

Real Estate EDA

For the area King County (WA)



Agenda

A wooden block house, constructed from light-colored wooden planks, sits on a patch of green grass. The house has a simple A-frame roof and a small arched entrance at the base. To the right of the house, a series of five green horizontal bars, each topped with a small green house icon, descend in a staircase pattern from the top right towards the bottom left, mirroring the shape of the wooden house.

Introduction

Exploratory data analysis

Results

Key Takeaways

Outlook

Introduction

What have the following companies in common?

SLALOM
CONSULTING

K2

QFC
The Omni Group

QFC
Quality Food Centers

napster
a Rhapsody® company

gettyimages®

tableau

JUNO
THERAPEUTICS

ClearSign

nanoString
TECHNOLOGIES

CRAY
THE SUPERCOMPUTER COMPANY

SP
BUCKER PUNCH

ZYMOGENETICS

a

Wizards
OF THE COAST®

Starbucks

VALVE

Microsoft

zulily

PEMCO
Insurance

f5

Atossa
GENETICS
INCORPORATED

VULCAN

Solutions for an application world.

avanade

KREI

PopCap
GAMES

BIG FISH
APP DEVELOPMENT STUDIO

COSTCO
WHOLESALE

whitepages
SEARCH. FIND. KNOW.

Introduction

They are all based in the King County (WA)



- **Area** 2,307 sq mi (5,980 km²)
- **Population (estimate 2019)** 2,252,782
- **35 cities, 4 towns and 14 ghost towns**

Introduction

Benefits

- Better understand of the real estate market in general.
- Indicators for low and high property prices
- Insights into the process of data analysis.



The dataset

- 21,597 successfully sold properties
 - 21 features each
- } = 453,537 feature characteristics
- Dated between winter 2014 and autumn 2015
 - Prices from 78,000 up to 7,700,000 USD
 - From bungalows up to 3 floor mansions

Exploratory data analysis

Bad news

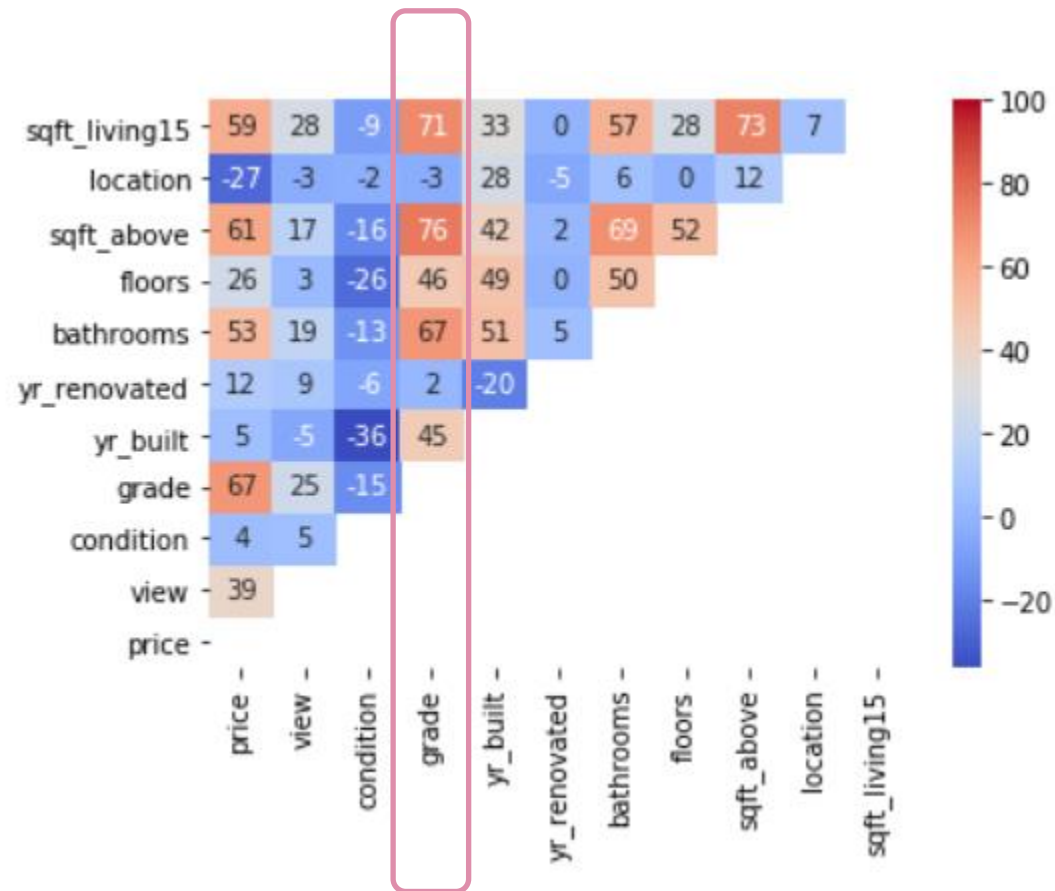
- Missing feature characteristics
- $\frac{3}{4}$ bathrooms and $\frac{1}{2}$ floors
- Incorrect data
- Little data for some features
- Partially missing description

Good news

- Numerous helpful information
- No duplicates
- Formatting mostly accurate



Exploratory data analysis



Exploratory data analysis / Results

Grouping the data

- Houses with 2,5 floors were more expensive
- Most houses are graded between 7 and 9
- Houses with up to eight bathrooms
- The condition of the house has a smaller impact than grade
- Most of the houses was tagged with no view and no waterfront



A small, rustic wooden house with a gabled roof, constructed from light-colored wood planks, sits on a patch of green grass. The house has a chimney on the left side and a small arched entrance.

