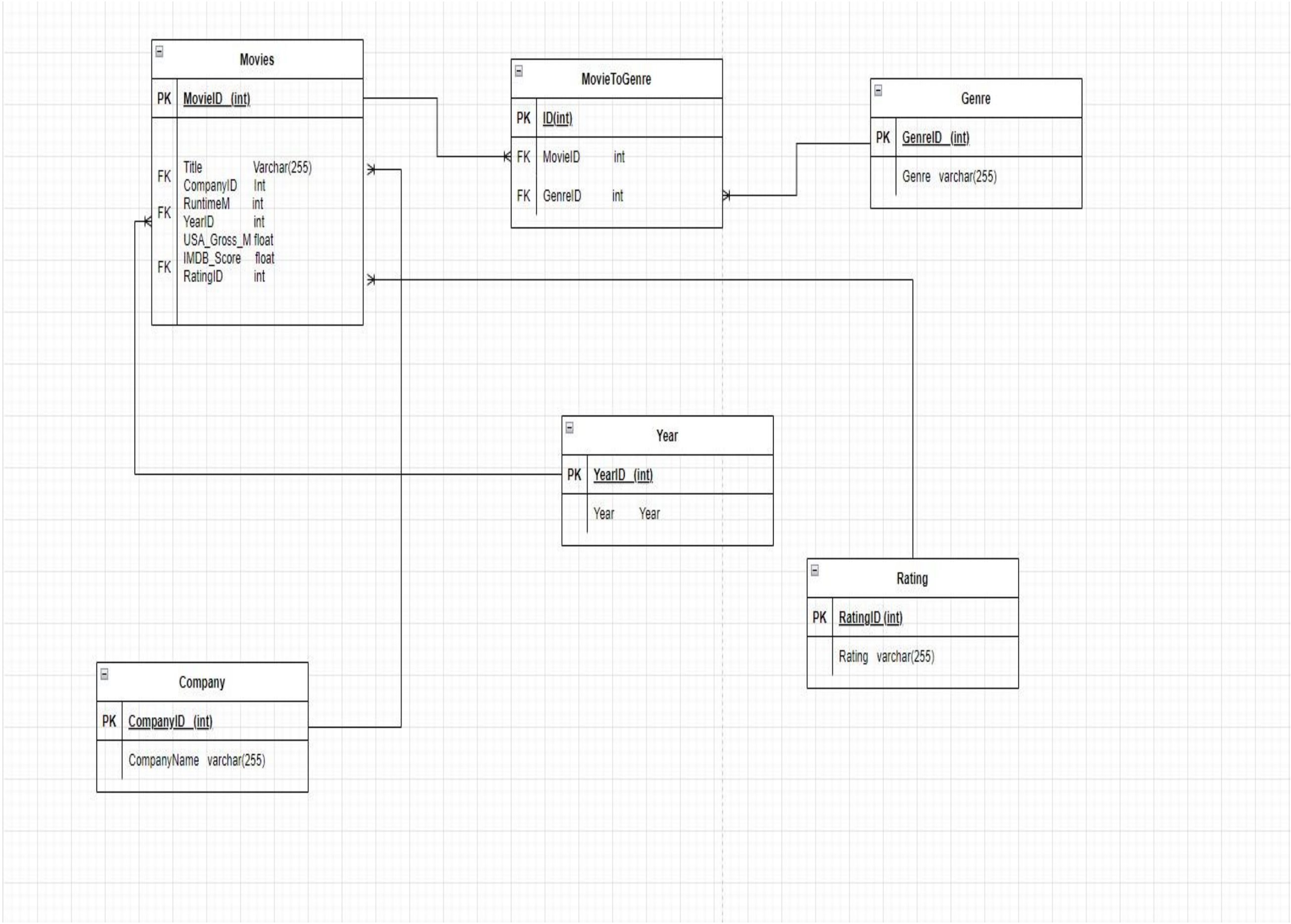
Comic Book Movie Analysis
By: Hamza Asad

KaggleSet: https://www.kaggle.com/datasets/hetulmehta/marvel-vs-dc-imdb-dataset

Database Documentation/ER Diagram:



Movies Table

Field Name	Data Type	Values	Notes
MovieID	Int	Numbers for movie entry	Primary Key
CompanyID	Int	Numbers for company entries	Foreign Key
Runtime M	Int	How long the movie is	

YearID	Year	The Year the movie was released	Foreign Key
USA_Gross_M	Float	How much the movie made in US sales	
IMDB_Score	Float	Audience rating of the movie	
RatingID	int	Rating of the movie	Foreign Key

Company Table:

Field Name	Data Type	Values	Notes
CompanyID	Int	Unique designation for each comic book film	Primary Key
CompanyName	varchar(255)	Name of the company	

Rating Table:

