



OUT OF THE BOX:

A Dynamic Pricing Model for Professional Sports Teams

Student researchers attempt to optimize tickets sold while maximizing revenue for an NFL partner organization.

ABSTRACT

We developed a dynamic ticket pricing model that will attempt to optimize tickets sold while maximizing revenue for our NFL partner organization. The motivation behind our research was the rapid adoption of dynamic ticket pricing by NFL organizations. In collaboration with an NFL franchise, we developed a model that takes in primary and secondary ticket market data from our NFL partner organization and identifies the highest probability settings for specific seats for maximizing revenue for specific events, all while minimizing the probability that a ticket goes unsold. By utilizing this model, organizations can refine their ticket pricing strategies to recapture sales that they have been losing to the secondary ticket market.

INTRODUCTION

As technology advances and data analytics proliferate, sports organizations must rethink their ticket pricing strategies. Organizations across professional sports have attempted to increase their revenue through ticket sales but have run into a major issue: their tickets are resold on secondary market at large premiums. These premiums introduced in the secondary market are uncaptured revenue for these organizations. Additionally, organizations also struggle with empty stadiums, which forces them to sell tickets below fan reservation prices. This clear swing in ticket value represents a stark reality: the value placed on a team's events is unique to each fan and fluctuates over the course of a season.

According to Grand View Research Inc., the total online event ticketing market was valued at \$46.6 billion in 2017 and is expected to reach \$68 billion by 2025 (Durgin, 2018). The worldwide secondary

ticket market is enormous and is set to continue growing. Given these changes, it's imperative for these organizations to find a way to extract the increasing value of tickets currently lost to the secondary market.

The solution to this problem is dynamic ticket pricing, the use of predictive and optimization models to dynamically adjust the prices of tickets to specific events to offset changes in supply and demand. According to *CIO* magazine, in 2009 the San Francisco Giants, a professional baseball team, piloted a dynamic pricing model on roughly 2.5% of the entire seating capacity (Overby, 2011). The model was a success and the organization ultimately increased its yearly revenue by \$500,000. The team went on to utilize the dynamic pricing system on the entire stadium, leading to a \$7 million dollar increase in ticket revenue.

In collaboration with an NFL organization we utilized data analytics to construct a dynamic ticket pricing model to increase the organization's revenue by filling out stadiums and competing with the secondary ticket market. We used internal data from the partner organization consisting of its own personal data as well as data about the secondary ticket markets. Using historical ticket sales records and the secondary market information, we used the model to determine the probability that a given ticket will sell on the primary ticket market. We then used these probabilities to maximize the expected revenue for the organization. We did so with an optimization model designed to take the predicted probabilities from our model to find the recommended price that would achieve the highest expected revenue and minimize unsold tickets across the event.

Through our research, we will attempt to answer the following questions:

Research Question 1: Can an accurate and interpretable predictive model be developed to determine how likely a ticket will be sold on the primary market?

Research Question 2: Can an optimization model be developed that maximizes the expected revenue of an event, given market conditions, while minimizing unsold tickets?

DATA

The final dataset used for this model was composed of five individual datasets:

1. Unsold inventory data
2. Primary ticket sale data
3. Secondary ticket sales
4. NFL opponent data
5. NFL fan and attendance data

The unsold, primary, and secondary datasets comprise data relevant to the sale of a ticket. Ticket sales were placed in their respective dataset based on their respective sales channel. These sets have 21 columns that served as the primary drivers in our model to both predict the probability of a ticket selling on the primary market and to optimize the price of that ticket. The ticket data spans the years 2012–2019. The NFL fan and attendance data comprises data found at Pro-Football-Reference.com. It contains relevant data on the characteristics of each team's fanbase as well as information on each team's attendance. The fan and attendance dataset spans the years 2015–2019.

METHODOLOGY

Our experimental model used a logistic regression model as a probabilistic classifier that then fed into an optimization model. A probabilistic classifier is designed to predict the probability that a certain event would occur. In our predictive model we used 10-fold cross-validation, with each fold containing 90% training data and 10% test data, to avoid overfitting our dataset. We also zeroed in on a specific section of the stadium where the partner organization most wanted to see results. Once the data was cut down to a single section it typically contained only a couple thousand observations, making it the perfect size to pilot our solution.

Our optimization model maximized the expected revenue resulting from Equation 1. The logistic regression predictive model supplied the probability that a ticket would be sold on the primary market by

factoring in specific traits associated with the ticket. Given this prediction, the optimization model found the price that maximized expected revenue on a given ticket and calculated the probability the ticket is sold. Both the price and probability interact with each other in this model, meaning that a higher price will typically correlate with a lower probability of being sold, and vice versa.

$$\text{Expected Revenue} = \sum \text{Price} * \text{Pr}(\text{Ticket Sold on Primary})$$

Equation 1. Revenue optimization equation.

With this expected revenue equation we relied heavily on the probability of a ticket being sold on the primary market to be accurate to correctly determine the price. The predicted probabilities need to be effective, so sensitivity and specificity were used as our performance metrics. Sensitivity measured our classifier's effectiveness in correctly predicting primary ticket sales for tickets sold on the primary market. Specificity was used to measure our model's effectiveness in correctly not classifying tickets as being sold on the primary market.

An illustration of these metrics is shown in Equation 2.

$$\text{Sensitivity} = \frac{\text{Number of Tickets Predicted as Sold in the Primary Market}}{\text{Actual Total Number of Tickets Sold in the Primary Market}}$$

$$\text{Specificity} = \frac{\text{Number of Tickets Predicted as Sold in the Non-Primary Market}}{\text{Actual Total Number of Tickets Sold in the Non-Primary Market}}$$

Equation 2. Sensitivity and specificity equations.

The methodology can be seen in Figure 1.

MODEL

Our optimization model required one predictive model, a probabilistic classifier. The probabilistic classification problem consisted of estimating the probability a ticket is sold on the primary market from variables in the ticket dataset as well. The primary candidate model was a logistic regression model that predicted the probability a ticket is sold on the primary market. The logic behind this selected predictive model is that the logistic regression model outputs well-defined and interpretable coefficients, or betas, that can be readily used for optimization. The betas of the model are the coefficients assigned to each predictive variable in the model. Therefore, by using logistic regression we obtained each coefficient for each predictor variable and used this to optimize the organization's expected revenue. Since the final optimization model was calculated using Excel Solver, clearly defined betas were necessary as con-

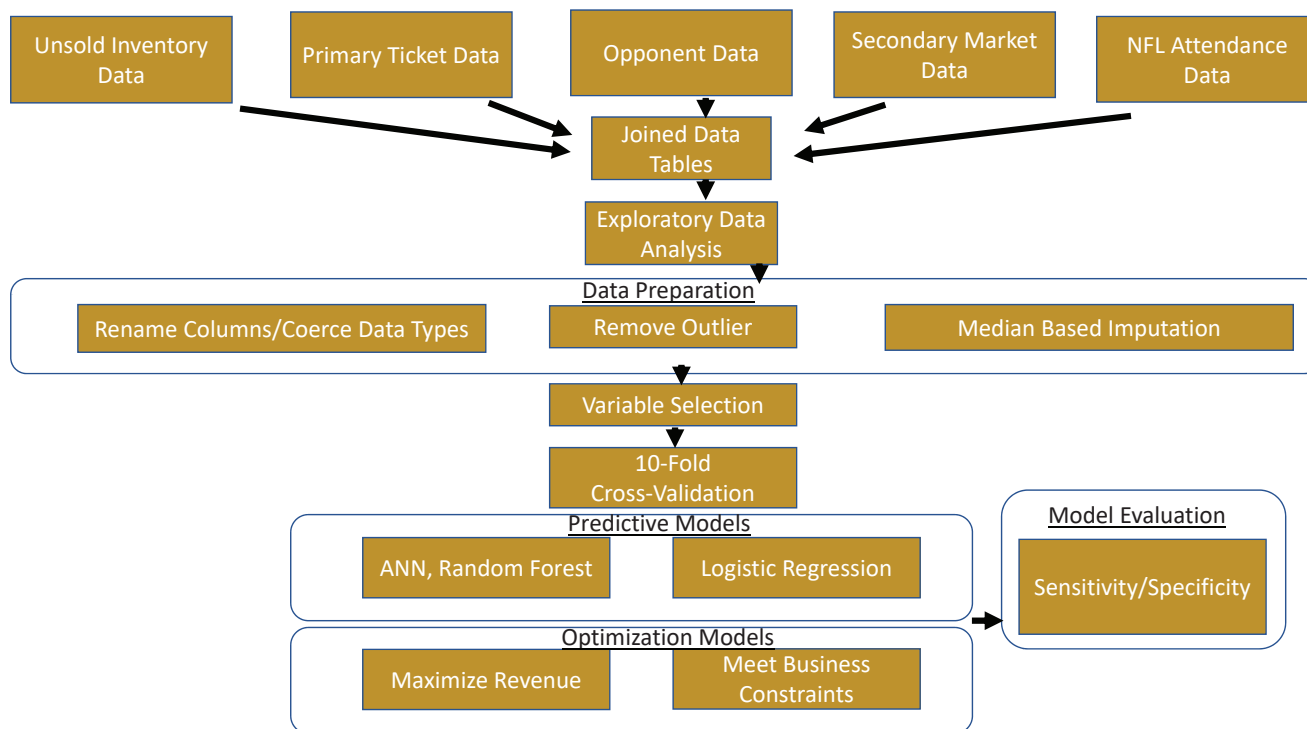


Figure 1. Methodological workshop.

starts to calculate the maximum expected revenue. Other probabilistic classification models tested were a boosted classification tree and a neural network; however, these were less interpretable.

Our final optimization model was set to maximize expected revenue for a game. The full optimization model can be seen in Equation 1 in the methodology section above. This model used a given price and multiplied it by the probability a ticket was sold on the primary market to calculate expected revenue. The probability is directly affected by the price at which a ticket is sold and thus was the decision variable for the optimization model. All other predictors stemming from the logistic regression model were held constant, although they do still contribute to the probability prediction of a ticket selling.

RESULTS

The initial logistic regression model included 61 variables. These initial predictors were selected using backward selection for logistic regression. To increase the effectiveness of the logistic regression model, the 6 predictors with the highest corresponding betas were selected to develop a new logistic regression model. The predictors and their beta can be viewed in Table 1. The final 6 predictors besides

final price were all created through feature engineering. WinRatio is an index of the opposing team's win to loss ratio; seatQuality is an index composed of information like seat row and section to score the quality of a ticket and its location within the stadium. The teamQuality predictor indexes information on the opposing team to give it a rating of quality, qualitySpending indexed information on how much an opposing team's fans typically spend. Home_out_contention is simply a true or false variable to indicate whether the home team is out of playoff contention at the time the game is scheduled to take place.

Running this model resulted in the probability calibration plot shown in Figure 2. As can be observed, the logistic regression model calibrated a bin of probabilities from around 8% to 12% well. There is also a point from around the 50% to 52% range that is relatively well calibrated. While this model has poor calibration for many bins of probability, it does show success for small intervals of probabilities. Two other metrics to represent the effectiveness of the logistic regression as a classifier in predicting whether a ticket is sold on the primary market are the specificity and sensitivity. Table 2 illustrates these performance metrics.

