

Modeling UFC Pay Per View Sales

Ryan O'Connell

Introduction

Since its creation in 1993, the UFC has primarily relied on a pay per view model to generate revenue. This makes them very different from most other sports organizations, as most of their revenue in a year comes from just a few nights where they do well. Thus, not only is it extremely important for fighters to have success in the cage, but it is also critical that they are able to generate pay per view sales for the company. In this paper, I am going to be examining the attributes of fighters that impact revenue from pay per view sales.

Lit Review

Watanabe, Nicholas M. "Sources of Direct Demand: An Examination of Demand for the Ultimate Fighting Championship.":

This paper examines various factors and their impact on UFC demand. It examines the demand for both live attendance and tv viewership, to try to determine what factors are relevant and how they differ. It utilizes 2 different OLS regression models, one of which to estimate in person and one for TV viewership.

The model utilizes PPV buys as the dependent variable with price, MSA income, MSA population, capacity, days from previous fight, a time trend and time trend squared variable, a holiday dummy, the number of fights on cable as well as a dummy variable for a title fight in each weight class.

The PPV model had an r^2 of 0.59, indicating it accounted for slightly over half of the variability in PPV buys. The holiday dummy as well as the ones for featherweight, welterweight, light heavyweight, and heavyweight championships are all also statistically significant in this model. The holiday coefficient is negative, and all weight class coefficients outside of

featherweight are positive. Only heavyweight is significant at 0.01, with light heavyweight and featherweight just missing this threshold and the welterweight and holiday dummy's are only significant at the 0.1 level. The heavyweight one (and maybe light heavyweight showing some smaller significance) seems like it could make sense as heavyweight is commonly believed to create the most demand, especially when examining the history of boxing.

While this paper did somewhat answer the research question, there are also some questions with the results. I think there may be some takeaways, but the predictors still aren't great. In terms of the other weight class dummy's, I don't think much can be taken away from them unless a logical explanation were provided for their effect on demand. The effect of the holiday dummy is difficult to interpret, as you would think different holiday weekends have different effects. However, the statistically significant negative coefficient does seem to show some level of decreased demand on holidays.

Robbins, Thomas R., and James E. Zemanek Jr. "UFC pay-per-view buys and the value of the celebrity fighter."

This paper tries to examine the value of the celebrity fighter within the UFC. While all sports generally have some aspect of star power, where star players will add significantly more value than other players, this is especially the case within combat sports as there are no teams and it is all about the individual fighters.

This paper utilizes pay per view buy numbers to determine fighter value, as this perfectly correlates with how much money they bring to the company. There is an extremely high level of variability in these numbers, with many observations around 100,000, and about 10% of the observations above 1,000,000 pay per views. An OLS model was created with a dummy variable

for each of the 57 fighters that appeared in at least 4 main events as well as other indicators such as whether it was a title fight or there were multiple title fights. A forward selection model was ran, and only 8 variables remained with all of them being individual fighter dummy variables. All of these were statistically significant at the 0.05 level, 6 of which at the 0.01 level as well. The model has an R^2 value of 61.3%. These 8 fighters were Conor McGregor, Brock Lesnar, Ronda Rousey, Georges St-Pierre, Forrest Griffin, Jon Jones, Anderson Silva, and Demetrious Johnson.

Intuitively, the fighters on this list seem to make sense. Conor McGregor has the largest coefficient with 973,169, followed by Brock Lesnar with 518,021 and Ronda Rousey with 401,426. Conor McGregor is somewhat of an anomaly in MMA, and nobody has elevated themselves to the same level of fame through the sport so it would be concerning if the McGrgeor coefficient was not this large. The next 2 are Georges St-Pierre with a coefficient of 325,595 and Forrest Griffin with a coefficient of 247,292. Jon Jones and Anderson Silva and were both in between 100,000-200,000. Demetrious Johnson's coefficient was -168,450, indicating he was the only fighter included who had a negative impact on pay per view demand.

These findings do seem to confirm the impact of the celebrity fighter within MMA. Between the top 2 fighters in pay per view coefficient (McGregor and Lesnar) there is 1 total title defense, meaning they did not have dominant careers. While they were both good fighters who won a belt, this is not why they drew eyes. Conor McGregor has been a media sensation like we've never seen and Lesnar brought over his large fanbase from the WWE. Following these 2 fighters are Ronda Rousey and Georges St-Pierre. While these 2 were both extremely dominant and some of the best fighters ever, we can attribute a lot of Ronda's influence to being a pioneer of women's MMA and becoming very famous through media similar to McGregor. St-Pierre, on

the other hand, was not known for his media presence and likely sold pay per views with his fighting skills. Next on the list is Forrest Griffin, who never won a belt and his coefficient of nearly 250,000 can almost certainly be traced to the fame he gathered in the Ultimate Fighter Season 1.

