SciencePG
Science Publishing Group

Analyzing the Impact of COVID 19 on Global Trends and Predicting Future Cases

Gyeongseung Han[1], Jeewon Han[2], Seungmin Han[3], Hyunkyung Jeong[1]

[1]Chadwick International, Incheon, South Korea

[2]Cranbrook Kingswood Upper School, Bloomfield Hills, MI, United States

[3]Korean Minjok Leadership Academy, Hoengseong, South Korea

Email address:
ryan.gyeongseung.han@gmail.com (G. Han)

Abstract: COVID-19 is a highly contagious, potentially lethal respiratory disease caused by a strain of coronavirus. Utilizing sets of data collected by Johns Hopkins University, this research paper analyzes the global trends of COVID-19 and the pandemic's effect on the global economy. The aim of this paper is to provide accurate information of COVID-19 through comparing the situation of the world before and after the pandemic. All visual representations have been created using python, a programming language, and each figure is accompanied with a thorough breakdown of its cumulative data. Its analysis shows the impact of COVID-19 globally and regionally. Machine learning is utilized to predict the future trends in the number of cases, from which it can be forecasted that the world will see a continuous increase in COVID-19 cases with an exception of a few countries where cases of COVID-19 have been declining consistently. Polynomial regression has been used to predict the future trend of COVID-19. Observing the numbers used in this paper such as card usage, job posts, and confirmed cases provides evidence to the fact that COVID-19 has negatively impacted the global economy. Statistics can show the relationship between COVID-19 and the global economy, nonetheless providing evidence on why certain events are occuring during this pandemic.

Keywords: COVID-19, Disease, Analysis, Prediction

1. Introduction

The infamous COVID-19 is a respiratory disease caused by a strain of coronavirus. The virus is commonly referred to as the 2019 Novel Coronavirus (although it has been renamed SARS_CoV_2), indicating that it has not been previously documented [1]. Coronavirus precedents have surfaced before in human societies in the form of Severe Acute Respiratory Syndrome (SARS) or Middle East Respiratory Syndrome (MERS), but the most recent one undeniably has the most lasting effects. Though the first person to retract the virus has not been identified, epidemiological data backtracks the first contamination to have occurred in Wuhan's Huanan seafood market located in China. As for the origin of the virus, scientists concur that bats were the initial reservoir, and the virus was passed on to humans through an intermediary. During the initial stages of

its outbreak, the estimated mortality rate for novel coronavirus was between 1 and 2 percent, relatively low compared to its counterparts, but real time data prove it to be around 2.79 % worldwide at 10:37 PM eastern time on October 18th, 2020 [2]. Approximately one month after its discovery on December 27th of 2019, the WHO declared the COVID-19 outbreak to be a public health emergency of International Concern [3]. From then on, nearly every country in the world has been devastated by the virus in various ways. Over 30 million have been diagnosed globally, with more than 690,000 deaths as of the 20th of September [4]. The suffering and passing of loved ones are the tip of the iceberg, as COVID-19 has had socio-economic effects as well. The IMF predicts that the global economy will shrink by 3% this year [5], and UNESCO estimates that close to 900 million learners have been affected by the closure of educational institutions, being forced to resort to online learning and placed in social isolation [6]. For those of a

disadvantaged or rural family without access to technology, the pursuit of education has become difficult. Cambridge University predicts the pandemic to cause the global economy to lose around $82 trillion across five years [7]. Needless to say, putting an end to this pandemic as quickly as possible is at the forefront of humanity's concerns. Some countries were able to keep the number of infected low, while others were not so successful. According to the data collected by Johns Hopkins University, the total number of coronavirus cases in the United States were way past the eight million mark (Around 2.52% of the total US population [8]) by October 16th and the death toll neared 224,282 [9]. In contrast, certain nations such as South Korea, where the total number of cases was 25,199 as of October 18 [10] (Around 0.05% of the total South Korean population [11]) and a death toll of merely 444 [10], appeared to maintain hold of the crisis. The purpose of this paper is to analyze the trends in COVID-19 cases, changes in job posts, and card usage to examine its economic impact on the United States. All data used in the study has been collected by Johns Hopkins University, and analysis will be done using Python, a programming language. Furthermore, we will be investigating card usages to study the economic impact on individual states within the United States, as card usage successfully reflects economic conditions because the majority of American adults own one or more cards [12]. Changes of job postings within certain states of the United States will also be analyzed, as it reveals how healthy an economy is. Finally, polynomial regression will be utilized to predict the future number of COVID-19 cases.

The following paper has been organized into five sections to enhance the neatness and composition of the paper. The first section is the Introduction, which is where this paragraph is located. It introduces COVID-19 and the format of this research paper. The second section is Methodology. The third section is Prediction. The fourth section is the "Conclusion & Discussion." The fifth section is the Reference.

2. Methodology

2.1. Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) is an approach for data analysis to find patterns, extract important variables, detect anomalies, test hypotheses, and check assumptions using summary statistics and graphic representation [9]. In this section, we will be performing an EDA of COVID-19 data of the World, South Korea, and the United States.

2.2. Global COVID-19 Situation

COVID-19 has quickly spread throughout the world. The earliest cases of COVID-19 were documented in December 2019 at a wholesale food market in Wuhan, China. A month later, the first death due to COVID-19 occurred in that very city. Since then, there have been more than 28 million COVID-19 cases around the world, and more than 917,000 fatalities.

Below are two graphs that show the globally confirmed cases of COVID-19, and the global recovery rate (percentage ratio of those who have recovered from COVID-19 to the total number of confirmed cases of the world) of COVID-19.
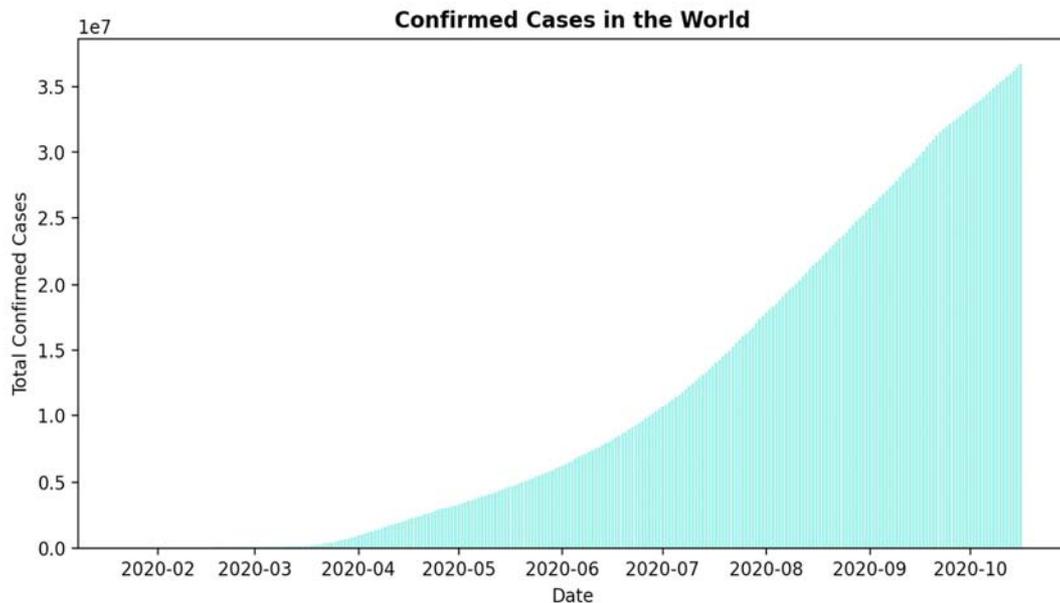

Figure 1. Number of total confirmed cases in the world (1/22/2020-10/16/2020).

The graph in figure 1 shows the number of confirmed cases of COVID-19 in the world. It is evident that the cases have skyrocketed since May, and are steadily increasing since. Since May 1st to October 1st, confirmed cases of COVID-19 have increased by 897% in less than a year. This shows that the world as a whole isn't able to reduce the

number of coronavirus cases around the world, and is   struggling to prevent the spread of COVID-19.
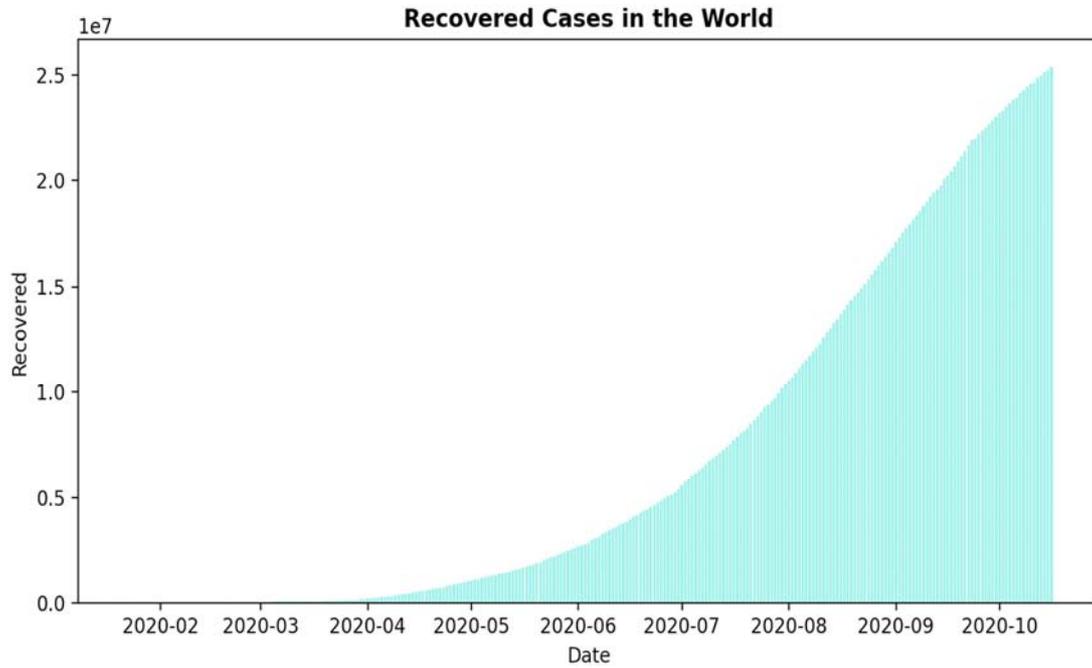

Figure 2. *Number of total recovered cases in the world (1/22/2020-10/16/2020).*

The graph in figure 2 shows the number of recovered cases in the world. According to the graph, the recovery rate has been rapidly increasing over time. The slope of the graph gets steeper as time passes, which means that the recovered cases will continue to increase around the globe. As of October 16th, there have been 25,392,559 COVID-19 cases in the world. As of this data, we can presume that the coronavirus cases will continue to increase rapidly.

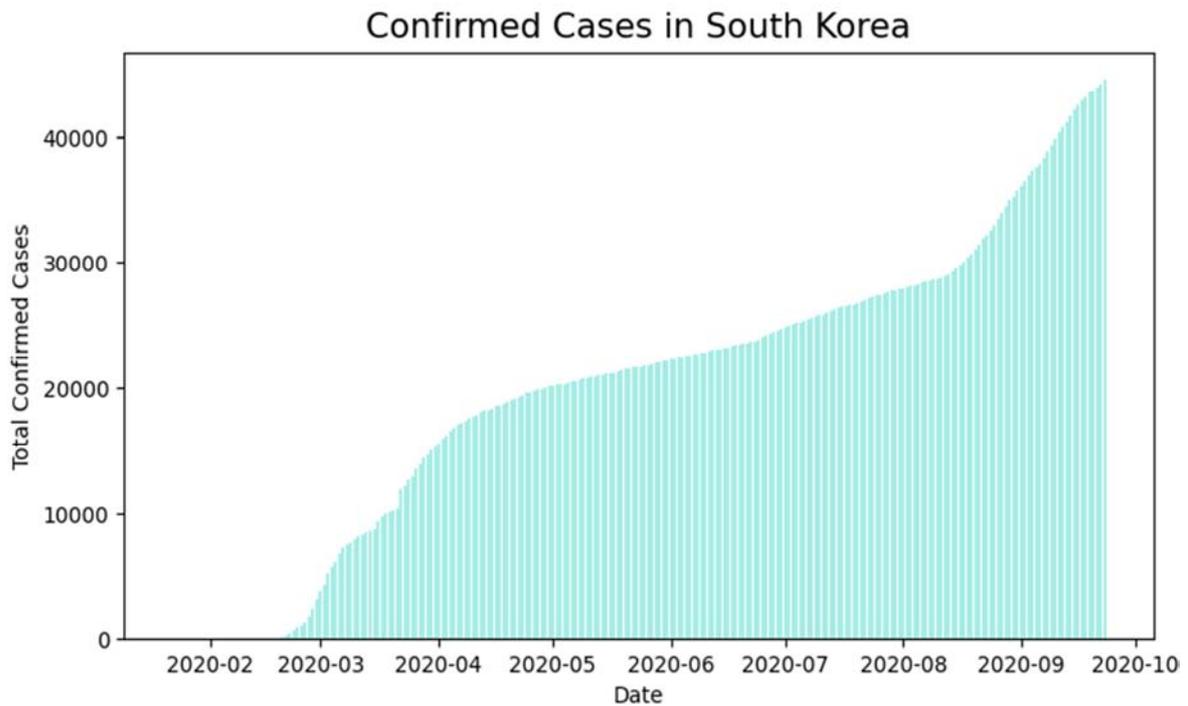### 2.3. COVID-19 Trend in United States and South Korea

Figure 3. *Number of Total Confirmed Cases in South Korea. (1/22/2020-10/16/2020).*
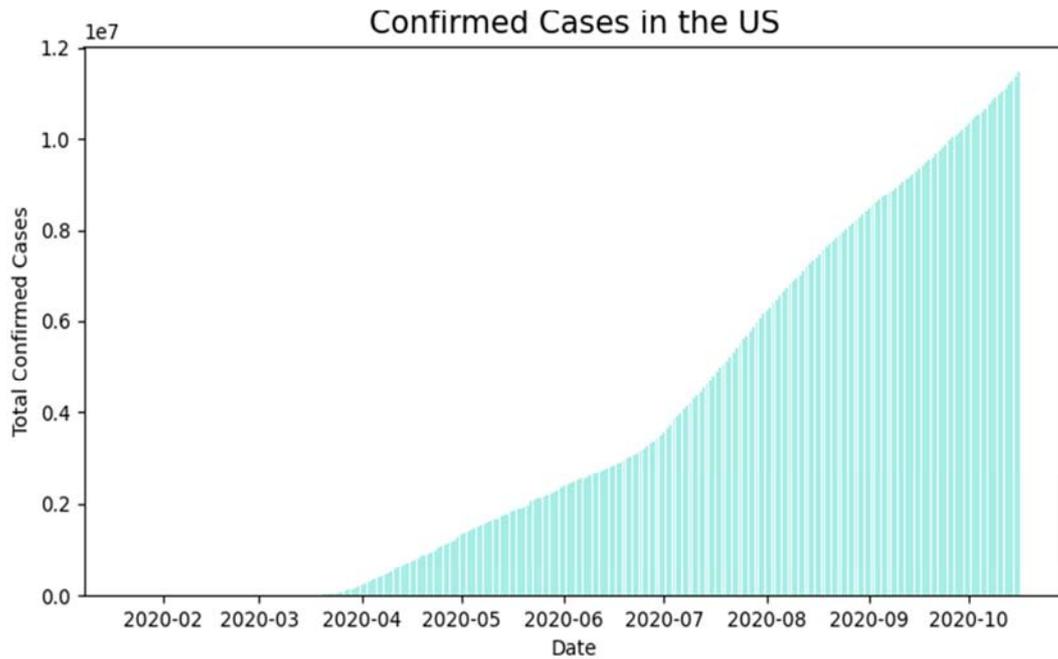
Figure 4. *Number of Total Confirmed Cases in the United States. (1/22/2020-10/16/2020).*

The curve in figure 3 flattens over time, depicting that the number of new confirmed cases is declining in South Korea. On the other hand, the curve in figure 4 becomes steeper, signaling the increase of new confirmed cases in the US. By August, the US also has significantly more cases compared to South Korea despite its curve starting much later than the latter. Calculated from the data collected from January 22, 2020 to October 16, 2020, the world average death rate was 4.830% and the average recovery rate was 48.489%. South Korea had a lower death rate of 2.042% and an extremely high recovery rate of 76.257%. However, the United States had dramatically different results, with a striking death rate of 26.982% and a recovery rate of 25.831%.

2.4. COVID-19 Trend by State - United States

Analyzing the shape of the curves can help discover and understand how each respective country has dealt with COVID-19. For example, the curve of the graph in figure 1 and figure 4 show similarities, as they both have a steep curve starting in June. However, this curve is not seen in the graph of figure 3. The curve of figure 3 is shown early in mid March, and ends early in mid April. After April, the graph shows that the cases have been steadily increasing but it is evident that it has a less steeper slope, which means that there are less confirmed cases per day. Therefore, it can be inferred from the graphs that South Korea has been able to manage the outbreak of COVID-19 earlier compared to the United States and the rest of the world.
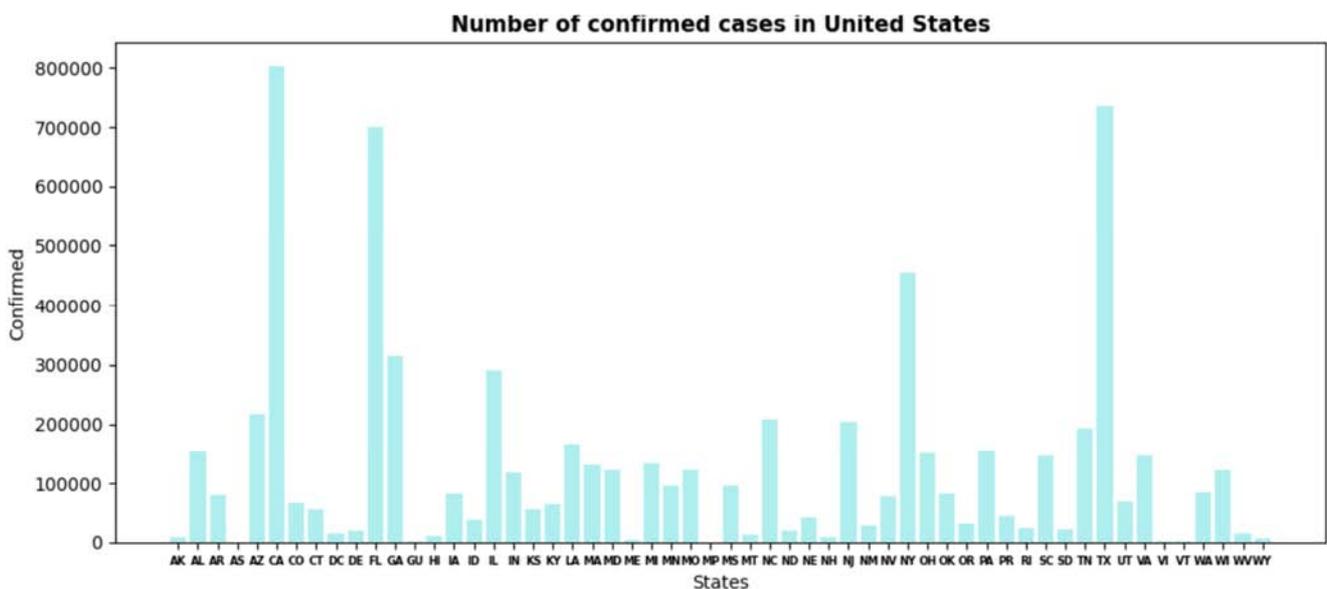

Figure 5. *Number of Total Confirmed Cases in the United States By State (1/22/2020-9/27/2020).*
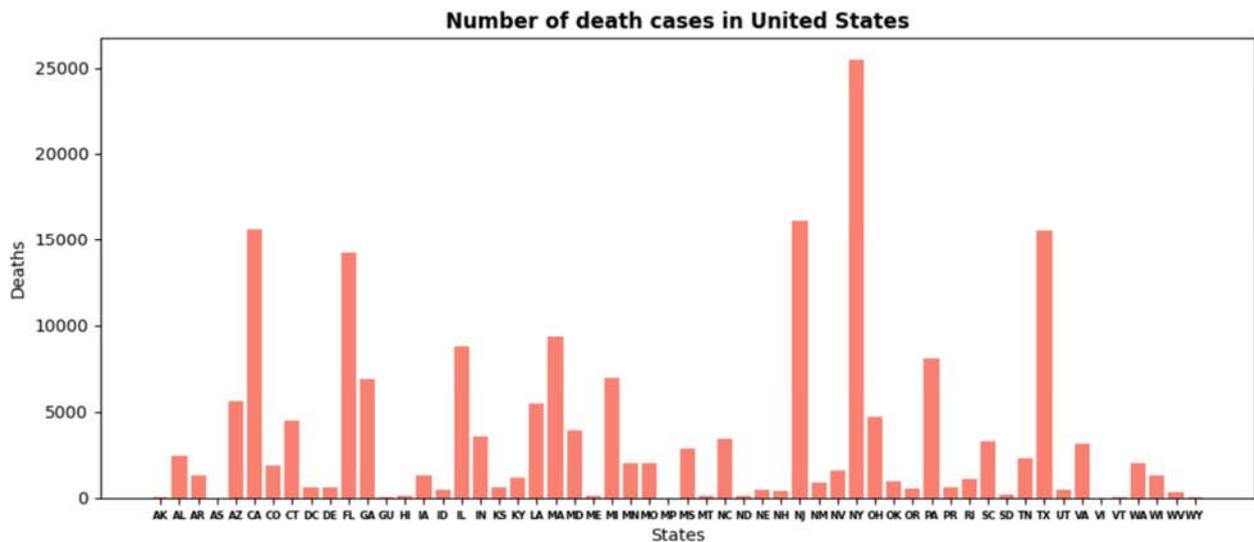
Figure 6. *Number of Total Death Cases in the United States By State (1/22/2020-9/27/2020).*

The graph in figure 5 shows that people are more susceptible to COVID-19 in metropolitan counties than in non-metropolitan ones. New York City, the nation's largest city, recorded 422,703 confirmed cases, representing 8.17% of the nation's total infections. It can also be seen that the number of deaths shows less connection in between the number of confirmed cases in the United States. States such as Florida

have recorded the second most confirmed cases, although they have significantly lower death cases in their respective state. Calculated from the data collected from January 22, 2020 to September 27, 2020, the average death rate of New York residents was 5.98% while other states like Florida and California had an average death rate around 1.72%.

2.5. Economic Impact (Card Usage) - United States

Table 1. *Monthly Average of Change of Credit and Debit Card Usage in the United States (3/1/2020-9/27/2020).*

	ACF	AER	APG	GRF	HCS	TWS	All
Mar	-0.0012	-0.0005	-0.0005	-0.0006	-0.0005	-0.0183	-0.00007
Apr	-0.345	-0.464	-0.0044	-0.0007	-0.0849	-0.514	-0.0052
May	-0.203	-0.122	-0.00016	-0.0020	-0.0024	-0.407	-0.0016
Jun	-0.160	-0.159	-0.00008	-0.0003	-0.0011	-0.359	-0.0001
Jul	-0.0729	-0.269	-0.00002	-0.0006	-0.0001	-0.309	-0.00002
Aug	-0.0481	-0.133	-0.0005	-0.0018	-0.0011	-0.306	-0.0001
Sep	-0.0128	-0.0343	-0.0007	-0.0012	-0.00002	-0.156	-0.00002

In the table above, ACF stands for Accomodation and food services, AER for art, entertainment, and recreation, APG for general merchandise stores, apparel, accessories, GFR for food and grocery stores, HCS for healthcare and social assistance, and TWS for transportation and warehousing. Changes in each category are calculated in comparison to January. In all categories, card usage has declined, which translates to a decline in the economy. Spending in Transportation and Warehousing has declined the most out of all, while spending in general merchandise stores, apparel, and accessories were the least affected. All categories except

GFR have experienced the largest drop in April, when COVID cases started to surge rapidly in the United States, as indicated in Figure 4. GFR experienced the largest drop in May. In total, April had a significantly larger drop of -0.0052 compared to other months, where card usage only dropped an average of around -0.0001.

Of all categories in April, TWS experienced the largest drop, reflecting the decrease of activity due to COVID, which also links to the decline of card usage in other categories such as AER or ACF.

Table 2. *Monthly average of card usage and those of their respective fields in New York.*

Month	Accommodation and food service	Art, entertainment and recreation	All merchant categories	Health care and social assistance	Transportation and warehousing
Mar	-0.0427	-0.0506	-0.0014	-0.0207	-0.1180
Apr	-0.6130	-0.7310	-0.2770	-0.5350	-0.7100
May	-0.4930	-0.6410	-0.1370	-0.4310	-0.5970
Jun	-0.3880	-0.5110	-0.0385	-0.1890	-0.4730
Jul	-0.3480	-0.5770	-0.0714	-0.0489	-0.4810
Aug	-0.3230	-0.5710	-0.0578	-0.1090	-0.4720

The uses of the card are divided into five groups according to their relevant fields in the table above so that it is possible to view the monthly average of card usage in New York by different categories. The usage of cards started decreasing from March and reached the peak in April, and started to increase by July and August.

Like the US as a whole, the category that was most affected by COVID-19 is transportation and warehousing and April had the largest drop in New York.

2.6. Economic Impact (Job Postings) - United States

Table 3. *Monthly Average Changes of Job Posts and those in their Specified Zones.*

Month	All	Job Zone 1&2	Job Zone 3,4&5
Mar	-0.027	-0.025	-0.035
Apr	-0.282	-0.283	-0.282
May	-0.328	-0.271	-0.36
Jun	-0.086	-0.034	-0.149
Jul	-0.152	-0.021	-0.222
Aug	-0.166	-0.012	-0.253

Table 2 shows the monthly flux in job posts in accordance with January, which is set at 1. A Job Zone is a group of occupations that are similar in how much education, related experience, and on-the-job training that is required for a person to do the work, with Job Zone 1 being occupations that require little to no preparation and Job Zone 5 being occupations that require extensive preparation. Job Zones with lower numbers (1, 2, an 3) usually require less experience and training, while Job Zones with higher numbers (4 and 5) require extensive skill, training, and experience. These jobs require a bachelor's degree and many years of extensive training. This table has divided jobs into 2 groups to show how the specific types of occupations were affected in the respective months.

With every month declining in total job posts, it can be reasonably inferred that the rate at which new jobs are being made is decreasing. It can be seen that there was a drastic peak in diminution in April and May but the numbers have become more stable since.

Table 4. *Monthly Average Changes of Job Posts and Those in Their Specified Zones in in New York.*

Month	All	Job Zone 1&2	Job Zone 3,4&5
Mar	-0.098	-0.145	-0.065
Apr	-0.413	-0.431	-0.396
May	-0.417	-0.475	-0.409
Jun	-0.268	-0.243	-0.295
Jul	-0.263	-0.053	-0.342
Aug	-0.304	-0.426	-0.394
Sep	-0.334	-0.277	-0.371
Oct	-0.267	-0.160	-0.310

California, Florida, and New York have one of the highest numbers of infected patients (according to figure 5), while Ohio and Illinois have a relatively low number. Overall, all states have had the largest drop of job postings around April and May when COVID began to spread rapidly within the US. However, for Job Zones 1 and 2, Ohio and Illinois experienced a sudden decline in the decrease of job posts between April and June. Ohio's changes of job posts rose from -0.471 in April to a -0.012 in June. In Illinois, the changes rose from -0.521 in April to a -0.047 in June. However, California, Florida, and New York did not experience such steep changes. Changes in the Job Posts of Job Zones 3, 4, and 5 had remained relatively constant compared to the other two zones for all states.

Table 5. *Monthly Average Changes of Job Posts and Those in Their Specified Zones in Ohio.*

Month	All	Job Zone 1&2	Job Zone 3,4&5
Mar	-0.330	-0.211	-0.389
Apr	-0.547	-0.471	-0.588
May	-0.470	-0.269	-0.587
Jun	-0.269	-0.012	-0.401
Jul	-0.263	-0.031	-0.466
Aug	-0.352	-0.074	-0.496
Sep	-0.335	-0.129	-0.460
Oct	-0.336	-0.123	-0.453

Table 6. *Monthly average of all job posts and those of their respective fields.*

	All	Manufacture	Finance	Professional & Business	Education & Health	Leisure & Hospitality
Mar	-0.027	-0.015	-0.011	-0.043	-0.035	-0.110
Apr	-0.282	-0.328	-0.335	-0.306	-0.274	-0.561
May	-0.328	-0.332	-0.379	-0.399	-0.338	-0.454
Jun	-0.086	-0.118	-0.076	-0.138	-0.129	-0.006
Jul	-0.152	-0.160	-0.239	-0.258	-0.137	-0.254
Aug	-0.166	-0.134	-0.179	-0.263	-0.167	-0.302

In the table above, jobs are divided into five groups according to their related fields. The groups are manufacturing, financial activities, professional and business services, education and health services, and leisure and hospitality. Therefore, it is possible to view the change in job posts as a whole but also in separate groups. Values are the difference of job posts calculated in comparison with January (which is set as 1). Total job posts started decreasing in March and reached its apex in May, but the number of job posts rose significantly in June and slightly declined in July and August. All state levels by sector follow the overall trend: the biggest drops in job posts are in April and May, and there is a significant rise in June before a slight decrease in the following months. The job types most heavily affected by COVID-19 are leisure and hospitality, as the decline of job posts in April is the biggest out of all the groups, and the

decrease remains biggest as of August.

3. Prediction

3.1. Techniques & Terms

a) Logistic Regression -7395030

In a logistic regression, the relationship between x and y is modeled as a logistic function. It is mostly used for binary classification problems, such as whether an email is spam or not [13].

b) Polynomial Regression

In a polynomial regression, the relationship between x and y is modeled as an nth degree polynomial. In nonlinear graphs, the line of best fit may not be as accurate as the actual curve. Polynomial regressions allow the line of best fit to curve and fit the shape similar to that of the actual graph, and therefore gives a more accurate prediction [14].

c) Root Mean Squared Error (RMSE)

The root mean squared error is the measure of difference in distance between the actual data and the estimated data point. Commonly referred to as RMSE, the Root Mean Squared Error shows the difference of the estimated data and real data. The RMSE usually measures the error of a model in predicting quantitative data [15]. The RMSE is used in many ways, such as analyzing outliers of a data or to define any unwanted data points.
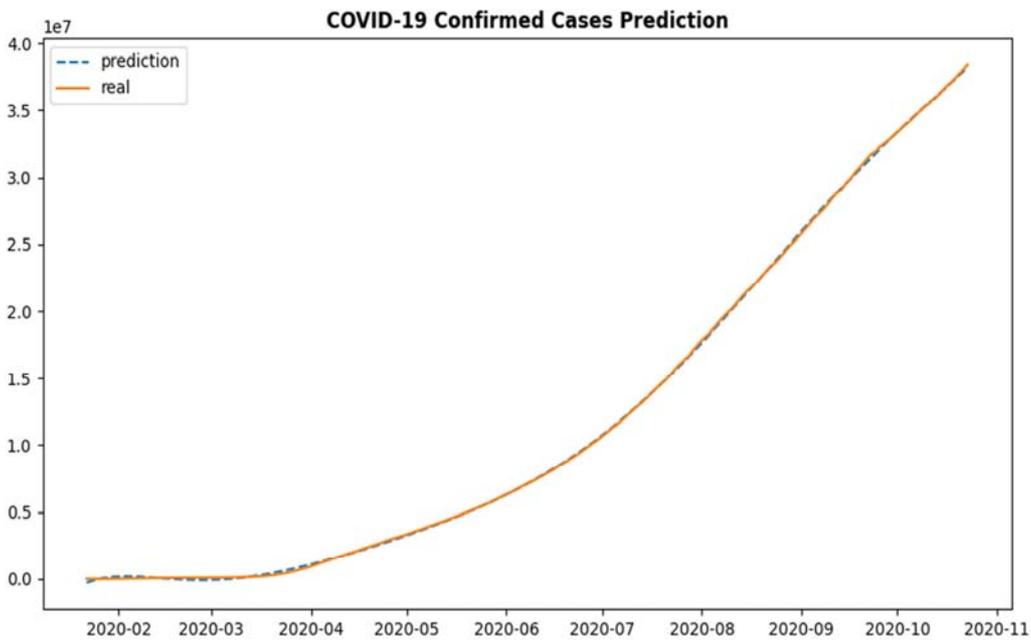
3.2. Prediction

Figure 7. *Prediction of COVID-19 Confirmed Cases Globally.*
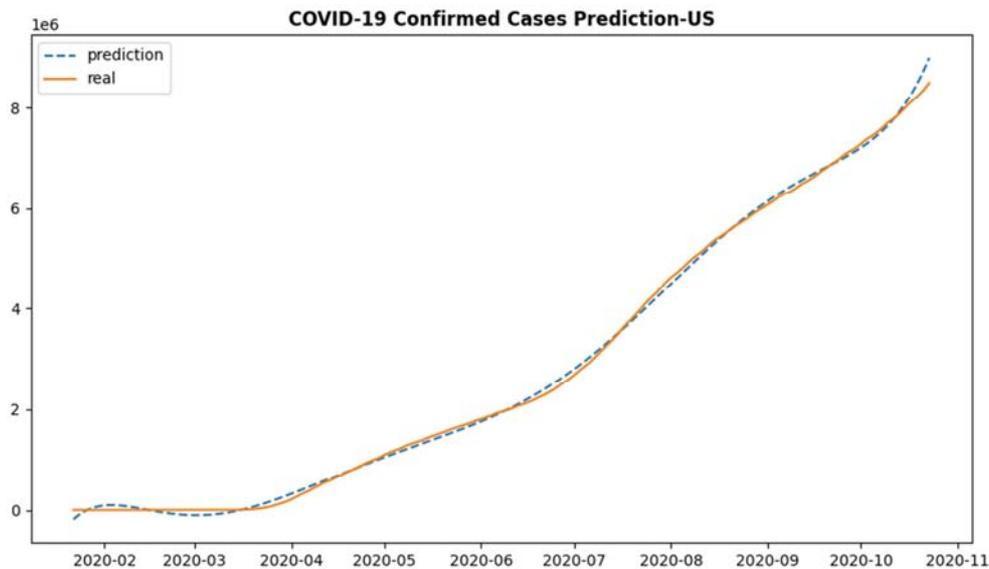

Figure 8. *COVID-19 Confirmed Cases Predictions for the United States.*

Figure 7 is a prediction for the number of COVID-19 cases globally from October 25th until November 2020. The prediction was done through using logistic and polynomial regression, then comparing the two's root mean square values. Figure 8 utilized the same methods to predict the number of COVID-19 cases until November exclusively in the United States. In Figure 7, the root mean square for using a polynomial regression was 112787, while for logistic regression, it was 7395030. Because the root mean square for using polynomial regression was significantly lower, it was used to predict. Figure 8 also utilized the lower mean square error value (357710 for polynomial regression). On October 25th, the number of confirmed cases in the United States are predicted to be approximately 9,140,000. The number of cases are to rapidly increase and by October 29th, is predicted to exceed 10,000,000. By November 1st, the United States is predicted to hit 10,747,514 cases. On November 9th, the number of cases are to exceed 13,300,000, which is more than a 37% increase compared to October 25th.

4. Conclusion

This research paper has shown the impact, future trend, and current situation of COVID-19 using various graphs and visual representations. Graphing of the data collected shows a rapid increase in the number of COVID cases globally. Some countries, such as South Korea, were able to flatten the curve somewhat, while other countries such as the United States experienced a constant rapid increase in the number of confirmed cases. Within the United States, California, Florida, and Texas had the 3 highest number of confirmed cases, with California nearing 800,000 infected patients. New York had the highest death rate of all states, with an approximately 5.98% death rate. Credit and debit card usage showed the changes in economic activities by categories, and displayed that transportation and warehousing had the largest overall decrease in spending since March. Analysis of changes in job posts by zones show that the largest decrease in posts were around April and May, when COVID-19 started to rapidly spread in the United States. Through analyzing the monthly average of all job posts and those of their respective fields, it was discovered that the job types most affected by COVID-19 were jobs related to leisure or hospitality. For all categories, the largest decrease in jobs was around April and May once again, signaling economic instability caused by the pandemic. Predictions of the number of cases made from previous data displays that there will be a constant increase of COVID-19 cases both globally and domestically within the United States.

References

[1] Qiu, Y., Chen, X., & Shi, W. (2020). Impacts of Social and Economic Factors on the Transmission of Coronavirus Disease 2019 (COVID-19) in China (Working Paper 494 [pre.]). GLO Discussion Paper.

[2] "Novel Coronavirus (2019-nCoV) situation reports, World Health Organization, October 2020.

[3] Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., Niu, P., Zhan, F., Ma, X., Wang, D., Xu, W., Wu, G., Gao, G. F., Tan, W., & China Novel Coronavirus Investigating and Research Team. (2020). A Novel Coronavirus from Patients with Pneumonia in China, 2019.

[4] "Outbreak Notification", National Health Commission (NHC) of the People's Republic of China. September 20, 2020.

[5] N. Donthu, A. Gutafsson "Effects of COVID-19 on business and research, Journal of Business Research, September 2020.

[6] Nicola, Maria, Zaid Alsafi, Catrin Sohrabi, Ahmed Kerwan, Ahmed Al-Jabir, Christos Iosifidis, Maliha Agha, and Riaz Agha. "The Socio-Economic Implications of the Coronavirus Pandemic (COVID-19): A Review." International journal of surgery (London, England). IJS Publishing Group Ltd. Published by Elsevier Ltd., April 17, 2020.

[7] Campbell, Charlie, and Alice Park. "Where Did Coronavirus Originate? Inside the Hunt to Find Out." Time. Time, July 23, 2020.

[8] Abouk, R., & Heydari, B. (2020). The Immediate Effect of COVID-19 Policies on Social Distancing Behavior in the United States. MedRxiv, 2020.04.07.20057356.

[9] Baccini, L., & Brodeur, A. (2020). Explaining Governors' Response to the Covid-19 Pandemic in the United States (SSRN Scholarly Paper ID 3579229). Social Science Research Network.

[10] Amy Dighe, Lo. Cattar, Steven Riley "Response to COVID-19 in South Korea and implications for lifting stringent interventions", October 2020.

[11] Shin young Park "Coronavirus Disease Outbreak in Call Center", Centers for Disease Control and prevention, August 2020.

[12] A. Brodeur, A. Islam, David Gray, and Suraiya Jabeen Bhuiyan "A Literature Review of the Economics of COVID-19", IZA Institute of Labor Economics, 2020.

[13] Joanne Peng "An Introduction to Logistic Regression Analysis and Reporting", The Journal of Educational Research, September 2020.

[14] Hsin-Hsiung Huang. "Nonlinear Regression Analysis." Statistics By Jim, January, 2010.

[15] Tianfeng, Chai. "Root mean square error (RMSE) or mean absolute error (MAE)?" Geoscientific Model Development, June, 2014.