DATA512 Part 4 - Project Report

Impact of COVID-19 on Socioeconomic Factors (Maricopa County, AZ)

Rohit Lokwani

December 12th, 2022

1. Introduction

COVID-19 has had a massive impact on the world in the last couple of years. The rapid spread of COVID-19 endangered lives, disrupted livelihoods, and impacted global trade, the economy, and businesses[1]. The world economy had started to experience significant disruptions and was and is still moving toward a severe recession and an unprecedented economic crisis. All nations have experienced challenges as a result of COVID-19, but many nations have had to deal with a more difficult situation because of their large populations, subpar health services, high rates of poverty[2], poor socioeconomic conditions, and inadequate social protection systems. According to Dr. William Cockerham, the Chief Sociologist at the University of Alberta, he says "Socioeconomic status is the strongest indicator of health, disease resistance and longevity in medical sociology". We try to corroborate his theory in our analysis.

The UN's Framework for the Immediate Socio-Economic Response to the COVID-19 Crisis warned that "The COVID-19 pandemic is far more than a health crisis: it is affecting societies and economies at their core. While the impact of the pandemic will vary from country to country, it will most likely increase poverty and inequalities on a global scale." (COVID-19 Socioeconomic Impact[9]). To educate and customize government and partner responses to the COVID-19 issue and ensure that no one is left behind in this effort, it is essential to assess the implications of the crisis on communities, economies, and vulnerable people. These factors can directly affect the ability to earn, the representation of race or caste in society, education, etc.

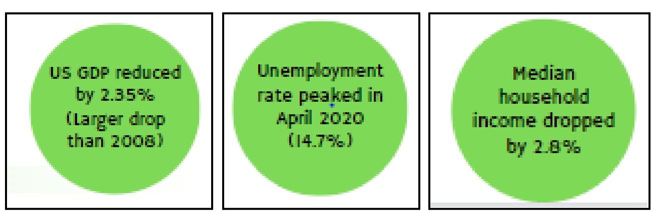

Figure 1.1: US Economic Statistics from Google

Hence, how did the pandemic impact such factors interest us the most? For this analysis, we target the socioeconomic factors primarily but focus mainly on economic ones. We try to do a fine-grained analysis of the socioeconomic implications of COVID-19 in Maricopa County, Arizona in the United States. We understand the implications at the individual level and industry level. The broader level question that this analysis tries to answer is:

What were the immediate and gradual impacts of COVID-19 on socioeconomic factors at the community and industrial level in Maricopa County in Arizona?

With the current macroeconomic situation, we know about the tech layoffs happening around. It is important to know what are the most vulnerable industries that could be facing layoffs. Also, it is important to understand the social impact in terms of education as this might entail the future of over a billion students (Figure 1.2). We do a broader study about different unemployment factors and social factors.

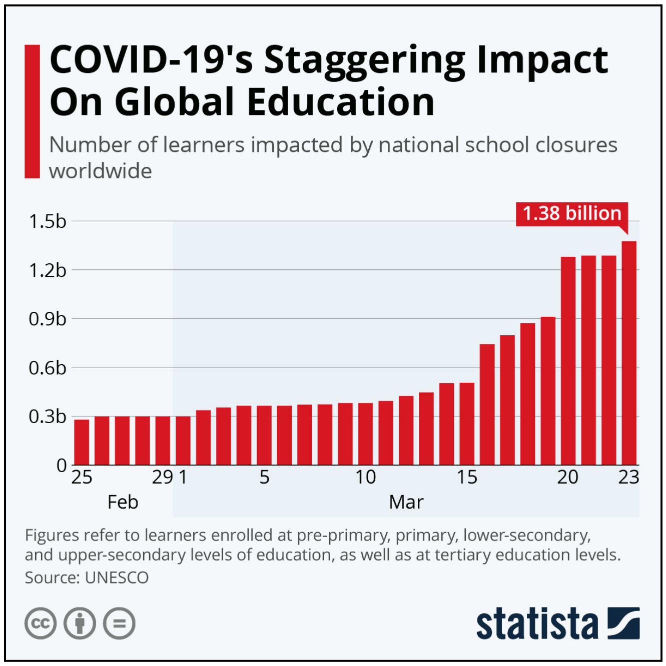

Figure 1.2: Impact of COVID-19 on Global Education from [World Economic Forum]

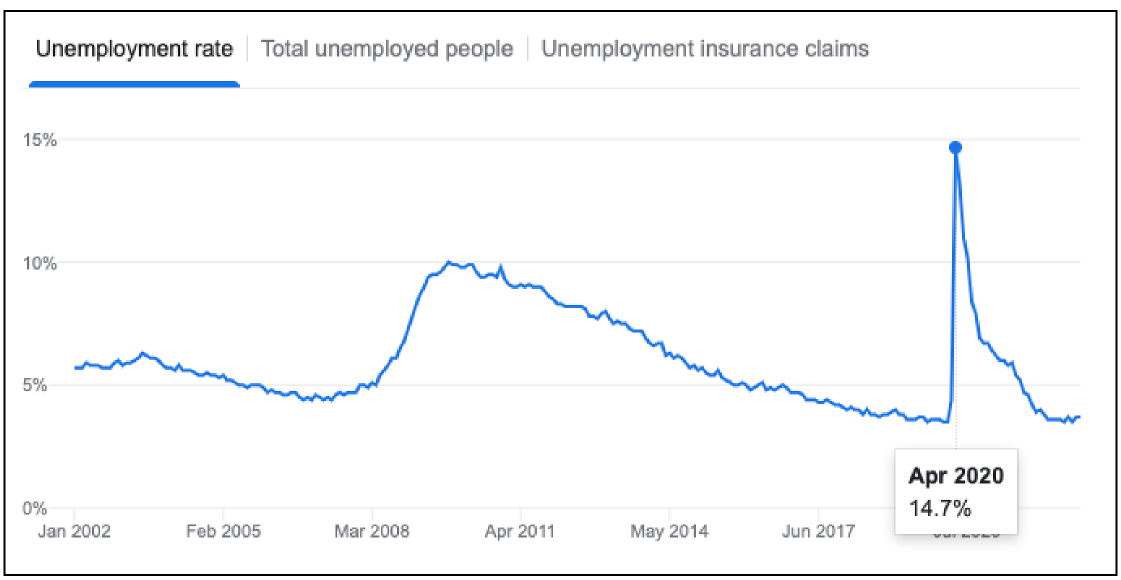

Figure 1.3: US Poverty Statistics from Google (Updated December 10th, 2022)

It is also important to measure the overall community risk of the specific county. This helps in prioritizing measures for the county. The end goal of this work is to be better prepared for such pandemics/epidemics in the future, hence it would be classified as a real problem over anything else.

2. Background/Related Work

This section covers our background research for this analysis project. Various papers and articles have covered this topic from different perspectives. We will be covering the most relevant ones that impacted our understanding of the problem at hand and informed us of the approach that we took to approach this project.

To estimate the impact of COVID-19 on poverty in Indonesia, Suryahadi et al. (2019) conducted simulations based on various economic growth scenarios. Using correlation analysis on income and expenditure they find that under the mildest COVID-19 impact on economic growth, the poverty rate will increase from 9.2 percent in September 2019 to 9.7 percent by the end of 2020. This implies that 1.3 million more people will be pushed into poverty[2]. Their analysis is limited to calculating the correlation between Income and Expenditures without considering the other community-level factors. Rasul et al.[5] study the economic impact on South Asia stating how because of its high labor intensity, agriculture-based rural economy and livelihoods are disrupted by COVID-19 and resultant quarantine, restrictions on the movement of goods and services, and closure of cross-border trade. Working on a more commercial US-based economy, we hypothesize that we could observe similar patterns in different industries and hence try to establish the impact of COVID-19 cases on these industries. According to Ramaswamy et al. (KPMG) [7], In India, around 250 million students were affected due to school closures at the onset of the lockdown induced by COVID-19. The pandemic posed several challenges in public and private schools which included an expected rise in dropouts, learning losses, and an increase in the digital divide. We test a variation of this hypothesis with our research question on the influence of the pandemic on education by analyzing if there was any change in the percentage of students graduating from college.

Coming to the usage of algorithms to a more human-centered approach, Baumer et. al. inform us of how having interpretable algorithms become important to understand the analysis of laypersons. It also informs us of design strategies like speculative, participatory, and theoretical[6]. We follow the later principles in this project and also use algorithms that are as interpretable as possible to aid user understanding. Based on this paper we also do not use a t-test because that could be a problem with normality and independence assumptions for our time-series use case. Andrew et. al. [3] use a SIR model which is a Markov model of the spread of an epidemic in a population in which the total population is divided into categories of being susceptible to the disease (S), actively infected with the disease (I), and recovered (or dead) and no longer contagious (R). How an epidemic plays out over time is determined by the transition rates between these three states. This model allows for quantitative statements regarding the tradeoff between the severity and timing of suppression of the disease through social distancing and the progression of the disease in the population. Although our analysis isn't directly related to the spread of COVID-19, we use masking policies to determine the COVID-19 spread and use better quantifiable and interpretable statistical models covering wider variables to reduce bias in our model. According to Flanagan et. al.[8], Social vulnerability refers to the socioeconomic and demographic factors that affect the resilience of communities. Studies have shown that in disaster events the socially vulnerable are more likely to be adversely affected, i.e. they are less likely to recover and more likely to die. Effectively addressing social vulnerability decreases both human suffering and the economic loss related to providing social services and public assistance after a disaster. We use a variant of this metric available in our dataset called the Overall Community Risk index and develop a prediction model. This paper helps us frame our 5th research question.

In summary, the five main research questions that we answer using this analysis are:

1. What was the effect of masking policies on the COVID-19 case count?
2. What was the influence of the pandemic on the unemployment rate and Civil Labor Force in the county?
3. How did the socioeconomic factors (Education, Median Household Income, and Gross Domestic Product) change during the early stage of the pandemic?
4. How did the COVID-19 cases affect the economic indexes of the county in different industries?
5. What are the other factors that help measure the overall community risk for similar epidemics/pandemics in the future?

3. Methodology

This section describes the end-to-end methods involved right from gathering, the tools, the analytical or statistical methods we foresee, and the presentation of findings.

