Building a Data Story

MidTerm Project

Movies Dataset

15 Nov 2022

Table of Contents

15 Nov 2022

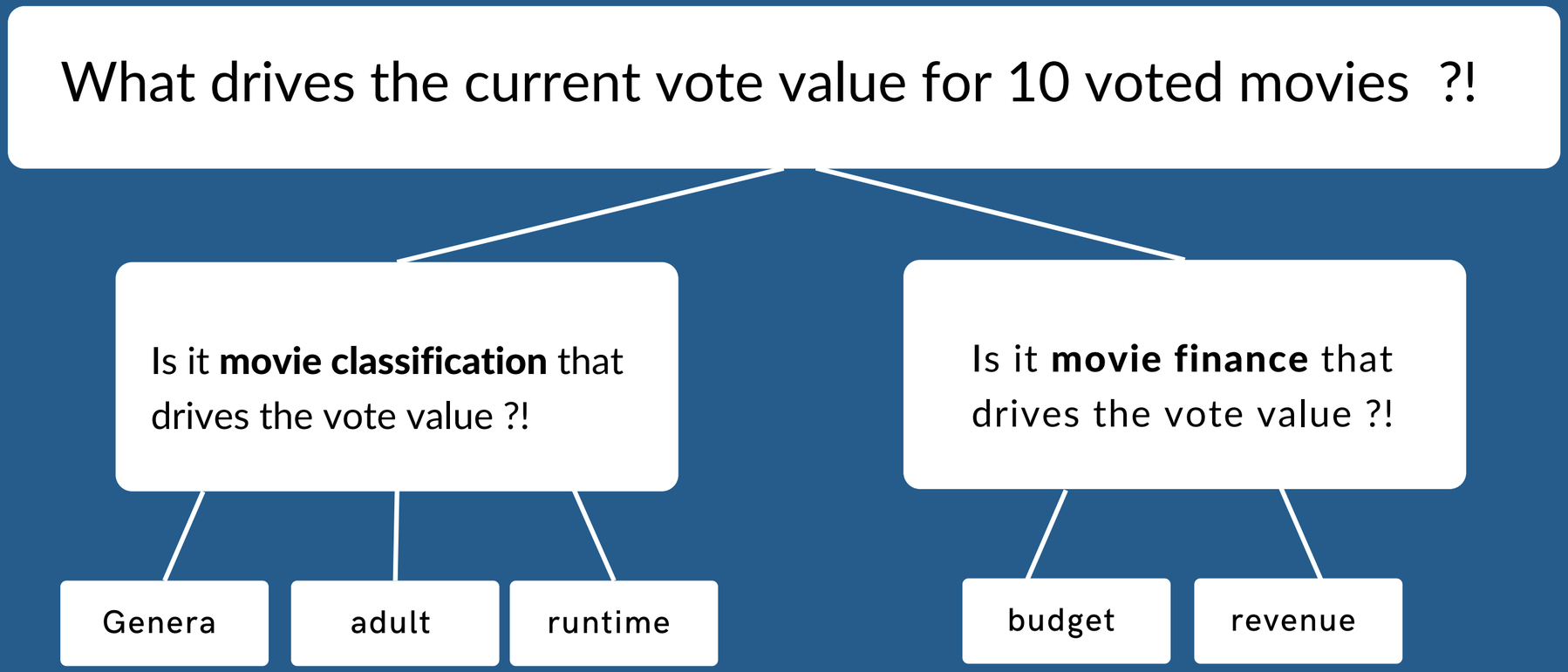
Executive Summary

The problem Statement was determining:

What drives the current vote value for 10 voted movies ?!

Our recommendation is produce movies with the highest potential

Issue Tree


What drives the current vote value for 10 voted movies ?!
Is it **movie classification** that drives the vote value ?!
Is it **movie finance** that drives the vote value ?!
Genera
adult
runtime
budget
revenue

Overview of Analysis

The problem Statement is:

What drives the current vote value for 10 voted movies ?!

My analysis focuses on drilling into two work streams:

- Is it movie classification (Genera/Adult/Runtime) that drives the vote value ?!
- Is it movie finance (Budget/Revenue) that drives the vote value ?!

10 Voted Movies

index	Title	
1	'Rameau's Nephew' by Diderot (Thanx to Denni..	10
2	A Kiss at Midnight	10
3	A Story of the Forest: Mavka	10
4	A Ticklish Affair	10
5	Aashiq	10
6	Acéphale	10
7	Almost Kings	10
8	American Hostage	10
9	American Sharia	10
10	An Apology to Elephants	10
11	Andrew Jenks, Room 335	10
12	At Ellen's Age	10
13	Avetik	10
14	Back to School with Franklin	10
15	Backyard Dogs	10
16	Bazodee	10
17	Bella Vita	10
18	Big Jay Oakerson: Live at Webster Hall	10
19	Birch Interval	10
20	Blessed Event	10
21	Book of Days	10
22	Bowery Battalion	10
23	Brave Revolutionary	10
24	Butterfly	10
25	Campaign of Hate: Russia and Gay Propaganda	10
26	Canal Zone	10
27	Canned Dreams	10
28	Carmen Miranda: Bananas Is My Business	10
29	Cattle Town	10
30	Children in the Wind	10
31	Chilly Scenes of Winter	10
32	Christopher Titus: Angry Pursuit of Happiness	10
33	Claymation Comedy of Horrors	10
34	Common Threads: Stories from the Quilt	10

Voted10 Movies

Across all years, there are **190** movies that go 10 out of 10 based on the vote average feature.

	# movies
Top Movies	190
Others	45,273

15 Nov 2022

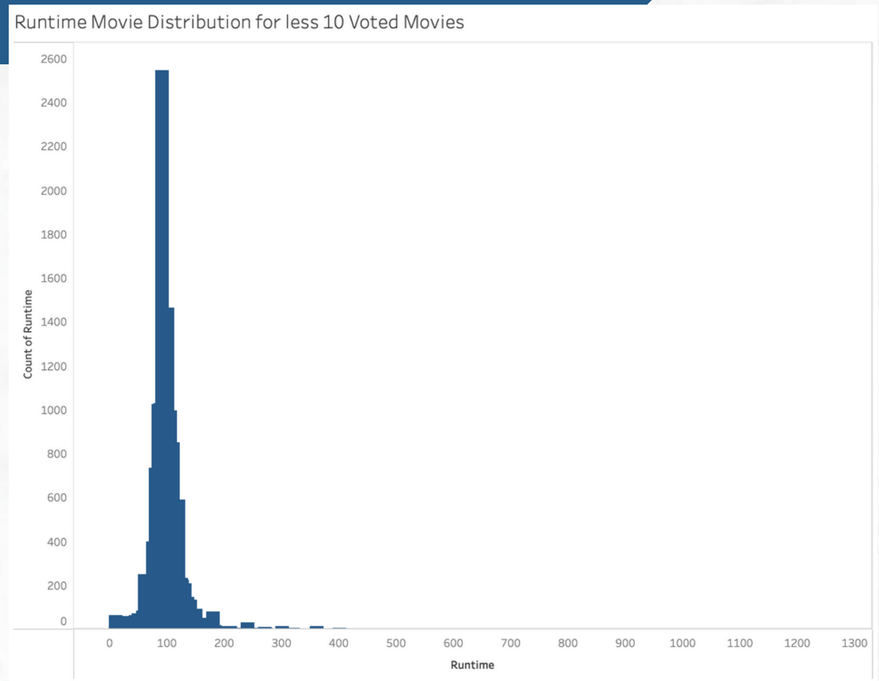
All 10 voted movies aren't adult movies

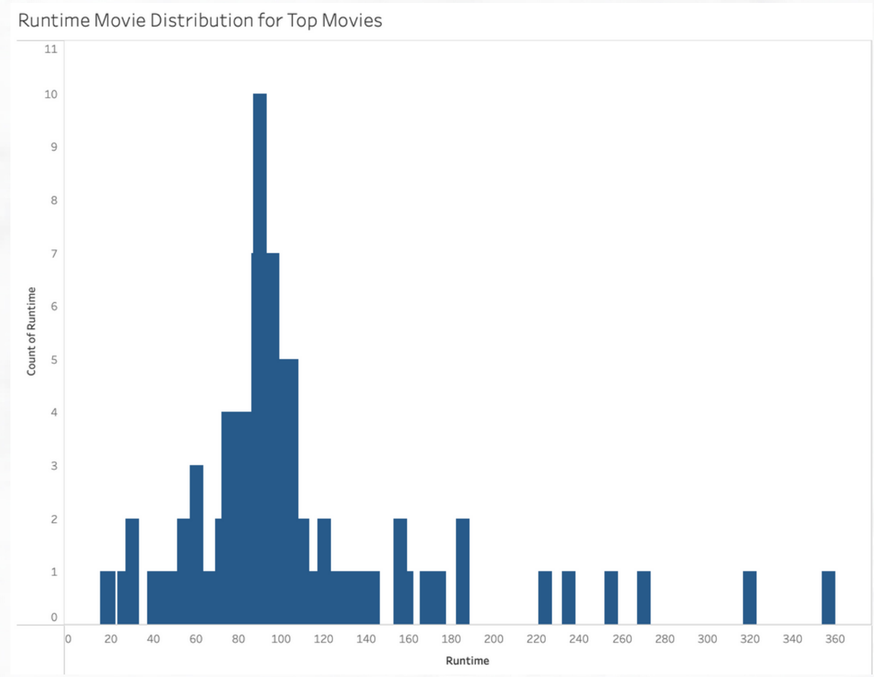
Adult Movie Distribution

Subset Labels	Adult	# Movies
Top Movies	False	190
Others	False	45,264
	True	9

Takeaways

10 voted movies aren't adult movies (100%), in other hand the less 10 voted movies are most often also not for adults.



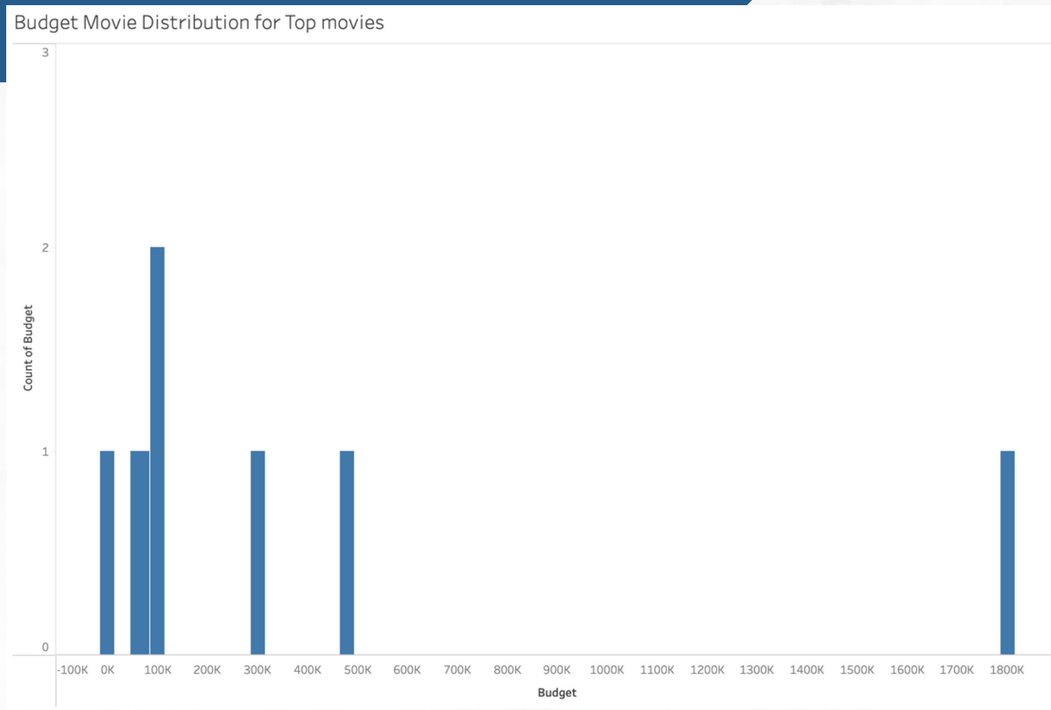


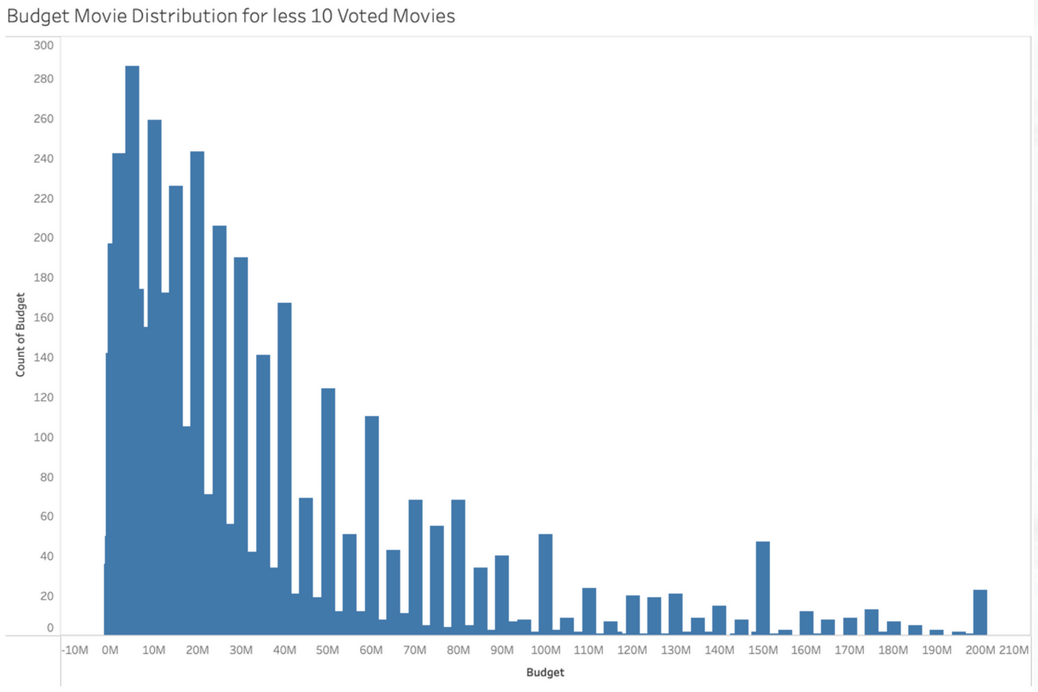
10 voted movies are less 10 minutes, on average compare to less 10 voted movies

Takeaways

There is no apparent difference in runtime feature between two groups. The runtime of 10 voted movies is 87 on average, while the less 10 voted movies is 94.

	Top Movies	Others
Min. Runtime	0	0
Percentile (25) of Runtime	71	85
Median Runtime	90	95
Avg. Runtime	87	94
Percentile (75) of Runtime	100	107
Upper Whisker (upper outlier)	143	140
Max. Runtime	357	1,256





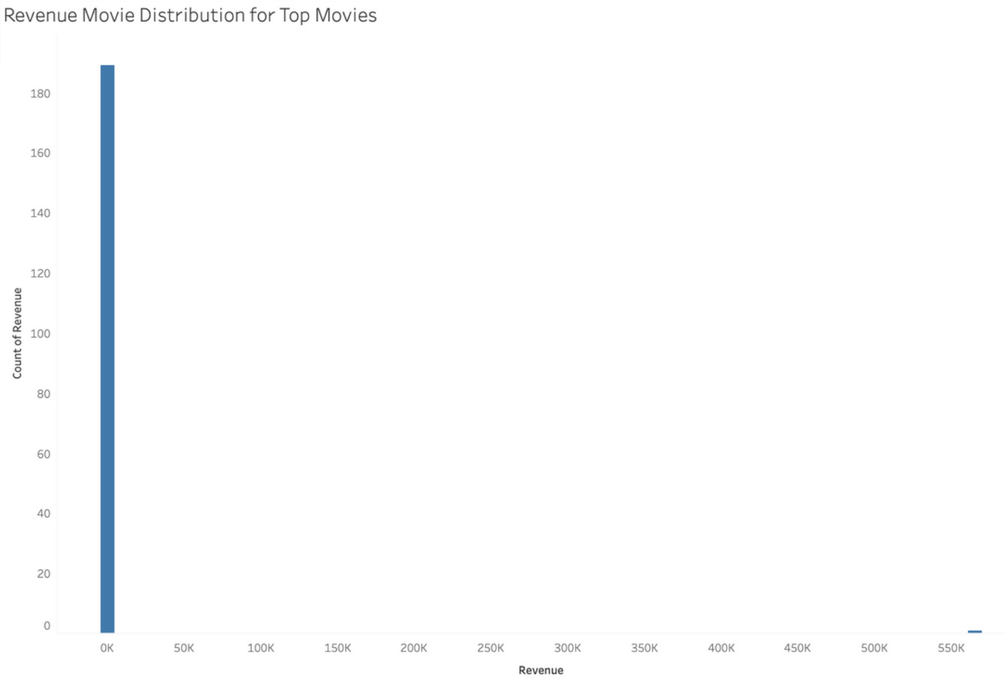
10 voted movies cost less than the rest Movies, on average $ 323,370

