Predicting Season Ticket Renewals for a Professional Sports Team

Griffin Herdegen, David Day, Kyle Betelak, Alejandro Isaac, Matthew A. Lanham

Purdue University Krannert School of Management

kherdeg@purdue.edu; day54@purdue.edu; kbetelak@purdue.edu; isaaca@purdue.edu; lanhamm@purdue.edu

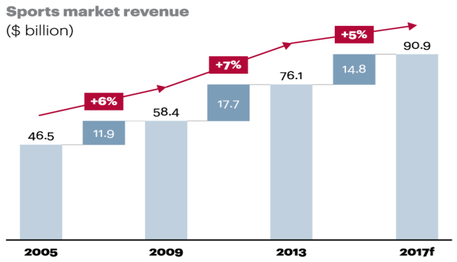
Abstract

In collaboration with a professional sports team, we developed a predictive model to identify why season ticket holders (STHs) renew year over year. The motivation for our research is that STHs are an essential revenue stream for sports teams. The interesting thing is that in sports, attendance and season ticket renewals tend to be a function of team performance and not necessarily business operational performance (e.g. ticketing, marketing, etc.). Our study is novel in that we identify demographic segments of fans and discuss actions that the team can take that lead to an increase in renewals.

Introduction

The sports industry is enormous and continues to grow. As such, industry leaders are determined to not only put out the best product but to discover which factors are most important when it comes to retaining fans, the industry's primary revenue stream. Season tickets are guaranteed seats in the stadium that will be paid for, and as such, are often a priority for these teams.

Figure 1
The sports market's next four-year cycle will bring continued growth



CAGR (%)

	2005–2009	2009–2013	2013–2017
Football	8%	9%	5%
U.S. sports	5%	5%	4%
Formula 1	3%	4%	4%
Tennis	2%	5%	3%
Golf	3%	2%	4%
Other	11%	9%	9%
Total	6%	7%	5%

Source: A.T. Kearney analysis

Research Question:

➢ How can we use machine learning to calculate which factors are most important to whether season ticket holders renew year-over-year?

Literature Review

Most past literature has been focused on various sub-industries (college athletics, Australian teams) that may or may not have the same drivers as American professional sports. Many of these studies have involved surveying season ticket holders regarding attendance behavior and willingness to renew.

Study	American sports	Professional teams	Survey	Recorded data	ANOVA	Regression	kNN	Decision tree	Demographic info
(Chen, 2009)	✓		✓						
(McDonald, 2010)		✓	✓			✓	✓		
(McDonald, 2013)		✓	✓	✓		✓			
(Warren, 2015)	✓	✓	✓		✓		✓		
Our study	✓	✓	✓	✓	✓	✓	✓	✓	✓

Figure 2. Literature comparison – sub-industry, data collection method, and methodologies used

Our study aims to be unique in that it takes a look at a broader array of potential factors influencing churn. Survey data is used, but in combination with recorded attendance data and demographic/financial information.

Methodology

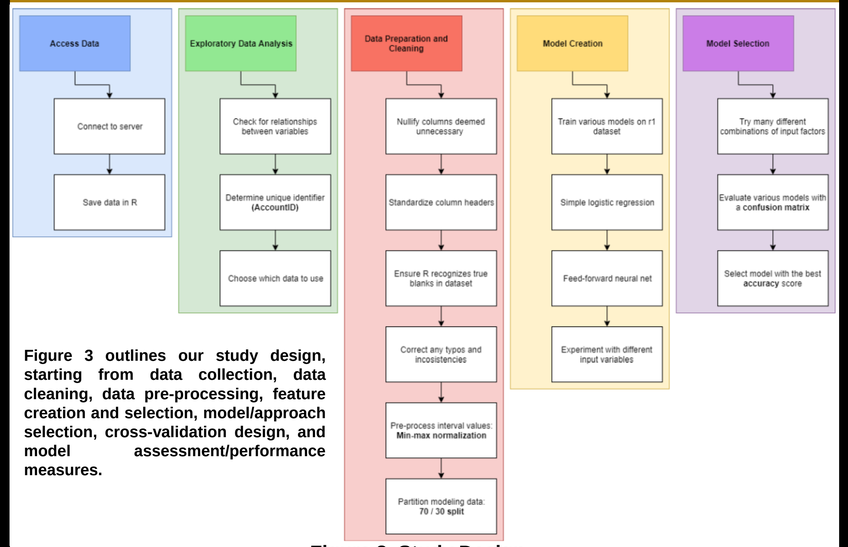


Figure 3 outlines our study design, starting from data collection, data cleaning, data pre-processing, feature creation and selection, model/approach selection, cross-validation design, and model assessment/performance measures.

Figure 3. Study Design

Data consisted of four datasets. However, we only found it appropriate to experiment with a single set – r1 (renewal information for the year 2018).

Data Cleaning & Pre-Processing

- Cleaned several typos in column heading
- Nullified data not used in predictive models
- Performed min/max normalization numeric features
- Made dummy variables

Feature Selection

We experimented with every feature included in the r1 dataset in order to determine which factors had the most influence on the variable.

Model Design

Partitioned the data into 70-30% train-test sets, used 3-fold cross-validation and trained various different types of models with our data.

Methodology Selection

Problem is a binary classification problem, and as such certain models are inappropriate for addressing the problem. Decided to try logistic regression and feed-forward neural networks..

Model Evaluation

Our models were evaluated using a confusion matrix and generating the model's accuracy. The model with the highest % accuracy is considered the most suitable model.

Results

After analyzing the various different models, we evaluated them based on their accuracy and ROC score. With an accuracy of 74% and an ROC of 78, we are confident that our model can guide management decisions to boost the sports organization's STHs identifiers, at an incredibly low cost.

Input	Estimate	Std. Error	Z-Value	P-Value
Intercept	-0.8139416	0.1909	-4.286	0.00001817
Number of Calls	0.1427293	0.03007	4.746	0.00000208
Number of Voicemails	0.9649775	0.03565	27.062	2.00E-16
Number of Emails	-0.9900365	0.04119	24.031	2.00E-16
Attendance Percentage	-0.008868	0.00218	-4.245	0.00002182
Distance from Stadium	0.0002899	0.0001438	2.016	0.0438

Figure 4. Statistically significant factors

Overall, the factors that we found to be the most significant and consistently correlated with a ticket holder's renewal habits were amount of contact points with representatives, attendance percentage, and distance from the stadium.

We experimented with two different models, which both performed similarly. Both our logistic regression and our neural network models had similar accuracy metrics. In our opinion, the logistic regression model is better suited for interpretative purposes, while the neural network may better serve the organization when it comes to predictive modeling.

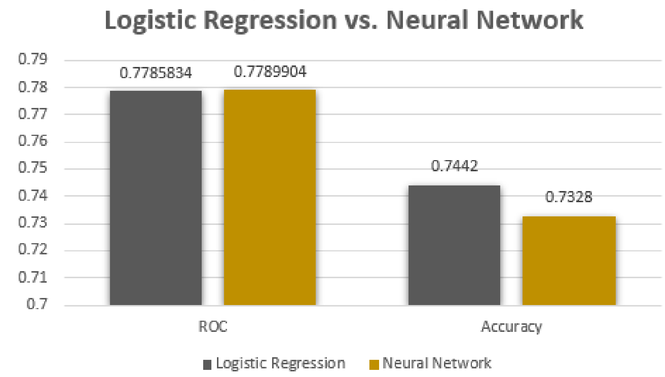


Figure 5. Model comparison

Conclusions

- There does not appear to be any standout demographic and/or financial indicators to whether a ticket holder is more likely to renew their tickets.
- Identify those ticket holders who are a sunk cost and highly unlikely to renew, and stop devoting resources to pursuing them.
- Target those uncontacted ticket holders with high attendance percentages, as those ticket holders are most likely to renew.
- Further studies include determining what factors are causing non-renewals (post-cancellation surveys, for example).

Acknowledgements

We thank Professor Matthew Lanham for guidance on this project.