ANALYSING THE MOVIES DATASET AND GIVING MICROSOFT GOOD RECOMMENDATIONS BASED ON THE ANALYSIS

Presenter: Ahmed Ali.

Business Understanding

With Microsoft recently branching out to make a new studio for making movies and not knowing about how to make movies put their idea to a halt.

The goals of this project is to conduct data analysis and give Microsoft recommendations on their new endeavor based on the datasets provided collected from trusted databases.

In order for them to become a competitive in the emerging streaming world and in the box office as well as generate profit.

DATA UNDERSTANDING

After being provided with the zipped data, I started with unzipping the dataset and found the relevant datasets which would help us in our project.

The main datasets which were helpful to us in this endeavor was ***im.db*** dataset which was in SQL language which means the datasets were contained in a number of tables which were zipped.

The other dataset was the ***bom.movie_gross.csv*** which was in a csv format that means comma-separated values which is much easier to work with unlike in SQL.

DATA ANALYSING

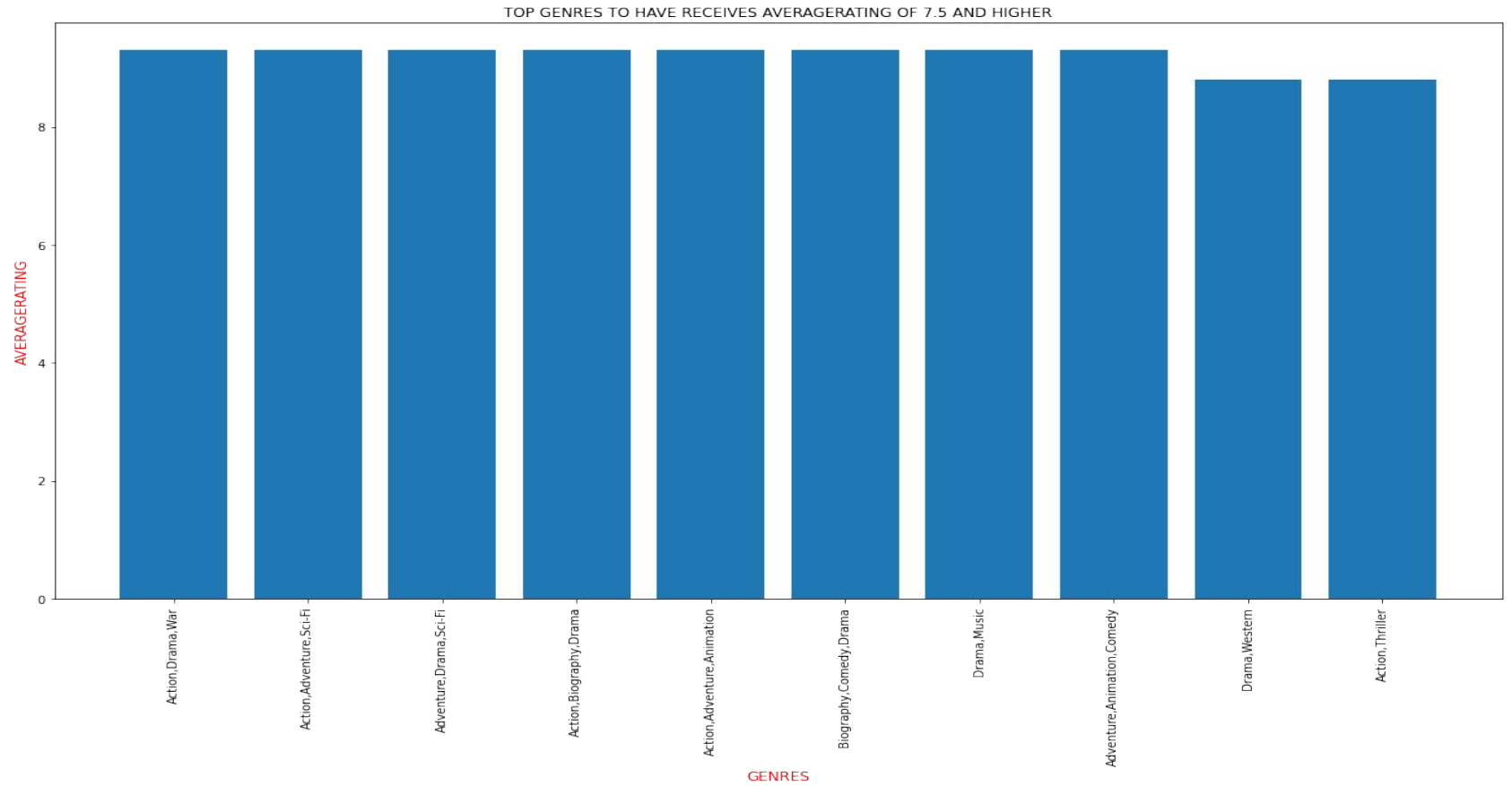
1. The IMDB Database

I started exploring the datasets by loading them, starting with the imdb SQL database. With the help of pandas, a popular python library I was able to first query the tables found in the database and then turn them into DataFrames.

I was able to extract the relevant data which were in ***persons, principals,movie_basics*** and ***movie_ratings***. I first of all was able to query by joining the ***movie_basics*** and ***movie_ratings*** tables based on which movie received the ratings above 7.5 and votes of over a hundred thousand to extract the genres and the primary title.

The following graph in the next slide summarizes it all.

GRAPH SHOWING GENRES WHICH RECEIVED RATINGS OF ABOVE 7.5 ON IMDB



Explaining the graph above:

