Using Machine Learning and StubHub Data to Estimate MLB Attendance and Ticket Pricing

Sumit Ilamkar, Erlan Konebayev, Braden Poe, Alex Toy & Kevin Van Lieshout[1]
University of Wisconsin Madison[2]

Dec. 12, 2019

Abstract

Stub Hub offers a great range of ticket data that can easily be utilized for analytical purposes. Within Major League Baseball, we have extracted every ticket sold/bought from 2007 to 2017 on Stub Hub with key characteristics to accurately predict attendances and magnify the best times to sell your tickets. From preliminary results, the best methods to predict prices have come through using neural nets and the best model to configuring optimal selling times is through k-means clustering.

Keywords: Forecasting and Predictive Modeling, Panel Data Models

JEL Classification Numbers: C53, C33

1 **Address: Department of Economics, University of Wisconsin Madison, 1180 Observatory Dr, Madison, WI 53706, email: ilamkar@wisc.edu, konebayev@wisc.edu, ajtoy@wisc.edu , bpoe@wisc.edu and kvanlieshout@wisc.edu**

2 **Special thanks to Lorenzo Magnolfi, Chris Sullivan and Dan Mcleod for teaching this course and giving advice on the project.**

I. Introduction

Stub Hub is one of the most reliable ticket traders in the world with thousands of exchanges every single day. They have changed the ability for fans to sell and buy tickets efficiently and affordably for both sides of the market. With the availability of data for users and companies, it opens up opportunities for ball clubs, managers, ticket offices and more to more accurately predict expected attendance and tickets sold from outside sources like StubHub in order to account and plan for seasons/games ahead. This dataset grants the opportunity to better understand what drives attendance, price and best time to sell and buy tickets. In this paper, we investigate which variables help accurately forecast game attendance and ticket prices. Additionally, our models will serve as a tool for sellers looking to optimize their sales strategy.

This dataset was initially collected and used by Andrew Sweeting[3] to study *"Dynamic Pricing Behavior in Perishable Goods Markets through the utilization of Major League Baseball tickets"*—i.e. price elasticity of ticket demand and supply as days-to-game fluctuates. His studies show that "sellers cut prices dramatically, by 40 percent or more, as an event approaches. The estimates also imply that dynamic pricing is valuable, raising the average seller's expected payoff by around 16 percent" (Sweeting 2012). The expectation of sellers and buyers is relevant to attendance and ticket prediction strategies as those expectations allow for behavioral aspects to be built into models. We take these acknowledgements and bridge them to the prediction of attendance and price while also developing a seller strategy.

Previous studies use a multi-modeling approach (Lim and Pedersen 2018) to find a relationship between ticket price, attendance, day of a game, local demographics. O'Hallarn et. Al. (2018) show twitter hashtags as a significant predictor of pricing in the secondary market using regression models, while Shapiro and Drayer (2014) show that fixed ticket price are on the low end of the price spectrum and secondary ticket prices are on the high end. According to Watanabe et. Al. (2013), the 2007 Stub-hub agreement has had a positive influence on price dispersion on ticket prices. In contrast to previous work, we uniquely try to determine an optimal seller's strategy using machine learning techniques.

In brief, a random forest was the best model for predicting attendance and prices of tickets based on a subset of optimal features. This machine learning was assisted by a marginal effects study as well. Secondly, we ran a K-means clustering analysis to determine the optimal time for sellers to upload tickets based on a variety of criteria. From that initial study in a K-means 5 clustering, the time frame of 47-50 days before the game yields the highest average ticket prices. These results will be thoroughly explained in the following parts of this paper.

In Section II, we explain characteristics of our dataset using a Milwaukee Brewers subset. Section III and Section IV contain our methodology and results, respectively. In these sections, we outline the machine learning methods we chose to use in our analysis and their overall performance. Finally, Section V concludes our paper by summarizing key findings, proposing new research possibilities, and discussing limitations to the project. All tables and figures not listed within the text can be found in the Appendix following Section V.

---

[3] **Andrew Sweeting, "Dynamic Pricing Behavior in Perishable Goods Markets: Evidence from Secondary Markets for Major League Baseball Tickets," Journal of Political Economy 120, no. 6 (December 2012): 1133-1172.**

II. Data

The database that we are using comes from the work of Andrew Sweeting which he compiled from Stub Hub. Our predictive analytics used the panel data version of the Stub Hub data which consisted of 2,124,235 observations of 162 variables. The year that this panel data was created on is 2007 and consists of the single game ticket sales for regular season Major League Baseball games. Each row represents a ticket sold and bought on Stub Hub during the 2007 season. Ticket price, original price and resell price are all in US Dollars. This dataset goes from 04-01-2007 to 09-30-2007, which is the entirety of the 2007 MLB regular season.

Our initial work done with the data was to clean the dataset by dropping numerous variables that were used only by Dr. Sweeting in his analysis and offered nothing for our analysis. In addition to this, we drop a number of indicator variables that apply to less than 1% of the dataset, and variables that display high levels of collinearity. We do these in order to ease the burden of computation, and to avoid potentially spurious results. Data specifically relevant to ticket prices included in the dataset such as standings information: home/away team games ahead in the division, home/away team games back in the division, wildcard standings for the home/away team, days until the game, original ticket prices, attendance of games, game of the season, and more, were used in our analysis.

First looking at some of the price variables included in this dataset. The tickets in this database had a regular price, the price listed by the original seller (i.e. MLB or an MLB Team) had a range of $5 at the lowest to $312 at the highest with an average of $39.95.
Sellers on average listed their tickets for $74.56 and buyers purchased tickets for an average price of $101.76. Looking at attendance metrics we see an average game attendance of 40,940. This varies by division and league, in the data we see 1,100,871 listings for American League home games, and 1,129,356 listings for National league games.

The AL home games had an average attendance of 41,522 with NL home games having an average attendance of 40,380. Looking at Divisions Table 1 outlines the average attendance by division.

Attendance by MLB Division

Division	Average Attendance
NL Central	**40,357**
NL West	**40,797**
NL East	**39,960**
AL Central	**34,916**
AL West	**36,485**
AL East	**47,166**

Table 1: Game attendance by division

Looking at a single team, we chose the Milwaukee Brewers a for which there were 88,774 observations in the overall Stub Hub dataset. In 2007 the Brewers finished the season with a record of 83-79, resulting in a 2nd place finish in the NL Central division two games

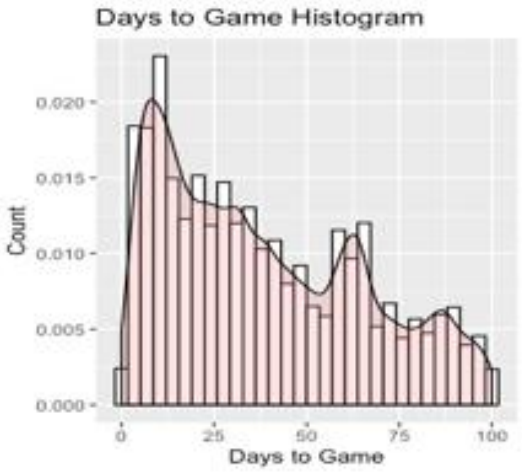
back from the Chicago Cubs. The Brewers are a unique team in the MLB as they are considered a small market team, ranking 27th out of 30 in market size[4] but they constantly draw high attendance numbers. In 2007 they drew an average attendance of 35,421, good enough for 12th in the MLB.[5]

When looking at the data for the Brewers tickets on average had an original price of $35.42 and games that they played in had an average attendance of 37,977 which accounts for home attendance and away game attendance. Tickets had transactions whether it be sales or price changes, on average 38.62 days before the game. With sellers on average listing a price of $63.54 and buyers purchasing on average tickets priced $87.37, meaning that lower priced tickets are likely to go unsold, this could be due to the fact that cheap tickets are often available directly from the team for a comparable price to those listed on Stub Hub.

The breakdown of Brewers' tickets listed was 20,084 for home games, or games taking place in Miller Park, and 68,690 for games in which the Brewers were the away team despite the Brewers playing an equal number of games home and away during the 2007 season. This discrepancy in tickets listed could be due to the park sizes, Miller Park, where the Brewers play has a capacity of 41,900 individuals, which ranks 16th among all MLB parks.

Looking deeper into the data there is a variable included in the dataset indicating whether or not a ticket was eventually sold, looking at this for the subset of the data for the Brewers we can see that 35,811 of the listings were sold eventually, which is less than half of all listed tickets. This may be due to when tickets are listed, below in Figure 1, we can see the distribution of the listings by days until the game. It looks as though there is a cluster of tickets listed between 10-20 days before the game, it could be that tickets listed this close to game day may not be listed early enough to be sold.

Figure 1: Number of ticket listings vs. days to game



One caveat to acknowledge when looking at attendance prediction is the possibility of omitted variables that may strongly impact attendance numbers.

[4] **https://bleacherreport.com/articles/961412-mlb-power-rankings-all-30-mlb-teams-by-market-size#slide3**

[5] **http://proxy.espn.com/mlb/attendance?year=2007**

An example of an omitted variable is weather[6] which can strongly impact attendance if the game is played in an outdoor venue but might have no impact if the game is played indoors, and due to this an unexpected drop in attendance could easily spike a prediction metric like MSE. Other examples of omitted variables might be ticket deals, such as student nights or veteran nights, or giveaways. Specifically, the Brewers for example in 2019 they held a Christian Yelich bobblehead giveaway which was the most attended home game of the season, despite it taking place only midway through the season, against a sub-par Pirates team. Additionally, the metric utilized to evaluate goodness of prediction could be heavily impacted by these effects, mean squared error (MSE) for example could blow up with an unexpected drop in attendance due to omitted variables.

III. Methodology

III.A - Attendance and Price Prediction

The goal for the first part of our project was to determine the most important factors that influenced attendance and ticket prices. For that purpose, we selected 57 and 56 covariates deemed to have the most predictive power, respectively—for easier interpretability, only linear terms were used. The summary statistics for attendance/ticket prices and their covariates can be seen in Appendix A for cases where "upto" equals to 0 or 1. "Upto" is a dummy variable that equals zero if a ticket was unsold and one if the ticket was sold. In many cases, the original ticket lister changed the ticket price multiple times, which generated multiple listings where "upto" equaled zero and a single listing where "upto" equaled one. We used 3 different estimation methods: cross-validated ridge and lasso, and random forests, with the parameter of interest being the normalized root MSE:

$$NRMSE = \frac{\sqrt{\sum_{n=1}^{N}(\hat{y_n} - y_n)^2/N}}{\sum_{n=1}^{N} y_n/N}$$

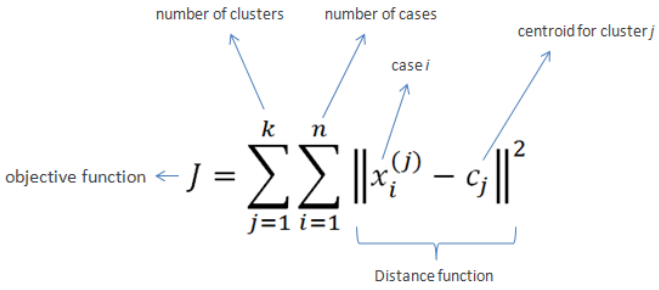
The models were trained and tuned on the training sample (about 70% of the dataset), and the quality of the fit was assessed based on test NRMSE, computed after using the final models on the test sample. The summary statistics can also be found in appendix A. We analyzed cases where all "upto" observations were included and also a subset where only "upto" observations equal to one were included. It was important to us that we differentiated between these two sets of tickets, since NRMSE's would likely be different. We will expound on notable differences in the results section.

