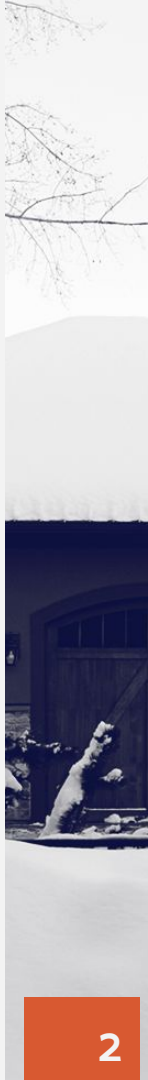
KING COUNTY

HOUSE PRICES ANALYSIS

Regression Analysis
of house sales prices

Fabian Werkmeister

TOPICS



AUDIENCE

REAL ESTATE AGENCY

HIGH PRICE SEGMENT

Operates for years in high price segment

Not specialized in luxury listings

No exclusive network or access to VIPs/celebrities

DATA BASIS

Little to no digital documentation of customer data

Heavily dependent on expertise and knowledge of skilled realtors

DATA SCIENCE

Want to get to know what data science is in order to evaluate if data science can be beneficial and/or reveal anything new to them

Propose an easy framework for sales predictions

ANALYSIS OBJECTIVES

FACTORS

What are the objective factors that determine sales prices in the high price real estate segment

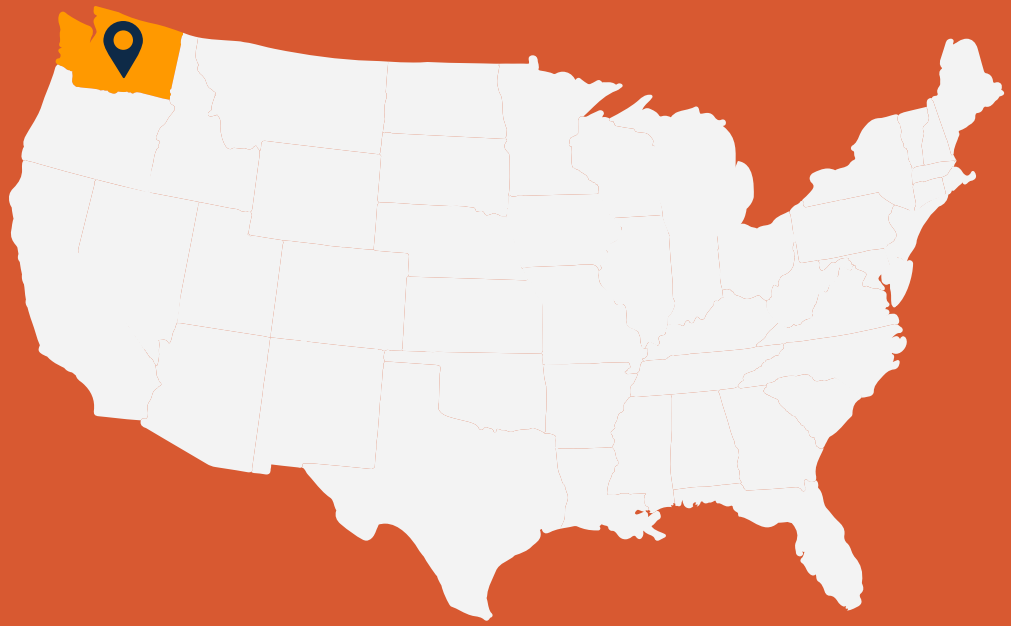
IMPACT

How and to which extent do these factors contribute to sales prices

LOCATION

Which neighbourhoods and locations in King County should be focused on

DATA BASIS



HOUSE SALE PRICES

MAY 2014 - MAY 2015

KING COUNTY, WA

21.613 OBSERVATIONS

DISTRIBUTION OF PRICES

DESCRIPTIVE STATISTICS HOUSE PRICES

Mean
540.296 $

Median
450.000 $

Minimum
78.000 $

Maximum
7.700.000 $

25 % of all prices are above
645.000 $

FACTORS WHICH DETERMINE THE SALES PRICE

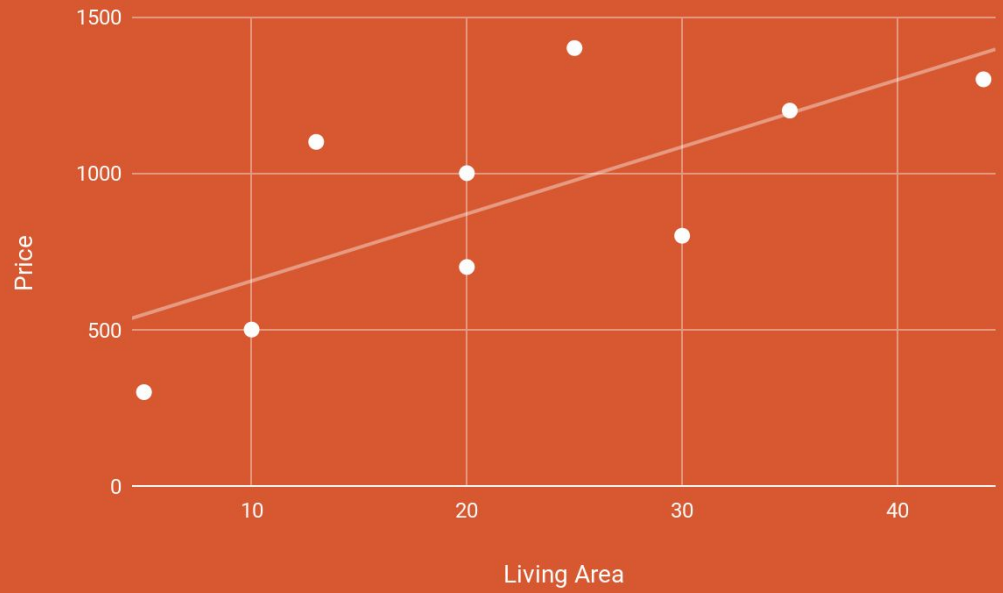
REGRESSION ANALYSIS

Linear regression quantifies the relationship between one or more predictor variables and one outcome variable.
→ is used for **predictive analysis**

Our case:
Outcome variable = “Price”

Predictor variables = Factors
(from previous slide)

Data Basis for Regression
→ all observations with sales prices ranging from **645.000 $ to 1.100.000**
(that way, unusual expensive houses, which we can’t get access to will be excluded from the analysis)



OUTCOMES OF MODEL ESTIMATION

Factors **increasing** house sales prices

Every additional **square foot of total living area** of the house increases the sales price on average by **44 $**

With **every year going backwards from now** the sales price increases on average by **619 $**

Every additional **bathroom** in the house increases the sales price on average by **12.317 $**

If the **house got renovated** the sales price increases on average by **55.240 $**

If the **house got a good quality grade** the sales price increases on average by **33.747 $**

If the **house has got a waterfront view** the sales price increases on average by **47.734 $**

EVALUATION OF MODEL

Checking for **statistical significance**
"Does my model make sense?"

The **overall Model** is **statistically significant**

The **model predictions deviate on average by 12%** in both directions from the average sales price
→ in value terms approx. **by +/- 100.000 $**

All Factors are statistically significant, i.e. **they contribute to predict the average sales price for houses**

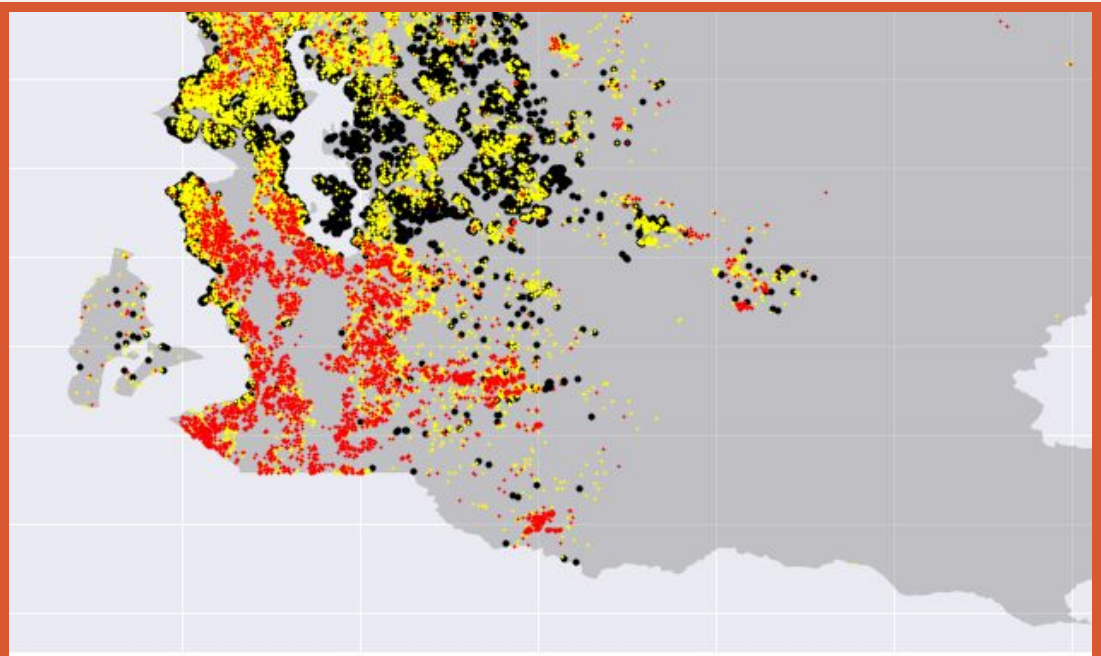
GEO-MAPPING

Visualization, where houses in King County have been sold to **which price**

→ lots of houses sold in high price segment (black) are at the waterfront

→ certain areas with high average sales prices (all zipcodes in appendix[2])

→ opportunity at Vashon Island



Legend:

Black :	high price segment (645.000 < price < 1.100.000 $)
Yellow:	medium price segment (322.000 $ < price < 645.000 $)
Red:	low price segment (price < 322.000 $)

IMPLICATIONS

To sell high
(derived from regression)

Houses in your portfolio should be in a good shape and/or renovated, which also lead to a good quality grade

To sell high
(derived from regression and geo-map)

Houses in your portfolio should have a waterfront view

New opportunity
(derived from geo-map)

Be the first and create a new exclusive neighbourhood on Vashon Island

FUTURE WORK

Variables weren't considered

Have a deeper look into the two "15"- variables

Detailed feature engineering

Categories were made to dichtotome variables. Perhaps create several dummies from these variables

Vashon Island

Further investigation, if this could be an actual opportunity, or if neighbourhood isn't that promising

THANK YOU!

Appendix

1: https://info.kingcounty.gov/assessor/esales/Glossary.aspx?type=r#g → under “GRADE”

2: