A Priceless Film: How budgeting and other aspects play into the success of a movie

Rishi Nair, Dongyeon Kang, Nathan Kizlinski, Abdulla Alnuaimi

Our Data

- Top 1000 Movies by IMDB Rating
    - 13 columns (Runtime, Genre, IMDb rating, Director, etc.)
    - 1000 entries
- Top 4000 Movies by Production Cost (Axel Torbenson)
    - 5 columns (Release Date, Movie Title, Production Budget, Domestic Gross, Worldwide Gross)
    - 4000 entries
- Using Python for EDA and data cleaning, using SQL for dataset integration and query implementation

Queries

- "Do higher budget costs always result in higher ratings?"
- "How does runtime affect gross?"
- "Which directors have the greatest box office success?"
- "Are certain genres more successful than others?"
- "Are movies getting better reviews over time?"

Methodology

- The initial data can tell us a lot, but the data isn't cleaned- we needed to run multiple functions to clean our dataset, such as: .dropna(), .replace(), .split(), and more

- There were missing values, unimportant columns, duplicates, and mismatch issues that needed handling

- Once our datasets were cleaned, we were ready to merge

Methodology

○ After merging our two cleaned datasets, the resulting dataset looks as follows:

	title	Released_Year	Runtime	Genre	IMDB_Rating	Director	No_of_Votes	ProductionBudget	DomesticGross	WorldwideGross
1	title									
2	12 years a slave	2013	134	Biography	8.1	Steve McQueen	640533	20.0	56671993	180765061
3	1917	2019	119	Drama	8.3	Sam Mendes	425844	100.0	159227644	370603379
4	2001 a space odyssey	1968	149	Adventure	8.3	Stanley Kubrick	603517	10.5	59936321	71854587
5	21 grams	2003	124	Crime	7.6	Alejandro G. Iñárritu	224545	20.0	16248701	59667625
6	300	2006	117	Action	7.6	Zack Snyder	732876	60.0	210614939	454161935
7	310 to yuma	2007	122	Action	7.7	James Mangold	288797	48.0	53606916	71171825
8	a beautiful mind	2001	135	Biography	8.2	Ron Howard	848920	78.0	170708996	317668058
9	a few good men	1992	138	Drama	7.7	Rob Reiner	235388	33.0	141340178	236500000
10	a star is born	2018	136	Drama	7.6	Bradley Cooper	334312	36.0	215288866	431870523
11	about time	2013	123	Comedy	7.8	Richard Curtis	303032	12.0	15323921	88538742
12	aladdin	1992	90	Animation	8.0	Ron Clements	373845	182.0	355559216	1046649706
13	alien	1979	117	Horror	8.4	Ridley Scott	787806	10.7	62005728	184694992
14	aliens	1986	137	Action	8.3	James Cameron	652719	17.0	85160248	183300764
15	almost famous	2000	122	Adventure	7.9	Cameron Crowe	252586	60.0	32522352	47371191

Final dataset has 10 columns and 370 rows (movies) & is ready for SQL implementation

Result: Query 1

"Do higher budget costs always result in higher ratings?"

```
25
26    SELECT
27      CASE
28        WHEN ProductionBudget < 50 THEN 'Low (<50M)'
29        WHEN ProductionBudget BETWEEN 50 AND 100 THEN 'Medium (50-100M)'
30        WHEN ProductionBudget BETWEEN 101 AND 200 THEN 'High (101-200M)'
31        ELSE 'Very High (>200M)'
32      END AS BudgetRange,
33      AVG(IMDB_Rating) AS AvgRating,
34      COUNT(*) AS MovieCount
35    FROM
36      merged
37    GROUP BY
38      BudgetRange
39    ORDER BY
40      BudgetRange;
41
```

	BudgetRange	AvgRating	MovieCount
1	High (101-200M)	7.94074074074074	54
2	Low (<50M)	7.92707423580786	229
3	Medium (50-100M)	7.98552631578948	76
4	Very High (>200M)	8.02727272727273	11

