VAR Analysis in Professional Soccer

Dylan Blechner

Introduction

November 18, 2009. It was the all-important second leg of the FIFA Men's World Cup Qualifying playoff between France and the Republic of Ireland where the winner would move on to South Africa to compete for the World Cup and the loser would return home. After France won the first leg 1-0, Ireland leveled the score in the Stade de France sending the match to extra time. What happened in extra time would change the course of soccer history forever. French legend Thierry Henry illegally hit the ball with his hand in the buildup to a William Gallas goal that ended up being the match winner. It was a blatant missed call by the referee and sent the Irish players into a frenzy. Henry later admitted to his "crime" and the debate to introduce video refereeing was fueled.

Ten years later, video assistant referees (VAR) are on hand for almost every professional soccer match and the technology has been used in the FIFA Men's and Women's World Cups, UEFA Champion's League, and each of the top five professional soccer leagues in Europe (English Premier League, German Bundesliga, Spanish La Liga, Italian Serie A, French Ligue 1). While blatant missed calls like the "Henry Incident" have been removed from the game, controversy remains due to the specific rules of VAR.

There are four categories of decisions that can be reviewed by VAR: a goal, a penalty kick, a red card, and mistaken identity (red or yellow card awarded to the wrong player). A specified video assistant referee looks at every event on the field and is in direct contact with the head on-field official during the match. If the video assistant referee feels the head official may want to view a particular call, a signal is made and the game is stopped while the referee looks at the replay. After review, the head official then makes his final call. The head official may also yield his judgement to the video assistant referee and agree with the decision he makes. The majority of controversy surrounding VAR has to do with the idea of a "clear and obvious error." Subjective refereeing errors may be reviewed with VAR, but they can only be overturned if the error made on the field was "clear and obvious." This has led to much controversy among soccer fans and pundits who feel erroneous decisions have been made.

I became very interested in this area of research while watching the 2019 FIFA Women's World Cup. It seemed there were many more VAR decisions than in other professional competitions and there was more controversy surrounding said calls. The Washington Post's Steven Goff[1] noticed this as well and wrote an article during the tournament on the subject of VAR. He noted that the number of VAR calls in the Women's World Cup was higher than those at the Men's World Cup (29 in 44 matches vs. 20 in 64 matches), which was worrying to some. FIFA kept track of their refereeing decisions and made the claim that through 44 matches of the Women's World Cup, referees had been correct on 98.2% their calls which is an improvement from the 92.5% accuracy without VAR. Even with this assertion that VAR is improving the game, some people are upset over the use of the technology to negate goals in tight decisions. I wanted to dive deeper into a relatively new area of the game of soccer to analyze VAR in a novel way.

This analysis will be broken down into two distinct portions that will look at different aspects of the new VAR technology. First, I will analyze the on-field impact of VAR and will test several hypotheses through the use of various data analyses. I believe that the introduction of VAR has decreased the number of fouls, penalties, and offsides calls per match. Referees feel more comfortable letting tight offsides and penalty decisions go because they have the safety blanket of VAR to check if they are wrong. Referees would rather let the play go on and come back to make a VAR decision rather than to call the play dead right away. I also hypothesize that VAR decisions tend to favor the home team. When decisions are too difficult to make definitively, the referees are more likely to default to the home team playing in front of their fans. Finally, I believe that the decisions that VAR can review are the ones that have the greatest impact on winning. If the calls reviewed by VAR are of high significance, it is important to get them right.

Second, I will analyze the sentiment associated with tweets about VAR. I once again have several hypotheses that will guide my research. The first hypothesis is that VAR decisions during professional soccer matches increases discussion about VAR on social media. This may be a simple hypothesis, but if it is not met, then it is impossible to analyze sentiment associated with VAR as very few people would be talking about it. Between the sources I have researched and my own experiences as a fan, I believe that the use of VAR has sparked controversy among supporters on social media which has led to more activity on sites such as Twitter. The second hypothesis is that the sentiment of tweets about VAR decisions is more negative rather than positive or neutral. I believe that while the use of VAR has improved the game of soccer, fans are not used to the types of calls that are made with the assistance of the technology. As a result, fans are usually upset with the calls made on the field and take to Twitter to vent their frustrations.

Through my research I hope to analyze VAR decisions in professional soccer to help improve its use in leagues and competitions. Because the use of VAR is quite subjective, I will show that analysis of fan sentiment can be used as a proxy to determine which calls may have been incorrect. There is a lot of work that can be done to improve VAR to make the refereeing of the game as fair and accurate as possible. A better use of this technology can improve television ratings in professional leagues and improve overall fan interest.

Literature Review

Academic VAR Data Analysis:

One of the only academic articles found about VAR usage in professional soccer was by Alex Bartiromo[2] in which he discussed the role that penalties play in adding luck to a Major League Soccer (MLS) match. After the introduction of VAR to the MLS in 2017, he analyzed its impact on the number of penalties called in a match. Bartiromo concluded that there has not been enough data gathered on the impact of VAR as it had only existed for one and a half seasons at the time of the analysis. He found that the number of penalties did increase from 109 to 138 from 2016 to 2018, but the number of penalties per match remained fairly consistent by year. He was more successful in his analysis of early game penalties as PKs called within the first 20 minutes grew from 11 to 24 from 2016 to 2018. Overall, there was not enough data to draw any important conclusions.

Table 1: *MLS Penalty Data by Season (2011-2018)*[2]

Year	Avg. min. of PK	PKs	PKs per team	PK <20 mins.	Matches played	PK per match
2011	51:04	84	4.7	12	306	0.27
2012	51:22	72	3.8	9	323	0.23
2013	49:59	81	4.3	11	323	0.25
2014	51:20	134	7.1	18	323	0.42
2015	51:24	108	5.4	17	340	0.32
2016	55:11	109	5.5	11	340	0.32
2017	50:31	119	5.4	20	374	0.32
2018	52:41	138	6.0	24	391	0.35

Another academic paper about the impact of VAR on professional soccer was written by Carlos Lago-Peñas, Rey Ezequiel & Kalén Anto[3] in 2019 and discussed the more basic effects of the technology. They analyzed data from the Italian Serie A as well as the German Bundesliga, which were the first two domestic leagues to introduce VAR during the 2017-2018 season. Variables such as fouls, goals, offsides, penalties, playing time in the first and second halves, total playing time, red cards, and yellow cards were analyzed for games before and after the implementation of VAR. The data was gathered from “WhoScored” (whoscored.com), a reliable source for advanced statistics in professional soccer. The authors found that there was an increase in the number of minutes played in the first half, but not the second half. They also found a decrease in offsides calls, fouls, and yellow cards after VAR was implemented. Both of these conclusions make sense as VAR is designed to allow referees the chance to take a second look at important plays, so they will be less likely to blow their whistle in certain situations. The inclusion of VAR does come with the caveat that the game will be longer as time is needed to look at the video replay.

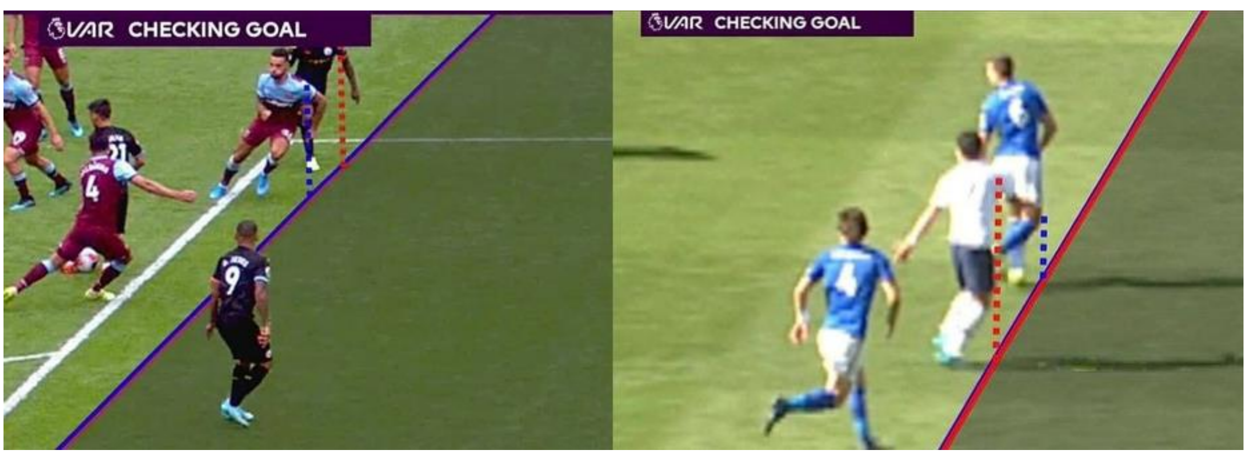
Non-Academic VAR Analysis:

Although little academic research has been done directly in the area of VAR and its various impacts on professional soccer matches and fan experience, many non-academic articles have been written discussing the accuracy of VAR decisions in key matches. One of the more popular articles was written by Dale Johnson[4] of ESPN FC who analyzed every VAR decision from the 2018 FIFA Men’s World Cup and rated the final decision out of 10. He studied not only referee decisions where the VAR technology was used, but also situations where he felt VAR should have been used. Overall, 33 VAR reviews were scored with an average rating of 7.61 out of 10. A brief summary of the incident was written, followed by the final rating and a reasoning of why the specific rating was given. This source was very useful as this was the first time VAR was used in a major international competition.

Dale Johnson[5] and others also wrote numerous articles about the use of VAR in the English Premier League during its first week of existence in 2019. The BBC[6] noted that there were 70 VAR checks that occurred behind the scenes during the first ten matches of the Premier League season. Prominent figures in the Premier League such as former England captain Alan Shearer, former referee Dermot Gallagher, and current Manchester City winger Raheem Sterling were asked about VAR and they gave their opinions about the technology. Shearer was very pleased with the Premier League during the opening weekend, saying, "they are using a high bar for VAR involvement and if it stays like that then it should be a success." Dermot Gallagher was very impressed with VAR's use and said "I wish I had it when I was a referee. I can look back at a few mistakes I made. As a referee, why would you not want to have it?" Even Raheem Sterling, whose Manchester City squad were disallowed a wonderfully worked team goal as a result of a fractional offside decision, said "if we're getting the decisions right then it can only be a good thing." While most people within the Premier League agree that VAR has been successfully used, some fans do not concur.

The opening weekend of the 2019-2020 Premier League season had some interesting VAR scenarios that put the technology to the test right off the bat. The most controversial decision from the weekend was in the Manchester City vs. West Ham United game where Gabriel Jesus' goal was disallowed after it was declared Raheem Sterling was marginally offsides in the buildup. This tight decision showed how the Premier League would judge offsides decisions: as a binary decision that will not be based on "clear and obvious" errors. This led to backlash from fans and much discussion around the league. Similar offsides decisions later in the season against Tottenham's Heung-Min Son and Liverpool's Roberto Firmino were also heavily criticized by supporters. Another call from the Manchester City match raised some eyebrows when Sergio Agüero was allowed to retake his weakly hit penalty because West Ham's Declan Rice stepped into the penalty box before Agüero made contact with the ball. Such a call had rarely been seen before the implementation of VAR, leading to more discussion around the league about how infrequently called infractions would be refereed under the new rules.

Figure 1: *Two marginal offsides decisions during the 2019-2020 English Premier League season including Manchester City's Raheem Sterling (left) and Tottenham's Heung-Min Son (right)*



Video Replay in Other Sports

I also studied published research in other sports where similar video refereeing has been used and a greater number of scholarly papers have been written. The Hawk-Eye technology in tennis, rugby, and cricket was the primary source of this research. Harry Collins and Robert Evans[7] wrote a paper a few years before VAR was implemented, and explained some of their concerns with the technology. The authors noted how the use of replay in certain sports has hurt the integrity of the game because of blatant mistakes that are shown to fans. VAR was primarily introduced to eliminate this situation where a "clear and obvious" refereeing error affects the game in a significant manner. Even after VAR's introduction, there are still scenarios where a decision can go either way, and the controversy that was supposed to be removed from the game of soccer has persisted.

Collins and Evans[8] wrote a follow-up paper where they compared and contrasted the research technology in cricket and tennis. The authors changed their tone and made recommendations for additional technologies in sports and concluded that the use of replay in cricket allows for a public discourse about the decisions made on the field. They argued that tennis should take a similar approach. This conclusion directly relates to VAR in soccer, as fans are now allowed to see many replays of key scenarios just as the head official does.

Overall, it seems that there are issues in other replay technologies in sports that have carried over to soccer. There are many pros and cons when it comes to allowing fans to see the mistakes of an official. While it can lead to a public discourse about the call, it may lead to controversy as well. VAR has the additional problem that the decisions are not as black-and-white as those with Hawk-Eye technology. Many additional rules were created for VAR decisions that are not needed with Hawk-Eye where the technology is only used to make decisions about whether the ball is called in or out. This is why so many fans and soccer pundits were concerned with the introduction of VAR technology.

Refereeing Issues in Professional Soccer

I will now discuss some of the issues with refereeing that explain the need for VAR in the current game. One paper by Gilis *et al*[9] looked at the various reasons for mistakes in offsides decisions in professional soccer. The authors cited the papers of Baldo et al. (2002) and Helsen et al. (2006) that consider the flash-lag effect as a reason for mistakes in offsides decisions. The flash-lag effect is an illusion where a flash and a moving object appear in the same location, but, in actuality, are not in the same place. They also examined decision-making by FIFA assistant referees as well as Belgian national referees. The results of the study supported the hypothesis as both FIFA and Belgian referees made errors in offsides decisions, especially when the attacker and defender in question were moving in the opposite direction. These calls are particularly difficult for professional referees and support the decision to implement VAR in the modern game.

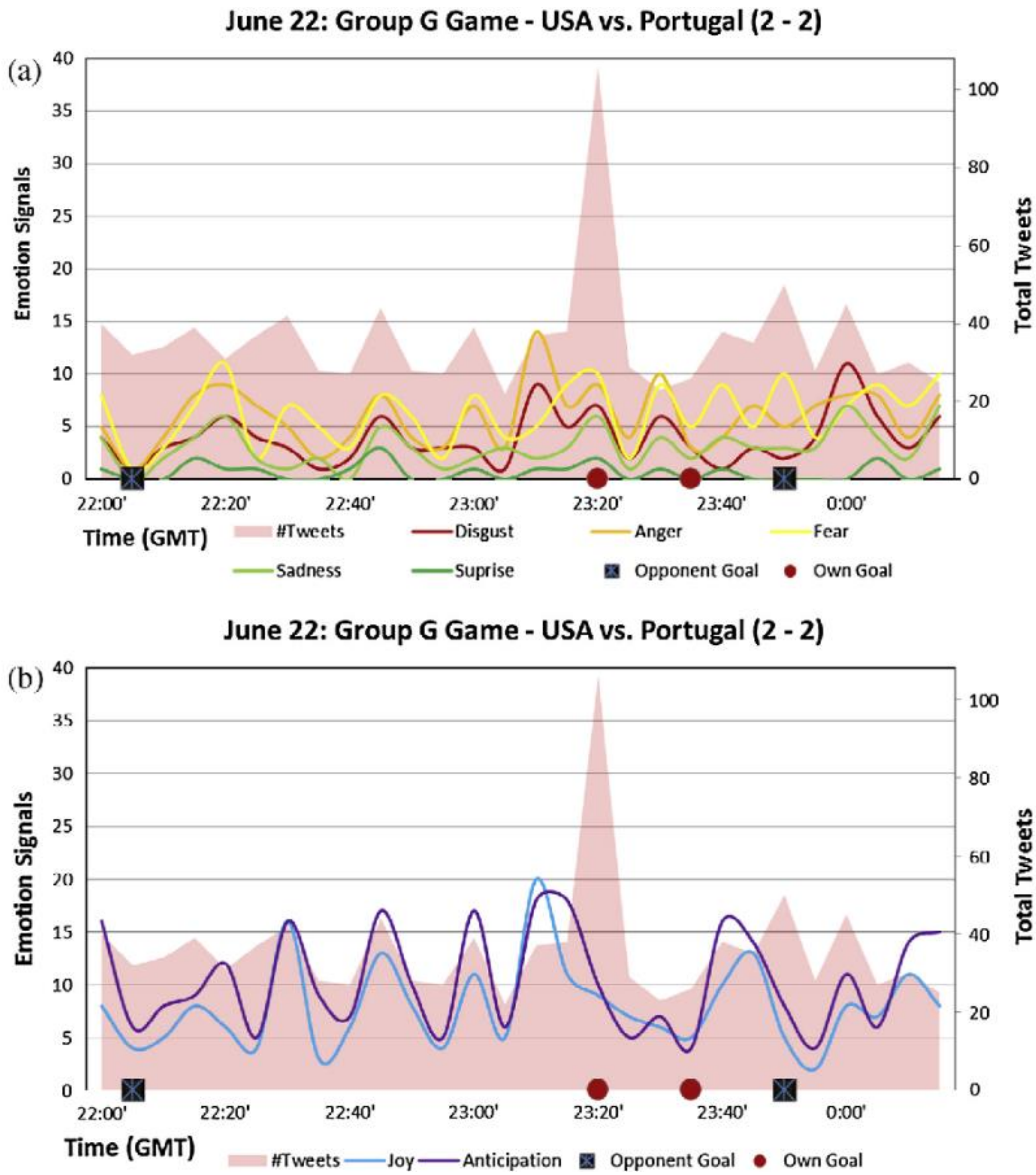
In another paper by Oudejans *et al*[10], three professional assistant referees judged 200 potential offsides situations in elite youth soccer. It was concluded that the referees made 40 errors. One possible explanation was that the referees shifted their view from the passer to the attacker and in that second of head movement, called the player incorrectly offsides. This was not the case as the referees wore cameras to track their head and body movement. A possible explanation is that when the referee is lined up slightly ahead of the last defender, it can give the perception that the attacker is offsides when, in reality, he is not. It was also discovered that when the attacker is closer to the referee than the last defender (in front of the last defender), the referee tended to make the correct call more often. This shows that referee placement and orientation on the field are extremely important when it comes to making the correct decision on the field. A similar paper by Mallo *et al*[11] looked to examine this effect of referee positioning on the accuracy of decisions during international soccer games. Research was conducted during the FIFA Confederations Cup in 2009. One key statistic to note from the study was that main officials were more prone to errors on plays in the middle of the field where help from assistant referees was limited. Also, more errors occurred for linesmen and main officials during the latter stages of matches. The distance of the linesman from an offsides decision did not impact accuracy, but the angle of viewing offsides did have an effect. From an angle between 46 and 60 degrees, linesmen called offsides more accurately. This paper certainly defends the use of VAR as it shows that referees fatigue similar to players.

Finally, a study by D'Ottavio and Castagna[12] examined the work-rate of Italian professional referees and its possible impact on decision-making. The analysis concluded that referees are put under incredible amounts of physical stress during matches, which makes their job even more difficult because they are not professional athletes. This is useful analysis to explain why VAR should be used as referees are physically and mentally drained throughout a match. Overall, these studies show that referees are prone to error for a variety of different reasons. By having the safety blanket of VAR, referees feel more comfortable making difficult calls as they have the ability to look at them again if need be.

Sentiment Analysis in Soccer

The new public discourse regarding referee decisions led me to research sentiment analysis during soccer matches. Is it possible to determine whether or not a VAR call is "correct" by using fan sentiment? In a paper by Yang Yu and Xiao Wang[13], the sentiment of United States soccer fans was analyzed during the 2014 FIFA Men's World Cup. The findings were to be expected as more fear and anger were shown after an opponent goal while joy was the primary sentiment indicator after a US goal. The authors concluded that using sentiment analysis was a generally accurate predictor of fan disposition during a soccer match.

Figure 2: *Sentiment analysis during the 2014 FIFA Men's World Cup Group Stage match between USA and Portugal*[13]



A paper by Corney, Martin, and Goker[14] analyzed matches during the 2012 and 2013 FA Cup Finals. Similar to the previous paper, they were able to identify key match events through spikes in Twitter usage and sentiment. They were also able to classify fans of specific teams through their tweets. The strategy used by both this paper and the previous were to analyze words, bigrams, and trigrams to see which phrases led to spikes in specific sentiment categories. Gratch *et al.*[15] used Twitter data from the 2014 World Cup to analyze fan sentiment during matches in the tournament. This paper specifically wanted to redefine what makes a sporting event exciting. The authors decided to classify all of the tweets into three categories: positive, neutral, and negative. They also studied the pre-match odds as well as in-game events to try and quantify excitement and used a classifier from the SemEval 2014 Twitter sentiment analysis challenge to classify their tweets. The classifier performed very well, and led me to believe that it would provide a good starting place for my own sentiment analysis.

Figure 3: *The number of Tweets from supporters from each team during the 2012 (left) and 2013 (right) FA Cup Finals at various points of the match*[14]

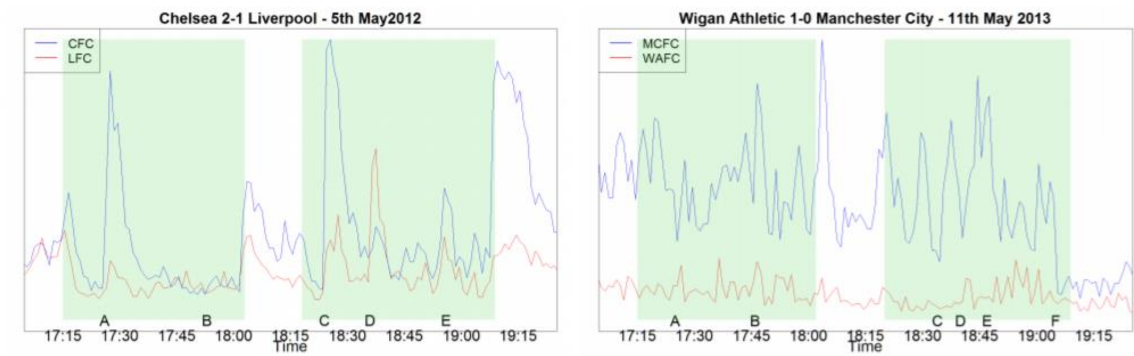


Table 2: *The corresponding labeled match events*[14]

Key	Team	2012 Final Event	Key	Team	2013 Final Event
A	CFC	Ramires scores	A	WAFC	MacManaman shot, misses
B	CFC	Mikel yellow card for a foul	B	MCFC	Tevez shot, saved
C	CFC	Drogba scores	C	MCFC	Zableta booked
D	LFC	Carroll scores	D	WAFC	MacManaman chance
E	LFC	Carroll shoots, saved	E	MCFC	Rodwell free-kick is saved
			F	WAFC	Watson scores

Possible Biases in VAR

In addition to analyzing sentiment during professional soccer matches to understand VAR calls, I also wanted to discover any possible biases in the usage of VAR. Does it favor the home team? The more skilled team? The team currently in the lead? A paper by Petersen-Wagner and Ludvigsen[16] examined possible biases of VAR during its inaugural use during the 2018 FIFA Men's World Cup. The authors were able to scrape over 300,000 YouTube comments from FIFA's official page to analyze fan sentiment during the World Cup. They created three unique categories of criticism towards VAR: Global North vs. Global South, Non-Neutrality of Technology, and VAR is Killing the Beautiful Game. They argue that VAR was operated in favor of European teams as their interpretation of the rules of soccer are more similar to those that VAR follows.

Win Probability Model in Soccer

Finally, I would like to analyze how a VAR decision impacts the win probability of a team during a match. A paper by Robberechts, Van Haaren, and Davis[17] created an in-game win probability model in soccer. They noted that such in-game models have been created in most other sports like baseball, basketball, and football, but because of the low-scoring nature of soccer, not much research has been done. They introduced a new Bayesian statistical model using 8 variables (Game Time, Score Differential, Rating Differential – ELO, Team Goals, Yellows, Reds, Attacking Passes – rolling average, Duel Strength – rolling average) to tally the running probabilities of soccer teams. They found that their model performed adequately which bodes well for future research along these lines.

Jay Boice[18] at FiveThirtyEight wrote an article explaining how soccer forecasts for the top five European soccer leagues are determined. They explained that their model is based on the Soccer Power Index (SPI) which was developed by Nate Silver. With SPI, every team is assigned an offensive and defensive rating which is determined based on expected goals scored and allowed. The model does not simply take into account goals scored in a match, but accounts for randomness with adjusted goals, shot-based expected goals, and non-shot expected goals. These variables account for how many goals “should” have been scored based on the shots taken (and non-shot actions) in the match. After coming up with these ratings, match outcomes are forecast using a Poisson model including the offensive and defensive ratings for both teams, home-field advantage, and the days of rest. They have also created tiers for the major soccer leagues by analyzing interleague play, which could be useful in my investigation. By using a combination of strategies from these two models, I will be able to analyze the importance of VAR decisions with respect to win probability added.

Data and Methodology

Data Summary

Data was gathered from several different sources for the separate analyses performed in this paper. For the sentiment analysis, Twitter data was gathered using the twitterscraper[19] module in Python. Tweets were collected in a two-step process. First, tweets written in English containing the term "var" or the hashtag "#var" were scraped for each day that a Premier League game occurred during the first half of the 2019-2020 season (between August 9, 2019 and December 27, 2019). Data was gathered in this manner to evaluate fan sentiment for English Premier League games specifically and eliminate any tweets that were not about VAR. Second, tweets containing the term "var" as well as the corresponding game-specific hashtag were scraped for each game day (e.g. "#LIVMCI" is the game-specific hashtag for a game between Liverpool and Manchester City). These tweets were gathered to analyze sentiment on a game level and determine if any individual VAR decisions were deemed by fans as incorrect. Because of the nature of the scraping process, many of the tweets in this dataset were also included in the first dataset. The game-specific hashtags as well as the EPL schedule were taken from the official Premier League website. Once these tweets were scraped, information about the authors was gathered including the number of followers and location by using the rtweet package in R. The account names of these authors will not be shared at any point during this analysis to preserve anonymity.

Commentary data was also gathered for VAR analysis from ESPN by using the fcscrapR[20] package in R. This commentary data includes information about key events that took place during a given match. Data was scraped for the top five professional soccer leagues in Europe (English Premier League, German Bundesliga, Spanish La Liga, Italian Serie A, French Ligue 1) between 2010-2019 which included each season that VAR was utilized. VAR decisions are specified in the commentary data and were then separated into the four categories of decisions that can be reviewed by VAR: a goal, a penalty kick, a red card, and mistaken identity. Although there are many VAR decisions during professional matches, ESPN only specifies those that caused a significant stoppage in play due to a referee check. Roster and performance data were gathered from WhoScored.com and betting odds were gathered from Football-Data.co.uk to evaluate any possible biases among VAR decisions. Finally, referee data was scraped from Transfermarkt to analyze the impact of individual referees involved with VAR decisions.

Summary Statistics: VAR

To begin my analysis of the impact of VAR on the game of soccer, I first wanted to compare summary statistics from the top five professional leagues in Europe before and after the introduction of VAR. Each of the top leagues have introduced VAR within the last three seasons of play (German Bundesliga and Italian Serie A in 2017-2018, French Ligue 1 and Spanish La Liga in 2018-2019, and the English Premier League in 2019-2020), so data was gathered from the 2012-2013 season to the 2019-2020 season for an adequate comparison.

The number of fouls committed per match was first analyzed to measure the impact VAR has had on the refereeing of matches. After running a two-sample t-test, I found that the average number of foul calls decreased after the implementation of VAR ($p < .001$). This can also be seen in the plots below. Referees are becoming more lenient with the assistance of VAR and would rather let a foul go than make an incorrect decision. Common fouls cannot be reviewed by VAR, but fouls in the box can be looked at. This is most likely why the difference in fouls is small, but noticeable. The number of fouls called in the Bundesliga has gone down more dramatically than in other leagues, suggesting more lenient refereeing in the German league over the past few seasons.

***Figure 4**: Number of fouls per game called by season in the top five European leagues*

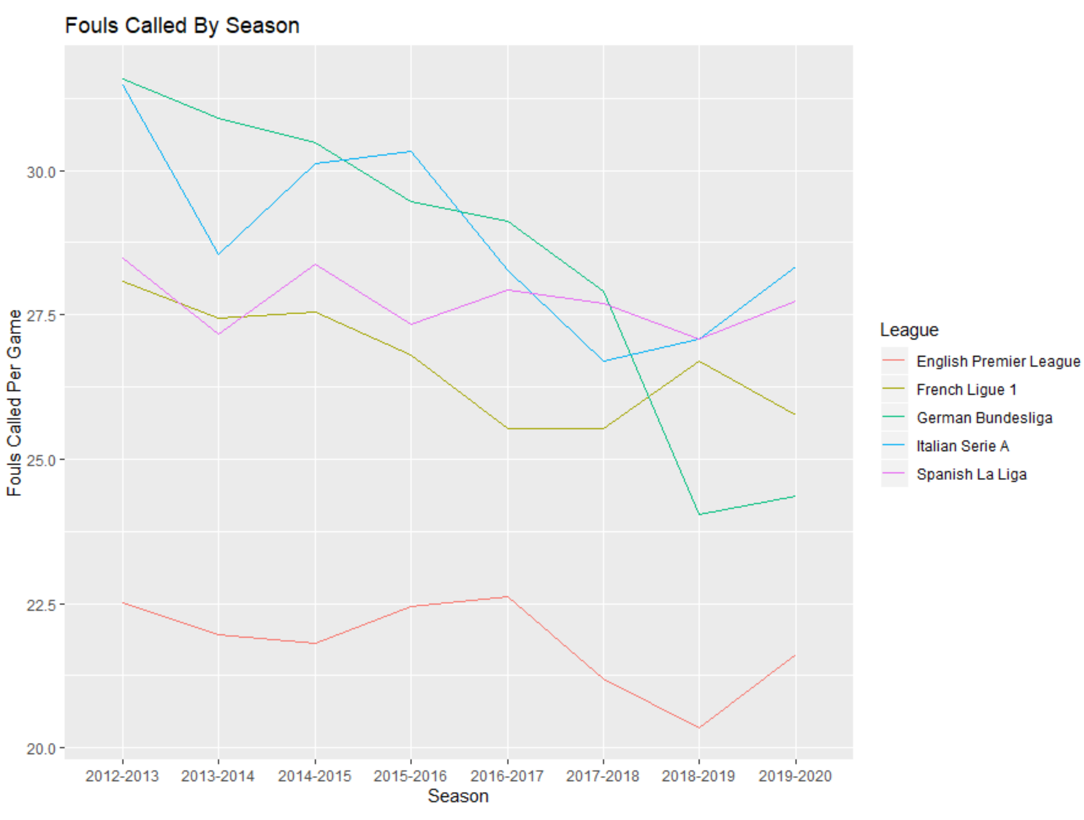
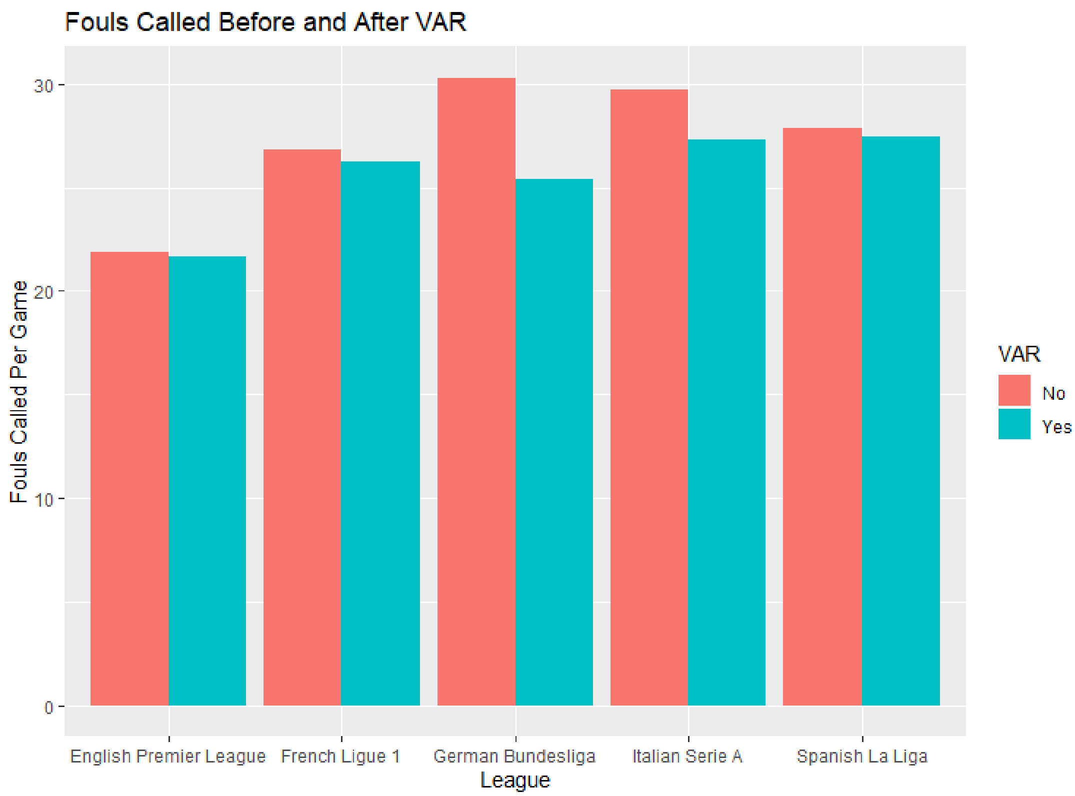


***Figure 5**: Number of fouls per game called before and after the introduction of VAR*



I then compared the number of goals scored per match before and after the introduction of VAR. I wanted to see if the increased leniency amongst referees as well as the ability to check offsides decisions would increase or decrease the number of goals scored per match. I found that there was no significant effect in any league. The number of goals per match does not seem to have a season-to-season pattern and has more to do with individual and team success. For example, the number of goals in the Bundesliga this season has been very high due to the goal-scoring prowess of top players such as Robert Lewandowski, Timo Werner, and Jadon Sancho. Also, the 2016-2017 La Liga saw a jump in goals per match due to the dominance of Barcelona and Real Madrid led by Lionel Messi, Cristiano Ronaldo, and Luis Suarez. I ran a two-sample t-test and found no significant difference between the average number of goals scored before and after the introduction of VAR ($p = .581$), most likely due to the individualistic effect of dominant goal scorers.

Figure 6: *Number of goals per match scored by season*

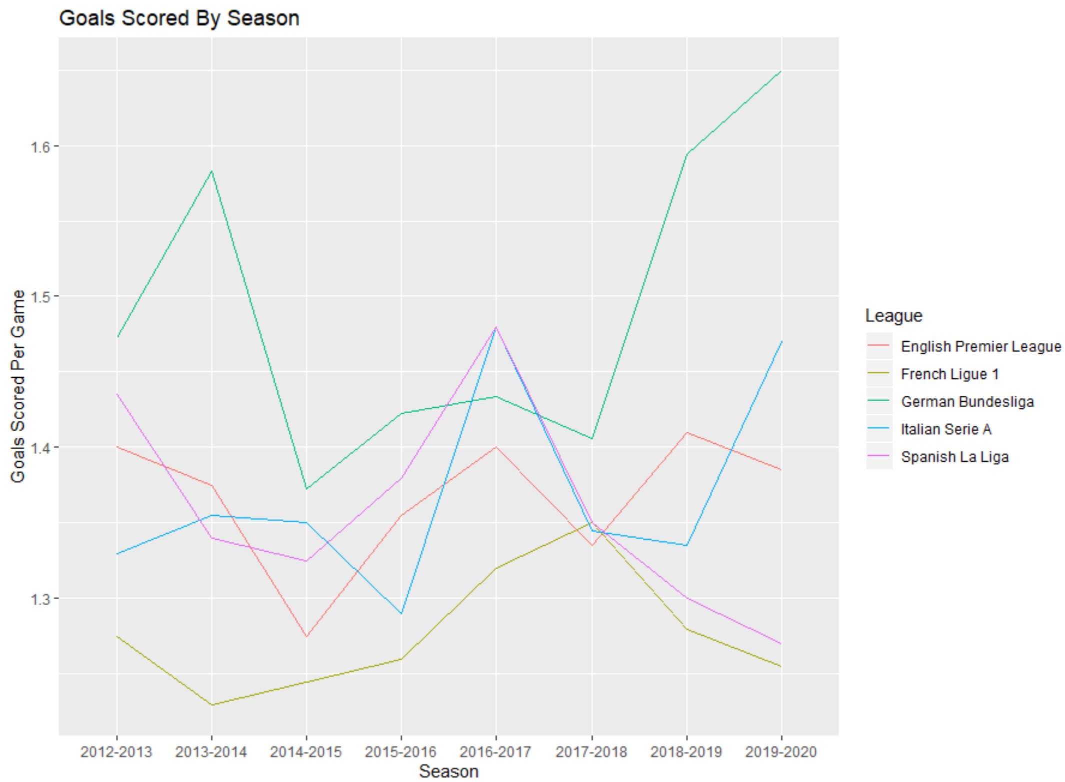


Figure 7: *Number of goals per match scored before and after the introduction of VAR*

Next, I analyzed the number of offsides decisions before and after the introduction of VAR. I found that there were significantly fewer offsides calls made after VAR was introduced. Linesmen were told when VAR was first brought into the game that they should keep their flag down on close offsides decisions and only raise their flag once the play has concluded or a goal has been scored. This was the case because if a player is called offsides, the play is ruled dead and any goal that would have been scored must be counted out regardless of whether or not the player was actually offsides. If the play is allowed to continue and a goal is scored, it can then be checked using VAR. Because of this instruction to the linesmen, fewer offsides calls have been made after the introduction of VAR, indicating the referees are following correct protocols. A two-sample t-test also showed that there were significantly fewer offsides calls made after VAR became a part of professional soccer ($p < .001$).

Figure 8: *Number of offsides calls per game by season*

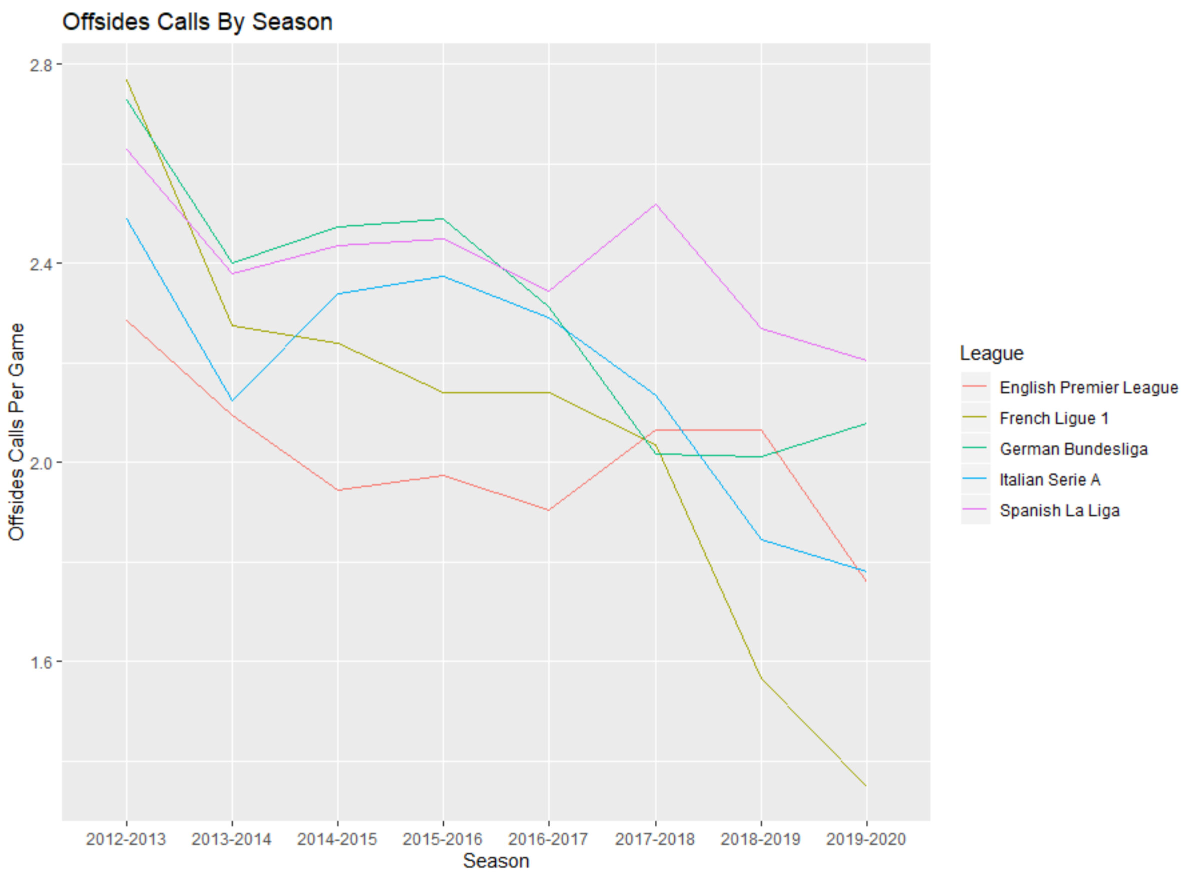
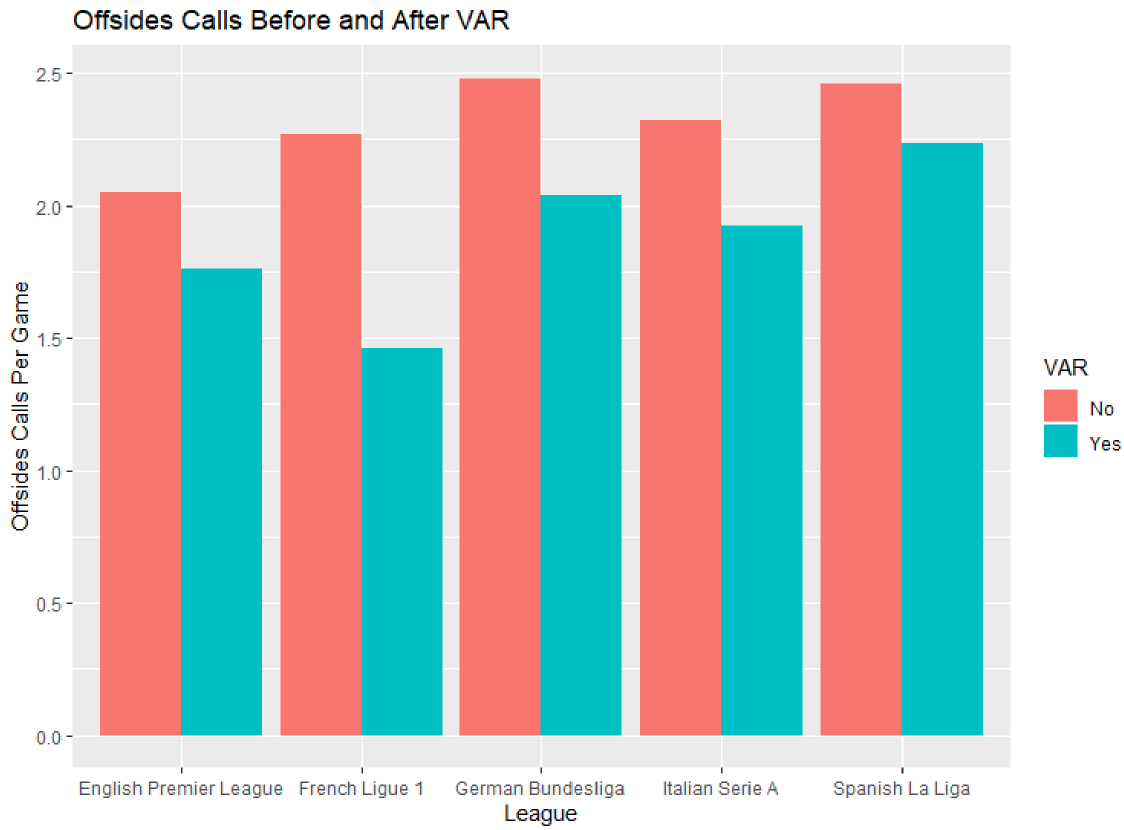


Figure 9: *Number of offsides calls per game before and after the introduction of VAR*



I also compared the number of penalty kicks given over time to see the impact of VAR on spot kicks. I found that the number of penalty kicks has increased after VAR came into play, but the number of penalties had been on the rise even before VAR existed. Referees have been encouraged to call more penalties for the past few seasons, especially when players are jostling for position on set pieces. A two-sample t-test showed that the number of penalties has significantly increased since the introduction of VAR ($p = .028$), but it is tough to say the technology has been the main cause.

Figure 10: *Number of penalties given per match by season*

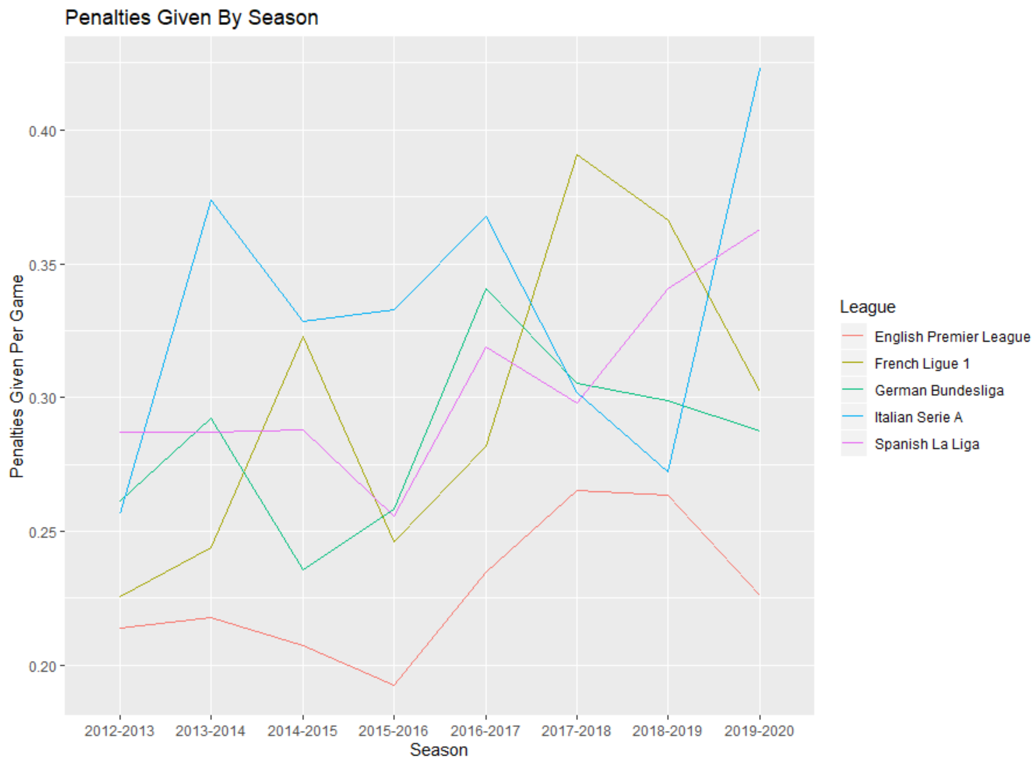
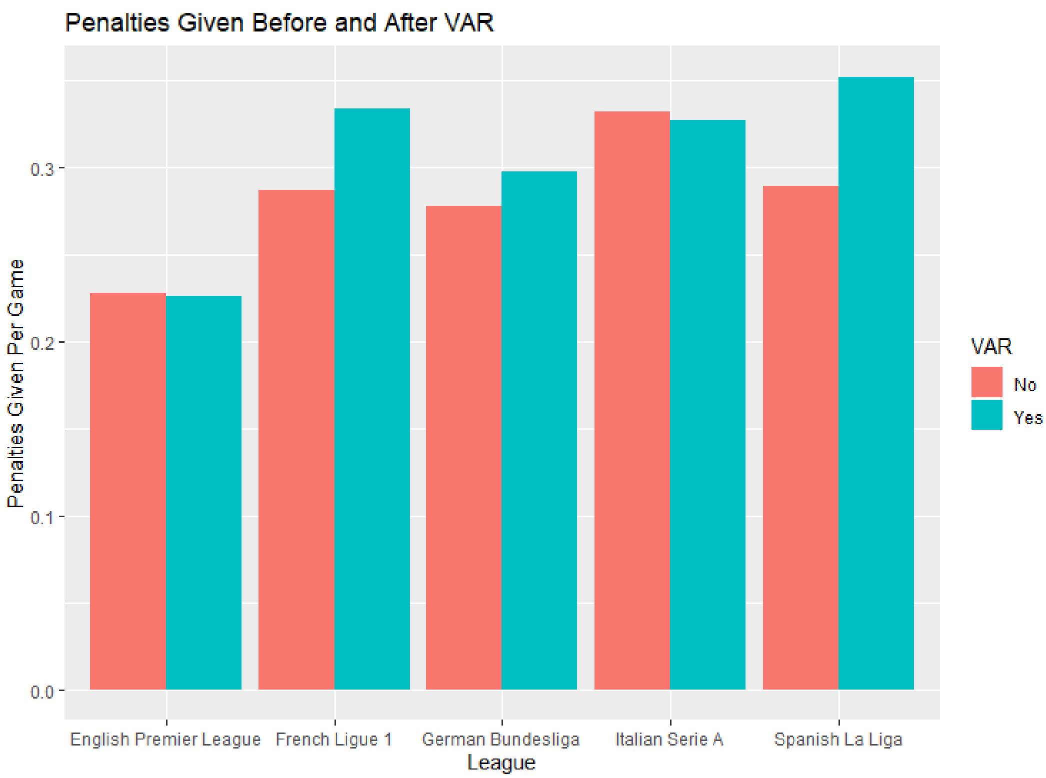


Figure 11: *Number of penalties given per match before and after the introduction of VAR*



I also wanted to see if the number of yellow and red cards given during matches was also impacted by the introduction of VAR. The number of yellow cards shown during matches was affected in different manners for the different leagues in Europe. In the Bundesliga, Serie A, and La Liga, the number of yellow cards went down as a result of VAR, but the number of yellows given was already quite high in these leagues to begin with. In the Premier League and Ligue 1, the number of yellows increased slightly. The number of yellow cards seems to be more affected by the nature of the league rather than VAR. For example, the La Liga is known for its tactical fouling and harsher refereeing which is why there are more yellow cards given each season. On the other hand, the Premier League is known for its lenient refereeing and tends to have the fewest yellow cards each season. A two-sample t-test showed that the number of yellow cards did not change significantly after the introduction of VAR ($p = .324$). It does seem that the number of red cards given has decreased since VAR has been utilized. Red card offenses can be reviewed by VAR, so referees do not have to be as quick to show a red card than they were before VAR existed. The referee may call just a common foul and only the most serious incidents would be considered red card offenses by VAR. This suggests that referees gave a number of red cards before the implementation of VAR that should not have been red cards. A two-sample t-test showed that the number of red cards did decrease significantly after the introduction of VAR ($p = .021$).

Figure 12: *Yellow cards shown per game by season*

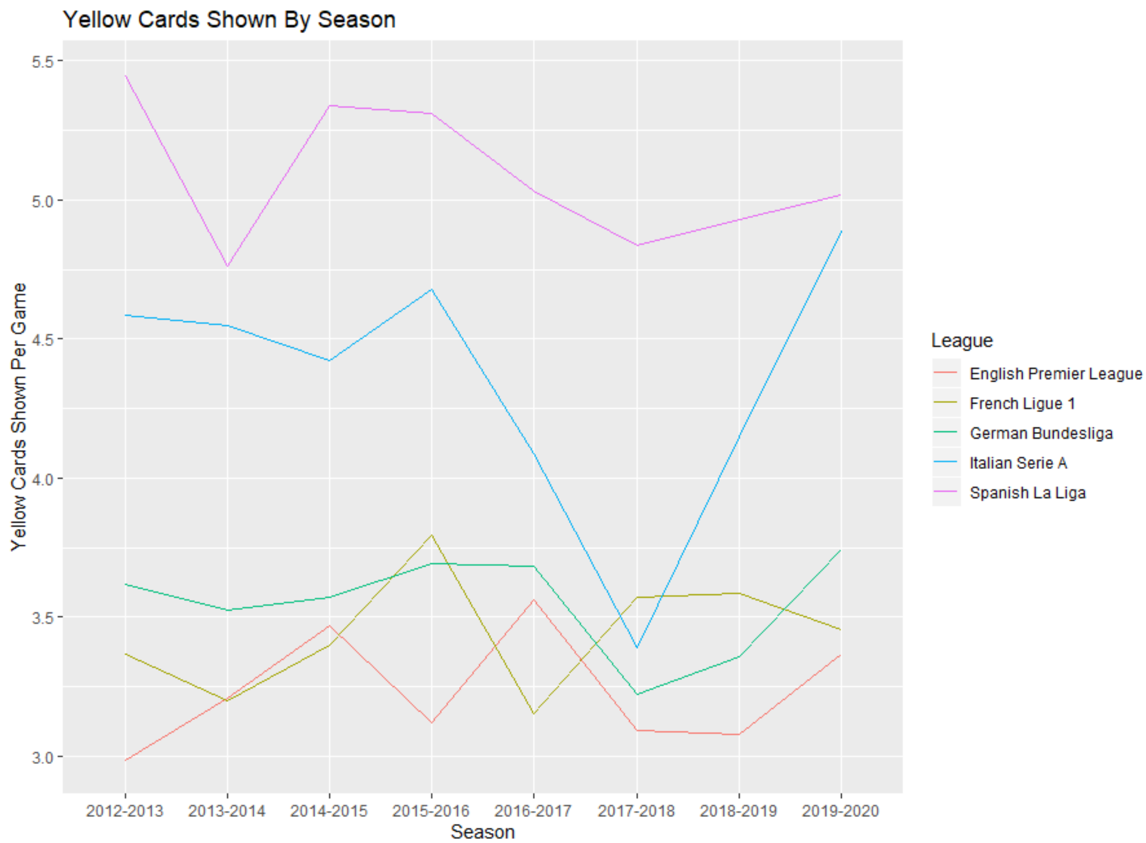


Figure 13: *Yellow cards shown per game before and after the introduction of VAR*

