Impact of COVID-19 on New York City's Restaurants

Introduction

New York City is known for its wide variety of food options, thanks to its diverse culture. With everything from street food to high-end dining, the city offers something for everyone.

To ensure that restaurants follow the necessary food safety protocols in this booming industry, the NYC Health Department conducts unannounced and procedural inspections of restaurants at least once a year.

In order to monitor the nutrition profile maintained by fast food chains across the USA, Menustat.org collects data on the nutritional composition of their menu items yearly.

In this study, we aim to utilize this data to examine how the COVID-19 pandemic has affected food chains and restaurants in NYC. We attempt to answer questions regarding how the restaurant inspections and menus have changed during the pandemic and what conclusions can be drawn from these adaptations, if any.

Problem Statement

Among the multitude of effects that the COVID pandemic has had on the restaurant industry, we have attempted to answer the following questions through this short study :

- Is the city conducting more inspections now?
- Are the inspections more thorough/strict?
- How have restaurant menus changed? Were the food choices also affected by disrupted supply chains during the pandemic? If so, which menu items were left out from most menus? How did the share of food categories change during this time?
- Are certain neighborhoods/types of restaurants that were affected more?

- What about the nutritional composition of the menu items for restaurants? Are healthier alternatives being promoted by restaurants now that the importance of eating healthy has been underscored? Are there any trends which are expected to continue (decreasing sugar in beverages - indication of sugar free options rising?)? (good - dietary fiber, protein; bad - cholesterol, sugar, trans fat, saturated fat, total fat ; informative - calories, carbohydrates, sodium,)

- Have health code violations changed? (new categories check for new ones in 2020 to 2023 vs 2019)

Related Work and References

Literature search: In order to gain a better understanding on this particular topic we studied various articles and studies to examine the research that has already been conducted and the results that have been found. Following are the links to some of the preliminary literature these :

[1]- Timeline of NYC Indoor Dining Curfews, capacity limits, reopening : https://www.alwaysthevip.com/all-the-facts-about-ny-2021-reopening-capacity-limits-and-curfews/

[2]- NYC restaurants closing permanently due to covid : https://ny.eater.com/2020/8/12/21336334/nyc-closings-lookback-coronavirus-pandemic-2020

[3]- Health inspector ratings of Asian restaurants during the early COVID-19 pandemic :
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9713539/

https://www.nyu.edu/about/news-publications/news/2022/december/study-finds-nyc-health-inspectors-gave-only-asian-restaurants-hi.html

[4]- COVID-19 Restaurant Impact Survey : https://restaurant.org/nra/media/downloads/pdfs/business/covid-19-restaurant-impact-survey-september-2021.pdf

[5]- Food delivery applications and fast-food consumption during COVID-19 pandemic: a cross-sectional study :
https://www.emerald.com/insight/content/doi/10.1108/NFS-02-2023-0030/full/html

[6]- Changes in Calorie Content of Menu Items at Large Chain Restaurants After Implementation of Calorie Labels

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8719240/

[7]- Trends in the Nutrition Profile of Menu Items at Large Burger Chain Restaurants

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8713464/

[8]- The impact of the COVID-19 pandemic on food price indexes and data collection

https://www.bls.gov/opub/mlr/2020/article/the-impact-of-the-covid-19-pandemic-on-food-price-indexes-and-data-collection.htm

Data Sources : While we came across various research studies employing surveys and alternative data collection techniques to illustrate the effects of COVID-19 on the restaurant industry, we encountered difficulties in locating datasets that supported claims such as reductions in restaurant reservations, growth in online orders and deliveries, and declines in restaurant revenues.

Food chains rarely release the market basket data publically. Moreover, the requirement of datasets covering the temporal period of 2018-2022 and the spatial location of New York further reduced the candidate datasets which could help us study the changes in people’s eating habits.

However, after thoroughly scouring the web for data that could be useful, we have located the following datasets that are relevant to our research.

NYC restaurant inspections
https://data.cityofnewyork.us/Health/DOHMH-New-York-City-Restaurant-Inspection-Results/rs6k-p7g6

Menu items and nutritional information for major US food chains by year
https://www.menustat.org/data.html

Open restaurant applications (June 2020 - 2023; business - location, seating requirements, time of submission)
https://data.cityofnewyork.us/Transportation/Open-Restaurant-Applications/pitm-atqc

We also explored the following datasets which could have been relevant to our study but they lacked either the required temporal granularity or the relevant spatial attributes:

US restaurant and menus on uber eats [No temporal variation]
https://www.kaggle.com/datasets/ahmedshahriarsakib/uber-eats-usa-restaurants-menus?select=restaurant-menus.csv

Yelp open dataset [Does not include NYC restaurants]:
https://www.yelp.com/dataset/download

Yelp APIs
https://docs.developer.yelp.com/docs/fusion-faq
Yelp reviews over time about the NYC restaurants could have been a good source for studying which foods New Yorkers are liking more and how the eating habits have changed.

Although Yelp has released two datasets containing information about reviews, restaurants and more. Unfortunately, these datasets do not include restaurants in New York City.

Yelp also offers programmatic access to its data. However, there is a limitation on the number of reviews that one can access for each restaurant. Currently only 3 reviews per restaurant can be fetched.

Covid statistics : [Granularity : day]
https://data.cityofnewyork.us/Health/COVID-19-Daily-Counts-of-Cases-Hospitalizations-an/rc75-m7u3

National food expenditure :
https://www.ers.usda.gov/data-products/food-expenditure-series/food-expenditure-series/#Current%20Food%20Expenditure%20Series

Final Collection:

1. **MenuStat** is a free nutritional database of thousands of foods served by the nation's largest chain restaurants. It aggregates nutrition information posted on restaurant websites since 2012.Nutrition information was manually collected from restaurant websites in January of each year through 2020. The process for data collection changed for the 2021 dataset, which was collected using web scraping.

   This data is stored in separate tabular csv files for each year. So the first step for utilizing this data was merging the different files to use a common schema for all the years.

   This data contains the following attributes :

Column	**Type**
menu_item_id	Text
year	Numeric Discrete
restaurant	Text
food_category	Text
item_name	Text
item_description	Text
serving_size	Numeric Continuous
serving_size_unit	Categorical Nominal
calories	Numeric Continuous
total_fat	Numeric Continuous
saturated_fat	Numeric Continuous
trans_fat	Numeric Continuous
cholesterol	Numeric Continuous
sodium	Numeric Continuous
carbohydrates	Numeric Continuous
protein	Numeric Continuous
sugar	Numeric Continuous
dietary_fiber	Numeric Continuous
kids_meal	Boolean
limited_time_offer	Boolean
regional	Boolean
shareable	Boolean

2. **NYC Health Inspections** dataset contains every sustained or not yet adjudicated violation citation from every full or special program inspection conducted up to three years prior to the most recent inspection for restaurants and college cafeterias in an active status on the RECORD DATE (date of the data pull). When an inspection results in more than one violation, values for associated fields are repeated for each additional violation record. Establishments are uniquely identified by their CAMIS (record ID) number. Keep in mind that thousands of restaurants start business and go out of business every year; only restaurants in an active status are included in the dataset.

   Records are also included for each restaurant that has applied for a permit but has not yet been inspected and for inspections resulting in no violations. Establishments with an inspection date of 1/1/1900 are new establishments that have not yet received an inspection. Restaurants that received no violations are represented by a single row and coded as having no violations using the ACTION field.

   This dataset is compiled from several large administrative data systems, it contains some illogical values that could be a result of data entry or transfer errors. Data may also be missing.

   This dataset and the information on the Health Department's Restaurant Grading website come from the same data source. The Health Department's Restaurant Grading website is here :
   http://www1.nyc.gov/site/doh/services/restaurant-grades.page

This data contains the following attributes :

Column	Type
CAMIS	Plain Text
DBA	Plain Text
BORO	Categorical Nominal
BUILDING	Plain Text
STREET	Plain Text
ZIPCODE	Categorical Nominal
PHONE	Plain Text
CUISINE DESCRIPTION	Plain Text
INSPECTION DATE	Date & Time
ACTION	Plain Text
VIOLATION CODE	Categorical Nominal
VIOLATION DESCRIPTION	Plain Text
CRITICAL FLAG	Boolean
SCORE	Number
GRADE	Categorical Ordinal
INSPECTION TYPE	Categorical Nominal
Latitude	Numerical Continuous
Longitude	Numerical Continuous

Cleaning The Data :

As is the case with most Data Analysis tasks, a significant portion of our time was spent in cleaning the datasets. Following is a brief description of the steps we have used to do so :

