



# Study of Small-Medium Business & Crime during COVID-19

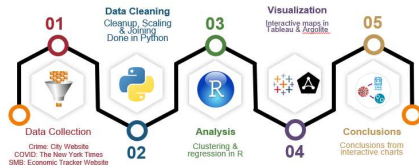
Antonio De Los Heros, Muhammad Ibraheem, Muhammad Kaleem Khan, Chandra Narasimhan, Raghvendra Trivedi, John Walker

## Introduction-Motivation

There is limited research done to study the interdependencies of economic shift and crime during global pandemics. This study will not only cover many insights, but it will also help users understand those insights using interactive data visualization. This study will help government agencies plan better for future pandemics and help many small-medium business owners in future challenges. It will help residents of these areas see benefits by the uncovered truths and realities of the COVID-19 situation for a given period.

## Approach

The benefit of this project is the ability to not only gain insights from each individual subject, however also looking at the interactions and correlations that exist amongst them; this is believed to be a true differentiator comparing to other in-depth analysis that have been done on crime, COVID-19, and small-medium businesses. Our proposed method ranges from data source collection, data clean up and integration, to performing a comprehensive analysis that provides us the insights and correlations between the data sources; we also provide an exploratory data analysis that will look to guide the broader dashboard concepts containing the most useful views.



## Datasets

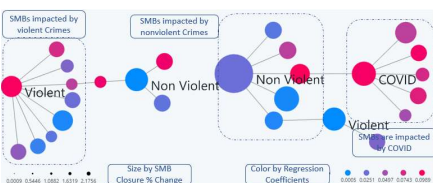
The Project consists of three data sources that have been gathered from different sites/APIs, crime, COVID-19, SMBs. Crime data is collected from each city's website in csv, raw data size was 56MB after processing it reduced to 45,693 rows. COVID-19 data is sourced from The New York Times in csv it has 6.5M rows & 280MB in size. SMB data sourced from Womply based on transactions, it has 55 files & 122MB in size.

## Experiments & Results

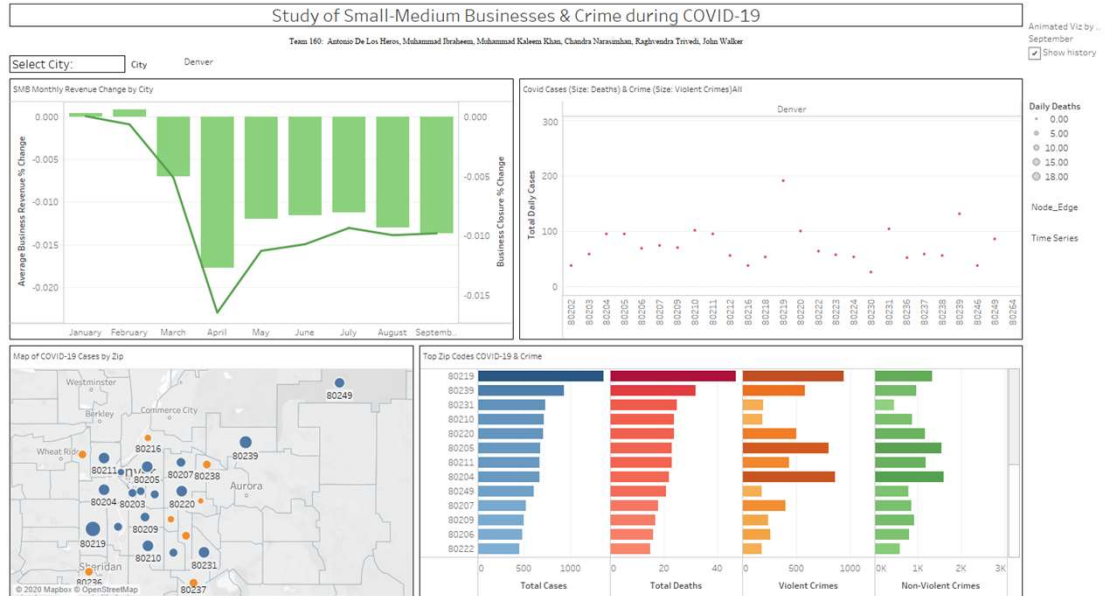
we sought to answer five questions with the results of our analysis and visualizations.

What are the correlations due to the interactions of COVID-19, crime, and SMB?

To answer this question, we turned to our linear regression analysis of significant variables and resulting node-edge diagram. Based on this analysis, we found that out of 161 zip codes:

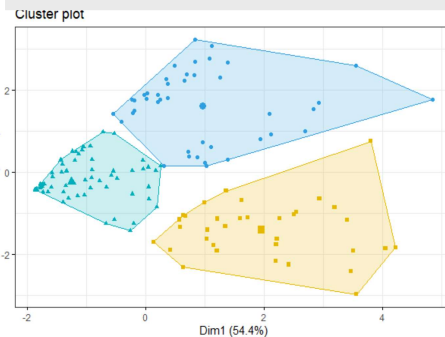


- 59 zip codes show no relationship
- 22 zip codes have less than 0.1% change
- 44 zip codes show strong relationship between SMB % change and COVID-19
- 37 zip codes show a strong relationship between SMB % change and crime
- 26 zip codes show a strong relationship between COVID-19 and crime



## Data Analysis & Visualization

Data cleanup and compilation is done in python, crime categories are standardized, data joined at zip code level, SMB data scaled at zip code level. Timeseries graph is created by converting all dataset in to timeseries. K-means clustering is done to find similar zip code where COVID, SMB and crime relationships are closely related this is used to generate zip code map. Regression is used to find most important factors in these relations and is used to generate network edge graph. A Tableau dashboard was created to link the visualizations mentioned above, incorporate additional visualizations that help in understanding these factors, and provide the user with one a central location to access all the information



slight increase in aggravated assault. As the COVID-19 cases peaked, so did the counts of three crimes (larceny, drug offense, and aggravated assault), however we noticed decreasing trend in fraud in the same time period

Does the number of violent crimes correlate with the number of COVID-19 cases in each ZIP code? Analyzing the "Top Zip Codes COVID-19 and Crime" chart in the Tableau dashboard, we found an interesting observation in terms of COVID-19 and crime. Zip codes with the highest number of cases typically have the most amount of deaths, however, when we compare COVID-19 and incidents of violent and non-violent crime, we can identify that the zip codes containing the most COVID-19 cases are not the zip codes with the highest crimes.



## Conclusions

In this project, we have combined large datasets from multiple sources, leveraged visualization tools such as Tableau and Argo Lite, and applied machine learning algorithms such as linear regression and k-means clustering to help us derive insights from how COVID-19 has impacted crime and SMB performance.

What type of impact has COVID-19 had on crime?

As per available data, we have found four types of crime which have shown movement during COVID-19 period from January 2020 to September 2020. From June to August, we see increasing trends in larceny, drug offenses, and a