



Analysis of Car Insurance Brokerage Services

Mirna Arivalagan - 220142881 | Descriptive Analytics and Visualisation – MIS771
Trimester 2 2021

Contents

Table of Contents	Error! Bookmark not defined.
Introduction	2
Question 1:	2
Question 2:	3
Question 3:	3
Question 4(A):	4
Question 4(B):	4
Question 5:	4
Conclusion:	5
Appendix	6

Email:

To: Edmond Kendrick

From: Mirna Arivalagan

Subject: Analysis of car insurance brokerage service

Thank you for your e-mail. I have completed a thorough analysis on the sample data set that has been provided to understand the relationship between insurance premium savings and the customers gender, location, valuation method, vehicle type and broker that has been used. I hope you find the following report useful in helping make some further decisions.

Introduction

Between 2016 and 2021, it is estimated that the value of motor vehicle insurance is around \$2.8 billion dollars in Australia. The motor vehicle insurance is the largest segment in Australia's general insurance market as of March 2021. The motor vehicle insurance market has been on a steady up trend since 2016 till March 2021.

In an ever-growing competitive market, it is important to understand what the impacts are of the increase in brokerage services in recent times to the customers. By understanding, how customers respond to the access to numerous brokerage services, a broker can use these learnings to ensure that the product that they are providing ticks all the consumers boxes, thus helping the broker remain relevant in a competitive market.

This report covers the analysis of about 400 randomly selected customers and types of brokerage services and insurance covers that the customers have used. It is designed to address if there are differences in premium savings across genders, valuation methods, vehicle types and across states in Australia. This report also analyses if there are differences in customer satisfaction depending on the locality and broker used.

Question 1:

In terms of understanding if there are differences in savings across genders, I have run some descriptive and inferential analysis (Table 1). Based on sample data, the overall average savings in premiums for customers under 30 years old is \$317.25. There is some variation in figures when this is split across genders, the average for females were higher than males, \$355.22 compared to \$290.23 (Graph 1). It appears that females tend to save more, however, to prove this, further testing was conducted which is covered below.

I have found that 50% of males saved \$143 or more, and 50% of females saved \$136 or more on premiums. It is also important to note that there is an average spread of \$413.20 from the average savings of \$290.23 for males, and a spread of \$473.19 from the average of \$355.22 for females, thus indicating that there is a wide variation in the sample data set provided. I've also found that the lowest amount of savings across genders is similar, -\$76 for males and -\$74 for females, however the highest amount in savings, there is quite a difference with males saving more at \$1956 compared to \$1603 for females (Table 1).

At a 5% significance level, there is insufficient evidence to conclude that females aged 30 and below saves more on insurance premiums compared to males (Table 2). So even though from the sample data set, it appears that females saved more, this isn't the case, and this could perhaps be attributed to the outliers that are in the data which can be seen in the box plot (Graph 2).

Question 2:

In terms of customer satisfaction across urban and rural areas, from the sample data set given, the proportion of dissatisfied rural customers was 38.95% which was much higher than the proportion of urban customers of 21.64% (Graph 3).

From the tests that were conducted on the sample data set provided, at 5% significance level, there is sufficient evidence to conclude that the true proportion of dissatisfied urban customers is less than rural customers. I am 95% confident that the true proportion of dissatisfied urban customers is between -7.24% to -27.37% less than that of rural customers (Table 3).

Question 3:

By analysing the sample data set by valuation methods to see if there is a difference in average savings between agreed valuations and market valuations, we can see that the overall average savings for all customers in the sample data set is \$235.45. When split out, agreed valuation customers on average saved more than market valuations, with an average of \$360.64 compared to \$219.97 (Graph 4). Overall, 50% of the agreed valuation customers saved \$257.50 or more and for market valuations 50% of customers saved \$105 or more. I've also found something interesting whereby for agreed valuations, the most common amount of savings is \$39, and market valuations, it is \$0, meaning a good amount of market valuation customers do no make a saving.

The minimum values are similar, -\$76 for agreed value compared to -\$87 for market value. There is however a difference in the highest amount of savings, market value customers highest amount is \$1956 compared to \$1651 for agreed valuations (Table 4). Again, there are a few outliers here that could be skewing the averages (Graph 5).

At 5% significance level, I have sufficient evidence to conclude that there is in fact a difference in premium savings between the agreed value valuation and market value valuation. Further to this, with 95% confidence, there is sufficient evidence to conclude that the true average savings across valuation methods differ by \$36.68 to \$244.64 (Table 5).

Question 4(A):

To understand if there is a difference in average savings on premiums between customers based in New South Wales (NSW), Victoria (VIC) and Queensland (QLD), we had to conduct an ANOVA analysis which is an analysis of variances. From the sample data, there was a total of 127 customers based in NSW, 89 in VIC & 61 in QLD. The highest average in savings was seen in NSW with an average of \$266.95, followed by VIC with \$242.34 and the lowest average was seen in QLD with \$240.36 (Graph 6). As with the others, there is also some outliers in this scenario that can be observed in the box plot (Graph 7).

At a 5% significance level, there is insufficient evidence to conclude that there is a significant difference in average savings across the 3 states. There is an 81% chance that the average savings are the same across the 3 states. (Table 6)

Question 4(B):

I have found that the brokerage with the highest number of satisfied customers was VChoose with 83% of their customers rating as highly satisfied, this is followed by UChoose with 76%. YChoose and IChoose had very similar percentages of satisfied customers. IChoose has 73% of their customers satisfied, and YChoose is 72% (Graph 8).