1. **MenuStat**:
    a. Uniqueness constraints were handled by dropping duplicate rows
    b. Value constraints were handled by filtering erroneous values : Non-numeric values have been replaced with null values for columns where numeric values are expected e.g. sugar
    c. Entity Resolution was done by standardizing entity names using fingerprint functions appropriate for the corresponding column - e.g. for restaurant names and food categories a set of case insensitive, normalized words makes the key for fingerprint function (thus marking “John’s Pizza” and “Johns John’s Pizza” as duplicates). However, for item names a list of such words is a better choice of fingerprint key as ‘extra large’ should be treated differently from ‘Extra extra large’
    d. Outlier detection - We ensured that there were no extreme outliers in the nutrition columns as they would affect the means for the corresponding columns which we are using to draw the relevant conclusions.
    e. Dropping non-informative columns : columns which contained a large number of null values (over 75 %) were dropped from the dataset
2. **NYC Health Inspections**:
    a. Uniqueness constraints were similarly handled by dropping duplicate rows
    b. Value constraints were handled by filtering erroneous values : Non-numeric values have been replaced with null values for columns where numeric values are expected e.g. score; non-descriptive restaurant names; years outside the expected range etc. were dropped as they formed a small part of the dataset
    c. Entity Resolution was done by standardizing entity names using fingerprint functions appropriate for the corresponding column - e.g. for restaurant names and cuisine descriptions, a set of case insensitive, normalized words makes the key for fingerprint function (thus marking “John’s Pizza” and “Johns John’s Pizza” as duplicates).
    d. Outlier detection - We ensured that there were no extreme outliers in the score column as they would affect the average scores which eventually form the basis for some of our conclusions. We have also ensured that all the zipcodes have the same length (5 characters) to ensure sound zip code based analysis
    e. Dropping non-informative columns : columns which contained a large number of null values (over 50 %) were dropped from the dataset

Methods:

In this section, we describe the design and methods used to answer each question identified above regarding the impact of the pandemic on restaurants in New York City (NYC).

- ### How have restaurant menus changed? Were the food choices also affected by disrupted supply chains during the pandemic? If so, which menu items were left out from most menus? How did the share of food categories change during this time?

    We utilize the cleaned restaurant menus dataset to calculate the following

    i. Total number of menu items available for each year
    ii. Total number of chain - restaurants in each year
    iii. For each food category, the year on year percentage change in terms of number of menu items available (e.g. if beverages increase from 70 to 140 from 2020 to 2021, there is an increase of 100%)
    iv. For each menu item, the year on year change in terms of number of occurrences (since the number of repetitions of menu items is low, we do not use a percentage change here)

- What about the nutritional composition of the menu items for restaurants? Are healthier alternatives being promoted by restaurants now that the importance of eating healthy has been underscored? Are there any trends which are expected to continue?

    We utilize the scaled (using MinMax Scaler) and cleaned restaurant menus dataset to calculate the average and variance for all of the nutrition related columns (‘calories','total_fat','saturated_fat','trans_fat','cholesterol','sodium','carbohydrates','protein','sugar','dietary_fiber') grouped by food_category and year and search for trends and outliers in this normalized data.

- ### Is the city conducting more inspections now?

    We counted the number of inspections for each year starting in 2019. We also counted the number of distinct restaurants for each year - indicated by the number of unique CAMIS IDs in the dataset. Both of these, when combined, will help determine whether there are more restaurants being inspected therefore increasing the number of total inspections. We also normalize the number of inspections for each year by dividing it with the number of unique restaurants inspected in the corresponding year to calculate the average number of inspections per restaurant.

- Are the inspections more thorough/strict after the pandemic ?

  In order to determine if the inspections have gotten more strict we used the following methods:

  1. We compared the number of restaurants that have closed due to the inspections failing using the “ACTION” column in the inspections data set. The action column consists of the following entry types:

ACTION	Count of Rows (Restaurant Citations /Restaur…
Violations were cited in the following area(s).	186,117
Establishment Closed by DOHMH. Violations were …	7,412
Establishment re-opened by DOHMH.	1,814
No violations were recorded at the time of this ins…	1,115
Establishment re-closed by DOHMH.	4

  We filter on the data set using this column only selecting the following:

  1.Establishment Closed by DOHMH. Violations were cited in the following area(s) and those requiring immediate action were addressed.
  2. Establishment re-closed by DOHMH.

  2. We checked the number of inspections with the “CRITICAL FLAG” column being flagged as critical. Critical violations are those most likely to contribute to food-borne illness and add the most points to an inspection.

  3. We compared the number of restaurant closures due to failed inspections, the number of inspections with critical flags, and the number of unique restaurants inspected in 2019 and 2022. These years were chosen because they represent pre- and post-pandemic conditions and provide the most accurate data in the dataset. Furthermore, we normalized the inspection and restaurant closure

counts by the number of unique restaurants inspected to ensure a fair comparison.

4. Furthermore, Since 2010, New York City has required restaurants to post letter grades that correspond to scores received from sanitary inspections. An inspection score of 0 to 13 is an A(Best possible grade) , 14 to 27 points is a B, and 28 or more points is a C(Worst possible grade). We checked the average score for each year between 2019 to 2023 to check if there was any increase in the score - which would indicate a decline in the restaurant’s safety standards or a corresponding increase in the thoroughness of inspections.

5. We examined the "Violation Code" column to identify any violations that were not cited in 2019 but were cited in 2022. The "Violation Code" column contains a code that corresponds to a specific type of violation. The presence of new violation codes in 2022 may suggest that inspectors are less tolerant during inspections and are issuing citations for violations that they previously overlooked. Below is an image to show different violation codes and their descriptions.

APPENDIX 23-C:
FOOD SERVICE ESTABLISHMENT AND NON-RETAIL FOOD PROCESSING ESTABLISHMENT PENALTY SCHEDULE

SCORED VIOLATIONS									
VIOLATION CODE	CITATION	CATEGORY	VIOLATION DESCRIPTION	CURE ACCEPTED OR $0 PENALTY FIRST-TIME VIOLATIONS	VIOLATION PENALTY CONDITION I*	VIOLATION PENALTY CONDITION II*	VIOLATION PENALTY CONDITION III*	VIOLATION PENALTY CONDITION IV*	VIOLATION PENALTY CONDITION V*
02A	NYCHC 81.09(c)	Public Health Hazard	Time and temperature control for safety ("TCS") hot food not heated to 140°F for 15 seconds					$400	$600
02A	NYCHC 81.09(c)(1)	Public Health Hazard	Poultry, poultry stuffing, parts and ground, comminuted poultry not heated to 165°F for 15 seconds					$400	$600
02A	NYCHC 81.09(c)(2)	Public Health Hazard	Pork/food containing pork not heated to 150°F for 15 seconds					$400	$600
02A	NYCHC 81.09(c)(3)	Public Health Hazard	Rare roast beef/steak not heated to minimum time/temperature					$400	$600
02A	NYCHC 81.09(c)(4)	Public Health Hazard	Ground, comminuted meat, foods containing ground, comminuted meat not heated to 158°F					$400	$600
02A	NYCHC 81.09(c)(5)	Public Health Hazard	Stuffed meats, fish, ratites and tenderized meats not heated to 165°F for 15 seconds; injected, mechanically tenderized meats not heated to 155°F					$400	$600
02A	NYCHC 81.09(c)(6)	Public Health Hazard	Shell eggs/food containing shell eggs not heated 145°F for 15 seconds					$400	$600
02A	NYCHC 81.09(c)(7)	Public Health Hazard	Raw animal food cooked in microwave not heated to 165°F, not covered, rotated or stirred, not held for 2 minutes					$400	$600
02B	NYCHC 81.09(a)	Public Health Hazard	Hot TCS food not held at 140°F or above		$250	$300	$350	$400	$500
02C	NYCHC 81.09(d)	Critical	Previously cooked and cooled TCS food not reheated to 165°F for 15 seconds within 2 hours		$200	$200	$250	$300	

- Are certain neighborhoods/types of restaurants affected more than others?

  To determine whether certain neighborhoods are worse than others in terms of violation count and severity of the violations we used the following methods:

1. Compare the total number of total inspections grouped by the column "CUISINE DESCRIPTION" for the years 2019 and 2022.

2. Based on the findings from above we saw that Coffee/Tea and Chinese restaurants were affected the most from 2019 to 2022. We used this finding to test the following hypothesis to further explore the data:
    a. Were Coffee/Tea restaurants closed more than others on average in 2020 due to most NYC employees working from home? Did they reopen the same amount as other cuisine types?
        i. To determine this we counted the year on year change of the number of unique Coffee/Tea restaurants and all other cuisine types.
    b. Were chinese restaurants targeted more by health inspectors due to a potential bias due to the food related origin of Covid pandemic in Wuhan, China?
        i. To determine this we first calculate the increase in the number of inspections for Chinese and Asian restaurants from 2019 to 2022 and compared to American and all other restaurants.

ii. We checked if certain neighborhoods were inspected more by checking the spatial distribution of inspections, to determine if there was an increase in inspections for zip codes with predominantly Asian populations.

iii. We calculated the average score for restaurants grouped by CUISINE TYPE and compared the values for Chinese restaurants against all others 2019 (pre-pandemic) and 2022(post pandemic).

MenuStat Results:

The results of our findings have been summarized below :

The change in number of menu items available on the restaurant menus is evident from Jan 2020 to Jan 2021. Through this period the number of restaurants was the same (106), however all the food categories face a negative change between -13 % to -33 % with the total number of available items going down from 33.2k to 25.6k. Categories which were most impacted by the disrupted supply chains during the pandemic include Entrees which require a larger variety of ingredients.

Figure 1 : Number of restaurants with detailed menu information

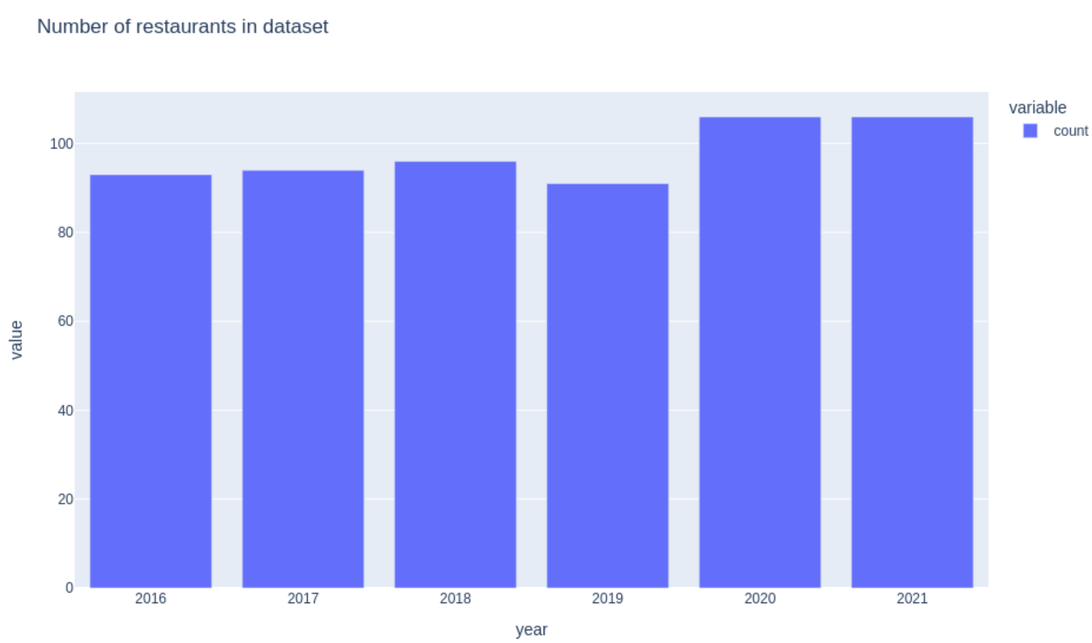


Figure 2 : Number of menu items available in dataset by year

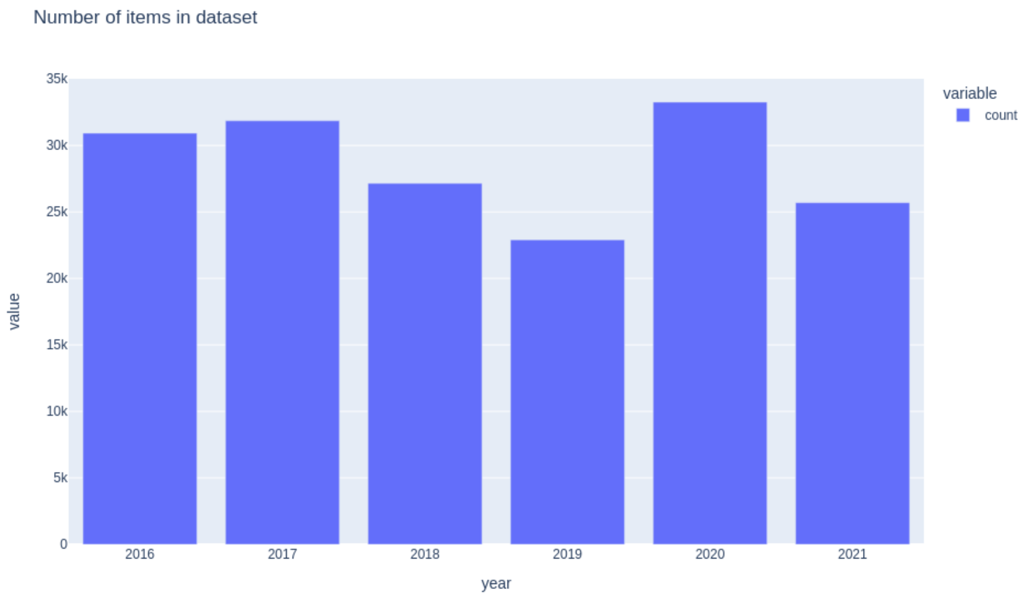


Figure 3 : Year on year change in percentage of menu items by category

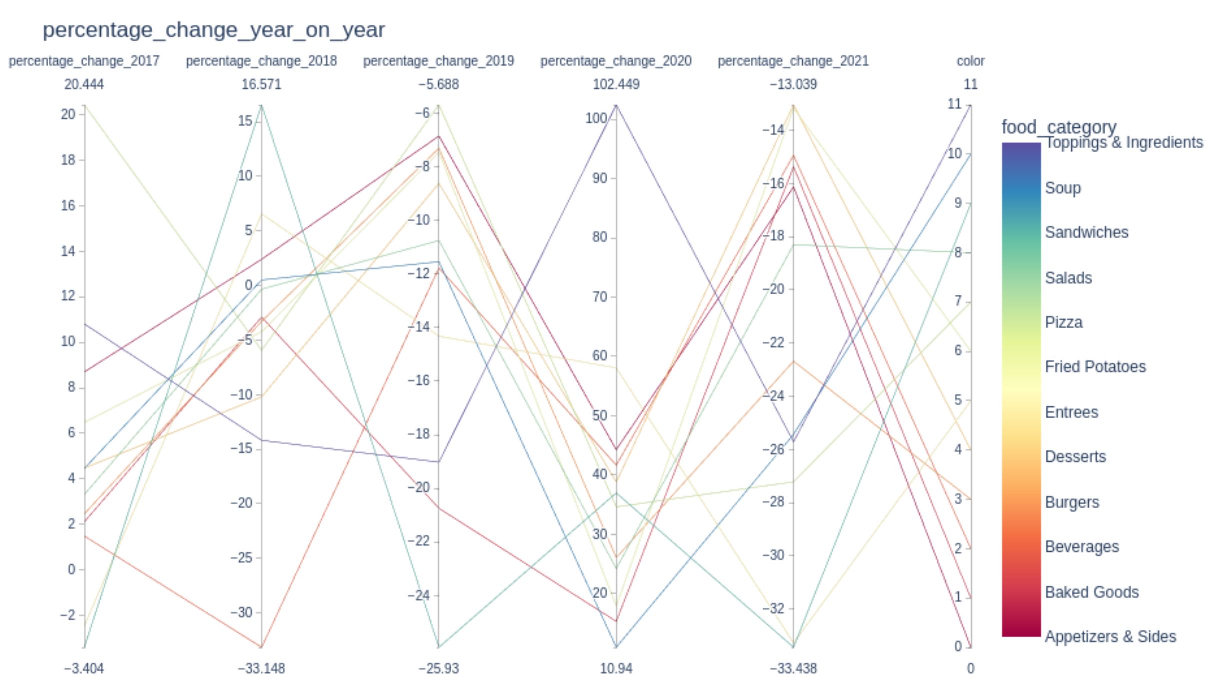
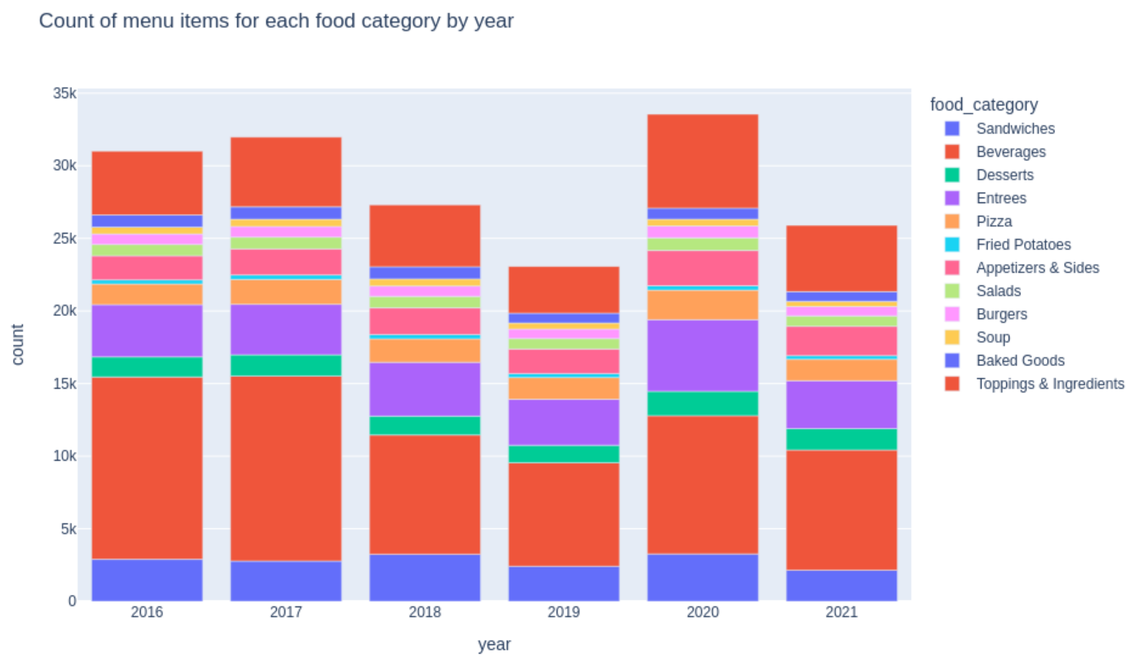


Figure 4 : Share of various categories in the menus by year



As confirmed by this report[8] from US Bureau of Labor Statistics, supply chains for meat, fish, dairy, and eggs were especially affected by the shifting economy brought on by the pandemic. This is demonstrated by the highest reduction in menu items involving cheese, chicken, bacon and milk for both years - 2020 and 2021 (excluding specific beverages).

