UDACITY DATA VISUALIZATION NANODEGREE

Mid Term Project

Kay Sun

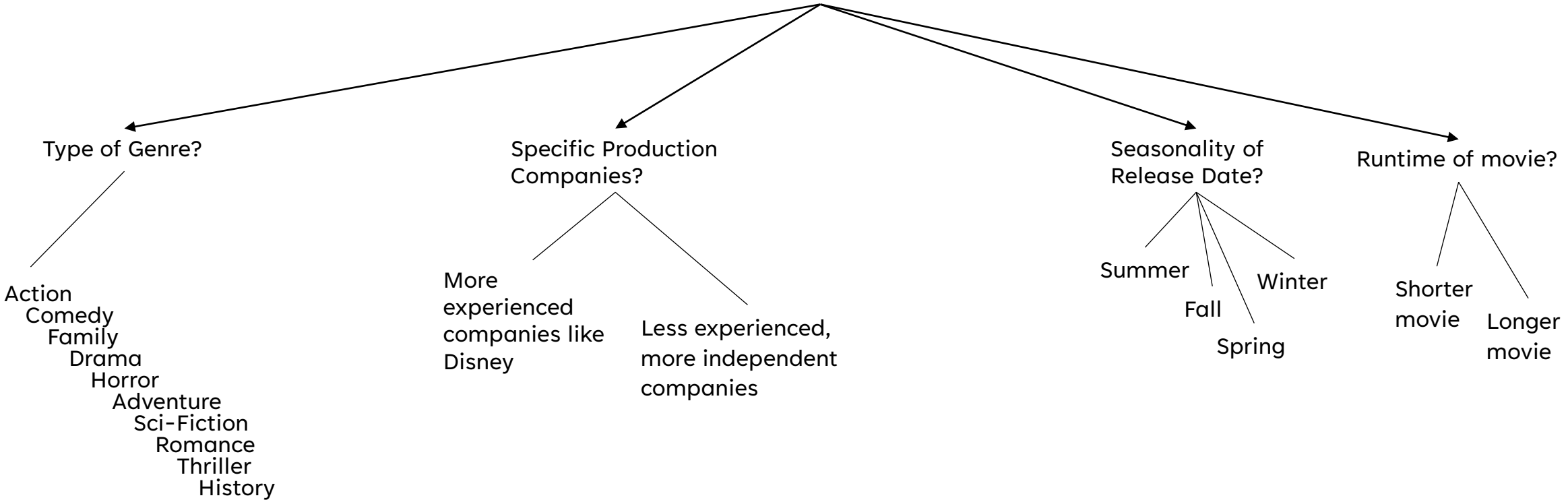
October 20 2023

PROBLEM STATEMENT

What are the **key movie features** the **top 5% movies** have compared to the **bottom 95% movies** when considering **profits**?

ISSUE TREE

What are the **key movie features** the **top 5% movies** have compared to the **bottom 95% movies** when considering **profits**?



SYNTHESIS

GENRE

Drama, Comedy, and Romance are most common genres to both Top and Bottom movies.

Adventure, Action, Animation are most profitable genres.

PRODUCTION COMPANY

Productive production companies make many movies.

Some are top grossing or most are not. Hit or miss.

RELEASE DATE

Q2 and Q4 are most profitable for Top Movies.

All quarters are loss for Bottom Movies

MOVIE RUNTIME

Median runtime of 100 mins similar for Top and Bottom Movies.

RECOMMENDATION

GENRE

Adventure, Action, Animation are most profitable genres.

Target these genres for next potential Top Movie if budget is at least $23 million.

PRODUCTION COMPANY

Disney, Columbia, Universal are a few of top production companies.

Beware they can also make the least profitable movies.

RELEASE DATE

Release movies in Q2 (Summer) and Q4 (Holidays).

MOVIE RUNTIME

Runtime is heavily influenced by human behavior and psychology.

Too long loses interest and too short can’t make the emotional investment.

Median runtime of 100 mins.

ANALYSIS OVERVIEW

What drives the profitability for the Top 5% of movies?

Profit = Revenue - Budget

GENRE

Compare profits by genres for Top and Bottom Movies to find the most common genres and most/least profitable.

PRODUCTION COMPANY

Compare profits by production companies for Top and Bottom Movies to find the most common genres and most/least profitable.

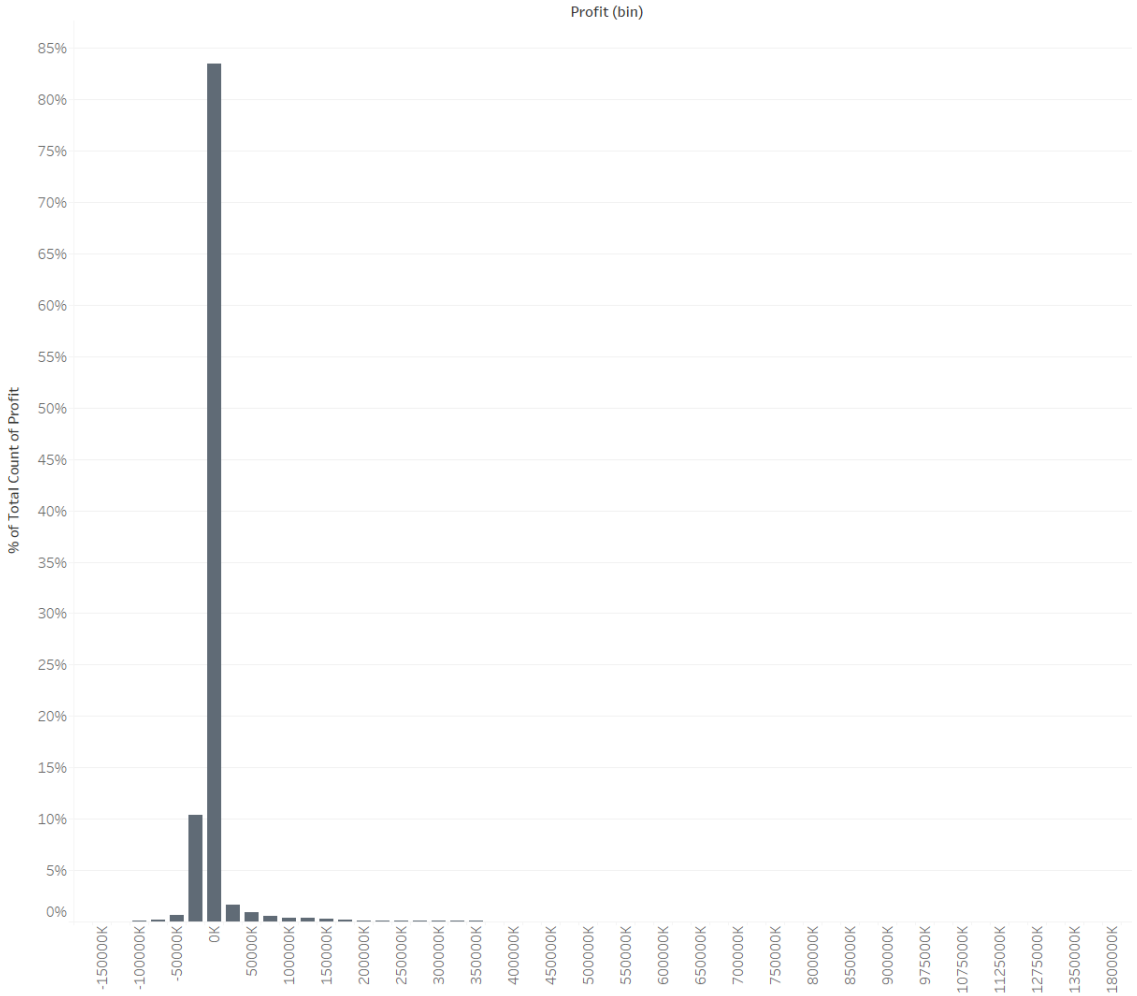
RELEASE DATE

Compare profits by quarter for Top and Bottom Movies to find the most profitable release date.

MOVIE RUNTIME

Compare profits by runtime for Top and Bottom Movies to find any correlation.

MOST MOVIES DO NOT MAKE ANY PROFITS



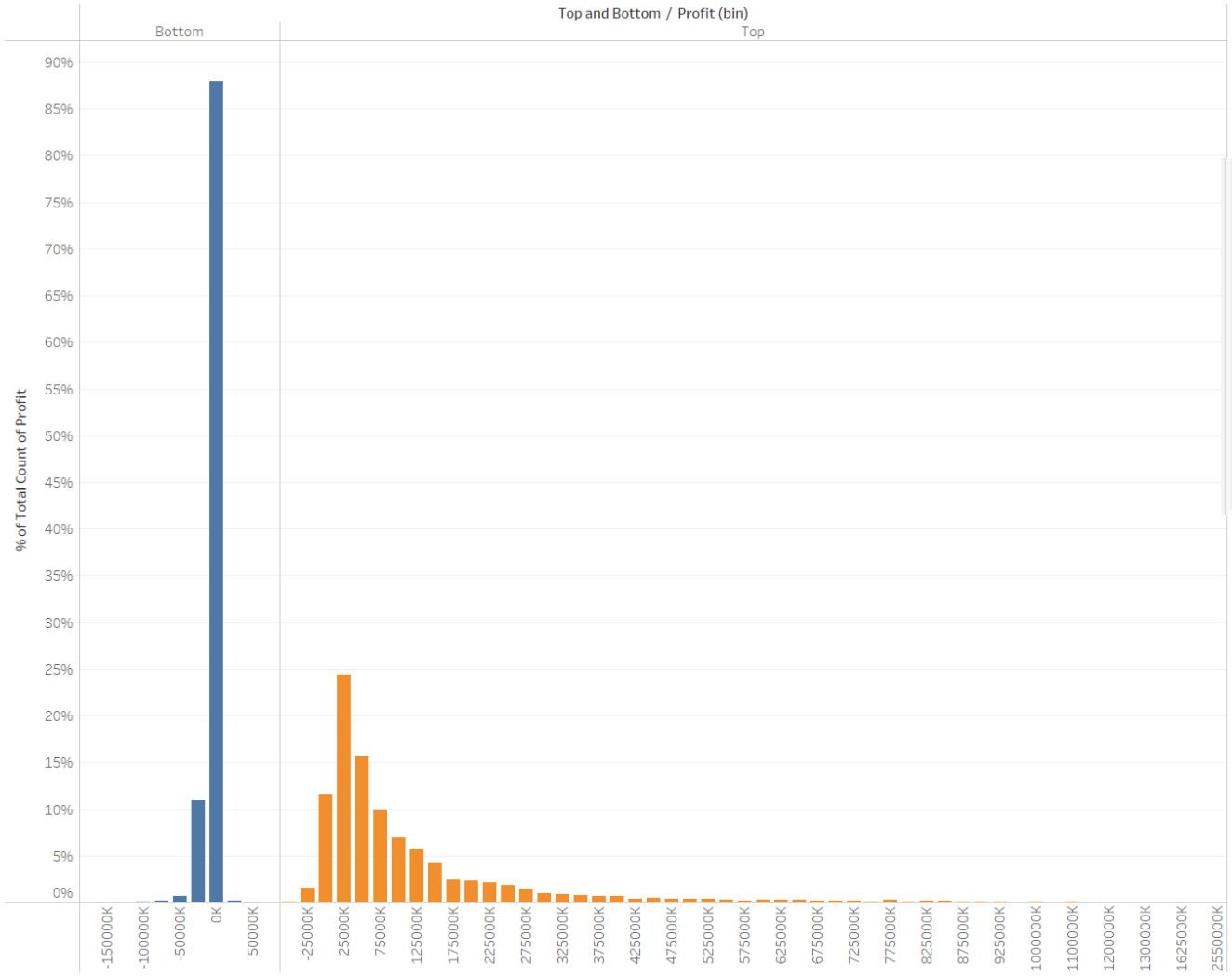
Sum:	317,514,957,998
Average:	6,984,490.94
Minimum:	-165,710,090
Maximum:	2,550,965,087
Median:	0.00
Standard deviation:	52,141,254
First quartile:	0.00
Third quartile:	0.00
Skewness:	13.96

Profit defined by Revenue - Budget.

Profit is highly skewed with all movies making a median of no profits but mean of $7 million.

Include extremes to explore why the Top 5% of movies have in common compared to the Bottom 95%.

TOP 5% OF MOVIES MAKES MUCH MORE PROFITS THAN BOTTOM 95% OF MOVIES



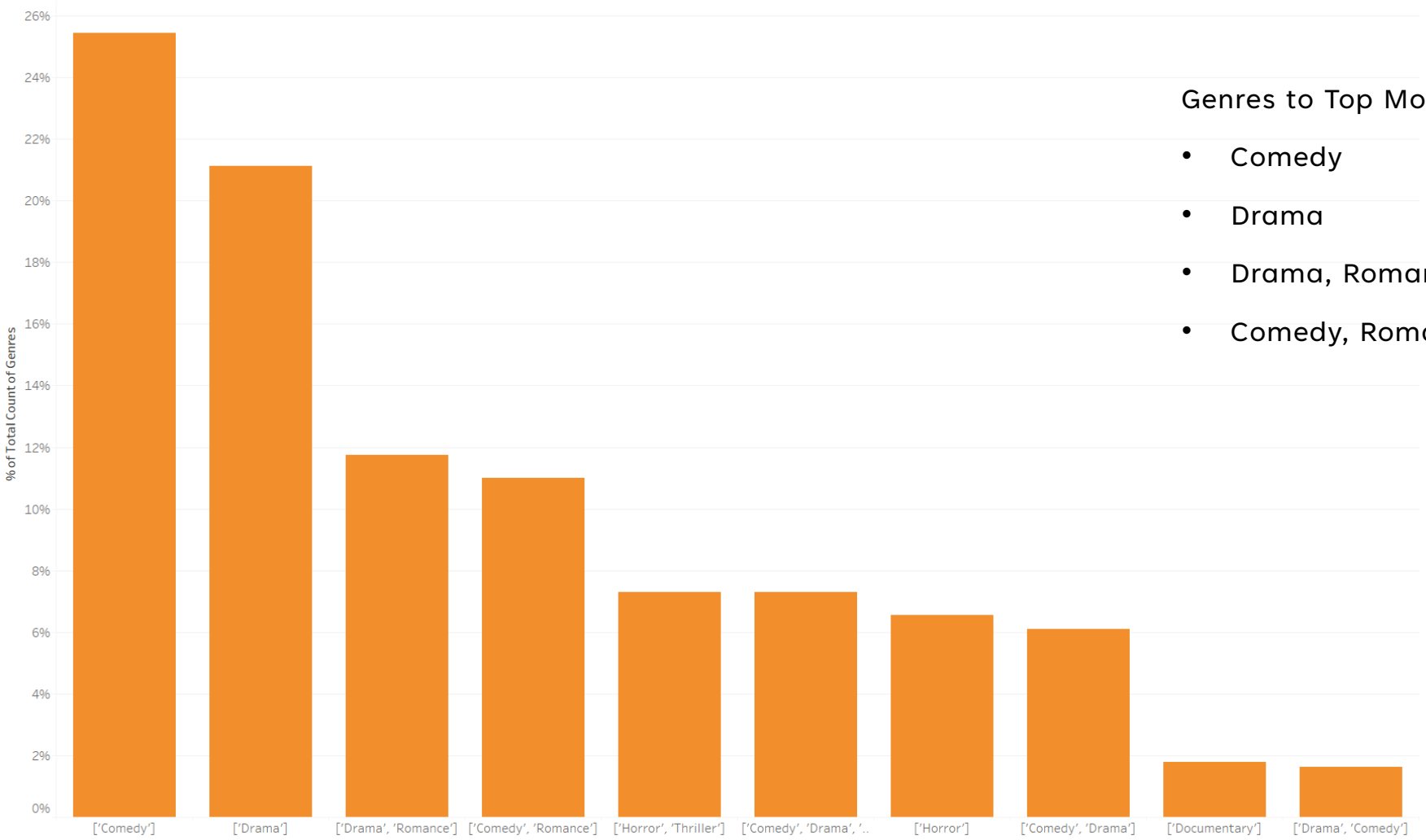
	Top	Bottom
Sum:	336,499,771,205	-18,984,813,207
Average:	126,885,283.26	-443,487.51
Minimum:	-34,208,170	-165,710,090
Maximum:	2,550,965,087	81,289,342
Median:	69,036,822.50	0.00
Standard deviation:	175,291,137	6,199,084
First quartile:	36,488,936.00	0.00
Third quartile:	146,223,962.25	0.00
Skewness:	3.94	-6.15

Split all movies into 2 cohorts – Top 5% and Bottom 95% by profits.

Many more Bottom Movies than Top Movies.

Median profit of Top Movies is much higher at $69 million while median Bottom Movies is $0.

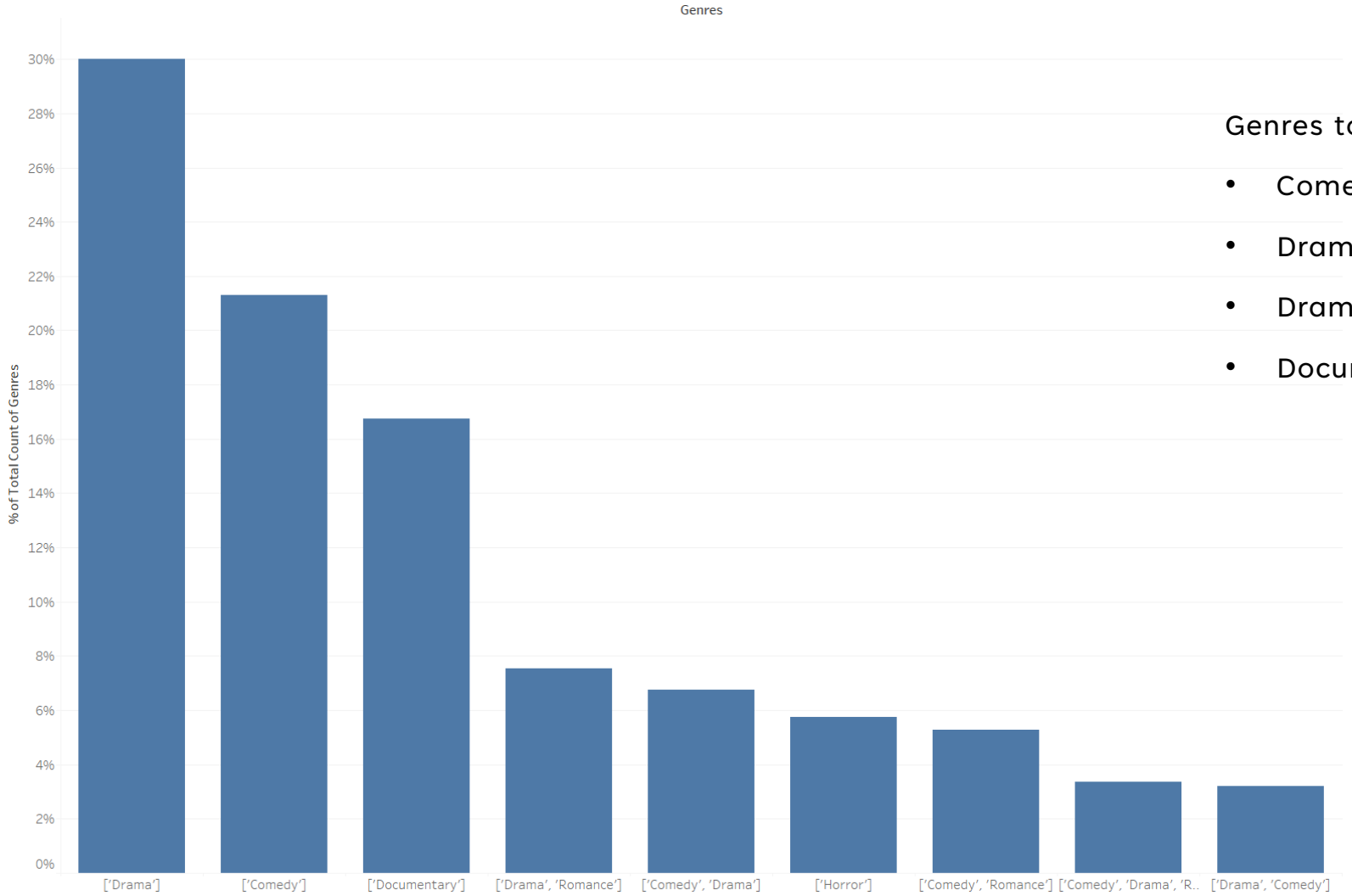
COMEDY, DRAMA, ROMANCE ARE MOST COMMON GENRES TO TOP 5% OF MOVIES



Genres to Top Movies

- Comedy
- Drama
- Drama, Romance
- Comedy, Romance

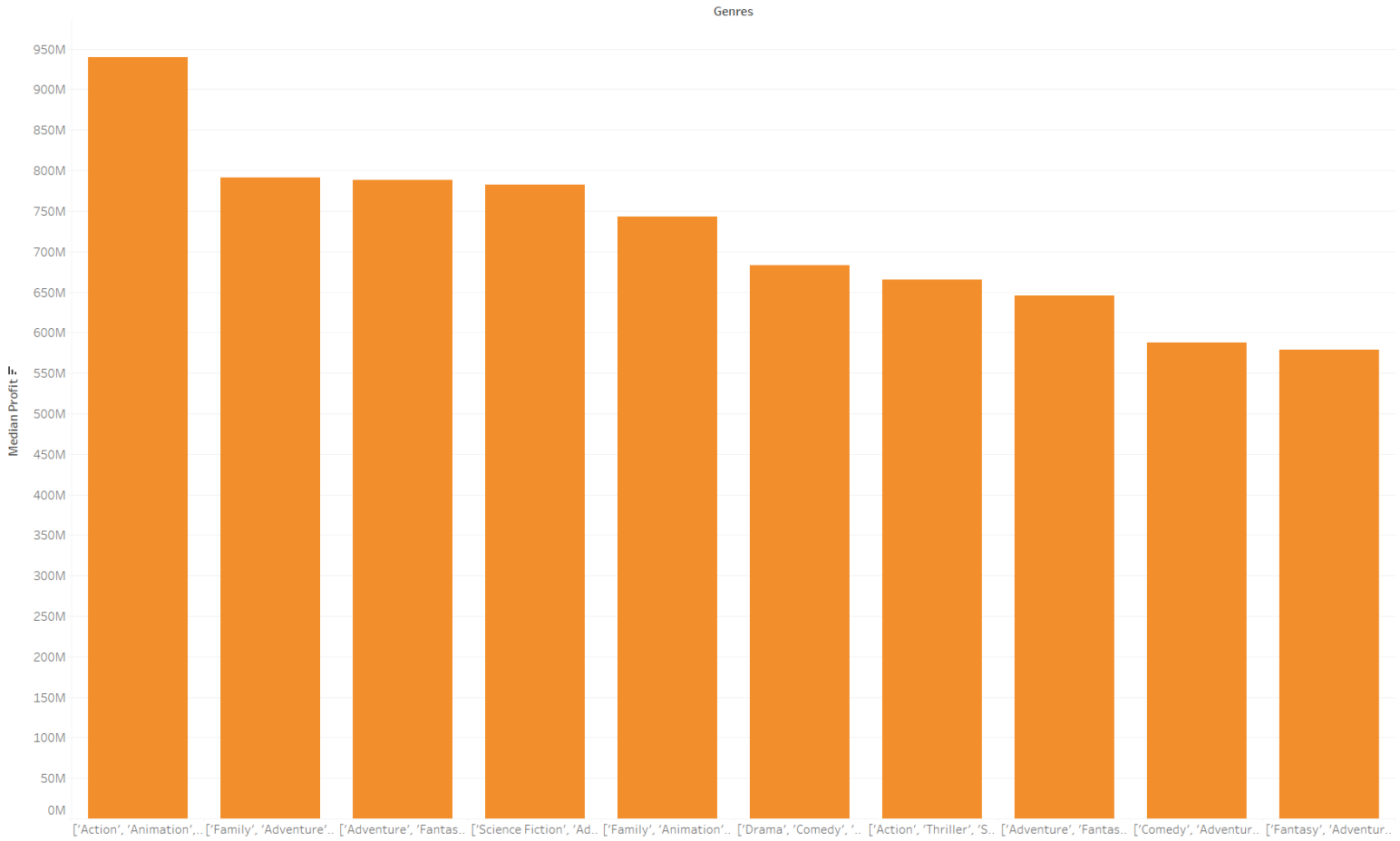
DRAMA, COMEDY, DOCUMENTARY, ROMANCE ARE MOST COMMON GENRES TO BOTTOM 95% OF MOVIES, SIMILAR TO TOP MOVIES



Genres to Bottom Movies similar to that of Top Movies

- Comedy
- Drama
- Drama, Romance
- Documentary

ADVENTURE, ACTION, ANIMATION GENRES ARE MOST PROFITABLE GENRES TO TOP 5% OF MOVIES



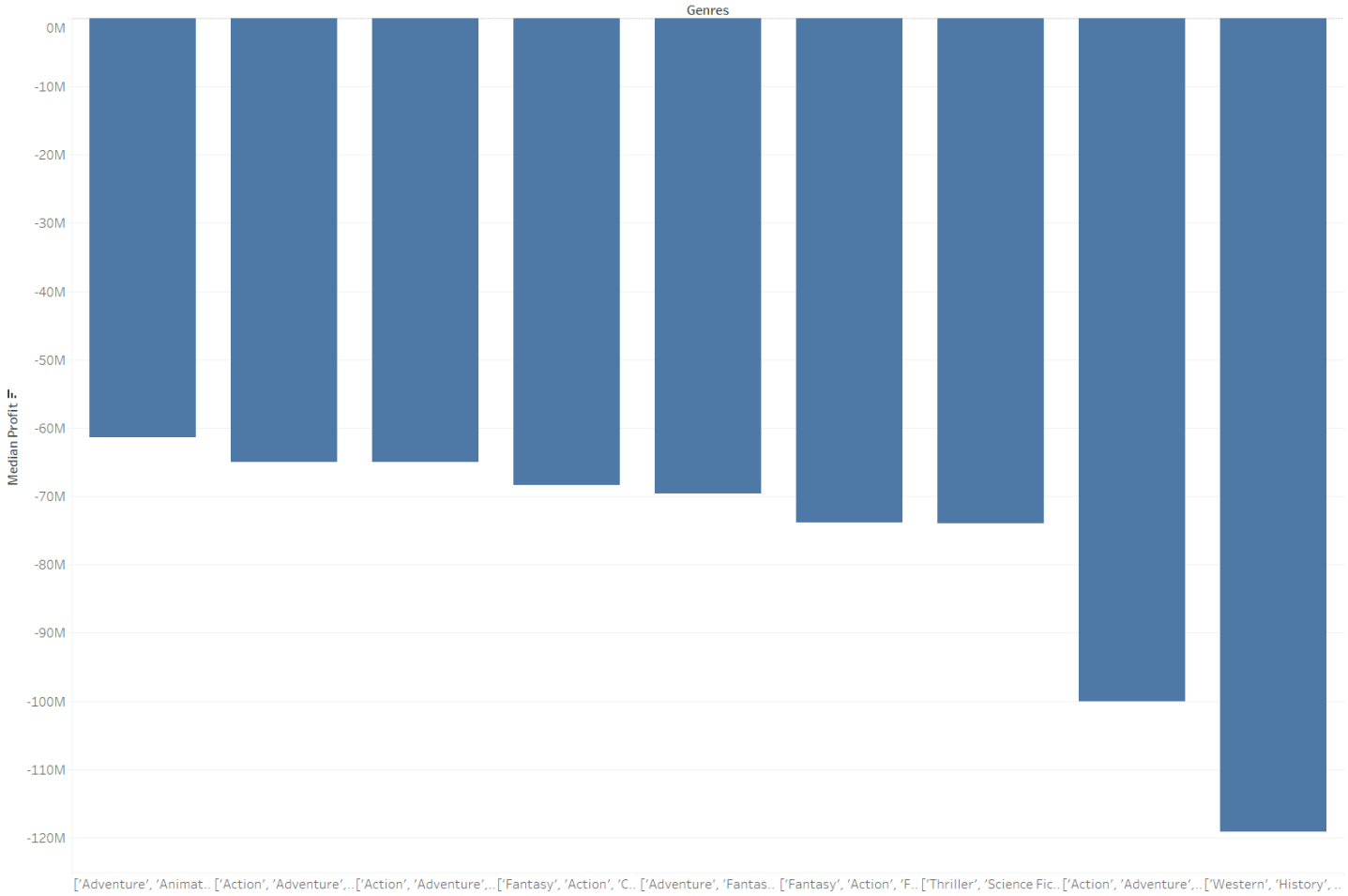
Most profitable genres to Top Movies

- Action, Animation
- Family, Adventure
- Adventure, Fantasy
- Science Fiction, Adventure

Target these genres for next potential Top Movie if budget is at least $40 million.

	Top Movies Budget
Average:	40,049,088.04
Minimum:	0
Maximum:	380,000,000
Median:	23,000,000.00
Standard deviation:	49,469,042
First quartile:	3,500,000.00
Third quartile:	57,000,000.00
Skewness:	1.90

DRAMA, CRIME, THRILLER ARE MOST PROFITABLE GENRES TO BOTTOM 95% OF MOVIES

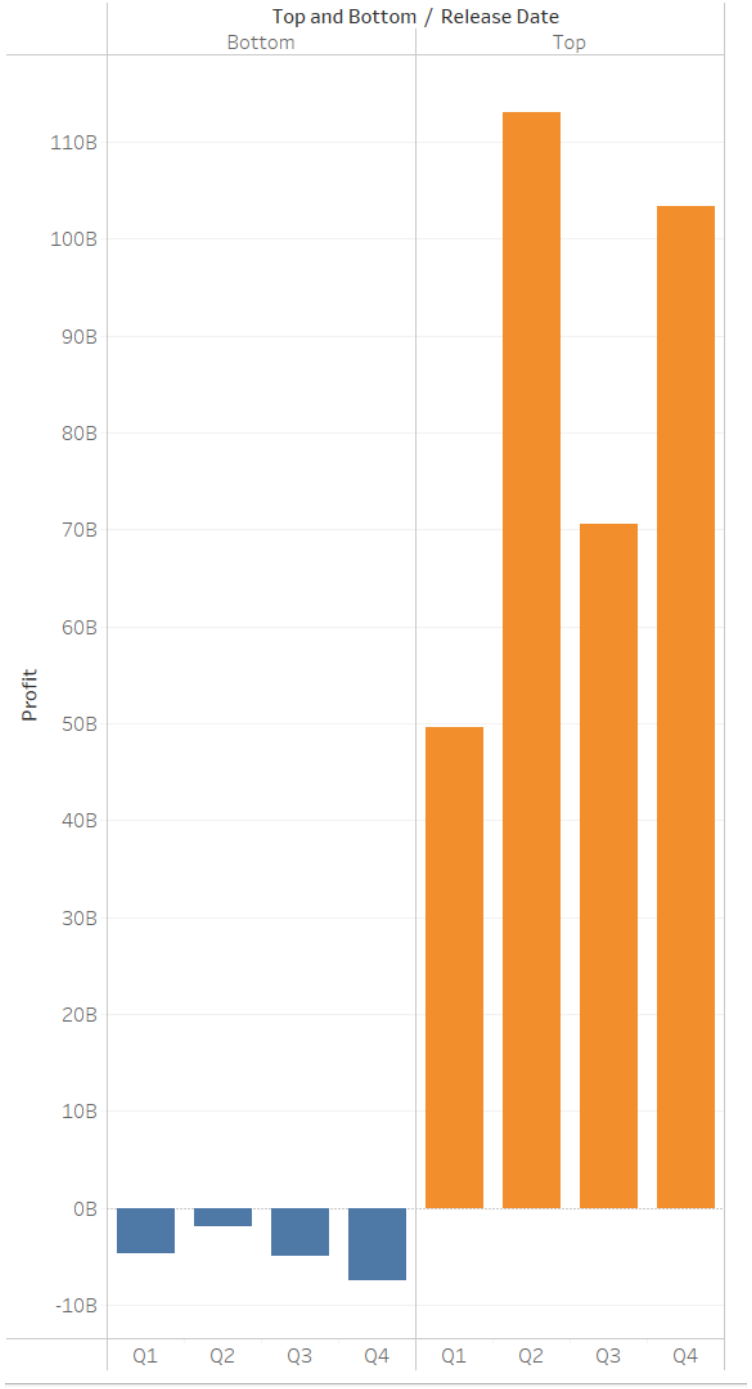


Genres to least profitable Bottom Movies

- Wester, History
- Action, Adventure
- Thriller, Science Fiction
- Fantasy, Action

Avoid these genres if budget less than budget is at $40 million.

	Bottom Movies Budget
Average:	2,005,509.37
Minimum:	0
Maximum:	260,000,000
Median:	0.00
Standard deviation:	9,295,892
First quartile:	0.00
Third quartile:	0.00
Skewness:	8.96



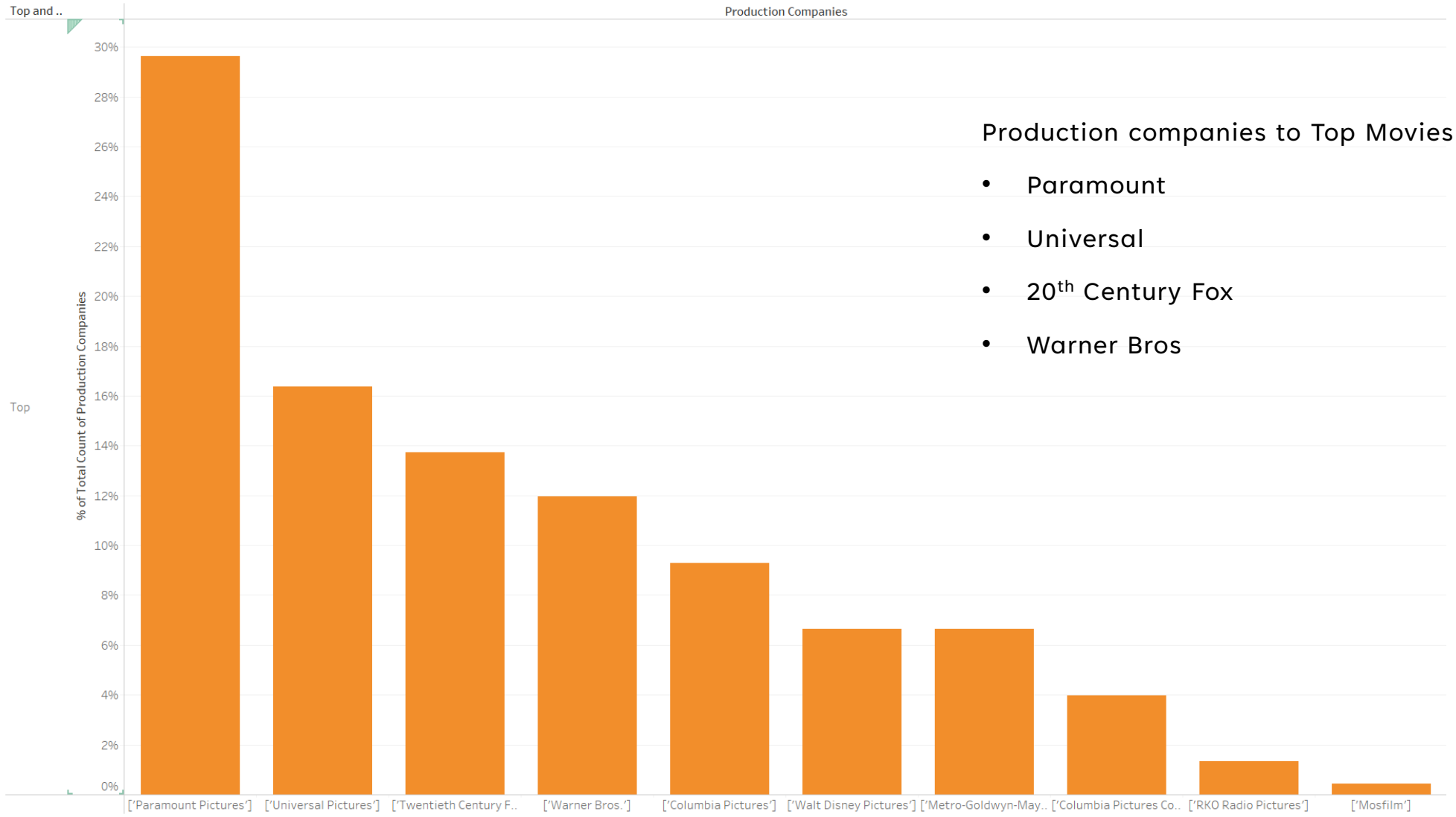
Q2 AND Q4 ARE MOST PROFITABLE FOR TOP 5% OF MOVIES ALL QUARTERS ARE LOSS FOR BOTTOM 95% OF MOVIES

Q2 and Q4 are more profitable for Top Movies.

Q4 is typically the worst profits for Bottom Movies.

Q2 is best profits for Bottom Movies, but all quarters are negative profits.

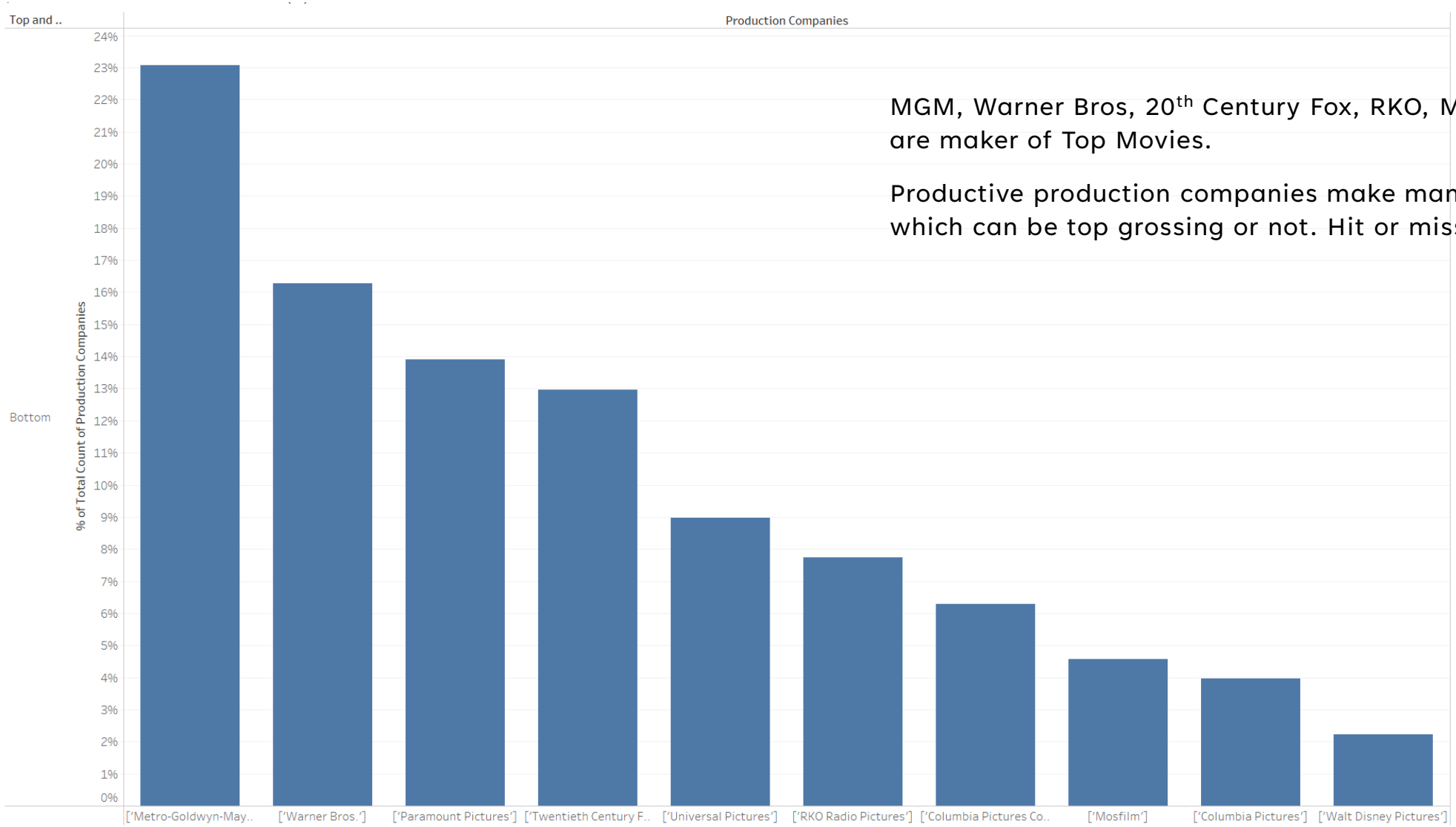
PARAMOUNT, UNIVERSAL, 20TH CENTURY FOX ARE THE COMMON PRODUCTION COMPANIES TO THE TOP MOVIES



Production companies to Top Movies

- Paramount
- Universal
- 20th Century Fox
- Warner Bros

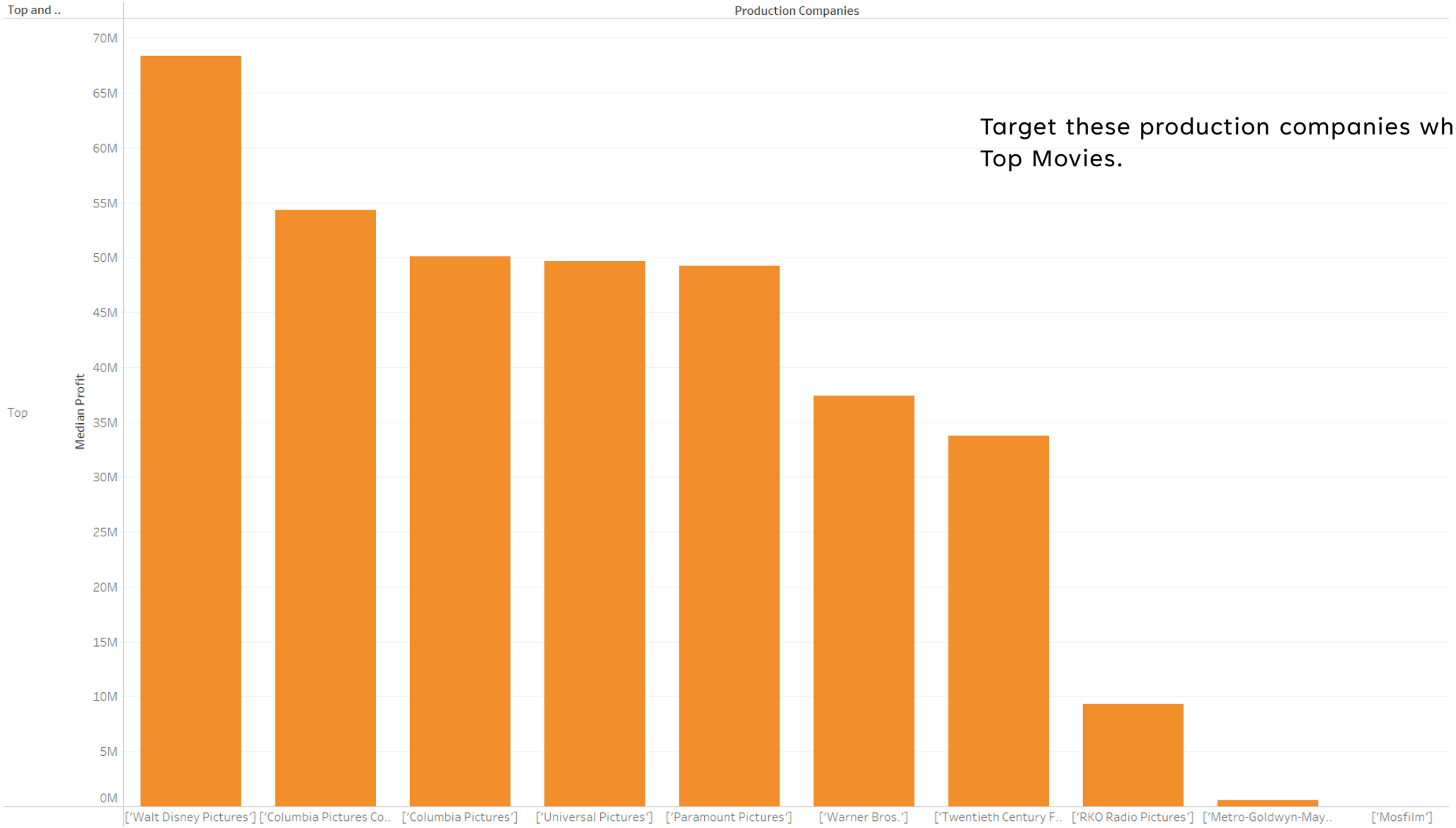
PRODUCTIVE PRODUCTION COMPANIES MAKE BOTH TOP AND BOTTOM MOVIES



MGM, Warner Bros, 20th Century Fox, RKO, Mosfilm also are maker of Top Movies.

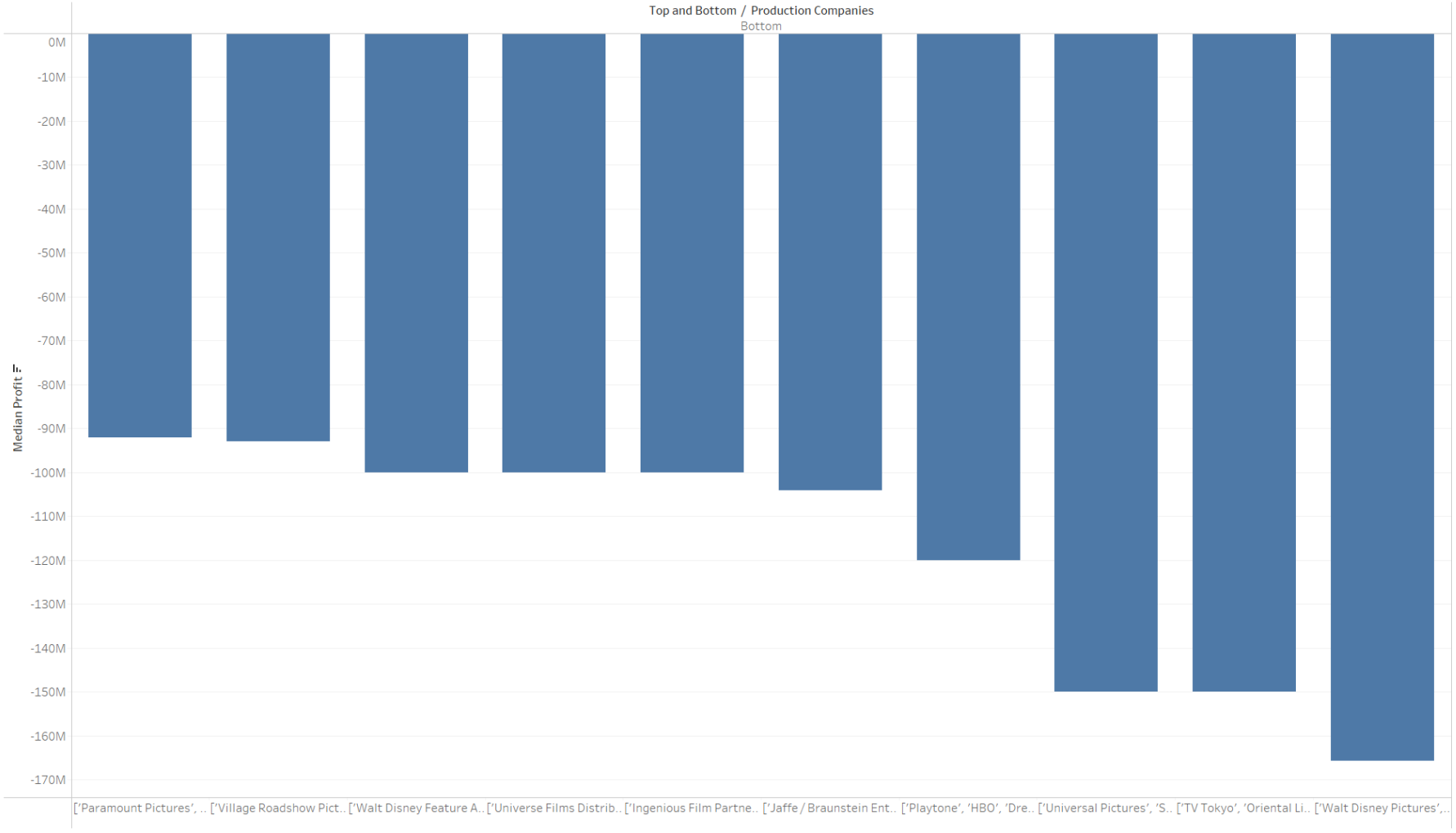
Productive production companies make many movies which can be top grossing or not. Hit or miss.

DISNEY, COLUMBIA, UNIVERSAL MAKE MOST MEDIAN PROFITS FOR THE TOP MOVIES



Target these production companies when making next Top Movies.

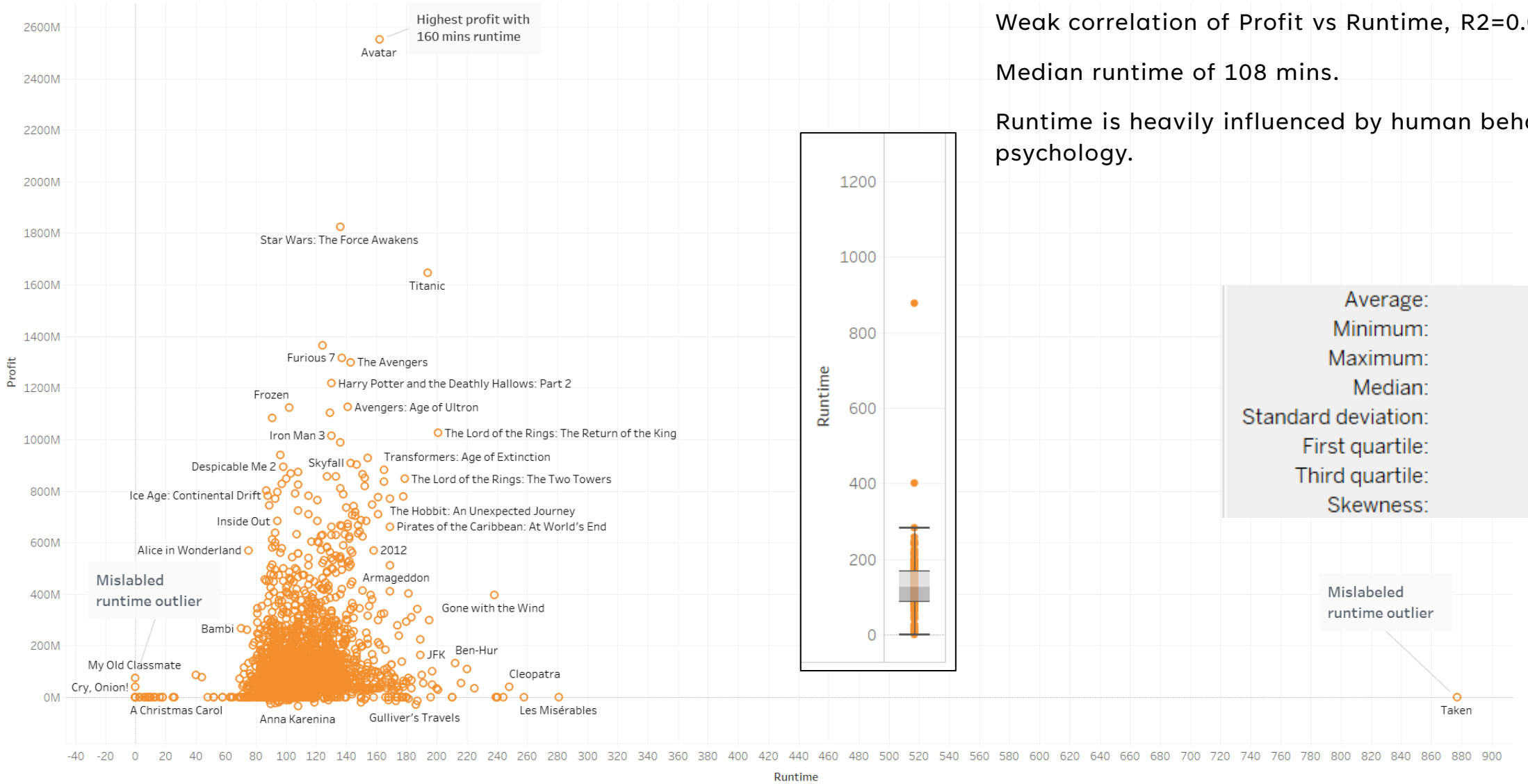
PARAMOUNT, DISNEY, RELATIVITY MAKE LEAST MEDIAN PROFITS FOR THE BOTTOM MOVIES



Disney and Universal also makes most and least profitable movies.

Hit or miss movie making.

LONGER MOVIES DOES NOT INDICATE MORE PROFITS FOR TOP MOVIES



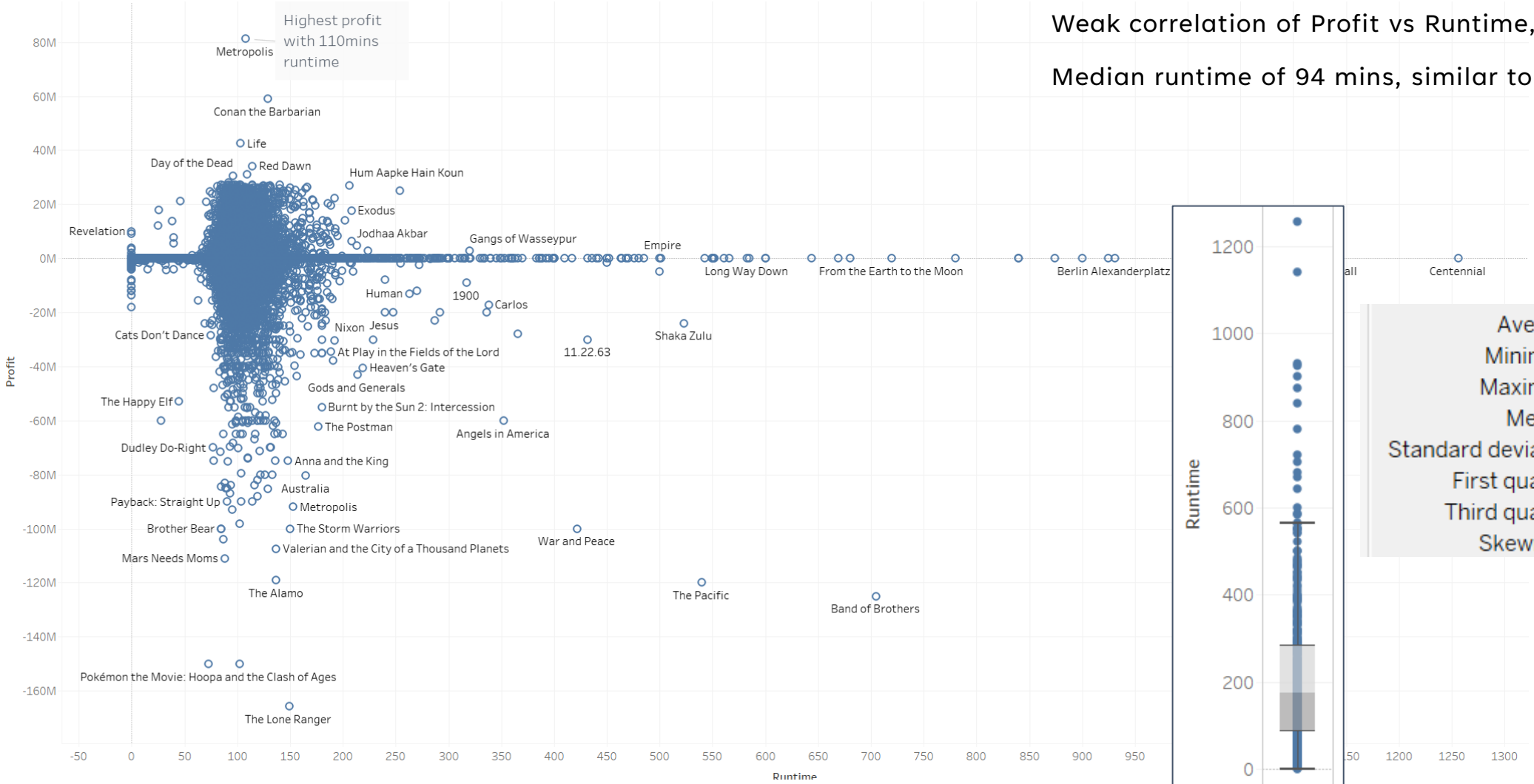
Weak correlation of Profit vs Runtime, R2=0.0272.

Median runtime of 108 mins.

Runtime is heavily influenced by human behavior and psychology.

Average:	111.6
Minimum:	0.0
Maximum:	877.0
Median:	108.0
Standard deviation:	29.6
First quartile:	96.0
Third quartile:	123.0
Skewness:	7.23

LONGER MOVIES DOES NOT INDICATE MORE PROFITS FOR BOTTOM MOVIES



Weak correlation of Profit vs Runtime, R2=0.0026.

Median runtime of 94 mins, similar to Top Movies.

Average:	93
Minimum:	0
Maximum:	1,256
Median:	94
Standard deviation:	39
First quartile:	85
Third quartile:	105
Skewness:	4.49

LIMITATIONS AND BIASES

Limitation

- Raw data is messy, have to be parsed to separate out like genre and production companies.

Data Collection

- Popularity and voter ratings were sourced from 1 website – GroupLens.
- Potential errors as dataset was assembled as part of an education coursework and not an official release with quality checks.

Data Processing

- Missing data for some fields.
- Mislabel data for some fields.

Data Insights

Profit defined by Revenue – Budget may not be present the P/E relationship. There could be extraneous factors like tax incentives, insurance payouts, etc.

NEXT STEPS

1. Impute missing values
2. Remove outliers.
3. Expand to include other datasets.
4. Expand to include analysis of other features and multi-variate analysis for combination of features.