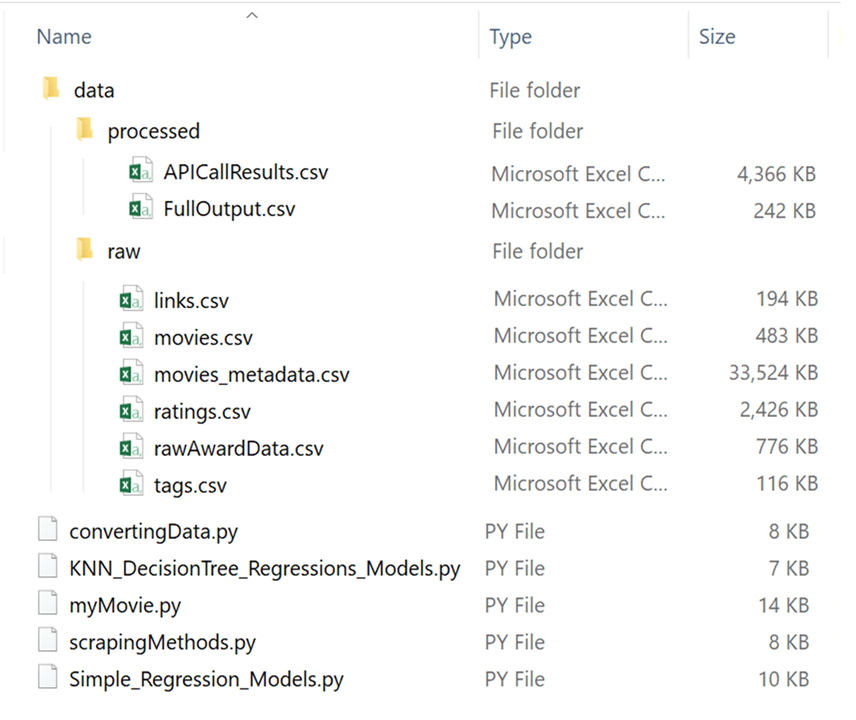
**About MyMovie v1.0:**

This compact application aims to help empowering producers to make successful movies based on publicly available historical data through our regression model, and allowing users to explore information and figures related to past successful films.

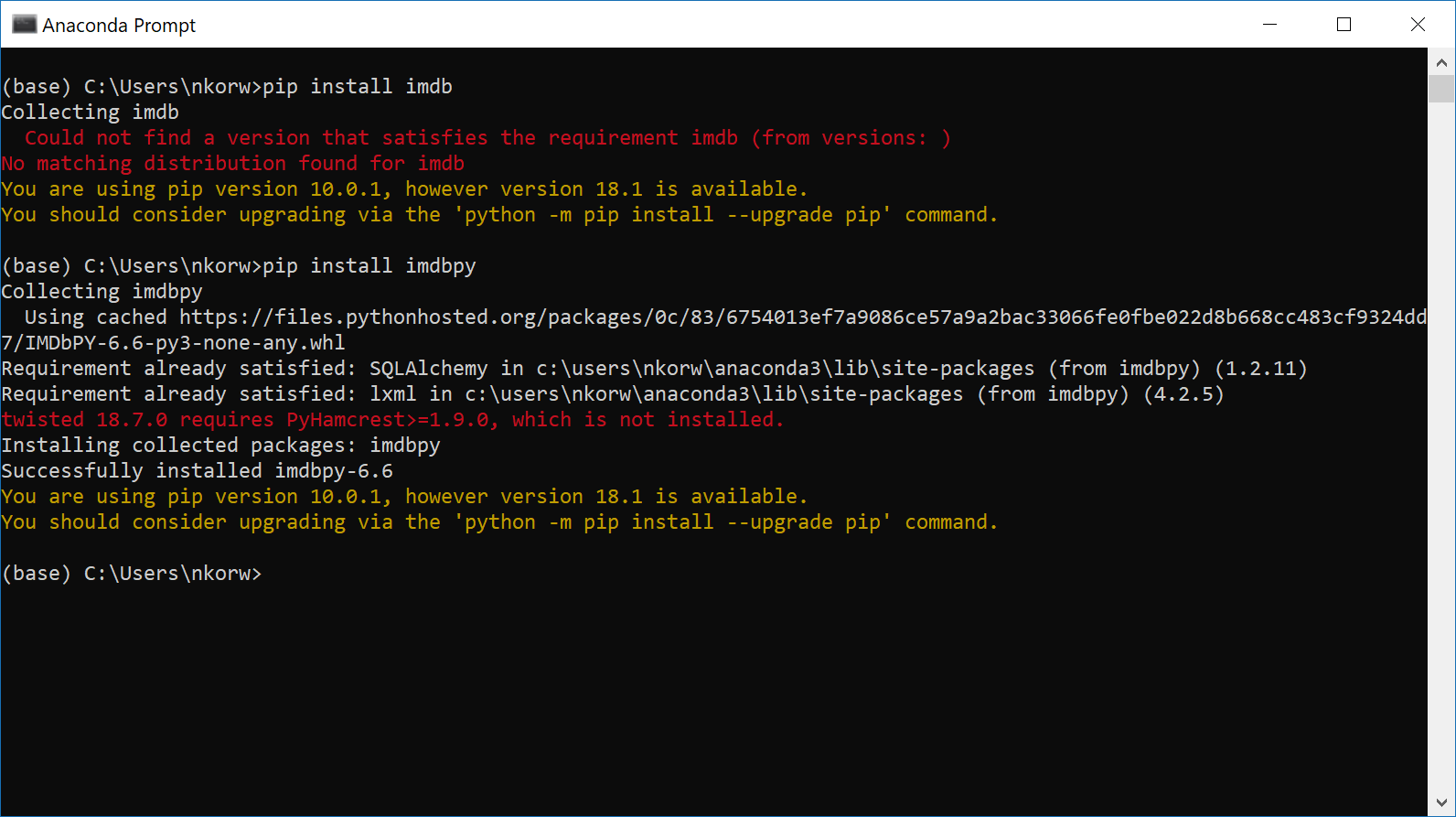
Created for 95888 Data Focused Python Course in Fall 2018 - Heinz College, Carnegie Mellon University

Contributors: Shayne Bement, Jeffrey Curran, Ghazal Erfani, Naphat Korwanich, Asvin Sripraiwalsupakit

**MyMovie Installation Instructions:**

1. Setup the file structure  
   When unzipping the file to your local computer, the structure should be as shown.  
   
2. Install Anaconda 3 version 5.3.0 or above  
   For computers without prior Python installation, Anaconda 3 should have all the necessary packages installed with only one additional Python package installation needed to be made (see Step 3)
3. Install the [IMDB](https://imdbpy.readthedocs.io/en/latest/) API package: IMDbPY 6.6

Launch Anaconda Prompt and type **pip install imdbpy**

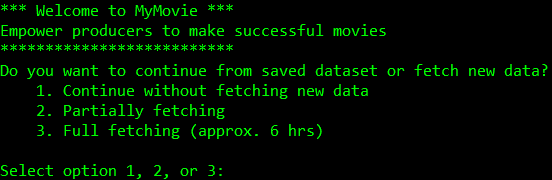


Note For users who might already have other versions of Python installed, it is advised to check for a complete Python package installation simply by running pip install command for the following packages

* NumPy pip install numpy
* Pandas pip install pandas
* Seaborn pip install seaborn
* sklearn pip install sklearn
* Matplotlib pip install matplotlib
* BeautifulSoup4 pip install beautifulsoup4

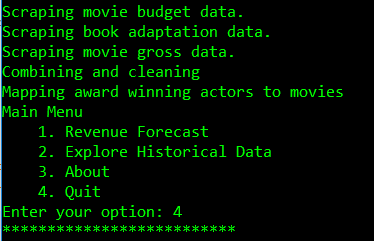
**Running the Application:**

1. To run the application, navigate to the folder you have extracted MyMovie.zip in the command line.
2. From here you can run the application by running the file myMovie.py. Note that if you have multiple versions of python installed, you will need to specify python 3.
3. When the program launches, you will be presented with this menu:



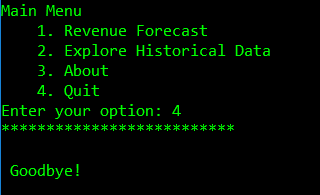
In this menu you have 3 options. Option 1 will use the preprovided set of data that comes with the application to produce the regression model and results. This option will not necessarily provide up to date information, and thus should be used only if there is not time for another option. Option 2 will do a quick refresh of the data, meaning it will scrape only the wikipedia, budget, and gross information, but use the actor data from a preprovided set, thus giving partial updates to the data. This updated will take roughly 5 minutes and the data will the replace the set we provided. Finally, option 3 does a full scrape, meaning it does the wikipedia, budget, and gross scrapes, as well as pinging the imdb api for actor data on all the movies we get from the webscraping. This process takes up to 6 hours, but also updates the preprovided dataset. The idea was that this could be run maybe once a week to ensure up to date base information, then use options 1 or 2 for day to day tasks.

1. For our walkthrough we will select option 2. However, the options will only differ visually up until the main menu selection, and these differences only show progress on scraping. We will use option 2, this is our quickscrape option. Selecting this option will take a few minutes to scrape the data, but we provide updates on the progress so you know the program isn’t hanging. When finished scraping and aggregating, you will get the following output:

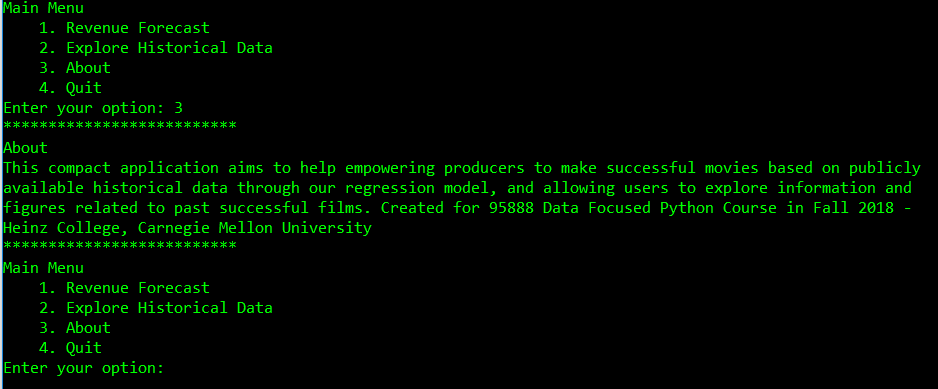


You can see that we got updates as the movie data was scraped from our various sources and the data was combined and cleaned. When this is finished we get the new option menu.

1. As seen in the menu, you have 4 options. We will start from the bottom and work our way up.
   1. Option 4: Quit
      1. This is self explanatory, quits the program completely. Should look like this:



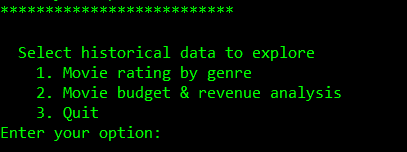
* 1. Option 3: About
     1. This selection will show the about message that is printed at the top of this read\_me file:



As you can see, this also prompts the user to continue if they’d like.

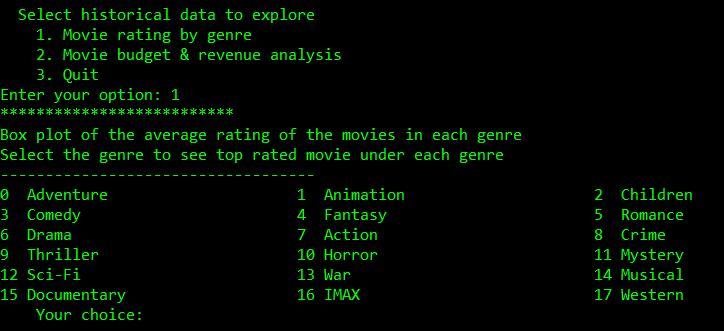
* 1. Option 2: Explore Historical Data
     1. This section uses data about movie performance to generate plots. This is explored in step 6.
  2. Option 1: Revenue Forecast
     1. This section provides the user the ability to input parameters for various regression and clustering algorithms to predict the gross of your potential movie. This is explored in step 7.

1. In this step we will look at the Explore Historical Data choice from the main menu. In this choice you will get plots describing historical trend in movie information. When you select this option, the following menu will appear:



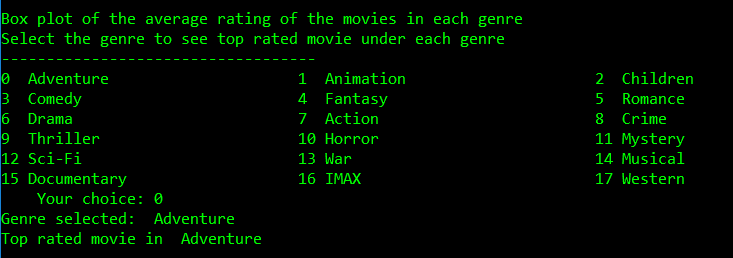
Here we see a few options. We can either get movie ratings by genre, or get a graphical view of movie budget and revenue analysis.

* 1. Option 1: Movie rating by genre
     1. Selecting this option will generate a few plots. First will be a box plot which appears in a separate window. After this plot, you will be prompted to enter a genre like the following:



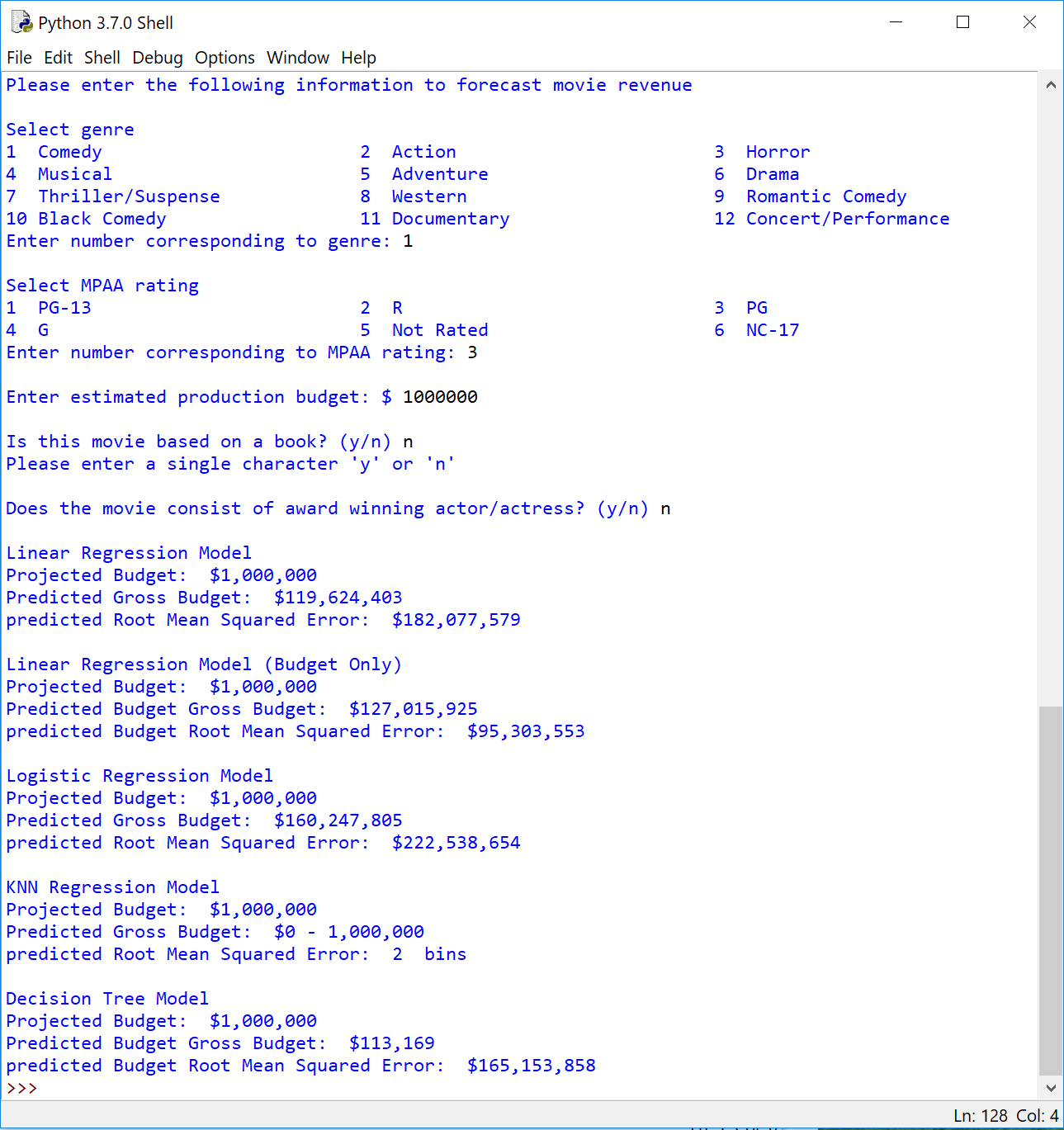
At this selection, you simply enter the number corresponding to the genre. So if I wanted to see the graph for Adventure movies, I would simply type in the number 0. When I do this additional plots will be generated for that genre.

* + 1. Selecting 0 generates a box plot and a bar plot of historically top performing movies in that genre.



* + 1. After these plots are generated, the program exits.
  1. Option 2 : Movie Budget and Revenue Analysis
     1. This option first generates 3 plots, an average revenue per movie trend, an average movie budget per movie trend plot, and finally a line plot of both of these trends. These plots are all given in millions of dollars. When the user closes these plots, another set of 2 box-plots is made, showing movie production and gross as a box plot.

1. In this step we will look into the Revenue Forecast option from the main menu. Selecting this option will generate a number of regression and clustering models. It will then use these models to predict a value of gross income for the inputs the user provides. When you select this option, you are given a series of prompts, shown below:



* 1. The first request is for the genre, simply input the number corresponding to the genre your movie falls under. For example, I have a comedy so I entered 1.
  2. Then, the program will ask for the project rating of your movie. Again entering a number, in my case 3 for PG.
  3. Next, it will ask for a budget in USD. I found some change in the carpet today, so I’m putting in a million dollars.
  4. Then you are asked if the movie is based on a book, simply enter y or n for yes or no.
  5. Finally, you are asked if the movie will have an award winning actor or actress involved. Again this is a boolean y or n for yes or no.
  6. After all these inputs, these variables are input into our regressions, and a series of potential revenue outcomes are presented to the user.
  7. After these outputs are given, the program closes

1. Rerun the program if you would like to try with different parameters.

Thank you - the MyMovie team.