And the bottom 3 on this list: Jones, Silva, and Johnson. Each of these fighters has a serious argument to be the greatest fighter of all time and are easily considered so in their weight class. While Jones and Silva have a positive effect on demand, it is still not to the level of the other fighters on the list, most of whom are not as good as these 2. While Jones and Silva are able to do good numbers based on their skills, they do not have the same celebrity fighter effect as some of the other people mentioned. Johnson, on the other hand, had a negative impact on demand. This makes sense, as while he was a great fighter he always struggled to get attention in the UFC and essentially ended up being traded away as they did not see the value in him.

Gift, Paul. "Moving the needle in MMA: On the marginal revenue product of UFC":

This paper is about the marginal revenues of fighters within the UFC, and how it relates to fighter pay. It tries to determine whether different types of UFC fighters are underpaid or overpaid relative to the revenue they generate. This is a very interesting topic, as it is a very hot issue right now that people talk about, but the conversation is very emotionally driven and the people discussing it likely know nothing about the actual number behind it.

The paper uses a linear regression model with pay per view buys as a function of main card fighter variables and other event characteristics. The model contains year fixed effects, a variable F for summed covariates of fighters (consumer interest or win metrics) as well as an additional variable that is a vector of event characteristics such as price, title fights, or outcome

uncertainty. Two different general models are examined: one where F contains win metrics and another where it contains consumer interest metrics.

For win metrics, 4 different models are utilized using either win percentage, wins in last 5 fights, total wins, and wins minus losses. Winning does not seem to be an indicator of consumer interest in these regressions, with the largest R^2 in the models being 0.43. Three regressions were examined for other consumer interest variables (not winning), and 2 of them yielded R^2 values of 0.84 and 0.76, indicating these other variables likely explain consumer interest better than winning. Both regressions found a statistically significant correlation with consumer interest in the main event. Additionally, interest in the main event is much larger than the interest in any other bout position. The model also found that main event fighter popularity has a much larger effect than uncertainty of outcome, and uncertainty of outcome hypothesis does not seem to be a factor here.

Fighter MRP's were calculated as their contribution to pay per view buys times the marginal effect of PPV buys on UFC revenue. It was found that 15 of the 509 main card fighters (3%) generated 50% of the total MRP. Fighters are classified into 3 types to determine how fighters are paid. Type 1 is fighters who appeared in a main card and ranked top 44 in terms of revenue generated. Type 2 is the fighters who appeared on a main card and are not in the top 44, and type 3 is fighters who never appeared on a main event.

The compensation gap for group 1 was found to be 1.93 million, and these fighters earned just 8.5% of their MRP. They are only overpaid about 1/3 of the time. Type 2 fighters have a compensation gap of 91,000, and their compensation is 31.8% of their MRP. They are overpaid 77% of the time. Type 3 fighters, on the other hand, had a MRP compensation gap of -16,000, making them always overpaid.

This gives us some new insights into the research question. A lot of the discussion around fighter pay centers around fighters on the bottom of a card not making enough, but this paper seems to indicate they are overpaid in economic terms. Because many of them make minimum salary, the UFC does not have an option of changing their pay, so their only option here would be to remove the fight from the card to eliminate this loss. However, it is through these fights that the future stars of the sport who will become type 1 fighters develop and rise, so the UFC is ok with paying a bit of a premium here to find future talent.

The argument that the top fighters in the sport are underpaid, however, does appear to hold in this paper. Many would think that these fighters have the most bargaining power, but it appears the UFC has most of the leverage here. This is likely due to the UFC dominating the MMA landscape, and other organizations such as Bellator or One FC likely do not have the money to pay these guys, so they are forced to accept whatever they get in the UFC.

Data & Methodology:

I decided the best way to approach the question of what affects UFC demand is to try to model UFC revenue utilizing various fighter attributes. Revenue will be examined instead of sale numbers to account for price increases. If prices are not considered, there would be various points in the data where it appears the demand is artificially decreased overnight. I want to include predictors in the model to capture attributes such as fighter pace, style (striking vs. grappling), and damage done to see what types of fighters sell more pay per views. I also want to include predictors that capture the fighters' legacy within the UFC as well as uncertainty of outcome bias.