3.1 Analysis Tools

Jupyter Notebook is used to script the analysis steps using Python 3. While Python data analysis libraries such as NumPy, Pandas, and Sklearn are used for data preprocessing, integration, wrangling, and executing correlations, Python's plotting libraries such as Matplotlib and Seaborn are used to plot the results.

3.2 Data Gathering and Preprocessing

In all the datasets mentioned above, the datasets are quite clean in general. All of them are structured files stored as CSVs or Excel spreadsheets. Since, for answering all the questions, our primary data would be the timestamp and the related features from our sparse dataset. By the looks of it, the data looks clean and quite well formatted, hence no preprocessing is required apart from changing column names for understandability. To create a consolidated dataset that is ready for analysis, the integration stage entails matching entries from all the processed datasets using state and county variables as unique keys. State and county names are included in each dataset. To map records from other datasets, these two fields serve as the primary keys. Some of the data is structured as monthly or in other cases as annual intervals. We adjusted our COVID-19 case data to these intervals in order to answer our questions. To be specific, the data is adjusted(rolling average) to monthly levels for answering questions 1,3 and annually for question 2. For answering question 4, we use all the counties in the US to develop a more representative and robust model.

3.3 Analysis Methods

To answer our research questions, we mainly used correlation, and time-lagged cross-correlations to compare two sets of interesting features. We also used bar charts and other basic plots for Exploratory Data Analysis. We used Linear Regression to answer our question 4.

3.3.1 Correlation Tests

Correlation analysis is a statistical method that gives insights into the existence of connections between quantitative variables and provides a metric to infer the strength of such relationships (Link). Hence, we use it to find the correlation between the economic index, and value added by the sector to the percentage change in employment. We use Pearson's correlation coefficient (Linear Relationship) and Spearman's Correlation Coefficient (Less stringent, looks for monotonicity) to determine that. We find Pearson's correlation works the best for us. We also plan to determine the relationship between different industrial sectors and present these as a heatmap.

3.3.2 Time-Lagged Cross-correlation

This is a method to determine if there's a correlation between two-time series with a lag of time interval. This usually is helpful when you hypothesize that the patterns in one time series occur after another instead of at the same time interval. We used this technique to determine the impact of COVID-19 on unemployment and Civil Labor Force participation as it is expected to occur with a time difference. We also use this for finding the correlation between different industries and their COVID-19 cases, which tells us about how an increase in COVID-19 cases could result in a reduction in force activities in these industries.

3.3.3 Linear Regression

In statistics, linear regression is a linear approach for modeling the relationship between a scalar response and one or more explanatory variables (also known as dependent and independent variables). The case of one explanatory variable is called simple linear regression; for more than one, the process is called multiple linear regression. We use multiple linear regression primarily to predict the overall community risk index from different factors like socioeconomic index, and economic index amongst others. This essentially improves the predictive power of our analysis and helps in measures curtailing the spread of future epidemics. The choice of algorithm is determined by the interpretability it offers, thus following human-centered algorithmic design principles.

3.3.4 Change Point Detection

In statistical analysis, change detection or change point detection tries to identify times when the probability distribution of a stochastic process or time series changes. In general, the problem concerns both detecting whether or not a change has occurred, or whether several changes might have occurred, and identifying the times of any such changes. We used the ruptures library in Python for doing this and then verified against the masking policy dates. We used this technique to answer our research question 1 for determining if the change points due to masking or vaccination policies had an impact on COVID-19 cases.

3.3.5 Exploratory Data Analysis

We do an exploratory data analysis of the entire data set to understand the data better. We also use this analysis without any specific statistical methods and draw conclusions for research question 2 from visualizing graphs like line graphs and bar plots. Using line charts, helps us study the trends in change of these variables/metrics.

3.3.6 Human-centeredness of the design

We try to preserve the human-centeredness of these analyses using a participatory design when it came to the masking survey. The algorithms that we use are considered human-centered in the sense that we built models which are explainable and interpretable. This makes the analysis more transparent and open. We also used representative data across the US for building the community risk prediction model. This helps us get rid of any underlying biases and prejudices related to specific communities and promote fairness for algorithm design.

4. Findings

This section explains the results/findings of our analysis.

4.1 Common analysis (Impact of Masking policies on COVID-19 cases)

The visualization (figure 4.1) shows the change points in the time series data for changes in daily confirmed cases data in Maricopa county in Arizona state in the United States. The X-axis represents the day of the data point. The Y-axis represents the daily infection rate change. The X-axis essentially did not require any pre-processing as the data available seemed complete and accurate. The Y-axis was derived using the confirmed cases data provided by John Hopkins Hospital. We took this value, then calculated the daily infection rate change by calculating the change/slope over the cumulative cases given by the data. The colors of the data points show the national-level masking policy by the CDC (as indicated in the legends) and the blue vertical lines indicate the change points calculated using the Pelt Search method. We use the CDC guidelines as the state mask mandates dates are missing and a masking survey shows 89% of people wear masks more than frequently. The best way to read the graph for the viewer is to go from left to right and see how a change in masking policy was causing a change point in the data after almost 1 month in at least 2 instances. The graph is also indicative of the fact that the CDC changed the policies to less strict ones when these infection rate changes were negative and more strict ones when the other way round.

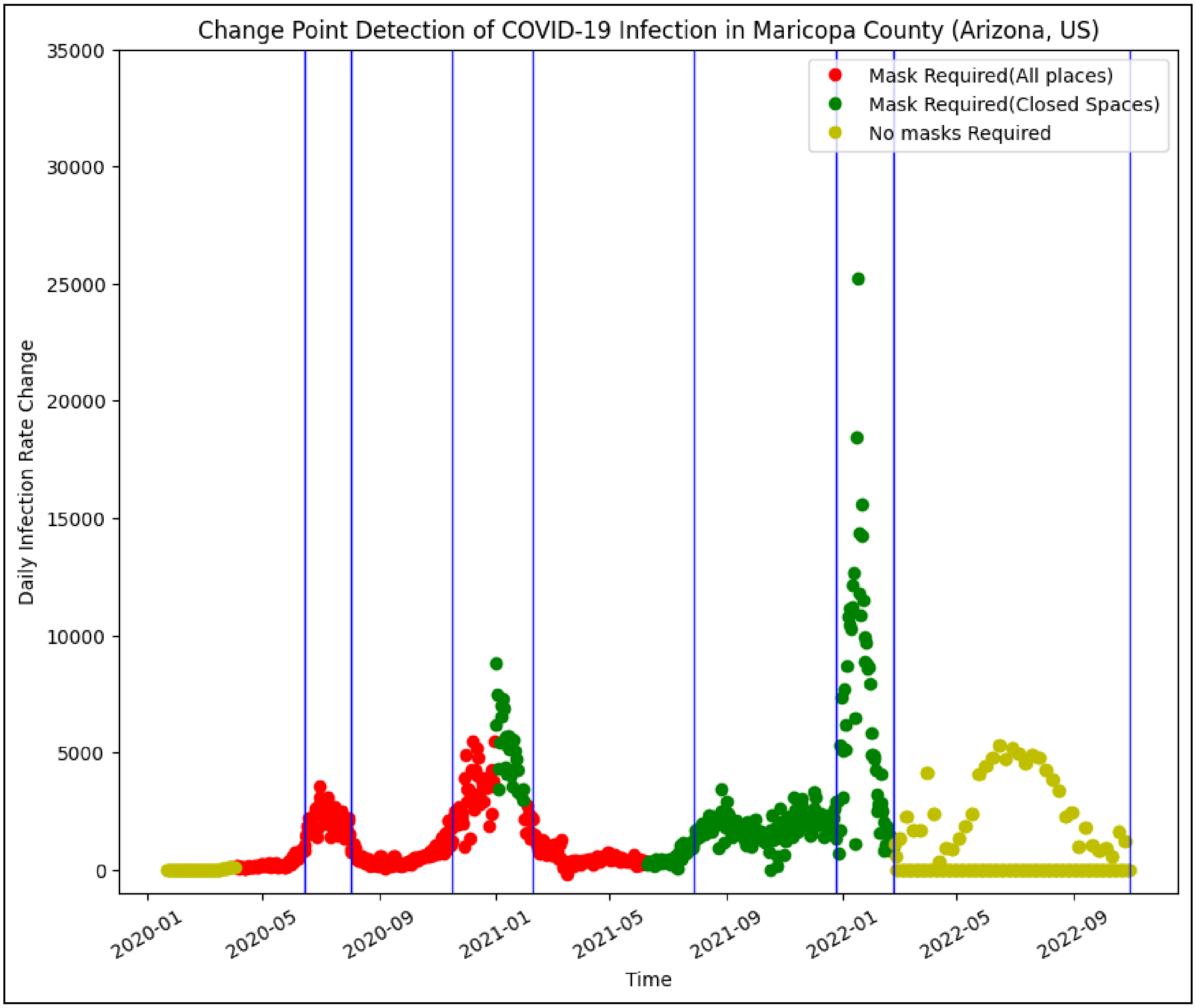

Figure 4.1: Impact of masking policies on COVID-19 cases

Overall, we see that the mask policies were changed during the increase in cases or when they just started to drop. The masking policies did actually help show some difference in daily infection rate reduction with almost a month's delay. One important thing to note here, is we do not consider hospitalizations, herd immunity, recovery, or other implicit factors in our analysis. One interesting finding is that after vaccines (Late December 2020) were available, the CDC changed to a closed-space masking policy but still saw an increase in cases reason being people traveling across states during Christmas and New Year. But as we see the 4 months following that had stricter masking policies but with more people getting vaccinated the infection rates dropped and masking policies were made less strict. In summary, if we stick to the timeframe of our research question from February 1, 2020, to 1st October 2021, we see that masking had some impact on daily infection rates with a 1-month delay. This delay includes infection, COVID-19 reports, and case count actually fed-to-database delay.

4.2 Impact of COVID-19 cases on unemployment and Civil Labor Force

We did a direct correlation of COVID-19 cases against the unemployment rate and civilian labor force. But did not find any direct correlation. For a time-lagged correlation, we see some moderate-to-high correlation between unemployment (-0.57) and the civilian labor force (0.67) against COVID-19 cases. Labor force participation is important as it tells you about the employed and unemployed people in the general populace apart from people with privileged jobs. Overall, these findings inform us of the gradual impact of COVID-19 on people's jobs and willingness to work.

