YeaCC4 – Research & Analytics
CA1 – Business Analytics

This individual CA uses analytics to interpret data sets to assist in solving business problems. Using the business data provided, you will write Python code to generate analytic-based results related to **four** analytics problems. These Python results will be used to provide recommended solutions to specific business problems. You will also provide a description of each main analytical code used to solve the analytics problems (*shown as #annotated descriptions in the Jupyter Notebook*).

This CA involves:
(1). Developing Python code to query 4 datasets to answer specific business problems;
(2.) Describing the main analytical tools/measures used in solving the 4 business problems.
(3.) Writing a report that interprets the Python findings for each of the 4 business problems.

Deliverables:

1. 1 written report structured around the following headings*:*

    - Business Problem 1 - Interpretation of Findings
    - Business Problem 2 - Interpretation of Findings
    - Business Problem 3 - Interpretation of Findings
    - Business Problem 4 - Interpretation of Findings

2. Four Python files with the code and a description of the code/analytic measures
    (Descriptions should be written as annotations in the appropriate area of the python code).

Marking Structure

Description of analytical measures	20%
Python Coding	40%
Interpretation of findings in relation to business problem	40%

CA Mark: 50%

Due Date: Sun 20th November, 2022 (Midnight)

Submission: *Blackboard (Turnitin) -1 report document.*

Email*: All 4 Python (.ipynb files) email to: tim.mcnichols@iadt.ie*

Use any research materials and sources available to support the information in your report.
All materials used must be properly cited and referenced using APA style. For style details refer to: \\Sideshowbob1\lecturers\Tim McNichols\CC4\CA Referencing Information

Analytics Problem 1:

MegaStore Ltd. needs the following business information based on their sales.
Using the data, ***megastore_sales.csv***, answer the following:
(displaying only relevant columns and include any visualisations/plots where possible)

1. Clean up the data by dropping any rows with missing values.
2. List the Top 5 salespeople in Sales according to each segment.
3. Create a calculated column called 'Price' (=Sales / Quantity) for each product.
4. Show the sum of 'Profit' for each region (highest first) (use .agg method).
5. Using groupby, display the top 3 sales for each category and subcategory.
6. Using a pivot table, show the sum and mean of 'Sales' in terms of Profit and Category and Shipped.
7. Find the Technology Accessories sold in New York City worth more than $300 and sold by Nathan Mautz in 2014.
8. Find all the products with ID #TEC-PH-10003273 in the 'Home office' segment that were shipped 'Second class' in 2014.
9. Any sales visualisation (graph) that you think would help the MegaStore management to interpret their sales data better.

In your report, explain how the information from each task can help support MegaStore sales and management decisions.

Analytics Problem 2:

MegaStore Ltd needs to target visitors to their website who are more likely to convert to a sale.
Using the data, ***online_behaviour.csv***, which shows data of the users' interaction with our website, answer the following:

1. Which of the website interactions are highly correlated with a user's likelihood (propensity) to purchase?
2. Which of the website interactions are highly correlated with a user's likelihood (propensity) to repeat purchase?
3. Predict how likely new customers are to buy based on a new test model (also test the accuracy of the model).
4. Based on real-time use of your model, decide whether a prospect has a propensity to convert for **each** website column clicked.
5. Plot any graph (or add extra code) that would help MegaStore Ltd management interpret their click behaviour data.

Discuss how this model can be used to improve MegaStore's online offering and how they can improve conversion of visitors to a sale.

Analytics Problem 3:

GameStore Ltd. needs to determine what other 'games' to recommend to other customers.
Using the data, ***games_bought.csv***, which includes the list of orders and the items they bought, answer the following:

1. When a customer views a series, we need an affinity score (for all games) to recommend other items to the customer.

2. Create an item list (for all games) that you can recommend to the user in order of preference (include the name of the Items, Prices, Publishers & Score).
3. Use merged data from games_catalogue.csv
4. Include any extra code (or plot any graph) that you think would help GameStore management interpret their data.

Discuss the importance of recommendations and how the GameStore management can use this information to support their customers.

Analytics Problem 4:

StreamMedia Ltd. wants to understand the viewing patterns and associations for streaming users. Using the data, ***stream_viewing.csv***, which lists all the 'shows' viewed based on the Users' ID, answer the following:

1. What were the 10 most popular shows (based on 'support')?
2. Display all the combinations of the 'show' sets (viewed together).
3. Identify the 'shows' most likely to be viewed by the same user ID (based on 'confidence').
4. List the top 3 combinations of 'shows' viewed together in the same transaction as opposed to separate transactions (based on 'lift').
5. Display the top 5 combined 'shows' in one transaction (based on 'lift').
6. Include any extra code (*or plot any graph)* that you think would help StreamMedia management interpret their behaviour data.

Discuss how each item above can be used to improve StreamMedia's offering for their users.

Learning Outcomes

	Excellent	Very good	Good	Satisfactory	Weak	Inadequate
Interpret the findings of analytical tools when applied to a specific business problem	Excellent interpretation of the code and findings related to the three business problems with no major errors/omissions	Very good interpretation of the code and findings related to the three business problems with very few major errors/omissions	Good interpretation of the code and findings related to the three business problems with little evidence of errors/omissions	Adequate interpretation of the code and findings related to the three business problems with some major errors/omissions	Weak interpretation of the code and findings related to the three business problems with many major errors/omissions	Limited or no interpretation of the code and findings related to the three business problems with substantial errors/omissions
Describe analytical and non-analytical measures used in Business Performance Management	Excellent description of the analytics and non-analytics used for business performance management with no major errors/omissions	Very good description of the analytics and non-analytics used for business performance management with very few errors/omissions	Good description of the analytics and non-analytics used for business performance management with little evidence of errors/omissions	Adequate description of the analytics and non-analytics used for business performance management with some major errors/omissions	Weak description of the analytics and non-analytics used for business performance management with many major errors/omissions	Limited or no description of the analytics and non-analytics used for business performance management with substantial errors/omissions