BT3017 - Tutorial 2

Data Analysis and Visualization using Pandas and Seaborn

A0248812B _ Huang Ping Mao

(a) What are the top three favourite sports of yours?

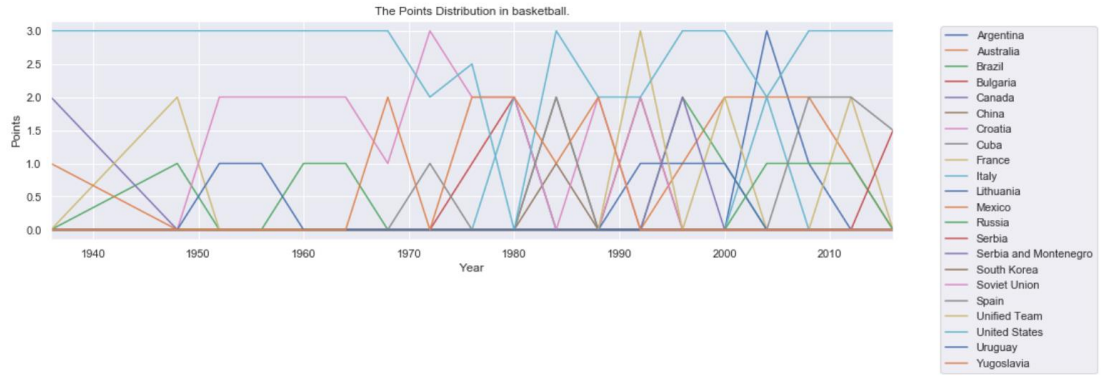
My top three favourite sports are the baseball, the basketball, and the table tennis.

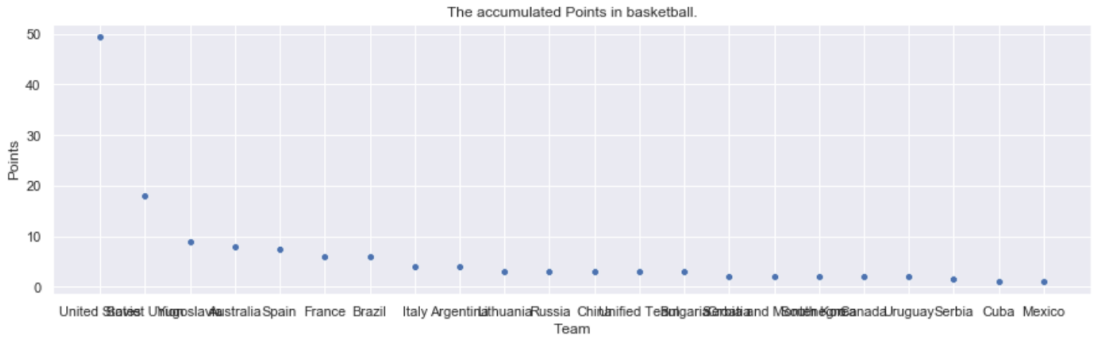
(b) What are the top three questions you want to ask regarding the data related to at least one of these three sports?

(i) Which countries get the top three points in basketball for the whole time if gold medal for three points, silver medal for two points, and bronze medal for one point?
(ii) Which country gets the longest winning streak in baseball? That is, Which country wins the longest in a row with the top three in baseball?
(ii) What is the probability to get other medals when winning one medal?

(c) Using Pandas and Seaborn, analyze and visualize the data to help answer your three questions raised in Step (b) above.

(i) Which countries get the top three points in basketball for the whole time if gold medal for three points, silver medal for two points, and bronze medal for one point?

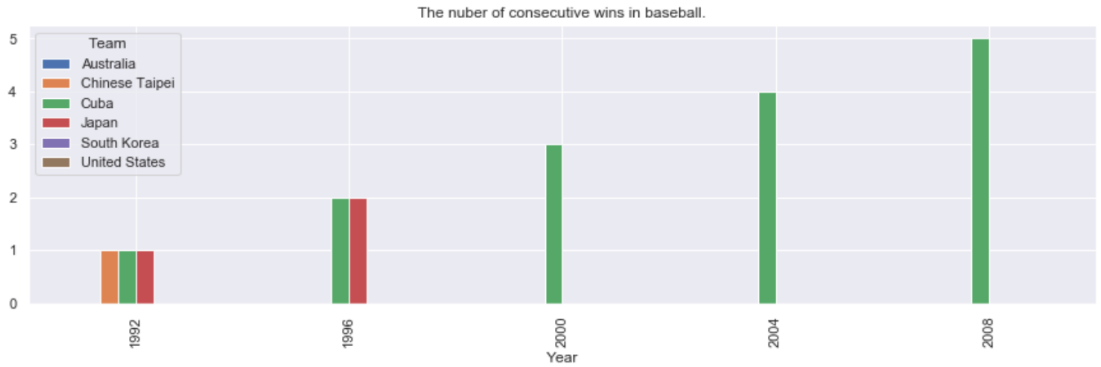




1	United States	49.5
2	Soviet Union	18.0
3	Yugoslavia	9.0

There are United States, Soviet Union, Yugoslavia with regard to 49.5, 18, and 9 points in basketball.
Note: There are 0.5 points appear because the games are split into playing with males and females.

(ii) Which country gets the longest winning streak in baseball? That is, Which country wins the longest in a row with the top three in baseball?



Cuba wins the longest in a row with the top three in baseball.

(iii) What is the probability to get other medals when winning one medal?
No country win all three games at the same time.
There are 4 countries win the baseball and the basketball games at the same time.

Year	Team
1996	United States
2000	United States
2004	Australia
2008	United States

There is 1 country wins the baseball and the table tennis games at the same time.

Year	Team
2008	South Korea

There are 2 countries win the basketball and the table tennis games at the same time.

Year	Team
1988	Yugoslavia
1992	China

When the country wins a medal in Baseball.
It is about 26.67 percent chance to earn another medal in Basketball.
It is about 6.67 percent chance to earn another medal in Table Tennis.
It is about 33.33 percent chance to earn another medal in Basketball or Table Tennis.

When the country wins a medal in Basketball.
It is about 5.48 percent chance to earn another medal in Baseball.
It is about 2.74 percent chance to earn another medal in Table Tennis.
It is about 8.22 percent chance to earn another medal in Baseball or Table Tennis.

When the country wins a medal in Table Tennis.
It is about 1.92 percent chance to earn another medal in Baseball.
It is about 3.85 percent chance to earn another medal in Basketball.
It is about 5.77 percent chance to earn another medal in Baseball or

Basketball.

It is the highest probability to earn another medal When winning a medal in Baseball among three sports games.

Reference:

drop Nan with specific columns:
https://stackoverflow.com/questions/13413590/how-to-drop-rows-of-pandas-dataframe-whose-value-in-a-certain-column-is-nan
fillna:
https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.fillna.html
sort_values:
https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.sort_values.html
Group by Sum as new column name:
https://stackoverflow.com/questions/45124992/group-by-sum-as-new-column-name
selecting top n rows:
https://stackoverflow.com/questions/41825978/sorting-columns-and-selecting-top-n-rows-in-each-group-pandas-dataframe
seaborn figure size: https://www.delftstack.com/howto/seaborn/size-of-seaborn-plot/
reset index:
https://pandas.pydata.org/docs/reference/api/pandas.Series.reset_index.html
change index: https://stackoverflow.com/questions/32249960/in-python-pandas-start-row-index-from-1-instead-of-zero-without-creating-additi

Place Legend Outside the Plot in Matplotlib:
https://www.delftstack.com/howto/matplotlib/how-to-place-legend-outside-of-the-plot-in-matplotlib/
plot:
https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.plot.html