

# MIS771 Descriptive Analytics and Visualisation



## Assignment One

### Background

This is an **individual** assignment. You need to analyse a given data set, and then interpret and draw conclusions from your analysis. You then need to convey your conclusions using plain language in a written report to a person with little or no knowledge of Business Analytics.

Percentage of the final grade	30%
The Due Date and Time	11.59 PM <b>Sunday 18<sup>th</sup> August 2019</b>

### 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 15<sup>th</sup> August 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.

The assignment uses the dataset file **A1.xlsx**, which can be downloaded from CloudDeakin. Analysis of the data requires the use of techniques studied in Module-1.

## Assurance of Learning

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

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

### 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 feedback will be released via CloudDeakin, **usually within 15 working days**. You are expected to refer and compare your answers to the feedback to understand any areas of improvement.

## Case Study

You are **Natalia Navarska**, a data analyst in the Research and Analysis group at *Financial Review Magazine*. Your primary role is to evaluate new products and services. You are often required to report outcomes of your analysis to senior editors at the Magazine who have little or no knowledge of data analysis.

Of specific interest to *Financial Review* magazine are the increasing numbers of companies that offer brokerage services for car insurance and potentially what this means for consumers. **An insurance broker is an independent insurance agent who works with many insurance companies to find the very best available policies for his or her customers.** Most of these brokers are advertising that they can save vehicle owners hundreds of dollars each year on insurance premiums.

Just recently, your research and analysis group secured a dataset from the Insurance Brokers Association (IBA), which is **a random sample of 400 customers** who obtained the services of car insurance brokers. You have performed an exploratory analysis and have emailed the results (see pages 6-7) to **Edmond Kendrick**, one of the senior editors of *Financial Review Magazine*.

Edmond has replied to your email regarding the Insurance Brokers. His email is reproduced next page:

## Email from Edmond

---

**To:** Natalia Navarska

**From:** Edmond Kendrick

**Subject:** Analysis of car insurance brokerage services

---

Hi Nat,

Thank you for the comprehensive analysis and notes. Now I am more curious about what else could we learn from analysing the dataset.

1. From what I can gather from your notes, iChoose was able to save their customers more money than other brokers. Can I now conclude that iChoose, on average, can save more on insurance premiums than uChoose?
2. Your analysis of 400 customers showed that the proportion of dissatisfied (i.e. either 'Dissatisfied' or 'Very Dissatisfied') urban customers is smaller than the proportion of dissatisfied rural customers. Can we argue that this difference would hold across all urban and rural customers?
3. I did my own analysis of the sample and came to the following conclusions:
  - a. The average savings on insurance premiums differ between rural and urban customers.
  - b. On average, customers with 'Agreed Value' policy saved more on their insurance premiums than the customers with 'Market Value' policy;
  - c. The proportion of female customers with a diamond level no claim bonus rating (NCBR) is less than male customers with a diamond level no claim bonus rating (NCBR);What would be great is if you can verify my findings and tell me how much the difference is in each of the three scenarios mentioned above.
4. I would like you to expand the analysis and look at whether:
  - a. The average savings on insurance premiums significantly differ between Victoria, NSW and Queensland.
  - b. The average savings on insurance premiums significantly differ between 4WD, Luxury and Sports car.
5. Does the proportion of customers who approached their insurance provider before reaching out to a broker differ between the insurance providers?
6. I asked Raj to design an experiment to see the effects of the valuation method and the vehicle type on savings on insurance premiums, he sent me a table with some numbers (see Appendix-A). Can you complete the analysis?

I look forward to your response.

Regards

*Eddie*

### ***Appendix- A: Data for the experiment prepared by Raj***

---

	Vehicle Type					
Valuation Method	4WD	Family	Sport	Luxury		
Agreed Value	1068	169	1799	966		
	128	150	680	1144		
	98	-59	373	893		
	560	22	143	1144		
	429	108	442	629		
Market Value	104	54	99	1273		
	72	0	156	247		
	311	94	1084	357		
	146	84	357	676		
	135	-10	131	366		

