



Predicting housing prices

Using venue data

Contents



Introduction



Data



Methodology



Results



Conclusion

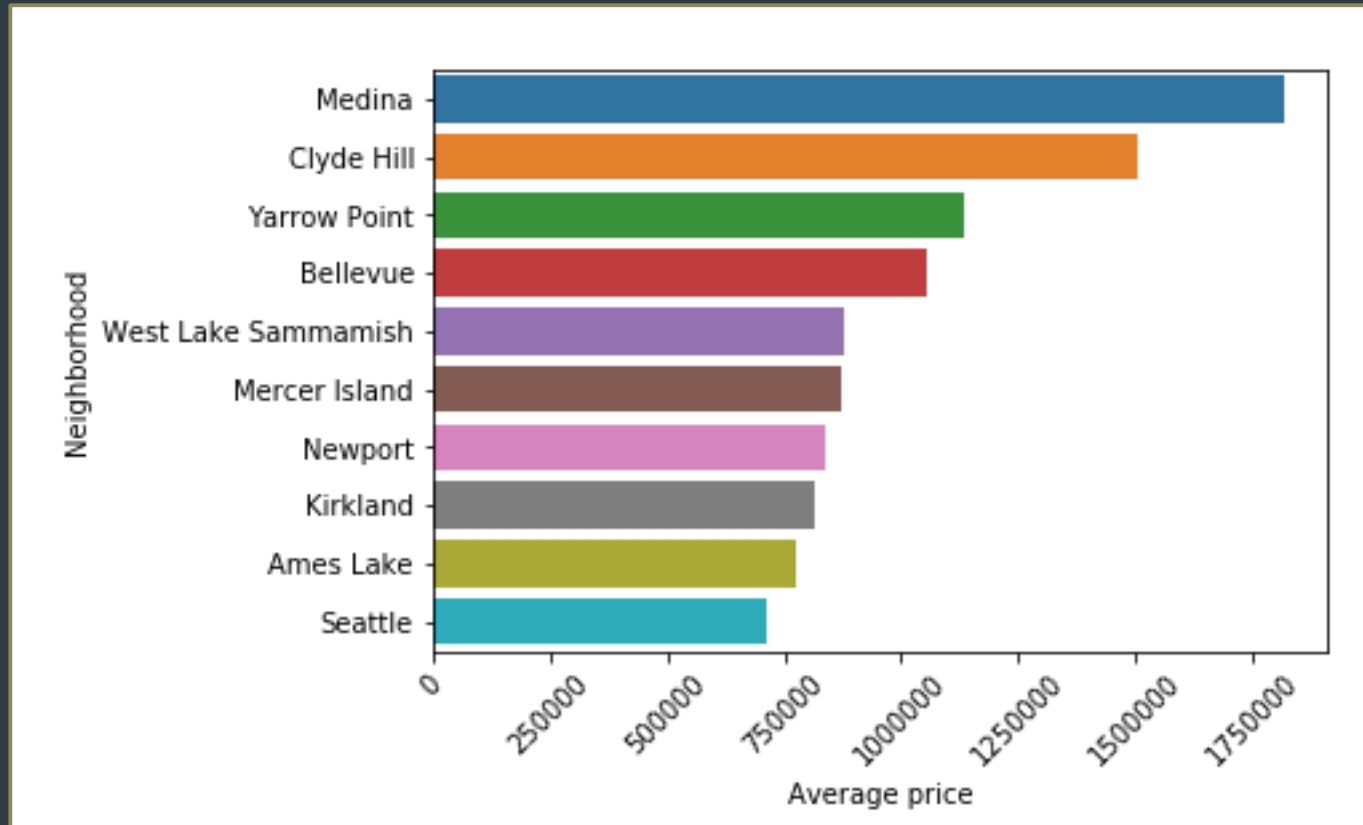
Introduction

- ▶ Housing market is very inefficient
- ▶ House prices are subjective
- ▶ Our goal is to quantify features of houses to predict prices
- ▶ How? ...

Data

- ▶ We propose a novel method to use venue data in order to predict house prices
- ▶ We look at the city King County. Washington. USA

Top 10 most expensive neighborhoods



Top 10 most expensive neighborhoods

Neighborhood	Price (x 10 ⁵)	Floors	Bedrooms	Bathrooms	Grade
Medina	18.17	1.47	3.86	2.90	9.17
Clyde Hill	15.06	1.38	4.07	2.67	8.76
Yarrow Point	11.33	1.56	3.59	2.41	8.57
Bellevue	10.54	1.43	3.66	2.41	8.49
West Lake Sammamish	8.770	1.54	3.79	2.57	8.75
Mercer Island	8.741	1.45	3.63	2.29	8.11
Newport	8.381	1.28	3.72	2.40	8.19
Kirkland	8.145	1.50	3.52	2.35	8.05
Ames Lake	7.771	1.84	3.79	2.81	9.15
Seattle	7.101	1.70	3.22	2.09	7.72

We see little relations across price and features, but maybe venue data can help us out...

Methodology

- ▶ 3 different models
- ▶ Linear regression
- ▶ Ridge regression
- ▶ Random forest

Results

Model	Without venue features	With venue features
Linear regression	4.012	3.672
Ridge regression	4.012	3.672
Random forest	2.259	2.105

Values in table are mean squared errors (MSE). The lower the better.

Discussion

- ▶ Restaurants and bars in the center increase the price of a house
- ▶ Studios decrease the price of a house
- ▶ Might be that more restaurants and bars are more located in the centre of the city
- ▶ Random forest model is best in terms of MSE in predicting housing prices

Advice

- ▶ Use venue data to predict house prices of houses that aren't sold yet.
- ▶ Further research can be done:
 - ▶ Expand the different types of categories
 - ▶ Look at the distance of each house to venues
 - ▶ Look at number of venues around a house