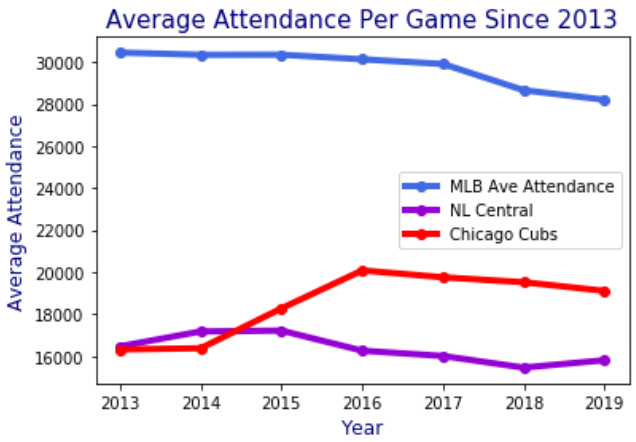
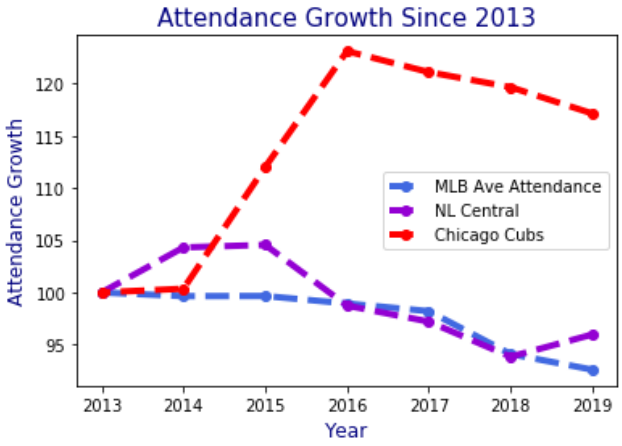
James Casey
MSDS 457 SEC 55
Week 7 Assignment: Revenue Analysis

Since 2007, Major League Baseball (MLB) has seen a steady decrease in the accumulative attendance for all teams. With a total attendance of 79.5 million in the 2007 season, the MLB has fallen below 70 million fans in the past couple seasons. This trend can also be seen in the average attendance per game for the National League Central Division. However, the trend is not as apparent in the Chicago Cubs' attendance due to a spike in 2016, the year they won the World Series. Below depicts the average attendance per game in the entire MLB, NL Central Division, and Chicago Cubs since 2013 (the year the Houston Astros left the Central Division) along with their respective growth, or decline, represented by indexed data:

