King County House Prices

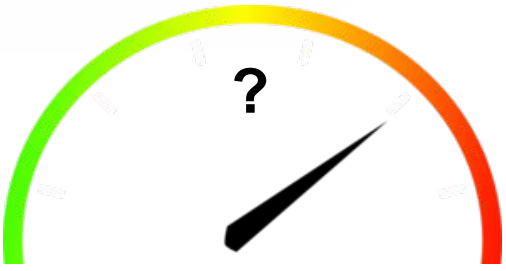
- Recommendations for Home Sellers & Buyers -

by Karima Chakroun
11/04/19

Welcome to my presentation about King County House prices - Recommendations for Home Sellers and Buyers.

- image source: https://storage.googleapis.com/kaggle-datasets-images/128/270/d149695d1f9a97ec54cf673be6430ad7/dataset-card.jpg

Problem statement



$ 78 000 **$ 7 700 000**

renovated grade waterfront
bathrooms floors basement
living area bedrooms location

As we all know, houses come in different flavors, which can range from nothing more than a simple hut to a king-sized palace with pools and gardens. Accordingly, also the selling price of a house can have an extreme range. For example, historical house sale prices in King County ranged from around 70 000 $ for the cheapest house up to over 7 Million dollars for the most expensive object.

Now, you want to buy or sell a house in King County and are faced with the difficult question: Where is my house located in this extreme price range?
The good news is that there are several features that determine the sales price of a house and can be used for price prediction. And this is exactly what I have done in this project. I explored the impact of different features on the house price with the goal to come up with some useful recommendations for you as a home seller or buyer.

- image sources:
  - https://www.flickr.com/photos/dyamasaki/463569490,
  - https://rdcnewscdn.realtor.com/wp-content/uploads/2018/02/KY-home-02-628x354.jpg

Business value

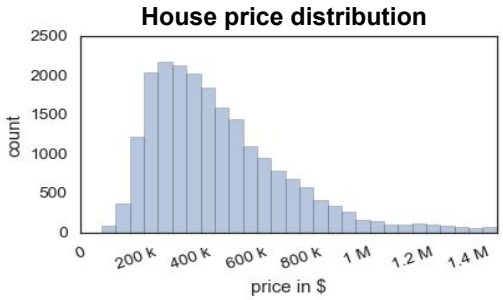
- Prediction of house sale prices
- Recommendations to increase sales price

Hence, the goal and added business value of this project is twofold:
First, to use the knowledge gained from the historical data to help home sellers & buyers to accurately predict the sales price of their house.
Second, to bring useful recommendation to home buyers to increase their house's sales price.

- image source: https://meine-bank-vor-ort.de/wp-content/uploads/2015/02/steigende-Preise.jpg

Methodology

- Dataset:
  - >20 000 houses in King County
  - 20 features to predict price
- Exploratory data analysis
- Linear Regression (simple, multiple)

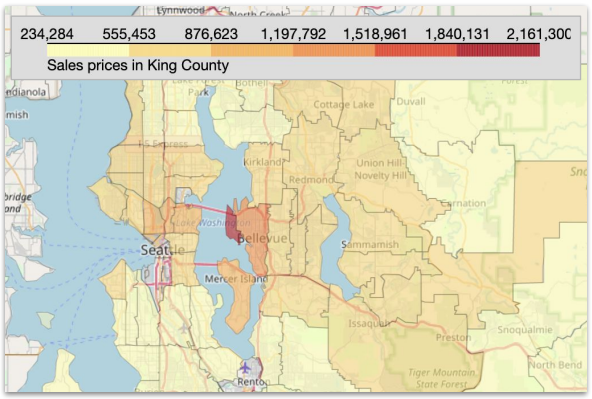


The dataset I used for this project included historical sales prices of more than 20 000 houses in King County, sold between 2014 and 2015, along with 20 features to predict price. Most of these houses were located in the price range between 100 thousand and 1 million dollar, but some of them also ranged up to several million dollars, as we saw before.

The project methodology included first a comprehensive exploratory data analysis and visualization, and second linear regression modeling, both simple and multiple regression, meaning I build mathematical models that linearly predict house prices based on one or several features.

- map source: https://de.wikipedia.org/wiki/King_County_(Washington)

Location (zipcode)



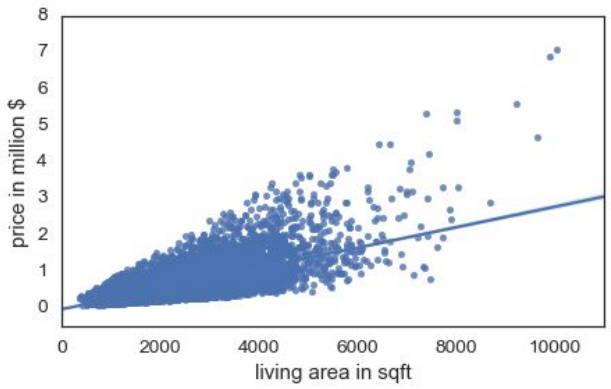
- explains **>40%** of price
- up to 10-fold price difference!

Recommendation 1: Choose the right location!

The first feature that had a major impact on price was found to be the location, coded in the dataset by the zipcode. Crucially, the zipcode alone already explained over 40% of the sales price. As you can see on this map, light yellow areas had a mean price around 200 thousand dollars, while the dark red area shows a mean price of over 2 million dollars - which is up to a 10-fold price difference explained by location alone.

Hence, my **first recommendation** (to home buyers) is to choose the right location for your house - and if you aim for very expensive objects, choose the areas of Mercer Island or Bellevue.

Living area

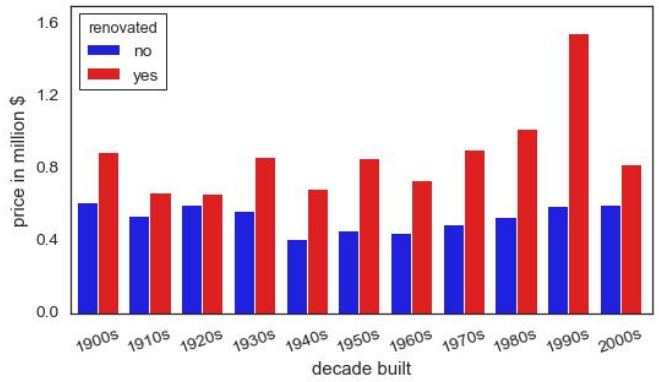


- explains **~50%** of price
- every 10 sqft raise price by **$ 2800**

Recommendation 2: Increase the interior living area!

The second important feature is the house's total interior living area (including both basement and above), which alone explained around 50% of price. More specifically: For every additional 10 sqft of interior living area, the mean price increased by around 2800 dollars.
Hence, my **second recommendation** to raise a house's price is to increase the house's living area, for example by building a winter garden, additional floors etc.

Renovation status



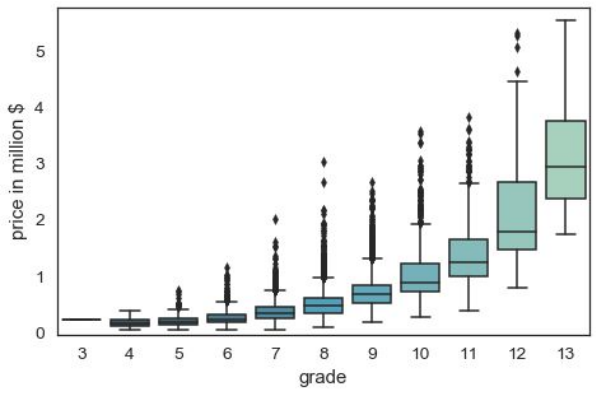
- renovated houses bring **$ 238 000** more
- < 5% of houses renovated

Recommendation 3: Renovate the house!

Third, the renovation status of a house is also a factor that significantly impacts the sales price of a house. Renovated houses (shown here in red) were on average worth 238 000 $ more than non-renovated houses (shown in blue). Note also that renovation raises the price not only for very old houses, but also for houses built only some 20 to 50 years ago.

Hence, my **third recommendation** would be to renovate the house, e.g. by modernizing electricity, thermal isolation, bathrooms and so on.

Grade



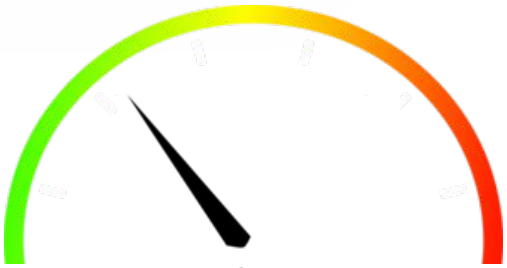
- explains **>50%** of price
- 8 to 9: **+$ 230 000**
- 11 to 12: **+$ 700 000**
- 12 to 13: **+$ 1 500 000**

Recommendation 4: Improve the house's grade!

Last but not least, King County has a house grading system and the grade of a house has a major impact on price, explaining alone more than 50% of the price. In fact, regression modeling revealed that the selling price shows not only a linear, but a quadratic increase with grade. For example, while increasing the grade from 8 to 9 brings only a raise in 230 000 $, an upgrade from 11 to 12 brings already a price jump of 700 000 $ and from 12 to 13 of more than a million $.

My **last recommendation** is therefore: try to increase the grade of the house, which could be done by renovations. For example, based on King County's grading system, an upgrade from 10 to 11 can be achieved by adding amenities of solid woods and luxurious options to your house.

