Predictive Churn Analysis

Amanda Wright

17th June 2025

Table of Contents

Executive Summary

SY has conducted an analysis aimed at identifying factors influencing customer contract renewals, using data from a recent telemarketing campaign. The goal being to develop a logistic regression model that predicts the probability of customer renewal in future campaigns and to assess the effectiveness of a targeted marketing approach compared to previous untargeted methods.

The analysis concludes that a targeted marketing approach, informed by predictive modelling, can significantly enhance customer renewal rates and profitability. Key factors influencing renewals at SY include tenure, monthly charges, and contract type. The methodology and findings provide a framework for future campaigns and strategies to reduce churn.

For marketing managers, it is recommended to focus on renegotiating costs and refining the targeted approach for ongoing campaigns to maximize profitability and customer retention.

For senior executives, it is recommended that further research be undertaken to uncover what may be influencing increased churn of longer-term customers and reduced churn rates of those who spend more on telecommunications at SY, among strategically incentivizing churn reduction across the organization.

Research Background and Objectives

Profitability of Reducing Churn

Definition of Churn – "The number of customers who decide to stop using a service offered by one company and to use another company, usually because it offers a better service or price"(Ossian, 2020)

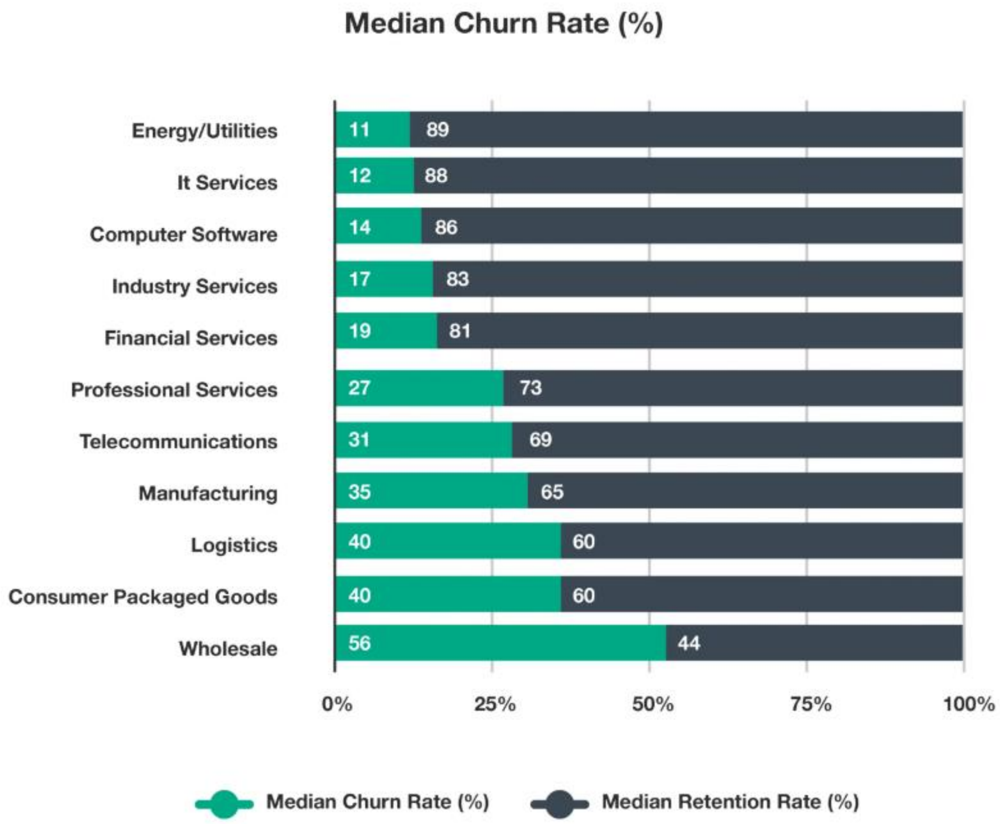

Figure 0-1 Median Churn Rate By Industry (What's the Average Churn Rate by Industry?, 2025)

The advantages of reducing churn, and increasing retention, of customers are well-documented in the business world.

Studies suggest the cost of customer acquisition is five to twenty-five times the cost of retaining your existing customers, and improving retention by just 5% can lead to significantly higher profit margins from 25% to 95%.(Gallo, 2014)

The telecommunications industry has formidable competition for customers with one of the highest rates of churn across all industries (Fig. 1).

Research Goals

Our goal is to build a model to:

1. Identify factors contributing to customer contract renewal at SY, based on the data from the last telemarketing campaign.
2. Predict the probability of renewal for SY customers in the new campaign to classify customers as “targeted” and “untargeted” whilst covering the advertising-to-sales ratio.

Analyse the success of the strategy by:

3. Comparing the predicted profitability in the new campaign’s targeted approach vs. actual profitability of the previous campaign’s untargeted approach.
4. Using the model to quantify the impact of the cost in upcoming re-negotiations with the call centre.

Report Structure

- A description of the sample data and the steps needed for preparation of the data for analysis in R.
- A detailed description of the methodological approach undertaken to build the logistic regression model, advantages of logistic regression modelling, and evaluation of the model outputs.

- Draw conclusions on what factors influence customer renewal at SY based on the model.
- Use the model to create a “targeted approach” for the new campaign based on the probability a customer will renew their contract with us.
- Discuss the business implications of the modelling including comparing key success metrics of the two marketing approaches.

Description of Sample

Last Campaign Summary

In the last telemarketing campaign:

- SY cold called *all* 4140 customers to renew their contract via a call centre, with 1086 customers renewing their contract
- Renewal rate = 26.23% last campaign, below the industry average retention of 69% in 2025 in Fig. 1

Renewals Last Campaign

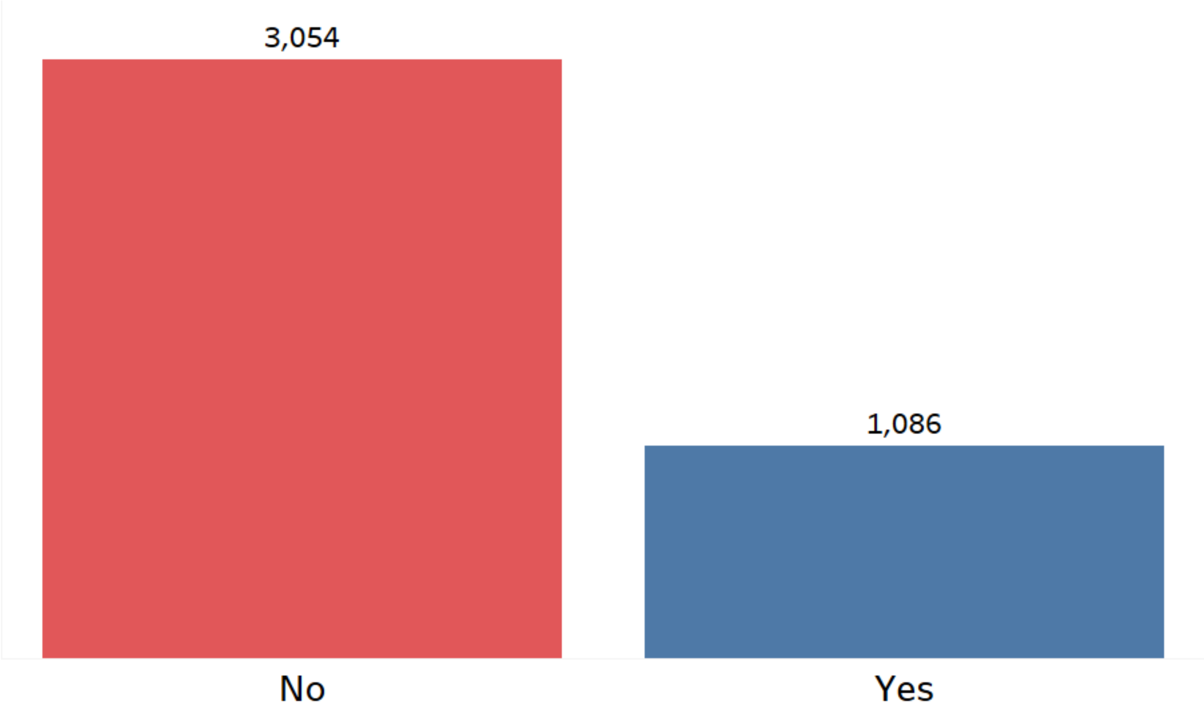

Figure 0-1 Renewals last campaign 1086/4140 = 26.23%

- The dataset from the last campaign included 4140 observations, a statistically valid sample size, with 12 variables as per Table 1.

Variable	Description
Customer ID	Identification of a customer
SeniorCitizen	Whether the customer is retired (Yes/No)
Partner	Whether the customer is married (Yes/No)
Dependents	Whether the customer has dependents (Yes/No)
Tenure	Number of months a person has been a customer of the company
PhoneService	Whether the telephone service is connected (Yes/No)
PaperlessBilling	Whether the customer uses paperless billing (Yes/No)
MonthlyCharges	Current monthly payment
TotalCharges	The total amount that the client paid for the services for the entire time
Gender	Customer gender (Male / Female)
Contract	Type of customer contract (Long Term/Short Term)
Renew	**Whether the customer renewed contract after call centre contact (Yes/No)**

Table 0-A Description of variables captured in the last campaign

Model Predictor Variables

Table 1 describes the variables available for selection in the model to predict "Renew" including:

- The variable name
- The variable description

Model Dependent Variable

Dependent Variable = "Renew"

- "Renew" is the *dependent variable* we want to predict in the logistic regression model.
- Modelling will give us insights into the factors that affect whether our customers "Renew" or not.

➢ Whilst reducing churn is our ultimate goal, for the purposes of calculating profits, and additionally the “Renew” variable is in the dataset already, predicting “Renew” will be the aim of the model.

➢ We can infer that increasing “Renew” will ultimately reduce churn.

Data Preparation

Step 1: Recode the categorial variables in the dataset to numerical variables for analysis in R.

```
# Renew: Whether the customer renews the contract (Binary variable: Yes/ No).
# Convert it to a number Renew = No as 0 and Renew = Yes as 1
telecomm_df$Renew <-ifelse(telecomm_df$Renew=="Yes",1,0)

# Recoding of Predictor variables convert to binary 0 or 1.
telecomm_df$Gender_IsMale <-ifelse(telecomm_df$gender=="Male",1,0)
telecomm_df$Partner <-ifelse(telecomm_df$Partner=="Yes",1,0)
telecomm_df$Dependents <-ifelse(telecomm_df$Dependents=="Yes",1,0)
telecomm_df$PhoneService <-ifelse(telecomm_df$PhoneService=="Yes",1,0)
telecomm_df$Contract_LT <-ifelse(telecomm_df$Contract=="Long term",1,0)
telecomm_df$PaperlessBilling <-ifelse(telecomm_df$PaperlessBilling=="Yes",1,0)
```

Figure 0-2 Recoding of variables in R for Logistic Regression Modelling

➢ “Gender” and “Contract” have descriptive names “Male/Female” and “LongTerm/ShortTerm” rather than just “Yes/No

➢ Recoded to “Gender_IsMale” and “Contract_LT” for easier reference in future analyses

Step 2: Remove the following variables:

➢ “Customer ID”, is a unique identifier and not impactful on “Renew”

➢ “Gender” and “Contract” has been recoded as “Gender_IsMale” and “Contract_LT”, therefore can be removed

```
LastCampaign <- telecomm_df[, !names(telecomm_df) %in% c("customerID",
"Contract", "gender")]
```

Figure 0-3 Drop unnecessary columns, assign to "LastCampaign"

- Data preparation is complete with all variables data types numeric/integer

```
> str(LastCampaign)
'data.frame':	4140 obs. of  11 variables:
 $ SeniorCitizen   : int  1 0 0 0 1 0 0 0 0 0 ...
 $ Partner         : num  1 1 0 1 1 1 0 1 0 1 ...
 $ Dependents      : num  0 0 0 1 0 1 0 0 1 0 ...
 $ tenure          : int  38 70 39 30 60 50 1 14 52 62 ...
 $ PhoneService    : num  1 0 0 0 1 1 1 0 1 1 ...
 $ PaperlessBilling: num  1 1 0 0 1 0 1 0 0 1 ...
 $ MonthlyCharges  : num  75 49.9 35.5 51.2 99 ...
 $ TotalCharges    : num  2870 3370 1309 1562 6018 ...
 $ Renew           : num  1 0 0 1 0 0 0 1 0 0 ...
 $ Gender_IsMale   : num  0 0 1 0 1 1 1 1 1 1 ...
 $ Contract_LT     : num  0 1 0 0 0 0 0 0 1 1 ...
```

Figure 0-4 Sample of finalised "LastCampaign" predictor variables - 4140 observations and 11 variables

Next steps are:

- build a logistic regression model
- evaluate the effectiveness of the model

Methodological Approach

Logistic Regression Modelling

Logistic regression uses predictor variables to model the probability of the occurrence of a binary outcome – for example yes or no, and to renew or not to renew.

The process entails building, evaluating, refining and re-evaluating the model until it satisfies specific statistical criteria.

Advantages of Logistic Regression

If we can predict the probability a customer says yes to "Renew", then we can:

1. "Target" ***only*** the customers with a predicted probability to "Renew" above the advertising-to-sales ratio of $7/$50 = 0.14, leaving the remainder "untargeted."
2. Maximise the efficiency of our marketing efforts and improving profits at SY with less customers for the call centre to call, whilst covering our advertising-to-sales-ratio of $7 a contact for $50 revenue.

Model Building Process

Model 1

- First attempt will include all the variables in the modified dataset "LastCampaign" (Table 2), to predict "Renew"

```
model1<-glm(Renew ~ .,data=LastCampaign, family=binomial)
```

Figure 0-1 Build" model1" to predict "Renew" with all variables from "LastCampaign" dataset (noted by Renew ~ .)

Variable	Description	Outcomes
SeniorCitizen	Whether the customer is retired	No = 0 Yes = 1
Partner	Whether the customer is married (Yes/No)	No = 0 Yes = 1
Dependents	Whether the customer has dependents (Yes/No)	No = 0 Yes = 1
Tenure	Number of months a person has been a customer of the company	N/A
PhoneService	Whether the telephone service is connected (Yes/No)	No = 0 Yes = 1
PaperlessBilling	Whether the customer uses paperless billing (Yes/No)	No = 0 Yes = 1
MonthlyCharges	Current monthly payment	N/A
TotalCharges	The total amount that the client paid for the services for the entire time	N/A
Gender_IsMale	Customer gender (Male / Female)	No = 0 Yes = 1
Contract_LT	Type of contract (Long Term/Short Term)	No = 0 Yes = 1

Table 0-A "LastCampaign" Predictor Variables to predict "Renew"

Local Model Evaluation Criteria Model 1

Criteria at the local level:

1. Evaluating the Variance Inflation Factors (VIF) to ensure no variables have a strong relationship with another variable, which would adversely affect the validity of the model, or no VIF > 10.

```
> vif(model1)
  SeniorCitizen          Partner       Dependents           tenure
       1.116493         1.372159         1.274910        14.531705
   PhoneService PaperlessBilling   MonthlyCharges     TotalCharges
       1.262509         1.108569         2.685743        18.512729
  Gender_IsMale      Contract_LT
       1.000997         1.154643
>
```

Figure 0-2 VIF Model 1 shows "Tenure" > 10 and" TotalCharges" > 10, they have multicollinearity

```
> cor(LastCampaign$tenure, LastCampaign$TotalCharges)
[1] 0.8300946
```

Figure 0-3 Correlation analysis "Tenure" and "TotalCharges" confirms high correlation close to 1

2. Ensuring all predictor variables coefficient estimates in the model to be statistically significant in the variance of "Renew", |z| < 0.05 (denoted by * - ***)

```
> summary(model1)

Call:
glm(formula = Renew ~ ., family = binomial, data = LastCampaign)

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)    
(Intercept)      -8.679e-01  1.961e-01  -4.425 9.65e-06 ***
SeniorCitizen     4.957e-01  1.056e-01   4.694 2.68e-06 ***
Partner           5.201e-03  9.782e-02   0.053  0.95760    
Dependents       -1.747e-01  1.139e-01  -1.533  0.12530    
tenure           -6.687e-02  7.635e-03  -8.758  < 2e-16 ***
PhoneService     -9.464e-01  1.527e-01  -6.196 5.78e-10 ***
PaperlessBilling  4.407e-01  9.287e-02   4.746 2.08e-06 ***
MonthlyCharges    2.537e-02  2.531e-03  10.023  < 2e-16 ***
TotalCharges      2.521e-04  8.592e-05   2.934  0.00335 ** 
Gender_IsMale    -1.374e-02  8.181e-02  -0.168  0.86662    
Contract_LT      -1.374e+00  2.152e-01  -6.387 1.69e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 4764.9  on 4139  degrees of freedom
Residual deviance: 3613.5  on 4129  degrees of freedom
AIC: 3635.5

Number of Fisher Scoring iterations: 6
```

Figure 0-4 Summary Output Model 1 in R

Local Model Evaluation Summary

Local model evaluation results in Table 3 suggests Model 1 is *not* a good model for predicting "Renew" and needs adjustment.

Model Evaluation Criteria	Interpretation	Results	Pass/Fail
Variance Inflation Factors (VIF) < 10	VIF > 10 = Variable multicollinearity	"Tenure" = 14.53 "TotalCharges" = 18.51	Fail
Correlation analysis of high VIF variables	0 = No correlation 1 = Perfect correlation	"Tenure" vs "TotalCharges" = 0.83	Fail
All predictor variables significance level < 0.05	Predictor variables are < 0.05	"Partner" > 0.05 "Gender_IsMale" > 0.05 "Dependents" > 0.05	Fail

Table 0-A Local Model Evaluation Criteria Summary Model 1

Actions for Model 1

- Drop *either* "Tenure" or "Total Charges" due to multicollinearity
  - Drop "Total Charges", highest VIF and the lowest significance in the model
- Drop "Partner", "Dependents", "Gender_IsMale", insignificance in Model 1

Model 2

- Refine the model to only include the following predictor variables to predict "Renew":

```
# Adjust to Model 2 with new set of independent variables
model2<-glm(Renew ~ SeniorCitizen+
                PhoneService+
                PaperlessBilling+
                MonthlyCharges+
                Contract_LT+
                tenure,
            data=LastCampaign, family=binomial)
```

Figure 0-5 Adjusted variable set Model 2

Variable	Description	Recode as Numerical
SeniorCitizen	Whether the customer is retired	No = 0 Yes = 1
Tenure	Number of months a person has been a customer of the company	N/A
PhoneService	Whether the telephone service is connected (Yes/No)	No = 0 Yes = 1
PaperlessBilling	Whether the customer uses paperless billing (Yes/No)	No = 0 Yes = 1
MonthlyCharges	Current monthly payment	N/A
Contract_LT	Type of contract (Long Term/Short Term)	No = 0 Yes = 1

Table 0-A Model 2 variables to predict "Renew"

Local Model Evaluation Criteria Model 2

As before:

1. Evaluate the Variance Inflation Factors (VIF).
2. All predictor variables in the model need to be statistically significant.

```
> vif(model2)
   SeniorCitizen      PhoneService PaperlessBilling    MonthlyCharges 
        1.056521          1.253626         1.101682          1.708401 
     Contract_LT            tenure 
        1.132855          1.451103 
```

Figure 11 Variance Inflation Factor(VIF) for Model 2

```
> summary(model2)

Call:
glm(formula = Renew ~ SeniorCitizen + PhoneService + PaperlessBilling + 
    MonthlyCharges + Contract_LT + tenure, family = binomial, 
    data = LastCampaign)

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)    
(Intercept)      -1.281731   0.146793  -8.732  < 2e-16 ***
SeniorCitizen     0.529161   0.102978   5.139 2.77e-07 ***
PhoneService     -0.905067   0.148608  -6.090 1.13e-09 ***
PaperlessBilling  0.436422   0.092462   4.720 2.36e-06 ***
MonthlyCharges    0.030158   0.002042  14.769  < 2e-16 ***
Contract_LT      -1.340813   0.212526  -6.309 2.81e-10 ***
tenure           -0.046592   0.002444 -19.062  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 4764.9  on 4139  degrees of freedom
Residual deviance: 3625.6  on 4133  degrees of freedom
AIC: 3639.6

Number of Fisher Scoring iterations: 6
```

Figure 12 Summary output for Model 2

Local Model Evaluation Model 2 Summary

Model Evaluation Criteria	Interpretation	Results	Pass/Fail
Variance Inflation Factors (VIF) < 10	VIF > 10	Nil > 10	Pass
Correlation analysis of high VIF variables	0 = No correlation 1 = Perfect correlation	N/A	Pass

All predictor variables significance level < 0.05	Predictor variables are < 0.05	All predictor variables < 0.05	Pass

Table 0-A Model 2 Local Evaluation Criteria for Model 2

Global Model Evaluation

Nagelkerke R^2 - Logistic Regression

- When a dependent variable is categorical, Nagelkerke R^2 can we used as a pseudo-R^2 which approximates the proportion of the variance in the dependent variable explained by the model from 0 to 1, with higher values indicating a better fit. (*IBM SPSS Statistics*, 2021)
- A Nagelkerke R^2 0.10 - 0.50 is considered acceptable if the predictor variables are significant in research involving social sciences due to the unpredictability of human behaviour. (Ozili, 2022)
- Model 2 explains approximately 35% of the variation in “Renew”.

```
> PseudoR2(model2, which = "Nagelkerke")
Nagelkerke
  0.351895
```

Figure 13 Nagelkerke R^2 indicating 35% of variation in “Renew” explained by Model 2

ANOVA test

- ANOVA (Analysis of Variance) tests whether adding certain predictors significantly improves the model

- A significance level < 0.05 after the last variable is added, “Tenure”, indicates Model 2 is significant in explaining the variance in “Renew”

```
> anova(model2, test="Chisq")
Analysis of Deviance Table

Model: binomial, link: logit

Response: Renew

Terms added sequentially (first to last)


                 Df Deviance Resid. Df Resid. Dev  Pr(>Chi)    
NULL                              4139     4764.9              
SeniorCitizen     1   101.64      4138     4663.2 < 2.2e-16 ***
PhoneService      1     0.18      4137     4663.1    0.6704    
PaperlessBilling  1   114.08      4136     4549.0 < 2.2e-16 ***
MonthlyCharges    1    57.56      4135     4491.4 3.271e-14 ***
Contract_LT       1   422.29      4134     4069.1 < 2.2e-16 ***
tenure            1   443.55      4133     3625.6 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 14 Model 2 ANOVA Test, last variable "Tenure" and model is significant < 0.05

Hosmer-Lemeshow Test

- The Hosmer-Lemeshow test assesses the goodness of fit for logistic regression models. (‘Hosmer–Lemeshow test’, 2025)
- A p-value of > 0.05 indicates the predicted values of Model 2 and the actual values of the last campaign are not significantly different, or they are similar.
- Unfortunately, the Hosmer-Lemeshow test indicates value 0.01828, indicating the model is *not* a great predictor of “Renew”
- Hosmer-Lemeshow literature suggests that it does not take over-fitting into account, and has limitations in describing goodness- of-fit for logistic regression (‘Hosmer–Lemeshow test’, 2025)

```
        Hosmer and Lemeshow goodness of fit (GOF) test

data:  LastCampaign$Renew, fitted(model2)
X-squared = 18.421, df = 8, p-value = 0.01828
```

Figure 15 Model 2 - Hosmer-Lemeshow goodness-of-fit test results

Predictive Accuracy of Model 2

- The confusion matrix of *predictions vs. actuals* on the last campaign data gives us Model 2 predictive accuracy = (1007 + 1537) / 4140 = 61.45%
- The cut-off set at 0.14 means customers below the advertising-to-sales ratio $7/$50 won't be contacted.

```
> Conf(model2, cutoff = 0.14)

Confusion Matrix and Statistics

          Reference
Prediction    1    0
         1 1007 1517
         0   79 1537

              Total n : 4'140
             Accuracy : 0.6145
```

Figure 16 Model 2 - Confusion matrix reflecting predictive accuracy

- Model 2 predicted 1007 customers would renew, and 1537 customers as non-renewals, both accurately
- The model would have incorrectly predicted 1517 renewals as non-renewals, and 79 renewals who did not renew, both incorrectly.
- The model is very good at predicting non-renewals at 1537/(1537+79) = 95%, however only 1007/(1007+1517) = 39.8% accurate at predicting renewals.
- The swing here is because we have set the cut-off at 0.14, which gives us more customers to contact, opportunity for more renewals whilst covering costs, which is what we want the model to do.

Results

Overall Model 2 Evaluation

Criteria	Interpretation	Results	Pass/Fail
Nagelkerke R2	0.10 – 0.50	0.35	Pass
ANOVA test	Significance < 0.05	< 2.2e-16	Pass
Homer-Lemeshow	Significance > 0.05	0.0128	Fail*
Variance Inflation Factors (VIF) < 10	VIF > 10	Nil > 10	Pass
Correlation analysis of high VIF variables	0 = No correlation 1 = Perfect correlation	N/A	Pass
All predictor variables significance levels < 0.05	Predictor variables are < 0.05	All predictor variables< 0.05	Pass
Predictive Accuracy	Accuracy to predict based on training data	61.45%	Pass

Table 0-A Local and global model evaluation Model 2. Failure on Homer-Lemeshow does not indicate failure of the model.

Model 2 is a good model, and statistically useful to predict "Renew" for our research goals

Research Goal 1

Identify the factors that contribute to customers renewing contract at SY, using the data from the last telemarketing campaign

```
> summary(model2)

Call:
glm(formula = Renew ~ SeniorCitizen + PhoneService + PaperlessBilling +
    MonthlyCharges + Contract_LT + tenure, family = binomial,
    data = LastCampaign)

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)      -1.281731   0.146793  -8.732  < 2e-16 ***
SeniorCitizen     0.529161   0.102978   5.139 2.77e-07 ***
PhoneService     -0.905067   0.148608  -6.090 1.13e-09 ***
PaperlessBilling  0.436422   0.092462   4.720 2.36e-06 ***
MonthlyCharges    0.030158   0.002042  14.769  < 2e-16 ***
Contract_LT      -1.340813   0.212526  -6.309 2.81e-10 ***
tenure           -0.046592   0.002444 -19.062  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 4764.9  on 4139  degrees of freedom
Residual deviance: 3625.6  on 4133  degrees of freedom
AIC: 3639.6

Number of Fisher Scoring iterations: 6
```

Figure 17 Summary output of Logistic Regression Model to predict "Renew"

We will interpret the output of model 2 in two ways:

1. Odds-Ratio (OR) – how the coefficients for each predictor can be translated in how they impact likelihood of a renewal to occur.
2. Coefficient Standardisation – how we can compare the importance of the coefficients on “renew” by placing them on the same scale.

Impact of Predictor Variables on Odds-Ratio for “Renew”

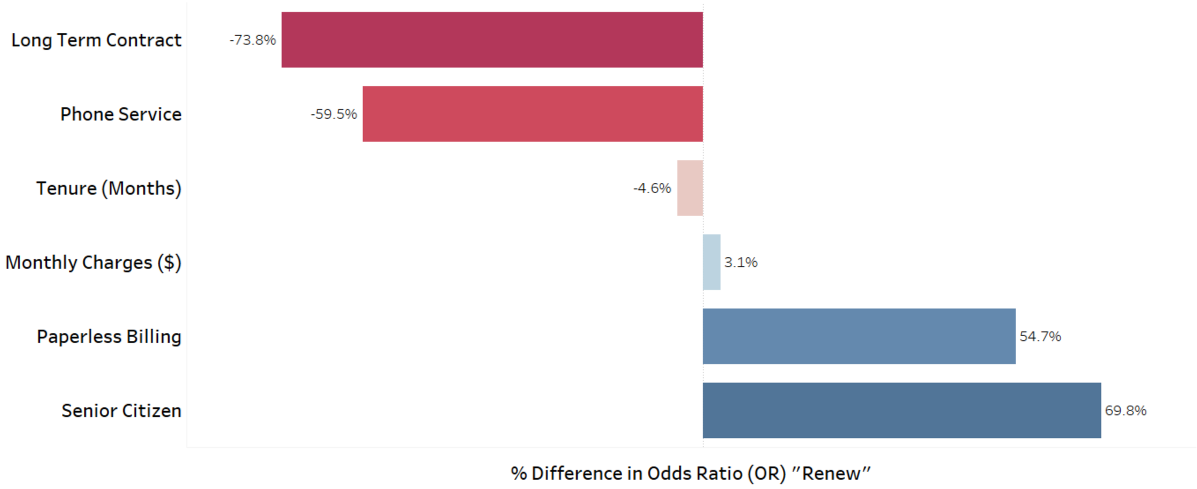

Figure 18 Multiplicative effect on odds to "Renew" by a 1-Unit increase in the predictor variable

- Being retired means 70% higher odds to renew than someone not retired
- Customer on a long-term contract is 75% lower odds to renew than a short-term contracted customer
- Tenure and monthly charges are measured by an increase in 1 month and $1 change accordingly

Relative Importance of Predictor Variables

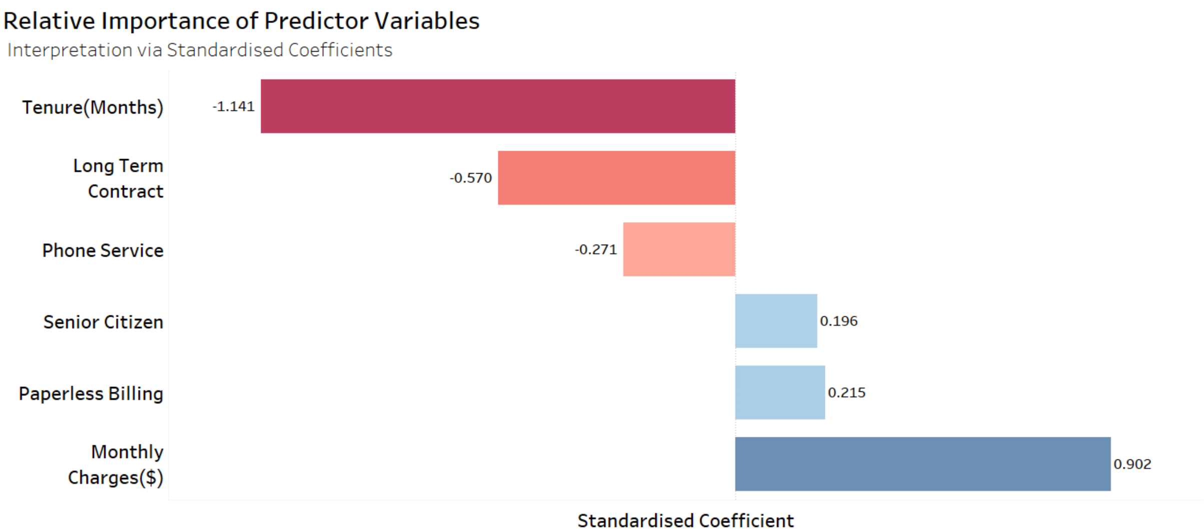

Figure 19 The effect of a variance of one standard deviation increase in the predictor variable and its relative impact on "Renew"

Standardization of the model coefficient estimates in Figure 19 allows us to:

1. Compare the impact on **“Renew”** of continuous variables like **“Tenure”** and **“Monthly Charges”** on the same scale as categorical variables, such as **“SeniorCitizen”**.
2. Assess which variables have the *<u>strongest effect</u>* on the likelihood of renewal
   - **Positive standardized coefficient** → A **higher-than-average** value of this predictor **increases** the likelihood of renewal

   - **Negative standardized coefficient** → A **higher-than-average** value of this predictor **decreases** the likelihood of renewal
   - **Magnitude of coefficient** → Larger absolute values indicate stronger effects on renewal probability

Conclusion

The variables that contribute the most to contract renewal at SY are:

- **<u>Tenure</u> - longer tenure strongly decreases the probability of the renewal.**
- **<u>Monthly Charges</u> - higher monthly charges have the strongest positive effect on the probability renewal**
- **<u>Long Term Contract</u> - longer contracts decrease the probability of the renewal.**

Research Goal 2

Predict the probability of renewal for SY customers in the new campaign to classify customers as "targeted" and "untargeted" whilst covering the advertising-to-sales ratio.

Data Preparation and Modelling

- Read in the "NewCampaign" dataset of 1835 observations
- Prepare categorical data as numerical variables as before,
- Apply Model 2 to "New Campaign" data as "Predicted.Renew"

```
NewCampaign$Predicted.Renew <- predict(model2, newdata = NewCampaign, type =
"response")
```

Figure 20 Calculating probability of renewal based on the model

```
> str(NewCampaign)
'data.frame':	1835 obs. of  7 variables:
 $ SeniorCitizen   : int  0 0 0 0 0 0 1 0 0 0 ...
 $ tenure          : int  72 44 4 2 33 1 55 52 32 51 ...
 $ PhoneService    : num  1 1 1 1 1 0 1 1 1 1 ...
 $ PaperlessBilling: num  0 1 1 0 1 1 1 1 1 1 ...
 $ MonthlyCharges  : num  24.1 88.2 55.9 53.5 90.7 ...
 $ Contract_LT     : num  1 0 0 0 0 0 0 0 0 0 ...
 $ Predicted.Renew : num  0.00212 0.24194 0.43759 0.33893 0.3649 ...
```

Figure 21 "NewCampaign" including "Predicted.Renew"

Classification of Target Customers

The campaign will *only* target customers that cover the advertising-to-sales ratio:

- Each customer will be "targeted" if their predicted probability of renewal is greater than the breakeven threshold of the advertising-to-sales ratio of $7 cost to $50 revenue, or > 0.14

- Remaining customers will be "Untargeted" as the model predicts probability of renewal < 0.14 and likely an unprofitable contact

Predicted Renewal Rate By Customer – New Campaign

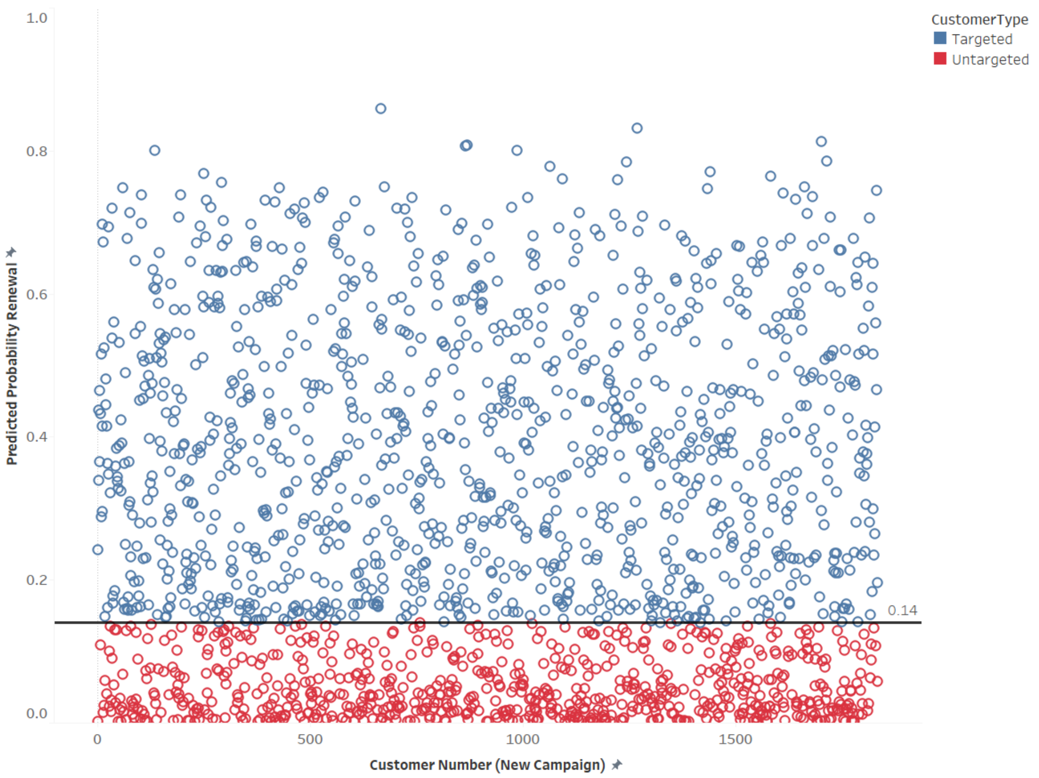

Figure 0-1 Classification of 'targeted" v "untargeted" for new campaign

Conclusion

The new campaign will target 1118 (60.9%) of customers, with 717 (39.1%) untargeted.

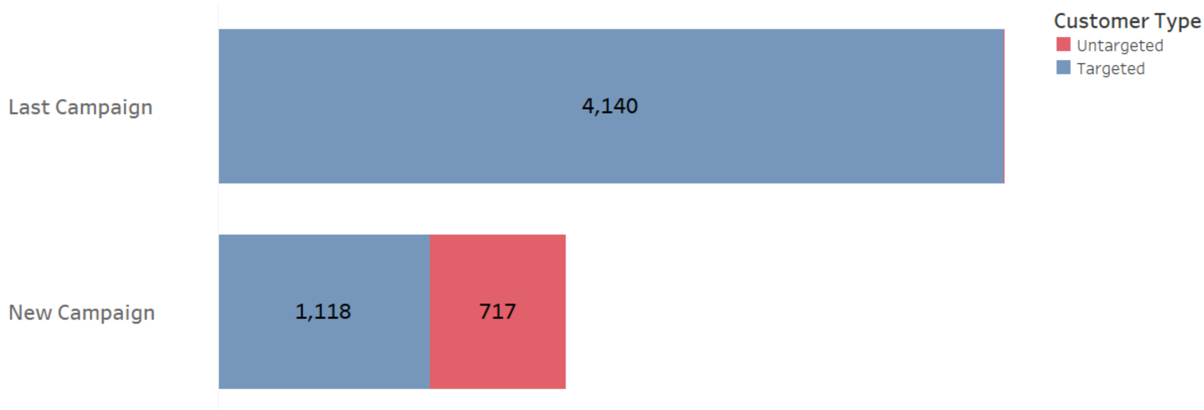

Figure 23 Efficiency of the new campaign in time and costs

Implications for Campaign

Research Goal 3

Compare predicted profitability of the new campaign to the previous campaign's actual profits.

Total Profit By Campaign

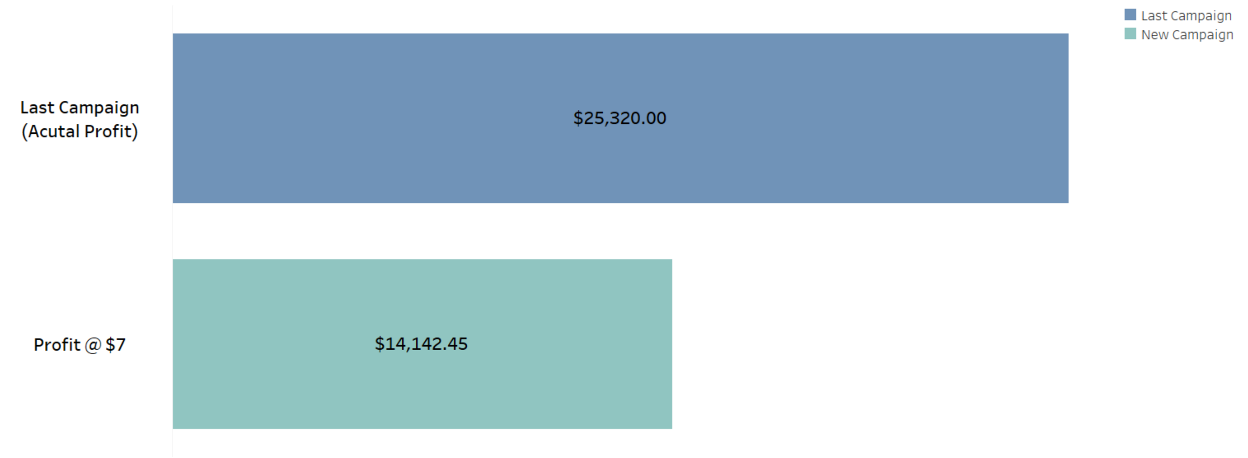

Figure 24 Actual Profits vs. Predicted Profits using logistic regression model

- The previous campaign was ***more*** profitable overall, however ***fewer*** customers were available to be contacted (1835 vs. 4140) in the new campaign and ***only 1118 "targeted" by the model***
- Therefore, comparison of ***profit by "targeted" customers*** gives greater insight

Profit Per Customer By Campaign

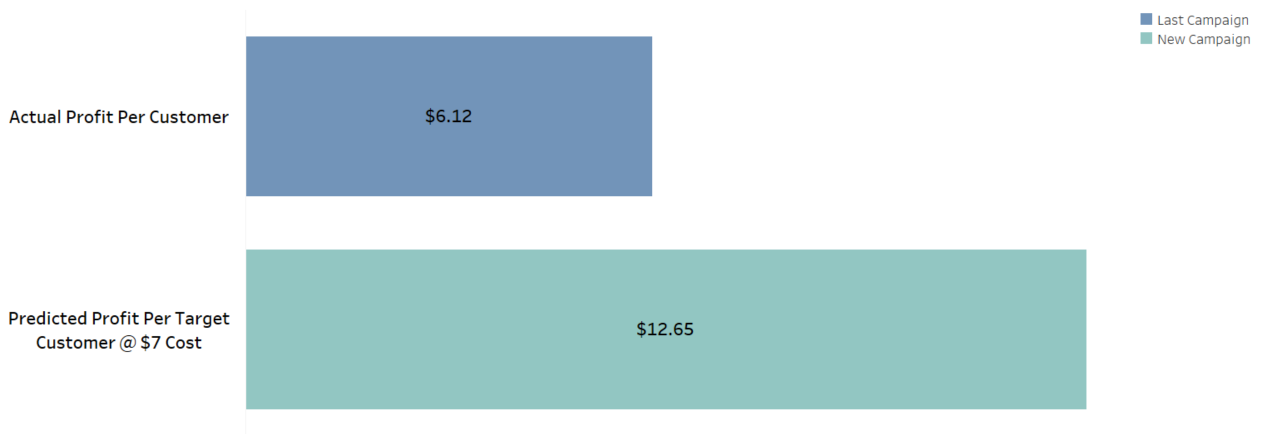

Figure 25 Predicted Profit By "Targeted" Approach more than 2 times "Untargeted" approach

Conclusion

Predicted profits per target customer in the new campaign are more than double the last campaigns actual profits per customer due to efficiency of the process.

[**For a full financial analysis refer to Appendix C**]

Research Goal 4

Use the data in re-negotiations for costs with the call centre.

Effect of Costs on Number of "Targeted" Customers

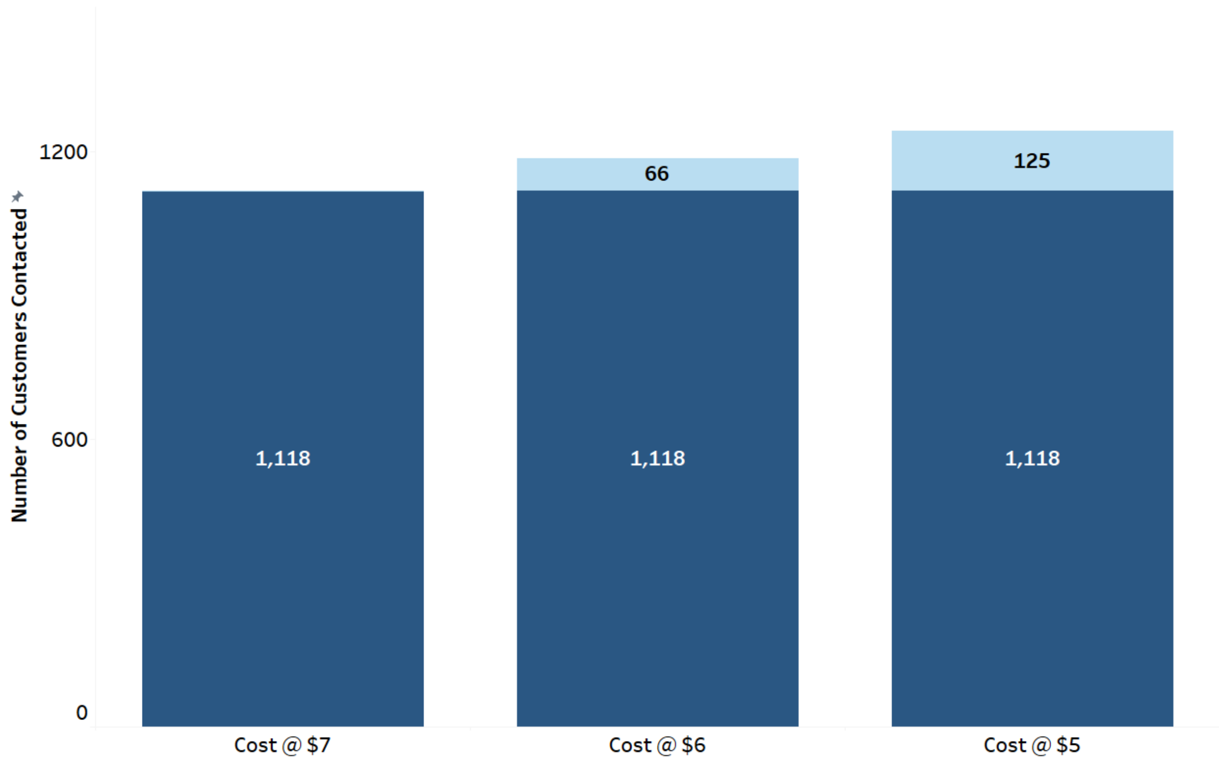

Figure 0-16 Lower costs = lower threshold to breakeven = higher number of target customers

- The model calculates the number of customers contacted in the upcoming campaign by covering the advertising-to-sales ratio.

➢ The predicted profits at $6 a customer contact and $5 a customer contact are compared below, to both the current $7 contract and last campaign profit per customer.

Effect of Call Centre Costs vs. Predicted Profit Per Customer

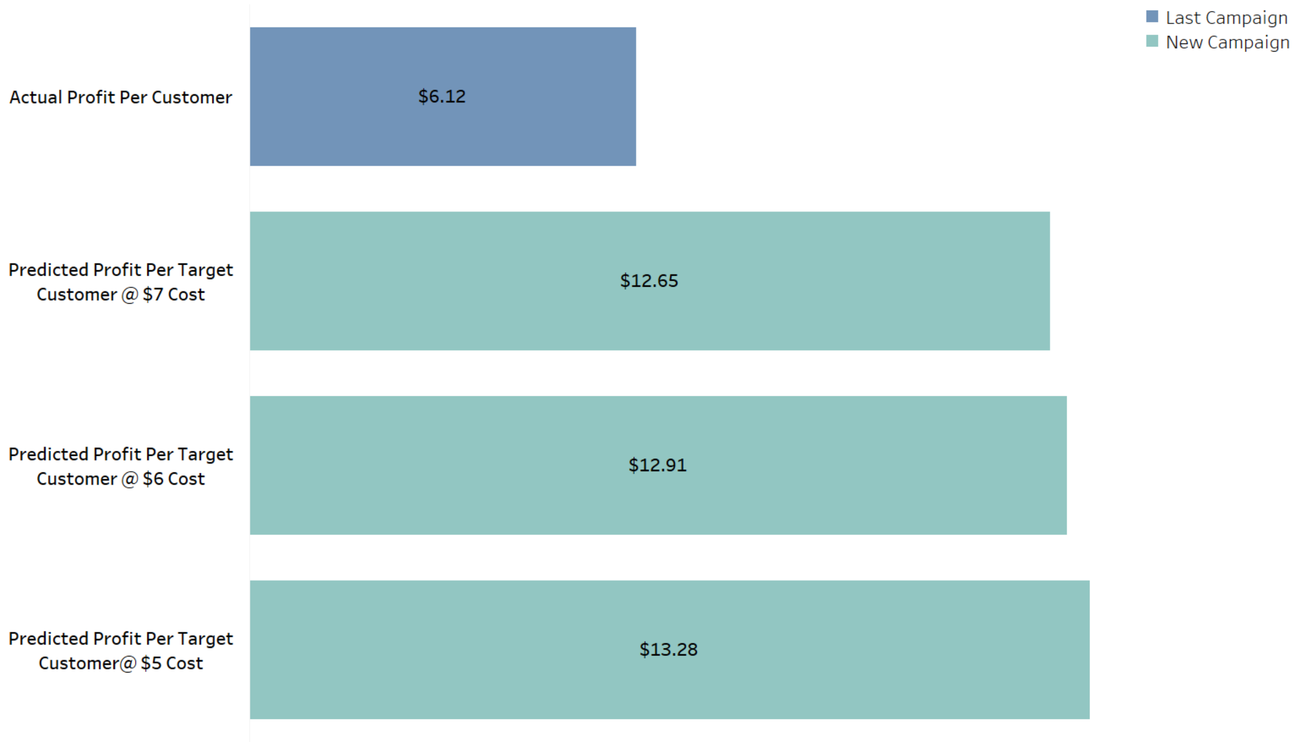

Figure 0-27 Impact of call centre costs on predicted profit per customer

Conclusion

Re-negotiation of costs below $7 with the call centre results in increasing the number of "target" customers whilst covering advertising-to-sales ratio.

By contacting more customers, there are more chances for customers to "Renew" and inflates predicted profits greater than the cost saving.

Full financial analysis refer Appendix D

Conclusion

Summary

- Modelling suggests a **targeted approach can increase profits per target customer by more than double** compared to an untargeted approach.
- The variables that contribute the most to contract renewal at SY are:

  Tenure - Longer tenure strongly decreases the probability of customer renewals.

  Monthly Charges - Higher monthly charges have the strongest positive effect on the probability customer renewals.

  Long Term Contract - longer contracts decrease the probability of a customer renewal.

For Marketing Managers

1. **Re-negotiation of costs below $7**

- Lower costs results in increasing the number of “targeted” customers therefore providing more chances for customers to renew and therefore inflates predicted profits greater than the cost saving.
- Potentially look to bring the contacting of customers internally to increase engagement with customers and lower costs simulataneously.

2. **<u>Fine Tune "Targeted" Approach</u>**

The methodological approach, and the R-code in Appendix G, can be re-utilized after each new campaign to update the model futher, and keep it meaningful and current as customer dynamics change.

3. **<u>Measure Success</u>**

After each new campaign measure the success of the "targeted approach" looking for an increase in renewals as a % of customers contacted, therefore reducing churn.

4. **<u>Capture Defection Data</u>**

Look to capture defection reason data when customers don't renew after being contacted by the call centre, enhancing the model further.

For Senior Management

5. **<u>Drill Down Deeper</u>**

Drill down on the factors influencing renewal ratios and answer the following questions:

<u>Tenure and Long-Term Contract</u> – Why are long term customers leaving us? Are they leaving us at the most profitable time in their lifetime value?

<u>Monthly Charges</u> – Why do our high monthly spend customers more likely to renew? Are we great value on premium products?

6. **<u>Focus on the Right Customers</u>**

Research Customer Lifetime Value (CLV) and Customer Referral Value (CFV) metrics at SY, and the customer acquisition strategy, are we acquiring the right customers with the highest lifetime value to us?

7. Measure and Incentivize the Right Behaviours

Research ways to incentivize customer churn measures at SY, and make it become a KPI in the business, incentivized for all levels given the profitability of reducing churn.

Limitations of the Analysis

- The analysis cannot be interpreted as a “cause and effect” relationship but rather an association between the predictor variables and their impact on renewal.
- There are variables that have not been included such as age of customer, distance from a metro area and many other factors that may have been important predictors but omitted.
- We can only say that based on the predictor variables available and the customers from the last campaign, this is our “best guess” on what may happen in the next campaign.

References

Gallo, A. (2014) ‘The Value of Keeping the Right Customers’, *Harvard Business Review Digital Articles*, pp. 2–5.

‘Hosmer–Lemeshow test’ (2025) *Wikipedia*. Available at: https://en.wikipedia.org/w/index.php?title=Hosmer%E2%80%93Lemeshow_test&oldid=1292036248 (Accessed: 8 June 2025).

IBM SPSS Statistics (2021). Available at: https://www.ibm.com/docs/en/spss-statistics/25.0.0?topic=model-r-squared-statistics (Accessed: 8 June 2025).

Ossian (2020) ‘Churn Rate vs Retention Rate - Comparison Guide’, *ReliaBills*, 17 July. Available at: https://www.reliabills.com/blog/churn-rate-vs-retention-rate/ (Accessed: 10 June 2025).

Ozili, P.K. (2022) ‘The acceptable R-square in empirical modelling for social science research’.

What’s the Average Churn Rate by Industry? (2025) *CustomerGauge*. Available at: https://customergauge.com/blog/average-churn-rate-by-industry (Accessed: 6 June 2025)

.

Appendices

Appendix A –Last Campaign Variables

Variable	Description	Recode as Numerical
Customer ID	Identification of a customer	N/A
SeniorCitizen	Whether the customer is retired	No = 0 Yes = 1
Partner	Whether the customer is married (Yes/No)	No = 0 Yes = 1
Dependents	Whether the customer has dependents (Yes/No)	No = 0 Yes = 1
Tenure	Number of months a person has been a customer of the company	N/A
PhoneService	Whether the telephone service is connected (Yes/No)	No = 0 Yes = 1
PaperlessBilling	Whether the customer uses paperless billing (Yes/No)	No = 0 Yes = 1
MonthlyCharges	Current monthly payment	N/A
TotalCharges	The total amount that the client paid for the services for the entire time	N/A
Gender	Customer gender (Male / Female)	Gender_IsMale = 0 Gender_IsMale = 1
Contract	Type of customer contract (Long Term/Short Term)	Contract_LT = 0 Contract_LT = 1
Renew	Whether the customer renewed contract after call centre contact (Yes/No)	No = 0 Yes = 1

Table 0-A Last Campaign Variables

Appendix B – Final Model

