Chris Westendorf
IBM Data Science
Capstone Project
August 11, 2020

COVID-19 NYC: FOOT TRAFFIC IMPACT

1. Introduction

1.1 Background

The global pandemic, Covid-19, has infected much of the world. As of the today, 17 million cases have been reported worldwide and 670,000 deaths have been reported. New York City is an economic capital of the world that hosts many transient people and goods. New York City was recently the global epicenter of the virus having confirmed 224,863 cases and 23,525 deaths. Through New York's diligent effort it was able to curtail the virus. However, since the virus has subdued the city has relaxed it's safety measures in phases and in response citizens have begun neglecting the safety protocols that helped them once curtail the virus's spread.

Understanding New York City's COVID-19 Phases

Shelter in Place Orders

1. Effective at 8pm on Sunday, March 22, all non-essential businesses statewide will be closed
2. Non-essential gatherings of individuals of any size for any reason (e.g. parties, celebrations or other social events) are canceled or postponed at this time
3. Any concentration of individuals outside their home must be limited to workers providing essential services and social distancing should be practiced
4. When in public individuals must practice social distancing of at least six feet from others
5. Businesses and entities that provide other essential services must implement rules that help facilitate social distancing of at least six feet
6. Individuals should limit outdoor recreational activities to non-contact and avoid activities where they come in close contact with other people
7. Individuals should limit use of public transportation to when absolutely necessary and should limit potential exposure by spacing out at least six feet from other riders
8. Sick individuals should not leave their home unless to receive medical care and only after a tele-health visit to determine if leaving the home is in the best interest of their health
9. Young people should also practice social distancing and avoid contact with vulnerable populations
10. Use precautionary sanitizer practices such as using isopropyl alcohol wipes

Businesses that qualified as "essential businesses" under the stay-at-home order included but were not limited to:

- utility companies
- banks
- pharmacies
- laundromats
- gas stations
- grocers, restaurants, and convenience stores
- liquor stores
- hardware stores
- auto repair shops
- delivery services
- skilled contractors like plumbers
- health care providers
- warehouses
- manufacturers
- construction companies
- animal-care providers

June 8, Phase One
Allowed construction, manufacturing, and wholesale supply-chain businesses to reopen, as well as many retailers for curbside pickup, in-store pickup, or drop-off. Phase-one retail categories included

clothing and shoes, electronics and appliances, web and mail order, florists, jewelry, luggage, and sporting goods, among others. Malls remained closed. (Many nonessential retail businesses throughout the state were already offering curbside pickup ahead of the reopening, but not all.)

Agriculture, forestry, fishing, and hunting businesses can also resume, as can landscaping and gardening businesses, drive-in movie theaters, and low-risk recreational activities like socially distant sports such as tennis.

June 22, Phase Two
Allowed a greater range of businesses to reopen, including: offices, outdoor dining, places of worship (at 25 percent capacity), and storefront retailers and businesses in the professional-services, finance and insurance, administrative support, and real-estate and rental-leasing industries.

Salons and barbershops were also allowed able reopen in phase two with limited capacity, as were car dealerships. For retail, malls will remain closed.

July 6, Phase Three
Focuses on the hospitality industry, allowing restaurants and other food-service businesses to reopen for dine-in service at 50 percent capacity. (In New York City, however, indoor dining will not be allowed in this phase.) Diners are required to be separated by at least six feet or by a barrier when that's not possible, and must wear masks until they sit down. Gatherings of as many as 25 people, up from 10, are also allowed under this phase.

July 19, Phase Four
The final phase, allows schools and low-risk arts, entertainment, and recreation businesses to reopen — all with social distancing required — but not indoor dining, movie theaters, shopping malls, or gyms. Gatherings of up to 50 people will also be allowed.

1.2 Problems to be resolved

1. Is there a way to inform New York City inhabitants of the risk factor in their foot traffic choices?
2. Can the foot traffic reports of New York City from before and after the City's lock down response provide a meaningful analysis and classification of the types of venue's role in the pandemic economics?

1.3 Audience

The goal of this analysis is to provide answers to these questions for New York City inhabitants. And since Dr.Anthony Stephen Fauci, the director of the National Institute of Allergy and Infectious Diseases, has suggested that New York's reaction to Covid-19 should serve as a guide to other cities and states this analysis should also help support phase decision making for other communities as well.

2. Data acquisition and cleaning

2.1 Data sources and data description

We will be using a combination of location data and foot traffic data from Foursquare's 13 million users.
New York City has a total of 5 boroughs and 306 neighborhoods. In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the the latitude and longitude coordinates of each neighborhood. This dataset exists here: https://geo.nyu.edu/catalog/nyu_2451_34572

Foursquare has been compiling and analyzing COVID-19 Foot Traffic Data since 02/19/2020 and it can be interfaced with through amazon web services. This data set contains indexed foot traffic to 19 categories of venues. The indexed data is broken out geographically, with included data for National, SF, NYC, LA, and Seattle. The data is normalized against U.S. Census data to remove age, gender and geographical bias. They use indexed foot traffic to demonstrate the relative decline in visits to different types of places, where visits on the first day are 100. They analyze data on a rolling 7-day basis to reduce the effects of foot traffic trends influenced by certain days of the week (for example, bars and clubs experience an uptick on Fridays and Saturdays).

So, to explain this in an example, an 81 index to airports in Seattle for March 6 indicates that foot traffic between February 29 - March 6 (a 7-day rolling period) is 19% lower than the first 7 days of analysis, February 13 - February 19.

They used February 13 - February 19 as the first 7-day period benchmark for analysis because February 19 is when they last estimated foot traffic to be roughly normal for the categories analyzed.

