Understanding the Impact of the Nightlife Industry on Urban Mobility in NYC and Charting COVID-19 Related Disruption

Progress Report 2

Kaifu Ren (kr2516), Nicholas LiCalzi (nl949)

Yingyuan Zhang (yz6378), Yutong Zhu (yz6080)

New York University CUSP

May 3, 2020

Sponsor: VibeLab

CUSP Mentor: Kim Mahler

Abstract

The nightlife industry (bars, restaurants, clubs, music venues, and more) is a major economic and cultural engine for New York City. The goal of our research is to capture and represent the vibrancy of New York City and its venues at night through the lens of its transportation network and the people that use it, whether by bike, car, bus, or train. In doing so, we might be able to offer targeted recommendations for ensuring that nightlife activities are adequately serviced by transit and help stakeholders both in the industry and in government collaborate in a productive, informed manner. Our original research question sought to answer whether we can leverage transportation data to better understand the spatial, cultural, and economic impacts of nightlife activities in NYC; with the arrival of the COVID-19 pandemic, we've expanded it to explore whether that same data reveals that nightlife venues and businesses have been disproportionately harmed by the arrival of COVID and ensuing business closures? To answer these questions, we will leverage transportation, built environment, business typology/opening hours, health, and other datasets across varying geographic scopes. We expect to find specific clusters of mode choice preference throughout the city, with some neighborhoods experiencing higher rates of nighttime bike ridership than others, for example, and hope to be able to correlate that with some of our collected built environment or demographic data to explain the phenomenon. Our work will demonstrate that nightlife activity is geographically diffuse throughout the city, while some areas act as particularly dense hotspots, decreasing the travel time/distance required for both consumers and workers, and highlighting bars and clubs as local foci of culture and economic activity across neighborhoods.

Introduction

The nightlife industry makes up a significant portion of New York City's economy and culture. In recent years, and following significant organizing by local activists, the NY City Council passed legislation forming the Mayor's Office of Nightlife in an attempt to foster communication and collaboration between industry stakeholders and the city government, in recognition of nightlife's importance (Feuer, 2018). However, since the appointment of Ariel Palitz, the city's first night mayor, one of the major sponsors of the legislation creating the office has criticised the mayor for under-staffing and -funding the office and their work (Sanders, 2019). VibeLab, our project sponsor and an international nightlife advocacy consultancy, approached us with the goal of performing research that nightlife industry stakeholders might use to demonstrate the value and importance of their work to the city government as an advocacy tool. We hope that our work might be of use to citizens, visitors, patrons, workers, owners, and other participants in/beneficiaries of the city's rich night-time tapestry.

The arrival and rapid dissemination of COVID-19 across NYC, resulting in closures and massive disruptions to daily life, has forced a reframing of our research. While NYC will certainly suffer a huge economic loss due to the forced shutdown and bans on public gatherings and venues closed citywide; the nightlife industry in NYC, largely fueled by freelancers, is experiencing an existentially threatening financial fallout. Having established a snapshot of the urban transportation network and its night time utilization during "normal" times, the research team will then inspect the impact of COVID-19 on the city's nightlife economy, looking for any indications that nightlife has been disproportionately affected by the closures. We hope to capture and depict the scale of losses experienced by the transportation system and nightlife businesses in NYC, and hope that our work might be used to assist policymakers and city managers in deliberations with industry stakeholders and possibly the allocation of resources for financial aid.

Literature Review

Studies of nightlife in cities have often focused on the so called "Night Time Economy" (NTE), and the utility of nightlife as an economic development indicator or engine for tourism, but a more useful framing is one that operates through an equity lens-- considering differences in service and experience based on gender, economic status, and accessibility to understand how various groups interact with the transportation network, particularly at night (McArthur et al., 2019) (Plyushteva & Boussaw, 2020). Researchers have proposed various framings, including those of racial segregation, environmental exposure, and accessibility, that might impact human mobility, individual accessibility, and spacetime travel patterns in cities (Kwan, 2013), and used those factors to advocate for new, more holistic transportation planning and decisionmaking processes to improve equity outcomes (Jones & Lucas, 2012). Others have focused primarily on time, and how the opening hours of public services like libraries might variably affect individuals based on certain personal or household attributes-- something that's interesting to consider, given NYC's somewhat unusual 24 hour provision of transportation services (Neutens et al, 2010) (Delafontain et al, 2011).

We have also found it useful to explore the literature around some of our methods in order to validate their utility. There is a large body of work around the specific issue of ZIP codes being particularly poorly suited to spatial analysis and the need for that data to be disaggregated before it is usable (Grubesic & Matisziw, 2006). One method that many have found useful is the use of areal interpolation (Comber & Zeng, 2019) for the refitting of data into dasymetric maps, though there continues to be some disagreement over what exactly constitutes dasymetric mapping (Eicher & Brewer, 2001). Other researchers have found success using geographically weighted expectation-maximization algorithms for such interpolation (Schroeder & Van Riper, 2013), a technique that we may explore for our processes of dis-aggregation and re-aggregation to a hexbin scheme (Langton & Solymosi, 2019).

Problem Statement

We seek to capture and analyze the relative utilization rates of various transportation methods across the city, both before the COVID-19 pandemic and during it, in order to understand the state of nighttime travel pre-COVID and how it might have been uniquely affected during the crisis. We seek to answer the following questions and others: Where are the major nexuses of nightlife activity in NYC? How do people normally reach them? Are there any obvious gaps in service? Are there spatial differences in transportation mode preferences? Are there differences in ridership based on land use patterns, venue density, or any number of other data-points captured in our search? Do any detected patterns or clustering hold true for both the pre- and mid-COVID periods? If not, can we find some reason they might deviate?

Timeline of Government Interventions and Our Expectations

Since the detection of the first COVID-19 case in the US in early March, NYC and NYS have taken a variety of efforts to attempt to contain the spread of the disease. On March 7th, Governor Cuomo declared a state of emergency, and over the next few days the government requested that people avoid dense crowds and take alternate methods of transportation to buses, cars, and subways. On March 12th, the state banned events with over 500 people in attendance, and on March 17th the governor signed an executive order shuttering all restaurants, bars, and clubs to all activity besides takeout and delivery (Cuomo, 2020) . On March 22nd, all non-essential businesses were ordered shut. (De Blasio, 2020).

We have certain hypotheses that we expect to see reflected in the transportation data, given the above timeline of measures that were implemented. We expect to see a decline in public transit ridership, with March 8th as an inflection point, with a corresponding increase in Citibike and FHV utilization. Perhaps nighttime transit usage will begin to decline as people are still required to go to work, but socializing in large groups at night becomes more of a risky proposition. Overall transportation utilization should drop on March 12th and 17th as more and more categories of businesses are shuttered, especially the full shuttering of nighttime gathering spots.

Data

A core part of our project involves spatial analysis, so we have a plethora of geospatial data that we are relying upon. We have a file with X, Y coordinates of nightlife venues in the city collected by VibeLab as part of their Creative Footprint project, and we've supplemented that with data previously scraped from Google Maps, containing geospatial records of all music venues, lounges, and bars in the city, along with their opening hours, in an attempt to rectify any imbalances (geographical or selection-specific) in the community survey data. We will also use the DCP PLUTO tax lot dataset for information about the built environment, Census data for population and demographic information, and more.

Referring to Appendix I, the primary traffic data used for analysis are MTA real-time turnstile data, TLC monthly trip record, and Citibike monthly trip records. MTA data provides accumulated entry and exit counts indexed by each subway station and subway line in NYC for every four hour interval. However, the geospatial coordinate of each station is not included. TLC provides trip records for operating yellow taxi, green taxi and for hired vehicles including rideshare services like Uber or Lyft. For data privacy, the pickup and dropoff locations for each trip has been anonymised into NYC taxi zones[1]. TLC release their trip records every half of year which makes the near real-time data processing impractical. Citibike provides their trip records based on the starting and ending bike stations with spatial coordinates. Time information up to seconds is shown with each row of TLC and Citibike trip records.

The limitation with the structure of these three datasets can be observed. Firstly, they are aggregated to different geometries including entrance location of the station for MTA, bike racks location for Citibike, and taxi zone polygons for TLC. Secondly, they provide time information in different formats such as four-hour intervals for MTA, and complete time series for TLC and Citibike.

Starting on March 31th 2020, NYC Health released the daily-updated number of Covid-19 cases aggregated by NYC zip codes. Again, the data source needs to convert to the same geometries to have the joint analysis with other datasets.

Methods

Stated as the limitation in the data section, we need to reconcile the geometries for different data sources in the same format. Considering only TLC data were aggregated based on taxi zone polygons and to avoid conflict and loss of information, we decided to perform spatial joins for the other two data sources into NYC taxi zones. We used booth ID in the MTA data to map the spatial coordinate for each subway station. Thus, spatial joins operation can affix data points in MTA and

[1] Taxi zones are defined based on NYC Department of City Planning's Neighborhood Tabulation Areas (NTAs). (Data.gov)

Citibike to taxi zones respectively. However, as discussed in the literature review, taxi zones are not the ideal geospatial geometry due to regular census boundaries, we will try to perform spatial joins of traffic data into hexbins in the next step. The success of our project hinges on our ability to accurately process multiple vector based geodatasets, varying in size and shape and all overlapping, into a uniform raster structure that can then be analysed with typical machine learning and spatial analysis algorithms for clustering, outlier, and pattern detection to drive our insights. We will use areal interpolation and dasymetric mapping techniques to disaggregate our data before building it back up, accurately and at as high a resolution as possible.

Trip records provided by different sources are in gigabytes. No single computer can handle these large data tables and store them in memory. To perform adequate analysis over the huge amount of data, we need to rely on big data techniques. Streaming technique which allows us to process trip records row by row and doesn't consume the memory tends to be a good fit for spatial counts based on the extracted data timeline. Our objective is trying to record the time up to hour of day for each instance arriving in the selected taxi zone (traffic inflow), leaving for that zone (traffic outflow) and traveling inside that zone (flow inside).

Traffic mode share analysis can help us to understand how New Yorkers access these music venues via different means of transportation. We can calculate the proportion of any transportation usage in selected modes at a certain time point or timeframe created by the previous streaming techniques. Comparing the changes in traffic mode share can help decision makers to adjust the traffic management system. Manipulating the processed spatial counts in the previous step, we can monitor hourly, daily or nightly changes of inflow and outflow for different mobility usage in selected taxi zones or music venues' hotspots. Time series analysis can help to get insights for anomaly identification and trend summary for selected timelines.

Comparison study will be carried out in two-folds. We will compare the observation in 2020 with the baseline model in 2019 to show differences in mobility demands potentially caused by the outbreaks of Covid-19. On the other hand, we will overlay 2020 daily mobility usage with key government actions and daily reports of COVID-19 cases in selected areas to draw the implication of those measures over people's travel behaviors.

Updated Project Timeline

Referring to the project timeline in Appendix II, we have completed the data collection and various data analysis parts. For the next step, we will reconcile all the trip information based on the modifiable areal unit and perform time series analysis. Our next big push is the consolidation of our datasets and a renewed effort towards data analysis and visualization.

Project Risks and Mitigation Strategies

The first project risk is related to data limitation. TLC publishes their trip records dataset every half of year. The earliest date we can get those taxi trip records will be early July this year. Currently, we are using trip records in 2019 as a substitute and to perform baseline modeling. We have built the data processing pipeline ready. When the data arrives, it can be processed immediately.

Secondly, cities are complex systems. Passengers take different means of transportation for different purposes. Assumptions to construct the causal relationship between traffic volumes and nearby music venues tend to be unjustified. Specifically, it’s hard to define how much change in traffic volume is caused by nightlife events. We will conduct traffic pattern analysis to depict the influences. In order to keep the balance of the research generalization, we need to limit our assumptions in reasonable manners .

We are working around the Modifiable Areal Unit Problem where data sources are collected in different geometries. We need to outsource extra information like living and working population, building units and footprints to disaggregate and organize those data in the right geometries.

Team Collaboration Statement

Kaifu Ren: TLC data engineering and cleaning; building big data pipeline for processing; baseline analysis; data visualization

Nicholas LiCalzi: venue data collection and aggregation; spatial analysis and geovisualization; literature review on background and methods; compiling, writing, and editing report

Yutong Zhu: MTA data engineering and cleaning; Covid-19 related literature research, research hypothesis; data visualization

Yingyuan Zhang: Citi Bike data engineering and cleaning; transportation system news research; data visualization

Bibliography

1. Comber, Alexis, and Wen Zeng. “Spatial Interpolation Using Areal Features: A Review of Methods and Opportunities Using New Forms of Data with Coded Illustrations.” *Geography Compass*, vol. 13, no. 10, June 2019, doi:10.1111/gec3.12465.
2. Delafontaine, Matthias, et al. “The Impact of Opening Hours on the Equity of Individual Space–Time Accessibility.” *Computers, Environment and Urban Systems*, vol. 35, no. 4, 2011, pp. 276–288., doi:10.1016/j.compenvurbsys.2011.02.005.
3. Eicher, Cory L., and Cynthia A. Brewer. “Dasymetric Mapping and Areal Interpolation: Implementation and Evaluation.” *Cartography and Geographic Information Science*, vol. 28, no. 2, 2001, pp. 125–138., doi:10.1559/152304001782173727.
4. Feuer, Alan. *New York City Appoints Its First Nightlife Mayor*. The New York Times, 7 Mar. 2018, www.nytimes.com/2018/03/07/nyregion/nyc-nightlife-mayor.html.
5. Grubesic, Tony H, and Timothy C Matisziw. “On the Use of ZIP Codes and ZIP Code Tabulation Areas (ZCTAs) for the Spatial Analysis of Epidemiological Data.” *International Journal of Health Geographics*, 5 58, 13 Dec. 2006, doi:10.1186/1476-072X-5-58.
6. “How Mayor De Blasio Can Make Biking to Work a Possibility for More New Yorkers.” *Transportation Alternatives*, 9 Mar. 2020, www.transalt.org/news/releases/12718.
7. Hu, Winnie. “A Surge in Biking to Avoid Crowded Trains in N.Y.C.” *The New York Times*, 14 Mar. 2020, www.nytimes.com/2020/03/14/nyregion/coronavirus-nyc-bike-commute.html.
8. Jones, Peter, and Karen Lucas. “The Social Consequences of Transport Decision-Making: Clarifying Concepts, Synthesising Knowledge and Assessing Implications.” *Journal of Transport Geography*, vol. 21, 2012, pp. 4–16., doi:10.1016/j.jtrangeo.2012.01.012.
9. Kwan, Mei-Po. “Beyond Space (As We Knew It): Toward Temporally Integrated Geographies of Segregation, Health, and Accessibility.” *Space-Time Integration in Geography and GIScience*, 2014, pp. 39–51., doi:10.1007/978-94-017-9205-9_4.
10. Langton, Samuel, and Reka Solymosi. “Cartograms, Hexograms and Regular Grids: a Commentary on Minimising Misrepresentation in Spatial Data Visualisations.” Feb. 2019, doi:10.31235/osf.io/t6agd.
11. McArthur, Jenny, et al. “Socio-Spatial and Temporal Dimensions of Transport Equity for Londons Night Time Economy.” *Transportation Research Part A: Policy and Practice*, vol. 121, 2019, pp. 433–443., doi:10.1016/j.tra.2019.01.024.
12. Neutens, Tijs, et al. “Evaluating the Temporal Organization of Public Service Provision Using Space-Time Accessibility Analysis.” *Urban Geography*, vol. 31, no. 8, 2010, pp. 1039–1064., doi:10.2747/0272-3638.31.8.1039.
13. “During Novel Coronavirus Briefing, Governor Cuomo Announces New Mass Gatherings Regulations.” Governor Andrew M. Cuomo, New York State, 17 Mar. 2020, www.governor.ny.gov/news/during-novel-coronavirus-briefing-governor-cuomo-announces-new-mass-gatherings-regulations.
14. “Mayor De Blasio Issues New Guidance to New Yorkers.” *The Official Website of the City of New York*, 20 Mar. 2020, www1.nyc.gov/office-of-the-mayor/news/173-20/mayor-de-blasio-issues-new-guidance-new-yorkers.
15. “Statement From Mayor De Blasio on Bars, Restaurants, and Entertainment Venues.” *The Official Website of the City of New York*, 15 Mar. 2020, www1.nyc.gov/office-of-the-mayor/news/152-20/statement-mayor-de-blasio-bars-restaurants-entertainment-venues.
16. Plyushteva, Anna, and Kobe Boussauw. “Does Night-Time Public Transport Contribute to Inclusive Night Mobility? Exploring Sofias Night Bus Network from a Gender Perspective.” *Transport Policy*, vol. 87, 2020, pp. 41–50., doi:10.1016/j.tranpol.2020.01.002.
17. Sanders, Anna. *Mayor's Office of Nightlife 'Drastically Underfunded and Understaffed,' Pol Says*. New York Daily News, 28 Aug. 2019, www.nydailynews.com/news/politics/ny-mayors-office-nightlife-underfunded-and-understaffed-20190828-5rvycmj3r5arnhrej62iquuyka-story.html.
18. Schroeder, Jonathan P., and David C. Van Riper. “Because Muncies Densities Are Not Manhattans: Using Geographical Weighting in the Expectation-Maximization Algorithm for

