CHAPMAN University

Department of Computational and Data Sciences
CS501 Introductory Computation for Scientists
Fall 2019
Homework#6

Date Given: Sep 25, 2019 Due Date: Oct 1, 2019

==========

There are 10 problems in this homework assignment. Write a program in Python to solve these problems. Use 'datascience' package to write the Python code.

To install 'datascience' package on your computer, specify the following command at the 'Anaconda prompt' level (Windows).

```
pip install datascience
```

Documentation of the 'datascience' package is available at the following URL.

http://data8.org/datascience

In the first cell of your Jupyter notebook, please load the libraries using the following script.

```
from datascience import *
import numpy as np
import matplotlib.pyplot as plots
plots.style.use('fivethirtyeight')
%matplotlib inline
```

Use the data file (top_movies_by_title.csv) for all problems of this assignment. This file is uploaded on Blackboard available for download adjacent to the homework assignment#6 (HW06.pdf) file.

	A	B	C	D	E
1	Title	Studio	Gross	Gross (Adjusted)	Year
2	101 Dalmatian:	Disney	144880014	869280100	1961
3	2001: A Space (	MGM	56954992	377027700	1968
4	9 to 5	Fox	103290500	334062200	1980
5	A Star Is Born (	Warner Bros	80000000	326760600	1976
6	Air Force One	Sony	172956409	327752300	1997
7	Airport	Universal	100489151	575168200	1970
8	Aladdin	Buena Vista (	217350219	456248400	1992
9	Alice in Wonde	Buena Vista (	334191110	365718600	2010
10	American Graf	Universal	115000000	571714300	1973
11	American Snipe	Warner Bros	350126372	374796000	2014
12	An Officer and	Paramount	129795554	379814600	1982

The 'Gross (Adjusted)' amount is the inflation adjusted gross amount. For example, '101 Dalmatians' movie did $144 Million in 1961, it is equivalent to $869 Million in today's dollars.

1. Read the ‘top_movies_by_title.csv’ file into a table and name it ‘table1’.

2. Select ‘Title’ and ‘Year’ of the ‘table1’, into another table and name is ‘table2’.

==========

3. Select only the movies from ‘table1’ where the gross revenues generated by the movie exceeds 100 million.

4. Select only the movies from ‘table1’ where the gross revenues generated by the movie is between 100 and 200 million.

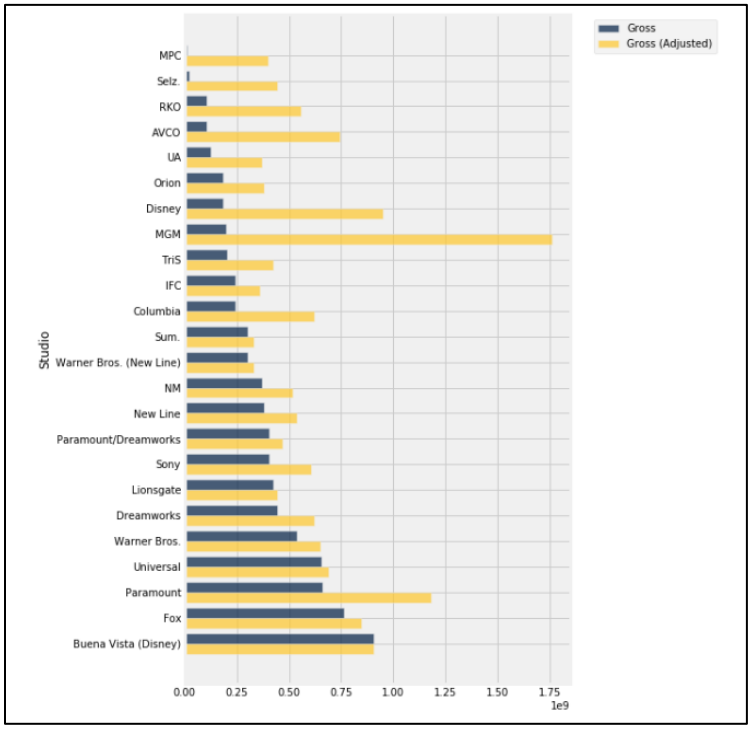
5. Select movies where the word ‘the’ appears in the title of the movie.

==========

6. Sort the movies in ‘table1’ in descending order (highest value first) by their ‘Gross’ revenues.
   Which movie made the most amount of money based on ‘Gross’ column?
   Which movie made the most amount of money based on ‘Gross (Adjusted)’ column?

7. Select the highest grossing movies of each studio (hint: first sort by highest grossing movies and then sort again by studio name with ‘distinct=True’ option).
   Which is the highest grossing movie of ‘Columbia’ and ‘Fox’ studios?

8. The following bar graph shows studios and movies as a function of ‘Gross’ and ‘Gross (Adjusted)’ figures. Generate the following graph using this data file.



9. Compute the ratio between ‘Gross (Adjusted)’ and ‘Gross’ for all movies. Next sort the table by the ratio in descending ratio. Display the top 10 ratios.

10. Group all the movies by the year in which they were released. In which year has the maximum number of movies released?