Real Estate EDA

For the area King County (WA)

Agenda

- Introduction
- Exploratory data analysis
- Results
- Key Takeaways
- Outlook

Introduction

What have the following companies in common?

SLALOM CONSULTING · K2 · The Omni Group · QFC Quality Food Centers · napster a Rhapsody® company

gettyimages · +ableau · JUNO THERAPEUTICS · ClearSign

nanoString TECHNOLOGIES · CRAY THE SUPERCOMPUTER COMPANY · SP SUCKER PUNCH · ZYMOGENETICS

a · Wizards of the Coast · VALVE · Microsoft

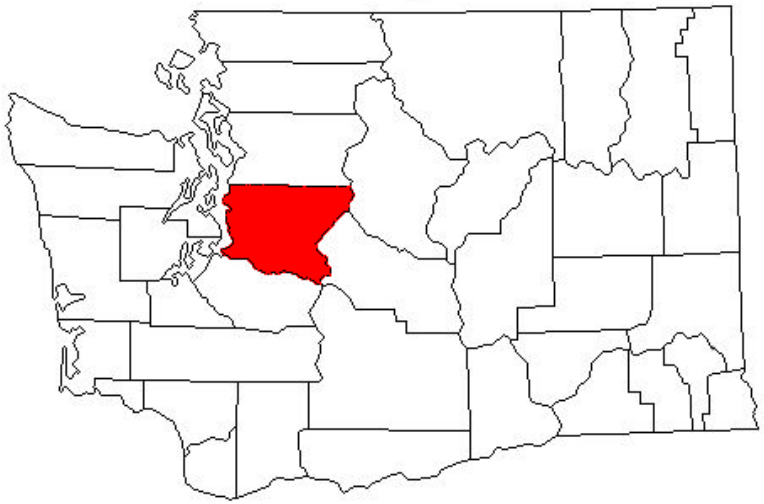
zulily · PEMCO Insurance · f5 · Atossa GENETICS INCORPORATED · VULCAN

Solutions for an application world.

avanade · Whitepages SEARCH. FIND. KNOW. · REI · PopCap GAMES · BIG FISH APP DEVELOPMENT STUDIO · COSTCO WHOLESALE

Introduction

They are all based in the King County (WA)



King County

- **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 spring 2014 and spring 2015**
- **Prices from 78,000 up to 7,700,000 USD**
- **From bungalows up to 3 floor mansions**

Exploratory data analysis

Bad news

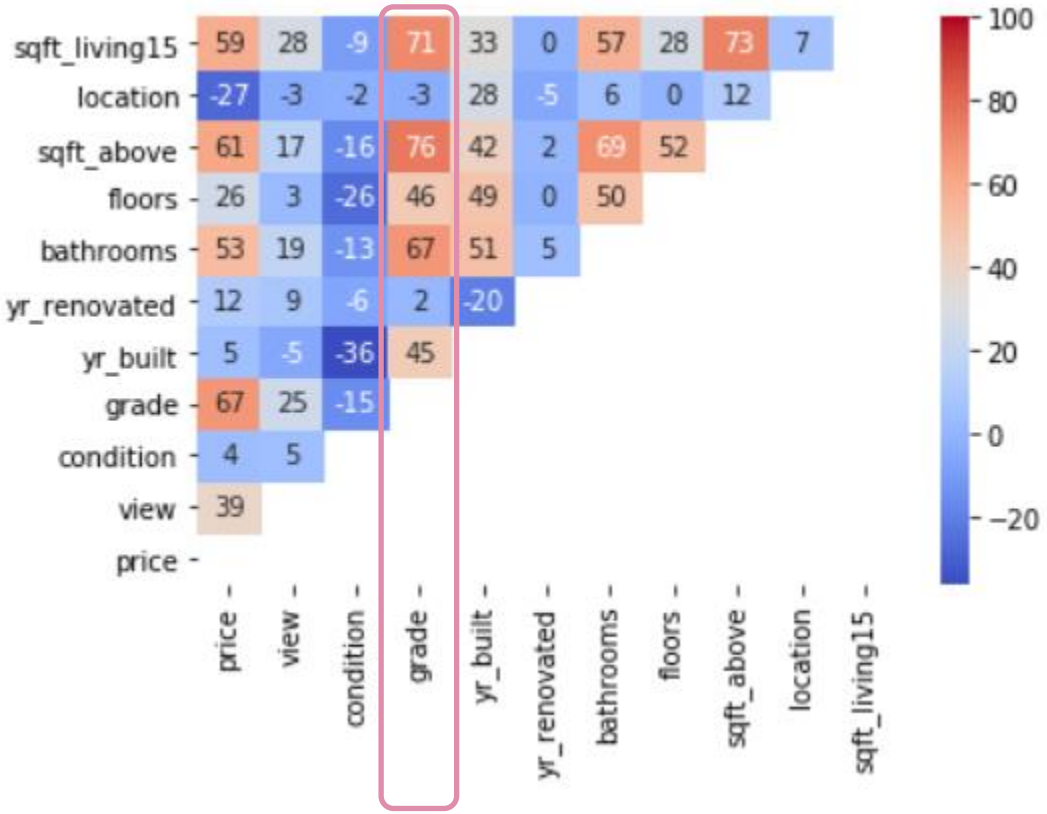
- Missing feature characteristics
- ¾ bathrooms and ½ floors
- Incorrect data
- Little data for some features
- Partially missing description

Good news

- Numerous helpful information
- No duplicates
- Formating mostly accurate

Exploratory data analysis

Correlation between the variables



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**

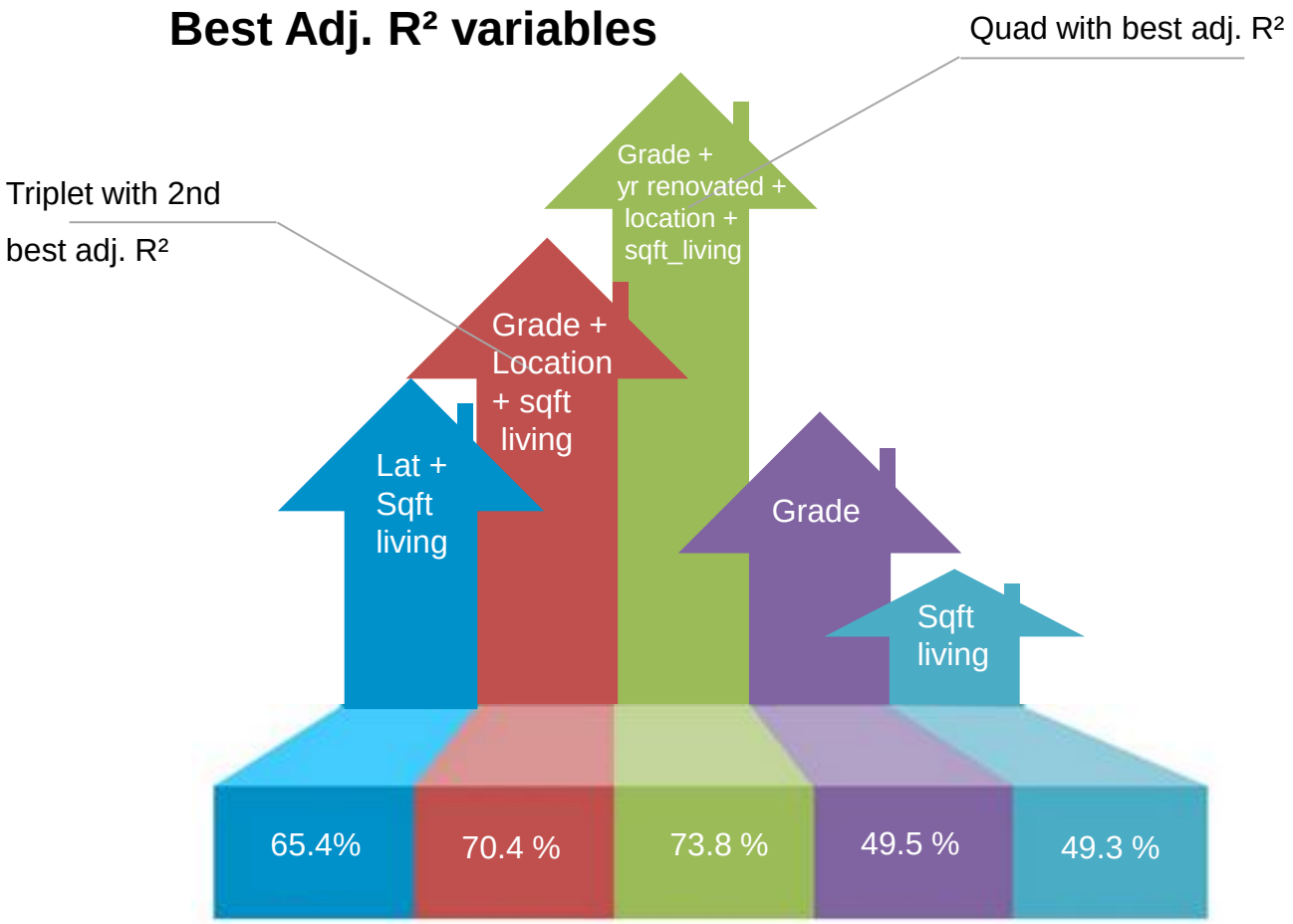
Exploratory data analysis

(Multiple) Linear Regression

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

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

Results

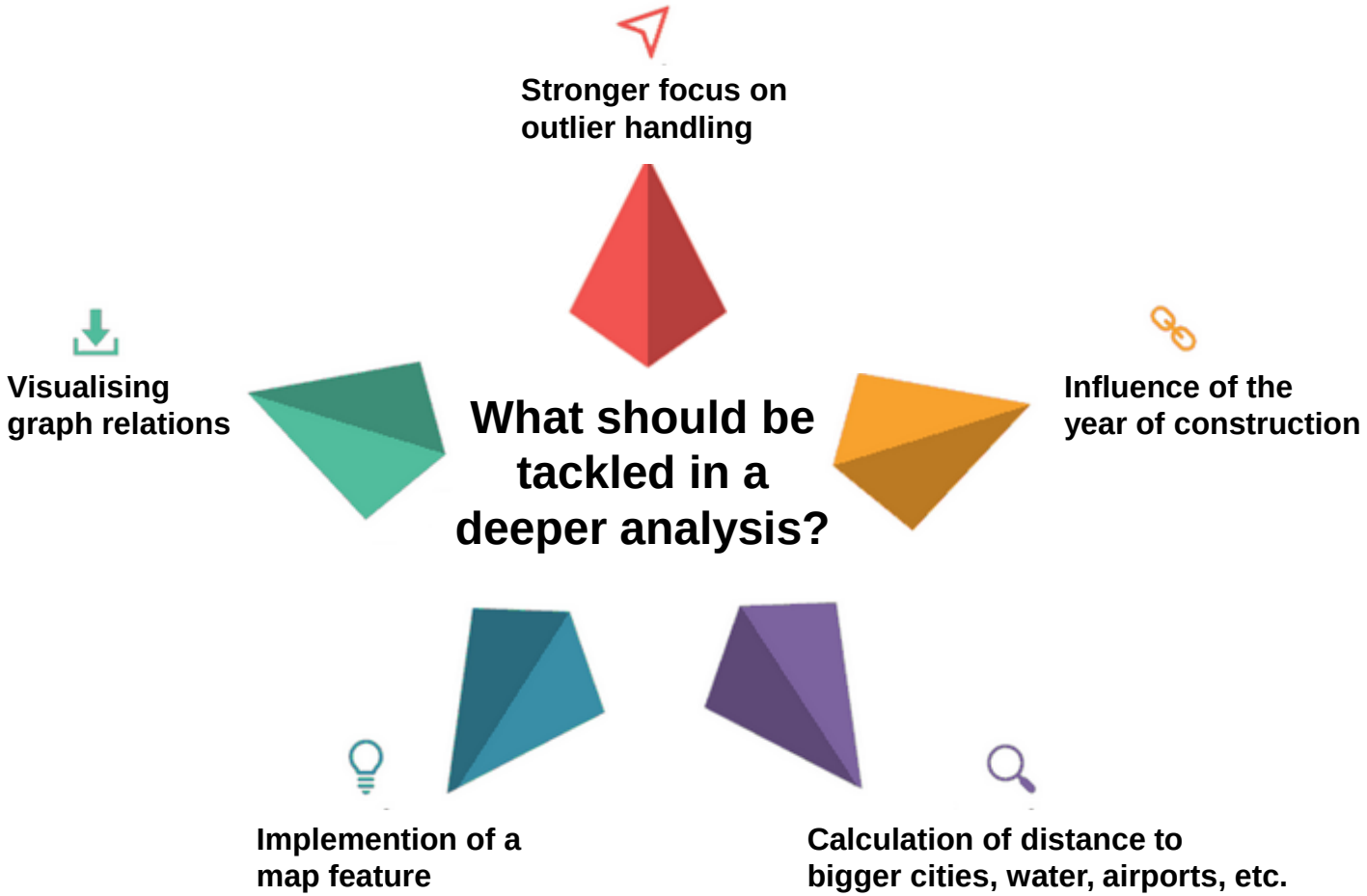


Key Takeaways

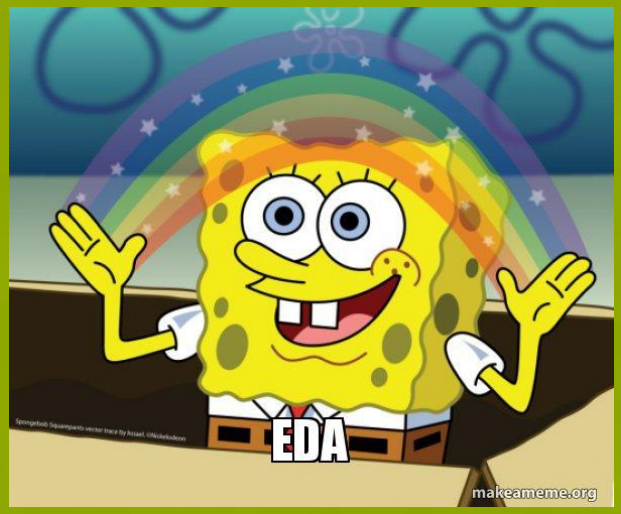
- **Do not consider the condition as a key indicator for your choice**
- **Houses graded 7 and 8 has the best value for price**

- **The amount of bedrooms had only a small impact on the grading**
- **Houses with more than one bathrooms will increase the grading signifcantly**
- **Houses with a view cost at least twice as much as without view**
- **Houses with a waterview will at least triple the price**

Outlook


Stronger focus on outlier handling
Visualising graph relations
What should be tackled in a deeper analysis?
Influence of the year of construction
Implemention of a map feature
Calculation of distance to bigger cities, water, airports, etc.

Liked the presentation?



Visit my **GIT** and take a look at the EDA in the juypter notebook.