

# PROJECT GOALS

The EDA process was directed by the following questions:

1. What are the highest grossing film genres Microsoft can invest in?
2. What are the best combinations (if any) of the highest grossing films by genre
3. Who are the best directors for that job?
4. BONUS: An animated visualization of trends of movies since 1980

# DATA

1. SQL Database: There is a im.db relational database. The SQL file was provided in the instructions. It contains a comprehensive collection of movie-related data, including information on genres, directors, and more.
2. TSV Files: Two TSV (Tab-Separated Values) files were used, presumably containing supplementary data for the analysis. One has reviews
3. CSV Files: Two CSV (Comma-Separated Values) files were included, likely containing additional information about movies. One of the files was downloaded from Kaggle.  
<https://www.kaggle.com/danielgrijalvas/movies>

# EDA PROCEDURES

- Importing all necessary packages i.e
  - numpy,
  - pandas,
  - sqlite3,
  - matplotlib.pyplot & pyplot
- Loading the data into dataframes
- Cleaning the data
  - This process entailed
  - Handling missing values
  - Handling duplicates
  - Data transformation as needed (to strings or floats where applicable)
  - Outlier handling i.e. one dataset contained data from year 2115

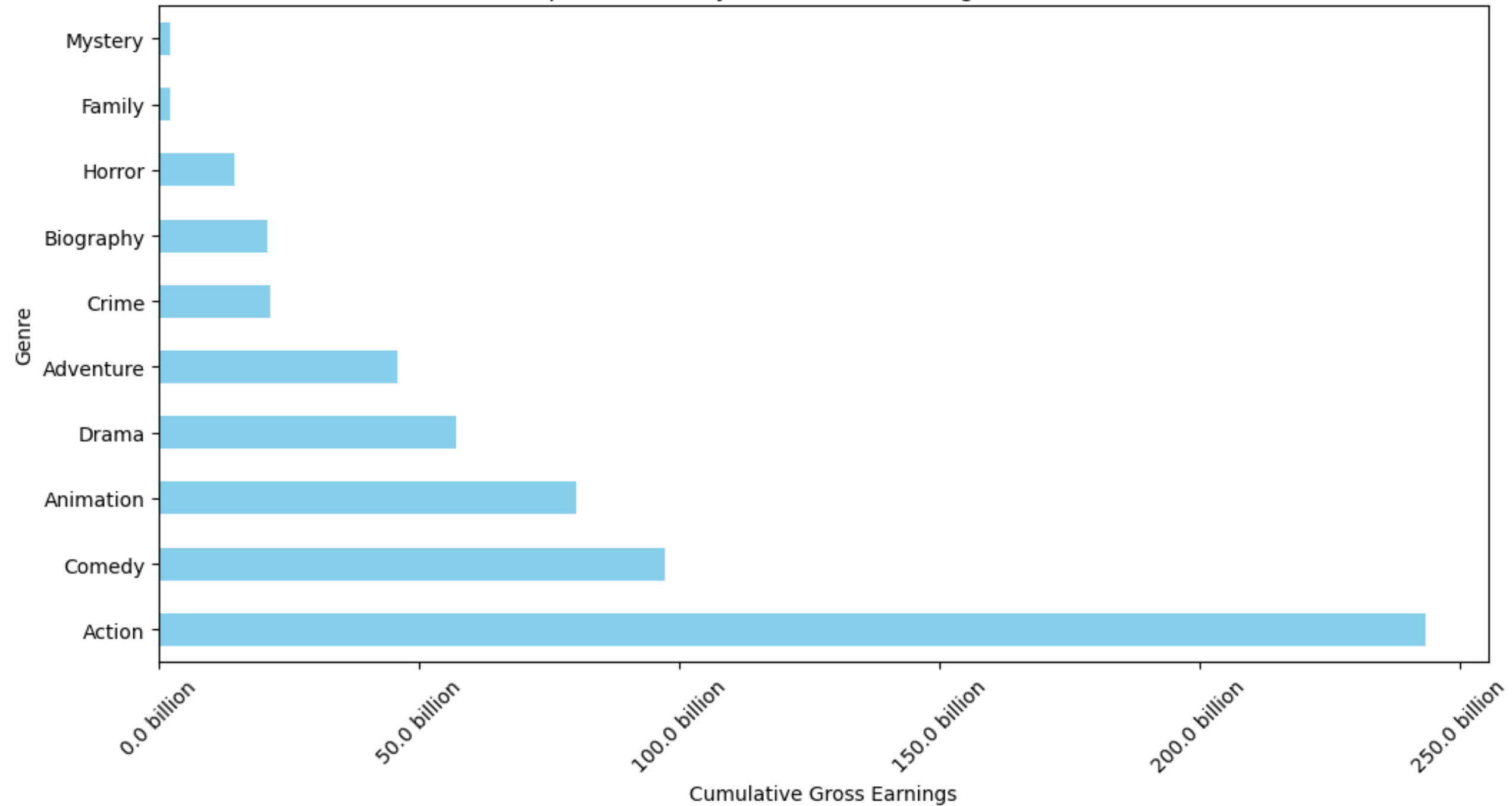
# EDA METHODS

- Added relevant columns as needed
  - New columns were created to aid in analysis. For example, the to calculate profit, you create a new column from gross profit and budget
  - Data cleaning following this process since the created column of profit contained negative figures. The negative values were converted to 0
- Combined DataFrames
  - A mask was applied to one DataFrame i.e sql\_df and filtered then combined with data from movies\_df. Movies DF was inner joined. The aim was to calculate the gross revenue per genre. One DF contained most genres and they were inner joined with movies\_df which had the earnings data

# RESULTS – Question 1

- What are the highest grossing film genres Microsoft can invest in?
  - The top 3 highest grossing genres are Action, Comedy and Animation
  - Most Movies do not fall into a single genre. They have multiple genres.

Top 10 Genres by Mean Gross Earnings since 1980

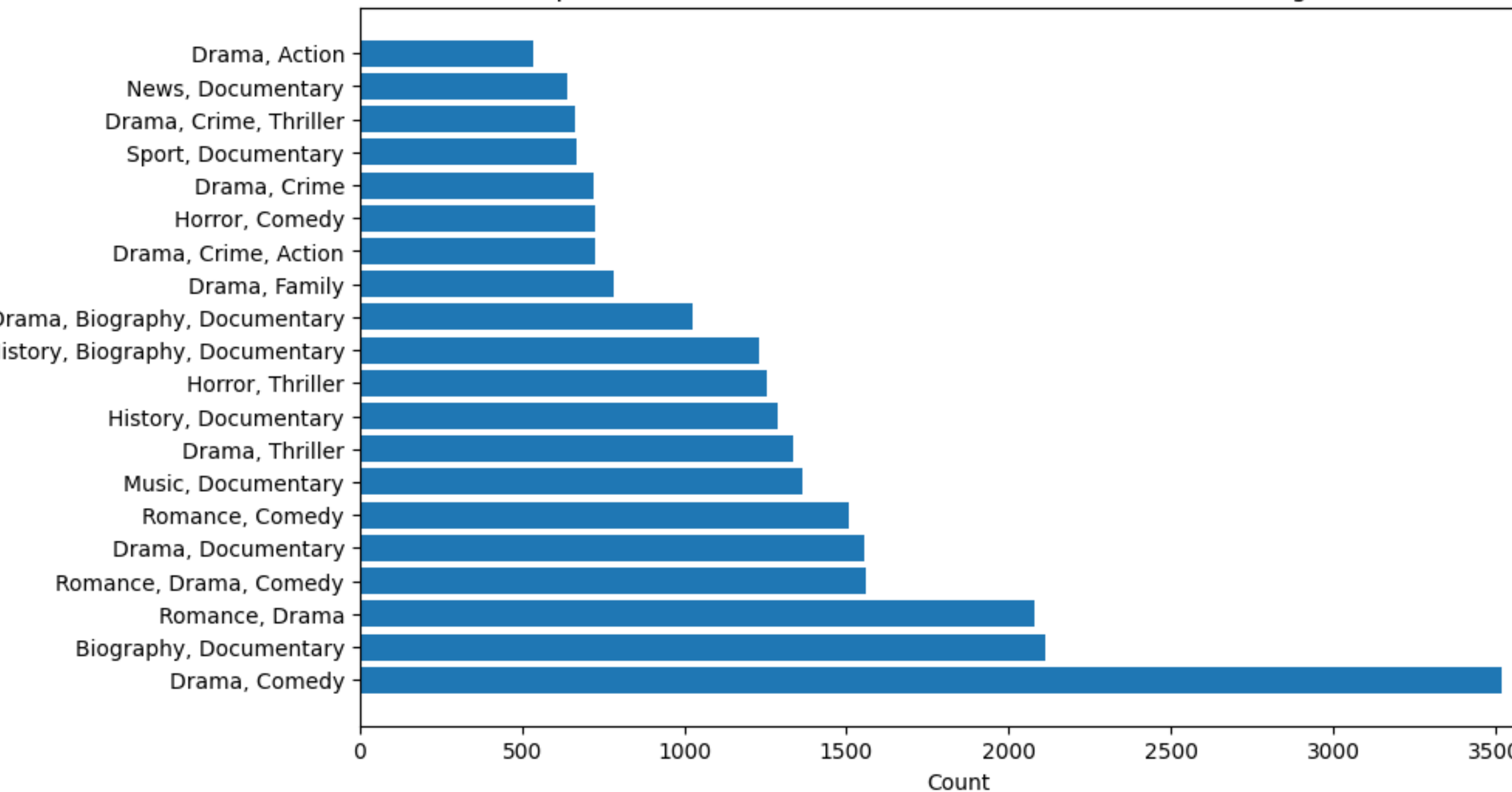


# RESULTS: Question 2

- What are the best combinations (if any) of the highest grossing films by genre
  - The most common genre combinations are Comedy and Drama
  - So we'd advise Microsoft to create a movie on comedy and or drama



Top 20 Most Common Genre Combinations (with 2 or more genres)



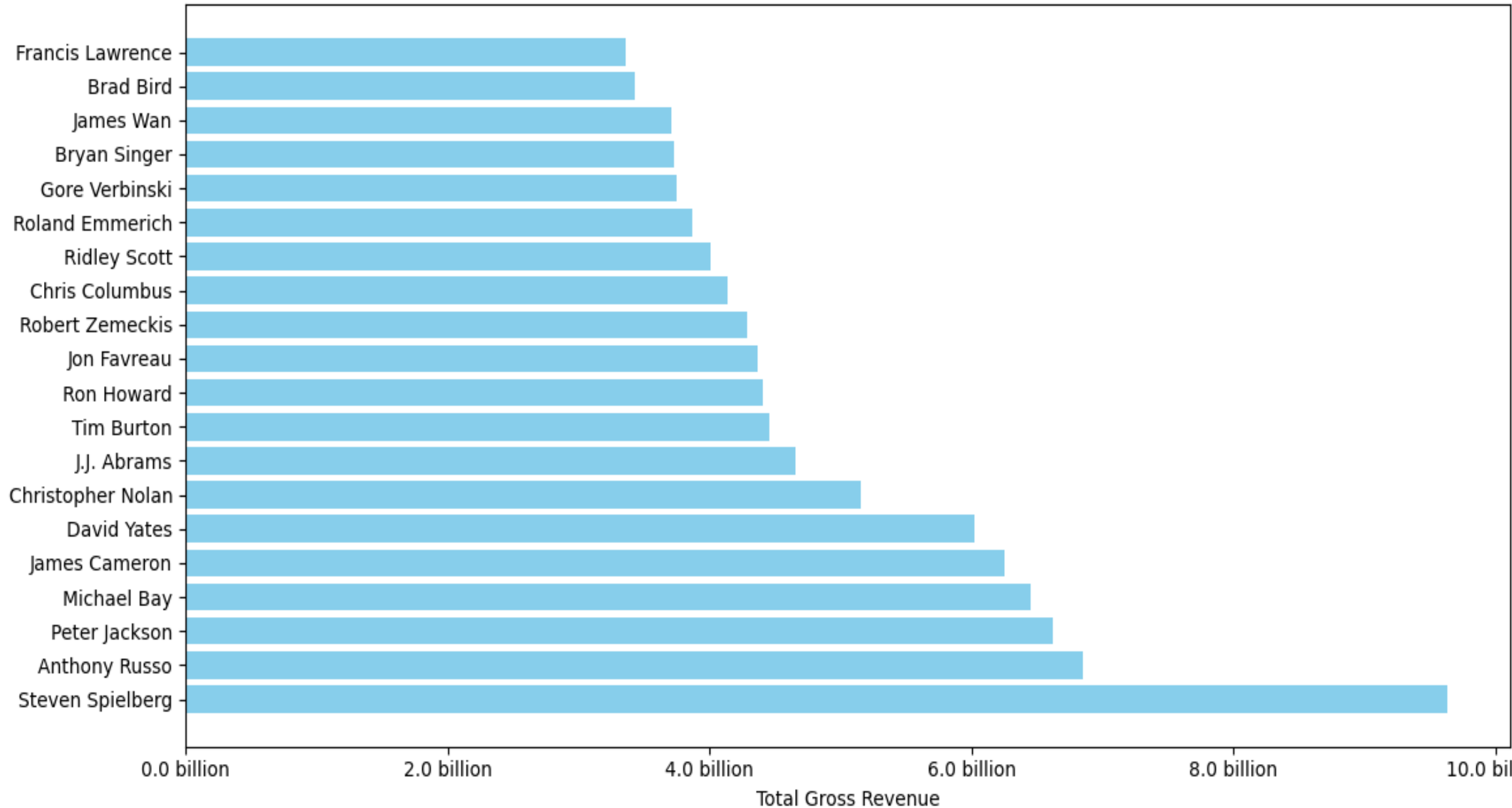
# RESULTS – Question 3

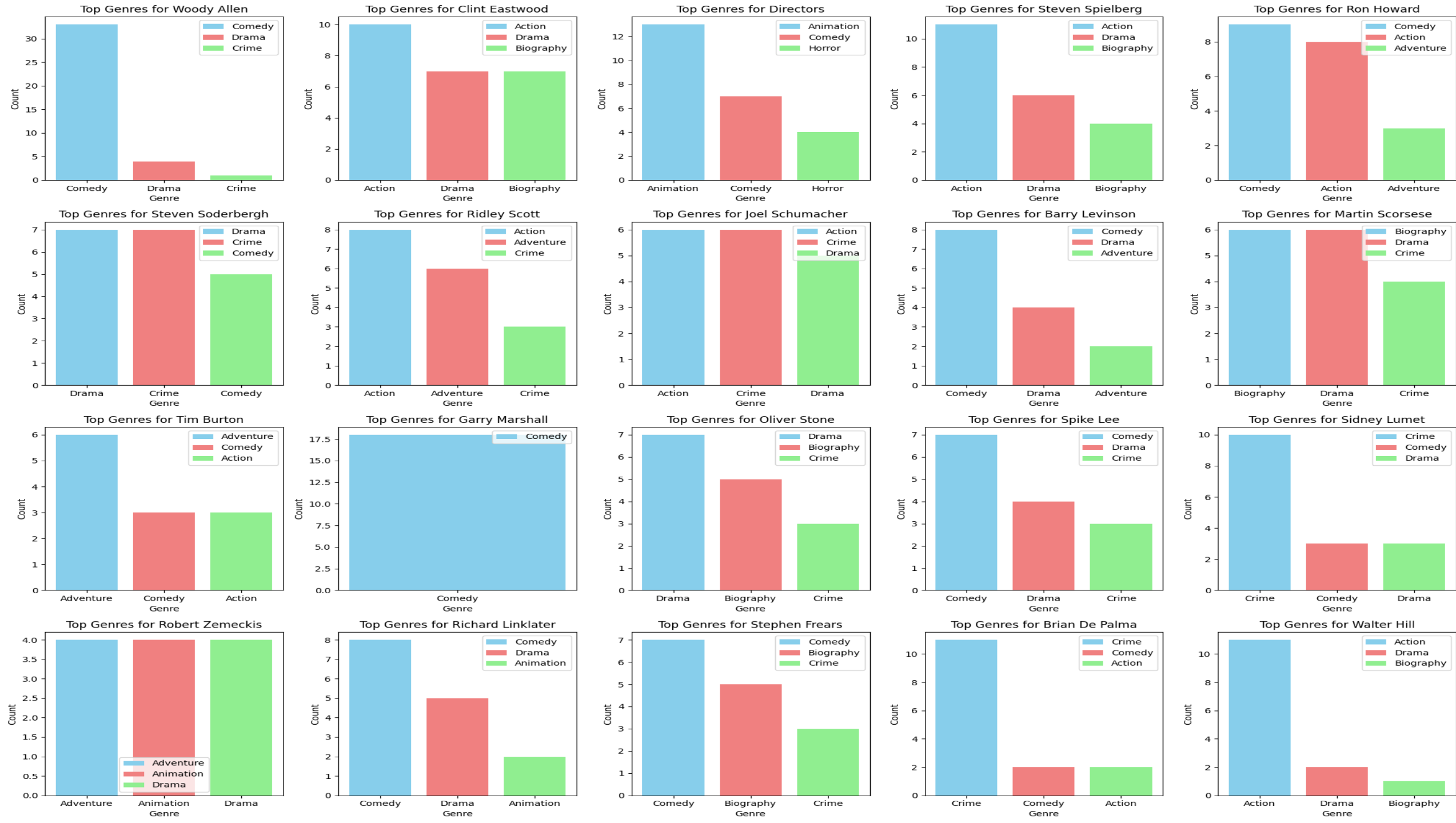
- Who are the best directors for that job?
- Since we understand that the best genres are Animation, Comedy and Drama,
- The directors with highest experience (by number of movies) and most successful by their gross income in these areas are:
  - Comedy - Garry Marshall, Woody Allen
  - Drama - Barry Levinson
  - Animation – ‘Directors’
    - NB: Under animation, there are assorted directors who have created the best animated movies. But if it is a name, Robert Zemeckis and Richard Linklater. Each has 4 and 2 animated movies respectively

-

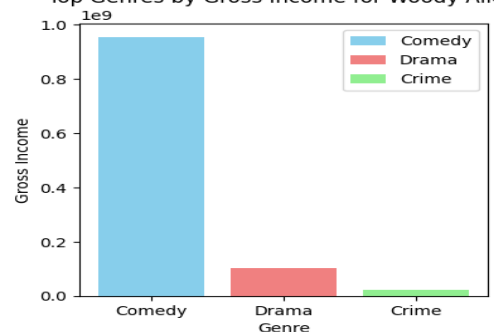
Top 10 Directors by Gross Revenue

Director

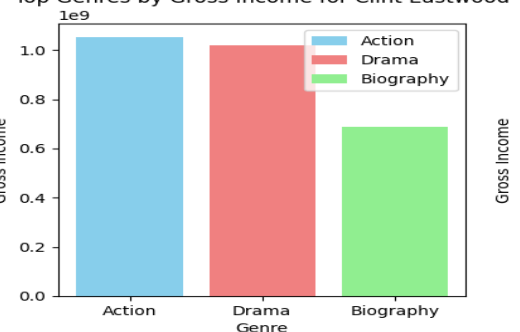




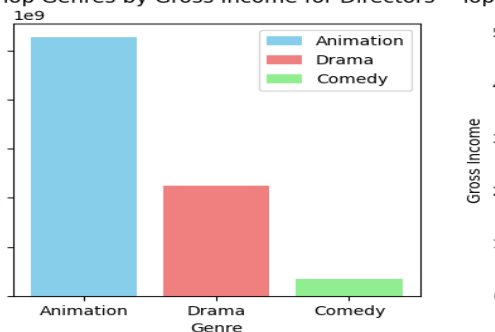
Top Genres by Gross Income for Woody Allen



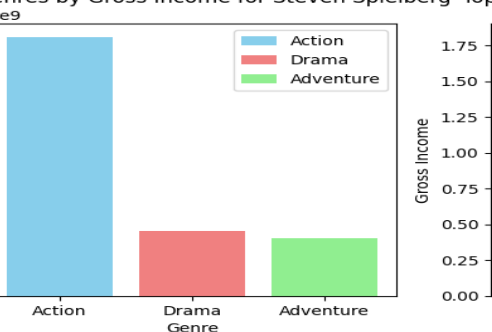
Top Genres by Gross Income for Clint Eastwood



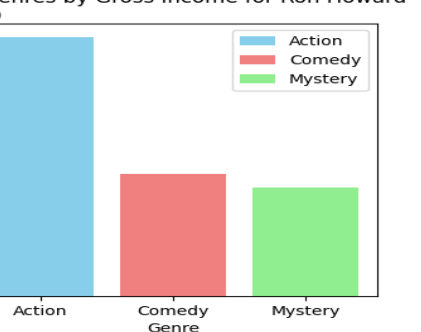
Top Genres by Gross Income for Directors



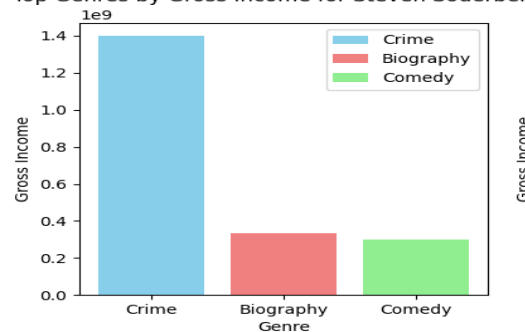
Top Genres by Gross Income for Steven Spielberg



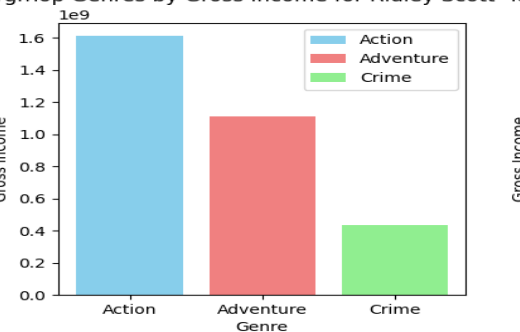
Top Genres by Gross Income for Ron Howard



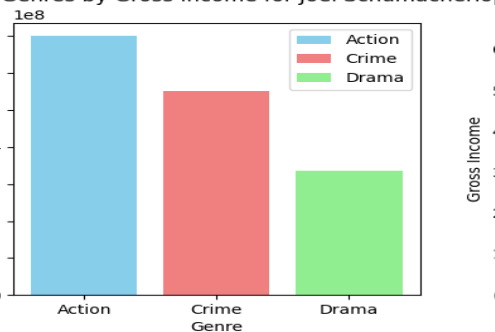
Top Genres by Gross Income for Steven Soderbergh



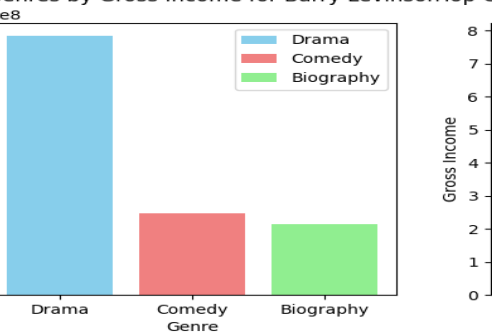
Top Genres by Gross Income for Ridley Scott



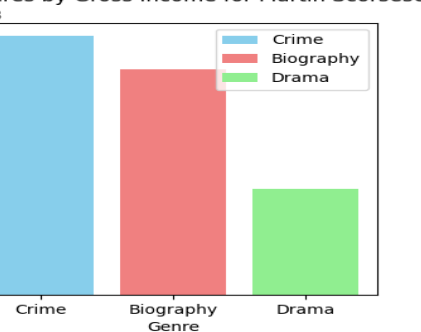
Top Genres by Gross Income for Joel Schumacher



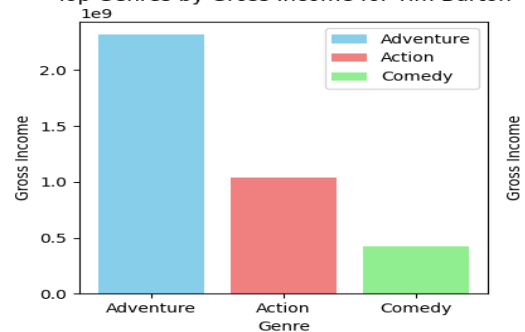
Top Genres by Gross Income for Barry Levinson



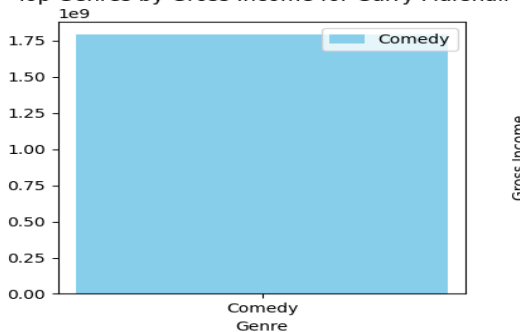
Top Genres by Gross Income for Martin Scorsese



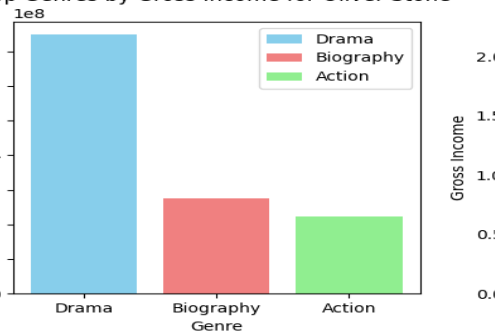
Top Genres by Gross Income for Tim Burton



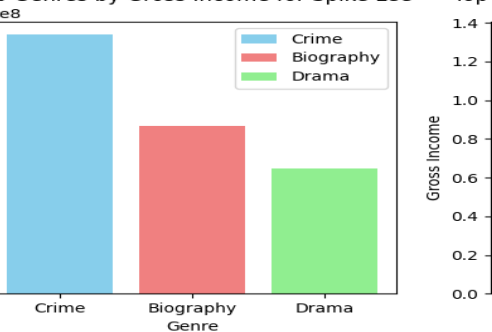
Top Genres by Gross Income for Garry Marshall



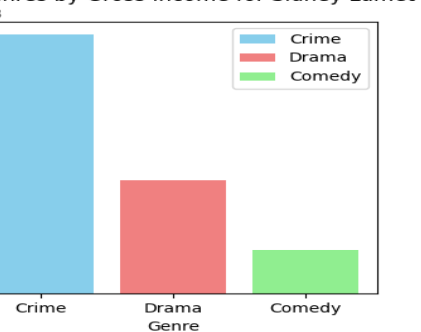
Top Genres by Gross Income for Oliver Stone



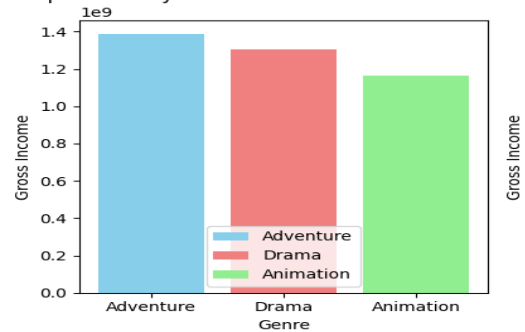
Top Genres by Gross Income for Spike Lee



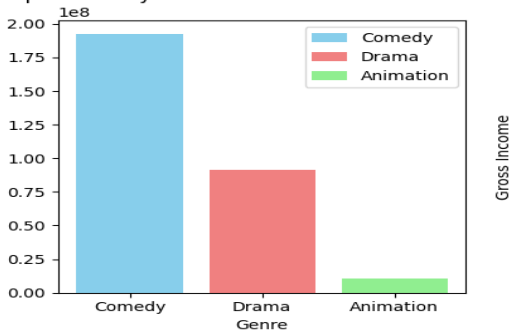
Top Genres by Gross Income for Sidney Lumet



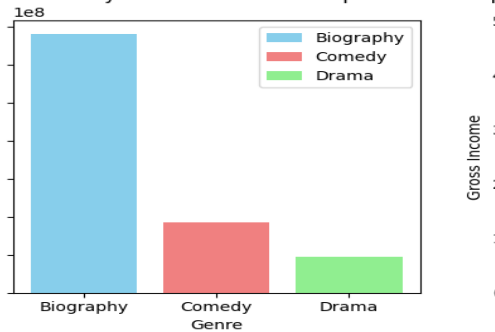
Top Genres by Gross Income for Robert Zemeckis



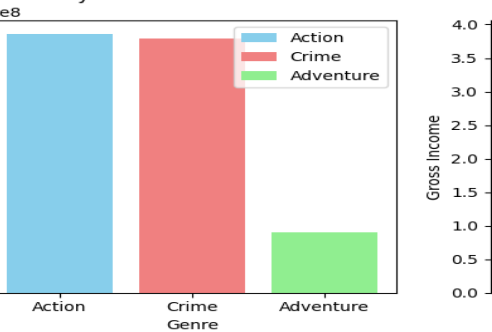
Top Genres by Gross Income for Richard Linklater



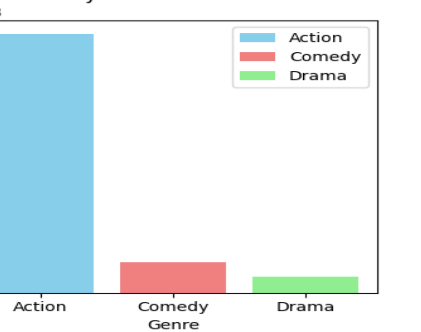
Top Genres by Gross Income for Stephen Frears



Top Genres by Gross Income for Brian De Palma



Top Genres by Gross Income for Walter Hill



# BONUS

- This animation shows that there has been an increase in Budget and Earnings from each movie since 1980s
  - There is a general trend in increase in cost of producing the movies
  - Animations and comedy seems to be some of the highest grossing movies since 2010
- Please Note that this is an animation from the .ipynb

