---

title: Analysis - Covid with other factors influencing unemployment

rate

author: "Jaya Johnson"

date: "4/19/2021"

output: pdf\_document

---

\*\*Summary of Proposed Analysis\*\*.

In this section we are analyzing the impact of various factors that may or may not contribute to the overall unemployment rate. Some of the contributing factors are:

1) Race (we are taking into consideration the unemployment rate of African Americans, Hispanic Americans, White Americans, Asian Americans).

2) Gender (unemployment rate for males and females).

3) Overall monthly job loss percent.

4) Monthly Covid Numbers.

As a part of this analysis we would like to understand the relationship between the different contributing factors and employment rate.

We would take the most influential parameters, come up with different hypothesis and come up with a multiple regression analysis to predict the unemployment rate.

\*\*Data Analysis\*\*

```{r}

library(readr)

Covid\_and\_Unemployment\_US <- read\_csv("Covid and Unemployment US - Sheet1.csv")

```

```{r}

unemployment\_df <- data.frame(Covid\_and\_Unemployment\_US)

unemployment\_df[is.na(unemployment\_df)] = 0

unemployment\_df$Date <- as.Date(unemployment\_df$Date , format = "%d/%m/%y")

head(unemployment\_df)

```

```{r}

library(funModeling)

library(tidyverse)

library(Hmisc)

```

```{r}

print("Raw Data Metrics")

status(unemployment\_df)

```

```{r}

plot\_num(unemployment\_df)

```

```{r}

library(ggplot2)

library(dplyr)

```

```{r}

library(cowplot)

gp1<-ggplot(data = unemployment\_df, aes(x = Date, y = Unemployment.Rate)) +

geom\_line(color = "#fc03f0", size = 1) +

xlab("") +

ylab("Unemployment Rate")+

theme(

axis.title.y = element\_text(color="blue", size=6, face="bold"),

)

gp2<-ggplot(data = unemployment\_df, aes(x = Date, y = Unemployment.Rate.AfricanAmerican)) +

geom\_line(color = "#FC4E07", size = 1) +

xlab("") +

ylab("UnemploymentRate - African Americans")+

theme(

axis.title.y = element\_text(color="blue", size=6, face="bold"),

)

gp3<-ggplot(data = unemployment\_df, aes(x = Date, y = Unemployment.Rate.White)) +

geom\_line(color = "#8403fc", size = 1) +

xlab("") +

ylab("UnemploymentRate - White")+

theme(

axis.title.y = element\_text(color="blue", size=6, face="bold"),

)

gp4<-ggplot(data = unemployment\_df, aes(x = Date, y = Unemployment.Rate.Asian)) +

geom\_line(color = "#036ffc", size = 1) +

xlab("") +

ylab("UnemploymentRate - Asian")+

theme(

axis.title.y = element\_text(color="blue", size=6, face="bold"),

)

gp5<-ggplot(data = unemployment\_df, aes(x = Date, y = Unemployment.Rate.Hispanic)) +

geom\_line(color = "#03e7fc", size = 1) +

xlab("") +

ylab("UnemploymentRate - Hispanic")+

theme(

axis.title.y = element\_text(color="blue", size=6, face="bold"),

)

gp6<-ggplot(data = unemployment\_df, aes(x = Date, y = Unemployment.Rate.Men)) +

geom\_line(color = "#03fcd3", size = 1) +

xlab("") +

ylab("UnemploymentRate - Male")+

theme(

axis.title.y = element\_text(color="blue", size=6, face="bold"),

)

gp7<-ggplot(data = unemployment\_df, aes(x = Date, y = Unemployment.Rate.Women)) +

geom\_line(color = "#98fc03", size = 1) +

xlab("") +

ylab("UnemploymentRate - Female")+

theme(

axis.title.y = element\_text(color="blue", size=6, face="bold"),

)

gp8<-ggplot(data = unemployment\_df, aes(x = Date, y = Percent.JobLosers)) +

geom\_line(color = "#fce303", size = 1) +

xlab("") +

ylab("Percent of Job Losers")+

theme(

axis.title.y = element\_text(color="blue", size=6, face="bold"),

)

gp9<-ggplot(data = unemployment\_df, aes(x = Date, y = CovidCases)) +

geom\_line(color = "#fc6f03", size = 1) +

xlab("") +

ylab("Covid Cases")+

theme(

axis.title.y = element\_text(color="blue", size=6, face="bold"),

)

plot\_grid(gp1, gp2, gp3, gp4, gp5, gp6, gp7, gp8, gp9, labels = c('','','','','','','','',''), label\_size = 4)

```

```{r}

library(viridis)

gp1<-ggplot(data = unemployment\_df, aes(x=Unemployment.Rate, y=Unemployment.Rate.AfricanAmerican)) +

geom\_point(aes(colour=Unemployment.Rate.AfricanAmerican))+

scale\_color\_viridis(option = "D")

gp1

gp2<-ggplot(data = unemployment\_df, aes(x=Unemployment.Rate, y=Unemployment.Rate.Asian)) +

geom\_point(aes(colour=Unemployment.Rate.Asian))+

scale\_color\_viridis(option = "A")

gp2

gp3<-ggplot(data = unemployment\_df, aes(x=Unemployment.Rate, y=Unemployment.Rate.White)) +

geom\_point(aes(colour=Unemployment.Rate.White))+

scale\_color\_viridis(option = "B")

gp3

gp4<-ggplot(data = unemployment\_df, aes(x=Unemployment.Rate, y=Unemployment.Rate.Hispanic)) +

geom\_point(aes(colour=Unemployment.Rate.Hispanic))+

scale\_color\_viridis(option = "C")

gp4

gp5<-ggplot(data = unemployment\_df, aes(x=Unemployment.Rate,

y=Unemployment.Rate.Men)) +

geom\_point(aes(colour=Unemployment.Rate.Men))+

scale\_color\_viridis(option = "D")

gp5

gp6<-ggplot(data = unemployment\_df, aes(x=Unemployment.Rate, y=Unemployment.Rate.Women))+

geom\_point(aes(colour=Unemployment.Rate.Women))+

scale\_color\_viridis(option = "E")

gp6

gp7<-ggplot(data = unemployment\_df, aes(x=Unemployment.Rate,

y=Percent.JobLosers))+

geom\_point(aes(colour=Percent.JobLosers))+

scale\_color\_viridis(option = "F")

gp7

gp8<-ggplot(data = unemployment\_df, aes(x=Unemployment.Rate, y=CovidCases))+

geom\_point(aes(colour=CovidCases))+

scale\_color\_viridis(option = "G")

gp8

```

Analysis:

* Analysis of how the market suffered when the COVID 19 spreaded and how the Dow Jones Industrial average index dropped with the increase spread of the virus.
* Analysis of the stock market indices that were affected by COVID 19 in some countries in the south of US such as Brazil and Argentina and in the north of the US such as Mexico.

Section 3:

\*\*Summary:

Data1:

-The first graph shows the COVID cases for the period for the period from May 2020 till April 2021 in Brazil. The second graph describes how the main stock market index in Brazil changed monthly from February 2020 till October 2020. It was decreasing then started to increase even with the increase of the COVID cases number.

Chart, line chart

Description automatically generatedChart, line chart

Description automatically generated

Chart, histogram

Description automatically generatedChart, histogram

Description automatically generated

The first graph describes the histogram of the Brazil stock market, and the second graph describes the daily increase rate of the COVID cases. For the stock market index, the mean value is 94640 , and it has a multi modal distribution. There are not outliers points. The COVID cases is a timeseries, increasing with time, and has an exponential pattern.

Chart, scatter chart

Description automatically generated

The above graph shows the correlation between stock market index in Brazil and the COVID cases number. The correlation seems to be a positive after a certain value of the cases.

Model:

Call:

lm(formula = Brazil\_Index ~ log(Brazil\_C), data = ALL\_data[log(ALL\_data$Brazil\_C) >

8, ], na.action = na.omit)

Residuals:

Min 1Q Median 3Q Max

-9330.4 -3002.0 561.2 2643.5 10546.0

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 30474.7 2372.9 12.84 <2e-16 \*\*\*

log(Brazil\_C) 4585.4 173.4 26.44 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4036 on 137 degrees of freedom

(252 observations deleted due to missingness)

Multiple R-squared: 0.8361, Adjusted R-squared: 0.8349

F-statistic: 699.1 on 1 and 137 DF, p-value: < 2.2e-16

Chart, scatter chart

Description automatically generatedChart, line chart

Description automatically generated

Chart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generated

Hypothesis: there is a significant correlation between the COVID cases and the stock market index, but there are also a lot of leverage points that affect the model.

Data 2:

The first graph shows the COVID cases for the period from May 2020 till April 2021 in Argentina. The second graph describes how the main stock market index in Argentina changed monthly from February 2020 till October 2020. It was decreasing then started to increase even with the increase of the COVID cases number.

Chart, line chart

Description automatically generatedGraphical user interface, chart, line chart

Description automatically generated

* The first graph describes the histogram of the Argentina stock market, and the second graph describes the daily increase rate of the COVID cases. For the stock market index, the mean value is 40037 , and it has a multi modal distribution. There are not outliers points. The COVID cases is a timeseries, increasing with time, and has an exponential pattern

Chart, histogram

Description automatically generatedChart, histogram

Description automatically generated

The below graph shows the correlation between stock market index in Argentina and the COVID cases number. The correlation seems to be a positive after a certain value of the cases.

Chart, scatter chart

Description automatically generated

Model:

Call:

lm(formula = Argentina\_Index ~ log(Argentina\_C), data = ALL\_data[log(ALL\_data$Argentina\_C) >

6, ], na.action = na.omit)

Residuals:

Min 1Q Median 3Q Max

-7778.2 -2951.7 -225.2 3206.3 9925.8

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 12137 1786 6.798 3.19e-10 \*\*\*

log(Argentina\_C) 2663 160 16.644 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3944 on 134 degrees of freedom

(255 observations deleted due to missingness)

Multiple R-squared: 0.674, Adjusted R-squared: 0.6716

F-statistic: 277 on 1 and 134 DF, p-value: < 2.2e-16

Chart, scatter chart

Description automatically generatedChart, line chart

Description automatically generated

Chart, scatter chart

Description automatically generatedChart

Description automatically generated

Hypothesis: there is a significant correlation between the COVID cases and the stock market index in Argentina, but there are also a lot of leverage points that affect the model.

Data3:

The first graph shows the COVID cases for the period from May 2020 till April 2021 in Mexico. The second graph describes how the main stock market index in Mexico changed monthly from February 2020 till October 2020. It was decreasing then started to increase even with the increase of the COVID cases number.

Chart, line chart

Description automatically generatedGraphical user interface, chart, application, line chart

Description automatically generated

* The first graph describes the histogram of the Mexico stock market, and the second graph describes the daily increase rate of the COVID cases. For the stock market index, the mean value is 38064 , and it has a multi modal distribution. There are not outliers points. The COVID cases is a timeseries, increasing with time, and has an exponential pattern.

Chart, histogram

Description automatically generatedChart, histogram

Description automatically generated

* The below graph shows the correlation between stock market index in Mexico and the COVID cases number. The correlation seems to be a positive after a certain value of the cases.

Chart, scatter chart

Description automatically generated

Model:

Call:

lm(formula = Mexico\_Index ~ log(Mexico\_C), data = ALL\_data[(ALL\_data$Mexico\_C) >

6, ], na.action = na.omit)

Residuals:

Min 1Q Median 3Q Max

-2889.0 -869.4 56.7 787.3 4823.4

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 34308.46 446.33 76.867 < 2e-16 \*\*\*

log(Mexico\_C) 222.83 38.64 5.767 4.37e-08 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1319 on 152 degrees of freedom

(255 observations deleted due to missingness)

Multiple R-squared: 0.1795, Adjusted R-squared: 0.1741

F-statistic: 33.26 on 1 and 152 DF, p-value: 4.368e-08

Chart, scatter chart

Description automatically generatedChart, line chart

Description automatically generated

Chart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generated

Hypothesis: there is a significant correlation between the COVID cases and the stock market index in Mexico, but there are also a lot of leverage points that affect the model.

Data 4:

* The first graph shows the COVID cases for the period from May 2020 till April 2021 in US. The second graph describes how Dow Jones Industrial average index in US changed weekly from February 2020 till October 2020. It was decreasing with the increase spread of the virus.

Chart, line chart

Description automatically generatedChart, line chart

Description automatically generated

* The first graph describes the histogram of the US Dow Jones Industrial average index, and the second graph describes the daily increase rate of the COVID cases. For the Dow Jones Industrial average index, the mean value is 27989 , and it has a multi modal distribution. There are not outliers points. The COVID cases is a timeseries, increasing with time, and has an exponential pattern.

Chart, histogram

Description automatically generatedChart, histogram

Description automatically generated

* The below graph shows the correlation between the Dow Jones Industrial average index in US and the COVID cases number. The correlation seems to be a positive after a certain value of the cases.

Chart, scatter chart

Description automatically generated

Model:

Call:

lm(formula = US\_Index ~ log(US\_C), data = ALL\_data[(ALL\_data$US\_C) >

8, ], na.action = na.omit)

Residuals:

Min 1Q Median 3Q Max

-5258.5 -1850.3 -195.3 1975.2 5878.2

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 22483.54 1350.61 16.647 < 2e-16 \*\*\*

log(US\_C) 381.42 90.79 4.201 8.95e-05 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2755 on 60 degrees of freedom

(381 observations deleted due to missingness)

Multiple R-squared: 0.2273, Adjusted R-squared: 0.2144

F-statistic: 17.65 on 1 and 60 DF, p-value: 8.946e-05

Chart, line chart

Description automatically generatedChart, scatter chart

Description automatically generated

Chart

Description automatically generatedChart, line chart

Description automatically generated

Hypothesis: there is a significant correlation between the COVID cases and the Dow Jones Industrial average index in US, but there are also a lot of leverage points that affect the model