Austin Nguyen A13749314
Jingyi Tay A19901470
Sammy Said A12792805

COGS 109
December 9th, 2019

Effects of Crime on Airbnb Listing in Chicago

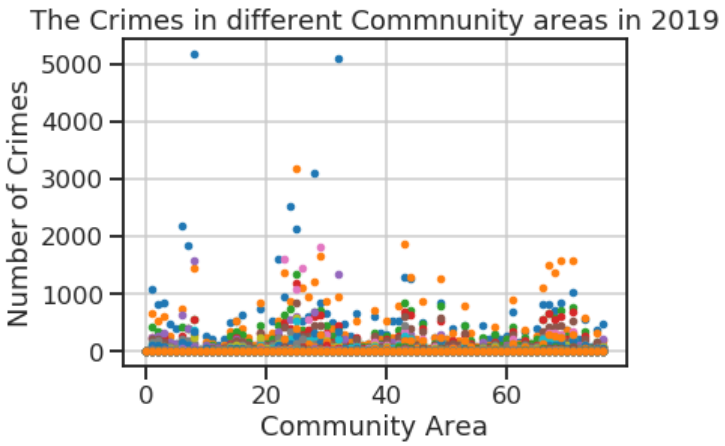
Abstract

How does certain features of a city's ecosystem, namely the crime rate, affect businesses and tourism in Chicago? More specifically, is there a relationship between the types of criminal activities (theft, burglary, battery, ...) and the number of Airbnb listings? Airbnb is an online marketplace where individuals can arrange lodging, primarily homestays, or tourism experiences. The company has a readily available data set of its listsing's that can be combined with Chicago's public records to explore if there is a correlation between the crime rate and the number of Airbnb listings. The patterns of crimes' impact on Airbnb industry hopes to provide security suggestions to protect the safety of lodging guests, hosts, and the property from being victims of crime. However, no relationship was found that can provide insights into how to improve businesses, tourism, and public safety as a whole in Chicago. However through regression models, it is shown that there is hardly anything correlation between Airbnb listings in the city of Chicago and crime rates and a key idea behind our results is that visitors to Chicago are not deterred from visiting Chicago by the high crime rate that occurs in certain areas of the city and will still visit Chicago by staying in safe places, which shows why the number of listings in the years culminating at 2019 do not decrease.

Introduction

According to The New York Post Chicago is among the worst in the world for violence since 2016, and is known as America's mass-shooting capital. Through the end of October, there have been 2,242 shooting victims and 424 murders in Chicago. Last year, we had 2,462 shooting victims and 478 murders over the same period.

Even though Chicago experienced a substantial decrease in numbers of shooting and murder crime in the past couple years, it still does not justify the safety of citizens as a huge number of people are still victims of crime. The city's overall crime rate, especially the violent crime rate is higher than the US average. Chicago was responsible for nearly half of 2016's increase in homicides in the US, though the nation's crime rates remain near historic lows.

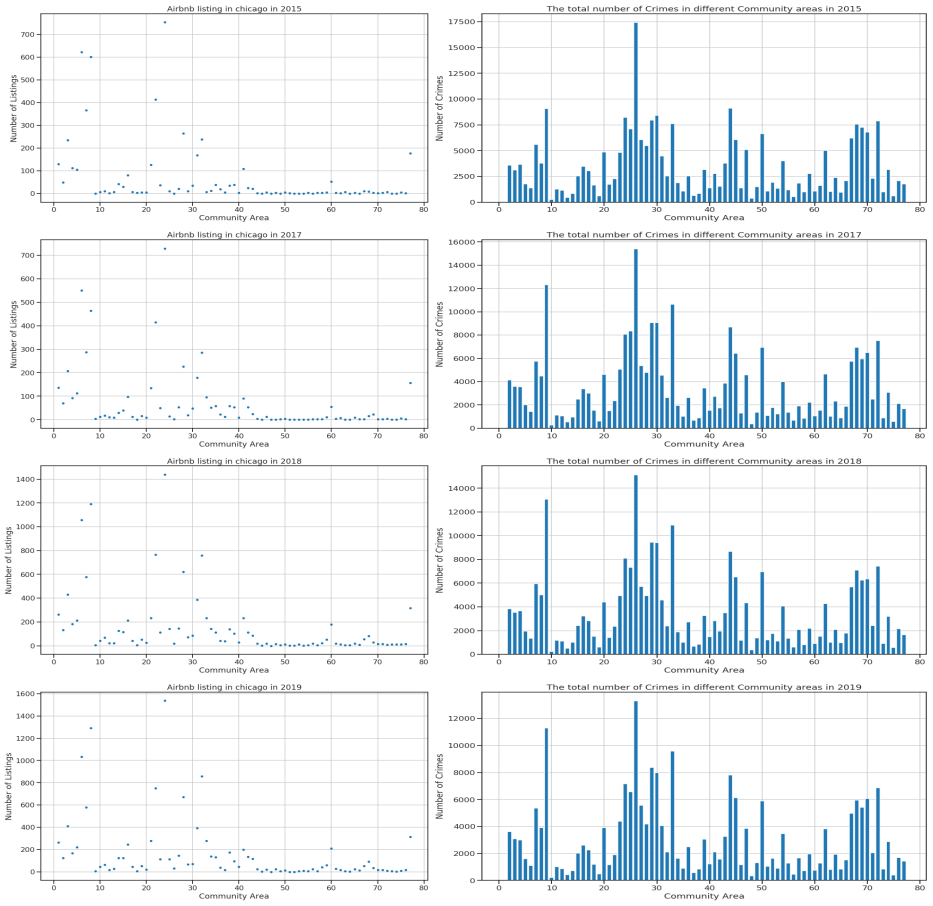


In a 2011 article by Launch.co, it was reported that residents in high crime area were told not to rent their rooms to white people for fear of racially motivated crime that may negatively affect the reputation of the city nationwide and internationally. Thus we hope to find out whether there is any significant correlation between the crime rate in Chicago and the number of Airbnb listings to see if there are patterns that we could retrieve that could potentially improve the safety of Airbnb consumers, like avoiding areas like Near West Side and Near North Side) and policy

making. It would also be significant to know if the label as america's crime city after 2016 did anything to affect the number of Airbnb listings in Chicago negatively.

Methods

For our project, we decided to utilize datasets collected by the city of Chicago and Airbnb. Specifically, we were interested in reported incidents of crime and the number of Airbnb listings in Chicago to achieve our goal. We let each row represent an observation by utilizing the following grouping: 1 of the 77 distinct community areas subdivided to Chicago's urban planning paired with the year starting from 2015 up to 2019, the most current year. The arbitrary grouping is necessary to combine the two data datasets from the city of Chicago and Airbnb together. For the columns, we let the left-most columns be our features and the right-most column be our predictor. Therefore, we let the number of each type of criminal incident be our explanatory variable in an attempt to capture the number



of Airbnb listings, our predictor. To extract the features, we had to first clean the dataset accumulated from the city of Chicago. We began by inspecting the 'crimes.csv' file and found that the dataset compiled all incidents in Chicago dating back to the year 2001. We then dropped any rows whose incident happened before the year 2015. This is because the Airbnb dataset only had dataset listings starting in the year of 2015. We then dropped all columns besides 'Date', 'Primary Type', and 'Community Area'. The reason is that these three columns will allow us to extract our features for the final data set as well as let us combine it with the Airbnb dataset . We then had to order and partition the rows by year and had to further subdivide by community area. This allowed us to count the different types of crime incidents that occurred in a community area in any given year.

Furthermore, to extract our predictors we had to clean the dataset amassed from the Airbnb website. This is done so by first knowing that the two data sets, the types of crimes committed in any given community in a given year and the total amount of listings any given year, has to be joined by the given year. Second, there are random updates throughout the years on the website insideairbnb.com where the data is retrieved from, but each update comes with a set of dataset that mentions the Calendar Data of each listing from that day of the year to the next (like the balance sheet of a company), and more importantly, the listing data, where it contains the information of all the listings present at the date, such as key details like address and community area. So the most accurate portrayal of the number of listings that happen in each year would be to aggregate the listing datasets that were updates throughout a particular year and remove duplicate listings and match the neighbourhood name with their community area code and summarize the number of listings of each community area code. This is done iteratively for

the other years too. Thus, by combining the two datasets together with the grouping mentioned above, a picture of the head of our data frame can be seen below.

After cleaning our dataset, one can easily see that from our exploratory data set that this would be a regression over classification problem. Therefore, we performed multiple linear regression on the combined data set, as well as, LASSO for regularization in an attempt to enhance the interpretability and prediction accuracy of the statistical model produced by minimizing the testing error.

MOTOR VEHICLE THEFT	ROBBERY	...	STALKING	INTIMIDATION	OBSCENITY	CONCEALED CARRY LICENSE VIOLATION	NON - CRIMINAL	NON-CRIMINAL	PUBLIC INDECENCY	HUMAN TRAFFICKING	OTHER NARCOTIC VIOLATION	Number of Listings
0	0	...	0	0	0	0	0	0	0	0	0	263
102	144	...	1	1	0	1	1	0	0	0	0	134
171	133	...	3	4	1	3	0	0	0	0	0	431
123	106	...	1	3	0	2	0	0	0	0	0	183
77	39	...	0	5	0	0	1	0	0	0	0	213

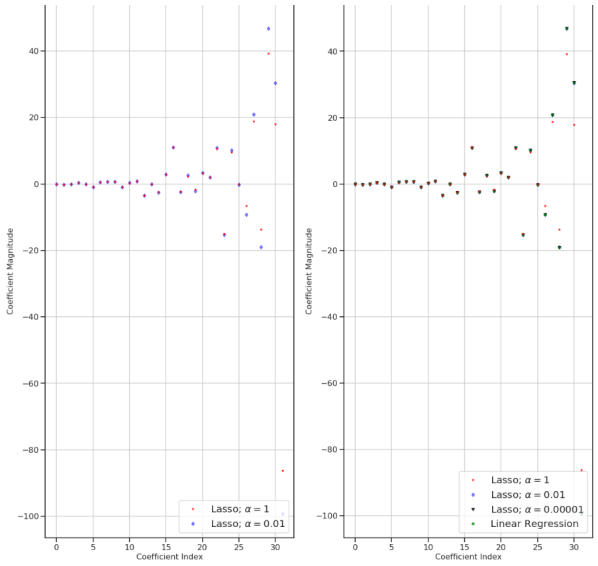
Results

The results seen are represented using the graphs below. It is seen that the R-squared value of the multiple linear regression model showed that there is not a strong correlation between Airbnb listing and the crime predictors. The P-values of most of our predictors were

OLS Regression Results

Dep. Variable:	Number of Listings	**R-squared (uncentered):**	0.394
Model:	OLS	**Adj. R-squared (uncentered):**	0.323
Method:	Least Squares	**F-statistic:**	5.601
Date:	Thu, 05 Dec 2019	**Prob (F-statistic):**	2.30e-16
Time:	22:14:24	**Log-Likelihood:**	-2062.3
No. Observations:	308	**AIC:**	4189.
Df Residuals:	276	**BIC:**	4308.
Df Model:	32		
Covariance Type:	nonrobust		

>0.05 which showed the lack of correlation between our predictors and our desired variable, implying that our model is weak, but there were some predictors that did carry some significance in the model. Predictors such as Battery, Interference with a Police Office, Deceptive Practic,e and sex offense with P-Values > 0.005. So in an effort to improve our model we performed Lasso Regression to reduce the weight of the predictors that do not carry significance. The act of doing so showed no difference in model as the test error remain the same throughout the process. The test error for all methods which include using LASSO with an alpha values of 0.01, 0.00001 and Linear Regression remained the same at -0.347. As seen on the graph to the left there was barely any in the coefficient values between LASSO and linear regression. As a result we decided to stick with our original model to make our conclusion about our model.



Discussion

The next steps of this project would be to either look for a different set of predictors to use that may affect Airbnb listing in chicago or continue to use our current crime data and look at using an unsupervised learning method such as PCA. What could have possibly made our model less accurate could be the way we organized our data. Deciding to organize our predictors based on community and not based on the population size of a community could have added some

unwanted bias in our model. We may have also has some duplicate Airbnb listing per community due to the nature of how the data was downloaded.

Conclusion

After performing a multiple linear regression on the combined dataset it was seen that the final model was not able to properly predict the number of Airbnb listings. The final model used the different kinds of crimes as predictors and the number of listings as the outcome variable. Based on the results, it was concluded that there is no evidence supporting our initial hypothesis. This implied that we have no evidence to reject the null hypothesis that the decrease in crime across the years as 2015-2019 does not affect the changes in the Airbnb listings that have during those same years. In other words, the rate of crime does not affect tourism airbnb rentals in Chicago in general from 2015-2019. In other words, the rate of crime does not affect tourism Airbnb rentals in Chicago in general from 2015 - 2019.

	coef	std err	t	P>|t|	[0.025	0.975]
THEFT	-0.0512	0.078	-0.654	0.514	-0.205	0.103
BATTERY	-0.5301	0.189	-2.805	0.005	-0.902	-0.158
CRIMINAL DAMAGE	0.1827	0.276	0.663	0.508	-0.360	0.725
NARCOTICS	0.7268	0.427	1.703	0.090	-0.113	1.567
OTHER OFFENSE	0.0671	0.263	0.255	0.799	-0.450	0.585
ASSAULT	-0.7018	0.438	-1.604	0.110	-1.563	0.160
DECEPTIVE PRACTICE	0.3069	0.096	3.212	0.001	0.119	0.495
BURGLARY	0.7254	0.299	2.425	0.016	0.137	1.314
MOTOR VEHICLE THEFT	0.2356	0.410	0.574	0.566	-0.572	1.043
ROBBERY	-0.9251	0.468	-1.977	0.049	-1.846	-0.004
CRIMINAL TRESPASS	0.2268	0.497	0.456	0.648	-0.751	1.205
WEAPONS VIOLATION	0.3822	0.546	0.700	0.485	-0.693	1.457
PUBLIC PEACE VIOLATION	-2.5829	1.661	-1.555	0.121	-5.853	0.687
OFFENSE INVOLVING CHILDREN	0.9482	1.330	0.713	0.476	-1.669	3.566
CRIM SEXUAL ASSAULT	0.2568	2.253	0.114	0.909	-4.179	4.693
PROSTITUTION	1.2099	2.028	0.597	0.551	-2.782	5.202
INTERFERENCE WITH PUBLIC OFFICER	11.5649	2.468	4.687	0.000	6.707	16.423
SEX OFFENSE	-1.2544	0.580	-2.162	0.031	-2.396	-0.112
HOMICIDE	5.4288	3.744	1.450	0.148	-1.941	12.799
ARSON	2.4903	4.239	0.587	0.557	-5.855	10.835
GAMBLING	6.9269	4.362	1.588	0.113	-1.661	15.515
LIQUOR LAW VIOLATION	4.1899	4.086	1.025	0.306	-3.854	12.234
KIDNAPPING	9.5446	7.417	1.287	0.199	-5.056	24.145
STALKING	-3.0733	7.809	-0.394	0.694	-18.445	12.298
INTIMIDATION	1.0019	8.223	0.122	0.903	-15.186	17.190
OBSCENITY	-3.0173	4.023	-0.750	0.454	-10.938	4.903
CONCEALED CARRY LICENSE VIOLATION	-0.3783	12.328	-0.031	0.976	-24.648	23.891
NON - CRIMINAL	25.9549	18.427	1.409	0.160	-10.319	62.229
NON-CRIMINAL	-4.5767	30.804	-0.149	0.882	-65.218	56.065
PUBLIC INDECENCY	20.3634	31.171	0.653	0.514	-41.000	81.727
HUMAN TRAFFICKING	42.7582	39.651	1.078	0.282	-35.298	120.814
OTHER NARCOTIC VIOLATION	-53.6775	39.172	-1.370	0.172	-130.792	23.437

The results could also be interpreted as the tourism industry in Chicago is largely unaffected by the label of the most dangerous crime city in US after 2016 and one primary reason might be that people generally avoid high crime areas and populate low crime areas. This could be seen by the graphs below where many areas with high Airbnb listings does indeed have low crime rate with the exception of 1-2 places. There are also shifts in patterns in the types of crimes committed through the period of 2015-2019, where Community Area 9 shows a large increase in crime after 2015. More exploratory analysis could be done on the variations of crime between neighborhoods and the types of crimes provoking them.

References

http://insideairbnb.com/get-the-data.html

https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present-Dashboard/5cd6-ry5g

https://nypost.com/2019/11/09/chicagos-crime-proves-black-lives-dont-matter-to-democrats/

https://launch.co/blog/airbnb-hosts-chicago-neighborhood-most-murders-in-the-city-6.html

https://en.wikipedia.org/wiki/Crime_in_Chicago

https://www.citylab.com/equity/2019/10/chicago-crime-rate-statistics-data-violence-segregation/599317/

https://time.com/4497814/chicago-murder-rate-u-s-crime/

https://www.usnews.com/news/articles/2016-09-19/chicago-drives-uptick-in-murders-national-crime-rate-stays-near-historic-lows

https://en.wikipedia.org/wiki/Airbnb