We will also be interfacing with the foursquare api to pull venue information for New York City. For reproduction purposes we will stay within the free developer account parameters:

- 99,500 Regular calls / day
- 500 Premium calls / day
- 2 Photos per venue
- 2 Tips per venue

These exploration calls will yield:

- Venue name
- A unique venue Id
- Venue category(s)
- Venue location attributes such as neighborhood, latitude, longitude, and postal address
- Venue popularity

3. Methodology

3.1 Data Cleaning

The Foursquare Covid-19 Dataset indexed foot traffic for 26 venue categories. Not all the columns declared New York City the same way, which required visual inspection and then creating a loop to check for different spellings so that we could extract the relevant columns. The data contained: 105 NaN values for Casual Dining Chains, 77 NaN values for Furniture Stores, 126 NaN values for Movie Theaters, 164 NaN values for Nail Salons. After reviewing the null values from Foursquare's Covid-19 Impact data the following choices were made:

- Drop Movie Theaters because they have been closed throughout Covid-19.
- Drop Nail Saloons for two reasons:(1)we only have national data and New York City has atypical foot traffic pattern for Nail Saloons (2) Nail Saloons have only recently opened with strict safety protocols.
- As of March 18, Casual Dining Chains stopped reporting their data. Averaging the index score for Casual Dining Chains for the proceeding 4 weeks, an observed normalization period of the foot traffic, afforded replacing the NaN values.
- Drop Auto Dealerships to avoid uniformed speculative remapping and thus garbage data effecting the analysis. This will have the most effect on Hell's Kitchen because that is where Manhattan's dealerships are concentrated, for the most part.
- Furniture Stores have missing data between May 2 and July 17, remapping would be a complete guess so this category was dropped as well

Issues with venue category IDs, not all the venue categories were attributed to one specific category ID, such as Casual Dining Chains. When asked, Foursquare declined to elaborate on its category ID attribution. These attribution issues caused the venue categories to be reduced to 19 when we were associating the index scores to the nearby venues in each neighborhood. The second attribution issue came from querying the API. Even when specific category IDs are queried, the results reference child category IDs as the primary ID instead of the parent ID. This caused 139 unique category IDs to be pulled into the dataset, which required the further reduction of the dataset to the attributable venue categories.

The next step was extrapolating the venue data so that each venue has an index score each day. This required instantiating an ordered dictionary for filling by a nested loop that iterates through each row of the Foursquare Venue data frame and the Covid-19 Impact data frame. This data frame is then exported to an Excel document for replication and validation purposes.

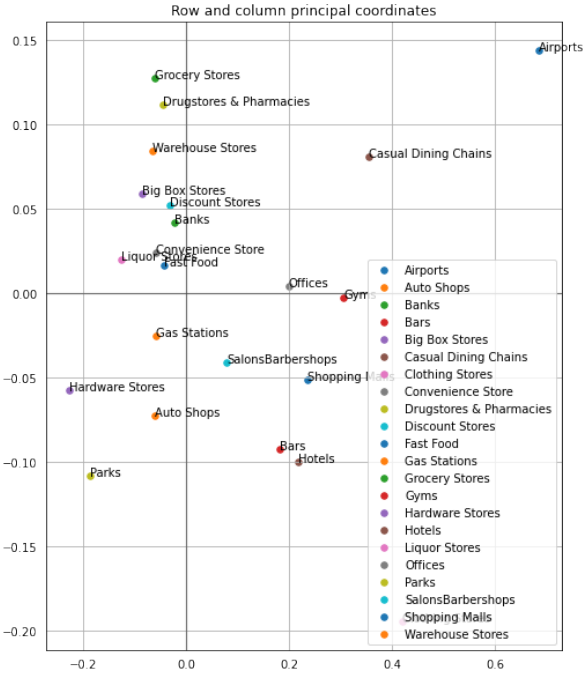
4. Exploratory Data Analysis (Results & Discussion)

4.1 Multiple correspondence analysis & Principal Component Analysis

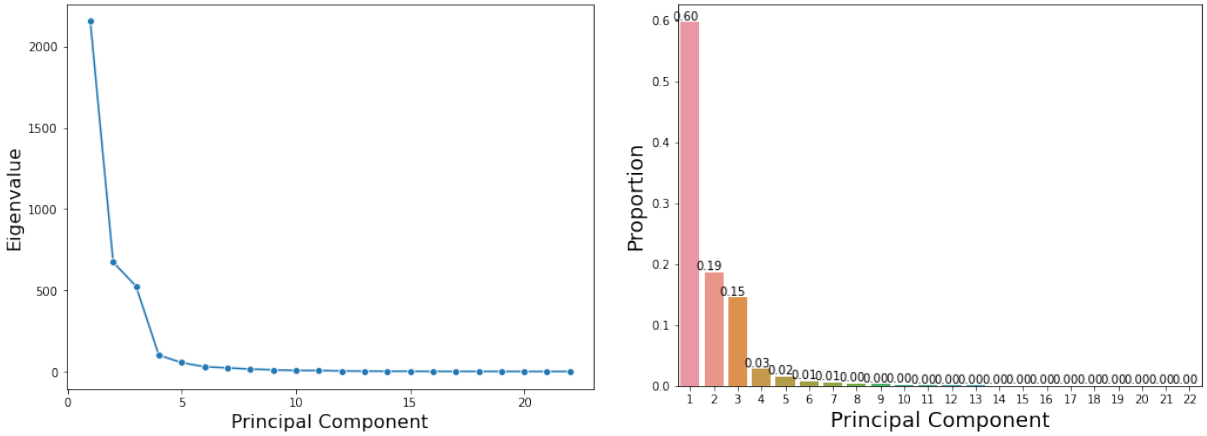
Since the dataset involves nominal categorical data, simply running unsupervised clustering would be uninformative since there isn't an inherent underlying data structure to categorical data. Luckily in statistics, multiple correspondence analysis (MCA) is a data analysis technique for nominal

categorical data, used to detect and represent underlying structures in a data set. It does this by representing data as points in a low-dimensional Euclidean space. The procedure thus appears to be the counterpart of principal component analysis for categorical data. MCA can be viewed as an extension of simple correspondence analysis (CA) in that it is applicable to a large set of categorical variables. Ultimately, these analyses yield principal components that can provide euclidean values for classifying neighborhoods and venue categories. These principal components can then be used in algorithms like k-means and k-mode to cluster and classify areas.

