

Fuel Consumption Ratings

Context

Canadians have built a society that is the envy of the world, but in the process, we've developed a serious problem. Canada uses more energy per capita than almost any other country in the world. This is well illustrated in the [Energy Consumption Per Capita Around the World visualization by Visual Capitalist](#) (using data from [Our World in Data](#)). One of the reasons for this is our reliance on the automobile. Statistics Canada reports that [Canadians own about 24 million light-duty vehicles](#) including cars, vans and light-duty trucks, and typically drive more than 300 billion kilometers (km) per year. With close to one vehicle for every two Canadians, [we have one of the highest ratios of car ownership in the world](#).

To a degree, our energy consumption in the transportation sector can be explained by our climate, the vast size of the country and the locations of our population. A great deal of fuel is also wasted in Canada. This happens when we make uninformed purchasing decisions, practice inefficient driving behaviors and fail to properly maintain our vehicles. It's not just a waste of energy – it's a huge waste of money, too. By some estimates, Canadian motorists could save hundreds of dollars per year in fuel and maintenance costs by adopting fuel-efficient practices.

Therefore, gaining a comprehensive understanding of model-specific fuel consumption ratings and the estimated carbon dioxide emissions for new light-duty vehicles intended for retail sale in Canada becomes vitally significant.

Although most Canadians would prioritize the purchase of a hybrid or electric model for their next vehicle purchase, the high price tag for these types of cars and trucks puts them out of reach for many. Understanding which models are the most affordable within each class would therefore also be helpful.

Problem Statement

- Exploratory Data Analysis (EDA): Explore the dataset to understand the distribution of fuel consumption ratings across different vehicle attributes. Identify trends and patterns that could affect fuel efficiency.
- Segmentation Analysis: Segment the dataset based on vehicle types, manufacturers, or other relevant factors. Compare fuel consumption trends among different segments and provide recommendations tailored to each segment.

- Visualization: Create compelling visualizations to communicate your findings effectively to a non-technical audience. Visualizations could include fuel efficiency trends over time, comparison between different vehicle types, and more.

Have you noticed any trends or unusual observations about fuel consumption?

Scope of solution space

Whether this is your first hackathon or your hundredth, I'm committed to ensuring that everybody has a great time and can grow. There are no stupid questions! That being said, we're all in different places on our coding journeys, so it will be helpful if the more experienced team members could help mentor the less experienced.

I've started to code our project in Python, which is a very versatile, easy to learn language. It's also great for machine learning and data science. If you don't know Python, or would rather code in a different language, go for it! Just make sure that your code is clean and well-documented so that others can understand what it does.

In order to simplify our task, I've gone through some preliminary data cleaning steps. The clean data is available in our [GitHub repository](#). In order to focus our analyses, I've selected data from 2023 model vehicles only.

In general, we'll be going through the following steps together (the list isn't exhaustive, and we may decide to go in a specific direction that isn't listed here depending on our team's expertise and interests):

1. Exploratory Data Analysis (EDA)
 - a. What are the different data fields and how do they relate to each other?
 - b. How much data is missing?
 - c. Are there specific trends in the data that could be worth exploring further?
 - d. Are there similar fields in both types of datasets (fuel economy and price)?
 - e. Can the two datasets be combined?
2. Data Engineering
 - a. How are we going to handle missing data?
 - b. Can we engineer some data fields in the fuel economy data so that they might be easier to work with?
3. Data Analysis
 - a. Which trends did we identify in the EDA phase that we'd like to explore in more detail?
 - b. Is it worthwhile building models to explain trends in our data or is a descriptive analysis sufficient?
 - c. How should our analyses be visualized?
4. Finalizing Deliverables
 - a. What are the key take-home messages from our analyses?
 - b. How should take-home messages be presented?

Definition / Data Dictionary

Model

- 4WD/4X4 = Four-wheel drive
- AWD = All-wheel drive
- FFV = Flexible-fuel vehicle
- SWB = Short wheelbase
- LWB = Long wheelbase
- EWB = Extended wheelbase

Transmission

- A = automatic
- AM = automated manual
- AS = automatic with select shift
- AV = continuously variable
- M = manual
- # = number of gears

Fuel type

- X = regular gasoline
- Z = premium gasoline
- D = diesel
- E = ethanol (E85)
- N = natural gas
- B = electricity
- For hybrid vehicles, an * means that in testing, this vehicle did not use any gasoline during electric mode operation. However, depending on how you drive the vehicle, you may use gasoline during electric mode operation following a full charge.

Fuel consumption

- City and highway fuel consumption ratings are shown in liters per 100 kilometers (L/100 km) - the combined rating (55% city, 45% hwy) is shown in L/100 km and in miles per imperial gallon (mpg)
- For electric vehicles, **Le** is gasoline liter equivalent. One liter of gasoline contains the energy equivalent to 8.9 kWh of electricity.

CO2 emissions

- the tailpipe emissions of carbon dioxide (in grams per kilometer) for combined city and highway driving

CO2 rating

- the tailpipe emissions of carbon dioxide rated on a scale from 1 (worst) to 10 (best)

Smog rating

- the tailpipe emissions of smog-forming pollutants rated on a scale from 1 (worst) to 10 (best)

Range

- For electric vehicles, the estimated driving distance (in kilometers) on a fully charged battery
- For hybrid vehicles, the estimated driving distance (in kilometers) on a fully charged battery or full tank of fuel

Recharge time

- For electric vehicles, the estimated time (in hours) to fully recharge the battery at 240 volts

Data

<https://www.kaggle.com/datasets/freddyclchen/fuel-consumption-ratings>

<https://www.kaggle.com/datasets/eringill/2023-american-vehicle-prices>

Reading material and more -

<https://www.diva-portal.org/smash/get/diva2:1255660/FULLTEXT01.pdf>