

INFT6201 – BIG DATA

ASSESSMENT 2: PRESENTATION



OVERVIEW

- Weighting: 30%
- Due date: Ongoing (Weeks 8–11 during Lab)
- Method of submission: Lab Presentation
- Content: Presentation (**individual** assessment)
- Length of submission: Presentation (12 minutes) + Q&A (2-3 minutes)

DESCRIPTION

This assessment encourages students to expand and deepen their conceptual knowledge of big data in real-world applications (e.g., business, health). They achieve this through a discussion of a data analytics concept (e.g., big data frameworks, data visualization, natural language processing) in practice. In their presentation, students are required to provide evidence of extensive research on the concept, drawing from resources such as academic journals, professional press, and popular media. They are expected to demonstrate both reflection and analysis related to the chosen data analytics concept, and to produce a clear and concise response that conveys an evidence-based understanding of the topics.

PRESENTATION

In your presentation, we ask you to provide a practical case study on a specific big data analytics application. Each student (or group) will select their own data analytics project, identifying relevant data and a case study. The presentation should begin with an overview of the project, its significance, and the potential impact or applications it might have. Finding a suitable dataset and identifying relevant applications is a key part of this assessment. Students must outline interesting data science questions they aim to explore, address data quality issues, discuss data modeling approaches, identify key stakeholders, and present data visualisation or communication results.

Your presentation should follow this structure:

1. **Title Slide** (1 slide: Project title and student names)
2. **Datasets** (1-2 slides: Background on the datasets and any data quality issues)
3. **Case Study** (1-2 slides: Specific application examples including interesting data science questions)
4. **Data Modelling** (1-2 slides: Data models and their suitability for the case study)
5. **Data Visualisation/Communication of Results** (1-2 slides: Appropriate data visualisation techniques to communicate results and support the project)
6. **Preliminary Results** (1-2 slides: Include data characterization, distribution, and initial data modelling results)
7. **References** (1-2 slides: Proper citation of data sources, articles, or other publications)

Students aiming for high marks should ensure their presentation aligns with the criteria outlined in the marking rubric. Presentations will be delivered during a lab session. For those unable to attend, alternative arrangements can be made.

MARKING CRITERIA FOR ASSESSMENT 2: PRESENTATION

Criterion	Mark	High Distinction (HD) 5-4.25 pts	Distinction (D) 4.25 – 3.75 pts	Credit (C) 3.75 – 3.25 pts	Pass (P) 3.25 - 2.5 pts	Fail (F) 2.5 – 0 pts
1) Project Ideation: Rationale for the big data analytics project and its relevance to real-world applications.	/5	Clearly articulates the rationale for the big data analytics project, with connections to real-world applications such as (1) business impact, (2) societal impact, and (3) Sustainable Development Goals. Articulation of the impact has been demonstrated with specific examples to illustrate these connections effectively.	Well-balanced explanation of the rationale for the big data analytics project, with connections to real-world applications; with some articulation of the impact, such as (1) business impact, (2) societal impact, and (3) Sustainable Development Goals. Specific examples have not been provided to demonstrate these impacts effectively.	An adequate explanation of the rationale for the big data analytics project, with connections to some real-world applications, though the impact is vague.	A basic explanation of the rationale for the big data analytics project, with connections to some real-world applications.	No or very limited explanation and shallow elaboration of the big data analytics project or its impact.
2) Datasets: Relevance to the Project and Considerations for Data Quality	/5	A clear explanation of the dataset's relevance to the project, with well-articulated justification for why this dataset can help answer the project's data science questions, and a thorough investigation of data quality issues.	A balanced explanation of the dataset's relevance to the project, with either well-articulated justification for why this dataset can help answer the project's data science questions, or a thorough investigation of data quality issues.	An adequate explanation of the dataset's relevance to the project, with some justification for why this dataset can help answer the project's data science questions, and some investigation into data quality issues.	A basic explanation of the dataset's relevance to the project, with no justification for why this dataset can help answer the project's data science questions, and a limited investigation into data quality issues.	No or very limited explanation of the dataset's relevance to the project, and no or limited investigation into data quality issues.
3) Data Modelling: Data models and their suitability for the project	/5	A well-balanced explanation of the model's relevance to the project is provided, with a well-articulated justification of why this model can help answer the project's data science questions, along with a thorough investigation of the model's suitability for the dataset.	A balanced explanation of the model's relevance to the project is provided, with either a well-articulated justification of why this model can help answer the project's data science questions or a thorough investigation of the model's suitability for the dataset.	An adequate explanation of the model's relevance to the project is provided, though with limited justification for why this model can help answer the project's data science questions and a limited investigation of the model's suitability for the dataset.	A basic explanation of the model's relevance to the project is provided, with no justification for why this model can help answer the project's data science questions and only a limited investigation of the model's suitability for the dataset.	A data model has been chosen, but its relevance to the project has not been justified.

4) Data Visualisation and Communication of Results	/5	Types of data visualisations and model performance metrics adopted for the project have been clearly articulated and justified, with connections to the project's applications and its end users.	Types of data visualisations and model performance metrics adopted for the project have some justification with connection to the project's applications and its end users.	Types of data visualisations and model performance metrics adopted for the project either have some justification or show connection to the project's applications and its end users.	A basic justification of the types of data visualisations and model performance metrics adopted has been provided.	No or very limited explanation has been provided regarding the types of data visualisations and model performance metrics that would be adopted for the project.
5) Preliminary Results	/5	A preliminary analysis of data characterization, distribution, and initial data modeling results has been provided, and a clear connection has been established to the data science questions.	A preliminary analysis of data characterization, distribution, and initial data modeling results has been provided, with some connection to the data science questions.	A preliminary analysis of either data characterization, distribution, or initial data modeling results has been provided, with some connection to the data science questions.	A preliminary analysis of either data characterization, distribution, or initial data modeling results has been provided.	No or limited analysis of data characterization, distribution, and initial data modeling results has been provided.
6) References	/5	Excellent 5-3.5 points		Good 3.5-2 pts		Basic 2-0 pts
		High quality and relevant references for the datasets, project idea or applications, scientific articles, white papers, industry outlets, public code repositories, and generative AI tools used are provided and are in a consistent format.		Relevant references for the datasets, project idea or applications, scientific articles, white papers, industry outlets, public code repositories, and generative AI tools used are provided and are in a consistent format.		Some references for the datasets, project idea or applications, scientific articles, white papers, industry outlets, public code repositories, and generative AI tools used are provided and are in a consistent format.