

Analysis of Smart phone data

Assignment-1

MIS771 – Descriptive Analytics and Visualisation

Shantanu Gupta

Student Id: 218200234

Introduction:

This report conduct a preliminary analysis of smartphone usage in Australia. In Australia there are close to twenty three million business buyers. For 2017, the number of smartphone users in Australia is estimated to have 16.69 million. These differences are important in understanding the behaviour of business services and consumer satisfaction and consumer markets. So, to delve deeper I have perform a series of descriptive and inferential analyses on the 150 randomly customer data and produce one report based on all the findings I got. This report is especially useful for the telecom companies, or those companies who wanted to step their telecom business in Australia. By reading this report, the companies can improve their operations, marketing strategy, customer churn etc., in this digital and competitive world. The report also includes on the user's expenditure, usage patterns, satisfaction levels and demographics.

This report submitted in fulfilment of the requirements for publishing the article in Business Insider in collaboration with BI Intelligence. BI Intelligence is a business insider paid research service. It was aimed to cater reports on key digital areas like mobile industry, specifically for the needs of big companies, investors and tech reader.

Body:

To: Sam Edmondson

From: Lee Slattery

Subject: Analysis of Smartphone data

Dear Sam,

Thank you for your e-mail. After properly reviewing your concerns I have analysed and came up with the following observations on your findings. Hope this fulfils your concerns.

Q1.

The average Australian spends in smartphone data of 150 customers participated in the survey is around \$68 in the monthly bill .If you want more precise figure then it is \$64. Therefore, 50% of the Australians have the bill of \$64 or less monthly bill and 50 % of the Australians have more than \$64 or more. I can say that the most common monthly bill is \$50. The minimum monthly bill is \$11 & the maximum monthly bill is \$216. The distribution is pointed in the middle. Average Australian monthly bill spending is \$64 as data contains high degree of asymmetry. The monthly bill difference between the highest bill and the lowest bill was

\$205. Furthermore, there was \$31.18 average spread away from the mean monthly bill of \$67.65. This indicates quite a bit of variation in the spending of monthly bill. More insights that I have found is 25% of the monthly bill is less than or equal to \$49 and on 75% have the monthly bill is greater than \$49. More deeply I have found that 75% of the monthly bill is less than or equal to \$84 and 25% of is greater than \$84. The majority of the data that will clustered around average value will lie between these regions \$36.42 & \$98.78. Most of the times we get the monthly bill within this interval. The middle 50% of our data is within \$50 to \$84.5. The lower 25% of the data is between \$85 and \$216, while upper 25% of the data is between \$11 to \$49.62. As an approximation when we select any person monthly bill we get the error of $\pm 2.5\%$. If we visualize data of monthly bill in the form of a graph then we find that there are most of the values are concentrated in the lower portion of the distribution. The tail on the right side is longer or fatter than the left side. The mass of the distribution is concentrated on the left of the figure. Distribution is shorter, tails are thinner than the normal distribution. The peak is lower, which means that data are light-tailed. There was two customer whose average monthly bill is higher than the rest of participants in the survey and thus, considered as an unusual case.

Q2.a.

We are 95 percent confident that the true average percentage use of smartphones for work-related activities is between 14.19% to 19.41% of the entire customer base. In other words, I have found that the percentage of all smart phone owners in Australia that use their phone as a work-related activities is 16.79% with the margin of error is $\pm 2.6057\%$. 95 percent of the time the data would match this results.

Q2.b.

We are 95 percent confident that the true proportion of all smartphone users who are classified as geo tribe 'Crusaders' is between 14.78% to 27.89% of the entire customer base. In other words, I have found that the proportion of all smart phone owners in Australia who are classified as geo tribe 'Crusaders' is 21.06% with the margin of error is $\pm 6.83\%$. 95 percent of the time the data would match this results.

I think no. of calls, SMS, and data allowance column should be included in the dataset to make the better profile of the user. This extra column will allows us to get more insight how different geo tribes people are using

smartphone services. For example, younger would prefer to spend more in data services than call services as comparison to old people because he wanted to surf internet more like Facebook, YouTube etc. But old people just use the services to call only their relatives.

Q3.a.

At 5 percent significance level, there is sufficient evidence to conclude that the average monthly bill of all Australian smart mobile phone user is less than the \$72. In other words, average monthly bill of Australian users is less than \$72. The industry report claim is correct.

Q3.b.

At 5 percent significance level, there is not enough evidence to conclude that 75% of all Australian smartphone customer are either 'Very Satisfied' or 'Moderately Satisfied' with their service provider. There would be a 7.32% chance of making an error where we think that more or less than (not equal to) 75% Australian smartphone customers are either 'Very Satisfied' or 'Moderately Satisfied' with their smartphone services. The last year similar study report is correct. This statement is valid.

Q3.c.

At 5 percent significance level, there is no sufficient evidence to conclude that there is a difference in the level of satisfaction between male and female customers about their satisfaction for smartphone service provider. In other words, both have the same level of satisfaction between male and female customer towards smartphone services. At 95% confidence, no conclusive results could be drawn regarding the true difference in proportion of the level of satisfaction between male and female customers. As the range contains both negative and positive values it is inconclusive. Unable to say whether one mean is different to the other. Because the interval estimate contains zero, you can conclude that there is insufficient evidence of a difference in the population means. There is insufficient evidence of a difference in the proportion of level of satisfaction between male and female users.

Q3.d.

Yes, the data support this proposition. At 5 percent significance level, there is sufficient evidence to conclude that there is a difference in the

average monthly bill amounts between the users who are using smartphone as a payment device and those users who are not using smartphone as a payment device. Thus, with 95% confidence, the mean difference in average monthly bill between the users who are using the smartphone as a payment device and those users who are not using smartphone as a payment device is between -\$0.83 and \$22.55. As the range contains both negative and positive values it is inconclusive. Unable to say whether one mean is different to the other. Because the interval estimate contains zero, I can conclude that there is a sufficient evidence of a difference in the population means.

Q3.e.

