



VENUES' RELATION TO APARTMENT PRICES IN HELSINKI, FINLAND

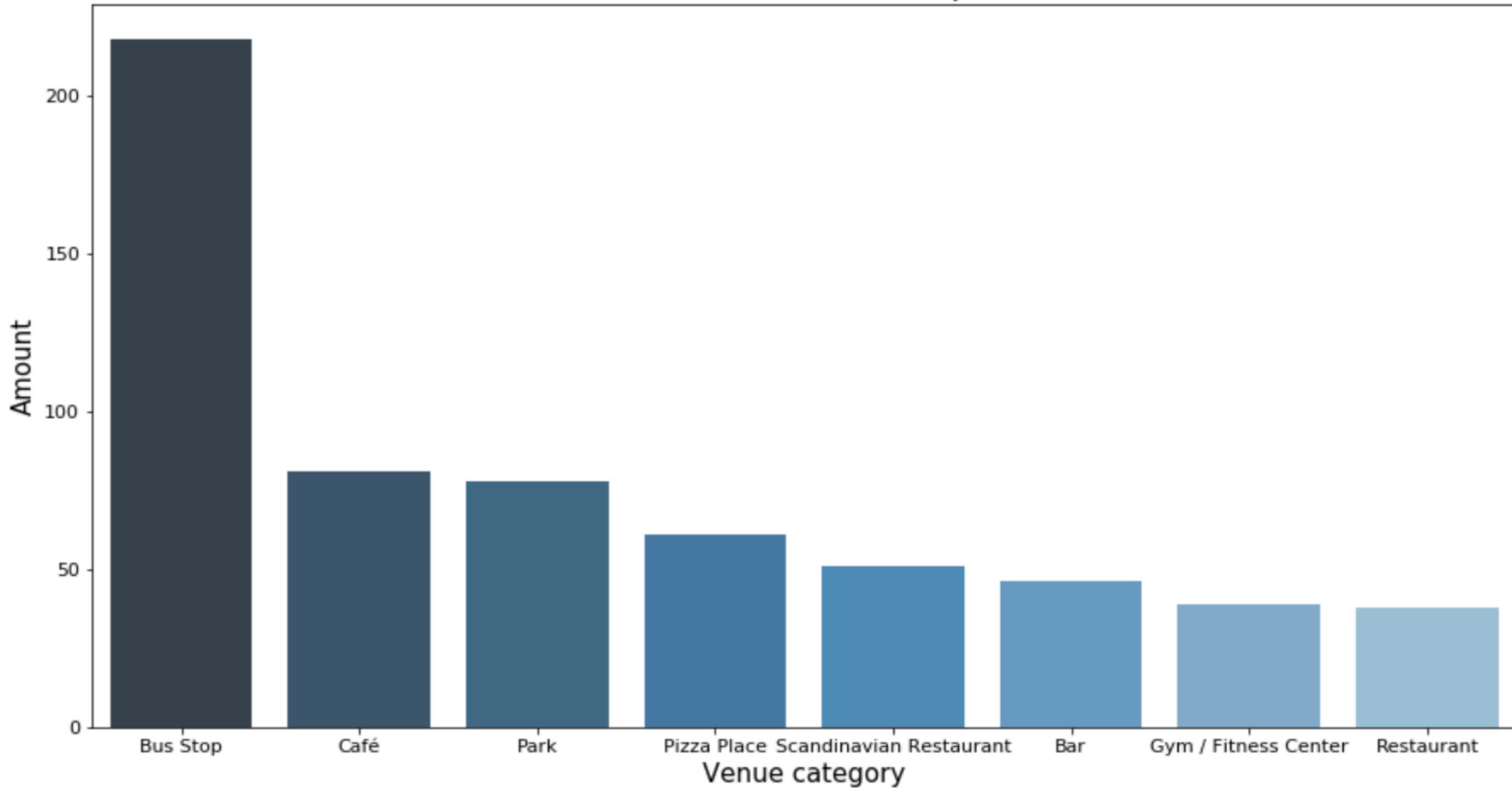
INTRODUCTION AND GOAL

- To find more data for real estate investors for making better decisions
- Outcome to be the list of venues with highest relevance to square meter prices in Helsinki apartments
- Also to have machine learning model that predicts the square meter price

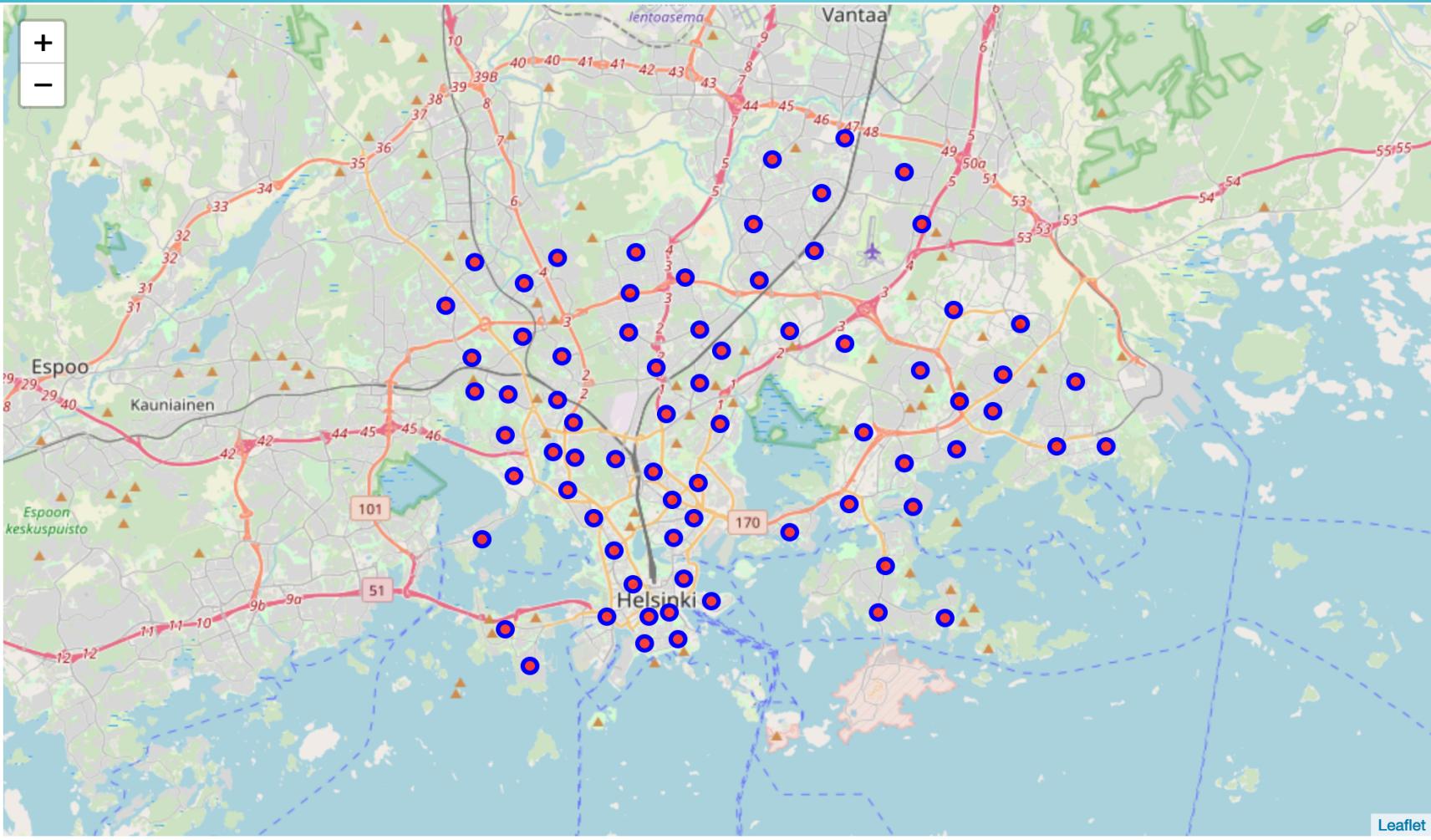
METHODOLOGY

- 3 Data sources
 - Apartments' square meter prices in different post code areas.
 - Location data for postcodes and streets
 - Foursquare for venue data
- Multiple linear regression model with backward elimination to find the most relevant variables

Most common venues in helsinki postcode areas



HELSINKI POSTCODE AREAS



DATAFRAME BEFORE MACHINE LEARNING FITTING

```
: df_sa2['Total_venues'] = df_sa[col_list].sum(axis=1)  
df_sa2.head()
```

	Square_meter_price	Postcode	Art Gallery	Art Museum	Bakery	Bar	Beach	Beer Bar	Bistro	Buffet	...	Soccer Field	Supermarket	Sushi Restaurant	Thai Restaurant	Theater	Tra Static
0	7038	00100	2	3	3	2	0	0	0	0	...	0	1	2	0	0	0
1	7526	00120	3	0	4	4	0	5	1	0	...	0	0	3	0	0	1
2	7938	00130	2	1	0	3	0	2	0	0	...	0	0	2	0	0	0
3	7945	00140	0	1	0	0	0	1	2	0	...	0	0	1	1	1	1
4	7663	00150	0	0	3	2	1	3	2	0	...	0	0	2	0	0	2

5 rows x 64 columns

```
: df_sa2.shape  
(73, 64)
```

OUTCOME

- Here is a complete list of venues that has highest correlation with apartment's square meter price in Helsinki
 - P-value of 0.05 was the limit for this list
- Total venues means total number of venues in one postcode area
- We see a theme with cultural venues, activites, and sea side locations. Also wide range of restaurants present on the list

```
[ 'Art Gallery',
  'Bakery',
  'Beach',
  'Bistro',
  'Buffet',
  'Dog Run',
  'Falafel Restaurant',
  'Flea Market',
  'Gastropub',
  'Gym',
  'Gym / Fitness Center',
  'Harbor / Marina',
  'Himalayan Restaurant',
  'Middle Eastern Restaurant',
  'Music Venue',
  'Pharmacy',
  'Plaza',
  'Sandwich Place',
  'Scenic Lookout',
  'Soccer Field',
  'Supermarket',
  'Sushi Restaurant',
  'Thai Restaurant',
  'Wine Bar',
  'Total_venues' ]
```

OUTCOME

- Prediction model's accuracy is not high enough to use it as a stand alone product for apartment price estimations
 - Average difference of \hat{y} and y_{test} was 27% per predicted price (avg_error %)
- Correlation was found with venus and this model
 - Possible use cases alongside other machine learning models to make them more accurate

	predicted	test_set	predicted / test	error2	average distance	error	avg_error %
0	5649.685649	3325	1.699154	0.699154	0.101688	0.699154	0.27401
1	2810.750501	2053	1.369094	0.369094	0.101688	0.369094	0.27401
2	4231.726563	3271	1.293710	0.293710	0.101688	0.293710	0.27401
3	3062.041386	3791	0.807713	-0.192287	0.101688	0.192287	0.27401
4	3509.731768	3395	1.033794	0.033794	0.101688	0.033794	0.27401
5	3298.859945	2849	1.157901	0.157901	0.101688	0.157901	0.27401
6	2493.491677	2396	1.040689	0.040689	0.101688	0.040689	0.27401
7	2944.147721	2657	1.108072	0.108072	0.101688	0.108072	0.27401
8	6007.271138	5554	1.081612	0.081612	0.101688	0.081612	0.27401
9	10615.077250	7354	1.443443	0.443443	0.101688	0.443443	0.27401
10	4405.920478	3550	1.241104	0.241104	0.101688	0.241104	0.27401
11	5660.556363	4417	1.281539	0.281539	0.101688	0.281539	0.27401
12	1467.426914	3618	0.405591	-0.594409	0.101688	0.594409	0.27401
13	5281.985228	3377	1.564106	0.564106	0.101688	0.564106	0.27401
14	4099.311419	5951	0.688844	-0.311156	0.101688	0.311156	0.27401
15	4222.783670	2740	1.541162	0.541162	0.101688	0.541162	0.27401
16	2766.904172	3380	0.818611	-0.181389	0.101688	0.181389	0.27401
17	3602.042567	3151	1.143143	0.143143	0.101688	0.143143	0.27401
18	3603.315833	4694	0.767643	-0.232357	0.101688	0.232357	0.27401
19	4444.033156	2920	1.521929	0.521929	0.101688	0.521929	0.27401
20	2878.559822	7017	0.410227	-0.589773	0.101688	0.589773	0.27401
21	4007.390334	3129	1.280726	0.280726	0.101688	0.280726	0.27401
22	2510.121364	3186	0.787860	-0.212140	0.101688	0.212140	0.27401
23	8833.942240	7663	1.152805	0.152805	0.101688	0.152805	0.27401
24	4052.218146	3334	1.215422	0.215422	0.101688	0.215422	0.27401
25	4557.379590	3139	1.451857	0.451857	0.101688	0.451857	0.27401
26	6329.675253	7938	0.797389	-0.202611	0.101688	0.202611	0.27401
27	4710.495648	4768	0.987940	-0.012060	0.101688	0.012060	0.27401
28	3275.892079	3230	1.014208	0.014208	0.101688	0.014208	0.27401

CONCLUSION AND DISCUSSION

- Successfull project
- More accurate model could be built with deep comparison between regression models and more accuratevenues data
 - This was unfortunately out of scope for this IBM's project, but interesting idea for future study