

Business Analytics – Assessment Task 3 – Individual

DUE DATE: Monday, 29 May 2023, by 5:00pm (Melbourne time)

PERCENTAGE OF FINAL GRADE: 40%

Submission: You will submit to unit site:

- one Excel file, with your analysis, and
 - one Word file, with your written report
-

Description

The assignment requires that you analyse a data set, interpret, and draw conclusions from your analysis, and then convey your conclusions in a written report. The assignment must be completed individually and must be submitted electronically in Cloud by the due date. When submitting electronically, **you must check** that you have submitted the work correctly by following the instructions provided in Cloud. Hard copies or assignments submitted via email will NOT be accepted.

The assignment uses the file *DreamCruise dataset A3.xlsx* which can be downloaded from Cloud. The assignment focuses on materials presented up to and including Week 11. The Excel file which has been provided has different worksheets explaining and containing the DreamCruise dataset. For confidentiality reasons actual data has not been used in the assessment task. Following is an introduction to this scenario and detailed guidelines.

Context/Scenario: DreamCruise Passenger Analysis

DreamCruise¹ is an established holiday company that specializes in providing leisure cruises from various ports in Australia to islands in the South Pacific. Before the Covid-19 pandemic, DreamCruise was a thriving business with a loyal customer base. However, like many other companies in the tourism industry, DreamCruise has been severely impacted by the pandemic and has had to suspend its operations.

¹ DreamCruise is a fictitious company and not meant to bear any resemblance to any existing company. All data and any individuals mentioned are fictitious and have been produced by the Unit team.

As the world looks forward to a post-Covid era, DreamCruise is now exploring ways to relaunch itself and rebuild its business.

Assume that you are a business analyst recruited by DreamCruise. You have received an email from **Maria Rodriguez**, DreamCruise's Director of Analytics. Your response will be used as part of a report to the DreamCruise Board of Directors. Maria's email together with guidelines (shown in blue) are presented below:

Email from Maria Rodriguez

To: You
From: Maria Rodriguez, Director of Analytics, DreamCruise
Subject: Analysis of the DreamCruise's booking passengers dataset

Hi ...,

We are very happy with the strong interest and business generated from the DreamCruise's booking passengers dataset. The Board wants a detailed understanding of some of the key aspects of the bookings. I have attached an Excel file with key data and included some guidelines (shown in blue) to direct your work.

Please provide answers to the following questions. Return the Excel file to me. As I have training in business analytics, I am comfortable with technical language. The Board wants a report from you which explains the outcome of your analysis. As they do not have the benefit of training in business analytics your report must present the results of your analysis in plain, straight-forward language. I have provided a template for you to use.

1. Hypothesis Testing (*consider $\alpha = 5\%$*)

The Board is concerned about the average passenger satisfaction. It has been suggested at a recent Board meeting that the average Passenger Satisfaction for every Booking Type, is now less than 70. Does the data confirm this hypothesis?

To answer this question, you will need to conduct an appropriate hypothesis test for Passenger Satisfaction for each Booking Type.

2. Multiple Linear Regression Modelling (*consider $\alpha = 5\%$*)

Passenger Satisfaction is an important measure for DreamCruise, as it represents a major element of the company's marketing strategy. Build a multiple regression model to predict **Passenger Satisfaction**. Your model should provide insights into which factors have a significant influence on passenger satisfaction, as well as the ability to predict Passenger Satisfaction for various scenarios.

For this analysis, you will need to build a multiple regression model using Passenger Satisfaction as the dependent variable. All other variables in the DreamCruise dataset should be included in the model, except ID, Age Band, Satisfaction Band, and Spending Band (i.e., exclude ID, Age Band, Satisfaction Band, and Spending Band from your regression model).

Follow the model building process introduced in the lecture and seminars. Carefully consider the following:

- (a) Transform categorical variables into suitable **dummy variables** (i.e., Gender, Cabin, and Booking Type).

Copy the DreamCruise Dataset (excluding ID, Age Band, Satisfaction Band, and Spending Band from your regression model) to the “Correlation” sheet in the Excel file that has been provided (no earlier than Column W - be careful not to overwrite the Conclusion, Correlation Table and Scatter Diagram frames).

- i. When transforming **Gender** into dummy variables, consider Female as the *baseline category*; meaning the created dummy variables for Gender should only include Male (Yes and No) and Other/Prefer not to disclose (Yes and No)
- ii. When transforming **Cabin** into dummy variables, consider Suite as the *baseline category*; meaning the created dummy variables for Cabin should only include Luxury (Yes and No), Porthole (Yes and No), and Internal (Yes and No).
- iii. When transforming **Booking Type** into dummy variables, consider Solo as the *baseline category*; meaning the created dummy variables for Booking Type should only include Double (Yes and No), Group (Yes and No), and Family (Yes and No).

Complete the Dummy Variables Summary table which is in the Conclusion section of the Correlation worksheet. The table summarises the results of your transformation of categorical variables into dummy variables.

- (b) Using the DreamCruise dataset (which now includes transformed dummy variables) as your reference, complete the following steps:

- i. **Correlation** – in the section marked “Correlation Table” (below the Conclusion section on the “Correlation” worksheet) generate a correlation table. Use the “Correlation” option in Excel’s Data Analysis ToolPak.
- ii. On the correlation table, identify and clearly indicate the Independent Variables which are (virtually) uncorrelated with the Dependent Variable (i.e., all IVs which have a correlation coefficient with the DV of between -0.05 and 0.05). These IVs are to be removed from the model prior to running the first iteration of the regression model.
- iii. Complete the Uncorrelated Independent Variables summary table which is in the Conclusion section of the Correlation worksheet. This table summarises which

Independent Variables are to be eliminated from the regression model due to being (virtually) uncorrelated with Passenger Satisfaction (DV).

- iv. **Multi-collinearity** - review the correlation table for instances of multi-collinearity between Independent Variables (IV). In cases of multicollinearity, identify and clearly indicate the IVs with the weakest correlation with the Dependent Variable. These IVs are to be removed from the model prior to running the first iteration of the regression model.
 - v. Complete the Multi-Collinearity summary table which is in the Conclusion section of the Correlation worksheet. This table summarises which Independent Variables are to be eliminated from the regression model due to multi-collinearity.
 - vi. **Scatter diagrams** - in the section marked "Scatter Diagrams" (below the Correlation Table section on the "Correlation" worksheet) generate three scatter diagrams, for:
 - Passenger Satisfaction (Dependent Variable, DV) and the numerical (not categorical) Independent Variable (IV) which has the highest correlation with the DV. Include a calculation of the correlation coefficient. Format the diagram, and include a linear trendline,
 - Passenger Satisfaction (DV) and the numerical (not categorical) Independent Variable (IV) which has the lowest (i.e., most negative) correlation with the DV. Include a calculation of the correlation coefficient. Format the diagram, and include a linear trendline, and
 - Passenger Satisfaction (DV) the Independent Variable (IV) that is closest to being uncorrelated with the DV (i.e., correlation coefficient closest to zero). Include a calculation of the correlation coefficient. Format the diagram and include a linear trendline.
- (c) Using the DreamCruise dataset as your reference complete the following steps, on the "**Regression Model**" spreadsheet in the Excel file that has been provided (the data set includes the dummy variables you have created and excludes the Independent Variables which have been eliminated due to multi-collinearity or being uncorrelated with the Dependent Variable):
- i. Using the "Regression" option in Excel's Data Analysis ToolPak build a multiple regression model.
 - Assess the model for overall significance (F test with alpha set at 0.05, i.e., Confidence Level = 95%).

- ii. If your first iteration of the overall model is found to be significant, in a stepwise fashion, sequentially (one at a time) remove the Independent Variables that are least likely to be contributing to any significant change in the Dependent Variable.
 - You will need to conduct t-tests with alpha set at 0.05 to determine the significance of the various IVs you exclude and include in your model.
- (d) Once you have created a regression model where all the remaining Independent Variables are contributing significantly to a change in Passenger Satisfaction, copy the Summary Output of your final multiple regression model and paste it into the Output section of the “Regression Model” spreadsheet in the Excel file that has been provided, i. In the **Conclusion** section of the “Model” spreadsheet,
- Write the (final) multiple regression equation. Use the format: $\hat{Y} = b_0 + b_1X_1 + b_2X_2...$
 - Explain (interpret) the (final) multiple regression equation/model.
- (e) Using the final multiple regression equation (from step (d)(i)),
- i. In the **Predictions** section of the “Regression Model” spreadsheet in the Excel file that has been provided, for the scenario outlined below:
 - Calculate a Point Estimate for Passenger Satisfaction (DV),
 - Calculate a Prediction Interval for Passenger Satisfaction (DV),
 - Calculate a Confidence Interval for Passenger Satisfaction (DV), ii. In the **Conclusion** section of the “Regression Model” spreadsheet in the Excel file that has been provided, for the scenario outlined below:
 - Interpret the Point Estimate
 - Interpret the Prediction Interval
 - Interpret the Confidence Interval

Independent Variables	Scenario
Age	52
Gender	Female
Children	3
Passengers	5

Cabin	Luxury
Booking Type:	Family
Cabin Cost	\$1950
Side Trips	\$270
Insurance	\$880
Food and Drink	\$450
Entertainment	\$20
Merchandise	\$270
Total Spending	\$3840

I look forward to receiving details of your analysis and your report. Sincerely,

Maria

Data description

The provided data file includes multiple sheets, labelled “Data Description”, “DreamCruise Data Set” and a worksheet for your dashboard. The “Data Description” sheet describes all the variables used in the “DreamCruise Data Set” and is copied below for your convenience.

Variable	Description
Passenger ID	ID of the passenger making the booking (“Booking Passenger”)
Age	Age of the Booking Passenger
Age Band	Booking passengers have been allocated to one of five age bands: Young (<30 years), Core (30 – 45 years), Prime (46 – 59 years), Mature (60 – 69 years), Senior (>70 years)
Gender	Gender of Booking Passenger: Male, Female, or Other/Prefer Not to Disclose
Children	Number of children included in the booking (included in Passengers)
Passengers	Number of passengers included in the booking (including children)
Cabin	Suite, Luxury, Porthole, Internal
Booking Type:	Solo, Double (couple or two friends travelling together), Group, Family
Cabin Cost	Cost per passenger on Cabin (included in Total Spending)

Side Trips	Spending per passenger on side trips
Insurance	Spending per passenger on travel insurance
Food and Drink	Spending per passenger on food and drinks
Entertainment	Spending per passenger on entertainment
Merchandise	Spending per passenger on merchandise
Total Spending	Total Spending per passenger (sum of all cost/spending categories)
Spending Band	Budget (<\$2,500pp), Medium (\$2,501 - \$3,499), High (\$3,500 - \$4,499pp), Premium (>=\$4,500pp).
Satisfaction	Passenger rating of satisfaction with their experience: 0 – 100 (lowest to highest)
Satisfaction Band	Unhappy (<50), Unimpressed (50 – 59), Acceptable (60 – 69), Happy (70 – 79), Delighted (>80)

Assignment instructions

The assignment consists of two parts.

Part 1: Data Analysis

Your data analysis must be performed on the Assignment 3 Excel file. The file includes tabs for:

- Data Description
- DreamCruise Survey Data Set
- Analysis for Hypothesis Testing
- Analysis for Correlation
- Analysis for Regression Model Building

When conducting the analysis, you need to apply techniques learnt in the lectures and seminars. The analysis section you submit should be limited to the Hypothesis Testing, Correlation, and Regression worksheets of the Excel file. These are the only worksheets which will be marked. Your analysis should be clearly labelled and grouped around each question. Poorly presented, unorganised analysis or excessive output will be penalised.

In the **Conclusion** section of each worksheet there is space allocated for you to write a succinct response to the questions posed in Maria's email (above). When drafting your Conclusion, make sure

that you directly answer the questions asked. Cite (state) the important features of the analysis in your Output section. Responses in the Conclusion section will be marked.

Use the **Output** section for your analysis to complete the analysis as directed in Maria's email and supports your response to his questions (which you will write in the Conclusion section). Analysis in the Output section will be marked. Make sure your analysis and process complete, clear, and easy to follow. You may need to add (or widen/narrow) rows or columns to present your analysis clearly and completely. Poorly presented, disorganised analysis or excessive output will be penalised. It is useful to produce both numerical and graphical analysis. Sometimes something is revealed in one that is not obvious in the other.