Conclusion



$ 78 000 **$ 7 700 000**

>80%

renovation **grade** waterfront

bathrooms floors basement

living area bedrooms **location**

In conclusion, I presented you today several features that were found to have a major impact on house selling prices in King County, the most important ones being location, living area, renovation status and house grade. Including these and further relevant features in a final multiple regression model, these predictors could together explain more than 80% of the variance in housing selling prices in King County.

Hopefully, I could also give you some useful recommendations not only to determine an appropriate price for selling or buying a house in King County, but also some tips for home sellers on how to significantly boost the selling price of your house.

Future work

- Feature engineering:
  - mathematical transformations
  - grouping to new categories

- Additional location data:
  - Foursquare, Google maps
  - nearby shops, schools, parks, …

Some prospects for future work include, first, feature engineering, meaning we could take the features we already have to build new features, which might be even better predictors of price. For example, numerical features could be transformed mathematically by taking the logarithm or some power function of a feature (as we have seen for the feature ‘grade’ before). Or, categorical variables could be grouped into new, meaningful categories. For example, we might combine several zip codes to larger areas.

A second idea for future work would be to use of additional location data, which we could get e.g. via Foursquare or Google maps, to build new features to predict price. For example, using our longitude and latitude data of each, we could get nearby venues or that house (shops, schools, parks, …) and include these new features in our regression model.

- image source: https://geospatialmedia.s3.amazonaws.com/wp-content/uploads/2018/04/multi-location.jpg

Thank you!

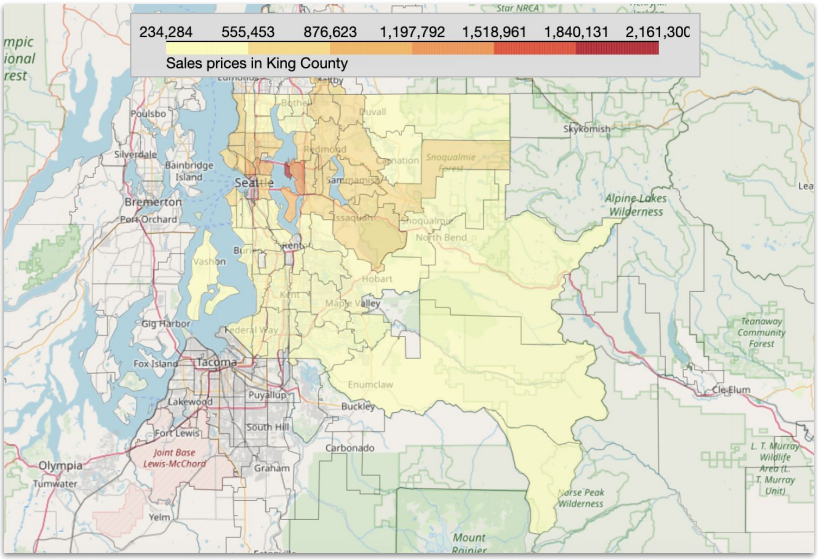
Eli
Dirk, Larissa
Bootcamp team!

Thank you for listening, and also a big thank you to Eli, Dirk, Larissa and the whole bootcamp team for your support in this project.

Appendix

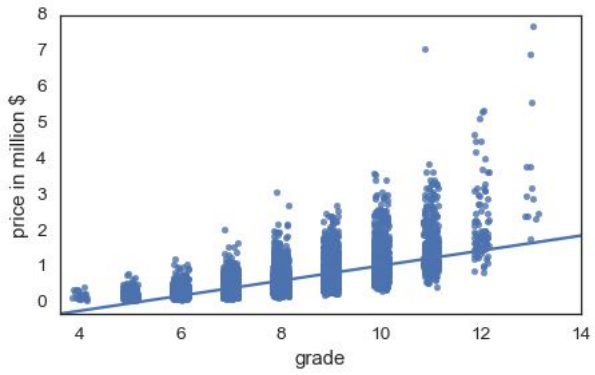
1) Full price map of King County
2) Grade as quadratic predictor
3) King County Grading System
4) Heatmap - Correlation analysis
5) GitHub project repository

Full price map of King County

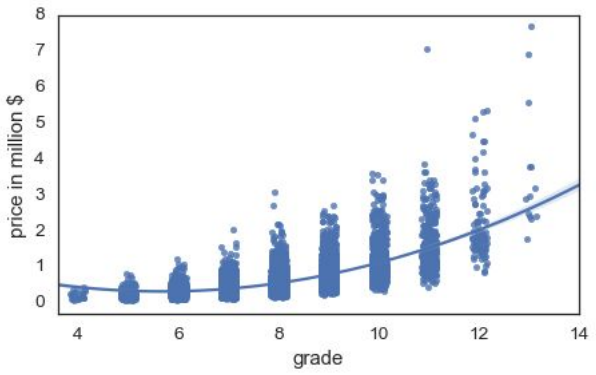


- map shows the mean sales price for each of the zipcodes (included in the dataset)
  - lowest mean price: zipcode 98002 ($ 234 284) - Auburn, Washington
  - highest mean price: zipcode 98039 ($ 2161300) - Medina, Washington