Figure 5 : Year on year change by item names in 2020 (top 20 items with most negative change)

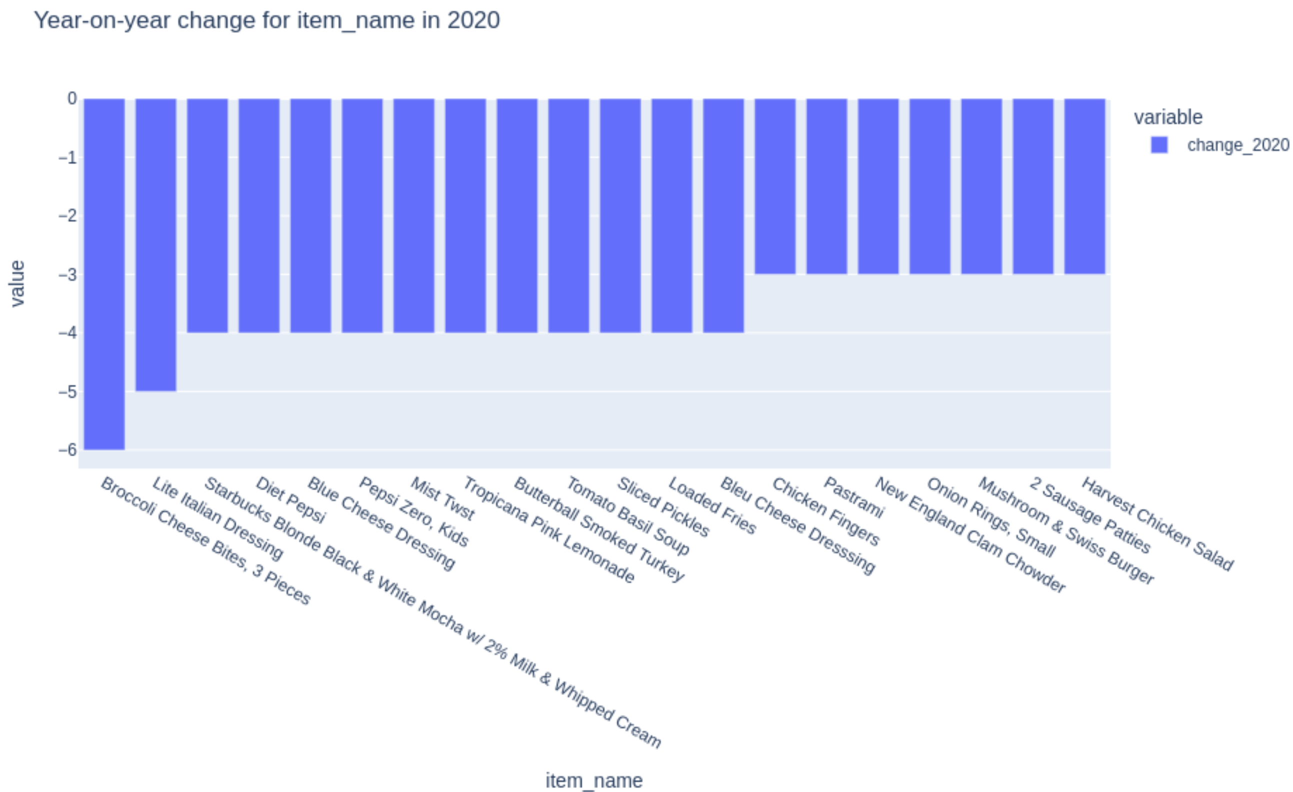
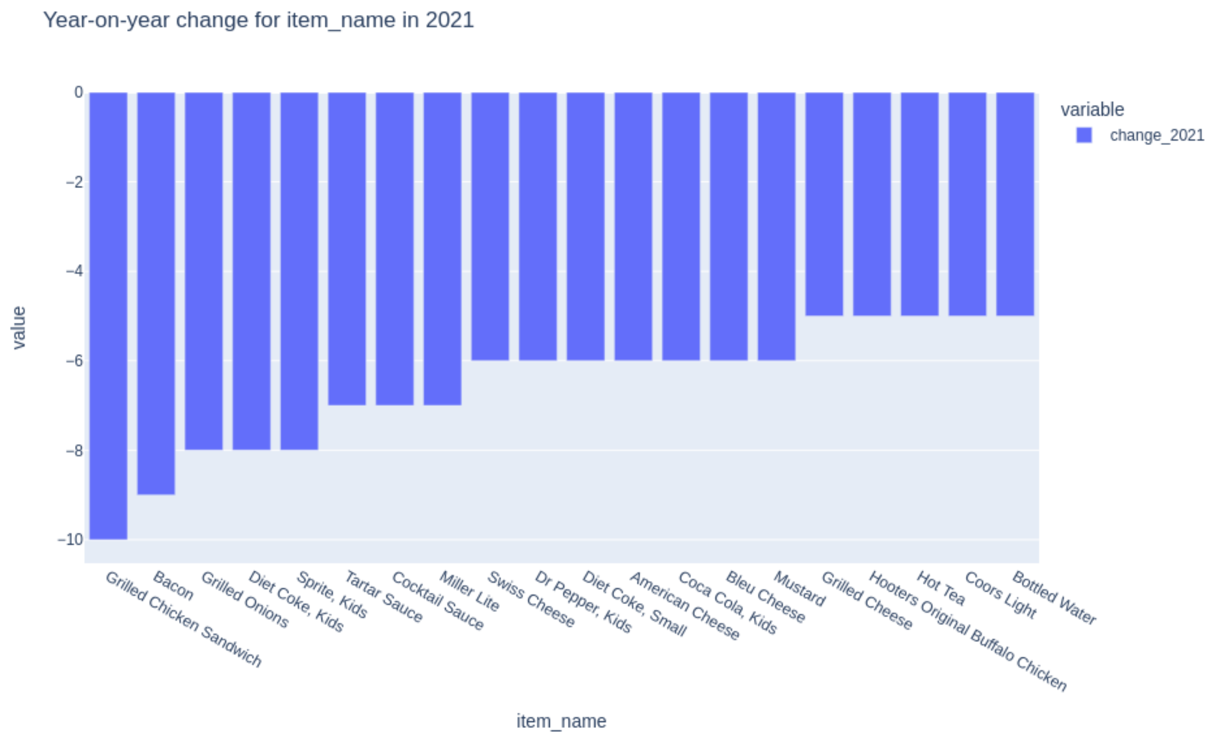


Figure 6 : Year on year change by item names in 2021 (top 20 items with most negative change)



Consistent with [6] we do not observe any significant change in the average calorie content of any of the food categories. Nor does the calorie profile (relative ordering of categories in terms of average calories) of menus change across food categories.

Figure 7 : Yearly average scaled calories by food category

Amount of sugar in beverages has been on a continuous decline in recent years, indicating more availability of sugar free options. This has been confirmed by a sharp increase in the number of items which include ‘zero sugar’ or ‘sugar free’ in their descriptions.

Figure 8 : Yearly average scaled sugar content by food category

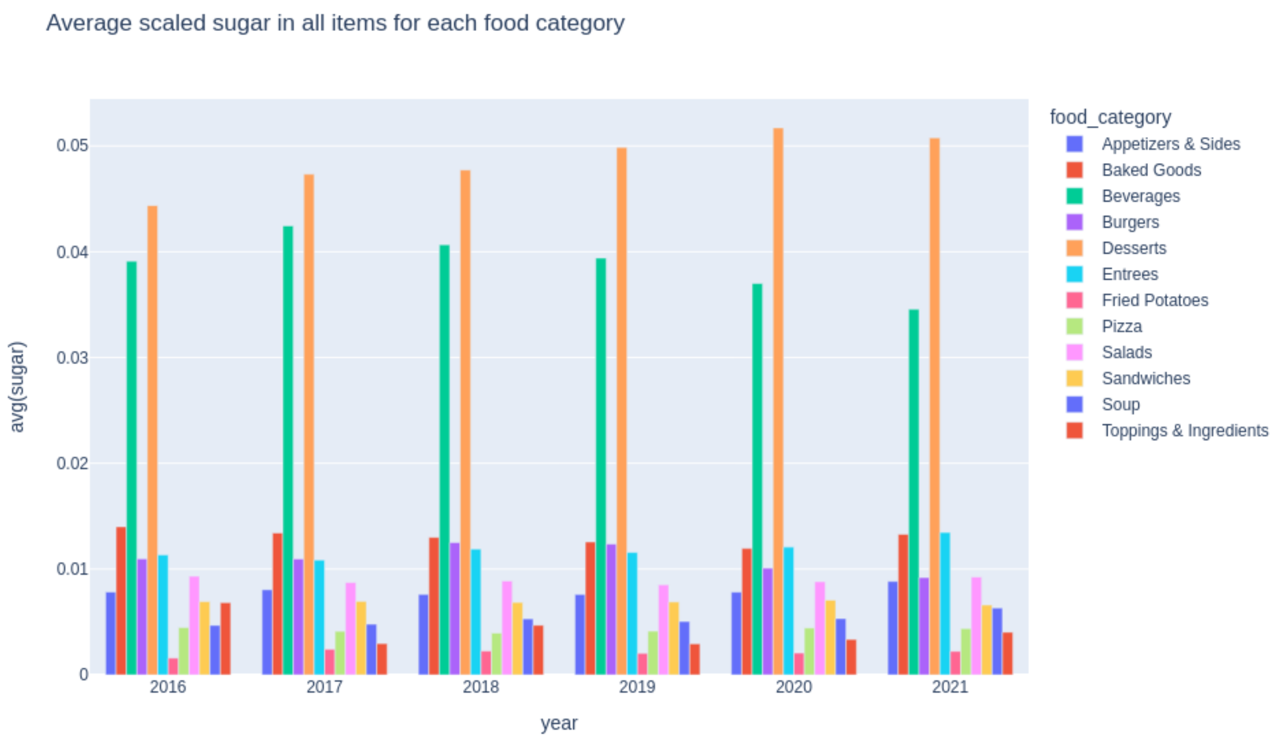
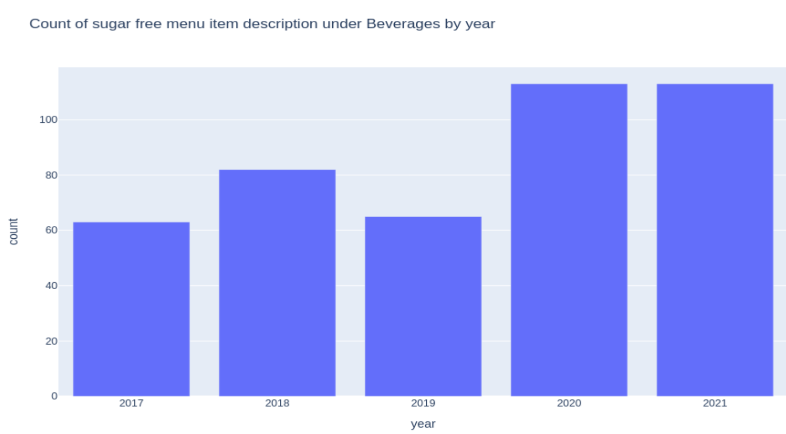


Figure 9 : Count of menu items mentioning ‘zero sugar’ or ‘sugar free’ by year



Entrees showed a significant increase in the fat content accompanied by a large increase in the cholesterol content:

Figure 10 : Yearly average scaled total fat content by food category

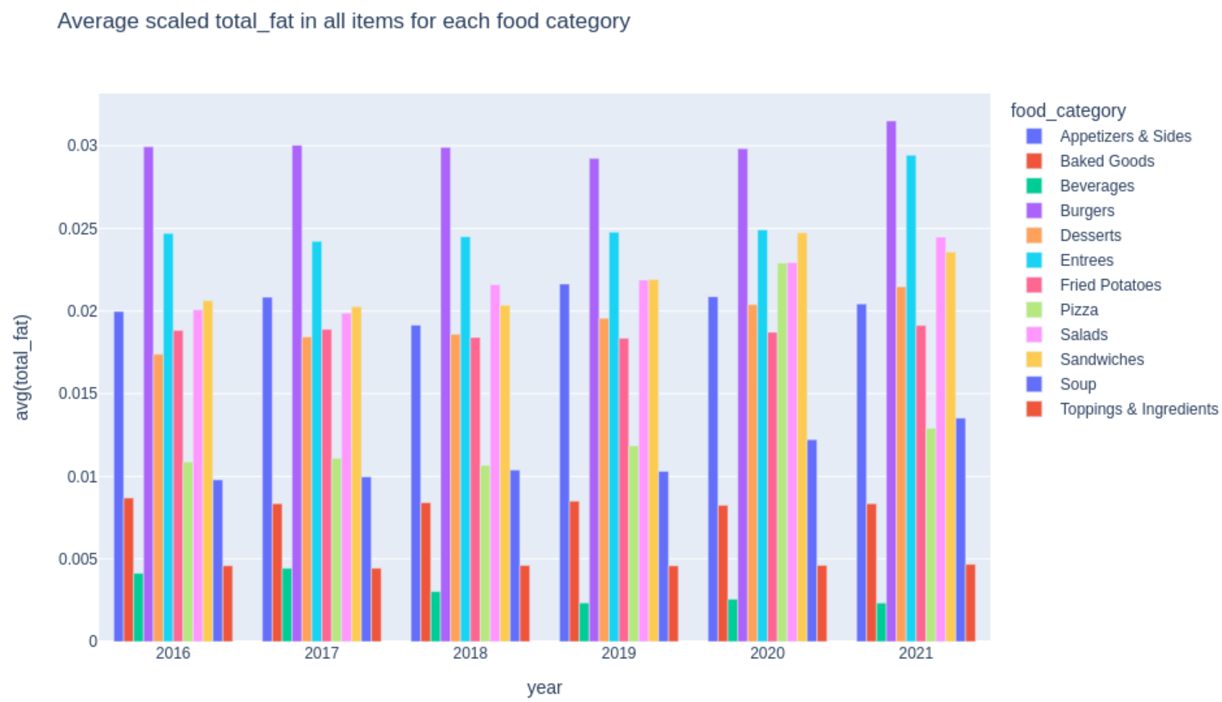
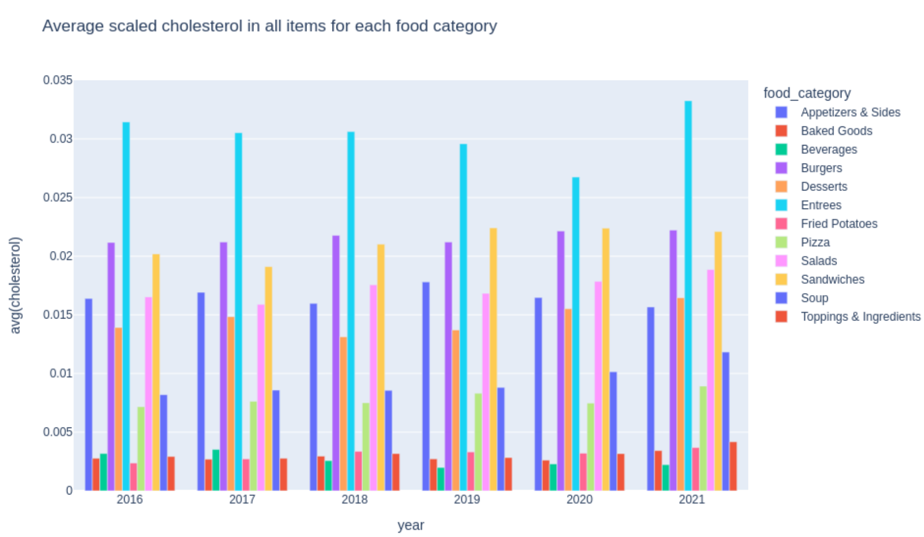


Figure 11 : Yearly average scaled cholesterol content by food category



A large change in the variance can indicate changing distribution of menu items. E.g. large variance in 2021 for protein content in Entrees due to affected supply chains of protein rich items (e.g. Meats) or smaller variance in Appetizers' protein content potentially due lack of choices in the corresponding menu items (confirmed by the largest decrease of 33% in number of Entree menu items from 2020 to 2021):

Figure 12 : Yearly variance in scaled protein content by food category

Inspection Dataset Results:

We conducted an analysis of the number of inspections and number of unique restaurants inspected annually from 2018 to 2023. The resulting data revealed a similar trend for both metrics. Specifically, there was a decline in both the number of inspections and unique restaurants inspected between 2019 and 2020, which can be attributed to the shutdown of indoor dining by the New York State in April 2020. There was a slight increase observed between 2020 and 2021 as businesses started to reopen. In 2022, there was a significant increase in both metrics, which can be linked to the resumption of indoor dining in NYC after the pandemic. This indicates that the city took measures to enforce restaurants comply with health and safety regulations during this period.

Figure 13: Number of Inspections Per Year

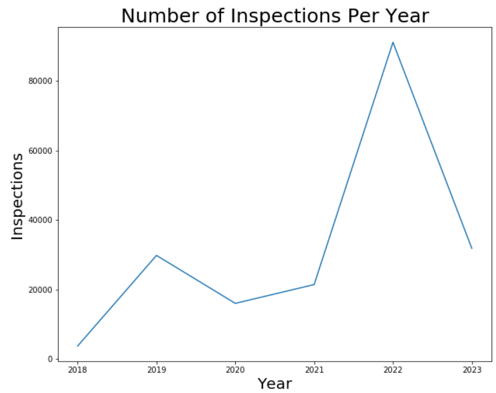
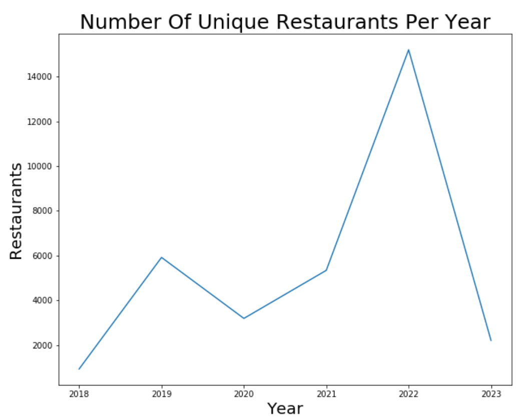
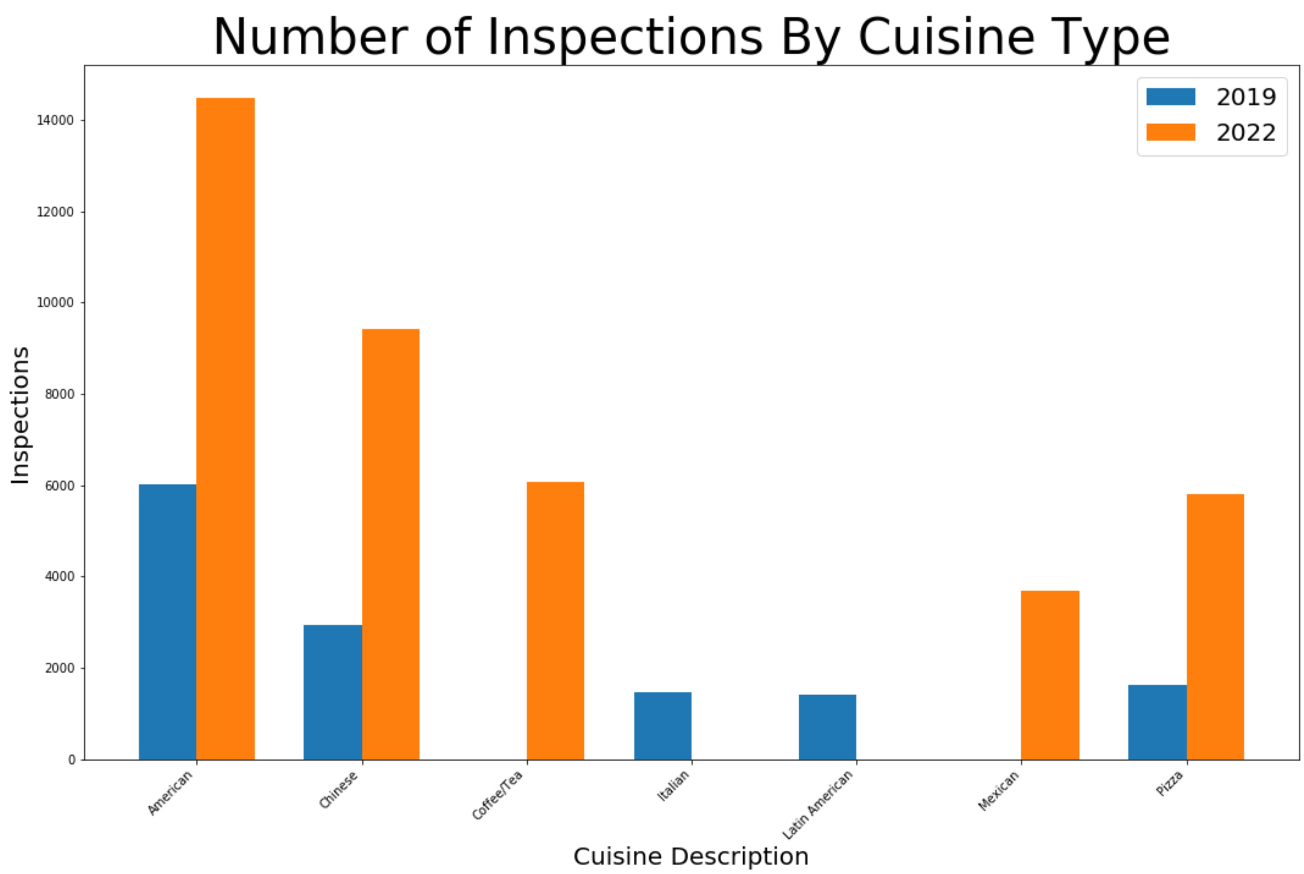


Figure 14: Number of Unique restaurants Per Year



We analyzed the number of inspections for different cuisine types in both 2019 and 2022 and identified the most frequently inspected types for each year.

Figure 15: Number of Inspections Per Cuisine Type



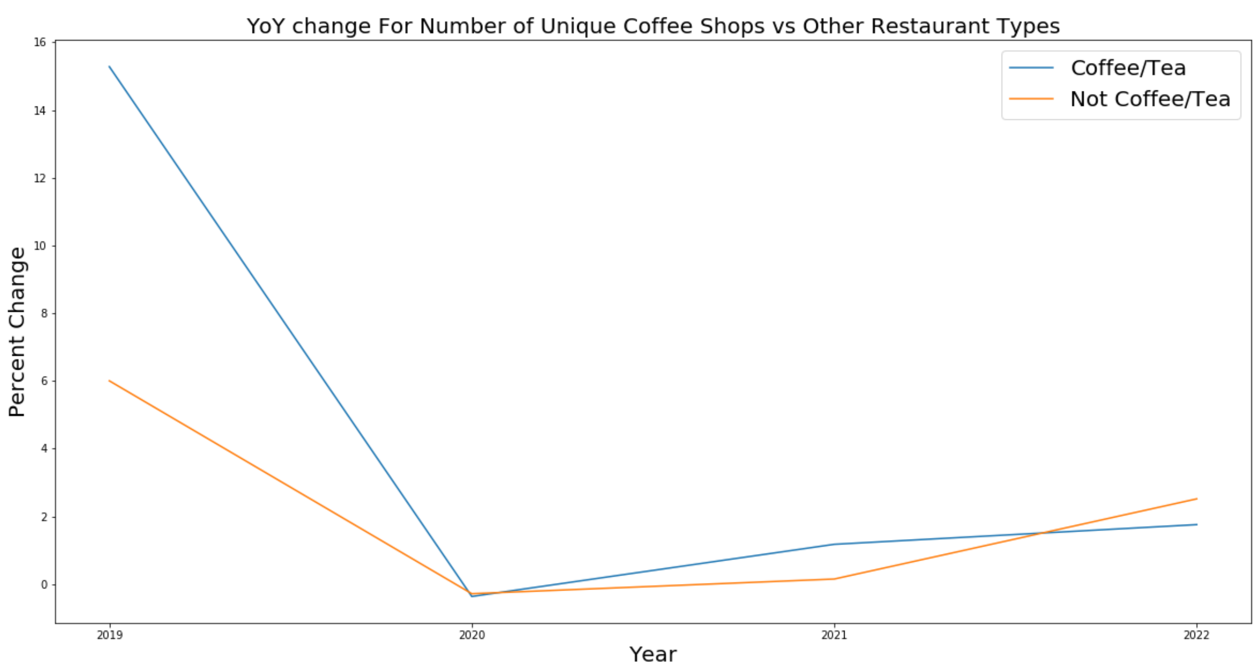
To normalize the data, we calculated the percentage increase per restaurant since there may have been an increase in the number of unique restaurants in a particular cuisine type, which would increase the total number of inspections for that cuisine type. The results of our analysis are as follows:

- Coffee/Tea: experienced a 17.03% increase in inspections from 2019 to 2022.
- Chinese: experienced a 9.47% increase in inspections from 2019 to 2022.
- Pizza: experienced a 7.41% increase in inspections from 2019 to 2022.
- American: experienced a 3.66% increase in inspections from 2019 to 2022.
- Mexican: experienced a -5.91% decrease in inspections from 2019 to 2022.
- Latin American: experienced a -6.64% decrease in inspections from 2019 to 2022.
- Italian: experienced a -16.11% decrease in inspections from 2019 to 2022.

This analysis provides a more accurate comparison of inspection frequency across cuisine types and suggests that Mexican, Latin American, and Italian restaurants had fewer inspections in 2022 compared to 2019, while Coffee/Tea, Chinese, Pizza, and American restaurants had increased inspections during the same period.

Considering the results mentioned earlier, we focused our investigation on whether Coffee/Tea restaurants experienced more closures than other cuisine types in 2020 due to the increased work-from-home trend and reduced commuting. Our analysis revealed a significant decrease in the total number of unique restaurants inspected from 2019 to 2022, suggesting that many restaurants had to close down during this period. Moreover, compared to the average increase in unique restaurants inspected for other cuisine types, Coffee/Tea restaurants had fewer inspections. This can indicate that they were more severely impacted by the pandemic, which is likely due to the shift in work patterns caused by remote work and the shutdowns by the government.

Figure 16: Year over year change for number of unique coffee/tea vs all other restaurants.



We then investigated the possibility of discrimination against Chinese/Asian restaurants during the pandemic. Through our literature review, we came across articles and studies suggesting that Chinese restaurants were unfairly targeted during the early stages of the pandemic. To explore this issue further, we conducted a comparative analysis of changes in inspections for Chinese restaurants and other cuisine types. Our findings revealed a larger increase in the percentage change of inspection frequency from 2019 to 2022 for Chinese restaurants, particularly when compared to American cuisine types. Additionally, we observed that the average inspection score for Chinese restaurants increased by 17.5 points, which was significantly higher than the average increase for all other cuisine types (less than 10 points). These results are consistent with similar studies that were conducted for shorter periods. In conclusion, our analysis indicates that the Chinese/Asian community may have been subjected to greater scrutiny and discrimination than other cuisine types, especially American restaurants.

Figure 17: Increase in Inspections by cuisine types.

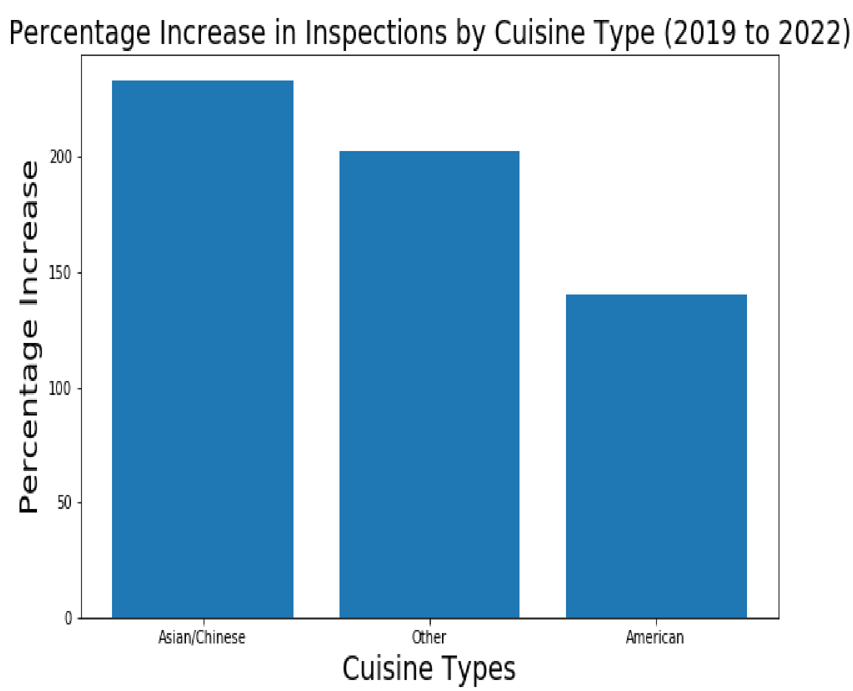
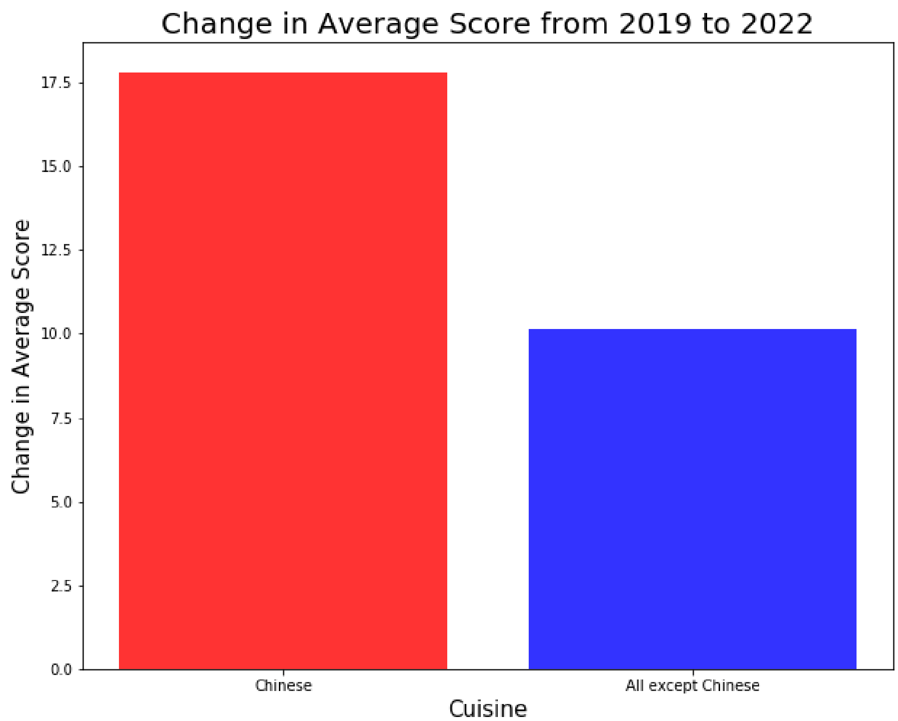


Figure 18: Increase in average score for Chinese restaurants vs all others.



Initially, we aimed to determine if certain neighborhoods in NYC were affected more than others or if there was an increase in inspections for specific areas. To achieve this goal, we identified the top three most frequently inspected neighborhoods in 2019 and 2022. Interestingly, our analysis provided additional evidence that supported our initial problem statement. Specifically, we observed that the zip code corresponding to Chinatown was not among the top three neighborhoods with the highest frequency of inspections in 2019, but it became the most frequently inspected area in 2022. Meanwhile, East Village and Midtown West remained in the top three neighborhoods, but Midtown was replaced by Chinatown.

