

1. What data did you collect? How many different sources did you use? How many data samples did you collect? (Similar to Homework 2 Q2, but now you should answer based on your collection progress, rather than just your plan). 5 points.

- The data that I collected can be divided into four categories: car information, car features, car prices, and services information, which all are in the terms of the year(2021 and 2022). Car information contains tables of basic information about cars such as car makes and car models; Car features contain the body feature of the model; car prices contain the price of the cars in the years 2021 and 2022; services' information contains tables of current services prices of the model.
- I used two different sources for now.
 - <https://carapi.app/> (the API for car information, car features, and car prices)
 - <https://api.carmd.com/member/docs#intro> (provides services information that can be used in decision-making models such as the available fields, maintenance information, and repair information.)
- The data folder has a total of 8 CSV files with more than around 5000 samples in total.

3. a & b

The original data format is JSON format. After I requested the API, I collect the keywords that I want and convert them into a data frame format and store them in the CVS.

The keywords that I collect: make, model, price, trim, repair, total_price, and maintenance.