Areal Interpolation.” *Geographical Analysis*, vol. 45, no. 3, 2013, pp. 216–237., doi:10.1111/gean.12014.

19. Spivack, Caroline. “Here's How the Coronavirus Pandemic Is Affecting Public Transit.” *Curbed NY*, Curbed NY, 24 Mar. 2020, ny.curbed.com/2020/3/24/21192454/coronavirus-nyc-transportation-subway-citi-bike-covid-19.
20. Lapin, Tamar. “Coronavirus in NY: A Timeline of How the Disease Spread through the Metro Area.” *New York Post*, New York Post, 18 Mar. 2020, nypost.com/2020/03/12/coronavirus-in-ny-a-timeline-of-how-the-disease-spread-through-the-metro-area/.

Appendix I Data Sources

System	Date Modified	Link	Frequency	Action
Citibike-tripdata	Feb 2020	https://s3.amazonaws.com/tripdata/index.html	Monthly Updated	Available
MTA Real-time Turnstile Data	Apr 04 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	Make request to TLC
MTA Bus Ridership Data	To present	https://bustime.mta.info/wiki/Developers/Index	Daily Updated	Make MTA API calls
Geographical Features (Subway lines, bus stops, taxi zones, zip codes, etc.)	-	NYC open data portal	-	Available
Music Venues & Information (Creative Footprint NYC - VibeLab)	2019	https://drive.google.com/drive/u/1/folders/1fKMt3mrJVU0vn5YVUeq10W88FufhwYfH	One time survey	Available
Event Information (Ticketmaster, Eventbrite)	To present	-	Daily Updated	Make API calls
COVID-19 dataset	To present	https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset#COVID19_line_list_data.csv	Daily Updated	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

Appendix II Project Plan

	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12	W13	W14	W15	W16	W17	W18
	Mar. 30	**Apr. 6**	**Apr. 13**	**Apr. 20**	**Apr. 27**	**May 4**	**May 11**	**May 18**	**May 25**	**Jun. 1**	**Jun. 8**	**Jun. 15**	**Jun. 22**	**Jun. 29**	**Jul. 6**	**Jul. 13**	**Jul. 20**	**Jul. 23**
Preparation & Data Acquisition																		
Structuring Weekly Working Log w/ Mentor																		
Literature Review																		
Proprietary datasets (Creative Footprint, NYC Mayor's Office of Nightlife)																		
Open data sources (TLC, MTA, Citibike, Census)																		
Scraped data sources (Eventbrite, Resident Advisor, Google Maps)																		
Outreach / Interviews (Govt. & Industry Stakeholders)																		
Data Analysis																		
Exploratory Data Analysis (Jupyter Notebooks)																		
Time Series Analysis for Taxis (Jupyter Notebook)																		
Time Series Analysis for MTA Turnstiles (Jupyter Notebook)																		
Time Series Analysis for CitiBike (Jupyter Notebook)																		
Spatial Analysis (Autocorrelation & other methods)																		
Data Visualization																		
Mapping nightlife hotspots (Venue Density Isopleth)																		
Mapping night time transit activity near nightlife hotspots (pre-COVID)																		
Mapping night time transit activity near nightlife hotspots (early COVID)																		
Mapping night time transit activity near nightlife hotspots (after closures)																		
Visualization of spatial autocorrelation analysis																		
Discussion of Findings & Implications																		
DUE: Second Report & Presentation					x (Apr. 29)													
DUE: Third Report & Presentation													x (Jun. 22)					
Finalization of Project GitHub Repo (with descriptive README)																		
Development of Project Website																		
Design Final Presentation																		
DUE: Finalize Technical Project Report																		x (Jul. 23)