The Netflix Movie Tracker

Description of the project: Python is a powerful tool to conduct analysis on large data sets and visualize various attributes of those data sets. The objective of this project is to give you insight into this aspect of programming using Python. You will write a program that reads a data set in CSV format, analyzes the data set. Next, you will answer the questions provided in the following section. You are allowed to use any libraries of your choice for this project.

Data set: The data set is a CSV file (2016_movie_data.csv) that contains several attributes of the movies released in the year 2016. The data set is available here: https://drive.google.com/drive/folders/1MpQNDaLqNydJxKL-ILRKZSZpxNvA1jX?usp=sharing

The data set has the following attributes:
- Movie name
- Release date
- Distributor
- Genre
- MPAA
- Tickets sold

You are required to answer the following questions by writing a python program. Your program should generate graphs. Your graphs should include the necessary label for the axes, the title of the plots and the legends.

1. Find out the details of the data set and print them. Details should include:
   a. The total number of movies released
   b. The number of different genres
   c. The number of different MPAA ratings
   d. The number of different distributors
   e. The total number of tickets sold

2. Plot a bar chart of the number of movies released in different months. In which month was the highest number of movies released?

3. Plot a line chart of the number of tickets sold in different months. In which months were the highest number of tickets sold?

4. Calculate the percentage of tickets sold by each distributor. There are some distributors whose percentage would be less than 1%. Group those under the ‘Others’ category. Print the distributors in descending order of their percentage. Plot a pie chart to show their percentage.
5. Plot a line chart that shows the number of movies from drama, horror, action and comedy genres released in different months of 2016. You need to plot these in the same graph, with proper legends.

**Deliverables:**

1. **The code:** Document your code properly. Every function must have a documentation string.
2. **The project report:** In your report, include the following.
   There are 5 questions provided in the previous section (1, 2, 3, 4, 5). Give a concise explanation of your thought process to answer each of these questions. If the question asks for a visualization, include the graph with its title in the appendix of your report.
   Briefly describe the trend that you observe in each graph. Your graphs and explanation should match the output of your submitted code.

**Demo video:** [https://youtu.be/Lk2Ums6sYyw](https://youtu.be/Lk2Ums6sYyw)

**Sample output:**
[https://drive.google.com/file/d/1ms8YkRfkJ4VsEGxj99VMtPUyt7TOux7i/view?usp=sharing](https://drive.google.com/file/d/1ms8YkRfkJ4VsEGxj99VMtPUyt7TOux7i/view?usp=sharing)
INSTRUCTIONS:

1. Project Submission Guidelines:
The last day to submit your project is Tuesday, December 10th, 2019 at 11:59 PM CST.
You may resubmit your project as many times as you need until the due date. Only the most recent submission is graded. Submit all your files on Gradescope.
No late submissions will be accepted.

The project submission should include:
   a. Your written project report in pdf
   b. Your project file(s)
      You are required to include the following lines in the header of all your code files:
      File: filename.py
      Author: Team members names
      Date: 12/01/2019
      E-mail: Team members emails
      Description: Project name, and description

2. Group Work and Peer-review
This is a group project of 3 members per team. Every team member is required to participate. You will be graded partially by your team members using a peer-review evaluation. If you need additional help with your project, you should ask the TAs or the course instructor. The TAs are available during regular lab time and office hours. You can also post general questions on Piazza.

3. Project report Guidelines:
The report should be typed. Neatness counts. We expect a coherent and professional looking report that includes a table of contents, a narrative, and a conclusion. For conciseness, your report should not exceed 3 pages. The limit of 3 pages does not include the cover page, the table of contents page and the appendix.

Your report must include the following:
Introduction
Give an overview of the project and indicate how the paper is organized.

Body
● Start with a narrative that describes the functionality of your program.
● Include all the answers to the project questions.

Conclusion
● What was your experience writing this program?
● Review and summarize your work including the challenges that you encountered and possible improvements that could be added.
Here is a project report template:
https://docs.google.com/document/d/1LXJh1dmRzxKCKpznpsfgP2uFgwnvn8wMBrltx9D_6RU/edit?usp=sharing
4. **Academic Dishonesty:**

*Your code will be verified for plagiarism, and authenticity of the work submitted.* You MUST do your own work; treat this project as a take-home examination. We will treat any cases of plagiarism seriously. As this project can take on many forms in its implementation, if we find two or more submissions that are substantially similar, the teams involved will receive a grade of zero for the project.

5. **Grading**

This final project is worth **8%** of your total grade. We will use the following guidelines for grading:

<table>
<thead>
<tr>
<th>Description</th>
<th>Poor Report</th>
<th>Average Report</th>
<th>Exceptional Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>The program does not work at all</td>
<td>10 pts</td>
<td>20 pts</td>
<td>30 pts</td>
</tr>
<tr>
<td>The program works partially and fails more than 50% of the test cases.</td>
<td>30 pts</td>
<td>40 pts</td>
<td>50 pts</td>
</tr>
<tr>
<td>The program works partially and passes more than 50% of the test cases.</td>
<td>60 pts</td>
<td>70 pts</td>
<td>80 pts</td>
</tr>
<tr>
<td>The project works and passes all test cases</td>
<td>70 pts</td>
<td>80 pts</td>
<td>100 pts</td>
</tr>
</tbody>
</table>