

Analyzing the Attendance Patterns of the 2019 Johnson City Cardinals

Ben Wachtel

SAL 213

Dr. Jeremy Losak

December 9, 2020

Introduction

For the Johnson City Cardinals, 2019 marked a successful year on and off the field. As the team finished 35-33, made the playoffs and eventually won the Appalachian League championship, they were also setting franchise attendance records. By welcoming over 80000 fans in 2019, the Johnson City Cardinals broke their franchise record of about 69000 fans that they set just a year prior. This paper will analyze what impacted their attendance and look into what led them to setting this new attendance mark. This paper will not only examine factors under the team's control, such as promotions, but will also investigate how factors like weather, opponent, and day of week factored into their 2019 attendance numbers. With the changing structure of the Appalachian League in 2021, it will become even more important to understand what drives attendance in Johnson City.

Summary of Data and Variables Used

| Summary Statistics for Quantitative Variables | | | | | | |
|---|------------|-------------------|-------------|----------|-------|--------------------------|
| | Attendance | Johnson City Win% | Temperature | Humidity | Wind | Total Precipitation (in) |
| Mean | 2540.121 | 52.169 | 74.336 | 76.948 | 2.985 | 0.149 |
| Standard Error | 147.638 | 2.487 | 0.426 | 1.05 | 0.232 | 0.064 |
| Median | 2751 | 54.05 | 74.3 | 78.7 | 2.7 | 0 |
| Standard Deviation | 848.114 | 14.288 | 2.448 | 6.031 | 1.335 | 0.367 |
| Minimum | 866 | 0 | 68 | 63.9 | 1.5 | 0 |
| Maximum | 4095 | 69.23 | 78.7 | 88.3 | 6.2 | 1.84 |
| Count | 33 | 33 | 33 | 33 | 33 | 33 |

In order to model attendance, many types of categorical and quantitative variables that may have an impact were gathered. In the table above, summary statistics for each quantitative variable are shown. Starting on the left with attendance, this variable is the response variable for this model, but it is still important to understand it in greater detail. The variable to the left of attendance is Johnson City's win percentage. The purpose of this variable is to measure the overall strength of the Johnson City Cardinals when the home game was played. For example, if

Johnson City were 4-4 heading into a home game, their win percentage would be logged as 50%. The next four variables are all weather based variables gathered from Weather Underground. The values entered for temperature, humidity, and wind were the average value from the day the game was played.

| Frequency Table for Categorical Variables | | Frequency Table of Opponents | |
|---|-------|------------------------------|-----------------|
| Variable | Count | Team | # of Home Games |
| Fireworks | 6 | Bluefield | 3 |
| Giveaway | 8 | Bristol | 6 |
| Food and Drink Discount | 17 | Burlington | 3 |
| Special Events | 10 | Danville | 3 |
| Sunday | 6 | Elizabethton | 4 |
| Monday | 5 | Greeneville | 6 |
| Tuesday | 3 | Kingsport | 3 |
| Wednesday | 6 | Princeton | 3 |
| Thursday | 3 | Pulaski | 2 |
| Friday | 5 | | |
| Saturday | 5 | | |

The frequency table on the left simply shows a count of how many times a certain promotion occurred and a count of how many home games were played on a certain day of the week. While most days of the week either had five or six home games, it was interesting to find that only three home games were played on a Tuesday or Thursday. Promotions were grouped into four separate categories, fireworks, giveaways, food and drink discounts, as well as special events. While fireworks and giveaways are self-explanatory, the other promotion categories consist of multiple types of slightly different promotions. For example, the food and drink discount category consists of four separate promotions, Dollar Mondays, Tacos & Tall Boys, All You Can Eat Wednesdays, and Thirsty Thursdays. Lastly, the special events category consists of one off promotions that did not appear multiple times. For example, this includes promotions such as a celebration of the 150th anniversary of Johnson City, Star Wars Night and Fan

Appreciation Night. It is also important to note that on certain nights, there were multiple promotions which is why the total amount of promotions is greater than the amount of home games. The frequency table on the right shows the amount of home games Johnson City had against each opponent. Notably, they only faced the Pulaski Yankees at home twice, while they had a minimum of three home games against every other team.

Model Overview

$$\begin{aligned}
 \textit{Attendance} = & \alpha + \beta_1\textit{Bristol} + \beta_2\textit{Burlington} + \beta_3\textit{Danville} + \beta_4\textit{Elizabethton} + \\
 & \beta_5\textit{Greenville} + \beta_6\textit{Kingsport} + \beta_7\textit{Princeton} + \beta_8\textit{Pulaski} + \beta_9\textit{JohnsonCityWin}\% + \\
 & \beta_{10}\textit{Temperature} + \beta_{11}\textit{Humidity} + \beta_{12}\textit{Wind} + \beta_{13}\textit{TotalPrecipitation} + \beta_{14}\textit{Monday} + \\
 & \beta_{15}\textit{Wednesday} + \beta_{16}\textit{Thursday} + \beta_{17}\textit{Friday} + \beta_{18}\textit{Saturday} + \beta_{19}\textit{Sunday} + \\
 & \beta_{20}\textit{Special Event} + \beta_{21}\textit{Giveaway} + \beta_{22}\textit{Fireworks} + \varepsilon
 \end{aligned}$$

In this model attendance was the response variable with 22 independent variables included to predict this value. The first eight variables represent the opponent that Johnson City faced, with the Bluefield Blue Jays being excluded. With this team excluded, it means that the coefficients for the remaining teams are relative to Bluefield. The next five variables are the quantitative variables that were summarized earlier. Next, are the days of the week when the game was played with Tuesday excluded. With this day excluded, it means that these six coefficients are relative to a game being played on Tuesday. Lastly, in regard to the promotional category, the promotion type “Food and Drink Discount” was excluded, meaning these coefficients are relative to a game where this type of discount occurred.

Model Results

| Attendance Regression Output Table | | | | |
|------------------------------------|----------|----------------|-------------|---------|
| Variable | Estimate | Standard Error | T-Statistic | P-Value |
| Intercept | -3382.94 | 3787.198 | -0.893 | 0.392 |
| Bristol | 1050.759 | 386.471 | 2.719 | 0.022 |
| Burlington | 1239.869 | 656.899 | 1.887 | 0.088 |
| Danville | 425.354 | 528.431 | 0.805 | 0.44 |
| Elizabethton | 1311.185 | 515.769 | 2.542 | 0.029 |
| Greeneville | 1340.18 | 391.302 | 3.425 | 0.006 |
| Kingsport | 2014.093 | 451.165 | 4.464 | 0.001 |
| Princeton | 1251.52 | 630.314 | 1.986 | 0.075 |
| Pulaski | 2196.158 | 2699.214 | 0.814 | 0.435 |
| Johnson City Win% | 5.789 | 37.385 | 0.155 | 0.88 |
| Temperature | 12.185 | 34.817 | 0.35 | 0.734 |
| Humidity | 37.011 | 36.477 | 1.108 | 0.333 |
| Wind | -132.261 | 133.28 | -0.992 | 0.344 |
| Total Precipitation | -767.366 | 293.188 | -2.617 | 0.026 |
| Monday | 885.842 | 323.914 | 2.735 | 0.021 |
| Wednesday | 565.556 | 504.279 | 1.122 | 0.288 |
| Thursday | 1489.264 | 413.582 | 3.601 | 0.005 |
| Friday | 596.291 | 929.543 | 0.641 | 0.536 |
| Saturday | 1089.553 | 685.547 | 1.589 | 0.146 |
| Sunday | 299.502 | 604.837 | 0.495 | 0.631 |
| Special Event | 572.613 | 271.845 | 2.106 | 0.061 |
| Giveaway | 561.363 | 267.436 | 2.099 | 0.062 |
| Fireworks | 1097.334 | 811.177 | 1.353 | 0.206 |

R-Squared = 94.75%

| Key |
|--------------------------------|
| Significant at $\alpha = 0.10$ |
| Significant at $\alpha = 0.05$ |
| Significant at $\alpha = 0.01$ |

This regression produced an R^2 value of 94.75%, which while encouraging it should be noted that this may be a product of the sheer amount of x-variables in this model. Furthermore, it is important how limited these results actually are. This dataset only contains 33 observations from one season, meaning that any statistically significant variables can only be applied to this dataset. Attempting to apply results from this model to data outside of the 2019 Johnson City Cardinal's season would lead to extrapolation and is an incorrect usage of this model.

Noting the limited scope of this data, we can still note from the output above that eleven of the variables included were significant at a minimum of one commonly used significance level. Of these significant variables, six were opponents that Johnson City faced throughout the year. It is important to note that these significant opponents are in terms of the Bluefield Blue Jays, meaning that when Johnson City played any of these six teams, they saw a statistically significant increase in attendance relative to a home game against Bluefield. The five remaining statistically significant variables included certain promotions, days of the week, and weather. Both the special event and giveaway promotion categories were significant, but similarly to opponents this is relative to the food and drink discount promotion. This means, relative to game with a food discount, when either of these promotions were held there was a statistically significant increase in attendance. Furthermore, a home game occurring on a Monday or Thursday saw a statistically significant increase in fans relative to a game on Tuesday. Lastly, the total precipitation variable proved to be the only weather variable that was significant. This variable has a more straightforward interpretation in which a one inch increase in precipitation led to a roughly 770 fan decrease in that games attendance.

There were certain variables that surprisingly did not have a statistically significant impact on attendance. For instance, fireworks are traditionally seen as a major draw for fans, but in this case, fireworks were not a statistically significant promotion relative to food and drink discounts. Furthermore, weekend days did not see a statistically significant change in attendance relative to a game on a Tuesday, which is surprising because weekend games are also seen as major draws for fans. However, more testing with a larger dataset should be conducted to see if these variables are not significant over a longer period of time.

Model Diagnostics

In order to ensure this model's effectiveness multiple diagnostic tests were conducted to see if there were concerns for multicollinearity, autocorrelation, and heteroscedasticity in this model.

| Attendance Model VIFs | |
|-----------------------|-------|
| Variable | VIF |
| Opponent | 1.852 |
| Johnson City Win% | 8.695 |
| Temperature | 1.387 |
| Humidity | 3.581 |
| Wind | 2.897 |
| Total Precipitation | 1.753 |
| Monday | 1.92 |
| Wednesday | 3.215 |
| Thursday | 1.965 |
| Friday | 5.509 |
| Saturday | 4.063 |
| Sunday | 3.856 |
| Special Event | 2.065 |
| Giveaway | 1.894 |
| Fireworks | 5.172 |

To check for multicollinearity, we can look at the variance inflation factors (VIFs) on each variable in this model. Looking at the table of VIFs for this attendance model, none of these variables met were greater than or equal to ten, which is a major indicator that there is collinearity. However, multiple values are greater than five which while less problematic than if it is greater than ten, is still a minor cause for concern. The variables fireworks, Friday, and Johnson City Win% all met this threshold, with Johnson City Win% having the highest VIF in the model at 8.695. One point to note is that VIFs were calculated for all opponents together, rather than individually, with a degrees of freedom of eight for that VIF measure.

| Durbin - Watson Test Results | | Breusch - Godfrey Test Results | |
|------------------------------|-------|--------------------------------|-------|
| DW Statistic | 2.094 | Chi Square Statistic | 5.491 |
| P-Value | 0.817 | P-Value | 0.064 |

Next, to check for autocorrelation, both a Durbin – Watson (DW) Test as well as a Breusch – Godfrey (BG) Test were conducted. Above, from the results of the DW Test we can see that the DW statistic was 2.094 with a p-value of 0.817. This tests' null hypothesis is that there is no autocorrelation in this model, so with this p-value we fail to reject this hypothesis. To help reinforce this belief that autocorrelation is not an issue in this model a BG Test was also conducted. In this test the null hypothesis is also that there is no autocorrelation. From the results above, the p-value of 0.064, while much smaller than in the previous test, is still not enough to reject this null hypothesis at the 0.05 significance level. These tests both helped show us that autocorrelation was not a concern for this model.

| Breusch - Pagan Test Results | | White's Test Results | |
|------------------------------|--------|----------------------|-------|
| BP Statistic | 23.393 | Test Statistic | 29.6 |
| Degrees of Freedom | 22 | P-Value | 0.953 |
| P-Value | 0.38 | | |

Lastly, to check for heteroscedasticity, both a Breusch – Pagan (BP) Test and a White's Test were conducted. In a BP – Test the null hypothesis is that there is no heteroscedasticity, more specifically, it is that our model is homoscedastic. From the results above, we can see that there was a BP statistic of roughly 23 and a p-value of 0.38, which is above the significance threshold of 0.05. This means we fail to reject this null hypothesis of no heteroscedasticity. With White's Test we are hoping to see the same results and reaffirm the belief that heteroscedasticity is not an issue for this model. This test has a null hypothesis that there is no heteroscedasticity in the model. From the results, we can see that the p-value was 0.953, which is above the

significance threshold of 0.05. So, from these tests, we can confirm that heteroscedasticity is not an issue in this attendance model.

Conclusion

Certain variables that would have been interesting to include were unavailable and therefore not part of this project. For instance, it would be very interesting to have demographic information regarding the fans who attended each game. This information, theoretically, would include age and gender information for the fans in attendance for each game. This data would be helpful in seeing how specific promotions affected attendance for specific groups. For example, it would allow us to see how a food discount impacted the amount of males below 25 attending that game. While this variable would be nice to have, it likely does not mean the current model has an endogeneity problem. The exclusion of this information means that we must model using attendance as a whole as the response variable and are unable to model how certain sects of the population respond to certain factors.

While our results do provide helpful insights into what specific factors impacted Johnson City's attendance, it is also important to reinforce that there is a limited sample size. This dataset only considered the 33 home games from the 2019 season. Further testing and modeling needs to be conducted with a larger amount of data to gain a clearer sense of what drives Johnson City's attendance. Adding more data would also allow us to see any historical trends and changes in how certain variables affect attendance. Lastly, with the Appalachian League converting from a short season minor league to a college wood bat league in 2021, it will be interesting to see what changes are made by the teams to boost attendance. Hopefully, by conducting these types of

analyses, teams like Johnson City will be able to smoothly transition into this new era of Appalachian League baseball and keep fans coming to the ballpark.

Sources

“Johnson City Cardinals Set Single-Season Club Attendance Record.” Ballpark Digest, 22 Aug.

2019, <https://ballparkdigest.com/2019/08/22/johnson-city-cardinals-set-single-season-club-attendance-record/>.

“Johnson City Schedule: Schedule.” *MiLB.com*, 2019, www.milb.com/johnson-city/schedule/2019/fullseason.

“Johnson City, TN Weather .” *Weather Underground*,
www.wunderground.com/weather/us/tn/johnson-city.