**Logo, company name

Description automatically generated**

Odette School of Business, University of Windsor

Odette Consulting Group Project Management

BSMM – 8720 (Section 2): Data Analytics & Project Management

Summer 2022

Prof. Gokul Bhandari

August 2, 2022

**Group 10**

Vighnesh Thakur (110084731)

**Executive Summary**

Consulting firm offers services such as review and planning, computer software and hardware evaluations and selection, and custom software development. The Odette Consulting Group (OCG) is a similar project management consulting group located in Windsor, Ontario. The consulting group focuses on the customer satisfaction to expect the business growth. The not known information about consulting group is demand and supply. Consulting firm needs to take a more conscious look at their competition and human resource management policies. For tomorrow’s consulting firm, the areas such as recruitment, training and development, and incentives are particularly important (Graubner et al.,2003). The research questions provide information about what is important to the business leaders. The management can understand factors that affect most to the client satisfaction based on exploratory analysis. The exploratory analysis identified multicollinearity between multiple independent variables, which were remove during the data modelling process. Predictive analysis utilized various independent variables to predict the dependent variables such as client satisfaction and potential future partnership. The predictive analysis explains linear and logistic regression with different combinations of independent variables. The linear regression uses expertise, implementation, recommendation, consulting fee, customer type, industry type, firm size, region, and partnership to predict satisfaction score. Three models were created for linear regression. Model 1 provided R-Square value of 74%, model 2 accounted for 82%, and finally model 3 provided slight decrease in R-square value. Therefore, we used model 2 as our final model. The logistic regression uses expertise, implementation, recommendation, consulting fee, and satisfaction to predict potential future partnership with OCG. The logistic regression gave 87% accuracy while predicting the outcome variable.

**Literature Review**

First, the literature review explains known information in the area of research. Consulting firm usually offer services such as review and planning, computer software and hardware evaluations and selection, and custom software development. The Odette Consulting Group (OCG) is a similar project management consulting group located in Windsor, Ontario. It conducted a random survey to assess the satisfaction of 200 clients. The ultimate goal of consulting group is to increase revenue. The industry experts use various metrics such as customer satisfaction, likelihood of recommendation, loyalty score and share-of-wallet to predict sales growth. The likelihood to recommend business has been widely adopted by managers as a measure of customer mindset and predictor of sales growth (Baehre et al., 2022). There are some contrary results which indicate that recommend intention alone will not suffice as a single predictor of customer’s future loyalty behaviour. The use of multiple indicator model performs better in predicting customer satisfaction (Timothy et al. 2007). The survey results include information about customer type, industry, firm size, region, ability to manage client portfolio, innovative project management, responsiveness, expertise, competitive consulting fees, communication skills, ability to implement project, client satisfaction, likelihood of recommending to others, and potential partnership with OCG. The results obtained about variables from survey can be utilized to predict client satisfaction.

Second, the consulting group should be focused on the customer loyalty and satisfaction to expect the business growth. Both practitioners and academics understand that customer loyalty and satisfaction are linked inextricably. Although loyal customers are most typically satisfied, satisfaction does not universally translate into loyalty. A conclusion from investigative analysis explains that loyalty cannot be achieved or pursued as a reasonable goal by many providers because of the nature of product category or consumer disinterest. For some firms, satisfaction is the only feasible goal for which they should strive; thus, satisfaction remains a worthy pursuit among the consumer marketing community (Richard, 1999).

Third, the not known information is about information consulting, developments, and trends. The rapid evolution and spread of information technology and how it has created a new set of managerial tasks. The consulting companies should enhance their profitability based on factors like speed, efficiency, quality results, and value creations (Basil et al., 1997). The given variables lack information about efficiency, quality of results, and additional value creation. Also, it is a better to look over next 3 to 5 year about how to manage Odette Consulting Group. On the supply side, competition in the consulting industry has been growing as many companies have entered the market. On the demand side, clients are becoming selective. Consulting firm needs to take a more conscious look at their competition and human resource management policies. For tomorrow’s consulting firm, the following areas such as recruitment, training and development, incentives, and HR administration are particularly important (Graubner et al.,2003). Thus, the information about individual working on the project, experience of employee working in the project, training and development of employee, incentives and motivation can be included to better understand the actual situation of Odette Consulting Group’s projects.

Fourth, there are different measurement tools for IT consulting firms. In an attempt to measure the perceived quality of consulting services, the researchers developed a diagnostic tool of SERVQUAL based on solid theoretical background, which can specifically be applied to the IT consulting service sector. It is composed of 6 dimensions of reliability, responsiveness, assurance, empathy, process and education, and the former four are derived from IT SERVQUAL and the latter two are added to consider the characteristics of IT consulting service. The research indicated that six different areas in the measurement tool are significantly related with level of the customer satisfaction (Yoon et al., 2004).

**Possible Research Questions**

The Odette Consulting Group has recently conducted a random survey of its 200 clients. The following research questions can be answered from the analysis of the survey data:

What factors contribute most to the client satisfaction?

What is the average customer satisfaction score by industry type (IT and non-IT)?

What is the average customer satisfaction score by firm size (fewer than 50 employees and more than 50 employees)?

What is the average customer satisfaction score by region (Ontario and Alberta)?

What percentage of customers are likely to recommend OCG to others?

What is the ratio of customers who would consider OCG as a potential future consulting partner?

What is the overall average score for client satisfaction?

Can we interpret that the overall client satisfaction leads to likelihood of recommending OCG to others and potential future partnership with OCG?

