|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| A book with a flag and a car on it  Description automatically generated Optmizing Vehicle Choices: A Data-Driven Approach to Fuel Consumption and CO2 Emissions in Canada | | | | |
|  | |  |  | |
| Project Proposal | | | | |
|  | Group-5  Algorithm Avengers | | |

|  |  |  |
| --- | --- | --- |
| A car driving on a road with a city in the background  Description automatically generated | |  | | --- | | TABLE OF CONTENTS |   [Introduction 3](#_Toc189237973)  [1. Executive Summary 5](#_Toc189237974)  [2. Problem / Rational Statement 5](#_Toc189237975)  [3. Business Goal 5](#_Toc189237976)  [4. Company Description 6](#_Toc189237977)  [5. Stakeholder/client 6](#_Toc189237978)  [6. Data Description 7](#_Toc189237979)  [7. Data Analysis Approach 9](#_Toc189237981)  [8. Implementation / Production 11](#_Toc189237985)  [References 12](#_Toc189237988) |

|  |
| --- |
| Introduction |

**Source Of the Dataset**

The dataset used in this project originates from the Government of Canada’s Open Data Portal, specifically from the **EnerGuide** Vehicle Fuel Consumption Ratings database. This dataset provides model-specific fuel consumption ratings and estimated carbon dioxide (CO₂) emissions for new light-duty vehicles available for retail sale in Canada.

The data includes vehicles from 1995 to 2025, with fuel consumption ratings for older models (1995–2014) adjusted to reflect modern 5-cycle testing methods. These adjustments provide approximate values that allow for consistent comparisons across different model years.

For vehicles from 1995 to 2014, the dataset does not include smog or CO₂ ratings, as these were not consistently recorded. Therefore, we are focusing on data from **2015 -2025**, which provides complete and reliable information on fuel consumption, CO₂ emissions, and smog ratings for more accurate analysis and modeling.

The dataset is publicly available on the Government of Canada’s Open Data website: Fuel Consumption Ratings Dataset.

* Dataset Link: <https://open.canada.ca/data/en/dataset/98f1a129-f628-4ce4-b24d-6f16bf24dd64>

**EnerGuide for Vehicles** – A labeling system that provides basic fuel efficiency and CO₂ emission data for new vehicles.

* EnerGuide Link: <https://natural-resources.canada.ca/energy-efficiency/energuide/energuide-vehicles/21010>

**Fuel Consumption Ratings Tool** – An interactive government tool that helps users compare vehicle fuel consumption but lacks direct carbon tax estimates and cost breakdowns.

* Fuel consumption ratings search tool Link:<https://fcr-ccc.nrcan-rncan.gc.ca/en>

**Why We Selected This Dataset**

* We chose this dataset because it provides **critical insights into fuel efficiency, emissions, and carbon tax impact**, essential for both consumers and policymakers. It is part of **Canada’s EnerGuide system**, offering standardized fuel consumption and CO₂ emissions data. The dataset’s **rich details**, including engine size, transmission type, and vehicle class—make it ideal for **predictive modeling and classification**.

**Importance Of the Dataset For Society**

* This dataset is valuable for:
* **Environmental Awareness** – Helps consumers understand vehicle emissions and fuel efficiency.
* **Policy Support** – Assists governments in refining **carbon tax policies** based on emission data.
* **Consumer Empowerment** – Enables data-driven decision-making for **cost-effective and eco-friendly vehicle choices**.

**How Implementing This Project Can Be Beneficial**

* **Predict Fuel Efficiency** – Machine learning models will estimate fuel consumption based on vehicle attributes.
* **Estimate Carbon Tax Costs** – Provides consumers with a **clear financial impact of CO₂ emissions**.
* **Promote Sustainability** – Encourages **adoption of greener vehicles**, reducing carbon footprints.
* **Support Data-Driven Policies** – Offers insights to enhance **sustainable transportation regulations**.

