Case Study Project - Property Rentals

An overview of the project and business goals

Busines case : Inn the Neighborhood is an online platform that allows people to rent out their properties for short stays. At the moment, only 2% of people who come to the site interested in renting out their homes start to use it.

Projects and business goals : develop a way to predict how much someone could earn from renting their property that could power the application taking on count that we need to avoid estimating prices that are more than 25 dollars off of the actual price, as this may discourage people

Data validation

- first thing we cleaned the data by removing the rows that contains empty columns we got only 14 in our case so dropping it will not affect the outcome of future predictions

then we need to change object types to numeric types because our models works only with numeric variables :

- our problem is a regression type so the price column should be a numeric type so we removed the dollar sign and changed to float

- both bathrooms and bedrooms should be integer type

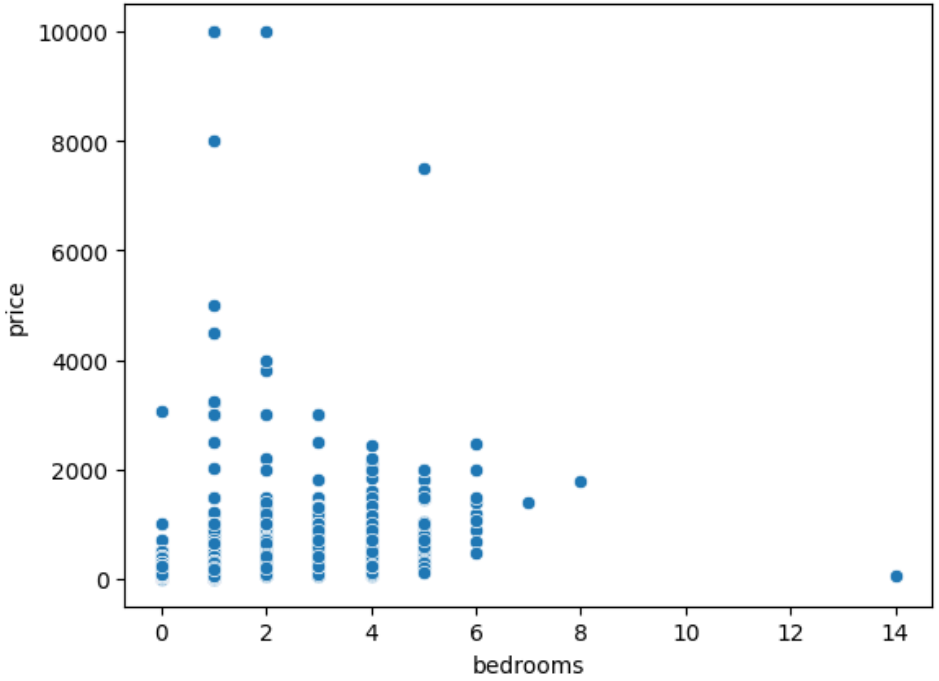
- the property type column need to become a numerical it's not an ordinal column and there are many types so it can’t be changed to dummy variables(0,1,0) that’s why I decided to drop it