- Not a strong correlation between budget costs and ratings, unless the movie has a budget of over $200 million
- "Low" budget range movies on average only earned ~0.1 less of a rating then other movies
- Note that most movies had a budget of less than $50 million

Result: Query 2

How does runtime affect gross?

```
46    --How does runtime affect gross?
47    SELECT
48      CAST(Runtime AS INTEGER) AS Runtime,
49      CAST(AVG(WorldwideGross) AS INTEGER) AS AvgWorldwideGross,
50      COUNT(*) AS MovieCount
51    FROM
52      merged
53    GROUP BY
54      Runtime
55    ORDER BY
56      Runtime DESC;
```

	Runtime	AvgWorldwideGross	MovieCount
1	242	7129670	1
2	229	5576072	1
3	228	69995047	1
4	220	85433929	1
5	212	91669648	1
6	209	910234	1
7	202	52734955	2
8	201	1120214046	1
9	197	86020194	2
10	194	2207986545	1
11	193	21500213	1
12	191	127767889	1
13	189	248050687	2
14	188	48446802	1
15	183	50057795	1
16	181	1611004684	2
17	180	197207743	2
18	179	919148764	1
19	178	315190517	4
20	172	62500000	1
21	170	71574497	3
22	169	717322832	3
23	168	152366784	1
24	167	263502914	1
25	166	329631958	1

Result: Query 2
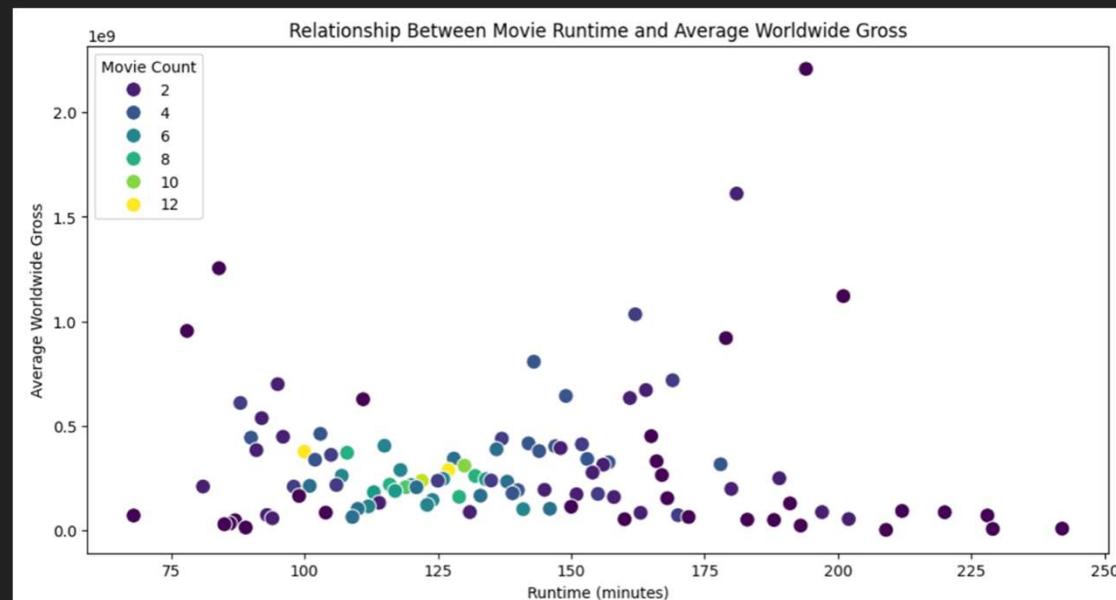
No real correlation between the runtime and worldwide gross, note that most movies are around the 125 min mark (~2 hrs)

Result: Query 3

- As shown, Anthony Russo makes the most on average per movie, grossing over $1.6 billion per movie (Mostly consisting of his Marvel work)

- Note that Roger Allers has one of the highest grossing movies in our dataset with about $1.65 billion from The Lion King

- Peter Jackson, producer of the Lord of the Rings movie series, made about $980 million per movie

Which directors have the greatest box office success?

```
56
57   SELECT
58       Director,
59       CAST(AVG(WorldwideGross) AS INTEGER) AS AvgWorldw
60       COUNT(*) AS MovieCount
61   FROM
62       merged
63   GROUP BY
64       Director
65   ORDER BY
66       AvgWorldwideGross DESC;
67
```

	Director	AvgWorldwideGross	MovieCount
1	Anthony Russo	1677165374	4
2	Roger Allers	1654367425	1
3	James Cameron	1438144433	4
4	Gary Trousdale	1253802559	1
5	Josh Cooley	1073080329	1
6	Byron Howard	1004629935	1
7	Peter Jackson	980975323	5
8	Wolfgang Reitherman	953554418	1
9	Andrew Stanton	936094852	1
10	Lee Unkrich	933272973	2
11	David Yates	929411069	1
12	Ron Clements	840807633	2
13	Bryan Singer	817451692	2
14	David Leitch	786362370	1
15	Tim Miller	784157111	1
16	Joss Whedon	777709825	2
17	James Gunn	770882395	1
18	Todd Phillips	768997550	2
19	Pete Docter	715196054	3

Result: Query 4

Are certain genres more successful than others?

```
70   -- Are certain genres more successful than others?
71   SELECT
72     Genre,
73     CAST(AVG(WorldwideGross) AS INTEGER) AS AvgWorldwideGross,
74     COUNT(*) AS MovieCount
75   FROM
76     merged
77   GROUP BY
78     Genre
79   --removed genres with total movies less than 10
80   HAVING
81     COUNT(*) >= 10
82   ORDER BY
83     AvgWorldwideGross DESC;
```

	Genre	AvgWorldwideGross	MovieCount
1	Animation	601992602	36
2	Action	455324738	89
3	Adventure	302942595	36
4	Drama	187550416	82
5	Comedy	157341870	29
6	Crime	152237026	35
7	Biography	141899237	54

- Animation movies have the highest success, making on average $600 million
- Action movies are the most popular to produce, which makes sense if they average $455 million

Result: Query 5

Are movies getting better reviews over time?

```
92    --Grouped the movies into 5-year intevals
93    SELECT
94       (Released_Year / 5) * 5 AS YearStart,
95       ((Released_Year / 5) * 5) + 4 AS YearEnd,
96       ROUND(AVG(IMDB_Rating), 2) AS AvgIMDBRating,
97       COUNT(*) AS MovieCount
98    FROM
99       merged
100   WHERE
101      Released_Year IS NOT NULL
102      AND Released_Year >= 1920
103   GROUP BY
104      (Released_Year / 5) * 5
105   --removed the rows with moviecount less than 3
106   HAVING
107      MovieCount >= 3
108   ORDER BY
109      YearStart;
```


Average IMDB Rating Over Time
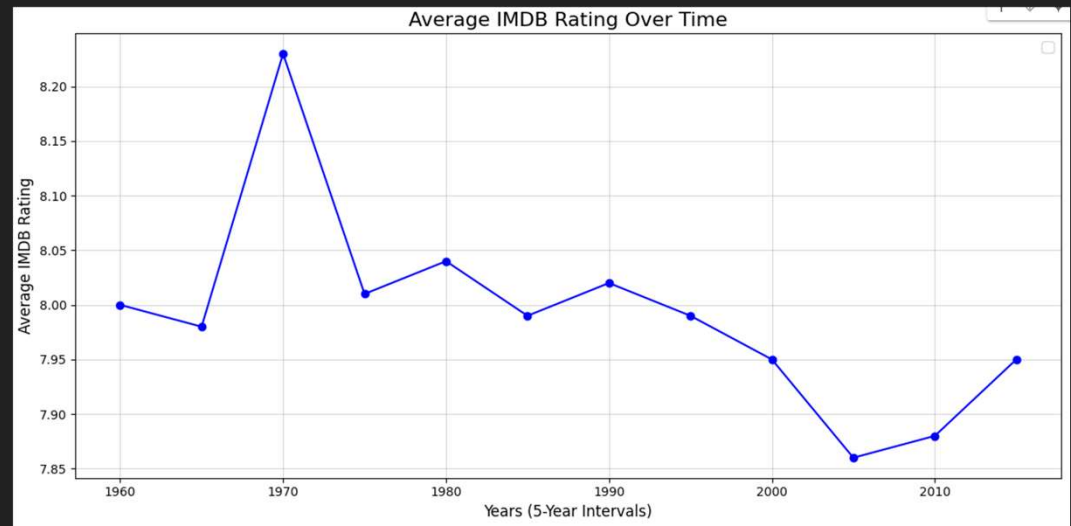
- Noticeable decline in average ratings after a peak in the 1970s
- Assume that people were being more critical about movies as time goes on

Analysis

- Do higher budget costs result in higher ratings?
    - High budgets achieved a lower avg rating than medium budgets
    - Higher budget costs do not always result in higher ratings.
- How does runtime affect gross?
    - There is no correlation between runtime and worldwide gross, indicating that movie success depends more on content, audience engagement, and external factors like marketing.
    - Most movies are around the 125 min mark (~2 hrs), indicating that this is the ideal runtime for audiences.
- Which directors have the greatest box office success?
    - Directors who direct blockbuster franchises tend to dominate the box office. Some of these franchises are Anthony Russo's *Avengers: Infinity War* and *Endgame,* James Cameron's *Avatar* and *Titanic,* Peter Jackson's *Lord of the Rings* and *The Hobbit* trilogies.
    - One hit wonders like Roger Allers (The Lion King) who achieved high rankings through a single iconic movie.

Analysis (cont'd)

Average Worldwide Gross by Genre
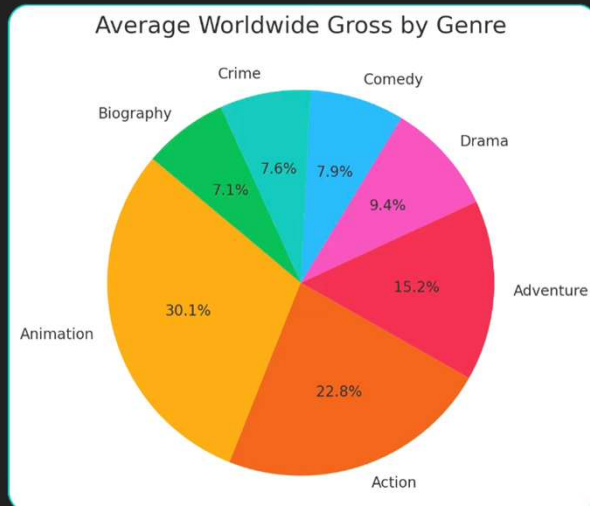
- ○ Are certain genres more successful than others?
  - o Animation dominates in success which can be attributed to factors like family appeal, attracting all age groups
  - o Action/Adventure is popular to produce and highly successful due to its entertainment value like stunts, visual effects and intense story lines.
- ○ Are movies getting better reviews over time? -
  - o Movie ratings peaked in the 1970s due to innovative filmmaking.
  - o Declined afterwards but began to recover slightly in the 2010s, driven by streaming platforms.

Conclusion

- Movie success is influenced by a combination of story quality, genre, global appeal, and director influence rather than just budget or runtime.

Future Work

- Several areas for further exploration:
    - the impact of streaming platforms on modern movies
    - incorporating audience demographics could provide a better understanding of how different groups influence the success of various genres.
    - Exploring the role of marketing budgets and their correlation with box office performance.