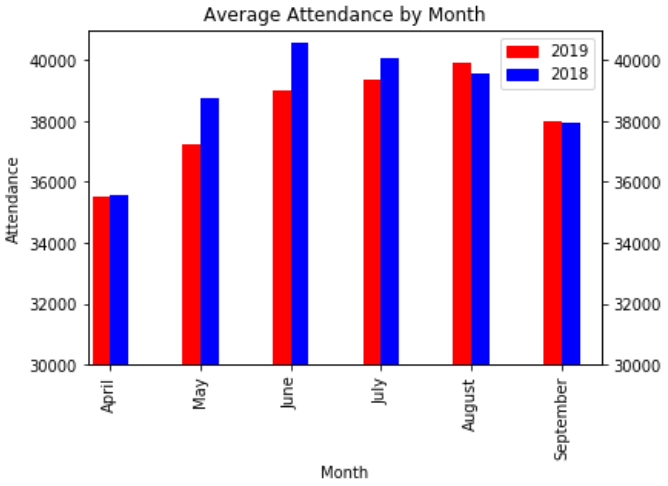


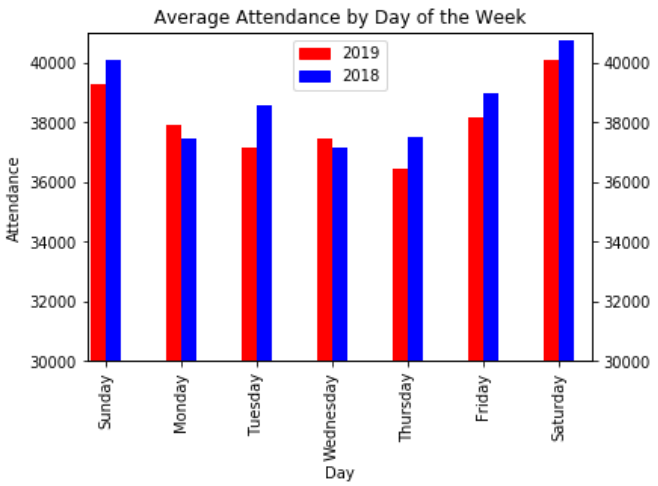


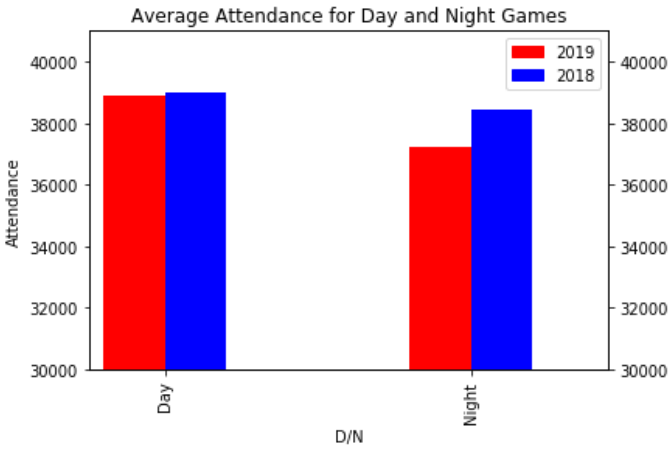
One can notice that MLB attendance is trending downwards. The Chicago Cubs saw growth from 2014 to 2016, but started to decline since then. Although attendance has declined, Moran notes in his article that the MLB is not actually losing fans. In fact, when looking into local ratings, national ratings, and digital properties, it suggests that the league is witnessing growth. Therefore, it will be worthwhile to look into factors that may affect attendance and recommend a dynamic pricing model that may help promote ticket sales.

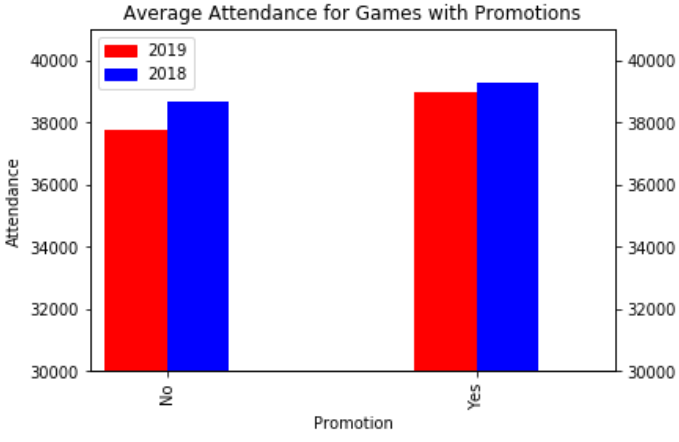
Attendance from the 2018 and 2019 home games of the Chicago Cubs will be analyzed with the key variables of interest being whether there was a **promotion**, the **month**, the **day of the week**, and whether it was a **day or night** game. Weather, scheduling, and giveaways all are important factors in determining whether a fan may attend a ball game. These variables may help reveal information on increasing ticket sales. Furthermore, these variables were selected since the data was readily available from BaseballReference.com and due to Thomas Miller's Los Angeles Dodgers example in *Modeling Techniques in Predictive Analytics with Python and R*. In his work, he details how he analyzed these factors and how they impacted the Dodgers' attendance. A similar method will be employed for the Chicago Cubs and their attendance.

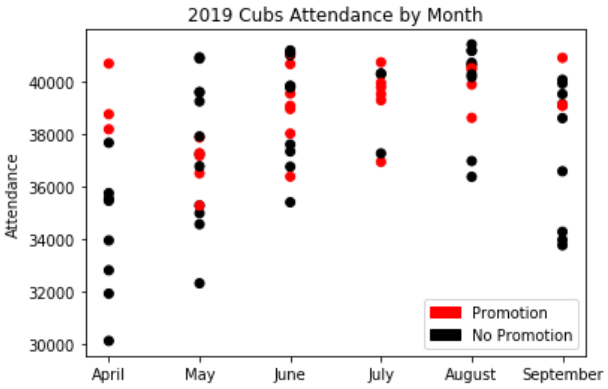
In order to determine the effects of the variables of interest on attendance of the Chicago Cubs in the 2018 and 2019 season, the following plots/charts were made:

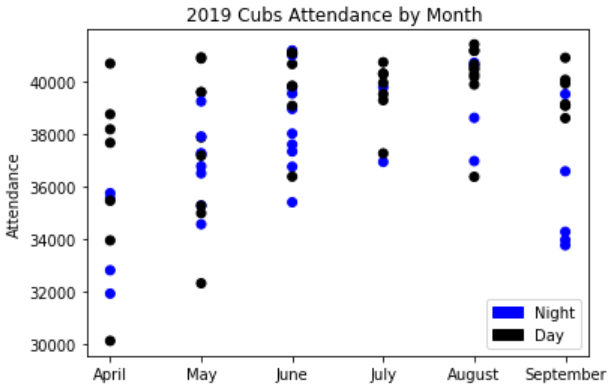


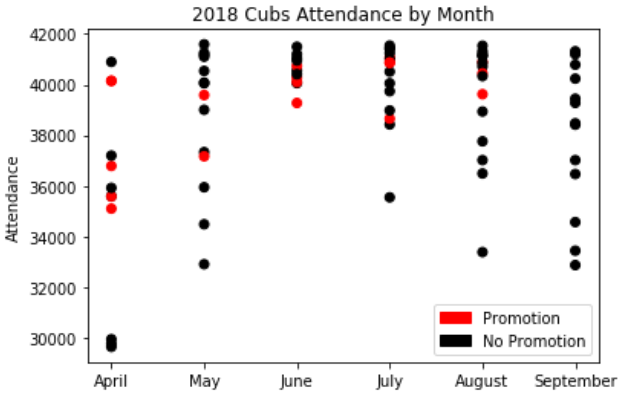


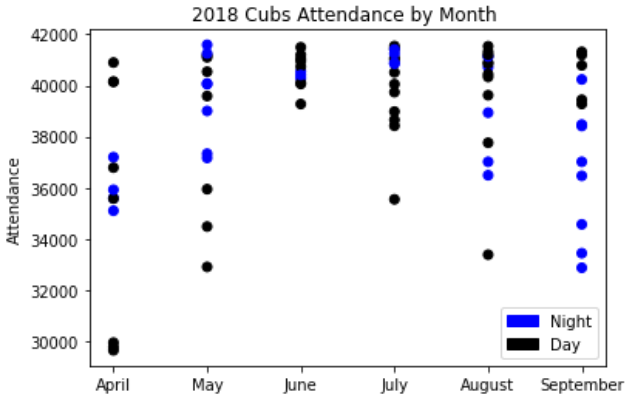












Year	Number of Promotions	Number of Day Games
2019	30	48
2018	18	50

There are some key takeaways that one can draw from these data visuals:

- **Average attendance** is **<u>down</u>** from 2018 to 2019.
- Attendance is **<u>highest</u>** in the months of **June, July, and August**. The warmer weather of the summer months may draw more fans to Cubs games. In 2019, the most popular month to attend a game was in August, as opposed to June, which had the highest average in 2018.
- On average, **<u>more fans</u>** attend games on **Saturdays and Sundays**.
- **Day games** attract **<u>more fans</u>** than night games. This is more apparent in 2019.
- **Promotions <u>increased</u>** the average amount of attending fans in both 2018 and 2019.
- In 2018 and 2019, most games with promotions tended to have a higher attendance. However, they do not necessarily have the highest attendances in each month. Particularly, in 2019, games with promotions appear to have a wider range of attendances.
- In 2018, night games tended to have the lowest attendances with the exception of May which had multiple games with high attendances.

- In 2019, night games were more inclined to have the lower crowds compared to 2018.

In summary, these plots suggest that games with promotions, during the day, on the weekend, and in summer months, all help lead to higher attendance numbers. These trends are found in both years' data. It appears that the Cubs organization tried to boost ticket sales by adding more promotions in 2019. However, the attendance still decreased from 2018. Perhaps, a dynamic pricing strategy based on predicted attendances can help increase ticket sales.

Therefore, linear regression models were employed to see if predictions could be made for game attendances with these factors being varied for each year. The linear regression models used each variable of interest along with the attendance of each game. The data sets for each year were split into two groups in which 95% of the data would be used for building a model, while the left over data points were used to test each model. The following chart displays the predictions of random data points with the actual attendances:

2018						
Month	**Day of the Week**	**Time of Day**	**Promotion**	**Actual Attendance**	**Predicted Attendance**	**Difference**
May	Saturday	Day	No	39598	37727	**1871**
April	Sunday	Day	Yes	38181	37754	**427**
June	Wednesday	Night	Yes	38017	37936	**81**
June	Wednesday	Night	No	35395	36893	**(1498)**
June	Tuesday	Night	No	36753	36812	**(59)**
			Average	**37588.8**	**37424.4**	**787.2**
			RMSE	**1089.42**		

2019						
Month	**Day of the Week**	**Time of Day**	**Promotion**	**Actual Attendance**	**Predicted Attendance**	**Difference**
July	Sunday	Day	Yes	39291	39365	**(74)**
July	Monday	Night	Yes	36935	38226	**(1291)**
August	Friday	Day	No	40276	39170	**1106**
May	Monday	Night	No	37909	35994	**1915**
August	Monday	Night	No	40721	37492	**3229**
			Average	**39026.4**	**38049.4**	**1903.75**
			RMSE	**1843.09**		

The model does a fairly good job in predicting the attendance of the games. In 2018, the average difference between games was about 787 fans, while the average difference was 1903 fans in 2019. This suggests that the linear regression may not be as well suited for the 2019 data since the predictions are not as accurate. A statistic that can assess this is Root Mean Square Error (RMSE). RMSE measure the average deviation from the regression line. A high RMSE indicates that the data is spread out, while a low RMSE would be observed if the data points were near the regression line. As seen in the RMSE values from the charts above, the attendance from 2019 was more volatile and deviated from the regression line more than the 2018 attendance. It appears that it may be getting more difficult to predict attendances, and new techniques may be needed to attract fans.

The Cubs organization's pricing model has multiple issues in its current state. First, the prices are never discounted below the original prices released at the beginning of the season. The prices will only increase as time passes, especially for games that are expected to sell well, as they want to encourage fans to buy tickets early in order to ensure seats are sold. Furthermore, in 2019, they implemented a system in which fans can purchase eight game ticket packages. Some of the packages include "The Rival Pack", "The Sunday Pack," "The Promo Game Pack," "The Summer Pack," and "The Friday Night Pack." Fans have expressed dismay from this system due to misleading or unclear costs for the tickets.

There are two apparent issues with these policies. First, by not allowing discounts, it will be harder to attract sales for less desirable games. Next, the ticket packages are a good idea that is executed poorly. The packages that are offered are for games that already attract fans. Instead they should try to focus on selling more tickets for:

- Games in April, May, and November.
- Games that take place at night.

- Weekday games.

In order to do this, some adjustments to their pricing strategy can be made. This includes:

- Packaging less desirable games with games that are predicted to perform well in terms of sales/attendance.
- Running more promotions on non-summer games, night games, and weekday games to hopefully boost attendance on those games.
- Allow the price of tickets to fall below original price for games that are predicted to perform poorly. This discount could potentially encourage day-of sales.
- In packaging plans, allow for ticket exchanges for fans who cannot attend games due to unforeseen circumstances. Ticket can be swapped for available lesser valued games. This will allow fans to feel a sense of security, while the Cubs can ensure attendance on those games. This has worked well for the Minnesota Twins organization in the past.

In conclusion, although the Cubs are outperforming the attendance rates for the average team in the NL Central, there is still room for improvement. The Cubs need to focus on the games that are predicted to underperform by developing a pricing policy that attracts fans to these games.

References

Bauernfeind, John. “Ticketing Strategies That Work.” *Sports Business Daily*, 9 Jan. 2017, www.sportsbusinessdaily.com/Journal/Issues/2017/01/09/In-Depth/Ticketing-strategy.aspx.

Lacques, Gabe. “Baseball's Future: Declining Attendance – and Shrinking Stadiums to Match.” *USA Today*, Gannett Satellite Information Network, 8 Aug. 2019, www.usatoday.com/story/sports/mlb/2019/08/08/mlb-attendance-stadiums-future/1941614001/.

Miller, Thomas W. *Modeling Techniques in Predictive Analytics with Python and R: a Guide to Data Science*. Pearson Education, 2015.

Moran, Eddie. “Another Year Of Declining Attendance: How Worried Should MLB Be?” *Front Office Sports*, 7 Aug. 2020, frontofficesports.com/mlb-attendance-2019-2/.

Muret, Don. “For Cubs, Dynamic Pricing's a One-Way Street.” *Sports Business Daily*, 18 Mar. 2013, www.sportsbusinessdaily.com/Journal/Issues/2013/03/18/Facilities/Cubs.aspx.

Price, Satchel. “How Cubs' Bizarre Ticket-Pricing System Turns $128 Package into $504 at Checkout.” *Times*, Chicago Sun-Times, 10 Mar. 2019, chicago.suntimes.com/2019/3/9/18392472/how-cubs-bizarre-ticket-pricing-system-turns-128-package-into-504-at-checkout.

Reddy, Durgin. “How Dynamic Pricing Is Changing Sports Ticketing.” *ACCESS*, ACCESS, 10 Dec. 2018, access.intix.org/Full-Article/how-dynamic-pricing-is-changing-sports-ticketing.