Use the **Workings** section for calculations and workings that support your analysis. The Workings section will not be marked.

Part 2: Report

Having analysed the data, including answers (in technical terms) to the Data Analysis questions from Part 1 you are required to provide a formal report which can be placed before the DreamCruise Board of Directors. Assume that none of the directors on the Board have any training in statistics; they will only be familiar with broad generally understood terms (e.g., average, correlation, proportion, and probability). They will need you to explain more technical terms, such as quartile, mode, standard deviation, coefficient of variation, correlation coefficient, and confidence interval, etc.

In section 1 of the report a short interpretation of your findings to each question. In section 2 of the report, **Make TWO (2) recommendations that the DreamCruise Board could consider to maximise Passenger Satisfaction.** Your recommendations can be based on analysis in this assignment, analysis from previous assignments and any other analysis that you consider is relevant and adds impact to your recommendations.

Thoughts to consider in framing your recommendations include:

- Specific actions DreamCruise could take to maximise Passenger Satisfaction based on the outcomes of your regression model.
- Specific actions DreamCruise could take to maximise Passenger Satisfaction based on the outcomes of your analysis from Assignment 1 and Assignment 2.

- Specific actions DreamCruise could take to maximise Passenger Satisfaction based on the outcomes of any additional analysis you perform.
- Recommending targeting a group that DreamCruise could pursue that maximises Passenger Satisfaction.
- The impact of maximising Passenger Satisfaction on the other important outcome of Total Spending.
- Considering the impact on Passenger Satisfaction of the variables not specifically included in your regression model.
- Recommending targeting a group that DreamCruise could pursue that provides an optimal balance in terms of all DreamCruise's corporate KPIs, including Total Spending and Passenger Satisfaction.

Make sure that all your recommendations are directly informed by your data analysis. Do not include any commentary that is not supported by your data analysis.

Highest marks will be awarded to students who draft distinct (i.e., different) recommendations, and whose recommendations take into account a broad range of (data-supported) considerations.

When exploring data, we often produce more results than we eventually use in the final report, but by investigating the data from different angles, we can develop a much deeper understanding of the data. This will be valuable when drafting your written report.

It is useful to produce both numerical and graphical statistical summaries. Sometimes something is revealed in one that is not obvious in the other.

You are allowed approximately 1,000 words (950 to 1,050 words) for your report. Remember you should use font size 11 and leave margins of 2.54 cm.

A **template** is provided for your convenience. Carefully consider the following points:

- Your report is to be written as a stand-alone document. Assume that your Excel file is for Maria's use only and that Maria will only pass your written report directly to the Board.
- Keep the English simple and the explanations clear. Avoid the use of technical statistical jargon. Your task is to convert your analysis into plain, simple, easy to understand language.

- Follow the format of the template when writing your report. Delete the report template instructions (in purple) when drafting your report.
- **Do not include any charts, graphs, or tables into your Report.**
- Include a succinct introduction at the start of your report, and a conclusion that clearly summarises performance against KPIs.
- Marks will be deducted for the inclusion of irrelevant material, poor presentation, poor organisation, poor formatting, and reports that exceed the word limit.

When you have completed drafting your report, it is a useful exercise to leave it for a day, and then return to it and re-read it as if you knew nothing about the analysis. Does it flow easily? Does it make sense? Can someone without prior knowledge follow your written conclusions? Often when rereading, you become aware that you can edit the report to make it more direct and clearer.

Learning Outcomes

This task allows you to demonstrate your achievement towards the Unit Learning Outcomes (ULOs) which have been aligned to the Graduate Learning Outcomes (GLOs). GLOs describe the knowledge and capabilities graduates acquire and can demonstrate on completion of their course. This assessment task is an important tool in determining your achievement of the ULOs. If you do not demonstrate achievement of the ULOs you will not be successful in this unit. You are advised to familiarise yourself with these ULOs and GLOs as they will inform you on what you are expected to demonstrate for successful completion of this unit.

The learning outcomes that are aligned to this assessment task are:

Unit Learning Outcomes (ULO)	Graduate Learning Outcomes (GLO)
ULO1: Apply quantitative reasoning skills to analyse business problems.	GLO1: Discipline-specific knowledge and capabilities
ULO2: Create data-driven/fact-based solutions to complex business scenarios.	GLO5: Problem solving
ULO3: Analyse business performance by implementing contemporary data analysis tools.	GLO3: Digital literacy
ULO4: Interpret findings and effectively communicate solutions to business problems	GLO2: Communication

Submission

You must submit your assignment in the Assignment Dropbox in the unit Cloud site on or before the due date.

Your submission will comprise of two files:

1. A Microsoft Excel workbook file containing your Analysis (Part 1), on the relevant tabs, and
2. A Microsoft Word document containing your report (Part 2) to Maria.

When uploading your assignment, your submission files should be named:

Word file: MIS171_T2_YOURStudentID.doc (or .docx), and

Excel file: MIS171_T2_YOURStudentID.xls (or .xlsx).

Submitting a hard copy of this assignment is not required. You must keep a backup copy of every assignment you submit until the marked assignment has been returned to you. In the unlikely event that one of your assignments is misplaced you will need to submit your backup copy.

Any work you submit may be checked by electronic or other means for the purposes of detecting collusion and/or plagiarism and for authenticating work.

When you submit an assignment through your Cloud unit site, you will receive an email to your email address confirming that it has been submitted. You should check that you can see your assignment in the Submissions view of the Assignment Dropbox folder after upload and check for, and keep, the email receipt for the submission.

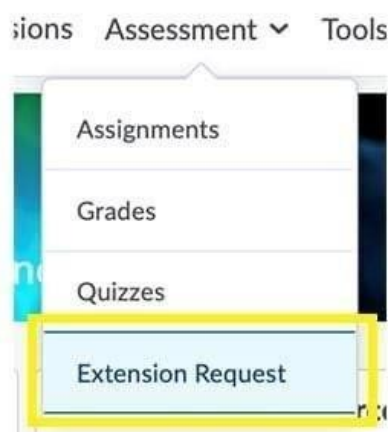
Marking and feedback

The marking rubric indicates the assessment criteria for this task. It is available in the Cloud unit site in the Assessment folder, under Assessment Resources. Criteria act as a boundary around the task and help specify what assessors are looking for in your submission. The criteria are drawn from the ULOs and align with the GLOs. You should familiarise yourself with the assessment criteria before completing and submitting this task.

Students who submit their work by the due date will receive their marks and feedback on Cloud 15 working days after the submission date.

Extensions

Extensions can only be granted for exceptional and/or unavoidable circumstances outside of your control. Requests for extensions must be made by 12 noon on the submission date using the online Extension Request form under the Assessment tab on the unit Cloud site. All requests for extensions should be supported by appropriate evidence (e.g., a medical certificate in the case of ill health).



Applications for extensions after 12 noon on the submission date require University level [special consideration](#) and these applications must be submitted via StudentConnect in your Sync site.

Late submission penalties

If you submit an assessment task after the due date without an approved extension or special consideration, 5% will be deducted from the available marks for each day after the due date up to seven days*. Work submitted more than seven days after the due date will not be marked and will receive 0% for the task. The Unit Chair may refuse to accept a late submission where it is unreasonable or impracticable to assess the task after the due date. *'Day' means calendar day for electronic submissions and working day for paper submissions.

An example of how the calculation of the late penalty based on an assignment being due on a Monday at 8:00pm is as follows:

- 1 day late: submitted after Monday 11:59pm and before Tuesday 11:59pm – 5% penalty.
- 2 days late: submitted after Tuesday 11:59pm and before Wednesday 11:59pm – 10% penalty.
- 3 days late: submitted after Wednesday 11:59pm and before Thursday 11:59pm – 15% penalty.
- 4 days late: submitted after Thursday 11:59pm and before Friday 11:59pm – 20% penalty.
- 5 days late: submitted after Friday 11:59pm and before Saturday 11:59pm – 25% penalty.
- 6 days late: submitted after Saturday 11:59pm and before Sunday 11:59pm – 30% penalty.

- 7 days late: submitted after Sunday 11:59pm and before Monday 11:59pm – 35% penalty.

The Dropbox closes the Monday after 11:59pm AEST/AEDT time.