Socio-Economic Impacts of COVID-19 on Brownsville and Brooklyn Communities

Final Report

Earl Lin
Jerome Louison
Yushu Rao
Xinran Zhao

Sponsor: Civ:Lab, Public Sentiment
CUSP Mentor: Stanislav Sobolevsky

Abstract

The goal of this project is to leverage real-time urban data to help New York City government and community organizations identify community-level response to the Covid-19 epidemic within a low-income neighborhood disproportionately affected by the disease, using Brownsville, Brooklyn as the study case. We determined that in fact there was a statistically significant difference in the use of government services (including a 7% drop in 311 calls, and 51% decrease in MTA ridership) as a result of the "stay-at-home" order. Additionally, we identify the Brooklyn community districts (CD) and zip codes most similar to Brownsville (using a cosine similarity matrix), and the "informal" attitudes of Brownsville residents (by sentiment analysis of Twitter data). Lastly, we conducted OLS regressions to determine the correlation between the dataset variables with cases of Covid-19. All regressions conducted showed a high level (R^2 > 0.84) of explanatory power with relation to Covid-19 related variables, and the correlation between independent and dependent variables was statistically significant (past p=0.05 threshold) for most factors. Deliverables included a data visualization tool (ie: webpage) of results, and one-page summary memo. Our methodology could potentially serve as a blueprint for low-cost, real-time analysis to inform public policy for Brownsville, as well as similarly situated neighborhoods in NYC.

Introduction

COVID-19 is a disease resulting from the SARS-CoV2 virus, which has reached pandemic status globally, according to the World Health Organization. Starting in March, New York City became one of the global epicenters. Stay-at-home orders were put in place, forcing businesses to shut down, which increased unemployment rates while decreasing access to essential services for many New Yorkers. Social distancing policies have also been enforced, changing the social behaviors of residents.

Data collected throughout the United States during the pandemic has shown that Black and Hispanic residents are disproportionately affected by Covid-19, a disparity prevalent in NYC as well. According to the CDC, people at higher risk for Covid-19 include those with pre-existing conditions and older adults. According to the NYC Department of Health’s 2015 Community Health Profile, Brownsville ranked near the bottom of several key health indicators. In a community that is 73% Black and 23% Hispanic, average life expectancy is 74.1 years, one of the lowest in the city. Brownsville also ranked 1st in premature mortality rate, with 367.1 premature deaths (before the age of 65) per 100,000 people.For Brownsville residents, who rank 6th in NYC for adult asthma hospitalizations, these results may result in a more vulnerable population than the average NYC neighborhood. Over 60% of the residents have a household income of less than $35,000, with over 43% living in poverty.

Using publicly available data sources, the team will analyze social and economic impacts of Covid-19 on Brownsville to help city government and organizations identify these issues. Factors including 311 calls, transportation, public sentiment dynamics, and environment will be assessed. In collaboration with the non-profit groups Civ:Lab and Public Sentiment, our team will study the behavior of Brownsville residents during this quarantine. Project deliverables include a website and report detailing the research findings. An additional summary memo for policymakers will also be provided.

Literature Review

Studies regarding the socioeconomic impacts of pandemics are scarce but existing studies have shown that low-income populations are more susceptible to epidemics due to crowded and unhygienic living conditions (Bouye, et al). In New York City, reports have shown that low-income people and people of color are at increased risk to Covid-19.

Comparative analysis between neighborhoods in Brooklyn will be conducted to determine the disparate effects of Covid-19. Research has shown that neighborhoods can be characterized by using socioeconomic and demographic data obtained from ACS and 311 data (Wang, et al). Similarities are determined in relation to CD 16, which contains Brownsville, by representing each community district with ACS data and using cosine similarity. The community districts are then ranked and visualized by their similarity to CD 16 (Preotuic-Pietro, et al.).

Social sentiment can be analyzed through 311 and Twitter data. Research has shown that 311 service requests can be used to understand the sociodemographic characteristics of a neighborhood; similarities in 311 service requests correlate to similarities in socioeconomic characteristics of the neighborhood (Wang, et al). Natural language processing and sentiment analysis has been successfully used to measure sentiment on Twitter, categorizing tweets as positive, negative, or neutral (Agarwal, et al.). Sentiment analysis can also be used to quantify the degree of negativity in sentiment (Namugera, et al).

Mobility data is another method to measure the socioeconomic impacts of Covid-19 on a community. Mandates on social distancing have greatly reduced the demand for public transportation and result in a drop in physical and social activity that negatively impact mental and physical well-being (De Vos). Changes to public transportation ridership can be used to understand the well-being and workforce profile of a neighborhood.

The effects of Covid-19 can be exacerbated by a neighborhood's environmental conditions like pollution and lack of access to green space. Long-term exposure to air pollution, specifically PM2.5 particulates, can increase death rates from Covid-19 (Wu, et al.). The pandemic has highlighted the benefits of access to parks and green space, which can improve mental and physical health and mitigate the effects of pollution, but there exists inequalities in access to these parks in New York City (Surico).

Data

The project team analyzed the effects of COVID-19 by using publicly available data that captures different aspects of socioeconomic conditions. COVID-19 data (number of cases, percent positive, and number of deaths) is used in the dependent variable in OLS regressions to determine the impacts from socioeconomic and environmental conditions like race, transportation methods, occupation, air quality,and park space. 311 data is used to determine the effects of COVID-19 on citizen behavior and engagement with the platform. Turnstile entries data from MTA Turnstile Data is used to analyze the impacts of shelter-in-place on subway ridership and to quantify the differences in ridership between different neighborhoods. Twitter data is extracted from 01/01/2020 to 06/30/2020. to measure sentiment of users during the pandemic. The dataset of Brooklyn contains 42,631 tweets, the dataset of Brownsville contains 3,347 tweets. Additional information about the data structure is provided in Table 8.

Datasets	**Source**
MTA Turnstile Data	NY Open Data
Covid-19 Data, Health Facilities	NYC Department of Health
311 Service Requests, Air Quality, Parks, Trees	NYC Open Data
Twitter Data	Twitter, NYU CUSP
American Community Survey	US Census Bureau

Table 1: Summary of datasets and sources

Methodology

In order to gain a clearer picture of the behavior of Brownsville residents, analysis was split into two parts: descriptive and predictive. The descriptive stage included visualizing the datasets as presented. Cosine similarity, the cosine of the angle between two n-dimensional vectors in an n-dimensional space, is used to rank the most and least similar neighborhoods to Brownsville. The results of the cosine similarity inform which community boards will be compared to analyze the disparate impacts of COVID-19 on different neighborhoods.

Ordinary Least Square (OLS) multivariate regression was used to determine the effects of urban factors on percent positive cases, rate of positive cases, and rate of death of COVID-19 cases. The OLS regression was performed on three distinct categories of features to allow for better interpretability: occupation, transportation methods, and environmental, neighborhood factors. OLS regression allows for easily interpretable results and a means of comparison between the different features.

Temporal analysis focused on the period from March 1 to June 30 2020, compared against March 1 to June 30 2019 to see the impacts of Covid-19 on 311 and mobility. Spatial analysis used Geographic Information Systems (GIS) and geopandas to compare and contrast the disparities in environmental conditions and impacts of Covid-19 across Brooklyn neighborhoods. This included the percentage change in 311 service requests per zip code, the lowest level of granularity available.

TWINT was used to scrape COVID-19 related Twitter data by searching keywords near Brooklyn: covid OR covid-19 OR COVID-19 OR Coronavirus OR Corona. The aim was to see how the public sentiment to COVID-19 changes over time. Text data analysis was performed to remove miscellaneous symbols, URL, and account information from text. Stop words are removed as they do not have impact on the text meaning. Sentiment analysis was conducted on the processed data to try to understand people's opinions expressed by the text. Using a sentiment analysis model from Kaggle, the sentiment behind each Tweet was quantified with a numerical score. A score between 0 and 0.3 indicates overall negative sentiment, between 0.3 and 0.7 indicates overall neutral sentiment, and above 0.7 indicates overall positive sentiment.

Results

Health and Environment

The geospatial visualizations in the Appendix reveal the disparities between Brooklyn neighborhoods and OLS regressions reveal the impact of different features on COVID-19 cases. CD16, which contains Brownsville, and CD17, the community district most similar to CD16, have more cases, deaths, and greater percentage of positive tests than CD 6, the community district least similar to CD16. The independent and dependent variables regression are summarized in Table 6.

The transportation regression in Table 3 shows that working at home has a negative impact on all the analyzed variables of COVID-19, making it the most effective commute method to avoid COVID-19. The occupation regression in Table 4 shows that service workers are least affected by COVID-19, likely due to their status as non-essential workers. Essential workers like maintenance and transportation personnel show a higher chance of testing positive for COVID-19. The environmental regression in Table 5 shows that median income of a neighborhood has an inverse relationship to the three features of COVID-19 that were analyzed. Park space and density of trees also have inverse relationships to COVID-19. Many of the environmental factors can be related to the wealth of the neighborhood, for example, wealthier neighborhoods have more access to parks, green space and medical facilities.

311 Analysis

Between the dates of March 1 and June 30, comparing 2019 to 2020, there was an approximately 7% percent decrease in the number of 311 calls made in Brownsville. This is consistent with the wider trend of a call decrease in the entire borough of Brooklyn (~10%), as well as New York City (~18%). Choropleth maps highlight this fact (Figure 17). Time series analysis indicated that the trend of percent decrease in 311 calls is consistent. Starting in June, however, the number of calls in 2020 increased quite drastically, correlated with the preponderance of fireworks throughout the city (Figure 18).

For Brownsville, distinct changes were observed in the complaint types (Figure 19). For 2019, the top call types included “Heat-Hot Water”, “Noise-Residential”, “Unsanitary Condition”, and “Plumbing”. For 2020, the number one call type was “Noise - Residential”, accounting for over 25% of calls. Additionally, there was the emergence of two new categories: “Non-Emergency Police Matter” (~15%), and “Illegal Fireworks”(~8%).

Noise complaints seem fairly self explanatory. The calling of police for non-emergency matters may be correlated with the rise in cases of Covid-19, as residents look to police departments for assistance with sick family members. Additionally, the spike in consumer complaints could potentially be correlated with the rise in “price gouging” among businesses for essential supplies.

These results also match the findings that the New York City Police Department (NYPD) responded to over 50% of all 311 calls in 2020 for the time period mentioned above. They were followed by the Department of Housing Preservation and Development (25%), and Consumer Affairs (7%) (Figure 20).

Transportation

The transportation analysis for Brownsville shows two drastic decreases in ridership; one on the week of March 16, corresponding to public school closures, and the second one the following week, on the week of March 23, corresponding to stay-at-home orders. While the same decreasing trend can be observed in CD17 and CD6, the extent of decrease is different. The percent difference in ridership is quantified by comparing ridership of March to June 2020 against March to June 2019 for the three community boards shown in Table 7. CD16 and CD17, which are majority low-income, people of color communities, have similar COVID-19 cases and decreases in ridership, -51% and -55% respectively. This is in contrast to CD6 which has more affluent, white residents, significantly less COVID-19 cases, and a larger decrease in ridership of -70.5%. The results indicate that residents of low-income neighborhoods, who make up a large majority of essential workers, lack the ability to work at home and must rely on public transportation to get to work. The OLS results show that these occupations and reliance on public transportation put these residents at increased risk to COVID-19.

Twitter Data Analysis

The trend of weekly number of tweets posted within Brooklyn was shown in Figure 26. Regarding the number of tweets, it shows that before the advent of the pandemic, there are less tweets posted. While entering into the 11th week, the number of tweets has a huge increase in the next two weeks and then decreases until 22th week. After 22th, the number of tweets increases again while NYC enters into the reopening phase. Figure 27 shows the correlation between the number of real infected cases and the number of tweets published each day. The observed positive trend for the correlation is confirmed by the coefficient equal to 0.315.

Figure 28 summarizes the number of top hashtags for Brooklyn. The hashtags related to search term such as #COVID-19, #COVID-19, #COVID, #Coronavirus and #Corona have been removed as these keywords served for data extractions. These top hashtags revealed some events or situations happening during the pandemic such as stay at home, quarantine, save the world, billionshields (a challenge that uses recycle empty soda bottles into face shields to against COVID-19) etc.

Figure 29 shows the 41% tweets are neutral, 27% show negative emotions and 32% shows positive emotions. Combined with the number of tweets, the Figure 30 shows the number of tweet trends by sentiment category during the 1H of 2020.

Based on word frequencies for positive tweets and negative tweets separately, word cloud for each of the two categories was drawn in Figure 31 and 32. The positive tweets contain positive words like “new york”, “great”, “thank” while the negative tweets are mainly expressing worries about people’s health, need and the virus trend.

Figure 33, 34, 35, 37 represent the tweets analytical Insight of Brownsville. Though the amount of tweets is limited, the overall sentiment in Brownsville tends to be more positive.

Limitations and challenges

There was a lack of availability of data and a condensed timeline for this project. Economic data (i.e. unemployment) was not publicly available and thus could not be included. Twitter data related to Covid-19 in Brownsville was limited, therefore, it was hard to get accurate emotions of how Brownsville residents react to the pandemic. Additionally, since COVID-19 data was reported on the zip code level, the OLS regression had to be analyzed on the zip code level which could lead to some inaccuracies; neighborhood tabulation or census tract would have been preferred.

Unfortunately, the initial purpose of the project was also not fully implemented; our team was to use the insights of this analysis to compare against a survey taken by Brownsville residents. As the survey didn't go out, we were left with the analysis from these publicly available datasets, without direct input from residents. Future projects could incorporate resident concerns into the analysis.

Conclusion

The results from this project supports many of the findings reported by the news and scientific community. Low-income communities of color are disproportionately affected by Covid-19, with more overall cases, more deaths, and a higher percentage of positive cases. The effects of Covid-19 are further exacerbated by transportation methods, occupations, and environmental factors like lack of access to green space and parks. Analysis of 311 shows a consistent trend of less use of government services, with residents shifting concern to noise and Covid-19 related issues.

Further research would entail studying 311 calls by time of day, and studying the content of each call for sentiment analysis and comparing it with Twitter data. For future Twitter data extraction improvements, the study can be focused on building a config based pipeline: a) crawl Twitter data by location/search-term two dimensions, b) after data collection, enable auto trigger the model evaluation and result forward. Furthermore, explore the possibility to construct the feedback loop based on the model evaluation results, which can drive crawlers to get more substantial data. Other avenues for future analysis can include economic data like regarding unemployment and small businesses.

Actions can be taken by individuals and the city to become more resilient to the effects of Covid-19. Individuals can follow CDC guidelines to wear masks and maintain social distancing.

Employers should allow employees to work remotely and allow flexible work hours or staggered work schedules, only requiring employees to work in the office if necessary. The city can create more open, green space by expanding parks, opening streets for pedestrians, and planting more trees. Parks and sidewalks can be used as outdoor work spaces. The city can create a platform that allows citizens to voice their needs in real-time, allowing government agencies to plan and allocate resources for communities with reduced delays.

References

Agarwal, Apoorv, et al. “Sentiment Analysis of Twitter Data.” *Proceedings of the Workshop on Language in Social Media (LSM 2011)*, 2011, www.aclweb.org/anthology/W11-0705.pdf.

Bouye, Karen, et al. “Pandemic Influenza Preparedness and Response Among Public-Housing Residents, Single-Parent Families, and Low-Income Populations.” *American Journal of Public .Health*, U.S. National Library of Medicine, 2009, pubmed.ncbi.nlm.nih.gov/19797740/.

Braveman, P., et al. “Building a Culture of Health at the Neighborhood Level Through Governance Councils.” *Journal of Community Health*, Springer US, 2020, link.springer.com/article/10.1007/s10900-020-00804-0.

King L, Hinterland K, Dragan KL, Driver CR, Harris TG, Gwynn RC, Linos N, Barbot O, Bassett MT. Community Health Profiles 2015, Brooklyn Community District 16: Brownsville; 2015; 40(59):1-16.

Namugera, Frank, et al. “Text Mining and Determinants of Sentiments: Twitter Social Media Usage by Traditional Media Houses in Uganda.” *Computational Social Networks*, 2019, link.springer.com/article/10.1186/s40649-019-0063-4.

Preotiuc-Pietro, Daniel, et al. “Exploring Venue-Based City-to-City Similarity Measures.” *University of Pennsylvania* , UrbComp, 2013, www.sas.upenn.edu/~danielpr/files/cities13urbcomp.pdf.

Surico, John. “In a Pandemic, the Parks Are Keeping Us Alive.” *Bloomberg.com*, Bloomberg, 9 Apr. 2020, www.bloomberg.com/news/articles/2020-04-09/in-a-pandemic-the-parks-are-keeping-us-alive.

Vos, Jonas De. “The Effect of COVID-19 and Subsequent Social Distancing on Travel

Behavior.” *Transportation Research Interdisciplinary Perspectives*, 2020, www.ncbi.nlm.nih.gov/pmc/articles/PMC7180344/.

Wang, Lingjing, et al. “Structure of 311 Service Requests as a Signature of Urban Location.” *PLOS ONE*, Public Library of Science, 2017, journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0186314.

Wu, Xiao, et al. “Exposure to Air Pollution and COVID-19 Mortality in the United States: A Nationwide Cross-Sectional Study.” *MedRxiv*, Cold Spring Harbor Laboratory Press, 27 Apr. 2020, www.medrxiv.org/content/10.1101/2020.04.05.20054502v2.

Data Sources

Health Facility General Information (New York State Wide)

https://health.data.ny.gov/Health/Health-Facility-General-Information/vn5v-hh5r/data

NYC Open Data Portal

https://opendata.cityofnewyork.us/

NYC Health Covid-19 Data Archive

https://www1.nyc.gov/site/doh/covid/covid-19-data-archive.page

https://github.com/nychealth/coronavirus-data

NYC Turnstile Data 2020

https://data.ny.gov/Transportation/Turnstile-Usage-Data-2020/py8k-a8wg

NYC Population Finder

https://popfactfinder.planning.nyc.gov/#12.25/40.724/-73.9868

NYC Community Board Profiles

https://communityprofiles.planning.nyc.gov/

Subway stations

https://data.cityofnewyork.us/Transportation/Subway-Stations/arq3-7z49

Twitter Data

Extracted from https://twitter.com/explore though Twint

Twitter Sentiment Analysis

https://www.kaggle.com/paoloripamonti/twitter-sentiment-analysis/notebook

Appendix

Team Member	Roles and Responsibilities
Earl Lin	Project coordination; geospatial, health, and environmental data analyses; socioeconomic research; literature review; memo
Jerome Louison	Project coordination; 311 data analysis; literature review; memo
Yushu Rao	Twitter data analysis; literature review; Github
Xinran Zhao	Website lead developer; transportation analysis; literature review

Table 1: Team Members Roles and Responsibilities

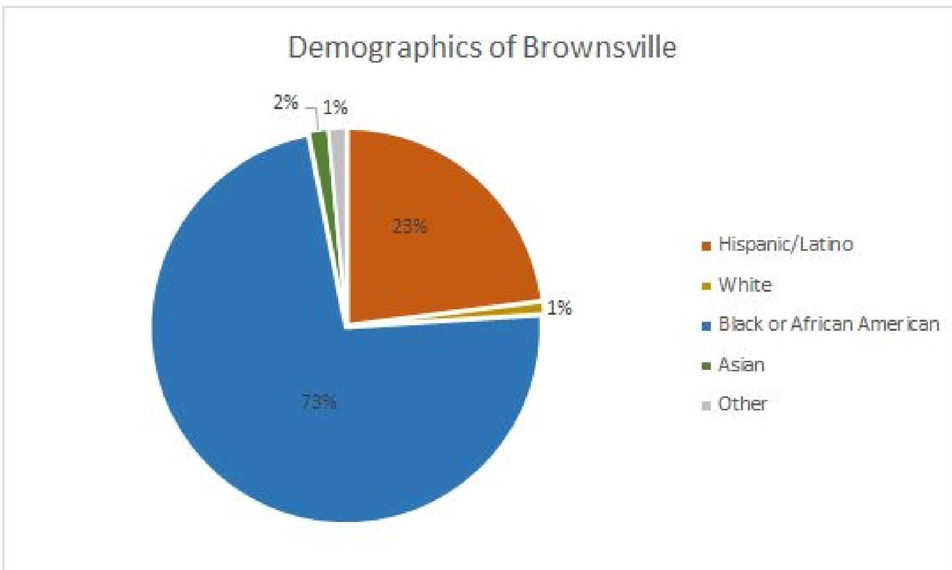

Figure 1: Demographics of Brownsville
Source: NYC Planning Population FactFinder

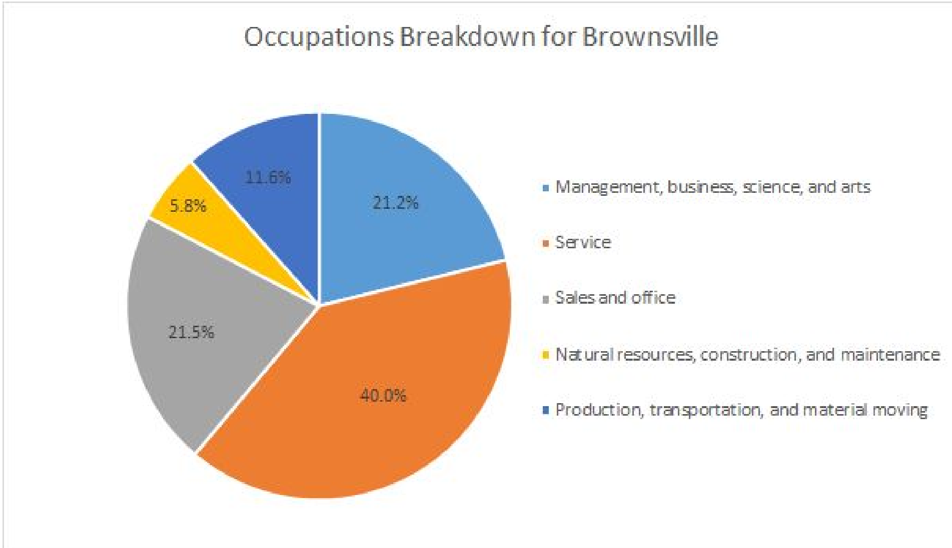

Figure 2: Occupations Breakdown for Brownsville
Source: NYC Planning Population FactFinder

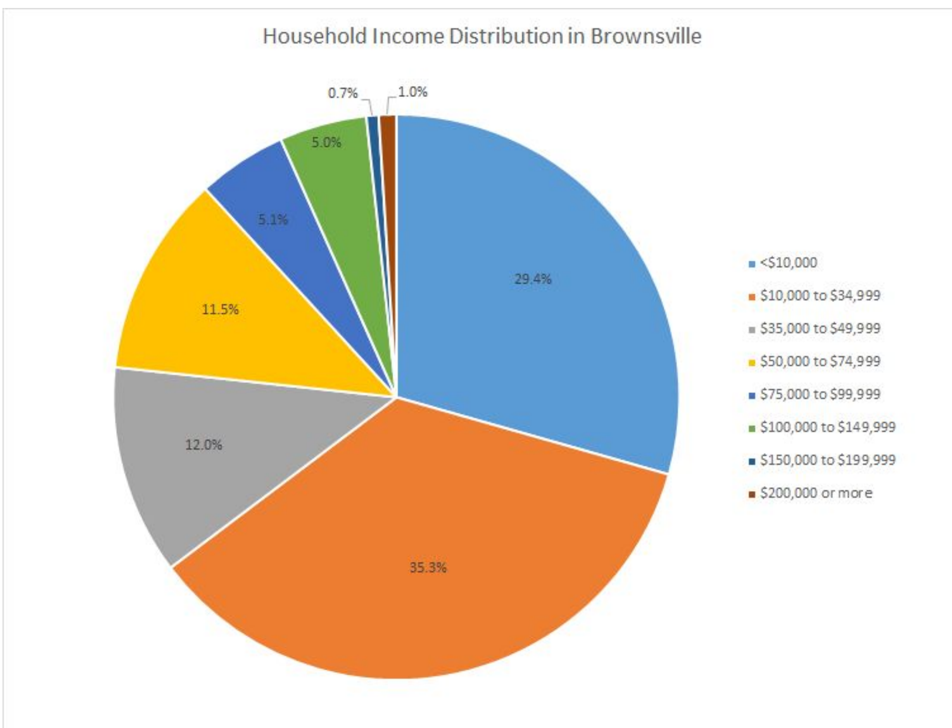

Figure 3: Household Economic Distribution in Brownsville
Source: NYC Population FactFinder

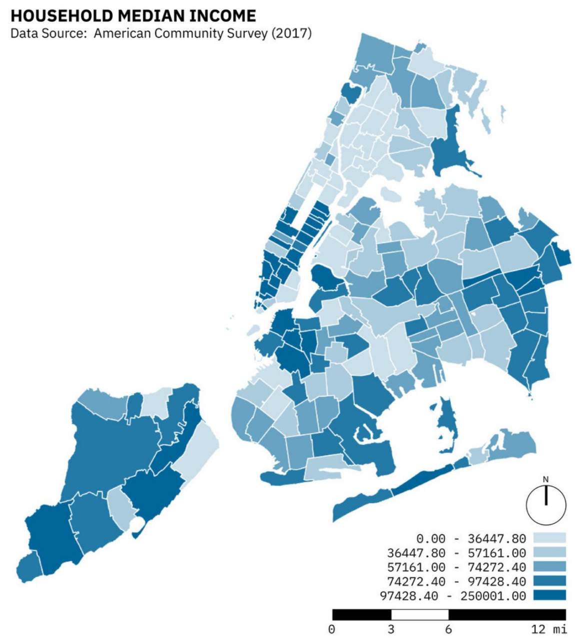

Figure 4: Median Household Income by NYC zip code
Source: Urban Systems Lab

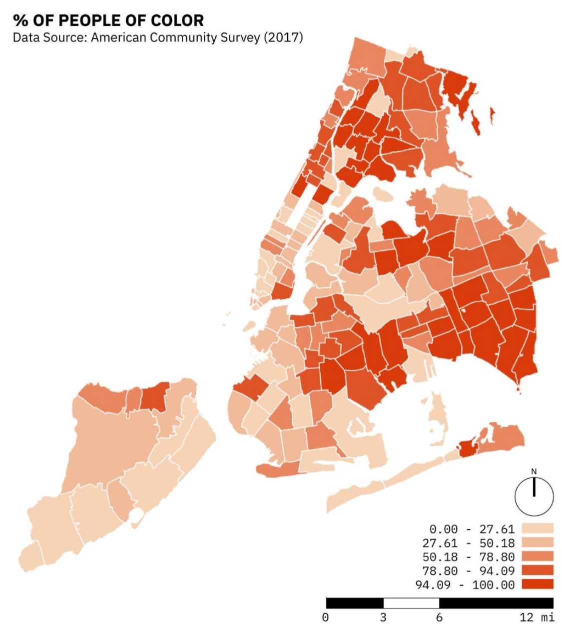

Figure 5: % People of Color by NYC zip code
Source: Urban Systems Lab

Deadline	Status/Notes	Task/Deliverable/Milestones	T/D/M	Responsibility
27-Jul	Not Started	Final NYU-CUSP Report and Presentation	D	Team
27-Jul	Not Started	Complete Visualization Tool	T	Xinran
27-Jul	Not Started	Complete final memo deliverable to clients	D	Earl & Jerome
20-Jul	Not Started	Draft final memo deliverable to clients	M	Earl & Jerome
20-Jul	Not Started	Draft final report and final presentation	M	Team
14-Jul	Not Started	Start creating Github Repository	M	Yushu
14-Jul	Not Started	Start building website and visualization	M	Xinran
14-Jul	In Progress	Complete all analysis	M	Team
TBD	Not Started	Incorporate Survey Results	T	Yushu
22-Jun	Completed	Midpoint Progress Report	D	Team
27-May	In Progress	Complete list of (non-open source) additional data sets	M	Earl & Jerome
13-May	In Progress	Meeting about Next Steps (with all teams)	T	Team

Table 2: Project Schedule

Cosine Similarity

The socioeconomic features used for cosine similarity include percent of racial composition (black, white, hispanic, asian, other), poverty rate, percent of population with a bachelor's degree, poverty rate, crime rate, and unemployment rate. Community District 16 (where Brownsville is located in) is most similar to Community Districts 5 and 17 (ex. East New York and Flatbush) and most dissimilar to Community Districts 6 and 12 (ex. Carroll Gardens, Park Slope, and Borough Park).