From the test, at a 5% significance level, there is insufficient evidence to conclude that the proportion of satisfied customers is not the same across all insurance brokers. There is a 97.73% chance that the proportion of satisfied customers are equal across all brokerages (Table 7).

Question 5:

As requested, I have run an experiment to understand if there is an interaction between valuation methods and vehicle types that impacts the average savings.

From the analysis done, I have found that at a 5% significance level, there is sufficient evidence to conclude that there is a difference in average savings between valuation methods. From the graph (Table 8), using the sample data set, we can visualize and identify that the average savings of the agreed valuation is higher than that of the market valuation. This is clearly prominent in the line chart, there is a large difference in 4WD & Sports customers, the difference is lowered for Family and Luxury customers (Graph 10).

Further to this, I have also found that at a 5% significance level, that there is a difference in the true average savings across vehicle types. Sports and Luxury had the highest amount of savings, with an average of \$647.90 and \$636.60. Family had the lowest average savings at \$72.40 and 4WD was in the middle with an average of \$364 (Table 8, Graph 9).

Finally, we tested to see if there was an interaction between the vehicle types and valuation methods, at a 5% significance level, we have insufficient evidence to conclude that there is an interaction across valuation methods and vehicle types that are increasing or reducing the average savings on insurance premiums (Table 8, Graph 10).

Conclusion:

To conclude, from the sample data provided, we were not able to find if there is a significant difference in savings across genders for customers aged under 30 years old. There were a few outliers in the data set, perhaps being able to analyse a larger data set would provide a clearer outcome. With 95% confidence, we were able to find that rural customers are on average less satisfied than urban customers.

Further analysis is needed to understand the relationship between savings on insurance premiums and valuation methods as the sample sizes in the data provided was unequal, market valuation had 356 customers which was more prominent than agreed valuation customers at 44. With 95% confidence, we were also able to conclude that the average savings between the states were similar. From the dataset, we have also found that customer satisfaction across the 4 brokerages is similar, with most of the customers rating the services as satisfied, which is a good sign that consumers are happy to use these services.

Finally, we have also found that luxury cars and sports cars that cost a lot more tend to have higher savings on insurance premiums, this is to be expected as the premiums for these cars are much higher compared to a non-luxury car, thus a small % on discount will have a higher dollar value. As highlighted through this report, there are some limitations with the tests that was run, these include outliers being present in the data set, unequal sample sizes, and when analysing the savings by vehicle types, the sample size was smaller than 30, hence we cannot assume that the data is normally distributed.

Through this experiment, we have also gone with the assumption that since the standard deviation is less than double of each other, the variances are then assumed to be equal. Further testing is required to support these assumptions. I would suggest running these tests again with a larger and equally distributed sample set, to support the findings in this report.

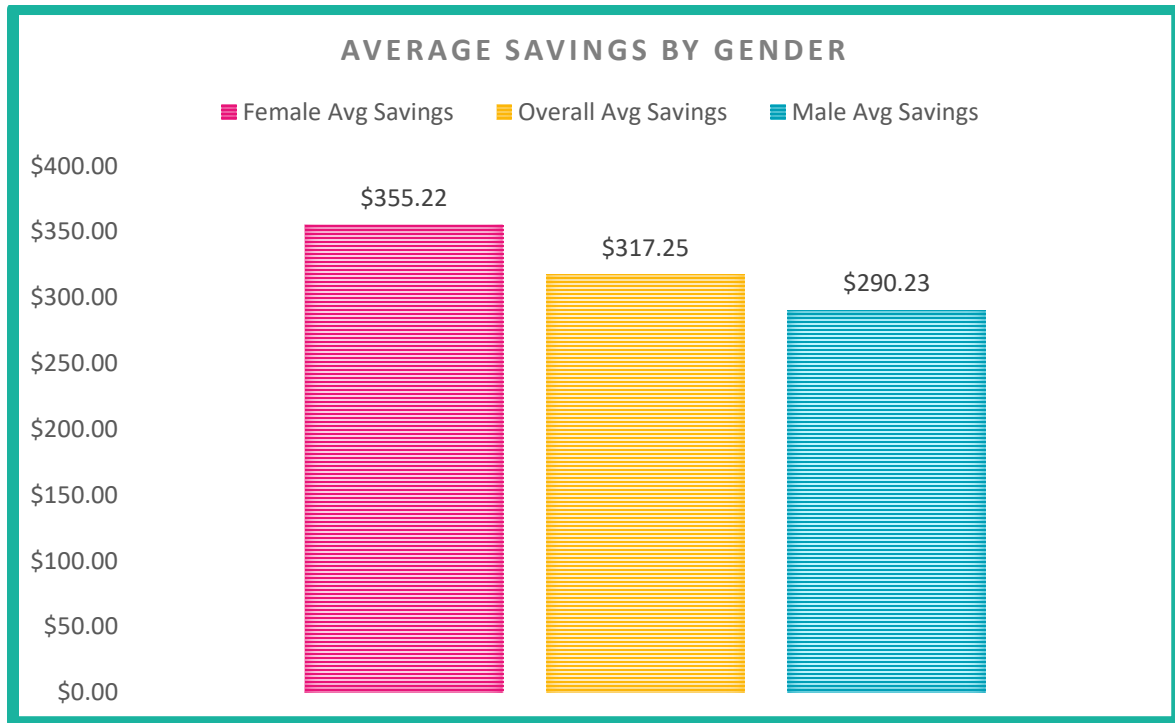
Appendix

Question 1:

Table 1:

	<i>Male Savings</i>	<i>Female Savings</i>
Mean	\$290.23	\$355.22
Standard Error	57.30	77.79
Median	143.00	136.00
Mode	366.00	#N/A
Standard Deviation	413.20	473.19
Sample Variance	170730.30	223910.12
Kurtosis	7.07	0.52
Skewness	2.43	1.33
Range	2032.00	1677.00
Minimum	-76.00	-74.00
Maximum	1956.00	1603.00
Sum	15092.00	13143.00
Count	52.00	37.00
Q1	34.75	28
Q3	368.75	560
IQR	334	532
Upper fence	869.75	1358
Lower fence	-466.25	-770
Outliers	Yes	Yes

Graph 1:



Graph 2:

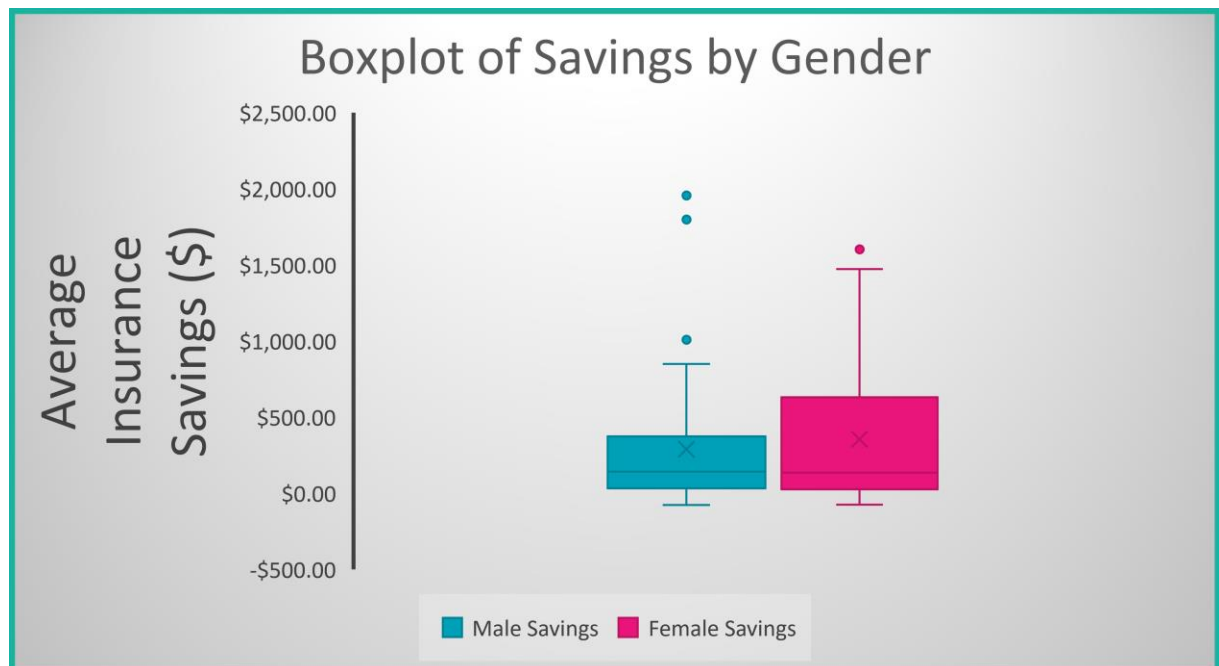


Table 2:

Hypothesis Test for $\mu_1 - \mu_2$ (independent, equal variances)			
Hypotheses			
Null Hypothesis	$\mu_1 - \mu_2$	\leq	0
Alternative Hypothesis	$\mu_1 - \mu_2$	$>$	0
Test Type			Upper
Level of significance			
			α 0.05
Critical Region			
Degrees of Freedom			87
Critical Value			1.6626
Sample Results			
Sample 1 Female			
Sample Standard Deviation			473.19
Sample Mean			355.22
Sample Size			37
Sample 2 Male			
Sample Standard Deviation			413.20
Sample Mean			290.23
Sample Size			52
Pooled Variance			192737.50
Standard Error of the Mean			94.4225
t Sample Statistic			0.6883
p-value			0.2466
Decision			
Fail to reject Null Hypothesis			

H0: Females aged under 30 save less than or equal to males on the insurance premiums
H1: Females aged under 30 saves more than males on the insurance premiums

Question 2:

Graph 3:

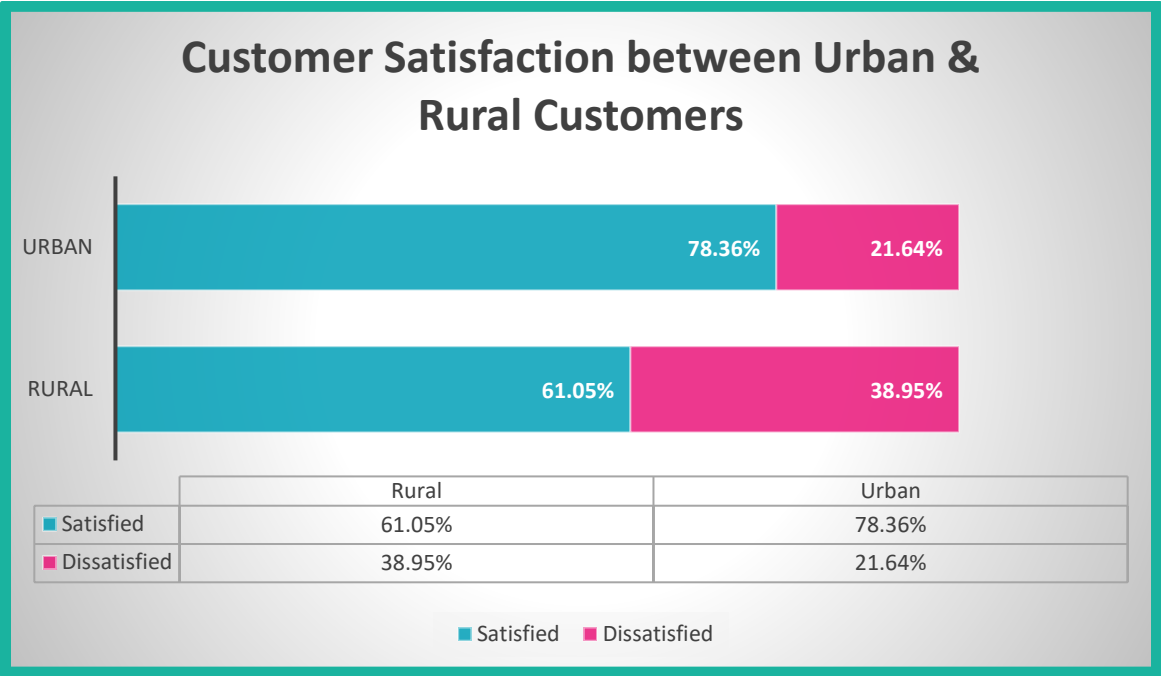


Table 3:

Hypothesis Test for $\pi_1 - \pi_2$			
Hypotheses			
Null Hypothesis	$\pi_1 - \pi_2$	\geq	0%
Alternative Hypothesis	$\pi_1 - \pi_2$	$<$	0%
Test Type			Lower
Level of significance			
		α	0.05
Critical Region			
Critical Value			-1.6449
Sample Data			
Sample 1 (Urban)			
Sample Size			305
Count of 'Successes'			66
Sample proportion, p_1			21.64%
Sample 2 (Rural)			
Sample Size			95
Count of 'Successes'			37
Sample proportion, p_2			38.95%
Pooled estimate of proportion			25.75%
Standard Error			5.14%
z Sample Statistic			-3.3689
p-value			0.0004
Decision			
Reject Null Hypothesis			
<p>Ho: The true Proportion of dissatisfied urban customers is equal or greater than the proportion of dissatisfied rural customers</p> <p>H1: The true Proportion of dissatisfied urban customers is less than the proportion of the dissatisfied rural customers</p>			

Confidence Interval for $\pi_1 - \pi_2$	
Level of Confidence	
Level of Confidence	95%
Sample Results	
Sample 1 Urban	
Sample Size	305
Count of 'Successes'	66
Sample proportion, p_1	21.64%
Sample 2 Rural	
Sample Size	95
Count of 'Successes'	37
Sample proportion, p_2	38.95%
Intermediate Calculations	
Pooled estimate of proportion	25.75%
Standard Error	5.14%
z value	1.9600
Confidence Interval for $\pi_1 - \pi_2$	
Interval Lower Limit	-27.38%
Interval Upper Limit	-7.24%

Question 3:

Graph 4:

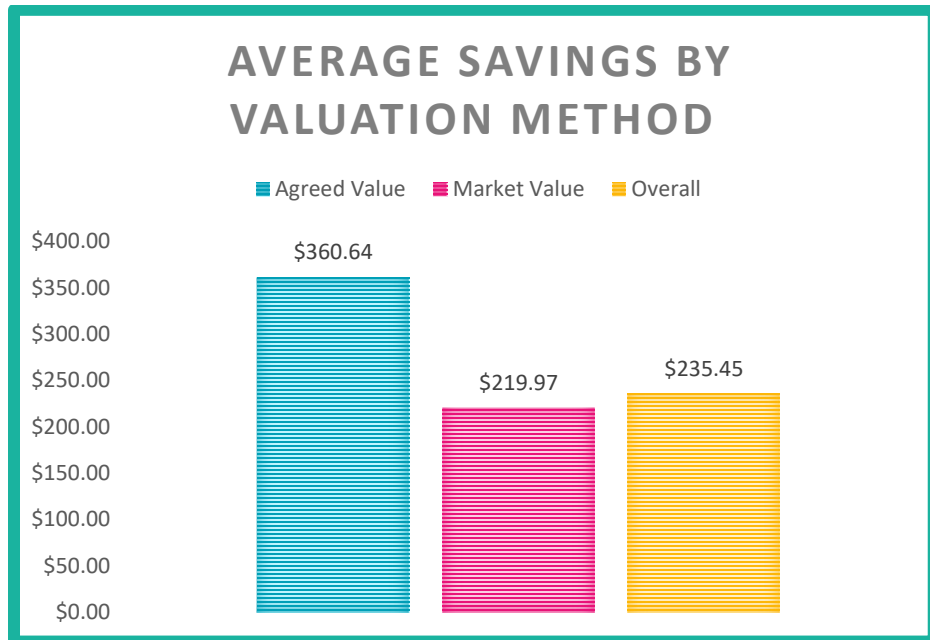


Table 4:

	Agreed Value	Market Value
Mean	\$360.64	\$219.97
Standard Error	63.10	16.89
Median	257.50	105.00
Mode	39.00	0.00
Standard Deviation	418.59	318.73
Sample Variance	175217.82	101590.79
Kurtosis	2.13	6.07
Skewness	1.53	2.20
Range	1727.00	2043.00
Minimum	-76.00	-87.00
Maximum	1651.00	1956.00
Sum	15868.00	78311.00
Count	44.00	356.00
Q1	69.25	0.00
Q3	494.75	342.25
IQR	425.50	342.25
Upper fence	1133.00	855.63
Lower fence	-569.00	-513.38
Outliers	Yes	

Graph 5:

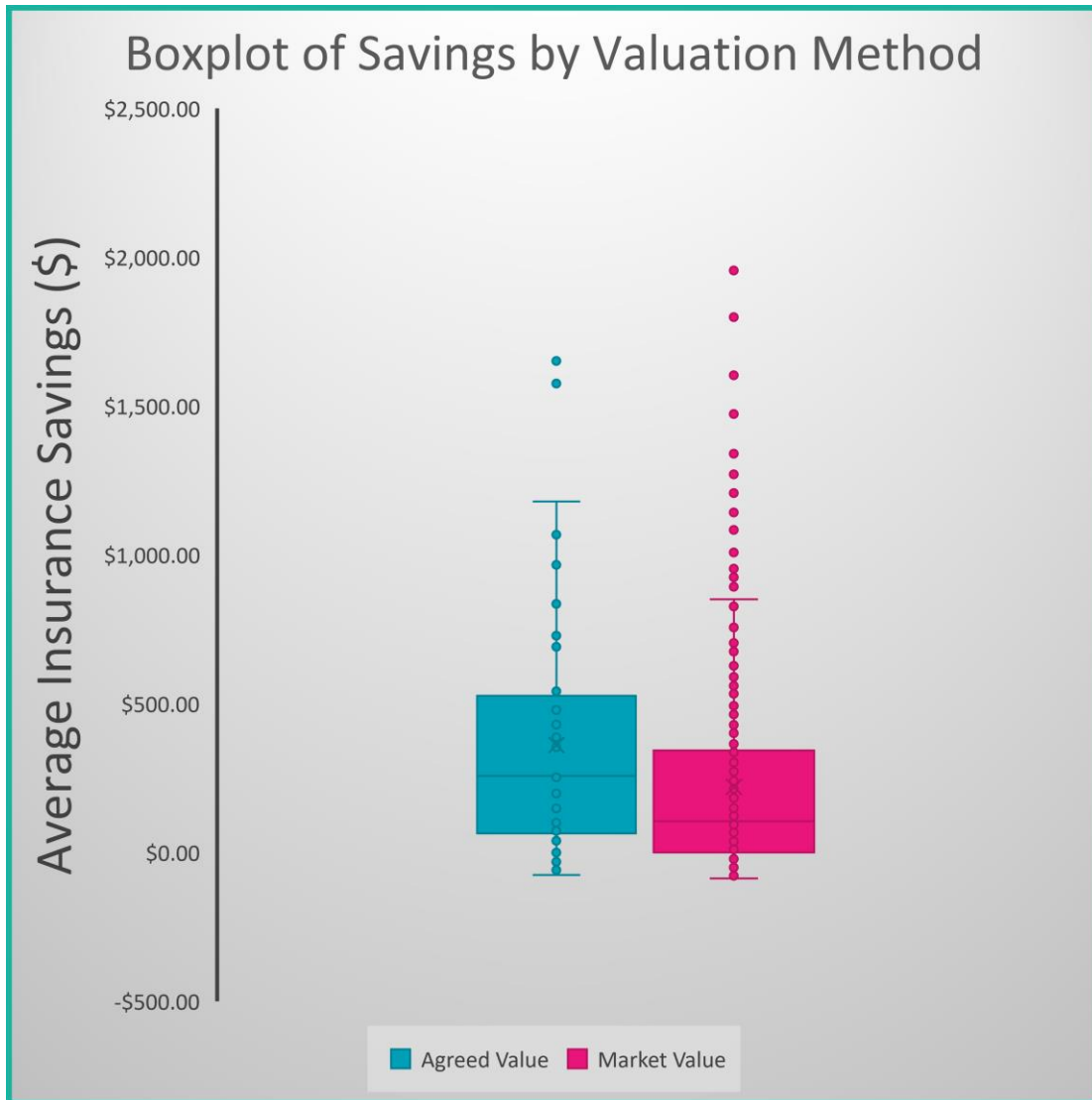


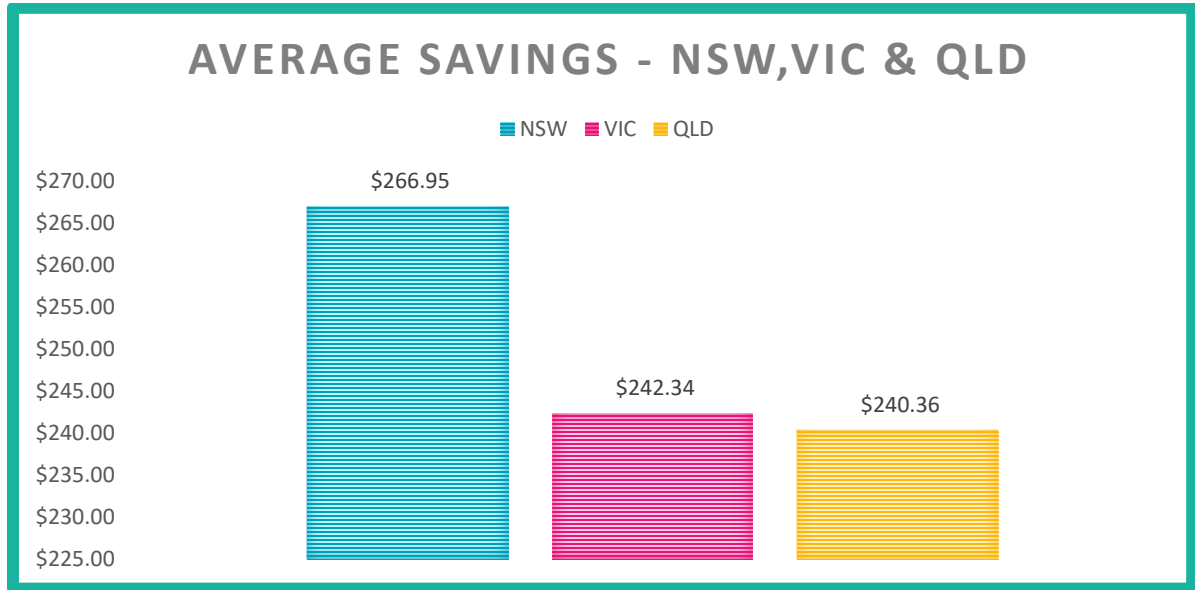
Table 5:

Hypothesis Test for $\mu_1 - \mu_2$ (independent, equal variances)			
Hypotheses			
Null Hypothesis	$\mu_1 - \mu_2$	=	0
Alternative Hypothesis	$\mu_1 - \mu_2$	\neq	0
Test Type			Two
Level of significance			
		α	0.05
Critical Region			
Degrees of Freedom			398
Lower Critical Value			-1.9659
Upper Critical Value			1.9659
Sample Results			
Sample 1 (Agreed Value)			
Sample Standard Deviation			418.59
Sample Mean			360.63
Sample Size			44
Sample 2 (Market Value)			
Sample Standard Deviation			318.73
Sample Mean			219.97
Sample Size			356
Pooled Variance			109545.41
Standard Error of the Mean			52.8903
t Sample Statistic			2.6594
p-value			0.0081
Decision			
Reject Null Hypothesis			
H0: The true average savings between both valuation methods is the same			
H1: The true average savings between both valuation methods is not the same			

Confidence Interval for $\mu_1 - \mu_2$ (independent, equal variances)	
Level of Confidence	
Level of Confidence	95%
Sample Results	
Sample 1 (Agreed Value)	
Sample Standard Deviation	418.5902748
Sample Mean	360.63
Sample Size	44
Sample 2 (Market Value)	
Sample Standard Deviation	318.733
Sample Mean	219.974
Sample Size	356
Intermediate Calculations	
Degrees of Freedom	398
Pooled Variance	109545.41
Standard Error of the Mean	52.8903
t value	1.9659
Confidence Interval for $\mu_1 - \mu_2$	
Interval Lower Limit	36.68
Interval Upper Limit	244.64

Question 4(A):

Graph 6:



Graph 7:

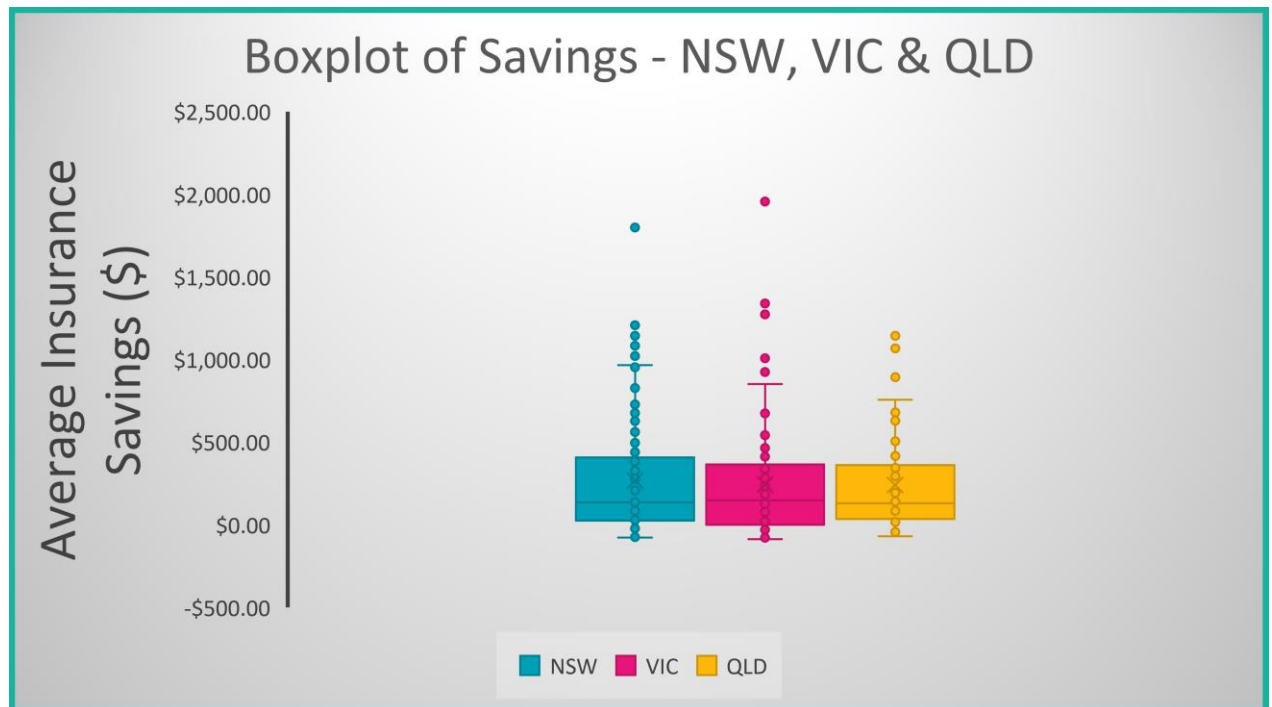


Table 6:

	NSW	Victoria	Queensland
Mean	\$ 266.95	\$ 242.34	\$ 240.36
Standard Error	30.04	35.98	37.81
Median	136.00	147.00	131.00
Mode	0.00	0.00	357.00
Standard Deviation	338.56	339.46	295.30
Sample Variance	114626.25	115231.04	87200.57
Kurtosis	3.11	8.05	1.29
Skewness	1.63	2.43	1.38
Range	1877.00	2043.00	1213.00
Minimum	-78.00	-87.00	-69.00
Maximum	1799.00	1956.00	1144.00
Sum	33903.00	21568.00	14662.00
Count	127.00	89.00	61.00
Q1	27.00	0.00	38.00
Q3	404.00	365.00	357.00
IQR	377.00	365.00	319.00
Upper fence	969.50	912.50	835.50
Lower fence	-538.50	-547.50	-440.50
Outliers	Yes		

Anova: Single Factor						
SUMMARY						
Groups	Count	Sum	Average	Variance		
NSW	127	33903.00	266.95	114626.25		
Victoria	89	21568.00	242.34	115231.04		
Queensland	61	14662.00	240.36	87200.57		
ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups/SSP	44578.57	2.00	22289.28	0.20	0.81	3.03
Within Groups/SSW	29815273.67	274.00	108814.87			
Total/SST	29859852.24	276.00				

Question 4(B):

Graph 8:

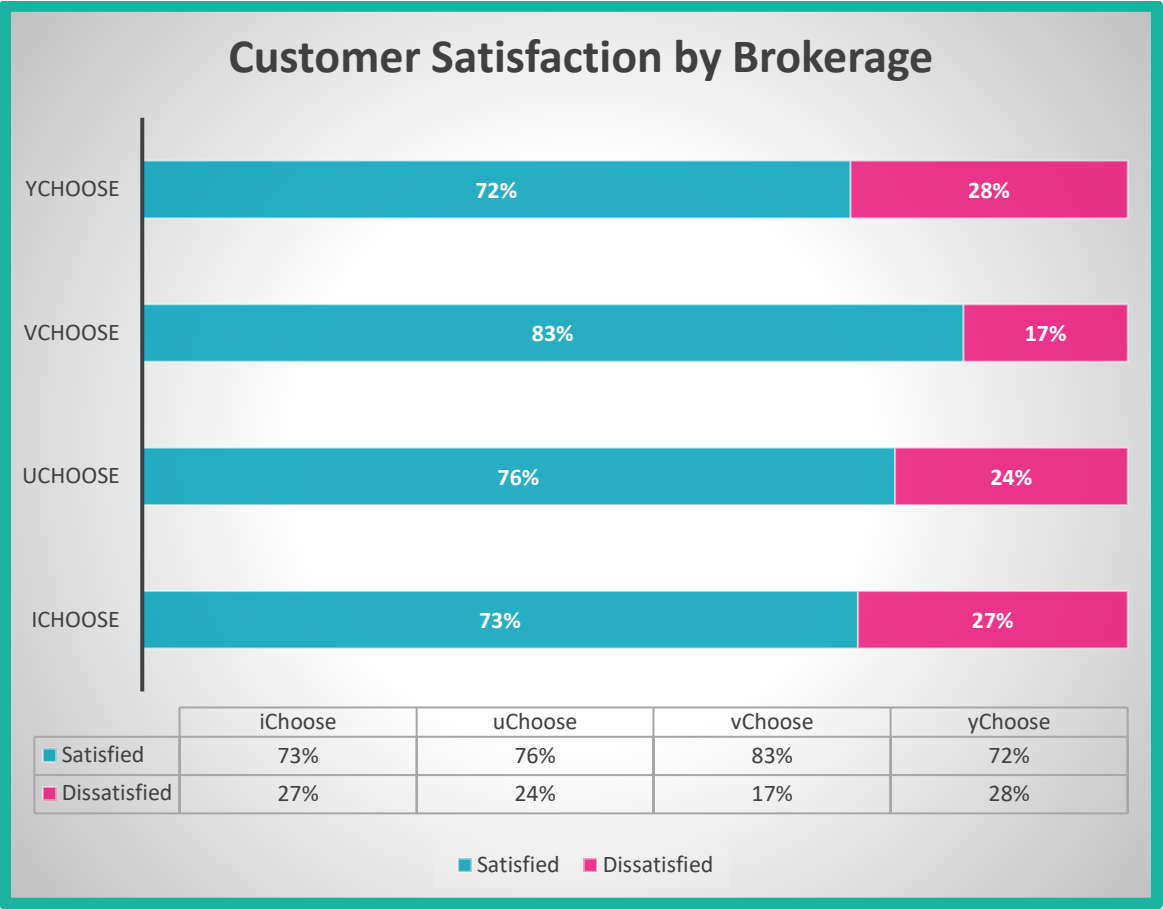


Table 7:

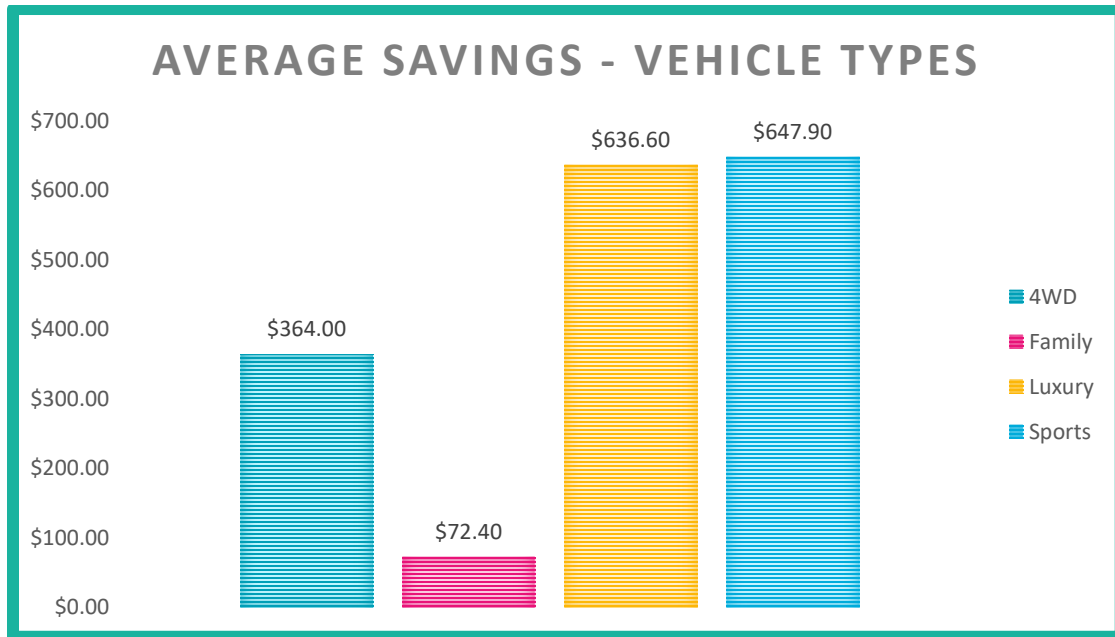
Chi-Square Test					
Observed Frequencies					
	Column variable				
Row variable	iChoose	uChoose	vChoose	yChoose	Total
Satisfied	138	55	35	69	297
Dissatisfied	52	17	7	27	103
Total	190	72	42	96	400
Expected Frequencies					
	Column variable				
Row variable	iChoose	uChoose	vChoose	yChoose	Total
Satisfied	141.0750	53.4600	31.1850	71.2800	297
Dissatisfied	48.9250	18.5400	10.8150	24.7200	103
Total	190	72	42	96	400
Data					
Level of Significance	0.05				
Number of Rows	2				
Number of Columns	4				
Degrees of Freedom	3				
Results					
Critical Value	7.8147				
Chi-Square Test Statistic	0.2603				
p-Value	96.73%				
Do not reject the null hypothesis					
Expected frequency assumption is met.					

Question 5:

Table 8:

Anova: Two-Factor With Replication						
SUMMARY	4WD	Family	Luxury	Sports	Total	
Agreed Value						
Count	5.00	5.00	5.00	5.00	20.00	
Sum	2865.00	482.00	3449.00	4418.00	11214.00	
Average	573.00	96.40	689.80	883.60	560.70	
Variance	126076.00	13171.30	51435.70	231554.30	177486.01	
Market Value						
Count	5.00	5.00	5.00	5.00	20.00	
Sum	775.00	242.00	2917.00	2061.00	5995.00	
Average	\$155.00	\$48.40	\$583.40	\$412.20	\$299.75	
Variance	8127.00	1798.80	173689.30	146398.20	116116.41	
Total						
Count	10.00	10.00	10.00	10.00		
Sum	3640.00	724.00	6366.00	6479.00		
Average	\$364.00	\$72.40	\$636.60	\$647.90		
Variance	108180.22	7293.38	103200.27	229706.10		
ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Sample	680949.03	1.00	680949.03	7.24	0.01	4.15
Columns	2223975.28	3.00	741325.09	7.88	0.00	2.90
Interaction	345468.28	3.00	115156.09	1.22	0.32	2.90
Within	3009002.40	32.00	94031.33			
Total	6259394.98	39.00				

Graph 9:



Graph 10:

