**Suyash Shrivastava – G22901161 – Sports Analytics DNSC 6327 – Assignment 1**

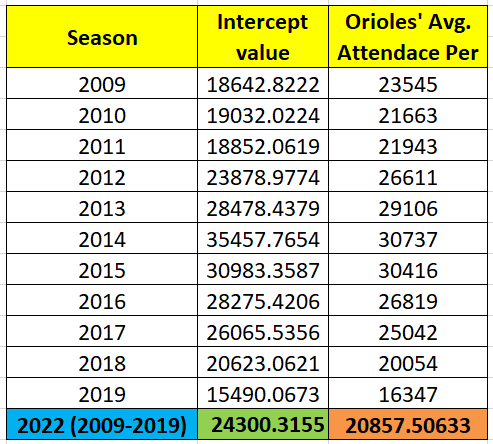
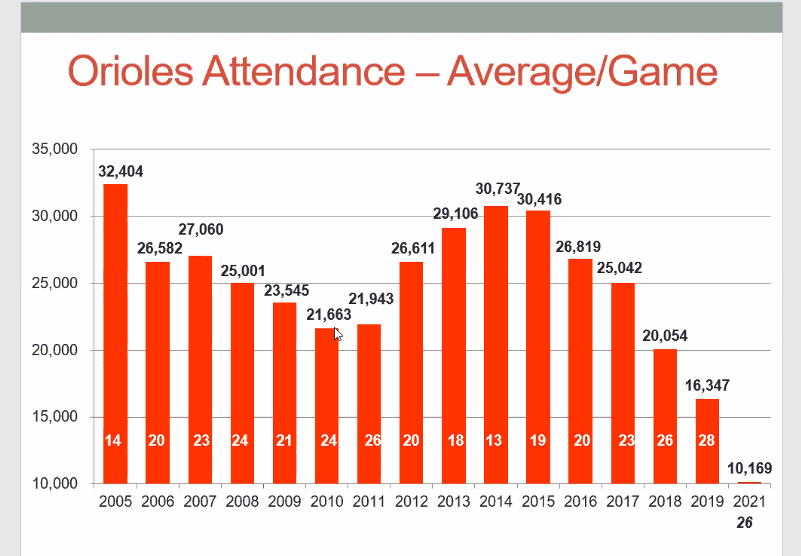
**SUMMARY**

After running the models and calculating attendance per game, the average attendance for Orioles’ season according to my forecast will be **20857.50**. As per the assignment, we can omit earliest seasons in the data and therefore, I omitted seasons from 2001 to 2008 and ran regression models on seasons 2009-2019 individually as well as combined.

I ran regression analysis on Excel for seasons from 2009-2019 (individual season and combined season data from 2009-2019) using independent variables (columns in the original dataset) as Bobblehead, Holiday/ Memorial Day, Opening Day, NYY, BOS, WSH, NIGHT, SAT, Makeup Game, Jun, Interleague since they are in binary.

I ran another regression analysis in R for the same season range (2009-2019) but here, I used independent variables (columns in the original dataset) like Opponent, Bobblehead, Day, Promo/Giveaway Item/Rain-Make-up Game Note to do my analysis.

I compared each season’s (2009-2019) individual regression model’s ‘Intercept Coefficient’ in Excel (using the same independent variables as described above for Excel) with the average Orioles Attendance per game from 2009-2019 and the numbers seemed close.



Therefore, I took the base attendance number as **24300.** All the values of the regression analysis which I did on R including intercept, R square value, p value and all the other values are stored in the file ‘lm\_2009to2019.txt’. The code is included with all the files. While doing the regression in R, I tried several combinations of various independent variables and rejected some of them completely because the data (estimate) didn’t seem correct. Sometimes the estimate was greater than the ballpark capacity upon running the model in R with some independent variables. I excluded the independent variables (columns in the excel dataset) like ‘Holiday (Memorial Day/4th of July)’, ‘Weather’ (because it’s hard to predict weather), ‘Month’, ‘Opening Day’, ‘Night’, ‘SAT’ and a few others.

Now, referencing the values from this file, I got all the ‘Estimate’ values for all the opponents, promos, days and promos. Since there are not many games happening on Friday in the given dataset, I have estimated the addition in attendance on Friday as 1500 (more than Thursday’s estimate and less than Saturday’s estimate).

I analyzed the dataset for all the seasons individually and found that on occasions like ‘Opening day’, 25th anniversary, 50th anniversary, Memorial day, Father’s day, Mother’s day etc. the attendance has always been north of 40,000. Similarly, on occasions like field days, fireworks, Hawaii t-shirt giveaway, student night etc, the attendance has been north of 30,000-35,00 (leaning towards 35,000 or more on weekends). I have mentioned my estimates on the sheet titled ‘*Est. Attndc on Promo days*’. For the initial days, I have not taken the estimate values of the opponents from the regression results from R assuming fans will come in numbers at the starting of the season.

Therefore, to summarize my calculations for each game:

* *Row 34: Game 31 – 4th June – Saturday – CLE –Promos (Fireworks):*

Attendance = **35,000** (average attendance whenever there are Fireworks in previous seasons, from sheet ‘*Est. Attndc on Promo days*’)

* *Row 55: Game 52 – 27th July – Wednesday – TB –Promos (None):*

Base Attendance (**24300** from Excel regression of combined data from 2009-2019) + Wednesday **(‘-2011’**: fromlm\_2009to2019.txt - ‘dayWed’ from R) + TB (‘**-14578’** : fromlm\_2009to2019.txt - ‘opponentTB’ from R) = **7711**

Rest of the calculations for each game is done accordingly. Therefore, summarizing the results, we will see an increase in attendance from the previous two seasons. A suggestion to the Orioles to increase their attendance will be to increase their promo events and offer some more in game entertainment in the ball park.

The exhibits and screenshots from Tableau are referenced and explained below. The forecasted data for the 2022 season was added in a new sheet in the excel (2001-2019 (2)) which contains the initial dataset as well.

