CSE 216 – Programming Abstractions (Fall 2019)

Programming Assignment # 3

Submission deadline: Sunday 10 November 11:59 PM

In this assignment, you will extend the software program you have developed as part of Assignment 2 to display a particular statistics requested by a user in a graphical form. The program will be developed in Python language using PyCharm IDE (preferred). The data presented in this assignment is taken from The Movies Dataset[1] made available by Rounak Banik on Kaggle – which is an online community of data scientists. These files contain metadata for about 45,000 movies released on or before July 2017. Some information about movies in production or planned is given as well. We will use a file named movies_metadata from this dataset that contains information on the following:

1) adult: TRUE/FALSE
2) genres: Zero or multiple genres. It can be considered as a list of dictionaries in Python. E.g. [{'id': 80, 'name': 'Crime'}, {'id': 35, 'name': 'Comedy'}]
3) id: ID of the movie in this dataset.
4) Imdb_id: ID of the movie as assigned by IMDB
5) Original_language: Original language of movie as given by two letter code (e.g. es for Spanish)
6) Original_title: Original title of the movie
7) Popularity: Some value showing popularity of the movie – higher is better
8) Production_companies: Dictionary list containing names of production companies. E.g. [{'name': 'Universal Pictures', 'id': 33}, {'name': 'Largo Entertainment', 'id': 1644}, {'name': 'JVC Entertainment Networks', 'id': 4248}]
9) Production_countries: Dictionary list containing names of production countires. E.g. [{'iso_3166_1': 'US', 'name': 'United States of America'}]
10) Release_date: Date of release in MM/DD/YYYY format
11) Revenue: Revenue obtained (in $?)
12) Runtime: length of movie in minutes
13) Spoken_language: Language spoken in the movie
14) Status: One of cancelled, In Production, Planned, Post Production, Released
15) Title: Movie title

Your program will accept a calendar period from the user and should be able to answer all of the following questions using graphs:

1) "Number of movies released per month in a particular period." (bar chart)
2) Most popular top 20 movies: (bar chart)
    a. "Most popular top 20 movies in a particular period."
    b. "Most popular top 20 movies released in a particular language."
    c. "Most popular top 20 movies released in a particular genre type."
    d. "Most popular top 20 movies produced in a particular country."
3) Highest earning top 20 movies: (bar chart)

[1] The Movies Dataset, https://www.kaggle.com/rounakbanik/the-movies-dataset.

a. "Top 20 highest earning movies in a particular period."
b. "Top 20 highest earning movies released in a particular language."
c. "Top 20 highest earning movies released in a particular genre type."
d. "Top 20 highest earning movies released in a particular country."

4) Cumulative charts: (Pie chart)
a. "Number of movies produced per language."
b. "Number of movies produced per genre type."
c. "Number of movies produced per country."

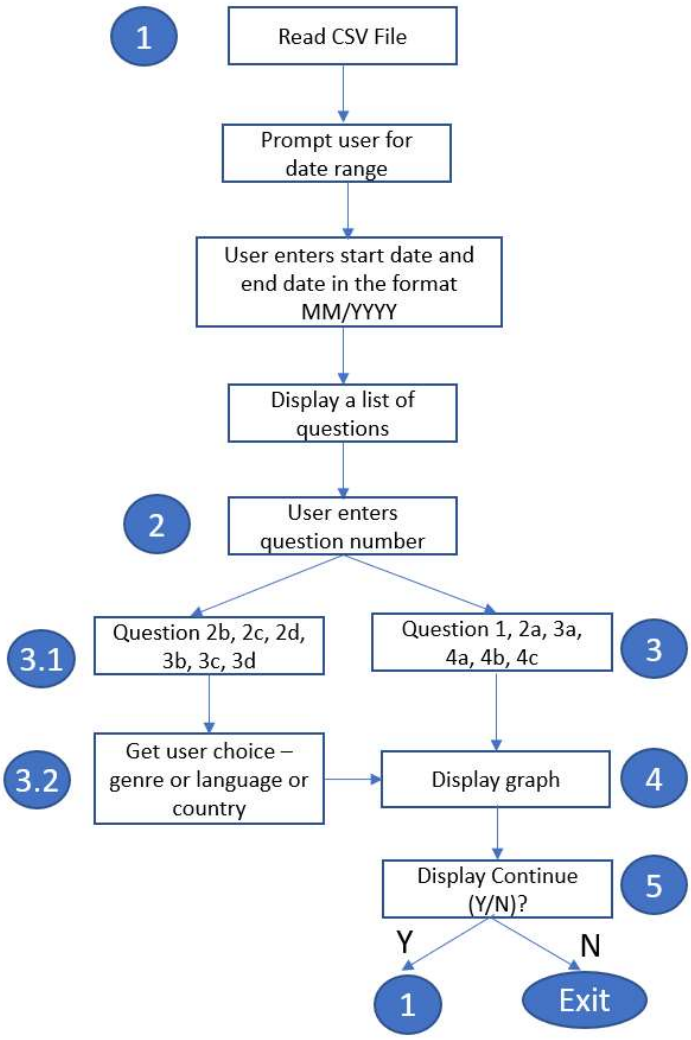

Figure 1: User interaction workflow

Your tasks are the following:

A. Install and learn Python MatPlotlib package:

MatPlotlib is a very powerful plotting library useful for those working with Python. The most used module of Matplotib is Pyplot which provides an interface like MATLAB but instead, it uses Python and it is open source. Follow a short and concise MatPlotlib tutorial made available at

https://towardsdatascience.com/matplotlib-tutorial-learn-basics-of-pythons-powerful-plotting-library-b5d1b8f67596.

B. **User interaction:** You will extend the Assignment 2 project (I have provided my solution, but you are free to use your solution as well) to provide the user interaction shown in Figure 1.

C. Submission:

Submit all the Python scripts you will write as separate files. Do NOT submit the csv data file.

D. **Evaluation:** The graduate TA will download and run your program and it will be checked for correctness using a few use-cases. If required, the graduate TA will contact you for clarifications.

Evaluation Rubric: (Total points: 40)

1. Proper submission: All necessary Python scripts: 5 pts
2. Reaching point 2 of user interaction: +3 pts
3. Reaching point 3.2 of user interaction: +3 pts
4. Displaying proper graphs (point 4): +2 pts each (24 points total)
5. Reaching point 5 of user interaction: +3 pts
6. Restarting from point 1: +2 pts

E. Sample output will be provided soon.

Note:

If your code does not compile, it will not be graded.

Late submissions will not be accepted under any circumstances. Submit whatever you can if you were not able to fully complete the assignment.

To be safe, always, ALWAYS, prepare to submit ahead of time, not exactly AT last moment!

Submission deadline: Sunday 10 November 11:59 PM