Motor Group
23	2.108025	Transmission_CVT
24	1.986935	Transmission_Dual-Clutch
16	1.954462	Make_Nissan-Renault Alliance
13	1.687611	Make_Honda
25	1.475923	Transmission_Manual
18	1.437744	Make_Tata
21	1.184554	Make_Volvo
10	1.045263	Make_Ferrari

Data Transformation

Data Preparation & Encoding Overview

- ✓ **Passenger Mapping:** Most cars have seating capacities matching their type, improving dataset consistency.
- ✓ **Handling Special Cases:** Special-purpose vehicles assigned specific values, enhancing accuracy.
- ✓ **Cleaned Data Structure:** Reduced redundancy by removing the **Vehicle class** column, resulting in **31 columns** and **8,296 rows**.
- ✓ **Parent Firms:** Parent companies like **Toyota** and **Volkswagen Group** merged under the **Make** column.
- ✓ **Transmission Types:** Streamlined into **CVT, Dual-Clutch, Automatic, Manual, Automated Manual** categories.
- ✓ **Fuel Types:** Categorized into **E, X, Z, D** for easier analysis.
- ✓ **Machine Learning Compatibility: One-Hot Encoding** used to convert categorical data into numerical form, adding binary columns for brands, transmission types, and fuel types.
- ✓ **Numerical Features:** Key characteristics like **mileage, CO_2 emissions, fuel consumption**, and **engine size** remained unchanged.

Data Types (Post-Transformation)

- ✓ **Categorical Columns:** Vehicle Make, Transmission Type, Fuel Type encoded as **binary (0 or 1)** using **int32** data type.
- ✓ **Numerical Columns:** Fuel Consumption, Engine Size, CO_2 Emissions represented as continuous data with **int64** and **float64** data types.
- ✓ **Summary:** The dataset includes both categorical and numerical variables, with categorical variables encoded as binary and numerical variables represented as continuous values.

```
Final dataset shape: (8296, 31)
   Model year               Model  Engine size (L)  Cylinders  City (L/100 km)
0        2017                 ILX              2.4          4              9.4
1        2017      MDX Hybrid AWD              3.0          6              9.1
2        2017          MDX SH-AWD              3.5          6             12.6
3        2017    MDX SH-AWD Elite              3.5          6             12.2
4        2017                 NSX              3.5          6             11.1

   Highway (L/100 km)  Combined (L/100 km)  Combined (mpg)  \
0                 6.8                  8.2              34
1                 9.0                  9.0              31
2                 9.0                 11.0              26
3                 9.0                 10.7              26
4                10.8                 11.0              26

   CO2 emissions (g/km)  CO2 rating  ...  Make_Toyota  Make_Volkswagen Group
0                   192           7  ...            0                      0
1                   210           6  ...            0                      0
2                   259           4  ...            0                      0
3                   251           5  ...            0                      0
4                   261           4  ...            0                      0

   Make_Volvo  Transmission_Automatic  Transmission_CVT  \
0           0                       0                 0
1           0                       0                 0
2           0                       0                 0
3           0                       0                 0
4           0                       0                 0

   Transmission_Dual-Clutch  Transmission_Manual  Fuel type_E  Fuel type_X  \
0                         1                    0            0            0
1                         1                    0            0            0
2                         0                    0            0            0
3                         0                    0            0            0
4                         1                    0            0            0

   Fuel type_Z
0            1
1            1
2            1
3            1
4            1
```

Modeling Approach

Methods Overview:

We evaluated a variety of machine learning models and finalized the following three for our regression task of predicting CO_2 emissions:

1. **Linear Regression**
2. **Ridge Regression**
3. **XGBoost Regressor**

Rationale for Selection:

Linear Regression

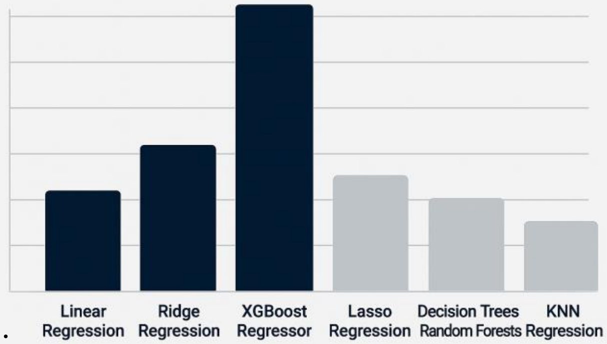
- Chosen as a **baseline model** for its simplicity and interpretability, assuming a linear relationship between independent variables and CO_2 emissions.
- Helps establish a performance benchmark for more advanced models.
- Easily explains relationships between input variables and the target,

Ridge Regression

- Ideal for datasets with **multicollinearity** (as confirmed by Spearman correlation & VIF).
- Adds **L2 regularization**, which penalizes large coefficients and prevents overfitting.
- Improves model generalization while retaining interpretability.

XGBoost Regressor

- Selected as a **robust non-linear model** to capture complex relationships in the data.
- Performs automatic feature interaction and handles skewed or noisy data well.
- Known for **high accuracy, speed, and regularization** capabilities.
- Performs better on datasets with a **mix of numerical and encoded categorical features**.



Models Which Were Not Chosen

- **Lasso Regression** was tested but overly shrunk important coefficients due to L1 regularization.
- **Decision Trees** and **Random Forests** offered less consistent performance and interpretability compared to XGBoost.
- **KNN Regression** was avoided due to computational inefficiency on larger datasets and poor handling of high-dimensional encoded features.

Model Outputs

Model Performance:

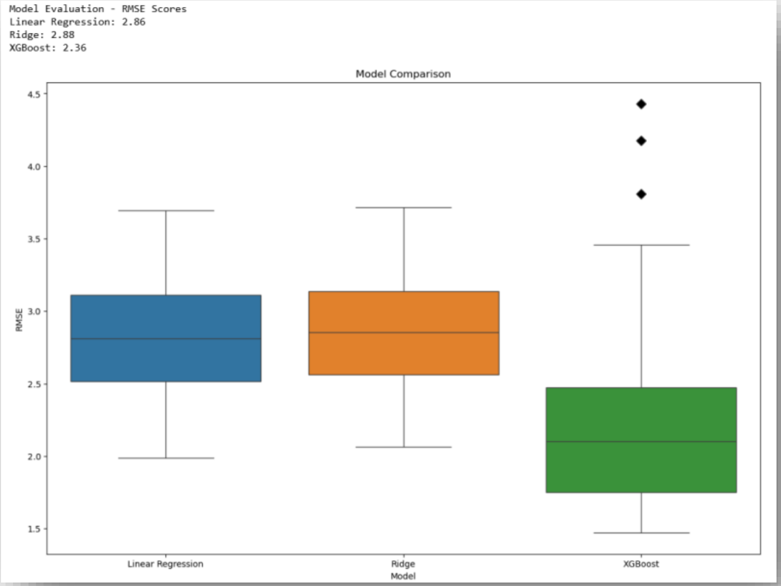
- ✓ **Linear Regression:** Although Linear Regression achieves a high test R^2 score of 0.9966, it lacks regularization, which makes it potentially sensitive to multicollinearity or noise in the data. This can lead to less stable performance in real-world applications compared to models like Ridge, which are specifically designed to handle such issues.
- ✓ **Ridge:** It performs similarly to Linear Regression, but with added regularization to prevent overfitting. In this regularization had little effect due to the dataset's fit.
- ✓ **XGBoost:** It achieved a high R^2 score but with slightly higher RMSE. However, its MAE is lower than Linear and Ridge Regression, indicating it makes smaller errors on average

Learning Curves:

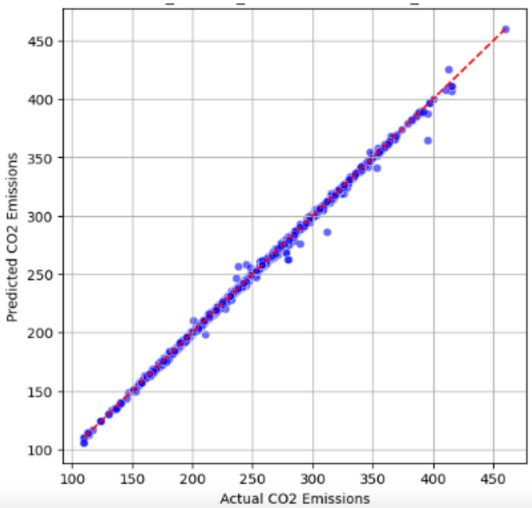
The learning curves, plotted after hyperparameter tuning, help visualize model performance and learning behavior, indicating whether more data could improve the model or if the model is underfitting/overfitting.

- ✓ **Linear Regression:** Training and validation RMSE are very close and remain relatively flat across all sample sizes.
- ✓ **Ridge:** Similar to Linear Regression, but with slightly improved generalization
- ✓ **XGBoost:** Training RMSE is much lower than validation RMSE; validation RMSE decreases with more data.

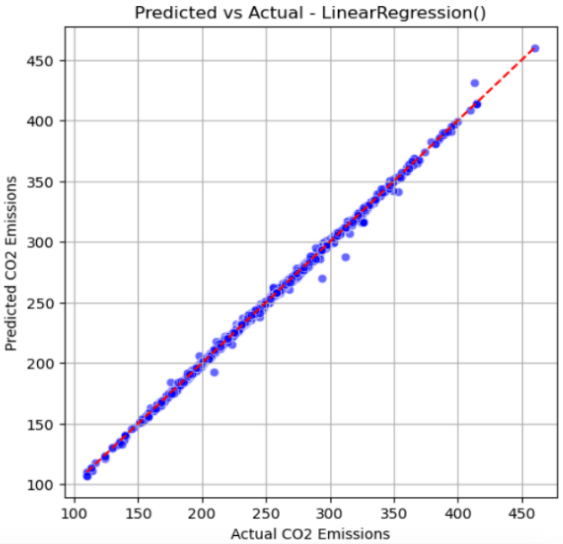
Metric	Linear Regression	Ridge Regression	XGBoost
Train R^2	0.9987	0.9987	0.9997
Test R^2	0.9966	0.9966	0.9959
Train RMSE	2.0095	2.0099	0.9146
Test RMSE	3.6252	3.6435	3.9703
Train MAE	1.1608	1.1638	0.6277
Test MAE	1.5447	1.5511	1.3001



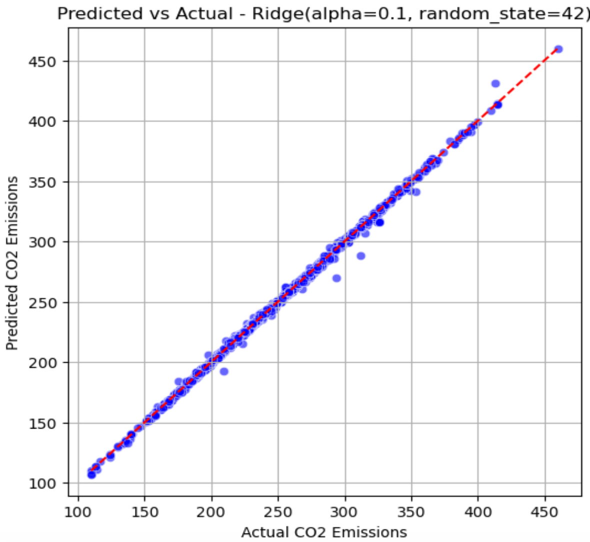
```
Running GridSearchCV for XGBoost

 ***Optimized Model: XGBRegressor(base_score=None, booster=None, callbacks=None,
             colsample_bylevel=None, colsample_bynode=None,
             colsample_bytree=None, device=None, early_stopping_rounds=None,
             enable_categorical=False, eval_metric=None, feature_types=None,
             gamma=None, grow_policy=None, importance_type=None,
             interaction_constraints=None, learning_rate=0.2, max_bin=None,
             max_cat_threshold=None, max_cat_to_onehot=None,
             max_delta_step=None, max_depth=5, max_leaves=None,
             min_child_weight=None, missing=nan, monotone_constraints=None,
             multi_strategy=None, n_estimators=300, n_jobs=None,
             num_parallel_tree=None, random_state=42, ...)***
Best Parameters: {'clf__learning_rate': 0.2, 'clf__max_depth': 5, 'clf__n_estimators': 300}
R2 Score: 0.999
RMSE: 2.112
MAE: 1.035
```