III.B - Seller Strategy

In addition to predicting price and attendance, we also created a profit-maximizing seller strategy that focused on the 2007 Milwaukee Brewers. We approached this task with k-means clustering algorithms since this is considered an unsupervised learning problem where there is no "outcome" variable besides an optimal price. The variables that were important to place in the cluster were seller price, upto (an indicator for whether the variable was =1, sold, or =0, still on the site), the month of the game, how many home games were left,

[6] **https://www.forbes.com/sites/maurybrown/2018/04/15/weather-woes-taking-bite-out-of-mlb-attendance/#21a3fc2e445e**

how many games the home team was ahead, how many games the away team was back, the home record, the away record, day of the week, days the ticket was sold relative to the game and a variable for price inflation. By utilizing these variables, we uncovered many trends and recommendations for ticket sellers to follow. The methodology for a k means algorithm is expressed below[7]:



For a representative size, we decided to group the Milwaukee ticket data into 5 clusters, which provided enough variation without overclassifying the data. The k means clustering technique uses Hartigan and Wong (1979) as its algorithm placing tickets into their clusters. The values for each metric are the mean of the values in that cluster.

There is intuition in how we looked at this unsupervised problem. "Seller price", for obvious reasons, is a huge factor in determining clustering of tickets. "Month of the season" allows us to control for seasonality, since playoff/late-season games may carry higher ticket prices or vice versa for each season games. "Upto" is a metric of whether a ticket was sold on StubHub or left on the site unbought. For cheaper games in a week or early games in a season, we felt that these tickets are more likely to be left on StubHub unpurchased, since customers may feel the need to purchase expensive tickets from a secure site rather than purchase from scalpers. "Home games to go" indicates proximity to playoffs and "Home games ahead in a divisional race" allows us to capture team quality/dominance. By pairing those two variables with "Away games back", "day of week", and "home/away records" we gain an adequate proxy for ticket demand at any given time. The last variable we utilize is "price inflation", which measure of how much the price of a ticket has increased from its face value. We recognize correlation with seller price and inflation, but we feel that price inflation captures sellers who are attempting to capitalize on high market demand.

IV. Results

IV.A - Attendance and Price Prediction

The NRMSE for all three models we implemented can be seen in the tables on the following page. The table to the left displays the errors for all observations while the table on the right displays the error for the subset where "upto" equals 1.

[7] **Algorithm explanation provided by https://www.saedsayad.com/clustering_kmeans.htm**

Normalized Root MSE			
	Ridge	*Lasso*	*Random Forest*
Attendance	0.205	0.203	0.123
Price	0.762	0.760	0.601

Table 2: NRMSE for data where "UpTo" = 0 & 1

Normalized Root MSE			
	Ridge	*Lasso*	*Random Forest*
Attendance	0.077	0.072	0.035
Price	0.426	0.420	0.254

Table 3: NRMSE for data where "UpTo" = 1

From these results, it follows that random forests are the most accurate in terms of explaining the attendance and ticket prices. Judging from the magnitude of the NRMSE values, all three models produce accurate predictions of attendance and price, with attendance being slightly more accurate. Table 3 indicates that our algorithms performed especially well on the "upto" subset. We attribute the performance improvement to the decreased noise within the subset of tickets where "upto" equals 1. Despite differences in predictive power, both models are useful depending on which side of the market an individual stands on. Buyers will be more interested in the "upto" subset, since it only includes final ticket sale prices. Sellers, on the other hand, will benefit from viewing both analyses, since the unrestricted set captures market dynamics that occur between initial listing and final sale.

In addition to NRMSE, it is also helpful to understand how important certain covariates are in predicting both price and attendance.[8] The tables below show 10 most important covariates for both of the random forest models, which we highlight because they were the best performing.

Variable importance, ticket prices

Variable	Scaled importance (0 to 1)
Regular price	1.00
Number of section	0.168
Home games ahead	0.111
Row number	0.107
Days to go	0.081
Home record	0.075
Day of week	0.062
Number of seats	0.061
Away record	0.057

Table 4: Variable importance for ticket prices

Variable importance, attendance

Variable	Scaled importance (0 to 1)
Home wild card games back	1.0000000000
Regular price	0.9636695856
Home record	0.9530183904
Home games back	0.9236441664
Section number	0.6467728946
Home games to go	0.3311824091
Day of the week	0.3276340910
Away games to go	0.3031127083
Row number	0.3013642180
Away record	0.2376991367

Table 5: Variable importance for stadium attendance

[8] **We only include tables for the unrestricted "upto" dataset, because it mirrors the results from the restricted subset.**

Finally, we evaluate the marginal effects of those covariates to understand the relationship between the most influential variables and predictions. For ease of computation, we first ran random forests on same datasets with only the top 10 covariates as in the table above, picked 4 most important covariates, and computed the marginal effects for those 4 using a smaller subset of the test sample. The partial dependence plots in the Appendix describe those marginal effects, again for both random forest models.

In the case of attendance, it seems like it decreases with the number of home games and home wild card games back, and the higher regular ticket price and team home record tend to be associated with higher attendance. As for the ticket price, it obviously is very dependent on the regular ticket price, and both section number and home games ahead are associated with higher ticket price. The row number doesn't seem to have a very clear effect on price, but a slightly negative trend can still be observed.

IV.B - Seller Strategy

In the seller strategy analysis, our K-means algorithm grouped tickets into 5 different clusters for 82030 Milwaukee Brewers tickets. Below are the summary statistics of these clusters:

K Means Clustering Analysis

Cluster Number	1	2	3	4	5
Seller Price	$ 47.18	$ 122.41	$ 43.03	$ 42.06	$ 250.08
Month	5.83	5.59	6.86	3.85	5.38
Up To	0.50	0.35	0.39	0.40	0.50
Home Game to Go	96.73	103.76	72.21	147.75	110.25
Home Game Ahead	1.52	0.46	1.30	0.53	0.93
Away Game Back	1.88	0.56	2.10	0.57	1.14
Day of Week	3.11	3.41	3.12	2.85	3.44
Away Record	0.55	0.56	0.53	0.55	0.55
Home Record	0.49	0.47	0.49	0.50	0.47
Day to Game	69.26	49.03	20.24	34.65	47.81
Price Inflation	0.89	2.21	0.76	0.73	4.82

Note: These are mean values for the cluster

Table 6: K Means Clustering of Milwaukee Brewers ticket listings

Using the Brewers 2007 data, our focus is brought to cluster 5 because of the high ticket price, which is optimal for sellers. The cluster mean for month is 5.38, which places us roughly in the middle of the season. Initially, this month mean seems odd. Intuition says it should be higher, because tickets should be most in-demand—and therefore have the highest price—near end of the season when competition heats up. However, we think risk-averse buyers, anticipating this consumer demand spike, decide to buy their August and September tickets earlier in the summer to avoid the demand surge. The day-to-game mean tells us that cluster 5 tickets are precisely the ones that fall in August and September. Because so many risk-averse individuals think similarly, they effectively create the surge in demand that they were hoping to avoid, but purchase tickets anyways due to risk-averse tendencies. Another explanation relates to day-of-week, which is at its highest at 3.44. This maximum indicates that more Friday and Saturday tickets are being purchased, likely due to better summer weather. A variable controlling for weather could help us justify this assumption.

A logical criticism to the previous analysis is that sellers are simply posting tickets at high prices hoping to capitalize on risk-aversion, but buyers are not taking the bait. However, this criticism does not stack up because we observe "upto", which measures percentage of

listed tickets being sold, reaching a maximum of 50%. Cluster 1 also achieves the same "upto" percentage, but at a substantially lower price. Because of the similarity across other covariates, we conclude that cluster 1 represents low-value, late-season tickets (bleacher or upper deck seats), while cluster 5 represents high-value, late season tickets (suites and lower deck seats, or front row bleacher and upper deck seats). Cluster 2 is somewhat of an outlier, so we do not use it in construction of our seller strategy. The combination of a low sale percentage (upto = 35%) and high price inflation, indicates there is a group of sellers hoping to luckily make a large profit from tickets that truly belong in cluster 1.

Given the K-means cluster analysis and our interpretation of results, we advise that ticket sellers do their research in the offseason to determine how competitive the Brewers will be in the following year. Prior to the season, sellers should buy front-row tickets in the lower deck, upper deck, and suite-level for games in August and September. If sellers can purchase these tickets at face-value or near face-value, then they should list tickets on StubHub roughly 5 months into the season and at a price inflation at or around 400%. Maintaining a price slightly below the mean inflation level should maximize sale percentage and sale price, which will garner the highest profit for sellers. If the Brewers wind up being less competitive than expected, then we advise sellers to re-post tickets roughly a month away from the game. This strategy maximizes sellers' potential for gain, while minimizing potential for loss.

V. Conclusion

The Stub Hub data set provided us with rich data to apply machine learning algorithms and draw insights. Through the process, we developed procedures to predict attendance of games and ticket prices in 2007 while also generating a seller strategy in order to maximize profits of when users should sell their tickets.

The prediction of attendance was best suited to a random forest model which yielded the best accuracy measure in the set of models we ran. This was followed by the use of marginal effect studies determining the most important variables we see influencing attendance and prices. This was a type of analysis that gives good insight for what to look for and use in terms of accurate predictions.

In terms of the seller strategy, we utilized a K-means clustering exercise to find representative samples of tickets based on covariates that we believed best help distinguish the price of a ticket. There were very useful variables that we were able to apply in this to give insight as to when sellers should look to sell, or even perhaps when buyers should look to buy. An important variable to take away from this study is the "upto" variable, which is equal to 1 if the ticket was sold at that price, or 0 the ticket was not sold at that price and was re-listed at another price. We utilized "upto" in our study to control for re-listing, since there are many tickets that either get changed multiple times or left unsold in general for games. This control helps us figure out on average, how often these cluster of tickets get sold, changed or left up on Stub Hub without a buyer. The main takeaway was that, more expensive tickets typically have a higher sold rate since Stub Hub is a secure and reliable site to purchase an expensive ticket. In contrast, we feel that cheaper tickets are typically left unsold on Stub Hub more often because consumers are more confident in their ability to scalp a ticket or buy one at the stadium.

This project gave a lot of good insight on prediction analysis and StubHub strategy, and provides a foundation for analysis with richer data sets. The one limitation we would like to address comes in the forecasting of attendance. Most variables in the dataset are excellent for determining price but inadequate for determining attendance since characteristics on a ticket tell you about the ticket itself but not the nature of the game someone is attending. This gives less than ideal insight into predicting attendance for games, despite NRMSE signaling excellent models. Going forward, dummy variables that indicate whether a game was "special" like a student night or a free t shirt night would help us predict attendance more precisely.

References

Hartigan, J A, and M A Wong. 1979. “Algorithm AS 136: A K-Means Clustering Algorithm.” *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28 (1): 100–108. https://doi.org/10.2307/2346830.

Lim, Namhun, and Paul M Pedersen. 2018. “Examining Determinants of Sport Event Attendance: A Multilevel Analysis of a Major League Baseball Season.” *Journal of Global Sport Management*, December, 1–18. https://doi.org/10.1080/24704067.2018.1537675.

O’Hallarn, Brendan, Stephen L. Shapiro, and Ann Pegoraro. 2018. “Hashmoney: Exploring Twitter Hashtag Use as a Secondary Ticket Market Price Determinant.” *International Journal of Sport Management and Marketing* 18 (3): 199–219. https://doi.org/10.1504/IJSMM.2018.091754.

Shapiro, Stephen L., and Joris Drayer. 2014. “An Examination of Dynamic Ticket Pricing and Secondary Market Price Determinants in Major League Baseball.” *Sport Management Review* 17 (2): 145–59. https://doi.org/10.1016/j.smr.2013.05.002.

Sweeting, Andrew. 2012. “Dynamic Pricing Behavior in Perishable Goods Markets: Evidence from Secondary Markets for Major League Baseball Tickets.” *Journal of Political Economy* 120 (6): 1133–72. https://doi.org/10.1086/669254.

Watanabe, Nicholas, Brian Soebbing, and Pamela Wicker. 2013. “Examining the Impact of the StubHub Agreement on Price Dispersion in Major League Baseball.” *Sport Marketing Quarterly* 22 (3): 129.

VI. Appendix

Summary statistics: attendance, price, selected covariates

Statistic	N	Mean	St. Dev.	Min	Max
Attendance	2,043,502	40,971.630	9,431.075	8,201	56,438
Price	2,043,502	86.562	80.567	0	903
Hour	2,043,502	8.424	5.395	0	23
Dummy for ticket sale	2,043,502	0.370	0.483	0	1
Auction dummy	2,043,502	0.002	0.047	0	1
Piggy	2,043,502	0.0003	0.019	0	1
Aisle	2,043,502	0.029	0.167	0	1
Parking	2,043,502	0.008	0.088	0	1
Multirow	2,043,502	0.007	0.082	0	1
Face value	2,043,502	39.962	28.549	5	312
General admission	2,043,502	0.010	0.101	0	1
Month	2,043,502	5.452	1.579	1	9
Home team record	2,043,502	0.502	0.101	0.000	1.000
Home games back	2,043,502	4.725	5.065	0	29
Home games to go	2,043,502	107.590	37.902	2	162
Home wild card games back	2,043,502	3.758	4.274	0	26
Home games ahead	2,043,502	0.683	1.965	0	12
Home wild games ahead	2,043,502	0.068	0.403	0	6
Home prior to first game	2,043,502	0.128	0.334	0	1
Away team record	2,043,502	0.503	0.101	0.000	1.000
Away games back	2,043,502	4.928	5.506	0.000	29.000
Away games to go	2,043,502	107.261	37.993	2	162
Away wild card games back	2,043,502	4.053	4.755	0	26
Away games ahead	2,043,502	0.740	2.124	0	12
Away wild games ahead	2,043,502	0.071	0.401	0	6
Away prior to first game	2,043,502	0.128	0.334	0	1
Days to go	2,043,502	41.262	27.335	0	100
Price change (dummy)	2,043,502	0.832	0.374	0	1

Table 7a: Summary statistics for selected covariates used in price and attendance predictions.

Summary statistics: attendance, price, selected covariates (ctd.)

Statistic	N	Mean	St. Dev.	Min	Max
First row	2,043,502	0.110	0.313	0	1
Second row	2,043,502	0.094	0.292	0	1
Row number	2,043,502	8.086	6.706	0	26
Section number	2,043,502	5,085.607	4,921.214	1	24,228
No row listed	2,043,502	0.076	0.265	0	1
Month of game	2,043,502	6.854	1.587	4	9
Day of week	2,043,502	3.030	2.070	0	6

Table 7b: Summary statistics (cont.) for selected covariates used in price and attendance predictions.

Figure 2a: Partial Dependencies for the 4 most important covariates in predicting attendance when "UpTo" = 0 and 1.

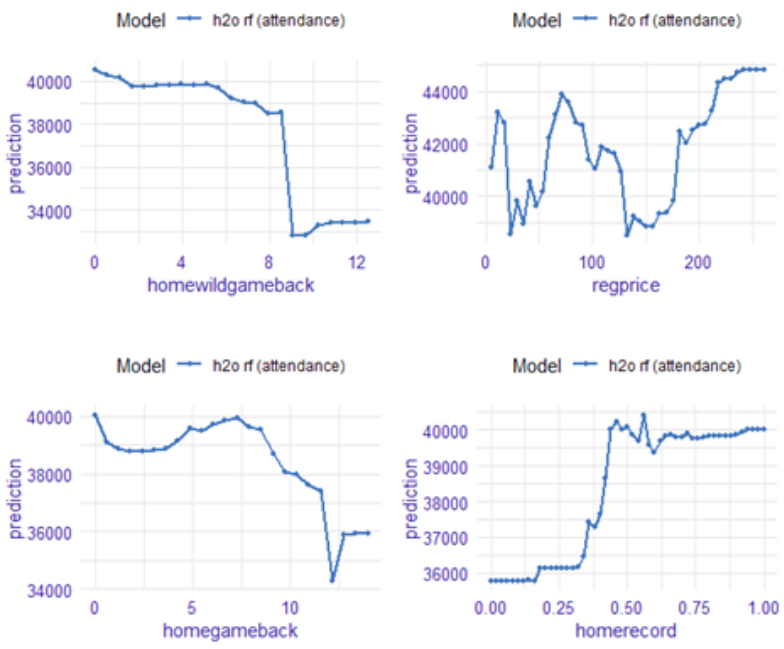


Figure 2b: Partial Dependencies for the 4 most important covariates in predicting price when "UpTo" = 0 and 1.

