Paul Nguyen

3/7/13

CSE 140

Final Project part 1

<p align="center">Comparison between NFL player stats and Salaries</p>

Research Question:

1. Is there a correlation between the National Football League player stats and their salaries? I want to see if it's just the player's stats over a previous ten years (2000-2009) affects the players' salaries of that last year.

   Answer: There isn't a correlation between the quarterbacks rating and his salary that I could find. After doing more research I discovered that there are many factors leading to a quarterback being signed a contract of ten million dollars vs. a couple of hundred thousands of dollars.

2. How can I determine a players overall score. For quarterback I will be computing the quarter back rating and for the running back I will be determining a method to rate their performance based on yards gained, fumbles and TD

   Answer: I determined his rating by using a formula and I couldn't do the running backs data because of time constants.

Motivation and Background

Football has been one of my pastimes either playing myself at the IMA or Watch College or pro leagues. I was always interested on why players get paid so much and how teams determine how much each player is worth. I know that many players have endorsements but the bulk of their income comes from the National Football League. I have looked at how much each team pays their players overall and it looks like the focus is offense, defense, then special teams. Theses stats will not necessarily help me understand the game more but will give me a broader picture of how each player is paid and how their stats affect their score. This is also relevant to the fantasy football league that I my brothers and I at my fraternity are starting up. These computations like the quarter back stats and running back stats will help me determine how to pick players. Having learned how to compute stats about football players will ultimately help me beat my brother next season when trading and choosing my players.

Data sets

The data set I'm using is on http://www.playerfilter.com/nfl/passing-leaders/#/g,r;comp;att;pct;pyds,desc,r;ptd;ptdr;py_a;intc;intr;qbrat;tdintc;fpts;&s=t&gtf=2012

This website as all the stats I would need to calculate a score for each player. For example for the quarter back they have games played, completed passes, attempted passes, passing yards, and a lot

more data I can use to compare the players to one another. This website also has links to rushing yards by the running back, tail back, and any other positions that gained positive yards rushing. This website also has data on the salaries of each player. My program will print instructions on how to download each file that will be used and test whether the CSV file exist within the current working directory.

There are two files you will need to run this program the first is

http://www.playerfilter.com/nfl/passing-leaders/#/g,r;comp;att;pct;pyds,desc,r;ptd;ptdr;py_a;intc;intr;qbrat;tdintc;fpts;&s=t&gtf=2012

Just click convert to CSV and rename the file Quarterback_info.csv and put into the same folder as the python program

The next one is located

http://www.playerfilter.com/nfl/salary/#all-players/team,u,r;pos,r;g;salcap,desc,r,u;salt;salb;salsb;salob;&gtf=2009

Same instructions as above and I will also be uploading the CSV but this website is a great place to find data sets that might be a great how assignment for next quarter. Remember to name it playersalaries.csv

Methodology

I will use four calculations to convert the raw stats to a quarterback rating which I will graph to show a visual representation from their stats and their salaries. I will use a tuple from a player's name to their score and another tuple of their name to their salary.

Percentage of Completions — 324 of 461 is 70.28 percent. Subtract 30 from the completion percentage (40.28) and multiply the result by 0.05. The result is a point rating of 2.014.
Note: If the result is less than zero (Comp. Pct. less than 30.0), award zero points. If the results are greater than 2.375 (Comp. Pct. greater than 77.5), award 2.375.

Average Yards Gained Per Attempt — 3,969 yards divided by 461 attempts is 8.61. Subtract three yards from yards-per-attempt (5.61) and multiply the result by 0.25. The result is 1.403.
Note: If the result is less than zero (yards per attempt less than 3.0), award zero points. If the result is greater than 2.375 (yards per attempt greater than 12.5), award 2.375 points.

Percentage of Touchdown Passes — 35 touchdowns in 461 attempts is 7.59 percent. Multiply the touchdown percentage by 0.2. The result is 1.518.
Note: If the result is greater than 2.375 (touchdown percentage greater than 11.875), award 2.375.

Percentage of Interceptions — 10 interceptions in 461 attempts is 2.17 percent. Multiply the interception percentage by 0.25 (0.542) and subtract the number from 2.375. The result is 1.833.
Note: If the result is less than zero (interception percentage greater than 9.5), award zero points.

The sum of the four steps is (2.014 + 1.403 + 1.518 + 1.833) 6.768. The sum is then divided by six (1.128) and multiplied by 100. In this case, the result is 112.8. This same formula can be used to determine a passer rating for any player who attempts at least one pass.

After doing those computations I decided to look to another approach to see the correlation between the quarterbacks completion percentage and his yards per attempt. I took the CSE course staffs advice and tried to look for a linear combination between the two data set I just said. I did something like this with x being percent completion and y being average yards gained per attempt.

x
x + y
x + 2y
x + 3y
x + 4y
x + 5y
2x
2x + y
2x + 2y
2x + 3y
…
3x
3x + y
…

Results

Question 1: There isn't a correlation between the quarterback score and his salary that I could find and that is because 1. My data set was too small, I only looked at a single season and one season only has a maximum of 16 games. This is the problem the quarterback named Shaun Hill had a qb rating of 172 which is insanely high but he was being paid a faction of what the mainstream guys were being paid. (millions) that is because he is a backup and he only played one game. The one game he played in he did extremely well which is why his score is superb plus he could have been playing a not so talented team. I conclude that because a quarterback score isn't just on the QB it is on his the rest of his team to put him in a position to get a high score that there isn't really a correlation between a QB rating and his salary. I determined this by graphing the QB scores with his salaries as data points when you run my program.

After doing some research I found that a quarterback's salary is determined by his agent who negotiates his contract and how well he's played in the past or his overall resume of football. A new player like Robert Griffin III and Russell Wilson had higher scores than Drew Brees of the Saint and Tom Brady two famous and great quarterback but because they are rookies and fresh from the NFL draft they can't be paid on the level of season veterans. I determined this by printing out the 2012-2013 season of quarterback sorted by their score highest to lowest. This prints when I run the program. The most interesting thing that I discovered was that if a person has a really good agent and does well in the playoffs he will be paid higher if his team won than someone with a higher QB rating on a losing team. An example would be Joe Flacco who received the highest NFL contract in history at 120 million over 5 years. That means even if next season he drops the ball and goes out with a QB score of 70 he will still earn 20 million dollars before bonuses and sponsorships.

The next point of my research is trying to find a correlation between average yards per attempt and pass percentage. I used the CSE staff suggestion and my program graph the data with different coefficient. I

made a Microsoft excel documents with some of the data. I didn't put all the data on the excel document because after the first twenty or so test it seem clear that all the data was pointing to the same answer. That there isn't a correlation between that QB score system and his salary.

There can be many explanations but one I found was that at the level of skill all the QB of the NFL has are so high and every player is the best of the best that they all can score a high QB score over the season if the rest of his team is also skilled. And that rookies and free agents don't make as much as veterans do even if they do really well. That just means that that once the rookie's contracts up he will receive a pay grade from the team if they want to keep him. Russell Wilson only get 700,000 dollars a year during his first contract but once that one is up he will received a pay grade to match those of the veterans from other teams.

Reproducing your results

In the command shell run the python program

If that doesn't work open the python program and run it everything will pop up

1. Frist is the printed results of QB scores in the 2009 season
2. The last seasons QB and his score will print out next ranked from highest score to lowest
3. The first graph will be of the 2009 QB score VS his salary
4. The next barrage of graph will be of the 2009 season but each graph will be of different coefficents, the number of coefficient goes in this pattern
   a. Y will be 1 with x being incremented by .5 nine times
   b. Then Y will be incremented by .5 with X being reset to 1 and being incremented by .5 with the Y being constant

   This will create 72 graphs will varying coefficients

Collaboration

I worked by myself on this project and have no help on this assignment so far.

Reflection

I learned that there's more to football than meets the eye and if I want to know how the QB salaries are determined I have to talk to a football agent or a manager of a team. I wouldn't do anything differently and I would suggest finding a data set that interest you so coding won't be as boring.