**[place holder for Abstract]**

1. **Introduction and Problem Statement**

As of December 6th, 2020 COVID-19 had infected over 66m people, caused over 1.5m deaths, and devastated living standards across the world, according to the COVID Tracking Project. While governments around the world have largely responded by mobilizing resources to tame the immediate public health crisis, the indirect damages in the form of economic scaring, a deterioration in mental hygiene, diminished learning outcomes, and exacerbation of social inequalities, among other legacies, stemming from the pandemic-induced shutdowns deployed by governments will arguably pose no less a challenge to public authorities.

Big cities were particularly vulnerable to a virus that thrives within societies characterized by high population density, exemplified by the course of virus in New York City (NYC), which with ~3% of the United States’ population, accounts for ~9% of the deaths officially caused by the disease (Covid Tracking Project). Furthermore, given NYCs largely service driven economy and reliance on crowded public transportation, city-wide mandated shut-downs and social distancing requirements have left the city facing a formidable socioeconomic crisis just as the city’s fiscal position - largely underpinned by sales, personal income, and property taxes - comes under severe strain with personal income tax revenue expected to drop by $2 billion this fiscal year, sales tax revenue down by ~11% year-to-date, and a ~4.7% rise in the rate of delinquent property tax payments that were due July 1st, culminating in a $9 billion revenue deficit and a $7 billion reduction in the city’s annual budget (Reuters, 2020). Without state and federal funds, itself not a forgone conclusion with New York State facing its own deficit challenges and federal legislators at a stalemate in negotiations over further stimulus measures, the city faces the prospect of spending cuts just when the opposite is needed to help rebuild the cities vibrancy.

Complicating the situation further is the disproportionate impact the pandemic has had on communities of color which are overrepresented in several of the low-wage service sectors that were decimated by the pandemic, including bricks-and-mortar retail, leisure, and hospitality, all of which do not lend themselves to remote working, thereby exposing these communities to a higher risk of infection and/or unemployment.

Given the unprecedented scale of the challenges ahead and the need to allocate scare fiscal funding efficiently, this study seeks to quantify how COVID-19 has impacted various socioeconomic outcomes in NYC in order to inform a targeted, high-impact fiscal policy moving forward. We will focus our study on investigate the following hypotheses:

* COVID-19 has likely resulted in an above average rise in poverty
* COVID-19 has likely resulted in an above average deterioration in health care security
* COVID-19 has likely resulted in an above average deterioration in public safety
* COVID-19 has likely resulted in an above average rise in income insecurity
* COVID-19 has likely impeded student learning

(Covid Tracking Project, 2020)

1. **Literature Review and Hypotheses**

For the present study, several related works have been chosen as a basis for further exploration.

* 1. *Poverty and Income*

(Han, Meyer, & Sullivan, August 2020) use high frequency data from the Basis Monthly Current Population Survey, validated by comparing historical estimates derived using this source to those based on official surveys, to measure income distribution and poverty during the pandemic. Using income as a measure of poverty, the paper finds that, leading up to the pandemic, poverty had been steadily decreasing, falling by 0.9 percentage points between November 2019 and February 2020 but subsequently accelerated to 1.5 percentage points (~14%) between February and June. Within demographic groups, poverty was found to have fallen by 17.1%, 16.1%, and 11.1% for individuals aged 65+, 18-64, and 0 – 17, respectively while declines were observed across all racial and gender groups with those categorized as other (neither white nor black) experiencing a ~25% decline in poverty. The paper attributes the observed decline in poverty primarily to government assistance, including the CARES Act signed into law on March 27th which, through the Economic Impact Payments, Pandemic Unemployment Compensation, and Pandemic Unemployment Assistance programs supported the incomes of millions of American individuals and families.

(Furceri, Loungani, Ostry, & Pizzuto, 2020) investigate previous pandemics and major epidemics over the past two decades and their impact on income inequality. Focusing on the SARS (2003), H1N1 (2009), MERS(2012), Ebola(2014), and Zika (2016) pandemics/epidemics and a sample of 175 countries impacted, the authors find that five years after each event, on average, both the market and net Gini coefficients, widely used measures of inequality , were above the pre-shock trends by approximately 0.75% and 1.25%, respectively, which the paper characterizes as quantitatively significant since Gini coefficients typically vary slowly over time. The authors cite these findings to support the premise that COVID-19 could intensify income inequality both within and across countries. Therefore, we postulate the following:

*H1: COVID-19 has likely resulted in an above average rise in poverty*

*H2: COVID-19 has likely resulted in an above average rise in income insecurity*

* 1. Health Safety

The COVID-19 pandemic has had a profound effect on general health care and have affected the way patients are using health-care facilities( Mercier, Grégoire et al., 2020). Numerous studies made since the onset of pandemic have noted a significant decline in hospital admission for elective and emergency procedures. Sharma et al. (2020) in their study noted a significant decline of 40-60% in acute myocardial infarction admissions in the United States during the onset of pandemic in March. Similar findings were echoed by Mesiner et al. (2020) in their study to evaluate the effect of nationwide lockdown in France on admissions to hospital for acute myocardial infarction.

Building on the above findings, we seek to investigate the same in our study. We undertake an assessment of the healthcare infrastructure availability for varying medical needs like adult, pediatric, ventilators and traumatic injury healthcare under the heightened strain exerted on medical infrastructure due to pandemic. Furthermore, recognizing the debilitating economic effects of COVID shutdowns, the study also assess access to healthcare facilities by analyzing enrollment in Healthcare-safety net programs such as Medicare and Children's Health Insurance Program (CHIP). Therefore, we postulate:

*H3: COVID-19 has likely resulted in an above average deterioration in health care security.*

