**PROJECT REPORT:**

**Comparing Quarterback & Defensive Impact in the NFL**

Billy Bourke

**GitHub URL:** <https://github.com/billybourke/UCDPA_billybourke>

*Word count (excl. references/quick reference table): 1,998*

**Abstract**

This report will look at American football’s National Football League, and an emerging discussion as to whether one single position (quarterback) holds more importance than that of the eleven total players that make a team’s defensive unit due to rule changes emphasising higher point totals and passing the ball over running to boost viewer ratings.

We will look at how statistical performance for quarterbacks and defences correlate with the likelihood of their team winning games, identify key statistical categories that show the strongest correlations, will look at some of the standout quarterbacking and defensive seasons of the last decade and their performances in these areas, and will examine any potentially noteworthy changes within these categories in that timeframe. Both sides do of course prove important to a teams’ success, sometimes in obvious ways, and sometimes in others that are slightly less expected.

**Introduction**

If unfamiliar with American football, please see this 80 second video: <https://www.youtube.com/watch?v=3t6hM5tRlfA&ab_channel=NFLUK>

In addition to the video above, some additional notes:

1. When the offense has the ball, it is referred to as a ‘possession’ or a ‘drive’.
2. The video discusses teams punting the ball down the field if they are on their fourth (last down). However, a turnover (intercepted pass or recovered fumble, by the defence) will see the opposing team take control of the ball from where the turnover was made, which can make them extremely valuable for field position.
3. Penalties can be given against the offense or defence for committing fouls, which can see yards given away (e.g. making a 1st down and 10 yards into a 1st down and 5 or 15 yards, depending on who committed it).
4. QBR (Quarterback Rating) and passer rating, which will each be referred to frequently are based on formulae to attempt to quantify the performance of a quarterback when passing the ball.

Due to rule changes over the last twenty-or-so years which have made it easier to pass the ball in an effort to produce more points, more highlights, more TV ratings, and thus more profit, a team will very rarely build itself around a running game these days, and instead use it more as a balancing act to keep the defence ‘on their toes’ and uncertain of what to expect from them. This has led to the quarterback position becoming far more valuable than previously.

Seventeen of the last twenty (and nine of the last ten) Most Valuable Player awards have gone to quarterbacks, compared to 11.5 in the twenty years between 1982 and 2001 (1997 being shared). Quarterbacks can take up to as much as 23% of their team’s salary cap space despite rosters holding 50-53 players. This has seen the debate change from whether a good offense or a good defence is the best method to success, to whether a good quarterback is more important to winning than a good defence in its entirety.

**Quick links to examples of plays:**

Pass play/touchdown: <https://www.youtube.com/watch?v=akn2Yq8he-E&ab_channel=NFL>

Run play/touchdown: <https://www.youtube.com/watch?v=DrDk506-avY&ab_channel=NFL>

Sack: <https://www.youtube.com/watch?v=55ozo2QqUds&ab_channel=NFL>

Fumble recovery: <https://www.youtube.com/watch?v=BkYE011kCkQ&ab_channel=GamedaySports>

Interception: <https://www.youtube.com/watch?v=ctpfQ3_PNgo&ab_channel=NFL>

A glossary of statistical terms is located on the last page of this report.

**Dataset**

The dataset used is from Pro Football Reference, which tracks all statistics from the NFL over several decades in a far more extensive and uniform manner than others such as NFL.com. It is particularly useful not only because of this, but because of the format in which they display their data, that lends itself very well to web scraping and CSV extraction, and for this reason is regularly used by outlets such as ESPN, Bloomberg, New York Times and Forbes for their articles on the sport.

Specific datasets used were annual passing and defence statistics, plus standings (from 2012-2021):

1. Passing: <https://www.pro-football-reference.com/years/2021/passing.htm>
2. Defence: <https://www.pro-football-reference.com/years/2021/opp.htm>
3. Standings: <https://www.pro-football-reference.com/years/2021/index.htm>

**Implementation Process**

**For ‘quarterbacks’ (Section 1):**

1. Passing statistics for each year 2012-21 were web-scraped and concatenated together.
2. Some columns renamed and ‘NaN’ results replaced with ‘0’ as this is what a blank cell would denote.
3. ‘Win-loss-tie’ column separated into three individual columns, and ‘wins’ divided by ‘games started’ to give win/loss percentage (a QB is not credited with a win in the NFL unless they **start** the game as opposed to being substituted in during it).
4. Any players with less than one game started, or whose position is not quarterback (occasionally a team will do a ‘trick play’ where a non-QB passes the ball to catch opposition defences off guard) were filtered out, and index set to win/loss percentage in descending value.
5. All columns outside of QB and team names set to numeric, as some were appearing as strings.
6. All special characters (including spaces) were removed as players who win awards or make all-star teams typically have a ‘\*’ or ‘+ placed beside their name, which was causing later issues.
7. Reindexed to make the most relevant reference information (Player, Team, Year) immediately visible.

**For ‘defences’ (Section 2):**

1. Defensive statistics for each year (2012-21) were imported from downloaded CSV files.
2. Defensive statistics only had full team names (e.g. ‘Miami Dolphins’) while ‘standings’ files only listed team abbreviation (e.g. ‘MIA’). Teams were reindexed alphabetically, and a new column with their corresponding abbreviations created.
3. All special characters were removed from the team names, as with quarterbacks.
4. Imported annual ‘standings’ CSVs for each team’s win/loss record; inner merged with defence.
5. All years were concatenated, and unnecessary columns dropped.
6. Columns reindexed for easier referencing, and all statistical columns set to numeric.
7. Several columns renamed, either for clarity or to remove suffixes assigned from the earlier inner merge.
8. Index set to win/loss percentage, with values sorted by the same in descending order.

**For charts (Sections 3-7):**

1. **Chart 1:** Two lots of scatter plots created, labels edited for clarity and x/yticks edited for better visualisation.
2. **Chart 2:** Each statistic’s correlation to win/loss record calculated using ‘.corr’ function, placed in dataframes, and changed all to positive with ‘.abs’ function as some (e.g. point allowed) would correlate to wins based on lower totals rather than higher. Bar subplots created; xticks manipulated for consistency/better visualisation (up to full 100% correlation rate).
3. **Chart 3:** Used ‘.quantile’ query function (set to ‘>0.75’ and ‘<0.25’) and calculated mean W/L records when set to each of these based on four key statistics for quarterbacks and defences from Chart 2. Created stacked barplot, set yticks at 0.1 intervals from 0.0 – 1.0, and added a horizontal axis line at halfway point (0.5).
4. **Chart 4:** Used ‘.nlargest’ function to filter the top 10 win/loss ratios of quarterbacks and defence and placed in dataframe with main statistics, created a bottom row of each using .loc with the mean of each statistic over 10 years, used ‘.astype’ function to convert W/L record and ‘year’ to string, and combined into a single column along with ‘Player’. Divided each bar statistic by the mean to get percentage of mean. Used .melt’ function to break each quarterback/defence down into four statistics, changed xticks for uniformity across each bar chart.
5. **Chart 5:** Converted all columns for key statistics to ‘float’, created a dictionary of mean score for each key statistic in each year. Created three line-charts: quarterback, defence, and defence points against (as the much higher figure was impacting the other three defensive statistics). Adjusted legend to outside of chart, to not get in the way, changed xticks to annual (2012-21), and yticks as appropriate.

**For remainder (sections 8-11):**

1. Used iterrows to generate a ‘scorecard’ for each QB/defence in key statistics in each year (sec 8 + 9).
2. Defined reusable function using filtering (top 150 performances, of 315 for QBs and 320 for defences), as well as ‘.nlargest/.nsmallest’ functions and groupings to track which quarterbacks and teams have had the most success in each category.

**Results**

**Chart 1:** Comparing scatter charts of each statistical category for quarterbacks and defences.

Calendar

Description automatically generated

Calendar

Description automatically generated

**Chart 2:** Getting an exact bar plot reading of the correlation with wins/losses for each statistical category.

Chart

Description automatically generated

**Chart 3:** Looking at the average win/loss percentage for QBs and defences that rank in both the top and bottom quartile of the four statistics with the greatest win/loss correlation.

Chart, bar chart

Description automatically generated

**Chart 4:** Comparing the top 10 winning-percentage seasons for both quarterbacks and defences, and their performance relative to the league average across the decade in the four key statistics identified in Chart 2.

Chart

Description automatically generated

**Chart 5:** Looking for fluctuation in league averages for key statistics over time.

Graphical user interface, chart

Description automatically generated

**Insights**

**Chart 1 (scatter plots):** These help us identify the distribution of performance to win/loss ratios, outliers such as a sole QB Passer Rating with a top quartile win ratio despite a bottom quartile QB Passer Rating, anomalies like one quarterback’s particularly high percentage of passes intercepted, and multiple areas where finishing in the top quartile for performance appears a certainty to avoid finishing the bottom quartile of win/loss ratio (and vice-versa).

**Chart 2 (bar plot):** Some are obvious, such as points allowed on defence (concede less points, win more games), while others are more surprising such as yards per pass completed holding the lowest correlation of all QB statistics despite adjusted net yards per attempt being fourth highest, or that penalty yards conceded by a defence holds almost no correlation whatsoever to teams’ win/loss records.

**Chart 3 (tiered bar plot):** We can see from this that teams with top-end quarterback play in categories tend to hold a slightly higher win percentage ceiling than their defensive equivalents, while those with bottom quartile defensive play appear to lose more than those with bottom quarter QB play. Elite quarterback play appears to lead to a higher ceiling than defensive, while poor defensive play can result in a lower floor.

**Chart 4 (multi bar plot):** Well-rounded quarterback play in these key categories is pivotal to success (touchdown percentage most of all), suggesting that taking risks to put up higher scores rather than settling for field goals is worth it if you have the talent at the position to do so. Alex Smith’s 2020 season presents a fascinating anomaly, being the only quarterback to finish below average in any category – and actually managing to do so in all of them – some by quite a distance.

Defensively, the importance of turnovers to high win percentages is very noticeable, as it immediately gives your team the ball from wherever the turnover occurred - hence the NFL adage ‘sacks kill drives, turnovers kill games.’

Points against were lower than average for all teams but one – the 2013 Denver Broncos, whose offense had the highest points total in NFL history. For scoring percentages against however, most on this list are closer to the mean – suggesting that some of their lower points against totals may be linked to better ball retention from their offensive counterparts, or them giving up more 3pt field goals and less 7pt touchdowns. Finally, while these teams typically perform well on net yards per pass attempt, it does not appear as much of a prerequisite to the highest levels of success as the other three factors despite its’ correlation to general success.

**Chart 5 (line graph):** We can see a slight drop in league average QBR scores in recent years by about 10%, but an increase in both scoring percentages per-drive and average points per game by a similar amount. The exact reasons for this are unclear in the statistics we have explored, though may suggest a slight return of the relevance of the running game, or an increased dependence on field goal kicking. Otherwise, all other statistics have remained quite consistent throughout.

**Conclusion:** While the discrepancies are not particularly large, it does appear that elite play at the quarterback position can provide a higher likelihood of reaching the highest levels of success, though it also appears that standout levels of defensive play can give a team a higher baseline of success, and if coupled with a high turnover ratio can likewise lead to a similarly elite levels of success as seen by top-tier quarterback play.

**References**

1. <https://pandas.pydata.org> – for assistance with various functions, syntax and capabilities.
2. <https://matplotlib.org/stable/gallery> – for chart templates, and assistance in using additional functions.
3. <https://seaborn.pydata.org/examples/index.html> – for chart templates, and assistance in using additional functions.
4. <https://stackoverflow.com> – for assistance in various issues relating to troubleshooting functions and syntax.
5. <https://www.geeksforgeeks.org/python-programming-language> – various useful articles on functions and combining different methods to achieve a goal.

|  |  |
| --- | --- |
| **Quick reference guide for functions required in report** | |
| **Importing data (web scrape):** | Section 1 |
| **Importing data (CSV):** | Section 2 |
| **Sorting:** | multiple (first = Section 1, line 81) |
| **Indexing:** | multiple (first = Section 1, lines 66 + 81) |
| **Grouping:** | Section 10, line 10 + Section 11, line 9 |
| **Replace missing values:** | Section 1, lines 71 + 72 |
| **Slicing/loc/iloc:** | loc = multiple; first = Section 6, lines 11 & 23  iloc = Section 6, lines 30, 31, 32, 33, 38, 39. 40 & 41 |
| **Iterrows:** | Sections 9 + 10 |
| **Merge DataFrames:** | merge = Section 2, lines 14, 26, 38, 50…  concat = Section 1, line 61 + Section 2, line 124 |
| **Define custom function to create reusable code:** | Sections 10 + 11 (all) |
| **Numpy:** | Section 2, lines 63, 64 & 65 |
| **Dictionaries/lists:** | Section 2, lines 7 , 19 , 31, 43. 55, etc |

|  |  |  |
| --- | --- | --- |
| **Glossary of NFL Statistical Categories** | | |
| **Quarterbacks:** | | |
| QB | *Cmp%* | Percentage of pass attempts successfully caught by a teammate. |
| QB | *Pass Yards* | Cumulative yards gained over the season from completed passes. |
| QB | *TD%* | Percentage of passes attempted to result in a touchdown (score). |
| QB | *Int%* | Percentage of passes attempted to result in an interception (turnover). |
| QB | *1st Down %* | Percentage of passes attempted to result in a first down (beyond the yellow line on the field). |
| QB | *Y/A* | Average yards gained per pass attempted:  Passing yards / Passes attempted |
| QB | *NY/A* | Net yards gains per pass attempted:  (Passing Yards - Sack Yards) / (Passes Attempted + Times Sacked) |
| QB | *AY/A* | Adjusted yards gained per pass attempt:  (Passing Yards + 20 \* Passing TD - 45 \* Interceptions) / (Passes Attempted) |
| QB | *Y/C* | Average yards gained per successfully completed pass attempt. |
| QB | *4QC* | Successful drive in the fourth-and-final quarter of a game to put your team ahead. |
| QB | *GWD* | Successful drive in the fourth quarter that puts your team ahead, and results in a victory. |
| **Defences:** | | |
| Def | *PA* | Total points allowed over the course of the season. |
| Def | *Sc%* | Percentage of opposition possessions/drives ending in a score (FG or TD). |
| Def | *PassTD* | Total passing touchdowns allowed over the season. |
| Def | *RunTD* | Total running touchdowns allowed over the season. |
| Def | *Ply* | Total numbers of plays/downs over the season. |
| Def | *Yds* | Total yards allowed over the season (passing and running). |
| Def | *Y/G* | Average total defensive yards allowed per game. |
| Def | *Y/P* | Average yards allowed per play (passing and running) |
| Def | *PassYd* | Total passing yards allowed over the season. |
| Def | *RunYd* | Total running yards allowed over the season. |
| Def | *Y/P* | Average yards allowed per play (passing and running) |
| Def | *Y/A* | Average yards allowed per run attempt. |
| Def | *NY/A* | Net average yards per pass attempt:  (Passing Yards - Sack Yards) / (Passes Attempted + Times Sacked) |
| Def | *TO%* | Percentage of opposition possessions/drives resulting in a turnover (interception, fumble). |
| Def | *INT* | Interception turnover (catching a pass from the opposing quarterback). |
| Def | *F/L* | Fumble recovered (gaining possession of the ball if dropped by an opposing player). |
| Def | *Pen* | Penalties (fouls) committed over the season. |
| Def | *PenYds* | Total yards conceded as a result of penalties (fouls) committed over the season. |