Prediction of Housing Prices

Sara Maxwell

Problem

Social benefits of homeownership:

- Greater happiness and prosperity levels
- Higher civic participation rates
- Better education outcomes.

What do prospective buyers look for? Not just price, but also:

- Square footage
- Number of bedrooms
- Number of bathrooms

Main Clients

Homebuyers

- Can see how much house they'd get for their money
- More informed decisions about purchasing a house

Businesses

- Homeownership rates associated with greater stability

Government officials

- Come up with policies that improve housing affordability

Overview

The steps involved in this analysis include:

- Data cleaning and wrangling
- Feature engineering and preprocessing
- Exploratory data analysis
- Machine learning

Steps in the Analysis

- Data Cleaning
  - Isolate useful values and rename them
  - Convert values such as “poor” and “good” to numeric
  - Replace “nan” values with zeros
- Feature Engineering
  - Calculate age: 2011 - year built
  - Combine half and full bathrooms
  - First floor + second floor + basement = overall square footage

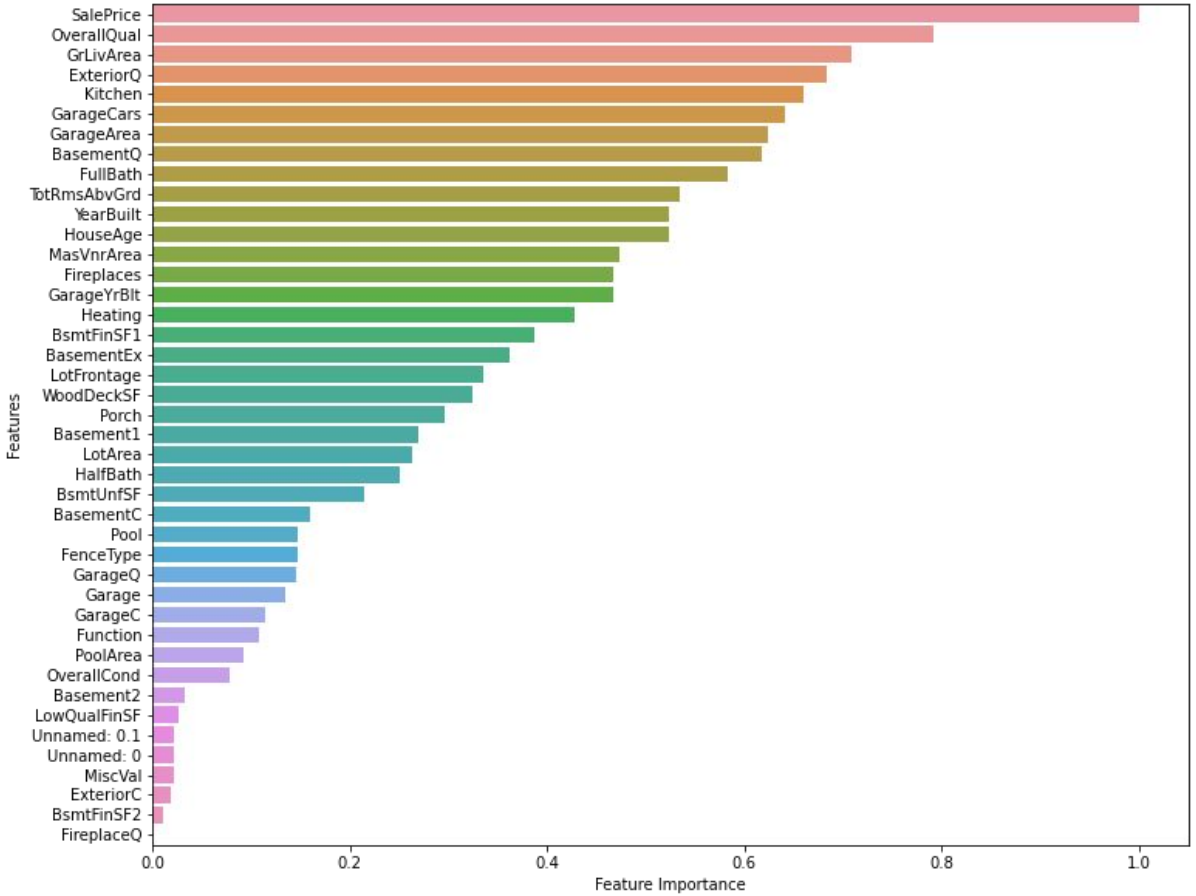
Steps in the Analysis (cont.)

- Exploratory Data Analysis
  - Checking for correlations with heatmap
  - Plotting sale price against other variables
  - Hypothesis testing
- Machine Learning
  - Run regression analyses

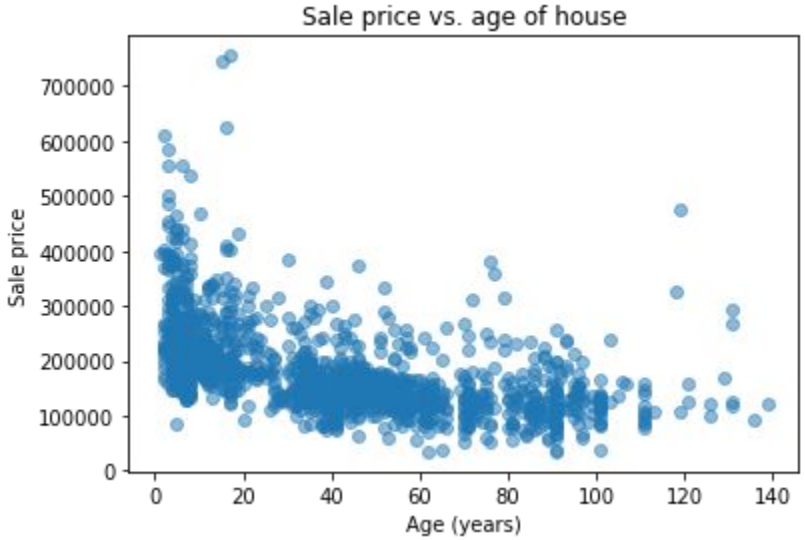
Feature Selection

Strongest features:

- Overall quality
- Living area
- External quality
- Kitchen
- Garage area



Exploratory Data Analysis

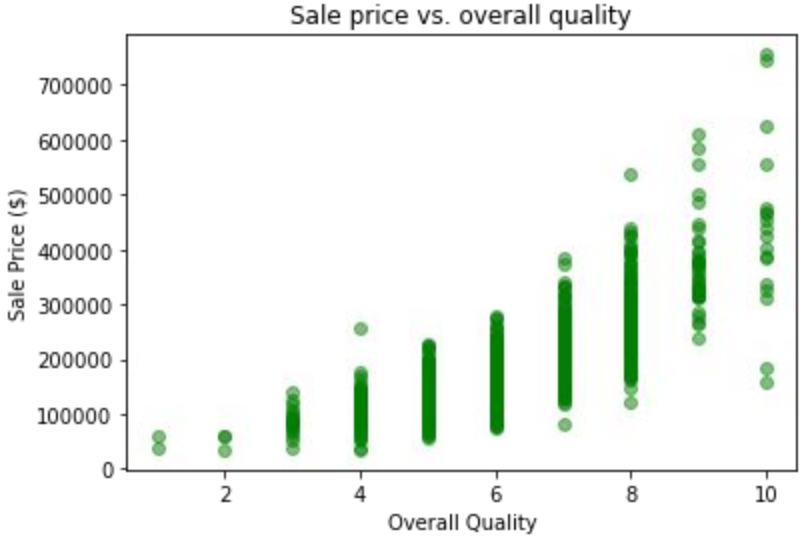


Plotting the age of the house against the sale price. Age of the house is negatively correlated with sale price.

Exploratory Data Analysis (cont.)

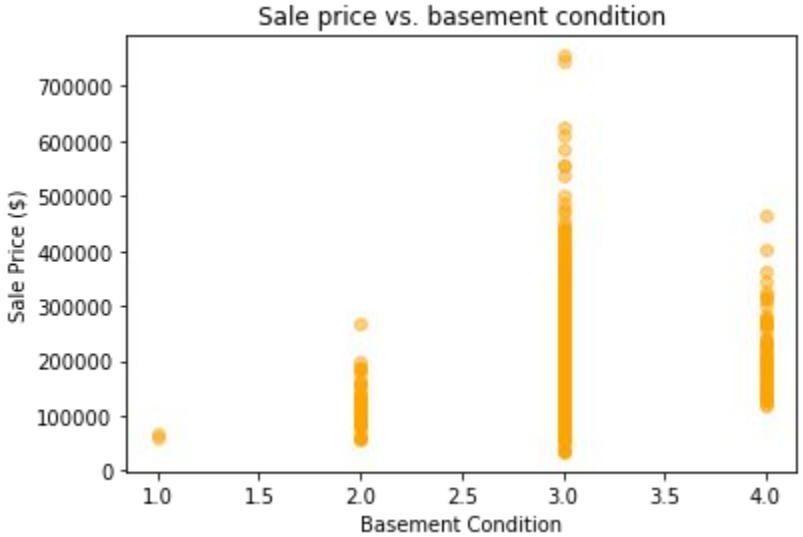
Plotting sale price against greater living area A strongly positive correlation can be seen between the living area and sale price.

Exploratory Data Analysis (cont.)



Plotting overall quality with sale price. Quality is positively correlated with sale price.

Exploratory Data Analysis (cont.)



Plotting basement condition against sale price. Houses with higher sales prices had basements with condition values of 3.

Machine learning

Comparing different regression models:

- Linear Regression
  - R-squared: 83.6%
- Ordinary Least Squares
  - R-squared: 80.0%
- Random Forest Regressor
  - R-squared: 90.0%
- SVM
  - R-squared: -0.03%

Conclusions

Most important features

- Overall quality
- Living area
- Exterior quality
- Kitchen
- Garage area

Conclusions (cont.)

Other external factors can influence housing prices:

- Geography
- Population
- Crime rates
- Proximity to schools

Dataset >10 years old, so more recent data may help.