* 1. Public Safety

To stop the spread of the COVID-19 virus statewide mandatory stay-at-home and business closures orders are being instituted across the United States. As a result of the ensuing economic turmoil and heightened unemployment rates, the lockdowns have had a cascading impact on public safety which were not originally planned. Recent Studies abound on both side of the spectrum wherein Jacoby et al., [2020](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7340780/#CR11) reported that weekly calls for service dropped “at least” 12% between February 2 and March 28, 2020 across 30 police agencies. On the other hand Ashby ([2020](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7340780/#CR1)) in his analysis on over a dozen U.S. cities of various sizes and compared crime rates across six serious crimes concluded the results about lowered crime rates as inconclusive. Current research investigates the same premise over a larger time-frame. Therefore, we postulate,

*H4: COVID-19 has likely resulted in an above average deterioration in public saftey.*

* 1. Education

[From R]

2.5 Mental Health

The COVID-19 healthcare pandemic has taken a severe toll on the physical health of human populace across the world. With over 1.5m deaths and millions hospitalized, the physical impact on human life has been unprecedented. To exacerbate the situation, non-pharmaceutical interventions (NPIs), although essential to halt transmission of the virus, have led to physical isolation, closure of schools (with untold effects on the development and wellbeing of children), widespread job losses and reduced public safety. Coagulation of these factors has been detrimental to the mental wellbeing of the population as noted by WHO (2020) in their survey across 130 member counties. Similar findings were echoed in a KFF Tracking Poll conducted in mid-July wherein 53% compared to 32% of adults in March in the United States reported that their mental health has been negatively impacted due to worry and stress over the coronavirus. Owing to the multipronged nature of COVID-19’s impact on physical, health social and economic wellbeing influencing the metal wellness of the society, the study aims to investigate the ensuing impacts. Hence, we hypothesize:

*H5: COVID-19 instituted deterioration in health care security has negatively impacted mental wellness.*

*H6: COVID-19 instituted deterioration in economic security has negatively impacted mental wellness.*

*H7: COVID-19 instituted deterioration in public safety has negatively impacted mental wellness.*

*H8: COVID-19 instituted deterioration in education enrollments has negatively impacted mental wellness.*

1. **Research Methods, Architecture, and Design**
   1. *Data*

The following datasets, sourced primarily from the NYC Open Data repository were used to support our analysis:

1. Daily count of NYC residents who tested positive for SARS-CoV-2, who were hospitalized with COVID-19, and deaths among COVID-19 patients. (COVID-19 Daily Counts of Cases, Hospitalizations, and Deaths, 2020)
2. Daily number of families and individuals residing in the Department of Homeless Services (DHS) shelter system and the daily number of families applying to the DHS shelter system since 2013 provided by Department of Homeless Services. (DHS Daily Report, 2020)
3. Total number of enrollees in Medicaid since [2016] provided by the Human Resource Administration.(Citywide HRA- Administered Medicaid Enrollees, 2020)
4. Quarterly report showing the number of individuals served meals in food pantries and soup kitchens since [2013] provided by the Human Resources Administration. (Emergency Food Assistance Program (Quarterly Report), 2020)
5. Seasonally adjusted employment data for New York City since 1990 provided by the Mayor's Office of Management & Budgets. (New York City Seasonally Adjusted Employment, 2020)
6. Breakdown of every arrest effected in NYC by the NYPD since 2006 provided by Police Department (NYPD). (Police Department (NYPD), 2020)
7. Daily listing of students enrolled, present, absent or released statistical count by district, borough and school provided by NYC Department of Education. (2018-2021 Daily Attendance by School, 2020)

|  |  |  |  |
| --- | --- | --- | --- |
| Indicator | Dataset | Source | Size(R\*C) |
| Healthcare Security | Medicaid Enrollees |  | 164\*6 |
| Chip Enrollees |  | 229386 \* 9 |
| Healthcare Infrastructure Availability |  | 283851\* 35 |
| Covid Cases | COVID Case-Hospitalization-Deaths |  | 269\* 39 |
| Public Saftey | Shooting Incidents |  | 21626 \* 19 |
| Arrest Counts |  | 5012956 \* 19 |
| Mental Health | Mental Health |  | 44\*12 |

* 1. *Methodology*

Our study adopted a variety of quantitative approaches to test for the hypothesized relationships between the pandemic and the aforementioned socioeconomic outcomes.

1. Excess Quantities

Borrowing from the concept of excess mortality used in epidemiology and public health to approximate the number of deaths from all causes during a crisis in excess of what would have been expected under normal circumstance, our study generalizes this model, applying it, where appropriate, to derive a measure known as the p-score to approximate the pandemics impact on enumerable socioeconomic measures. We define the p-score as follows:

1. Pearson Correlation Method

Within each socioeconomic segment, our study computed coefficient values for the following features:

|  |  |
| --- | --- |
| Poverty (homelessness) | CVD-19 cumulative cases, CVD-19 cumulative hospitalizations, CVD-19 cumulative deaths, number of adults in homeless shelters, number of children in homeless shelters |
| Poverty  (income insecurity) | CVD-19 cumulative cases, CVD-19 cumulative hospitalizations, CVD-19 cumulative deaths, total non-farm payroll |
| Health Care Security (Infrastructure Availability) | CVD-19 cumulative cases, CVD-19 cumulative hospitalizations, Nursing Home Beds Availability, Ventilator Beds Availability |
| Health Care Security (Healthcare Safety-Net Insurance Enrollments) | CVD-19 cumulative cases, Medicare Enrollments, Children's Health Insurance Program (CHIP) enrollments |
| Public Safety | CVD-19 cumulative cases, arrest counts, shooting incidents |
| Education | [From R] |

1. Linear Regression:

To establish causation, our study employed multi-variate linear regression models with COVID cumulative cases as the independent variable and the features as the dependent variables. Further, we also built a multiple linear regression to study the impact of COVID instituted shut-down’s economic, social, healthcare and public saftey turmoil on mental well being of human populace

|  |  |
| --- | --- |
| number of people employed, arrest counts, medicare and CHIPs enrollments, Nursing Home Beds Availability, Ventilator Beds Availability | Mental Wellbeing |

* 1. *Architecture/design*

|  |  |
| --- | --- |
| Data Quality Issues | Solution |
| Missing Data Points | For all but case, mean values of the respective were imputed in place of missing values. For Cases, 0 was imputed as COVID did not exist prior to March'20) |
| Number as Strings | Convereted to Float,Integer as required. |
| Column Headers | Column Headers were standardized to be in the format first\_second wherein any special characters were replace with "\_" for ease of usabilitiy |
| Data format | Standardized from varying formats to YYYY-MM-DD |

1. Extraction

Python’s data analysis and manipulation library was used to profile the raw datasets in our repository in order to extract metadata including schema semantics, data types, data statistics, and value distributions, thereby facilitating the discovery of the following data quality issues and adopted remedies:

* Duplicated rows

Remedy: deletion from dataset during data integration

* Null values for certain days in the time period of analysis

Remedy: replaced by appropriate mean value (see second bullet point in integration / aggregation)

* Inconsistent date formats

Remedy: converted all dates to the following format – “YYYY-MM-DD”

* Inconsistent formatting of column headers (e.g., new line characters, extra spaces, etc.)

Remedy: stiped all white space, new line characters and replaced with “\_”

* [others?]

[ ]

1. Integration / Aggregation

Our study primarily employed the Apache Spark SQL engine on NYU’s 48-node DUMBO cluster to perform schema matching, integration, and aggregation, thereby facilitating the following data transformations:

* Joining the COVID-19 datasets with each of the datasets used for our analysis, resulting in an intermediate dataset with the following schema:

where is one of the quantities of interest mentioned in [Table #]

* Creation of additional columns to compute the cumulative counts of cases, hospitalizations, and deaths as well as a running average of each feature over an appropriate period, as defined below, used to replace null values, resulting in the following schema

where is the 7-day running average for DHS data; [….]

1. Data analysis

* Time Series P-scores

Using the transformed datasets, our study again relied on the Apache Spark SQL engine to, for each feature, compute its average on the date in question going back as far as the data permitted but in no case more than 10 years as well as its p-score as defined in .

* Regression

[from AK]

1. Visualization

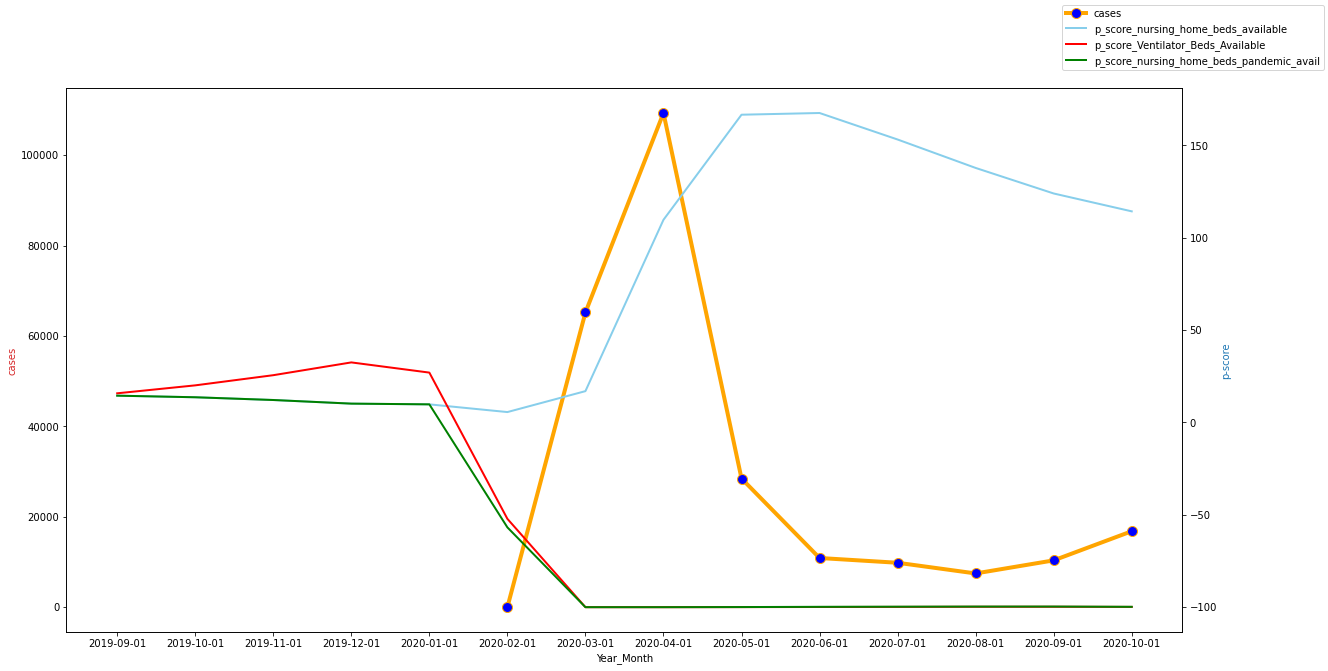
The results of our data analysis were visualized using Tableau software and Python’s Matplotlib modules.

1. Analysis and Findings

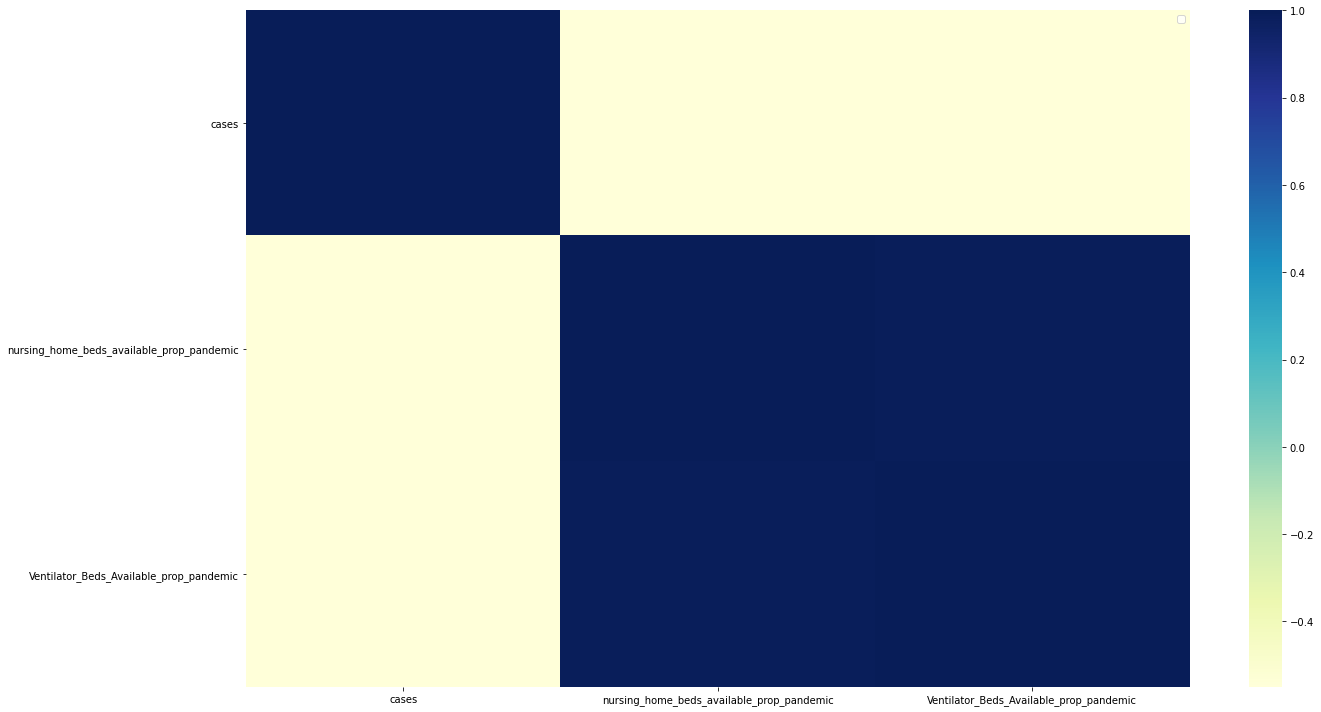
4.1 Impact of Covid-19 on healthcare safety

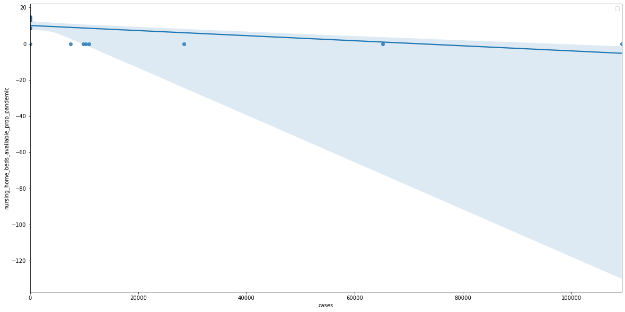
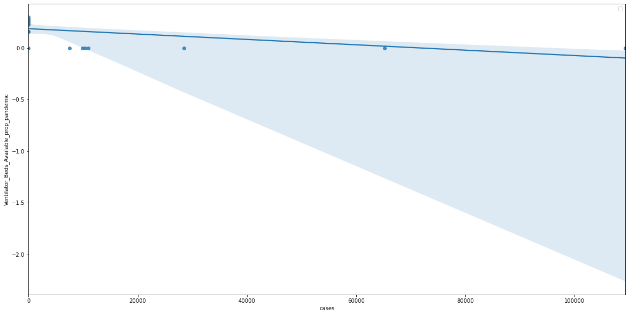
Healthcare Infrastructure Availability

A Timeseries graph of healthcare infrastructure availability over the period extending from September’19 to October’20 was plotted to understand the impact of COVID-19. Infrastructure availability was assessed by analyzing availability of nursing home infrastructure (being repurposed during pandemic) to treat COVID patients. The availability has been measured using the p-scores adjusted for number of cases to give a percentage of increase/decrease in the availability. As seen in the graph, there was an overall increase in the medical bed availability as nursing homes were incentivized by federal and state agencies to admit COVID-19 patients(NPR,2020). However, the repurposing was not found to be sufficient to keep up with the increased numbers of hospitalizations as cases have surged. The number of available ventilator and general care beds plummeted significantly as COVID-19 hospitalizations increased and stay low owing to the lingering impacts of the first wave.



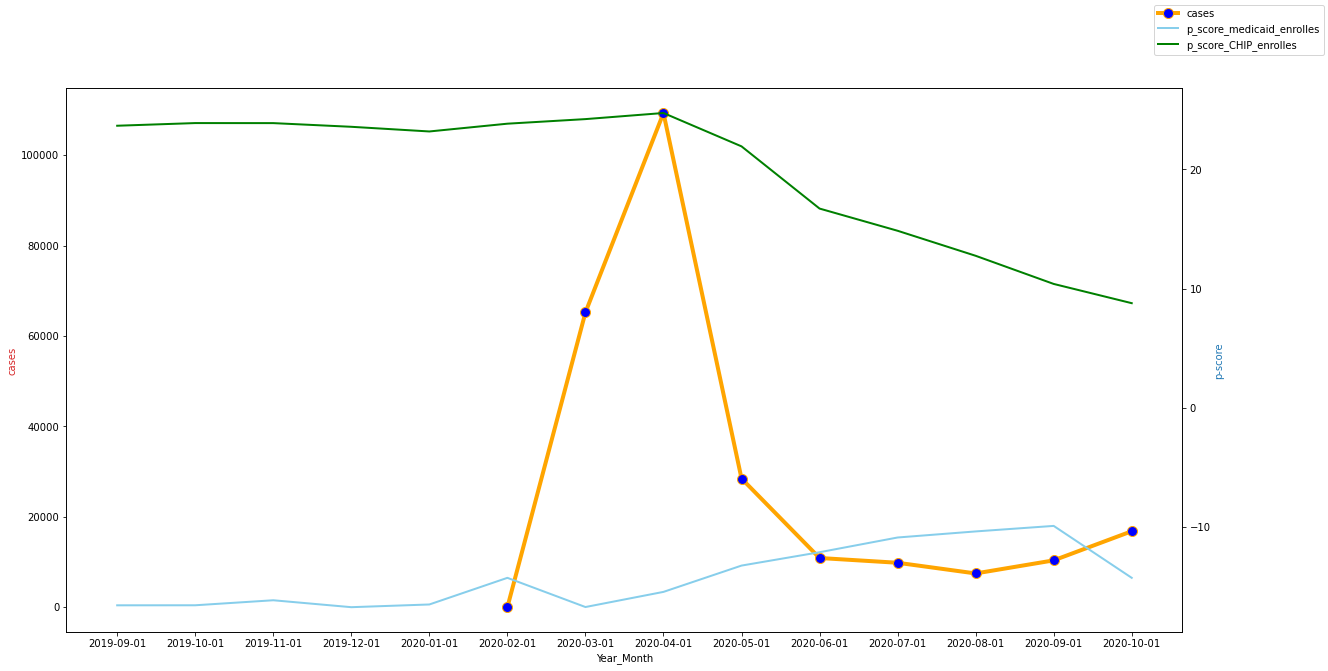
To quantify the visualized behavior we calculated correlation between cases and healthcare infrastructure availability and found the relationship to be negatively correlated. Furthermore, to establish a possible causation as COVID cases driving healthcare infrastructure unavailability we ran a multi-variate linear regression mode. Both indicators pediatric bed and nursing home bed availability were found to have statistically significant inverse relationship with COVID cases. The results attained above are in congruence with studies made by Mercier, Grégoire et al., (2020)



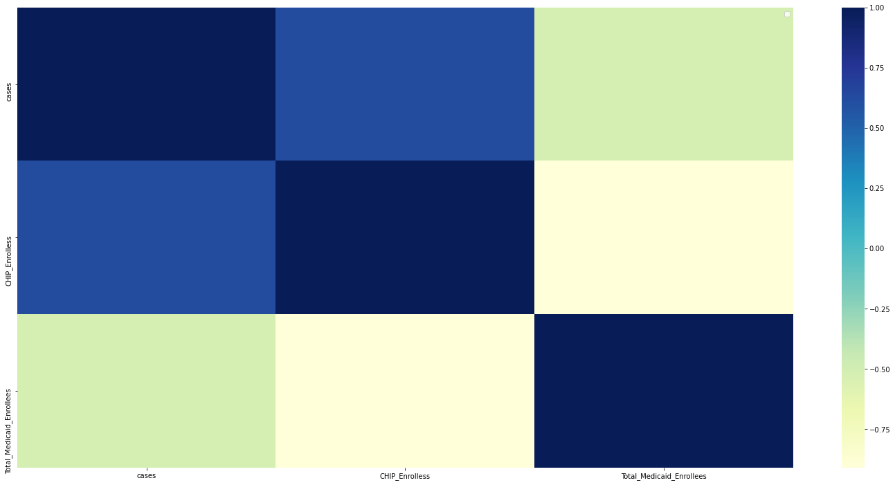
 

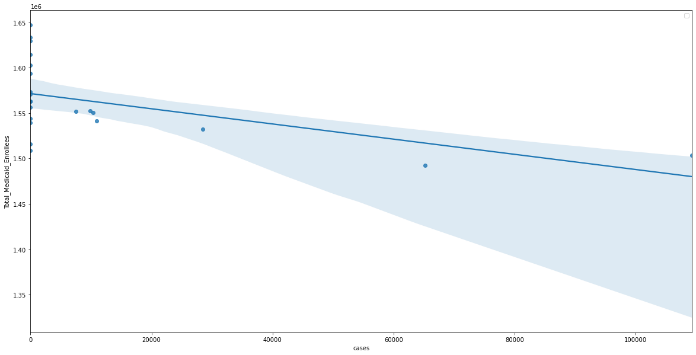
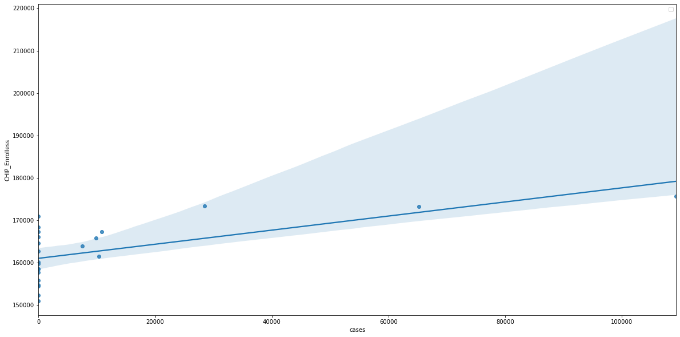
Healthcare Insurance Enrollments:

A Timeseries graph of the enrollment in health safety-net insurance schemes namely Medicare & CHIPSs over the period extending from September’19 to October’20 was plotted to understand the impact of COVID-19. Access to healthcare services a key indicator of Healthcare safety was assessed by analyzing the enrollments in Medicare and CHIPS programs built for financially vulnerable. The enrollments have been measured using the p-scores to give a percentage of increase/decrease in the enrollments for the corresponding month. As seen in the graph, there was an overall increase in the Medicare enrollments as COVID institute shutdowns led to job losses. CHIP enrollments at the same time decreased as schools were shut-down leading to increased dropout of the poorer students who would have otherwise qualified. Increased enrollment in the safety-net programs have long been a topic of debate as providing a lower standard of healthcare when compared to the pricier Employment based and private insurance schemes.