**Screenshots/ exhibits from Tableau**

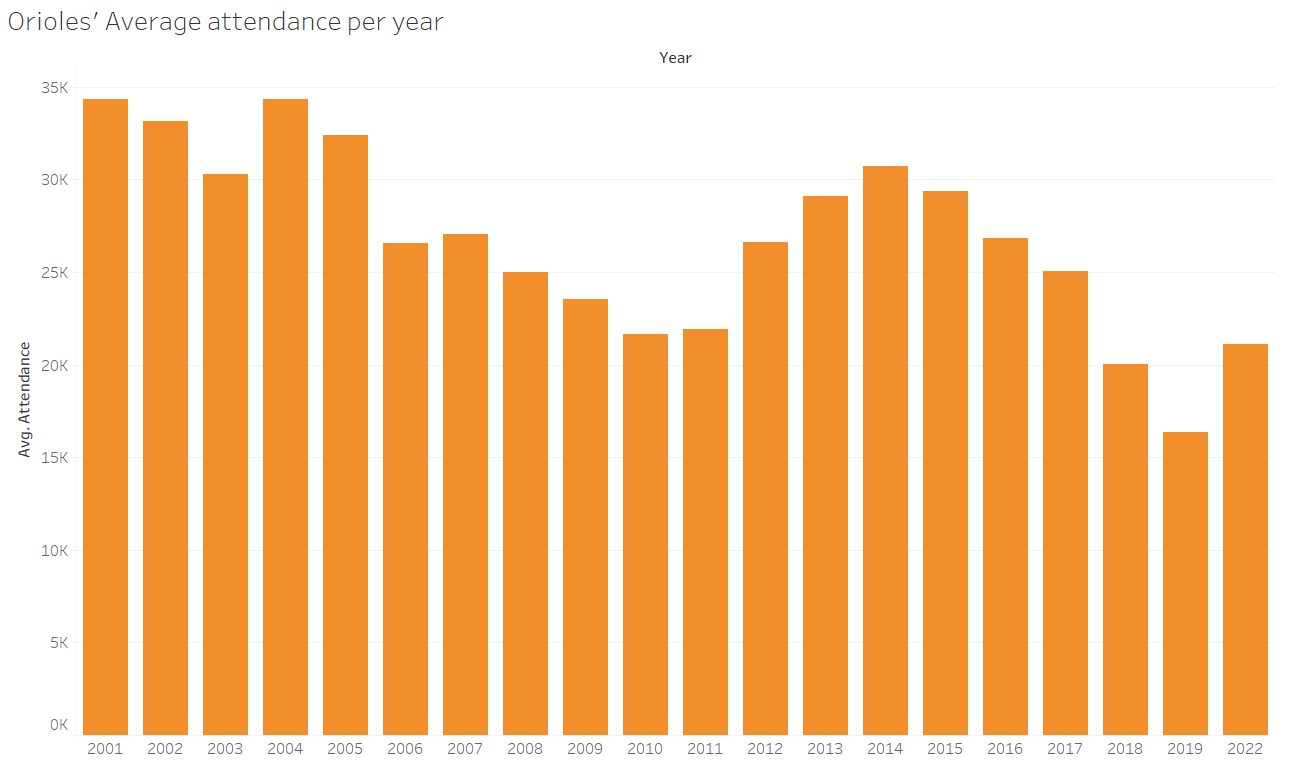


Exhibit 1: Average attendance per year (from Sheet 1 in the Tableau workbook)

In this particular exhibit we can see the average attendance per year at Oriole’s games. At the very end of graph on the right, we can see the forecasted attendance for the year 2022 from the analysis done in Part A1 and A2. The forecasted average attendance is more than the last 2 years (2018 and 2019). There is an increase of 29.27% from 2019’s average attendance and 5.38% increase from 2018.

