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.

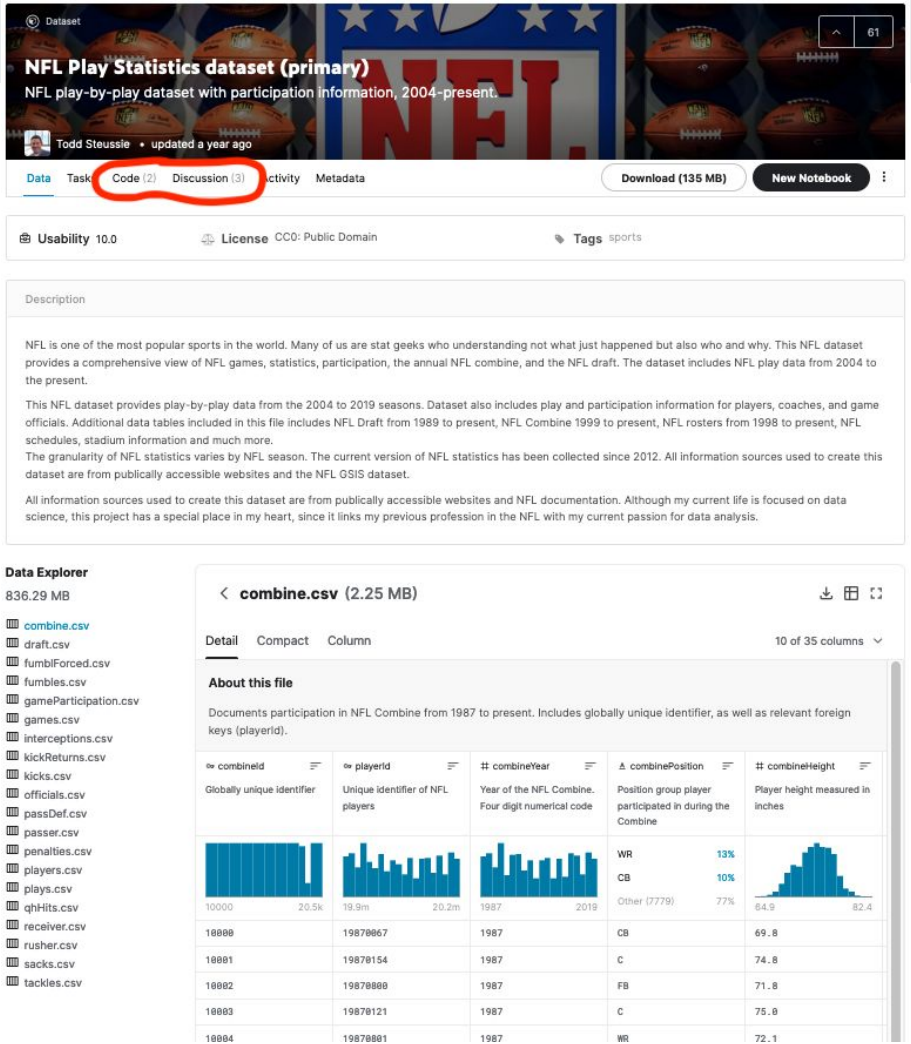
For this project, we used data for Wide Receivers from the combine, draft, and regular season games to analyze what metrics may be better predictors of NFL performance than others.

Data selection

Play level NFL data.

20 files each in 2NF+

Sanity Checks



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

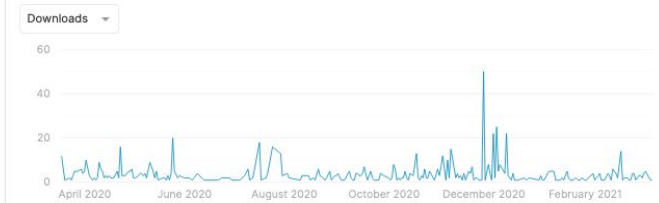
Be careful with unvetted data

Data Tasks Code (2) Discussion (3) Activity Metadata — Download (135 MB) — New Notebook

Card data was last refreshed just now

Activity stats

9249 Views	1475 Downloads
0.16 Download per view ratio	5 Total unique contributors



Top contributors

- DSchaller — 1 total upvote
- Todd Steussie — 0 total upvotes
- Brandon CORTESE — 1 total upvote

Top contributor countries

Kernel stats

2 Kernels	0 Kernel comments
0 Upvote per kernel ratio	0 Kernel upvotes

Kernels

Top kernels

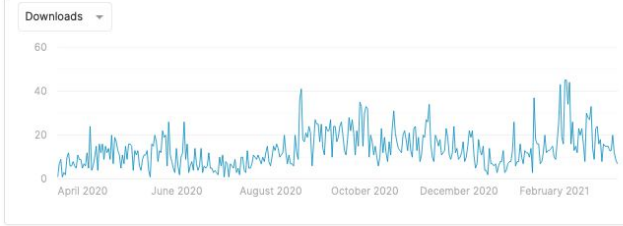
- 0 — Starter: NFL Play Statistics dataset 564c1af4-c — 2 years ago
- 0 — Starter: NFL Play Statistics dataset 59d67966-b — 2 years ago

Data Tasks Code (1,272) Discussion (25) Activity Metadata — Download (274 MB) — New Notebook

Card data was last refreshed 7 hours ago

Activity stats

129k Views	19.8k Downloads
0.15 Download per view ratio	1281 Total unique contributors



Top contributors

- Rachael Tatman — 1023 total upvotes
- Alexis Cook — 236 total upvotes
- beluga — 17 total upvotes

Top contributor countries

Kernel stats

1272 Kernels	405 Kernel comments
1.29 Upvote per kernel ratio	1646 Kernel upvotes

Kernels

Top kernels

- 1023 — Data Cleaning Challenge: Handling missing values — 3 years ago — games, dailychallenge
- 236 — Handling Missing Values — 5 months ago
- 17 — Exploratory Data Analysis & External Data — 2 years ago — sports, exploratory data analysis, data cleaning, data visualizati...

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

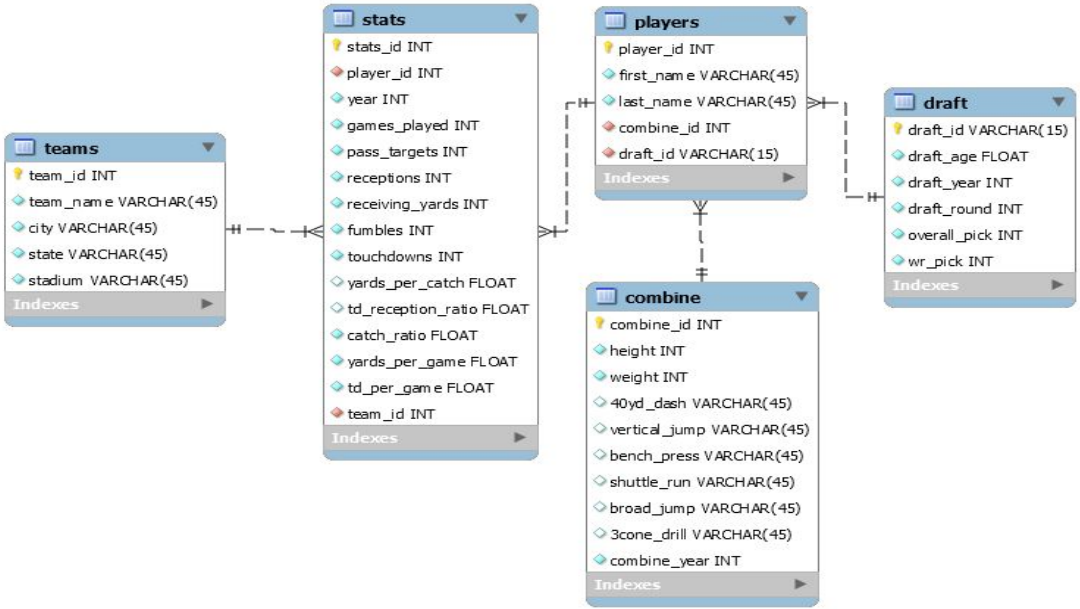
```
SET sql_safe_updates=0;

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

Step 4: Foreign Keys

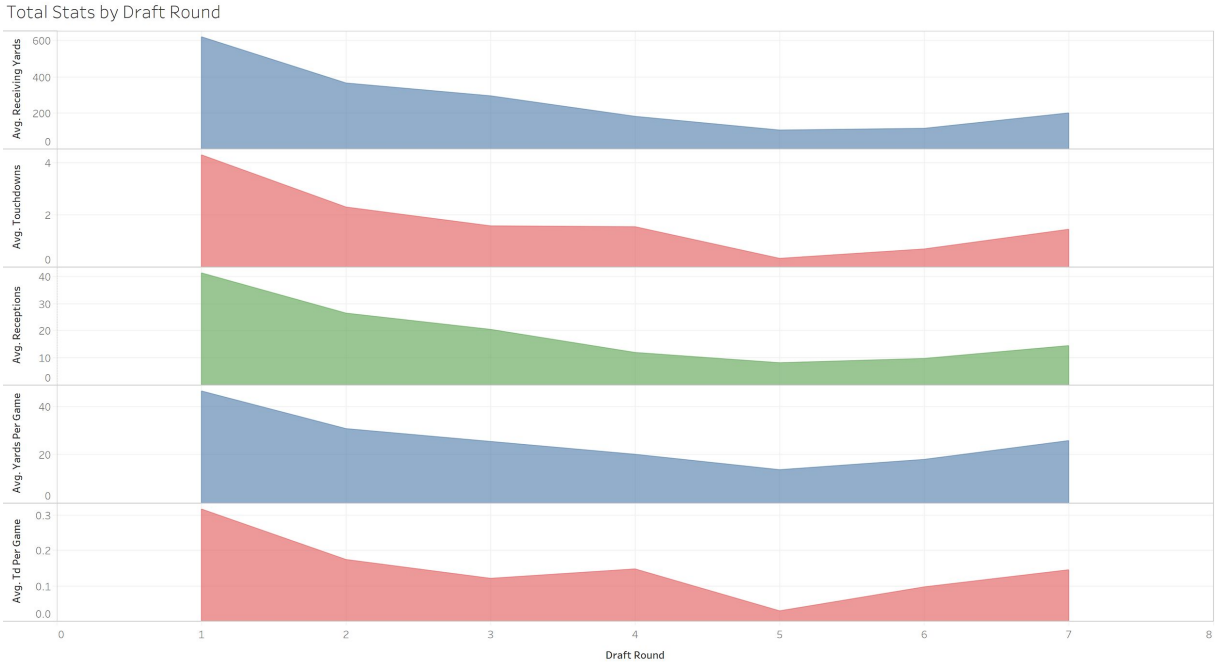
```
ALTER TABLE nfl.players
ADD CONSTRAINT `fk_players_draft1`
FOREIGN KEY (`draft_id`)
REFERENCES `nfl`.`draft` (`draft_id`)
ON DELETE NO ACTION
ON UPDATE NO ACTION;
```

Database Creation: EER Diagram

- Model Summary
- Relationships Overview
- Data Challenges

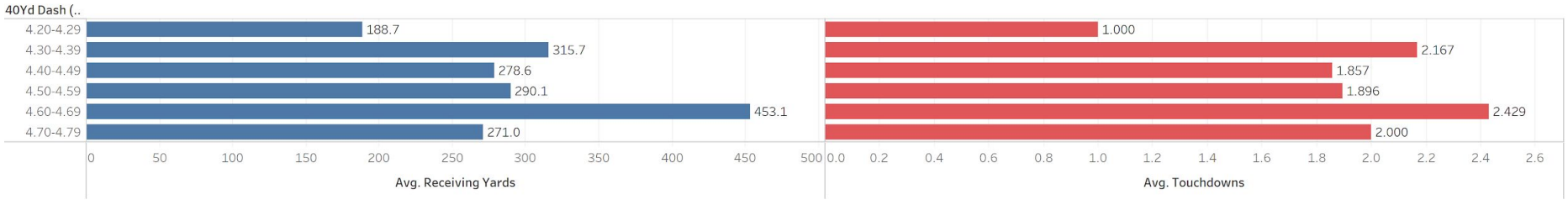


Insights & Findings

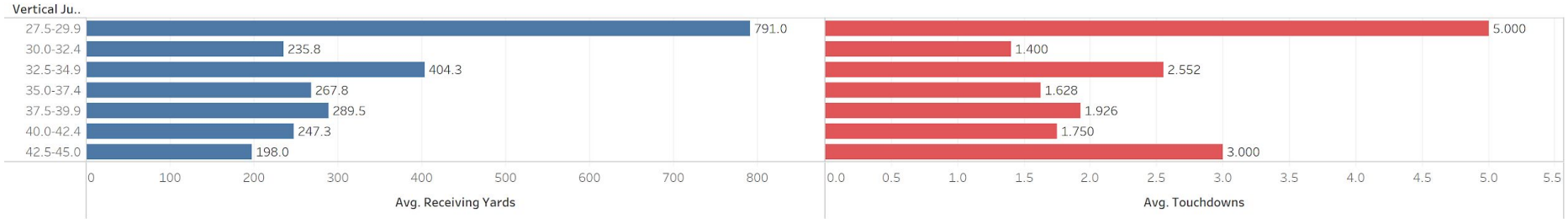


Insights & Findings

by 40 yard dash time

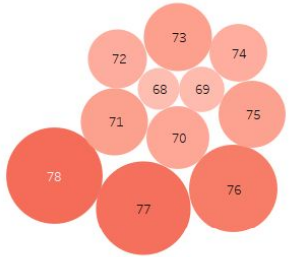


by vertical jump

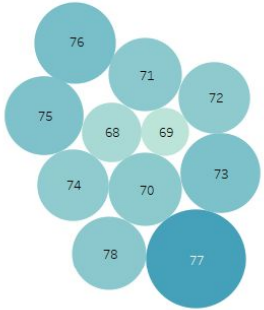


Insights & Findings

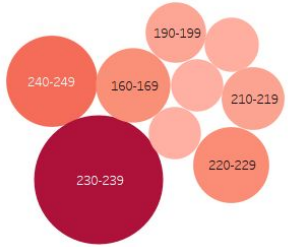
Touchdowns, by combine height



Yards, by combine height



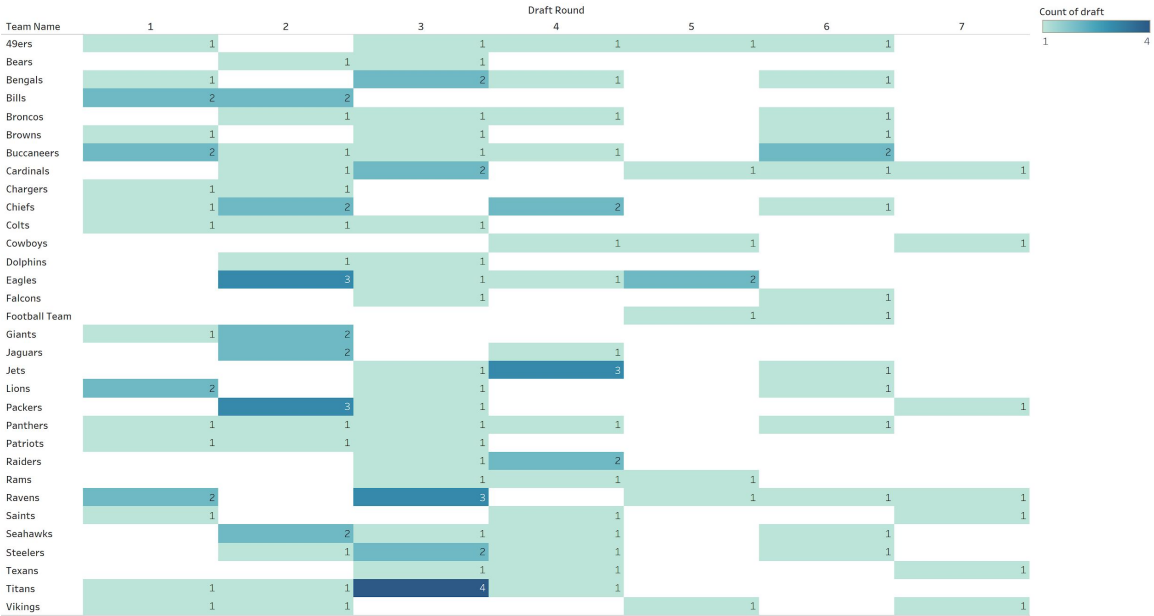
Touchdowns, by combine weight



Yards, by combine weight

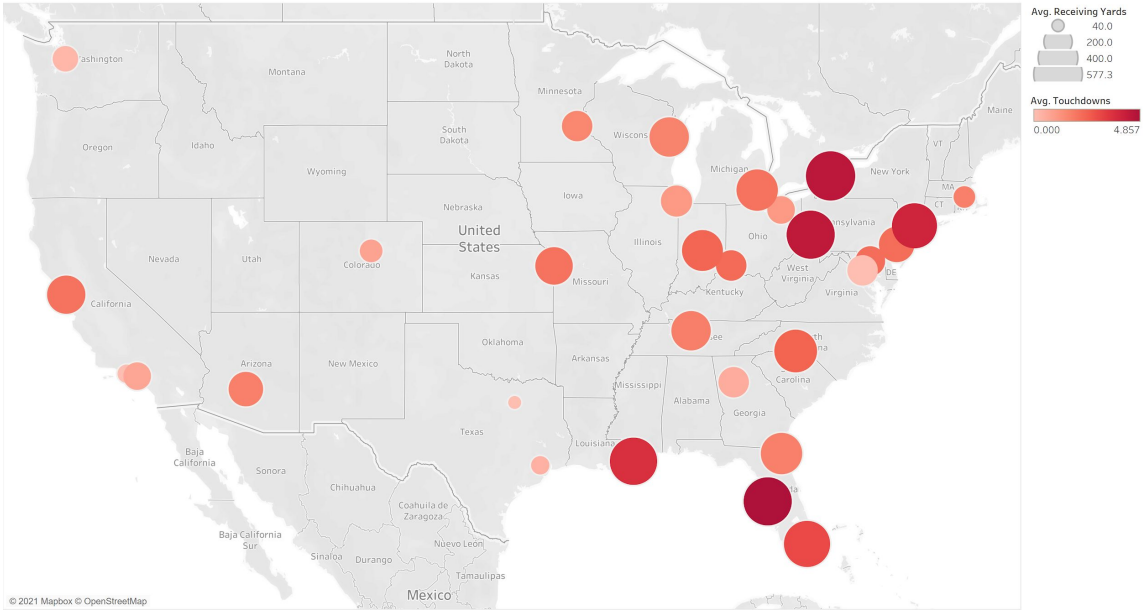
Insights & Findings

Draft Picks, by Team



Insights & Findings

By Team



Questions?