Predictor	Beta
winRatio	4.12397
seatQuality	-7.09938
teamQuality	-2.91929
qualitySpending	5.78219
Home_out_contention	15.68487
finalPrice	7.23588

Table 1. Final betas used for predictive modeling

	Training Set	Testing Set
Sensitivity	.9113	.8414
Specificity	.5689	.5698

Table 2. Logistic regression sensitivity and specificity

	Training Set	Testing Set
Sensitivity	1.000	.9512
Specificity	.9880	.9704

Table 3. Sensitivity and specificity values for boosted classification tree

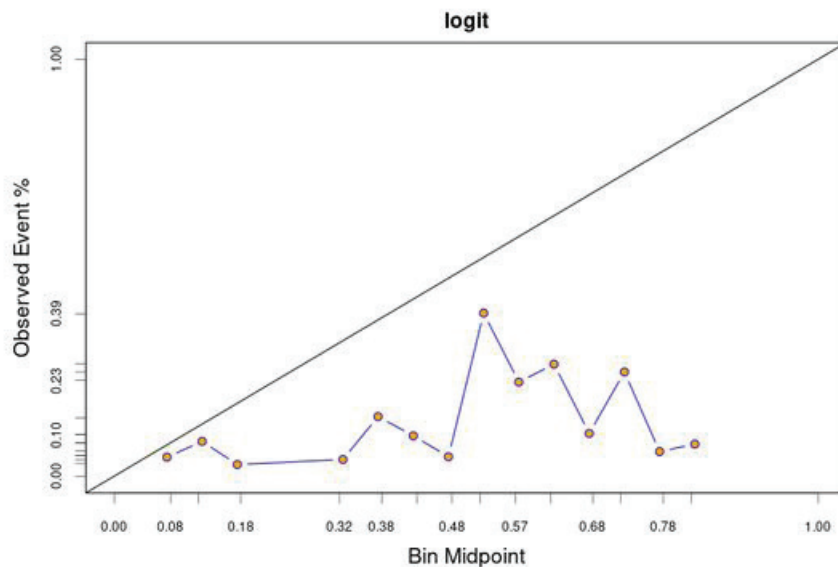


Figure 2. Logistic regression probability calibration plot.

The sensitivity and specificity results grant two important insights. First, the model does not appear to be overfit as there is not a drastic drop in sensitivity or specificity as the model moved from the training to the test set. Second, the logistic regression model is relatively efficient at predicting true positives in the dataset. As can be seen in Equation 2, the sensitivity is the proportion of true positives the model predicted to the total number of positives. On the test set, the logistic regression model can accurately predict 84% of all tickets sold on the primary market.

A boosted classification tree was run using the same 6 predictors from Table 1. Much like the logistic regression, for most parts the probability plot displays subpar calibration. However, there is good cali-

bration performance in the 2% to 5% probability bin and around the 58% to 60% probability bin. Also, the sensitivity and specificity results show that the model has not been overfit and suggest a high rate of successful classification of true positives (sensitivity) and a very high rate of predicted true negatives divided by overall true negatives (specificity). These performance metrics, seen in Table 3, suggest that the boosted classification tree is a superb classifier for predicting ticket selling probabilities.

The final probabilistic classifier trained was an artificial neural network. Figure 4 affirms that this classifier obtained the best probability calibration of all three classifiers trained. The plot indicates that the neural network is calibrated in the probability bins ranging from approximately 5% to 22% as well as

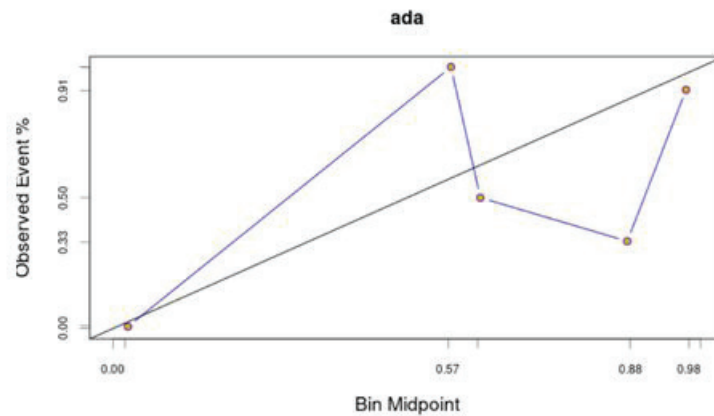


Figure 3. Boosted classification tree probability calibration plot.

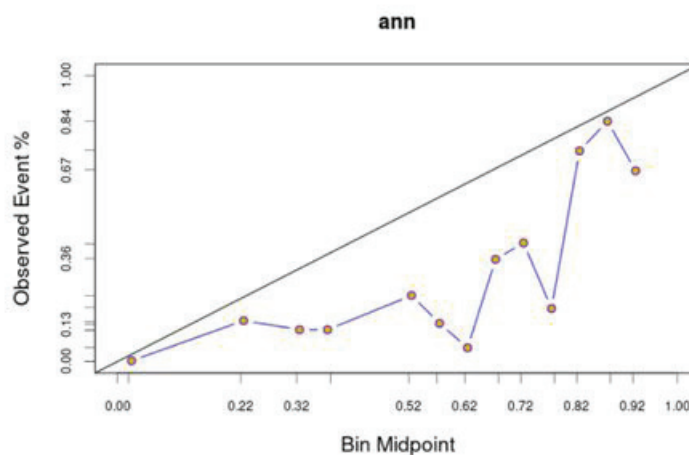


Figure 4. Artificial neural network probability calibration plot.

the 80% to approximately 87% probability bin. The sensitivity and specificity performance metrics both display similar results in sensitivity to the logistic regression and boosted classification tree model. The sensitivity in the testing set indicates very accurate prediction of the model. Sensitivity is predicting tickets sold on the primary market divided by the number of tickets sold on the primary market. Overall, the artificial neural network accurately predicts around 95% of true occurrences of primary tickets being sold. The sensitivity of the neural network also displays the model's ability to properly predict the tickets that are not sold on the primary market. Of the overall tickets not sold on the primary market, the neural network is able to predict approximately 82%.

While each model had good results over evaluation of the specified performance metrics, sensitivity and specificity, no model showed superior quality in prediction calibration. Each model had certain bins of probabilities that were calibrated much better than other bins. This draws the conclusion that a reliabil-

ity probability binning approach is an appropriate implementation for this dynamic pricing model.

A reliability probability binning approach uses certain models to capture different levels of probabilities when optimizing ticket prices. For the pricing model, the logistic regression, boosted classification tree, and artificial neural network each cover different bins of probabilities. Table 5 represents the different bins of probabilities for which each model had a good calibration. If the model has good calibration in a certain bin, then we classify it as the most reliable model to make predictions for that bin of probabilities.

Assigning each model to a probability prediction reliability bin ensures that the optimization model has better calibrated probabilities to gain more accurate insights on maximizing expected revenue. In an optimization model this is applied by having all three models determine the probability a ticket is sold on the primary market. For example, if the neural network model predicts the probability a ticket is sold

	Training Set	Testing Set
Sensitivity	.9839	.9512
Specificity	.8457	.8161

Table 4. Sensitivity and specificity values for artificial neural network.

	Logistic Regression	Boosted Classification Tree	Artificial Neural Network
Well-Calibrated Probability Bins	8%–12%, 50%–52%	2%–5%, 58%–60%	5%–22%, 80%–87%

Table 5. Binned probabilities for each model.

as 82%, then this prediction would be added to the optimization model because the neural network has proved reliable in the 80%–87% probability bin, and the expected value would be obtained through the optimization model shown in Equation 1.

We did this because the probability calibration plot (see Figure 4) shows that the neural network has a visually reliable calibration for the probability bin covering from 80% to 87%. However, if the neural network made a predicted probability of 62% that the ticket is sold on the primary market, we would not add this to the optimization model as expected revenue because the neural network is not well calibrated at that probability. Rather, that prediction would be delegated to the model with the reliability probability bin that covers that probability.

With the high interpretability of the logistic regression model, we optimized the expected revenue with the reliability bin from 8% to 12%, meaning that the price that maximizes expected revenue based on our logistic regression model must have a probability between 8% and 12%. For section 642 in the NFL stadium studied, the tickets observed as being sold in the primary market were 11.9% of all tickets sales for section 642 in the dataset. Using the median price for section 642 of \$80 per ticket, we calculated the expected revenue using the current static ticket pricing technique as \$9.52. This means under the current static ticket pricing, the team expects to capture \$9.52 from every ticket sale in section 642. Equation 3 illustrates the fundamentals of this calculation. The actual observed occurrences of tickets sold on the primary market takes the place of $\Pr(\text{Ticket Sold on Primary})$ given it is the empirical probability distribution for this primary market ticket distribution.

$$\text{Expected Ticket Revenue} = \text{Ticket Price} * \Pr(\text{Ticket Sold on Primary})$$

Equation 3. Expected revenue of single ticket equation.

Using Excel Solver as an optimization tool, we set up the optimization function with the logistic regression betas observed in Table 1 to calculate the probability of the ticket being sold on the primary market. Ticket price becomes the decision variable that interacts with the probability that a ticket is sold on the primary market given all else remains equal. The expected revenue is then maximized by the function observed in Equation 1. The optimal ticket price obtained from this maximization problem is \$121.58 at a probability of 12% being sold on the primary market. Equation 3 calculates the expected revenue using the optimized logistic regression model to be \$14.59. Equation 4 illustrates that this is a 53% jump in expected revenue for each ticket in section 642, moving from static ticket pricing to this dynamic ticket pricing model.

$$\text{Percent Change} = \frac{\text{Dynamic Ticket Expected Revenue} - \text{Static Ticket Expected Revenue}}{\text{Static Ticket Expected Revenue}}$$

Equation 4. Percent change in expected ticket revenue formula.

CONCLUSION

In conclusion, we managed to create a two-tiered predictive and optimization model that provides accurate and interpretable insights to make strategic pricing decisions. With our model, organizations can recoup and even hedge against ticket sales that are lost to the secondary market.

REFERENCES

- Durgin, T. (2018, August 21). Online ticketing to reach \$68 billion by 2025, primarily due to mobile growth. *ACCESS*. <https://access.intix.org/Article-Admin/online-ticketing-to-reach-68-billion-by-2025-primarily-due-to-mobile-growth>
- Overby, S. (2011, June 29). For San Francisco Giants, dynamic pricing software hits a home run. *CIO*. <https://www.cio.com/article/2406673/for-san-francisco-giants--dynamic-pricing-software-hits-a-home-run.html>
-

Cain, B., Saporoschetz, N., & Ginting, T. (2020). Out of the box: A dynamic pricing model for professional sports teams. *Journal of Purdue Undergraduate Research*, 10, 112–118.

Student Authors



with 3M as a supply chain analyst.

Brian Cain is a senior at Purdue University's Krannert School of Management, working toward a BS in supply chain, information, and analytics with a focus in business analytics. He has interned as an industrial engineer at UPS and has an upcoming internship



information systems. He began his career this past June as a supply chain management development associate for Nestlé.

Nikolai Saporoschetz graduated from Purdue University's Krannert School of Management this past May. He earned his BS in supply chain, information, and analytics, focusing in business analytics with a minor in business economics and a concentration in management



Purdue, pursuing an MS in business analytics and information management.

Theo Ginting graduated from Purdue University's Krannert School of Management with a BS in supply chain, information, and analytics with a minor in business economics this past May. He will continue his study at the graduate level this coming fall semester at

Mentor



and teaches Krannert's Data Mining, Predictive Analytics, Using R for Analytics, and Industry Practicum courses. These courses are highly experiential in nature and often interface with Purdue's Krenicki Center for Business Analytics & Machine Learning, where he serves as associate director for student engagement.

Matthew Lanham is a member of the Quantitative Methods faculty in Purdue University's Krannert School of Management. He serves as academic director for the MS in Business Analytics and Information Management program, where he coordinates