# 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?

General analysis of housing market, best investment locations.

* 1. **What is the problem you are trying to solve?** 
     1. Find most profitable house/flat locations in Lithuania.
     2. Predict most profitable future districts in Lithuania.
     3. Find the cheapest properties in most profitable districts that will provide the most financial returns in n years.
     4. Find anomalies in topmost paid districts. Find houses that should be worth more than their asking price.
  2. **What metrics will you use to measure your data to achieve your objective?**
     1. (2.2.1) Profitability measurement by dividing (price m2 / rent m2) and taking into account future price fluctuations.
     2. (2.2.2) Forecast profitability measurements using (price m2 / rent m2).
     3. (2.2.2) Investigate correlation between general prices of properties compared to bread index, oil index and inflation. Investigate if properties prices trail indexes. Answer questions about how these indexes can assist predict future property prices.
     4. (2.2.4) Most biased houses whose price is extremely lower than districts average adjusted by m2. Abstract formula: districts avg. m2 price \* house m2 – house price
     5. (2.2.4) Find similar houses by features using k-means algorithms in the most profitable districts. Then find properties that has lowest prices in each k cluster. Compare it to the average using m2 price \* house m2 – house price.
     6. (2.2.4) Create house price prediction model and buy those houses whose prediction is higher than actual listed value. Just to be safe compare it to the average districts price using m2 price \* house m2 – house price.
  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.

Scraping aruodas

Calculate average price increase and most biased property (avg timeseries combined with most profitable and least profitable districts)

After investigating top districts, rank best houses/flats to buy. Compare to that district average. Remember how average is computed – there is always lower prices than average. Im not interested in those. Only in those that are listed lower than average even though actual value is supposed to be higher. Filter into avg street price, room quantities and so on. No, instead make house price prediction model and buy those houses whose prediction is higher than actual value. Just to be safe compare it to the average districts price using m2 price \* house m2 – house price.

Is there are differences between the rising prices in 2012-2018 and 2020-2022

Make nice graphs in tableu: highlight richest district

Pitfalls: 1 dollar investment after 10 years returns 10 dollars. 100 Dollars investment after 10 years returns 1000 dollars. Return coefficients are the same but second option is still better, it returns more money. So I should filter by price brackets and then by return index. Ex: homes from [50k-100k]… Basically nonsense statistics… Just use price with no fancy return coefficients.

Group houseSale and houseRent data (not together) by year, city, district and call it PotentialPropertySale, PotentialPropertyRent. Filter data that has more than 5 or more data points and spans longer than 1 year. Take linear regression of each PotentialPropertySale, PotentialPropertyRent, find predicted values at 2022-10-01 and find profitability measurement by using predicted 2022-10-01 (price m2 / rent m2) formula.