**Business Objective (from the Word document)**

The fundamental goal is to **model CO2 emissions** as a function of several **car engine features**.  
This means understanding how vehicle attributes like engine size, fuel type, and fuel consumption influence **CO2 emissions** and potentially predicting emissions for new vehicles.

**Dataset Overview (from co2\_emissions.csv)**

* **Rows (instances):** 7,385
* **Columns (variables):** 12
* **No missing values** found

**Variables:**

1. **make** → Car brand (e.g., ACURA, BMW)
2. **model** → Specific car model
3. **vehicle\_class** → Car body type (e.g., COMPACT, SUV - SMALL)
4. **engine\_size (L)** → Engine size in liters
5. **cylinders** → Number of cylinders
6. **transmission** →
   * "A" = Automatic
   * "AM" = Automated manual
   * "AS" = Automatic with select shift
   * "AV" = Continuously variable
   * "M" = Manual
7. **fuel\_type** →
   * "X" = Regular gasoline
   * "Z" = Premium gasoline
   * "D" = Diesel
   * "E" = Ethanol (E85)
   * "N" = Natural gas
8. **fuel\_consumption\_city (L/100km)** → Fuel consumption in city driving
9. **fuel\_consumption\_hwy (L/100km)** → Fuel consumption on highway
10. **fuel\_consumption\_comb(l/100km)** → Combined city/highway fuel consumption
11. **fuel\_consumption\_comb(mpg)** → Combined city/highway fuel consumption in miles per gallon
12. **co2\_emissions (g/km)** → **Target variable** (grams of CO2 emitted per km)

**Dataset vs. Objective**

* **Target Variable:** co2\_emissions
* **Predictors (features):** All other columns (engine specs, fuel consumption, fuel type, transmission, etc.).
* The dataset provides all the necessary variables to **build predictive models** (e.g., linear regression, decision trees, or ensemble models).