



TEAM BOOMSLANG PRESENTS

MOVIE LOOPS

Avni, Carlos, and Sarah

Topics

- Cleaning the data
- Looking at genres being produced per decade
- Genre vs. Production Budget
- Genre vs. Profit
- Genre vs. Runtime
- Genre vs User Rating
- Budget vs User Rating



Where did the Data Come from?

CSV DATA SET FROM KAGGLE

Data set started out with 5000 films contained in a csv file

<https://www.kaggle.com/tmdb/tmdb-movie-metadata>

OMDB API CALLS

To get the IMDb user rating and MPAA rating information we used the OMDb API also used in class to add it to our dataframe

	budget	genres	homepage	id	keywords	original_language	original_title	overview	popularity	production_companies	production_countries	release_date	revenue
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}]	http://www.avatarmovie.com/	19995	[{"id": 1463, "name": "culture clash"}, {"id": 1464, "name": "culture clash"}]	en	Avatar	In the 22nd century, a paraplegic Marine is di...	150.437577	[{"name": "Ingenious Film Partners", "id": 289}], [{"name": "United States o..."}]	[{"iso_3166_1": "US", "name": "United States o..."}]	2009-12-18	2787965
1	300000000	[{"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}]	http://disney.go.com/disneypictures/pirates/	285	[{"id": 270, "name": "ocean"}, {"id": 726, "name": "ocean"}]	en	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	139.082615	[{"name": "Walt Disney Pictures", "id": 2}], [{"name": "United States o..."}]	[{"iso_3166_1": "US", "name": "United States o..."}]	2007-05-19	961000
2	245000000	[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}]	http://www.sonypictures.com/movies/spectre/	206647	[{"id": 470, "name": "spy"}, {"id": 818, "name": "spy"}]	en	Spectre	A cryptic message from Bond's past sends him o...	107.376788	[{"name": "Columbia Pictures", "id": 5}], [{"name": "United Kingdom..."}]	[{"iso_3166_1": "GB", "name": "United Kingdom..."}]	2015-10-26	880674
3	250000000	[{"id": 28, "name": "Action"}, {"id": 80, "name": "Adventure"}]	http://www.thedarkknighttrises.com/	49026	[{"id": 849, "name": "dc comics"}, {"id": 853, "name": "dc comics"}]	en	The Dark Knight Rises	Following the death of District Attorney Harve...	112.312950	[{"name": "Legendary Pictures", "id": 923}], [{"name": "United States o..."}]	[{"iso_3166_1": "US", "name": "United States o..."}]	2012-07-16	1084939
4	260000000	[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}]	https://movies.disney.com/john-carter	49629	[{"id": 818, "name": "based on novel"}, {"id": 819, "name": "based on novel"}]	en	John Carter	John Carter is a war-weary, former military ca...	43.926995	[{"name": "Walt Disney Pictures", "id": 2}], [{"name": "United States o..."}]	[{"iso_3166_1": "US", "name": "United States o..."}]	2012-03-07	284139
5	250000000	[{"id": 14, "name": "Fantasy"}, {"id": 28, "name": "Action"}]	http://www.sonypictures.com/movies/spider-man3/	559	[{"id": 851, "name": "dual identity"}, {"id": 852, "name": "dual identity"}]	en	Spider-Man 3	The seemingly invincible Spider-Man goes up ag...	115.699814	[{"name": "Columbia Pictures", "id": 5}], [{"name": "United States o..."}]	[{"iso_3166_1": "US", "name": "United States o..."}]	2007-05-01	890871
6	260000000	[{"id": 16, "name": "Animation"}, {"id": 10731, "name": "Animation"}]	http://disney.go.com/disneypictures/tangled/	38757	[{"id": 1562, "name": "hostage"}, {"id": 2343, "name": "hostage"}]	en	Tangled	When the kingdom's most wanted-and most charmi...	48.681969	[{"name": "Walt Disney Pictures", "id": 2}], [{"name": "United States o..."}]	[{"iso_3166_1": "US", "name": "United States o..."}]	2010-11-24	591794
7	200000000	[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}]	http://marvel.com/movies/movie/193/avengers_ag...	99061	[{"id": 8828, "name": "marvel comic"}, {"id": 8829, "name": "marvel comic"}]	en	Avengers: Age of Ultron	When Tony Stark tries to jumpstart a dormant p...	134.279229	[{"name": "Marvel Studios", "id": 420}], [{"name": "United States o..."}]	[{"iso_3166_1": "US", "name": "United States o..."}]	2015-04-22	1405403
8	250000000	[{"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}]	http://harrypotter.warnerbros.com/harrypottera...	767	[{"id": 616, "name": "witch"}, {"id": 2343, "name": "witch"}]	en	Harry Potter and the Half-Blood Prince	As Harry begins his sixth year at Hogwarts, he...	98.885637	[{"name": "Warner Bros.", "id": 6194}], [{"name": "United Kingdom..."}]	[{"iso_3166_1": "GB", "name": "United Kingdom..."}]	2009-07-07	933959
9	250000000	[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}]	http://www.batmanvsupermandawnofjustice.com/	209112	[{"id": 849, "name": "dc comics"}, {"id": 2092, "name": "dc comics"}]	en	Batman v Superman: Dawn of Justice	Fearing the actions of a god-like Super Hero l...	155.790452	[{"name": "DC Comics", "id": 429}], [{"name": "United States o..."}]	[{"iso_3166_1": "US", "name": "United States o..."}]	2016-03-23	873260

Cleaning

REMOVED UNNECESSARY COLUMNS

CLEANED GENRE COLUMN TO ONLY INCLUDE
THE FIRST GENRE LISTED

SPLIT THE RELEASE DATE INTO MONTH,
DAY AND YEAR COLUMNS

WE ALSO CUT THE DATA OFF AT 1960 SO
ANYTHING BEFORE THAT WAS DROPPED.
WE ALSO DROPPED ANY FILM THAT HAD
A BUDGET LISTED UNDER \$1M OR THAT
WAS FORMATTED INCORRECTLY

```
#Clean Genre column
for i in range(len(movie_df_4)):
    try:
        movie_genre_2 = list(movie_df_4["Genre dirty"][i].split(" "))
        movie_genre_2_filter = filter(str.isalnum, movie_genre_2[3])
        movie_genre_3 = "".join(movie_genre_2_filter)
        movie_genre_4.append(movie_genre_3)
    except (IndexError, KeyError):
        movie_genre_4.append("No Genre")

#add new column to DF
movie_df_4["Genre"] = movie_genre_4
movie_df_4=movie_df_4.drop(columns=['Genre dirty'])
```

