EDA project

Andrea Schrader

12.05.2022

The stakeholder

social
footprint

renovate

win

WIN

buy   sell

Name: Erin Robinson
Type: Buyer
Description: Invest in poor neighborhood, buying & selling, costs back + little profit, socially responsible.

King County population

U.S. Decennial Census[3], 1790–1960[4] 1900–1990[5], 1990–2000[6] 2010–2020[1]

Table extracted from

https://en.wikipedia.org/wiki/King_County,_Washington with the

following references

[1] State & County QuickFacts". United States Census Bureau. Retrieved April 5, 2022.
[2] Annual Estimates of the Resident Population for Counties: April 1, 2020 to July 1, 2021". Retrieved April 5, 2022.
[3] "U.S. Decennial Census". United States Census Bureau. Retrieved January 7, 2014.
[4] "Historical Census Browser". University of Virginia Library. Retrieved January 7, 2014.
[5] "Population of Counties by Decennial Census: 1900 to 1990". United States Census Bureau. Retrieved January 7, 2014.
[6] "Census 2000 PHC-T-4. Ranking Tables for Counties: 1990 and 2000" (PDF). United States Census Bureau. Retrieved January 7, 2014.

|	Historical population	
Census	**Pop.**	**%±**
1860	302	—
1870	2,120	602.0%
1880	6,910	225.9%
1890	63,989	826.0%
1900	110,053	72.0%
1910	284,638	158.6%
1920	389,273	36.8%
1930	463,517	19.1%
1940	504,980	8.9%
1950	732,992	45.2%
1960	935,014	27.6%
1970	1,156,633	23.7%
1980	1,269,749	9.8%
1990	1,507,319	18.7%
2000	1,737,034	15.2%
2010	1,931,249	11.2%
2020	2,269,675	17.5%
2021 (est.)	2,252,305[2]	16.6%

The dataset

"King County House Data" on home sales in King County (USA)

