Attached pre-linked observation data, containing a random sample of 30,000 observations. You have two full hours for the following tasks:

1. Which are the most common 5 makes (brands) having largest number of obs ? Show as chart and table, both frequency and % share of entire data. Make clear and informative visualisation.

1. Use the long make-model description field.

Make model description has number of doors (like 4ov). Create new numerical column doors having numerical value for number of doors. Show as chart the distribution of number of doors separately for registration years 2021 and 2022.

1. A) Describe visually the relationship of engine size and CO2 as graph by fuel type.

B) Make prediction model that predicts the CO2 based on the displacement and fuel type. Show data having those predictions (real CO2 and prediction). Report the accuracy of the model as R2 and as correlation plot between actual CO2 and prediction. Interprete the results.

You can get points for

1. Correct answer and accuracy (40%)
2. Clear visual and table reports (35%)
3. Clearness of presentation in English (25 %)

Use Python. After 2 hours, send Python code and the output by replying to this email.