Applied Statistics Project

Aniruddhan, Jack, Jaehyuk, Soumyajyoti

November 2025

1 Introduction: Questions to answer and null hypothesis

Understanding how NFL teams allocate their salary cap spending and draft capital across positions is critical for evaluating organizational strategy and predicting competitive success. NFL operates under a strict salary cap that creates a zero-sum environment: every dollar spent on one position represents an opportunity cost elsewhere. Similarly, draft capital is a scarce resource that teams must strategically deploy to build sustainable rosters.

This analysis examines three fundamental questions about NFL resource allocation from 2013 to 2024:

1. How are allocations changing over time? The evolution of the game, including rule changes favoring passing offenses, the rising importance of mobile quarterbacks, and shifting defensive schemes, may drive systematic changes in how teams value different positions. We investigate whether league-wide spending patterns show temporal trends that reflect these strategic adaptations.

2. How different are allocations with respect to teams? While all teams operate under the same salary cap constraints, organizational philosophies vary considerably. Some teams may prioritize premium positions (QB, pass rusher) while others distribute resources more evenly. We examine the heterogeneity in team allocation strategies and whether certain teams consistently deviate from league norms.

3. How do allocations relate to outcomes? Here we address the key question of how allocation decisions translate into winning. We model the relationship between positional spending (both salary cap and draft capital) and team win percentage to identify which allocation strategies, if any, are associated with competitive success.

For our primary outcome analysis (Question 3), we test the null hypothesis that positional resource allocation has no significant relationship with team win percentage. Specifically:

Null Hypothesis (H_0): Team win percentage is independent of the distribution of salary cap and draft capital across positions. Formally, in a regression framework:

$$H_0 : \beta_{\text{cap}_p} = 0 \text{ and } \beta_{\text{draft}_p} = 0 \text{ for all positions } p$$

where β_{cap_p} and β_{draft_p} represent the coefficients for salary cap and draft capital allocation to position p, respectively.

Rejecting this hypothesis would suggest that how teams allocate resources, not merely how much they spend, meaningfully influences competitive outcomes, providing actionable insights for NFL decision-makers.

2 Identification of dataset

For this project we used data from [5] and used the nfl read py ([4]) package to read the data into python rather than R. This data set contains volumnious data about the NFL ranging from play-by-play data from all NFL games, to the referees, weather, and roster for that particular game. We choose to analyze data from 2013 to 2024. The quality of contract data is much higher past 2013 and a labor dispute in the NFL in 2011 created abnormal cap spending which would make analysis more difficult.

For our project we are interested in analyzing how NFL teams' use their spending (cap) and draft resources to allocate spending across positions. To that end we used the contracts and draft data sets that are available from within the nfl read py ([4])package. In order to analyze outcomes we looked at regular season win percentage, which was obtained from the schedule dataset also in the nfl read py package.

The contracts data contains information on nfl player contracts for each season in the nfl going back to at least our cutoff year of 2013. For this data we are primarily interested in the percentage of the salary cap that is being spent on each position. This data is present for almost all contracts going back to our cut off year of 2013. The data set used in [4] is sourced form [2], a website which diligently tracks all of the information about nfl player contracts.

The drafts data contains information on the NFL draft for each season going back to 1983. We chose just to look at the NFL draft going back to 2013 to stay consistent with the contracts data. Draft picks were assigned a value based on the [6]. This chart was chosen because it was hosted on the same website as the contracts data, was easy to scrape, and had an understandable methodology.

3 Existing Literature

We were unable to find many recently published article regarding our questions, however this has been studied. [1] studies the impact of inequity in player pay within teams and its effect on performance. This study does not combine both salary and draft capital spending to get a full picture of how NFL teams are allocating fixed resources.

While it is difficult to find journal articles about this question, media coverage of the NFL in America is ubiqutotous and many articles, discussions, tv shows, and game discussions are rife with discussion of contract and draft capital allocations. This is most evident in the coverage of the NFL draft and articles like [3] which discusses how NFL teams should handle players and their contracts and the opprotunity costs inherent with their current trajectory. This shows that there is wide and varied interest in questions of this type.

4 Concerns about Data

4.1 Inconsistencies between the datasets

Between all data sets, one of the concerns was inconsistency regarding team names. NFL teams have relocated, causing their name to be listed slightly differently across years. Additionally the same team might be named the 'Los Angeles Chargers' in one data set and 'LAC' in another for example. This was fixed by looking at all unique team names across all 3 data sets and standardizing it.

Another inconsistency was with the positions. The contracts data listed positional data ata a more granular level then the drafts data, for example listing RT and LT instead of just T (for the tackle position). This was fixed by creating a standardized position listing and collapsing all of the lineman positions into one position, for Offensive Line and Defensive Line respectively.

4.2 Problems specific to: Contracts Data

For the contracts data the biggest problem was that some players had multiple positions. Thus the percentage of their salary cap would be double counted when we would group by position. For example a player could be listed at both running back and receiver and receive 5% of the salary cap. Then it would look like that same 5% was allocated to the running back position and the wide receiver position. This was fixed by simply dividing the allocation by the total number of positions for that player, and so in this example 2.5% would be given to the running back and receiver position respectively.

4.3 Problems specific to: Drafts Data

Outside of the inconsistencies noted above, the drafts data was easy to work with.

4.4 Problems specific to: Schedule Data

For the schedule data, the only concern was how to handle ties. In the NFL, ties count as half a win and half a loss, so once ties are properly accounted for, they will not be hard to include in the win percentage.

4.5 Problems specific to: Draft Value Data

The concern with the draft value data is that draft pick value was assigned based on a statistical analysis of contracts given out to players drafted at that position. Thus our analysis of draft pick allocation can only be as good as the analysis done to assign values to each draft pick. Due to time constraints we could not fully analyze the analysis performed as there was an entire book written about it ([7]).

5 Plan of analysis

5.1 Cleaning Data

5.1.1 Contracts Data

Contracts data originally comes in a listing where each row is a contract, and this was difficult to untangle due to the complexity of the NFL salary cap, as it was not immediately clear how much of each contract was allocated to each team by the year. Additionally, players can switch teams throughout the duration over their contract if they are traded. Luckily there was a column in the data set called 'cols' which broke down for each player, how much they were paid, and which team paid it, for each season. For each row in the original data set (which was by player, contract), this column was a dictionary which contained all the information for the player for each year, regardless of the contract information on that row.

By extracting information from this column for each row we were able to construct a data set which had for each player, and for each season, the percentage of the salary cap that they took up. For each player, we also had the position that they played. This created a challenge because when looking at allocation across position, we could potentially double count the amount that a player was making, since it would count the total percentage of the cap that they were taking up and add it to the allocation for each position. This was solved by creating a uniform positional mapping. This solved many of the dual listings by collapsing all offensive lineman into one position for example. It also had the advantage of creating consistency between the draft data set and the contracts data set. For positions that

The final challenge with the contracts data was that the contracts data did not contain information about veteran and undrafted free agent signings, dead cap, and unused cap each season. We knew that this was related to the dead cap by cross referencing teams which had a very low percentage of their salary cap used and our knowledge of nfl history and saw that nfl teams with a low percentage of cap used also had a high dead cap hit that season. Thus when summing up the percent of salary cap spent on each position per team we would very rarely get exactly 1. (See figure in exploratory data analysis section). To resolve this we created a dummy position that represented, for each season, the percentage of the cap that was unaccounted for in the data, knowing that this represented the amount that the team spent on veterans, dead cap, and unrestricted free agent. This allows us to have clean data and creates another interesting variable for analysis.

Using the dummy VDU position we were able to get a percentage of team salary cap spent by position and also a total percentage of league wide salary cap spent on each position (including VDU). Additional tasks doen to clean the data was to standardize team names, including fixing the team name for the Washington team and joining to the standardized team name file.

5.1.2 Drafts Data

The drafts data was cleaned by joining draft picks, their teams, and the position drafted to the draft value chart. The draft value chart was scraped from the html on its website. The draft values were then summed up for each year and for each team to get a percentage of draft capital allocation by team and by the league from which we could analyze league wide trends. Additional tasks doen to clean the data was to standardize team names, including fixing the team name for the Washington team and joining to the standardized team name file.

5.1.3 Schedule Data

NFL schedule data was cleaned in order to get a data set containing the winning percentage by team and by season. The original data listed each game and the score for the home and away team. I added additional columns which indicated the winning team, losing team, and listed both teams in the event of a tie. Then, using a deduplicated data set for each team, we joined to the schedule data set and got a count of all of the results for each team. Then we divided by the games played to get the winning percentage.

5.2 Analyzing Cleaned Data

The plan to analyze the cleaned data is to apply various linear models to two cleaned data sets. One which has the position and allocation in terms of draft capital and salary by year for the entire league and another that has it broken out by team. Then consider the significance of our variables and answer our major questions.

6 Exploratory data analysis

Our exploratory data analysis looked at the percentage of salary cap ($) spent by each team to see how necessary it was to create the VDU position. We looked at just the year 2017 as an example. The chart shows each teams cap allocation. The cap allocated to players whose contracts we were able to track is orange, and the remaining share of the cap, up to 100% is in blue.. As one can see in the chart, in the year 2017 most teams spent a relatively small percentage of the cap on VDU, however, some teams, like the Browns and 49ers, spent an excessive amount. For the year 2017 in particular the Browns and 49ers had a lot of dead

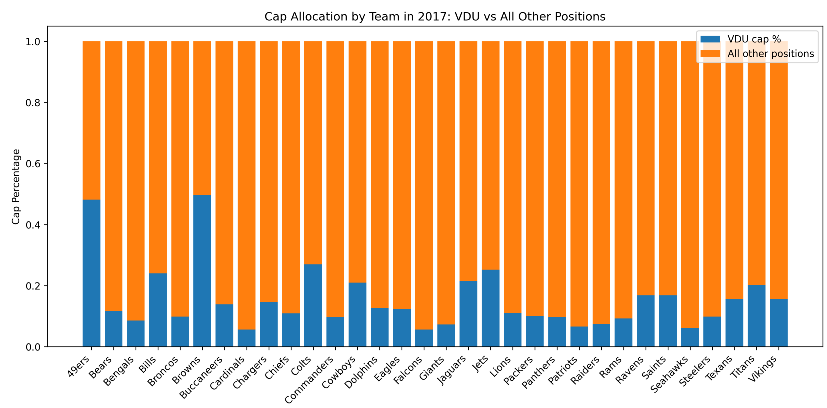

Figure 1: Cap Allocation by Team in 2017: VDU vs all other positions

cap spend due to having to pay players no longer on their roster. These cases illustrate why the VDU category is necessary. Without it, the distribution of cap spending across positions would be distorted by the amount spent on dead or unused cap space. By treating VDU as its own position we create a consistent basis for comparing cap allocation across teams and can further identity situation where cap usage reflects mismanagement and inefficiency.

References

[1] Richard Borghesi. Allocation of scarce resources: Insight from the NFL salary cap. *Journal of Economics and Business*, 60(6):536–550, 2008.

[2] Over The Cap. Contracts — nfl salary cap, player contracts, salaries and bonuses. `https://overthecap.com/contracts`, 2025. Accessed: 2025-11-15.

[3] IMissFatRex. Resource allocation and new york jets cornerback sauce gardner: Should the jets consider trading him?, Dec 2023. Accessed: 2025-11-15.

[4] nflverse development team. nflreadpy: A python package for downloading nfl data from nflverse repositories.

[5] nflverse development team. nflreadr: Efficiently download nflverse data (underlying data repositories).

[6] OverTheCap. "fitzgerald-spielberger nfl draft trade value chart". https://overthecap.com/draft-trade-value-chart, 2025. Accessed: 2025-11-15.

[7] Brad Spielberger. *The Drafting Stage: Creating a Marketplace for NFL Draft Picks*. CreateSpace Independent Publishing Platform, 2020. Kindle edition available on Amazon. Accessed: 2025-11-15.