Powerlifting

March 28, 2023

With this short jupyterlabs project, I'll look through some weightlifting data and identify some of the maximums lifted then connect these max's to the lifter and the meet it was achieved at.

Since the data for this is spread over two csv files, I'll use sql (convert csv to pandas dataframe briefly then convert to sql file) in order to connect data from the two files.

[1]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import sqlite3
```

[2]:
```
%load_ext sql
```

[3]:
```
import os
os.getcwd()
```

[3]:
```
'C:\\Users\\jflieder\\Desktop\\Code\\Data Science Portfolio\\Data Science
projects'
```

[4]:
```
os.chdir(r"C:\Users\jflieder\Desktop\Code\Data Science Portfolio\Data Science␣
↪projects") #make sure I'm working in the same directory as the target dataset
os.getcwd()
```

[4]:
```
'C:\\Users\\jflieder\\Desktop\\Code\\Data Science Portfolio\\Data Science
projects'
```

[5]:
```
#create a sqlite database:
conn = sqlite3.connect("lifter_meets.db")
```

[6]:
```
#Connect this jupyterlab to the sql database:
%sql sqlite:///lifter_meets.db
```

[7]:
```
#Put the datasets in csv format into dataframe format:
meets = pd.read_csv("meets.csv")
lifters = pd.read_csv("openpowerlifting.csv")
```

[8]:
```
#Take a look at the meets dataframe:
meets.head()
```

[8]:
```
   MeetID        MeetPath Federation        Date MeetCountry MeetState  \
0       0  365strong/1601  365Strong  2016-10-29         USA        NC
1       1  365strong/1602  365Strong  2016-11-19         USA        MO
2       2  365strong/1603  365Strong  2016-07-09         USA        NC
3       3  365strong/1604  365Strong  2016-06-11         USA        SC
4       4  365strong/1605  365Strong  2016-04-10         USA        SC

    MeetTown                                              MeetName
0  Charlotte  2016 Junior & Senior National Powerlifting Cha…
1      Ozark                  Thanksgiving Powerlifting Classic
2  Charlotte                             Charlotte Europa Games
3  Rock Hill                  Carolina Cup Push Pull Challenge
4  Rock Hill                               Eastern USA Challenge
```

[9]:
```
#Take a look at the lifters dataframe:
lifters.head()
```

[9]:
```
   MeetID             Name Sex   Equipment   Age     Division  BodyweightKg  \
0       0  Angie Belk Terry   F       Wraps  47.0    Mst 45-49         59.60
1       0       Dawn Bogart   F  Single-ply  42.0    Mst 40-44         58.51
2       0       Dawn Bogart   F  Single-ply  42.0  Open Senior         58.51
3       0       Dawn Bogart   F         Raw  42.0  Open Senior         58.51
4       0      Destiny Dula   F         Raw  18.0   Teen 18-19         63.68

  WeightClassKg  Squat4Kg  BestSquatKg  Bench4Kg  BestBenchKg  Deadlift4Kg  \
0            60       NaN        47.63       NaN        20.41          NaN
1            60       NaN       142.88       NaN        95.25          NaN
2            60       NaN       142.88       NaN        95.25          NaN
3            60       NaN          NaN       NaN        95.25          NaN
4          67.5       NaN          NaN       NaN        31.75          NaN

   BestDeadliftKg  TotalKg Place   Wilks
0           70.31   138.35     1  155.05
1          163.29   401.42     1  456.38
2          163.29   401.42     1  456.38
3             NaN    95.25     1  108.29
4           90.72   122.47     1  130.47
```

[10]:
```
#I left out the keyword entry of method='multi' to prevent a 'too many␣
 ↪variables' type of error:
meets.to_sql("meets", conn, if_exists='replace', index=False)
```

[10]:
```
8482
```

[11]:
```
lifters.to_sql("lifters", conn, if_exists='replace', index=False)
```

[11]:
```
386414
```

[12]:
```
#take a ppek at the meets data:
%sql select * From meets Limit 5
```

