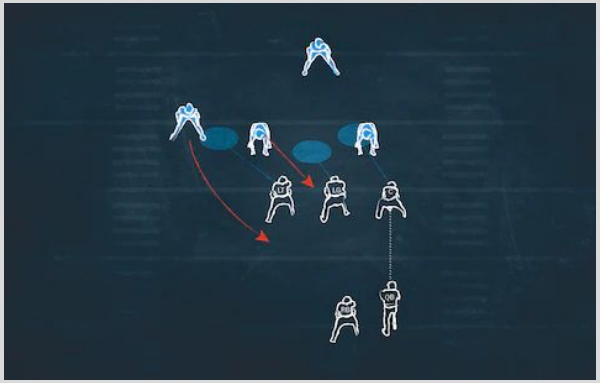
NFL Linemen Performance Analysis

Dennis Kreger, Robert Loesch, Liam Selfors, & Greg Winkelman

Topic Selection

- Our team has selected to participate in a Kaggle competition for the NFL Big Data Bowl 2023 and is planning to submit an entry into the online competition.
- The topic of the 2023 NFL Big Data Bowl competition is Linemen on Passing Plays.



  - Quarterbacks may get the glory, but some of the most important work takes place a few feet in front of them. The offensive line protects the passer, providing precious seconds to find receivers downfield. At the same time, the opposing team's defensive line attempts to find a disruptive path. If a defender sneaks through, it can mean a sack, a blocked pass, or even a turnover. Some of the game's most important plays happen on the line and this competition examines the data behind the hardest workers in football.

Why NFL Data Analysis?

- The team is enthusiastic about sports and football in particular.
- Some team members strive to be Sports Data Scientists, and the opportunity to present findings and results to elite athletes/organizations would be a life changing.
- Getting exposure to Kaggle (and the potential to win the large competition) would showcase our technical skills and would be helpful for future job opportunities.
- Robert has experience playing elite-level football in Germany and Greg's career is in sports medicine and performance, so we are hoping to utilize that experience if questions about the game, players, or movements arise.
- In addition, there is a $10,000 prize that the team members could win!

Data Sources

Teams in the competition were given access to:

- **NFL Next Gen Stats**
  - Player tracking (Key variable = ‘gameId’, ‘playId’, ‘nflId’)
  - Plays = ‘gameId’, ‘playId’)
  - Games = ‘gameId’)
  - Player information = ‘nflId’)
- **Pro Football Focus (PFF) scouting data**
  - Weeks 1-8 of 2021 NFL season
  - Passing plays only (Key variable = ‘gameId’, ‘playId’, ‘nflId’)

Supplemental:

- **NFL Combine dataset**
  - Deeper dive into the maximum performance metrics of players when they enter the league

Links to Data Files on Cloud Storage:

https://storage.googleapis.com/big-data-bowl/games.csv

https://storage.googleapis.com/big-data-bowl/pffScoutingData.csv

https://storage.googleapis.com/big-data-bowl/players.csv

https://storage.googleapis.com/big-data-bowl/plays.csv

https://storage.googleapis.com/big-data-bowl/week1.csv

https://storage.googleapis.com/big-data-bowl/week2.csv

https://storage.googleapis.com/big-data-bowl/week3.csv

https://storage.googleapis.com/big-data-bowl/week4.csv

https://storage.googleapis.com/big-data-bowl/week5.csv

https://storage.googleapis.com/big-data-bowl/week6.csv

https://storage.googleapis.com/big-data-bowl/week7.csv

https://storage.googleapis.com/big-data-bowl/week8.csv

https://storage.googleapis.com/big-data-bowl/nflcombine.csv

We hope to answer:

- The team will participate in the “Metric track” of the competition
- We hope to create a metric to assess performance and/or strategy - (*this can be focused on offensive or defensive players, and on teams or individuals).*

Main Objective(s):

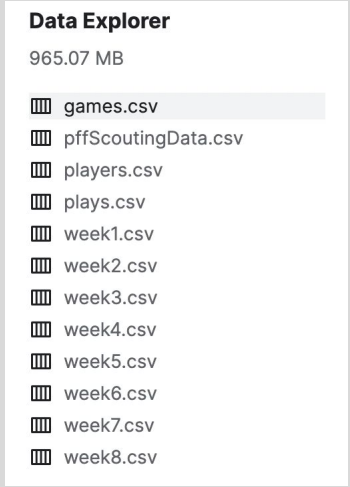
- Predict positive or negative outcomes of the offensive line performance before a play in the Red Zone (inside the opponents 20-yard line) based on the defensive formation and the outcome of the play.
    - **Positive factors**: Touchdown, First down, 3rd/4th down conversion, holding blocks for extended time, picking up a blitz, and making a positive play when there is defensive pressure.
    - **Negative factors**: Beaten by defender, Sack allowed, QB hit/hurried, Penalty, Sack or scramble play-type, not converting on 3rd/4th down.

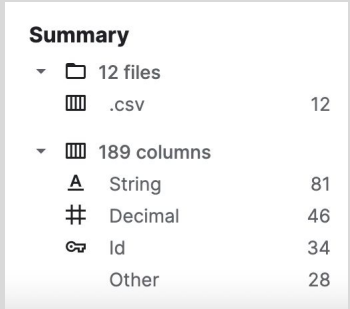
    - We also plan to further analyze player performance data (speed, acceleration, “Positive factors”, sacks allowed, team success, and combine performance) to predict matchup outcomes on an individual and team basis that will be interactive on our dashboard.

Data Exploration

Luckily, a majority of our data was given to us with direct download links posted on Kaggle (NFL Big Data Bowl 2023 - *link*)

- Due to the competition format, we don't have to use any other data if we choose not to
- We chose to incorporate all of the data sources given on the Kaggle page and also add a dataset with all of the players NFL Combine data on it
    - Further insight to athletic performance insights to use in our data
    - These Big Data Bowl datasets have millions of rows of data due to capturing every frame (0.10 sec) of a play.(For every player in every game - weeks 1-8)





Data Exploration (cont.)

Exploratory findings with the datasets:

- Combining yardage columns to find absolute yards per play
- Finding “QB Zone” threshold for distance from a defender
- Counting “event” frequencies of each play/result
- Linemen matchup threshold (proximity and play result)
- Defining Red Zone plays

Analysis

Along with the exploration findings listed on the previous slide, the data was further analyzed for linemen specifically.

- Datasets were merged and then separated out by offensive line position (center, guard, tackle)
- Using PFF data from specific plays, we can use the defensive formation and play outcome to determine the success rate of the offensive line as a unit, and also as individuals
- PostgreSQL, Python, and Pandas have all been used thus far to analyze the data, create new tables for the data, and perform some exploratory analyses on the linemen data

Analysis (cont.)

Specifics:

- Key values amongst datasets: ‘gameId’, ‘playId’, & ‘nflId’
- Analyzing positive versus negative performance from offensive linemen
  - Negative objects measured: Hit/sack allowed, beaten by defender, interception/scramble, penalty, failing to convert a 3rd/4th down
  - Positive objects measured: Touchdown, 1st down, 3rd/4th down conversion, line is able to hold off pressure, able to block for >3 seconds, and protection from a blitz
- Using player tracking metrics (i.e. acceleration, speed, distance covered) to factor into the highest performing “positive performing” linemen

```
w1_T_df = pd.read_csv('Resources/tackle_w1.csv')
w1_T_df
```

	nflId	height	weight	displayName	gameId	playId	frameId	team	s	a
0	48455	6-7	299	Matt Nelson	2.021091e+09	3444.0	15.0	DET	1.60	6.51
1	43640	6-6	313	Rashod Hill	2.021091e+09	3977.0	29.0	MIN	4.05	5.69
2	52418	6-5	320	Jedrick Wills	2.021091e+09	797.0	8.0	CLE	0.82	5.28
3	43648	6-5	315	Justin Murray	2.021091e+09	2982.0	38.0	ARI	0.78	4.51
4	43640	6-6	313	Rashod Hill	2.021091e+09	672.0	14.0	MIN	2.83	4.49
...	...	...	...	...	...	...	...	...	...	...
6913	46084	6-8	325	Kolton Miller	2.021091e+09	1344.0	7.0	LV	0.00	0.00
6914	46084	6-8	325	Kolton Miller	2.021091e+09	1344.0	6.0	LV	0.00	0.00
6915	38553	6-6	305	Riley Reiff	2.021091e+09	4849.0	6.0	CIN	0.00	0.00
6916	44846	6-6	320	Cam Robinson	2.021091e+09	2245.0	6.0	JAX	0.00	0.00
6917	29550	6-4	328	Jason Peters	2.021091e+09	69.0	6.0	CHI	0.00	0.00

6918 rows × 55 columns

Sample data frame for Tackles that played in week 1 to analyze

Dashboard

The following slides show the relative layout that we are hoping to accomplish with our dashboard.

Details:

- Public dashboard will be hosted via HTML, CSS, and Javascript
- Bootstrap template will be used to help format page and style
- Interactive Elements:
  - A navigation bar on the side of the page to go between “Game”, “Player”, and “Plays” visualizations
  - Matplotlib visualizations of any play contained in the dataset (aerial view from above)
  - Text entry boxes for determining the predicted winner of a game based on analysis
  - Drop down menus to select and predict specific player matchups, and also to customize a play and predict the success of the offensive line
  - Visualization showing the acceleration and ability to hold blocks (stay engaged with a defender) with any play in the dataset

(Dashboard Blueprint)

Game

Player

Plays

Performance Of Linemen on Pass Plays

Home Team: TB

Away Team: DAL

Team Info

Division: NFC South

Team Info

Division: NFC East

Win Prediction

(Dashboard Blueprint)

Game

Player

Plays

Performance Of Linemen on Pass Plays

Matchup Prediction

Offensive Player: Joe Smith

Defensive Player: Eric Jordan

Player Info

College: Auburn (Div I-A)
Age: 21
Height: 6’6”
Weight: 321

Player Info

College: Wisconsin (Div I-A)
Age: 28
Height: 6’8”
Weight: 350

Player Matchup Trouble Meter

(Dashboard Blueprint)

Game

Player

Plays

Performance Of Linemen on Pass Plays

Play Setup

Down: 3

Yards to Go: 7

Sideline: HOME

Offensive Play Selection:

Formation: SHOTGUN

Personnel: 1 RB, 1 TE, 2 WR

Block Type: Pass Protection

Drop Back Type: TRADITIONAL

Defensive Play Selection:

Cover: ZONE

Personnel: 4 DL, 2 LB, 5 DB

Coverage Type: Prevent

(Dashboard Blueprint)

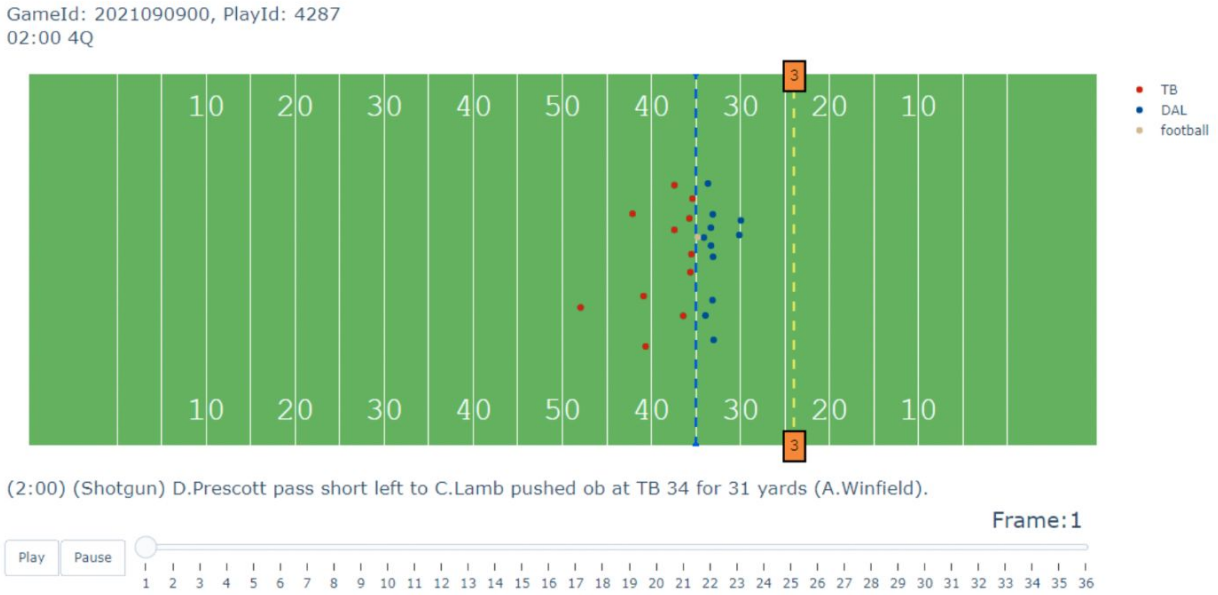
Game

Player

Plays

Performance Of Linemen on Pass Plays (continued)

Play Simulation



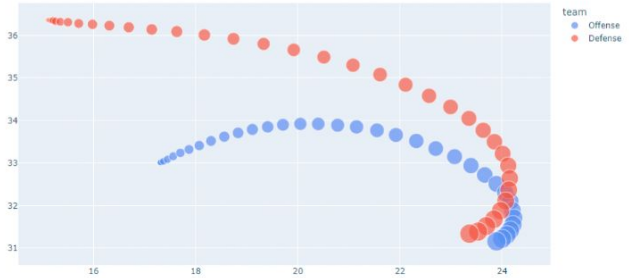
(Dashboard Blueprint)

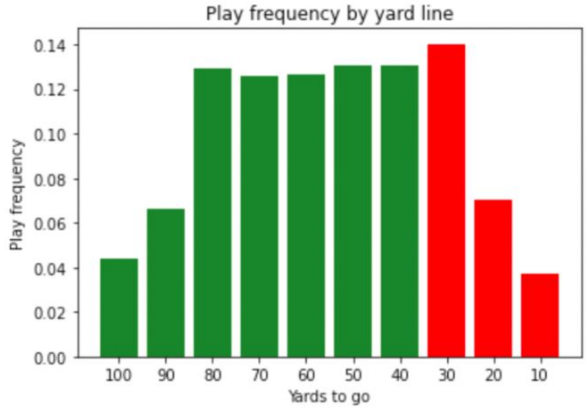
Game

Player

Plays

Performance Of Linemen on Pass Plays (continued)





Images

- https://www.google.com/imgres?imgurl=https%3A%2F%2Fkinexon.com%2Fuploads%2Fimages%2FSports%2F_800x533_crop_center-center_82_line%2FLoad-Management_website.jpg&imgrefurl=https%3A%2F%2Fkinexon.com%2Fproducts%2Famerican-football%2F&tbnid=RTmgyCojORr3yM&vet=12ahUKEwiwy-z1w7T7AhXRkWoFHfUDCQIQMygOegUIARDWAQ..i&docid=-jxml2lc6P8XSM&w=800&h=533&q=nfl%20performance%20tracking&client=safari&ved=2ahUKEwiwy-z1w7T7AhXRkWoFHfUDCQIQMygOegUIARDWAQ
- https://www.google.com/imgres?imgurl=https%3A%2F%2Fstorage.googleapis.com%2Fkaggle-competitions%2Fkaggle%2F15696%2Flogos%2Fthumb76_76.png%3Ft%3D2019-10-04-16-17-46&imgrefurl=https%3A%2F%2Fwww.kaggle.com%2Fc%2Fnfl-big-data-bowl-2020&tbnid=NAPKnaBhOrhlRM&vet=12ahUKEwiN-Izax7T7AhUwnGoFHfkpDN4QMygAegUIARCzAQ..i&docid=wdwEgMxpvWu1XM&w=240&h=240&q=nfl%20big%20data%20bowl&client=safari&ved=2ahUKEwiN-Izax7T7AhUwnGoFHfkpDN4QMygAegUIARCzAQ
- https://www.google.com/imgres?imgurl=https%3A%2F%2Fi.ytimg.com%2Fvi%2F-hOmfBTIJtg%2Fmaxresdefault.jpg&imgrefurl=https%3A%2F%2Fwww.youtube.com%2Fwatch%3Fv%3D-hOmfBTIJtg&tbnid=LNKq2JGQcFdszM&vet=12ahUKEwipuJSqxbT7AhVJmGoFHcNUBa8QMygCegUIARDtAQ..i&docid=-bTA-2A7EDpITM&w=1280&h=720&q=nfl%20passing%20pocket%20animated&client=safari&ved=2ahUKEwipuJSqxbT7AhVJmGoFHcNUBa8QMygCegUIARDtAQ
- https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.washingtonpost.com%2Fsports%2F2020%2F10%2F16%2Funderstanding-pass-blocking-schemes-football%2F&psig=AOvVaw1nRh4KFrqfTVyzQFhWnaJ_&ust=1668751885926000&source=images&cd=vfe&ved=0CA8QjRxqFwoTCOi4hcHHtPsCFQAAAAAdAAAAABAD
- https://www.google.com/imgres?imgurl=https%3A%2F%2Fmiro.medium.com%2Fmax%2F995%2F1*WBfAodFsdvqZkHUPTTsXdA.png&imgrefurl=https%3A%2F%2Fmedium.com%2Fanalytics-vidhya%2Fguide-to-the-kaggle-nfl-big-data-bowl-eb20819566c&tbnid=sSbCaYPKtaNEJM&vet=12ahUKEwiN-Izax7T7AhUwnGoFHfkpDN4QMygJegUIARDHAQ..i&docid=IF6L05D_-_RNzYM&w=995&h=331&q=nfl%20big%20data%20bowl&client=safari&ved=2ahUKEwiN-Izax7T7AhUwnGoFHfkpDN4QMygJegUIARDHAQ
- https://www.google.com/imgres?imgurl=https%3A%2F%2Fbostonsportsmedia.files.wordpress.com%2F2014%2F06%2Fpfflogo.jpg&imgrefurl=https%3A%2F%2Fbostonsportsmedia.com%2F2014%2F06%2F04%2Fcan-pro-football-focus-stats-be-blindly-trusted%2F&tbnid=DeByskEJnfairM&vet=12ahUKEwjOv5zjyLT7AhUPomoFHfZmDe4QMyglegUIARDbAQ..i&docid=_uFC1fh6IAdlSM&w=1079&h=302&q=pro%20football%20focus%20&client=safari&ved=2ahUKEwjOv5zjyLT7AhUPomoFHfZmDe4QMyglegUIARDbAQ
- https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.nfl.com%2Fvideos%2F2021-big-data-bowl-joe-andruzzi&psig=AOvVaw1BoSRL9r15ETp7-aUIfvgW&ust=1668752586132000&source=images&cd=vfe&ved=0CA8QjRxqFwoTCLCui_3JtPsCFQAAAAAdAAAAABAD
- https://www.kaggle.com/competitions/nfl-big-data-bowl-2023/data?select=games.csv
-