Table of Contents

[Introduction 2](#_Toc61887)

[1. Provide an overall summary of two variables 3](#_Toc61888)

[a. Sales 3](#_Toc61889)

[b. Online Sales Channel 3](#_Toc61890)

[2. Potential factors that may influence sales 4](#_Toc61891)

[2.1 An appropriate statistical technique could be used here to identify a list of possible factors 4](#_Toc61892)

[2.2 Build a model to estimate Sales 4](#_Toc61893)

[2.3 Interaction Effect 6](#_Toc61894)

[3. Opening an online sales channel 7](#_Toc61895)

[3.1 To calculate the probability of stores 8](#_Toc61896)

[3.2 Visualizing the probability 8](#_Toc61897)

[4. Time-Series Model 8](#_Toc61898)

[Conclusion 9](#_Toc61899)

[Appendix 10](#_Toc61900)

# Introduction

The objective of this report is to provide a model that will estimate Furphy sale across each channel. We are using a random sample of 150 stores which is being used for the analysis. The report compares the online channel sales against the total sales and looks at the average sales by the stores. It identifies the potential factors that may influence sales and builds the regression model to draw the analogies. The report also underlines the relation between the number of competitors and sales by using the interaction effect. It highlights the factors that can influence the store to open an online sales channel. The report also highlights the construction of a forecasting model to predict the sales for the next quarters. This empowers you to bring the shrewd arrangements that could affect the future business arrangements.

# 1. Provide an overall summary of two variables

## a. Sales

* The average sales for the Furphy is 11.65 million dollars. The lowest (minimum) and the highest (maximum) sales for the Furphy is 5.9 million and 23.5 million dollars. We can see that 25% of the store have sales between 5.6-8.8 million dollars and 25% of the stores have sales more than 14 million dollars. *[Appendix A.1.a]*
* The sales difference between the lowest and the highest is 17.6 million dollars. Moreover, there is a 3.57 average spread away from the mean score of 11.65. The tail of the distribution is towards the right and hence is positively skewed. The peak of the distribution says that the tails are longer and fatter, and its central peak is higher and sharper. The data provided indicates that there is quite a bit of variation in the average sales of the Furphy stores. There were few stores who had average sales bit odd when compared with the rest and can be considered as an unusual case. *[Appendix A.1.a]*

## b. Online Sales Channel

* The average sales for the Furphy online channel is 13.30 million dollars. The lowest (minimum) and the highest (maximum) sales for the Furphy online channel is 7.4 million and 23.5 million dollars. We can see that 25% of the online store have sales between 7.4-9.55 million dollars and 25% of the online stores have sales more than 16.275 million dollars. *[Appendix A.1.b]*
* The sales difference between the lowest and the highest is 16.1 million dollars. Moreover, there is a 3.95 average spread away from the mean score of 11.65. The tail of the distribution is towards the right and hence is positively skewed. The peak of the distribution says that the tails are shorter and thinner, and its central peak is lower and broader. The data provided indicates that there is quite a bit of variation in the average sales of the Furphy online stores. *[Appendix A.1.b]*
* We can see from the distribution of channel bar chart, that the number of online channel stores are 62 compared to the offline channel store of 88. *[Appendix A.1.b]*

We can now compare the online sales with the overall total sales using a box plot. From the box plot we can see that, average sales for the online channel ($13.30m) is higher than the overall total sales ($11.65 m). Online store channel has more stores having sales below than the average sales of $13.30 m, whereas when we look at the broader picture, we can say that overall stores are having sales more than the average sales of $11.65 m. It means for total sales, more values lie in the 3rd quartile which is Q3, whereas for the online channel sales more value lie in the 1st quartile that is Q1. In few cases, we have more sales done by the offline stores. But overall, we can say that online stores have performed better than the offline stores. *[Appendix A.1.c]*

# 2. Potential factors that may influence sales.

### 2.1 An appropriate statistical technique could be used here to identify a list of possible factors.

Since our target variable is sales and we need to identify the factors that have an influence on the sales, we must use the correlation statistical technique between our target variable sales and the independent variables. We are using correlation because it helps us to measure the strength and the direction of the linear relationships between the pairs of the variables. Correlation value always lie between +1 and -1. Positive value means it has a positive linear relationship and negative means it has a negative linear relationship. r value (correlation) being 1 gives the perfect relationship. Relationship can be established using the below method:

* If r>0.7, it depicts they are strongly correlated.
* If 0.4 <r< 0.7, it depicts they are moderately relationship.
* If r<0.3, it depicts they are weakly correlated.

We can see the correlation value from the chart of each variable with the sales variables. Wages, Adv.$’000 are highly correlated. %ChangeBasket, OnlineChannel are moderately correlated. Sundays, Union% are weakly correlated. There are variables like Mng\_Train, Mng-Age, Competitiors and Age which have a negative linear relationship with the Sales *[Appendix A.2.1]*.

After performing the correlation, we can see the relationship between pairs of variables. This helps us to understand which variables are important and would help us in identifying the potential factors. *[Appendix*

*A.2.1]*

The table on the left highlights the variables with the low correlation values. But we are going to remove the variables which have multi collinearity. Due to high multicollinearity between the no. staff and wages, we are removing the no. staff has it has less correlation with the sales variable. Due to high multicollinearity between the Basket:2016 and Basket:2017, we are removing the Basket:2106 as it has less correlation with the sales variable. Hence the table on the right shows the final variables which will be used to build the model. *[Appendix A.2.1]*

### 2.2 Build a model to estimate Sales.

To build the model, we need to follow the steps as below:

* From our correlation analysis, we saw there were independent variables which had high correlation between them. To remove the variables, we check the multi collinearity and removed the variables which had less correlation value with the Sales like in our case we removed the No. Staff and the Baasket:2016. Hence the final variables used for the regression model are shown in the table. *[Appendix A.2.1]*
* The reason, we did not remove the variables which had low correlation value is because, we don’t want to end up removing those variables which could have been useful. Hence, to avoid that risk, we are going to follow the backward elimination method and remove the variables one at a time and see how our regression model performs.
* Now, we will apply our independent and the target variables to the build the model using the regression analysis. The regression output highlights the relationship between the target and the independent variables. Here, we need to check the null hypothesis and the alternative hypothesis using the p-value. Using the significance F value, we can talk about the predicting power of the model and using the individual p-values of each independent variables we can see if they are statistically significant. *[Appendix A.2.2]*
* Assumptions that are to be made for the model is:
  + Ho: B1 (Null Hypothesis) - Testing the population slope to determine whether there is a relationship between Sales (target variable) and independent variables.
  + Ha: B1≠0 (Alternative Hypothesis) – If p-value > alpha (0.05), do not Reject Ho and conclude independent variable is not statistically significant in the model.
* Therefore, null hypothesis will be only rejected when our p-value is either 0.05 or greater than that. If the p-value is low, it signifies the importance of the model. Hence, to obtain the best model, we will run or iterate our regression model multiple times, eliminating higher p-values (meaning p-value > alpha) one at a time as they are not statistically significant and lessen the overall predictive power of the model.
* From the first regression result, we can see that based on the significance F value, the model has a great predictive power, but there are many independent variables which have high p-value making them statistically insignificant. Hence, we will remove the %ChangeBasket as it has the highest p-value and re-run the regression model after eliminating this variable.
* Hence, on performing the several iterations of the regression model, removing variables one at a time we build a final model. The order in which the variables are removed are OnlineChannel(pvalue=0.737), Union%(p-value=0.655), Competitiors(p-value=0.500), CarSpaces(p-value=0.497), MngTrain(p-value=0.343), Sundays(p-value=0.233), Mng-Gender(p-value=0.148) and Age (p-value=0.119). We can see that based on our assumptions we are eliminating the variables. We can say all the individual independent variables are individually statistically significant. *[Appendix A.2.2]*
* The final model is build using the remaining variables and the regression output is shown. *[Appendix*

*A.2.2]*

The final regression model builds to estimate sales consists of the independent variables which are shown in the above table. Now, I will discuss about the regression statistics: *[Appendix A.2.2]*

* Multiple R – This is our correlation coefficient and since the value is 0.974 which is very close to 1, we can say that our model has a very strong linear relationship.
* R square – This is our coefficient of determination and tells us about the points falling on the regression line. We can say that 94.9% of variation of y-values around the mean are explained by the x-values. We can also say that 94.9% of the variation in sales can be explained by our regression model. And approximately, 5.1% of the variation in the sales would be explained by other factors not included in the model. Since the value is very high, we can say it is a strong predictive model.
* Standard Error – The final predictive value can have a variation of 0.823.
* Observation – 150 number of observations participated in the analysis.

* There are outliers in the residuals as can be seen from the scatterplots of each individual variables. From the normal probability plot, we can see that nearly all the points fall on the line, except few which can be seen at the edge of the line. They are basically outliers in our data and we can say that data is not approximately normally distributed, or we can say is slightly positively skewed. But overall, there are no apparent problems in the plot and they satisfy the model assumptions. *[Appendix A.2.2]*
* Following interpretations can be made on the independent variables:
  + Coefficient for Wages (0.82) tells us that, assuming no change in other independent variables in the model, for a unit increase in Wages, sales increases by 0.82, on average.
  + Coefficient for Gross Profit (0.22) tells us that, assuming no change in other independent variables in the model, for a unit increase in Gross Profit, sales increases by 0.22, on average.
  + Coefficient for Mng-Age (-0.02) tells us that, assuming no change in other independent variables in the model, for a unit increase in Mng-Age, sales decreases by 0.02, on average.
  + Similarly, we can explain the same for the remaining independent variables.

### 2.3 Interaction Effect

Here, we are going to predict the sales using the number of competitors and stores open on Sundays. When we perform the regression analysis with these independent variables and the interaction term, we get the regression output. *[Appendix A.2.3]*

Our independent variable here is competitors and the moderator variable is Sundays which will be used to predict the dependent variable that is sales. From the regression output, we can interpret this:

Anova: p-value < 0.05 therefore Reject Ho. Overall, the model has some predictive power. All individual independent variables (including the interaction term) are individually significant at p < .05. Significance of the interaction term indicates that Sundays interacts with competitors in predicting sales. In other words, Sundays moderates the relationship between competitors and sales.

At low levels of competitors, open Sundays are on average having more sales than the close Sundays. Competitors positively affect sales across both Sunday groups. At high level of competitors, close Sundays outperform open Sundays in terms of their sales. Compared to open Sundays, close Sundays can have more sales as the number of competitors increases. Overall, Sundays interact with competitors and sales (moderates the relationship between competitors and sales), such that the relationship is significantly stronger for close

Sundays than for the open Sundays. *[Appendix A.2.3]*

# 3. Opening an online sales channel.

The variables involved here are Mng-Gender, Mng-Age, Mng-Exp and OnlineChannel. Since our predictor variable that is the dependent variable is binary, we will use logistic regression to predict the likelihood of opening an online sales channel. Let’s interpret the logistic regression output:

|  |  |  |  |
| --- | --- | --- | --- |
| Classification Table | |  |  |
|  |  |  |  |
|  | Suc-Obs | Fail-Obs |  |
| Suc-Pred | 43 | 16 | 59 |
| Fail-Pred | 19 | 72 | 91 |
|  | 62 | 88 | 150 |
|  |  |  |  |
| Accuracy 0.693548 0.818182 0.766667 | | | |
|  |  |  |  |
| Cutoff | 0.5 |  |  |

Success observations are 62 out of which 43 were correctly classified and the success rate accuracy is 69.35%. Failure observations are 88 out of which 72 were correctly classified and failure rate accuracy is 81.81%. 76.75% is the overall classification accuracy, but since the observations are different for groups we need to check it against the PCC and the rule of thumb. Since this value is greater than above mentioned values, we can say our model is practically significant. We can also use Maximum Chance Criterion, Proportional Chance Criterion and Rule of thumb. We were able to improve the classification for each category beyond complete randomness.

To check if our model was statistically significant we look at the Chi Square. LL0 and LL1 are the sum of squared residuals in the multiple regression and we want to reduce this as much as

|  |  |  |
| --- | --- | --- |
| LL0 | -101.707 |  |
| Chi-Sq | 66.767 | 77 |
| LL1df | -68.3234 | 3 |
| p-value | 2.1E-14 | |
| alpha | 0.05 | |
| sig | yes | |
|  |  | |
| R-Sq (L) | 0.328235 | |
| R-Sq (CS) | 0.359252 | |
| R-Sq (N) | 0.483948 | |
|  |  | |
| Hosmer | 120.7688 | |
| df | 135 | |
| p-value | 0.804438 | |
| alpha | 0.05 | |
| sig | no | |

|  |  |  |
| --- | --- | --- |
| MCC |  | 41.3% |
| PCC |  | 51.5% |
| Rule |  | 64.4% |

possible. Also, LL1 should be less than LL0 which is true in our case. From the chi square we can see p-value is less than alpha, hence we can say that the model is statistically significant and tells us about the predictive power of the model.

|  |  |
| --- | --- |
| R-Sq (L) | 0.328235 |
| R-Sq (CS) | 0.359252 |
| R-Sq (N) | 0.483948 |

R-Sq (L) – We can say 32.8% of the variation in dependent variable can be explained by the variation in the independent variable.

R-Sq (CS) - We can say 35.9% of the variation in dependent variable can be explained

by the variation in the independent variable.

R-Sq (N) - We can say 48.39% of the variation in dependent variable can be explained by the variation in the independent variable.

Now we will discuss about the magnitude and the direction of the relationship. Mng-Gender, Mng-EXp are positively related and Mng-Age is negatively related to the likelihood of opening an online sales channel. 1 unit of increase in positive feeling towards Mng-Gender increase the likelihood of opening an online sales channel by 163.69%. 1 unit of increase in positive feeling towards Mng-Age decrease the likelihood of opening an online sales channel by 12.12%. 1 unit of increase in positive feeling towards Mng-Exp increase the likelihood of opening an online sales channel by 45.64%. *[Appendix A.3.1]*

ROC curve helps us to visualize the ability of our LR model to discriminate the success and failure categories. This line should cover the entire area, but in our case it doesn’t appear to do so, hence we can say the model has not that good discriminatory power. *[Appendix A.3.1]*

### 3.1 To calculate the probability of stores.

Here, we are using Age, Experience and Gender to calculate probabilities of stores to open an online sales channel. We have used age in mid-thirties which in my case is 34,35 and 36. We are varying the experience from 2 years to 16 years. Hence, we have computed probabilities for different genders, for different ages and for different years of experience.

### 3.2 Visualizing the probability

The below graph shows the probability of opening the online channel against the different attributes. We can say that male have a higher probability than females for opening an online sales channel as they outperform them. We can see that as the age is increasing the probability is going down for both the males and females. As the years of experience is increasing the chances are high to open the online sales channel. We can say that Gender and years if experience are positively related with the online sales channel, but age is negatively related. It is better for the organization to recruit the tech-savvy young managers who have more years of experience and are in mid-thirties. It would be better if the recruit manager is male as they have high probability of opening an online sales channel. *[Appendix A.3.2]*

# 4. Time-Series Model

We are using multiplicative technique because we have seasonal components, trend components and irregularities. From the trend line we can see an upward trend and see irregularities. *[Appendix A.4]*