Figure 19: Top 3 zip codes with most inspection in 2019

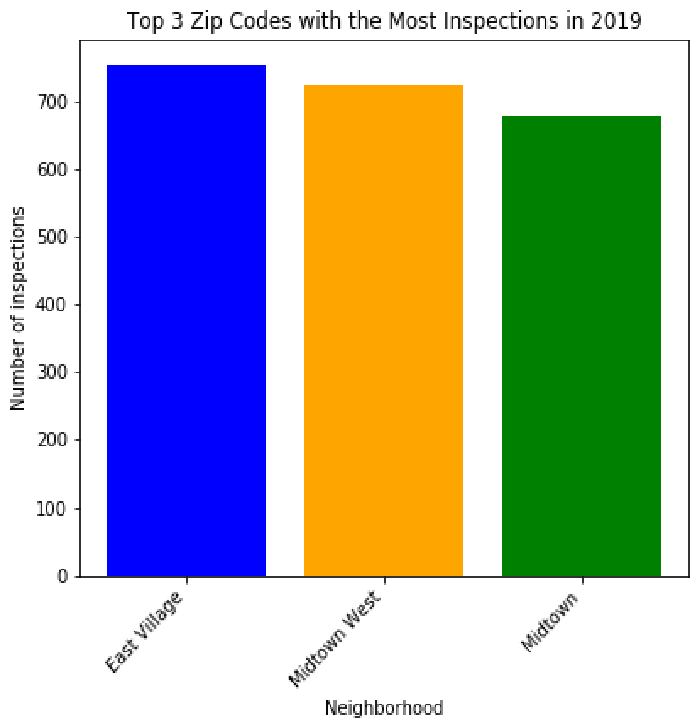
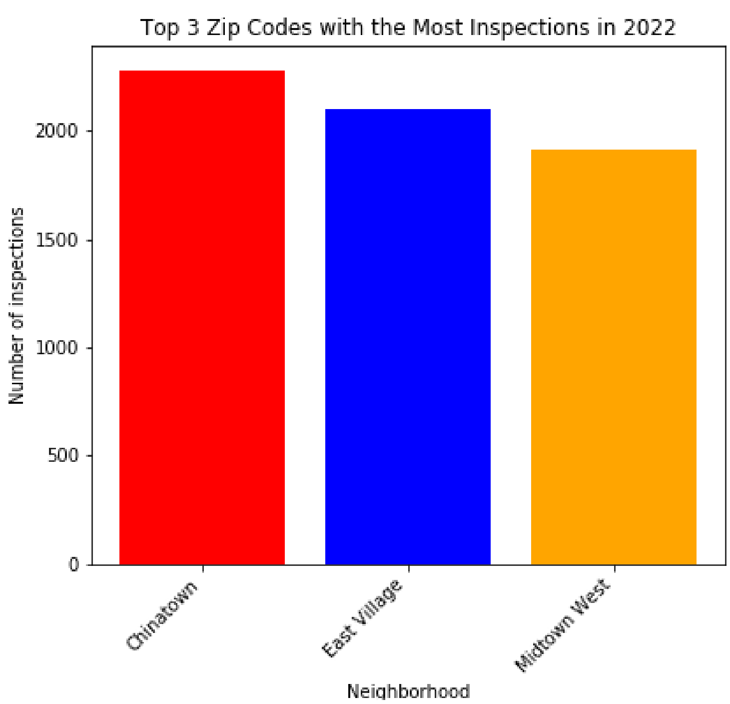
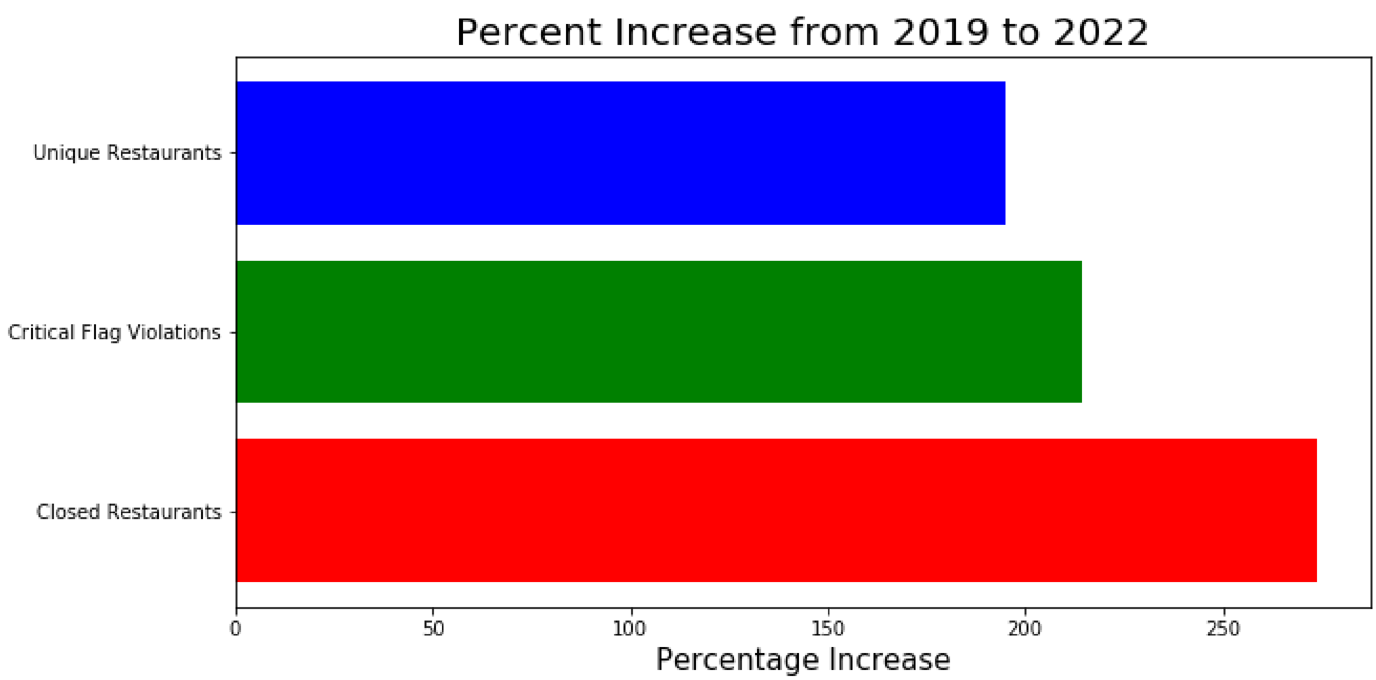


Figure 20: Top 3 zip codes with most inspection in 2022



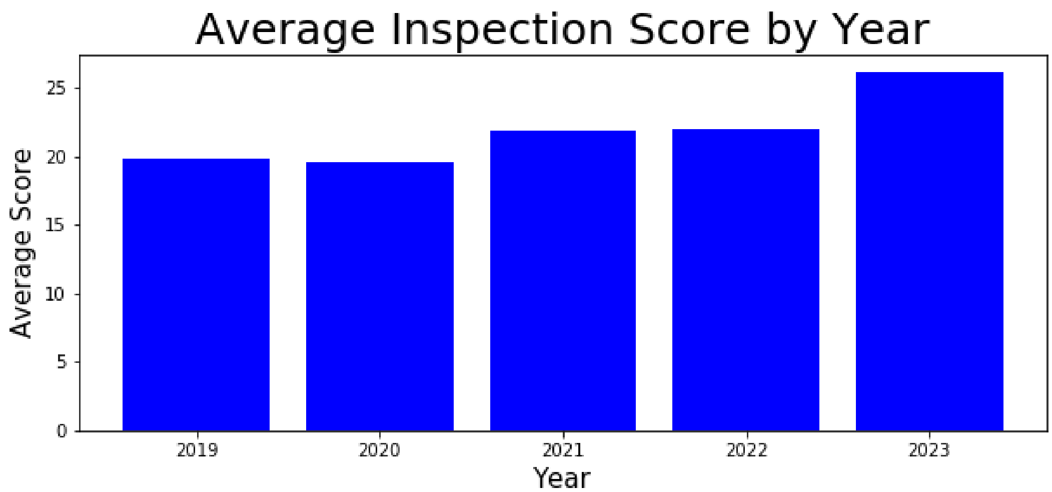
Subsequently, we examined whether inspections became more rigorous after the pandemic, as the government may have aimed to reduce the number of people falling ill due to the virus, considering that hospitals were already overwhelmed with COVID-19 patients. We analyzed the number of restaurants that were closed or received critical flags during inspections. To ensure a fair comparison, we normalized the data by comparing the increase in the number of closed restaurants and critical flags to the increase in the number of unique restaurants inspected. Our analysis revealed that the number of closed restaurants and critical flags increased more than the increase in the number of unique restaurants inspected, indicating that inspections became more stringent after the pandemic.

Figure 21: Percentage increase from 2019 to 2022 in unique restaurants, critical flag violations, and closed restaurants.



To further support this hypothesis, we analyzed the average score year over year. Each inspection is provided with a resulting score, in which the lower the score the better the inspection was for that particular restaurant. The results were quite clear as we saw an increase in the average score per year indicating there were possibly more stricter inspections.

Figure 22: Average Inspection Score Per Year



In the final stage of our analysis, we investigated the types of violations being issued and whether any new violations were introduced in 2022 that were not present in earlier inspections. Our assumption was that the identification of such new violations could indicate that inspectors were more lenient in the past and did not issue citations for these violations. We identified the most frequent violations that were absent in inspections prior to 2020 but were present in inspections conducted after 2020:

1. 20-04 -Alcohol use during pregnancy sign not posted
2. 20-06 -Grade card not posted after reopening when FSE ordered closed by BFSCS
3. 09E - Wash hands sign not posted
4. 20-01-Allergy poster not posted or not in correct location
5. 28-06-Contract with pest management professional, record of pest exterminations, activities not kept on premises

Based on our analysis of the average score, the number of closed restaurants, the number of critical flags, and the appearance of new violation types, it is evident that inspections became more stringent after the pandemic.

Limitations and Future work

- In order to answer questions regarding how the eating habits of people have changed during the pandemic, we require reliable market basket data which isn't yet publicly available for NYC restaurants for the time period between 2018 to 2022. Therefore the corresponding questions could not be answered. Online reviews for restaurants in NYC from the last 5 years could be scraped to understand how people's eating habits changed
- Given more time, we would like to test the proposed hypotheses to determine their statistical significance
- We would like to study how violations and scores correlate with restaurant ratings on Yelp/Google. We also intend to analyze if restaurants have adapted to the new trends such as takeouts and contactless orders and evaluate whether these trends are going to be long lasting even after the pandemic
- [7] studies the changes in nutrition profile of top burger restaurants from 2012 to 2016 using Nutrition Profile Index. We would like to extend this work to the larger set of restaurants till 2021 to judge whether restaurants have improved the nutrition profile of their menus in recent years
- Explore potential correlations between COVID-19 case statistics and the observed changes in inspections dataset

Code Repository and Data

https://github.com/himanshu1196/big_data_project

Configuration used for running :

1- Restaurant_Menu_Analysis.ipynb :

Environment :

Google Colab CPU with

MemTotal: 13297192 kB memory

108 GB available disk space of which 26 GB was used

Operating System :

Distributor ID: Ubuntu

Description: Ubuntu 20.04.5 LTS

Release: 20.04

Codename: focal

2- Inspection_Data_Analysis.ipynb :

Environment :

NYU HPC Cluster with

MemTotal: 123658260 kB memory

>= 10GB of disk space (Less than 2 GB utilized)

Operating System :

Distributor ID: Debian

Description: Debian GNU/Linux 10 (buster)

Release: 10

Codename: buster