</becode>

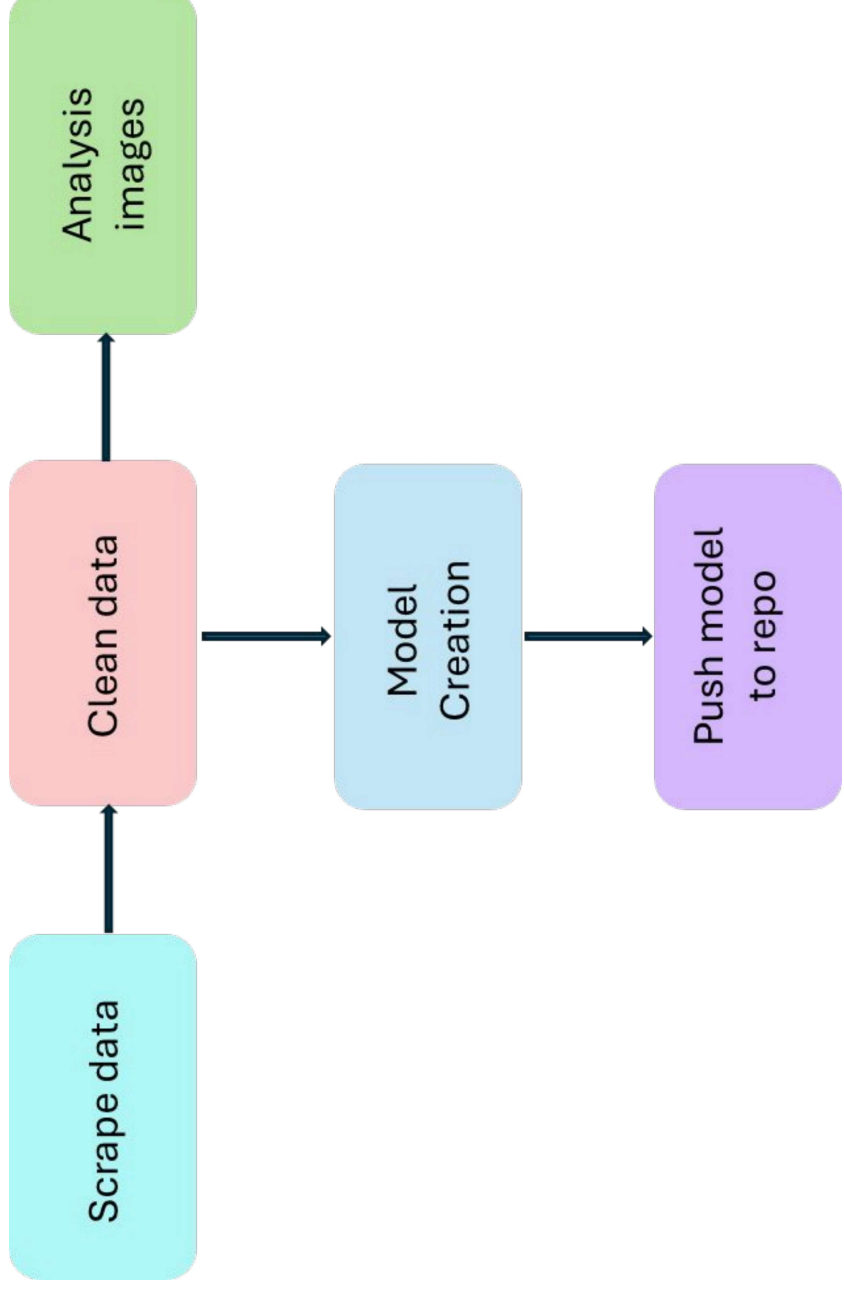# Immo Eliza pipeline automatization
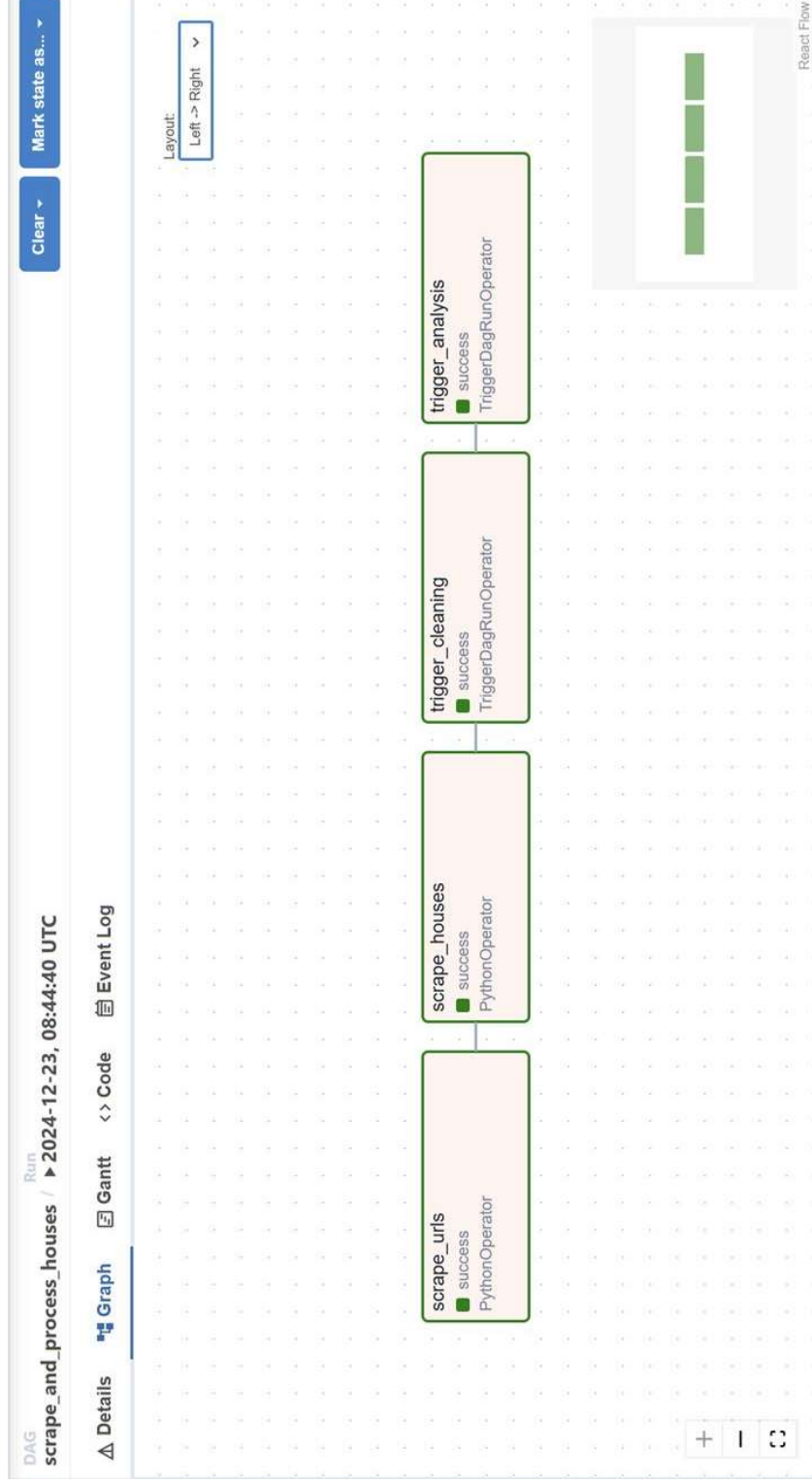
Rasmita Damaraju
Nicolaas De Clercq
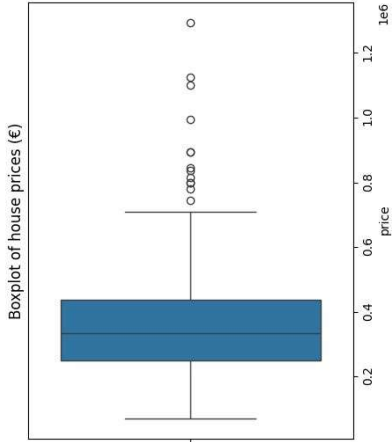Adheeba Thahsin

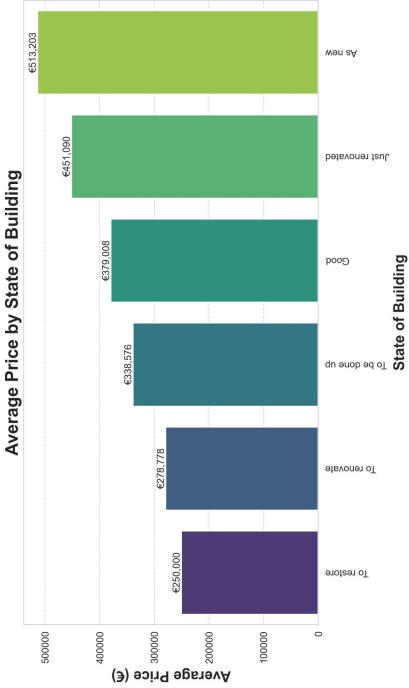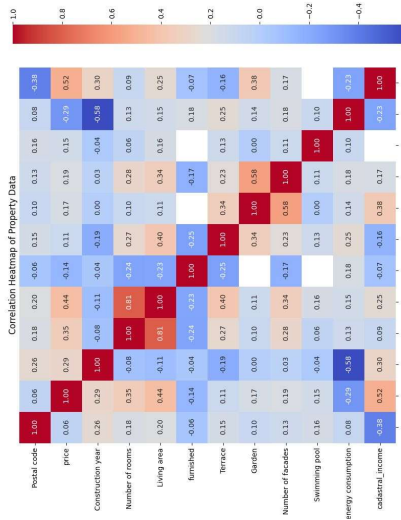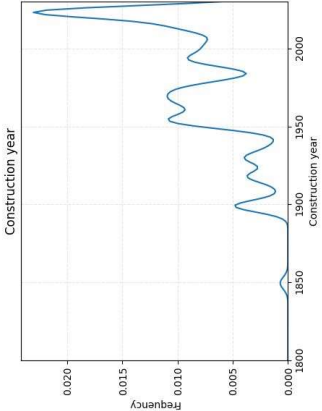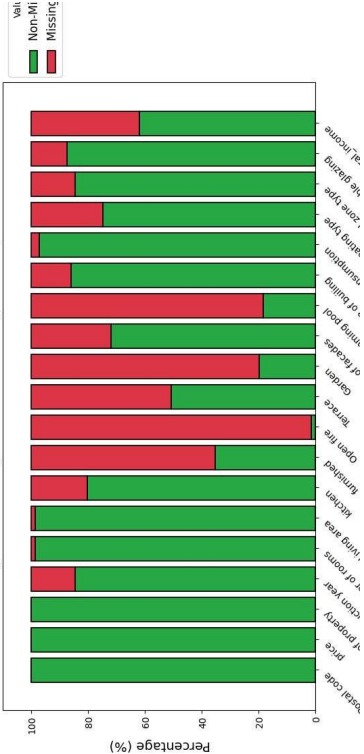# </becode>   Goal of the project

