Explorative data analysis of housing market data in King County

Franziska Schulze Bockeloh

24.02.2023

Agenda

- Introduction
    - Dataset
    - Client
    - Hypotheses
- EDA to test hypotheses
- Summary and recommendations
- Outlook

Introduction - About the dataset

- Housing market in King County
- House details
- Sale details
- 21597 entries from 2014 to 2015

Introduction - General Data cleaning

- Dropping columns that are not required for further analysis
- Adjust data types
  - date of sale to dateformat
  - floats to integer
  - floats to categorical value (zipcode)
- Create new columns
  - extract year and month information from date of sale
  - defining neighborhood score: sqft_living15 / sqft_lot15
- Check for empty values

Introduction - The Client

“I am looking for a house in King County!”

“I am a single woman, I will live alone. A house size between 700 and 2000 sqft will be fine.”

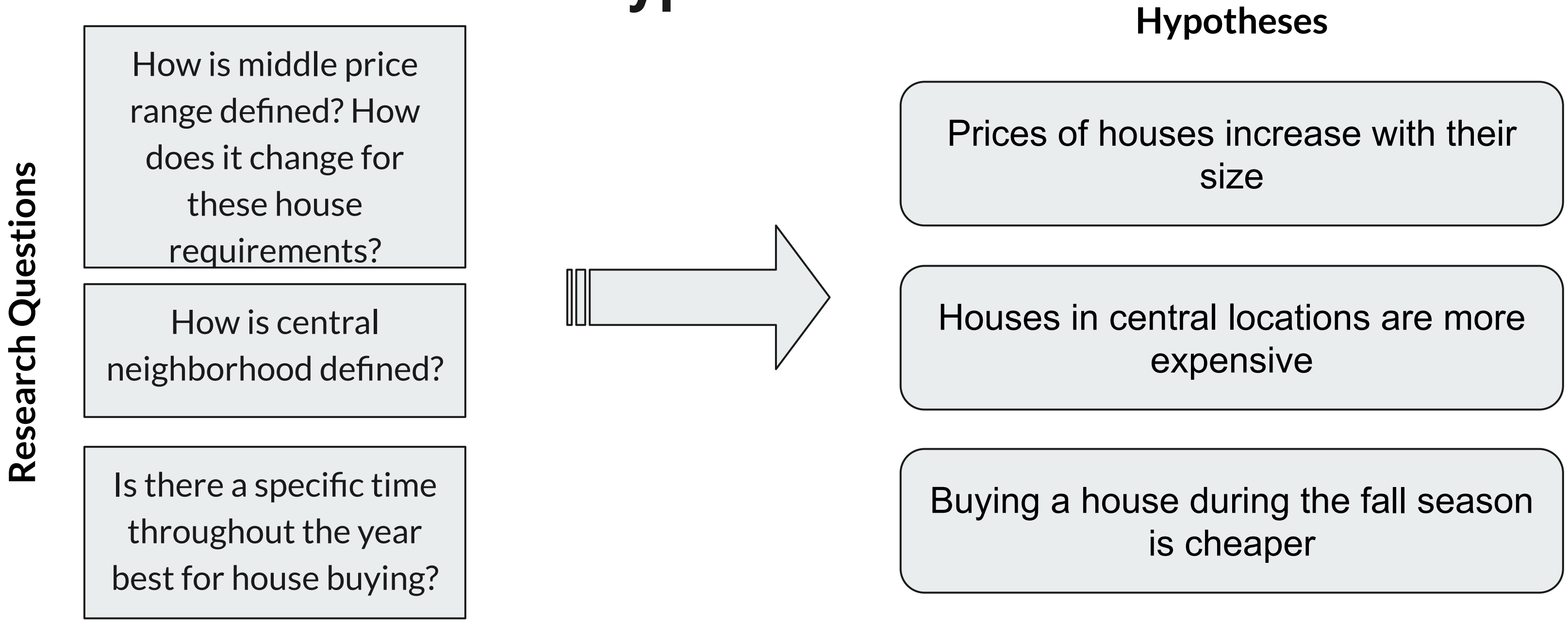
“It should be in a lively and central neighborhood.”

Nicole Johnson

“I want to spend a middle price range. With how much should I calculate?”

“I am flexible for the right timing. When would be the best time to buy?”

Introduction - The Hypotheses



Research Questions

How is middle price range defined? How does it change for these house requirements?

How is central neighborhood defined?

Is there a specific time throughout the year best for house buying?

Hypotheses

Prices of houses increase with their size

Houses in central locations are more expensive

Buying a house during the fall season is cheaper

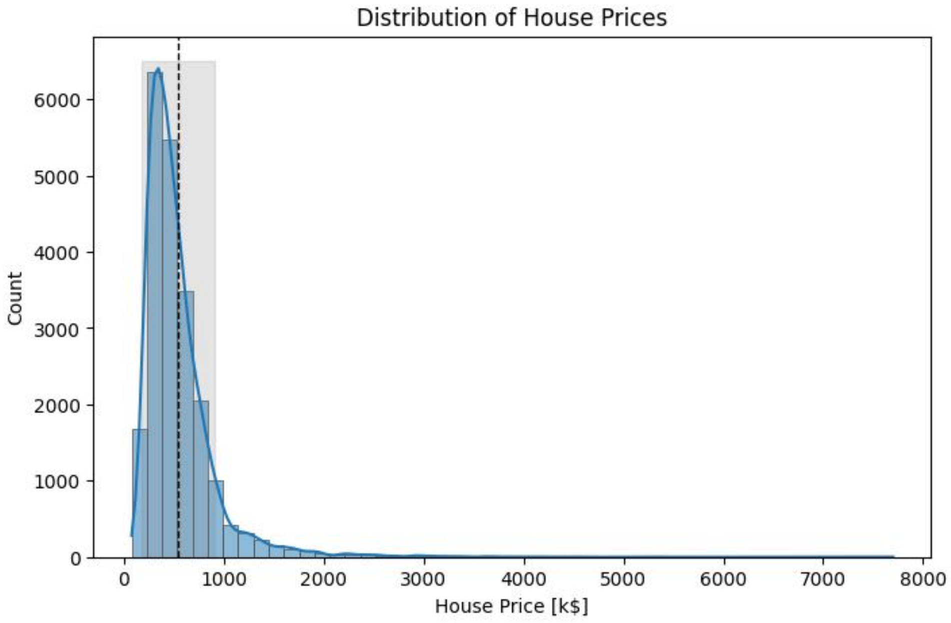
Hypothesis 1 - Methodology

Prices of houses increase with their size

- Price variable and house size variable
- Define price range for all houses
- Filter houses for specific house size and define price range
- Compare price ranges and calculate correlation coefficient of house price and house size

Price range - all houses

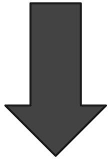
- 21597 houses were sold in the years 2014 and 2015
- Right skewed distribution
- Price range between 78 k$ and 7700 k$
- Mean price of 540.16 k$

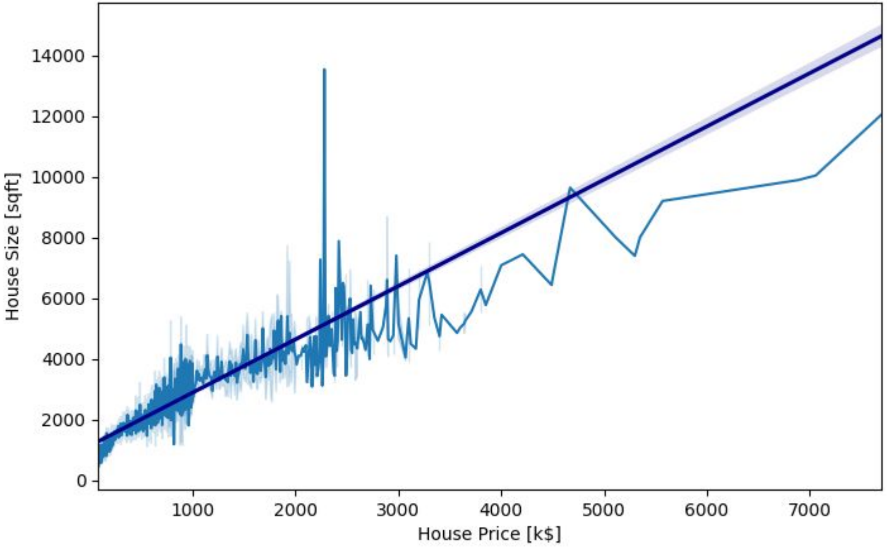


Filter for house size

- Size between 700 and 2000 sqft
- 11429 houses (53%)
- Mean price of 387.19 k$
- Middle price range between 232 -542.6 k$