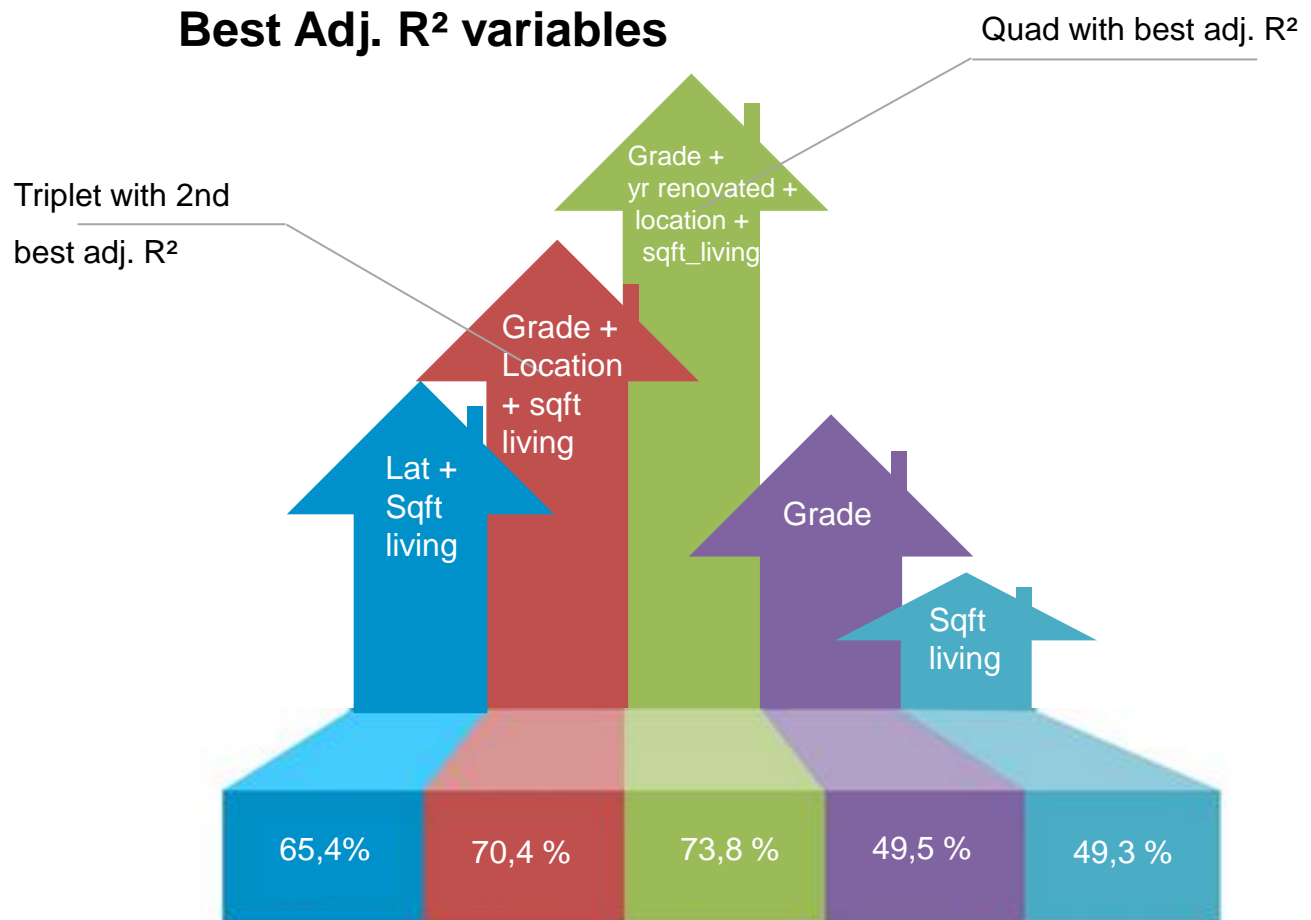
Exploratory data analysis

(Multiple) Linear Regression

$$R^2 = 1 - \frac{\text{Unexplained Variation}}{\text{Total Variation}}$$

$$R_{adj}^2 = 1 - \left[\frac{(1 - R^2)(n - 1)}{n - k - 1} \right]$$

Results

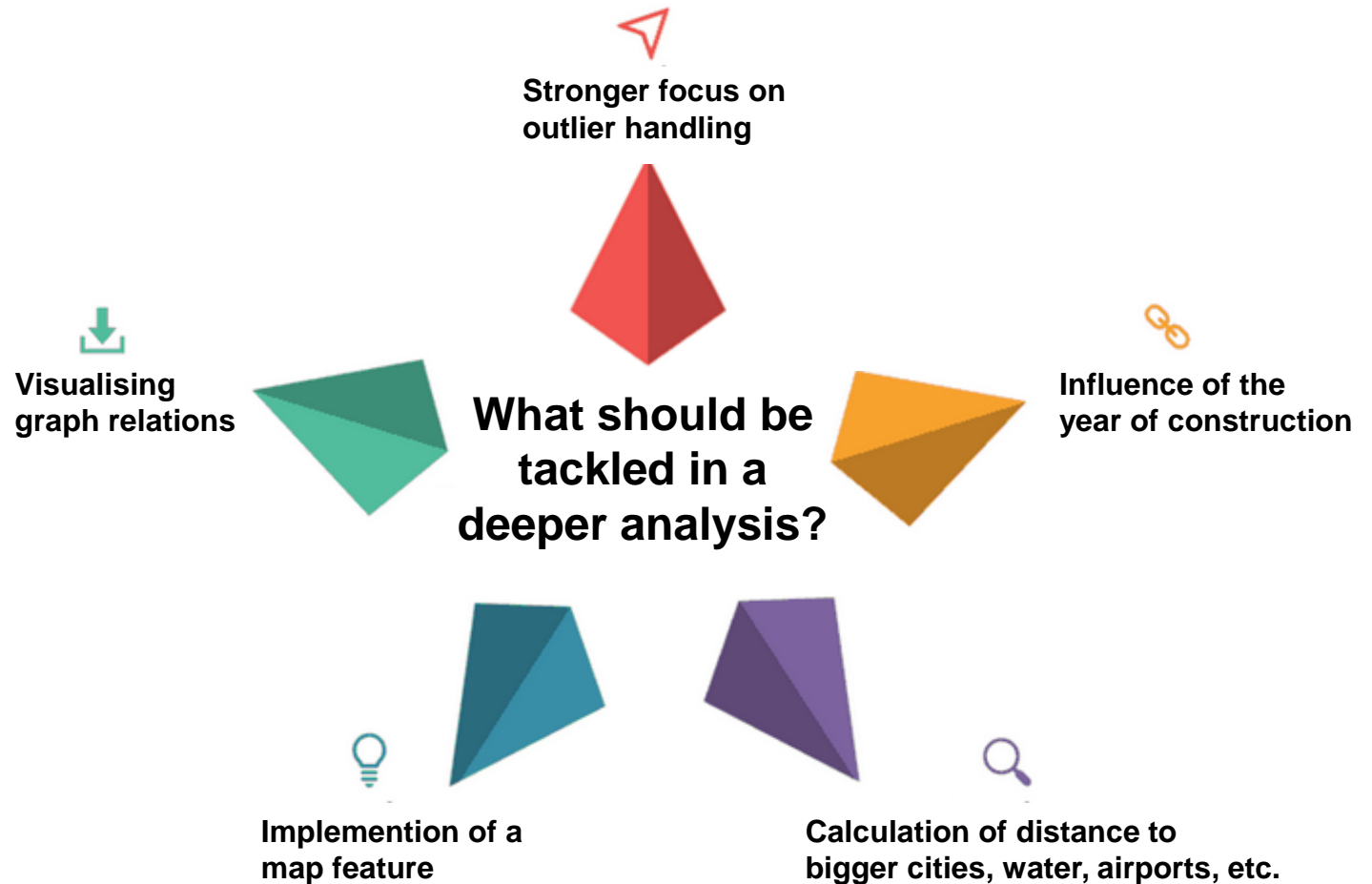


Key Takeaways

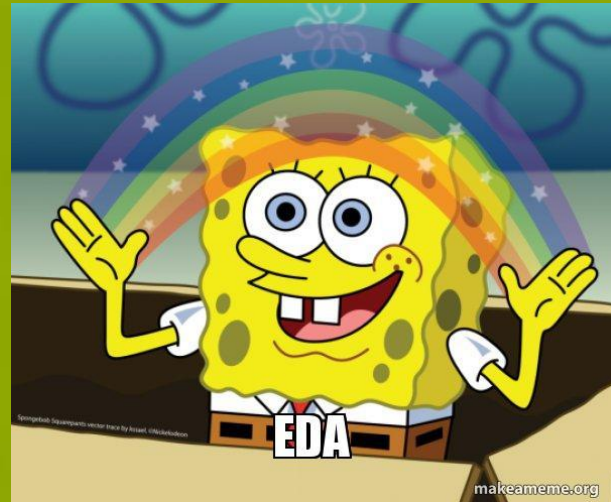
- Do not consider the condition as a key indicator for your choice
- Houses graded 3 and 4 has the best value for price
- The amount of bedrooms only a small impact on the grading
- Houses with more than one bathrooms will increase the grading significantly
- Houses with a view cost at least twice as much as without view
- Houses with a waterview will at least triple the price



Outlook



Liked the presentation?



Visit my [GIT](#) and take a look at the EDA in the jupyter notebook.