DATA SCIENCE PROJECT
Phase 1 Project

BY: ROSALINE WANGARI MUNGAI

Movie making business is currently one of the most lucrative businesses to start in this age of technology and online devices. Several factors come into play namely: movie title, gross income, genre, ratings and number of votes by customers .Here is a snippet of some of the factors considered from analyzed data .

	title	studio	domestic gross	year	tconst	original_title	genres	Average rating	numvotes	
0	Toy Story 3	BV	415000000.0	2010	tt0146592	Pál Adrienn	Drama	6.8	451	
1	Toy Story 3	BV	415000000.0	2010	tt0154039	Oda az igazság	History	4.6	64	
2	Toy Story 3	BV	415000000.0	2010	tt0162942	A zöld sárkány gyermekei	Drama	6.9	120	
3	Toy Story 3	BV	415000000.0	2010	tt0230212	The Final Journey	Drama	8.8	8	
4	Toy Story 3	BV	415000000.0	2010	tt0312305	Quantum Quest: A Cassini Space Odyssey	Adventure,Animation,Sci-Fi	5.1	287	

The top 10 most watched movies are shown below based on their average rating and the number of votes. This also shows how much gross income their make and the release year. This forms a big part of decision making in terms of the genres to go with and how much revenue it generates

title	studio	domestic_gross	year	tconst	original_title	genres	averagerating	numvotes
Toy Story 3	BV	415000000.0	2010	tt1375666	Inception	Action,Adventure,Sci-Fi	8.8	1841066
Alice in Wonderland (2010)	BV	334200000.0	2010	tt1375666	Inception	Action,Adventure,Sci-Fi	8.8	1841066
Harry Potter and the Deathly Hallows Part 1	WB	296000000.0	2010	tt1375666	Inception	Action,Adventure,Sci-Fi	8.8	1841066
Inception	WB	292600000.0	2010	tt1375666	Inception	Action,Adventure,Sci-Fi	8.8	1841066
Shrek Forever After	P/DW	238700000.0	2010	tt1375666	Inception	Action,Adventure,Sci-Fi	8.8	1841066
The Twilight Saga: Eclipse	Sum.	300500000.0	2010	tt1375666	Inception	Action,Adventure,Sci-Fi	8.8	1841066
Iron Man 2	Par.	312400000.0	2010	tt1375666	Inception	Action,Adventure,Sci-Fi	8.8	1841066
Tangled	BV	200800000.0	2010	tt1375666	Inception	Action,Adventure,Sci-Fi	8.8	1841066
Despicable Me	Uni.	251500000.0	2010	tt1375666	Inception	Action,Adventure,Sci-Fi	8.8	1841066
How to Train Your Dragon	P/DW	217600000.0	2010	tt1375666	Inception	Action,Adventure,Sci-Fi	8.8	1841066

Action, Adventure, Sci-Fi is the most watched genre according to the number of votes.
This data also shows correlation between the year of release vs rating .The correlation is negative because there is no correlation there .However, there is correlation between the average rating of the movie by the number of votes it gets .This is because watchers get to improve the rating of a movie based on the votes.

MOST WATCHED GENRE

genres	averagerating	numvotes
Action,Adventure,Sci-Fi	8.8	1841066
Action,Adventure,Sci-Fi	8.8	1841066
Action,Adventure,Sci-Fi	8.8	1841066
Action,Adventure,Sci-Fi	8.8	1841066
Action,Adventure,Sci-Fi	8.8	1841066

