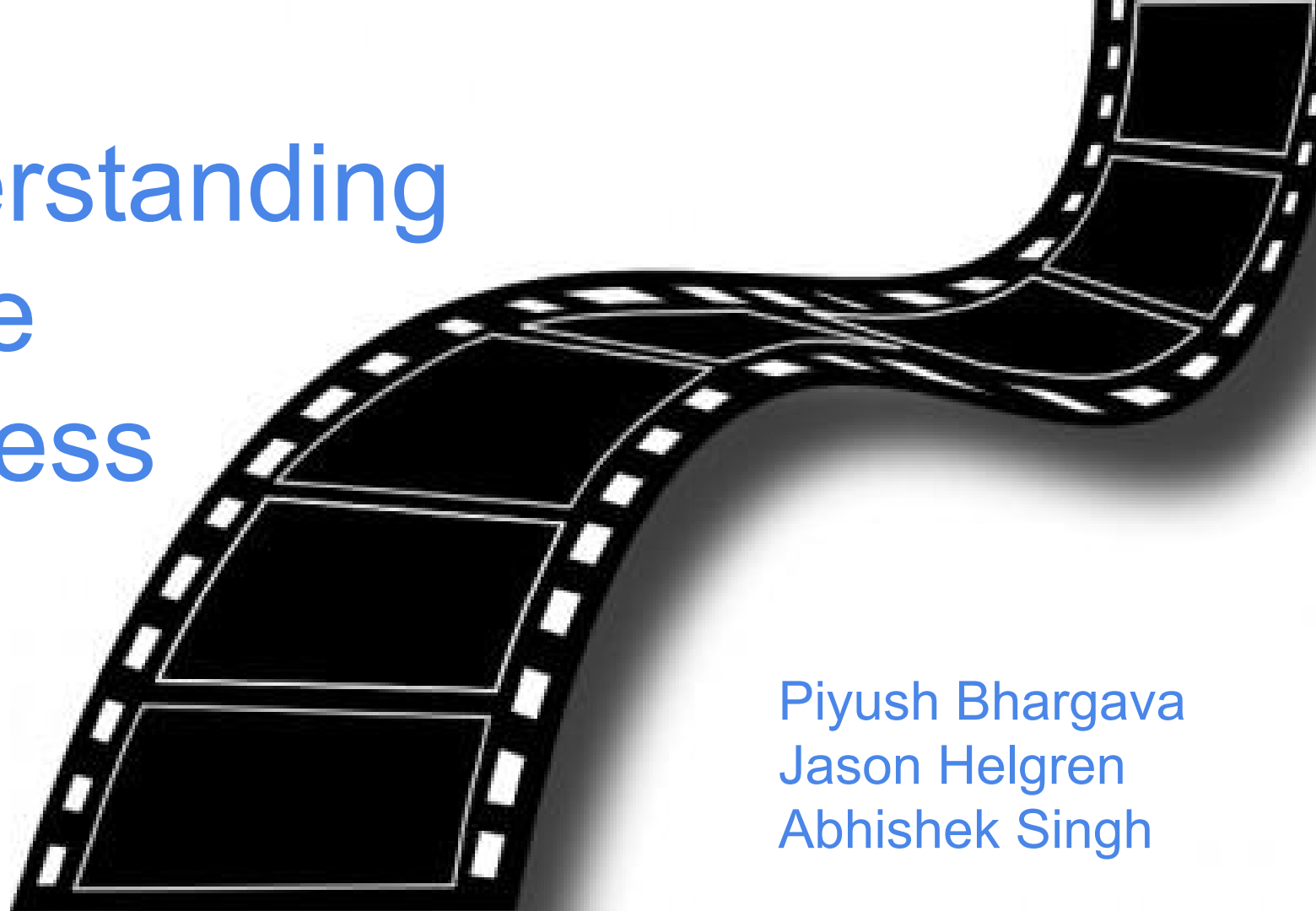


Understanding Movie Success



Piyush Bhargava
Jason Helgren
Abhishek Singh

The Questions

How do we calibrate a movie's success? What are the indicators?

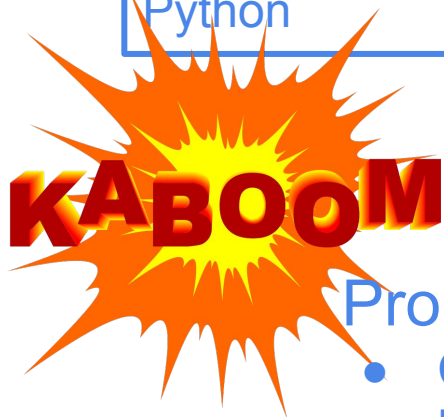
How do other variables correlate with these indicators?

Other interesting insights?

Methodology



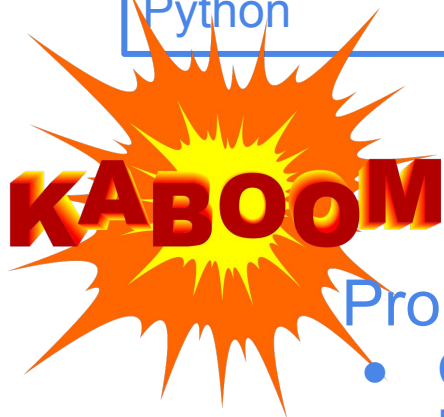
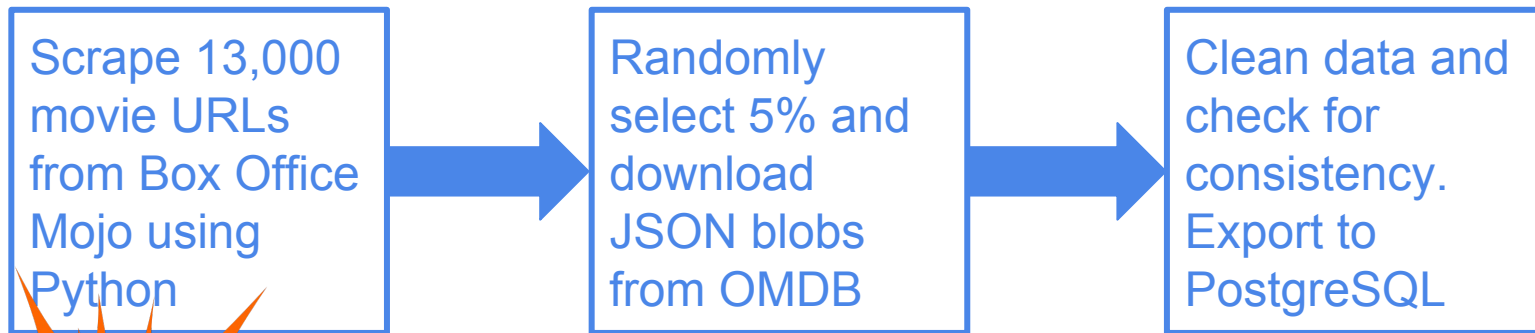
Methodology



Problems along the way:

- Couldn't get an API key
- Lots of missing values
- Not all URLs had JSON files

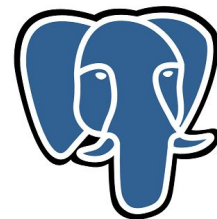
Methodology



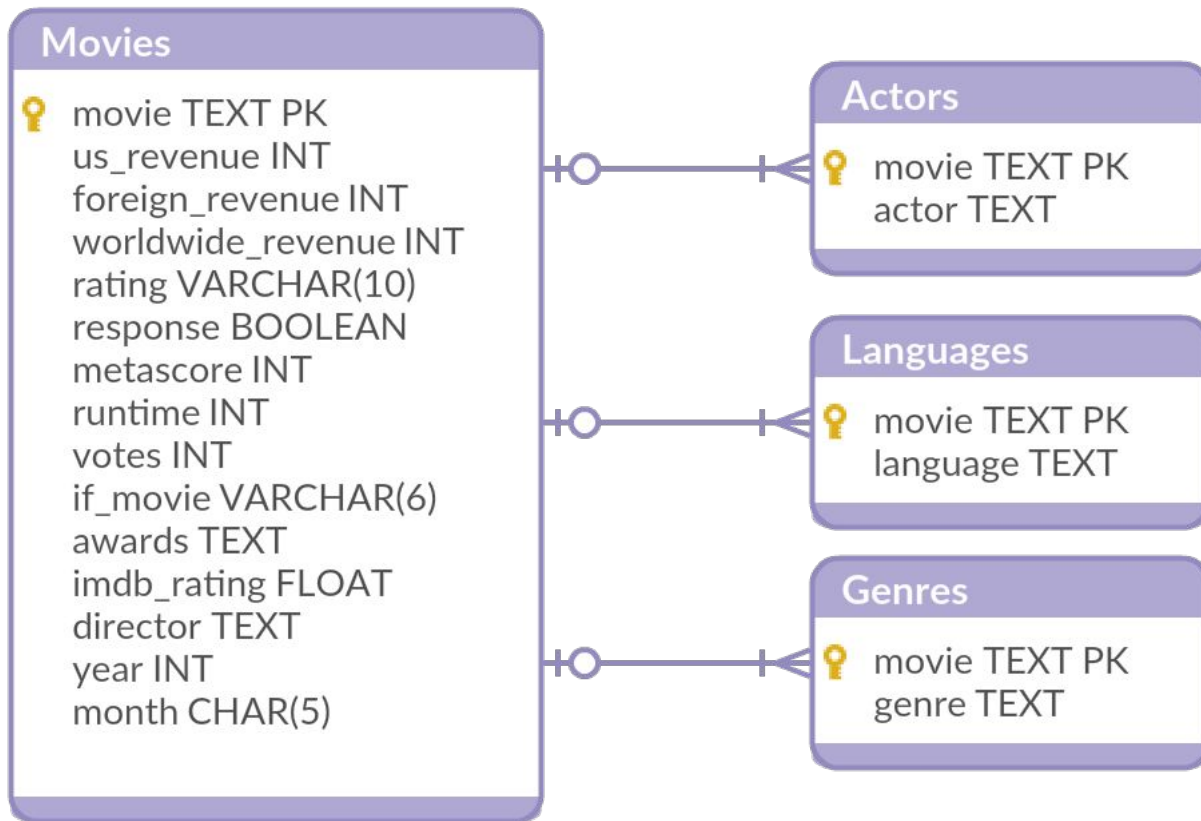
Problems along the way:

- Couldn't get an API key
- Lots of missing values
- Not all URLs had JSON files

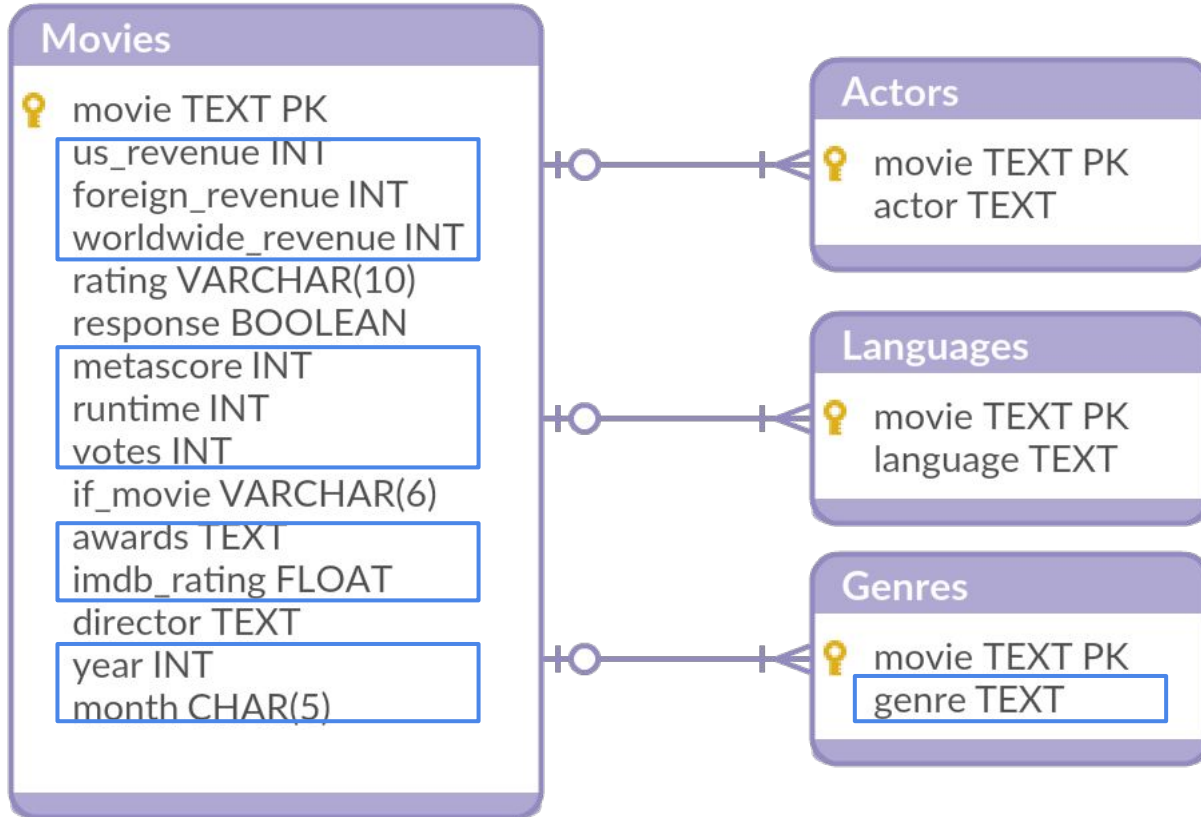
Final Result:
database of
about 500
movies



ER Diagram



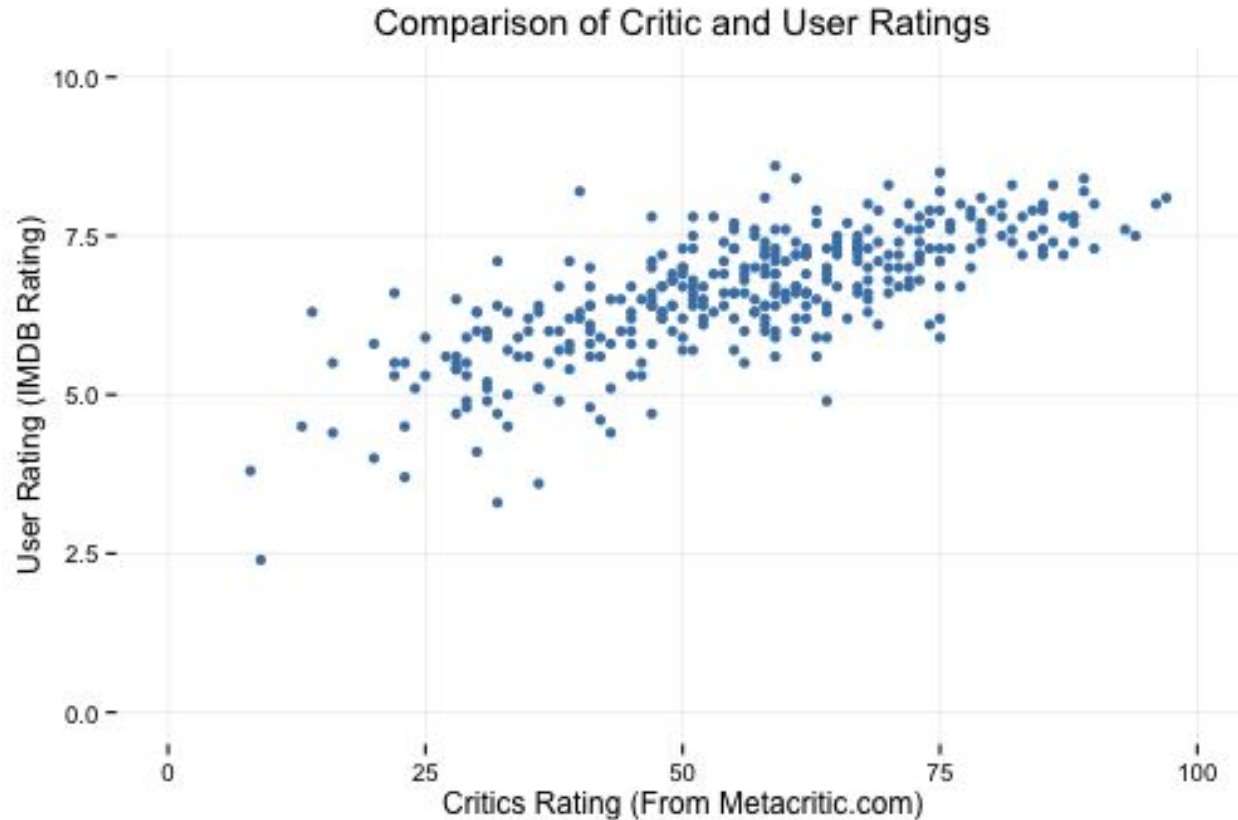
ER Diagram



Sample Query: User vs Critic Rating

```
SELECT AVG(metascore) AS critic_avg,  
  
STDDEV_POP(metascore) AS critic_std_dev,  
  
AVG(imdb_rating) AS imdb_avg,  
  
STDDEV_POP(imdb_rating) AS imdb_std_dev,  
  
CORR(metascore, imdb_rating) AS correlation  
  
FROM movies;
```


User vs Critic Rating



Sample query: Summer blockbuster effect?

```
SELECT sum(1) AS count, month, genre FROM

(SELECT lhs.movie, lhs.month, rhs.genre FROM

(SELECT movie, month FROM movies) AS lhs

LEFT JOIN

(SELECT movie, genre from genre) AS rhs

USING(movie)) AS iq WHERE genre

IN ('Action', 'Comedy', 'Drama', 'Romance', 'Adventure', 'Crime')

GROUP BY month, genre ORDER BY month;
```

Sample Query: Summer Blockbuster Effect?

```
SELECT sum(1) AS count, month, genre FROM
```

```
(SELECT lhs.movie, lhs.month, rhs.genre FROM
```

```
(SELECT movie, month FROM movies) AS lhs
```

```
LEFT JOIN
```

```
(SELECT movie, genre FROM genres) AS rhs
```

```
USING(movie)) AS iq WHERE genre
```

```
IN ('Action', 'Comedy', 'Drama', 'Romance', 'Adventure', 'Crime')
```

```
GROUP BY month, genre ORDER BY month;
```

Sample query: Summer blockbuster effect?

2

```
SELECT sum(1) AS count, month, genre FROM
```

```
(SELECT lhs.movie, lhs.month, rhs.genre FROM
```

```
(SELECT movie, month FROM movies) AS lhs
```

1

```
LEFT JOIN
```

```
(SELECT movie, genre FROM genres) AS rhs
```

```
USING(movie)) AS iq WHERE genre
```

```
IN ('Action', 'Comedy', 'Drama', 'Romance', 'Adventure', 'Crime')
```

```
GROUP BY month, genre ORDER BY month;
```

Sample query: Summer blockbuster effect?

2 `SELECT sum(1) AS count, month, genre FROM`

`(SELECT lhs.movie, lhs.month, rhs.genre FROM`

`(SELECT movie, month FROM movies) AS lhs`

1 `LEFT JOIN`

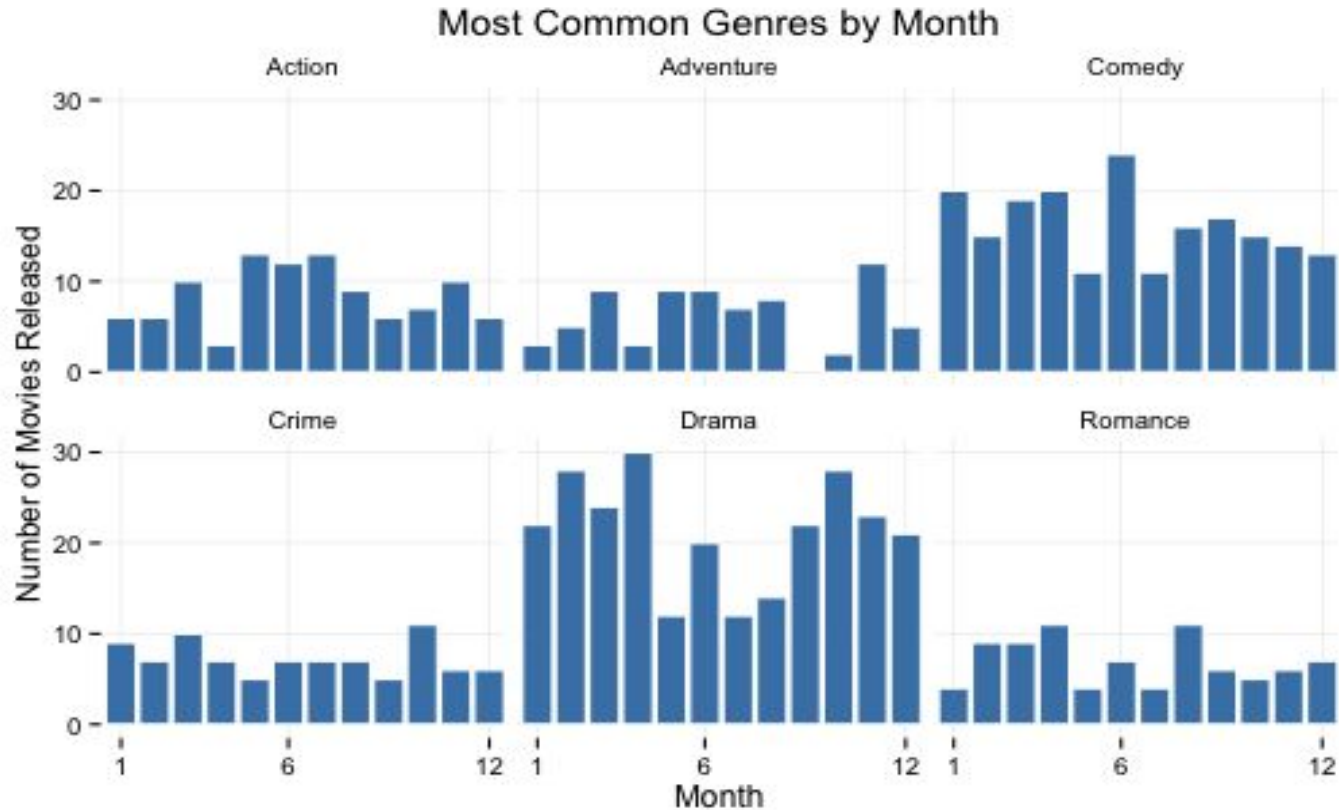
`(SELECT movie, genre FROM genres) AS rhs`

`USING(movie)) AS iq WHERE genre`

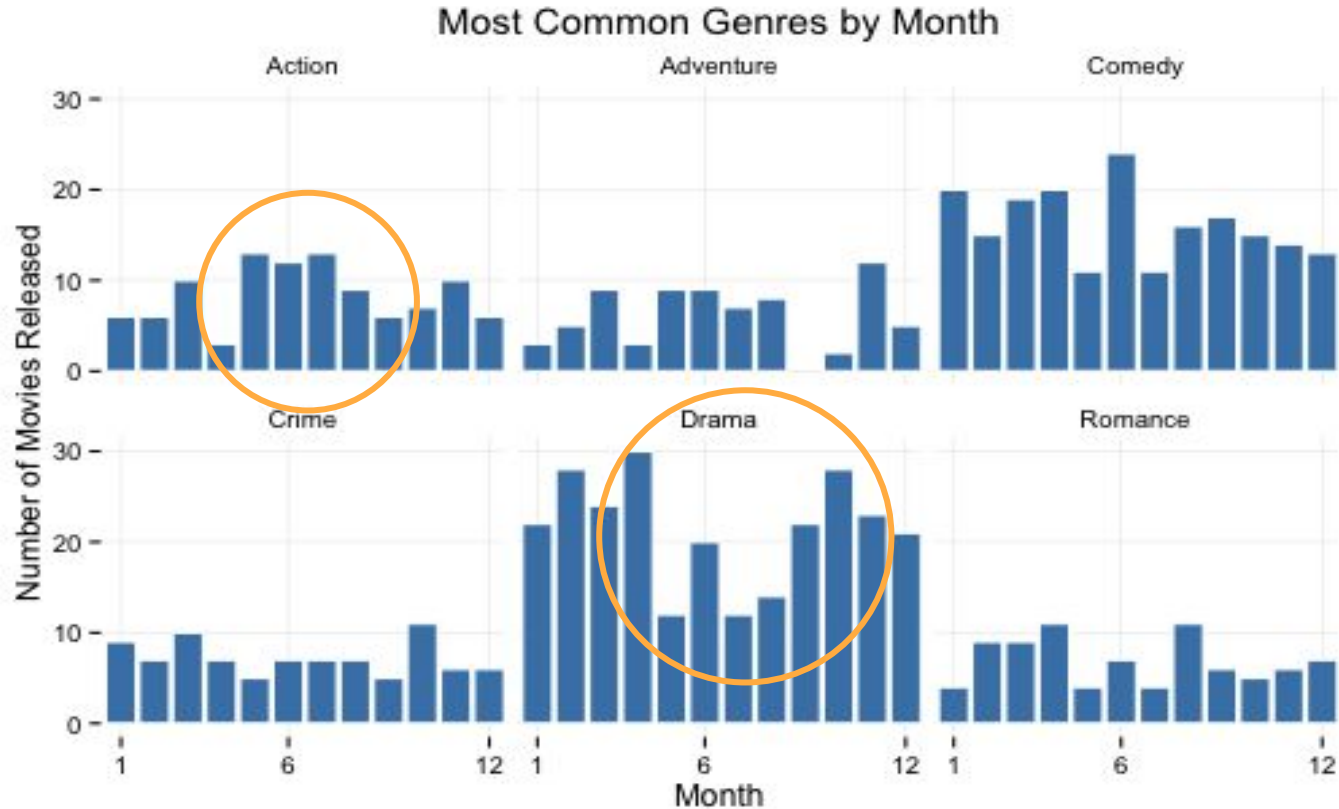
3 `IN ('Action', 'Comedy', 'Drama', 'Romance', 'Adventure', 'Crime')`

`GROUP BY month, genre ORDER BY month;`

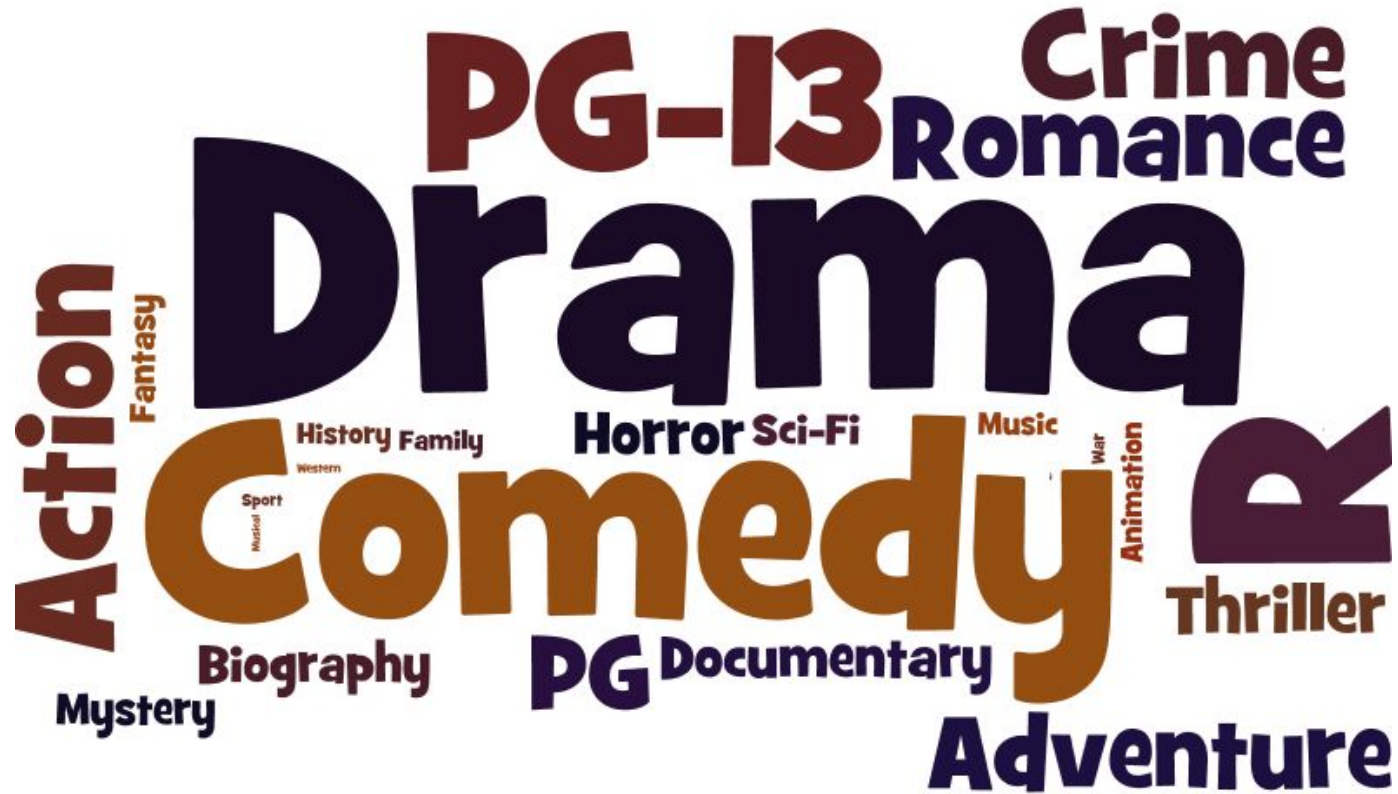
Summer Blockbuster Effect



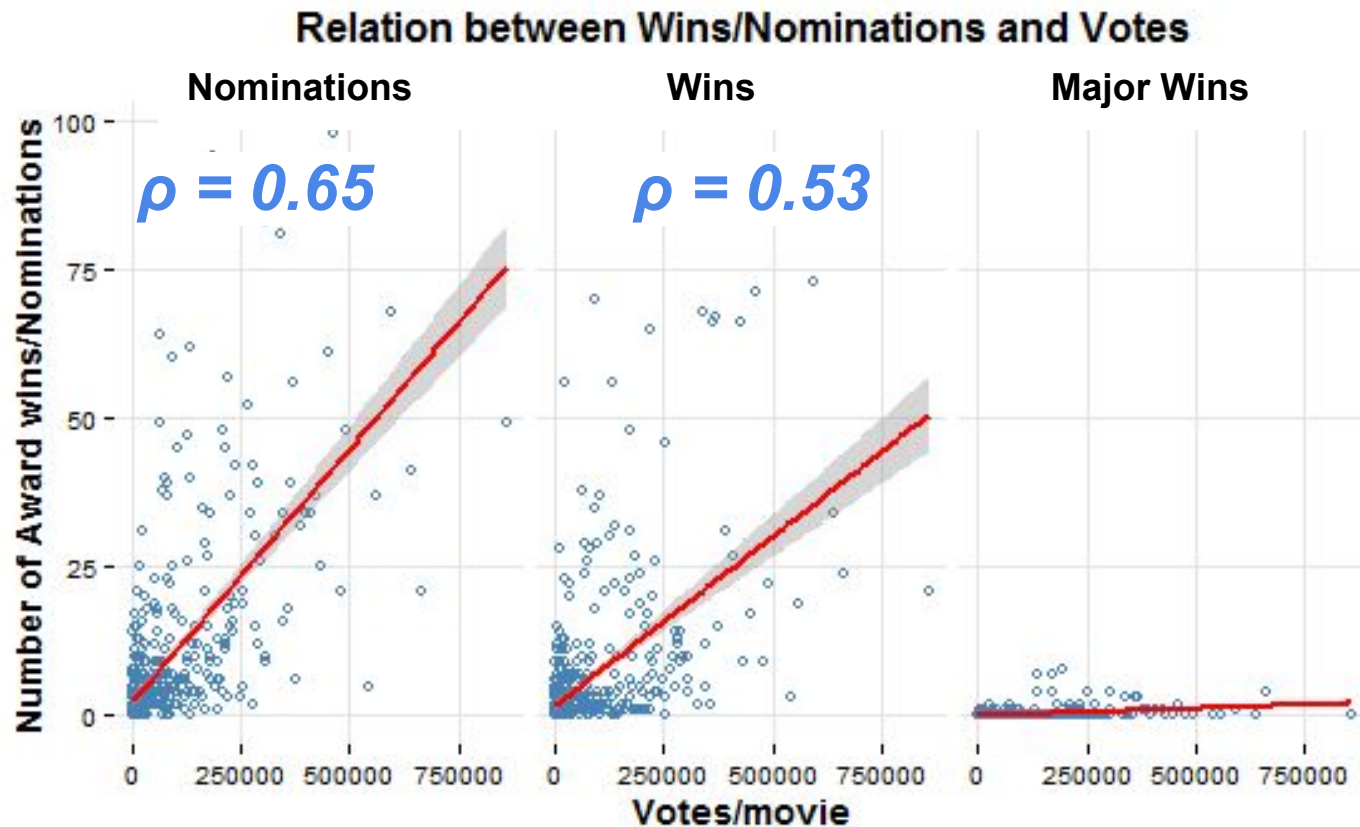
Summer Blockbuster Effect



Most Common Genres and MPAA Rating

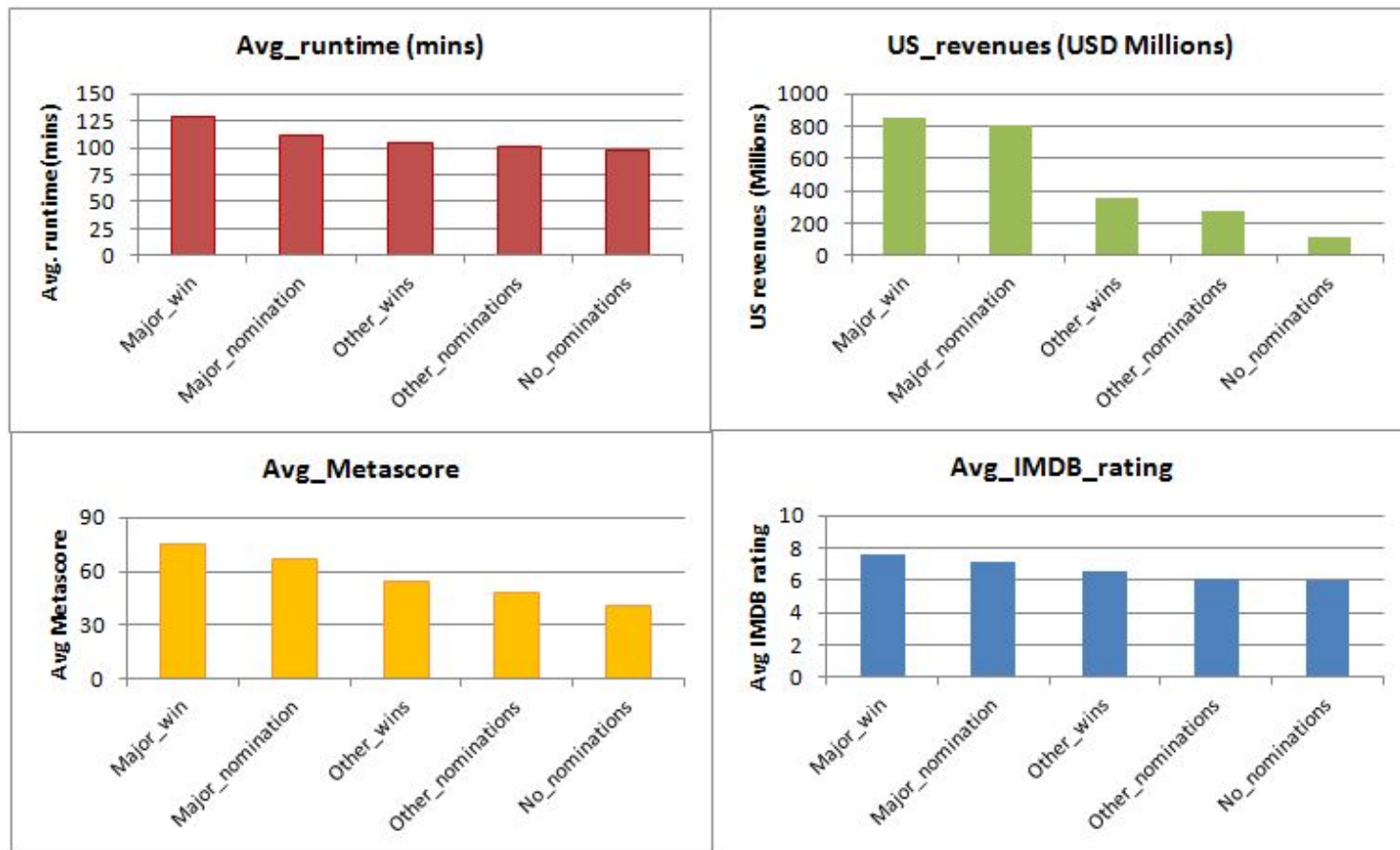


Awards/Nominations Versus Votes



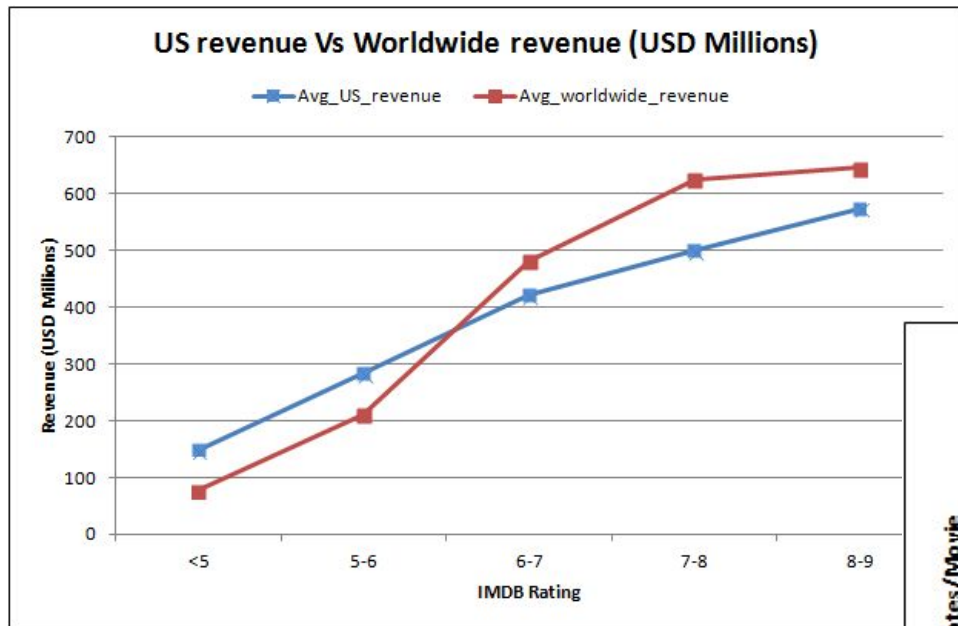
*Major Win - Oscar/BAFTA/Golden Globe Wins

Awards/Nominations vs other variables

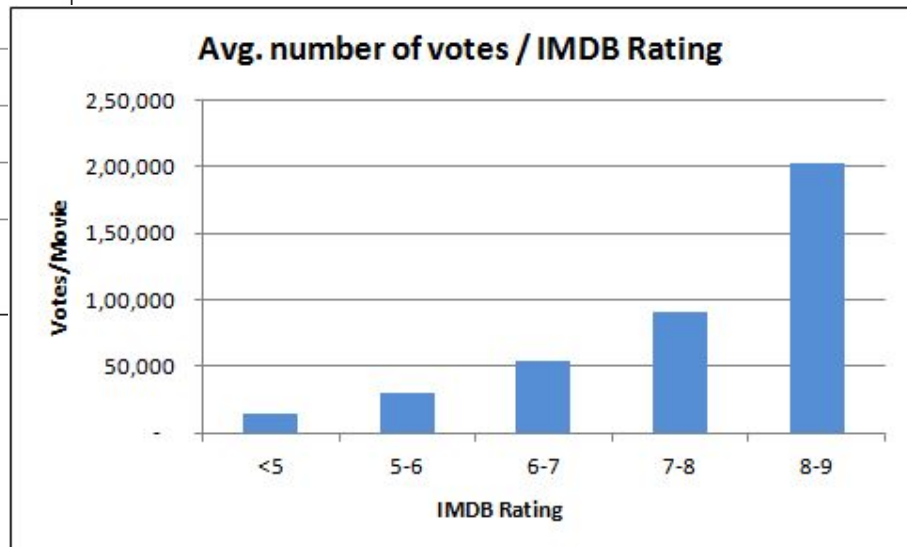


*Major Win/Nomination - Oscar/BAFTA/Golden Globe Wins/Nominations

Revenue Vs IMDB Rating



- Higher rated movies earn more worldwide than in US.
- Worldwide movies are potentially influenced by IMDB rating/number of votes.



Takeaways and Next Steps

- Scrape more data to get a normal distribution!
- Build a model to predict movie success.

