New York University

Center for Urban Science + Progress

Mapping Nighttime Urban Mobility in NYC and Charting COVID-19 Related Disruptions to Nightlife

Nicholas LiCalzi, Kaifu Ren, Yingyuan Zhang, Yutong Zhu[1]

Sponsor: Vibelab

CUSP Mentor: Kim Mahler

July 20, 2020

[1]Email contacts: *nl949@nyu.edu*, *kr2516@nyu.edu*, *yz6378@nyu.edu*, *yz6080@nyu.edu*

Abstract

Our research aims to capture the vibrancy and expansiveness of New York City's (NYC) nightlife landscape by interrogating how people access it: with Citibike, subways, or for-hire vehicles (FHV). We establish an understanding of the night-time transportation landscape to enable targeted recommendations around (increasing) transportation usage and mode choice that can ensure the continued survival of nightlife establishments throughout the city, the employment of people who work in them, and their accessibility and importance to their communities and patrons. Our initial research goal was to use multi-modal time series transportation data as an exploratory and advocacy tool to help local stakeholders in the nightlife industry (dance clubs, music venues, bars, and more) demonstrate their contribution to NYC's citywide and hyper-local night-time economies. With the arrival of the COVID-19 pandemic, we expanded our research to examine how the night economy has been uniquely impacted by the crisis in the hopes of enabling policymakers to develop a nuanced understanding of the role of nightlife in the City and develop a uniquely targeted package of policies and aid that ensure its continued vitality. We built a baseline model of winter day and nighttime transportation with data from 2019, and expect to see significant variation given the arrival of the COVID-19 pandemic and related disruptions. We posit that the disruptions have severely affected night-life venues given their in-person nature and expect to see that reflected in the travel data for nightlife hotspots versus in the city as a whole. To empower future research we are sharing the complete dataset, consisting of aggregated venue data, travel data by mode, and more. Data analysts, city planners, venue owners, and others should be able to build on our work and better advocate for the needs of crucial players in NYC's urban nightscape.

1 Introduction

The nightlife industry (bars, restaurants, clubs, music venues, and more) is a major economic and cultural engine for New York City (hereafter NYC). The idea for this research came in collaboration with Vibelab, our project sponsor and an international consultancy that aims to keep cities vibrant and flourishing after dark. They previously developed an initiative called Creative Footprint (CFP) that quantifies the impact of nightlife and cultural activity on cities. CFP has conducted surveys and mapping projects on the nightlife community in New York, Berlin, and Tokyo, producing impact studies for nightlife activists to use in advocating for the value of their scene to their city. In their reporting, they emphasize the exclusion of small to medium sized nightlife venues from policy-making processes that affect them, and attribute that exclusion to "a lack of detailed data collected at the grassroots level... there is little documentation of the character, modi operandi, and outlook of the City's smaller and more progressive venues." We pick up the thread of their work and seek to answer some key questions to supplement their process: Where does NYC's nightlife occur? Is it centralized or diffuse? Are there specific (or surprising) clusters? How do people access nightlife in NYC and are there any predictors of difference in mode choice?

Our research aims to capture the vibrancy and expansiveness of NYC's nightlife landscape by interrogating how people access it: with Citibike, subways, or for-hire vehicles (FHV). Nightlife is a major economic and cultural driver for NYC, and our initial research goal was to use multi-modal time series transportation data as an advocacy tool to help local stakeholders in the nightlife industry (dance clubs, music venues, bars, and more) demonstrate their contribution to NYC's citywide and hyper-local night-time economies. With the arrival of the COVID-19 pandemic, we've expanded our research goal to examine how the night economy has been uniquely impacted by the crisis in the hopes of enabling policymakers to develop a nuanced understanding of the role of nightlife in the City and develop a uniquely targeted package of policies and aid that ensure its continued vitality. The resulting drop in public transportation usage presents urban scientists with an incredible research opportunity– we've built a baseline model of the transportation landscape as it looked in comparable months from 2019 and will explain exactly how each mode (bike, bus, subway, and taxi/for hire vehicle) were affected at each step of the onset of the pandemic.

2 Literature Review

Our literature review uncovered many resources related to the "Night-time Economy" and the utility of nightlife as an economic development indicator or engine for tourism, but we were only able to scant work specifically looking at how people get to and from their nightlife activities. There has been much work in looking at differences in transportation service and experience based on gender, economic status, and accessibility: the experience of nighttime workers [1], the experience of women on night buses [2], etc. Other researchers have proposed various framings (like racial segregation or environmental exposure) that impact human mobility and spacetime travel patterns in cities [3], and used those factors to advocate for new, more holistic transportation planning and decisionmaking processes to improve equity outcomes [4]. Others have focused primarily on time, and how the opening hours of public services might variably affect individuals based on certain personal or household attributes[5][6]– something that is important to consider given NYC's somewhat unusual provision of 24/7 public transportation services, and the threat to that round the clock operation presented by COVID-19 related budget shortfalls.

We have also supplemented our spatial analysis methodology with procedures from our review of the relevant literature. There is a large body of work concerning the specific issue of ZIP codes and other large, spatially arbitrary geographical units being poorly suited for visualization/analysis and requiring disaggregation as a best practice [7]. Some researchers tout the use of areal interpolation for the refitting of data into dasymetric maps [8], while others agree in principle but disagree over what exactly dasymetric means [9]. Our research specifically benefits from the disaggregation of various geodatasets to small, uniform hexagonal bins [10].

3 Data

CFP worked with the nightlife community in NYC to produce a survey mapping nightlife venues across the city, and captured 494 individual venues with a slew of survey data about attendance and more, as well as geographic coordinates for each. They expressed to us a concern about sampling bias in their resulting dataset as it only captured establishments that had participated in the survey and that were largely clustered in downtown Manhattan and the parts of Brooklyn close to it. In order to supplement their findings and extend the analytical capacity of the dataset, we relied on a scraped dataset from Google Maps that one of our team members had previously captured (March 2019). The Google data was filtered to consist of bars, music venues, and nightclubs, and in doing so we were able to add another 3,601 venues that host nightlife to our dataset as well as capture significantly more venues in the outer boroughs, extending the utility of the CFP dataset citywide. In order to preserve the anonymity/safety of the venues in the CFP study (not all of them are operating in a fully legal/licensed manner), as well as to enable further spatial analysis and the joining of other datasets, we aggregated the venues into a hexbin scheme.

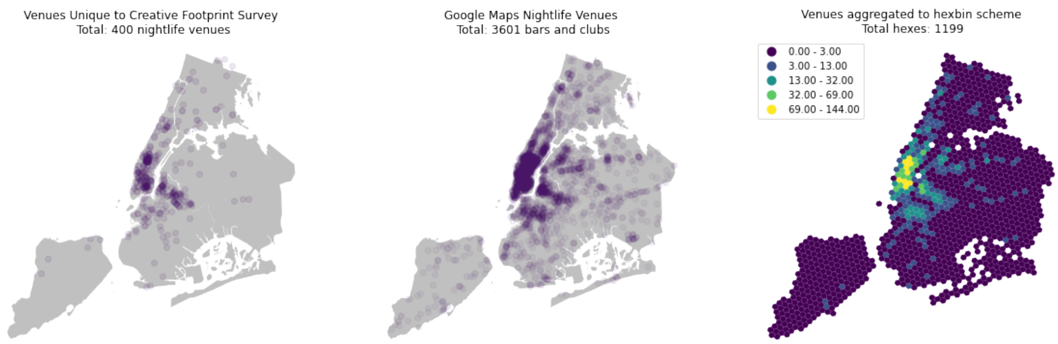

Figure 3.1: Venues by Data Source and Venue Aggregation

We relied on open-source data for our transportation sets: using the MTA's subway turnstile entries/exits dataset, NYC Taxi and Limousine Corporation's (TLC) For-Hire Vehicle (FHV) pickups/drop-offs by taxi zone dataset, and Citibike's trip origin/destination point-file with the temporal information as the primary datasets. We took advantage of Oak Ridge National Labs' recently open-sourced Land-Scan dataset, which represents day and night population counts across the globe[11]. In the interest of tuning several regression models, we also collected data relating to the built environment, daily weather information, and more (Appendix 7.1). Our final dataset is available via Github and available for public use and research (https://github.com/nlicalzi/cusp_capstone).

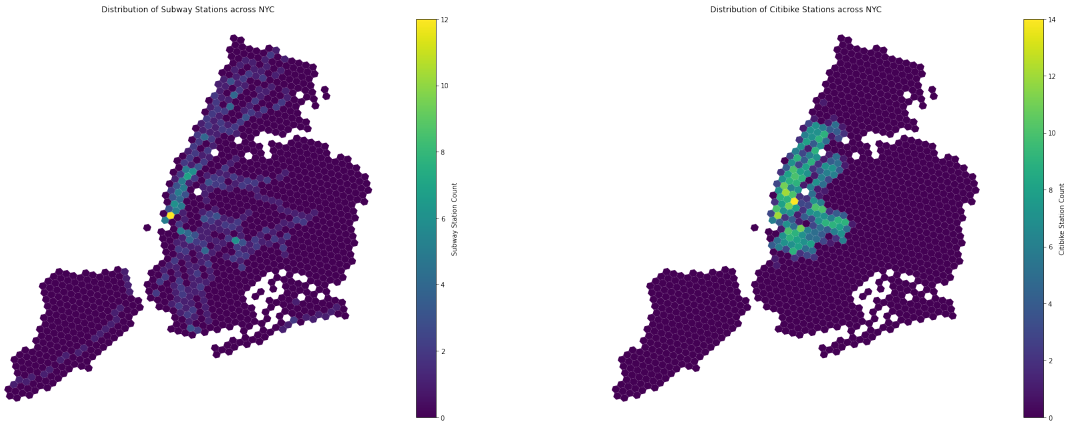

Figure 3.2: Visualization of MTA Subway and Citibike Station Distribution Across NYC

4 Methodology

4.1 Spatial Analysis

Given the locations of nightlife venues across the city, one of the most critical components of our research was identifying which parts of the city were nightlife hotspots, and what might serve as the threshold for declaring an area a hotspot. It is often difficult to analyze/interpret data about NYC because the magnitude of population and built area in Manhattan can often obscure interesting trends in other boroughs, a bias-related concern similar to CFP's about their survey results. We can see this reflected in the first map in Figure 4.1, which depicts the citywide hexes with a z-score (number of standard deviations (σ) from the mean(μ)) of 2 or greater. Given that $\mu = 3$ nightlife venues and that $\sigma = 11$ venues, that means each of the hexes depicted contains at least 25 nightlife venues. That's quite a few venues, considering that the hexbins represent just a few square blocks (they each have an area of about 17 acres, or around 4 square blocks in NYC), and it's unsurprising that that kind of density doesn't exist outside of Manhattan or the parts of Brooklyn closest to it along the L train.

Figure 4.1: Methods of Identifying Nightlife Clusters in NYC

In an attempt to de-bias our nightlife clusters from merely tracking co-linearly with density (and therefore prioritizing Manhattan), we standardize the data by the share of businesses in a given hexbin that are venues rather than the absolute count. In order to avoid misleading results (for example, consider a hexbin with only 2 businesses in 4 square blocks, one of which is a nightclub– would one consider this a nightlife hotspot?) we filter the hexes to only include those with a minimum of 150 total businesses and having a venue share z-score of at least .5 (or 2.5% of the businesses categorized as nightlife). The application of those filters produces the second map in Figure 4.1, which we can see captures a much more geographically diverse spread of nightlife (and closer to the ground-truth in neighborhoods across the City) than the previous methodology.

A key component of our work involved figuring out how to compare data aggregated at multiple geographic levels. The taxi pickups/dropoffs, for example, were aggregated to Taxi Zones, which itself is a derivative of the City's Neighborhood Tabulation Area (NTA) shapefile, while other data like the LandScan population set were aggregated at the census block level. This leaves analysis performed on that data susceptible to the Modifiable Areal Unit Problem (MAUP)[12] a classic form of statistical bias in spatial analytics, where individual spatial phenomena (in this case, taxi origin/destination data) are aggregated into geometries of arbitrary shape and size and the resulting aggregates are subsequently misleading or obfuscating. We attempt to minimize this bias by decomposing the data through areal interpolation; this process is illustrated in Figure 4.2– the Taxi Zones (on the left) vary wildly in shape and size, and tessellate very poorly, leaving them poorly suited for the conducting of various spatial analysis techniques. To accomplish the interpolation we used a pre-release version of the Tobler Python library from PySAL currently in development by a team at UC Riverside led by Sergio Rey, a leading computational geographer[13].

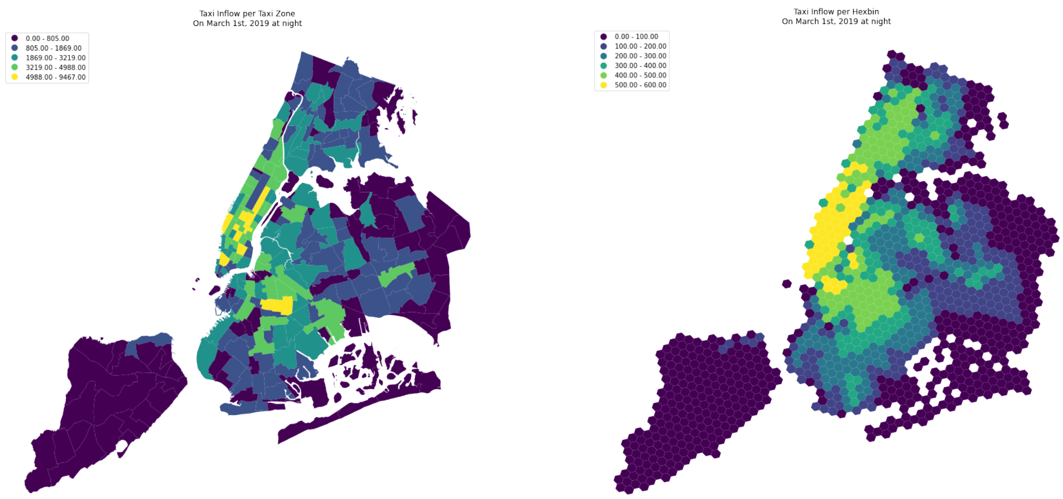

Figure 4.2: Disaggregating FHV Data from Disparate Taxi Zones to Uniform Hexbins

The sheer size of the trip records in some of the travel datasets (Citibike produces unaggregated data for every trip, for example) presents an analysis challenge. To perform adequate analysis over the huge amount of data, we relied on big data techniques. Streaming the data allowed us to process trip records row by row without monopolizing RAM on our personal computers, while aggregating by day and time of day (day/night boolean). We processed daily day and night data intervals ranging from Jan. 1 -Apr. 30 for the years 2019 and 2020.

4.2 Data Visualization

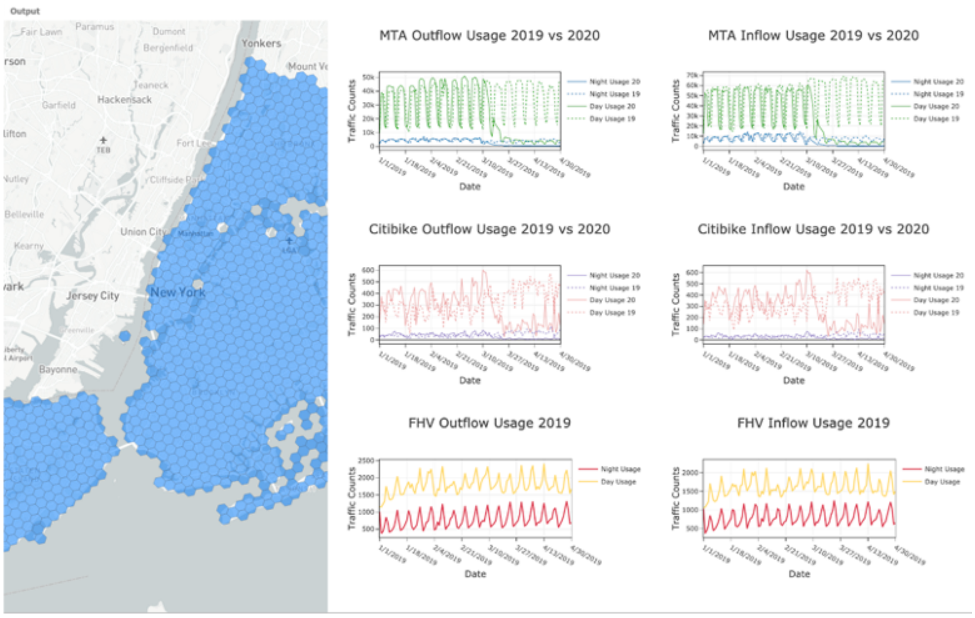

Figure 4.3: Project data visualization interface (*Link* : https://nlicalzi.github.io/cusp_capstone/)

We use D3.js[2] and the MapboxGL Javascript API[3] to construct a data visualization interface (Figure 4.3). Importing the prepared data, users can monitor daily or nightly changes of inflow and outflow for different mobility usage in hexagons or music venues' hotspots by clicking on the map. Time series analysis can help to get insights for anomaly identification and trend summary for selected cases. On the other hand, the interface allows users to compare the observation in 2020 with the baseline in 2019 to show differences in mobility demands potentially caused by the outbreaks of COVID-19. The changes in demands can be correlated to the milestones of government actions.

4.3 Regression Analysis

To further analyze the hidden factors over the changes in daily usage of transportation tools during night-time, we initiated a multivariate regression. In the regression model, we separated three main transportation mode during night-time and research for the correlation with demographic features, built environment features (retail area, commercial area, residential area), weather features (temperature, precipitation), temporal factors (day of the week, month, COVID-19), and traffic-related factors (station counts in the area).

[2] Web link: https://d3js.org/
[3] Web link: https://docs.mapbox.com/api/

To boost the growth of the nightlife industry as well as increase the new-registered night-time venues, we conduct a multivariate regression based on the venue counts. Taking business counts, transportation inflow data (Subway, Citibike and FHV), location data (boroughs) as our independent variables.

From statistics that the regression models have, we keep features with strong confidence and apply feature selections to make sure the R-squared score is good to be trusted. During the process for feature selection, we filter out the factory area, number of buildings for the final regression. Moreover, we separated the population feature and environmental built feature into sub-models to test regression separately.

Regression Equation:(Detailed explanation can be found in Appendix)

$$DailyInflow_i = \beta_0 + \beta_1 DayoftheWeek_i + \beta_2 AveTemp_i + \beta_3 Precipitation_i + \beta_4 NightDummy + \varepsilon_i$$

$$\begin{aligned} VenueCounts_i =& \beta_0 + \beta_1 BuseinssCount_i + \beta_2 TMTAInflows_i + \beta_3 CitiBikeInflow_i + \\ & \beta_4 FHVInflows_i + \beta_5 Stationcounts_i + \beta_6 boroDummy + \varepsilon_i \end{aligned}$$

$$\begin{aligned} NightDailyInflow_i =& \beta_0 + \beta_1 StationCount_i + \beta_2 BusinessCount_i + \beta_3 LotsArea_i + \\ & \beta_4 ResidentialArea_i + \beta_5 OfficeArea_i + \beta_6 RetailArea_i + \\ & \beta_7 NightPopulation_i + \beta_8 BoroughDummy + \varepsilon_i \end{aligned}$$

5 Results and Discussion

Are there variations in preferred night-time transportation mode in different boroughs of the city? From the night-time transportation inflow regression, we found that during night-time, New Yorkers prefer using taxi as their transportation mode.(Figure 7.2, Figure 7.5, Figure 7.8) The result reveals that if the destination is Manhattan, FHV will have more night-time inflows as well as CitiBike. But not the subway. This might be caused by the fact of less traffic congestion at night and safety issues. The borough that people love taking MTA at night is Staten Island with a 166.7 increase. This could be explained by the fact that the taxi's price to Staten Island is relatively higher than MTA's $2.75 ride.

Looking at the entire NYC, we could easily find that Manhattan has the greatest impact on the transportation inflows as it holds the relatively high absolute value of coefficient in all three transportation's regression results.

Moreover, if we would like to attract more people to visit certain areas, we should increase the amount of stations in the area. This could apply on MTA as well as Citibike. However, people may use mixed of transportation modes for their round trips. Thus, the more stations (both MTA and Citibike) in the area will increase the night-time taxi trips in the area.

Once COVID-19 hits on NYC, fewer people are visiting Brooklyn and Manhattan at night. As there are more residential areas in Bronx, Queens and Staten Island, there is an increase in people visiting these three areas at night time during COVID-19.(Figure 7.11, Figure 7.14) Overall, the population and built environment factors have minor impact on the transportation inflows.(Figure 7.7, Figure 7.16)

How does the nightlife venue grows? Although people will not consider the place's location as their primary factor when choosing to visit, Manhattan still has a strong influence on venue count. As we know, Manhattan is the heart of NYC and lots of famous shows are presented in Manhattan. Compared with other transportation tools, Citibike has a higher positive coefficient on venue counts.(Figure 7.20) NYC may have traffic congestion at night time, thus some people may choose Citibike as their transportation tool to avoid this or use it as their last mile solution. Moreover, different land use has less impact on venue counts. Overall, venue locations have strong relationships with boroughs, transportation station locations and little relationship with populations.(Figure 7.21, Figure 7.22)

What is the citywide and borough-wide decrease in transportation utilization due to COVID-19? To find out this, we define three time periods with the same number of days and run a comparison study for both years. On March 1st, the first confirmed case was reported in NYC. On March 17th the governor De Balsio signed an executive order shuttering all restaurants, bars, and clubs to all activity besides takeout and delivery ([14]). The first period, between Feb 12 and Feb 28, marks the normal condition. The second period, between Mar 1 and Mar 17, is when people started to be aware of the outbreaks while business and venues were operating. The third period, between Mar 18 and Apr 3, is when all nightlife had been closed and the city started to lock down.

MTA usage presents a slight loss in citywide and borough-wide statistics in period 1, however it tends to remain in the normal fluctuation (Table 5.1, Table 5.2). In period 2, it shows an increased reduction around 10-15% in general. Losses at night are greater than day-time in every sector. Also, Manhattan holds the biggest loss for both day and night transit, followed by the nightlife clusters which indicates that people rely on taking subways to access those venue places. Citibike performance shows a reversed trend. There is a boost in usage around 30% in both period 1 and 2 (Table 5.3, Table 5.4). The outbreak of COVID-19 doesn't seem to interfere with bike riding in the first place. Part of the reasons for this phenomenon is NYC's great investment in improving the biking experience and providing more protected bike lanes. Moreover, people tend to avoid mass transit and prefer biking. After venues were closed and all business have been shut down, MTA and Citibike show the significant loss in around 80% and 60% respectively. MTA presents a gradual reduction among three periods, while Citibike usage drops sharply from period 2 to 3.

	Citywide		Nightlife Cluster	
	Day	Night	Day	Night
Period 1 ('19)	51,975,855	10,908,923	9,020,153	2,351,575
Period 1 ('20) %	-3.28	-2.77	0.47	2.58
Period 2 ('19)	50,564,654	12,107,212	8,825,378	2,661,869
Period 2 ('20) %	-14.97	-22.23	-14.85	-23.84
Period 3 ('19)	54,441,850	13,615,099	9,487,134	2,870,466
Period 3 ('20) %	-83.42	-85.37	-86.89	-89.83

Figure 5.1: Subway usage Δ(%) 2019-2020

	Bronx		Brooklyn		Manhattan		Queens		Staten Island	
	Day	Night	Day	Night	Day	Night	Day	Night	Day	Night
Period 1 ('19)	5,927,941	639,374	13,690,109	1,755,900	25,666,065	7,704,481	6,684,529	806,287	7,211	2,881
Period 1 ('20) %	-5.19	-5.92	-1.19	3.7	-4.78	-4.96	-0.07	6.5	-19.27	-9.44
Period 2 ('19)	5,699,645	781,246	13,345,254	1,995,352	24,975,884	8,428,520	6,536,322	899,358	7,549	2,736
Period 2 ('20) %	-8.16	-10.76	-10.74	-12.49	-20.58	-26.61	-8.11	-12.79	-14.04	-14.94
Period 3 ('19)	6,225,826	998,281	14,711,417	2,267,534	26,398,321	9,306,346	7,096,639	1,040,432	9,647	2,506
Period 3 ('20) %	-73.01	-69.53	-81.6	-77.33	-87.68	-89.57	-80.47	-80.59	-73.25	-64.96

Figure 5.2: Subway usage Δ(%) 2019-2020

	Citywide		Nightlife Cluster	
	Day	Night	Day	Night
Period 1 ('19)	393756	54401	131583	22280
Period 1 ('20) %	29.92	34.27	28.79	30.96
Period 2 ('19)	456412	64891	152838	26326
Period 2 ('20) %	33.39	22.39	31.28	15.04
Period 3 ('19)	604294	89059	199420	36107
Period 3 ('20) %	-62.64	-69.07	-62.86	-73.52

Figure 5.3: Citibike usage Δ(%) 2019-2020

	Bronx		Brooklyn		Manhattan		Queens		Staten Island	
	Day	Night	Day	Night	Day	Night	Day	Night	Day	Night
Period 1 ('19)	243	77	73933	12703	311219	40079	8361	1542	0	0
Period 1 ('20) %	50.61	58.44	25.14	31.53	31.09	34.73	28.09	43.77	N/A	N/A
Period 2 ('19)	291	87	86412	15002	360192	47802	9517	2000	0	0
Period 2 ('20) %	54.63	34.48	34.3	29.32	32.67	19.61	51.47	36.4	N/A	N/A
Period 3 ('19)	435	140	115009	21754	474723	64114	14127	3051	0	0
Period 3 ('20) %	-17.24	-27.85	-57.11	-72.07	-64.68	-68.51	-40.75	-61.45	N/A	N/A

Figure 5.4: Citibike usage Δ(%) 2019-2020

What is the relationship between transport mode choice and weather, weekday and weekends and day or night? From regression analysis, we could conclude that New Yorkers are preferred to use FHV during weekends and the subway during the weekdays(Figure 7.19, Figure 7.17). Citibike does not have a dramatic difference on weekends and weekdays.(Figure 7.18) Once it is nighttime, all transportation usage decrease, especially the subway. Rainy days will have a bigger impact on subway ridership but minor impact on Citibike(Figure 7.19, Figure 7.18, Figure 7.17).For FHV, we do not have significant evidence to indicate the relationship with weather(Figure 7.19).

6 Conclusion

Our expectation is that this work will serve as an inspiration/facilitation of future scholarship on this issue. We have outlined a process for identifying nightlife hotspot loci in an urban environment, as well the process by which we can combine and analyze disparate datasets to attempt to explain ridership phenomena. Potential extensions of our work might include using measures of local/global spatial autocorrelation to identify clusters of nightlife activity, the use of ridge/LASSO regressions to attempt to neutralize any colinearity in our datasets, and placing a greater emphasis on walkability/last mile transit options (incl. Revel scooters, and others). The COVID-19 pandemic has had devastating effects on the City as a whole, but particularly on the nightlife sector, and we hope that the research and processes contained in this project may be of use to interested parties.

References

[1] Jenny McArthur, Enora Robin, and Emilia Smeds. "Socio-spatial and temporal dimensions of transport equity for London's night time economy". In: *Transportation Research Part A: Policy and Practice* 121 (2019), pp. 433–443. DOI: 10.1016/j.tra.2019.01.024.

[2] Anna Plyushteva and Kobe Boussauw. "Does night-time public transport contribute to inclusive night mobility? Exploring Sofia's night bus network from a gender perspective". In: *Transport Policy* 87 (2020), pp. 41–50. DOI: 10.1016/j.tranpol.2020.01.002.

[3] Mei-Po Kwan. "Beyond Space (As We Knew It): Toward Temporally Integrated Geographies of Segregation, Health, and Accessibility". In: *Space-Time Integration in Geography and GIScience* (2014), pp. 39–51. DOI: 10.1007/978-94-017-9205-9_4.

[4] Peter Jones and Karen Lucas. "The social consequences of transport decision-making: clarifying concepts, synthesising knowledge and assessing implications". In: *Journal of Transport Geography* 21 (2012), pp. 4–16. DOI: 10.1016/j.jtrangeo.2012.01.012.

[5] Tijs Neutens et al. "Evaluating the Temporal Organization of Public Service Provision Using Space-Time Accessibility Analysis". In: *Urban Geography* 31.8 (2010), pp. 1039–1064. DOI: 10.2747/0272-3638.31.8.1039.

[6] Matthias Delafontaine et al. "The impact of opening hours on the equity of individual space–time accessibility". In: *Computers, Environment and Urban Systems* 35.4 (2011), pp. 276–288. DOI: 10.1016/j.compenvurbsys.2011.02.005.

[7] Tony H Grubesic and Timothy C Matisziw. ""On the Use of ZIP Codes and ZIP Code Tabulation Areas (ZCTAs) for the Spatial Analysis of Epidemiological Data."". In: *International Journal of Health Geographics* 5.1 (Dec. 2006), p. 58. DOI: 10.1186/1476-072x-5-58.

[8] Alexis Comber and Wen Zeng. "Spatial interpolation using areal features: A review of methods and opportunities using new forms of data with coded illustrations". In: *Geography Compass* 13.10 (2019). DOI: 10.1111/gec3.12465.

[9] Cory L. Eicher and Cynthia A. Brewer. "Dasymetric Mapping and Areal Interpolation: Implementation and Evaluation". In: *Cartography and Geographic Information Science* 28.2 (2001), pp. 125–138. DOI: 10.1559/152304001782173727.

[10] Samuel Langton and Reka Solymosi. "Cartograms, hexograms and regular grids: a commentary on minimising misrepresentation in spatial data visualisations". In: *SocArXiv* (2019). DOI: 10.31235/osf.io/t6agd.

[11] Budhendra Bhaduri et al. "LandScan USA: a high-resolution geospatial and temporal modeling approach for population distribution and dynamics". In: *GeoJournal* 69.1-2 (2007), pp. 103–117. DOI: 10.1007/s10708-007-9105-9.

[12] Stan Openshaw. "The Modifiable Areal Unit Problem". In: *Concepts and Techniques in Modern Geography* 38 (1984).

[13] Renan Xavier Cortes et al. *pysal/tobler: Version 0.3.0*. Apr. 2020. DOI: 10.5281/ZENODO.3757526. URL: https://github.com/pysal/tobler.

[14] Bill de Blasio. *Statement on Bars, Restaurants, and Entertainment Venues*. 2020 (accessed July 3, 2020). URL: https://www1.nyc.gov/office-of-the-mayor/news/152-20/statement-mayor-de-blasio-bars-restaurants-entertainment-venues.

7 Appendix

7.1 Data Sources

System	Date Modified	Link	Frequency	Action
Citibike-tripdata	Jul 2020	https://s3.amazonaws.com/tripdata/index.html	Monthly Updated	Available
MTA Real-time Turnstile Data	Jul 2020	http://web.mta.info/developers/turnstile.html	Weekly Updated	Available
TLC Monthly Data (Yellow, Green Taxi, FHV trips)	Dec 2019	https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page	Annually Updated	Unavailable for 2020
Day and night demographics in NYC	May 2020	LandScan Census Data https://www.ornl.gov/news/gis-landscan-goes-public	N/A	Available
Geographical Features (Shapefile of subway lines, taxi zones, zip codes, etc.)	N/A	NYC open data portal	N/A	Available
Built Environment Features (Commercial Area, Residential Area, Number of Buildings, etc.)	N/A	PLUTO, NYC Department of City Planning	N/A	Available
Music Venues & Information (Creative Footprint NYC - VibeLab)	2019	https://drive.google.com/drive/u/1/folders/1fKMt3mrJVU0vn5YVUeq10W88FufhwYfH	One time survey	Available
NYC Businesses (scraped from Google Maps)	March 2019	Able to be shared only in an obscured form (tbd)	Collected as a citywide snapshot	Available
COVID-19 dataset	To present	https://github.com/thecityny/covid-19-nyc-data	Daily Updated	Available
Weather information (Temperature and Rainfall)	To present	https://www.wunderground.com/history/daily/KLGA/date/2020-6-19	Daily collected	Available

Figure 7.1: Data Source

7.2 Regression

```
The following result is 2019 mta night-time inflow regression
                            OLS Regression Results
==============================================================================
Dep. Variable:      mta_inflow_counts   R-squared:                       0.781
Model:                            OLS   Adj. R-squared:                  0.781
Method:                 Least Squares   F-statistic:                 6.411e+04
Date:                Sun, 19 Jul 2020   Prob (F-statistic):               0.00
Time:                        01:31:50   Log-Likelihood:            -1.2522e+06
No. Observations:              143880   AIC:                         2.504e+06
Df Residuals:                  143871   BIC:                         2.504e+06
Df Model:                           8
Covariance Type:            nonrobust
========================================================================================
                           coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------
Intercept             -430.9263      4.813    -89.534      0.000    -440.360    -421.493
venue_count             78.7181      0.587    134.055      0.000      77.567      79.869
subway_station_count   593.8213      5.686    104.430      0.000     582.676     604.966
bike_station_count    -131.0410      2.692    -48.670      0.000    -136.318    -125.764
business_count           1.6937      0.006    280.379      0.000       1.682       1.706
Bronx                  -99.1198      9.232    -10.736      0.000    -117.215     -81.025
Brooklyn              -238.5262      7.495    -31.827      0.000    -253.215    -223.837
Manhattan             -324.8862     16.448    -19.753      0.000    -357.123    -292.649
Queens                  30.0437      6.876      4.370      0.000      16.567      43.520
Staten_Island          201.5622      8.609     23.414      0.000     184.690     218.435
==============================================================================
Omnibus:                    95714.088   Durbin-Watson:                   2.081
Prob(Omnibus):                  0.000   Jarque-Bera (JB):         34342957.498
Skew:                           2.025   Prob(JB):                         0.00
Kurtosis:                      78.579   Cond. No.                     2.29e+18
==============================================================================
```

Figure 7.2: MTA night-time inflow in 2019

```
The following result is 2019 mta night-time inflow regression
                            OLS Regression Results
==============================================================================
Dep. Variable:      mta_inflow_counts   R-squared:                       0.781
Model:                            OLS   Adj. R-squared:                  0.781
Method:                 Least Squares   F-statistic:                 6.411e+04
Date:                Sun, 19 Jul 2020   Prob (F-statistic):               0.00
Time:                        01:31:50   Log-Likelihood:            -1.2522e+06
No. Observations:              143880   AIC:                         2.504e+06
Df Residuals:                  143871   BIC:                         2.504e+06
Df Model:                           8
Covariance Type:            nonrobust
========================================================================================
                           coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------
Intercept             -430.9263      4.813    -89.534      0.000    -440.360    -421.493
venue_count             78.7181      0.587    134.055      0.000      77.567      79.869
subway_station_count   593.8213      5.686    104.430      0.000     582.676     604.966
bike_station_count    -131.0410      2.692    -48.670      0.000    -136.318    -125.764
business_count           1.6937      0.006    280.379      0.000       1.682       1.706
Bronx                  -99.1198      9.232    -10.736      0.000    -117.215     -81.025
Brooklyn              -238.5262      7.495    -31.827      0.000    -253.215    -223.837
Manhattan             -324.8862     16.448    -19.753      0.000    -357.123    -292.649
Queens                  30.0437      6.876      4.370      0.000      16.567      43.520
Staten_Island          201.5622      8.609     23.414      0.000     184.690     218.435
==============================================================================
Omnibus:                    95714.088   Durbin-Watson:                   2.081
Prob(Omnibus):                  0.000   Jarque-Bera (JB):         34342957.498
Skew:                           2.025   Prob(JB):                         0.00
Kurtosis:                      78.579   Cond. No.                     2.29e+18
==============================================================================
```

Figure 7.3: MTA night-time inflow in 2019 with Area Variable

The following result is 2019 mta night-time inflow regression with population variables

OLS Regression Results

Dep. Variable:	mta_inflow_counts	R-squared:	0.802
Model:	OLS	Adj. R-squared:	0.802
Method:	Least Squares	F-statistic:	5.829e+04
Date:	Sun, 19 Jul 2020	Prob (F-statistic):	0.00
Time:	01:31:51	Log-Likelihood:	-1.2449e+06
No. Observations:	143880	AIC:	2.490e+06
Df Residuals:	143869	BIC:	2.490e+06
Df Model:	10		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-140.9966	5.786	-24.371	0.000	-152.336	-129.657
venue_count	78.5151	0.558	140.607	0.000	77.421	79.610
subway_station_count	668.4430	5.713	117.002	0.000	657.245	679.641
bike_station_count	-113.5545	2.595	-43.753	0.000	-118.641	-108.468
business_count	0.9413	0.011	84.204	0.000	0.919	0.963
Bronx	-42.5529	8.896	-4.784	0.000	-59.988	-25.117
Brooklyn	-152.0322	7.184	-21.163	0.000	-166.113	-137.952
Manhattan	52.4051	16.170	3.241	0.001	20.712	84.098
Queens	-10.6606	6.552	-1.627	0.104	-23.502	2.180
Staten_Island	11.8440	8.518	1.390	0.164	-4.851	28.539
day_pop	0.0432	0.001	76.968	0.000	0.042	0.044
night_pop	-0.0530	0.001	-102.548	0.000	-0.054	-0.052

Omnibus:	105981.049	Durbin-Watson:	2.130
Prob(Omnibus):	0.000	Jarque-Bera (JB):	44968910.355
Skew:	2.396	Prob(JB):	0.00
Kurtosis:	89.476	Cond. No.	2.81e+19

Figure 7.4: MTA night-time inflow in 2019 with Population Variable

The following result is 2019 Citi Bike night-time inflow regression

OLS Regression Results

Dep. Variable:	citi_inflow_counts	R-squared:	0.647
Model:	OLS	Adj. R-squared:	0.647
Method:	Least Squares	F-statistic:	3.293e+04
Date:	Sun, 19 Jul 2020	Prob (F-statistic):	0.00
Time:	01:32:25	Log-Likelihood:	-5.3570e+05
No. Observations:	143880	AIC:	1.071e+06
Df Residuals:	143871	BIC:	1.072e+06
Df Model:	8		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	1.2478	0.033	37.701	0.000	1.183	1.313
venue_count	0.9079	0.004	224.838	0.000	0.900	0.916
subway_station_count	-1.4430	0.039	-36.900	0.000	-1.520	-1.366
bike_station_count	2.3661	0.019	127.789	0.000	2.330	2.402
business_count	-0.0021	4.15e-05	-51.107	0.000	-0.002	-0.002
Bronx	-1.6889	0.063	-26.601	0.000	-1.813	-1.564
Brooklyn	-2.2773	0.052	-44.186	0.000	-2.378	-2.176
Manhattan	8.4104	0.113	74.357	0.000	8.189	8.632
Queens	-1.8950	0.047	-40.077	0.000	-1.988	-1.802
Staten_Island	-1.3014	0.059	-21.983	0.000	-1.417	-1.185

Omnibus:	192591.941	Durbin-Watson:	1.766
Prob(Omnibus):	0.000	Jarque-Bera (JB):	139898520.364
Skew:	7.202	Prob(JB):	0.00
Kurtosis:	155.080	Cond. No.	2.29e+18

Figure 7.5: Citibike Night-time inflow in 2019

```
The following result is 2019 Citi Bike night-time inflow regression with area variables
                            OLS Regression Results
==============================================================================
Dep. Variable:     citi_inflow_counts   R-squared:                       0.661
Model:                            OLS   Adj. R-squared:                  0.661
Method:                 Least Squares   F-statistic:                 2.335e+04
Date:                Sun, 19 Jul 2020   Prob (F-statistic):               0.00
Time:                        01:32:25   Log-Likelihood:            -5.3278e+05
No. Observations:              143880   AIC:                         1.066e+06
Df Residuals:                  143867   BIC:                         1.066e+06
Df Model:                          12
Covariance Type:            nonrobust
=========================================================================================
                            coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------------
Intercept                 0.5282      0.044     12.102      0.000       0.443       0.614
venue_count               0.9205      0.004    223.408      0.000       0.912       0.929
subway_station_count     -1.2189      0.040    -30.760      0.000      -1.297      -1.141
bike_station_count        2.2387      0.019    118.709      0.000       2.202       2.276
business_count            0.0020   7.32e-05     27.628      0.000       0.002       0.002
Bronx                    -1.6508      0.063    -26.343      0.000      -1.774      -1.528
Brooklyn                 -2.4365      0.051    -47.987      0.000      -2.536      -2.337
Manhattan                 7.2756      0.119     61.043      0.000       7.042       7.509
Queens                   -1.8410      0.047    -38.980      0.000      -1.934      -1.748
Staten_Island            -0.8192      0.061    -13.528      0.000      -0.938      -0.700
LotArea                4.071e-09   5.43e-10      7.499      0.000    3.01e-09    5.13e-09
ResArea                8.447e-08   1.02e-08      8.310      0.000    6.45e-08    1.04e-07
OfficeArea            -9.994e-07   1.62e-08    -61.556      0.000   -1.03e-06   -9.68e-07
RetailArea            -2.323e-06   1.01e-07    -22.995      0.000   -2.52e-06   -2.12e-06
==============================================================================
Omnibus:                   191532.275   Durbin-Watson:                   1.725
Prob(Omnibus):                  0.000   Jarque-Bera (JB):         134847642.782
Skew:                           7.134   Prob(JB):                         0.00
Kurtosis:                     152.298   Cond. No.                     6.32e+22
==============================================================================
```

Figure 7.6: Citibike night-time inflow in 2019 with Area Variable

```
The following result is 2019 Citi Bike night-time inflow regression with population variables
                            OLS Regression Results
==============================================================================
Dep. Variable:     citi_inflow_counts   R-squared:                       0.650
Model:                            OLS   Adj. R-squared:                  0.650
Method:                 Least Squares   F-statistic:                 2.671e+04
Date:                Sun, 19 Jul 2020   Prob (F-statistic):               0.00
Time:                        01:32:25   Log-Likelihood:            -5.3505e+05
No. Observations:              143880   AIC:                         1.070e+06
Df Residuals:                  143869   BIC:                         1.070e+06
Df Model:                          10
Covariance Type:            nonrobust
========================================================================================
                           coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------
Intercept                0.8723      0.042     20.936      0.000       0.791       0.954
venue_count              0.9072      0.004    225.593      0.000       0.899       0.915
subway_station_count    -1.4579      0.041    -35.434      0.000      -1.539      -1.377
bike_station_count       2.3623      0.019    126.385      0.000       2.326       2.399
business_count       -4.973e-05   8.05e-05     -0.618      0.537      -0.000       0.000
Bronx                   -1.6961      0.064    -26.474      0.000      -1.822      -1.571
Brooklyn                -2.4039      0.052    -46.463      0.000      -2.505      -2.303
Manhattan                7.8698      0.116     67.577      0.000       7.642       8.098
Queens                  -1.8371      0.047    -38.935      0.000      -1.930      -1.745
Staten_Island           -1.0604      0.061    -17.285      0.000      -1.181      -0.940
day_pop                 -0.0001   4.04e-06    -29.708      0.000      -0.000      -0.000
night_pop             8.369e-05   3.72e-06     22.482      0.000    7.64e-05     9.1e-05
==============================================================================
Omnibus:                   190748.222   Durbin-Watson:                   1.748
Prob(Omnibus):                  0.000   Jarque-Bera (JB):         131931211.186
Skew:                           7.082   Prob(JB):                         0.00
Kurtosis:                     150.669   Cond. No.                     2.81e+19
==============================================================================
```

Figure 7.7: Citibike night-time inflow in 2019 with Population Variable

```
The following result is 2019 FHV Taxi night-time inflow regression
                            OLS Regression Results
==============================================================================
Dep. Variable:      fhv_inflow_counts   R-squared:                       0.781
Model:                            OLS   Adj. R-squared:                  0.781
Method:                 Least Squares   F-statistic:                 6.399e+04
Date:                Sun, 19 Jul 2020   Prob (F-statistic):               0.00
Time:                        01:32:28   Log-Likelihood:            -9.4663e+05
No. Observations:              143880   AIC:                         1.893e+06
Df Residuals:                  143871   BIC:                         1.893e+06
Df Model:                           8
Covariance Type:            nonrobust
========================================================================================
                           coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------
Intercept               96.8106      0.576    168.167      0.000      95.682      97.939
venue_count             14.1384      0.070    201.299      0.000      14.001      14.276
subway_station_count    20.4962      0.680     30.135      0.000      19.163      21.829
bike_station_count      46.1969      0.322    143.450      0.000      45.566      46.828
business_count           0.0273      0.001     37.829      0.000       0.026       0.029
Bronx                   15.6294      1.104     14.154      0.000      13.465      17.794
Brooklyn                 6.3462      0.896      7.080      0.000       4.589       8.103
Manhattan              233.4191      1.967    118.650      0.000     229.563     237.275
Queens                 -59.6966      0.822    -72.588      0.000     -61.308     -58.085
Staten_Island          -98.8875      1.030    -96.039      0.000    -100.906     -96.869
==============================================================================
Omnibus:                   117194.179   Durbin-Watson:                   1.550
Prob(Omnibus):                  0.000   Jarque-Bera (JB):         14469498.151
Skew:                           3.258   Prob(JB):                         0.00
Kurtosis:                      51.694   Cond. No.                     2.29e+18
==============================================================================
```

Figure 7.8: FHV Night-time inflow in 2019

```
The following result is 2019 FHV Taxi night-time inflow regression with area variables
                            OLS Regression Results
==============================================================================
Dep. Variable:      fhv_inflow_counts   R-squared:                       0.804
Model:                            OLS   Adj. R-squared:                  0.804
Method:                 Least Squares   F-statistic:                 4.933e+04
Date:                Sun, 19 Jul 2020   Prob (F-statistic):               0.00
Time:                        01:32:28   Log-Likelihood:            -9.3834e+05
No. Observations:              143880   AIC:                         1.877e+06
Df Residuals:                  143867   BIC:                         1.877e+06
Df Model:                          12
Covariance Type:            nonrobust
========================================================================================
                           coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------
Intercept               47.7141      0.731     65.251      0.000      46.281      49.147
venue_count             14.7286      0.069    213.359      0.000      14.593      14.864
subway_station_count     7.7225      0.664     11.632      0.000       6.421       9.024
bike_station_count      35.2618      0.316    111.596      0.000      34.642      35.881
business_count           0.0399      0.001     32.595      0.000       0.038       0.042
Bronx                    2.8617      1.050      2.725      0.006       0.804       4.920
Brooklyn                 8.5077      0.851     10.001      0.000       6.840      10.175
Manhattan              146.5424      1.997     73.382      0.000     142.628     150.456
Queens                 -44.8766      0.791    -56.710      0.000     -46.428     -43.326
Staten_Island          -65.3210      1.015    -64.383      0.000     -67.310     -63.332
LotArea              -1.435e-07    9.1e-09    -15.774      0.000   -1.61e-07   -1.26e-07
ResArea               1.876e-05    1.7e-07    110.154      0.000    1.84e-05    1.91e-05
OfficeArea           -3.203e-06   2.72e-07    -11.774      0.000   -3.74e-06   -2.67e-06
RetailArea           -2.408e-05   1.69e-06    -14.230      0.000   -2.74e-05   -2.08e-05
==============================================================================
Omnibus:                   129988.972   Durbin-Watson:                   1.504
Prob(Omnibus):                  0.000   Jarque-Bera (JB):         21141020.916
Skew:                           3.798   Prob(JB):                         0.00
Kurtosis:                      61.896   Cond. No.                     6.32e+22
==============================================================================
```

Figure 7.9: FHV night-time inflow in 2019 with Area Variable

```
The following result is 2019 FHV Taxi night-time inflow regression with population variables
                            OLS Regression Results
==============================================================================
Dep. Variable:      fhv_inflow_counts   R-squared:                       0.809
Model:                            OLS   Adj. R-squared:                  0.809
Method:                 Least Squares   F-statistic:                 6.092e+04
Date:                Sun, 19 Jul 2020   Prob (F-statistic):               0.00
Time:                        01:32:29   Log-Likelihood:            -9.3668e+05
No. Observations:              143880   AIC:                         1.873e+06
Df Residuals:                  143869   BIC:                         1.873e+06
Df Model:                          10
Covariance Type:            nonrobust
========================================================================================
                           coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------
Intercept               36.7791      0.679     54.142      0.000      35.448      38.111
venue_count             14.3193      0.066    218.398      0.000      14.191      14.448
subway_station_count    -6.6591      0.671     -9.927      0.000      -7.974      -5.344
bike_station_count      39.8710      0.305    130.837      0.000      39.274      40.468
business_count           0.0257      0.001     19.602      0.000       0.023       0.028
Bronx                   -5.5480      1.045     -5.312      0.000      -7.595      -3.501
Brooklyn                -9.4744      0.844    -11.232      0.000     -11.128      -7.821
Manhattan              162.7351      1.899     85.711      0.000     159.014     166.456
Queens                 -52.0083      0.769    -67.608      0.000     -53.516     -50.501
Staten_Island          -58.9252      1.000    -58.916      0.000     -60.885     -56.965
day_pop                  0.0002   6.59e-05      3.672      0.000       0.000       0.000
night_pop                0.0088   6.07e-05    145.349      0.000       0.009       0.009
==============================================================================
Omnibus:                   131983.642   Durbin-Watson:                   1.494
Prob(Omnibus):                  0.000   Jarque-Bera (JB):         22945976.123
Skew:                           3.878   Prob(JB):                         0.00
Kurtosis:                      64.379   Cond. No.                     2.81e+19
==============================================================================
```

Figure 7.10: FHV night-time inflow in 2019 with Population Variable

```
The following result is 2020 mta night-time inflow regression
                            OLS Regression Results
==============================================================================
Dep. Variable:      mta_inflow_counts   R-squared:                       0.533
Model:                            OLS   Adj. R-squared:                  0.533
Method:                 Least Squares   F-statistic:                 1.842e+04
Date:                Sun, 19 Jul 2020   Prob (F-statistic):               0.00
Time:                        01:32:30   Log-Likelihood:            -1.2710e+06
No. Observations:              145079   AIC:                         2.542e+06
Df Residuals:                  145069   BIC:                         2.542e+06
Df Model:                           9
Covariance Type:            nonrobust
========================================================================================
                           coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------
Intercept             -156.7377      6.117    -25.624      0.000    -168.727    -144.749
venue_count             44.2114      0.620     71.325      0.000      42.996      45.426
subway_station_count   350.4150      6.002     58.378      0.000     338.650     362.180
bike_station_count     -81.1152      2.842    -28.540      0.000     -86.686     -75.545
business_count           1.0713      0.006    167.993      0.000       1.059       1.084
Bronx                    1.6293      9.769      0.167      0.868     -17.518      20.777
Brooklyn               -62.7987      7.941     -7.909      0.000     -78.362     -47.235
Manhattan             -359.4390     17.375    -20.687      0.000    -393.494    -325.383
Queens                  81.3815      7.290     11.164      0.000      67.093      95.670
Staten_Island          182.4892      9.113     20.026      0.000     164.629     200.350
covid19               -329.6804      8.108    -40.662      0.000    -345.572    -313.789
==============================================================================
Omnibus:                   159521.045   Durbin-Watson:                   0.119
Prob(Omnibus):                  0.000   Jarque-Bera (JB):        152211607.422
Skew:                           4.771   Prob(JB):                         0.00
Kurtosis:                     161.395   Cond. No.                     3.31e+17
==============================================================================
```

Figure 7.11: 2020 MTA Night-time Inflow Regression

```
The following result is 2020 mta night-time inflow regression with area variables
                            OLS Regression Results
==============================================================================
Dep. Variable:      mta_inflow_counts   R-squared:                       0.540
Model:                            OLS   Adj. R-squared:                  0.540
Method:                 Least Squares   F-statistic:                 1.418e+04
Date:                Sun, 19 Jul 2020   Prob (F-statistic):               0.00
Time:                        01:32:31   Log-Likelihood:            -1.2700e+06
No. Observations:              145079   AIC:                         2.540e+06
Df Residuals:                  145066   BIC:                         2.540e+06
Df Model:                          12
Covariance Type:            nonrobust
========================================================================================
                           coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------
Intercept             -135.5092      6.788    -19.962      0.000    -148.814    -122.204
venue_count             38.5751      0.641     60.194      0.000      37.319      39.831
subway_station_count   363.0696      6.163     58.908      0.000     350.990     375.150
bike_station_count     -42.2374      2.933    -14.399      0.000     -47.987     -36.488
business_count           0.8047      0.011     70.721      0.000       0.782       0.827
Bronx                    6.6918      9.747      0.687      0.492     -12.412      25.796
Brooklyn              -101.4452      7.897    -12.845      0.000    -116.924     -85.966
Manhattan              -67.0399     18.539     -3.616      0.000    -103.375     -30.704
Queens                  -6.5240      7.346     -0.888      0.374     -20.922       7.874
Staten_Island           32.8082      9.419      3.483      0.000      14.348      51.268
LotArea               7.069e-07   8.44e-08      8.373      0.000    5.41e-07    8.72e-07
ResArea              -6.508e-05   1.58e-06    -41.165      0.000   -6.82e-05    -6.2e-05
OfficeArea             3.63e-05   2.53e-06     14.374      0.000    3.14e-05    4.13e-05
RetailArea               0.0005   1.57e-05     29.891      0.000       0.000       0.001
==============================================================================
Omnibus:                   144216.284   Durbin-Watson:                   0.121
Prob(Omnibus):                  0.000   Jarque-Bera (JB):        152245409.630
Skew:                           3.890   Prob(JB):                         0.00
Kurtosis:                     161.509   Cond. No.                     1.25e+22
==============================================================================
```

Figure 7.12: MTA night-time inflow in 2020 with Area Variable

```
The following result is 2020 mta night-time inflow regression with population variables
                            OLS Regression Results
==============================================================================
Dep. Variable:      mta_inflow_counts   R-squared:                       0.543
Model:                            OLS   Adj. R-squared:                  0.543
Method:                 Least Squares   F-statistic:                 1.722e+04
Date:                Sun, 19 Jul 2020   Prob (F-statistic):               0.00
Time:                        01:32:31   Log-Likelihood:            -1.2696e+06
No. Observations:              145079   AIC:                         2.539e+06
Df Residuals:                  145068   BIC:                         2.539e+06
Df Model:                          10
Covariance Type:            nonrobust
========================================================================================
                           coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------
Intercept             -111.2241      6.359    -17.492      0.000    -123.687     -98.761
venue_count             44.0466      0.614     71.771      0.000      42.844      45.250
subway_station_count   400.8047      6.279     63.833      0.000     388.498     413.111
bike_station_count     -69.3168      2.852    -24.301      0.000     -74.908     -63.726
business_count           0.6344      0.012     51.636      0.000       0.610       0.658
Bronx                   12.2800      9.777      1.256      0.209      -6.883      31.443
Brooklyn               -36.1424      7.895     -4.578      0.000     -51.617     -20.667
Manhattan             -149.5998     17.772     -8.418      0.000    -184.432    -114.767
Queens                  28.0378      7.201      3.894      0.000      13.925      42.151
Staten_Island           34.2003      9.362      3.653      0.000      15.852      52.549
day_pop                  0.0250      0.001     40.595      0.000       0.024       0.026
night_pop               -0.0331      0.001    -58.239      0.000      -0.034      -0.032
==============================================================================
Omnibus:                   147816.881   Durbin-Watson:                   0.121
Prob(Omnibus):                  0.000   Jarque-Bera (JB):        164190392.042
Skew:                           4.066   Prob(JB):                         0.00
Kurtosis:                     167.607   Cond. No.                     5.80e+18
==============================================================================
```

Figure 7.13: MTA night-time inflow in 2020 with Population Variable

```
The following result is 2020 Citi Bike night-time inflow regression
                            OLS Regression Results
==============================================================================
Dep. Variable:     citi_inflow_counts   R-squared:                       0.628
Model:                            OLS   Adj. R-squared:                  0.628
Method:                 Least Squares   F-statistic:                 2.725e+04
Date:                Sun, 19 Jul 2020   Prob (F-statistic):               0.00
Time:                        01:32:32   Log-Likelihood:            -5.1585e+05
No. Observations:              145079   AIC:                         1.032e+06
Df Residuals:                  145069   BIC:                         1.032e+06
Df Model:                           9
Covariance Type:            nonrobust
========================================================================================
                          coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------
Intercept               1.5116      0.034     45.040      0.000       1.446       1.577
venue_count             0.6692      0.003    196.773      0.000       0.663       0.676
subway_station_count   -1.2014      0.033    -36.479      0.000      -1.266      -1.137
bike_station_count      1.9306      0.016    123.804      0.000       1.900       1.961
business_count         -0.0008    3.5e-05    -21.779      0.000      -0.001      -0.001
Bronx                  -1.2833      0.054    -23.941      0.000      -1.388      -1.178
Brooklyn               -1.8154      0.044    -41.668      0.000      -1.901      -1.730
Manhattan               6.9025      0.095     72.403      0.000       6.716       7.089
Queens                 -1.4022      0.040    -35.057      0.000      -1.481      -1.324
Staten_Island          -0.8900      0.050    -17.800      0.000      -0.988      -0.792
covid19                -1.4201      0.044    -31.922      0.000      -1.507      -1.333
==============================================================================
Omnibus:                   193732.367   Durbin-Watson:                   0.276
Prob(Omnibus):                  0.000   Jarque-Bera (JB):        112769879.556
Skew:                           7.260   Prob(JB):                         0.00
Kurtosis:                     138.810   Cond. No.                     3.31e+17
==============================================================================
```

Figure 7.14: Citibike Night-time inflow in 2020

```
The following result is 2020 Citi Bike night-time inflow regression with area variables
                            OLS Regression Results
==============================================================================
Dep. Variable:     citi_inflow_counts   R-squared:                       0.638
Model:                            OLS   Adj. R-squared:                  0.638
Method:                 Least Squares   F-statistic:                 2.133e+04
Date:                Sun, 19 Jul 2020   Prob (F-statistic):               0.00
Time:                        01:32:32   Log-Likelihood:            -5.1389e+05
No. Observations:              145079   AIC:                         1.028e+06
Df Residuals:                  145066   BIC:                         1.028e+06
Df Model:                          12
Covariance Type:            nonrobust
========================================================================================
                          coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------
Intercept               0.1862      0.037      5.033      0.000       0.114       0.259
venue_count             0.6778      0.003    194.031      0.000       0.671       0.685
subway_station_count   -1.0846      0.034    -32.282      0.000      -1.150      -1.019
bike_station_count      1.8047      0.016    112.859      0.000       1.773       1.836
business_count          0.0023    6.2e-05     36.520      0.000       0.002       0.002
Bronx                  -1.4191      0.053    -26.708      0.000      -1.523      -1.315
Brooklyn               -2.0613      0.043    -47.881      0.000      -2.146      -1.977
Manhattan               5.6709      0.101     56.115      0.000       5.473       5.869
Queens                 -1.4635      0.040    -36.544      0.000      -1.542      -1.385
Staten_Island          -0.5408      0.051    -10.533      0.000      -0.641      -0.440
LotArea              4.431e-09    4.6e-10      9.627      0.000    3.53e-09    5.33e-09
ResArea              1.187e-07   8.62e-09     13.776      0.000    1.02e-07    1.36e-07
OfficeArea          -7.502e-07   1.38e-08    -54.498      0.000   -7.77e-07   -7.23e-07
RetailArea          -1.561e-06   8.56e-08    -18.230      0.000   -1.73e-06   -1.39e-06
==============================================================================
Omnibus:                   189687.683   Durbin-Watson:                   0.284
Prob(Omnibus):                  0.000   Jarque-Bera (JB):        103773056.848
Skew:                           6.981   Prob(JB):                         0.00
Kurtosis:                     133.276   Cond. No.                     1.25e+22
==============================================================================
```

Figure 7.15: Citibike night-time inflow in 2020 with Area Variable

```
The following result is 2020 Citi Bike night-time inflow regression with population variables
                            OLS Regression Results
==============================================================================
Dep. Variable:     citi_inflow_counts   R-squared:                       0.629
Model:                            OLS   Adj. R-squared:                  0.629
Method:                 Least Squares   F-statistic:                 2.460e+04
Date:                Sun, 19 Jul 2020   Prob (F-statistic):               0.00
Time:                        01:32:32   Log-Likelihood:            -5.1571e+05
No. Observations:              145079   AIC:                         1.031e+06
Df Residuals:                  145068   BIC:                         1.032e+06
Df Model:                          10
Covariance Type:            nonrobust
========================================================================================
                           coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------
Intercept                0.6591      0.035     18.715      0.000       0.590       0.728
venue_count              0.6684      0.003    196.643      0.000       0.662       0.675
subway_station_count    -1.1787      0.035    -33.896      0.000      -1.247      -1.111
bike_station_count       1.9357      0.016    122.532      0.000       1.905       1.967
business_count           0.0011    6.8e-05     16.077      0.000       0.001       0.001
Bronx                   -1.3809      0.054    -25.503      0.000      -1.487      -1.275
Brooklyn                -2.0269      0.044    -46.354      0.000      -2.113      -1.941
Manhattan                6.3938      0.098     64.962      0.000       6.201       6.587
Queens                  -1.4800      0.040    -37.114      0.000      -1.558      -1.402
Staten_Island           -0.8470      0.052    -16.336      0.000      -0.949      -0.745
day_pop                 -0.0001   3.42e-06    -31.523      0.000      -0.000      -0.000
night_pop              6.31e-05   3.15e-06     20.057      0.000    5.69e-05    6.93e-05
==============================================================================
Omnibus:                   191686.879   Durbin-Watson:                   0.277
Prob(Omnibus):                  0.000   Jarque-Bera (JB):        107280816.104
Skew:                           7.121   Prob(JB):                         0.00
Kurtosis:                     135.455   Cond. No.                     5.80e+18
==============================================================================
```

Figure 7.16: Citibike night-time inflow in 2020 with Population Variable

```
The following result is 2019 MTA inflow regression with Weather and time factors
                            OLS Regression Results
==============================================================================
Dep. Variable:      mta_inflow_counts   R-squared:                       0.022
Model:                            OLS   Adj. R-squared:                  0.022
Method:                 Least Squares   F-statistic:                     713.8
Date:                Sat, 18 Jul 2020   Prob (F-statistic):               0.00
Time:                        21:10:57   Log-Likelihood:            -2.9516e+06
No. Observations:              287760   AIC:                         5.903e+06
Df Residuals:                  287750   BIC:                         5.903e+06
Df Model:                           9
Covariance Type:            nonrobust
=================================================================================
                    coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------
Intercept      2041.3627     45.199     45.164      0.000    1952.775    2129.951
Temp_Avg          5.1364      1.194      4.301      0.000       2.796       7.477
Precipitation  -190.9686     45.319     -4.214      0.000    -279.793    -102.145
Monday          387.4297     32.085     12.075      0.000     324.545     450.315
Tuesday         473.5121     31.388     15.086      0.000     411.993     535.031
Wednesday       554.0703     32.069     17.277      0.000     491.216     616.925
Thursday        555.8469     32.158     17.285      0.000     492.819     618.875
Friday          555.7554     32.569     17.064      0.000     491.921     619.590
Saturday       -116.5270     32.557     -3.579      0.000    -180.338     -52.716
Sunday         -368.7249     32.706    -11.274      0.000    -432.828    -304.622
night         -1933.3409     25.701    -75.225      0.000   -1983.714   -1882.968
==============================================================================
Omnibus:                   462211.166   Durbin-Watson:                   1.553
Prob(Omnibus):                  0.000   Jarque-Bera (JB):        370145047.387
Skew:                          10.542   Prob(JB):                         0.00
Kurtosis:                     177.432   Cond. No.                     3.57e+16
==============================================================================
```

Figure 7.17: MTA inflow with Weather and Time Factor in 2019

```
The following result is 2019 Citi inflow regression with Weather and time factors
                            OLS Regression Results
==============================================================================
Dep. Variable:     citi_inflow_counts   R-squared:                       0.018
Model:                            OLS   Adj. R-squared:                  0.018
Method:                 Least Squares   F-statistic:                     584.5
Date:                Sat, 18 Jul 2020   Prob (F-statistic):               0.00
Time:                        21:10:58   Log-Likelihood:            -1.6843e+06
No. Observations:              287760   AIC:                         3.369e+06
Df Residuals:                  287750   BIC:                         3.369e+06
Df Model:                           9
Covariance Type:            nonrobust
=================================================================================
                    coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------
Intercept         8.1321      0.553     14.715      0.000       7.049       9.215
Temp_Avg          0.3850      0.015     26.366      0.000       0.356       0.414
Precipitation    -3.5112      0.554     -6.336      0.000      -4.597      -2.425
Monday            1.6974      0.392      4.327      0.000       0.928       2.466
Tuesday           2.5405      0.384      6.619      0.000       1.788       3.293
Wednesday         3.1941      0.392      8.146      0.000       2.426       3.963
Thursday          2.0220      0.393      5.142      0.000       1.251       2.793
Friday            1.0700      0.398      2.687      0.007       0.289       1.851
Saturday         -0.3708      0.398     -0.931      0.352      -1.151       0.409
Sunday           -2.0211      0.400     -5.054      0.000      -2.805      -1.237
night           -21.0277      0.314    -66.914      0.000     -21.644     -20.412
==============================================================================
Omnibus:                   455712.919   Durbin-Watson:                   1.202
Prob(Omnibus):                  0.000   Jarque-Bera (JB):        234928581.693
Skew:                          10.454   Prob(JB):                         0.00
Kurtosis:                     141.407   Cond. No.                     3.57e+16
==============================================================================
```

Figure 7.18: Citibike inflow with Weather and Time Factor in 2019

```
The following result is 2019 FHV inflow regression with Weather and time factors
                            OLS Regression Results
==============================================================================
Dep. Variable:      fhv_inflow_counts   R-squared:                       0.020
Model:                            OLS   Adj. R-squared:                  0.020
Method:                 Least Squares   F-statistic:                     664.5
Date:                Sat, 18 Jul 2020   Prob (F-statistic):               0.00
Time:                        21:10:58   Log-Likelihood:            -2.2563e+06
No. Observations:              287760   AIC:                         4.513e+06
Df Residuals:                  287750   BIC:                         4.513e+06
Df Model:                           9
Covariance Type:            nonrobust
=================================================================================
                    coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------
Intercept       307.2272      4.034     76.157      0.000     299.320     315.134
Temp_Avg         -0.1570      0.107     -1.473      0.141      -0.366       0.052
Precipitation     2.1624      4.045      0.535      0.593      -5.765      10.090
Monday            8.9422      2.864      3.123      0.002       3.329      14.555
Tuesday          16.2817      2.801      5.812      0.000      10.791      21.773
Wednesday        19.9086      2.862      6.955      0.000      14.299      25.519
Thursday         35.7727      2.870     12.464      0.000      30.147      41.398
Friday           66.5673      2.907     22.900      0.000      60.870      72.265
Saturday        101.2542      2.906     34.845      0.000      95.559     106.950
Sunday           58.5006      2.919     20.040      0.000      52.779      64.222
night          -166.3859      2.294    -72.534      0.000    -170.882    -161.890
==============================================================================
Omnibus:                   350129.514   Durbin-Watson:                   0.993
Prob(Omnibus):                  0.000   Jarque-Bera (JB):         49167022.819
Skew:                           6.637   Prob(JB):                         0.00
Kurtosis:                      65.646   Cond. No.                     3.57e+16
==============================================================================
```

Figure 7.19: FHV inflow with Weather and Time Factor in 2019

```
The following result is 2019 Venue counts regression with Business Counts
                            OLS Regression Results
==============================================================================
Dep. Variable:            venue_count   R-squared:                       0.677
Model:                            OLS   Adj. R-squared:                  0.677
Method:                 Least Squares   F-statistic:                 7.529e+04
Date:                Thu, 16 Jul 2020   Prob (F-statistic):               0.00
Time:                        23:19:42   Log-Likelihood:            -9.4587e+05
No. Observations:              287760   AIC:                         1.892e+06
Df Residuals:                  287751   BIC:                         1.892e+06
Df Model:                           8
Covariance Type:            nonrobust
======================================================================================
                         coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------
Intercept             -0.0585      0.016     -3.667      0.000      -0.090      -0.027
citi_inflow_counts     0.0196      0.000     83.610      0.000       0.019       0.020
mta_inflow_counts     -0.0001   3.03e-06    -38.419      0.000      -0.000      -0.000
fhv_inflow_counts      0.0066   4.38e-05    150.183      0.000       0.006       0.007
business_count         0.0045   1.93e-05    235.882      0.000       0.005       0.005
Bronx                 -1.0529      0.028    -36.947      0.000      -1.109      -0.997
Brooklyn              -0.2610      0.023    -11.198      0.000      -0.307      -0.215
Manhattan              1.4468      0.053     27.333      0.000       1.343       1.551
Queens                -0.1814      0.022     -8.346      0.000      -0.224      -0.139
Staten_Island         -0.0100      0.028     -0.363      0.717      -0.064       0.044
==============================================================================
Omnibus:                   345604.042   Durbin-Watson:                   1.915
Prob(Omnibus):                  0.000   Jarque-Bera (JB):         99582303.769
Skew:                           6.149   Prob(JB):                         0.00
Kurtosis:                      93.301   Cond. No.                     4.01e+19
==============================================================================
```

Figure 7.20: Venue Counts with business counts

```
The following result is 2019 Venue counts regression with Area variables
                            OLS Regression Results
==============================================================================
Dep. Variable:            venue_count   R-squared:                       0.701
Model:                            OLS   Adj. R-squared:                  0.701
Method:                 Least Squares   F-statistic:                 5.617e+04
Date:                Fri, 17 Jul 2020   Prob (F-statistic):               0.00
Time:                        00:06:47   Log-Likelihood:            -9.3473e+05
No. Observations:              287760   AIC:                         1.869e+06
Df Residuals:                  287747   BIC:                         1.870e+06
Df Model:                          12
Covariance Type:            nonrobust
======================================================================================
                         coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------
Intercept             -0.0805      0.020     -4.043      0.000      -0.119      -0.041
citi_inflow_counts     0.0198      0.000     86.235      0.000       0.019       0.020
mta_inflow_counts     -0.0001   2.94e-06    -50.482      0.000      -0.000      -0.000
fhv_inflow_counts      0.0061   4.51e-05    135.808      0.000       0.006       0.006
LotArea             1.836e-09   2.44e-10      7.533      0.000    1.36e-09    2.31e-09
ResArea            -2.175e-07    4.5e-09    -48.371      0.000   -2.26e-07   -2.09e-07
OfficeArea         -3.968e-07   7.28e-09    -54.532      0.000   -4.11e-07   -3.83e-07
RetailArea          6.469e-06   4.35e-08    148.649      0.000    6.38e-06    6.55e-06
business_count         0.0033    3.3e-05    100.014      0.000       0.003       0.003
Bronx                 -1.0218      0.028    -37.135      0.000      -1.076      -0.968
Brooklyn              -0.4774      0.023    -21.177      0.000      -0.522      -0.433
Manhattan              2.0489      0.054     37.849      0.000       1.943       2.155
Queens                -0.4123      0.021    -19.375      0.000      -0.454      -0.371
Staten_Island         -0.2178      0.027     -7.933      0.000      -0.272      -0.164
==============================================================================
Omnibus:                   323643.942   Durbin-Watson:                   1.966
Prob(Omnibus):                  0.000   Jarque-Bera (JB):         76007089.680
Skew:                           5.501   Prob(JB):                         0.00
Kurtosis:                      81.855   Cond. No.                     2.31e+23
==============================================================================
```

Figure 7.21: Venue Counts with Built Environment features

```
The following result is 2019 Venue counts with Population Variables
                            OLS Regression Results
==============================================================================
Dep. Variable:            venue_count   R-squared:                       0.677
Model:                            OLS   Adj. R-squared:                  0.677
Method:                 Least Squares   F-statistic:                 6.024e+04
Date:                Thu, 16 Jul 2020   Prob (F-statistic):               0.00
Time:                        23:19:45   Log-Likelihood:            -9.4586e+05
No. Observations:              287760   AIC:                         1.892e+06
Df Residuals:                  287749   BIC:                         1.892e+06
Df Model:                          10
Covariance Type:            nonrobust
======================================================================================
                         coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------
Intercept             -0.1124      0.019     -5.769      0.000      -0.151      -0.074
citi_inflow_counts     0.0198      0.000     83.057      0.000       0.019       0.020
mta_inflow_counts     -0.0001   3.05e-06    -38.219      0.000      -0.000      -0.000
fhv_inflow_counts      0.0065   4.59e-05    141.719      0.000       0.006       0.007
business_count         0.0045   3.63e-05    123.969      0.000       0.004       0.005
night_pop           7.366e-06   1.68e-06      4.382      0.000    4.07e-06    1.07e-05
day_pop             2.857e-06   1.84e-06      1.554      0.120   -7.47e-07    6.46e-06
Bronx                 -1.0725      0.029    -37.247      0.000      -1.129      -1.016
Brooklyn              -0.2809      0.024    -11.865      0.000      -0.327      -0.234
Manhattan              1.3907      0.054     25.639      0.000       1.284       1.497
Queens                -0.1742      0.022     -7.995      0.000      -0.217      -0.132
Staten_Island          0.0244      0.028      0.860      0.390      -0.031       0.080
==============================================================================
Omnibus:                   345770.745   Durbin-Watson:                   1.917
Prob(Omnibus):                  0.000   Jarque-Bera (JB):         99349911.658
Skew:                           6.156   Prob(JB):                         0.00
Kurtosis:                      93.191   Cond. No.                     1.31e+20
==============================================================================
```

Figure 7.22: Venue Counts with Demographic features