House sales by year

4
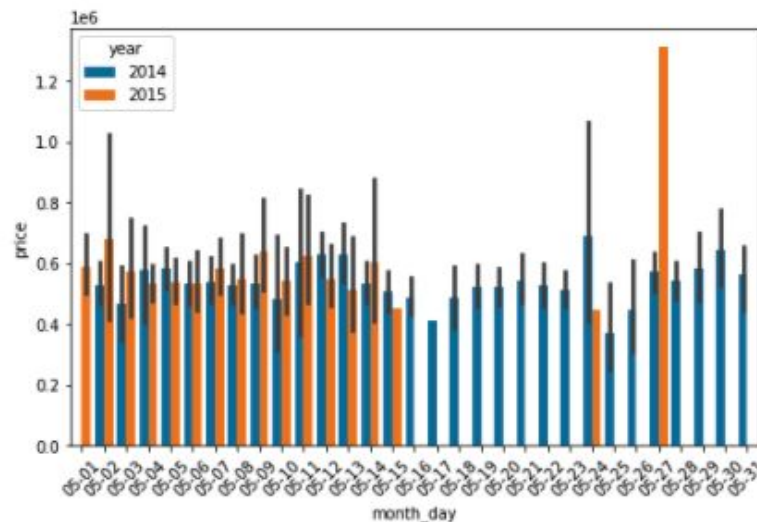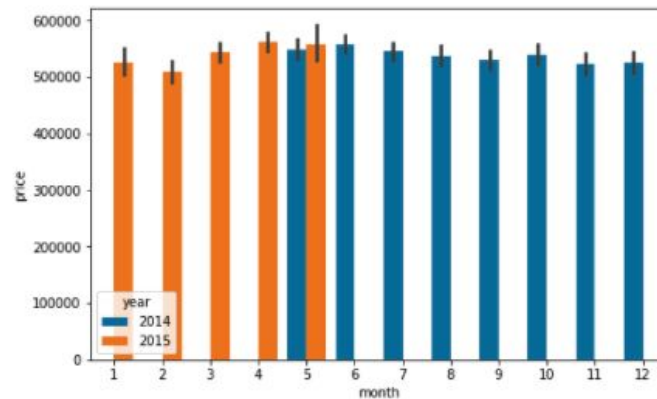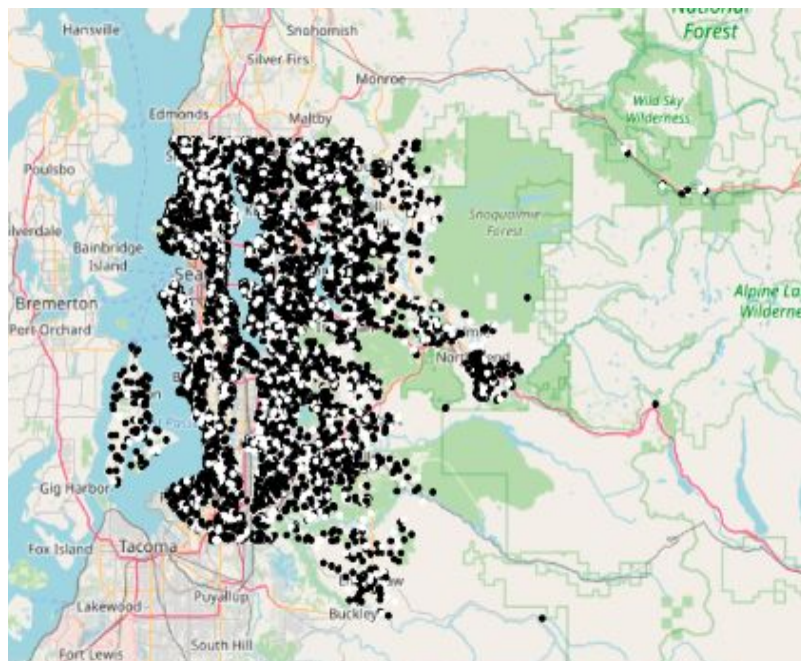
Question 1: Where are most houses with low prices located?
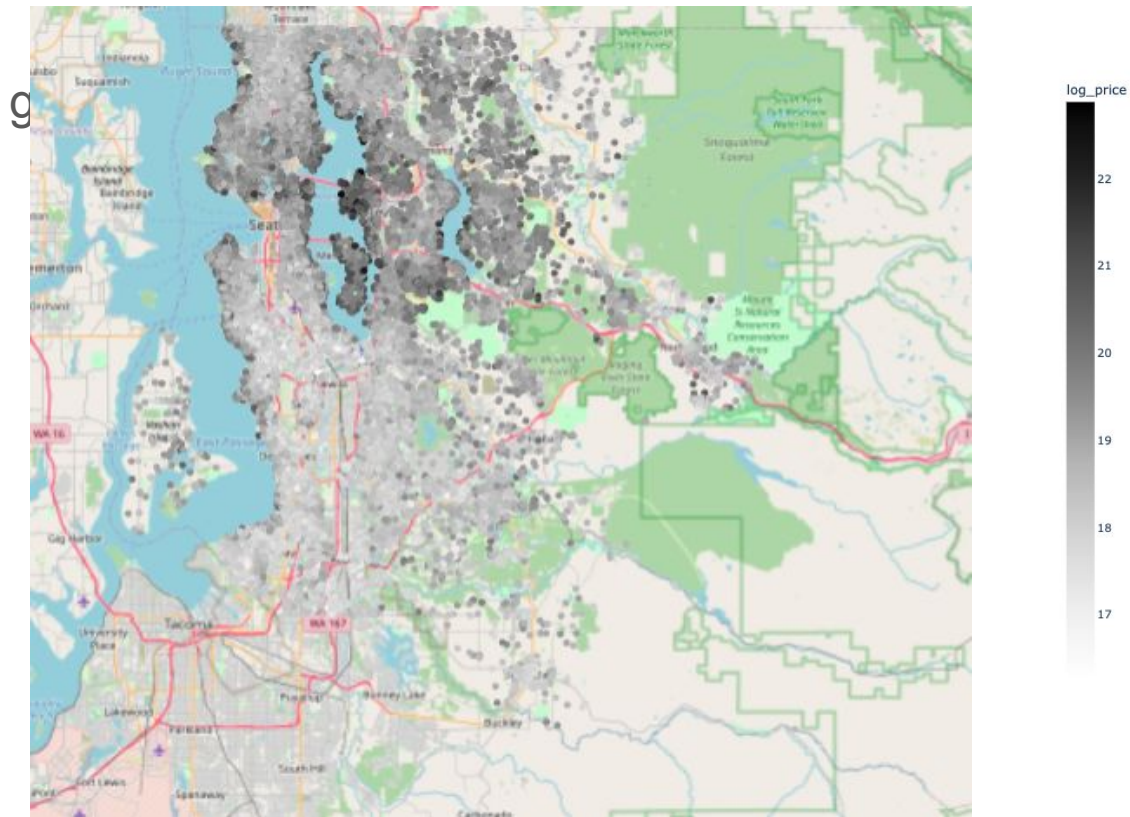
Question 2: How much money can be spent for renovations in comparably poor areas to ensure that costs are refunded and a little profit is achieved based according to the data?