```
 * sqlite:///lifter_meets.db
Done.
```

[12]:
```
[(0, '365strong/1601', '365Strong', '2016-10-29', 'USA', 'NC', 'Charlotte',
'2016 Junior & Senior National Powerlifting Championships'),
 (1, '365strong/1602', '365Strong', '2016-11-19', 'USA', 'MO', 'Ozark',
'Thanksgiving Powerlifting Classic'),
 (2, '365strong/1603', '365Strong', '2016-07-09', 'USA', 'NC', 'Charlotte',
'Charlotte Europa Games'),
 (3, '365strong/1604', '365Strong', '2016-06-11', 'USA', 'SC', 'Rock Hill',
'Carolina Cup Push Pull Challenge'),
 (4, '365strong/1605', '365Strong', '2016-04-10', 'USA', 'SC', 'Rock Hill',
'Eastern USA Challenge')]
```

[13]:
```
#take a ppek at the lifters data:
%sql select * From lifters Limit 5
```

```
 * sqlite:///lifter_meets.db
Done.
```

[13]:
```
[(0, 'Angie Belk Terry', 'F', 'Wraps', 47.0, 'Mst 45-49', 59.6, '60', None,
47.63, None, 20.41, None, 70.31, 138.35, '1', 155.05),
 (0, 'Dawn Bogart', 'F', 'Single-ply', 42.0, 'Mst 40-44', 58.51, '60', None,
142.88, None, 95.25, None, 163.29, 401.42, '1', 456.38),
 (0, 'Dawn Bogart', 'F', 'Single-ply', 42.0, 'Open Senior', 58.51, '60', None,
142.88, None, 95.25, None, 163.29, 401.42, '1', 456.38),
 (0, 'Dawn Bogart', 'F', 'Raw', 42.0, 'Open Senior', 58.51, '60', None, None,
None, 95.25, None, None, 95.25, '1', 108.29),
 (0, 'Destiny Dula', 'F', 'Raw', 18.0, 'Teen 18-19', 63.68, '67.5', None, None,
None, 31.75, None, 90.72, 122.47, '1', 130.47)]
```

Here are the divisions with the most entries of competitors:

[14]:
```
%sql select Division, count(*) as Entrant_Amount from lifters group by Division␣
↪order by count(8) desc limit 20
```

```
 * sqlite:///lifter_meets.db
Done.
```

[14]:
```
[('Open', 68618),
 ('Boys', 59641),
 ('R-O', 28667),
 (None, 15843),
 ('Amateur Open', 9396),
 ('R-JR', 7849),
 ('Open Men', 7487),
```

```
 ('Junior', 7391),
 ('Junior 19-23', 6695),
 ('Junior 20-23', 6255),
 ('Teen 14-18', 5348),
 ('O', 4963),
 ('Master 40-49', 4899),
 ('Pro Open', 4767),
 ('Juniors', 4437),
 ('Master 1', 3231),
 ('Master 50-59', 2740),
 ('R-T3', 2397),
 ('Sub-Junior', 2289),
 ('Master 40-44', 2248)]
```

Next, I'll find the highest meet total by a lifter.

[15]:
```
%sql select max(TotalKg) from lifters
```

```
 * sqlite:///lifter_meets.db
Done.
```

[15]:
```
[(1365.31,)]
```

Now I'll dig into the available information about the occurrence of that highest total.

[16]:
```
#Get the information on the highest total (lifter name, meetid, etc.)
%sql select * from lifters where TotalKg = ( select max(TotalKg) from lifters )
```

```
 * sqlite:///lifter_meets.db
Done.
```

[16]:
```
[(1412, 'Dave Hoff', 'M', 'Multi-ply', None, 'Pro Open', 131.09, '140', None,
560.19, None, 437.72, None, 367.41, 1365.31, '1', 771.07)]
```

Now I'll find the highest Wilks achieved by a lifter then dig into the available information about the occurrence of that highest Wilks.

[17]:
```
#Find the highest meet Wilks by a lifter
%sql select max(Wilks) from lifters
```

```
 * sqlite:///lifter_meets.db
Done.
```

[17]:
```
[(779.38,)]
```

[18]:
```
#Get the information on the highest Wilks (lifter name, meetid, etc.)
%sql select * from lifters where Wilks = ( select max(Wilks) from lifters )
```

```
 * sqlite:///lifter_meets.db
Done.
```

[18]:
```
[(1329, 'Dave Hoff', 'M', 'Multi-ply', 25.0, 'Heavyweight', 122.97, '125', None,
548.85, None, 442.25, None, 371.95, 1363.05, '1', 779.38)]
```

Interesting! The highest total and the highest Wilks were achieved by the ostensibly the same lifter (both with multiply equipment) at different weight classes.

Now I'll display some of the data of the lifter on those respective days alongside the meet's name and date. I'll use %%sql instead of %sql to spread the query over multiple lines. The easiest way I found to limit the total to that of the highest in the lifters data file is to create a custom table while inner joining to it on that criteria.

[19]:
```
%%sql select lifters.Name, lifters.Age, lifters.Division, lifters.BodyweightKg,␣
 ↪lifters.WeightClassKg, lifters.TotalKg, meets.MeetName, meets.Date
from lifters
inner join meets on meets.MeetID = lifters.MeetID
inner join (select max(TotalKg) total from lifters) custom on custom.total =␣
 ↪lifters.TotalKg
```

```
 * sqlite:///lifter_meets.db
Done.
```

[19]:
```
[('Dave Hoff', None, 'Pro Open', 131.09, '140', 1365.31, 'Champion of Champions
2', '2017-10-07')]
```

[20]:
```
%%sql select lifters.Name, lifters.Age, lifters.Division, lifters.BodyweightKg,␣
 ↪lifters.WeightClassKg, lifters.Wilks, meets.MeetName, meets.Date
from lifters
inner join meets on meets.MeetID = lifters.MeetID
inner join (select max(Wilks) total from lifters) custom on custom.total =␣
 ↪lifters.Wilks
```

```
 * sqlite:///lifter_meets.db
Done.
```

[20]:
```
[('Dave Hoff', 25.0, 'Heavyweight', 122.97, '125', 779.38, '2013 IPA Westside
Invitational Pro Open Powerlifting Championships', '2013-08-17')]
```

Next, I'll find the same data but out of the female lifters.

[21]:
```
%sql select max(TotalKg) from lifters where Sex = 'F'
```

```
 * sqlite:///lifter_meets.db
Done.
```

[21]:
```
[(816.47,)]
```

[22]:
```
#I'll need to first filter for Sex = 'F' in the first where statement to get␣
 ↪the highest total that a female lifted, then again at the end to filter out␣
 ↪men that lifted that same amount.
```

```
%sql select * from lifters where TotalKg = ( select max(TotalKg) from lifters␣
↪where Sex = 'F' ) and Sex = 'F'
```

```
 * sqlite:///lifter_meets.db
Done.
```

[22]:
```
[(5479, 'Laura Phelps-Sweatt', 'F', 'Multi-ply', None, 'Pro', 75.0, '75', None,
337.93, None, 240.4, None, 238.14, 816.47, '1', 776.17)]
```

[23]:
```
%sql select max(Wilks) from lifters where Sex = 'F'
```

```
 * sqlite:///lifter_meets.db
Done.
```

[23]:
```
[(776.17,)]
```

[24]:
```
%sql select * from lifters where Wilks = ( select max(Wilks) from lifters where␣
↪Sex = 'F' ) and Sex = 'F'
```

```
 * sqlite:///lifter_meets.db
Done.
```

[24]:
```
[(5479, 'Laura Phelps-Sweatt', 'F', 'Multi-ply', None, 'Pro', 75.0, '75', None,
337.93, None, 240.4, None, 238.14, 816.47, '1', 776.17)]
```

So the highest female total and the highest female Wilks were achieved by the same lifter but, unlike those for the male side, they were achieved on the same meet! Here is the merging of the data for this performance, as I did previously for the male total and wilks maximums. In attempting to filter for only female lifters (since the highest overall was performed by a male), I found an easier way to find this that does not involve an inner join:

[25]:
```
%%sql select lifters.Sex, lifters.Name, lifters.Age, lifters.Division, lifters.
↪BodyweightKg, lifters.WeightClassKg, lifters.TotalKg, meets.MeetName, meets.
↪Date
from lifters, meets
where lifters.TotalKg = ( select max(TotalKg) from lifters where Sex = 'F' )␣
↪and lifters.Sex = 'F' and lifters.MeetID = meets.MeetID
```

```
 * sqlite:///lifter_meets.db
Done.
```

[25]:
```
[('F', 'Laura Phelps-Sweatt', None, 'Pro', 75.0, '75', 816.47, '2011
PowerStation Pro/Am', '2011-08-20')]
```

The same could be done for each of the individual lifts. The data could also be isolated for different meets or divisions. This is just a general exploration of the data.