Name: Solutions

SA433 · Data Wrangling and Visualization Fall 2024 · Uhan

Exam 2 – Part 1 – 11/6/2024

Problem	Max	Score
1	5	
2	5	
3	5	
4	5	
Total		/ 20

Instructions

- This part is worth 20 points total. The exam (all three parts) is worth 100 points total.
- You have 50 minutes to complete Parts 1 and 2 of the exam.
- For Parts 1 and 2 of the exam, you may not use any outside assistance. These parts of the exam are closed book, closed notes, and closed internet.
- **No collaboration allowed.** All work must be your own.
- You must turn in Part 1 before beginning Part 2.
- **Do not discuss the contents of this exam with any midshipmen until it is returned to you.**

Background

This part is based on the DataFrame below, called `track_df`, which provides the time (in seconds) for six different athletes in two events (400 meter and 800 meter) over two track meets: "meet1" and "meet2":

	athlete	team	meet1_400	meet2_400	meet1_800	meet2_800
0	Fazio	Blue	60.7	59.1	135.6	132.0
1	Martell	Gray	62.2	62.6	139.5	133.2
2	Ohlsson	Blue	56.7	54.7	132.9	133.2
3	Tovcach	Blue	63.0	63.2	121.9	122.8
4	OConnell	Gray	61.6	58.5	135.0	133.4
5	Acosta	Gray	59.5	60.4	125.1	123.7

Assume the following code has already been run to import Pandas and read in the DataFrame:

```
import pandas as pd

track_df = pd.read_csv('data/track.csv')
```

The four problems below consist of writing lines of code in a single method chain that begins with `track_df` and creates a new DataFrame, `blue_meet1_df`, which focuses on the Blue team's performance in the 400 meter race in the first meet. The new DataFrame is displayed below the code for reference.

1. Filter the rows for athletes on the Blue team.
2. Change the name of column `meet1_400` to `400m`.
3. Sort the rows of the DataFrame according to the times in the `400m` column so that the fastest time is first.
4. Pare down the columns so that only the three columns shown below remain (there are 2 ways to do this–either way is fine).

```
blue_meet1_df = (
    track_df
    .query('team == "Blue"')                    # Problem 1
    .rename(columns = {'meet1_400': '400m'})    # Problem 2
    .sort_values('400m')                        # Problem 3
    [['athlete', 'team', '400m']]               # Problem 4
)
```

or

```
blue_meet1_df = (
    track_df
    .query('team == "Blue"')                                  # Problem 1
    .rename(columns = {'meet1_400': '400m'})                  # Problem 2
    .sort_values('400m')                                      # Problem 3
    .drop(columns = ['meet1_800', 'meet2_400', 'meet2_800'])  # Problem 4
)
```