- Introduction of multi-sale class

  multi-scale class houses: 176

  single-sale class houses: 21244

- location
- safe investment?
- best data background in the dataset
- definition, sold again within a year
- location matches roughly
- gain? how much for renovation?
- rise in price plot (bar)

Comparison of house sale prices:
- single-sale vs. multi-sale class of sales -

Question 2: How much money can be spent for renovations in comparably poor areas to ensure that costs are refunded and a little profit is achieved based according to the data?
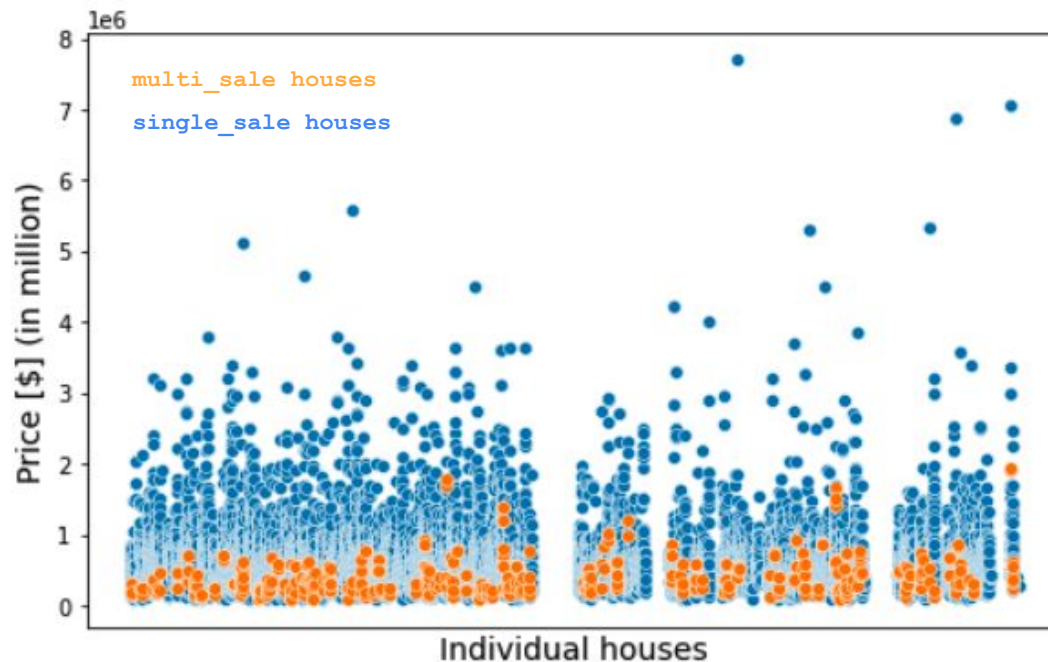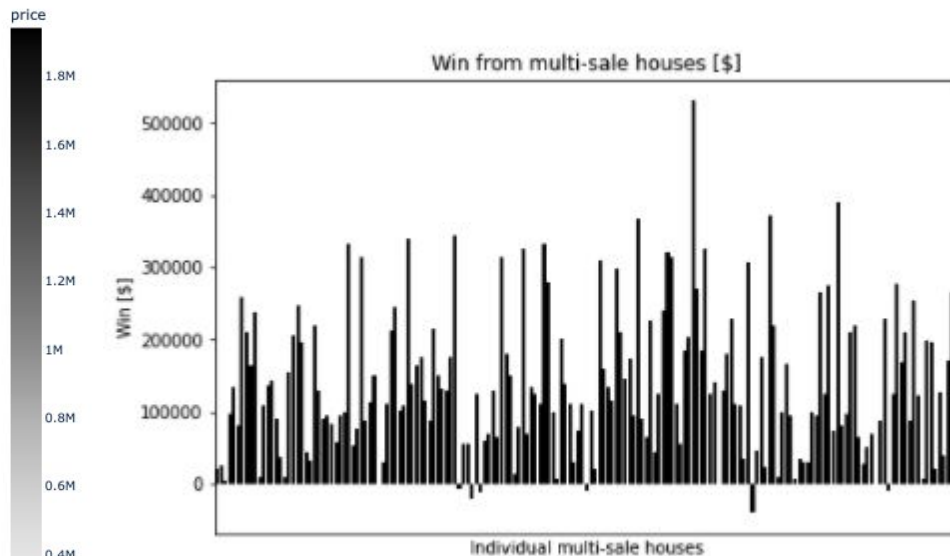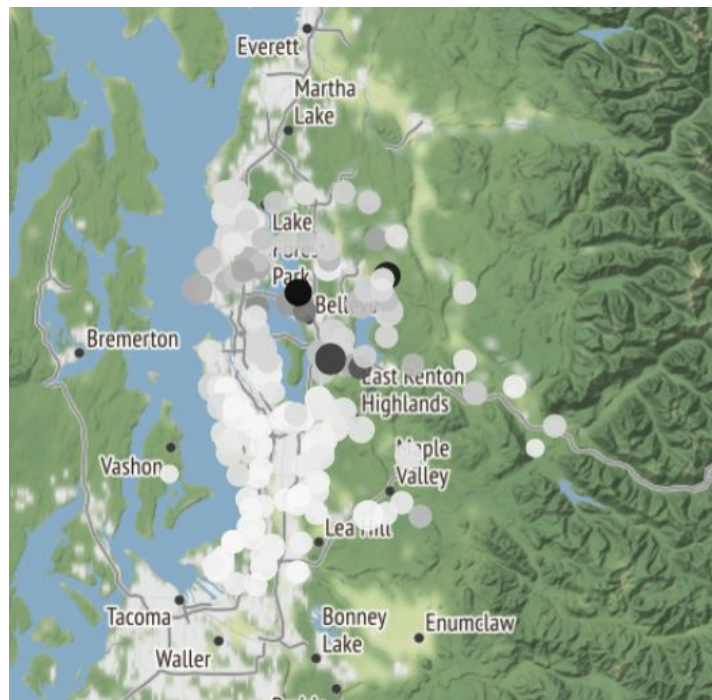
Question 3: Are there additional risks due to the condition of these houses and can these be overcome by renovations?

MEDIAN

	features_single	feature_name	features_multi
1	451000.000	price	340000.000
2	3.000	bedrooms	3.000
3	2.250	bathrooms	1.750
4	1920.000	sqft_living	1600.000
5	7606.000	sqft_lot	7980.000
6	1.500	floors	1.000
9	3.000	condition	3.000
10	7.000	grade	7.000
11	1570.000	sqft_above	1290.000
12	1975.000	yr_built	1959.000
13	0.000	yr_renovated	0.000
17	1840.000	sqft_living15	1610.000
18	7620.000	sqft_lot15	7920.000

1/25

Andrea Schrader - 12.05.2022 - **neuefische - DataScience bootcamp - 2022-2**

Insights and Recommendations

1) Where are most houses with low prices located?

Answer: In the outer areas of the city, especially in the south and not at the water.

2) How much money can be spent for renovations in comparably poor areas to ensure that costs are refunded and a little profit is achieved based according to the data?

Answer: Most parts of the hypothesis are true. While the area for living was indeed smaller and multi-sale houses had less floors, the area of the lot was surprisingly larger. this implies that there might also be more room for extending the living area if allowed.

3) Are there additional risks due to the condition of these houses and can these be overcome by renovations?

The quality of these houses is comparable to the median of the single-sale houses.

With respect to the number of renovations conducted, the hypotheses is also not supported by the data. However, the dataset is small for renovations and errors might have occurred for the "0" values. One might consider to asl those collecting the data or, if not accessible,to suggest to conduct a new survey focussing on this aspect in case a stakeholder requests more secure recommendations based on this feature.

Outlook

- predict houses which will be in the same class in the futur
- maybe add more data on renovations
- succeed competitors
- tightly monitor the market and development of houses with the same characteristics and in the neighborhood of known multi-sale houses

Should the class of multi-sale houses be kept in analysis pipelines in the future?

YES, this class of houses can be recommended to the stakeholder and can be used for future predictions for similar strategies.

Thank you

for your attention

&

of course …   our patient coaches :)