COVID-19 Closure Impacts on Food-related, Non-chain, Businesses in Select New York City Communities

Final Report

Ari Lewenstein (al6740)

Shu Wang (sw5034)

Haochen Xing (hx2099)

Sponsor: Urban Modeling Group

Mentor: Prof. Debra Laefer

New York University CUSP

November 22nd, 2021

Abstract

The COVID-19 pandemic created devastating effects for small businesses, despite large, targeted relief programs. As such, this project uses the Spring 2020 New York State COVID-19 PAUSE order as a stressor for business closure to identify small business resiliency factors to inform both urban planners and business owners. To do this, three modeling approaches were tested to identify socio-economic, spatial, and business characteristics associated with business closures. This was done for 2,186 food related retail businesses in eight zip codes in New York City. Footfall data from cell phone signals were used to identify small businesses, their type, operational status, and monthly customer visits. Supplementary analysis was done with data from path tracking of 2,215 individuals leaving eight healthcare facilities. Our three methods allowed us to predict the accuracy of business closures at rates between 86% and 95%. Lower commercial diversity and larger distances to subway stops correlated with a higher risk of business closures. Using path data, we found that businesses between healthcare facilities and subway stops appeared less likely to close than others. These findings show the value of diverse commercial development near transit areas.

1. Introduction

1.1 Motivation

The arrival of the COVID-19 virus in New York City (NYC) on February 29th, 2020 created a devastating, complex crisis for NYC, which resulted first in advisories against densely packed public transit, then prohibitions of gatherings exceeding 50 people, and subsequently the closure of schools, bars, and in-person dining (March 17th), followed by the Governor of New York's PAUSE order, which shuttered all non-essential businesses as of March 22nd (Adcroft, 2021). Despite 15 billion dollars in aid to NYC businesses, an estimated 12% of them closed permanently within a year of COVID-19's onset in NYC

(Braun 2021, Fonrouge 2021). Many were small businesses and restaurants that have been key contributors to NYC's economy, generating nearly 25% of new hires over the past two decades (NYC Small Business First, 2018). As such, this report uses the Governor's PAUSE order as a stressor to investigate resiliency of food-related retail businesses in NYC. Specifically, the goal was to understand how urban planning considerations could facilitate small business resiliency.

1.2 Literature Review

COVID-19's impact on small businesses has been predominantly researched through surveys (Bartik 2020, Dai 2020, Hassan 2020), including the use of the Current Population Survey (CPS) to trace long-term trends in unemployment rates and determinants of business ownership (Fairlie and Fossen 2018, Wang 2019). Based on the CPS, the United States lost 3.3 million (or 22%) active business owners in April 2020 (Fairlie, 2020). In a survey of more than 5,800 small businesses, those able to operate remotely (e.g., in finance, professional services, and real estate-related businesses) experienced less disruption, while those in the construction, hospitability, transportation, and in-person personal services reported employment declines exceeding 50% (Bartik et al. 2020).

Footfall data from companies like Safegraph (Safegraph 2020) have been used to quantify America's response to stay-at-home orders and the effects on small businesses. SafeGraph data showed a dramatic increase in individuals staying at home during the first two months of the pandemic (Safegraph 2020). The impact of this decreased mobility varied significantly, even within seemingly similar businesses. For example, in Columbus, Ohio, Dollar stores and local convenience stores experienced smaller drops in business during the initial lockdown, but larger box stores and chain grocery stores rebounded more quickly as the economy reopened (Kar et al., 2021). Grocery stores in urban centers experienced more significant decreases in visits than non-urban grocery stores (Kar et al., 2021).

1.3 Problem Statement

This capstone sought to identify factors affecting small, food-related retail business survivability with respect to socio-economic factors, retail characteristics, built environment features, and changes in nearby foot traffic. Second, the capstone considered how hyper-local patterns related to the presence of significant institutions could be used as further predictors of business survivability.

2. Data

2.1 Data Source

The research employed datasets based on (1) anonymized cell phone detection and (2) first-hand observations. The former was obtained from SafeGraph through their January 2019 – May 2021 Core Places (business listings) and Patterns (aggregated cell phone detection). Those data contain monthly geospatial records of point of interest (POI) information such as location name, address, category, and the place visits and demographic aggregations from anonymous mobile devices. These data were used to identify closed businesses and analyze changes in visit levels over time. The secondary dataset was based on firsthand observers recording the paths of individuals leaving select NYC healthcare facilities.

2.2 Data Ethics and Privacy

Safegraph obtains data by partnering with third party mobile applications that collect location information about their users and then aggregates this information. Recently, Safegraph's data collection methods have come under scrutiny in the media and by Google (Valentino-Devries, 2018; Robertson, 2021). Independent analysis has shown the ability to use Safegraph's data to track specific individuals from place to place (Valentino-Devries, 2018), which could be misused by governments or individual bad actors. These potential risks and subsequent negative effects on user privacy led Google to ban Safegraph's location collection practices in August 2021. This capstone did not employ the Safegraph datasets in this way. Thus, no specific ethical concerns were raised through their use.

3. Methodology

Figure 1 provides a visual overview of this capstone's methodology summarizing the data acquisition and integration processes.

