1. 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.

2. STEP 1 – ASK

2.1. What is the problem you are trying to solve?

2.1.1.Predict most profitable districts in Lithuania.

2.1.2. Find the cheapest properties in most profitable districts that will provide the most financial returns.

2.2. What metrics will you use to measure your data to achieve your objective?

2.2.1. (2.1.1) Forecast profitability measurements using (flat price m^2 / rent m^2).

2.2.2. (2.1.2) Create house price prediction model and buy those houses whose prediction is higher than actual listed value.

2.3. Visualizations

2.3.1. Graph that showcases most profitable districts: Measurement: buy price m^2 / rent

2.4. Who are the stakeholders?

2.4.1. Real estate executive team.

2.5. Who is your audience?

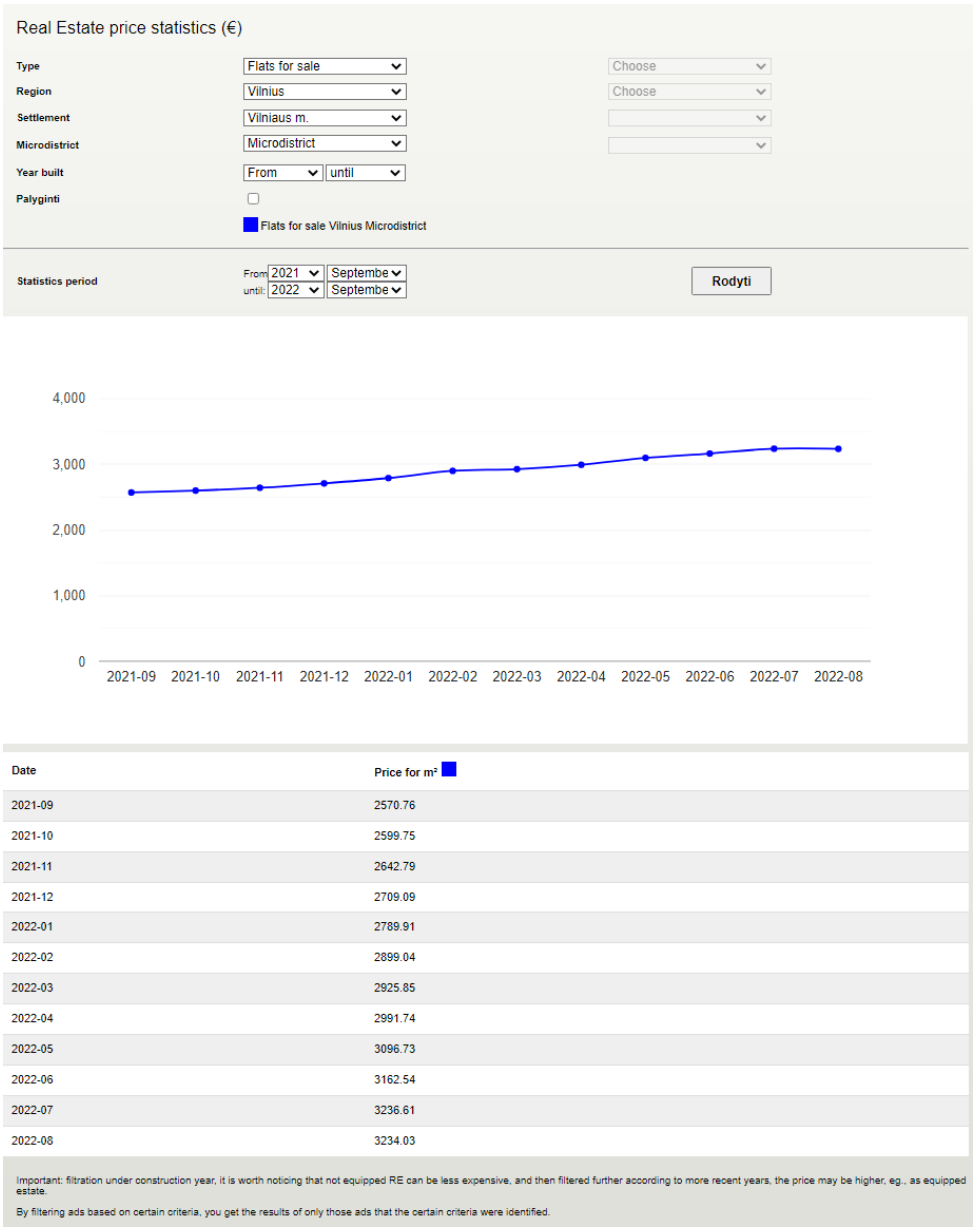
2.5.1. The audience for this presentation is real estate executive team.

2.6. How can your insights help your client make decisions?

2.6.1. It will assist in locating the cheapest houses that will yield the most return in the long run.

3. 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.



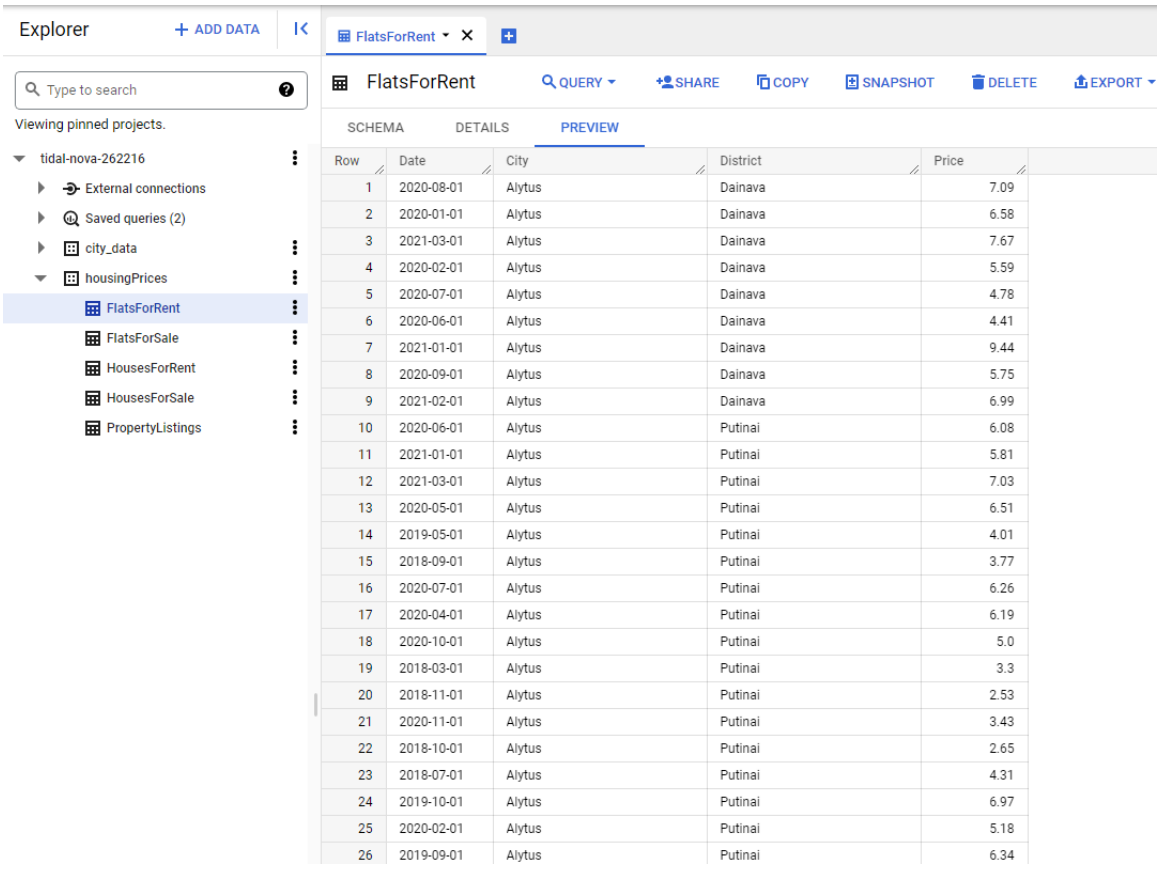
Date	Price for m²
2021-09	2570.76
2021-10	2599.75
2021-11	2642.79
2021-12	2709.09
2022-01	2789.91
2022-02	2899.04
2022-03	2925.85
2022-04	2991.74
2022-05	3096.73
2022-06	3162.54
2022-07	3236.61
2022-08	3234.03

Important: filtration under construction year, it is worth noticing that not equipped RE can be less expensive, and then filtered further according to more recent years, the price may be higher, eg., as equipped estate.