Running the MCA analysis yielded the following principal component relationship:



Running the MCA analysis created an unverifiable “blackbox” data relationship because the data behind the index scores is not accessible. Next, PCA was conducted in a further attempt a discovery of an underlying data relationship and inertia.

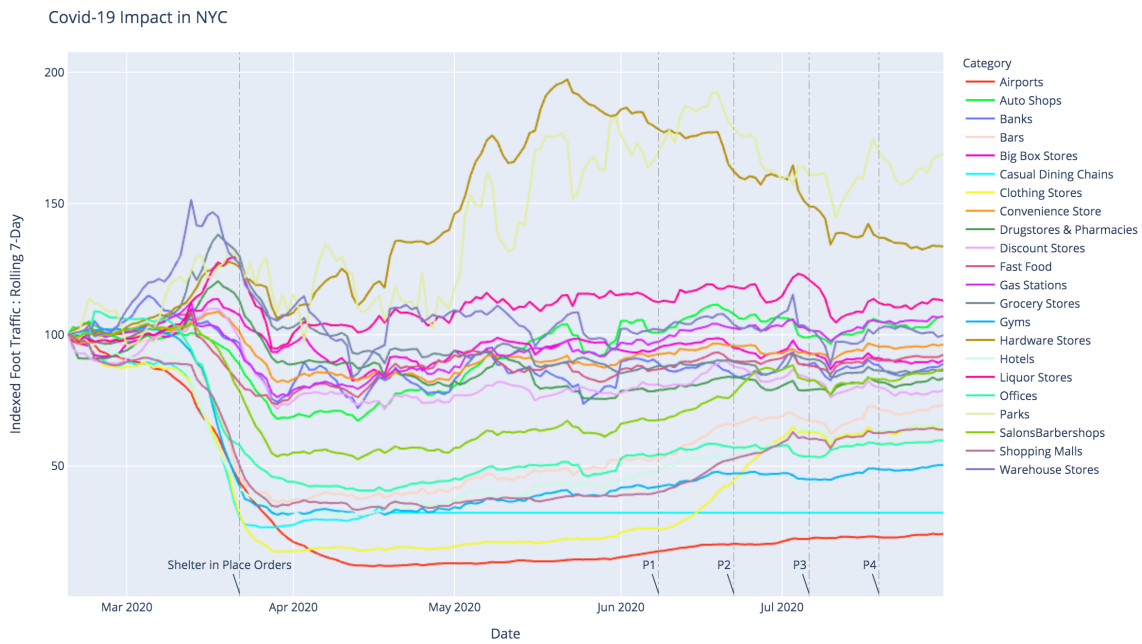


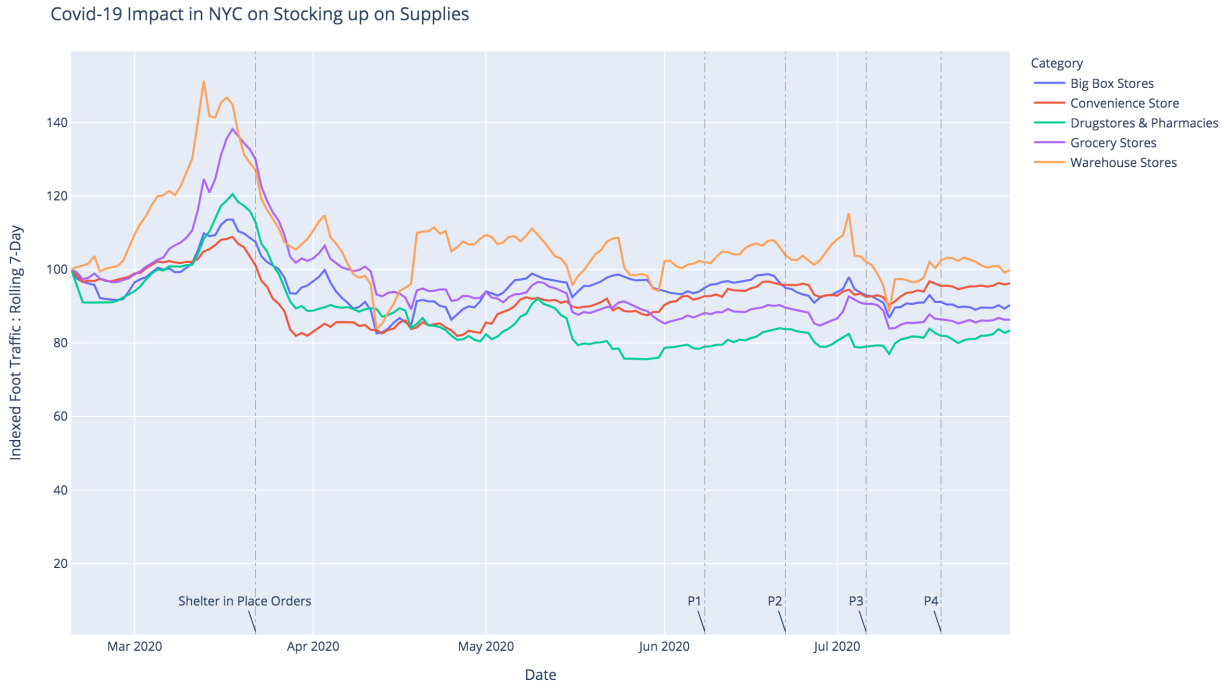
	0	1	2	3	4	5
Airports	0.933801	0.248285	-0.166353	-0.065421	-0.160093	0.007105
Auto Shops	0.500915	-0.836825	0.137811	-0.034304	-0.077331	0.021643
Banks	0.921256	0.064132	0.081587	-0.163625	0.082555	-0.230113
Bars	0.939284	-0.252178	-0.210800	0.059476	-0.013496	0.044757
Big Box Stores	0.516261	0.187953	0.763855	-0.205306	-0.156313	0.167666
Casual Dining Chains	0.903558	0.195320	-0.260625	-0.231316	-0.084374	-0.032048
Clothing Stores	0.904500	-0.199240	-0.263190	0.236148	0.039872	0.071296
Convenience Store	0.870865	-0.228093	0.377065	0.139889	-0.045047	-0.079737
Discount Stores	0.867427	0.123965	0.434484	0.014798	-0.071727	-0.045696
Drugstores & Pharmacies	0.602407	0.627820	0.428291	0.134697	-0.041150	-0.122923
Fast Food	0.849620	-0.242969	-0.135809	-0.283899	0.309734	-0.059387
Gas Stations	0.675519	-0.694718	0.140322	0.114601	0.131280	0.021701
Grocery Stores	0.444656	0.681185	0.558301	0.101701	-0.069289	0.004432
Gyms	0.960606	0.033777	-0.218573	-0.128000	-0.078074	0.000320
Hardware Stores	-0.468835	-0.682373	0.372185	-0.386195	-0.101909	-0.010786
Hotels	0.923037	-0.225562	-0.291330	0.031804	-0.039412	0.045118
Liquor Stores	-0.116906	-0.484388	0.803848	0.176262	0.055908	-0.129360
Offices	0.958483	0.014064	-0.214447	-0.130048	-0.093920	-0.009654
Parks	-0.286368	-0.868478	0.304234	0.046215	-0.126177	0.020926
SalonsBarbershops	0.947095	-0.238984	-0.036573	0.180099	0.014420	0.042811
Shopping Malls	0.961979	-0.076855	-0.195383	0.145761	-0.003048	0.057239
Warehouse Stores	0.553397	0.352093	0.650574	-0.112836	0.296768	0.189459

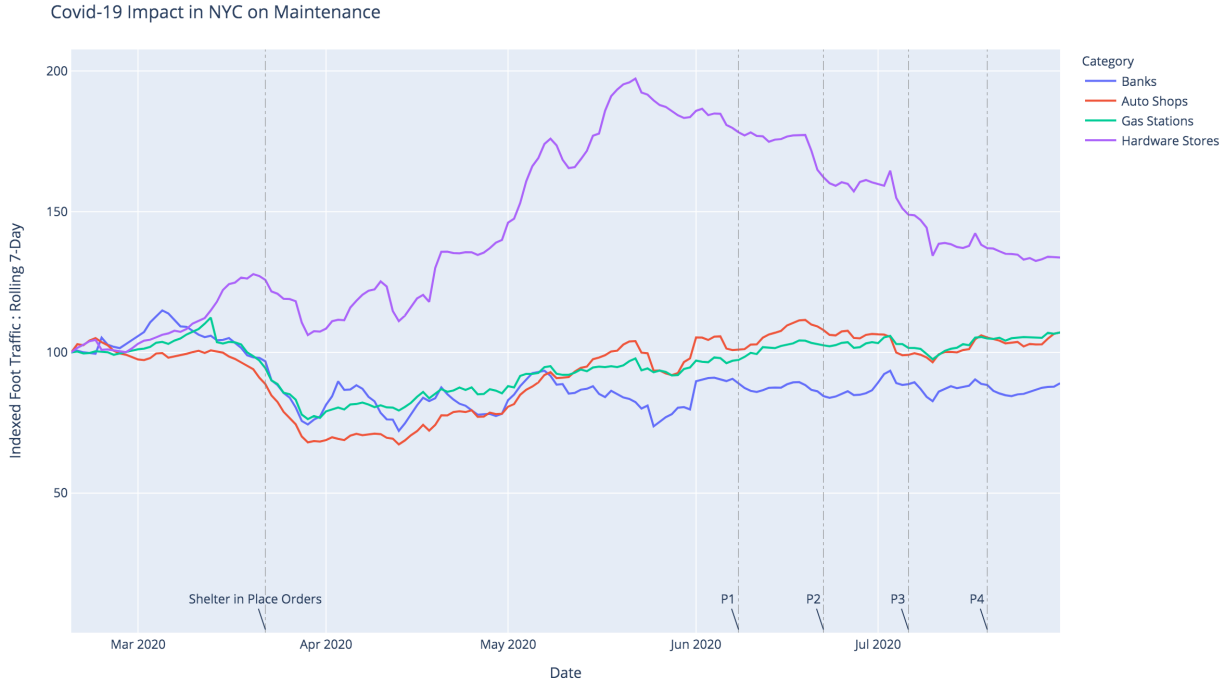
While these exploratory analyses yielded results they weren't verifiable given the data set. Therefore, running predictive algorithms based on these results wouldn't guarantee the answers to our problems.

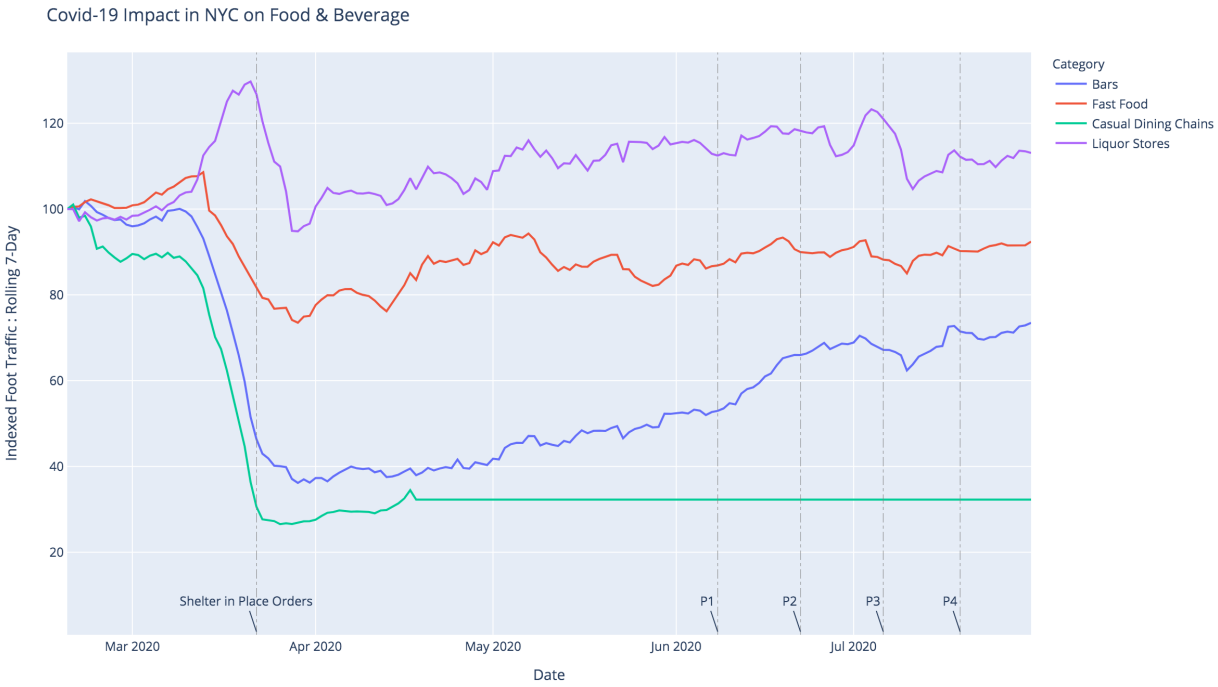
3.2 Visual Plotting and Charting

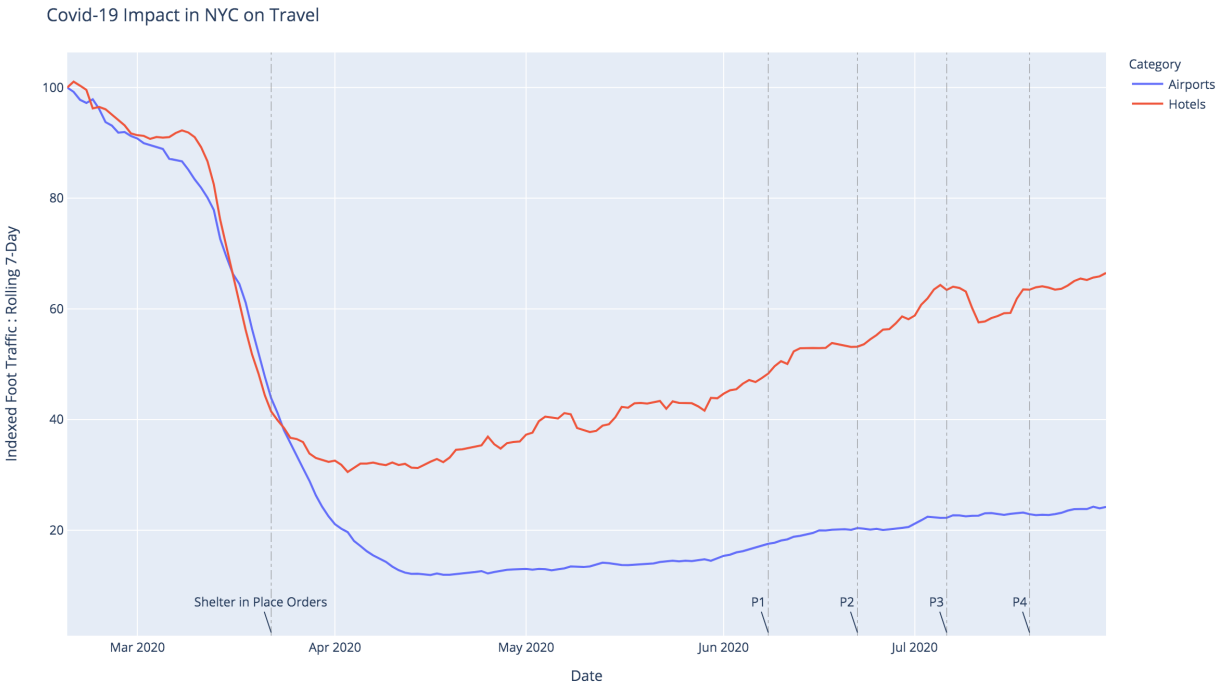
Since the data does not lend itself to algorithmic classification, the next step was to produce charts for observational analysis of foot traffic. This method will allow us to create observable hypotheses for the 2021 commercial landscape and to qualify the size of the impact of Covid-19 on the urban fabric of New York City.

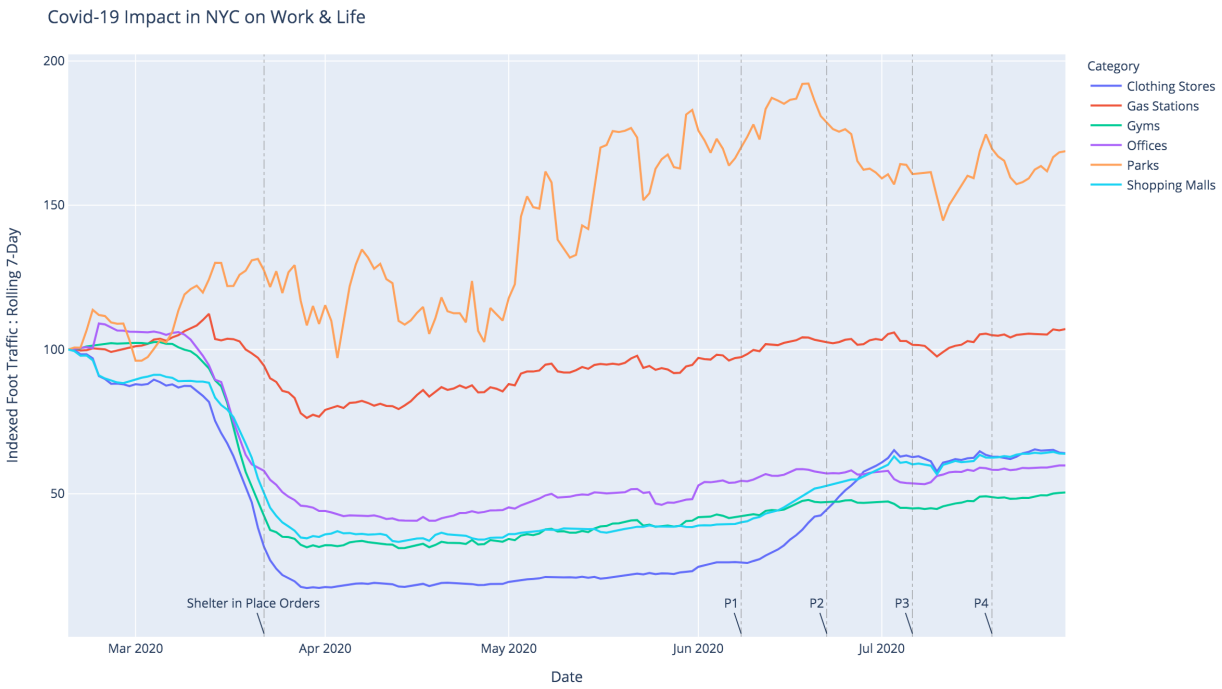


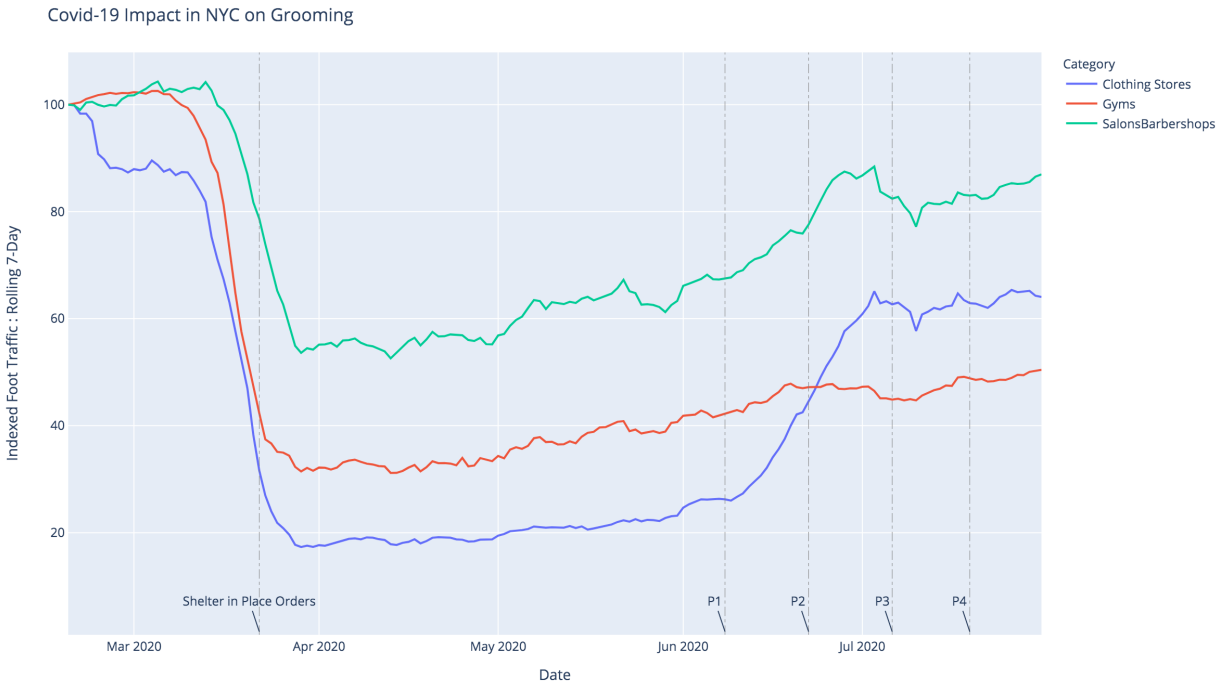

Covid-19 Impact in NYC on Stocking up on Supplies
Indexed Foot Traffic : Rolling 7-Day
140
120
100
80
60
40
20
Shelter in Place Orders
P1
P2
P3
P4
Mar 2020
Apr 2020
May 2020
Jun 2020
Jul 2020
Date
Category
Big Box Stores
Convenience Store
Drugstores & Pharmacies
Grocery Stores
Warehouse Stores

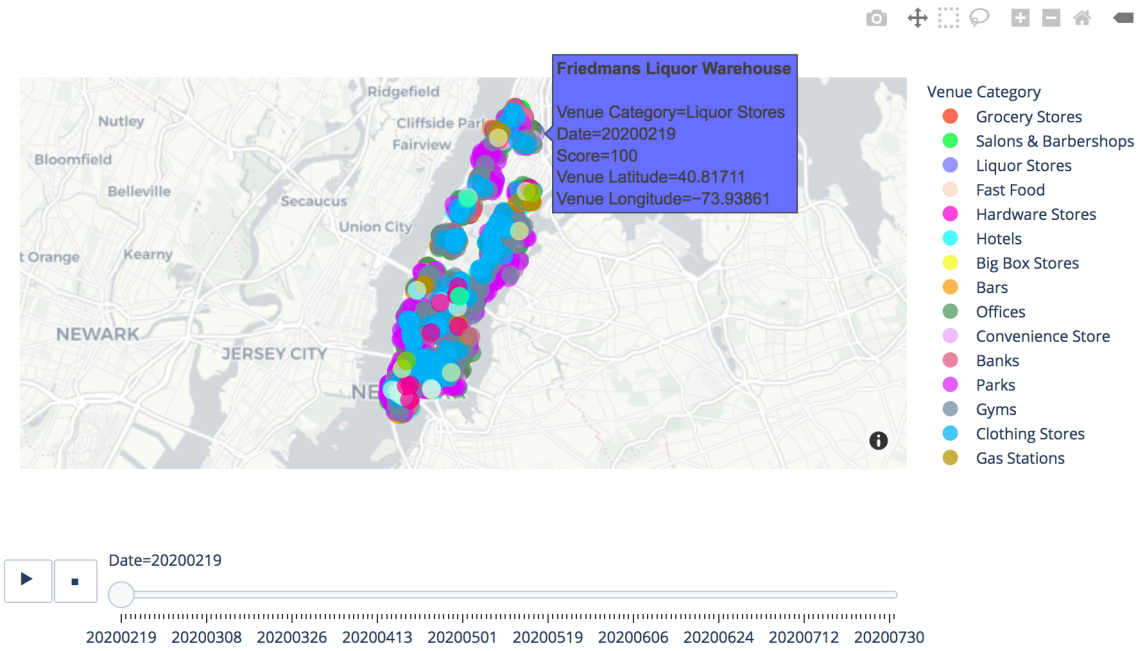

Covid-19 Impact in NYC on Maintenance
Indexed Foot Traffic : Rolling 7-Day
200
150
100
50
Shelter in Place Orders
P1
P2
P3
P4
Mar 2020
Apr 2020
May 2020
Jun 2020
Jul 2020
Date
Category
Banks
Auto Shops
Gas Stations
Hardware Stores

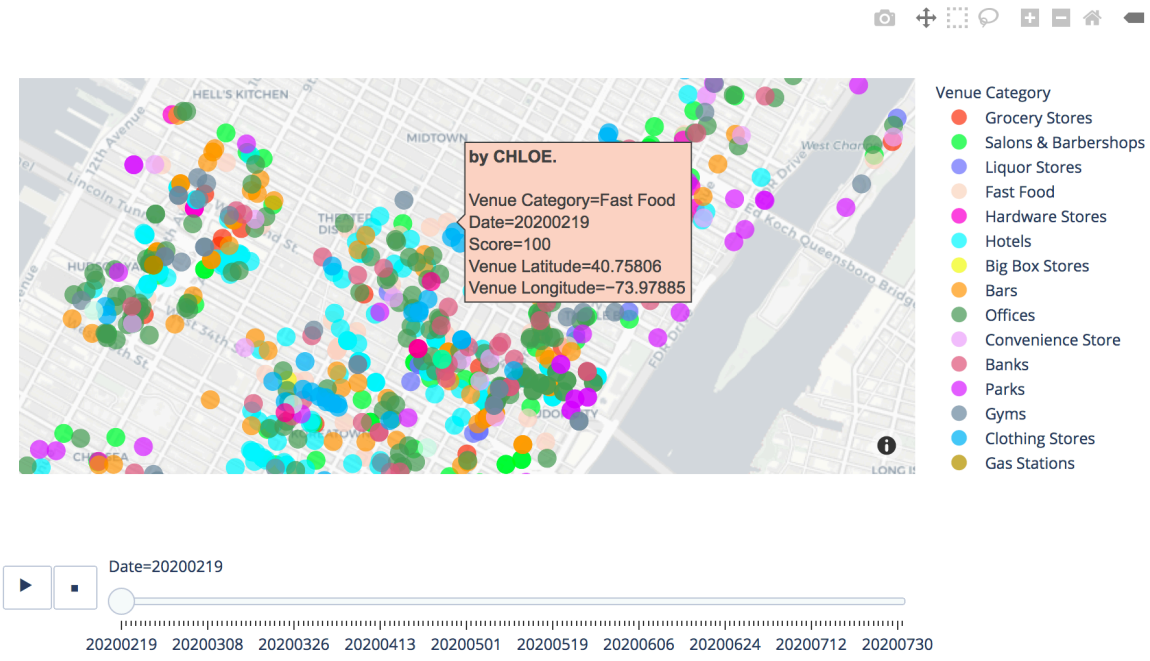

Covid-19 Impact in NYC on Food & Beverage
Indexed Foot Traffic : Rolling 7-Day
120
100
80
60
40
20
Shelter in Place Orders
P1
P2
P3
P4
Mar 2020
Apr 2020
May 2020
Jun 2020
Jul 2020
Date
Category
Bars
Fast Food
Casual Dining Chains
Liquor Stores


Covid-19 Impact in NYC on Travel
Indexed Foot Traffic : Rolling 7-Day
100
80
60
40
20
Shelter in Place Orders
P1
P2
P3
P4
Mar 2020
Apr 2020
May 2020
Jun 2020
Jul 2020
Date
Category
Airports
Hotels


Covid-19 Impact in NYC on Work & Life
Indexed Foot Traffic : Rolling 7-Day
200
150
100
50
Shelter in Place Orders
P1
P2
P3
P4
Mar 2020
Apr 2020
May 2020
Jun 2020
Jul 2020
Date
Category
Clothing Stores
Gas Stations
Gyms
Offices
Parks
Shopping Malls


Covid-19 Impact in NYC on Grooming
Indexed Foot Traffic : Rolling 7-Day
100
80
60
40
20
Shelter in Place Orders
P1
P2
P3
P4
Mar 2020
Apr 2020
May 2020
Jun 2020
Jul 2020
Date
Category
Clothing Stores
Gyms
SalonsBarbershops


Friedmans Liquor Warehouse
Venue Category=Liquor Stores
Date=20200219
Score=100
Venue Latitude=40.81711
Venue Longitude=-73.93861
Ridgefield
Nutley
Cliffside Park
Fairview
Bloomfield
Belleville
Secaucus
Union City
Orange
Kearny
NEWARK
JERSEY CITY
Venue Category
Grocery Stores
Salons & Barbershops
Liquor Stores
Fast Food
Hardware Stores
Hotels
Big Box Stores
Bars
Offices
Convenience Store
Banks
Parks
Gyms
Clothing Stores
Gas Stations
Date=20200219
20200219
20200308
20200326
20200413
20200501
20200519
20200606
20200624
20200712
20200730


by CHLOE.
Venue Category=Fast Food
Date=20200219
Score=100
Venue Latitude=40.75806
Venue Longitude=-73.97885
HELL'S KITCHEN
MIDTOWN
Venue Category
Grocery Stores
Salons & Barbershops
Liquor Stores
Fast Food
Hardware Stores
Hotels
Big Box Stores
Bars
Offices
Convenience Store
Banks
Parks
Gyms
Clothing Stores
Gas Stations
Date=20200219
20200219
20200308
20200326
20200413
20200501
20200519
20200606
20200624
20200712
20200730

4. Conclusion

Preparing Financially, banks saw a meaningful uptick in foot traffic around March 5, with visits up nearly 20% since the week ending February 19. People may have gone to their local bank branches to withdraw cash, apply for a loan, or otherwise make financial arrangements as the outbreak worsened. From March 6 through the week ending March 20, foot traffic to banks seems to have gradually returned to more 'normal' levels.

Visits to grocery stores and warehouse stores are on the rise in New York City since the outbreak of COVID-19, as people shop for food, drinks, and supplies. People seem to be buying in bulk, with visits to warehouse stores like Sam's Club and Costco where foot traffic to warehouse stores was up more than 51% in the beginning of the pandemic. This traffic leveled out once shelter in place orders were put in place and New York shifted to third party shoppers through services to instacart as grocery stores were required to limit shopping behavior due to safety protocols.

Visits to hardware stores increased in New York in particular, with visits up 100% at one point as residents fix up their properties to afford the new shelter in place and work from home daily normal. Hardware stores seeming have benefited the most from the pandemic in New York City.

Drinking at Home, traffic to liquor stores has increased since the outbreak of COVID-19 as people opt to drink at home rather than going to the bar and once bars closed entirely. We've seen a particularly large increase in liquor store visits in the New York area, with visits up 29% from the week ending February 19 to the week ending March 20. The traffic then dropped off to levels lower than before the pandemic. Eventually the traffic picked back up to above the normal traffic pattern as people continue to restock.

Visits to quick serve restaurants, casual dining restaurant chains, restaurants, and bars saw a dramatic drop in New York as the pandemic set in and shelter in place orders took effect. Bars have climbed the first back during the phased re-opening as bars were allowed to offer to-go curbside drink service similar to New Orleans.

Fitness has taken a huge hit during the pandemic as shelter in place orders closed them for the foreseeable future. There is no indication that New York City will re-open fitness centers in 2020. Gyms have pivoted to offering outdoor classes as people begin to shake up their work out routines.

Visits to New York airports declined 48% by the end of March 20 as people cancelled their work trips and travel plans. The Airports still have not picked back up to normal but did rebound slightly as people began to leave the city likely because they no longer had to report into offices. Hotels similarly took a huge hit as people cancelled their trips to New York City when it was the epicenter of the virus. Many Hotels became boarding houses for the medical community that flocked to New York to help.

New Yorkers also took a break from grooming themselves and shopping as shelter in place orders took effect. Eventually in the phase 1 re-opening, in person shopping picked up as the city afforded curbside pick up and eventually limited in-store occupancy. However, Saloons and

Barbershops are still suffering as the previously well groomed city is hesitant to risk grooming for its inactive social life.

In a city that has some of the highest commercial rent prices in the United States, the fabric of the city will be permanently changed by Covid-19 as a significant percentage of businesses will obviously be forced under as they default on their rent obligations. In particular, it seems clothing, restaurants, bars, and cosmetic shops will be impacted the most.

5. Future Direction

This report was limited to the number of API calls that could be made and the indexed scores provided by Foursquare. A much more granular report could be made if we had access to venue specific foot traffic reports and we could build a dataset with all the businesses in New York City. Furthermore, a detailed picture of the commercial landscape could also be drawn if we layered in commercial lease period for each business. If we had this information, then we could forecast commercial real estate opportunities for new businesses to take advantage of as space opens up and market completion declines.