Data Science Skills Challenge

By: Connor Tivadar

Defense Wins Championships?

- One of the most common phrases you will hear in regards to American Football is “Defense wins championships”
- So of course, a large goal among NFL general managers is to build as strong of a defense as possible
- While a simple solution to this may seem to be “spend as much money as possible buying the best players available”, this isn’t very realistic because of the NFL’s Salary Cap limit and the need to also make sure other aspects of the team have proper players

How Can Data Help Us?

- We need to discover a way for NFL General Managers to build a championship caliber defense while still being as effective as possible in how they spend money
- To do so, we will analyze the different aspects of the defense, specifically the secondary (cornerbacks and safeties) and the defensive line (defensive tackles and defensive ends), and see if this aspect of the defense is more correlated with having a successful defense
- By knowing which aspect of the defense is correlated with a successful defense, we can recommend what type of players NFL General Managers should prioritize spending money on and which they shouldn't

Questions Our Data Needs to Answer:

- Which aspect of the defense, the secondary or the defensive line, is more impactful in creating an overall strong defense?
- What are the types of statistics that let us judge how well a secondary unit or defensive line unit is performing?
- How can we give an effective recommendation to NFL General Managers based on what types of players they need to sign to both effectively spend and have a powerful defense?

How We Can Go About Investigating This?

- Step 1: Finding a database that gives us proper statistics in order to determine what makes a aspect of the defense "successful"
- Step 2: Sort and Clean the dataset
- Step 3: Import the dataset into Tableau in order to create visualizations and collect statistical data
- Step 4: Analyze Results, Report, and Give Answers to our Questions

Step 1: Finding a Dataset

- In order to find a dataset we need to decide on what statistical factors we will use to determine whether or not the defensive line and secondary are playing well
- I decided on 8 total: 4 for the secondary and 4 for the defensive line
    - The four statistical factors for the secondary are Passing Touchdowns Allowed, Net Yards Gained per Pass Attempt, 1st Downs By Passing, and Yards After Catch on Completions
    - The four statistical factors for the defensive line are 1st Downs By Rushing, Rushing Yards per Attempt, Rushing Touchdowns Allowed, and QB Pressures per Dropback
- Then I had to decide on a success factor, which would be a variable which allowed us to know whether a defense in general was good or bad
    - For this, I used one of the most important statistics in the NFL, Points Allowed per Game, where a lower value for Points Allowed per Game indicates a good defense and vice versa

Step 1: Finding a Dataset

- Now that we know what variables we want to prioritize, I can go about finding my dataset
- I chose to export datasets from the website Pro Football Reference, from their section of data that described NFL defenses in 2023
- I chose to use these datasets because not only does it include all the variables I'm interested in with a good format, but also provides more variables that can be used in future analysis and discovery
- Datasets Credit: https://www.pro-football-reference.com/years/2023/opp.htm#team_stats::1

Step 2: Sorting and Cleaning Data

- Pro Football Reference already provides fairly clean datasets for me to use, however for this project I wanted a little bit more clear data set with an emphasis on the important variables for use in Tableau
- To do so, I imported each of the Pro Football Reference datasets into Google Sheets and then cut up and reformatted the datasets in order to create one overall dataset to use
- The final dataset I used in Tableau for this project can be found here: https://docs.google.com/spreadsheets/d/10myr81JOGwREoMtXN08GZ7wdkQSo5FAI/edit?usp=drive_link

Step 3: Tableau and Collecting Data

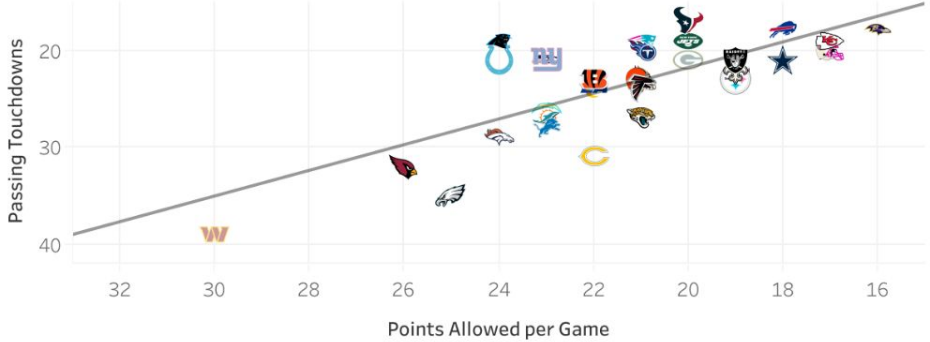
- Once the dataset was imported into Tableau, I began by using their functionality to start creating graphs and looking for correlations
- What Worked: Early on we were able to see a lot of correlation between certain variables, and producing strong graphs with interesting trend lines to see a relationship
- What Didn't: However, some of our variables we had originally chosen we realized we had to reevaluate, since some of them represented the performance of both the defensive line and secondary at the same time. So we had to go back to Pro Football Reference and choose different variables that made more sense for our project

Step 3: Tableau and Collecting Data

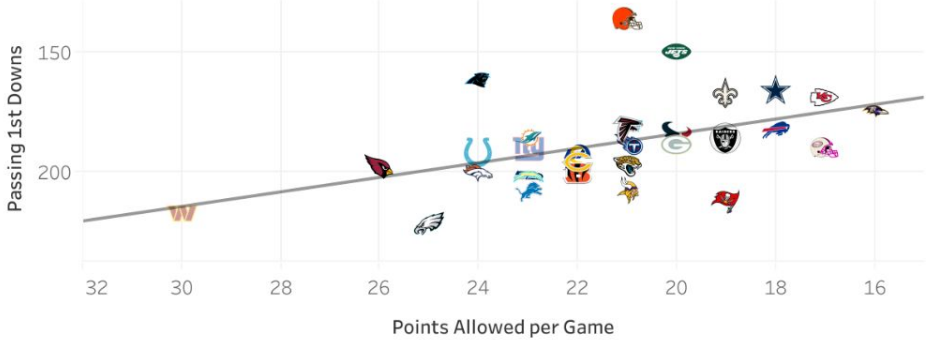
- Once we were able to create our graphs comparing the different variables to the variable of Points Allowed per Game, we used Tableau to get trend lines for each graph as well as R squared values and P-values in order to analyze correlation
- We also edited the ranges and domains of our axis in order to get a strong graph to give a much clearer representation of the data
- Finally, imported NFL logos as shapes into Tableau and used these images as the shapes of our data points so people can clearly understand the data points without needed to look at the dataset and overall improve our visualizations

Results: Graphs for Secondary

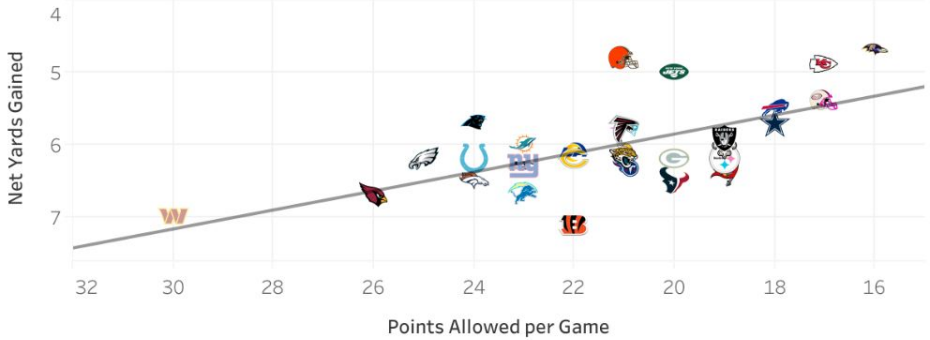
Pasing Touchdowns vs Points Allowed per Game



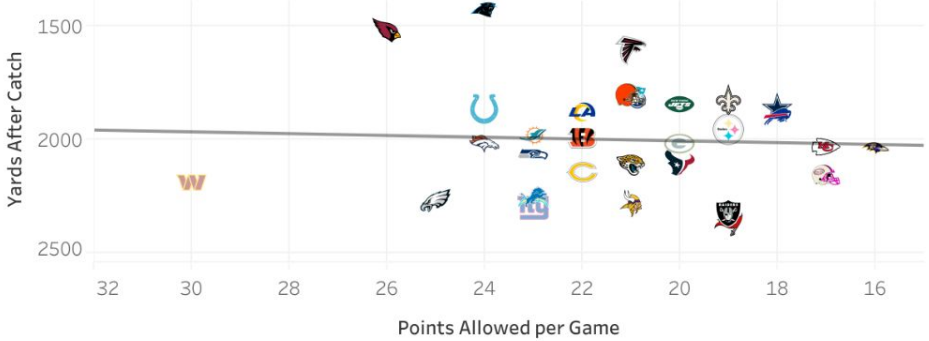
1st Downs By Passing vs Points Allowed



Net Yards Gained per Pass Attempt vs Points Allowed

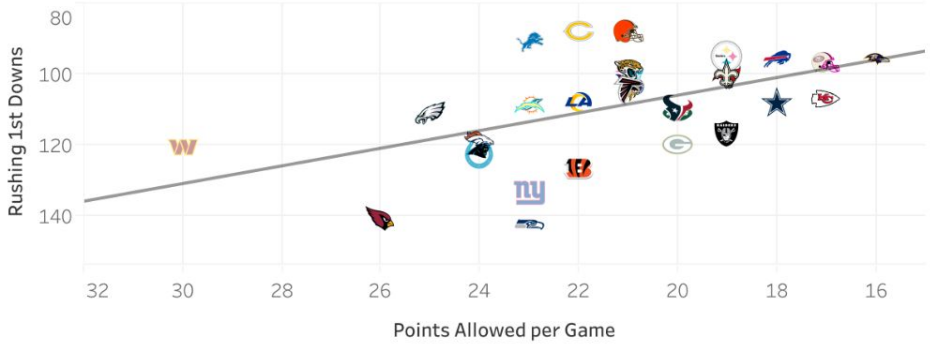


Yards After Catch on Completions vs Points Allowed

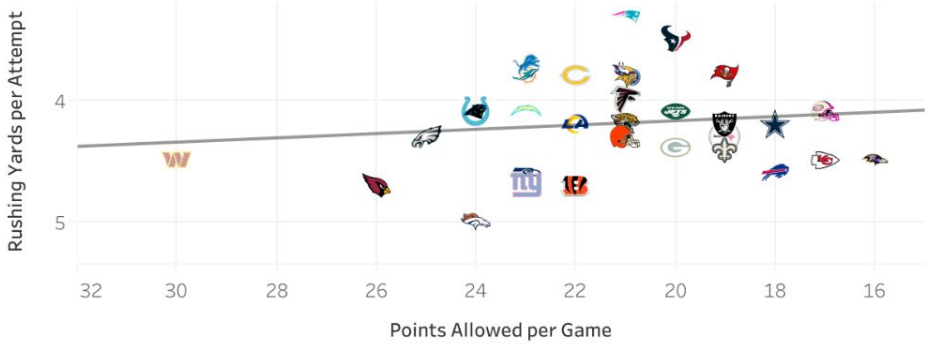


Results: Graphs for Defensive Line

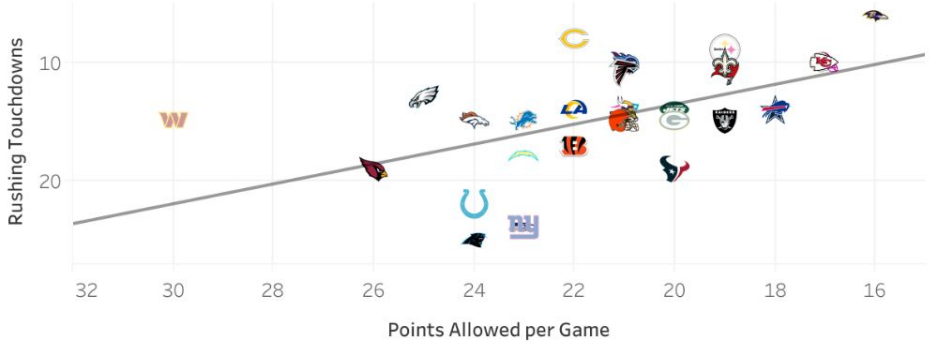
1st Downs By Rushing vs Points Allowed



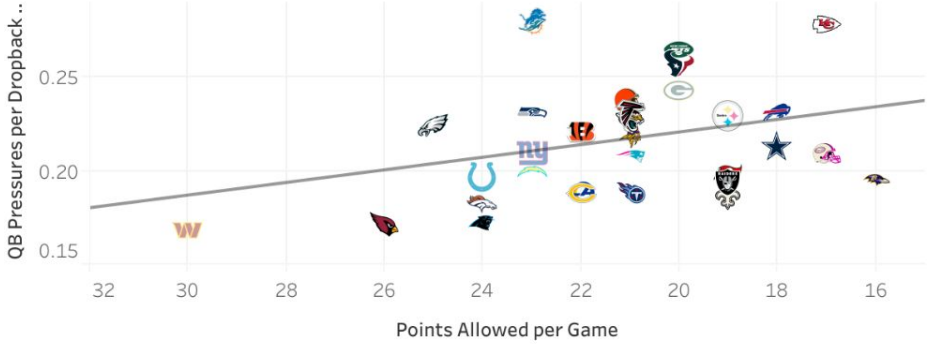
Rushing Yards per Attempt vs Points Allowed



Rushing Touchdowns vs Points Allowed



QB Pressures per Dropback vs Points Allowed



Results: R Squared Values

- R Values for Secondary Graphs:
  - Passing Touchdowns - 0.5550
  - 1st Downs by Passing - 0.2242
  - Net Yards Gained per Pass Attempt - 0.4116
  - Yards After Catch on Completions - 0.0024
- R Values for Defensive Line Graphs:
  - 1st Downs by Rushing - 0.2641
  - Rushing Yards per Attempt - 0.0180
  - Rushing Touchdowns - 0.2789
  - QB Pressures Per Dropback - 0.0961

Results: P Values

- P Values for Secondary Graphs:
  - Passing Touchdowns - < 0.0001
  - 1st Downs by Passing - 0.0062
  - Net Yards Gained per Pass Attempt - < 0.0001
  - Yards After Catch on Completions - 0.7880
- P Values for Defensive Line Graphs:
  - 1st Downs by Rushing - 0.0026
  - Rushing Yards per Attempt - 0.4640
  - Rushing Touchdowns - 0.0018
  - QB Pressures Per Dropback - 0.0828

Step 4: What These Results Mean

- As we can see from the graphs and the R squared values, in general the variables I looked at for the secondary had a stronger correlation with Points Allowed per Game then the variables I looked at for the defensive line
    - The only variable for the secondary that had a fairly low R squared value was the "Yards After Catch on Completions" variable, however this variable also had a very large p-value meaning that it could've been just by chance that we saw results like this, and they may not be representative of how much that variable correlates with Points Allowed per Game
- All the other variables within the secondary with a high correlation had a very low p-value, indicating that these could be significant results and not due to chance

What can we Gather from This

- By looking at the visualizations and data, we can see that the **Secondary players** statistics are more correlated with a defense that allows less points per game than the defensive line statistics
- This shows us that when trying to determine what type of players a General Manager should focus on paying, they should choose **Secondary players** rather than defensive line players because secondary players are more important when it comes to having an overall championship level defense
- General Managers in the NFL can use this information to spend less money on their defensive line, making it possible for teams to have a championship caliber defense while staying under the salary cap or being able to have the spending money to bolster other aspects of their team such as offense and special teams

References

Linebacker Image:
https://pngimg.com/image/37788

Lombardi Trophy Image:
https://www.fineawards.com/articles/super-bowl-trophy-the-history-of-the-vince-lombardi-football-trophy

Google Slides Theme:
https://slidesgo.com/theme/innovate-business-plan#search-Professional&position-22&results-4909