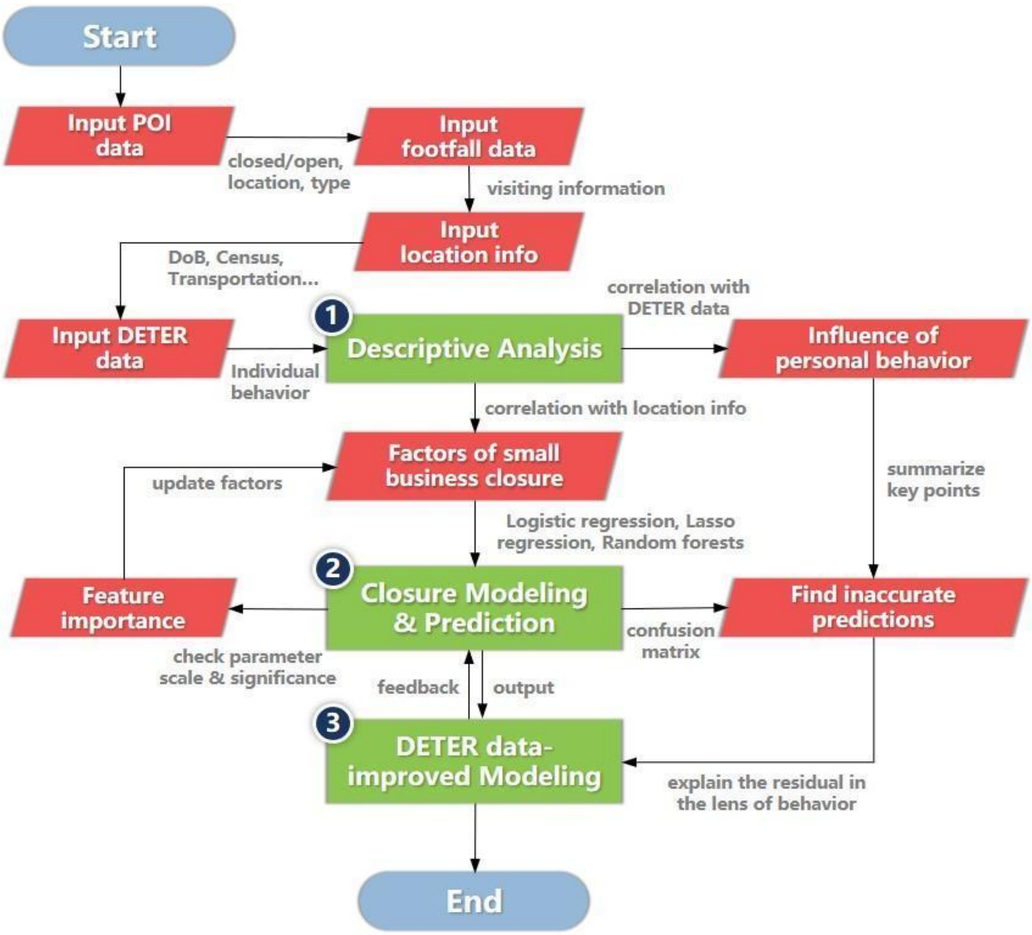

Figure1. Process of data preparation and analysis

3.1 Subsetting and Auditing Safegraph Data

The study was limited to food-related retailers in eight zip codes. As food-related businesses were permitted to operate throughout the PAUSE order, they were selected as the focus of this study. The eight sites were based on the availability of the DETER data, which provided a certain set of similarities, as the eight sites had two locales in NYC's four most populous boroughs with a hospital and an urgent care site within each borough (Figure 2). Using the Safegraph business subcategories, all businesses that were labeled convenience stores, full-service restaurant, limited-service restaurant, snack and non-alcoholic beverage bar, or supermarkets or other grocery store were initially included for all eight zip codes. This was then reduced to "small businesses" which then excluded chain stores or well-known brands.

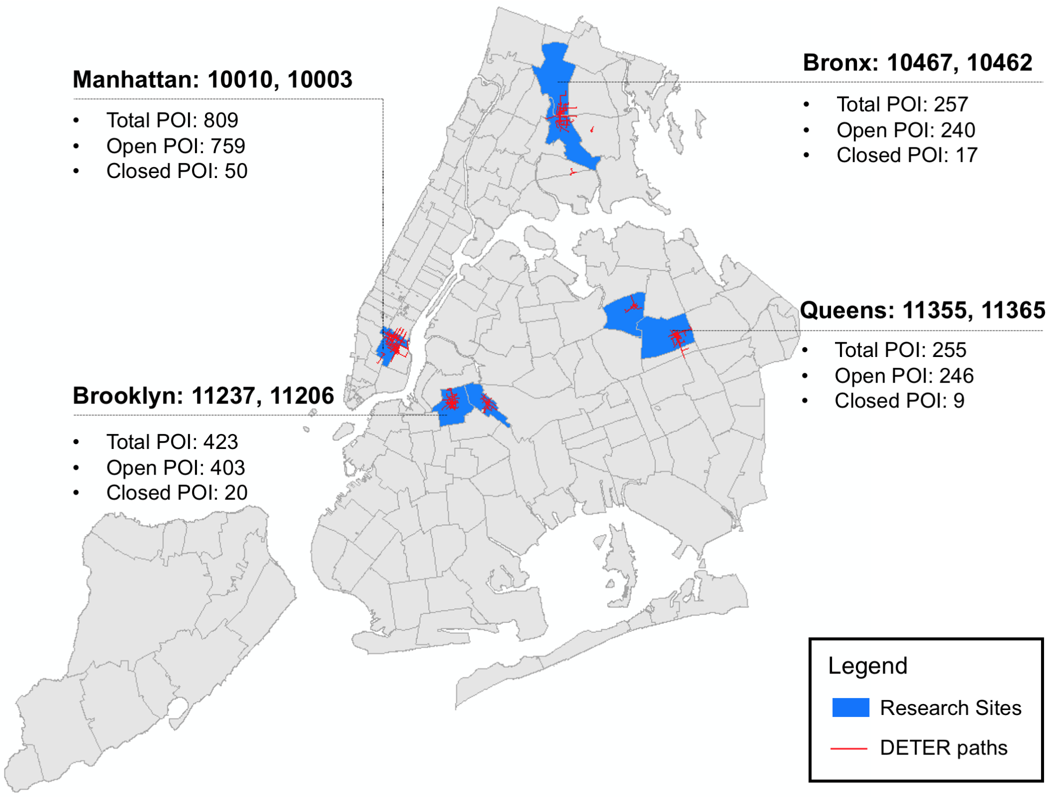

Figure2. Eight Study Sites (two in all NYC boroughs except Staten Island)

Closed small businesses were identified according to their presence in the "closed_on" column in the Safegraph data. A total of 93 were found to have closed across the 8 zip codes in the period of March 2020- May 2021. To check the reliability of that data, we used internet searches to verify that Safegraph accurately labeled closed businesses. The Safegraph labeling was found to be 94% accurate (95% confidence interval: 89-99%).

Finally, our group limited business closures to closures occurring in February 2020 (based on Safegraph documentation indicating closure date had a margin of error of two months) to focus our analysis exclusively on businesses that were open at the onset of the Covid-19 pandemic. Businesses that closed earlier than this were excluded, but were used as part of the audit to check the validity of Safegraph's closure identification.

3.2 Modeling and Prediction

Three easily interpretable methods were used to create a binary classification of open or closed, and to cross-check the validity of the most predictive features. First, a logistic regression method was used. This required the selection of a probability threshold maximizing the area under a receiver operating characteristic curve to generate binary predictions from the likelihood estimates produced by our logistic regression. Second, a ridge regression method was employed, which also involved selection of a probability threshold using the same method as the logistic regression. Finally, a Random Forest classifier was adopted.

In addition to footfall data from Safegraph, predictive variables about area income level, unemployment, physical characteristics, and commercial diversity were drawn from the American Community Survey and NYC Open Data (see **Appendix** IA for a complete list of variables)

3.3 Integration of DETER data

After using regression analysis and machine learning to describe which features best predicted business' risk of closure, the pathway selection from the DETER data was considered to explore how the impact of a certain kind of business (i.e. healthcare) could impact the survivability of other types of business, and whether the scale of that other business was influential. For this, the paths at the DETER study locations were overlaid on the business open/closure data.

4. Results

4.1 Descriptive Analysis

4.1.1 POI Closure Status

A total of 2,186 food-related small businesses were identified in the 8 zip codes. Of these 93 were listed as permanently closed between March 2020 and May 2021, and 570 small businesses were permanently closed before February 2020. To make an accurate prediction, we excluded the 570 small businesses closed before the COVID-19 PAUSE order in our modeling. The five business categories showed slight closure differences, with convenience stores having the highest closing rate -- around 6.9%; Limited service restaurants have a relatively lower closing rate -- approximately 4.8% (see **Appendix** B for closing rate figures). To understand the operation situation of POI in each zip code, we calculated the closing rates respectively and compared them with the social vulnerability index (SVI). Social vulnerability refers to the potential adverse effects on communities caused by external stresses on human health (CDC, 2015). The higher the SVI score, the more social vulnerability in that area, meaning that the area may need more resources to thrive. From Figure3, we can find that the closing rates in the zip codes of Bronx, Brooklyn, and Queens are consistent with the SVI scores. In contrast, the zip codes of Manhattan have a pretty high closing rate, indicating that Manhattan stores have experienced a more considerable impact by COVID-19.

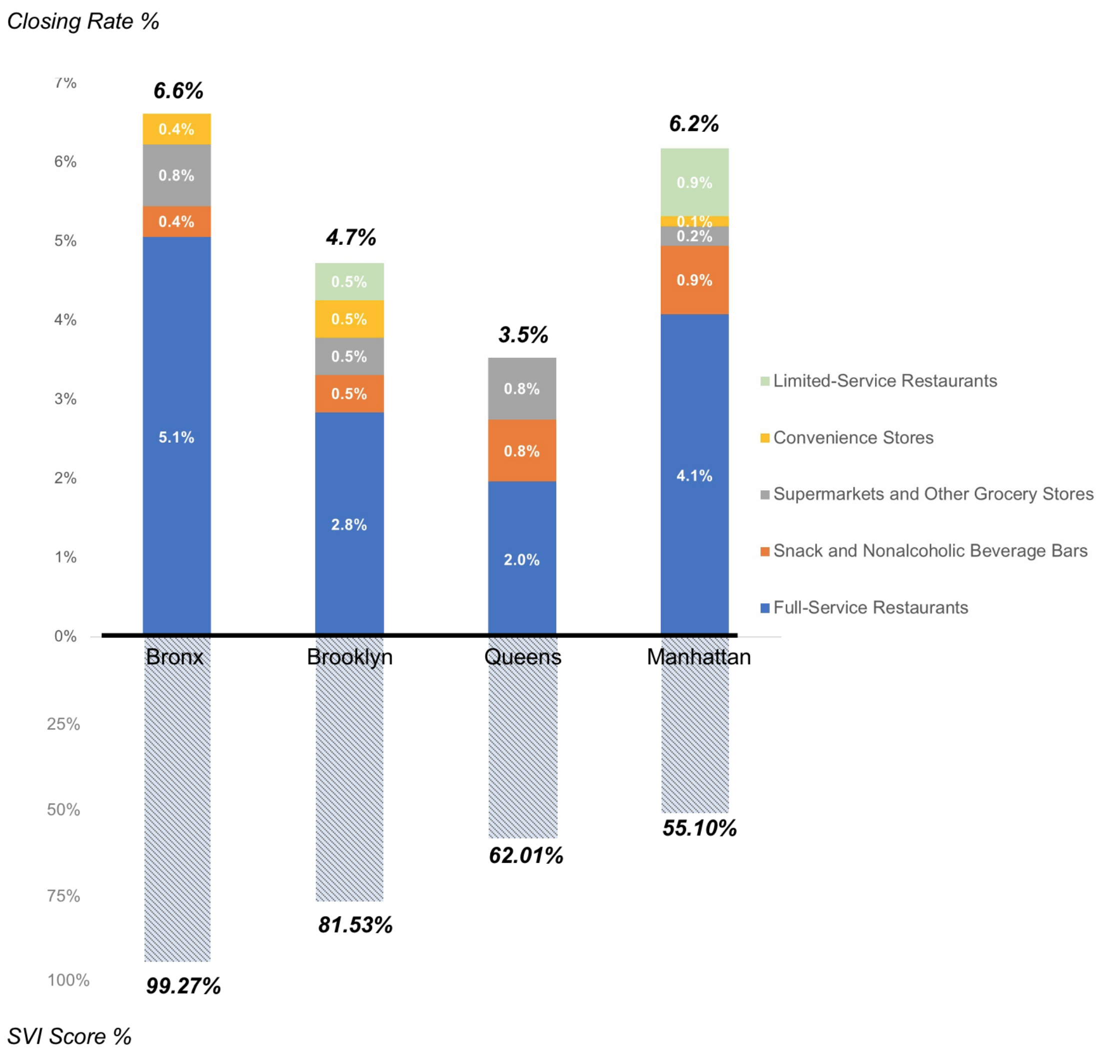

Figure3. Closing rate by zip code & SVI score

4.1.2 Comparison of Closed and Open POI

Besides the small business category, we further compared the spatial features of open and closed POI based on 12 variables from multiple sources of built environment data. From Figure 6, we can find that changing visits and distance to the subway are the main differences between open and closed POI: POI with a more significant decrease of visits and closer to the subway is more likely to close due to COVID-19. The results make sense since the drop in visits is one of the most critical factors causing business closures, and COVID-19 decreased crowd flow, especially around transit stations. As for other factors, it is hard to differentiate open and closed POI using these average values, indicating that a detailed case study might be necessary at a smaller scale.

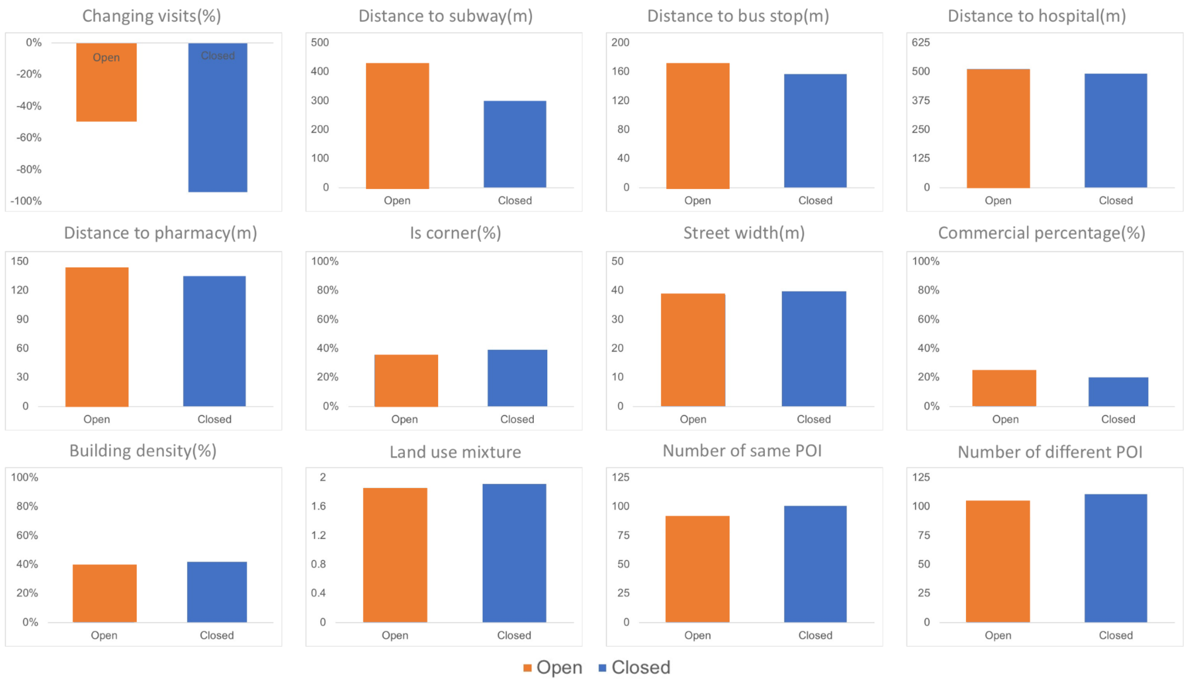

Figure4. Spatial feature of closed and open POI

4.1.3 POI Distribution and Individual Path Heat map

As shown in Figure 5, we plotted open and closed POI on maps to reveal detailed spatial information. Moreover, we visualized the individual paths based on DETER data to explore the relationship between small business closures and individual behaviors. Road segments with wider orange tracks indicate a more significant number of individuals passed through during the observation. Though the DETER paths failed to cover our whole research area, primarily due to the high collection cost of DETER data, it's intuitively believed that characteristics of DETER paths can provide a new perspective under the lens of individual behaviors, which contributes to a deeper understanding of business closure.

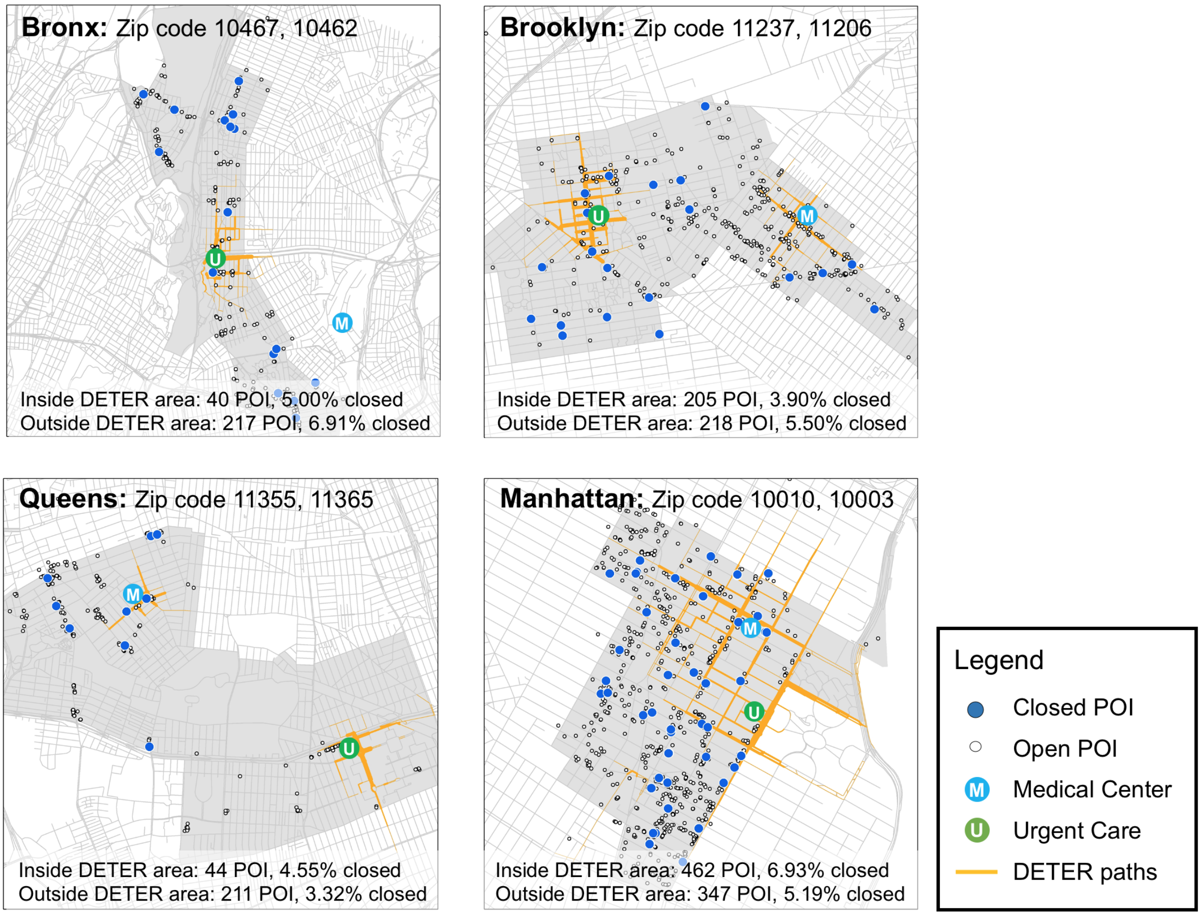

Figure5. Distribution of POI and individual path heat map

4.2 Predictive results

4.2.1 Logistic Regression

In the logistic model, only four features were statistically significant. The decrease in a business' visits had the most considerable impact on a business's likelihood of survival. The commercial percentage of a neighborhood also correlated with a slightly decreased risk of closure, as did a business' proximity to a subway station, and the number of different types of businesses within 500 meters of the business. Convenience stores were used as the reference category for the regression. A complete graph of the logistic regression's coefficients can be seen below. Using this model, we created binary predicted classifications for each POI for whether it would close or remain open and were able to achieve an 86%

accuracy rate (see **Appendix** C, **Table C.1** for a confusion matrix of logistic regression result).

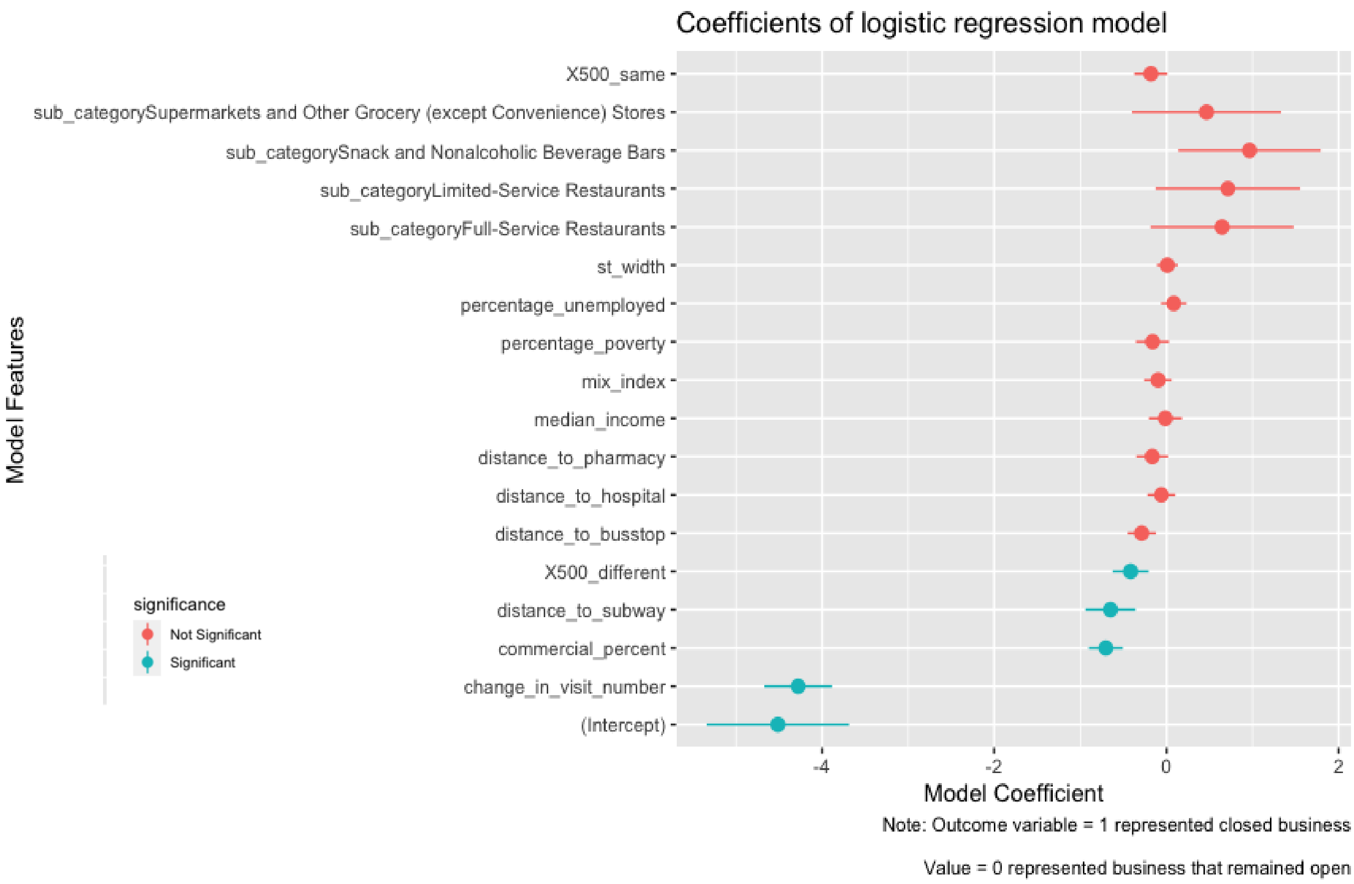

Figure6. Logistic regression modeling result

4.2.2 Lasso Model

In addition to logistic regression to create a binary classification system, we used penalized regression to restrict the number of features in our final model, while still maximizing the accuracy of our model. To do this, we used ten fold cross-fold validation across our training data, and fit a model by selecting the model within one standard error of the lambda that minimized error. This model also showed that the number of visits recorded in safegraph was the most crucial feature in the model. A summary of our coefficients can be seen below. The model performed roughly as well as the logistic regression, with similar accuracy levels on the test data set (see **Appendix** C, **Table C.2** for a confusion matrix of lasso regression results).

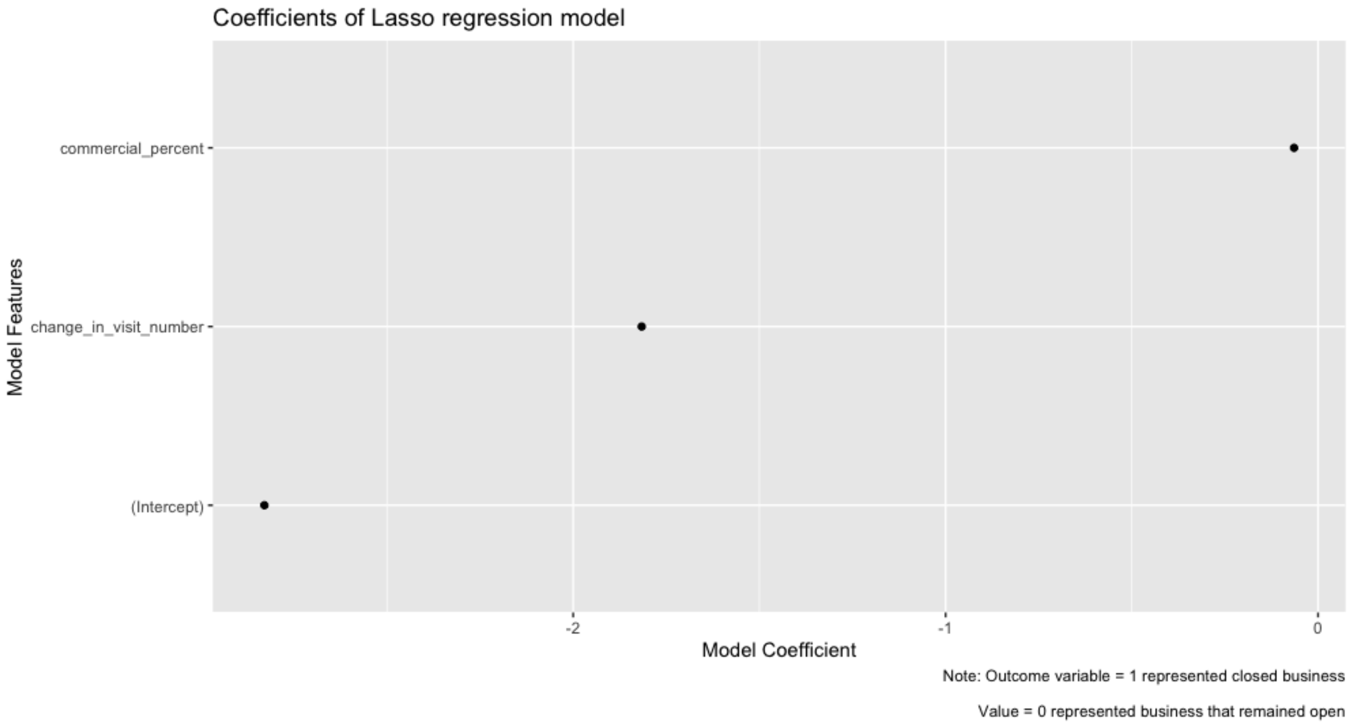

Figure7. Lasso regression modeling result

4.2.3 Random Forest

We used a random forest model to create binary classification categories with an R2 score of 0.14. We find the best model for RF is min_samples_leaf=13 and min_samples _split=3. In the RF model we developed, there were only two most important features--Average monthly visits after (0.40) and Average monthly visits before (0.18). The model performed roughly as well as the logistic regression, with similar accuracy levels on the test data set (see **Appendix** C, **Table C.3** for a confusion matrix of random forest results).

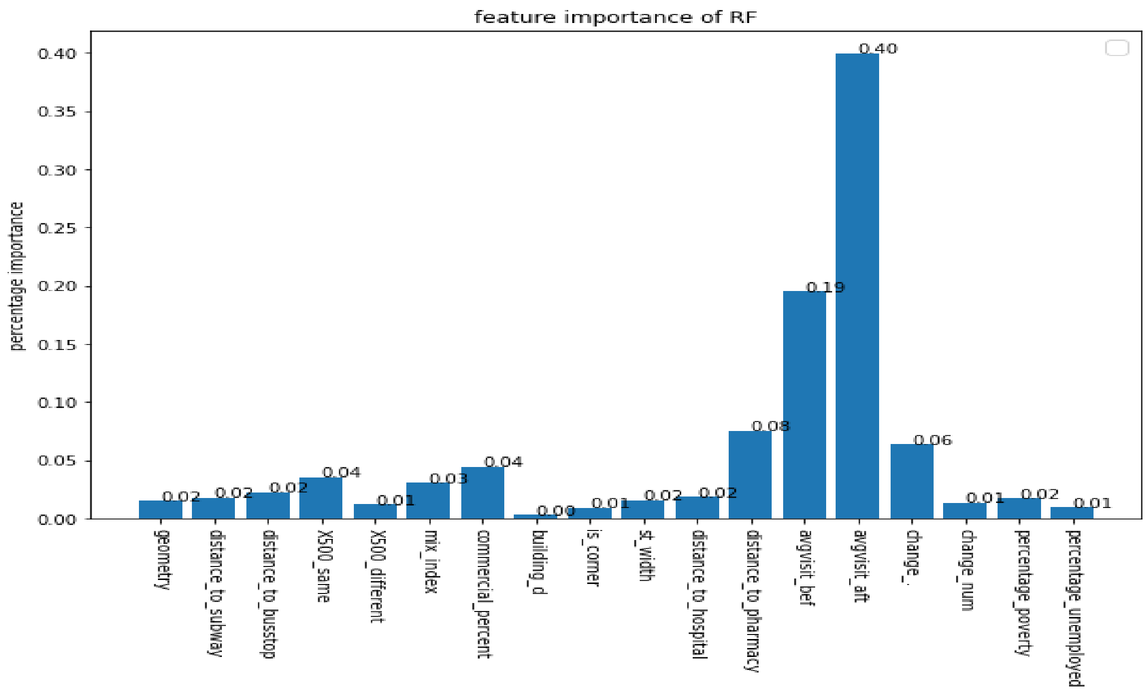

Figure8. Random forest modeling results by variable

4.3 DETER Analysis

The map above shows exit patterns at Mt. Sinai Hospital in Manhattan. At this location, there is a clear pattern of individuals leaving the hospital to head directly to the subway station at the adjacent corner. Businesses outside of this path, the opposite direction from the hospital, were much more likely to close, as were businesses past the subway stop. These patterns show the value of hyper-local, path-based analyses. Two businesses may be in similar socio-economic neighborhoods and may even be equidistant from public transportation. However, their specific location matters, as the difference between a consumer being willing to cross a street to go to a business may be the difference between a business surviving and folding during a crisis.

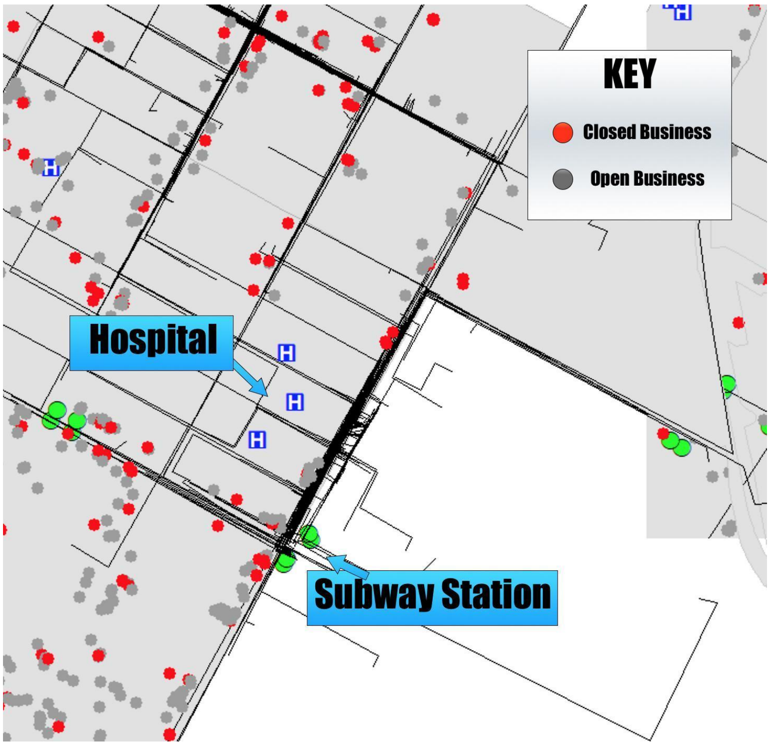

Figure9. An example of DETER path (Mt. Sinai Hospital in Manhattan)

5. Conclusions

Our group's findings suggest that foot traffic, combined with a complex combination of urban features promotes business resiliency. A change in foot traffic likely reflects a loss in revenue for a business. For the foodservice restaurants our study considered, this could include grocery delivery, delivery food, or a decision to forego eating out altogether.

Beyond foot traffic, we found that commercial proximity and subway proximity both were associated with resilient businesses. The commercial density would be a proxy for how many different types of businesses are available in an area. At the onset of the pandemic, public health guidance suggested that individuals limit their trips out of the house to essential trips. For shopping trips, this would encourage people to go to areas where they could accomplish multiple goals in one trip, rather than making several trips. Because of this, businesses located near other types of businesses were better positioned to attract customers during the public health crisis. The impact of transit proximity, while small in our regression analysis, may be more difficult to measure. Our analysis of DETER paths showed that businesses between major employers and subway locations appeared less likely to close than businesses that were equally close to the subway, but where an individual would need to travel past the transit stop to get to the business. This microdata could be useful to urban planners as they consider what business locations are best for small businesses.

Our analysis included limitations. Our data is limited to 8 zip codes and data limitations restrict our SafeGraph POI data to five store categories. These limitations restrict the generalizability of our findings. Given the complexity of New York City, and the variation between neighborhoods, the findings in our report may not translate to other types of POIs and other neighborhoods. Future research could consider similar questions of whether microdata about small business location and pedestrian activity could help explain business closure patterns.

.

Team Roles and Contributions

- **Ari Lewenstein** : SafeGraph data acquisition and cleaning, regression modeling, socio-economic variable construction, DETER data analysis
- **Shu Wang** : SafeGraph data, DETER data, and other geodata cleaning and processing, descriptive graphing and analysis, spatial variable construction
- **Haochen Xing** : Random Forest

Reference

1. Adcroft, P., Toor F., (2021). Timeline: How Covid-19 Changed NYC. NY1
2. Bartik, A. W., M. Bertrand, Z. Cullen, E. L. Glaeser, M. Luca and C. Stanton (2020). "The impact of COVID-19 on small business outcomes and expectations." Proceedings of the National Academy of Sciences **117**(30): 17656-17666.
3. Bartik, A. W., M. Bertrand, Z. B. Cullen, E. L. Glaeser, M. Luca and C. T. Stanton (2020). How are small businesses adjusting to COVID-19? Early evidence from a survey, National Bureau of Economic Research.
4. Braun, M. (2021). NYC Got Billions in Federal Aid Dollars. Now One Man Wants to Know Where it Went. Bloomberg https://www.bloomberg.com/news/articles/2021-10-19/nyc-isn-t-keeping-track-of-aid-dollars-says-comptroller-nominee
5. Centers for Disease Control and Prevention (CDC). Planning for an Emergency: Strategies for Identifying and Engaging At-Risk Groups. A guidance document for Emergency Managers: First edition. Atlanta (GA): CDC; 2015.
6. Dai, R., H. Feng, J. Hu, Q. Jin, H. Li, W. Ranran, R. Wang, L. Xu and X. Zhang (2020). "The impact of covid-19 on small and medium-sized enterprises: evidence from two-wave phone surveys in china." Center for Global Development Working Paper **549**.
7. Fairlie, R. (2020). "The impact of COVID-19 on small business owners: Evidence from the first three months after widespread social-distancing restrictions." Journal of economics & management strategy **29**(4): 727-740.
8. Fairlie, R. W. and F. M. Fossen (2018). "Opportunity versus necessity entrepreneurship: Two components of business creation."
9. Fonrouge, G. Hicks, N, Fenton, R. (2021). New York City Businesses are Barely Holding On. New York Post
10. Kar, A., Motoyama, Y., Carrel, A., Miller, H., and Le, H. (2021). "Covid-19 exacerbates unequal food access" Applied Geography **134**
11. Hassan, T. A., S. Hollander, L. Van Lent, M. Schwedeler and A. Tahoun (2020). Firm-level exposure to epidemic diseases: Covid-19, SARS, and H1N1, National Bureau of Economic Research.
12. NYC Small Business First (2015) https://www1.nyc.gov/assets/smallbizfirst/downloads/pdf/small-business-first-report.pdf.
13. Robertson A. (2021) Google Bans Tracking Tool That Sold Users' Location Data. The Verge
14. SafeGraph (2020). "Foot Traffic Patterns by State and Industry." https://www.safegraph.com/data-examples/reopening-the-economy-foot-traffic?s=NY&d=11-22-2020&i=all
15. Senate Committee on Small Business & Entrepreneurship (2020). "Senate Committee on Small Business & Entrepreneurship guide to the CARES Act." https://www.sbc.senate.gov/public/_cache/files/2/9/29fc1ae7-879a-4de0-97d5-ab0a0cb558c8/1BC9E5AB74965E686FC6EBC019EC358F.the-small-business-owner-s-guide-to-the-cares-act-final-.pdf
16. U.S. Small Business Administration (2018). "NY small business profiles 2018." https://www.sba.gov/sites/default/files/advocacy/2018-Small-Business-Profiles-NY.pdf
17. Valentino-DeVries J., Singer N., Keller M., Krolik A., (2018), Your Apps Know Where You Were Last Night, And They're Not Keeping It Secret. The New York Times
18. Wang, C. (2019). "Tightened immigration policies and the self-employment dynamics of Mexican immigrants." Journal of Policy Analysis and Management **38**(4): 944-977.
19. Zhang, F., Wang, Y., Fei, Y., & Chen, X. Las Vegas Business Closure Prediction Model.

Appendix A - Variable Description

Table A.1 Variable Source and Description

Variable	Description	Data source
distance_to_subway (meters)	Distance between each POI and its nearest subway station	Subway Entrances https://data.cityofnewyork.us/Transportation/Subway-Entrances/drex-xx56
distance_to_busstop (meters)	Distance between each POI and its nearest bus stop	Bus Stop Shelters https://data.cityofnewyork.us/Transportation/Bus-Stop-Shelters/qafz-7myz
500_same	Number of the same subcategory POIs within 500 meters (* 500m is around the diameter of each poi cluster)	SafeGraph Data *(5 sub-categories: Convenience Stores, Full-Service Restaurants, Limited-Service Restaurants, Snack and Nonalcoholic Beverage Bars, Supermarkets and Other Grocery (except Convenience) Stores)*
500_different	Number of different sub-category POIs within 500 meters (* 500m is around the diameter of each poi cluster)	SafeGraph Data *(5 sub-categories: Convenience Stores, Full-Service Restaurants, Limited-Service Restaurants, Snack and Nonalcoholic Beverage Bars, Supermarkets and Other Grocery (except Convenience) Stores)*
mix_index	Land use mixture of the census tract each POI locates, using Shannon's entropy formula is: H(x) = -Σni=1 [P(xi) * logbP(xi)]	PLUTO data https://www1.nyc.gov/site/planning/data-maps/open-data/dwn-pluto-mappluto.page
commercial_percent	Proportion of commercial land in the census tract each POI locates, using commercial area/total area	PLUTO data https://www1.nyc.gov/site/planning/data-maps/open-data/dwn-pluto-mappluto.page
building_d	Building density of the census tract each POI locates, using building gross area/total area	Building Footprints https://data.cityofnewyork.us/Housing-Development/Building-Footprints/nqwf-w8eh
Is_corner (Boolean)	Whether each POI is located in the corner or not	NYC Street Centerline (CSCL) https://data.cityofnewyork.us/City-Government/NYC-Street-Centerline-CSCL-/exjm-f27b
st_width (meters)	The width of the street where each POI locates	NYC Street Centerline (CSCL) https://data.cityofnewyork.us/City-Government/NYC-Street-Centerline-CSCL-/exjm-f27b
distance_to_hospital (meters)	Distance between each POI and its nearest hospital	SafeGraph Data *(sub_category: General Medical and Surgical Hospitals)*
distance_to_pharmacy (meters)	SafeGraph Data (sub_category: General Medical and Surgical Hospitals)	SafeGraph Data *(sub_category: Pharmacies and Drug Stores)*
Change_in_foot_traffic	(avgvisit_aft - avgvisit_bef) / avgvisit_bef	SafeGraph Data
sub_category	Description of the type of business	Safegraph data
percentage_poverty	Percentage of population under poverty level income in census tract	2019 American Community Survey, variables B05010_002 (population under poverty level) and B01003_001 (total population)
percentage_unemployed	Percentage of population unemployed in census tract	2019 American Community Survey, variables C18120_006 (unemployed population) and C18120_002 (total labor force population)
median_income	Median income in census block	2019 American Community Survey, variable B06011_001

We used Python and ArcGIS to compute these variables, including 11 variables from multi-source spatial data, 2 variables from Safegraph data, and 3 variables from community survey data. Of these, distance to subway, distance to bus stop, distance_to_hospital, and distance_to_pharmacy were calculated as the shortest straight line connecting each POI and its nearest facilities (subway, bus stop, hospital, and pharmacy). 500_same and 500_different were calculated by counting the number of POI of the same and different subcategories within 500 meters of each POI. Mix_index refers to land use mixture and was generated by using Shannon's entropy. The formula is: $H(x) = -\Sigma ni=1 [P(xi) * logbP(xi)]$. We also calculated the average visits to closed and open POI before and after March 2020, and computed the changing rate. For POI closed between March 2020 and May 2021, we only included the monthly visits when the stores were still open.

Appendix B. Descriptive FIgures

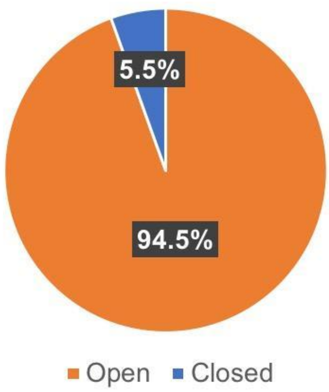

Figure B.1 Number of closed and open POI

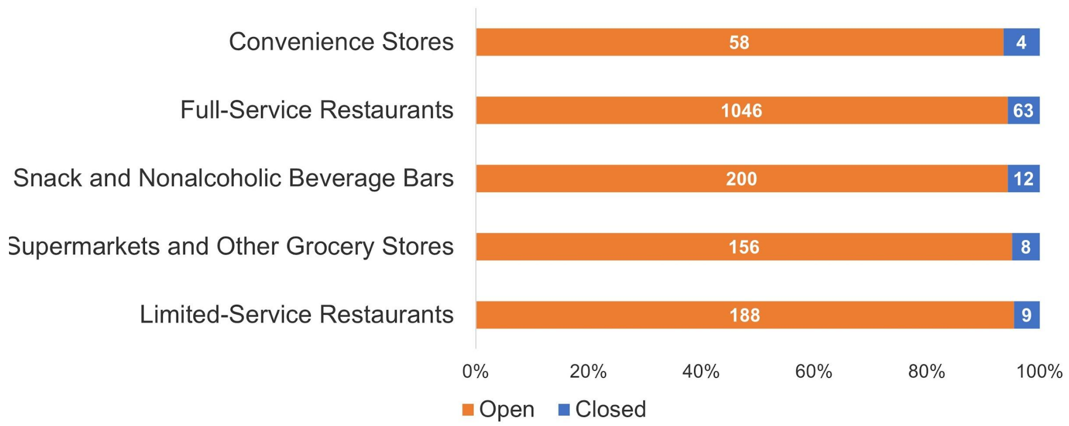

Figure B.2. Number of closed POI by category

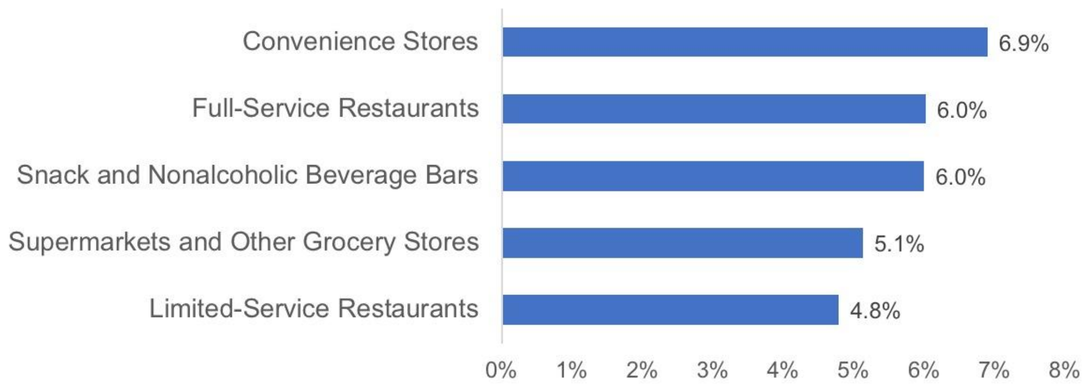

Figure B.3. Closing rate by category

Appendix C. Predictive Modeling Results

Table C.1. Confusion matrix of logistic regression result

	Remained Open	Closed	Predicted Totals
Predicted Open	1,311	15	1326 (82%)
Predicted Closed	212	78	290 (18%)
Actual Totals	1,523 (94%)	93 (6%)	1,616

Accuracy = 0.860, Sensitivity = 0.861, Specificity = 0.839, F1 = 0.407

Table C.2. Confusion matrix of lasso regression result

	Remained Open	Closed	Predicted Totals
Predicted Open	283	1	284 (88%)
Predicted Closed	27	12	39 (12%)
Actual Totals	310 (96%)	13 (4%)	323

Accuracy = 0.913, Sensitivity = 0.913, Specificity = 0.923, F1 = 0.461

Table C.3. Confusion matrix of random forest result

	Remained Open	Remained Closed	Predicted Totals
Predicted Open	1523	75	1598(99%)
Predicted Closed	0	18	18 (1%)
Actual Totals	1523 (94%)	93 (6%)	1,616

Accuracy = 0.954