Bryce Rasmussen

ECO316

NFL Cap hit Dataset Assignment

Dataset Overview:

This dataset provides data regarding the Cap Hits of all players in the National Football League between the years of 2011 and 2023. Data includes player name, position, team, team statistics, individual salary cap hit, and total salary cap for the given year. This data allows for investigations into questions surrounding the impact of player cap hits on areas such as team performance, salary changes over time, etc. This dataset is particularly useful as it is difficult to find ways to access this type of data in a way that allows exploration and experimentation.

Applications:

An interesting question this data could be used for is a study of the perceived value of players in the quarterback position over time, based on cap hit changes. There are multiple approaches to answering this question. One would be to regress player cap hit as a percentage of the total salary cap on years since 2011, years in the league, and player fixed effects. To do this the number of years in the league would need to be added, which could be found by subtracting the season year from the player's draft year. Draft year could added through a source like ESPN NFL draft Prospect data. The coefficient of importance here is that on years since 2011. We would expect to see a positive value if quarterbacks have on average increased in the percent of the salary cap that they take up.

Another question to explore is if with the decline of run focused offenses over time if the percentage of the salary cap going to primary running backs decreases. The relevant regression would be to regress the aggregated total salary cap hit percentage of the 3 highest paid running backs on a team in a given year on team fixed effects, years since 2011, and rushing yards as a percentage of total team yards. The coefficient of importance here is that on years since 2011. We would expect to see a negative value coefficient when accounting for team changes and performance differences if the theory is correct.

A third question to consider is an investigation of contract efficiency amongst teams. Using a combination of the highest paid positions of a certain set (ex Top 1 QB, 2 RB, 3 WR, 1 TE) regress this value as a percentage of total cap hit against total offensive yards with year and team fixed effects. From this examine the coefficients of the team fixed effects and see if over time certain teams tend to be more efficient than others when it comes to paying for some sort of efficiency, for example total offensive yards. I include an example of a simple regression investigating this topic at the end. There we see that accounting for other variables, certain teams tend to perform worse offensively. The most significant example of this was the Pittsburg Steelers who during this timeframe had had roughly 700 less offensive yards on average than many of the other teams in the league. This would suggest that they make worse salary decisions with their players or that they rely on bringing in more expensive core offensive players.

Creation Details:

The core piece of this dataset was scraped by hand from the year pages on Spotrac NFL Salary Rankings (2021 cap hit example). I highlighted and copied the data roughly 100 players at a time over to an excel spreadsheet where I had to remove the visuals by hand. This was a many hours long process for me to collect all the data from the earliest on the site starting in 2011 to the most recent season 2024. Each season's data was stored in its own csv file such that labeling could occur at a later time. From here I wrote python code to process the raw csv that consisted of a single column of data arranged in the same order for each entry but not always having rank because of ties. The code puts this data together into usable rows for each entry with extra remnants of the copying removed and data properly separated and joined. Here season was also added.

At this time I merged on team performance data from each year found on Kaggle. This data required some amount of cleaning to bring it into line with the naming conventions of the existing dataset I was working with and was also reduced to what I deemed relevant features before merging. Then I collected the NFL yearly salary cap limit to be merged onto the dataset which I collected by hand from Spotrac (Cap History). This was used to create a feature that examined each players cap hit as a percentage of the total salary cap. During this time I also added some additional features to data, examples including salary rankings by position and year, unit, and simplified positions. Lastly, I added a feature called "Total Cap Hit of Scoring 7 (Top 1 QB, 2 RB, 3 WR, 1 TE)". This is meant to provide inspiration for types of features others can create when using this dataset. This specifically was created as it took the total cap hit of the highest paid players of important offensive positions for use in regressions around scoring player pay and team success.

Notes on CBA:

Before the 2020 season a new Collective Bargaining Agreement was made between the NFL Players Association and the team owners. The changes in the agreement created a number of changes that affect the data within this set. An additional regular season game was added in this agreement, bringing the total number of games to 17. The active roster was expanded to 55 from the previous 53 player limit. Additionally, there was created a new "four year player benefit" where up to an additional $1.25 million in player salary can be excluded from the cap of up to 2 players.

GitHub Link:

https://github.com/BryceRas/ECO316_dataset.git

Features:

** Team statistics reflect reduced dataset to reflect variations amongst teams.

- Name: Full name of player
    - Source: Spottrac (Cap Hits)
    - Unique Entries: 7,857

- Team: Team name abbreviation. Based on team location during season.
    - Source: Spottrac (Cap Hits)
    - Unique Entries: 35
- Position: All listed player positions. Players with multiple listed positions in form Pos1/Pos2.
    - Source: Spottrac (Cap Hits)
    - Unique Entries: 62
- Player Cap Hit Rank (Same Year): The rank of the players cap hit for that season. 1 being the highest hit and decreasing from there.
    - Source: Spottrac (Cap Hits)
    - Min: 1
    - Max: 2,549
- Cap Hit: The amount of the total salary cap their contracts consumes in the given season.
    - Source: Spottrac (Cap Hits)
    - Mean: 2,072,587.13
    - Std: 3,242,354.72
    - Min: 11,500
    - 25%: 465,000
    - Median: 780,000
    - 75%: 22,500,000
    - Max: 38,600,000
- Season: Year of season. For example 2022-2023 season listed as 2022.
    - Source: Spottrac (Cap Hits)
    - Min: 2011
    - Max: 2023
- wins: Number of games won by player's team during season.
    - Source: Kaggle (NFL Team Performance)
    - Mean: 8.08
    - Std: 3.08
    - Min: 0
    - 25%: 6
    - Median: 8
    - 75%: 10
    - Max: 15
- losses: Number of games lost by player's team during the season.
    - Source: Kaggle (NFL Team Performance)
    - Mean: 8.08
    - Std: 3.07
    - Min: 1
    - 25%: 6
    - Median: 8
    - 75%: 10
    - Max: 16

