



1000 MOVIES Data Analysis EDA Project for INSAID

AUGUST 2022 COHORT
TERM 3 - 6th November 2022
BY - SOUMYA KANTI KUNDU
EMAIL- SOUAMYAKUNDU55@GMAIL.COM

HELLO!

I AM SOUMYA KANTI KUNDU

I am a fresher from BCA, and
a student at INSAID.

I've also been teaching
Martial Arts for Body & Brain
for 6 & 1 years respectively.

You can find me on socials at
[@Soumyak4](https://twitter.com/Soumyak4)



“

- Everybody has a favorite movie line.
- Disney's Alan Horn likes, "I'll have what she's having," from When Harry Met Sally.
- Fox's Stacey Snider picks "You complete me," from Jerry Maguire.
- Netflix's Ted Sarandos choose "I'm going to make him an offer he can't refuse," from The Godfather.

PRESENTATION ROADMAP

INTRODUCTION

UNDERSTANDING
DATA

SUMMARIZED
CONCLUSION

PROBLEM
SCENARIO

EXPLORATORY
DATA ANALYSIS

ACTIONABLE
INSIGHTS

1

3

5

2

4

6

INTRODUCTION



- ❖ For many of us, Entertainment is a major part of our lives where we spend much of our free time to relax and chill.
- ❖ Watching movies in time-off is can be good for your mental health and general wellbeing.
- ❖ When we enjoy a good movie, we get lost in a movie and get a break from the everyday stressful life.



- ❖ IMDB is an online database of information related to films, television programs, home videos, video games, and streaming content online – including cast, production crew and personal biographies, plot summaries, trivia, ratings, and fan and critical reviews.
- ❖ Metascore is the rating given by another movie rating company called Metacritic. However, unlike IMDB, they get ratings from registered well known rating agencies and calculates a weighted average of those ratings.



PROBLEM SCENARIO

Is there any secret formula for movie making in order
to generate more Revenue ?



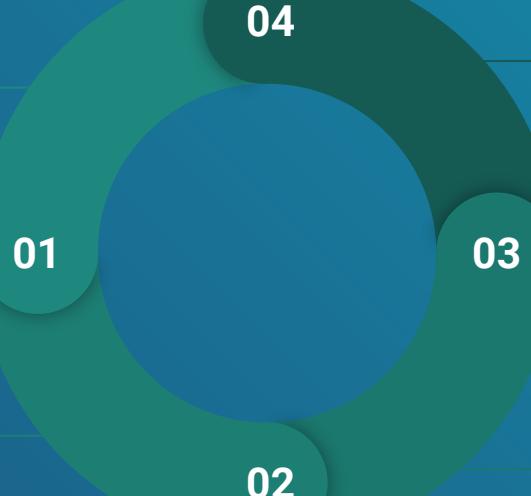
- ❖ In 2017, a production company ABC decides to produce movies that will earn the best in terms of revenue & popularity.
- ❖ This company approaches SOUMYA, and asks him to come up with characteristics of movies that will help them achieve their purpose.
- ❖ He is given the IMDB database of 1000 movies for the last 10 years (2006-2016).
- ❖ His task is come up with the characteristics of movies that are well appreciated by critics and audience & earn the highest revenue.





What changes a director/producer must make in order to make movie a success in terms of revenue and popularity.

Making movies according to audience preferences in terms of genre,cast, runtime etc.



How a positive rating and metascore impacts overall movie business.

What movies connect to global audience more.

UNDERSTANDING DATA



BEFORE DATA CLEANING

Total Datapoints	TotalFeatures
1000	12

<u>Id</u>	<u>Feature</u>	<u>Description</u>
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

AFTER DATA CLEANING

Total Datapoints (Rows)	Total Features (Columns)
999	11

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

A QUICK LOOK AT A SAMPLE OF THE DATASET

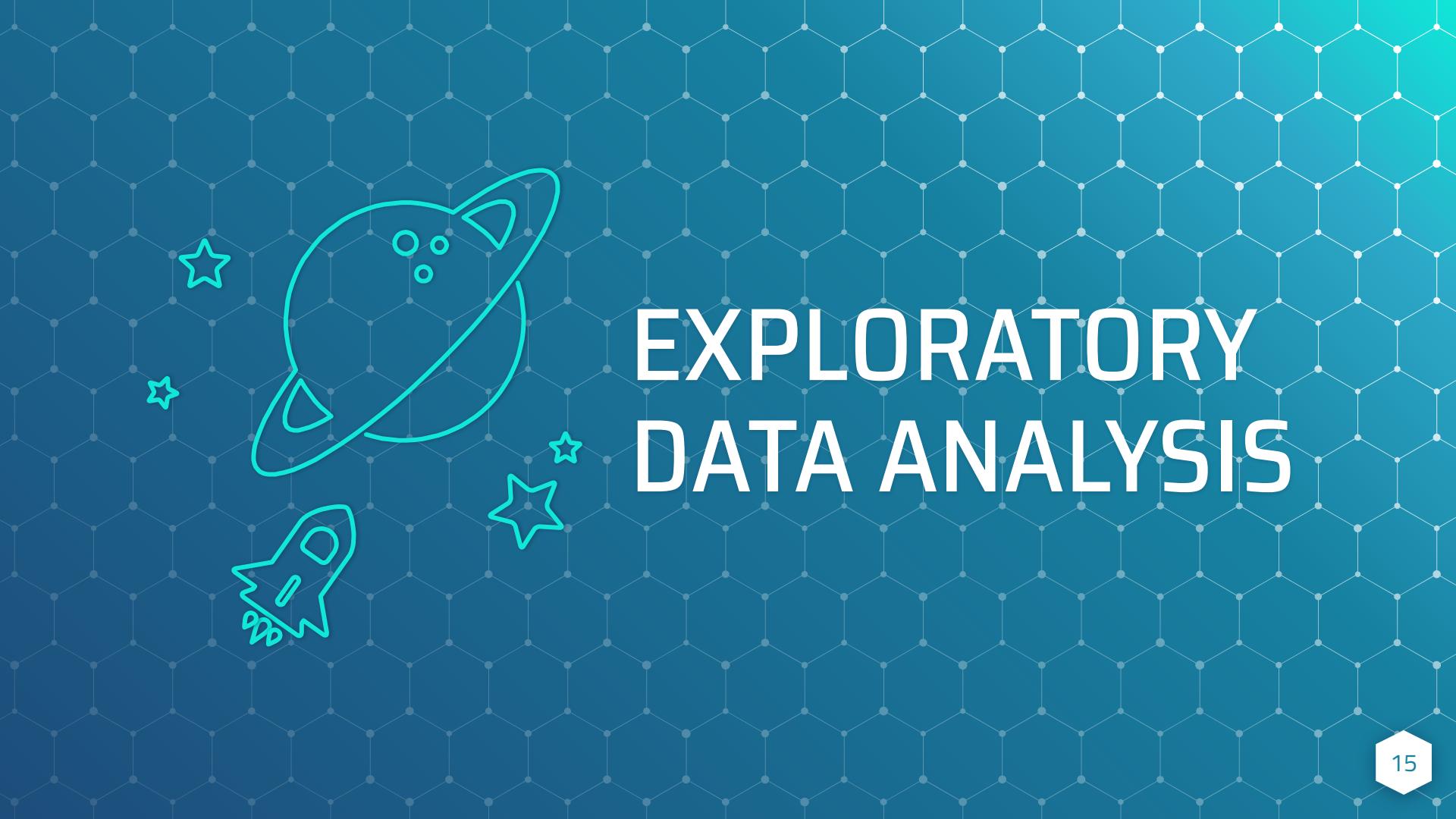
Rank	Title	Genre	Director	Actors	Year	Runtime_Mins	Rating	Votes	Revenue_Mils	Metascore	
141	142	Diary of a Wimpy Kid	Comedy,Family	Thor Freudenthal	Zachary Gordon, Robert Capron, Rachael Harris,...	2010	94	6.20	34184	64.00	56
984	985	Max	Adventure,Family	Boaz Yakin	Thomas Haden Church, Josh Wiggins, Luke Kleint...	2015	111	6.80	21405	42.65	47
442	443	Coraline	Animation,Family,Fantasy	Henry Selick	Dakota Fanning, Teri Hatcher, John Hodgman, Je...	2009	100	7.70	156620	75.28	80
366	367	The Twilight Saga: Breaking Dawn - Part 2	Adventure,Drama,Fantasy	Bill Condon	Kristen Stewart, Robert Pattinson, Taylor Laut...	2012	115	5.50	194329	292.30	52
252	253	Straight Outta Compton	Biography,Drama,History	F. Gary Gray	O'Shea Jackson Jr., Corey Hawkins, Jason Mitch...	2015	147	7.90	139831	161.03	72

A QUICK LOOK AT THE STATISTICS OF THE DATASET

	Rank	Year	Runtime_Mins	Rating	Votes	Revenue_Mils	Metascore
count	999.00	999.00	999.00	999.00	999.00	999.00	999.00
mean	500.77	2012.78	113.19	6.72	169974.92	78.56	58.99
std	288.84	3.21	18.81	0.94	188783.56	97.13	16.64
min	1.00	2006.00	66.00	1.90	61.00	0.01	11.00
25%	251.50	2010.00	100.00	6.20	36317.00	17.50	47.50
50%	501.00	2014.00	111.00	6.80	110825.00	47.98	59.00
75%	750.50	2016.00	123.00	7.40	240047.50	99.46	71.00
max	1000.00	2016.00	191.00	9.00	1791916.00	936.63	100.00

#	Column	Non-Null Count	Dtype
0	Rank	999 non-null	int64
1	Title	999 non-null	object
2	Genre	999 non-null	object
3	Director	999 non-null	object
4	Actors	999 non-null	object
5	Year	999 non-null	int64
6	Runtime_Mins	999 non-null	int64
7	Rating	999 non-null	float64
8	Votes	999 non-null	int64
9	Revenue_Mils	999 non-null	float64
10	Metascore	999 non-null	int64

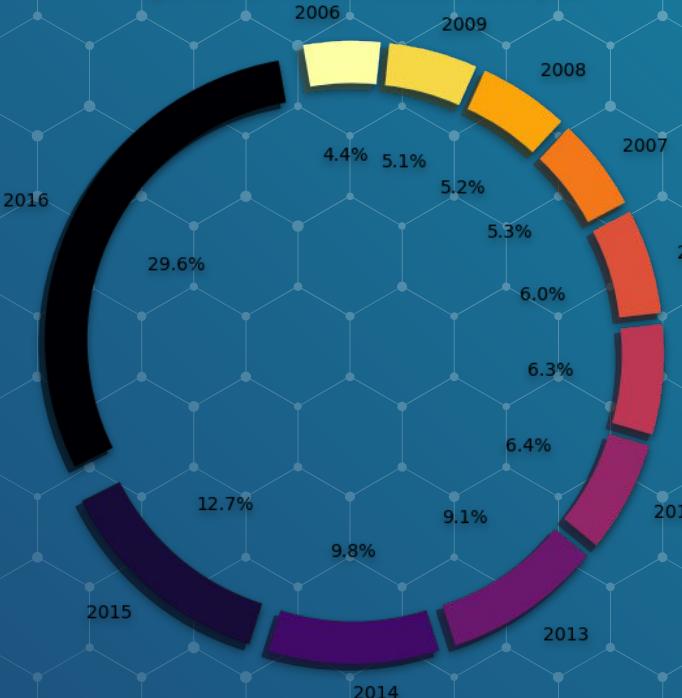
	Title	Genre	Director	Actors
count	999	999	999	999
unique	998	207	643	995
top	The Host	Action, Adventure, Sci-Fi	Ridley Scott	Jennifer Lawrence, Josh Hutcherson, Liam Hemsw...
freq	2	50	8	2



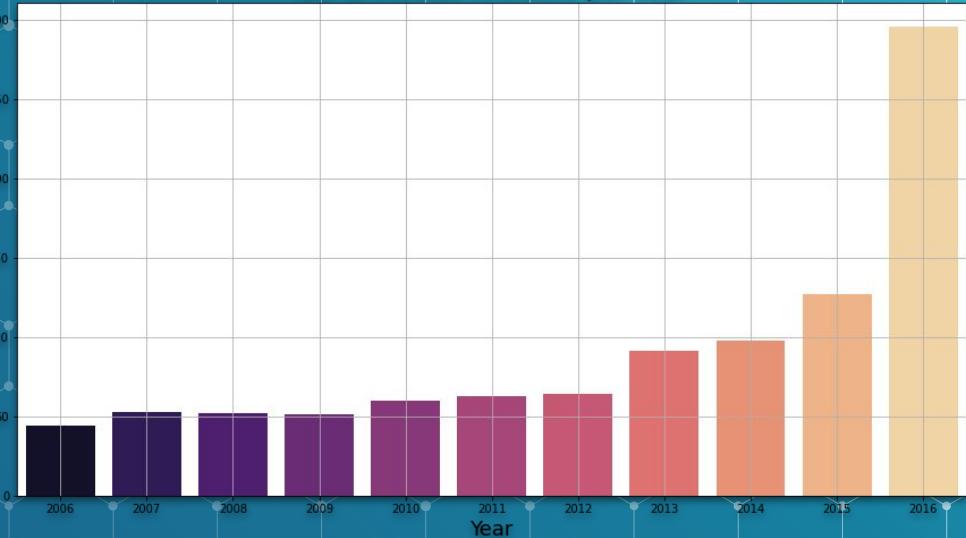
EXPLORATORY DATA ANALYSIS

Q1 - WHAT ARE THE NUMBER OF MOVIES RELEASED YEAR WISE?

Proportion of Movies released each year

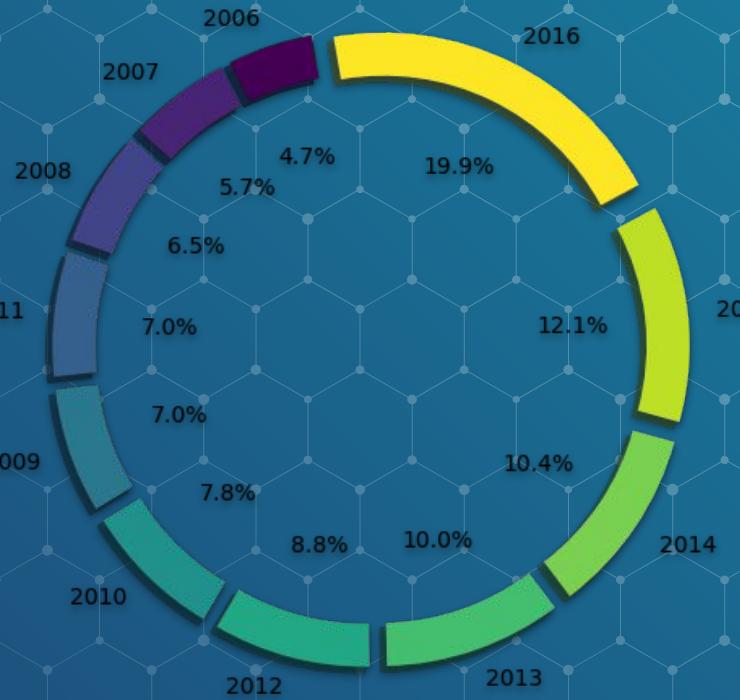


No. of Movies released per Year

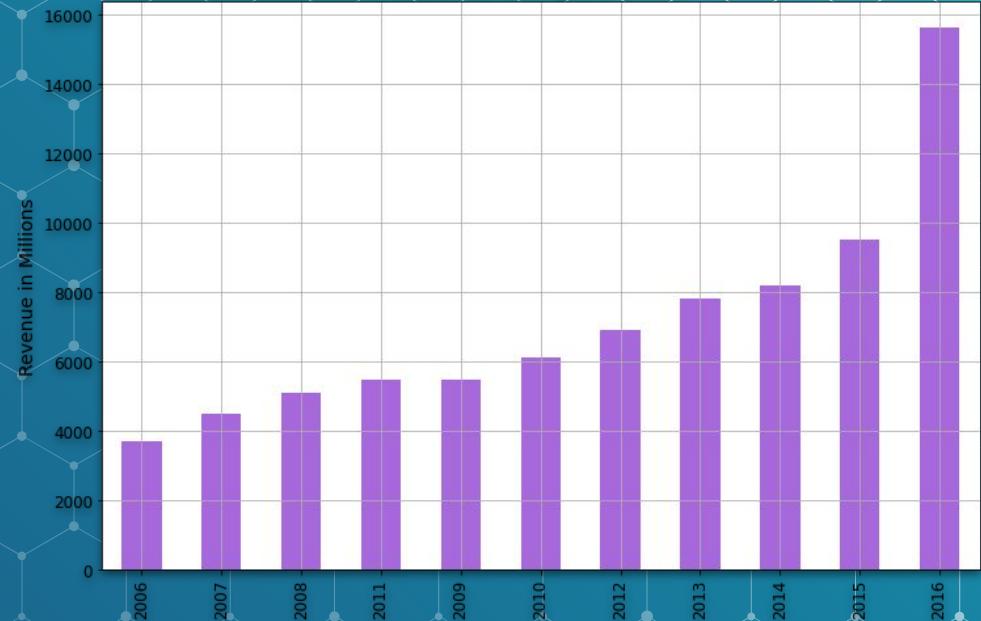


Q2 - WHICH YEAR GENERATED THE MOST REVENUE?

Proportion of Movies revenue each year

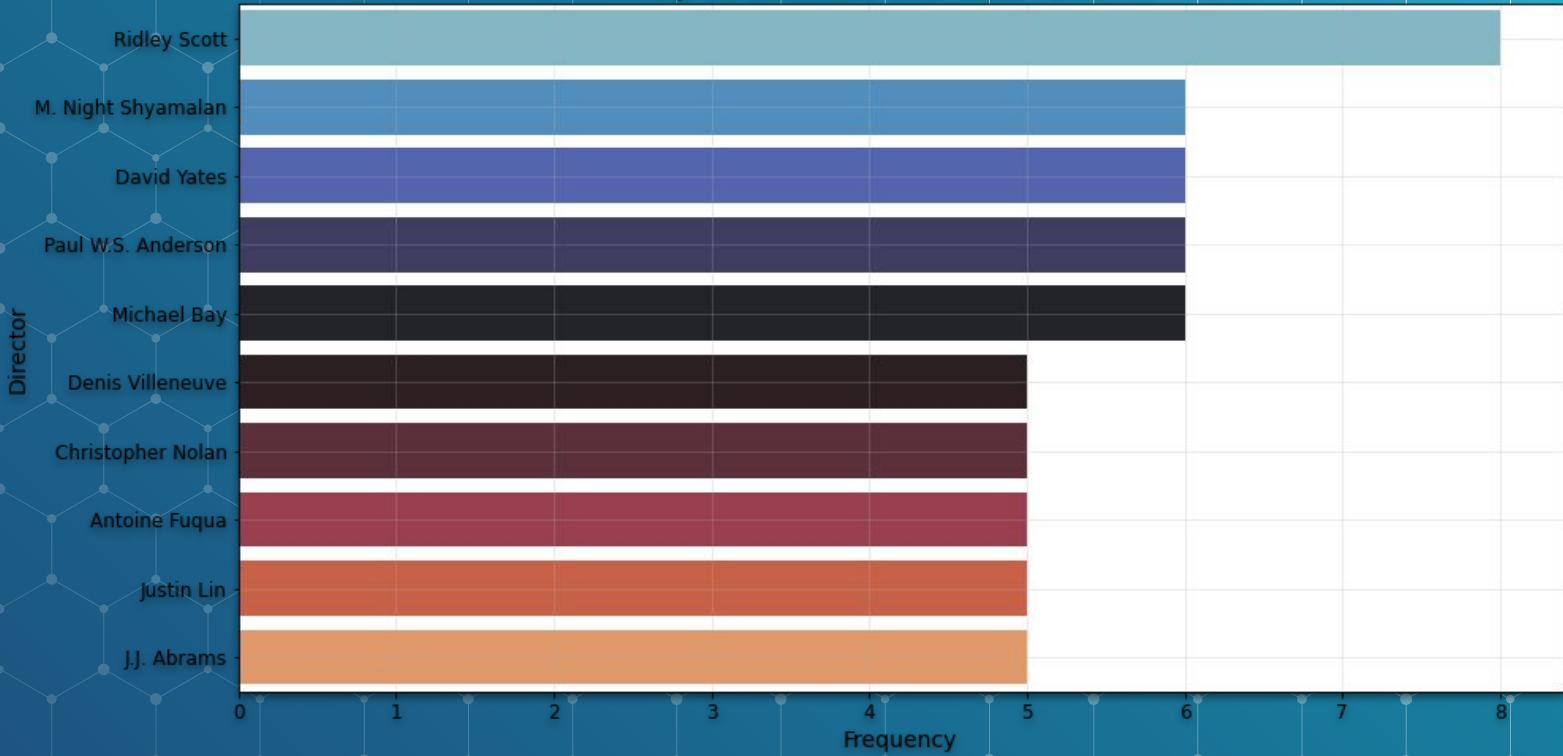


Revenue along the years



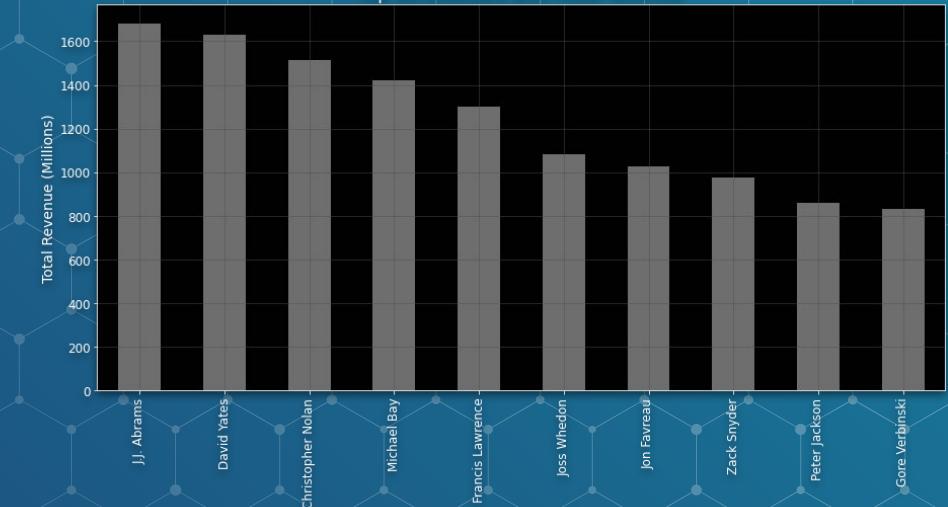
Q3 - WHO ARE THE DIRECTORS WHO MADE THE MOST MOVIES?

Top 10 Directors with most releases

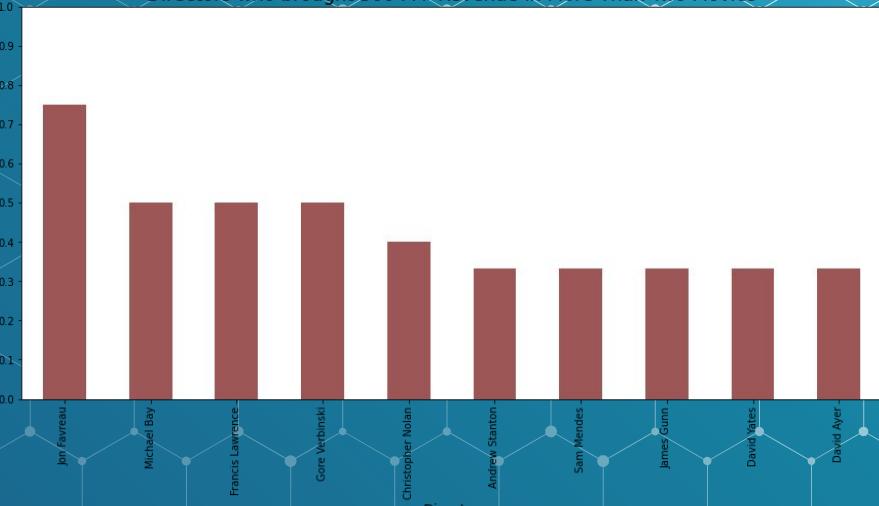


Q4 - WHO ARE THE DIRECTORS OF MOST PROFITABLE MOVIES?

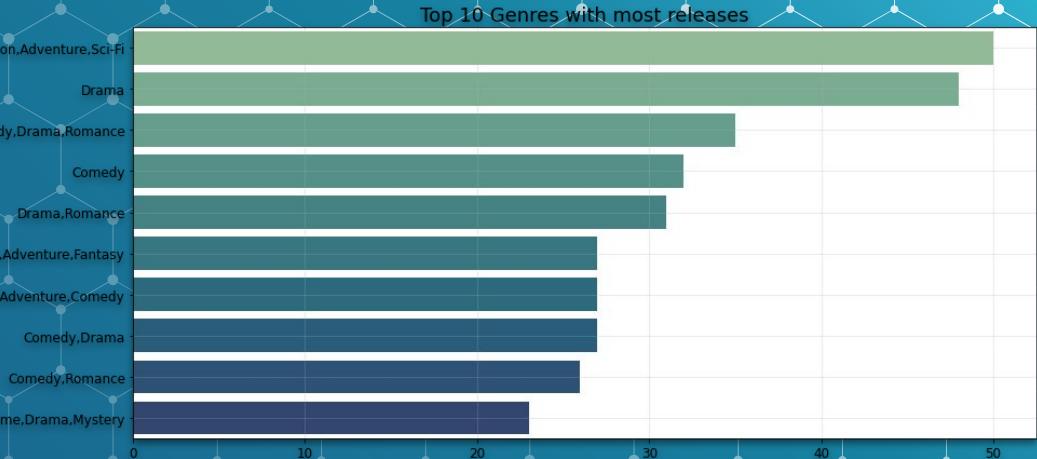
Top 10 Directors with most revenue



Directors who brought 300 M+ Revenue in More Than Two Movies

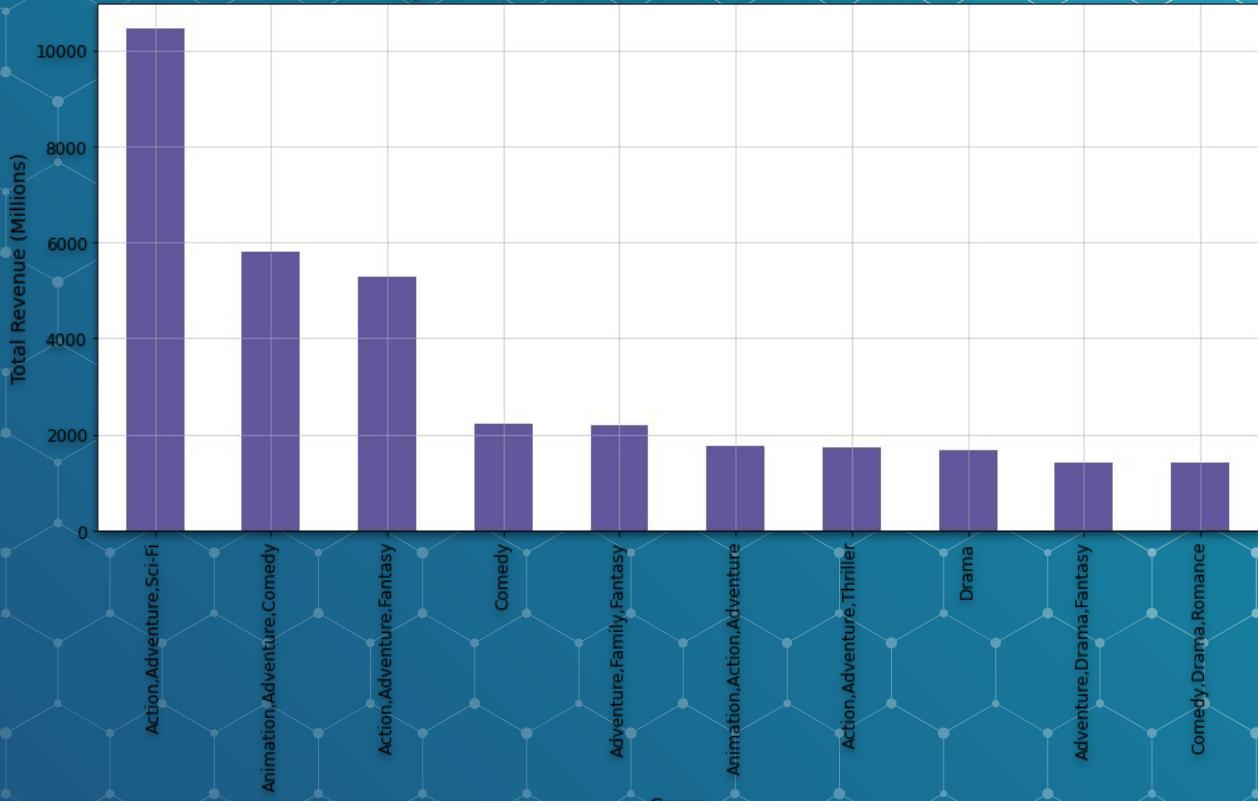


Q5: WHAT ARE THE TOP GENRE COMBINATIONS THAT HAS MOST MOVIE RELEASES?

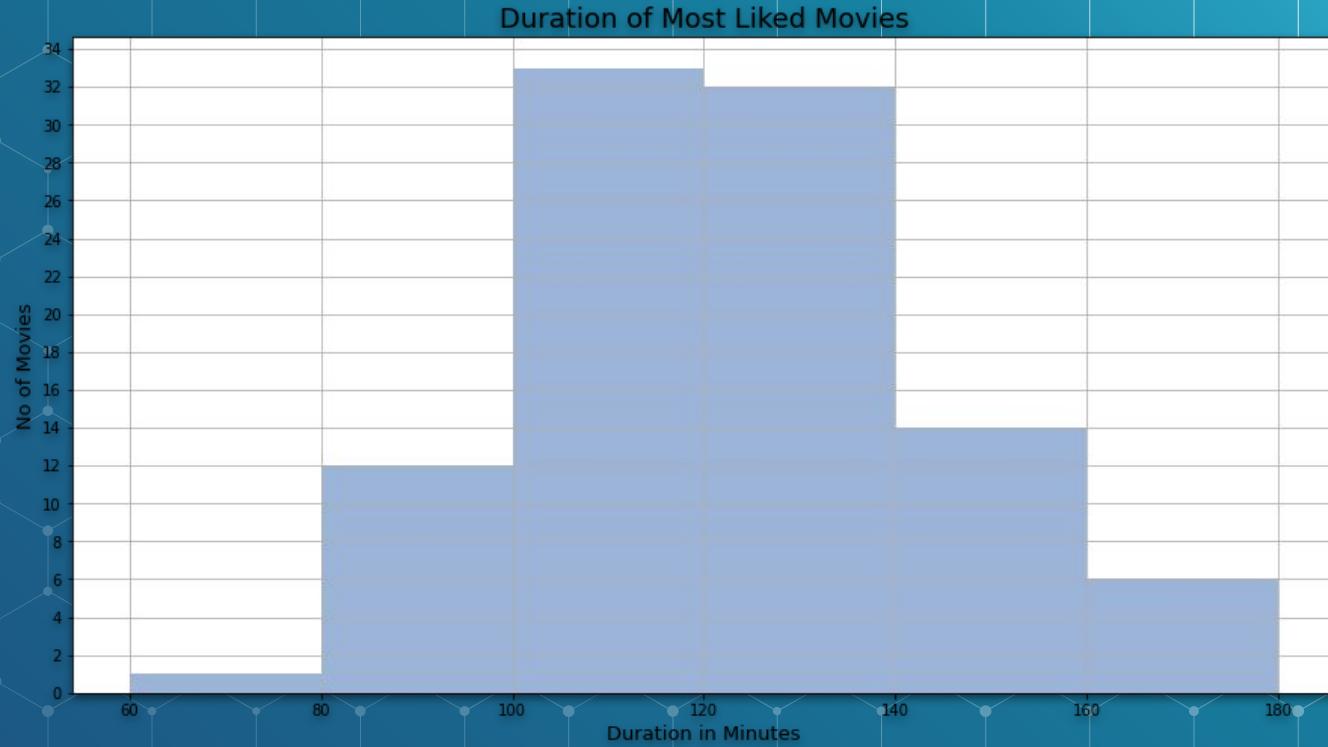


Q6: WHAT ARE THE TOP GENRE COMBINATIONS THAT GAINED THE MOST REVENUE.

Top 10 Genre combinations with most revenue

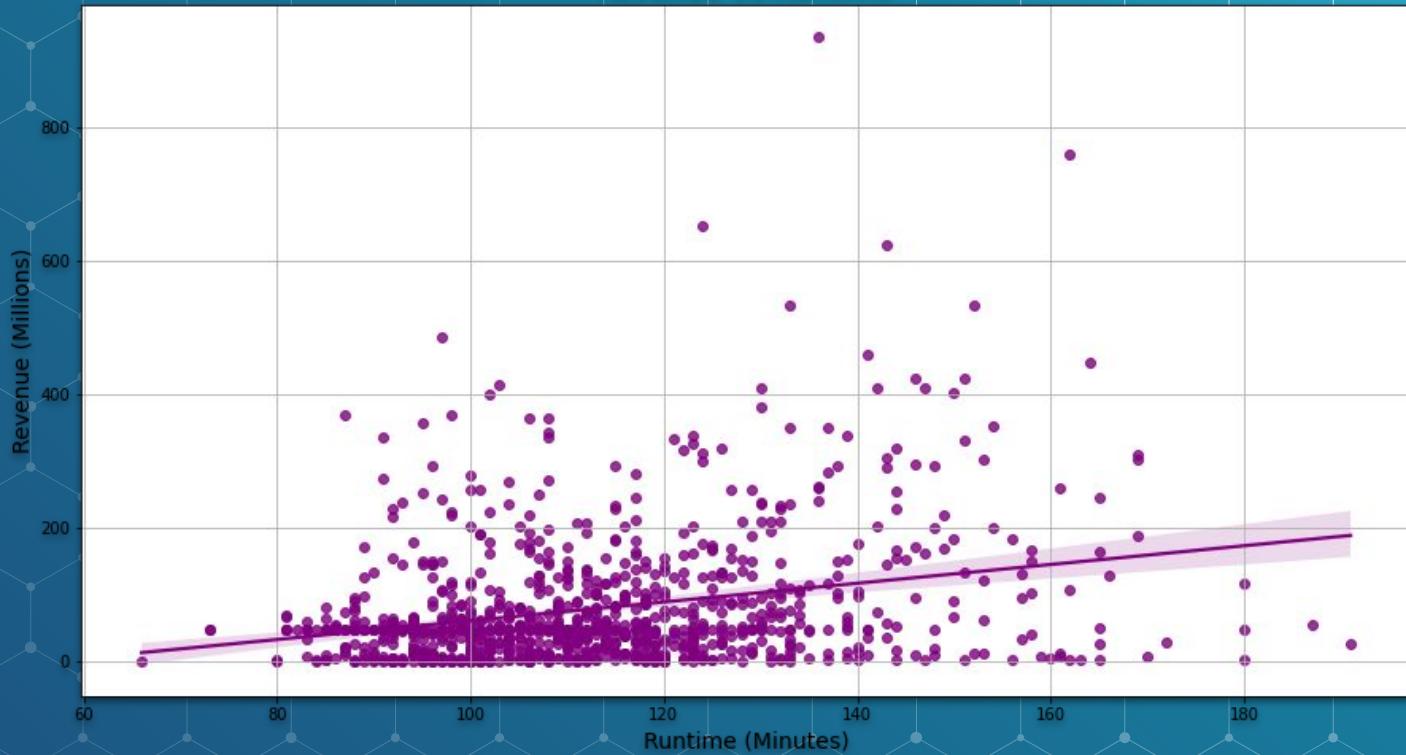


Q7: What is the Average duration of movies most favored by audience and CRITICS?

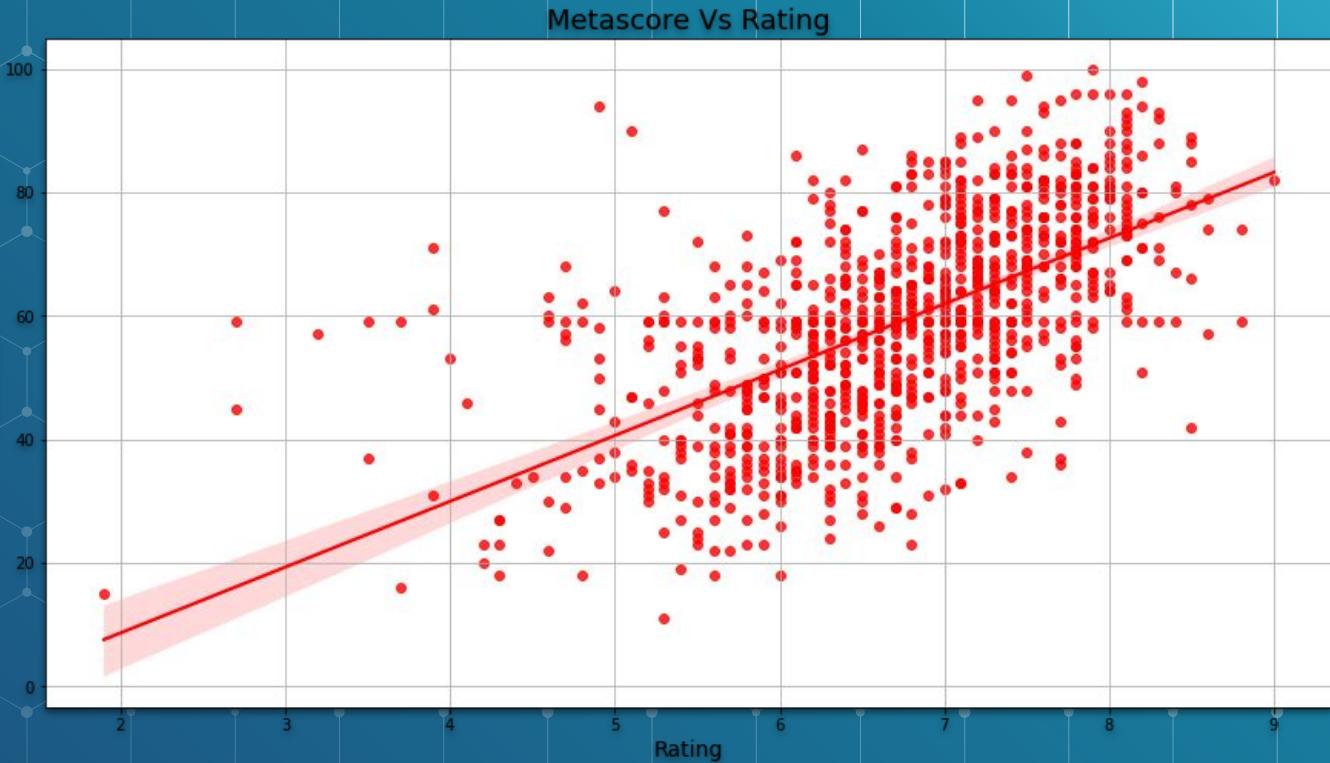


Q8: How is the runtime of a movie related to the movie's REVENUE?

Revenue Vs Runtime

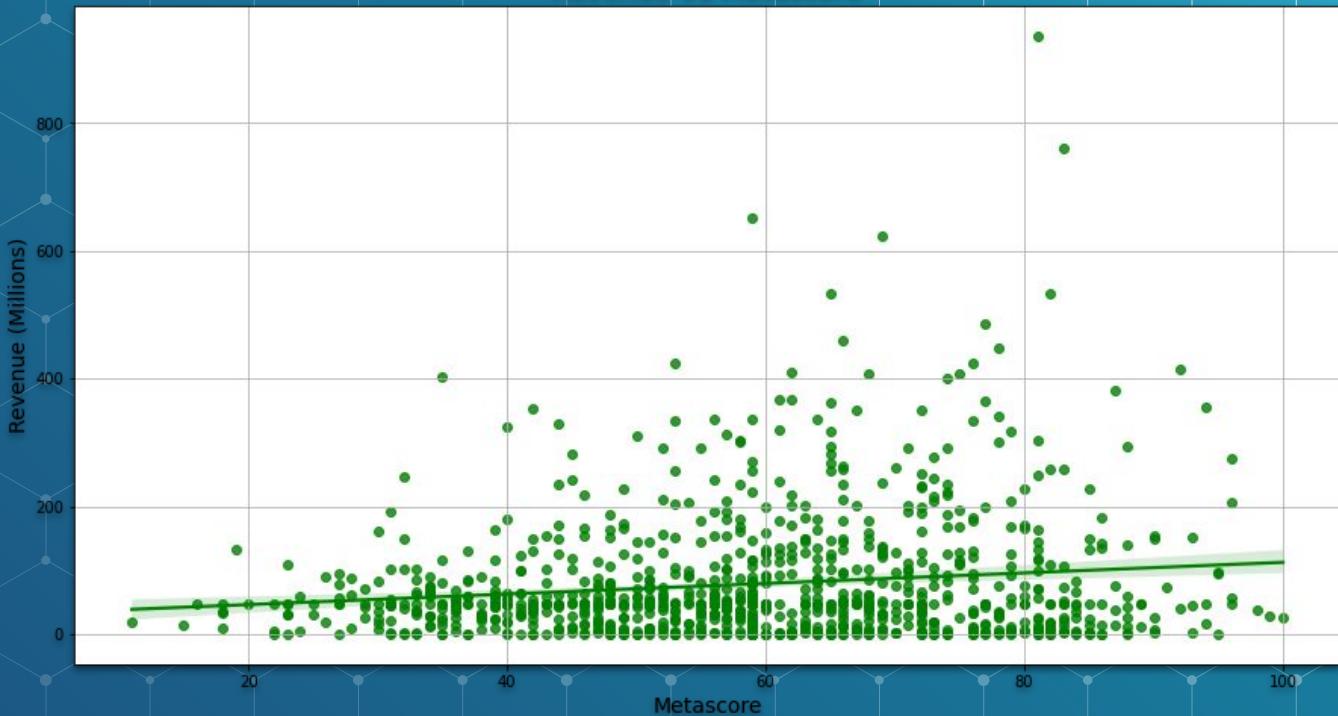


Q9: How is the rating of a movie related to the movie's METASCORE?



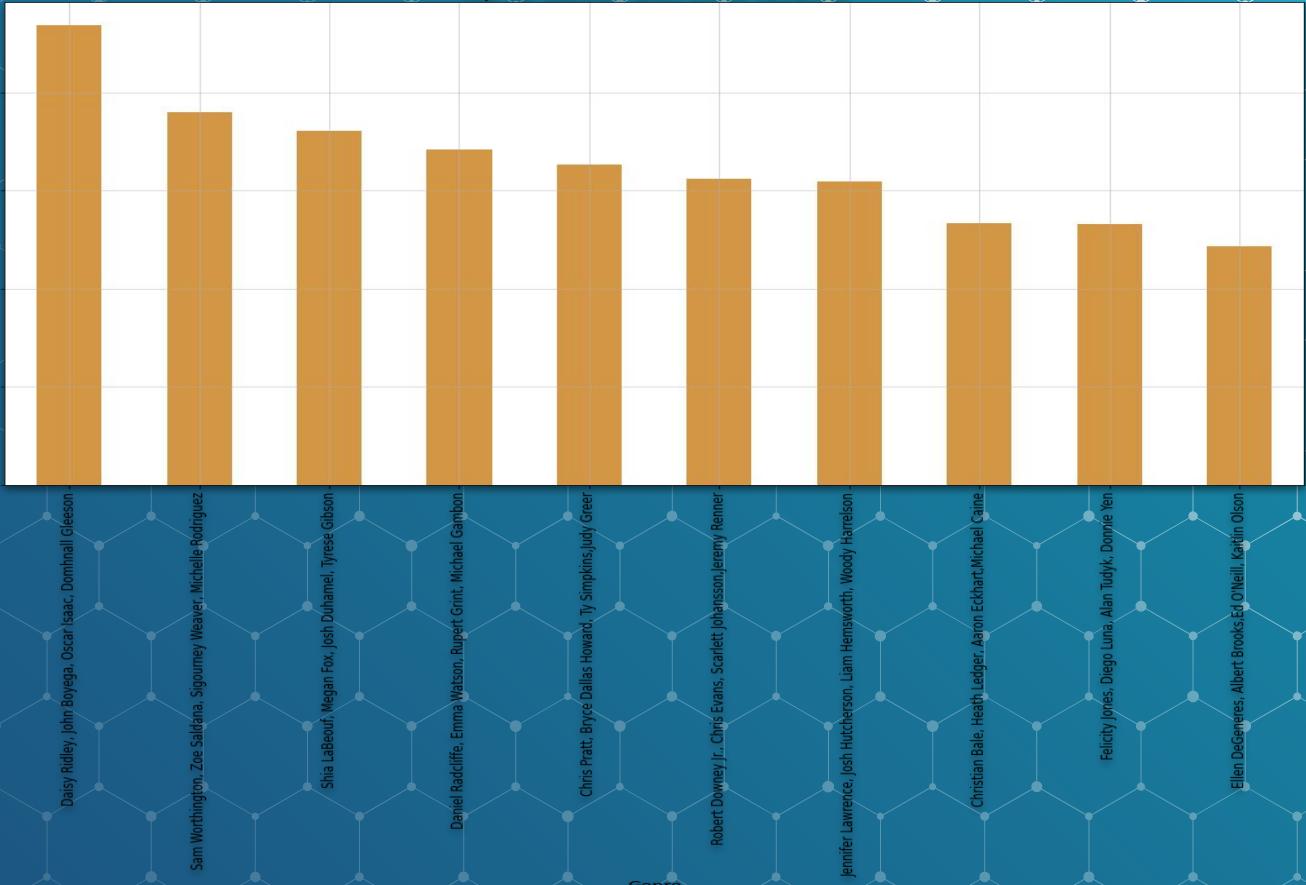
Q10: Does Metascore affect the REVENUE?

Revenue Vs Metascore

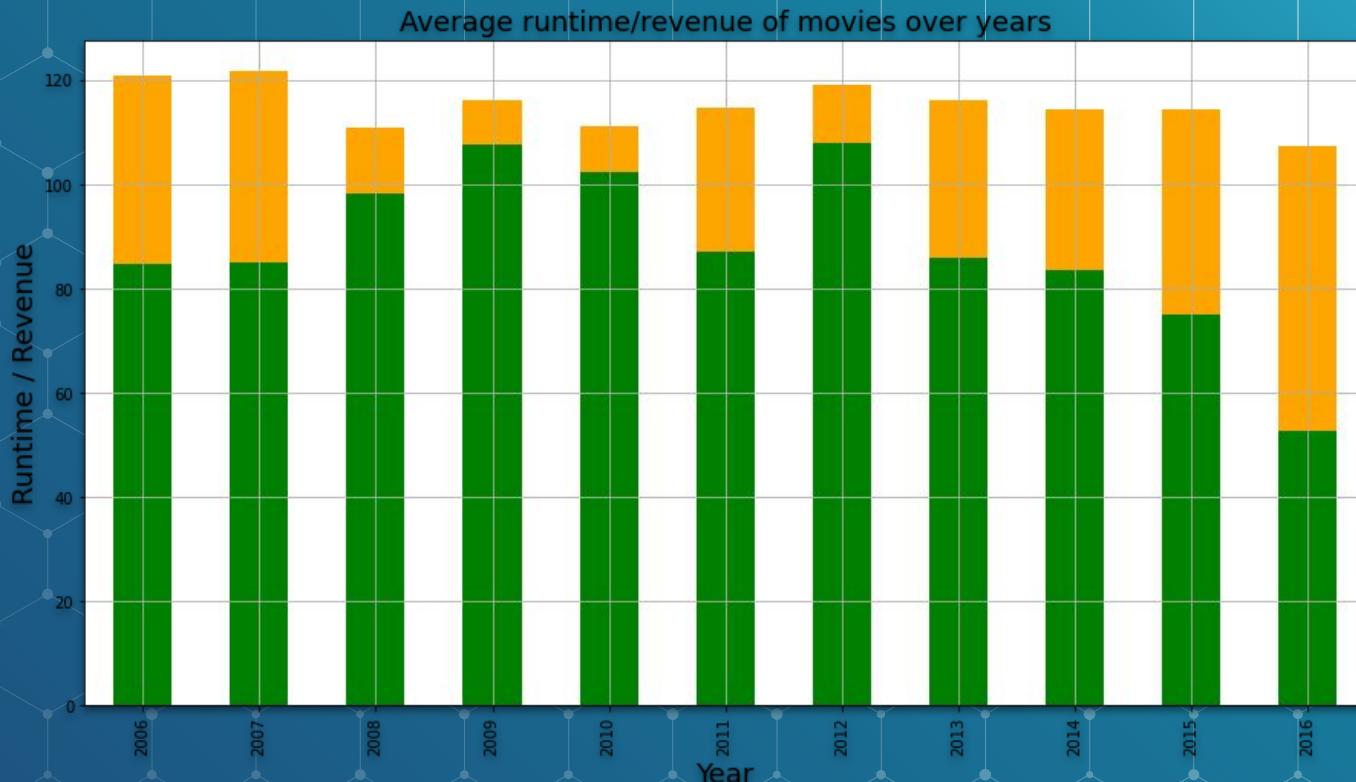


Q11: WHICH ACTORS COMBINATION MADE THE MOST REVENUE?

Top 10 Genres with most revenue

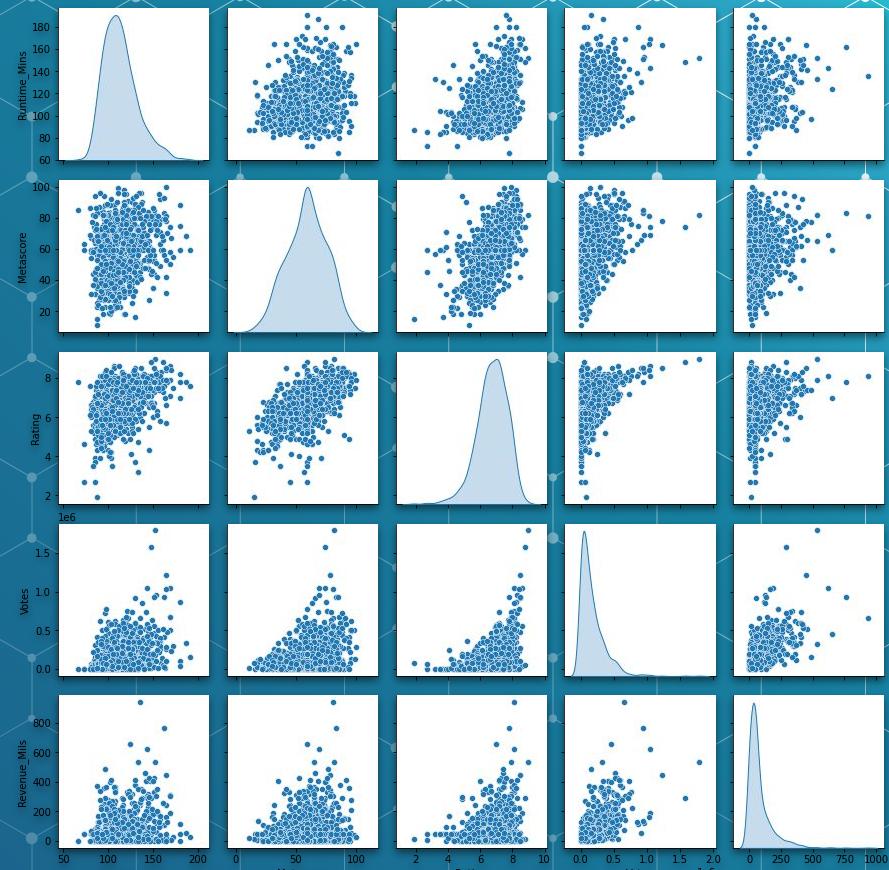
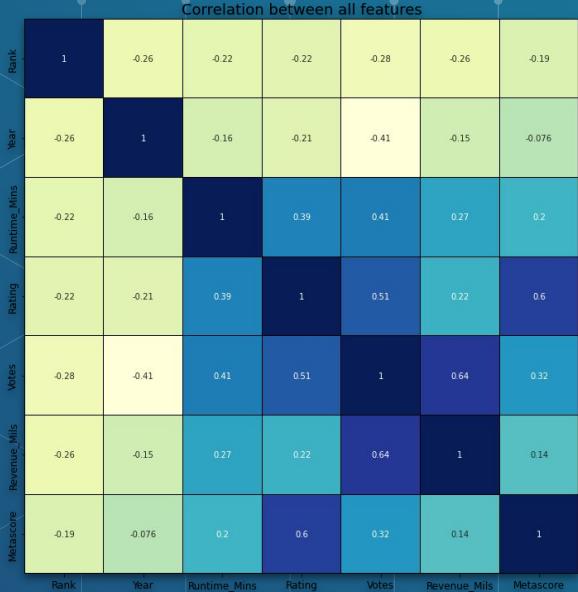


Q12: AVERAGE RUNTIME & REVENUE OF MOVIES OVER YEARS.

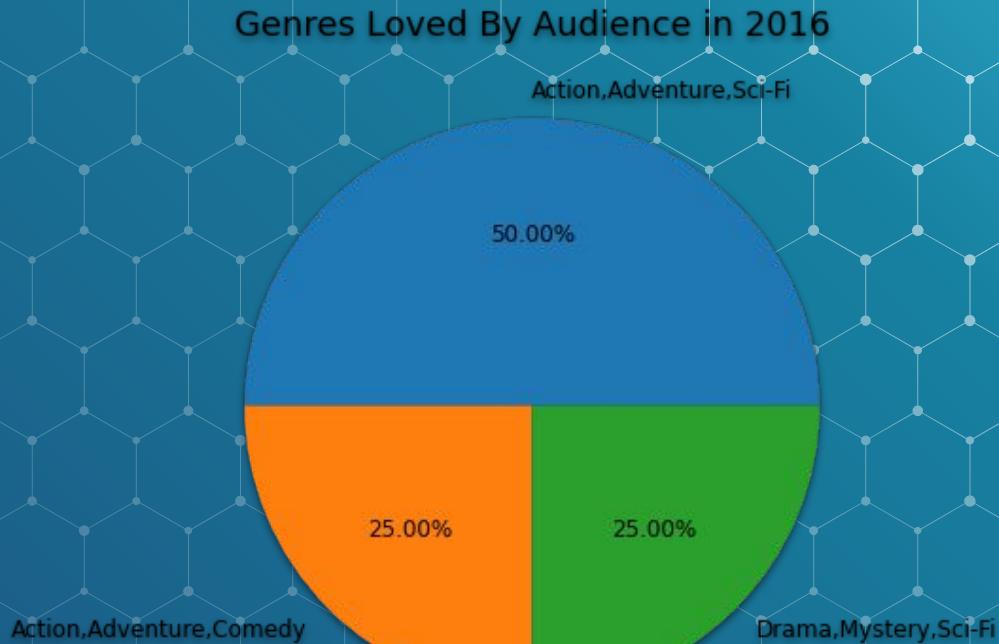


Runtime_Mins
Revenue_Mils

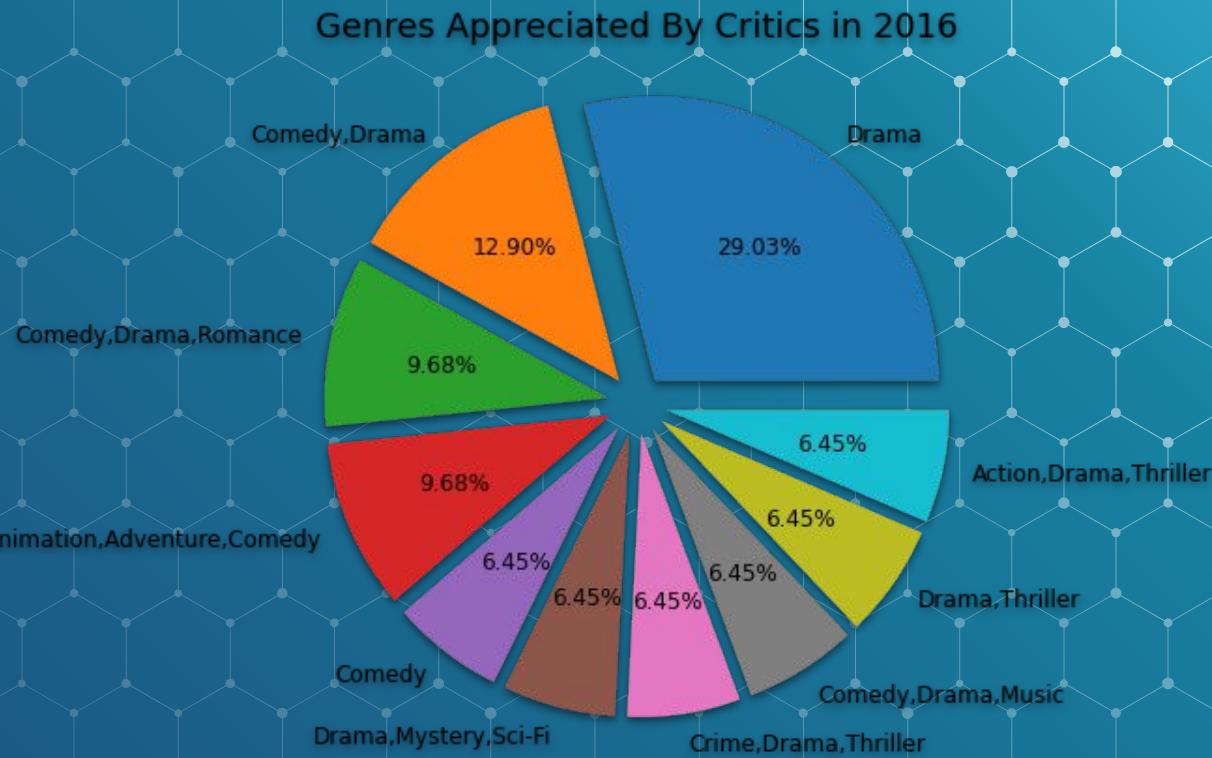
Q13: CORRELATION BETWEEN FEATURES.



Q14: WHICH GENRE OF MOVIES THE AUDIENCE LOVED MOST IN 2016?

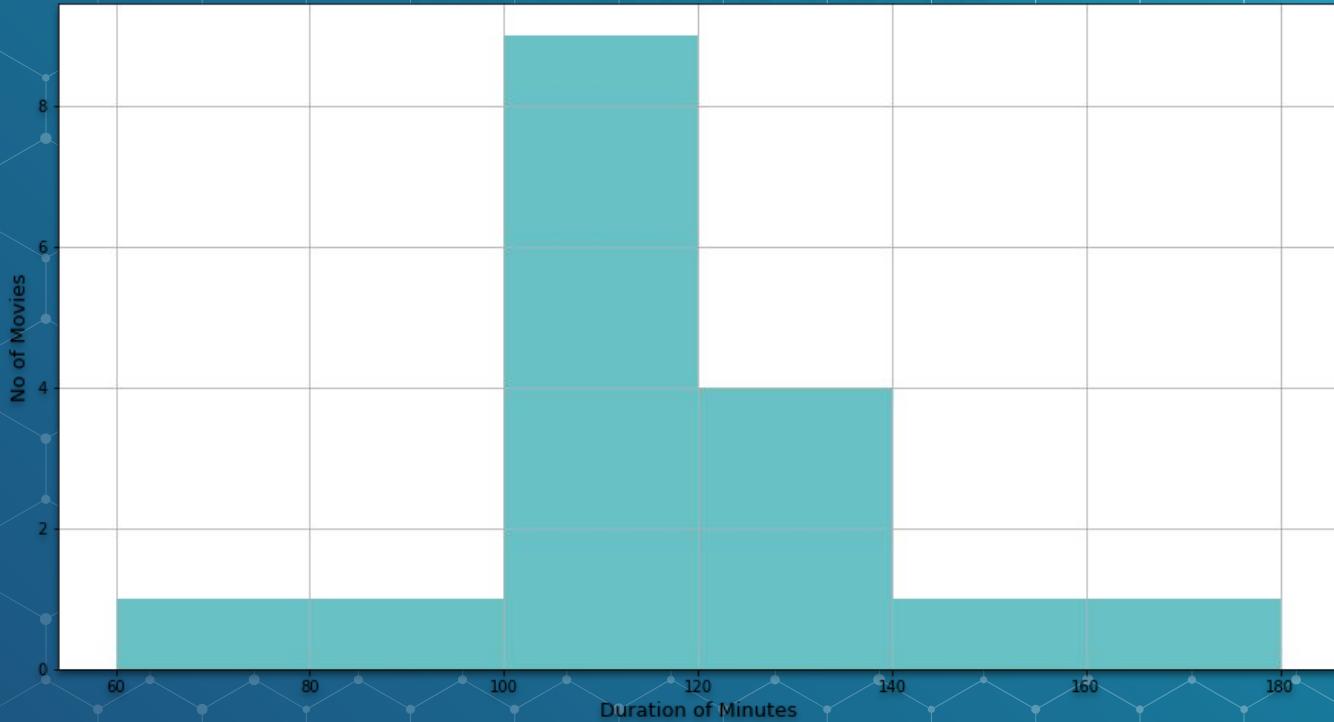


Q15: WHICH GENRE OF MOVIES WERE APPRECIATED BY THE CRITICS IN 2016?



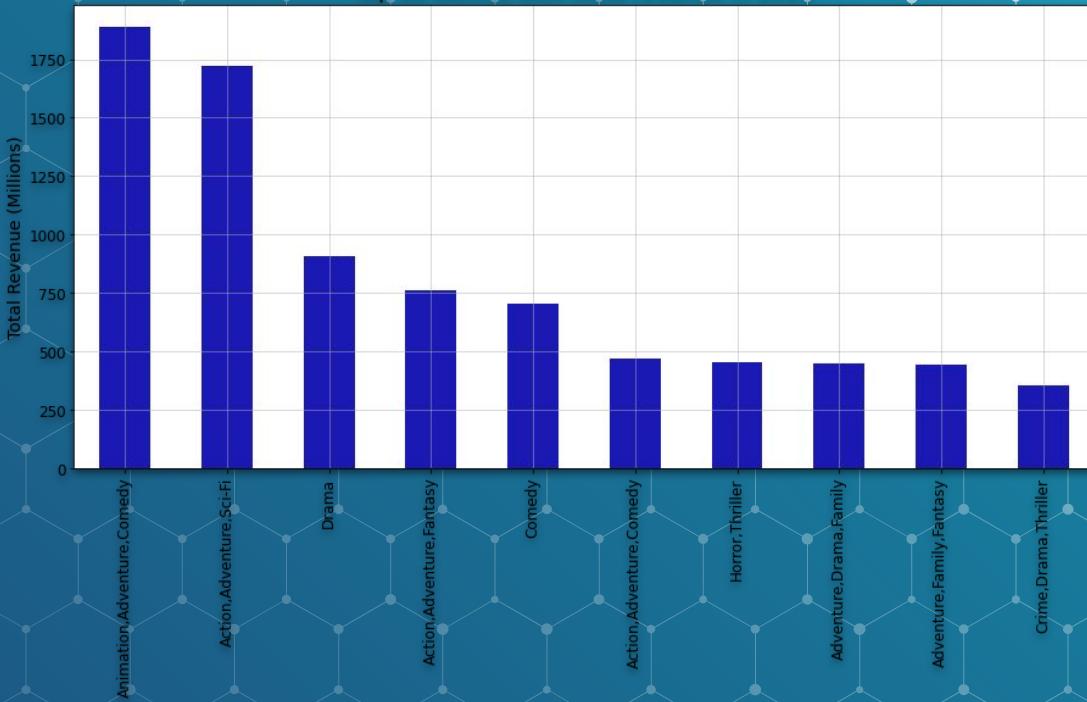
Q16: WHICH DURATION OF MOVIES WERE APPRECIATED BY THE CRITICS AND AUDIENCE IN 2016?

Duration of Most Favoured Movies in 2016



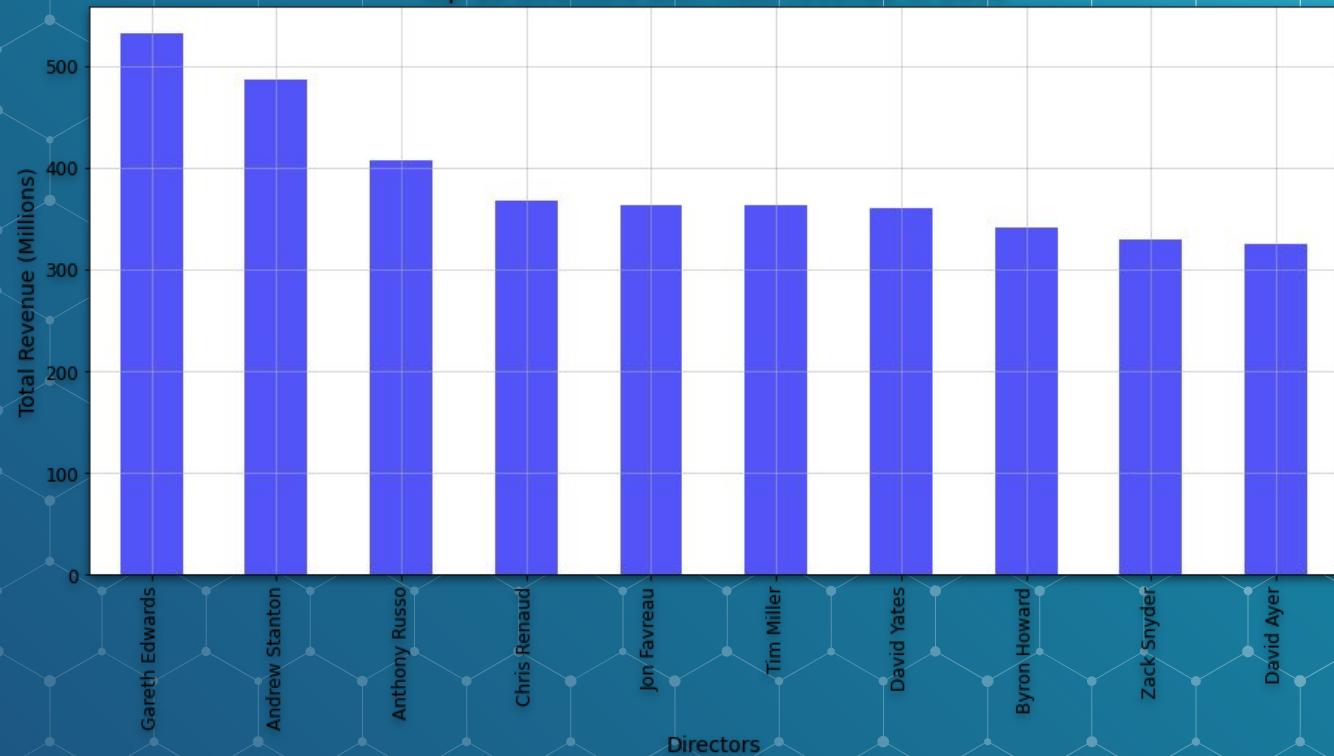
Q17: WHAT ARE THE TOP GENRE COMBINATIONS THAT GAINED THE MOST REVENUE IN 2016.

Top 10 Genres with most revenue in 2016

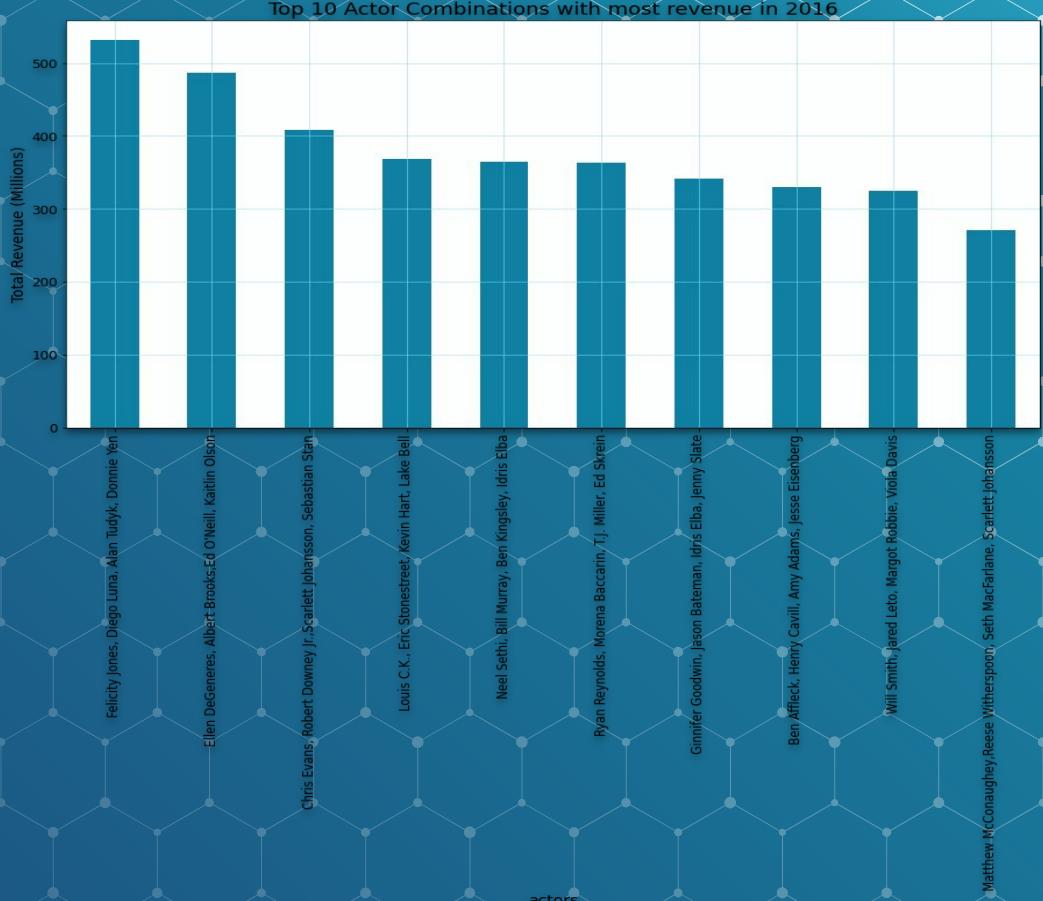


Q18: WHO ARE THE DIRECTORS OF MOST PROFITABLE MOVIES IN 2016?

Top 10 Directors with most revenue in 2016



Q19: WHICH ACTORS COMBINATION MADE THE MOST REVENUE IN 2016?





SUMMARIZED CONCLUSION

2016 produced most no. of movies, hence got the highest total revenue.

Movies that has a duration of approx 100 to 140 minutes are most favoured by audience and critics.

Just because a director made more number of movies doesn't mean he/she has generated more revenue.

Revenue greater than 400 million dollars are the ones with Duration between 120 minutes and 170 minutes.

Most Movies Are From This Following Genre Combination, So They Also Gained The Most Revenue Over The Years "Action, Adventure, Sci-Fi"

Metascore and Rating are directly proportional to each other.



SUMMARIZED CONCLUSION

Revenue greater than 400 million dollars are the ones with Metascore greater than 60

Average Runtime is slightly decreasing over the years & the average Revenue is decreasing in recent years.

Runtime of the Movie is an considerable entity which indirectly affects the Votes, Metascore and Rating which In turn affects the Revenue.

The highest positive correlation 0.64 is between votes and revenue.

ACTIONABLE INSIGHTS



1. Pick the directors who made high revenues in more than 2 movies:
2. Make movies in the following genre combinations: 'Animation, Adventure, Comedy' & 'Action, Adventure, Sci-Fi' as they've have gained significantly more revenue than other genre combinations in the latest year.
3. Duration should be between 120 to 140 minutes.
4. Movies with combination of 2/3 Genres are preferred by viewers compared to sole genre.
But, Drama movies as an exception, are appreciated by critics and also scored good ratings from maximum audience.
5. Votes do play a role in getting better rating and in turn making more revenue, so the filmmakers should do better promotions.

THANK YOU!

You can find me at:

@SoumyaK4

soumyakundu55@gmail.com

Please share any feedbacks over
mail.





CREDITS

Special thanks to all the faculties and classmates at
INSAID.