|  |
| --- |
| Executive Summary |

|  |
| --- |
| Problem / Rational Statement |

This project aims to enhance vehicle fuel efficiency analysis by integrating machine learning into an improved EnerGuide-style dashboard. Existing government tools provide basic fuel consumption data but lack insights into annual fuel costs, carbon tax impact, and vehicle efficiency rankings. Our model will analyze historical vehicle data to predict fuel costs, CO₂ emissions, and tax burden, helping consumers make informed, cost-effective, and eco-friendly vehicle choices

The automotive industry faces growing pressure to improve fuel efficiency and reduce environmental impact, yet consumers lack clear insights into the long-term financial and carbon tax implications of their vehicle choices. Existing tools like EnerGuide for Vehicles and the Fuel Consumption Ratings Tool provide only basic fuel efficiency data without predictive cost analysis.

|  |
| --- |
| Business Goal |

Our project will develop a machine learning-powered dashboard to estimate annual fuel costs, carbon tax impact, and fuel efficiency scores. We will evaluate model performance using metrics like Precision, Recall, and F1-score, aiming for an accuracy rate above 85% to ensure reliable consumer insights and promote sustainable vehicle choices.

The aim of the business is to improve the EnerGuide dashboard by adding a model that computes the carbon tax for each vehicle and suggests the optimal vehicle based on fuel usage and emissions. This solution will provide clients with data-driven insights, allowing them to make decisions that are informed by environmental and financial considerations. The project seeks to promote sustainability initiatives and enhance consumer decisions by increasing transparency and usability.

|  |
| --- |
| Company Description |

|  |
| --- |
| Stakeholder/client |

Our client is a government agency in Canada that regulates vehicle fuel consumption and emissions. Managing a dataset of cars from 2015-2025, the agency provides data-driven insights to shape policies, guide eco-friendly vehicle choices, and support cost-effective fuel use. It also helps customers make informed purchases and aids manufacturers in improving fuel efficiency and reducing emissions

**Government Agencies & Regulatory Bodies (Primary Client)**

* Canadian organizations are tasked with environmental policy, carbon tax regulation, and vehicle fuel efficiency norms.
* Guarantee adherence to sustainability objectives and offer updated regulatory frameworks.

**Manufacturers (Automotive Industry):**

* Identify inefficiencies in current models and improve fuel efficiency and emissions.
* Benchmark their vehicles against competitors to meet regulatory standards and market demands.

**Business & Data Science Team**

* Participated in the validation of models and their integration into current systems.
* Take measures to guarantee that calculations of the carbon tax and suggestions for vehicles based on fuel efficiency and emissions data are correct.

**Customers**

* People and companies are looking for vehicle choices that save fuel and are eco-friendly.
* End users of the EnerGuide dashboard depend on its insights to make informed buying choices

**Environmental Organizations**

* Access data on vehicle emissions and fuel efficiency to support advocacy efforts.
* Monitor progress toward environmental goals and promote eco-friendly policies.

|  |
| --- |
| Data Description |

The dataset includes environmental impact information and vehicle fuel consumption ratings for a range of models from 2015 to 2025.

* **Dataset Summary:**
* Total Rows: 8,297
* Total Columns: 15
* **Data Types:**
* **Numerical:** 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
* **Categorical:** Make, Model, Vehicle class, Transmission, Fuel type
* **Null Values:** No missing values in any column.
* **Column Names:**
* **Model year:** The manufacturing year.
* **Make:** The car's brand, such as Acura, Ford, or Toyota.
* **Model:** The name of the particular model.
* **Vehicle class:** A vehicle is categorized according to its weight and size (e.g., compact, sport utility vehicle, two-seater).
* **Engine displacement** in liters is the engine size (L).
* **Cylinders:** The engine's cylinder count.
* **Fuel type:** The kind of fuel that is utilized (for example, Z stands for premium gasoline, X for regular gasoline, D for diesel, and E for E85).
* **City (L/100 km):** The amount of fuel used for driving in cities, expressed in liters per 100 kilometers.
* **Highway (L/100 km)**: The amount of fuel used for highway driving, expressed in liters per 100 kilometers.
* **Combined (L/100 km):** The weighted average of fuel consumption on the highway and in cities.
* **Combined (mpg)**: Miles per gallon, or fuel efficiency (the higher the better).
* The quantity of carbon dioxide released per kilometer is measured as CO2 emissions (g/km).
* **CO2 rating**: Based on emissions, this number ranges from 1 (worst) to 10 (best).
* **Smog rating:** Based on the pollutants that cause smog, a score ranging from 1 (worst) to 10 (best).

### **Transmission Types & Fuel Efficiency**

* **Automatic (A)** – Shifts gears automatically; **less efficient** than manual but improving.
* **Manual (M)** – Requires shifting; **more fuel-efficient** but less common.
* **Automated Manual (AM)** – Combines manual efficiency with automatic convenience.
* **Automatic with Select Shift (AS)** – Auto transmission with manual control option.
* **Continuously Variable (AV/CVT)** – Uses belts instead of gears; **better fuel economy**.
* **Dual-Clutch (DCT)** – Faster shifts and **higher efficiency** than standard automatics.
* **Electric Drive (E/B)** – Found in EVs; **most efficient** with no gears.

Vehicle Types:

* **Two-Seater (T)** – **2 passengers**, mainly sports cars or compact electric vehicles.
* **Minicompact (I)** – **4 passengers**, smaller than subcompact cars.
* **Subcompact (S)** – **4 to 5 passengers**, common in city cars.
* **Compact (C)** – **4 to 5 passengers**, offering a balance between fuel efficiency and space.
* **Mid-Size (M)** – **5 passenger seats**, suitable for small families.
* **Full-Size (L)** – **5 to 6 passengers**, offering a spacious interior.
* **Station Wagon - Small (WS)** – Typically holds **5 passengers**, with extra cargo space.
* **Station Wagon - Mid-Size (WM)** – Seats **5 to 7 passengers**, making it ideal for families.
* **Pickup Truck - Small (PS)** – **2 to 5 passengers**, often used for light-duty work.
* **Pickup Truck - Standard (PL)** – **2 to 6 passengers**, with larger towing capacity.
* **SUV - Small (US)** – **5 passengers**, balancing fuel efficiency and space.
* **SUV - Standard (UL)** – **5 to 7 passengers**, with a larger cargo area.
* **Minivan (V)** – **7 to 8 passengers**, commonly used for family transport.
* **Van - Cargo (VC)** – **2 passengers**, primarily used for commercial purposes.
* **Van - Passenger (VP)** – **7 to 15 passengers**, often used for group transport.
* **Special Purpose Vehicle (SP)** – Varies in seating capacity, typically **2 to 7 passengers**, depending on use.

|  |
| --- |
| Data Analysis Approach |

We are going to use the following steps to analyze the dataset and derive actionable insights for stakeholders -customers, government, and manufacturers. The focus will be on feature engineering, exploratory data analysis (EDA), and integrating these insights into the existing implementation plan.

**Data Cleaning and Preprocessing**

We are going to clean and preprocess the dataset to ensure accuracy and consistency.

* **Handling Missing or Inconsistent Data:**
  + We will use Python's Pandas library to identify and handle missing or inconsistent data, such as missing fuel consumption values or inconsistent vehicle class labels.
  + Missing values will be imputed using mean, median, or predictive models where necessary.
* **Standardizing Units:**
  + Fuel efficiency metrics will be converted for consistency, ensuring uniform representation (e.g., L/100 km to mpg).
  + CO₂ emissions data will be standardized across all records.
* **Creating Derived Variables Using Feature Engineering:**
  + **Fuel Efficiency Score:** We will derive the Fuel Efficiency Score using the combined fuel consumption values from the dataset. This will help rank vehicles from most to least fuel-efficient.
  + **Annual Carbon Tax Calculation:** Using CO₂ emissions (g/km) and a predefined annual driving distance, we will estimate yearly carbon emissions. This will be multiplied by the applicable carbon tax rate to determine the financial impact per vehicle.
  + **Fuel Cost per Passenger-Kilometer:** By using fuel consumption values, average fuel prices, and assumed passenger count per vehicle, we will estimate the cost per kilometer per passenger for a more accurate cost analysis.

### **Exploratory Data Analysis (EDA)**

We will conduct an in-depth exploratory data analysis to extract key insights.

* **Analyzing Trends Over Time:** Using Matplotlib and Seaborn, we will visualize trends in fuel consumption, emissions, annual carbon tax, and fuel efficiency scores from 2015 to 2025.
* **Comparing Across Vehicle Classes, Engine Sizes, and Fuel Types:** We will generate bar charts and box plots to compare fuel efficiency scores, emissions, and annual carbon tax across different vehicle categories, including compact cars, SUVs, and trucks.
* **Identifying Most Fuel-Efficient and Eco-Friendly Vehicles:** We will rank vehicles based on Fuel Efficiency Score, CO₂ emissions, and annual carbon tax to identify top-performing and eco-friendly options.

### **Statistical Analysis**

* **Correlation Analysis:** We will use Pandas and Seaborn to compute and visualize correlations between key variables such as engine size, fuel efficiency score, number of cylinders, emissions, and annual carbon tax.
* **Regression Analysis:** Using Scikit-Learn, we will build regression models to predict fuel consumption, emissions, annual carbon tax, and fuel efficiency scores based on vehicle attributes like engine size, number of cylinders, and transmission type.

### **Integration with Implementation Plan**

* **Backend Development:** We will use FastAPI to serve processed data, ML model predictions, and recommendations.
* API Endpoints will be developed for:
  + Fetching vehicle specifications, efficiency data, environmental impact, carbon tax, and fuel efficiency scores.
  + Predicting fuel cost, emissions, annual carbon tax, fuel cost per passenger-kilometer, and fuel efficiency scores.
  + Comparing vehicles and generating recommendations based on user input.
* **Frontend Integration:**
  + We will integrate our REST API with a frontend dashboard using React or Vue.js.
  + The dashboard will feature dynamic, real-time visualizations and insights tailored to user preferences.

|  |
| --- |
| Implementation / Production |

### **Implementation Plan for Vehicle Fuel Efficiency & Emissions Analytics**

* **Step 1: Data Processing & Machine Learning (Backend)**
  + Python (for data processing & ML)
  + Pandas & NumPy (data processing & feature engineering)
  + Scikit-Learn & PyTorch (ML models for prediction)
  + PostgreSQL/MySQL (database for storing vehicle data and insights)
* **Step 2: REST API for Data & ML Model Access**
  + Vehicle specs, efficiency data, and environmental impact.
  + ML model predictions for fuel cost, CO2 rating, and vehicle comparisons.
  + Recommendations for customers, government, and manufacturers.
* **Step 3: Frontend Dashboard for Different Stakeholders**
  + REST API integration (to fetch data dynamically)

### **Features for Each Stakeholder:**

* **Customers (Improvement in the existing Energuide)**
  + Search & Compare Vehicles (fuel efficiency, emissions, fuel cost, carbon tax per year)
  + Interactive Graphs (fuel economy vs. emissions, cost over time)
* **Government (Regulatory Insights)**
  + Vehicles Data (fuel efficiency, emissions, fuel cost, carbon tax per year).
  + Predict whether a vehicle complies with future emissions standards.
  + Predictions for Future Emission Compliance
* **Manufacturers (Business Insights)**
  + Identify Inefficient Models & Suggest Fixes.

|  |  |  |  |
| --- | --- | --- | --- |
| STAKE HOLDERS | FEATURE | ALGORITHM | USE-CASE |
| Customers | Interactive Graphs (fuel economy vs emissions, cost over time) | XGBoost Regression | Predict fuel cost, emissions, Carbon tax per year and visualize trends. |
| Government | Search & Compare Vehicles | Gradient Boosting (XGBoost/LightGBM) | Compare vehicles based on fuel efficiency, emissions, etc. |
| Predictions for Future Emission Compliance | Random Forest Classification | Predict whether a vehicle complies with future emissions standards. |
| Manufacturers | Identify Inefficient Models & Suggest Fixes | |  | | --- | | Decision Trees | |  | | Identify inefficiencies |

|  |
| --- |
| 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>