Portfolio Assessment

This assessment builds upon the work you completed in Assignment A.

This assessment is individual [No groups allowed]

With this assessment, you will complete exercises for some of the main topics covered during the second part of the module. These can be used to showcase your work as a Data Analyst to prospective employers.

Required Tasks

You are to complete a full example of applying the following approaches to a suitable data set, including details of analysis and reflections on findings.

Select two of the following four topics for this Portfolio Submission.

1. AutoML – A comparison of using AutoML for the same data set used in Assignment A, and compare the models and outcomes identified by AutoML to the model selected in Assignment A. **Two** AutoML libraries/packages should be evaluated.
2. Association Rules
3. Clustering
4. Text Mining

Important: You are responsible for selecting what datasets to use. The datasets should be **non-trivial** (eg. Iris data set and similar), and should **not** include any of the datasets and examples used throughout the module i.e. use a different dataset.

Data sets: See links to various websites on the module webpage. Also, check out this post with lots of links to various dataset repositories. https://oralytics.com/2019/04/18/data-sets-for-analytics/

Software: All analysis and code should be completed using Python. No other tool or language should be used e.g. Excel.

References & Citations: Include a References section for each Task. Each reference should be cited in your text. Use a referencing/citation method you are used to or select APA or Numeric Style. Three to Five references for each topic/notebook.
https://tudublin.libguides.com/APA_quick_guide
https://www.dit.ie/media/library/documents/Numeric.pdf

Notebooks: For each topic, create a new separate Jupyter Notebook. All code, analysis, insights, documentation, recommendations, etc should be entered into the Notebook. Submit these notebooks on BrightSpace.

Deliverables

You will be required to document your work. Create a separate notebook for each topic. Add code and documentation/comments using MarkDown and/or as a code comment.

Submission Details

The assignment is due on Week 12 - Sunday 4th May @23:00

You will need to submit your assignment on **BrightSpace VLE**. You cannot submit your assignment via email.

Important Information

Each Jupyter Notebook must contain your name, student number, class, course (**TU??**) and year information. **Failure to give this information will incur a 10% penalty.**

TU257 – Data Analytics - Assignment B (60%)

Due Before : (week 12)
Sunday 4th May @23:00

The assignment must be performed **individually.**

Each submission must be original work as **plagiarism** will result in a **zero** mark (0%).

There will be a 10% penalty deduction applied for each day the assignment is late.
There is no penalty for submitting early.

Plagiarism

Each submission must be original work as plagiarism will result in a **zero** mark (0%).
Each student is responsible for ensuring they are compliant with the TU Dublin General Assessment Regulations and the TU Dublin Plagiarism Policy
TU Dublin Plagiarism Policy: https://tudublin.libguides.com/c.php?g=674049&p=4794713
https://www.tudublinsu.ie/advice/exams/breachesofregulations/
Ensure you are compliant with TU Dublin policy on usage of Large Language Models (LLMs), (e.g. ChatGPT, etc) and other similar tools and software. The usage of such tools and software is not permitted for this assessment.

Assignment Feedback

I will endeavor to mark the assessments and provide feedback via Brightspace VLE, within two to three (working) weeks of the assignment submission date. This will consist of a mark and a short comment on the assignment. This comment will include areas where you did well and areas where you lost marks. All marks are provisional, are subject to change and are reviewed by the external examiner(s).

Marking Rubric

Each topic/notebook will have 30% allocated to them.

Achievement	Excellent	Satisfactory	Basic	Unsatisfactory
% of Marks Available	>75%	55-75%	40-55%	<40%

	Weighting				
Dataset, Problem Definition	10	Complex data sets selected for all tasks, well defined problem definition, justifications for selections	Mostly moderately to complex data sets, with corresponding problem definition, medium to well defined	Trivial, Simple and Moderate complex datasets chosen, Problem defined, missing details and impact	Trivial dataset chosen, poorly defined problem, etc
Data Insights & Data Preparation	25	Excellent focused insights from dataset select, good explanation, no trivial data analysis, selection of appropriate data preparation, explanations given,	Useful insights with explanations, impact of these on problem solution, selection of some appropriate data preparation, explanations given	Some insights given, limited details, limited data preparations, appropriateness of data prep, minimum explanations given	No or poorly selected data insights, limited or no data preparation
Application of Algorithms	25	Suitable algorithms selected, in-depth details on these and why, good details on experimentation, insights from experimentation, reflections, and discussion	Suitable algorithms selected, details of selection and why, some details of algorithm experimentation, some discussion of experimentation	Suitable algorithms selected, limited details of selections given, limited details of application of algorithms given, limited details of algorithm settings and tuning	Limited or no details of selection and application of algorithms for data and problem. No explanations
Analysis of Results	25	Excellent detailed analysis of results and excellent insights of these. Clearly demonstrates impact and outcomes	Good detailed analysis of results, good level of insights on these, what then mean, their impact and outcomes	Some discussion of results, at a basic level with little insights	Little, no or very limited analysis of results and outcomes from tasks
Learning for work	15	Excellent level of insights, brings together details through work, 4-8 citations comparing related work in each task, clear identification of improvements, reflection on learning outcomes from tasks	Good level of discussion of results, identify some areas for improvement, 3-5 citations used to compare results in each task, identification of some improvements with limited discussion	Some discussion and evaluation of work, some comparison with related research, limited number of citations used	Little or no discussion or work, no comparison with related research