Module:	Data Analytics
Assignment No: 2	Data Analytics Using the CRISP-DM Model
Hand-Out Date:	Friday 17th October 2025
Hand-In Date:	Friday 5th December 2025
Lecturer:	Andrew Shields (andrew.shields@staff.ittralee.ie)

Overview

Working individually, you are required to analyse a dataset using the CRISP-DM (Cross Industry Standard Process for Data Mining) methodology.
Your submission will consist of:

1. A Jupyter Notebook (or Google Colab Notebook) implementing the CRISP-DM process.

2. A short research paper (4 pages, conference-style) summarising your analysis, methodology, and findings.

Requirements

You must demonstrate applied and theoretical understanding of data analytics through:

Notebook file /Code

- Use of **Python**, **Pandas**, **Scikit-Learn**, **Matplotlib/Seaborn**, or other relevant tools.

- A clear application of the **six stages of the CRISP-DM process**:

    1. **Business Understanding:** Define the problem, objectives, and success criteria.

    2. **Data Understanding:** Describe the dataset, perform exploratory analysis, identify issues.

    3. **Data Preparation:** Clean, transform, and prepare data for modelling.

    4. **Modelling:** Apply one or more data mining or machine learning techniques.

    5. **Evaluation:** Assess the results in relation to the original objectives.

    6. **Deployment:** Present findings, discuss limitations, and potential applications.

Each section should include **clear markdown documentation**, tables, figures, and commentary.

Research Paper (4 pages)

You must write a short, academic-style research paper (approx. 2000 words, excluding references) summarising your project. The paper should resemble a **conference publication format** (e.g., IEEE, ACM, or Springer-style layout) and include:

Suggested Structure

1. **Title and Abstract (150–200 words)** – concise summary of aims, dataset, methods, and key results.

2. **Introduction** – outline the problem, its relevance, and objectives of your analysis.

3. **Methodology** – summarise how you applied the CRISP-DM model, highlighting key stages and analytical methods.

4. **Results and Discussion** – present main findings, visualisations, and interpretation.

5. **Conclusion and Future Work** – summarise key insights and propose potential extensions.

6. **References** – minimum of **4 peer-reviewed academic sources**, formatted in **MTU Harvard style**.

Tip: Review conference proceedings (e.g., IEEE Big Data, ACM SIGKDD, or Springer Lecture Notes in AI) to understand formatting and tone.

Submission Components

Submit the following via **Canvas or GitHub/Colab link**:

- **Notebook file** (.ipynb) implementing CRISP-DM.

- **Research Paper (PDF)** in conference format.

- **All supporting datasets or code files** (OneDrive or GitHub if files are large)**.**

Plagiarism

All submissions will be screened for originality. Plagiarism or AI-generated content without attribution will result in penalties under MTU academic integrity policies.

Grading Rubric

Assessment and Grading Rubric (100 Marks Total)

Category	Marks	Description
Active Engagement	10	Evidence of consistent progress and participation throughout the project.
Technical Setup	10	Efficient configuration and correct use of Python environment and libraries.
CRISP-DM Application	35	Clear, structured execution of all six CRISP-DM stages, with logical transitions and depth.
Research Paper	25	Quality, clarity, and academic rigour of the 4-page paper (structure, analysis, references).
Analytical Rigour	10	Depth of analysis, correctness of methods, and interpretation of results.
Reflection and Broader Context	5	Discussion of real-world applications, limitations, and learning outcomes.
Working Files and Submission Quality	5	Well-organised, complete, and reproducible submission.

Detailed Rubric (Summary)

Criterion	Excellent (80–100%)	Good (60–79%)	Satisfactory (40–59%)	Unsatisfactory (<40%)
CRISP-DM Application (35%)	All six stages fully documented and executed with high analytical depth and coherence.	Most stages completed with good explanation and logical flow.	Some stages incomplete or weakly linked.	Major omissions or superficial application.
Research Paper (25%)	Professional academic format; well-structured and supported by quality sources.	Clear structure and relevant references, minor formatting or depth issues.	Basic structure, limited academic support or weak discussion.	Poor structure, lacks references or originality.
Analytical Rigour (10%)	In-depth statistical and model evaluation; results clearly justified.	Sound analysis but limited interpretation.	Basic or partial analysis.	Minimal or incorrect analysis.
Technical Setup (10%)	Environment correctly configured; reproducible code.	Minor issues not affecting outcomes.	Several errors or missing dependencies.	Incomplete setup affecting analysis.
Active Engagement (10%)	Consistent progress and innovation evident.	Steady progress; some initiative.	Limited engagement or late work.	Minimal contribution or effort.
Reflection & Context (5%)	Insightful reflection, connects findings to real-world context.	Some reflection with moderate depth.	Minimal reflection or unclear relevance.	No reflection.
Working Files (5%)	All materials complete, organised, and clearly labelled.	Minor omissions or disorganisation.	Missing components or unclear structure.	Major missing files or confusion.