Movie Database Analysis

Isaac Muturi

11-Mar-2023

Overview

- Business Understanding
- Data Understanding
- Data Analysis
- Recommendations
- Next Steps
- Thank You
- prompt for questions as well as your contact information (name and LinkedIn profile)

Business Understanding

The movie industry is complex and involves various stakeholders. Success depends on several factors, including script quality, star power, marketing, release date, and competition.

To analyze the industry, we need data from multiple sources and use data science techniques. A combination of quantitative and qualitative analysis techniques and domain expertise is necessary to make informed decisions for producing successful movies and improving the industry's profitability.

The First Dataset

Name: tn.movie_budgets.csv

Type: csv

Time frame of the data: 1915 to 2020

Total data points : 5782 records

Data columns (total 6 columns):

#	Column	Dtype
0	id	int64
1	release_date	object
2	movie	object
3	production_budget	object
4	domestic_gross	object
5	worldwide_gross	object

dtypes: int64(1), object(5)

Data Understanding

The second dataset

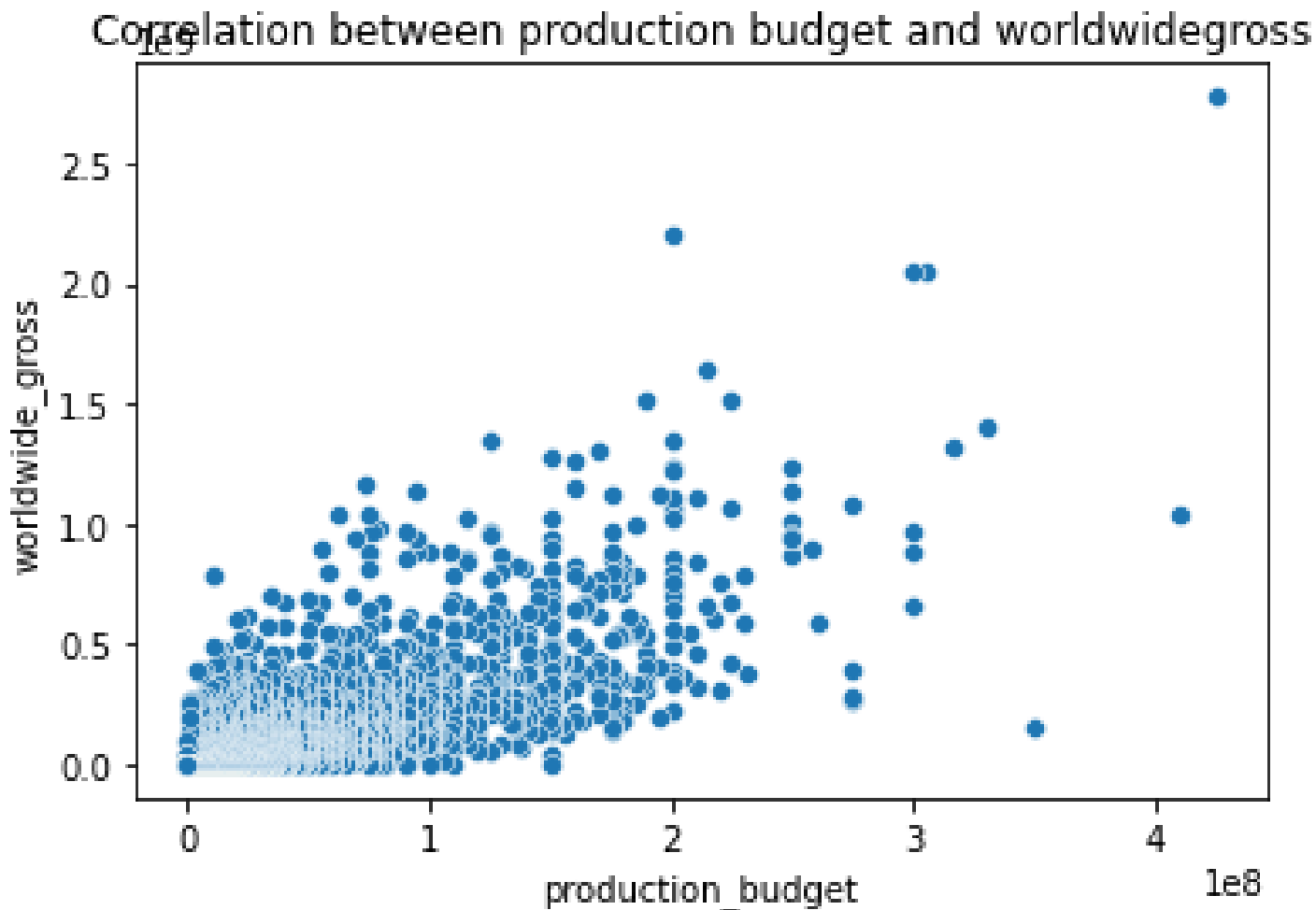Name: im.db
Type: sqlite database

| Table Names	columns
0 movie_basics	(movie_id, primary_title, original_title, start_year, runtime_minutes, genres)
1 directors	(movie_id, person_id)
2 known_for	(person_id, movie_id)
3 movie_akas	(movie_id, ordering, title, region, language, types, attributes, is_original_title)
4 movie_ratings	(movie_id, averagerating, numvotes)
5 persons	(person_id, primary_name, birth_year, death_year, primary_profession)
6 principals	(movie_id, ordering, person_id, category, job, characters)
7 writers	(movie_id, person_id)

