

<https://github.com/pilipentseva/predictions>

**Goal of the project:** Predict prices for small and medium apartments in San Diego

**Data set:** for my project I used a data set from Airbnb Inside. Initial dataset has 9673 observations and is collected for the data 25th of September 2021.

**Data cleaning and transformations:** Firstly, I got rid of the columns which didn't have any meaning for my analysis. Secondly, I deleted the columns which repeated information listed in another column or stating the obvious information. After this I looked at data types and transformed those who belonged to a mistaken data type. I also evaluated the missing values. Depending on the logic and structure of the data, I took decisions on how to replace missing values. For some observations I replace them with 0, for others with mean or mode. Last but not least, I encoded the categorical variables.

**Model building:** I built 3 models. One with basic predictors. Second one with basic predictors and variables for reviews. Third with all features included in the data set. I was supposed to run a random forest to select the best model, which I didn't.