



EDA Project - King County House Prices

Exploratory Data Analysis

Ricarda Albers



Initial situation

Client: Erin Robinson

Business expectation:

- invest in economically weak neighborhoods
- buy and sell
- cost back and small profit
- socially responsible



Goal of this EDA

The purpose of this exploratory data analysis is to enable socially responsible, small-margin real estate transactions (purchases and sales) in economically weak urban neighbourhoods.

- To achieve this, economically weak districts within King County are identified.
- Selection criteria for buying and selling houses are named, which contribute to the achievement of low profit as well as to social responsibility.



Hypothesis

The starting point of the present EDA are the following hypotheses:

1. There are economically weak districts in King County.
2. Selling date has an impact on houses price.
3. Houses with more bedrooms are more costly.



First Hypothesis

- There are economically weak districts in King County



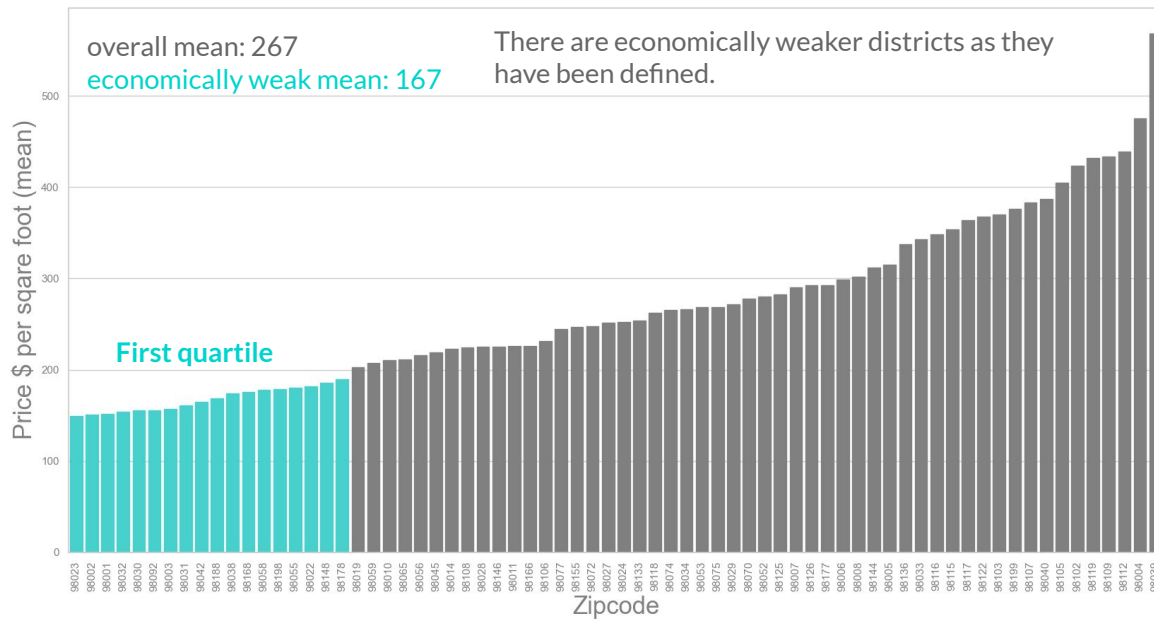
Economically weak districts

In order to define economically weak districts, the average (mean) price per square feet living has been used. It has been calculated for each zip code.

The lower quartile of this ratio sets the searched range.

Those are the neighborhoods, where it is recommended to our client to buy houses.

Economically weak districts

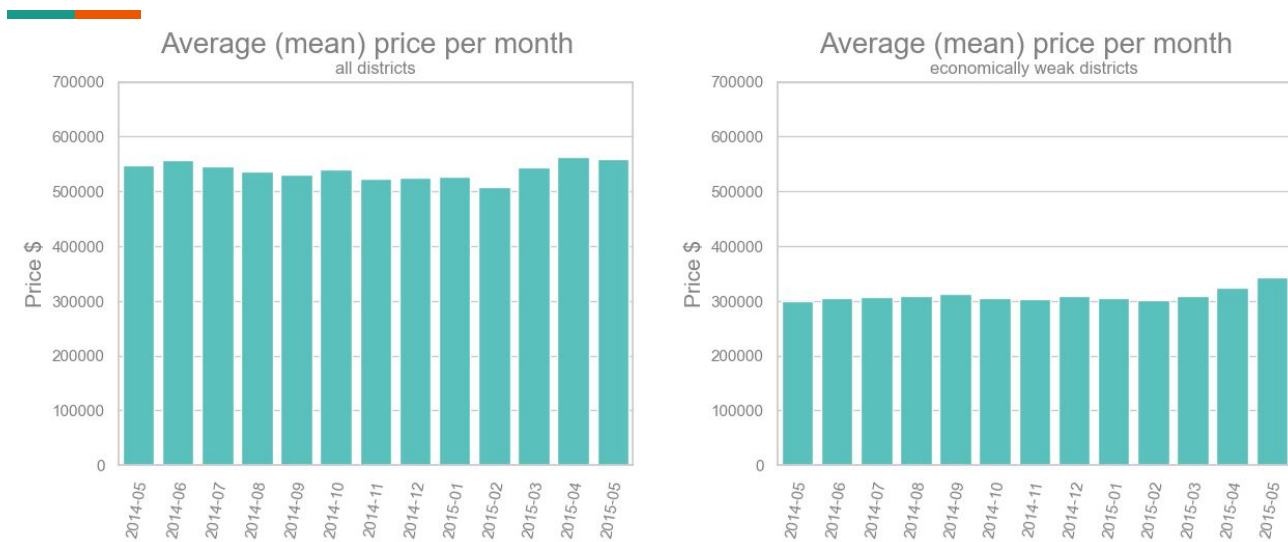


Zip codes:

98002
98003
98022
98023
98030
98031
98032
98038
98042
98055
98058
98092
98148
98168
98178
98188
98198

Second Hypothesis

- Selling date has an impact on houses price

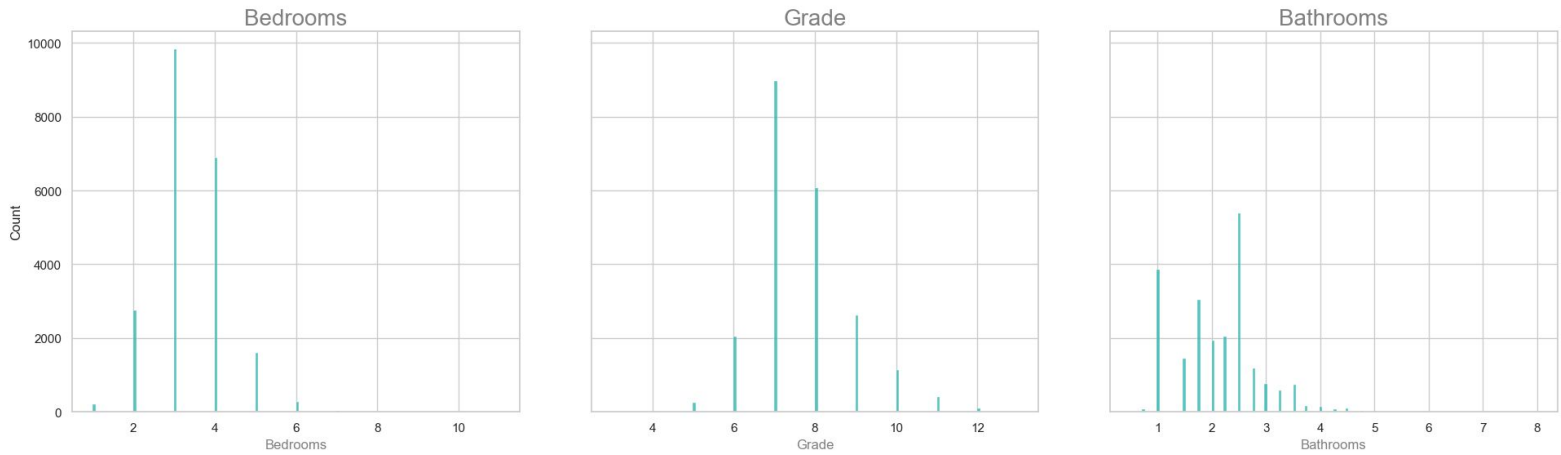


There are no significant differences between months in terms of selling price. In the relevant - economically weak - districts, those differences are even weaker.

Third Hypothesis

- Houses with more bedrooms are more costly

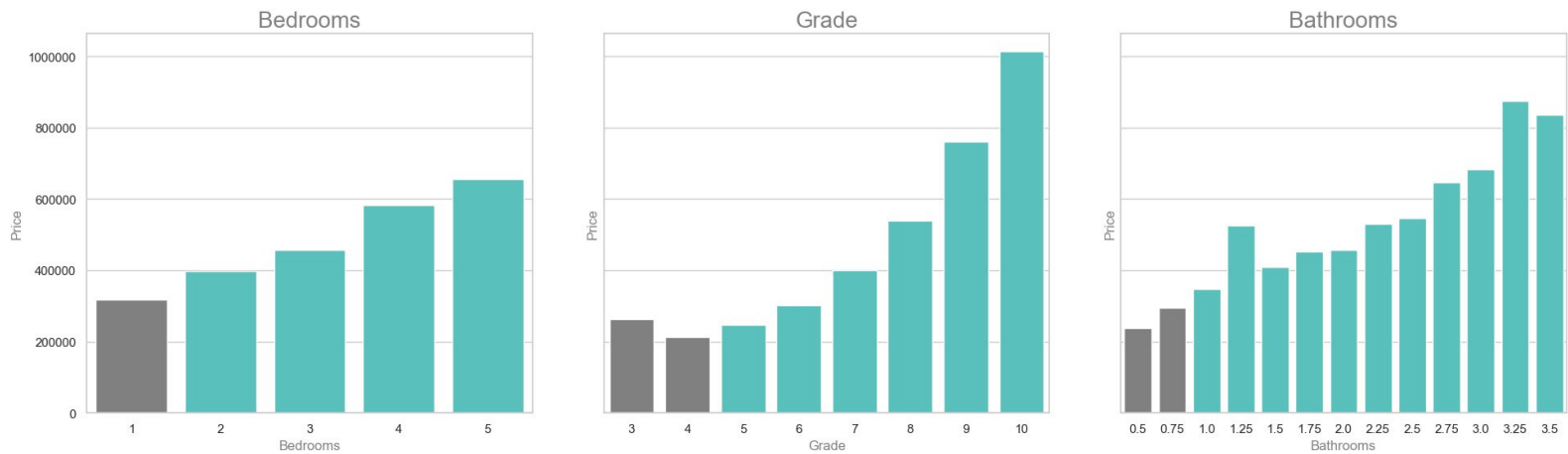
Frequency distribution - all districts



Frequency distribution of categorical variables with relevant impact on price.

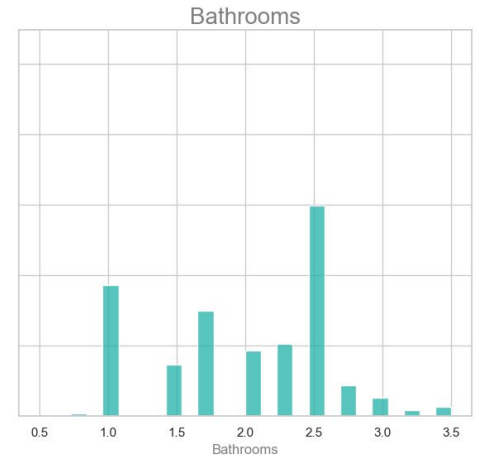
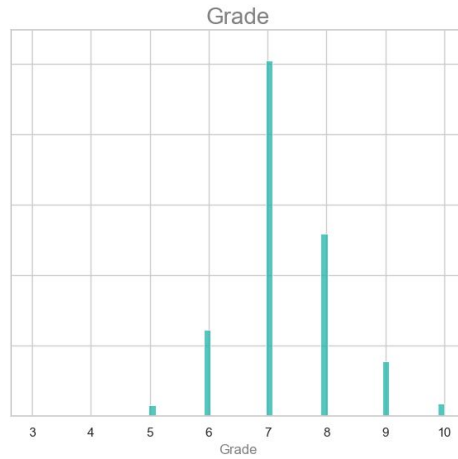
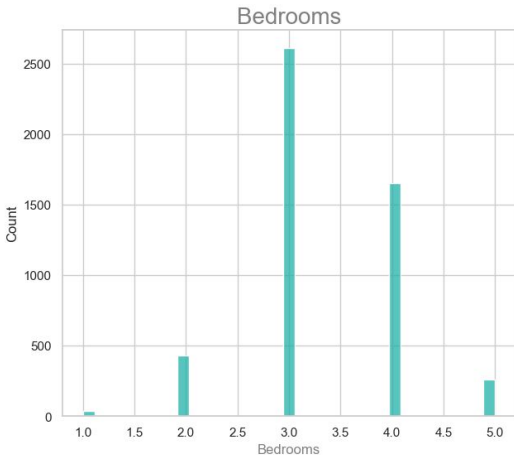
11

Correlation with price - all districts



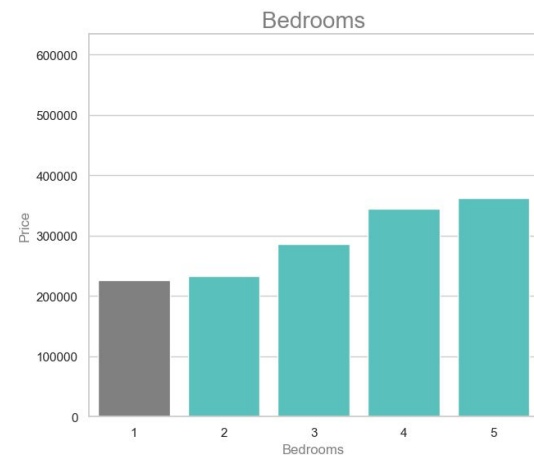
12

Frequency distribution - economically weak districts

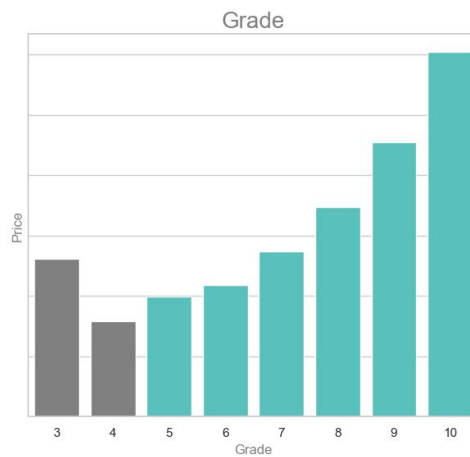


13

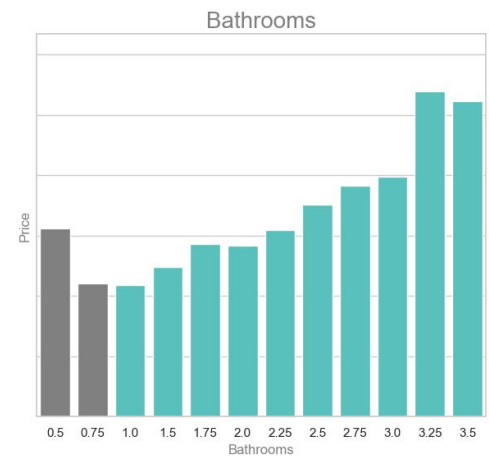
Correlation with price - economically weak districts



One bedroom more costs on average
48.612\$



One grade more costs on average
84.269\$



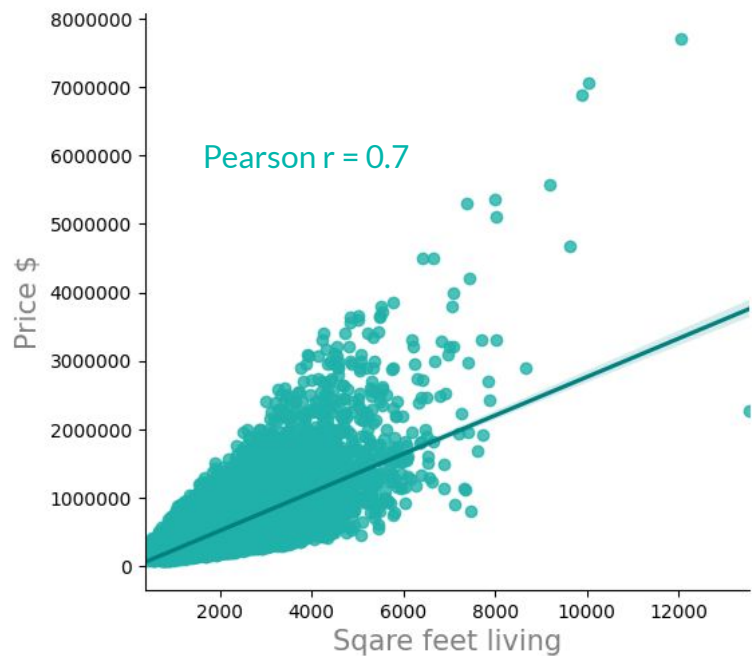
0.25 bathrooms more costs on
average 34.615\$ (1 -> 124.618\$)

14



There is the strongest positive correlation between price and living space in the present dataset.

Correlation price and living space



15



Broken down into economically weak and non-economically weak districts, the correlation of price and living space is slightly stronger for economically weak districts.

Correlation price and living space



16



Recommendations I

- Districts to buy and sell in:

98002, 98003, 98022, 98023, 98030, 98031, 98032, 98038, 98042, 98055, 98058, 98092, 98148, 98168, 98178, 98188, 98198 are economically weak districts.

- The date is not relevant for the real estate transactions to be carried out. There has no special timing to be prioritised.

17



Recommendations II

- Square foot living, grade and the number of bathrooms correlate more strongly with price than the number of bedrooms. It is recommended to buy houses and renovate them.

Especially adding an additional bathroom to houses with just one bathroom is recommended, as it can affect the grade, too. In addition, social responsibility is supported, since a higher number of bathrooms also increases the suitability for families.

18



Data

Data source: open data from King County government (<https://data.kingcounty.gov>)

Cleaning :
overall good data quality
2 columns removed (waterfront, view)
one outlier removed
NaNs within square foot basement have been calculated
Features: price per Square foot, economically weak (bool)