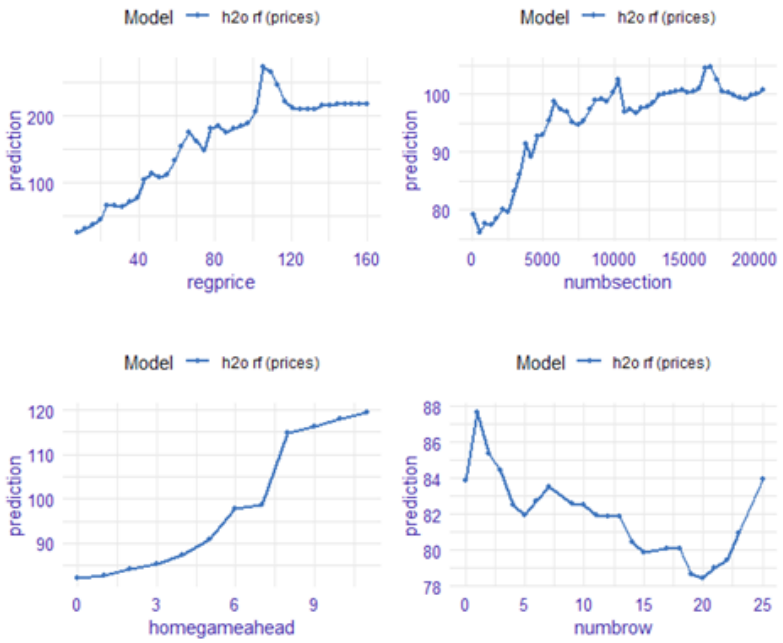


Figure 3a: Partial Dependencies for the 4 most important covariates in predicting attendance when "UpTo" = 1.

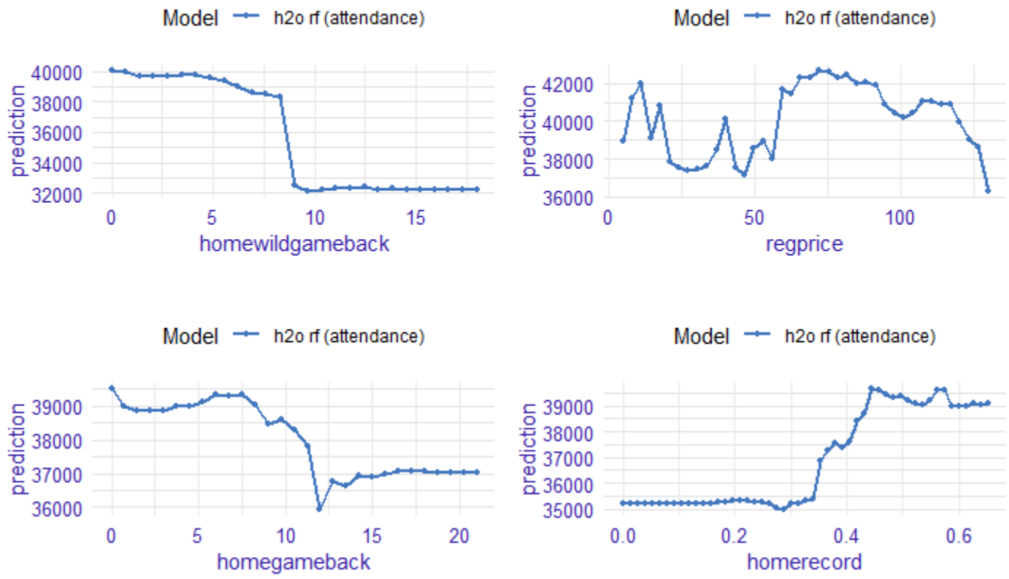


Figure 3b: Partial Dependencies for the 4 most important covariates in predicting attendance when "UpTo" = 1.