House Price Analysis in King County

Paul Kotyrba
September 22, 2023

Agenda

- Client
- Methodology
- Data Insights
- Finding houses
- Larry's house
- Final Thoughts

Client:
Larry Sanders

Requirements

Larry

Buyer

Limited Budget

Requirements

Larry

Buyer

Limited Budget

House

Waterfront access

Needs place for kids

Requirements

Larry

Buyer

Limited Budget

House

Waterfront access

Needs place for kids

Area

Nice & isolated, but central

Neighborhood without many children

Methodology #1

Data Collection

Gather relevant data on houses in King County.

Clean data from irrelevant and false data.

Data Overview

Explore data, find patterns, define hypotheses.

Data Insights

Hypotheses results, phrase the findings, raise relevant questions.

Methodology #2

Define approach   Match data insights with stakeholder requirements: Translate requirements into searching for the right house.

Filter data   Identify affordable houses that meet the requirements.

House selection   Provide recommendations.

Database: Houses in King County (US)

Features

Price

Size of house & lot

Rooms

House condition

…

Location

Zip codes

Altitude, longitude

Waterfront access

Neighbor house size

…

Additional

House viewed?

Year sold

Year built

Year renovated

…

Hypotheses

1 If houses have **waterfront** access, then the **prices** are **higher** .

2 The **larger** the house, the higher the **price** of the house.

3 The **more rooms** the house, the higher the **price** of the house.

4 The higher the **grading** of the house, the higher the **price** of the house.

5 The larger the lot of neighbors, the higher the price.

6 (The larger the house of neighbors, the less kids in an area.)
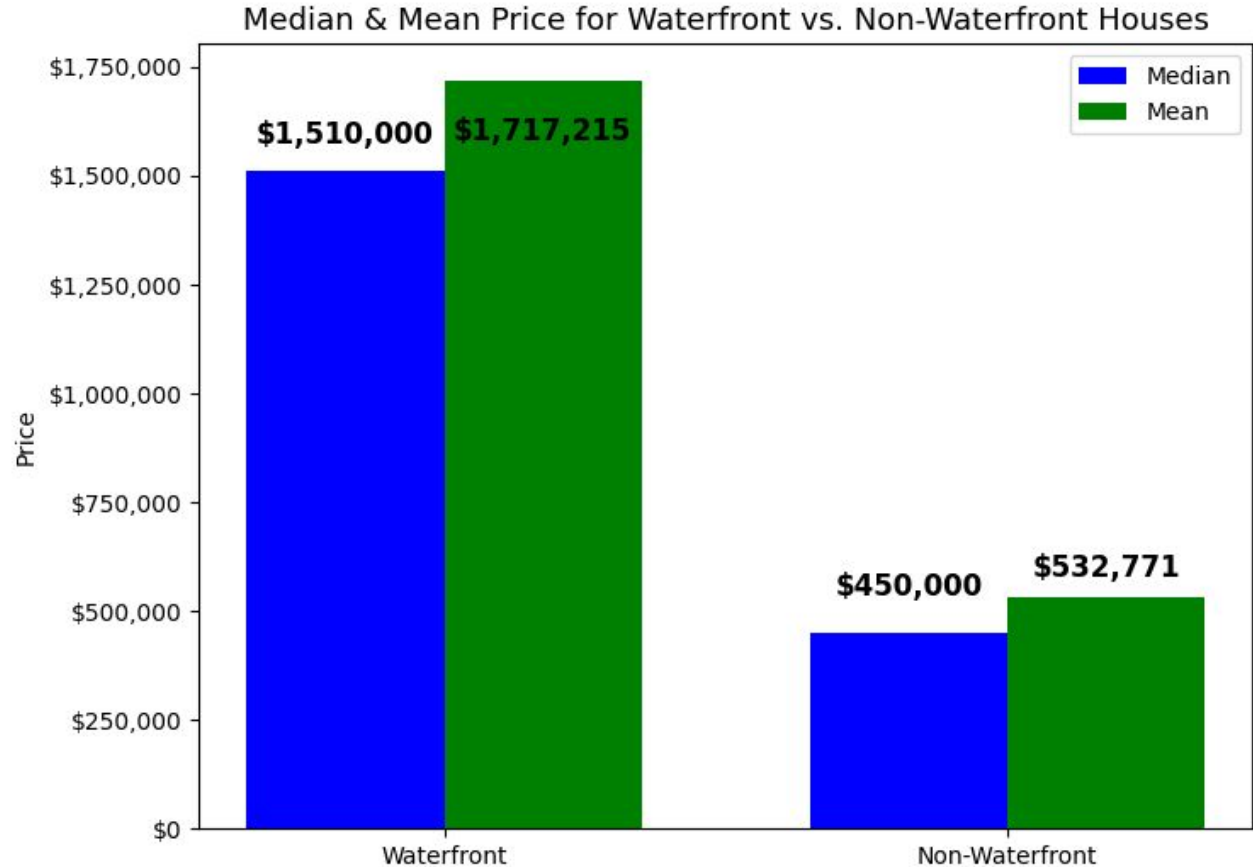
Data insights

146 houses have **waterfront access**.

That's **0.68%** of all houses in the database.
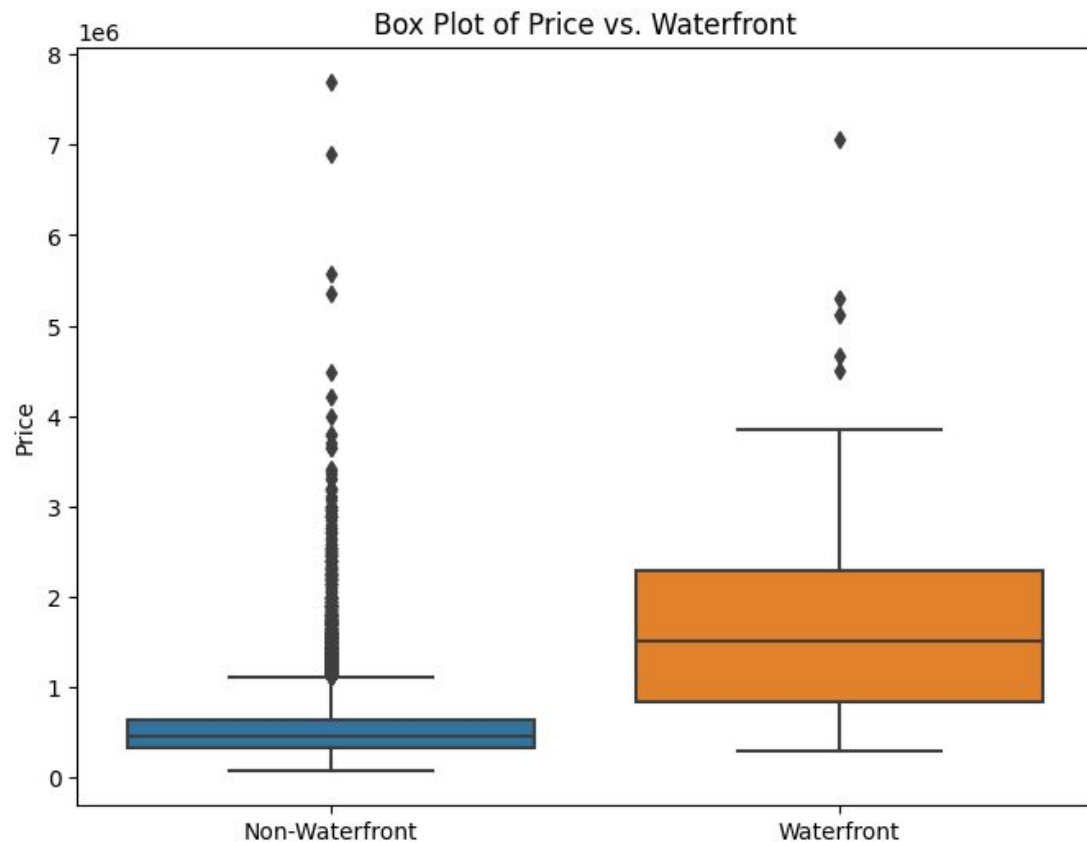
Data insights

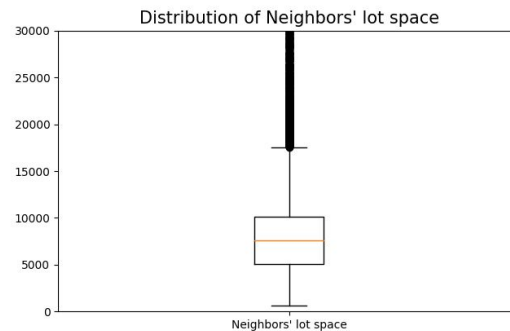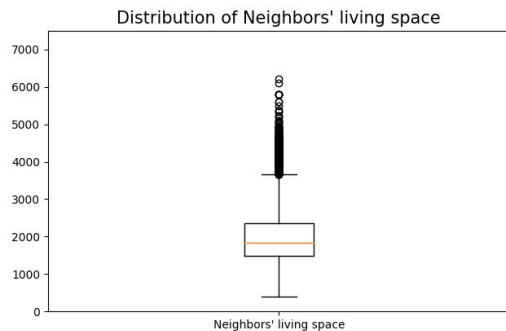Houses **with waterfront**

> 3x higher price on average.

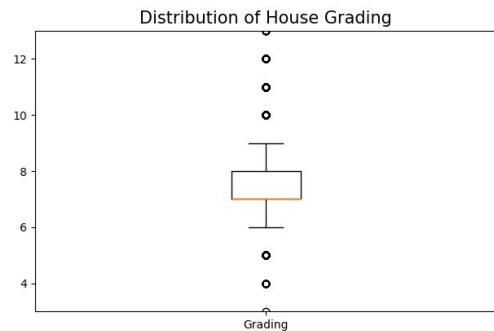Median & Mean Price for Waterfront vs. Non-Waterfront Houses

Data insights

Houses **with waterfront**

> 3x higher

price on average.

Box Plot of Price vs. Waterfront
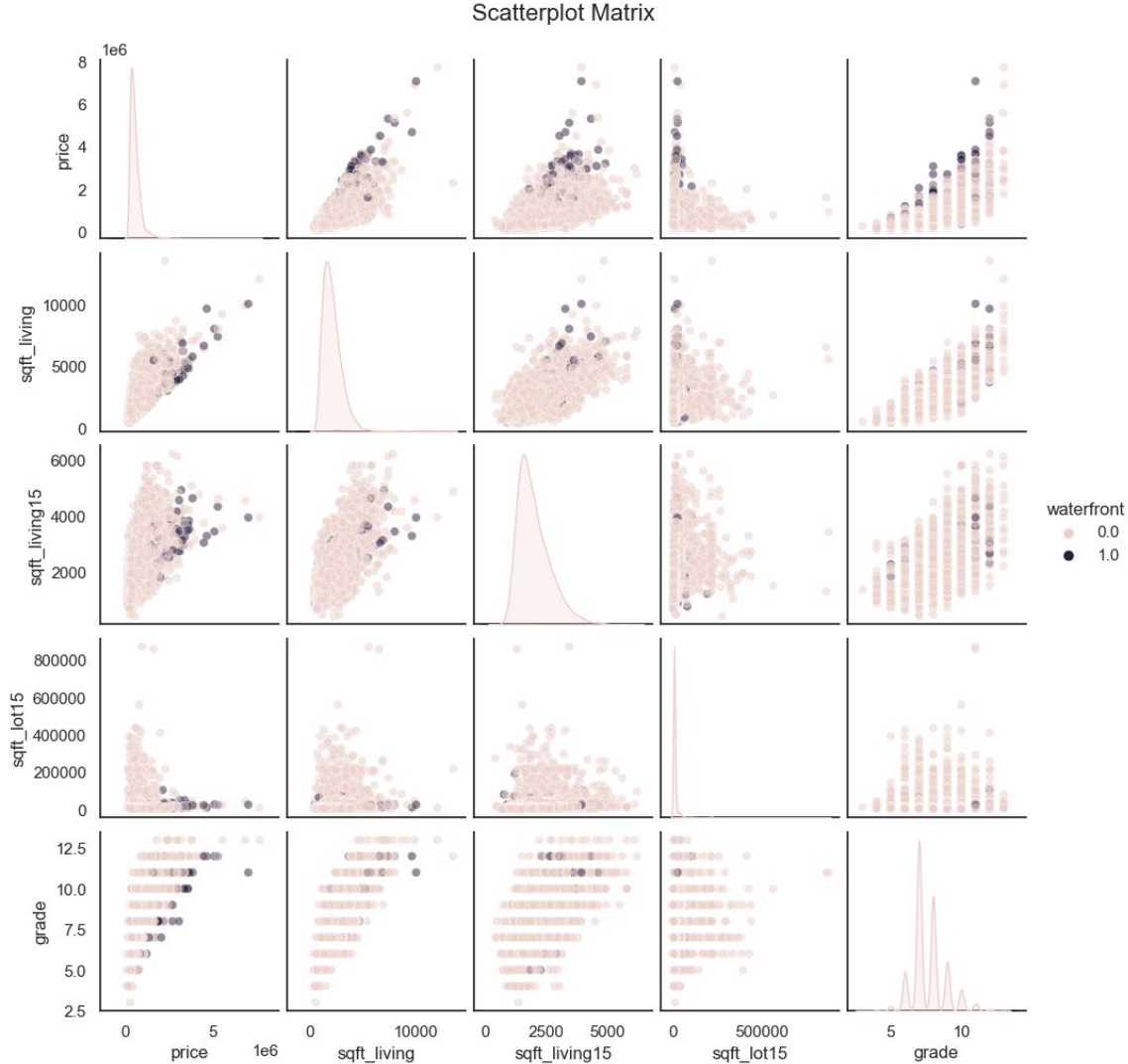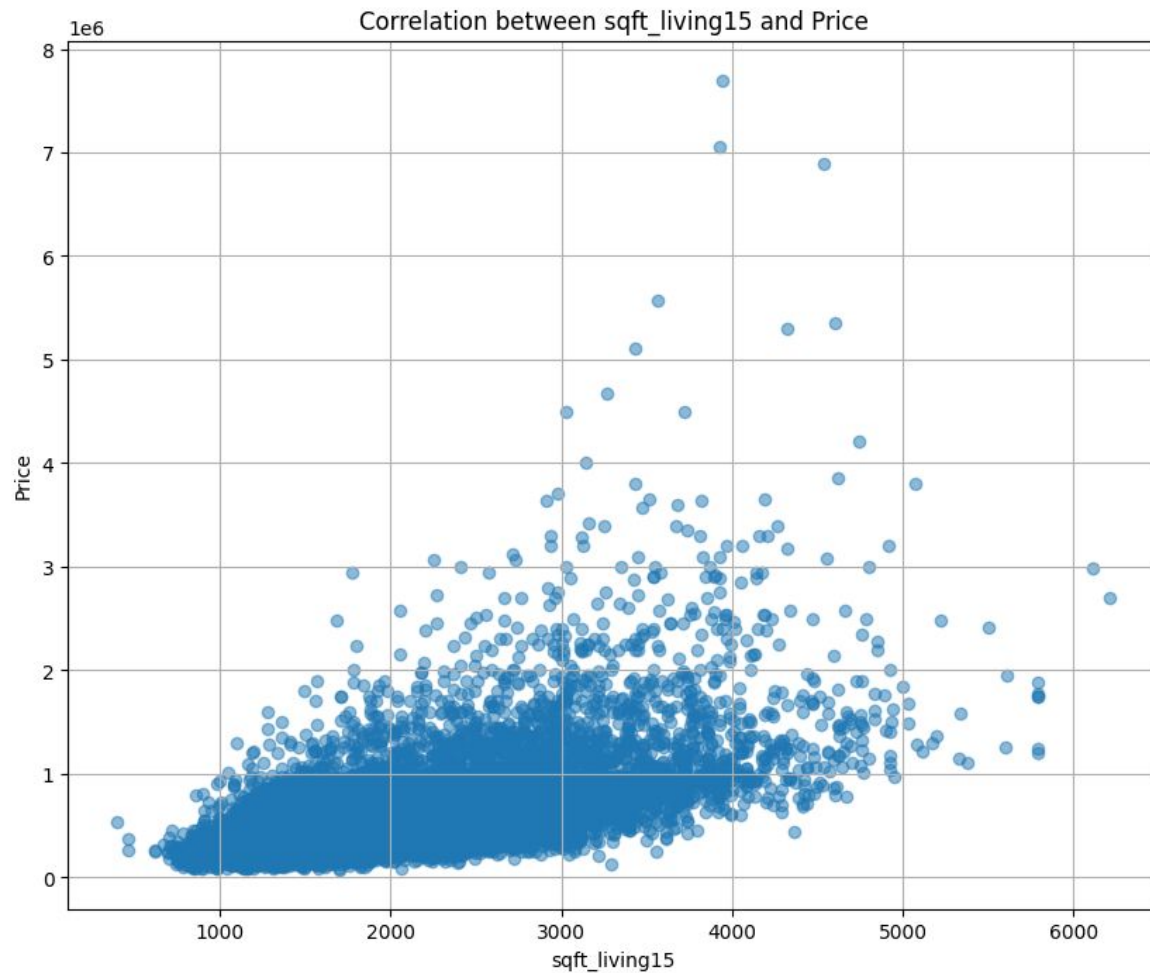
Data insights

Distribution of numeric columns
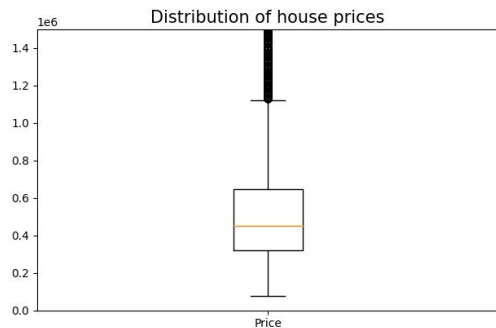
Data insights

Relationships between:

- house price
- living space
- neighborhood living space
- neighborhood lot space
- house grading

Scatterplot Matrix

Data insights

Correlation between

Price and the size of the **neighbor's living space**

Correlation between sqft_living15 and Price

Data insights

Correlation between

Price and the size of the **neighbor's living space**

⇒ the larger the houses in neighborhood, the higher the price.

Correlation between sqft_living15 and Price
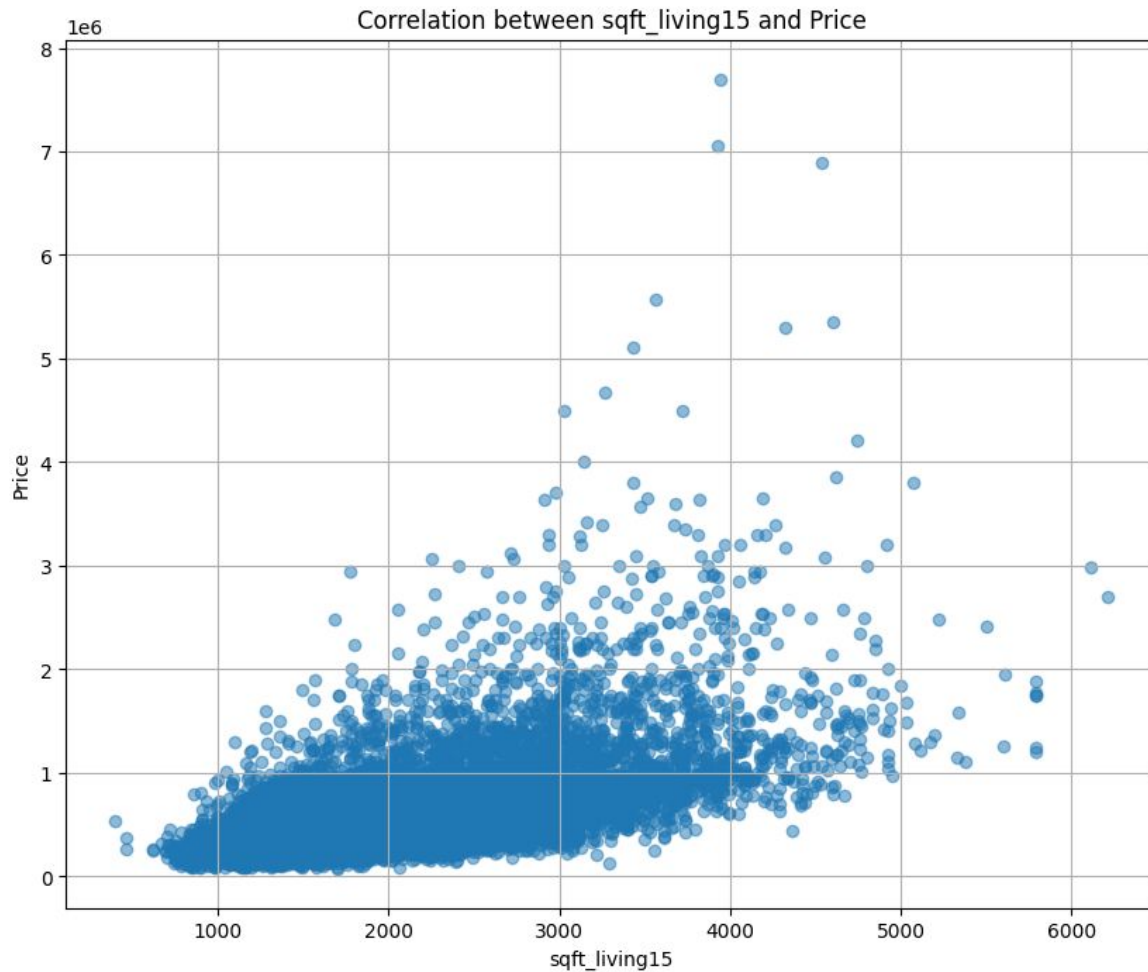
Data insights

Correlation between

Price and the size of the **neighbor's lot space**

⇒ **NO correlation**

Correlation between sqft_lot15 and Price
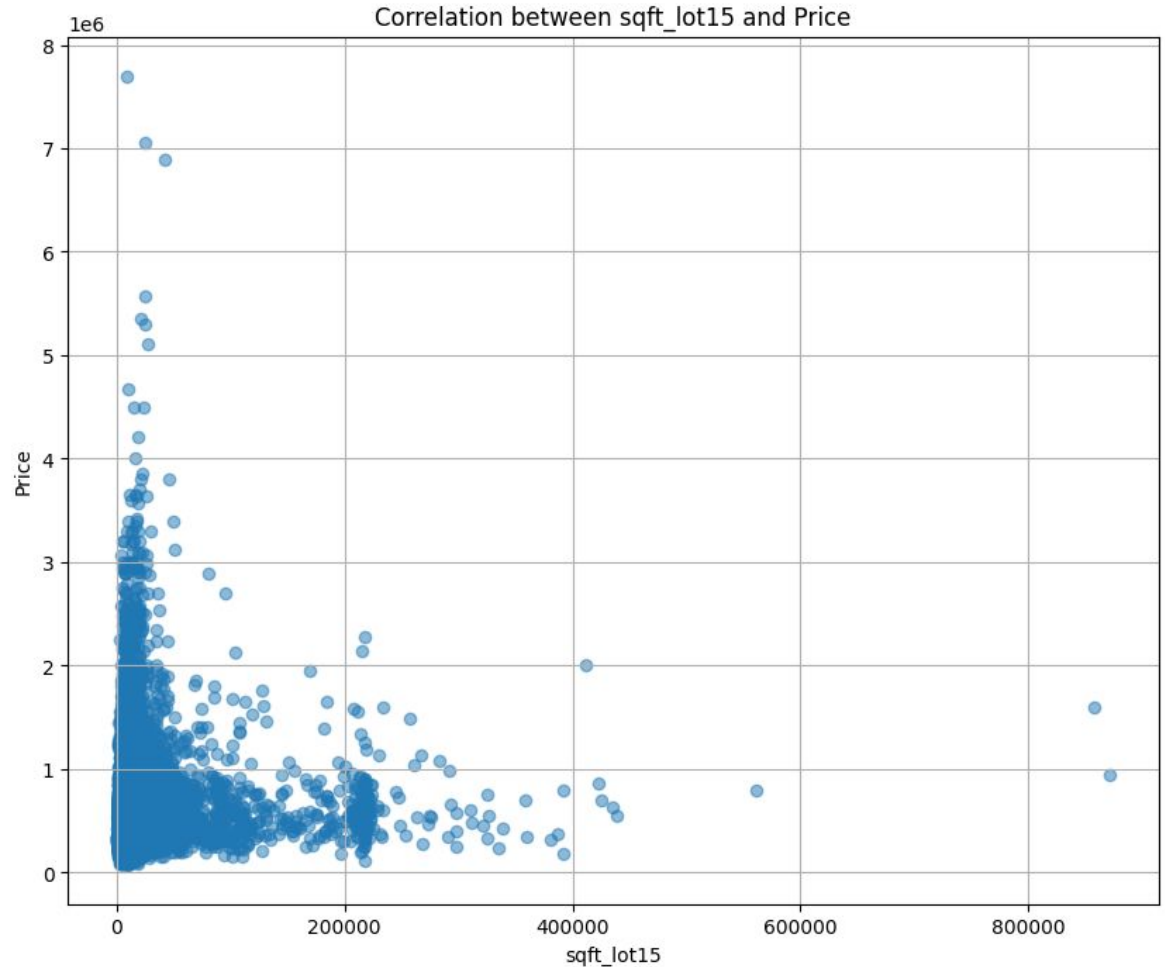
Hypotheses results

1. If houses have **waterfront** access, then the **prices** are **higher** .   **YES**

2. The **larger** the house, the higher the **price** of the house.   **YES**

3. The **more rooms** the house, the higher the **price** of the house.   **YES**

4. The higher the **grading** of the house, the higher the **price** of the house.   **YES**

5. The larger the **lot** of **neighbors**, the higher the **price**.   **NO**

6. ~~(The larger the house of neighbors, the less kids in an area.)~~

Reminder: Client Requirements

Larry

Buyer

Limited Budget

House

Waterfront access

Needs place for kids

Area

Nice & isolated, but central

Neighborhood without many children

Finding houses

1

Goal

Fulfilling as many requirements as possible

Within the budget

2

House Selection

Price: max $650.000

Bedrooms: 3+

Waterfront

Decent house condition & grading (average or better)

3

Area

Next to city centre

Avoid dense areas

Avoid areas with good children's infrastructure

House selection - top 5

House features								Location		Distances to		
House ID	**Price**	**Bedrooms**	**Bathrooms**	**Water-front**	**Living Space (ft²)**	**Lot (ft²)**	**Floors**	**House Condition**	**House Grading (1-13)**	**Zip code** **comment**	**School (in miles)**	**Next City Centre (in miles)**	**Seattle (in miles)**
7631800110	$380,000	3	2	Yes	1,980	17,342	2	Good	10	98166	0.8	1.1	13.6
121039083	$629,000	3	1	Yes	1,460	12,367	2	Very Good	8	98023	0.9	1.7	16.3
3222049055	$650,000	3	1	Yes	2,800	19,386	1	Fair	8	98198	0.8	1.4	17.2
2013802030	$357,000	3	2	Yes	2,460	53,882	1	Fair	7	98198	0.8	0.9	17.6
2623029003	$635,000	3	1	Yes	1,940	167,125	1	Very Good	7	98070 Ferry needed to Seattle	0.3	4.3	11.2

Side note: Finding distances to relevant locations

1   Use house coordinates

2   Find coordinates for closest preschool, school, playground, city centre, Seattle

3   Calculate distances

4   Add distances to dataset

Recommendation for Larry

1

2 houses **fulfilling all requirements** have a decent price tag

<= $380k

3 alternatives available

2

Focus at **houses with waterfront**

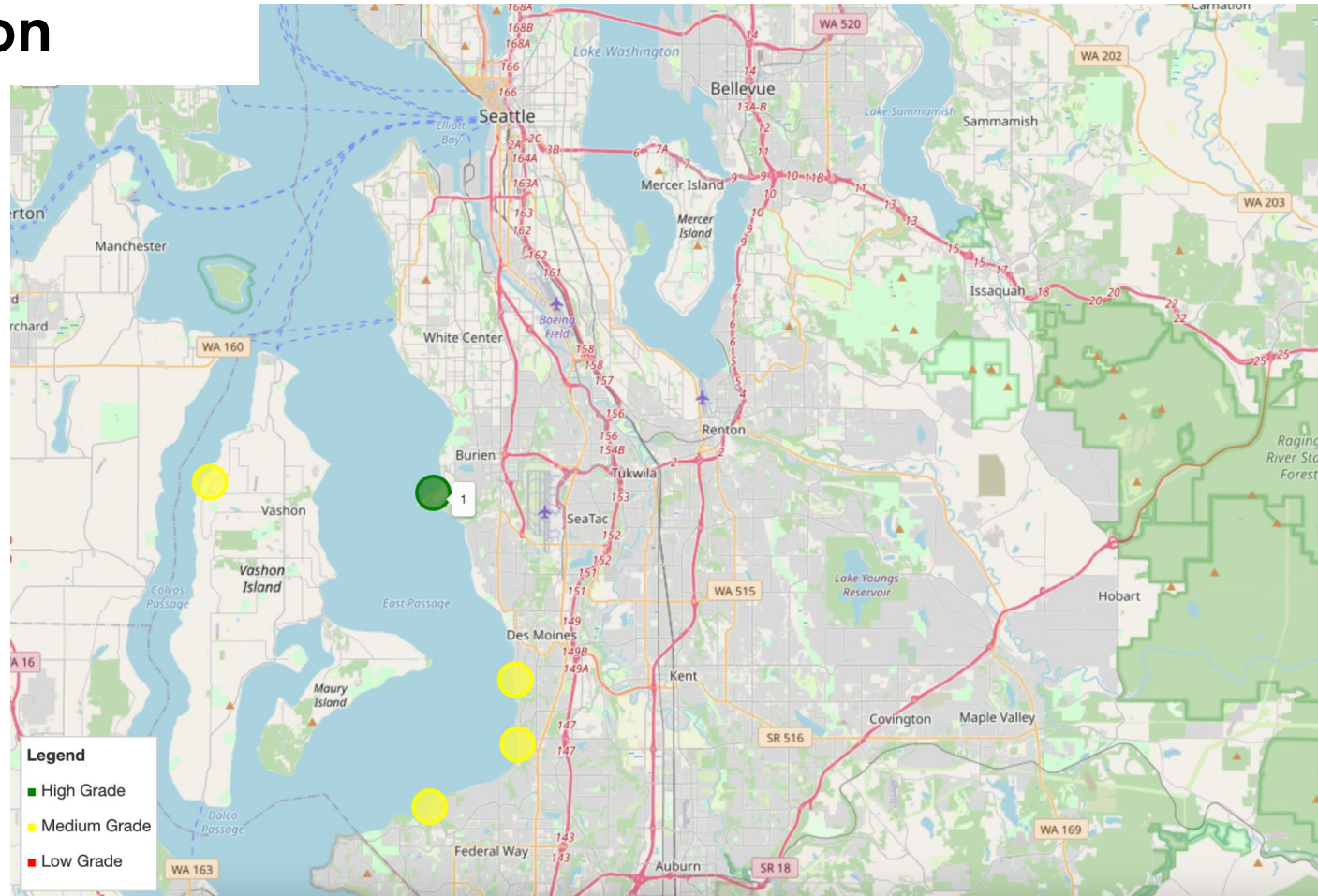despite general price differences to houses without, there are houses which fulfill the waterfront requirement.
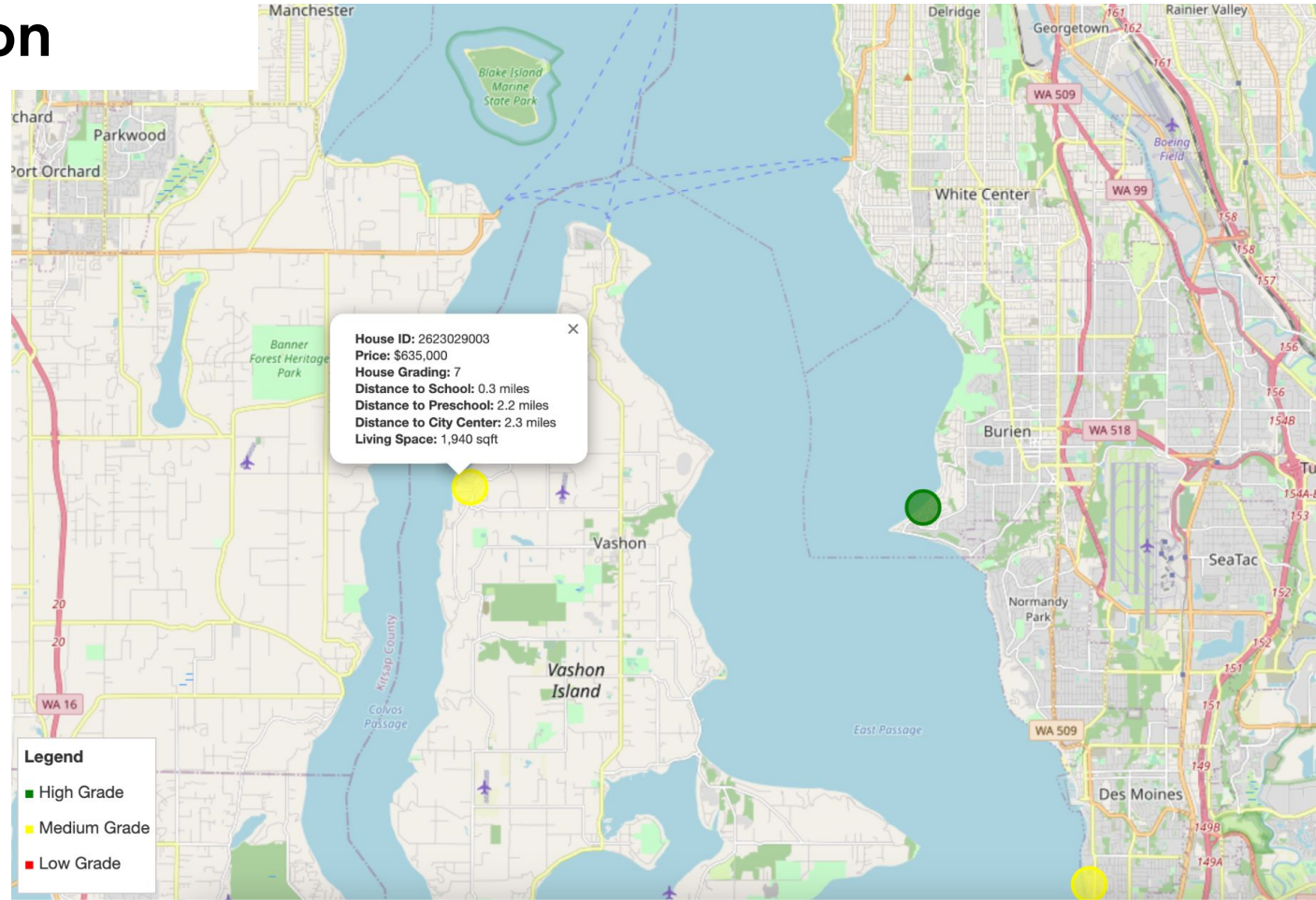
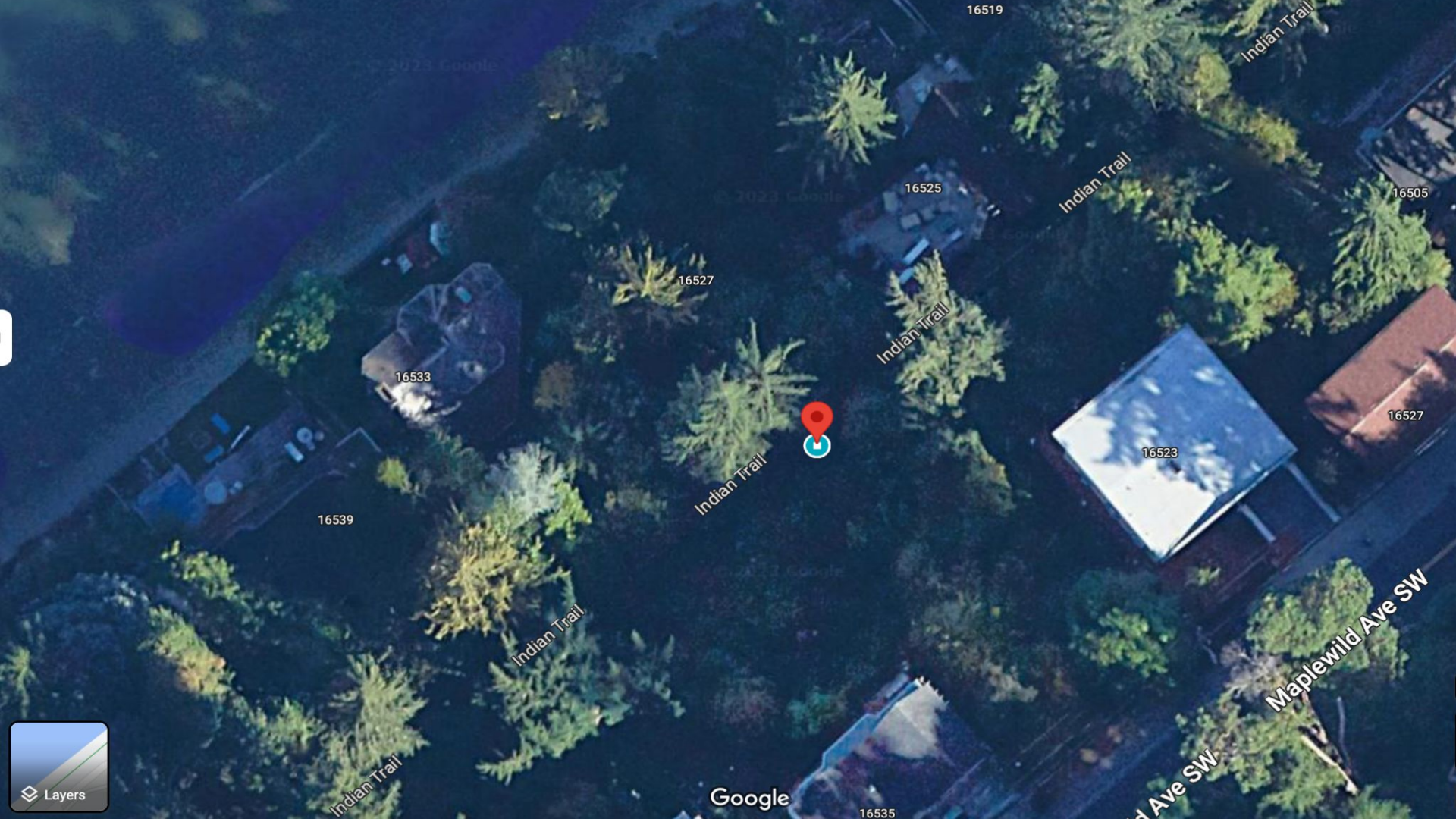**3**

House no. 1

Southwest of Burien (House ID: 7631800110)

$ 380k
3 bedrooms
highest grading
best location

House location

Legend
- ■ High Grade
- ● Medium Grade
- ■ Low Grade

House location

House ID: 2623029003
Price: $635,000
House Grading: 7
Distance to School: 0.3 miles
Distance to Preschool: 2.2 miles
Distance to City Center: 2.3 miles
Living Space: 1,940 sqft

Legend
- High Grade
- Medium Grade
- Low Grade

Final Thoughts

1. The house selection is based on assumptions (i.e. 2 kids ⇒ 3 bedrooms), therefore **additional feedback** and requirements would improve the quality of the house selection ⇒ Ask Larry

2. **Implement weighting** of the requirements (i.e. how important is waterfront or # of rooms)

3. Use **database for KC children's infrastructure**, include data into new database for all houses, consider weighting of distances (see 2)

4. **Additional results** possible: i.e. include NaN results for waterfront, group per zip code, add houses with "relative" closeness to water

5. ToDos: Optimise readme, github, notebook