



NFL WR Analysis

Gina Champion, Peter Cullen, Nathan Deweerdt



Executive Summary

Each year, hundreds of new players enter the NFL draft with the hopes of becoming the next league MVP. Because college players entering the draft could have unique backgrounds in terms of team, conference, schedule, etc., the NFL combine is held each spring before the draft to standardize metrics related to player athletic ability. This project will analyze these metrics historically to gauge the accuracy of combine results & draft position on the impact the players have made in the NFL.



Business Case & Objectives

- A lot of pressure rides on NFL teams to select the right player at the right time in the draft.
- Comparing NFL combine statistics & draft position to actual NFL career statistics to help understand the relevant indicators of top-tier, NFL talent.

For this project, we used data for Wide Receivers from the combine, draft, and regular season games to analyze what metrics may be better predictors of NFL performance than others.

Data selection

Play level NFL data.

20 files each in 2NF+

Sanity Checks

The screenshot shows the Kaggle dataset page for "NFL Play Statistics dataset (primary)". The page header includes the dataset name, a description ("NFL play-by-play dataset with participation information, 2004-present"), the creator's name (Todd Steussle), and the update date (updated a year ago). Below the header is a navigation bar with tabs: Data, Task, Code (2), Discussion (3), Activity, and Metadata. The "Code" tab is highlighted with a red circle. To the right of the tabs are buttons for "Download (135 MB)" and "New Notebook". Below the navigation bar is a section for "Usability 10.0", "License CC0: Public Domain", and "Tags sports". The "Description" section contains three paragraphs of text about the dataset's scope and sources.

NFL Play Statistics dataset (primary)
NFL play-by-play dataset with participation information, 2004-present

Todd Steussle • updated a year ago

Data Task **Code (2)** Discussion (3) Activity Metadata

Download (135 MB) New Notebook

Usability 10.0 License CC0: Public Domain Tags sports

Description

NFL is one of the most popular sports in the world. Many of us are stat geeks who understand not what just happened but also who and why. This NFL dataset provides a comprehensive view of NFL games, statistics, participation, the annual NFL combine, and the NFL draft. The dataset includes NFL play data from 2004 to the present.

This NFL dataset provides play-by-play data from the 2004 to 2019 seasons. Dataset also includes play and participation information for players, coaches, and game officials. Additional data tables included in this file includes NFL Draft from 1989 to present, NFL Combine 1999 to present, NFL rosters from 1998 to present, NFL schedules, stadium information and much more.

The granularity of NFL statistics varies by NFL season. The current version of NFL statistics has been collected since 2012. All information sources used to create this dataset are from publically accessible websites and the NFL GSIS dataset.

All information sources used to create this dataset are from publically accessible websites and NFL documentation. Although my current life is focused on data science, this project has a special place in my heart, since it links my previous profession in the NFL with my current passion for data analysis.

Data Explorer

836.29 MB

- combine.csv
- draft.csv
- fumbleForced.csv
- fumbles.csv
- gameParticipation.csv
- games.csv
- interceptions.csv
- kickReturns.csv
- kicks.csv
- officials.csv
- passDef.csv
- passer.csv
- penalties.csv
- players.csv
- plays.csv
- qhHits.csv
- receiver.csv
- rusher.csv
- sacks.csv
- tackles.csv

The screenshot shows the "Data Explorer" interface for the "combine.csv" file (2.25 MB). The interface has a sidebar with a list of files, including "combine.csv", "draft.csv", "fumbleForced.csv", "fumbles.csv", "gameParticipation.csv", "games.csv", "interceptions.csv", "kickReturns.csv", "kicks.csv", "officials.csv", "passDef.csv", "passer.csv", "penalties.csv", "players.csv", "plays.csv", "qhHits.csv", "receiver.csv", "rusher.csv", "sacks.csv", and "tackles.csv". The main area displays the "Detail" view for "combine.csv". It includes a table of data with columns: combineId, playerId, combineYear, combinePosition, and combineHeight. The table shows data for the first 10 rows of the file. Below the table, there are five charts: a bar chart for "combineId", a bar chart for "playerId", a bar chart for "combineYear", a bar chart for "combinePosition", and a bar chart for "combineHeight". The "combinePosition" chart shows a distribution of positions, with WR at 13%, CB at 10%, and Other (7779) at 77%.

