MIS771 Descriptive Analytics and Visualisation	DEAKIN BUSINESS SCHOOL	AACSB ACCREDITED	EFMD EPAS ACCREDITED

Assignment Two

Background

This is an **individual** assignment, which requires you to analyse a given data set, interpret and draw conclusions from your analysis, and then convey your conclusions in a written technical report to an expert in Business Analytics.

Percentage of final grade	**35%**
The Due Date and Time	**11.59 PM Sunday 15th September 2019 (AEST)**

Submission instructions

The assignment must be submitted by the due date electronically in CloudDeakin. When submitting electronically, you must check that you have submitted the work correctly by following the instructions provided in CloudDeakin. Please note that we will NOT accept any paper or email copies, or part of the assignment submitted after the deadline.

No extensions will be considered unless a written request is submitted and negotiated with the unit chair before Thursday 12th September 2019, 5:00 PM. Please note that assignment extensions will only be considered if you attach your draft assignment with your request for an extension.

You must keep a backup copy of every assignment you submit (that is, the work you have done to date) until the assignment has been marked. In the unlikely event that an assignment is misplaced, you will need to submit your backup copy. Work you submit will be checked by electronic or other means to detect collusion and/or plagiarism.

When you submit an assignment through your CloudDeakin unit site, you will receive an email to your Deakin email address confirming that the assignment 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.

Penalties for late submission: The following marking penalties will apply if you submit an assessment task after the due date without an approved extension: **5%** will be deducted from available marks **for each day 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 days or part thereof. 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.

The assignment uses the file *supermart.xlsx,* which can be downloaded from CloudDeakin. Analysis of the data requires the use of techniques studied in Module 2.

Assurance of Learning

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

Graduate Learning Outcome (GLO)	Unit Learning Outcome (ULO)
GLO1: **Discipline-specific knowledge and capabilities** - appropriate to the level of study related to a discipline or profession. **GLO3**: **Digital Literacy** - Using technologies to find, use and disseminate information **GLO5: Problem Solving** - creating solutions to authentic (real-world and ill-defined) problems.	**ULO 1**: Apply quantitative reasoning skills to solve complex problems. **ULO 2**: Use contemporary data analysis and visualisation tools and recognise the limitation of such tools.

Feedback before submission

You can seek assistance from the teaching staff to ascertain whether the assignment conforms to submission guidelines.

Feedback after submission

An overall mark together with suggested solutions will be released via CloudDeakin, **usually within 15 working days**. You are expected to refer and compare your answers to the suggested solutions to understand any areas of improvement.

Case Study (Background to Mitchies Superstore)

Mitchies Superstore is one of Australia's leading supermarket chains. There are 500 stores in the chain. Originating from a family-based chain of general stores, Mitchies Superstore now has stores all over Australia, with the first one being established 55 years ago. Regarding operation, individual store management has wide-ranging powers about the day-to-day operations of their stores. However, Mitchies Superstore's strategic planning and direction take place in the company Head Office in Dandenong.

Despite their successful operations and strong financial turnovers in the last two years, Mitchies Superstore is forecasting a shift in the business climate within the next five years. This is a result of ever-increasing competition in the grocery supermarket sector. Now more than ever, Mitchies Superstore management team feels the need to ensure a good understanding of their business performance. The Mitchies Superstore Head Office is slightly confused about the lack of enthusiasm of store managers to open their stores on Sundays or to open an online sales channel given that Mitchies Superstore Head office has invested heavily on digital platforms, self-checkout machines and staff. Also, they are planning to put in place a formal procedure to forecast their Sales.

Subsequently, Mitchies Superstore has approached BEAUTIFUL-DATA (a market research company) and asked them to conduct a study to understand the characteristics of Mitchies Superstore's stores and their business performance.

The Data

For this study, BEAUTIFUL-DATA has collected two sets of Data:

1. The data related to stores were extracted from the company's datamarts. It is a random sample of 150 stores in the supermarket chain. A complete listing of variables, their definitions, and an explanation of their coding are provided in Working Sheet "Stores-Variable Description.
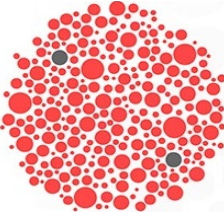2. Time-Series data is available on Working Sheet "Quarterly Sales".

Your Role as a BEAUTIFUL-DATA Data Analyst Intern

You are a postgraduate student doing an internship at BEAUTIFUL-DATA. The research team manager (Todd Nash, with a PhD in Data Science and a Master Degree in Digital Marketing) has asked you to lead the data analysis process for the Supermarket project and directly report the results to him. You and Todd just finished a meeting wherein he briefed you on the vital purpose of the project.

Todd explained that a model should be built to estimate Sales. Therefore, the first goal is to identify key factors that influence Sales. The second goal is to understand the relationship between "number of competitors" and" Sales". He is also interested in gaining insights into factors that influence stores to open an online sales channel. The final goal is to construct a forecasting model, which forecast Sales in the upcoming four quarters. From these insights, Todd and consequently Mitchies Superstore will be in an excellent position to develop plans for the next financial year.

Todd also allocated relevant research tasks and explained his expectations from your analysis in the meeting. Minutes of this meeting are available on the next page.

Now, your job is to review and complete the allocated tasks as per this document.

	BEAUTIFUL-DATA, 727 Collins St, Docklands VIC 3008 Phone: (+61 3 212 66 000) info@beautiful-data.com.au	Reference	PH-102 Supermarket Project
		Revised	25th August 2019
		Level	Expert Analysis

Meeting Chair	Todd Nash				
Date	25th August 2019	Time	11:00 AM	Location	BEAUTIFUL-DATA F3.101
Topic	Supermarket Project – Analytics Details				

Meeting Purpose:	**Specifying and Allocating Data Analytics Tasks**	
Discussion items:	• Variable(s) <u>description</u>. • Predict <u>Sales</u>. • Predict <u>the likelihood of a store opening an online sales channel</u>. • Forecast <u>Sales </u>in the upcoming four quarters. • Produce <u>a technical report</u>.	

Detailed Action Items	Who:
	
	Graduate Intern
		What:
		1. Provide an overall summary of the following variables:
		   1.1. Sales
		   1.2. Open on Sundays
		   1.3. Online Sales Channel
		
		2. Using an appropriate model building process, build a model to predict Sales.
		
		3. Todd has done a separate regression analysis to predict Sales using the number of competitors and stores open on Sundays. He believes that the relationship between Sales and the number of competitors should be weaker for those stores that are open on Sundays. Your task here is to test Todd's assumption by modelling the interaction between the predictors mentioned above and the target variable. Comment whether there is sufficient evidence that the interaction term makes a significant contribution to the model.
		
		4. Using an appropriate model building process, build a model to predict the likelihood of a store opening on Sunday.
		
		5. Finalise Todd's model to predict the likelihood of a store opening an online sales channel:

		5.1. Todd has already done an initial analysis for this task. Based on his analysis, he has narrowed down the key predictors of the likelihood of a store opening an online sales channel to "Manager's Age, Experience and Gender". Your task is to continue his work and develop a predictive model to ascertain the "likelihood of a store opening an online sales channel".
		5.2. Todd is specifically interested in understanding the probability of stores which meet the following criteria to open an online sales channel: Those stores with managers, a) in their mid-thirties; b) with varying levels of Management Experience (i.e. 2-16 years?); c) and across both, male and female store managers. He believes that the age, experience and the gender of the store manager may influence the decision to open an online sales channel. Therefore, it is essential for the client's management team to know whether effort and money should be put into recruiting tech-savvy young managers. Accordingly, your job is to visualise the predicted probability of a store opening online sales channel with the attributes described earlier.
		6. Develop a time-series model to forecast Sales in the next four fiscal quarters.
		7. Produce a written report detailing ALL aspects of your analysis. Your description should be as detailed as possible and should describe ALL critical outputs of your analysis. Make sure to provide recommendations to the client's management team that will guide them to improve their decision-making. The results of your analyses should drive your recommendations/insights.
Next meeting	Monday 07th October	

Appendix: Explanatory Notes

To accomplish allocated tasks, you need to examine and analyse the dataset (**supermart.xlsx**) thoroughly. Below are some guidelines to follow:

Task 1 – Summarising Dependent Variables

The purpose of this task is to analyse and explore key features of these variables individually. At the very least, you should thoroughly investigate relevant summary measures/charts and graphs of these variables. Proper visualisations should be used to illustrate key features.

Your **technical report** should describe ALL key aspects of each variable.

Task 2. – Model building (predicting Sales)

You should follow an appropriate model building process. All steps (**including pre and post model diagnostics**) of the model building process should be included in your analysis. You can have as many Excel worksheets (tabs) as you require to demonstrate different iterations of your regression model (i.e., 2.2.a., 2.2.b., 2.2.c. etc.). You must make reasonable/realistic/practical assumptions about the parameters you are working with in Task 2.

Your **technical report** should clearly explain why the model might have undergone several iterations. Also, you must provide a detailed interpretation of ALL elements of the **final** model/regression output.

Task 3. – Interaction effect

To accomplish this task, you need to develop a regression model using **ONLY** the factors discussed in the meeting (Task 3). In other words, this section of analysis is separate from the regression model constructed in Task 2. You must make reasonable/realistic/practical assumptions about the parameters you are working with in Task 3.

Your **technical report** should clearly explain the role of each variable included in the model. A proper visualisation technique should be used. Make sure you interpret all relevant outputs in detail and provide managerial recommendations based on the results of your analysis.

Task 4 – Model building (likelihood of a store opening on Sunday)

You should follow an appropriate model building process. All steps (**including pre and post model diagnostics**) of the model building process should be included in your analysis. You can have as many Excel worksheets (tabs) as you require to demonstrate different iterations of your regression model. You must make reasonable/realistic/practical assumptions about the parameters you are working with in Task 4.

You are required to discuss all details of your predictive model/logistics regression output.

Task 5.1 – Model building (likelihood of a store opening an online sales channel)

You should start building the predictive model by including **ONLY** the variables listed in the 'minutes of the meeting – Task 5.1'. You must make reasonable/realistic/practical assumptions about the parameters mentioned in Task 5.

You are required to discuss all details of your predictive model.

Task 5.2. – Visualising and interpreting predicted probabilities

Your **technical report** must include the predicted probability visualisation and be supplemented by practical recommendations to Superstore's Management. These recommendations should answer the following question:

"How changes in Experience, Age and the Gender of the manager may affect the predicted probability of opening an online sales channel."

Task 6 – Forecasting Sales

Mitchies Superstore's past quarterly Sales are given in the Excel file. Your job is to develop a proper forecasting model to predict turnover for **the next four quarters**.

In your **technical report**, you must explain the reason for selecting the forecasting method to predict future sales. The report also must include a detailed interpretation of the **final** model (e.g. a practical interpretation of the time-series model, choices about smoothing techniques etc.).

Task 7. – Technical report

Your **technical report** must be as comprehensive as possible. ALL aspects of your analysis and final outputs must be described/interpreted in detail. Remember, your audience are experts in analytics and expect **a very high standard of work** from your report. High standards means **quality content** (demonstrated attention to details) as well as an **aesthetically appealing report**.

Note: The use of technical terms is acceptable in this assignment.

Your report should include an **introduction** as well as a **conclusion**. The introduction begins by highlighting the main purpose(s) of analysis and concludes by explaining the structure of the report (i.e., subsequent sections). The conclusion should highlight the key findings and explain the main limitations.

Submission Guide

The assignment consists of **two** parts: 1) *Analysis* and 2) *Technical Report*. You are required to submit both your technical report (**Word.docx** document only) and the analysis (**Excel.xlsx** file only).

1) Analysis (excel.xlsx)

The analysis should be submitted in the appropriate worksheets in the Excel file. Each step in the model buildings should be included in a separate tab (e.g. 2.2.a., 2.2.b., …; and 3.2.a. 3.2.b., …). Add more worksheets if necessary.

Before submitting your analysis make sure it is logically organised, and any incorrect or unnecessary output has been removed. Marks will be deducted for poor presentation or disorganised/incorrect results. Your worksheets should follow the order by which tasks are allocated in the minutes of the meeting document.

> **Note**: Give the Excel file the following name **A2_YourStudentID.xlsx** (use a short file name while you are doing the analysis.

2) Technical Report (word.docx)

Your technical report consists of four sections: **Introduction**, **Main Body**, **Conclusion**, and **Appendices**. The report should be approximately 2,500 (± 300) words.

Use proper headings (i.e., 1., 2.1., 2.2., …) and titles in the main body of the report. Use sub-headings where necessary.

Visualisations / statistical output allowed in the report are:

1. Interaction effect plot
2. Predicted probability plot.

All other visualisations should be ideally in the *Appendices* (appendices are not included in the word count).

Make sure these outputs are **visually appealing**; have **consistent formatting style,** and **proper titles** (title, axes titles etc.); and are **numbered correctly**. Where necessary, refer to these outputs in the main body of the report.

> **Note**: Give the report the following name **A2_YourStudentID.docx**.

Sample Rubric

Criteria Name	Criteria Weight		Not Attempted	Needs Improvement	Satisfactory	Good	Very Good	Exemplary
Analysis (35%) GLO1 GLO3	Task 1	5%	Does NOT use any appropriate descriptive analysis tool.	Use irrelevant or inappropriate descriptive analysis tool.	Use appropriate descriptive analysis tool BUT there are errors in the analysis.	Most relevant descriptive analysis tools are used BUT there are minor errors in the analysis.	All relevant descriptive analysis tools are used with minor errors in the analysis.	Skilful and comprehensive descriptive analysis of all relevant variables using variety of techniques.
	Task 2 and 3	10%	Does NOT use any appropriate bivariate exploratory data analysis tool.	Use irrelevant or inappropriate bivariate analysis tool.	Use appropriate bivariate analysis tool to identify IVs, BUT there are errors in the analysis.	Appropriate bivariate analysis tool is used, BUT NOT all relevant IVs are identified.	All relevant IVs are identified using proper bivariate analysis technique, BUT minor issues noted.	Skilful and comprehensive analysis of bivariate relationships is presented and all relevant IVs are identified.
			Either inappropriate predictive model is developed and/or analysis lacks All steps of mode-building process missing.	Relevant IVs are included in the predictive model, BUT some steps of model-building process missing.	A predictive model is developed with All model-building steps included, BUT the final model is incorrect and/or there are many errors in the analysis.	An appropriate predictive model is developed with All model-building steps presented BUT there are minor errors in the analysis.	The final model includes those IVs that have predictive power with All steps in model-building process clearly presented.	Model-building process is presented in logical/comprehensive manner AND the final model is correct.
			Interaction analysis is missing.	Interaction analysis is incorrect.	Analysis of interaction effects is presented BUT there are many errors.	Interaction analysis is done correctly BUT wrong visualisation technique is used.	Interaction analysis is presented accurately with proper visualisation technique BUT with minor errors.	Masterful analysis of interaction effects supplemented by a correct visualisation.
	Task 4 and 5	10%	Either inappropriate predictive model is developed and/or analysis lacks All steps of mode-building process missing.	Relevant IVs are included in the predictive model, BUT some steps of model-building process missing.	A predictive model is developed with All model-building steps included, BUT the final model is incorrect and/or there are many errors in the analysis.	An appropriate predictive model is developed with All model-building steps presented BUT there are minor errors in the analysis.	The final model includes those IVs that have predictive power with All steps in model-building process clearly presented.	Model-building process is presented in logical/comprehensive manner AND the final model is correct.
			Predicted probabilities are not calculated and/or a visualisation is missing.	Not All Predicted probabilities are calculated and/or a visualisation is missing.	All predicted probabilities are calculated and a visualisation is presented BUT there are many errors in the analysis.	All predicted probabilities are calculated and a visualisation is presented BUT there are minor errors in the analysis.	All predicted probabilities are calculated correctly and a proper visualisation is presented.	A skilful and comprehensive analysis of predicted probabilities is presented along with a well-structured visualisation.
	Task 6	10%	Does not use any appropriate time-series techniques.	Uses irrelevant or inappropriate techniques to analyse the time-series and/or there are many errors in the analysis.	A relevant time-series model developed but there are many errors in the analysis.	A relevant time-series model is developed and but there are minor errors in the analysis.	Time-series model is developed correctly and relevant measure(s) for evaluating the model quality is presented.	Time-series model developed correctly and presented in a clear and logical fashion including relevant visualisations.
Interpretation GLO1 GLO5	Task 7	60%	Does not communicate any of the main findings of the analysis in an accurate or meaningful way.	Interpretation and communication of findings is at a basic level or does not adequately explain the main findings of the analysis.	Explains the main findings of the analysis accurately and enables reader to draw some reasonable conclusions.	Provides an accurate description of the most - BUT NOT ALL - important features of the analysis, with appropriate conclusions.	Provides very detailed and accurate descriptions of the most important features of the analysis.	Provides an outstanding description and conclusion of All relevant analysis/visualisation outputs. Interpretation of results are novel and insightful.
Technical Report GLO1 GLO5	-	5%	The technical report is poorly structured and/or few sections missing with a poor use of technical language.	The technical report is poorly structured. Only few analysis outputs are presented in appendix. Language is difficult to follow with many grammatical errors noted.	The technical report is well-structured with All required sections included. Most relevant analysis outputs are included in appendix. Communication is NOT clear throughout the report and grammatical errors noted.	The technical report is well-structured with All sections included. All relevant analysis outputs are included in appendix. Communication is clear with NO grammatical errors noted.	The technical reports on par with a professional report. All relevant analysis outputs are presented in appendix in a logical order. Written communication is clear, easy to follow and has a structure.	The technical report is masterfully structured. All relevant analysis outputs are included in appendix. Outputs are visually appealing, and follow a consistent formatting style. Language is truly professional and easy to follow
OVERALL	100% (Equivalent of 35 Marks)		0-29%	30-49%	50-59%	60-69%	70-79%	80-100%
Overall Description			Fail (N)		Pass (P)	Credit (C)	Distinction (D)	High Distinction (HD)