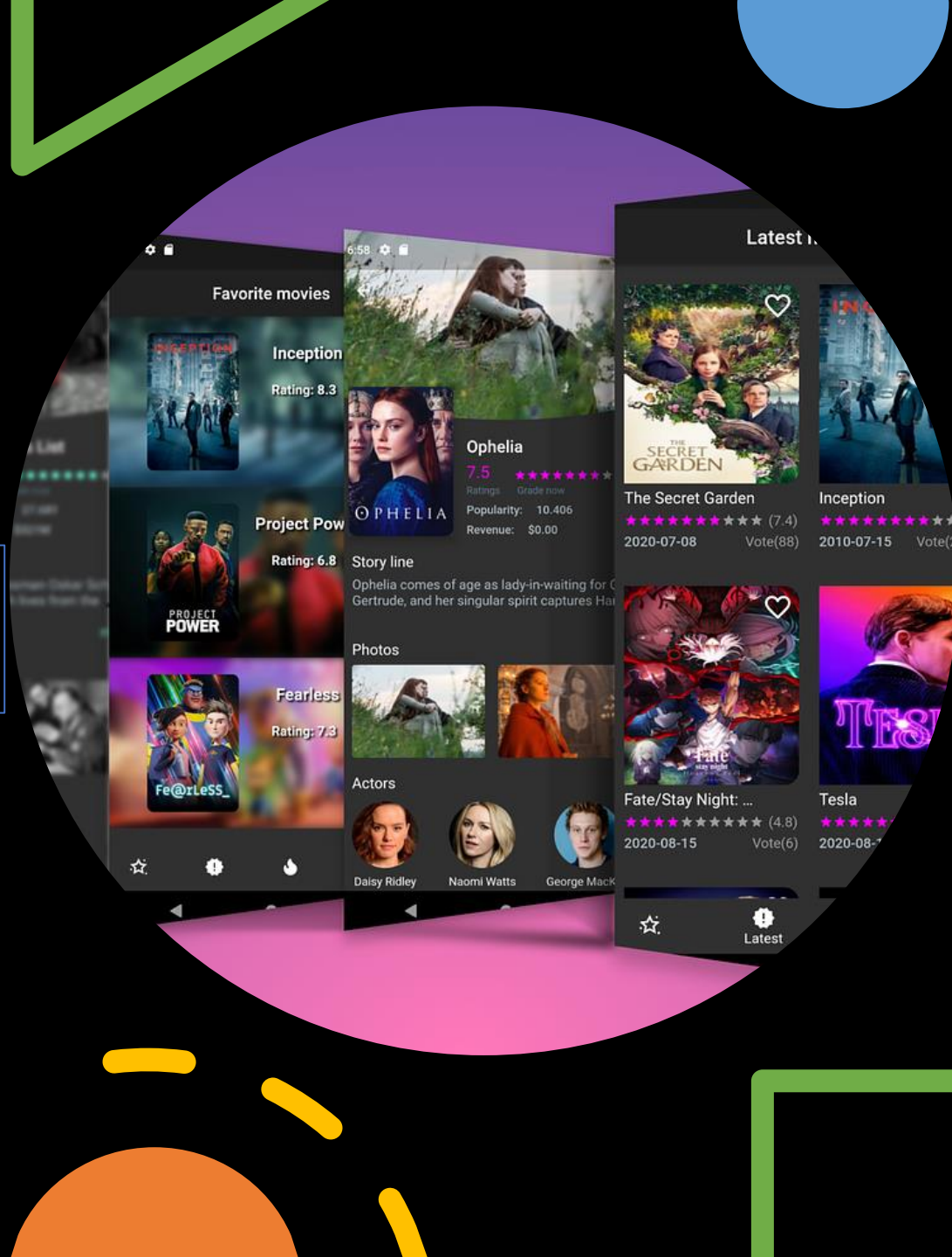# TMDB MOVIE DATA ANALYSIS

BY NAINCY KOTANGLE

- Objective:
- Analyze Movie Dataset to Understand Movie Success in Cinemas.
- Identify Successful Genres, Correlate Budget vs. Revenue, Analyze Ratings'.
- Impact, Assess Star Power, and Develop Predictive Model.
- Importance of Data-Driven Decision-Making in Film Industry.

Dataset
4803 rows × 20 columns

Columns present in dataset

```
In [66]:   1  df.columns

Out[66]:  Index(['budget', 'genres', 'homepage', 'id', 'keywords', 'original_language',
                 'original_title', 'overview', 'popularity', 'production_companies',
                 'production_countries', 'release_date', 'revenue', 'runtime',
                 'spoken_languages', 'status', 'tagline', 'title', 'vote_average',
                 'vote_count'],
                dtype='object')
```

Description for numerical columns

```
1  df.describe()
```

	budget	id	popularity	revenue	runtime	vote_average	vote_count
count	4.803000e+03	4803.000000	4803.000000	4.803000e+03	4801.000000	4803.000000	4803.000000
mean	2.904504e+07	57165.484281	21.492301	8.226064e+07	106.875859	6.092172	690.217989
std	4.072239e+07	88694.614033	31.816650	1.628571e+08	22.611935	1.194612	1234.585891
min	0.000000e+00	5.000000	0.000000	0.000000e+00	0.000000	0.000000	0.000000
25%	7.900000e+05	9014.500000	4.668070	0.000000e+00	94.000000	5.600000	54.000000
50%	1.500000e+07	14629.000000	12.921594	1.917000e+07	103.000000	6.200000	235.000000
75%	4.000000e+07	58610.500000	28.313505	9.291719e+07	118.000000	6.800000	737.000000
max	3.800000e+08	459488.000000	875.581305	2.787965e+09	338.000000	10.000000	13752.000000

Titles and genres of first 50 movies

	genres	original_title
0	[Action, Adventure, Fantasy, Science Fiction]	Avatar
1	[Adventure, Fantasy, Action]	Pirates of the Caribbean: At World's End
2	[Action, Adventure, Crime]	Spectre
3	[Action, Crime, Drama, Thriller]	The Dark Knight Rises
4	[Action, Adventure, Science Fiction]	John Carter
5	[Fantasy, Action, Adventure]	Spider-Man 3
6	[Animation, Family]	Tangled
7	[Action, Adventure, Science Fiction]	Avengers: Age of Ultron
8	[Adventure, Fantasy, Family]	Harry Potter and the Half-Blood Prince
9	[Action, Adventure, Fantasy]	Batman v Superman: Dawn of Justice
10	[Adventure, Fantasy, Action, Science Fiction]	Superman Returns
11	[Adventure, Action, Thriller, Crime]	Quantum of Solace
12	[Adventure, Fantasy, Action]	Pirates of the Caribbean: Dead Man's Chest
13	[Action, Adventure, Western]	The Lone Ranger
14	[Action, Adventure, Fantasy, Science Fiction]	Man of Steel
15	[Adventure, Family, Fantasy]	The Chronicles of Narnia: Prince Caspian
16	[Science Fiction, Action, Adventure]	The Avengers
17	[Adventure, Action, Fantasy]	Pirates of the Caribbean: On Stranger Tides
18	[Action, Comedy, Science Fiction]	Men in Black 3
19	[Action, Adventure, Fantasy]	The Hobbit: The Battle of the Five Armies

Null values

```
In [12]:    1  df.isnull().sum()

Out[12]:  budget                     0
          genres                     0
          homepage                3091
          id                         0
          keywords                   0
          original_language          0
          original_title             0
          overview                   3
          popularity                 0
          production_companies       0
          production_countries       0
          release_date               1
          revenue                    0
          runtime                    2
          spoken_languages           0
          status                     0
          tagline                  844
          title                      0
          vote_average               0
          vote_count                 0
          dtype: int64
```

null value treatment

```
In [78]:  1  # Replace null values with "Not Mentioned"
          2  df_filled = df.fillna("Not Mentioned")
          3  print(df_filled)
```

```
            budget                                        genres  \
0        237000000  [Action, Adventure, Fantasy, Science Fiction]
1        300000000                    [Adventure, Fantasy, Action]
2        245000000                       [Action, Adventure, Crime]
3        250000000               [Action, Crime, Drama, Thriller]
4        260000000          [Action, Adventure, Science Fiction]
...            ...                                            ...
4798        220000                    [Action, Crime, Thriller]
4799          9000                             [Comedy, Romance]
4800             0           [Comedy, Drama, Romance, TV Movie]
4801             0                                            []
4802             0                               [Documentary]

                                       homepage       id  \
0                   http://www.avatarmovie.com/    19995
1        http://disney.go.com/disneypictures/pirates/      285
2         http://www.sonypictures.com/movies/spectre/   206647
3              http://www.thedarknightrises.com/    49026
4              http://movies.disney.com/john-carter    49529
```

In the dataset, there are some movies for which the budget and revenue columns have the value 0 which mean unknown values .Removed the rows with value 0 from both the budget and revenue columns.

```
1  df_filled
```

	budget	genres	homepage	id	keywords	original_language	original_title	overview	popularity	productio
0	237000000	[Action, Adventure, Fantasy, Science Fiction]	http://www.avatarmovie.com/	19995	[culture clash, future, space war, space colon...	en	Avatar	In the 22nd century, a paraplegic Marine is di...	150.437577	[{"nam Film
1	300000000	[Adventure, Fantasy, Action]	http://disney.go.com/disneypictures/pirates/	285	[ocean, drug abuse, exotic island, east india ...	en	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	139.082615	[{"name' Picture
2	245000000	[Action, Adventure, Crime]	http://www.sonypictures.com/movies/spectre/	206647	[spy, based on novel, secret agent, sequel, mi...	en	Spectre	A cryptic message from Bond's past sends him o...	107.376788	[{"nam Pic
3	250000000	[Action, Crime, Drama, Thriller]	http://www.thedarkknightrises.com/	49026	[dc comics, crime fighter, terrorist, secret i...	en	The Dark Knight Rises	Following the death of District Attorney Harve...	112.312950	[{"name Pictures",
4	260000000	[Action, Adventure, Science Fiction]	http://movies.disney.com/john-carter	49529	[based on novel, mars, medallion, space travel...	en	John Carter	John Carter is a war-weary, former military ca...	43.926995	[{"name' Pic

Removed the rows with value 0 from the revenue columns

```
1  df_filled = df_filled.drop(df_filled[df_filled['revenue'] == 0].index)
```

```
1  df_filled
```

	budget	genres	homepage	id	keywords	original_language	original_title	overview	popularity	pr
0	237000000	[Action, Adventure, Fantasy, Science Fiction]	http://www.avatarmovie.com/	19995	[culture clash, future, space war, space colon...	en	Avatar	In the 22nd century, a paraplegic Marine is di...	150.437577	
1	300000000	[Adventure, Fantasy, Action]	http://disney.go.com/disneypictures/pirates/	285	[ocean, drug abuse, exotic island, east india ...	en	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	139.082615	
2	245000000	[Action, Adventure, Crime]	http://www.sonypictures.com/movies/spectre/	206647	[spy, based on novel, secret agent, sequel, mi...	en	Spectre	A cryptic message from Bond's past sends him o...	107.376788	
3	250000000	[Action, Crime, Drama, Thriller]	http://www.thedarkknightrises.com/	49026	[dc comics, crime fighter, terrorist, secret i...	en	The Dark Knight Rises	Following the death of District Attorney Harve...	112.312950	P
4	260000000	[Action, Adventure, Science Fiction]	http://movies.disney.com/john-carter	49529	[based on novel, mars, medallion, space travel...	en	John Carter	John Carter is a war-weary, former military ca...	43.926995	
...	...	...		...	...	...	...	...	...	
4773	27000	[Comedy]	http://www.miramax.com/movie/clerks/	2292	[salesclerk, loser, aftercreditsstinger]	en	Clerks	Convenience and video store clerks Dante and R...	19.748658	
		[Horror			[gay, trailer park			Notorious		

movie categories that have a budget greater than $220,000

```
1  movies_higher_budget
```

	budget	genres	homepage	id	keywords	original_language	original_title	overview	popularity	productio
0	237000000	[Action, Adventure, Fantasy, Science Fiction]	http://www.avatarmovie.com/	19995	[culture clash, future, space war, space colon...	en	Avatar	In the 22nd century, a paraplegic Marine is di...	150.437577	[{"nam Film
1	300000000	[Adventure, Fantasy, Action]	http://disney.go.com/disneypictures/pirates/	285	[ocean, drug abuse, exotic island, east india ...	en	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	139.082615	[{"name" Pictures
2	245000000	[Action, Adventure, Crime]	http://www.sonypictures.com/movies/spectre/	206647	[spy, based on novel, secret agent, sequel, mi...	en	Spectre	A cryptic message from Bond's past sends him o...	107.376788	[{"name Pic
3	250000000	[Action, Crime, Drama, Thriller]	http://www.thedarkknightrises.com/	49026	[dc comics, crime fighter, terrorist, secret i...	en	The Dark Knight Rises	Following the death of District Attorney Harve...	112.312950	[{"name Pictures",
4	260000000	[Action, Adventure, Science Fiction]	http://movies.disney.com/john-carter	49529	[based on novel, mars, medallion, space travel...	en	John Carter	John Carter is a war-weary, former military ca...	43.926995	[{"name" Pic
...	...	...		...	...		...	...	...	...
4680	450000	[Crime, Horror, Thriller]	http://www.malevolencemovie.com/	54702	[hostage, psychopath, abandoned house, serial	en	Malevolence	It's ten years after the kidnapping of Martin	1.077321	

movie categories where the revenue is greater than $961,000,000

```
In [18]:   1   movies_higher_revenue = df_filled[(df_filled["revenue"]>961000000)]
```

```
In [19]:   1   movies_higher_revenue
```

original_title	overview	popularity	production_companies	production_countries	release_date	revenue	runtime	spoken_languages	status	tagline	ti
Avatar	In the 22nd century, a paraplegic Marine is di...	150.437577	[{"name": "Ingenious Film Partners", "id": 289...	[{"iso_3166_1": "US", "name": "United States o...	10-12-2009	2787965087	162.0	[{"iso_639_1": "en", "name": "English"}, {"iso...	Released	Enter the World of Pandora.	Ava
The Dark Knight Rises	Following the death of District Attorney Harve...	112.312950	[{"name": "Legendary Pictures", "id": 923}, {"...	[{"iso_3166_1": "US", "name": "United States o...	16-07-2012	1084939099	165.0	[{"iso_639_1": "en", "name": "English"}]	Released	The Legend Ends	The Da Knight Ris
Avengers: Age of Ultron	When Tony Stark tries to jumpstart a dormant p...	134.279229	[{"name": "Marvel Studios", "id": 420}, {"name...	[{"iso_3166_1": "US", "name": "United States o...	22-04-2015	1405403694	141.0	[{"iso_639_1": "en", "name": "English"}]	Released	A New Age Has Come.	Avenge Age of Ultr
Pirates of the Caribbean: Dead Man's	Captain Jack Sparrow works his	145.847379	[{"name": "Walt Disney Pictures", "id": 2},...	[{"iso_3166_1": "JM", "name": "Jamaica"},...	20-06-2006	1065659812	151.0	[{"iso_639_1": "en", "name": "English"},...	Released	Jack is back!	Pirates of Caribbe Dead Ma

the top 10 movies with the least budget

```
3]: 4238              Modern Times
    3611      A Farewell to Arms
    3372            Split Second
    3419            Bran Nue Dae
    4608            The Prophecy
    3131            Hross í oss
    3137              Nurse 3-D
    2933                F.I.S.T.
    1912          Angela's Ashes
    1771          The 51st State
    Name: original_title, dtype: object
```

the top 10 movies with the highest revenues

```
top 10 highest revenue movies:

 0                              Avatar
25                              Titanic
16                         The Avengers
28                       Jurassic World
44                            Furious 7
 7            Avengers: Age of Ultron
124                              Frozen
31                          Iron Man 3
546                             Minions
26          Captain America: Civil War
Name: original_title, dtype: object
```

How are popularities of movies related with the movie budgets? Are they correlated or totally uncorrelated with each other? Write the interpretation of your analysis
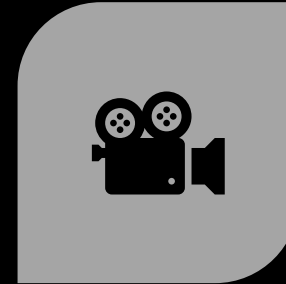

Relationship between Movie Budgets and Popularity
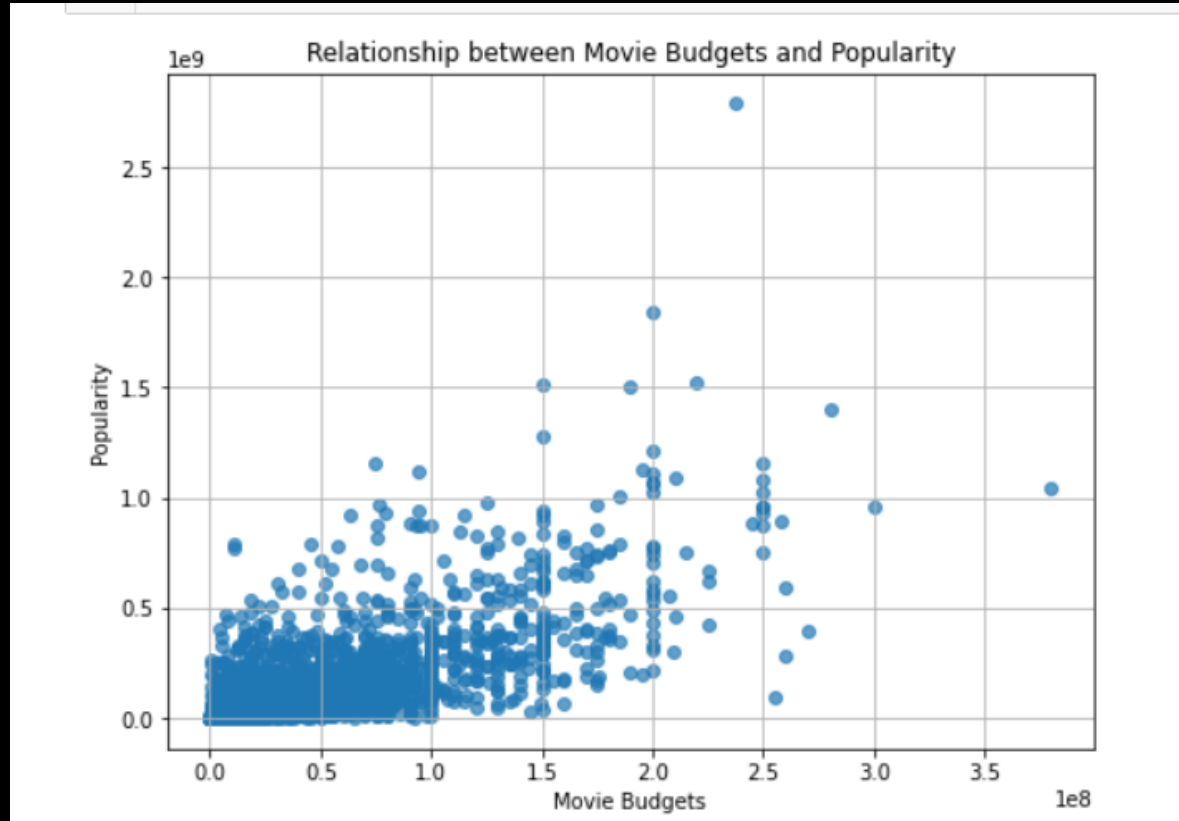
- Correlation between Budget and Popularity: 0.7053993306781382

interpretation

POSITIVE CORRELATION: HIGHER BUDGETS OFTEN ALLOW FILMMAKERS TO INVEST IN BETTER PRODUCTION VALUES, SPECIAL EFFECTS, STAR ACTORS, AND EXTENSIVE MARKETING CAMPAIGNS.

THESE ELEMENTS CAN ENHANCE THE OVERALL QUALITY AND PRESENTATION OF THE MOVIE, ATTRACTING A LARGER AUDIENCE AND POTENTIALLY LEADING TO HIGHER POPULARITY.
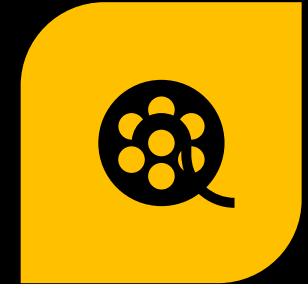
QUALITY MATTERS: REGARDLESS OF BUDGET, THE QUALITY OF THE MOVIE PLAYS A VITAL ROLE IN DETERMINING ITS POPULARITY. A WELL-MADE FILM, WITH A COMPELLING STORY AND STRONG PERFORMANCES, CAN GAIN POPULARITY THROUGH POSITIVE WORD-OF-MOUTH AND CRITICAL ACCLAIM, EVEN WITH A MODERATE BUDGET.

To display production companies along with the number of times they have appeared in the datase

```
1   production_counts
```

```
351
Paramount Pictures
58
Universal Pictures
45
New Line Cinema
38
Columbia Pictures
37

...
New Line Cinema, Juno Pix, Cecchi Gori Pictures
1
WingNut Films, TriStar Pictures, Key Creatives, Block / Hanson, Canadian Film or Video Production Tax Credit (CPTC), The Depart
ment of Trade and Industry of South Africa, QED International, Province of British Columbia Production Services Tax Credit, Dis
trict 9, New Zealand Post Digital and Visual Effects Grant        1
Paramount Pictures, Nickelodeon Movies, United Plankton Pictures, Paramount Animation
1
Village Roadshow Pictures, Malpaso Productions, NPV Entertainment, Warner Bros.
1
rusty bear entertainment, lucky crow films
1
Name: production_companies, Length: 3697, dtype: int64
```

To display the names of the top 25 production companies based on the number of movies they have produced, you can use the previously calculated sort it in descending order.

```
                                                                 351
Paramount Pictures                                                58
Universal Pictures                                                45
New Line Cinema                                                   38
Columbia Pictures                                                 37
Metro-Goldwyn-Mayer (MGM)                                         32
Twentieth Century Fox Film Corporation                            31
Warner Bros.                                                      27
Walt Disney Pictures                                              27
Touchstone Pictures                                               26
Dimension Films                                                   17
Columbia Pictures Corporation                                     16
Miramax Films                                                     16
DreamWorks Animation                                              12
United Artists                                                    12
Walt Disney Pictures, Pixar Animation Studios                     11
Fox 2000 Pictures                                                 10
Fox Searchlight Pictures                                           9
Imagine Entertainment, Universal Pictures                          9
Walt Disney Pictures, Walt Disney Feature Animation                9
Marvel Studios                                                     8
Blue Sky Studios, Twentieth Century Fox Animation                  8
Lions Gate Films                                                   8
Hollywood Pictures, Cinergi Pictures Entertainment                 7
United Artists, Eon Productions, Danjaq                            7
Name: production_companies, dtype: int64
```

Top 25 Production Companies by Number of Movies Produced
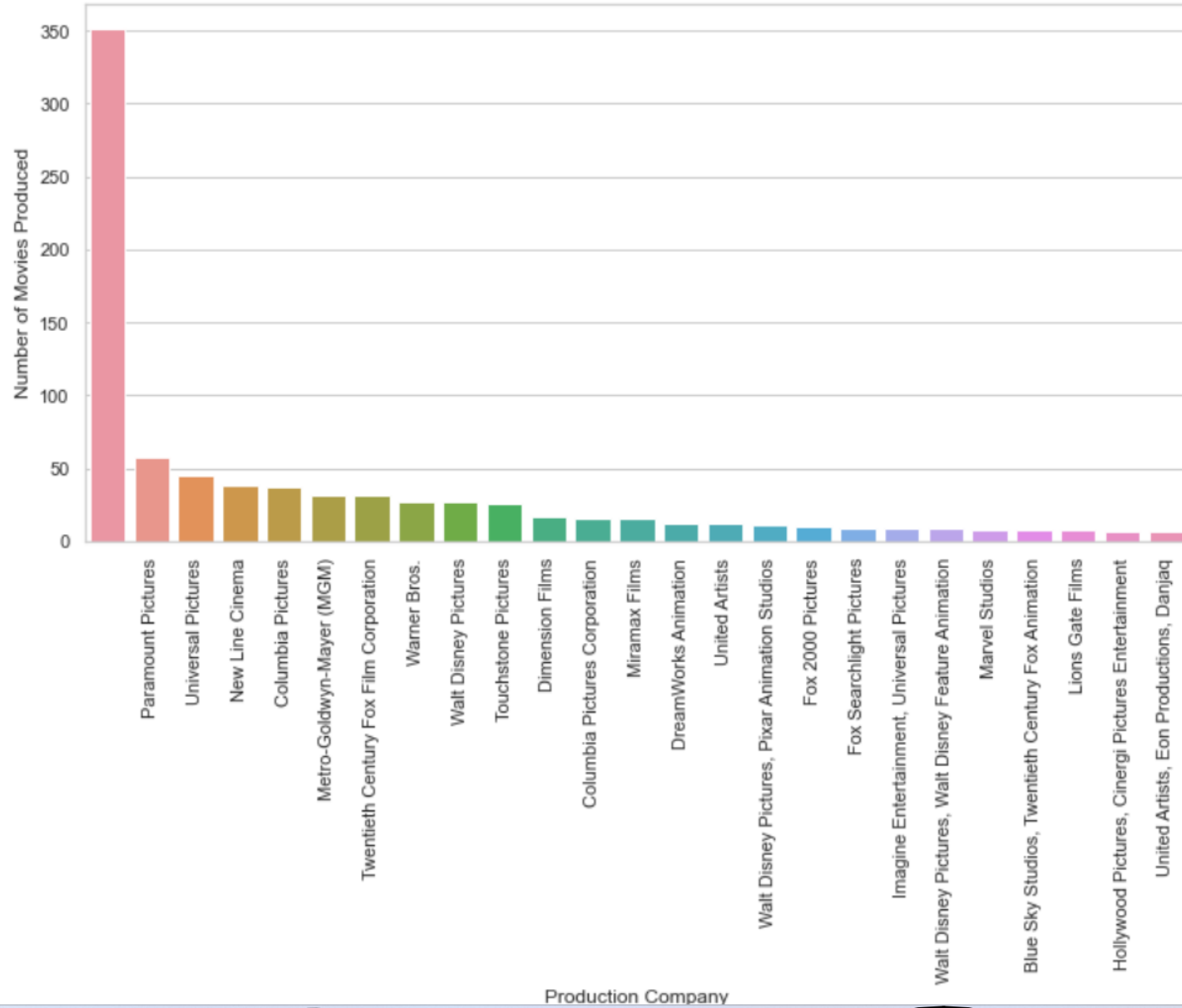
interpretation

Paramount Pictures and **Universal Pictures** stand out as the two most prolific production companies, having produced 58 and 45 movies, respectively. This suggests that both companies have been actively involved in the film industry and have contributed significantly to the production of a wide range of movies.

Diversified Portfolios: Several production companies, such as **New Line Cinema**, **Columbia Pictures**, **Metro-Goldwyn-Mayer (MGM)**, and **Twentieth Century Fox Film Corporation**, have produced a substantial number of movies (ranging from 31 to 38). This indicates that these companies maintain diversified portfolios and are involved in various film genres and projects.

Big Players in Animation: Walt Disney Pictures, **DreamWorks Animation**, and **Pixar Animation Studios** (a subsidiary of Walt Disney Pictures) are prominent players in the animation industry. Their

Outlier analysis for three columns
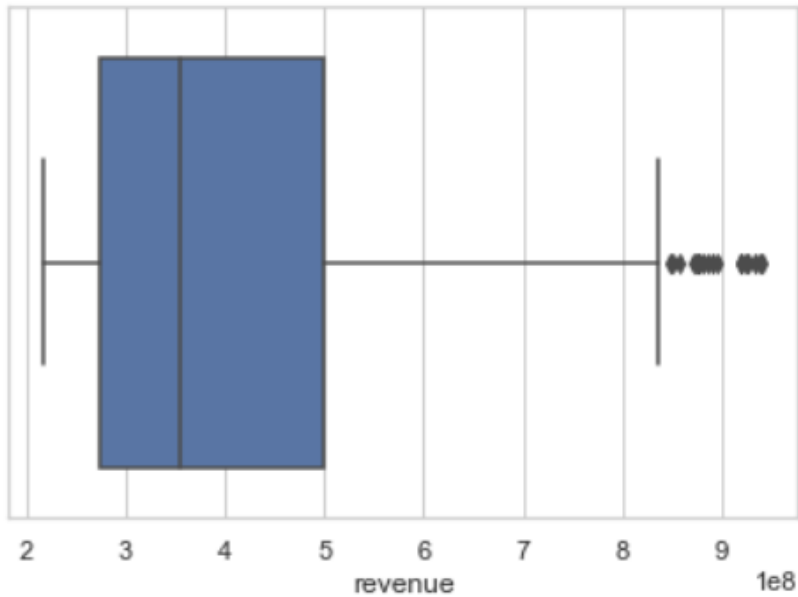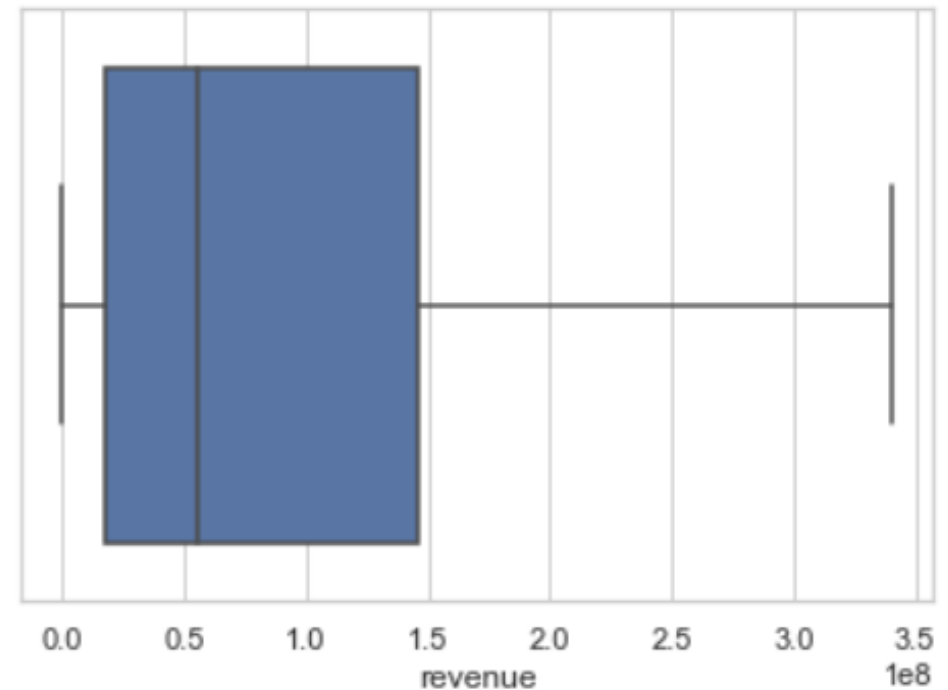
Before and after outlier treatment of budget column
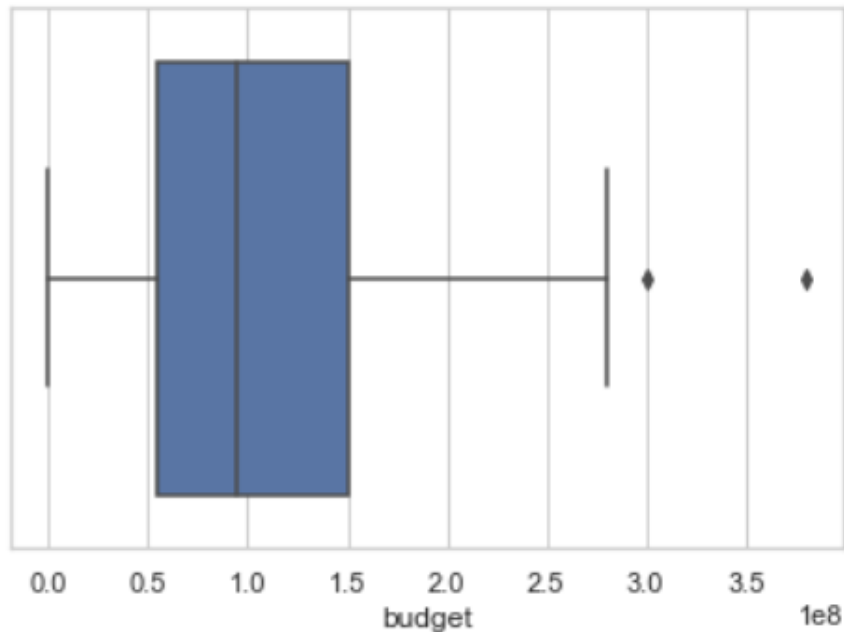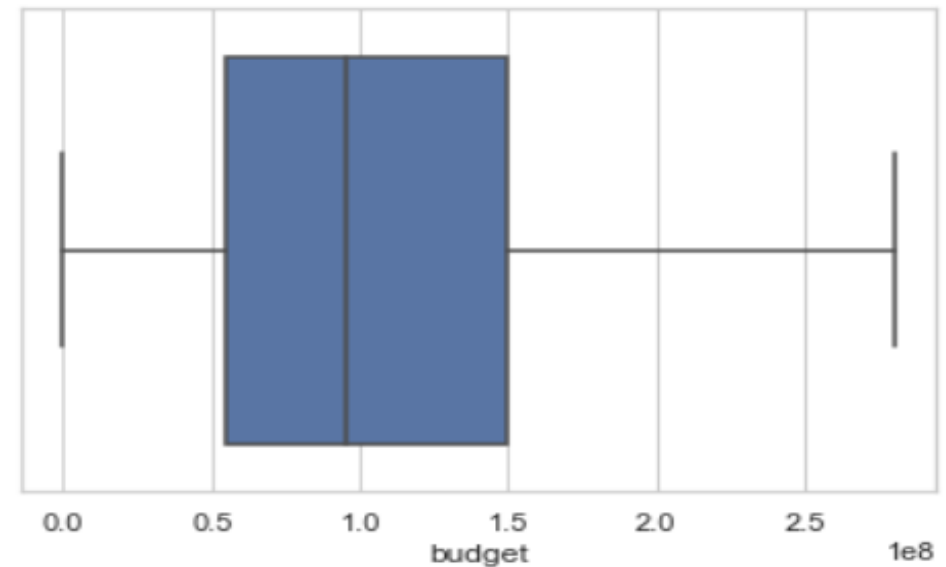
Before and after outlier treatment of runtime column
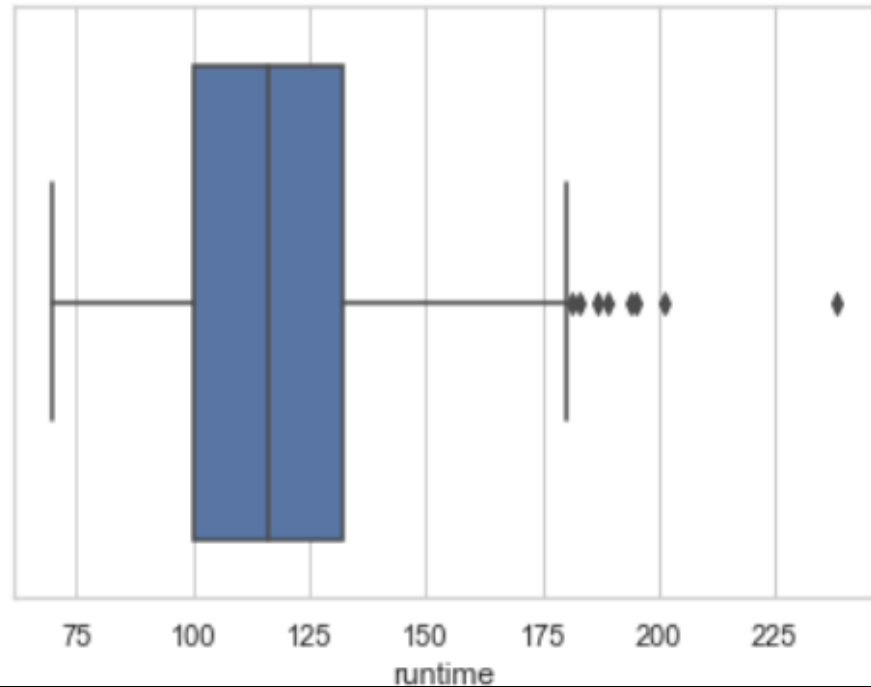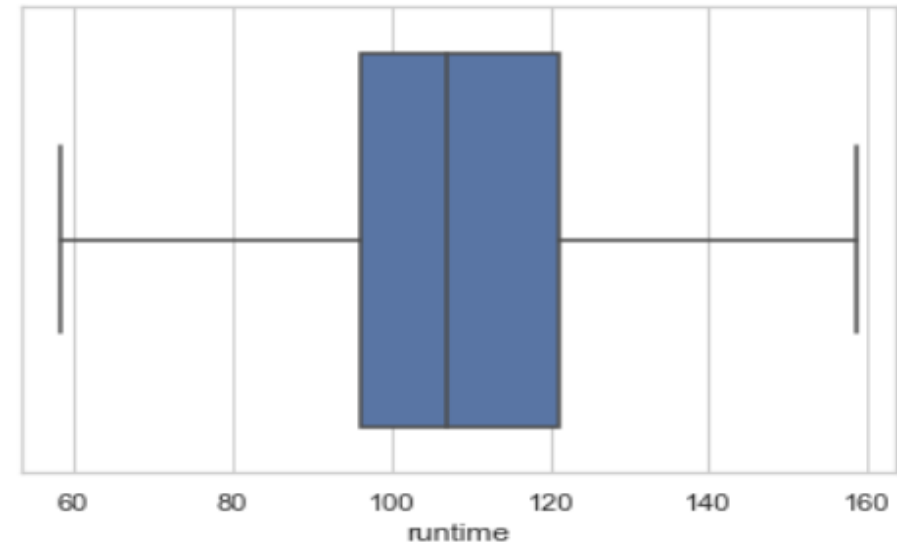
THANK YOU

ANY QUESTIONS

- Feel free to ask