Tasks:
- Scrape housing data for sales every night
- Train a regression model and evaluate its performance.
- Update your API/streamlit app with the new data

</becode>

DAG flow overview

```
┌──────────────┐         ┌──────────────┐          ┌──────────────┐
│              │         │              │─────────▶│   Analysis   │
│  Scrape data │────────▶│  Clean data  │          │    images    │
│              │         │              │          │              │
└──────────────┘         └──────┬───────┘          └──────────────┘
                                │
                                ▼
                         ┌──────────────┐          ┌──────────────┐
                         │    Model     │─────────▶│  Push model  │
                         │   Creation   │          │   to repo    │
                         └──────────────┘          └──────────────┘
```

</becode>

Airflow configuration

</becode>

Sample of reports generated

Boxplot of house prices (€)

Average Price by State of Building

Construction year

Correlation Heatmap of Property Data

Percentage of Missing and Non-Missing Values per Column

</becode>   What were the challenges?

Problems:

- Integrating code
- Blocker with scraping
- Different columns in individual code models

‹/becode›

Enhancements

Problems:

- Train the model on the data scraped from the last 2 weeks instead of the last day.
- Fix model deployment integration.
- Build dashboard for the reports generated in streamlit.
- Push data files into the git repo.

</becode>

Thank you!

Q&A