

Instructions

- This homework assignment is worth 38 points.
- Please submit a **.ipynb** file to Blackboard.
- One submission per team.
- **Please strive for clarity and organization.**
- **Due Date: September 10, 2021 by 11:59 pm.**

Exercise 1

(5 points) What is your favorite sports? List and describe the common metrics that are recorded in your favorite sports.

Exercise 2

(5 points) Histograms are by far the most popular visualization tool of numerical variables. What are the key features that practitioners can learn from a numerical variable when they create a histogram?

Exercise 3

This is a list of every UFC fight in the history of the organization. Every row contains information about both fighters, fight details and the winner. Each row is a compilation of both fighter stats. Fighters are represented by “red” and “blue” (for red and blue corner). So for instance, red fighter has the compiled average stats of all the fights except the current one. The stats include damage done by the red fighter on the opponent and the damage done by the opponent on the fighter (represented by “opp” in the columns) in all the fights this particular red fighter has had, except this one as it has not occurred yet (in the data). Same information exists for blue fighter. The target variable is “Winner” which is the only column that tells you what happened. Here are some column definitions:

- R_ and B_ prefix signifies red and blue corner fighter stats respectively
- _opp_ containing columns is the average of damage done by the opponent on the fighter
- KD is number of knockdowns
- SIG_STR is no. of significant strikes “landed of attempted”
- SIG_STR_pct is significant strikes percentage
- TOTAL_STR is total strikes “landed of attempted”
- TD is no. of takedowns

- `TD_pct` is takedown percentages
- `SUB_ATT` is no. of submission attempts
- `PASS` is no. times the guard was passed?
- `REV` is the no. of Reversals landed
- `HEAD` is no. of significant strikes to the head “landed of attempted”
- `BODY` is no. of significant strikes to the body “landed of attempted”
- `CLINCH` is no. of significant strikes in the clinch “landed of attempted”
- `GROUND` is no. of significant strikes on the ground “landed of attempted”
- `win_by` is method of win
- `last_round` is last round of the fight (ex. if it was a KO in 1st, then this will be 1)
- `last_round_time` is when the fight ended in the last round
- `Format` is the format of the fight (3 rounds, 5 rounds etc.)
- `Referee` is the name of the Ref
- `date` is the date of the fight
- `location` is the location in which the event took place
- `Fight_type` is which weight class and whether it’s a title bout or not
- `Winner` is the winner of the fight
- `Stance` is the stance of the fighter (orthodox, southpaw, etc.)
- `Height_cms` is the height in centimeter
- `Reach_cms` is the reach of the fighter (arm span) in centimeter
- `Weight_lbs` is the weight of the fighter in pounds (lbs)
- `age` is the age of the fighter
- `title_bout` Boolean value of whether it is title fight or not
- `weight_class` is which weight class the fight is in (Bantamweight, heavyweight, Women’s flyweight, etc.)
- `no_of_rounds` is the number of rounds the fight was scheduled for
- `current_lose_streak` is the count of current concurrent losses of the fighter
- `current_win_streak` is the count of current concurrent wins of the fighter
- `draw` is the number of draws in the fighter’s ufc career

- `wins` is the number of wins in the fighter's ufc career
- `losses` is the number of losses in the fighter's ufc career
- `total_rounds_fought` is the average of total rounds fought by the fighter
- `total_time_fought(seconds)` is the count of total time spent fighting in seconds
- `total_title_bouts` is the total number of title bouts taken part in by the fighter
- `win_by_Decision_Majority` is the number of wins by majority judges decision in the fighter's ufc career
- `win_by_Decision_Split` is the number of wins by split judges decision in the fighter's ufc career
- `win_by_Decision_Unanimous` is the number of wins by unanimous judges decision in the fighter's ufc career
- `win_by_KO/TKO` is the number of wins by knockout in the fighter's ufc career
- `win_by_Submission` is the number of wins by submission in the fighter's ufc career
- `win_by_TKO_Doctor_Stoppage` is the number of wins by doctor stoppage in the fighter's ufc career

Consider the `UFC_data.csv` file. **In Python**, answer the following:

- (4 points) Using the pandas library, read the csv file and create a data-frame called `ufc`.
- (4 points) Using the appropriate commands, remove all the observations in which `weight_class` is equal to `OpenWeight`.
- (4 points) Using the appropriate commands, report the corner with highest winning rate (is it Red, Blue or draw?)
- (6 points) Using the appropriate commands, report the corner with highest winning rate (is it Red, Blue or draw?) across the different weight classes. Is this consistent with your answer from part (c)? Explain.
- (6 points) Using the appropriate commands, report the corner with highest winning rate (is it Red, Blue or draw?) across the different weight classes and whether or not the fight is a title bout. Is this consistent with your answers from parts (c)-(d)? Explain.
- (4 points) Using your answers from parts (c)-(e), what can you conclude from that analysis? Explain.