# PAD – final project

**Ultimate Goal (Not all of this will be implemented in this project)**:

A websites to monitor the prices of apartments in the Warsaw’s districts over long range of time, over a few years.

Build a scrapper which scraps data from different apartment builder’s / seller’s website, along with other popular aggregator websites like Facebook Market-place, OLX, OtoDom.pl, etc.

The supposed users of this platform are retail home-buyers who can filter their desired type of living place,

for eg: { type = “APARTMENT”, price-range = 500k-2000k, district=[WwOCHOTA, WwWOLA], placeOfInterest = [(Longitude-Latitude), (51.12313, 41.32131), (51.3241, 41. 2431)] ,

and can visualize:

1. current listings of houses
2. History of averaged-price for that filtered

**This Project’s goal**:

1. Build scrapper for at least one website.
2. Clean it.
3. Display it on Warsaw map
4. Basic dashboard
5. Train a model (please suggest what can I do here. I am thinking to use a model which will suggest a reasonable price for a house I want to buy).

**DESCRIPTION:**

The project as part of the PAD course consists in collecting data along with cleaning and analysis, creating a dashboard presenting the collected data, and inserting them into the selected classification/regression model (depending on the selected problem).

Project stages:

1. Selecting a topic that is interesting to you
2. Collecting/finding data
3. Cleaning data/missing values
4. Data analysis – their dependencies
5. Creating a dashboard - charts, tables, with descriptions so that you can learn more about the selected topic, data, etc.
6. \*Deciding whether the problem is classification or regression and with the help of the sklearn library, selecting any model and its evaluation using accuracy or F1 score (depending on the data distribution).

**Requirements:**

On 3: Steps 1-5 and data found on Kaggle/elsewhere

On 4: Steps 1-6 and data found on Kaggle/elsewhere

On 5: Stages 1-6 and your own data obtained through web scraping

**Form of devotion:**

You should create either a jupyter notebook that will contain a dashboard or a website created using the streamlit library, upload the source code and submit it as a repository on github.

Additionally, at 4 and 5 there will be a short presentation of the results during the PAD exam. The project is an exam in the subject.

You can use all the techniques learned in the subject. If someone wants to use a library that was not discussed during classes, please consult the instructor first.