**Instructions:** Two Excel datasets are provided for the contest, namely Covid-19 and hypothyroidism. The goal of the contest is to estimate the missing values within the data. The Covid-19 data contains 6,000 de-identified patients. The hypothyroidism data contains 517 de-identified patients. Both datasets have missing values so you can use any technique to predict the missing values. You are allowed to use any software for the data analysis. However, I recommend R software since it is a powerful programming language for data analysis and freely available for a personal computer.

**Final Report:** At the end of the contest, I expect a report from you, which can include the summary of data, type of data, number of missing values, your method(s) used to estimate the missing values, etc. Either you can use R to create a professional report or you can create a report manually on the word document and send it to me at [pandiyab@uww.edu](mailto:pandiyab@uww.edu).

If you don’t know R software, it is okay – you can still learn and participate in this competition. Here I included the helpful video links to learn the software.

**Download & Install:**

To download and install R and Rstudio software in your personal computer, use the following links. **Note:** R is the command prompt while Rstudio is user-graphical interface.

**R:** [**https://www.r-project.org/**](https://www.r-project.org/)

**Rstudio:** [**https://www.rstudio.com/**](https://www.rstudio.com/)

**Installing a package:**

To install any package, use the syntax on the console

install.packages(“package name”)

**Reading an Excel file:**

To read an excel file and store in the dataframe, use the following codes:

library(readxl)

df<-as.data.frame(read\_excel("C:/Users/pandiyab/Desktop/MAAContest/Covid-19.xlsx"))

View(df)

str(df) #To view the structure of the data

**DataFrames:**

Explore the dataframe**:** [**https://www.youtube.com/watch?v=9f2g7RN5N0I**](https://www.youtube.com/watch?v=9f2g7RN5N0I)

Subset, extend & sort dataframes:[**https://www.youtube.com/watch?v=Nh6tSD4i4qs**](https://www.youtube.com/watch?v=Nh6tSD4i4qs)

**Writing a file into the folder:**

To write the data in an excel file, use the following codes

library(writexl)

write\_xlsx(data, location with the name of your excel file)

**Create a professional report of your final data:**

library(DataExplorer)

create\_report(your data name)

**YouTube R videos for handling missing values:**

[**https://www.youtube.com/watch?v=sNNoTd7xI-4**](https://www.youtube.com/watch?v=sNNoTd7xI-4)

[**https://www.youtube.com/watch?v=q1IbzRXqrtI**](https://www.youtube.com/watch?v=q1IbzRXqrtI)

[**https://www.youtube.com/watch?v=MpnxwNXGV-E**](https://www.youtube.com/watch?v=MpnxwNXGV-E)

[**https://www.youtube.com/watch?v=toRjEdNf0Hs**](https://www.youtube.com/watch?v=toRjEdNf0Hs)

[**https://www.youtube.com/watch?v=sUAMiAIUhcI**](https://www.youtube.com/watch?v=sUAMiAIUhcI)

**Intro to R videos:**

**Meet R:** [**https://www.youtube.com/watch?v=SWxoJqTqo08**](https://www.youtube.com/watch?v=SWxoJqTqo08)

**Basic Datatypes:** [**https://www.youtube.com/watch?v=hxlHQ2AtLUk**](https://www.youtube.com/watch?v=hxlHQ2AtLUk)

**Vectors:** [**https://www.youtube.com/watch?v=w5dOALbZ9HE**](https://www.youtube.com/watch?v=w5dOALbZ9HE)

**Subsetting Vectors:** [**https://www.youtube.com/watch?v=o\_ldRKmgvHo**](https://www.youtube.com/watch?v=o_ldRKmgvHo)

**Textbook:** [**https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf**](https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf)