**Odette Consulting Group**

Table of Contents

[Executive Summary 3](#_Toc110295303)

[Literature Review 4](#_Toc110295304)

[Research Questions 6](#_Toc110295305)

[Data Pre-Processing 7](#_Toc110295306)

[Data Quality assessment 7](#_Toc110295307)

[Data Transformation 7](#_Toc110295308)

[Data Reduction 7](#_Toc110295309)

[Data Analysis 8](#_Toc110295310)

[Exploratory Data Analysis 8](#_Toc110295311)

[Exploratory Data Analysis: Linear Regression 8](#_Toc110295312)

[Exploratory Data Analysis: Logistic Regression 13](#_Toc110295313)

[Predictive Analysis 14](#_Toc110295314)

[Linear Regression 15](#_Toc110295315)

[Logistic Regression 17](#_Toc110295316)

[K Nearest Neighbour 19](#_Toc110295317)

[Classification Tree 20](#_Toc110295318)

[Conclusion 21](#_Toc110295319)

[Insights and Recommendation 21](#_Toc110295320)

[References 22](#_Toc110295321)

[Appendix 23](#_Toc110295322)

# Executive Summary

The Odette Consulting Group is a consultant group located in Windsor, Ontario with clients from various sectors including IT and non-IT sector. The company conducted a survey from its 200 clients to measure their satisfaction from the services provided by OCG group.

The literature review suggested that the Canadians are not the easiest clients to satisfy but a reputable company, Statistics Canada, has a customer satisfaction of 87%. Therefore, we have two benchmarks to evaluate the performance of OCG group with respect to average satisfaction level of Canadian and the other a professional company that provide similar type of consultancy services and has an extremely satisfied customer base. In order to accurately measure the client’s satisfaction, we have considered two dependent variables, recommendation and partnership as they can provide accurate measure of client’s satisfaction because a dissatisfied client with neither plan to continue its partnership with the company nor will recommend it to someone.

The data was processed including data quality assessment, data transformation and data reduction before the analysis. The analysis included explanatory data analysis to understand the distribution and patterns of data, and predictive analysis to predict and classify the results. For this purpose, we conducted linear regression, logistic regression and K nearest neighbour. The dependent variables are recommendation and partnership while independent variables are customer type, industry type, firm size, region, client portfolio management, innovative project management, responsiveness, expertise, competitive consulting fee, communication, and implementation. The results show that the satisfaction has a positive relationship with the dependent variable viz. Recommendation and Potential of future Partnership.

# Literature Review

The World Happiness Report is the UN’s annual report on the world happiness. It shows how satisfied the residents of different countries of the world are. The Canadians are not among the top 10, they are ranked at 14th position meaning that the Canadian’s are not the easiest people to satisfy. The Odette Consulting Group (OCG) is operating from Windsor, Canada so they have a demanding customer base. Their customer satisfaction survey should account those Canadian customers are hard to satisfy. The client’s satisfaction is a tough job as service providers must deal with a range of different emotions from customers on several different channels at the same time. On top of that, the expectations of customer are increasing. In fact, 93% of professionals agree that customers today have higher expectations than ever before. the American Customer Satisfaction Index gathered the average customer satisfaction levels for several industries in the year 2021 and found that top three industries with highest customer satisfaction are Real Estate 96%, IT Services and Consultancy 95%, and Healthcare. 94%. The worst three are Retail and Wholesale 82%, social media 78%, and Entertainment and Arts 77%. This implies that the customer belonging to real estate or IT services are easy to satisfy then the customer in retail, wholesale, and media etc. According to Accenture's tenth annual Global Consumer Pulse Research, over one-quarter of Canadian consumers (28%) feel very loyal toward their providers and similar percent are willing to recommend them to others. These figures have remained consistent over the past several years. The 38% are open to purchasing products and services offered by non-traditional providers. This highlight the fact that the Canadian clients are more open to switching services and are not much loyal to their brands. Moreover, the percentage of recommendation is also on the lower end for the Canadian customers.

When it comes to company size, the stats revealed some unexpected results, including the existence of a “customer satisfaction valley” that correlated to the size of the company delivering the service. The survey showed that smaller companies - those with one to nine employees - had a customer satisfaction average of 91%. Larger organizations, with a headcount between 500 and 4,999, had an average customer satisfaction rating of 93%. The largest companies (5,000+ employees) averaged 90% on the customer satisfaction scale. But mid-sized companies fell short. Companies with 10-99 employees averaged 84%, and those with 100-499 staff posted 88% satisfaction rates. Customer satisfaction survey can be derived by following formula:

Satisfaction = Scale \* Efficiency \* Quality

A similar type of survey was conducted by Statistics Canada, the national statistical office, one the most reputable survey provider. The purpose of the Statistics Canada Client Survey was to measure satisfaction with product and/or service quality, delivery, and use. The greatest number of respondents were from international organizations and the Federal Government. 84% of clients were satisfied with the overall quality of the product or service received. Respect, courtesy, and response of the staff received the highest levels of satisfaction. 71% of clients considered the product or service useful. Areas of improvement include cost and amount of time to get the product or service as well as the level of detail and supporting documentation.

Therefore, while analysing the performance of Odette Consulting Group, we will consider the overall satisfaction index of the average Canadians as well as the results of survey conducted by Statistics Canada as benchmark for comparative analysis.

# Research Questions

A well-defined research question is necessary as it exactly defines what you want to find out and gives your work a clear focus and purpose. The researcher can then take multiple approaches to establish if his hypothesis is correct or not. Without a proper question, the researcher will only play around data without any conclusive insight from it. From our literature review of Satisfaction Index of the clients in Canada, we found out that they are not the easiest customer to satisfy. Meanwhile, the Statistic Canada which is a reputable company and specialize in surveys have a high satisfaction from its clients. Therefore, we have two benchmarks to gauge the OCG’s client satisfaction with average satisfaction of clients and an upper limit. If the company performs somewhere in between, we can conclude that they have a satisfied customer base. We are also aware that survey is prone to have different biases especially framing and anchoring. In order to evaluate the customer satisfaction of OCG’s clients we decided to monitor the relationship of client’s recommendation and intent to continue the partnership with the consulting group. These are two dependent variables that will have a positive relation only if the clients are satisfied with the services. A dissatisfied client will never recommend the company to others or will have no intention to continues its partnership. Hence, these are two indirect but reliable variables to identify the client’s satisfaction.

Therefore, our research question is “How many of the clients plan to continue the partnership with our company, Odette Consulting Group, and how many are willing to recommend the company to others to measure the satisfaction level of company’s services to its existing clients?

# Data Pre-Processing

Data pre-processing is the idea of transforming the raw data into a clean data set. Pre-processing data makes it simpler to analyse and use. The accuracy of our model is improved by removing data discrepancies or duplicates. The actions we took to prepare the dataset are listed below.

## **Data Quality assessment**

To begin with, we verified the data's accuracy by assessing whether the information was true. We also made that the data was comprehensive, had no missing values, and was reliable. However, there were three qualitative customer type categories: less than one year, between one and five years, and more than five years. We split the three client categories mentioned above into two types to perform linear regression in the exploratory analysis and compare recommendations across various variables. Additionally, it was emphasized in class that when the first two variables are known, the third variable could be automatically captured. Moreover, this helps in effectively analysing and computing the correlation matrix as well.

## **Data Transformation**

To perform logistic regression and to compare the variation of partnership across the various variables like customer type, industry type, firm size, region, etc. we substituted the binary variables 0s and 1s with valid numerical labels for the current trend to be captured.

## **Data Reduction**

There was no room for data reduction in our study because we used every piece of information available in the survey data. In this way, we made sure that the data quality was boosted and that no values were inaccurate or missing, which will lead to high-quality mining results.

# Data Analysis

To find answers to the research questions, we have classified our data analysis in two categories named as exploratory data analysis and predictive analysis. In exploratory data analysis we understand the distribution of data and identify any patterns that can be derived by creating various graphs and charts, it also helps to understand the relationship between the variables in our dataset. In predictive analysis we will create models which will help us in prediction or classification of the results. Based on the analysis we will be further able to provide insights and recommendations for the business or stakeholders.

## Exploratory Data Analysis

We have conducted exploratory data analysis for linear regression which will help us predict the likelihood of Recommendation which is the dependent variable in this case and logistic regression which will help us determine potential future partnership which will be our dependent variable for this purpose, hence we needed to understand the effect of all other independent variables on them. In the data we had 5 qualitative variables Customer type, Industry type, Firm size, Region, and Partnership. The other 9 variables are quantitative viz. Client portfolio management, Innovative project management, Responsiveness, Expertise, Competitive consulting fee, Communication, Implementation, Overall client satisfaction, and Likelihood of recommending OCG to others.

### Exploratory Data Analysis: Linear Regression

Since we have conducted a survey and have asked to input ratings, we will understand the effect of ratings of different parameters. To understand the pattern, we have plotted average ratings across all the qualitative variables viz. Customer type, Industry type, Firm Size, and Region with the quantitative parameters. It will help us identify the pattern and the relationship they have, and how it will affect the dependent variable.

#### Customer Type

The average rating of the ability to handle client’s overall portfolio across all customer type is quite low with 3.813 being the highest for the customers with length of relationship between 1 and 5 years. All customer types are rated similarly highly on the capacity to handle projects innovatively, with 5.91 being the highest rating for clients with relationships longer than five years. The customers with a related term of between 1 and 5 years placed top with the highest average rating of how responsive OCG is to a client's requirement, with 4.40 being the highest. All client types have a relatively low average level of project management expertise, with customers with relationships lasting between one and five years having the longest on average 5.57. The OCG's consulting costs are competitive, and it's noticeable that their average across all clients is quite high, with clients with relationships lasting less than a year paying the highest rate of 7.59. With 4.69 being the highest for clients with relationships lasting between one and five years, the average rating of project managers' communication skills across all clients is the highest. With a difference of 2% between the two, the average ratings of the ability to implement projects successfully are nearly identical for client types 2 and 3. The average rating of the overall client satisfaction is quite high among all the customer types with 7.85 being the highest for the customers with a length of relationship longer than 5 years. Calculated results show that all client types have nearly the same average likelihood of referring OCG to others, with the highest average likelihood being 7.52 for customers with relationships lasting longer than five years.

#### Industry Type

The average rating of the ability to handle client’s overall portfolio with respect to the type of client industry is also below average, however for IT clients the average rating 3.876 is slightly better than the non-IT clients. The capacity to handle projects innovatively is determined to have a poor average rating depending upon the type of client industry, however for IT clients, the average rating is 5.422, which is marginally better than the non-IT clients. In terms of the client industry, the average rating for OCG's responsiveness to client needs is also below average; however, for IT clients, the average rating is 4.068, which is only marginally better than for non-IT clients. The average evaluation of OCG's project management proficiency for IT clients is observed to be 5.254, which is higher than the rating for non-IT clients. The average rating of consulting fees across industries is found to be significantly higher in the IT industry than in the non-IT industry, with a rating of 7.174. The non-IT industry has a reasonably modest grade for project managers' communication abilities, with a rating of 4.250, which is somewhat higher than the IT industry. With a range of roughly 3.82, the average assessment of the ability to carry out projects successfully is discovered to be relatively better in the IT industry than the non-IT industry. The non-IT sector consistently rated much higher on average for customer satisfaction, with the IT sector scoring the highest at 7.044. In contrast to the IT industry, the non-IT industry has a higher average likelihood of referring OCG to others, with 6.958 being the highest.

#### Region

The average rating of the ability to handle client’s overall portfolio for Ontario is 3.893 which is slightly better than that of Alberta which has an avg. rating of 3.577. However, the overall rating is low across both the regions. Overall, both areas scored lower on the capacity to manage projects innovatively than the national average, with Alberta scoring 5.37, slightly higher than Ontario's average of 5.36. When the level of OCG's client responsiveness is averaged out, Ontario appears to have a rating of 4.27, which is higher than Alberta on average. With a range of 5.6, Ontario's average rating for OCG's project management proficiency is found to be significantly higher than Ontario's. Calculating OCG's average consulting costs for the two regions, Ontario has higher fees than Alberta does, with an Ontario average consulting fee rating of 7.71. With a range of 4.25 and 4.24, respectively, both regions—Ontario and Alberta have relatively low averages in terms of the project managers' communication skills. The average rating for Ontario's capacity to carry out projects successfully is 3.82, which is marginally higher than Alberta's average grade of 3.80. Nevertheless, both regions' overall ratings are low. Overall customer satisfaction ratings in both regions are on the moderate side, with Alberta leading the pack with an average rating of 7.21, which is superior to Ontario. With a range of 7.04, which is significantly higher than Ontario, the average likelihood of recommending OCG to others is higher in the Alberta region.

#### Firm Size

The average rating of the ability to handle client’s overall portfolio for both the categories of firm sizes is low. Although, the avg. rating for the firm size with more than 50 employees is 3.865, which is slightly greater than the firms having less than 50 employees. The capacity to handle projects innovatively has a reasonable average rating across both categories of firm sizes, with the firm with more than 50 people scoring somewhat higher on average 5.549 than the firms with fewer than 50 employees. Between the two companies, the OCG's responsiveness to client needs is rated relatively poorly on average. The average rating for companies with more than 50 employees is 4.290, which is marginally higher than the rating for companies with fewer than 50 employees. The average rating of project management expertise for the two companies varies significantly, with the firm with more than 50 workers receiving an average rating of 5.57 and the firm with fewer employees receiving only 4.906. With a rating of 7.446—nearly a full point higher than the other firm the average rating of OCG's fees for the firm with more than 50 employees is high. The project managers at both companies receive a relatively lower average rating for their communication skills. However, the average rating for companies with more than 50 employees is 4.293, which is marginally higher than the average rating for companies with fewer than 50 employees. For both categories of business sizes, the capacity to undertake projects successfully receives a poor overall rating. However, the average rating for companies with more than 50 employees is 3.939, which is marginally higher than the average rating for companies with fewer than 50 employees. With the larger firm's average customer satisfaction rating assessed to be 7.187 and the other firm's to be 6.707, the overall average rating of client satisfaction across both businesses is significantly higher than the other factors. According to computations, the average rating for the chance of referring OCG to others for businesses with more than 50 employees is 7.267, which is much higher than the rating for businesses with fewer employees.

Bivariate Analysis

To understand the relationship between the dependent and the quantitative variables we have utilized the bivariate analysis by creating scatter plots. This analysis will explain the variation of the independent variables and help us understand the correlation.

The overall recommendation rating with the ability to handle the overall client’s portfolio can derived by looking at the variation between them, it suggests that it has a weak positive linear relationship with the magnitude of 0.304. By examining the variation between them, it is possible to determine the overall recommendation rating and the capacity to manage the project creatively; they have a positive linear relationship with a magnitude of 0.433. When comparing them, one can see that there is a positive weak linear association with a magnitude of 0.233 between the total recommendation rating and OCG's capacity to meet client needs. A positive weak linear regression with a magnitude of 0.387 is detected between the overall recommendation rating and OCG's project management skills. A negative weak linear relationship with a value of -0.19, which is noticed to be the lowest, exists between the total recommendation rating and the ability of OCG's consulting fees. By examining the differences between them, the total recommendation rating for the project managers at OCG is calculated. It has a positive weak linear association with a magnitude of 0.423. A positive linear relationship with a value of 0.474 is established between the overall recommendation rating and the capacity to carry out projects successfully. By comparing their variance, it is possible to determine the total client satisfaction, which can be distinguished from the recommendation rating; it has a positive linear association with a magnitude of 0.761, which is found to be the greatest of all.

### Exploratory Data Analysis: Logistic Regression

To understand the potential of future partnerships across various factors we have plotted charts to capture the variation.

The variation of partnerships across customer types suggests that the customers who are associated for a longer period with OCG are more likely to get involved in the potential future partnerships. Approximately 93% of customers who are associated with OCG for less than one years are reluctant to form a partnership, whereas 70% of customers who are engaged for more than 5 years are more inclined to form a partnership. The variation in partnerships between the two types of industries shows that the variation is essentially the same for the IT and non-IT sectors. While only 54% of the IT businesses are not looking to partnering with OCG, over 60% of the non-IT industries connected to OCG are unlikely to do so. The variation in partnerships among the firm sizes shows that the company with more than 50 workers has been affiliated with OCG for a longer length of time and is more likely to participate in potential future collaborations. Surprisingly, it is shown that the proportion for both decisions is the same for business sizes with more than 50 employees. Contrary to what was stated above, nearly 63 percent of businesses with fewer than 50 employees are unwilling to create a partnership. According to the differences in partnerships between regions, only 39% of clients in Alberta are willing to build a relationship with OCG, compared to nearly 72% of clients in Ontario who are hesitant to do so.

The average rating of the ability to handle overall client’s portfolio management across the potential future partnerships is low, but the avg. rating for forming future partnerships is 4.036 which is slightly better when compared to the firms not forming partnership. Although the average rating for creating future partnerships is 4.32, which is marginally better than the firms not creating partnerships, the average rating for OCG's capacity to address customer demands across all prospective future partnerships is nevertheless relatively low. The ability of OCG to successfully implement projects across all potential future collaborations is found to have a low average rating; nevertheless, the average rating for creating new partnerships is 4.2, which is marginally better than the enterprises that do not develop partnerships. The ability of OCG to manage projects creatively across all potential future collaborations is likewise determined to have a low average rating; however, the average rating for creating new partnerships is 4.2, which is slightly better than the organizations that do not develop partnerships. The ability of OCG to manage projects across prospective future collaborations is found to have a reasonable average rating, but the average rating for creating future partnerships is 5.67, which is noticeably better than the enterprises not creating partnerships. The average rating for developing future partnerships is found to be high at 7.9, which is noticeably better when compared to the firms not forming partnerships. The average rating for overall client satisfaction across the prospective future partnerships is determined to be high. The average rating for OCG's consulting fees across all prospective future partnerships is determined to be high, with the highest rating 7.18 found for not forming future partnerships, which is noticeably superior to the firms’ forming partnerships.

## Predictive Analysis

Based on the survey data, and conducting an initial exploratory data analysis, we have identified Recommendation as a dependent variable which can be predicted based on the other independent variables. Linear Regression is performed on this dataset as it is appropriate algorithm to predict the quantitative variable that is recommendation in our case. In Addition, to classify the potential future partnership we have applied logistic regression, as it helps us with the classification., in our case Partnership will be the classification variable.

### Linear Regression

Linear regression is used to predict the values of dependent variable based on the values of the independent variables. This method estimates the coefficients of the linear equation with one or more independent variables that can help us predict the value of dependent variable. Linear regression can be represented by the following equation:

y = b0 + b1\*x1 + b2\*x2

Where y is the dependent variable, b0 is the intercept, b1 and b2 are coefficient for the independent variables x1 and x2 respectively.

First Iteration:

After conducting the pre-processing of data and converting Customer type into the dummy variables viz. customer type 1 and customer type 2, further utilizing the Data analysis capability of excel we created the first iteration of linear regression model using all the independent variables to create a base model. This base model had the R square value of 0.626 and Adjusted R Square as 0.600. After studying the p-values of all the variables and the correlation matrix we identified Implementation, Innovation and Expertise to be removed from the model, as their respective p-values were 0.993, 0.190 and 0.583 also Implementation had strong correlation with Innovation and communication with the strength of 0.877 and 0.773 respectively, similarly Innovation had strong correlation with Communication, Implementation and Satisfaction with the strength of 0.740, 0.877 and 0.596 respectively, also Expertise had strong correlation with Portfolio management and Responsiveness with the strength of 0.788 and 0.626 respectively hence we decided to remove these variables from the base model and perform another iteration to improve the model.

Second Iteration:

Post removing the above identified variables from the base model, we created the second iteration. The second model had R-square value of 0.620 and Adjusted R-square of 0.600. In this iteration, after analysing the p-values we decided to remove Customer type 2, Region, and Communication from the model as they had very high p-values of 0.793, 0.996 and 0.703 respectively. After removing these variables, a third iteration will be created to further refine the model.

Third Iteration:

In third iteration, the generated model has R-square value of 0.620 and Adjusted R-square value of 0.606. We can still find several variables which still have a higher p-value than 0.05. We removed Customer type 1, Industry type, Portfolio management and Consulting Fee as they have relatively high p-values of 0.302, 0.274, 0.070 and 0.211 respectively. We have kept Responsiveness for the next iteration as its p-values is very close to 0.05 with 0.055, as removing the above variables might impact Responsiveness. We created another iteration to further enhance the model.

Fourth Iteration:

In fourth iteration, the model had R-square value of 0.607 and Adjusted R-square of 0.60. Removing the above variables impacted the p-value of Responsiveness and it is now 0.171 which is significantly greater than 0.05. Hence, we will remove Responsiveness and generate another iteration to further refine the model.

Fifth Iteration:

This also turned out to be the final iteration and a final model was generated. This model had the R-square value of 0.603 and Adjusted R-square as 0.599. The remaining variables have the p-value as less than 0.05, Firm size and Satisfaction are found to be the significant variables impacting the Recommendation with p-values of 0.001 and 0.000 respectively. Hence, this will be our final model where we can say that Recommendation can be studied with the variation of Firm size and Satisfaction.

In terms of the linear regression equation, the intercept is 2.344 and the co-efficient for Firm size and Satisfaction is 0.335 and 0.638. This means that for Firm size with more than 50 employees, with 1 unit increase and Satisfaction, Recommendation will increase on average by 0.335 and 0.638 respectively. These variables can now be used to predict Recommendation in future.

### Logistic Regression

Logistic regression helps with estimating the probability of an event occurring based on a given dataset of independent variables. The outcome of the dependent variable is restricted to either 0 or 1 as the outcome is a probability. In logistic regression, a logit transformation is applied on the odds—that is, the probability of success divided by the probability of failure. Logistic regression can be represented by following equation:

Logit(pi) = 1/ (1+ exp(-pi))

ln(pi/(1-pi)) = b0 + b1\*x1 + … + bk\*xk

where logit(pi) is the dependent or response variable and x is the independent variable. The beta parameter, or coefficient, in this model is commonly estimated via maximum likelihood estimation.

First Iteration:

After performing the pre-processing of data and converting Customer type into the dummy variables viz. customer type 1 and customer type 2, further utilizing Analytic Solver we generated the initial logistic regression model to help us classify the Potential future partnerships. This initial base model possessed Specificity of 0.850, Sensitivity (Recall) of 0.837, Precision of 0.808, and F1 score of 0.822.

Specificity can be understood as true negative rate which means that it is a measure of proportion of the negatives which are correctly identified and is complementary to false positive rate. It can also be derived with following formula:

Specificity = True negative / (True negative + False positive)

Sensitivity or recall can be understood as the true positive rate which is the measure of proportion of actual positive which are identified and is complementary to false negative rate. It can also be derived with the following formula:

Sensitivity = True positive / (True positive + False negative)

Precision can be understood as the measure of the ratio of the True positive observations to the total predicted positive observations. It can also be derived using the following formula:

Precision = True Positive / (True positive + False Positive)

F1 score can be understood as the measure of weighted average of Precision and Sensitivity (Recall). It can be derived from following formula:

F1 score = 2 (Recall \* Precision) / (Recall + Precision)

In addition, we studied the p-values of all the variables, we decided to remove Customer Type 2, Firm size, Innovation, Expertise, and Implementation as these variables have significantly higher p-values of 0.836, 0.727, 0.358, 0.555, and 0.527 respectively. In this iteration we removed Innovation as it had strong correlation with communication, implementation, and satisfaction. After removing the above variables another iteration is generated to further improve the model.

Second Iteration:

In this iteration, the model had the Specificity, Sensitivity, Precision and F1 score of 0.859, 0.813, 0.813 and 0.813 respectively. Additionally, we analysed the p-values of all the variables and decided to remove Industry type, Firm size, Consulting fee, and Communication as they had higher p-values compared to all other variables with values of 0.197, 0.200, 0.241 and 0.784 respectively. After removing these variables another iteration was created to further refine the model.

Third Iteration:

This was the final iteration resulting in our final model, this model had the Specificity, Sensitivity, Precision and F1 score of 0.833, 0.825, 0.788, and 0.806 respectively. In addition, we studied the p-values for the following variables Customer type 1, Portfolio management, Responsiveness, and Satisfaction and found them as 0.075, 0.004, 0.036, and 0.000. Although, Customer type 1 is having a p-value of 0.075, we have kept it in our model as this value is very close to 0.05 and would help us in explaining the variation in future partnerships. This suggests that these variables are significant factors to determine the potential future partnerships.

In terms of logistic regression equation, the intercept is -15.509 and the coefficient of Customer type 1 and Responsiveness are -1.451 and -0.456 whereas the coefficient of Portfolio management and Satisfaction is 1.147 and 1.834 respectively. This means that Customer type 1 and Responsiveness has negative influence on the future partnerships whereas Portfolio management and Satisfaction have positive effect on the outcome of the future partnerships.

### K Nearest Neighbour

K nearest neighbour or KNN utilizes proximity to make classifications or predictions about the grouping of a particular data point. It works on a basic assumption that similar points can be found near one another. For the purpose of classification, a class variable is assigned on the basis of majority vote which typically means that the label that is more frequently represented around a given data point.

In this classification, we have utilized the variables that have shown an impact on the potential future of partnership and created a model using the Analytic Solver in Excel. The output parameters like Specificity, Sensitivity, Precision, and F1 score are 0.947, 1, 0.934 and 0.966 respectively which are better than the final model of logistic regression. We can see that there are no False negatives found in the output and it is giving us accurate results for True positive. Hence, this will be a better classifier than the logistic regression based on the above matrix.

### Classification Tree

Classification Tree also known as decision tree algorithm is used to solve problems pertaining to regression and classification. The aim of the classification tree is to form a training model that can then be utilized to predict the class of the dependent variable by learning simple decision rules deduced from training data.

We have created a classification tree using the Analytic Solver in Excel and found that Specificity, Sensitivity, Precision and F1 score are 1, which implies that this model is an ideal model. Since there are less data points in the classification tree model, we cannot fully rely on this model to predict the dependent variable. Hence this is not reliable in this particular case.

# Conclusion

## Insights and Recommendation

By analysing the results of Linear Regression, we can infer that Firm Size and Satisfaction are the significant drivers for any firm to provide recommendation to OCG, both these have positive effect on it and Firm size which have more than 50 employees are more likely to provide OCG recommendation and higher the satisfaction rating more likelihood of getting a positive recommendation.

Similarly, by analysing the results of logistic regression, we infer that Customer type 1, Portfolio Management, Responsiveness and Satisfaction ratings are significantly driving the classification whether a firm will form a potential future partnership with OCG. Customer type 1 which are customers who are less than 1 year old are highly likely to not form partnership as compared to firms that have more than 5+ years of tenure with OCG who are highly likely to provide recommendation and hence form future partnership with OCG. In Ontario Region, there is highly probability that more firms will be driven to not form a partnership. If OCG receives better average ratings for Portfolio management and Satisfaction than it will be able to make more partnership with other firms in future.

Additionally, the classification generated using KNN utilizing Responsiveness and Satisfaction is giving us better results than that of logistic regression, as there are no False negative in the confusion matrix and provide us more accurate results for True positives.

By combining the above findings, it is inferred that overall Satisfaction ratings is the most significant variable that can be used to predict the likelihood of Recommendation and classification of forming the potential future Partnerships for OCG. Hence, OCG should focus more on Satisfying the overall client needs in order to enhance its consulting practice.

# References

A. (2018, December 24). Consumer Satisfaction Rates in Canada Decline amid Market Disruption and Today’s “Always On” Customer, Accenture Finds. Accenture. https://www.newswire.ca/news-releases/consumer-satisfaction-rates-in-canada-decline-amid-market-disruption-and-todays-always-on-customer-accenture-finds-517475681.html

Happiest Countries in the World 2022. (2022). World Population Review. https://worldpopulationreview.com/country-rankings/happiest-countries-in-the-world

Savitz, E. (2012, April 20). Customer Satisfaction By The Numbers: An Industry Breakdown. Forbes. https://www.forbes.com/sites/ciocentral/2012/04/19/customer-satisfaction-by-the-numbers-an-industry-breakdown/?sh=711085ab580f

Statistics Canada Client Survey 2019. (2022). StatsCan. https://www.statcan.gc.ca/en/consultation/2019/sccs2019

# Appendix

Variation in average ratings across Customer type

Graphical user interface

Description automatically generated

Variation in average ratings across Industry type

Graphical user interface, application, table, Word, Excel

Description automatically generated

Variation in average ratings across region

Graphical user interface, application, table

Description automatically generated

Variation in average ratings across Firm Size

Graphical user interface, application, table, Word, Excel

Description automatically generated

Bivariate analysis: Linear relations of dependent with independent variables

Diagram

Description automatically generated

Correlation Matrix

Table

Description automatically generated

Linear Regression Model

Iteration -1

Table

Description automatically generated

Iteration – 2

Table

Description automatically generated

Iteration – 3

Table

Description automatically generated

Iteration – 4

Table

Description automatically generated

Iteration – 5

Table

Description automatically generated

Variation of Partnership across the independent variables

Chart, bar chart, waterfall chart

Description automatically generated

Graphical user interface, application, table, Word, Excel

Description automatically generated

Logistic Regression Model

Iteration 1:

Table

Description automatically generated

Table

Description automatically generated

Table

Description automatically generated

Chart

Description automatically generated

Iteration 2:

Table

Description automatically generated

Table

Description automatically generated

Table

Description automatically generated

Chart

Description automatically generated

Iteration 3:

Table

Description automatically generated

Table

Description automatically generated

Table

Description automatically generated

Chart

Description automatically generated

KNN Classification:

Table

Description automatically generated with low confidence

Classification Tree:

Table

Description automatically generated