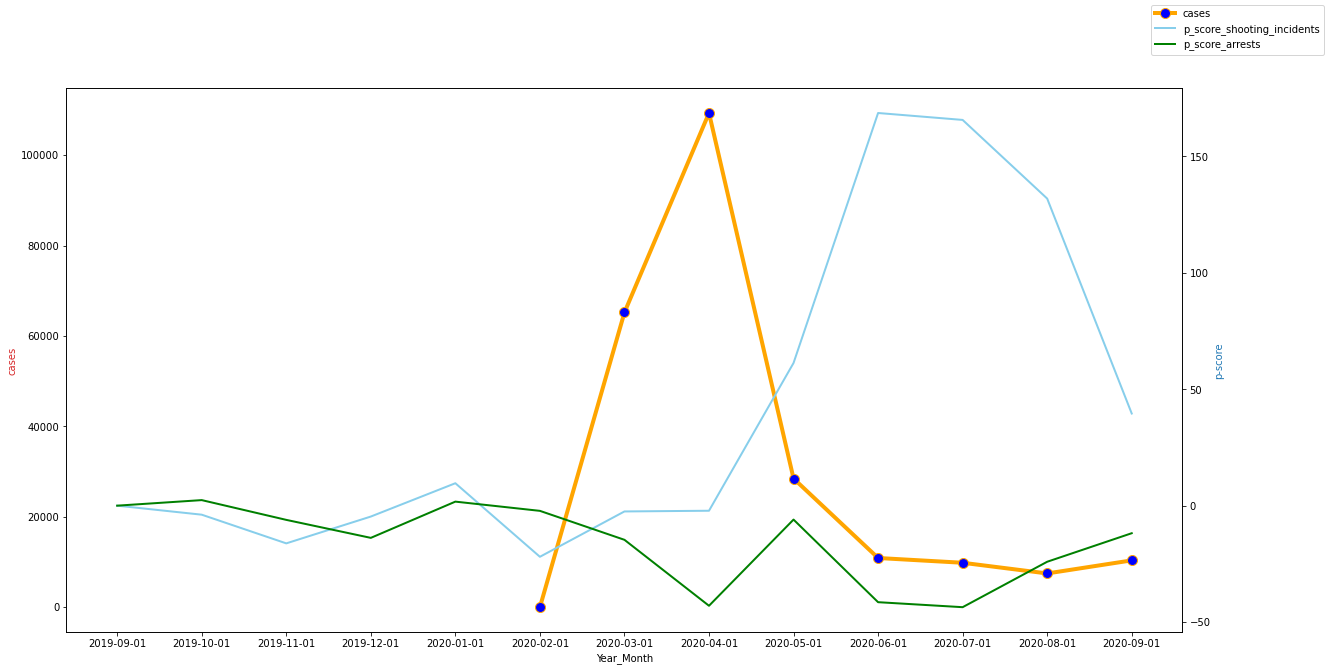
To quantify the visualized behavior we calculated correlation between COVID cases and healthcare insurance enrollments and found the relationship to be negatively correlated with Medicare enrollments. Medicare enrollments lagged initially during March-April as qualifying criterion required income under a set threshold for a month and then picked up gradually. Post the reopening as employees began returning to work there was a corresponding fall in the Medicare enrollments in October’20 (Blumenthal, Fowler, Abrams, Collins, 2020). On the other hand, CHIP enrollments were found to have a positive correlation which can be attributed to underlying data timeframe. CHIP enrollments were stable prior to pandemic and began declining in may along with COVID cases as poorer students were not able to keep up with remote learning requirements. Furthermore, to establish a possible causation we assessed the impact of COVID cases on saftey-net insurance enrollments by using multi-variate regression. We found Medicare to have a statistically significant inverse and CHIP enrollments to have a statistically significant direct relationship with COVID cases.

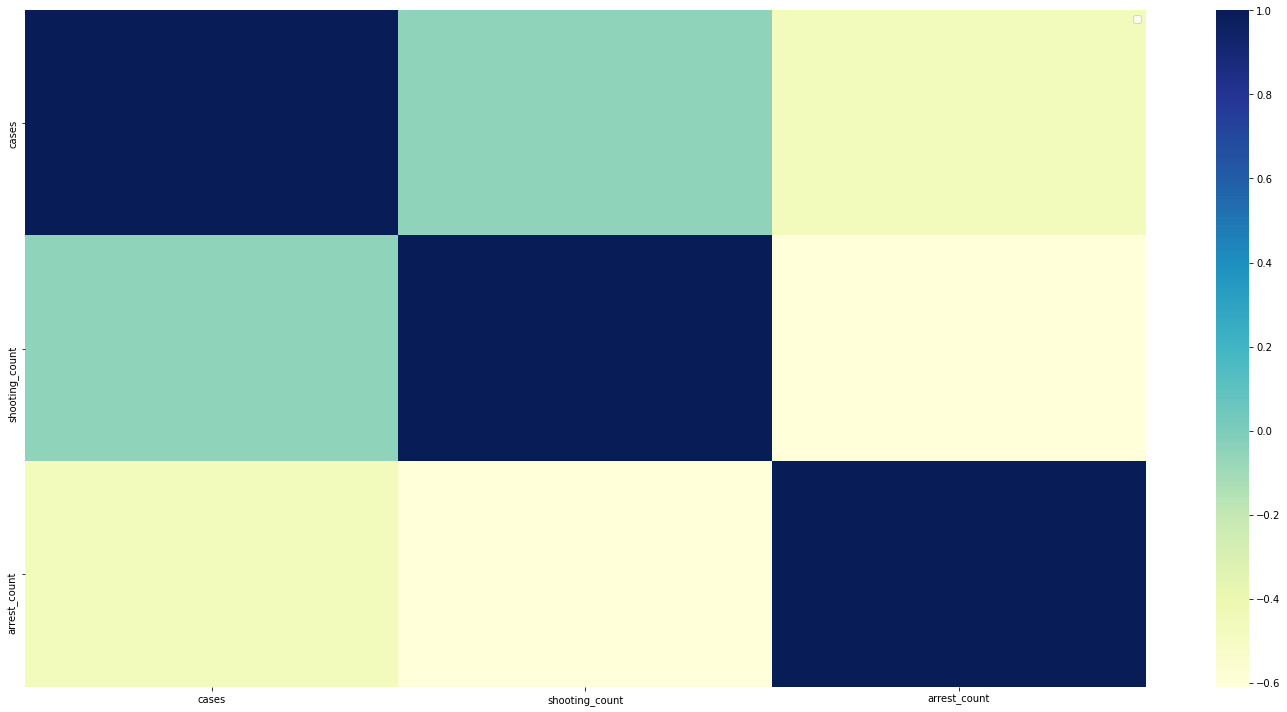


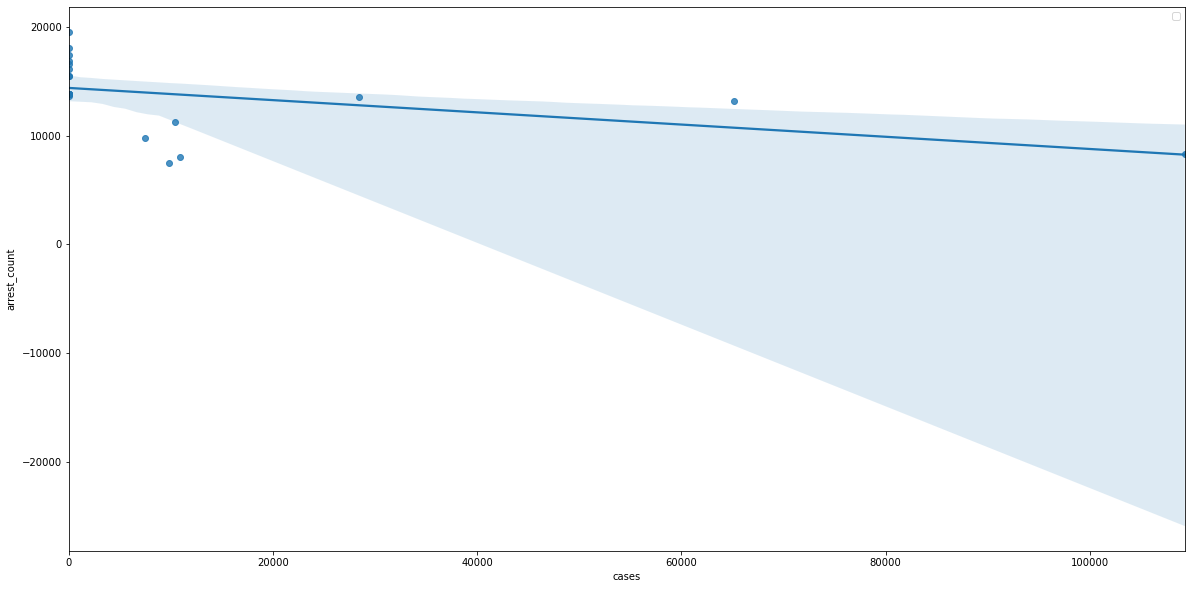
 

4.2 Impact of Covid-19 on public safety

A Timeseries graph of arrests and shooting incidents over the period extending from September’19 to October’20 was plotted to understand the impact of COVID-19 on public safety. Public safety was assessed by analyzing both the count of arrests made and shooting incidents. Both the metrics have been measured using the p-scores to give a percentage of increase/decrease in the occurrence. As seen in the graph, there was an overall increase in the shooting incidents during the period May-August and have been gradually coming down since july. The period corresponds with high rate of job losses resulting in heightened poverty and income insecurity. However, the same period also saw Black Live Matter (BLM) protests in wake of racial injustice leading to many violent protests. Arrests counts showed a fluctuating pattern through this period with more organized crime being driven down while protests and COVID inflicted financial insecurity driving the arrest counts up.



To quantify the visualized behavior we calculated correlation between COVID cases and public safety enrollments and found the relationship to be negatively correlated with arrest counts. Furthermore, to establish a possible causation we assessed the impact of COVID cases on both arrest and shooting incident counts and found only arrest count to have a statistically significant inverse relationship with COVID cases. Mixed results observed therefore, echo similar findings made by Ashby (2020) in his analysis and require future investigations with a broader indictor set to elucidate the impact.

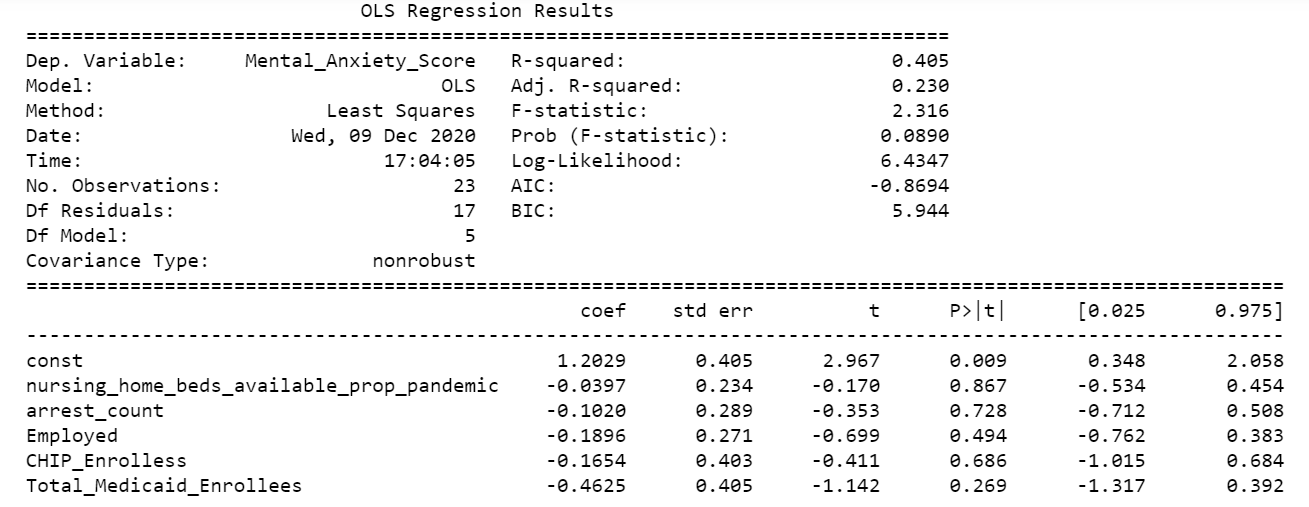




4.1 Impact of Covid-19 instituted effects on public safety, healthcare security, poverty, and income insecurity

To understand the impact of COVID and the ensuing non-pharmaceutical interventions on the mental wellbeing of the populace we modeled mental health as a function of economic, healthcare and public/physical wellbeing safety. As established in the previous sections COVID has had a significant impact on all of these aspects. We therefore ran a linear regression model to test the proposed causation. Prior to regression for any possible Multi-collinearity among the indicators by calculating for variance inflation factors (VIFs). Nursing home bed availability and Ventilator bed availability was found to have higher values of VIF owing to being drawn from same dataset. Therefore, ventilator beds availability is being removed for the model.

Model was subsequently ran 5 independent variables however, all the relationships was found to be statistically insignificant with mental health well being indicators. The results although insignificant need to be understood with caution as mental health is a taboo topic. There is an inherent tendency of participants to misrepresent their mental health (Panchal et al., 2020) and even society does not hold a positive view on the mental disorders. As a future workstream, an analysis of the “NYC Cause of Death” dataset can be undertaken to understand the trend of suicide cases per and post COVID. As of date, the data has been only updated until March’20.



# References

Covid Tracking Project. (2020, December 6). Retrieved from Covid Tracking Project: https://covidtracking.com

Department of Health and Mental Hygiene. (2020, November 30). *COVID-19 Daily Counts of Cases, Hospitalizations, and Deaths*. Retrieved from NYC Open Data: https://data.cityofnewyork.us/Health/COVID-19-Daily-Counts-of-Cases-Hospitalizations-an/rc75-m7u3

Department of Homeless Services (DHS). (2020, November 30). *DHS Daily Report*. Retrieved from NYC Open Data: https://data.cityofnewyork.us/Social-Services/DHS-Daily-Report/k46n-sa2m

Furceri, D., Loungani, P., Ostry, J. D., & Pizzuto, P. (2020, May 1). Will Covid-19 affect inequality? Evidence from past pandemics. *Covid Economics, Vetted and Real-Time Papers*, pp. 138-157.

Han, J., Meyer, B. D., & Sullivan, J. X. (August 2020). Income and Poverty in the COVID-19 Pandemic. *NBER Working Paper No. 27729*, JEL No. H53,I32,J65.

Human Resource Administration (HRA). (2020, November 27). *Citywide HRA- Administered Medicaid Enrollees*. Retrieved from NYC Open Data: https://data.cityofnewyork.us/Social-Services/Citywide-HRA-Administered-Medicaid-Enrollees/33db-aeds

Human Resources Administration (HRA). (2020, October 19). *Emergency Food Assistance Program (Quarterly Report)*. Retrieved from NYC Open Data: https://data.cityofnewyork.us/Social-Services/Emergency-Food-Assistance-Program-Quarterly-Report/mpqk-skis

Mayor's Office of Management & Budgets (OMB). (2020, November 20). *New York City Seasonally Adjusted Employment*. Retrieved from NYC Open Data: https://data.cityofnewyork.us/City-Government/New-York-City-Seasonally-Adjusted-Employment/5hjv-bjbv

NYC Department of Education. (2020, July 17). *2018-2021 Daily Attendance by School*. Retrieved from NYC Open Data: https://data.cityofnewyork.us/Education/2018-2021-Daily-Attendance-by-School/xc44-2jrh

NYC Office of the Comptroller. (2020). Retrieved from New York City Comptroller: https://comptroller.nyc.gov/

Police Department (NYPD). (2020, November 5). *NYPD Arrest Data*. Retrieved from NYC Open Data: https://data.cityofnewyork.us/Public-Safety/NYPD-Arrest-Data-Year-to-Date-/uip8-fykc

Reuters. (2020, September 23). *New York City mayor announced more furloughs to counter budget shortfall*. Retrieved from Reuters: https://www.reuters.com/article/us-new-york-budget/new-york-city-mayor-announces-more-furloughs-to-counter-budget-shortfall-idUSKCN26E2JM

Sharma M, Lioutas V, Madsen T, et alDecline in stroke alerts and hospitalisations during the COVID-19 pandemicStroke and Vascular Neurology 2020;svn-2020-000441. doi: 10.1136/svn-2020-000441

Hospital admissions for acute myocardial infarction before and after lockdown according to regional prevalence of COVID-19 and patient profile in France: a registry study

Mesnier, Jules et al. The Lancet Public Health, Volume 5, Issue 10, e536 - e542

#### Understanding the effects of COVID-19 on health care and systems

Mercier, Grégoire et al. The Lancet Public Health, Volume 5, Issue 10, e524

Ashby MPJ. Initial evidence on the relationship between the coronavirus pandemic and crime in the United States. Crime Science. 2020;9(6):1–16.

Jacoby, K., Stucka, M., & Phillips, K. (2020, April 16). Crime rates plummet amid the coronavirus pandemic, but not everyone is safer in their home. Retrieved May 28, 2020, from <https://www.usatoday.com/story/news/investigations/2020/04/04/coronavirus-crime-rates-drop-and-domestic-violence-spikes/2939120001/>

**The intersection of COVID-19 and mental health**

The Lancet Infectious Diseases,

The Lancet Infectious Diseases, Volume 20, Issue 11, 1217

WHO 2020. The impact of COVID-19 on mental, neurological and substance use services

Accessed on https://www.who.int/publications/i/item/978924012455

Kaiser Family Foundation, Panchal, N., et al. (2020). The Implications of COVID-19 for Mental Health and Substance Use.

# Nursing Home Residents Moved Out To Make Way For COVID-19 Patients

Accessed on 12/08/20 using https://www.npr.org/sections/coronavirus-live-updates/2020/08/04/897239846/nursing-home-residents-moved-out-to-make-way-for-covid-19-patients

Blumenthal D, Fowler EJ, Abrams M, Collins SR. Covid-19 - Implications for the Health Care System. N Engl J Med. 2020 Oct 8;383(15):1483-1488. doi: 10.1056/NEJMsb2021088. Epub 2020 Jul 22. Erratum in: N Engl J Med. 2020 Jul 23;: PMID: 32706956.