No, this is not a valid statement. At 5 percent significance level, there is not enough evidence to conclude that the use of Smartphones for online purchases is higher for users who are highly active in social media than for those users who are moderately active in social media.

Q4.a.

At 5 percent significance level, there is not enough evidence to conclude that the proportion of female users is higher than that of male users who are moderately active in social media. Based on these results, there is insufficient evidence to say that the proportion of female users is higher than that of male users who are moderately active in social media.

Q4.b.

At 5 percent significance level, there is not enough evidence to conclude that the average monthly bill amount for female users is higher than the male users in which both the groups are moderately active in social media. In other terms, the average monthly bill of male users is more than that of female users who are moderately active in social media.

Sam, I hope that this provides more information to your concerns and these observations will provide you a help to write up your article. Please let me know in case of any other concerns or any further information required from the data set. I will be happy to help you.

Sincerely

Lee Slattery

Conclusion:

Improvements in every area of the sector are needed if the company or new entrant in telecom sector is, in the first instance, to survive and then grow. The key areas of reform are the average monthly bill, using smartphone as a payment device and improve customer satisfaction with the telecom providers. Management must address these areas simultaneously if the company is to overcome and make a good profit.

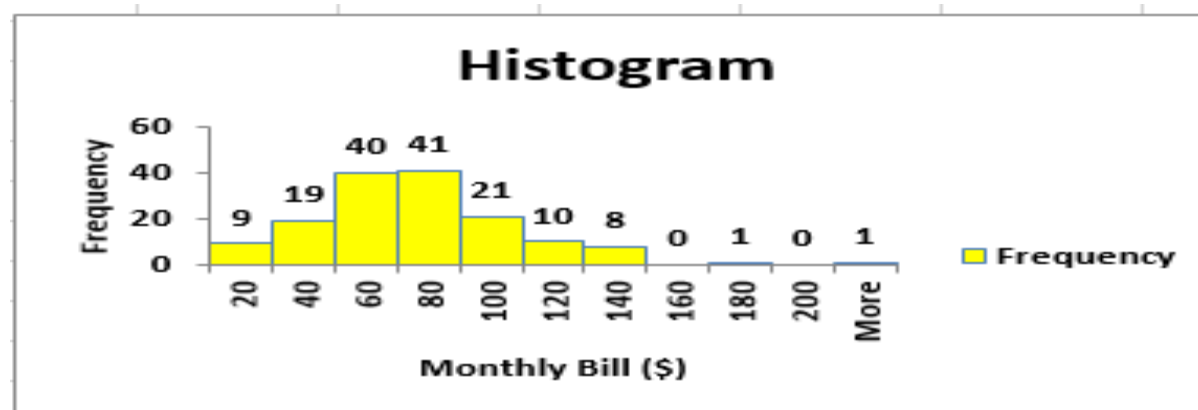
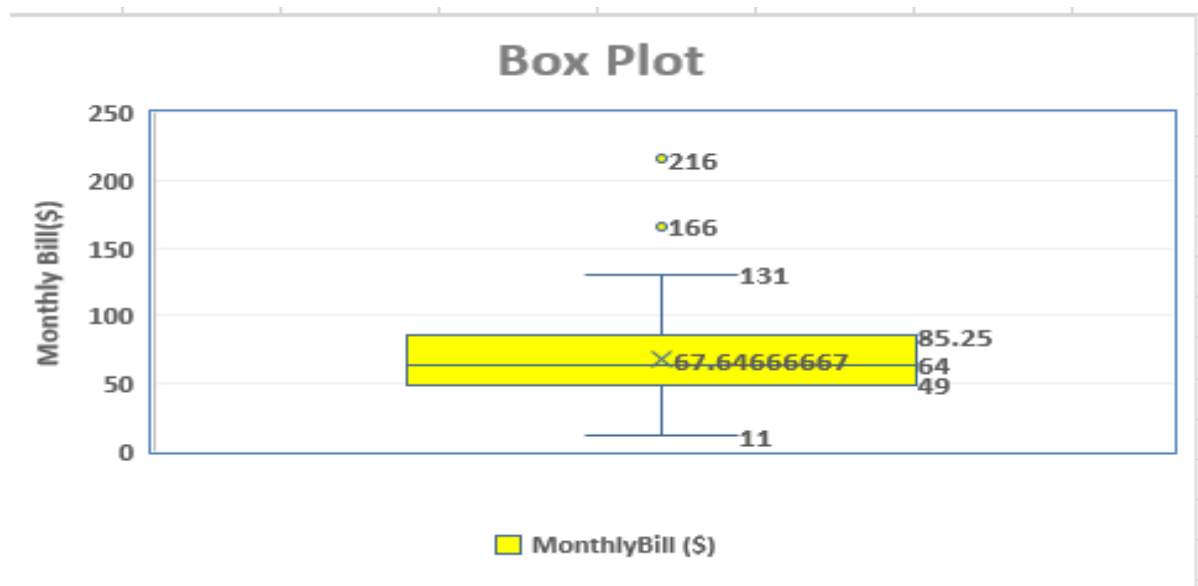
Limitations:

It must be remembered that this analysis is limited: a greater depth of understanding and evaluation can only occur if we have the bigger dataset and have more information about the user like no. of calls, SMS and data allowance. We are also look into other resources such as comparisons with other company's data, the government regulation and many more factor in telecom industries. Only after this process can a full appreciation of the company's current situation and possible future occur.

Appendices:

Q1.

MonthlyBill (\$)	
Mean	67.65
Standard Error	2.55
Median	64.00
Mode	50.00
Standard Deviation	31.18
Sample Variance	972.46
Kurtosis	2.82
Skewness	1.05
Range	205.00
Minimum	11.00
Maximum	216.00
Sum	10147.00
Count	150.00
Q1	49.25
Q3	84.50
IQR	35.25
LF	-3.63
UF	137.38
Potential Outliers	Yes

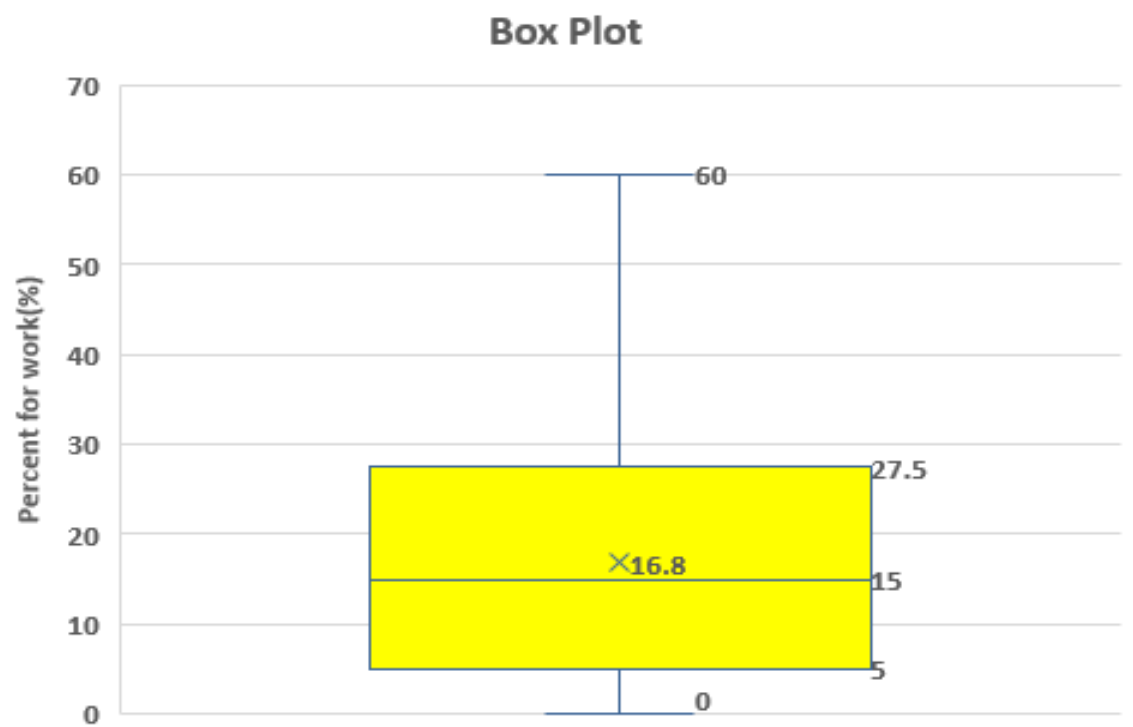
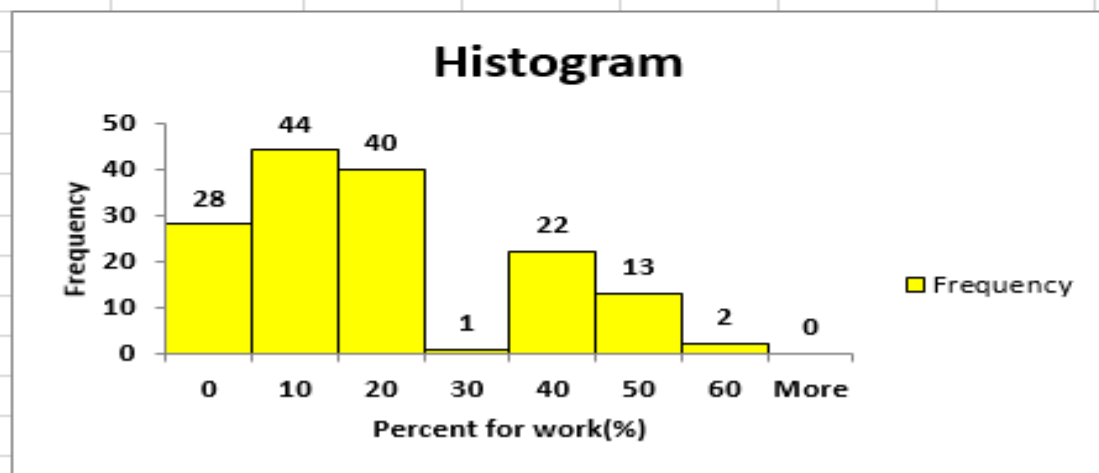


Q2.a

PercentForWork%	
Mean	16.80
Standard Error	1.32
Median	15.00
Mode	5.00
Standard Deviation	16.15
Sample Variance	260.83
Kurtosis	-0.22
Skewness	0.93
Range	60.00
Minimum	0.00
Maximum	60.00
Sum	2520.00
Count	150.00
Q1	5.00
Q3	23.75
IQR	18.75
LF	-23.13
UF	51.88
Potential Outliers	Yes