< combine.csv (2.25 MB)

Detail Compact Column 10 of 35 columns

About this file

Documents participation in NFL Combine from 1987 to present. Includes globally unique identifier, as well as relevant foreign keys (playerId).

combineId	playerId	combineYear	combinePosition	combineHeight
Globally unique identifier	Unique identifier of NFL players	Year of the NFL Combine. Four digit numerical code	Position group player participated in during the Combine	Player height measured in inches
10000	19.9m	1987	WR 13%	64.9
20.5k	20.2m	2019	CB 10%	82.4
18888	19870867	1987	Other (7779) 77%	69.8
18881	19870154	1987	CB	74.8
18882	19870888	1987	C	71.8
18883	19870121	1987	FB	75.0
18884	19870801	1987	C	72.1
			WR	



Python Preprocessing and Aggregation

- No Touchdown Attribute
 - PlayStats (string of play description) included TDs called back and turnover TDs
 - scorePossession included turnovers immediately resulting in a score
 - `final_df["play_end_in_TD"]=(final_df["yards_match"]==0.0) & (final_df['playStats'].str.contains('TD')) & (final_df["turnover"]==0) & (final_df['scorePossession']>5) & (final_df["playType2"]=="pass, complete")`
- Duplicate Plays
 - Seemed like an easy fix, but...
 - Duplicates represented missing data
 - Attempted to integrate an alternative data source

Be careful with unvetted data

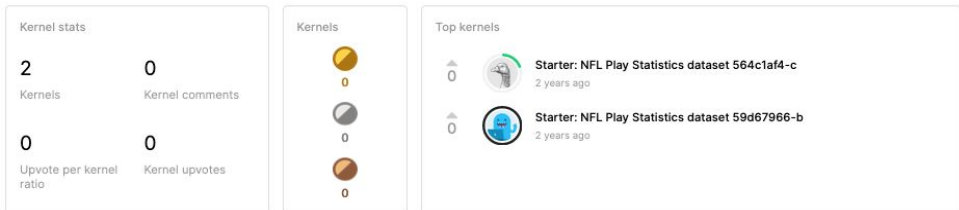
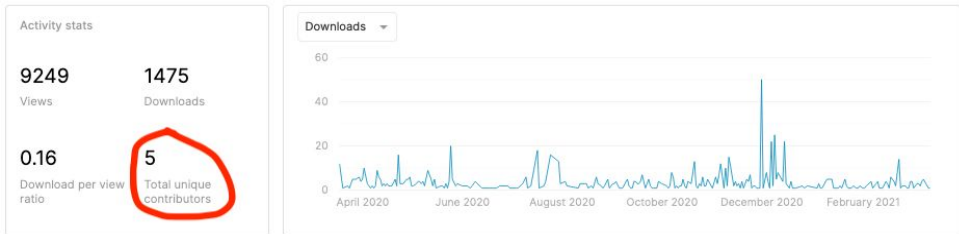
NFL Play Statistics dataset (primary)
NFL play-by-play dataset with participation information, 2004-present.

Todd Steussle • updated a year ago

Data Tasks Code (2) Discussion (3) **Activity** Metadata

Download (135 MB) New Notebook

Card data was last refreshed just now



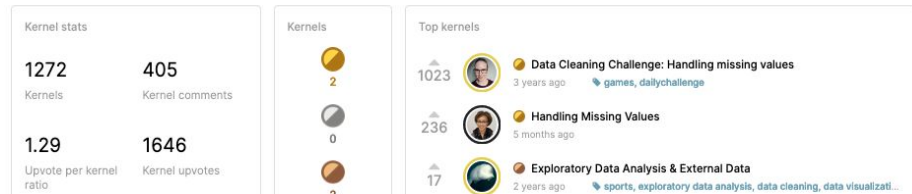
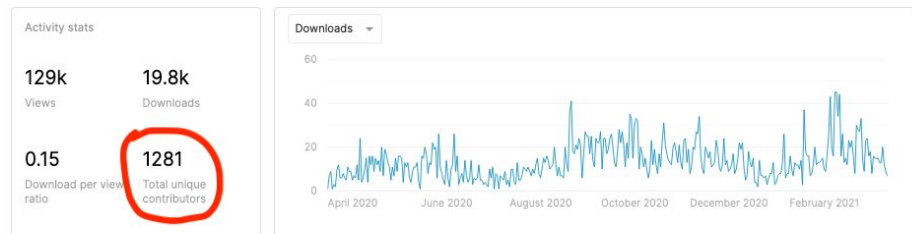
Detailed NFL Play-by-Play Data 2009-2018
nflscraper generated NFL dataset with expected points and win probability

Max Horowitz • updated 2 years ago (Version 6)

Data Tasks Code (1,272) Discussion (25) **Activity** Metadata

Download (274 MB) New Notebook

Card data was last refreshed 7 hours ago



Database Creation: Creating Tables in MySQL

Step 1: Table Creation

```
DROP TABLE IF EXISTS `nfl`.`draft` ;
```

```
CREATE TABLE IF NOT EXISTS `nfl`.`draft` (  
  `draft_id` VARCHAR(15) NOT NULL,  
  `draft_age` FLOAT NOT NULL,  
  `draft_year` INT NOT NULL,  
  `draft_round` INT NOT NULL,  
  `overall_pick` INT NOT NULL,  
  `wr_pick` INT NOT NULL,  
  PRIMARY KEY (`draft_id`))  
ENGINE = InnoDB  
DEFAULT CHARACTER SET = utf8;
```

Step 2: Data Import

```
-- Draft Table
```

```
INSERT INTO draft(draft_id,draft_age,draft_year,draft_round,  
overall_pick,wr_pick) VALUES (497110,22.137,2010,3,89,12);  
INSERT INTO draft(draft_id,draft_age,draft_year,draft_round,  
overall_pick,wr_pick) VALUES (497190,21.6712,2010,2,36,3);  
INSERT INTO draft(draft_id,draft_age,draft_year,draft_round,  
overall_pick,wr_pick) VALUES (497218,24.0027,2010,6,197,24);  
INSERT INTO draft(draft_id,draft_age,draft_year,draft_round,  
overall_pick,wr_pick) VALUES (497222,22.074,2010,7,227,29);  
INSERT INTO draft(draft_id,draft_age,draft_year,draft_round,  
overall_pick,wr_pick) VALUES (497270,21.6329,2010,2,39,4);
```

Step 3: Further Cleaning

```
SET sql_safe_updates=0;
```

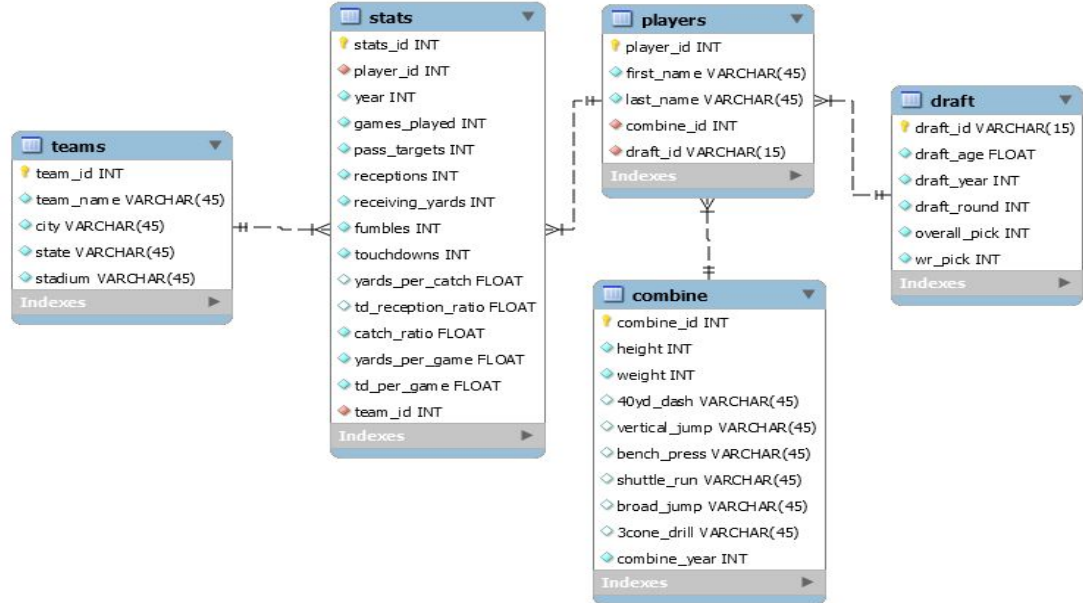
```
DELETE FROM nfl.players  
WHERE draft_id NOT IN  
(SELECT draft_id  
FROM nfl.draft);
```

Step 4: Foreign Keys

```
ALTER TABLE nfl.players  
ADD CONSTRAINT `fk_players_draft1`  
FOREIGN KEY (`draft_id`)  
REFERENCES `nfl`.`draft` (`draft_id`)  
ON DELETE NO ACTION  
ON UPDATE NO ACTION;
```

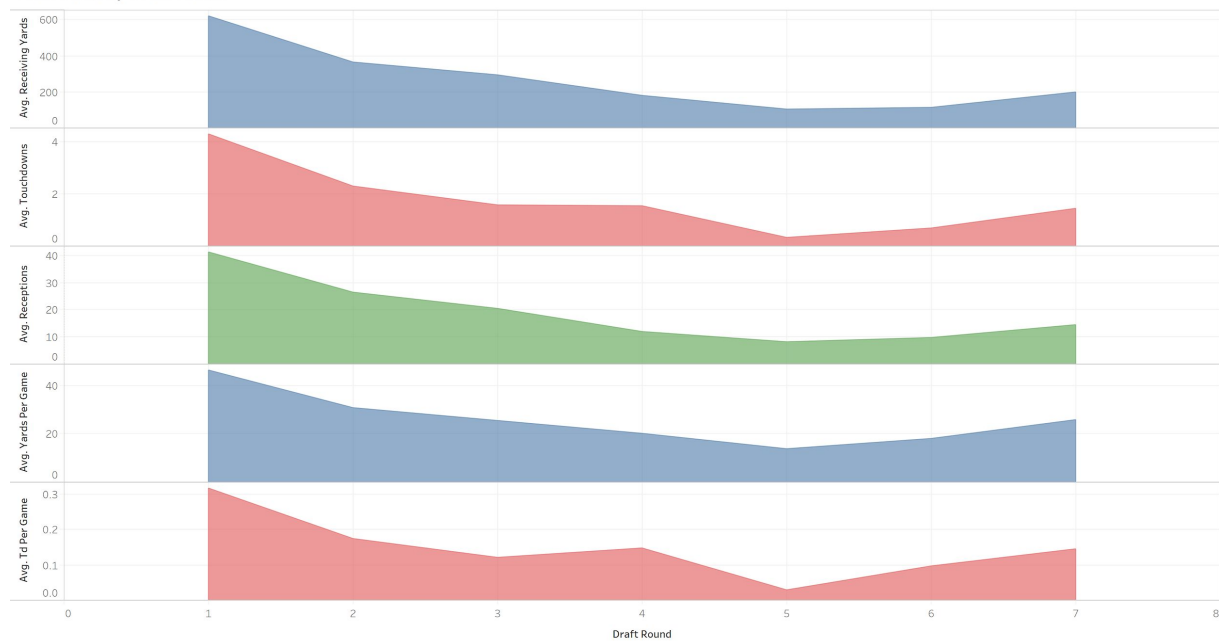
Database Creation: EER Diagram

- Model Summary
- Relationships Overview
- Data Challenges



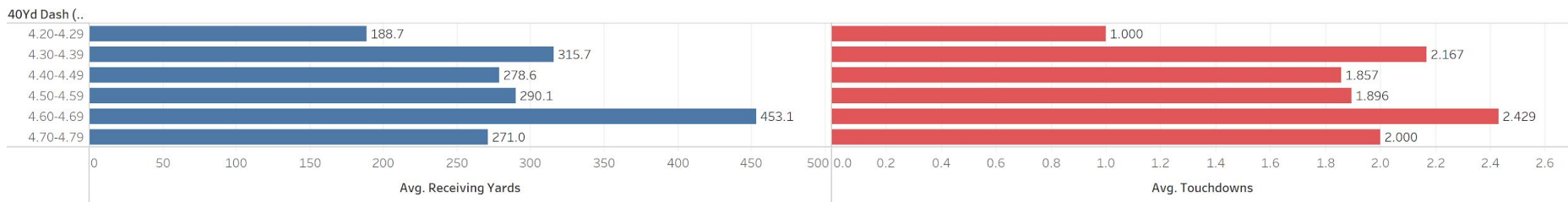
Insights & Findings

Total Stats by Draft Round

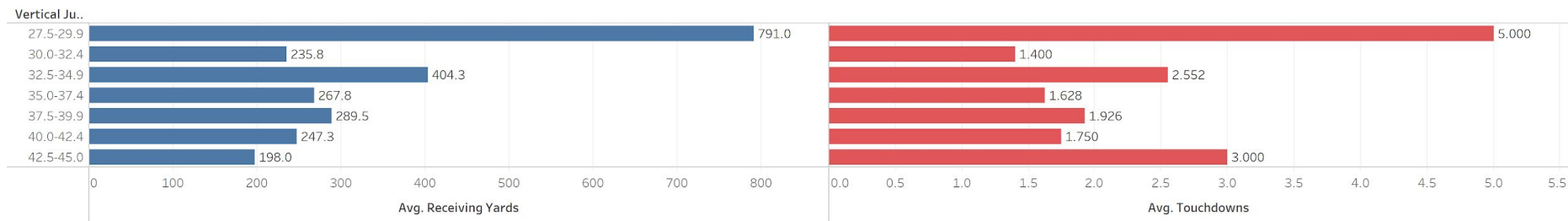


Insights & Findings

by 40 yard dash time



by vertical jump



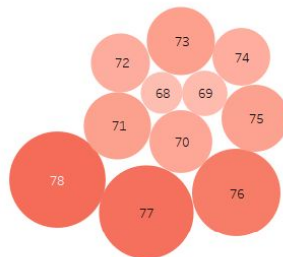
Insights & Findings



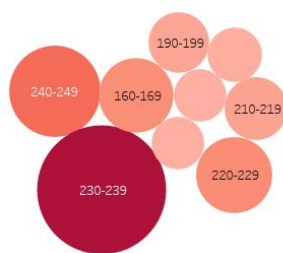
Avg. Receiv... 122.9 928.0

Avg. Touch... 0.750 8.000

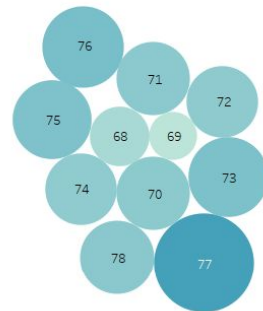
Touchdowns, by combine height



Touchdowns, by combine weight



Yards, by combine height

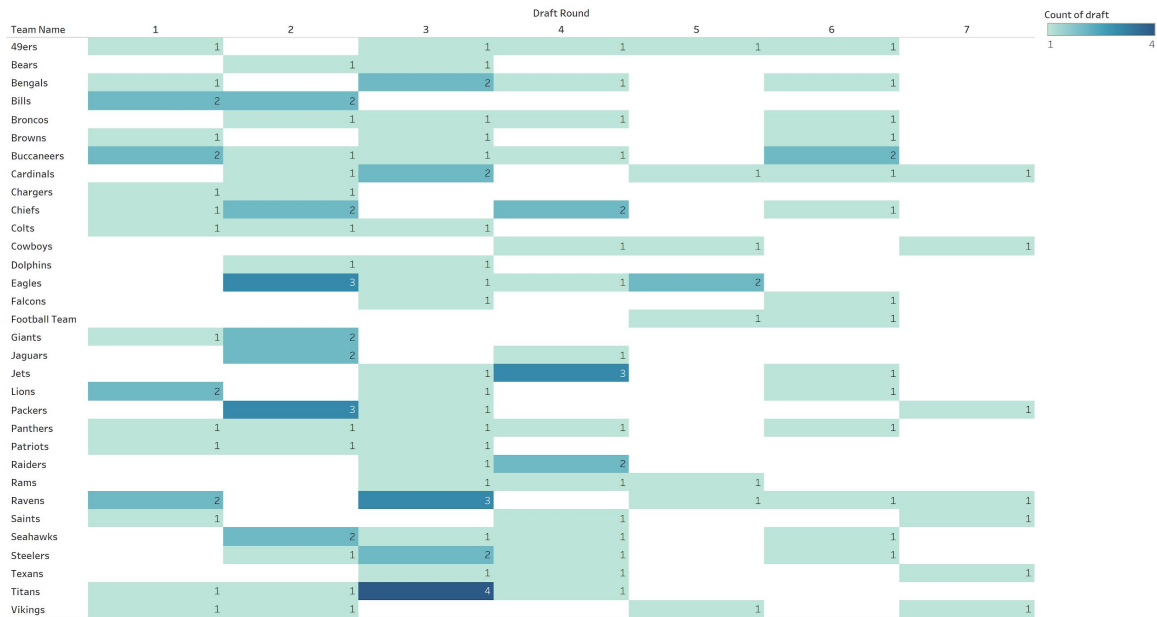


Yards, by combine weight



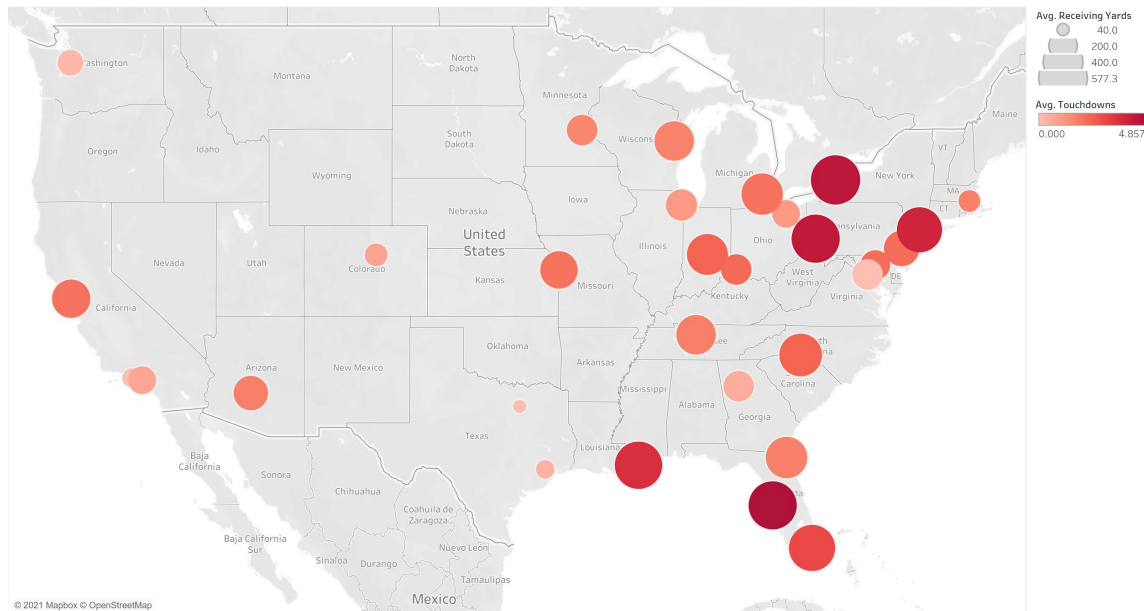
Insights & Findings

Draft Picks, by Team



Insights & Findings

By Team





Questions?