DATA BIRD

REAL STATE ANLYSIS

HAWAII

Exploratory Data Analysis (EDA) -Real State Analysis

About Dataset

Original dataset consists of 809 rows (deals) and 27 columns.

Below table shows the number of missing values and data types of each feature in the dataset.

```
 #   Column                                                                                Non-Null Count  Dtype
---  ------                                                                                --------------  -----
 0   SALE TYPE                                                                             809 non-null    object
 1   SOLD DATE                                                                             801 non-null    object
 2   PROPERTY TYPE                                                                         809 non-null    object
 3   ADDRESS                                                                               808 non-null    object
 4   CITY                                                                                  809 non-null    object
 5   STATE OR PROVINCE                                                                     809 non-null    object
 6   ZIP OR POSTAL CODE                                                                    809 non-null    object
 7   PRICE                                                                                 809 non-null    int64
 8   BEDS                                                                                  552 non-null    float64
 9   BATHS                                                                                 533 non-null    float64
 10  LOCATION                                                                              801 non-null    object
 11  SQUARE FEET                                                                           536 non-null    float64
 12  LOT SIZE                                                                              720 non-null    float64
 13  YEAR BUILT                                                                            537 non-null    float64
 14  DAYS ON MARKET                                                                        0 non-null      float64
 15  $/SQUARE FEET                                                                         536 non-null    float64
 16  HOA/MONTH                                                                             355 non-null    float64
 17  STATUS                                                                                801 non-null    object
 18  NEXT OPEN HOUSE START TIME                                                            0 non-null      float64
 19  NEXT OPEN HOUSE END TIME                                                              0 non-null      float64
 20  URL (SEE https://www.redfin.com/buy-a-home/comparative-market-analysis FOR INFO ON PRICING)  809 non-null    object
 21  SOURCE                                                                                801 non-null    object
 22  MLS#                                                                                  801 non-null    float64
 23  FAVORITE                                                                              809 non-null    object
 24  INTERESTED                                                                            809 non-null    object
 25  LATITUDE                                                                              809 non-null    float64
 26  LONGITUDE                                                                             809 non-null    float64
dtypes: float64(13), int64(1), object(13)
memory usage: 170.8+ KB
```

According to the above table, it shows some of the variables have missing values and they are replaced using the forward fill method.

Summary statistics

Continuous variables

The average values, standard deviations and 5 summary statistics are obtained for the continuous variables. The average price of a real estate is $454,110. 50% of the houses are higher than $182,000. More characteristics of the variables will be discussed in future sections.

There are some variables that are missing all the values (days of market, next open house start time, next open house end time) as squared they are removed from the dataset before doing further analysis.

	PRICE	BEDS	BATHS	SQUARE FEET	LOT SIZE	YEAR BUILT	DAYS ON MARKET	$/SQUARE FEET	HOA/MONTH	NEXT OPEN HOUSE START TIME	NEXT OPEN HOUSE END TIME	MLS#	LATITUDE	LONGITUDE
count	8.090000e+02	552.000000	533.000000	536.000000	7.200000e+02	537.000000	0.0	536.000000	355.000000	0.0	0.0	8.010000e+02	809.000000	809.000000
mean	4.541100e+05	2.235507	1.733583	1128.283582	1.310433e+05	1982.923650	NaN	516.440299	531.873239	NaN	NaN	7.510187e+07	20.459630	-156.524951
std	6.107641e+05	1.794013	0.929031	783.162290	3.787302e+05	20.228816	NaN	345.825476	371.519955	NaN	NaN	9.737080e+07	0.928859	1.368855
min	5.264000e+03	0.000000	1.000000	29.000000	9.920000e+02	1920.000000	NaN	17.000000	4.000000	NaN	NaN	3.913990e+05	19.032161	-159.714066
25%	9.000000e+04	1.000000	1.000000	545.500000	8.344750e+03	1970.000000	NaN	282.750000	243.000000	NaN	NaN	6.610580e+05	19.535107	-157.841293
50%	1.820000e+05	2.000000	1.500000	922.000000	2.150000e+04	1978.000000	NaN	471.500000	498.000000	NaN	NaN	6.643950e+05	20.846039	-156.452392
75%	7.250000e+05	3.000000	2.000000	1614.000000	8.713100e+04	1996.000000	NaN	656.000000	714.000000	NaN	NaN	2.022097e+08	21.309331	-155.103069
max	6.750000e+06	18.000000	10.000000	6838.000000	4.295321e+06	2023.000000	NaN	3500.000000	2095.000000	NaN	NaN	2.022237e+08	22.226329	-154.892380

Categorical Variables

The following table shows the count of observations for each variable and the unique shows how many unique categories are in that feature. When there is 1 unique category that feature is removed as all rows have the same category.

	SALE TYPE	SOLD DATE	PROPERTY TYPE	ADDRESS	CITY	STATE OR PROVINCE	ZIP OR POSTAL CODE	LOCATION	STATUS	URL (SEE https://www.redfin.com/buy-a-home/comparative-market-analysis FOR INFO ON PRICING)	SOURCE	FAVORITE	INTERESTED
count	809	801	809	808	809	809	809	801	801	809	801	809	809
unique	1	76	7	689	51	1	76	168	1	698	3	1	1
top	PAST SALE	August-25-2022	Single Family Residential	4th Ave (awapuhi)	Honolulu	HI	96778	Waikiki	Sold	https://www.redfin.com/HI/Waianae/84-770-Kili-...	HI Information Service	N	Y
freq	809	25	284	5	122	809	87	62	801	2	388	809	809

List of unnecessary columns:

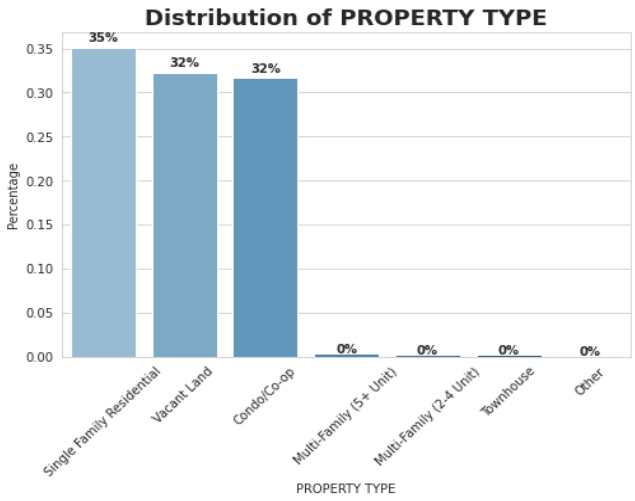
- SALE TYPE
- STATE OR PROVINCE
- FAVORITE
- INTERESTED
- STATUS
- DAYS ON MARKET
- NEXT OPEN HOUSE START TIME
- NEXT OPEN HOUSE END TIME

Univariate Analysis

Property type

The number of real estate vs property type are shown in the following table.

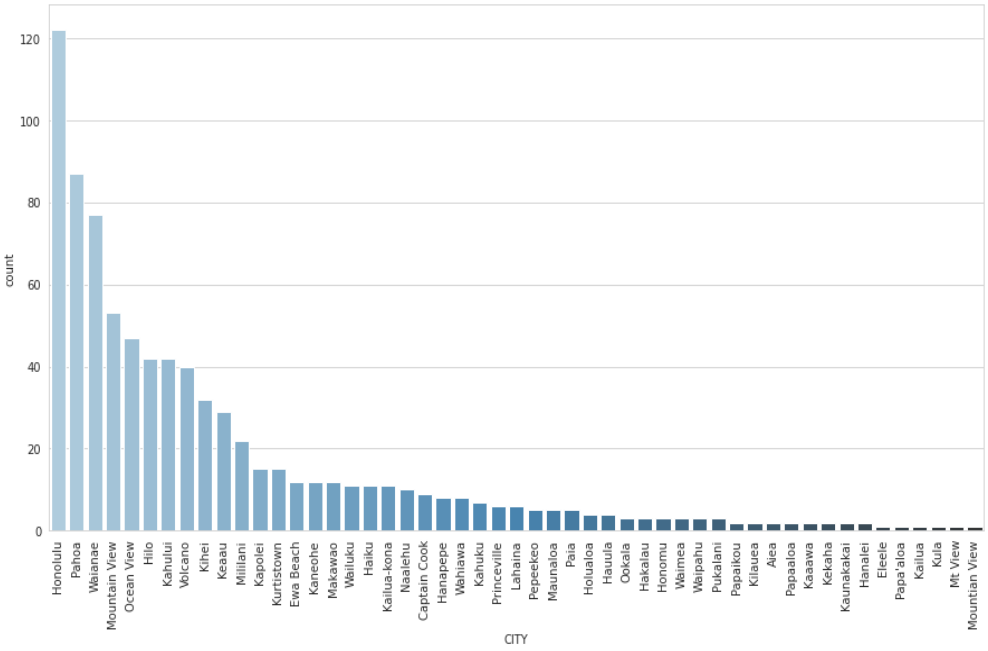
index	PROPERTY TYPE
Single Family Residential	284
Vacant Land	261
Condo/Co-op	256
Multi-Family (5+ Unit)	3
Multi-Family (2-4 Unit)	2
Townhouse	2
Other	1



Most people are looking for single family type real estate.

Most of them (99%) are looking for Single Family Residential, Vacant Land or Condo/Co-op.

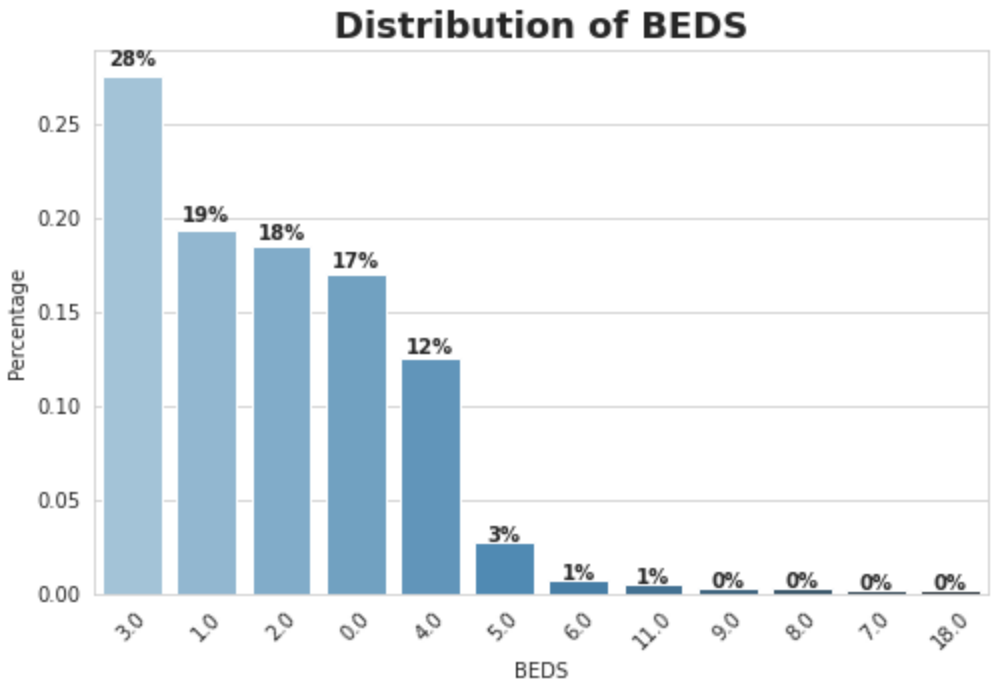
City



Majority of residences are chosen Honolulu, Pahoa and Waianae and lowest residences are chosen in Eleele, Papa’alola, Kailua, Kula, Mt view and Mountain View areas.

Beds

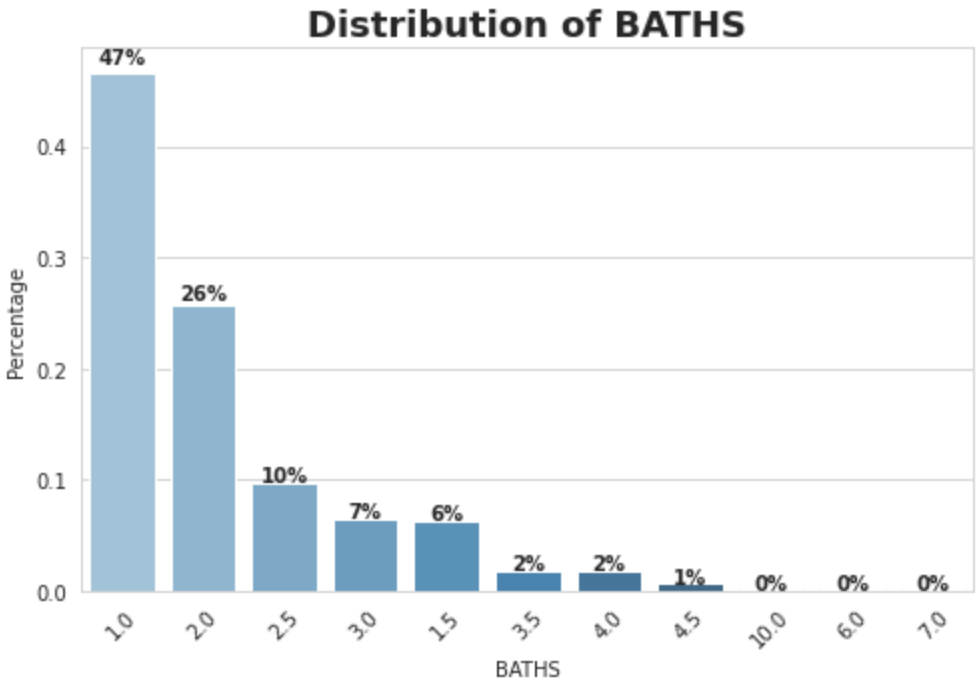
The bed variable is assumed to indicate the number of bedrooms in the property.



Almost 65% of the houses have 1-3 bedrooms and surprisingly 17% of the real estate have not any beds. We can assume these areas are vacant lands. Only a few houses have 18 beds in this is an outlier.

Baths

The baths variable is assumed to indicate the number of bathrooms in the property.



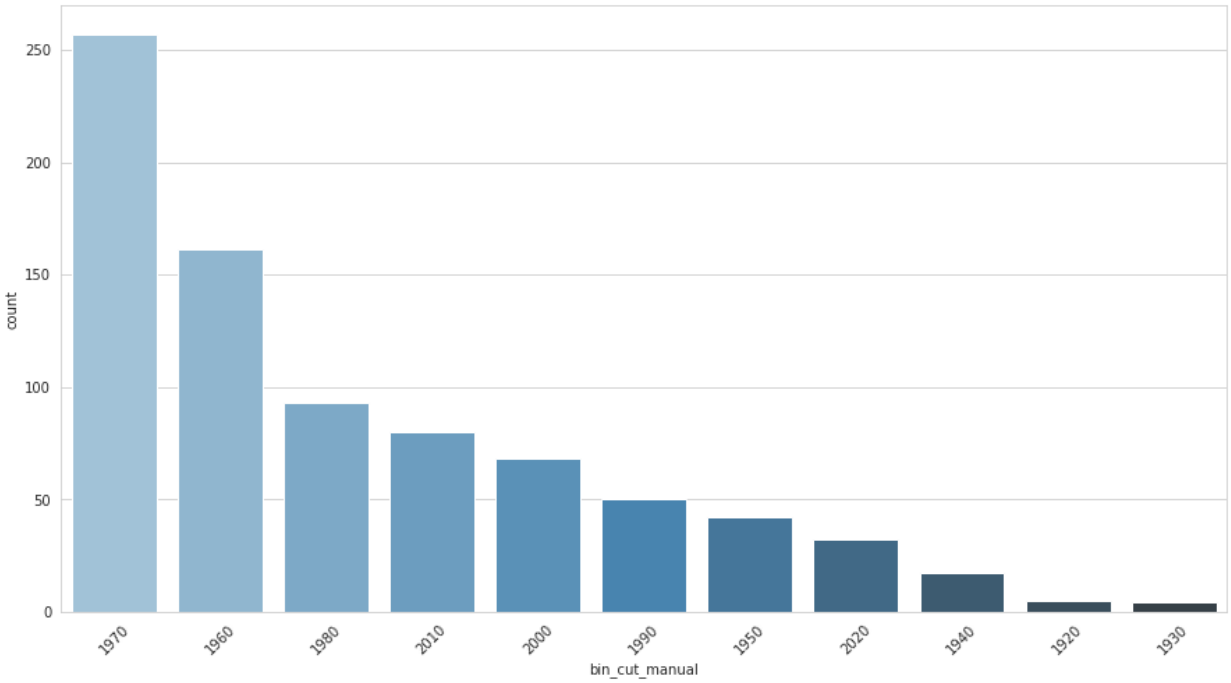
Every real estate has a Baths and almost 75% of the real states have 1 - 2 baths.

1 full bathroom [toilet, sink, shower head & tub]

0.5 baths = two of them in the above list

Year Built

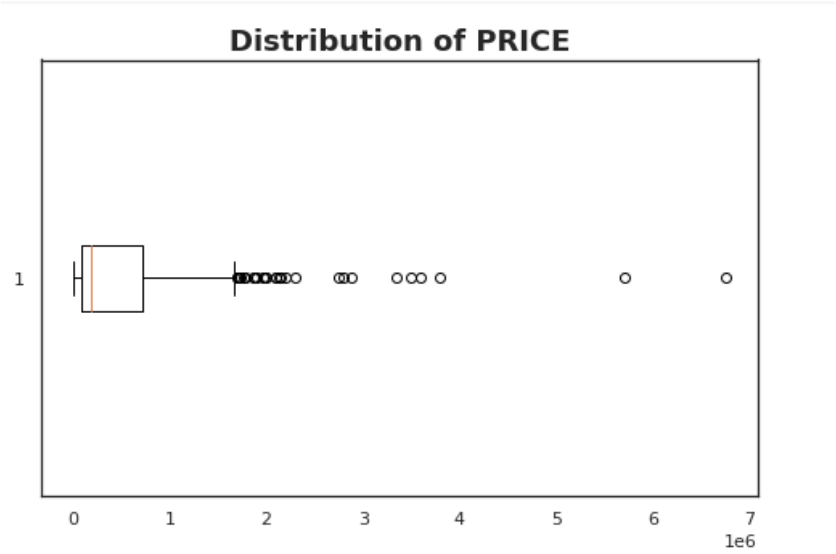
The build of the house was binned for decades.



We can see most of the houses were built between 1960 to 1980. Which means these houses are older than one generation; they are about 40-60 years old houses. Also, this dataset contains houses built in 1930 as well as the modern houses which were built after 2020.

Price

	index	PRICE
0	count	8.090000e+02
1	mean	4.541100e+05
2	std	6.107641e+05
3	min	5.264000e+03
4	25%	9.000000e+04
5	50%	1.820000e+05
6	75%	7.250000e+05
7	max	6.750000e+06



About 75% of the houses cost less than $725,000 (7.25e5 = 7.25*10^5) and the average price of the sold real estate is approximately $454,110.

Minimum value of a real estate is $5264 and highest sold price of a real estate is $ 6,750,000.

In the next step the prices will be binned as the PRICE due to huge variance and binning will reduce the variance of the data.

The bin ranges used are as follows

5264-100,000 → <100K

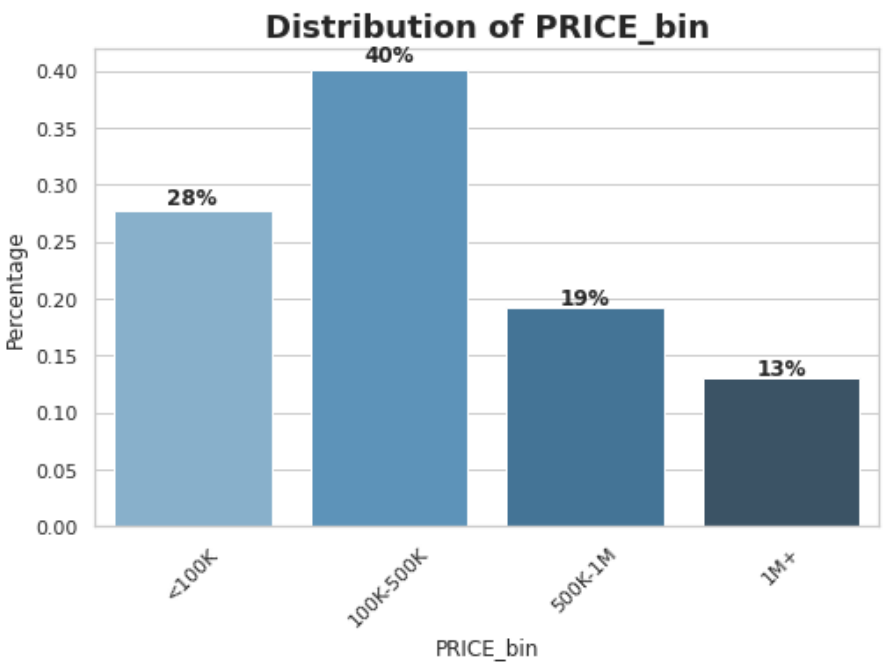
100,000 – 500,000 → 100K-500K

500,000-1,000,000 → 500K-1M

1,000,000> →1M+

The count for each bin are represented by price_bin column.

index	PRICE_bin
100K-500K	324
<100K	225
500K-1M	155
1M+	105



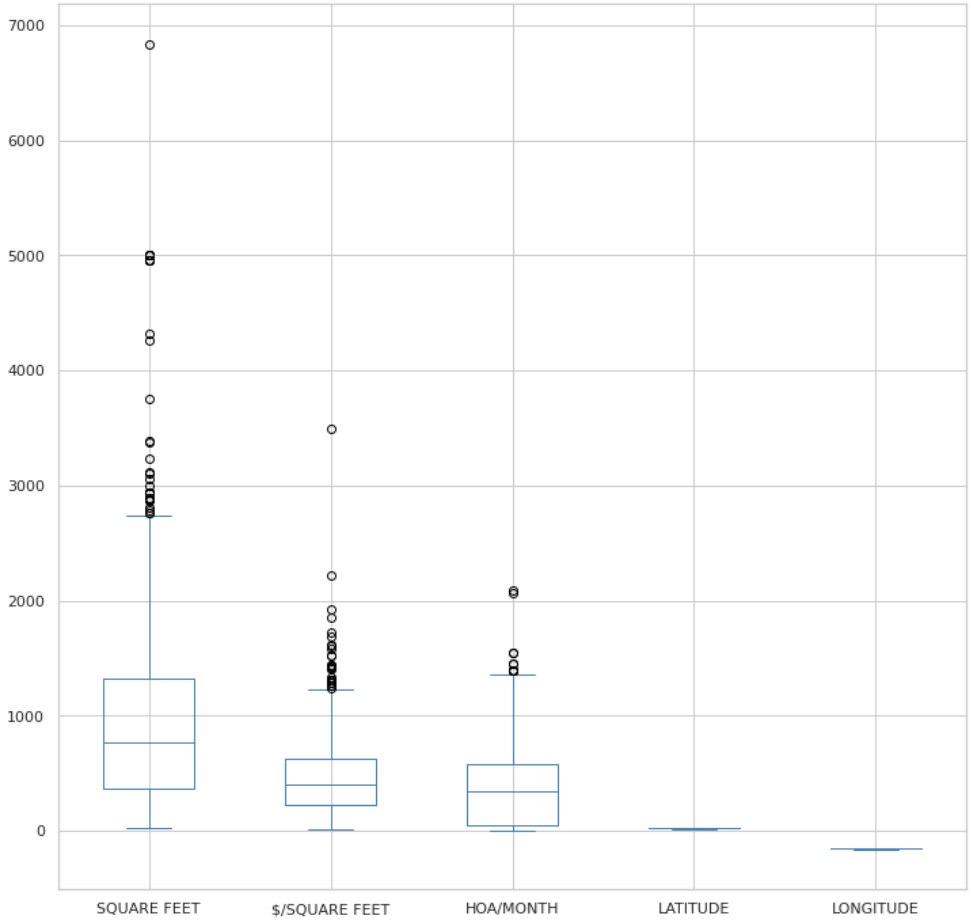
We can see only 32% of the sold real estate are worth more than $500K and 28% of the real states are sold less than $100K.

Square ft, $/Square ft/ HOA Month/ Longitude and Latitude

The average HOA is $531 and 75% of the less than $714 per month.

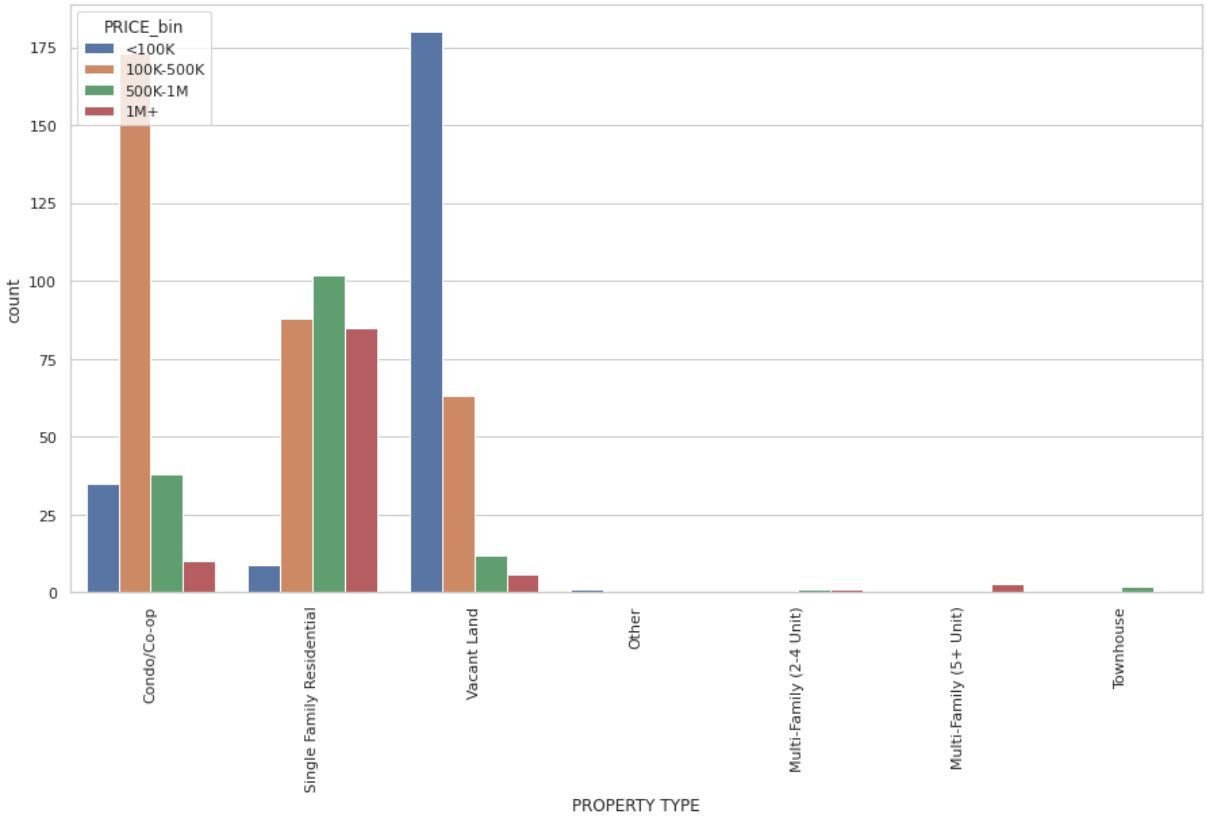
The average size of a house is about 1128 square ft, and 50% of the houses are larger than 922 square ft.

The average price per sq is approximately $516.



Bi Variate Analysis

Property type vs the price



People looking for vacant land the most frequent range is less than $100, 000.

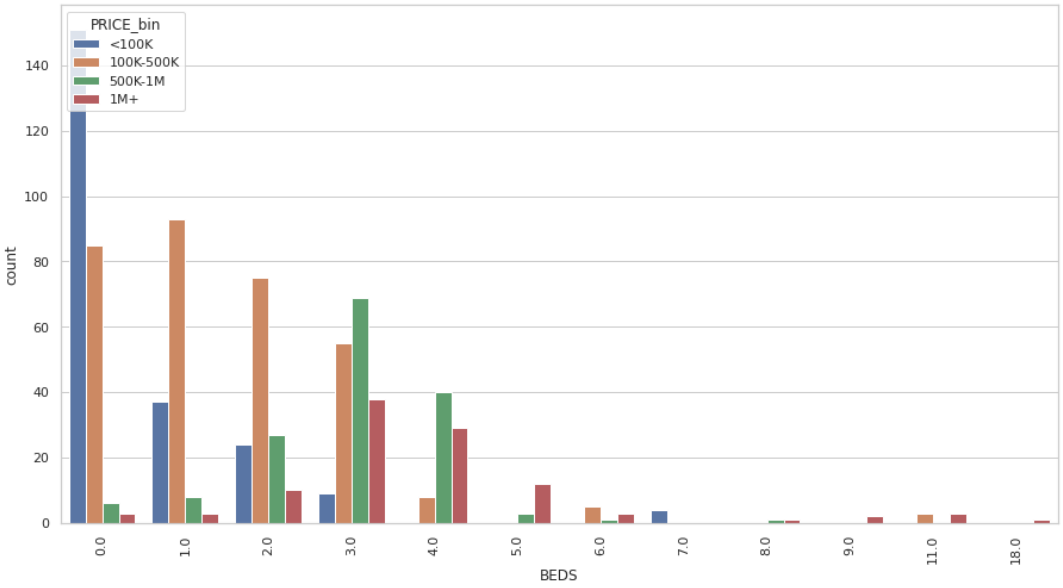
People looking forCondo/ Co-op the most frequent range is between $100K-500K.

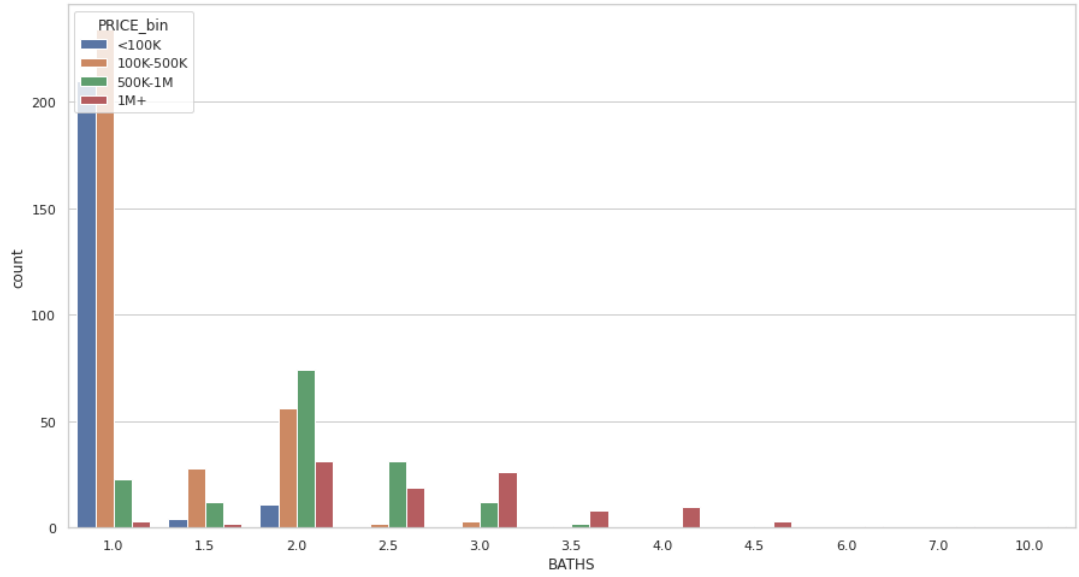
People looking at single family residential are looking for any range.

Number of Bedrooms/ Bathrooms vs Price

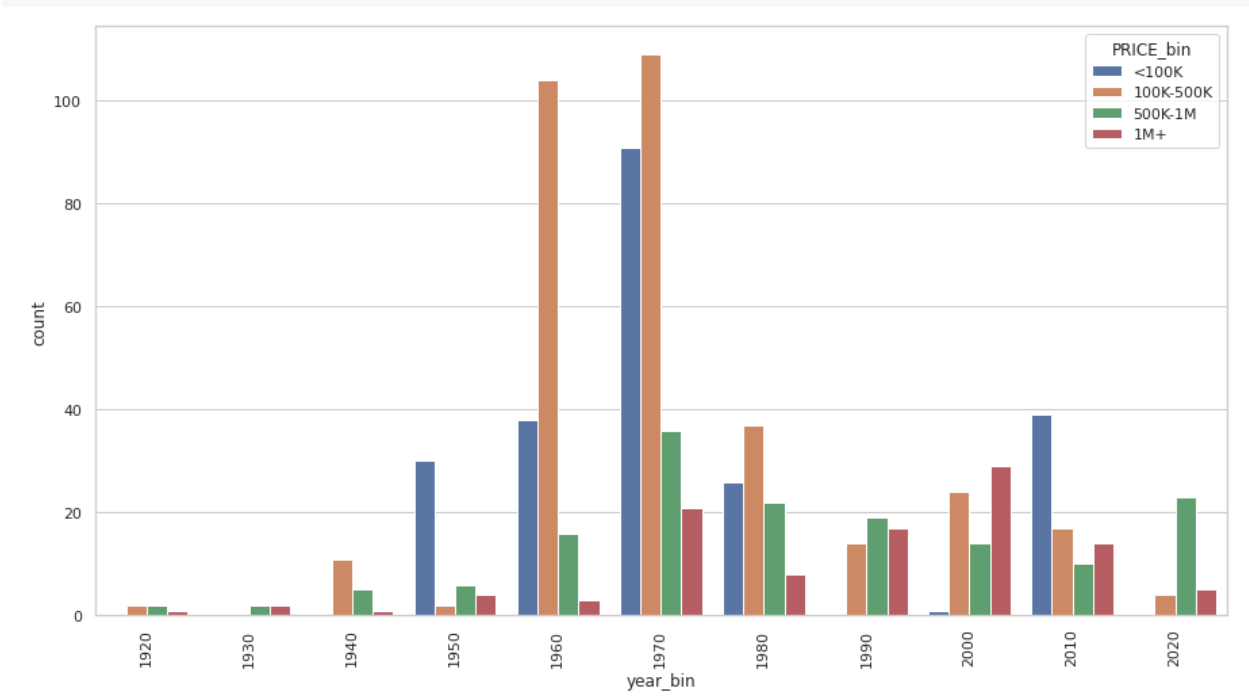
As the number of bedrooms and bathrooms increases the price range also increases.

If there are more than 2 bathrooms and 4 bedrooms, then the price can be recommended more than $500K.





Price Vs Year Built



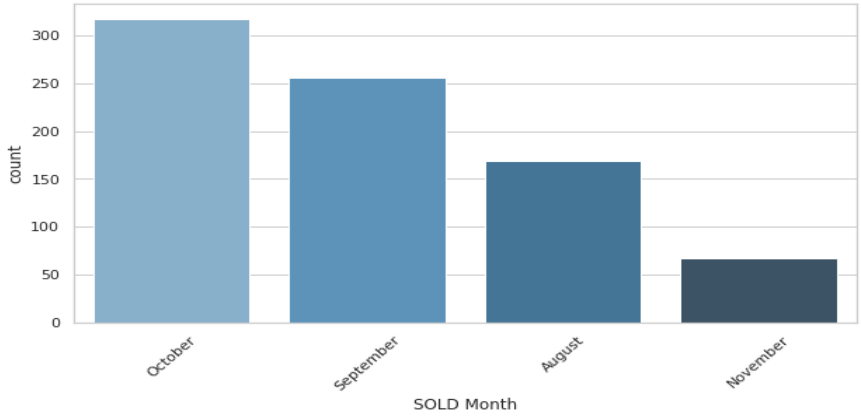
Average price by decade

	year_bin	PRICE
0	1930	1.105000e+06
1	1990	9.357390e+05
2	2000	8.754926e+05
3	2020	8.399239e+05
4	1920	6.587110e+05
5	1980	4.640043e+05
6	1940	4.267059e+05
7	2010	4.048218e+05
8	1970	3.479158e+05
9	1950	3.428571e+05
10	1960	2.475584e+05

In 1930 built houses has the highest average selling price compared to others. This can be obvious because old houses have an ancient and historical value as well. But modern houses (built after 1990) has second highest average selling price.

SOLD Month

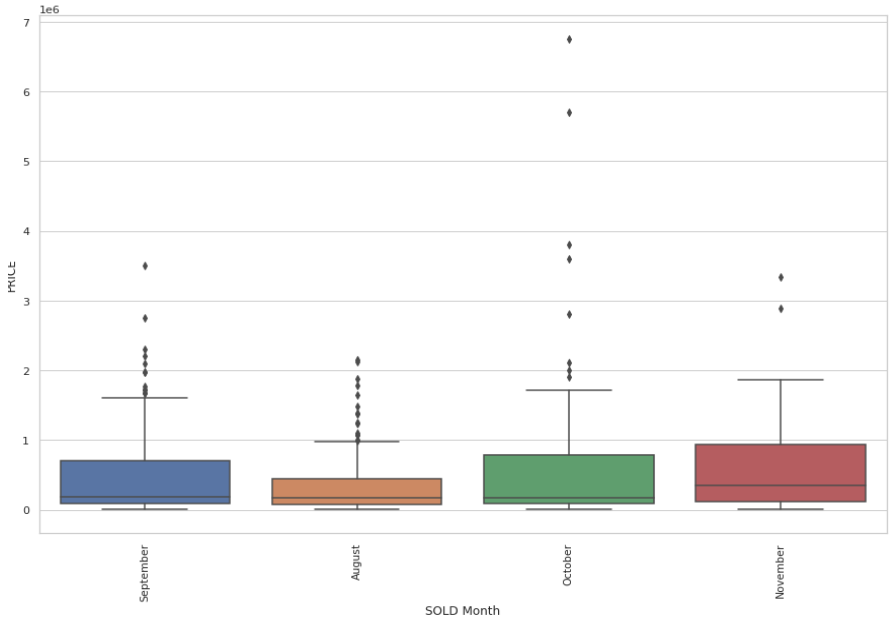
Most sales have been happened at the end of the year; October month has the highest number of sales (Average sale price of October is $476,900). Although November have less number of sales than other months the average price is highest in November, the reason could be the high demand.



Average sold price per month

	SOLD Month	PRICE
0	November	608129.85075
1	October	476900.89590
2	September	454972.25000
3	August	348992.92308

The distribution of prices for each month are shown as below boxplots.

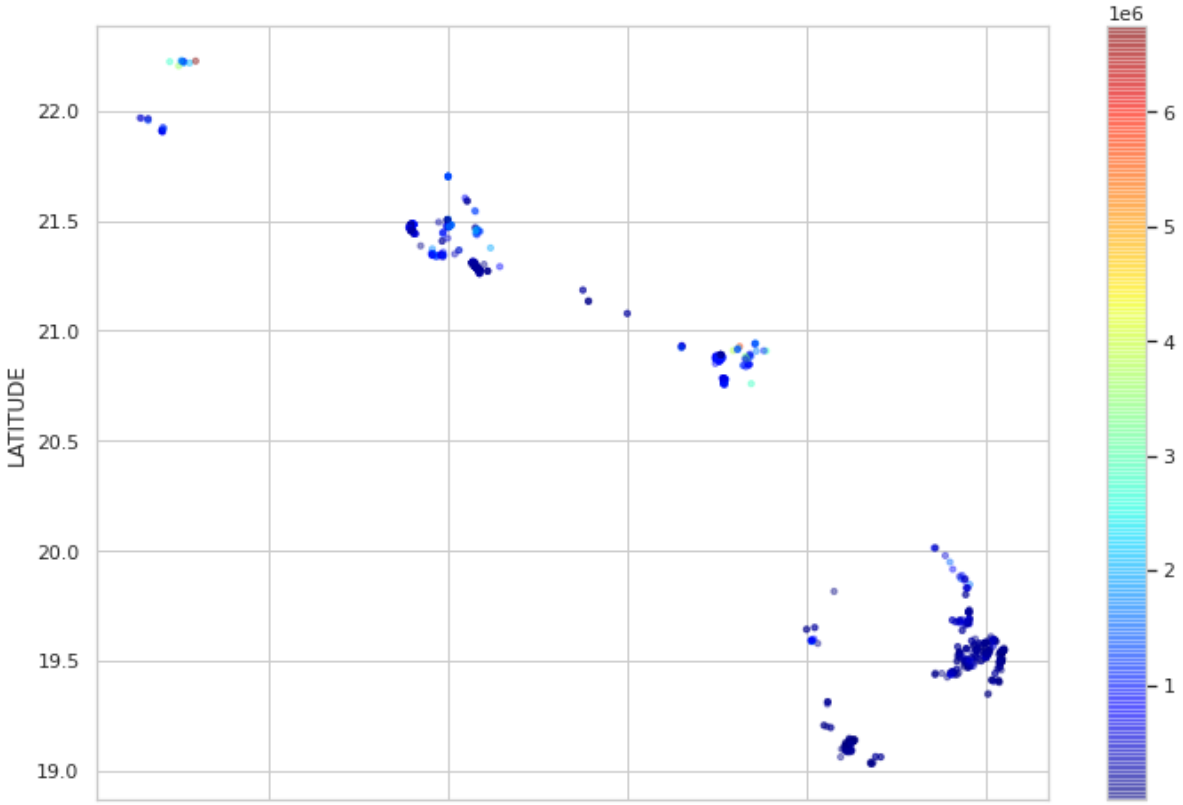


Multivariate Analysis

What is the most expensive location?

The most expensive houses are shown by red points longitude latitude wise.

The graph provides the exact location of the houses.



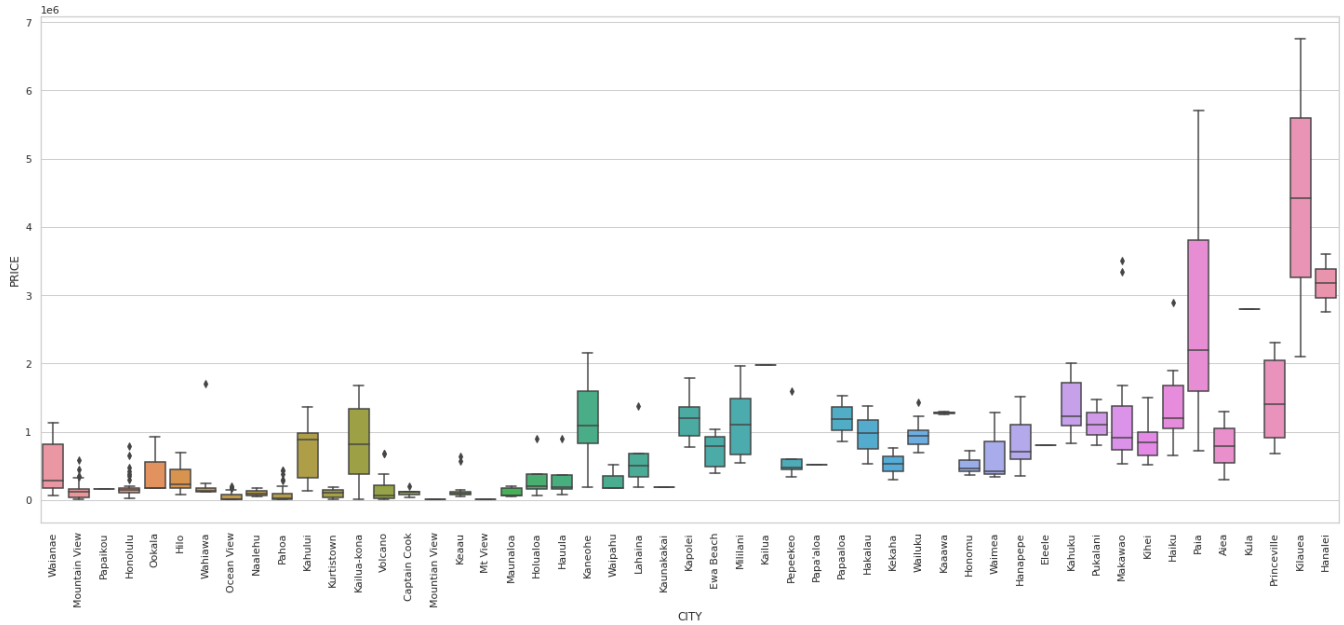
Top 10 city wise average prices

Average prices of each city (best 10 results)

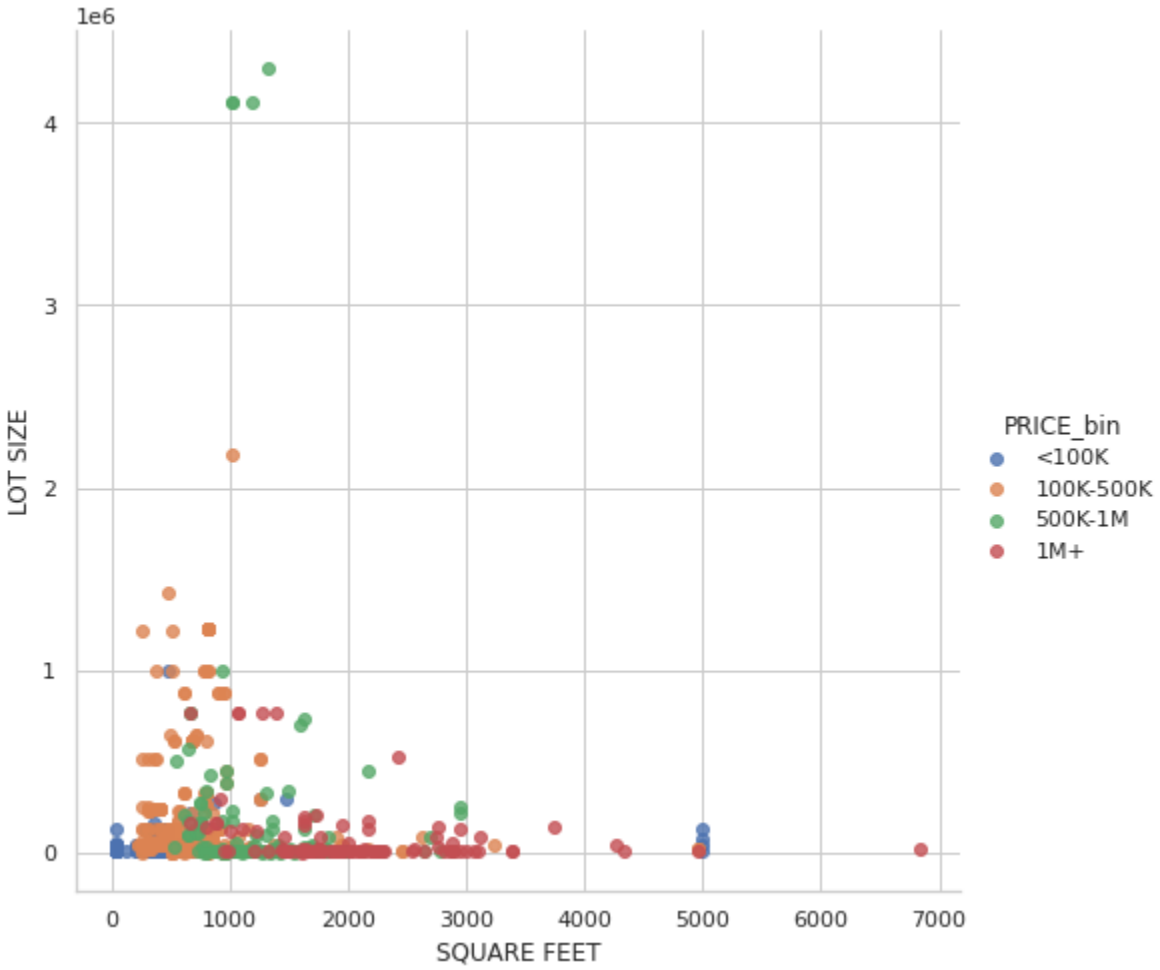
The top three expensive cities are Kilauea, Hanalai and Paia.

	CITY	PRICE
0	Kilauea	4425000.00000
1	Hanalei	3175000.00000
2	Paia	2804000.00000
3	Kula	2800000.00000
4	Kailua	1980000.00000
5	Princeville	1463583.33333
6	Haiku	1427090.81818
7	Kahuku	1379057.00000
8	Makawao	1338416.66667
9	Kaaawa	1277500.00000

The price distribution for each city are as follows. The highest variations are observed in the top three expensive cities which mentioned before.



LOT vs Square FT

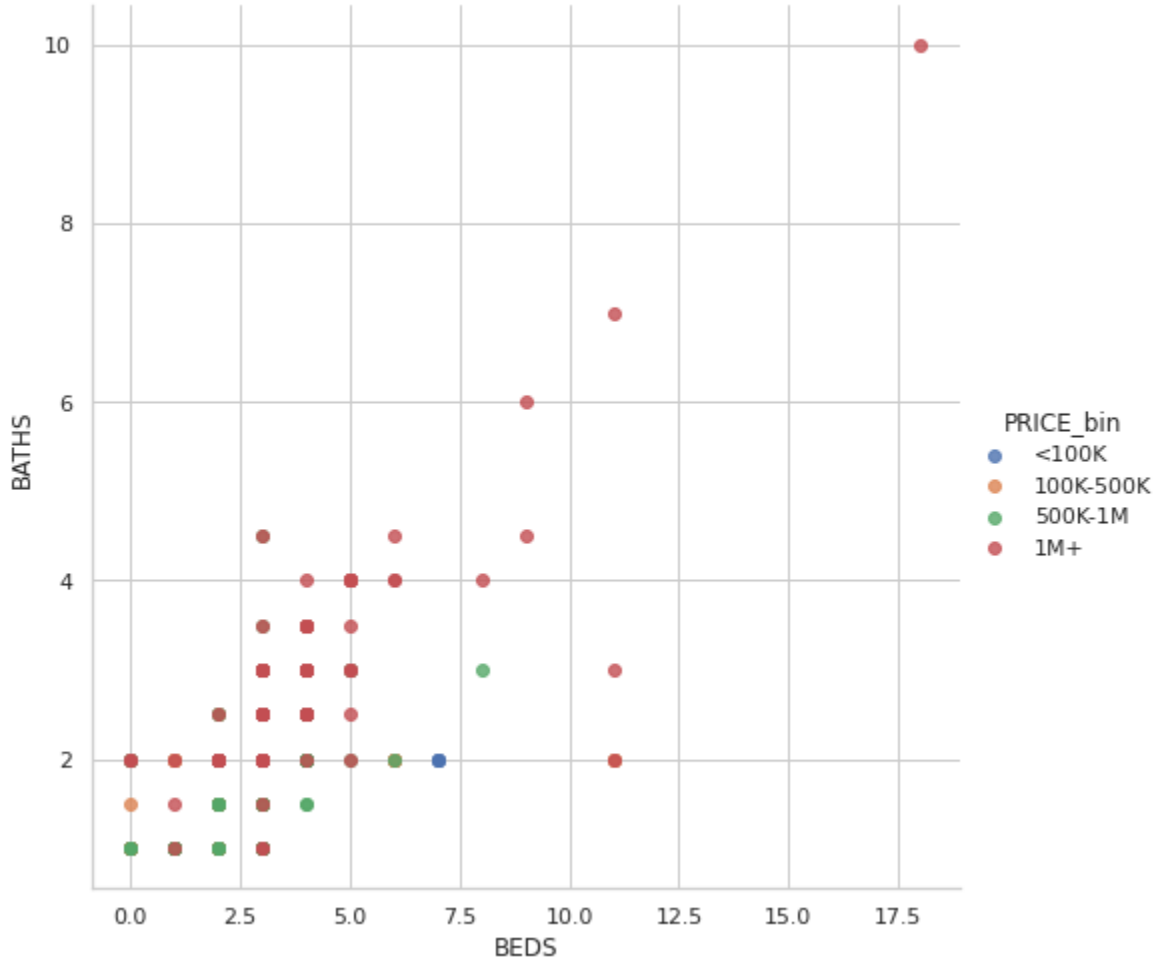


LOT size indicates the size of the piece of land where the property is situated. There is no strong relationship between the LOT size and the square ft.

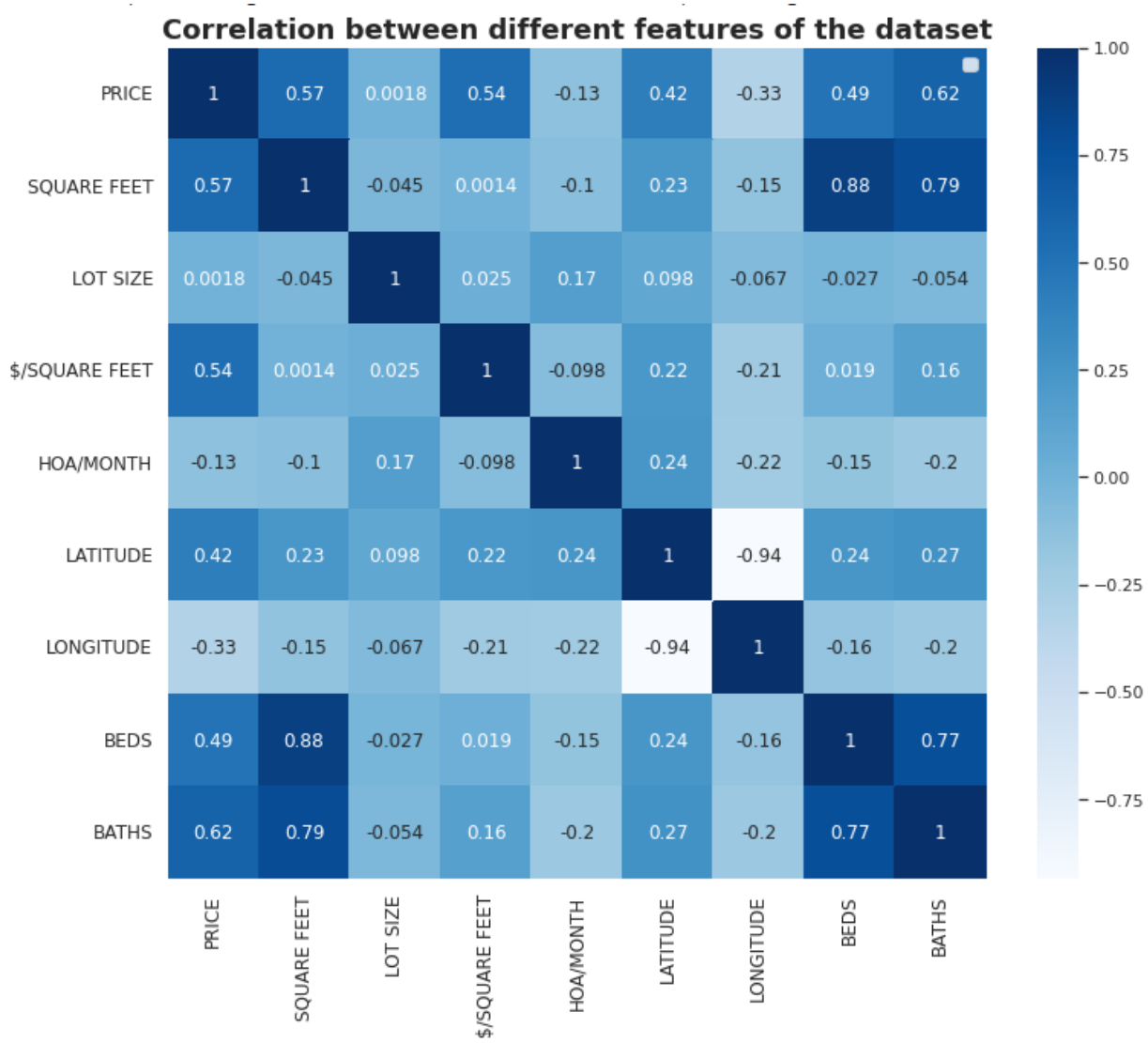
Number of bathrooms/ bedrooms vs Price

There is a strong relationship between the number of bedrooms and bathrooms.

If the number of bedrooms are greater than 8 and the bathrooms are greater than 4 the house can be sold for a price more than $1M.

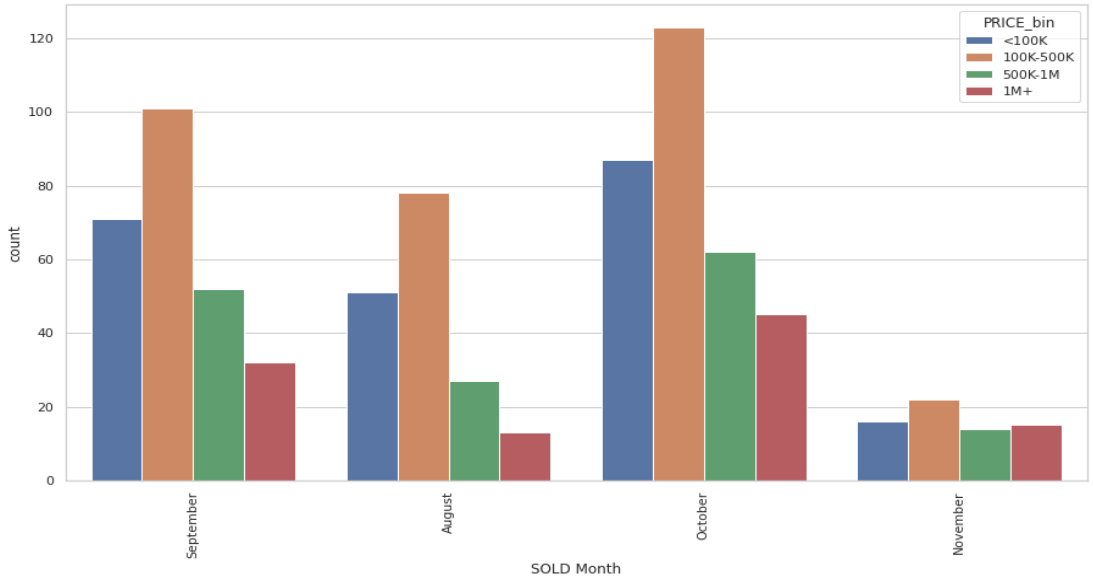


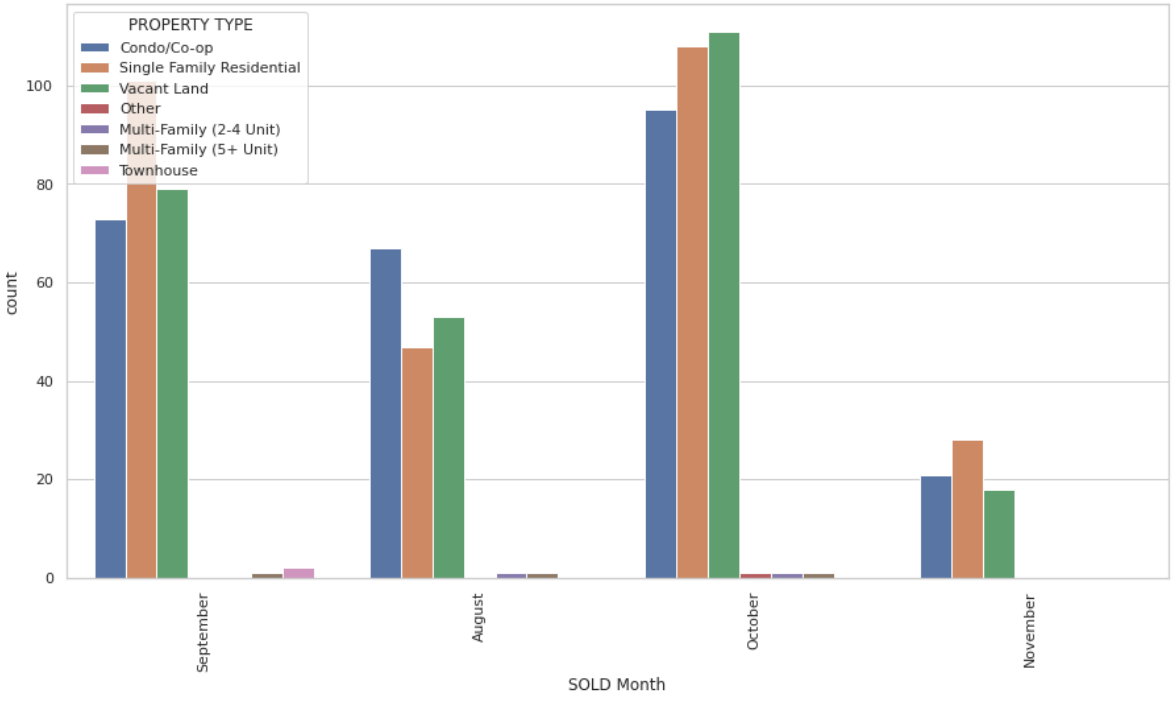
The Heatmap

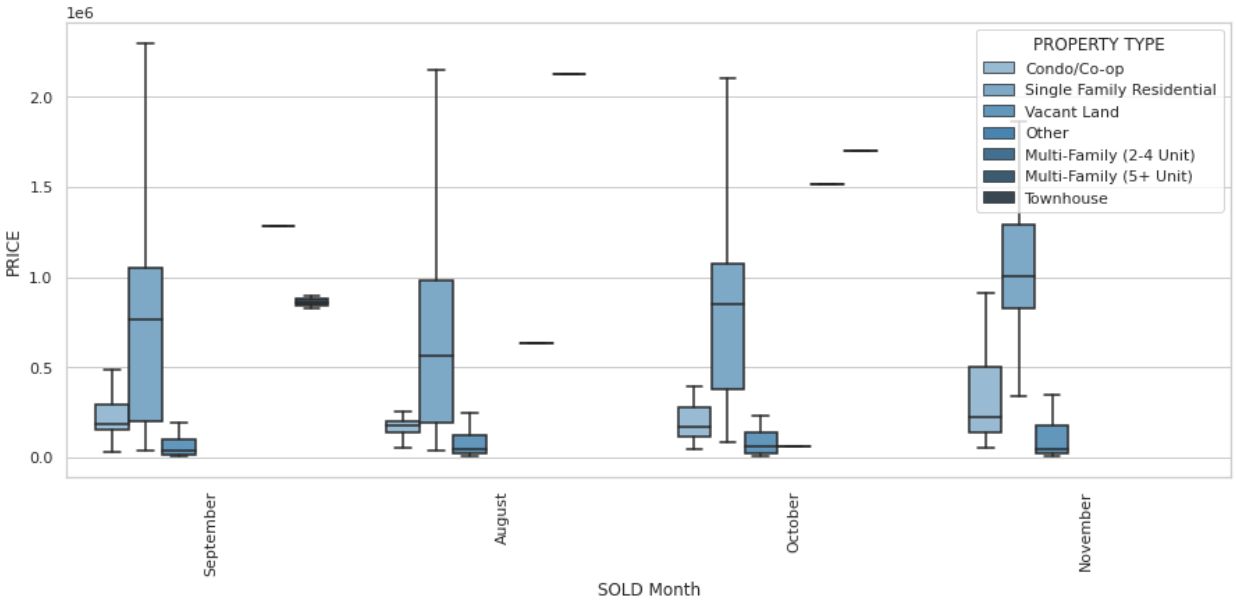


The highest correlated features with the price variable are number of bathrooms, the size of the house, price per square ft, number of bedrooms and latitude. Square ft and number of bedrooms are highly correlated which is obvious.

Other Findings






1e6
2.0
1.5
1.0
0.5
0.0
PRICE
PROPERTY TYPE
Condo/Co-op
Single Family Residential
Vacant Land
Other
Multi-Family (2-4 Unit)
Multi-Family (5+ Unit)
Townhouse
September
August
October
November
SOLD Month

Conclusion

Original dataset consists of 809 rows (deals) and 27 columns.

But there are few unnecessary columns in the dataset (SALE TYPE, STATE OR PROVINCE, FAVORITE, INTERESTED, STATUS, DAYS ON MARKET, NEXT OPEN HOUSE START TIME, NEXT OPEN HOUSE END TIME).

If your real estate has below characteristics, you have a better chance to sell your property. These figures are based on the average values and the counts of the attributes.

- Location : Honolulu, Pahoa and Waianae
- Property Type: Single Family Residential, Vacant Land or Condo/Co-op
- No of bedrooms : 1 - 3
- No of bathrooms : Should be have at least one bathroom and ideally 1 - 2 bathrooms
- HOA per month : At around $532
- Price per sq : $516
- Square feet of the house : 1128 square ft
- Built year : 1960 - 1980 or after 2020
- Starting month of advertising campaign : October
- Price : $454,110
- Price range : 100K - $725K (75%-28% = 47% of selling rate)

You can increase the value of your property by considering following factors:

- The top three expensive cities are Kilauea, Hanalai and Paia.
- If there are more than 2 bathrooms and 4 bedrooms, then the price can be recommended for more than $500K.
- If the number of bedrooms are greater than 8 and the bathrooms are greater than 4 the house can be sold for a price more than $1M.
- In 1930 built houses has the highest average selling price compared to others. But modern houses (built after 1990) has second highest average selling price.
- If the property will sell in November, you can increase the value of the property.
- People looking for vacant land the most frequent range is less than $100, 000.
- People looking for Condo/ Co-op the most frequent range is between $100K-500K.
- People looking at single family residential are looking for any range.

How to demand my property?

If you really need to increase the selling price, you must increase the size and the quality of these features as they have higher correlation to price.

- number of bathrooms
- the size of the house
- price per square ft
- number of bedrooms
- Latitude