Bin	Frequency
0	28
10	44
20	40
30	1
40	22
50	13
60	2
More	0

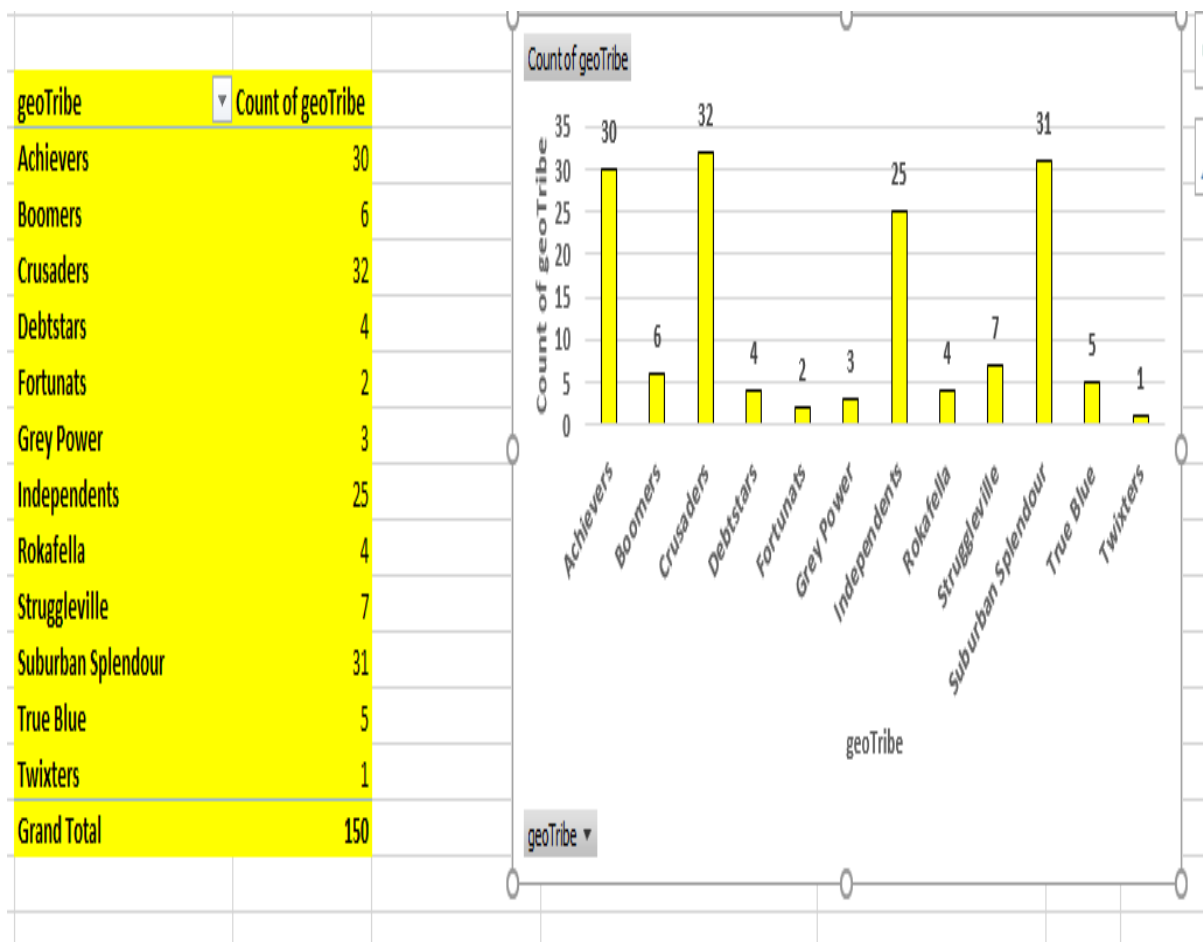
Bin
0
10
20
30
40
50
60



Confidence Interval for mean	
Data	
Sample Standard Deviation	16.15
Sample Mean	16.8
Sample Size	150
Confidence Level	95%
Intermediate Calculations	
Standard Error of the Mean	1.3186
Degrees of Freedom	149
t Value	1.9760
Margin of Error	2.6057
Confidence Interval	
Interval Lower Limit	14.19
Interval Upper Limit	19.41

Q2.b

Confidence Interval for proportion	
Data	
Sample Size	150
Count of Successes	32
Confidence Level	95%
Intermediate Calculations	
Sample Proportion	0.213333333
Z Value	1.9600
Standard Error of the Proportion	0.033448689
Margin of Error	0.0656
Confidence Interval	
Interval Lower Limit	14.78%
Interval Upper Limit	27.89%

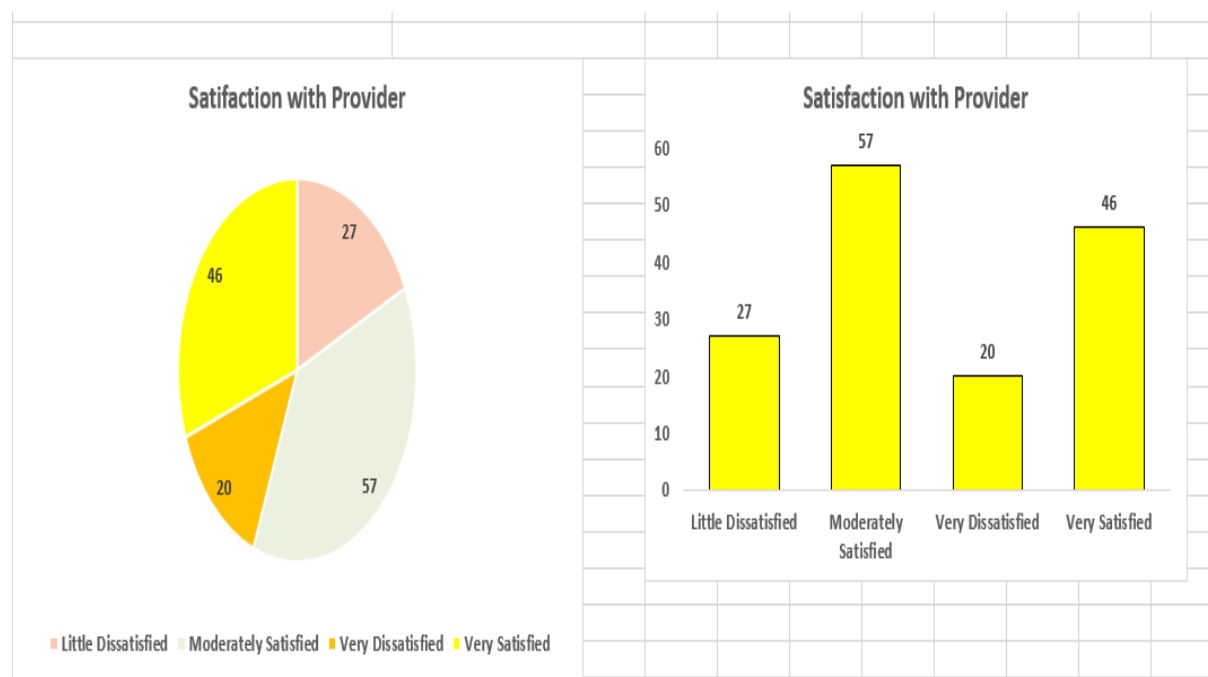


Q3.a

MonthlyBill (\$)	
Mean	67.65
Standard Error	2.55
Median	64.00
Mode	50.00
Standard Deviation	31.18
Sample Variance	972.46
Kurtosis	2.82
Skewness	1.05
Range	205.00
Minimum	11.00
Maximum	216.00
Sum	10147.00
Count	150.00
Q1	49.25
Q3	84.50
IQR	35.25
LF	-3.63
UF	137.38
Potential Outliers	Yes

