PREDICTING SEASON TICKET HOLDER RENEWAL

Team R.P.T.
Varsha Prabhakar, Sudeep Kurian, Mayank Daga, Abhishek Talwar
Purdue University, Krannert School of Management, West Lafayette, IN 47907
prabhakv@purdue.edu; kurians@purdue.edu; mdaga@purdue.edu; talwar2@purdue.edu

ABSTRACT

Season ticket holders are strategically and financially the most important to professional sports teams. A major concern for these teams is the churn of existing season ticket holders. They are always concerned if the marketing strategy for their team is perfectly aligned to maximize the number of their fans buying season tickets. Actual renewal data for the season ticket holders of an NFL team was tracked against previous year's data to find out any pattern for non-renewal of season tickets. Demographic data of approximately three thousand account holders were integrated to identify the effectiveness of these features in determining churn. Survey data of a thousand season ticket holders were analyzed using Microsoft Azure's text analytics API to identify difference between the feedback of churners as well as loyalists. Exploratory data analysis suggests that middle-aged population is highly likely to renew their tickets and married males constitute a majority portion of this population. Most renewals have occurred during the November month and decreases in each consequent month. Also, most people who renew their tickets don't contact Indianapolis Colts Sales representatives but those who do are more probable to renew their tickets. A logistic model identifies the probability of renewals with an accuracy of 78% at present. Our team, in the final phase, intends to perform more sophisticated modelling and A/B testing to identify if the new ticketing scheme has got better response from its fans and how could the team utilize these findings to actively manage its season ticket holders to reduce churn.

INTRODUCTION

Most teams are employing analytics to retain its season and partial-season ticket holders and identifying indicators that would help them maximize revenues. Understanding the relationship between sport consumer and the sport team and what drives the consumer to be loyal is significant for sports teams to generate revenue. Firms should have a balance between acquiring new customers and retaining old ones. New customers are effective in generating word-of-mouth about their experience with the firm while it is easier to predict utilization patterns for existing customers leading to lower servicing costs (East, Hammond, & Gendall, 2006). Funk and James (2001) developed the Psychological Continuum Model (PCM) to explain how sport fans progress through four stages of team identification, namely awareness, attraction, attachment and allegiance. Research on season ticket holders suggests that external circumstances (McDonald & Stavros,

2007) as well as internal constraints and lack of attachment (Kim & Trail, 2010) may influence churn rates. Technology has also changed the way fans bought season tickets. A 2015 article by Forbes states that secondary ticket arena has become increasingly popular because of the flexibility it provides in buying tickets for any game. However, different sports teams have responded by bringing in loyalty programs, experiential opportunities of insider access, ticket reselling features using mobile apps to provide a better user experience to fans that any other ticket reselling facility can provide. Therefore, identifying the right metric to predict the probability of churn for season ticket holders becomes a crucial exercise for sports teams. Lee (2003) stated that areas such as price of tickets, game experience, promotions/giveaways, stadium features, and the teams involved are among the few reasons tickets may be purchased. In this paper, we sought to find the drivers of season tickets renewals every year for an NFL team and build a predictive model to find the probability of their renewal. We also identify the effectiveness of the new ticketing scheme and how fans have responded with respect to the old method of renewal of tickets. The results of this research would help the sports team to gain insight into consumer behavior and develop more comprehensive marketing strategy to market their season ticket packages.

LITERATURE REVIEW

Season ticket holders are widely viewed and analyzed as fans in the sport marketing literature. McDonald (2010) describes STHs as highly involved consumers who attend games in an effort to express themselves and their core values. In an industry where fan loyalty is remarkably high, season ticket packages offer guaranteed revenue streams for organizations involved in spectator sport. Since customers rarely redirect their support over to other sports franchises, it becomes essential for organizations to understand what motivates season ticket holders (STH) to renew their season tickets. A study conducted to understand season ticket renewal across different account types using k means clustering and MANOVA gave us six STH clusters of statistically significant differences (Warren CJ, 2015).

STHs provide substantial direct revenue through ticket purchases, while indirect revenue through merchandise, food and beverage sales. Years in membership (tenure), games attended, and overall satisfaction were found to be closely related to renewal behavior by applying a combination of descriptive analysis, logistic Regression, k-means clustering (Heath J. McDonald, 2010).

STH renewal could also be understood in terms of geographical locations. The analysis involved detailed stadium attendance information on 13,892 STH of a professional German Bundesliga team inclusive of basic socio-demographic STH data. In addition to these, the potential role of increasing opportunity costs resulting from larger home-stadium distances. Using this dataset, a binary probit model was estimated to analyze potential determinants of STH stadium attendance demand (Loyalty). Loyalty via geographical location suggest that STH stadium admission increases for those STHs located in close or distant proximity to the stadium. (Schreyer, Schmidt, Torgler, 2017).

Martin and Goldman (2015) talked about a sport brand detachment process following a four-part sequential process of a breakdown trigger, iterative decline, disengagement incident and exit phase.

In our study, we are attempting to segment customers based on volume of seats booked and corresponding revenue generated through these customers.

Study	Motivation	Algorithms	Results
McDonald, Heath, Karg, Adam J. and Leckie (2014)	Predicting which season ticket holders will renew and which will not	Multi-variable regression model	Utilization, satisfaction and tenure are closely aligned with actual renewal behavior
Warren CJ (2015)	Industrial Marketing in Sport: Understanding Season Ticket Renewal Across Account Types	k-means clustering, MANOVA	Multivariate null hypothesis of equality of the means across the six STH clusters indicated statistically significant differences
Heath J. McDonald (2010)	The Factors Influencing Churn Rates Among Season Ticket Holders: An Empirical Analysis	Descriptive analysis, Logistic Regression, k-means clustering	Years in membership (tenure), games attended, and overall satisfaction are closely related to renewal behavior
Schreyer, Schmidt, Torgler (2017)	Predicting season ticket holder loyalty using geographic information	Probit Regression, Generalized Linear Model	Loyalty via geographical location suggest that STH stadium admission increases for those STHs located in close or distant proximity to the stadium. Beyond that we observe that stadium admission increases for those holding fewer season tickets and those who pay higher prices.
Sam Healy (2018)	Analysis of Consumer Demand for American Sports Leagues	Linear Regression	NFL's median ticket price include away team playoff team, the home team city's population, average attendance, and the team's net worth.
Martin and Goldman (2015)	A process model of sport fan detachment	Qualitative research design on behavioral observations	Provide evidence of sport brand detachment process following a four-part sequential process of a breakdown trigger, iterative decline, disengagement incident and exit phase.
Matt Butler (2013)	Season Seat Holder Retention in Minor League Baseball	Descriptive statistics and paired t-tests	Entertainment followed by aesthetics and family are the top motivating factors that contribute to a sport fan's motivation to attend sporting events
Daniel C. Funk, Kevin Filo, Anthony A. Beaton, Mark Pritchard (2009)	Measuring the Motives of Sport Event Attendance	MANOVA, Multiple linear regression	Five facets of motivation: Socialization, Performance, Excitement, Esteem, and Diversion (SPEED). Multiple linear regression results indicate three facets explain 30% of the variance in the frequency of game attendance.
Mark Brown, Darian Misko, Dennis Lee (2009)	Membership retention in professional sports organizations	Partial Least Square Analysis	Team performance, sportscape features, and media advertising significantly influence decisions to renew season tickets. Fan identification and the presence of star players do not appear to have any such influence.

Table 3.1 Literature Review

DATA

The data sources that we used for this study are renewal data for 2019 and 2018, different demographics for each account and experience feedback collected for some of the accounts. The renewal dataset contais information about the base revenue, distance from stadium, tenure, attendance percentage and other factors. Different demographics of each account such as occupation, education, estimated household income and net worth.

Finally, a dataset containing the survey of about 1000 customers was also used for sentiment analysis. Each of these datasets are explained in the table below:

Variable	Type	Description
Renewed	Categorical	1: Season ticket member has renewed 0: Season ticket member has not yet renewed
Tenure	Numeric	Number of years being a season ticket member
AttendancePercentage	Numeric	The 2018 attendance percentage for the season games
DistanceFromStadium	Numeric	The distance of the account holder's zip code from the stadium
NumberOfCalls	Numeric	Number of calls a customer service rep has made to this account during the 2019 renewal program
NumberOfVoicemails	Numeric	Number of voicemails a customer service rep has made to this account during the 2019 renewal program
NumberOfEMails	Numeric	Number of emails a customer service rep has sent to this account during the 2019 renewal program
BaseRevenue	Numeric	Their 2018 season ticket account value
BaseNumSeats	Numeric	Season ticket number of seats that is up for renewal for the 2019 season (their 2018 season ticket number of seats)
Gender	Categorical	Gender of the account holder (M/F)
LengthOfResidence	Numeric	How long the account holder has lived in their current residence
IncomeSector	Categorical	A feature created based on the income data, with three buckets created: High, medium, low

Field	Type	Description
TicketingSystemAccountID	Numeric	Ticketing AccountID number (same one as the Renewal Dataset AcctId)
AgeIncrements	Numeric	Age Range of the account holder
Occupation	Categorical	Occupation the account holder currently partakes in

Education	Categorical	Education level of the account holder
MaritalStatus	Categorical	Marital status of the account holder
EstHhldIncome	Categorical	Estimated household income level of the account holder
HomeOwnerRenter	Categorical	Indicator for if the account holder owns or rents their place of residence
LengthofResidence	Categorical	How long the account holder has lived in their current residence
HomePropertyTypeDetail	Categorical	What type of property the account holder currently resides in
HomeMarketValue	Categorical	Market value of the account holder's current residence
Vehicle	Categorical	What type of vehicle the account holder currently owns
DiscretionaryIncomeIndex	Numeric	Indicator for how much discretionary income the account holder has (the higher the value the more discretionary income they have to spend on non-necessities) - Keep in mind our product is driven by this portion of income
HomeAssessedValueRanges	Categorical	What the account holder's home was assessed at
NetWorthGold	Categorical	The net worth of the individual

Table 4.1 Data Dictionary

METHODOLOGY

In this paper, we sought to find the drivers of one of the NFL team fans to renew their season tickets every year and build a predictive model to find the probability of their renewal. We also identify the effectiveness of the new ticketing scheme and how the fans have responded with respect to the old method of renewal of tickets. Hence the purpose for our predictive modelling is twofold:

- Prediction: Predicting if a customer Account ID will renew Season Tickets given demographics and marketing details with the highest accuracy
- Interpretation: Interpreting which characteristics play an important role for members who renew/ don't renew tickets to make better marketing strategies

Data Pipeline

We used the following pipeline for generating our predictive models.

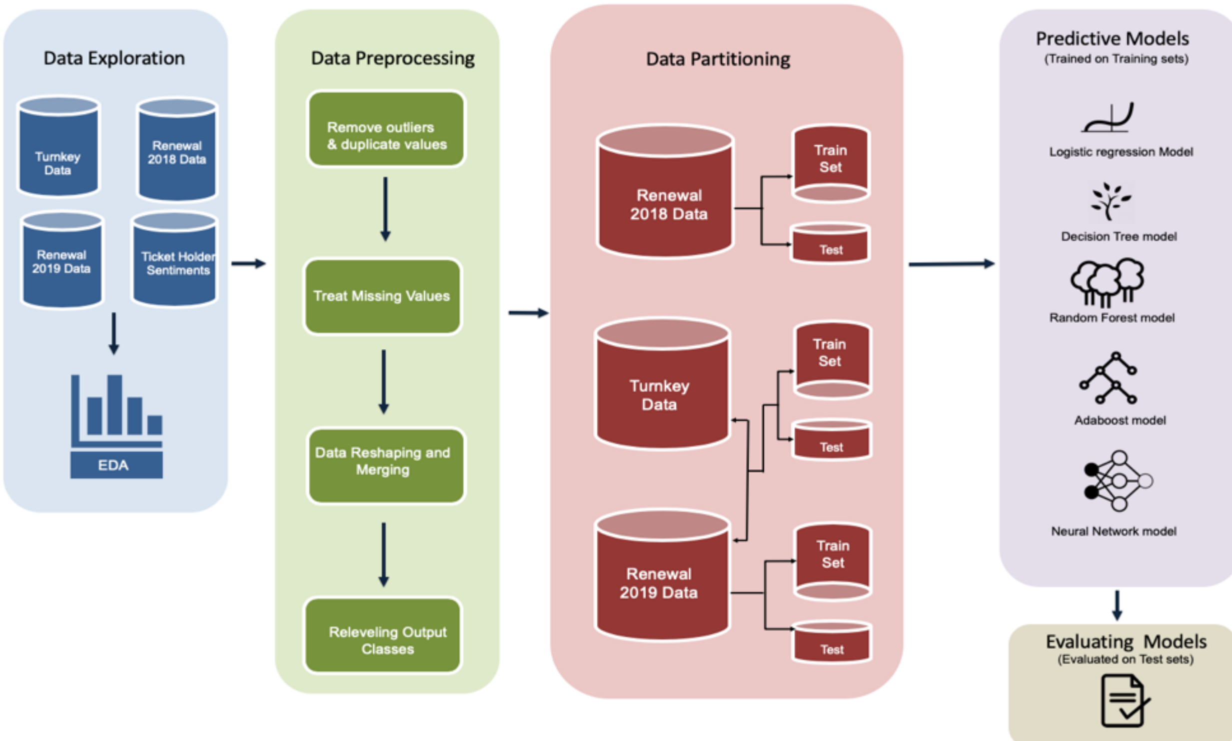

Figure 5.1 Process Flow

For our model, we used three datasets, the 2018 Renewal dataset, 2019 Renewal dataset (for renewals from November 2018 to January 2019) and the Turnkey dataset. Both datasets had duplicate values which were removed. Then, both datasets were merged on Account Id. Upon merging, we had close to 4,000 data points with the merged information.

Pre-Processing

After merging the data, we Pre-Processed the features. We first created a response variable feature, called 'Renewed' for Renewal 2018 dataset. The assumption for creating this feature was that, if the Renewal Revenue for any account Id was 0, that meant that the Account holder had not renewed for the next year's season tickets.

We had a few rows which had missing data. These rows were removed for the current basic model. The future scope for this problem could be imputing values intuitively. We also created a new feature called 'incomeSector'. It checked on the base income level for an account and converted it to three distinct sector groups: low, medium and high.

Cleaning the data also involved some string operations. An example of which would be, the length of residence was a string value which contained values like '10 Years'. These text values had to be converted to numeric values. Post this Pre-Processing, we split the dataset into training and test sets.

Splitting the data

For splitting the data, we used R's Caret package's 'createDataPartition' function. This function creates a stratified sample which would be representative of all the classes for our response variable (Renewed). Training data consisted of 75% of data while the test dataset consisted of the remaining 25% data. Increasing the training data always adds information and should improve the fit. We split the train and test dataset in such a way because we wanted more data points for training our model to represent majority of the data.

Releveling Output classes

Upon close inspection of the output classes, we found out that the output classes of renewals(X1) and non-renewals(X0) were highly unbalanced as shown in the illustration below. We up-sampled the data to have a proportionate balanced sample of outputs to generate our predictive classification model on.

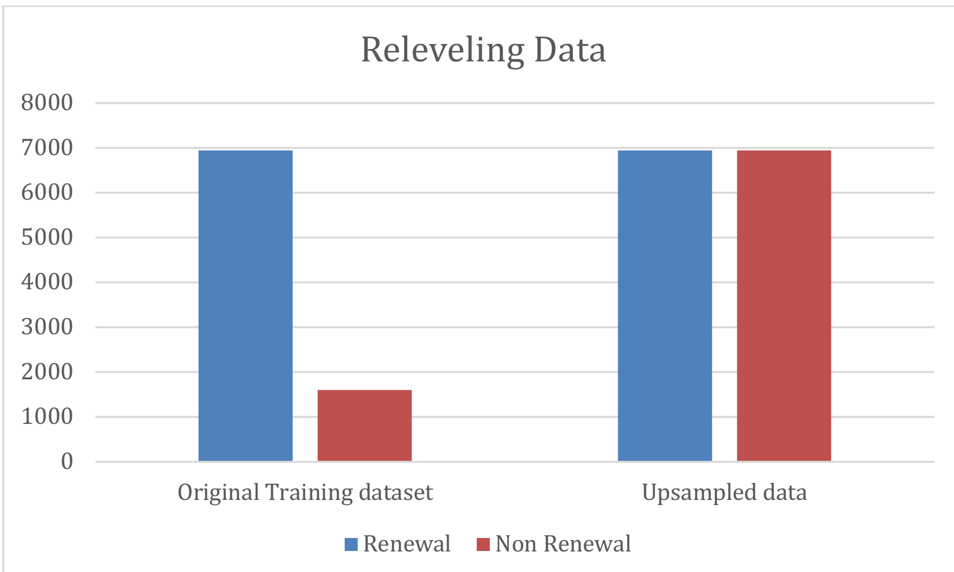

Figure 5.2 Releveling output classes

In order to evaluate our model, we will be using following metrices to gauge different models:

- AUC (Area under the Curve) for the ROC curve: This will let us know how the effective is the model overall. A higher value of AUC tells us that our model performs really well in predicting both classes.
- Sensitivity: This metric will tell us how accurate our model is in predicting the people who will actually renew their season tickets out of all the people who renew their season tickets. The business problem at hand is to increase number of people who renew their season tickets and increase revenue generated out of existing season ticket holders who will be renewing their tickets in the coming season. A high sensitivity score will enable us to target those who will renew their tickets effectively and try to increase the revenue generated out of this segment.

MODELS

Logistic regression model on merged data

Logistic regression is one of the most robust and interpretable classification models in the list of classification models. We used Logistic regression for interpreting the relationship between various parameters and the probability of renewing season tickets. The purpose of this model was to interpret the data. Because logistic regression models the predicted probability of class, we can generate insights by looking at the coefficients of various features and understand how they affect the probability of ticket renewal. This model was generated on the 4000 common data points among the turnkey data and 2018 Renewal data. Because the number of merged rows was less, we can't depend much on the accuracy of this model to perform well on real time data. Here are the results obtained from the Logistic Regression Model.

Datasets Used: Turnkey data, 2018 Renewal

Features Used: Tenure, Attendance Percentage, Distance from Stadium, Number Of Calls, Number Of Voicemails, Number Of Emails, Base Revenue, Base Number Seats, Gender, Length of Residence and Income Sector

```
Confusion Matrix and Statistics

          Reference
Prediction   0   1
         0 128  34
         1  86 273

               Accuracy : 0.7697
                 95% CI : (0.7311, 0.8052)
    No Information Rate : 0.5893
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.506
 Mcnemar's Test P-Value : 3.23e-06

            Sensitivity : 0.5981
            Specificity : 0.8893
         Pos Pred Value : 0.7901
         Neg Pred Value : 0.7604
             Prevalence : 0.4107
         Detection Rate : 0.2457
   Detection Prevalence : 0.3109
      Balanced Accuracy : 0.7437

       'Positive' Class : 0
```

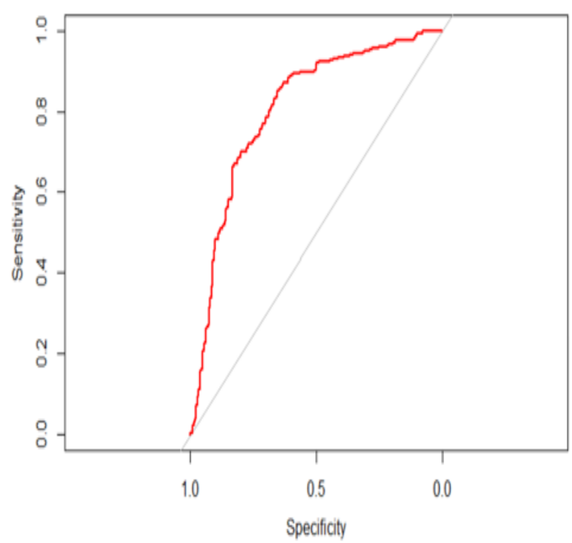

Figure 6.3 ROC Curve

The following is the ROC curve with an Area Under the Curve of 0.8054

Logistic regression model on 2019 Renewal data

Another Basic logistic regression model was generated on 2019 renewal dataset. This dataset included Members who renewed their season tickets from November 2018 to January 2019. Hence the data was a little skewed towards non-renewals. The non-renewals did not mean that the account ID failed to renew. That particular member might just renew in the future. We used up-sampled data to relevel the responses to have a robust model. Here are the results obtained from this model.

Datasets Used: 2019 renewal

Features Used: Tenure, Attendance Percentage, Distance from Stadium, Number of Calls, Number Of Voicemails, Number Of Emails, Base Revenue, Base Number of Seats

```
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0  628  458
         1  219 1202

               Accuracy : 0.73
                 95% CI : (0.7121, 0.7473)
    No Information Rate : 0.6621
    P-Value [Acc > NIR] : 1.545e-13

                  Kappa : 0.4354
 Mcnemar's Test P-Value : < 2.2e-16

            Sensitivity : 0.7414
            Specificity : 0.7241
         Pos Pred Value : 0.5783
         Neg Pred Value : 0.8459
             Prevalence : 0.3379
         Detection Rate : 0.2505
   Detection Prevalence : 0.4332
      Balanced Accuracy : 0.7328

       'Positive' Class : 0
```

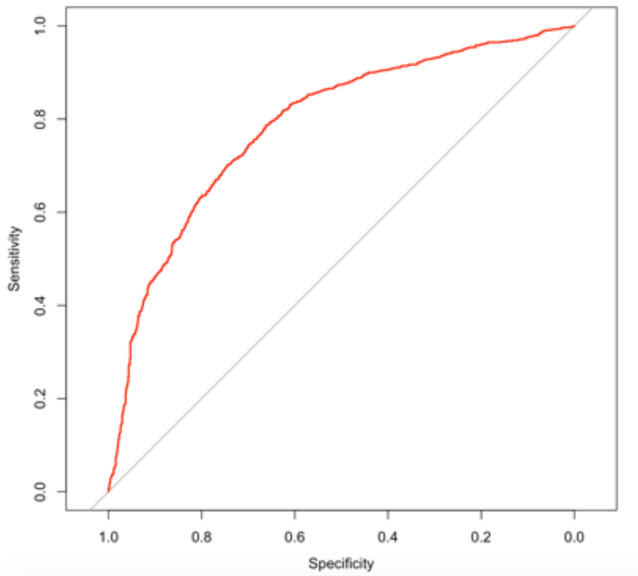

Figure 6.4 ROC Curve

The following is the ROC curve with an Area Under the Curve of 0.78

Basic Decision Tree Model with 2019 Data

We used a decision tree model for another easily interpretable model. Decision tree diagrams are really easy to interpret. However, it should be noted that decision trees can become really overfit and hence we used 10-fold cross validation and generated trees based on Gini Index node purity. The trees were also pruned to have much robustness. The following is a diagram of our decision tree model.

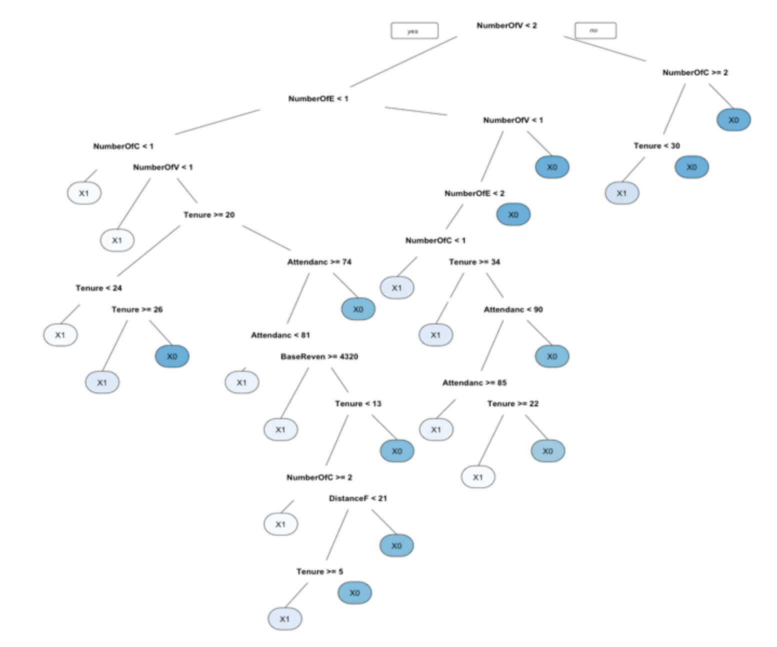

Figure 6.5 Decision Tree

It can be seen that the major branch cuts off at number of voicemails. In general, if the number of voicemails is large, there are higher chances of the person not renewing. This can be indicative of the fact that the account ids with higher voice mails are not really interested in renewing their tickets and repeated targeting on these customers is going in vain. Many other insights can be generated out of this basic model.

```
Confusion Matrix and Statistics

          Reference
Prediction   X1   X0
        X1 1123  298
        X0  364  722

               Accuracy : 0.7359
                 95% CI : (0.7182, 0.7531)
    No Information Rate : 0.5931
    P-Value [Acc > NIR] : < 2e-16

                  Kappa : 0.4584
 Mcnemar's Test P-Value : 0.01153

            Sensitivity : 0.7552
            Specificity : 0.7078
         Pos Pred Value : 0.7903
         Neg Pred Value : 0.6648
             Prevalence : 0.5931
         Detection Rate : 0.4479
   Detection Prevalence : 0.5668
      Balanced Accuracy : 0.7315

       'Positive' Class : X1
```

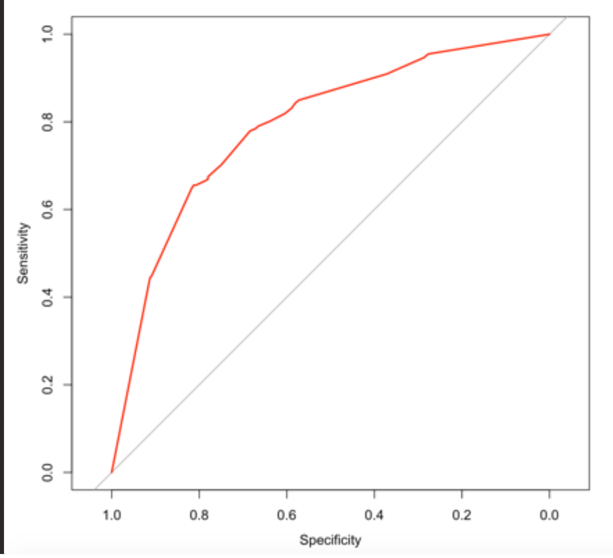

Figure 6.6 ROC Curve

The following is the ROC curve with an Area Under the Curve of 0.79

Random Forest Model on 2019 Dataset

Training a random forest model will be useful for models with good predictive power. We tuned the hyperparameters to obtain the best random forest tree by trying out various number of trees in a forest and varying the number of parameters and length of the tree. As we can see the ROC for the random forest model is one of the best with 0.8047.

Datasets Used: 2019 renewal

Features Used: Tenure, Attendance Percentage, Distance from Stadium, Number of Calls, Number Of Voicemails, Number Of Emails, Base Revenue, Base Number of Seats

```
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0  637  228
         1  449 1193

               Accuracy : 0.73
                 95% CI : (0.7121, 0.7473)
    No Information Rate : 0.5668
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.4366
 Mcnemar's Test P-Value : < 2.2e-16

            Sensitivity : 0.5866
            Specificity : 0.8395
         Pos Pred Value : 0.7364
         Neg Pred Value : 0.7266
             Prevalence : 0.4332
         Detection Rate : 0.2541
   Detection Prevalence : 0.3450
      Balanced Accuracy : 0.7131

       'Positive' Class : 0
```

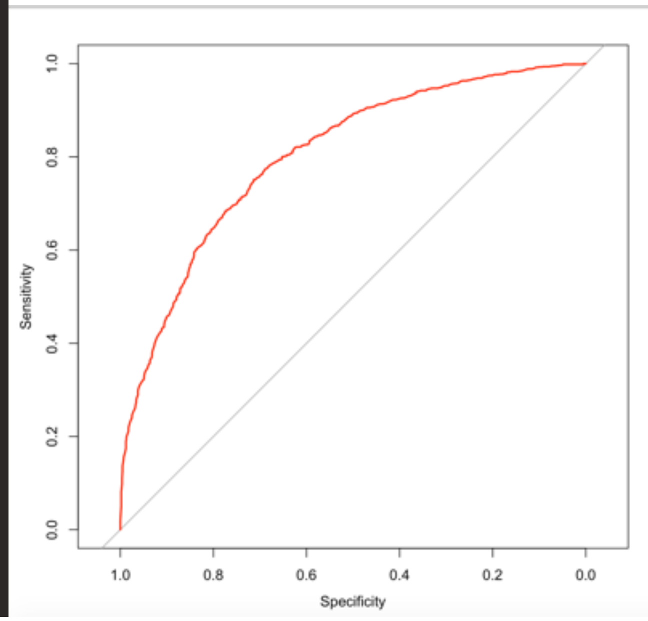

Figure 6.7 ROC Curve

However, the Sensitivity and specificity of this model isn't as great as the simpler logistic regression model we used above. The screenshot above has the positive class leveled as 0, hence the sensitivity and specificity values in the screenshot are interchanged.

Adaboost Model

Adaboost is short for Adaptive boosting and is a class of boosted tree algorithm which focuses on improving the predictive power of a week classifier by learning from various weak classifiers to generate a strong classifier. Here are the results of the Adaboost classification on the 2019 dataset.

Datasets Used: 2019 renewal

Features Used: Tenure, Attendance Percentage, Distance from Stadium, Number of Calls, Number Of Voicemails, Number Of Emails, Base Revenue, Base Number of Seats

```
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0  650  234
         1  436 1187

               Accuracy : 0.7327
                 95% CI : (0.715, 0.75)
    No Information Rate : 0.5668
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.4436
 Mcnemar's Test P-Value : 8.145e-15

            Sensitivity : 0.5985
            Specificity : 0.8353
         Pos Pred Value : 0.7353
         Neg Pred Value : 0.7314
             Prevalence : 0.4332
         Detection Rate : 0.2593
   Detection Prevalence : 0.3526
      Balanced Accuracy : 0.7169

       'Positive' Class : 0
```

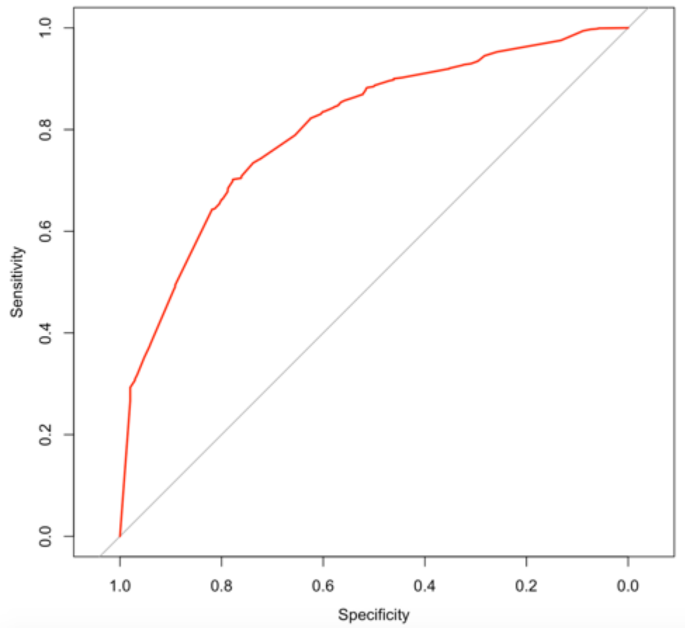

Figure 6.8 ROC Curve

The Adaboost Model gives an AUC of 0.802 for the ROC curve.

Random Forest Model on 2018 Dataset

The 2018 Dataset is complete and consist of a year's renewal stats. This dataset is more accurate for modeling; however, it has a few missing features. The following is the test statistics for a random forest model with uptrained training data.

```
          Reference
Prediction    0    1
         0  200  562
         1  319 1719

               Accuracy : 0.6854
                 95% CI : (0.6678, 0.7025)
    No Information Rate : 0.8146
    P-Value [Acc > NIR] : 1

                  Kappa : 0.1177
 Mcnemar's Test P-Value : 3.545e-16

            Sensitivity : 0.38536
            Specificity : 0.75362
         Pos Pred Value : 0.26247
         Neg Pred Value : 0.84347
             Prevalence : 0.18536
         Detection Rate : 0.07143
   Detection Prevalence : 0.27214
      Balanced Accuracy : 0.56949

       'Positive' Class : 0
```

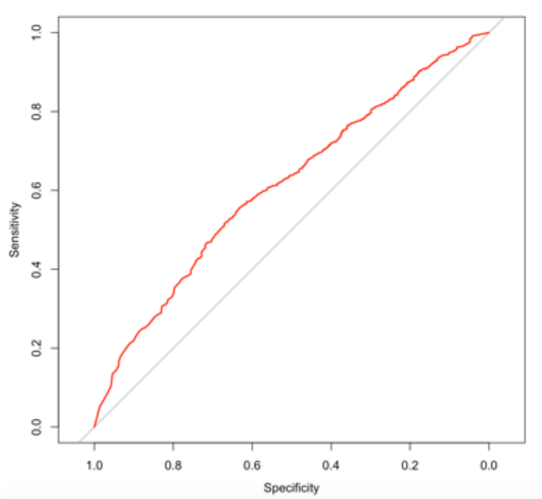

Figure 6.9 ROC Curve

The above Random Forest model provides an AUC of 0.6127. This is because the model doesn't have as many features as the Random Forest model trained before.

New Feature Creation for Analysis

Besides the predictive modeling, customer analysis was also carried out to segment the consumers based on their renewal behavior. Two features were created by comparing the renewed features (Number of seats, Revenue) with Base features (Number of seats, Revenue)- Ternary Revenue Change and Ternary Num of Seats Change. The value of these features are as follows: -

Ternary Revenue Change (TRC): -1 when Renewed Revenue> Base Revenue
0 when Renewed Revenue=Base Revenue
+1 when Renewed Revenue>Base Revenue

Ternary Num of Seats Change (TSC): -1 when Renewed Seats> Base Seats
0 when Renewed Seats=Base Seats
+1 when Renewed Seats>Base Seats

These two features were created for both the years 2018 and 2019. However, the customers were segmented based on 2018 features since it represented the overall behavior in a season. The segmentation table for customers are as follows:

Customer Segment	Name given	Criteria
A	Hungry Howie	TRC=-1,0 TSC=1
B	Touchy Ted	TRC=1 TSC=1
C	Pricey Penney	TSC=-1 TRC=1,0
D	Richie Rich	TSC=-1 TRC=-1
E	Stingy Steve	TRC=-1,0 TSC=0
F	Lucky Lucy	TRC=1 TSC=0

Table 6.1 Customer Segmentation

RESULTS

Base Revenue and Renewed Revenue figures are the estimated and the actual revenue generated by the NFL team. The 2018 and ongoing 2019 figures are mentioned in the table below:

Revenues	2018	2019 Ongoing
Base Revenue	$30,175,060	$29,738,131
Renewed Revenue	$29,361,085 (3% drop from base)	$17,125,040 (42% more to go)

Table 7.1: Total Revenues

Seats	2018	2019 Ongoing
Base Seats	29,201	28,415
Renewed Seats	28,089	16,373
Reduction in seats	1,112 (4% decrease in base seats)	12,042 (58% more to go)

Table 7.2: Total Seats sold

The reduction in revenues for the NFL team can be attributed to two factors:

- Lesser number of seats purchased by the customers
- Renewing the season tickets at a lower price point that before

Based on the ternary seat change and revenue change parameters, contingency tables for customer distribution, aggregate revenue distribution, and revenue change distribution were plotted for 2018 and 2019. The tables are mentioned below:

Revenue Change/ Seats Change	-1	0	1	Grand Total
-1	488	168	5	661
0		8,066	1	8,067
1	5	478	97	580
Grand Total	493	8,712	103	9,308

Revenue Change/ Seats Change	-1	0	1	Grand Total
-1	87	28	1	116
0		5,360	1	5,361
1		6	19	25
Grand Total	87	5,394	21	5,502

Table 7.3a: Customer Distribution Table 2018

Table 7.3b: Customer Distribution Table 2019

The customer distributions for 2018 and 2019 (Table 7.3a and 7.3b) show that more customers have reduced their expenditure on buying season tickets but most of them have remained consistent with their spending.

Revenue Change/ Seats Change	-1	0	1	Grand Total
-1	$1,632,824	$559,431	$20,802	$2,213,057
0		$25,002,770	$2,680	$25,005,450
1	$18,744	$1,748,304	$375,530	$2,142,578
Grand Total	$1,651,568	$27,310,505	$399,012	$29,361,085

Table 7.4a: Aggregate Revenue Distribution Table 2018

Revenue Change/ Seats Change	-1	0	1	Grand Total
-1	$309,750	$138,760	$5,640	$454,150
0		$16,537,270	$1,980	$16,539,250
1		$15,620	$116,020	$131,640
Grand Total	$309,750	$16,691,650	$123,640	$17,125,040

Table 7.4b: Aggregate Revenue Distribution Table 2019

The aggregate revenue distribution tables (Table 7.4a and 7.4b) depict the distribution of revenue for the NFL team with respect to customer behavior. The value pairs of (Revenue Change, Seat Change) indicate the revenue and seat purchase trends for the customer. While the major chunk of revenue has remained as the base revenue figures for 2018 and 2019, however, revenue share generated from negative growth customers is higher than the strong growth customers. This may be an issue of concern for the management of the NFL team.

Revenue Change/ Seats Change	-1	0	1	Grand Total
-1	($1,222,496)	($84,039)	($3,178)	($1,309,713)
0		$0	$0	$0
1	$1,934	$301,344	$192,460	$495,738
Grand Total	($1,220,562)	$217,305	$189,282	($813,975)

Table 7.5a: Revenue Change Distribution Table 2018

Revenue Change/ Seats Change	-1	0	1	Grand Total
-1	($222,300)	($7,203)	($72)	($229,575)
0		$0	$0	$0
1		$3,000	$47,944	$50,944
Grand Total	($222,300)	($4,203)	$47,872	($178,631)

Table 7.5b: Revenue Change Distribution Table 2019

Revenue Change Distribution tables (7.5a and 7.5b) show the revenue change distribution with respect to the number of seats change and revenue change. The overall negative figures in both the years show that customers have not been renewing their season tickets either at the same price point or have been renewing lower number of season tickets than before. The surplus revenue generated from consumers spending more than the base values is not enough to offset the negative impact. It can be inferred from tables 5a and 5b that 10% and 4% revenue loss for the NFL team accounts for consumers switching for lower price point tickets. But, a huge chunk of revenue loss is because customers have lowered their renewal number of seats.

Months/ Revenue Change	-1	0	1	Account ID Count
November	78	3,988	18	4,084
December	28	1,123	5	1,156
January	10	250	2	262
Grand Total	116	5,361	25	5,502

Table 7.6a: Month-wise renewal counts

Months/ Revenue Change	-1	0	1	Total Revenue Change
November	($146,039)	$0	$39,414	($106,625)
December	($63,766)	$0	$2,530	($61,236)
January	($19,770)	$0	$9,000	($10,770)
Grand Total	($229,575)	$0	$50,944	($178,631)

Table 7.6b: Month-wise renewal revenue change

Table 7.6a depicts the count of renewals that have been lower, equal and greater than the base revenue (expected) for the accounts for November, December and January. Table 7.6b shows the change in revenue figures for these months in each ternary revenue change category. The figures show that most people buy in November compared to December and January. This is also depicted by Figure 7.1 below

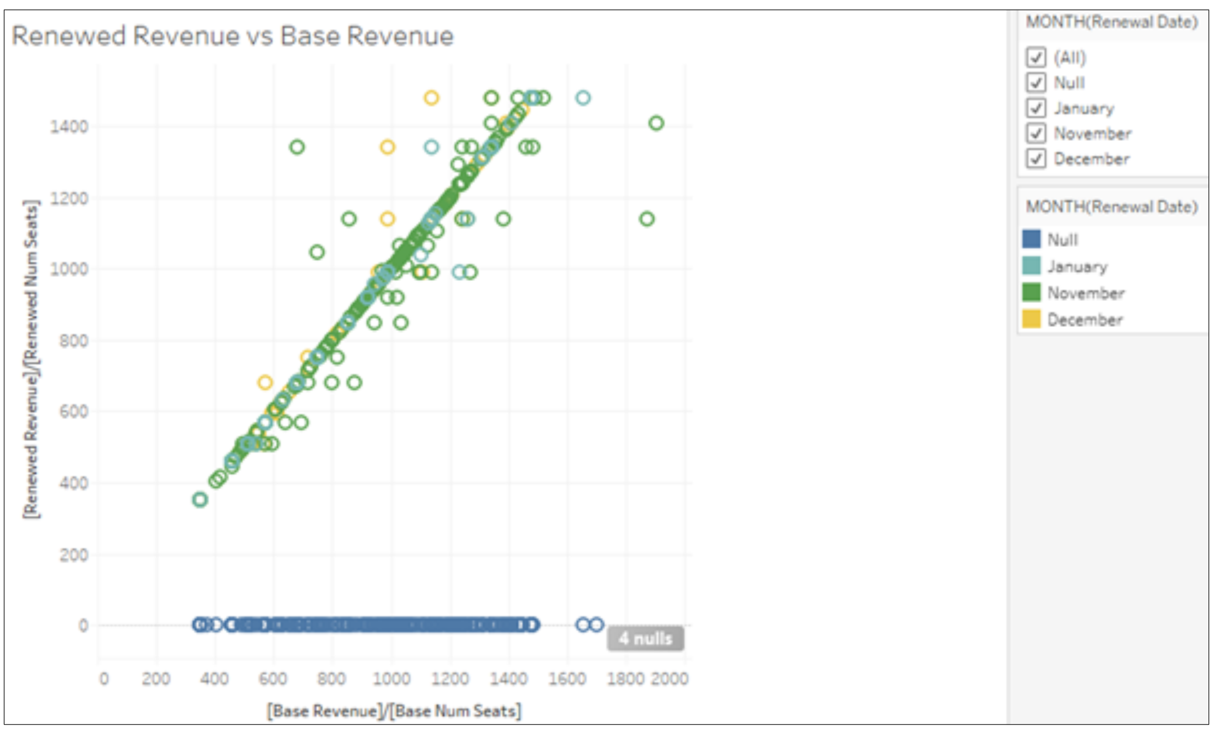

Figure 7.1

After analyzing the revenue figures and the trend, the renewals were classified based on the Revenue Change behavior in 2018 and 2019. Table 7.7 depicts this categorization.

Revenue Change 2019/2018	-1	0	1	Grand Total
-1	18	82	16	116
0	376	4615	370	5361
1	6	15	4	25
Grand Total	400	4712	390	5502

Number of customers	
Bad/Good to Worse	**486**
Good to Better	**401**

Table 7.7: Renewal classification based on Renewal Change behavior

The orange color code represents the consumers who are disillusioned or have exhibited a negative revenue growth trend from the base figures. Similarly, the green color coded cells denote consumers who have exhibited higher revenue growth compared to the base figures. The number of consumers who have shown a downward trend is more than the number of consumers who have shown an upward growth in terms of revenue for the NFL team.

Methodology: Segmentation of customer, explain revenue change ternary,

Based on the segmentation of consumers discussed in Methodology section, the revenue change numbers are shown in figure below:

Customer Segments	Aggregate Revenue Change 2018	Aggregate Revenue Change 2019
A	($3,178)	$2,426
B	$192,460	$8,212
C	$1,934	$0
D	($1,222,496)	($21,285)
E	($84,039)	($166,928)
F	$301,344	($1,056)
Grand Total	($813,975)	($178,631)

Table 7.8: Revenue Change with respect to different customer segments

This table shows that the greatest decline of revenues has been from D and E segment customers and therefore the marketing strategy for the NFL team company should be more focused on these consumer segments

Turnkey Analytics and Survey Data revealed the following information about customer segments.

Survey Data Analysis Results

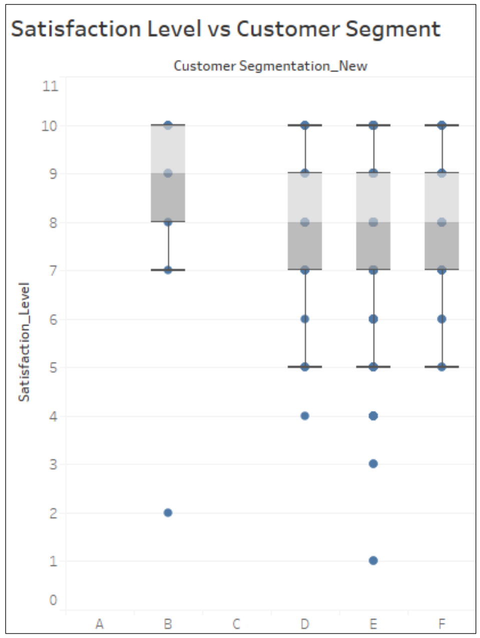

Figure 7.2: Satisfaction score vs Customer Segment

The survey satisfaction scores are not representative of the revenue growth numbers of each of the customer segment. This can also be seen in the Table 7.9, which describes the average satisfaction scores based on ternary revenue change for season ticket holders in 2018. The satisfaction scores are similar for negative as well as positive revenue growth scenarios implying that survey responses are not effective in determining consumer characteristics.

Ternary Revenue Change 2018	Revenue Change 2018	MSM Experience Score	Reselling tickets satisfaction score	STM Satisfaction Score	Gameday Experience Score	Price Paid Satisfaction Score	One year experience Score	STM Prestige Score	Being valued customer score	Gifts and Events Satisfaction Score	Recommendation Score
-1	($83,259.00)	8.15	5.62	7.55	7.55	6.24	6.09	6.35	6.82	5.51	7.18
0	$0.00	8.09	4.76	8.07	7.97	6.43	6.54	6.52	7.06	6.34	7.66
1	$50,242.00	8.05	5.11	8.16	8.02	6.98	6.68	7.05	7.25	6.88	7.88
Grand Total	**($33,017.00)**	**8.09**	**4.85**	**8.04**	**7.94**	**6.45**	**6.52**	**6.55**	**7.05**	**6.32**	**7.64**

Table 7.9: Customer Satisfaction scores based on Survey Data in 2018

Turnkey Demographics Results

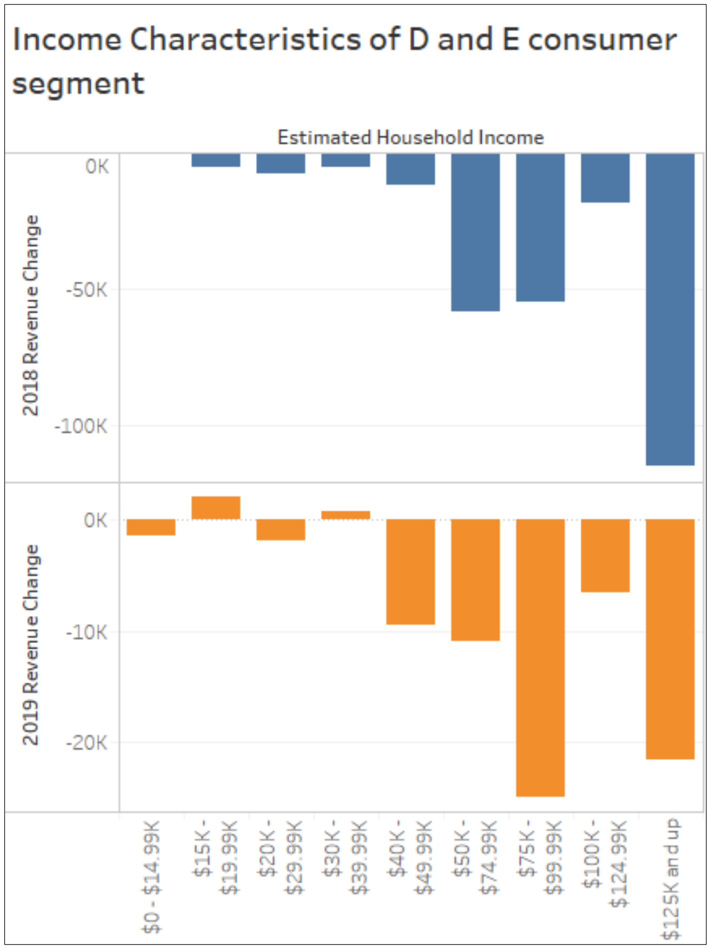

Figure 7.3: Household Income vs Revenue Change for 2018 and 2019 (D and E consumer segments)

The demographics data for D and E consumer segments in Figure 7.3 show that most consumers belong to the higher income segment and the renewal revenue for the NFL team has decreased from these income groups.

D and E consumer segment Age distribution

Age	
Age 22-31	35
Age 32-41	357
Age 42-51	699
Age 52-61	890
Age 62-71	852
Age 72-81	348
Age 82-91	105
Age 92 and above	24

Table 7.10: Age Group for D and E segments

The D and E consumer group comprises of people mostly in age group 42-71, that is, middle-aged and old.

Predictive Modelling Results

Model	Dataset	AUC	Sensitivity
Logistic Regression	Renewal 2018+ turnkey	0.8054	0.88
Logistic regression	Renewal 2019	0.78	0.74
Decision Tree	Renewal 2019	0.79	0.75
Random Forest	Renewal 2019	0.8047	0.83
Adaboost	Renewal 2019	0.83	0.802
Logistic Regression	Renewal 2018	0.66	0.77
Random Forest	Renewal 2018	0.61	0.75
Adaboost	Renewal 2018	0.67	0.76
Neural Network	Renewal 2018	0.501	0.67

Table 7.11: Model Results

In our predictive models, the renewal probability of consumers is best explained by Adaptive Boosting Model.

Below is a word cloud created for two different segments of customers, one for which there has been a decrease in the renewed amount and the other for which there has been an increase in the renewed amount in 2019. We see that there is a general positive sentiment for accounts where there has been an increase in the amount. On the other hand, there is a neutral sentiment for the accounts with the decreased renewal amount.

CONCLUSIONS

Predicting season ticket renewal by employing analytics on user data can help Sports teams to maximize their revenues. In this dataset, we used demographic information and other predictors such as number of times of contact instances, survey information, etc. to identify association of renewal behavior with season ticket holders. The results can be used by the Sports team to align their marketing strategy as per the needs of different segments of fans/customers. Our preliminary logistic regression model is able to identify renewal behavior with an accuracy of 77%. However, we intend to increase the accuracy of our model by exploring other sophisticated modeling techniques.

In our analysis, we have assumed that the behavior of the overall population of season ticket holders is similar to the behavior of approximately 10,000 season ticket holders present in the Renewal dataset. Also, there were only ~3400 user accounts that contained demographic information and we assume that their characteristics are representative of the entire population of the fans of the Sports team.

REFERENCES

Lee, J. &. (2003). A structural equation modeling for factors influencing attendance at professional sports events. International Journal of Applied Sports Sciences, 15(2), 28-39

Funk, D.C. & James, J.D. (2001). The Psychological Continuum Model (PCM). A conceptual framework for understanding an individual’s psychological connection to sport. Sport Management Review, 4, 119–150.

Kim, Y.K. & Trail, G. (2010). Constraints and motivators: A new model to explain sport consumer behavior. Journal of Sport Management, 24(2), 190- 210.

McDonald, H. & Stavros, C. (2007). A defection analysis of lapsed season ticket holders: A consumer and organizational study. Sport Marketing Quarterly, 16(4), 218-229.

Butler, M. (2007). Season seat holder retention in Minor League Baseball. Sport Management Undergraduate. Paper 93

Martin, L., & Goldman, M. M. (2015). A process model of sport fan detachment. African Journal for Physical, Health Education, Recreation and Dance, 21(3:2), 1030-1046

Healy, S. (2018). Analysis of Consumer Demand for American Sports Leagues. CMC Senior Theses. 1921

McDonald, Heath, Karg, Adam J and Leckie, Civilai (2014). Predicting which season ticket holders will renew and which will not. European sport management quarterly, vol. 14, no. 5, pp. 503-520

Schreyer, D., Schmidt, S., Torgler, B. (2018). Predicting season ticket holder loyalty using geographical information. Applied Economic Letters, vol 25, no. 4

Clinton, W. (2015). Industrial marketing in Sport: Understanding season ticket renewal across account type. International Journal of Sport Management Recreation & Tourism, Vol.20, p.1-19

Mark Brown, Darian Misko, Dennis Lee (2009). Membership retention in professional sports organizations, Vol 6, issue 1

Daniel C. Funk, Kevin Filo, Anthony A. Beaton, Mark Pritchard (2009). Measuring the Motives of Sport Event Attendance