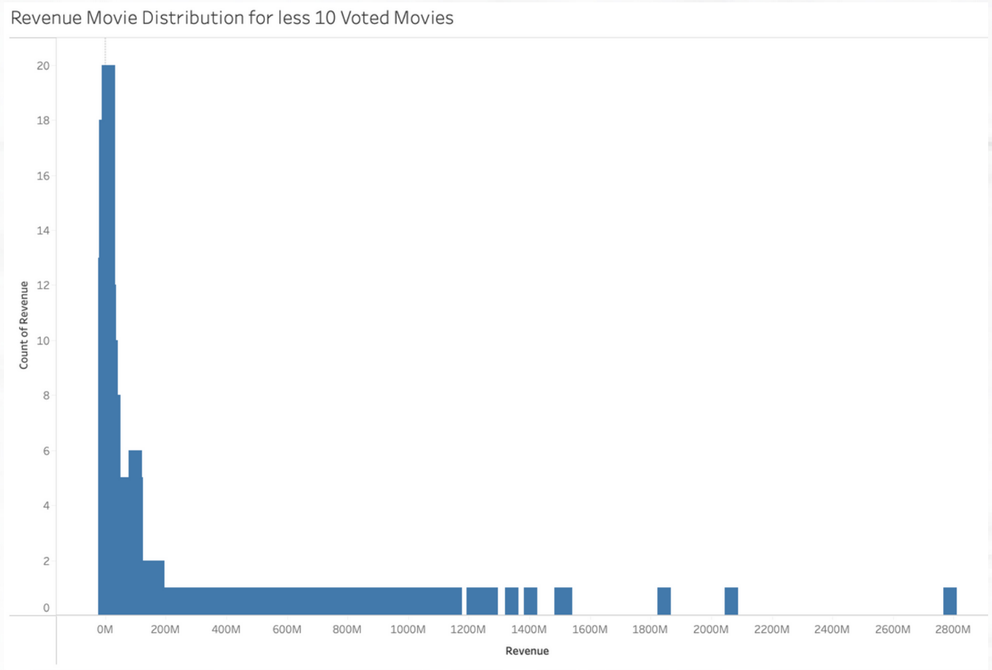
Takeaways

The budget of the top 10 voted movies is less than the less 10 voted movies by $ 21,496,570 on average. All 10 voted movies's budget are under $1,800,000.

- I excluded budget values less than $10 from the below statistical analysis.

	Top Movies	Others
Min. Budget	130	10
Lower Whisker (low outlier) B..	130	10
Percentile (25) of Budget	60,000	2,000,000
Median Budget	100,000	8,500,000
Avg. Budget	323,370	21,819,940
Percentile (75) of Budget	300,000	25,000,000
Upper Whisker (high outlier) ..	480,000	59,000,000
Max. Budget	1,800,000	380,000,000





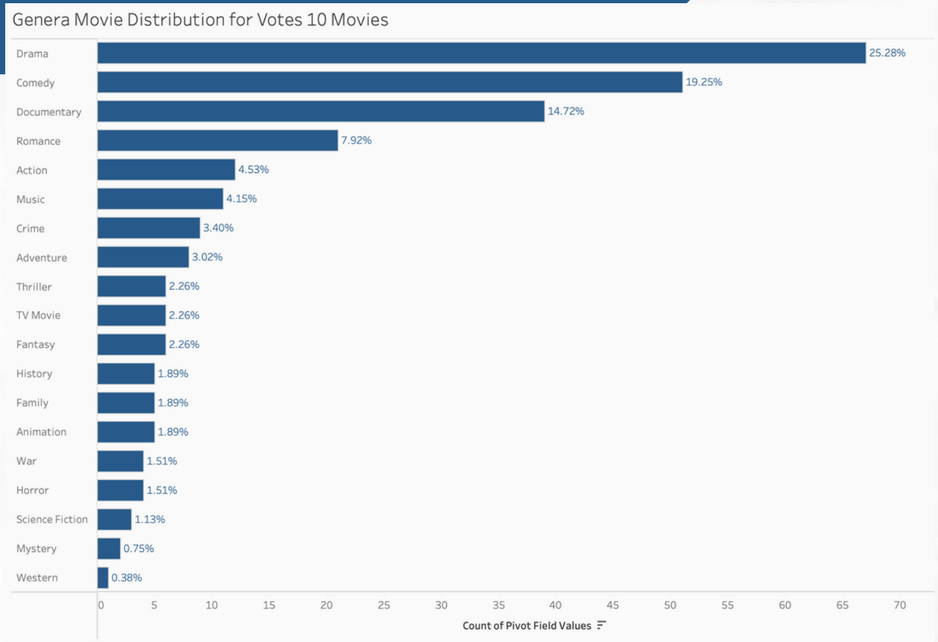
We don't have enough info about revenue of 10 voted movies