Hypothesis Test for μ			
Hypotheses			
Null Hypothesis	μ	\geq	72
Alternative Hypothesis	μ	$<$	72
Test Type			Lower
Level of significance			
		α	0.05
Critical Region			
Degrees of Freedom			149
Critical Value			-1.6551
Sample Data			
Sample Standard Deviation			31.18
Sample Mean			67.65
Sample Size			150
Standard Error of the Mean			2.5458
t Sample Statistic			-1.7087
p-value			0.0448
Decision			
Reject Null Hypothesis			

Q3.b



Hypothesis Test for π			
Hypotheses			
Null Hypot	π	=	75%
Alternative	π	\neq	75%
Test Type			Two
Level of significance			
		α	0.05
Critical Region			
Lower Critical Value			-1.9600
Upper Critical Value			1.9600
Sample Data			
Sample Size			150
Count of 'Successes'			103
Sample proportion, p			68.67%
Standard Error			3.54%
z Sample Statistic			-1.7913
p-value			0.0732
Decision			
Fail to reject Null Hypothesis			

SatisfactionWithProvider	Count of SatisfactionWithProvider
Little Dissatisfied	27
Moderately Satisfied	57
Very Dissatisfied	20
Very Satisfied	46
Grand Total	150

Q3.c

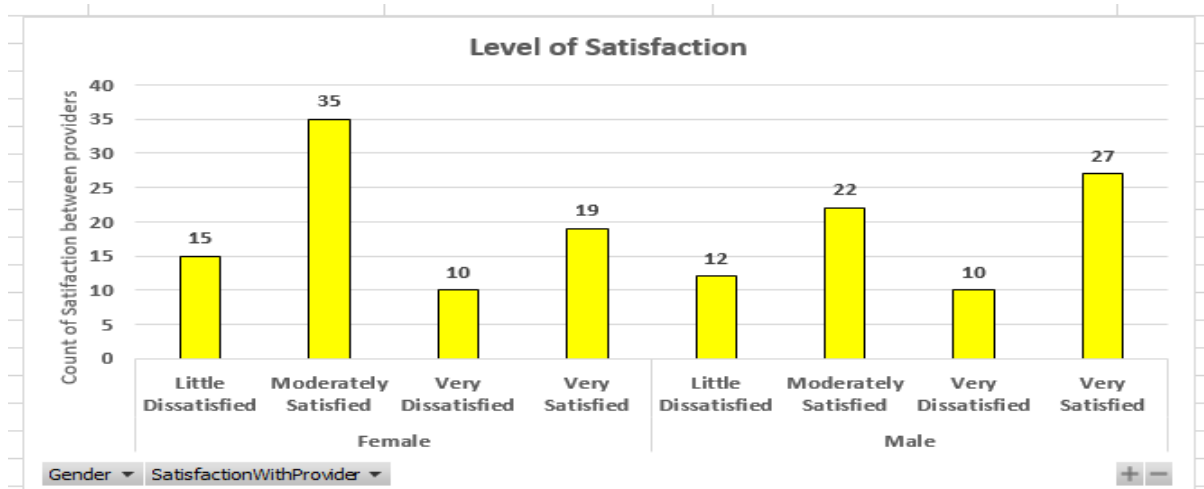
Gender	Count of Gender
Female	79
Male	71
Grand Total	150

SatisfactionWithProvider	Gender	Count of SatisfactionWithProvider
Little Dissatisfied	Female	15
	Male	12
Little Dissatisfied Total		27
Moderately Satisfied	Female	35
	Male	22
Moderately Satisfied Total		57
Very Dissatisfied	Female	10
	Male	10
Very Dissatisfied Total		20
Very Satisfied	Female	19
	Male	27
Very Satisfied Total		46
Grand Total		150

Count of Successes	Value
Male	49
Female	54

Hypothesis Test for $\pi_1 - \pi_2$			
Hypotheses			
Null Hypothesis	$\pi_1 - \pi_2$	=	0%
Alternative Hypothesis	$\pi_1 - \pi_2$	<>	0%
Test Type			Two
Level of significance			
		α	0.05
Critical Region			
Lower Critical Value			-1.9600
Upper Critical Value			1.9600
Sample Data			
Sample 1 Data(Level of Satisfaction-Male)			
Sample Size			71
Count of 'Successes'			49
Sample proportion, p_1			69.01%
Sample 2 Data(Level of Satisfaction-Female)			
Sample Size			79
Count of 'Successes'			54
Sample proportion, p_2			68.35%
Pooled estimate of proportion			68.67%
Standard Error			7.59%
z Sample Statistic			0.0870
p-value			0.9307
Decision			
Fail to reject Null Hypothesis			

Confidence Interval for $\pi_1 - \pi_2$	
Level of Confidence	
Level of Confidence	95%
Sample Results	
Sample 1 Data(Level of Satisfaction-Male)	
Sample Size	71
Count of 'Successes'	49
Sample proportion, p_1	69.01%
Sample 2 Data(Level of Satisfaction-Female)	
Sample Size	79
Count of 'Successes'	54
Sample proportion, p_2	68.35%
Intermediate Calculations	
Pooled estimate of proportion	68.67%
Standard Error	7.59%
z value	1.9600
Confidence Interval for $\pi_1 - \pi_2$	
Interval Lower Limit	-14.21%
Interval Upper Limit	15.53%



Gender	SatisfactionWithProvider	Count of SatisfactionWithProvider
Female	Little Dissatisfied	15
	Moderately Satisfied	35
	Very Dissatisfied	10
	Very Satisfied	19
Female Total		79
Male	Little Dissatisfied	12
	Moderately Satisfied	22
	Very Dissatisfied	10
	Very Satisfied	27
Male Total		71
Grand Total		150

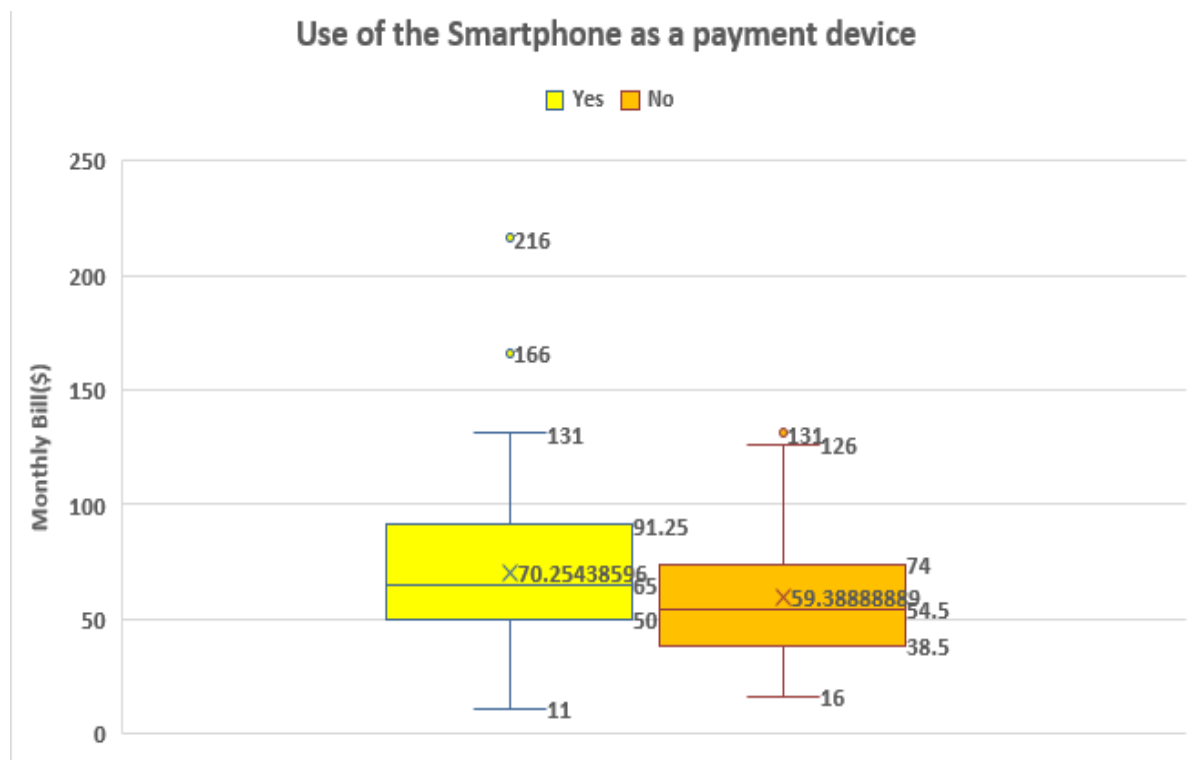
Q3.d

UsedForPayment	Average of MonthlyBill (\$)
No	59.39
Yes	70.25
Grand Total	67.65
UsedForPayment	Count of UsedForPayment
No	36
Yes	114
Grand Total	150
UsedForPayment	StdDev of MonthlyBill (\$)
No	29.36
Yes	31.41
Grand Total	31.18

Hypothesis Test for $\mu_1 - \mu_2$ (independent, equal variances)			
Hypotheses			
Null Hypot	$\mu_1 - \mu_2$	\leq	0
Alternative	$\mu_1 - \mu_2$	$>$	0
Test Type			Upper
Level of significance			
		α	0.05
Critical Region			
Degrees of Freedom			148
Critical Value			1.6552
Sample Results			
Sample 1 Data(Yes)			
Sample Standard Deviation			31.41
Sample Mean			70.25
Sample Size			114
Sample 2 Data(No)			
Sample Standard Deviation			29.36
Sample Mean			59.39
Sample Size			36
Pooled Variance			957.13
Standard Error of the Mean			5.9146
t Sample Statistic			1.8361
p-value			0.0342
Decision			
Reject Null Hypothesis			

Confidence Interval for $\mu_1 - \mu_2$ (independent, equal variances)	
Level of Confidence	
Level of Confidence	95%
Sample Results	
Sample 1 Data(Yes)	
Sample Standard Deviation	31.41
Sample Mean	70.25
Sample Size	114
Sample 2 Data(No)	
Sample Standard Deviation	29.36
Sample Mean	59.39
Sample Size	36
Intermediate Calculations	
Degrees of Freedom	148
Pooled Variance	957.13
Standard Error of the Mean	5.9146
t value	1.9761
Confidence Interval for $\mu_1 - \mu_2$	
Interval Lower Limit	-0.83
Interval Upper Limit	22.55

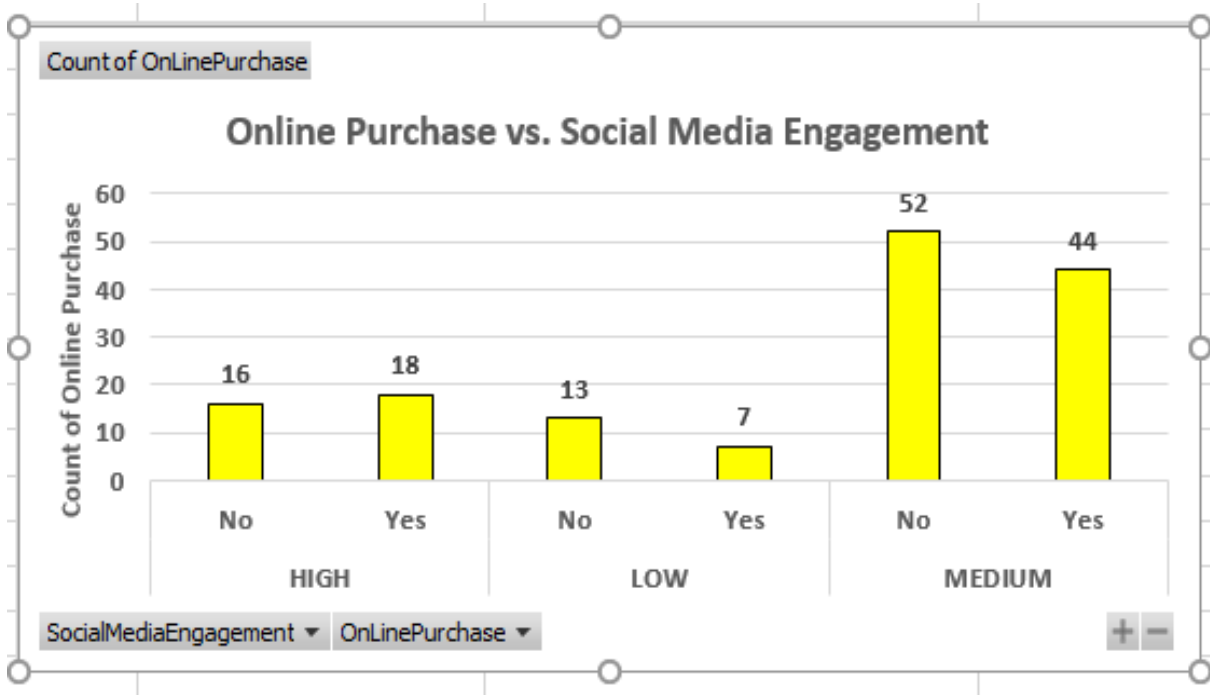
Used of Smartphone as a paymen	Yes	No
Mean	70.25	59.39
Standard Error	2.94	4.89
Median	65.00	54.50
Mode	50.00	40.00
Standard Deviation	31.41	29.36
Sample Variance	986.72	861.90
Kurtosis	3.44	0.48
Skewness	1.12	0.87
Range	205.00	115.00
Minimum	11.00	16.00
Maximum	216.00	131.00
Sum	8009.00	2138.00
Count	114.00	36.00
Q1	50.00	39.50
Q3	89.50	74.00
IQR	39.50	34.50
LF	-9.25	-12.25
UF	148.75	125.75
Potential Outliers	Yes	Yes



Q3.e

SocialMediaEngagement	OnLinePurchase	Count of OnLinePurchase
HIGH	No	16
	Yes	18
HIGH Total		34
LOW	No	13
	Yes	7
LOW Total		20
MEDIUM	No	52
	Yes	44
MEDIUM Total		96
Grand Total		150
SocialMediaEngagement	Count of SocialMediaEngagement	
HIGH	34	
LOW	20	
MEDIUM	96	
Grand Total	150	

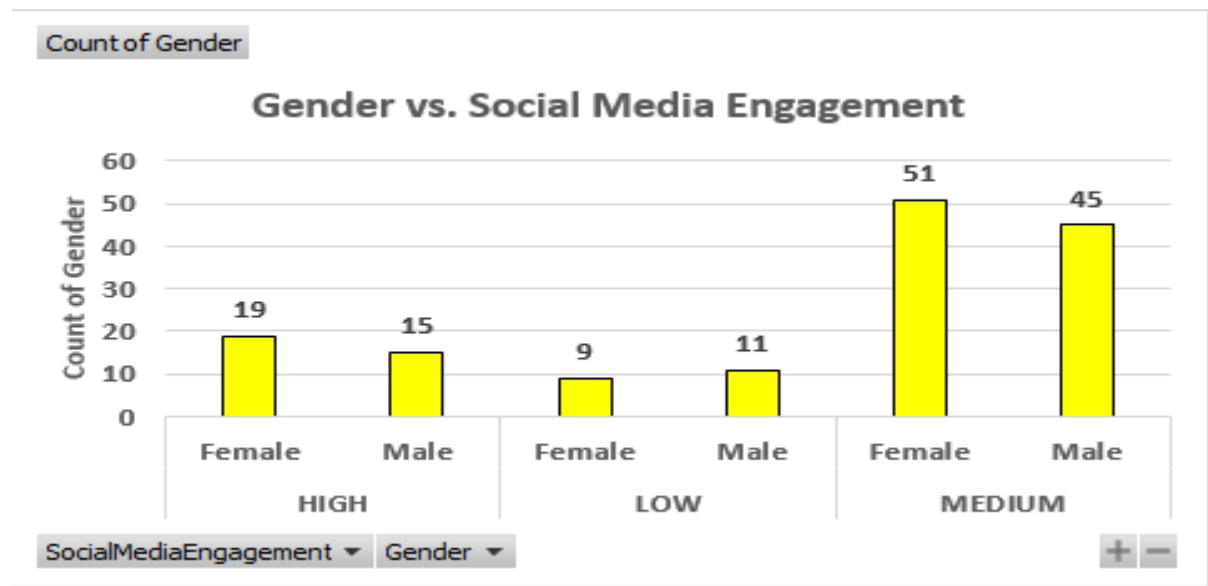
Hypothesis Test for $\pi_1 - \pi_2$			
Hypotheses			
Null Hypothesis	$\pi_1 - \pi_2$	\leq	0%
Alternative Hypothesis	$\pi_1 - \pi_2$	$>$	0%
Test Type			Upper
Level of significance			
		α	0.05
Critical Region			
Critical Value			1.6449
Sample Data			
Sample 1 Data(Highly Active)			
Sample Size			34
Count of 'Successes'			18
Sample proportion, p_1			52.94%
Sample 2 Data(Moderately Active)			
Sample Size			96
Count of 'Successes'			44
Sample proportion, p_2			45.83%
Pooled estimate of proportion			47.69%
Standard Error			9.97%
z Sample Statistic			0.7131
p-value			0.2379
Decision			
Fail to reject Null Hypothesis			



Q4.a

SocialMedia	Gender	Count of Gender
HIGH	Female	19
	Male	15
HIGH Total		34
LOW	Female	9
	Male	11
LOW Total		20
MEDIUM	Female	51
	Male	45
MEDIUM Total		96
Grand Total		150
Gender	Count of Gender	
Female	79	
Male	71	
Grand Total		150

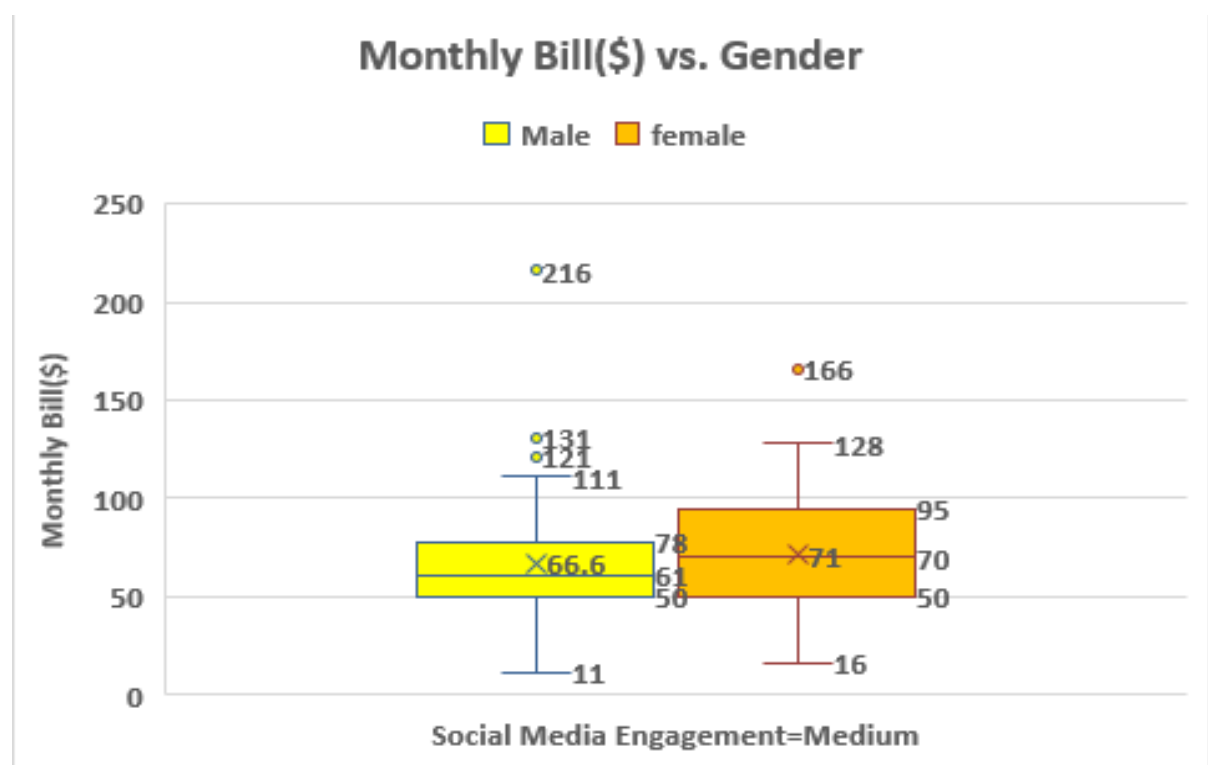
Hypothesis Test for $\pi_1 - \pi_2$			
Hypotheses			
Null Hypot	$\pi_1 - \pi_2$	\leq	0%
Alternative	$\pi_1 - \pi_2$	$>$	0%
Test Type			Upper
Level of significance			
		α	0.05
Critical Region			
Critical Value			1.6449
Sample Data			
Sample 1 Data(Moderately active in Social Media-Female)			
Sample Size			79
Count of 'Successes'			51
Sample proportion, p_1			64.56%
Sample 2 Data(Moderately active in Social Media-Male)			
Sample Size			71
Count of 'Successes'			45
Sample proportion, p_2			63.38%
Pooled estimate of proportion			64.00%
Standard Error			7.85%
z Sample Statistic			0.1499
p-value			0.4404
Decision			
Fail to reject Null Hypothesis			



Q4.b

MonthlyBill (\$)-Female	
Mean	71.00
Standard Error	4.40
Median	70.00
Mode	78.00
Standard Deviation	31.40
Sample Variance	985.96
Kurtosis	0.41
Skewness	0.42
Range	150.00
Minimum	16.00
Maximum	166.00
Sum	3621.00
Count	51.00
Q1	52
Q3	95
IQR	43
LF	-12.5
UF	159.5
Potential Outliers	Yes

MonthlyBill (\$) - Male	
Mean	66.60
Standard Error	4.96
Median	61.00
Mode	50.00
Standard Deviation	33.28
Sample Variance	1107.38
Kurtosis	8.49
Skewness	2.24
Range	205.00
Minimum	11.00
Maximum	216.00
Sum	2997.00
Count	45.00
Q1	50
Q3	77
IQR	27
LF	9.5
UF	117.5
Potential Outliers	Yes



SocialMedia	Gender	Count of Gender
HIGH	Female	19
	Male	15
HIGH Total		34
LOW	Female	9
	Male	11
LOW Total		20
MEDIUM	Female	51
	Male	45
MEDIUM Total		96
Grand Total		150
Gender	Count of Gender	
Female	79	
Male	71	
Grand Total	150	

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		94	
Critical Value		1.6612	
Sample Results			
Sample 1 Data(Female)			
Sample Standard Deviation		31.40	
Sample Mean		71.00	
Sample Size		51	
Sample 2 Data(Male)			
Sample Standard Deviation		33.28	
Sample Mean		66.60	
Sample Size		45	
Pooled Variance		1042.88	
Standard Error of the Mean		6.6048	
t Sample Statistic		0.6662	
p-value		0.2535	
Decision			
Fail to reject Null Hypothesis			