Correlation between production budget and worldwidegross

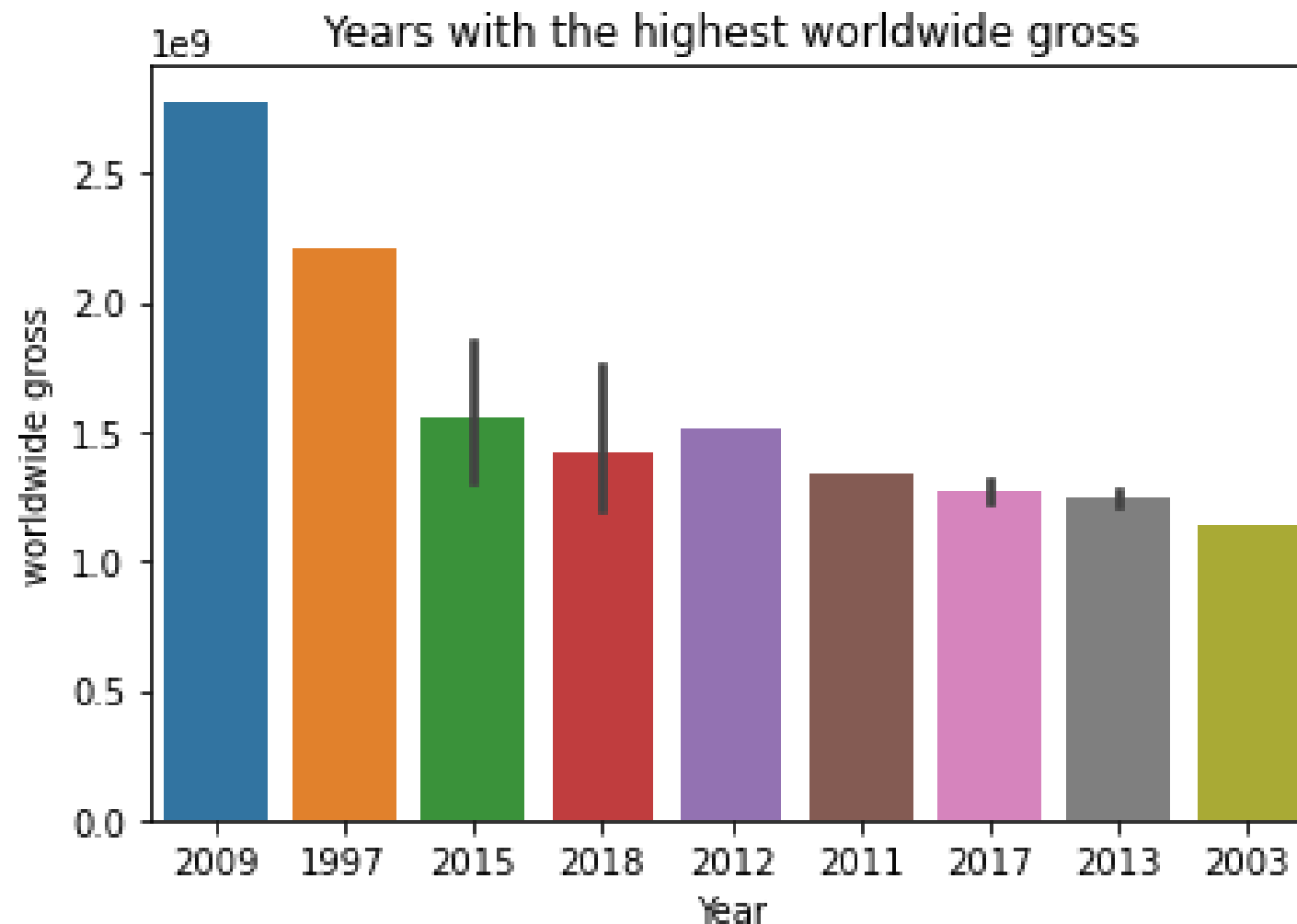There is a positive relationship between production budget and worldwide gross

Histogram of production_budget
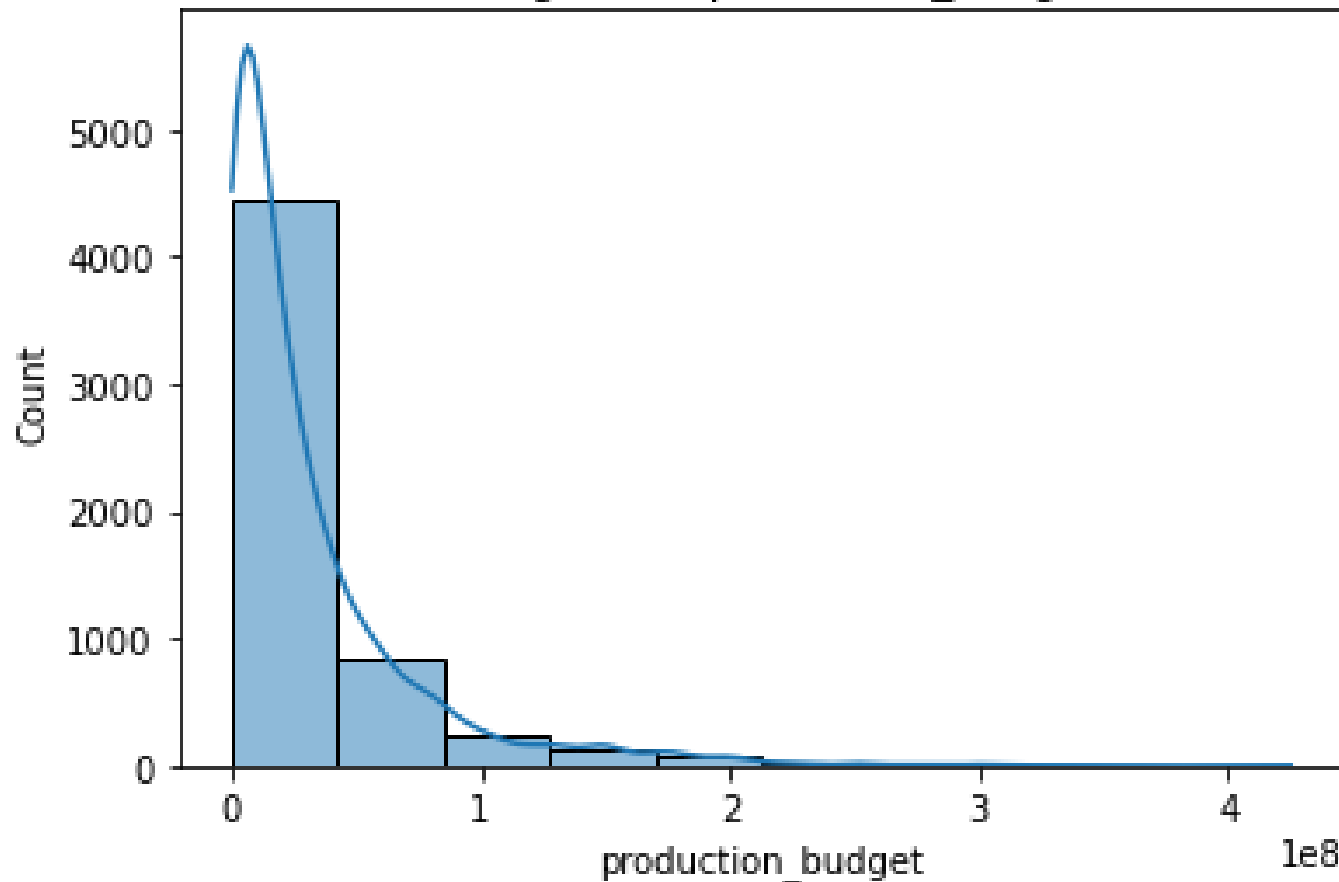
The number of movies increases as the production budget increases.

Which are the highest grossing years

Years with the highest worldwide gross

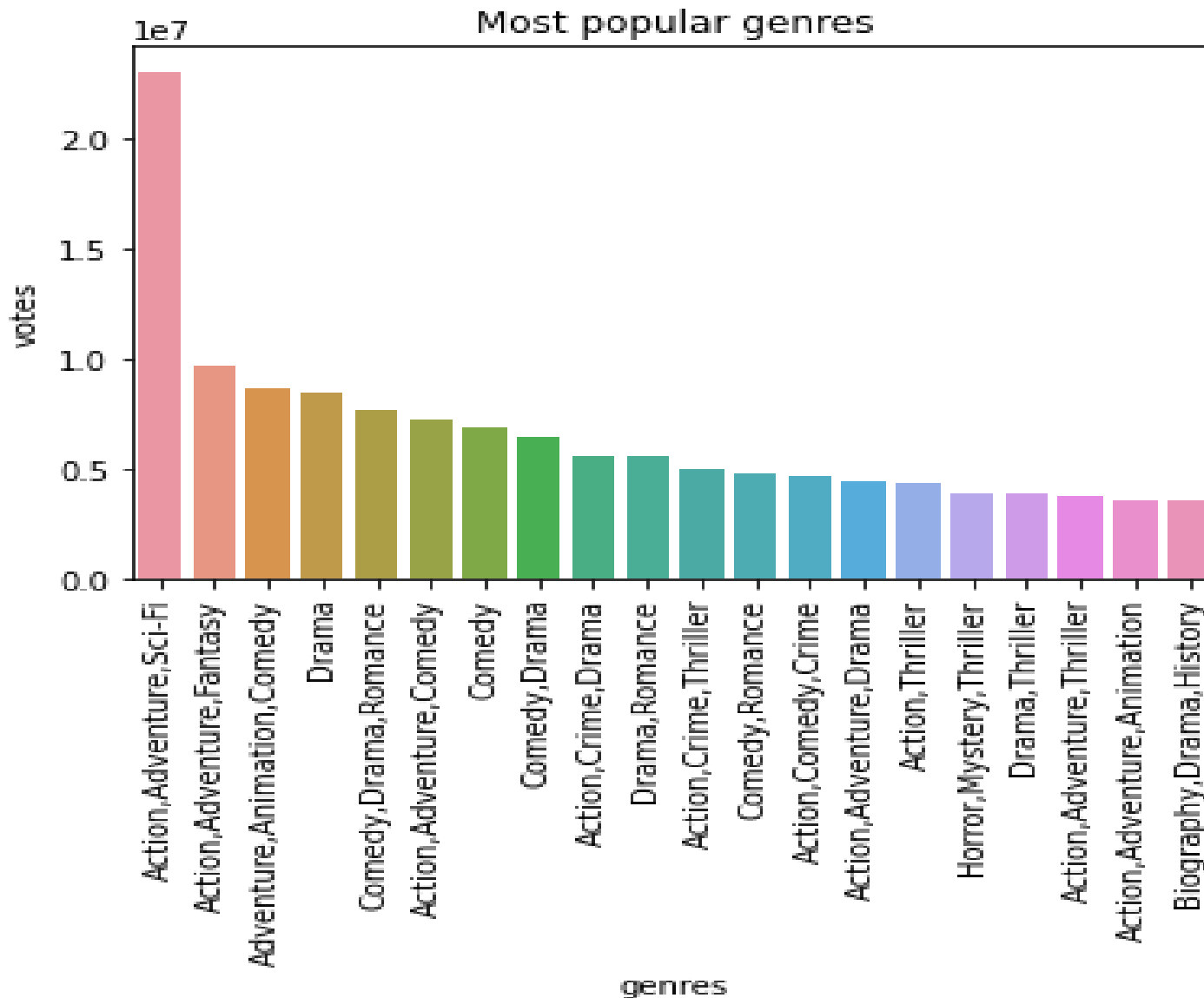There is no correlation between year of movie release and worldwide gross.
