King County House Sales

First Project EDA by Tjade Appel

Agenda

Business Case 1
Best price / sq ft

Business Case 2
Value of King County grade
in relation to price

Business Case 3
Rich, old golf player wants
house near downtown

Linear Regression Model

Methodology

Business Case 1

Best price / sq ft

The overall median price / sq ft in King County is 245$

Zip code	City	Median price / sq ft
98023	Federal Way	145$
98032	Kent	146$
98001	Federal Way	147$

Business Case 2

Value of King County grade
in relation to price

- **Median price: 450.000$**

- **Quadratic correlation**

- **High linear correlation with other indicators of a valuable property**
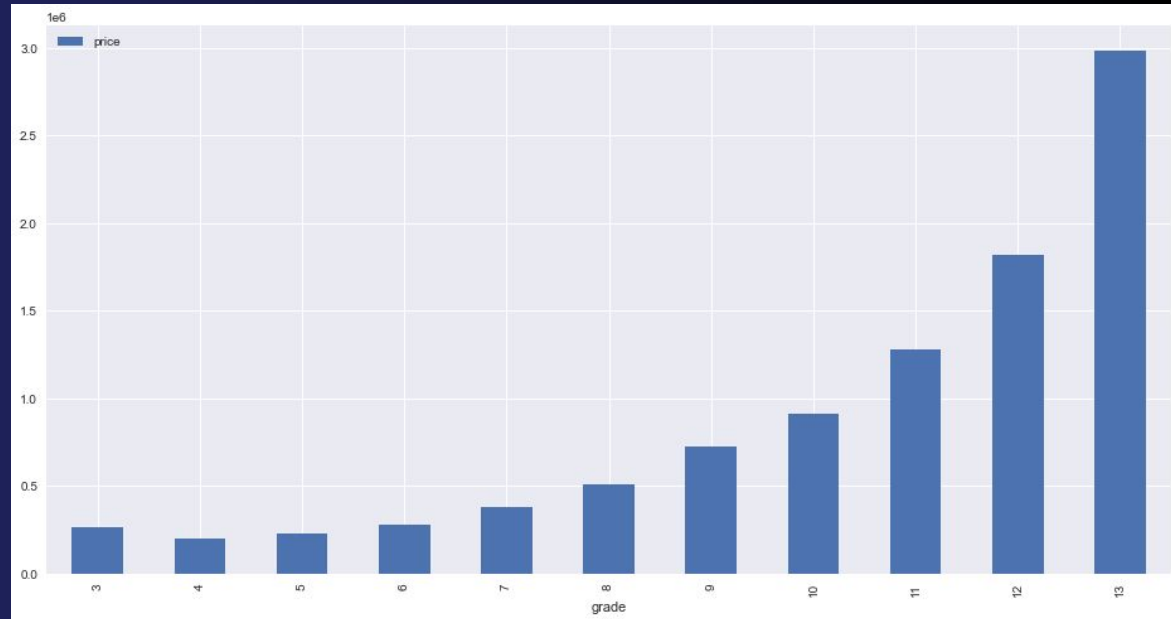
Business Case 2

Value of King County grade
in relation to price

- **Median price: 450.000$**

- **Quadratic correlation**

- **High linear correlation with other indicators of a valuable property**

Correlation coefficient of grade with feature	Feature
0.76	Living space
0.71	Living space (15 nearest neighbours)
0.66	Bathrooms
-0.45	Years since last modernisation

