Project Team: Rakshita (rr3653), Abhishek Kalra (ak7468), Fola Tubi (ft718)

Project Choice	Data Analysis : Understanding the Impact of the COVID-19 Pandemic
Previous Work and References	1. Martin, A., Markhvida, M., Hallegatte, S. et al. Socio-Economic Impacts of COVID-19 on Household Consumption and Poverty. EconDisCliCha 4, 453–479 (2020). https://doi.org/10.1007/s41885-020-00070-3 2. Osofsky, J. D., Osofsky, H. J., & Mamon, L. Y. (2020). Psychological and social impact of COVID-19. Psychological Trauma: Theory, Research, Practice, and Policy, 12(5), 468-469. http://dx.doi.org/10.1037/tra0000656 3. D Furceri, P Loungani, J Ostry, P Pizzuto Will covid-19 affect inequality? evidence from past pandemics. Covid Economics, volume 12, p. 138 – 57 Posted: 2020 4. Don Bambino Geno Tai, Aditya Shah, Chyke A Doubeni, Irene G Sia, Mark L Wieland, The Disproportionate Impact of COVID-19 on Racial and Ethnic Minorities in the United States, Clinical Infectious Diseases, , ciaa815, https://doi.org/10.1093/cid/ciaa815 5. Snyder, B. F., & Parks, V. (2020). Spatial variation in socio-ecological vulnerability to COVID-19 in the contiguous United States. Health & place, 66, 102471.
Problem Description and Goal:	COVID-19 has infected and killed over 12m and 250k people, respectively in the U.S., with New York alone accounting for ~13% of deaths, the highest in the country. Beyond the public health emergency, public officials will need to come to terms with the unprecedented strain on the city's social safety net. Using NYC Open Data repository, our goal is to quantify how COVID-19 has impacted socioeconomic outcomes in NYC and what that might imply for fiscal policy moving forward. We will seek to investigate the following hypotheses: 1. COVID-19 has likely resulted in an above average rise in poverty 2. COVID-19 has likely resulted in an above average deterioration in health care security 3. COVID-19 has likely resulted in an above average deterioration in public safety 4. COVID-19 has likely resulted in an above average rise in income insecurity 5. COVID-19 has likely impeded student learning
Relevant Datasets	COVID-19 Daily Counts of Cases, Hospitalizations, and Deaths DHS Daily Report Citywide HRA- Administered Medicaid Enrollees Emergency Food Assistance Program New York City Seasonally Adjusted Employment NYPD Arrests Data 2018-2021 Daily Attendance by School M/WBE, LBE, and EBE Certified Business List Legally Operating Businesses
Method/approach	Our approach entails using the map-reduce/Apache Spark framework to process the relevant datasets with a view to identifying "excess" quantities, above and beyond what we would have expected to see under "normal" conditions for the various socioeconomic indicators identified above.
Evaluation Criteria	• Let X denote some observed quantity. Therefore, we will evaluate the stated hypotheses "excess" quantities using P-scores, defined as follows: $$\frac{X - E(X)}{E(X)} \; where \; E(X) \; is \; the \; expected \; value \; of \; X$$ • $E(X)$ will be approximated using historical averages (e.g. 10-year average) • Where appropriate, we will also derive Pearson's Coefficient values to measure the correlation between COVID-19 case counts and each indicator's P-score. Visualizations: a) Time series plots to be generated to visualize the impact of COVID on the respective metrics identified. b) Scatter plots to be generated to understand/identify the impact of COVID on the socio-economic parameters .
Weekly Schedule	See milestones.txt.