EDA project presentation

Christian Reimann

15.07.2024

The client: Larry Sanders

- Buyer with limited budget
- Houses on the waterfront only
- Isolated, but central neighborhood
- Has kids

The task: Get insights on the data, create and proof at least three hypothesis and find houses that match the client’s criteria.

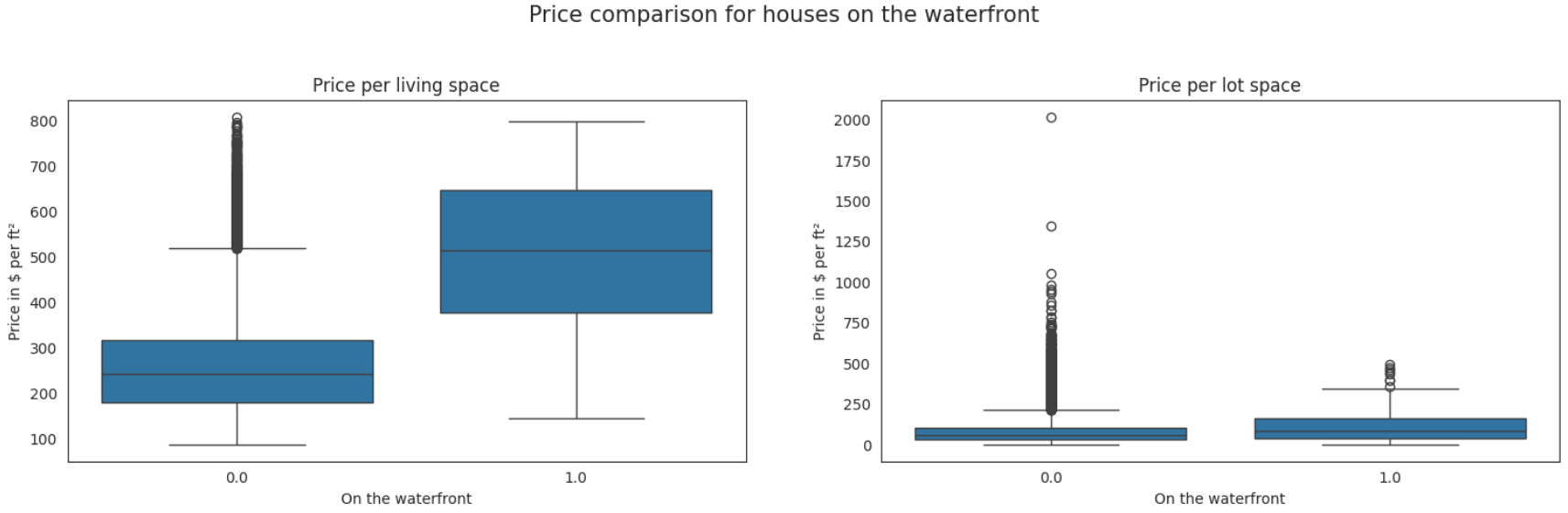
The dataset

- King County Housing Data: 21420 objects, 21 columns, 3.5+ MB
- Missing values for waterfront (11%) and yr_renovated (18%)
- No duplicate rows, no rows were dropped
- Primary focus on price, living space, lot space and distance to the city center
- **Correlations**:
  Grade ~ Living space (0.76)
  Bathrooms ~ Living space (0.76)
  Living space ~ Living space 15 (0.76)
  Lot space ~ Lot space 15 (0.72)
  Price ~ Living space (0.70)

Hypothesis

1. Locations at the waterfront are more expensive than those on the landside.
2. Houses closer to the city center are more expensive than those farer away.
3. Houses closer to the city center have less lot space.

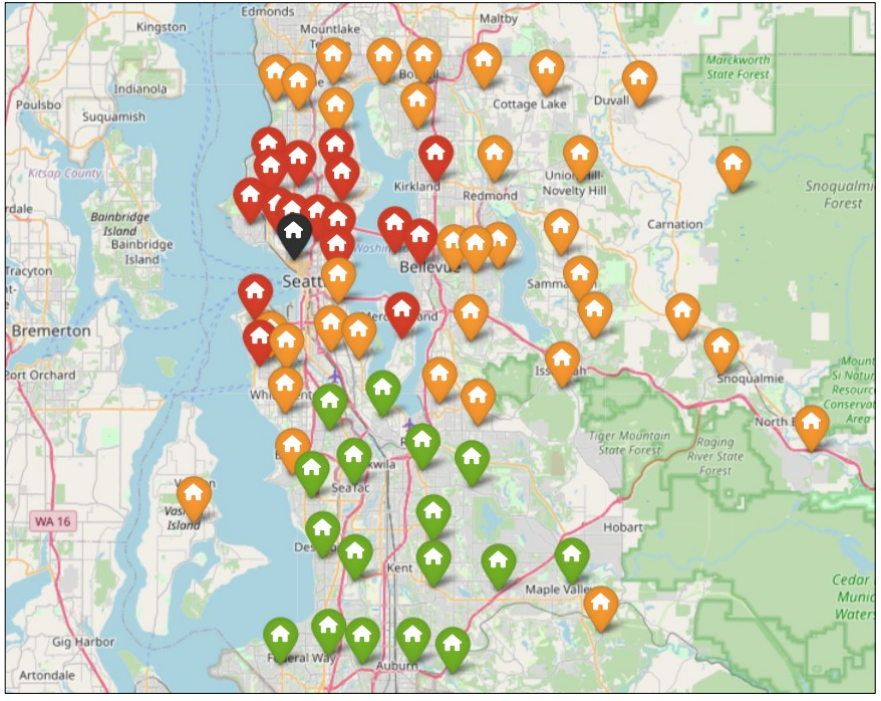
Locations at the waterfront are more expensive than those on the landside



Difference of 250 $ per ft². Possibly inaccurate results due to weak data on waterfront houses (only 146 objects).

Houses closer to the city center are more expensive than those farer away

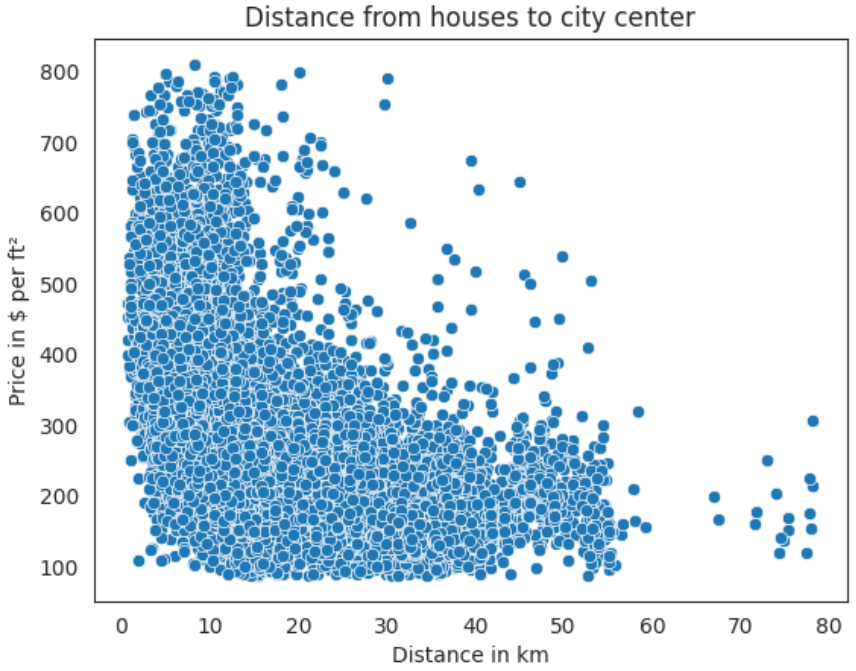
Median price per ft² living space by zipcode



Percentile	Category	Color
p < P25	cheap	**green**
P25 <= p < P75	normal	**orange**
p >= P75	expensive	**red**