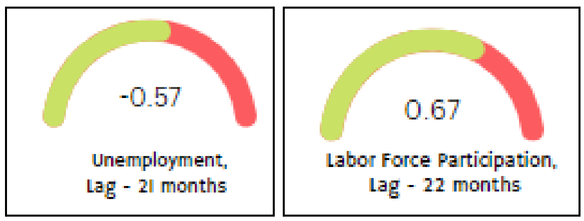

Figure 4.2: Correlation of unemployment and labor force participation against COVID-19

4.3 Immediate Impact of COVID-19

4.3.1 Education (Percentage of people with degrees)

We use the data from 2008 to 2020 for this analysis. We see that there was no particular change in the increasing trend of the percentage of degrees obtained by the people before the pandemic and at the start of the pandemic. One of the reasons could be that these degrees are mostly more than 1-year degrees and the graduation dates are usually over the summer (June). Hence, assuming COVID-19 started to hit at the start of January, 6 months Is a very short time to gauge any impact, this can be seen as the limitation of the data. But as an immediate impact of COVID-19, there wasn't a case of people dropping out immediately at the first sight of COVID-19.

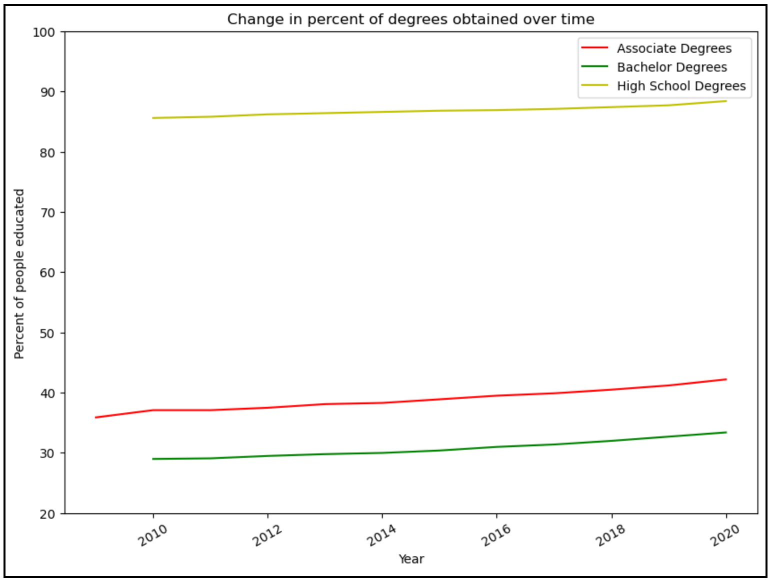

Figure 4.3 Line chart for change in percent of degrees obtained over time

4.3.2 Change in GDP

We see that the slope changes quite a bit from 2019-2020 as compared to the timeline previous to that. This shows that COVID-19 had started impacting the GDP in some manner, at least showing some slowdown in the economy if not showing a downturn.

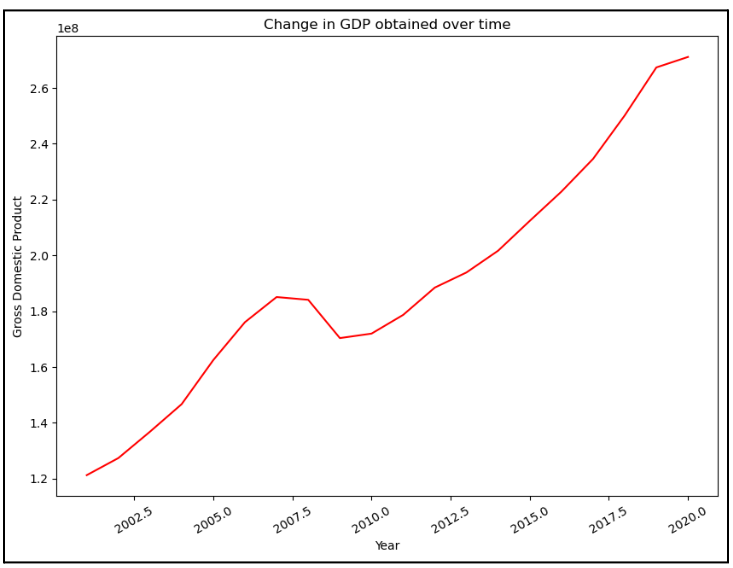

Figure 4.4 Line chart for change in GDP over time

4.3.3 Change in median household income

Household income is used by economists to arrive at conclusions about the economic health of a given county. We do not see any difference in the trends for it. We see that the household income was stable for the most part of 2020. Although, that is surprising because of the shutdowns in the country this value could have changed unless it equally affected the people in the lower part or the higher part of the distribution.

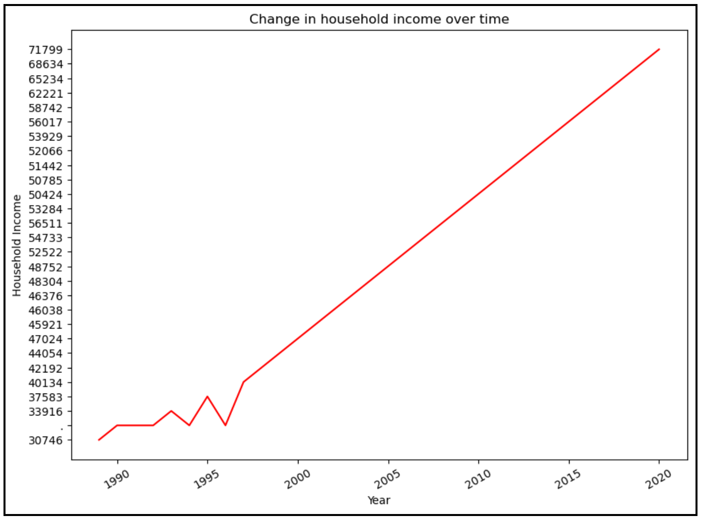

Figure 4.5 Line chart for change in median household income over time

4.4 Gradual Impact of COVID-19

In this experiment, we tried collating the industries and understanding how they were affected by the increase or decrease in COVID-19 cases. We saw that they did show some moderate-to-strong lagged cross-correlation with COVID-19 cases. The top industries which showed a negative correlation with the percentage change in employment, are as follows:

Agriculture, Forestry, Fishing and Hunting (18, -0.7620846308776791)	
Utilities	(18, -0.7132145659484582)
Oil and gas extraction	(17, -0.7271901496246335)
Pipeline transportation	(15, -0.7973103729829343)
Telecommunications (17, -0.7583358976455519)	
Insurance carriers and related activities	(14, -0.7003278642996527)
Lessors of nonfinancial intangible assets (15, -0.760855182812672)	
Computer and peripheral equipment	(16, -0.7841826245310932)

Figure 4.6 Table showing industries and their time lag and correlation data

The results are interpreted as Agriculture, Forestry, Fishing, and Hunting had a crosscorrelation of -0.7621 with COVID-19 cases with a lag of 18 months. So if the COVID-19 cases increase by 100% then there would be a 76.21% reduction in employment 18 months down the line. We can interpret the results for other industries similarly.

4.5 Overall impact

We find a correlation between the economic index and with value added by the industry. We see value added against the percentage change in unemployment. Both of these experiments show a very strong correlation without lag (0.97 and 0.99) respectively.

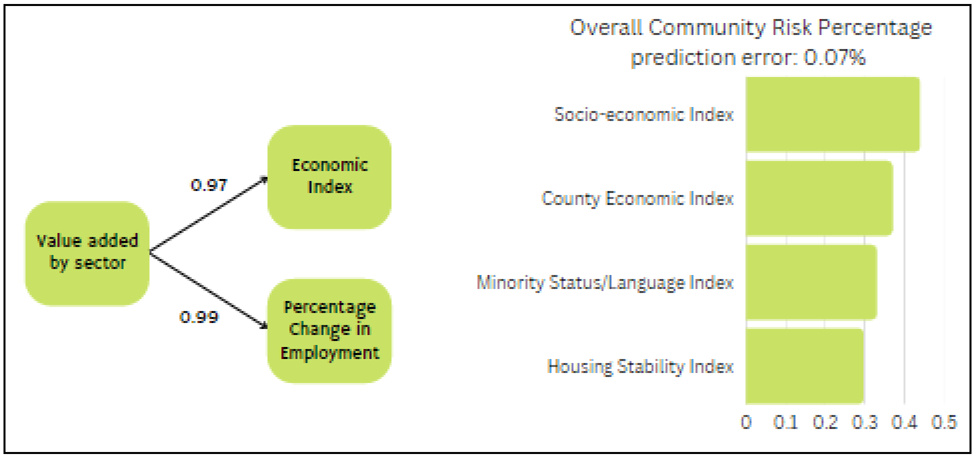

Figure 4.7 (Left) Correlation of economic index with value added by sector and percentage change in employment, (Right): Coefficient values of variables with overall community risk index

We then predict overall community risk prediction using five indexes and get that these specific indexes show how important they are in predicting the overall community risk. Per unit change, the socioeconomic index showed a change of 0.43 overall community risk index keeping all other variables constant. Similarly, the model's results can be interpreted for others, the county economic index, minority status/ language index, and housing stability index. There would be certain factors having some sort of correlation with these respective indexes, which is a part of the future part of the analysis. We tried exploring the first two here. We get the mean absolute percentage error of the overall community risk (measured on a scale of 0 to 5) as 0.07% which is fairly accurate.

5. Discussion/Implications

Examining the socioeconomic impact of COVID-19 helps in better gauging its impact on the economic downturn. As figure 5.1 from Sciencedirect shows us that socioeconomic status comprises occupation, income, and education. We have uncovered the facts in terms of the Occupational impact of COVID-19 in the longer term but Income and Education in the shorter term. One aspect of the future work of this project could lie in uncovering the impact of COVID-19 on Income and Education in the longer term. This analysis could then go deeper into understanding how discrimination, behaviors, stigma, and biases played a role in determining this impact directly or indirectly (Remember: As a starting point, the Minority Status/Language Index was a part of the Overall Community Risk Index prediction model). Understanding all of these aspects is the best way to describe the human-centered motivation of this project.

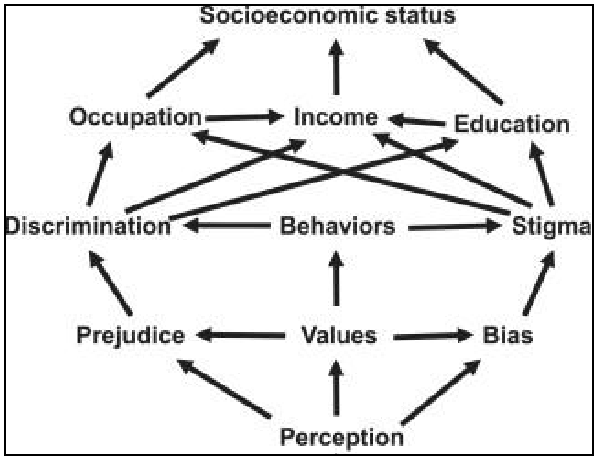

Figure 5.1: Socioeconomic factors hierarchy from Sciencedirect

We initially proposed trying to build the models that could have been much more complicated than the current ones. But the human-centered approach helped us pick more interpretable approaches as they result in more transparency. The data we used was also about fairness and hence used as many as communities possible over using specific communities or counties when building the overall community risk index model. On the other hand, a deeper analysis of all the factors involved and prediction of the impact/of COVID-19 using an autoregressive model can be further areas of improvement for this work. Hence, we do not claim this to be a comprehensive analysis but see it as a good starting point to understand the socioeconomic impact of COVID-19.

Overall, our work shows the most vulnerable industries in terms of their employment. It also shows how overall community risk for epidemics/pandemics in the future. This prediction model can directly be used to predict this in the future given the other indexes. This would help prioritize curtailment measures for specific communities in case of crisis again.

6. Limitations

In all the datasets used, the datasets are quite clean in general. The only concern is understanding the features accurately, which required researching socioeconomic terminologies in order to avoid drawing incorrect conclusions. Another concern is here, we are directly correlating the variables like economic impact in industrial sectors and the effect of COVID-19, but it's important to note that these variables might be impacted by various other factors like International relations amongst others. For example, the CDC’s Agency for Toxic Substances and Disease Registry Data measure the impacts of drug consumption and diseases on the economic indexes, but we are just correlating the impact with the diseased part.

Some of the other limitations would include, these results and analysis for Maricopa county in Arizona, US. These results might not hold for other counties in the country. When building the model overall community risk we added certain factors and left out some like the Local

Government Revenue index as we thought they would be redundant and as they correlated highly with County Economic Index. But we think adding those features might change the predictive power of the model. Also, many times there are factors like state economic indexes which would be impacting the local index of the communities and they might have some contribution in each of the counties. We consider all the counties as independent of each other which might violate the Independence assumption of linear regression. On the other hand, the time-lagged correlation of the industries with COVID-19 cases was done on the basis of their months. We took the sum of the cases, whereas the other value is a percentage change. Although on the outside, this looks correct, a statistically more appropriate way of doing it could be the percentage change in COVID-19 cases against the percentage change in unemployment. The COVID-19 masking data we used could be incorrect as those are the CDC guidelines and not state mandates. On top of it, the masking survey data is sampled from a specific set of people, extrapolating that could be a simplification assumption that might not be right. Another limiting aspect could be that we do not gauge the long-term impact of Social factors like Education as its the limitation of data available at hand.

7. Conclusion

We started with a human-centered approach to this project trying to corroborate Dr. Cockerham's theory of socioeconomic status being the primary indicator of health and impact. We uncovered the socioeconomic impacts in immediate and gradual perspectives at the individual/community level and industry level. All of this was done keeping in mind the guidelines of code reproducibility, transparency, openness, and fairness for any analysis project.

Some of the interesting findings were: The impact of masking policies on the COVID-19 cases showed some impact with a lag of 1 month in some cases but that does not guarantee that there was any direct impact. Hence, in the absence of state mandates we use CDC mask recommendation guideline dates but do not use masking as the chief basis of this analysis. We do the correlation analysis using COVID-19 cases instead. Given its effectiveness in decreasing infection, public mask-wearing might still be an appropriate policy response to future outbreaks. Overall, the study provides suggestive evidence about the benefits of wearing masks in public during the various stages of the COVID-19 pandemic. The study also highlights the relevance of public mask-wearing for the ongoing pandemic, where the vaccination rate is precarious, and access to vaccines is still limited in many counties.

On the economic front, Unemployment and Civil Labor Force participation showed a time-lagged cross-correlation of 0.57 and 0.67 at 21 and 22 months respectively. We did not see any immediate impact of COVID-19 on socioeconomic factors like Median Household Income and Education. On the other hand, the GDP suggested a relative economic slowdown. The gradual impact showed Agriculture and Fisheries industry having a strong negative correlation(-0.71) with COVID-19 at a lag of 18 months amongst other industries. We see that the overall economic index, value added by the sector, and the percentage change in employment have a very strong correlation (>0.97). We also predict the overall community risk index with a mean absolute percentage error of 0.07%. We identify the important factors that are most important in predicting that which include the Socioeconomic Index and County Economic Index followed by the Minority Status/Language Index. Overall, these findings would help us be better prepared for determining the communities at risk and the most vulnerable industries in the future.

8. References

1. Pokhrel, S., & Chhetri, R. (2021). A Literature Review on Impact of COVID-19 Pandemic on Teaching and Learning. Higher Education for the Future, 8(1), 133–141. https://doi.org/10.1177/2347631120983481
2. Suryahadi, Asep, Ridho Al Izzati, and Daniel Suryadarma. "The impact of COVID-19 outbreak on poverty: An estimation for Indonesia." *Jakarta: The SMERU Research Institute* 12 (2020): 3-4
3. Atkeson, Andrew (2020) 'What Will be the Economic Impact of COVID-19 in the US? Rough Estimates of Disease Scenarios.' NBER Working Paper No. 26867. Cambridge, MA: National Bureau of Economic Research.
4. Eichenbaum, Martin S., Sergio Rebelo, and Mathias Trabandt (2020) 'The Macroeconomics of Epidemics.' NBER Working Paper No. 26882. Cambridge, MA: National Bureau of Economic Research
5. Rasul G, Nepal AK, Hussain A, Maharjan A, Joshi S, Lama A, Gurung P, Ahmad F, Mishra A and Sharma E (2021) Socio-Economic Implications of COVID-19 Pandemic in South Asia: Emerging Risks and Growing Challenges. Front. Sociol. 6:629693. doi: 10.3389/fsoc.2021.629693
6. Baumer, E. P. (2017). Toward human-centered algorithm design. Big Data & Society, 4(2). https://doi.org/10.1177/2053951717718854
7. Ramaswamy (2020), The impact of COVID-19 on school education and the road to recovery. Paper
8. Flanagan, Barry E.; Gregory, Edward W.; Hallisey, Elaine J.; Heitgerd, Janet L.; and Lewis, Brian (2011) "A Social Vulnerability Index for Disaster Management," Journal of Homeland Security and Emergency Management: Vol. 8: Iss. 1, Article 3. DOI: 10.2202/1547-7355.1792
9. COVID-19 Socioeconomic Impact
10. John Hopkins University COVID-19 data
11. Masking mandates by county
12. Mask compliance survey
13. Unemployment Rates in Maricopa County
14. Civilian Labor Force (Total employed)
15. Housing impact
16. CDC's Agency for Toxic Substances and Disease Registry Data
17. US Counties Economic Information
18. Wikipedia.org
19. Britannica

9. Data Sources

Following are the datasets that we used for all of the questions mentioned above.

1. **Covid-19 data:**

   We plan to use the COVID-19 masking dataset as provided in the Common Analysis part (Question 1). The COVID-19 cases dataset is used for all the questions.

   a. John Hopkins University COVID-19 data (License: Attribution 4.0 International (CC BY 4.0))
   b. Masking mandates by county (NCHS: Can be used for Statistical reporting and analyses)
   c. The New York Times mask compliance survey data (Copyright 2021 by The New York Times Company, used for non-commercial purposes)

2. **Federal Reserve Economic Data, FRED Monthly Data:**

   The dataset is licensed under FRED® Services General License and is allowed to be used none other than for statistical analysis purposes. This dataset has data points starting from 1990 to 2022 for unemployment rates, Civilian Labor Force participation. It is a two-dimensional dataset with timestamps and the respective measures in either case. We understand the effect of the pandemic (COVID-19 cases) on these two measures. It helps in answering question 2.

   The links to the dataset are as follows:

   a. Unemployment Rates in Maricopa County
   b. Civilian Labor Force (Total employed)

3. **CDC's Agency for Toxic Substances and Disease Registry Data:**

   This dataset keeps a track of the social vulnerability of counties given the diseases or abuse of toxic substances. Social Vulnerability Index (SVI) indicates the relative vulnerability of every U.S. Census tract. Census tracts are subdivisions of counties for which the Census collects statistical data. SVI ranks the tracts on 16 social factors, including unemployment, racial and ethnic minority status, and disability. We have the index values for all of these themes.

   The National Center for Health Statistics (NCHS), and Centers for Disease Control and Prevention (CDC), conduct statistical and epidemiological activities under the authority granted by the Public Health Service Act. NCHS survey data are protected by Federal confidentiality laws including Section 308(d) Public Health Service Act and the Confidential Information Protection and Statistical Efficiency Act or CIPSEA. These confidentiality laws state the data collected by NCHS may be used only for statistical reporting and analysis.

Since it is the annual patterned data we use it to answer our 3rd and 5th question about the state of the socioeconomic variables before the pandemic and after the pandemic. Median household income and Gross Domestic Product, Education, and Poverty Estimates in different groups. The dataset can be found in United States Counties.

4. **Argonne National Laboratory Data for Different Sector Information**

   This dataset contains the indexes for different industrial sectors across the country. We use it to research question 4. Since this dataset is available from January 2020-April 2022. We will be able to study the impact from almost the start of the pandemic. The data is spread out monthly for each of the counties as the index column and different sectors as the subindex. This dataset is again allowed to be used only for Statistical Analysis purposes and is licensed under DEAR 970.5204. The link for the dataset is US Counties Economic Information.

Overall the data summary looks as below: