CSc 496, Homework #5: Analysis of first down play calling in the NFL.

Due date: October 24th, 2024. **No late assignments will be accepted.**

In this assignment you will learn about whether teams run or pass on first down.

Football fans have an irrational love of running plays. A common criticism of coaches is that they pass too often, especially on first down. However, a lot of such criticism ignores the situation. Specifically, overall pass/run rates are biased towards the pass because must-pass situations occur more often than must-run situations. For example, on a final drive where the clock is critical, a team must often pass on every play.

In this assignment you will compute run-pass ratios for all teams for the 2023 NFL season, subject to game situations. Specifically, you will take as input (from the command line), in this order: `yardlineStart`, `yardlineEnd`, `timeRemaining`, `WPstart`, and `WPend`. The first three are integers, and the last two are floats that are between 0 and 1. (Assume the input is error-free.)

Then, for each team, you will count their run and pass plays for any play that is between `yardlineStart` and `yardlineEnd`, with more than `timeRemaining` seconds left in the game, and the win probability is between `WPstart` and `WPend`. In addition, for all of the plays, sum the expected probability added (EPA) for each team.

The data comes from the library `nfl_data_py`. Please see `https://github.com/nflverse/nfl_data_py` for details. There are a *lot* of fields in the Pandas data frame that is returned by the library call `nfl.import_pbp_data`; the fields you will need are `play_type`, `down`, `ydstogo`, `yardline_100`, `half_seconds_remaining`, `wp`, and `epa`. For more details on the fields, see this guide: `https://cran.r-project.org/web/packages/nflfastR/nflfastR.pdf`; it is the R version but the explanation of the fields is the same. Note that you will need to make sure that you do not include two-point attempts as plays.

Your output should be 32 lines, one per NFL team, and sorted by highest pass ratio. In addition, produce a scatterplot that has on the x-axis the EPA and on the y-axis the run ratio. Label the four quadrants as high/low and EPA/ratio.

Submit your python files on lectura using the `turnin` command; for this program, use the assignment name `csc496-f24-hw5`. Your program should be called `firstDownRunPass.py`.