What are the different variables that can be utilized to understand business performance?

What are the areas of improvement that can increase the satisfactions of a client?

What benchmarks do we need to establish for business growth?

What should be the priority for this business unit during next fiscal year?

What is unique about Odette Consulting group?

What are the steps to increase repeat business rate?

Can you provide any strategic recommendation for enterprise cost reduction?

How to predict client satisfaction for a particular client?

**Data Pre-processing**

Data pre-processing is very prevalent in the real world because data in most cases are not found in a streamlined format, they must be turned into a format that is more understandable. Before using the dataset, we investigated it thoroughly to understand how to pre-process it. We went through the following steps, as stated by Chauhan( 2019) which seemed appropriate for our dataset:

1. **Checking the missing values:** We checked the missing values using the “MATCH” function to see if it returns any N/A error, but there were not any.
2. **Identifying variables:** The variables in our model can be broadly defined as nominal and ordinal variables.
3. **Encoding the categorical variables:** Interpreting categorical variables by using them directly in our model is not intuitive. Therefore, to make the results more interpretable, we make dummies for each category within the categorical variables. However, the number of dummies is one less than the total number of categories for a particular variable, and then the results are interpreted in comparison to the base state (which is the variable value for the category for which the dummy variable is not made).
4. **Spelling check:** We checked for any spelling errors by clicking on the Review tab and clicking on spelling. There were no spelling errors found either.

The quality of the data was great since there were no errors or issues that we had to face while processing it. Given how the data was organized and labelled clearly, which aided in our understanding of the variables.

**Data Analysis**

**Exploratory Analysis**

In this section of the report, graphics and visualizations were used for exploration and analysis of the data set containing 200 clients. We performed several sub-group analyses to answer the various possible research questions as enlisted above. In answering those research questions, we would be enabled to have a better understanding of the data set. We looked at the following data analysis parameters: Correlation, Variance, Pivot Tables, Conditional Formatting, Box Plots, Scatter Plots and Outliers. In the processes, we noticed several relationships between and among various variables and will give examples in this section.

A simple conditional formatting revealed several observations that were important to our analysis. Regardless of the company type (IT or non-IT), the company with the highest expertise tends to have its clients more satisfied. In addition, the higher the expertise a company has, the higher the likelihood of recommendation. For example, an IT company with an expertise rating of 7.8/10, had 7.6/10 satisfaction rate and 7.9/10 likelihood to be recommended to other potential clients *(Figure 8).* The high expertise rate correlating with high rates of both satisfaction and recommendation rates could be said of the non-IT companies. Companies that have higher expertise tend to know how to do their job well. They tend to understand better what is needed of them and what the clients need. Services that come from people with high expertise will most likely be of high standards and probably long-lasting solution. If deliverables meet the expectations and the needs of the clients, the clients will be contented with the services and that will make them happy and satisfied. If clients are satisfied with services and that happens constantly over a period, they would be more likely to recommend the consulting group’s services to other people, and this seems to be true going by the model built from the data set presented.

From the analysis it can be observed that, the longer a company stays with the Consulting Group, the higher the rate of satisfaction. Through the pivot tables we created, we looked at the customer type, which has three subcategories – Those who have been with OCG for less than one year, those who have been with OCG between one and five years and those who have been OCG for more than five years. The companies that have been with the consulting group for more than five years have posted higher average satisfaction ratings at 7.853 *(Figure 9)*. Furthermore, those who have stayed with the consulting company for less than one year, posted an average rate of satisfaction of 5.729 *(Figure 9)*. Although on a scale of 10, this rating is above average, it is relatively low when put against the other periods. For example, for the companies that have stayed with the consulting group for a period of between one and five years, they have a rating of 7.294 *(Figure 9)* which is well higher than that of those who have stayed with the consulting group for less than a year and its lower than the satisfaction rating of the customers that have stayed with the consulting group for more than five years. One of the explanations that can be attributed to the high ratings for satisfaction as customers stay with the consulting group more is that, over time, misunderstandings and other issues negatively affecting the good working relationship between the two entities, can be ironed out. Once those issues are sorted out, it is more likely that both parties would be happy, hence the high rate of satisfaction. The same pattern was also observed on the numbers of partnership, given the period customers have been with the consulting group and the level of satisfaction. The longer the period customers are with the consulting group, the higher the number of partnerships. For example, the customers that had stayed with the consulting group for less than a year, between one year and five years and more than five had 5, 34 and 47 *(Figure 9)* partners respectively. It is logical because, if a client is happy with the services, they are satisfied then more people would be attracted to such a consulting group. It is therefore true to say that the more you get to know a service provider and you are happy with their service deliver, it wouldn’t be necessary to start a search for new consulting services. Many will stick with their current service providers. The services provider giving higher satisfaction to their client, the more clients would want to partner with them.

The correlation analysis using the correlation matrix revealed some interesting statistics. Only innovation, communication and implementation scored a correlation coefficient of above 0.500 when examined against satisfaction. Of the three variables that seemed to positively highly correlate with satisfaction, implementation was the highest with a score of 0.630 *(Figure 13)*. With implementation, it should be one of the most important variables that would mean satisfaction and that should be highly correlated with implementation. I would have thought the coefficient to be higher than the observed one. However, it could also be argued that, may be the number is not as high as it would have been thought initially because, even though the consulting group does implement the tasks given, the deliverables could be not of high standards. Although, the consulting group does the work, the clients aren’t always happy with the end services and that explains the not so high correlation coefficient value as shown in the table below. Innovation is another variable that seems to be highly positive correlated with satisfaction at a score of 0.596 *(Figure 13)*. This means, as innovation increases at the consulting group, there is also a rise in the level of satisfaction among clients. This makes sense because innovation allows for better and efficient service delivery which in turn meets the client’s expectation and that drives their satisfaction up. Although, we have to be a bit skeptical about the significance of communication to score such a high correlation coefficient of 0.540 *(Figure 13)*, it is not surprising. Communication is a key pillar in building a strong working relationship between the client and the consulting group. This is what builds trust and ultimately satisfaction over time. Whenever and wherever issues rise there should be swift and timely communication between the two parties. The earlier issues are communicated, the sooner they are likely to be resolved and everything move forward timely. Satisfaction and recommendation are highly correlated with a correlation coefficient of 0.762 *(Figure 13)*, meaning that the higher the rate of satisfaction among clients, the higher the likelihood that the consulting group is to be recommended to other people. When clients are happy, they spread the good news about the consulting group and that will more likely attract the attention of other people and would want to know more about the consulting group being talked about. The same can also be said about satisfaction against partnership. The more customers are satisfied, the more likely they are to partner with Odette Consulting Group. The two variables have a correlation coefficient of 0.693 *(Figure 13)* which is high, illustrating that customers that are more satisfied are 69% likely to partner with the consulting group.

There are some variables that seemed to have weak correlation between them and satisfaction that, logically, should have had strong correlation. Variables such as expertise, responsiveness, and portfolio management. These are variables that were earlier observed to influence the levels of satisfaction among clients. With the correlation model, the earlier observation made in the pivot tables do not seem to marry each other. In the earlier model in the pivot table, the level of expertise showed to reflect with satisfaction. However, in this correlation matrix, although the coefficient is in the positive, it remains below the value of 0.5 which means the correlation between the two variable is a weak one. It would have minimal effect to have a weak correlation between satisfaction and responsiveness because responsiveness is not a major aspect in the working relationship of the two entities. Because with good communication, the responsiveness aspect would be taken care of and would be less of a factor in whether customers are satisfied or not.

**Predictive Analysis**

**Methodology**

Based on our correlation analysis, we could see that variables such as portfolio management, responsiveness and expertise have very high correlation with each other. Similarly, innovation, communication and implementation also seem to be highly associated with each other. Due to these correlations between the mentioned variables, while making the regression models we make sure that all these highly correlated sets of variables do not appear together, as that shall lead to multi collinearity in the regression models. To decide which one of the variables we pick from these pairs, we select the variables that have the maximum correlation with our dependent variables – which is satisfaction in case of linear regression and partnership in case of logistic regression.

In both cases, it is the expertise and implementation variables that have the maximum association with our dependent variables and hence we use both variables in our predictive and classification regression models. In addition to these two variables, we also use the recommendation variable as a client’s likelihood of recommending OCG to others shall have a significant impact upon customer satisfaction and the likelihood of a resulting partnership. Additionally, we also try to use the categorical variable dummies in our regression models as the pivot table analysis of these variables suggested that they have high impact upon both satisfaction and partnership. After the preliminary linear and logistic regression models that only use the numerical variables having high correlation with the dependent variables, we try to improve our models by adding categorical variables or by eliminating the variables that are not significant in previous models (i.e., they have p values less than five percent).

**Linear Regression Models**

***Model 1***

This is a base model where we only take the ordinal variables that have the highest impact upon satisfaction, based on our correlation analysis. The regression equation for this model is as follows:

*Satisfaction = 0.263\*Expertise + 0.531\*Implementation + 0.545\*Recommendation*

*– 0.171\*ConsultingFee + 0.946*

The regression equation suggests that keeping all other factors constant, on average a one unit increase in the ability to implement projects successfully causes an increase of 0.531 in the satisfaction level. Similarly, project management expertise and the likelihood of recommending OCG to others also increase satisfaction. However, an increase in consulting fee is likely to decrease the satisfaction level by 0.171.

The detailed regression results (figure 1 of the appendix) show that all the independent variables are significant as they have p values less than five percent. P values of less than five percent show that we can reject the null hypothesis at five percent significance level in a t-test for each of the coefficients. The overall model is also significant as the p value for the f-stat is also less than five percent.

Despite all the variables being significant, our model has an adjusted R-square value of approximately seventy four percent only. This shows that the variables we took in our model only account for almost seventy four percent of the variation in satisfaction levels. Hence, to improve this we try a new model that also uses categorical variable dummies as our independent variables.

***Model 2***

Based on our pivot table and scatter plot analysis, we could see that categorical variables like customer type, industry type, firm type and region also have a considerable affect upon customer satisfaction. Therefore, we try to add these variables in our model. However, to be able to clearly interpret regression coefficients for these variables, we make dummies for each category within these variables and use those dummies in our model. The regression equation for this model is as follows:

*Satisfaction = 0.320\*Expertise + 0.238\*Implementation + 0.287\*Recommendation – 0.037\*ConsultingFee – 0.924\** *CustType\_LessThanOneYr – 0.352\** *CustType\_BtwnOne&FiveYr – 0.229 \* Industry\_IT – 0.241\** *FirmSize\_LessThanFifty – 0.439\*Region\_Ontario + 0.410 Partnership\_Yes + 3.375*

The regression equation suggests that keeping all other factors constant, on average if a person has been a customer for less than a year as compared to more than five-year-old customers, then that causes the satisfaction to go down by 0.924. Similarly, customers having a firm size of less than fifty or those living within Ontario cause satisfaction levels to go down by 0.241 and 0.439 respectively. A customer considering partnership with OCG will be a cause for an increment in the satisfaction level by 0.410. Variables like expertise, implementation and recommendation continue to cause an increment in satisfaction levels in this model too.

The detailed regression results (figure 2 of the appendix) show that all the independent variables (except consulting fees) are significant as they have p values less than five percent. The overall model is also significant as the p value for the f-stat is also less than five percent. The model has an adjusted R-square value of approximately eighty two percent and is a significant improvement from the previous model. This shows that the variables we took in our model now account for almost eighty two percent of the variation in satisfaction levels.

***Model 3***

Our previous model showed better results, but it had one variable that was not significant in terms of an effect upon satisfaction. Hence, we run another model but this time removing the insignificant variable – Consulting Fee – from the model. The regression equation for this model is as follows:

*Satisfaction = 0.317\*Expertise + 0.231\*Implementation + 0.294\*Recommendation*

*– 0.961\** *CustType\_LessThanOneYr – 0.379\** *CustType\_BtwnOne&FiveYr – 0.249 \* Industry\_IT – 0.227\** *FirmSize\_LessThanFifty – 0.487\*Region\_Ontario + 0.411 Partnership\_Yes + 3.160*

The regression equation is giving almost similar coefficient values as the previous model. The detailed regression results (figure 3 of the appendix) suggest that there is not much improvement in the adjusted R square for this model. In fact, the adjusted R square has gone down by 0.1 precent in this model. All the variables and the model overall are still significant based on p values.

Since this model does not add significant improvement to our previous model, we shall go ahead with model 2 as our final model as consulting fee has high association with satisfaction (with a correlation of almost 0.5 between the two) and can be an important predictor for satisfaction.

**Logistic Regression Models**

***Model 1***

This is a base model where we only take the ordinal variables that have the highest impact upon the partnership, based on our correlation analysis. Partnership the dependent variable is used in a multivariate setting to investigate if the partnership can be accomplished based on Expertise, Implementation, Recommendation, Consulting Fee and Satisfaction. The logit model used in this study is:

*Logit (Partnership) = -0.015\*Expertise + 0.12\*Implementation + 1.619\*Recommendation – 0.003\*Consulting Fee + 1.452\*Satisfaction – 22.674*

*To calculate odds ratio, Odds = e(coefficient)*

*For example,*  *Odds (Recommendation)= e1.619 = 5.049*

Partnership, the dependent variable, is a dichotomous variable with a value equal to 1 if the firm is identified as a suitable partner. The odds ratio for the above equation are 0.985, 1.127, 5.049, 0.996, and 4.273 respectively. The regression equation suggests that keeping all other factors constant, on average a one unit increase in the ability to implement projects successfully causes an increase of the odds ratio in the willingness of successful partnership. For instance, the odds ratio for recommendation is 5.049[4.579 – 5.519], which indicates that the odds of partnership with OCG will be 5 times higher if a recommendation is received. With a confidence of 95% in the range stated above. And the odds ratio of partnership is likely to quadruple if there is an increase in satisfaction.

The detailed regression results (figure 4 of the appendix) show that some independent variables are significant as they have p values less than five percent. P values of less than five percent show that we can reject the null hypothesis at five percent significance level in a chi square-test. This suggests that variable Recommendation is significant.

To further analyze this, we make use of the classification summary of our test dataset (figure 5 of the appendix); here we observe the confusion matrix, the error report, and the logistic regression metrics. By observing the error report, we can see that the class 1 error here is around fifteen percent and the class 0 error is approximately eleven percent. Our model has an accuracy of 86.67%, meaning the current model can predict roughly 87 percent of the predictions correctly. Of the total class 1 predictions generated by our model, approximately 90% predict correctly (positive partnership results). The sensitivity suggests that the model predicts 85 percent of the times actual class 1 values correctly. While the specificity shows how the model is able to detect the actual negative partnership 89% of the time. Our model's F1 score of 0.875 indicates the model is proficient at balancing both the captured positive cases (recall) and its accuracy with the cases it does capture (precision). The Area under the curve (ROC curve) of 93% suggests that the model is proficient in distinguishing between the positive and negative classes.

***Model 2***

Based on our pivot table and scatter plot analysis, we could see those categorical variables like customer type, industry type, firm type and region also have a considerable impact upon partnership. Therefore, we try to add these variables in our model. However, to be able to clearly interpret regression coefficients for these variables, we make dummies for each category within these variables and use those dummies in our model. The regression equation for this model (figure 6 of appendix) is as follows:

*Logit (Partnership) = 0.057\*Expertise - 0.203\*Implementation + 1.707\*Recommendation – 0.126\*Consulting Fee + 1.368\*Satisfaction –1.563\*CustomerType\_1 - 0.729\*CustomerType\_2 - 0.883 IndustryType\_0 – 19.5\*FirmSize\_0 - 1.14\*Region\_0 - 0.21*

The odds ratios are around 1.05, 0.816, 5.513, 0.881, 3.927, 0.209, 0.482, 0.413, 0, and 0.456 for the equation above, respectively. According to the regression equation, when all other variables are held constant, an average unit increase in the capacity to carry out projects successfully leads to an increase in the likelihood of successful partnership. For instance, the odds ratio for a recommendation is 5.51[5 - 6.02], which means that getting a recommendation will increase the chances of partnering with OCG by a factor of 5. 95% certain within the aforementioned range. The likelihood of a partnership increasing four times if there is client satisfaction are two other important variables in this equation along with firm size. However, a higher consulting price is likely to reduce the likelihood of a partnership by 12%. And if the client is from Alberta, the likelihood of a partnership drops dramatically by up to 54 percent.

After looking at the metrics for the logistic regression (figure 7 of the appendix), the confusion matrix, and the error report from the classification summary. Dummy category values were added, and as a result, the class 1 error was increased by three percent, to eighteen. Additionally, the class 0 error rate increased by 4 percent making it fifteen percent now. Our model's accuracy has plunged marginally by 4% making it now 83 percent, implying that it now predicts only about 83 percent of observations. 87 percent of the class 1 predictions produced by our model overall were accurate (positive partnership results). The improvised model's sensitivity has decreased somewhat by 3%, which means that it now predicts less actual class 1 values than before now predicting only 82 percent of the time. The F1 score of 0.84 for our model shows that it is less effective at balancing both the number of positive examples it captures (recall) and the accuracy of the cases it does catch (precision). The Area under the curve (ROC curve) remains unchanged suggesting that the model is still proficient in distinguishing between the positive and negative classes. Thus, we can say that introducing categorical values has not helped in our analysis with logistic regression because the overall metrics have deteriorated when compared to our first model.

**Findings and Conclusion**

In conclusion we are Odette consultancy group who are identifying good employees and project together. We will start from the very basic of knowing what the clients want. Then we will find out what hardware and software will best fit that need, and additionally also provide custom software development. Our main agenda for which we are doing all the analysis is to increase revenue while retaining and attracting old & new customer groups.

The main metric that we have used is customer satisfaction because we are banking on positive word of mouth. If our customers are satisfied with the service we provide to them, they will recommend our company and we will get new clients. At the end of the day whether a customer is going to be loyal to us or not depends on factors such as if they can find another company who are giving them equal quality service at a cheaper rate or not. It is prevalent that we know how much they are willing to spend on us, i.e., how much share we will have on their wallet.

While some researchers advocate for ‘likelihood to recommend’ business as the best predictor for sales growth, many says that there are multiple variables at play which is a better way of forecasting the satisfaction rate. Odette Consultancy Group feels that just using one variable may cause us to become myopic since this is a factor that we don’t have much control over as we cannot force our customers to recommend our company to others. At the same time there will be other competitors in the market who will provide similar services at a lower price. Therefore, we have decided to use multiple metrics: Likelihood of recommendation & satisfaction score.

After thoroughly going through the data set, we have realized that there were no problems with the dataset. We checked for missing values, encoded the categorical variables, checked to see if there were any duplicates or spelling errors. The data set was labelled, and the description of the categorical variables were clear. There were no outliers or data discrepancy. The steps taken to pre-process the data set has been mentioned in detail in the ‘Data Pre-processing’ section.

For exploratory analysis we used four variables for prediction: Customer type, industry type, firm size, region as the independent variables and satisfaction as the dependent variable. We found that the longer a customer company was with us, the higher was their average rate of satisfaction as compared to those who stayed for a short period of time. The pattern was the same for number of partnerships. The highest correlation with satisfaction was implementation with a score of 0.63 followed by innovation and communication with a score of 0.5. Expertise, responsiveness, & portfolio management surprisingly has a weak correlation. Resultantly we have used expertise and implementation variables for our predictive analysis.

We have used three models for linear regression, of which Model 1 accounts for 74% of the variation in satisfaction, Model 2 accounts for 82%, & lastly Model 3 does not account for significant improvement. Hence, we used Model 2 as our final model due to the high association between consultancy fee & satisfaction. For logistic regression we have used two models. Model 1 has a F1 score of 0.875 and Model 2 has a F1 score of 0.84 so both are effective at balancing both the number of positive examples it captures (recall) and the accuracy of the cases it does catch (precision). Additionally, the overall accuracy of the model drops by 4 percent when categorical variables were introduced. Thus, we can say that introducing categorical values has not helped in our analysis with logistic regression because the overall metrics have deteriorated when compared with our first model.

Lastly, it is important to include as many relevant factors as required for analysis of customer satisfaction as it is a very dynamic variable in and of itself. It is important to give customers good value for money so that they consider us to be their first choice when it comes to partnership. Additionally, this survey should be done in contrast with other competitor companies as well so that we can weigh the variables that are giving them an advantage and try to implement those for ourselves. Summing up, it is important for a company to regularly update their surveys and data so that they don’t fall behind.

**References**

Baehre, S., O’Dwyer, M., O’Malley, L. et al. (2022). The use of Net Promoter Score (NPS) to predict sales growth: insights from an empirical investigation. *Journal of the Academy of Marketing Science Vol* 50, 67–84. <https://doi.org/10.1007/s11747-021-00790-2>

Bansal, S. (2021, May 10). *10 super neat ways to clean data in Excel spreadsheets*. Trump Excel. <https://trumpexcel.com/clean-data-in-excel/>

Basil P.,Yen D.C., Tang H.-L., (1997). Information consulting: developments, trends and suggestions for growth. *International Journal of Information Management, Vol* 17, Issue 5, Pages 303-323 <https://doi.org/10.1016/S0268-4012(97)00013-3>.

Chauhan, N. S. (2019, July 19). *From data pre-processing to optimizing a regression model performance*. KDnuggets. <https://www.kdnuggets.com/2019/07/data-pre-processing-optimizing-regression-model-performance.html>

Graubner, M., & Richter, A. (2003). Managing tomorrow's consulting firm. *Consulting to Management, Vol* 14(3), 43-50. <https://www.proquest.com/scholarly-journals/managing-tomorrows-consulting-firm/docview/215897671/se-2>

Richard L. Oliver, (1999). Whence Consumer Loyalty? *Journal of Marketing, Vol.* 63, 33-44 <https://doi.org/10.1177/00222429990634s105>

Timothy L. Keiningham, Bruce Cooil, Lerzan Aksoy, Tor W. Andreassen, Jay Weiner. (2007). The value of different customer satisfaction and loyalty metrics in predicting customer retention, recommendation, and share‐of‐wallet, *Managing Service Quality: An International Journal*,

*Vol*. 17 No. 4, 361-384 <https://www.emerald.com/insight/content/doi/10.1108/09604520710760526/full/html>

Yoon, S., Suh, H. (2004). Ensuring IT Consulting SERVQUAL and User Satisfaction: A Modified Measurement Tool. *Information Systems Frontiers Vol* 6, 341–351. <https://doi.org/10.1023/B:ISFI.0000046376.10364.16>

**Appendix**

Table

Description automatically generated

*Figure 1: Detailed Regression Results (Linear Model 1)*

Table, Excel

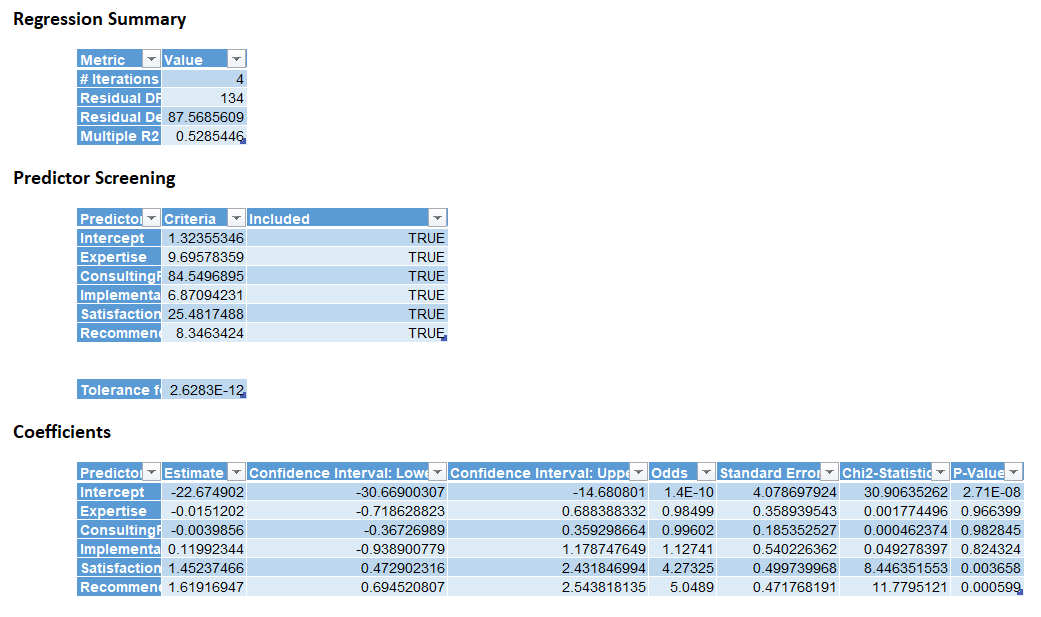
Description automatically generated

*Figure 2: Detailed Regression Results (Linear Model 2)*

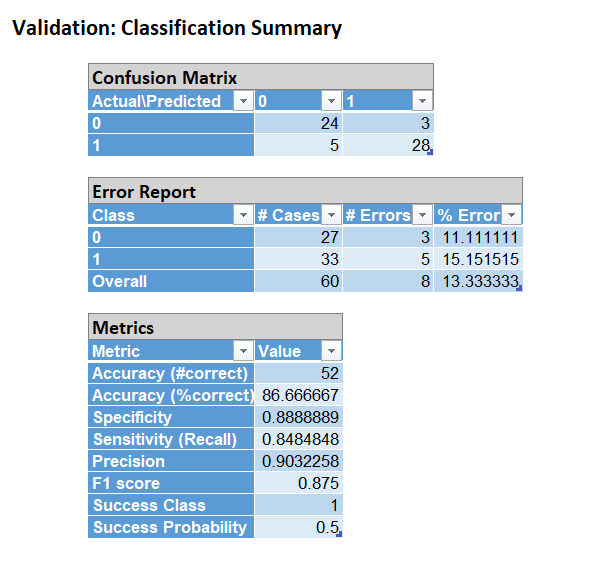
Application, table, Excel

Description automatically generated

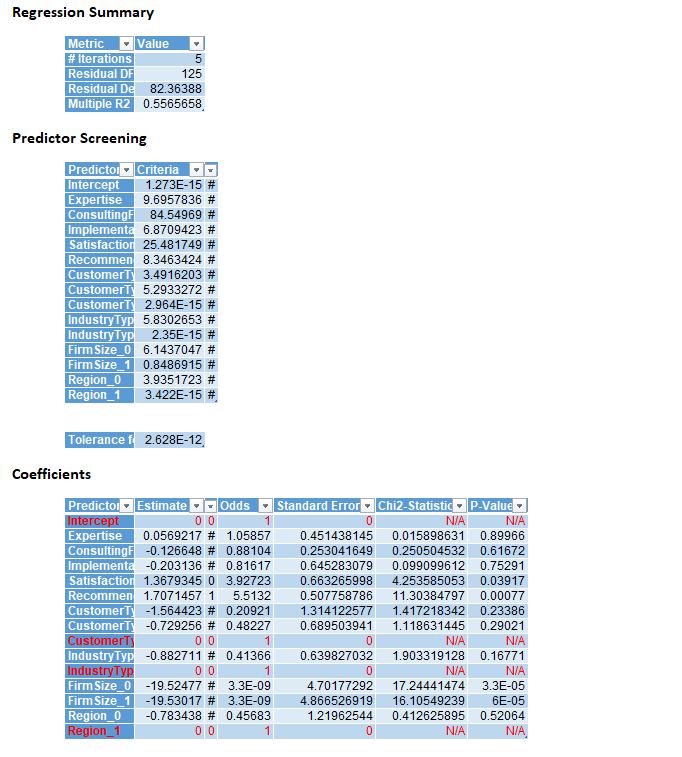
*Figure 3: Detailed Regression Results (Linear Model 3)*



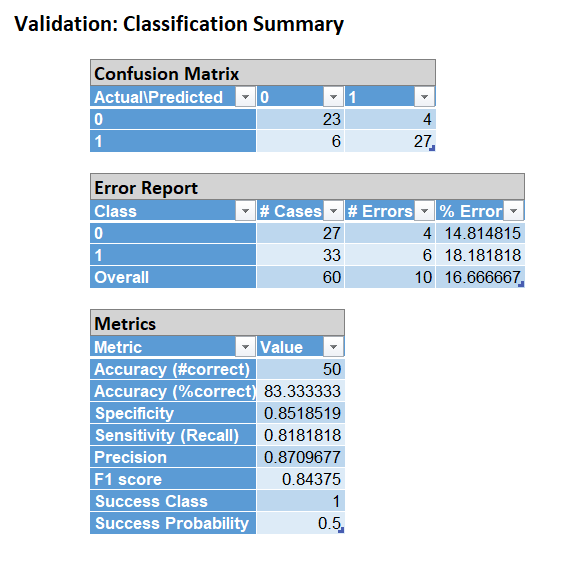
*Figure 4: Detailed Regression Results (Logistic Model 1)*



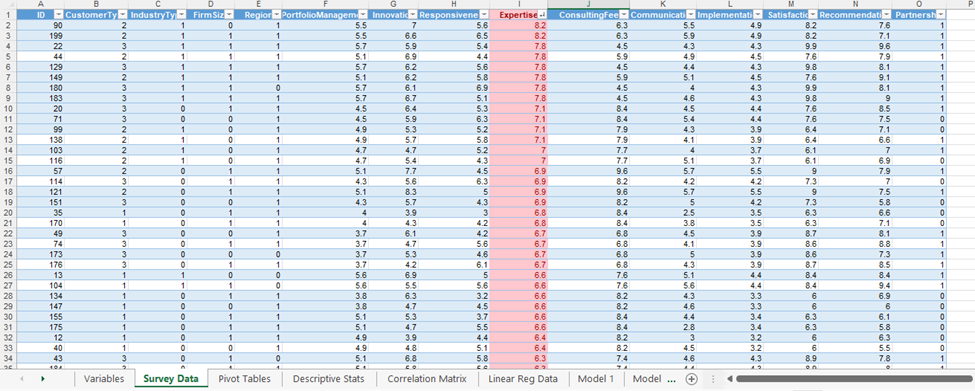
*Figure 5: Classification Summary (Logistic Model 1)*



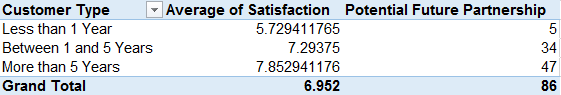
*Figure 6: Detailed Regression Results (Logistic Model 2)*



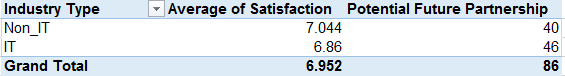
*Figure 7: Classification Summary (Logistic Model 2)*



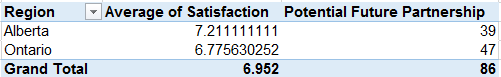
*Figure 8: Illustration of Conditional Formatting (Exploratory Data Analysis)*



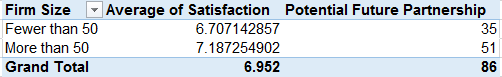
*Figure 9: Customer type*



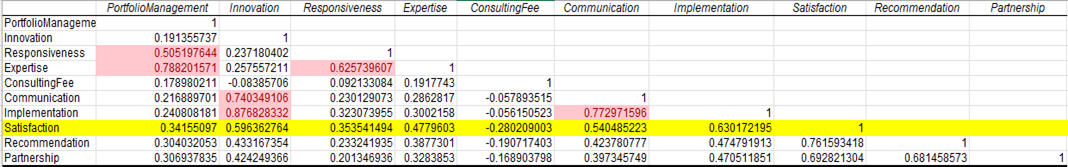
*Figure 10: Industry Type*



*Figure 11: Region*

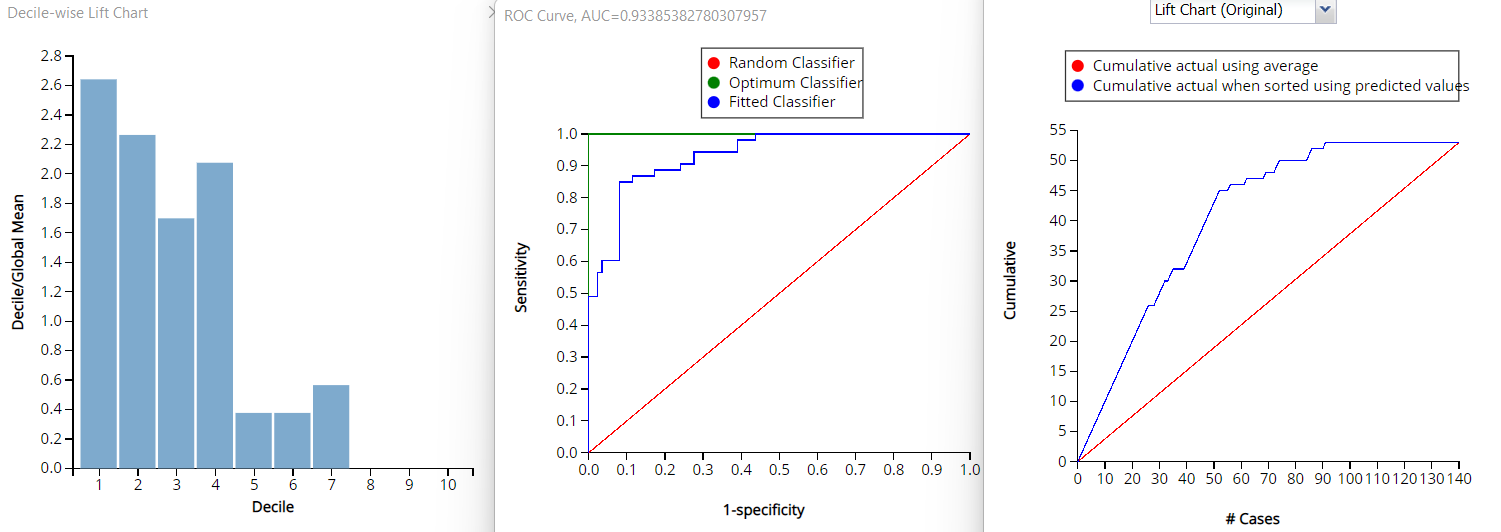


*Figure 12: Firm Size*

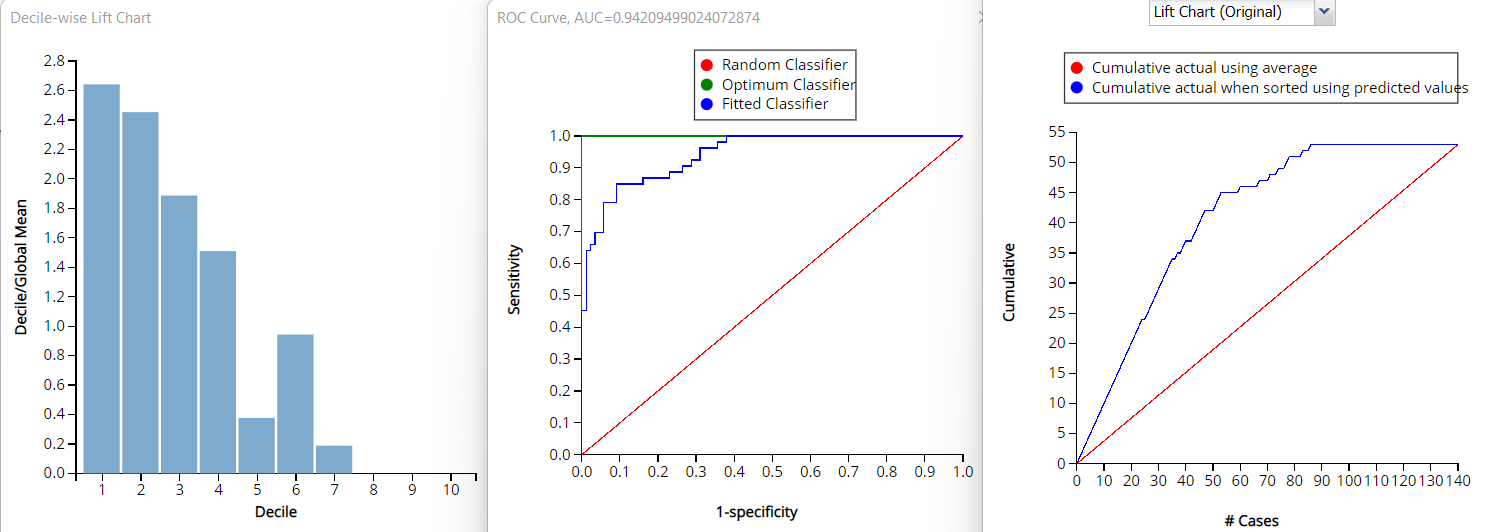


*Figure 13: Correlation between Satisfaction and various Independent Variables*

*Figure 13: Correlation Matrix*



*Figure 14: Decile Chart, ROC Curve and Lift Chart Logistic Regression Model 1*



*Figure 15: Decile Chart, ROC Curve and Lift Chart Logistic Regression Model 2*