* We are going to quantify our trend components, seasonal components and get rid of irregularities.
* We are going to slice up these 3 components and then quantify them. To identify which moving average techniques to use, we look at our chart and see we have different years and each having 4 quarters in each season, hence we are using the 4 moving average technique. We first find the 4-moving average and then 4 centered moving average and we call it as our baseline.
* Now we quantify our seasonality by calculating the ration (observation/moving average) and it contains both season variations and irregularities. To get rid of irregularities, we do the indexing by using the ratio values. We see for 2014-Q4, the actual observation is below the baseline by 12.32%.
* We now calculate the average ratio of each quarter and get the indexes. Similarly, we calculate it for the other quarters and sum the averages up.
* The sum should be equal to the number of period in each season which in case is 4. Since the sum of values is less than 4, we normalize the averages value for each quarter. We now got the index for our seasonal component.
* Now we will remove the season components from actual values, basically we are deseasonalizing our data. We draw a trendline for the deseasonalized stage and get the equation, which will be used to forecast based on the trend component. Equation is y = 4.1671\*x + 45.308
* Now we do the slice up or decouple the components together to forecast. Now we calculate the error factor that is MAPE which is 4.62% we may commit by using this technique using the historical data.

# Conclusion

Following conclusions can be made based on the key findings of the analysis:

* Average sales for online sales channel is higher than the offline sales channel.
* We used the correlation analysis to identify the potential factors that contribute towards the prediction of sales. We used the multicollinearity and the regression model to build the model that has predictive power. Also, 94.9% of the variance in the sales can be explained by our model.
* When we have low competitors, open Sundays have more sales but as the number of competitor increases, close Sundays have better sales. We can say the interaction effect exists.
* We performed the logistic regression to predict the likelihood of opening an online sales channel. We found that our model is practically and statistically significant and has a predictive power.
* We found that male have high probability than females and if the experience is more than the probability is high. But we saw, as the age increases the probability to open an online sales channel is decreasing.
* We have used the multiplicative technique for forecasting sales as we have seasonality, trends and irregularities. We have used the 4 moving averages and used the indexing by normalizing it to remove the irregularities. From our trend equation, we have computed the forecasting sales.

# Appendix

A.1.a Summary Statistics of Sales.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | *Summary Statistics for Total Sales $m* | | | Mean | 11.65933333 | | Standard Error | 0.291569995 | | Median | 10.95 | | Mode | 8.1 | | Standard Deviation | 3.570988558 | | Sample Variance | 12.75195928 | | Kurtosis | 0.007741155 | | Skewness | 0.744799581 | | Range | 17.6 | | Minimum | 5.9 | | Maximum | 23.5 | | Sum | 1748.9 | | Count | 150 | | Q1 | 8.825 | | Q3 | 14 | | IQR | 5.175 | | LF | 1.0625 | | UF | 21.7625 | | Potential Outlier | Yes | | 1  74  55  18  2  0  10  20  30  40  50  60  70  80  1-6  6-11  11-16  16-21  21-26  Frequency  Total Sales $m  Overall view of sales  Total |

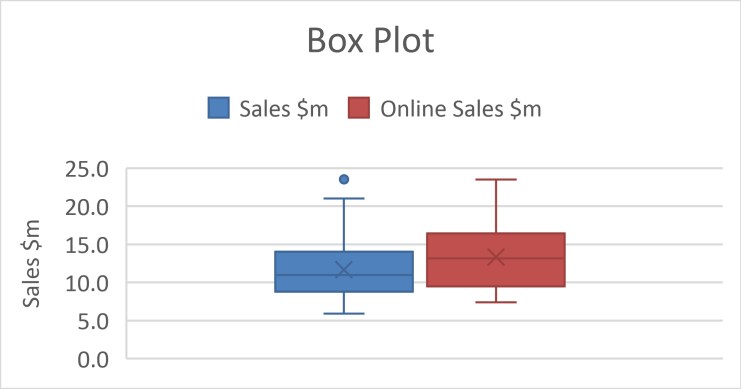
A.1.b Summary Statistics of Online Sales Channel

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | *Summary Statistics for Online Sales Channel $m* | | | Mean | 13.3016129 | | Standard Error | 0.502481899 | | Median | 13.15 | | Mode | 17.1 | | Standard Deviation | 3.95654643 | | Sample Variance | 15.65425965 | | Kurtosis | -0.731711231 | | Skewness | 0.392217938 | | Range | 16.1 | | Minimum | 7.4 | | Maximum | 23.5 | | Sum | 824.7 | | Count | 62 | | Q1 | 9.55 | | Q3 | 16.275 | | IQR | 6.725 | | LF | -0.5375 | | UF | 26.3625 | | Potential Outlier | No | | 88  62  0  20  40  60  80  100  No Online Channel  Online Channel  Distribution of Channel  Total  22  22  16  2  0  5  10  15  20  25  6-11  11-16  16-21  21-26  Frequency  Online Sales $m  Overall view of Sales for  Online Channel  Total |

**Row Labels Count of Sales $m**

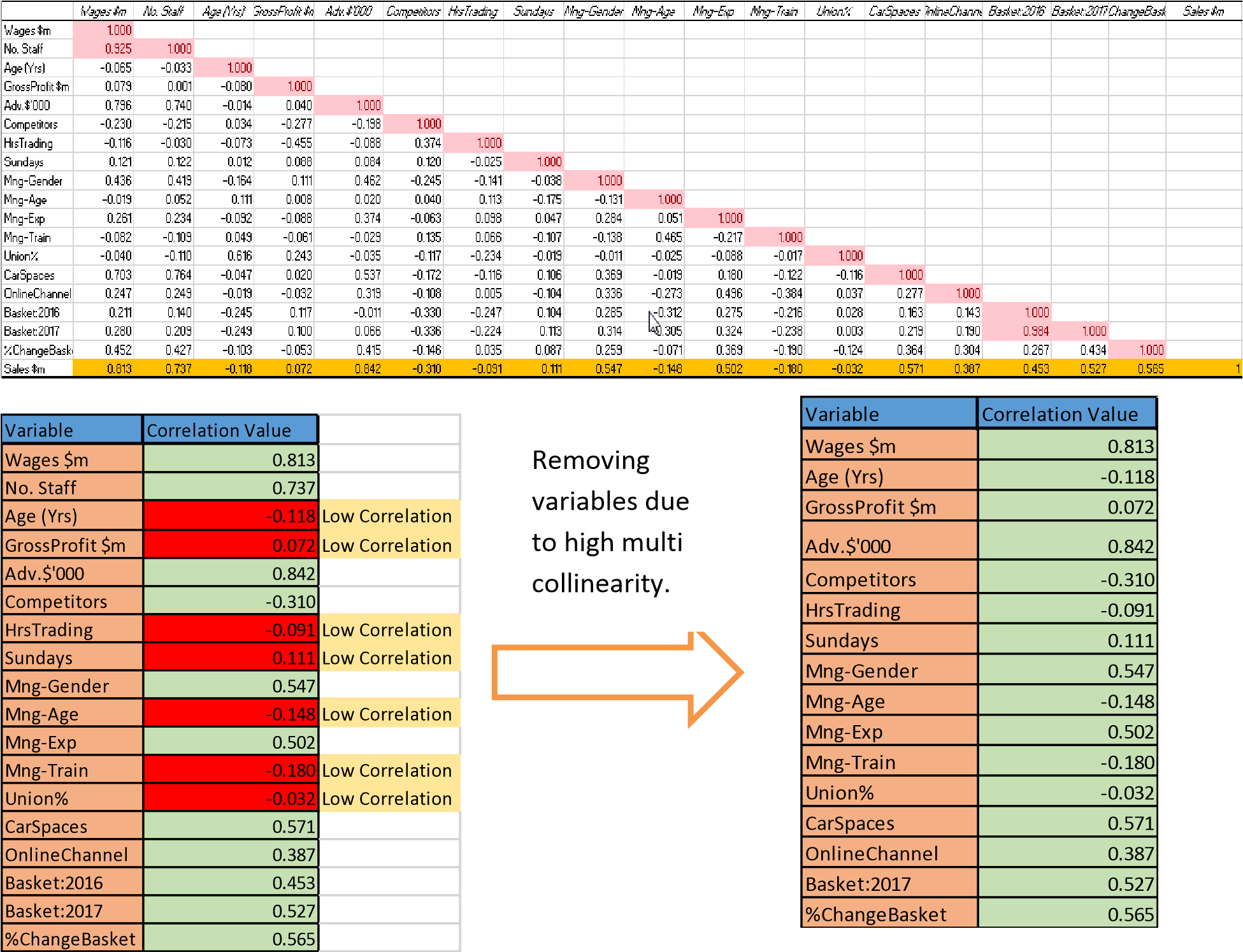
|  |  |  |
| --- | --- | --- |
|  |  |  |
| 6-11 |  | 22 |
| 11-16 |  | 22 |
| 16-21 |  | 16 |
| 21-26 |  | 2 |
| **Grand Total** |  | **62** |

A.1.c Boxplot



A.2.1 Correlation and Multi Collinearity

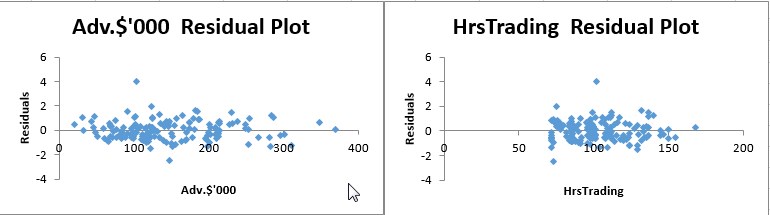
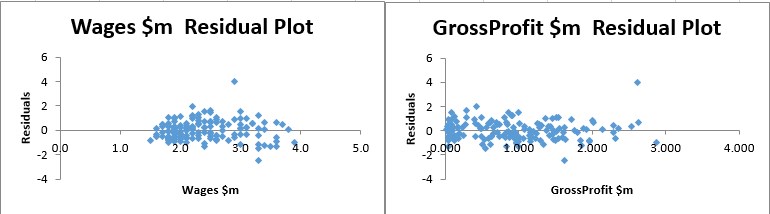
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Correlation of Varaibles with Sales   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | B  B  Onli  M | asket:2017 asket:2016 neChannel  CarSpaces Union% |  |  |  |  | |  |  |  | |  |  | |  |  | |  |  |  | | | Mng-Train Mng-Exp | |  |  |  | | Mng-Age ng-Gender  Sundays HrsTrading | |  |  |  | |  | | | Co  Gros | mpetitors Adv.$'000 sProfit $m | |  |  |  |  | | | Age (Yrs)  No. Staff | |  |  |  |  |   %ChangeBasket  Variables  Wages $m  -0.400 -0.200 0.000 0.200 0.400 0.600 0.800  Correlation Value | 1.000 |

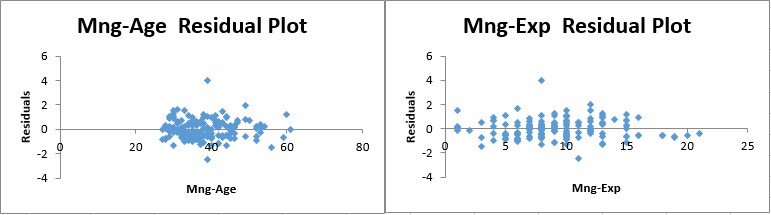


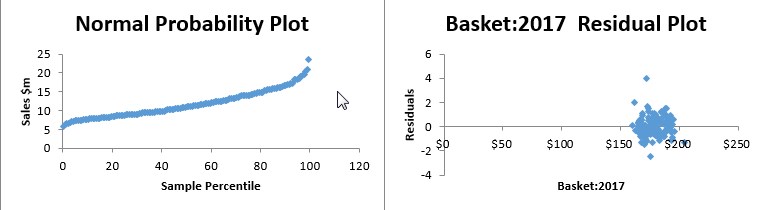
A.2.2 Regression Analysis and Residual Plots

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| SUMMARY OUTPUT | |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| *Regression Statistics* | |  |  |  |  |  |  |  |
| Multiple R | 0.976 |  |  |  |  |  |  |  |
| R Square | 0.952 |  |  |  |  |  |  |  |
| Adjusted R Square | 0.946 |  |  |  |  |  |  |  |
| Standard Error | 0.826 |  |  |  |  |  |  |  |
| Observations | 150.000 |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |  |  |  |
| Regression | 16.000 | 1809.215 | 113.076 | 165.579 | 0.000 |  |  |  |
| Residual | 133.000 | 90.827 | 0.683 |  |  |  |  |  |
| Total | 149.000 | 1900.042 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
| Intercept | -27.579 | 2.188 | -12.604 | 0.000 | -31.907 | -23.251 | -31.907 | -23.251 |
| Wages $m | 0.730 | 0.255 | 2.865 | 0.005 | 0.226 | 1.233 | 0.226 | 1.233 |
| Age (Yrs) | 0.028 | 0.018 | 1.528 | 0.129 | -0.008 | 0.063 | -0.008 | 0.063 |
| GrossProfit $m | 0.258 | 0.123 | 2.089 | 0.039 | 0.014 | 0.502 | 0.014 | 0.502 |
| Adv.$'000 | 0.036 | 0.002 | 17.761 | 0.000 | 0.032 | 0.040 | 0.032 | 0.040 |
| Competitors | -0.037 | 0.054 | -0.676 | 0.500 | -0.143 | 0.070 | -0.143 | 0.070 |
| HrsTrading | 0.019 | 0.004 | 4.740 | 0.000 | 0.011 | 0.027 | 0.011 | 0.027 |
| Sundays | -0.178 | 0.154 | -1.150 | 0.252 | -0.483 | 0.128 | -0.483 | 0.128 |
| Mng-Gender | 0.197 | 0.168 | 1.172 | 0.243 | -0.135 | 0.529 | -0.135 | 0.529 |
| Mng-Age | -0.022 | 0.012 | -1.764 | 0.080 | -0.046 | 0.003 | -0.046 | 0.003 |
| Mng-Exp | 0.058 | 0.024 | 2.403 | 0.018 | 0.010 | 0.106 | 0.010 | 0.106 |
| Mng-Train | -0.057 | 0.071 | -0.805 | 0.422 | -0.197 | 0.083 | -0.197 | 0.083 |
| Union% | -0.007 | 0.018 | -0.395 | 0.693 | -0.042 | 0.028 | -0.042 | 0.028 |
| CarSpaces | 0.003 | 0.005 | 0.637 | 0.525 | -0.006 | 0.012 | -0.006 | 0.012 |
| OnlineChannel | -0.065 | 0.188 | -0.348 | 0.728 | -0.437 | 0.306 | -0.437 | 0.306 |
| Basket:2017 | 0.170 | 0.012 | 14.584 | 0.000 | 0.147 | 0.194 | 0.147 | 0.194 |
| %ChangeBasket | 1.310 | 9.313 | 0.141 | 0.888 | -17.111 | 19.731 | -17.111 | 19.731 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | | SUMMARY OUTPUT | |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  | | *Regression Statistics* | |  |  |  |  |  |  |  | | Multiple R | 0.974 |  |  |  |  |  |  |  | | R Square | 0.949 |  |  |  |  |  |  |  | | Adjusted R Square | 0.947 |  |  |  |  |  |  |  | | Standard Error | 0.823 |  |  |  |  |  |  |  | | Observations | 150.000 |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  | | ANOVA |  |  |  |  |  |  |  |  | |  | *df* | *SS* | *MS* | *F* | *Significance F* |  |  |  | | Regression | 7.000 | 1803.758 | 257.680 | 380.029 | 0.000 |  |  |  | | Residual | 142.000 | 96.284 | 0.678 |  |  |  |  |  | | Total | 149.000 | 1900.042 |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  | |  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* | | Intercept | -27.828 | 1.876 | -14.833 | 0.000 | -31.537 | -24.119 | -31.537 | -24.119 | | Wages $m | 0.827 | 0.219 | 3.784 | 0.000 | 0.395 | 1.259 | 0.395 | 1.259 | | GrossProfit $m | 0.223 | 0.113 | 1.975 | 0.050 | 0.000 | 0.446 | 0.000 | 0.446 | | Adv.$'000 | 0.036 | 0.002 | 19.423 | 0.000 | 0.033 | 0.040 | 0.033 | 0.040 | | HrsTrading | 0.017 | 0.004 | 4.419 | 0.000 | 0.009 | 0.024 | 0.009 | 0.024 | | Mng-Age | -0.022 | 0.010 | -2.172 | 0.031 | -0.042 | -0.002 | -0.042 | -0.002 | | Mng-Exp | 0.060 | 0.021 | 2.895 | 0.004 | 0.019 | 0.101 | 0.019 | 0.101 | | Basket:2017 | 0.171 | 0.010 | 17.500 | 0.000 | 0.151 | 0.190 | 0.151 | 0.190 | | |  |  | | --- | --- | | Variable | Correlation Value | | Wages $m | 0.813 | | GrossProfit $m | 0.072 | | Adv.$'000 | 0.842 | | HrsTrading | -0.091 | | Mng-Age | -0.148 | | Mng-Exp | 0.502 | | Basket:2017 | 0.527 | |

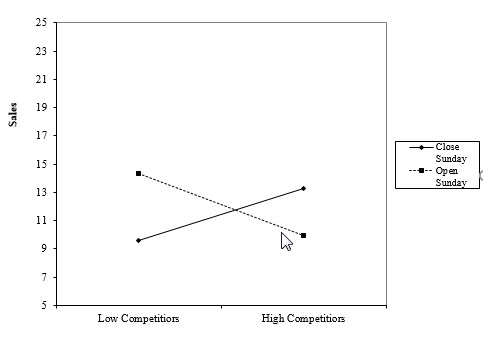






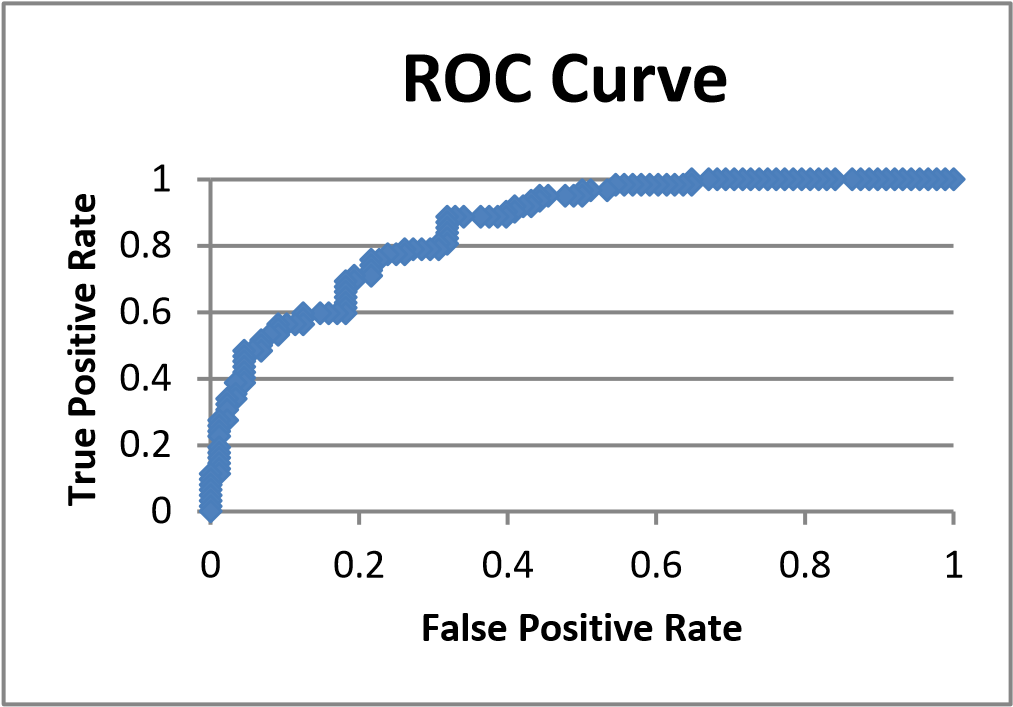
A.2.3 Interaction Effect

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| SUMMARY OUTPUT |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| *Regression Statistics* | |  |  |  |  |  |  |  |
| Multiple R | 0.599 |  |  |  |  |  |  |  |
| R Square | 0.359 |  |  |  |  |  |  |  |
| Adjusted R Square | 0.346 |  |  |  |  |  |  |  |
| Standard Error | 2.888 |  |  |  |  |  |  |  |
| Observations | 150.000 |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |  |  |  |
| Regression | 3.000 | 682.237 | 227.412 | 27.264 | 0.000 |  |  |  |
| Residual | 146.000 | 1217.805 | 8.341 |  |  |  |  |  |
| Total | 149.000 | 1900.042 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
| Intercept | 8.154 | 0.854 | 9.549 | 0.000 | 6.467 | 9.842 | 6.467 | 9.842 |
| Competitors | 1.247 | 0.317 | 3.930 | 0.000 | 0.620 | 1.874 | 0.620 | 1.874 |
| Sundays | 7.857 | 1.037 | 7.579 | 0.000 | 5.808 | 9.906 | 5.808 | 9.906 |
| Competitors\*Sundays | -2.717 | 0.367 | -7.402 | 0.000 | -3.442 | -1.991 | -3.442 | -1.991 |

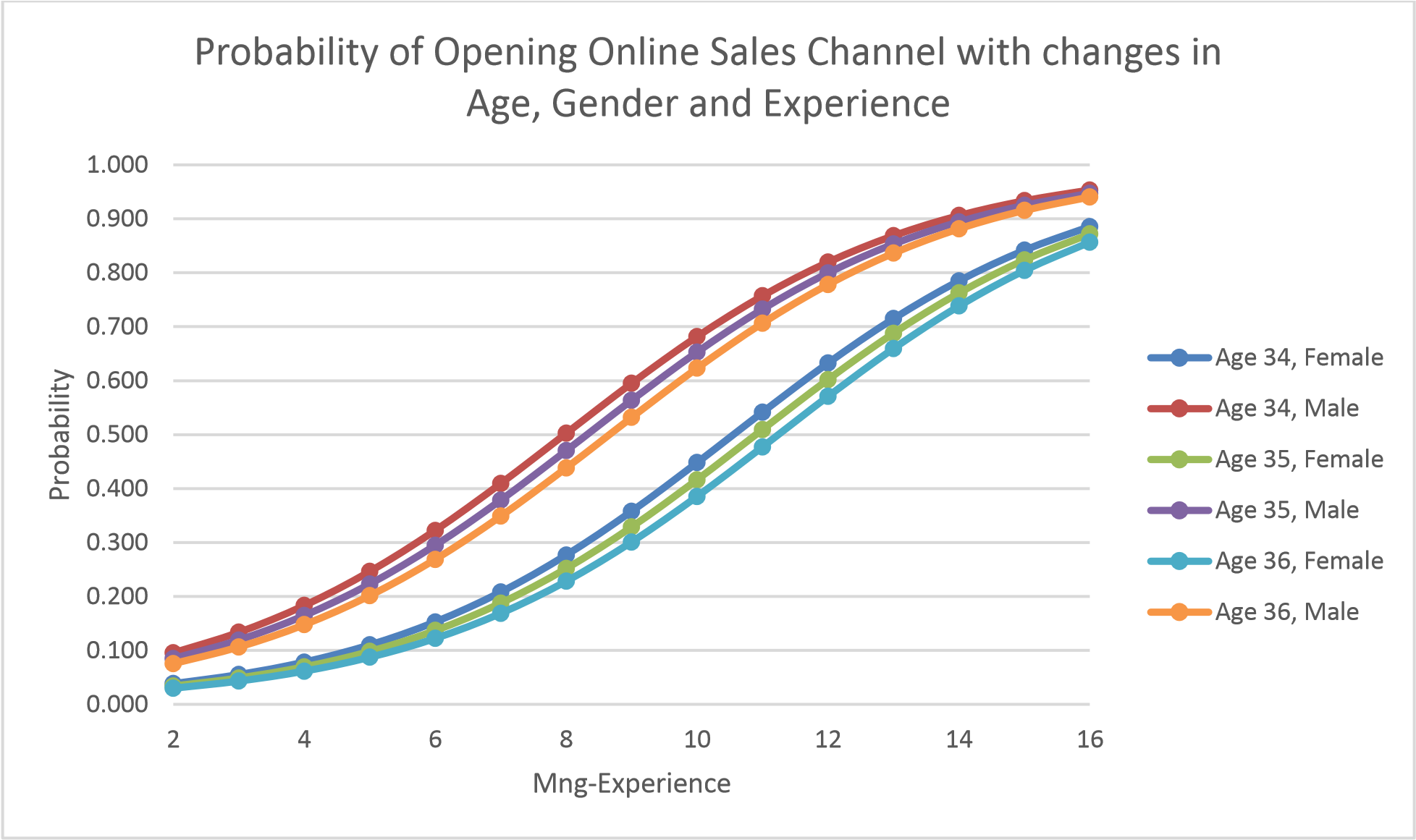


A.3.1 Logistic Regression

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | coeff b | s.e. |  | Wald | p-value | exp(b) | lower | upper | %change in odd = (exp(b)-1)\*100 |
| Intercept | 0.417 |  | 1.303 | 0.102 | 0.749 | 1.518 |  |  |  |
| Mng-Gender | 0.970 |  | 0.428 | 5.126 | 0.024 | 2.637 | 1.139 | 6.104 | 163.696 |
| Mng-Age | -0.129 |  | 0.035 | 13.346 | 0.000 | 0.879 | 0.820 | 0.942 | -12.120 |
| Mng-Exp | 0.376 |  | 0.074 | 25.955 | 0.000 | 1.456 | 1.260 | 1.683 | 45.646 |



A.3.2 Probability Curve



A.4 Forecasting