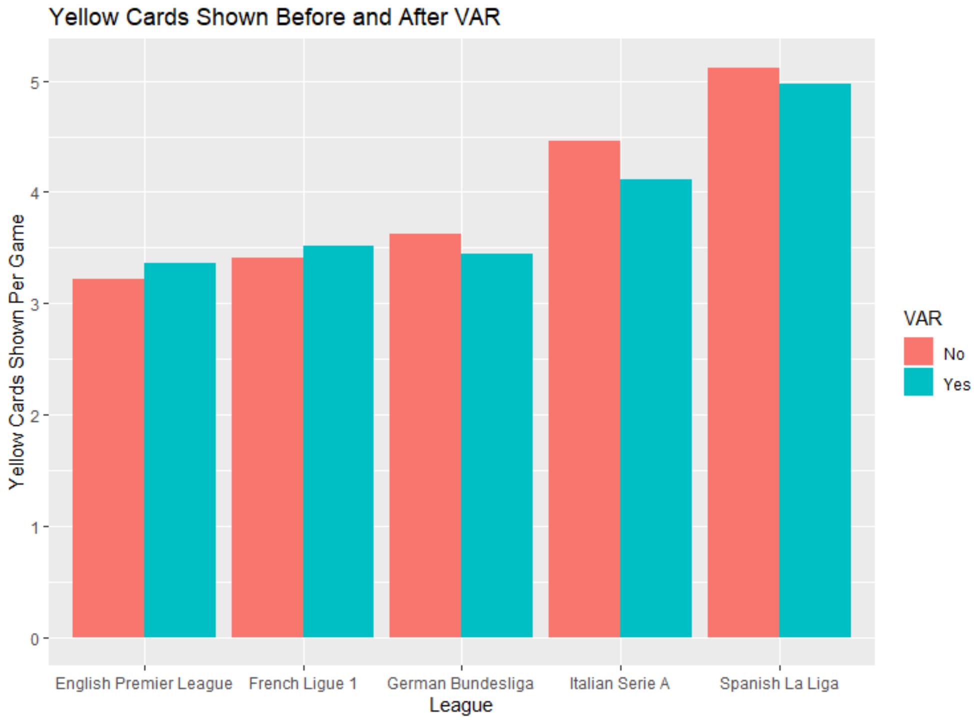


Figure 14: *Red cards shown per game by season*

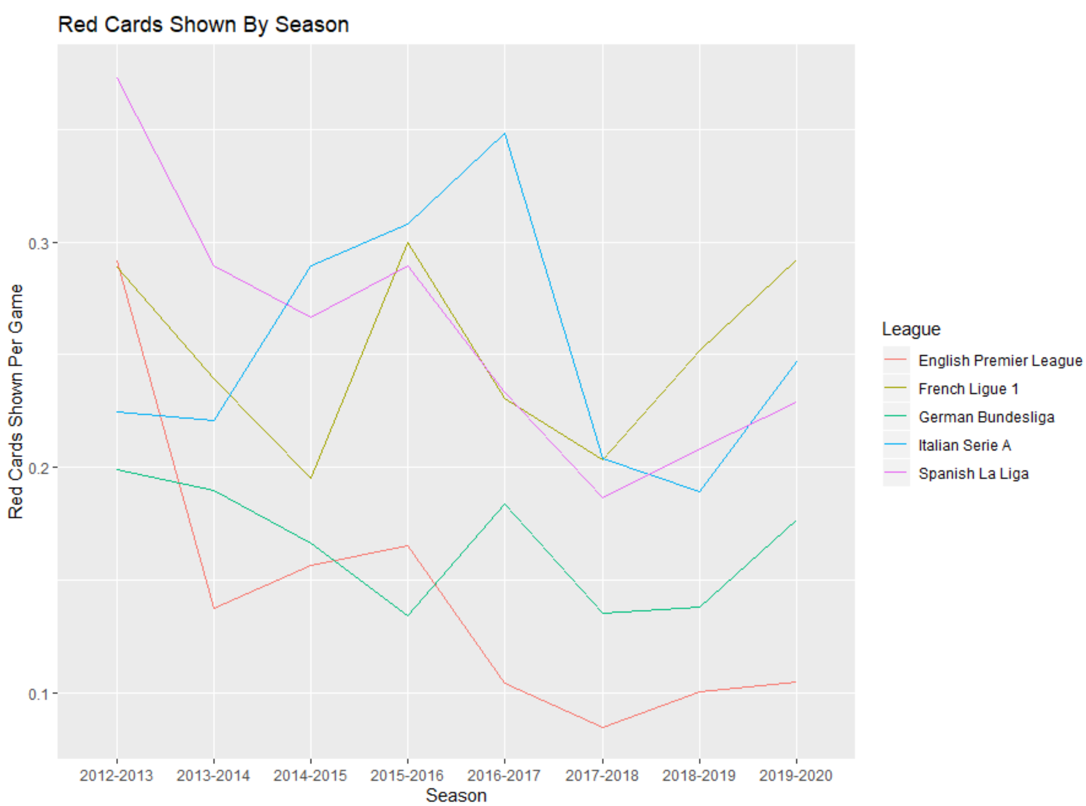
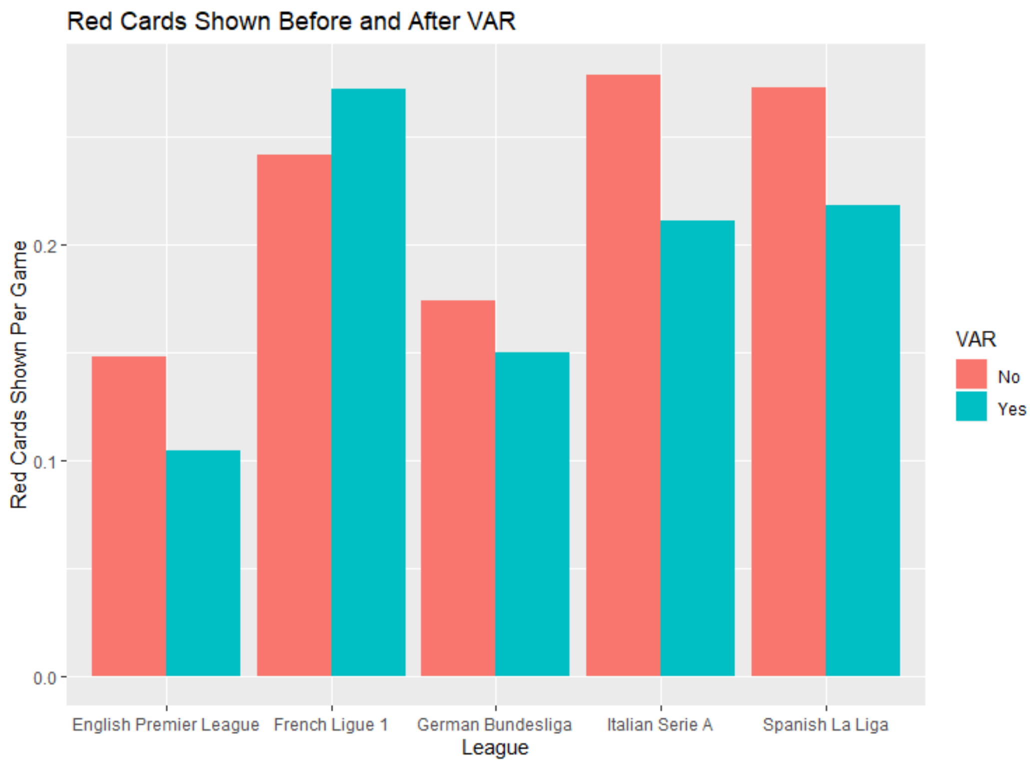


Figure 15: *Red cards shown per game before and after the introduction of VAR*



Finally, I wanted to analyze the amount of stoppage time in matches before and after the implementation of VAR. For matches after the introduction of VAR, the amount of stoppage time in the first half, second half, and full match all increased significantly. The biggest drawback to using VAR is that the match must be stopped to adequately analyze the decision to come up with the correct call. Leagues across Europe have struggled to utilize VAR without adding more stoppage time to matches. A two-sample t-test showed that the amount of stoppage time in the first and second halves as well as the total match time has increased after the introduction of VAR ($p < .001$ for each test). It is clear that VAR has impacted refereeing and the pace of play of professional soccer.

Figure 16: *First half stoppage time per game by season*

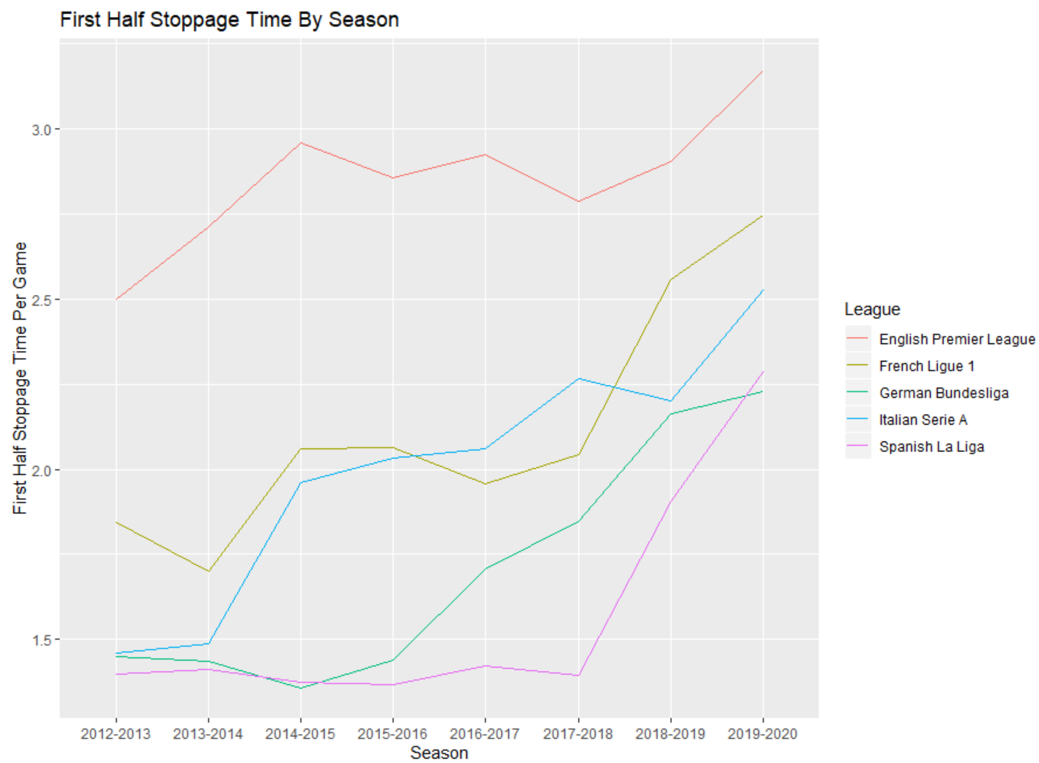


Figure 17: *First half stoppage time per game before and after the introduction of VAR*

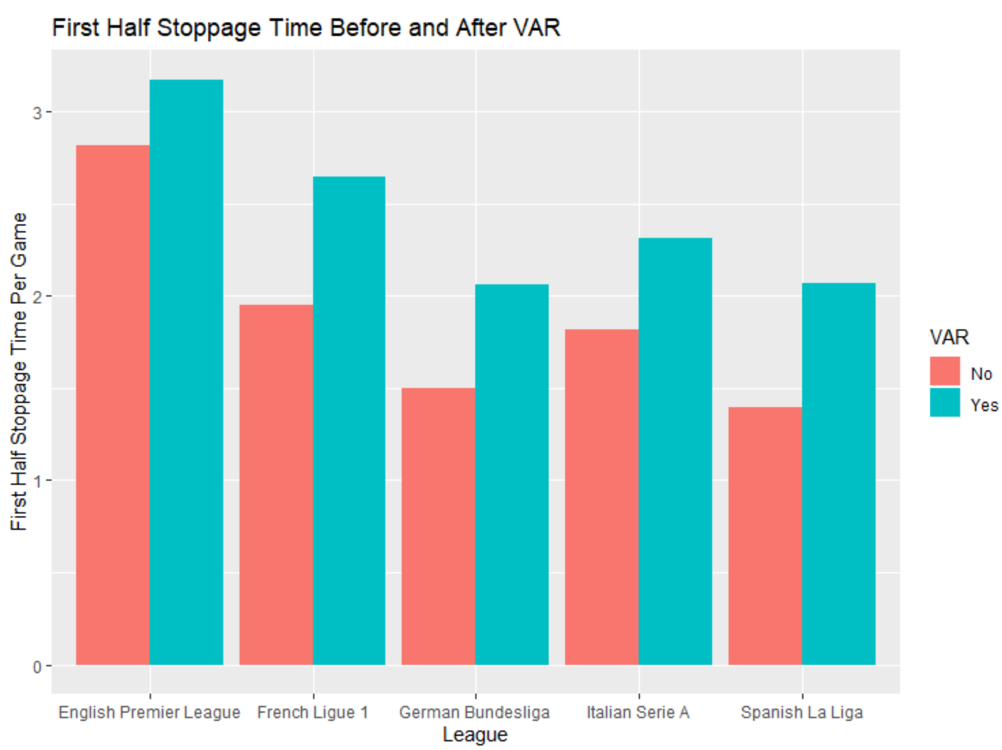


Figure 18: *Second half stoppage time per game by season*

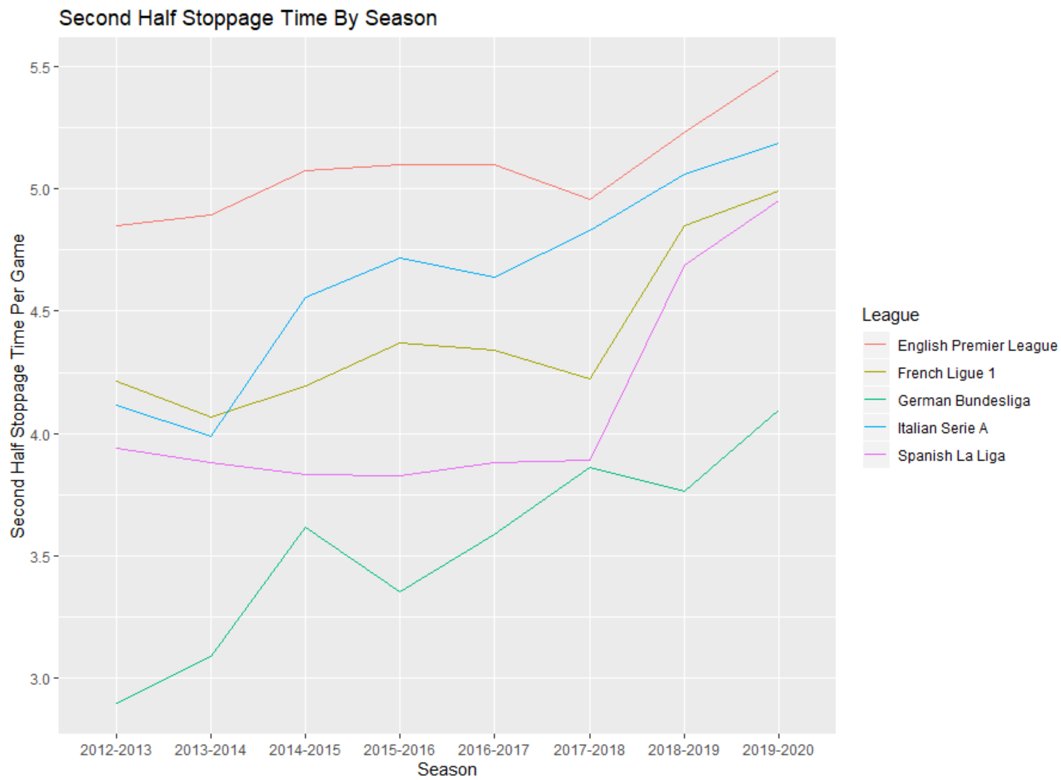


Figure 19: *Second half stoppage time per game before and after the introduction of VAR*

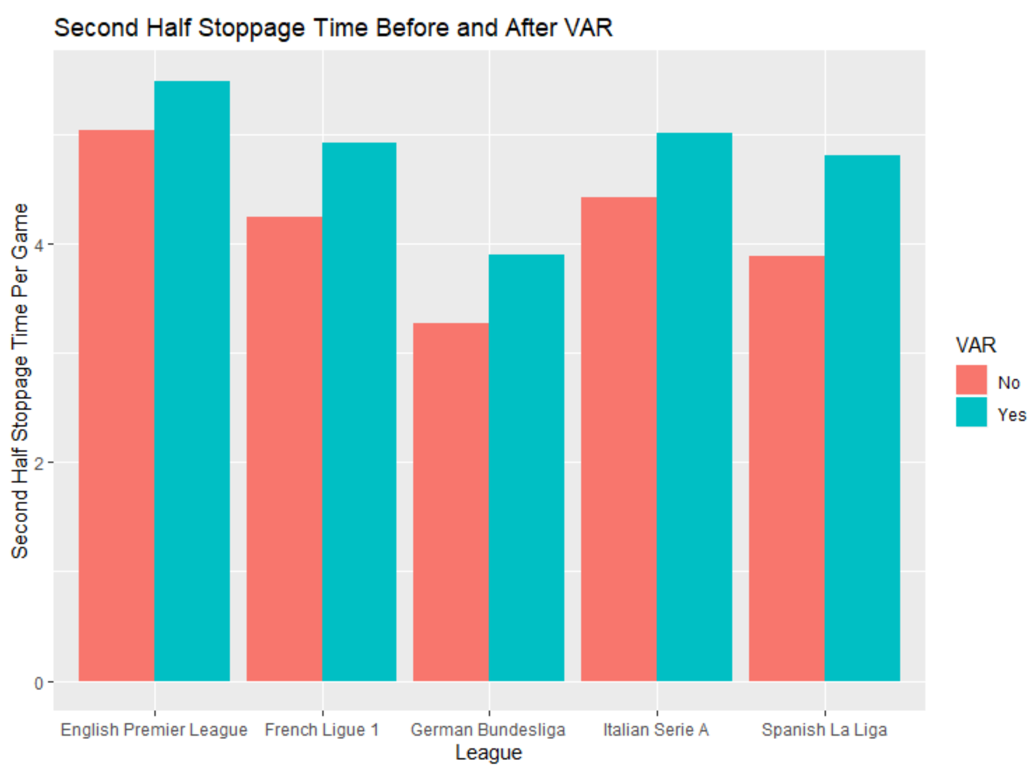


Figure 20: *Full match stoppage time per game by season*

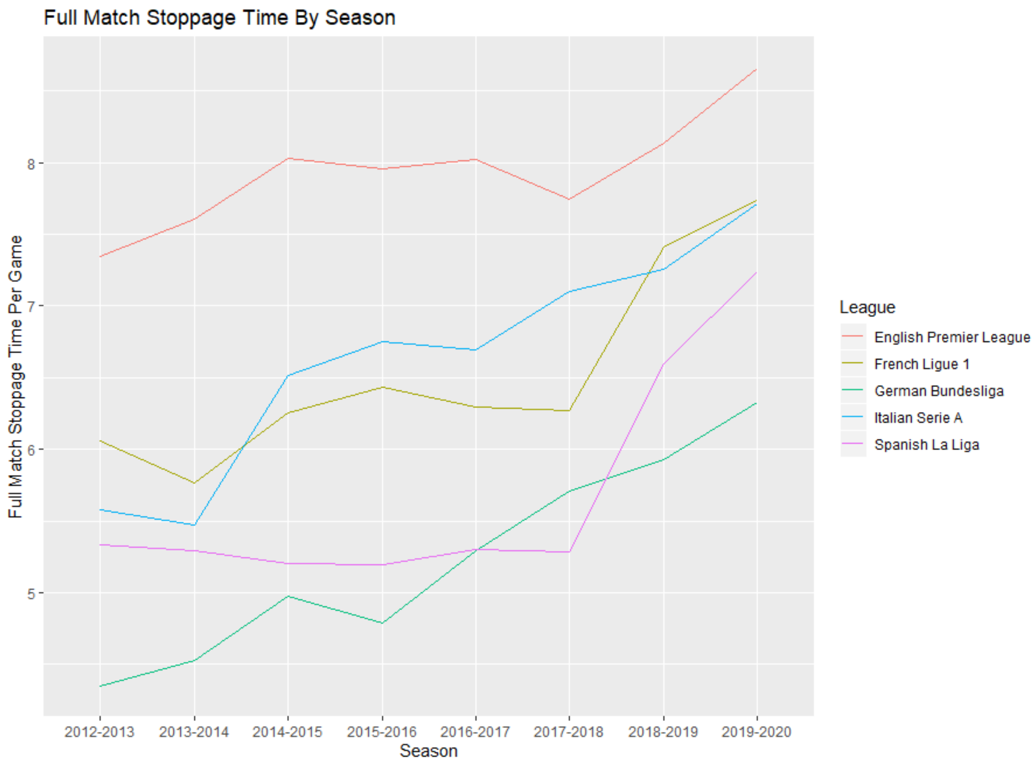
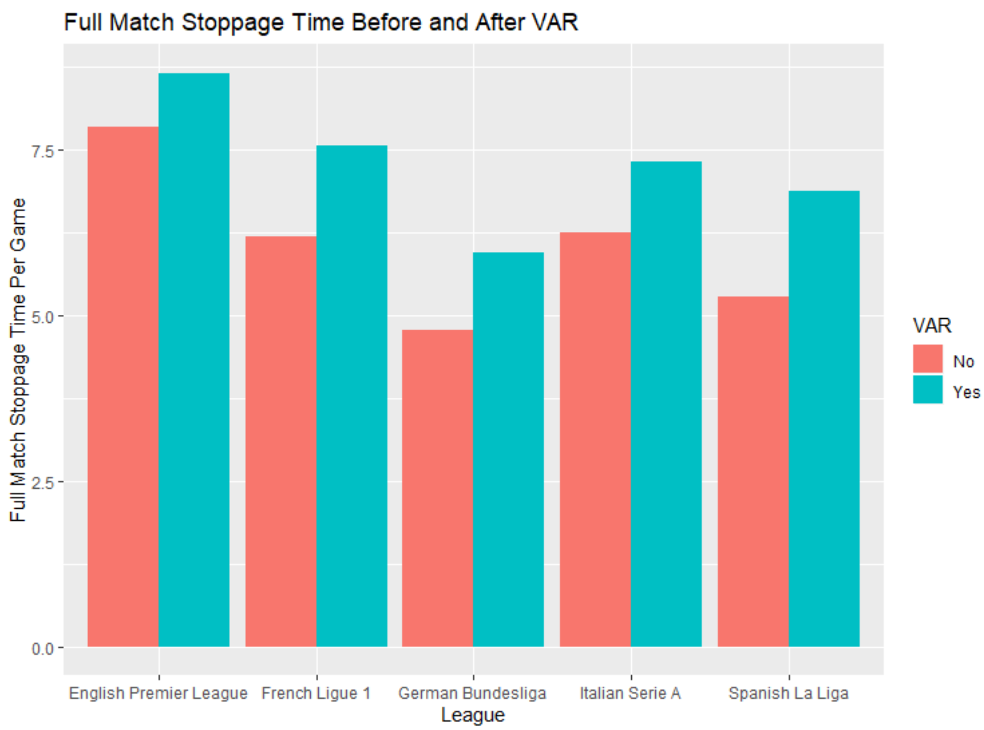


Figure 21: *Full match stoppage time per game before and after the introduction of VAR*

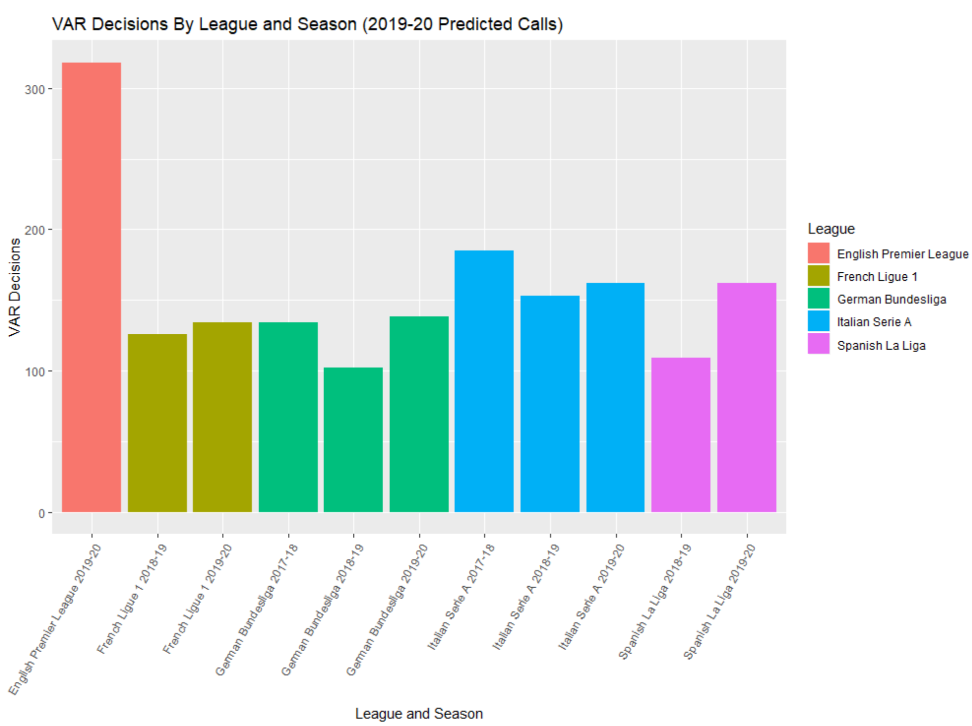


Potential Biases of VAR

After analyzing how VAR has impacted the officiating and pace of play in professional soccer, I wanted to further look into the individual VAR decisions made in matches over the past three seasons. Data was collected from ESPN commentary data which indicates any VAR decision that was made during a match. As was mentioned previously, ESPN only specifies VAR decisions that led to a significant stoppage in the match, so the actual number of VAR checks is larger than the number provided by ESPN. Key information for each VAR decision was included in the commentary data such as the type of event (goal, penalty kick, red card, and mistaken identity), whether the ruling on the field was upheld or reversed, and the point in the match when the call was made. I also included other important information such as the team that benefitted from the VAR decision, the head referee of the match, and the pregame win-loss record of the teams involved in the match. Each of these variables helped to analyze potential biases in terms of VAR decisions made by referees.

I first calculated the number of VAR calls made in each of the top five European leagues to compare the use of the technology in each country. I included predicted VAR calls for the 2019-2020 season by simply doubling the number of VAR calls made through the halfway point of the year. It is clear to see that the English Premier League has utilized the VAR technology more than the other leagues in Europe. The Premier League had 158 VAR calls through the halfway point of the season, which was more than every full season of VAR other than the 185 call 2017-2018 season of the Serie A. This large discrepancy may be due to reporting bias with many more VAR decisions being recorded in the Premier League even though each league has a similar number of calls. I believe this explains part of the large difference in calls, but I also think the Premier League has made it a point to utilize VAR more than other leagues as the technology continues to improve.

Figure 22: *VAR decisions by league including predicted VAR calls for the 2019-2020 season*



I then grouped the VAR decisions by the type of call to see which event requires a second-look more than others. VAR decisions regarding a penalty kick were reviewed the most frequently, followed by goals and red cards. Cases of mistaken identity are very rare and only two such scenarios occurred in the past three seasons. Penalty kicks very often require VAR assistance as tackles made in the box may be difficult to judge as fouls during the run of play. Handballs in the penalty area have also become a topic of discussion and it is hard to tell if a player's hand is in an "unnatural position" that would lead to a penalty kick. Goals are also commonly reviewed at length with VAR and, much like professional American football, every scoring play is looked at a second time before the match is restarted. Many of the goal checks involve an offsides decision which can be very tight and tough to spot without the use of technology. I also analyzed how frequently VAR calls stood and how frequently the call was cancelled. Approximately 70% of VAR calls were upheld, which is higher than the 63% of calls that have been upheld in the NFL between 1999-2016[21]. It seems that referees in soccer have been more accurate than their counterparts in American football. This may be explained by the fact that only very specific events can be changed in soccer while more can be changed in football. It also may be explained by the fact that early on in the use of VAR, more plays were prematurely called dead before the technology could be utilized. As the use of VAR continues to evolve over the years, more calls will be overturned just as it did in the NFL which had an overturn rate of just 13% in the late 1980s.

Figure 23: *VAR decisions by event in the top five European leagues*

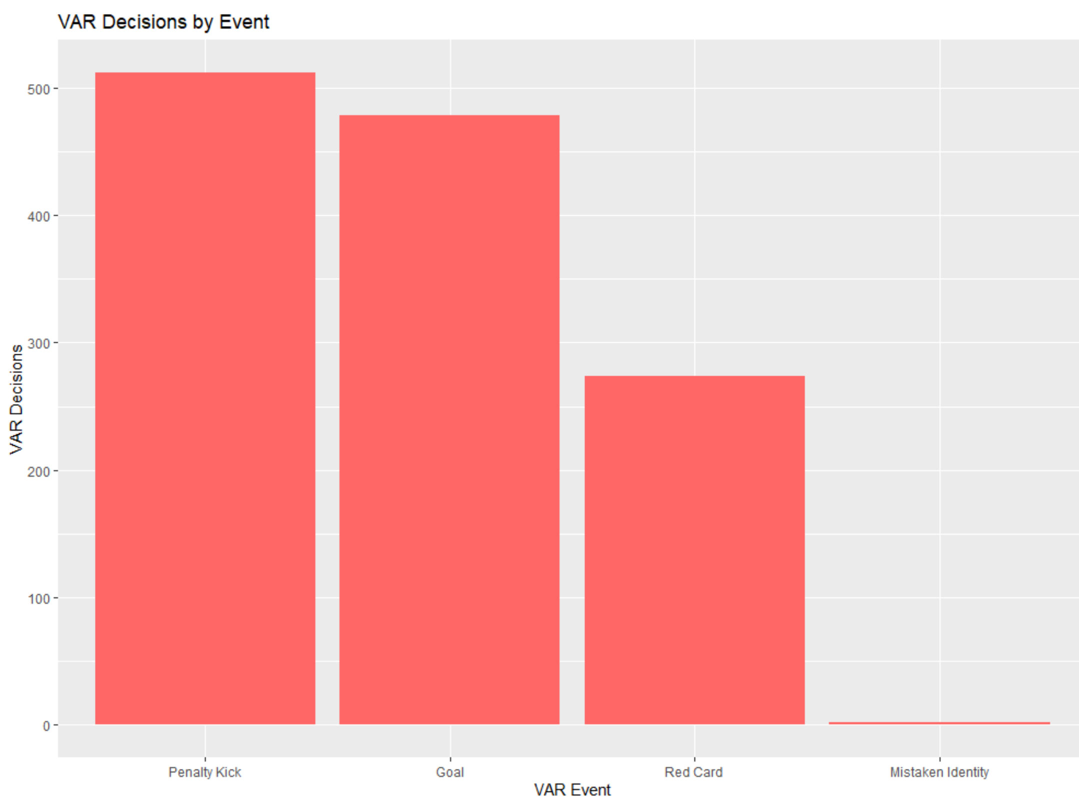
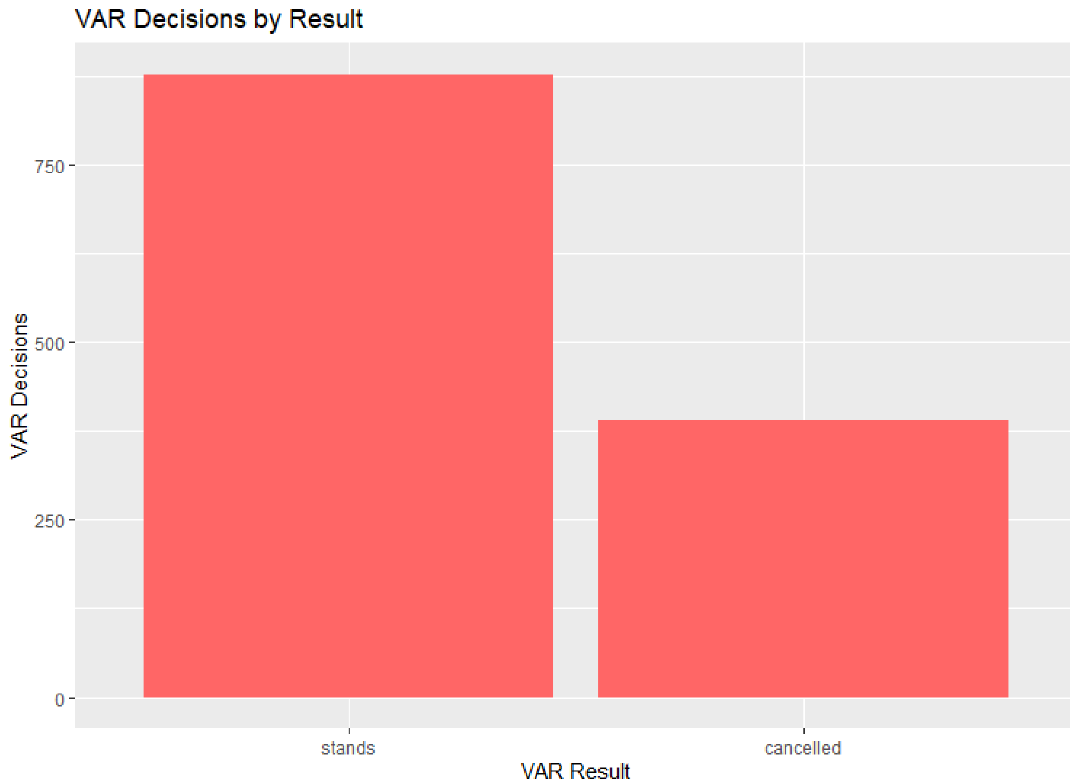


Figure 24: *VAR decisions by the end result of the call*



I hypothesized that the number of VAR calls would be greater for the home team than the away team and the calls would tend to favor the home team as well. I found that 61% of VAR calls involved the home team which is not a surprise. This is the case because the home team has a distinct competitive advantage and is more likely to be involved in goal-scoring opportunities that would lead to VAR decisions. If there was no home team bias for VAR, it would be expected that the home team would not benefit from calls any more than the away team would. The home team benefitted from 51% of the VAR calls made, showing that there is very little home-field bias. The small discrepancy can be explained by the fact that the home team does not only have a home-field advantage in terms of fan support and familiarity of the grounds, but also in how the fans can influence a referee's decision-making. Fans are quick to voice their displeasure towards an official after a suspect call, and referees may be hesitant to make a call against the home team as a result. VAR calls are of upmost importance in a match as they involve goal-scoring opportunities, penalty kicks that frequently lead to goals, and red card scenarios that can put a team at a severe disadvantage for the remainder of a game. Referees for the most part have done a good job in eliminating this possible bias.

Figure 25: *VAR decisions for the home and away team*

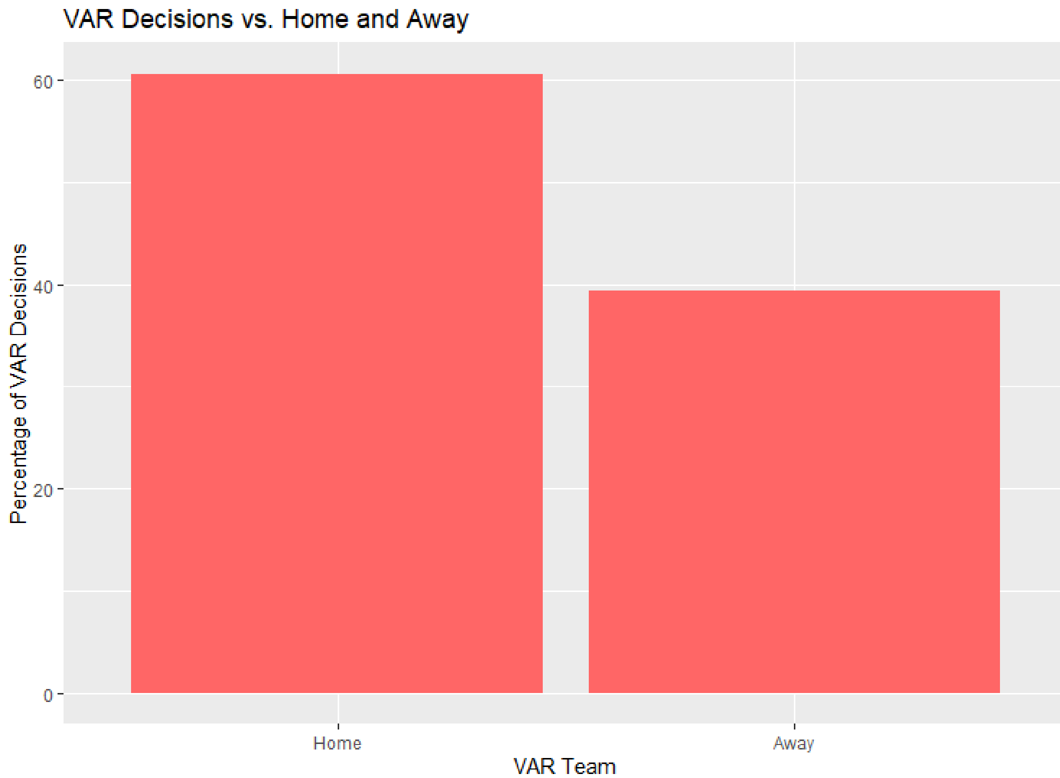
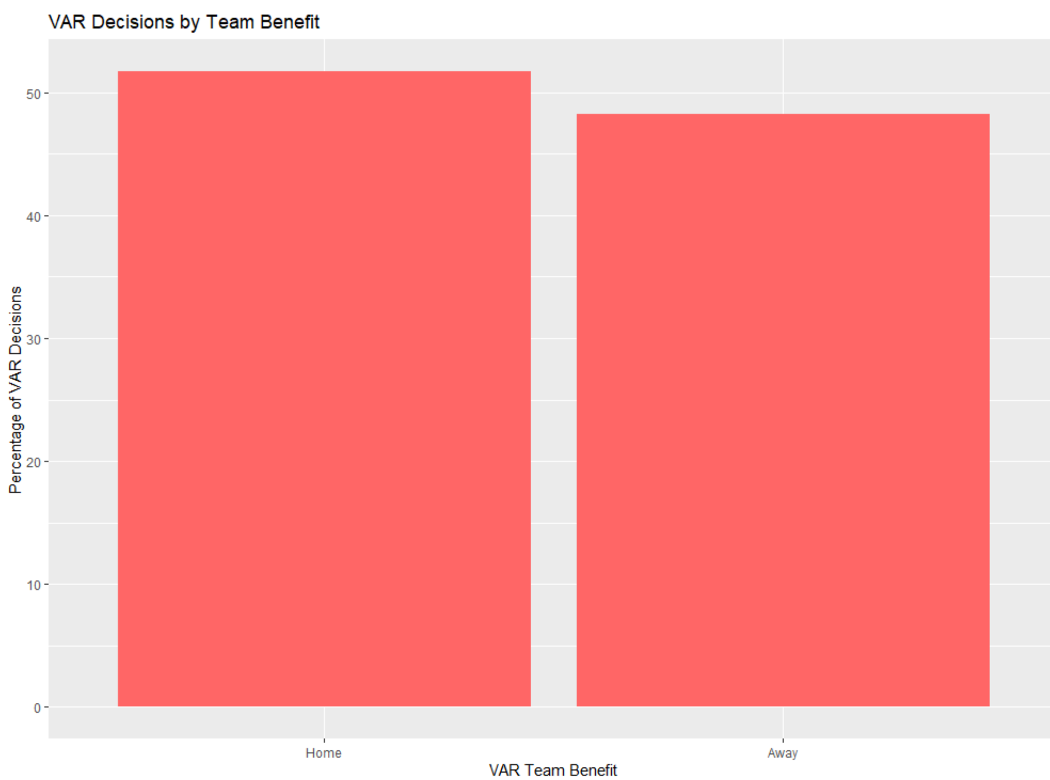


Figure 26: *VAR decisions that benefitted the home and away team*



I next wanted to take a look at individual teams involved with VAR decisions to see if any patterns emerged. I found that even though the Premier League in 2019-2020 had a very large number of VAR decisions, only Aston Villa and Manchester United were amongst the top 15 in terms of VAR calls involved in. Many of the teams involved in the most VAR decisions were Serie A teams that finished in the top half of the table during the season. AC Milan, Inter Milan, Lazio, Atalanta, and Napoli all finished in the top seven during the 2017-2018 Serie A season, suggesting that better performing teams are involved in VAR decisions more frequently than underperforming teams. This is not surprising as the more successful teams are more likely to get themselves in goal-scoring scenarios that require the use of VAR. When looking at patterns among team who benefitted from VAR decisions, no discernible pattern can be seen. Relatively few of the teams that were involved in many VAR decisions benefitted from them and new teams such as Tottenham and Wolves led the way. It seems that more successful teams are involved in VAR situations more frequently, but that does not necessarily lead to a greater benefit from VAR.

Figure 27: *VAR decisions by team involved*

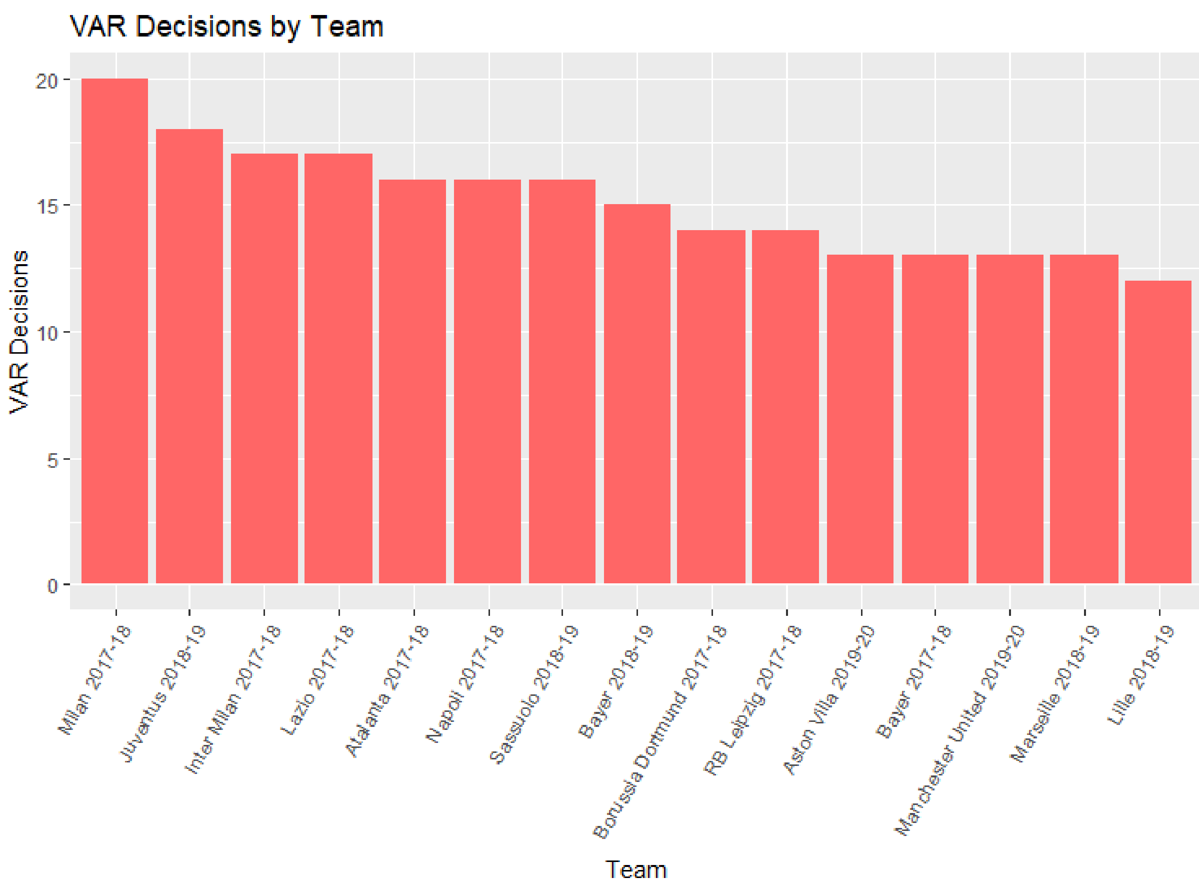
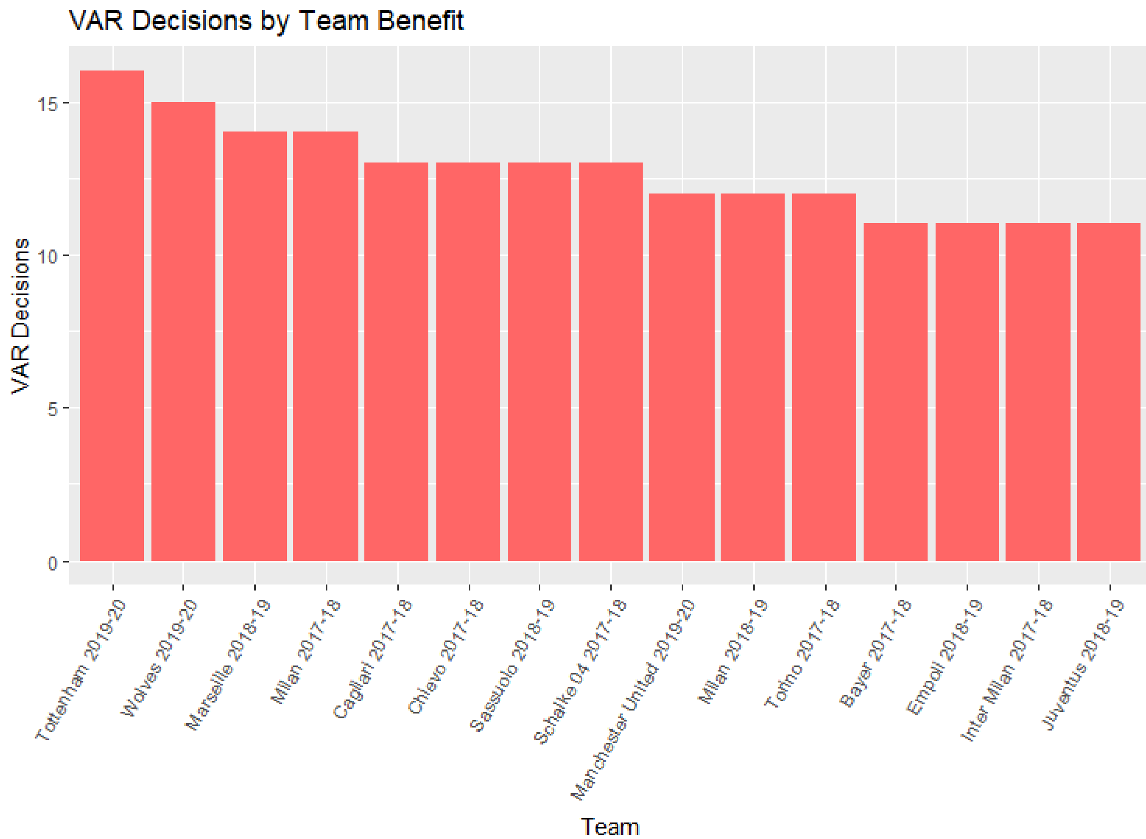
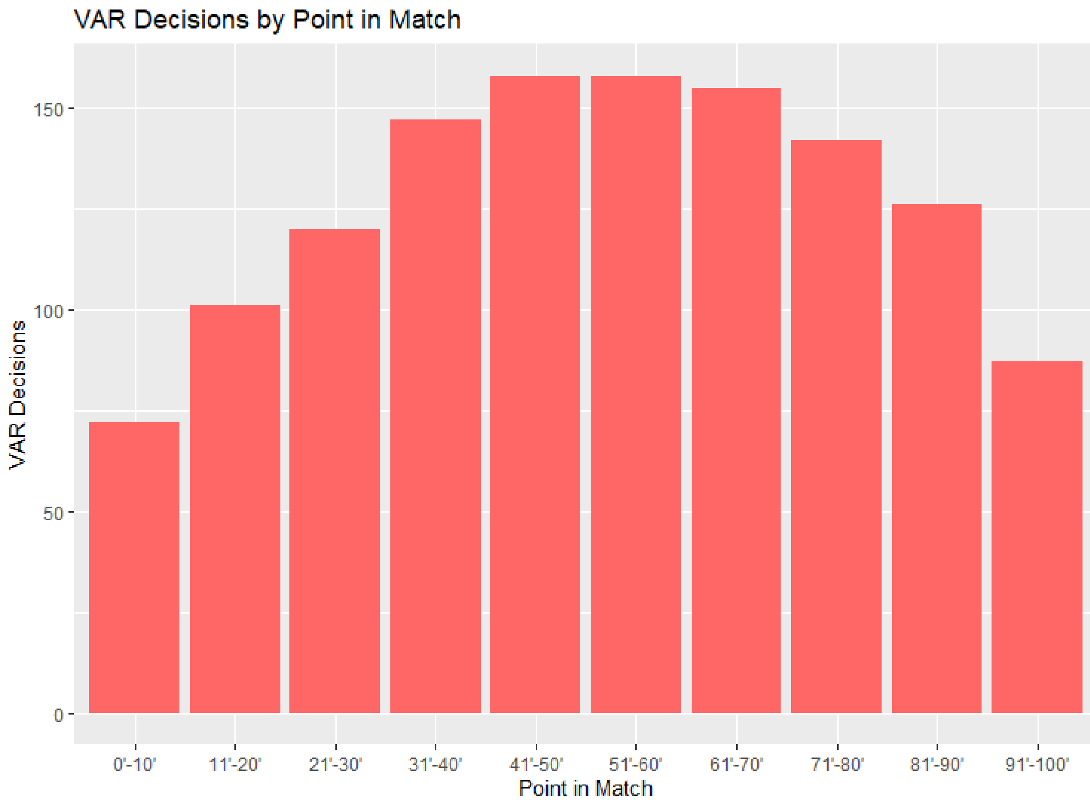


Figure 28: *VAR decisions by team benefitting from the call*



I also analyzed the VAR decisions by grouping them by minute of the match. I created ten groups of ten-minute periods and counted the number of VAR decisions in each bin. I found that the distribution of calls was relatively normal with the largest proportion of decisions being made between the 41^{st} and 60^{th} minutes of the match. Slightly more calls are made in the second half with 56% of VAR decisions being made after the break. There are two reasons why this is the case. First, players tend to use the earlier portion of the match to feel the opposition out and play a more possession-based game during this timeframe. Fewer goals and goal-scoring opportunities occur in the early portion of the game which would also lead to fewer VAR decisions. Second, referees are more cautious about making game-changing decisions during the first few minutes of a match as they can have a significant impact on the rest of the game. As a result, most calls are made in the first 15 minutes of the second half after teams make adjustments and play a more attacking style of soccer.

Figure 29: *VAR decisions by minute of match*



I also wanted to see if there was any individual referee bias in the use of VAR. I collected referee data from Transfermarkt for each match that utilized VAR in the top five European leagues. I found that all of the referees that had many VAR decisions per match were in the Premier League or Serie A, which were the two leagues that utilized the technology the most. The only referee to have over one VAR decision per match was Martin Atkinson in the Premier League. Atkinson has been at the center of several controversial decisions over the course of his lengthy career, including a VAR decision this season in a fixture between Manchester United and Liverpool[22]. Marcus Rashford scored the only goal for Manchester United in a 1-1 draw which VAR saw to be the right decision after Liverpool's Divock Origi was brought down in the build-up. Liverpool manager Jürgen Klopp criticized Atkinson's role in the call saying "this isn't a VAR decision; it is Martin Atkinson's decision." This raises an interesting point because the head official isn't directly involved with the final decision-making process with VAR, but the head referee does decide whether or not to look at a play. In future research, I would like to analyze the number of VAR calls made by individual VAR officials, but it is interesting to see the role a head referee has in making a decision. Also, many of the referees who made the fewest number of VAR calls per match were from the Bundesliga and La Liga: two leagues with fewer VAR decisions per match.

Figure 30: *Top 10 referees in terms of VAR decisions per match*

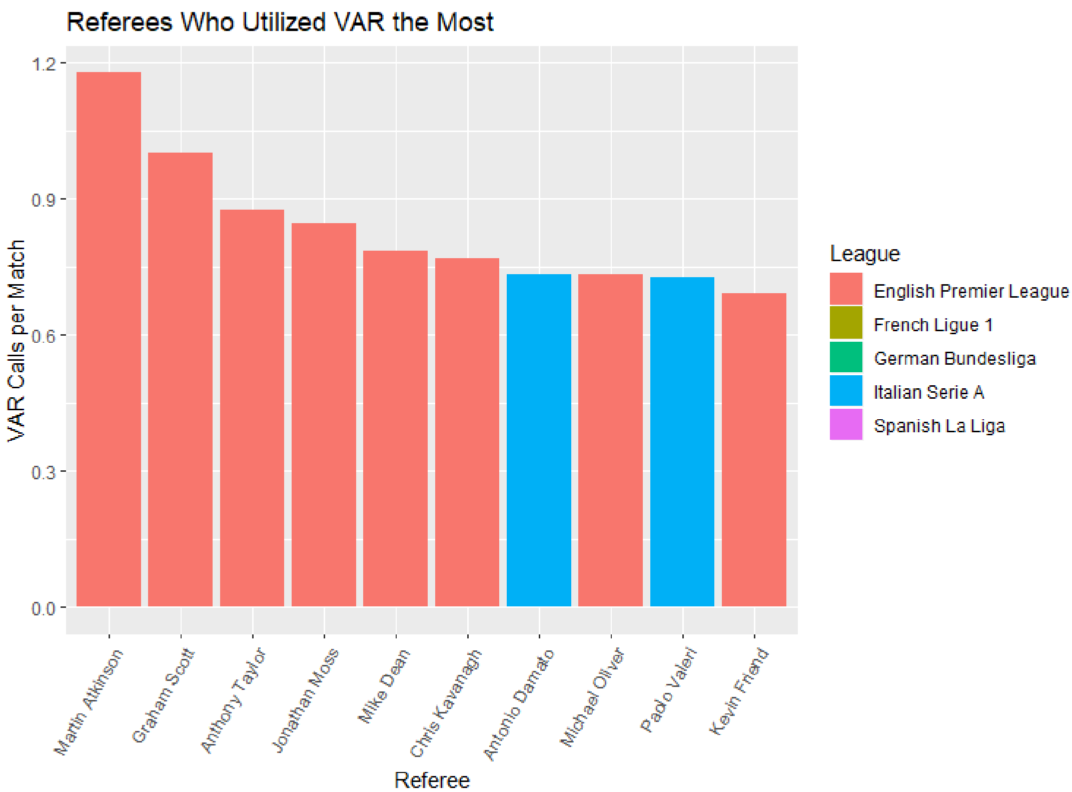
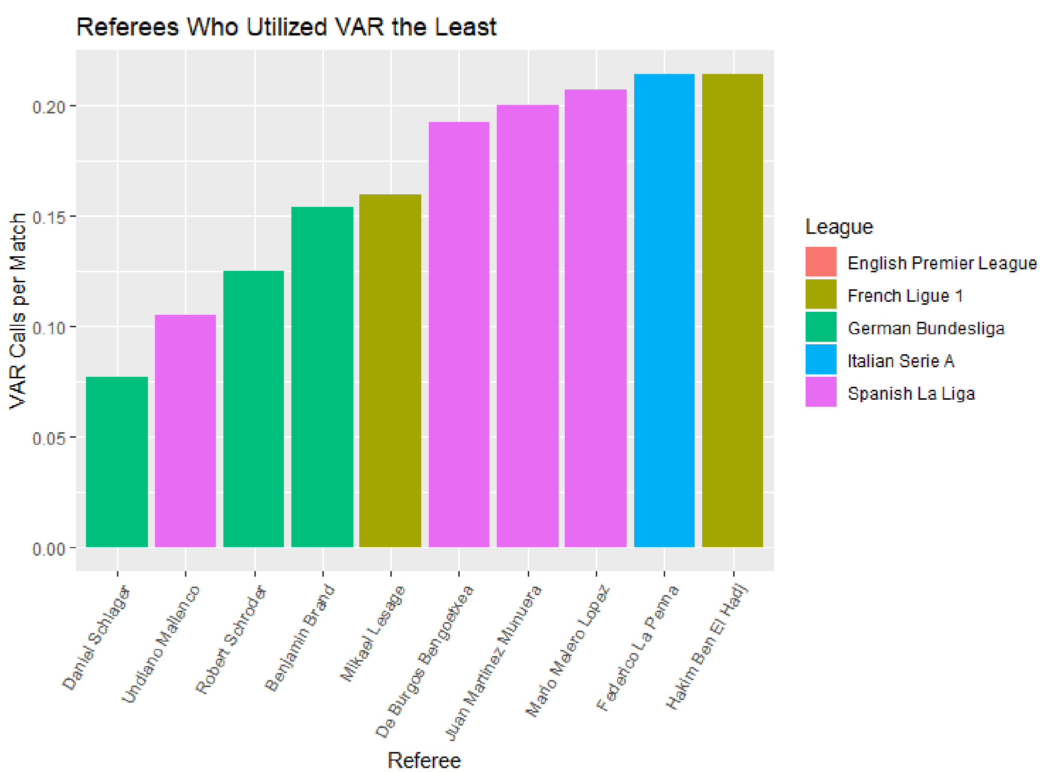


Figure 31: *Bottom 10 referees in terms of VAR decisions per match*



Finally, I wanted to see if the winning or better team benefitted more from VAR decisions. I found that in matches where the score was not tied, the team that was winning tended to benefit more from VAR decisions than the losing team. This may be the case because the winning team was playing better and found themselves with more goal-scoring opportunities than the losing team. It may also be the case that in tight decisions, the referee favored the team currently winning the match as they would be most likely to go on and win the match regardless. There might be less controversy from giving a call to the winning team as the fans of the losing team would feel that they were going to lose even if the call had gone their way. I found a similar effect when analyzing the end of season records as the better team benefitted more than the worse team. This shows that season performance has a much bigger impact on a VAR decision than current match performance. This makes sense because a team could have an impressive performance any given weekend, but the team that is usually better is more likely to be more affected by a wrong VAR decision. A bad call could have serious title implications and referees seems to err on the side of caution when it comes to making a VAR call in the favor of the worse team. Overall, there are several additional factors that can impact a VAR decision other than the scenario that led to the call itself. Referees and VAR officials should work to eliminate these biases to improve the video replay system even more.

