FACULTY OF BUSINESS AND LAW

Department of Information Systems and Business Analytics

MIS770 – Foundation Skills in Data Analysis

Assignment 2

Particulars

Due: Week 9, **11:59pm**, Thursday 23rd January, 2020
Marks: 30%
Words: 2000
Submission: Online to the MIS770 assignment TWO drop box in the Cloud Unit site
Email submissions will **not** be accepted
Note: This assignment is to be completed individually

Assurance of Learning

This assignment assesses following Graduate Learning Outcomes and related Unit Learning Outcomes:

Graduate Learning Outcome (GLO)	Unit Learning Outcome (ULO)
GLO4: Critical thinking: evaluating information using critical and analytical thinking and judgment	**ULO2**: Manipulate and summarise data that accurately represents real world problems
	ULO3: Interpret and appraise statistical output to assist in real-world decision making

Overview

The purpose of this assignment is to investigate a dataset utilising the knowledge learned in Modules One and Two. This will enable conclusions to be drawn that ultimately assist in decision making.

The assignment requires you to analyse a given dataset, interpret the results, and then draw conclusions such that you are able to reply to specific questions being asked of you in the form of a report. (These questions are asked in the following memorandum).

The aims of the assignment is to:

- provide you with some examples of the application of data analysis within an organisation
- test your understanding of the material in the relevant topics
- test your ability to analyse and interpret your results
- test your ability to effectively communicate the results of your analysis to others

Before tackling the assignment, make sure you have prepared yourself well. As a ***minimum***, please read the relevant sections of the prescribed text and listen/watch the pre-recorded material for Modules 1 and 2.

Scenario

Note: All data, reports, people and scenarios in this assignments are either fictitious or have been modified from their original state.

You play the role of Emma Thomson, a data analyst in the Research and Analysis Group at the Real Estate Institute of Victoria (REIV). You are often required to report outcomes of your analysis to senior managers at the Institute who have little or no knowledge of data analysis. Of specific interest to the REIV is the increasing house prices in Melbourne and potentially what this will mean for buyers yet to enter the market.

A recent analysis, which received publicity in the media, showed that there were more Melbourne suburbs with a median house price of over $1 million, with this trend only set to continue as evidenced from recent auction and private sales figures (refer to the documents million dollar suburbs.pdf and buyers smash records.pdf – sourced from the Melbourne's Herald Sun). Just recently you have been working on a pilot survey from the City of Kingfisherbay concerning house prices, rental costs and returns, as well as other related data.

In light of the above, the REIV is very interested to see where the City of Kingfisherbay, and more specifically its suburbs are at when it comes to house prices and affordability. Paul Anderson, Director of the Housing Affordability Division at the Real Estate Institute has written to you regarding the Kingfisherbay housing survey. His memo to you is reproduced below:

Memorandum

Memorandum	
To:	Emma Thomson
From:	Paul Anderson
Subject:	Analysis of Kingfisherbay's housing and rental data

Dear Emma,

You will have seen the recent media articles about how Melbourne's house prices are increasing and how more Melbourne suburbs now have a median house price of over $1 million. I am concerned about the direct effect this will have on housing affordability in Melbourne, especially for buyers yet to enter the market. I know that you recently undertook a survey of house sales in the City of Kingfisherbay.

Could you please provide me with answers to the following questions:

1. An overall summary of the house prices in Kingfisherbay? In addition, the media articles focus on median house prices and not the mean. I have never been able to understand why this is so; surely the "average" is the "average". Can you clarify that for me?

2. Does there appear to be any difference in the house prices between the suburbs in Kingfisherbay?

3. Can you supply me with a brief summary on the condition of houses in Kingfisherbay? In addition, can you further analyse to see whether there appears to be any differences among the suburbs in terms of condition?

4. Is house price related to any of the variables, namely: Rooms, Land Size, House Area and Weekly rent that you have collected in the sample? A short description of the relationship would be fine.
   Is there any one which stands out as having a stronger influence on house prices?
   Which of these variables has the least influence on house prices?

I realise that the survey relates to a random sample of 120 houses sold in the (fictitious) Melbourne City of Kingfisherbay, and that this information can be used to draw inferences about all houses in the City of Kingfisherbay. With that in mind, I hope you are able to provide me with answers to the following questions:

5. The newspaper article "Buyers smash records" mentions how new land is in high demand and the ridiculous situation of buyers having to camp out for days to purchase land. I'm interested in whether there is the opportunity to make more land available in areas such as Kingfisherbay by subdividing. Generally if a land size is at least $1000m^2$ or more then it can be subdivided.
   Can you work out what percentage of all houses in Kingfisherbay are suitable for subdivision?
   Also, what is the estimated average land size across all of Kingfisherbay?

6. As well as housing affordability, the REIV is also concerned about the lack of rental properties available right across Melbourne including areas such as Kingfisherbay. At a recent meeting it was claimed by a senior manager that the proportion of rental properties that were vacant had dropped below 25%. Furthermore, it was recently reported that the average selling price for bay side houses, that is, for all cities and suburbs located around the Bay in Melbourne, had now exceeded $825,000.
   Can you check whether these two claims are correct for all of Kingfisherbay?

7. For next year's study, we would like to be able estimate the true proportion of houses of land size $1000m^2$ or more to within 2% and the average house price to within $20,000 with a high level of confidence
   How many houses would be needed to be included in next year's survey to achieve these two requirements?
   Finally, assuming we take a big enough sample from Kingfisherbay next year, would it be possible to generalise the results to all of Melbourne? This will save us a lot of work duplicating the study in other suburbs.

I look forward to your responses.
Sincerely,

Paul Anderson

Requirements:

- Your report should be no longer than 2000 words and there is no need to include, Charts and Tables, or Appendices in the report
- Your Charts/Graphics and Tables are only to be placed in the Data Analysis file i.e. the Excel spreadsheet
- The report is to be written as a stand-alone document (assume Paul will only read your report). Thus, you should not have any references in the report to your data analysis output. Eg. "According to Table 1 in the analysis…"
- Your report must have an informative title
- Your report must contain an executive summary that explains in plain language what the report is for and summarises the main findings. The executive summary should be no more than a page
- The body of your report must be set out in the same order as in the originating memorandum from Paul Anderson, with each section (question) clearly marked
- Use plain language and your explanations succinct. Avoid the use of technical or statistical jargon as Paul Anderson will not necessarily understand statistical terms. As a guide to the meaning of "Plain Language", imagine you are explaining your findings to a person without any statistical training (e.g. someone who has not studied this unit). What type of language would you use in this case?
- Marks will be lost if you use unexplained technical terms, irrelevant material, or have poor presentation/ organization
- All Microsoft Excel data analysis output associated with each question in the Memorandum is to be placed in the corresponding tab in the ***yourstudentid.xlsx*** file

Data Analysis Instructions/Guidelines

In order to prepare a reply to Paul's memorandum, you will need to examine and analyse the dataset **KingfisherbayData.xlsx** thoroughly.

Paul has asked a number of questions and your Data Analysis output (i.e. your charts/tables/graphs) should be structured such that each question is answered on the separate tab/worksheet provided in your Excel document. There are also extra tabs in **KingfisherbayData.xlsx** called CI, SampleSize and HT and you should use the various templates contained in these tabs in your "CI_Mean/CI_Proportion", "SampleSize" and "HT_Mean/HT_Proportion" answers.

In order to effectively answer the questions, your Data Analysis output needs to be appropriate. Accordingly, you'll need to establish which of the following techniques are applicable for each question:

- Numerical Summary Measures (Inc. Outlier detection)
- Suitable tables and charts or graphics (Module One) that will illustrate more clearly, other important features of a variable
  Numerical variables can be converted to categorical variables
- Comparative Summary Measures, Scatter diagrams, Correlation analysis and Cross Tabulations (sometimes called Contingency Tables), used to establish the relationships (dependencies) between two variables
- Confidence Intervals: You can assume that a 95% confidence level is appropriate. We use Confidence Intervals when we have no idea about the population parameter we are investigating. Additionally, we would use Confidence Intervals if we are asked to provide an estimate of a population parameter.
- We Use Hypothesis Tests when we are testing a Claim, a Theory or a Standard. Use 5% significance in any hypothesis tests you perform, and provide a summary of your conclusions. Where appropriate, make comparisons with other levels of significance (2%, 1%).
- Sample size calculation: You can assume that a 95% confidence level is appropriate. You may include comparisons for 90% and 99% and a recommendation for the appropriate sample size.
- To answer some questions you may need to make certain assumptions about the data set we are using. Mention these in your data analysis, where relevant. There is no need to mention this in the memo.

***Note**: There is an Appendix at the end of each Chapter of the Prescribed Textbook which describes the basic Excel steps associated with that Topic. Chapters 1 to 9 are applicable for this assessment.*

Submission

Your completed assignment should be in **two** separate files:

- **Data Analysis (Part A):** An Excel document containing separate tabs/ worksheets with charts/ tables/ graphs for each question
- **Report (Part B):** A Word document of no more than 2000 words which is **not** to contain any charts/ tables/ graphs
- All interpretations should be presented in your "Report" and the excel document should only contain your *intermediate* analysis and final output

The assignment is to be submitted to the MIS770 assignment two dropbox in Deakin's Cloud Campus **before 11:59pm on Thursday 23rd January 2020**.
Note: The Cloud Unit site is the ONLY method of submission acceptable.

Please name/rename each file using the format *yourstudentid.docx* and *yourstudentid.xlsx*. Failure to follow this naming convention may lead to a delay in receiving feedback and marks.

When you are required to submit an assignment through your CloudDeakin unit site, you will receive an email to your Deakin 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.

Extensions

Extensions will only be considered if a draft assignment is attached with your request for an extension which shows progress has been made, and documentary evidence for the extension. Applications after **5pm** on **Thursday 23rd January 2020** will not be accepted, and submissions after the due date/time without an approved extension will be considered late.

Extensions are only granted in extreme circumstances, such as ongoing health, personal hardship or work-related problems. *Temporary illnesses, normal work pressures, multiple assignments due at the same time, failure to keep backups, technology failure, etc. are not reasons for an extension.*

If your extreme circumstance prevented you from applying for an extension before the due date, then you must apply for special consideration for the assignment.

Notes:

Penalties for late submission: The following *marking penalties* will apply if you submit an assessment task after the due date and time without an approved extension:

- **5% will be deducted from available marks for each day (or part thereof) up to five days**, and work that is submitted more than five days after the due date will not be marked. You will receive 0% for the task. 'Day' means calendar day for electronic submissions.
- The Unit Chair may refuse to accept a late submission where it is unreasonable or impracticable to assess the task after the due date.
- For more information about academic misconduct, special consideration, extensions, and assessment feedback, please refer to the document "**Your rights and responsibilities as a student in this Unit"** in the first folder next to the Unit Guide of the Resources area in the CloudDeakin unit site.

Building evidence of your experiences, skills and knowledge (Portfolio) - Building a portfolio that evidences your skills, knowledge and experience will provide you with a valuable tool to help you prepare for interviews and to showcase to potential employers. There are a number of tools that you can use to build a portfolio. You are provided with cloud space through OneDrive, or through the Portfolio tool in the Cloud Unit Site, but you can use any storage repository system that you like. Remember that a Portfolio is YOUR tool. You should be able to store your assessment work, reflections, achievements and artefacts in YOUR Portfolio. Once you have completed this assessment piece, add it to your personal Portfolio to use and showcase your learning later, when applying for jobs, or further studies. Curate your work by adding meaningful tags to your artefacts that describe what the artefact represents.

Supporting Materials

You will find the data in the file called **KingfisherbayData.xlsx**, on CloudDeakin for download. Please ensure you read the relevant sheets in that file. In particular, ensure you read the sheet called "Data Description". Save a copy for your assignment purposes. Also ensure you read the information which follows.

Marking Rubric

	Poor/Needs Improvement	Satisfactory	Good	Very Good	Exemplary
Executive summary **(Marks: 10)**	0 points Does not communicate any of the main findings of the analysis in an accurate or useful way, or the findings are basic. 0 – 4.5 Marks	5 points Explains most of the main findings of the analysis accurately and enables reader to draw some reasonable conclusions. 5 - 5.5 Marks	6 points Explains nearly all of the main findings of the analysis accurately and enables reader to draw mostly reasonable conclusions. 6 - 6.5 Marks	7 points Provides detailed and accurate descriptions of the most important features of the analysis along with appropriately qualified conclusions. 7 - 7.5 Marks	10 points Provides an outstanding descriptions and conclusions that is carefully considered and insightful. 8 - 10 Marks
Data Analysis and Graphics **(Marks: 40)**	0 points Uses irrelevant or inappropriate techniques to analyse the data, or Data analysis and visualisation tools have been used to analyse the data but in an incomplete or inaccurate manner. A very poor presentation of the analysis, or the analysis does not follow principles of good graphical display. 0 – 19.5 Marks	20 points Uses appropriate data analysis and visualisation tools to analyse the data but there are several errors in the analysis. The presentation of the analysis is satisfactory. 20 – 23.5 Marks	24 points Uses appropriate data analysis and visualisation tools to analyse the data but there are some errors in the analysis. The presentation of the analysis is of a respectable standard. 24 – 27.5 Marks	28 points Comprehensive analysis of the data using appropriate techniques, but there are some minor errors in the analysis. Uses data visualisations to understand the patterns in data. The analysis is well organised and follows principles of good graphical display. 28 – 31.5 Marks	40 points Skilful and comprehensive analysis of data using many different techniques. Uses data visualisations to produce novel insights. An exemplary presentation of the analysis. 32 – 40 Marks
Interpreting Results and Communication **(Marks: 40)**	0 points Does not communicate any of the main findings of the analysis in an accurate and/or useful way, or the interpretation and communication of findings is at a basic level. The written communication is unprofessional or difficult to follow and contains numerous spelling or grammatical errors. 0 – 19.5 Marks	20 points Explains most of the main findings of the analysis accurately and enables the reader to draw some reasonable conclusions. The written communication is clear and easy to follow but it contains minor spelling or grammatical errors. 20 – 23.5 Marks	24 points Explains nearly all of the main findings of the analysis accurately and enables the reader to draw mostly reasonable conclusions. The written communication is clear and easy to follow and generally free of spelling or grammatical errors. 24 – 27.5 Marks	28 points Provides detailed and accurate descriptions of the most important features of the analysis along with appropriately qualified conclusions. The written communication is professional, easy to follow and has a good structure. 28 – 31.5 Marks	40 points Provides an outstanding descriptions and conclusions that is carefully considered and insightful. The written communication is truly professional, logical and easy to follow. 32 - 40 Marks
Overall Assignment Presentation **(Marks: 10)**	0 points A very poor presentation of the analysis or it is mostly disorganised. The memorandum is un-professional, difficult to follow and/or contains numerous spelling or grammatical errors. 0 – 4.5 Marks	5 point The presentation of the analysis is satisfactory. The memorandum is clear and easy to follow but it contains minor spelling or grammatical errors. 5 -5.5 Marks	6 points The presentation of the analysis is of a respectable standard. The memorandum is clear and easy to follow and generally free of spelling or grammatical errors. 6 -6.5 Marks	7 points The analysis is well organised and follows principles of good graphical display. The memorandum is professional, easy to follow with good structure. 7 -7.5 Marks	10 points An exemplary presentation of the analysis. The memorandum is truly professional, logical and easy to follow. 8 - 10 Marks