-46,41%

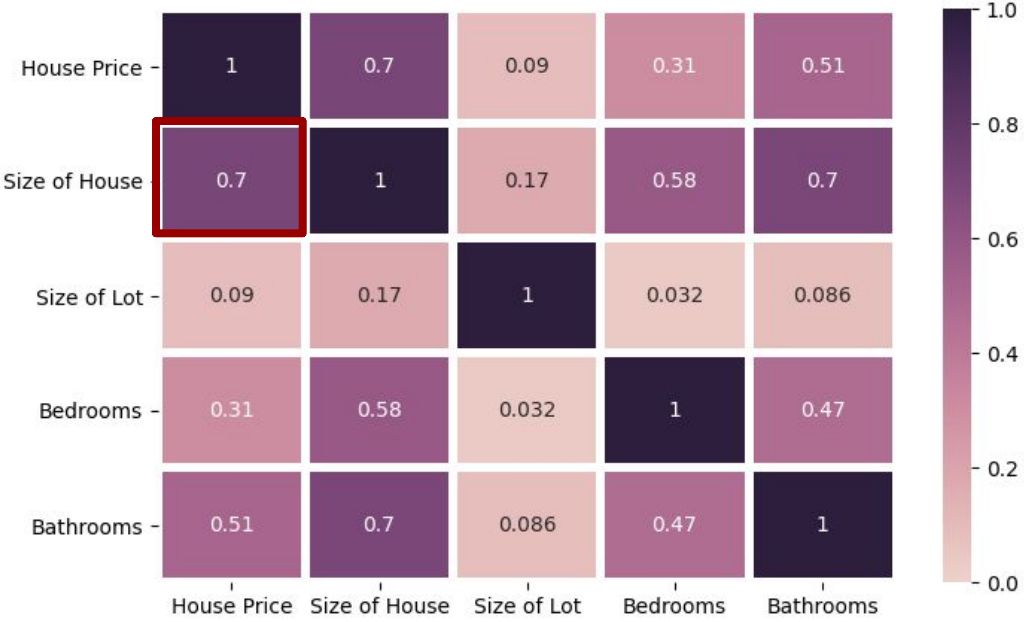




Hypothesis 1

Prices of houses increase with their size

- Correlation between house size and price
- Further parameters having impact on the price



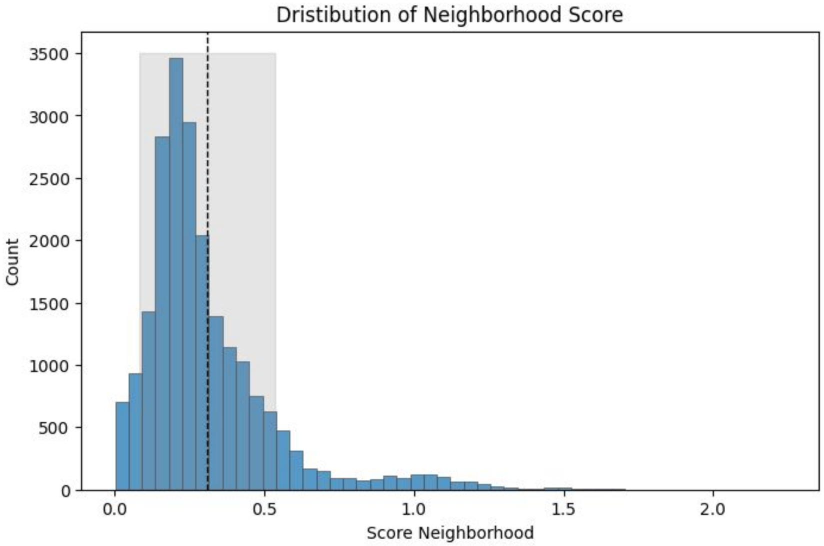
	House Price	Size of House	Size of Lot	Bedrooms	Bathrooms
House Price	1	0.7	0.09	0.31	0.51
Size of House	0.7	1	0.17	0.58	0.7
Size of Lot	0.09	0.17	1	0.032	0.086
Bedrooms	0.31	0.58	0.032	1	0.47
Bathrooms	0.51	0.7	0.086	0.47	1

Hypothesis 2 - Methodology

Houses in central locations are more expensive

Score Neighborhood
=
Living space of neighborhood / Lot space of neighborhood

- High score defines central neighborhood
- Calculate mean value of neighborhood rating for zip codes
- Score above 0.3 considered as central neighborhood
- Define correlation between neighborhood and price



Price range for houses

- 25 zip codes were defined as central
- 5065 houses in central neighborhoods (23.5%)
- Mean price of 454.71 k$
- Middle price range between 305 and 604 k$

All Houses

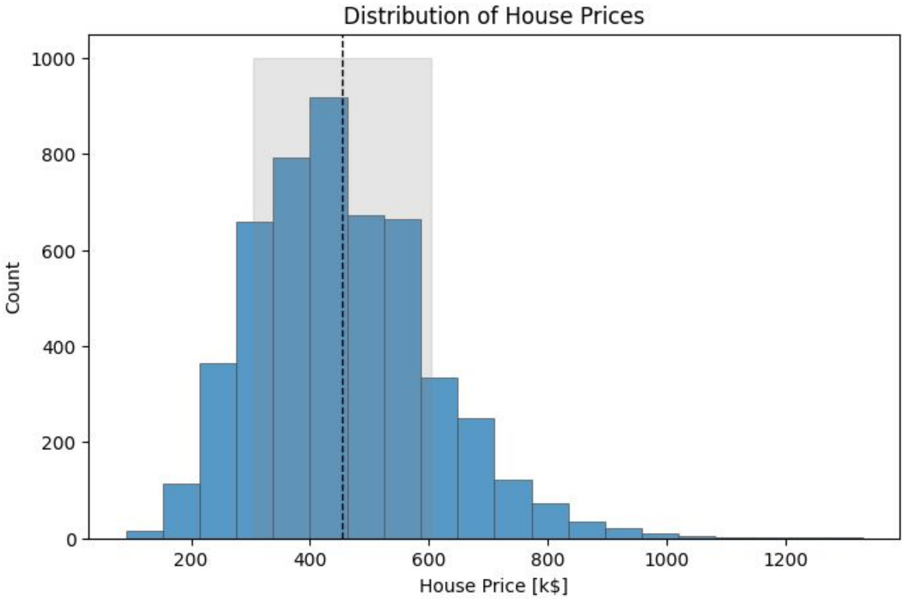
↑

+16.1%

Size Filtered

↑

+36.37%



Hypothesis 2

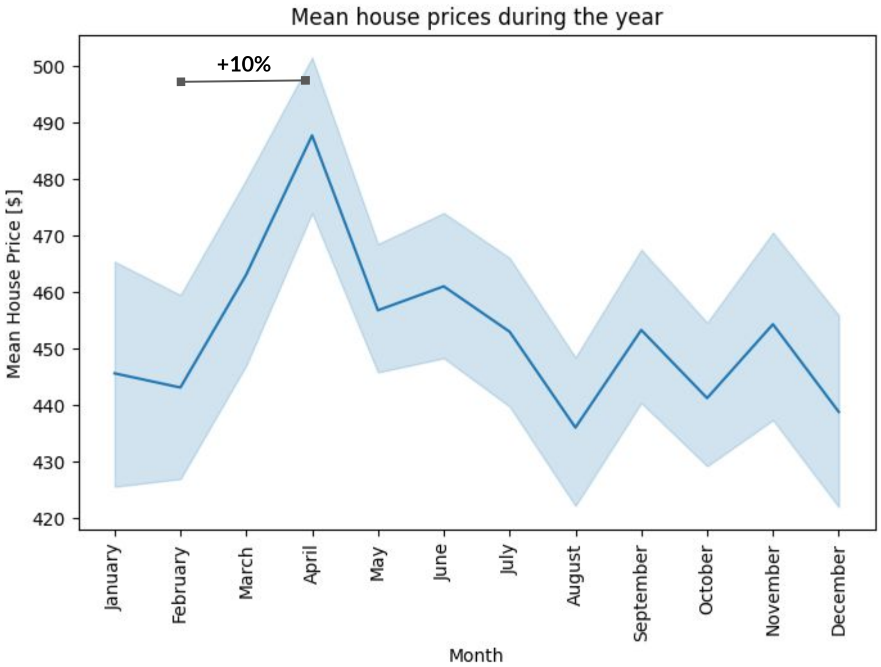
Houses in central locations are more expensive

- Only slight correlation between neighborhood score and house price (0.48)
- Other parameters might influence price change more dramatically

Hypothesis 3

Buying a house during the fall season is cheaper

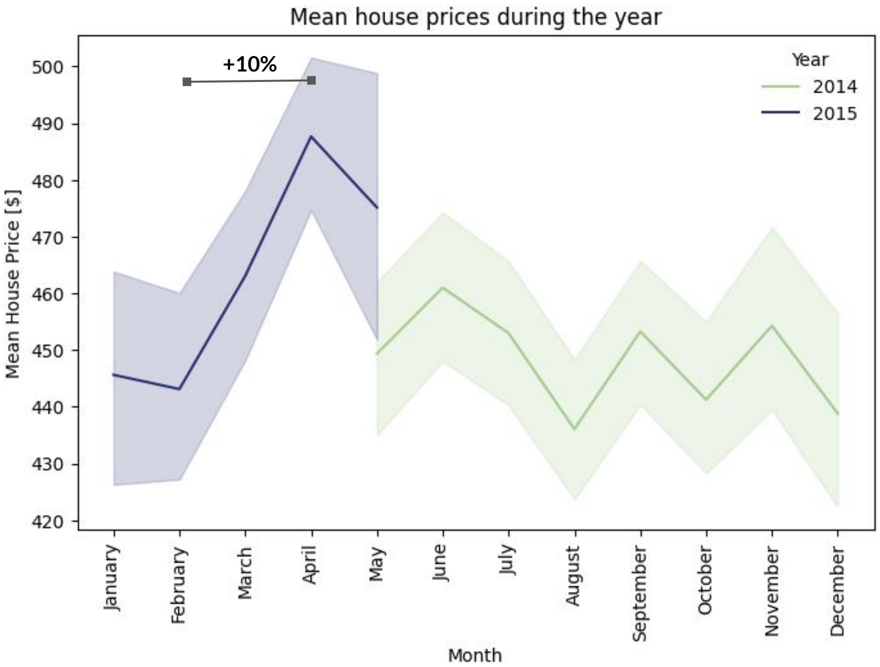
- Fluctuation of prices throughout the year
- Lowest prices in February, August and October
- Price increase during spring



Hypothesis 3

Buying a house during the fall season is cheaper

- Fluctuation of prices throughout the year
- Lowest prices in February, August and October
- Price increase during spring
- More data required for further analysis

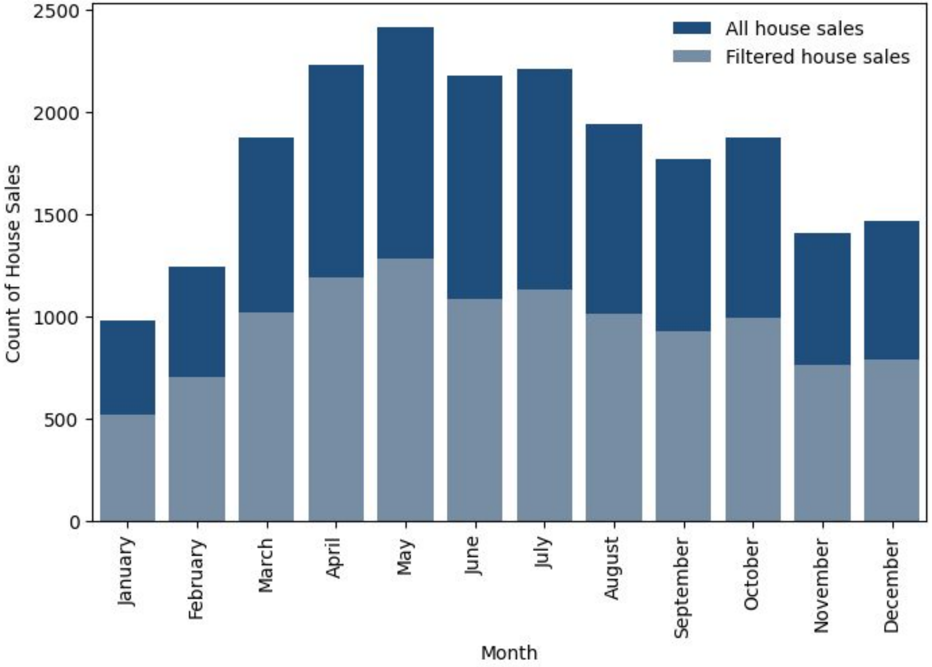


Hypothesis 3

Buying a house during the fall season is cheaper

Availability of houses

- less houses available during winter



Summary

Prices of houses increase with their size

Houses in central locations are more expensive

Buying a house during the fall season is cheaper

Summary - Recommendations for Client

- Search for smaller houses, they are cheaper
- No clear influence of neighborhood centrality on price
    - Check other parameters
- Calculate with a price range between 305 - 604 k$
- Prices might increase during spring, wait for fall season to buy a house

Outlook

- Include more parameters like grading, condition and proximity to shops or infrastructure
- Map central zip codes on geographical plot
- Collect more data for analysis of season influence

Thank you for your attention

Questions?