

# Predicting Prices in German Real Estate Market

Hanna ()  
Konrad Wartke (6481209)

June 24, 2020

## 1 Original Data Sources

We started with two datasets from Kaggle, one with rent price and one with house prices for the German Real Estate market. The Rent data was scraped from Immoscout24 on three different dates (22.09.2018, 10.05.2019, 08.10.2019). The house prices were also taken from Immoscout24. While no date is given for the creation of this dataset, it is most likely from novembre 2017. The name of the house price dataset is misleading, it features houses from all over Germany instead of Hesse only.

## 2 Cleanup and Merging

Since both datasets are compiled from the same original source, they share most of the features. For several features, the names are in German in one dataset, but English the other, so we had to translate the names.

The most important feature is the price, which is influenced by most features. For example, all other things equal, having a garden increases the price. To eliminate the influence of several features, we trained a linear regression model. The feature weight is then used to calculate a clean price as if the real estate had no extras. This price can be negative, which means money has to be invested to bring the house into a standard state.

At the end, all rent and buying prices where grouped by the zip code and the average rent and buying price per square metre is calculated. Also included are the percentages of real estate units for which a binary feature is true.

### 3 Prediction

#### 3.1 kNN

#### 3.2 Linear Regression

### 4 Results

### 5 Discussion