Figure 32: *VAR decision benefit by the current match standing (winning or losing)*

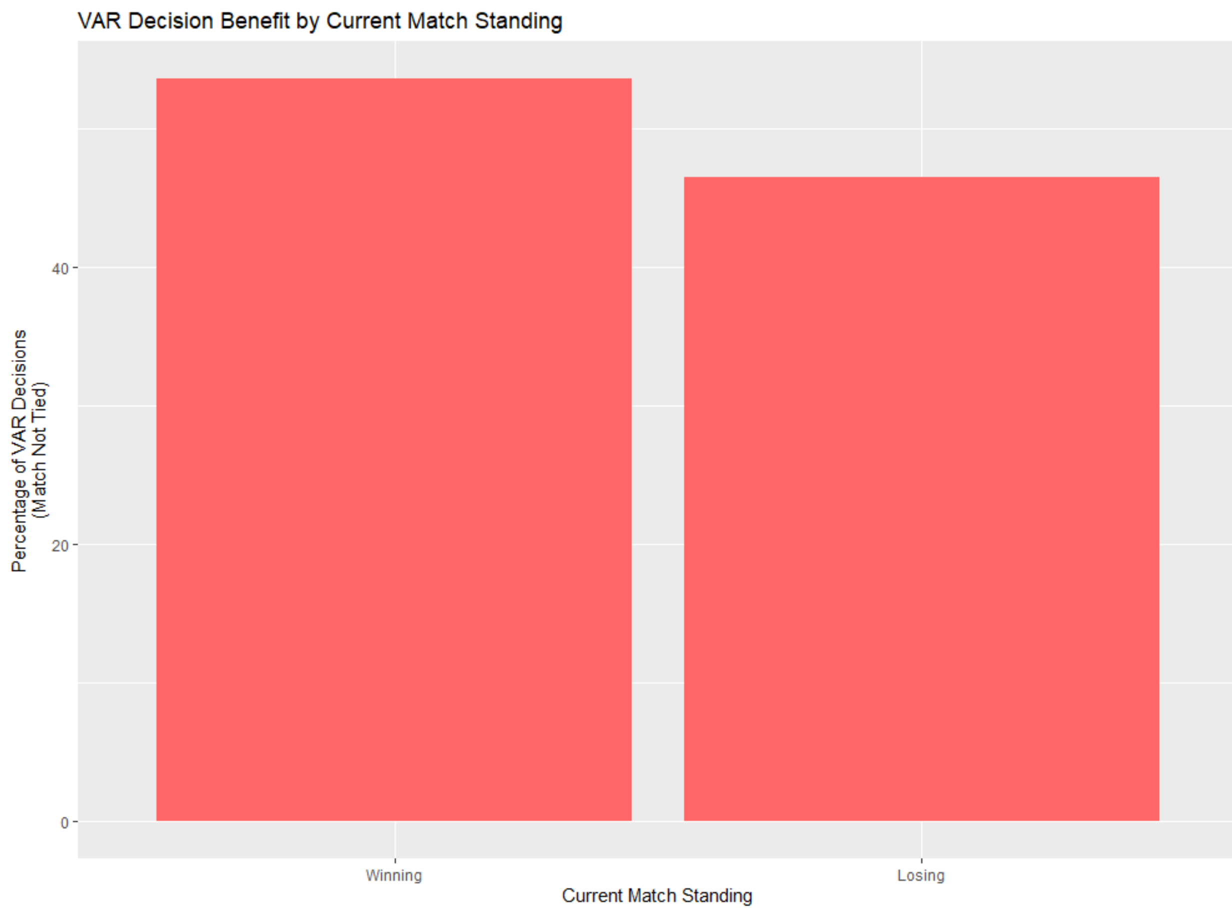
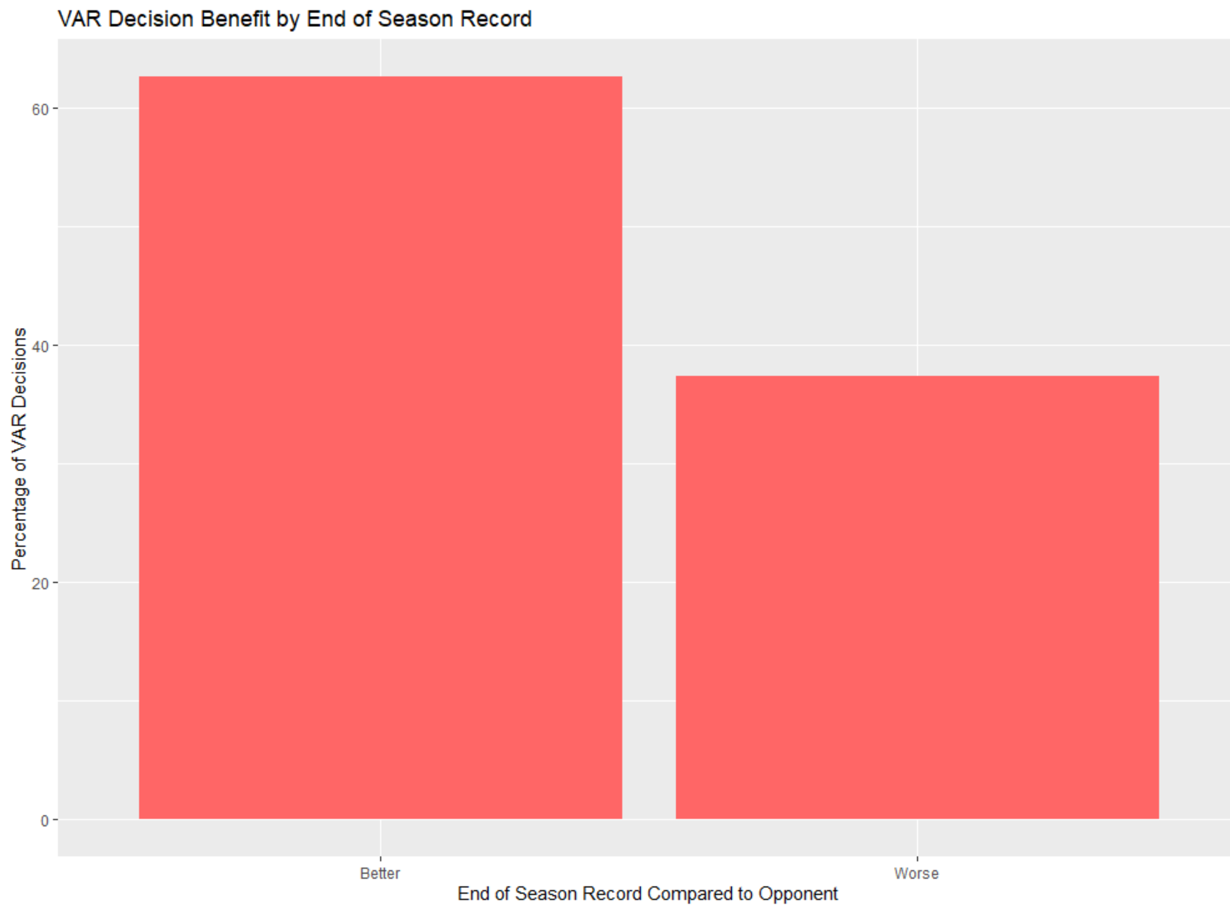


Figure 33: *VAR decision benefit by end of season record*



Win Probability Model

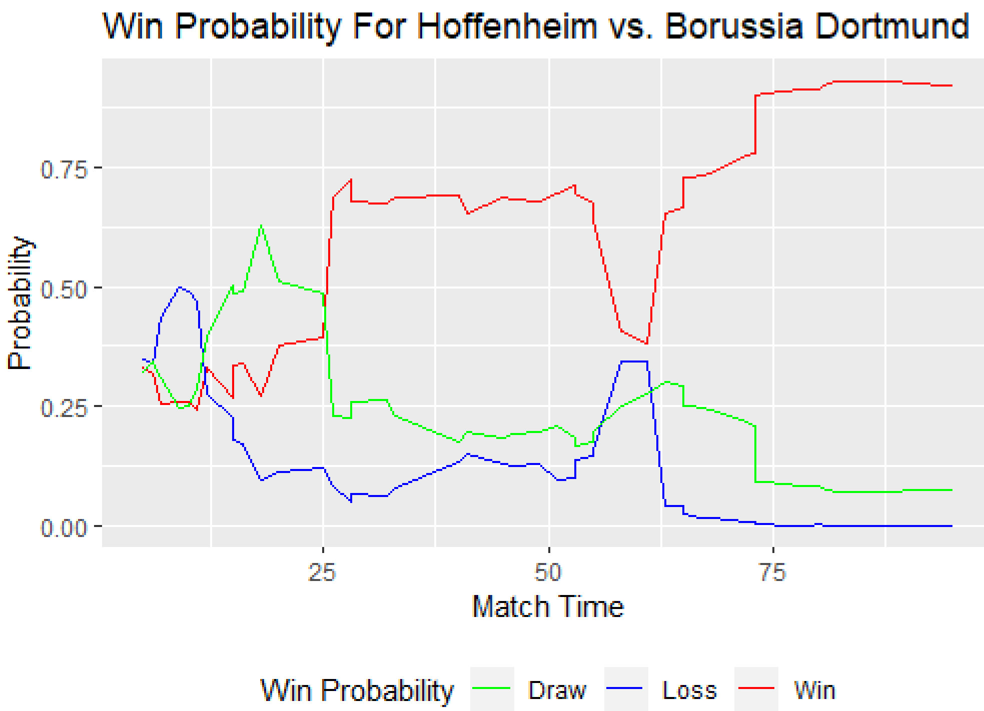
Through the use of the ESPN commentary data, I hoped to create an in-game win probability model to evaluate the significance of VAR decisions in terms of win probability added. I based my model off of the one created by Robberechts, Van Haaren, and Davis[17] who introduced a Bayesian statistical model using 8 variables (Game Time, Score Differential, Rating Differential – ELO, Team Goals, Yellows, Reds, Attacking Passes – rolling average, Duel Strength – rolling average) to predict win probability during a specific match. Because ESPN commentary data only provides information about key events that took place during a match rather than minute by minute updates, I was forced to go about creating the model in a slightly different manner. In the future I hope to use my model with tracking data to better predict win probability during a match.

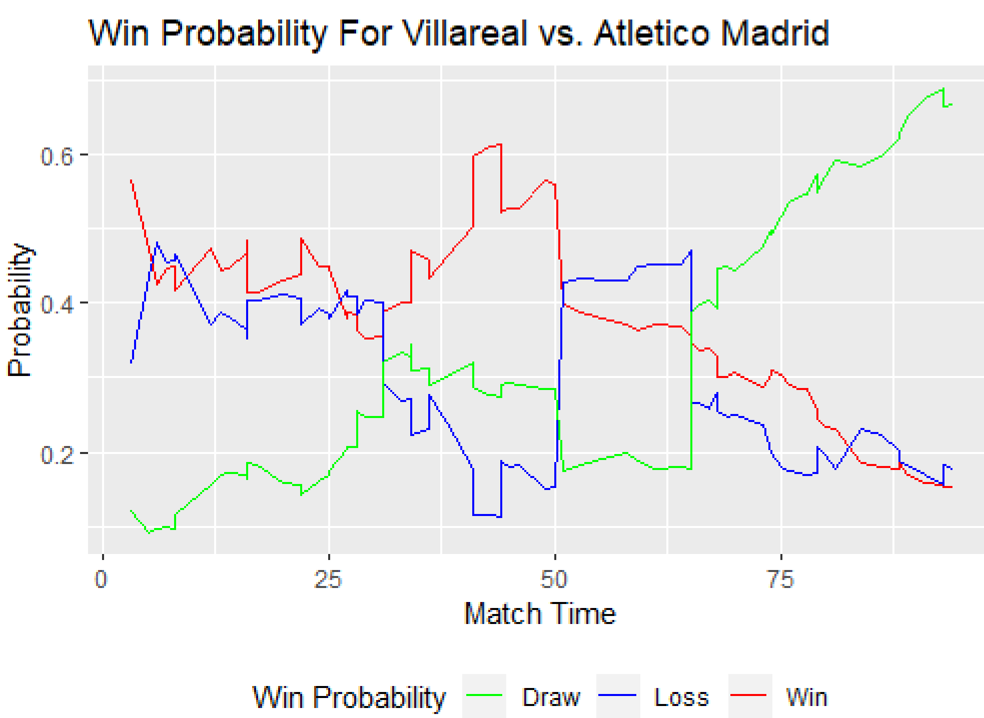
I used many of the same variables from the model from Robberechts, Van Haaren, and Davis, but did not have access to statistics such as attacking passes and duel strength. I included match time as a percentage of the total time in the match to account for differences in match lengths due to stoppage time. I included the scores of the home and away team as well as the number of yellow and red cards for each team at the specific match time. As a substitute for ELO ratings, I used pre-match betting odds from "bet365" who provide reliable odds for a home win, draw, and loss, which I included in the model. As a substitution for attacking passes and duel strength, I used a rolling average of shots attempted, corner kicks taken, and penalties taken for both teams throughout the match. The dependent variable for the model was the home team result which could take three forms: win, draw, or loss.

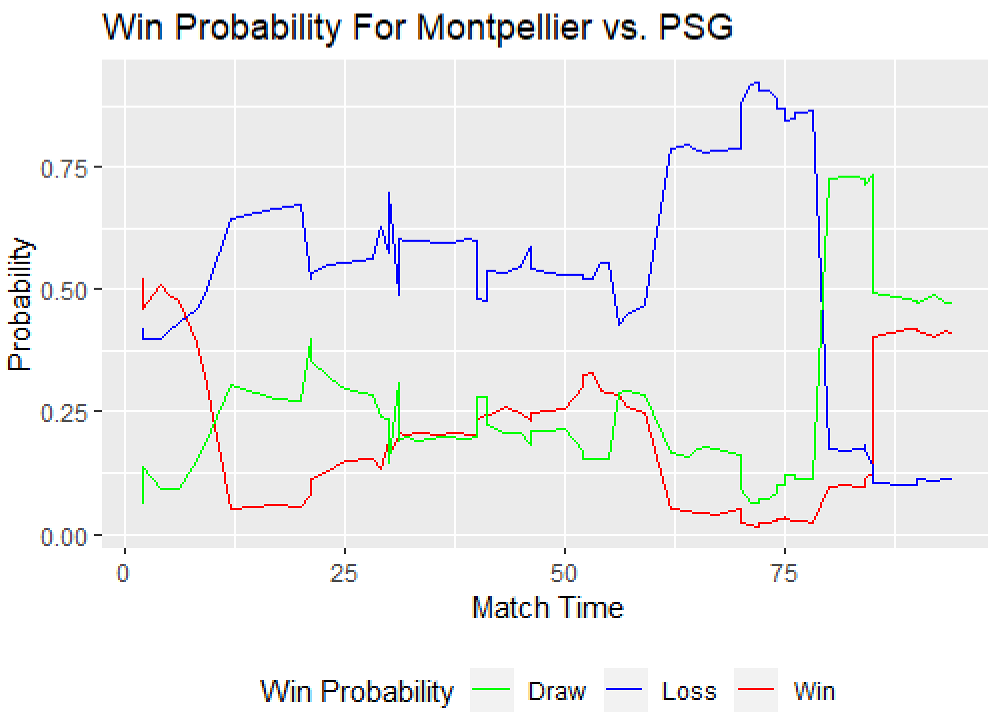
I ran the win probability model using two machine learning techniques and compared the overall results to determine the better model. I first ran a multinomial logistic regression using the "nnet" package in R with the model parameters mentioned previously. I did this because the dependent variable took three forms, so a standard logistic regression could not be used. I used a 10-fold cross validation to evaluate the model and found a mean accuracy of 66.29%. The training and test sets were split by game ID, so each full match was included in either the training or test data. I also found that the multinomial regression was 92.02% accurate in predicting the final outcome of a match. I also ran a random forest model with 10-fold cross validation once again. Although the mean accuracy was slightly smaller at 64.13%, the model could predict the final outcome of the match with 96% accuracy. Because of the improved end-of-match accuracy, I decided to utilize the random forest model to evaluate the win probability associated with VAR.

After running the random forest model, I used a 75%-25% train-test split to predict the in-game win probability for matches. I did this to evaluate win probability throughout matches and see what factors provided the largest increases and decreases in win probability throughout a match. First, I took a look at a match between Hoffenheim and Borussia Dortmund on the final day of the 2017-2018 Bundesliga campaign. Hoffenheim came into the match as slight home underdogs as they were just three points behind Dortmund in the table. A large jump in win probability can be seen at around the 25th minute when Andrej Kramaric struck first for Hoffenheim. A dip and spike can be seen around the 60th minute when Marco Reus equalized for Dortmund and then minutes later Adam Szalai grabbed the lead back for Hoffenheim. Just before the 75th minute a final spike can be seen as Hoffenheim secured the win with a third goal courtesy of Pavel Kaderabek. While this match shows the home team winning, the second match between Villareal and Atletico Madrid ended in an exciting tie. Villareal came into the match as home favorites and the match stayed scoreless until around the 50th minute when Filipe Luis scored for Atletico. This goal came against the run of play as Villareal seemed to be having more opportunities before the goal was scored. Mario Gaspar equalized in the 65th minute and the match seemed to be heading for a draw. As the game went on after the second goal, the probability of a draw continued to rise until the match ultimately ended at 1-1. Finally, a match between Montpellier and PSG showcased an upset win by the home team. Montpellier came into the match as slight favorites according to the model, but that was quickly wiped away after the match begun. PSG controlled the game and were helped out by an own goal by Montpellier in the 12th minute. PSG returned the favor in the 21st minute and scored an own goal to tie the match at 1-1. PSG continued to look like clear favorites and Angel Di Maria scored to put them up 2-1 in the 61st minute. This looked like the end for Montpellier, but amazingly they scored two goals in the last ten minutes of the match to completely change the tie. As the match ended, the model still gave hope for PSG to score an equalizer in the dying moments, but Montpellier held on for the upset win.

Figure 34: *Three example matches showing the change in win probability for the home team: (1) Hoffenheim vs. Borussia Dortmund, (2) Villareal vs. Atletico Madrid, (3) Montpellier vs. PSG*

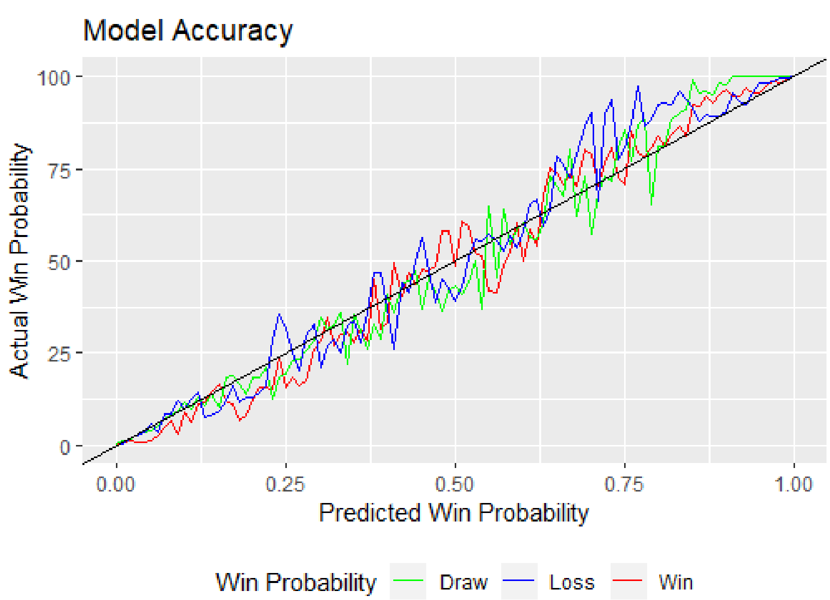






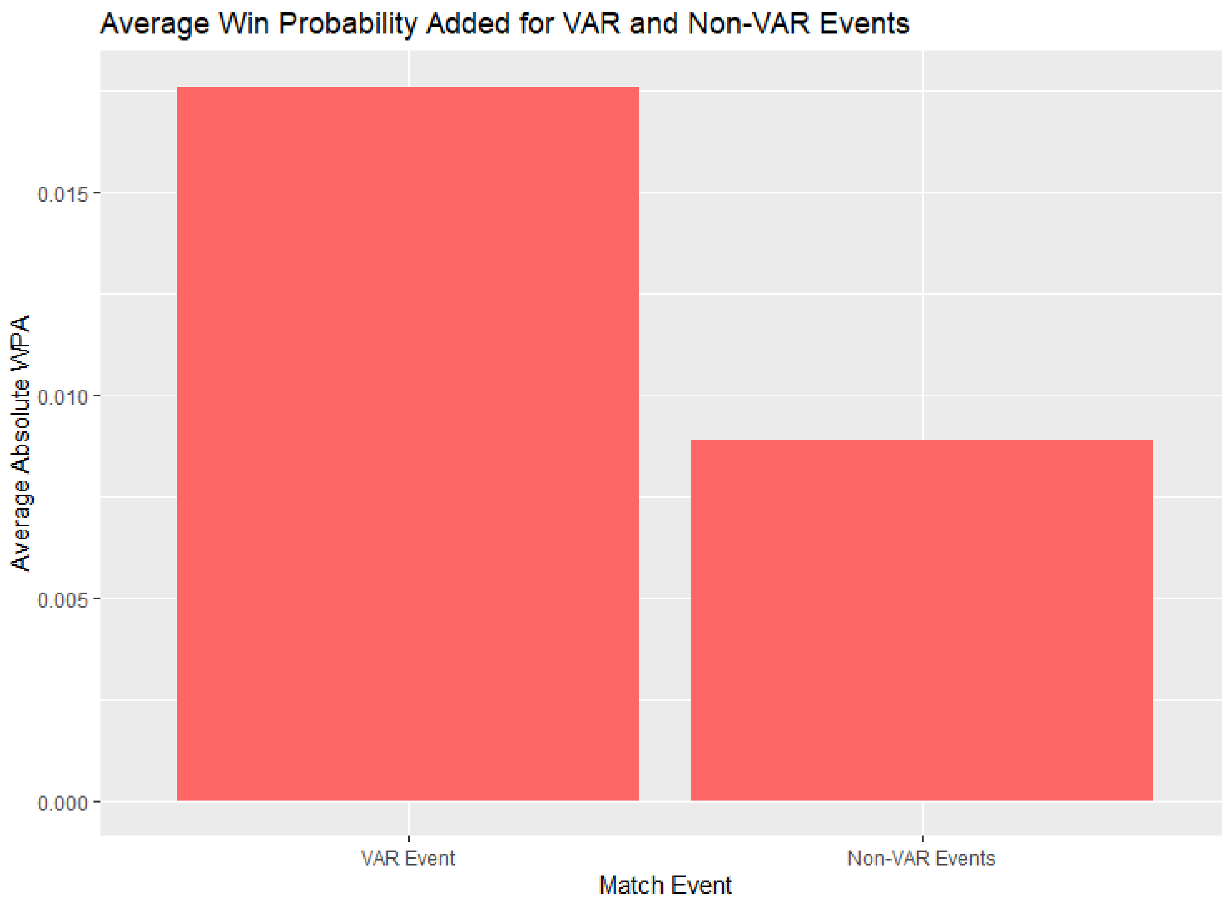
The following plot shows the model accuracy of the random forest model. The x-axis plots the predicted win probability given by the model and the y-axis plots the actual win probability. The actual win probability is measured by calculating the percentage of matches that were won when the predicted win probability was a specific value. If the model performed with 100% accuracy, the plot would be a straight line with a slope of one and the predicted win probability would always equal the actual win probability. This plot shows that, overall, the random forest model performed quite well and did not deviate too far from the ideal slope.

Figure 35: *Model accuracy of the random forest model*



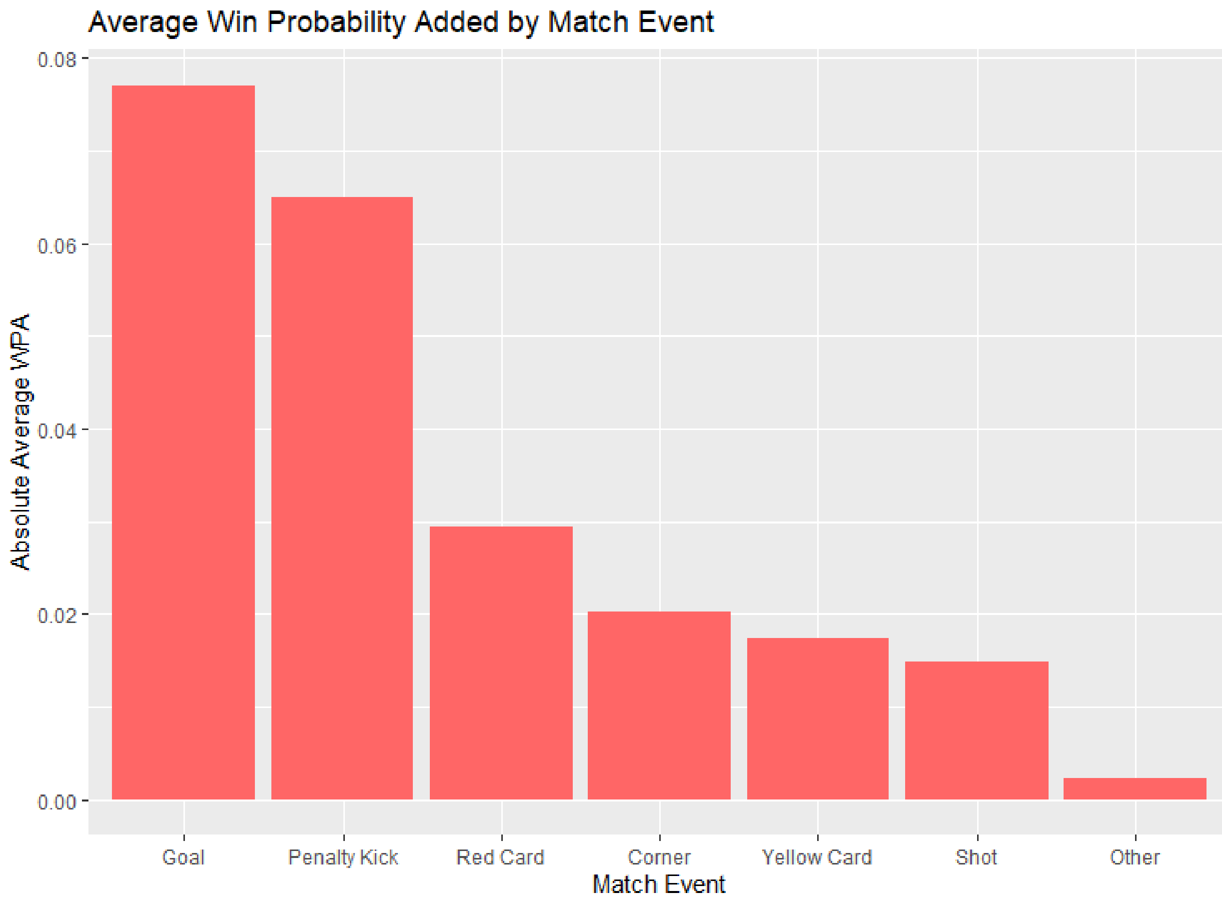
After predicting the win probabilities of all matches in the VAR era, I wanted to calculate the win probability added for VAR events. To calculate win probability added, I simply subtracted the current win probability from the win probability from the previous match event. I then filtered out the VAR events and compared the average WPA (win probability added) between VAR and non-VAR events. I used the absolute value of WPA to put instances where the home team became more likely to win and instances where the away team became more likely to win on the same scale. It is clear to see that events that went to VAR for a second look were of higher importance to those that did not use VAR. This makes intuitive sense because events that went to VAR involved goals, penalties, and red cards which can all swing the tide of a match. I noticed when doing this evaluation that the only VAR events that led to high WPA values were those calls that were upheld on the field. This is the case because if a goal was ruled out by VAR due to offsides, the original goal was never included in the commentary data. Because of this, I wanted to also look at the WPA based on match events.

Figure 36: *Win probability added for VAR and non-VAR events*



After looking at the win probability added for all match events in the ESPN commentary data, I found that the three events that led to the greatest increase in win probability added were goals, penalty kicks, and red cards. Unsurprisingly, these are the three events that can be reviewed by VAR. This drives home the point that the decisions made with the use of VAR are the most important decisions made during a match. It is of upmost importance to get these calls correct as they can have serious consequences and can be the difference between a team winning and losing a match.

Figure 37: *Win probability added by match event*



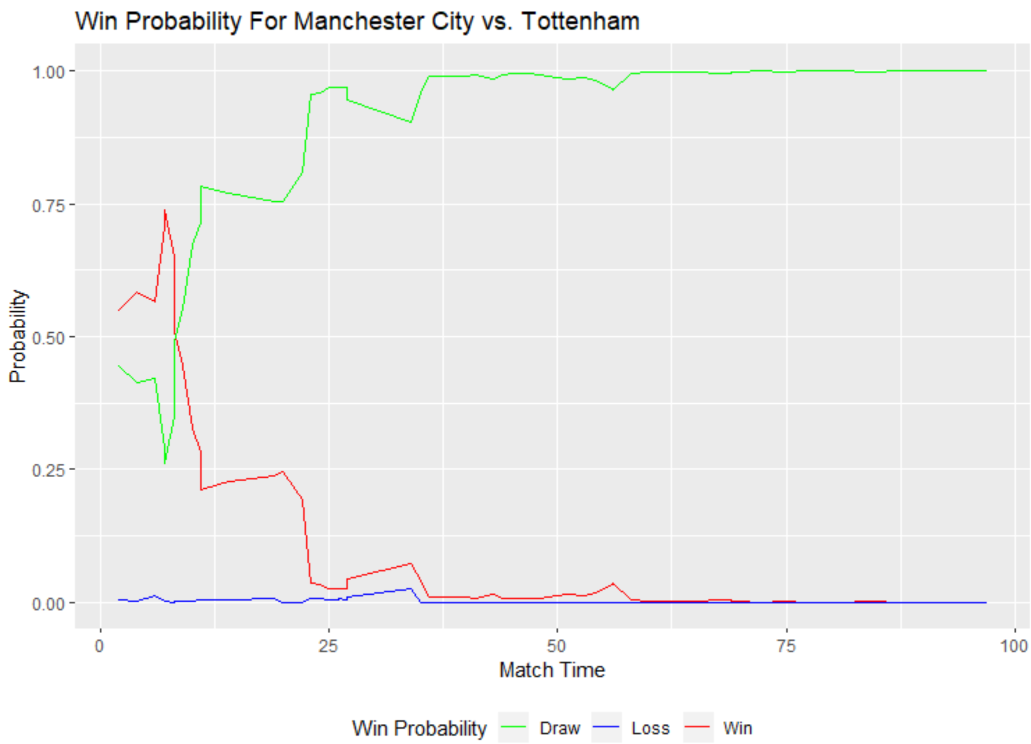
The three events in *Table 3* are examples of VAR decisions that were upheld that led to large swings in WPA. First, in a match between Inter Milan and Fiorentina, Danilo D'Ambrosio scored what would be the game-winning goal for Inter in the 77th minute. The goal was checked for offsides by VAR and it was determined that the goal would stand. This was a very crucial usage of VAR and if the opposite decision would have been made, Inter Milan's probability of winning would have been quite different. Second, in a match between Cagliari and Benevento, Pietro Iemmello converted a penalty to equalize the match for Benevento after Massimo Coda was brought down in the box. The decision was once again reviewed by VAR and much to the chagrin of the home Cagliari supporters, the penalty on the field stood. Although Cagliari miraculously scored a winner in the last minute of stoppage time, their chances of winning the match decreased tremendously after the controversial penalty was given so late in the match. Finally, in a match between Düsseldorf and Nürnberg, Matheus Pereira was shown a red card for an off-the-ball incident in only the second minute of the match. As a result, Nürnberg were forced to play the rest of the match with only ten men. This VAR decision had serious consequences for Nürnberg and Düsseldorf was projected to be 7.4% more likely to win even though it was only the second minute of the match. These calls show the significance of VAR decisions that were upheld, but what about a decision that was overturned?

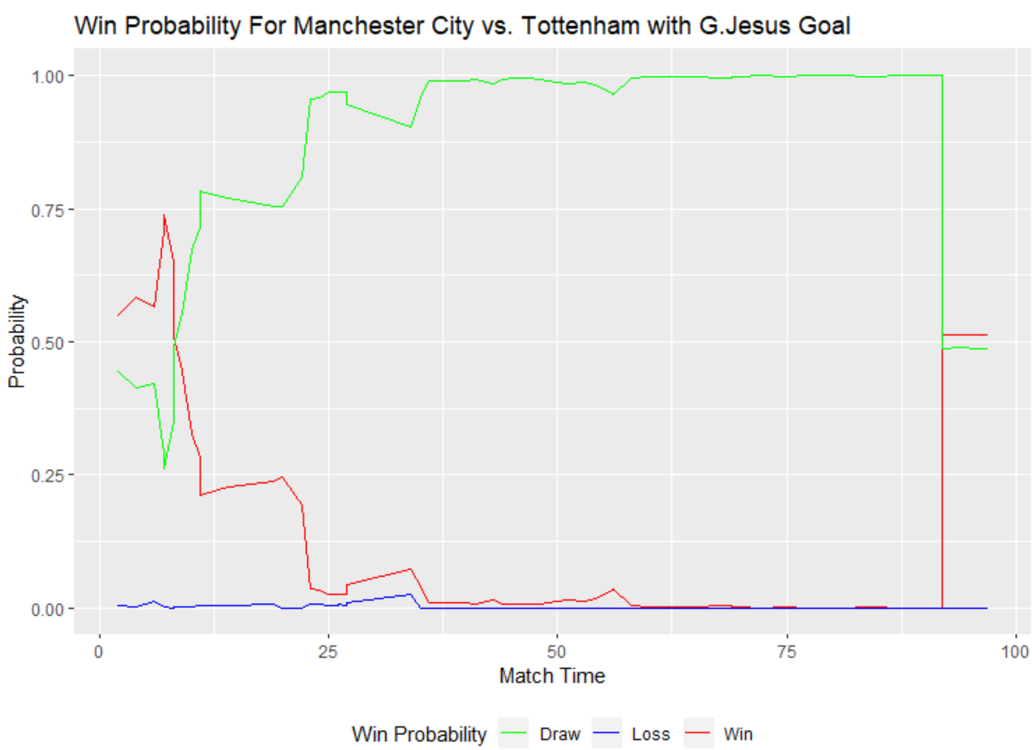
Table 3: *Three VAR events that led to large swings in WPA*

Home	Away	FTHG	FTAG	Match Time	Event	WPA
Inter Milan	Fiorentina	2	1	77	Goal! Inter Milan 2, Fiorentina 1.	0.396
Cagliari	Benevento	2	1	94	Goal! Cagliari 1, Benevento 1.	-0.226
Düsseldorf	Nürnberg	2	1	2	Red Card Matheus Pereira (Nürnberg).	0.074

The two plots below show the win probability graphs for a match between Manchester City and Tottenham at the start of the 2019-2020 Premier League season. Manchester City came into the match as clear favorites as they had dominated throughout the 2018-2019 campaign and came off a 5-0 thrashing of West Ham United. All seemed to be going right when Raheem Sterling scored in the 20th minute for Manchester City, but a quick equalizer from Tottenham's Erik Lamela changed things and increased the likelihood of a draw. Sergio Aguero gave Manchester City the lead before the break, but Lucas Moura scored for Tottenham to level the contest at 2-2 in the 56th minute. Although the match seemed to be destined for a draw after the goal by Lucas, Manchester City thought they had the winner in the 92nd minute after a goal by Gabriel Jesus off a corner kick. The goal was checked by VAR after appeals of handball in the build-up, and after indications that the ball hit Aymeric Laporte's arm before the goal was scored, VAR cancelled the goal and the match finished as a draw. I used the win probability model to estimate the change in win probability if the goal was given instead of called out. I found that the goal by Gabriel Jesus increased Manchester City's win probability by 51.4%, which is quite significant. The VAR decision made had significant implications to the end result of the match and it shows the importance that such calls can have, especially in crucial moments of key matches. With VAR decisions occurring in key matches such as the 2018 World Cup Final and the 2019 Champions League Final, the importance of getting these calls correct cannot be understated and can have drastic implications in terms of win probability added or lost.

Figure 38: *The WPA for a match between Manchester City and Tottenham as it truly was, and with an added goal by Gabriel Jesus that was ruled out by VAR*



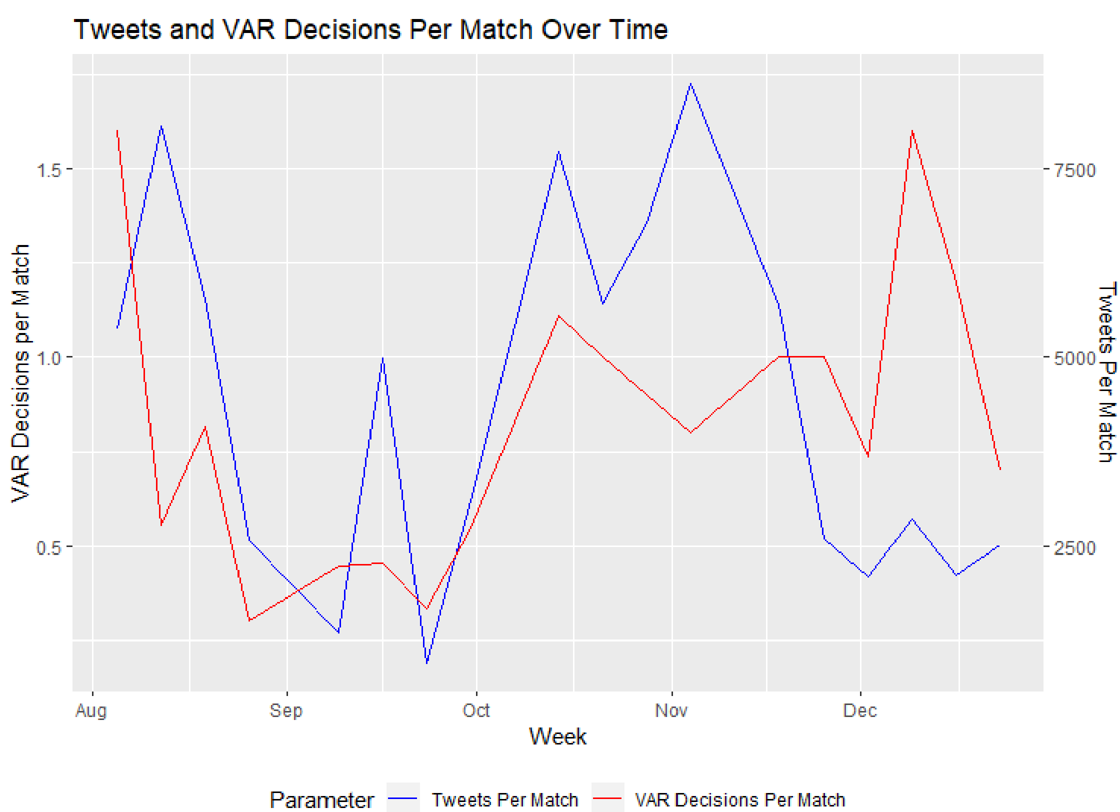


Sentiment Analysis: VAR

After analyzing the impact that VAR has had on professional soccer, I would like to transition into examining the fan experience of watching the game after the introduction of VAR. The one aspect of VAR that I could not track or measure in the previous section was how frequently a call made by VAR was the correct one. This is not an easily identifiable measure and it is one that is highly subjective. My goal is to analyze fan sentiment using Twitter to identify possible aspects of VAR that fans dislike and areas of improvement for the technology itself.

There were 805,548 tweets containing the term "var" or the hashtag "#var" gathered during the first half of the 2019-2020 English Premier League season. Using these tweets, I was able to perform sentiment analysis to evaluate how fans felt about the new use of VAR in the Premier League. First, I looked at how many tweets were posted each week to understand how fans felt about VAR over time. *Figure 39* shows the average number of tweets per match by week from August to December as well as the number of VAR decisions made during that same time period (according to ESPN). There were 158 VAR decisions made during the first half of the Premier League season. It is clear that fans were quick to tweet about the use of VAR through the early portion of the season, especially since a significant number of calls were being made during that time. There is a weak positive correlation between the number of VAR calls and the number of tweets about VAR during a given week (R = .206). This means that we would generally expect fans to tweet more about VAR when there are more decisions that involve the technology, which makes sense.

Figure 39: *Number of tweets per match and number of VAR decisions per match over the course of the 2019-2020 English Premier League season*



Twitter Sentiment Background

Next, I calculated the word frequency of every term from the tweets containing "var" in order to take a closer look into the content of such tweets[23]. After tokenizing the text and removing common stop words, I created a list of the top 15 most common words and a word cloud of the top 250 terms (excluding "var" which was the most frequent term). Among the most frequent terms were popular words in soccer such as "goal", "game", "penalty", and "ref" as well as the team names of the top two clubs in the English Premier League: "liverpool" and "city" (short for Manchester City). The casual and informal nature of Twitter shows as profanity was used by tweeters when discussing VAR.

Figure 40: *Word frequency plot of the top 15 most common terms in tweets containing "var"*

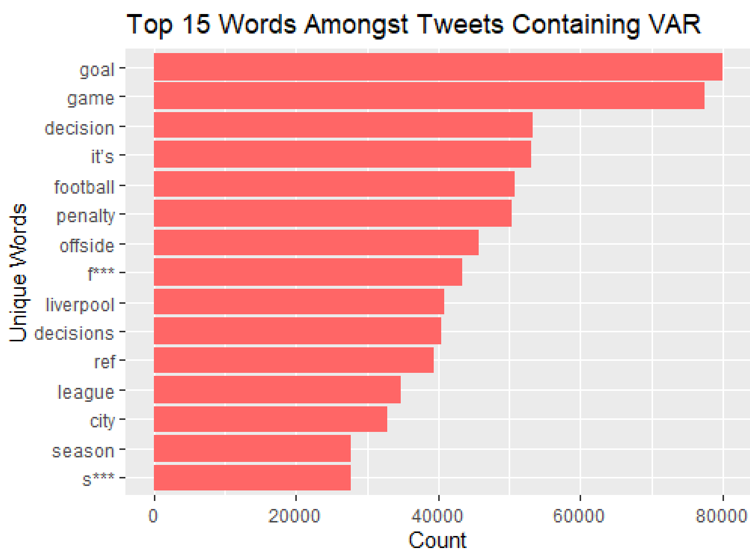
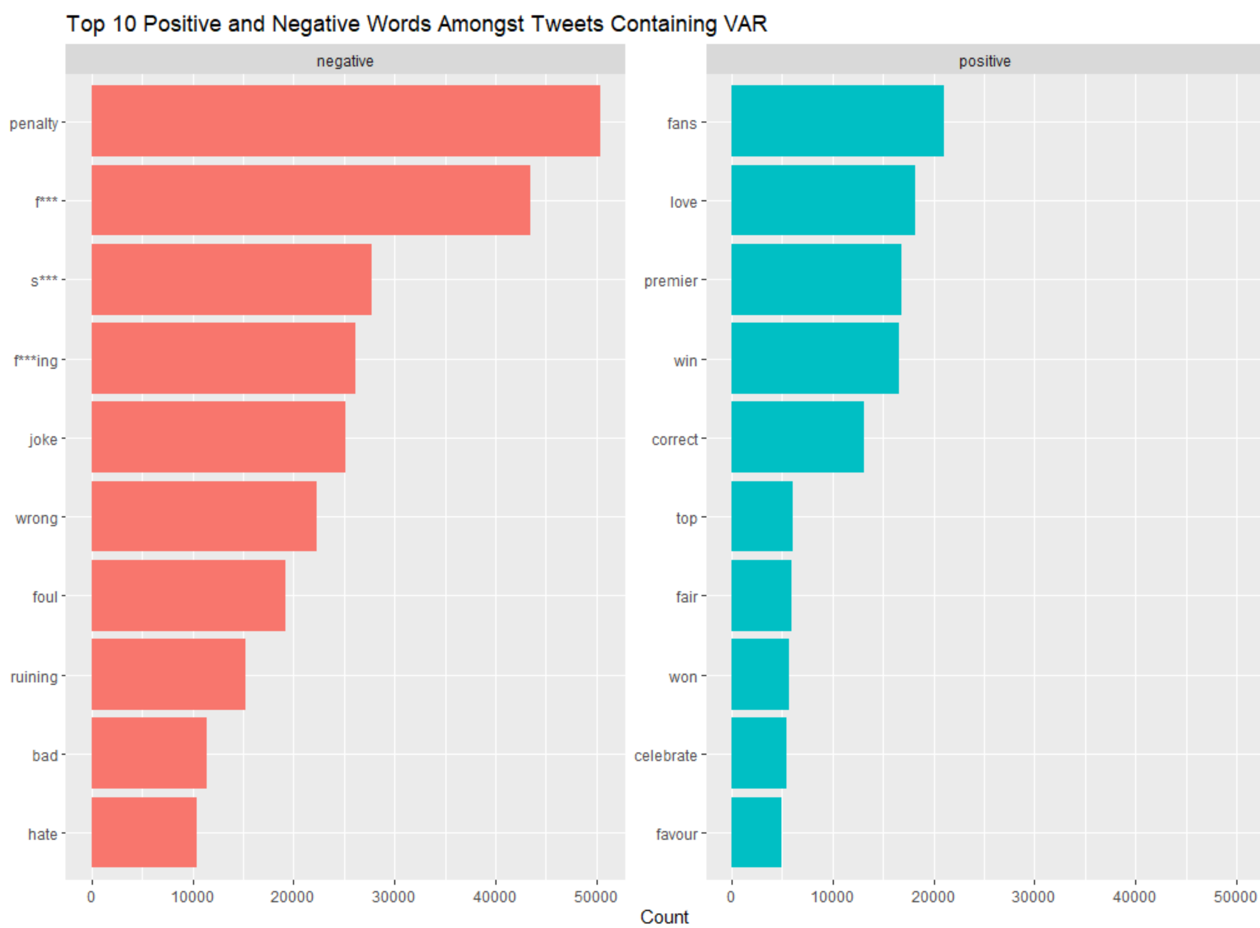


Figure 41: *Word cloud of 250 most frequent terms*

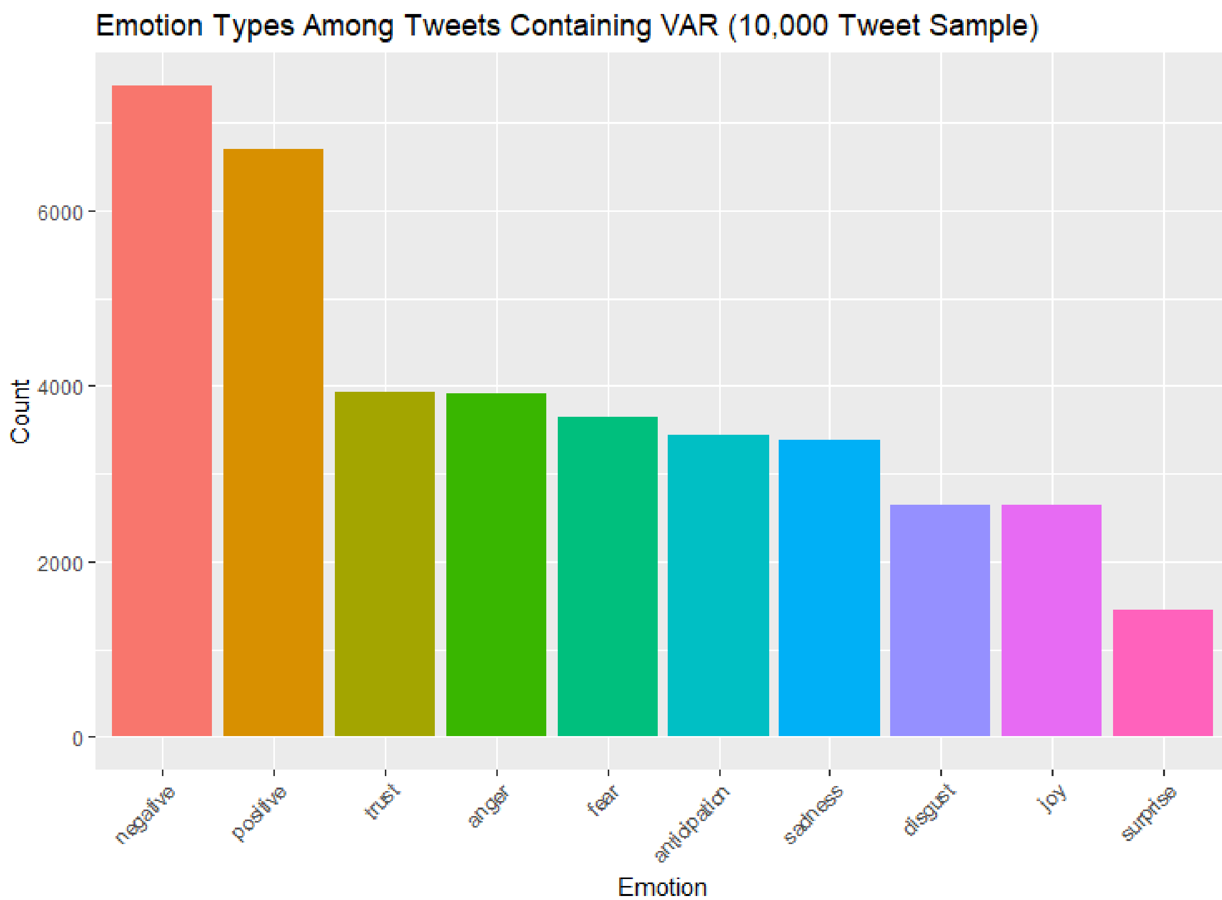
By using the Bing Sentiment Lexicon which categorizes 6,786 words as either positive or negative, we can see how frequently positive and negative words occur in tweets. After merging the Bing Lexicon with the tokenized terms, the word frequency graphs in *Figure 42* can be created with negative terms on the left side and positive terms on the right. Several new terms appear when using the Bing Lexicon. Frequent negative terms such as "joke", "wrong", "bad", and "hate" suggest that some fans are discouraged with the use of VAR and may feel that decisions are not going the right way. Some of the positive terms such as "love", "correct", and "fair" also show the opposite viewpoint of VAR where fans are appreciative of the new technology. It seems that there is a larger portion of negative words in the plot, but this is the case only because the Bing Lexicon is made up of 70% negative words and only 30% positive words. Similarly, 70% of the terms in the VAR tweets picked up by the Bing Lexicon were negative.

Figure 42: *Word frequency plots of top 10 most common positive and negative terms*



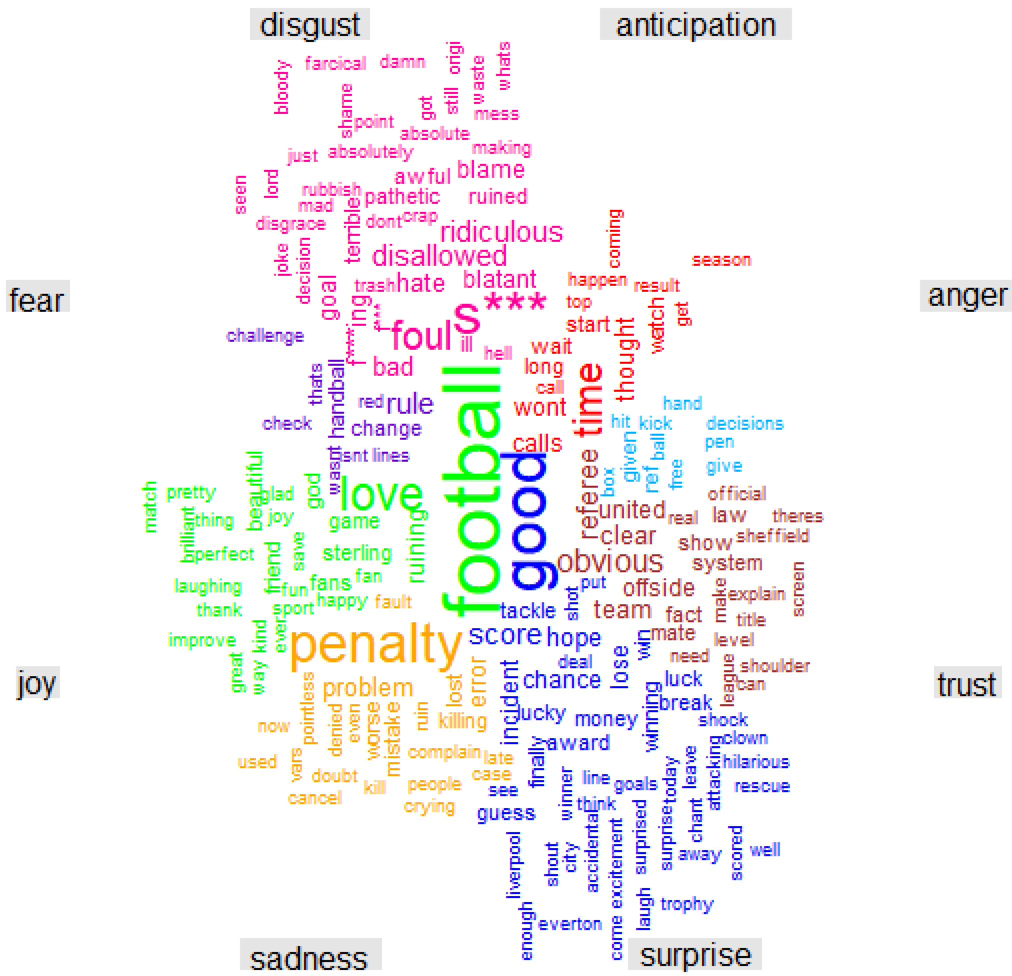
Sentiment can be broken down even further than just positive and negative through the use of the NRC Lexicon which categorizes words into eight additional categories: anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. A word may fall into zero or multiple categories as well. Because of the large number of tweets in the dataset which all have many terms that must be broken down in the NRC Lexicon, I analyzed a random sample of 10,000 tweets to see which emotions were most prevalent in the data. *Figure 43* shows that the most basic sentiments of negative and positive appeared most frequently in the sample. There were 7,426 occurrences of negative terms followed by 6,697 positive terms which once again displays the negativity shown by tweeters towards VAR. One of the additional sentiment categories that appeared frequently was anger with 3,911 terms showing that a good portion of tweeters do not just have negative feelings towards the use of VAR, but are also angry with the decisions being made. Tweeters showed trust towards VAR as well, which demonstrates the divide amongst fans about the use of the technology. It must be noted that there may be bias in these findings of sentiment because fans are more likely to tweet about the use of VAR if it negatively affects their favorite team. Fans are very unlikely to tweet when they are unaffected by the use of VAR, which may be the cause of high frequencies of positive and particularly negative sentiment. This similar bias has been noted in political tweets as well where there is considerable negative rhetoric amongst opposing politicians[24].

Figure 43: *Frequencies of terms by NRC Lexicon Category*



The word cloud in *Figure 44* takes the sampled tweets above and shows some of the most frequent terms associated with the NRC Lexicon. Although there were not as many terms that fell into the disgust category, there were many words that appeared in the word cloud. Words of disgust such as "awful", "ridiculous", "pathetic", and other profane terms suggest fans are more creative in their choice of wording when tweeting a negative remark. It should be noted that context is key when evaluating these tweets and words themselves cannot necessarily determine the sentiment of a phrase.

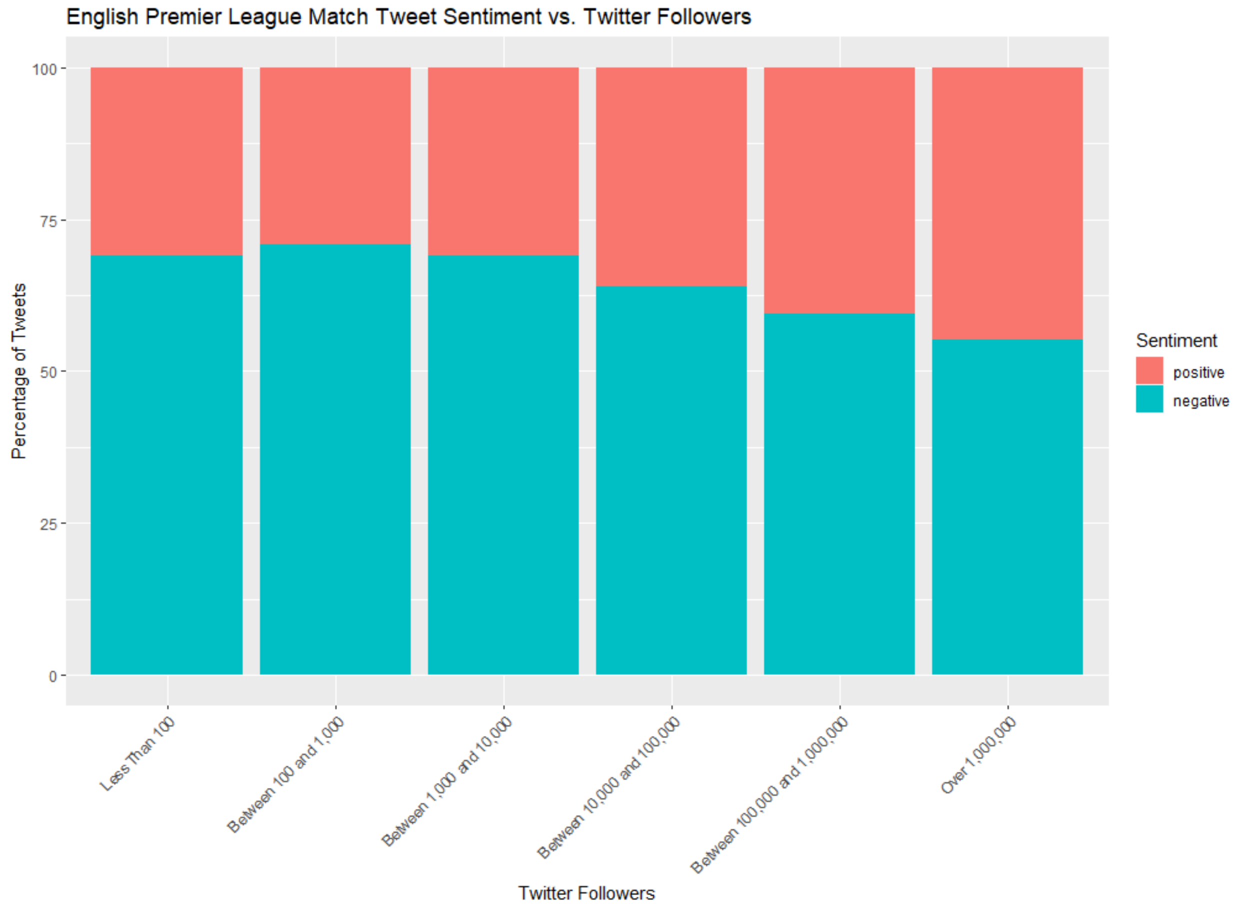
Figure 44: *Word cloud comparing the frequencies of terms associated with the NRC Lexicon*



Twitter Followers and Likes:

I also wanted to take a look at the impact of the number of followers on VAR tweet sentiment. For each tweet that contained VAR during the English Premier League season, I also scraped the user's corresponding information which includes number of followers at the time of the scrape. I once again utilized the Bing Lexicon to analyze the positive and negative tweets made by each user. *Figure 45* shows the percentage of positive and negative tweets for users in several bins of follower counts. There is a clear relationship between number of followers and tweet sentiment: the more followers a user has, the more positive they are. This makes sense because many users with a large number of followers that discuss the use of VAR on Twitter are employed by the Premier League and do not want to make any harsh claims about the technology used in their own league. Common fans with fewer followers do not share this constraint and are free to tweet in a more negative fashion. This disconnect between the common fans and the higher-ups in the Premier League is important because it may make ordinary people feel that they do not have a say in the decisions the league makes. While several current and former players such as Raheem Sterling, Gary Linekar, İlkay Gündoğan, and Peter Schmeichel have spoken about the negatives of VAR, many large corporate accounts have not done the same.

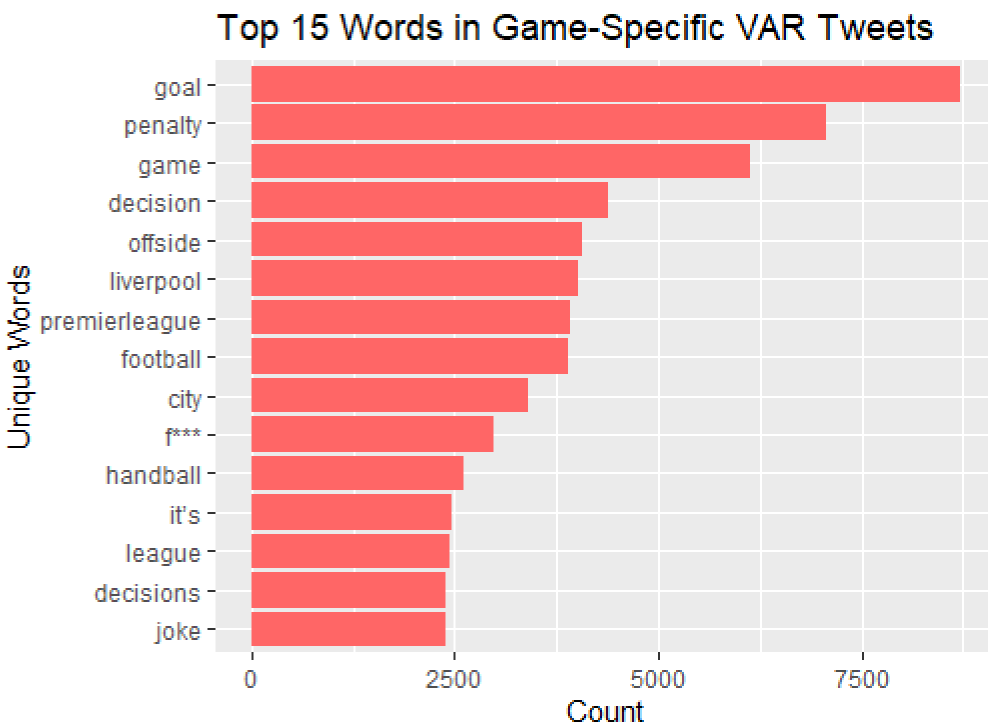
Figure 45: *Twitter sentiment by number of followers*



Sentiment Analysis: Game Hashtags

There were 61,127 tweets containing the term "var" and a game-specific hashtag gathered during the first half of the 2019-2020 English Premier League season. Using these tweets, I was able to perform sentiment analysis to evaluate how fans felt about the new use of VAR in specific games during the Premier League season. *Figure 46* shows a word count of the 15 most frequent words in the subset of tweets about VAR that contained a game hashtag. Many of the same terms that frequently appeared in these tweets appeared in the VAR tweets as well, which makes sense considering many of the tweets with a game hashtag were also included in the larger dataset.

Figure 46: *Word frequency plot of terms in tweets containing a game-specific hashtag and "var"*

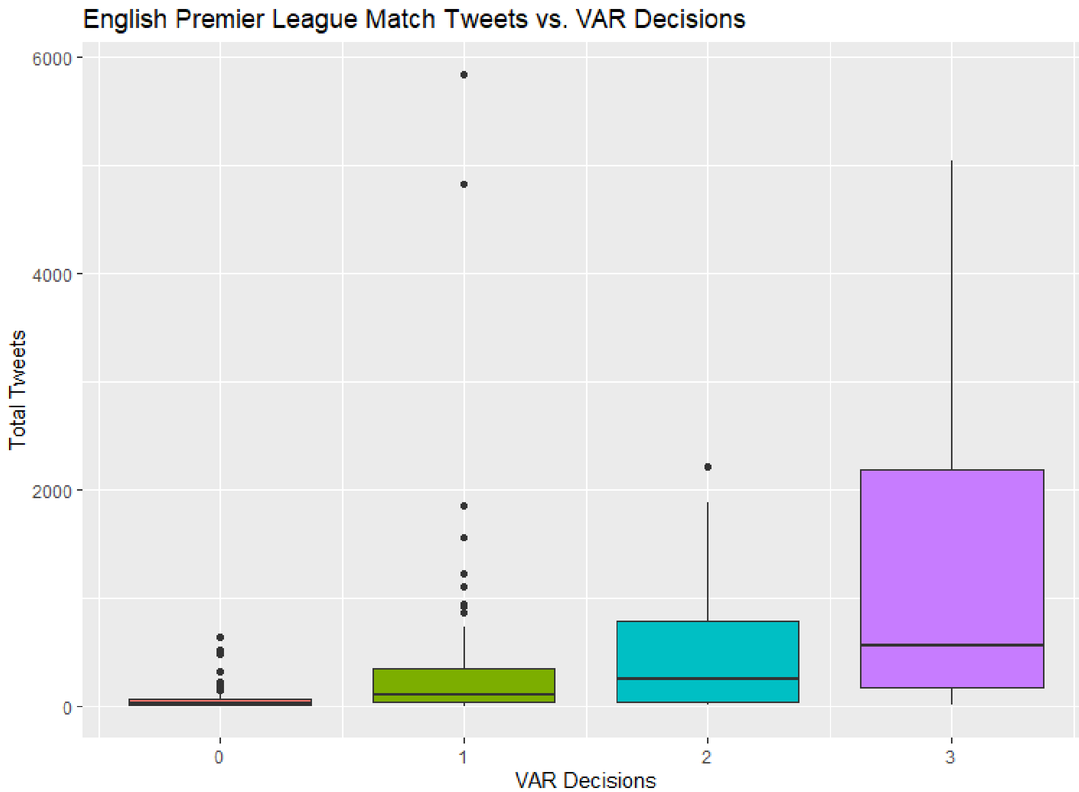


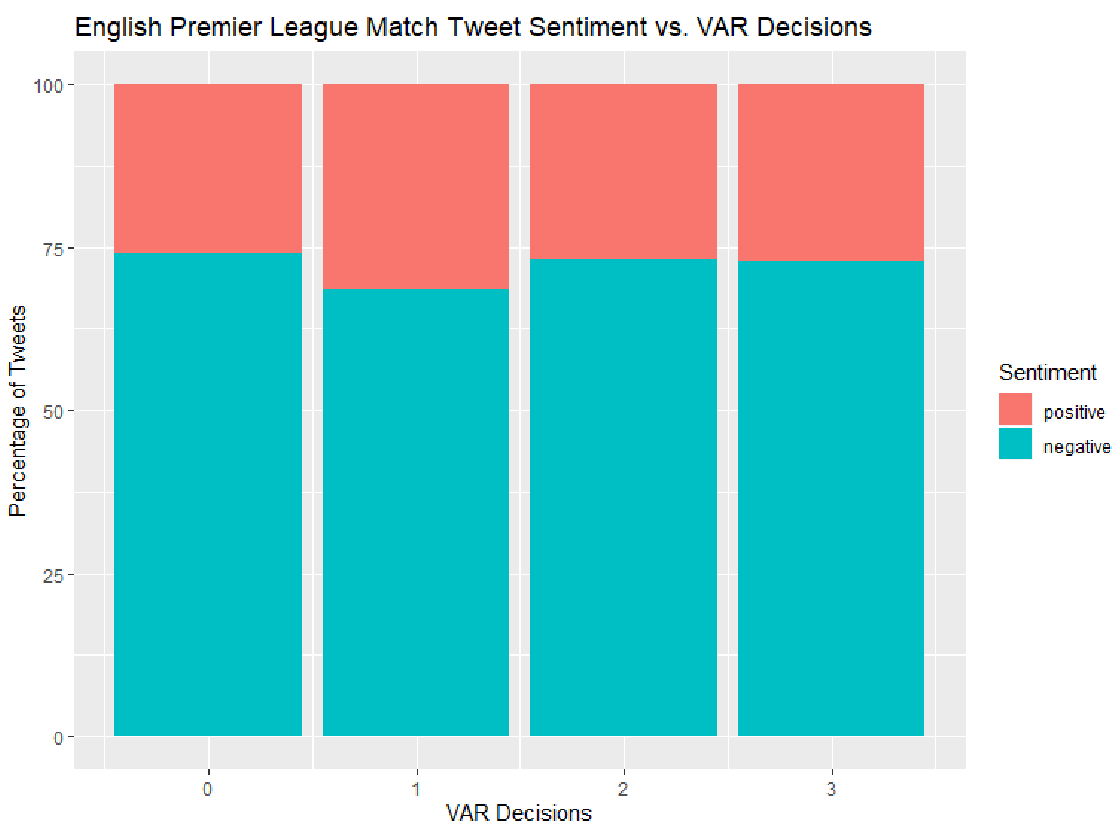
By looking at tweets on a game level, we can analyze how the number of VAR decisions in a match affects fan sentiment. The top graph in *Figure 47* plots the total number of tweets against the number of VAR decisions made in the match. A clear relationship can be seen as the number of tweets grows as the number of VAR decisions made in a match increases. The correlation between total tweets and VAR decisions is 0.357 suggesting a weak positive correlation between the two variables. Intuitively, this relationship makes sense as fans are more likely to voice their opinions about VAR when a call occurs during the match they are watching. *Figure 47* also plots the percentage of tweets that contain positive and negative terms against the number of VAR decisions. The number of positive and negative terms is once again calculated by using the Bing Sentiment Lexicon. There does not seem to be a strong relationship between positive and negative sentiment and VAR calls. One may think that a fan may show more negative sentiment towards VAR if more calls are made during the match, but it seems that fans tend to be generally negative no matter how many calls are made.

Table 4: *Average number of tweets and total matches by number of VAR decisions*

VAR Decisions	Average Tweets	Total Matches
0	65.90	82
1	415.18	68
2	465.30	27
3	1240.50	12

Figure 47: *Total tweets and percentage sentiment by number of VAR decisions*

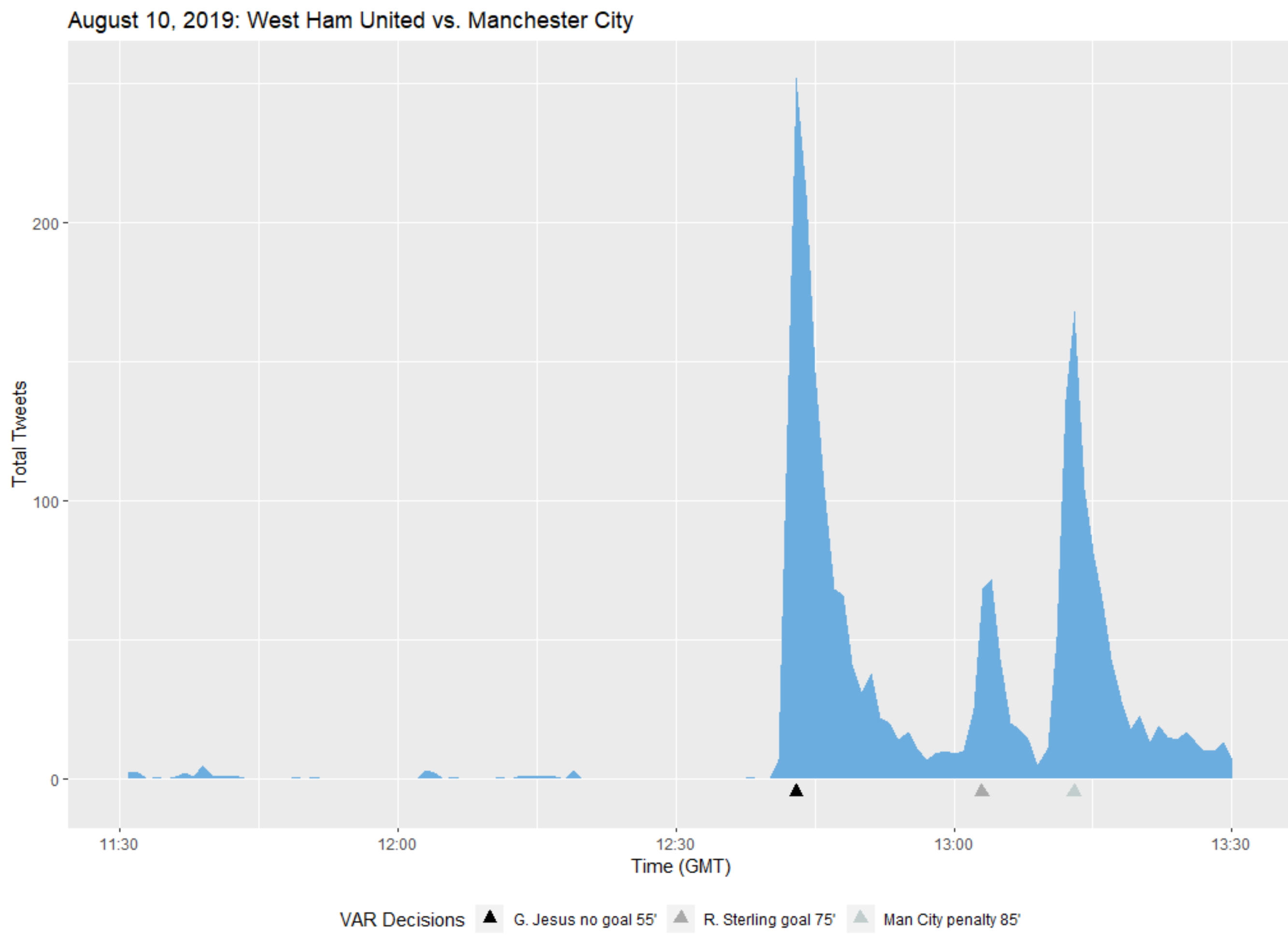




In-game Analysis:

By using Twitter data from specific games, sentiment can be measured over the course of a match. We can see how fans react when a VAR call is made and evaluate whether a call may have been the correct one. *Figure 48* shows the breakdown of tweets during a selected match from the 2019-2020 English Premier League season. The match was between West Ham United and Manchester City and will go down in Premier League history as the first match to utilize VAR technology. As can be seen in the plot, there was very little talk on Twitter about the use of VAR during the match in the first half as the technology was not needed. A huge spike in tweets can be seen after the 55th minute when Manchester City striker Gabriel Jesus scored the opening goal after a clever assist from Raheem Sterling. VAR was needed to check for offsides in the build-up to the goal and many on Twitter were excited to talk about the use of the technology. Controversially, Sterling was called offsides (see *Figure 1*) and the goal was overturned. Two more instances of VAR were needed during the match including a less-controversial Sterling goal and a penalty retake by Manchester City. Both these instances saw spikes in Twitter activity which was expected.

Figure 48: *Breakdown of tweets during the August 10, 2019 match between West Ham United and Manchester City, the first match in the Premier League where a VAR decision was made*



We can go one step further by analyzing sentiment using the NRC Lexicon to evaluate positive and negative emotions which can be seen in *Figure 49* and *Figure 50*. The same spikes in sentiment occurred during the VAR decisions during the match. It is clear that surprise does not play a significant role in terms of sentiment during these VAR decisions as it does not reach the same high sentiment value as the other categories. Fans did reach high peaks of anger, sadness, and disgust after the first VAR decision, suggesting that this was a bad call. It was a very tight decision and Manchester City supporters as well as neutrals may have disliked the nature of the call. They were less concerned with the second call and reached higher levels of positivity rather than negativity. Interestingly enough, fans reached a peak in terms of fear after the third VAR decision, possibly because of the way the call was made. Manchester City were allowed to retake their penalty because of Declan Rice's early entry into the penalty box. Because such a call had rarely been made before the implementation of VAR, fans may have been fearful that the use of VAR would completely change how matches would be officiated.

Figure 49: *Breakdown of tweets by negative sentiment types during the match*

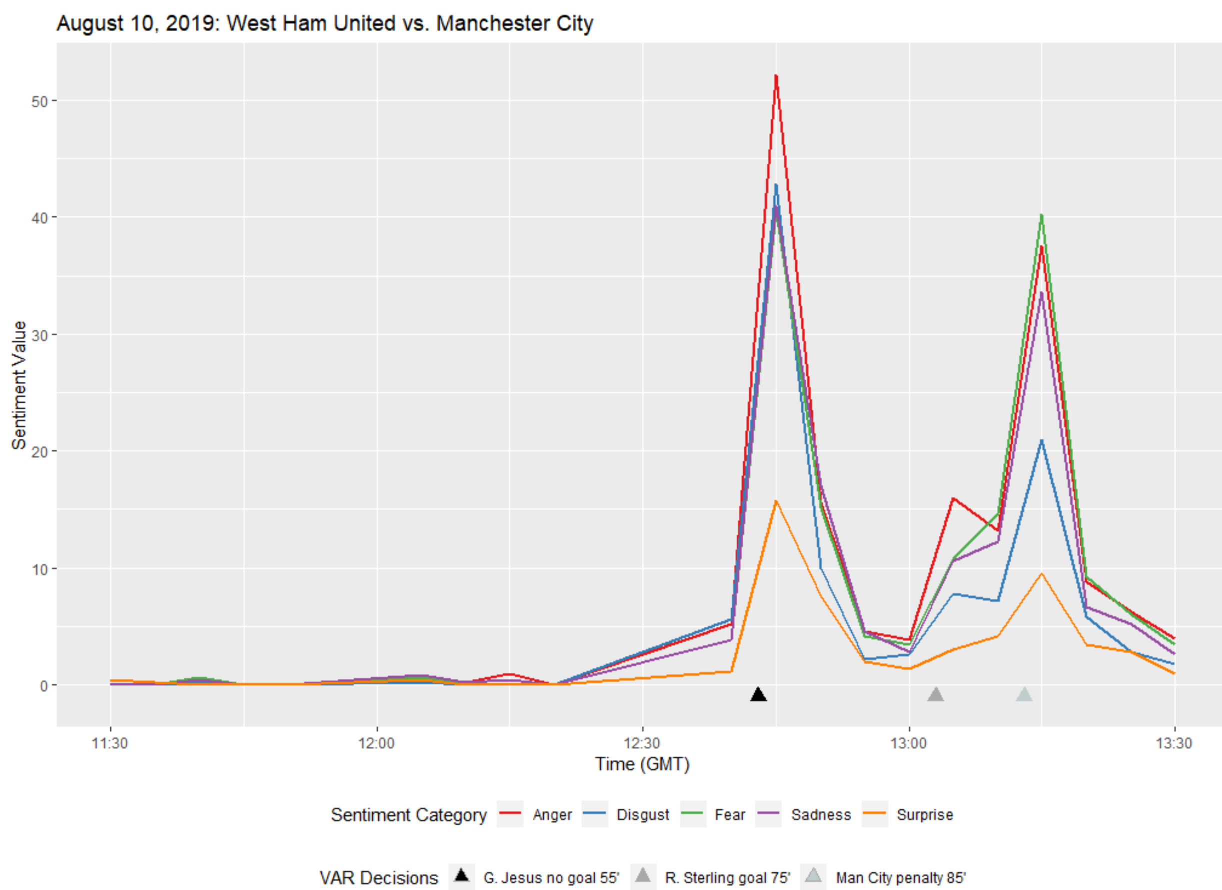
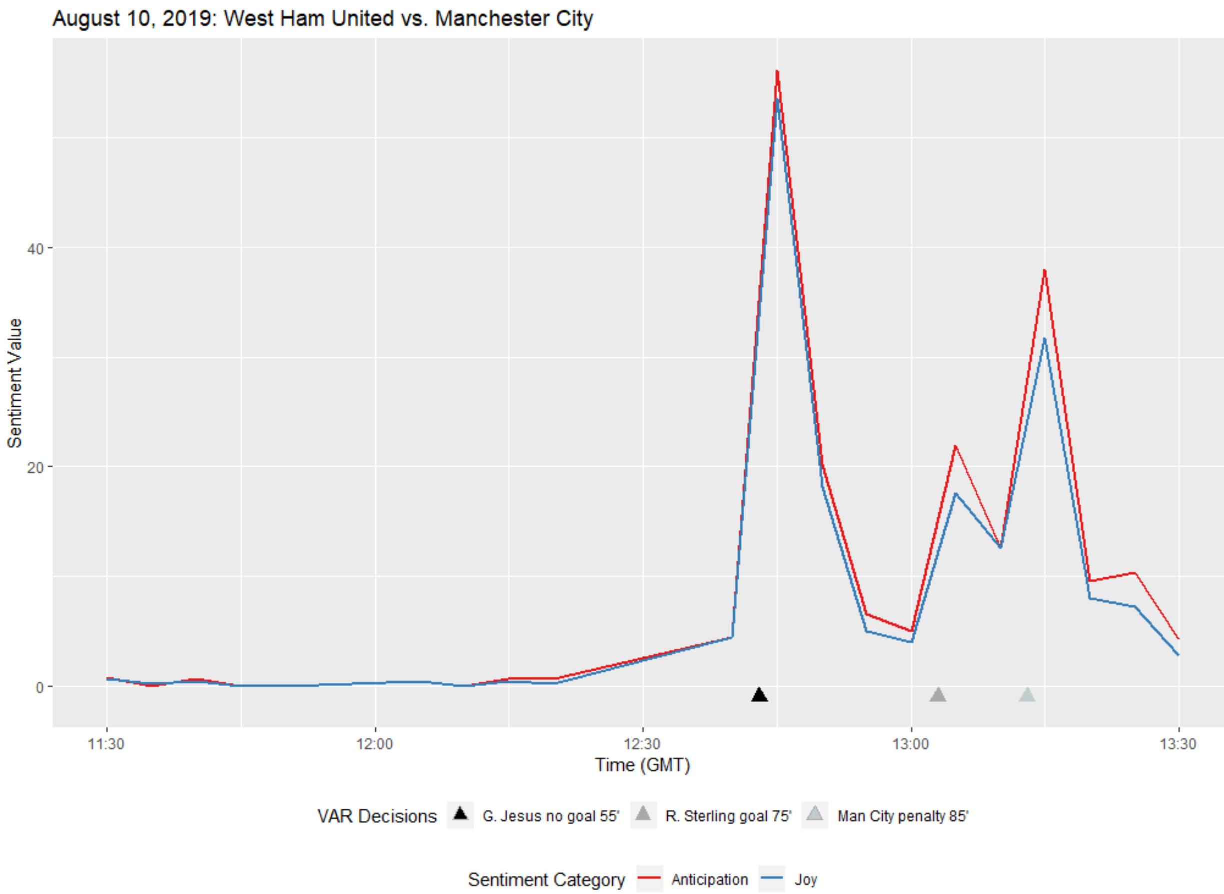


Figure 50: *Breakdown of tweets by positive sentiment types during the match*



Natural Language Processing: Predicting Sentiment

Now that we have analyzed fan sentiment on Twitter through the use of lexicons such as Bing and NRC, I wanted to utilize Natural Language Processing techniques to create a model to predict the sentiment of future tweets concerning VAR. To train my model, I used Twitter data from the SemEval 2014 competition which consisted of approximately 10,000 tweets that were categorized as either positive, negative, or neutral. I ran a Naïve Bayes Classification model in Python with the NLTK package using feature sets to predict the sentiment of a given tweet.

To create my model, I first tokenized the tweets by splitting each phrase into its corresponding words. For each of these strings of tokens, I then removed common stop words (ex. "the", "a", "and") and negation words (ex. "no", "not", "never") that would not have too big an impact on sentiment. Next, I found the 1500 most common words in the SemEval tweets and used those as word features. I originally used the 2000 most common words, but changed to 1500 because I noticed in previous steps that some of the most common words were not used very frequently. The vocabulary used on Twitter is more expansive than that used in normal writing because Twitter uses emoticons, handles, and slang terms. I then wanted to filter out a more specific list of stop words that were more relevant to Twitter. I utilized a list of texting and online chat abbreviations on Webopedia[25] which I filtered out of the 1500 most common features. Then, I used some of the options in the TweetTokenizer[26] function in NLTK to try and improve my model. The default of the "strip_handles" parameter in the TweetTokenizer function is False, but when changed to True, all Twitter handles are removed from the tweet. This removes the Twitter handles in replies to other tweets which are not necessary in this analysis. The default of the "red_length" parameter is False, but when changed to True, all words with more than three consecutive letters are trimmed down to a maximum of three (ex. "waaaaaayyyyyyy" changes to "waaayyy"). This was useful in this analysis as it increased the frequency of some terms and caused them to become common features. Finally, I converted all of the words to lowercase and removed all hyperlinks that would not be helpful in the analysis. I also tried removing hashtags, emoticons, punctuation, and at-symbols, but this led to decreased accuracy in my model. At the end of this preprocessing stage, my model was 58% accurate in predicting the sentiment of a tweet, which is very good, but could be improved even more.

In the experimental portion of my feature engineering, I utilized several popular sentiment lexicons used in Natural Language Processing. I first used the Subjectivity lexicon which analyzes the polarity of subjective expressions. Expressions were labeled as positive, negative or neutral by annotators in the Subjectivity lexicon. I added a "positivecount", "negativecount", and "neutralcount" feature for each word in a given tweet that was also in the Subjectivity lexicon at the noted polarity. Next, I used the LIWC (Linguistic Inquiry and Word Count) Lexicon which lists positive and negative words in a similar fashion to the Bing Lexicon. I did this to catch some of the words that were not in the Subjectivity lexicon. If the word was contained in the positive lexicon, the "positivecount" feature would increase by one, while one would be added to the "negativecount" for the negative words. Finally, I used the VADER[27] (Valence Aware Dictionary and sEntiment Reasoner) Lexicon which specializes in sentiment analysis in social media. In Python, VADER can be used to return a positive, negative, neutral, and compound (aggregate score of sentiment in other popular lexicons) score for a given word or phrase. The sentiment scores for positive, negative, and neutral all add up to one. VADER is useful for Twitter data because it deals with emoticons, punctuation, and slang terms. For example, for the emoticon :) a positive sentiment score of 1 is returned, and for the emoticon :(a negative sentiment score of 1 is returned. Also, the sentiment score for punctuation can differ depending on the number of punctuation symbols in a row (ex. ! is different than !!!). Finally, slang terms that would not be recognized in a sentiment lexicon like Subjectivity or LIWC would be picked up in the VADER lexicon. I added the VADER lexicon to the feature set function by taking the highest sentiment score for each word in a tweet and adding one to the respective feature set ("positivecount", "negativecount", "neutralcount"). For example, the sentiment scores for the word "happy" are {'neg' : 0.0, 'neu' : 0.0, 'pos' : 1.0}, so one would be added to the positive count feature set. I also tried to use part of speech feature sets, but part of speech tagging is difficult with Twitter data. I also attempted to use bigram feature sets, but there were not enough common bigrams to have a significant effect.

In my preliminary analysis of the model, I split the tagged tweets so 90% were in the training set and 10% were in the test set. I then ran the Naïve Bayes Classifier on the training data and evaluated the 30 most informative features.

Table 5: *The 30 most informative features in the Naïve Bayes Classification model*

Feature	Relationship	Feature	Relationship
f***	neg : neu = 66.4 : 1	negativecount = 9	neg : neu = 19.5 : 1
:)	pos : neg = 51.4 : 1	thanks	pos : neu = 19.4 : 1
fun	pos : neu = 49.5 : 1	missing	neg : pos = 18.3 : 1
excited	pos : neu = 47.2 : 1	negativecount = 10	neg : neu = 18.1 : 1
sad	neg : neu = 43.9 : 1	cool	pos : neu = 16.8 : 1
sorry	neg : neu = 43.9 : 1	cant	pos : neu = 15.1 : 1
positivecount = 12	pos : neu = 42.5 : 1	brilliant	pos : neu = 15.1 : 1
great	pos : neu = 38.9 : 1	funny	pos : neu = 14.2 : 1
happy	pos : neu = 35.7 : 1	exciting	pos : neu = 14.2 : 1
amazing	pos : neu = 35.7 : 1	wrong	neg : neu = 14.2 : 1
luck	pos : neu = 28.8 : 1	hate	neg : neu = 13.9 : 1
thank	pos : neu = 25.4 : 1	negativecount = 8	neg : pos = 13.8 : 1
:(	neg : pos = 25.3 : 1	can't	neg : neu = 13.4 : 1
injury	neg : pos = 23.5 : 1	interesting	pos : neu = 13.3 : 1
awesome	pos : neu = 21.9 : 1	pavol	neg : pos = 13.1 : 1

Some of the top features are what we would expect in such a classifier. The :) emoticon and words like "fun", "excited", and "great" led to tweets with more positive sentiment than negative. Words like the f-word, "sad", and "sorry" led to more negative sentiment. Tweets with 12 positive words were usually tagged as positive and tweets with 8-10 negative words were usually tagged as negative. This shows that although there are certain words with strong sentiment that are good predictors of overall tweet sentiment, the number of positive or negative words in a tweet can also be a tool to evaluate sentiment.

I then used cross-validation to find precision (positive predicted value), recall (true positive rate), and F1 (balance between precision and recall) scores. Using 10 folds, I found a mean accuracy of 0.687 or 68.7%.

Table 6: *Sentiment model accuracy scores*

Sentiment Category	Precision	Recall	F1
Neutral	0.693	0.660	0.676
Negative	0.490	0.500	0.495
Positive	0.658	0.695	0.676

As can be seen, the classifier performed worse on tweets with negative sentiment than those with positive or neutral. This is most likely the case because it is harder to pick up on sarcasm in tweets. Overall, the model performed very well and can be used to predict the sentiment of the tweets about VAR in the English Premier League.

After testing the model using cross validation, I wanted to test it on the game-specific tweets from the English Premier League to evaluate the predicted sentiment. *Figure 51* shows that the model predicted a large number of neutral sentiment tweets during the season, followed by negative sentiment and finally positive sentiment. The results are as expected as many tweets about the use of VAR simply state that a decision was made without any additional opinions about the technology. The sentiment model, like most models, struggles to pick up sarcasm in a tweet, and may classify these tweets as neutral because they have positive text with a negative underlying message.

Figure 51: *Predicted sentiment for every game-specific tweet from the first half of the 2019-2020 English Premier League season*

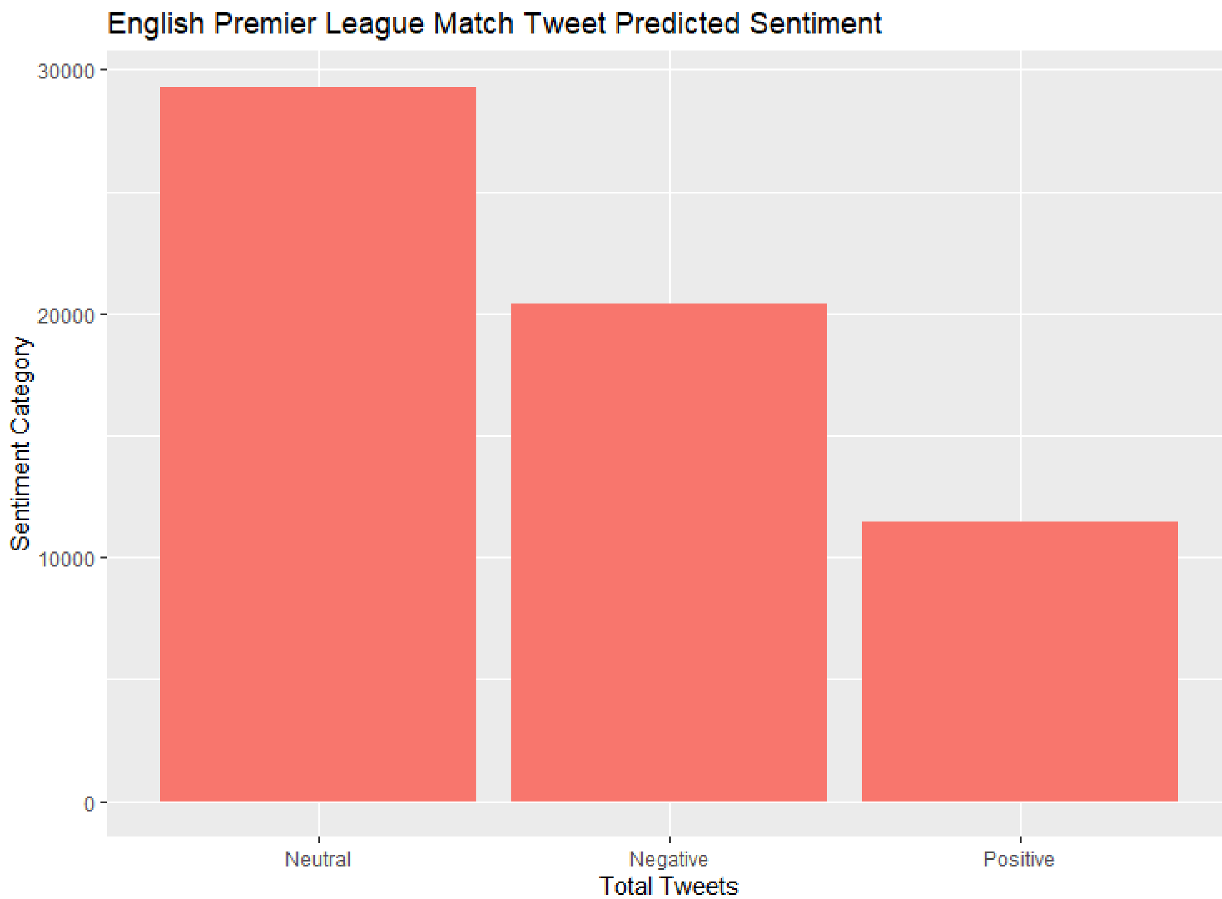


Table 7 shows several examples of tweets that were classified by the model. The positive, neutral, and negative tweet examples seem like they were correctly classified given the context of the tweets. The positive tweet says that VAR has been used correctly in the Premier League and shows belief in the technology. The neutral tweet simply states that a VAR decision has been made giving a goal to Arsenal and has no sentiment associated with it. The negative tweet states that the VAR decision was incorrect and that particular part of the game does not need to be officiated so closely. The table also shows examples of team fans that may be biased in their sentiment towards VAR in given scenarios. The positive home fan praises the use of VAR which has ruled out a goal for Wolves and keeps the match at 0-0. The negative away fan is disgusted by the use of VAR in the Manchester City vs. Tottenham match that ended level after a Gabriel Jesus winner was ruled out by the technology. Both tweets show sentiment toward VAR in support of their favorite teams and does not tell us if the call on the field was right or wrong. Finally, the positive sarcastic tweet shows where the model can incorrectly classify a tweet. The tweet says VAR is brilliant, but that opinion may change several times throughout a match. The

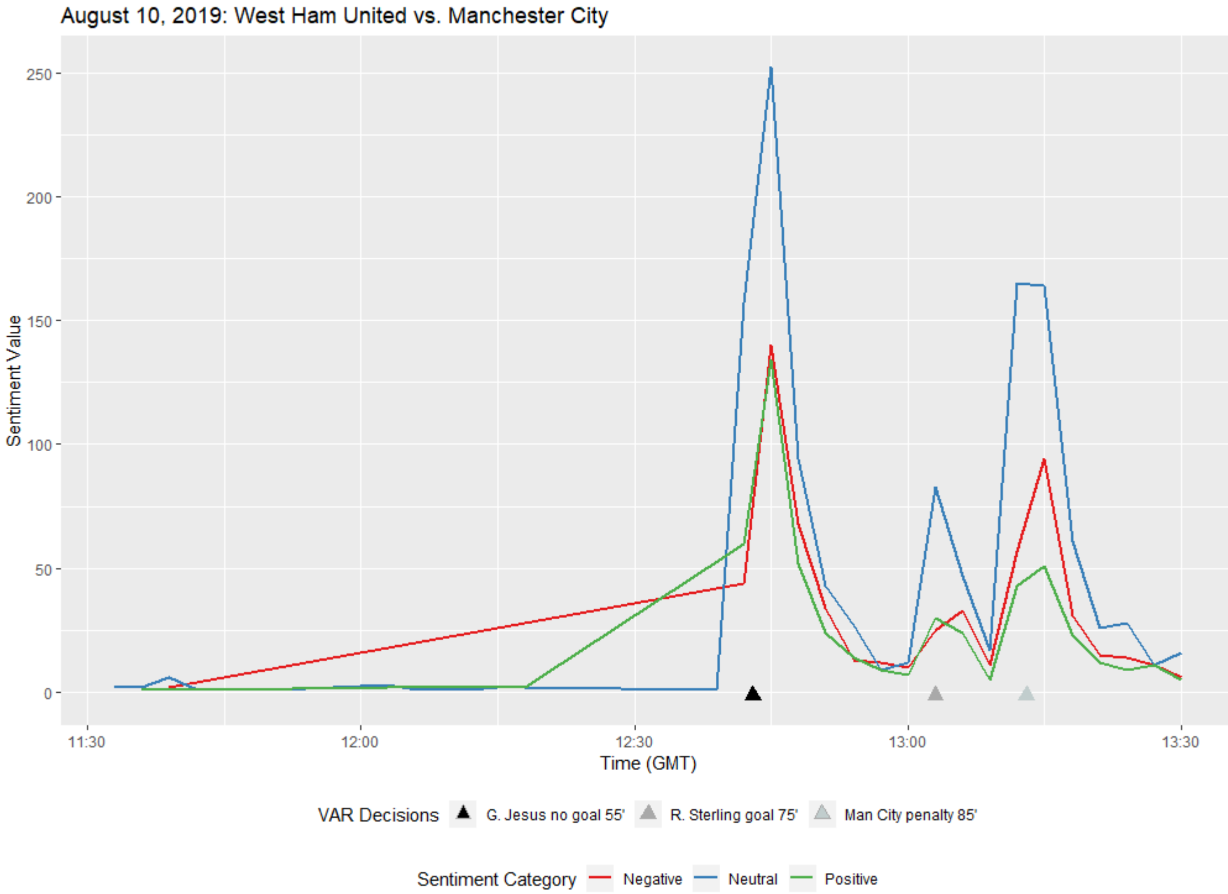
tweet seems to have positive sentiment towards VAR, but in reality, it would be better classified as negative or neutral because of the context surrounding the tweet.

Table 7: *Examples of classified tweets*

Sentiment	**Tweet**
Positive	"I actually think they've done VAR right in the PL. You can tell they've done it with speed in mind. Not perfect but a start. #TOTAVL"
Positive Home Fan	"We love you VAR we do We love you VAR we do We love you VAR we do Oh VAR we love you #LCFC #LEIWOL"
Positive Sarcasm	"#VAR is clearly brilliant *#LEIWOL*opinion may change several times a game"
Neutral	"VAR say goal, 1-1 #MUNARS"
Negative	"Ew. Not fond of the VAR involvement in the penalty. Doesn’t seem like that’s an area of the game crying for oversight. #WHUMCI"
Negative Away Fan	"Football is dead #VAR #MCITOT VAR must be removed immediately @ManCity board you quickly respond and tackle the FA about what happened and what we can expect more rubbish calls against City. I said and I will say it again, VAR is the perfect weapon to stop Manchester City."

I also wanted to take a look at the breakdown of tweets during a match to see if there was more negative sentiment after VAR decisions. If we look at the Manchester City vs. West Ham game once more, we can see that there were large spikes in neutral sentiment after each VAR decision. This is expected as there would be many tweets reporting that a VAR call was made, but not as many criticizing or praising the technology. It can be seen that the first decision was split evenly in terms of positive and negative sentiment suggesting that fans were possibly undecided on whether or not the decision was correct which is a bit different than what the NRC Lexicon suggested. The decision was a very tight offsides call and was the first VAR decision in Premier League history, which may have led to the indecision among fans. The second decision was very similar to the first, but the third call was met with more negative sentiment than positive. Fans were not pleased with the third call because of the nature of the decision which led to the Manchester City penalty being retaken for a minor infringement. This information can be very useful to the Premier League as they continue to improve the use of VAR. If fans do not want to see VAR used to retake penalties for slight violations of the rules, the Premier League may want to rethink their use of VAR in such situations.

Figure 52: *Predicted sentiment throughout the Manchester City vs. West Ham United match*



Future Work

In future work with this project, I will continue to analyze the sentiment associated with VAR decisions in professional soccer. Through the use of my own sentiment model and the Bing and NRC lexicons, I hope to be able to classify VAR decisions into good and bad calls to evaluate how the Premier League can change the review system to improve the game of soccer. I will then be able to analyze specific VAR decisions that were deemed incorrect and investigate any similar traits among them. I would also like to analyze VAR decisions in other top professional leagues and competitions such as the Champions League, World Cup, and MLS to see how the use of VAR differs. After finding that the better team tends to receive the benefit of more VAR decisions, I would also like to see if the team that spends more money on transfers tends to get the benefit of VAR. Do highly successful teams that are big spenders such as Manchester City, PSG, and Barcelona have an advantage over smaller clubs? I would also like to analyze VAR officials specifically rather than the head official to see if some VAR referees tend to have different biases than the head referee. Finally, I hope in the future to have access to tracking data to improve my win probability model. I would like to include other factors that I believe are important such as possession and attacking passes, but am limited with the use of commentary data. I believe having data on a second-by-second basis for a match would only improve my model.

Conclusion

In conclusion, I have utilized Twitter and commentary data to analyze Video Assistant Referee decisions as well as fan sentiment associated with such decisions in professional soccer. I found that the introduction of VAR into professional soccer has decreased the number of fouls and offsides calls in the game, but has increased the number of penalty kicks. There has been an upward trend in terms of penalties given in professional soccer and VAR has not impacted this trend. I also discovered that VAR decisions tend to favor the home team slightly, but tend to favor the team with the better end-of-season record more significantly. Referees seem to feel that the better team will be more greatly affected by a VAR decision and make fewer calls that will benefit the worse team. Also, VAR calls are the most important decisions made during the match in terms of win probability added. Goals, penalty kicks, and red cards impact the game the most significantly and change the course of the match most dramatically. I also found that VAR decisions during professional soccer matches does in fact lead to more tweets concerning VAR. As the number of VAR decisions in a match increases, we would generally expect the number of tweets about VAR to increase as well. I also found that the sentiment of tweets about VAR decisions is more negative rather than positive or neutral. Through the use of the Bing and NRC Lexicons it is clear that fans show more negativity towards the use of VAR than positivity or indifference. I was also successful in creating a model to predict the sentiment associated with a given tweet concerning VAR through the use of Natural Language Processing. The model helped to emphasize the negative nature of tweets about VAR in the English Premier League.

By looking at tweets at a game-level, patterns arose about the negative sentiment associated with VAR. Many of the tweets with high levels of negative sentiment occurred in three aspects of the game. First, fans were upset with some of the extremely tight offsides decisions such as the ones shown in *Figure 1*. Offsides has been classified by the Premier League as "black and white" decisions that will not be upheld due to a lack of "clear and obvious" evidence. As a result, attackers have been called offsides by less than three centimeters. Although the technology used by VAR is advanced and technically sound in terms of determining the body position of players in three dimensions, the frame rate of the cameras used by VAR will always lead to some margin of error. The determination of when the ball was kicked as well as the furthest point of the player is a bit subjective, and small differences in the offsides line placement can rule out important goals. Second, fans were not pleased with decisions that involved the retake of a penalty due to early entry into the penalty box by a defender or the goalkeeper being off his line. Such decisions were rarely made before the introduction of VAR because of the difficulty referees had watching every player during a penalty kick. Fans believe this type of decision, much like a tight offsides call, is a case of "over-refereeing" the game. Soccer purists would much rather a call not be made just as it was before the introduction of VAR. Even though the call may be the correct decision, fans are discouraged by the extra time wasted to check VAR on these tight decisions that may not have a big impact on the play. Finally, fans have not been happy with the consistency of VAR on calling penalties. Some key penalties in big matches were not called during the first half of the season that led to supporters criticizing the technology. This subjectivity is part of the VAR technology and will continue to be a part of the game. However, referees must stay consistent with their use of VAR on penalties throughout the season to avoid any potential bias.

My main suggestion to the Premier League and other professional soccer leagues using VAR would be to put regulations in place to limit the number of tight decisions being made in matches. One suggestion would be to change offsides decisions only if a "clear and obvious" error has been made. Liverpool manager Jürgen Klopp has proposed making the offsides line thicker to create a greater margin of error for the attacker. If a set margin of error is listed for offsides decisions, then there would be less controversy surrounding the calls. Another suggestion would be to have more leniency on early intrusion into the penalty box. If the player clearly enters the box early or the goalkeeper is well off his line before the kick is taken, then the penalty should be retaken, but if the encroachment is minor then a call should not be made. Finally, a time limit should be set on VAR decisions to keep the pace of play fast in professional soccer. For example, if a decision cannot be made in 30-45 seconds, then the call on the field should stand. Some may argue for a longer time frame in key scenarios, but if such a call takes longer than 30-45 seconds to make, then it most likely was called correctly on the field the first time. Such implementations to the VAR technology could satisfy the demands of fans and players alike and lead to improved ratings for the Premier League. VAR technology has already improved the game of soccer, and by repairing the flaws in the system, it will continue to help the beautiful game.

References

1. Goff, S. (2019, June 26). Analysis | VAR is working at World Cup, FIFA says. That's not wrong, but it could work better. Retrieved September 10, 2019, from https://www.washingtonpost.com/sports/2019/06/26/var-is-working-world-cup-fifa-says-thats-not-wrong-it-could-work-better/?noredirect=on.
2. Bartiromo, A. (2019, June 13). The Impact of VAR on Penalties. Retrieved September 10, 2019, from https://www.americansocceranalysis.com/home/2019/6/6/the-impact-of-var-on-penalties.
3. Lago-Peñas, C., Ezequiel, R., & Anto, K. (2019, July 23). How does Video Assistant Referee (VAR) modify the game in elite soccer? Retrieved October 11, 2019, from https://www.tandfonline.com/doi/full/10.1080/24748668.2019.1646521.
4. Johnson, D. (2018, July 15). VAR at the World Cup: Assessing every major decision in Russia. Retrieved September 10, 2019, from https://www.espn.com/soccer/fifa-world-cup/4/blog/post/3533707/var-at-the-world-cup-a-timeline-of-the-tournament.
5. Johnson, D. (2019, August 10). VAR in the Premier League: The big decisions explained. Retrieved August 13, 2019, from https://www.espn.com/soccer/english-premier-league/story/3917260/var-in-the-premier-league-the-big-decisions-explained.
6. VAR in the Premier League: How did first weekend go for technology in top flight? (2019, August 11). Retrieved August 13, 2019, from https://www.bbc.com/sport/football/49307496.
7. Collins, H., & Evans, R. (2008, May 22). You cannot be serious! Public understanding of technology with special reference to "Hawk-Eye" - Harry Collins, Robert Evans, 2008. Retrieved September 12, 2019, from https://journals.sagepub.com/doi/pdf/10.1177/0963662508093370?casa_token=BBf9rZRTT_AAAAAA:a4bQ9hHiEnKD6-PT78jzV3xoOPp-nhewuVSoQhdm1L2GqMEQ4RzLhhqUDZK-8X6KW22BibAaRc0.
8. Collins, H., & Evans, R. (2011, July 29). Sport-decision aids and the "CSI-effect": Why cricket uses Hawk-Eye well and tennis uses it badly. Retrieved September 12, 2019, from https://journals.sagepub.com/doi/pdf/10.1177/0963662511407991?casa_token=h2AUNS

C4Ve0AAAAA:tU4n1aR48VcNzMmDXPFdhUnWGA--TyX5Gp2G70ahVThvVB03bs3rft_-wNMqXOyjC0hxpAXdJtc.

9. Gilis, B., Helsen, W., Catteeuw, P., Van Roie, E., & Wagemans, J. (2008, December 18). Interpretation and application of the offside law by expert assistant referees: Perception of spatial positions in complex dynamic events on and off the field. Retrieved September 12, 2019, from https://www.tandfonline.com/doi/full/10.1080/02640410802702178.
10. Oudejans, R. R. D., Verheijen, R., Bakker, F. C., Gerrits, J. C., Steinbrückner, M., & Beek, P. J. (2000, March 2). Errors in judging 'offside' in football. Retrieved September 12, 2019, from https://www.researchgate.net/publication/12600290_Errors_in_judging_'offside'_in_football.
11. Mallo, J., Gonzalez Frutos, P., Juárez, D., & Navarro, E. (2012, August 6). Effect of positioning on the accuracy of decision making of association football top-class referees and assistant referees during competitive matches. Retrieved September 12, 2019, from https://shapeamerica.tandfonline.com/doi/abs/10.1080/02640414.2012.711485?casa_token=qOW_tqPOa_kAAAAA:jnVCju_e25BVwax7Un3x4CPRK2lJXSZ-3g-cAdrE81mFzSpHqSgToQ8F2A1sbQlZ7y3NeFjOZZ4#.XYgB5UZKg2x.
12. D'Ottavio, S., & Castagna, C. (2001, January 1). Analysis of Match Play Activities in Elite Soccer Referees During Actual Match Play. Retrieved September 23, 2019, from https://www.researchgate.net/profile/Stefano_Dottavio/publication/256615318_Analysis_of_Match_Play_Activities_in_Elite_Soccer_Referees_During_Actual_Match_Play/links/0deec523802a01cee7000000.pdf.
13. Yu, Y., & Wang, X. (2015, February 20). World Cup 2014 in the Twitter World: A big data analysis of sentiments in U.S. sports fans' tweets. Retrieved October 3, 2019, from https://www.sciencedirect.com/science/article/pii/S074756321500103X.
14. Corney, D., Martin, C., & Goker, A. (2014, April 13). Spot the Ball: Detecting Sports Events on Twitter. Retrieved October 3, 2019, from http://dcorney.com/papers/ECIR_SpotTheBall.pdf.
15. Gratch, J., Lucas, G., Malandrakis, N., Szablowski, E., Fessler, E., & Nichols, J. (2015, September 21). GOAALLL!: Using Sentiment in the World Cup to Explore Theories of Emotion. Retrieved October 5, 2019, from http://ict.usc.edu/pubs/GOAALLL! Using Sentiment in the World Cup to Explore Theories of Emotion.pdf.
16. Petersen-Wagner, R., & Ludvigsen, J. A. (2019, July 2). VAR (video-assistant-referee) as Neo-Colonialism? Reading the FIFA 2018 Mens World Cup YouTube Video Comments. Retrieved October 11, 2019, from http://eprints.leedsbeckett.ac.uk/6029/.
17. Robberechts, P., Van Haaren, J., & Davis, J. (2019, June 12). Who Will Win It? An In-game Win Probability Model for Football. Retrieved October 5, 2019, from https://arxiv.org/pdf/1906.05029.pdf.
18. Boice, J. (2017, January 19). How Our Club Soccer Projections Work. Retrieved October 10, 2019, from https://fivethirtyeight.com/features/how-our-club-soccer-projections-work/.
19. Taspinar, A. (2019, November 4). TwitterScraper. Retrieved December 20, 2019, from https://github.com/taspinar/twitterscraper
20. Yurko, R. (2018, July 5). FCScrapR. Retrieved August 1, 2019, from https://github.com/ryurko/fcscrapR

21. Upon Further Review… (n.d.). Retrieved September 5, 2019, from https://operations.nfl.com/the-game/history-of-instant-replay/
22. O'Neill, C. (2019, October 21). Martin Atkinson Blamed for Controversial Man U Goal by Former Ref. Retrieved April 8, 2020, from https://www.liverpoolecho.co.uk/sport/football/football-news/liverpool-manchester-united-var-debate-17119828
23. Wasser, L., & Farmer, C. (2017, April 19). Sentiment Analysis of Colorado Flood Tweets in R. Retrieved January 8, 2020, from https://www.earthdatascience.org/courses/earth-analytics/get-data-using-apis/sentiment-analysis-of-twitter-data-r/
24. Mejova, Y., Srinivasan, P., & Boynton, B. (2013, February 1). GOP Primary Season on Twitter: "Popular" Political Sentiment in Social Media. Retrieved January 14, 2020, from https://dl.acm.org/doi/10.1145/2433396.2433463
25. Beal, V. (n.d.). Huge List of Texting and Online Chat Abbreviations. Retrieved September 15, 2019, from https://www.webopedia.com/quick_ref/textmessageabbreviations.asp
26. TweetTokenizer. (n.d.). Retrieved September 15, 2019, from https://kite.com/python/docs/nltk.TweetTokenizer
27. Hutto, C. J. (2019, September 6). VADER Sentiment Analysis. Retrieved September 15, 2019, from https://github.com/cjhutto/vaderSentiment