P25: 182 $ per ft²
P75: 318 $ per ft²

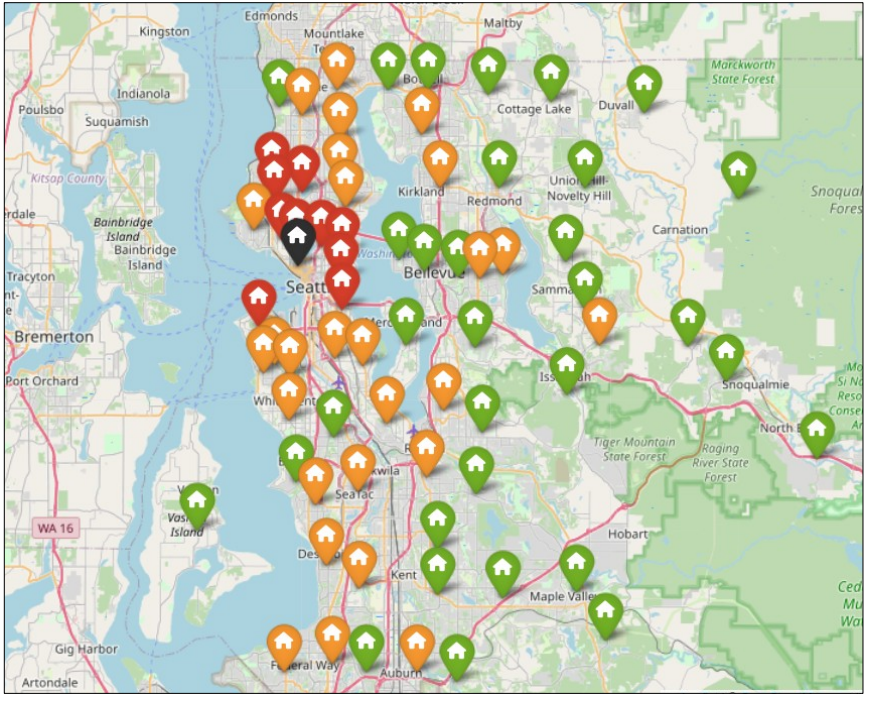
Houses closer to the city center are more expensive than those farer away



- Haversine distance
- Seattle Space Needle as city center
- Plotting distance against Price per ft² looks similar to f(x)=1/x

Houses closer to the city center have less lot space

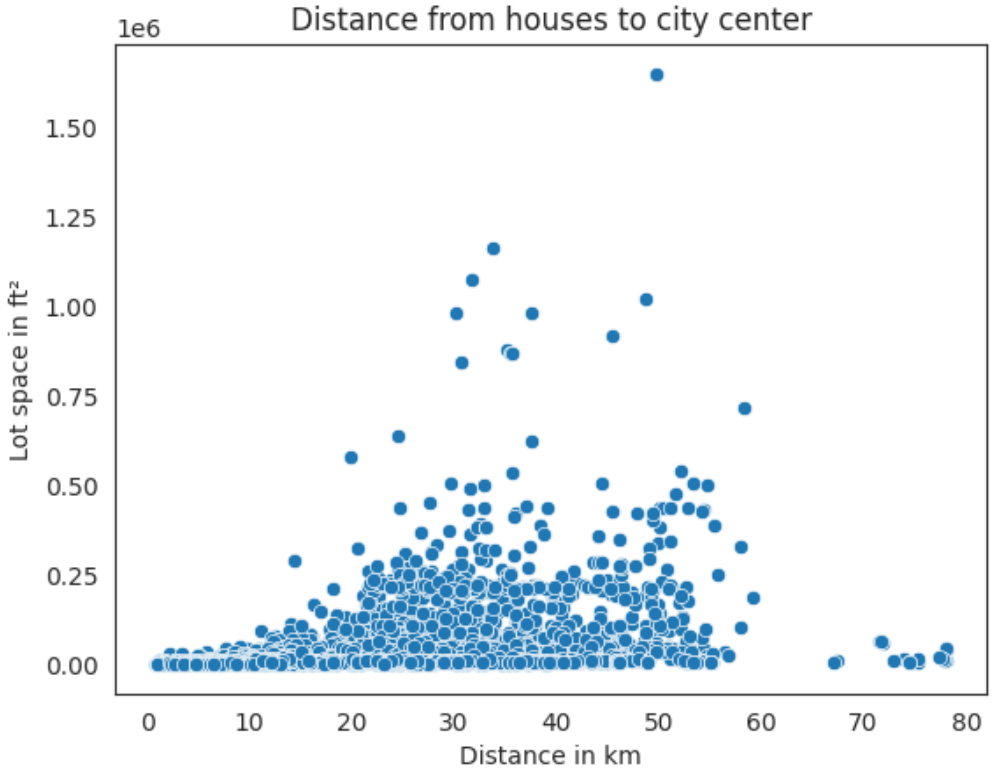
Mean lot space by zipcode



Percentile	Category	Color
s < P25	small	**red**
P25 <= s < P75	normal	**orange**
s >= P75	large	**green**

P25: 5040 ft², **P75**: 10685 ft²

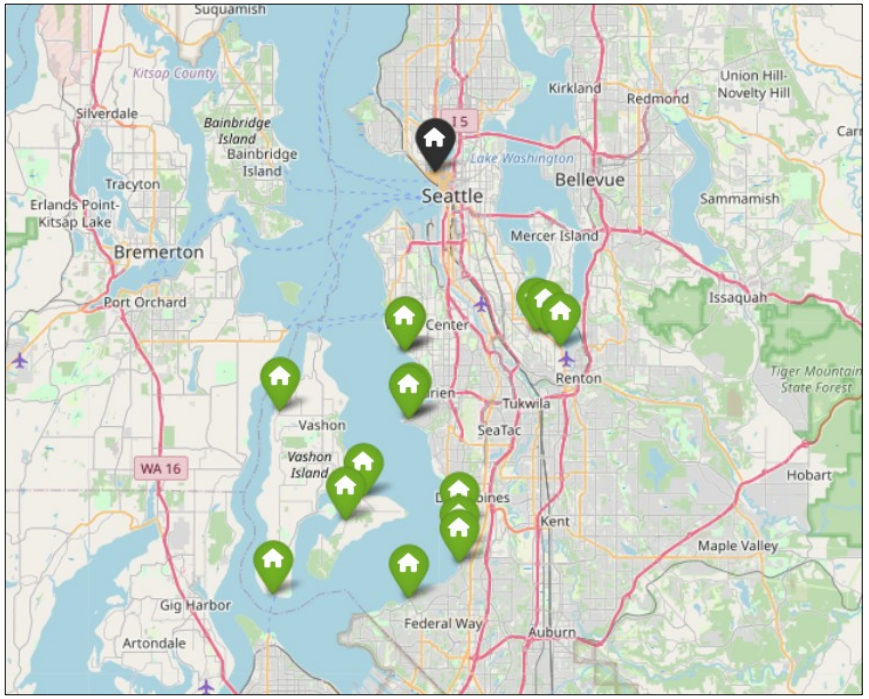
Houses closer to the city center have less lot space



Recommendations

- Budget of 800.000 $
- Didn’t tell the number of kids, so at least three bedrooms are required
- House has to be on the waterfront
- Larry is picky: Only grade 7 houses or better

Recommendations



- Resulting dataset contains 15 objects
- Further narrowing it down to the houses with the smallest distance
- Sort by the largest lot space
- Select the top five.

Recommendations

Price in $	Distance in km	Built	Grade	Bedrooms	Living space ft^2	Lot space ft^2	Living space ft^2 15	Lot space ft^2 15
540000	13.6	1912	8	3	2600	23361	1700	14700
750000	13.6	1960	8	4	2520	21834	1700	8100
770000	18.7	1930	9	3	2050	21744	2300	12200
380000	18.5	1984	10	3	1980	17342	2060	17313
750000	16.2	1954	8	5	2640	13290	2400	11942

Thank you.