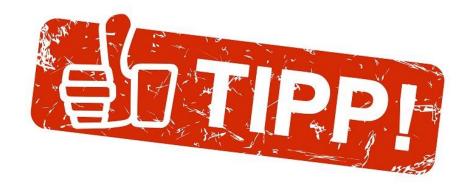


CASE STUDY 045 [Python] Proving the Birthday Paradox with 2016 Olympics Athletes





Here are some clues in case you are stuck with the case study:

- 1. For this exercise we just need to use Pandas.
- 2. To convert the column dob in a date format, use the function pd.to_datetime, passing besides the dob column, format='%m/%d/%y' and errors='ignore'. Assign this to column called birthday.
- 3. The values that couldn't be converted will have a null value due to the parameter errors='ignore'. So you have to filter nulls, like:

```
athletes = athletes[~athletes['birthday'].isnull()]
```

4. To extract a sample from the dataframe, use the sample() method, like: athletes.sample(sample size)

5. After sampling, you have to reset_index() of the sample dataframe. You can do both steps at once, returning or assigning the result to a variable:

athletes.sample(sample size).reset index()