Python Project Baseball

Assignment Overview

This assignment focuses on the design, implementation and testing of Python programs to process data files using lists and tuples, as described below.

Assignment Specifications

1. You will develop a Python program to manage information about baseball players. The program will maintain the following information for each player in the data set:

    player's name (string)
    team identifier (string)
    games played (integer)
    at bats (integer)
    runs scored (integer)
    hits (integer)
    doubles (integer)
    triples (integer)
    homeruns (integer)
    batting average (real)
    slugging percentage (real)

The first nine items will be taken from a data file; the last two items will be computed by the program.   The following formulas will be used for those computations:

    singles = hits - (doubles + triples + homeruns)
    total bases = singles + 2*doubles + 3*triples + 4*homeruns
    batting average = (hits) / (at bats)
    slugging percentage = (total bases) / (at bats)

A player with zero at bats is defined to have a batting average and slugging percentage of zero.

2. The program will recognize the following commands:

    QUIT
    HELP
    INPUT filename
    TEAM identifier
    REPORT n HITS
    REPORT n BATTING
    REPORT n SLUGGING

The program will recognize commands entered with any mix of upper and lower case letters.

The program will be operated interactively: it will prompt the user and accept commands from the keyboard.

If the user enters an invalid command, the program will display an appropriate message and prompt the user to enter another command.

3. The "QUIT" command will halt execution.

4. The "HELP" command will display information to the user about the commands recognized by the program.

5. The "INPUT" command will be followed by a string representing the name of an input file. The program will discard the current data set stored in memory, and then process the input file as the source for a new data set (open the file, read the file once and store the records in a list of tuples, then close the file).

An input file contains zero or more player records, where each record consists of the nine fields listed above. The records are separated by newlines, and the fields are separated by semicolons.

If the user enters an invalid file name, the program will display an appropriate message and prompt the user to enter another file name. The program will halt after the user enters an invalid file name three consecutive times.

6. The "TEAM" command will be followed by a string representing a team identifier. The program will display all information about all players on that team, in alphabetical order.

The information will be displayed in tabular form. The fields will be identified using column headers, and the fields will be aligned beneath the headers.

If the user enters an invalid team identifier, the program will display an appropriate message and prompt the user to enter another command; the program will not display an empty table.

7. The "REPORT" command will be followed by an integer number and a string (one of "HITS", "BATTING" or "SLUGGING").

If the user enters an invalid command, the program will display an appropriate message and prompt the user to enter another command; the program will not display an empty report.

For each report, the program will display all information about the top "n" players in a given category:

    HITS -- number of hits
    BATTING -- batting average
    SLUGGING -- slugging percentage

The reports will be sorted from highest to lowest value in the category. Players with the same number of hits (or same batting average, or same slugging percentage) will be displayed in alphabetical order.

The information will be displayed in tabular form. The fields will be identified using column headers, and the fields will be aligned beneath the headers. All real values will be displayed with three digits of precision following the decimal point.

8. The program will display appropriate messages to inform the user about any unusual circumstances.

9. The program will consist of at least four meaningful functions. Communication between functions will occur via parameters and return values; you may not use any global variables.

Assignment Notes

1. The file named "mlb.2013.txt" contains information about players on the teams in Major League Baseball in 2013.

2. The file named "mlb.part.txt" contains information about 20 players on teams in Major League Baseball in 2013; that file will be useful for your initial development. The first few lines of that file are shown below:

```
De Aza, Alejandro; CWS; 153; 607; 84; 160; 27; 4; 17
Hunter, Torii; DET; 144; 606; 90; 184; 37; 5; 17
Hamilton, Josh; LAA; 151; 576; 73; 144; 32; 5; 21
Choo, Shin-Soo; CIN; 154; 569; 107; 162; 34; 2; 21
Upton, Justin; ATL; 149; 558; 94; 147; 27; 2; 27
Cabrera, Miguel; DET; 148; 555; 103; 193; 26; 1; 44
Posey, Buster; SF; 148; 520; 61; 153; 34; 1; 15
Suzuki, Ichiro; NYY; 150; 520; 57; 136; 15; 3; 7
Holliday, Matt; STL; 141; 520; 103; 156; 31; 1; 22
Headley, Chase; SD; 141; 520; 59; 130; 35; 2; 13
```

3. As noted above, your program must consist of at least four meaningful functions; you certainly may develop more than four functions. You will have to decide how to decompose the overall program into smaller tasks. Note that it would be natural to implement some (or even all) of the subtasks as functions.

Each function must be declared at the global level. Each function should have a coherent purpose which can be expressed succinctly in a line or two.

You may not use any global variables in your program. All communication between the functions which constitute your program will be done using parameters and return values.

4. Note that you must sort the data set (your list of tuples) to prepare the various reports. There are a number of different strategies that you could use, including:

   Develop your own sorting function
   Use the built-in function `sorted`
   Use the list method `sort`

There is a nice introduction to sorting on the Python website:

   https://wiki.python.org/moin/HowTo/Sorting#Sorting_Mini-HOW_TO

Please note the use of function "`itemgetter`" in the section on "Operator Module Functions".

Reminders

- Solve the problem using pencil and paper first.

- Develop a simple version of the program, then add functionality incrementally, testing your program thoroughly after each addition.