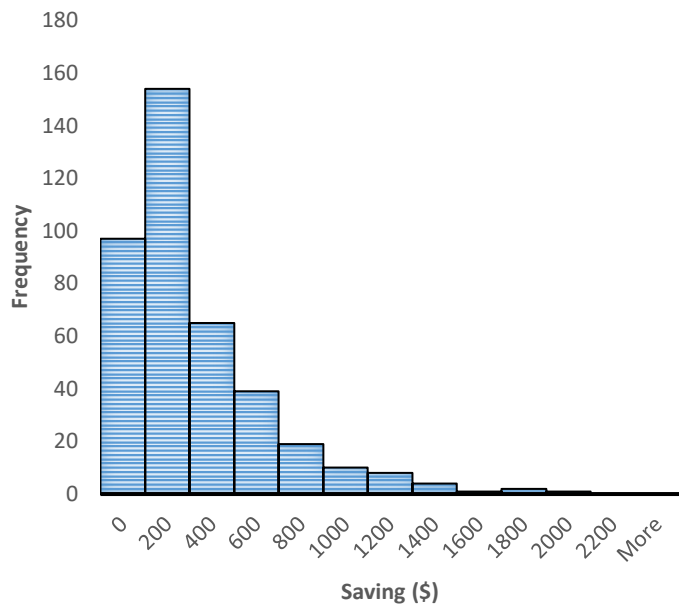
## An Extract of the Analysis and Notes Prepared by Nat

- A summary of Savings:**

Savings	
Mean	229.64
Standard Error	16.03
Median	113
Mode	0
Standard Deviation	320.56
Sample Variance	102759.59
Kurtosis	5.46
Skewness	2.08
Range	2043
Minimum	-87
Maximum	1956
Sum	91857
Count	400
Q1	12
Q3	357
IQR	345
LF	-505.5
UF	874.5
OUTLIERS	YES

Saving Outcome	Count of Customers
Not benefited from (saving < 0)	72
Neither benefited nor lost (saving = 0)	25
Benefited from (saving > 0)	303

### HISTOGRAM: SAVING



- Summary of Saving by Broker**

	iChoose	uChoose	vChoose	yChoose
Mean	262.442	230.847	137.381	204.188
Standard Error	25.883	36.672	14.330	31.575
Median	127	94.5	123.5	100
Mode	0	0	294	0
Standard Deviation	356.766	311.169	92.868	309.368
Sample Variance	127281.930	96825.934	8624.437	95708.659
Kurtosis	4.121	4.678	-0.461	6.102
Skewness	1.826	1.934	0.442	2.210
Range	2034	1645	392	1738
Minimum	-78	-69	-31	-87
Maximum	1956	1576	361	1651
Sum	49864	16621	5770	19602
Count	190	72	42	96
Q1	0	24	65.5	0
Q3	412.5	388.75	200	338
IQR	412.5	364.75	134.5	338
LF	-618.75	-523.125	-136.25	-507
UF	1031.25	935.875	401.75	845
OUTLIERS	YES	YES	NO	YES

- *Customer Satisfaction*

Customer Satisfaction	Count of Customers
Very Dissatisfied	35
Dissatisfied	57
Satisfied	174
Very Satisfied	134
Total	400

- *Customer Satisfaction by Area*

Area	Satisfaction				Total
	Very Dissatisfied	Dissatisfied	Satisfied	Very Satisfied	
Rural	10	23	32	30	95
Urban	25	34	142	104	305
Total	35	57	174	134	400

### ***Notes to Edmond***

#### **Savings:**

From a sample of 400 customers,

- On average, car insurance brokers saved their customers \$113 (median).
- The middle 50% of customers saved between \$12 and \$357; a quarter of the customers saved at most \$12; three-quarter of the customers saved no more than \$357.
- The savings ranged from a loss of \$87 to a substantial gain of \$1956.
- Almost 40% of the customers, saved between \$1 and \$200 on their current insurance premiums; car insurance brokers have shown their ability to find an appropriate policy for most of their customers.
- The bulk of the customers have relatively low (in few cases none at all) annual savings on premium, with a relatively small number having high savings. 89% of customers saved up to \$600; Only 4% of consumers saved between \$1000 and \$2000; with only 1%; shows that brokers have the ability to save consumers a massive amount (more than \$1600) on their annual premiums but the prospect of making such savings is low.
- 24% of consumers paid a higher premium than previously or did not save on their annual premium.
- 18% of customers made a loss; the brokers are claiming to save most customers hundreds of dollars, but the discussion about the possibility of customers paying more money for the insurance is missing.

