Final Project Name: Lin Li Uniqname: lllinda

Github Repo: https://github.com/lllinda123/FinalprojectW21

Demo Link: https://www.youtube.com/watch?v=OS0tpSmW7TA

README:

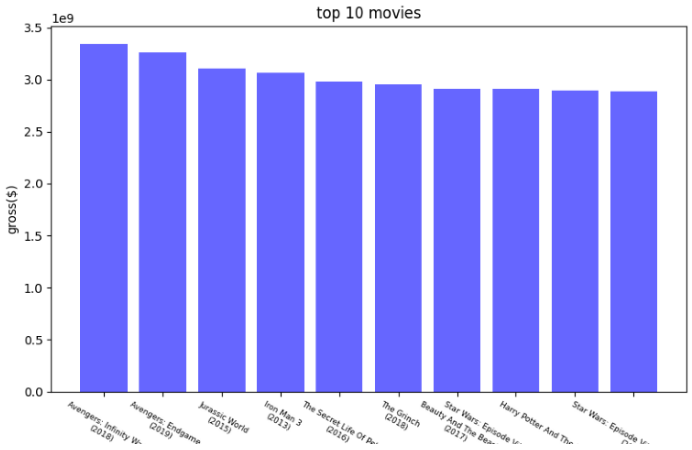
Packages needed: bs4, requests, matplotlib, pandas, numpy, sqlite3, flask, time, json, webbrowser, os

Interaction and Presentation Plans:

1. When running the program for the very first time (before storing the data in the database), information of all the movies fetched will be displayed in the format of " The #1 release of <year> <quarter> is <name> "

```
--------------------------------------------------
Q1 Historical Box Office
--------------------------------------------------
The #1 release of 2021 q1 is Tom And Jerry
The #1 release of 2020 q1 is Bad Boys For Life
The #1 release of 2019 q1 is Captain Marvel
The #1 release of 2018 q1 is Black Panther
The #1 release of 2017 q1 is Beauty And The Beast
The #1 release of 2016 q1 is Deadpool
The #1 release of 2015 q1 is American Sniper
The #1 release of 2014 q1 is The Lego Movie
The #1 release of 2013 q1 is Oz The Great And Powerful
The #1 release of 2012 q1 is The Hunger Games
The #1 release of 2011 q1 is The King'S Speech
The #1 release of 2010 q1 is Avatar
The #1 release of 2009 q1 is Paul Blart: Mall Cop
The #1 release of 2008 q1 is Horton Hears A Who!
The #1 release of 2007 q1 is 300
```

2. The program will ask: “Would you like to check the bar chart of the top 10 movies with highest cumulative_gross to help you to pick a movie you may be interested in? (answer: yes/no):” If ‘yes’, a bar chart will pop up to help users more intuitively understand which movie may be better.



3. After displaying the initial information, the program will ask the user 'Did you find a movie you are interested in? (answer: yes/no/leave) '.
   If 'yes', the program will ask 'Enter the name of the movie, I will show you more info:' and display more detailed information in a table. At the same time, it will exit the program automatically.

```
--------------------------------------------------
Enter a year that you want to know more about the rank of release(range: 1978-2021)
or "leave" if you do not need that:Leave
--------------------------------------------------
Did you find a movie you are interested in? (answer: yes/no/leave):yes
--------------------------------------------------
Enter the name of the movie, I will show you more info:Galaxy Quest
--------------------------------------------------
Fetching
Saving data to database...
Data has been successfully saved.
+-------------------+--------------------------------------------------+
|       Title       |                  Galaxy Quest                    |
|      Released     |                  25 Dec 1999                     |
|      Runtime      |                    102 min                       |
|       Genre       |            Adventure, Comedy, Sci-Fi             |
|      Director     |                  Dean Parisot                    |
|       Actors      |Tim Allen, Sigourney Weaver, Alan Rickman, Tony...|
|      Language     |                    English                       |
|      Country      |                      USA                         |
|       Awards      |             7 wins & 14 nominations.             |
+-------------------+--------------------------------------------------+
```

4. If 'no', then the program will continue to help them find movies they may like. The specific way is: the user can enter any year (between 1978-2021), and then a webpage will pop up, which will display all the movie rankings of that year. This process is repeated until the user chooses to leave. After that, I assume that the user has found his favorite movie, so he will be asked again if he finds a movie that he's interested in, and the details will be displayed for him.
   Note: if the user does not find a favorite movie in the end, he can enter "leave" when the question 'Did you find a movie you are interested in? ' appears to leave the program.

```
--------------------------------------------------
Would you like to check the bar chart of the top 10 movies with highest cumulative_gross
to help you to pick a movie you may be interested in? (answer: yes/no):no
--------------------------------------------------
Did you find a movie you are interested in? (answer: yes/no/leave):no
Maybe I can help you to find one :)
--------------------------------------------------
Enter a year that you want to know more about the rank of release(range: 1978-2021)
or "leave" if you do not need that:Leave
--------------------------------------------------
Did you find a movie you are interested in? (answer: yes/no/leave):no
Maybe I can help you to find one :)
--------------------------------------------------
Enter a year that you want to know more about the rank of release(range: 1978-2021)
or "leave" if you do not need that:2000
--------------------------------------------------
Launching
https://www.boxofficemojo.com/quarter/q1/2000/?ref_=bo_ql_table_22
in web browser...
--------------------------------------------------
```

Domestic Box Office For Q1 2000

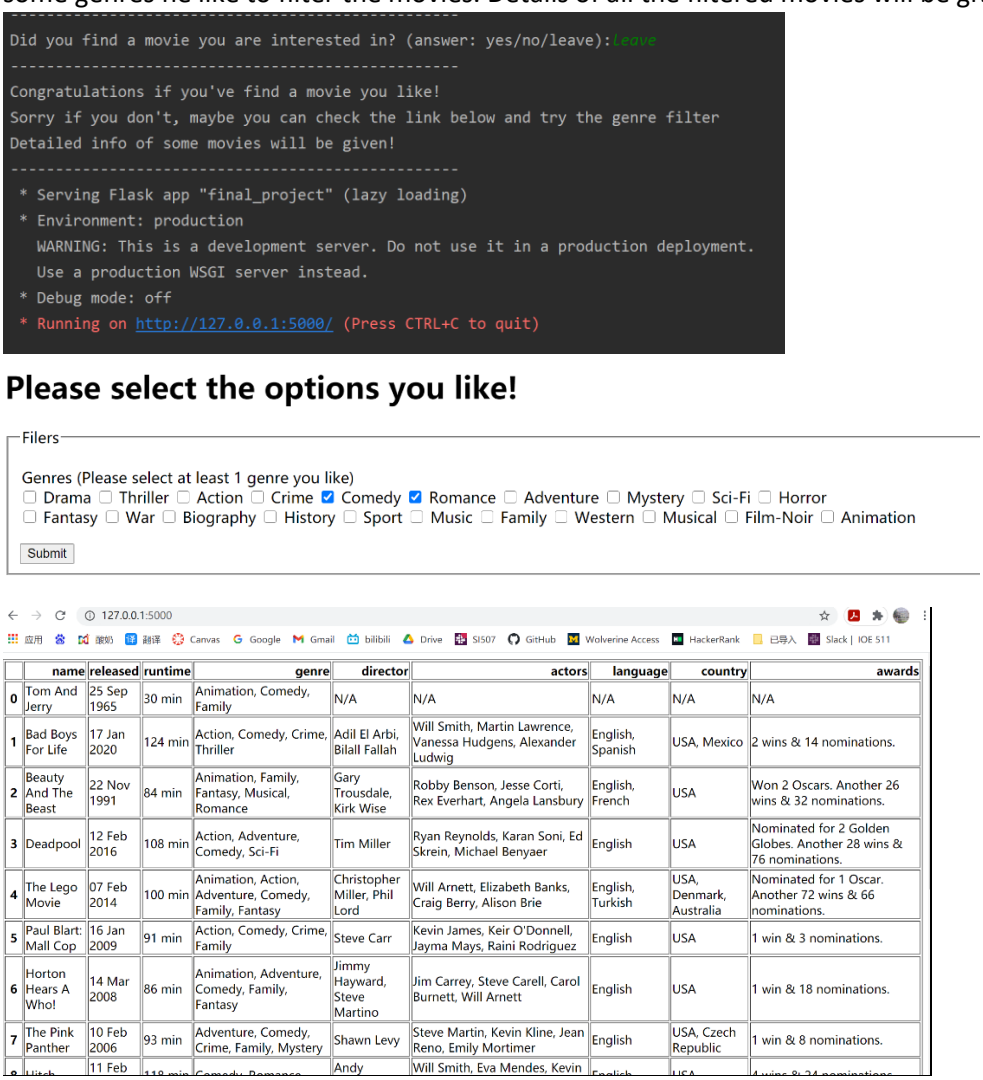
By Quarter | Q1 | 2000 | Calendar grosses

Q1 2000 runs Jan 1-Mar 31

Rank	Release	Gross	Theaters	Total Gross	Release Date	Distributor
1	Scream 3	$86,680,240	3,467	$89,143,175	Feb 4	Dimension Films
2	Stuart Little	$71,566,121	3,151	$140,035,367	Dec 17	Sony Pictures Entertainment (SPE)
3	The Green Mile	$68,261,221	2,875	$136,801,374	Dec 10	Warner Bros.
4	Erin Brockovich	$66,214,660	3,070	$125,595,205	Mar 17	Universal Pictures
5	Snow Day	$58,216,267	2,717	$60,020,107	Feb 11	Paramount Pictures
6	Next Friday	$56,163,192	1,420	$57,328,603	Jan 12	New Line Cinema
7	The Whole Nine Yards	$54,978,150	2,910	$57,262,492	Feb 18	Warner Bros.
8	Mission to Mars	$52,128,480	3,101	$60,883,407	Mar 10	Walt Disney Studios Motion Pictures
9	The Talented Mr. Ripley	$51,320,743	2,369	$81,298,265	Dec 25	Paramount Pictures
10	Galaxy Quest	$50,395,869	2,450	$71,583,916	Dec 25	DreamWorks Distribution

5. <u>If the user leave the program without checking detailed information of any movie, I will assume he didn't find a good movie and offer one last option.</u> He can go to the webpage and select some genres he like to filter the movies. Details of all the filtered movies will be given.

```
--------------------------------------------------
Did you find a movie you are interested in? (answer: yes/no/leave):Leave
--------------------------------------------------
Congratulations if you've find a movie you like!
Sorry if you don't, maybe you can check the link below and try the genre filter
Detailed info of some movies will be given!
--------------------------------------------------
 * Serving Flask app "final_project" (lazy loading)
 * Environment: production
   WARNING: This is a development server. Do not use it in a production deployment.
   Use a production WSGI server instead.
 * Debug mode: off
 * Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)
```



	name	released	runtime	genre	director	actors	language	country	awards
0	Tom And Jerry	25 Sep 1965	30 min	Animation, Comedy, Family	N/A	N/A	N/A	N/A	N/A
1	Bad Boys For Life	17 Jan 2020	124 min	Action, Comedy, Crime, Thriller	Adil El Arbi, Bilall Fallah	Will Smith, Martin Lawrence, Vanessa Hudgens, Alexander Ludwig	English, Spanish	USA, Mexico	2 wins & 14 nominations.
2	Beauty And The Beast	22 Nov 1991	84 min	Animation, Family, Fantasy, Musical, Romance	Gary Trousdale, Kirk Wise	Robby Benson, Jesse Corti, Rex Everhart, Angela Lansbury	English, French	USA	Won 2 Oscars. Another 26 wins & 32 nominations.
3	Deadpool	12 Feb 2016	108 min	Action, Adventure, Comedy, Sci-Fi	Tim Miller	Ryan Reynolds, Karan Soni, Ed Skrein, Michael Benyaer	English	USA	Nominated for 2 Golden Globes. Another 28 wins & 76 nominations.
4	The Lego Movie	07 Feb 2014	100 min	Animation, Action, Adventure, Comedy, Family, Fantasy	Christopher Miller, Phil Lord	Will Arnett, Elizabeth Banks, Craig Berry, Alison Brie	English, Turkish	USA, Denmark, Australia	Nominated for 1 Oscar. Another 72 wins & 66 nominations.
5	Paul Blart: Mall Cop	16 Jan 2009	91 min	Action, Comedy, Crime, Family	Steve Carr	Kevin James, Keir O'Donnell, Jayma Mays, Raini Rodriguez	English	USA	1 win & 3 nominations.
6	Horton Hears A Who!	14 Mar 2008	86 min	Animation, Adventure, Comedy, Family, Fantasy	Jimmy Hayward, Steve Martino	Jim Carrey, Steve Carell, Carol Burnett, Will Arnett	English	USA	1 win & 18 nominations.
7	The Pink Panther	10 Feb 2006	93 min	Adventure, Comedy, Crime, Family, Mystery	Shawn Levy	Steve Martin, Kevin Kline, Jean Reno, Emily Mortimer	English	USA, Czech Republic	1 win & 8 nominations.
8	Hitch	11 Feb	118 min	Comedy, Romance	Andy	Will Smith, Eva Mendes, Kevin	English	USA	4 wins & 24 nominations.

To sum up, other than simple print( ) statements, my program can produce 4 different presentation.

Data source:

https://www.boxofficemojo.com/quarter/q1/?grossesOption=calendarGrosses

format: HTML (scrape new single page, caching used challenge score:4)

records available & # records retrieved = 185

I will actually scrape 4 pages, each of them contains data of the #1 release box office of a quarter, and each of them has about 50 records (1972-2021, one record for each year).

Year	Cumulative Gross	%± LY	Releases	Average	#1 Release	Gross	% of Total
2021	$222,694,767	-87.6%	109	$2,043,071	Tom and Jerry	$30,752,852	13.8%

I will save year, cumulative gross, average gross, name, rank_of_year_link (a link to a webpage in which shows all the movie rankings of that year) and the corresponding quarter in my database.

http://www.omdbapi.com/

format: JSON (Web API haven't used that requires API key, caching used challenge score:4)

records available = inf # records retrieved = 185

I will find the detailed information of the corresponding movie on this website according to the movie name fetched in the first table. I will save title, released(date), runtime, genre, director, actors, language, country, awards, imdbRating in my database.

Database:

```python
def save_to_database(data_in_rows, table_name="Movie_Rank"):
    print("Saving data to database...")
    conn = sqlite3.connect(DATABASE_NAME)
    data_in_rows.to_sql(table_name, conn, index=False, if_exists='replace')
    conn.close()
    print("Data has been successfully saved.")


def save_to_database_append(data_in_rows, table_name="Movie_Info"):
    print("Saving data to database...")
    conn = sqlite3.connect(DATABASE_NAME)
    data_in_rows.to_sql(table_name, conn, index=False, if_exists='append')
    conn.close()
    print("Data has been successfully saved.")
```

```python
# save data into 2 tables
if os.path.exists(DATABASE_NAME):
    pass
else:
    quarters = ['q1', 'q2', 'q3', 'q4']
    data_in_rows = []
    for q in quarters:
        d = extract_box_office_data(q)
        data_in_rows.extend(d)
    df = pd.DataFrame(data_in_rows, columns=['year', 'quarter', 'name', 'cumulative_gross',
                                             'per_release_average_gross', 'rank_of_year_link'])
    save_to_database(df, table_name="Movie_Rank")
    r = read_from_rank(2)
    fetch_list = [r.name[i] for i in range(len(r))]
    # print(fetch_list)
    data_in_rows1 = []
    for t in fetch_list:
        # print(t[0], t[1])
        data1, list_of_tuple1 = get_movie_info(t)
        data_in_rows1.append(data1)
        df = pd.DataFrame(data_in_rows1, columns=['name', 'released', 'runtime', 'genre', 'director',
                                                  'actors', 'language', 'country', 'awards'])
        save_to_database(df, table_name="Movie_Info")
```

Data in tables: