# Scenario

A real estate management company wants to determine if they should expand into a new geographic area. Use public data from Aruodas to answer questions about the house market in a specific area. Investigate how neighborhoods or amenities influence house prices.

# STEP 1 – ASK

* 1. **What is the problem you are trying to solve?** 
     1. Predict most profitable districts in Lithuania.
     2. Find the cheapest properties in most profitable districts that will provide the most financial returns.
  2. **What metrics will you use to measure your data to achieve your objective?**
     1. (2.1.1) Forecast profitability measurements using (flat price m2 / rent m2).
     2. (2.1.2) Create house price prediction model and buy those houses whose prediction is higher than actual listed value.
  3. **Visualizations**
     1. Graph that showcases most profitable districts: Measurement: buy price m2 / rent
  4. **Who are the stakeholders?**
     1. Real estate executive team.
  5. **Who is your audience?**
     1. The audience for this presentation is real estate executive team.
  6. **How can your insights help your client make decisions?**
     1. It will assist in locating the cheapest houses that will yield the most return in the long run.

# Step 2 & 3 & 4 & 5 – Prepare, Process, Analyze, Share

Using selenium and BeautifulSoup libraries in python I scraped <https://en.aruodas.lt/kainu-statistika/> webpage for each month’s average flat selling and renting price in all districts.

Graphical user interface, application, Teams

Description automatically generated

Scraping code can be found in Scripts/DistrictPriceWebCrawling/RealEstatePriceStatistics.ipynb

All data points are put inside pandas dataframe an saved as .csv file in “Scripts/DistrictPriceWebCrawling”. Utilizing SQL queries, each dataframe was inserted into the proper BigQuery tables using Scripts/DistrictPriceWebCrawling/SaveRealEstatePrices.ipynb helper functions.

**Table

Description automatically generated**

Figure 1 Flats for rent saved data for future analysis

Using SQL queries, I created new .csv file containing only eligible districts, who has enough data points for further analysis. This table is saved in “Scripts\DistrictPriceAnalysis\EligibleFlatsForAnalysis.csv”.

Text, application

Description automatically generated

Figure 2 SQL querie for finding eligible districts for further analysis

Table

Description automatically generated

Figure 3 Eligible districts

Using linear regression I predicted next month sale and rent prices per m2. Dividing Sale

Using polynomial regression and anomaly detection methods, I find most profitable districts that are sorted and saved in Scripts\DistrictPriceAnalysis\FlatsPriceDataPivot.csv file. The lower RelativePriceToRent number is, the more profitable the district. It is calculated by dividing the predicted average next month's Sale\_m2 by the predicted Rent\_m2 value, as we can see in the next image.

Table

Description automatically generated

Figure 4 FlatsPriceDataPivot.csv data table

In the same excel file there is pivot table that summarizes each district.

Chart, bar chart

Description automatically generated

Figure 5 Pivot table for FlatsPriceDataPivot data

Most profitable districts search was done in /Scripts/DistrictPriceAnalysis/DistrictPriceAnalysis.ipynb file.

Based on most profitable districts we found, we scrape <https://en.aruodas.lt/> all profitable districts and save each listings properties in Scripts\FlatWebCrawling\AllListings.csv files for most profitable listing.

A picture containing application

Description automatically generated

Figure 6 AllListings.csv data table with each flats properties.

Scraping was done in \Scripts\FlatWebCrawling\FlatWebCrawling file.

To predict flat prices, I was training a linear regression and neural network models using AllListings.csv data as inputs. Having made a prediction, I compared it to the actual prices of flats in the most profitable districts, looking for the most undervalued property to purchase.

Text

Description automatically generated

Figure 7 Most profitable flats to buy in city Kaunas, district Aleksotas

Each flats price prediction has to be double checked to avoid models mistakes. Prediction models can be found in \Scripts\FlatAnalysis\FlatAnalysis

# Step 6 – Act

After feeding all listings into the price prediction model and receiving each flat's predicted price, we can identify a flat that is significantly underpriced. These flats can be recommended for purchase to real estate executive team.

A screenshot of a computer screen

Description automatically generated with medium confidence

Figure 8 Most profitable flats to buy and rent in Lithuania