## SUBMISSION

The assignment consists of **two** parts: **Analysis** and **Report**. You are required to submit both your written report and analysis.

### **Guidelines for Data Analysis**

Read the case study and questions asked by Edmond carefully. Then spend some time reviewing the data to get a sense of the context. The analysis required for this assignment involves material covered in **Module 1**, with the corresponding tutorials being a useful guide.

The analysis should be submitted in the appropriate worksheets in the Excel file. Each question from the email should be analysed in a separate tab (e.g. Q1, Q2 ... or Q3.1, Q3.2 ...). You need to add these. Before submitting your analysis, make sure it is logically organised, and any incorrect or unnecessary output has been removed. Marks will be penalised for poor presentation or disorganised/incorrect results.

For all questions in the email, you can assume that:

- **95 % confidence level** is appropriate for confidence intervals and;
- **5.0 % level of significance** (i.e.  $\alpha = 0.05$ ) is appropriate for any hypothesis tests.

You can complete all data analysis using the Excel templates provided in the assignment data file. In choosing the technique to apply for a given question, keep the following in mind:

- Are we dealing with a numerical variable or categorical variable?
- Are we dealing with one sample, two samples or more than two samples situation?
- Are we dealing with independent samples or paired-samples situation?
- Each question must be answered using the most appropriate technique.
- For all hypothesis questions, please formulate your hypotheses, and state them in both notation and words clearly.
- Even though question(s) may lead you to inferential technique, consider conducting a descriptive analysis of the sample data first.

### **ATTENTION!**

- When you have established that there is a difference between two means or proportions, we expect you to estimate and report the difference.
- When you have established that there is a difference between two or more means or proportions, we expect you to follow up with an appropriate multiple comparison procedure.

You may need to make certain assumptions about the dataset we are using to answer some questions. For other questions, there will be technical/statistical assumptions that you need to make; for example, whether to use an equal or an unequal variance test...etc. *You need to consider and incorporate any violations of assumptions such as unequal sample sizes as limitations of your analysis in your report.*

**Note:** Give the Excel file the following name **A1\_YourStudentID.xlsx** (use a short file name while you are doing the analysis).



### ***Guidelines for your Business Report***

Once you have completed your data analysis, you need to summarise the key findings for each question and write a response to Edmond in a report format. Your business report consists of four sections: **Introduction**, **Main Body**, **Conclusion**, and **Appendices**. The report should be around 1,500 words.

Use proper headings (e.g. Q1, Q2 ... or Q3.1, Q3.2...) and titles in the main body of the report. Use sub-headings where necessary.

Keep the language plain and the explanations brief. **That is, avoid the use of any unnecessary technical statistical jargon.** Your reader may not necessarily understand even the simplest statistical term. Thus your task is to convert your analysis into plain, easily understandable expressions.

#### ***General instructions:***

- You **MUST** report both **descriptive** and **inferential** analysis results. Otherwise, marks will be deducted.
- The report is to be written as a **stand-alone document** (assume Edmond will only read your written report). Thus, you should not have any direct references in the report to your analysis.
- Your report may include relevant excel outputs including templates, tables, charts, and graphs **but ONLY as Appendices** (appendices are not included in the word count).
- Make sure these outputs in the Appendix are **visually appealing**, have a **consistent formatting style** and **proper titles** (title, axes titles, etc.), and are **numbered correctly**.
- **The introduction** begins by highlighting the main purpose(s) of the analysis and concludes by explaining the structure of the report (i.e., subsequent sections). **The conclusion** should highlight the key findings of the analyses and **explain the main limitations** (if any).
- **Marks will be deducted for the use of technical terms**, irrelevant material, and poor presentation/organisation.

When you have completed the report, it is a useful exercise to leave it for a day, return to it and then re-read. Does it flow easily? Does it make sense? Can someone without prior knowledge follow your written conclusions? Often, on re-reading, you become aware that you have made some points clumsily, and you find that you can re-phrase them much more clearly.

**Note:** Give the report the following name **A1\_YourStudentID.docx** or **pdf**.