INT303 Big Data Analytics
Coursework and marking schema
Semester 1, 2020-2021

1.   General Information

A. This coursework is the replacement of the closed/open book final examination due to covid19.

B. It worth 70% of the total marks for this module and there is no resit and retake.

C. This coursework will require approximately 1200 hours (number based on group of 7 students – individual contribution around 150 -170 hours (three to four full weeks).

D. You will be penalized for late or non-submission according to Xi'an Jiaotong- Liverpool University regulations.

E. You are reminded of the need to comply with Xi'an Jiaotong-Liverpool University's guidelines to Academic Integrity.

F. Reports should be written in Word or PDF.

G. An electronic copy of each submission should be submitted to Learning Mall, and a hard copy of each submission should be submitted to Dr. Gangmin Li's submission box (located at the 4th floor of SD) using the official Coursework Submission Cover Sheet.

2.   Learning Outcomes Addressed

A. Demonstrate a solid understanding of processes and issues related to Big Data Analytics;

B. Identify applications of BDA that can help improve business operations;

C. Determine the appropriate use of technologies, tools, and software packages to support data analysis involving practical scenarios;

D. Be proficient with at least one data analytics software package.

3.   Feedback

The formal feedback for this assignment will be available after the exam board at the end of the semester. However, informal feedback can be provided during the lab sessions or tutorial time if you are asking.

4.   Instructions

This is a group assignment. Each group will consist of **a maximum of 7** students within your lab practice group. You can choose to work on your own with a risk of overload works due to new safety regulations. As part of this exercise you must look into the literature, textbook, and all the extra materials provided in the source. You need identify a problem and a relevant dataset, choose a suitable methodology and tools, analyze the data against the problem, and make meaningful conclusions through a clearly

identified and traceable process.

You are expected to work in groups but you should write report individually. You are expected to submit **two reports** and **one video recording of your presentation.**

The **three components** will be individually marked based on **100 marks.** Their percentage on the final mark is indicated in the braces.

If you choose work on your own, you will have no students peer review marks. The 10 percent of your peer review marks will be awarded by the TA who is responsible for your lab group and based on you general performance.

If you are join a team, you are required to provide one individual 100 marks for your fellow team mates.

Your personal section in the final project report to assess your fellow team mate's contribution in 100 marks

4.1 Report No. 1 -- Project Proposal (20% of 100 marks)

Prepare a three-page document detailing your plan. This does not need to be too detailed, but needs to at least contain:

1. Individuals in the group.
2. Details about the problem:
    A. What the problem is.
    B. Why the problem is interesting – refer to the literature.
    C. Relevant work – refer to the literature.
3. Information about the data source:
    A. What data you plan to use.
    B. Where you plan to get it from.
4. Proposed methodology (including subtasks, methods used in the analysis, tools, processes etc.).
5. Final evaluation methods and criteria.
6. Potential limitations and challenges.

Please note that it is quite likely that the instructor will provide feedback and alter or modify your proposed plans. This can either happen during the lab sessions or will come in feedback on a specific proposal.

See below for a list of 50 useful data sources: https://learn.g2.com/open-data-sources

Deadline: 11th of December 2020, 23:59 (Beijing Time)

4.2 Report No 2 -- Final Report (60% of 100 marks)

Your final report should be **seven** pages long. However, you will be allowed an unlimited number of additional pages for references and appendices. This needs to contain at least the following:

1. Explain the problem and motivation. You can borrow some material from your proposal if you have not changed your plan.

2. Explain what data you explored, where it came from, and how you understand it.

3. Explain what you did to preprocess with the data. You should present the ideas in words instead of cut- paste your codes.

4. Explain what method you used to analysis the data.

5. How do you interpret your result.

6. Discuss limitations of your approach.

7. Discuss what you would do differently in the future.

A separate section in your final report addresses your personal contribution and learning. The section must not be longer than **500 words** and **must be attached to the Final Report**. It must contain a discussion of the following issues from your individual perspective.

a. A personal perspective of the problem addressed by the project. This may include discussing the aims of the project, the benefits the project might offer and to whom if it were successfully carried out, and the main benefits that pursuing the project would offer the team members.

b. A critical appraisal of the methods applied to address the problem and manage the project. This may include an evaluation of the methods used from your perspective, a justification of why particular approaches have been taken and why alternatives have not been used, a discussion of how effective the selected approaches were and if, in hindsight, alternatives might have been better, a discussion of how changes in the project plan and execution have been handled and if the project execution progressed as initially expected, an evaluation of how successful the project was and whether the results could be used to continue the work, and suggestions for improving the approach to solving the problem and the project management. Write this from your personal perspective but relate this to the overall project.

c. Reflections on your contributions to the project. Discuss which parts of the project you carried out/contributed to, how you approached these tasks and how you interacted with other members, both in sharing your results and in organizing the team's activities. Also consider how your personal experience of the project compared to your expectations and experience before you started the project, how well your existing skills were utilized and what new skills you have learnt. Justify why you should get the full percentage of the project mark.

d. Lessons learnt and advice for the future. Discuss what lessons you learnt from executing the project about your discipline, project management and

teamwork. Consider how and where you might apply this in the future. The content of your report must reflect the main items above. It is up to you how to structure the report. You must write a reflective report about your group project from your individual perspective. Whilst the above details are intended to help you with deciding what to put in the report, they are not necessarily complete nor should be used as section headings.

e. **Peer review.** Briefly discuss the contributions of each member and give a grade for each of your team members (100 marks maximum).

Note, the report must not copy material from any of the team reports but may refer to the content of the team reports.

Deadline: 27th of December 2020, 23:59(Beijing Time)

4.3 Presentation (10% of 100 marks)

You are required to submit a video recording of your presentation no more than 10 minutes with any recoding software. I hope to see the following elements in your talk.

1. What is the problem and data you worked on?
2. What were the key ideas in your approach?
3. What techniques did you use?
4. What conclusions you came up with?
5. How do you evaluate and interpret your result?
6. What did you learn from the project?

Deadline: 5th of Jan 2021, 23:59 (Beijing Time)

5.   Marking Criteria for the Assessment

Credit will be awarded based on marker's academic judgment based on the following criteria.

5.1 Reports:

A. **Report No 1 – Project Plan:**

1. How well you understand the problem and the base data?

2. How well you understand the analytical problem is a detective work?

3. How well you understand data analyzing is to find insight of data?

4. How well you understand the solution is never 100% satisfactory?

B. **Report No 2 - Final report:**

1. How well you can demonstrate you understand the data process

procedure?

2. How well you can assess data quality and quantity?

3. How well you can manage data preparation methods?

4. How well you can chose from many data analysis methods and use one you choose well?

5. How well you can make data analyzing report?

6. How well you can interpret your results?

7. Did you consider legal, social, ethical and professional issues to justify your choices and evaluate their results?

8. How well justified conclusions and concise discussion of future work?

5.2 Presentation:

1. Coherent and detailed presentation of the solution to the problem, focusing on major challenges of the particular tasks in the project.

2. Evidence of testing and evaluation with clear statements of what the solution is provided and what conclusion is reached.

3. How well you did in the project in comparison with other students?

4. How well you can work with teammates and you can appreciate others contribution?

6. Overall Mark Calculation

Your final mark will be calculated based on the following formula:

Final mark = Mark_Report1 × 0.2 + Mark_Repor2 × 0.6 + M_Video_pre × 0.1 + Average (student mark_project) × 0.1