EDA King County Real Estate

Felix Behrendt

Outline

1) Dataset
2) Stakeholder
3) Quality of the data
4) Insights:
    a) Real Estate Industry Activity
    b) Central Neighbourhood
    c) Relationship of Centrality and Price on condition
5) Recommendations
6) Outlook

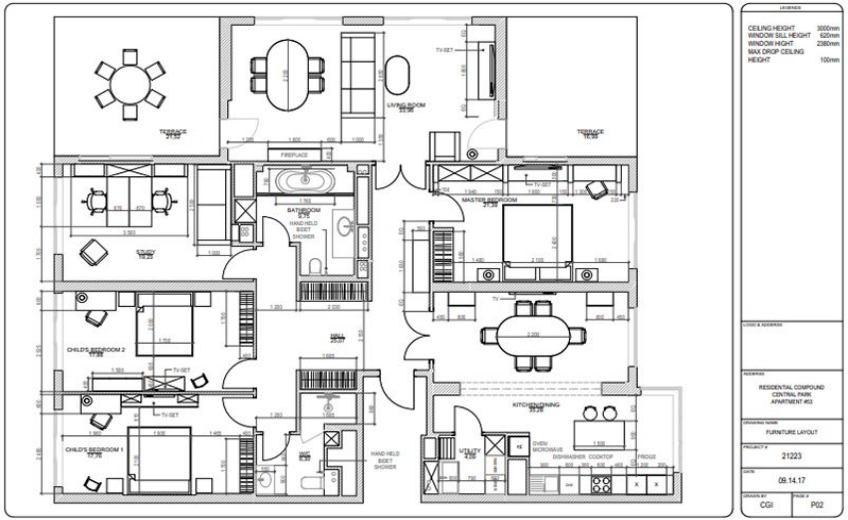
Dataset

Object sales in the real estate industry in King County (USA).

The dataset contains 21597 entries and 20 features (excluding the id).

relevant features:

- date
- price
- sqft_lot15 (land lots of the nearest 15 neighbors in square footage)
- condition of the object
- coordination (lat, long)

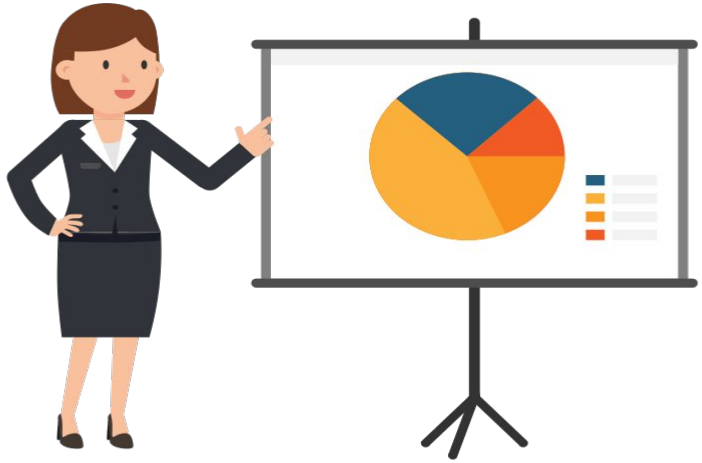


Source URL

Stakeholder

The Nicole Johnson is a Buyer, who looks for an object in the real estate industry of King County (USA).

Characteristics of the object:
- lively and central neighborhood
- middle price range
- right timing (within a year)



Source URL

Quality of the data (for my analyses)

missing data in features

- view to a waterfront
- number of views
- renovation year

Source URL

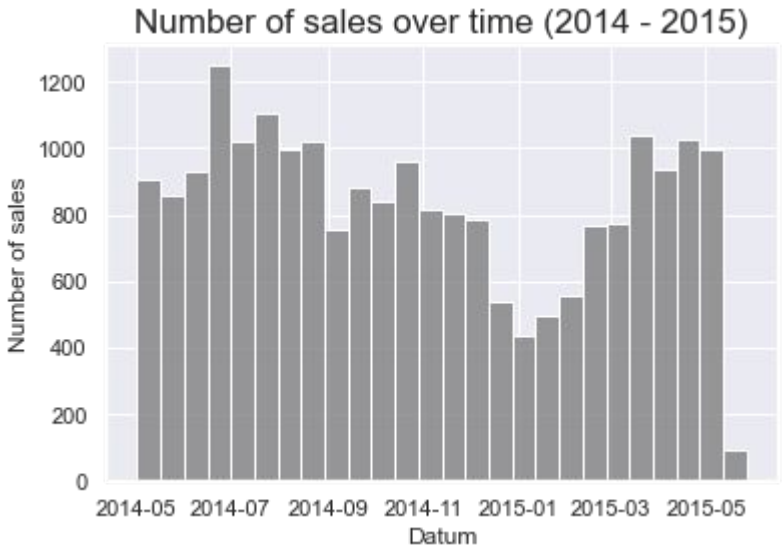
Quality of the data (for my analyses)

missing data in columns

- view to a waterfront
- number of views
- renovation year

Date:

- unusually few observations in May 2015



Insight 1: Real Estate Industry Activity

For the stakeholder, the timing of the purchase is relevant factor. Under the 'right timing' I assume it refers to the activity of the real estate market. This raises the following questions for the data set:

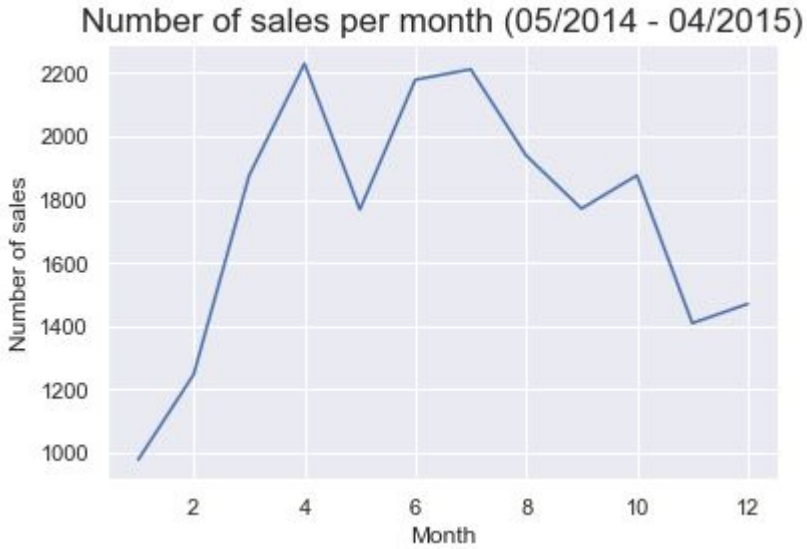
How does the time affect the activity of the real estate market?

Is there a variation in activity of the real estate market?

If so, which months show a high activity ?

Insight 1: Results

Number of sales are higher during the summer season than in the winter season.



Insight 1: Conclusion

Is there a variation in activity of the real estate market?

There is variation in activity of the real estate market. More properties are sold in the summer months than in the winter months. From this I assume that the number of properties on offer is also higher.

Is there a variation in activity of the real estate market?

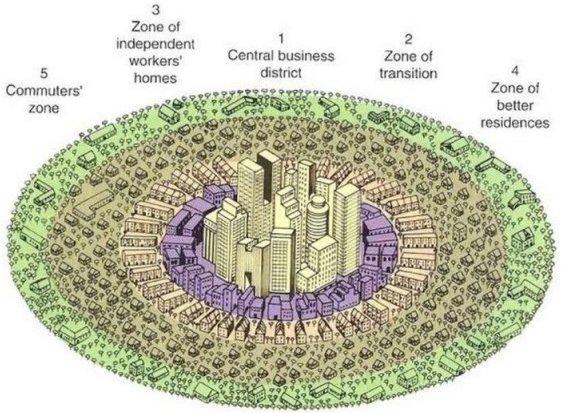
The months with the highest activity are:

- April
- June
- July

Insight 2: Central Neigborhood

The stakeholder wishes for an object in a central neighborhood. In general the real estate objects close to the center are smaller (area) and more expensive. Therefore I ask the following question:

How does the centrality of the objects affect the price and the size of neighboring objects?



The closer an object is to the city center, the smaller the size of neighboring objects.

The closer an object is to the city center, the higher the price.

Source URL

Insight 2: Challenge of the price distribution

This skewed distribution causes visual problems when displaying the price. In addition, the stakeholder calls for proposals in the middle price range.

Therefore it is advisable to group the costs into intervals using quantiles (Q25, Q50, Q75).

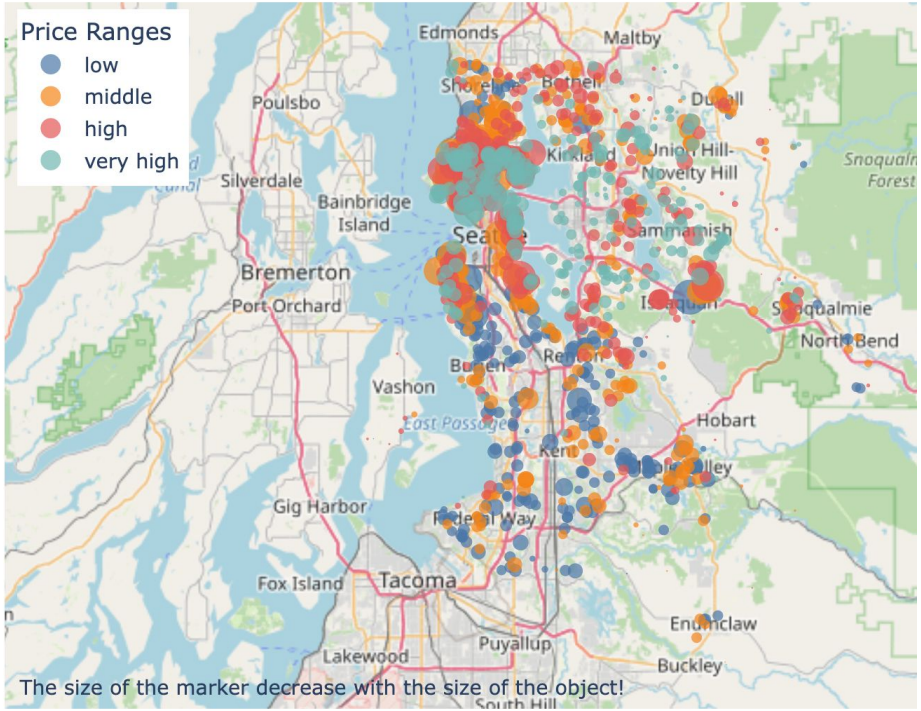
Price Interval	Price Range in $
low prices	(77999.999, 322000.0]
middle prices	(322000.0, 450000.0]
high prices	(450000.0, 645000.0]
very high prices	(645000.0, 7700000.0]

Insight 2: Results

This means the size of the neighboring objects are smaller in the city centre of Seattle

Expensive and very expensive objects are dominant in the urban core of Seattle.

In addition to the city centers, there are other clusters of very expensive properties in Bellevue and Mercer Island.



Insight 2: Conclusion

The closer an object is to the city center, the smaller the size of neighboring objects.

The closer an object is to the city center, the higher the price.

There is no clear relationship between distance to the city center and property size or price.

But an object closer to a center tends to be of higher price and smaller size. Inside a city can be multiple centers, which varies in size of area. In smaller cities this rule does not apply.

Insight 3: Centrality / Price and Condition

The stakeholder wishes for an object in a central neighborhood. In general the real estate objects close to the center are smaller (area) and more expensive. Therefore I ask the following question:

How does the centrality and the price range affect the condition of the objects?

Objects with a high price or very high price and high centrality are in better conditions.	—--

Insight 3: Defining Centrality

Based on some visual sampling and an educated guess I define the centrality class on these intervals.

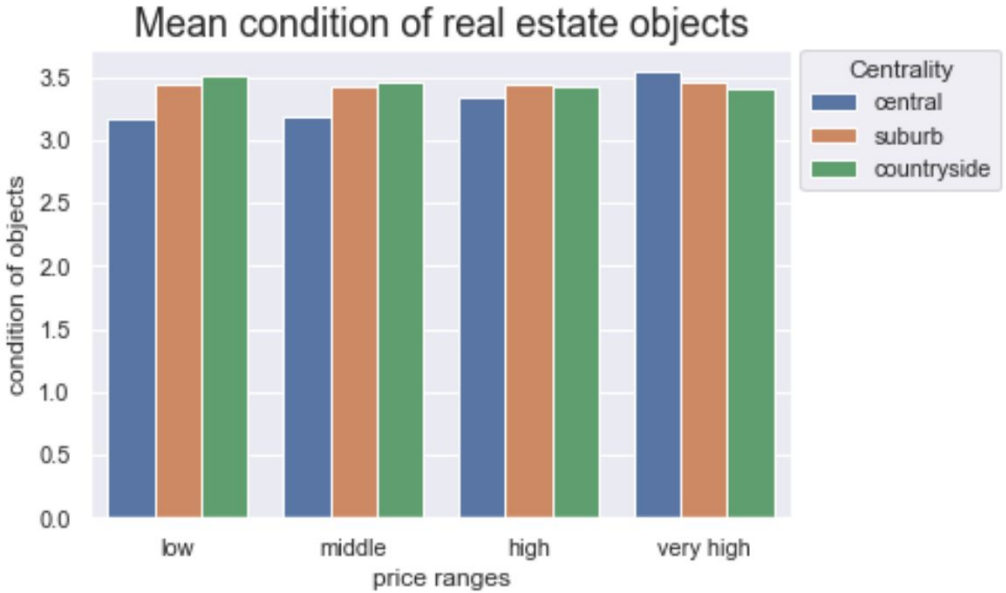
Intervals for centrality of neighborhood:

for centrality of neighborhood	land lots in square footage of the 15 neighbors
central	0 - 5000
suburbs	5000 - 12000
countryside	> 12000

Insight 3: Results

The average condition of the objects does not vary strongly depending on the centrality or the price ranges.

The most stable average conditions are in the suburb area.



Insight 3: Conclusion

Objects with a high price or very high price and high centrality are in better conditions.

—-

Objects with a high price or very high price and high centrality are not in better conditions than others.

Neither does the centrality or the price range provides valuable information about the condition of an object.

Recommendations for Nicole

- **Nicole should look for an object from March to September.**
- **Nicole should look for an object with a zipcode between 98103 to 98199.**

- **Nicole should look for the objects with the id:**

    - 9297800165
    - 6145600865
    - 1250201194
    - 3524049083
    - 2734100065

Outlook

Improving spatial analyses:

1) Using clustering algorithm to determine centers of a city
2) Define centrality on basis of distance to the closest cluster centre

Thanks for your attention!