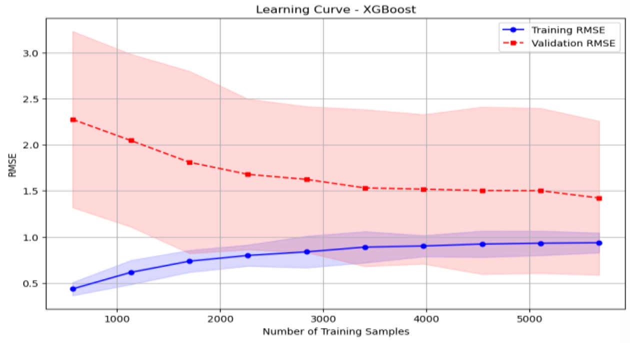


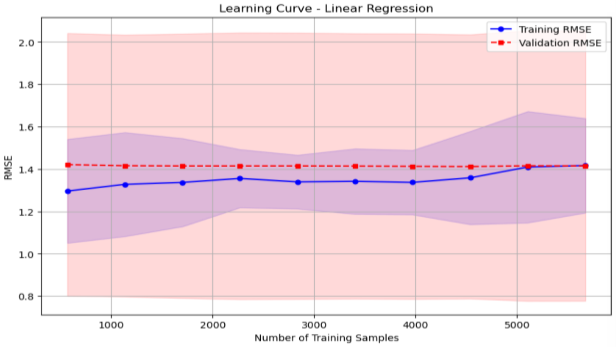

```
Running GridSearchCV for Linear Regression

 ***Optimized Model: LinearRegression()***
Best Parameters: {}
R2 Score: 0.999
RMSE: 2.001
MAE: 1.162
```

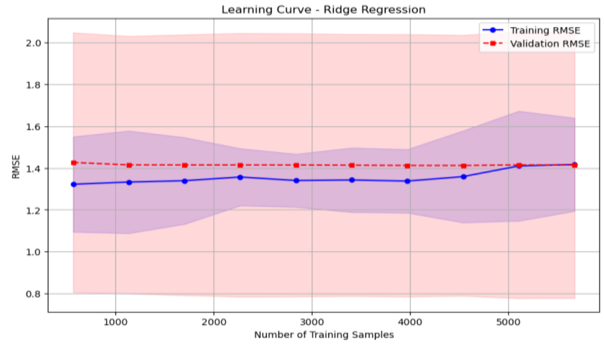



```
Running GridSearchCV for Ridge

 ***Optimized Model: Ridge(alpha=0.1, random_state=42)***
Best Parameters: {'clf__alpha': 0.1}
R2 Score: 0.999
RMSE: 2.002
MAE: 1.168
```









Final Model Selection - XGBoost

Selected Model- XGBoost

- Captures complex, non-linear relationships in the dataset.
- Outperformed Ridge and Linear Regression in **Mean Absolute Error (MAE)**.
- Handles skewed data, multicollinearity, and categorical variables effectively.
- Robust performance after **outlier removal** and **hyperparameter tuning**.

Final Model Performance (Test Set)

- **R^2 Score**: 0.9959
- **RMSE**: 3.97
- **MAE**: 1.30
- **Train RMSE**: 0.91 (High accuracy with controlled overfitting)

Key Takeaways

- XGBoost provides a **balance of accuracy and feature importance**.
- Slight variance, but better at generalization after tuning.
- Chosen for **deployment in the live dashboard** for real-time predictions.