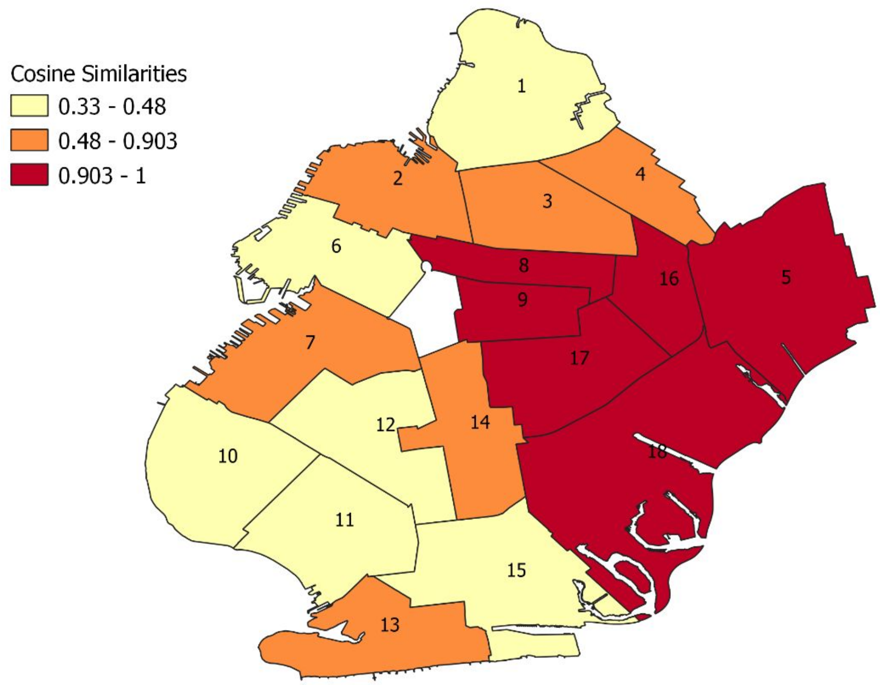

Figure 6: Cosine similarity of Brooklyn Community Districts

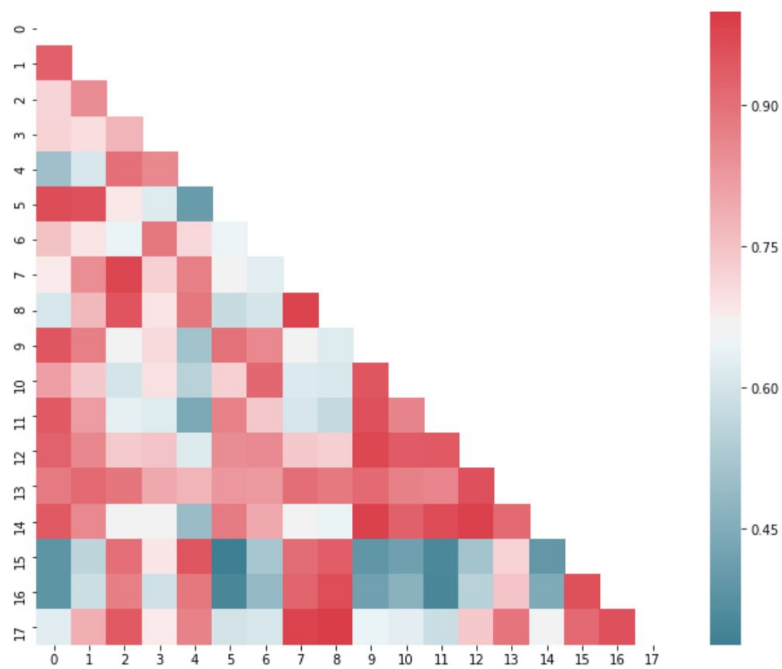

Figure 7: Cosine similarity matrix for Brooklyn Community Districts

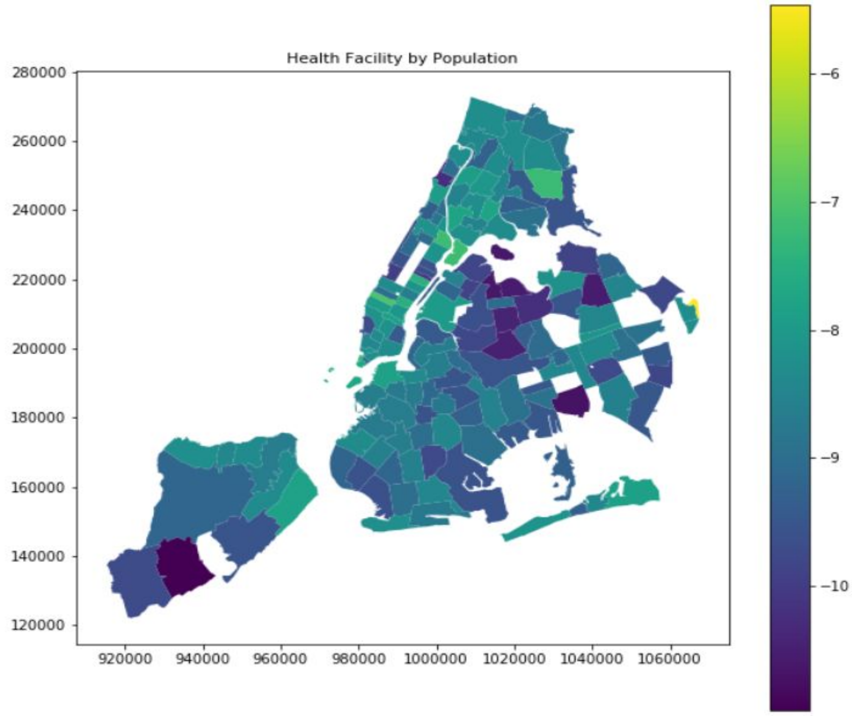

Figure 8: Heat Map showing the health facility per 100,000 people

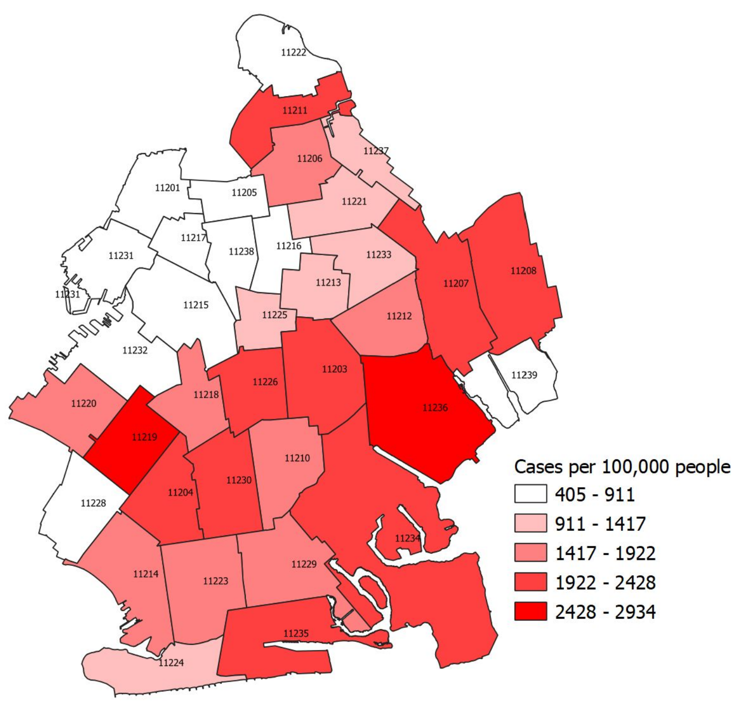

Figure 9: Cases per 100,000 people by zip code

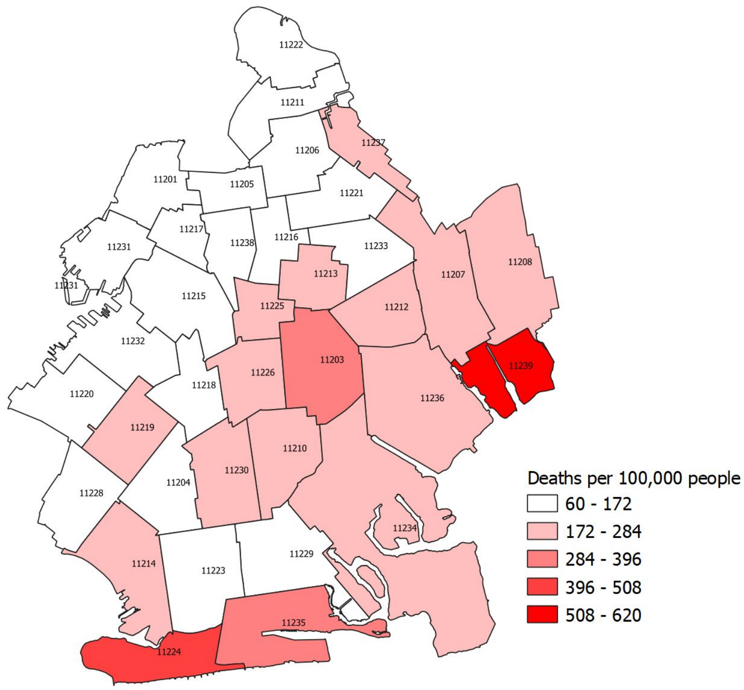

Figure 10: Deaths per 100,000 people by zip code

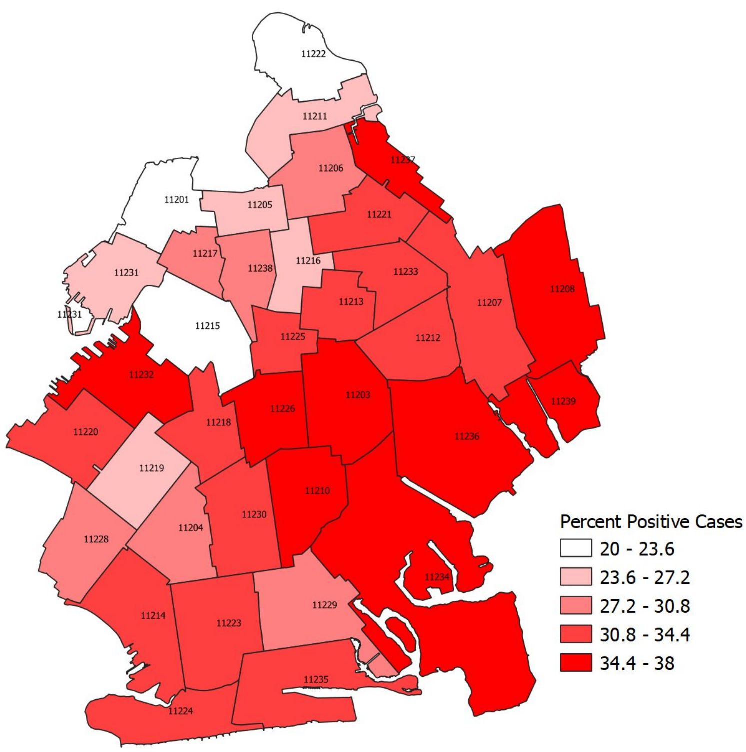

Figure 11: Percent positive cases by Zip Code

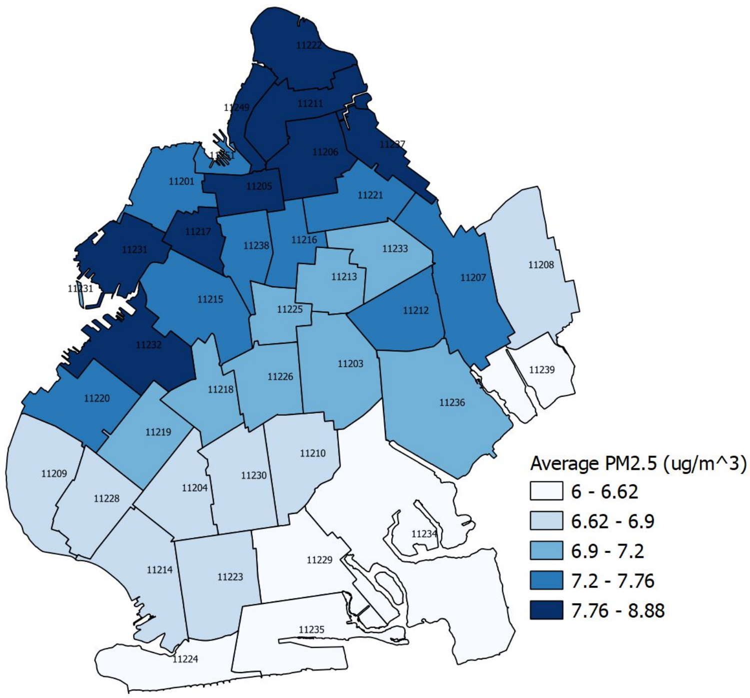

Figure 12: Average levels of PM2.5 by Zip Code

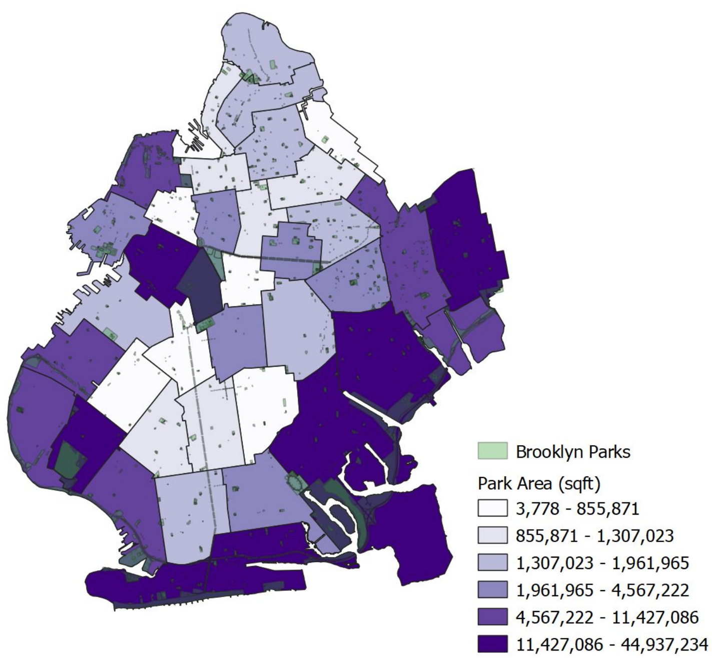

Figure 13: Park Area by Zip Code

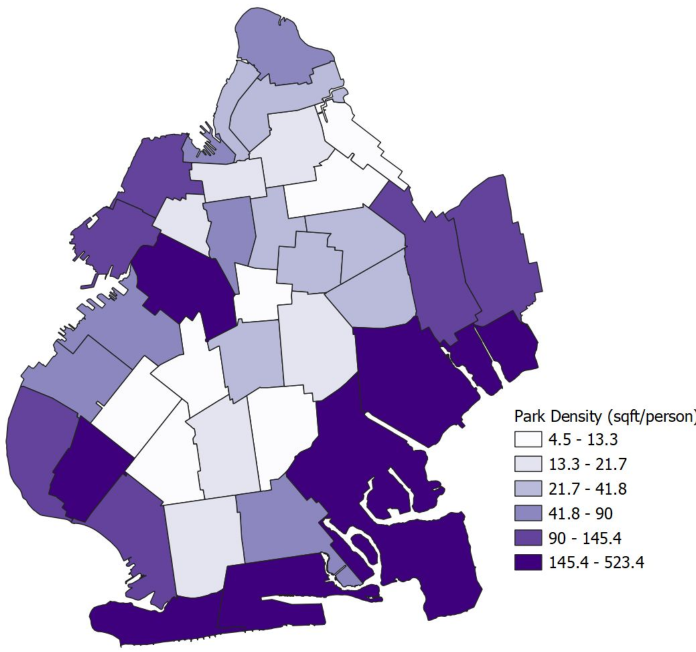

Figure 14: Park Density by Zip Code

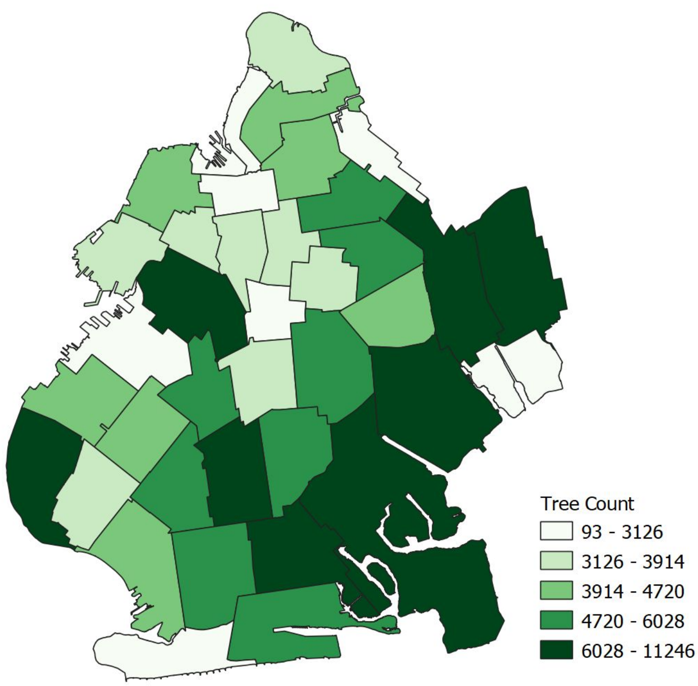

Figure 15: Street Tree count by Zip Code

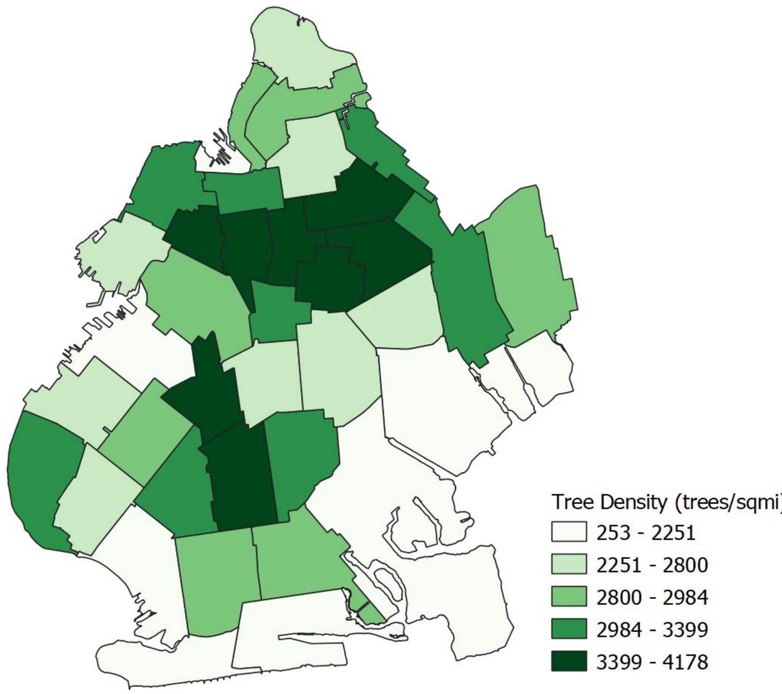

Figure 16: Street Tree Density by Zip Code

	Percent Positive		COVID-19 Case Rate		COVID-19 Death Rate	
Features	coef	p-value	coef	p-value	coef	p-value
% drove alone	0.2949	0.000	41.4305	0.000	2.8930	0.000
% public trans	0.2615	0.000	33.1274	0.000	3.7185	0.000
% walked	0.1105	0.002	19.1064	0.004	1.5441	0.112
% worked at home	-1.1313	0.000	-154.0118	0.000	-19.5727	0.000
# of obs.	177		177		177	
R-squared	0.967		0.936		0.841	

Table 3: OLS Regression for Commute to Work Method for New York City zip code

	Percent Positive		COVID-19 Case Rate		COVID-19 Death Rate	
Features	coef	p-value	coef	p-value	coef	p-value
% Service	0.1705	0.000	30.2496	0.000	5.5963	0.000
% Sales, Office	0.3742	0.000	47.4910	0.000	3.6623	0.000
% Natural resources, construction, maintenance	0.4429	0.000	57.7564	0.005	-2.1314	0.491
% Production, transportation, material, moving	0.4170	0.000	42.8633	0.062	2.6059	0.449
# of obs.	177		177		177	
R-squared	0.979		0.946		0.856	

Table 4: Occupation OLS Regression for New York City zip code

	Percent Positive		COVID-19 Case Rate		COVID-19 Death Rate	
Features	coef	p-value	coef	p-value	coef	p-value
Park Area	-11.1931	0.043	-3512.2811	0.000	-484.0154	0.001
Park Density	7.85e+05	0.004	2.043e+08	0.000	3.877e+07	0.000
PM2.5	1.9681	0.000	200.1093	0.004	17.3611	0.116
Street Tree Count	0.0024	0.000	0.4284	0.000	0.0269	0.057
Street Tree Density	-0.0032	0.031	-0.6484	0.005	-0.0513	0.160
Population Density	0.0002	0.007	0.0396	0.002	0.0030	0.137
Medical Facilities	-0.0230	0.761	6.6049	0.560	2.5819	0.560
311 Call Volume	-0.0006	0.139	-0.1014	0.071	0.0053	0.560
Median Household Income	-9.78e-05	0.003	-0.0117	0.14	-0.0020	0.010
# of obs.	44		44		44	
R-squared	0.975		0.965		0.918	

Table 5: Environmental and Select Socioeconomic Factors OLS Regression for Brooklyn zip code

Feature	Unit
Percent Positive	%
COVID-19 Case Rate	Cases per 100,000 people
COVID-19 Death Rate	Deaths per 100,000 people
% drove alone	%
% public trans	%
% walked	%
% worked at home	%
% Service	%
% Sales, Office	%
% Natural resources, construction, maintenance	%
% Production, transportation, material, moving	%
Park Area	Square miles
Park Density	Square mile / population of associated zipcode
PM2.5	ug/m^3
Street Tree Count	Number of street trees
Street Tree Density	Number of street trees / square mile of zip code
Population Density	Number of people / square mile of zip code
Medical Facilities	Number of medical facilities / 100,000 people
311 Call Volume	Number of 311 Calls in 2020
Median Household Income	$

Table 6: Summary of all features used in OLS Regression and their units

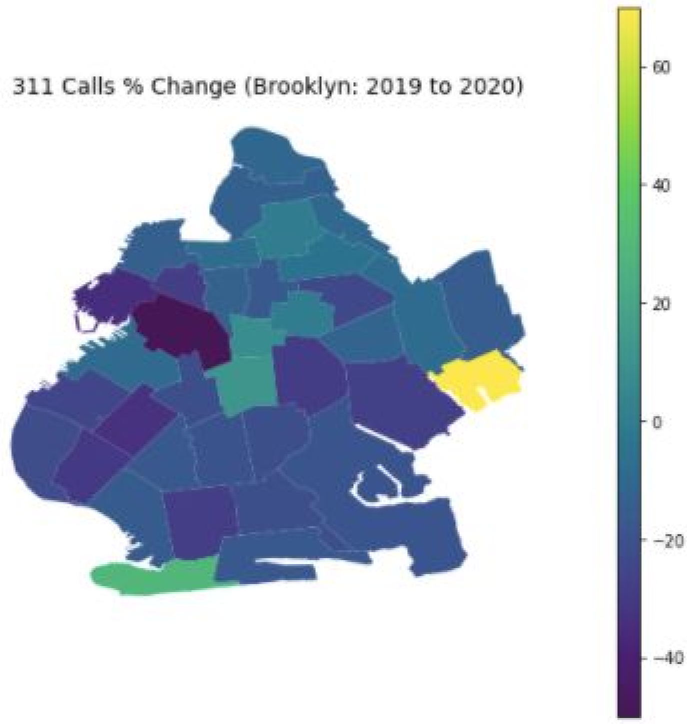

Figure 17: Choropleth map indicating % change in NYC 311 calls by zip code March-June 2019 to 2020 (Brownsville zip code highlighted in white circle).
Colors range from dark purple (percent decrease) to light yellow (percent increase)

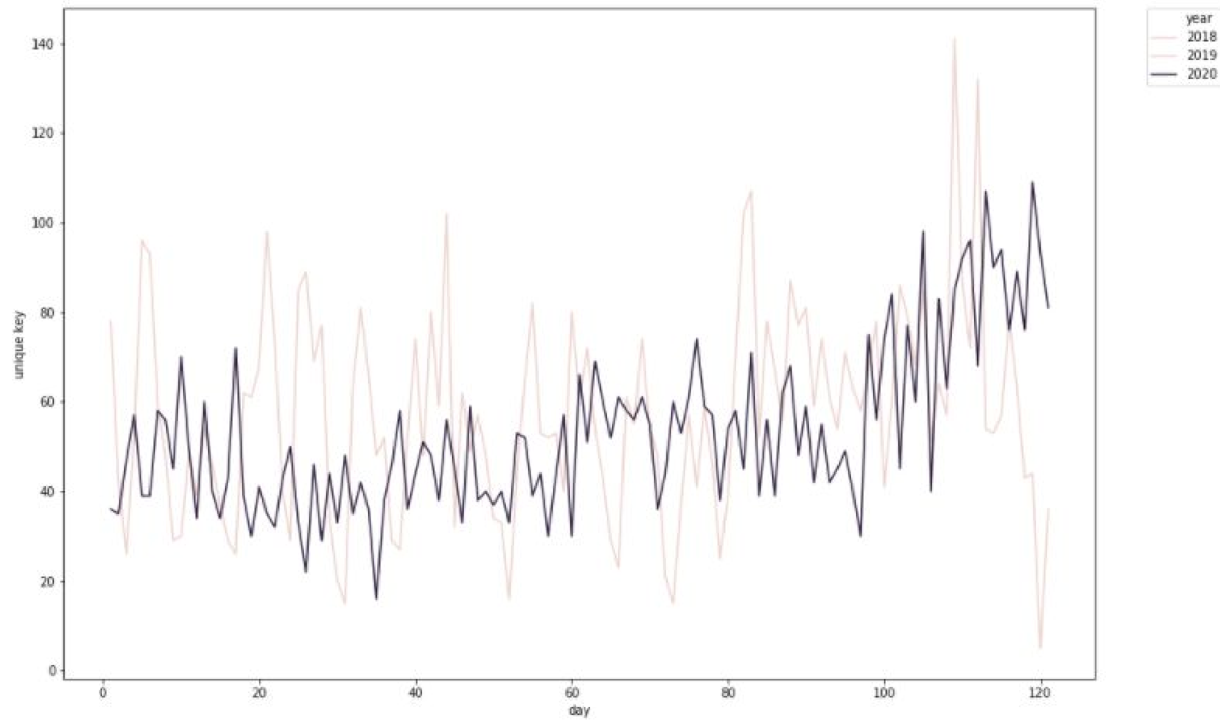

Figure 18: Time series chart showing daily 311 calls in Brownsville, Brooklyn March-June 2019 vs 2020

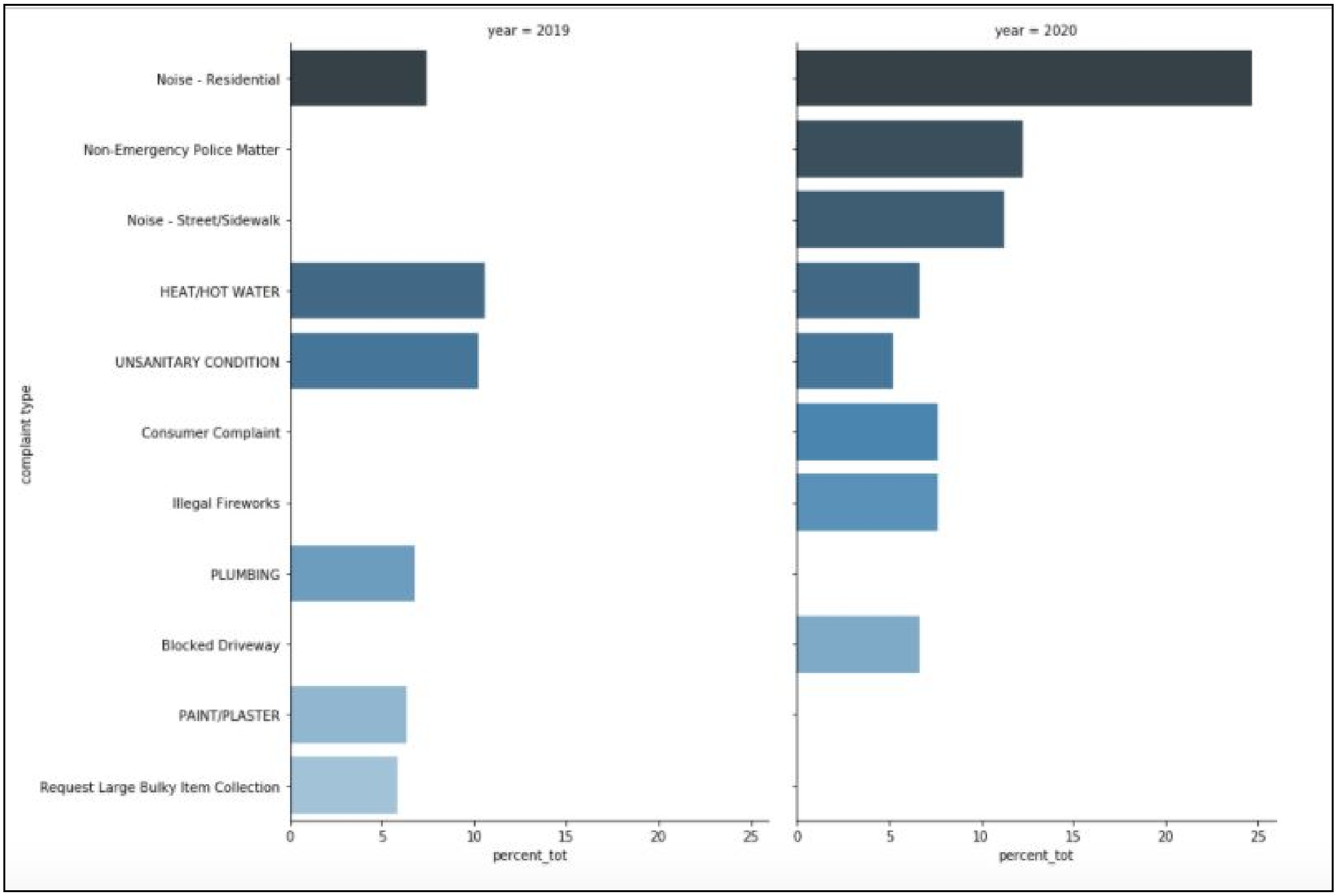

Figure 19: Bar chart for % of total 311 calls by complaint type for Brownsville March-June, 2019 (left) vs 2020 (right)

Figure 20: Bar chart for % of total 311 calls by department for Brownsville March-June, 2019 (left) vs 2020 (right)

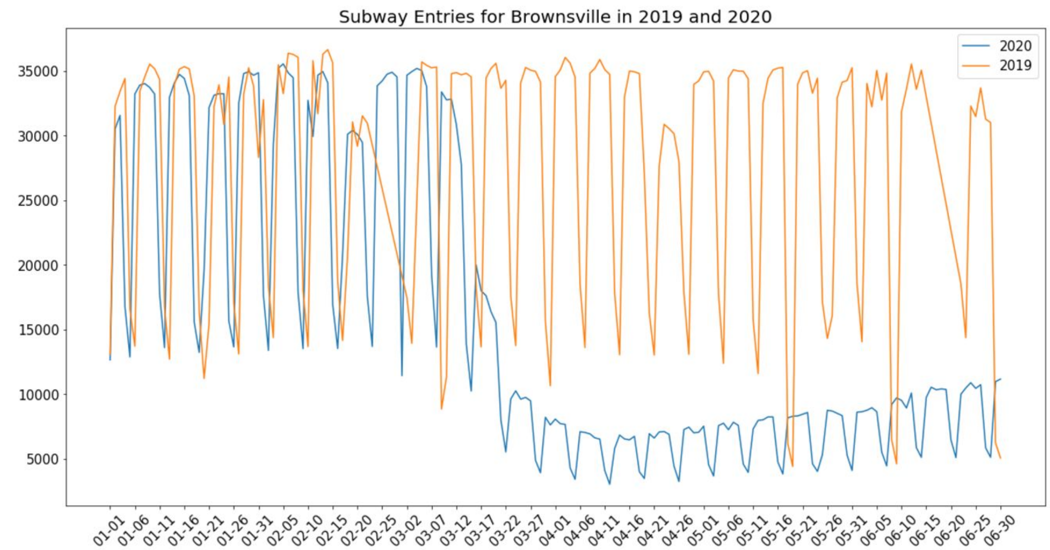

Figure 21: Total Entries for Brownsville in 2019 and 2020

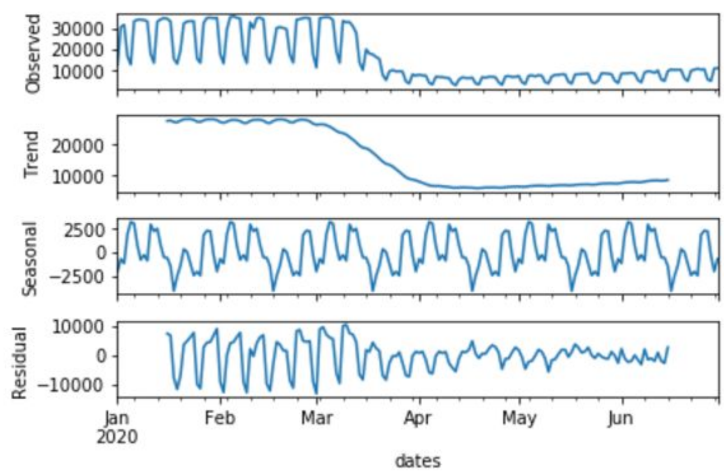

Figure 22: Time Series of Brownsville Subway Ridership

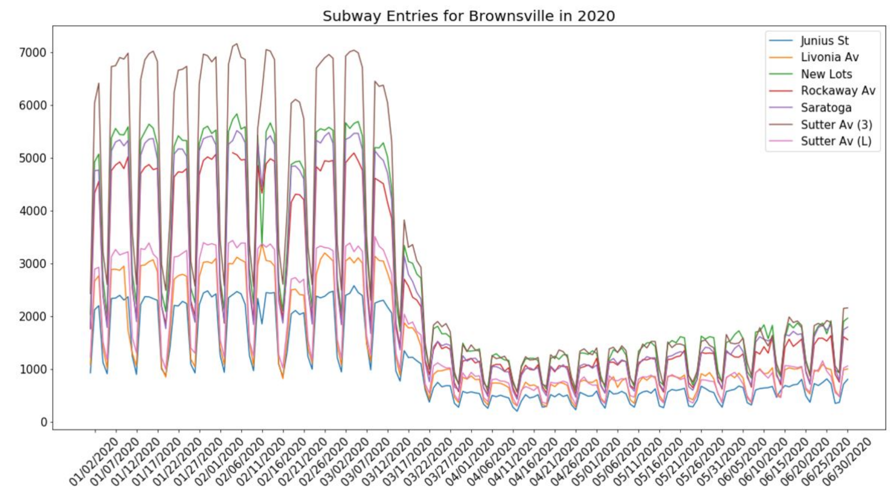

Figure 23: Entries for Brownsville Subway Stations in 2020

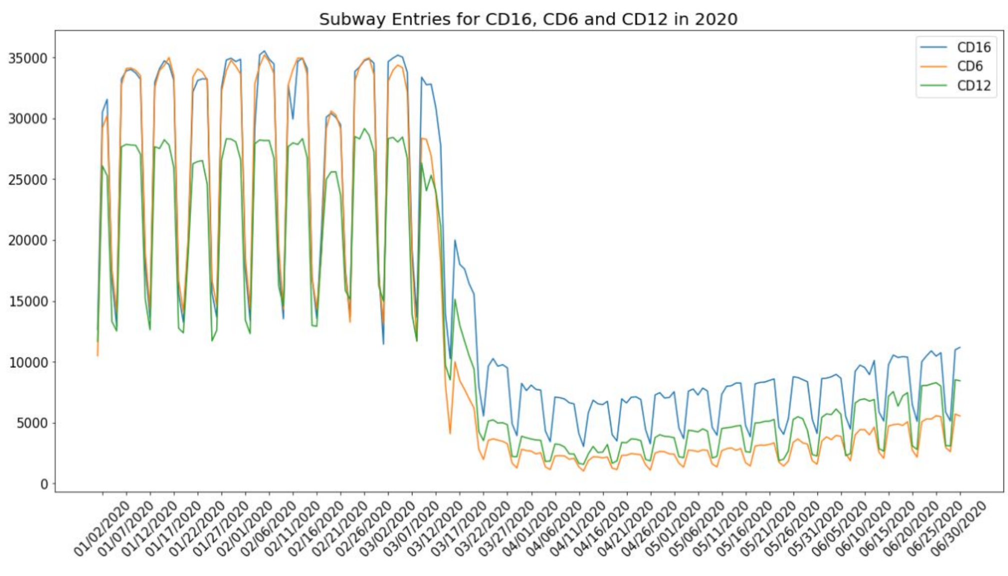

Figure 24: Entries for CD16 (Brownsville), CD6 and CD12 in 2020

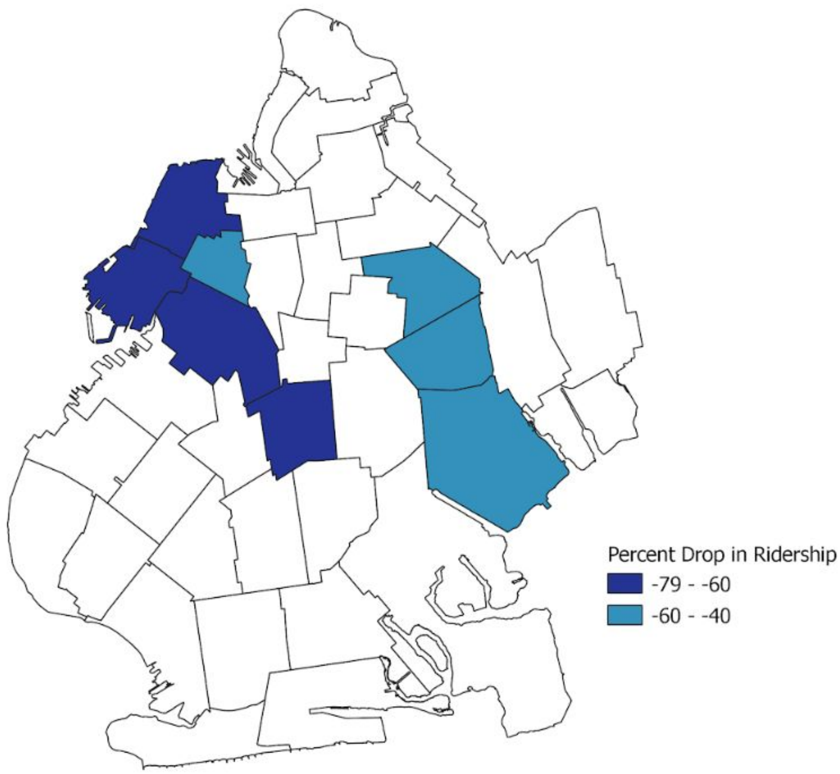

Figure 25: Percent Drop in Ridership for zip codes in select Community Boards

	CD16	CD17	CD6
Average Change	-51%	-55%	-70.5%
Average % Positive	20.33%	22.32	9.63
Average COVID-19 Rate	2135.69	2745.56	1166.26
Average COVID-19 Death Rate	236.11	311.64	120.38

Table 7: Subway Ridership Change Compared to COVID-19 Data by Community Board

id	date	time	timezone	user_id	username	tweet	hashtags	geo
1278091980535470000	6/30/20	15:24:08	PDT	23037281	iddyeganthor	ue #aphrodite üíô @ Clown Heigh	['#aphrodite']	40.6652,-73.9125,1.6k
1278030761992410000	6/30/20	11:20:52	PDT	1521356120	oods_valenc	:OCK #QUEEENOFMASSAGE #Ql	eepdimples', '#dimplesrc	40.6652,-73.9125,1.6k
1278029606872920000	6/30/20	11:16:17	PDT	40626897	nizthemoney	atch What The ü§¨ I Do,Äù by @sr	['#nowstreaming']	40.6652,-73.9125,1.6k

Table 8: Illustration of the twitter dataset structure

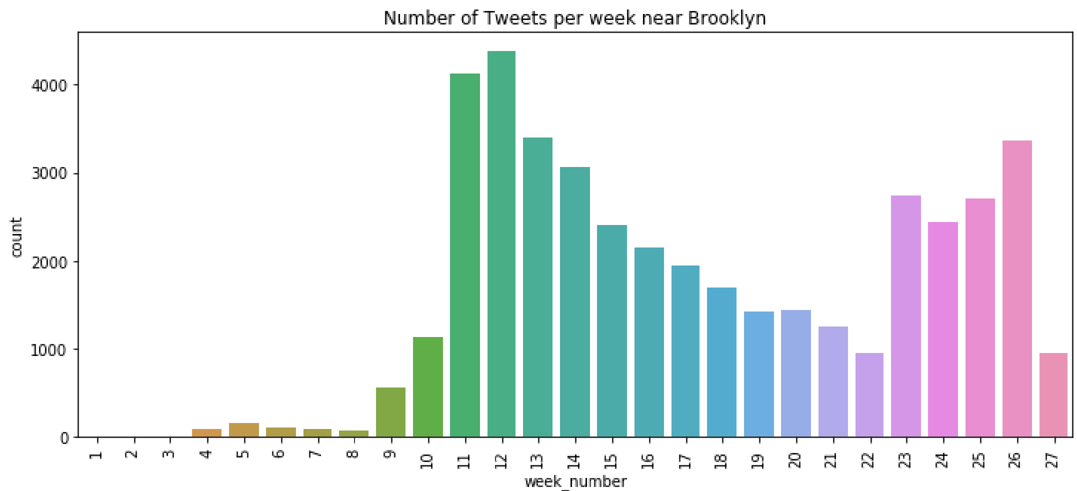

Figure 26: Number of Tweets Per Week Within Brooklyn

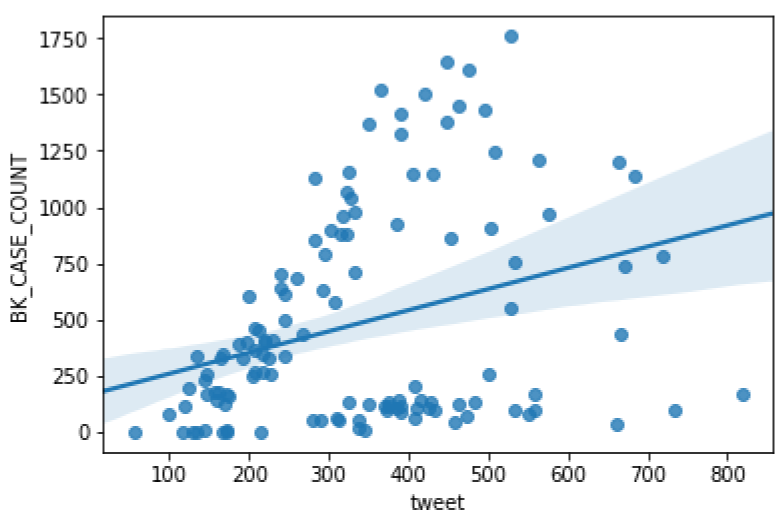

Figure 27: Correlation between the number of new cases and the number of tweet in Brooklyn