The graph shows us the the main genres which captivated people hearts that led them to rate higher ratings were a combination of the following genres:

1. Action
2. Drama
3. War
4. Adventure
5. Sci-Fi
6. Animation

The top celebrities involved in performing movies.

I then queried the principals,persons and movie ratings tables by joining them on which movie received the 7.5 and above and votes of over hundred thousand to extract the top directors,actors,actresses and writers.this are the following lists.

The top actors are:

1. Murat Arkin
2. Bedii Akin
3. Ozan Agaç
4. Chris Hemsworth
5. Mark Ruffalo
6. Ken Watanabe
7. Joseph Gordon-Levitt
8. Robert Downey Jr.
9. Leonardo DiCaprio
10. Chris Evans
11. Matthew McConaughey
12. Mahershala Ali
13. Aamir Khan
14. François Cluzet
15. Omar Sy
16. Shameik Moore
17. Jake Johnson
18. J.K. Simmons
19. Miles Teller
20. Paul Reiser

The top actresses are:

1. Eylül Arular
2. Ellen Page
3. Mackenzie Foy
4. Anne Hathaway
5. Jessica Chastain
6. Fatima Sana Shaikh
7. Sanya Malhotra
8. Hailee Steinfeld
9. Anne Le Ny
10. Audrey Fleurot
11. Sakshi Tanwar
12. Melissa Benoist
13. Mone Kamishiraishi
14. Alanna Ubach
15. Kerry Washington
16. Aoi Yûki
17. Sareh Bayat
18. Linda Cardellini
19. Leila Hatami
20. Joan Cusack

The top directors are:

1. Alper Caglar
2. Anthony Russo
3. Joe Russo
4. Christopher Nolan
5. Rodney Rothman
6. Peter Ramsey
7. Bob Persichetti
8. Olivier Nakache
9. Éric Toledano
10. Nitesh Tiwari
11. Damien Chazelle
12. Lee Unkrich
13. Adrian Molina
14. Quentin Tarantino
15. Makoto Shinkai
16. Peter Farrelly
17. Asghar Farhadi
18. Thomas Vinterberg
19. Denis Villeneuve
20. Martin McDonagh

The top writers are:

1. Christopher Markus
2. Jack Kirby
3. Stephen McFeely
4. Stan Lee
5. Jonathan Nolan
6. Phil Lord
7. Rajshri Sudhakar
8. Shreyas Jain
9. Piyush Gupta
10. Nikhil Mehrotra
11. Philippe Pozzo di Borgo
12. Jason Katz
13. Bob Kane
14. David S. Goyer
15. Matthew Aldrich
16. Clark Cheng
17. Brian Hayes Currie
18. Nick Vallelonga
19. Tobias Lindholm
20. Valérie Beaugrand-Champagne

