DS 220 Project 2
NFL Data Analysis Group Project

Group Members: Lucas Sadoulet and Cameron Willis

Overview:

This project overviews NFL teams between the period of 1999-2019, covering the last 20 years of statistics. Generally reviewing major stats such as touchdowns, wins, interceptions, fumbles, and overall success of teams.

Motivation:

Athletic analytics is a well documented practice, as professional sports is one of the consistently and most thoroughly documented industries in data science. This means we can conduct data analytics across a fairly significant period of time to map trends. We are able to use this data to verify trends and corroborate stereotypes surrounding teams and individuals.

Problem Space:

This kaggle dataset outlines all 32 NFL teams and their aggregate seasonal stats for all years between 1999 - 2019. We had to conduct some preprocessing before answering the questions, the most significant being dividing up the code, which consisted of year and team code into two separate columns

Questions to Address:

Some trends from fans are more general, speaking on a team or sport as a whole. Some questions we wanted to look at were these changes over time across all teams, like which teams have improved and worsen the most since 1999, and does this align with our understanding of these teams? Other questions in this category include are defenses growing stronger over time. We also wanted to test a significant issue we heard brought up, that the sport has gotten more violent in recent years, which would be a significant issue if our data confirm such.

Another group of questions is breaking down a team's groups, between the offense, the defense, and the quarterback. We used differing measures to answer all three and tie into the defining characteristics of a team which fans are able to recall. This also allows us to compare teams and evaluate their trends over time.

For some of our questions, we wanted to target certain events and teams, attempting to verify preceded changes with our data. An example is the perception that Kansas City Chiefs improved significantly after hiring a new coach in 2013, which we aim to back up or refute with the data. Another question of this type is when the best streak of a given team was, a time most fans would remember of their team, and compare that to the statistical best streak.

Data:

Understanding of Dataset:

Our dataset is made up of rows, which each represent a team and years given as a teams code, in the format "abc1999". Each row is made of four groupings of data points, the team's total and average data points, and the opposition's of all of the games total and average statistics. This aggregate is useful for the problems we are looking at over the individual statistics of each game since the data is scaled multiple, not a singular, season. Despite the fact that teams change their names often, when put under new management, the creator of the dataset replaces all previous names of the teams with the current (as of 2019) name, allowing us to avoid manually fitting the data.

Exploratory Data Analysis:

We use the EDA to gain a better understanding of our data as a whole and this allows us to confirm the data with our basic understanding of the sport as a whole. We also used this EDA to ensure we understood the home vs opposition variables and that they align across seasons. Finally we used EDA to confirm that our dataset was complete, that teams used the same names and were present in all years, which was already accounted for as some teams have changed their indicator over our data's timespan. We did find that some of th earlier years (prior to 2001) were missing teams, and for more of our analysis in that time span we removed teams which hadn't been present during that period

Questions:

1) Which team has improved the most since their first appearance?

We took the teams average wins during the 1999-2003 timeframe and compared that to their wins from 2015-2019. The goal was to find which teams were most improved over the time period. The New England Patriots came out as the most improved team, with a lot of validity behind this. The margian they improved by was exactly 3 games per season. This helped the team reach the playoffs every season and win an outstanding 3 Super Bowls during the period from 2015-2019. The 1999-2003 period was the beginning of the Tom Brady dynasty being built, and the data reflects the success overtime during the 2015-2019 period. This team's defense during that period allowed for 282.2 total season points. On the other hand, The Tampa Bay Buccaneers experienced a falling off from the start of the decade, being super bowl champions and having a well established offense and their outstanding defense including Derrick Brooks, Simeon Rice, and Warren Sapp. Tampa Bay later struggled with their defense heavily allowing opposing teams to rack up a total of 415.8 points against them.

2) How did the Kansas City Chiefs stats change with a new coach in 2013?

There is a common perception that the Kansas City Chiefs had one of the largest rebounds as a team as they replaced their coach for the 2013 season and went on to do significantly better. We wanted to see if we could replicate this change with our dataset. We took 6 years after and before 2013, to balance out the data, and were looking for a jump in the data. In addition, we also graphed the wins over time to see if the data also showed their.

During the time period from 2006-2013: the Kansas City Chiefs won an average 5.43 games, which is a losing record per year on average. After the coaching change to Andy Reid, the Chiefs won an average of 11 games per season which is a winning record per season. Overall, with Andy Reid being introduced they experienced an increase of 5.57 more games won per season.

Number of Wins by Year for Kansas City Chiefs
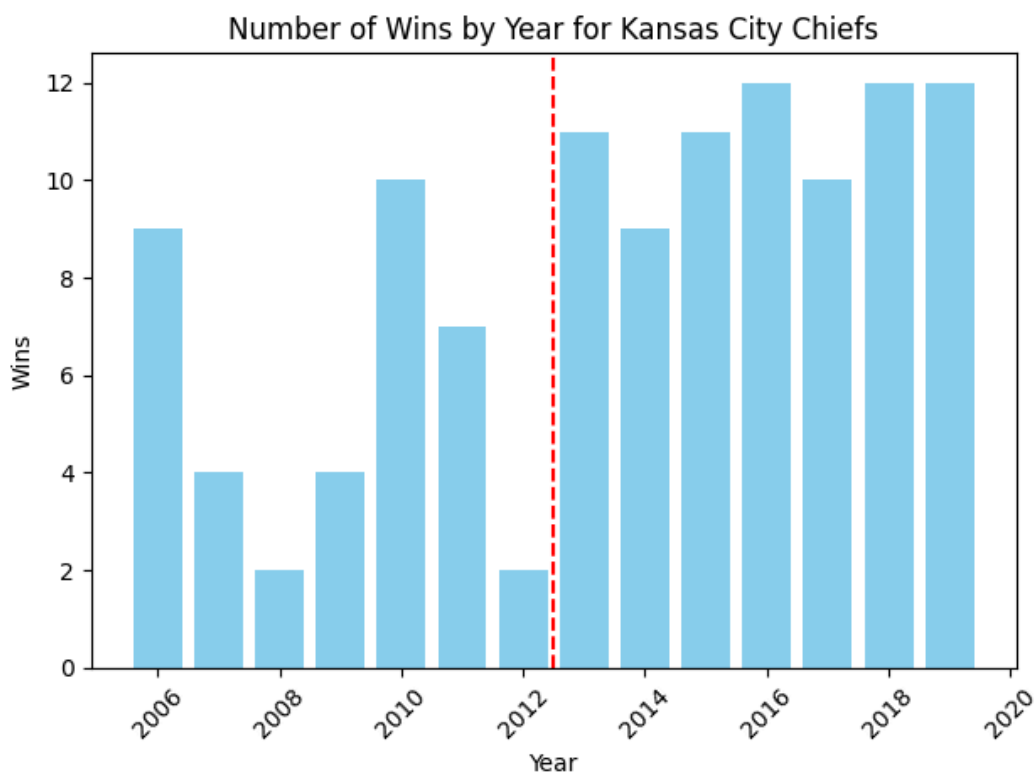
3) Which teams get the most favor on penalty calls?

Most fans get most riled up during penalty calls and often accuse others and referees of being biased in favor of their opponents. We suspect that most of these claims are unfounded, although we still wanted to use our dataset to see if there were teams receiving significantly more penalties called in their favor. We see certain discrepancies through data such as the Cleveland Browns who receive the most penalty yards but fail to have winning records or playoff success. Even still, the data seems to indicate there is not a significant difference between the lowest and highest teams to the point that any one team is receiving an advantage.

4)   Who has the best running record by team and season by the wins?

Fans often fondly remember when their team seemed to always win, to have a great coach or QB, or for any other reason. We wanted to see if each team's best 5-year seasonal window would line up with what fans remember. For this statistic we wanted to run this script over all teams and seasons, so we could determine if a team has had several strong seasons over the past few years. These periods stick with fans who would recall eras when their team was unstoppable or when their rival always seemed to win. We wanted to compare these assumptions to our data by mapping out the strongest 5 year streak for a team.

Different teams such as the New England Patriots, Green Bay Packers, and Pittsburgh Steelers showed evidence by their results and superbowl wins/appearances. The Pittsburgh Steelers experienced major success in their Quarterback Position with Ben Roethlisberger who attributed to an outstanding winning record and two Super Bowl wins. Aaron Rodgers also attributed to overall success during a period

5)   Which teams had the strongest / weakest offense as measured by fumble percentage?

We wanted to measure what were the strongest and weakest offensive lines in recent seasons (2015 - 2019) to determine if this aligned with fan's perceptions of a team's offensive performance over the period. We used the fumble ratios, i.e. the number of times the ball was lost to the opposing team on the offense's throw, which is one of the most important indicators for an offense's strength without measuring the ability of the QB, whose stats are often separated out from the rest of the offensive players. The strongest offense which was displayed by the Patriots supports the data we found, which is further proven by super bowl appearances and wins. The Green Bay Packers experienced multiple super bowl appearances and wins through 2010-2015 due to Quarterback play by Aaron Rodgers at a Most Valuable Player level during 2011 and 2014.

6)   Which teams had the strongest / weakest defense as measured by opponents PF?

We wanted to conduct measurements on the defense of teams, as this is independently one of the most important characteristics of a team. We want to verify perception about the strongest and weakest team's defense compared to the data at hand. We found that these results generally align with the fan's idea of the team's defense. New England was the strongest defense by a good margian, as the points scored by opposing teams could be lower due to the hyper powered offense that scored and won a lot of games for the team. The teams that struggled including the Cleveland Browns and Tampa Bay Buccaneers had high fumble ratios and negative records through the 2015-2019 seasons.

7)   Are defenses getting stronger over time?

We hypothesised that teams' defenses are growing stronger over time as it feels games are having less touchdowns and overall scoring. We wanted to see if I assumptions were true, by charting the total opposition scoring in a league, which would indicate that the teams defense in stopping less scoring for their opponents. After graphing the data, we determined our initial prediction was wrong, that teams seem to be scoring more over time indicating defenses are becoming weaker compared to offense. This was quite interesting as the data proved the initial guess wrong.
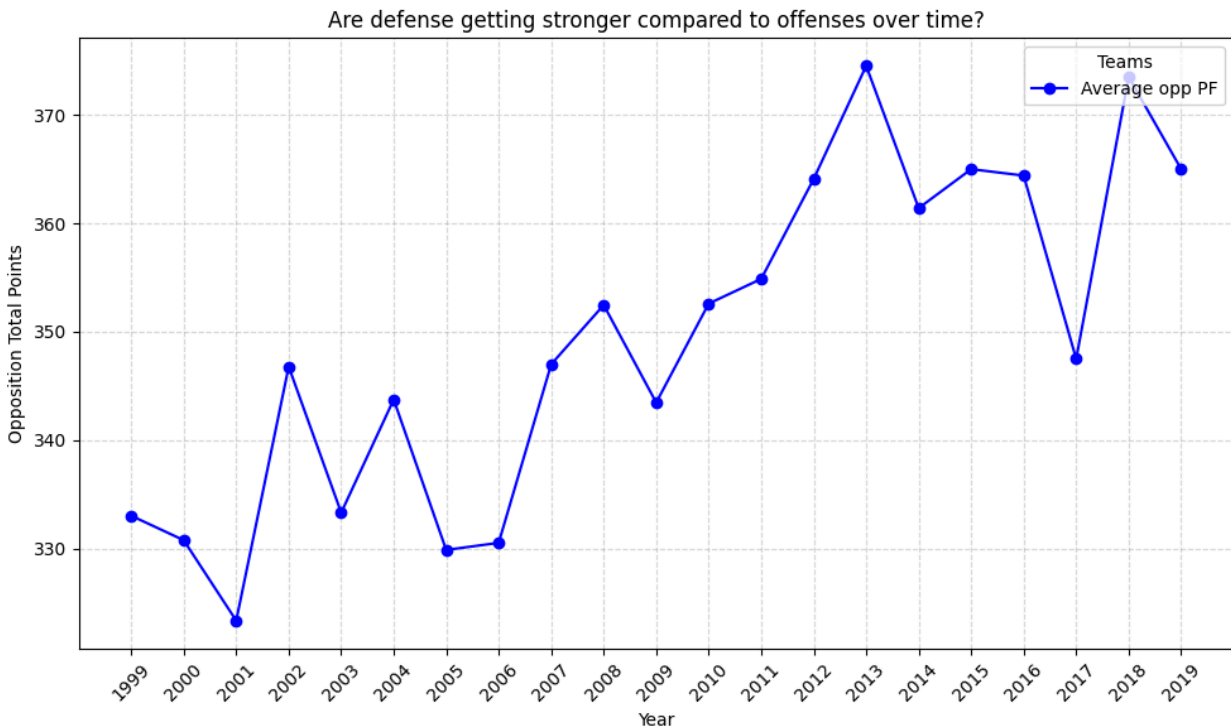

8)   Which team's season had the strongest td/int?

A team's touchdown per interception is a strong indicator used to determine the strength of a team's quarterback, the player responsible for throwing the ball each play. We would assume that quarterbacks famous for their skill would show up with their team in this graph. Besides the quarterback the wider receivers on a team are a direct indicator of effectiveness of overall offense performance. The wide receiver's inability to catch passes can lead to the cornerback interception and can overall affect the performance of a team. The reason the New England Patriots were dominating was because of great quarterback play and completion by the tight ends and wide receivers.

9) Is the sport getting more violent?

Some people we spoke to thought the sport itself had gotten more violent in recent years. We attempted to test this by measuring the number of penalties by year, which should partially show the trend for the overall aggressiveness of the games. This is an important factor since a significant issue for football has always been players being hurt and the prevention that it is a dangerous sport. An increase in dangerous plays would be a negative trend for the game that fans perceive. Our graph indicates that the trend has not increased significantly in recent years, which aligns with what we found online saying.

Year vs Penalties
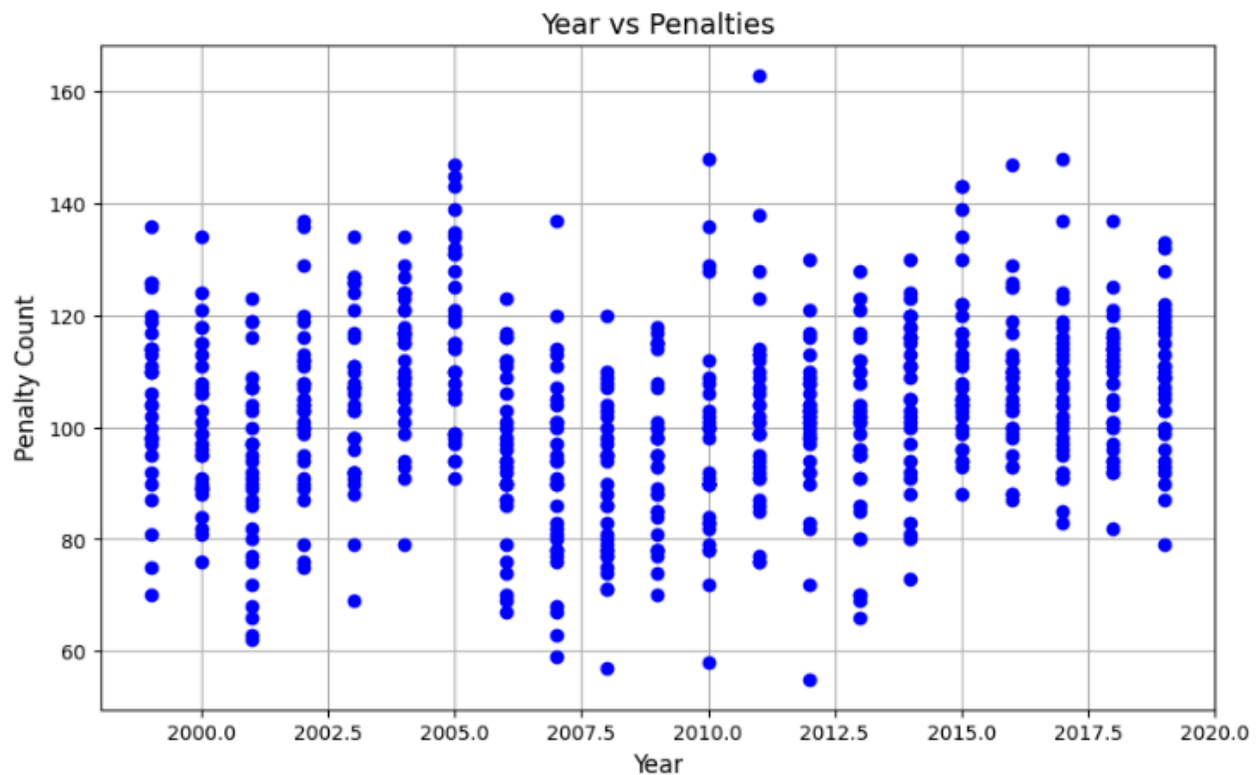
Presentation of Insights:

Overall the data we collected show that fans' perceptions on many of these claims are based in reality and showed that teams are often correctly analysed by fans even those not conducting data analysis. We ran into some interesting discoveries along the way though, like disproving the idea that defenses are getting stronger or that any one team is getting special treatment on penalty calls. We assume most fans constructed these data points in the spur of the moment and are less likely to follow long term trends during these calls. We found different teams dominating and improving and other teams struggling and declining aligned well with the general fans' understanding of them. Overall, most of the claims made by fans seem to be fairly correct and can be supported by the data.