1st Project - Exploratory Data Analysis

Dataset: King County Houses

Name: Jonas Bechthold
Date: 2020-09-17
Location: Hamburg, neue fische

AGENDA

01 **Introduction**
Where to buy a house?

02 **Data Set**
General Overview

03 **Where can you find a good house?**
Criteria for selection and buying

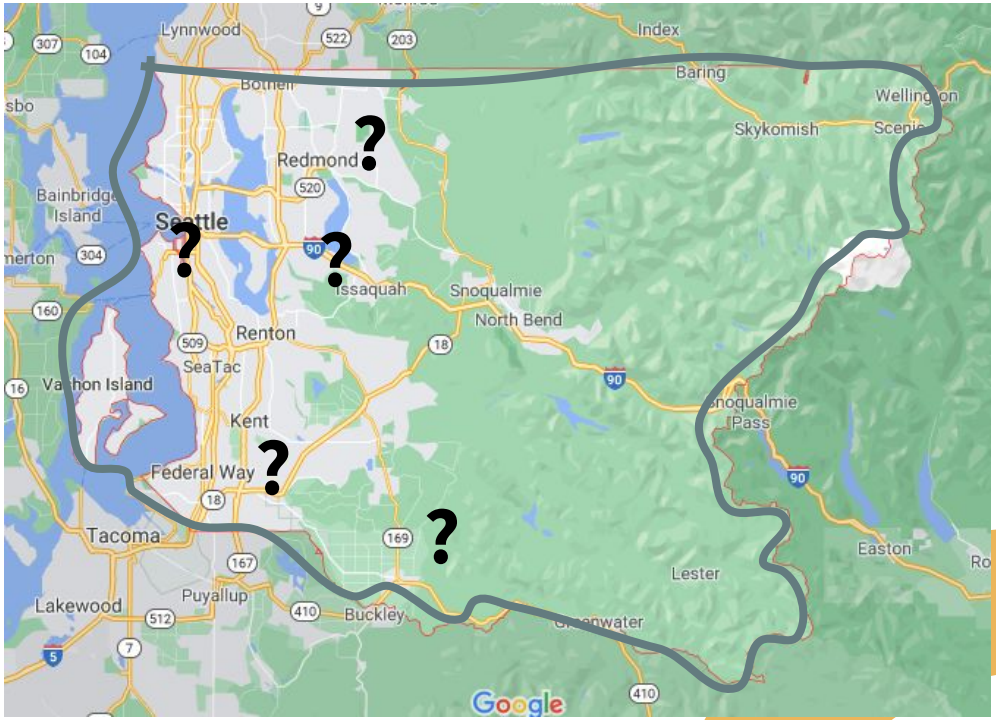
04 **Price Prediction Tool**
Estimate the price by using different parameter

01 | Introduction & Motivation

- King County (Washington)
- 2 Areas
  a. Seattle Area
  b. Countryside

- Buying a house:
  - City and/or countryside
  - Transportation/roads
  - Neighbourhood
  - **PRICE!!!**

- **How to find out where?**



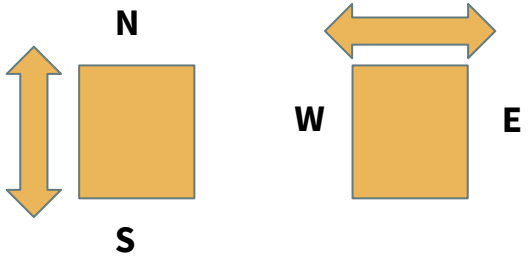
02 | Data Set

King County House Sales: May 2014 to May 2015

- 21597 Sales with a long list of items

21 different features:

1. **Price**
2. **Dates**: date of sale, year built, year renovated
3. **House**: Area (lot, living, basement), bedrooms, bathrooms, floors
4. **Location:** Zip, latitude, longitude
5. **Ratings**



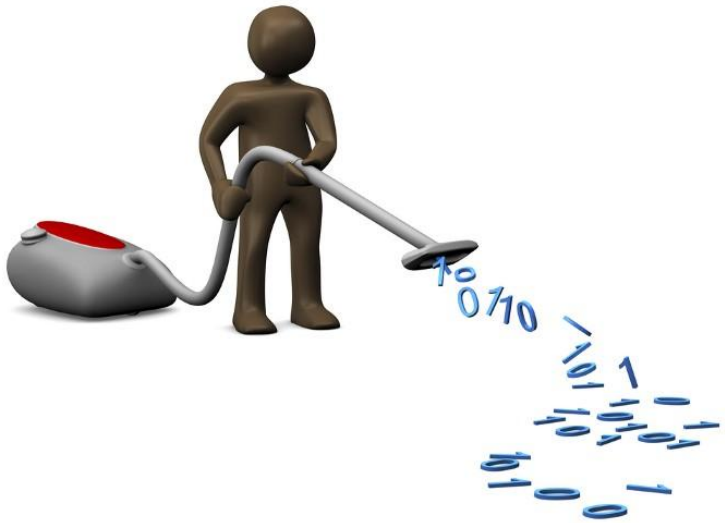
02 | Data Set

Start: Clean Data

- use statistics of non missing data to fill them -> add
- unable to rebuilt data -> delete

Examples with this data set:

- total living area = living area + basement area
- missing renovation year = no renovation
- delete part of the dataset if non-reconstructable



Source: https://medium.com/cracking-the-data-science-interview/an-introduction-to-big-data-data-cleaning-a238725a9b2d

03 | Where can you find a good house?

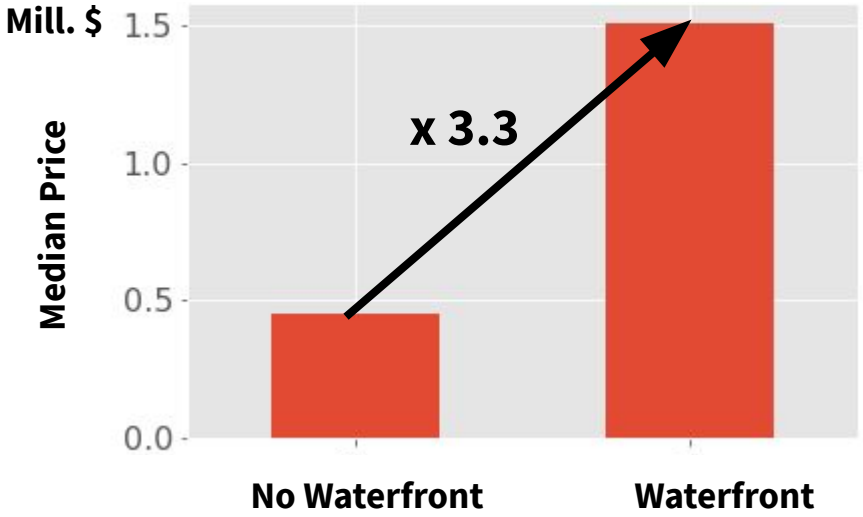
A family wants to find a new house in King County

Space: parents + three children
Location: as close by the city as possible (Infrastructure, school/work distance, leisure)
Price: fair or cheap price
Optional: renovation work (if pays off)