Field Name	Data Type	Values	Notes
Rating ID	Int	Unique designation for each Rating	Primary Key
Rating	varchar(255)	Rating Type (PG,PG-13,R,TV-MA)	

Year Table:

Field Name	Data Type	Values	Notes
YearID	Int	Unique designation for each year	Primary Key
Year	Year	Years	

Genre Table:

Field Name	Data Type	Values	Notes

GenreID	Int	Unique designation for each Genre	Primary Key
Genre	varchar(255)	Name of the Genre	

MoviesToGenre Table:

Field Name	Data Type	Values	Notes
ID	int	Unique ID for each movie to genre pairing	Primary Key
MovieID	int	ID of the Movie	Foreign Key
GenreID	int	ID of the Genre	Foreign Key

Letter From Someone at Fictional Company:

Hey Borromean Data team! Our newest client, the National Comic Book Association of North America has given us access to their datasets and wants us to do some data analysis. They are interested in movies inspired by comics and how successful they are. Some of the questions they want us to answer include:

1.) Which comic book company is the most successful in the film industry? In other words, which company has the highest average audience score and USA box office? Please also include the count of how many movies they've released.
2.) What are the Top 10 movies by USA Box office sales and what company do they belong to?
3.) In an industry dominated by big competitions and household brand names is it possible for Originals not based on previous IP or source material to be successful? How well do original movies do when compared to larger companies such as Marvel and DC in terms of box office and audience score? What are the top 5 Original films by Imdb score and Box office?
4.) People often talk about superhero movie fatigue. It seems that most tentpole movies that come out are comic book movies. What year did the most movies come out? And what year has the highest average box office sales?
5.) A lot of studios are hesitant to experiment with movies outside of pg-13. Are they right to stick with pg-13 or can other ratings such as R and PG be successful? Which ratings tend to get the highest average box office?
6.) Can rated R movies be successful? What are the top rated R movies by USA Gross, including Year and Audience score.
7.) Which Genres are overrepresented in the comic book movie landscape and which Genres are underrated?
8.) For the longest time it was assumed that movies with a runtime longer than 2 hours wouldn't be successful and that audiences simply don't have the patience to sit for longer movies. Recently however Zack Snyder's Justice League film released a 4 hour cut that

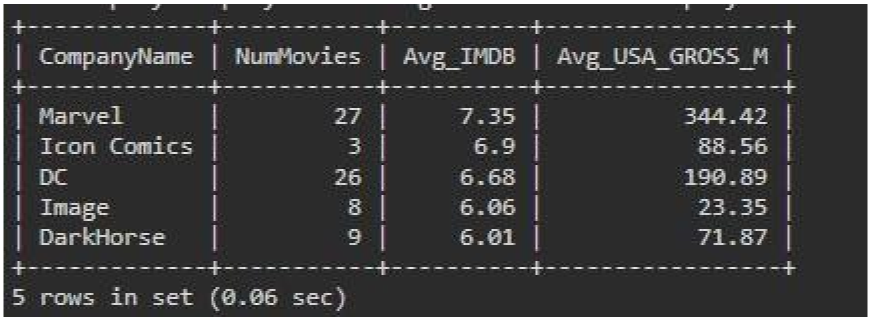
was beloved by fans. Since then the Batman also played a 3 hour theatrical cut. Is it possible for movies to surpass the 2 hour mark and still be successful? What are the Top 10 movies by runtime and USA gross?

Please have this report on my desk by Tuesday morning. Thanks, Raj.

Letter From Someone at Fictional Company:

Hey Raj! I hope you are doing well. Here is the report you requested attached below. Thanks, and have a great day!

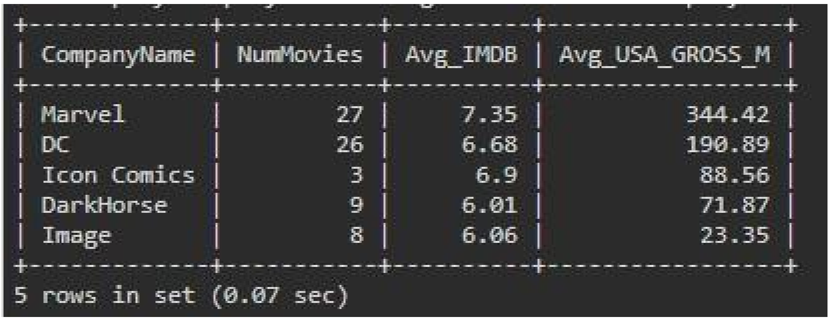
Figure 1. To answer the first question. Based on this figure it appears Marvel is the most successful when it comes to audience scores.



```
+-------------+----------+----------+-----------------+
| CompanyName | NumMovies | Avg_IMDB | Avg_USA_GROSS_M |
+-------------+----------+----------+-----------------+
| Marvel      |        27 |     7.35 |          344.42 |
| Icon Comics |         3 |      6.9 |           88.56 |
| DC          |        26 |     6.68 |          190.89 |
| Image       |         8 |     6.06 |           23.35 |
| DarkHorse   |         9 |     6.01 |           71.87 |
+-------------+----------+----------+-----------------+
5 rows in set (0.06 sec)
```

SELECT Company.CompanyName, COUNT(Movies.CompanyID) as NumMovies,ROUND(AVG(IMDB_Score),2) as Avg_IMDB,ROUND(AVG(USA_Gross_M),2) as Avg_USA_GROSS_M FROM Movies INNER JOIN Company ON Movies.CompanyID=Company.CompanyID WHERE Company.CompanyName!='Original' GROUP BY CompanyName ORDER BY AVG(IMDB_Score) DESC;

Figure 2. It looks as though Marvel is still the most popular when it comes to Avg Box office Gross.

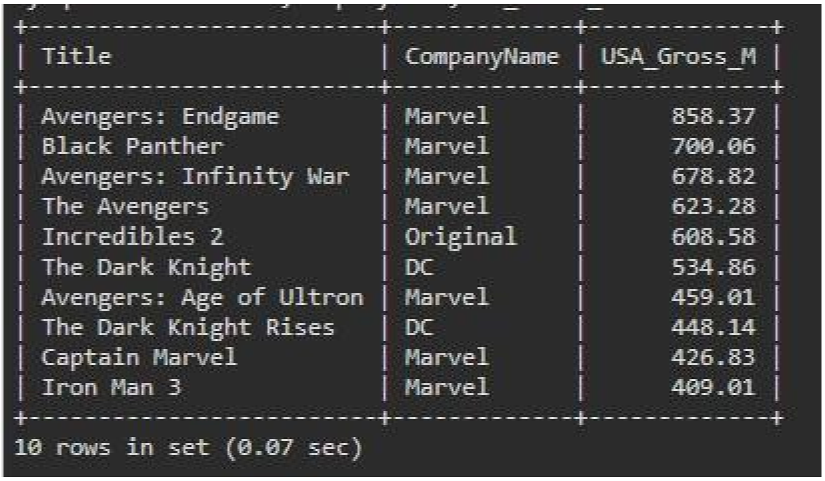


```
+-------------+----------+----------+-----------------+
| CompanyName | NumMovies | Avg_IMDB | Avg_USA_GROSS_M |
+-------------+----------+----------+-----------------+
| Marvel      |        27 |     7.35 |          344.42 |
| DC          |        26 |     6.68 |          190.89 |
| Icon Comics |         3 |      6.9 |           88.56 |
| DarkHorse   |         9 |     6.01 |           71.87 |
| Image       |         8 |     6.06 |           23.35 |
+-------------+----------+----------+-----------------+
5 rows in set (0.07 sec)
```

SELECT Company.CompanyName, COUNT(Movies.CompanyID) as NumMovies,ROUND(AVG(IMDB_Score),2) as Avg_IMDB,ROUND(AVG(USA_Gross_M),2) as Avg_USA_GROSS_M FROM Movies INNER JOIN

Company ON Movies.CompanyID=Company.CompanyID WHERE Company.CompanyName!='Original' GROUP BY CompanyName ORDER BY Avg_USA_GROSS_M DESC;

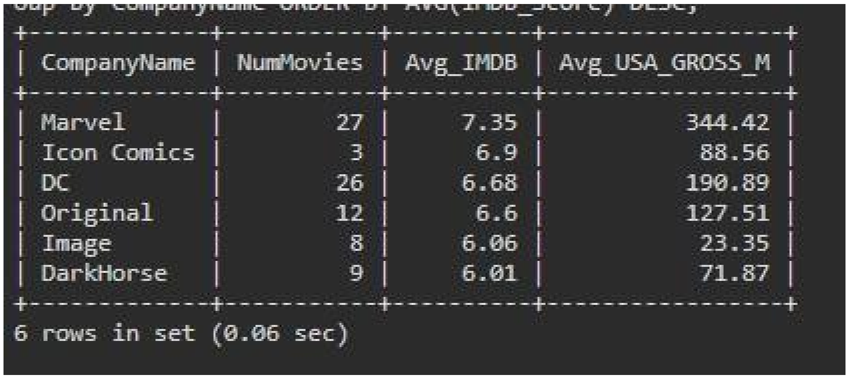
Figure 3. To answer the second question these are the top ten movies by box office draw. It appears as though Marvel is dominant with 7 films in the top 1.



```
+-------------------------+-------------+------------+
| Title                   | CompanyName | USA_Gross_M |
+-------------------------+-------------+------------+
| Avengers: Endgame       | Marvel      |      858.37 |
| Black Panther           | Marvel      |      700.06 |
| Avengers: Infinity War  | Marvel      |      678.82 |
| The Avengers            | Marvel      |      623.28 |
| Incredibles 2           | Original    |      608.58 |
| The Dark Knight         | DC          |      534.86 |
| Avengers: Age of Ultron | Marvel      |      459.01 |
| The Dark Knight Rises   | DC          |      448.14 |
| Captain Marvel          | Marvel      |      426.83 |
| Iron Man 3              | Marvel      |      409.01 |
+-------------------------+-------------+------------+
10 rows in set (0.07 sec)
```

SELECT Title,CompanyName,USA_Gross_M From Movies M INNER JOIN Company C ON M.CompanyID=C.CompanyID ORDER BY USA_Gross_M desc limit 10;

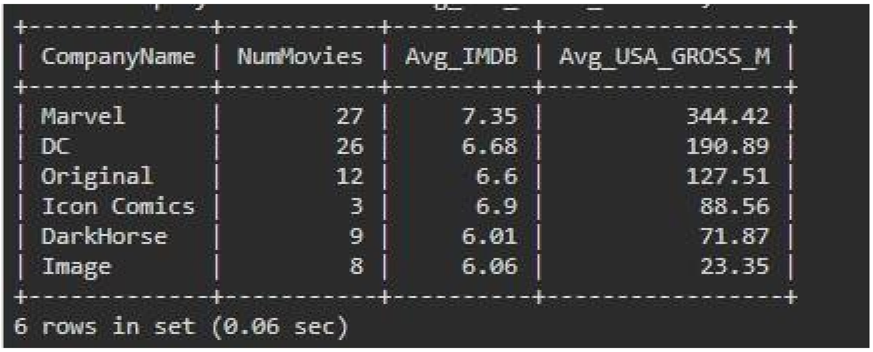
Figure 4. To answer the third question it appears that Originals rank fourth in terms of audience scores.



```
oup by CompanyName ORDER BY AVG(IMDB_Score) DESC;
+-------------+-----------+----------+-----------------+
| CompanyName | NumMovies | Avg_IMDB | Avg_USA_GROSS_M |
+-------------+-----------+----------+-----------------+
| Marvel      |        27 |     7.35 |          344.42 |
| Icon Comics |         3 |      6.9 |           88.56 |
| DC          |        26 |     6.68 |          190.89 |
| Original    |        12 |      6.6 |          127.51 |
| Image       |         8 |     6.06 |           23.35 |
| DarkHorse   |         9 |     6.01 |           71.87 |
+-------------+-----------+----------+-----------------+
6 rows in set (0.06 sec)
```

SELECT Company.CompanyName, COUNT(Movies.CompanyID) as NumMovies,ROUND(AVG(IMDB_Score),2) as Avg_IMDB,ROUND(AVG(USA_Gross_M),2) as Avg_USA_GROSS_M FROM Movies INNER JOIN Company ON Movies.CompanyID=Company.CompanyID GROUP BY CompanyName ORDER BY Avg_IMDB DESC;

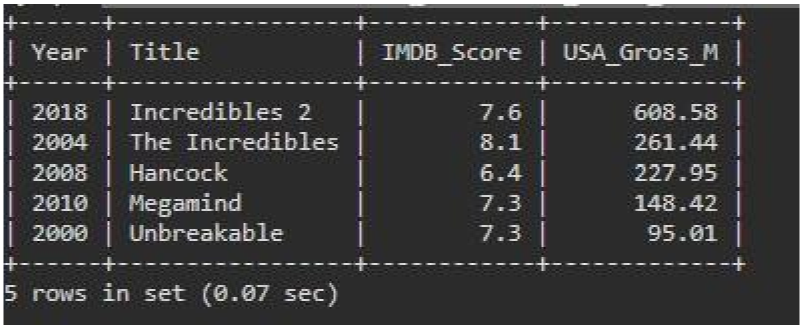
Figure 4. Continuing with the third question, Originals rank in third when ranking based off Avg USA gross.



```
+--------------+------------+----------+-----------------+
| CompanyName  | NumMovies  | Avg_IMDB | Avg_USA_GROSS_M |
+--------------+------------+----------+-----------------+
| Marvel       |         27 |     7.35 |          344.42 |
| DC           |         26 |     6.68 |          190.89 |
| Original     |         12 |      6.6 |          127.51 |
| Icon Comics  |          3 |      6.9 |           88.56 |
| DarkHorse    |          9 |     6.01 |           71.87 |
| Image        |          8 |     6.06 |           23.35 |
+--------------+------------+----------+-----------------+
6 rows in set (0.06 sec)
```

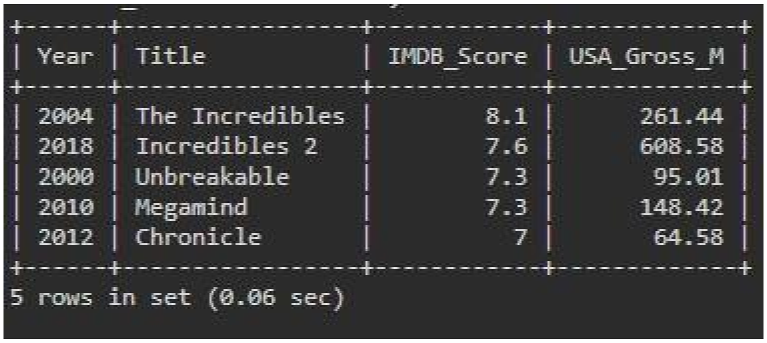
SELECT Company.CompanyName, COUNT(Movies.CompanyID) as NumMovies,ROUND(AVG(IMDB_Score),2) as Avg_IMDB,ROUND(AVG(USA_Gross_M),2) as Avg_USA_GROSS_M FROM Movies INNER JOIN Company ON Movies.CompanyID=Company.CompanyID GROUP BY CompanyName ORDER BY Avg_USA_GROSS_M DESC;

Figure 5. This is the second part of the third question they asked for.



```
+------+-----------------+------------+------------+
| Year | Title           | IMDB_Score | USA_Gross_M |
+------+-----------------+------------+------------+
| 2018 | Incredibles 2   |        7.6 |     608.58 |
| 2004 | The Incredibles |        8.1 |     261.44 |
| 2008 | Hancock         |        6.4 |     227.95 |
| 2010 | Megamind        |        7.3 |     148.42 |
| 2000 | Unbreakable     |        7.3 |      95.01 |
+------+-----------------+------------+------------+
5 rows in set (0.07 sec)
```

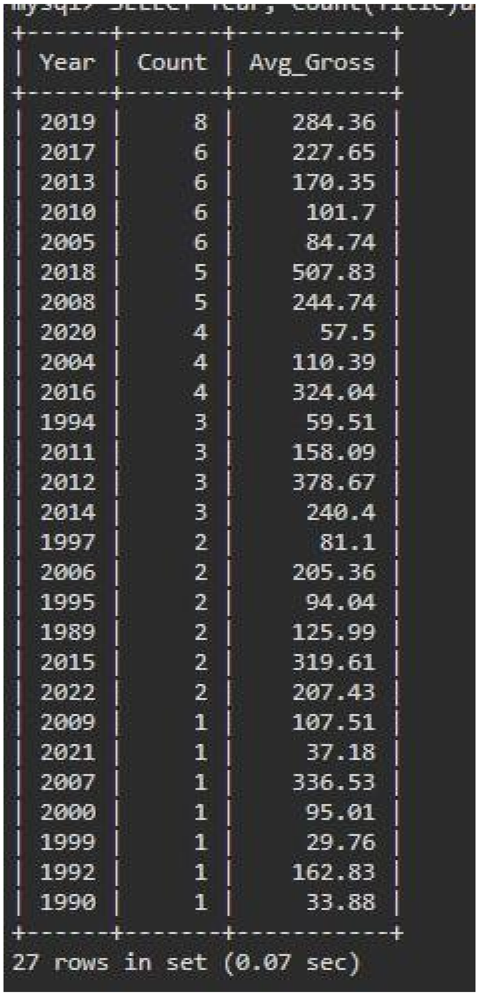
SELECT Year,Title,IMDB_Score, USA_Gross_M from Movies M INNER JOIN Company C ON M.CompanyID=C.CompanyID JOIN Year ON M.YearID=Year.YearID WHERE CompanyName='Original' ORDER BY USA_Gross_M desc limit 5;



```
+------+-----------------+------------+-------------+
| Year | Title           | IMDB_Score | USA_Gross_M |
+------+-----------------+------------+-------------+
| 2004 | The Incredibles |        8.1 |      261.44 |
| 2018 | Incredibles 2   |        7.6 |      608.58 |
| 2000 | Unbreakable     |        7.3 |       95.01 |
| 2010 | Megamind        |        7.3 |      148.42 |
| 2012 | Chronicle       |          7 |       64.58 |
+------+-----------------+------------+-------------+
5 rows in set (0.06 sec)
```

SELECT Year,Title,IMDB_Score, USA_Gross_M from Movies M INNER JOIN Company C ON M.CompanyID=C.CompanyID JOIN Year ON M.YearID=Year.YearID WHERE CompanyName='Original' ORDER BY IMDB_Score desc limit 5;

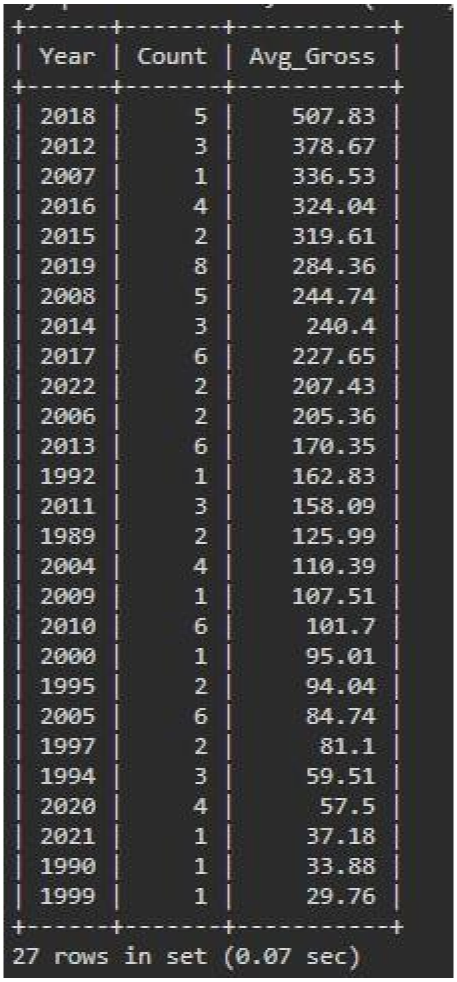
Figure 6. To answer the fourth question it appears that superhero movies are indeed increasing in frequency. 2019 had the highest number of movies.



```
mysql> SELECT Year, Count(Title)a
+------+-------+-----------+
| Year | Count | Avg_Gross |
+------+-------+-----------+
| 2019 |     8 |    284.36 |
| 2017 |     6 |    227.65 |
| 2013 |     6 |    170.35 |
| 2010 |     6 |     101.7 |
| 2005 |     6 |     84.74 |
| 2018 |     5 |    507.83 |
| 2008 |     5 |    244.74 |
| 2020 |     4 |      57.5 |
| 2004 |     4 |    110.39 |
| 2016 |     4 |    324.04 |
| 1994 |     3 |     59.51 |
| 2011 |     3 |    158.09 |
| 2012 |     3 |    378.67 |
| 2014 |     3 |     240.4 |
| 1997 |     2 |      81.1 |
| 2006 |     2 |    205.36 |
| 1995 |     2 |     94.04 |
| 1989 |     2 |    125.99 |
| 2015 |     2 |    319.61 |
| 2022 |     2 |    207.43 |
| 2009 |     1 |    107.51 |
| 2021 |     1 |     37.18 |
| 2007 |     1 |    336.53 |
| 2000 |     1 |     95.01 |
| 1999 |     1 |     29.76 |
| 1992 |     1 |    162.83 |
| 1990 |     1 |     33.88 |
+------+-------+-----------+
27 rows in set (0.07 sec)
```

SELECT Year, Count(Title)as Count, Round(AVG(USA_Gross_M),2) as Avg_Gross FROM Year Y INNER JOIN Movies M ON Y.YearID=M.YearID Group By Year Order by Count desc;

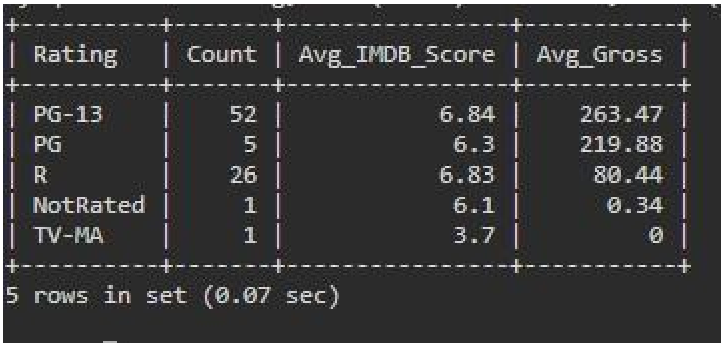
Figure 7. Continuing with the question four 2018 had the highest average gross.



```
+------+-------+-----------+
| Year | Count | Avg_Gross |
+------+-------+-----------+
| 2018 |     5 |    507.83 |
| 2012 |     3 |    378.67 |
| 2007 |     1 |    336.53 |
| 2016 |     4 |    324.04 |
| 2015 |     2 |    319.61 |
| 2019 |     8 |    284.36 |
| 2008 |     5 |    244.74 |
| 2014 |     3 |     240.4 |
| 2017 |     6 |    227.65 |
| 2022 |     2 |    207.43 |
| 2006 |     2 |    205.36 |
| 2013 |     6 |    170.35 |
| 1992 |     1 |    162.83 |
| 2011 |     3 |    158.09 |
| 1989 |     2 |    125.99 |
| 2004 |     4 |    110.39 |
| 2009 |     1 |    107.51 |
| 2010 |     6 |     101.7 |
| 2000 |     1 |     95.01 |
| 1995 |     2 |     94.04 |
| 2005 |     6 |     84.74 |
| 1997 |     2 |      81.1 |
| 1994 |     3 |     59.51 |
| 2020 |     4 |      57.5 |
| 2021 |     1 |     37.18 |
| 1990 |     1 |     33.88 |
| 1999 |     1 |     29.76 |
+------+-------+-----------+
27 rows in set (0.07 sec)
```

SELECT Year, Count(Title)as Count, Round(AVG(USA_Gross_M),2) as Avg_Gross FROM Year Y INNER JOIN Movies M ON Y.YearID=M.YearID Group By Year Order by Avg_Gross desc;

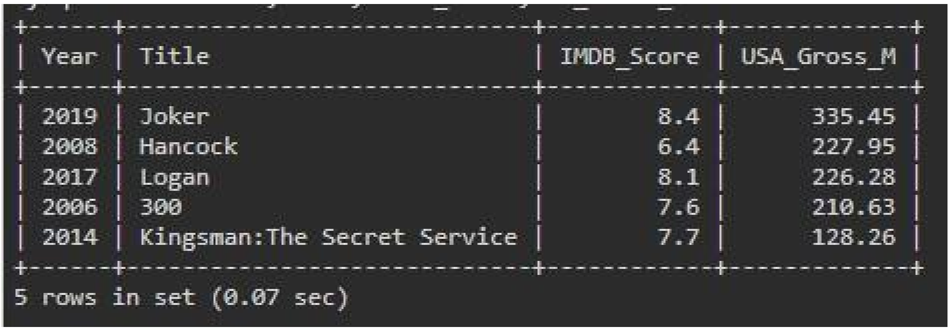
Figure 7. Answering question 4 PG 13 tends to get the highest average earnings but they also have the most films. PG and R tend to do pretty well.



```
+----------+-------+----------------+-----------+
| Rating   | Count | Avg_IMDB_Score | Avg_Gross |
+----------+-------+----------------+-----------+
| PG-13    |    52 |           6.84 |    263.47 |
| PG       |     5 |            6.3 |    219.88 |
| R        |    26 |           6.83 |     80.44 |
| NotRated |     1 |            6.1 |      0.34 |
| TV-MA    |     1 |            3.7 |         0 |
+----------+-------+----------------+-----------+
5 rows in set (0.07 sec)
```

SELECT Rating,COUNT(Title) as Count , ROUND(AVG(IMDB_Score),2) as Avg_IMDB_Score,ROUND(AVG(USA_Gross_M),2) as Avg_Gross FROM Rating R INNER JOIN Movies M ON R.RatingID=M.RatingID GROUP BY Rating ORDER BY Avg_Gross desc;

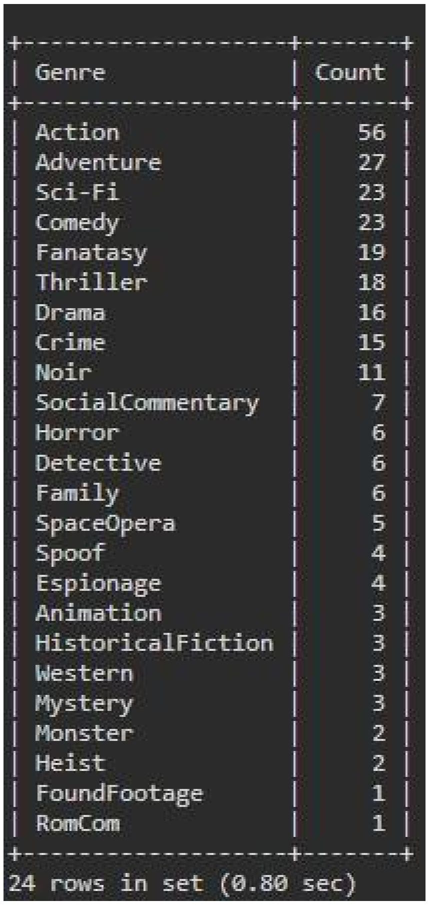
Figure 8. Answering question 6 it does appear that Rated R comic book movies are financially viable and beloved by audiences.



```
+------+---------------------------+-----------+-------------+
| Year | Title                     | IMDB_Score | USA_Gross_M |
+------+---------------------------+-----------+-------------+
| 2019 | Joker                     |       8.4 |      335.45 |
| 2008 | Hancock                   |       6.4 |      227.95 |
| 2017 | Logan                     |       8.1 |      226.28 |
| 2006 | 300                       |       7.6 |      210.63 |
| 2014 | Kingsman:The Secret Service |     7.7 |      128.26 |
+------+---------------------------+-----------+-------------+
5 rows in set (0.07 sec)
```

SELECT Year,Title, IMDB_Score,USA_Gross_M FROM Movies M INNER JOIN Year Y ON M.YearID=Y.YearID INNER JOIN Rating R ON M.RatingID=R.RatingID WHERE Rating='R' ORDER BY USA_Gross_M Desc limit 5;

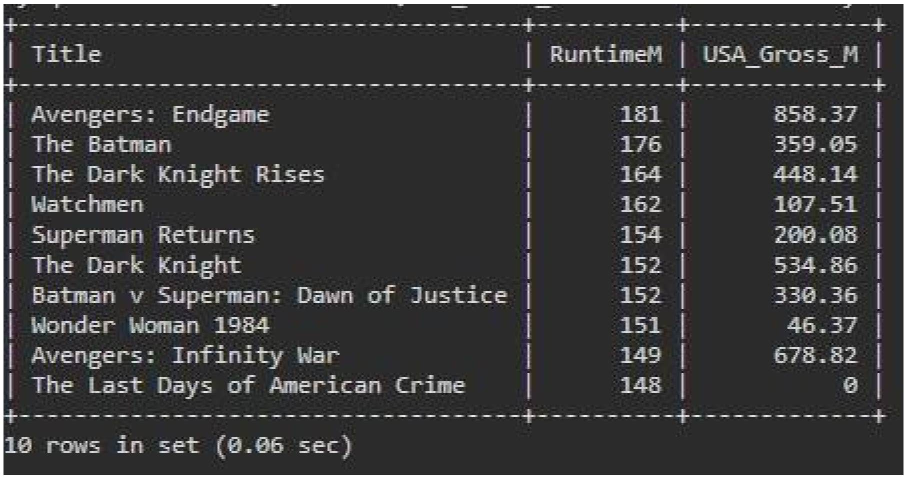
Figure 9. To answer question 7 Action Movies seem to have significant representation in the comicbook film landscape leaving much room to be desired for Heist, Monster, FoundFootage, and Romcoms.



```
+-------------------+-------+
| Genre             | Count |
+-------------------+-------+
| Action            |    56 |
| Adventure         |    27 |
| Sci-Fi            |    23 |
| Comedy            |    23 |
| Fanatasy          |    19 |
| Thriller          |    18 |
| Drama             |    16 |
| Crime             |    15 |
| Noir              |    11 |
| SocialCommentary  |     7 |
| Horror            |     6 |
| Detective         |     6 |
| Family            |     6 |
| SpaceOpera        |     5 |
| Spoof             |     4 |
| Espionage         |     4 |
| Animation         |     3 |
| HistoricalFiction |     3 |
| Western           |     3 |
| Mystery           |     3 |
| Monster           |     2 |
| Heist             |     2 |
| FoundFootage      |     1 |
| RomCom            |     1 |
+-------------------+-------+
24 rows in set (0.80 sec)
```

SELECT Genre, Count(Title) as Count FROM Genre G INNER JOIN MovieToGenre MG ON G.GenreID=MG.GenreID INNER JOIN Movies M ON MG.MovieID=M.MovieID GROUP BY Genre ORDER BY Count DESC;

Figure10. To answer question 8 it appears that having a long runtime isnt necessarily a deterrent for USA Gross.



```
+------------------------------------+----------+-------------+
| Title                              | RuntimeM | USA_Gross_M |
+------------------------------------+----------+-------------+
| Avengers: Endgame                  |      181 |      858.37 |
| The Batman                         |      176 |      359.05 |
| The Dark Knight Rises              |      164 |      448.14 |
| Watchmen                           |      162 |      107.51 |
| Superman Returns                   |      154 |      200.08 |
| The Dark Knight                    |      152 |      534.86 |
| Batman v Superman: Dawn of Justice |      152 |      330.36 |
| Wonder Woman 1984                  |      151 |       46.37 |
| Avengers: Infinity War             |      149 |      678.82 |
| The Last Days of American Crime    |      148 |           0 |
+------------------------------------+----------+-------------+
10 rows in set (0.06 sec)
```

SELECT Title,RuntimeM,USA_Gross_M from Movies Order By RuntimeM desc limit 10;

Outline of the Database Design Choices Made:

When designing the database I came up with the questions that I was interested in looking into before designing the database,and then tried to construct the database revolving around those questions. For instance, I wasn’t very interested in questions relating to cast and crew members so I excluded the director and cast columns. I was interested in questions related to company, genre, and rating and it only seemed natural to give those categories their own tables because each of those categories divide into further categories. For instance there are different Ratings such as PG,PG-13,R,TV-MA and it seemed the natural thing to do was to give each one its own unique designation and key.The same was true for the company, genre, and year columns. I also decided to do the same for the Year category. In regards to the design of Genre I realized that it was a multiple to multiple relationship as one movie could have multiple genres so I needed to design an intermediary between the genre and movie tables. Thus I created a Movie to Genre table. Every other relationship was a one to many relationship so there were no other intermediary tables necessary.