DSCI 510 Final Report

Description

The project name is *The Impact of Inflation on Car Prices and Purchase Decisions*. Inflation has maintained high for months, leading to higher gas and car prices. Under this macro scenario, how can people make better decisions when purchasing a car? This project will focus on the automotive industry, using datasets to overview the current price situation of different car models to see inflation effects, and analyze the datasets for decision-making (the best car to buy with users' input options).

The availability data sources will be an API for car prices of different models over the years and an API for users to find a car price by choosing the desired features, such as make, model, color, and mileage. The analysis method will be data preprocessing (data fetching + data cleaning) and data processing, including plotting car prices and finding the best-selling cars by data analysis.

Code Running

Link to the GitHub repository: https://github.com/fearofhelicopter/dsci_510_final_project

See the requirements.txt file in dependency.

Open terminal:

```
# need a python environment >= 3.6
git clone https://github.com/fearofhelicopter/dsci510_project_cars.git
cd dsci510_project_cars
pip install -r requirements.txt
cd code
python main.py
```

to play with the decision making system

python data_decision_making_system.py // under code file

Data Collection

The general nature of the collected data is the car price of different models.

There are two APIs available for fetching data: 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 7 CSV files with more than 5000 samples.

From the first API, 6 files about the basic information about the cars were generated, including car makes, car models, and car prices in 2021 and 2022.

- all_make_table _2021.csv (all makes in the market in 2021) [43 * 2]
- all_make_table _2022.csv (all makes in the market in 2022) [43 * 2]

	make_id	make_name
0	1	Acura
1	24	Alfa Romeo
2	44	Aston Martin
3	2	Audi
4	25	Bentley
5	3	BMW

- all_models_table_2021.csv (all models in the market in 2021) [353 * 3]
- all_models_table_2022.csv (all models in the market in 2022) [385 * 3]

	model_id	model_name	make_id
0	97	ILX	1
1	98	NSX	1
2	99	RDX	1
3	100	TLX	1
4	101	GIULIA	24
5	102	STELVIO	24
6	395	DB11	44

- models_trims_table_2021.csv (model prices and invoices in 2021) [2794 * 4]
- models_trims_table_2022.csv (model prices and invoices in 2022) [2916 * 4]

	trim_id	model_id	invoice	price
0	3484	97	26912	27950
1	3485	97	28771	29950
2	3486	97	25315	25950
3	3487	97	30593	31850
4	3488	97	28882	30000
5	3489	98	146232	157500

The second API generated the user's maintenance costs, indicating the issuance price and potential repair costs for the first 30k miles.

- service_cose.csv

year	model_id	maintenance_cost

It would be an empty file at the beginning and used as the log file for the decision-making system (DMS). When the DMS looks for a suitable car, it calls the second API for maintenance cost and generates one new row of data in this file.

In the initial data processing, attempts were made to collect data for each month from 2021 to 2022. Each month's price plots were graphed, and how prices change with inflation was analyzed. With monthly data, the decision-making process is more reliable. However, the APIs only provide yearly data, which means it can only compare data between 2021 and 2022. To make the decision process more convincing, a decision was made to find a new API approximating the model's insurance and potential repair fee. The cost can be added to the model's price in the decision-making system, making the data more authentic for those asking for advice.

Analyses and Visualizations

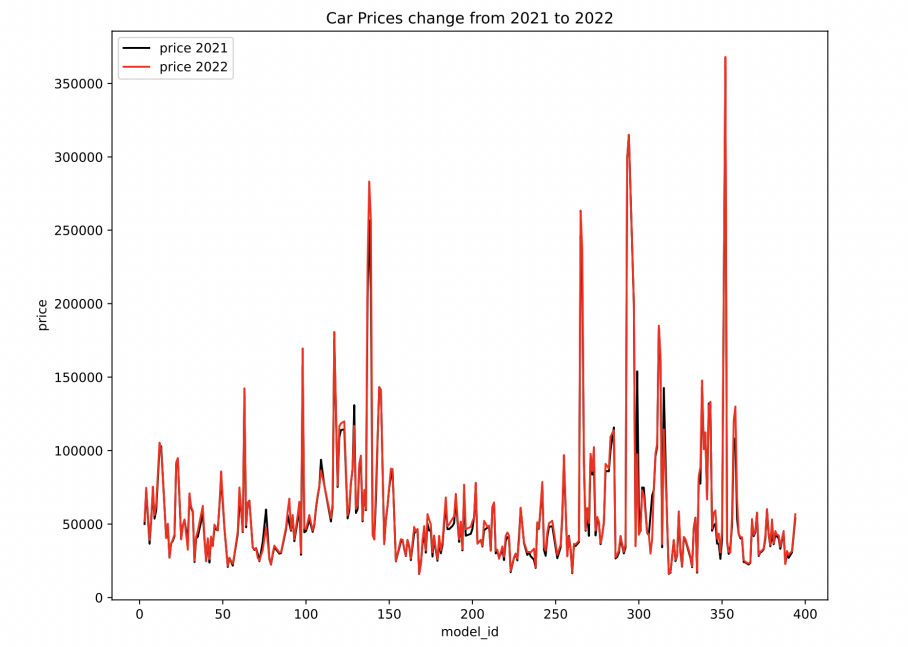
The figure compares the percentage change in prices between 2021 and 2022. Two data frames were used, prices for 2021 and prices for 2022. Two data frames were merged into one by the car's model id. The percentage change in price in two years was calculated.

To see which model impacts the most, datasets were manually split into three parts: models priced below 40k were named cheap cars, cars priced between 40k and 80k were named median cars, and those priced above 80k were named luxury cars. 4 figures were made:

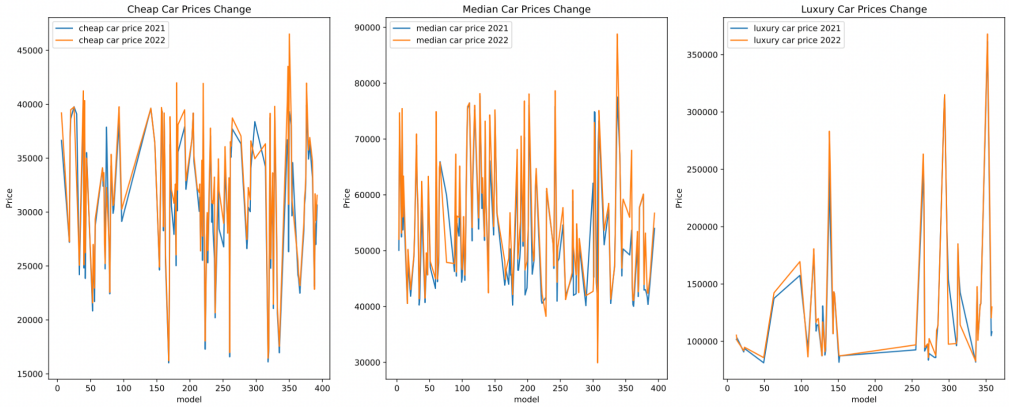
- *Car Price Changes from 2021 to 2022* is the figure that shows the price changes of different models. The red line is the 2022 price, and the black line is the 2021 price. It can be found that the red line is mostly above the black line, indicating price increases.

- *Car Price Changes for Different Classes* is the figure that shows the price changes for different categories of models. Compared with affordable and economical cars, the price of luxury cars changes very little.
- *Car Price Increase Rate from 2021 to 2022* is the figure that shows the growth rate of the prices of different models.
- *Car Price Increase Rate for Different Classes* is the figure that shows the growth rate of different categories of car prices.

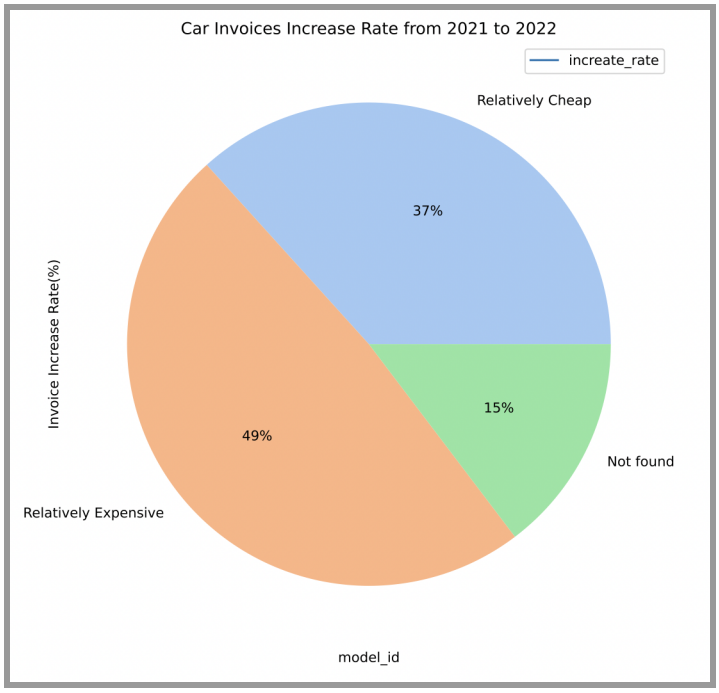
The prices of different car models were collected to analyze price changes with inflation from 2021 to 2022. In the analysis, two data samples (2021 and 2022 car prices) were merged using their unique model ids. The growth rate was calculated with the formula R = (last year/ this year) - 1 to get the increase rate of the different models. It can be easily found that the average increase in car prices from 2021 to 2022 reaches 3.13%, and around 80% of models have positive increases, which strongly supports the initial hypothesis that inflation impacts the car market and leads to rising car prices.



In order to deeply analyze the consumption behavior in the car market during the high inflation period, the data set was divided into three clusters: cheap cars, median cars, and luxury cars. It can be found that the average increase for both cheap and median car prices stays high at around 3.3%, while for luxury cars is only around 2%. The results make sense because when inflation is high, people are more willing to pay less for a cheaper car. Therefore, the demand for luxury cars does not increase much, and the increasing rate stays relatively low.



The one change is that the invoice prices were also analyzed by making a pie chart to do the comparison. Invoice prices are the prices that dealers get for each model and the actual prices of the cars. If the invoice price rises less than the actual price, the dealer makes less money selling the car, so the model is relatively cheap for the public. Otherwise, a smaller increase in invoice prices means dealers are making more money selling a relatively expensive model to the public.



It is an important change to help the decision-making system because if people know whether a dealership is selling a car at a relatively high or low price, they can know which car is fair to buy. Therefore, the final decision-making system will count on 3 aspects to make a good suggestion: market price, invoice price, and insurance and potential fee in the coming year, and give them a weight of 0.6, 0.3, and 0.1.

In the decision-making system, the file will generate a score for each model using a weighted matrix [0.6 0.3 0.1]^T based on the degree of price change, invoice change, and insurance cost information. The final score of each model helps the system advise the user. Users can either choose to enter a model name, and the system will tell users if the model is fair to buy, or enter a price range, and the system will tell users which model is the best to buy within this range.

```
Choice Your Options:
1. Enter a Model (system will give you advise)
2. Enter a Price Range (system will give you advise)
Other numbers to quit
Option: 1
Enter a Model: A6
The better model for similar price to buy is 4 SERIES BMW
```

Choose A6 as an example.

```
Choice Your Options:
1. Enter a Model (system will give you advise)
2. Enter a Price Range (system will give you advise)
Other numbers to quit
Option: 2
Start Price: 20000
End Price: 30000
The best model to buy in this range is ENCORE GX Buick
```

Choose a price range as an example.

to play with the decision making system can simply run:

python data_decision_making_system.py // under code file

Future Work

With more time, more APIs could be used to collect monthly price data. More precise price predictions can be made by applying a simple regression model to help the decision-making system. More predictors can be used to make more precise predictions. For example, the price can be predicted based on the colors, internal furniture, and appearance of the car body for different models. Thus, the decision system can provide more precise advice for decision-making.