Source: https://www.huntergalloway.com.au/deposit-calculator/

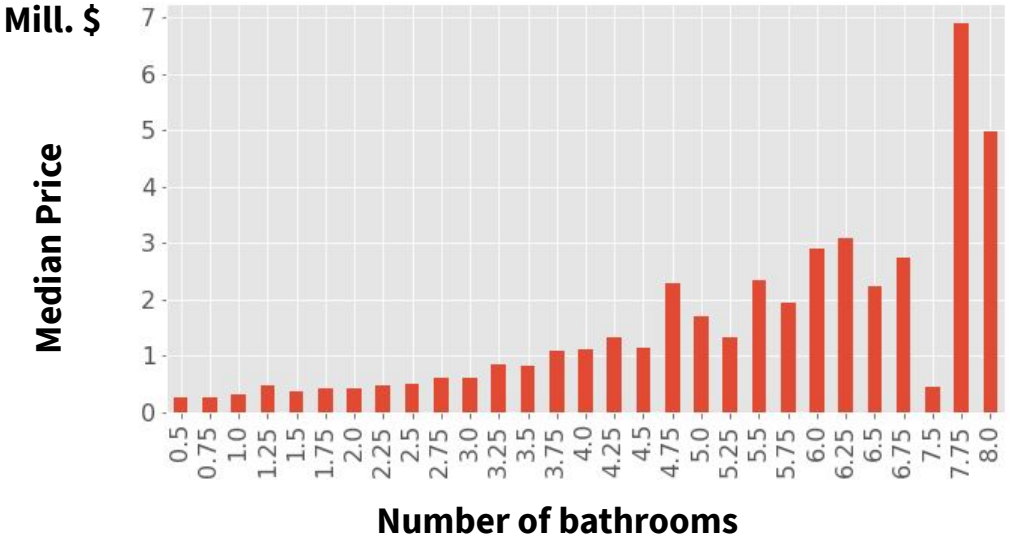
03 | Where can you find a good house?

Recom. #1 - Avoid the waterfront



03 | Where can you find a good house?

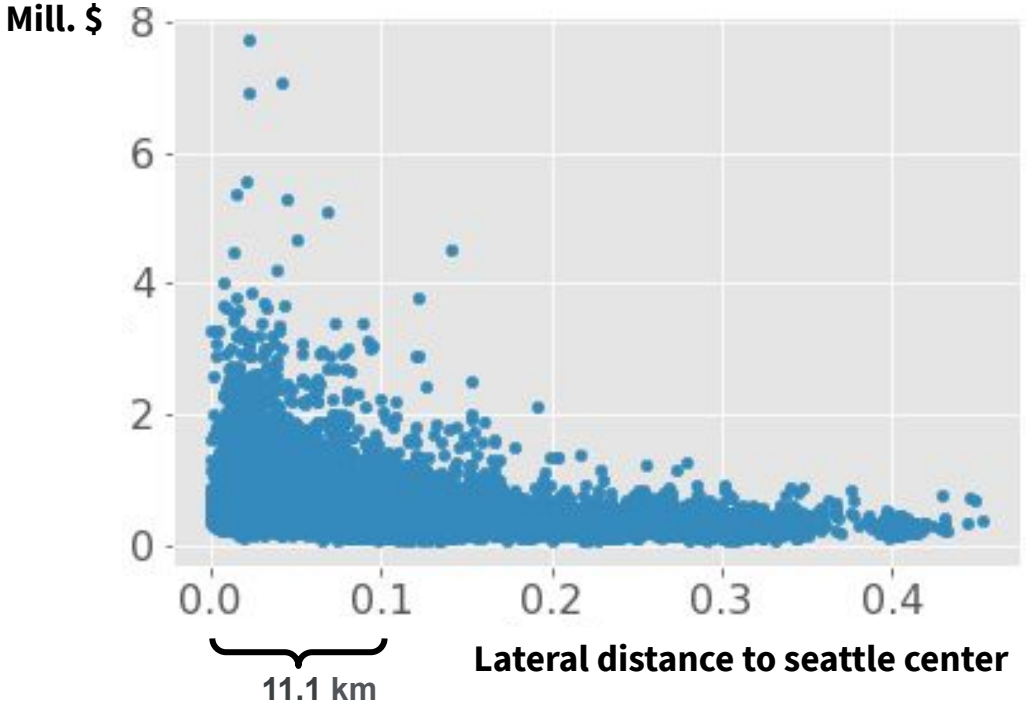
Recom. #2 - Think about how many bathrooms you need!



Bathrooms seems to be very expensive.

