City University of Hong Kong
The Department of Management Sciences

MS 4252 Big Data Analytics
(2024 Spring)

Project, due on **11:59pm April 19, 2024**

TASKS

1. Form a group of at most five students.
2. Find an interesting data set, analyze it, and illustrate your results. Any statistical methods taught in the lecture notes (text mining and social network) are allowed. You can use SAS, R, or Python.
3. Prepare a PPT and present it during April 9th, or April 16th.
4. Write a short report to summary the results.
5. Submit the PPT and report on the due day.

GUIDELINES OF REPORT

File Submission:

1. Upload your report, program code and data file on Canvas by the deadline.

Basic Report Formatting Rules:

1. Page limit: at most 8 pages main content (excluding the cover page and the appendices; at most 10 pages for appendices.)
2. Font size: use 12-point type in the main text of the report.
3. Spacing: 8 points of lading (about 1 one-and-a-half spacing).
4. Margins: left and right margin should be l" (2.5 cm) on both sides.

HOW TO WRITE REPORTS

Here are some suggestions for you to write your reports.

1. **Project Outline**
    - Title Page: Must contain
        - Project title.
        - Name of group members and group number.
        - Date.
        - Executive Summary. (very short and to the point).
          The executive summary contains a brief account of your conclusion. Write the results not the description of the problem. Say: I found that the IBM price stock was at a higher level in October than in November, in October the mean price was .....

    - Introduction: If any, must be very short. Describe your project and give the background information. Try to be brief but at the same time do not leave out the relevant information. Here you include the basic descriptive statistics, graphs and summaries of the data. Please if you have several graphs or large tables put them in an appendix and refer to them on the text.

    - Analysis & Results: State your points or hypotheses and prove them or disprove them. Go point by point showing performing the corresponding hypothesis tests and interpreting the results. If you use a computer software that generates a sizable output file, put the output file as an appendix and refer to the output here for performing your hypothesis tests.

    - Conclusion: Like the executive summary but longer.

    - References: List of books or articles that are cited on the text. If you do not cite any then omit it (you are not required to have it).

    - Appendix with computer output, computer generated or hand- made graphs. Try to underline and number the places in the output that are quoted in the above text and use those numbers to refer to that piece of output in the text.

2. **Writing your report**
    - Spelling and grammar are very important.

    - Paragraph Structure: When you write a paragraph try to be organized. A way to do this is to start by saying the thesis, then follow with a list of arguments that prove it, quoting the output, performing hypothesis tests, etc. Finally reinforce the result.
        - Say the thesis.
        - Prove it.
        - Say it again.

    - Another way to construct paragraphs is by using concatenation: from A prove B, from B prove C, and so on.

    - The reports that you will write in this class will be no more than 3 or 4 pages of text plus a few more of output. The output should be organized in tables + charts, but never append a thick programming printout. However, a reasonable report should be as brief as possible.

3. **Graphs and charts are essential to display information**
    - Graphs and charts are essential to display information.