CSE 216 – Programming Abstractions (Fall 2019)

Programming Assignment # 2

In this assignment, you will create a software program that shows a particular statistics requested by a user. 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 the questions like:

1) Provide a monthly graph statistic about number of movies released in a particular language.
2) Provide a monthly graph statistic about number of movies released in a particular genre type.
3) Provide a monthly graph statistic about number of movies produced in a particular country.
4) List top 20 movies which earned highest revenue in a particular period.
5) List 20 movies according to their runtime in a particular period.
6) What are the production companies of top 20 movies in a particular period?
7) What are most popular top 20 movies in a particular period?

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

We will use MatPlotLib library to plot these charts according to requirement. See for details: https://matplotlib.org/.

In this assignment, I have provided you a code to read the metadata file line by line. You need to extend this code to complete desired functionality. I will also provide couple of examples on how to draw matplotlib graphs.

You will extend this basic project to provide the following user interaction (to be updated):

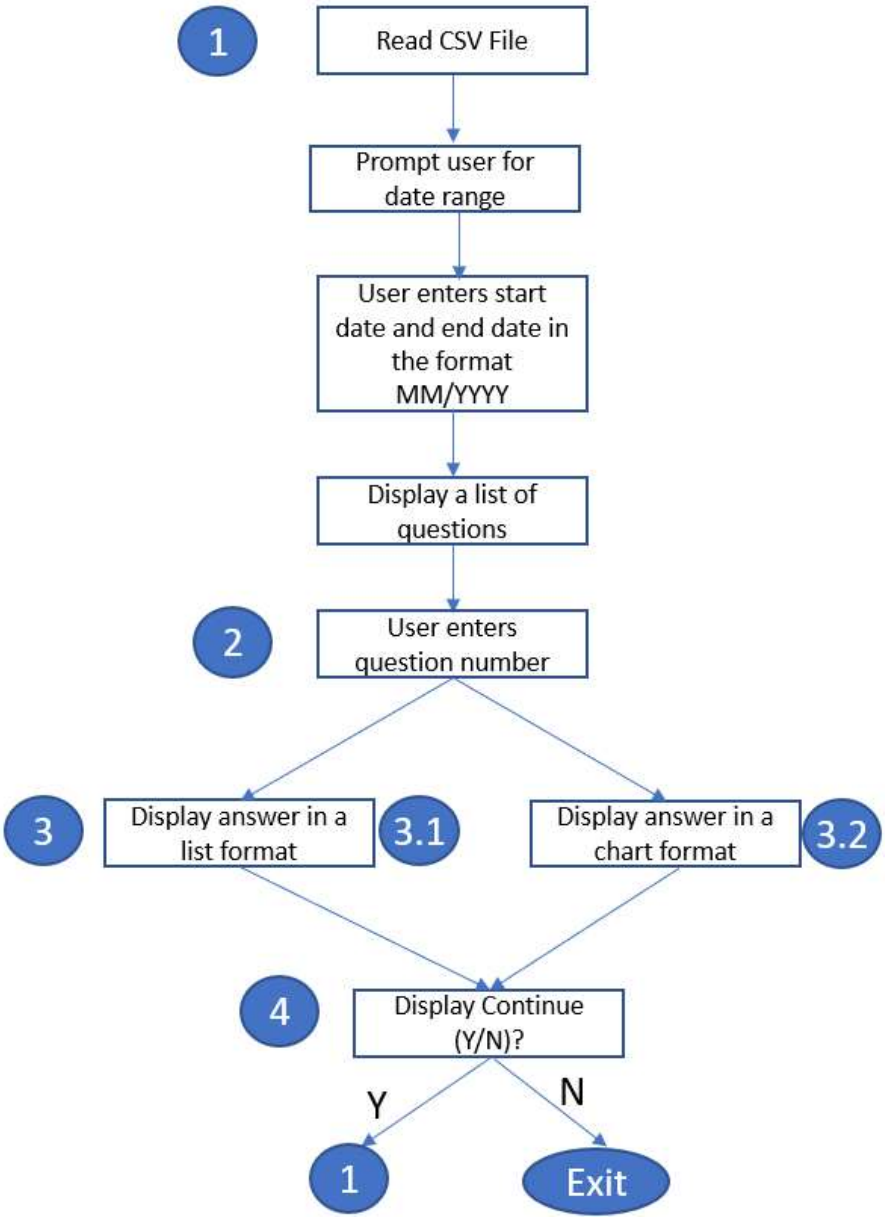

Figure 1: User interaction workflow

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: Monday 30 September, 11:59 PM