Mini Project in R

15/10/2022

Mini Project

This should be an analysis report of the Novel Coronavirus (COVID-19) around the world, to demonstrate data processing and visualization, insights and prediction with the packages you have learned in this course.

Some guidelines:

- You can use **only** those packages which are part of the course.
- The scope of the project will be restricted to the course contents.
- The analysis should be using the data supplied in the code given below. If you are using *any* extra data or secondary data, that should be mentioned clearly in the final document. The data should also be attached during the submission.
- Even though all are working on the same dataset, it is very much possible to analyze the same data in different aspect.
- The deadline for project submission will be **November 19th 2022**. No extension will be given after this date.
- After the submission you are required to give a brief presentation on the objectives and outcomes achieved via the mini project.
- The evaluation will be done on the novelty, objective, insights and proper usage of all packages learned in the course.
- Plagiarism in **any form** will lead to **zero marks** in the project.
- The final document should be a pdf or html document that should be submitted in the institute LMS. No other formats will be accepted. The rmd file you may submit as a zip file.
- The mode of submission is **only** via institute LMS
- The project will be evaluated on the basis of novelty, Objectives, application of concepts learned via hands on session.
- The template for submission can be found as a separate document with the name *miniproject_template*. It is mandatory to stick to the template.
- The dataset can be obtained from the code shown below
- You may extract and use the data from February 2020 to April 2022.

```
library(tidyverse)
library(lubridate)
library(ggplot2)
library(tidyr)
library(dplyr)
filenames <- c('time_series_covid19_confirmed_global.csv',
               'time_series_covid19_deaths_global.csv',
               'time_series_covid19_recovered_global.csv')
```

```
url.path <- paste0('https://raw.githubusercontent.com/CSSEGISandData/COVID-19/',
                   'master/csse_covid_19_data/csse_covid_19_time_series/')
## download files to local
download <- function(filename) {
  url <- file.path(url.path, filename)
  dest <- file.path('./', filename)
  download.file(url, dest)
}
bin <- lapply(filenames, download)
## load data into R
raw.data.confirmed <- read.csv('time_series_covid19_confirmed_global.csv')
raw.data.deaths <- read.csv('time_series_covid19_deaths_global.csv')
raw.data.recovered <- read.csv('time_series_covid19_recovered_global.csv')
```

- In addition to the above data secondary datasets could also be used for your analysis For example, as in : https://github.com/owid/covid-19-data/tree/master/public/data#%EF%B8%8F-download-our-complete-covid-19-dataset--csv--xlsx--json
- Please note that any secondary datasets that is used for the Mini project should be submitted along with the final Rmd file.

Wish you all the very best for the Mini Project !!!