Analysis of Movie Ratings

Introduction:

The dataset used for this project is Movielens dataset. The dataset consists of 1 Million ratings applied to 3900 movies from 6040 users. The datasets which helps the project are as follows:

1) Movies.dat:
   It contains information such as movieID, movie name, year of release and genre
2) Users.dat:
   It contains information such as UserID, Gender, age, job_code, zipcode
3) Ratings.dat:
   It contains information such as UsersID, movieID, rating
4) TMDB_crew.csv:
   It contains information such as cast, crew, movie
5) TMDB_movies.csv:
   It contains information such as budget, revenue, tag, release date.

Data Preprocessing:

The original data was in the format of 'dat'. These files were converted to csv files. Then they are merged together. Irrelevant columns were removed. Missing were removed. Proper column headers were added. Relevant information were extracted and data is modified according to the requirement.

Data Analysis:

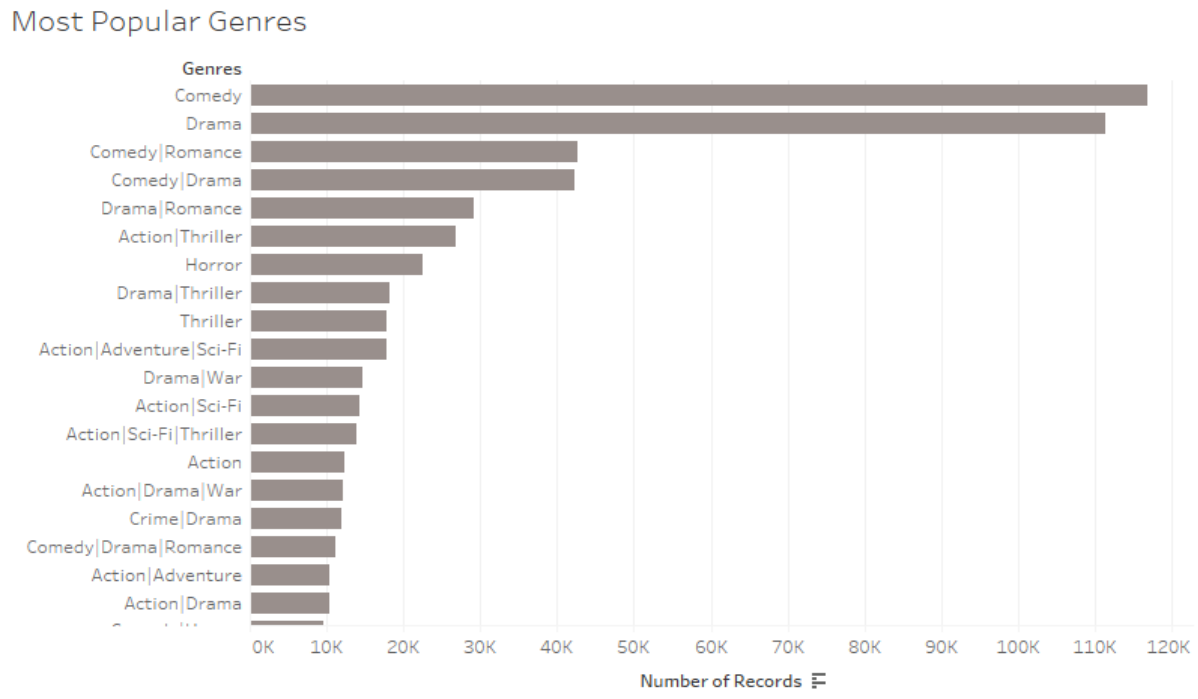


Fig.1

The Fig.1 depicts the bar graph representing Genre vs Number of records. It shows the most popular genres. We can find that Drama is the most popular genre followed by Comedy.

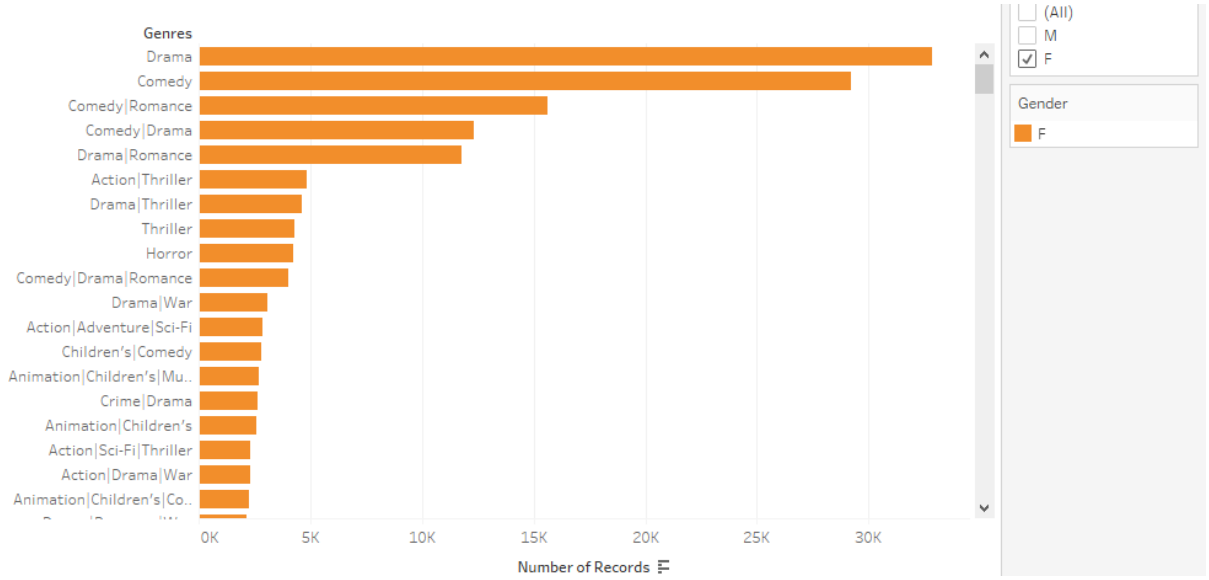


Fig.2

The Fig.2 depicts the bar graph representing Genre vs Number of records filtered by Female. It shows the most popular genres for the female population. We can find that males are more interested towards drama films.

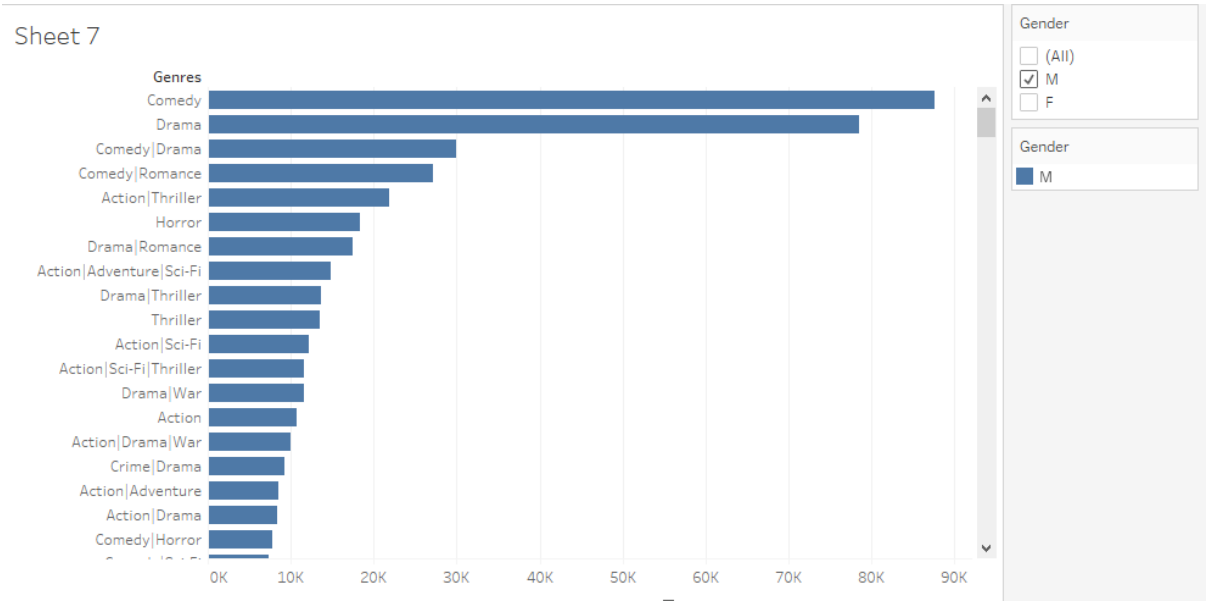


Fig.3

The Fig.3 depicts the bar graph representing Genre vs Number of records filtered by Male. It shows the most popular genres for the male population. We can find that males are more interested towards comedy films.

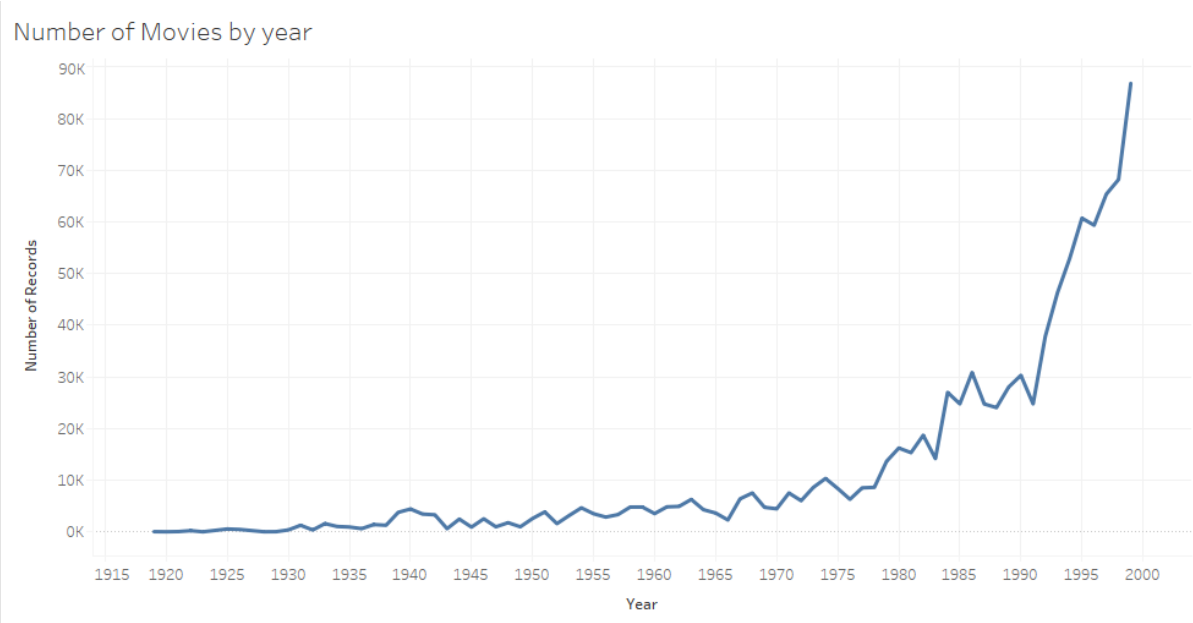


Fig. 4

The Fig.4 demonstrates the Year vs Number of records graph. The production of movies is increasing with each year.

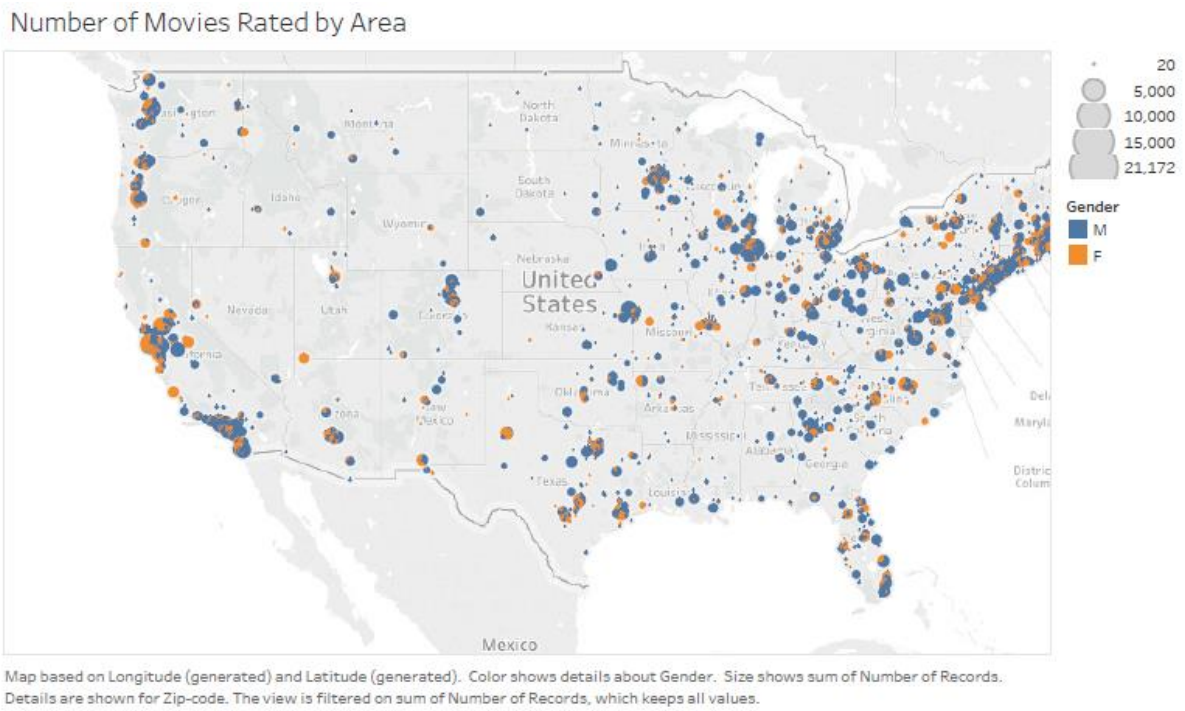


Fig.5

The Fig. 5 represents the number of movies rated based on the locality. We can find that people belonging to the east are watching most movies.

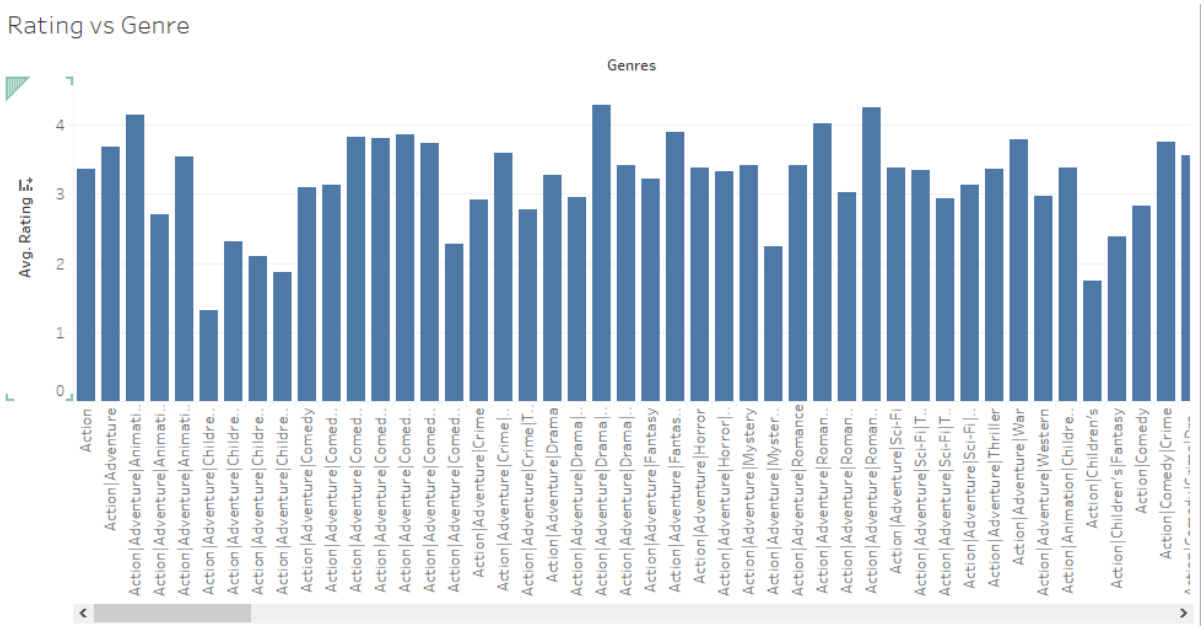


Fig. 6

The Fig. 6 represents the rating vs genre bar graph.

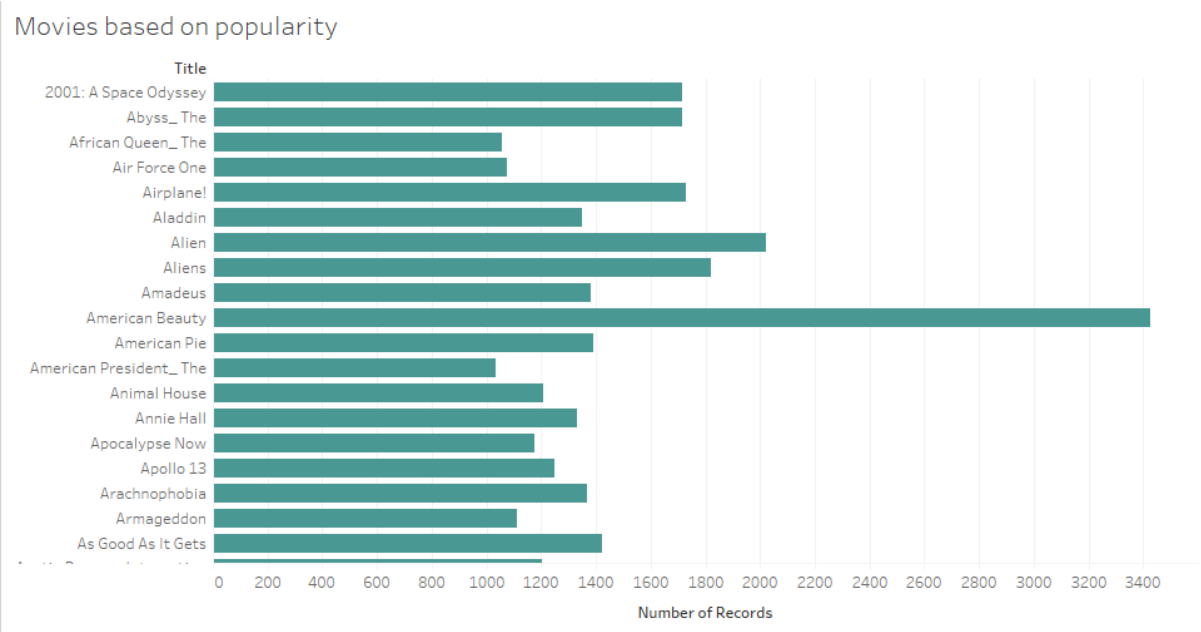


Fig.7

The Fig. 7 shows the movies according to their popularity.

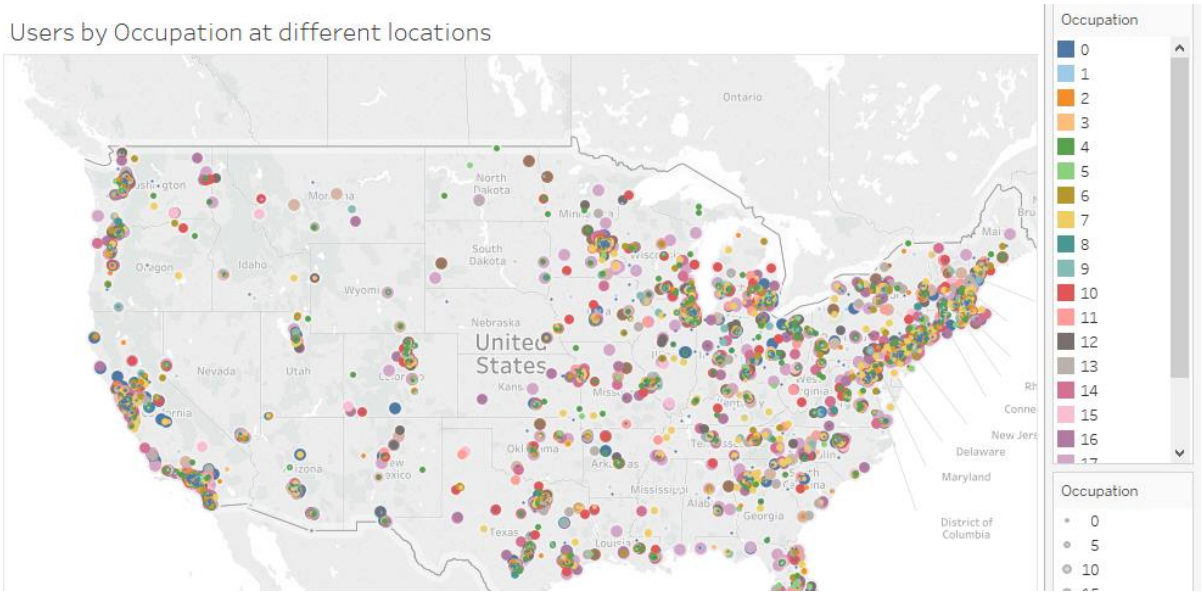


Fig. 8

The Fig.8 shows the locations at where the people of various occupations are located at.

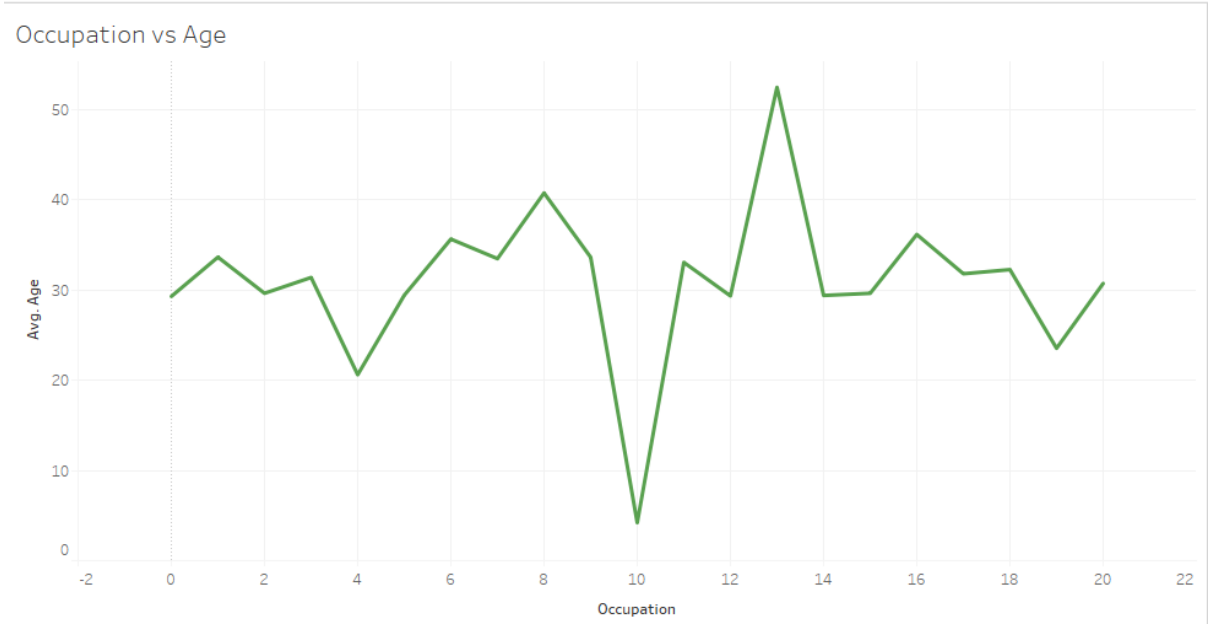


Fig. 9

From Fig. 9 we can observe that the average people who are having job code 14 are the older people where as, people having job code 10 are younger.

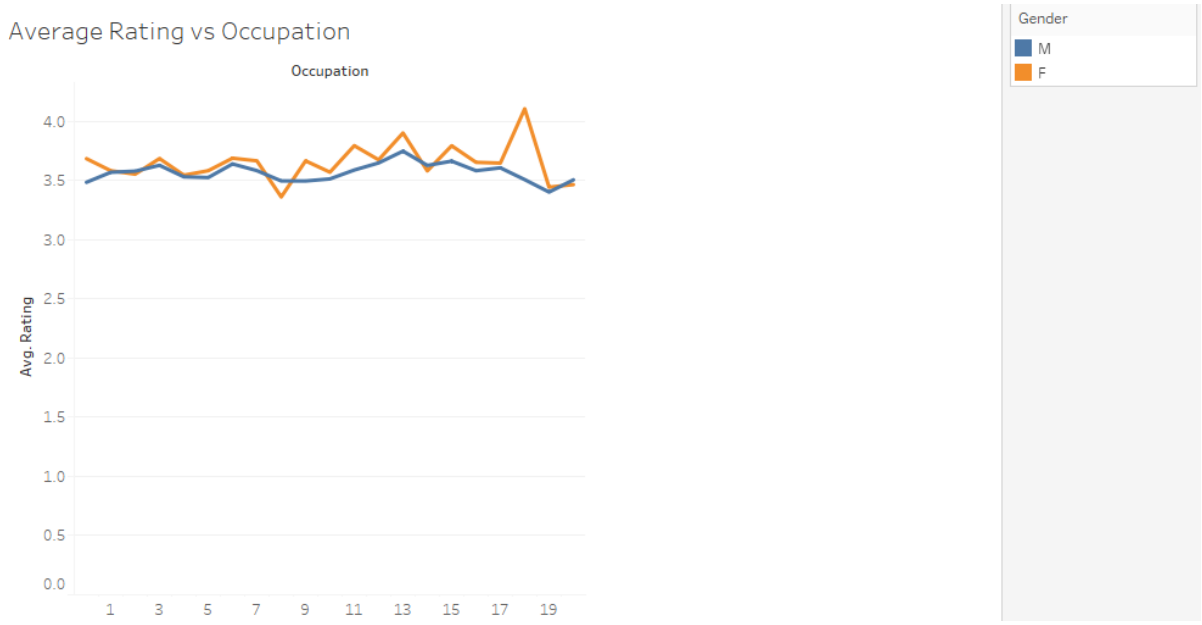


Fig.10

Fig. 10 represents average rating vs occupation graph

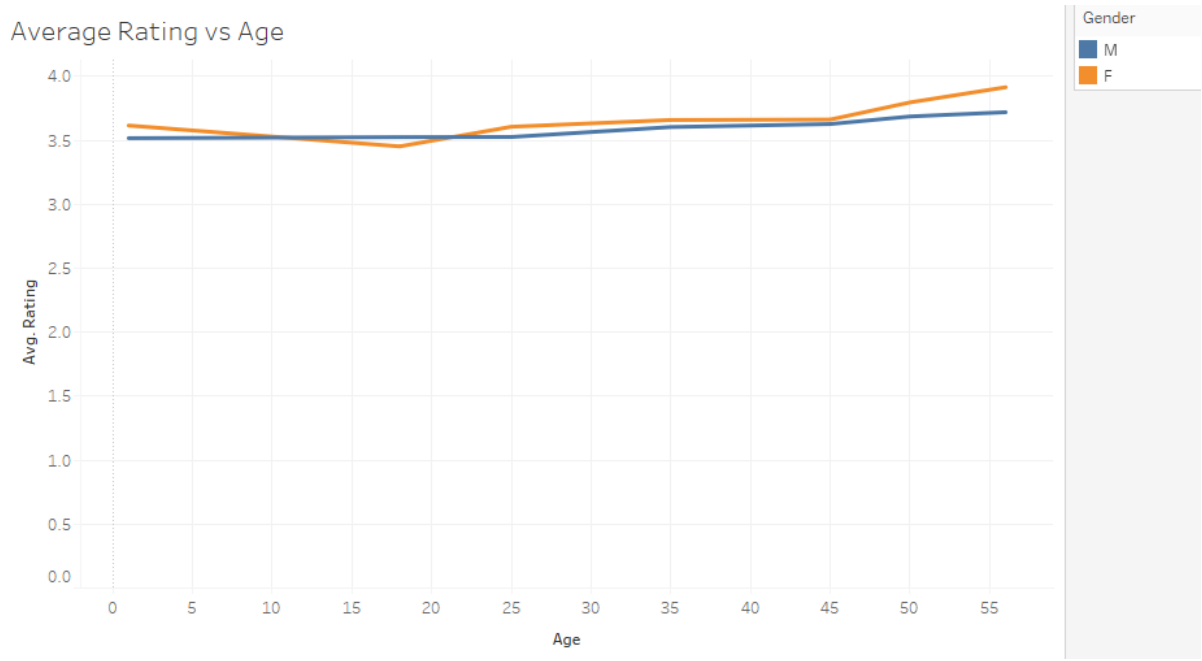


Fig.11

Fig. 11 represents average rating vs occupation graph

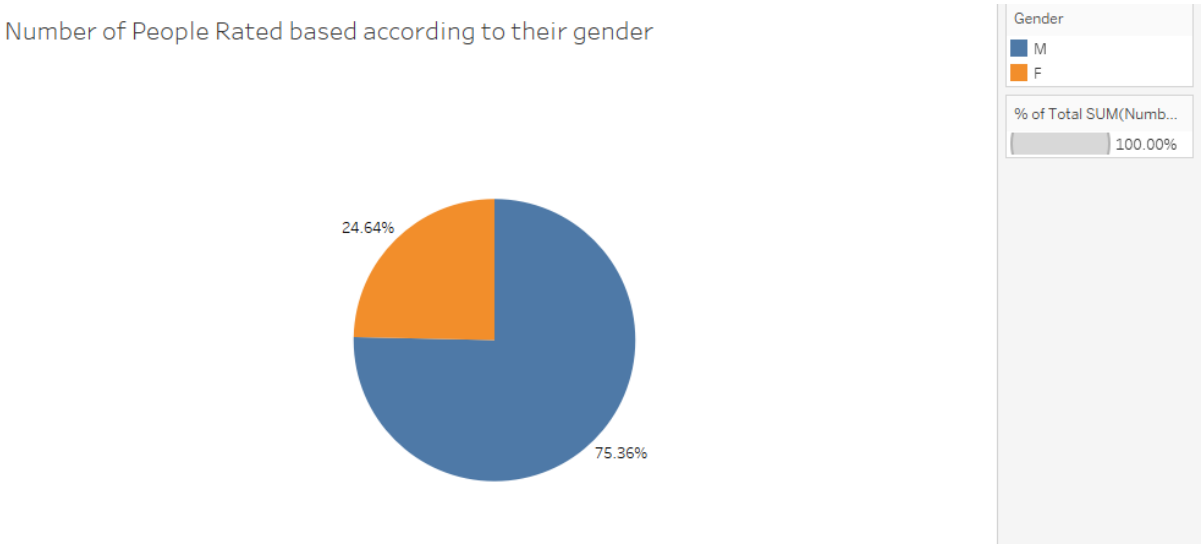


Fig.12

From the pie chart in Fig. 12 we can observe that very large number of males are rating the movies and less number of females are rating.

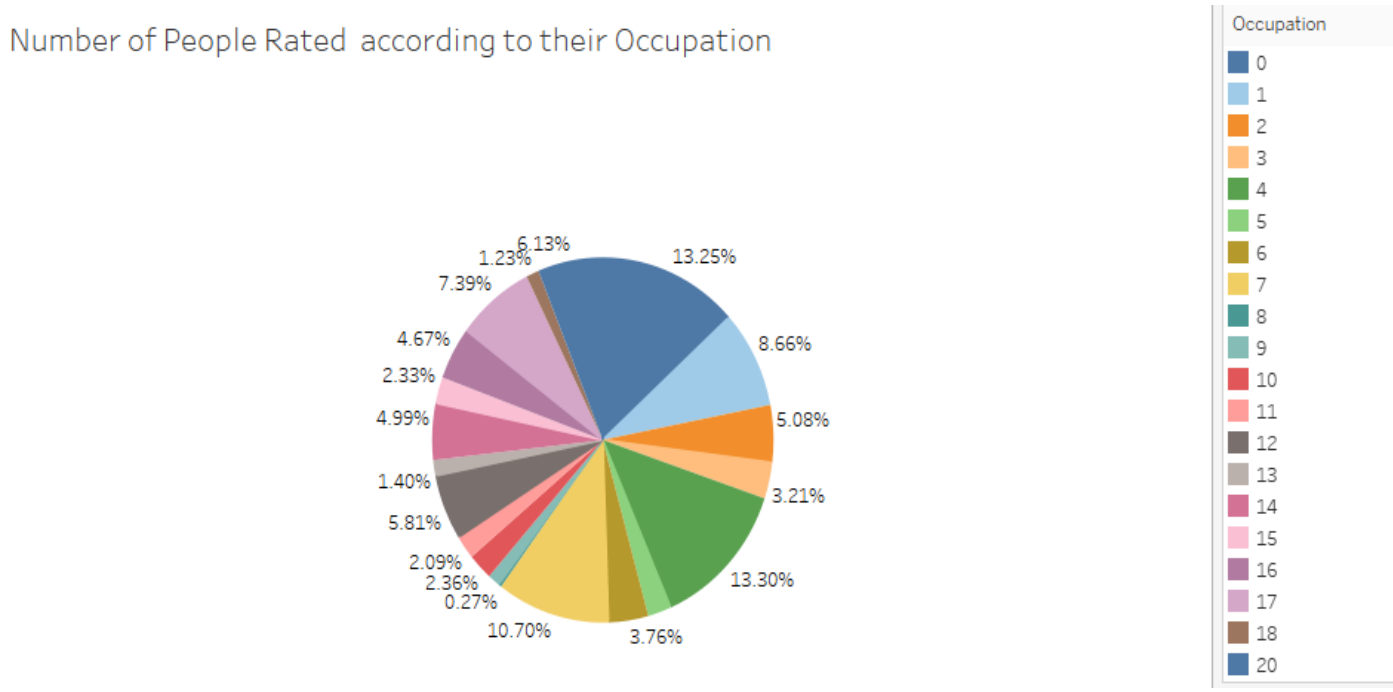


Fig.13

From the above fig. we can observe that people with no job are watching more movies.

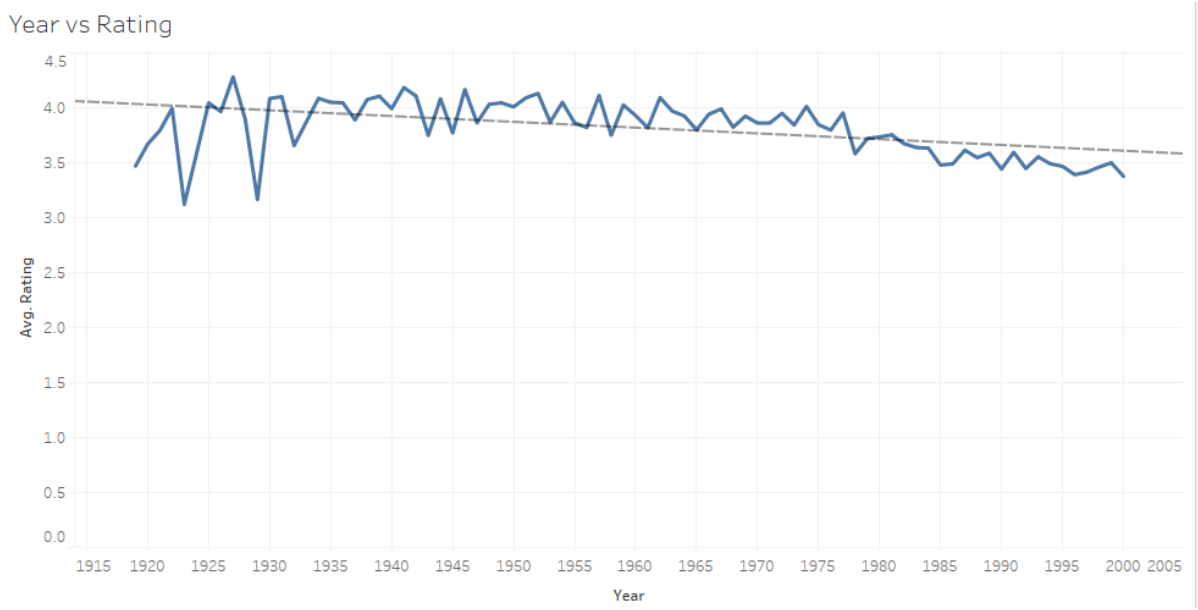
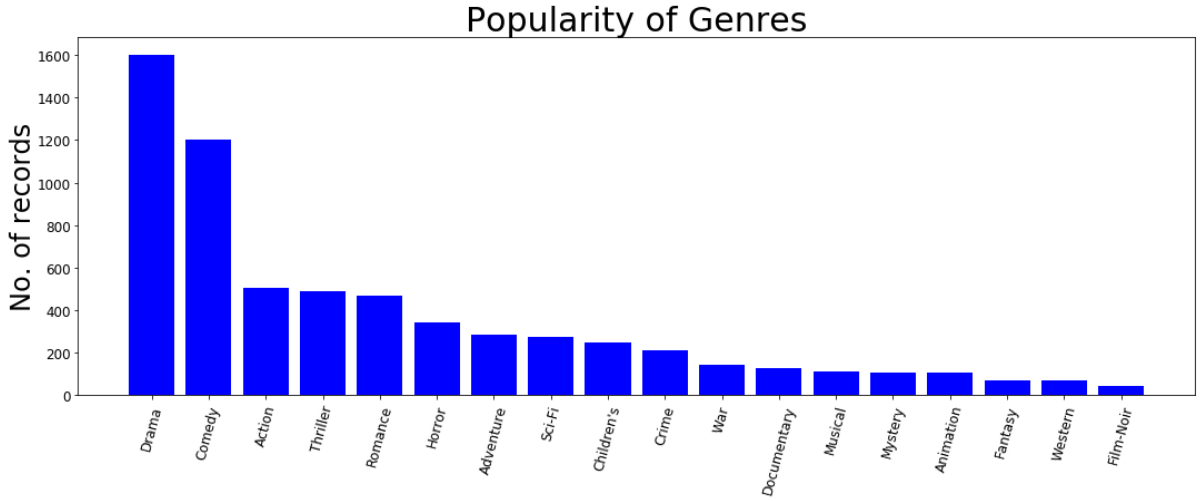
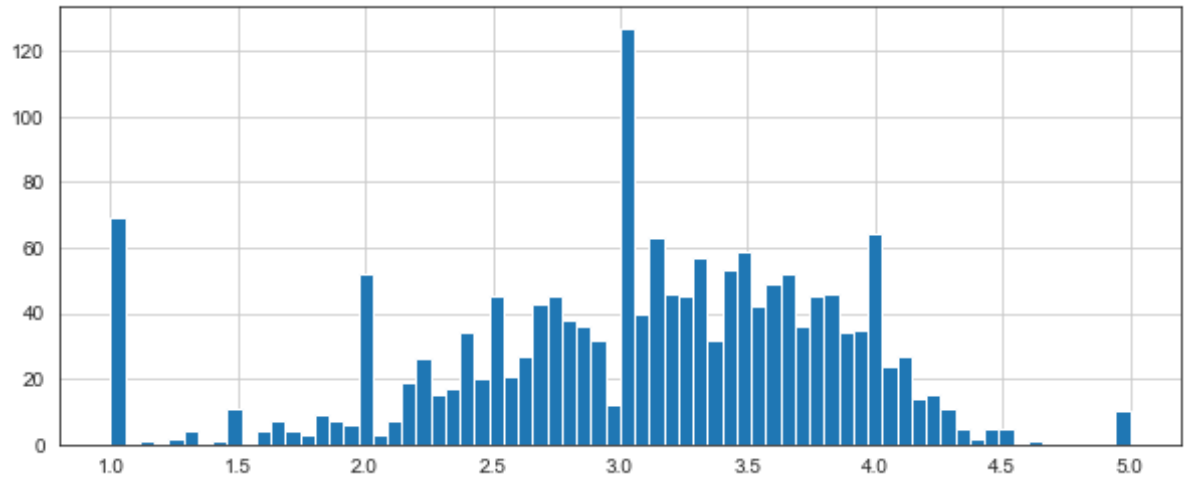


Fig. 14

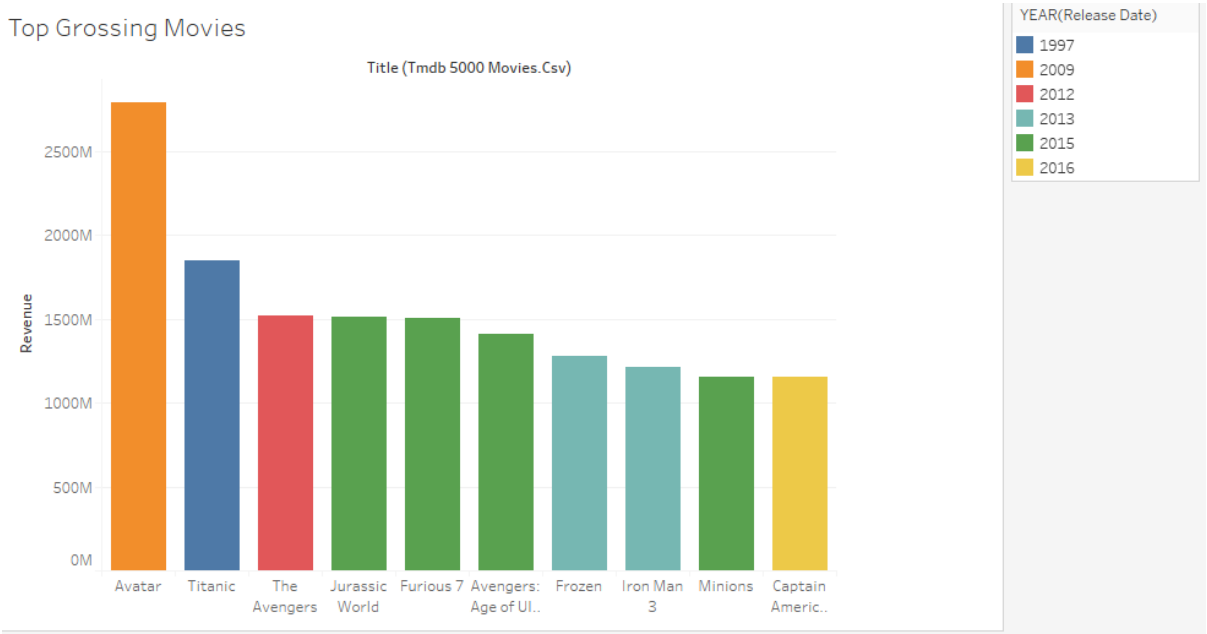
The rating is decreasing with each year



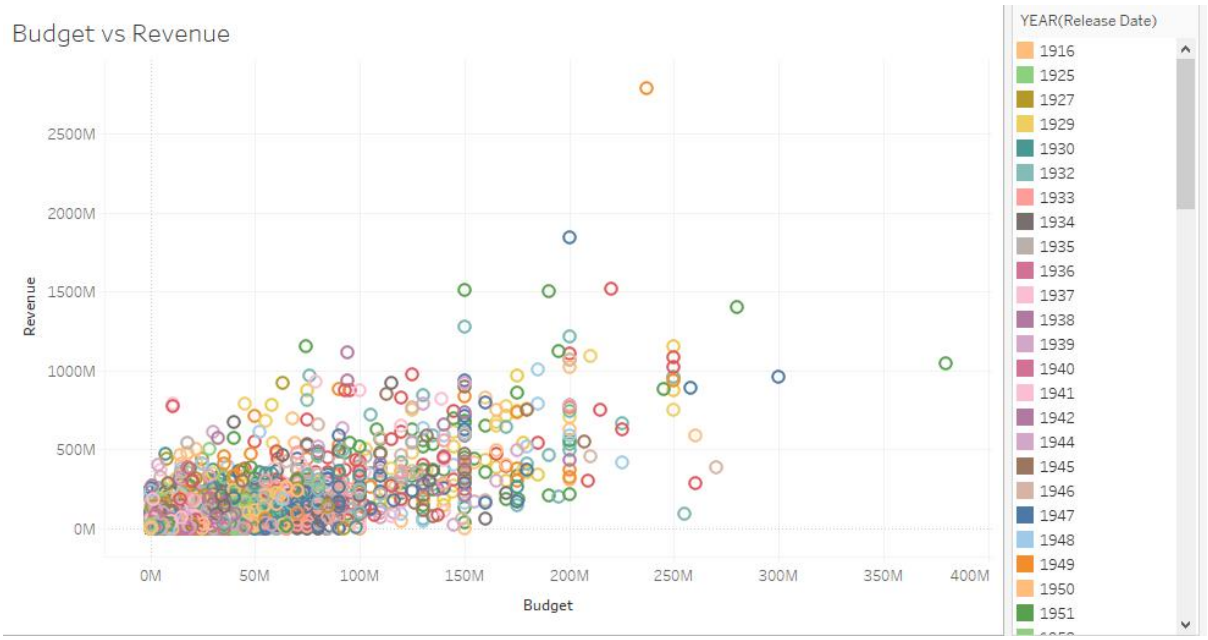


Rating vs Number of Records

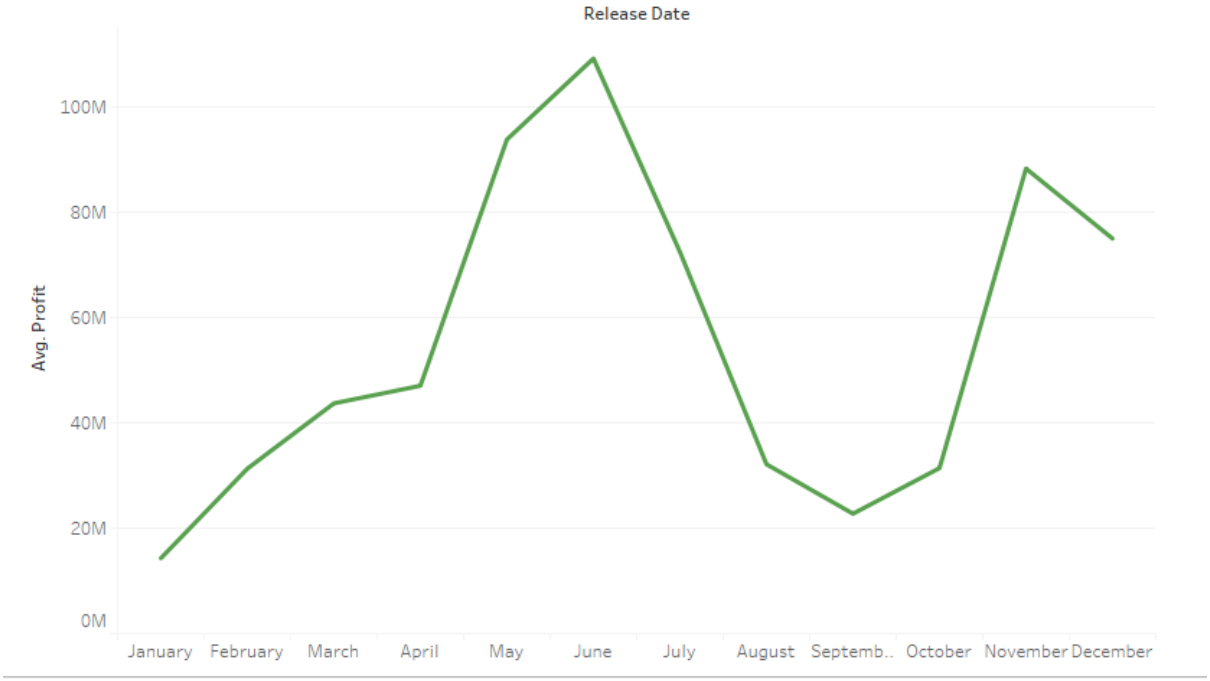
Top Grossing Movies



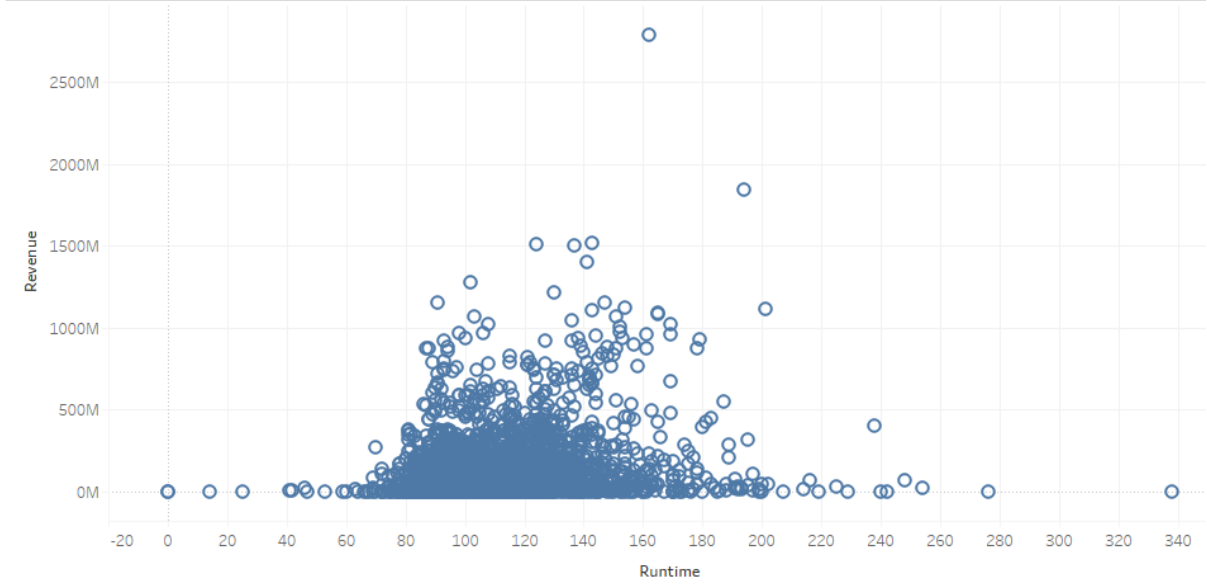
Budget vs Revenue



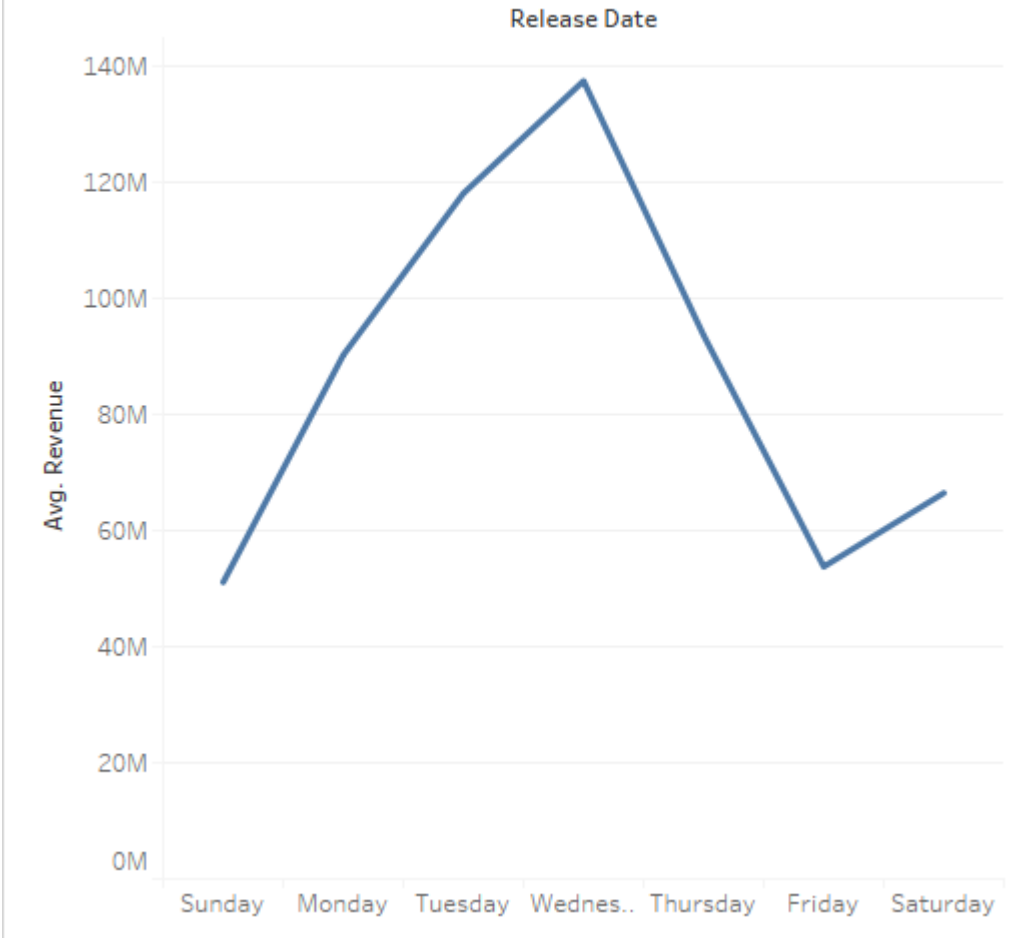
Avg. Profit vs Month



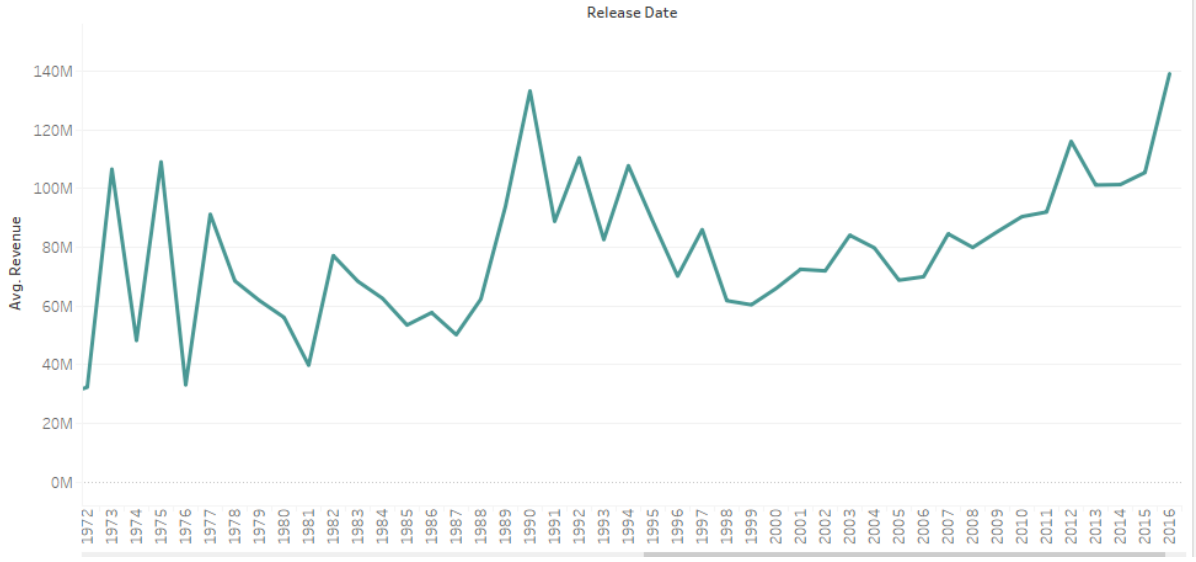
Revenue vs Runtime

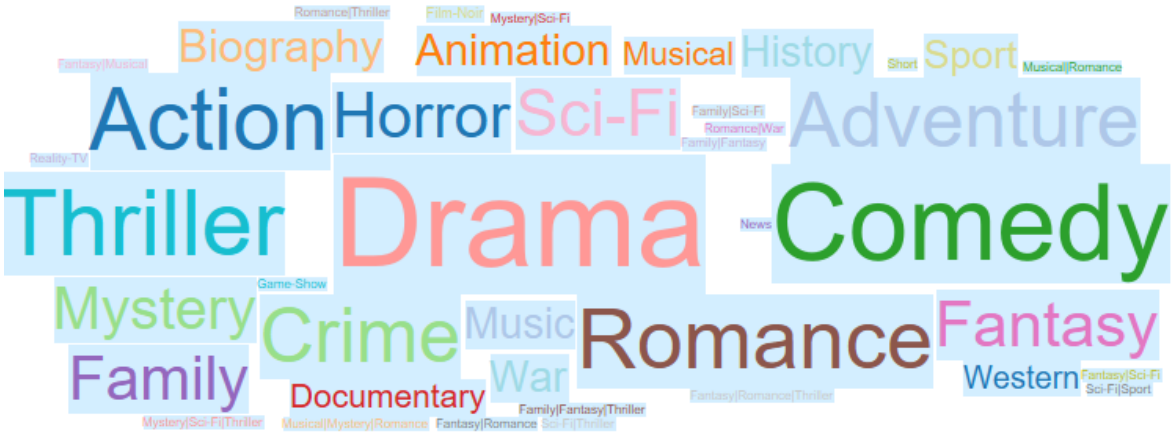


Weekday vs Revenue

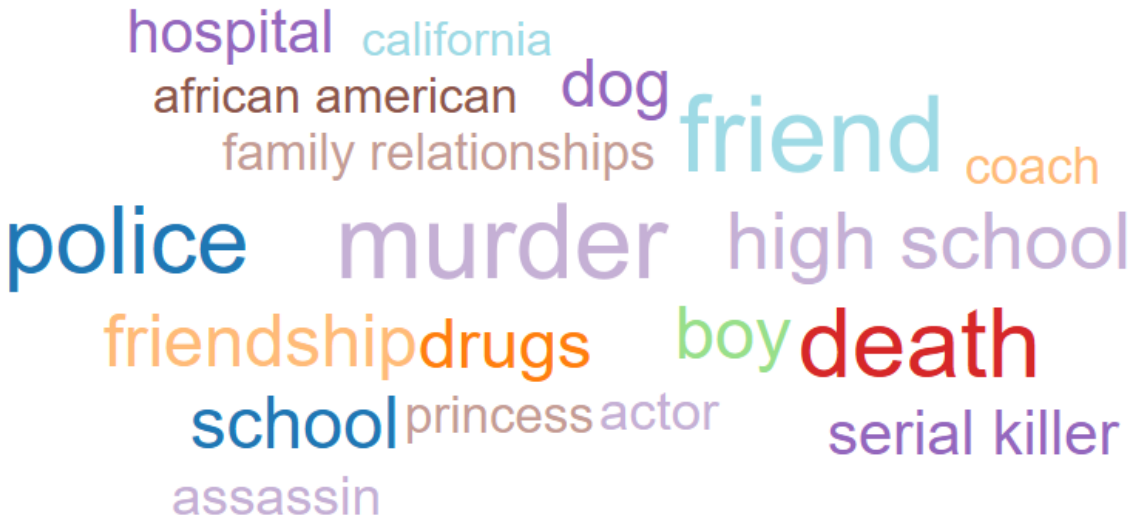


Revenue vs Year

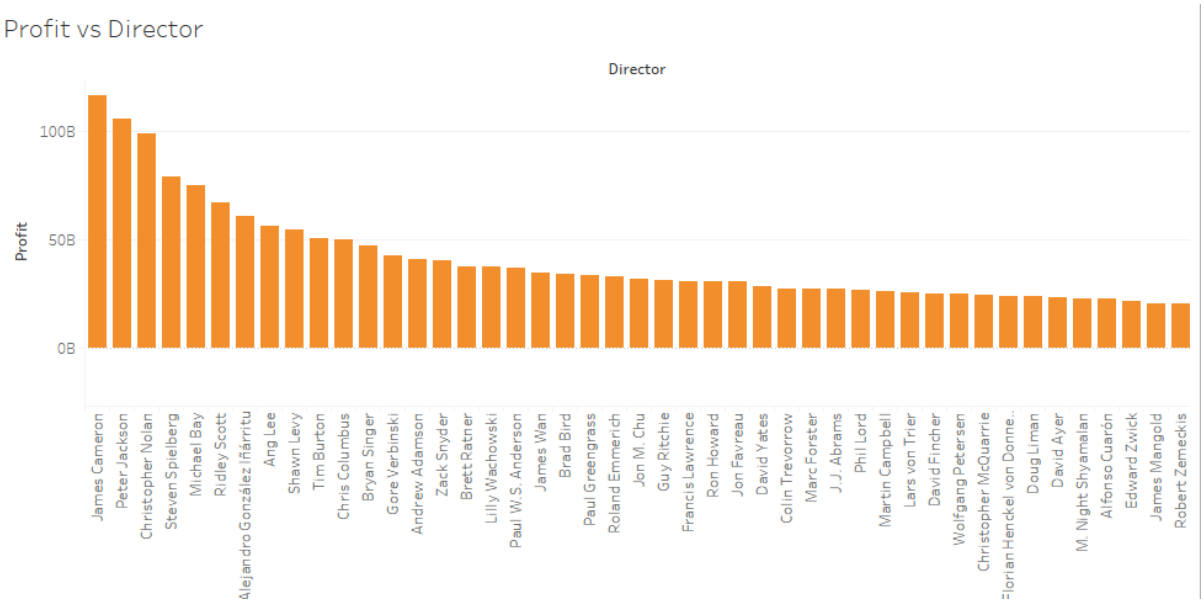




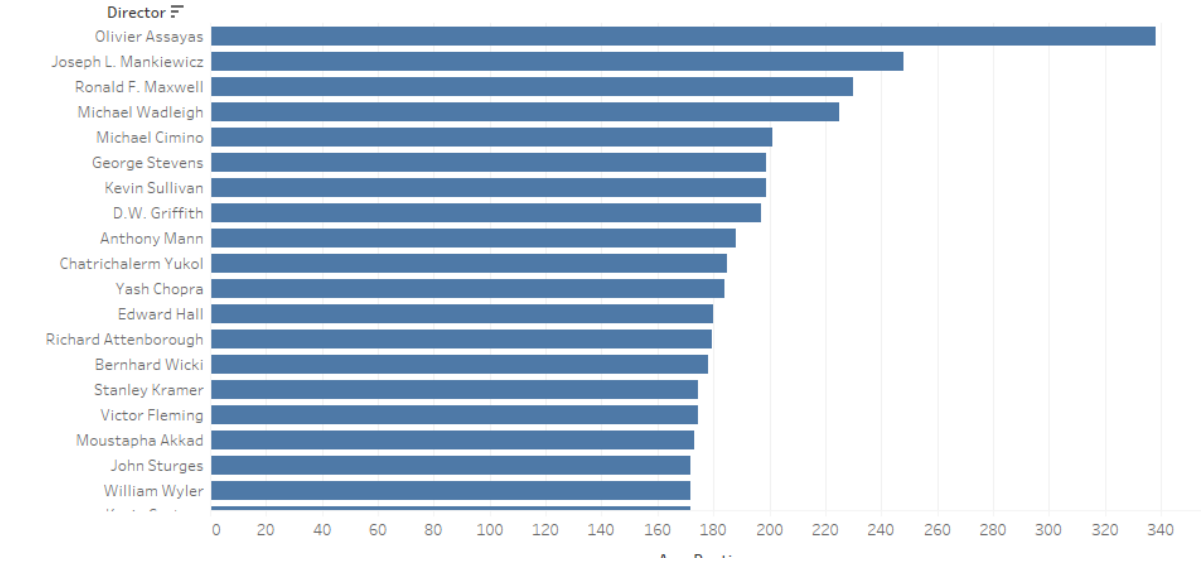
Highly Profitable Words



Profit vs Director



Director vs Runtime



title	'Til There Was You (1997)	1-900 (1994)	101 Dalmatians (1996)	12 Angry Men (1957)	187 (1997)	2 Days in the Valley (1996)	20,000 Leagues Under the Sea (1954)	2001: A Space Odyssey (1968)	3 Ninjas: High Noon At Mega Mountain (1998)	39 Steps, The (1935)	...	Yankee Zulu (1994)	Year of the Horse (1997)	You So Crazy (1994)	Young Frankenstein (1974)	Young Guns (1988)	Young Guns II (1990)	Pois Han The
user_id																		
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	
1	NaN	NaN	2.0	5.0	NaN	NaN	3.0	4.0	NaN	NaN	...	NaN	NaN	NaN	5.0	3.0	NaN	
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.0	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	
3	NaN	NaN	NaN	NaN	2.0	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	

title	Correlation
Star Wars (1977)	1.000000
Empire Strikes Back, The (1980)	0.748353
Return of the Jedi (1983)	0.672556
Raiders of the Lost Ark (1981)	0.536117
Austin Powers: International Man of Mystery (1997)	0.377433