

1. Source data & web scrapers

2. Data processing and storage

3. Web application



Coordination via
Airflow

A. Housing data

Datapoints to scrape

1
Zoopla

1. Housing price
2. Subway station and the distance
3. Housing features

2
Dexters


BeautifulSoup-scraping

B. Location data



1. Zone
2. Postcode
3. Station name
4. Latitude
5. Longitude

...
API-based scraping

C. Data enrichment

3
London Air

1. Air pollution

Database Schema

Zoopla data

Dexters data

Station name map

Air quality

Air Sites

Location Information


PostgreSQL




Query data via WP plugin




Wep-App based on WP

Zoopla.co.uk




£ 3,400,00

- 3 bed flat for sale
- Room description
- Address
- Subway information



£ 2,400,00

- 2 bed flat for sale
- Room description
- Address
- Subway information




£ 2,600,00

- 3 bed flat for sale
- Room description
- Address
- Subway information


1 Results page

Dexters.co.uk




£ 3,400,00

- 3 bed flat for sale
- Room description
- Address



£ 2,400,00


- 2 bed flat for sale
- Room description
- Address




£ 2,600,00

- 3 bed flat for sale
- Room description
- Address

1 Results page





£ 3,400,00

- ...
- ...
- Subway information

2 Detail page

1. Web scraper

2. Data processing and storage

3. Web app



Coordination via
Airflow

A. Housing data¹

Datapoints to scrape

1 **Zoopla**

1. Housing price
2. Subway station and the distance
3. Bed rooms
4. Bath rooms
5. Living rooms

2 **Dexters**

B. Location data²



Google Maps

1. Zone
2. Postcode
3. Station name
4. Latitude
5. Longitude

C. Enrichment data²

3 **London Air**

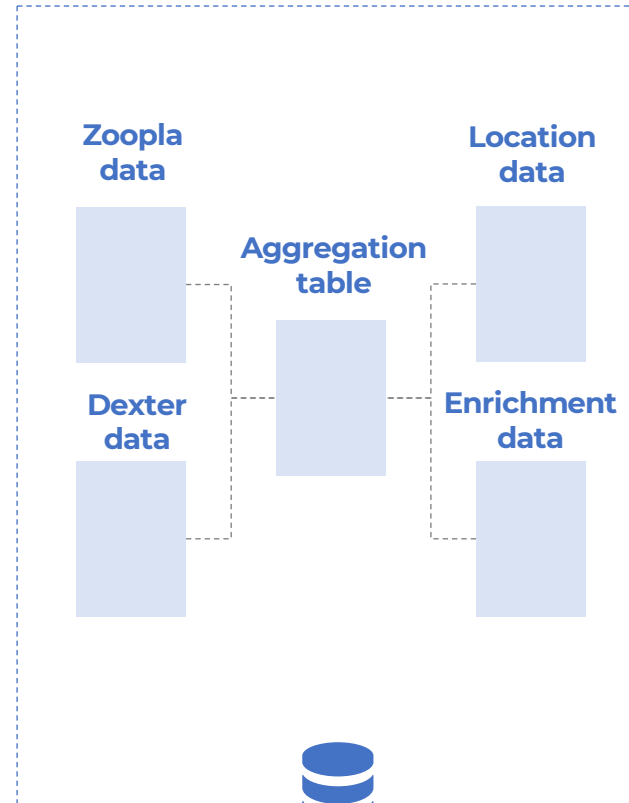
1. Air pollution rate



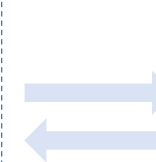
BeautifulSoup-scraping



API-based scraping



PostgreSQL



Query data via WP plugin



Web-App



WordPress