Grade as quadratic predictor



Linear relationship: explains 44.6% of price



Quadratic relationship: explains 51.1% of price

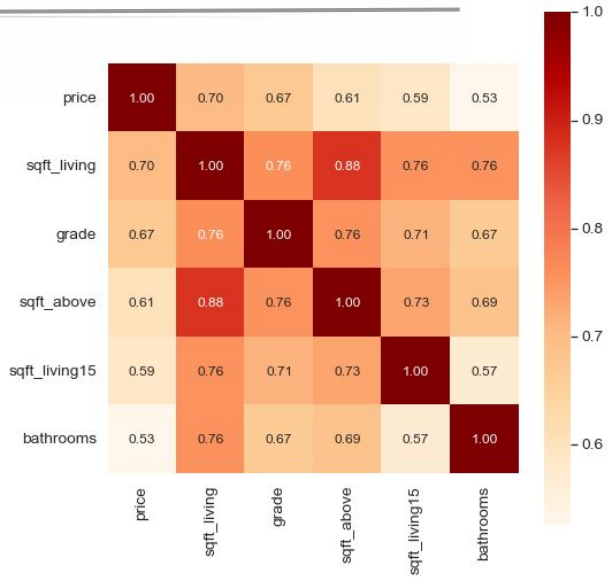
- linear regression modeling shows that the relationship between grade and price is quadratic rather than linear
  - model including only linear relationship between grade and price: explains 44.6% of price
  - model including quadratic relationship between grade and price: explains 51.1% of price

King County Grading System

- **Grade 1-3:** Falls short of minimum building standards. Normally cabin or inferior structure.
- **Grade 4:** Generally older low quality construction. Does not meet code.
- **Grade 5:** Lower construction costs and workmanship. Small, simple design.
- **Grade 6:** Lowest grade currently meeting building codes. Low quality materials, simple designs.
- **Grade 7:** Average grade of construction and design. Commonly seen in plats and older subdivisions.
- **Grade 8:** Just above average in construction and design. Usually better materials in both the exterior and interior finishes.
- **Grade 9:** Better architectural design, with extra exterior and interior design and quality.
- **Grade 10:** Homes of this quality generally have high quality features. Finish work is better, and more design quality is seen in the floor plans and larger square footage.
- **Grade 11**: Custom design and higher quality finish work, with added amenities of solid woods, bathroom fixtures & more luxurious options.
- **Grade 12:** Custom design and excellent builders. All materials are of the highest quality and all conveniences are present.
- **Grade 13:** Generally custom designed and built. Approaching the Mansion level. Large amount of highest quality cabinet work, wood trim and marble; large entries.

Link:
https://www.kingcounty.gov/depts/assessor/~/media/depts/Assessor/documents/AreaReports/2018/Residential/015.ashx

Heatmap - Correlation analysis



- heatmap showing the Pearson correlation coefficients between sales price and the five numerical (non-categorical) features that correlate highest with price
    - ‘grade’: overall grade given to the housing unit, based on King County grading system
    - ‘sqft_living’: square footage of total interior housing living space (both basement and above)
    - ‘sqft_above’: square footage of interior housing living space above ground
    - ‘sqft_living15’: square footage of interior housing living space for the nearest 15 neighbors
    - ‘bathrooms’: number of bathrooms

GitHub project repository

This repository contains a data science project based on the King County House Sales dataset. The dataset can be found in the file "kc_house_data.csv" in this repository, and a description of the corresponding column names can be found in the “column_names.ipynb” file.

The Data Science Life Cycle goals for this project include: Data Cleaning, Data Exploration, Data Visualization and Predictive Modeling (by linear regression).

All parts of the data analysis are documented in the notebook “House_Prices_Project_Karima.ipynb” and include the following steps:

1) Load Packages and Dataset
2) Data Cleaning:
    a) Data types (deals with conversion of data types)
    b) Missing data (handles missing values)
3) Data Exploration & Visualization
    a) Overview of all features (using scatter_matrix plots)
    b) Exploring features one by one (histograms, pie/bar charts, scatter/regression plots)
    c) Correlation analysis (heatmaps)
    d) Folium map (mean price per zipcode)
4) Multiple Regression Model
    a) Model fitting and summary statistics
    b) Multicollinearity (checking variance inflation factors)
5) Customize Plots for Business Presentation

- readme of the GitHub project repository
- link to the GitHub project repository: https://github.com/KarimaCha/House_Prices_Project