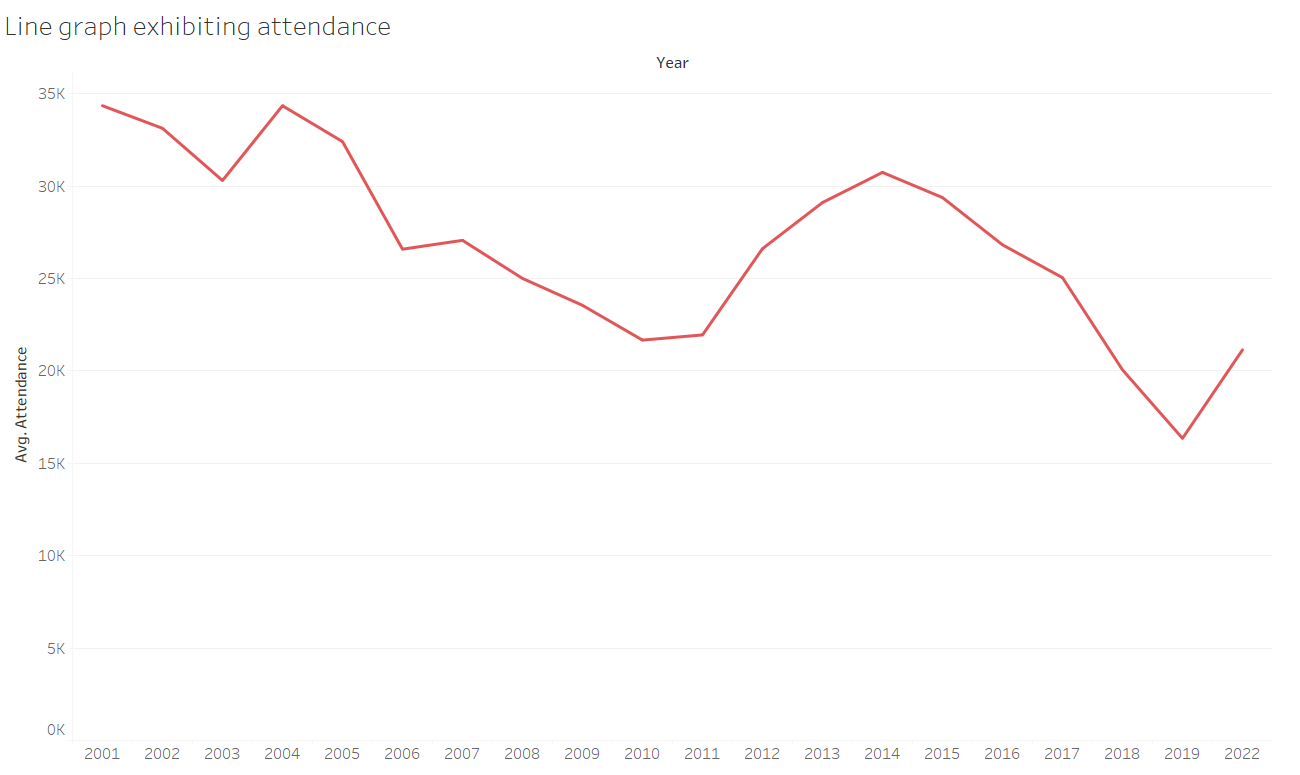


Exhibit 2: Line graph showing average attendance per year (from Sheet 2 in the Tableau workbook)

Exhibit 2 shows us a line graph showing average attendance per year. From the line graph we can observe that the attendance was on a slump from 2004 till 2011. After 2011, we can see that there is an increase in the attendance till 2015 and then there is a sharp dip till 2019. Since our forecasted data for year 2022 predicts an average attendance of 20857, we can see that the graph tends to head up.

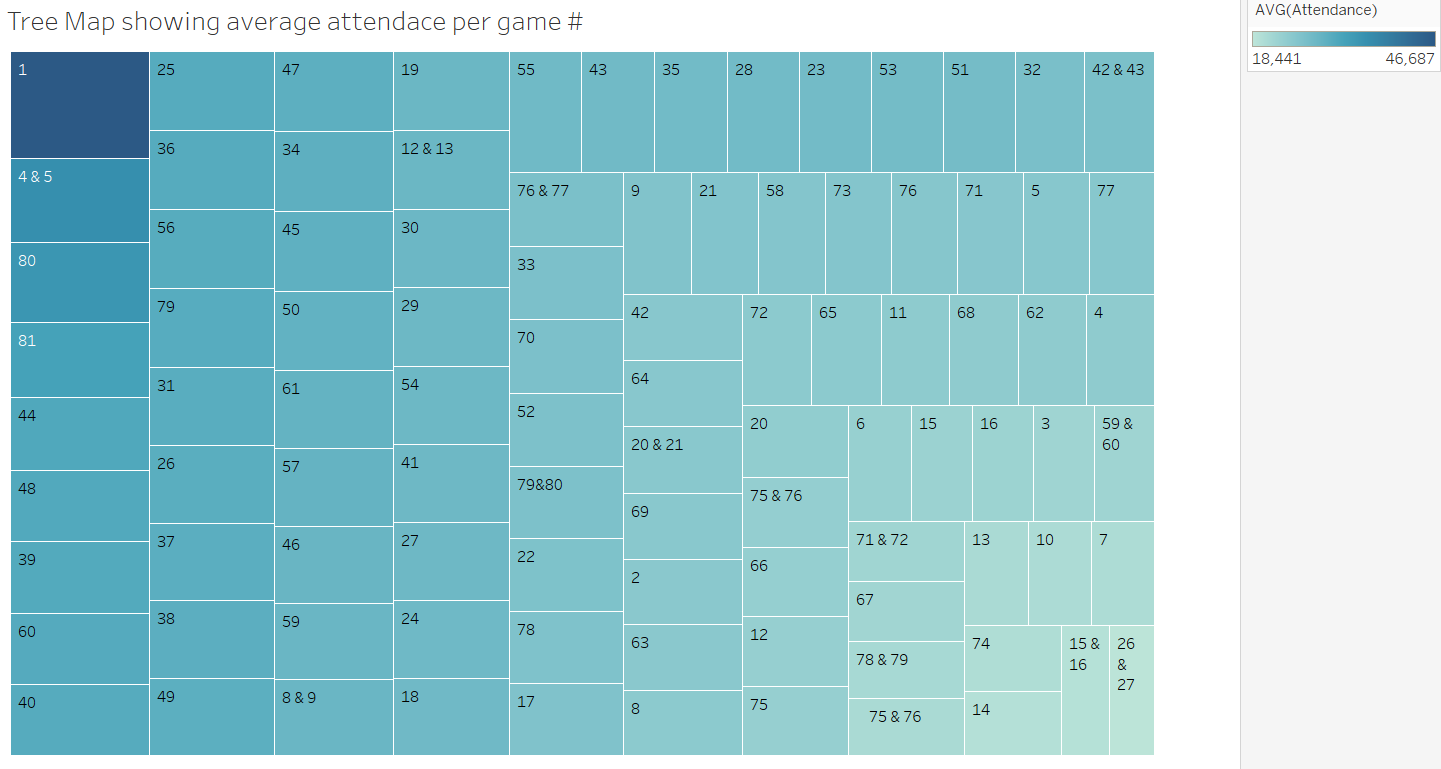


Exhibit 3: Tree map showing average attendance per game # (from Sheet 3 in the Tableau workbook)

Here in exhibit 3, we can see a tree map which displays a heat map showing which game number has the highest number of attendance. As expected, game number 1, that is the opening day, will have the highest number of spectators attending the game. From the tree map we can also see that game number 80 and 81 are also attended in numbers by the Orioles’ fans. From the tree map we can also observe that games 26 and 27 are the least attended by the fans.

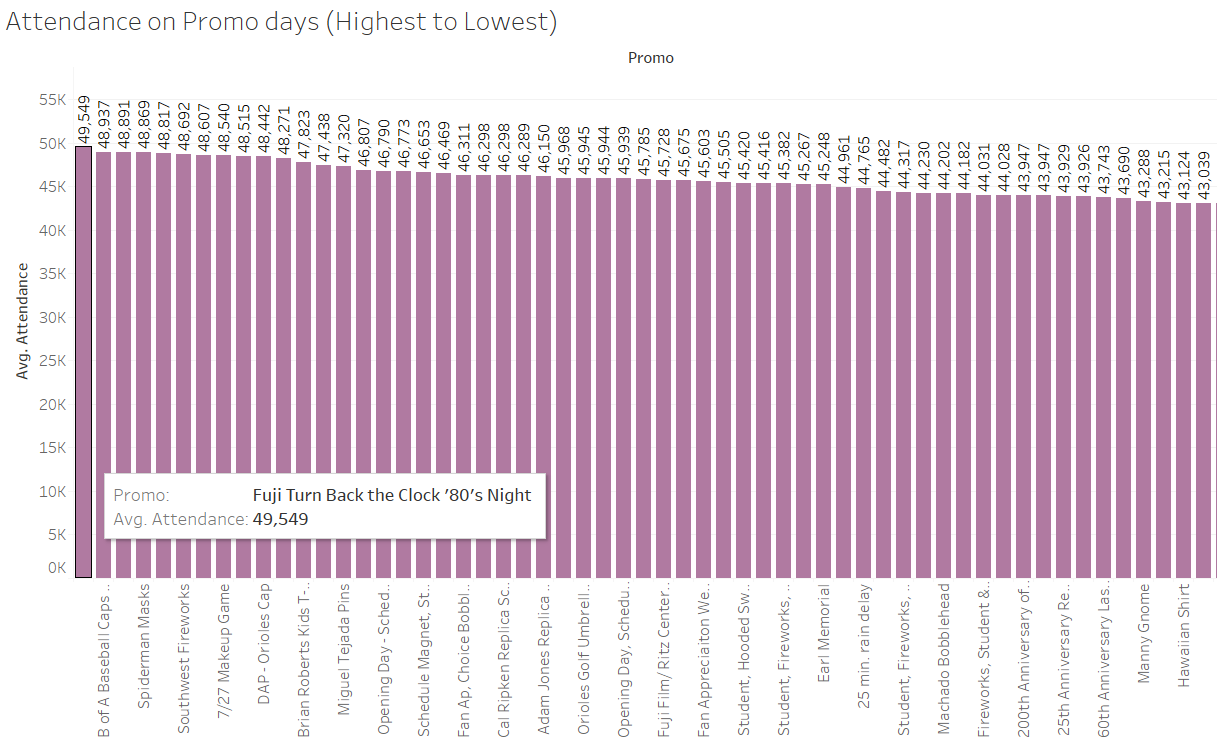


Exhibit 4: Horizontal bars showing attendance (highest to lowest) based on Promos (from Sheet 4 in the Tableau workbook)

In Exhibit 4, we can see the highest to lowest attended games when a promotional event was occurring in a match. We can observe from Exhibit 4 and Sheet 4 respectively in the Tableau workbook that when the promo event ‘Fuji Turn Back the Clock ‘80’s Night’ happened in a game, the attendance was 49,549 on that particular event, which is the highest attendance recorded when a promotional event took place. Similarly, we can check out all the other promotional events which had large attendance.

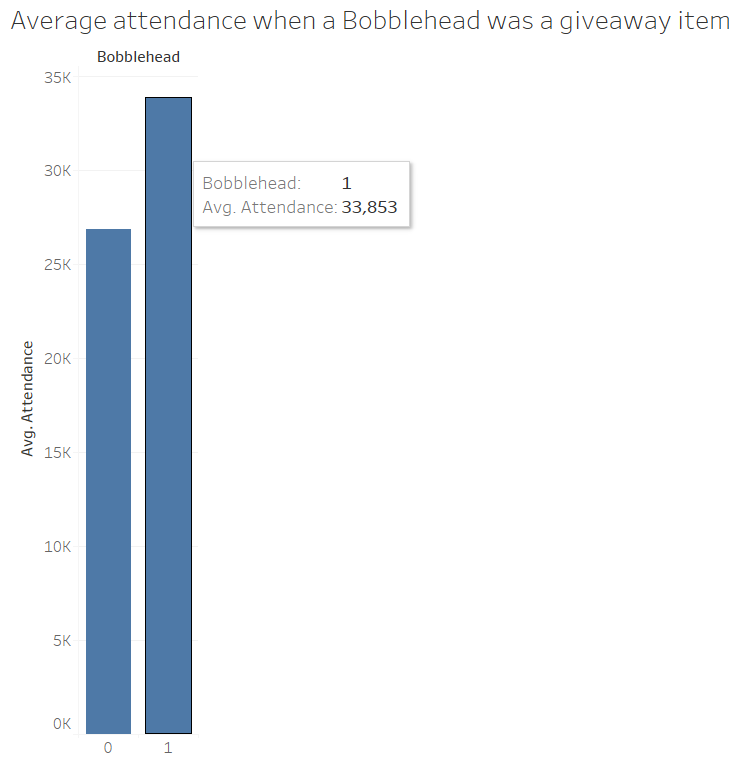


Exhibit 5: Vertical bars showing average attendance when a Bobblehead was a promo/ giveaway item (from Sheet 5 in the Tableau workbook)

In Exhibit 5, we can see that the average attendance goes north of 30,000 whenever there is a Bobblehead as a Promo/ Giveaway item. Since the Bobblehead column is in binary, therefore the result is a vertical bar graph with only 2 vertical bars.