REAL ESTATE RENTAL PRICE PREDICTION

Kaarel Vesilind
Marten Vainult
Tarmo Pungas

Predicting real estate rental prices based on data from kv.ee   Source code

INTRODUCTION

Putting your real estate up for rent can be troublesome. There are a lot of factors to consider: when it was built, how good the infrastructure is, the number of rooms, etc. Even the smallest of things can either bump up or decrease the price. And for the same reasons it is difficult to decide whether a property is a good deal or not. Most people in the renter's scenario go to an advert portal where they can compare their property with others. It provides a good estimate on the price since the price is determined mostly by the market average. As for the interested tenant, there might be hidden gold up for rent, but it's hard to come by and takes a lot of effort to find.

GOAL

Our goal was to make life easier for both the renter and tenant by creating a handy tool which evaluates a property based on its features. The renter would insure a fair price and the interested tenant would find a great deal.

We will measure our results by testing the tool with average and professionally priced properties to see how closely the price is predicted. Successful model would predict with the accuracy of at least ±10%

DATA

The dataset consisted of 24 000 real estate advertisements with their metadata, description and links to cached images. The data was scraped from kv.ee on Nov 3, 2020 by Taivo Pungas. Properties were located all over Estonia. However, to narrow the scope of this project, only to-let properties in Tallinn were chosen. After cleaning up and filtering the data, 1807 ads remained for analysis.
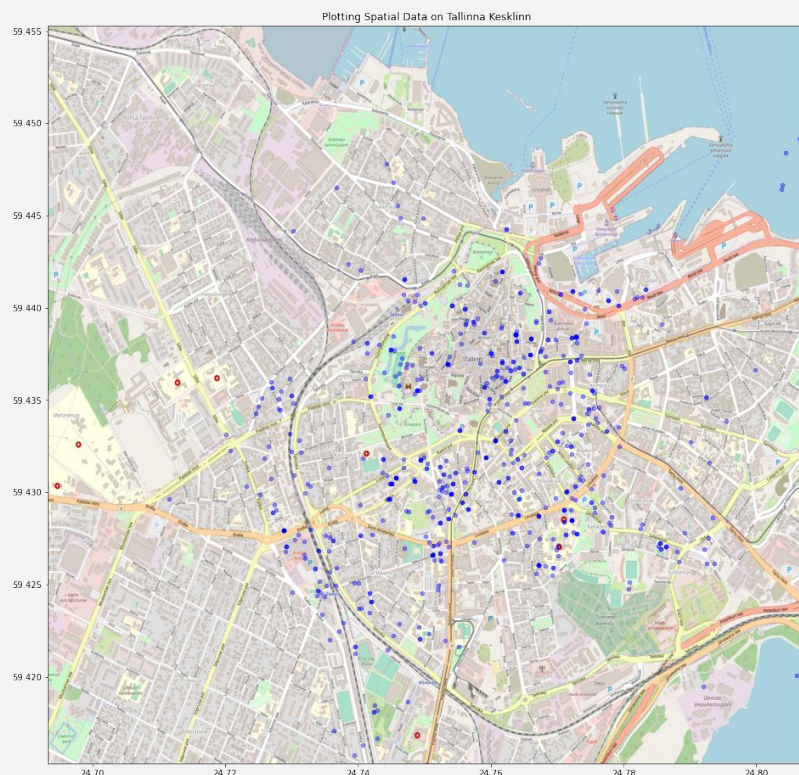

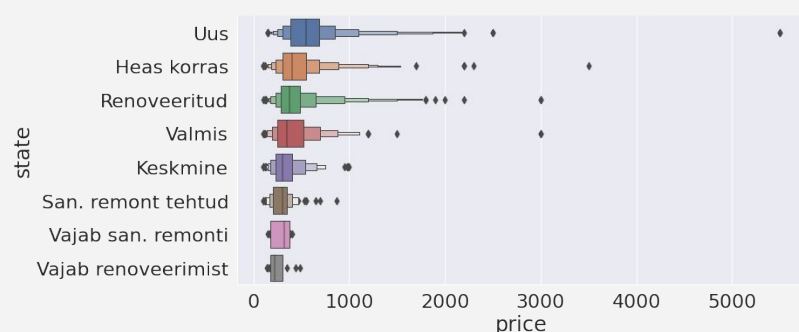Figure 1. Apartments for let in the centre of Tallinn

Figure 2. Apartments price to constructional state ratio.

MACHINE LEARNING

After cleaning and optimizing the data the following features were used for training models:
 number of rooms, area of property, year of construction, state of property, energy class, cluster of property and calculated distance from town center by coordinates
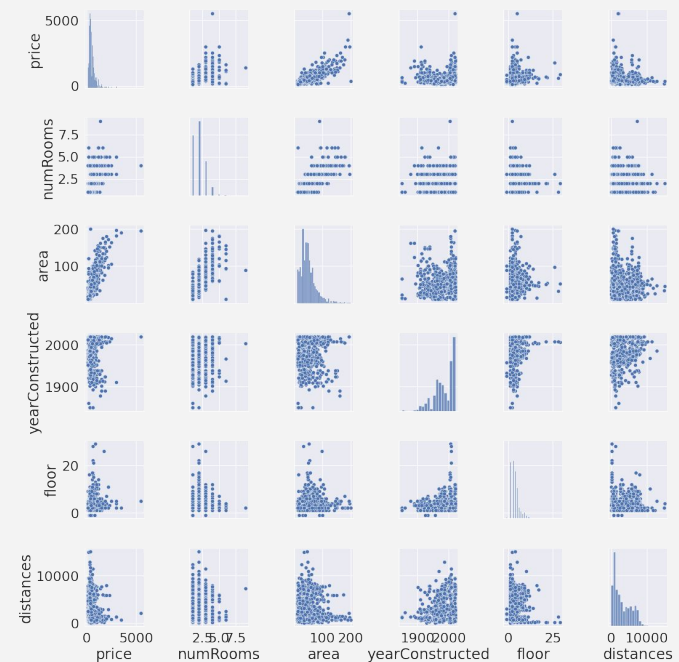

Figure 2. Pairplots of the data

RESULTS

The best results were obtained by Random Forest Regression (max_features="sqrt"). with the obtained accuracy of 17% on the Tallinn subset of the data. This is less than hoped for, but not a bad result, considering the data. The accuracy could likely be improved by a few percentages by extensive work on the data (e.g. detecting and removing more outliers).

We also tried with Deep Neural Network, Ridge and Lasso regression. All these three had similar error percent ~24%