

Case-Study Assignment: CO2 emissions in Canada

Answer the questions given in different sections

Data Description Statistics

- What is the structure (shape) of the dataset?
- Show the min, max, and mean of CO2 emission, Fuel Consumption ?(Hint: Pandas function that shows for all the columns at once is available.)
- Comment briefly about the distribution of columns – Model, Fuel Consumption and CO2 emissions.

Pre-processing

- Encode the categorical columns: Make, Model, Vehicle Class, Transmission, Fuel Type.

Data Visualization

- Make Histogram, and whisker plots to understand the meaning of the encoding.

Hypothesis Testing

- Perform Z-statistics test on various columns? And do Hypothesis Test
(**H0**: average passenger vehicle emits 251 grams of CO2 per km.
H1: average passenger vehicle emits more than 251.)
- Is there any multi-collinearity between the columns? If yes, give the column name and its corresponding VIF value.
- Perform correlation Analysis.

Modelling

- Build a Linear Regression Model
- Give the coefficients of the model($y=mx+c$)

For Example:

Engine Size: E

CO2: C

Linear Regression Model: $C = 133.78 + 37.12 * E$

- Comparing predictions and accuracy.