Assignment 1

DSCI 222 – Data Science Workflow using Python
School of Mathematical and Data Sciences
West Virginia University

Instructions

This is an individual assignment — an opportunity to demonstrate your technical and soft skills. All material needed for the assignment can be found in Github Enjoy it!

- All deliverables must be in a folder created in your Google Drive account. The folder name should follow this format: `DS_X_LastName_FirstName`, where **X** represents the assignment number.

  Files included in your deliverables folder

  - Deliverable 1 Video Recording
  - Deliverable 2 Colab Notebook
  - Deliverable 2 Report
  - The two .csv files provided
  - Anything additional that contributed to the assignment

- Include your **full name** as a Python comment at the top of the notebook and at the top of your report.
- All data manipulation should occur within the Python script. No manipulation of the .csv file(s) should occur prior to importing the file(s) into your script. You may check your work by manually performing data analysis.
- Include your report in PDF format, written in LaTeX.
- Set share folder permissions so chood@mix.wvu.edu can access and run every file and notebook.
- Everything counts! Include as much as you want in your deliverables, even if the activity is not fully complete by the deadline. *Important:* Review the grading policy and course policies in the online syllabus.
- **Total: 100 points.**

Activity 1: Video Recording (50 points)

Task

Create a video tutorial on **"How to import NFL Projections from a .cvs file"**. No knowledge of football is needed for this activity, but if you have questions, feel free to ask. Here is a short video explaining how fantasy football works if you are interested.

Requirements

1. In the Assignment 1 folder in Github, you will find three scripts that import stat projections for NFL players in different ways. They are named `option_1.py`, `option_2.py`, and `option_3.py`
2. Choose one of the three options and explain how the code works. You will be using the option you select for Activity 2 as well.
3. You will also find `FLX_Projections.csv` and `QB_Projections.csv` files. You may use either or both. By default, the three options use QB projections.
4. Your video should clearly explain what each line of code is doing in its entirety. It is expected for every aspect of the script to be addressed and nothing should be ignored.
5. Your video must be a screen recording of the code as it is being explained, to help follow along. You may include yourself in the video as well, but you do not have to.

Deliverable 1:

Video recording

Activity 2: Python Source Code (50 points)

Scenario

You want to use NFL player projections to make data-driven decisions that give you a competitive edge in your fantasy football league. You have been able to import your data using the script from Activity 1 and now you are ready to do some initial analysis.

Projection Data

- `FLX_Projections.csv`
  - Contains season-long statistical projections for all FLX (Flex) players. This includes Running Backs (RB), Wide Receivers (WR), and Tight Ends (TE)
- `QB_Projections.csv`
  - Contains season-long statistical projections for Quarterbacks (QB)

Problems

Problem 1: Using list comprehension, create a list of all quarterbacks that are projected to throw for over 4000 yards.

Problem 2: What is the average touchdown to interception ratio for all quarterbacks projected to have at least 500 passing attempts?

Problem 3: Each rushing yard (RUSH_YDS) awards .1 fantasy points and each rushing touchdown (RUSH_TDS) awards 6 points. Determine which quarterback has the largest percentage of their projected fantasy points from these two statistics. Include in the output the player's name and the team he plays for.

Problem 4: Using list comprehension, create an output of all FLX players that play for the Pittsburgh (PIT) Steelers. Include the player's name, the position they play, and the projected fantasy points (FPTS).

Problem 5: Create three histograms of projected fantasy points (FPTS). You may use your judgment for how many bins to include. Use the following criterion to determine which players are included in each histogram.

- Histogram 1: Running backs that are projected to have at least 150 rushing attempts
- Histogram 2: Wide receivers that are projected to have at least 75 receptions
- Histogram 3: Tight ends that are projected to have at least 500 receiving yards

Deliverable 2:

Google Colab notebook

- Include all of your code for problems 1 through 5
- Clearly identify your answer to each problem
- Before each block of code, include text briefly explaining how the block of code works
- Any libraries, functions, methods, etc. used that are not explicitly discussed in class must clearly be explained

Deliverable 3:

You don't currently have a favorite NFL team to root for. Write a 2-3 paragraph report explaining how you could write a script that uses the data provided in this assignment to help find which team you should pick to cheer on. You do not need to be able to write the actual code, so be creative!