CORRELATION MATRIX

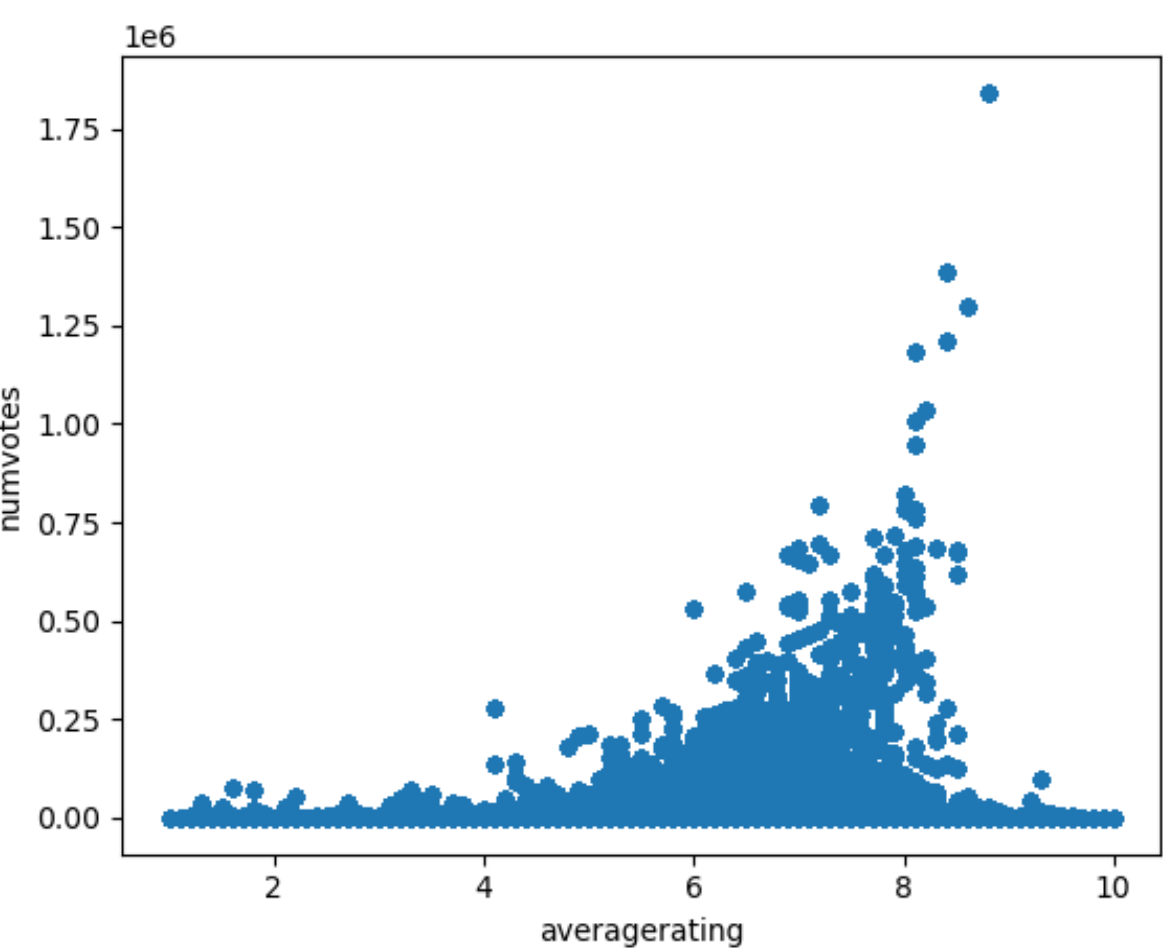|	year	averagerating	numvotes
year	1.000000	0.026788	-0.024855
averagerating	0.026788	1.000000	0.046086
numvotes	-0.024855	0.046086	1.000000

MOST USED PRODUCTION STUDIO

```
1  # find the most used production cmpany in studio column
2  final_df['studio'].mode()
```
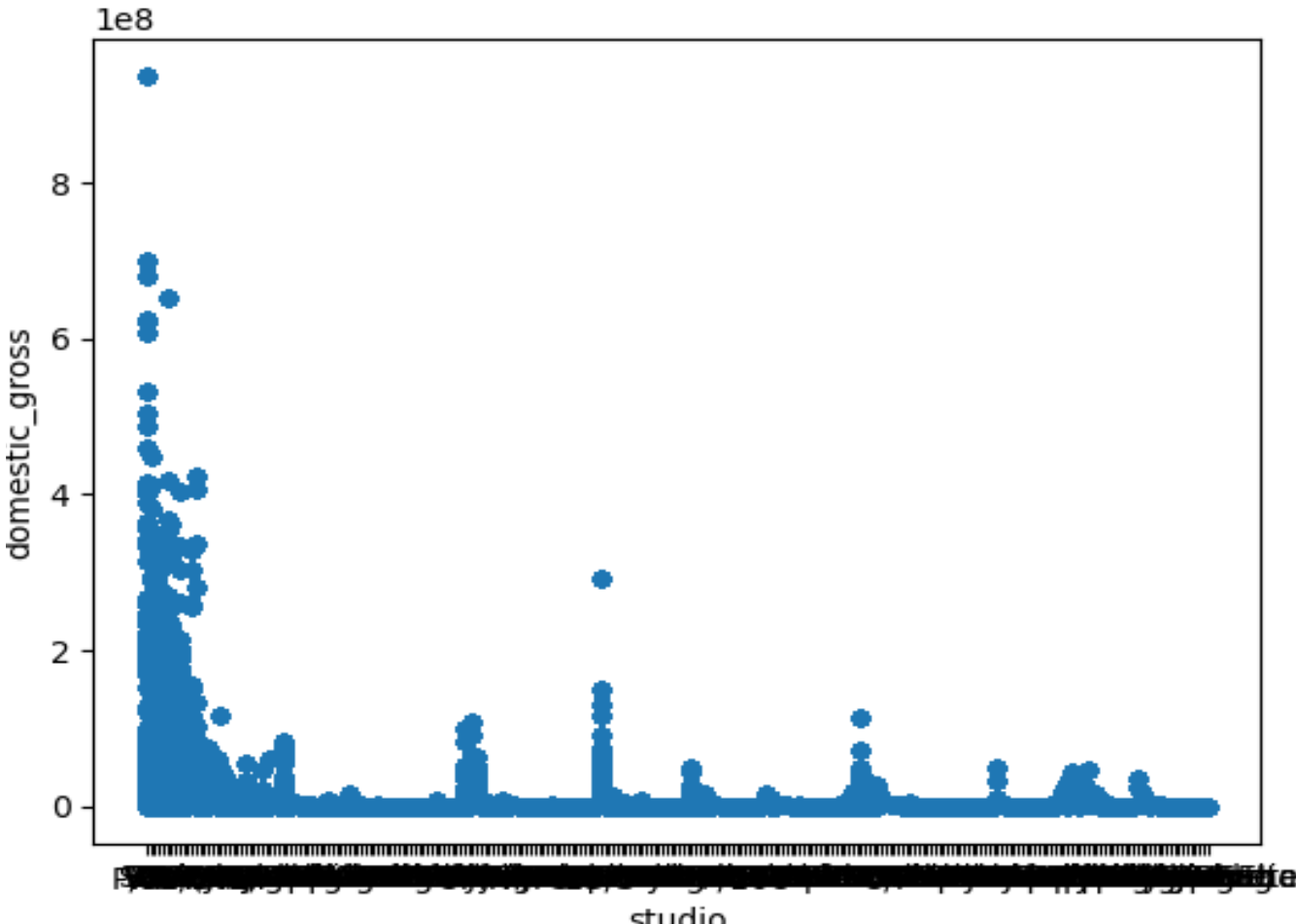
0   IFC

Scatter plot diagrams showing the correlation between votes and rating and correlation between studio types and the domestic gross income
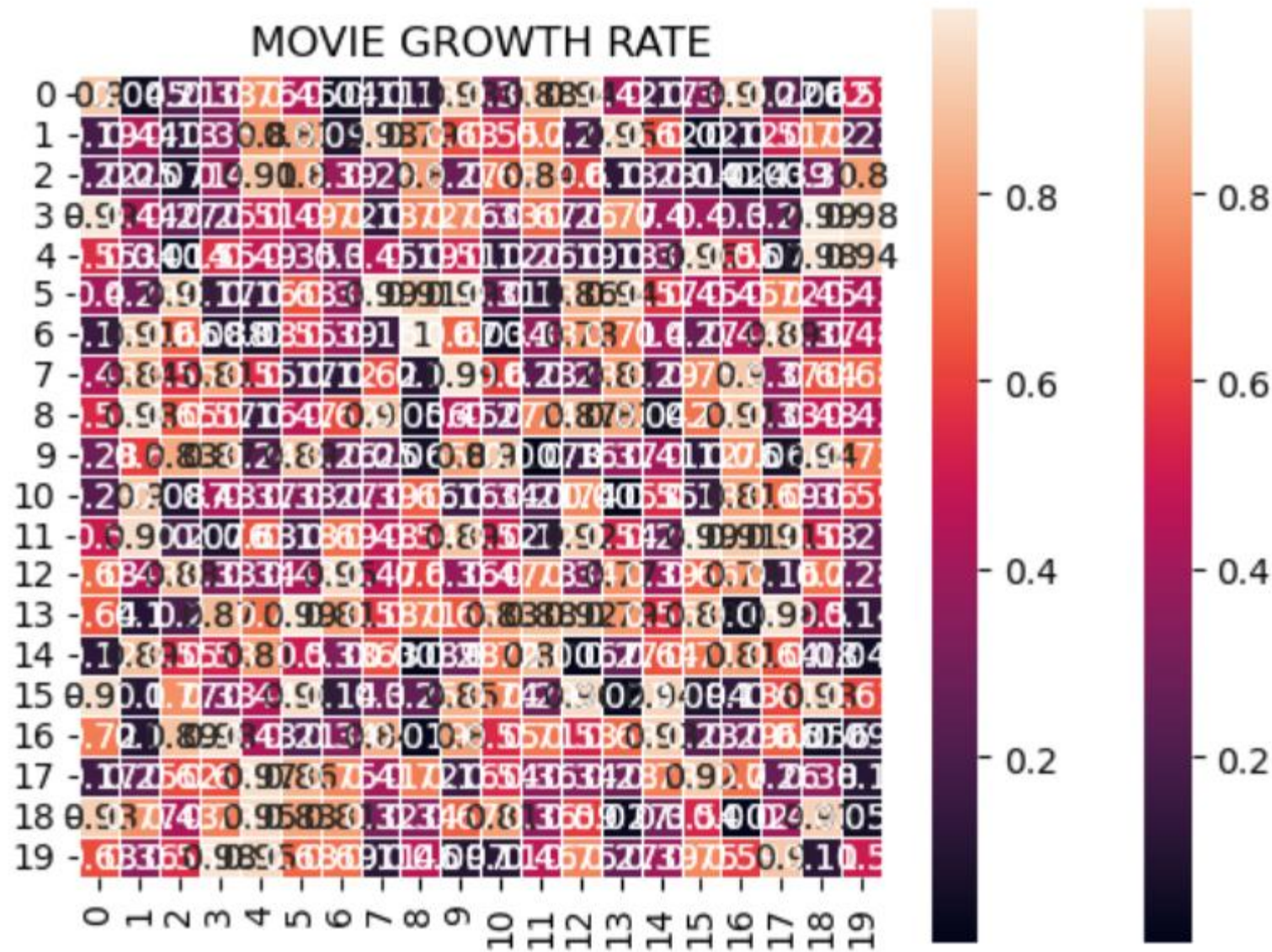
SCATTER PLOT SHOWING CORR BETWEEN VOTES AND RATING

SCATTER PLOT SHOWING CORR BETWEEN STUDIO AND GROSS INCOME

Heat Map showing the correlation between all the variables .

MOVIE GROWTH RATE

INSIGHTS AND RECOMMENDATIONS

INSIGHTS

- This is a good venture to dive into
- It is a good source of generating revenue especially based on how many people it gets to.
- This means building a site where all the movies are easily accessible for a cheaper fee.
- The year of movie production does not really matter as much as the genre and the ratings it gets .

RECOMMENDATIONS

- Focus mostly on the genre that is most famous which is scifi, action and adventure content .
- Work with a studio brand that has helped the majority of productions which is IFC.