Exploratory Data Analysis on Movies

~ Ishita Patnaik

Agenda

- Introduction
- Data
- Assumptions
- Analysis
- Summary
- Actionable Insights

Introduction

- Producers Guild of America and Hollywood

- Challenges faced by new producers:
  - Difficult to understand the trends in the movie industry
  - Difficult to identify the right director and cast for the movie
  - Years of experience needed to release blockbuster

- Problem Statement:
  - Create analysis to help new producers answer questions like:
    - What genre of movies make the most money at the box office?
    - Which actors and directors make the most successful movies?
    - What are the factors that help make a movie blockbuster?

Data

Existing Features

Id	Feature	Description
01	Rank	Movie Rank
02	Title	Title of the movie
03	Genre	The various Genre that the movie can be associated with
04	Description	Short description about the movie
05	Director	Director of the movie
06	Actors	Main actors in the movie
07	Year	Year in which the movie was released
08	Runtime (minutes)	Total movie playing time
09	Rating	Movie rating
10	Votes	Votes for the movie
11	Revenue (Millions)	Revenue by the movie (in millions)
12	Metascore	Is the score of the movie on the metacritic website by critics

**Data is from years 2006 to 2016.

New Features

- Length (on Runtime)
  - Short: less than 2 hrs
  - Medium: 2 to 3 hrs
  - Long: greater than 3 hrs
- Quality(on Rating)
  - Excellent (8 or above)
  - Good (6 or above,less than 8)
  - Average (4 or above, less than 6)
  - Poor (0 or above, less than 4)
- Blockbuster(on Revenue)
  - True if Revenue > $221.78 million

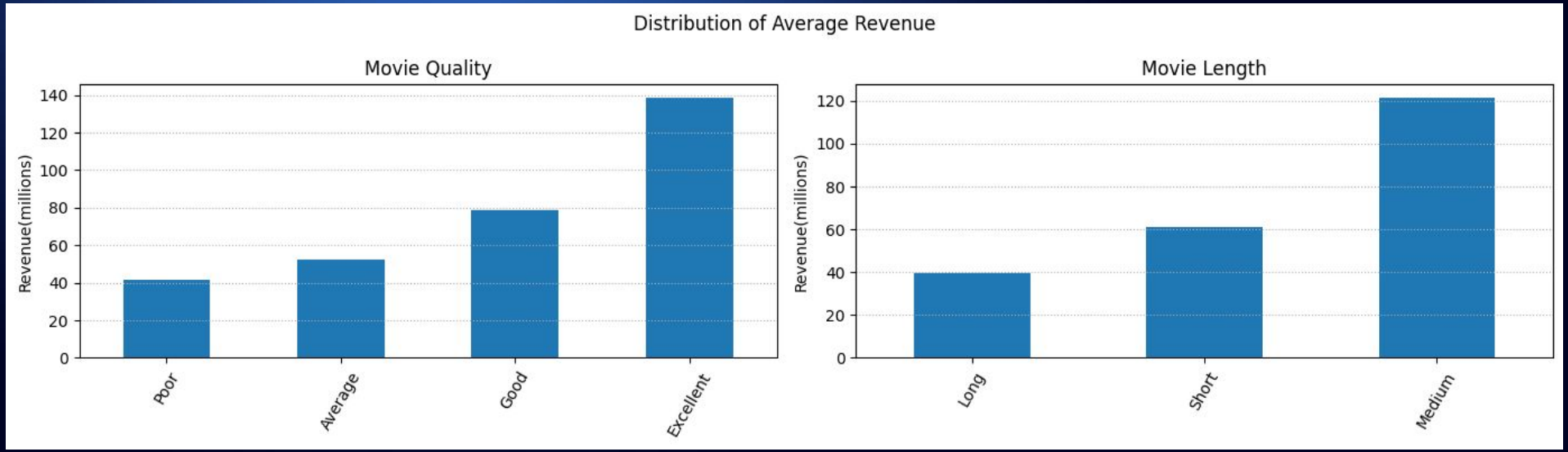
Assumptions

- Each genre and actor in the combination has contributed equally to the revenue of the movie.
- Order of genres and actors in combination doesn't matter.
- Number of votes is considered as viewership for the movie.
- Missing values are filled with central values for complete analysis.
- “Highest revenue” in all contexts refers to highest revenue per movie(on average).

Analysis

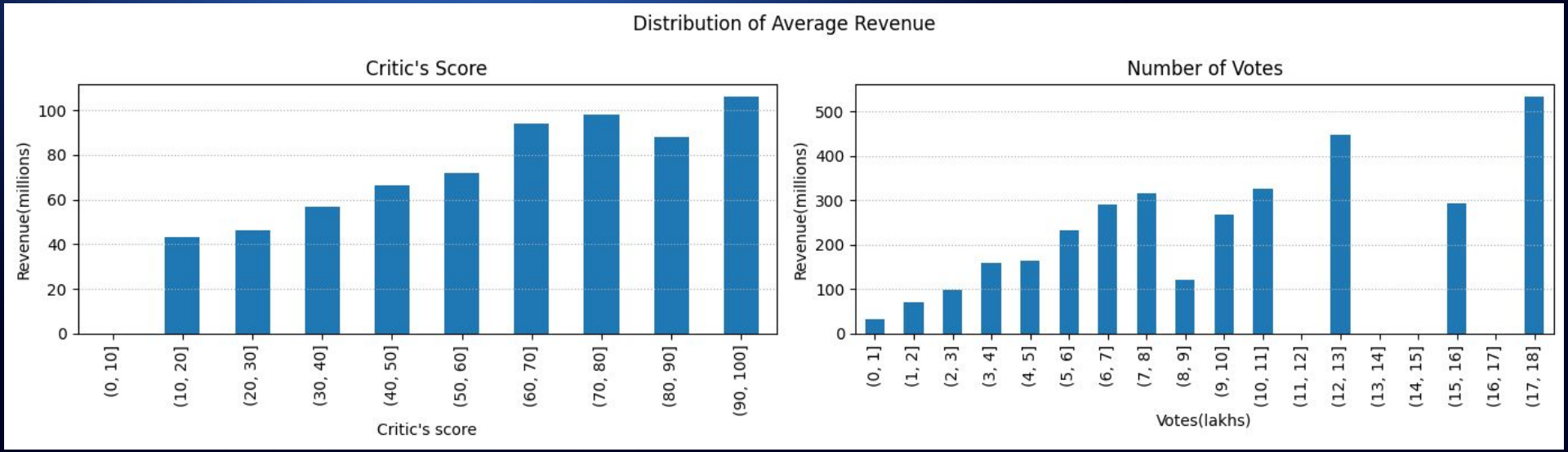
Revenue

What is the relationship between movie quality and movie length with revenue?



