

Dockerized Flask Real Estate API with frontend

This is a service that provides users with detailed and downloadable data about real estate adverts. Source of the data is my web scraping program that searches real estate listings and produces raw JSON files that are processed into MySQL tables. The service also has a simple and user friendly frontend with form to download data.

The entire app is dockerized and after cloning its git repository on:
github.com/Bogsatchio/real_estate_dockerized_API
it can be launched with simple command: ***docker compose up -d***

It consists of two container with database and the service itself

```
[+] Building 0.0s (0/0)
[+] Running 3/3
✓ Network prod_api_realestate_default Created
✓ Container mysql Started
✓ Container api_service Started
```

Technologies used:

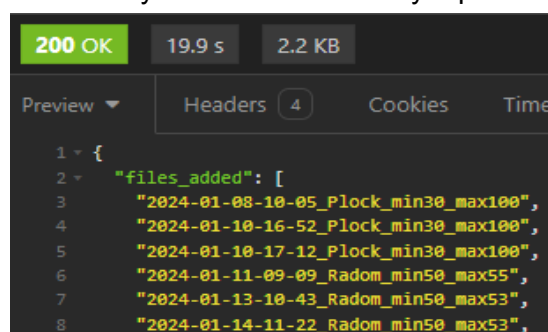
- Selenium & BeautifulSoup for webscraping
- Flask for a setup of a general service
- Pandas for data manipulation and data cleaning
- MySQL for database
- Jinja for HTML templating

Detailed app architecture and functionality:

Web scraping selenium program in a folder *local_realestate_scrap* saves the JSON data into *data* folder. From this folder the running service can load the files into a database.

data_prep_and_insert.py is a main script that handles the database connectivity and cleaning as well as transforming the data.

/refresh endpoint activates the function that first scans the data folder and the metadata table *processed_files* and loads only the files that aren't yet present into the database.



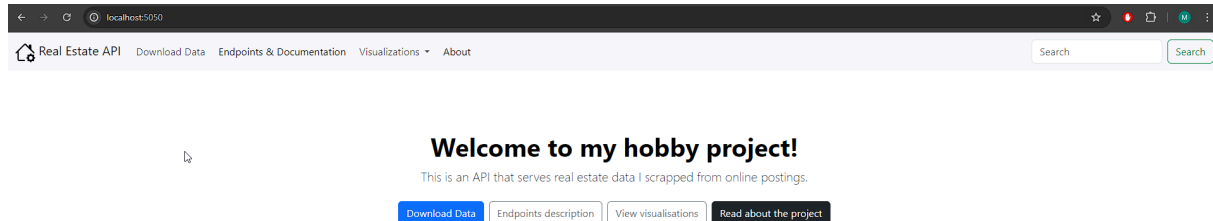
```
200 OK 19.9 s 2.2 KB
Preview Headers 4 Cookies Timeline
1 {
2   "files_added": [
3     "2024-01-08-10-05_Plock_min30_max100",
4     "2024-01-10-16-52_Plock_min30_max100",
5     "2024-01-10-17-12_Plock_min30_max100",
6     "2024-01-11-09-09_Radom_min50_max55",
7     "2024-01-13-10-43_Radom_min50_max53",
8     "2024-01-14-11-22_Radom_min50_max53",
```

`/truncate_all` endpoint cleans the entire database

Data in the app persists because of docker volume meaning they don't need to be loaded with every launch.

The API endpoints can be accessed directly via endpoint in a browser or other program as well as via simple frontend interface on:

<http://localhost:5050>



The endpoint aggregated data can be downloaded in CSV or JSON formats by filling out special form with dynamic field validation based on data present in the database

| City | min size | max size |
|---------------------|----------|----------|
| Plock | 30 | 100 |
| Opole | 50 | 65 |
| Warszawa-Wola | 50 | 65 |
| Targowek | 50 | 70 |
| Torun | 50 | 60 |
| Radom | 50 | 65 |
| Wroclaw-Srodmiescie | 50 | 65 |
| Wroclaw-Gaj | 50 | 70 |

API endpoints can also be viewed in any suitable program

| | | | | | | | | | |
|------------------------------------------------------------------------------|---------------------|-------------------|-----------------|--------------------|----------------------------|--------|-----------|--------|----------|
| GET http://127.0.0.1:5050/general/market_comparison/csv | | | | | Send | 200 OK | 26.3 ms | 1071 B | Just Now |
| Parameters | | | | | Body | Auth | Headers 1 | Docs | |
| URL PREVIEW | | | | | | | | | |
| http://127.0.0.1:5050/general/market_comparison/csv | | | | | | | | | |
| QUERY PARAMETERS | | | | | | | | | |
| Add Delete All Toggle Description | | | | | | | | | |
| Name Value | | | | | | | | | |
| PATH PARAMETERS | | | | | | | | | |
| Path parameters are url path segments that start with a colon ':' e.g. ':id' | | | | | | | | | |
| Preview | | | | | Headers 3 Cookies Timeline | | | | |
| city | market | average_price_sqm | average_size_m2 | average_build_year | | | | | |
| Plock | prywatny | 8512 | 58 | 2000 | | | | | |
| Plock | biuro nieruchomości | 7297 | 54 | 1988 | | | | | |
| Plock | piewotny | 8368 | 56 | 2024 | | | | | |
| Radom | prywatny | 7373 | 56 | 1983 | | | | | |
| Radom | biuro nieruchomości | 6726 | 57 | 1981 | | | | | |
| Radom | piewotny | 8325 | 57 | 2024 | | | | | |
| Torun | prywatny | 9378 | 55 | 1991 | | | | | |
| Torun | piewotny | 9286 | 55 | 2024 | | | | | |
| Torun | biuro nieruchomości | 8774 | 56 | 1987 | | | | | |
| Opole | biuro nieruchomości | 9290 | 58 | 1977 | | | | | |
| Opole | prywatny | 9756 | 57 | 1995 | | | | | |
| Opole | piewotny | 10853 | 56 | 2024 | | | | | |

The list of endpoint and their documentation can be found in swagger UI that is also implemented by the application and available on:
<http://localhost:5050/swagger-ui>

| | | |
|---------------------------------------------------------------------------------------------|--------------------------------|---|
| localhost:5050/swagger-ui | | ☆ |
| Real Estate API v1 OAS 3.0 | | |
| /openapi.json | | |
| general Utilities and queries for the general data concerning aggregates for all the cities | | |
| POST | /refresh | ⌵ |
| DELETE | /truncate_all | ⌵ |
| GET | /general/overview | ⌵ |
| GET | /general/overview/csv | ⌵ |
| GET | /general/price_change | ⌵ |
| GET | /general/price_change/csv | ⌵ |
| GET | /general/market_comparison | ⌵ |
| GET | /general/market_comparison/csv | ⌵ |

Visualisations on frontend site are still work in progress