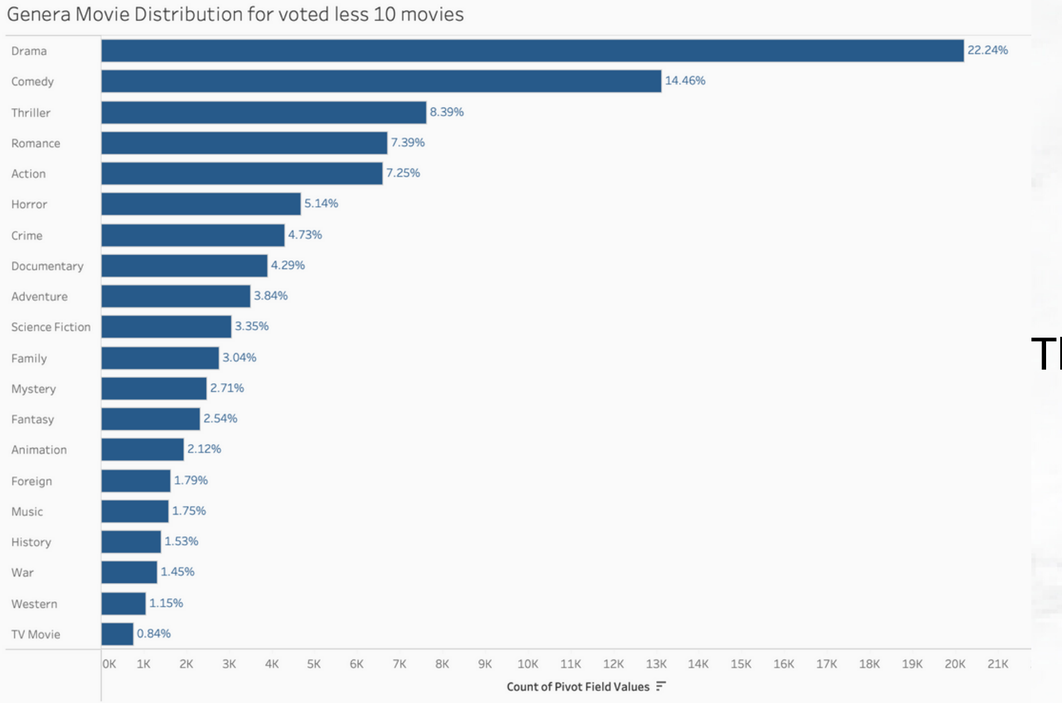
Takeaways

Revenue feature isn't useful to know which features drives the current vote value, since all revenue values for 10 voted movies are zeros except one movie.

- I excluded revenue values less than $10 from the below statistical analysis.

	Top Movies	Others
Min. Revenue	565,592	10
Lower Whisker (low outlier) R..	565,592	10
Percentile (25) of Revenue	565,592	2,500,000
Median Revenue	565,592	17,007,327
Avg. Revenue	565,592	69,217,117
Percentile (75) of Revenue	565,592	67,979,665
Upper Whisker (high outlier) ..	565,592	165,615,285
Max. Revenue	565,592	2,787,965,087





The genera of 10 voted movies are most often drama, comedy, documentary, romance

Takeaways

The drama genera is the top in both groups (voted 10 movies and the other).

The top genera for 10 voted movies:

- Drama 22.28%
- Comedy 19.25%
- Documentary 14.72%
- Romance 7.92%

The top genera for less 10 voted movies:

- Drama 22.4%
- Comedy 14.46%
- Thriller 8.39%
- Romance 7.39%

Limitation and biases

In Data Collection

- All revenue values of 10 voted movies are zero except one movies.
- Data is uncleaned, there is a JSON code in multiple fields.
- Under coverage bias since the respondents were only from those who visited the IMDB site.

In Data Processing

- Outlier in genera, budget, revenue and runtime features
- Missingness in genera, budget, revenue and runtime feature

Next Steps

- Provide information about revenue of 10 voted movies.
- Analyze the entire data set by examining other features like release date, production company and country.
- Normalize data to address skewed data.

Thank you!

Manal Almehmadi
mmehmadi94@gmail.com

11 Nov 2022