



MIS771 – DESCRIPTIVE ANALYTICS AND VISUALIZATION
Trimester 1, 2020

TASSPAPERMILL (TPM) RESEARCH PROJECT.

Assignment 2 - Individual

SUBMITTED BY: Ruth Audu
STUDENT ID: 218031013
EMAIL ID: raudu@deakin.edu.au

DUE DATE AND TIME: 17th MAY 2020
UNIT CHAIR: Dr Dilal Saundage

Table of Contents

TASSPAPERMILL (TPM) RESEARCH PROJECT.	1
1. INTRODUCTION	3
2. ANALYSIS	3
2.1 EXPLORING KEY DEPENDENT VARIABLES	3
2.2. MODELLING QUANTITY ORDERED FROM TPM.	4
2.2.1. Identified factors Influencing Quantity of purchased products.	4
2.2.2. Model Building.	4
2.3. THE INTERACTION EFFECTS OF BRAND IMAGE PERCEPTION ON PRODUCT QUALITY AND ORDER QUANTITY	5
2.3.1. The Interaction Effect	5
2.3.2. Managerial Recommendations	6
2.4. MODELLING THE LIKELIHOOD OF CUSTOMER SIGNING A CONTRACT WITH TPM	6
2.4.1. Model Classifications	6
2.4.2. Establishing Practical Significance.	6
2.4.3. Establishing Statistical Significance	6
2.4.4. The percentage of change in odds	7
2.4.5. The Interpreted and Predicted Probabilities	7
2.5. FORECASTING TURNOVER FOR NEXT FOUR QUARTERS.	8
3. CONCLUSION	8
4. APPENDICES	9
Appendix a	9
Appendix b	9
Appendix c	10
Appendix d	10
Appendix e	10
Appendix f	11
Appendix g	11
Appendix h	11
Appendix i	12
Appendix j	12
Appendix k	12
Appendix l	13

1. INTRODUCTION

This is a technical report for TassPaperMill, a company that sells paper products to newspaper and magazine market segments. Due to the advent of social media and increase in online activity, there has been a shift in customer preferences affecting sales, leading to a downturn in the company's financial performance. Using data gotten from TPM's data warehouse, decision support system and online customer survey, the objective of this study is to analyse 200 observations of this data ,create models, gather findings and draw conclusions to help the company better understand its client's characteristics, perceptions and preferences and offer recommendations based on findings. This would give TPM an edge in serving their customer's better; increasing sales, profitability and helping them gain competitive advantage. The use and application of Multiple linear regression, logistic regression, time series and forecasting aided the analysis.

Subsequent sections in the report highlights overall research and findings from the exploration of key data variables, possible factors that influences customer purchase quantity of TPM's product and likelihood of a customer to sign a contract with the company as well as the forecasting result for TPM's next four quarters.

2. ANALYSIS

2.1 EXPLORING KEY DEPENDENT VARIABLES

The key dependent variables for the analysis is the order quantity and contract. Using relevant charts and graphs, the exploration of these individual variable, gives summarized information about their key features

- ORDER QUANTITY

The order quantity refers to quantity of products purchased by TPM's customers. As a numerical variable, summary measure and histogram was used to describe the variable (refer to appendix a). These are the findings:

- 1533 is the total sum of the products purchased by the 200 TPM's customers.
- The average number of purchased products made by these customers is 7.67, having the minimum and maximum number of purchased products amongst them to be 4.3 and 9.9 respectively.
- Most of the customers ordered approximately 7 products.
- 25% and 75% of the quantity purchased by the customers are approximately less than or equal to 7 and 8 percent respectively.

- CONTRACT

The frequency distribution table, pie chart and bar graph (refer to appendix b) used to analyse this categorical variable revealed that:

- Amongst the 200 observations, 101 of the customers signed a contract with TPM, nominating the company as their preferred supplier while the remaining 99 signed no contract.
- Customers who signed a contract with TPM as their nominated supplier are just 1% higher than those who did not choose TPM to be their supplier.

2.2. MODELLING QUANTITY ORDERED FROM TPM.

There were strict steps followed to model the number of purchased products of the 200 observations.

2.2.1. Identified factors Influencing Quantity of purchased products

Scatter plot diagrams were generated to determine and better understand the relationship of the independent variables with the dependent variable (order quantity) (refer to appendix c). All independent variables appeared to have a positive linear relationship except TPM's customer type, their location (region), the extent to which customers believe the company offers competitive pricing (competitive pricing) and the perceived willingness of TPM to negotiate product prices (price flexibility). The use of correlation matrix also helped determine the magnitude of the linear relationships and checked for multi-collinearity. The matrix revealed that contract and shipping costs of the products have positive moderate linear relationships with the dependent variable, with other independent variables having positive weak linear relationships ($< .4$) excluding customer type, region, competitive pricing, price flexibility which have negative weak linear relationships (refer to appendix d).

Using 0.80 cut-off, potential multi-collinearity problem was identified between the delivery cost (shipping cost) and the perception of time taken to deliver products after order confirmation (shipping speed). The shipping speed variable was dropped due to its weaker correlation with the dependent variable.

At the end of the checks and explorations, the possible factors identified to influence the order quantity was Loyalty, Distribution Channel, Quality, Product line, brand image, order fulfillment, shipping speed, advert, Customer type, region, competitive pricing, price flexibility, contract, shipping cost and social media presence. These variables were used to build the first model.

2.2.2. Model Building.

Using multiple linear regression, in the first model built, approximately, 56% of variation in the quantity ordered can be explained by the regression model and the remaining 44% of the variation would be explained by other factors not included in the model (refer to appendix e). Although, the model has some predictive power, not all the independent variables (customer type, region, distribution channel, social media presence, advert, product line, competitive pricing, order fulfillment and price flexibility) were statistically significant. Multiple iterations were undertaken to achieve a final model having all statistically significant independent variables.

The Final Regression Model

Using multiple linear regression, overall, the model has some predictive power. Approximately, 54% of variation in the quantity ordered can be explained by the regression model and the remaining 46% of the variation would be explained by other factors not included in the model (refer to appendix f). The independent variables statistically significant to the model are the length of customer purchase time (Loyalty), the perceived level of product quality (product quality), perceived overall brand image (brand image), perception of delivery costs (shipping cost) and the signed contract nominating TPM as the preferred supplier and(contract).

The model predicted that:

- For a new TPM customer with no level of quality, brand image and delivery costs perceptions as well as no signed contract, the average quantity of products ordered is expected to be 3.53
- All other predictors held constant, for each unit of increase in the length of customer purchase time, the quantity of orders by a customer increases by .05 units on average.

- Holding all other variables constant, order quantity increases by an average of .21 units for every unit increase in the perceived level of product quality.
- Holding all other variables constant, order quantity increases by an average of .13 units for every unit increase in the perceived overall brand image.
- All other predictors held constant, for each unit of increase in the perception of delivery costs, the quantity of orders by a customer increases by .22 units on average.
- All other predictors held constant, for each unit of increase in signed contracts, the quantity of orders by a customer increases by .34 units on average.

The Residual Analysis Assumptions

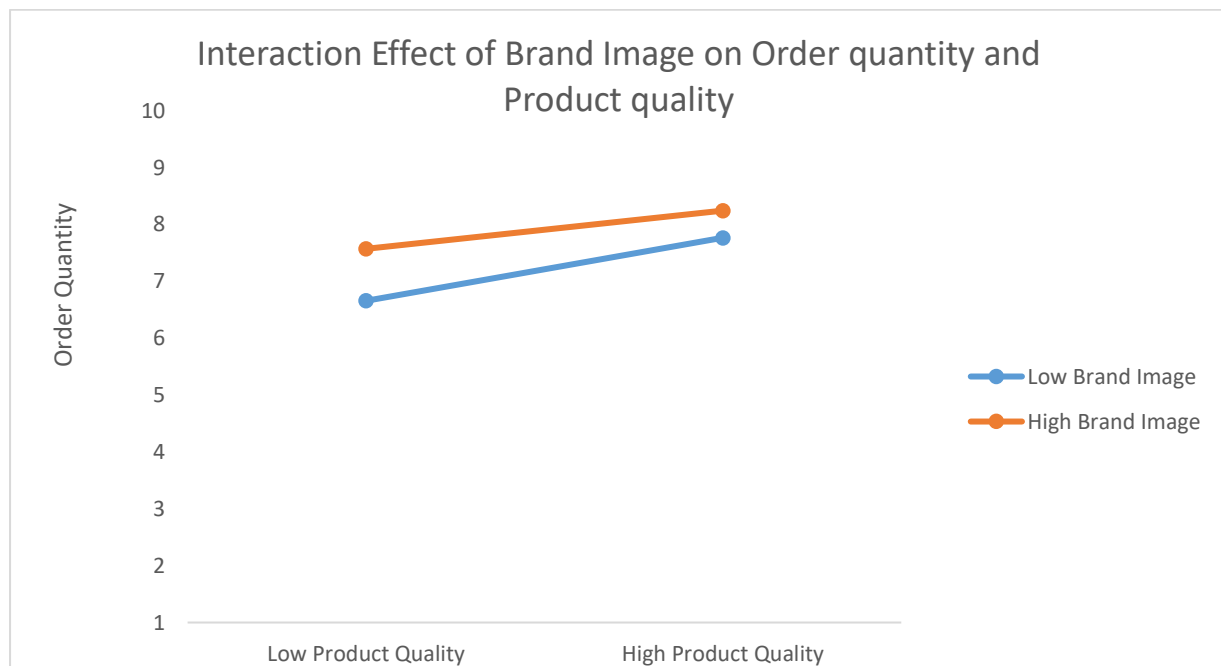
The model meets the normality, linearity, independence of error, and equal variance assumptions (refer to appendix g). The significant independent variables (Loyalty, product quality perception, brand image, shipping cost and contract) in the final regression model has a linear relationship with the dependent variable, no evident pattern, residuals lie within a thick stand and the normal probability plot is almost a straight line.