Pay per view data was obtained from tapology.com. Fight statistics were scraped from ufcstats.com for every UFC fight, and this data will be used to calculate attributes for fighters. Fight odds were scraped from bestfightodds.com. Pay per views since UFC 100 will be examined in the model. There are 145 cards included in the data, with 42 cards since UFC 100 missing due to unavailable pay per view data.

The first step is to determine the different predictors to be used, starting with the fighter attributes. Each of these attributes is calculated for each fighter, and the average of the 2 will be used in modeling. I decided significant strikes landed per minute (lands per minute) seems like a good measure of pace, and this was calculated as the total career significant strikes landed for a fighter divided by total fight time. Takedowns and control time per minute seem like good measures of style, as higher numbers of these would indicate more of a grappler. Knockout percentage and submission percentage were also included to see if finishing fights affects sales. The final fight metric included is knockdowns per 15 minutes (knockdown rate), as this seems like an aspect of fights fans want to see more of. The following table showcases the summary statistics for each of these attributes:

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
lnd_min	145	3.854	1.003	1.142	3.165	4.412	6.769
td_min	145	0.126	0.066	0.000	0.079	0.161	0.400
ctrl_min	145	0.252	0.119	0.017	0.174	0.312	0.707
kd.15	145	0.663	0.478	0.000	0.341	0.879	3.413
ko.pct	145	0.422	0.181	0.000	0.315	0.548	0.909
sub_pct	145	0.160	0.136	0.000	0.056	0.236	1.000

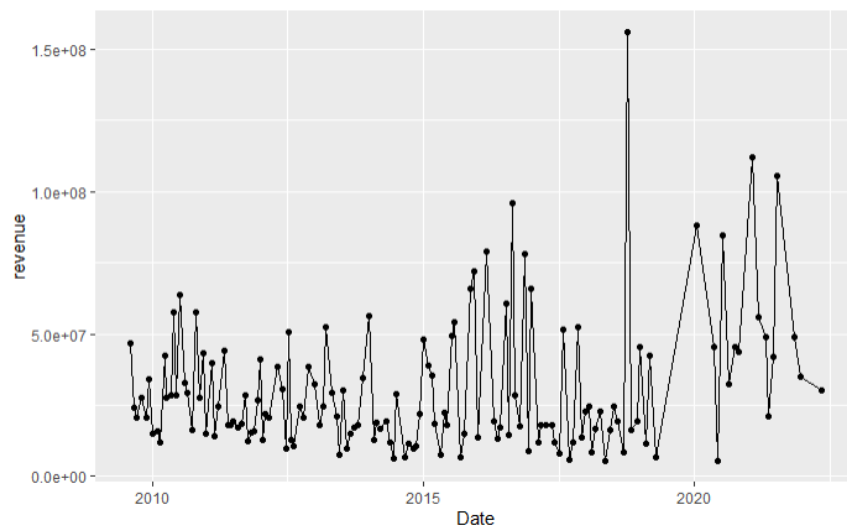
To capture legacy, I chose to include past title fight wins for each fighter. It is important to note that the red fighter is usually the defending champion, so this fighter will often have multiple of these. Because the blue corner is the challenger, many of these fighters will not have any previous title wins. 137 out of the 145 observations in the red corner had a previous title

fight win, while this was only the case for 69 of the 145 fighters in the blue corner. The average previous defenses were 3.88 for the red corner, and 1.36 for the blue corner.

Weight class was also included in the model. There were the most light-heavyweight events in this data with 30, and the least was flyweight and women's featherweight with 4 in each of these. Odds were calculated as the average of the favorite and underdog's betting odds, so higher numbers are fights with a clear favorite and numbers closer to 100 are close fights.

Observations by weight class	
f	n
Bantamweight	6
Catch weight	2
Featherweight	10
Flyweight	4
Heavyweight	20
Light Heavyweight	30
Lightweight	21
Middleweight	18
Welterweight	21
women's Bantamweight	9
women's Featherweight	4

The final variable to be considered in the model is a time trend as it is very possible to think that demand can be moving in a certain direction over time. The following graph shows the total revenue for UFC fights throughout the of the data:



The graph is not extremely clear, but it appears that there may be a trend over time in this data. The numbers of the higher selling pay per views clearly increase over time, but the revenue for the pay per views that perform poorly seems to be decreasing. I will begin modeling with a time trend, and later determine if it will stay in the model.

Results and Discussion:

The purpose of the first model ran is to determine what fights will be included for each pay per view. This model includes all of the predictors mentioned above, and has the fighter statistics for both the main event and co-main event fighters:

MODEL FIT: $F(41,103) = 2.72, p = 0.00$ $R^2 = 0.52$ $Adj. R^2 = 0.33$					me_b_title.wins	2570316.50	921211.19	2.79	0.01
<i>Standard errors: OLS</i>					cm_weightCatch Weight	3698203.42	24077565.00	0.15	0.88
					cm_weightFeatherweight	3657359.43	11341234.77	0.32	0.75
					cm_weightFlyweight	-475680.53	12311189.76	-0.04	0.97
					cm_weightHeavyweight	-6339568.48	10892327.29	-0.58	0.56
					cm_weightLight Heavyweight	3933630.13	10079081.93	0.39	0.70
					cm_weightLightweight	8049429.93	9971586.28	0.81	0.42
					cm_weightMiddleweight	-1497703.61	10325203.43	-0.15	0.88
					cm_weightUltimate Fighter	-17231175.99	30044660.69	-0.57	0.57
					Brazil 1 Middleweight				
					Tournament	5271404.47	9785033.94	0.54	0.59
					cm_weightWelterweight	11561097.43	12301039.58	0.94	0.35
					cm_weightwomen's Bantamweight				
					cm_weightwomen's Featherweight	11693994.93	17231633.34	0.68	0.50
					cm_weightwomen's Flyweight	-1242014.56	22966013.65	-0.05	0.96
					cm_weightwomen's Strawweight	-5598410.11	10950862.08	-0.51	0.61
					cm_titleTRUE	5002468.32	5055018.75	0.99	0.32
					cm_lnd_min	-1701625.79	2453465.93	-0.69	0.49
					cm_tld_min	-28520629.97	52972333.47	-0.54	0.59
					cm_ctrl_min	10890224.64	32265861.26	0.34	0.74
					cm_kd.15	-340465.93	4294964.45	-0.08	0.94
					cm_ko.pct	18328985.42	14436116.56	1.27	0.21
					cm_sub.pct	-18132522.18	16925800.56	-1.07	0.29
					cm_trend	91821.91	67684.54	1.36	0.18
					odds	-4151.92	9058.11	-0.46	0.65
					me_r_title.wins	1420470.00			

As you can see from the output, the co-main event attributes add very little predictive power to the model. The lowest p-value for these is 0.45, and most of them are around 0.9. The following model removes the co-main event attributes:

MODEL FIT: $F(23,121) = 4.30, p = 0.00$ $R^2 = 0.45$ $Adj. R^2 = 0.35$									
<i>Standard errors: OLS</i>									
					Est.	S.E.	t val.	p	
					(Intercept)	-14171456.93	12211908.81	-1.16	0.25
					weightCatch weight	-7311589.86	18160884.65	-0.40	0.69
					weightFeatherweight	1266495.63	10418685.39	0.12	0.90
					weightFlyweight	-3901169.84	13307926.61	-0.29	0.77
					weightHeavyweight	-10780808.42	10479884.68	-1.03	0.31
					weightLight Heavyweight	6832394.79	9632668.57	0.71	0.48
					weightLightweight	18705947.69	9397960.15	1.99	0.05
					weightMiddleweight	765447.52	10282633.44	0.07	0.94
					weightWelterweight	15135406.63	9978040.54	1.52	0.13
					weightwomen's Bantamweight	14305258.88	11546184.24	1.24	0.22
					weightwomen's Featherweight	31082.85	13167584.62	0.00	1.00
					titleTRUE	-6850448.42	4871035.70	-1.41	0.16
					lnd_min	3929766.99	2504938.07	1.57	0.12
					td_min	-7219983.98	42651001.66	-0.17	0.87
					ctrl_min	36894185.10	23023970.96	1.60	0.11
					kd.15	20366450.40	5376507.44	3.79	0.00
					ko.pct	-9564493.39	13265420.60	-0.72	0.47
					sub.pct	-2484692.43	14058293.31	-0.18	0.86
					me_r_title.wins	1449898.15	641298.39	2.26	0.03
					me_b_title.wins	2284533.06	822733.50	2.78	0.01
					cm_titleTRUE	1784466.63	4261251.35	0.42	0.68
					third_titleTRUE	12763304.37	8345417.32	1.53	0.13
					trend	50770.71	59152.79	0.86	0.39
					odds	-5276.15	8214.16	-0.64	0.52

When the co-main event attributes are removed, the R^2 value does not decrease by a lot. Based on the little predictive power they offer as well as Paul (2020) finding that main event's drive most of the UFC's demand, I will only include the fighter statistics for main event fighters. I will, however, still include variables for if the second or third fight on the card was a title fight, as having multiple title fights on a card is a big deal and adds value to it.

Before further simplifying the model, I am going to re-evaluate the criteria for the fighter statistics and see if they can be improved. As I mentioned earlier, these were calculated using career statistics for fighters. However, this may not be the best way to go about doing this. Many fighters will change how they fight over time, so including data within these calculations that is over a decade old for some fighters may just be adding noise.

I will now consider calculating the fighter statistics in the model for only recent fights. I will run the model I just created with different sets of data; each calculating these statistics with a different number of years of data before the fight. I will then compare the models using residual sum of squares. The following table shows the residual sum of squares for each of these models.

=====		(*10 ¹⁶)
Years	RSS	

Full Career	4.37	
12	4.32	
11	4.30	
10	4.28	
9	4.24	
8	4.29	
7	4.31	
6	4.36	
5	4.36	
4	4.67	

Starting from the full data, the RSS decreases as only more recent data is examined. It seems more than 9 years being included is not helpful to the model and is instead adding in unnecessary data. This makes sense as this is extremely old data from somebody's career. However, as less than 9 years are included the RSS is also decreasing. It appears removing this data from the calculations is not improving the model, and the ideal number of years of data to include in the data seems to be 9. This is the output with the new data:

MODEL FIT:
 $F(23,121) = 4.30$, $p = 0.00$
 $R^2 = 0.45$
Adj. $R^2 = 0.35$

Standard errors: OLS

	Est.	S.E.	t val.	p
(Intercept)	-14171456.93	12211908.81	-1.16	0.25
weightCatch weight	-7311589.86	18160884.65	-0.40	0.69
weightFeatherweight	1266495.63	10418685.39	0.12	0.90
weightFlyweight	-3901169.84	13307926.61	-0.29	0.77
weightHeavyweight	-10780808.42	10479884.68	-1.03	0.31
weightLight Heavyweight	6832394.79	9632668.57	0.71	0.48
weightLightweight	18705947.69	9397960.15	1.99	0.05
weightMiddleweight	765447.52	10282633.44	0.07	0.94
weightWelterweight	15135406.63	9978040.54	1.52	0.13
weightWomen's Bantamweight	14305258.88	11546184.24	1.24	0.22
weightWomen's Featherweight	31082.85	13167584.62	0.00	1.00
titleTRUE	-6850448.42	4871035.70	-1.41	0.16
lnd_min	3929766.99	2504938.07	1.57	0.12
td_min	-7219983.98	42651001.66	-0.17	0.87
ctrl_min	36894185.10	23023970.96	1.60	0.11
kd.15	20366450.40	5376507.44	3.79	0.00
ko.pct	-9564493.39	13265420.60	-0.72	0.47
sub.pct	-2484692.43	14058293.31	-0.18	0.86
me_r_title.wins	1449898.15	641298.39	2.26	0.03
me_b_title.wins	2284533.06	822733.50	2.78	0.01
cm_titleTRUE	1784466.63	4261251.35	0.42	0.68
third_titleTRUE	12763304.37	8345417.32	1.53	0.13
trend	50770.71	59152.79	0.86	0.39
odds	-5276.15	8214.16	-0.64	0.52

Now that I have figured out the data to use it is time to further simplify the model. I will run a forward selection process on the model, beginning with only an intercept. The following output resulted from the stepwise regression:

MODEL FIT:
 $F(16,128) = 6.40$, $p = 0.00$
 $R^2 = 0.44$
Adj. $R^2 = 0.37$

Standard errors: OLS

	Est.	S.E.	t val.	p
(Intercept)	-23346546.31	11029689.65	-2.12	0.04
kd.15	20411220.71	4125198.65	4.95	0.00
weightCatch weight	-10768475.10	15861197.60	-0.68	0.50
weightFeatherweight	1146516.50	9732924.67	0.12	0.91
weightFlyweight	-5099003.43	12216182.77	-0.42	0.68
weightHeavyweight	-12760242.84	9407015.21	-1.36	0.18
weightLight Heavyweight	6029752.63	8628553.94	0.70	0.49
weightLightweight	19441483.63	8713838.35	2.23	0.03
weightMiddleweight	1832498.03	9309197.54	0.20	0.84
weightWelterweight	17385456.32	8811630.92	1.97	0.05
weightWomen's Bantamweight	13324437.16	10423221.66	1.28	0.20
weightWomen's Featherweight	1035212.63	12039394.10	0.09	0.93
me_b_title.wins	2541777.33	777981.58	3.27	0.00
me_r_title.wins	1229105.70	605877.89	2.03	0.04
lnd_min	4607945.44	1950934.39	2.36	0.02
third_titleTRUE	14425507.08	7528282.60	1.92	0.06
ctrl_min	29773752.93	15644131.18	1.90	0.06

This model appears to be much better than the previous model. 7 predictors were removed from the model, and the R^2 only decreased by about 0.03. The predictors removed were takedowns per minute, knockout and submission percent, the main and co-main title indicator, time trend, and odds. It seems like a good call to remove either takedowns or control time due to multicollinearity, and knockout and submission percentage did not seem to add much to the model. The time trend and odds also added little, and I will leave these out as well. The only predictor I would consider putting back in is the co-main event title indicator, mostly due to the variable for a third title being included. However, I am still going to leave this out of the model as I think there may be an explanation for this. There were two title fights in 42 of the 145 cards in the data, and this is something the UFC frequently does. They do not necessarily do this to try to stack cards, but instead because these fights likely would not do well as main events. Three title fights, however, only occurred in 8 of the 145 cards. The UFC rarely puts this many title fights on a card, and the few times they do this they seem to try to stack cards and make it a big deal. Because of this, I will stick with the predictors chosen in the stepwise selection process and not add back in the variable for two title fights.

In the new model, lands per minute and third title are both now statistically significant at the 5% alpha level. Additionally, control per minute is now significant at the 10% alpha level. Knockdown rate is by far the strongest predictor, and this number increasing by 1 is estimated to add about \$20 million in revenue. Lands per minute also had a strong positive effect on the model. I am not surprised by the impact these have on the model, as it is generally believed that fans like to watch striking matches with a lot of action. That is exactly what an increase in these attributes is indicative of, so the model seems to back up this idea.

I was a little bit surprised that control time had a positive coefficient. Because of the reasons that I just said, I would think fans prefer fights with less control time as that would generally be a striking match. This could be indicative of fans liking higher level fighters in general, as better grapplers will get more control time.

The positive impact of previous titles won for both corners obviously make sense, as many fighters become superstars by racking up title defenses. It is important to remember that the red corner is usually the defending champion, so the red corner generally has previous title fight wins. As the challenger, however, it is much less likely that this fighter has previous title fight victories. This makes it a bigger deal when there is a fighter in the blue corner who has title defenses, as this means both fighters have been champions at some point. The model seems to back this up as well, as the coefficient for blue corner title defenses is about twice as large.

Conclusion

Overall, I have formed some new insights as well as reinforced some previous ideas about what affects UFC demand. By running a linear regression model containing fighter attributes for both the main and co-main event of pay per views, I reaffirmed the belief that the majority of UFC demand is driven purely by the main event. The only time when other fights seemed to have a significant impact is with multiple title fights, and even then, the difference is marginal with only 2 title fights on a card.

Through stepwise selection various predictors were removed from the model. The removal of odds seemed to indicate that uncertainty of outcome bias was not present.

Additionally, the time trend did not seem applicable here, indicating the revenue does not seem to inherently increase over time.

The results of the final model mostly reinforced my beliefs about what drives UFC fight demand. I figured that fans are more interested in paying for high-paced, high-action stand up fighters, and the effect of strikes per minute and knockdowns per 15 minutes showcase this. This is interesting, as if a fighter wanted to potentially try to maximize their profits, they could keep this in mind and try to fight in a way that is more exciting to the fans as opposed to just trying to win (some would argue fighters such as Michael Chandler are already doing this).

Overall, I am satisfied with the results of these models. Some further research that could be interesting is combining this model with some measures of popularity outside the cage, such as google trends or social media following, to try to see how the effect of these compares to the effects I found with fighting attributes.

References

- Gift, Paul. "Moving the needle in MMA: On the marginal revenue product of UFC fighters." *Journal of Sports Economics* 21.2 (2020): 176-209.
- Robbins, Thomas R., and James E. Zemanek Jr. "UFC pay-per-view buys and the value of the celebrity fighter." *Innovative Marketing* 13.4 (2017): 35.
- Watanabe, Nicholas M. "Sources of Direct Demand: An Examination of Demand for the Ultimate Fighting Championship." *International Journal of Sport Finance* 10.1 (2015).