- points: Total points scored by player's team in regular season.
    - Source: Kaggle (NFL Team Performance)
    - Mean: 396.1
    - Std: 70.96
    - Min: 193
    - 25%: 320
    - Median: 366
    - 75%: 416
    - Max: 606
- points_opp: Total points allowed by player's team in regular season.
    - Source: Kaggle (NFL Team Performance)
    - Mean: 370.29
    - Std: 55.96
    - Min: 225
    - 25%: 328.75
    - Median: 368.5
    - 75%: 407
    - Max: 519
- total_yards: Total yards gained by player's team over season.
    - Source: Kaggle (NFL Team Performance)
    - Mean: 5,617.55
    - Std: 589.27
    - Min: 3,865
    - 25%: 5,192.75
    - Median: 5,602.5
    - 75%: 6,009
    - Max: 7,474
- turnovers: Total fumbles lost and intercepted passes by player's team:
    - Source: Kaggle (NFL Team Performance)
    - Mean: 22.96
    - Std: 5.98
    - Min: 8
    - 25%: 18.75
    - Median: 23
    - 75%: 27
    - Max: 44
- fumbles_lost: Total fumbles lost by player's team:
    - Source: Kaggle (NFL Team Performance)
    - Mean: 9.22
    - Std: 3.15
    - Min: 2
    - 25%: 7
    - Median: 9

    - 75%: 11
    - Max: 22
- pass_cmp: Total passes completed by player's team.
    - Source: Kaggle (NFL Team Performance)
    - Mean: 354.79
    - Std: 47.33
    - Min: 217
    - 25%: 321
    - Median: 356
    - 75%: 383
    - Max: 499
- pass_att: Passes attempted by player's team.
    - Source: Kaggle (NFL Team Performance)
    - Mean: 562.34
    - Std: 60.19
    - Min: 377
    - 25%: 519
    - Median: 566.5
    - 75%: 605
    - Max: 751
- pass_yds: Total passing yards by player's team.
    - Source: Kaggle (NFL Team Performance)
    - Mean: 3769.15
    - Std: 579.54
    - Min: 2179
    - 25%: 3338.25
    - Median: 3751
    - 75%: 4136
    - Max: 5444
- pass_td: Number of passing touchdowns by player's team.
    - Source: Kaggle (NFL Team Performance)
    - Mean: 24.86
    - Std: 7.67
    - Min: 8
    - 25%: 19.75
    - Median: 24
    - 75%: 29
    - Max: 55
- pass_int: Number of interceptions thrown by player's team.
    - Source: Kaggle (NFL Team Performance)
    - Mean: 13.75
    - Std: 4.58
    - Min: 2

    - 25%: 11
    - Median: 13
    - 75%: 17
    - Max: 30
- pass_net_yds_per_att: Average net passing yards per attempt by player's team.
    - Source: Kaggle (NFL Team Performance)
    - Mean: 6.28
    - Std: 0.74
    - Min: 4.2
    - 25%: 5.7
    - Median: 6.2
    - 75%: 6.8
    - Max: 8.4
- rush_att: Number of rush attempts by player's team.
    - Source: Kaggle (NFL Team Performance)
    - Mean: 433.51
    - Std: 20.9
    - Min: 321
    - 25%: 396.75
    - Median: 429.5
    - 75%: 468.25
    - Max: 596
- rush_td: Number of rushing touchdowns by player's team.
    - Source: Kaggle (NFL Team Performance)
    - Mean: 13.6
    - Std: 5.08
    - Min: 3
    - 25%: 10
    - Median: 13
    - 75%: 17
    - Max: 32
- score_pct: Percent of drives ending in a score by player's team.
    - Source: Kaggle (NFL Team Performance)
    - Mean: 35.50
    - Std: 6.61
    - Min: 18.9
    - 25%: 30.85
    - Median: 35.4
    - 75%: 39.9
    - Max: 52.6
- Salary Cap: NFL salary cap in US Dollars for given season.
    - Source: Spotrac (NFL Salary Caps)
    - Mean: 164,786,538.46

    - Std: 33,791,308.01
    - Min: 120,375,000
    - 25%: 133,000,000
    - Median: 167,000,000
    - 75%: 188,200,000
    - Max: 224,800,000
- Cap Hit %: Percent of salary cap taken by player's cap hit.
    - Source: calculated
    - Mean: 1.26
    - Std: 1.93
    - Min: 0.0053
    - 25%: 0.31
    - Median: 0.46
    - 75%: 1.38
    - Max: 20.88
- Player Cap Hit Rank by Position (Same Year): Rank of player cap hit by year and position.
    - Source: calculated
    - Min: 1
    - Median: 58
    - Max: 320
- Simplified Position: Player positions grouped into larger position groups.
- Unit: Player positions grouped into Offense, Defense, and Special Teams. Players listed on multiple units based on position will be placed on a single unit with priority in order of Offense, Defense, Special Teams.
- Total Cap Hit of Scoring 7 (Top 1 QB, 2 RB, 3 WR, 1 TE): Total cap hit of highest x cap hit players at listed positions. Included as example variable to represent ways in which players' cap hits can be combined for comparisons.