DSCI-510
Fall 2025

The NFL Running Back Value Debate

John Minard

What is the Running Back Debate?

The Running Back Debate

With Running Backs holding out for training camp and even games, the question has to be asked; are running backs truly less valuable than other skill positions? Or are teams reacting to the perceived risk and productivity of the running back position?

Goal of this project

The goal of this project is to use contracts and performance data to evaluate running backs over time, and determine whether or not they are undervalued.

Core questions to address

1. Does Running Back production correlate with team wins?
2. Do Running back contracts align with performance?
3. Are teams justified in paying the Running Back position less?

Data Sources

These are the primary data sources used due to the data they provide *(production, money, team outcomes)*

1

ESPN Public API

This provides season statistics from a team level (wins, losses, and offensive metrics)

2

Pro-Football-Reference

Pro-Football-References (PFR) provides player season stats such as rush attempts, rushing yards, TDs, receptions, fumbles

3

OverTheCap

OverTheCap provides data on contracts and the contract structure (AAV, guarantees, length)

Data Obtained

	ESPN API (Obtained)	Pro-Football-Reference (Planned)	OverTheCap (Planned)
DESCRIPTION	Season level statistics	Player level statistics	Player contract data
What has been collected	• Team ID, Team Name, and Abbreviations • Record (Wins and Losses) • Team-level stats (rushing offense yards, passing offense yards)	Will scrape all RB player-seasons Metrics will be: - Attempts - Yards - TDs - YPC - Receptions - Fumbles	Will pull player contracts and contract structure Metrics will be: - Average Annual Value - Guaranteed $ - Cap Hit - Length
Purpose + Validation	Purpose • Establish baseline relationships between success and run-game performance • Will integrate with PFR data Validation • Test file (tests/test_api.py) passed	This data will enable regression and clustering	This data will be used to link salary to performance/usage/longevity

Planned Models + Analysis

1

Regression Models

Predict salary using:

- Rushing Yards
- Touchdowns
- Yards per Carry
- Receptions
- Fumbles
- Wins

Purpose: Identify the value of Running Backs

2

Clustering

This will group Running Backs into “Value Tiers” based on their production + salary

3

Longevity / Workload Analysis

This will help to show the relationship between touches, age, and decline curves

This will help to support the argument about Running Back career length and longevity

Market trends

Expected Results

Main Hypothesis:
Teams are justified in paying Running Backs less overall, there are exceptions to dual-threat Running Backs and they may be undervalued

Hypothesis

Running Backs are less correlated with wins than Quarterbacks or Wide Receivers.

Running Back production peaks early during their first 4-5 years (rookie contracts) heavy usage accelerate their decline

Expectations

Clear segmentation between high-value and low-value Running Back profiles

Evidence around diminishing returns for high workload Running Backs

Current Progress:

Repo structured

Working ESPN API integration

Progress Report

Dataset built constructed

Next Steps:

1. Complete PFR and OTC scraping
2. Merge Data
3. Build regression and Clustering models
4. Produce visuals
5. Finalize final project

Final report may have some slight alterations to how data is used (i.e., how often running backs are worth the value, establishing running back value, determining the validity of rookie contracts and then letting running backs walk)