Main Files For Running Modeling Code

Note: Supporting functions to not need to be run to generate csv files to run modeling scripts. These csv files are included in the zip folder for ease of use.

Modeling

- Get the Modeling csv file for the specific season using Create_Modeling_CSV notebook (For ease of use these file have been generated and are is included in the zip folder)
- Change the file directory in Specifying year to run Modeling on and Files to Import block of code. There is a variable named year_to_run_modeling and this variable is only used for file naming and not for actually modeling date range specification
- All modeling is done for 2015 year. If you want to change the year go to models run under *Running Model Variations* block and change the date range to desired dates
- Be careful to change the directory location for the following blocks code titles"
    - Specifying year to run Modeling on and Files to Import
    - Function to plot each modeling case result on single plot
    - Function to plot each modeling case accuracy result on single plot
    - Function to plot all accuracy cases for all modeling result on one plot
    - Function to plot a specific accuracy case for all modeling result on single plot

Playing_A_Season

- Get the Modeling csv file for the specific season using Create_Modeling_CSV notebook (For ease of use these file have been generated and are is included in the zip folder)
- In the block of code titled *Import input file*, specify for the following fields:
    - year = Year to play the season for
    - file_path = where the modeling file is saved
    - Csv = remove to use correct direction of the slashes for Mac, Windows, or Linux
    - output_file_save_location = Output file location to save results in jpeg
    - num_of_parallel_games = Number of parallel games to be played
- In the block of code titled *Function to create a plot of model,* there exists a field called *file_save_full_path* for which you would need to also change direction of slashes based on operating system
- Now that you have updated all the required fields, go to *Cells* tab of jupyter notebook and click *Run All Cells*
- The output plots should be displayed at end of the notebook and saved at specified location

Supporting Files

Data_Scrapping

- There are 3 main websites that are scrapped: Savant, Baseball-Reference, and Swish Analytics
- For all these websites there is usually a get csv file, a code to loop over every day, and sometimes an helper function.
- In order to run this file you will need to go through each block and change any file paths as necessary.
- Also, remember to change the location to your local chrome driver

Birthday List

- Ensure data exists in the file structure outlined in the Jupyter Notebook
- Change path to correct directory
- Run full notebook

Build_Season_DataFrame

- Ensure data exists in the file structure outlined in the Jupyter Notebook
- Change path to correct directory for reading files from outermost directory
- Change path to correct directory for writing output files
- Define year desired to generate csv for
- Run full notebook

Create_Modeling_CSV

- Change path to correct directory for reading master data frame and ERA files
- Change path to correct directory for writing output files
- Define year desired to generate csv for
- Run full notebook

Data_Analysis

- Change path to correct directory for reading modeling csv files
- Define year desired to generate plots for
- Run full notebook

Player_File_Box_Office_Data

- Ensure data exists in the file structure outlined in the Jupyter Notebook
- Change path to correct directory for reading files from outermost directory
- Change path to correct directory for writing output files
- Define year desired to generate csv for
- Run full notebook

Savant_Player_CSV_File

- Ensure data exists in the file structure outlined in the Jupyter Notebook
- Change path to correct directory for reading files from outermost directory
- Change path to correct directory for writing output files

- Define year desired to generate csv for
- Run full notebook