Business Case 3

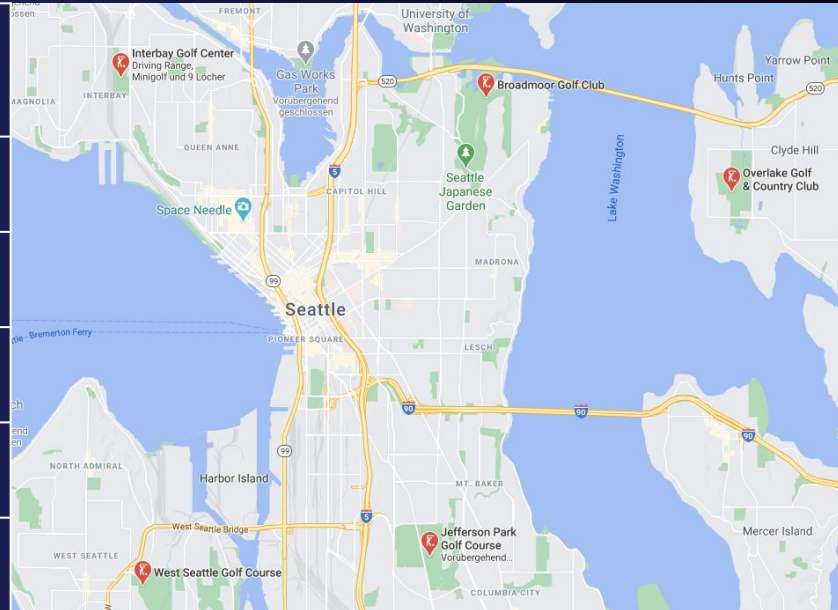Rich, old golf player wants
house near downtown

Possible zip codes
98126
98108
98039
98112
98119

Business Case 3

Rich, old golf player wants
house near downtown

Possible zip codes	Median			
	Price / sq ft	Years since last modernisation	Sq ft basement	No. of floors
98126	565$	28.5	0	2
98108	208$	61.5	305	1
98039	425$	83.5	442.5	2
98112	417$	76	285	2
98119	274$	69	0	1

Linear Regression Model

Methodology

First model included every single feature → R^2-value = 0.844

Not useful for house buyers, because most features are not known exactly

Not useful, because many of the features are highly correlated with each other

Boiled down model only includes the living space in feet, the zip code and the number of years since the last modernisation → R^2-value = 0.731

Linear Regression Model

Methodology

Boiled down model trained on 99.5% of data and tested on 0.5%
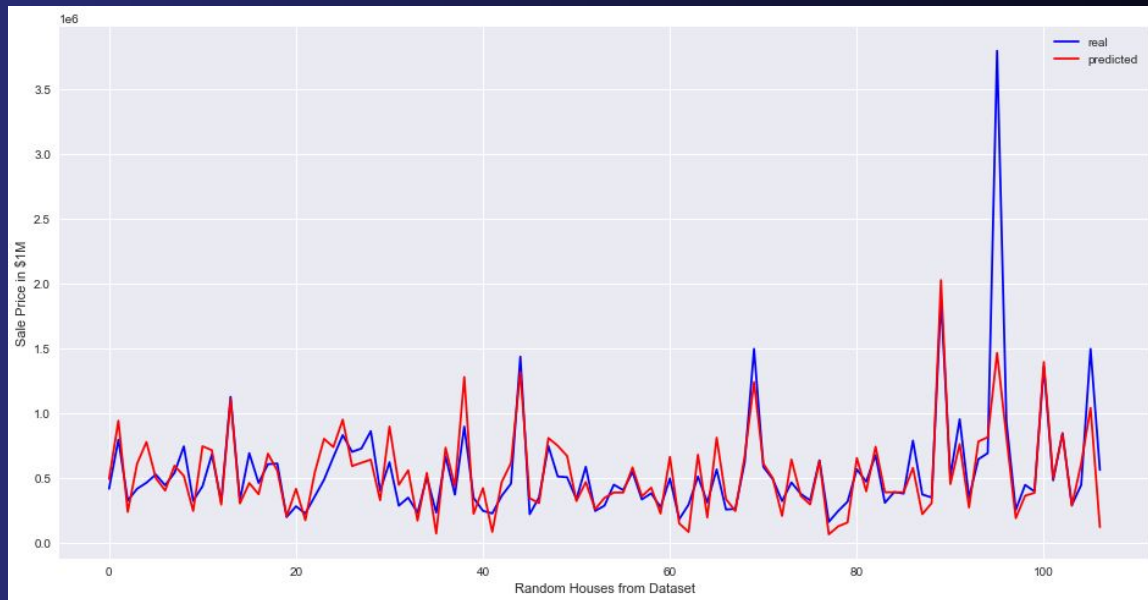
Thanks

Do you have any questions?