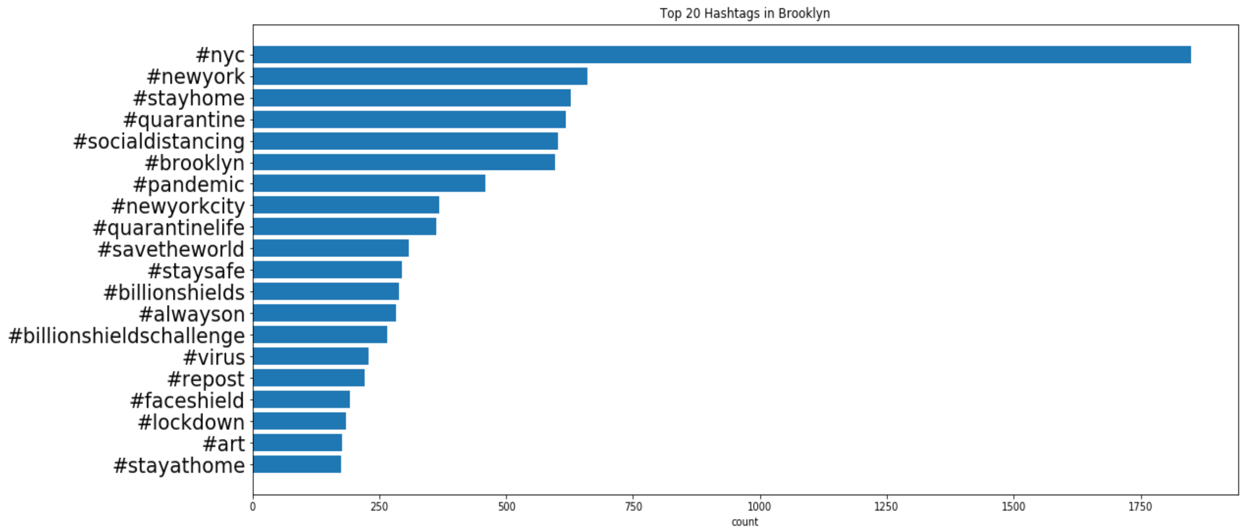

Figure 28: Top 20 Hashtags Within Brooklyn

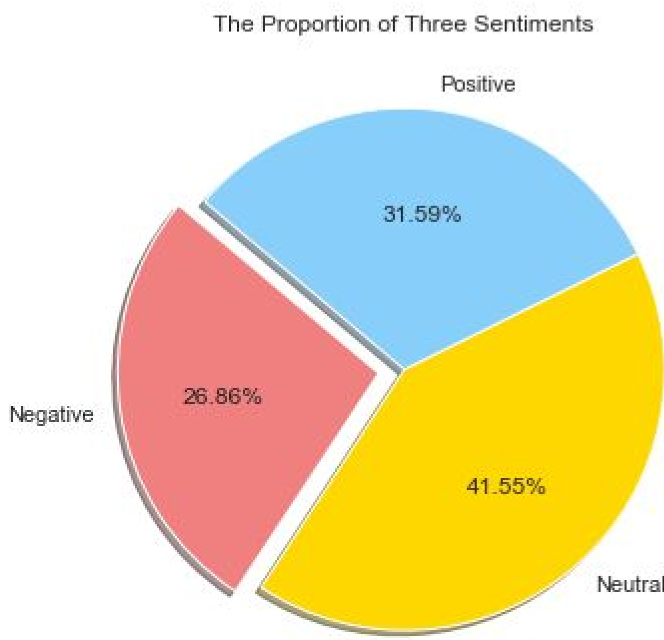

Figure 29: The Proportion of Three Sentiments Within Brooklyn

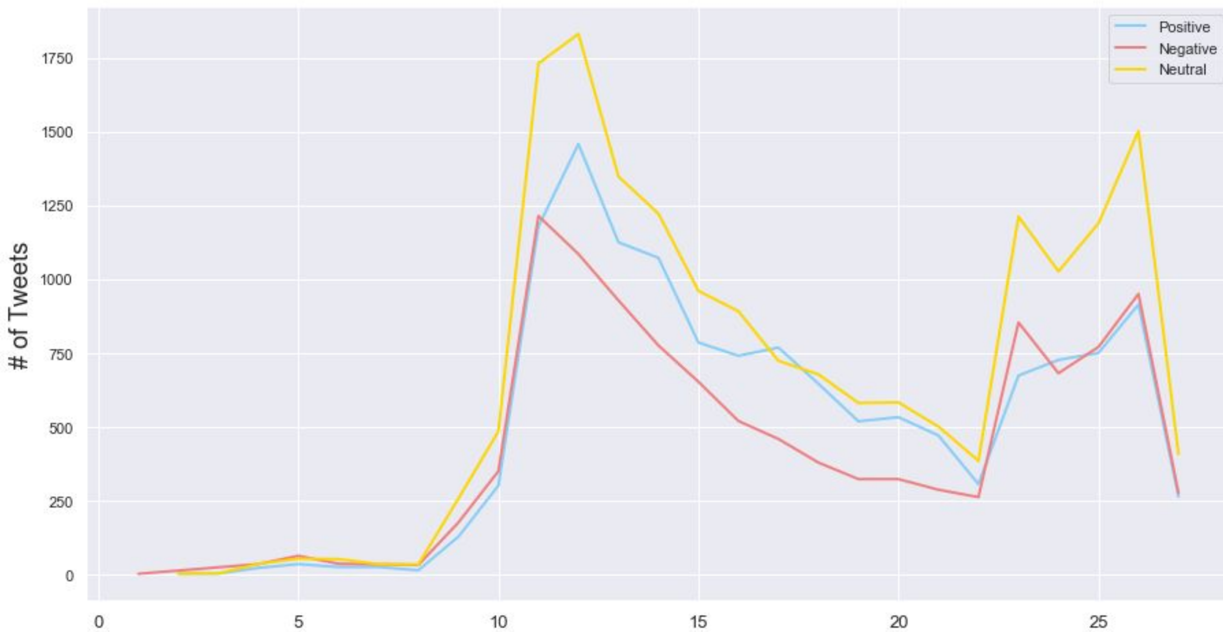

Figure 30: Number of Tweets By Sentiment Per Week Within Brooklyn

Figure 31: Word Clouds in Positive Category

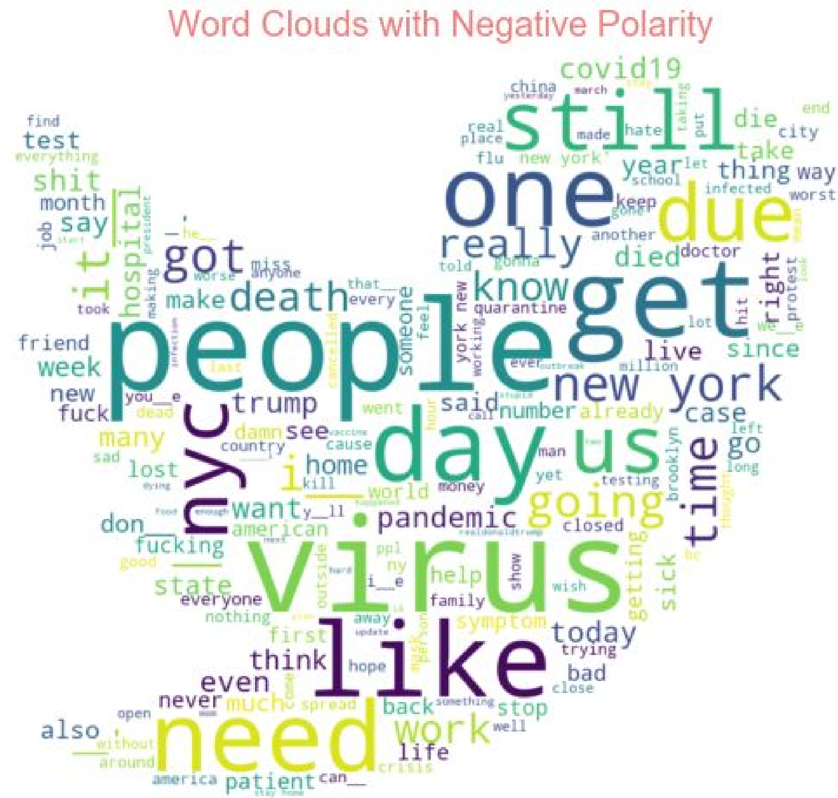

Figure 32: Word Clouds in Negative Category

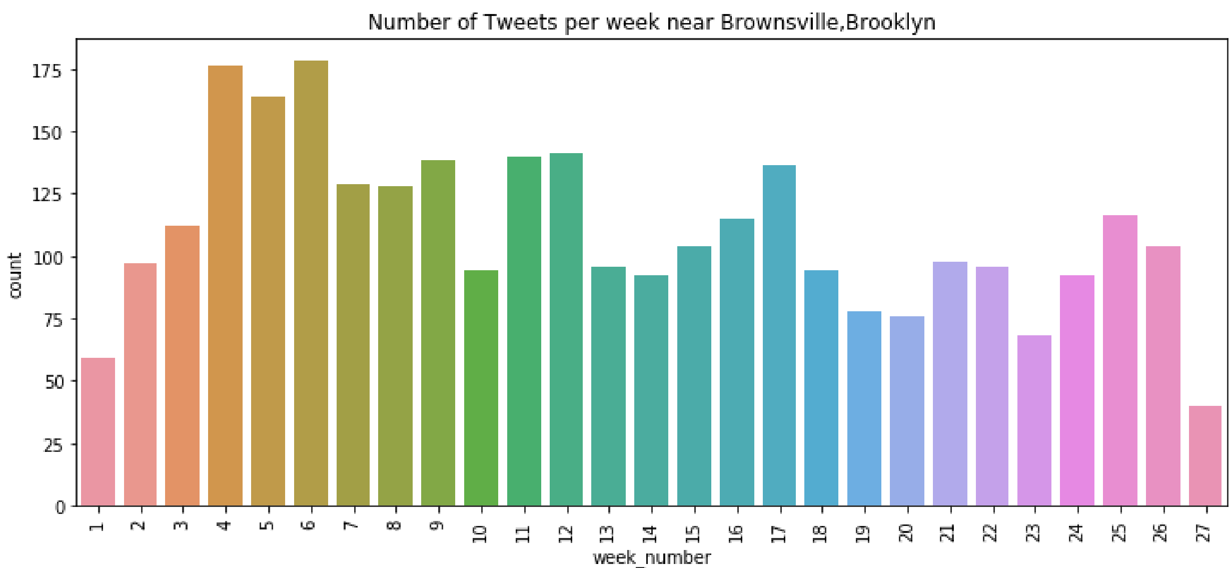

Figure 33: Number of Tweets Per Week Within Brownsville

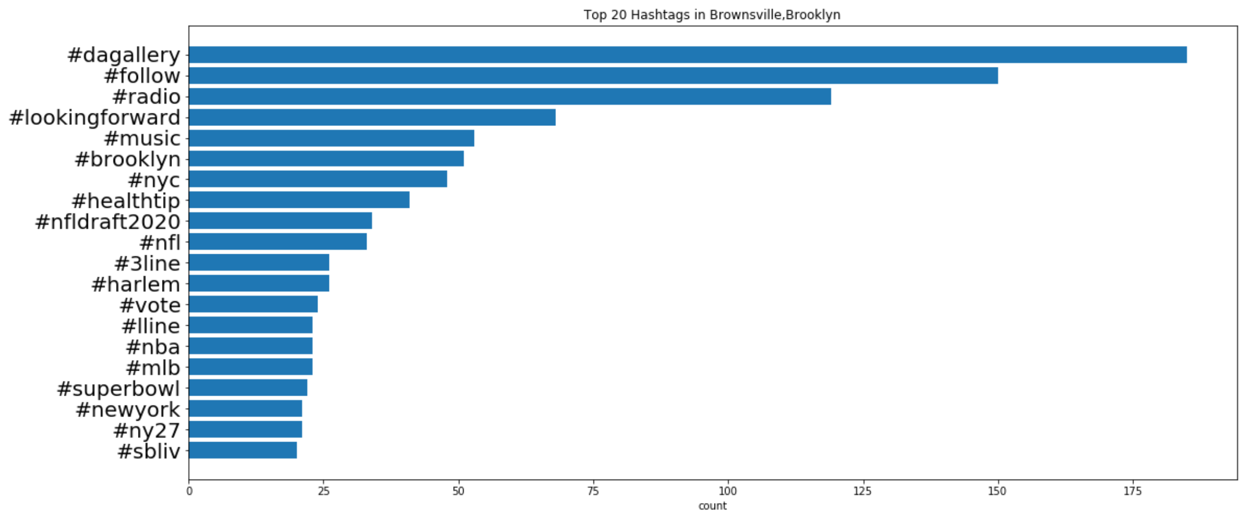

Figure 34: Top 20 Hashtags Within Brownsville

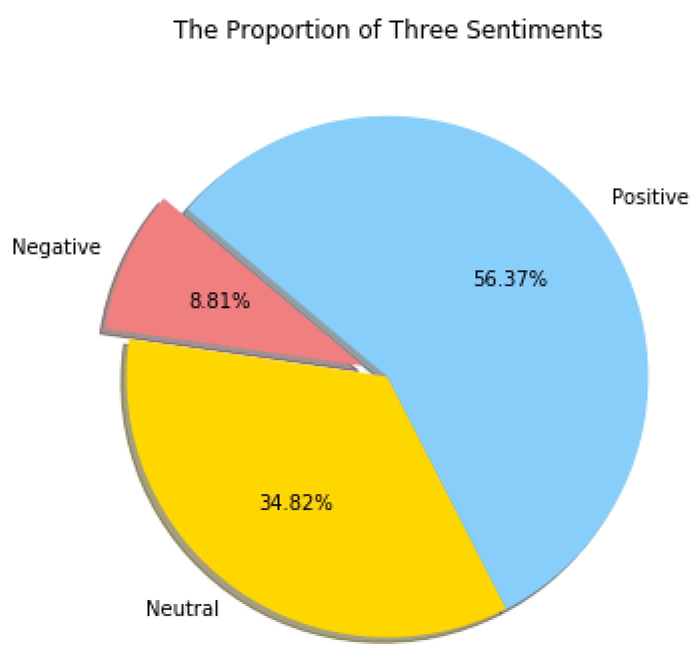

Figure 35: The Proportion of Three Sentiments Within Brownsville

Figure 36: Number of Tweets By Sentiment Per Week Within Brownsville

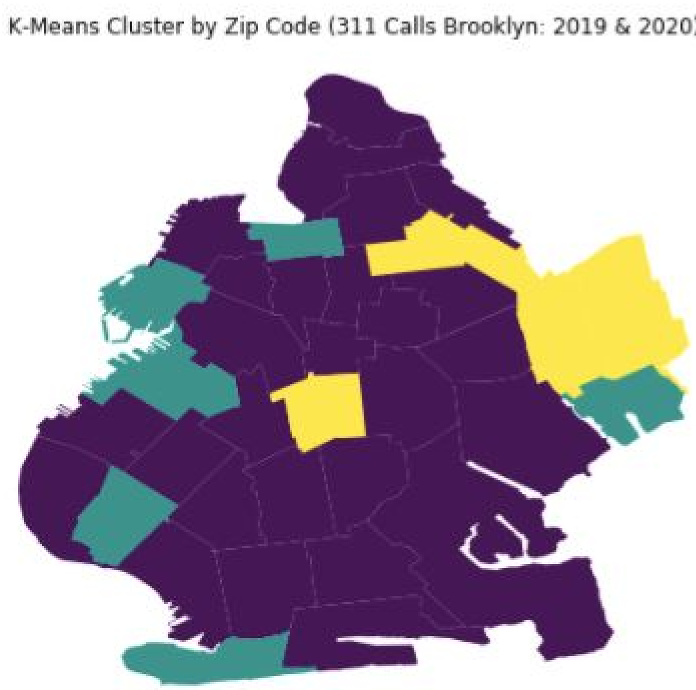

Figure 37: K-means clustering of Brooklyn zip codes by number of 311 calls