```
> summary(model2)

Call:
glm(formula = Renew ~ SeniorCitizen + PhoneService + PaperlessBilling +
    MonthlyCharges + Contract_LT + tenure, family = binomial,
    data = LastCampaign)

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)      -1.281731   0.146793  -8.732  < 2e-16 ***
SeniorCitizen     0.529161   0.102978   5.139 2.77e-07 ***
PhoneService     -0.905067   0.148608  -6.090 1.13e-09 ***
PaperlessBilling  0.436422   0.092462   4.720 2.36e-06 ***
MonthlyCharges    0.030158   0.002042  14.769  < 2e-16 ***
Contract_LT      -1.340813   0.212526  -6.309 2.81e-10 ***
tenure           -0.046592   0.002444 -19.062  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 4764.9  on 4139  degrees of freedom
Residual deviance: 3625.6  on 4133  degrees of freedom
AIC: 3639.6

Number of Fisher Scoring iterations: 6
```

Table B-1 Summary output from the final model

Variable	Description	Recode as Numerical
SeniorCitizen	Whether the customer is retired	No = 0 Yes = 1
Tenure	Number of months a person has been a customer of the company	N/A
PhoneService	Whether the telephone service is connected (Yes/No)	No = 0 Yes = 1
PaperlessBilling	Whether the customer uses paperless billing (Yes/No)	No = 0 Yes = 1
MonthlyCharges	Current monthly payment	N/A
Contract_LT	Type of contract (Long Term/Short Term)	No = 0 Yes = 1

Table B-2 Final predictor variables for final model

Appendix C – Model Interpretations

Odds-Ratio (OR) Model 2

```
> exp(coef(model2))
    (Intercept)   SeniorCitizen    PhoneService PaperlessBilling
      0.2775565       1.6975076       0.4045147        1.5471609
 MonthlyCharges     Contract_LT          tenure
      1.0306169       0.2616329       0.9544763
```

Figure C-1 Exponentiated coefficients for the final model

Standardisation of Variable Coefficients Model 2

```
> model2.std

Call:
glm(formula = c("Renew ~ SeniorCitizen.z + PhoneService.z + PaperlessB
illing.z + MonthlyCharges.z + ",
"    Contract_LT.z + tenure.z"), family = binomial, data = data)

Coefficients:
                   Estimate Std. Error z value Pr(>|z|)    
(Intercept)        -1.62565    0.06061 -26.821  < 2e-16 ***
SeniorCitizen.z     0.19596    0.03814   5.139 2.77e-07 ***
PhoneService.z     -0.27068    0.04444  -6.090 1.13e-09 ***
PaperlessBilling.z  0.21459    0.04546   4.720 2.36e-06 ***
MonthlyCharges.z    0.90179    0.06106  14.769  < 2e-16 ***
Contract_LT.z      -0.57040    0.09041  -6.309 2.81e-10 ***
tenure.z           -1.14121    0.05987 -19.062  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 4764.9  on 4139  degrees of freedom
Residual deviance: 3625.6  on 4133  degrees of freedom
AIC: 3639.6

Number of Fisher Scoring iterations: 6
```

Figure C-2 Standardised coefficients for the final model

Appendix E - New Campaign Targeted Approach Key Measures

Key Term	Description	Last Campaign (Actual)	New Campaign (Expected)
Targeted Customers	Number of Contacted Customers	4140	1118
Untargeted Customer	Number not contacted	0	717
Cost of campaign	Number Of Targeted Customers × Cost Per Customer Contact ($7)	4140 x $7 = $28,980	1118 x $7 = $7,826
Profit	Revenue $50 per renewal - Cost	$25,320	$14,142
Profit Per Contact	(Revenue – Cost)/ Targeted Customers	$6.12	$12.65

Table E Targeted Approach Profitability

Appendix F - Campaign Cost Financial Analysis

Key Term	Description	Last Campaign (Actual)	New Campaign @ $7 Cost (Predicted)	New Campaign @ $6 Cost (Predicted)	New Campaign @$5 Cost (Predicted)
Breakeven Threshold	**Cost Per Customer Contact ÷ Value Of Retaining An Existing Customer**	**N/A**	*$7 / $50 = 0.14*	*$6 / $50 = 0.12*	*$5 / $50 = 0.10*
Targeted Customers	**Number of Contacted Customers**	**4140**	*1118*	*1184*	*1243*
Untargeted Customer	**Number not contacted**	**0**	*717*	*651*	*592*
Campaign Cost	**Number Of Targeted Customers × Cost Per Customer Contact ($7)**	**4140 x $7 = $28,980**	*1118 x $7 = $7,826*	*1184 x $6 = $7,104*	*1243 x $5 = $6215*
Campaign Profit	**Revenue $50 per renewal - Cost**	**$25,320**	*$14,142.45*	*$15290.61*	*$16502.84*
Campaign Profit Per Contact	**(Revenue – Cost)/ Targeted Customers**	**$6.12**	*$12.65*	*$12.91*	*$13.28*
% Profit Increase on Last Campaign	**Per Customer Contacted**	**Profit Per Customer / $6.12**	***106%***	***111%***	***117%***

Table F Financial Analysis Campaign Cost Impact

Appendix G - R-Code

```
# install.packages("DescTools")
# install.packages("reghelper")
library(DescTools)
library("ResourceSelection")
library(car)
library("reghelper")

#--------------------------------------------------------------
# Import data
# Last Campaign Data (sample size: 4140).
#--------------------------------------------------------------

# Pull data from the csv file
telecomm_df <-read.csv("Original_A3_Model_development.csv",
header=TRUE)

#--------------------------------------------------------------
# Get an overview of the variables
#--------------------------------------------------------------
names(telecomm_df)

#--------------------------------------------------------------
#  Client information
# 1. Customer ID - identification of a customer
# 2. Gender - customer gender (binary variable: male / female)
# 3. Senior Citizen - whether the customer is retired (binary
variable: Yes=1/ No=0)
# 4. Partner - whether the customer is married (binary
variable: Yes/ No)
# 5. Dependents - whether the customer has dependents (binary
variable: Yes/ No)

# Details about the customer's status with SY
# 1. Tenure - number of months a person has been a customer of
the company
# 2. Phone Service - whether the telephone service is
connected (binary variable: Yes/ No)
# 3. Contract - type of customer contract (binary variable:
Long term/ Short term)
# 4. Paperless Billing - whether the customer uses paperless
billing (binary variable: Yes/No)
# 5. Monthly Charges - current monthly payment
# 6. Total Charges - the total amount that the client paid for
the services for the entire
# time
#--------------------------------------------------------------

#--------------------------------------------------------------
```

```
# Recoding of Dependent variable (Renew) and Predictor
Variables
#-----------------------------------------------------------------

# Renew: Whether the customer renews the contract (Binary
variable: Yes/ No).
# Convert it to a number Renew = No -> 0 and Renew = Yes -> 1
telecomm_df$Renew <-ifelse(telecomm_df$Renew=="Yes",1,0)

# Recoding of Predictor variables convert to binary 0 or 1.
telecomm_df$Gender_IsMale <-
ifelse(telecomm_df$gender=="Male",1,0)
telecomm_df$Partner <-ifelse(telecomm_df$Partner=="Yes",1,0)
telecomm_df$Dependents <-
ifelse(telecomm_df$Dependents=="Yes",1,0)
telecomm_df$PhoneService <-
ifelse(telecomm_df$PhoneService=="Yes",1,0)
telecomm_df$Contract_LT <-ifelse(telecomm_df$Contract=="Long
term",1,0)
telecomm_df$PaperlessBilling <-
ifelse(telecomm_df$PaperlessBilling=="Yes",1,0)

#-----------------------------------------------------------------
# Create LastCampaign with only the columns we are interested
in
# Drop the unmeaningful "CustomerID"
# Drop "gender" column and "Contract" as recoded these columns
into
# Gender_IsMale and Contract_LT
#-----------------------------------------------------------------
LastCampaign <- telecomm_df[, !names(telecomm_df) %in%
c("customerID", "Contract", "gender")]

str(LastCampaign)

# ---------------------------------------------------------------
# Logistic Regression Model Development
#-----------------------------------------------------------------
# Step 1: Define research goal
#
# 1. Identify the probability of each customer to renew in the
new campaign. (Renew = 1)
# 2. Customers will be identified as "Targeted" if the
probability to renew is more than 0.14, else "Untargeted"
# 3. Compare the profit from the last campaign's "untargeted"
approach to the expected profit from the new campaign's
"targeted" approach.
#-----------------------------------------------------------------

# ---------------------------------------------------------------
# Step 2: Specify the model
```

```
# ------------------------------------------------------------

# First model - logistic regression model from LastCampaign
data (glm) and include all independent variables as "." to
predict dependent variable "Renew"
model1<-glm(Renew ~ .,data=LastCampaign, family=binomial)

#-----------------------------------------------------------
# Step 3: Model evaluation.
# First, we do local model evaluation. We check model variable
significance or not & multicollinearity issue
# Second, we do global model evaluation. We check Nagelkerke R
square & accuracy
# ------------------------------------------------------------

# Local Model Evaluation - Multicollinearity Test
# Check for VIF over 10, this indicates multicollinearity
amongst variables
vif(model1)

# Check correlation of these two variables
cor(LastCampaign$tenure, LastCampaign$TotalCharges)


# Local Model Evaluation - Significance Test
# Check significance levels of independent variables in the
model
summary(model1)

#-------------------------------------------------------------
# Model Evaluation Model 1

# Local Model Evaluation: VIFs are greater than 10 for
"tenure" and "TotalCharges" means there is multicollinearity
issue in model.
# Additionally "Partner" and "Dependents" have significance
level of well above 0.05 as does "Gender_IsMale" or not
significant in predicting chance of renewal.
# Overall this is not a good model due to VIF and non-
significant independent variables
#  To address Multicollinearity we will drop "Total Charges" as
lowest significance and highest VIF, we will leave tenure as
is significant.
# To Address this we will remove -  "TotalCharges",
"Gender_IsMale", "Partner", "Dependents"

# Global model evaluation  - We will skip for Model 1 due to
high VIFs and insignificant variables that need to be fixed
first
#-------------------------------------------------------------
```

```
# Adjust to Model 2 with new set of independent variables
model2<-glm(Renew ~ SeniorCitizen+
                PhoneService+
                PaperlessBilling+
                MonthlyCharges+
                Contract_LT+
                tenure,
            data=LastCampaign, family=binomial)

#------------------------------------------------------------
# Step 3: Model evaluation:
# First, we do local model evaluation. We check model
parameter significant or not & multicollinearity issue.
# Second, we do global model evaluation. We check Nagelkerke R
square & accuracy
# ------------------------------------------------------------

# Local Model Evaluation - Multicollinearity Test
vif(model2)


# Local Model Evaluation - Significance Test
summary(model2)

# ------------------------------------------------------------
# Global model evaluation
# ------------------------------------------------------------

# ------------------------------------------------------------
# Nagelkerke R square
# It is always between 0 and 1. The higher the better.
# ------------------------------------------------------------
PseudoR2(model2, which = "Nagelkerke")

# ------------------------------------------------------------
# Global model evaluation - Prediction Accuracy
# ------------------------------------------------------------


#Create a Confusion Matrix to determine prediction accuracy of
the model
#Cut off
Conf(model2, cutoff = 0.14)

# Prediction accuracy
print(paste('Prediction Accuracy Model 2:',
prediction_accuracy <- (1007+1537)/ 4140))
```

```
# # ANOVA tests whether the variance in a set of data
explained by the model is significantly greater than the
unexplained variance.
# # Thus, we want this test to be significant (p<.05).
# # The test proceeds in a step wise manner, adding one of the
independent variables
# # in each step. We are only interested in the value in the
last step.
# # Here it is significant so ok.
anova(model2, test="Chisq")

# # Hosmer Lemeshow test whether the predicted values and the
actual values are significantly different.
# # "Renew" identifies the observed binary, "fitted"model" the
predicted.
# # The test partitions the data into groups and compares for
each one
# # whether there are differences between predictions and
observations.
# # "g=10" is th edefault choice for the number of groups the
test uses.
# # For a good model performance we want them to be NOT
different.
# # Thus, we want the Hosmer Lemeshow Test to be INSIGNIFICANT
(>.05)!
# # Here it is significant, thats not a good model.
h3<-hoslem.test(LastCampaign$Renew, fitted(model2), g=10)
h3

#---------------------------------------------------------------
# Standardization of Model Coefficients
# For comparison of affect of independent variables on
dependent "Renew" to see which variable has the strongest
effect.
#---------------------------------------------------------------
model2.std <- beta(model2)
model2.std

# Write standardized coefficients to csv file for
visualization
write.csv((coef(model2.std)[,
"Estimate"]),"Standarised_Coefficients.csv")

#---------------------------------------------------------------
# Odds-Ratio Calculations
# To see how the log odds "Renew" affected by 1 unit increase
in variables in model, minus the intercept
#---------------------------------------------------------------
exp(coef(model2))
```

```
# Write odds-ratios estimates and confidence intervals to csv
file for visualisation
write.csv((cbind(exp(coef(model2)), exp(confint(model2)))),
"Odds_Ratio.csv")

# #--------------------------------------------------------------
# # New Campaign Calculation of  Target Customers at
probability of Renew <.14 = Targeted
# #--------------------------------------------------------------


# Read in NewCampaign Data
NewCampaign <-read.csv("A3_Prediction.csv", header=TRUE)

str(NewCampaign)

#---------------------------------------------------------------
# Re-coding of Independent variables in NewCampaign
#
# Currently, the independent variables are character format
# To run a logistic regression, we need to convert them to
binary 0 or 1
# We re code them by using if else function.
# Change Categorical values into 1 or 0 for each column in
NewCampaign
#--------------------------------------------------------------

NewCampaign$PhoneService <-
ifelse(NewCampaign$PhoneService=="Yes",1,0)
NewCampaign$Contract_LT <-ifelse(NewCampaign$Contract=="Long
term",1,0)
NewCampaign$PaperlessBilling <-
ifelse(NewCampaign$PaperlessBilling=="Yes",1,0)

NewCampaign <- NewCampaign[, !names(NewCampaign) %in%
c('customerID','gender','Partner', 'Dependents',
'TotalCharges', 'Contract')]



NewCampaign$Predicted.Renew <- predict(model2, newdata =
NewCampaign, type = "response")

str(NewCampaign)
#---------------------------------------------------------------
# Threshold Calculation Targeted vs. Untargeted for New
Campaign
#
# Advertising-to-sales ratio = Total advertising expenses /
Sales revenue
#
```

```
# When a threshold is 7 (cost per customer contact) / 50
(value of retaining
# an existing customer), 0.14 means breakeven point of not
churn. If the
# predicted probability of ‘renew’ is greater than 0.14 (or
predicted
#
probability of ‘churn’ is less than 0.86), then this customer
will be the
# targeted customer.
#--------------------------------------------------------------

print(paste('Target Renew Threshold for Future Campaign:', 7 /
50))


# Classify customers "targeted" vs. "untargeted" based on the
probabilities "of "Renew"
# If probability "Renew" < 0.14, then Target = 0
# If probability "Renew" >  0.14, then Target = 1

NewCampaign$Target.Customer <-
ifelse(NewCampaign$Predicted.Renew > 0.14, 1, 0)
str(NewCampaign)
write.csv(NewCampaign, "NewCampaign Data.csv")

#---------------------------------------------
# Step 5: Predictions

# Advertising Campaign Statistics Prediction This Month
#
#---------------------------------------------

# Print Number of customers who will be targeted (Probability
Renew > 0.14)
num_target_customers <- sum(NewCampaign$Target.Customer == 1)

print(paste('Number of Targeted Customers New Campaign:',
num_target_customers))

# Print Ratio of customers who are targeted vs untargeted new
campaign
print(paste("Target Customer Ratio New Campaign:",
num_target_customers/nrow(NewCampaign)))


# Number of targeted customer = 1118, and they are ordered
from lowest
# probability to the highest probability of churn (all will be
contacted)
```

```
targeted.probabilities <- sort(NewCampaign$Predicted.Renew,
decreasing = TRUE)[1:1118]

# Predicted Revenue = Probability of Renewal of Targeted
Customers × Worth for retaining an existing customer
predicted.revenue <- round(targeted.probabilities * 50, 2)

# Predicted Profit = Predicted Revenue - Predicted Cost
predicted.profits <- sum(predicted.revenue) -
(num_target_customers*7)
predicted.profits

#--------------------------------------------------------------
# Advertising Campaign Statistics Last Campaign vs. New
Campaign
#--------------------------------------------------------------

last_campaign_renewals <- sum(telecomm_df$Renew == 1)
num_customers_contacted <- nrow(telecomm_df)

print(paste('Total Cost of Advertising (Last Campaign): $',
num_customers_contacted* 7))

# Calculate Predicted Cost New Campaign
print(paste('Total New Campaign Cost (Number of Targeted
Customers x $7): $',num_target_customers* 7))

# Acutal Profit Last Campaign
print(paste('Total Profit of Last Campaign (Revenue - Cost):
$', last_campaign_renewals*50 - num_customers_contacted* 7))

# Predicted Profit for New Campaign
print(paste('Total Predicted Profit of New Campaign (Revenue -
Cost): $',sum(predicted.revenue), '- $',
sum(num_target_customers * 7), ' = $',predicted.profits))

# Acutal Profit per customer Last Campaign
print(paste('Total Profit Per Customer Contacted (Last
Campaign) $', round((((last_campaign_renewals * 50) -
(num_customers_contacted * 7))/num_customers_contacted), 2)))

# Predicted Profit per customer New Campaign
print(paste('Total Predicted Profit of New Campaign per
Contact: $', round((predicted.profits / num_target_customers),
2)))


# And for those we calculate the expected profits,
# which are $14142.45
# Expected Profit/Customer Contacted is $12.65
```

```
# Remember: In the initial sample with the untargeted
campaign,
# our profit amounted to $25320 only.
# Actual Profit Per Customer Contacted was $6.12

#--------------------------------------------------------------
# WHAT IF COST WAS $6
#--------------------------------------------------------------
print(paste('Target Renew Threshold for Future Campaign:', 6 /
50))
NewCampaign$Target.Customer6 <-
ifelse(NewCampaign$Predicted.Renew > 0.12, 1, 0)

# Print Number of customers who will be targeted (Probability
Renew > 0.14)
num_target_customers6 <- sum(NewCampaign$Target.Customer6 ==
1)

print(paste('Number of Targeted Customers New Campaign @$6:',
num_target_customers6))

# Number of targeted customer = 1118, and they are ordered
from lowest
# probability to the highest probability of churn (all will be
contacted)
targeted.probabilities6 <- sort(NewCampaign$Predicted.Renew,
decreasing = TRUE)[1:1184]

# Predicted Revenue = Probability of Renewal of Targeted
Customers × Worth for retaining an existing customer
predicted.revenue6 <- round(targeted.probabilities6 * 50, 2)

# Predicted Profit = Predicted Revenue - Predicted Cost
predicted.profits6 <- sum(predicted.revenue6) -
(num_target_customers6*6)
predicted.profits6

#--------------------------------------------------------------
# WHAT IF COST WAS $5
#--------------------------------------------------------------
print(paste('Target Renew Threshold for Future Campaign:', 5 /
50))
NewCampaign$Target.Customer5 <-
ifelse(NewCampaign$Predicted.Renew > 0.10, 1, 0)

# Print Number of customers who will be targeted (Probability
Renew > 0.14)
num_target_customers5 <- sum(NewCampaign$Target.Customer5 ==
1)
```

```
print(paste('Number of Targeted Customers New Campaign @$5:',
num_target_customers5))

# Number of targeted customer = 1118, and they are ordered
from lowest
# probability to the highest probability of churn (all will be
contacted)
targeted.probabilities5 <- sort(NewCampaign$Predicted.Renew,
decreasing = TRUE)[1:1243]

# Predicted Revenue = Probability of Renewal of Targeted
Customers × Worth for retaining an existing customer
predicted.revenue5 <- round(targeted.probabilities5 * 50, 2)

# Predicted Profit = Predicted Revenue - Predicted Cost
predicted.profits5 <- sum(predicted.revenue5) -
(num_target_customers5*5)
predicted.profits5
```