The Economic Impact of COVID-19

Abstract

In this report, we analyzed the economic impact of COVID-19, which was measured by unemployment rate and quarterly GDP. In particular, we studied several features that were relative to COVID-19, including the types of industry, the population density, and the total cases and deaths caused by this disease. Finally, we implemented principal component analysis and decision tree algorithm to figure out the cities that were most affected by COVID-19 and predict the influence in the next 1, 2, and 5 years.

Introduction

With the global outbreak of the COVID-19 pandemic, scholars from research institutions and companies have begun to study the impact of the pandemic on the economy. From the aspect of industry types, Dua and Ellingrud (2020) from McKinsey measured the vulnerability of US small businesses during the COVID-19 crisis and concluded that accommodation & food service, health care, and construction sectors are more vulnerable than other industries [1]. Besides, Tanaka and Higashide (2017) provided a bankruptcy model for forecasting the vulnerability of industrial economic activities [2].

This report was written from macro to micro, which started by comparing the impact between different countries. Since this project focused more on the United States, which is also one of the severest epidemic area, the study of the US was state-based while the others were country-based. For regions that were mostly impacted by the COVID-19, we chose their representative cities as the analysis result. Based on our analysis, a model was constructed to predict the economic impact of COVID-19 in different regions in the next 1, 2, and 5 years using the selected features. Our methods involved Principal Component Analysis (PCA), decision tree, feature importance, and data simulation.

Methodology

This report is mainly about visualization. We used Tableau and Python to build informative map plots, scatter plots, bar plots, pie charts, as well as line plots. Besides, we conducted feature selection based on these plots. Also, Principal Component Analysis (PCA), decision tree, feature importance, and data simulation are applied in this project to analyze the economic growth rate.

Analysis

Figure 1 shows the cumulative COVID-19 cases for each country around the world. Deeper color means more cases. The data source is [3]. It is obvious that the United States has the largest number of cases. To analyze the economic effect of COVID-19 around the world, we selected eight major countries including G7 countries and China, which is the origin of COVID-19. This choice is due to the fact that our major objective is to study economic influence of COVID-19, which will be easier to analyze in larger economic entity. **Figure 2** is plotted based on data [6], revealing that the United States has the largest decrease of GDP in the second quarter of 2020 among all the eight countries. This accords with the truth that it contains the largest number of COVID-19 cases.

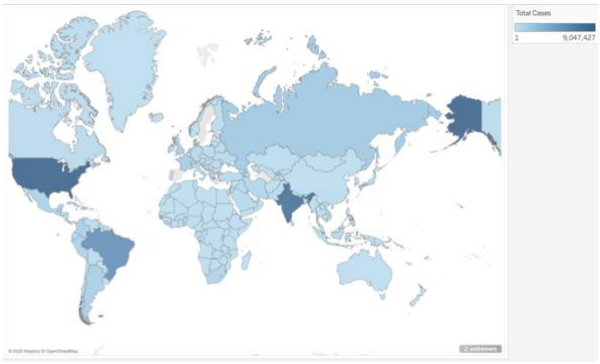

Figure 1: The world-wide distribution of COVID-19

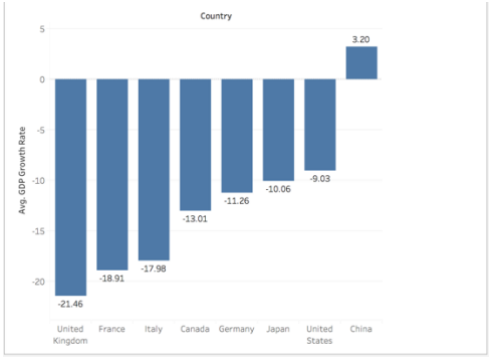

Figure 2: The second-quarter GDP growth rate of G7 countries and China (the origin of COVID-19)

Since the US has the largest number of cases and the greatest GDP reduction, the detailed analysis at the state level of the US needs to be worked on. **Figure 3** shows the cumulative COVID-19 cases in each state of the US, except Alaska and Hawaii. The red points denote the locations of the counties with the most COVID-19 cases. The data comes from [4] and [5]. It shows that California, Florida, and New York have large numbers of cases and are the clusters of the counties with the highest severity.

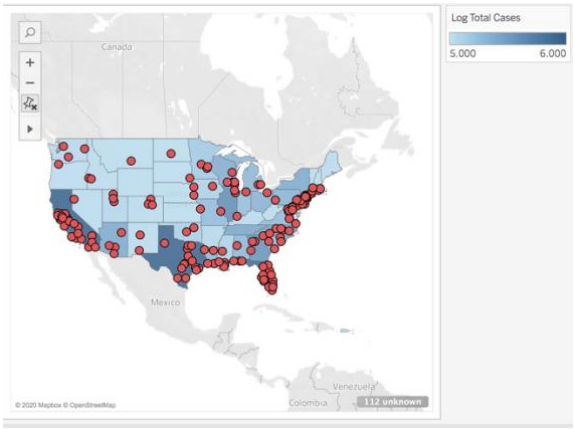

Figure 3: The severity of COVID-19 in the United States

To understand the economic impact of COVID-19, the industrial structure is an important feature to take into consideration. The COVID-19 can be spread through coughing, sneezing, or talking, thus it will intuitively have a strong impact on catering and tourist trade. To detailedly analyze the significance of industrial structure, **Figure 4** is visualized, where the x axis indicates the financial risk of each sector measured by the level of financial leverage, and the y axis calculates the degree affected by COVID-19. The sizes of the points correspond to their shares of GDP. If the points are at the upper right corner, it means that the corresponding sectors are more financially vulnerable and more negatively influenced by COVID-19. Based on this characteristic, among the 10 key industries, Health care and social assistance, Accommodation and food services, Administrative and support and waste, Transportation, Retail trade, and Construction are the key industries that we should focus on. The following *industry* feature is calculated by the proportion of the above 6 key industries in each country or state.

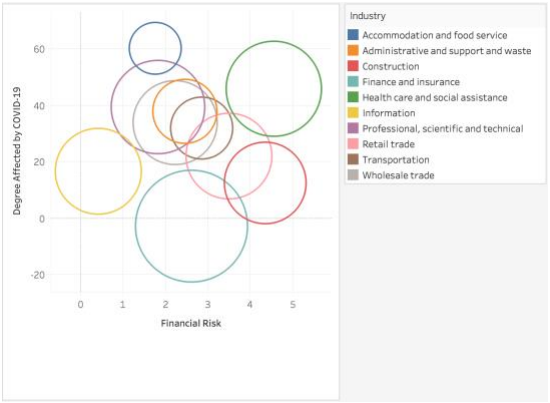

Figure 4: The industry sectors influenced by COVID-19

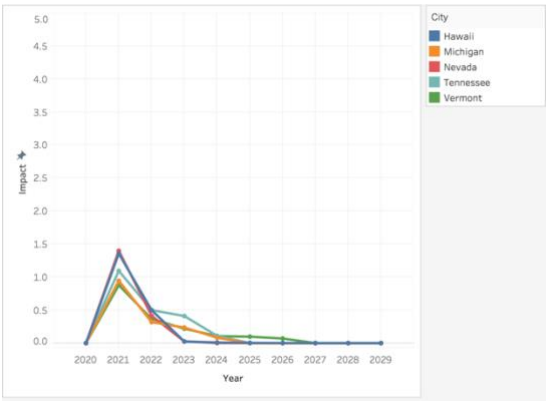

Figure 5: The COVID-19 impact in the next 1, 2, and 5 years for the 5 cities

Then, we build a model to predict the economic influence of COVID-19 for a few regions. The first step is to construct a factor that combines the information of second-quarter GDP (data source: [13]) and the unemployment rate (data source: [9]). The construction is via PCA method, which will explain more than 80% information of the original features. With the PCA results, which is shown in **Table 1**, the top regions that most economically suffer from COVID-19 can be selected. A lower PCA score identifies a more severe impact. The resulting regions are Nevada, Hawaii, Tennessee, Michigan and Vermont.

Table 1: PCA results of the regions that are most affected by COVID-19

State	PCA
Nevada	-13.99
Hawaii	-13.57
Tennessee	-10.94
Michigan	-9.45
Vermont	-8.78

Finally, we applied the decision tree algorithm to construct a prediction model based on features industry, population density, total COVID-19 cases, and *total COVID-19 deaths*, and also find out the relative feature importance during this process. According to the prediction model as well as the feature importance, we simulated the economic impact of COVID-19 for the 5 regions that mostly suffer from COVID-19 in the next 1, 2, and 5 years in **Figure 5**. As can be seen from the plot, the negative effect of COVID-19 reaches its highest point at the first year, and this influence is much lower in the second year. In the fifth year, this effect will almost disappear.

Findings

Firstly, based on the PCA method, we figure out the regions that are mostly affected by COVID-19 in terms of economic development. Although the previous analysis is at the state or country level, it doesn't lose generality to select the most representative city in the corresponding region, since COVID-19 is more severe in metropolis where the population density is higher and economic condition is more typical. As a result, the top five cities impacted by COVID-19 most acutely are Las Vegas, Honolulu, Nashville, Detroit, and Burlington, which are the capital or largest cities in the corresponding

states. The city that is negatively affected by COVID-19 to the highest extent is Las Vegas, the largest city in Nevada state. This makes sense due to Las Vegas' popularity in the gambling and tourism industry. Because of COVID-19, casinos are either closed or restricted and much fewer people would go gambling or taking a trip to Las Vegas. The PCA result also shows that the city that is least affected by COVID-19 is Beijing. This matches the reality perfectly as China is the only major country that has an increase in the second-quarter GDP in 2020.

Besides, the feature importance is visualized in **Figure 6**. The total number of deaths is the most important feature, indicating that the economic growth rate is mostly related to the fatality level; the number of COVID-19 cases ranks the second; the industry structure is the third important factor; the population density is the fourth. This finding is plausible because the number of deaths and the number of cases attract the most attention from people. The industry type also plays an important role since it reveals the vulnerability and sensitivity of certain regions towards the disaster. The population density is fairly vital as well, contributing to almost 20% of feature importance. For instance, NYC is one of the places that are struck by COVID-19 severely and it is probably related to its dense population.

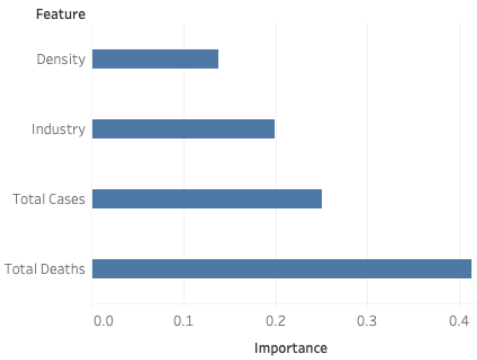

Figure 6: Feature importance

In the end, based on the decision tree model, the economic effect of COVID-19 is simulated for Las Vegas, Honolulu, Nashville, Detroit, and Burlington in the next 1, 2, and 5 years starting from 2020. As is shown in **Figure 5**, the economic impact of COVID-19 will be the severest in 2021 but can mitigate in 2022, and there will be almost no impact in 2025. This is very similar to the development of SARS in 2002.

Conclusion

Based on exploratory data analysis, we find that the world is suffering from COVID-19, and it will take a long time to recover. It is worth mentioning that there is an increasing trend in the number of cases in the United States and worldwide in recent days. However, based on the prediction result, we are confident that the world will succeed in recovering after two or three years. Even for the cities that are affected most severely, the recovery period will be less than half-decade. Based on the feature importance analysis, the number of deaths and cases of COVID-19 play a significant role in the economic impact, and the industry structure is not negligible as well. Therefore, it is suggested that governments, despite of the economic size, should make a diverse combination of industry sectors, so that their regions are more resistant to disasters like COVID-19.

Reference & Data Source

[1] Dua, Ellingrud (2020). Which small businesses are most vulnerable to COVID-19 and when. McKinsey & Company
[2] Tanaka, Higashide (2017). Forecasting the Vulnerability of Industrial Economic Activities: Predicting the Bankruptcy of Companies. Journal of Management Information and Decision Sciences, 1532-5806
[3] https://ourworldindata.org/coronavirus-data
[4] https://covid.cdc.gov/covid-data-tracker/#cases_casesper100k
[5] https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/
[6] https://stats.oecd.org/index.aspx?queryid=350
[7] https://github.com/nychealth/coronavirus-data
[8] https://fred.stlouisfed.org/release/tables?rid=331&eid=2088#snid=2122
[9] https://www.bls.gov/web/laus.supp.toc.htm
[10] https://csimarket.com/Industry/industry_Financial_Strength_Ratios.php?ind=903
[11] https://www.governing.com/gov-data/population-density-land-area-cities-map .html
[12] https://www.census.gov/data/datasets/2018/econ/cbp/2018-cbp.html
[13] https://www.bea.gov/news/2020/gross-domestic-product-state-2nd-quarter-2020

Appendix

The codes and datasets are uploaded to the following GitHub page:
https://github.com/Mingxuan-Yang/Datathon

The interactive plots are published through Tableau Public. The link is:
https://public.tableau.com/profile/mingxuan.yang