Analysis of Homelessness in NYC

Vishwajeet Shelar, Chuan-Heng Lin, Chunqing Xu, Dongjie Fan

NYU Center for Urban Science and Progress, NY, USA

vys217@nyu.edu, chl557@nyu.edu, cx495@nyu.edu, df1676@nyu.edu

Abstract

More New Yorkers are living on the streets and homeless shelters than ever before. There are thousands of homeless people sleeping in the shelters and on the streets. We are trying to figure out solutions that can help Department of Homeless Services (DHS) tackle this homelessness problem. Our research consists of three parts. We used data from 311 datasets, datasets provided by the DHS, NYPD, and the demographics dataset to create a supervised learning model which would predict the hot and cold spots in NYC to find homeless people. We also evaluated the effectiveness of the homeless drop-in centers using hypothesis testing and found that the drop-in centers may not be that effective. We analyzed the homeless data and discovered evidence on how policies affect the homeless population.

1. Introduction

Homelessness is a big problem in New York City. Every year the number of homeless people is increasing. In July 2016, there were 60,456 homeless people sleeping each night in the New York City Municipal shelter system and more than thousands of unsheltered homeless people sleeping on the streets, subways and other public places. The primary causes of homelessness include lack of affordable housing and employment opportunities, health and mental health challenges. These factors need to be analyzed and steps need to be taken to reduce the number of homeless people on streets as well as in the shelters.

We analyze the time series graph of the number of homeless people sleeping in the shelters and on the streets which gave insight about the patterns in homeless people. In NYC, the homeless people need to go to the drop-in centers to find a shelter. So, the placement of the drop-in centers should be fitting. In our research, we analyze the effectiveness of the drop-in centers.

We created a supervised machine learning model using homeless people's statistics, crime data and demographic parameters that can be linked to loss of home, to predict the areas where homeless people are clustered. We have included major factors like crime rate, income per capita, the percentage of higher education, unemployment. All of which are important reasons that relate to homelessness. The machine learning model will help DHS to figure out the hot and cold spots for homelessness in NYC

The rest of the paper is organized as follows. Section 3 provides explanatory analysis of the 311 homeless complaints dataset and the DHS daily shelter report dataset. The methodology and technical aspects of the model are discussed in Section 4. Finally, we have presented the results in Section 5 and discussed them in Section 6.

2. Dataset

We analyze the homeless complaints recorded in the New York City's 311 open dataset during 2013 to 2016. New York City is divided into 5 boroughs and 2166 unique census tracts. We analyzed the homeless complaints received from the 2166 census tracts. The complaint types in 311 data-set that we have considered in our analysis are Homeless Encampment and Homeless personal assistance. The datasets regarding the municipal homeless shelters, shelter scorecard, homeless shelters and the drop-in centers are obtained from DHS as well as 'www.coalitionforthehomeless.org'. There are 5 homeless drop-in centers in NYC. The demographics dataset which includes the ethnicity, employment, income, age, sex data is taken from data2go.nyc. The crime data provided by NYPD is taken as an indicator of homelessness.

3. Relevant Studies

Homelessness is a wicked problem. The city agency, nonprofit organizations like ‘Coalition for Homeless’, BRC and many other have collected data about homeless people and created statistics that identify the trends in the homeless population in NYC. These organizations have analyzed the policies and provided ideas that can help reduce the homeless people. In our research, we have also analyzed the trends in homeless population, but our research is unique because we have used machine learning techniques to predict the homeless people patterns and provide a model that can help the DHS.

4. Methodology

The aim of this work was threefold as mentioned in the introduction. In the exploratory analysis part, we explore the homeless data to find trends, we do this by time series analysis. The hypothesis testing is used to gauge the effectiveness of the drop-in centers. We use the machine learning model to predict the hot and cold spots of the homeless people in NYC.

4.1 Exploratory Analysis

The data about the count of homeless people in the municipal shelter obtained from the coalitionforthehomeless.org has entries from 1984 to 2016; there are some assumptions[‘Appendix 2’] related to the dataset, same is applicable to the 311 homeless data obtained from the NYC Open Data portal. The data had to be pre-processed to remove the outliers. We used python for cleaning the data and plotted the time series using rolling mean. Rolling mean helped to smooth out the time series making it clear to visually observe the trends and compensate for the outliers as well. The DHS shelter scorecard for the month of august 2016 was summarized and it displayed some interesting facts about the dataset.

4.2 Hypothesis Testing for effectiveness of the homeless drop-in centers

The second part of our research is to quantify the effectiveness of the drop-in centers by using the 311 homeless complaint data and the location data of the drop-in centers. To measure the effectiveness, we followed the below steps:

1. The average length of 20 Blocks in NYC is 1 miles. So, the length of 1 block is around 250 feet. So, we create buffers of radius 250 and 500 feet around the 5 drop-in centers.

2. Calculate the density of the number of homeless complaints from the 311 data-set. Density = #No. Of Complaints in the buffer area / Area of the buffer

3. Compare the densities in the different buffer circles

The figure 7 and 8 shows the locations of the drop-in centers in NYC and the buffers around them.

4.2.1 Hypothesis Test

For our research question, we have decided the significant value (alpha) to be 0.05, i.e. with a p-value smaller than 0.05, the observed values would be considered statistically significant to reject the null hypothesis.

The null hypothesis mathematically is defined as ‘The density of the homeless people (# No. of complaints from 311 – assuming that each complaint received is for an individual homeless person) in the buffer with small radius is greater than or equal to the density of homeless people in the buffer of larger radius.

Null Hypothesis: The drop-in center is ineffective

H0 = Density_Buffer (smaller radius) - Density_Buffer (larger radius) >= 0

The alternate hypothesis mathematically is defined as ‘The density of the homeless people (# No. of complaints from 311 - assuming that each complaint received is for an individual homeless person) in the buffer with a small radius is less than the density of homeless people in the buffer of larger radius.

Alternate Hypothesis: The drop-in center is effective

Ha = Density_Buffer (smaller radius) - Density_Buffer (larger radius) < 0

We completed the hypothesis test for the 5 different drop-in centers together and to test the statistical significance, we used student-t two-sample test.

4.3 Predictive Model to classify the hot and cold spots for homelessness in NYC

We create a classification model to predict the hot and cold spots of homeless complaints in NYC. The modeling in done in two steps. First, we create the labels for the hot and cold clusters from the number of homeless complaints using spatial analysis. In the second part, a predictive model is generated based on machine learning models (Support Vector Machine and Random Forest Algorithm)

4.3.1 Hot and Cold Spots using spatial analysis

Tobler's Law says that there is spatial autocorrelation in a variable if observations that are closer to each other in space have related values. We start our analysis considering the same principle. The places with a high number of homeless people would be clustered together, i.e there are spatial dependencies between the count of homeless people in an area and its neighborhood.

We use the ‘k-nearest neighbor’ algorithm to find the 5 nearest neighbors of each of the census tracts. The spatial lag is also calculated. In the next step, calculate the "Local indicators of spatial association" (LISA) to evaluate the clustering in the census tracts by calculating Local Moran's I and evaluating the statistical significance for each I_i. For our research, we want to cluster the census tracts in 5 groups, so we have set 2 levels of significance, one at a p-value of 0.01 and the other one at 0.05. The clusters of the hot spots and the cold spots are defined as follows:

Hot Spots (Label 1) (p-value < 0.01) - The pattern observed is statistically significant at the level 1 (alpha = 0.01) and the quadrant location is 1 i.e. ‘HH’.

Hot Spots (Label 2) (p-value >0.01 and <0.05) - The pattern observed is statistically significant at the level 2 (Label 4) (alpha = 0.05) and quadrant location is 1 i.e. HH

Cold Spots (Label 3) (p-value < 0.01) - The pattern observed is statistically significant at the level 1 (alpha = 0.01) and the quadrant location is 3 i.e. ‘LL’.

Cold Spots (Label 4) (p-value >0.01 and <0.05) - The pattern observed is statistically significant at the level 2 (alpha = 0.05) and quadrant location is 3 i.e. ‘LL’

In this way, all the 2166 census tracts are labeled as 1,2,3 or 4 and the census tracts which are not spatially autocorrelated are labeled as 0.

Figure 11 shows the hot spots and cold spots obtained from spatial autocorrelation. Using the labels obtained in this step, we further create a classification model which classifies the census tracts in two parts. One cluster will be the census tracts with no spatial autocorrelation i.e. label 0 and another cluster will be census tracts with hot and cold spots. I.e. labels 1,2,3 and 4.

4.3.2 Predictive Classification Model

We use two supervised machine learning models - Support Vector Machine and Random Forest. These two models are considered as robust decision models in the field of machine learning. We created the feature space with 172 features which include crime rate, ethnicity proportions, median income, the number of homeless shelters and others. The complete list can be found in Appendix [1].

Since many of the indicators are strongly correlated with each other, the next step is to perform dimensionality reduction using standard Principle Component Analysis (PCA). After PCA we get a reasonable selection of the top independent components. These selected principle components are then used as a feature space for teaching our model to classify. Scikit-learn package is used to perform SVM and random-forest. The parameters of the models are optimized using a function called gridsearchCV which is a function to iterate different combination of parameters. We used 67% of the dataset as training data and 33% as test data and did 10-fold cross validation. The models perform binary classification. The results about the performance of the models are discussed in Section 5.

5. Results

In this section, we discuss the results obtained from the exploratory analysis, the hypothesis tests, and the classification model.

5.1 Exploratory Analysis:

The time series plot shown in figure 1 shows that the number of homeless people has increased over the years. There are some other results obtained from the analysis of the time series data about homeless people.

1. In figure 3 which shows the trend in the number of individuals living in the municipal shelters, a large increase of 10% from 50135 to 53615 in the number of homeless people is observed around 2014. This was majorly due to Mayor Bloomberg's policy to remove the affordable housing placements and programs (section 8 housing)

2. Figure 2 shows the graph of number of single adults in the homeless shelters. The homeless single adults sleeping in the shelters has increased at an alarming rate of 11% from 12712 to 14147 between 2015 and 2016.

3. Figure 4 shows the complaints received from 311 about homeless person encampment or homeless person assistance. The data shows that there are homeless people who prefer living on the streets rather than the shelters. The NYC government has released scorecard of homeless shelters. Table 1 shows the violations in the shelters. The shelter conditions not being good may be one of the reason the homeless people prefer to sleep on the streets or other public places rather than the shelters.

4. Another interesting trend observed from the 311 homeless complaints data is that the number of complaints has increased substantially from 2015 to 2016 and the time series shows seasonality. Figure 6 shows the seasonal trend decomposed from the original time series. During the summer the homeless people prefer not to stay in the shelters. The DHS needs to come up with solutions to improve the shelter conditions and be prepared with extra shelter capacity during winters.

5.2 Hypothesis testing

We calculated the values of densities of homeless people for the different buffers. Figure 9 shows the densities of the homeless people at different radius for the 5 drop-in centers, it is observed that the density of the smaller buffer is larger than the density of the larger buffer.

The p-value obtained from the one-tail t-test is 0.1693, which means we do have sufficient evidence to reject the null hypothesis at a statistical significance of 95%. Which means that the null hypothesis that the drop-in centers are ineffective maybe true. However, more factors are supposed to be taken into consideration, like land use, population in buffer area and averaged income of residents nearby. Then we can draw much more convincing conclusion about the effectiveness of homeless drop-in centers in NYC.

5.3 Predictive Model to classify the hot and cold spots for homelessness in NYC

We perform the spatial analysis to find the hot spots and the cold spots for homeless people in NYC. The results show that the count of homeless people in census tracts are spatially autocorrelated. The Local spatial autocorrelation test results, Moran's I scatterplot is shown in figure 12. The figure shows the hot and cold spots distinguished by the p-values and the quadrant in which the points belong to.

Figure 13 shows the overall hot and cold spot clusters created when we use the complete NYC census tract dataset to do spatial analysis. When we spatially analyze the boroughs separately, we can see more granular spatial correlations between the census tracts in the boroughs. This is seen in figure 14.

As mentioned in section 4, we use PCA to find the important features. The PCA results return that almost 100% of the information is covered by 100 independent components. Figure 11 shows the results obtained from the PCA.

We create two classification models using SVM and random forest. The classification results obtained for the random Forest models is given below:

	precision	recall	f1-score	support
0	0.80	0.98	0.88	520
1	0.70	0.82	0.75	28
2	0.33	0.06	0.10	17
3	0.60	0.10	0.17	90
4	0.52	0.23	0.32	60
avg / total	0.74	0.78	0.72	715

The model created using random-forest shows ~80% confidence in classifying the given dataset. However, the SVM gives overfitting problem so, it is not considered.

The models we created using the demographics, socioeconomic, crime data is successful in classifying the census tracts into clusters without any spatial correlation and areas with hot and cold spots.

6. Conclusion

The main aim of the research is to provide DHS with data driven solutions. We analyzed the time series data and found patterns which can help DHS make policy decisions. We showed that placement of the drop-in centers maybe is not effective. However, the hypothesis testing can be further improved by adding more factors. The model we created to classify the hot and cold spots can be used by the DHS as a predictive model to target specific areas to find the homeless people in the future. The cold spots should also be given importance because the 311 complaints data is sensitive to bias. We need to consider the areas where people don't use the 311 service, such variables are not included in the model, the cold spots maybe the areas which are affected by this bias. An improved dataset can help to improve the classification model to classify the data into 5 groups.

Appendix 1

Figures

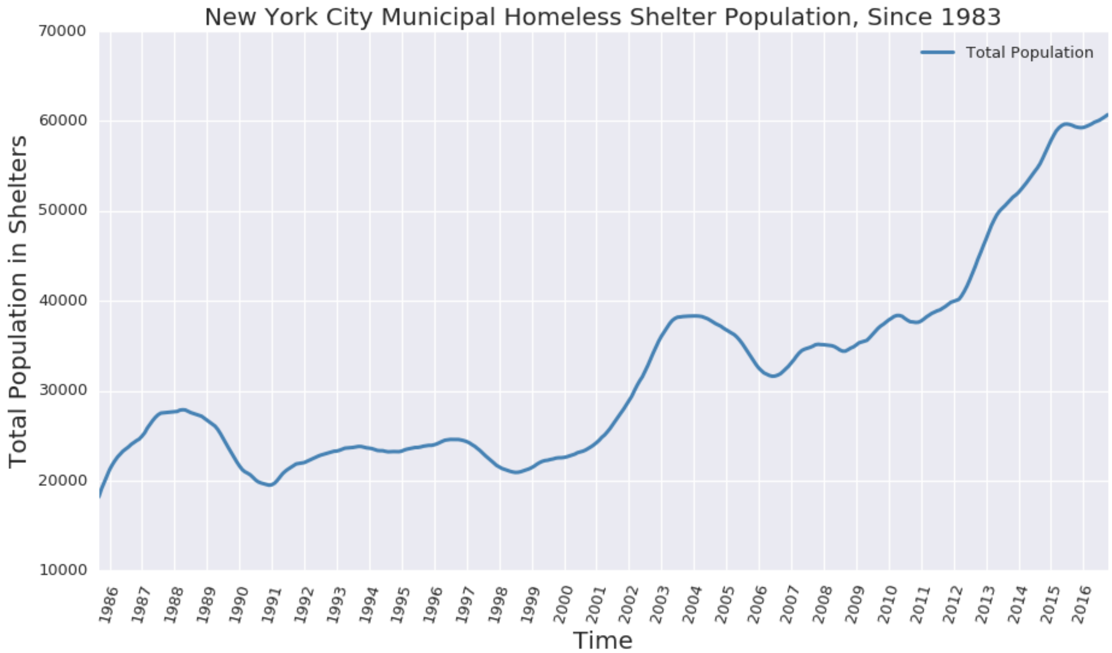


Figure 1: Total Population in Municipal Shelters of New York City Since 1983. In October 2016, there were 62,306 homeless people in the NYC Homeless shelter

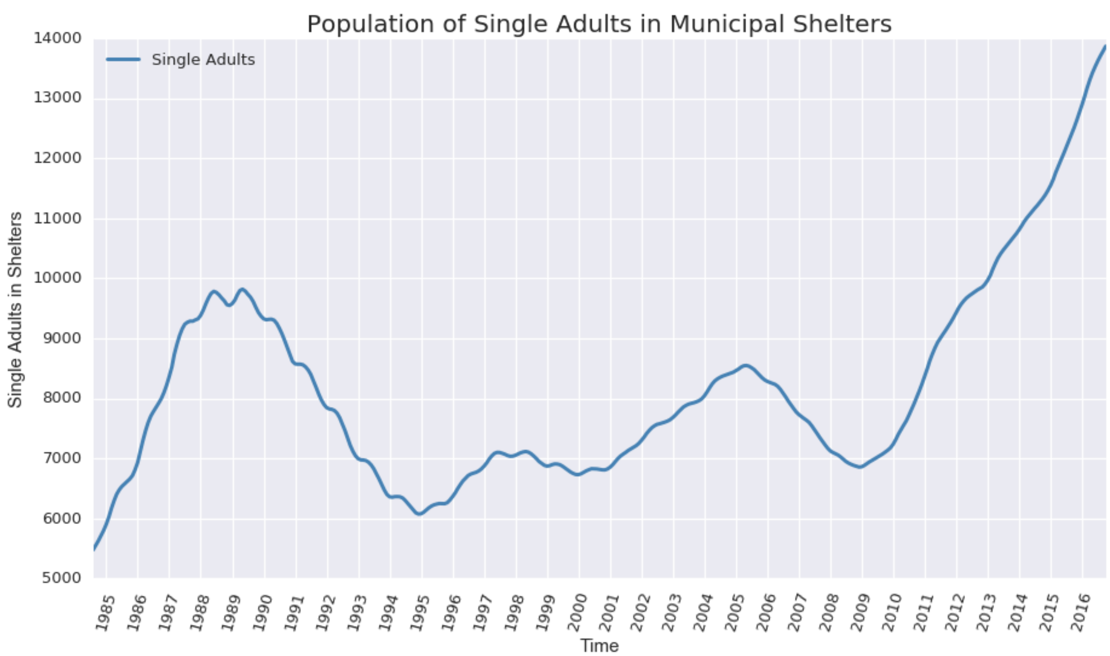


Figure 2: Population Single Adults in Shelters of New York City Since 1983. The population has increased at a high rate since 2010. There was 11% increase in the population of adults in homeless shelters from 2015 to 2016

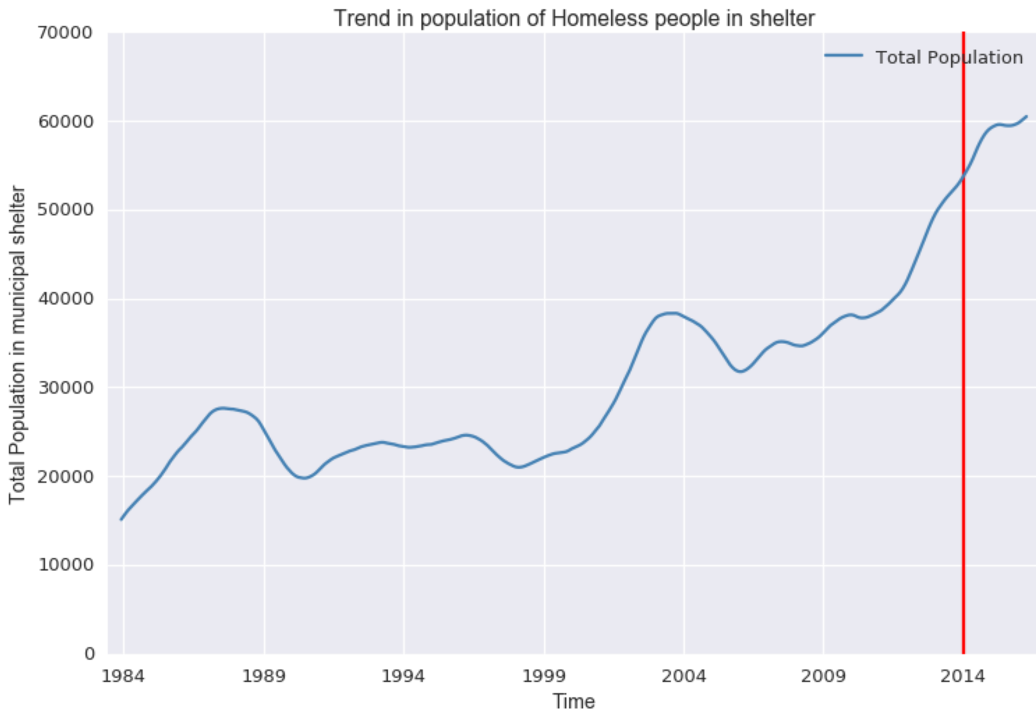


Figure 3: Increasing trend in the homeless population. The population has increased at higher rate around 2014 which is largely due to Mayor Bloomberg's policies to cancel the section 8 housing program

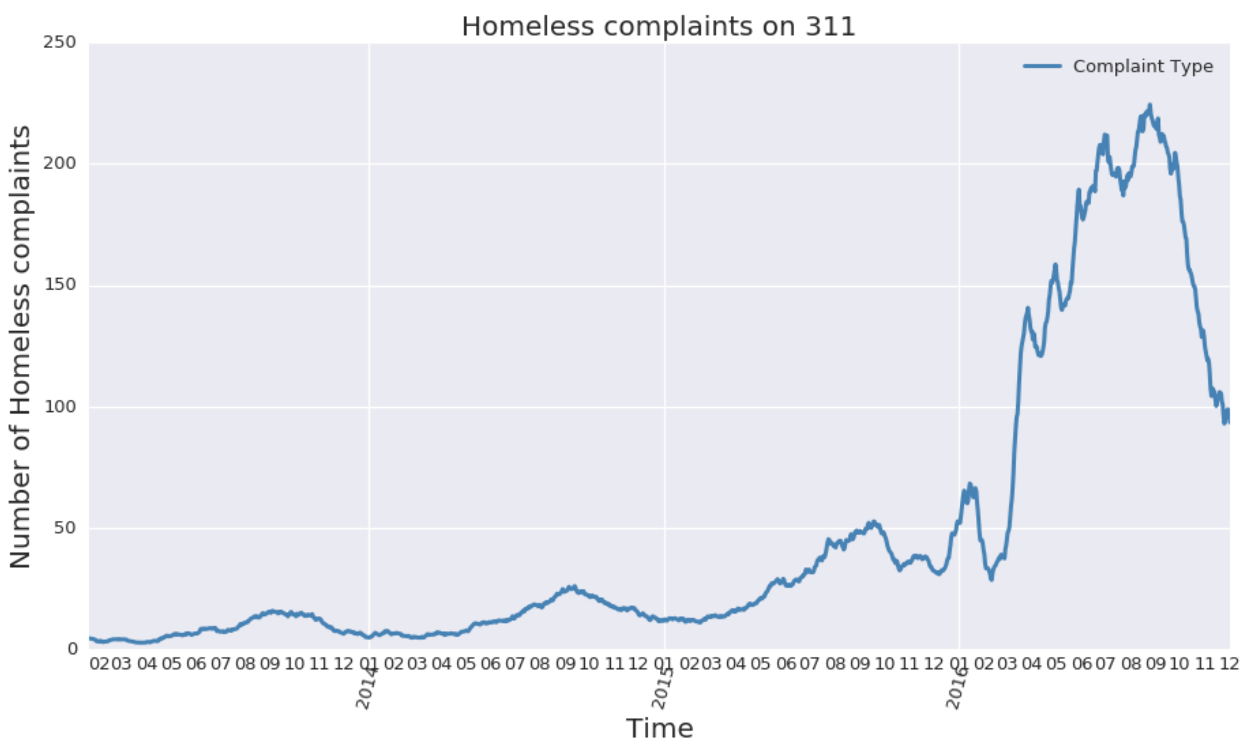


Figure 4: The homeless complaints received through 311. The number of complaints have increases in 2016 as people have started to use 311 and have become aware of the homelessness problem

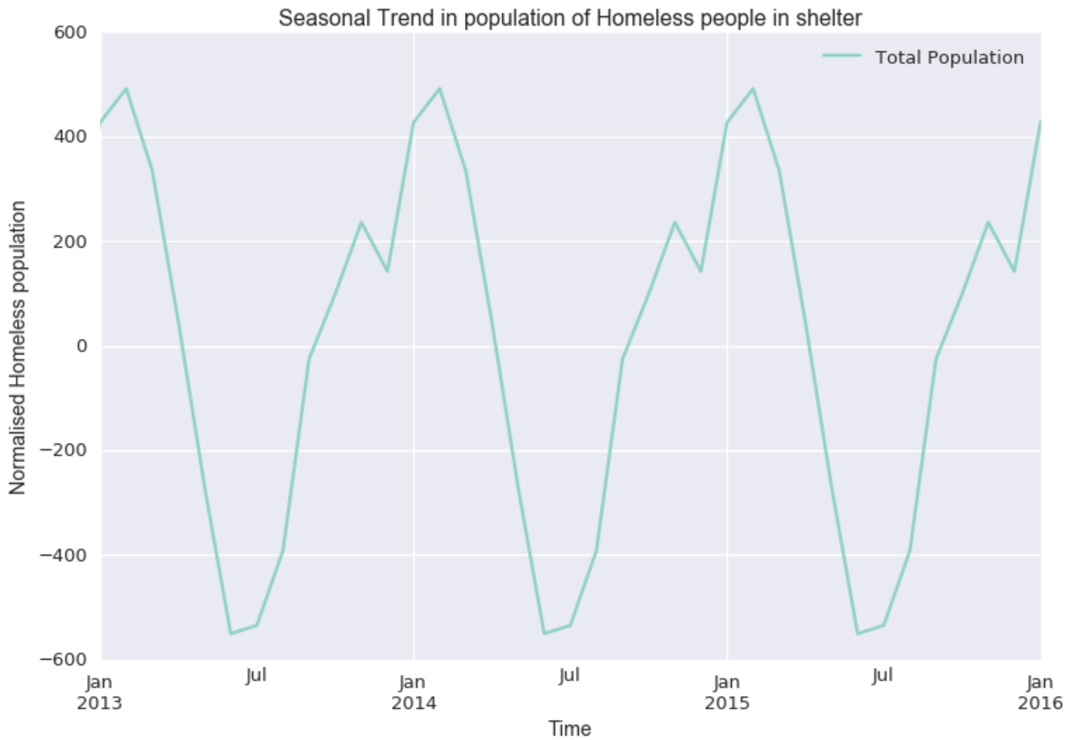


Figure 5: Seasonality in the time series graph of the population of Homeless people in shelters. The time series is periodic every season. The number of people in shelters increase every winter.

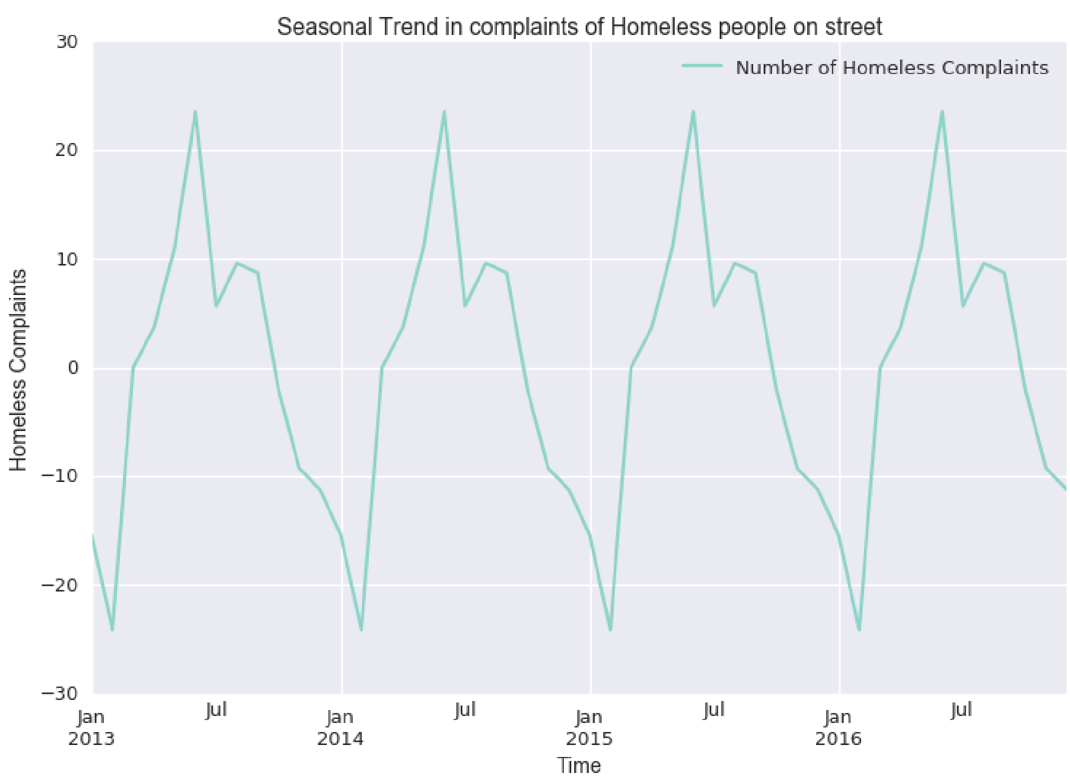


Figure 6: Seasonality in the time series graph of the complaints of Homeless people on streets. The time series is periodic every season. The number of complaints decreases every winter, this maybe because the homeless people tend to stay in the shelters during winters

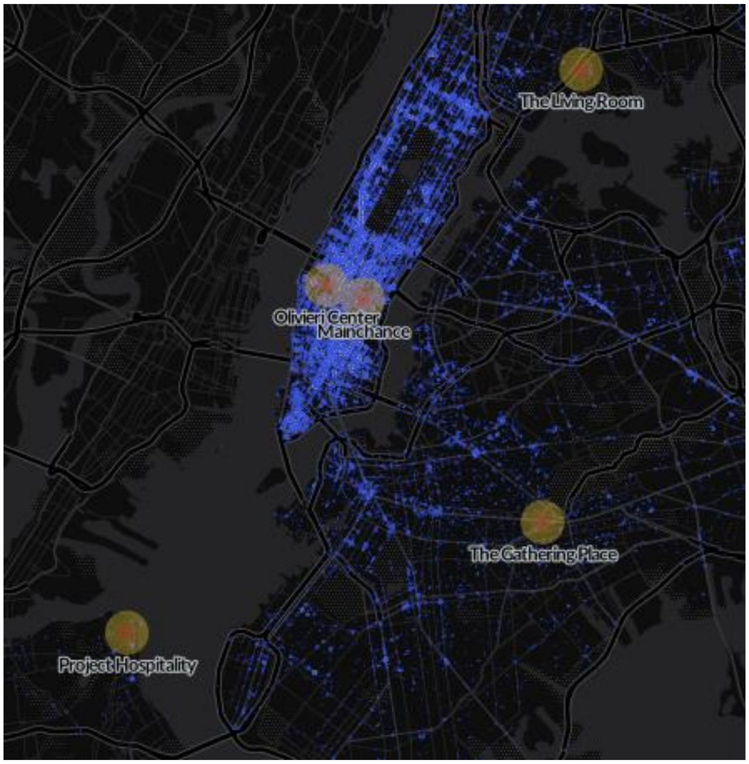


Figure 7: Drop-in Centers in New York City. There are 5 drop-in centers in NYC. The blue points represent the homeless complaints received

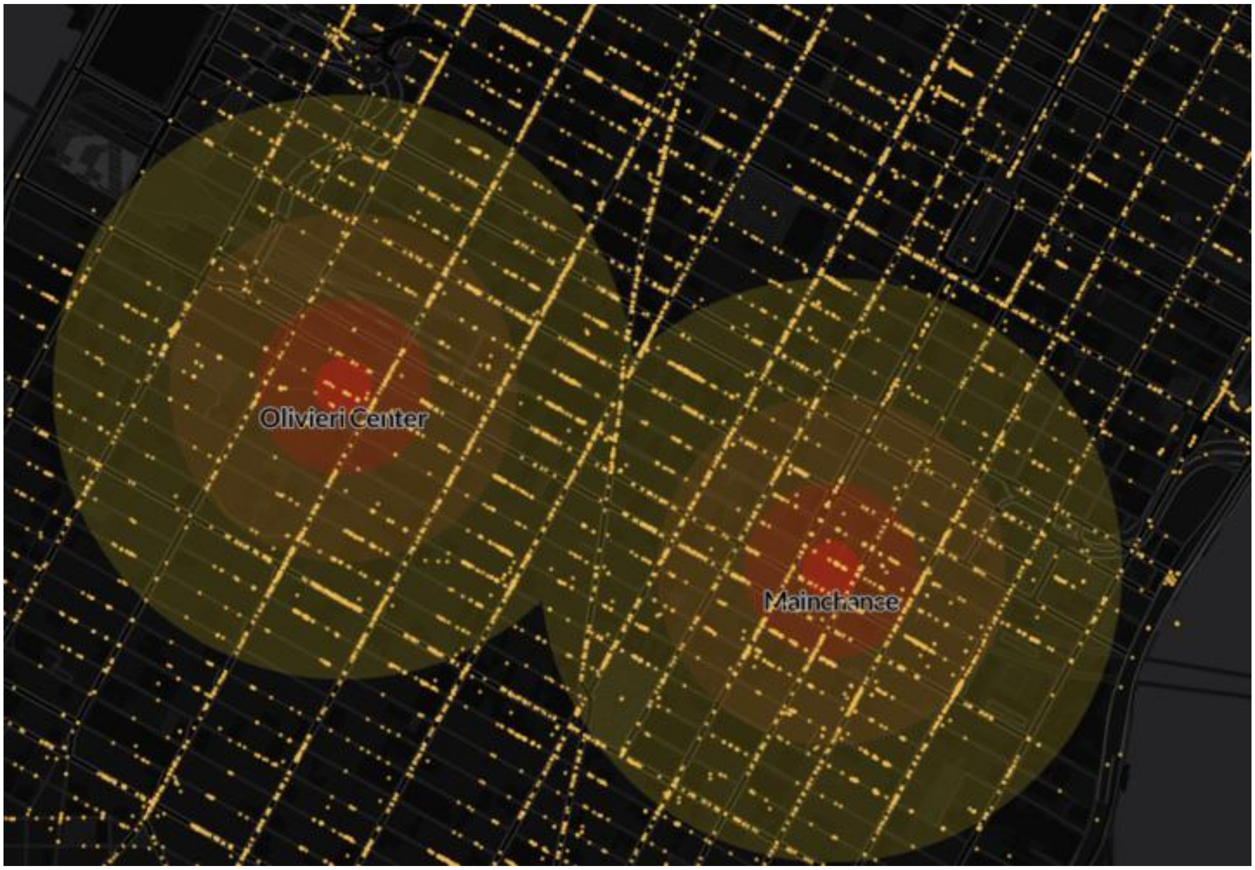


Figure 8: Drop-in Centers in Manhattan and the buffers with different radii.

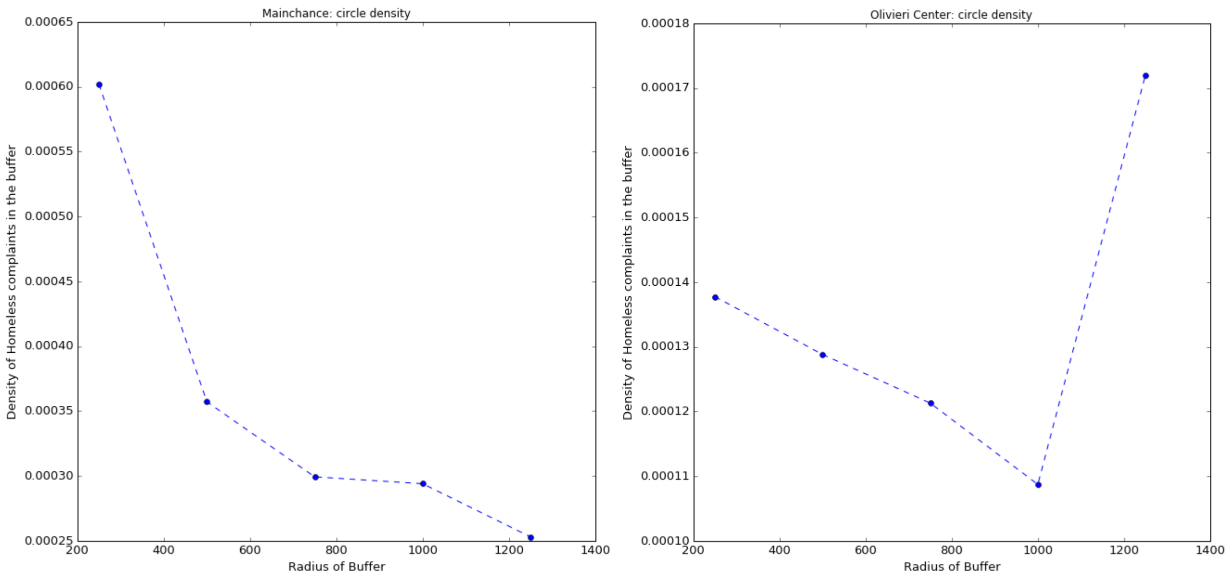


Figure 9: Cluster Density of the drop-in centers in Manhattan

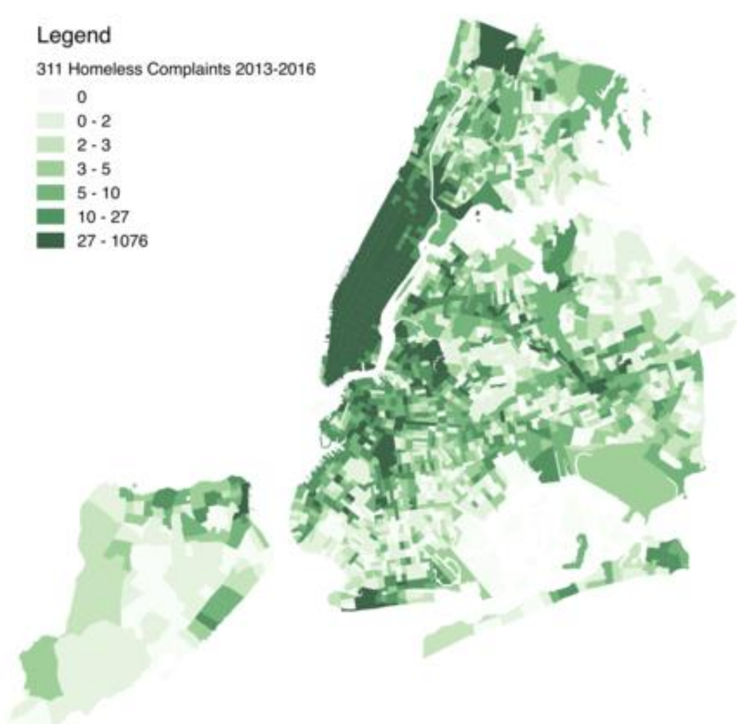


Figure 10: 311 Homeless Complaints in New York City, 2013-2016

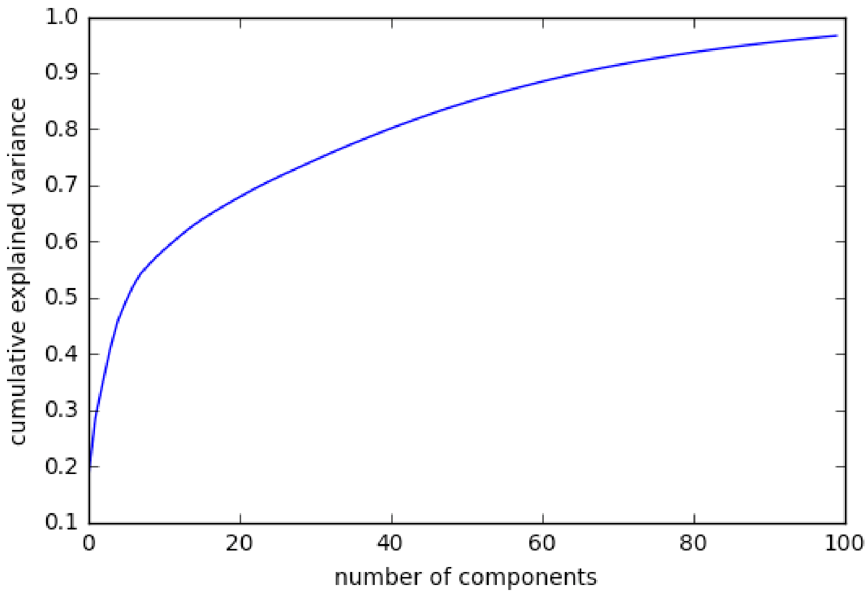


Figure 11: The first 100 of the total 172 principle components explain almost 100% of the total data variance

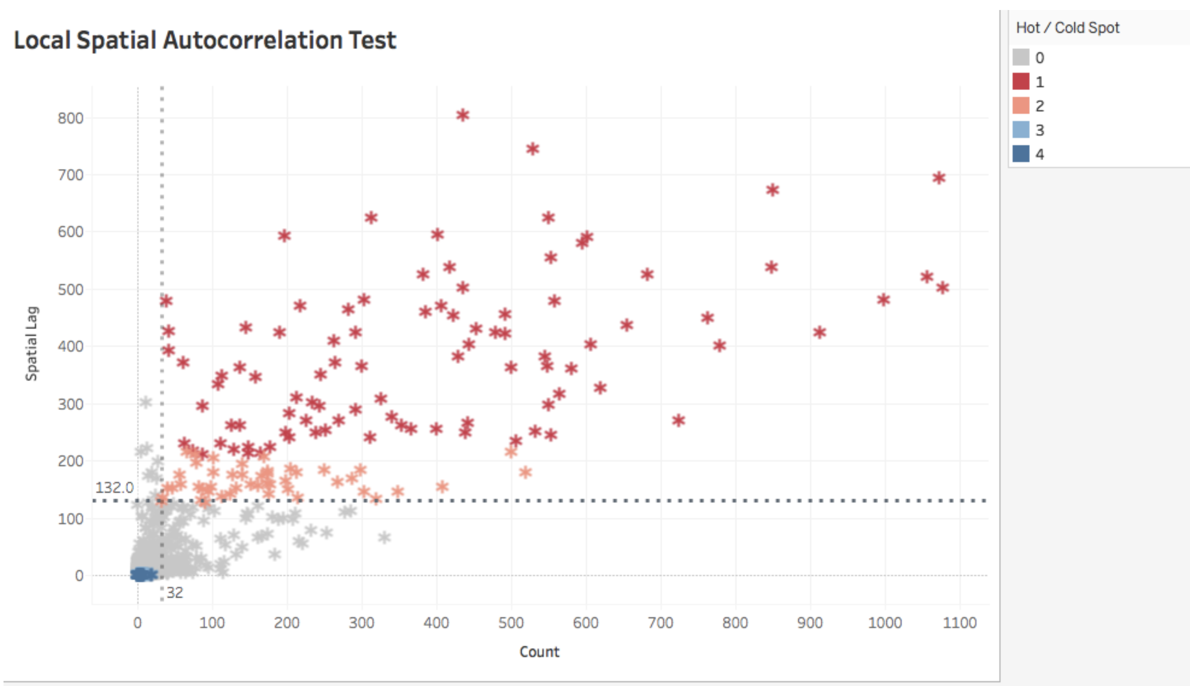


Figure 12: Scatter plot of the count of homeless people in census tract and its neighbors (spatial lag). The labels are generated per the LISA test, where 1&2 denote hot spots, 3&4 denote cold spots and 0 denotes no spatial autocorrelation

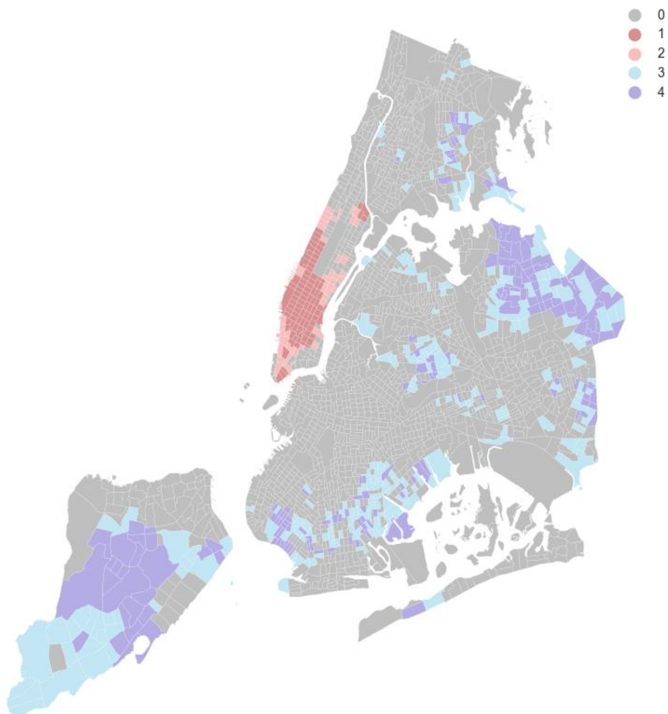
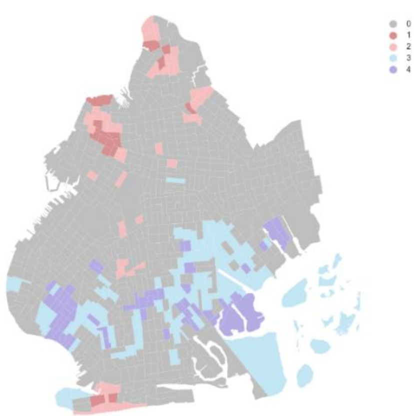
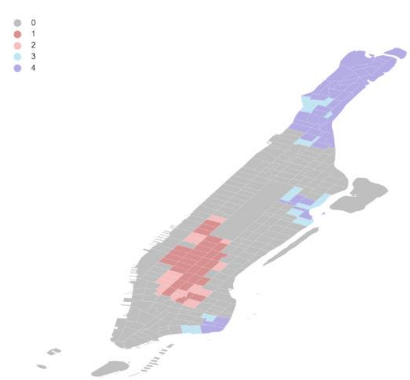


Figure 13: The hot and cold spots of the homeless complaints in NYC calculated using the spatial autocorrelation analysis on the complete NYC dataset





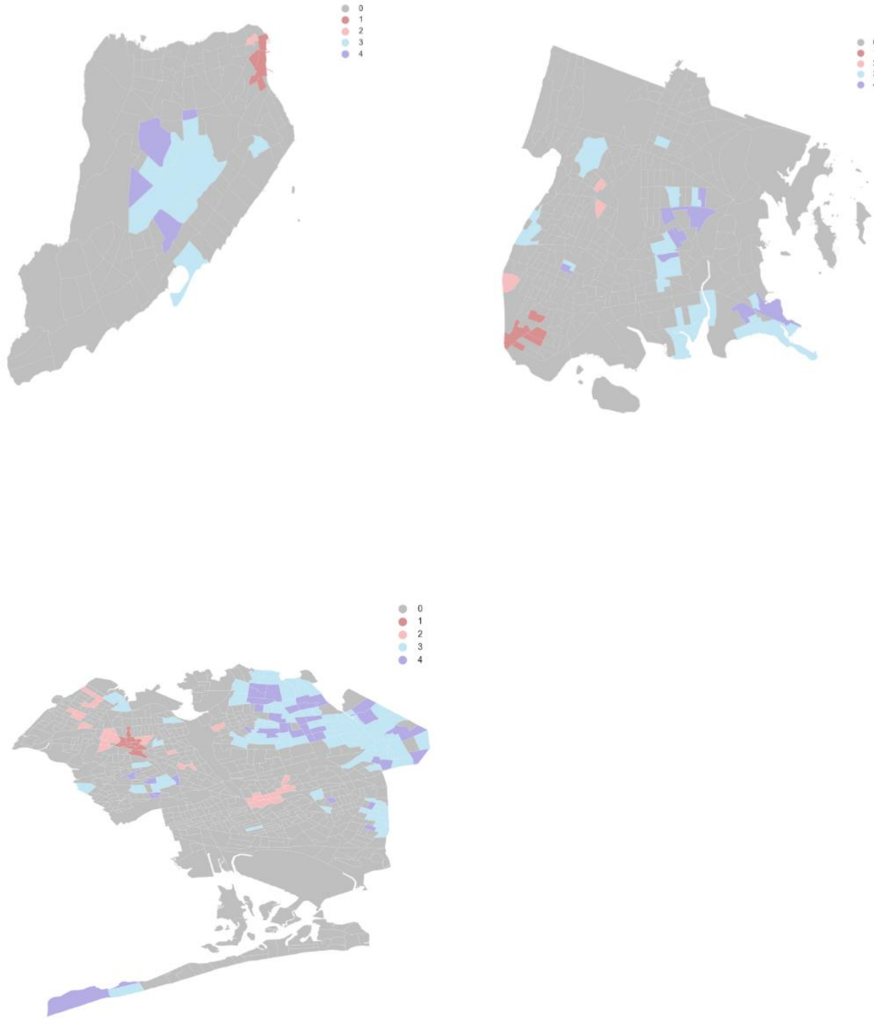


Figure 14: Individual boroughs show more granular hot and cold spots which were not visible in the complete NYC map

Tables

Type of Shelter	Count of Capacity	Sum of Violations
Adult Family Hotel	12	133
Adult Family Tier 2	13	145
Adult Shelter	103	669
Family Cluster	292	16026
Family Hotel	113	708

Family Tier 2	118	499
Late Arrival	2	19
Safe Haven	11	24
Veterans Short Term Housing	1	4
Grand Total	665	18227

Table 1: Report from the September Shelter Report Scorecard

	Radius = 250 ft	**Radius = 500 ft**
Mainchance	0.000602	0.000357
The Living Room	0.000010	0.000008
Olivieri Center	0.000138	0.000129
The Gathering Place	0.000010	0.000006
Project Hospitality	0.000005	0.000004

Table 2: The densities of the homeless complaints in buffers in per sq.ft

Feature Space

Number of Crimes in the census tract	Speak Spanish at Home (%)	Black: 2000 (% of total population)	Extreme Housing Burden (% of renters)
Number of Homeless shelters in the census tracts	Population (total #)	Foreign-Born, Caribbean (% of all foreign-born)	High Housing Burden (% of renters)
Area of the census tract	Population: 2000 (total #)	Foreign-Born, Central America (% of all foreign-born)	Lack Kitchen (% of housing units)
The total cpacity of the homless shelters in te census tract	Youth Population (% under 18)	Child Stability (% of children in same house 1 year ago)	Housing Units Occupied by Owners (%)
Adult Stability (% of adults living in same house 1 year ago)	Population under 5 (%)	Cohabitating Unmarried Couples (% of households)	Rental Vacancies (% of units)
Black (% of total population)	Foreign-Born, West Africa (% of all foreign-born)	Disabled (% of total population)	Housing Units Occupied by Renters (%)
Black Population (#)	Foreign-Born, Western Asia (% of all foreign-born)	Divorced (% ages 15+)	Industry Category: Agriculture, Forestry, Fishing, Hunting & Mining (% of all employed)

