UK Used Cars scraper

Autotrader:

For this website I used both requests and selenium libs to do the task. Requests is used to scrape the result page by inputting postal code, with each car listed selenium will load Firefox browser load the car page and scrape data moving to the next one and so on. This one turn to be the slowest so it will take too much time to load database

Data missing in Autotrader (column: replaced by):

Date: “0000-00-00”

Gears: 0

Cargurus:

Cargurus is the simplest one since it provides an open API. A simple request will gather all the data necessary to load in the database. It won’t take much time and with an offset of 36 the process can be done in less than an hour.

Data missing in cargurus:

Power: 0

Cylinder: 0

Gears: 0

Parkers:

Parkers didn’t provide API since it is an old structured website but python requests in able to grab data from it. For parkers search, I used cars make from a prebuilt file and page listing to get as much results as possible.

Data missing in parkers:

Date: “0000-00-00”

Body: “N/A”

Theaa:

Theaa might be the hardest to process on the machine since it loads two Firefox windows, the first one to load search results and the second one to load each car page from that search page, and it also have many missing data and small cars count.

Data missing in Theaa:

Date: “0000-00-00”

Power: 0

Consumption: 0Cylinder: 0

Gears: 0

Seat: 0

Installing driver and libraries:

For the python libraries it is sufficient to install them from the requirements file, but the Firefox driver is a bit more complicated:

1- Download GeckoDriver that suits your operating system from here:
https://github.com/mozilla/geckodriver/releases/tag/v0.31.0

2- Extract it to one of PATH directories or:

a. Right-click on My Computer or This PC.
b. Select Properties.
c. Select advanced system settings.
d. Click on the Environment Variables button.
e. From System Variables select PATH.
f. Click on Edit button.
g. Click New button.
h. Paste the path of GeckoDriver file.

3- Install Firefox 100 or later

4- Make a quick test with the code below:

```
i. from selenium import webdriver
j. driver=webdriver.Firefox()
k. driver.get("http://www.google.com")
```

The code above should open Firefox windows and load google on it

Note: The script and all it files should be on the same folder and ran from it.

Files description:

- **Cars.db**: the SQLite database for all scraped data.
- **Finalcodes**: postal codes used to scrape data in autotrader.
- **Makes**: cars makers list used for parkers.
- **Progress**: csv data for progress saving.
- **Requirements**: python packages used in script
- **Script.py**: main script to download all data from websites
- **Script_log.log**: log files to what happed during scraping process
- **Start.bat**: launcher for the script windows version (the one that should be scheduled)
- **Start.sh**: same as start.bat but for linux
- **Updater.py**: script to update database on a daily basis