***Executive Summary***

***Case 1***

**Objective:**

Our objective is to maximize the profit by finding more members with high propensity to buy.

**JMP Model:**

With all predictors included, JMPprovide a model suggests mailing 100,000 members, the highest profit of training data is $728,750 and the highest profit of testing data is $658,500.

**Our Best Model:**

We built a partition-based model (Decision Tree) that identify the most important factors that predict an outcome and use the tree to make prediction for new observations.

Our Objective is to create the following formula:

If Prob(Success = =1) > 0.20, then DecisionBuy1 =1

Else, DecisionBuy1 =0

Marketing department has suggested to mail the Products only to maximum of 100,000 members. We propose to send the mail to 99,500 members based on our model.

In our analysis, we found that Gender, Recency, Art party and Pool party play a prominent role in generating higher predictive power i.e., profitability while considering for 99,500 members. Based on our model we get the profitability of $889,000. We did beat the benchmark of $700,000 as prescribed by the marketing department*.*

**Why our model is better?**

We tried 35+ variations with different combinations in each iteration to maximize profit and reduce the number of emails sent. We identified new variables by both using business functionality as well as coming up random variables that are not directly related to business domain, p-values, and necessarily splitting further or pruning to check if that makes any difference in terms of business perspective/profit identified. The predictive power of our model is 12.7% and for the testing model is 9.3%. While these KPIs could increase based on parameter selection but they failed to achieve maximum profitability as per our iterations.

**Conclusion**

Our Model Success Evaluation:

1. Training R2 of 0.126, and testing R2 of 0.093
2. Accuracy rate of 80% in predictions
3. Training profit = $925,750
4. Testing profit = $889,000

The deviation of training profit and testing profit is less than 100,000. These numbers are way higher than the benchmark data. However, further iterations can improve profitability and other statistical KPI requirements.

**JMP Model**

**Build the Decision tree Model using JMP (Go option) on the following conditions,**

**Y = Success**

**X = All** **predictors**

**Cutoff Probability for mailing = 0.18**

1. **Statistical KPIs of JMP Model – From JMP Printout**

|  |  |  |  |
| --- | --- | --- | --- |
| **Measure** | **Training** | **Validation** | **Definition** |
| **Entropy RSquare** | 0.1048 | 0.0792 | 1-Loglike(model)/Loglike(0) |
| **Generalized RSquare** | 0.1405 | 0.1059 | (1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n)) |
| **Mean -Log p** | 0.3139 | 0.3093 | ∑ -Log(ρ[j])/n |
| **RMSE** | 0.3004 | 0.297 | √ ∑(y[j]-ρ[j])²/n |
| **Mean Abs Dev** | 0.1805 | 0.1775 | ∑ |y[j]-ρ[j]|/n |
| **Misclassification Rate** | 0.112 | 0.105 | ∑ (ρ[j]≠ρMax)/n |
| **N** | 1000 | 1000 | n |

1. **Statistical KPIs of JMP Model – From Excel Printout**

|  |  |  |
| --- | --- | --- |
|  | **Training** | **Validation** |
| **Accuracy %** | 82.70% | 82.70% |
|  |  |  |
| **True Positive Rate** | 36.61% | 36.61% |
| **False Positive Rate** | 11.49% | 11.49% |
|  |  |  |
| **Sensitivity ( True Positive Rate)** | 36.61% | 36.61% |
| **Specificity (True Negative Rate)** | 88.51 | 88.51% |

**a)Business KPIs of JMP Model – Training**

|  |  |  |
| --- | --- | --- |
| Predicted number of Buyer | = | 71500 |
| Upper limit for packages sent | = | 100000 |
| Actual number of packages sent | = | 71500 |
|  |  |  |
| Propensity to buy the Package | = | 28.671% |
| Propensity to not buy the Package | = | 71.329% |
|  |  |  |
| Total Profit | = | $ 728,750 |

**b) Business KPIs of JMP Model – Testing**

|  |  |  |
| --- | --- | --- |
| Predicted number of Buyer | = | 70500 |
| Upper limit for packages sent | = | 100000 |
| Actual number of packages sent | = | 70500 |
|  |  |  |
| Propensity to buy the Package | = | 26.950% |
| Propensity tonot buy the Package | = | 73.050% |
|  |  |  |
| Total Profit | = | $ 658,500 |

1. **Interpret the Model (decision tree) – From Business Point of view & Statistical Point of view**

Target the 2 groups of people.

The first group is people whose Rencency less than 16, have bought Art Party designs more than once and have bought block party designs more than twice.

The second group is people whose Rencency less than 16, have bought Art Party designs more than once and have bought block party designs less than twice.

**Graphical user interface, application, table

Description automatically generated**

**iv) Confusion Matrix for Training**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Not Buyer | Buyer |  |
| Not Buyer | 786 | 102 | 888 |
| Buyer | 71 | 41 | 112 |
|  | 857 | 143 | 1000 |

**iv) Confusion Matrix for Testing**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Not Buyer | Buyer |  |
| Not Buyer | 792 | 103 | 895 |
| Buyer | 67 | 38 | 105 |
|  | 859 | 141 | 1000 |

**v) Lift Table**

|  |  |  |
| --- | --- | --- |
| **Lift Table in Propensity** | Training | Testing |
| Lift with respect to Baseline - JMP Model | 2.55994006 | 2.406281662 |
| Lift with respect to Baseline - My Best Model |  |  |

|  |  |  |
| --- | --- | --- |
| **Lift Table in Dollars** | Training | Testing |
| Lift with respect to Baseline - JMP Model | 4.71988342 | 4.264896373 |
| Lift with respect to Baseline - My Best Model |  |  |
| Lift with respect to JMP Model - My Contribution | 1.137041332 | 1.177333082 |
| Overall Lift with respect to Baseline -My Best Model |  |  |

**vi) Attach JMP Printout**

|  | **RSquare** | **N** | **Number of Splits** |
| --- | --- | --- | --- |
| Training | 0.105 | 1000 | 4 |
| Validation | 0.079 | 1000 |  |



**Split History**



**My Best Model**

Build the Decision tree Model using JMP on the following conditions,

Y = Success

X = Gender, Recency, Art party and Pool party

**Cutoff Probability for mailing = 0.20**

1. **Statistical KPIs of JMP Model – From JMP Printout**

|  |  |  |  |
| --- | --- | --- | --- |
| **Measure** | **Training** | **Validation** | **Definition** |
| **Entropy RSquare** | 0.1275 | 0.0935 | 1-Loglike(model)/Loglike(0) |
| **Generalized RSquare** | 0.1697 | 0.1244 | (1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n)) |
| **Mean -Log p** | 0.3060 | 0.3045 | ∑ -Log(ρ[j])/n |
| **RMSE** | 0.2973 | 0.2950 | √ ∑(y[j]-ρ[j])²/n |
| **Mean Abs Dev** | 0.1769 | 0.1769 | ∑ |y[j]-ρ[j]|/n |
| **Misclassification Rate** | 0.1060 | 0.1050 | ∑ (ρ[j]≠ρMax)/n |
| **N** | 1000 | 1000 | n |

**Statistical KPIs of JMP Model – From Excel Printout**

|  |  |  |
| --- | --- | --- |
|  | **Training** | **Validation** |
| **Accuracy %** | 80.1% | 80.00% |
|  |  |  |
| **True Positive Rate** | 47.32% | 49.52% |
| **False Positive Rate** | 15.77% | 16.42% |
|  |  |  |
| **Sensitivity ( True Positive Rate)** | 47.32% | 49.52% |
| **Specificity (True Negative Rate)** | 84.23% | 83.58% |

1. **a) Business KPIs of JMP Model – Training**

|  |  |  |
| --- | --- | --- |
| Predicted number of Buyer | = | 96500 |
| Upper limit for packages sent | = | 100000 |
| Actual number of packages sent | = | 96500 |
|  |  |  |
| Propensity to buy the Package | = | 27.461% |
| Propensity to not buy the Package | = | 72.539% |
|  |  |  |
| Total Profit | = | $ 925,750 |

**b) Business KPIs of JMP Model – Testing**

|  |  |  |
| --- | --- | --- |
| Predicted number of Buyer | = | 99500 |
| Upper limit for packages sent | = | 100000 |
| Actual number of packages sent | = | 99500 |
|  |  |  |
| Propensity to buy the Package | = | 26.131% |
| Propensity tonot buy the Package | = | 73.869% |
|  |  |  |
| Total Profit | = | $ 889,000 |

**iii) Interpret the Model (decision tree) – From Business Point of view & Statistical Point of view**

Target the 4 groups of people.

* The first group is males whose Recency less than 8, have bought Art Party designs more than once,
* The second group is males whose Recency between 8 & 16, have bought Art Party designs more than once,
* The third group is females with Recency less than 16, have bought Art Party designs more than once,
* The fourth group is males with Recency between 8 & 16, have not bought Art Party design.

**Graphical user interface, table

Description automatically generated**

1. **Confusion Matrix for Training**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Not Buyer | Buyer |  |
| Not Buyer | 748 | 140 | 888 |
| Buyer | 59 | 53 | 112 |
|  | 807 | 193 | 1000 |

1. **Confusion Matrix for Testing**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Not Buyer | Buyer |  |
| Not Buyer | 748 | 147 | 895 |
| Buyer | 53 | 52 | 105 |
|  | 801 | 199 | 1000 |

**v) Lift Table**

|  |  |  |
| --- | --- | --- |
| **Lift Table in Propensity** | Training | Testing |
| Lift with respect to Baseline - JMP Model | 2.55994006 | 2.406281662 |
| Lift with respect to Baseline - My Best Model | 2.451887491 | 2.333094042 |

|  |  |  |
| --- | --- | --- |
| **Lift Table in Dollars** | Training | Testing |
| Lift with respect to Baseline - JMP Model | 4.71988342 | 4.264896373 |
| Lift with respect to Baseline - My Best Model | 5.995790155 | 5.757772021 |
| Lift with respect to JMP Model - My Contribution |  |  |
| Overall Lift with respect to Baseline -My Best Model |  |  |

**vi) Attach JMP Printout**

|  | **RSquare** | **N** | **Number of Splits** |
| --- | --- | --- | --- |
| Training | 0.127 | 1000 | 8 |
| Validation | 0.093 | 1000 |  |



**Split History**

**Chart, line chart

Description automatically generated**

**Key information 🡪 Cutoff Probability for mailing = 0.20**