► MI

```
#Split date into 3 columns
for i in range(len(movie_df_4)):
    try:
        split_date = movie_df_4["Release Date"][i].split("-")
        year.append(int(split_date[0]))
        month.append(int(split_date[1]))
        day.append(int(split_date[2]))
    except (KeyError):
        count.append(i)
```

API Calls

```
▶ M1  
  
for i in range(len(movie_df_5)):  
    try:  
        response = requests.get(url + movie_df_5["Title"][i] + api_key)  
        data = response.json()  
        mpaa_rating.append(data["Rated"])  
    except:  
        mpaa_rating.append("No Rating")  
    try:  
        user_rating.append(data["imdbRating"])  
    except:  
        user_rating.append("No Rating")  
print(len(user_rating))  
print(len(mpaa_rating))  
print(i)
```

```
2934  
2934  
2933
```

```
▶ M1  
  
movie_df_5["User rating"] = user_rating  
movie_df_5["Rated"] = mpaa_rating  
movie_df_5
```

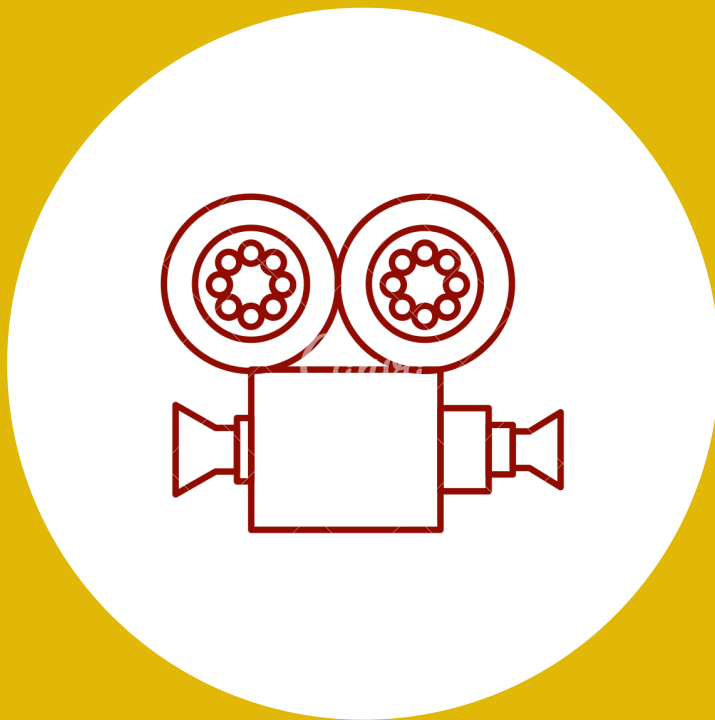
Using the OMDb API we were able to get the IMDb user rating as well as the MPAA rating of each film and added them to the dataframe

At this point we had 2934 films in our dataset

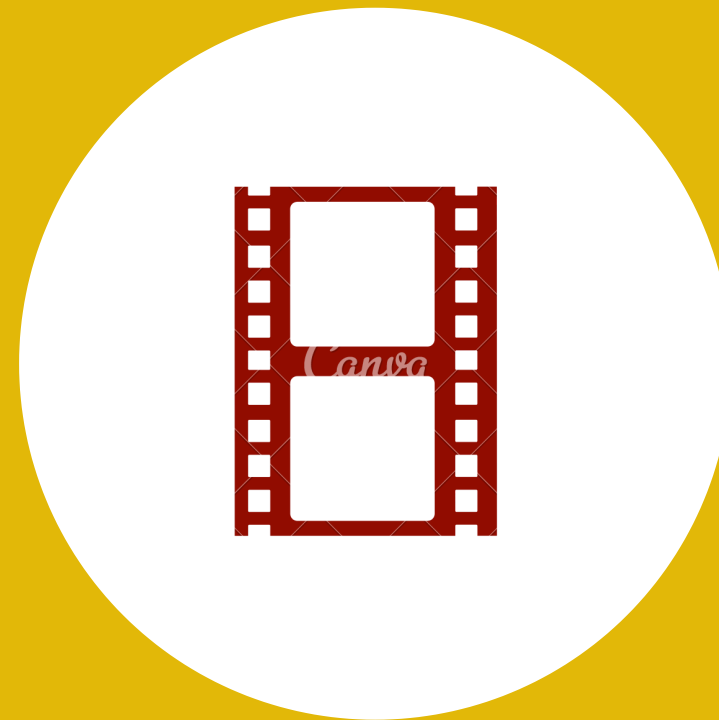
We also added a Profit column by subtracting the Budget column from the Revenue column

this is a very rudimentary profit figure because we don't account for advertising, points, etc. This is just based on the public figures available

Separating the Data



Separated the data using bins to allocate genres by decade



Used Groupby to find the mean runtime and user rating per genre



Plotted the data to get visual representation of the relationships we were trying to look at



Genre

ACCORDING TO RICK ALTMAN GENRE IS DEFINED USING
SEMANTIC AND SYNTACTIC ELEMENTS

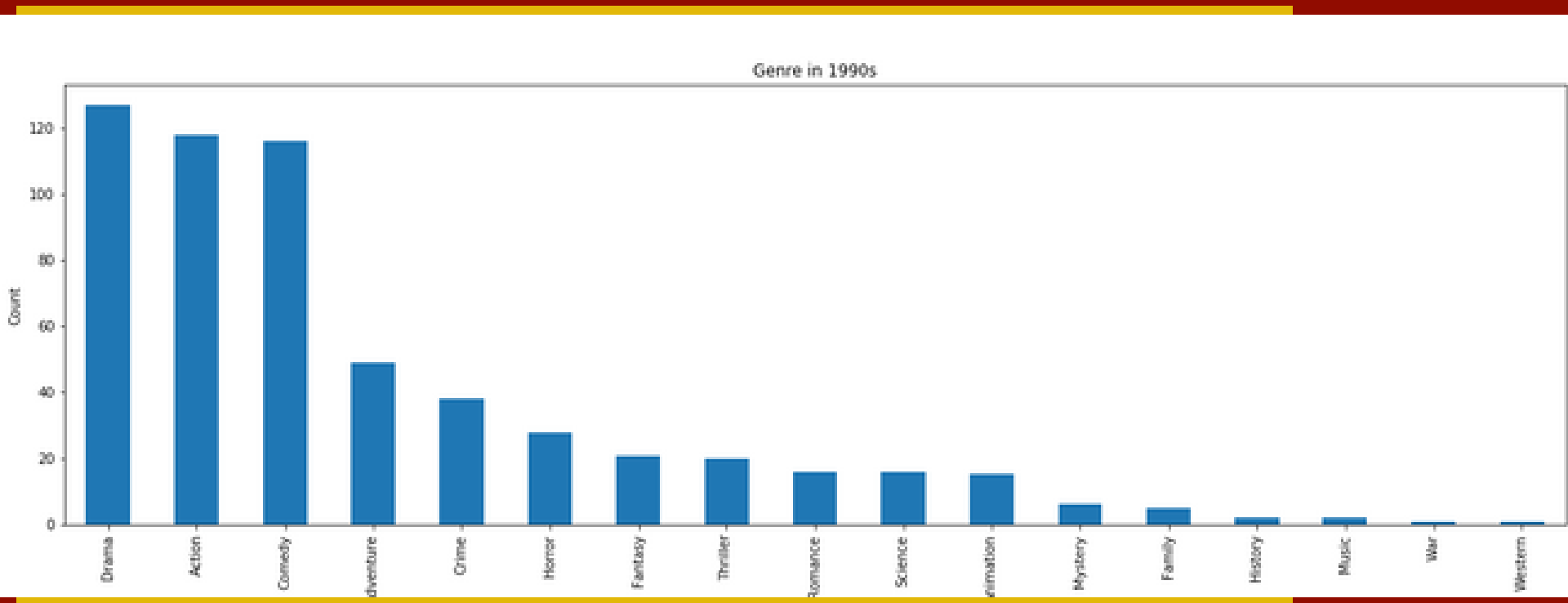
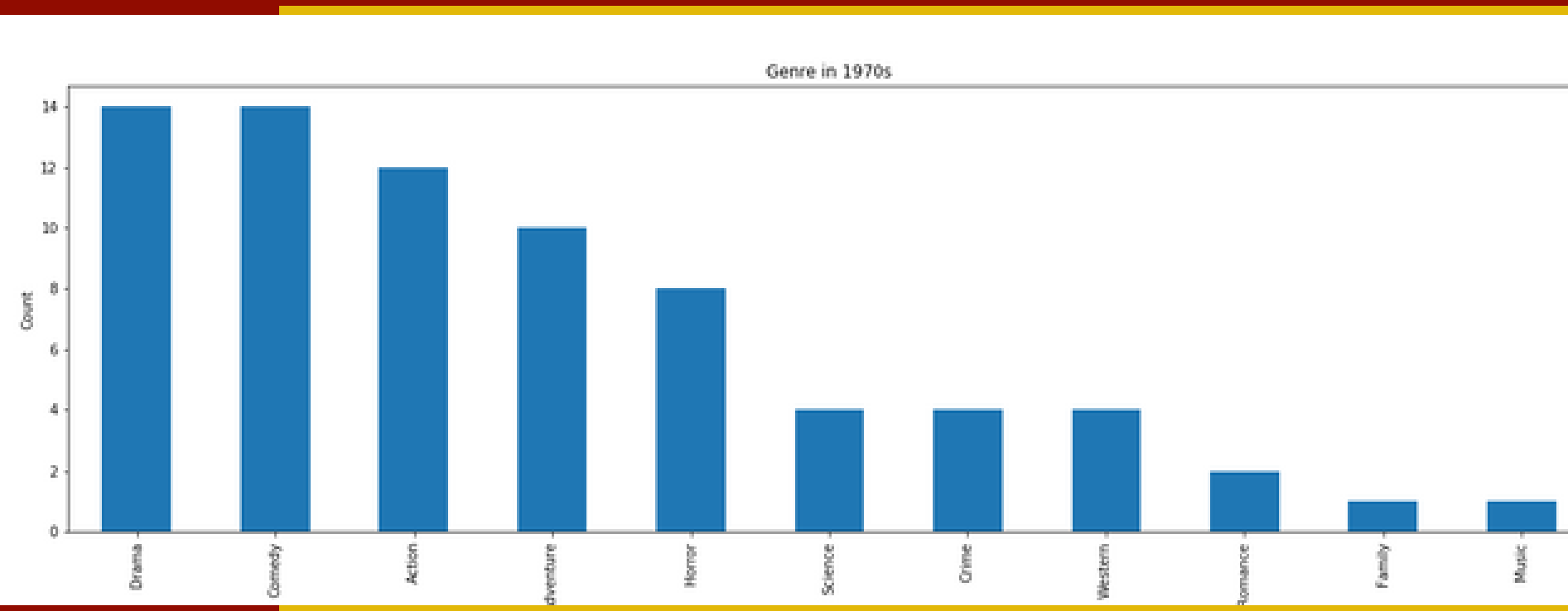
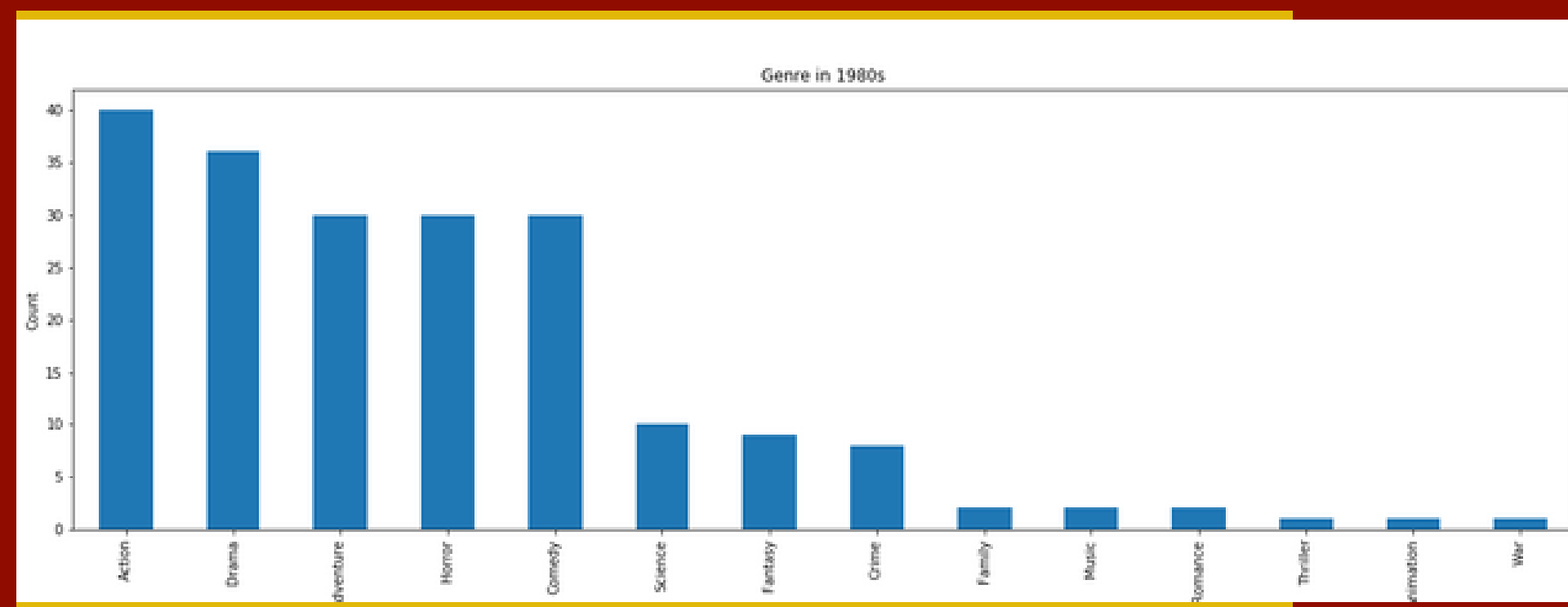
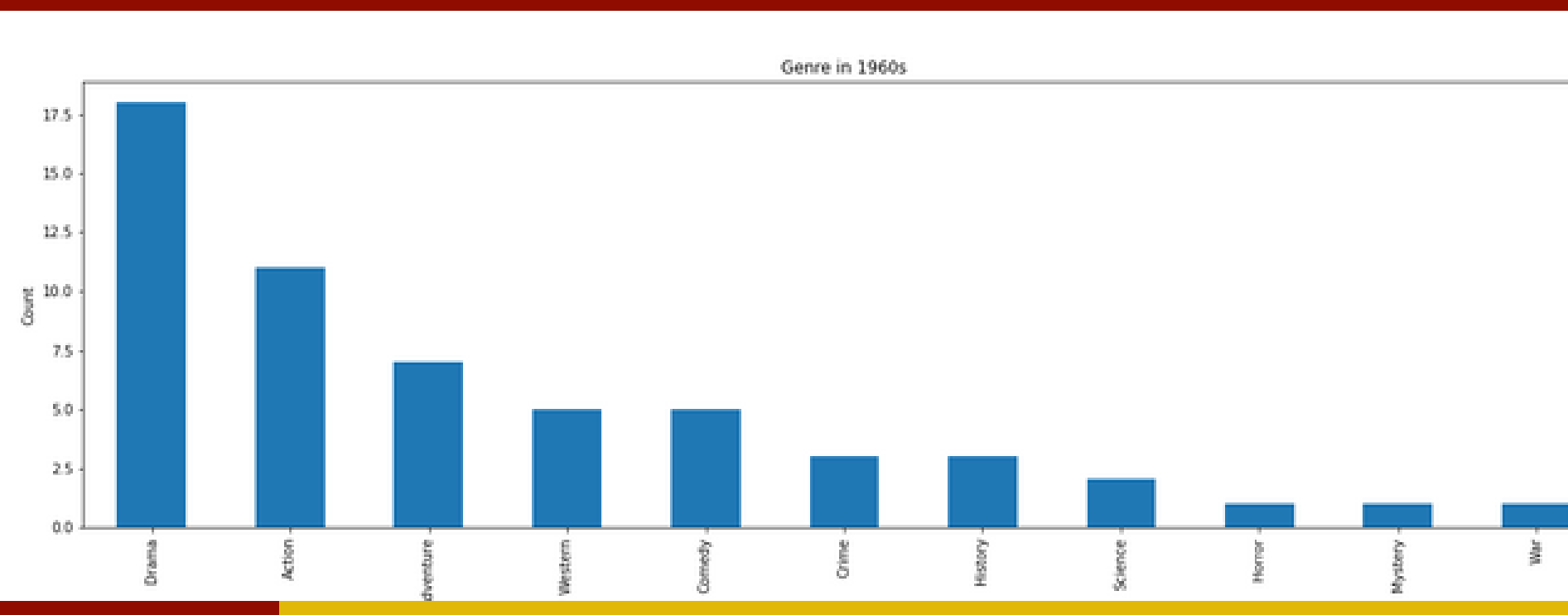
Semantic Elements

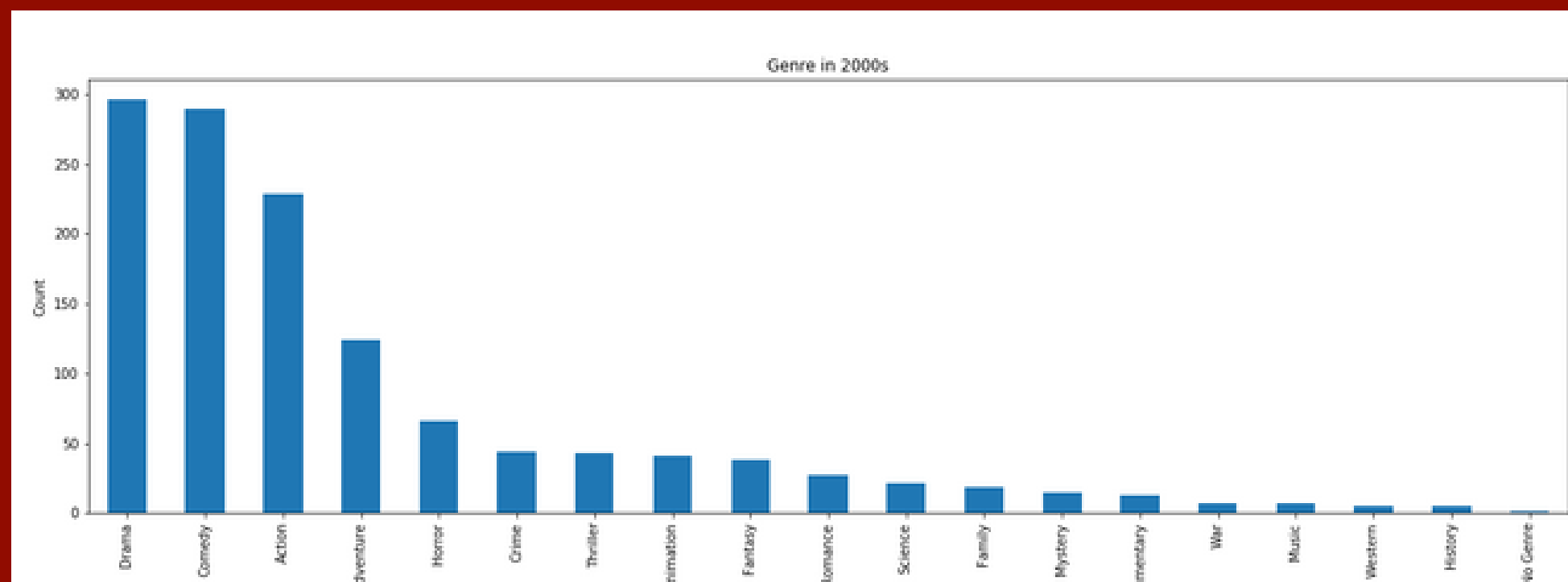
"A LIST OF COMMON TRAITS, ATTITUDES, CHARACTERS, SHOTS, LOCATIONS, SETS," I.E. THE GENRE'S "BUILDING BLOCKS THEMSELVES" (10).

Syntactic Elements

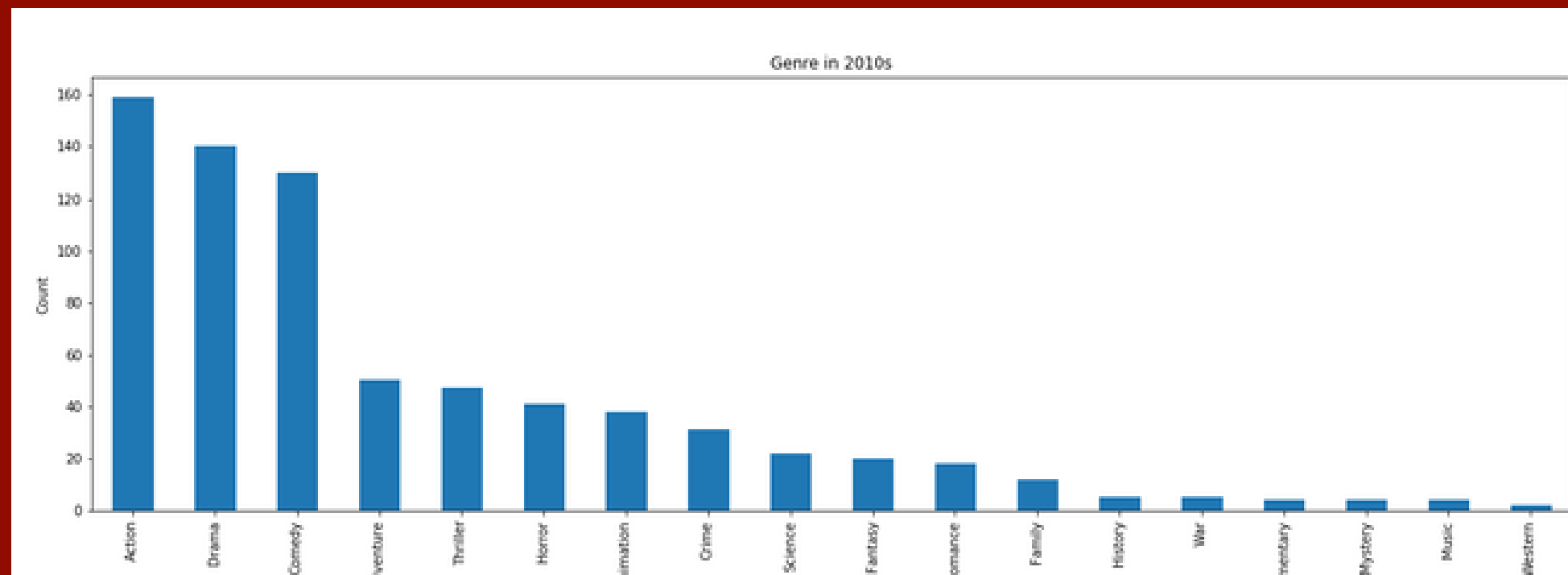
"CONSTITUTIVE RELATIONSHIPS BETWEEN UNDESIGNATED AND VARIABLE PLACEHOLDERS," OR THE "STRUCTURES INTO WHICH [THE BUILDING BLOCKS] ARE ARRANGED" (10).

GENRES OVER THE DECADES



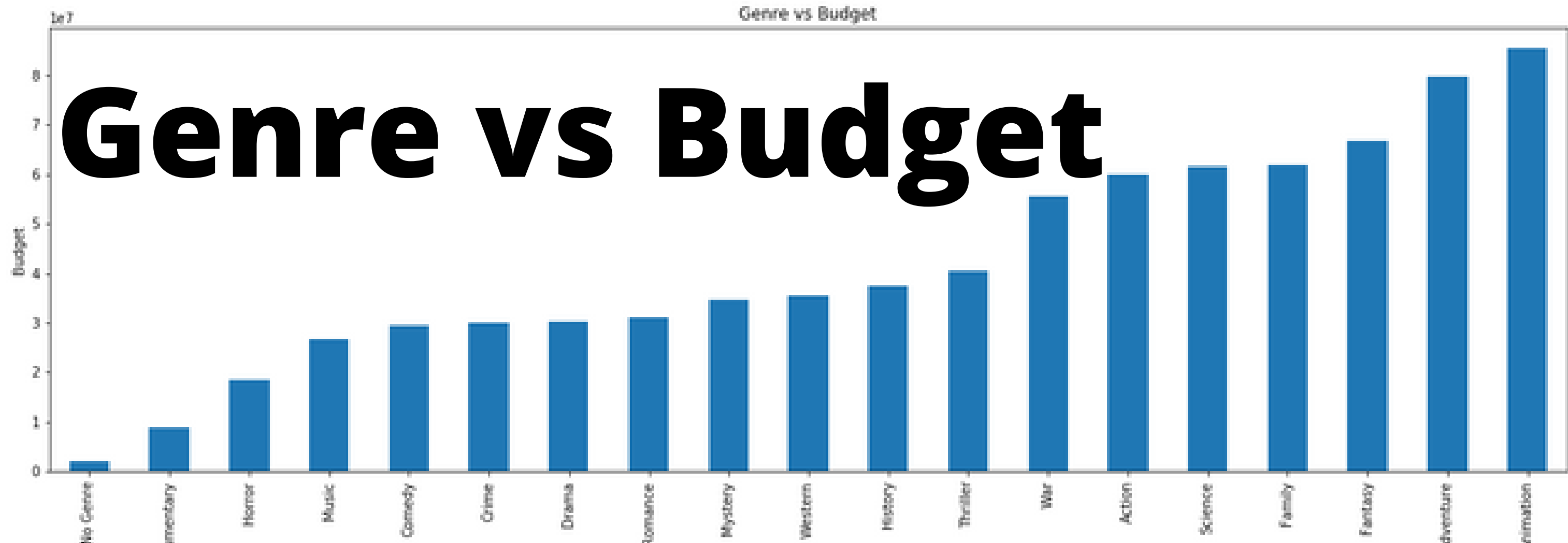


Action, Drama, Comedy and Adventure are the top four genres present from 1960 to 2017.

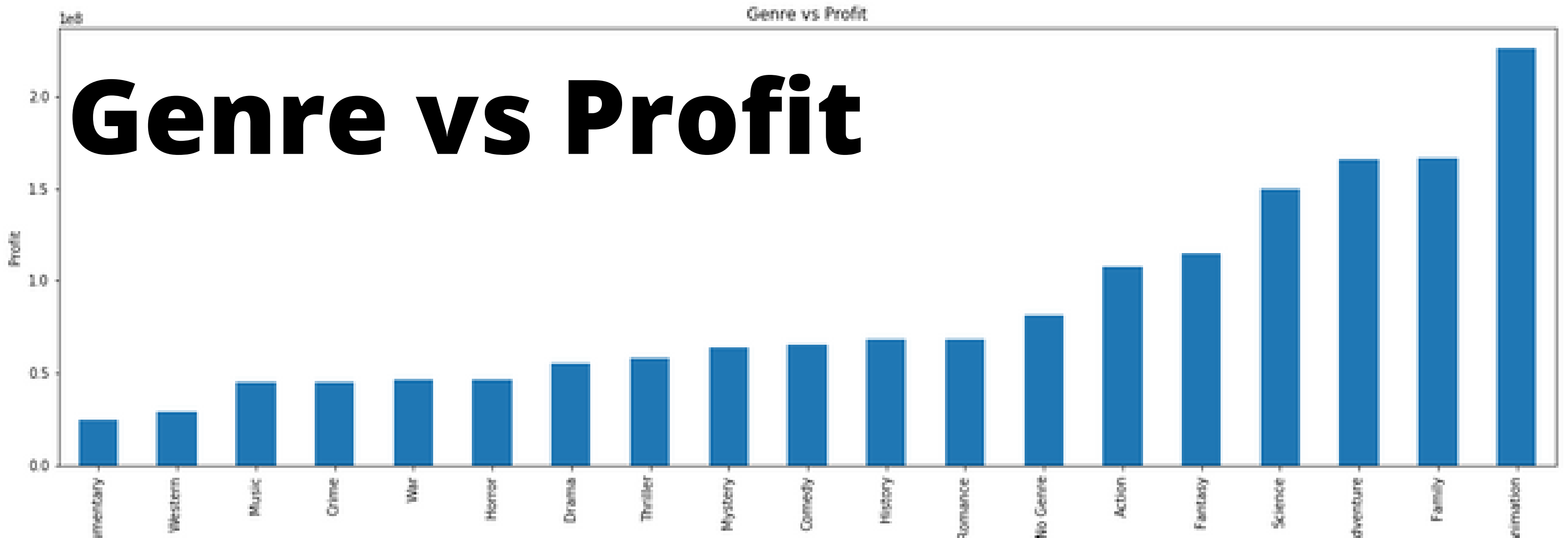


There is also a increase in the number of films being made as time passes.

Genre vs Budget



Genre vs Profit



Genre and \$\$\$

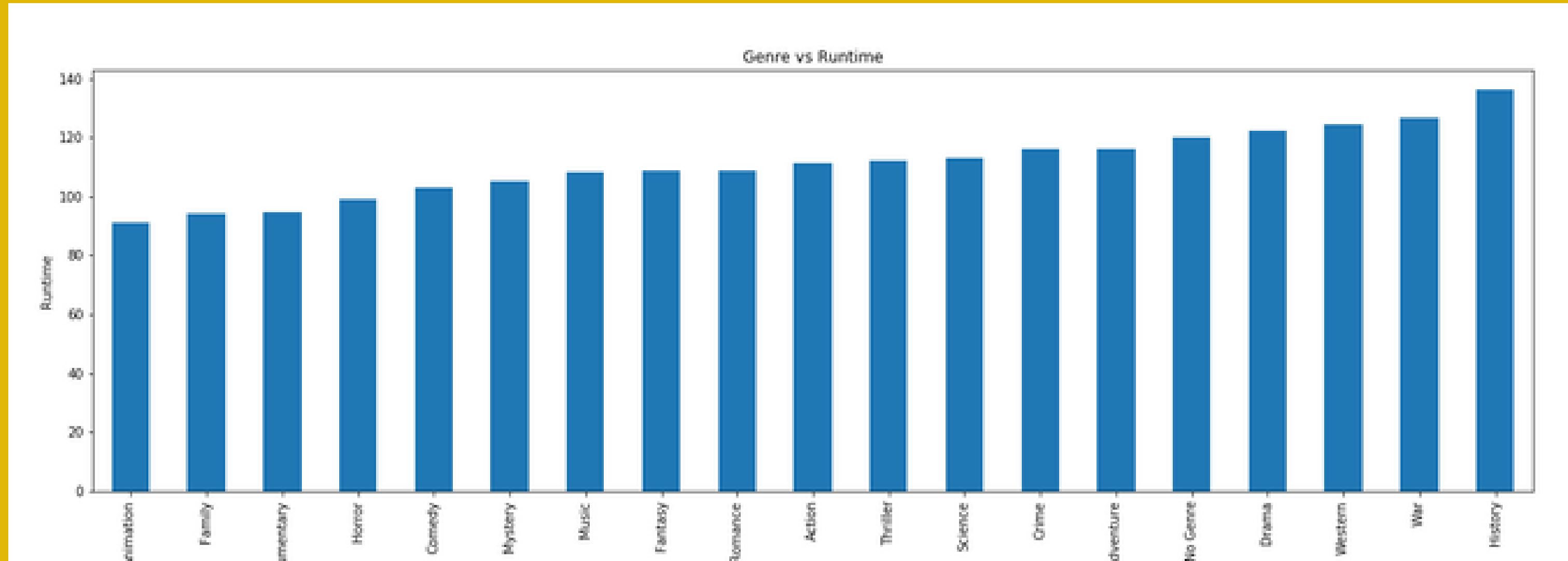
Animation costs on average the most, but also brings in the most profit.

Documentaries are shown to cost less on average but in turn are less profitable

Horror has a traditionally low budget but tend to make a decent profit

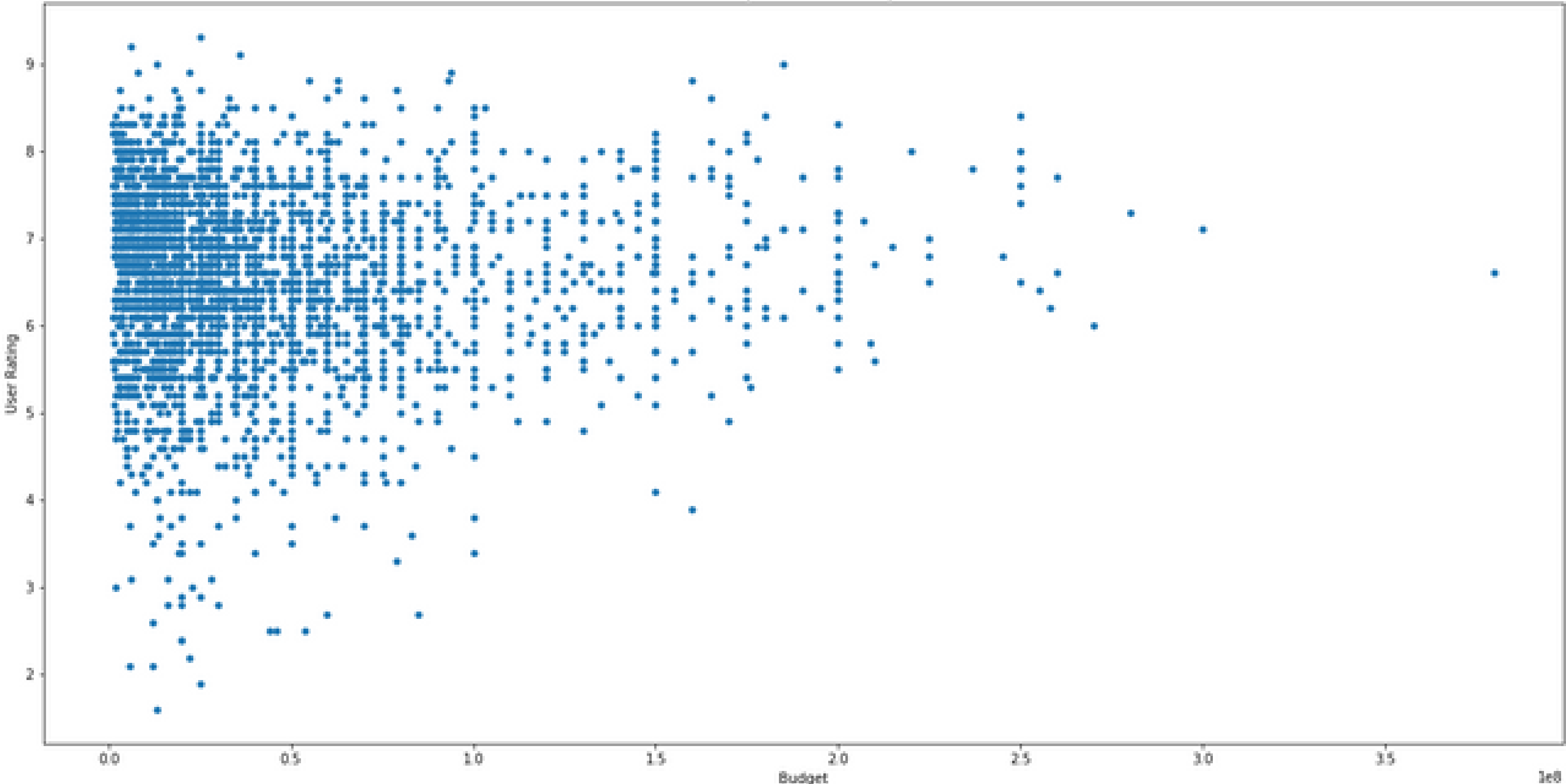
We are not taking into account genre hybrids in these graphs



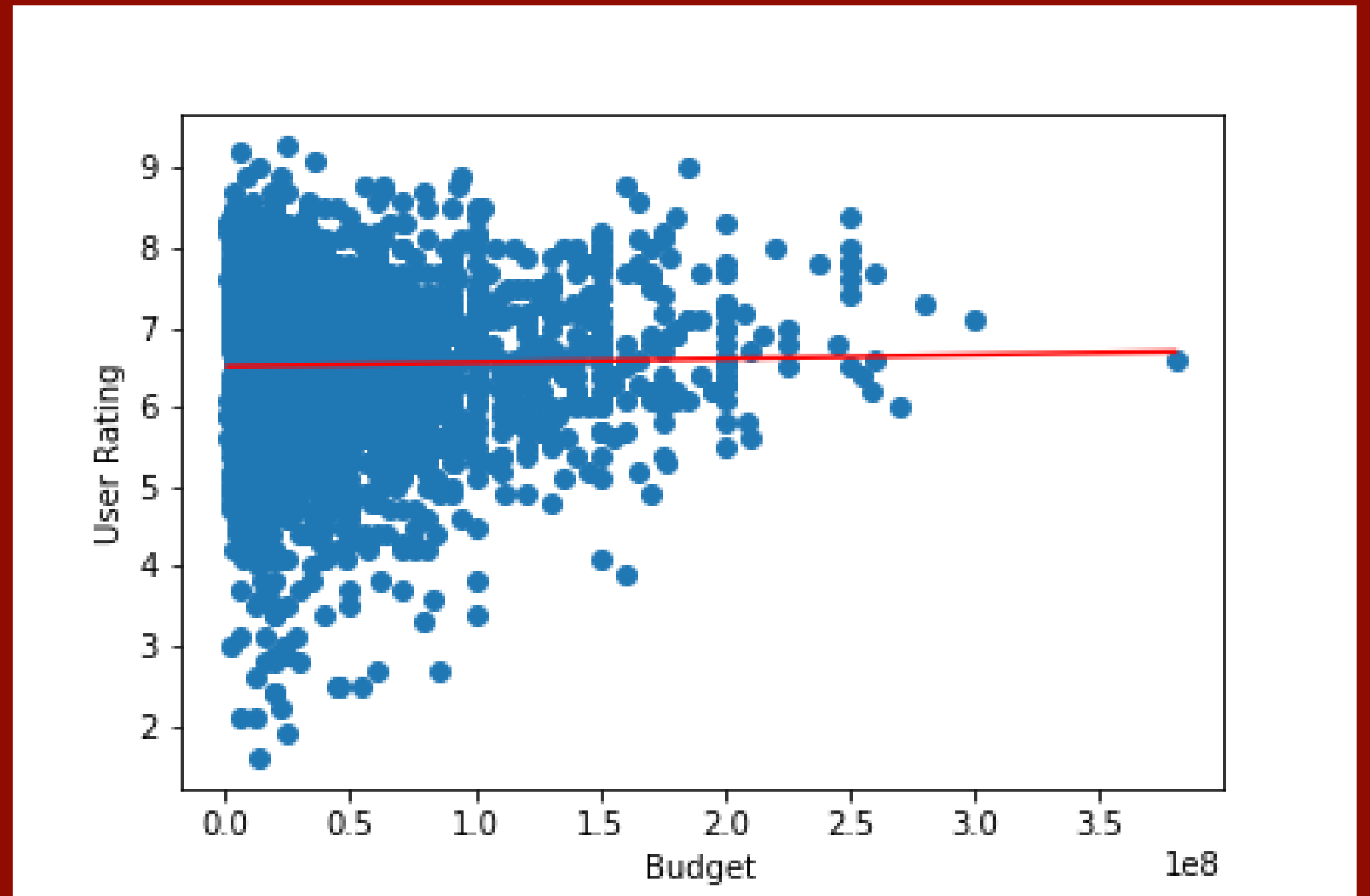


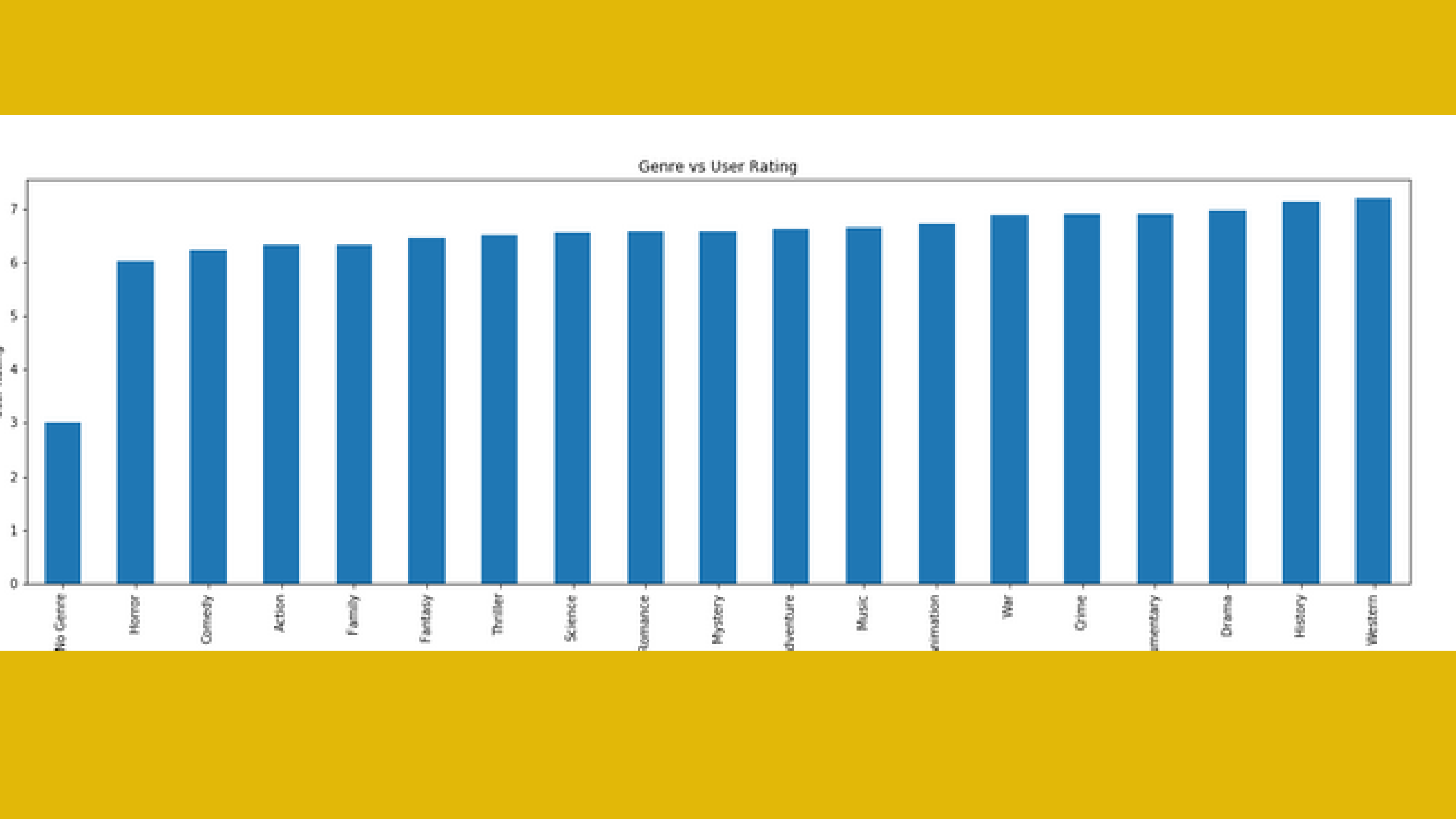
Animation and Family films have the shortest runtime while War and History Films run the longest

Budget vs User Rating



No
correlation in
this dataset
between how
much a film
cost and the
reception of
viewers







What does the viewer prefer?

The average viewer rating shows that genre as a whole does not indicate a preference of one genre over another all that much. This could be because despite fewer being made the Western may have more consistent ratings where Action films' ratings may be all over the place.