Females 15-24 (#)	Foreign-Born, Western Europe (% of all foreign-born)	Foreign-Born, Eastern Africa (% of all foreign-born)	Industry Category: Arts, Entertainment, Recreation, Accommodation and Food Services (% of all employed)
Females 25-34 (#)	White (% of total population)	Foreign-Born, Eastern Asia (% of all foreign-born)	Commute Time (average minutes each way)
Females 35-44 (#)	White: 2000 (% of total population)	Foreign-Born, Eastern Europe (% of all foreign-born)	Commute 60 Minutes Or More One Way (% of workers)
Females 45-54 (#)	White Population (total #)	Households with Children Under 18 (% of all households)	Industry Category: Construction (% of all employed)
Females 5-14 (#)	Veterans (total #)	Foreign-Born, Fiji (% of all foreign-born)	Industry Category: Educational Services, Health Care & Social Assistance (% of all employed)
Females 55-64 (#)	Veterans (total female)	Foreign-Born (% of total population)	Industry Category: Finance, Insurance, Real Estate, Rental & Leasing (% of all employed)
Females 65-74 (#)	Veterans (total male)	Grandparents Responsible for Grandchildren (% living with grandparents)	Employed Full-Time (% of workers)
Females 75-84 (#)	Veterans (%)	Latino (% of total population)	Work for Government (% of all employed)
Females 85+ (#)	Male Veterans (% of total veteran population)	Latino: 2000 (% of total population)	Gini Coefficient of Income Inequality
Females Under 5 (#)	Female Veterans (% of total veteran population)	Latino Population (total #)	Gini Coefficient of Income Inequality: 2006-2010
Males 15-24 (#)	Completed at Least Bachelor's Degree (% of adults 25+)	Living Alone (% of households)	Industry Category: Information (% of all employed)
Males 25-34 (#)	Completed at Least Bachelor's Degree: 2000 (% of adults 25+)	Elderly Living Alone (% of households)	Labor Force Participation Rate (% working or actively looking for work)
Males 35-44 (#)	Completed at Least High School (% of adults 25+)	Married With Children (% of households)	Occupational Category: Management, Business, Science & Arts (% of all employed)
Males 45-54 (#)	Completed at Least High School: 2006-2010 (% of adults 25+)	Married Without Children (% of households)	Industry Category: Manufacturing (% of all employed)
Males 5-14 (#)	Completed Bachelor's Degree (% of adults 25+)	Foreign-Born, Middle Africa (% of all foreign-born)	Median Household Income (2013 $)
Males 55-64 (#)	Completed High School or High School and Some College (% of adults 25+)	Overcrowding (% of units with more than one occupant per room)	Median Personal Earnings (2013 $)

Males 65-74 (#)	Graduate or Professional Degree (% of adults 25+)	Native American and Alaska Native (% of total population)	Median Personal Earnings: 2000 (2013 $)
Males 75-84 (#)	Did Not Complete High School (% of adults 25+)	Native American and Alaska Native Population (total #)	Median Home Value ($ for owner-occupied units)
Males 85+ (#)	Did Not Complete High School: 2006-2010 (% of adults 25+)	Native American: 2000 (% of total population)	Occupational Category: Natural resources, Construction & Maintenance (% of all employed)
Males Under 5 (#)	Preschool Enrollment (% of 3- and 4-year old)	U.S.-Born (% of total population)	Industry Category: Other Service (% of all employed)
Total Population (#)	School Enrollment (% of population ages 3 to 24 in school)	Foreign-Born, Africa, not elsewhere classified (% of all foreign-born)	Employed Part-Time (% of workers)
Speak Asian or Pacific Island Language at Home (%)	Veterans with Less than High School Degree (#)	Foreign-Born, Asia, not elsewhere classified (% of all foreign-born)	Elderly Poverty (# of adults 65 and older)
Asian or Pacific Islander (% of total population)	Veterans with a bachelor's degree (#)	Foreign-Born, Europe, not elsewhere classified (% of all foreign-born)	Elderly Poverty (% of adults 65 and older)
Asian or Pacific Islander: 2000 (% of total population)	Veterans with Less than High School Degree (%)	Foreign-Born, Oceania, not elsewhere classified (% of all foreign-born)	Poverty (# of individuals in households with incomes below poverty)
Asian or Pacific Islander Population (total #)	Veterans with a Bachelor's Degree (%)	Speak Language Other than English at Home (%)	Poverty (% in households with incomes below poverty)
Foreign-Born, Australia and New Zealand Sub region (% of all foreign-born)	SNAP Benefits (% households)	Nonfamily Households with Children Under 18 (% of all households)	Poverty: 1999 (% in households with incomes below poverty)
Speak English Less Than "Very Well" (%)	Adults with Medicaid (%)	Foreign-Born, Northern Africa (% of all foreign-born)	Child Poverty (# of children under 18 in households with incomes below poverty)
Black (% of total population)	Child Health Plus or Medicaid (% of children under 18)	Foreign-Born, North America (% of all foreign-born)	Child Poverty (% of children under 18 in households with incomes below poverty)
Foreign-Born, South America (% of all foreign-born)	Industry Category: Retail Trade (% of all employed)	Foreign-Born, Northern Europe (% of all foreign-born)	Private Wage and Salary Workers (% of all employed)
Foreign-Born, South Central Asia (% of all foreign-born)	Occupational Category: Sales & Office (% of all employed)	Two or More Races or Some Other Race (% of total population)	Occupational Category: Production, Transportation & Material Moving (% of all employed)
Foreign-Born, South Eastern Asia (% of all foreign-born)	Self-employed (% of all employed)	Two or More Races or Some Other Race: 2000 (% of total population)	Industry Category: Professional, Scientific, Management, Administrative & Waste Management (% of all employed)

Foreign-Born, Southern Europe (% of all foreign-born)	Occupational Category: Service (% of all employed)	Two or More Races or Some Other Race Population (total #)	Industry Category: Public Administration (% of all employed)
Employed Workers with Income Below Poverty Level (%)	Industry Category: Transportation, Warehousing & Utilities (% of all employed)	Prime Age Adults (% of total population ages 25-54)	TANF and General Assistance (% of households in past 12 months)
Industry Category: Wholesale Trade (% of all employed)	Unemployed (% ages 16 and older)	Single Father With Children (% of households)	Single Mother With Children (% of households)
Foreign-Born, Southern Africa (% of all foreign-born)	Unpaid Family Worker (% of all employed)	Elderly Population (% 65+)	

Appendix 2

Contribution

Vishwajeet Shelar (V.S) drafted the major part of the report. Chuan-Heng(Henry) Lin (C.H.L), Chunqing Xu (C.X) and Dongjie Fan (D.F) helped in drafting the methodology.

V.S, D.F, C.H.L and C.X collected datasets and cleaned the data individually. C.X cleaned the homeless shelter data and plotted the homeless shelter population time-series. V.S plotted and analysed the time-series (decomposition) of the 311 complaints. C.X and V.S worked together in analyzing the patterns in the time-series.

V.S and D.F worked on the hypothesis testing, D.F formulated the hypothesis test and implemented it. V.S contributed to the interpretation and conclusion.

D.F, C.X completed the spatial analysis of the 311 complaints data. C.H.L, V.S and C.X worked together to create feature space for the model. C.H.L implemented the models. D.F, C.H.L, V.S and C.X worked together to interpret the results of the models

Note on Data:

The analysis provided for unsheltered homeless people is based on the 311 complaint and maybe an underestimation of the count of homeless people. There is no accurate measurement of New York City's unsheltered homeless population, and recent City surveys significantly underestimate the number of unsheltered homeless New Yorkers. We have taken 311 data related to the homeless complaint types, there may be bias present in the data because generally most of the complaints received through 311 are from the areas with income which is on the higher scale.

Appendix 3

References

- http://www.coalitionforthehomeless.org/the-catastrophe-of-homelessness/state-of-the-homeless/
- http://www.nydailynews.com/new-york/horror-stories-show-homeless-avoid-nyc-shelters-article-1.2485417

- http://www.bisolutions.us/A-Brief-Introduction-to-Spatial-Regression.php
- https://en.wikipedia.org/wiki/Indicators_of_spatial_association
- Stanislav Sobolevsky, Emanuele Massaro, Iva Bojic, Juan Murillo Arias and Carlo Ratti, “Predicting Regional Economic Indices Using Big Data Of Individual Bank Card Transactions”, Senseable City Lab, Massachusetts Institute of Technology