MACQUARIE University
BUSINESS SCHOOL
SYDNEY·AUSTRALIA

BUSA8000 – Techniques in Business Analytics, Session 2, 2024

Assessment Task	Group Project
Due date	November 06 2024 – 23:45
Weight (%)	40%
Task description	Students will be given a dataset/business case and required to perform various analyses based on the techniques taught in classes. This is a group assessment.
Submission Method	Online iLearn and Turnitin
Feedback mechanism(s)	Via iLearn
Feedback available (anticipated date)	14 days after submission
Links to Unit Learning Outcomes	ULO4, ULO5, ULO6

INTEGRITY MATTERS

Integrity matters – at home, in your workplace and here at University. As a highly valued member of the Macquarie University Community, you carry great responsibility to uphold the good name of our institution. Our reputation is your reputation and will stay with you for life when your degree opens amazing opportunities for you.

If you are ever unsure whether your actions fall within the guidelines of Academic Integrity, please don't hesitate to reach out. Contact the Academic Literacies Unit or your Tutor/Unit Convenor.

DESCRIPTION

Job Description: Data Analyst for LuminaTech Lighting (Australia)

As a Data Analyst for LuminaTech Lighting, your role includes insights and performance recommendations often related to sales performance, customer demographics, and product inventory. You have been provided with a comprehensive dataset that includes variables such as accounting dates, fiscal and calendar year data, customer district codes, product codes, pricing adjustments, and currency types, order and invoice details.

Your responsibilities include cleaning and preparing data for analysis, creating meaningful visualisations, and conducting statistical analyses to identify valuable insights for the management team. You will be expected to provide insights that support decision-

making in sales strategy, customer retention, and business performance improvement. Furthermore, you will link these findings to actionable recommendations for enhancing operational efficiency and driving business growth.

Skills in focus for this project

- *Familiarity with Python packages and relevance to data manipulation and visualisation.*
- *Data importing.*
- *Understanding how to convert data types.*
- *Handling missing values, techniques for detecting, removing or imputing missing values.*
- *Ability to identify and correct errors in data such as outliers or inconsistent entries.*
- *Graphical representation skills for various types of visualisations.*
- *Understanding the basic statistics for data analysis.*
- *Critical thinking for analysing data, identifying patterns or trends and drawing meaningful conclusions.*
- *Knowledge of regression analysis.*
- *Interpretation of results of data analysis and visualisations.*
- *Familiarity with Jupyter development environment for Python programming.*

INSTRUCTIONS

Section 1: clean the dataset

You are expected to cleanse your data and, in your report, describe clearly the data cleaning process you undertook options considered and justify any transformations.

Section 2: Exploratory Insights

Using a method of your choice (such as visualisation, t-test, or others), provide five exploratory insights from the datasets provided. Examples of insights might include, but are not limited to, the time gap between order date and invoice date, customer retention rates, sales trends, profitability across various customer district codes, and the impact of currency on profit margins. It is up to your group to identify and develop these exploratory insights.

Each of the five insights should be accompanied by a relevant visualisation, as well as an explanation of the method used and the insights gained (e.g., what the findings reveal). Additionally, for each insight, please explain why it would be valuable for the management team.

Section 3: Test Sub Sample Differences
Develop two questions that require you to apply a two-sample test for each question and run these tests and derive insights from the data.

For each question, provide a clear explanation of what is being tested, how it is being tested, the results obtained, and why these results would be valuable for the management team. (Not you may use the t-test even if using a larger sample than necessary)

Section 4: Inference
Develop two questions, with each question requiring you to apply multiple regression analysis to examine the correlation between independent variables and a dependent variable of your choice.

For each question, explain the method used, present the results, robustness evidence and discuss why the insights are valuable for the management team.

Section 5: Prediction Model

Develop a prediction model that predicts the sales price in 2014.

Section 6: Higher Likelihood of Losing Customers

Identify features that result in a higher likelihood of losing customers (also known as customer churn analysis).

Submission files:

Below files should be submitted for this assignment

1. **Group Report** (No Python code) that includes the description of your analysis and results including the visualisations (in each step, why you used a particular method and what are the results) all the sections in the report should have **headings referring** to the section number and questions titles.
2. **Jupyter Lab file for the Python code** (the codes should have #comments and the file should have headings (in markup) for each step/question but an explanation of the method and the findings are not needed in the Jupyter file). The Jupyter file does not have a page limit or formatting limit.The code needs to run using the provided dataset. If the code fails the marker may award zero for the group report.
3. **Individual report** Written reflection on undertaking the project and your reflections. This may include when did the group work effectively and not so effectively and why, any change in understanding, perception or beliefs as a result, and possible future impact. Reflection on the impact teamwork/ collaboration had or did not have on achieving the task objective.

The Group Report page limit is 25 pages excluding references and title page.

The Individual Report page limit is 2 pages excluding references and title page.

The two reports to be 11-point Arial font with 1.5 line spacing. Each page of the report should be numbered and have at minimum 2.5 cm margins from the left and right edges and the top and bottom of the page.

The report should have the following structure:

1. A title page with the assignment title, your name and student number, as well as the word count of the text.
2. The body of the report is structured with paragraphs and with appropriate headings and citations, including page numbers.
3. Complete reference list of material cited in the text – References are not compulsory for this report; however, if a reference is cited in the report, the bibliography should be included. References should be cited if you use any materials from articles, books and other sources. Please use your own words even though references are cited.

Referencing style: Please use the American Psychological Association (APA) 7th edition reference style. This means that when referring to an article or book etc, you indicate the author's name followed by the year of publication (Jones, 1989). For multiple authors include all names (Jones, Smith, & Wilson, 1990). If you use a direct quote put the quoted words in "quotation marks" and include the page number with the reference (Jones, 1989, p. 76). A reference list should be presented in alphabetical order at the end of the paper. Further information on APA style is available here.

TIPS & FAQs (OPTIONAL)

TIPS

Start with a Plan: Before diving into coding, outline your analysis plan. Understand the objectives, the data you're working with, and the steps you might need to take to reach your conclusions.

Understand Your Data: Spend time exploring and understanding your data. Use summary statistics and initial visualisations to get a feel for the data's structure, variables, and potential quirks.

Keep Your Code Organised: Comment your code extensively and use consistent naming conventions for variables and functions. This makes your code easier to understand and debug, both for you and others.

Break Down the Task: Divide your assignment into smaller, manageable tasks (e.g., data import, cleansing, transformation, analysis, visualization). Tackle each task one at a time.

FAQs

Q: How do I choose the right type of visualisation?

A: The choice of visualisation depends on the nature of your data (categorical vs. numerical) and the story you want to tell.

Q: How many visualisations should I include in my report?

A: Include visualisations that add value to your analysis and help convey your findings clearly. Quality over quantity; each visualisation should have a clear purpose.

Q: Can I use other packages beyond what was learned?

A: Absolutely! Python has a vast ecosystem of packages. If you find one that suits your needs better, feel free to use it. Just ensure you understand it well and explain clearly your choice in your assignment.

USE OF RESOURCES AND TECHNOLOGIES INCLUDING GENERATIVE ARTIFICIAL INTELLIGENCE

For this assessment, students are permitted to use generative artificial intelligence tools (GAITs e.g., ChatGPT) to:

- clarify concepts, theories, ideas, etc., discussed in class
- generate preliminary ideas for writing and coding
- edit a working draft of the assessment
- read and summarise research and supporting evidence for the assessment

Students are **not** permitted to use GAITs to

- Generate definitions or writing used in their final submission.
- produce counter-arguments or refine thinking on their final submission
- Generate complete Python code in their final submission.

Any of these actions will constitute and be treated as a breach of academic integrity.

Don't's

1. DON'T ask a GAIT to complete an assessment for you. This is outsourcing your assessment and is a breach of academic integrity.
2. DON'T blindly trust GAIT information. GAIT outputs can be completely inaccurate and will often contain fake references.
3. DON'T rely on GAITs to replace your own thinking and creativity.

Acknowledgement Statement by students:

Please select one acknowledgment from the following

- ☐ I acknowledge that I have **not used** GAITs (e. g., ChatGPT) in drafting and proofreading of this assignment.

☐ I acknowledge that I have **only used** GAITs (e.g., ChatGPT) in drafting and proofreading this assignment, which is permitted in the assignment instructions.

LATE SUBMISSION

A maximum penalty of five percentage points of the total possible marks will be applied per day to late submissions, for up to a maximum of seven calendar days. Tasks that have not been submitted within the maximum number of additional late days will receive a mark of zero, unless otherwise specified in the late penalties section of the Unit Guide. Late submission for a task will only be permitted when specified in the unit guide. This provision does not apply to online exams or other assessment with a time-limit of less than 24 hours.

Where an application for Special Consideration is approved and the outcome is an extension to the due date of a task, submissions that are received after the new due date will be subject to late penalties that are calculated from the new due date. This only applies where the outcome is an extension to the due date – see the Special Consideration Policy for a schedule of all possible outcomes.