Ex2 - Filtering and Sorting Data This time we are going to pull data directly from the internet. Step 1. Import the necessary libraries In [1]: import pandas as pd Step 2. Import the dataset from this address. Step 3. Assign it to a variable called euro12. In [2]: euro12 = pd.read_csv('https://raw.githubusercontent.com/guipsamora/pandas_exercises/master/02_Filtering_%26_Sorting/Euro12/Euro_2012_stats_TEAM.csv', sep=',') euro12 % **Penalties Shots Shots** Total shots Saves-to-Out[2]: Hit Penalty Shooting Saves Fouls Yellow Red Subs Subs Players Fouls Offsides Team Goals Goalsoff on (inc. not shots Woodwork Won Conceded Cards Cards Accuracy goals made on off Used target target to-shots Blocked) scored ratio 2 0 51.9% 16.0% 32 0 0 0 13 41 62 9 0 9 9 16 Croatia 4 13 12 81.3% Czech 1 13 18 41.9% 12.9% 39 0 0 ... 60.1% 53 73 8 11 11 19 0 ... 2 4 10 10 50.0% 20.0% 27 1 0 10 66.7% 25 38 8 4 0 7 7 15 Denmark 3 5 11 18 50.0% 17.2% 40 0 0 0 ... 22 88.1% 43 45 6 5 11 11 16 England 0 3 22 37.9% 6.5% 65 0 0 ... 6 54.6% 36 51 5 6 0 11 11 19 4 France 24 1 2 0 ... 49 10 32 32 47.8% 15.6% 80 10 62.6% 63 12 4 0 15 15 17 Germany 1 ... 6 5 8 18 30.7% 19.2% 32 1 1 13 65.1% 67 48 12 9 1 12 12 20 Greece 2 0 ... 7 Italy 6 34 45 43.0% 7.5% 110 20 74.1% 101 89 16 16 0 18 18 19 8 Netherlands 2 12 36 25.0% 4.1% 60 2 0 0 ... 12 70.6% 35 30 3 5 0 7 7 15 2 0 0 0 ... 48 56 7 7 7 9 15 23 39.4% 5.2% 48 6 66.7% 3 17 Poland 0 ... 10 6 22 42 34.3% 9.3% 82 6 0 10 73 90 10 12 14 14 Portugal 71.5% 0 16 Republic of 11 1 7 12 36.8% 5.2% 28 0 0 0 ... 65.4% 43 51 6 10 10 17 11 1 17 5 31 12.5% 59 2 0 0 ... 43 6 0 7 7 12 Russia 9 22.5% 10 77.0% 34 4 16 13 12 42 33 16.0% 100 0 0 ... 15 102 83 19 11 17 17 Spain 55.9% 93.8% 0 18 7 14 5 17 19 47.2% 13.8% 39 3 0 0 ... 8 35 51 7 0 9 9 18 Sweden 61.6% 15 21.2% 6.0% 0 ... 48 9 18 Ukraine 26 0 13 76.5% 16 rows × 35 columns Step 4. Select only the Goal column. In [3] euro12["Goals"] Out[3]: 0 4 3 5 3 10 5 6 8 2 10 11 12 13 12 14 15 Name: Goals, dtype: int64 Step 5. How many team participated in the Euro2012? In [4]: euro12["Team"].nunique() Out[4]: 16 Step 6. What is the number of columns in the dataset? In [5]: euro12.shape Out[5]: (16, 35) Step 7. View only the columns Team, Yellow Cards and Red Cards and assign them to a dataframe called discipline In [6]: discipline = euro12[["Team", "Yellow Cards", "Red Cards"]] discipline Team Yellow Cards Red Cards Out[6]: 0 9 0 Croatia 1 Czech Republic 0 2 Denmark 3 England 4 6 France Germany 6 9 Greece 1 16 8 Netherlands 5 0 9 Poland 10 12 Portugal 11 Republic of Ireland 12 Russia 6 13 11 Spain 14 7 Sweden 15 0 Ukraine Step 8. Sort the teams by Red Cards, then to Yellow Cards In [7]: discipline.sort_values(by=["Red Cards", "Yellow Cards"], ascending=False) Out[7]: Team Yellow Cards Red Cards 9 6 Greece 1 Poland 11 Republic of Ireland 6 1 7 Italy 16 10 0 Portugal 12 13 11 Spain 9 0 Croatia 1 Czech Republic 0 14 Sweden France 12 6 0 Russia England 8 5 0 Netherlands 15 Ukraine 2 4 0 Denmark Germany Step 9. Calculate the mean Yellow Cards given per Team In [8]: discipline.groupby("Team")["Yellow Cards"].mean() Team Out[8]: Croatia Czech Republic Denmark England France Germany 9 Greece Italy 16 Netherlands 5 Poland Portugal 12 Republic of Ireland Russia Spain 11 Sweden 7 5 Ukraine Name: Yellow Cards, dtype: int64 Step 10. Filter teams that scored more than 6 goals In [9]: euro12[euro12["Goals"]>6] Out[9]: Shots Shots % **Total shots** Saves-to-Shooting Hit Penalty Penalties Saves Fouls Red Subs Subs **Players** Fouls Yellow Goals-Offsides Team Goals on off shots Woodwork made Won Conceded Cards off Used Accuracy goals not scored Cards on Blocked) target to-shots target ratio **5** Germany 10 32 32 47.8% 15.6% 80 2 0 ... 10 62.6% 49 12 4 0 15 15 17 0 ... 12 42 33 55.9% 16.0% 100 0 15 93.8% 102 83 19 11 17 17 18 13 Spain 2 rows × 35 columns Step 11. Select the teams that start with G In [10]: euro12[euro12["Team"].str.startswith("G")] Total shots Out[10]: Shots Shots % Saves-to-Shooting Saves Players Hit Penalty Penalties Fouls Fouls Yellow Red Subs Subs Offsides Team Goals Goalsshots off (inc. on Accuracy Woodwork goals not scored made Won Conceded Cards Cards on off Used target target to-shots Blocked) ratio 17 Germany 32 32 47.8% 15.6% 80 2 62.6% 63 12 15 15 10 0 10 49 0 67 12 12 Greece 18 30.7% 19.2% 32 13 65.1% 48 12 20 2 rows × 35 columns Step 12. Select the first 7 columns In [11]: euro12.columns[:7] Out[11]: Index(['Team', 'Goals', 'Shots on target', 'Shots off target', 'Shooting Accuracy', '% Goals-to-shots', 'Total shots (inc. Blocked)'], dtype='object') In [12]: euro12.iloc[:, :7] Team Goals Shots on target Shots off target Shooting Accuracy % Goals-to-shots Total shots (inc. Blocked) Out[12]: 0 32 Croatia 4 13 12 51.9% 16.0% 1 Czech Republic 13 18 41.9% 12.9% 39 50.0% 2 4 10 10 20.0% 27 Denmark 3 5 50.0% 17.2% England 11 18 40 3 22 4 France 24 37.9% 6.5% 65 32 5 10 32 80 Germany 47.8% 15.6% 6 5 8 18 19.2% 32 Greece 30.7% 7 6 34 45 Italy 43.0% 7.5% 110 8 2 12 36 Netherlands 25.0% 4.1% 60 9 2 15 23 Poland 39.4% 5.2% 48 10 6 22 42 9.3% 82 Portugal 34.3% 11 Republic of Ireland 7 12 36.8% 5.2% 28 12 Russia 5 9 31 22.5% 12.5% 59 13 12 42 33 100 Spain 55.9% 16.0% 5 17 19 39 14 Sweden 47.2% 13.8% 15 26 21.2% 38 Ukraine 6.0% Step 13. Select all columns except the last 3. In [13]: euro12.iloc[:, :-3] Out[13]: Saves-Shots Shots Total **Penalties** Shooting Goals-Hit Penalty to- Fouls Clean Goals Saves Fouls Yellow Red Team Goals shots (inc. Blocks off Offsides not on Woodwork **Sheets** conceded Won Conceded Cards Cards **Accuracy** togoals made shots target target Blocked) scored shots ratio 0 ... 0 Croatia 4 13 12 51.9% 16.0% 32 0 0 0 10 3 13 81.3% 41 62 2 9 0 Czech 4 13 41.9% 12.9% 39 0 0 ... 1 10 60.1% 53 8 0 1 18 0 6 9 73 Republic 0 2 Denmark 4 10 10 50.0% 20.0% 27 1 0 ... 1 10 5 10 66.7% 25 38 8 4 0 0 ... 3 5 11 18 50.0% 17.2% 40 0 0 2 29 3 22 43 45 6 0 England 88.1% 5 7 3 22 0 5 5 6 4 France 24 37.9% 6.5% 65 0 ... 1 6 54.6% 36 51 0 0 ... 10 32 32 47.8% 80 11 6 62.6% 63 49 12 0 Germany 15.6% 1 1 10 4 6 Greece 5 8 18 30.7% 19.2% 32 1 1 1 ... 1 23 7 13 65.1% 67 48 12 9 1 7 34 45 43.0% 110 0 0 ... 2 18 20 101 0 6 7.5% 74.1% 89 16 16 Italy 2 0 0 9 5 3 5 8 Netherlands 12 36 25.0% 4.1% 60 0 ... 12 70.6% 35 0 0 ... 9 15 23 48 0 0 0 8 3 66.7% 48 3 Poland 39.4% 5.2% 56 6 1 2 0 10 Portugal 22 42 34.3% 9.3% 82 6 0 0 ... 11 4 10 71.5% 73 90 10 12 Republic of 11 12 5.2% 28 0 0 ... 0 23 9 17 65.4% 43 11 1 7 36.8% 51 6 1 Ireland 0 ... 12 5 9 31 22.5% 12.5% 59 2 0 0 8 3 10 77.0% 34 43 4 6 0 Russia 0 5 8 19 0 13 Spain 42 33 55.9% 16.0% 100 0 ... 15 93.8% 102 83 11 0 ... 12 5 17 19 47.2% 13.8% 39 3 0 5 61.6% 35 51 7 7 0 14 1 8 Sweden 7 0 0 0 4 15 Ukraine 26 21.2% 6.0% 38 0 4 13 76.5% 48 31 16 rows × 32 columns Step 14. Present only the Shooting Accuracy from England, Italy and Russia euro12.loc[euro12.Team.isin(['England', 'Italy', 'Russia']), ['Team', 'Shooting Accuracy']] **Team Shooting Accuracy** Out[14]: 3 England 50.0% Italy 43.0% Russia 22.5% In []: