Mid-Bootcamp Project
Predicting
Real Estate Prices

by Janek Stein

01

INITIAL SITUATION

Expectations

TASKS

DATA EXPLORATION

Using different Python packages and SQL to get a better understanding about data structure and properties

BUILD A MODEL

Using different models like Linear Regression, KNN and Decision Tree Regression to build an insightful prediction model

VISUALIZE

Using Tableau and Python packages to dive deeper into the data

02

DATA

Information about the data used

LOCATION

- Dataset contains records of real estate purchases in King County, WA

21,597

OBSERVATIONS

22

FEATURES

21,420

UNIQUE OBSERVATIONS

22

FEATURES

16

INTEGERS

4

FLOATS

1

OBJECT

CORRELATION MATRIX

- Correlation coefficient values range from close to 0 to 0.7
- Signs for multicollinearity

SKEWED DISTRIBUTIONS

- Some features show skewed distributions

OUTLIERS

- Some features have outliers

03

RESULTS

Choice of prediction models
and properties

PREDICTION MODELS USED

LINEAR REGRESSION

KNN

DECISION TREE REGRESSION

ALSO: LASSO, RIDGE AND ELASTIC NET

COMPARING PREDICTION ACCURACY OF MODELS

	BASELINE	BEST KNN PREDICTION	FITTED MODEL WITH LEAST FEATURES
LINEAR REGRESSION	62.15	57.28	51.62
KNN	61.16	77.62 (k=6)	74.43 (k=6)
DECISION TREE	58.42	64.70	52.33

PROPERTIES OF MY PREDICTION MODEL

- **KNN Model**
- **7 Features:**
  'bathrooms', 'floors', 'condition', 'grade', 'zipcode', 'sqft_living15', 'sqft_lot15'
- **k=6**

04

ANALYSIS PROCESS

How I got to my results

PATH

Easy tasks:

- Pandas DataFrames and cleaning of data/manipulating
- Working with SQL

Medium

- Working with Tableau

Hard

- Choosing the right model
- Choosing the right selection of features

MY LEARNINGS

TRY UNTIL IT WORKS

Small steps to final results

Iterate!

THANKS!

When you have further questions,
feel free to contact me!

Janek Stein
LinkedIn: **@janekstein**
GitHub: **@jast92**