By filtering ads based on certain criteria, you get the results of only those ads that the certain criteria were identified.

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.



Row	Date	City	District	Price
1	2020-08-01	Alytus	Dainava	7.09
2	2020-01-01	Alytus	Dainava	6.58
3	2021-03-01	Alytus	Dainava	7.67
4	2020-02-01	Alytus	Dainava	5.59
5	2020-07-01	Alytus	Dainava	4.78
6	2020-06-01	Alytus	Dainava	4.41
7	2021-01-01	Alytus	Dainava	9.44
8	2020-09-01	Alytus	Dainava	5.75
9	2021-02-01	Alytus	Dainava	6.99
10	2020-06-01	Alytus	Putinai	6.08
11	2021-01-01	Alytus	Putinai	5.81
12	2021-03-01	Alytus	Putinai	7.03
13	2020-05-01	Alytus	Putinai	6.51
14	2019-05-01	Alytus	Putinai	4.01
15	2018-09-01	Alytus	Putinai	3.77
16	2020-07-01	Alytus	Putinai	6.26
17	2020-04-01	Alytus	Putinai	6.19
18	2020-10-01	Alytus	Putinai	5.0
19	2018-03-01	Alytus	Putinai	3.3
20	2018-11-01	Alytus	Putinai	2.53
21	2020-11-01	Alytus	Putinai	3.43
22	2018-10-01	Alytus	Putinai	2.65
23	2018-07-01	Alytus	Putinai	4.31
24	2019-10-01	Alytus	Putinai	6.97
25	2020-02-01	Alytus	Putinai	5.18
26	2019-09-01	Alytus	Putinai	6.34

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”.

Figure 2 SQL querie for finding eligible districts for further analysis

Query results

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS

Row	City	District
1	Alytus	Vidzgiris
2	Kaunas	Romainiai
3	Kaunas	Sargėnai
4	Kaunas	Žemutiniai Kaniūkai
5	Kaunas	Panemunė
6	Kaunas	Centras
7	Kaunas	Šilainiai
8	Kaunas	Aukštieji Šančiai
9	Kaunas	Freda
10	Kaunas	Dainava
11	Kaunas	Žemieji Šančiai
12	Kaunas	Aleksotas
13	Kaunas	Petrašiūnai
14	Kaunas	Vilijampolė
15	Kaunas	Žaliakalnis
16	Kaunas	Senamiestis
17	Kaunas	Kalniečiai
18	Kaunas	Eiguliai
19	Palanga	Palanga

Figure 3 Eligible districts

Using linear regression I predicted next month sale and rent prices per m^2. 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.

	A	B	C	D	E
1	City	District	Sale_m2	Rent_m2	RelativePriceToRent
2	Kaunas	Sargėnai	978	7.65	127.83
3	Vilnius	Naujoji Vilnia	1216.42	8.99	135.27
4	Kaunas	Aukštieji Šančiai	1280.45	8.86	144.57
5	Klaipėda	Žvejybos uostas	1096.83	7.07	155.05
6	Panevėžy	Centras	1030.73	6.56	157.08
7	Panevėžy	Žemaičiai	786.93	4.95	159.03
8	Klaipėda	Poilsis	1071.45	6.58	162.75
9	Šiauliai	Centras	1186.97	7.27	163.31
10	Kaunas	Aleksotas	1702.11	10.16	167.49
11	Vilnius	Balsiai	1532.1	9.06	169.17
12	Vilnius	Naujininkai	1865.58	10.84	172.11
13	Kaunas	Panemunė	1385.18	8.03	172.44
14	Vilnius	Žemieji Paneriai	1544.92	8.93	173.08
15	Kaunas	Žemieji Šančiai	1720.12	9.91	173.56
16	Kaunas	Romainiai	1452.92	8.29	175.29
17	Kaunas	Dainava	1502.13	8.48	177.03
18	Kaunas	Petrašiūnai	1431.24	8.07	177.26
19	Šiauliai	Dainiai	1058.72	5.95	178.04
20	Klaipėda	Mokykla	1396.73	7.82	178.65
21	Kaunas	Kalniečiai	1310.01	7.33	178.72
22	Vilnius	Vilkpėdė	1724.74	9.4	183.48
23	Vilnius	Rasos	1978.65	10.78	183.5

Figure 4 FlatsPriceDataPivot.csv data table

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

Row Labels	Average of Sale_m2	Average of Rent_m2	Average of RelativePriceToRent
Alytus	1107.69	5.21	212.68
Kaunas	1628.15	8.74	185.20
Klaipėda	1454.91	6.87	214.93
Palanga	3188.42	11.04	288.74
Panevėžys	1012.81	5.24	195.73
Šiauliai	1091.86	5.97	184.34
Vilnius	2328.16	10.81	212.43
Grand Total	**1816.31**	**8.74**	**205.68**

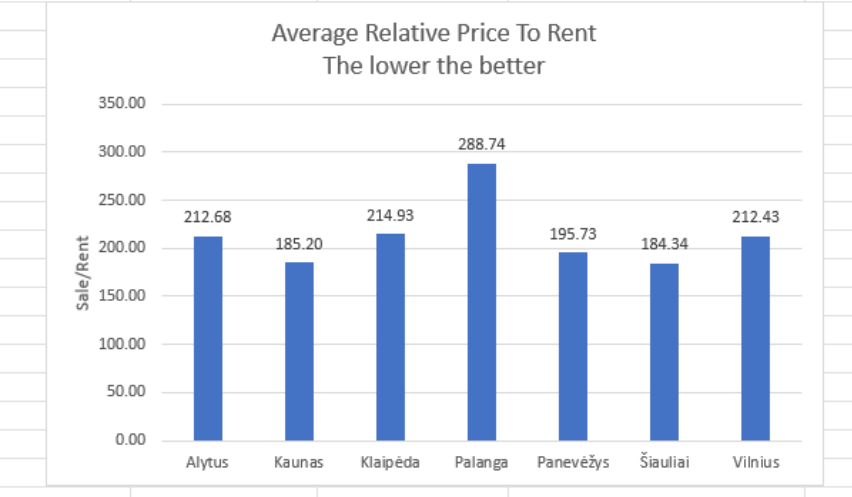


Figure 5 Pivot table for FlatsPriceDataPivot data

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	B	C	D	E	F	G	H	I	J	K	L	M
1	index	Price	PricePerM	Area	floor	Number o	Room cou	Built year	Heating	Equipmen	City	District	href
2	0	75000	1071	70	1	2	3	1/1/1965	Solid fuel	Fully equi	Kaunas	Sargėnai	https://en.aruodas.lt/butai-kaune-sargenuose-vandziogalos-pl-irengtas-b
3	1	199000	1487	133.82	3	4	4	1/1/2008	Gas, solid	Fully equi	Kaunas	Sargėnai	https://en.aruodas.lt/butai-kaune-sargenuose-tvenkinio-g-isskirtinis-4-ka
4	2	58000	2071	28	2	3	2	1/1/2021	Electric	Fully equi	Kaunas	Sargėnai	https://en.aruodas.lt/butai-kaune-sargenuose-vandziogalos-pl-parduodar
5	3	52900	1037	51	1	2	3	1/1/1996	Electric, s	Fully equi	Kaunas	Sargėnai	https://en.aruodas.lt/butai-kaune-sargenuose-vandziogalos-pl-parduodar
6	4	96000	1313	73.11	9	9	4	1/1/1989	Central	Fully equi	Vilnius	Naujoji Vi	https://en.aruodas.lt/butai-vilniuje-naujojoje-vilnioje-parko-g-skubiai-pa
7	5	61500	1847	33.3	6	6	2	1/1/1986	Central	Fully equi	Vilnius	Naujoji Vi	https://en.aruodas.lt/butai-vilniuje-naujojoje-vilnioje-karklenu-g-butas-t
8	6	150000	1835	81.75	5	5	5	1/1/1984	Central	Fully equi	Vilnius	Naujoji Vi	https://en.aruodas.lt/butai-vilniuje-naujojoje-vilnioje-linksmoji-g-patogi
9	7	85000	1049	81	1	2	2	1890-01-0	Aerotherr	Partial de	Vilnius	Naujoji Vi	https://en.aruodas.lt/butai-vilniuje-naujojoje-vilnioje-dumu-g-parduodar
10	8	46000	1603	28.69	4	4	1	1/1/1970	Electric	Fully equi	Vilnius	Naujoji Vi	https://en.aruodas.lt/butai-vilniuje-naujojoje-vilnioje-pramones-g-pardu
11	9	149000	2011	74.08	3	5	4	1/1/1975	Central	Fully equi	Vilnius	Naujoji Vi	https://en.aruodas.lt/butai-vilniuje-naujojoje-vilnioje-geniu-g-buto-priva
12	10	105000	1694	62	2	5	3	1/1/1985	Central	Fully equi	Vilnius	Naujoji Vi	https://en.aruodas.lt/butai-vilniuje-naujojoje-vilnioje-parko-g-triju-kamb
13	11	129900	2451	53	2	5	2	1/1/1983	Central	Fully equi	Vilnius	Naujoji Vi	https://en.aruodas.lt/butai-vilniuje-naujojoje-vilnioje-geroves-g-parduoc
14	12	36500	2028	18	3	3	1	1/1/1988	Gas, elect	Fully equi	Vilnius	Naujoji Vi	https://en.aruodas.lt/butai-vilniuje-naujojoje-vilnioje-dumu-g-parduodar
15	13	30400	760	40	2	2	2	1/1/1972	Other	Other	Vilnius	Naujoji Vi	https://en.aruodas.lt/butai-vilniuje-naujojoje-vilnioje-kloniu-g-parduoda
16	14	129000	1613	80	1	1	4	1/1/2022	Electric	Partial de	Vilnius	Naujoji Vi	https://en.aruodas.lt/butai-vilniuje-naujojoje-vilnioje-ateities-sodu-1-oji
17	15	57300	1910	30	1	2	2	1/1/1965	Electric, o	Fully equi	Vilnius	Naujoji Vi	https://en.aruodas.lt/butai-vilniuje-naujojoje-vilnioje-dumu-g-aukstos-lu
18	16	149999	1899	79	5	5	3	1/1/1996	Central	Fully equi	Vilnius	Naujoji Vi	https://en.aruodas.lt/butai-vilniuje-naujojoje-vilnioje-linksmoji-g-parduc
19	17	120000	1143	105	1	1	4	1/1/1940	Gas, solid	Fully equi	Vilnius	Naujoji Vi	https://en.aruodas.lt/butai-vilniuje-naujojoje-vilnioje-pramones-g-vilniu
20	18	55000	1319	41.7	1	2	2	1/1/1953	Solid fuel	Fully equi	Vilnius	Naujoji Vi	https://en.aruodas.lt/butai-vilniuje-naujojoje-vilnioje-a-kojelaviciaus-g-p
21	19	28000	1276	21.94	2	5	1	1/1/1967	Central	Fully equi	Vilnius	Naujoji Vi	https://en.aruodas.lt/butai-vilniuje-naujojoje-vilnioje-pergales-g-kambar
22	20	96000	2286	42	2	4	2	1/1/1962	Central th	Fully equi	Vilnius	Naujoji Vi	https://en.aruodas.lt/butai-vilniuje-naujojoje-vilnioje-parko-g-parduodar
23	21	96000	1959	49	1	3	2	1/1/1950	Central	Fully equi	Vilnius	Naujoji Vi	https://en.aruodas.lt/butai-vilniuje-naujojoje-vilnioje-parko-g-parduodar
24	22	87300	1375	63.51	5	5	3	1/1/1977	Central	Fully equi	Vilnius	Naujoji Vi	https://en.aruodas.lt/butai-vilniuje-naujojoje-vilnioje-geniu-g-parduodar
25	23	147000	2006	73.29	5	5	4	1/1/1986	Central	Fully equi	Vilnius	Naujoji Vi	https://en.aruodas.lt/butai-vilniuje-naujojoje-vilnioje-linksmoji-g-buto-p
26	24	32500	1083	30	2	4	1	1/1/1990	Electric, o	Partial de	Vilnius	Naujoji Vi	https://en.aruodas.lt/butai-vilniuje-naujojoje-vilnioje-pramones-g-be-tar
27	25	110000	1890	58.19	3	5	3	1/1/1966	Central	Fully equi	Vilnius	Naujoji Vi	https://en.aruodas.lt/butai-vilniuje-naujojoje-vilnioje-geroves-g-patogiai
28	26	97700	1378	70.92	1	2	3	1/1/1962	Electric, s	Fully equi	Vilnius	Naujoji Vi	https://en.aruodas.lt/butai-vilniuje-naujojoje-vilnioje-j-slovackio-g-vilnia
29	27	30000	1833	16.37	3	3	1	1/1/1968	Central	Partial de	Vilnius	Naujoji Vi	https://en.aruodas.lt/butai-vilniuje-naujojoje-vilnioje-parko-g-parduodar
30	28	27000	1240	21.77	4	4	1	1/1/1960	Central	Partial de	Vilnius	Naujoji Vi	https://en.aruodas.lt/butai-vilniuje-naujojoje-vilnioje-pergales-g-parduo

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.

	Price	PricePerM2	Area	floor	Number of floors	Room count	Built year	Heating	Equipment	City	District	href	Losses	PredictedPrice
386	27000	380	71.00	2	2	6	704553	Solid fuel	Partial decoration	Kaunas	Aleksotas	https://en.aruodas.lt/butai-kaune-aleksote-kal...	-6716.699219	33716.699219
379	77000	1038	74.15	1	1	3	737060	Central, solid fuel	Fully equipped	Kaunas	Aleksotas	https://en.aruodas.lt/butai-kaune-aleksote-kro...	-5191.429688	82191.429688
378	29900	622	48.06	0	3	2	711493	Central	Fully equipped	Kaunas	Aleksotas	https://en.aruodas.lt/butai-kaune-aleksote-ant...	486.394531	29413.605469
364	145000	2174	66.70	5	5	3	723181	Electric, gas, other	Fully equipped	Kaunas	Aleksotas	https://en.aruodas.lt/butai-kaune-aleksote-ant...	1521.328125	143478.671875
375	96000	1920	50.00	1	2	2	738156	Gas	Partial decoration	Kaunas	Aleksotas	https://en.aruodas.lt/butai-kaune-aleksote-vai...	2074.632812	93925.367188
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
372	185000	2256	82.00	2	3	3	732312	Gas	Fully equipped	Kaunas	Aleksotas	https://en.aruodas.lt/butai-kaune-aleksote-t-i...	8804.125000	176195.875000
384	107389	1700	63.17	2	3	3	738521	Aerothermal	Partial decoration	Kaunas	Aleksotas	https://en.aruodas.lt/butai-kaune-aleksote-k-d...	8836.890625	98552.109375
366	183000	2473	74.00	1	2	4	737425	Gas	Fully equipped	Kaunas	Aleksotas	https://en.aruodas.lt/butai-kaune-aleksote-zyl...	8842.046875	174157.953125
361	147500	1792	82.33	1	2	4	738156	Other	Partial decoration	Kaunas	Aleksotas	https://en.aruodas.lt/butai-kaune-aleksote-svi...	13219.406250	134280.593750
374	181425	2500	72.57	3	3	3	732312	Electric, gas	Fully equipped	Kaunas	Aleksotas	https://en.aruodas.lt/butai-kaune-aleksote-aka...	37465.562500	143959.437500

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

4. 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.

	Price	PricePerM2	Area	floor	Number of floors	Room count	Heating	Equipment	City	District	href	Losses	PredictedPrice	Year Built
3328	755000	4081	185.00	2	4	5	Central thermostat	Fully equipped	Vilnius	Naujamiestis	https://en.aruodas.lt/butai-vilniuje-naujamies...	-48737.12500	803737.12500	1900-01-01
2257	123000	3981	30.90	2	5	1	Central thermostat, other	Fully equipped	Vilnius	Baltupiai	https://en.aruodas.lt/butai-vilniuje-baltupiuo...	-43694.06250	166694.06250	2021-01-01
4211	235000	4700	50.00	1	2	2	Gas, other	Fully equipped	Palanga	Palanga	https://en.aruodas.lt/butai-palangoje-sermiesc...	-43577.34375	278577.34375	2022-01-01
3474	249900	3602	69.37	8	11	2	Other, electric, central thermostat	Fully equipped	Vilnius	Naujamiestis	https://en.aruodas.lt/butai-vilniuje-naujamies...	-41748.09375	291648.09375	2009-01-01
3845	470000	3790	124.00	1	2	5	Gas	Fully equipped	Vilnius	Senamiestis	https://en.aruodas.lt/butai-vilniuje-senamiest...	-38852.37500	508852.37500	1890-01-01

Figure 8 Most profitable flats to buy and rent in Lithuania