Supermarket Project - Analytics

Introduction:

Mitchies Superstore one of Australia's leading supermarket chains with 500 stores in the chain. Due to increasing competition in grocery supermarket, they foresee downsize in Sales revenue. Mitchies Supertore has heavily invested in digital platforms for online sales channel and self-check out machines. However, store managers are somewhat reluctant to operate on Sundays and use online sales channel.

Objective of this analysis is to-

1) Predict sales revenue.
2) Likelihood of stores operating on Sunday
3) Likelihood of stores capitalizing online sales channel
4) Forecasting sales for quarters 2019-Q2 to 2020-Q1

Random sample of 150 stores is collected with 19 variables. Further this report contains; A] Main body which includes details of model and its interpretation.
B] Conclusion with limitation, and C] Appendices which includes important artifacts.

Main body:

1. Summary of dependent variables:

1.1 Sales for the financial year ($million):

a) Total sales for 150 stores in sample is $17489.
b) Average sales across is $11.7.
c) Store number 65 has highest sales revenue of $23.5 and store number 40 has lowest sales revenue of $5.9.
d) Range is $17.6.
e) Sales is highly spread across stores from the mean sales as standard deviation is 3.57.
f) Median is $10.95 with IQR of $5.18.
g) Skewness is 0.74 which means sales across stores is more concentrated towards lower side.
h) Four stores have same sales of $8.10 which is the mode of the sales.
i) Sales data is not normally distributed which is evident from standard deviation, skewness, and Shapiro-Wilk Test.
j) Around 80% of stores have sales less than $16 mn.

1.2 Open on Sundays:

a) Out of 150 stores 96 stores (64%) are open on Sunday and 54 are closed (54%).

1.3 Online sales channel:

a) Out of 150 stores 62 stores (41%) have online sales channel and 88 stores (59%) do not have online sales channel.

2. Model to predict Sales:

Linear regression model with 5% alpha level is used to predict Sales revenue ($ million) for the financial year.

2.1. Pre-model diagnosis:

Variables having correlation coefficient between -0.2 and 0.2 are considered to have no linear relationship with Sales and thus excluded from the model. Sales has strong positive linear relationship with Advertisement expenses and Wage expenses. Store manager's gender and experience, Number of staffs, Car spaces, Cost basket (2017) and %Change in cost basket has moderate positive linear relationship with Sales. Online channel has weak positive linear relationship and Number of competitors have weak negative linear relation with. Rest of the independent variables have weak linear relationship with Sales i.e. between -0.2 and 0.2. No independent variable seems to have non-linear relationship with Sales.

(Refer 2.1 section in Appendices for scatter plots and correlation matrix).

2.2. Model specification:

Wage expenses has strong multicollinearity with Number of staffs with correlation coefficient of 0.92. Although wages have strong correlation with Sales, but ideally wages depends on Sales and not vice versa. Thus, Number of staffs is selected in the model instead of Wage expenses. Cost basket of 2016 and 2017 also have strong multicollinearity with correlation coefficient of 0.98. Data for 2017 is estimated and that estimation is captured in %Change in basket, thus, cost basket (2017) is dropped from the model. Potential independent variables for the first iteration of the model are Store manager's gender and experience, Online channel, Number of staffs, Advertise expenses, Number of competitors, Car spaces, Cost basket (2016), %Change in cost basket.

2.3. Model building:

Backward elimination based on highest p-value was used. Model has undergone 4 iterations and 5th model (2.3e) is the final model. Below is the estimated model 2.3e.

$\widehat{\text{Annual Sales}}$ = -27.962 + 0.061(Manager Experience) + 0.025(Number of staffs) + 0.037(Advertisement expense) + 0.178(Cost basket - 2016) + 36.303(%Change in cost basket)

2.4. Post-model diagnosis: Model 2.3e

2.4.1: Overall Significance:

Model 2.3e is overall significant as F statistic is 449.271 with p-value approximately 0. Thus, we conclude that this model has predictive power and explains significant proportion of the variation in Sales revenue and there is evidence that at least one independent variable has linear relationship with Sales in the population.

Goodness of fit, R^2 is 93.97%, thus we estimate that about 93.97% of the variation in Sales can be explained by the variation in independent variables. The remaining 1-R^2 = 6.02% of variation in Sales would be explained by other factors like location and customer demographics which is not included in the model. Similarly, based on Adjusted R^2 we estimate that 93.76% of the variation in Sales is explained by variation in these independent variables considering sample size of 150 stores and 5 independent variables. Based on Standard error, we estimate average error of 0.892% in predicting Sales using this model.

2.4.2: Significance of independent variables:

There is linear relationship in the population between Sales and independent variables in the model as p-value of t-statistics for five independent variables is less than 5% level of significance. Thus, we conclude that these independent variables have some power to

predict variation in Sales. Also, VIF for all variables is less than 5, thus we conclude that independent variables are not highly correlated with each other. (Refer Fig 2.3.1 for p-values)

2.4.3: Interpretation of independent variables:

1) Intercept (b_0):
   On an average, Sales for the financial year is -$27,962,351 for the stores whose Store manager has no experience, no staff, zero advertisement cost, cost basket (2016) is zero, and no change in cost basket. This coefficient has no practical sense as store incurring loss cannot continue its business.

2) Store manager's experience (years):
   b_1 is the regression coefficient of Store manager's experience = 0.061. Assuming no change in other independent variables, an increase in a year of Store manager's experience would increase Sales revenue for the financial year on an average by $61,197.

3) Number of staffs:
   b_2 is the regression coefficient of this variable = 0.025. Assuming no change in other independent variables, if 1 staff is added to workforce then Sales for the financial would increase on an average by $24,708.

4) Advertisement cost ($'000):
   b_3 is the regression coefficient of this variable = 0.037. Assuming no change in other independent variables, an extra $1000 spent on advertisement would increase Sales revenue for the financial year on an average by $37,099.

5) Cost basket (2016) ($):
   b_4 is the regression coefficient of this variable = 0.178. Assuming no change in other independent variables, an increase of $1 in Cost basket would increase Sales revenue for the financial year on an average by $178,197.

6) Change in cost basket (%):
   b_5 is the regression coefficient of this variable = 36.303%. Assuming no change in other independent variables, 1% increase in cost basket would increase Sales revenue for the financial year on an average by $363,030.

2.4.4: Residual analysis:

Residuals (e_i) are randomly and evenly spread across each independent variable (X_i) and Thus, assumption of linearity, independence, and homoscedasticity is not violated (Refer figures 2.4.4.1 to 2.4.4.5 in appendices). Residuals follow normal distribution as per QQ plot and as per cumulative percentage polygon, 98.67% of Standardized Residuals fall within range of -2 and +2. (Refer figure 2.4.4.7 and 2.4.4.8 in appendices)

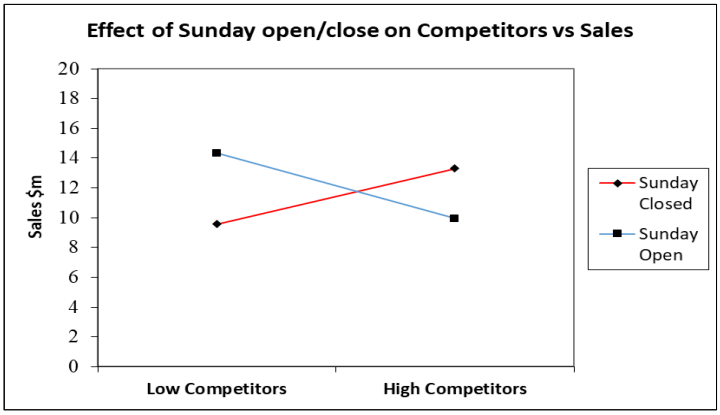
3. Interaction effect between competitors and Stores operating on Sunday:

Model without interaction term has 11.85% R^2 and 10.65% Adjusted R^2 with Sunday variable being insignificant. By adding interaction variable of Competitors and Sunday, R^2 increases to 35.91% and Adjusted R^2 to 34.59%. Also, model has some predictive power as p-value of F-statistic is less than 5% (significance level). With 5% significance level, we conclude that interaction effect does exist, thus, effect of number of Competitors on Sales revenue depends if store is open or closed on Sunday. Below is the estimated regression equation where b_1 is Competitors, b_2 is

Sunday (Open =1 and Close = 0), and interaction term (Competitors * Sunday) is b_3. (Refer below Fig 3.1, and Fig 3.2 to Fig 3.4 in appendices).

$\widehat{\text{Annual Sales}}$ = 8.154 + 1.247(Competitors) + 7.857(Sunday) - 2.717(Competitors x Sunday)

Fig 3.1: Interaction effect plot



At low level of competitors, stores operating on Sunday have higher Sales. At high level of competitors, stores closed on Sunday have higher Sales. Since, the interaction variable's regression coefficient is negative, stores which are closed on Sunday can have more Sales as number of Competitors increase.

4. Likelihood of a store opening on Sunday:

Logistic regression with cutoff of 0.5 and 5% level of significance of Wald statistic is used to calculate likelihood of a store opening on Sunday.

4.1 Pre-model diagnosis:

Includes scatter plot and box plot on each variable segregated based on Sunday open (1) and Sunday closed (0). Most of the variables have significant overlap in their values for Sunday open and close. However, Advertisement expense, Gross Profit, %Change in basket, CarSpaces, Sales, Number of staffs, Competitors, Manager age and experience have less overlap, but these variables have significant outliers, thus, resulting into higher p-value. Also, there is no correlation between any of the independent variable and Sunday; strongest being -17.5% between Manager age and Sunday. (Refer fig 4.1.1 to 4.1.15 in appendices)

4.2 Model building:

Backward elimination method based on highest p-value of Wald statistic was used. In first iteration all 18 variables were selected. Model has gone through 13 iterations and 14^{th} iteration is the final model (Model 4n). Below is the significance and interpretation of the final model.

4.3 Practical significance of the model:

a) The overall classification accuracy (Hit ratio) - 73.33% of stores in sample were accurately classified by the logistic regression model. The remaining miss-classification rate could have been captured accurately if more relevant IVs had been included in the sample.
b) Out of 96 stores who are open on Sunday (observed), 84 were accurately classified (predicted) as "Open on Sunday" (87.5% classification accuracy); and 12 were inaccurately classified as those who "Close on Sunday" (12.5% misclassification rate).

c) Out of 54 stores which are closed on Sunday (observed), 26 were accurately classified (predicted) as "Closed on Sunday" (48.1% classification accuracy); and 28 were inaccurately classified as those who "Open on Sunday" (51.9% misclassification rate).
d) Since, sample has unbalanced observations of success (Open on Sunday) and failure (Close on Sunday). Hit ratio is compared with PCC hit ratio and Standard hit ratio. Accuracy rate of 73.3% is greater than PCC hit ratio (53.9%) and Standard hit ratio (62.5%), providing evidence for practical significance of logistic model. That is, logistic model is significantly better than a random process (chance) in classifying observations.

4.4 **Statistical significance of the model:**

a) Overall fit - With p-value less than 5% level of significance of Chi-Sq test, we conclude that, compared to baseline model (LL0), the final logistic model (LL1) significantly reduced LL value, providing evidence for statistical significance of the overall model.
b) Variation in the dependent variable explained by the regression model is 12.9%, 15.5%, and 21.2% according to Hosmer and Lemeshow, Cox and Snell, Nagelkerke respectively.
c) Coupled with practical significance of the model, along with overal model fit, and poor correlation between independent and dependent variable, these R2 values are deemed as acceptable.

4.5 Contribution of predictors:

a) Online channel - The likelihood (odds) of a store operating on Sunday if store has an online sale channel is 78.26% lesser than that of a store not having online sale channel with no change in other independent variables.
b) Number of staff - For each increase in the number of staffs, likelihood (odds) of a store operating on Sunday increases by 3.72% with no change in other independent variables.
c) Gross profit - For $1 million increase in gross profit, likelihood (odds) of a store operating on Sunday increases by 88.28% with no change in other independent variables.
d) Competitors - For each increase in number of competitors, likelihood (odds) of a store operating on Sunday increases by 42.12% with no change in other independent variables.
e) Store manager age - One year increase in the age of a store manager decreases the likelihood (odds) of a store operating on Sunday by 9.17% with no change other independent variables.
f) Store manager experience - One year increase in the experience of a store manager increases the likelihood (odds) of a store operating on Sunday by 11.4% with no change in other independent variables.
g) ROC curve is distant from diagonal (0.50), indicating the model's ability to discriminate between success (Store open on Sunday) and failure (Store close on Sunday) is not due to chance.

Refer fig 4.2 and 4.3 in appendices for regression output.

5. Likelihood of a store opening an Online sales channel:

Logistic regression with cutoff of 0.5 and 5% level of significance of Wald statistic is used to calculate likelihood of a store opening an Online sales channel (Refer fig 5.1 and 5.2)

5.1 Practical significance of the model:

a) The overall classification accuracy (Hit ratio) - 76.7% of stores in sample were accurately classified by the logistic regression model. The remaining miss-classification rate could have been captured accurately if more relevant IVs had been included in the model.
b) Out of 62 stores who have Online sales channel (observed), 43 were accurately classified (predicted) as "Store has online sales channel" (69.4% classification accuracy); and 19 were inaccurately classified as those who "Stores do not have online sales channel" (30.6% misclassification rate).
c) Out of 88 Stores who do not have online sales channel (observed), 72 were accurately classified (predicted) as "Stores do not vote have online sales channel" (81.8% classification accuracy); and 16 were inaccurately classified as those who "Stores have online sales channel" (18.2% misclassification rate).
d) Since, sample has unbalanced observations of success (Stores with online sales channel) and failure (Stores with no online sales channel). Hit ratio is compared with PCC hit ratio and Standard hit ratio. Accuracy rate of 76.7% is greater than PCC hit ratio (51.5%) and Standard hit ratio (62.5%), providing evidence for practical significance of logistic model. That is, logistic model is significantly better than a random process (chance) in classifying observations.

5.2 Statistical significance of the model:

h) Overall fit - With p-value less than 5% level of significance of Chi-Sq test, we conclude that, compared to baseline model (LL0), the final logistic model (LL1) significantly reduced LL value, providing evidence for statistical significance of the overall model.
i) Variation in the dependent variable explained by the regression model is 32.8%, 35.9%, and 48.4% according to Hosmer and Lemeshow, Cox and Snell, Nagelkerke respectively.
j) Coupled with practical significance of the model, along with overal model fit, these R2 values are deemed as acceptable.

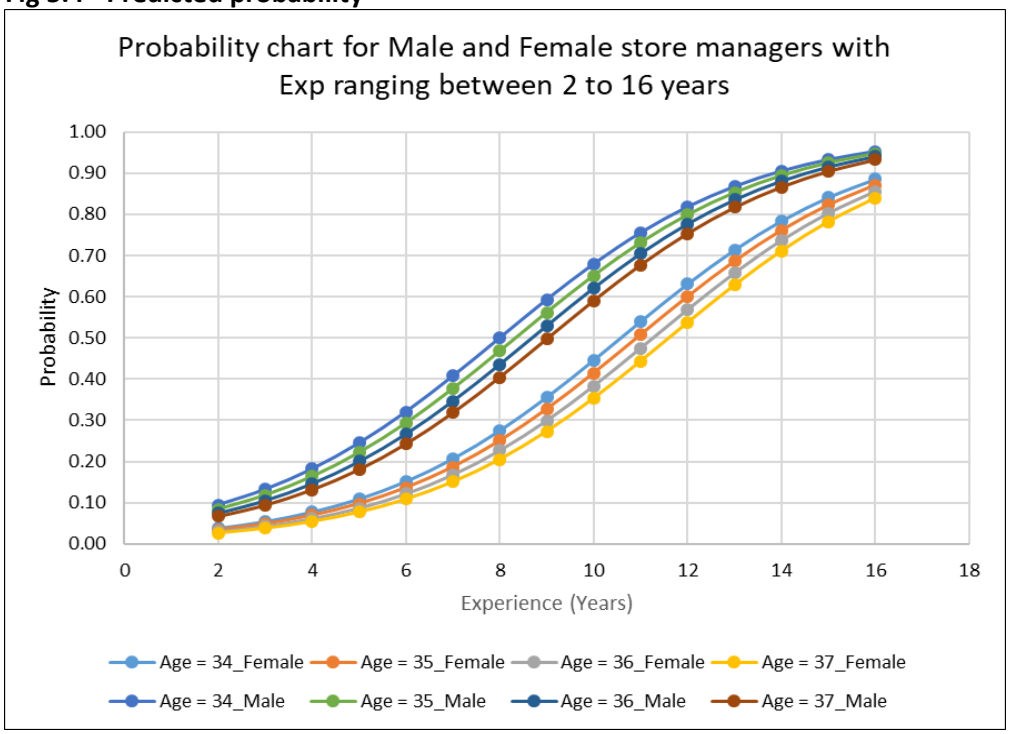
5.3 Contribution of predictors:

a) Store manager gender - The likelihood (odds) of a store opening an Online sale channel if store manager is male is 163.70% greater than that of a female store manager with no change in store manager age and experience.
b) Store manager age - One year increase in the age of a store manager decreases the likelihood (odds) of a store opening an Online sale channel by 12.12% with no change in store manager gender and experience.
c) Store manager experience - One year increase in the experience of a store manager increases the likelihood (odds) of a store opening and Online sale channel by 45.65% with no change in store manager gender and age.
d) ROC curve is distant from diagonal (0.50), indicating the model's ability to discriminate between success (Store opening an Online sale channel) and failure (Store not opening an Online sale channel) is not due to chance.
e)

5.4 Predicted probabilities w.r.t Store manager's experience, gender, and age (34 - 37):

Overall, probability of a store having an online sale channel is greater if store manager is male instead of female across experience (2-16 years) and age (34-37 years. Probability decreases as age increases from 34 to 37 for male and female manager. Probability increases as store manager's experience increases for both the gender. (Refer Fig 5.4).

Fig 5.4 - Predicted probability



6. **Forecasting sales:**

Sales data is quarterly starting from 2015-Q2 till 2019-Q1. Data has seasonal component where sales drops in Q1, gradually increases and peaks in Q4 to fall back in Q1 of next year. Seasonality does not remain constant over a period and has upward trend and irregular component. Thus, multiplicative model has been used to forecast sales for next 4 quarters. In sample MAPE of the forecast is 20.04% and forecast for 2019-Q2, 2019-Q2, 2019-Q3, and 2020-Q1 is $99.66 mn, $120.41 mn, $138.14 mn, and $160.45 mn respectively. (Refer fig 6 in appendices).

Conclusion:

Mitchies Superstore should spend more on advertisement as $1,000 invested in advertisement would increase sales revenue by $37,099. Stores should be obligated to have online sales channel as average sales for stores with online sales channel is $2.80 mn p.a. more than stores without online sales channel. Experienced store manager's (preferable male) should be appointed to increase probability of utilizing online sales channel. However, manager with age beyond mid 30's should be upskilled to utilize online sales channel. Stores with less than or equal to 3 competitors should operate on Sundays and stores with competitors greater than 3 should be discouraged to operate on Sunday.

Limitations - Variables were not adjusted for outliers and normality (linear regression). Customer demographics would have provided more insights to improve the model.

Appendices:

1. Summary of dependent variables:

Fig - 1.1 Distribution of Sales

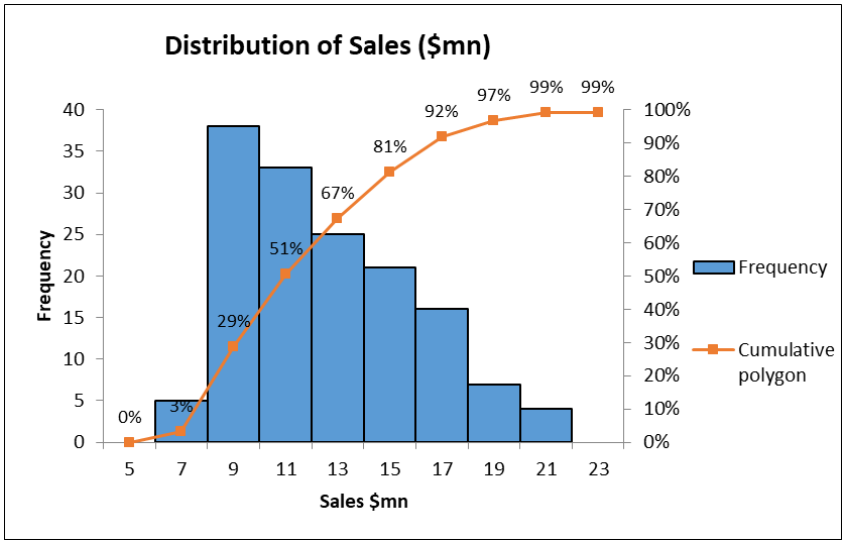


Fig - 1.2 Stores open on Sunday

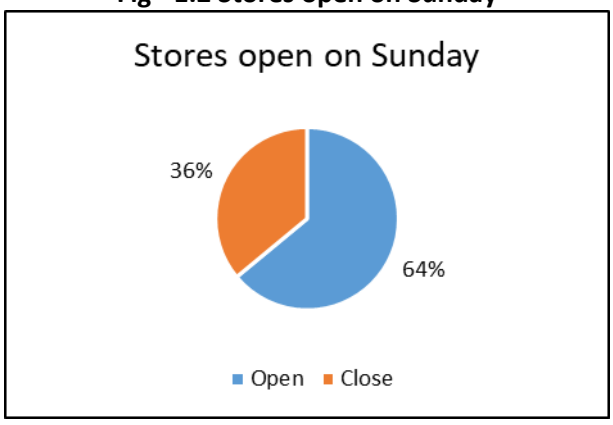


Fig - 1.3 Stores with online sales channel

2.1 Pre-model diagnosis: Scatterplots and correlation matrix

Fig - 2.1.1

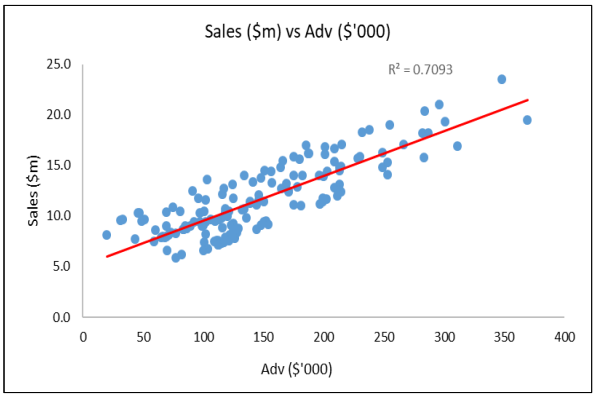


Fig - 2.1.2

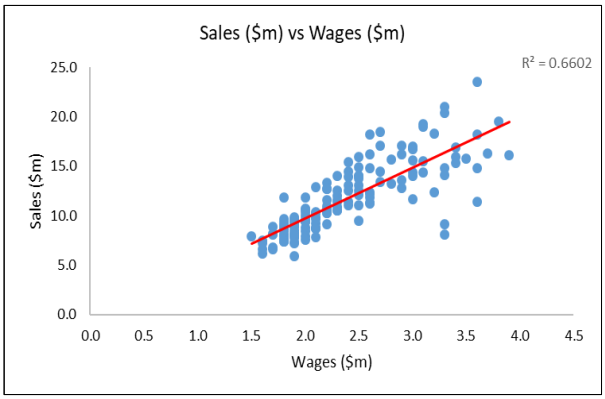


Fig - 2.1.3

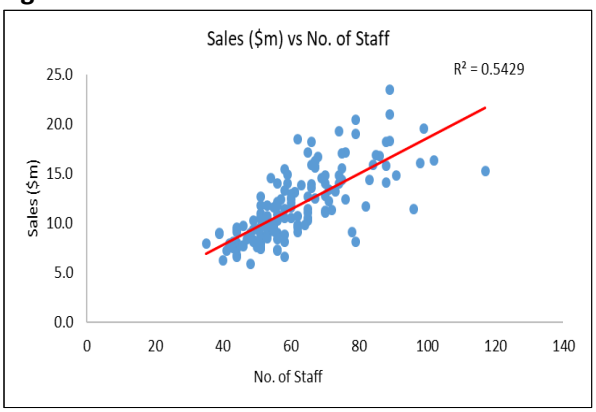


Fig - 2.1.4

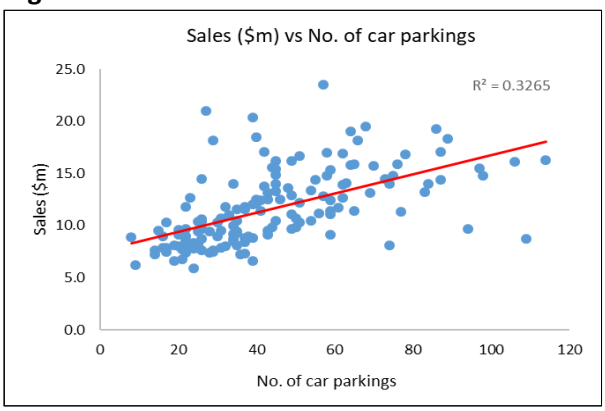


Fig - 2.1.5

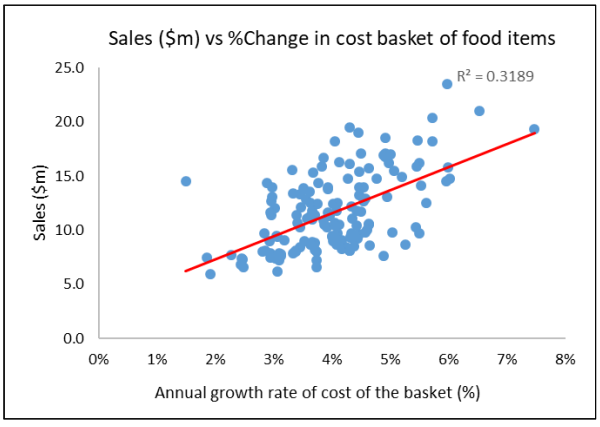


Fig - 2.1.6

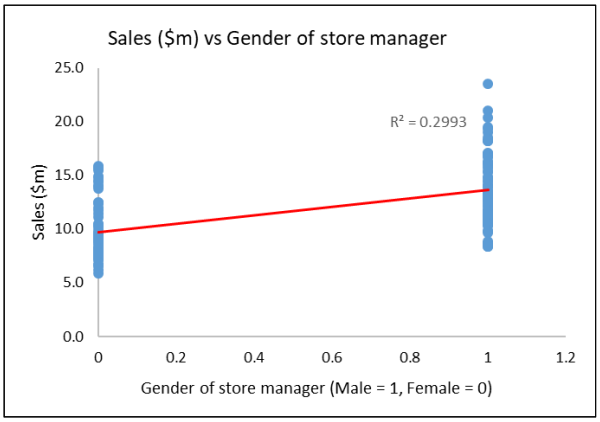


Fig - 2.1.7

Fig - 2.1.8

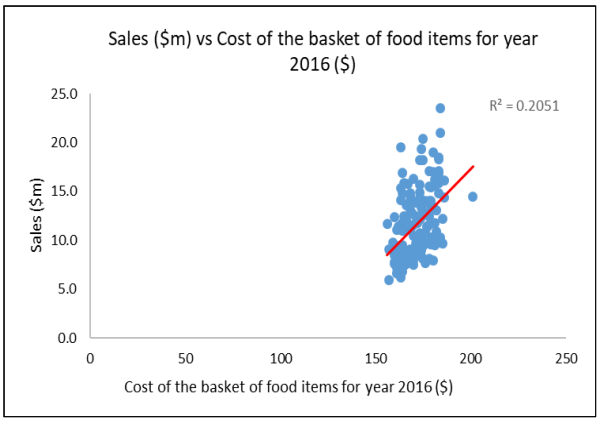


Fig - 2.1.9

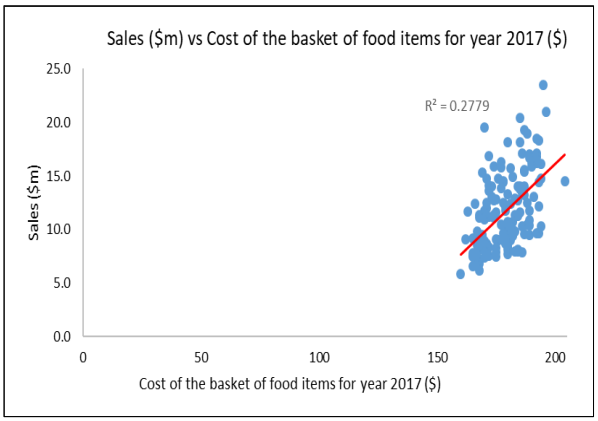


Fig - 2.1.10

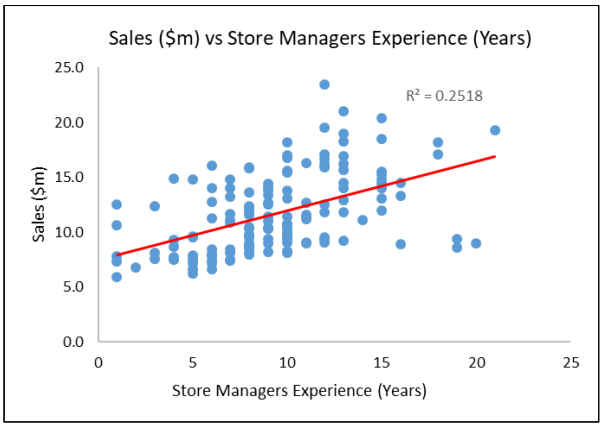


Fig - 2.1.11

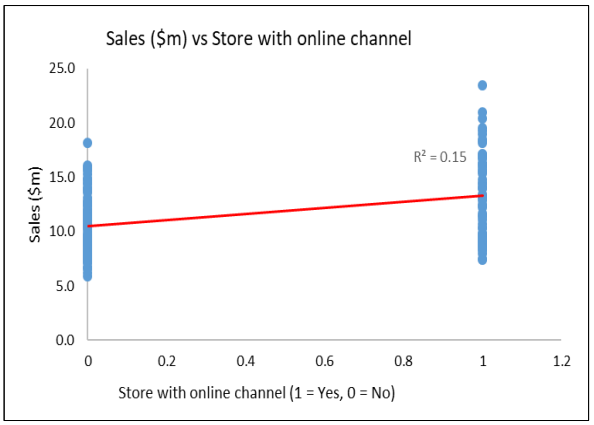


Fig - 2.1.12

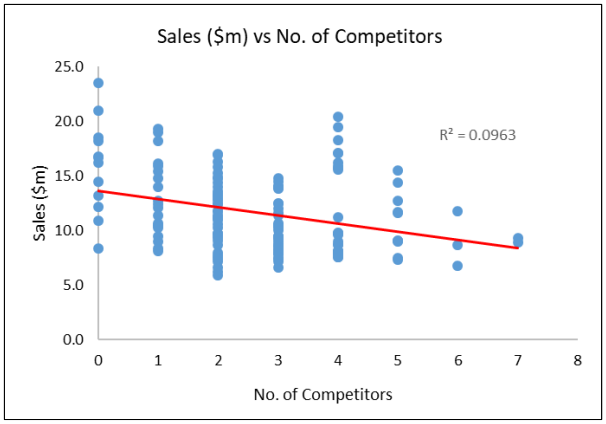


Fig - 2.1.13

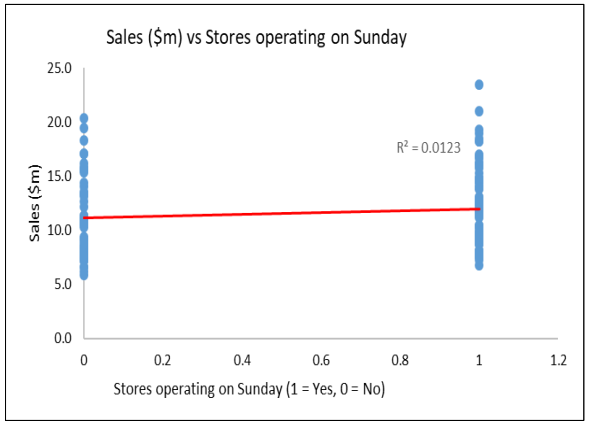


Fig - 2.1.14

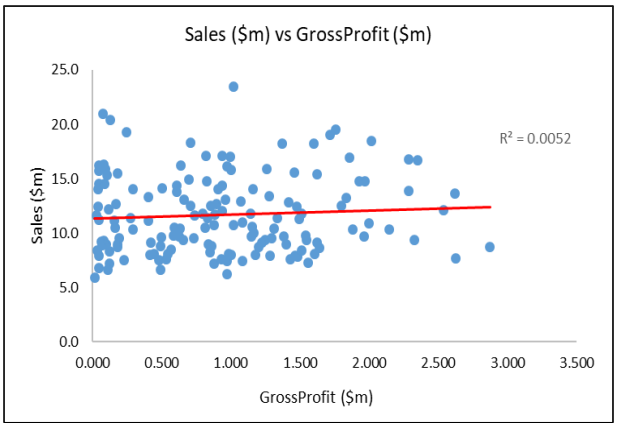


Fig - 2.1.15

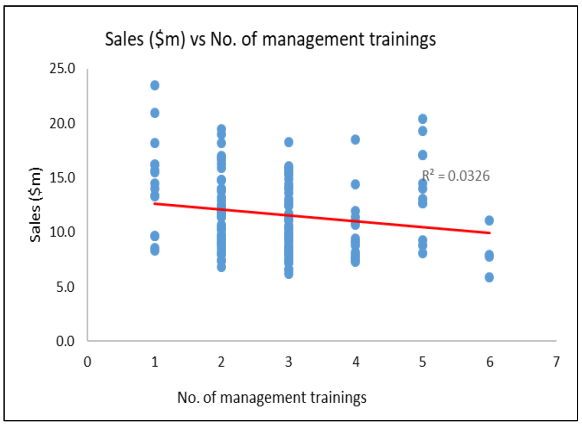


Fig - 2.1.16

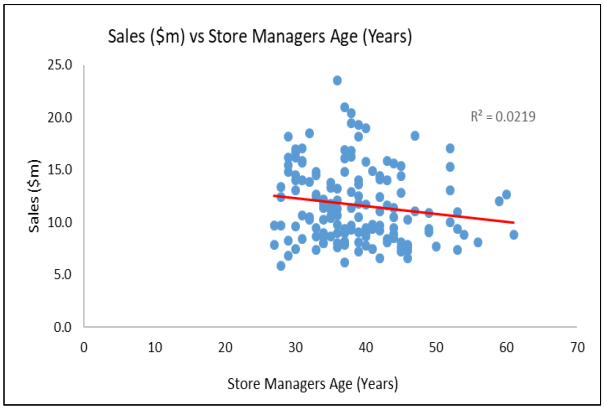


Fig - 2.1.7

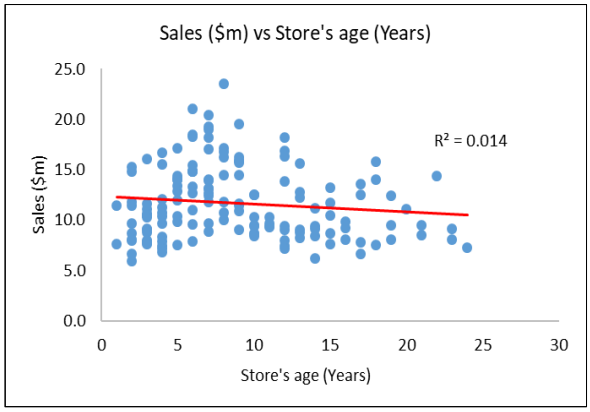
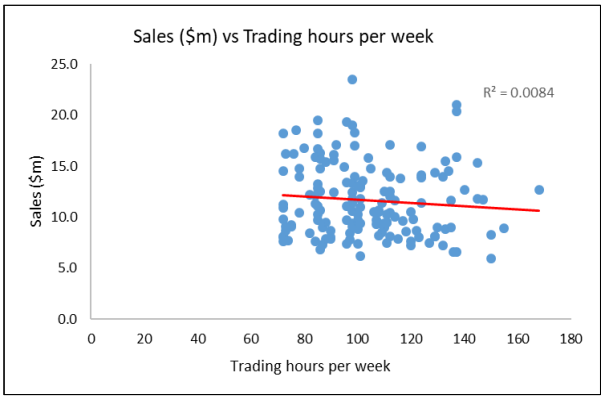


Fig - 2.1.18



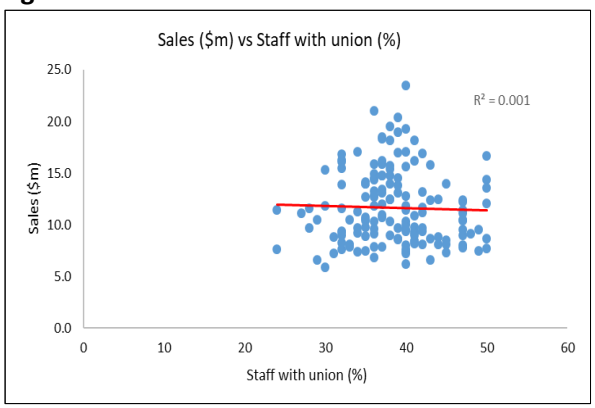


Fig - 2.1.19: Correlation matrix

Correlation matrix			Multicollinearity			Variables having weak linear correlation with Sales													
	Sundays	Mng-Gender	Online Channel	Wages $m	No. Staff	Age (Yrs)	GrossProfit $m	Adv.$'000	Competitors	HrsTrading	Mng-Age	Mng-Exp	Mng-Train	Union%	CarSpaces	Basket: 2016	Basket :2017	%Change Basket	Sales $m
Sundays	1.00																		
Mng-Gender	-0.04	1.00																	
OnlineChannel	-0.10	0.34	1.00																
Wages $m	0.12	0.44	0.25	1.00															
No. Staff	0.12	0.42	0.25	**0.92**	1.00														
Age (Yrs)	0.01	-0.16	-0.02	-0.06	-0.03	1.00													
GrossProfit $m	0.09	0.11	-0.03	0.08	0.00	-0.08	1.00												
Adv.$'000	0.08	0.46	0.32	0.80	0.74	-0.01	0.04	1.00											
Competitors	0.12	-0.24	-0.11	-0.23	-0.21	0.03	-0.28	-0.20	1.00										
HrsTrading	-0.02	-0.14	0.01	-0.12	-0.03	-0.07	-0.46	-0.09	0.37	1.00									
Mng-Age	-0.17	-0.13	-0.27	-0.02	0.05	0.11	0.01	0.02	0.04	0.11	1.00								
Mng-Exp	0.05	0.28	0.50	0.26	0.23	-0.09	-0.09	0.37	-0.06	0.10	0.05	1.00							
Mng-Train	-0.11	-0.14	-0.38	-0.08	-0.11	0.05	-0.06	-0.03	0.13	0.07	0.47	-0.22	1.00						
Union%	-0.02	-0.01	0.04	-0.04	-0.11	0.62	-0.06	-0.04	-0.12	-0.23	-0.03	-0.09	-0.02	1.00					
CarSpaces	0.11	0.37	0.28	0.70	0.76	-0.05	0.02	0.54	-0.17	-0.12	-0.02	0.18	-0.12	-0.12	1.00				
Basket:2016	0.10	0.29	0.14	0.21	0.14	-0.24	0.12	-0.01	-0.33	-0.25	-0.31	0.28	-0.22	0.03	0.16	1.00			
Basket:2017	0.11	0.31	0.19	0.28	0.21	-0.25	0.10	0.07	-0.34	-0.22	-0.30	0.32	-0.24	0.00	0.22	**0.98**	1.00		
%ChangeBasket	0.09	0.26	0.30	0.45	0.43	-0.10	-0.05	0.41	-0.15	0.04	-0.07	0.37	-0.19	-0.12	0.36	0.27	0.43	1.00	
Sales $m	**0.11**	**0.55**	**0.39**	**0.81**	**0.74**	**-0.12**	**0.07**	**0.84**	**-0.31**	**-0.09**	**-0.15**	**0.50**	**-0.18**	**-0.03**	**0.57**	**0.45**	**0.53**	**0.56**	**1.00**

Table 2.3.1: Model comparison

Model	R^2 (%)	Adjusted R^2 (%)	Standard Error (%)	Change in R^2	Change in Adj R^2	Change in Standard Error	Insignificant variables	F Statistic
2.3a	94.05	93.67	0.90	-	-	-	4	246.10
2.3b	94.05	93.72	0.90	0.00%	0.05%	-0.33%	3	278.83
2.3c	94.05	93.76	0.89	0.00%	0.10%	-0.89%	2	320.81
2.3d	94.03	93.78	0.89	-0.02%	0.12%	-0.89%	1	375.35
2.3e	93.98	93.77	0.89	-0.07%	0.11%	-0.89%	0	449.27

Fig - 2.3.1: Model 2.3e regression output (VIF is calculated using Real Statistics Data Analysis Tool)

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.969
R Square	0.940
Adjusted R Square	0.938
Standard Error	0.892
Observations	150

ANOVA 203.171

	df	*SS*	*MS*	*F*	*Significance F*
Regression	5	1785.579	357.116	449.271	0.000
Residual	144	114.462	0.795		
Total	149	1900.042			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*vif*
Intercept	-27.962	1.621	-17.255	0.000	-31.166	-24.759	
Mng-Exp	0.061	0.021	2.856	0.005	0.019	0.104	1.362
No. Staff	0.025	0.008	3.163	0.002	0.009	0.040	2.440
Adv.$'000	0.037	0.002	21.127	0.000	0.034	0.041	2.672
Basket:2016	0.178	0.010	17.772	0.000	0.158	0.198	1.222
%ChangeBasket	36.303	9.184	3.953	0.000	18.150	54.456	1.411

Fig - 2.4.4.1

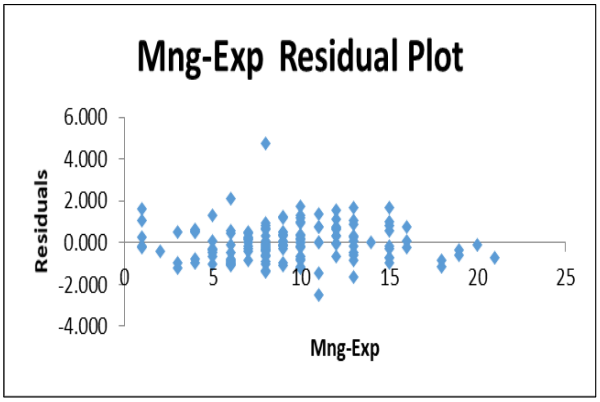


Fig - 2.4.4.2

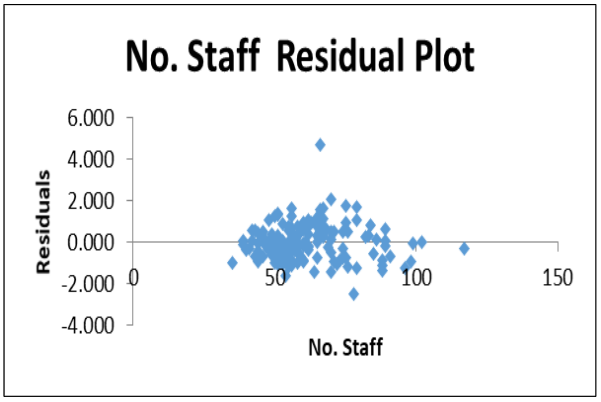


Fig - 2.4.4.3

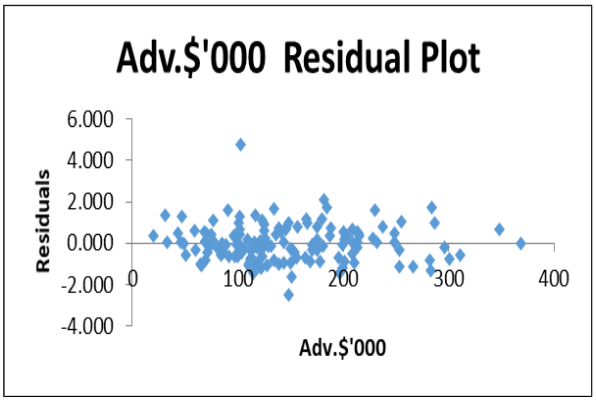


Fig - 2.4.4.4

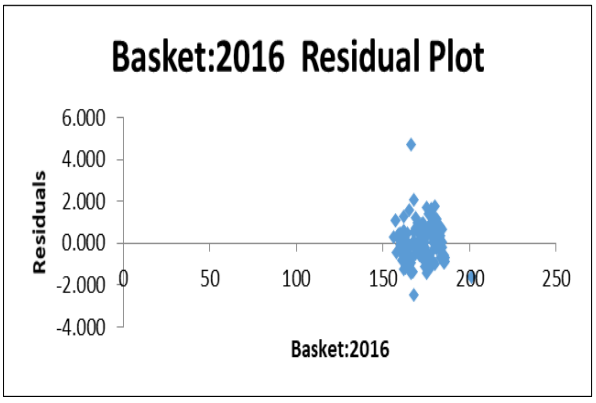


Fig - 2.4.4.5

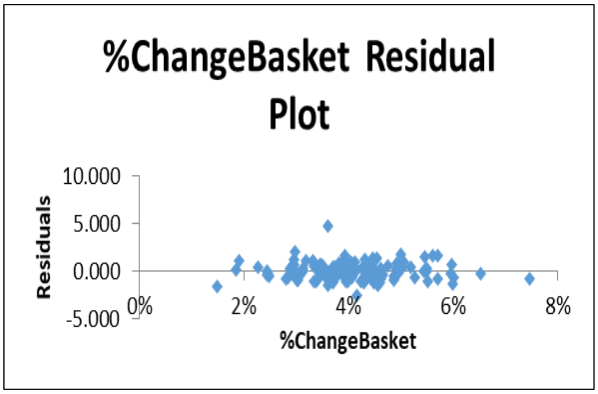


Fig - 2.4.4.6

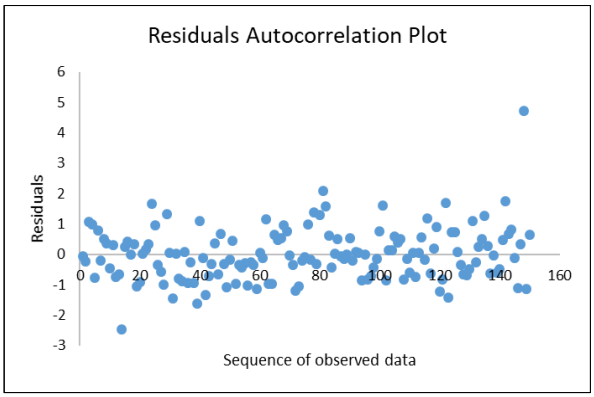


Fig - 2.4.4.7

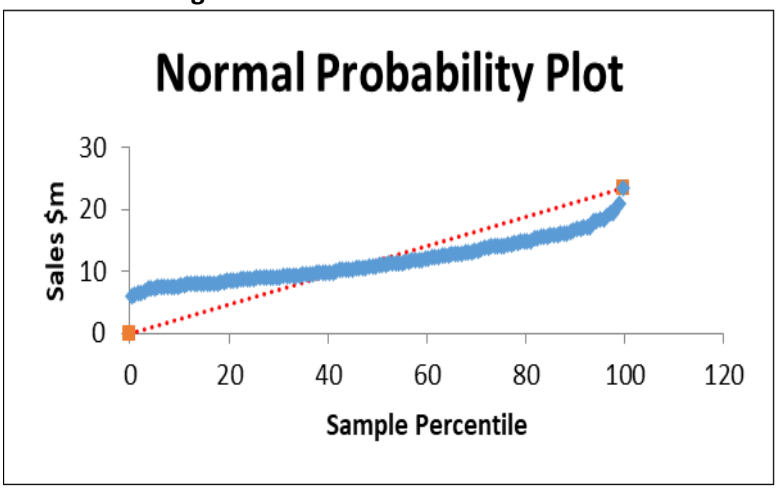


Fig - 2.4.4.8

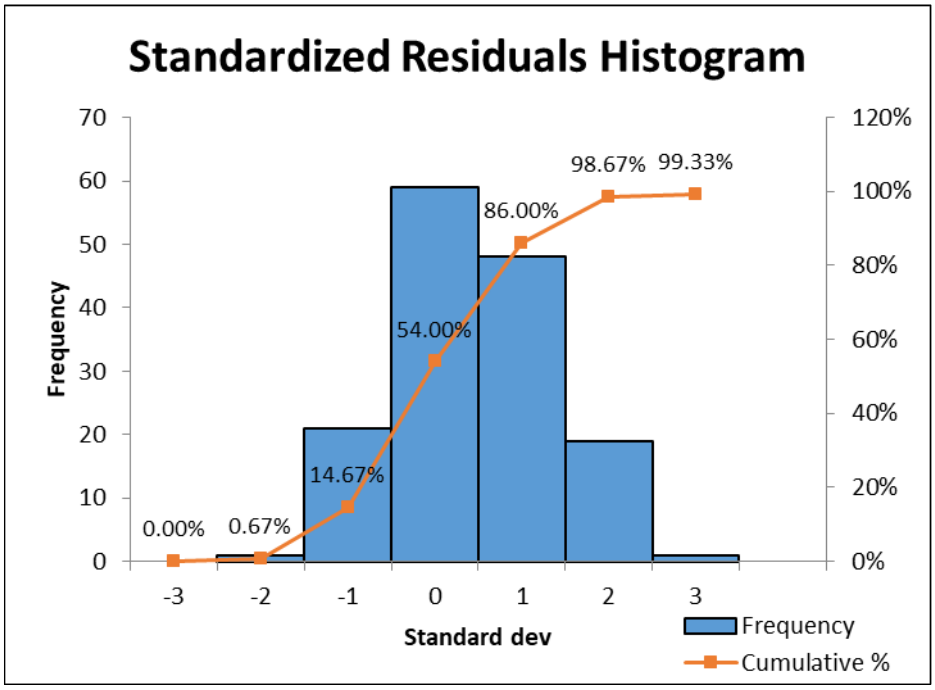


Fig 3.2: Linear Regression output

SUMMARY OUTPUT - REGRESSION WITHOUT INTERACTION TERM

Regression Statistics	
Multiple R	0.344
R Square	0.119
Adjusted R Square	0.107
Standard Error	3.375
Observations	150

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	225.189	112.595	9.882	0.000
Residual	147	1674.852	11.394		
Total	149	1900.042			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	13.006	0.640	20.337	0.000	11.742	14.270
Competitors	-0.784	0.186	-4.209	0.000	-1.152	-0.416
Sundays	1.113	0.578	1.925	0.056	-0.030	2.256

Fig 3.3: Regression output

SUMMARY OUTPUT - REGRESSION WITH INTERACTION TERM

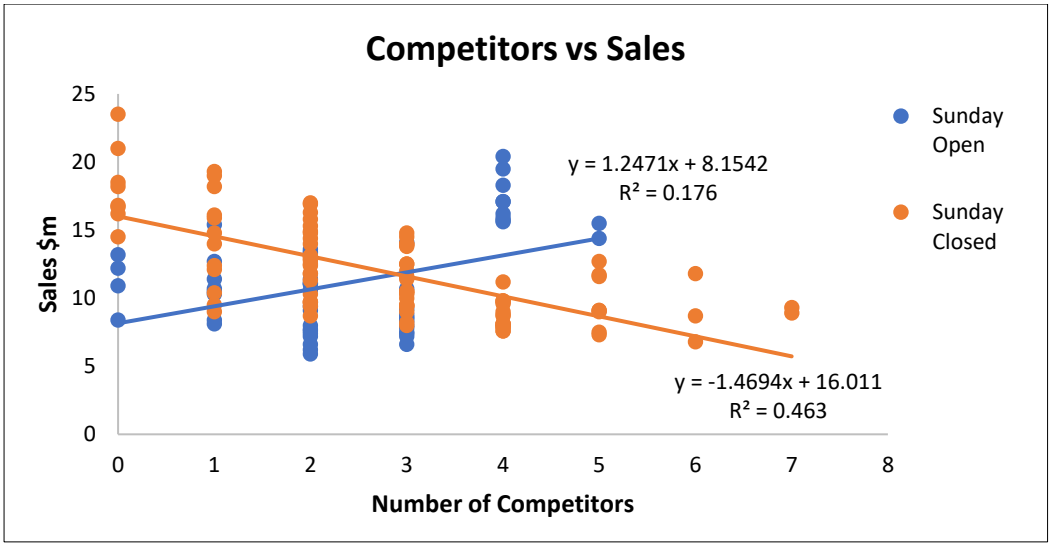
Regression Statistics	
Multiple R	0.599
R Square	0.359
Adjusted R Square	0.346
Standard Error	2.888
Observations	150

