**Project Instructions**

The CTO at "Car-ing is sharing", a car sales and rental company, hired you to help prototype a chatbot app that addresses diverse inquiries using LLMs. She proposed you piloting the following tasks:

* Use a pre-trained LLM to classify the sentiment of the five car reviews in the car\_reviews.csv dataset, and evaluate the classification accuracy and F1 score of predictions.
  + Store the model outputs in predicted\_labels, then extract the labels and map them onto a list of {0,1} integer binary labels called predictions.
  + Store the calculated metrics in accuracy\_result and f1\_result.
* The company is recently attracting customers from Spain. Extract and pass the *first two sentences* of the first review in the dataset to an English-to-Spanish translation LLM. Calculate the BLEU score to assess translation quality, using the content in reference\_translations.txt as references.
  + Store the translated text generated by the LLM in translated\_review.
  + Store the BLEU score metric result in bleu\_score.
* The 2nd review in the dataset emphasizes brand aspects. Load an extractive QA LLM such as "deepset/minilm-uncased-squad2" to formulate the question "What did he like about the brand?" and obtain an answer.
  + Use question and context for the two variables containing the LLM inputs: question and context.
  + Store the actual text answer in answer.
* Summarize the last review in the dataset, into approximately 50-55 tokens long. Store it in the variable summarized\_text.