Assignment 3

DA 219, Spring 2023

In this assignment, we will go over how to use the Pandas library of Python to load a real world dataset of IMDB movie ratings, perform data cleaning, and perform some basic data analysis along the way. We will answer questions like:

1. What is the longest duration of any movie in the dataset
2. How many directors are there
3. How many different movie genres are there
4. What is the number of different languages for the movies in the dataset
5. Which of the movies have the highest profit
6. Which are the Top 250 IMDb Movies

For this assignment, you will need to do all your work in an iPython notebook and submit that. You can fully complete this assignment in [Google Collaboratory](#). You do not need to download any special IDE for iPython locally on your machine (if you have a Google account). If you don't have one, you can create an account for free for this assignment. If you don't want to do that, you are welcome to download Anaconda and use Jupyter Notebooks locally on your computer.

In the zip folder for this assignment, you will find the handout, a `movies_data.zip` file, and a `HW3_movies_data_analysis.ipynb`. The `movies_data.zip` has the dataset, and `HW3_movies_data_analysis.ipynb` is the iPython notebook that has been populated partially and has the guidelines for what to do, and what questions to answer. If you are working on Google Collab, please be sure to upload both these files onto your Google Drive. Make sure to use the correct path of your `movies_data.zip file` in Step 0 of the assignment.

You can answer the text questions within the same iPython notebook by creating a new text cell. For cells that require you to write code, you will fill the cells as directed by the `TODO` prompts in the code cells. I have provided some soft guidance for how many lines of code are required for reach prompt. You are free to use more lines of code if you wish, but it should be in the ballpark of the guidance.

Please submit the completed `HW3_movies_data_analysis.ipynb` notebook on canvas. No need to zip it. Note that the points for each part of the assignment are stated in the assignment itself at the end of each prompt (in square brackets). Total points are 165.

Google luck with the assignment, and don't forget to have fun!