RMBI3010 20/21 Fall Term – Assignment 02

General Information:

- This is an **individual** assignment.
- Each student is required to 1) choose a website to scrape data from it, 2) conduct data analytics on the scraped data, 3) build a dashboard to demonstrate the discovered patterns or business insights, 4) write a report to present your method and findings.
- Each student is required to submit at least **four** files onto **Canvas** (**two .R** files respectively for web scraping and dashboard building, **one .csv** file for the data your extract and **one .pdf** for the written report).

Submission Deadline: **Dec 11 (Friday) midnight**. Late submission will NOT be accepted.
Weighting: 20% in Individual Assignment

Requirements

- You may choose a web source below or any other web source with data that interests you.

1	https://www.timeshighereducation.com/world-university-rankings/2020/world-ranking
2	https://www.monster.com/jobs/search/
3	https://www.imdb.com/search/title/?groups=top_1000&sort=user_rating
4	https://www.hkp.com.hk/en/list/transaction
5	https://www.hkex.com.hk/Market-Data/Securities-Prices/Equities?sc_lang=en

- You need to write R code to scrape at least 1,000 rows from the data source (cannot directly use a CSV/Excel/or any other data file). Please respect the web site's rules in its robots.txt when scraping, and do not overburden the website.
- For the ease of grading, you need to write the scraped data into a .csv file and submit it onto Canvas.
- You do not have to use all features from the downloaded data (which attributes of the data to use is dependent on your analytics process).
- You may use *flexdashboard* or *shinydashboard* packages, and the dashboard you build could be either static or dynamic, but dynamic ones are preferred.
- You don't have to build a machine learning or data mining model from the dataset, but the information/pattern/insights shown in the dashboard needs to make business sense or carry some business value.

Marking Scheme

- R Coding (10%)
  - Data scraping (e.g., regular expression, Xpath expression, etc.)
  - Data cleaning (e.g., missing values, normalization, categorical to numeric transformation, etc.)
  - Feature selection step, if any
  - Graphic design of the dashboard
- Report (10%)
  - Why choosing the web site/data source
  - Findings from the data analysis and their potential business values
  - Dashboard design consideration
  - Future extension of the work