This is the WORST Fanbase in Premier League History

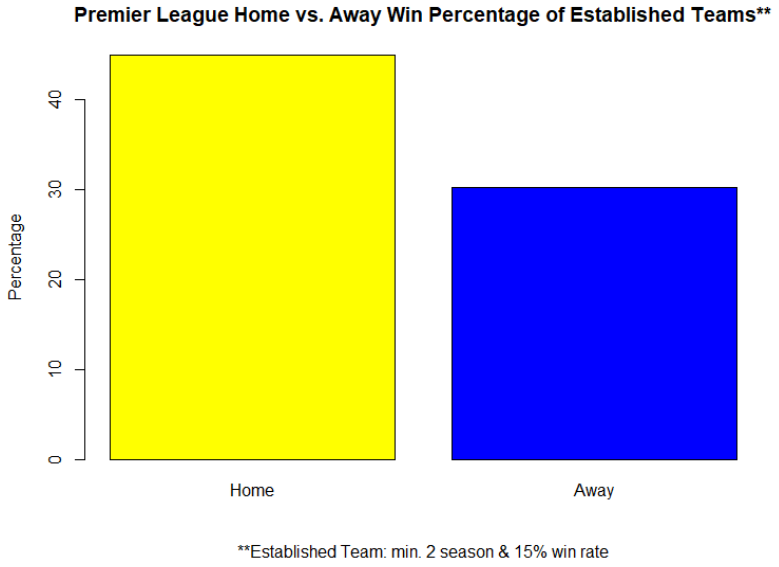
Alexander Mandryk *- Data 101 - Dr. Imielinski - 2/14/2024*

Since the introduction of regulated professional football in England, where teams like Sheffield United and Notts County represented the values of local communities, one question has plagued the sport: Is one fanbase objectively "*better*" than another?

For years, indicators such as ticket sales, consistency of away support, social media presence on various online communities, catchy yet occasionally contentious chants and songs, along with other factors have dominated the debate. Though, there is only one thing that can truly limit the biases and agendas against clubs and that is a statistical inference.

I retrieved over a decade of Premier League results containing the assigned referee, home team, number of cards given per team, goals, among other more important figures. This data provided the base to perform statistical analysis and mathematically prove which fanbase is the worst.

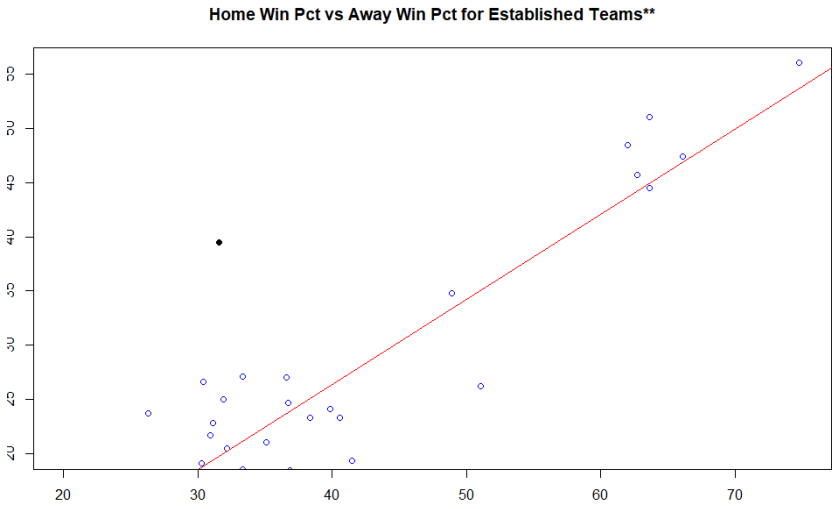
Though, in order to prevent so-called "yo-yo clubs" and other clubs with forgettable terms in the Premier League from skewing the data, I only looked at established clubs. An established Premier League team is one that has played a minimum of two seasons (totalling over 75 games) or one that maintained a win rate of at least 15%. This means that teams like Barnsley or Blackpool—who only spent one year in the Premier League—and a team like Sheffield United—which has a win rate of 13.3%—were all omitted from the dataset.



