Homework 1

1 Question 1

- Use Python/Pandas to analyze the data **"pitching.csv"**. The meaning of each column can be found from therespective readme file.

- Submit a ipython notebook file that answers the following questions with python commands. In the ipython notebook file that you submit, also include the original question as a comment before the code for each answer. (40 pts)

  - Read the data into a data frame.

  - Display all pitchers in 2015 under each team sorted by their last names.

  - Display all pitchers in 2015 sorted by their ERAs. This allows you to identify the best and worst pitchers.

  - Display all pitchers in 2015 under each team sorted by their ERAs.

  - Display all pitchers in 2015 under each team sorted by their strike outs (SO).

  - Calculate the average ERA for each team, list the teams by the average ERA in a descending order.

  - Calculate the total number of home runs allowed (HRs) for each team, and list the team by the count in an ascending order.

  - Calculate each team's average ERA from 2011 to 2015, and list the result as a table, where each column is a year, each row is a team.

  - Calculate the average ERAs for American League (AL) and National League (NL) from 2011 to 2015, and list the result as a table, where each column is a league and each row is a year.

  - List the pitchers who had the most wins in each of the years from 2011 to 2015.

2 Question 2

- Use Python and related packages (Pandas, Matplotlib, etc) to analyze the provided MLB baseball datasets **"players.csv", "pitching.csv" and "teams.csv"**, compute relevant statistics, and create visualizations. Themeaning of each column can be found from the respective readme file.

- Submit a ipython notebook file that answers the following questions with python commands. In the ipython notebook file that you submit, also include the original question as a comment before the code for each answer.

- Your analysis and visualization results will provide evidence for you to answer the following questions.

  - In the year of 2011-2015, Is a team's win-loss record related to its payrolls?

  - In the year of 2011-2015, Is a player's batting performance related to his team's win-loss record?

  - In the year of 2011-2015, is a team's win-loss record related to its pitching performance?

- To answer the questions above, you are asked to create the following tables and visualizations:

  - Task 1 (20 pts)

    1. Compute the total number of wins for each of the teams in MLB over 2011-2015, sort them in a descending order.

    2. Compute the average payroll per year for all teams over 2011-2015, sort them in a descending order.

    3. Create a visualization of your choice which will all allow you to show whether a team's winning record is related to its payroll. So is a team's winning record related to its payroll?

- Task 2 (20 pts)

    1. Compute the Batting Averages for each of the MLB teams over 2011-2015, sort them in a descendin-gorder. The Batting Average is defined as Hits/At Bats. The average is calculated from all players in each team.

    2. Create a visualization of your choice which will allow you to decide whether a team's batting average is related to its win-loss record. So is a team's winning record related to its batting average?

- Task 3 (20 pts)

    1. Display the average ERA (Earned Run Average) for each of the MLB teams in 2011-2015, sort them in a descending order. A lower ERA indicates a better pitching performance.

    2. Create a visualization of your choice which will allow you to decide if a team's win-loss record is related to its pitching performance. So is a team's winning record related to its pitching performance?