The Economic Impact of Covid-19 in the U.S

CS989: Big Data Fundamentals

Contents

Chapter 1- Introduction

It is fair to say that the Covid-19 pandemic caused the largest economic impact in more than a century which had a detrimental effect on the world economy. This severely affected the daily lives of people worldwide. An example of an economic impact Covid-19 has had globally relates to the resulting effect on the location of work for employees. The Office for National Statistics (2022) Report shows that there was an increase in the percentage of adults that were working from home whilst the Measures of Plan B were being implemented in England from the 10^{th} of December 2021 to 27^{th} January 2022. A significant economic impact of Covid-19 relates to the effect the virus had on the labour market, particularly unemployment and economic inactivity. For example, in the UK from between January and March 2020 to between October and December 2020 the number of people unemployed rose by approximately 400,000 in addition to the number of people that were economically inactive increasing by 327,000. The unemployment levels improved at the start of 2021, but economic inactivity continued to increase, hitting its peak between December 2021 and February 2021 at 8.89 million being economically inactive. This was almost 450,000 people more compared to the January - March 2020 period at the beginning of the pandemic (Devine et al; 2022).

The labour market in the U.S. is softening so far in an orderly manner. Although the job report of August 2023 shows a solid rise of payrolls and adding 187,000 new jobs to the labour market, the three-month moving average is 150,000 which is the slowest since January 2021. The vacancies to unemployed ratio also decreased to 1.5 from a peak of 2.0. Jobless claims are also historically low at 230,000 (Rasheed and Holzheu, 2023). U.S. labour market recovery has been exceptionally strong despite the unemployment rates skyrocketed up 11.1% from December 2019 while many European countries unemployment rate stayed relatively flat. In January, the unemployment rate in the U.S. reached more than a 50-year low (Harris and Sinclair, 2023). Revenue from air travel, indoor dining, and participation in large in-person gatherings fell by more than 50% during the first 30-months of the Covid-19 pandemic. Changes in the public's behavior, brought about by regulations and personal health concerns, caused the decline by 57.5%, 26.5% and 29.16%, respectively (Hlavka and Rose, 2023).

Chapter 2- Dataset

The U.S. was chosen for this analysis as this gives a large sample size to work from and draw any conclusions. The data extrapolated for this investigation was from the opportunity insights repository and highlights the influence of Covid-19 on the economic impact of various cities in the U.S. The key indicators from our data that illustrate this economic impact in which we will be analysing are the following: employment level, patterns observed through google mobility tracking, both the percentage change in number of small businesses opened and the percentage change in net revenue for small businesses and consumer spending levels. The data spans a range of 3 years from 2020-2022 to depict the influence of Covid-19 throughout the duration of the pandemic from its beginning, peaks and end to ascertain a holistic view of the effect of the virus on these economic factors.

Key column headings for the analysis were chosen from each economic factor from the opportunity's insight repository with the largest dataset of 64684 rows x 28 columns and condensed into new datasets ranging from 9724 rows x 17 columns to 102 rows x 11 columns. The Key variables used in this analysis can be seen in the table below.

Table	**Size**	**Source**	**Update Frequency**
Employment	9724 x 10 columns	Paychex, Intuit	Weekly
Consumer Spending	46354 rows x 3 columns	Affinity Solutions	Weekly
GPS Mobility	51145 rows x 11 columns	Google COVID-19 Community Mobility Reports, American Time Use Survey	Daily
Small Business Openings and Revenue	5777 rows x 4 columns	Womply	Weekly

COVID-19 Infections	61292 rows x 4 columns	The New York Times, The Johns Hopkins Coronavirus Resource Center	Daily

Variables**:**

Consumer spending

- spend_all: Spending in all merchant category codes.

Employment

- emp: Employment level for all workers.

Google Mobility

- gps_away_from_home: Time spent outside of home locations.
- gps_retail_and_recreation: Time spent at entertainment and retail locations.
- gps_grocery_and_pharmacy: Time spent at pharmacy and grocery locations.
- gps_workplaces: Time spent at places of work.
- gps_residential: Time spent at residential locations.

Small business openings and revenue data

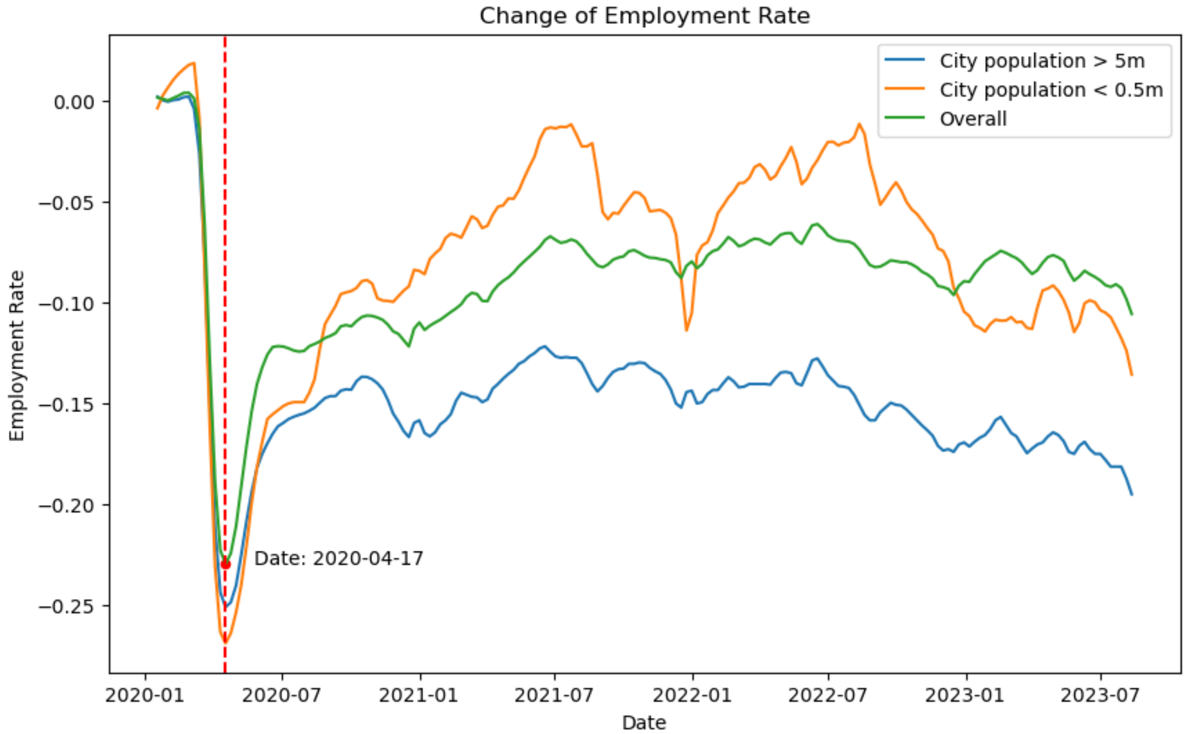
- revenue_all: Percent change in small businesses net revenue
- merchants_all: Percent change in number of small businesses open in the Covid-19 period
- case count: Confirmed Covid-19 cases.
- new_case_count: New confirmed Covid-19 cases.

Chapter 3 - Analysis of Data

3.1 Outline of Analysis

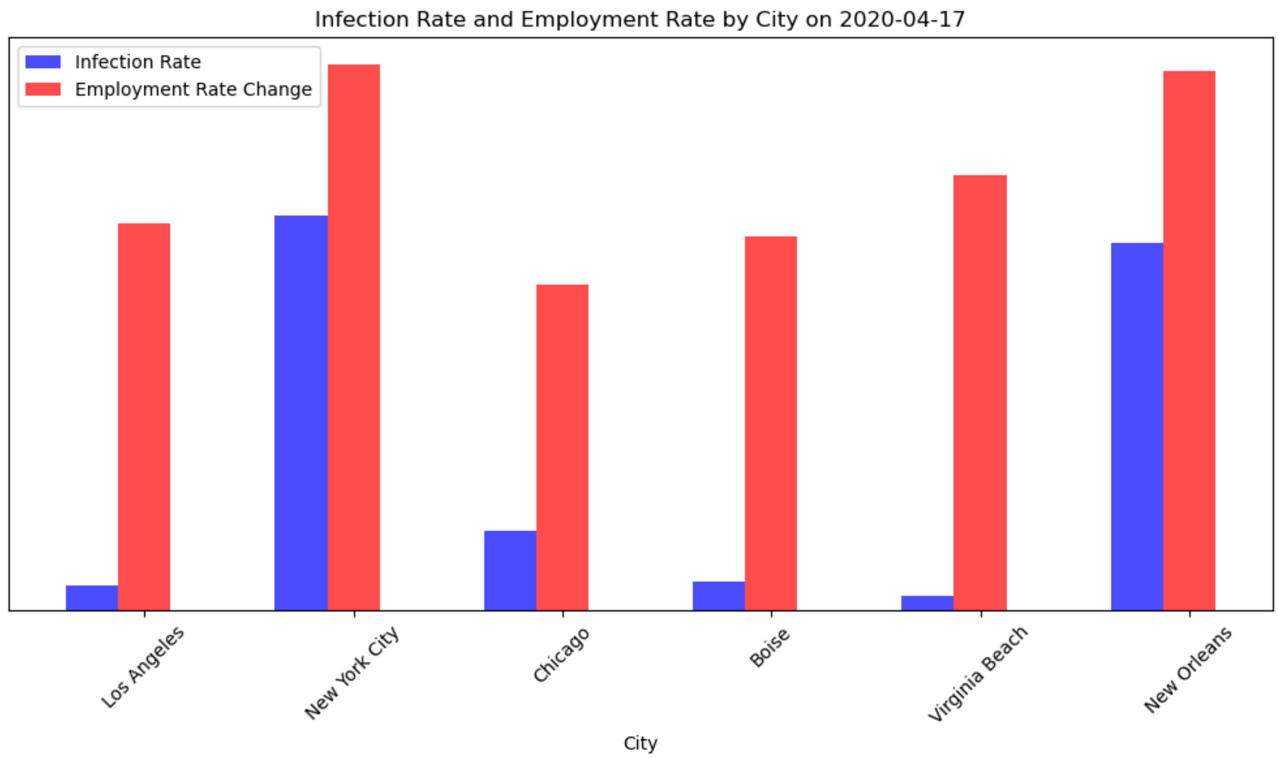
To carry out the analysis for employment rates both line and bar graphs were used to compare the differences in employment rates before, the peak and the end of the pandemic among small and large cities. In relation to the line graph the U.S. cities were grouped by population size of above 5 million, under 0.5 million and the total population of all the cities as the overall. In the bar graph, infection rate percentage for the cities with different sized population were derived to validate the impression of the link between Covid-19 and employment rates further. Thereafter the highest values from the charts were observed. All the variables used for each economic factor were illustrated in the form of a heatmap to show the exclusive correlation among each of these variables and the resulting behavioral changes of the people in the society. Another line graph was formulated to ascertain the change of workplace behavior by comparing the time spent at workplaces with the time elapsed during the duration of the Covid-19 period. The employment rate was added to link between employment and the number of people returning to the workplace to see if there was a correlation.

3.2 Change of Employment Rate



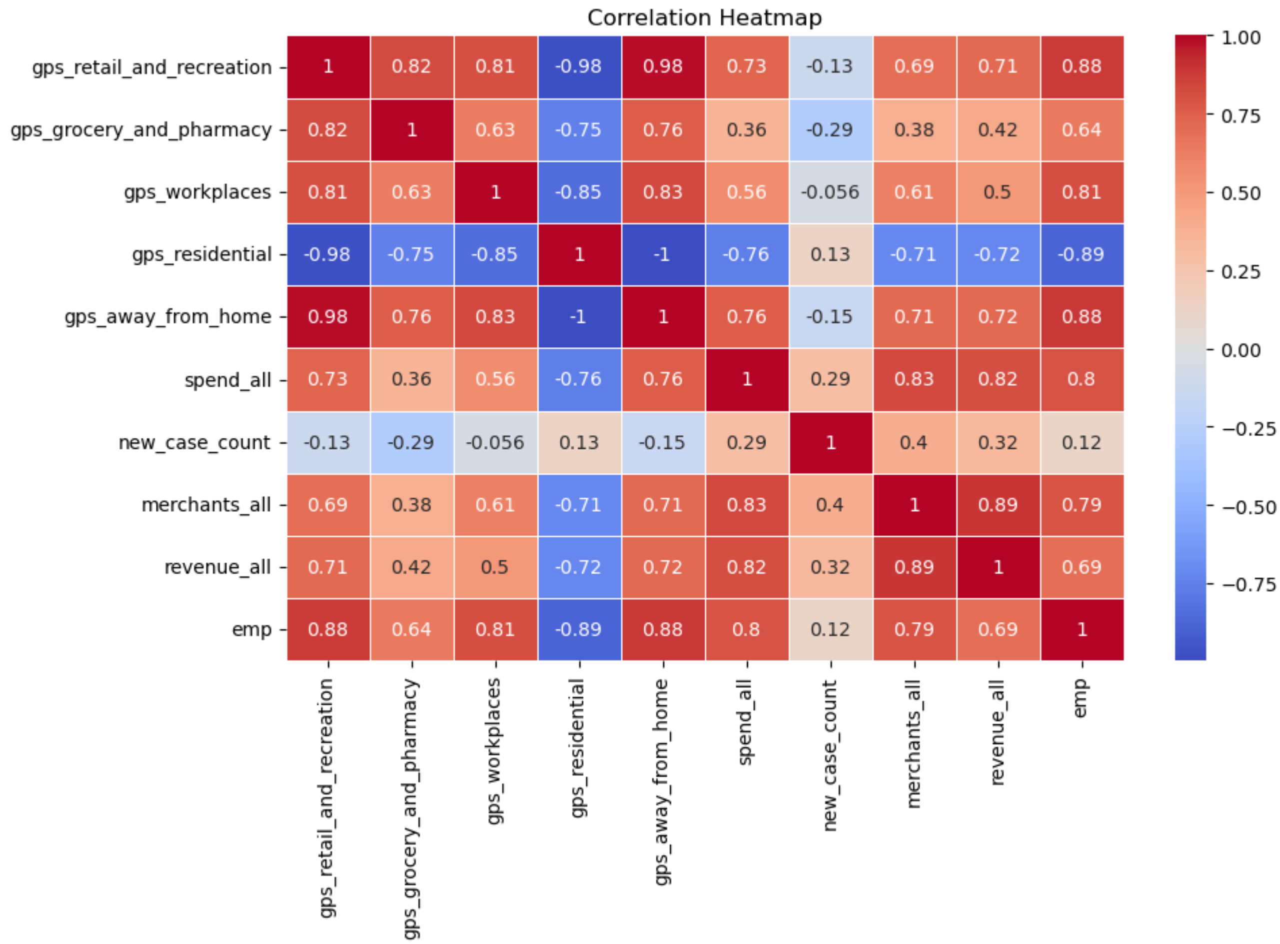
Just after the beginning of the pandemic in April 2020, the employment rate went down to close to a similar point in the case of all the variables small cities with a population of under 5 million had a figure of -26.9%, whereas large cities with a population of over 5 million was –25.1% and the overall population was –23%. In the course of time, small cities started to recover very quickly as the employment rate increased greatly from its lowest peak but did not stabilize until September 2023. On the contrary, the larger cities recovered slowly with not such a significant increase from their lowest peak but thereafter a more consistent trend of stability in employment rate can be seen compared to smaller cities. There is a slight downward trend in the employment rate for the last couple of months of this year in every city regardless of the population size. However, over this 42-month period the employment rate has never been on the same level in the case of any of the variables so not much can be ascertained from that particular small decline.

3.3 Infection Rate and Employment Rate by City



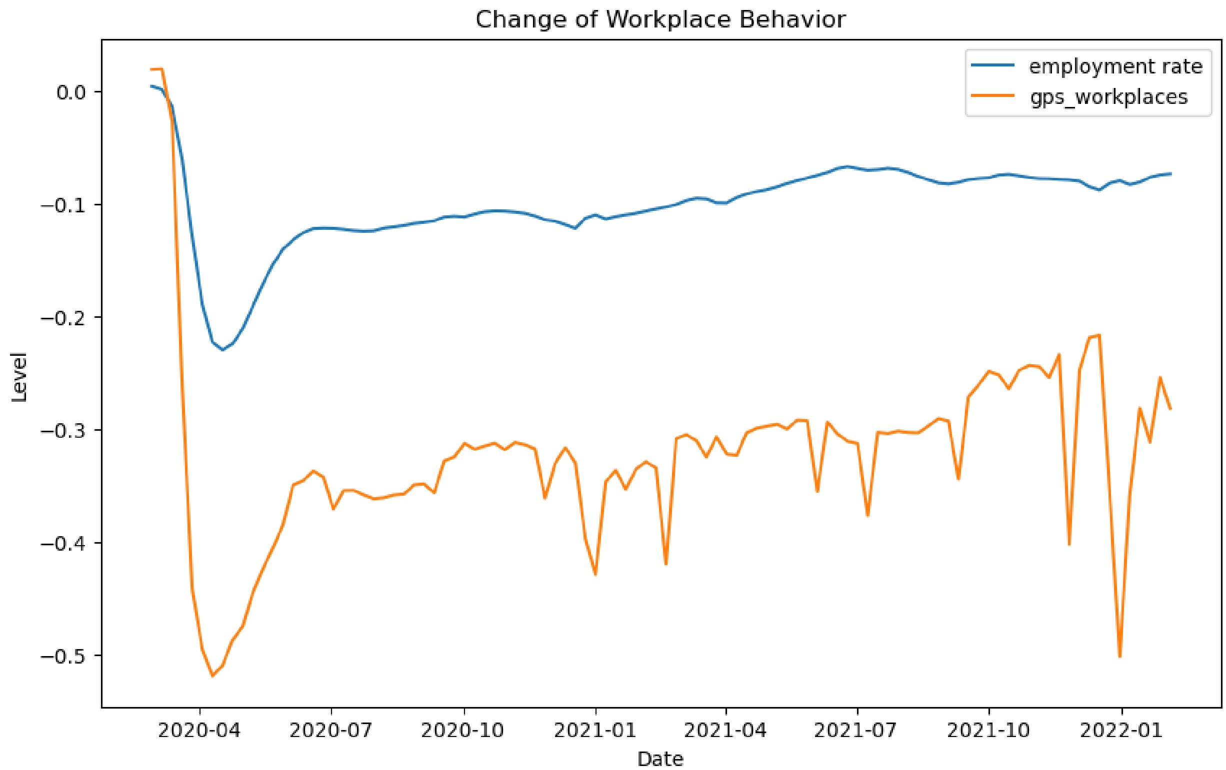
In the above figure, a correlation between employment rate and infection rate in two distinct types of cities was analysed. Los Angeles, New York City and Chicago are cities with a population size of more than 5 million. On the other hand, Boise, Virginia Beach and New Orleans are cities with a population size less than 0.5 million. It can be inferred from this graph that no matter the degree of infection rate regardless of the city size it still has a resultant significant effect on employment rate change. Thus, a large degree of infection rate for a particular city won't have a notable impact on the difference of employment rate change compared to a city with a smaller infection rate. This is because the impact of Covid-19 on employment rate change is only noticeable nationwide and so anlaysing its effect on a city-city basis won't result in a noteworthy correlation between the infection rate and employment rate change.

3.4 Correlation Heatmap



Whilst observing this heat map it can be seen that gps_retail and_ recreation (time spent at retail and entertainment locations) and emp (employment rate) have a correlation coefficient of 0.88. This shows the variables are directly proportional as there is a positive correlation and illustrates there is a strong relationship between these two variables as the value is close to 1. The gps_ residential (time spent at residential locations) and emp variables have a correlation coefficient of -0.89, this shows the variables are inversely proportional as there is negative correlation and depicts that there is a strong relationship between them as the value is close to –1. Similarly, there is also a strong negative correlation between gps_residential and gps_workplace (time spent at places of work) as it has a value of -0.85. There is a very weak correlation between new_case_count (New confirmed COVID-19 cases) and gps_residential having only a value of 0.13. The correlation between the percent change in net revenue for small businesses (revenue_all) and percent change in number of small businesses open (merchants_all) is also extraordinarily strong with a value of 0.89.

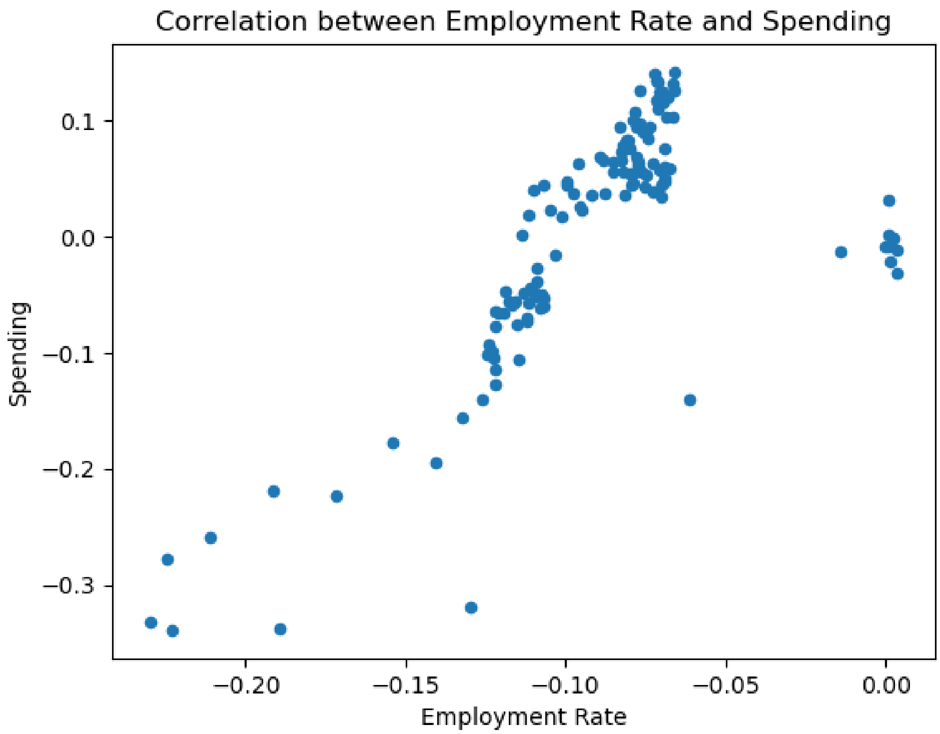
3.5 Change of Workplace Behavior



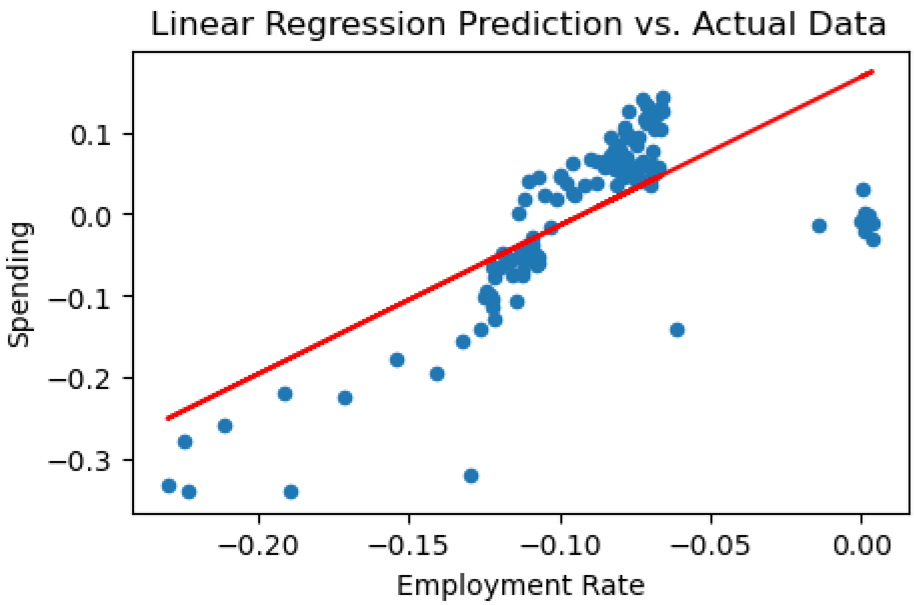
The above line graph is a detailed demonstration of the relationship of two variables from the heatmap which showed a strong positive correlation. Similarly, It can be seen from this line graph that there is also a strong positive correlation between the employment rate and gps_workplace holistically. However, at some points like January 2021 and January 2022 it is showing a somewhat negative correlation as the employment line is relatively flatter than the gps_workplace line which is fluctuating greatly.

Chapter 4- Supervised Learning

The objective of this section is to predict values based on independent and dependent variables. The scatterplot below demonstrates the relationship between weekly employment rate and spending. It illustrates that there is a strong positive correlation between these two variables.



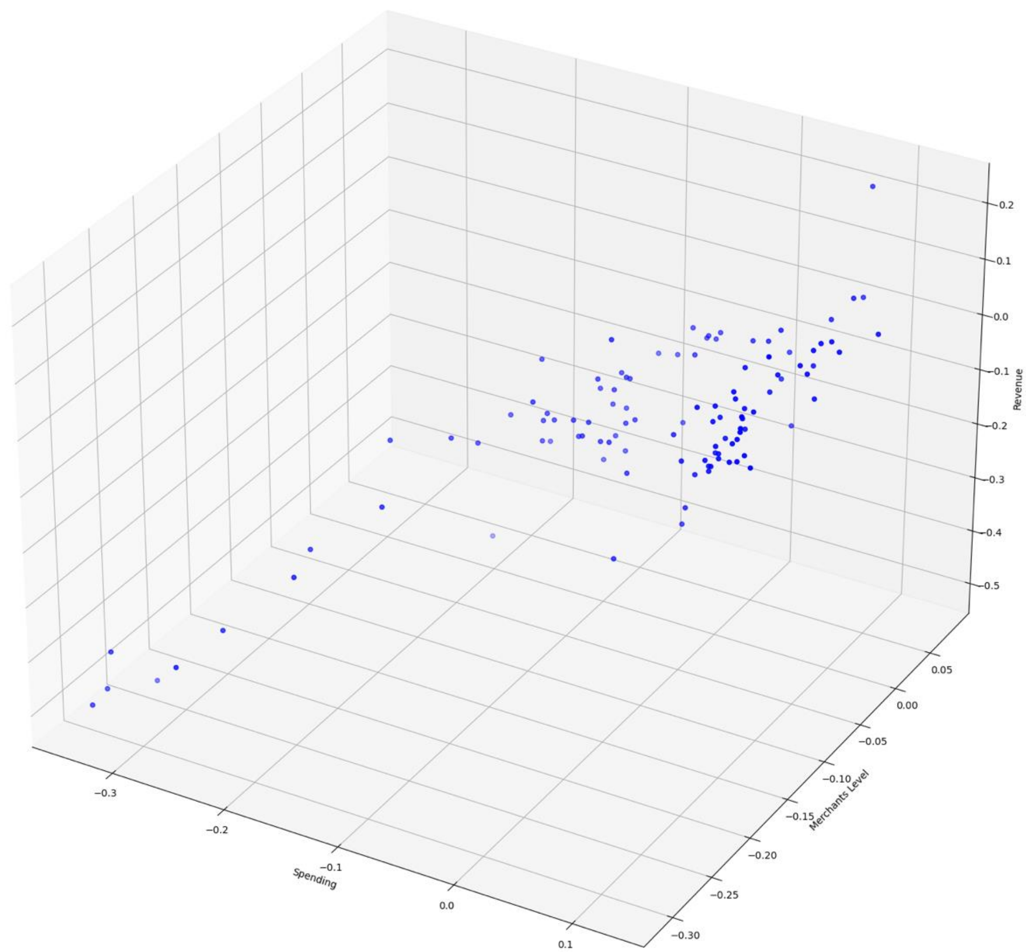
The linear regression approach from supervised learning was used as it's suitable for predicting independent and dependent variables.



The scatter plot above shows the correlation between the daily employment rate and spending. By applying linear regression, the following values are obtained:

- Model Coefficient: 0.4524886877685877
- Model Intercept: 0.0017391592763725122.

These two values illustrate the red line which represents the predicted relationship between the two variables which is that as the daily employment rate increases spending will also increase



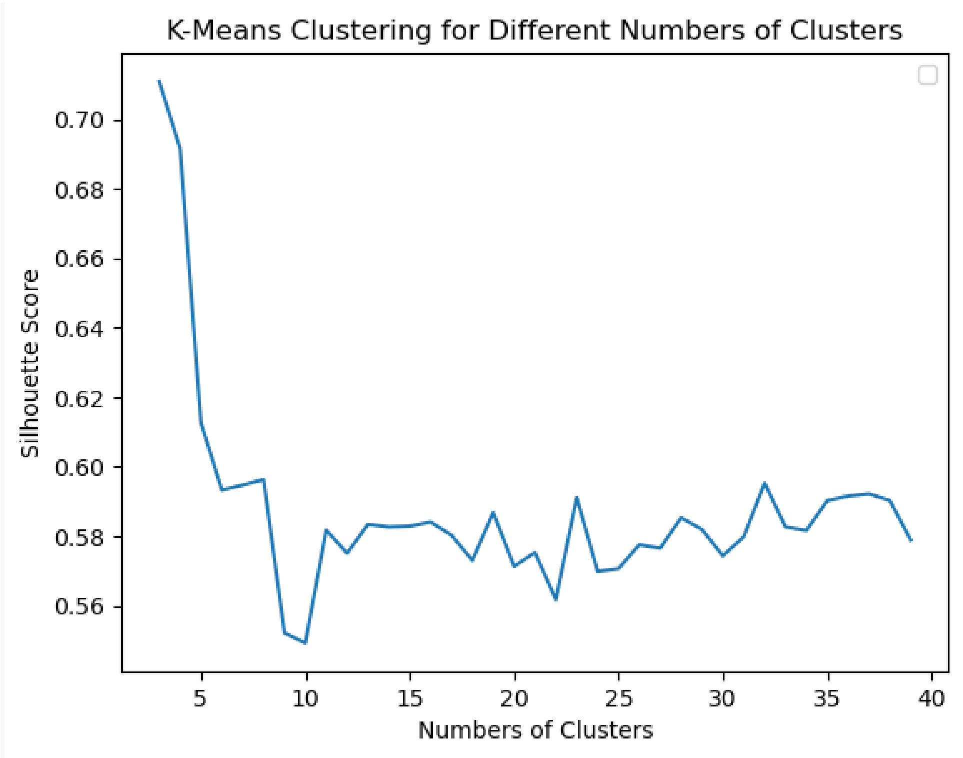
In order to show the correlation between three variables, such as spend_all, merchants_all, and revenue_all, a three-dimensional plot was created. As these plots may not fully represent the correlation when there are more than three variables used and the relationship between these variables are not able be properly visualised, therefore R squared, and Mean Squared values are introduced to accurately reflect the differences between the actual data and predicted results. The following results are obtained after splitting the data of merchants_all, revenue_all and gps_mobility data into a 7:3 ratio and contrasting the predicted results with the original results:

- Mean Squared Error: 0.0020895035738303703
- R-squared: 0.7280495082254383

Chapter 5- Unsupervised Learning

The objective in this section is to find the grouping of the behavior of the time people spent by clustering all the data for GPS mobility. The data used is on a continuous scale therefore it won't be suited to unsupervised clustering method.

The K-means clustering method was applied using different numbers of labels, to classify the data. K-means clustering is performed by 2 to 40 clusters on the data of gps_mobility and this is reflected in the graph below. The completeness and homogeneity are not able to be calculated as the original dataset that was utilised is not labelled.



The silhouette score is a metric to assess the performance of clustering algorithms which lies in a range from –1 to +1 where a higher value is an indication of well match of an object to its own cluster. The figure above shows that the silhouette score is lowest at the number 10 label. In between the 12 and 18 labels, a temporary stability is noticed around the silhouette score 0.58. However, a wavering is noticed in the subsequent ranges which were consistently around the score of 0.58.

Chapter 6- Reflections

6.1 Key Challenges

Data Integration was the initial challenge for the analysis process as there were multiple datasets with various dimensions of data among the datasets. As a result, to achieve the required goal and proper judgement, customization of content and integration was carried out. Sample results were examined with data from different dimensions and the results selected among these samples were based on the data results from them. Data visualisation was the next challenge. Several types of data visualisation graphs were created as the analysis advanced. In most cases, these visualisations were derived from initial assumptions. Some of these accurately represented data characteristics are not beneficial as relevant evidence for the analysis, thus, a subset of visualisations was considered for the analysis that are strongly believed most relevant with the objective.

6.2 Conclusions

The objective of this analysis to get an overview of the effects of Covid-19 on economic factors in the U.S. for the duration of the pandemic. In addition, to also ascertain how the economic impact from Covid-19 resulted in the change of people's behavior during this time. The first conclusion that can be drawn from this analysis from the range of graphs is that there is a correlation between the factors analysed. Moreover, it was found the linear regression from the supervised learning method was relatable for this investigation as the dataset utilised is on a continuous scale, thus the output obtained is closer and more representative to the actual value. Whilst using the dataset for the unsupervised learning method was not truly relatable for the purpose of this investigation due to classification purposes, certain connections can be drawn from the results of this method that will be able to give accurate and understandable evaluation with conducting more analysis Finally, after carrying out this investigation, it can be seen that there is indeed a correlation between the Covid-19 pandemic and certain economic factors.

Appendix A

Environment and Packages

Language: Python 3.11.4

Jupyter:6.5.4

Packages used:

- pandas
- numpy
- seaborn
- matplotlib.pyplot
- sklearn.linear_model
- sklearn.cluster
- sklearn.metrics
- sklearn.model_selection
- sklearn.linear_model
- mpl_toolkits.mplot3d

Bibliography

Harris, B. and Sinclair, T. (2023). *The U.S. Economic Recovery in International Context*. [online] U.S. Department of the Treasury. Available at: https://home.treasury.gov/news/featured-stories/the-us-economic-recovery-in-international-context-2023 [Accessed October 18th, 2023]

Hlávka, J. and Rose, A. (2023). *COVID-19's Total Cost to the U.S. Economy Will Reach $14 Trillion by End of 2023*. [online] USC Schaeffer. Available at: https://healthpolicy.usc.edu/article/covid-19s-total-cost-to-the-economy-in-us-will-reach-14-trillion-by-end-of-2023-new-research/#:~:text=Revenue%20from%20air%20travel%2C%20indoor [Accessed October 18th, 2023]

ONS. (2022). *Coronavirus and the social impacts on Great Britain - Office for National Statistics*. [online] www.ons.gov.uk. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandwellbeing/bulletins/coronavirusandthesocialimpactsongreatbritain/1april2022. [Accessed October 17th, 2023]

Powell, A., Francis-Devine, B. and Clark, H. (2022). Coronavirus: Impact on the Labour Market. *House of Commons Library*. [online] Available at: https://commonslibrary.parliament.uk/research-briefings/cbp-8898/#:~:text=The%20number%20of%20people%20in [Accessed October 17th, 2023]

Rasheed, M. and Holzheu, T. (2023). *US economic outlook: a cooler climate as the US economy enters Q4*. [online] Swiss Re Institute. Available at: https://www.swissre.com/institute/research/sigma-research/Economic-Outlook/us-economic-outlook-september-2023.html [Accessed October 18th, 2023]