

Progress Report

Project Scope Update

The project's scope is consistent with the original version which is to evaluate whether NFL running backs are undervalued through analysis of the relationship between performance, contracts, and team success. I have implemented a working ESPN API integration to collect team-level statistics. I will continue to utilize PFR and OTC for performance-to-value modeling.

Data Sources

ESPN Public API (currently implemented)

Provides Team names, wins and losses, and offensive yards
Obtained using the python script that is in src/api/espn_nfl.py
Command used “python -m src.api.espn_nfl --season 2024 --save data/espn_team_stats_2024.json”

Pro-Football-Reference and OvertheCap

PFR will provide player-level stats - this will anchor regression models
OTC will provide player contract data - this will help to show value of running backs and whether the salaries reflect production

API used

ESPN's public JSON API: (<https://site.api.espn.com/apis/site/v2/sports/football/nfl/teams>)

Issues / Difficulties

No official API for PFR or OverTheCap - this means that scraping will be used
Data merging challenges - player names, team names, abbreviations, and seasons may require some sort of fuzzy match to combine data sets.
Longevity modeling will require touches and age data, which may need to be drawn from additional scraping.
Salary data complexity will occur as contracts can differ in structure. There may be a need to simplify these contracts to total \$ and length.