Group Names: Aria Askaryar , Karla Cabrera, Lyba Batla
CPSC 375
Data Science Apache Project

a) RCode:

```
confirmed <-
read_csv("https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_confirmed_global.csv")
deaths<-
read_csv("https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_deaths_global.csv")
beds <- read_csv("C:/Users/Aria.A/Downloads/WHS6_102.csv")
demographics <- read_csv("C:/Users/Aria.A/Downloads/demographics (2).csv")
demographics2 <- demographics %>%select(-`Country Code`, - `Series Name`) %>%
pivot_wider( names_from= `Series Code`, values_from = YR2015)%>%mutate(SP.POP.80UP =
SP.POP.80UP.FE +SP.POP.80UP.MA) %>% mutate(SP.POP.1564 =
SP.POP.1564.MA.IN+SP.POP.1564.FE.IN) %>% mutate(SP.POP.0014=
SP.POP.0014.MA.IN+SP.POP.0014.FE.IN) %>% mutate(SP.DYN.AMRT=
SP.DYN.AMRT.FE+SP.DYN.AMRT.MA) %>% mutate(SP.POP.TOTL = SP.POP.TOTL.FE.IN+
SP.POP.TOTL.MA.IN) %>% mutate(SP.POP.65UP= SP.POP.65UP.MA.IN +
SP.POP.65UP.FE.IN)
demographics_tidied <- demographics2 %>% select(-c(5:16)) %>% rename(Country = `Country
Name`)
view(demographics_tidied)

confirmed2 <- confirmed %>%  select(-Lat, -Long) %>% rename(Country = `Country/Region`)
%>%
  pivot_longer(
    -c(Country,`Province/State`),
    names_to = "date",

    values_to = "confirmed"
  )
sumofconfirmednum <- confirmed2 %>% group_by(Country,date) %>%
summarise(sum(confirmed))
view(confirmed2)

deaths2 <- deaths %>%  select(-Lat, -Long) %>% rename(Country = `Country/Region`) %>%
  pivot_longer(
    -c(Country,`Province/State`),
    names_to = "date",
    values_to = "deaths"
  )
```

```
sumofdeathnum <- deaths2 %>% group_by(Country,date) %>% summarise(sum(deaths))
view(deaths2)

covid_table<- inner_join(sumofdeathnum,sumofconfirmednum)
view(covid_table)

beds2 <- beds %>% group_by(Country) %>% summarize(max(`Hospital beds (per 10 000 population)`))
view(beds2)

covid_table_with_beds <- inner_join(covid_table,beds2)
colnames(covid_table_with_beds)[3] <- "deaths"
colnames(covid_table_with_beds)[4] <- "confirmed"
colnames(covid_table_with_beds)[5] <- "beds"
view(covid_table_with_beds)

covid_table_complete <- inner_join(covid_table_with_beds,demographics_tidied)
view(covid_table_complete)

library(sparklyr)
sc <- spark_connect(master = "local")
myremotedata <- copy_to(sc, covid_table_complete)
mymodel <- ml_linear_regression(x = myremotedata , formula = Deaths ~ Cases + bed)
summary(mymodel)
spark_web(sc)
```

b) Output:

```
> #  Aria Askaryar
>
> confirmed <-
read_csv("https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_
19_data/csse_covid_19_time_series/time_series_covid19_confirmed_global.csv")

-- Column specification --------------------------------------------------------
cols(
  .default = col_double(),
  `Province/State` = col_character(),
  `Country/Region` = col_character()
)
i Use `spec()` for the full column specifications.

> deaths<-
read_csv("https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_
19_data/csse_covid_19_time_series/time_series_covid19_deaths_global.csv")

-- Column specification --------------------------------------------------------
cols(
  .default = col_double(),
  `Province/State` = col_character(),
  `Country/Region` = col_character()
)
i Use `spec()` for the full column specifications.

> beds <- read_csv("C:/Users/Aria.A/Downloads/WHS6_102.csv")

-- Column specification --------------------------------------------------------
cols(
  Country = col_character(),
  Year = col_double(),
  `Hospital beds (per 10 000 population)` = col_double()
)

> demographics <- read_csv("C:/Users/Aria.A/Downloads/demographics (2).csv")

-- Column specification --------------------------------------------------------
cols(
  `Country Name` = col_character(),
  `Country Code` = col_character(),
  `Series Name` = col_character(),
  `Series Code` = col_character(),
```

```
  YR2015 = col_double()
)

> demographics2 <- demographics %>%select(-`Country Code`, - `Series Name`) %>%
pivot_wider( names_from= `Series Code`, values_from = YR2015)%>%mutate(SP.POP.80UP =
SP.POP.80UP.FE +SP.POP.80UP.MA) %>% mutate(SP.POP.1564 =
SP.POP.1564.MA.IN+SP.POP.1564.FE.IN) %>% mutate(SP.POP.0014=
SP.POP.0014.MA.IN+SP.POP.0014.FE.IN) %>% mutate(SP.DYN.AMRT=
SP.DYN.AMRT.FE+SP.DYN.AMRT.MA) %>% mutate(SP.POP.TOTL = SP.POP.TOTL.FE.IN+
SP.POP.TOTL.MA.IN) %>% mutate(SP.POP.65UP= SP.POP.65UP.MA.IN +
SP.POP.65UP.FE.IN)
> demographics_tidied <- demographics2 %>% select(-c(5:16)) %>% rename(Country =
`Country Name`)
> view(demographics_tidied)
>
> confirmed2 <- confirmed %>%  select(-Lat, -Long) %>% rename(Country = `Country/Region`)
%>%
+   pivot_longer(
+     -c(Country,`Province/State`),
+     names_to = "date",
+
+     values_to = "confirmed"
+   )
> sumofconfirmednum <- confirmed2 %>% group_by(Country,date) %>%
summarise(sum(confirmed))
`summarise()` has grouped output by 'Country'. You can override using the `.groups` argument.
> view(confirmed2)
>
> deaths2 <- deaths %>%  select(-Lat, -Long) %>% rename(Country = `Country/Region`) %>%
+   pivot_longer(
+     -c(Country,`Province/State`),
+     names_to = "date",
+     values_to = "deaths"
+   )
> sumofdeathnum <- deaths2 %>% group_by(Country,date) %>% summarise(sum(deaths))
`summarise()` has grouped output by 'Country'. You can override using the `.groups` argument.
> view(deaths2)
>
> covid_table<- inner_join(sumofdeathnum,sumofconfirmednum)
Joining, by = c("Country", "date")
> view(covid_table)
>
> beds2 <- beds %>% group_by(Country) %>% summarize(max(`Hospital beds (per 10 000
population)`))
```

```
> view(beds2)
>
> covid_table_with_beds <- inner_join(covid_table,beds2)
Joining, by = "Country"
> colnames(covid_table_with_beds)[3] <- "deaths"
> colnames(covid_table_with_beds)[4] <- "confirmed"
> colnames(covid_table_with_beds)[5] <- "beds"
> view(covid_table_with_beds)
>
> covid_table_complete <- inner_join(covid_table_with_beds,demographics_tidied)
Joining, by = "Country"
> view(covid_table_complete)
>
> library(sparklyr)
> sc <- spark_connect(master = "local")
```

c) Summary Model:

http://127.0.0.1:4040/jobs/