Neighborhood Safety and Airbnb Short-term Housing Price in Boston

Irene Ham Liu, Yujia He and Yijie Zhi | UEP 293-6 Urban Analytics and Visualization

Tufts University. December 2019

Abstract:

The rapid-growth of sharing economy platforms like Airbnb and the prevailing of crime and safety issues, emphasize the importance of re-evaluating business policies around user protection and crime transparency. This study investigates the geographical correlation between price the price distribution through of the crime rates in the city of Boston, MA, USA and how they affect pricing for short-term and long-term housing. The linear regression models on price resulting from stepwise modelling evidences that short-term and long-term housing are affected by different types of crimes, the only two in common are vandalism and drug violence; which affect them in the same level of significance and trend. These results are confirmed through spatial analysis done in CARTO.

Keywords: crime, Airbnb, policy, safety

1. Introduction

In the recent years, a new socioeconomic system has emerged where platforms enable individuals to share properties, such as houses and cars, through fee-based services [1], known as 'Sharing Economy' or also 'Gig Economy'. Sharing economy has particularly made a significant impact on housing and hospitality industries, by allowing individuals and corporations alike to create and work in rapidly growing businesses enabled technological innovations and flexibility in terms of supply (hence the name 'Gig Economy'). These peer-to-peer platforms provide a wide range of different services, such as local accommodation (Airbnb), transportation (rideshare companies such as Uber, Lyft and Easy Taxi), delivery jobs

(Postmates, GrubHub, Uber Eats) and on demand jobs based on specific skills (TaskRabbit, Fiverr and Freelancer).

This study focuses on the local accommodation market, specifically Airbnb. Airbnb is a sharing lodging platform that offers temporary housing in local houses or apartments instead of options like hotels, with unique experiences which includes the accommodation, activities and the experience of living like a local. In other words, it enables the exchange between those individuals with extra spaces and those looking for accommodation in those locations.

In terms of user safety, in the US and other major countries Airbnb does not require any ID other an email address and phone number [2]. This means that the host could potentially be a convicted felon, a registered sex offender or a scam artist.

Price is another incurrent issue in the Airbnb market, as listings in this platform are not subject to any type of pricing policy. Prices they vary depending on location, demand, type of listing, etc. and hosts are free to assign the price to be charged for sharing their property. This makes it difficult to assess the market value based on the listings and the determinants of the price.

The main purpose of this study is to determine and compare which types of crimes in the area have the biggest effect on prices for Airbnb and long-term housing (in this case the unit area of study is terms of census tract) and get a better understanding on how non-residents and locals perceive neighborhood safety. This is done using two methods: for geo-spatial correlations, visualizes through mapping crime incidence locations and the locations of the accommodations along with their respective prices; for price determinants, develops a linear regression model having price as the dependent variable and types of crime only, it does not account for housing properties.

In terms of policy implications, this study seeks to suggest stricter screening policies to allow hosts to post their rooms on the Airbnb platform.

1.1 Literature Review

1.1.1 Spatial correlation of Airbnb Crime and Pricing

Some related research around this topic has been done previously [5,6]. In ‘Explore the Spatial Relationship between Airbnb Rental and Crime’ for the state of Florida [7], Xu et.al. aim to answer the question of whether there is a spatial relationship between the geographical locations of Airbnb rental sites and incidents of criminal activities by applying geographically weighted regression to explore the spatial relationships between the location of Airbnb facilities in Florida in 2017, the type of accommodation and the criminal activities data of the criminal activity counts of 2015 from the Florida Department of Law Enforcement.

1.1.2 Housing price and crime rates relation

In Urban Analytics, a great number of authors have researched and published findings on the relationship between the pricing of property with the crime rate of the location. In ‘Crime and Residential Choice: A Neighborhood Level Analysis of the Impact of Crime on Housing Prices’, George E. Tita, et.al.[8] point out that failure to distinguish among different neighborhood types within the city hinders the impact of crime varies across neighborhood types and thus leading to different housing market outcomes. Furthermore, the difference between the total number of crimes, property crimes and violent crimes may lead to inconsistences (due to underreporting) since the amount of property crimes is usually larger than violent crimes. Some common counterintuitive findings such as higher crimes rate with higher housing price are at least partially a function of the underreporting of crime that plagues official crime statistics, which happen most commonly in low-income neighborhoods.

Research questions

Question 1: Is there a relationship between the geographical locations of Airbnb rental prices (excluding all housing properties) and incidents of criminal activities?

Question 2: Using a stepwise regression model to determine significance, which types of crime are statistically significant to predict pricing and what effect do they have on it?

2. Data

- Airbnb data: *Inside Airbnb*

Sources: Listings in Boston, Massachusetts, USA

Description: This dataset contains housing and listing information such as room type, number of reviews, number of beds, bedrooms and bathrooms, accommodates etc. for Airbnb housing in 2018 in Boston. It does not provide square footage of the accommodation. The longitude and latitude information are used in this project to determine which census tract it belongs. The price per night is used as the dependent variable in our models.

- Crime Incidence data: *Analyze Boston*

Sources: Crime Incident Report (2018)

Description: This dataset shows the crime incidents in 2018 in Boston with its type, occurrence time (accurate to seconds), location (part, street, longitude and latitude) information etc. We select the geom-information for the same use as above and the crimes type as our independent variables, which will be aggregated on a census tract level.

- Housing data: *Mass GIS*

Sources: Standarized Assessors Parcels

Description: Similar to Airbnb dataset, except it only shows the detailed address down to the street and the assessed building for long term housing price in 2016 in Boston. It is necessary to transform the location from street to census tract for our use. Furthermore, it does have information about rooms nor host, instead it provides the area of the house.

- Census tracts: *Mass GIS*

Sources: M drive in Tufts GIS Data Server

Description: Dataset about the geo-information such as location, shape and area of each census tract in U.S. Census 2010 in Boston, which is the unit of our analysis and visualization model.

2.1 Data Pre-processing

First, we drop all the data without record (NAN) and record "0" in housing price and then We drop all the data without record (NAN) and record "0" in housing price and then select the location and price of each housing point by using Python.

Then using geocoding tools (https://geocoding.geo.census.gov/), transform data of street location to location in census tract for the long-term housing dataset.

We select the columns we want in Airbnb short-term housing (price per night and location in longitude and latitude), long-term housing (assessed building value and location in census tract) and crimes dataset (crime type and location in longitude and latitude) by using Python.

For the crimes data, we select several types of crimes (Aggravated Assault, Aircraft, Arson, Assembly or Gathering Violations, Auto Theft, Auto Theft Recovery, Ballistics, Bomb Hoax, Commercial Burglary, Criminal Harassment, Disorderly Conduct, Drug Violation, Explosives, Firearm Violations, HOME INVASION, Harassment, Homicide, Manslaughter, Offenses Against Child / Family, Residential Burglary, Robbery, Search Warrants, Simple Assault, Vandalism and Warrant Arrests) as our observed crimes type by using Python.

After, we geocode the crimes data into point, find the corresponding census tract of these types of crimes points to spatially join them together and then calculate the crimes level of each selected type in 2018 in each census tract in Boston by using ArcMap.

The same geocoding and spatial join operation is done to Airbnb short-term housing data to aggregate them into census tract level and then get the median value of Airbnb short term price per night in each census tract by using ArcMap.

We use attribute join to obtain the long-term housing assessed building value in census tract level and then get the median value of long-term price per night in each census tract by using ArcMap in the same way.

Finally, we merge these data together and then normalize them in order to get the final dataset for our analysis.

A table with variable definitions is provided in the Appendix I of this report.

2.2 Summary Statistics

	price	Sum_Commer	Sum_Firear	Sum_Search	Sum_Simple	Sum_Vandal
count	174.000000	174.000000	174.00000	174.000000	174.000000	174.000000
mean	126.591954	90.712644	88.00000	24.844828	1122.408046	709.798851
std	76.052937	497.760625	506.34201	52.659382	5094.069583	2032.953303
min	35.000000	0.000000	0.00000	0.000000	0.000000	4.000000
25%	76.125000	0.000000	0.00000	0.000000	100.500000	100.000000
50%	105.000000	10.000000	20.00000	0.000000	288.000000	322.000000
75%	162.625000	42.000000	70.00000	29.500000	896.000000	738.500000
max	628.000000	6412.000000	6641.00000	392.000000	65494.000000	25648.000000

Figure 1: Airbnb housing price and some related crimes type

	price	Sum_Assemb	Sum_Bomb_H	Sum_Commer	Sum_Disord	Sum_Drug_V	Sum_Homici	Sum_Offens	Sum_Vandal
count	172.00	172.00	172.00	172.00	172.00	172.00	172.00	172.00	172.00
mean	529028.63	969.22	250.69	2023.04	3483.39	21535.67	261.09	429.66	20117.05
std	944009.59	5098.34	1806.26	7560.54	16850.69	74795.47	727.37	675.84	33215.13
min	169600.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
25%	247512.50	0.00	0.00	0.00	32.75	2049.00	0.00	0.00	4896.00
50%	327650.00	0.00	0.00	553.50	1071.00	7614.50	0.00	0.00	12943.50
75%	469425.00	647.50	0.00	1683.00	2217.00	15186.25	0.00	681.25	24376.25
max	9373200.00	63232.00	23296.00	93184.00	216320.00	908544.00	4660.00	3339.00	372736.00

Figure 2: Long term housing price and some related crimes type

2.2Exploratory Data Analysis

2.2.1 Airbnb

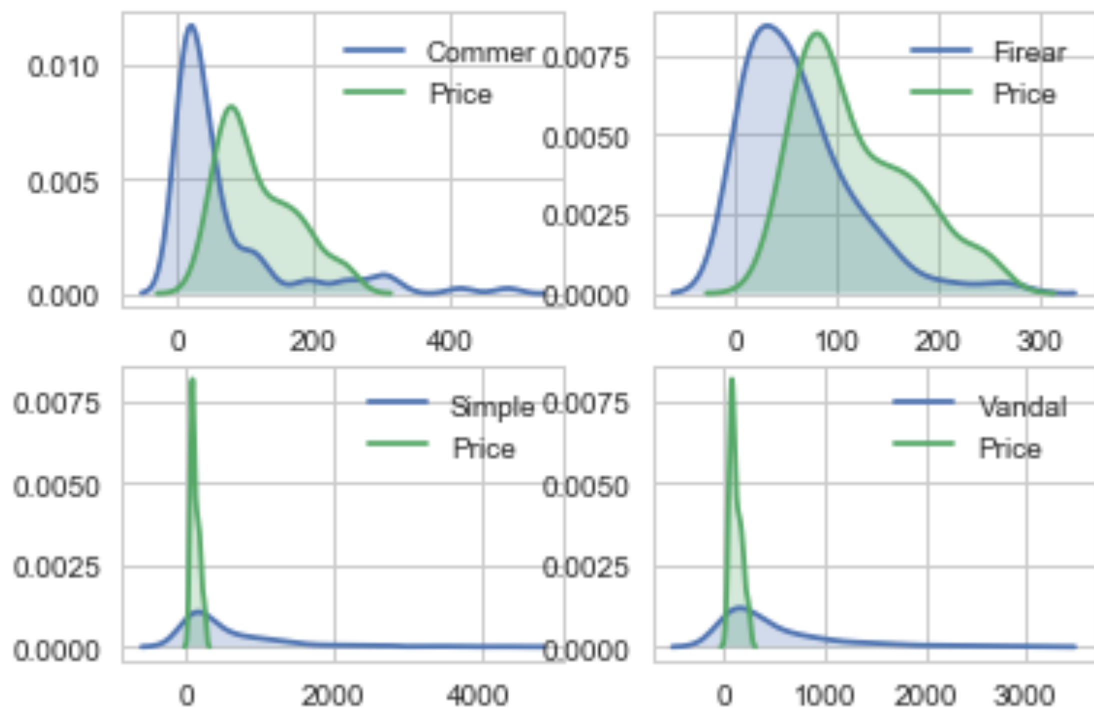

Figure 3: density plot of price and part of crime types

According to above graph, we can see density plot of price and part of crime types are skewness. As one of the assumptions for linear regression is data should follow the normal distribution or approximate normal distribution. Thus, we calculate the skewness value for all variables and select those values larger than 0.8, then take log for those skewness variables to ensure they are in the same normal distribution. However, according to the result of modeling, the variables are not significant enough compared to those variables without taking log, so finally we will use Airbnb price and crimes which are not modifies skewness problem and assume variables distribution are approximate normal.

For visualization of variable correlation analysis, it's better to normalize data if ranges of dependent variable and independent variables are large. Because In this way plot can shows the relationship clearer without points overlap.

As we show in *Figure1* ,the summary statistics for Airbnb&crime dataset, ranges of dependent variable and independent variables are large for example the mean of simple assaults

(Sum_Simple) is 1122.4 while the mean of Airbnb price is 126 which may cause the point overlap in the plot. Thus, we will perform z-score to normalize data. Summary statistics for normalized data is as follow:

	price	Sum_Commer	Sum_Firear	Sum_Search	Sum_Simple	Sum_Vandal
count	174.00	174.00	174.00	174.00	174.00	174.00
mean	0.00	0.00	-0.00	-0.00	-0.00	-0.00
std	1.00	1.00	1.00	1.00	1.00	1.00
min	-1.21	-0.18	-0.17	-0.47	-0.22	-0.35
25%	-0.67	-0.18	-0.17	-0.47	-0.20	-0.30
50%	-0.28	-0.16	-0.13	-0.47	-0.16	-0.19
75%	0.48	-0.10	-0.04	0.09	-0.04	0.01
max	6.61	12.74	12.98	6.99	12.67	12.30

Figure 4: Normalized Airbnb housing price and some related crimes type

2.2.2 Housing:

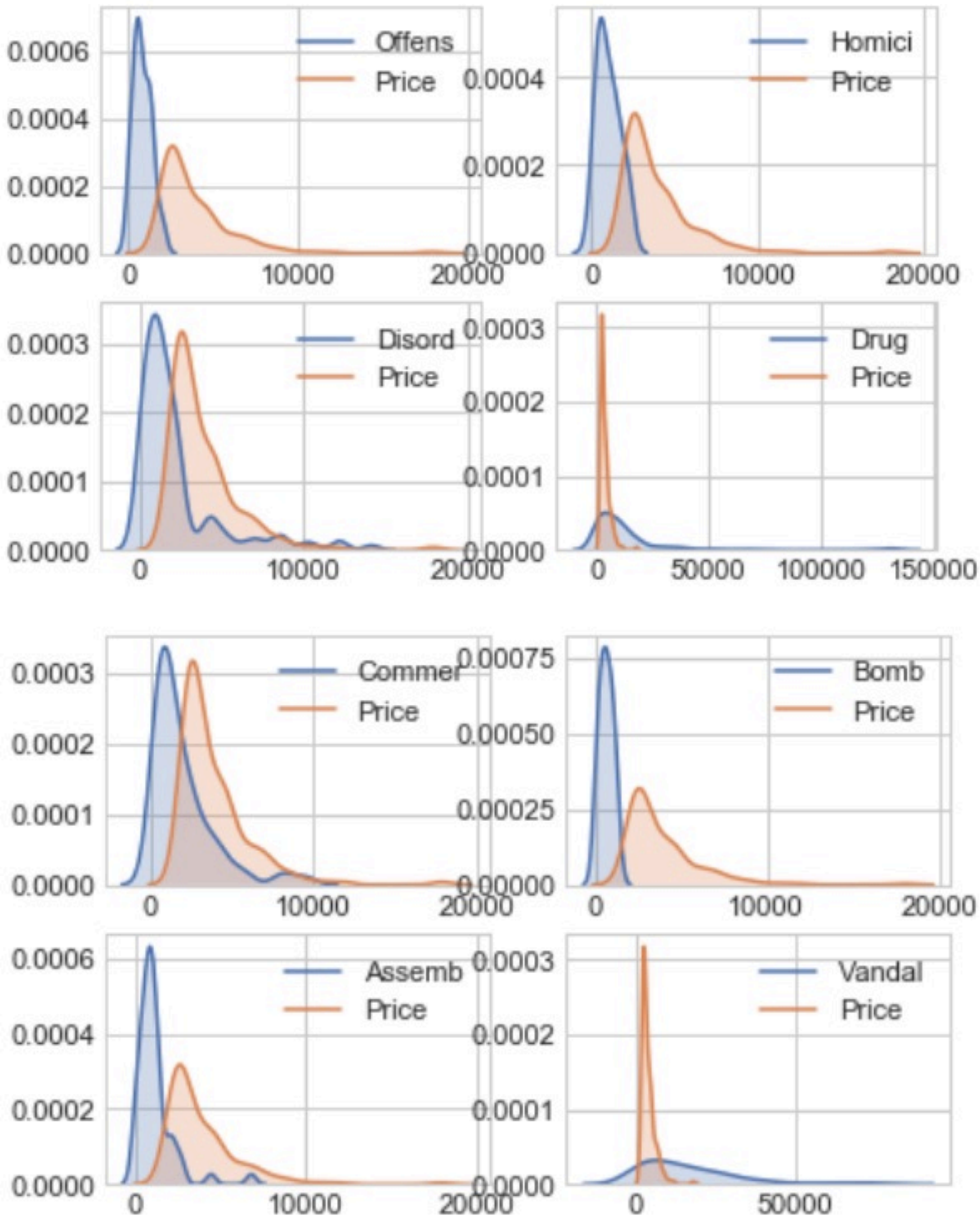

Figure 5: Density of price and part of crime types

According to graph, density of price and part of crime types are skewed, we calculated the skewness for all variables and selected those variables with skewness values are larger than 0.8, then take log to ensure all variables are approximate normal.

As we see (*Figure 2*), the range of price is significantly larger than the scales of crime variables, so it's necessary to normalize data for better visualization. Same as Airbnb data we utilize z-score for normalization and the normalized summary statistics is as following:

	price	Sum_Assemb	Sum_Bomb_H	Sum_Commer	Sum_Disord	Sum_Drug_V	Sum_Homici	Sum_Offens	Sum_Vandal
count	172.00	172.00	172.00	172.00	172.00	172.00	172.00	172.00	172.00
mean	-0.00	-0.00	-0.00	0.00	-0.00	0.00	-0.00	-0.00	-0.00
std	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
min	-0.38	-0.19	-0.14	-0.27	-0.21	-0.29	-0.36	-0.64	-0.61
25%	-0.30	-0.19	-0.14	-0.27	-0.21	-0.26	-0.36	-0.64	-0.46
50%	-0.21	-0.19	-0.14	-0.19	-0.14	-0.19	-0.36	-0.64	-0.22
75%	-0.06	-0.06	-0.14	-0.05	-0.08	-0.09	-0.36	0.37	0.13
max	9.40	12.25	12.80	12.09	12.67	11.89	6.07	4.32	10.65

Figure 6: Normalized long-term price and some related crimes type

3. Methods

For the analysis done in the research, the following hypothesis and assumption were taken into consideration:

- Both long term and Airbnb short term housing price in Boston in recent years will be significantly affected by crimes.
- Not all types of crimes will significantly affect long term and Airbnb short term housing price in Boston in recent years.
- We assume both the owners and buyers of long term and Airbnb short term housing in Boston in recent years pay attention to the level of the same types of crimes in this area.

3.1 Statistical Analysis

In the modeling part, we performed the Stepwise regression method which is an automatic procedure for feature selection in cases where there is a large number of potential explanatory variables. It is a combination of the forward and backward selection techniques and is also a good choice when multicollinearity is a problem. In R, we use stepAIC function to utilize the stepwise regression. We try to keep on minimizing the stepAIC value to come up with the final set of features. “stepAIC” does not necessarily means to improve the model performance, however it is used to simplify the model without impacting much on the performance. So AIC (Akaike Information Criteria.) quantifies the amount of information loss due to this simplification.

For data normalization, we use z-score which is the number of standard deviation from the mean a data point is. In Python I import sklearn.preprocessing module and use the build-in function StandardScaler. It standardizes features by removing the mean and scaling to unit variance. The standard score of a sample x is calculated as:

$$z_i = \frac{x_i - \bar{x}}{s}$$

For data processing, we use pandas and numpy module in Python to aggregate data based on census tract and calculate median housing and Airbnb price as well as the sum of the crime frequency in each census tract. Furthermore, for explore data analysis, we use matplotlib.pyplot and seaborn modules in Python. For modeling, we use the R packages of ‘MASS’ to call stepAIC function, ‘corrplot’to visualize correlation matrix.

3.2 Spatial Analysis

The objective of the spatial analysis done for this project is to visualize the distribution of prices of the Airbnb listings and the assessed value of the long-term housing buildings in the city of Boston. This is to compare the average city price to the one for areas with high index of the specific types of crimes that are indicated as significant from our stepwise model, as a method of verifying the results from the statistical analysis.

The spatial unit of analysis chosen is the census tract, as the Boston area only has about 50 neighborhoods which is not enough for our analysis. In other words, a research area too large may lead to wrong crimes or housing value summary (e.g. if crime incidents were clustered in a small district, the crimes level in this area cannot represent the whole neighborhood).

On the other hand, the problem with census block group is the opposite. Our data may lack of crimes incident records in some census block groups, so if the research was to be done on census block group level, we would need to collect more crimes data throughout several years.

3.2.1 Airbnb

The Boston Airbnb listings point data (colored by price) was mapped on top of the polygon data of the number of incidents for the significant crimes resulted from the stepwise model on a census tract level using CARTO(Figure 16-18) and the average city-price was obtained. To compare the city average price to those in the high crime areas, CARTO spatial analysis was used to intersect the Airbnb data points to the crime type incidence on a census tract level. Using this tool, the price of these Airbnb listings in the intersection were aggregated into an average price of Airbnb in each of these census tracts. The crime type incidence was then filtered to show only census tracts with the highest incidence of the crime type (top 30% values of incidence). The average price of the Airbnb listings in these high-crime areas was compared to the city-average to verify the effect of a subset of these crimes on the price based on the correlations obtained from the stepwise model(Figure 13).

3.2.2 Long-term housing

The same method as the spatial analysis for Airbnb listings was done for the long-term housing data, except the price of the building is the assessed value of it (figure 2). The crime incidence rates and price distribution of the housing buildings for the significant crimes resulted from the model (Figure 9) are presented in Figures 20-21.

4. Results

4.1 Statistical Analysis

4.1.1 Crimes

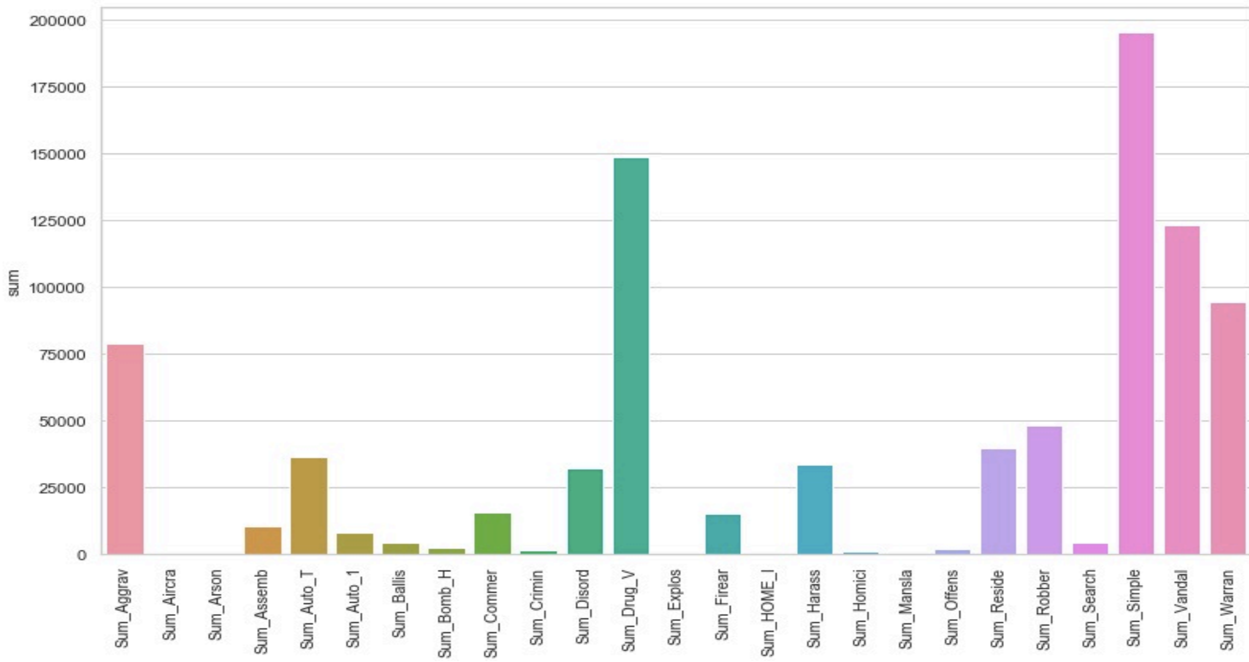

Figure 7: Frequency for each type of crimes

From Figure 7, we know that Simple Assault happens most frequently, next highly happened crime is Drug Violation and the third one is vandalism. While the least happened crimes are Aircraft, Explosives, and Manslaughter. However, we cannot make sure whether the more frequent crimes have more serious impact to both Airbnb and housing price. As there are totally 25 crimes types that may have potential impact to price, we have to selected more significant crimes using the statistical analysis and then utilize the special analysis to compare the difference crimes that affect either Airbnb price or housing price. The scope of analysis would be Boston area and the unit of data would be census tracts.

4.1.2 Housing and Crimes:

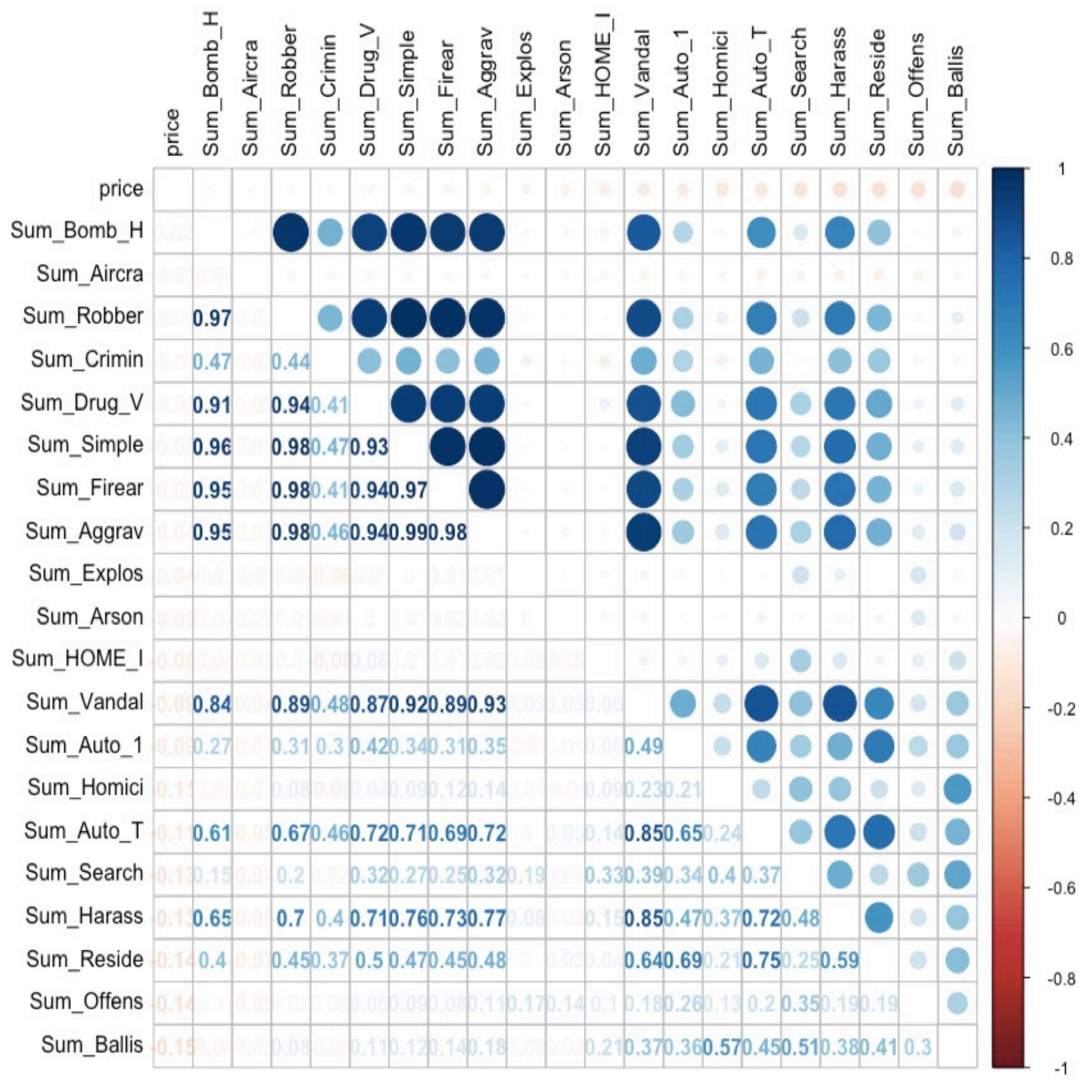

Figure 8: Correlation matrix for crimes and price

For housing data, according to the correlation matrix, including those correlation values of crimes and price are larger than absolute 0.01. We can see some correlation values are larger than 0.9, for example correlation value of Sum_Simple and Sum_Bomb_H is 0.96. Thus, there is multicollinearity problem. We build model by performing stepwise model to solve this problem.

```
Call:
lm(formula = price ~ Sum_Assemb + Sum_Bomb_H + Sum_Commer + Sum_Disord +
    Sum_Drug_V + Sum_Homici + Sum_Offens + Sum_Vandal, data = airbnb)

Residuals:
    Min      1Q  Median      3Q     Max
-1.4701 -0.2895  0.0004  0.2076  2.0677

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 14.35339    0.16889  84.988  < 2e-16 ***
Sum_Assemb   0.03401    0.01184   2.873  0.00461 **
Sum_Bomb_H   0.04366    0.01670   2.614  0.00980 **
Sum_Commer   0.01703    0.01100   1.549  0.12340
Sum_Disord   0.02723    0.01328   2.051  0.04188 *
Sum_Drug_V  -0.06341    0.01983  -3.198  0.00166 **
Sum_Homici  -0.02867    0.01402  -2.046  0.04240 *
Sum_Offens  -0.02669    0.01199  -2.226  0.02737 *
Sum_Vandal  -0.13804    0.02955  -4.672 6.21e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4613 on 163 degrees of freedom
Multiple R-squared:  0.4718,    Adjusted R-squared:  0.4459
F-statistic:  18.2 on 8 and 163 DF,  p-value: < 2.2e-16
```

Figure 9: Regression result of long-term housing price model

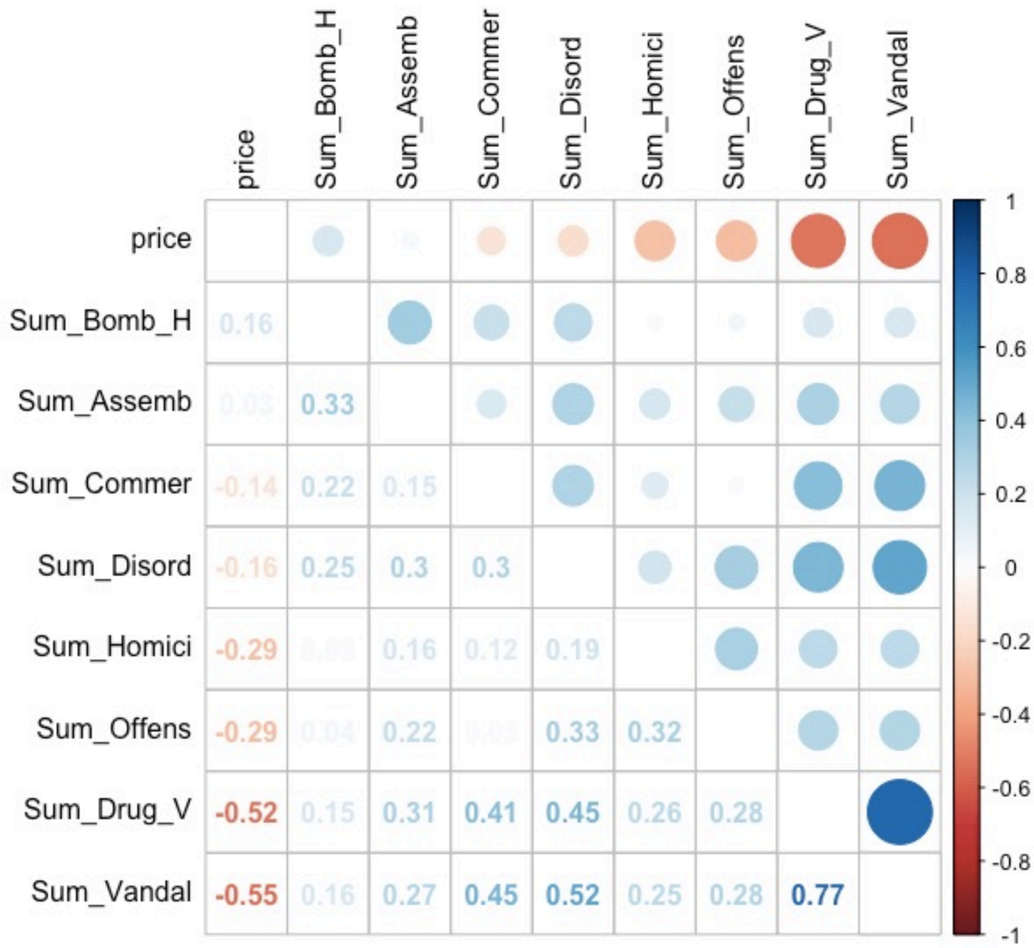

Figure 10: Correlation matrix of selected crims

We can see from Figure 10, multicollinearity is solved largely. Based on the stepwise variables selection result(Figure 9), there are 8 selected crime types which have significant impact to housing price: assembly or gathering, bomb hoax, commercial burglary, disorderly conduct, drug violation, homicide, offenses against, vandalism. And especially, vandalism has the most significant p-value and assembly or gathering, bomb hoax and drug violation have relevant significant impact. While commercial burglary has the least significant p-value. Then we can visualize the relationship between 8 selected crime types and price.


Important Feature VS Housing Price
Price
Commer
Price
Bomb
Price
Assemb
Price
Vandal
1.25
1.00
0.75
0.50
0.25
0.00
−0.25
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2
−0.5
1.5

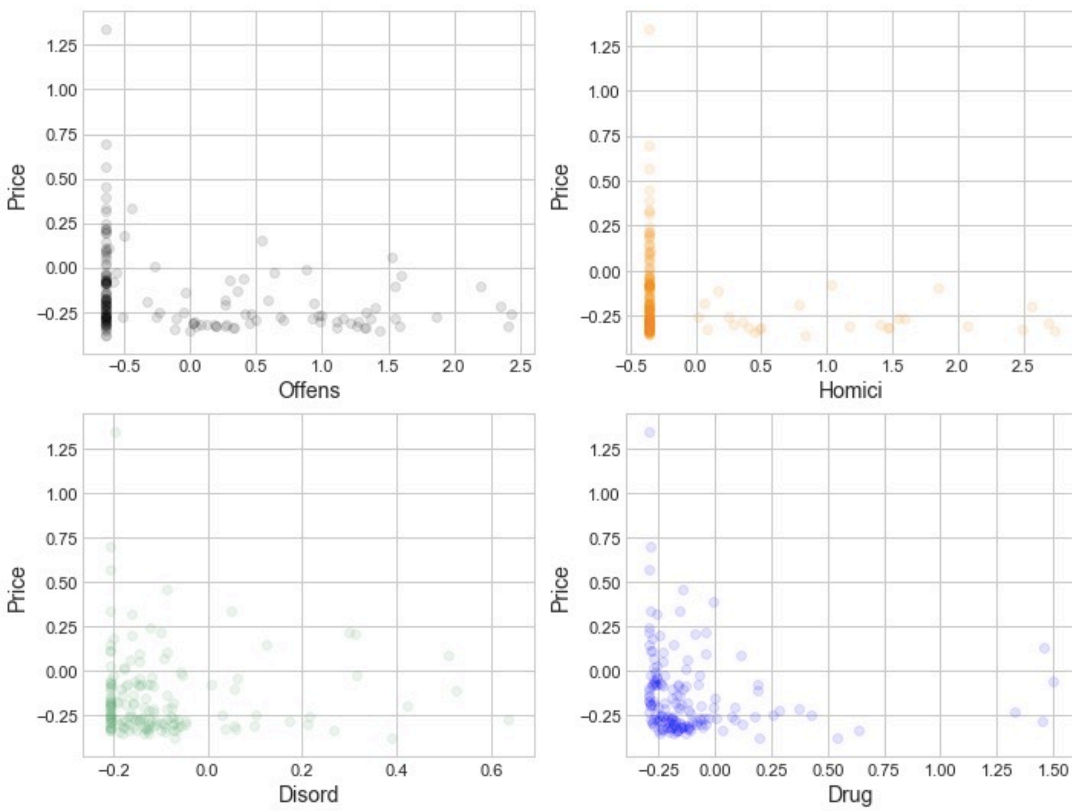

Figure 11: Important features VS long-term housing price

From Figure 11, we can see with increase of vandalism and drug violation, price trend shows apparently decrease in the plots, so we can confirm that vandalism and drug violation affect housing price in Boston area.

However, we notice that estimate coefficients of assembly or gathering, bomb hoax, commercial burglary, disorderly conduct are positive but as assumption, crimes should have negative relationship with price. Because when utilize the stepwise selection, more than one variable is selected, they may have interdependence, which may change the direction of some originally not strong correlations. And according to correlation matrix, assembly or gathering, bomb hoax, commercial burglary, disorderly conduct has very low correlation with price and are not shown in the plot. And we can confirm that vandalism and drug violation affect housing price largely in Boston area.

4.1.3 Airbnb Crimes:

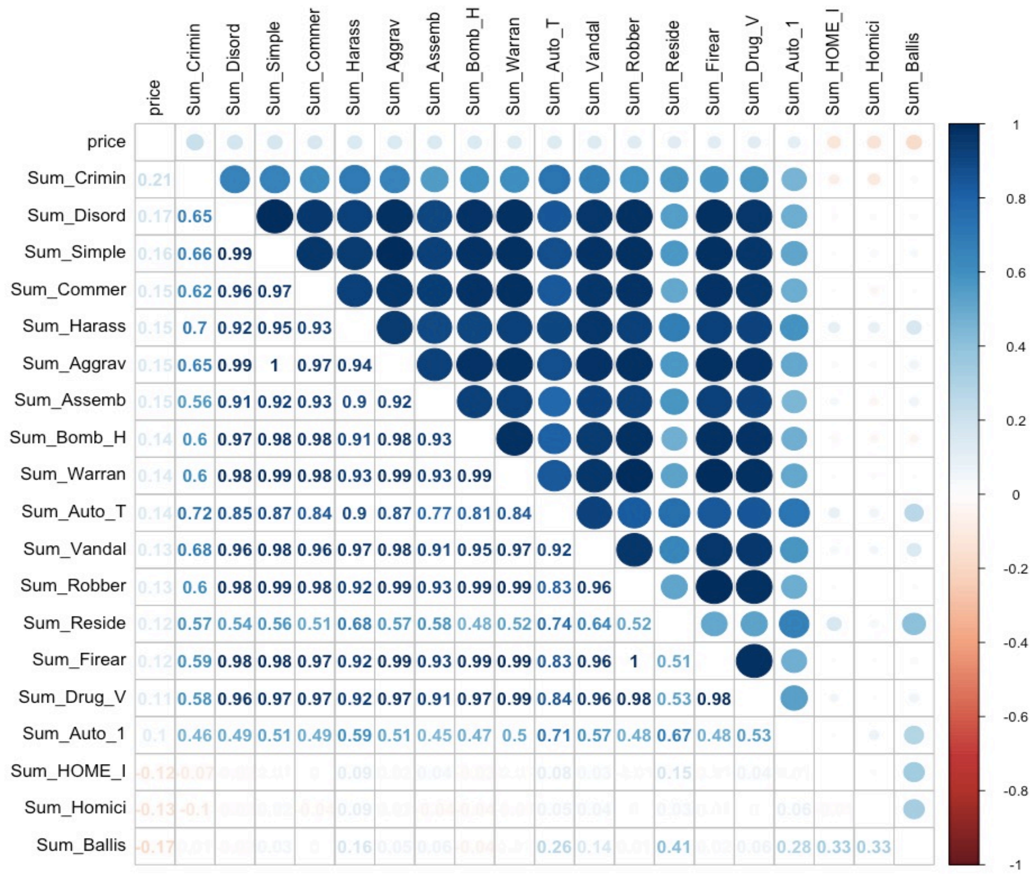

Figure 12: Correlation matrix between each crime type and Airbnb price

According to Figure 12, we can see there are many correlation values are larger than 0.9, thus multicollinearity is a problem. To solve this problem, we will perform variable selection such as stepwise regression model to avoid multicollinearity.

```
Call:
lm(formula = Price ~ Commercial.Burglary + Firearm + Simple.Assults +
    Vandalism + Search.Warrants, data = air_z)

Residuals:
    Min      1Q  Median      3Q     Max
-1.0677 -0.4632 -0.1030  0.3930  2.0260

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)          0.003893   0.076986   0.051  0.95973
Commercial.Burglary  1.234532   0.441085   2.799  0.00580 **
Firearm             -1.421104   0.653551  -2.174  0.03124 *
Simple.Assults       2.573256   0.599313   4.294 3.14e-05 ***
Vandalism           -0.768901   0.346562  -2.219  0.02801 *
Search.Warrants     -0.270780   0.098639  -2.745  0.00679 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6251 on 150 degrees of freedom
  (18 observations deleted due to missingness)
Multiple R-squared:  0.2214,    Adjusted R-squared:  0.1954
F-statistic: 8.531 on 5 and 150 DF,  p-value: 3.944e-07
```

Figure 13: Regression result of Airbnb housing price model

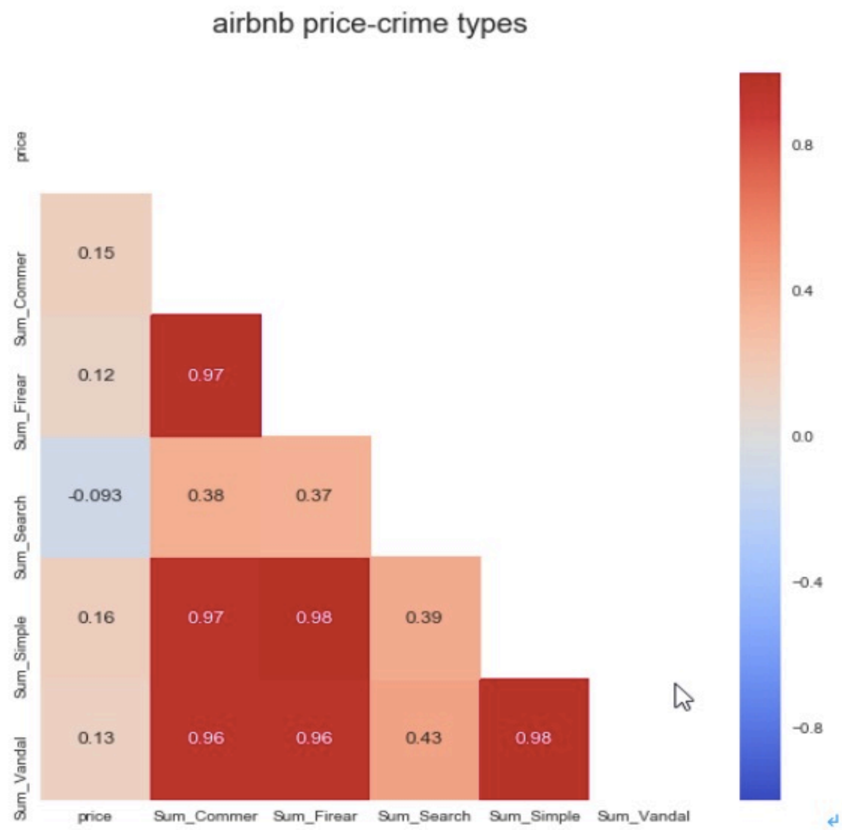

Figure 14: Correlation matrix of selected crimes

we can see from Figure 14, multicollinearity is solved in some extend. Based on the stepwise variables selection result(Figure 13), there are 5 selected crime types which have significant impact to Airbnb price: commercial burglary, firearm, simple assault, vandalism, search warrants. And especially, simple assault has the most significant p-value and commercial burglary, search warrants have relevant significant impact. While firearm and vandalism have the least significant p-value. Then we can visualize the relationship between 8 selected crime types and price. As we state at the housing price analysis part, we can only consider the crime with negative coefficient as for the crime types with positive coefficient may be affected by interdependency. And for Airbnb data, another guess for the reason why simple assaults is selected and have higher and positive coefficient is that it happens most frequently in the neighborhood and for some neighborhood has many Airbnb hosts is traveling hotpot which means more people activities around. And more frequent simple assaults happen will become reasonable.

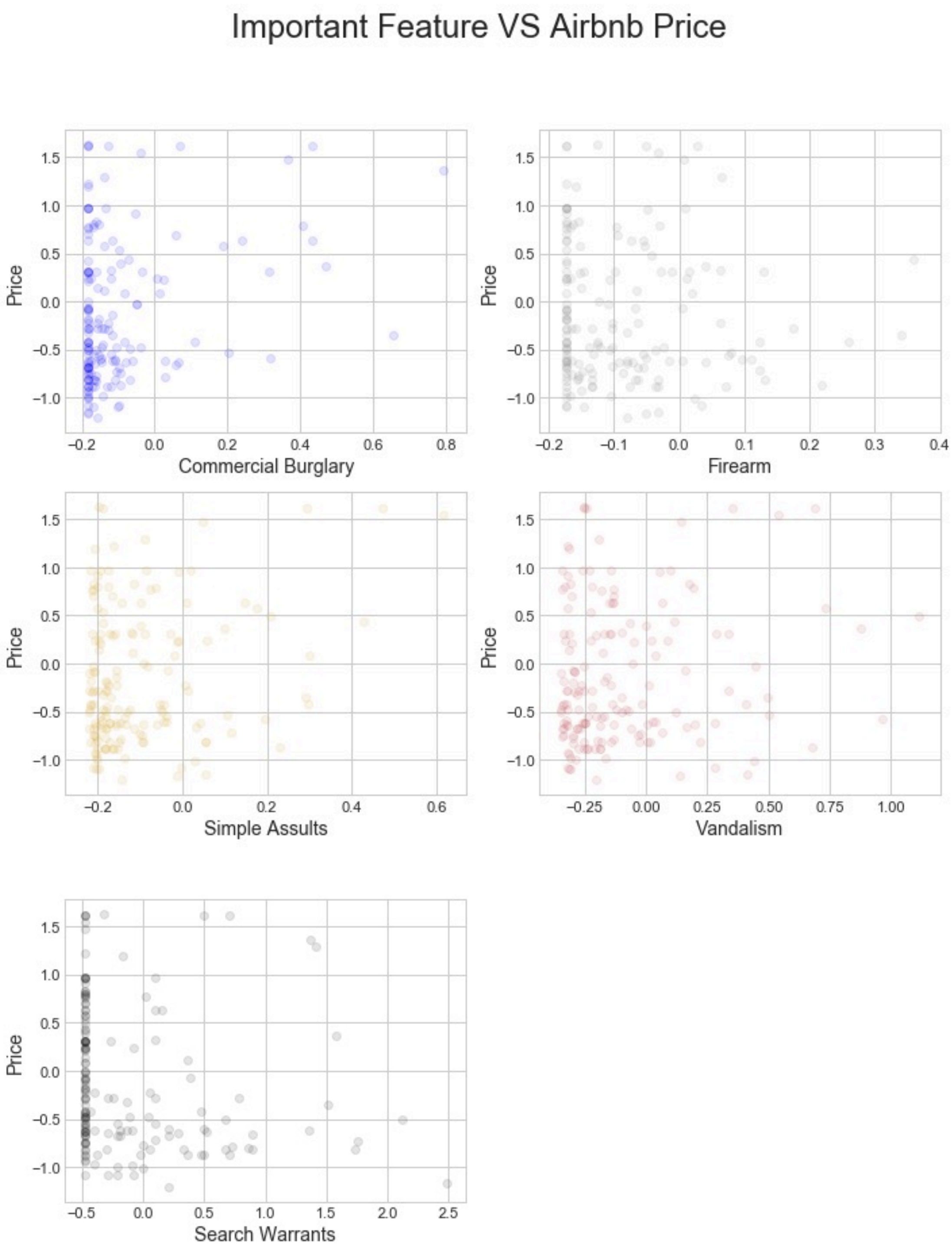

Figure 15: Important features VS Airbnb housing price

From the scatter plot, we can see with increase of search warrants, firearm and vandalism, price trend shows decrease in the plots. Thus, we can confirm Search warrants, firearm and vandalism affect Airbnb price.

4.2 Spatial Analysis

4.2.1 Airbnb

The average price of Airbnb listings in Boston, MA is of $190 a night (Figure 16), while the average price in census tracts with high incidence of search warrants (Figure 17) is $130/night, being lower than the city average matches the predicted effect based on the coefficient obtained (coefficient: -0.2800 , p-value: 0.0157*)in the statistical analysis.

The average price in census tracts with high incidence of simple assaults (Figure 18) is $202/night, being higher than the city average matches the predicted effect based on the coefficient obtained (coefficient: 0.0358, p-value: 1.47×10^{-5}***) in the statistical analysis.

The average price in census tracts with high incidence of firearms crimes (Figure 19) is $178/night, being lower than the city average matches the predicted effect based on the coefficient obtained (coefficient: -0.3269, p-value: 7.55×10^{-6}***) in the statistical analysis.

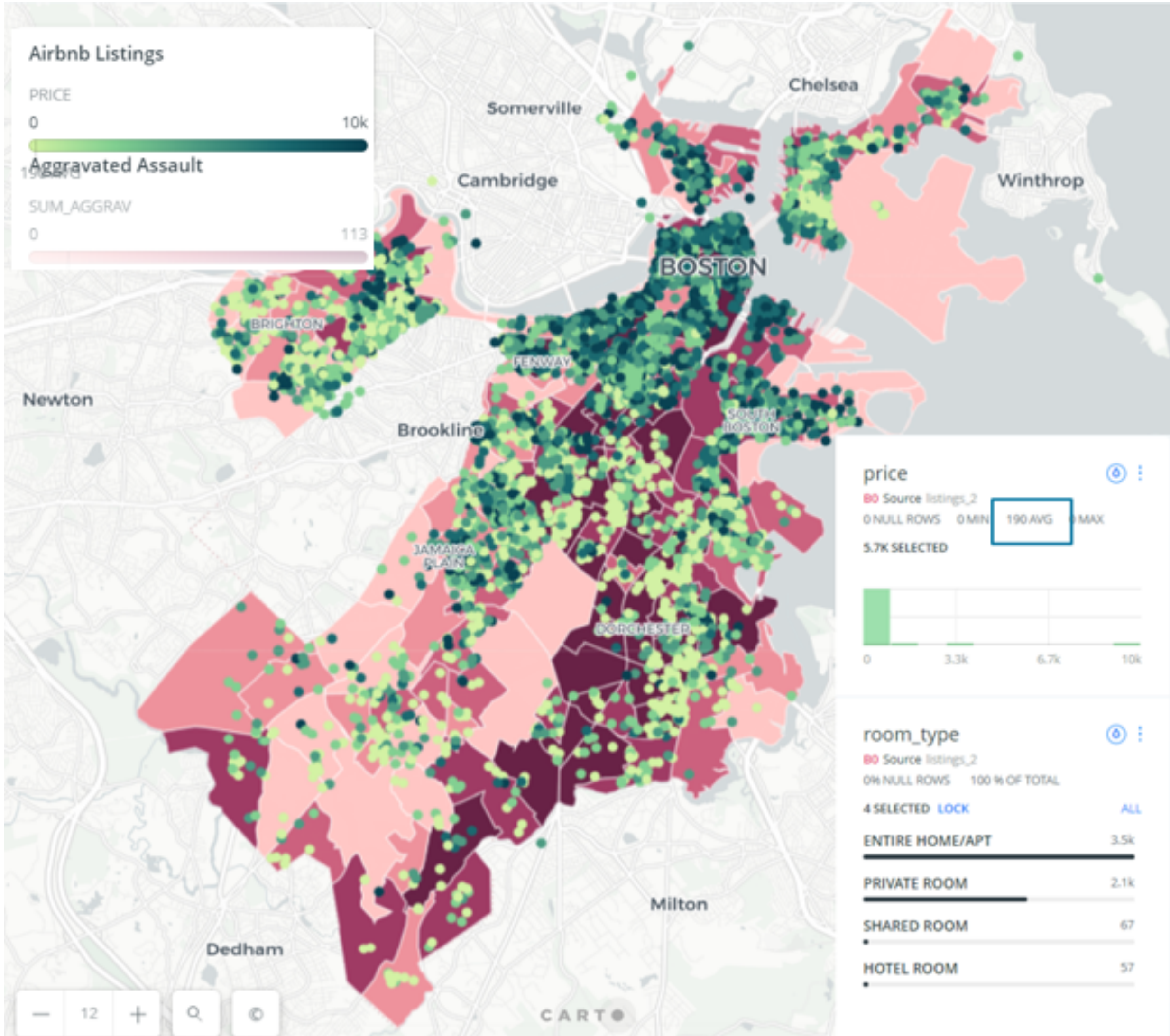

Figure 16: Price distribution of Airbnb listings and incidence of aggravated assaults in Boston, MA, 2018.

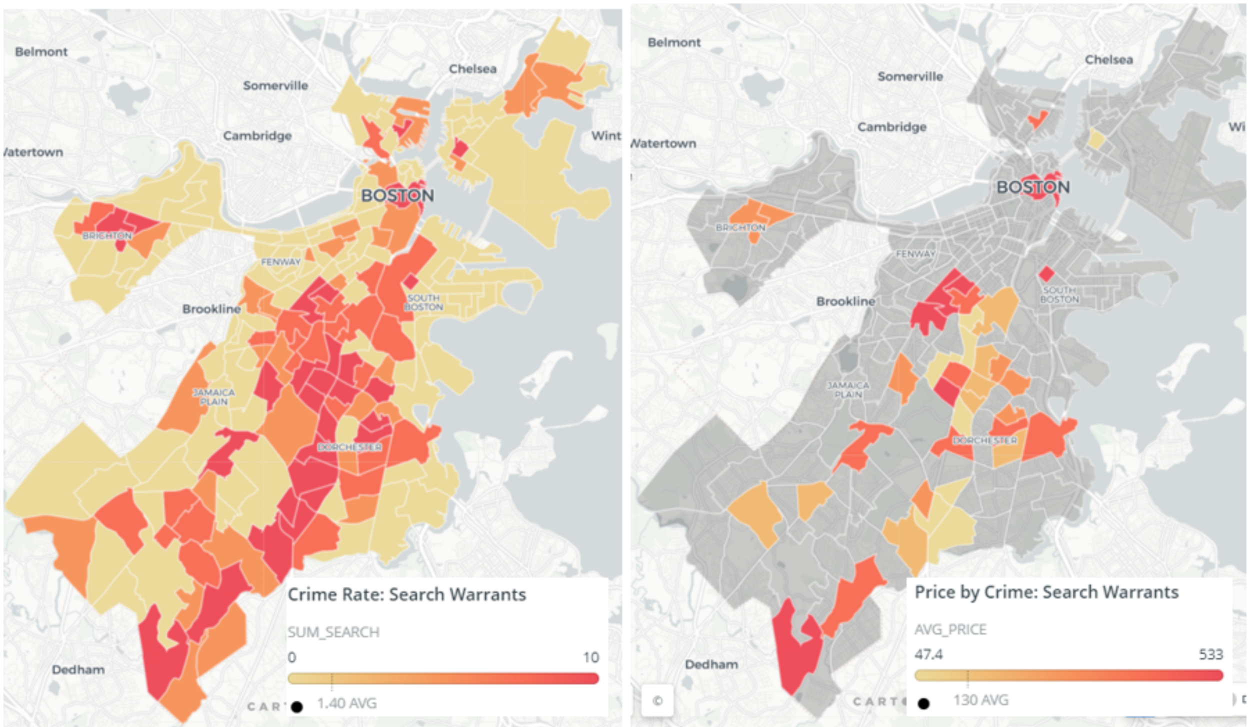

Figure 17: Search warrant crimes incidence rate (left) and price distribution of Airbnb listings in census tracts with high incidence rate of search warrants (right).

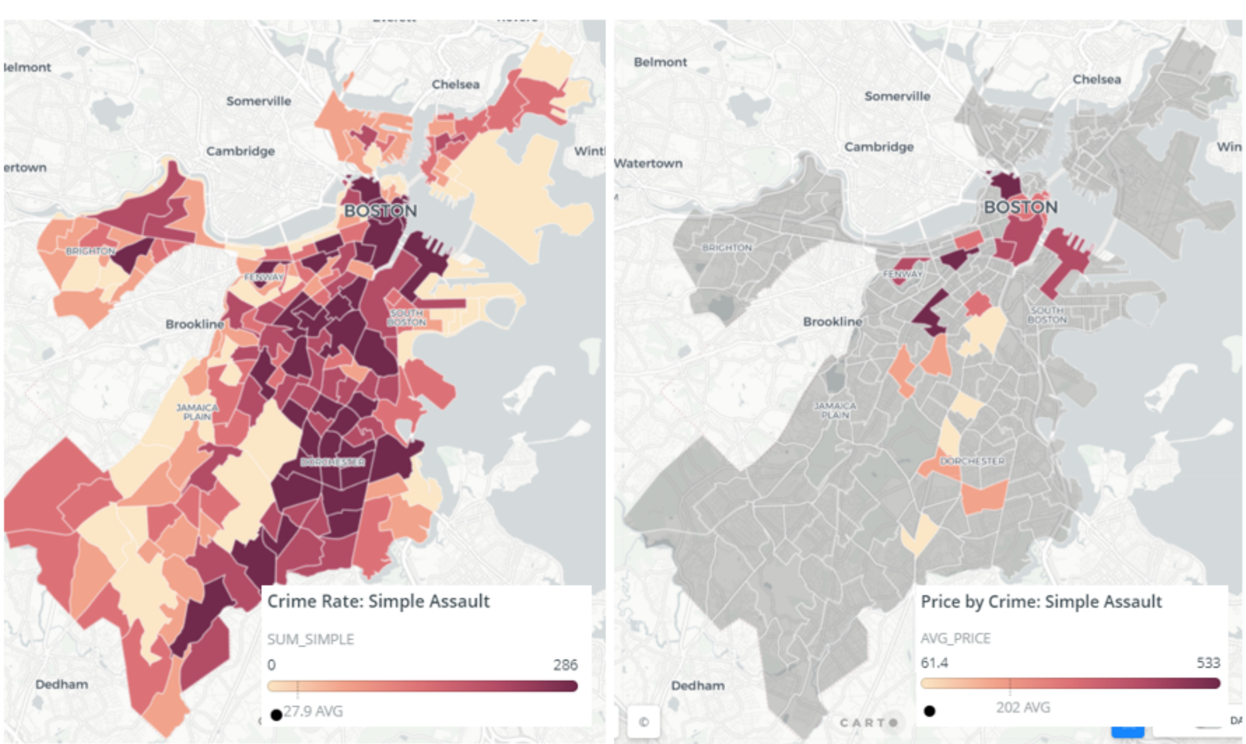

Figure 18: Simple assaults crimes incidence rate (left) and price distribution of Airbnb listings in census tracts with high incidence rate of simple assaults (right).

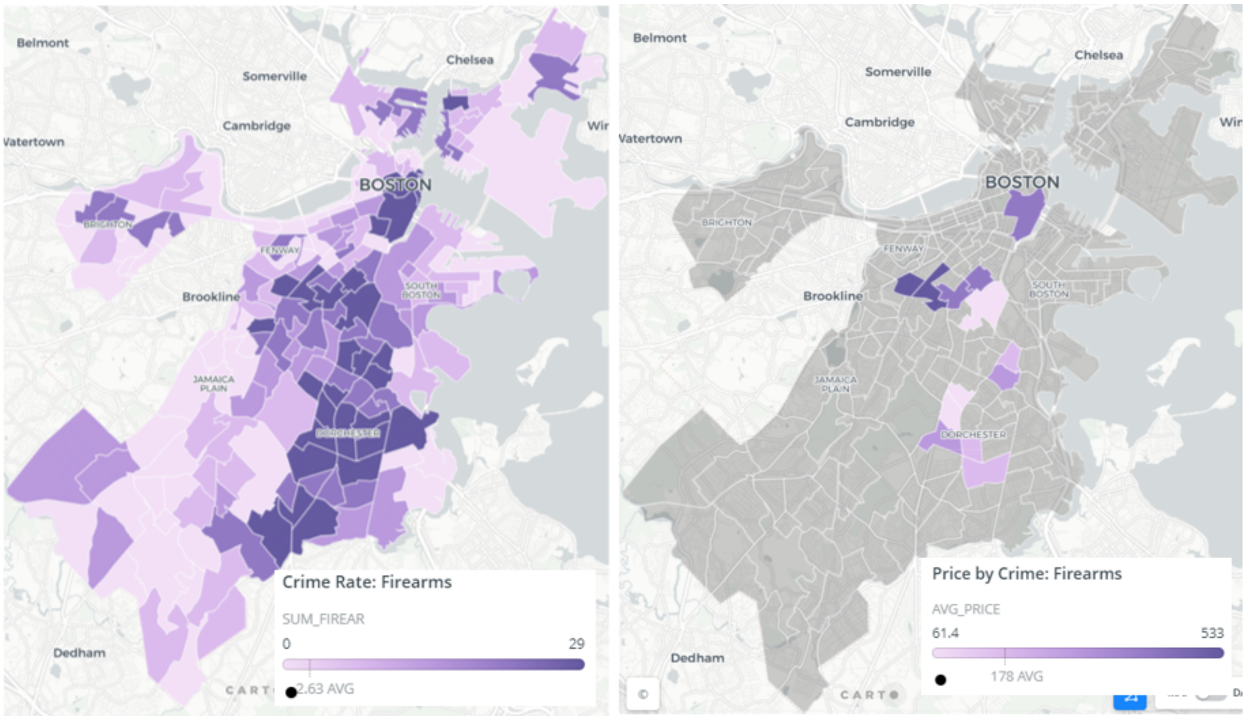

Figure 19: Firearms crimes incidence rate (left) and price distribution of Airbnb listings in census tracts with high incidence rate of firearm crimes (right).

4.2.2 Long-term housing

For this spatial analysis, the not-normalized assessed building price values were used to create the maps. The average assessed prices of housing buildings in Boston, MA is of $750, while the average assessed building price in census tracts with high incidence of vandalism (Figure 20) is $650K assessed building value, being lower than the city average matches the predicted effect based on the coefficient obtained (coefficient: -0.1380 , p-value: 6.21×10^{-6}***) in the statistical analysis.

The average price in census tracts with high incidence of drug violence (Figure 21) is $790K assessed building value, being lower than the city average which contradicts the predicted effect based on the coefficient obtained (coefficient: -0.0634, p-value: 0.0017**) in the statistical analysis. This is due to the fact the price used to map long-term housing is not normalized. When statistically normalizing the price data, this becomes a negative number which have no physical meaning and would make the maps difficult to understand. Further suggestions on how to normalize the data is discussed in the next section.

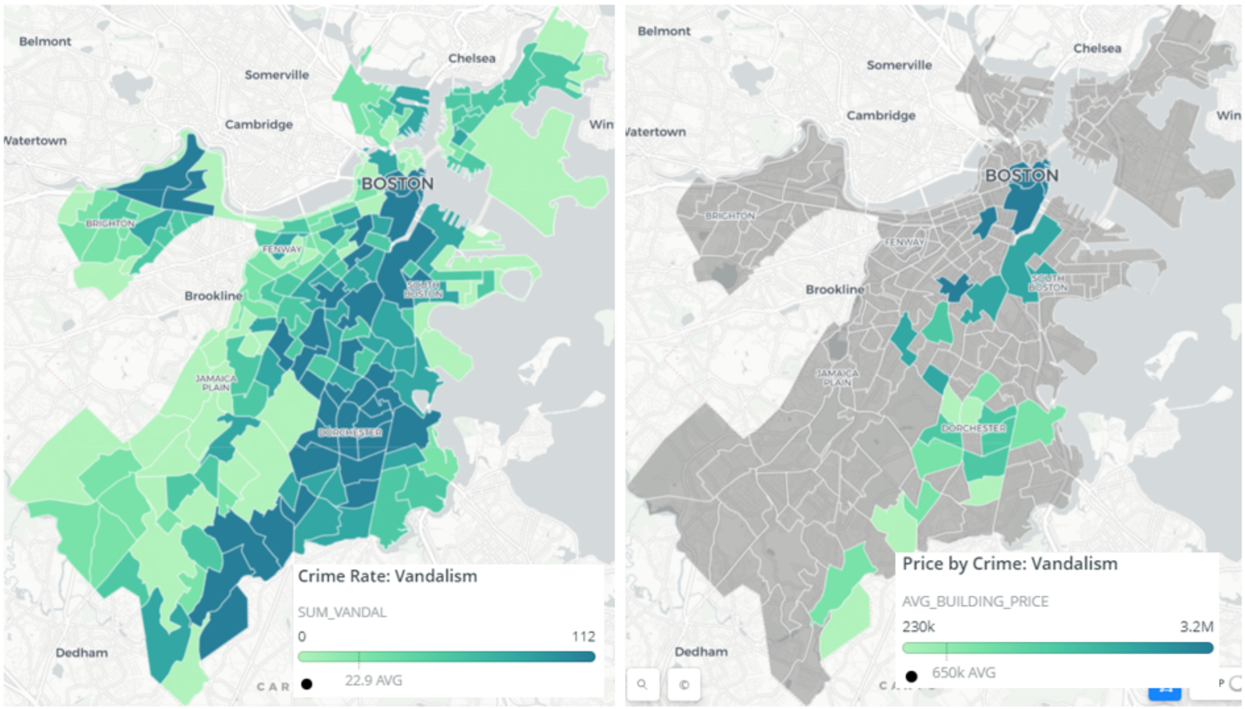

Figure 20: Vandalism crimes incidence rate (left) and price distribution of long-term housing buildings in census tracts with high incidence rate of vandalism (right).

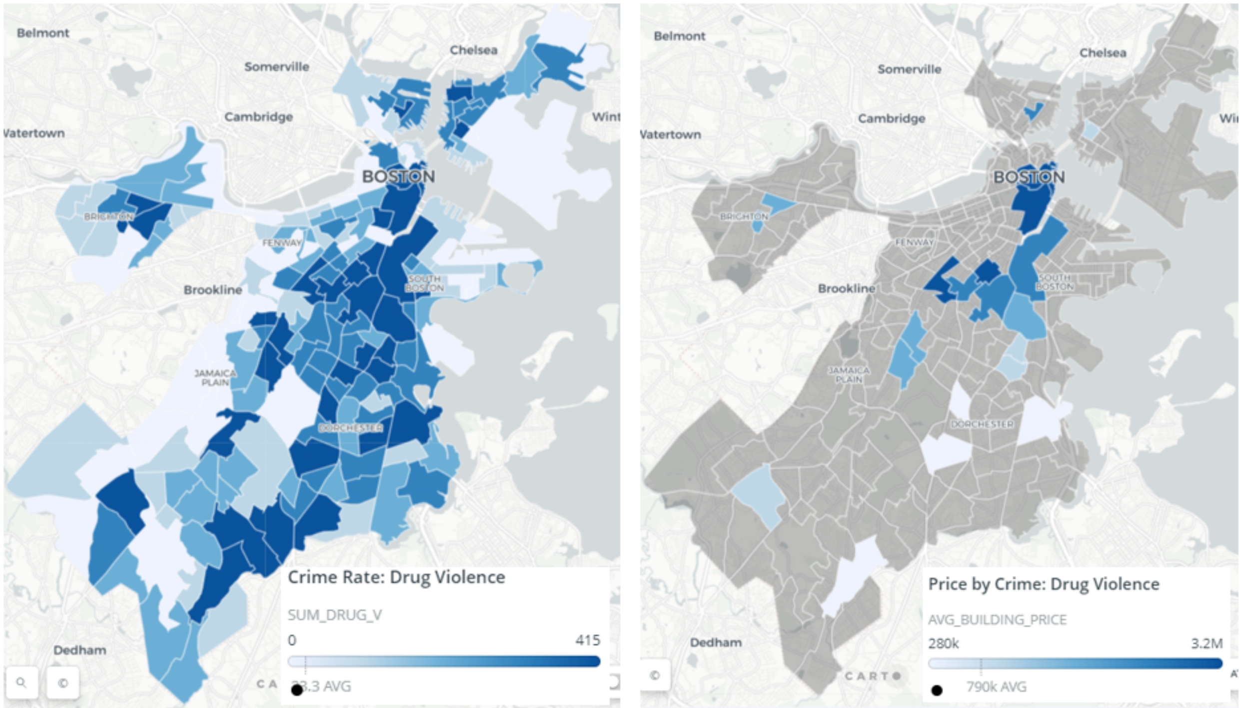

Figure 21: Drug violence crimes incidence rate (left) and price distribution of long-term housing buildings in census tracts with high incidence rate of vandalism (right).

5. Discussion and Conclusions

6.1 Policy Implications

It is useful to compare how crime rates influence short-term housing prices against long-term housing prices as residents in the city may have different knowledge of the neighborhood environment compared to the short-term visitors. Having this issue in mind, the policy implication of this project suggests stricter screening policies to allow hosts to post their rooms on the Airbnb platform.

6.2 Conclusion, Limitations and Future Work

Search warrants, firearm and vandalism affect Airbnb price largely in Boston area. These crimes are usually more serious and can be disclosure by media as a result make an effect to visitors, then demands of Airbnb will decrease as a result bring down the Airbnb price. While vandalism and drug violation affect housing price largely in Boston area. Drug violation is happened more frequently and only can be known by local. Drug violation and vandalism may be considered together as unsafety factors by local. In conclusion. Drug violation rate can be post on Airbnb hosts information as a reference for visitors.

Most types of crimes that significantly affect the pricing of short-term rentals and long-term housing are different, except for the case of two types of crimes which affect both in the same manner and level of significance: vandalism is has a positive correlation with pricing for Airbnb listings and housing assessed values with a low significance level; while drug violence crimes have a negative correlation to both pricings with a high level of significance.

The limitations of this work reside in the research of the effects on pricing due to crime incidences on a census tract level without considering other variables that might influence the price of long- and short-term housing, such as housing properties

In terms of future work and avenues of improvement, incorporating housing properties such as type of housing, square footage, number of bedrooms/bathrooms would make the model more robust and provide a wider context to better understanding of the magnitude of the effect of crime incidence in both types of housing scenarios.

Another improvement would be to normalize both short-term Airbnb and long-term housing data by square footage. This would require for us to use are more Airbnb dataset covering several years of data, as in 2018 only 32 observations have a reported square footage.

Appendix

Appendix I: Variable Definitions

Variable Name	Variable Description
price_s	Median value of price per night for Airbnb short term hosing in census tract
price_l	Median value of price per night for long term hosing in census tract
Sum_Aggrav	Number of Aggravated Assault crimes in census tract
Sum_Aircra	Number of Aircraft crimes in census tract
Sum_Arson	Number of Arson crimes in census tract
Sum_Assemb	Number of Assembly or Gathering Violations crimes in census tract
Sum_Auto_T	Number of Auto Theft crimes in census tract
Sum_Auto_1	Number of Auto Theft Recovery crimes in census tract
Sum_Ballis	Number of Ballistics crimes in census tract
Sum_Bomb_H	Number of Bomb Hoax crimes in census tract
Sum_Commer	Number of Commercial Burglary crimes in census tract
Sum_Crimin	Number of Criminal Harassment crimes in census tract
Sum_Disord	Number of Disorderly Conduct crimes in census tract
Sum_Drug_V	Number of Drug Violation crimes in census tract
Sum_Explos	Number of Explosives crimes in census tract
Sum_Firear	Number of Firearm Violations crimes in census tract

Sum_HOME_I	Number of HOME INVASION crimes in census tract
Sum_Harass	Number of Harassment crimes in census tract
Sum_Homici	Number of Homicide crimes in census tract
Sum_Mansla	Number of Manslaughter crimes in census tract
Sum_Offens	Number of Offenses Against Child / Family crimes in census tract
Sum_Reside	Number of Residential Burglary crimes in census tract
Sum_Robber	Number of Robbery crimes in census tract
Sum_Search	Number of Search Warrants crimes in census tract
Sum_Simple	Number of Simple Assault crimes in census tract
Sum_Vandal	Number of Vandalism crimes in census tract
Sum_Warran	Number of Warrant Arrests crimes in census tract

Appendix II: Resources

CARTO Maps:

- Airbnb and Crime Rates by Census Tract:
  https://eneririu.carto.com/builder/cf588bd5-e1f1-4362-a697-ef4a3a4a5192/embed
- Long-term Housing and Crime Rates by Census Tract:
  https://eneririu.carto.com/builder/8435efa5-4400-4b56-907b-30b7027abafe/embed

Appendix III: Organization

- Team members and task distribution:
  - Yijie Zhi: Data sources and pre-processing.
  - Yujia He: EDA and modelling.
  - Irene Ham Liu: Data analysis, mapping and reporting.

References

[1] Alsudais Abdulkareem. 2017. Quantifying the Offline Interactions between Hosts and Guests of Airbnb. Association for Information Systems, Virtual Communities and Collaboration. Available at: https://aisel.aisnet.org/amcis2017/Virtual/Presentations/16/

[2] Ert Eyal, Fleischer Aliza, Magen Nathan. 2016.Trust and reputation in the sharing economy: The role of personal photos in Airbnb. Tourism Management, Volume 55,Pages 62-73,ISSN 0261-5177. https://doi.org/10.1016/j.tourman.2016.01.013.

[3] Dan Wang, Juan L. Nicolau. 2017. Price determinants of sharing economy based accommodation rental: A study of listings from 33 cities on Airbnb.com. International Journal of Hospitality Management,Volume 62,Pages 120-131,SSN 0278-4319. DOI: https://doi.org/10.1016/j.ijhm.2016.12.007.

[4] Zhang Zhihua, Chen Rachel J.C. , Han Lee D., Yang Lu. 2017. Key Factors Affecting the Price of Airbnb Listings: A Geographically Weighted Approach. Sustainability 2017, 9, 1635.

[5] Coles, Peter A. and Egesdal, Michael and Ellen, Ingrid Gould and Li, Xiaodi and Sundararajan, Arun. 2017. Airbnb Usage Across New York City Neighborhoods: Geographic Patterns and Regulatory Implications. Forthcoming, Cambridge Handbook on the Law of the Sharing Economy. Available at: http://dx.doi.org/10.2139/ssrn.3048397

[6] Salvador Ruiz-Correa, Itzia Ruiz-Correa, Carlo Olmos-Carrillo, Fatima Alba Rendón-Huerta, Beatriz Ramirez-Salazar, Laurent Son Nguyen, and Daniel Gatica-Perez. 2019. Mi Casa es su Casa? Examining Airbnb Hospitality Exchange Practices in a Developing Economy. Trans. Soc. Comput. 2, 1, Article 2 (February 2019), 24 pages. DOI: https://doi.org/10.1145/3299817

[7] Xu Yu-Hua, Kim Jin-won, Pennington-Gray Lori. 2017. The Sharing Economy: A Geographically Weighted Regression Approach to Examine Crime and the Shared Lodging Sector. Journal of Travel Research, 58(7), 1193–1208. Available at https://doi.org/10.1177/0047287518797197 and https://scholarworks.umass.edu/cgi/viewcontent.cgi?article=2075&context=ttra

[8] George E. Tita, Tricia L. Petras and Robert T. Greenbaum. 2006. Crime and Residential Choice: A Neighborhood Level Analysis of the Impact of Crime on Housing Prices. J Quant Criminol (2006) 22:299–317. DOI: 10.1007/s10940-006-9013-z