Table of Content

Assignment Cover Sheet I

Introduction 1

Literature Review 1

Business understanding 2

Data Understanding 4

Data Preparation 6

Modelling: 7

‘Visit’ Model: 8

‘Spend’ Model: 8

Evaluation: 8

‘Visit’ Model: 8

‘Spend’ Model 10

Implementation 10

Final Results and Conclusion: 10

Reference 12

Appendix 1. Calculating expected profit and expected revenue from Universal Plus’ existing Email Marketing system and newly developed Email Marketing system 13

Introduction

Nowadays, digital marketing has developed into a crucial tool to boost activity and revenue in every organisation, and companies are now using marketing initiatives to strengthen their interactions with clients in a diverse range of commercial sectors. Direct marketing, by sending emails or messages containing newsletters, promotions, or seasonal coupons and discounts to specific groups of targeted users to drive sales (Bitter et al., 2021), is a powerful marketing tool. In this report, the CRISP-DM process will be followed to maximise the efficiency of email marketing through machine learning models.

Literature Review

For analysing large amounts of data to support objective marketing decisions, the cross-industry standard process for data mining (CRISP-DM) methodology is becoming the most favoured one in recent times. The paper drafted by Huifang Zeng and Ding Pan talks about the six phases of this methodology, that is, Business Understanding, Data Understanding and Preparation, Modelling, Evaluation and Deployment. Gaining knowledge about the CRISP-DM method to identify customers' future purchasing behaviour is vital to achieve the flow for the given business problem.

While intently focusing on the business objectives of "Universal Plus", we moved towards choosing relevant features from the given data that could help us equip our model. With the help of the research paper drafted by Derksen, S. and Keselman, H.J., various wrapper methods were discovered for feature selection and the comparison of all the wrapper techniques according to the proportion of noise variables selected.

The systematic approach towards data understanding and preparation helped us gain insight into the different machine-learning techniques that could be utilized to generate increased sales rates through effective email marketing (Redouan Abakouy, El Mokhtar EN-Naimi and Anass El Haddadi (2017)). This paper focuses on the growing importance of targeted email marketing campaigns and also cares about the efficient deliverability of the email for a higher conversion rate. Various classification and regression algorithms like Decision Trees, Artificial Neural Networks, SVM, and KNN were used for the predictive analytics.

The machine learning techniques discussed above are useful in predicting customers' behaviour towards direct marketing, or email marketing, as described in the article by Bitter, Tercan, et al (2021). Notably, the article suggests that while training the SVM model, a high volume of data causes longer training time. Bitter recommended evaluating the model from a business perspective, considering the potential profit expected from the user and the cost of delivering the campaign, which triggered further analysis.

Another supervised learning technique that played a major role in our analysis of data was the Random Forest (RF). The primary strength of the Random Forest (RF) model lies in maintaining accuracy even with inconsistent data and this method is easy to use, fast and robust (Gupta and Gupta, 2019). After reading this study thoroughly, it could be concluded that the RF model works well compared with the neural networks model, in this case, reinforcing the confidence to run the RF model for our data analysis.

The most important step of the CRISP-DM methodology is the model evaluation. Many classification models like the decision tree are usually followed by data collection and Sample Characteristics, Coding Attributes, and Coding Reliability Testing (Sann, Raksmey, et al, 2022). That means omitting the irrelative data and transferring the data to meaningful knowledge via a decoding mechanism. Overall accuracy, precision AUC, recall, specificity, and F-measure is used as an indicator to measure the decision tree's performance (ibid). Usually, the higher index implies better classification quality, indeed Area under Curve (AUC) greater than 0.8 is good (ibid).

Business understanding

Universal plus requested to develop and implement an email marketing system to target their customers effectively to improve sales and reduce marketing expenses. Before this, their email marketing was primarily carried out either at random or according to established rules of thumb, and the consequences were unsatisfactory. Henceforth, the business goal of this project is to create an email marketing system that takes targeting decisions for customers in a way that maximizes visits and overall profitability.

Before building the system, resources, risks, and contingencies must be assessed for both the client and the analyst firm. Companies may face the additional expense of inappropriate selection of analytics firms (for example, if the chosen model by the analytics firm cannot produce accurate prediction). In practice, businesses may also encounter contingency situations such as power outages, the inability of the analytical firm to complete projects according to the schedule, etc. Moreover, the analyst firm may encounter problems in its data analysis. For instance, clients may provide incomplete or non-standardised data records, increasing the analysis workload. Inevitably, if the client does not choose the analyst firm's proposal, the analyst firm will face the sunk cost of efforts. In terms of contingency, the analyst firm may encounter unexpected changes in deadlines or additional project requirements from the client. The analyst firm is obliged to take contingency measures, such as flexible changes to the schedule and applying agile development processes.

The approach to building this email marketing system is to use two machine learning models, one that will predict whether the customer will visit the store and another that will predict how much the customer will spend given he/she visits the store.

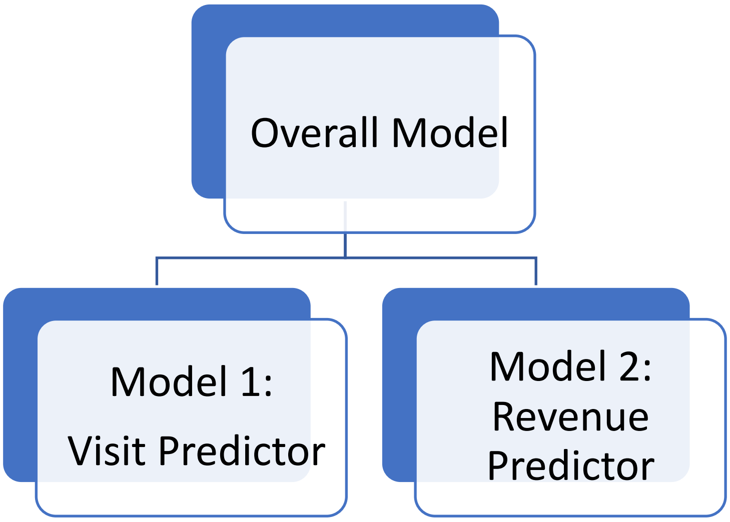
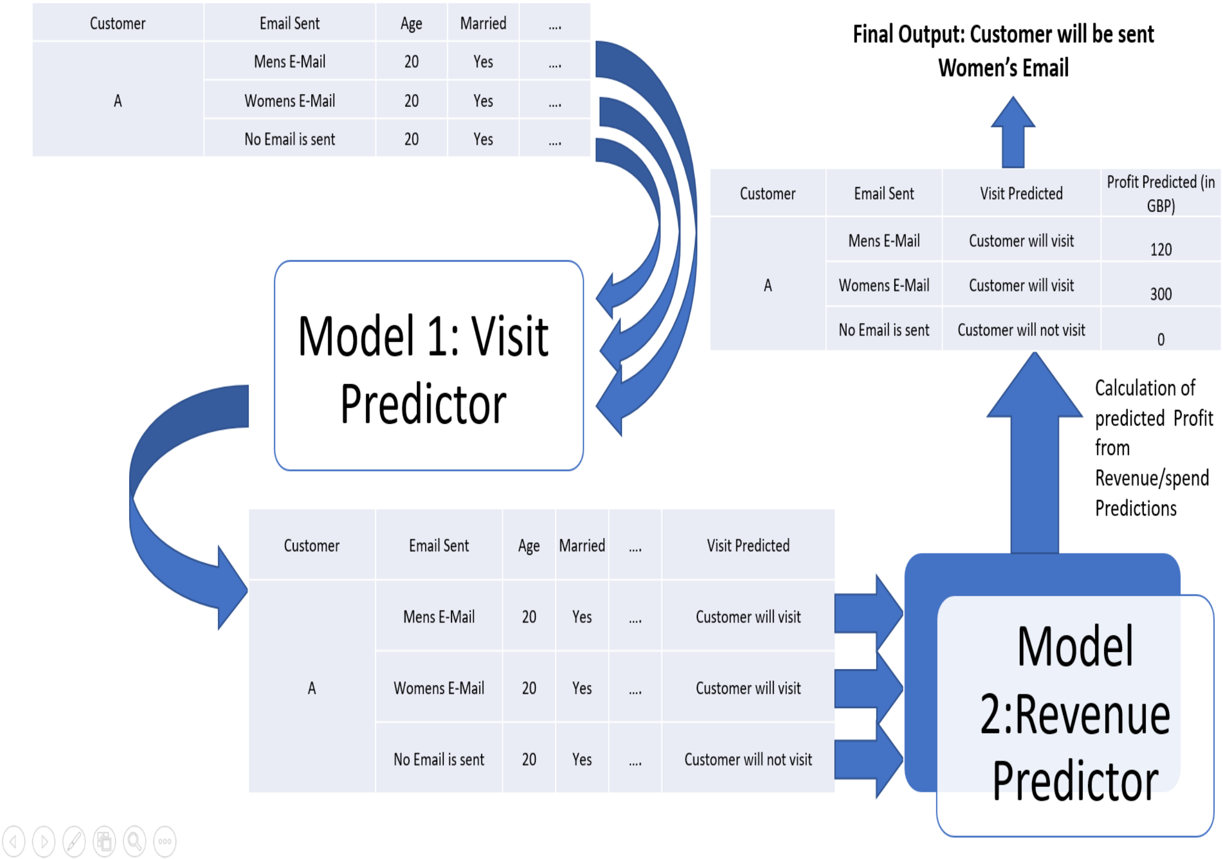


Figure 1: Flowchart of the Overall Model

Once the best model is selected for each, these two models should be used to determine what targeting decision should be taken for each customer to maximize visits and profitability.

Figure 2 illustrates the approach that was envisioned for achieving the business goal for this project. For each customer, for each of the three scenarios, the selected visit model should predict whether the customer will visit. The output should then be passed as input to the selected spend model, which will predict how much the customer will spend in response to each targeting decision. The predicted spend values will be converted to predicted profit values by subtracting the cost of sending an email, if applicable. The targeting decision that maximizes profit from the customer should be selected by this email marketing system.



Customer	Email Sent	Age	Married	....
A	Mens E-Mail	20	Yes	....
	Womens E-Mail	20	Yes	....
	No Email is sent	20	Yes	....

Model 1: Visit Predictor

Customer	Email Sent	Age	Married	....	Visit Predicted
A	Mens E-Mail	20	Yes	....	Customer will visit
	Womens E-Mail	20	Yes	....	Customer will visit
	No Email is sent	20	Yes	....	Customer will not visit

Model 2:Revenue Predictor

Calculation of predicted Profit from Revenue/spend Predictions

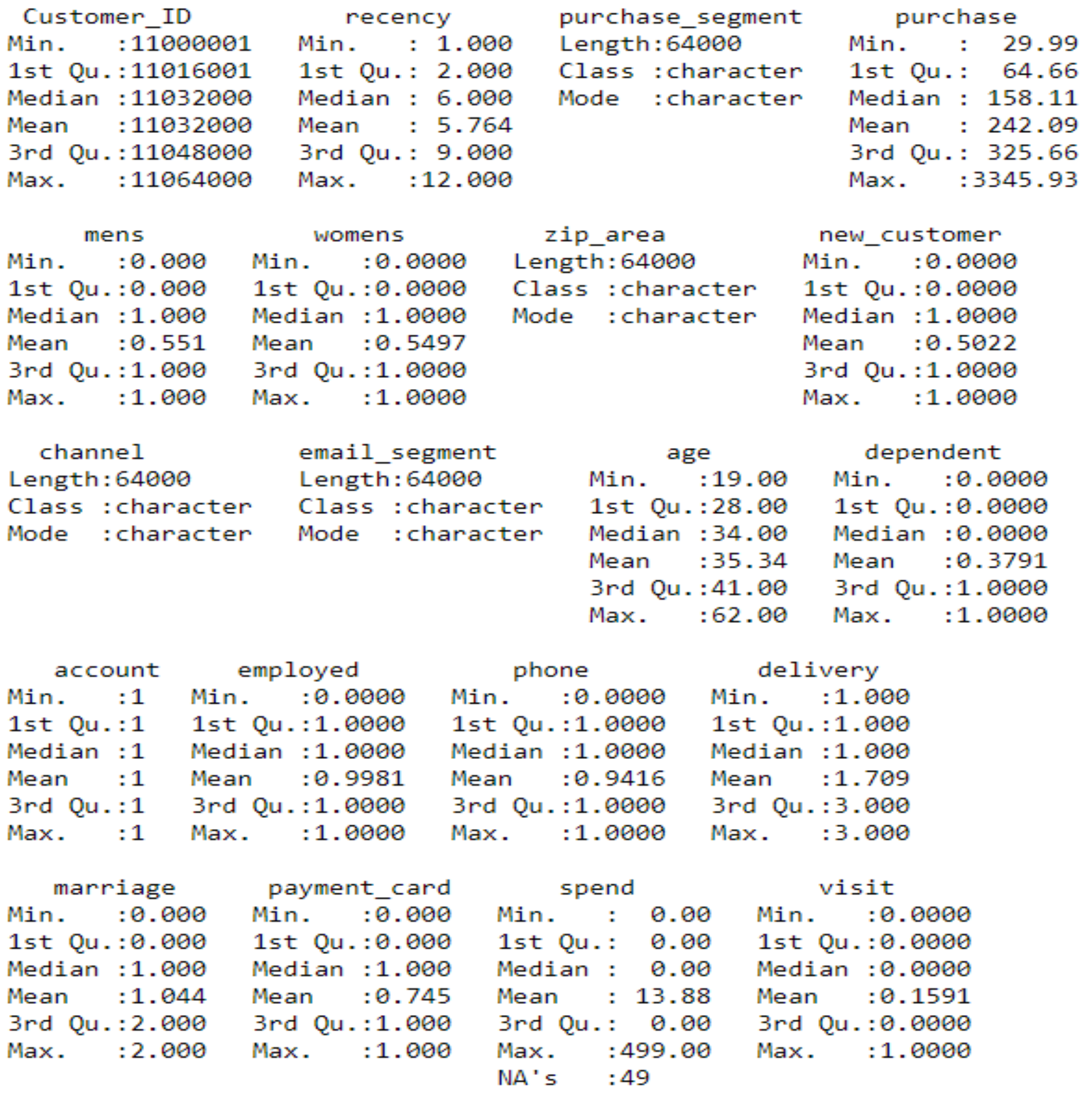
Customer	Email Sent	Visit Predicted	Profit Predicted (in GBP)
A	Mens E-Mail	Customer will visit	120
	Womens E-Mail	Customer will visit	300
	No Email is sent	Customer will not visit	0

Final Output: Customer will be sent Women's Email

Figure 2: Explanation of the envisioned data mining approach

Data Understanding

The data analysed in this report is collected from all Universal Plus customers after an email marketing campaign, containing 64000 records and 20 variables.



```
  Customer_ID           recency       purchase_segment       purchase      
 Min.   :11000001   Min.   : 1.000   Length:64000       Min.   :  29.99  
 1st Qu.:11016001   1st Qu.: 2.000   Class :character   1st Qu.:  64.66  
 Median :11032000   Median : 6.000   Mode  :character   Median : 158.11  
 Mean   :11032000   Mean   : 5.764                      Mean   : 242.09  
 3rd Qu.:11048000   3rd Qu.: 9.000                      3rd Qu.: 325.66  
 Max.   :11064000   Max.   :12.000                      Max.   :3345.93  

      mens           womens          zip_area          new_customer   
 Min.   :0.000   Min.   :0.0000   Length:64000       Min.   :0.0000  
 1st Qu.:0.000   1st Qu.:0.0000   Class :character   1st Qu.:0.0000  
 Median :1.000   Median :1.0000   Mode  :character   Median :1.0000  
 Mean   :0.551   Mean   :0.5497                      Mean   :0.5022  
 3rd Qu.:1.000   3rd Qu.:1.0000                      3rd Qu.:1.0000  
 Max.   :1.000   Max.   :1.0000                      Max.   :1.0000  

   channel          email_segment            age          dependent     
 Length:64000       Length:64000       Min.   :19.00   Min.   :0.0000  
 Class :character   Class :character   1st Qu.:28.00   1st Qu.:0.0000  
 Mode  :character   Mode  :character   Median :34.00   Median :0.0000  
                                       Mean   :35.34   Mean   :0.3791  
                                       3rd Qu.:41.00   3rd Qu.:1.0000  
                                       Max.   :62.00   Max.   :1.0000  

    account     employed          phone           delivery    
 Min.   :1   Min.   :0.0000   Min.   :0.0000   Min.   :1.000  
 1st Qu.:1   1st Qu.:1.0000   1st Qu.:1.0000   1st Qu.:1.000  
 Median :1   Median :1.0000   Median :1.0000   Median :1.000  
 Mean   :1   Mean   :0.9981   Mean   :0.9416   Mean   :1.709  
 3rd Qu.:1   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:3.000  
 Max.   :1   Max.   :1.0000   Max.   :1.0000   Max.   :3.000  

    marriage      payment_card        spend             visit       
 Min.   :0.000   Min.   :0.000   Min.   :  0.00   Min.   :0.0000  
 1st Qu.:0.000   1st Qu.:0.000   1st Qu.:  0.00   1st Qu.:0.0000  
 Median :1.000   Median :1.000   Median :  0.00   Median :0.0000  
 Mean   :1.044   Mean   :0.745   Mean   : 13.88   Mean   :0.1591  
 3rd Qu.:2.000   3rd Qu.:1.000   3rd Qu.:  0.00   3rd Qu.:0.0000  
 Max.   :2.000   Max.   :1.000   Max.   :499.00   Max.   :1.0000  
                                 NA's   :49                       
```

Figure 3: Summary of Imported Dataset

The necessary data cleaning steps were identified after understanding the summary. Moreover, It was also seen that the spend values were always 0 when visit = 0, implying that spending cannot happen without visiting the store. It was also important to check for Outliers in spend column, as existence of outliers will affect our machine learning models' evaluation metrics adversely. The spend column was filtered for visit =1, and was visualized. When checked, 0.6% of spend column values were upper bound outliers. However there were no lower bound outliers.

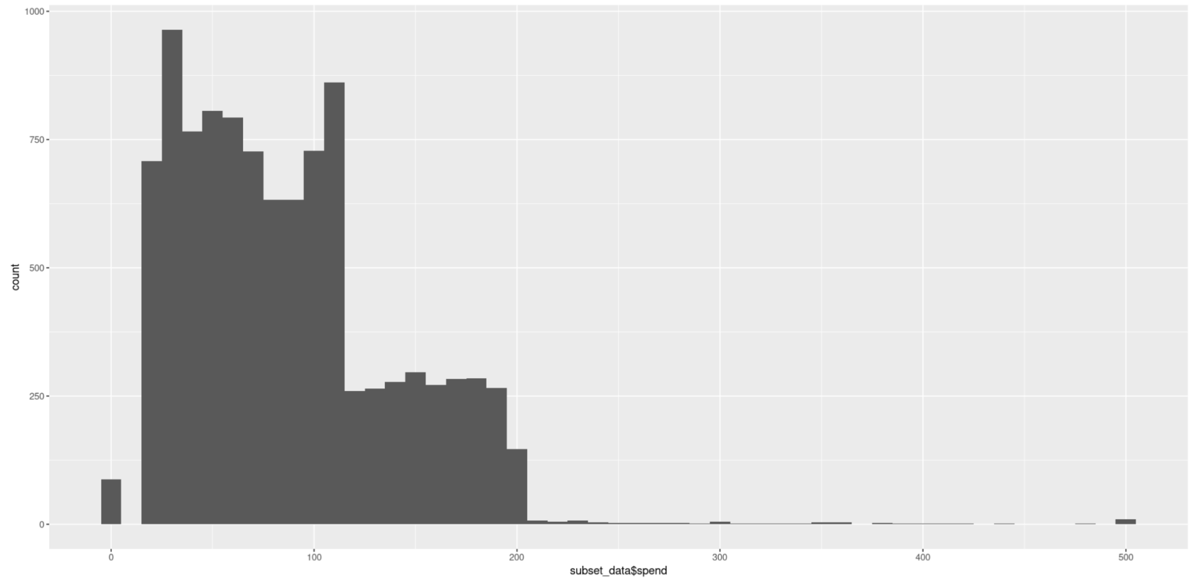
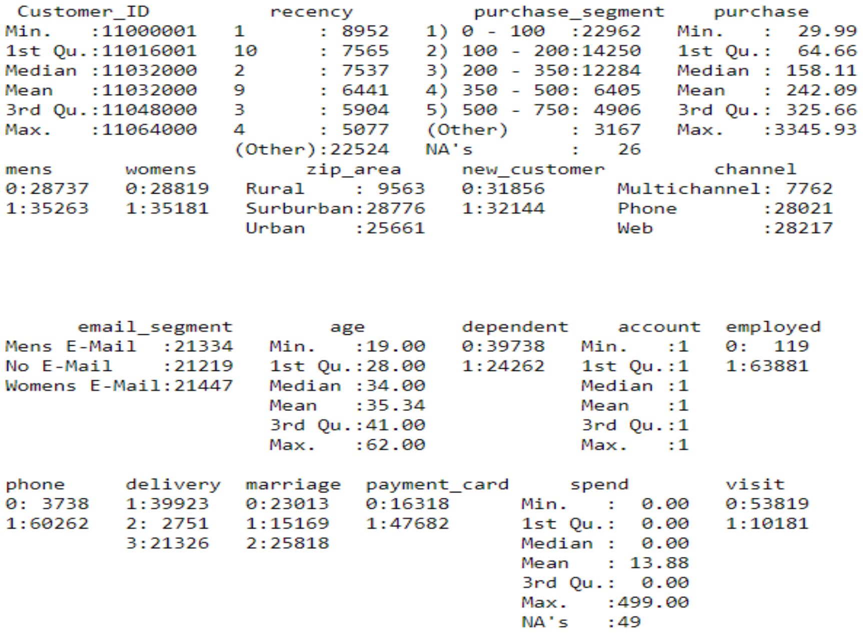


Figure 4: Frequency Distribution of ‘Spend’ column

Data Preparation

All variables of categorical nature were converted to factors. After checking the summary of the updated data, it was seen that purchase segment column had NA values, which were corrected based on values in purchase column.



```
  Customer_ID          recency        purchase_segment     purchase
 Min.   :11000001   1      : 8952   1) 0 - 100  :22962   Min.   :  29.99
 1st Qu.:11016001   10     : 7565   2) 100 - 200:14250   1st Qu.:  64.66
 Median :11032000   2      : 7537   3) 200 - 350:12284   Median : 158.11
 Mean   :11032000   9      : 6441   4) 350 - 500: 6405   Mean   : 242.09
 3rd Qu.:11048000   3      : 5904   5) 500 - 750: 4906   3rd Qu.: 325.66
 Max.   :11064000   4      : 5077   (Other)     : 3167   Max.   :3345.93
                    (Other):22524   NA's        :   26
 mens       womens          zip_area      new_customer         channel
 0:28737    0:28819    Rural     : 9563   0:31856        Multichannel: 7762
 1:35263    1:35181    Surburban:28776    1:32144        Phone       :28021
                       Urban     :25661                  Web         :28217



      email_segment          age         dependent      account     employed
 Mens E-Mail  :21334   Min.   :19.00   0:39738    Min.   :1   0:  119
 No E-Mail    :21219   1st Qu.:28.00   1:24262    1st Qu.:1   1:63881
 Womens E-Mail:21447   Median :34.00              Median :1
                       Mean   :35.34              Mean   :1
                       3rd Qu.:41.00              3rd Qu.:1
                       Max.   :62.00              Max.   :1

 phone      delivery   marriage   payment_card      spend          visit
 0: 3738    1:39923    0:23013    0:16318       Min.   :  0.00   0:53819
 1:60262    2: 2751    1:15169    1:47682       1st Qu.:  0.00   1:10181
            3:21326    2:25818                  Median :  0.00
                                                Mean   : 13.88
                                                3rd Qu.:  0.00
                                                Max.   :499.00
                                                NA's   :49
```

Figure 5: Summary of updated dataset

The outliers in the spend column were replaced with the IQR upper-bound value.