ARTICLE DATA SOURCE: Kaggle; Saif Uddin (linked here)

With this dataset of established Premier League sides, I analyzed the impact of home advantage created by the fan atmosphere. The results were *shocking*. All but one Premier League team had a home win percentage that bettered their away record. This unique tendency can be immediately spotted in the figure to the left.

The black dot clearly displays a team that completely avoids the trend displayed by the other 31 Premier League outfits. For a team to be the only one with a worse home record than away record, there must be a lackluster atmosphere generated by a far inferior group of fans.



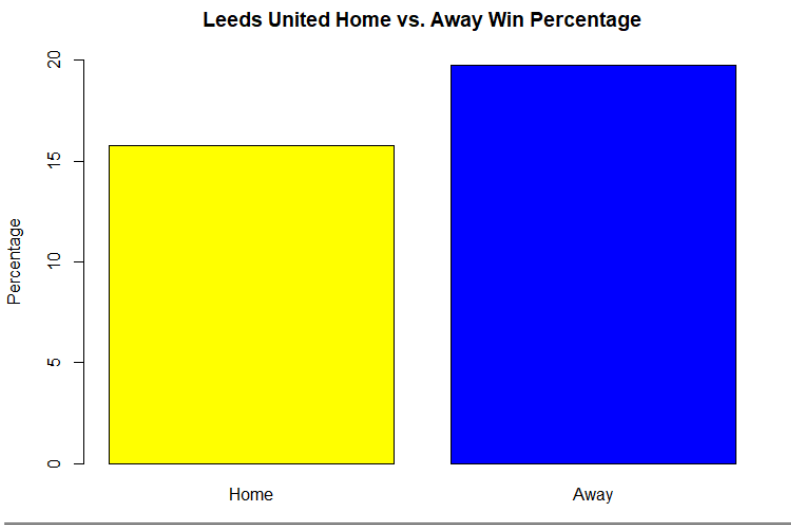
**Teams with >1 season & >15% win rate

Removed for readability:

Horizontal: Home Win Pct — *Vertical*: Away Win Pct

In the end, the data showed that the worst fanbase in Premier League history is **Leeds United Football Club**. This can be directly demonstrated by the complete dissimilarity between the graph on the right and the first graph displayed in this article. Premier League teams, rather, football teams as a whole are built around the notion of defending your fortress and backing your team when they play at home. Clearly, Leeds United fans are unable to live up to this footballing standard.

Although the overall plus-minus of home win percentage versus away win percentage vary from team to team, they are all positive. For example, Everton FC have a plus-minus of 12.61 which means their home win percentage is 12.61 percentage points greater than their away win percentage. This can be attributed to the true fortress mentality that comes from Goodison Park. At the other end of the spectrum lies Wigan Athletic, a team not renowned for winning nor having a strong home advantage but they still maintain the historic footballing trait of performing better at home rather than away.



Author's note: Despite all of the clear findings displayed from over a decades-worth of Premier League data, I must refute the results. I can back the mathematical process and the methods used to create the conclusions throughout but, as a supporter of Leeds United, I cannot support this finding. After spotting these findings, in the most stubborn Leeds United way possible, I try to change parameters and throw the blame on pure randomness. Unfortunately, the further I tested and scrutinized my hypothesis, the greater it became true. Alas, We Are Still MOT!

This is the most BIASED referee in the Premier League

Alexander Mandryk - *Data 101 - Dr. Imielinski - 2/14/2024*

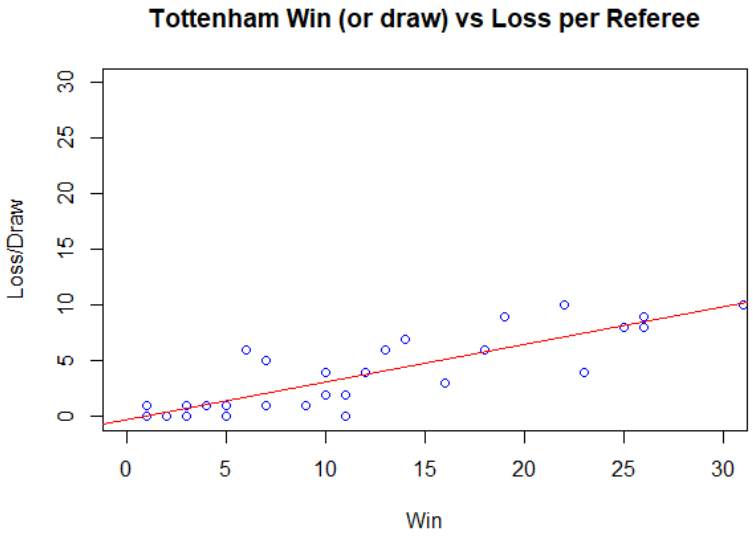
When football fans are asked to name the worst referees, it is safe to say that the list could go on forever. Though, we fans must watch and compile a list of the referees that seem to simply never get it right.

Using over a decade of Premier League games, I calculated the *worst* referee to grace the sporting competition.

When thinking of teams like Liverpool, Tottenham, and more recently Man City you can only help to think about the significantly greater winning percentage as compared to losing percentage. As a result, I compiled a list of Premier League teams and the wins and losses received by differing sporting officials.

Pictured to the right is over a decade of Tottenham Hotspurs games where each dot represents a referee. Horizontal axis is the number of wins from the respective referee and vertical represents the losses and draws. As shown by the data, there are no clear biases shown by any particular referee.

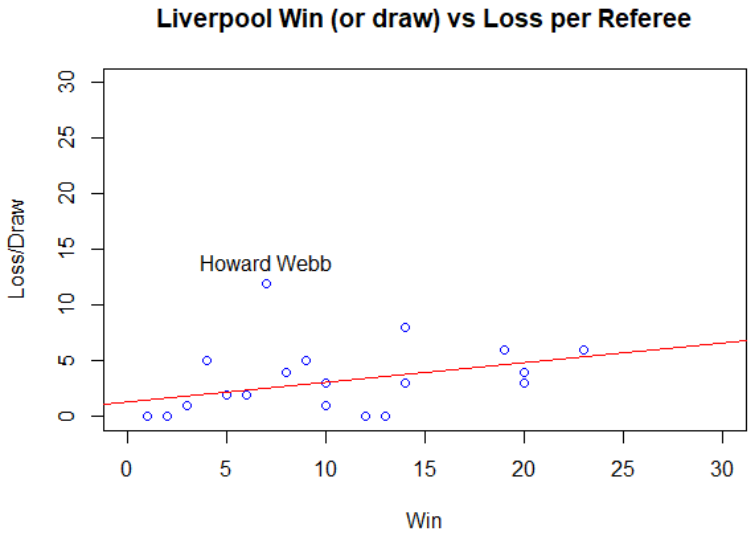
This distinct lack of bias cannot be observed by a particular referee.



ARTICLE DATA SOURCE: Kaggle; Saif Uddin (linked here)

For years, Liverpool's historic presence in the Premier League has led to officiating errors to be scrutinized on a disproportionate scale. Despite this constant feel of being done wrong, there has never been clarification about the promised virtue or possible bias contained within different referees. That is until now.

Looking at the graph below to the left, you can easily see a referee that stands out concretely. All other referees belong to a simply pattern that sees the number of wins generally outweigh all other outcomes. This referee is **Howard Webb**. Finally, a statistic that backs the claims of pleading Scousers.

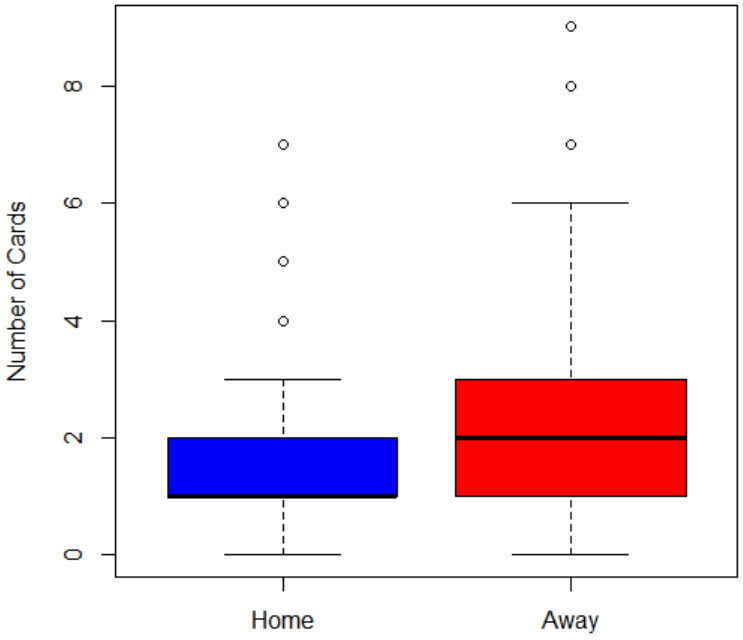


This outcome only seems to make more sense given the recent match against Arsenal or previous Merseyside Derbies where Webb was the official. Countless calls can be seen to go against the Reds.

Now, the most important question: What can be done about the evidently clear bias?

Other Findings

To begin, I created two vectors (home_cards and away_cards) that take the vectors of home and away cards (red and yellow). With these vectors I made a dataframe with a column of Cards and a column of Side which indicated whether it was home or away. From that I used tapply() to find the mean of cards with respect to the side. Home teams received an average of 1.56 cards per game whilst away teams received 1.83. Further the plot of the data shows a significant difference between home and away cards as the median, Q3, and max are all higher for away teams than home.



Prediction Model

Prediction: Man City will NOT lose a Premier League game if Jonathan Moss is officiating the game

To complete this prediction model, I created a subset of games where John Moss was the referee regardless of whether Man City was the home or away team. For viewing simplicity I utilized the dplyr library and shrunk the amount of columns in the data to three. With this new subset I created a vector that ran through the rows and made sure that Man City either:

- Won at home stadium ↠ HomeTeam==”Man City” & FTR==”H”
- Won at away stadium ↠ AwayTeam==”Man City” & FTR==”A”
- Drew ↠ FTR==”D”

This meant that all wins and draws were represented with TRUE and losses were FALSE. Following, I calculated the accuracy of my model by comparing it to a vector composed of TRUE (wins/draws). In the end, my prediction model was 96.6% accurate.

Hypothesis Tests

Alexander Mandryk *- Data 101 - Dr. Imielinski - 2/28/2024*

The following hypotheses are based on player statistics from the 2020/2021 Premier League Season (Linked).

Hypothesis 1: *Strong Alternative Hypothesis*

This is Why Leeds United Were REGULATED in 2022

Leeds United burst onto the Premier League scene in the 2020/2021 season following a sixteen year hiatus resulting from financial struggles. Unfortunately, this stardom was short lived following their relegation in the 2022/2023 season. This inability to maintain Premier League pedigree is incredibly common, though, what truly separates those that can establish themselves in the Premier League and those that are forever left to yo-yo between divisions? Following a tapply comparison of the number of interceptions completed by players with respect to the 20 different teams in the 2019/20 Premier League season, I found that Manchester City players averaged 98.041667 interceptions throughout the season and Leeds United averaged 5.583333. As a result, I had an alternative hypothesis that stated that Manchester City players complete more interceptions than Leeds United players; this rivals the null hypothesis that states that the number of interceptions completed by Leeds United players throughout the season is the same as Manchester City players. Since this comparison between the Club and Interceptions columns is the only hypothesis tested, the generic significance level of $\alpha = 0.05$ can be used. The permutation test with 10000 iterations and the Z-Test are as follows:

```
permutation_test(df[df$Position!="Goalkeeper",], 'Club', 'Interceptions', 10000, 'Leeds-United', 'Manchester-City')
```

```
z_test_from_data(df[df$Position!="Goalkeeper",], 'Club', 'Interceptions', 'Leeds-United', 'Manchester-City')
```

The permutation test yielded a p-value of 0.0005 and the Z-Test with a z-score of 3.544063 which corresponds to a p-value of 0.000197. Regardless of the chosen test the resulting p-value provided enough statistical evidence to reject the null hypothesis which claims a lack of difference between the number of interceptions completed by Leeds United and Manchester City players. This lack of interceptions completed by the Leeds United players can be connected to Leeds United's fate just a couple years later when they were relegated from England Premier Division. Clearly, the difference in defensive energy and strength is a deciding factor between England's best and those that struggle.

Hypothesis 2: *Close Call*

THIS is What Separates the Best and WORST Clubs in the Premier League

When comparing the aspirations of supporters of varying Premier League clubs, clubs like Manchester City, Liverpool, and Arsenal garner wishes for silverware and pole position finished whereas teams like Leeds United, Burnley, and Southampton simply hope to stay in the division. The real question is, other than points, what truly separates the English elite and those hoping to compete against them? Following a tapply comparison of the number of passes per match with respect to the 20 different teams in the 2019/20 Premier League season, I found that West Bromwich Albion players averaged 18.1625 passes per game and Arsenal averaged 35.194. As a result, I had an alternative hypothesis that stated that Arsenal players complete more passes per match than West Bromwich Albion players; this rivals the null hypothesis that

states that the number of passes completed by West Bromwich Albion players per match is the same as Arsenal players. Again, since this is the only comparison between the Club and Passes Per Match columns the significance level can be maintained at the generic $\alpha = 0.05$ value. The permutation test with 10000 iterations and the Z-Test are as follows

```
permutation_test(df, 'Club', 'Passes.per.match', 10000, 'West-Bromwich-Albion', 'Arsenal')

z_test_from_data(df, 'Club', 'Passes.per.match', West-Bromwich-Albion', 'Arsenal')
```

The permutation test yielded a p-value of 0.0463 and the Z-Test with a z-score of 1.69457 which corresponds to a p-value of 0.04508. Regardless of the chosen test the resulting p-value provided enough statistical evidence to reject the null hypothesis which claims a lack of difference between the number of passes per game completed by West Bromwich Albion and Arsenal players. Since West Bromwich Albion suffered relegation following the end of this season, the lack of ability to maintain possession and create passes could be attributed to this downfall. The inability to hold the ball and maintain good structure, resulting in high volume passing, appears to highlight the difference between those shooting for European football and those hoping to avoid Championship football.

Hypothesis 3: *Failed to Reject Null Hypothesis*

BREAKING: English Players Fail to Outshine Ivorians in the Premier League

Being based in England, English players are expected to provide a true showing and maintain the high standards required of a league that is credited as the best in the world. Though, they must fall short in some metrics. As a result, the question exists, in what areas do English players fail to stand out above other nations? Following a tapply comparison of the number of interceptions completed by players with respect to the numerous different nationalities present in the 2019/20 Premier League season, I found that players from Cote D'Ivoire (Ivory Coast) averaged 50.14286 interceptions throughout the season and English players averaged 80.5288. As a result, I had an alternative hypothesis that stated that English players complete more interceptions than Ivorian players; this rivals the null hypothesis that states that the number of interceptions completed by Ivorian players throughout the season is the same as English players. Since this comparison between the Nationality and Interceptions columns is the only hypothesis tested, the generic significance level of $\alpha = 0.05$ can be used. The permutation test with 10000 iterations and the Z-Test are as follows:

```
permutation_test(df[df$Position!="Goalkeeper",], 'Nationality', 'Interceptions', 10000, "Cote D'Ivoire", 'England')

z_test_from_data(df[df$Position!="Goalkeeper",], "Nationality', 'Interceptions', "Cote D'Ivoire", 'England')
```

The Z-Test yielded a z-score of 1.406715 which corresponds to a p-value of 0.079756 whilst the permutation test never went over 0.002. This means that the Z-Test does not find enough statistical evidence to reject the null hypothesis whilst the permutation test does. Regardless, the Z-Test shows that there is not enough statistical evidence to reject the null hypothesis which claims a lack of difference between the number of interceptions completed by Ivorian and English players. As a result, English players fail to outclass their Ivorian counterparts with regards to their defensive prowess shown by interceptions completed over a regular season.

Narrow Query: *eq2*

This is How French Defenders are Losing Their Impact in the Premier League

Comparing big chances created by nationality and position. Goalkeepers are omitted from this comparison since they have no statistics within this numerical variable. Mo and M are as follows:

Mo: `mean(bc$Big.chances.created)` = 8.171315

M: `mean(bc[bc$Nationality=='France' & bc$Position=='Defender',]$Big.chances.created)` = 3.461538

eq2 is satisfied since (8.171315 / 2) > 3.461538. The narrow query resulting from combining two separate categorical variable restrictions in a condition provides a small enough group that will significantly detract from the overall sample mean. The 'bc' dataframe is reduced from 502 players to just 13 when keeping only French Defenders.

BREAKING: Place Your Bet on This Team to Win When They Play at Home

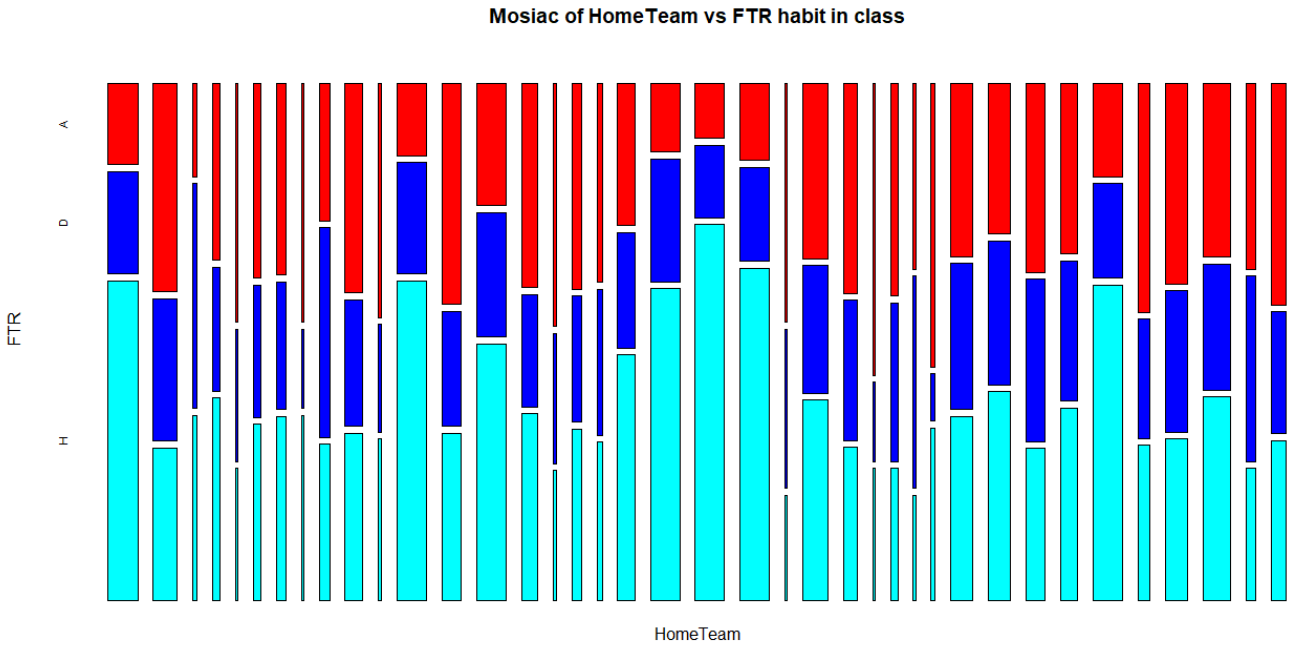
Alexander Mandryk - *Data 101 - Dr. Imielinski - 3/24/2024*

ARTICLE DATA SOURCE: Kaggle: Saif Uddin (linked here)

Is it even possible to outsmart the bookies? All of the statistical analysis, predictive models, and skewed odds that seem to favor betting companies endlessly might all be for nought. One major outlier exists that may guarantee prosperous dividends despite the slanted nature of major sports gambling as a whole.

Picture this: a team that persistently laughs in the face of probability, that scoffs at the notion of defeat. Is it a phenomenon? It is safe to say it will leave bookies scratching their heads and quaking in their boots. Here's the kicker: this anomaly isn't just a fluke. Rather, it's a pattern that appears to be as reliable as the sunrise.

Pictured below is a graph depicting the results of home teams throughout the history of the Premier League. Clearly, one team near the middle sets itself above the rest with a magnificent home record. Anomaly? Reliable? This team may be the solution to your negative betting balance.



Red: Home Defeat; Blue: Home Draw; Cyan: Home Victory

What team could possibly shine so unbelievably above the rest of the league so consistently throughout the years? The team your DraftKings account requires so desperately is **Manchester City Football Club**. Who's surprised? We most definitely saw this coming, too.

Now, you may be wondering how this mystery unfolded, but statistical evidence proves the benefit of putting your money on Manchester City. To begin, the question that must be asked is: *What are the odds of the home team winning if Man City are playing at home?*

Now, we have our observation that Manchester City are the home team along with a belief that the home team successfully won the match. The prior odds of this belief end up being approximately 82% (0.82) which shows a great skew towards the home team. Additionally, the True Positive and False Positive are 0.02, respectively. This clearly shows the dependency on Manchester City finding some way to win every game they appear to compete in.

These two statistics lead to a Likelihood Ratio of an astounding 4.0 which leads to Posterior Odds of 3.28 and a Posterior of 0.77. Such statistics leaning so

heavily in Manchester City FC must have convinced you by now, right?

If not, here is more financial evidence that will swing you into believing this statistical truth. At this moment in time, Manchester City are placed at +120 odds against Wolverhampton Wanderers when they face at the Etihad (Oddspedia). Utilizing Wolves' middle-of-the-park status allows this odd to be used as an interim average.

Of course this figure will fluctuate depending on the power of the team therefore using a mediocre team like Wolves will be sufficient.

Given a bettor placed $100 on every home game displayed in our statistical database at the average +120, they would complete their betting spree with $15,000. The numbers simply speak for themselves, do you listen or continue playing into the bookies hands?

This Team Has the Worst Traveling Fans in Premier League History

Alexander Mandryk - *Data 101 - Dr. Imielinski - 3/24/2024*

ARTICLE DATA SOURCE: Kaggle: Saif Uddin (linked here)

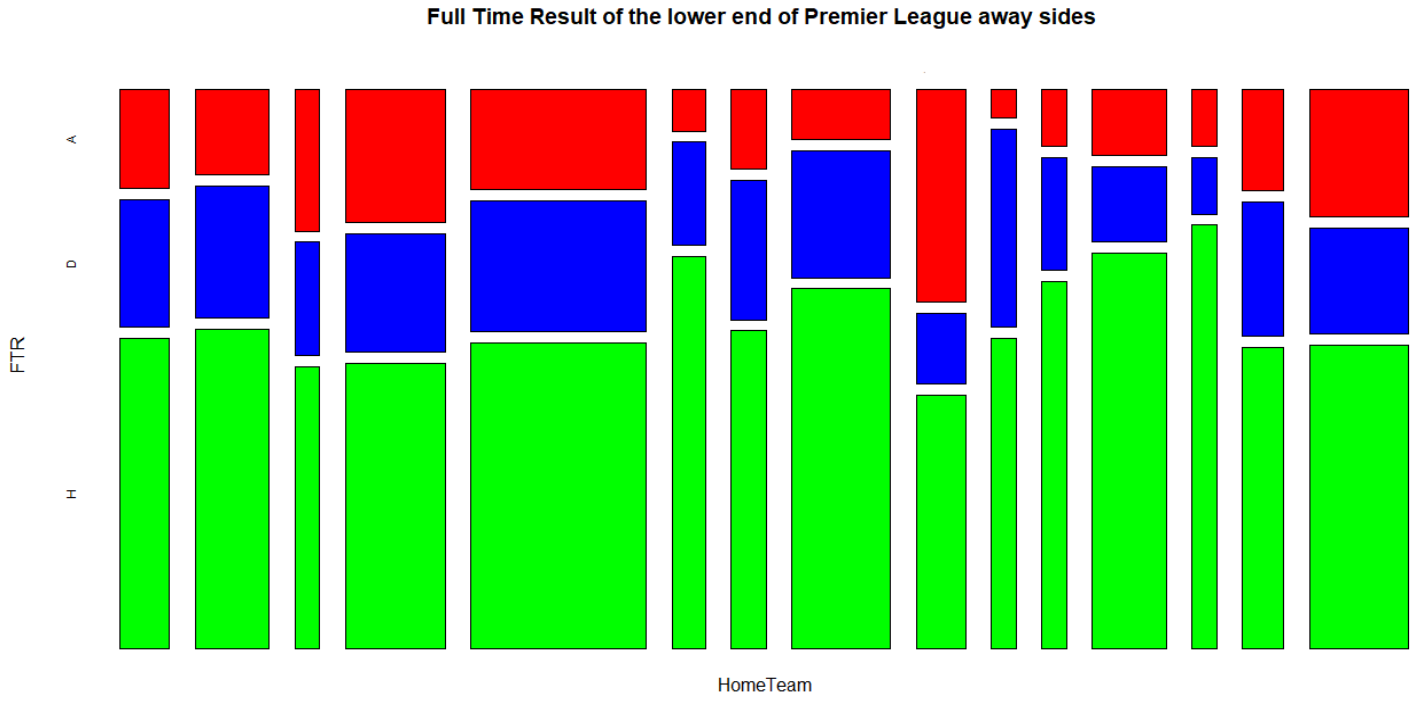
In the cutthroat world of football, every minute detail matters, even the mass of fans within the stands. Whilst some teams boast of fervent support no matter where they play—Leeds United, Everton, Liverpool—others seem to falter when they leave the comforts of their home turf.

They only started singing when they scored!

It's all gone quiet over there!

These are only a handful of the critiques given to poor traveling fanbases. In a league where each individual chance of an advantage counts—incoming FFP jokes about Everton, Leicester, and Forest—the traveling prowess of a team's supporters can make or break their success away from home.

Enter *this* Premier League 'powerhouse' that has statistically struggled throughout its history to find its footing when playing away matches. But just how much does their lack of support on the road impact their chances of victory? We shall dive into all of the numbers to unveil the single worst group of away fans to grace the Premier League.



Red: Away Victory; Blue: Away Draw; Cyan: Away Defeat

The graph above displays some of the teams when playing away from home. Yet again, near the far right there is one team that stands out above the rest of the pack. Is it Wigan? QPR? Bournemouth? What team could have fans this abysmal on the road.

According to recent data analysis completed at our in-house facilities, **Reading FC** present a prior odds of Away Defeat at a staggering 82%. A bookies nightmare? A supporter's single worst fear. When Reading are on the road, the True Positive of an away victory is a measly 0.7% (0.007) with a False Positive of 0.16% (0.0016). Even the occasional false win is nye-on impossible to come by.

As a result, the Likelihood Ratio is 4.59 with Posterior Odds of 3.75 and Posterior 0.79. Unbelievable. Reading fans, any words? Where were

you for the cold, rainy Tuesday night in Stoke?

Okay, what does this mean for Reading? It is perfectly clear that their struggles away from home extend beyond the pitch and into the stands. While their home matches offered some glimmer of hope (for 2 seasons that is) it is their road performances that truly defined their pitiful trip to the Premier League.

Had they fixed their traveling support or even magically reversed their away fortunes, they may have maintained their Premier League status. Though, for now they struggle to present their prowess whilst battling for promotion in EFL League One following a horrible relegation battle last season in the Championship. As a cry from fans of the beautiful game, Reading fans must fix this gap in their team and return to the place they rightfully belong.