Potential Outliers: From the standard residuals, 10 numbers, with five lesser than -2 and the remainder greater than 2 were found and identified as potential outliers.

2.3. THE INTERACTION EFFECTS OF BRAND IMAGE PERCEPTION ON PRODUCT QUALITY AND ORDER QUANTITY

2.3.1. The Interaction Effect

The interaction effect of brand image perception on the quality and the number of purchased products was identified. A model was built to test for the assumption that a favourable TPM brand image causes stronger relationship between the quality and the number of purchased products. The result of the regression model shows that all the individual independent variables (including the interaction term) are statistically significant.



From the interaction effect visualization, the chart shows that perception of brand image affects order quantity across low and high levels of product quality perceptions. At both low and high product quality, order quantity on average is higher when the perception of brand image is high than when

low. Overall, the brand image moderates the relationship between the perceived quality of the products and quantity of the products purchased. This means that the brand image interacts with the product quality in predicting quantity of orders by customers.

2.3.2. Managerial Recommendations

- Based on the interaction effect, it is recommended that TPM managers strategize on how best to improve the overall image of the company's brand.
- The quality of the products sold should also be improved as this might serve as a boost to the perceived overall image of the company's brand.

Adhering to these recommendations would cause

- Increase in the favourable perception of product quality and brand image
- Increase in customer satisfaction
- Increase in quantity of order requests leading to a growth in sales improving the financial performance of the organization positively.

2.4. MODELLING THE LIKELIHOOD OF CUSTOMER SIGNING A CONTRACT WITH TPM

Logistic regression was used for this model building since the dependent variable (contract) is categorical. Multiple iterations were done due to a statistically insignificant independent variable (competitive pricing) in the previous model built.

2.4.1. Model Classifications

Of the 101 customers who signed contracts with TPM, the model accurately classified 79 as "signed contract with TPM" and 22 were inaccurately classified as those with "no signed contract with TPM" having 78.2% classification accuracy. Of the 99 customers with no signed contracts with TPM, 74 were accurately classified as "no signed contract with TPM" and 25 were inaccurately classified as those who "signed contract with TPM" having 75% classification accuracy.

The overall classification accuracy (hit ratio) was 77%. 77 percent of the 200 TPM customers were accurately classified by the logistic regression model. However, the remaining mis-classification rate (23%) could have been captured accurately if more relevant Independent variables had been included in the model (refer to appendix h).

2.4.2. Establishing Practical Significance

The model significance can be established using the cut-off (50%), PCC hit ratio of 50% and/or standard hit ratio 63%. The accuracy rate of 77 is greater than the PCC hit ratio, cut-off and standard hit ratio providing evidence for practical significance of logistic model. This means that the logistic model is significantly better than a random process in classifying observations (refer to appendix h).

2.4.3. Establishing Statistical Significance

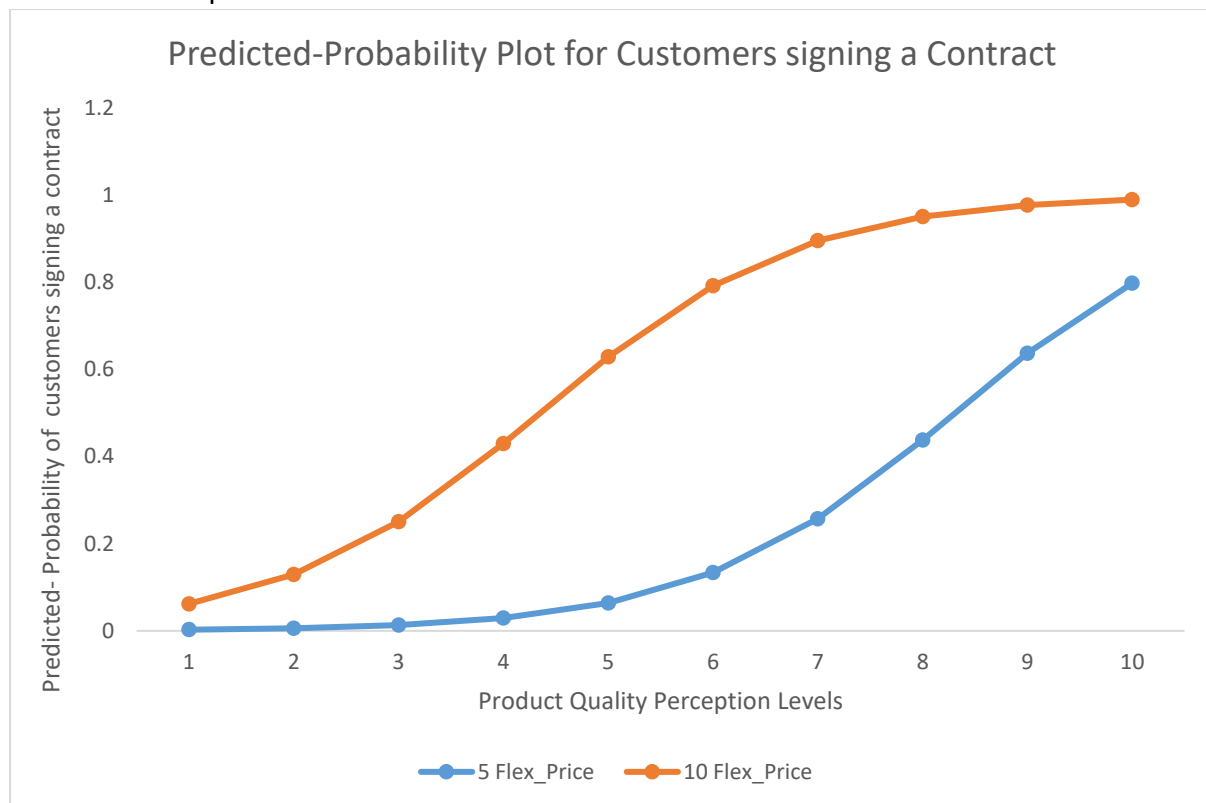
Compared to baseline model, the final logistic model significantly reduced LL value, providing evidence for statistical significance of the overall model (refer to appendix). According to Cox /Snell and Nagelkerke R^2 33.5% and 44.6% of variations in the dependent variable can be explained by the regression model respectively. These values are acceptable considering the overall model fit and the practical significance of the model (refer to appendix i).

The ROC Curve: The area under the curve is very close to 1.0, pointing to a model that fits data well (Refer to appendix j).

2.4.4. The percentage of change in odds

- For an individual with no level of product quality and brand image perceptions as well as the lack of product line that meets customer needs with no willingness from TPM for price negotiation, the likelihood (odds) of a customer signing a contract with TPM decreases by 16.8 percent (refer to appendix k).
- One unit of increase in the perception of product quality increases the likelihood (odds) of a customer signing a contract with TPM by 81.1
- One unit of increase in the depth and breadth of TPM product line to meet customer needs increases the likelihood (odds) of a customer signing a contract with TPM by 61.4 percent.
- One unit of increase in the perception of TPM brand image increases the likelihood (odds) of a customer signing a contract with TPM by 76.2 percent.
- One unit of increase in the perceived willingness of TPM to negotiate product prices increases the likelihood (odds) of a customer signing a contract with TPM by 64.2 percent

2.4.5. The Interpreted and Predicted Probabilities



For customers with neutral feelings towards brand image and product line, when there is a change in the product quality and price flexibility, it affects the likelihood of the customers signing a contract with TPM. An increase in the perception of product quality and willingness of TPM to negotiate product prices causes an increase in the likelihood of customers signing a contract that nominates the company as their supplier when there exist a neutral feeling about the overall brand image and TPM's product line meeting their needs. Under same neutral feelings, even with a high price flexibility, if product quality perception is below neutral, the odds of getting a customer to sign the contract is lower than 0.5.

MANAGERIAL RECOMMENDATIONS

Based on the findings, TPM should spend time and resources introducing quality products that meets their customer needs. This would not only increase the perception of their product line but also the perceived quality of their products and their overall brand image. The company should also be willing to negotiate their product prices in a way that it satisfies both the organization and the customer. This improvement would attract customers and increase the likelihood of customers signing TPM as their preferred supplier.

2.5. FORECASTING TURNOVER FOR NEXT FOUR QUARTERS.

A multiplicative model (seasonal forecast) was created to forecast TPM's turnover for the next four quarters using 2010 to 2019 financial year records. The multiplicative model was used because it is best suited for seasonal forecast with the magnitude of the seasonal pattern, dependent on the data magnitude.

The Forecasted Figures:

2020 Quarter 1: 1728.02 tonnes

2020 Quarter 2: 2128.02 tonnes

2020 Quarter 3: 1601.02 tonnes

2020 Quarter 4: 1767.81 tonnes

The model predicts that sales in the second quarter of the year will be higher than other quarters and quarter 3 would have the lowest turnover in that 2020 financial year (refer to appendix I).

The mean absolute percentage error that this model is likely to make is 4%.

3. CONCLUSION

During the analysis, few limitations were identified:

- The number of observations might be too small to accurately make predictions about all TPM's customers
- Some of the data used for predictions were from online surveys which might cause inaccuracy in the data if information provided lacks truth
- Identified potential outliers might have affected the results and accuracy of the predictions.

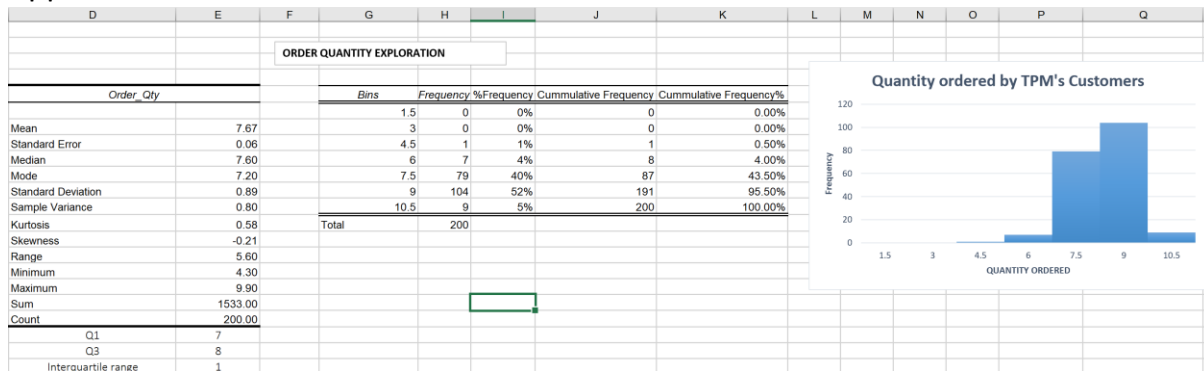
However, despite the limitations, there were findings and conclusions drawn from the analysis and predictions. From the overall findings, for TPM to improve the current financial state of the business, there must be changes made about the perception of the company and its products through the introduction of better and improved products that meet customer's needs. This gives a favourable perception of the breadth and depth of the product line, the quality of TPM's products and overall brand image.

The improved perceptions would give TPM:

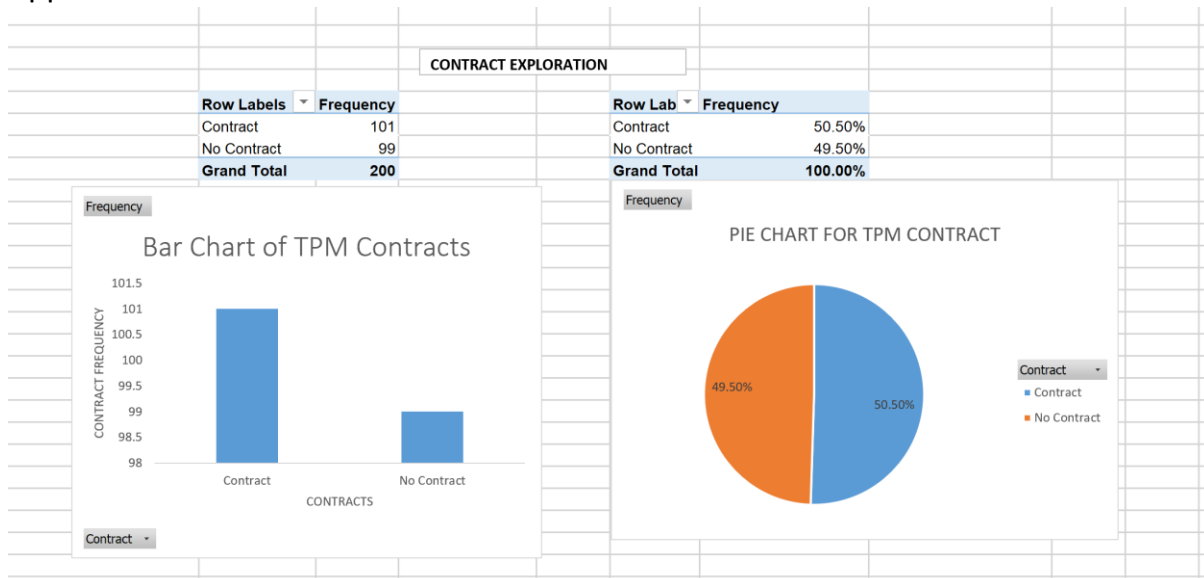
- Growth in customer base and improved customer loyalty and retention
- Better customer satisfaction
- Increase in the quantity of products ordered
- Increase in customer signed contracts nominating TPM as a supplier.

4. APPENDICES

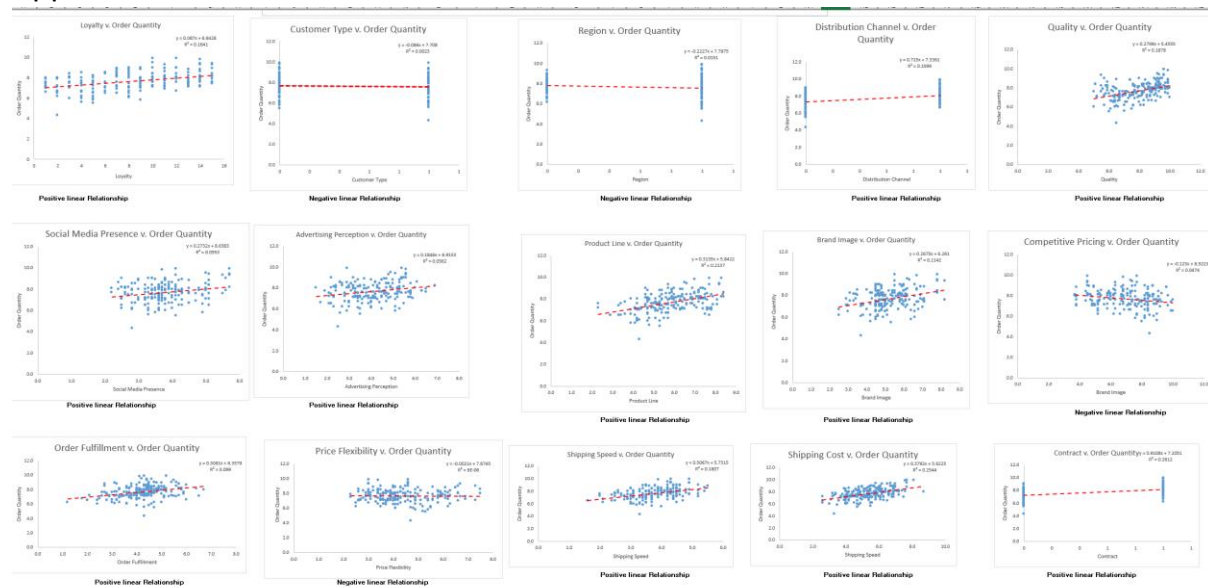
Appendix a



Appendix b



Appendix c



Appendix d

CORRELATION MATRIX SHOWING THE MAGNITUDE AND MULTICOLLINEARITY CHECK BETWEEN THE INDEPENDENT VARIABLES AND DEPENDENT VARIABLE.

	Loyalty	Cust_Type	Region	Dist_Chann	Quality	SM_Presence	Advert	Prdct_Line	Brand_Image	Comp_Pricing	Order_Fulfillment	Flex_Price	Shipping_Speed	Shipping_Cost	Contract	Order_Qty
Loyalty	1.000															
Cust_Type	-0.051	1.000														
Region	0.072	-0.031	1.000													
Dist_Chann	0.223	-0.140	-0.280	1.000												
Quality	0.084	-0.049	-0.544	0.382	1.000											
SM_Presence	0.190	0.145	0.203	0.301	-0.034	1.000										
Advert	0.259	0.006	0.223	0.167	-0.054	0.505	1.000									
Prdct_Line	0.153	-0.028	-0.514	0.457	0.509	0.077	0.117	1.000								
Brand_Image	0.258	0.005	0.379	0.290	-0.116	0.788	0.627	0.050	1.000							
Comp_Pricing	0.076	0.149	0.589	-0.332	-0.448	0.177	0.099	-0.483	0.200	1.000						
Order_Fulfillment	0.139	-0.009	0.005	0.263	0.083	0.217	0.230	0.444	0.284	-0.060	1.000					
Flex_Price	0.058	0.040	0.575	-0.217	-0.487	0.186	0.260	-0.349	0.272	0.470	0.419	1.000				
Shipping_Speed	0.196	0.016	0.014	0.254	0.067	0.241	0.323	0.512	0.299	-0.055	0.773	0.513	1.000			
Shipping_Cost	0.175	0.018	0.002	0.240	0.141	0.215	0.247	0.584	0.298	-0.094	0.696	0.358	0.840	1.000		
Contract	0.173	-0.010	-0.104	0.432	0.361	0.235	0.189	0.381	0.310	-0.130	0.315	0.058	0.398	0.378	1.000	
Order_Qty	0.405	-0.048	-0.123	0.400	0.433	0.235	0.237	0.462	0.338	-0.218	0.315	-0.003	0.424	0.504	0.511	1.000

Appendix e

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.75
R Square	0.56
Adjusted R Square	0.53
Standard Error	0.61
Observations	200.00

ANOVA

	df	SS	MS	F	Significance F
Regression	14.00	89.29	6.38	16.98	0.00
Residual	185.00	69.49	0.38		
Total	199.00	158.78			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	4.07	0.62	6.59	0.00	2.85	5.29	2.85	5.29
Loyalty	0.06	0.01	4.93	0.00	0.03	0.08	0.03	0.08
Cust_Type	0.01	0.09	0.12	0.91	-0.17	0.19	-0.17	0.19
Region	0.02	0.14	0.11	0.91	-0.27	0.30	-0.27	0.30
Dist_Channel	0.05	0.12	0.47	0.64	-0.17	0.28	-0.17	0.28
Quality	0.20	0.04	4.48	0.00	0.11	0.29	0.11	0.29
SM_Presence	-0.11	0.10	-1.16	0.25	-0.31	0.08	-0.31	0.08
Advert	-0.04	0.05	-0.73	0.47	-0.14	0.06	-0.14	0.06
Prdct_Line	0.03	0.07	0.37	0.71	-0.12	0.17	-0.12	0.17
Brand_Image	0.23	0.08	2.82	0.01	0.07	0.39	0.07	0.39
Comp_Pricing	-0.06	0.04	-1.58	0.12	-0.13	0.01	-0.13	0.01
Order_Fulfillment	-0.14	0.08	-1.89	0.06	-0.30	0.01	-0.30	0.01
Flex_Price	0.06	0.08	0.82	0.41	-0.09	0.21	-0.09	0.21
Shipping_Cost	0.25	0.07	3.51	0.00	0.11	0.39	0.11	0.39
Contract	0.31	0.11	2.79	0.01	0.09	0.53	0.09	0.53

INTERPRETATION

1. Anova: p-value < 0.05 then Reject Ho. Overall the model has some predictive power.

2. First Model: customer type, region, distribution channel, social media presence, advert, product line, competitive pricing, order fulfillment and price flexibility were not significant as their p-values > 0.05

3. R square = 56%. Approximately 56% of variation in the quantity ordered can be explained by the regression model. Approximately 44% of the variation in quantity ordered would be explained by other factors not included in the model.

Appendix f

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.74							
R Square	0.54							
Adjusted R Square	0.53							
Standard Error	0.61							
Observations	200.00							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	5.00	86.10	17.22	45.96	0.00			
Residual	194.00	72.68	0.37					
Total	199.00	158.78						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	3.53	0.40	8.87	0.00	2.75	4.32	2.75	4.32
Loyalty	0.05	0.01	4.94	0.00	0.03	0.08	0.03	0.08
Quality	0.21	0.03	5.93	0.00	0.14	0.28	0.14	0.28
Brand_Image	0.13	0.04	2.97	0.00	0.04	0.22	0.04	0.22
Shipping_Cost	0.22	0.04	5.48	0.00	0.14	0.30	0.14	0.30
Contract	0.34	0.10	3.24	0.00	0.13	0.54	0.13	0.54

Appendix g



Appendix h

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
	Classification Table																
		Success-Observed	Fail-Observed	Total													
	Success-Predicted	79	25	104													
	Fail-Predicted	22	74	96													
	Total	101	99	200													
	Accuracy	0.782178218	0.747474747	0.765	The overall classification accuracy (hit ratio) was 77% %, i.e. 77 percent of the 200 TPM customers were accurately classified by the logistic regression model. The remaining misclassification rate (23%) could have been captured accurately if more relevant IVs had been included in the model.												
	Cutoff			0.5													
		Of the 101 customers who signed contracts with TPM (observed), 79 were accurately classified (predicted) as "signed contract with TPM" and 22 were inaccurately classified as those with "no signed contract with TPM" (78.2% classification accuracy).	Of the 99 customers with no signed contracts with TPM (observed), 74 were accurately classified (predicted) as "no signed contract with TPM" and 25 were inaccurately classified as those who "signed contract with TPM" (75% classification accuracy).														
	Comparing accuracy rate with hit ratio																
	PCC hit ratio = $p^2 + (1 - p)^2$	50%															
	Standard (rule of thumb)	63%															
	Accuracy rate of 77 is greater than PCC hit ratio, providing evidence for practical significance of logistic model. That is, logistic model is significantly better than a random process (chance) in classifying observations.																
	Accuracy rate of 77 is greater than standard hit ratio (i.e. one-fourth greater than achieved by chance = .50), providing evidence for practical significance of logistic model. That is, logistic model is significantly better than a random process (chance) in classifying observations.																

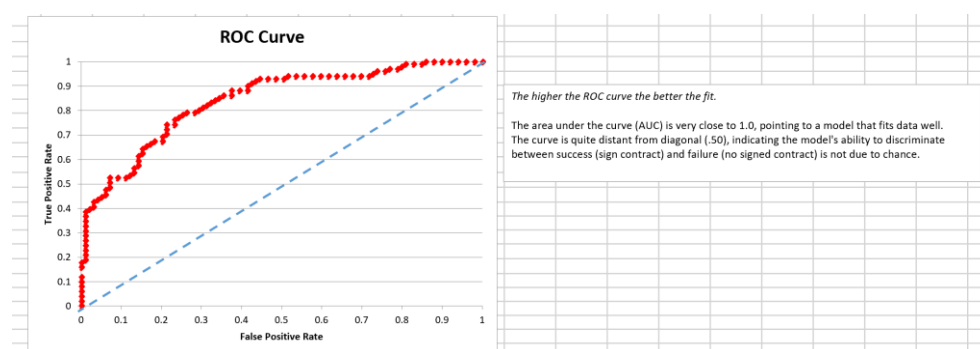
Appendix i

Comparing accuracy rate with hit ratio		
PCC hit ratio = $p^2 + (1 - p)$	50%	
Standard (rule of thumb)	63%	
LL0	-138.6194	
LL1	-97.8744	
Chi-Sq	81.4900	
df	4.0000	
p-value	0.0000	
alpha	0.0500	
sig	yes	Compared to baseline model, the final logistic model significantly reduced LL value, providing evidence for statistical significance of the overall model
R-Sq (L)	0.294	
R-Sq (CS)	0.335	According to Cox and Snell R^2 , 33.5 percent of variation in the dependent variable can be explained by the regression model.
R-Sq (N)	0.446	According to Nagelkerke R^2 , 44.6 percent of variation in the dependent variable can be explained by the regression model.
Coupled with practical significance of the model, along with overall model fit, these R^2 values are deemed as acceptable.		

Accuracy rate of 77 is greater than PCC hit ratio, providing evidence for practical significance of logistic model. That is, logistic model is significantly better than a random process (chance) in classifying observations.

Accuracy rate of 77 is greater than standard hit ratio (i.e. one-fourth greater than achieved by chance = .50), providing evidence for practical significance of logistic model. That is, logistic model is significantly better than a random process (chance) in classifying observations.

Appendix j



Appendix k

Percentage of Change in Odds			PERCENTAGE OF CHANGE IN ODDS INTERPRETATION
	coeff(b)	exp(b)	
Intercept	-16.827	0.000	For an individual with no level of product quality and brand image perceptions as well as the lack of product line that meets customer needs and no willingness for price negotiation, the likelihood (odds) of a customer signing a contract with TPM decreases by 16.8 percent. For an individual with no level of product quality and brand image perceptions as well as the lack of product line that meets customer needs and no willingness for price negotiation, the likelihood (odds) of a customer signing a contract with TPM decreases by 16.8 percent.
Quality	0.811	2.251	
Prdct Line	0.614	1.848	
Brand Image	0.762	2.143	One unit of increase in the perception of product quality increases the likelihood (odds) of a customer signing a contract with TPM by 81.1 percent.
Flex Price	0.642	1.901	One unit of increase in the depth and breadth of TPM product line to meet customer needs increases the likelihood (odds) of a customer signing a contract with TPM by 61.4 percent.
			One unit of increase in the perception of TPM brand image increases the likelihood (odds) of a customer signing a contract with TPM by 76.2 percent.
			One unit of increase in the perceived willingness of TPM to negotiate product prices increases the likelihood (odds) of a customer signing a contract with TPM by 64.2 percent.

Appendix I

