**[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**

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.

* 1. Health / Public Safety

[From AK]

* 1. Education

[From R]

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)
   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 | [From AK] |
| Public Safety | [From AK] |
| Education | [From R] |

Additionally, our study employed linear regression models with COVID cumulative cases as the independent variable and the aforementioned features as the dependent variables.

[More to come from AK on regression analysis on non-evident impact of COVID]

* 1. *Architecture/design*

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

* 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.

# 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