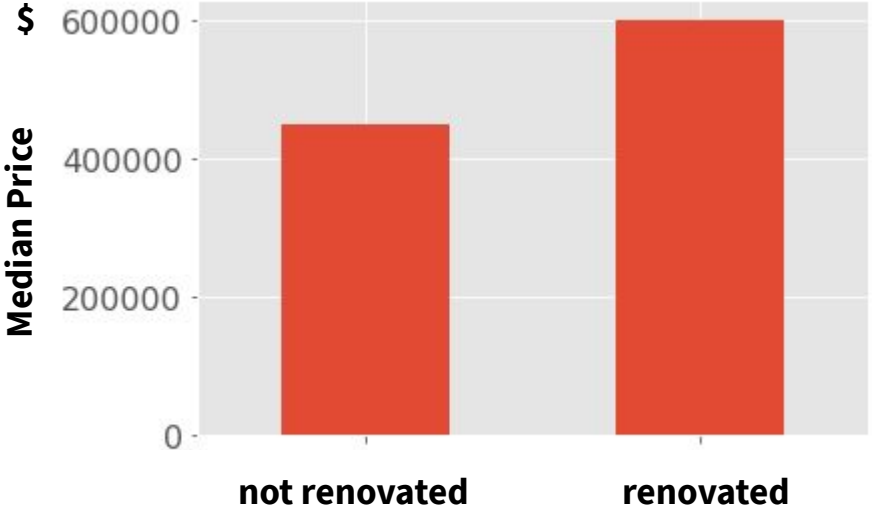
03 | Where can you find a good house?

Recom. #3 - A bit further north or south of Seattle center saves money !



03 | Where can you find a good house?

Recom. #4 - Self renovation could save money!



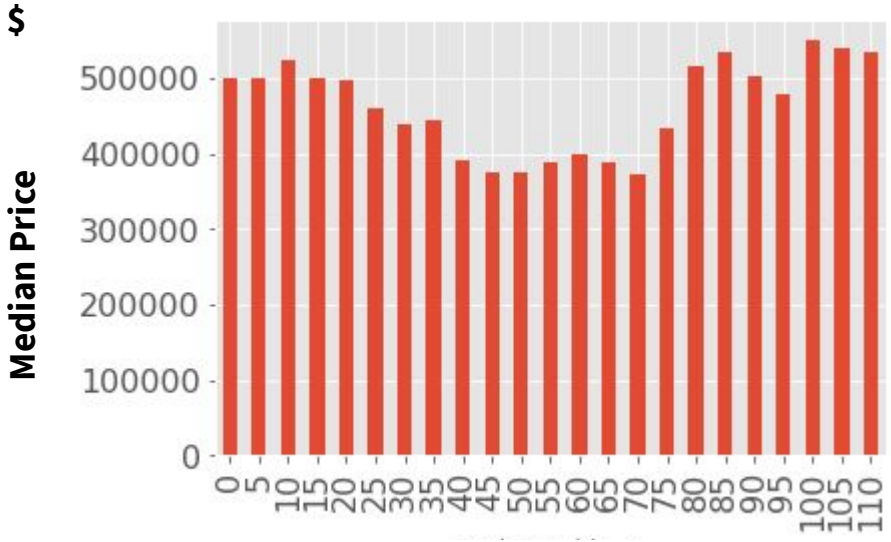
Invest the difference of

150000 $

to have a newly renovated house (increased value afterwards)

03 | Where can you find a good house?

Recom. #5 - Look at the “real age” of the house!



Threshold (20-25 years) to price decrease!

04 | Price Prediction Tool

A linear regression model was built based on multiple input parameter:

- *bedrooms* - number of bedrooms
- *floors* - number of floors
- *waterfront* - At Waterfront?
- *view* - Rating of the view
- *sqft_living* - total living area
- *grade* - gold county grading system
- *sqft_living15* - interior space of the 15 nearest neighbors

$$R^2 = \frac{\text{Variance explained by the model}}{\text{Total variance}}$$

R^2 = 0.715

Developed features:

- ***lat_dist_sea*** - distance to city center of Seattle (lat = 47.6 °)
- ***real_age*** - years from year of sale to last renovation
- ***was_renovated*** - renovated?
- ***years_old*** - how old is the house when sold?

Thank you for your attention!

Source: https://www.seattletimes.com/seattle-news/data/seattle-now-most-expensive-city-for-renters-outside-california-census-data-shows/