Final XGBoost Model Outputs

Explanation of Model Performance Metrics:

- **R^2 Score (0.9959)**
  Indicates that 99.59% of the variance in CO_2 emissions is explained by the model. Near-perfect fit.
- **RMSE (3.97)**
  Root Mean Squared Error shows the average deviation of predictions from actual values. Lower is better.
- **MAE (1.30)**
  Mean Absolute Error measures average magnitude of errors in predictions. XGBoost had the **lowest MAE** among all models.
- **Train RMSE (0.91)**
  Indicates excellent learning on the training data. Confirms the model has high predictive accuracy.

*XGBoost Model Equation:**

CO_2 Emissions (g/km) ≈
0.0000 + (0.8918 × **CO_2 Rating**) + (0.0628 × **Combined Fuel Consumption**) +
(0.0345 × **Fuel type_E**) + (0.0032 × **Make_General Motors**) +
(0.0022 × **Fuel type_X**) + (0.0012 × **Transmission_Automatic**) +
(0.0007 × **Make_Tata**) + (0.0006 × **Smog Rating**) + (0.0005 × **Fuel type_Z**) +
(0.0005 × **Engine Size**) + (0.0004 × **Model Year**) +
(0.0003 × **Make_Hyundai**) + (0.0003 × **Cylinders**) +
(0.0002 × **Make_Ford**) + (0.0001 × **Make_Toyota**) +
(0.0001 × **Make_Mercedes**) + (0.0001 × **Transmission_Dual-Clutch**)

Practical Application

Empowering Smarter Vehicle Choices

This model simplifies vehicle selection by helping consumers identify cars with lower fuel costs and CO_2 emissions based on their unique driving needs. By using our interactive carbon reduction efforts dashboard, users can:

- Make informed, eco-friendly decisions
- Reduce long-term fuel expenses
- Contribute to Canada's carbon reduction efforts

Even a 5% drop in emissions could lead to significant savings and improved air quality, supporting a greener lifestyle for all Canadians.

Faster Decisions with Forecasting

Consumers no longer have to wait for future models to assess performance:

- The dashboard can **predict fuel cost and emission impact** of unreleased vehicles based on their specifications
- Suggests **similar existing vehicles**, providing instant, data-driven results
- Enhances **buyer confidence, saves time**, and promotes responsible purchasing decisions

Insights for Automakers

Automakers can use the dashboard's visual analytics to:

- Understand **which configurations reduce emissions and operating costs**
- Innovate **more efficient engines and drivetrains** aligned with market demand
- Streamline R&D efforts, improve sustainability, and strengthen market competitiveness

Recommendations

Actionable Insights:

- Consumers should prioritize vehicles with alternative fuel types (E, X, Z) and lower fuel consumption (L/100 km) for cost savings and reduced emissions.
- Policymakers should expand incentives for electric and hybrid vehicle adoption to encourage a nationwide reduction in CO_2 output.
- Automakers should invest in improving engine efficiency, promoting compact engine sizes, and advancing transmission technology.

Future Enhancements:

- Incorporate real-world driving behavior, weather, and road conditions into the prediction model.
- Extend the model to evaluate lifecycle emissions for a more holistic environmental impact.
- Integrate solid-state battery and hydrogen fuel cell metrics for future-ready analysis.

Limitations:

- The model does not yet account for external real-time driving variables (e.g., traffic, terrain).
- Current emissions predictions rely on standardized data, not telematics or on-road sensor data.

Implementation

Frontend - User Interface and Interaction

Purpose: Provides the user interface for vehicle data input and prediction display.

Components:

- **HTML**: Form for input (Model Year, Make, Transmission, Fuel Type) and results display.
- **CSS**: Responsive design with circular progress bars and animations.
- **JavaScript**: Handles form submission, dynamic updates of predictions, and error handling.
- **Charts**: Shows fuel consumption and emission insights using Chart.js.
- Service Used – **Render.**

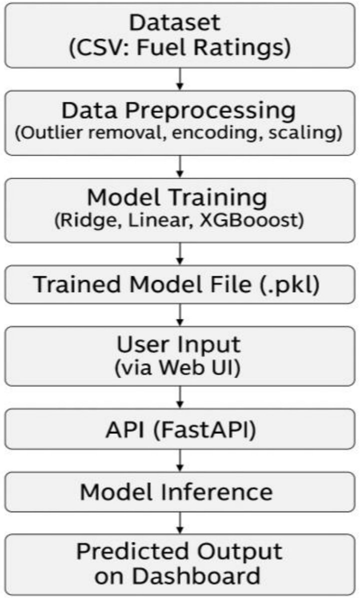
https://forecasting-fuel-efficiency.onrender.com/

Backend - API & Model Predictions

Purpose: Handles data processing and predicts vehicle metrics (fuel consumption, CO_2 emissions).

Components:

- **FastAPI**:
  - `/predict_json`: Receives vehicle data and returns predictions.
  - `/emission_insights`: Provides insights on emissions.
  - `/qr-code`: Generates QR code for easy sharing.
- **Machine Learning**: Ridge regression models predict fuel consumption and CO_2 emissions.
- **Data Preprocessing**: One-hot encoding.



Future Improvements

- **Model Enhancements:** Integrate **Deep Learning models** (e.g., Neural Networks) for complex patterns.
- **Frontend & UX:** Introduce **vehicle image previews** based on selected inputs.
- **Data & Insights**
  - ✓ Enable **real-time CO_2 statistics dashboard Using Apache Spark.**
  - ✓ Scaling the Project with Larger and broader Dataset.
- **Backend Enhancements :** Switch to **Dockerized deployment** for easy scaling.

CO2 Predictor

Home About Statistics Learn More

CANADA

ENERGUIDE

Gasoline Vehicle

Fuel Consumption Predictor

Model Year: 2020

Make: Nissan-Renault Alliance

Transmission: Automatic

Fuel Type: Regular Gasoline (X)

Predict Reset

Fuel Metrics

9.0	10.7	7.4	31
L/100 km Combined	L/100 km City	L/100 km Highway	MPG

Annual Fuel Cost

Based on 20,000 km annually

$1,962

Environmental Ratings

CO2 Emissions	CO2 Rating	Smog Rating	Eco Score
201.4	7	8	7.5

Conclusion

Project Summary:

- ✓ We built a **XGBoost Regression-based predictive model** to estimate CO_2 emissions, fuel costs, and efficiency scores for Canadian vehicles.
- ✓ The final model demonstrated high accuracy (**R^2 = 0.99**), helping users make environmentally and financially informed vehicle decisions.
- ✓ **Electric and hybrid vehicles** emerged as clear winners in reducing emissions.

Final Thoughts:

- ✓ This project showcases how **data science bridges the gap** between environmental goals and consumer decision-making.
- ✓ With further enhancements, the dashboard and model have the potential to serve as a **nationwide tool** for eco-conscious vehicle selection and climate-focused policy planning.
- ✓ As global climate concerns intensify, **predictive analytics will play a crucial role** in driving sustainable transportation solutions.

In closing, our project not only highlights the environmental impact of vehicle choices but also empowers consumers and stakeholders to take informed, data-driven action.

By harnessing machine learning and real-world data, we pave the way toward smarter, cleaner, and more sustainable transportation decisions.

The journey to a greener future starts with the right information and this project is a step in that direction.

THE END

References

- Canada, N. R. (2019, June 11). *EnerGuide for vehicles*. https://natural-resources.canada.ca/energy-efficiency/energuide/energuide-vehicles/21010
- De L'auteur Du Contenu, F. N. O. T. C. a. /. N. E. F. (n.d.). *Fuel consumption ratings search tool*. https://fcr-ccc.nrcan-rncan.gc.ca/en
- *Fuel consumption ratings - Open Government Portal*. (n.d.). https://open.canada.ca/data/en/dataset/98f1a129-f628-4ce4-b24d-6f16bf24dd64
- *Motor fuel prices*. (n.d.). Ontario.ca. https://www.ontario.ca/motor-fuel-prices/