JakeDineen_5.4

August 8, 2018

In [13]:
```
import os
import numpy as np
import pandas as pd
import csv
cwd = os.getcwd()
infile = os.listdir(cwd)[4]

# create new empty list
playersList = []

with open(infile, 'rU') as csvfile:
    # the csv file reader returns a list of the csv items on each line
    ALReader = csv.reader(csvfile, dialect='excel', delimiter='\t')

    # from each line, a list of row items, put each element in a dictionary
    #   with a key representing the data
    for line in ALReader:
        # skip lines without data
        if line[0] == '' or line[0].startswith('American') or line[0].startswith('Tea
              or line[0].startswith('Source'):
            continue
        else:
            try:
                # create a dictionary for each player
                player = {}
                # add each piece of data under a key representing that data
                player['team'] = line[0]
                player['name'] = line[1]
                player['sal'] = int(line[2].replace(',', ''))
                player['position'] = line[3]

                # add this player to the list
                playersList.append(player)

            except IndexError:
                print('Error: ', line)
csvfile.close()
```

```
        print("Read", len(playersList), "player data")

        # Write a report text file with a title and the average of the salaries
        # First create an output file name
        outfile1 = infile.replace('tsv', '') + 'report.txt'
        # open the file for writing
        fout1 = open(outfile1, 'w')

        # write title at top of file
        fout1.write("American League Baseball players average salary in 2003\n\n")

        # comput the average salary over all players
        total_salary = 0.0
        for player in playersList:
            total_salary += player['sal']
        average_salary = total_salary / len(playersList)

        # write a report line as a string to the file
        fout1.write('Average salary = ${:,.2f}'.format(average_salary))
        fout1.close()

        # Write a file with all the players who made more than $10 million.
        # We write a comma separated file, using the csv writer to quote the player names wit
        # first create an output file name
        outfile2 = infile.replace('tsv', '') + 'million.csv'

        # open the file
        with open(outfile2, 'w', newline='') as csvfileout:
            # create a csv writer for a comman sep file, with quoting as needed
            ALwriter = csv.writer(csvfileout, delimiter=',', quoting=csv.QUOTE_MINIMAL)
            # write the header row as a list of column labels
            ALwriter.writerow(['Player', 'Team', 'Salary'])
            for player in playersList:
                # select the players with salary under 310k
                if (player['sal'] < 310000):
                    # write the player as a list of data items
                    ALwriter.writerow([player['name'], player['team'], player['sal']])

        csvfileout.close()
```

```
Read 381 player data


C:\Users\jdine\Documents\Anaconda3\envs\tensorflowlatest\lib\site-packages\ipykernel_launcher.p
  # This is added back by InteractiveShellApp.init_path()
```

```
In [14]: pd.read_csv('albb.salaries.2003.million.csv')
```

Out[14]:

	Player	Team	Salary
0	Anderson, Jason	New York Yankees	300000
1	Callaway, Mickey	Anaheim Angels	302500
2	Ramirez, Julio	Anaheim Angels	300000
3	Shields, Scot	Anaheim Angels	305000
4	Wise, Matt	Anaheim Angels	302500
5	Gonzalez, Dicky	Boston Red Sox	300000
6	Lyon, Brandon	Boston Red Sox	309500
7	Person, Robert	Boston Red Sox	300000
8	White, Matt	Boston Red Sox	300000
9	Bard, Josh	Cleveland Indians	302100
10	Broussard, Benjamin	Cleveland Indians	303000
11	Davis, Jason	Cleveland Indians	301100
12	Hafner, Travis	Cleveland Indians	302200
13	Lee, Cliff	Cleveland Indians	300900
14	Myette, Aaron	Cleveland Indians	307500
15	Phillips, Brandon	Cleveland Indians	300900
16	Rodriguez, Ricardo	Cleveland Indians	302400
17	Sadler, Carl	Cleveland Indians	303200
18	Traber, Billy	Cleveland Indians	300000
19	Westbrook, Jake	Cleveland Indians	305500
20	Hendrickson, Mark	Toronto Blue Jays	302000
21	Lopez, Aquilino	Toronto Blue Jays	300000
22	Miller, Trever	Toronto Blue Jays	305000
23	Werth, Jayson	Toronto Blue Jays	300000
24	Bedard, Erik	Baltimore Orioles	300000
25	Leon, Jose	Baltimore Orioles	305000
26	Morban, Jose	Baltimore Orioles	300000
27	Abernathy, Brent	Tampa Bay Devil Rays	300000
28	Baldelli, Rocco	Tampa Bay Devil Rays	300000
29	Bierbrodt, Nick	Tampa Bay Devil Rays	300000
..	...	...	...
49	George, Chris	Kansas City Royals	303500
50	Harvey, Ken	Kansas City Royals	300000
51	Hernandez, Runelvys	Kansas City Royals	305500
52	Johnson, Rontrez	Kansas City Royals	300000
53	Lopez, Mendy	Kansas City Royals	300000
54	MacDougal, Mike	Kansas City Royals	301000
55	Cuddyer, Michael	Minnesota Twins	302500
56	Olivo, Miguel	Chicago White Sox	300000
57	Stewart, Josh	Chicago White Sox	300000
58	Bonderman, Jeremy	Detroit Tigers	300000
59	Cornejo, Nate	Detroit Tigers	303000
60	German, Franklyn	Detroit Tigers	300000
61	Infante, Omar	Detroit Tigers	300000
62	Knotts, Gary	Detroit Tigers	305000
63	Ledezma, Wilfredo	Detroit Tigers	300000
64	Maroth, Mike	Detroit Tigers	309000

65	Roney, Matt	Detroit Tigers	300000
66	Santiago, Ramon	Detroit Tigers	307000
67	Spurling, Chris	Detroit Tigers	305000
68	Bloomquist, Willie	Seattle Mariners	300000
69	Mateo, Julio	Seattle Mariners	302500
70	Snelling, Chris	Seattle Mariners	300000
71	Bowie, Micah	Oakland Athletics	304000
72	Byrnes, Eric	Oakland Athletics	300000
73	Ellis, Mark	Oakland Athletics	307500
74	Neu, Mike	Oakland Athletics	300000
75	Blalock, Hank	Texas Rangers	302500
76	Clark, Jermaine	Texas Rangers	300000
77	Garcia, Reynaldo	Texas Rangers	300000
78	Lewis, Colby	Texas Rangers	302500

[79 rows x 3 columns]