Xi'an Jiaotong-Liverpool University
西交利物浦大學

XJTLU Entrepreneur College (Taicang) Cover Sheet

Module code and Title	**DTS303TC Big Data Security and Analytics**
School Title	**School of AI and Advanced Computing**
Assignment Title	**Final Coursework**
Submission Deadline	**Wednesday, October 19th 23:59,2022 (China Time, GMT + 8)**
Final Word Count	**N/A**
If you agree to let the university use your work anonymously for teaching and learning purposes, please type **"yes"** here.	

I certify that I have read and understood the University's Policy for dealing with Plagiarism, Collusion and the Fabrication of Data (available on Learning Mall Online). With reference to this policy I certify that:

- My work does not contain any instances of plagiarism and/or collusion.
  My work does not contain any fabricated data.

By uploading my assignment onto Learning Mall Online, I formally declare that all of the above information is true to the best of my knowledge and belief.

Scoring – For Tutor Use	
Student ID	

Stage of Marking	**Marker Code**	**Learning Outcomes Achieved (F/P/M/D) (please modify as appropriate)**			**Final Score**
		A	**B**	**C**	
1st Marker – red pen					
Moderation – green pen	**IM Initials**	The original mark has been accepted by the moderator (please circle as appropriate):			Y / N
		Data entry and score calculation have been checked by another tutor (please circle):			Y
2nd Marker if needed – green pen					

For Academic Office Use			**Possible Academic Infringement (please tick as appropriate)**	
Date Received	**Days late**	**Late Penalty**	☐ **Category A**	Total Academic Infringement Penalty (A,B, C, D, E, Please modify where necessary) _____
			☐ **Category B**	
			☐ **Category C**	
			☐ **Category D**	
			☐ **Category E**	

Xi'an Jiaotong-Liverpool University
西交利物浦大學

DTS303TC Big Data Security and Analytics

Coursework 1 – Group Assignment

Submission deadline: 23:59, October 19th, 2022

Percentage in final mark: 40%

Learning outcomes assessed: A, B

Individual/Group: Group (max 3 students in a group)

Length: Group Report 5000 words (+/- 10%) + Application with Source Code and Recorded Individual Presentation (10 minutes +-10%). The assessment has a total of 100 marks (60 marks for group report and 40 marks for individual recorded presentation)

Late policy: 5% of the total marks available for the assessment shall be deducted from the assessment mark for each working day after the submission date, up to a maximum of five working days

Risks:

- Please read the coursework instructions and requirements carefully. Not following these instructions and requirements may result in loss of marks.
- The formal procedure for submitting coursework at XJTLU is strictly followed. Submission link on Learning Mall will be provided in due course. The submission timestamp on Learning Mall will be used to check late submission.

Task Summary

Conduct a Big data science study in your chosen domain, for example, business, industry, education, environmental, healthcare, etc. Write a group report on your Big data analytics project. The report should be written in a clear and concise manner (and be no more than 5000 words in length). You should start by exploring a domain that interests you. You need to identify a compact dataset (structured or unstructured) with a reasonable large size and number of attributes/variables in your chosen domain which can be used for the assessment. Your report should include the background of the chosen domain and the data analytics problem you attempt to solve, aims and objectives, significance of your study, and describe your analytics approach including the statistical method(s) and/or machine learning technique(s) you used to address the

problem. Members in the team are required to submit an individual recorded video presentation to the Mediasite or other source which will be informed before the submission date.

Context

In the Big data science process methodology, data scientists need to identify a data science problem from the chosen domain or industry and the approaches to the problem. After understanding the problem, data scientists need to define the analytic approach to solve the problem. They need to use statistical method(s) and/or machine learning to solve the data science problem. Data scientists also need to have the knowledge of different stages of the Big data life cycle, data analytics skills, some tools and programming languages (Python is one of the common and popular coding languages required in data analytics roles). Data scientists should be able to report and discuss their findings after performing data analytics. They also have the capability to provide an explanation or clarification needed for some aspects of the data model and the analysis results.

This assessment has been designed to help you build the necessary skills to achieve the following learning objectives to fulfil the learning outcomes of this module. After completing this assessment, you should be able to:

- demonstrate a solid understanding of concepts, processes and issues related to Big Data Analytics; and
- determine the appropriate use of analysis methods, algorithms, technologies, tools, and software packages to support data analysis involving practical scenarios;

By completing this assessment item, you will acquire the knowledge of Big data life cycle, data analytics and programming skills in Python to analyse the data from a chosen domain. You will also acquire the presentation skills necessary to present the analysis of the results in your report and recorded video to your audiences. This assessment will prepare you to address a Big data science problem in the real world.

Task Instructions

(1) Write a short group project proposal to describe your Big data analytics project. Your project proposal should be written in a clear and concise manner (no more than 800 words or 1-page A4 size). You start by exploring an area or domain which interests you. The project topic can be chosen from your favourite domain e.g., business, industry, healthcare, biomedical, etc. Show and discuss your proposal with the Teaching Assistant (TA) during the laboratory sessions. Please note that no mark will be given for this short proposal. However, this short proposal should serve as your first document to plan for your Big data analytics project.

(2) Write a report on your Big data analytics project. The report should be written in a clear and concise manner (and be no more than 5000 words in length). Your final report should be detailed and address the following areas:

- Clearly define the problem definition in your Big data project.
- Describe the significance of your Big data project in the chosen domain or area.
- Identify a compact dataset (structured or unstructured) with a reasonable large size and number of attributes/variables in your chosen domain.
- Highlight the project aim and objectives.
- Discuss the background of your chosen topic in the domain or area.
- Describe the analytics approach used.
- Describe how your analytics approach helped answer the problem and the statistical method(s) and machine learning technique(s) you used.
- Describe all the steps you took to analyse your data.
- Discuss the results of the analysis.
- **Include evidence, such as tables, graphs and plots** from the programming codes, to support your results.

(3) Prepare and record a short individual presentation (10 minutes) to introduce and explain your Big data science project and its significance. Your presentation should list the data science question or problem, describe your analytics approach and the statistical and/or machine learning method(s) you used to address the data science problem. Present and discuss the results of your analysis, and provide evidence (screenshots) from your programming codes to support the results. Your presentation should be clear, should be in no more than twelve PowerPoint slides, and you should not take more than 10 minutes to go through them.

Note: Students MUST use the tools and software packages in the lab sessions to support their data analytics involving practical scenarios.

Additionally, your final report should:

- be clearly structured (with well-organised content); and
- use the APA referencing style and include a reference list at the end.

For this assessment item, you are required to create programs using Python programming language in software packages from your lab sessions to analyse your data. You are also required to submit the programming source codes with the final report. Your programming source codes should be:

- written in Python programming language;

- well commented upon in relation to both the main program and each individual module, such as the function module; and
- free of errors, such as syntax errors, runtime errors, etc.

Report Format

- **Cover Page:** This should include the Assessment Number, Assessment Title, Student Name, Student ID and Student Email.
- **Body of the report:** This should include all the relevant section headings to address each aspect as indicated/highlighted in the question and the marking rubric.
- **References:** Both your in-text and the references included in the 'References' section the end of the report should adhere to the APA style.
- **Glossary (Optional):** This should include any terms frequently used in the report.

The following points are a general guide for the presentation of assessment items:

Assessments items should be typed;

- Use 1.5 spacing;
- Use a wide left margin (as markers need space to be able to include their comments);
- Use a standard 12-point font, such as Times New Roman, Calibri or Arial;
- Left-justify body text;
- Number your pages (excepting the cover page);
- Insert a header or footer that details your name and student number on each page;
- Always keep a copy (both hard and electronic) of your assessments; and
- Most importantly, always run a spelling and grammar check; however, remember, such checks may not pick up all errors. You should still edit your work manually and carefully.

Referencing

It is essential that you use appropriate APA style for citing and referencing research.

Xi'an Jiaotong-Liverpool University
西交利物浦大學

Assessment Rubric (Group – 60 marks)

Assessment Attributes	Fail 0–39%	Adequate 40–49%	Competent 50–59%	Comprehensive 60–69%	Very Good 70–79%	Excellent 80–100%
State the Big data science problem and dataset guiding your study. Briefly describe the analytics approach including the statistical method(s) and/or machine learning technique(s) and how your analytics approach helped answer the problem **Percentage for this criterion = 10%**	A statement of the Big data science problem and dataset and a brief description of analytics approach are not included. The explanation of analytics approach to answer the Big data science problem is not adequate.	A statement of the Big data science problem and dataset and a brief description of analytics approach are included. How the analytics approach helped answer the Big data science problem is to some extent explained.	A statement of the Big data science problem and dataset and a brief description of analytics approach are included. How the analytics approach helped answer the Big data science problem is competently explained.	A statement of the Big data science problem and dataset and a brief description of analytics approach are included. How the analytics approach helped answer the Big data science problem is comprehensively explained.	A statement of the Big data science problem and dataset and a brief description of analytics approach are included. How the analytics approach helped answer the Big data science problem is superbly explained but with modest gaps.	A statement of the Big data science problem and dataset and a brief description of analytics approach are included. How the analytics approach helped answer the Big data science problem is superbly explained.
Describe all the steps you took to analyse your data. **Percentage for this criterion = 10%**	The steps are not described in detail and are not systematic or logical or consistent with the selected data analytics technique.	The steps are described in some detail but are not systematic, logical or consistent with the selected data analytics technique.	A genuine attempt has been made to describe the steps and ensure that they are somewhat systematic, logical and consistent with the selected data analytics technique.	The steps are described in great detail and are clearly systematic, logical and consistent with the selected data analytics technique	The steps are described in full detail and are incredibly systematic, logical and consistent with the selected data analytics technique but with modest gaps.	The steps are described in full detail and are incredibly systematic, logical and consistent with the selected data analytics technique.
Discuss the results of the analysis. Provide evidence, such as tables, graphs and plots from the programming codes, to support your results.	The justification makes no sense. No evidence is provided to support the results in the report.	The justification makes some sense. Weak evidence is provided to support the results in the report.	The justification makes good sense. Sufficient evidence is provided to support the results in the report.	The justification makes great sense. Strong evidence is provided to support the results in the report.	The justification makes perfect sense. Comprehensive evidence is provided to support the results in	The justification makes perfect sense. Comprehensive evidence is provided to support the results in the report.

Percentage for this criterion = 20%					the report but with modest gaps.	
Programming source codes (Python) **Percentage for this criterion = 10%**	Source codes are not commented or are only lightly commented in the main program and for each module, such as functions. Codes contain errors.	Source codes are partially commented in the main program and for each module, such as functions. Codes are completely free from errors.	Source codes are mostly commented in the main program and for each module, such as functions. Codes are completely free of errors.	Source codes are well commented in the main program but are not well commented for each module, such as functions. Codes are completely free of errors.	Source codes are very well commented in the main program and for each module, such as functions (with modest gaps). Codes are completely free of errors.	Source codes are very well commented in the main program and for each module, such as functions. Codes are completely free of errors.
Content writing **Percentage for this criterion = 5%**	Rudimentary skills in expression and presentation of ideas. Not all of the material is relevant and/or is presented in a disorganised manner. The meaning is apparent, but the writing style is not fluent or well organised. Grammar and spelling contain many errors. Formal English is not used.	Some skills in the expression and presentation of ideas. The meaning is apparent, but the writing style is not always fluent or well organised. Grammar and spelling contain several careless errors. Formal English is rarely used.	Sound skills in the expression and clear presentation of ideas. The writing style is mostly fluent and appropriate to the assessment task/document type. Grammar and spelling contain a few minor errors. Formal English more or less used.	Well-developed skills in the expression and presentation of ideas. The writing style is fluent and appropriate to the assessment task/ document type. Grammar and spelling are accurate. Formal English is mostly used.	Highly-developed skills (with modest gaps) in the expression and presentation of ideas. The writing style is fluent and appropriate to the assessment task/document type. Grammar and spelling are accurate. Formal English is used throughout.	Highly-developed skills in The expression and presentation of ideas. The writing style is fluent and appropriate to the assessment task/document type. Grammar and spelling are accurate. Formal English is used throughout.
Uses the APA referencing style and provides a reference list **Percentage for this criterion = 5%**	Substandard (or no) referencing. Poor quality (or no) references.	Evidence of rudimentary referencing skills.	Good referencing in both the reference list and in-text citations. Good quality references.	Very good referencing in both the reference list and in-text citations. High quality references.	Faultless referencing (with modest gaps) in both the reference list and in-text citations. High quality references.	Faultless referencing in both the reference list and in-text citations. High quality references.

Xi'an Jiaotong-Liverpool University
西交利物浦大學

Assessment Rubric (Individual – 40 marks)

Assessment Attributes	Fail 0–39%	Adequate 40–49%	Competent 50–59%	Comprehensive 60–69%	Very Good 70–79%	Excellent 80–100%
Your presentation should be clear, should be in no more than twelve slides, and you should not take more than 10 minutes to go through them. **Percentage for this criterion = 10%**	The presentation is not clear and did not meet the number of slides and time constraints	The clarity of the presentation is OK but the presentation did not meet the number of slides and time constraints	The clarity of the presentation is acceptable and the number of slides and time constraints are sufficiently met	The clarity of the presentation is good and the number of slides and time constraints are closely met	The presentation is generally clear and understandable and precisely met the number of slides and time constraints	The presentation is extremely clear and completely understandable and precisely met the number of slides and time constraints
Your presentation should list the Big data science question or problem, discuss the results of your study, and provide evidence that supports the results. **Percentage for this criterion = 20%**	The results discussed do not address the Big data science question or problem, the results are not described in any detail and evidence is not provided	Some of the results discussed answer the Big data science question or problem, the results are somewhat described but evidence is not very convincing	A genuine attempt is made at ensuring the results discussed are consistent with the Big data science question or problem, the results are described in acceptable detail and some evidence are presented	The results discussed are sufficiently consistent with the Big data science or problem, the results are described in great detail and supported by solid evidence	The results discussed are generally consistent with the Big data science question or problem, the results are superbly described and evidence is exceptionally convincing	The results discussed are entirely consistent with the Big data science question or problem, the results are superbly described and evidence is exceptionally convincing
Brief describe the statistical and/or machine learning technique(s) you used **Percentage for this criterion = 10%**	The selected statistical and/or machine learning technique(s) are not adequately and are not appropriate for the Big data science question or problem	The selected statistical and/or machine learning technique(s) are to some extent described but are hardly appropriate for the Big data	The selected statistical and/or machine learning technique(s) are competently described and is to some extent appropriate for the	The selected statistical and/or machine learning technique(s) are properly described and is fairly appropriate for the Big data science question or problem	The selected statistical and/or machine learning technique(s) are for most parts superbly described and are fittingly appropriate for the Big	The selected statistical and/or machine learning technique(s) are superbly described and are fittingly appropriate for the

		science question or problem	Big data science question or problem		data science question or problem	Big data science question or problem