Movie Analysis

Phase 1 Project

Summary

With Microsoft as a leader within the technology space, the organization would like to embark in a new venture--this time into the big screen. This project focuses on gathering pertinent data to create an analysis on which direction the tech giant should focus on for their first major film.

Outline

- Business Problem
- Data
- Methods
- Results
- Conclusions

Business Problem

1. It's not about reinventing the wheel, but who can we model to replicate success in the first production.
2. Are online movie database ratings relevant to potential gross earning success.
3. What direction should the movie studio take to replicate a profitable lifetime gross earning success.

Data

API

- The Movie DB, we attempted to to do an API call, which was successful. This API was our initial dataset, however we decided to move towards a different direction.
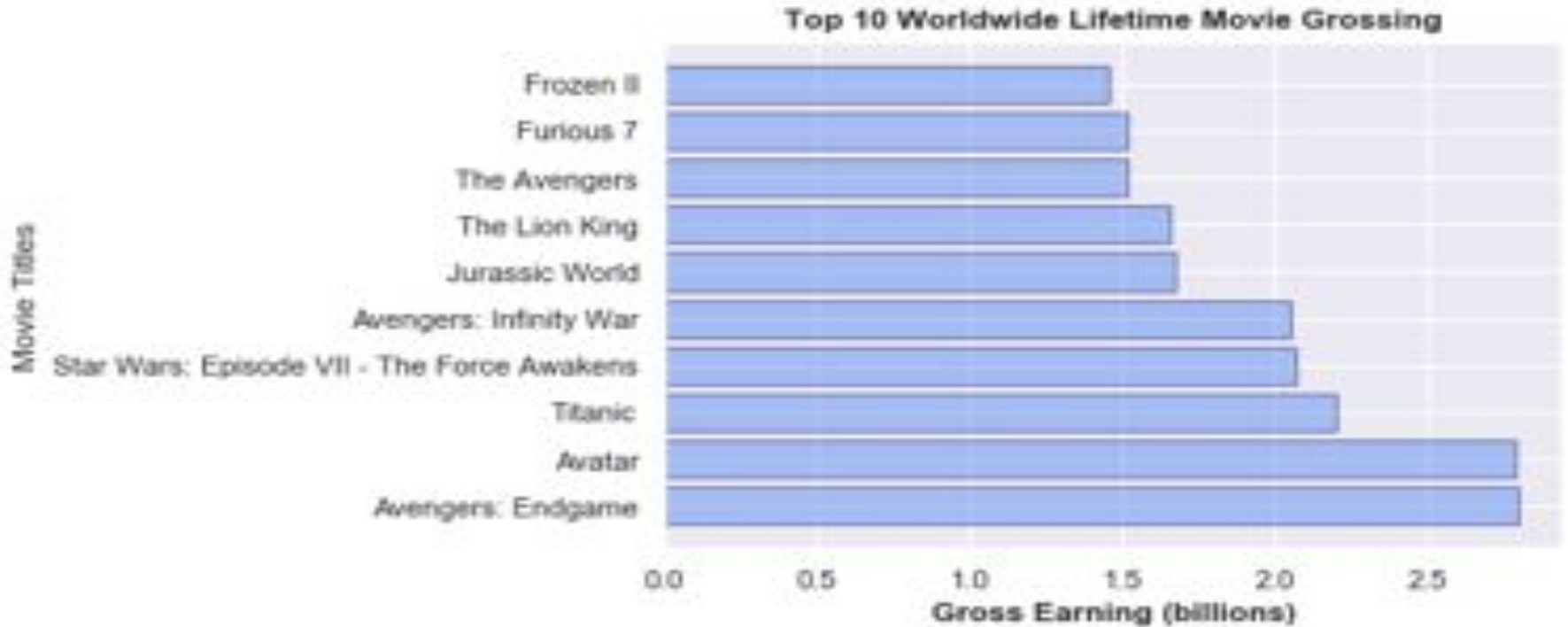
Web Scraping

- IMDB's Top 1000 movie list the following characteristics were used: **movie, year, time_minute, genre, imdb_rating, metascore, vote, and gross_earning**.
- Box Office Mojo Lifetime High Gross Earning Movie list of characteristics: **worldwide, domestic, and foreign**

Methods

- Finding the most successful movies from a global and domestic perspective.
- Finding the features of the top worldwide grossing movie and creating a dataframe that shows similar character attributes.
- Finding relationships based on character attributes.
- Coming up with inferences based on visualizations and select attributes within the data.

Box Office Mojo

Top 10 Worldwide Lifetime Movie Grossing

Movie Titles	Gross Earning (billions)
Frozen II	~1.45
Furious 7	~1.5
The Avengers	~1.5
The Lion King	~1.65
Jurassic World	~1.67
Avengers: Infinity War	~2.05
Star Wars: Episode VII - The Force Awakens	~2.07
Titanic	~2.2
Avatar	~2.8
Avengers: Endgame	~2.8

Box Office Mojo

Gross Domestic Contribution by Year

IMDB Analysis

```
#Checking summary stats
IMDB_df.describe()
```

	Unnamed: 0	year	time_minute	imdb_rating	metascore	vote	gross_earning
count	760.000000	760.000000	760.000000	760.000000	760.000000	7.600000e+02	760.000000
mean	420.088158	1994.986842	123.188158	7.931711	77.368421	3.349773e+05	75.128013
std	264.433162	19.517563	26.016821	0.289804	12.535002	3.436865e+05	112.765798
min	1.000000	1925.000000	66.000000	7.600000	28.000000	2.567700e+04	0.000000
25%	194.750000	1985.750000	104.000000	7.700000	70.000000	8.762125e+04	5.182500
50%	391.500000	2001.000000	120.000000	7.900000	78.000000	2.148220e+05	32.385000
75%	626.500000	2009.000000	136.000000	8.100000	86.000000	4.677370e+05	100.040000
max	985.000000	2020.000000	238.000000	9.300000	100.000000	2.304536e+06	936.660000

IMDB Analysis

```python
Movies_gross_earninsg_over75 = IMDB_df.loc[IMDB_df['gross_earning'] >= 75]
Movies_gross_earninsg_over75
```

	Unnamed: 0	movie	year	time_minute	genre	imdb_rating	metascore	vote	gross_earning
2	3	The Untouchables	1987	119	Crime, Drama, Thriller	7.9	79.0	276756	76.27
3	4	The Nightmare Before Christmas	1993	76	Animation, Family, Fantasy	8.0	82.0	293284	75.08
4	6	Harry Potter and the Sorcerer's Stone	2001	152	Adventure, Family, Fantasy	7.6	64.0	640187	317.58
7	9	Knives Out	2019	130	Comedy, Crime, Drama	7.9	82.0	413379	165.36
10	12	The Hunt for Red October	1990	135	Action, Adventure, Thriller	7.6	58.0	177745	122.01
...	...	...	...	...	...	...	...	...	...
562	615	Fiddler on the Roof	1971	181	Drama, Family, Musical	8.0	67.0	38781	80.50
613	690	Ray	2004	152	Biography, Drama, Music	7.7	73.0	136992	75.33
629	714	Fantasia	1940	125	Animation, Family, Fantasy	7.7	96.0	87520	76.41
652	764	The Muppet Movie	1979	95	Adventure, Comedy, Family	7.6	74.0	32249	76.66
666	789	On Golden Pond	1981	109	Drama	7.6	68.0	27283	119.29

228 rows × 9 columns

IMDB Analysis

	Unnamed: 0	movie	year	time_minute	genre	imdb_rating	metascore	vote	gross_earning
7	9	Knives Out	2019	130	Comedy, Crime, Drama	7.9	82.0	413379	165.36
12	14	Once Upon a Time... In Hollywood	2019	161	Comedy, Drama	7.6	83.0	527209	142.50
13	15	Coco	2017	105	Animation, Adventure, Family	8.4	81.0	363684	209.73
16	18	1917	2019	119	Drama, War	8.3	78.0	387911	159.23
18	21	Joker	2019	122	Crime, Drama, Thriller	8.5	59.0	892851	335.45
...	...	...	...	...	...	...	...	...	...
446	469	The Fighter	2010	116	Biography, Drama, Sport	7.8	79.0	336621	93.62
468	496	The Lego Movie	2014	100	Animation, Action, Adventure	7.7	83.0	321281	257.76
482	514	Straight Outta Compton	2015	147	Biography, Drama, History	7.8	72.0	177405	161.20
509	549	How to Train Your Dragon 2	2014	102	Animation, Action, Adventure	7.8	76.0	301289	177.00
558	610	Dawn of the Planet of the Apes	2014	130	Action, Adventure, Drama	7.6	79.0	408654	208.55

84 rows × 9 columns

The ideal Movie

	Unnamed: 0	movie	year	time_minute	genre	imdb_rating	metascore	vote	gross_earning
19	22	Avengers: Endgame	2019	181	Action, Adventure, Drama	8.4	78.0	778254	858.37

Average Gross Earning for IMDB Ratings
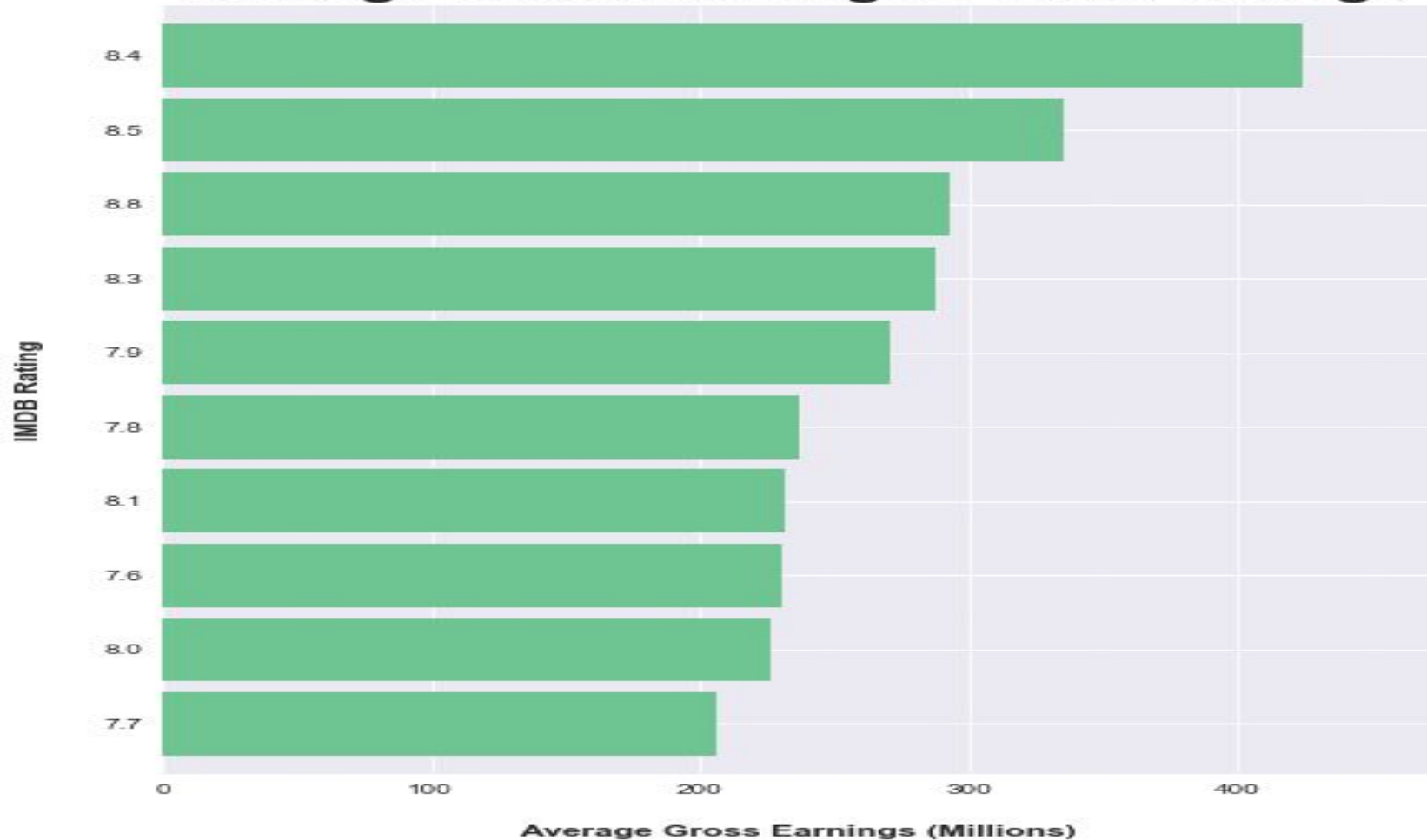
Relationship Between Voters to Gross Earning
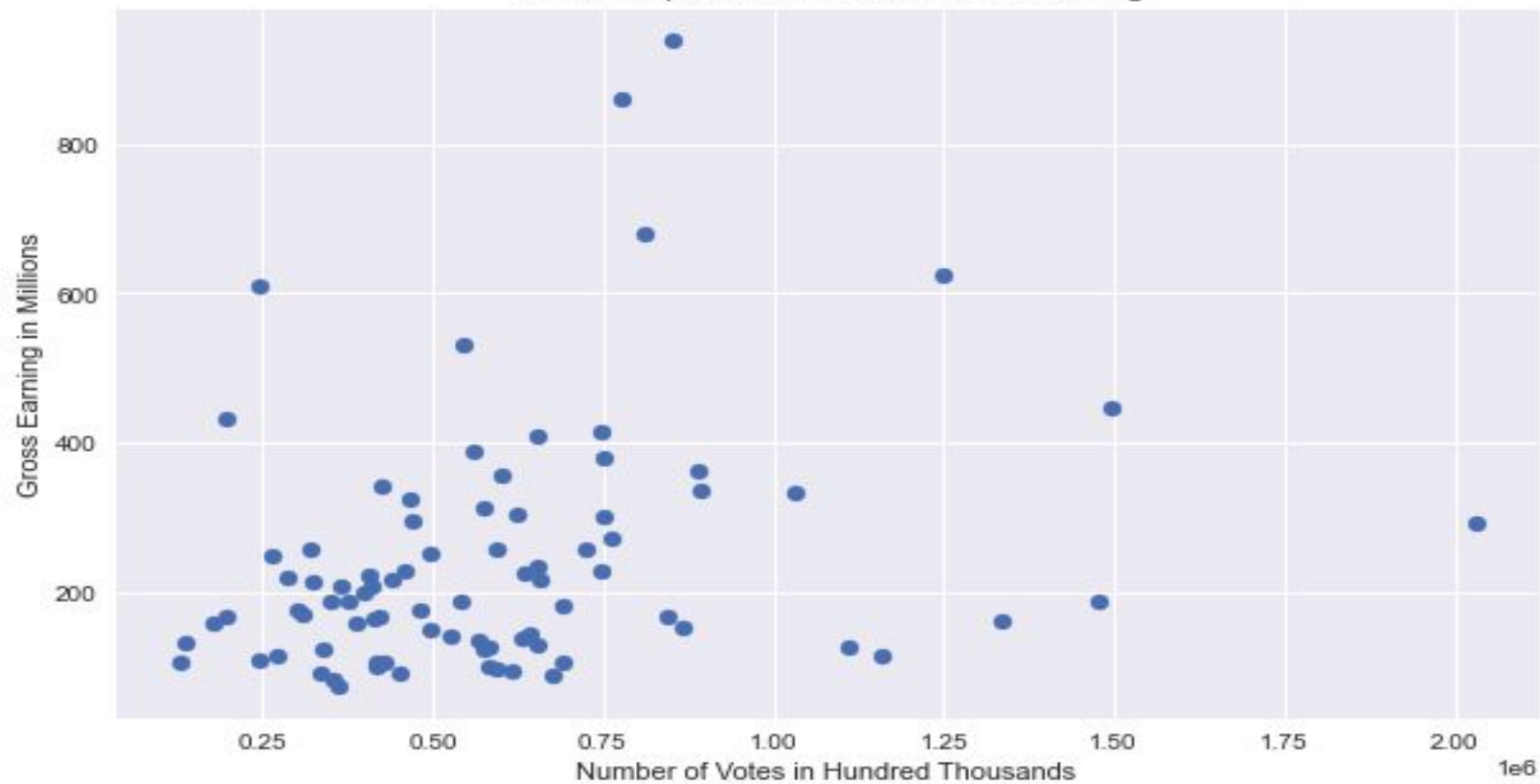
Top 5 Genres to Gross Earnings

Horizontal bar chart. Y-axis labeled "Genre", X-axis labeled "Average Gross Earnings (Millions)".

Genre	Average Gross Earnings (Millions)
Action, Adventure	~450
Action, Adventure, Drama	~420
Action, Adventure, Sci Fi	~385
Crime, Drama, Thriller	~335
Animation, Adventure, Comedy	~310
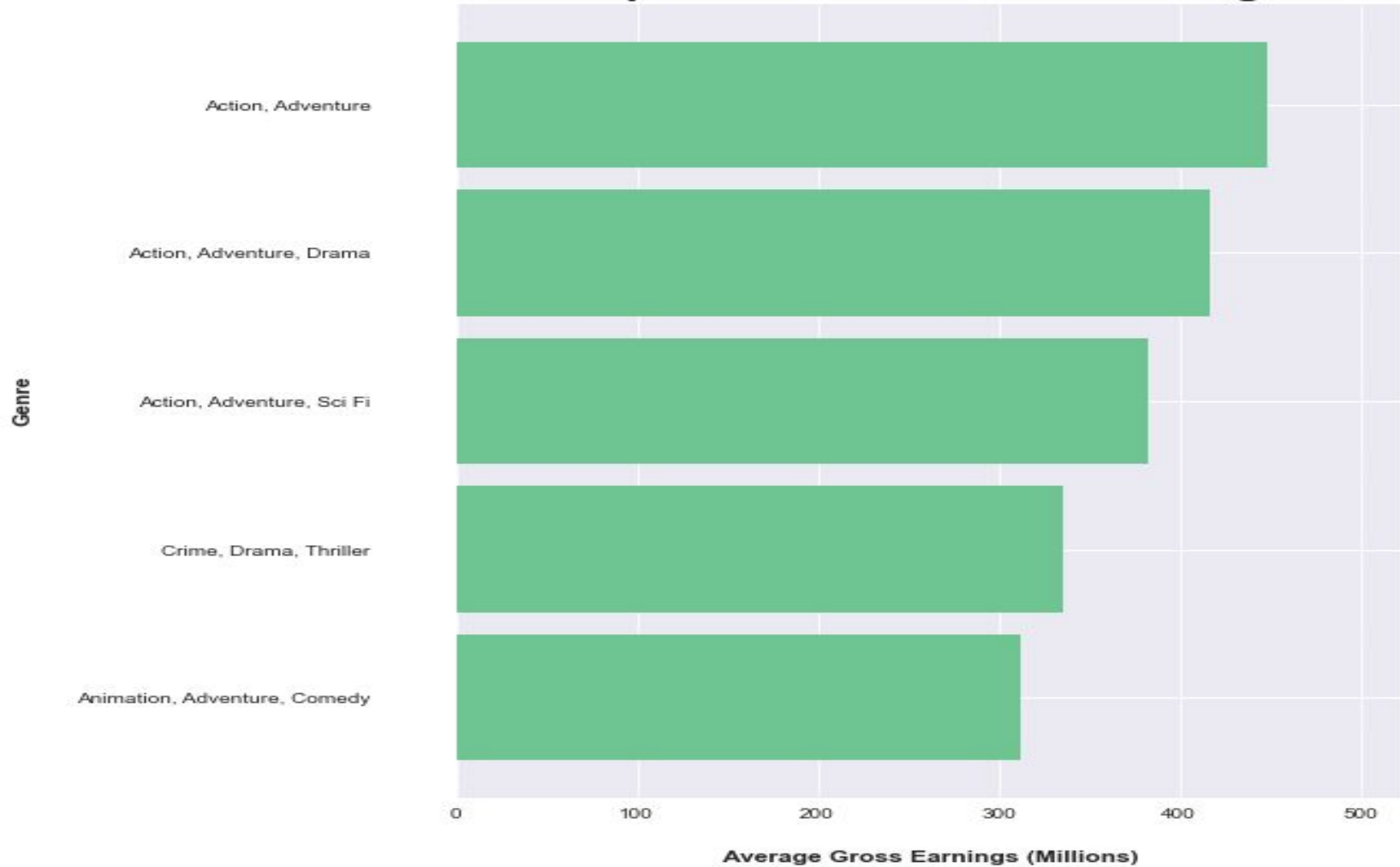
Average IMDB Rating for Genre

Genre	Average IMDB Rating
Crime, Drama, Thriller	
Action, Adventure	
Animation, Adventure, Family	
Adventure, Drama, Sci Fi	
Drama, War	
Biography, Comedy, Drama	
Biography, Crime, Drama	
Mystery, Thriller	
Drama, Mystery, Thriller	
Action, Drama, Sci Fi	
Action, Biography, Drama	
Comedy, Drama, Music	
Drama, Western	
Biography, Drama, Music	
Adventure, Drama, Fantasy	
Action, Drama, Mystery	
Drama	
Drama, Family	
Action, Adventure, Drama	
Drama, Thriller	
Action, Adventure, Sci Fi	
Action, Drama, History	
Drama, Mystery, Sci Fi	
Comedy, Crime, Drama	
Animation, Action, Adventure	
Animation, Adventure, Comedy	
Biography, Drama, History	
Biography, Drama, Thriller	
Action, Adventure, Comedy	
Adventure, Fantasy	
Adventure, Biography, Crime	
Crime, Drama, Mystery	
Drama, Romance	
Biography, Drama	
Adventure, Family, Fantasy	
Horror, Mystery, Thriller	
Drama, Sci Fi, Thriller	
Comedy, Drama, Romance	
Action, Adventure, Thriller	
Biography, Drama, Sport	
Drama, Sport	
Drama, Music, Romance	
Comedy, Drama	
Action, Crime, Music	
Animation, Comedy, Crime	

# Conclusions

1. Who can we mimic?
    a. Modeling the success of **Avengers: Endgame** would be the the most ideal route in in terms of box office success and rating.
2. What about the ratings?
    a. We did find that ratings to gross earning do share a relationship, but the number of votes do not necessarily contribute to a movies overall lifetime gross earnings
3. How are we going to mimic success.
    a. We found that a specific genres (Action and Adventure) are both profitable and popular, with an aim for an IMDB rating over 7.8

Next Steps

1. Further considerations in optimizing data and models.
   a. Financial conversations (budget and potential expenses)
   b. Casting recommendations
   c. Director recommendations
   d. Box office release locations (foreign vs domestic)
   e. Projecting earnings
   f. See next for preview for Relationship Between Movie Runtime to Gross Earning

Relationship Between Movie Runtime to Gross Earning
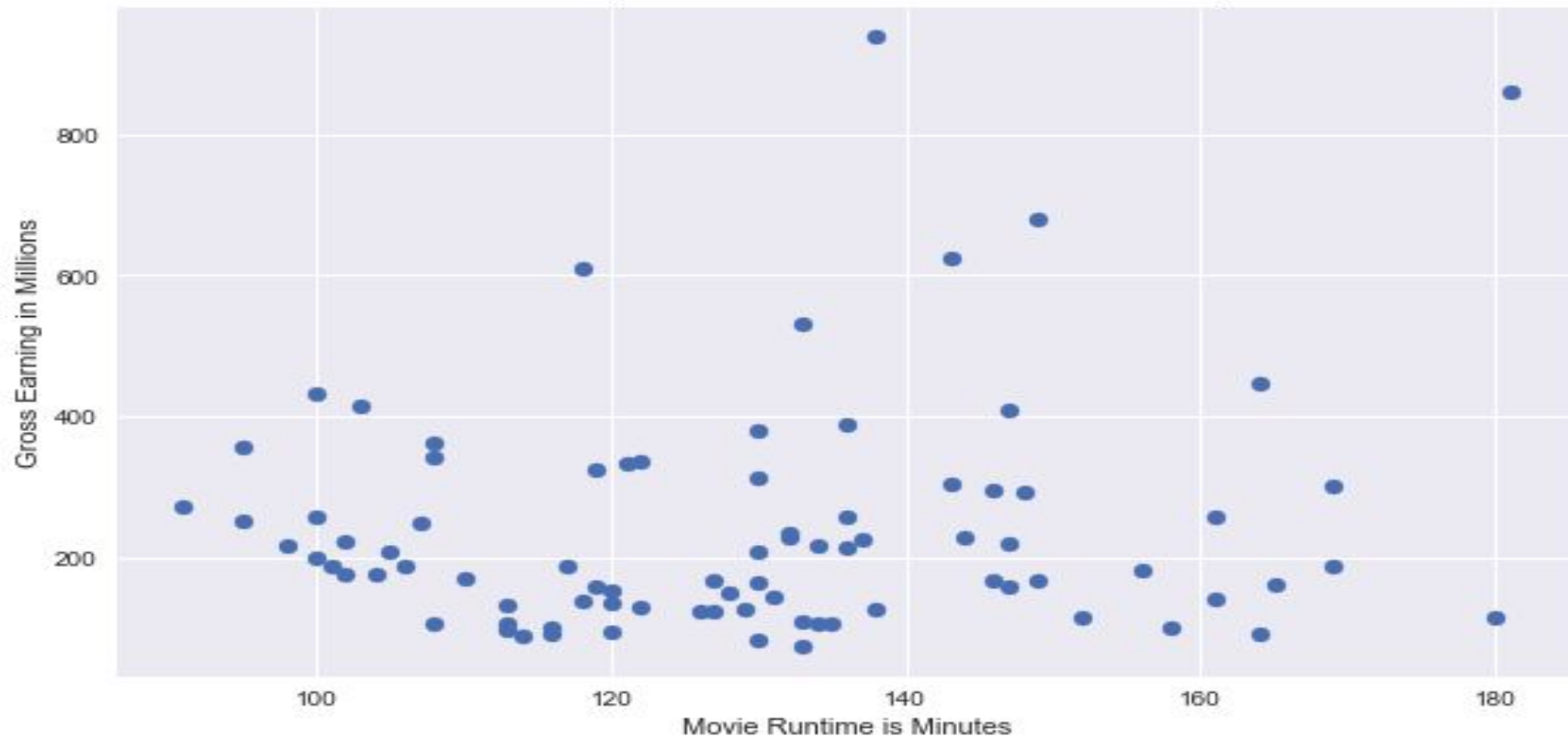
Thank You!

Christian Corrales
Ijeoma Akamnonu

GitHub: https://github.com/ccorrales1228/Phase_1_Project