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	682.237	227.412	27.264	0.000
Residual	146	1217.805	8.341		
Total	149	1900.042			

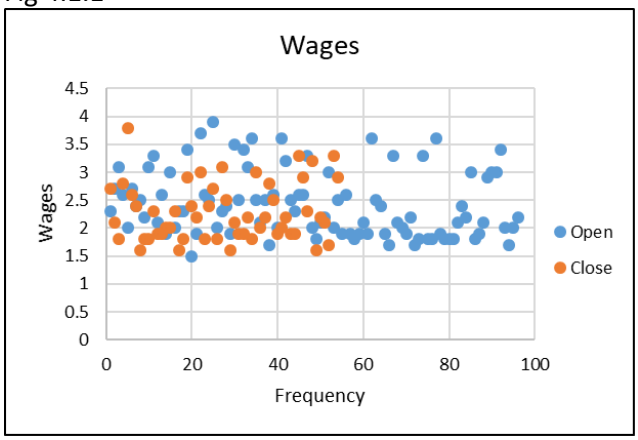
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	8.154	0.854	9.549	0.000	6.467	9.842
Competitors	1.247	0.317	3.930	0.000	0.620	1.874
Sundays	7.857	1.037	7.579	0.000	5.808	9.906
Sun_Comp	-2.717	0.367	-7.402	0.000	-3.442	-1.991

Fig 3.4: Scatter plot between competitors and Sales with effect of Sunday



4.1 Pre-model diagnosis-

Fig 4.1.1



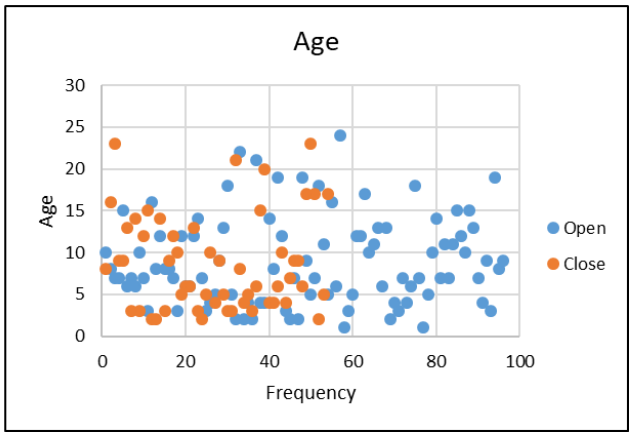
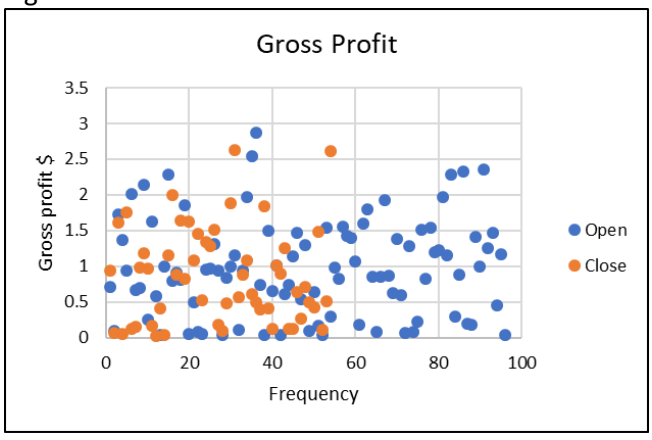


Fig 4.1.2



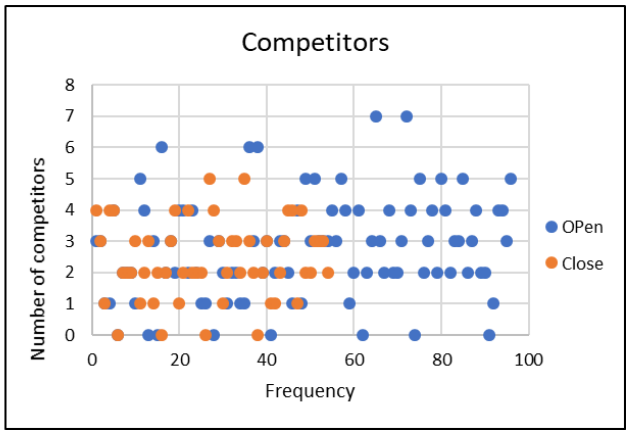
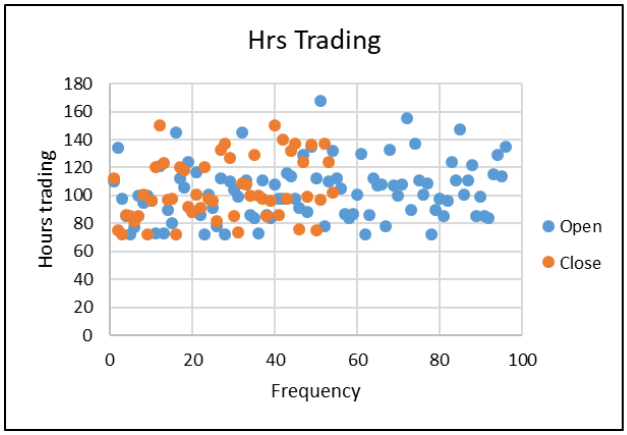


Fig 4.1.3



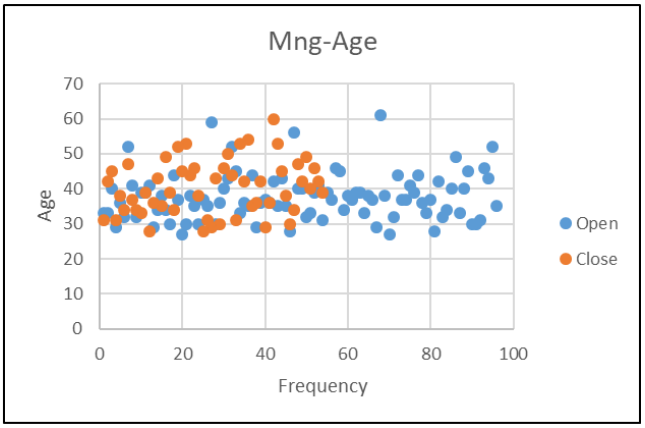
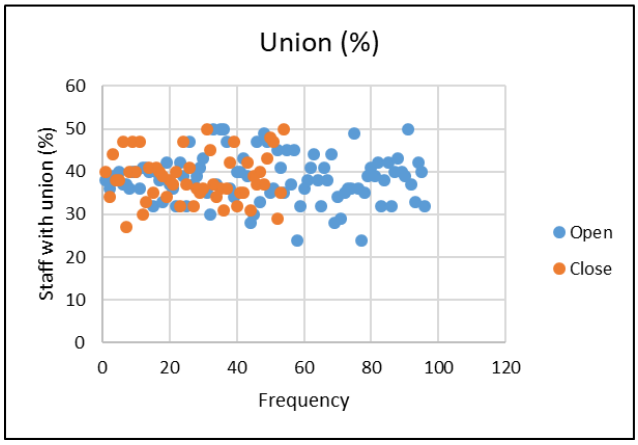


Fig 4.1.4



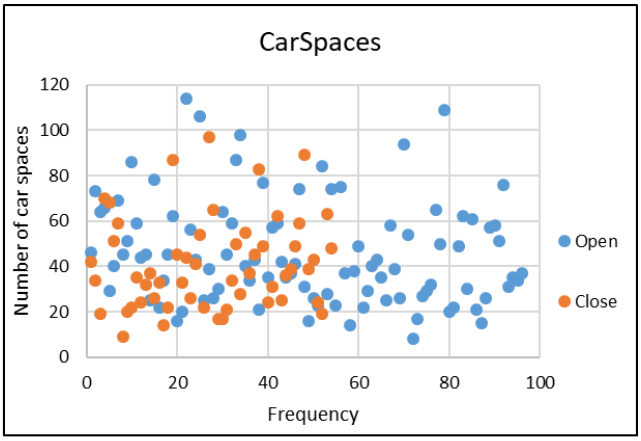
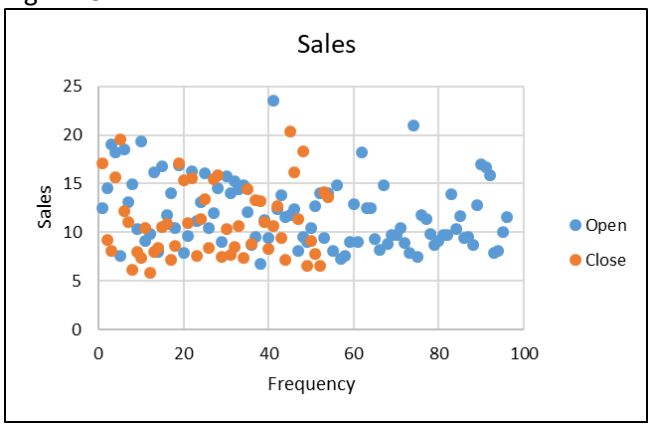


Fig 4.1.5



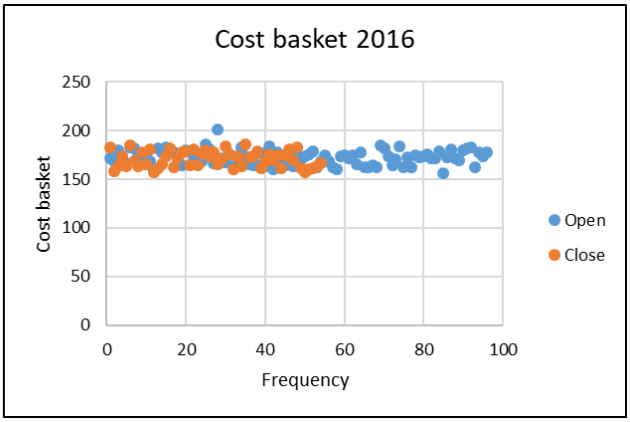
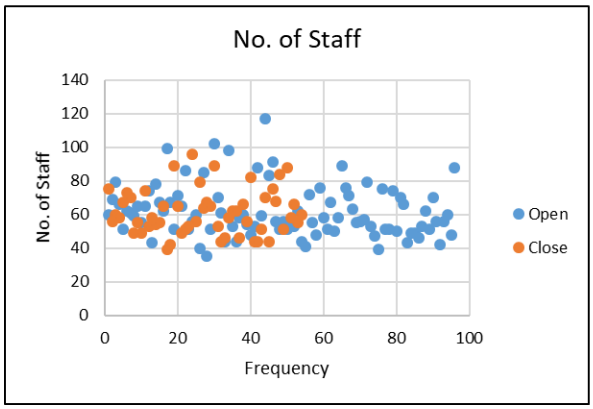


Fig 4.1.6



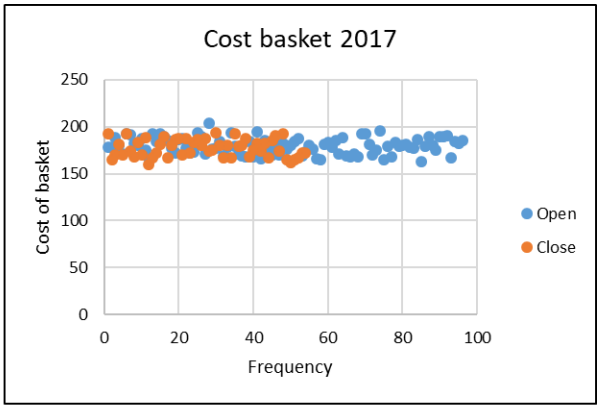
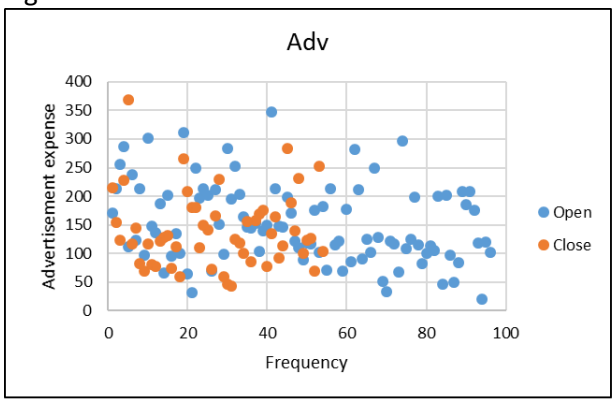


Fig 4.1.7



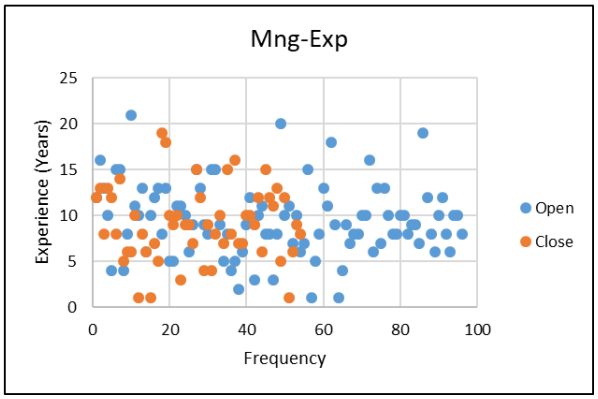
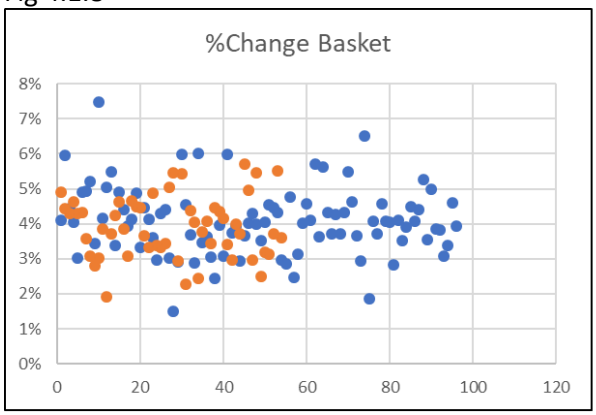


Fig 4.1.8



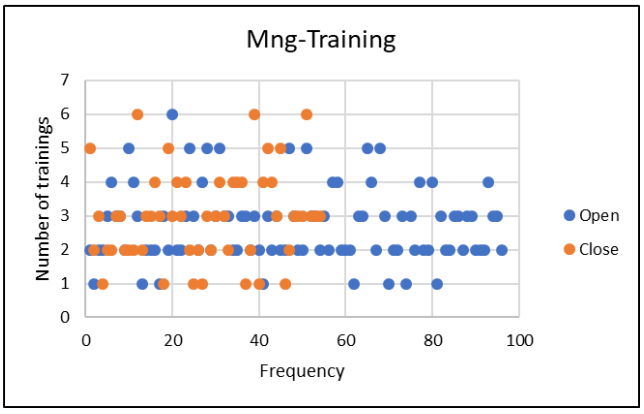


Fig 4.1.9

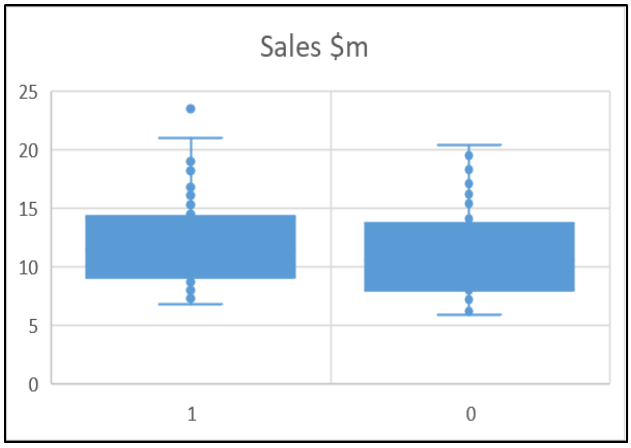
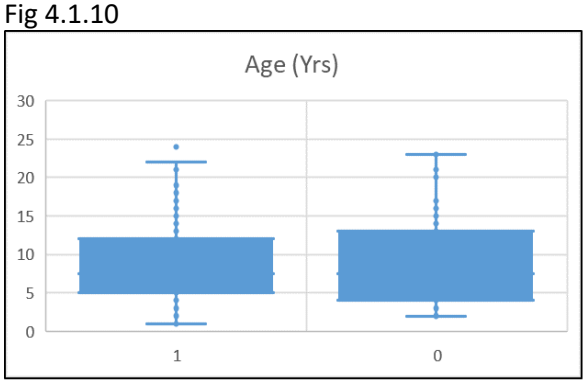


Fig 4.1.10



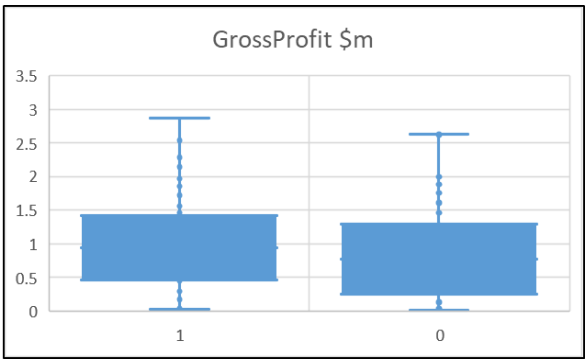
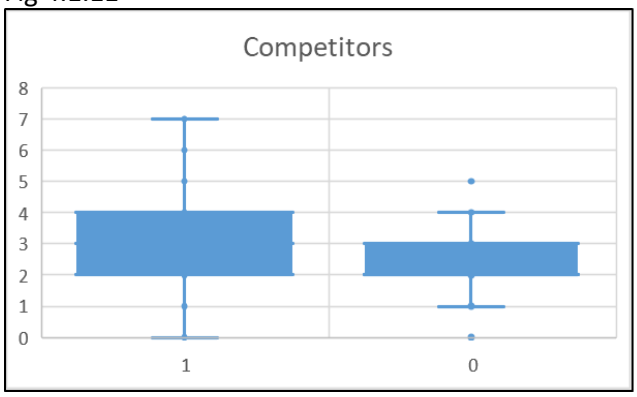


Fig 4.1.11



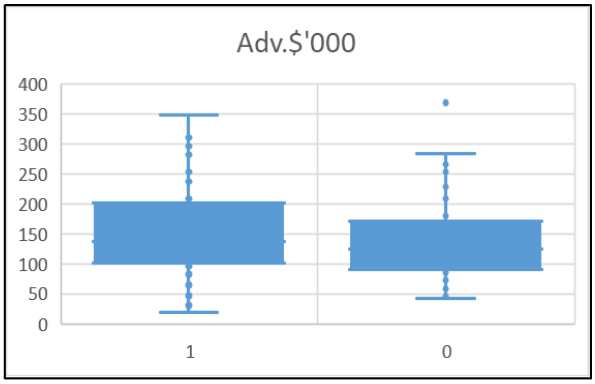


Fig 4.1.12

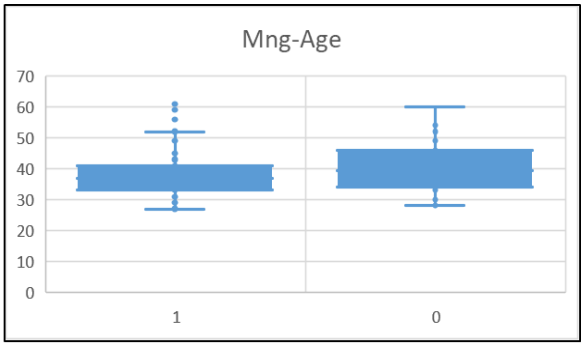
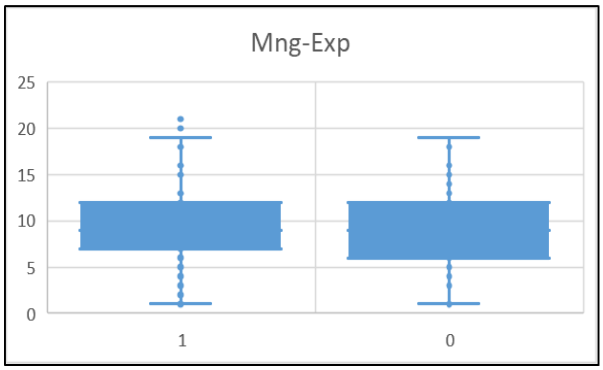


Fig 4.1.13



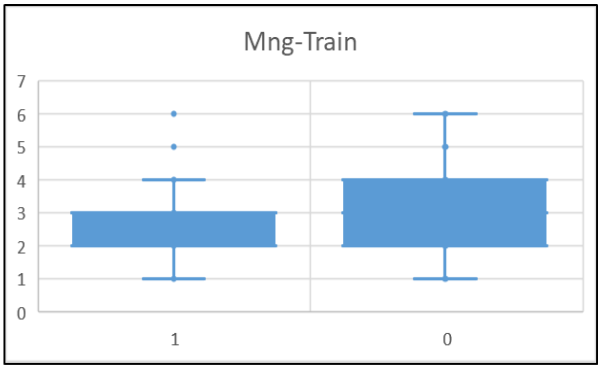
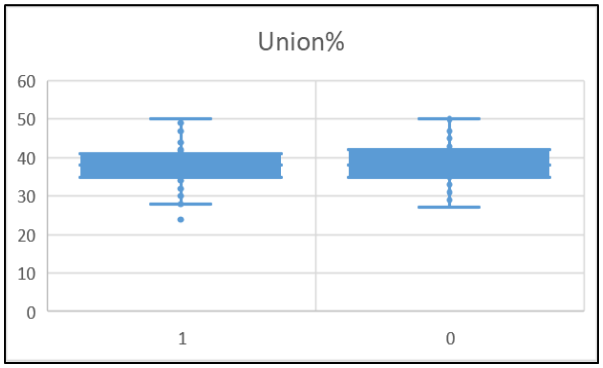


Fig 4.1.14



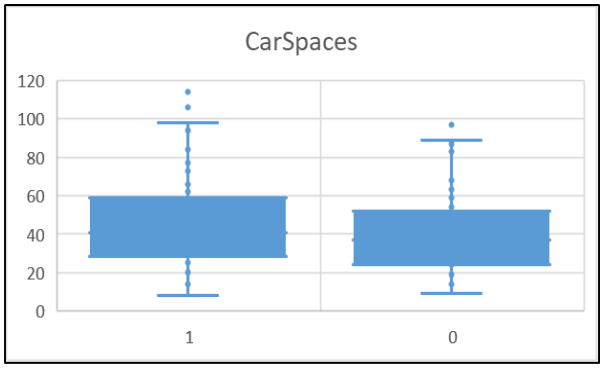
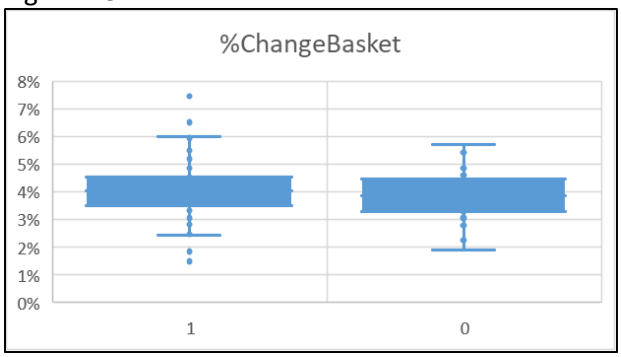


Fig 4.1.15



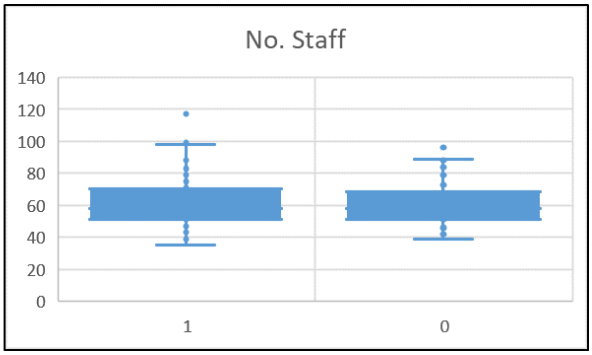


Fig 4.2 - Logistic regression output

LL0	-98.013	Classification Table			
LL1	-85.398				
			Suc-Obs	Fail-Obs	
Chi-Sq	25.230	Suc-Pred	84	28	112
df	6	Fail-Pred	12	26	38
p-value	0.000		96	54	150
alpha	0.05				
sig	yes	Accuracy	0.875	0.481	0.733
R-Sq (L)	0.129	Cutoff	0.5		
R-Sq (CS)	0.155				
R-Sq (N)	0.212	***Comparing accuracy rate with hit ratio***			
		PCC hit ratio = $p^2 + (1-p)^2$		0.539	
Hosmer	158.067	Standard (rule of thumb)		0.625	
df	148				
p-value	0.271				
alpha	0.05				
sig	no				

	coeff b	*s.e.*	*Wald*	*p-value*	*exp(b)*	*lower*	*upper*	*%Change in Odds*
Intercept	0.234	1.421	0.027	0.869	1.264			
OnlineChannel	-1.526	0.499	9.348	0.002	0.217	0.082	0.578	-78.26%
No. Staff	0.037	0.015	6.089	0.014	1.037	1.008	1.068	3.72%
GrossProfit $m	0.633	0.317	3.988	0.046	1.883	1.012	3.504	88.28%
Competitors	0.351	0.144	5.923	0.015	1.421	1.071	1.886	42.12%
Mng-Age	-0.096	0.029	11.213	0.001	0.908	0.859	0.961	-9.17%
Mng-Exp	0.114	0.055	4.208	0.040	1.120	1.005	1.249	12.02%

Fig 4.3 ROC Curve

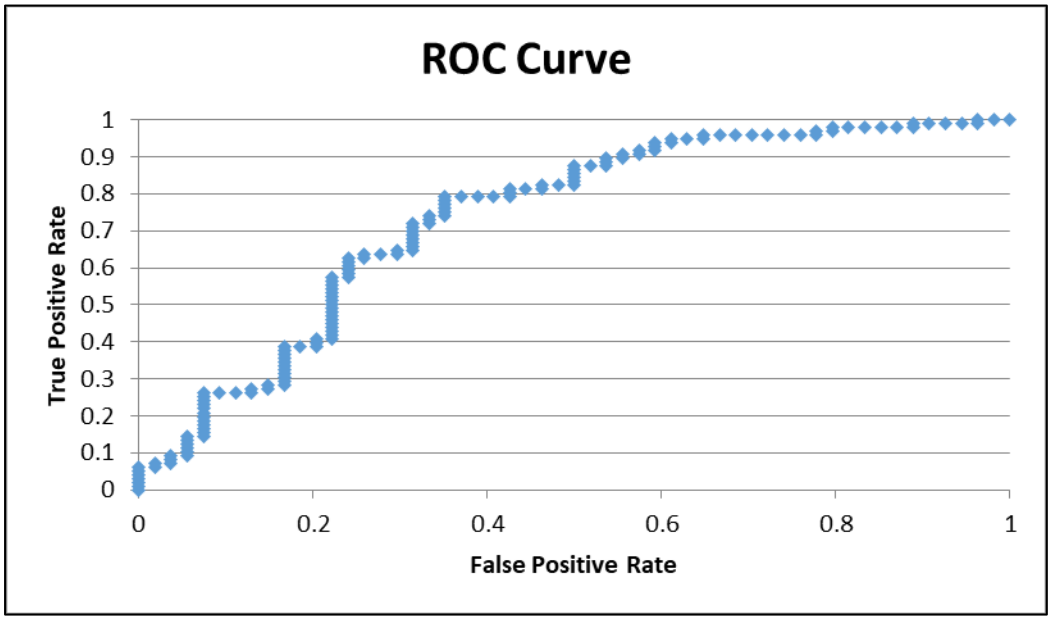


Fig - 5.1 - Logistic regression output for likelihood of store opening an online sale channel

LL0	-101.707	Classification Table			
LL1	-68.323				
			Suc-Obs	Fail-Obs	
Chi-Sq	66.768	Suc-Pred	43	16	59
df	3	Fail-Pred	19	72	91
p-value	0.000		62	88	150
alpha	0.05				
sig	yes	Accuracy	0.694	0.818	0.767
R-Sq (L)	0.328	Cutoff	0.5	18.2%	
R-Sq (CS)	0.359				
R-Sq (N)	0.484	***Comparing accuracy rate with hit ratio***			
		PCC hit ratio = p^2 + (1-		0.515	
Hosmer	120.769	Standard (rule of thur		0.625	
df	135				
p-value	0.804				
alpha	0.05				
sig	no				

	coeff b	*s.e.*	*Wald*	*p-value*	*exp(b)*	*lower*	*upper*	*%Change in Odds*
Intercept	0.417	1.303	0.102	0.749	1.518			
Mng-Gend	0.970	0.428	5.126	0.024	2.637	1.139	6.104	163.70%
Mng-Age	-0.129	0.035	13.346	0.000	0.879	0.820	0.942	-12.12%
Mng-Exp	0.376	0.074	25.955	0.000	1.456	1.260	1.683	45.65%

Fig - 5.2 - ROC curve of Logistic regression output for likelihood of store opening an online sale channel

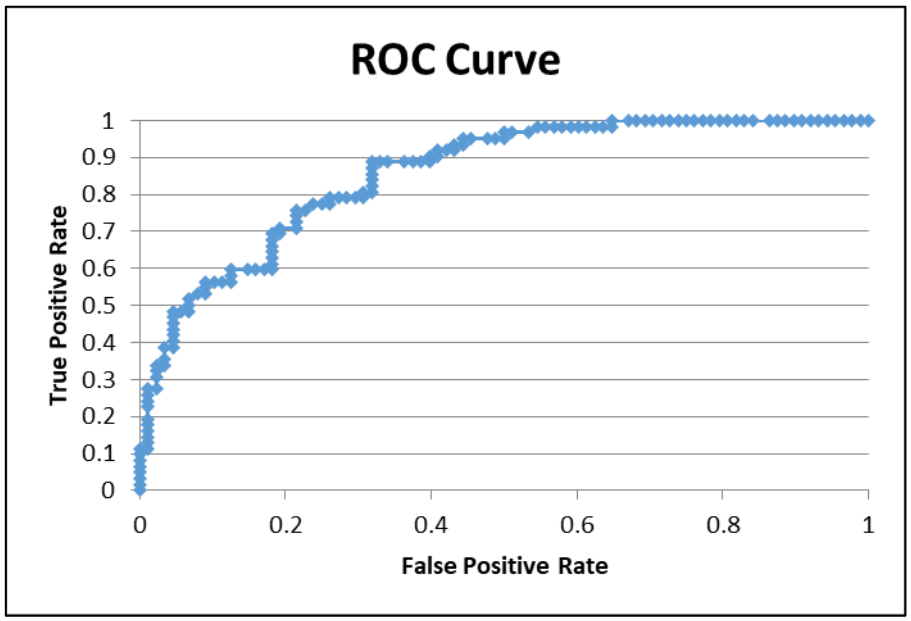


Fig 6 - Sales forecast for 2019-Q2 to 2020-Q1