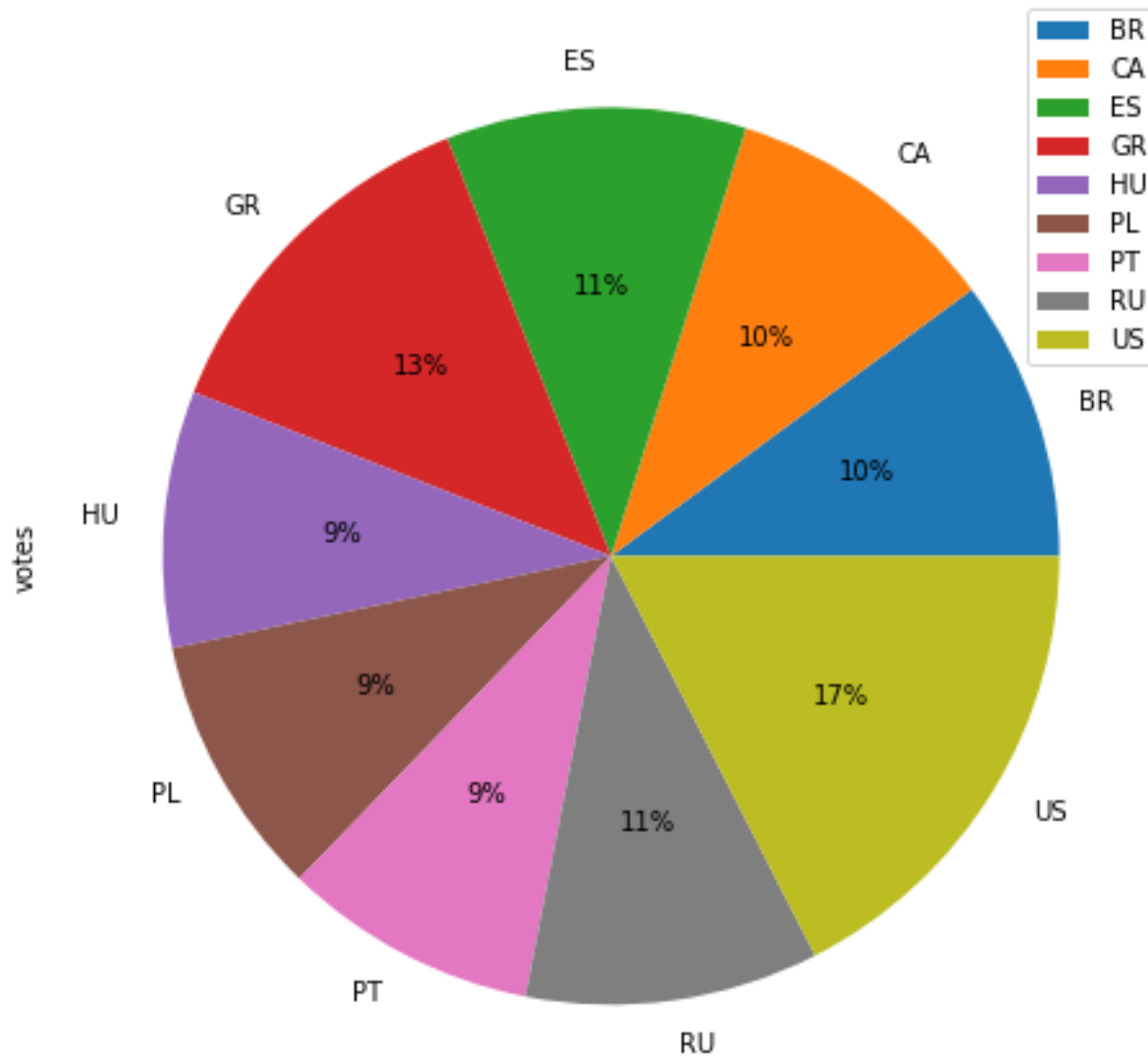
Most popular titles
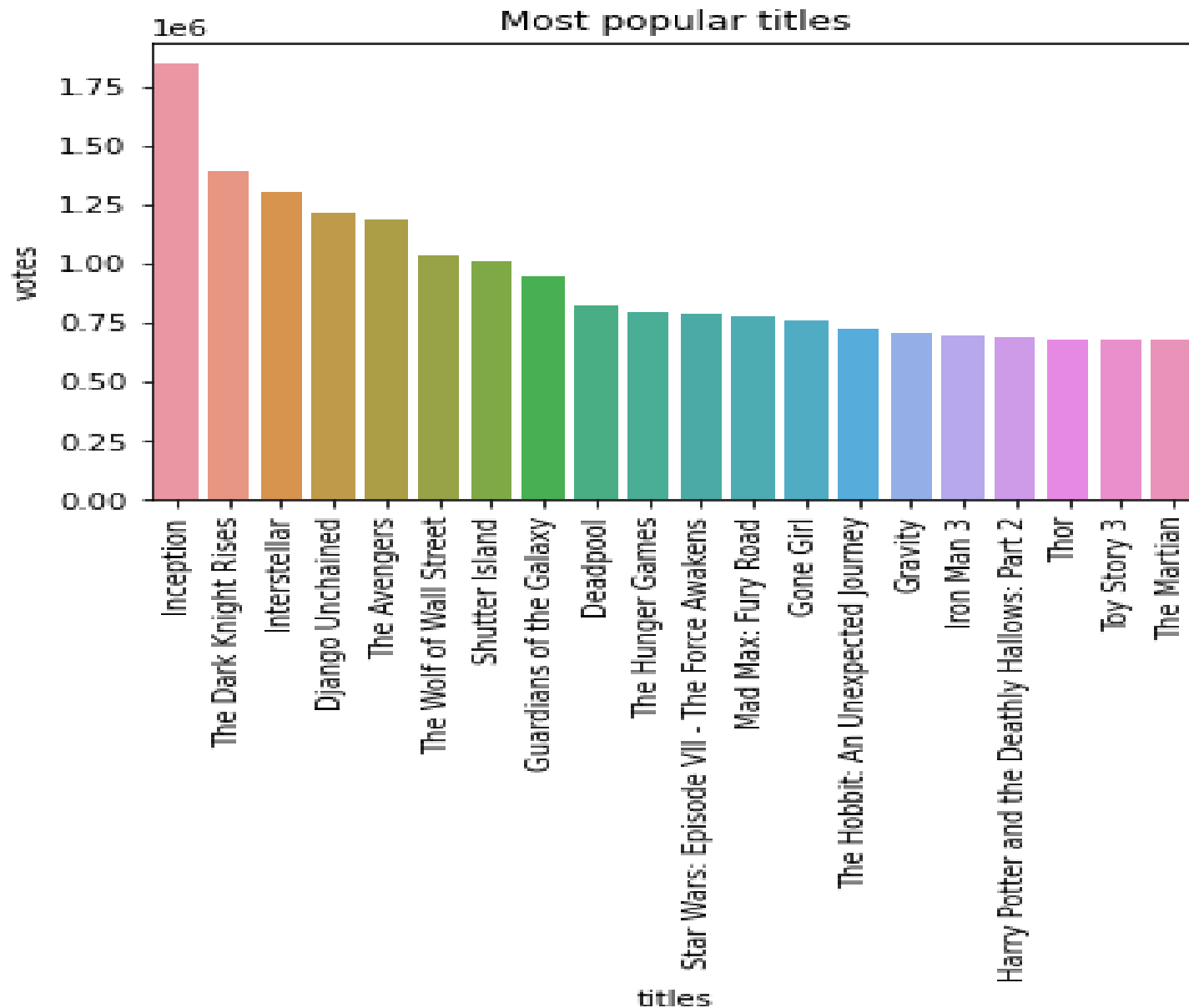
Movie sequels are very popular. They make a majority in this list

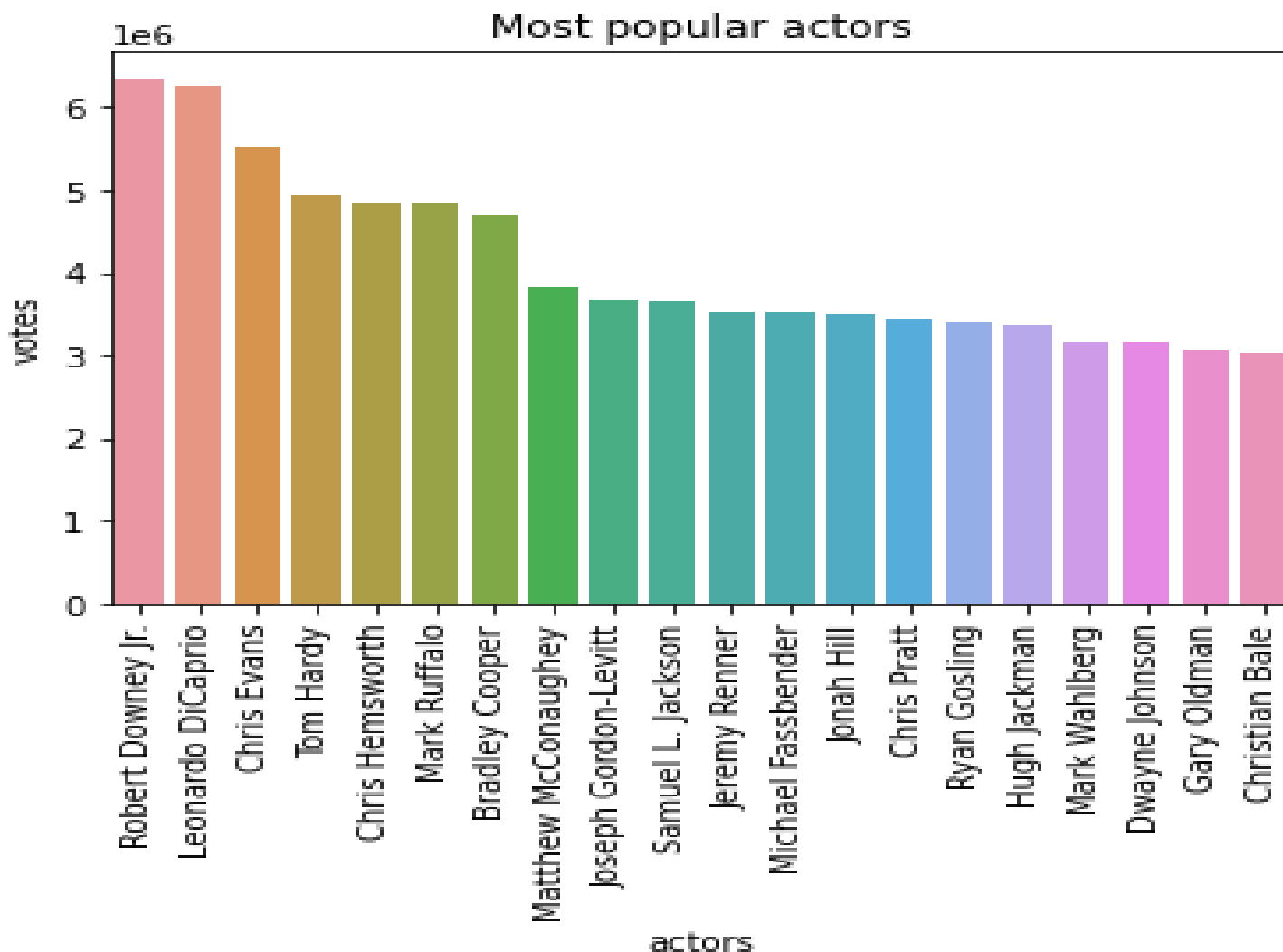# which are the most popular genres

A combo of Action, Adventure and Sci-Fi are significantly more popular than other genres

Which are the most popular regions or markets

US is the clear leader with 17 % of the market among the top 10 regions, followed by GR with 13% and then a tie at 10% by CA and BR

# Who are the most popular actors

Most popular actors

Robert Downey Jr leads the pack, followed closely by Leonardo DiCaprio, then Chris Evans and other popular movie actors

# Who are the most popular actresses

Most popular actresses
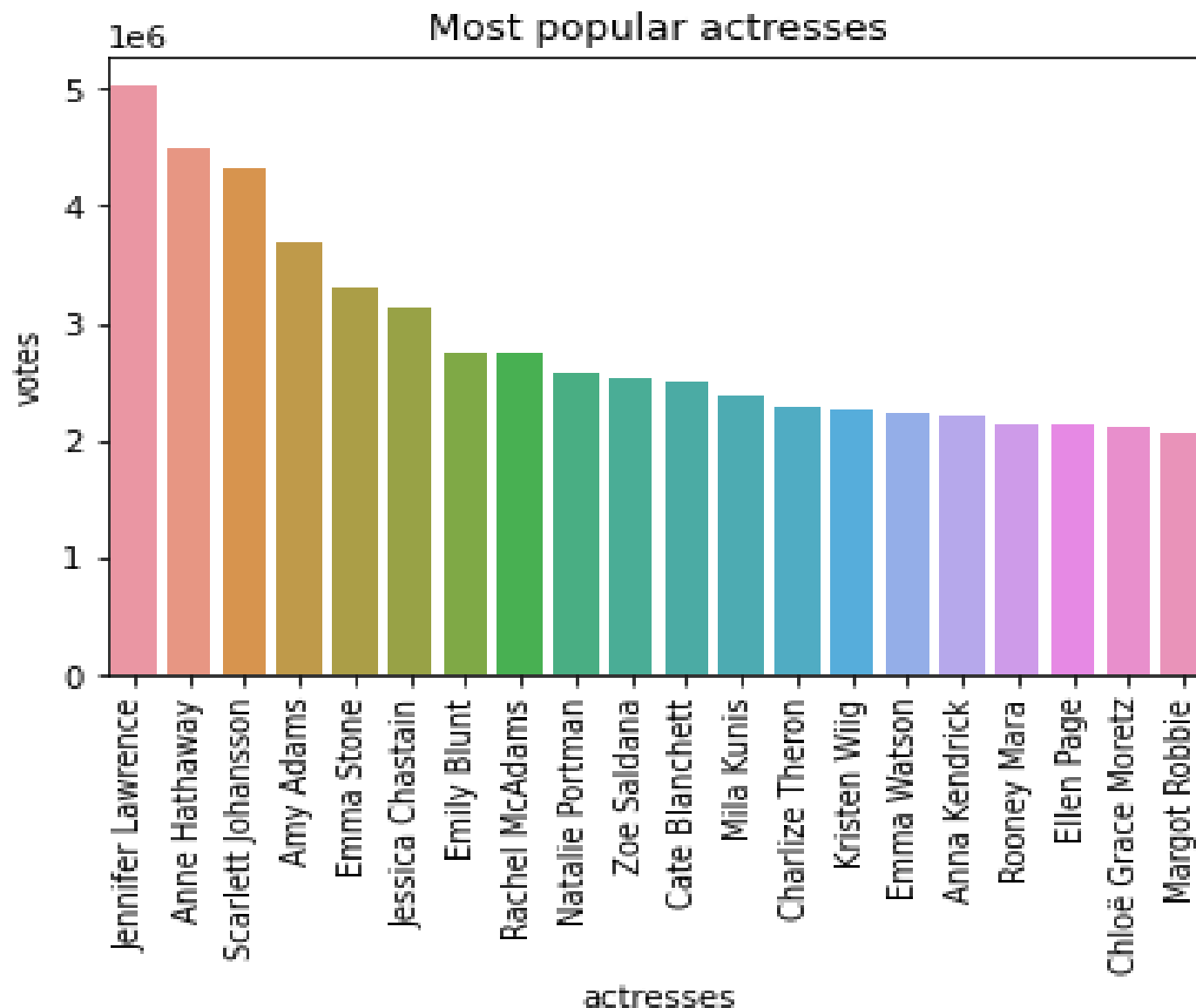
Jennifer Lawrence leads the pack, followed by Anne Hathaway and Scarlett Johansson and other popular actresses

# Who are the most popular directors

Most popular directors
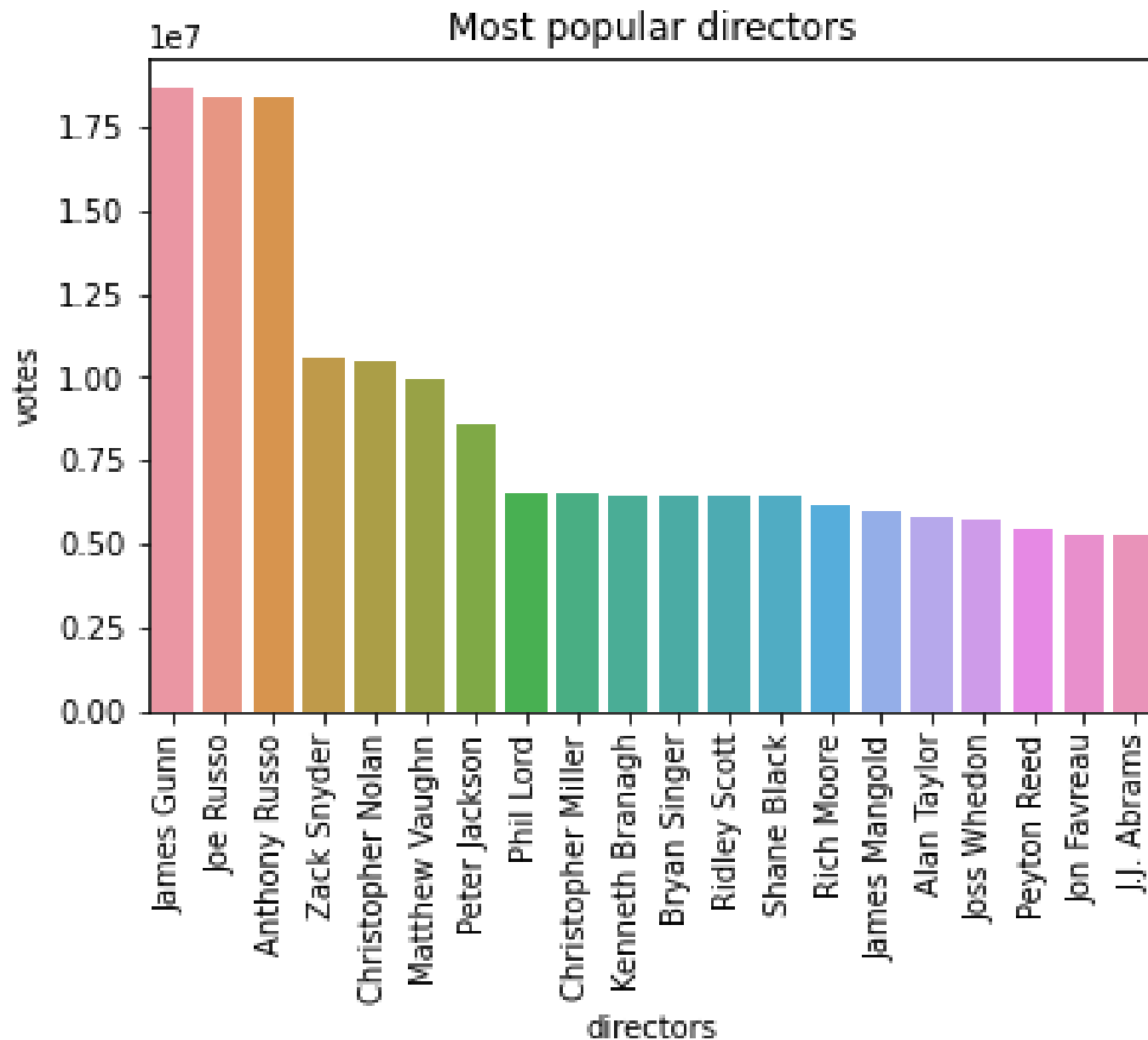
Three directors, James Gunn, Joe Russo and Anthony Russo lead significantly followed by other popular directors

What is the correlation between movie runtime and popularity

correlation between runtime and popularity
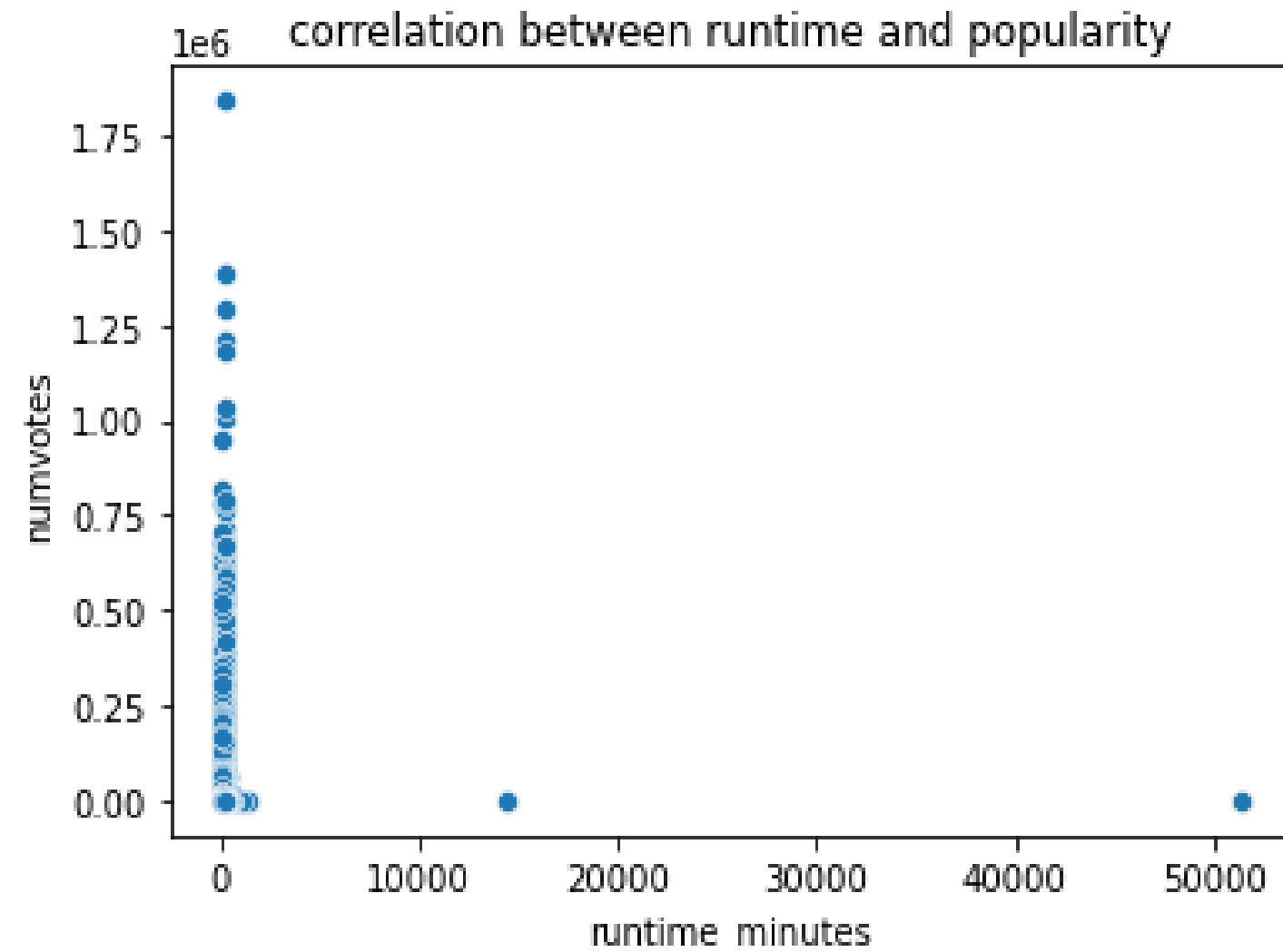
There is no correlation between runtime and popularity of movies

What is the correlation between movie ratings and popularity?

Correlation between ratings and popularity
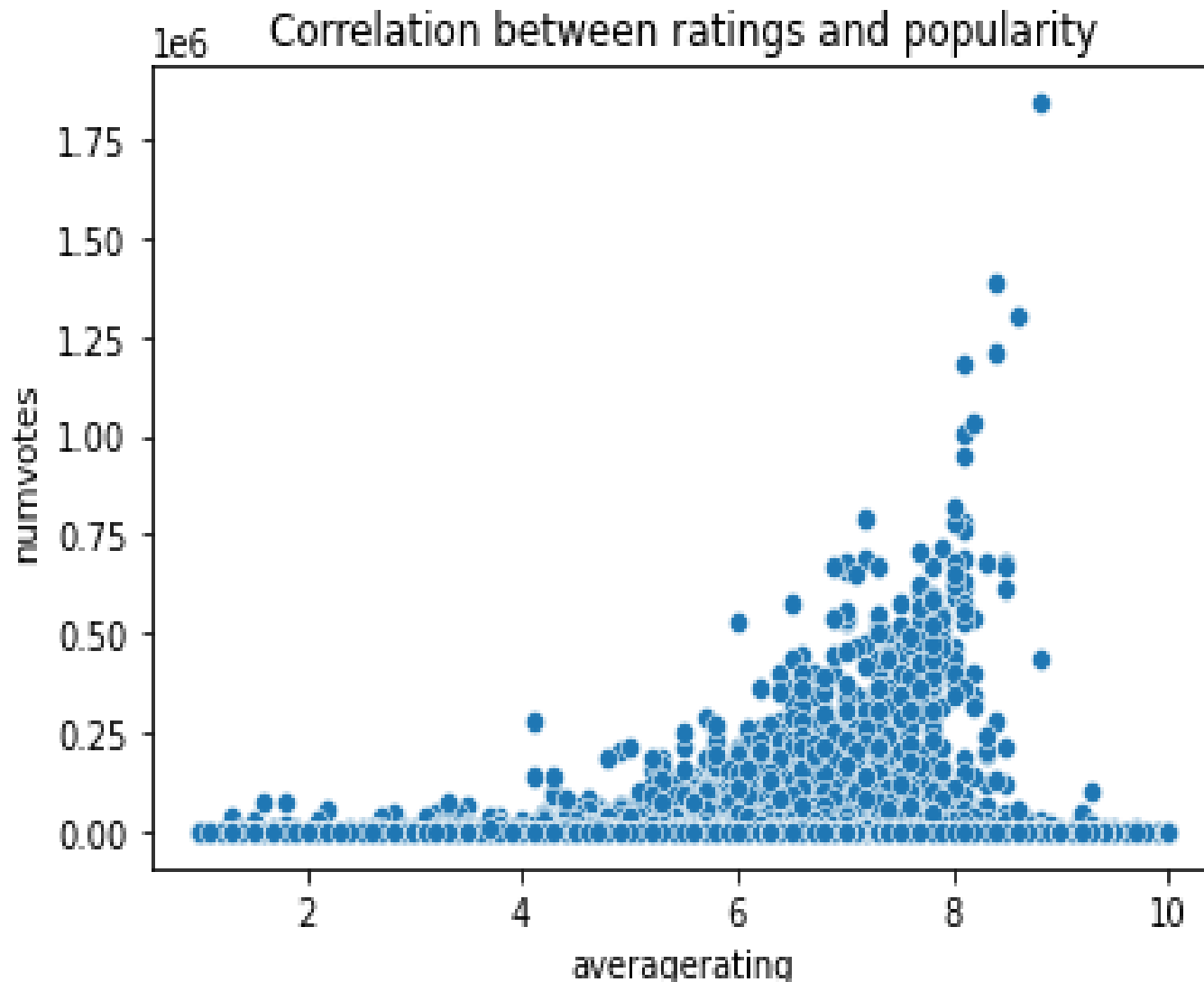
There is a positive relationship between movie ratings and popularity.

Recommendations

1.Collaborate with Popular Actors and Directors: Actors and directors who have a strong track record of success can be highly valuable to movie production companies.

2.Create Content Based on Audience Interests: By analyzing the data on ratings, it is possible to identify the genres and storylines that are most popular among audiences. A combo of Action, Adventure and Sci-Fi are significantly more popular than other genres.

3.It is important to consider this audience preferences which differ based on region when planning movie production and distribution. US is the clear leader with 17 % of the market among the top 10 regions, followed by GR with 13% and then a tie at 10% by CA and BR. You can decide to focus your content to cater for these more potent markets.

What are the next steps

Overall, the next steps for the movie dataset analysis project involve a combination of data cleaning, exploratory analysis, feature engineering, model selection, evaluation, and deployment. By following a structured approach and leveraging the latest tools and techniques, data scientists can generate valuable insights that can inform decision-making in the movie industry.

Thank You

This presentation is prepared by **Isaac Muturi**. If you have any enquiry, feel free to contact me at:

https://www.linkedin.com/in/ndirangu-muturi-3b6b2b237