

The key factor which impacts the house price in Kansas City

Github: <https://github.com/623059008/InfoVizProject>

Author: Tempest

Main Figure

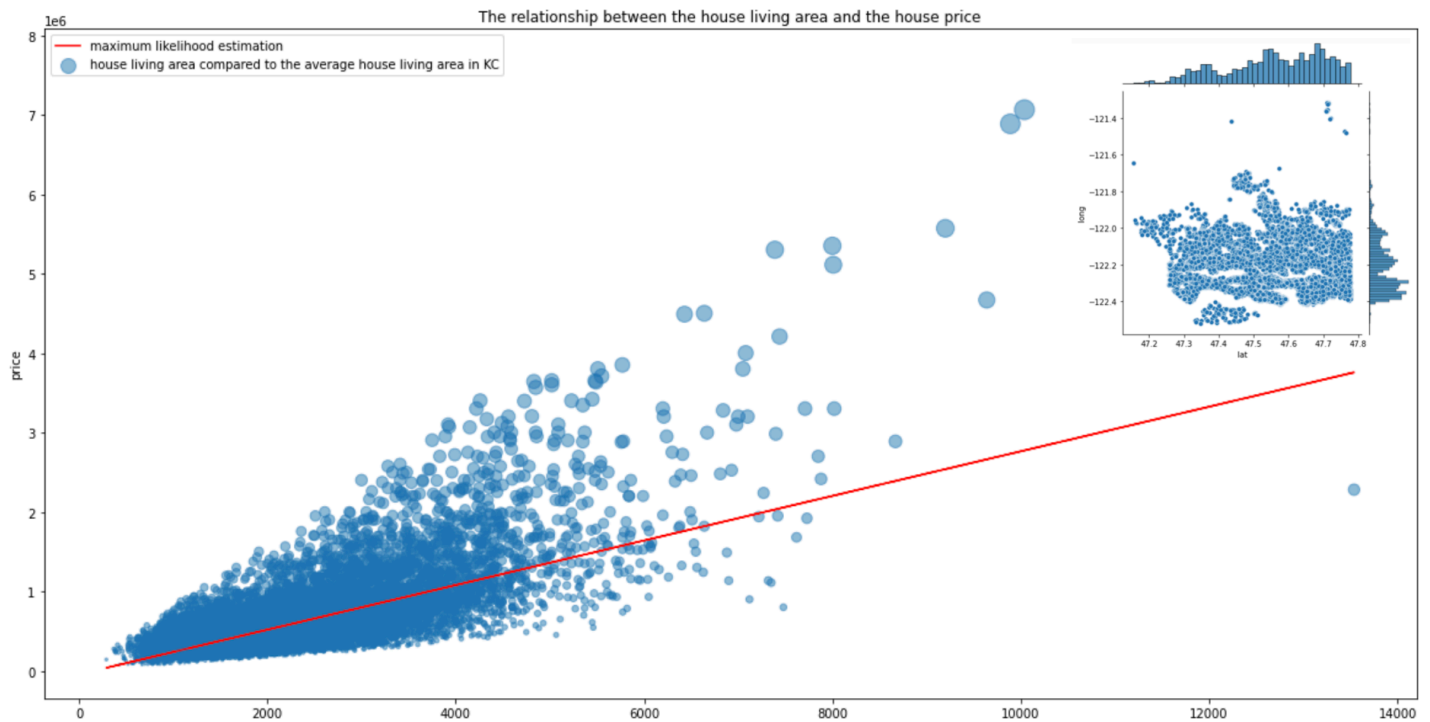


Figure Explanation

The top right small figure shows the geographic distribution of houses in KC. X-axis shows latitude and Y-axis shows longitude. Each point in this scatter represents a house.

The main figure shows the correlation between house living area and house price in KC. X-axis is the house living area and Y-axis is the price of the houses. On the left top corner, there is the legend of this figure. Each blue point represents a house and the size of the point represents the ratio of the house living area to the average house living area.

The red line represents the maximum likelihood estimation for house price to house area.

Findings

It shows housing prices are **positively correlated** with the house living area.

Using **linear regression**, it shows how the typical positive correlation is between house price and house living area.

There are also some acceptable abnormal points. Besides, the amount of houses with low price and small area is majority compared to the amount of houses with higher price and larger area.

DataSet

[Kansas City_house_data.csv](#)

This dataset downloads from Kaggle.

In this dataset, there are date, price, bedrooms, bathrooms, sqft_living, sqft_lot, floors, waterfront, view, sqft_basement, sqft_above, building year, renovated year, lat and long, zipcode.

	id	date	price	bedrooms	bathrooms	sqft_living	sqft_lot	floors	waterfront	view	...	grade	sqft_above	sqft_basement	yr_built
0	7129300520	20141013T000000	221900.0	3	1.00	1180	5650	1.0	0	0	...	7	1180	0	1955
1	6414100192	20141209T000000	538000.0	3	2.25	2570	7242	2.0	0	0	...	7	2170	400	1951
2	5631500400	20150225T000000	180000.0	2	1.00	770	10000	1.0	0	0	...	6	770	0	1933
3	2487200875	20141209T000000	604000.0	4	3.00	1960	5000	1.0	0	0	...	7	1050	910	1965
4	1954400510	20150218T000000	510000.0	3	2.00	1680	8080	1.0	0	0	...	8	1680	0	1987

	id	price	bedrooms	bathrooms	sqft_living	sqft_lot	floors	waterfront	view	condition	grade	sqft_above	sqft_basement	yr_built
id	1.00	-0.02	0.00	0.01	-0.01	-0.13	0.02	-0.00	0.01	-0.02	0.01	-0.01	-0.01	0.02
price	-0.02	1.00	0.31	0.53	0.70	0.09	0.26	0.27	0.40	0.04	0.67	0.61	0.32	0.05
bedrooms	0.00	0.31	1.00	0.52	0.58	0.03	0.18	-0.01	0.08	0.03	0.36	0.48	0.30	0.15
bathrooms	0.01	0.53	0.52	1.00	0.75	0.09	0.50	0.06	0.19	-0.12	0.66	0.69	0.28	0.51
sqft_living	-0.01	0.70	0.58	0.75	1.00	0.17	0.35	0.10	0.28	-0.06	0.76	0.88	0.44	0.32
sqft_lot	-0.13	0.09	0.03	0.09	0.17	1.00	-0.01	0.02	0.07	-0.01	0.11	0.18	0.02	0.05
floors	0.02	0.26	0.18	0.50	0.35	-0.01	1.00	0.02	0.03	-0.26	0.46	0.52	-0.25	0.49
waterfront	-0.00	0.27	-0.01	0.06	0.10	0.02	0.02	1.00	0.40	0.02	0.08	0.07	0.08	-0.03
view	0.01	0.40	0.08	0.19	0.28	0.07	0.03	0.40	1.00	0.05	0.25	0.17	0.28	-0.05
condition	-0.02	0.04	0.03	-0.12	-0.06	-0.01	-0.26	0.02	0.05	1.00	-0.14	-0.16	0.17	-0.36
grade	0.01	0.67	0.36	0.66	0.76	0.11	0.46	0.08	0.25	-0.14	1.00	0.76	0.17	0.45
sqft_above	-0.01	0.61	0.48	0.69	0.88	0.18	0.52	0.07	0.17	-0.16	0.76	1.00	-0.05	0.42
sqft_basement	-0.01	0.32	0.30	0.28	0.44	0.02	-0.25	0.08	0.28	0.17	0.17	-0.05	1.00	-0.13
yr_built	0.02	0.05	0.15	0.51	0.32	0.05	0.49	-0.03	-0.05	-0.36	0.45	0.42	-0.13	1.00

Method

- Firstly, I made some **assumptions** to explore the correlation between house price and other factors.
- Then I draw the **scatter figures** of house price and living area, lot area, longitude, amout of bedrooms, year built and year renovated.
 - After reviewing all of these figures, I found most of these factors are not directly correlative with the

house price except living area.

- Then I use **linear regression** model to show this positive correlation to validate my assumption.
- Finally, I finish the conclusion: housing prices are **positively correlated** with the house living area.

Importance

This topic is very important because it helps us find a more cost-effective house in KC. According to above conclusion, we need to consider the living area as main factor to choose our target house and other factors do not inapparently increase the price which means all these inapparent factors can be set as optimal.