Python

Airbnb Data Analysis

How Covid 19 affect Covid 19

Data Resource: Airbnb open data source – 04 December 2021, New York

Date Select: 2021/10 –2022/9
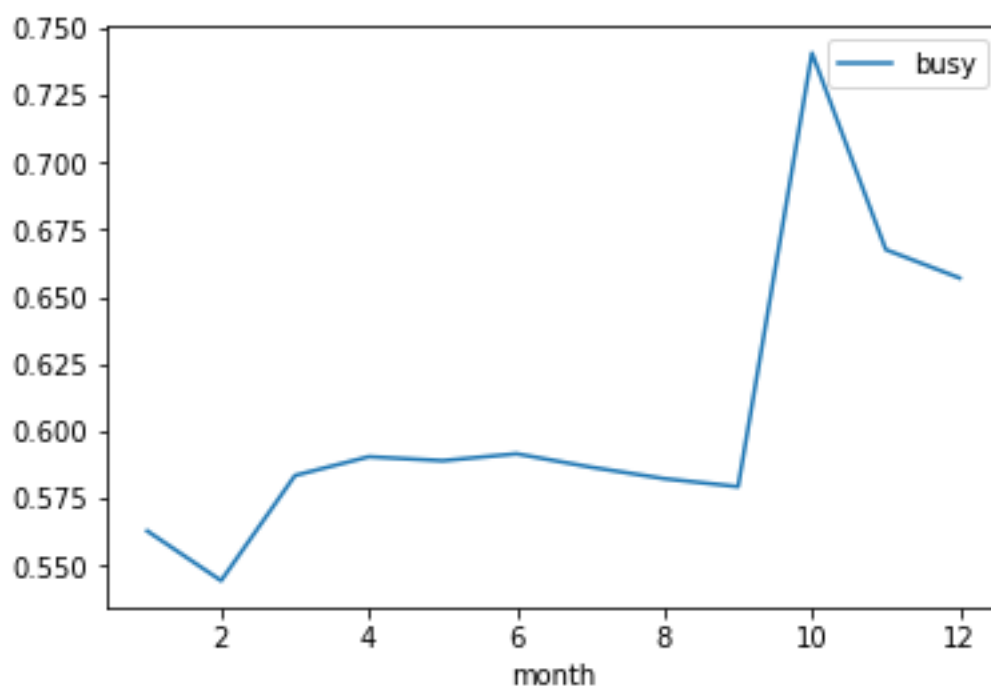
Structure of the data: total 3 csv file

Reviews.csv (all detailed customer review data of London)

Calendar.csv (all booking data of London)

Listing.csv (all Airbnb house of London)


Python: simpleplot.py

Analysis 1: Find the booking situation in different month (plot chart)

If there is already book, the busy is 1, if available for booking, the busy value is 0. In the 2021/12, the covid 19 situation has been increase, therefore the data is match our expectation that there are dropping demand on December 2021. After the covid 19 case decrease, the demand on house rent has been increase in March 2022.

Python: natural language.py

Analysis 2: Find the frequency keyword(Natural Language)
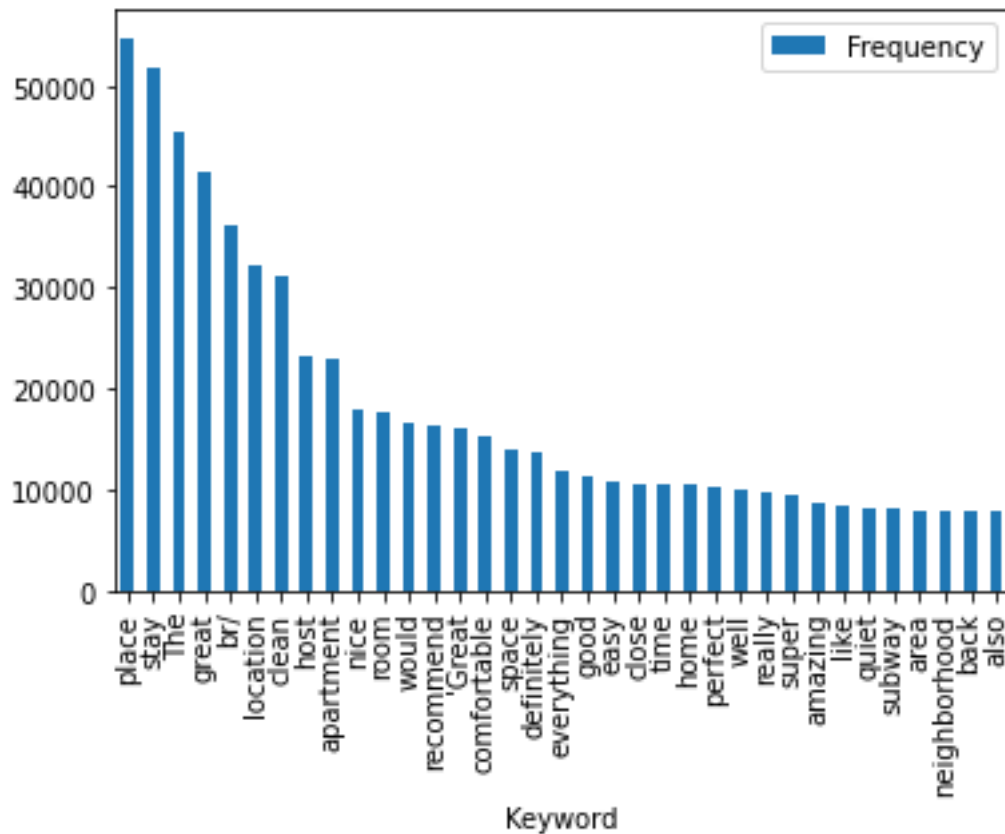
In the result, we can see that most of the people will focus on the location, clean, comfortable, close. If people want to invest household, they should focus on the house venue which is convenience, clean and comfortable design.

SAS

Analysis 3: find out what feature affect the price change(Regression Predictive)

Y = Price

X: date, room_type, neighbourhood, number_of_reviews, review_scores_rating, calculated_host_listings_count, instant_bookable

Hypothesis Testing for Coefficients

Price Prediction Function

$Price = 5.3117 + 0.0530 \times TI\_Instant\_Bookable1 + 0.0251 \times TI\_neighbourhood\_cleansed2 - 0.1526 \times TI\_room\_type1 + 0.1958 \times TI\_room\_type3 + 0.4914 \times TI\_room\_type4 - 0.00393 \times demand - 0.00316 \times host\_total\_listings\_count - 0.00110 \times number\_of\_reviews + 0.00833 \times reviews\_scores\_rating$

At 95% confidence level, we fail reject the null for predictors with P-value > 0.05

Conclusion:
TI_month2, TI_month3, TI_month4, TI_neighbourhood_cleansed 3 are no longer significant and do not help predict the Y variable (Price).
➡️ Remove these predictors from our equation

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	t Value	Pr > |t|
Intercept		1	5.3117	0.0380	139.78	<.0001
TI_instant_bookable1	0	1	0.0530	0.00378	14.03	<.0001
TI_month2	0	1	−0.00380	0.00487	−0.78	0.4359
TI_month3	0	1	−0.00395	0.00489	−0.81	0.4201
TI_month4	0	1	−0.00646	0.00480	−1.35	0.1780
TI_neighbourhood_cleansed2	0	1	0.0251	0.00426	5.89	<.0001
TI_neighbourhood_cleansed3	0	1	0.00541	0.00600	0.90	0.3670
TI_room_type1	0	1	−0.1526	0.00872	−17.51	<.0001
TI_room_type3	0	1	0.1958	0.00849	23.06	<.0001
TI_room_type4	0	1	0.4914	0.0119	41.18	<.0001
demand		1	−0.00393	0.000417	−9.42	<.0001
host_total_listings_count		1	−0.00316	0.000116	−27.35	<.0001
number_of_reviews		1	−0.00110	0.000047	−23.16	<.0001
review_scores_rating		1	0.00833	0.000388	21.49	<.0001

Select the P value have 99% confidence level, therefore TI_month2, TI_month3, TI_month4, TI_neighbourhood_cleansed 3 are no longer significant and do not help predict the Y variable (Price). We need remove these predictors from our equation.

Interpretation

Price on **Instant_bookable1** (instant_bookable: f) will be 0.0530 more expensive than our base Instant_bookable2 (instant_bookable: t), while holding other variables constant.

```
89
90                        Analysis of Maximum Likelihood Estimates
91                                                    Standard
92      Parameter                       DF   Estimate    Error   t Value   Pr > |t|
93
94      Intercept                        1     5.3117   0.0380    139.78     <.0001
        TI_instant_bookable1         0   1     0.0530   0.0378     14.03     <.0001
96      TI_month2                    0   1    -0.00380  0.00487    -0.78     0.4359
97      TI_month3                    0   1    -0.00395  0.00489    -0.81     0.4201
98      TI_month4                    0   1    -0.00646  0.00480    -1.35     0.1780
99      TI_neighbourhood_cleansed2   0   1     0.0251   0.00426     5.89     <.0001
100     TI_neighbourhood_cleansed3   0   1     0.00541  0.00600     0.90     0.3670
101     TI_room_type1                0   1    -0.1526   0.00872   -17.51     <.0001
102     TI_room_type3                0   1     0.1958   0.00849    23.06     <.0001
103     TI_room_type4                0   1     0.4914   0.0119     41.18     <.0001
104     demand                           1    -0.00393  0.000417   -9.42     <.0001
105     host_total_listings_count        1    -0.00316  0.000116  -27.35     <.0001
106     number_of_reviews                1    -0.00110  0.000047  -23.16     <.0001
107     review_scores_rating             1     0.00833  0.000388   21.49     <.0001
108
109
```

The significant attribute we base on the estimate value to do the analysis, if the estimate value is positive, then this attribute will increase the price. If the estimate value is negative, then this attribute will decrease the price.

Tableau:Airbnb.twb
Analysis 4: The price in different area

Airbnb location and price

© 2022 Mapbox © OpenStreetMap
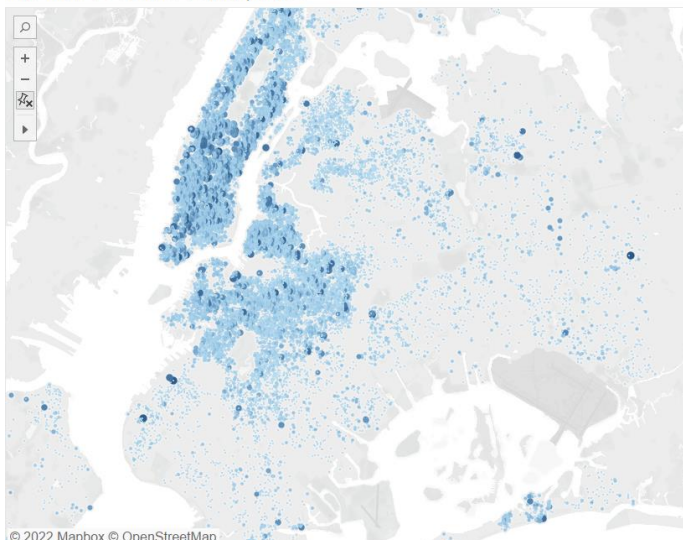
Use Tableau to generate the map, then set the price attribute to colour and size element. The biggest size of point, the price of house is expensive. The darker colour of the point, the price of house more expensive.