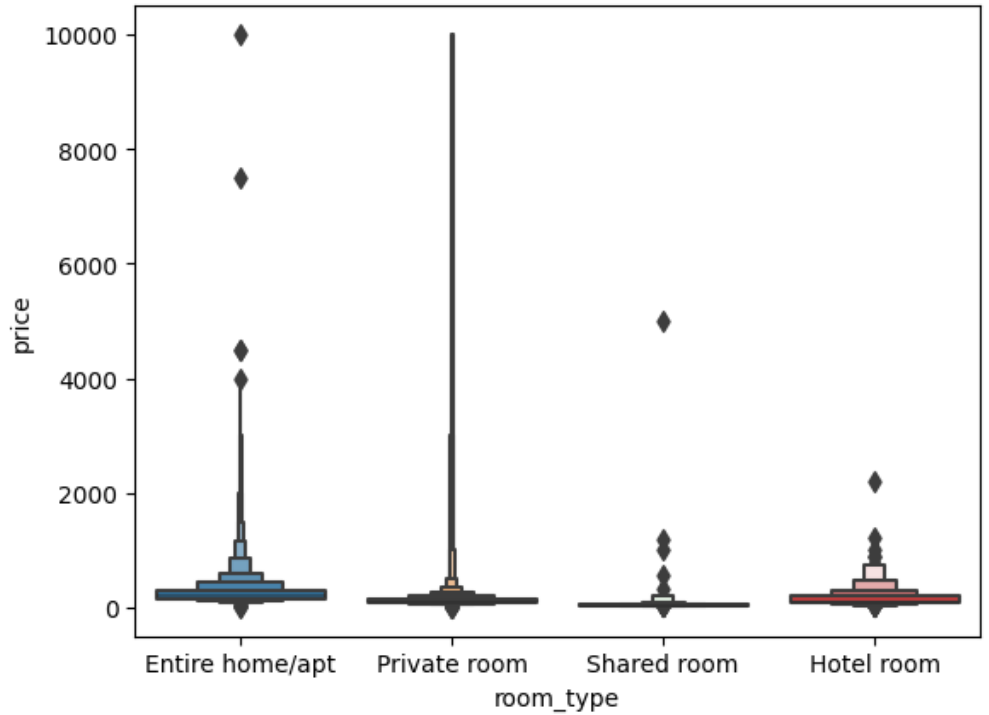
the room type is an ordinal type so we can go from home -> hothel room -> room(shared or private)

```
Entire home/apt    4767
Private room       2886
Shared room         235
Hotel room          207
Name: room_type, dtype: int64
```

```
0    4767
2    2886
3     235
1     207
Name: room_type, dtype: int64
```

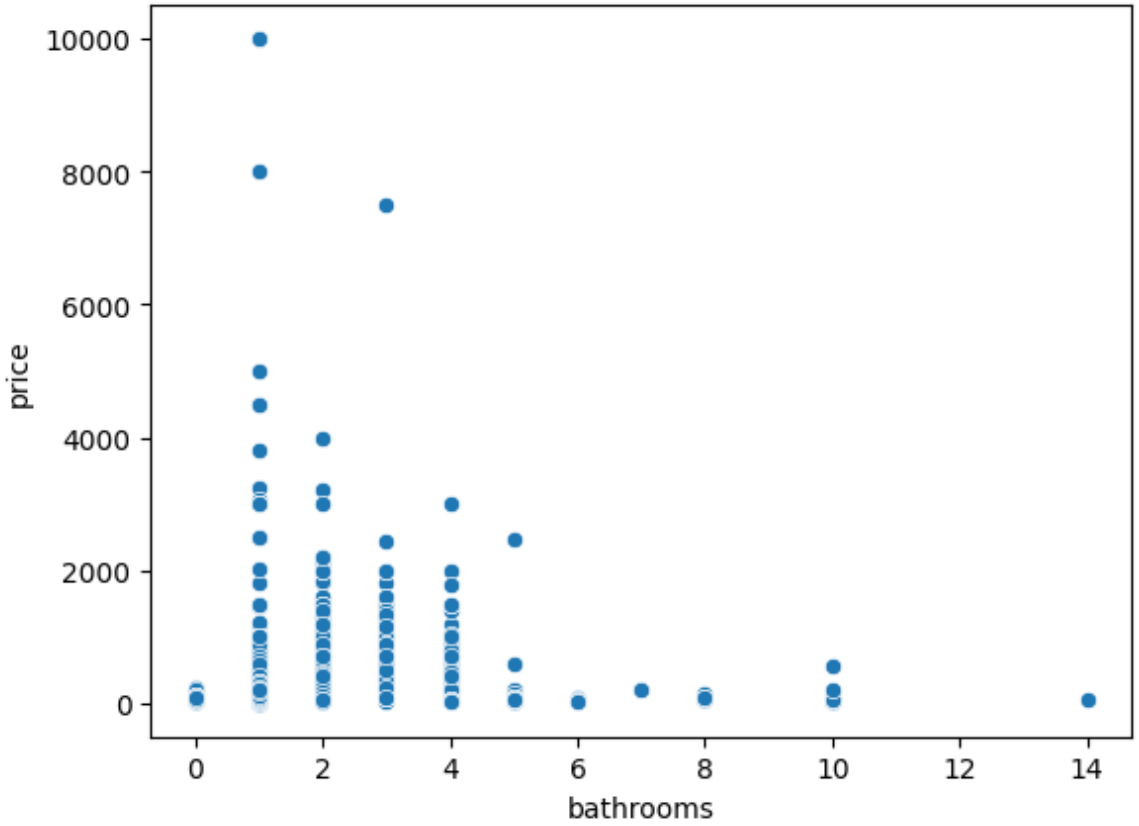
Key findings

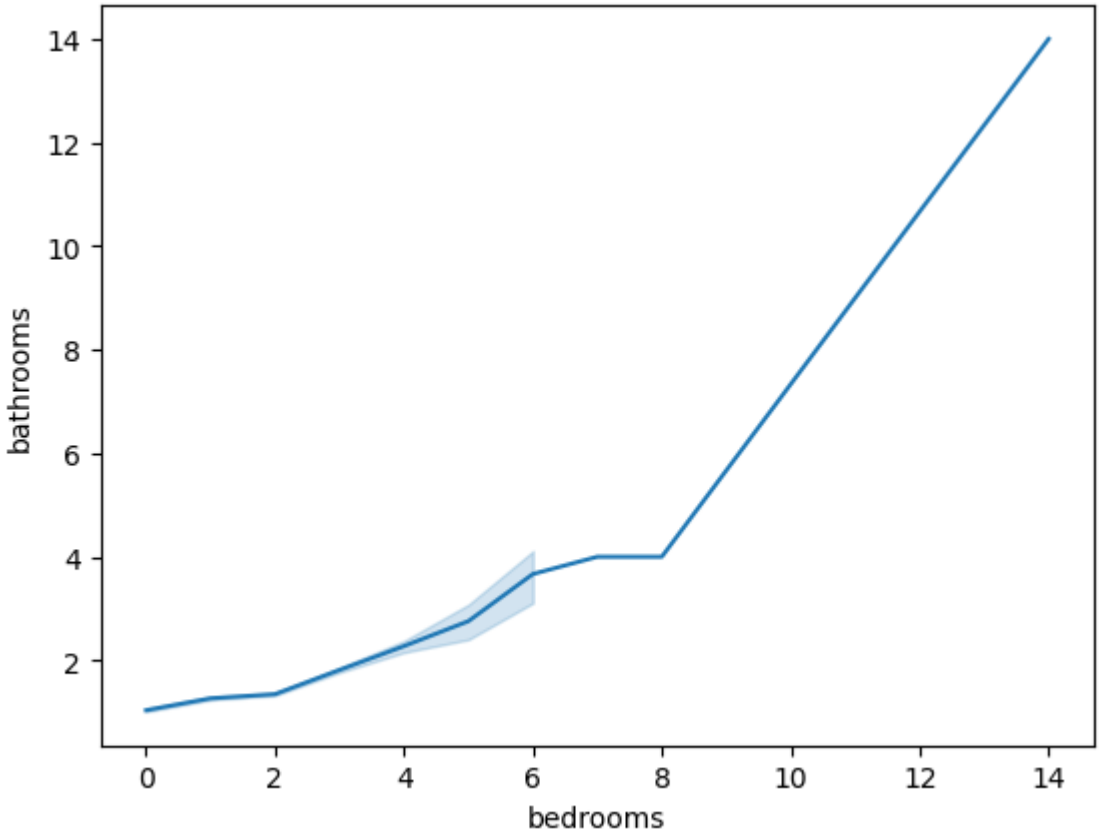




1 : looking at the first graph on the left we can see that prices can go up to 10000 even it includes one bedroom and a very low prices for 14 bedrooms , for the rest of data points the majority prices are all the same for different number of bedrooms presented

2 : for the second graph on the right the private room prices are competing with the ones of entire home/apt and for the shared rom and hotel room the prices are in the normal range





1 : the first graph on the left shows that bathrooms and price have the same correlation like the price and bedrooms

2 : for the second graph we can see the more bedrooms are present we got more bathrooms for individual use

31	114
7	74
32	60
60	40
6	36
90	31
180	18
14	10
365	9
13	4
120	4
21	4
10	4
183	3
360	3
50	3
80	3
18	2
29	2
45	2
58	2
8	2
75	2
40	2
100	1
1000	1
65	1
15	1
12	1
38	1

58	2
8	2
75	2
40	2
100	1
1000	1
65	1
15	1
12	1
38	1
25	1
11	1
150	1
188	1
59	1
100000000	1
85	1
1125	1
28	1
140	1
55	1
110	1
16	1
70	1
200	1
9	1
33	1

These two tables represents the different values of minimum_nights column we can clearly see there are many outliers that go up to 1000 and 10000000 and the majority of values are one month to 6 months

Customer Question

1 : Can you develop a way to predict how much someone could earn from renting their property that could power the application?

I found a way to predict the earnings from the rent using two regression models as you can see in the image

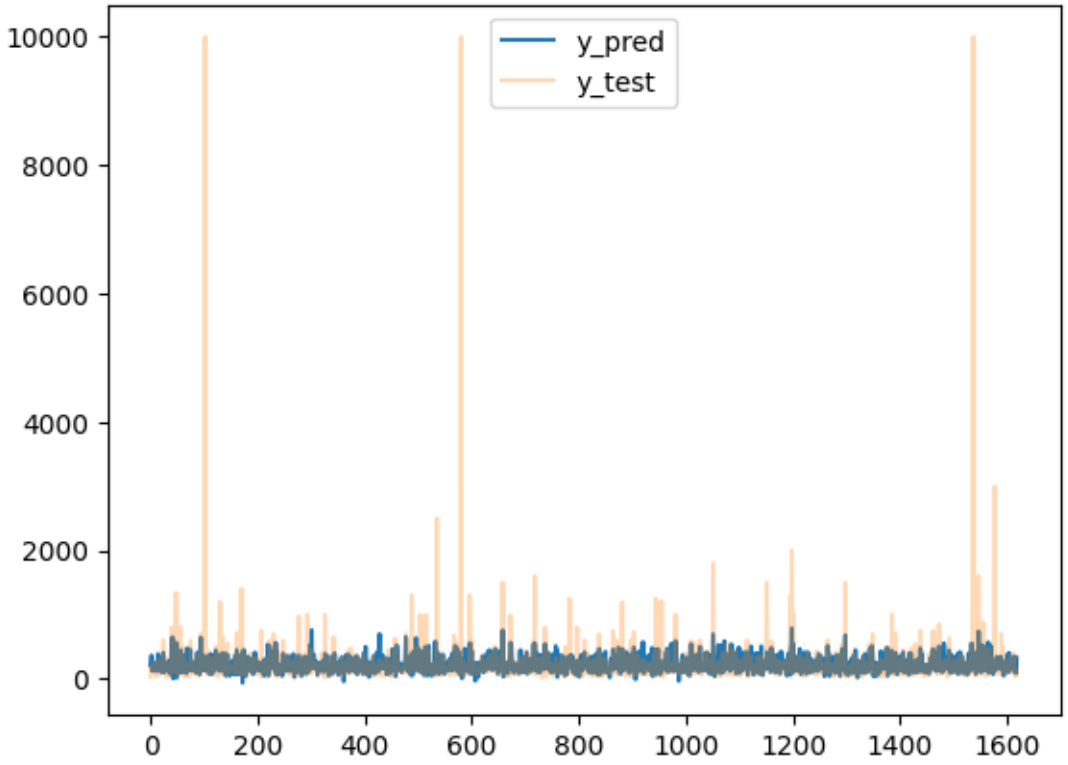
```
knn=KNeighborsRegressor(15)
knn.fit(X_train,y_train)
y_pred_knn=knn.predict(X_test)
print(mae(y_test,y_pred_knn))
```

```
146.24368952028001
```

```
linear=LinearRegression()
linear.fit(X_train,y_train)
y_pred_linear=linear.predict(X_test)
print(mae(y_test,y_pred_linear))
```

```
116.78031844008412
```

Comparing the result to the success criteria




```
diff_moy=(y_pred_linear-y_test)
np.mean(diff_moy)
```

```
0.5703638088226675
```

As we can see in the graph the predicted values were lower than the actual values and the mean difference =0.57
So we passed the success criteria where we avoided estimating prices that are more than 25 dollars off of the actual price

recommendations to the business for future work

1 : we can clearly see there are many outliers that go up to 1000 and 10000000 so we should the limits the values of the minimum nights to 6 months

2 : the prices for private rooms are compared to the prices of an entire home so I think it’s necessary to put a bar limit for the prices of each type of space

3 : put a bar limit for the prices for all the categories , 10000$ is quite big for renting