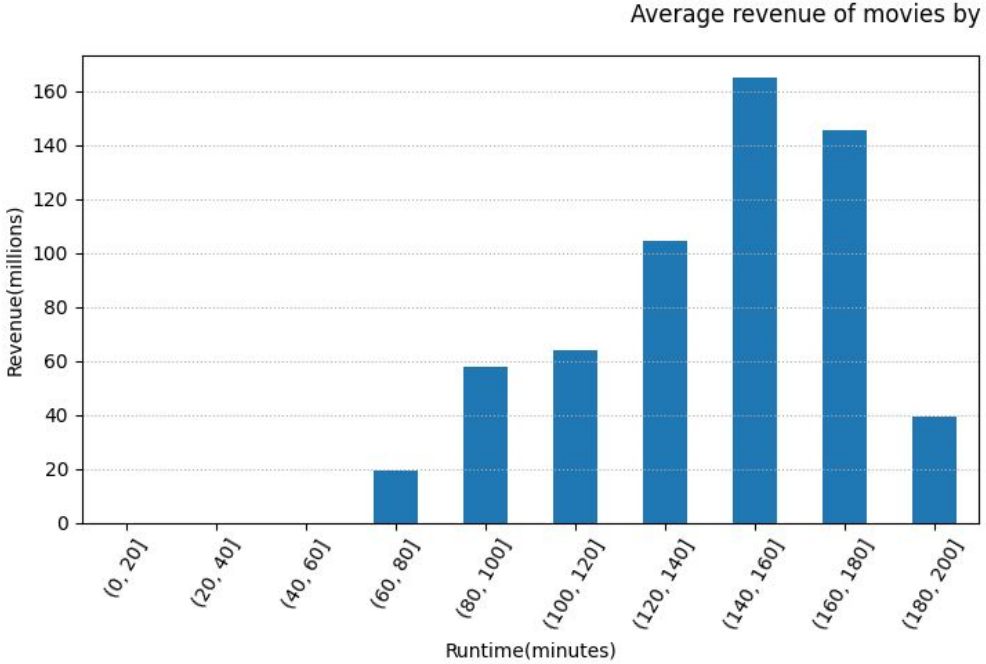
Revenue

What is the relationship between critic's scores and number of votes with revenue?



Runtime

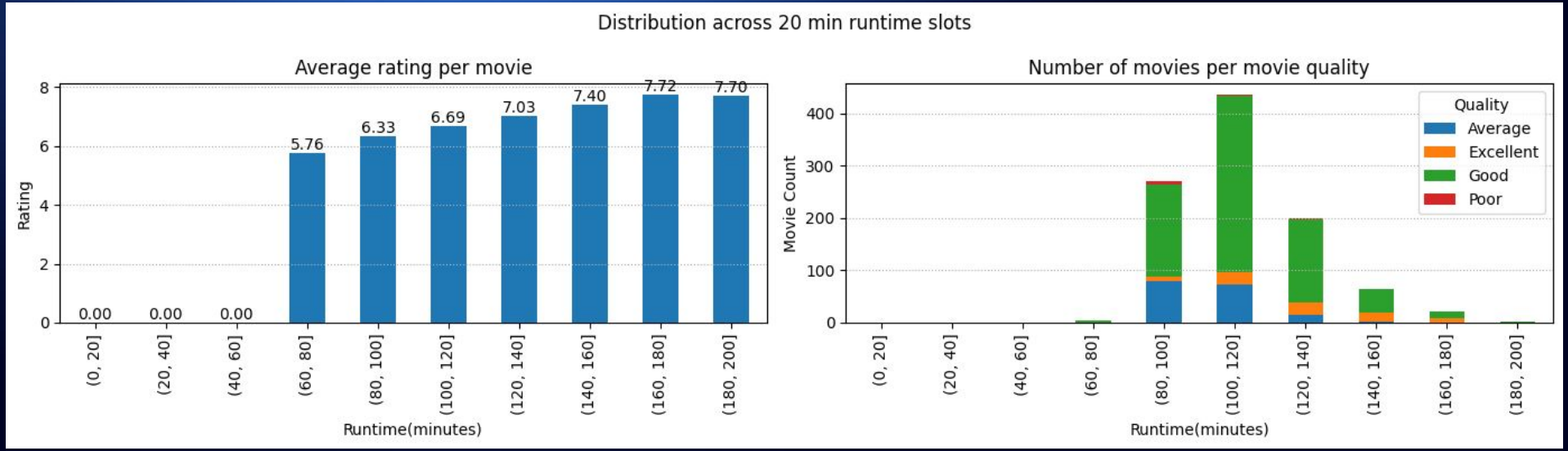
What is the runtime range where average revenue of movies is high?



Runtime(minutes)	Revenue(millions)	Movie Counts
(140, 160]	165.04	65
(160, 180]	145.81	22
(120, 140]	104.78	200
(100, 120]	63.88	435
(80, 100]	57.84	271
(180, 200]	39.58	2
(60, 80]	19.47	5
(0, 20]	nan	0
(20, 40]	nan	0
(40, 60]	nan	0

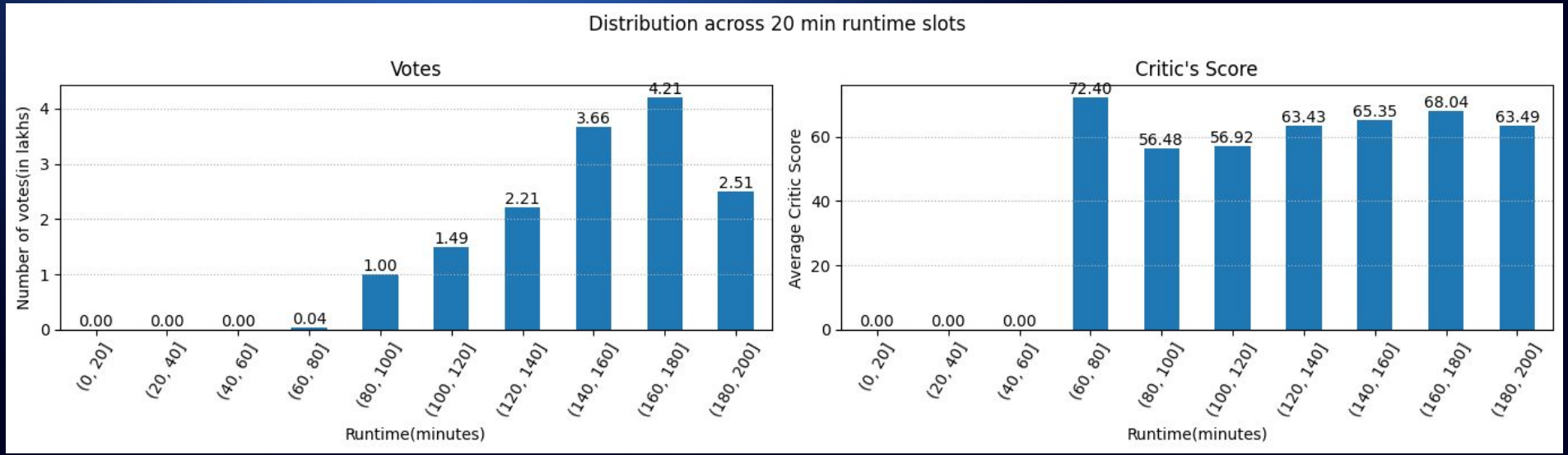
Runtime

How does the movie runtime affect rating?



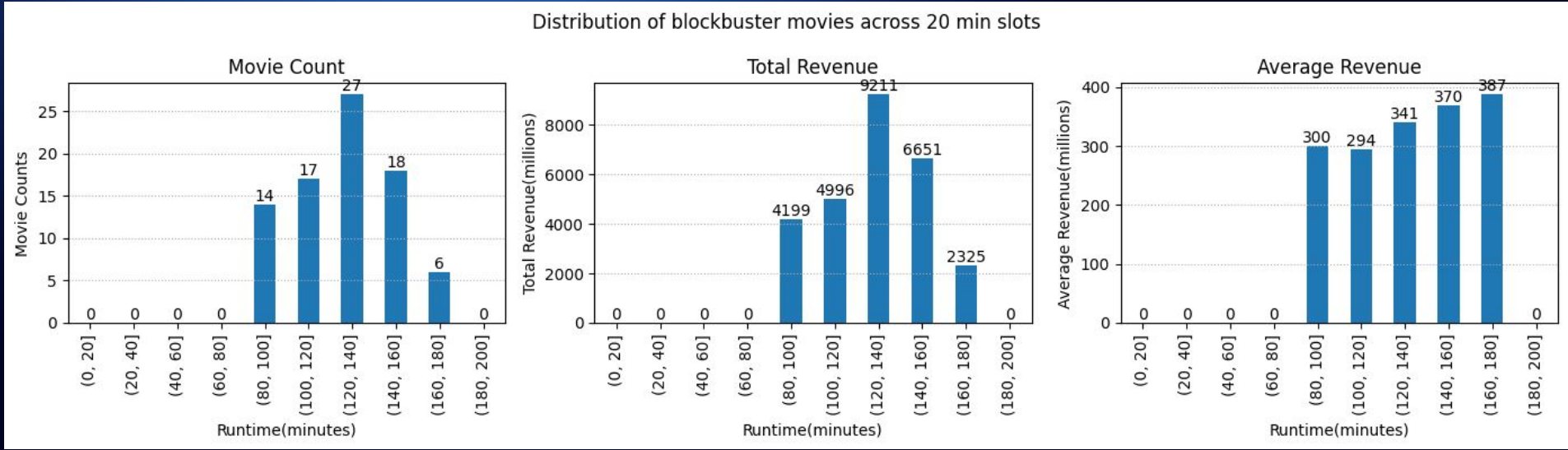
Runtime

How does the movie runtime affect votes and critic's score?



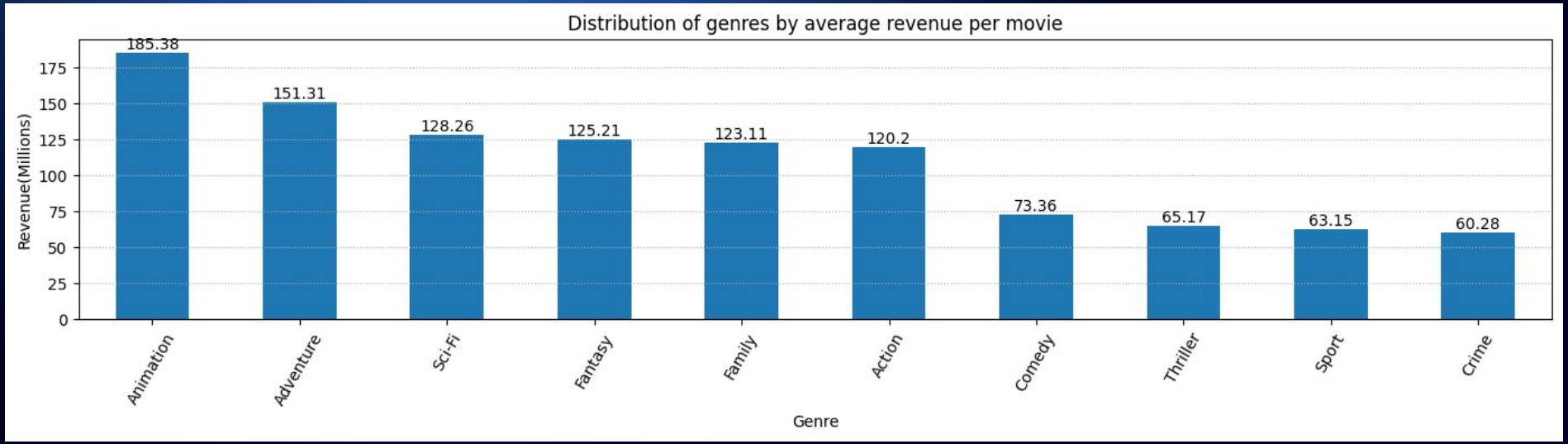
Runtime

Which runtime range gives more blockbusters than the others?



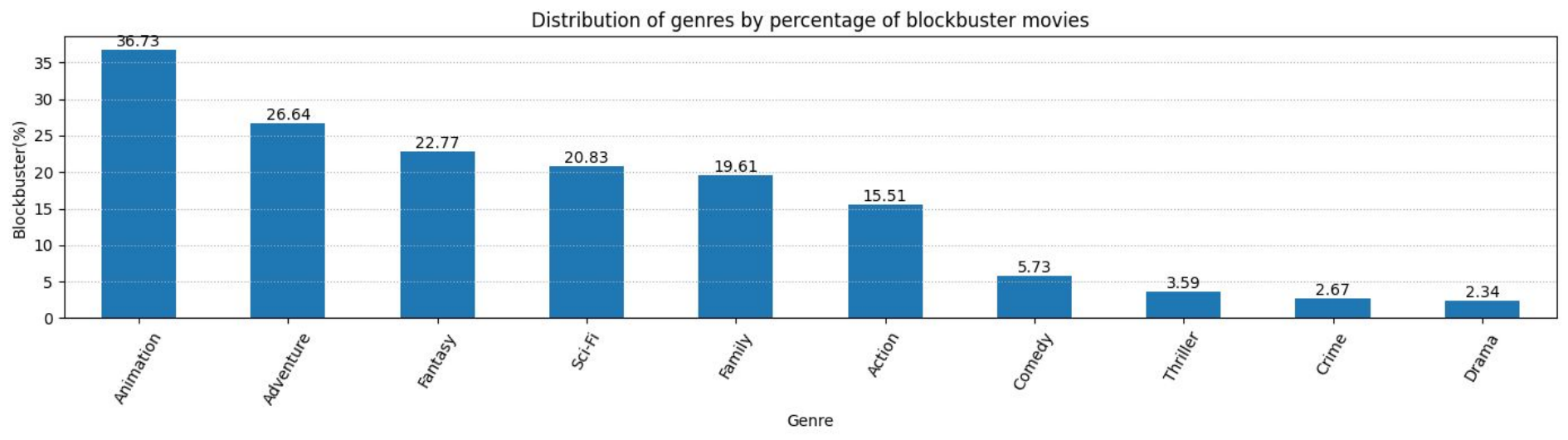
Genre

Which are the top 10 genres that make the highest revenue per movie?



Genre

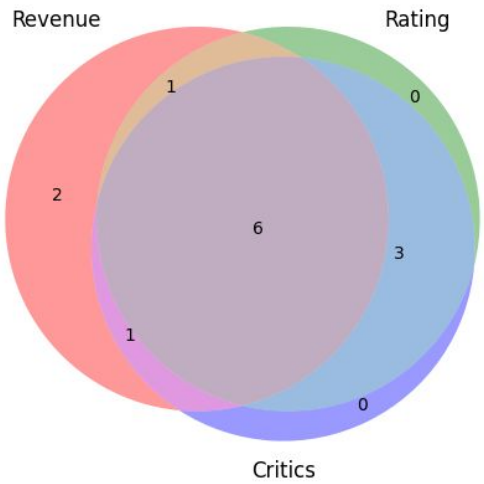
Which are the top 10 genre that make the maximum number of blockbuster movies(sorted by percentage)?



Genre

Which are the genres that are common in top 10 list of revenue, rating and critic's score?

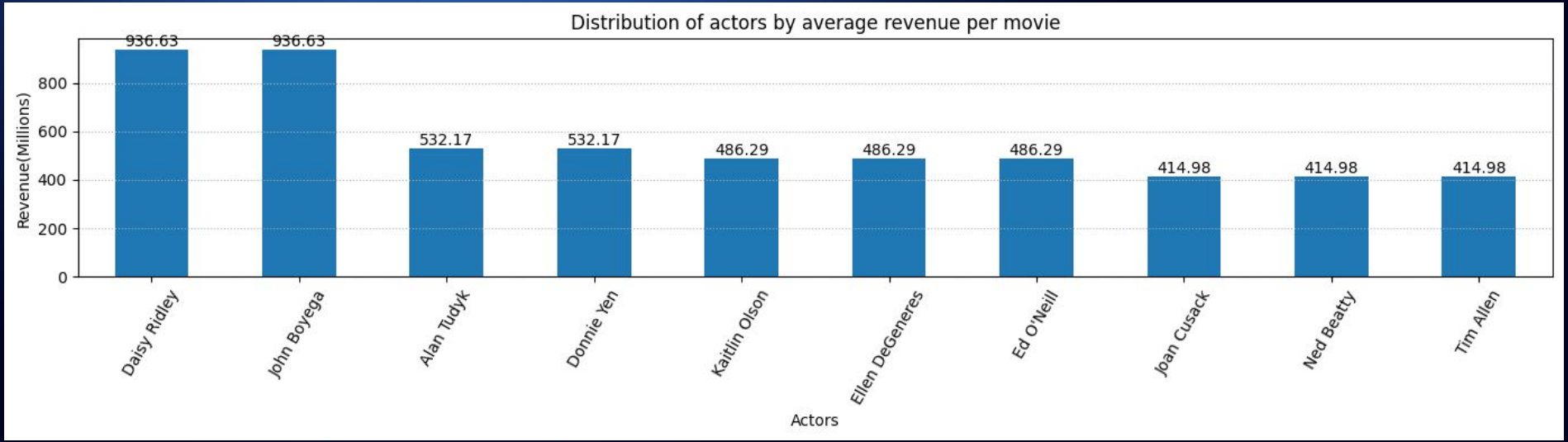
Top genres by revenue, rating and critic's score



Genre	Avg Revenue(millons)	Avg Rating	Avg Critic's Score
Adventure	151.31	6.77	57.72
Animation	185.38	7.32	70.06
Comedy	73.36	6.65	57.66
Crime	60.28	6.79	57.86
Sci-Fi	128.26	6.72	58.27
Sport	63.15	7.01	61.28

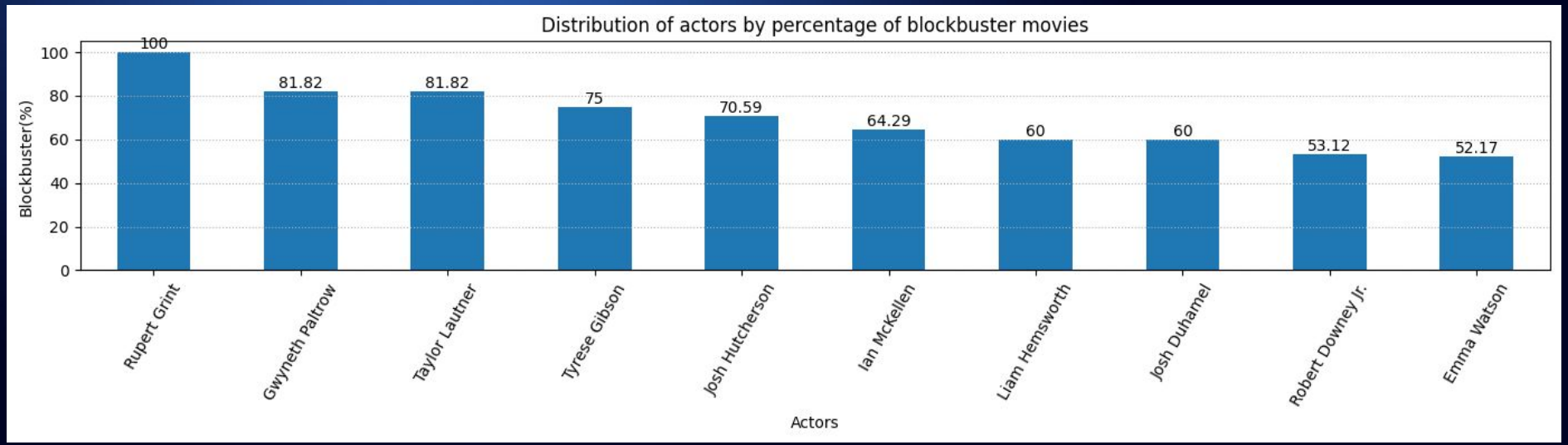
Actor

Who are the top 10 actors that make the highest revenue per movie?



Actor

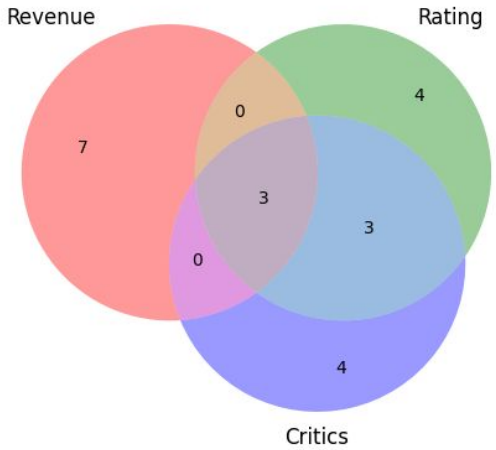
Who are the top 10 actors that make the maximum number of blockbuster movies(sorted by percentage)?



Actor

Who are the actors that are common in top 10 list of revenue, rating and critic's score?

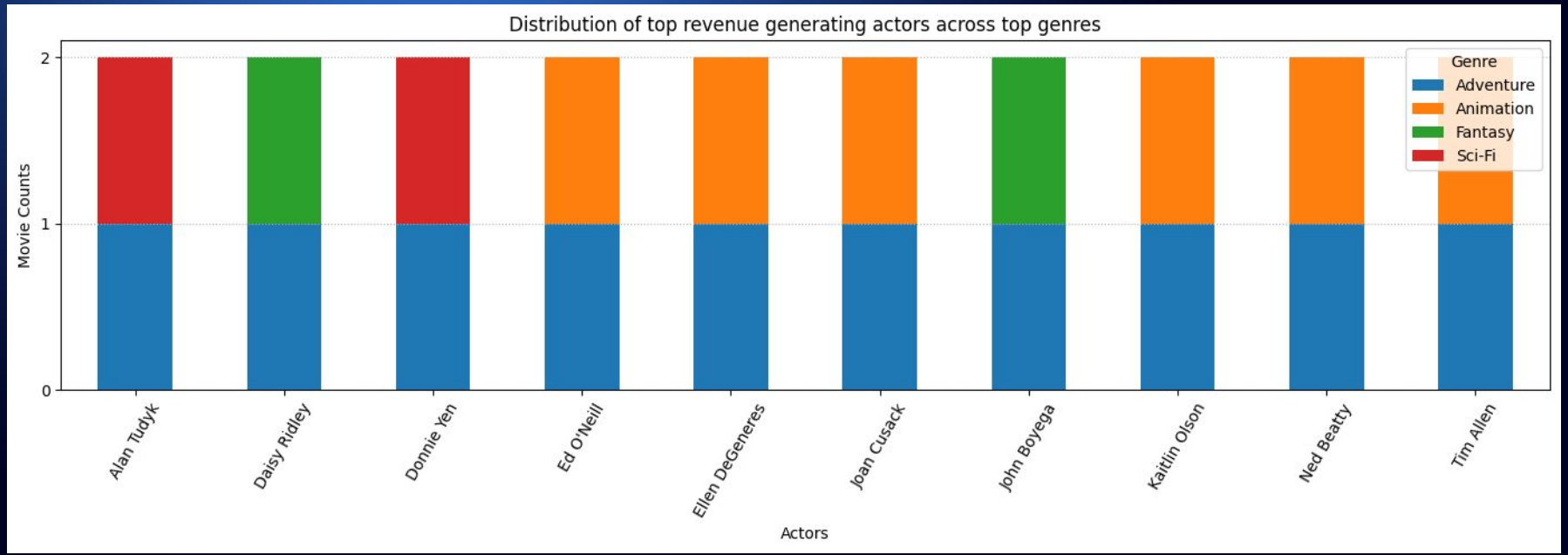
Top actors by revenue, rating and critic's score



Actors	Avg Revenue(millons)	Avg Rating	Avg Critic's Score
Joan Cusack	414.98	8.3	92.0
Ned Beatty	414.98	8.3	92.0
Tim Allen	414.98	8.3	92.0

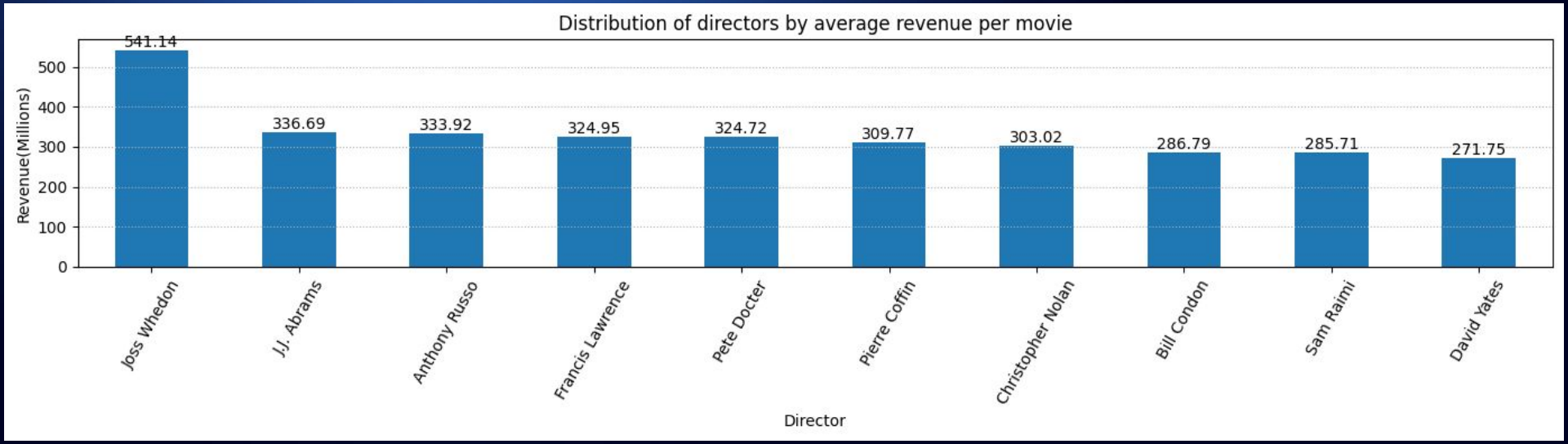
Actor

Which genres(top 5 highest revenue) do the top 10 actors perform in?



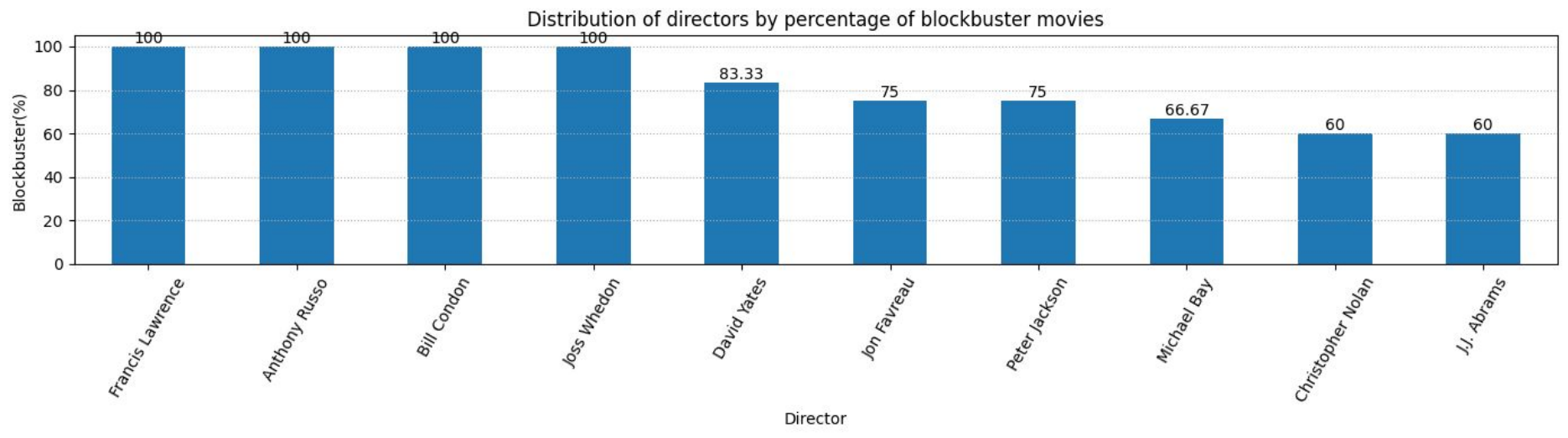
Director

Who are the top 10 directors that make the highest revenue per movie?



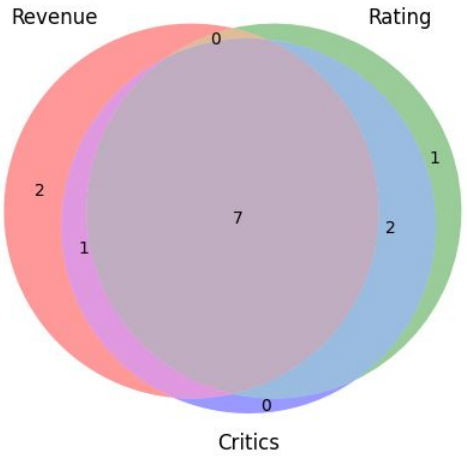
Director

Who are the top 10 directors that makes the maximum number of blockbuster movies(sorted by percentage)?



Director

Who are the directors that are common in top 10 list of revenue, rating and critic's score?

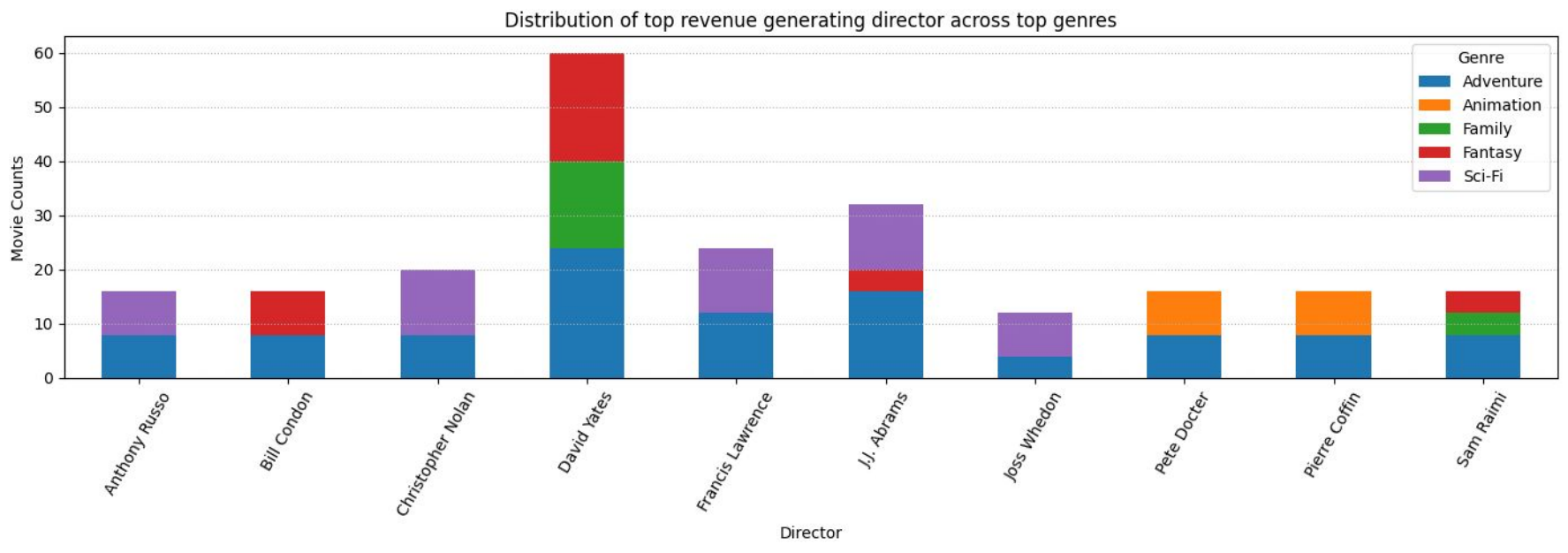


Top directors by revenue, rating and critic's score

Director	Avg Revenue(millons)	Avg Rating	Avg Critic's Score
Anthony Russo	333.92	7.85	72.5
Christopher Nolan	303.02	8.68	74.8
David Yates	271.75	7.43	68.5
J.J. Abrams	336.69	7.58	74.6
Joss Whedon	541.14	7.75	67.5
Pete Docter	324.72	8.25	91.0
Pierre Coffin	309.77	7.55	67.0

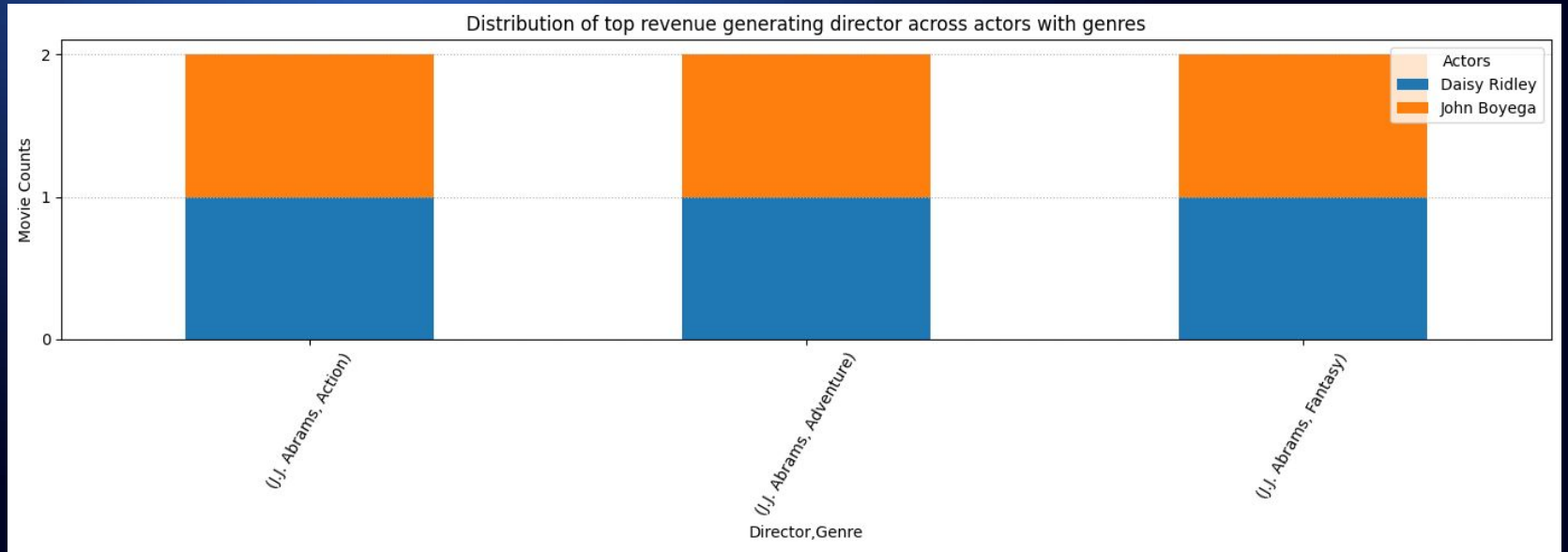
Director

Which genres(top 5 highest revenue) do the top 10 directors direct in?



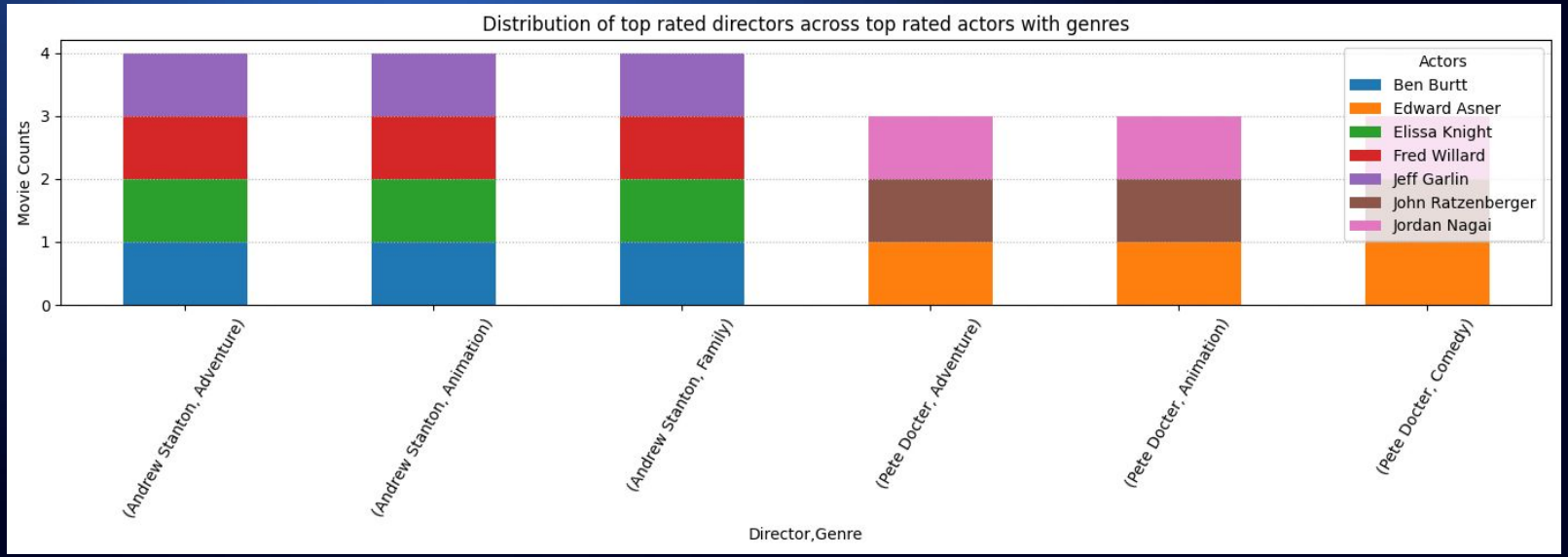
Director-Actor-Genre

Who are the highest revenue actors that the top 10 highest revenue directors have worked with along with genres?



Director-Actor-Genre

Who are the top rated actors that the top rated directors have worked with along with genres?



Summary

General:

- Revenue increases with Rating, Critic's score, Votes and Runtime.

- Highest revenue per movie (on average):
  - Runtime: 140 to 160 mins
  - Genre: Animation, Adventure and Sci-Fi
  - Actors: Daisy Ridley, John Boyega and Alan Tudyk
  - Directors: Joss Whedon, J.J. Abrams and Anthony Russo

Summary

Blockbusters Count:

- Runtime: 120 to 140 mins
- Genres: Animation, Adventure and Fantasy
- Actors: Rupert Grint, Gwyneth Paltrow and Taylor Lautner
- Directors: Francis Lawrence, Anthony Russo, Bill Condon and Joss Whedon

Actionable Insights

Movie recommendations:

- High revenue movie
  - Director: J.J. Abrams
  - Actors: Daisy Ridley or John Boyega
  - Genre: Action, Adventure or Fantasy
- High rating movie A:
  - Director: Andrew Stanton
  - Actors: Ben Burtt, Fred Willard, Jeff Garlin or Elissa Knight
  - Genre: Adventure, Animation or Family
- High Rating movie B:
  - Director: Pete Docter
  - Actors: Jordan Nagai, John Ratzenberger or Edward Asher
  - Genre: Adventure, Animation or Comedy
- Runtime: 120 to 140 mins

Other recommendations:

- Clustering algorithm can be used to cluster directors, actors and genres instead of scripts identifying the combinations

Thank You