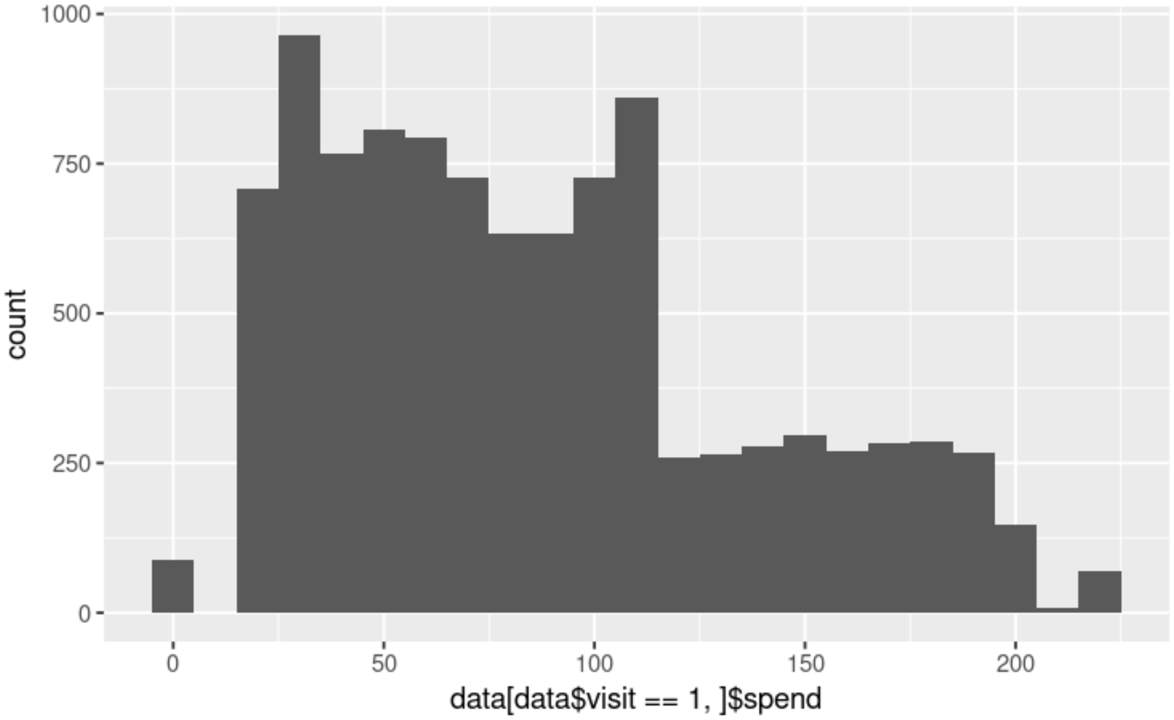


Figure 6: Frequency Distribution of 'Spend' after treating outliers

After initial pre-processing, two different datasets were created, one dataset (‘visit’ data) with the visit column without spend column, and the other dataset (‘spend’ data) which had spent column but not visit column, filtered with those rows wherein the visit were 1. This was done to build the two models, one for predicting visits, which will be built on the ‘visit’ data, and the other for predicting spend, which will be built on the ‘spend’ data.

The continuous variables in the spend data were normalized using Min-Max normalization technique so as to remove the effect of scale of the variables on our regression models. Using this technique, all the variables will be ranging from 0 to 1.

The ‘visit’ data and ‘spend’ data were further split into training and testing data at 70:30 standard split using stratified sampling. For visit data, stratified sampling held significance as this proportionality of 0s and 1s in both training and testing data was retained.

For data balancing, the oversampling method was applied on the training data derived from the visit data, to not delete any possibly valuable information.

For the training data derived from visit data, ultimately, some redundant variables were removed via ranking the information gain value to avoid overfitting and increase the generalization in final predictions.

A dataset having 3 duplicate records for each customer ID was also created, such that only the email segment values were different. This dataset will be used for the deployment of the models that would be built to determine which email should be sent to which customers.

Modelling:

'Visit' Model:

The team decided to put Random Forest (RF), Support Vector Machines (SVM), Logistic Regression (LogReg) and Decision Trees into use. The RF model was a powerful method for running efficiently on large databases, which was one significant feature of our dataset. However, its weakness was also distinct. If there were many wrong trees, the vote result may be unreliable and lead to inaccuracy of the final prediction (Gupta and Gupta, 2019). Decision Trees was easy to understand and interpret and could process both numerical and categorical data (Sann, Raksmey, et al, 2022). Meanwhile, when the tree was overly large, branches should be pruned back. In addition, both Log Reg and SVM were linear discriminant functions. The former worked well under structured data and the latter could handle abundant inputs. Nevertheless, Log Reg was sensitive to outliers and SVM was slow when running the model. Under precise comparison, the four models were suitable for our current data.

'Spend' Model:

Initially, Feature Selection using the subset wrapper method was conducted on the Multiple Linear Regression (MLR) model to select those features which increased the linear regression Adjusted R2 value the most while also reducing overfitting.

5 ML Models were trained first including all the features and then were trained separately only on those features selected by our wrapper feature selection algorithm described above. In this way, we got two models for each kind, one trained on all the features and the other one trained on the features selected by the wrapper method. The models used namely linear regression, SVR, RF regression, Decision Tree regression and KNN respectively.

Evaluation:

'Visit' Model:

All the models were tested on the testing data and ROC curves for each model were plotted.

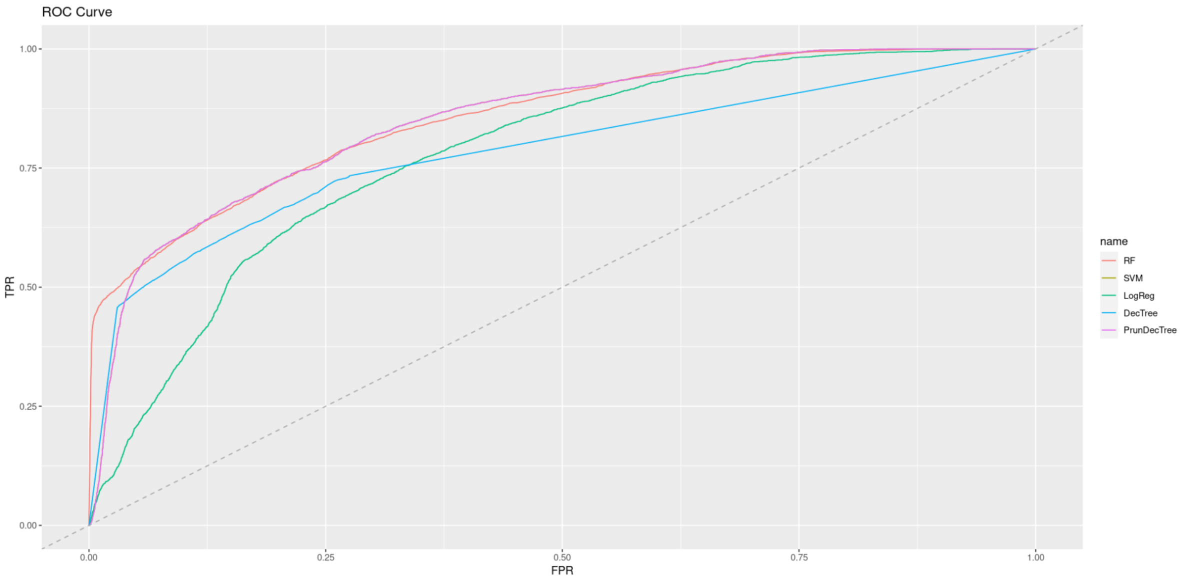


Figure 7: ROC Curves for all models

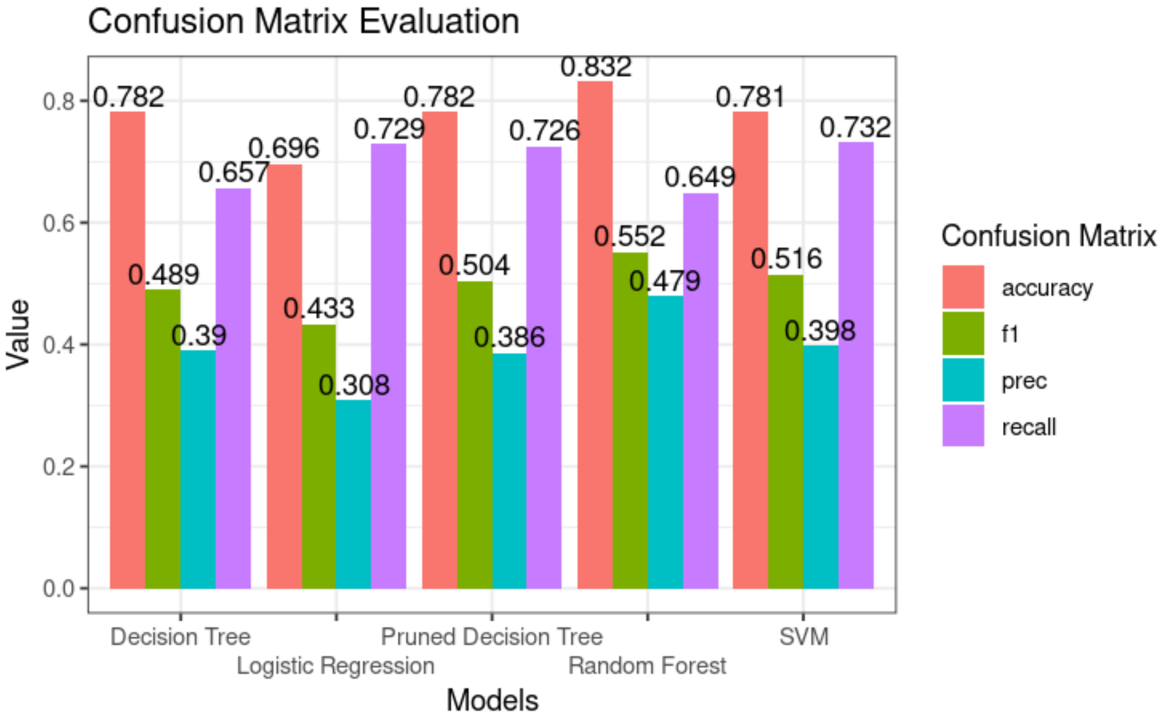
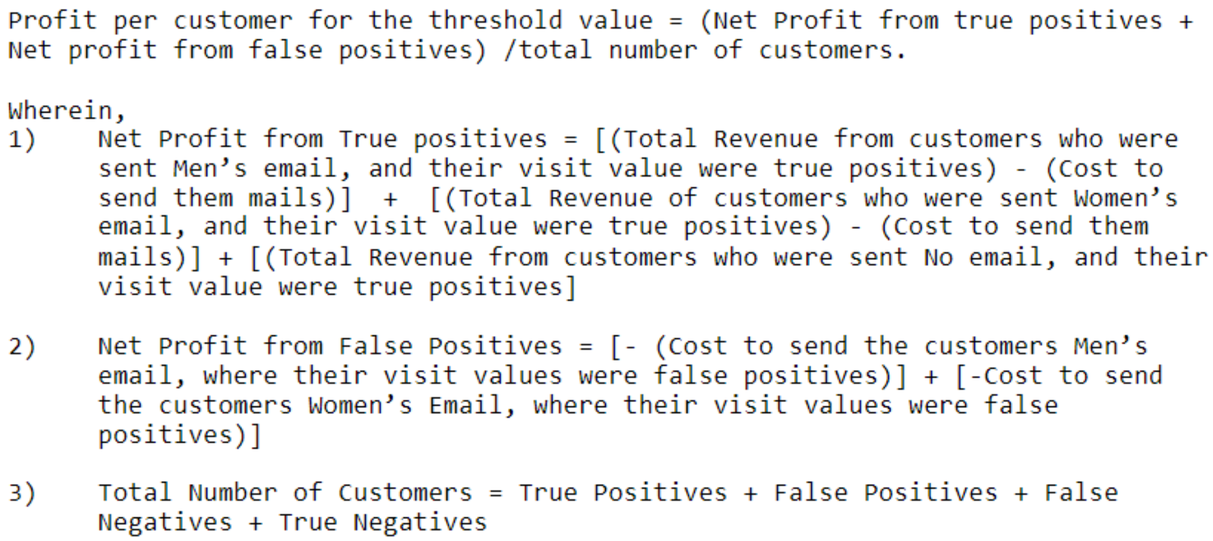


Figure 8: Confusion Matrix Evaluation for all models at 0.5 threshold value

The aim was to select that the model that generated the maximum expected profit in the test data at a particular optimum threshold value. Equation 1 was used to determine the expected profit at a particular threshold value:



Profit per customer for the threshold value = (Net Profit from true positives + Net profit from false positives) /total number of customers.

Wherein,

1) Net Profit from True positives = [(Total Revenue from customers who were sent Men’s email, and their visit value were true positives) - (Cost to send them mails)] + [(Total Revenue of customers who were sent Women’s email, and their visit value were true positives) - (Cost to send them mails)] + [(Total Revenue from customers who were sent No email, and their visit value were true positives]

2) Net Profit from False Positives = [- (Cost to send the customers Men’s email, where their visit values were false positives)] + [-Cost to send the customers Women’s Email, where their visit values were false positives)]

3) Total Number of Customers = True Positives + False Positives + False Negatives + True Negatives

Equation 1: Expected Value of Profit per Customer

The cost of 1 email was assumed as 1 GBP. Based on the cost of email, the model that generated the maximum expected profit was SVM, at a threshold value of 0.04. To prove SVM’s feasibility, the team checked the AUC (Area under ROC Curve) value. The SVM model’s AUC value was 0.8515, which was a great indicator of a feasible model.

‘Spend’ Model

The linear regression model had the least RMSE value for predicting the normalized spend (RMSE: 0.1614195) amongst all the other models. Hence this model was selected for predicting spend.

Implementation

The SVM model was used to predict whether the customer will visit or not in each of the scenarios, and correspondingly multiple linear regression was used to predict spending for those scenarios where it was predicted that customers will visit. For those scenarios where it was predicted that customers will not visit, the prediction for spend was made equal to 0. However, the spend predictions that the multiple linear regression model gave were scaled values ranging from 0 to 1, and hence those predicted values were converted back to their actual scale by reversing the min-max scale equation, to arrive at the actual values of spend. These values of spend for each record were subtracted with the cost of sending 1 email, wherever the records show that a particular mail was sent. The resulting values were the profit associated with each row.

Amongst every three duplicated rows for each customer that had different email segment values, a code was constructed that filtered that row having the maximum profit value. The resultant table was our Model's final output. The Email Segment column represented what targeting decision our Model had selected for each customer.

Final Results and Conclusion:

We calculated various Key Performance Indicators, firstly for the company's existing randomized targeting system and for our newly developed Email Marketing System using the approach and formulae described in Appendix 1. Using our Email Marketing System, it was found that the company can expect to increase its Expected Profit, Expected Revenue and Expected Visits by 15%, 17% and 16% respectively. It was henceforth concluded that the Email Marketing system our team built was successful in meeting Universal Plus' business objectives.

Reference

Abakouy, R., En-Naimi, E.M. and El Haddadi, A. (2017) "Classification and prediction based data mining algorithms to predict email marketing campaigns," *Proceedings of the 2nd International Conference on Computing and Wireless Communication Systems - ICCWCS'17*, pp. 1–5. Available at: https://doi.org/10.1145/3167486.3167520.

Bitter, C., Tercan, H., Meisen, T. et al. (2021) "When to Message: Investigating User Response Prediction with Machine Learning for Advertisement Emails," *2021 4th International Conference on Artificial Intelligence for Industries (AI4I).* IEEE. Available at: https://ieeexplore.ieee.org/abstract/document/9565531 (Accessed: November 16, 2022).

Derksen, S. and Keselman, H.J. (1992), "Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables". *British Journal of Mathematical and Statistical Psychology*, vol.45, pp.265-282. Available at: https://0-doi-org.pugwash.lib.warwick.ac.uk/10.1111/j.2044-8317.1992.tb00992.x (Accessed: December 2, 2022)

Gupta, A. and Gupta, G. (2018) "Comparative study of random forest and neural network for prediction in direct marketing," *Applications of Artificial Intelligence Techniques in Engineering. Advances in Intelligent Systems and Computing*, 697, pp. 401–410. Springer, Singapore. Available at: https://doi.org/10.1007/978-981-13-1822-1_37. (Accessed: November 20, 2022).

Sann, R., Lai, P., Liaw, S. & Chen, C. (2022) "Predicting online complaining behavior in the hospitality industry: Application of Big Data Analytics to online reviews," *Sustainability*, 14(3), p. 1800. Available at: https://doi.org/10.3390/su14031800. (Accessed: November 22, 2022).

Zeng, H. and Pan, D. (2010) "A knowledge discovery and data mining process model in e-marketing," *2010 8th World Congress on Intelligent Control and Automation [Preprint].* Available at: https://doi.org/10.1109/wcica.2010.5553834.

Appendix 1. Calculating expected profit and expected revenue from Universal Plus' existing Email Marketing system and newly developed Email Marketing system

Assume that the following table depicts the predictions of visits and correspondingly spend, for each of the scenarios for a particular customer with customer ID 1011. The following are the scenarios: When 1) Email Segment = "Men's E-Mail" 2) Email Segment = "Women's E-Mail" and 3) Email Segment = "No Email".

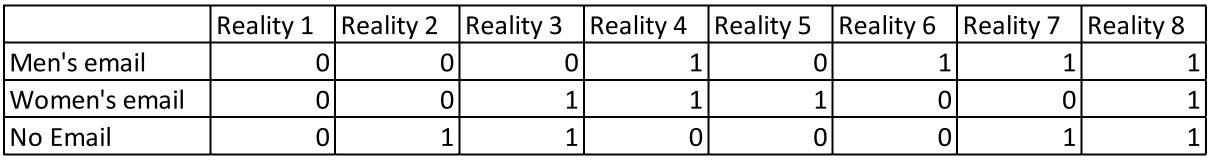
Customer ID	Email Segment	Probability for selecting email segment by the company's existing random pick model	Actual Model's visit prediction (as a proxy for what will actually happen)	Model's Spend prediction (as a proxy of what will actually happen)
1001	Men's email	0.33	1	100
1001	Women's email	0.33	1	120
1001	No Email	0.33	0	0
Base Case (if all the visits are 1s, what would the spend prediction be for each)				
Customer ID	Email Segment	Probability for selecting email segment randomly by the company	If all visits are 1s	What does the model predict for spend
1001	Men's email	0.33	1	100
1001	Women's email	0.33	1	120
1001	No Email	0.33	1	20

Assuming that the following ratios remain constant throughout the future:

1) Probability of the outcome of visit is actually 0 given that the model predicts 1 (Conditional Probability) =FP/(FP + TP)
2) Probability of the outcome of visit is actually 1 given that the model predicts 1 (Conditional Probability) = TP/(FP + TP)
3) Probability of the outcome of visit is actually 1 given that the model predicts 0 (Conditional Probability) = FN/(FN + TN)
4) Probability of the outcome of visit is actually 0 given that the model predicts 0 (Conditional Probability) = TN/(FN + TN)

We know that matrix of Prediction of visits according to Visit Model= $\begin{matrix}1\\1\\0\end{matrix}$

In spite of this, 8 possible actual realities can be there, which are the following:



	Reality 1	Reality 2	Reality 3	Reality 4	Reality 5	Reality 6	Reality 7	Reality 8
Men's email	0	0	0	1	0	1	1	1
Women's email	0	0	1	1	1	0	0	1
No Email	0	1	1	0	0	0	1	1

The probability of each of the outcome is the following, based on the prediction we got from our model:

Outcome 1)

Prediction	Actual Outcome	Probability of the actual outcome to happen given that model predicted 1,1,0
1	0	1) FP/(FP + TP)
1	0	1) FP/(FP + TP)
0	0	4) TN/(FN + TN)
total probability for this actual outcome to happen		1) * 1) * 4)

Outcome 2)

Prediction	Actual Outcome	Probability of the actual outcome to happen given that model predicted 1,1,0
1	0	1) FP/(FP + TP)
1	0	1) FP/(FP + TP)
0	1	3) FN/(FN + TN)
total probability for this actual outcome to happen		1) * 1) * 3)

Outcome 3)

Prediction	Actual Outcome	Probability of the actual outcome to happen given that model predicted 1,1,0
1	0	1) FP/(FP + TP)
1	1	2) TP/(FP + TP)
0	1	3) FN/(FN + TN)
total probability for this actual outcome to happen		1) * 2) * 3)

Outcome 4

Prediction	Actual Outcome	Probability of the actual outcome to happen given that model predicted 1,1,0
1	1	2) TP/(FP + TP)
1	1	2) TP/(FP + TP)
0	0	4) TN/(FN + TN)
total probability for this actual outcome to happen		2) * 2) * 4)

Outcome 5

Prediction	Actual Outcome	Probability of the actual outcome to happen given that model predicted 1,1,0
1	0	1) FP/(FP + TP)
1	0	1) FP/(FP + TP)
0	1	4) FN/(FN + TN)
total probability for this outcome to happen		1) * 1) * 4)

Outcome 6

Prediction	Actual Outcome	Probability of the actual outcome to happen given that model predicted 1,1,0
1	1	2) TP/(FP + TP)
1	0	1) FP/(FP + TP)
0	0	3) TN/(FN + TN)

total probability for this outcome to happen	2) * 1) * 3)

Outcome 7

Prediction	Actual Outcome	Probability of the actual outcome to happen given that model predicted 1,1,0
1	1	2) TP/(FP + TP)
1	0	1) FP/(FP + TP)
0	1	4) FN/(FN + TN)
total probability for this outcome to happen		2) * 1) * 4)

Outcome 8

Prediction	Actual Outcome	Probability of the actual outcome to happen given that model predicted 1,1,0
1	1	2) TP/(FP + TP)
1	1	2) TP/(FP + TP)
0	1	4) FN/(FN + TN)
total probability for this outcome to happen		2) * 2) * 4)

New Email Marketing System KPIs:

1) Expected Profit from New Email marketing System from this customer= Probability of outcome 1 * (0-cost of sending email) + Probability of outcome 2 * (0-cost of sending email) + Probability of outcome 3 * (120-cost of sending email) + Probability of outcome 4*(120-cost of sending email) + Probability of outcome 5*(0-cost of sending email) + Probability of outcome 6*(0-cost of sending email) + Probability of outcome 7*(0-cost of sending email) + Probability of outcome 8*(120-cost of sending email)

2) Expected Revenue from New Email Marketing System from this customer= Probability of outcome 1 * (0) + Probability of outcome 2 * (0) + Probability of outcome 3 * (120) + Probability of outcome 4*(120) + Probability of outcome 5*(0) + Probability of outcome 6*(0) + Probability of outcome 7*(0) + Probability of outcome 8*(120)

Old Email Marketing System KPIs:

1) Expected Profit from existing E-Mail Marketing System from this customer= 0.33 * [Probability of outcome 1 * (Profit from sending Men's Email in this actual outcome) + Probability of outcome 2 * (Profit from sending Men's Email in this actual outcome) + Probability of outcome 3 * (Profit from sending Men's Email in this actual outcome) + Probability of outcome 4*(Profit from sending Men's Email in this actual outcome) + Probability of outcome 5*(Profit from sending Men's Email in this actual outcome) + Probability of outcome 6*(Profit from sending Men's Email in this actual outcome) + Probability of outcome 7*(Profit from sending Men's Email in this actual outcome) + Probability of outcome 8*(Profit from sending Men's Email in this actual outcome)]

   +

   0.33 x [Probability of outcome 1 * (Profit from sending Women's Email in this actual outcome) + Probability of outcome 2 * (Profit from sending Women's Email in this actual outcome) + Probability of outcome 3 * (Profit from sending Women's Email in this actual outcome) + Probability of outcome 4*(Profit from sending Women's Email in this actual outcome) + Probability of outcome 5*(Profit from sending Women's Email in this actual outcome) + Probability of outcome 6*(Profit from sending Women's Email in this actual outcome) + Probability of outcome 7*(Profit from sending Women's Email in this actual outcome) + Probability of outcome 8*(Profit from sending Women's Email in this actual outcome)]

   +

   0.33 x [Probability of outcome 1 * (Profit from sending No Email in this actual outcome) + Probability of outcome 2 * (Profit from sending No Email in this actual outcome) + Probability of outcome 3 * (Profit from sending No Email in this actual outcome) + Probability of outcome 4*(Profit from sending No Email in this actual outcome) + Probability of outcome 5 Profit from sending No Email in this actual outcome) + Probability of outcome 6*(Profit from sending No Email in this actual outcome) + Probability of outcome 7*(Profit from sending No Email in this actual outcome) + Probability of outcome 8*(Profit from sending No Email in this actual outcome)]

2) Expected Revenue from existing Email Marketing System from this customer = 0.33 * [Probability of outcome 1 * (Revenue from sending Men's Email in this actual outcome) + Probability of outcome 2 * (Revenue from sending Men's Email in this actual outcome) + Probability of outcome 3 * (Revenue from sending Men's Email in this actual outcome) + Probability of outcome 4*(Revenue from sending Men's Email in this actual outcome) + Probability of outcome 5*(Revenue from sending Men's Email in this actual outcome) + Probability of outcome 6*(Revenue from sending Men's Email in this actual outcome) + Probability of outcome 7*(Revenue from sending Men's Email in this actual outcome) + Probability of outcome 8*(Revenue from sending Men's Email in this actual outcome)]

   +

   0.33 x [Probability of outcome 1 * (Revenue from sending Women's Email in this actual outcome) + Probability of outcome 2 * (Revenue from sending Women's Email in this actual outcome) + Probability of outcome 3 * (Revenue from sending Women's Email in this actual outcome) + Probability of outcome 4*(Revenue from sending

Women’s Email in this actual outcome) + Probability of outcome 5*(Revenue from sending Women’s Email in this actual outcome) + Probability of outcome 6*(Revenue from sending Women’s Email in this actual outcome) + Probability of outcome 7*(Revenue from sending Women’s Email in this actual outcome) + Probability of outcome 8*(Revenue from sending Women’s Email in this actual outcome)]

+

0.33 x [Probability of outcome 1 * (Revenue from sending No Email in this actual outcome) + Probability of outcome 2 * (Revenue from sending No Email in this actual outcome) + Probability of outcome 3 * (Revenue from sending No Email in this actual outcome) + Probability of outcome 4*(Revenue from sending No Email in this actual outcome) + Probability of outcome 5 *(Revenue from sending No Email in this actual outcome) + Probability of outcome 6*(Revenue from sending No Email in this actual outcome) + Probability of outcome 7*(Revenue from sending No Email in this actual outcome) + Probability of outcome 8*(Revenue from sending No Email in this actual outcome)]