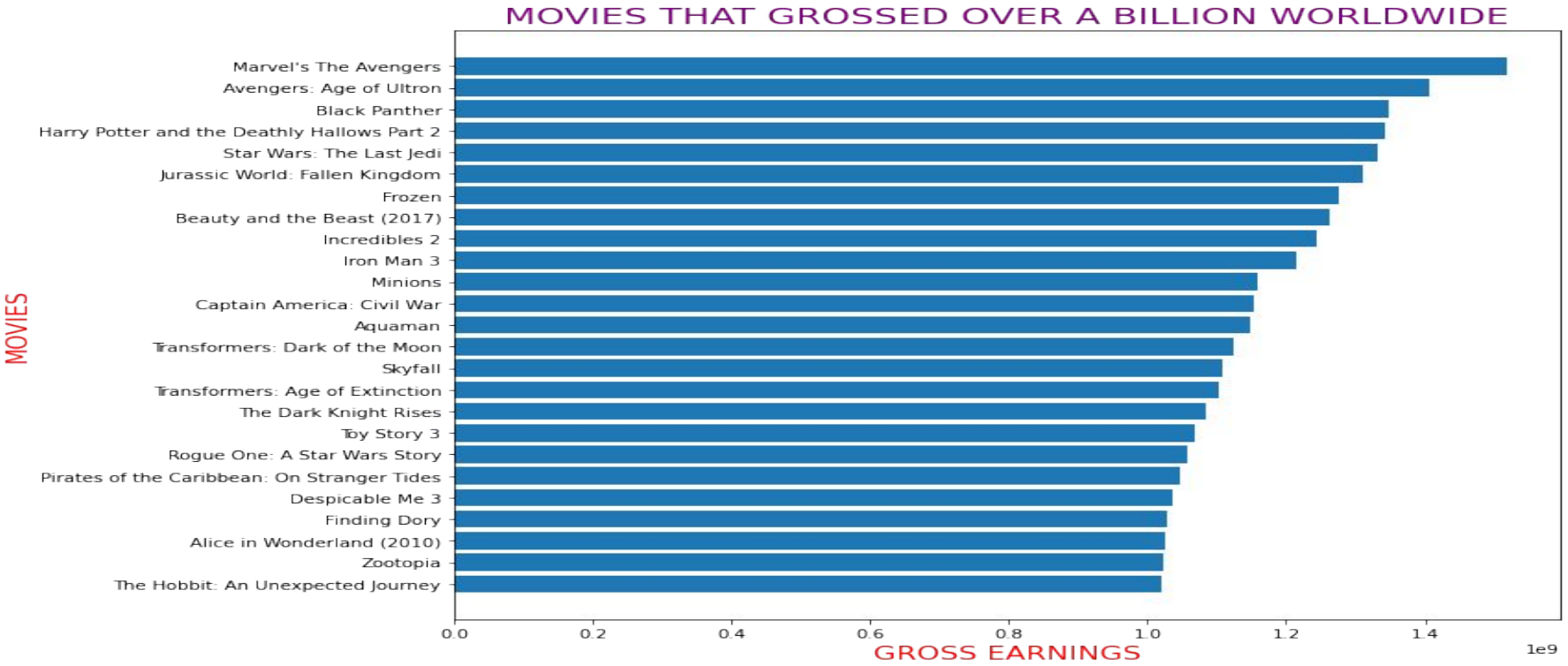
2.bom.movie_gross Dataset

As this file was in zipped csv for format, I unzipped it first and then turned it into a pandas DataFrame for better understanding and analysis.

I checked for duplicates but they were none. I turned the foreign_goross column which was in object data type to integer type so that i can add it to domestic_gross column to produce new column name combined gross which would give us how the earnings looked in a worldwide perspective.

The following graph in the next slide tells us about how individual movie performed on the box office worldwide.

Graph of movies against their worldwide earnings.



Explain the above graph:

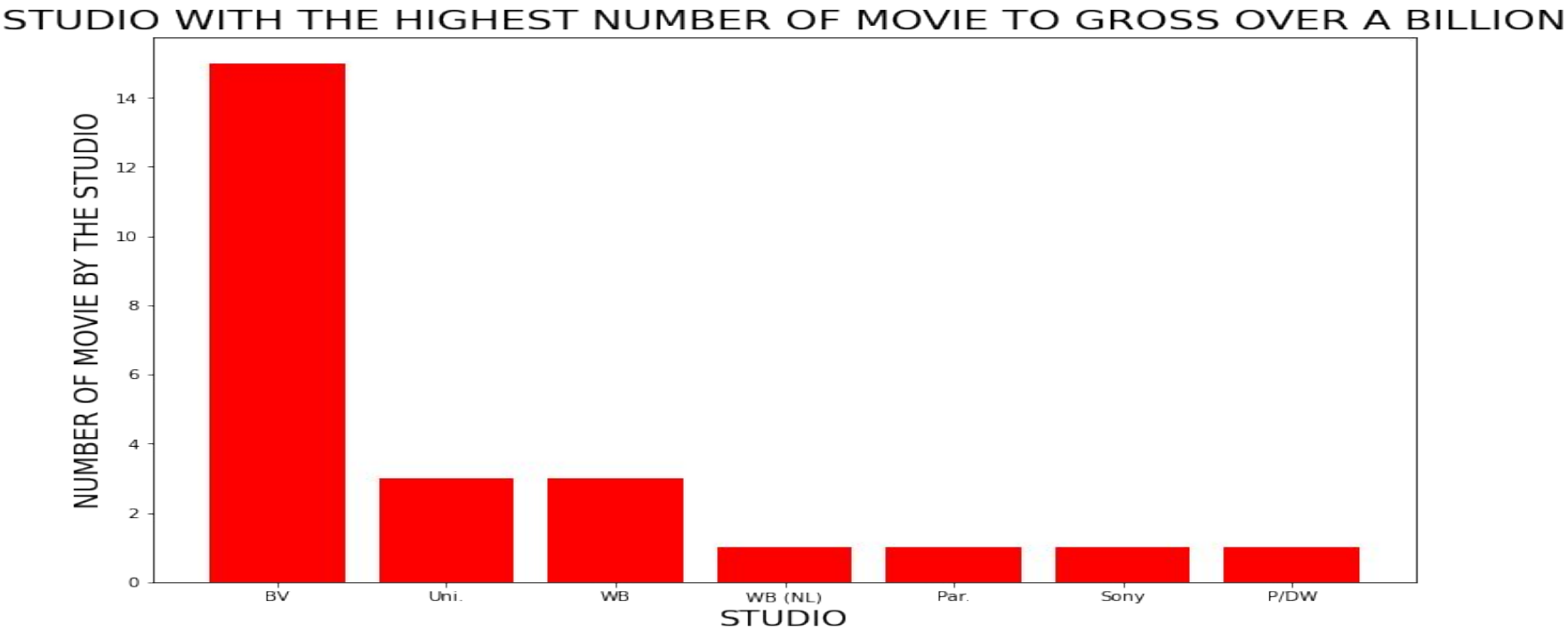
Movies that were able to perform in the box office worldwide resulting in over a Billion Dollar earnings followed a predictive pattern.

The pattern being an adaptation of comic book or novel mainly science fiction that were able to franchise over time which hooked people leading to the unbelievable numbers in the box office such as Avengers.

The other pattern being animated movies which resonated with both children and adults such as frozen and zootopia.

The other pattern being remakes of popular movies that existed since before and live action adaptation of animated movies.

Graph showing the best performing movies in box office:



Studios that directed the most performing movies.

The main studios which would give us a tough competition in the future if this business idea goes through are the followings:

1. BV
2. Universal
3. Warner Bros
4. Paramount
5. Sony

Recommendations

I would like to recommend to you to employ the list of personalities i provided in the seventh slide and consider possible collaboration with the studios.

I would also recommend to adapt comics or novels as they are more popular and pull more people in the theatres.

I would also recommend to start a streaming business as people nowadays stream more at the their homes in their comforts rather than going outside.

Thank you for choosing and entrusting me with this project.