Hypothesis Testing with Men's and Women's Soccer Matches

Goal: want to determine if the mean number of goals scored in women's international soccer matches is greater than men's using statistical tests

Technologies: Python, Pandas, Matplotlib.pyplot, pingouin

Description:

1. Read in the data through read_csv() for both women and men
2. Converted 'date' column into proper datetime datatype using to_datetime()
3. Filtered dataframe based on certain date and tournament by using subsetting techniques
4. Determined 'total_score' from sum of 'home_score' and 'away_score'
5. Visualized distribution of 'total_score' for both men and women and obtained non-normalized distribution, which indicated a non-parametric test should be implemented: Wilcoxon-Mann_Whitney
6. Used pandas.concat() to stack men and women's data and pivoted data from long to wide to feed into statistical test
7. Performed pingouin.mwu() to obtain p-values
8. Determined significance based on alpha level and delivered results