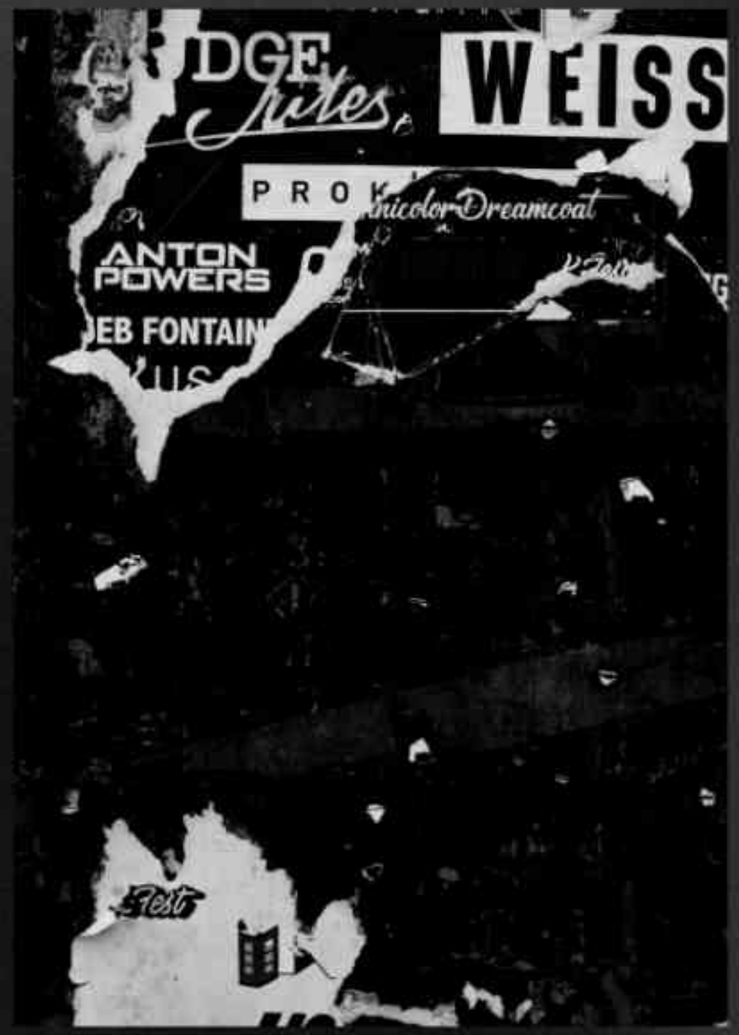

SDU
UNIVERSITY
MOVIES
TEMIRLAN AMANZHOLULY
210107005
TIMA.KING PICTURES

1.DATASETS
2.VISUALIZATION
3.HYPOTHESES
4.RECOMENDATIONS



TIMA.KING PICTURES


DATASET

DATASET

my dataset is about films from 1980 and all their characteristics

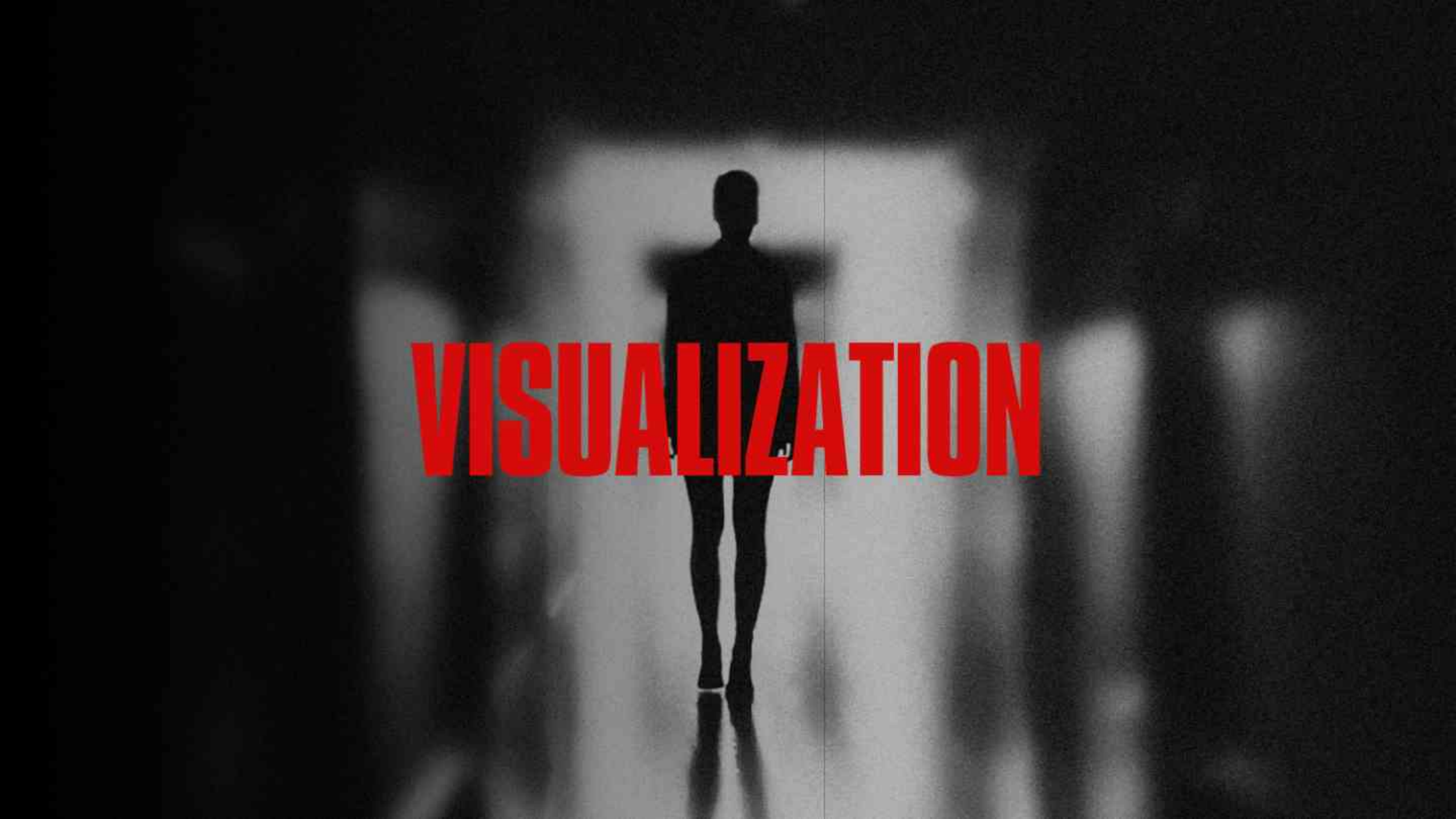
```
df.head()
```

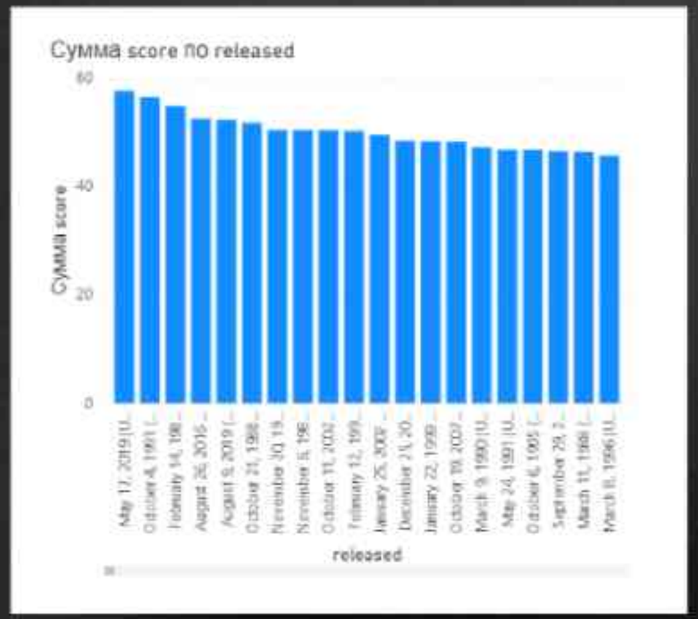
	name	rating	genre	year	released	score	votes	director	writer	star	country	budget	gross	company	runtime
0	The Shining	R	Drama	1980	June 13, 1980 (United States)	8.4	927000.0	Stanley Kubrick	Stephen King	Jack Nicholson	United Kingdom	19000000.0	46998772.0	Warner Bros.	146.0
1	The Blue Lagoon	R	Adventure	1980	July 2, 1980 (United States)	5.8	65000.0	Randal Kleiser	Henry De Vere Stacpoole	Brooke Shields	United States	4500000.0	58853106.0	Columbia Pictures	104.0
2	Star Wars: Episode V - The Empire Strikes Back	PG	Action	1980	June 20, 1980 (United States)	8.7	1200000.0	Irvin Kershner	Leigh Brackett	Mark Hamill	United States	18000000.0	538375067.0	Lucasfilm	124.0
3	Airplane!	PG	Comedy	1980	July 2, 1980 (United States)	7.7	221000.0	Jim Abrahams	Jim Abrahams	Robert Hays	United States	3500000.0	83453539.0	Paramount Pictures	88.0
4	Caddyshack	R	Comedy	1980	July 25, 1980 (United States)	7.3	108000.0	Harold Ramis	Brian Doyle-Murray	Chevy Chase	United States	6000000.0	39846344.0	Orion Pictures	98.0

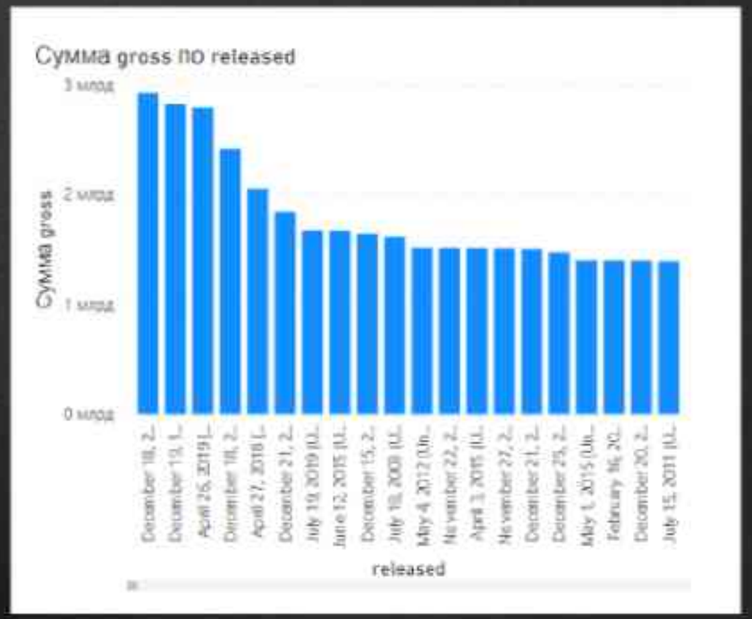
```
df = df.drop_duplicates()
df.shape
```

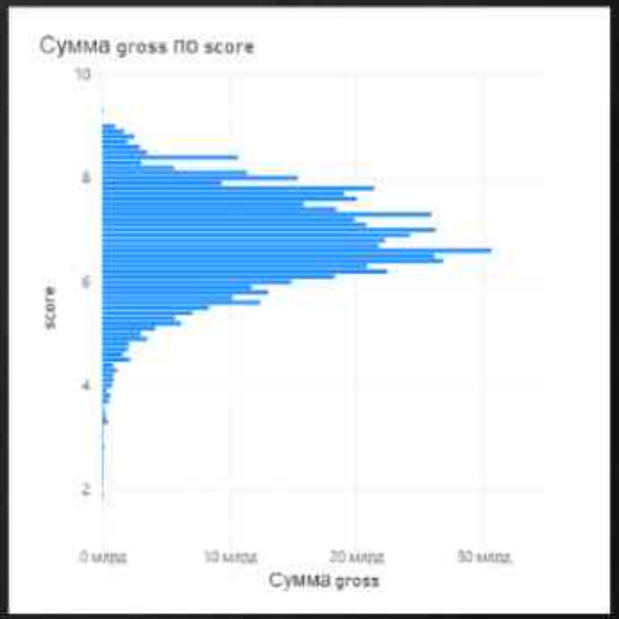
```
(7668, 15)
```

TIMA.KING PICTURES


VISUALIZATION



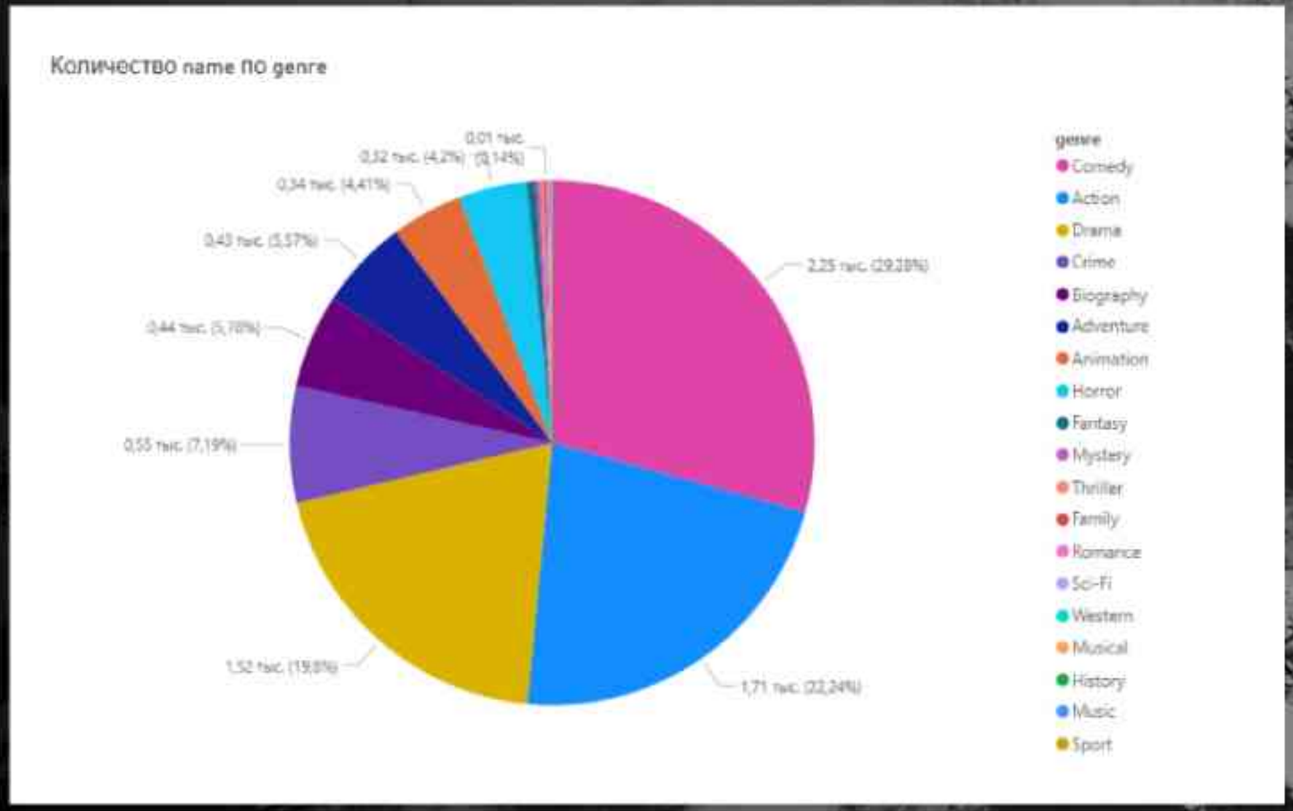


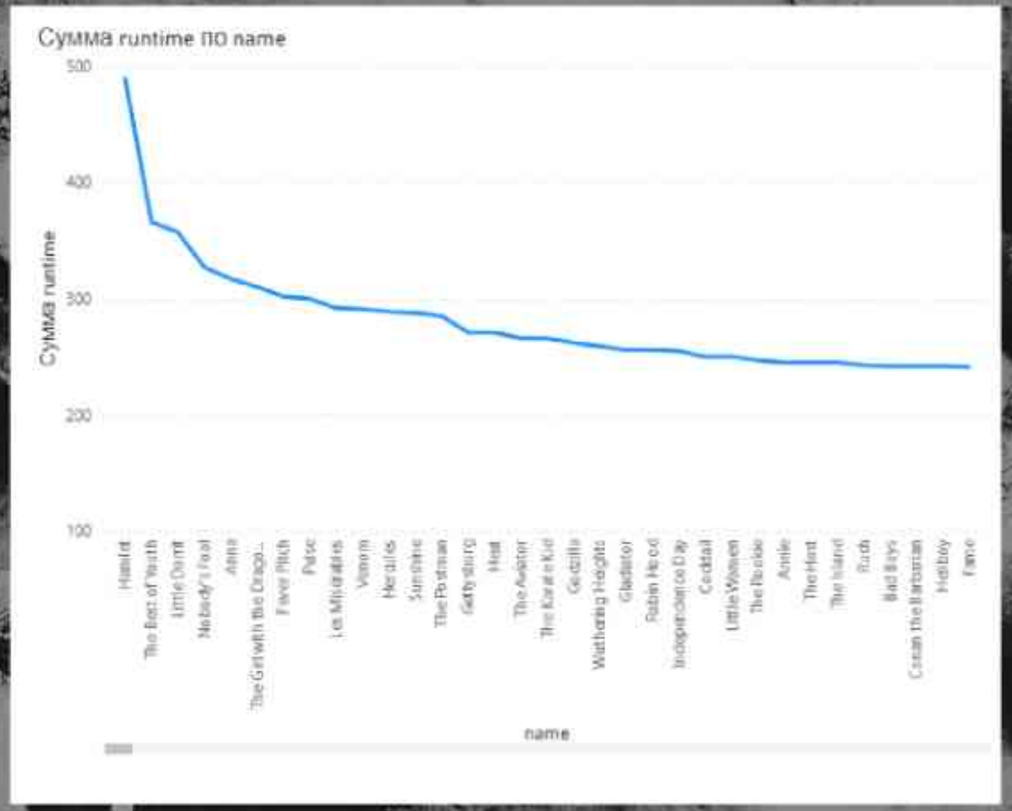


VISUALIZATION

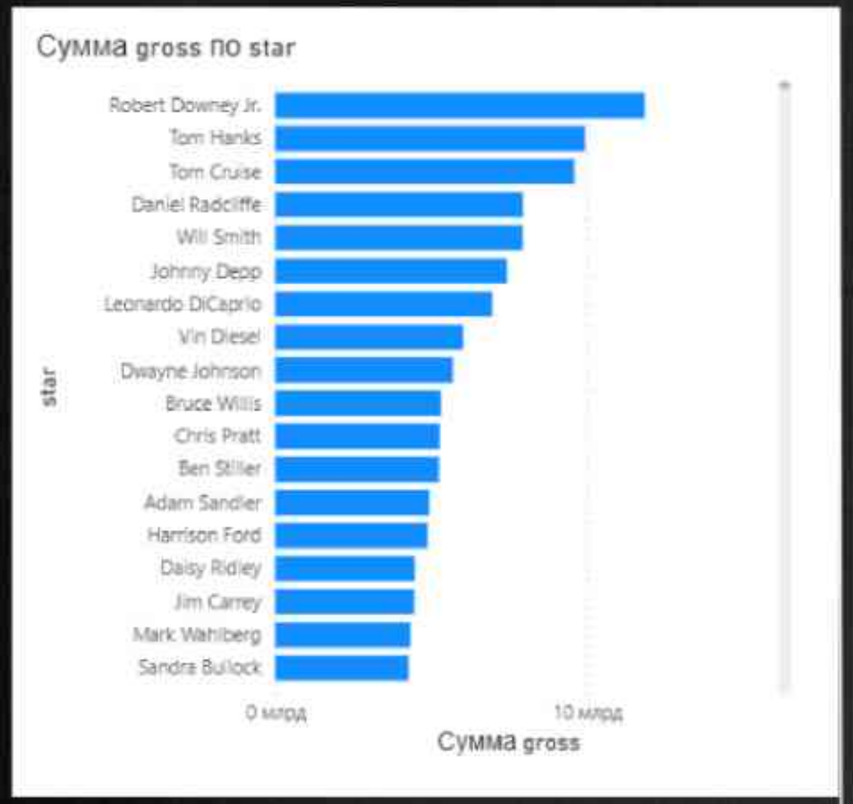
TIMA.KING PICTURES

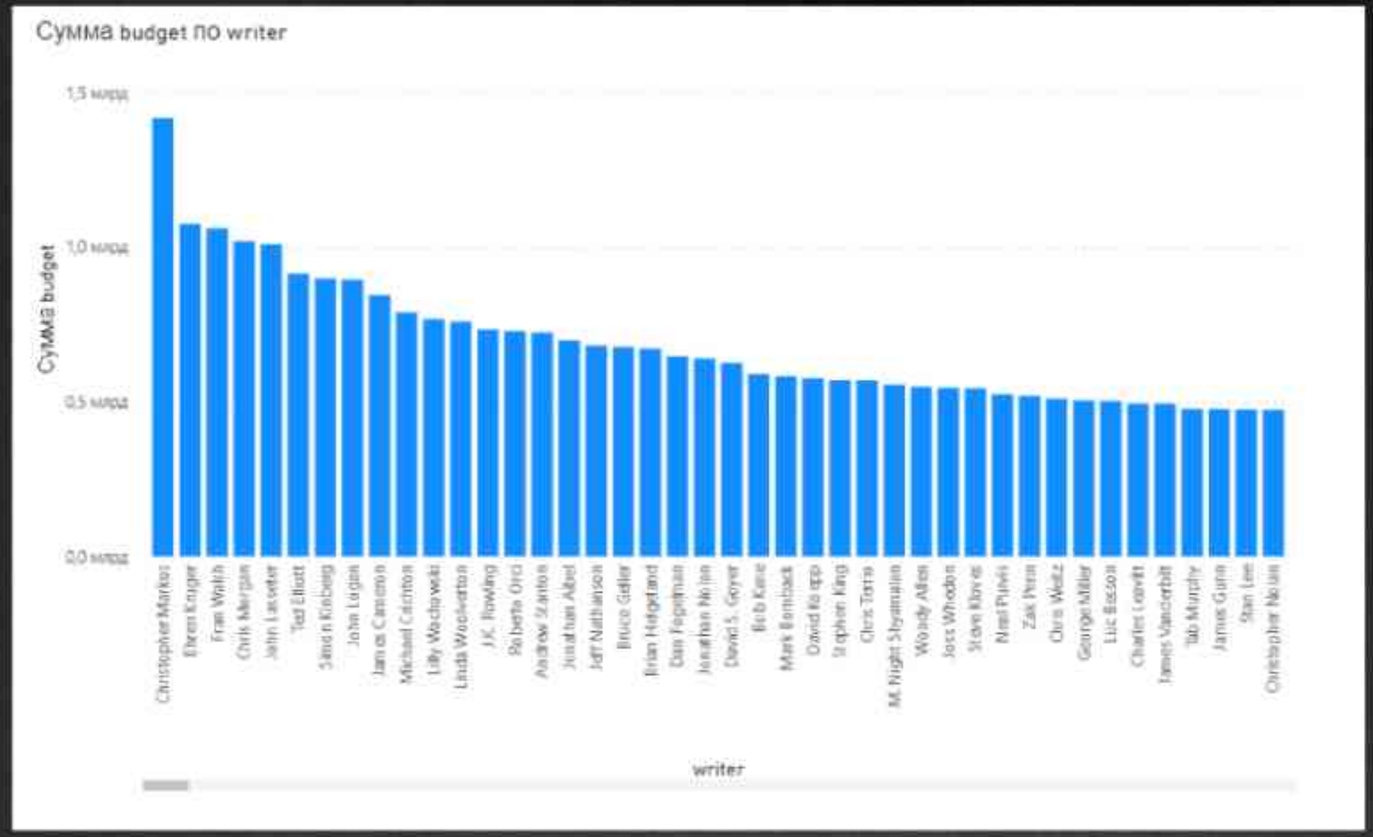
VISUALIZATION





VISUALIZATION






HYPOTHESES #1-2

HYPOTHESES

Hypothesis 1: Films of certain genres have a higher average rating than films of other genres.

Hypothesis 2: Films with a higher budget usually have a higher total revenue.

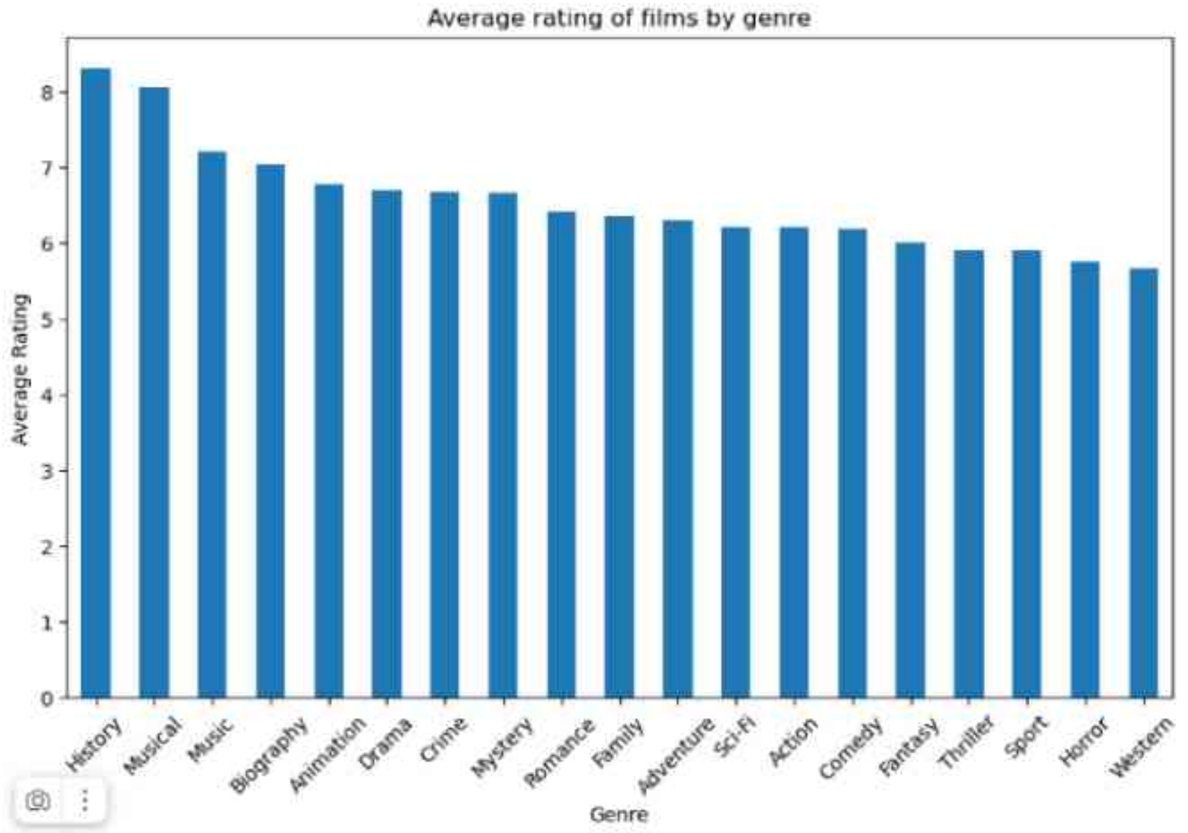
```python
import pandas as pd
import matplotlib.pyplot as plt

data = pd.read_csv('movies.csv')

# Hypothesis 1: Films of certain genres have a higher average rating than films of other genres.

genre_ratings = data.groupby('genre')['score'].mean().sort_values(ascending=False)
print("Average rating of films of various genres:")
print(genre_ratings)

genre_ratings.plot(kind='bar', figsize=(10, 6))
plt.title('Average rating of films by genre')
plt.xlabel('Genre')
plt.ylabel('Average Rating')
plt.xticks(rotation=45)
plt.show()

# Hypothesis 2: Films with a higher budget usually have a higher total revenue.
plt.figure(figsize=(10, 6))
plt.scatter(data['budget'], data['gross'], alpha=0.5)
plt.title('The relationship between budget and total movie revenue')
plt.xlabel('Budget')
plt.ylabel('Total revenue')
plt.show()
```

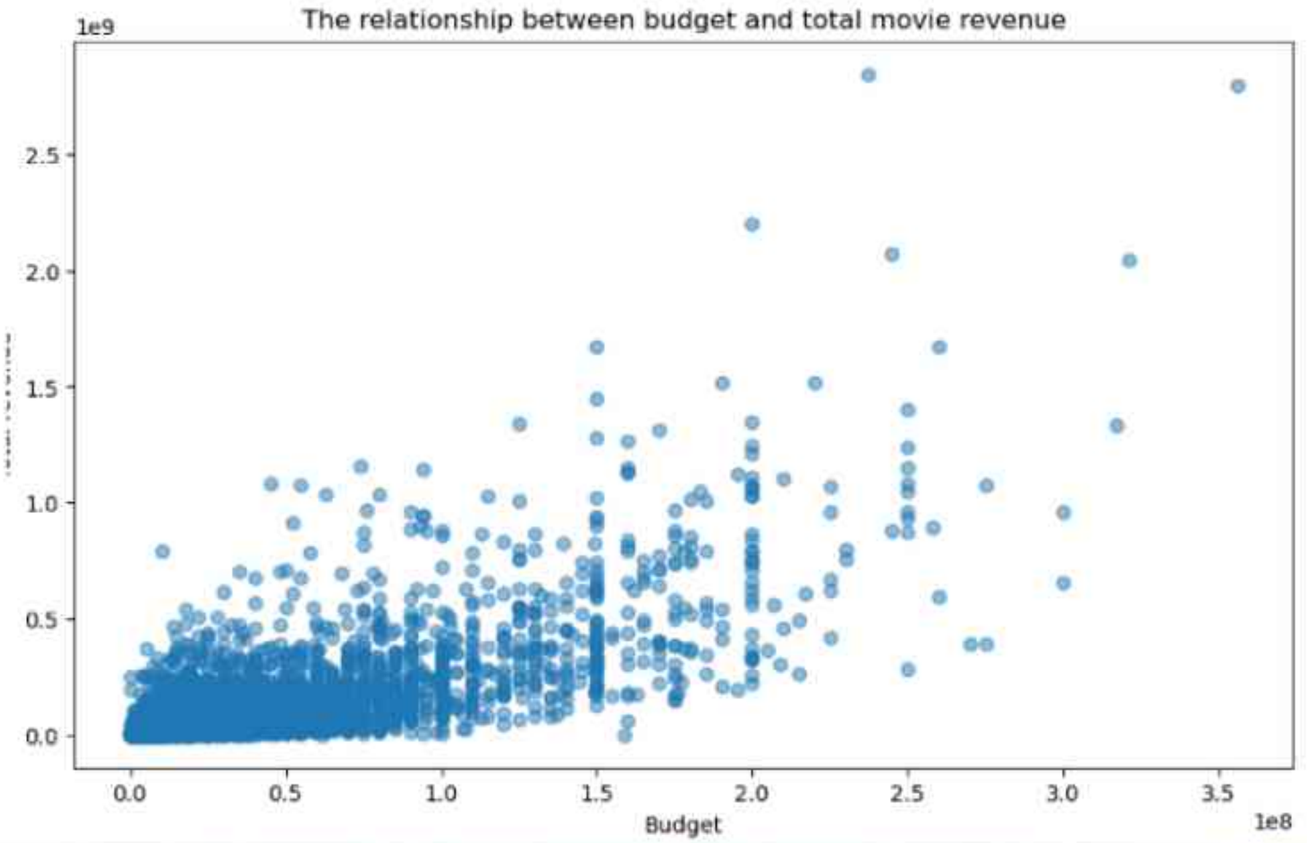
```
Average rating of films of various genres:
genre
History      8.300000
Musical      8.050000
Music        7.200000
Biography    7.030926
Animation    6.769231
Drama        6.693668
Crime        6.671506
Mystery      6.665000
Romance      6.410000
Family       6.363636
Adventure    6.291569
Sci-Fi       6.210000
Action       6.202817
Comedy       6.193987
Fantasy      6.006818
Thriller     5.912500
Sport        5.900000
Horror       5.750621
Western      5.666667
Name: score, dtype: float64
```

HYPOTHESES

Hypothesis 1: Films of certain genres have a higher average rating than films of other genres.

Hypothesis 2: Films with a higher budget usually have a higher total revenue.





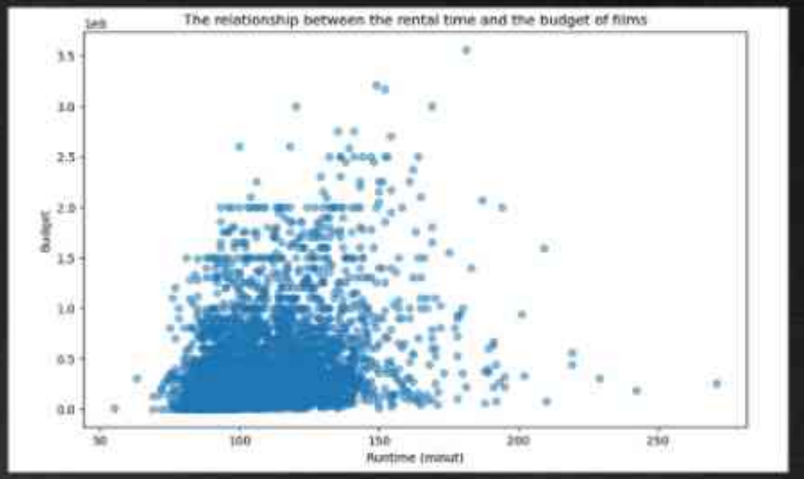

HYPOTHESES #3-4

HYPOTHESES

Hypothesis 3: Films with a long rental time usually have a higher budget.

Hypothesis 4: Films released in certain countries have different average ratings.

```python
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

data = pd.read_csv('movies.csv')

# Hypothesis 3: Films with a long rental time usually have a higher budget.

plt.figure(figsize=(10, 6))
plt.scatter(data['runtime'], data['budget'], alpha=0.5)
plt.title('The relationship between the rental time and the budget of films')
plt.xlabel('Runtime (minut)')
plt.ylabel('Budget')
plt.show()

# Hypothesis 4: Films released in certain countries have different average ratings.
country_ratings = data.groupby('country')['rating'].mean().sort_values(ascending=False)
print("Average rating of films in different countries:")
print(country_ratings)
```



HYPOTHESES

Hypothesis 3: Films with higher ratings usually have a higher total revenue.

Hypothesis 4: Films released in certain countries have different average ratings.

Average rating of films in different countrie

country	
Lebanon	8.400000
Libya	8.300000
Soviet Union	8.200000
Federal Republic of Yugoslavia	8.100000
Republic of Macedonia	8.000000
Iran	7.970000
Romania	7.900000
Turkey	7.866667
Indonesia	7.800000
Taiwan	7.742857
Brazil	7.700000
Argentina	7.687500
New Zealand	7.420000
Greece	7.400000
Colombia	7.400000
Kenya	7.400000
Russia	7.350000
Sweden	7.328000
Poland	7.275000
India	7.240323
Japan	7.224691
Yugoslavia	7.220000
West Germany	7.175000
South Korea	7.162857
Belgium	7.137500
Switzerland	7.130000
Denmark	7.100000
Spain	7.085106
Austria	7.020000
Hungary	6.983333
Norway	6.941667
Ireland	6.932558
Vietnam	6.850000
France	6.743728
Thailand	6.733333
Hong Kong	6.708889
Italy	6.704918
China	6.702500
Israel	6.700000
Australia	6.682609
Netherlands	6.666667
United Kingdom	6.642770
Chile	6.600000
Panama	6.600000
Malta	6.500000
South Africa	6.487500
Germany	6.436752
Iceland	6.400000
Portugal	6.350000
Mexico	6.345455
Philippines	6.333333
United States	6.257072
Jamaica	6.200000
Canada	6.134737
Serbia	6.100000
Czech Republic	6.075000
Finland	6.033333
United Arab Emirates	6.000000
Aruba	4.900000

Name: score, dtype: float64

STRATEGIC RECOMENDATIONS

- **Optimization of marketing strategies:** analyzing audience preferences will help the company choose the most effective channels for promoting films.

- **Developing genre diversity:** Researching audience preferences by genre will help the company expand its range of films and attract new audiences.

- **Optimizing the production budget:** Analyzing the relationship between budget and revenue will allow the company to make informed decisions about the allocation of funds for film production.

- **International expansion:** Given the differences in the average rating of films in different countries, the company may consider expanding into the international market and adapting content to local preferences.