

Data Collection

Link to the GitHub repository

https://github.com/fearofhelicopter/dsci510_project_cars

The general nature of data

- The collected data can be divided into 4 categories: car information, car features, car price, and services information, all in the years 2021 and 2022. Car information contains tables of basic information about cars, such as car makes and car models. Car features include the body features of the model. Car prices comprise 2021 and 2022 car prices. Service information incorporates tables of current service prices of the model.

Data Sources

1. <https://carapi.app/> (the API for car information, car features, and car prices)
 2. <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 7 CSV files with more than 5000 samples.

Description

The original data format is JSON.

After requesting the API, the keywords are collected and converted into a data frame format and saved as CSV files.

The collected keywords are make, model, price, trim, repair, total_price, and maintenance.