Data Science

The Movie Industry

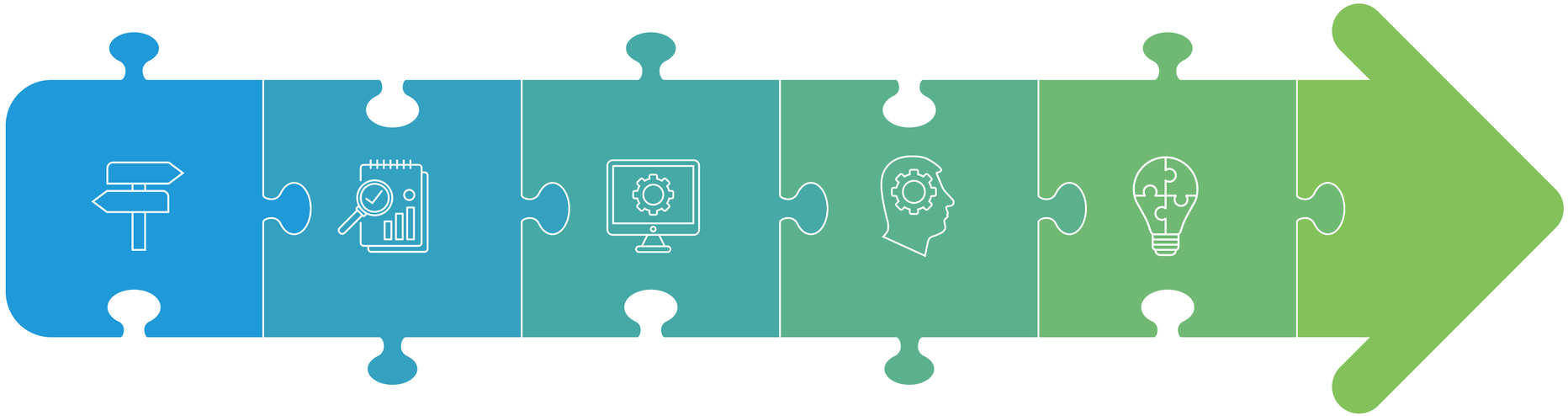
Bao Tram Duong

Research Process

Business
Objectives

Data
Cleaning & Preparation

Actionable Insights

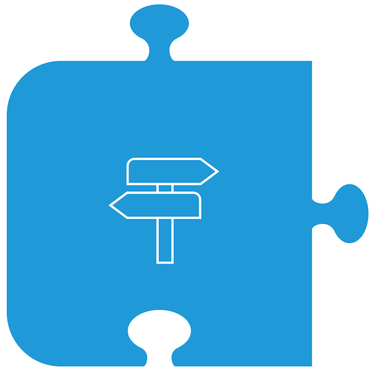


Data
Gathering

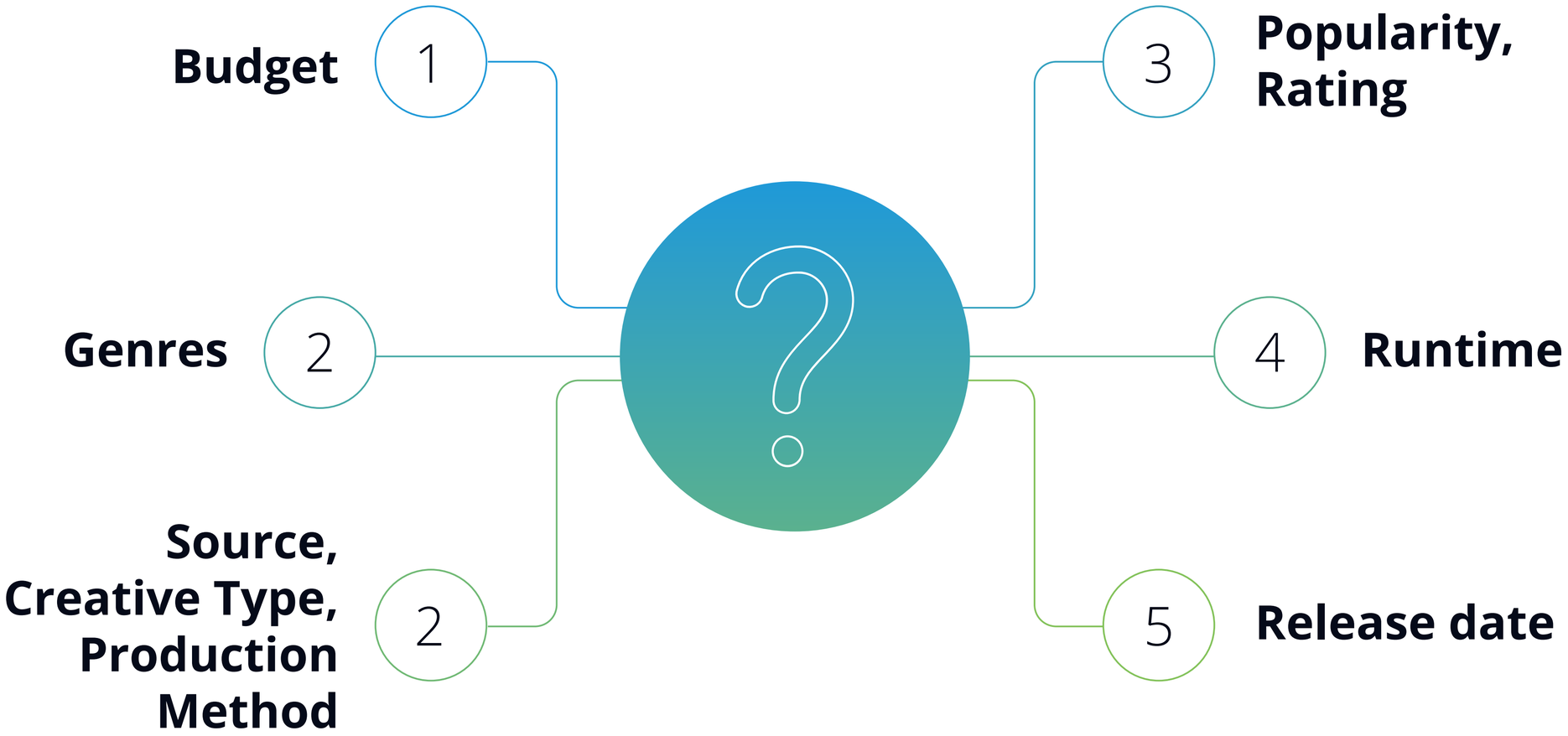
Data
Analysis & Visualization

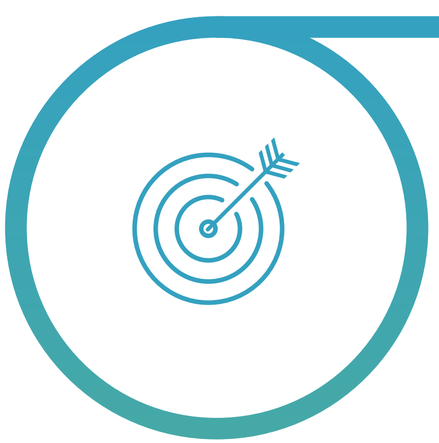
Research Process
Step 1

Business Objectives



Business Problems



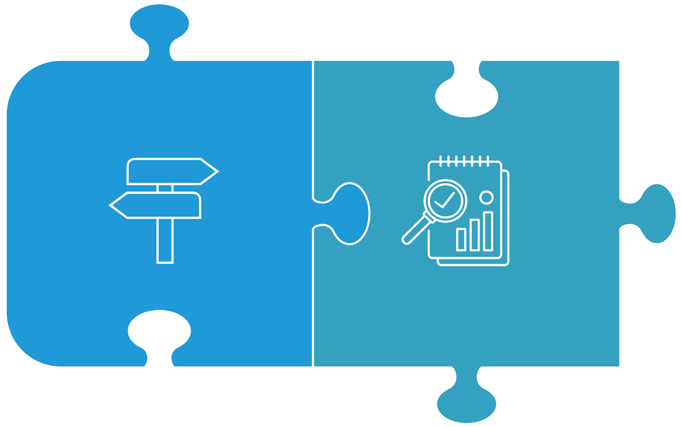


Business Statements

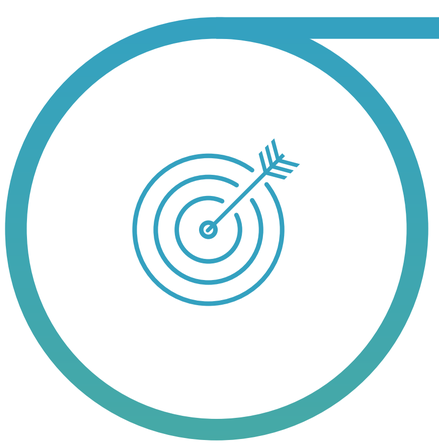
- **Q1.** Is there a correlation between production budget and profit? If so, how much should Microsoft invest into production?
- **Q2.** What kind of movie contents, in terms of genre, source, creative type, production method, perform the best?
- **Q3.** Is there a correlation between popularity and positive words of mouth (average rating) and profit? How do they affect the performance of a movie?
- **Q4.** When is the best time of year to release a movie?
- **Q5.** Is there a correlation between runtime and profit? What is the best runtime?

Research Process
Step 2

Business
Objectives

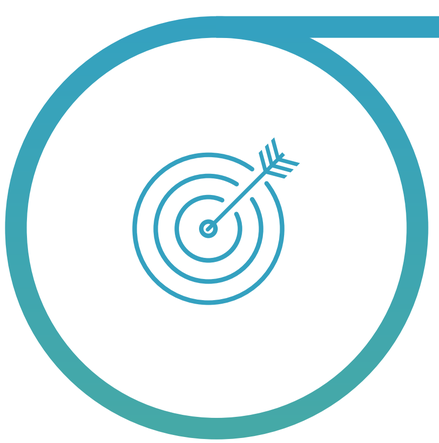


**Data
Gathering**



Features

- Production Budget
- Domestic Gross, International Gross, Worldwide Gross
- Genre, Source, Creative Type, Production Method
- Vote Count, Vote Average
- Popularity
- Release Time
- Runtime
- Studio



Data Sources

1. Box Office Mojo

2. IMDB

3. Rotten Tomatoes

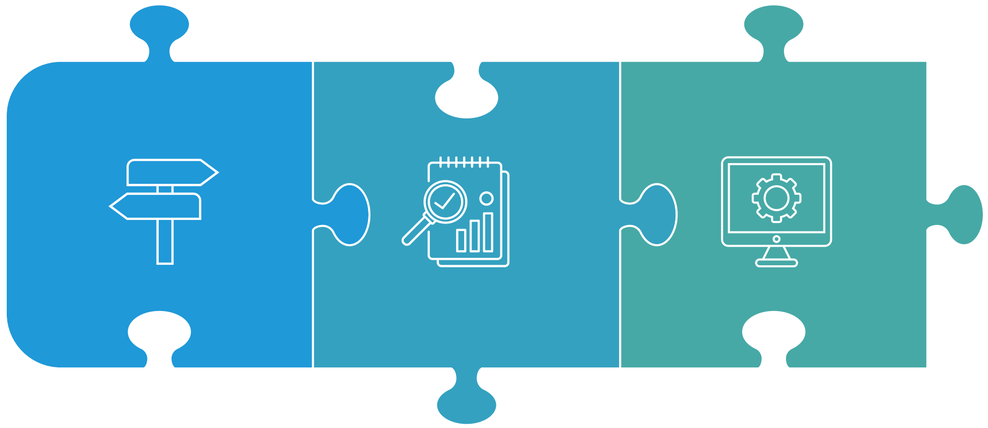
4. TheMovieDB.org

5.The-Numbers.com

Research Process
Step 3

Business
Objectives

**Data
Cleaning &
Preparation**

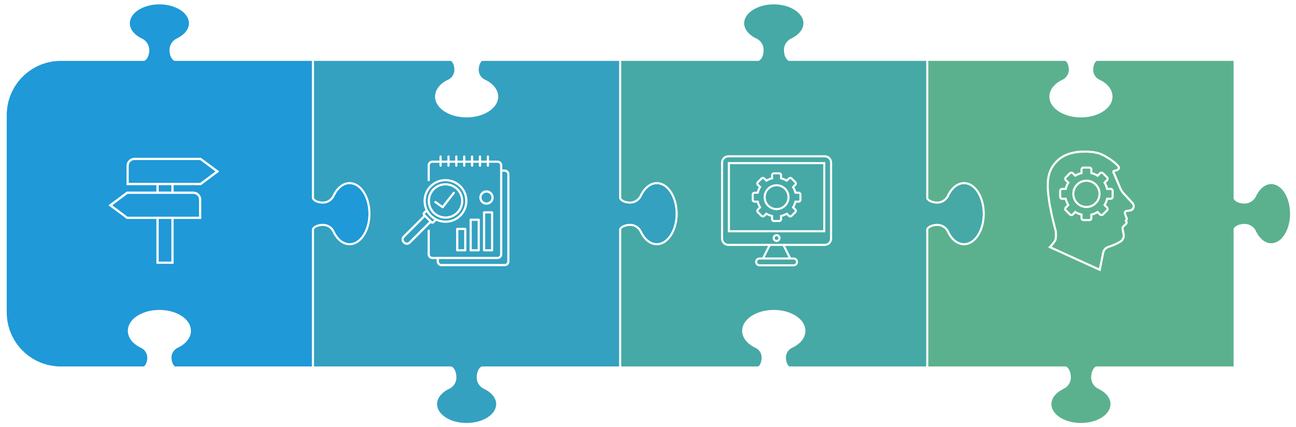


Data
Gathering

Research Process
Step 4

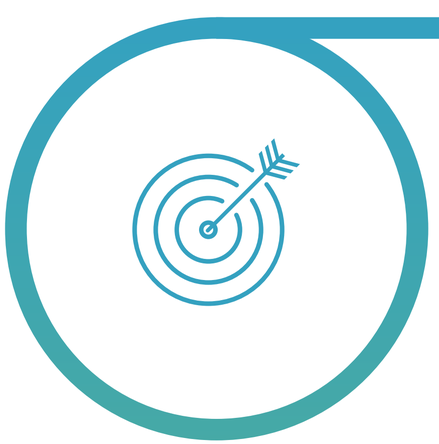
Business
Objectives

Data
Cleaning & Preparation



Data
Gathering

**Data
Analysis &
Visualization**

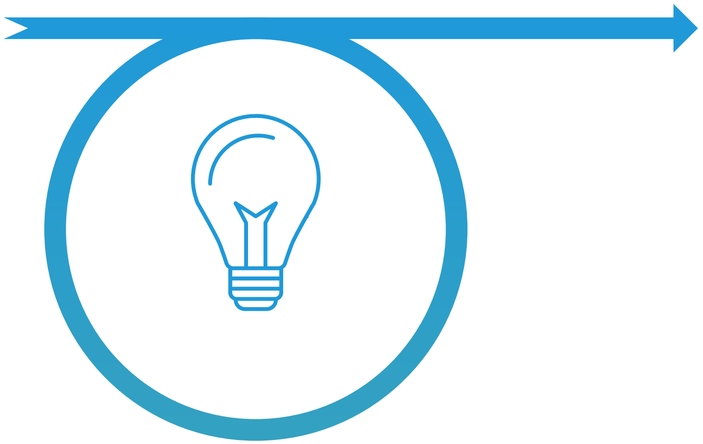


Methodology

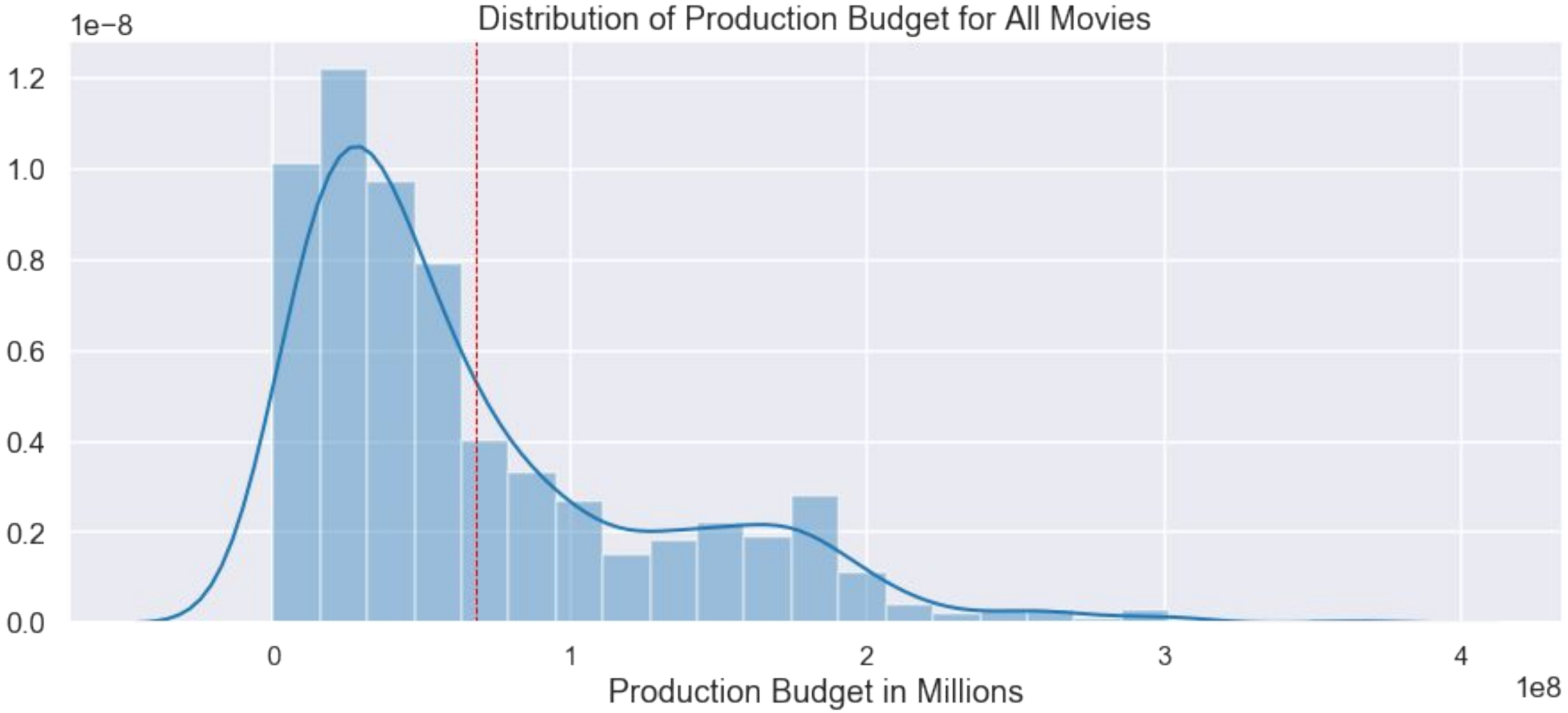
1. Get the **General Movies** trend/ distribution of all movies

2. Get distribution of **Top 100 Movies** to see what is done differently to reach higher success.

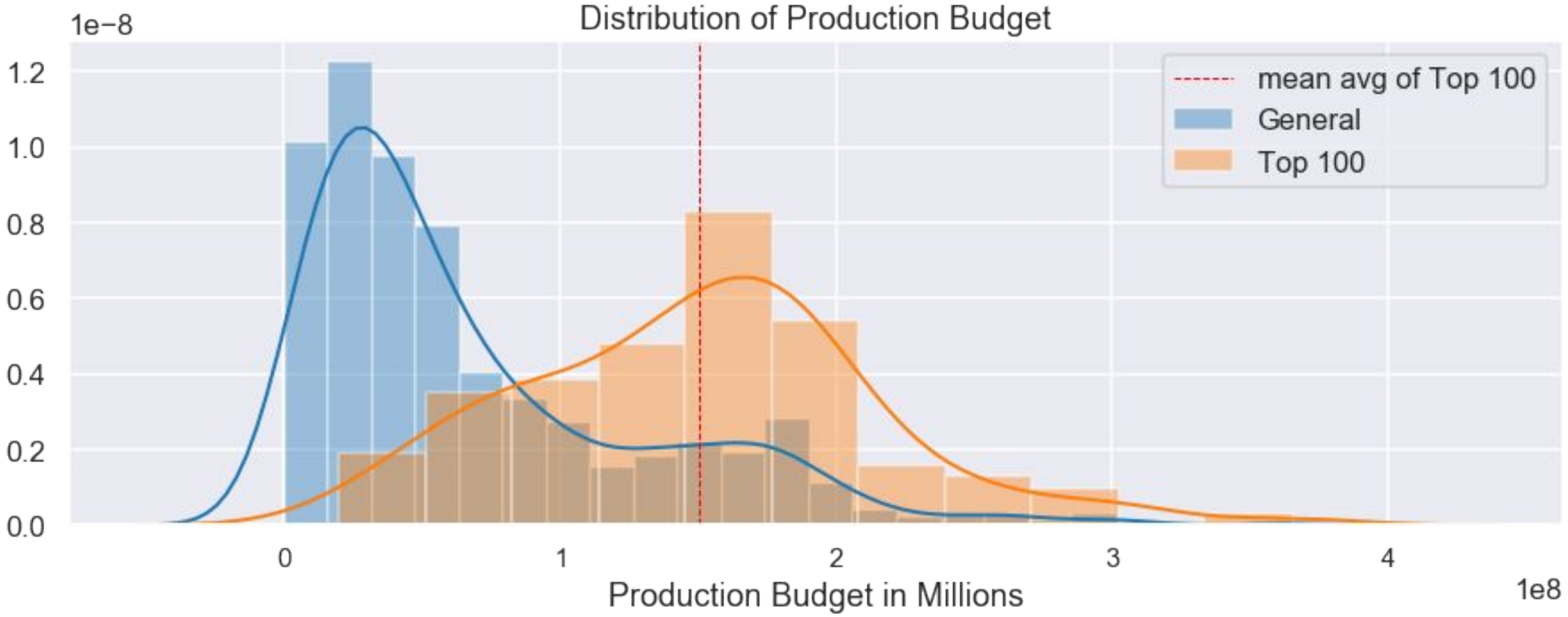
4. Analyze each attribute of the Top 100 Movies with respect to production budget and worldwide profit



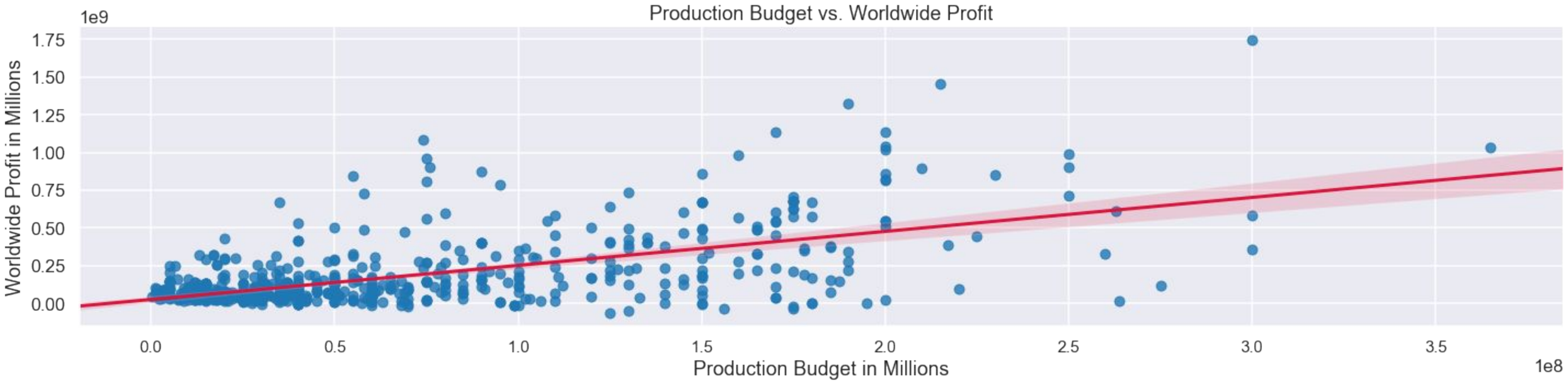
Business Solution
Production Budget


Distribution of Production Budget for All Movies
1e−8
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
1e8
Production Budget in Millions

$68M

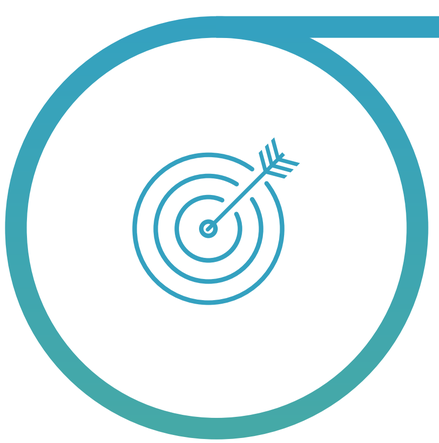


- The average production budget in the General Movies group is **$68M**.
- The average production budget in Top 100 Movies group is **$150M**.



$r^2 = 0.358$

- The average worldwide profit in General Movies group is **$179M**.
- The average worldwide profit of the Top 100 Movies group is **$624M**.

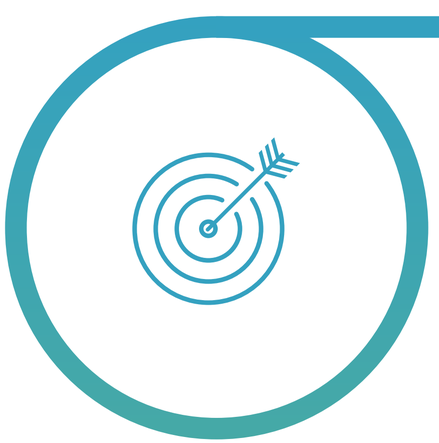


Examples

- ***Pirates of the Caribbean: On Stranger Tides*** (2011) - $379M budget, $1B earned

- ***Avenger: Age of Ultron*** (2015) - $365M budget, $1.3B earned

- ***Spectre*** (1997) - $300M budget, $879M earned

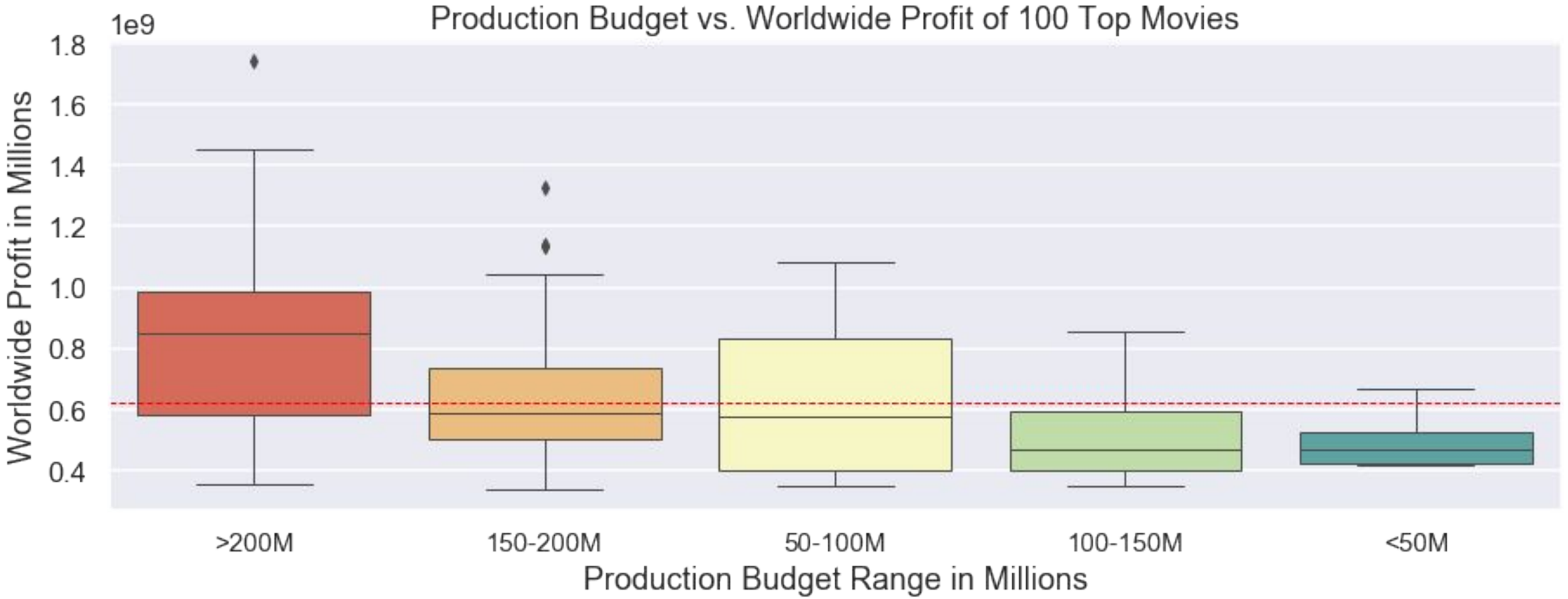


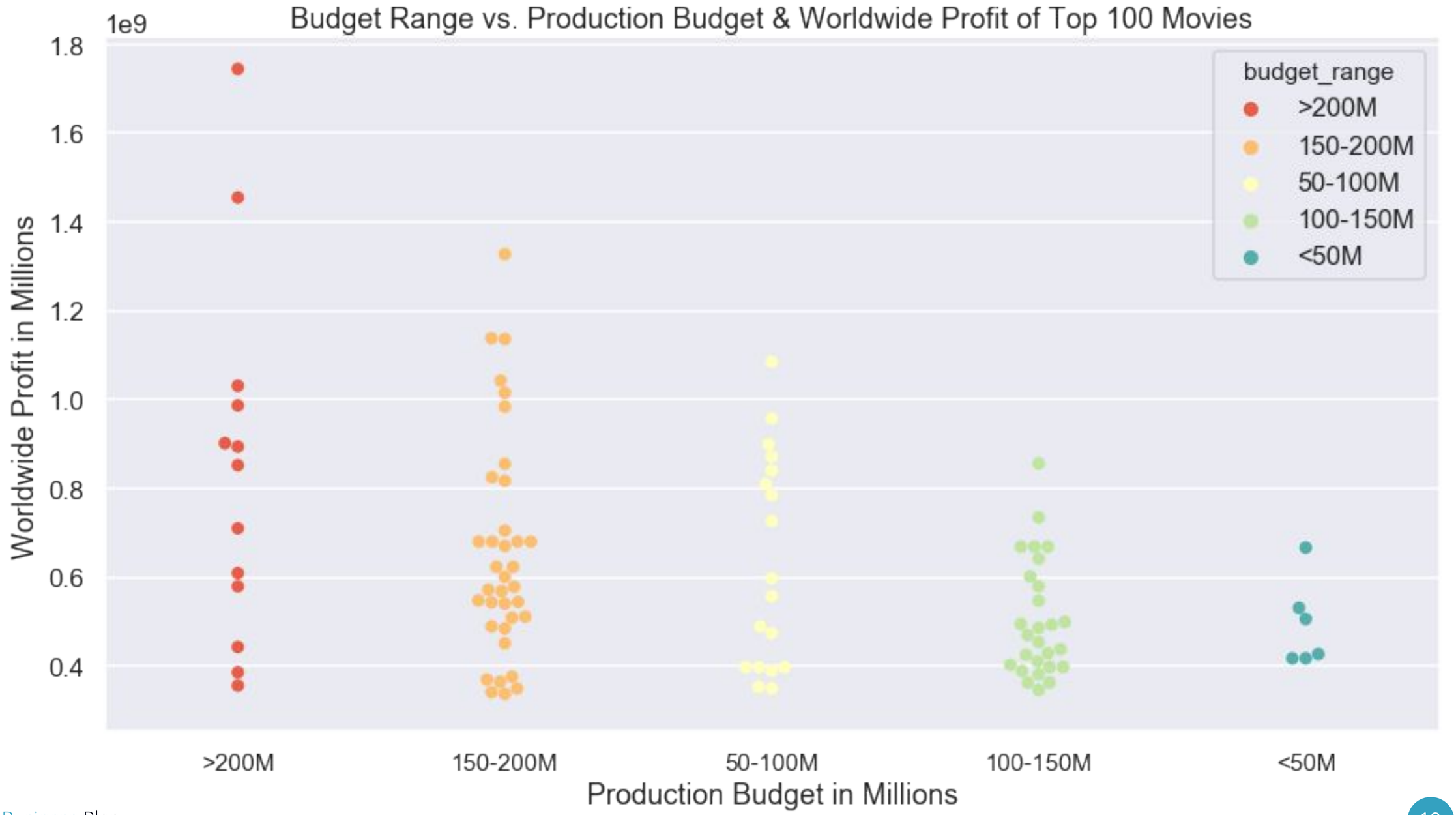
Examples

- ***Facing the Giants*** (2006) - $100,000 budget, $24.4M earned

- ***Once*** (2007) - $150,000 budget, Oscar nominations, $19M earned

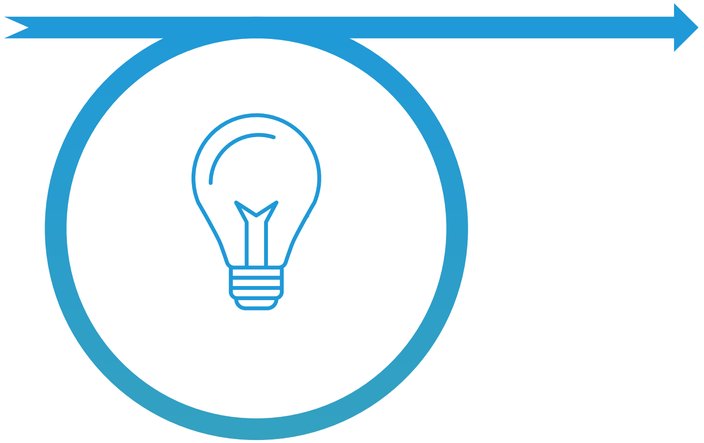
- ***Mad Max*** (1980), $200,000 budget, $49.6M earned


Production Budget vs. Worldwide Profit of 100 Top Movies
1e9
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
Worldwide Profit in Millions
>200M
150-200M
50-100M
100-150M
<50M
Production Budget Range in Millions


Budget Range vs. Production Budget & Worldwide Profit of Top 100 Movies
1e9
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
Worldwide Profit in Millions
budget_range
>200M
150-200M
50-100M
100-150M
<50M
>200M
150-200M
50-100M
100-150M
<50M
Production Budget in Millions

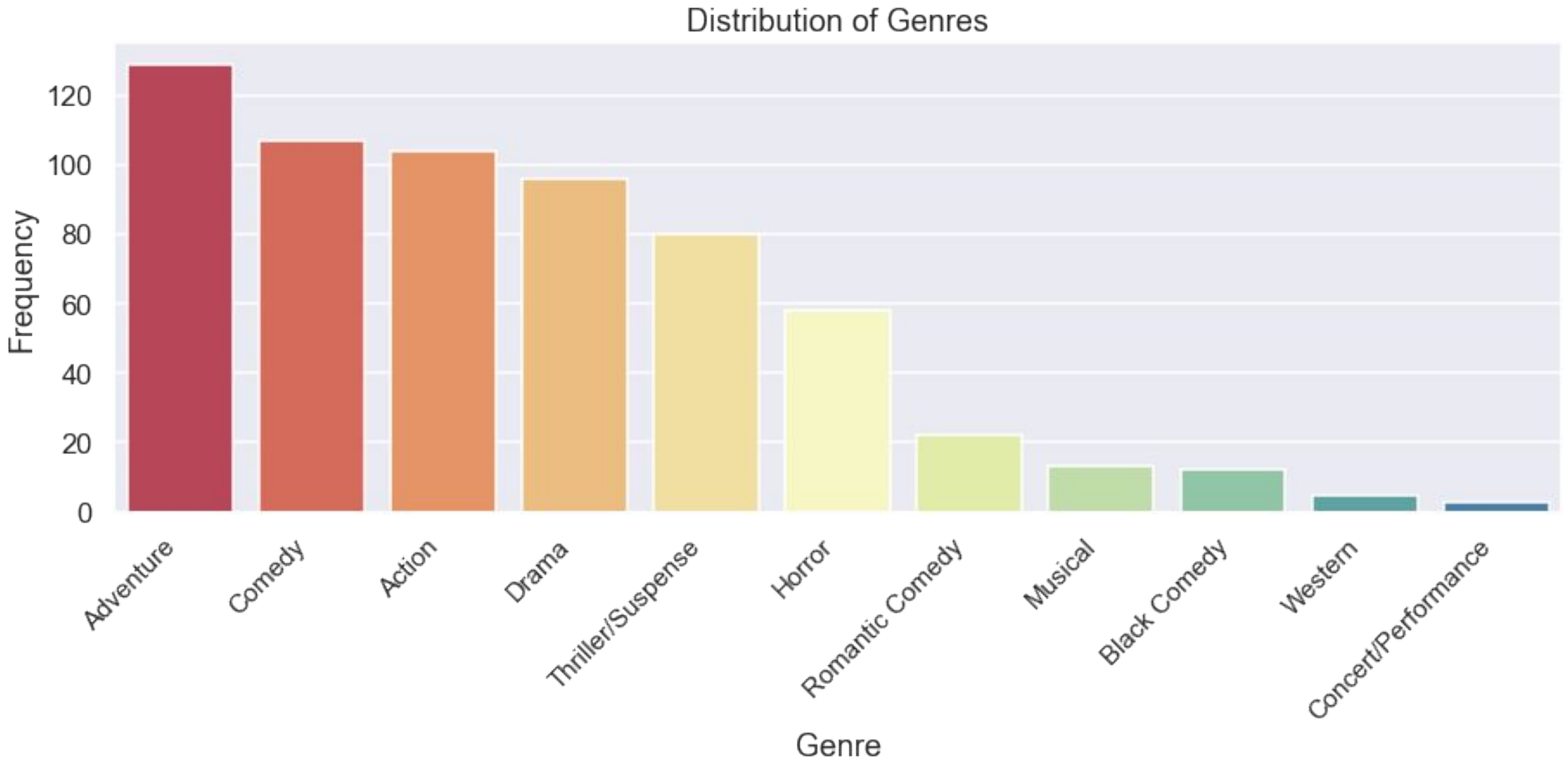
Conclusion

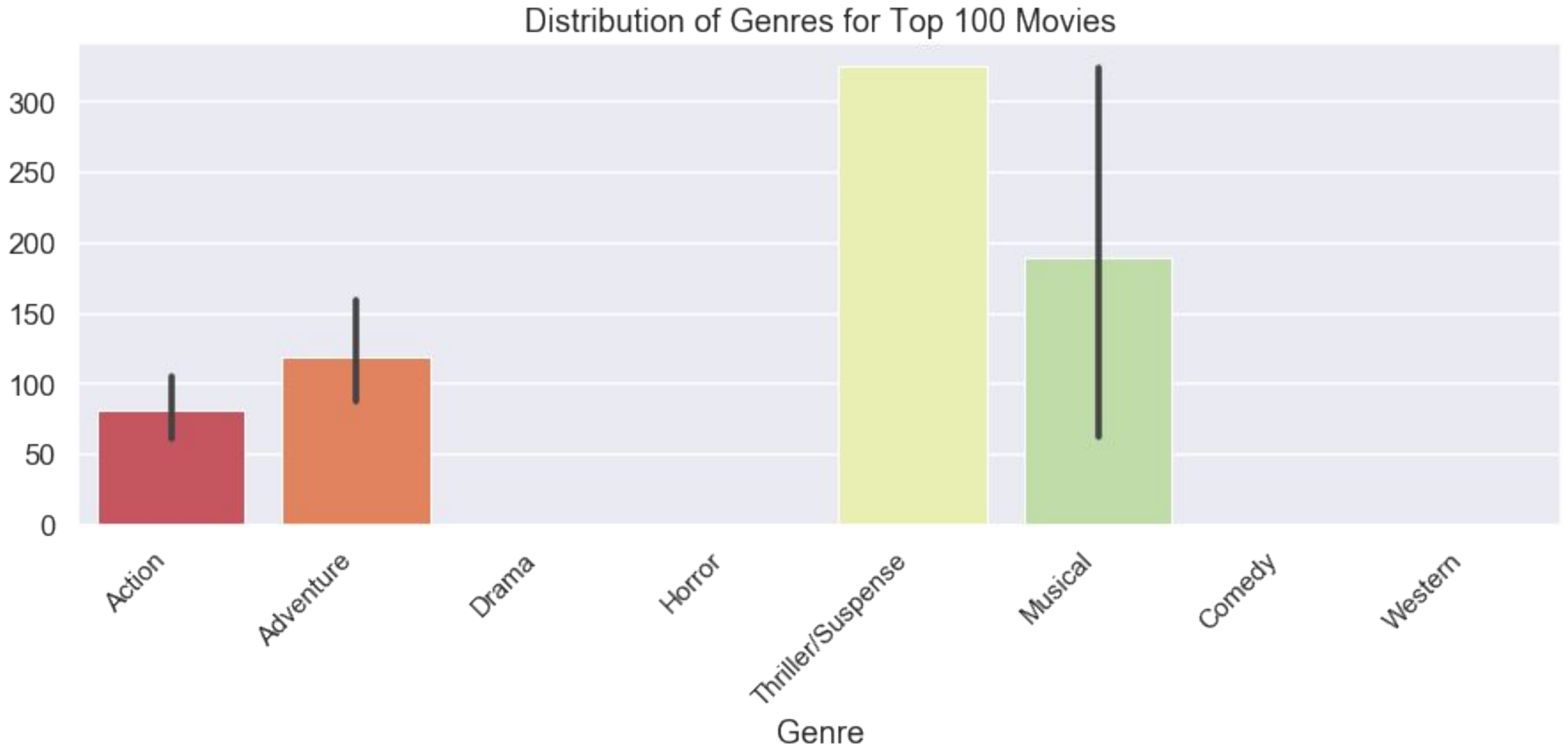
$150+ M

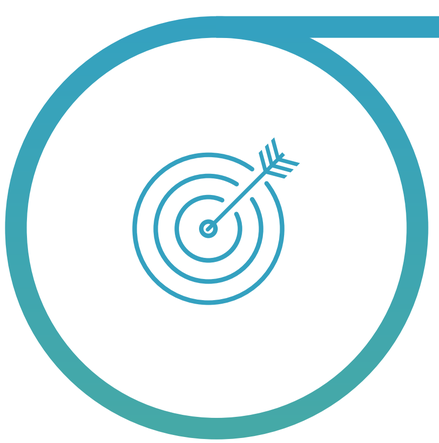


Business Solution

Genre

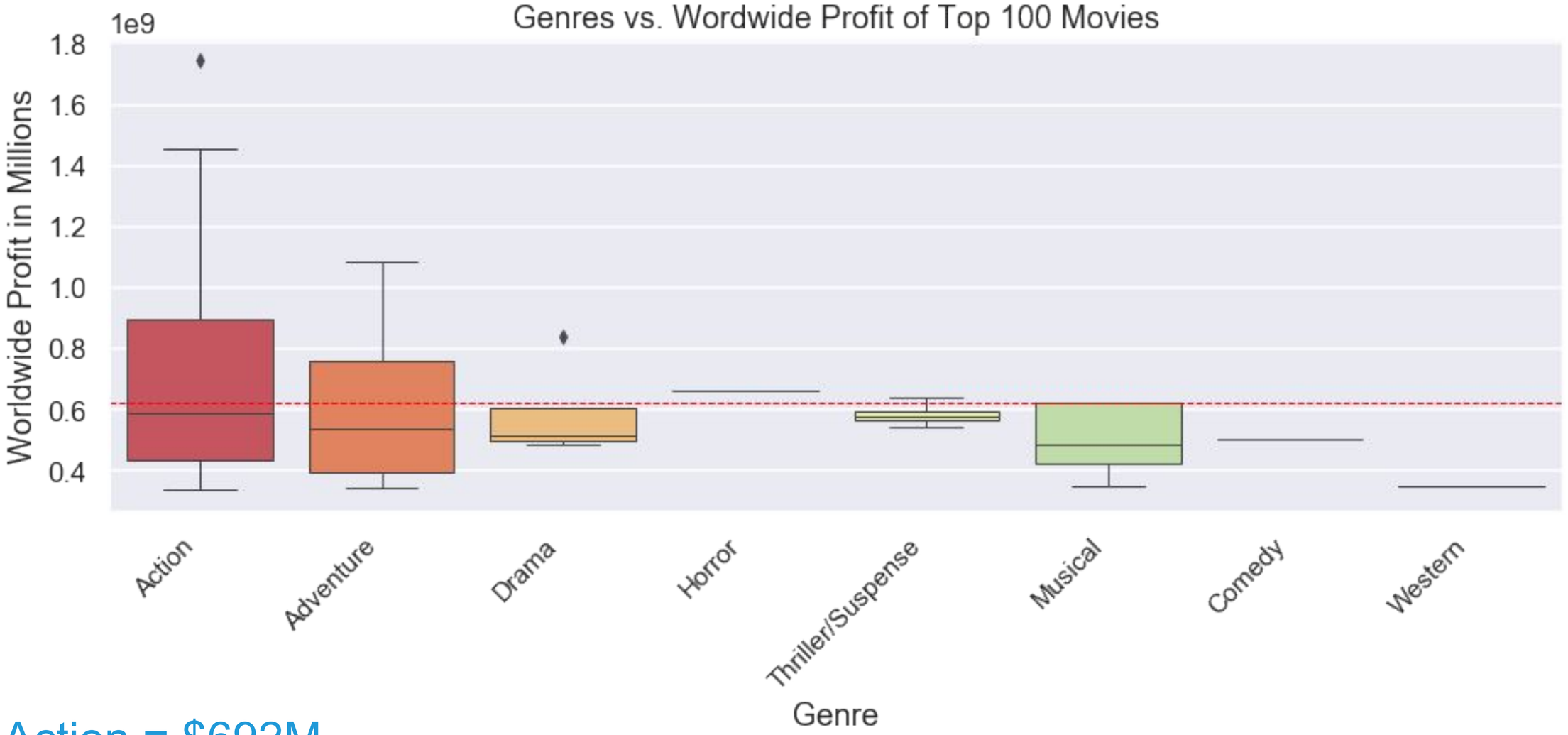

Distribution of Genres
Frequency
120
100
80
60
40
20
0
Adventure
Comedy
Action
Drama
Thriller/Suspense
Horror
Romantic Comedy
Musical
Black Comedy
Western
Concert/Performance
Genre


Distribution of Genres for Top 100 Movies
300
250
200
150
100
50
0
Action
Adventure
Drama
Horror
Thriller/Suspense
Musical
Comedy
Western
Genre

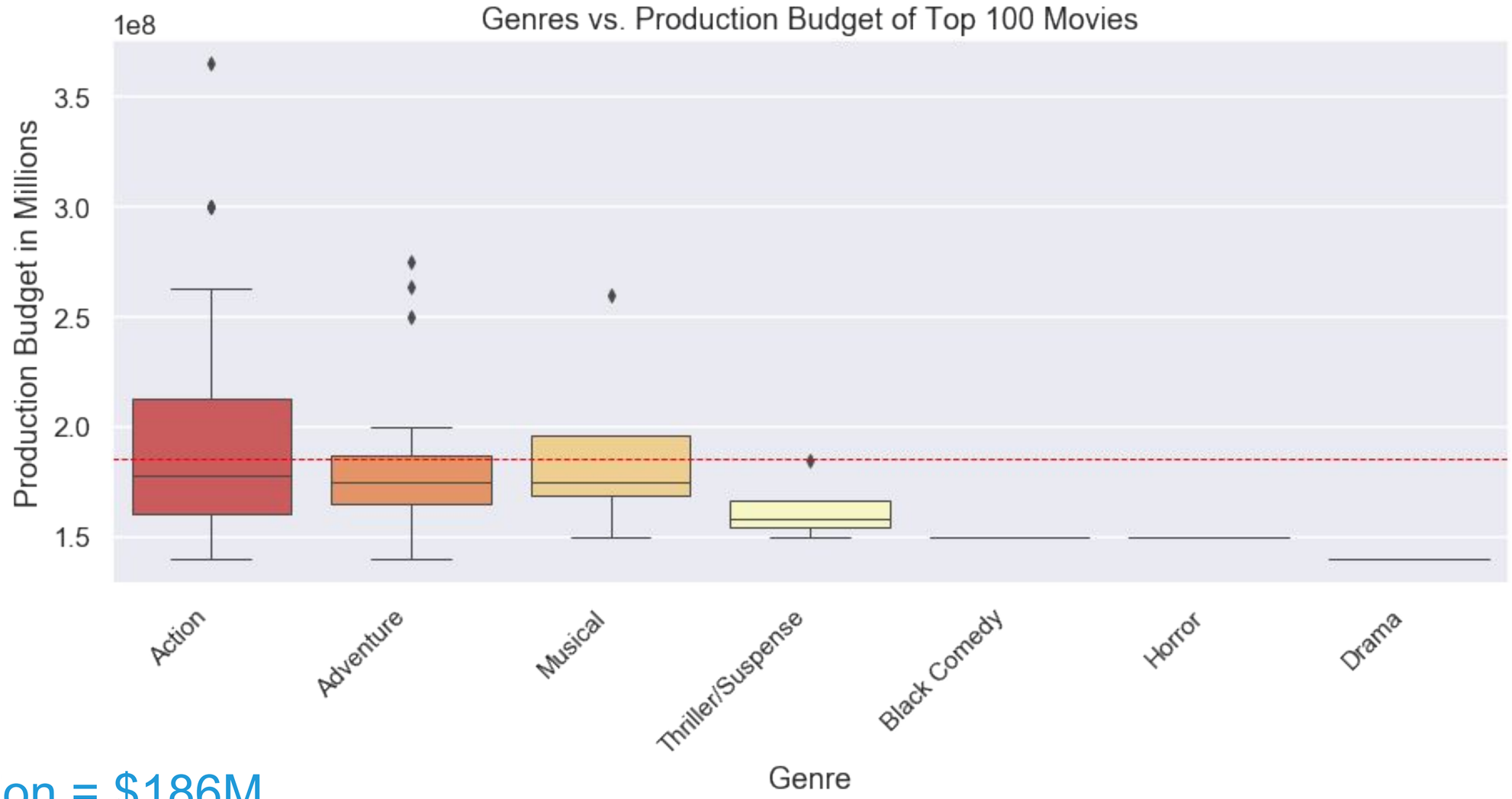


Examples

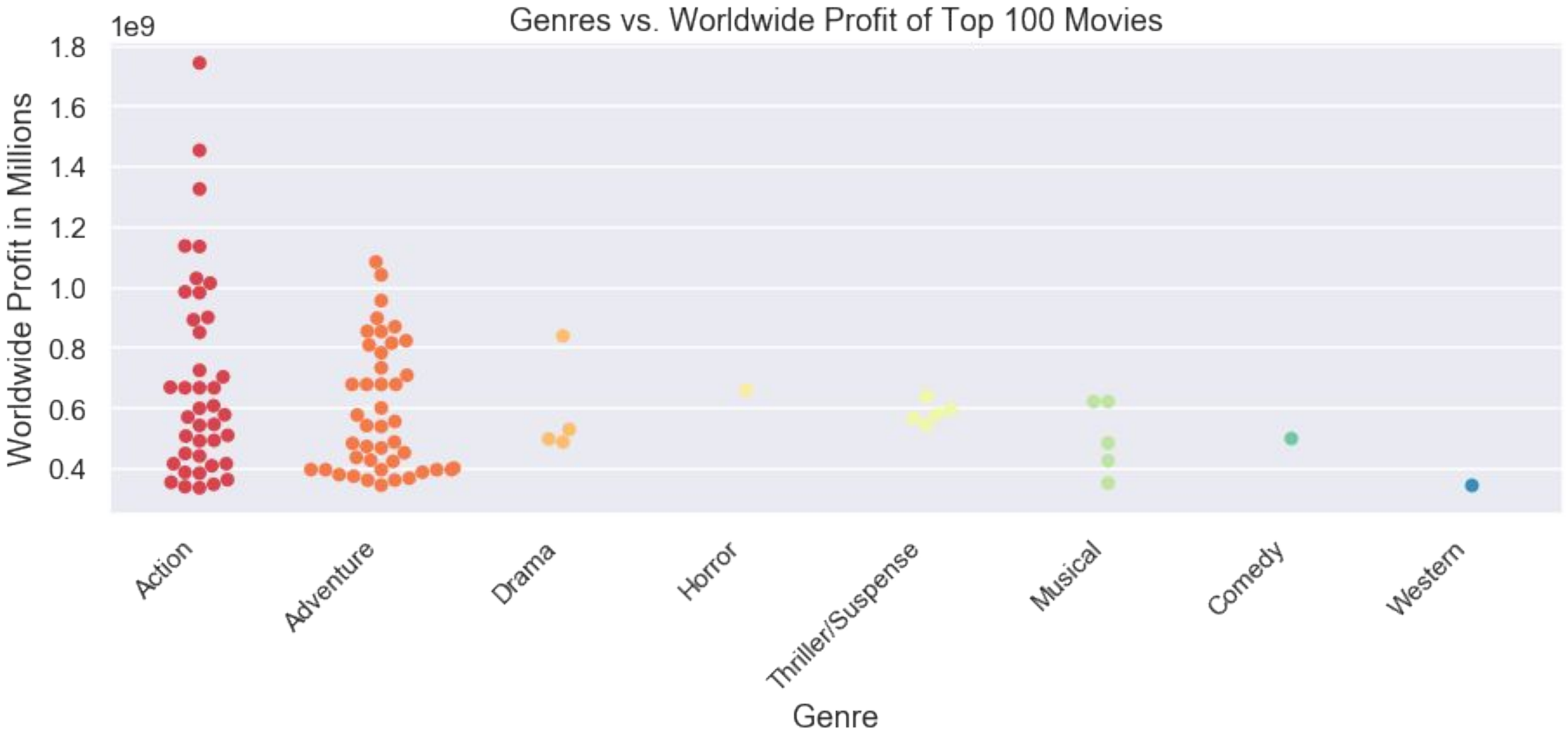
- ***Paranormal Activity*** (2007) — $15,000 budget, $162 millions profit, 539,336.30% ROI
- ***The Blair Witch Project*** (1994) — 20,591% ROI
- ***Night of the Living Dead*** (1990) — 13,057.89% ROI

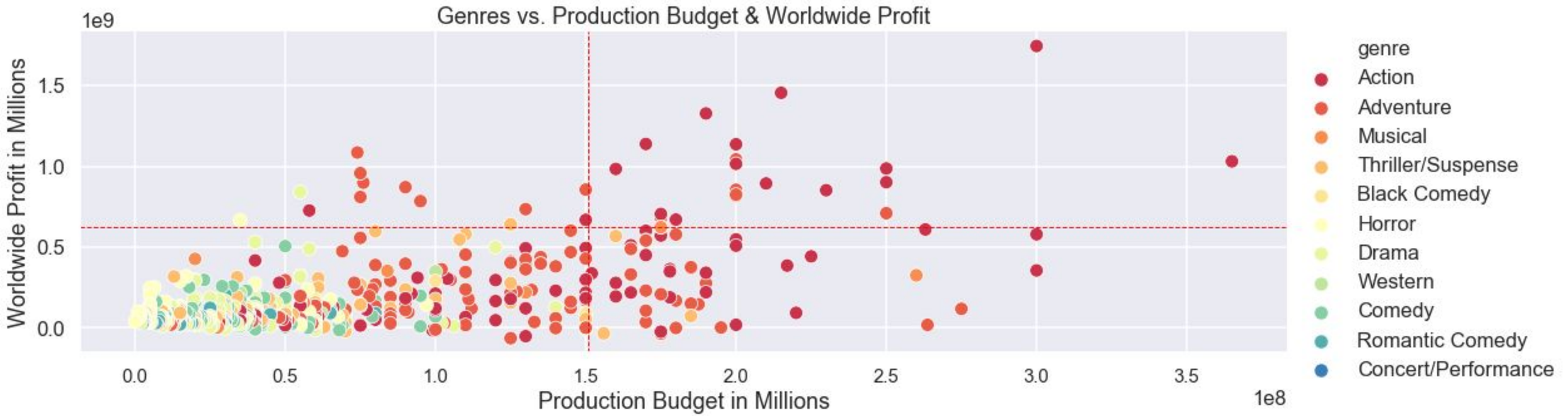


Action = $692M
Adventure = $591M
Thriller/Suspense = $587M



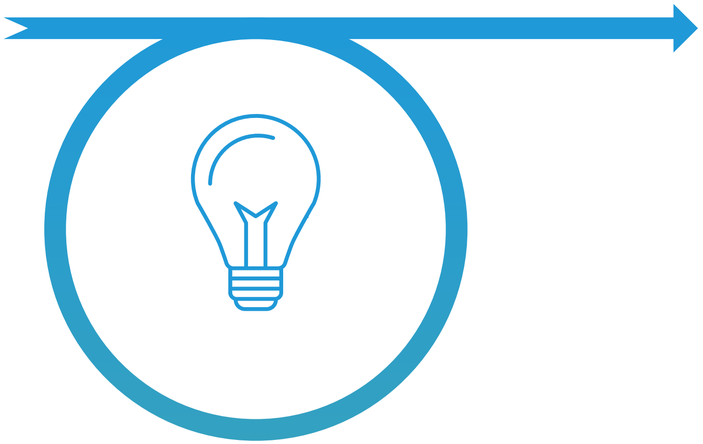
Action = $186M
Adventure = $138M
Thriller/Suspense = $116M


Genres vs. Worldwide Profit of Top 100 Movies
1e9
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
Worldwide Profit in Millions
Action
Adventure
Drama
Horror
Thriller/Suspense
Musical
Comedy
Western
Genre


Genres vs. Production Budget & Worldwide Profit
1e9
Worldwide Profit in Millions
1.5
1.0
0.5
0.0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
1e8
Production Budget in Millions
genre
Action
Adventure
Musical
Thriller/Suspense
Black Comedy
Horror
Drama
Western
Comedy
Romantic Comedy
Concert/Performance

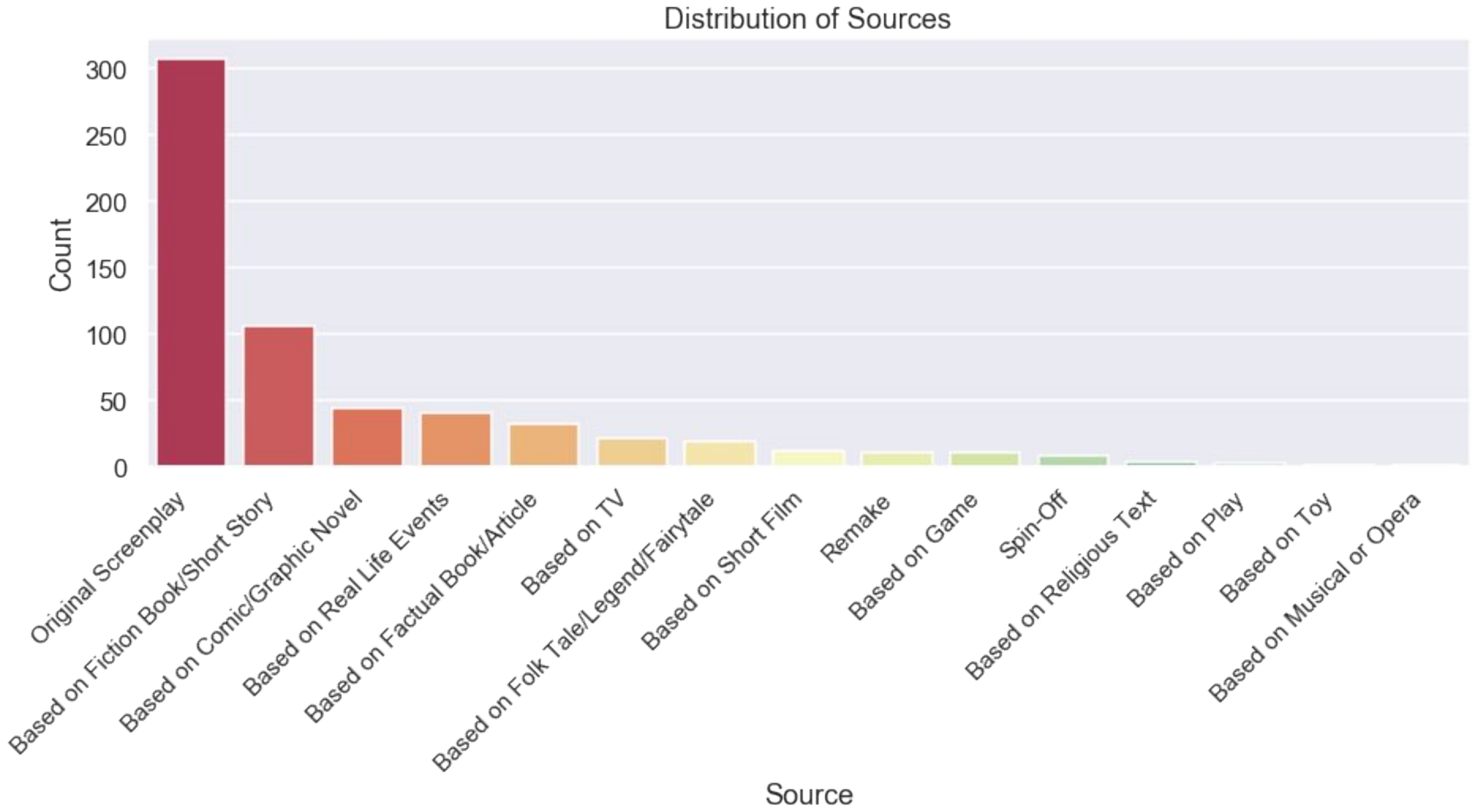
Conclusion

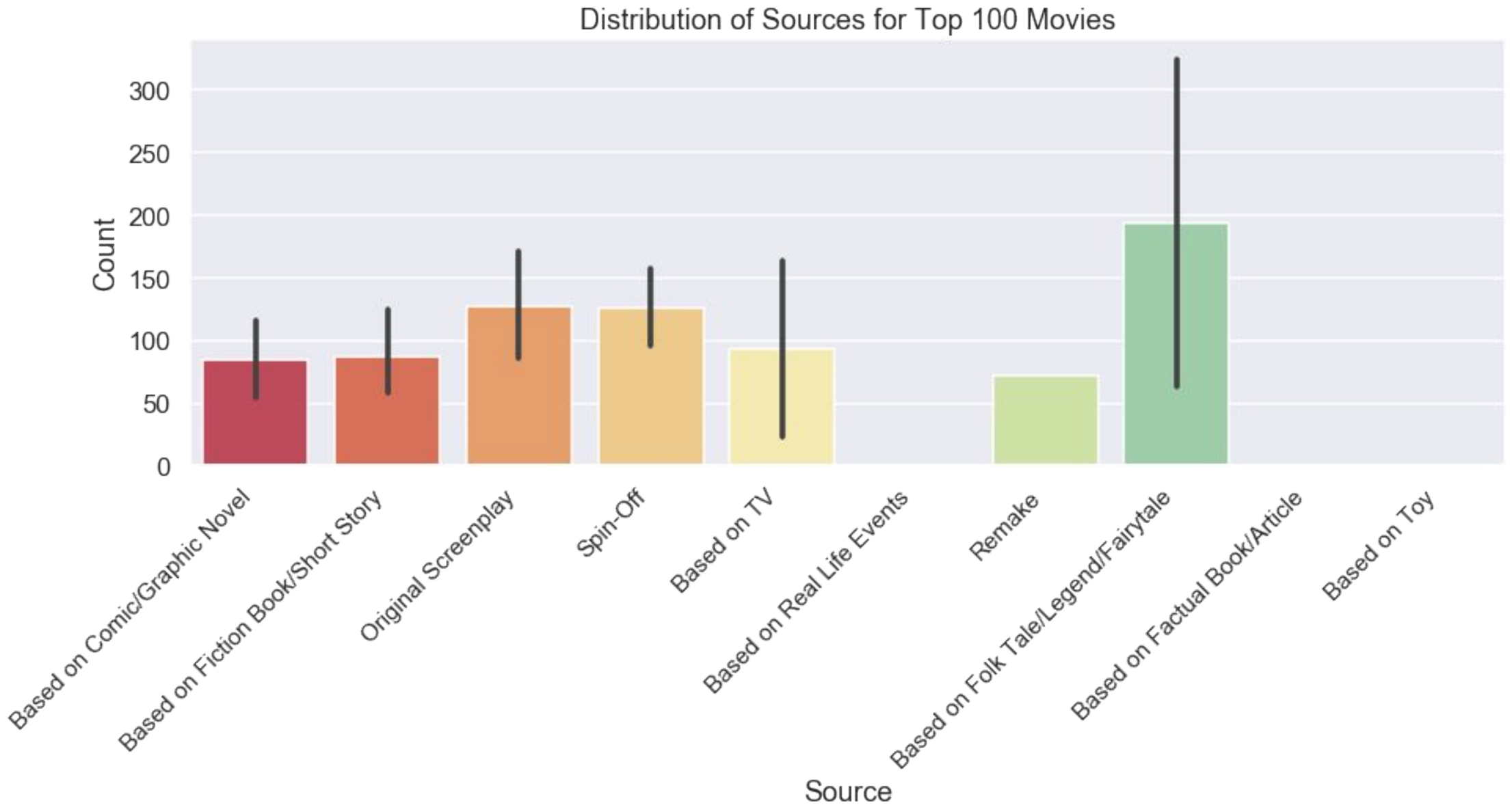
Action Adventure

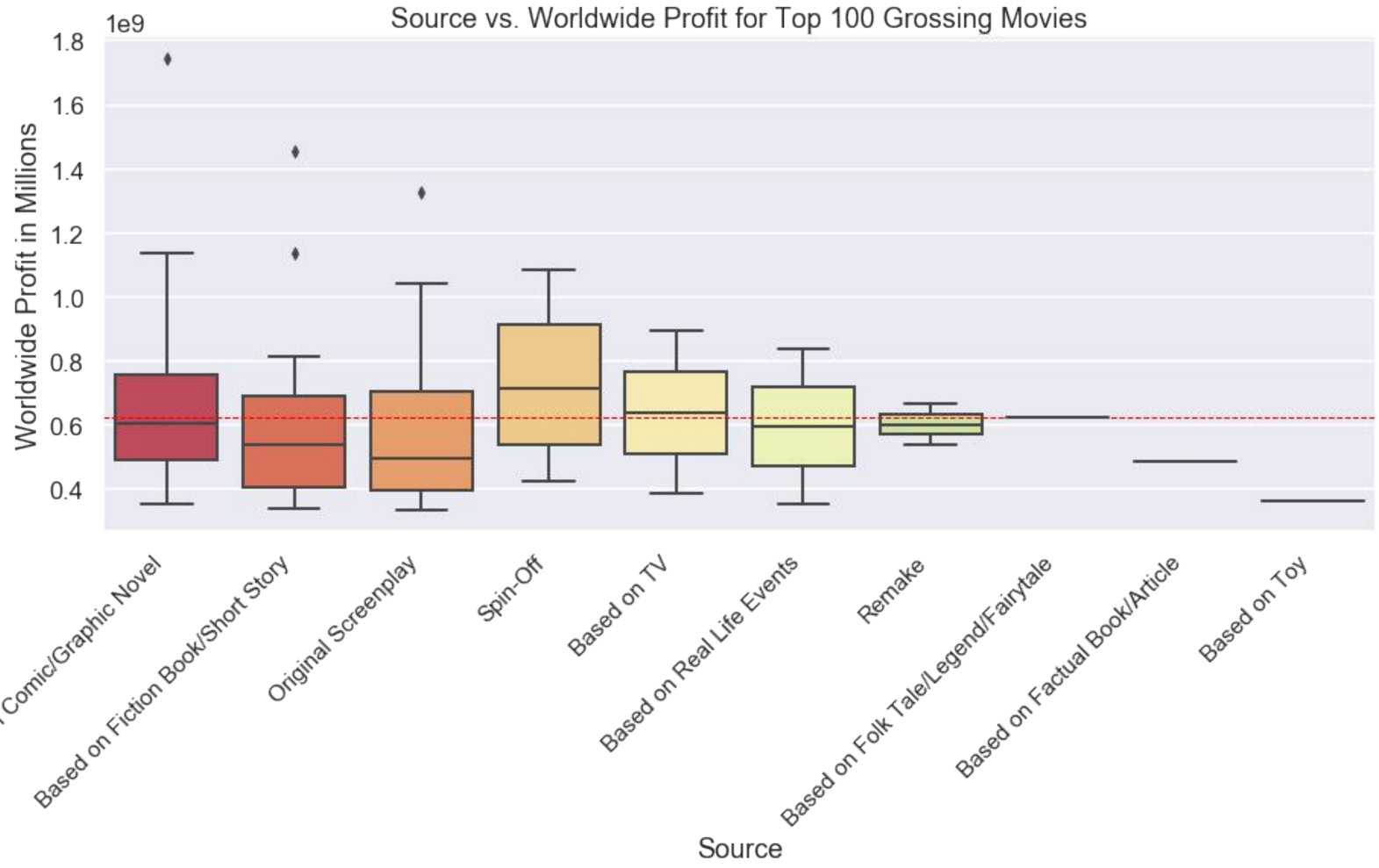


Business Solution

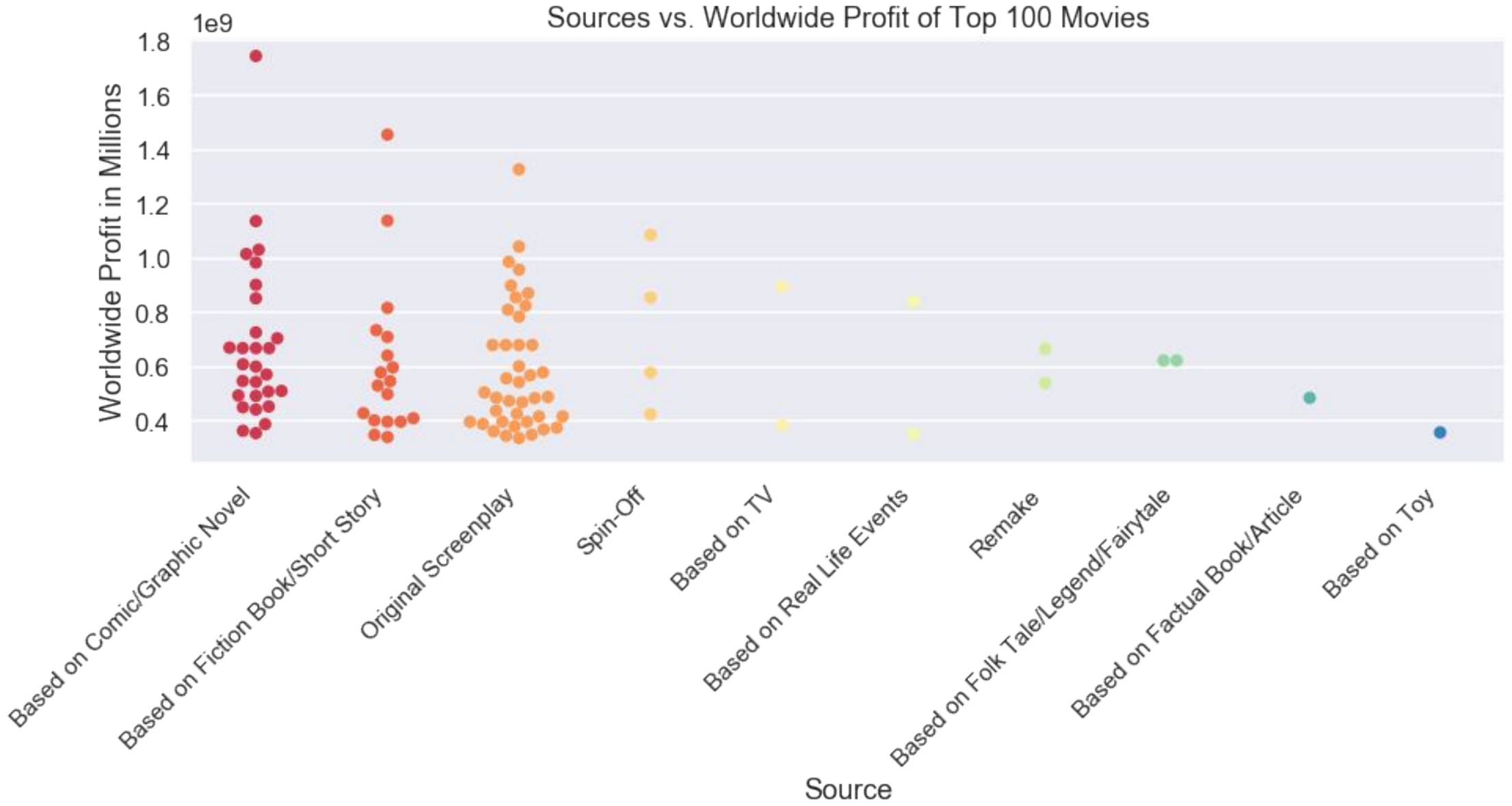
Source

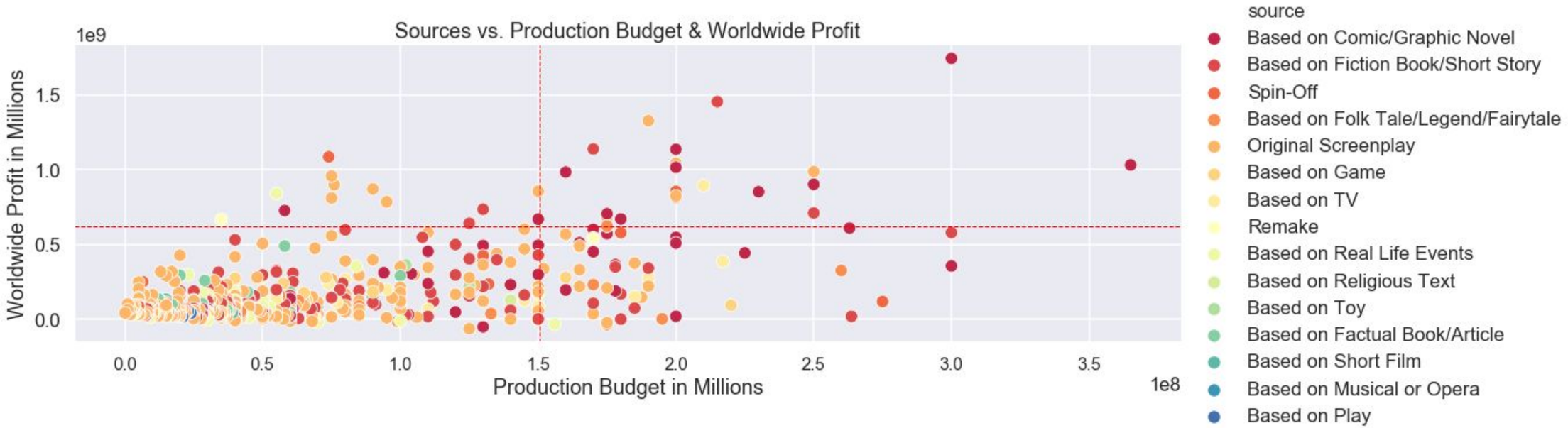

Distribution of Sources
Count
300
250
200
150
100
50
0
Original Screenplay
Based on Fiction Book/Short Story
Based on Comic/Graphic Novel
Based on Real Life Events
Based on Factual Book/Article
Based on TV
Based on Folk Tale/Legend/Fairytale
Based on Short Film
Remake
Based on Game
Spin-Off
Based on Religious Text
Based on Play
Based on Toy
Based on Musical or Opera
Source


Distribution of Sources for Top 100 Movies
Count
300
250
200
150
100
50
0
Based on Comic/Graphic Novel
Based on Fiction Book/Short Story
Original Screenplay
Spin-Off
Based on TV
Based on Real Life Events
Remake
Based on Folk Tale/Legend/Fairytale
Based on Factual Book/Article
Based on Toy
Source



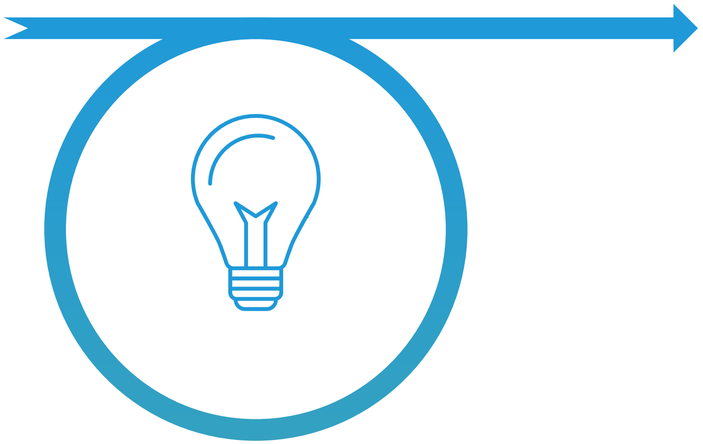
Comic = $682M
Fiction = $610
Screenplay = $591M


Sources vs. Worldwide Profit of Top 100 Movies
1e9
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
Worldwide Profit in Millions
Based on Comic/Graphic Novel
Based on Fiction Book/Short Story
Original Screenplay
Spin-Off
Based on TV
Based on Real Life Events
Remake
Based on Folk Tale/Legend/Fairytale
Based on Factual Book/Article
Based on Toy
Source


Sources vs. Production Budget & Worldwide Profit
1e9
Worldwide Profit in Millions
1.5
1.0
0.5
0.0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
1e8
Production Budget in Millions
source
Based on Comic/Graphic Novel
Based on Fiction Book/Short Story
Spin-Off
Based on Folk Tale/Legend/Fairytale
Original Screenplay
Based on Game
Based on TV
Remake
Based on Real Life Events
Based on Religious Text
Based on Toy
Based on Factual Book/Article
Based on Short Film
Based on Musical or Opera
Based on Play

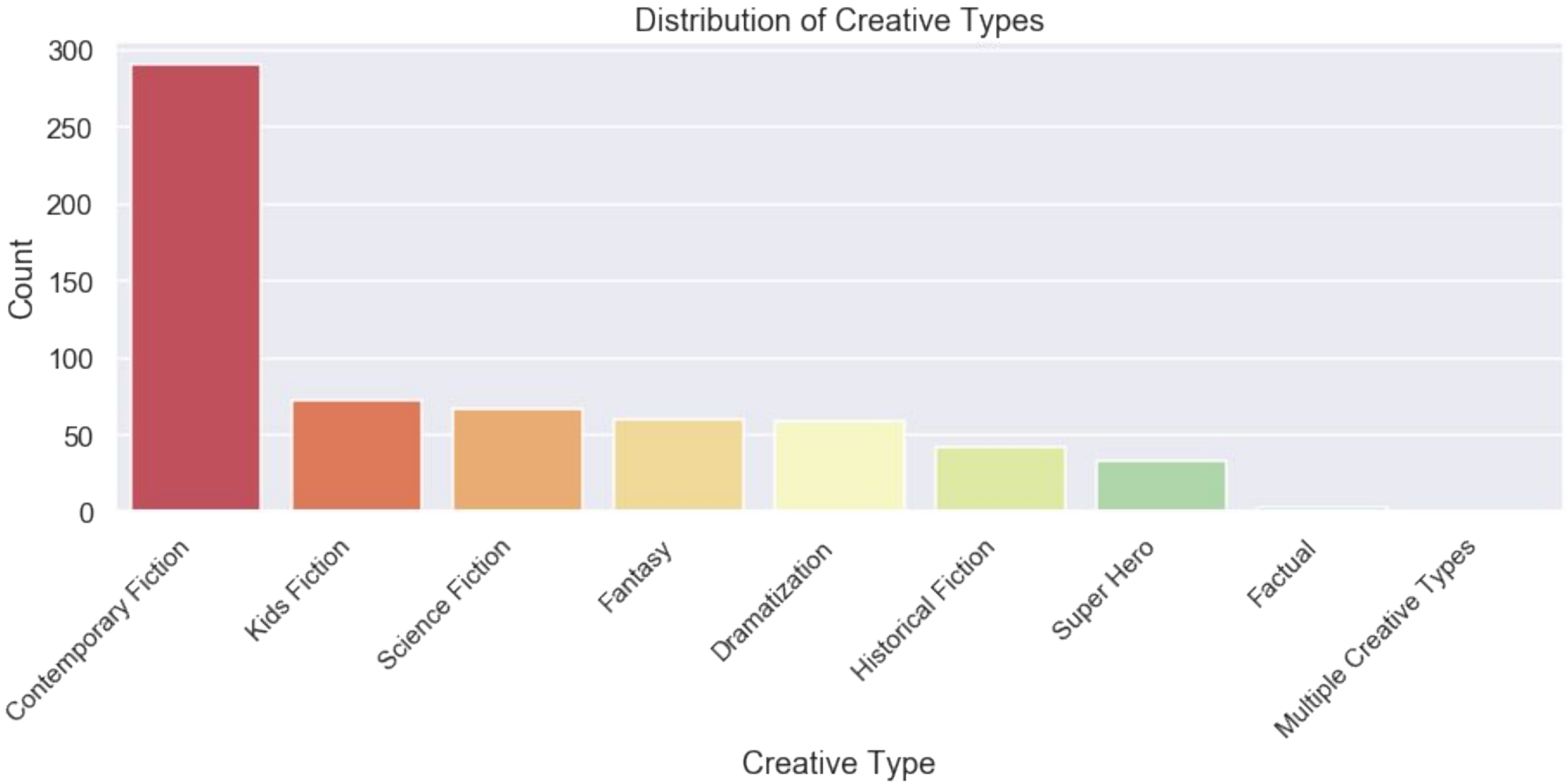
Conclusion

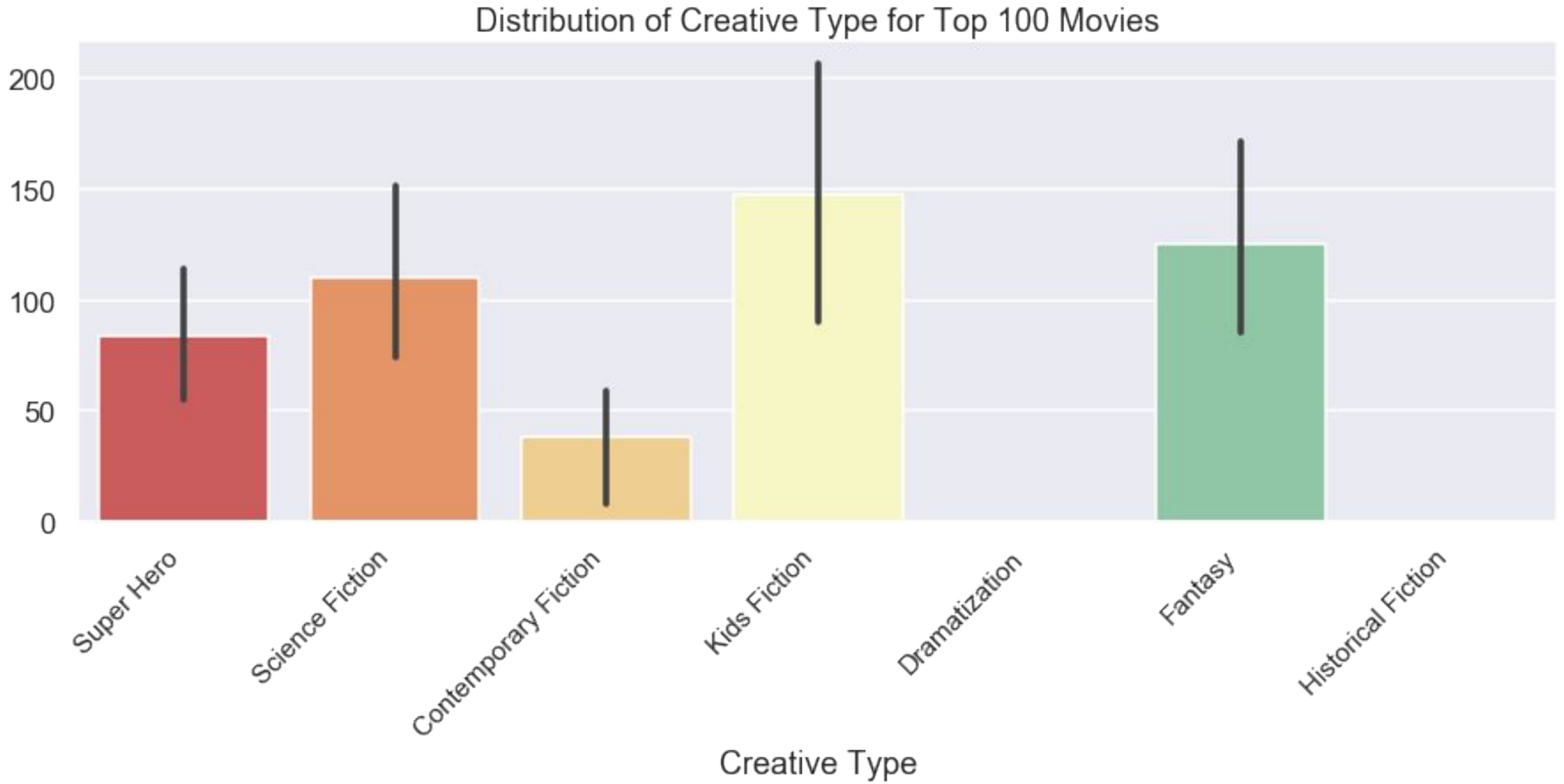
Comic Graphic/ Novel

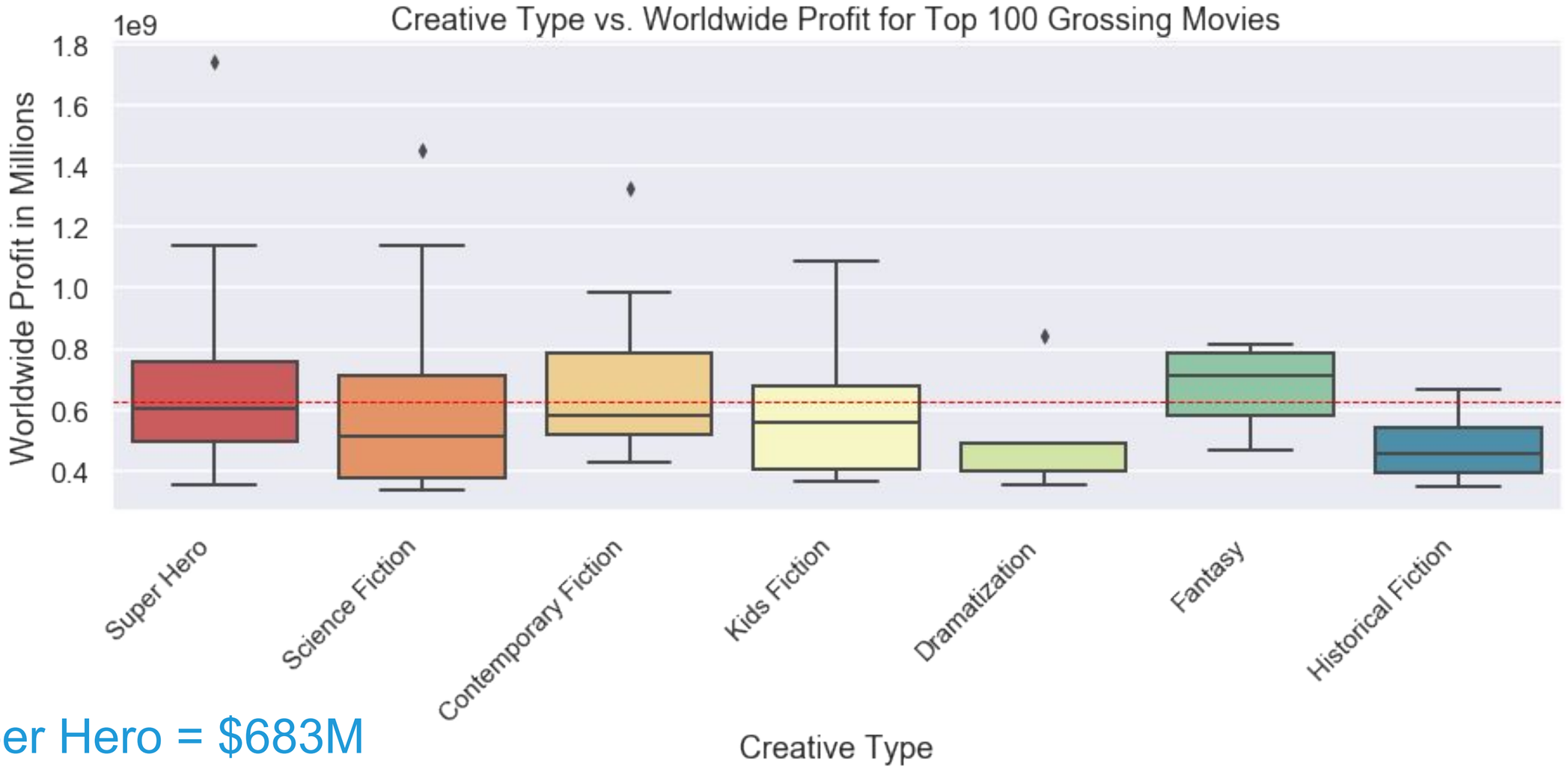


Business Solution

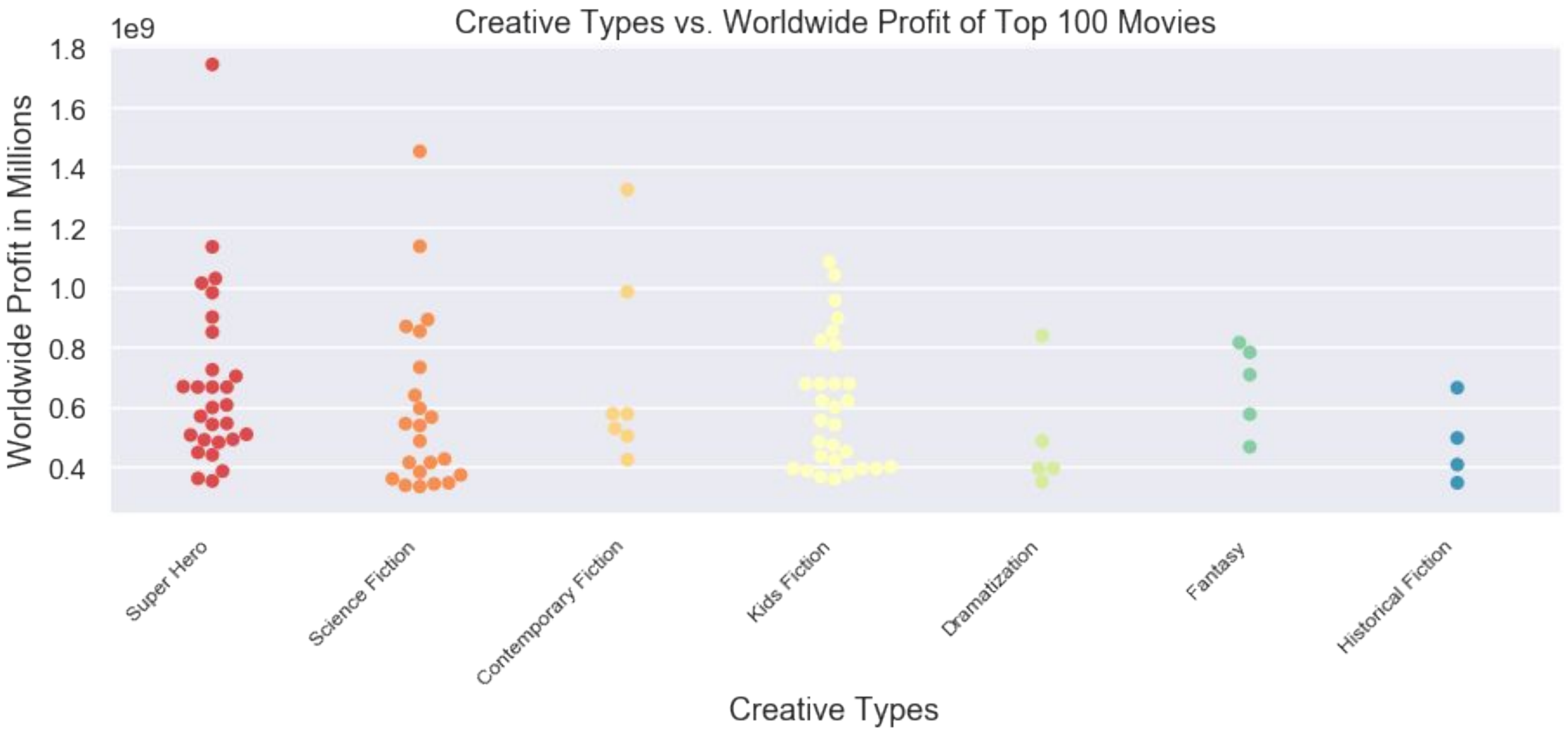
Creative Type

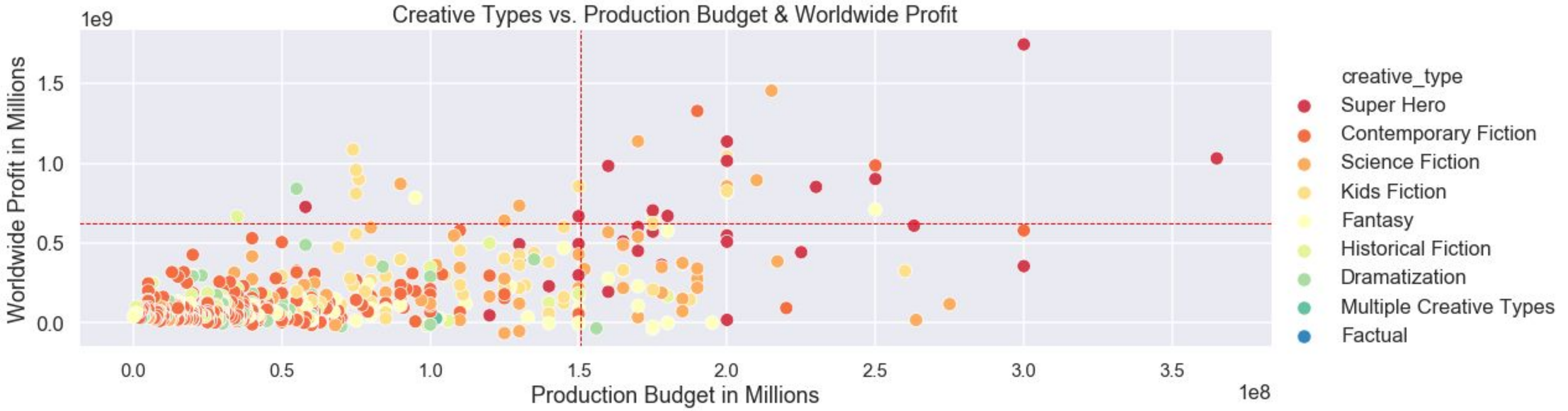

Distribution of Creative Types
Count
300
250
200
150
100
50
0
Contemporary Fiction
Kids Fiction
Science Fiction
Fantasy
Dramatization
Historical Fiction
Super Hero
Factual
Multiple Creative Types
Creative Type


Distribution of Creative Type for Top 100 Movies
200
150
100
50
0
Super Hero
Science Fiction
Contemporary Fiction
Kids Fiction
Dramatization
Fantasy
Historical Fiction
Creative Type



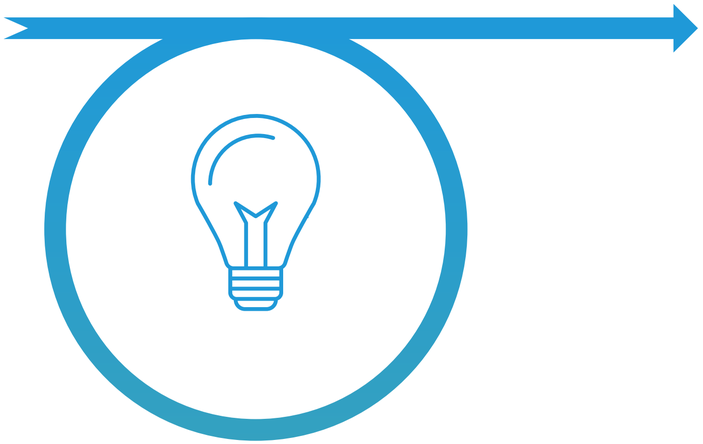
Super Hero = $683M
Science Fiction = $595M
Contemporary Fiction = $705M


Creative Types vs. Worldwide Profit of Top 100 Movies
1e9
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
Worldwide Profit in Millions
Super Hero
Science Fiction
Contemporary Fiction
Kids Fiction
Dramatization
Fantasy
Historical Fiction
Creative Types

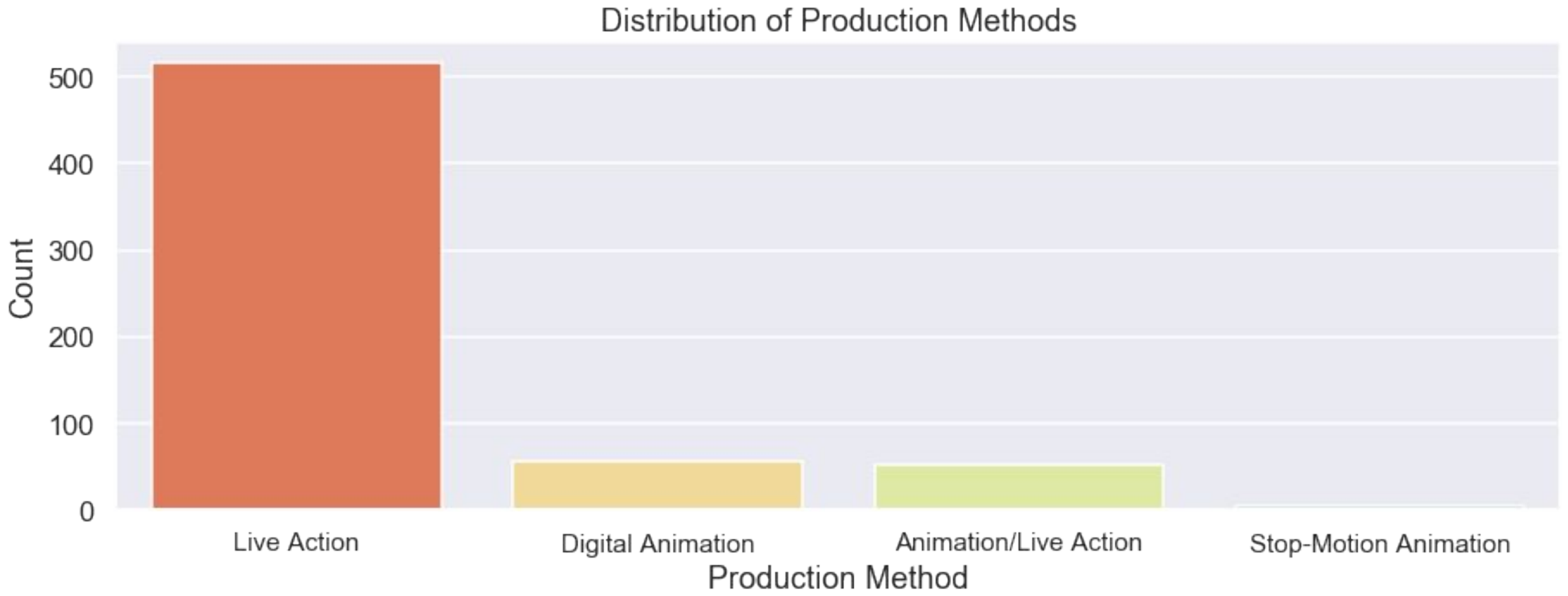

Creative Types vs. Production Budget & Worldwide Profit
1e9
Worldwide Profit in Millions
1.5
1.0
0.5
0.0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
1e8
Production Budget in Millions
creative_type
Super Hero
Contemporary Fiction
Science Fiction
Kids Fiction
Fantasy
Historical Fiction
Dramatization
Multiple Creative Types
Factual

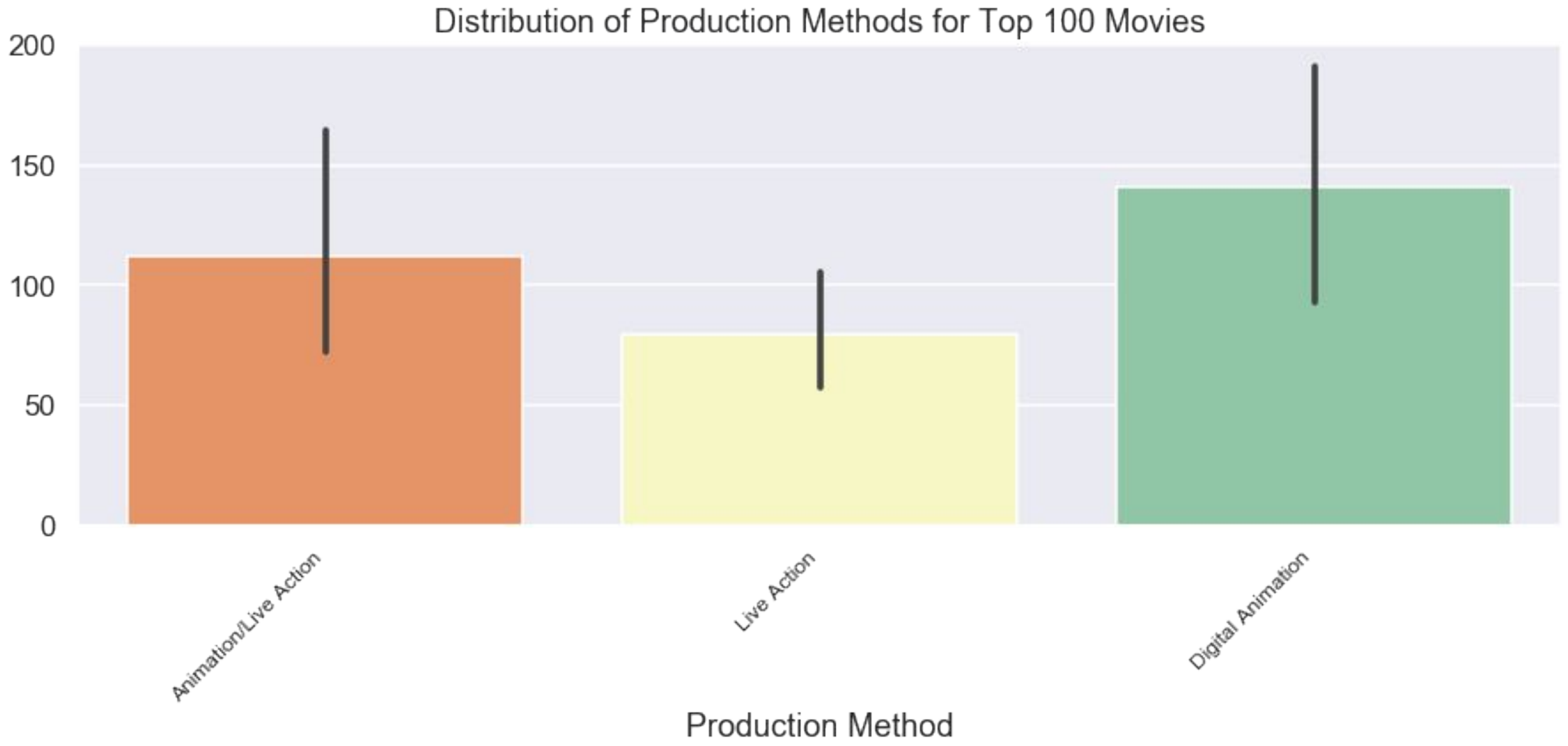
Conclusion

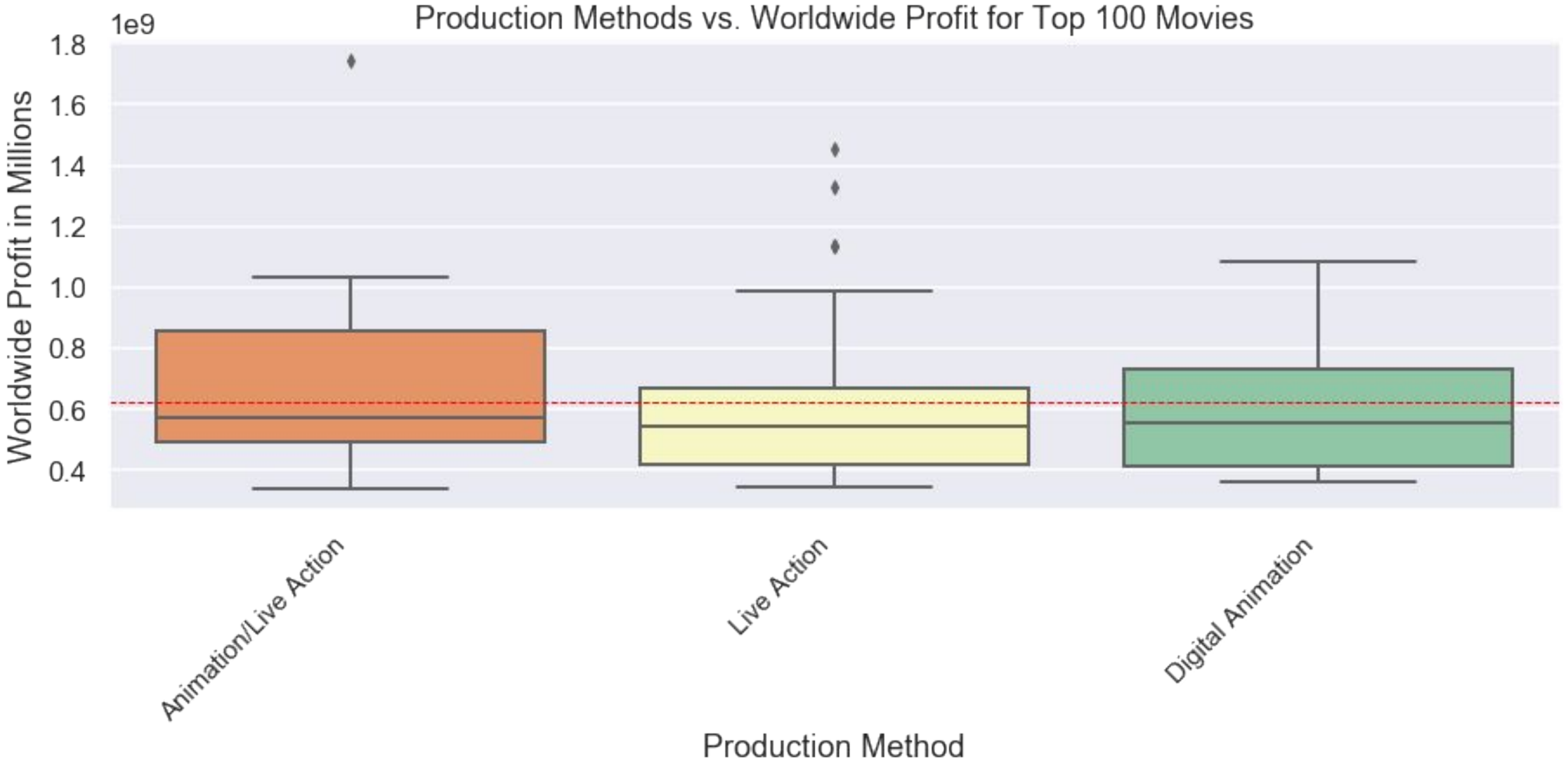
Super Hero



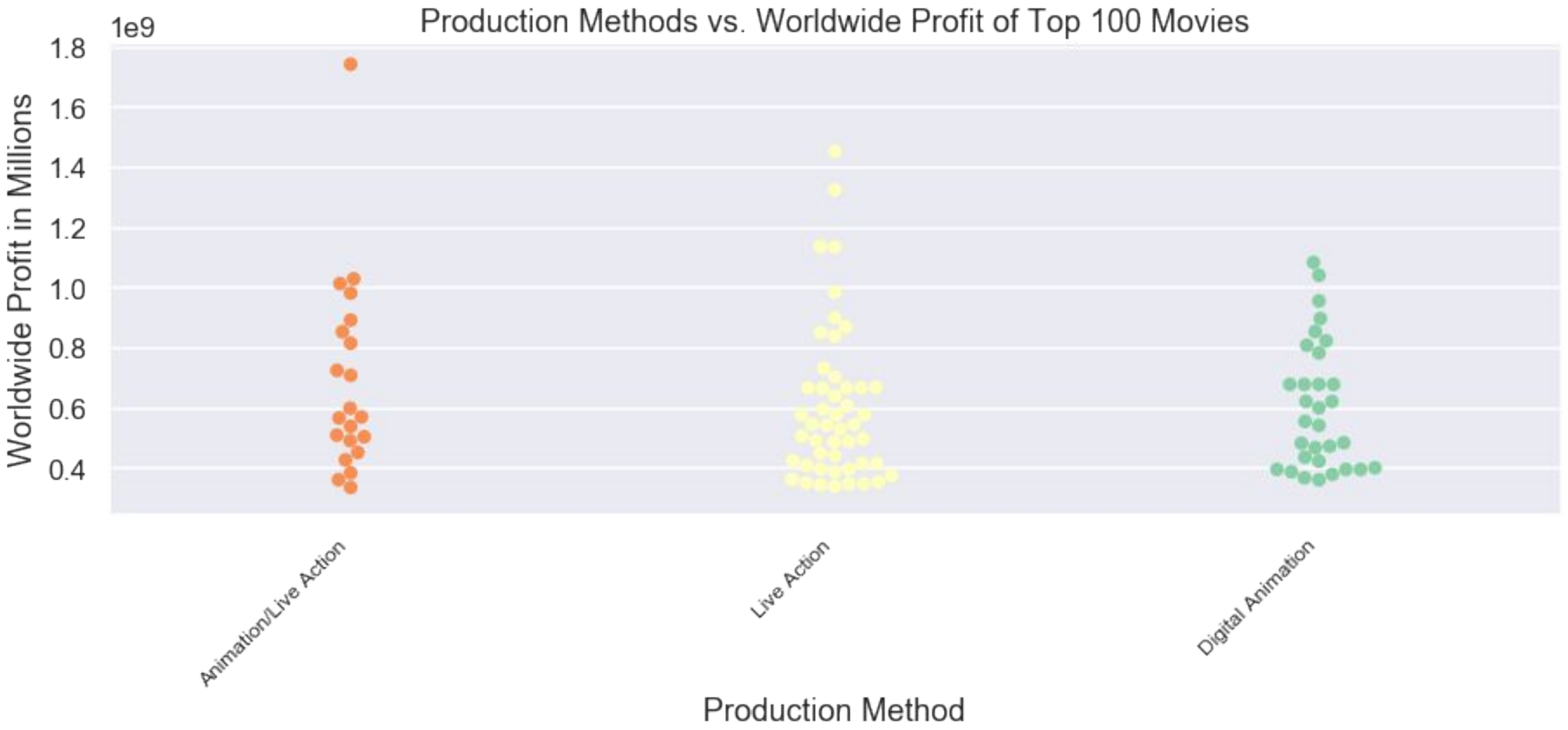
Business Solution
Production Method

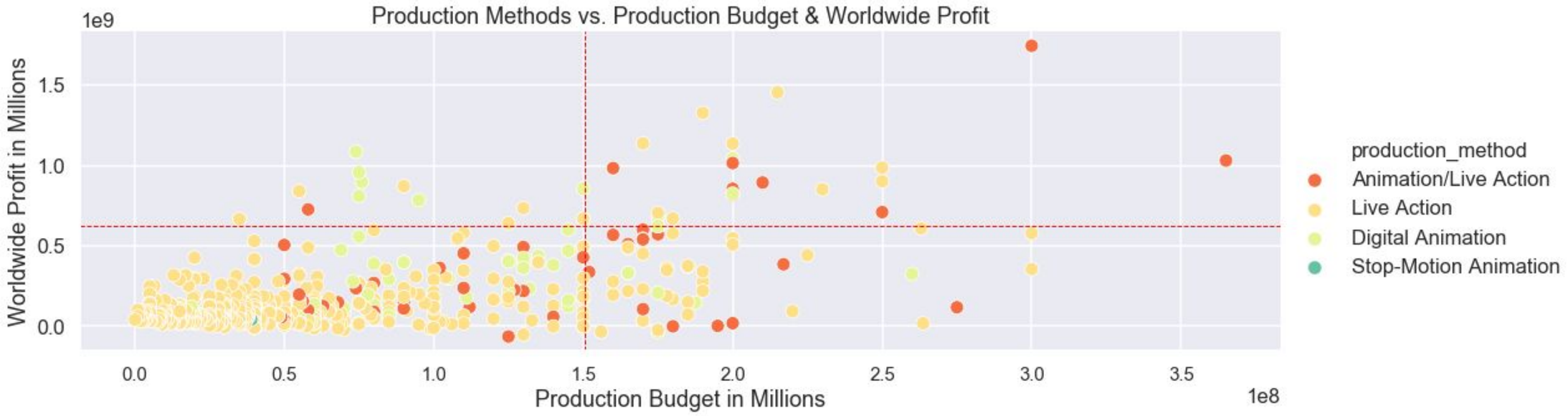

Distribution of Production Methods
Count
500
400
300
200
100
0
Live Action
Digital Animation
Animation/Live Action
Stop-Motion Animation
Production Method


Distribution of Production Methods for Top 100 Movies
200
150
100
50
0
Animation/Live Action
Live Action
Digital Animation
Production Method



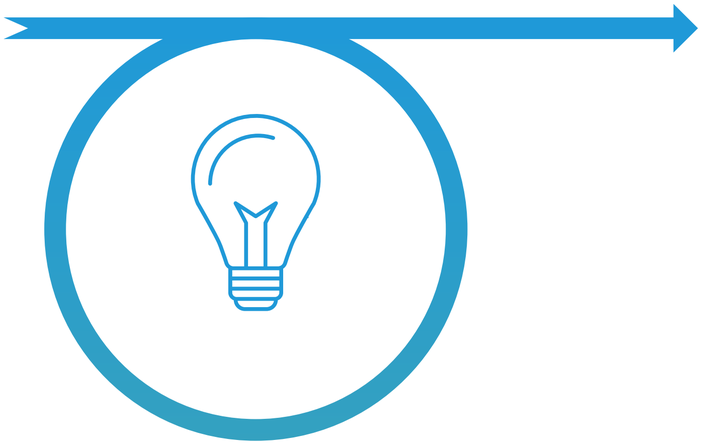
Animation/ Live Action = $692M
Live Action = $606M
Digital Animation = $606M


Production Methods vs. Worldwide Profit of Top 100 Movies
1e9
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
Worldwide Profit in Millions
Animation/Live Action
Live Action
Digital Animation
Production Method

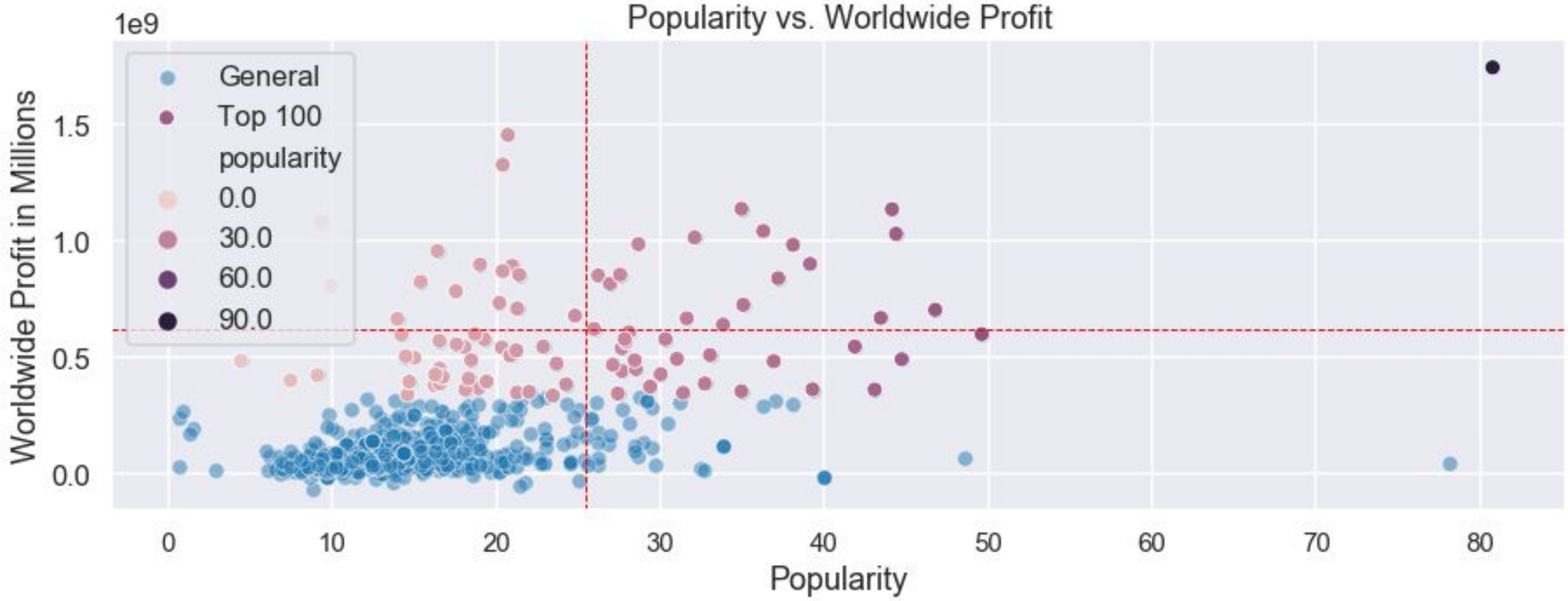

Production Methods vs. Production Budget & Worldwide Profit
1e9
Worldwide Profit in Millions
1.5
1.0
0.5
0.0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
1e8
Production Budget in Millions
production_method
Animation/Live Action
Live Action
Digital Animation
Stop-Motion Animation

Conclusion

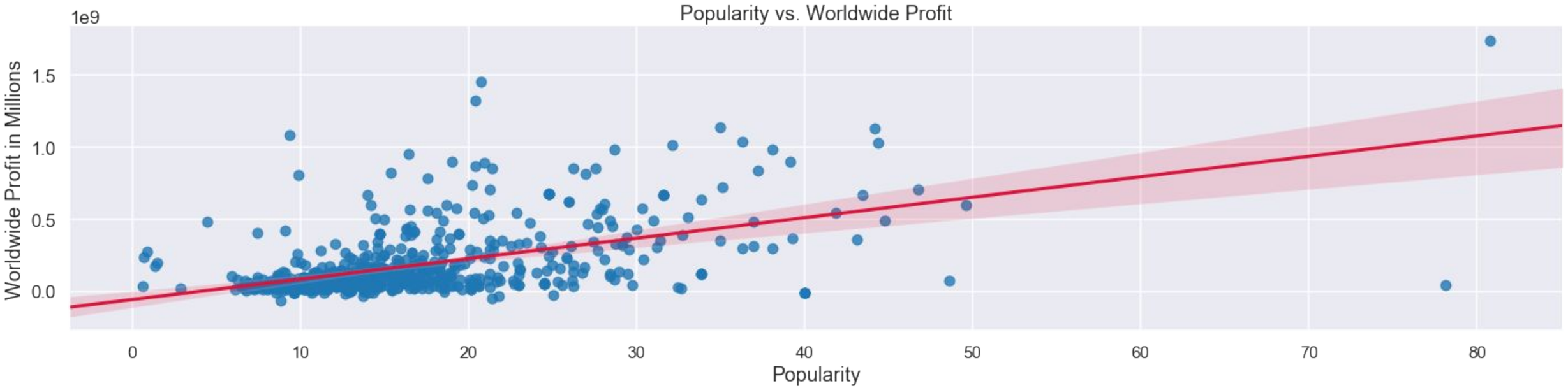
Animation/ Live Action



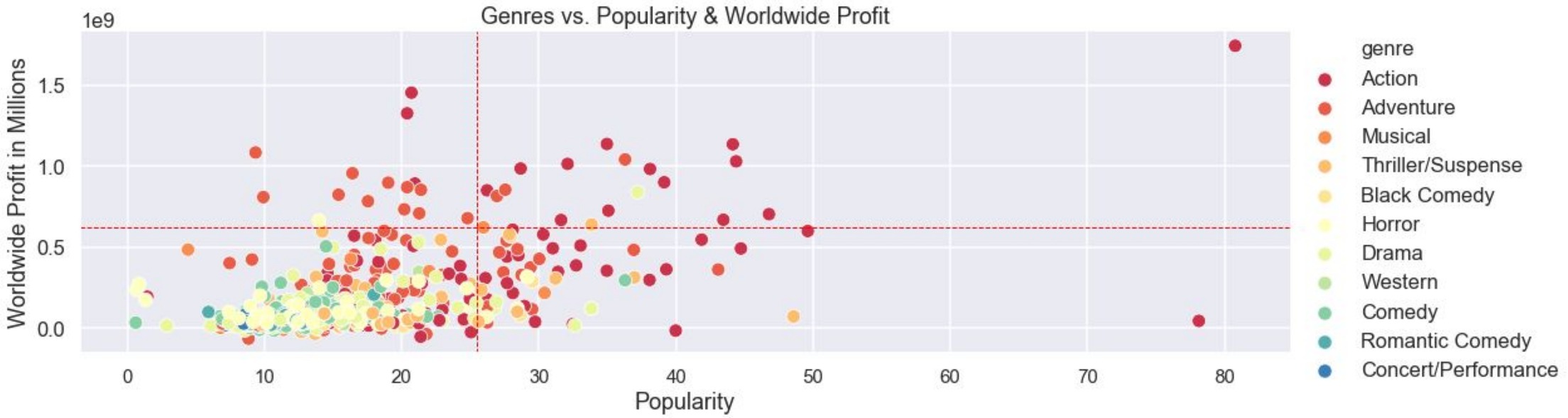
Business Solution
Popularity & Rating

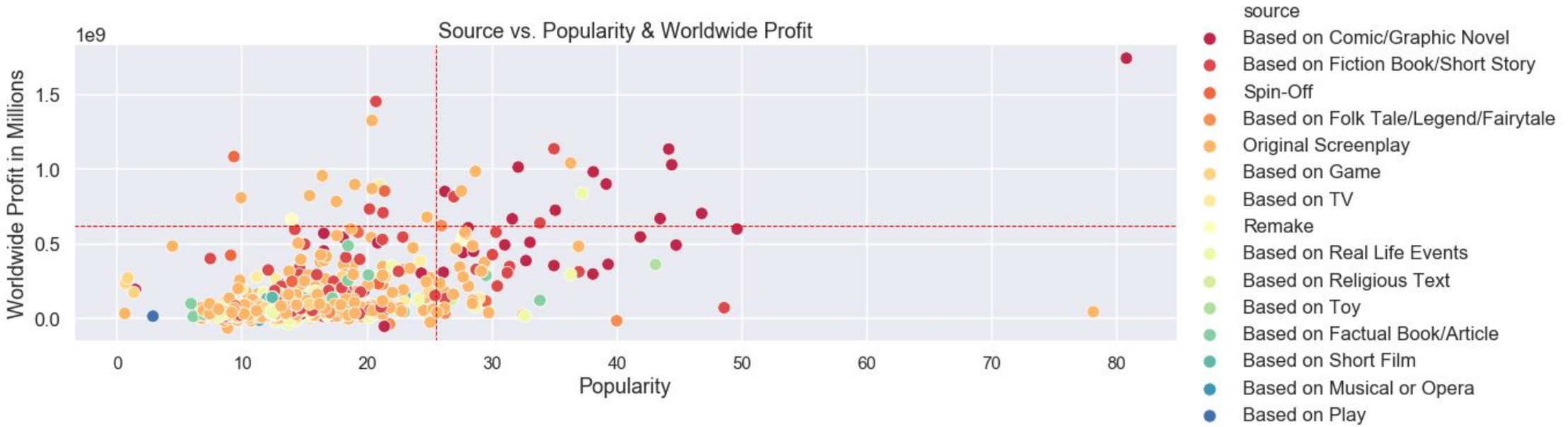


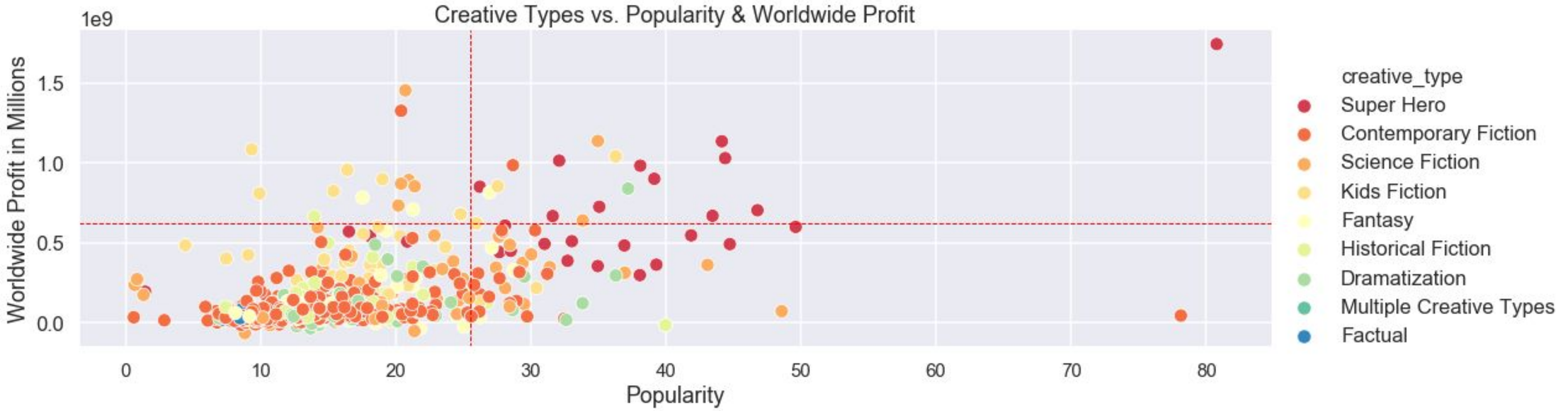
- Top 100 movies score a 25 on average.
- Bottom 100 movies score a 13.2 on average.

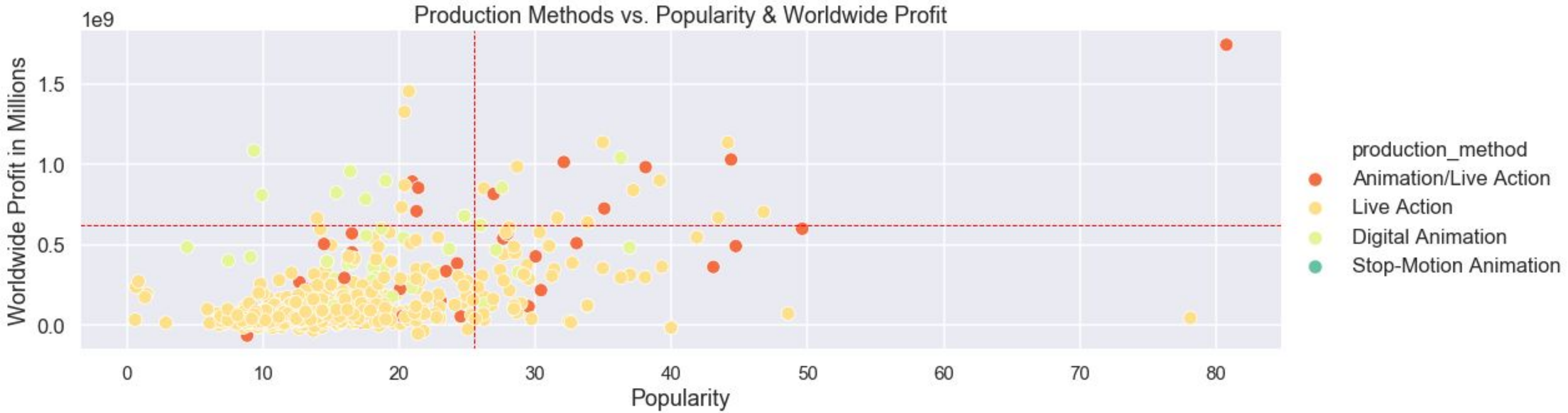

Popularity vs. Worldwide Profit
1e9
Worldwide Profit in Millions
1.5
1.0
0.5
0.0
0
10
20
30
40
50
60
70
80
Popularity

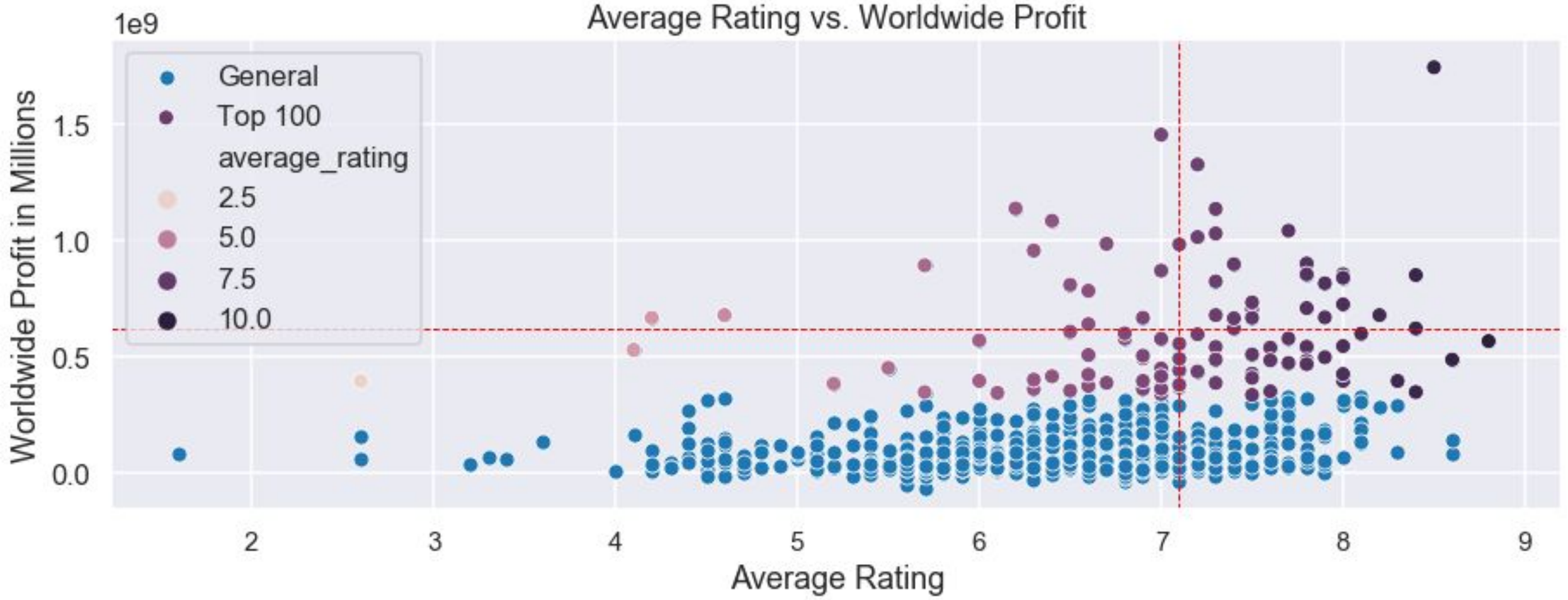
$r^2 = 0.274$


Genres vs. Popularity & Worldwide Profit
1e9
Worldwide Profit in Millions
1.5
1.0
0.5
0.0
0
10
20
30
40
50
60
70
80
Popularity
genre
Action
Adventure
Musical
Thriller/Suspense
Black Comedy
Horror
Drama
Western
Comedy
Romantic Comedy
Concert/Performance

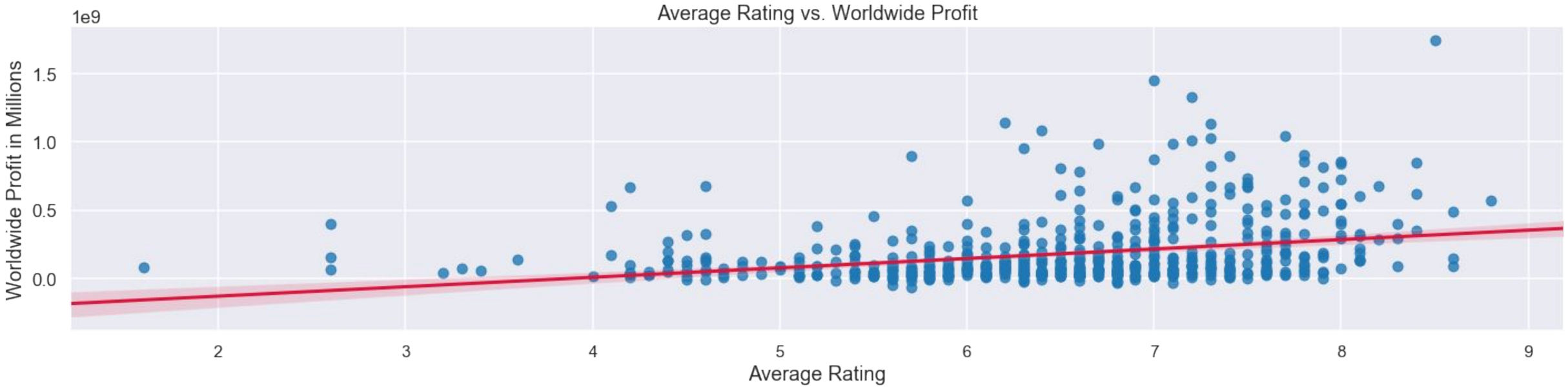

Source vs. Popularity & Worldwide Profit
1e9
Worldwide Profit in Millions
1.5
1.0
0.5
0.0
0
10
20
30
40
50
60
70
80
Popularity
source
Based on Comic/Graphic Novel
Based on Fiction Book/Short Story
Spin-Off
Based on Folk Tale/Legend/Fairytale
Original Screenplay
Based on Game
Based on TV
Remake
Based on Real Life Events
Based on Religious Text
Based on Toy
Based on Factual Book/Article
Based on Short Film
Based on Musical or Opera
Based on Play


Creative Types vs. Popularity & Worldwide Profit
1e9
Worldwide Profit in Millions
1.5
1.0
0.5
0.0
0
10
20
30
40
50
60
70
80
Popularity
creative_type
Super Hero
Contemporary Fiction
Science Fiction
Kids Fiction
Fantasy
Historical Fiction
Dramatization
Multiple Creative Types
Factual

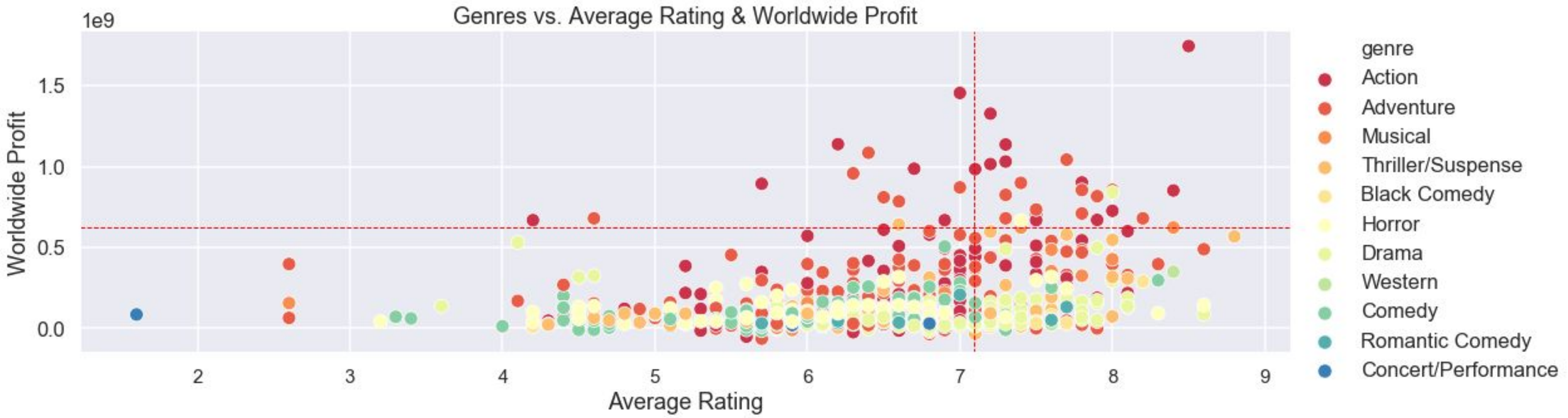

Production Methods vs. Popularity & Worldwide Profit
1e9
Worldwide Profit in Millions
1.5
1.0
0.5
0.0
0
10
20
30
40
50
60
70
80
Popularity
production_method
Animation/Live Action
Live Action
Digital Animation
Stop-Motion Animation



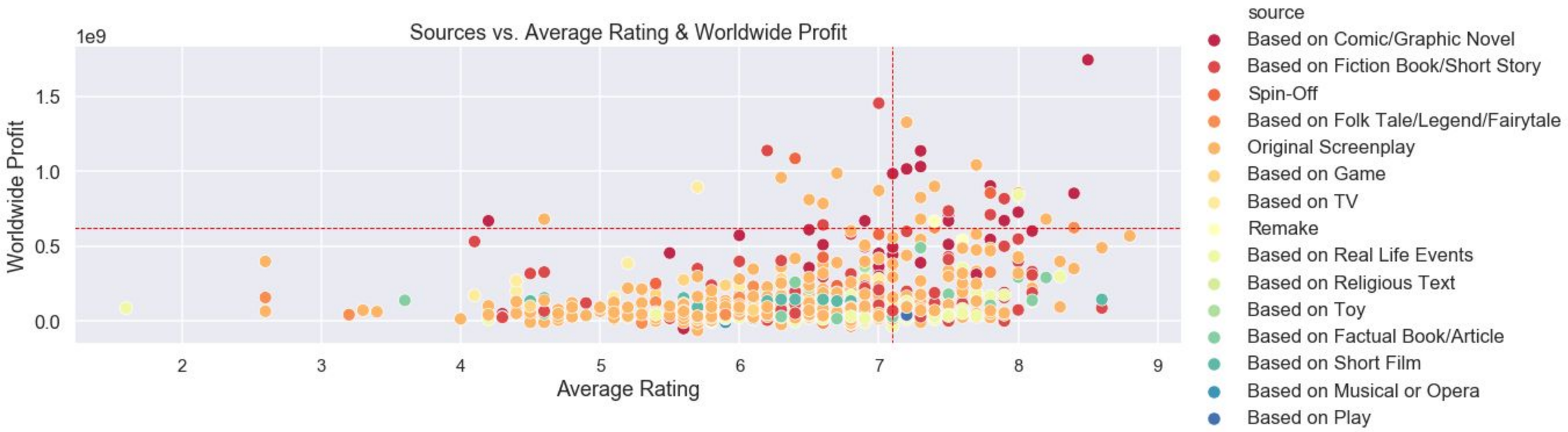
- Top 100 Movies score a 7.1 on average.
- General Movies score a 6.4 on average.
- Bottom 100 Movies score a 6.1 on average.


Average Rating vs. Worldwide Profit
1e9
Worldwide Profit in Millions
1.5
1.0
0.5
0.0
2
3
4
5
6
7
8
9
Average Rating

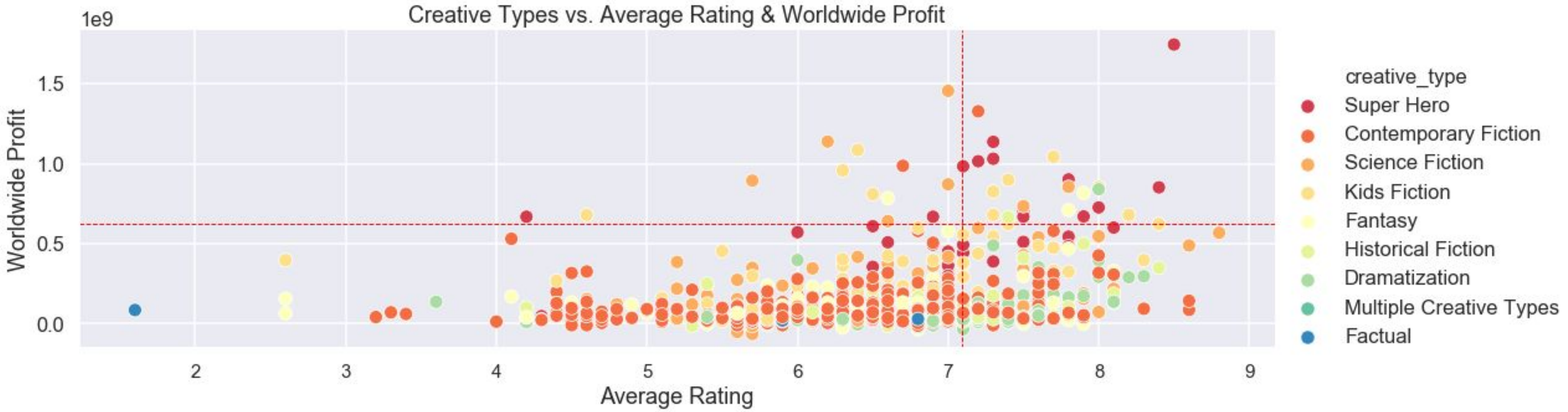
$r^2 = 0.089$



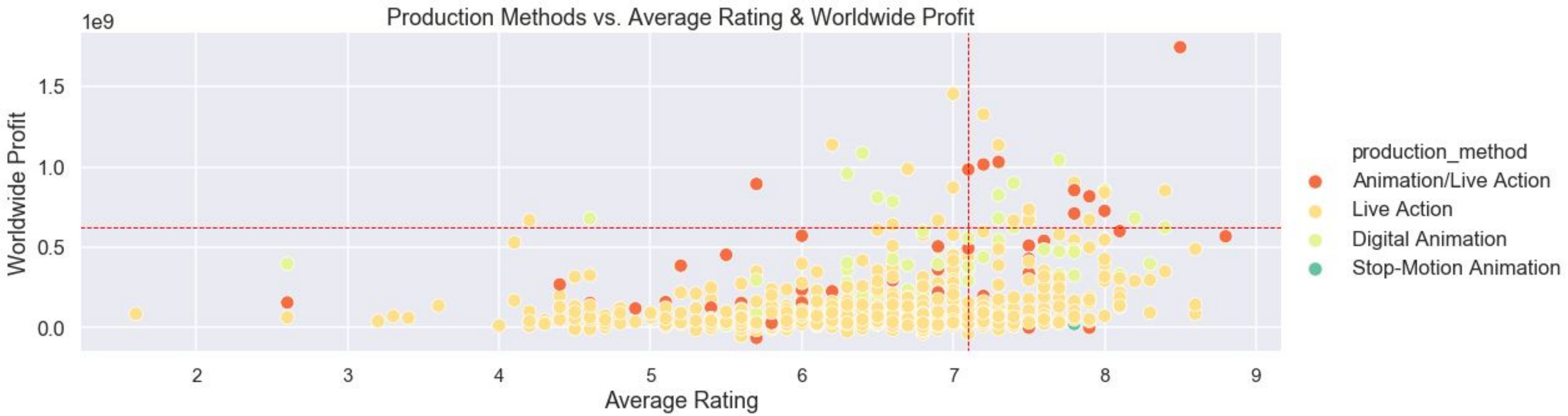
- Thriller/Suspense is highest rated, not Action/ Adventure



- Original Screen Play is highest rated, not Based on Comic Graphic/ Novel



- Science Fiction is highest rated, not Super Hero



- Only Animation/ Live Action remains the consistent.

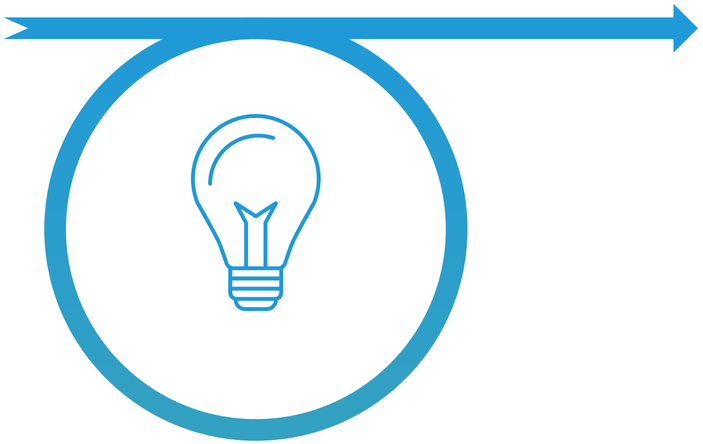
Conclusion

Popularity = ✓

Rating = ✓

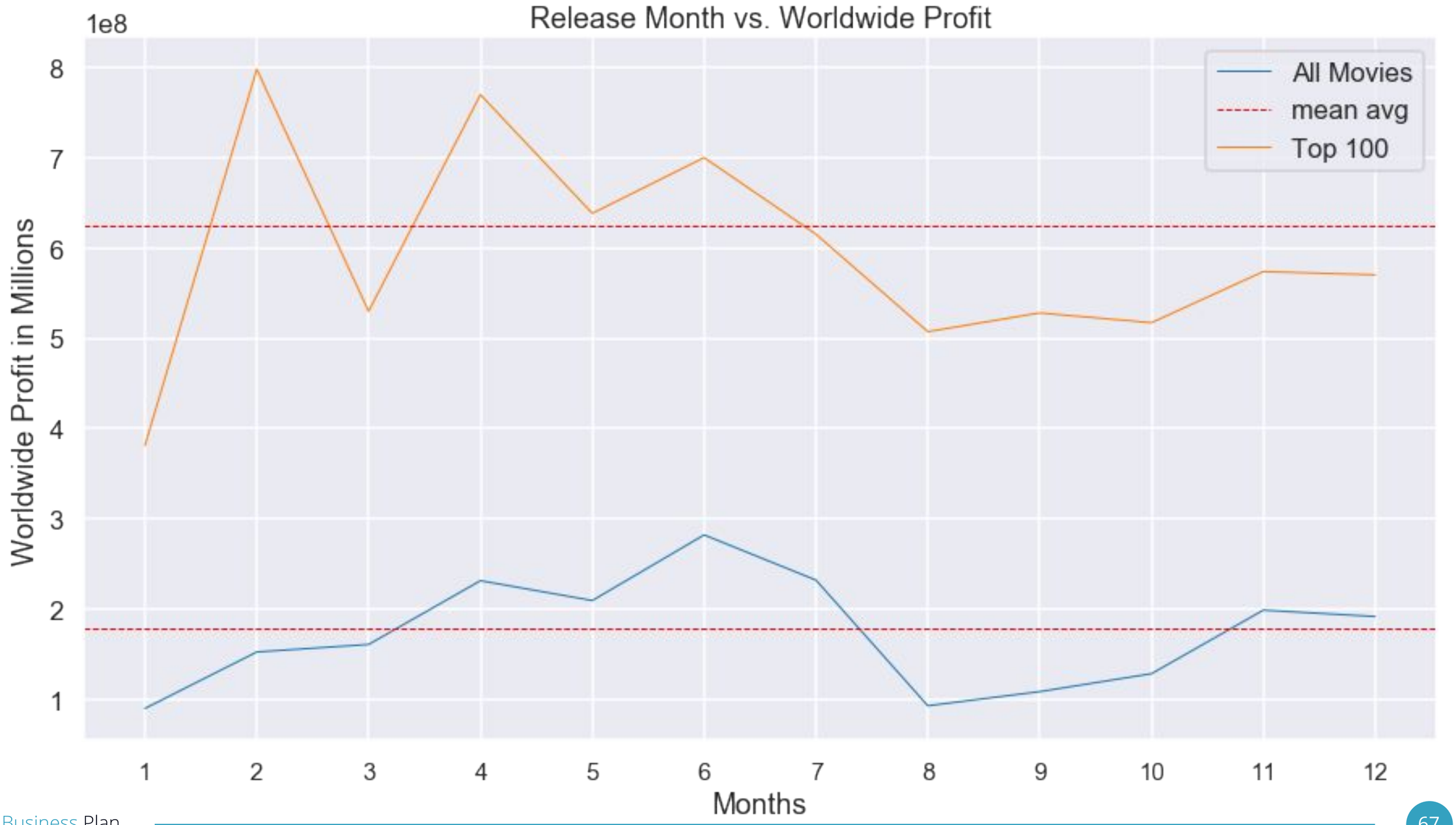
Examples

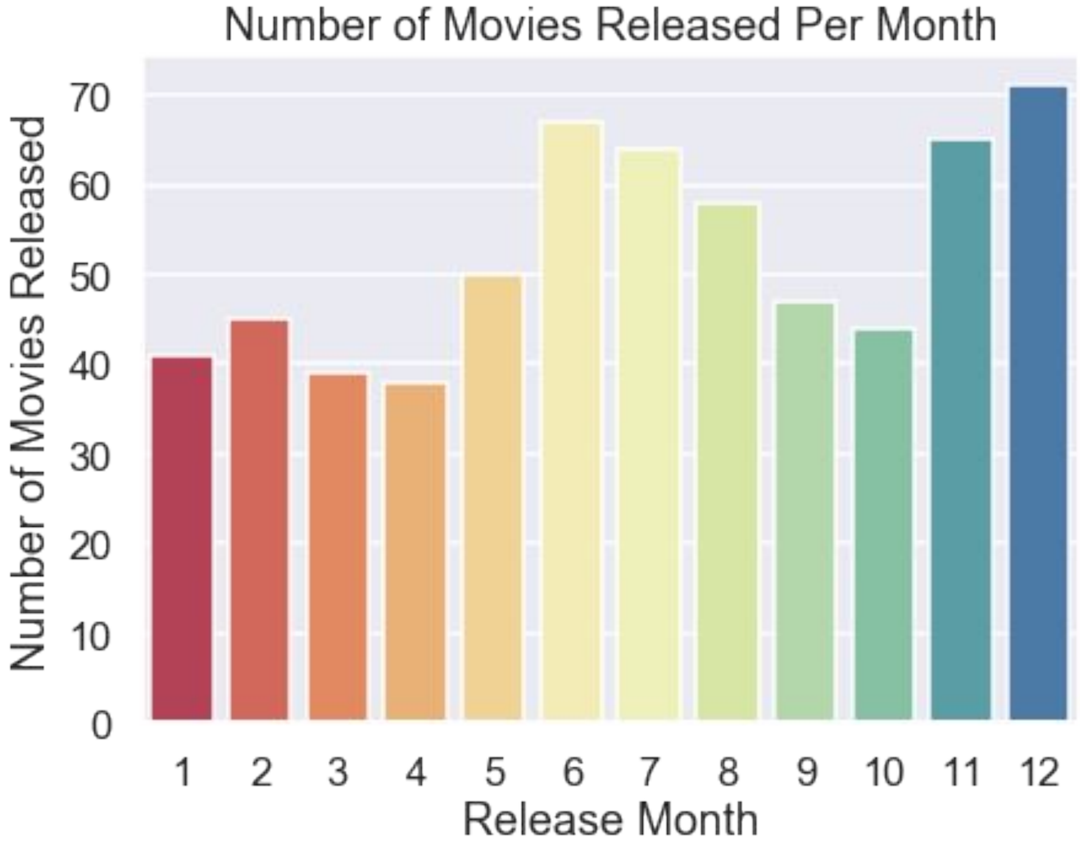
- Action + Science Fiction
- Based on a popular book
- Star power: Jennifer Lawrence
- Underserved need for strong female protagonists

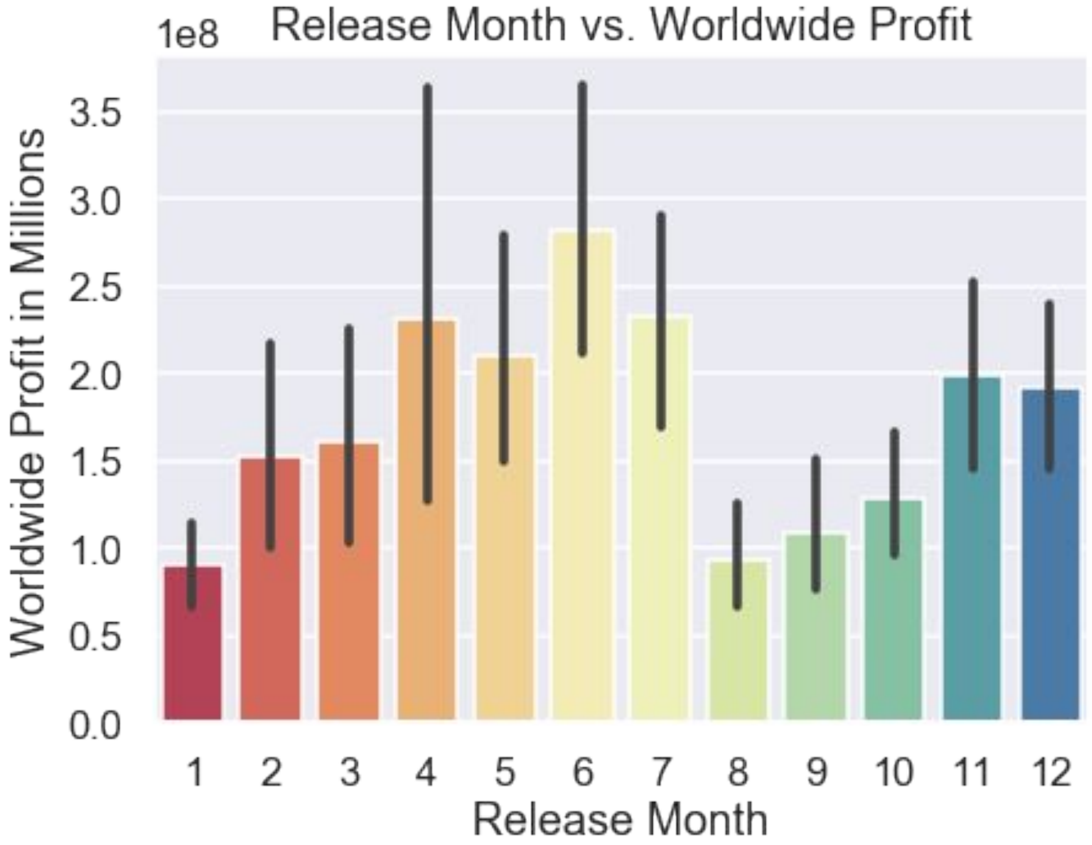


Business Solution

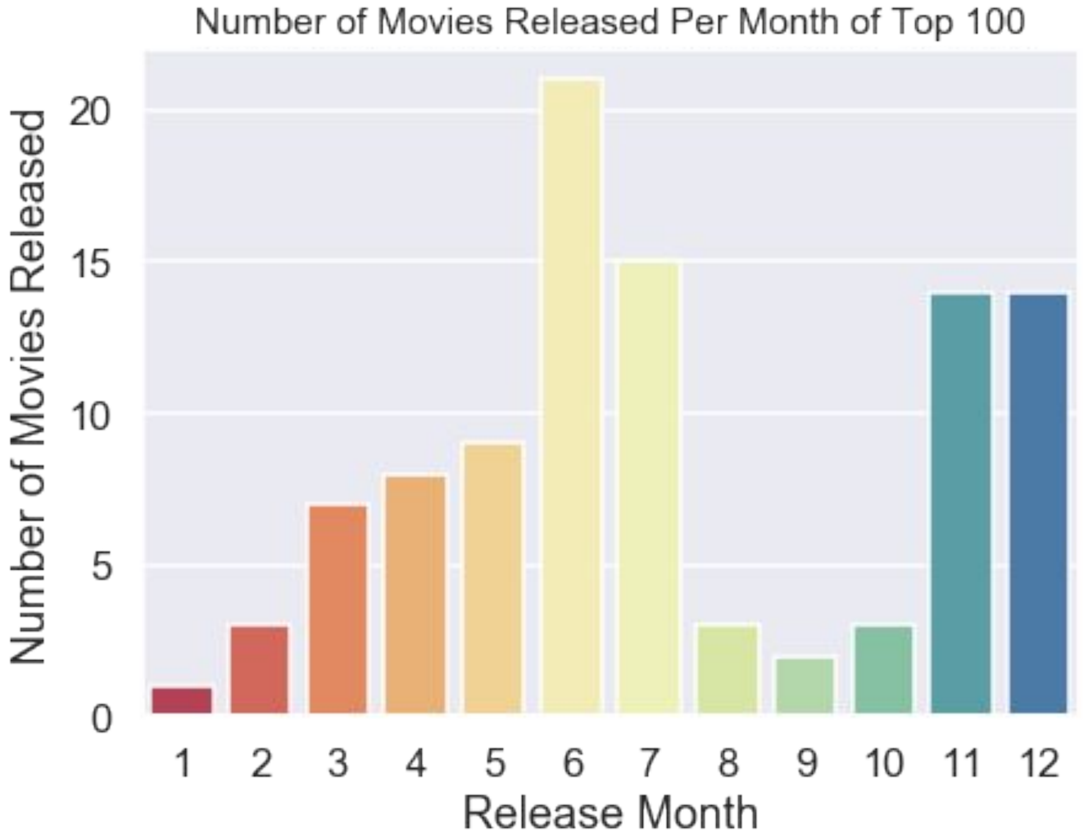
Release Month

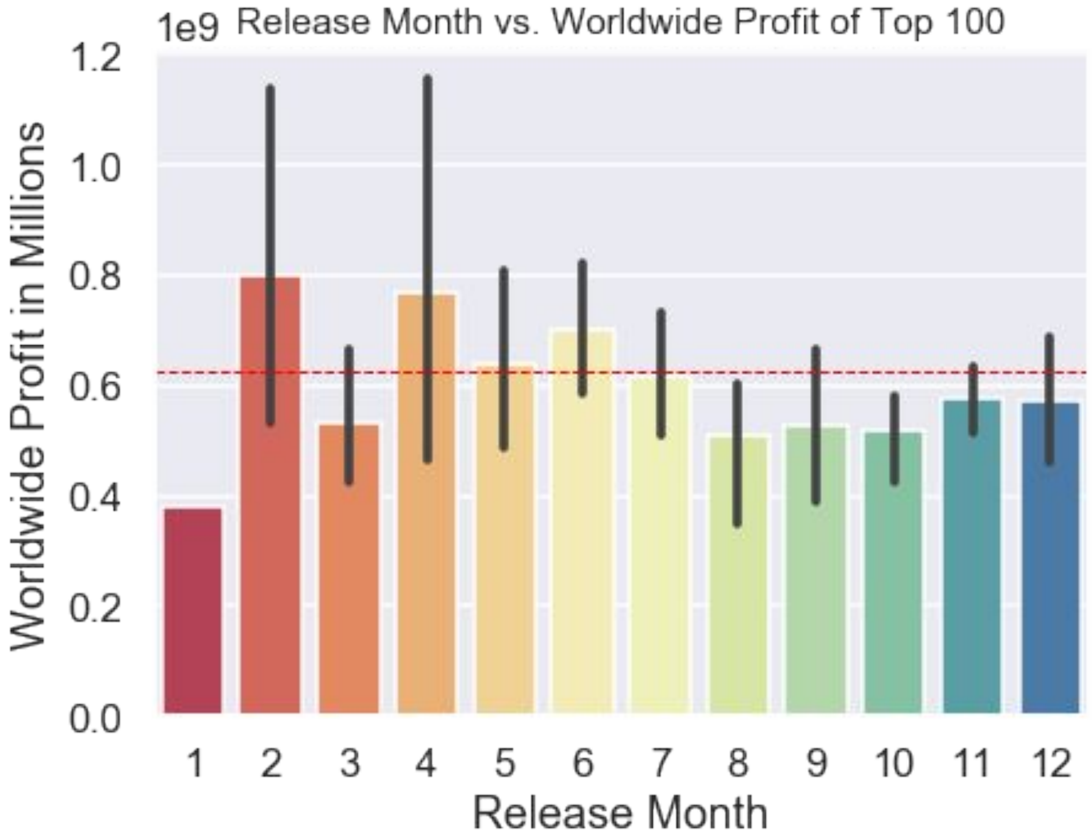

Release Month vs. Worldwide Profit
1e8
Worldwide Profit in Millions
Months
All Movies
mean avg
Top 100
1
2
3
4
5
6
7
8
9
10
11
12



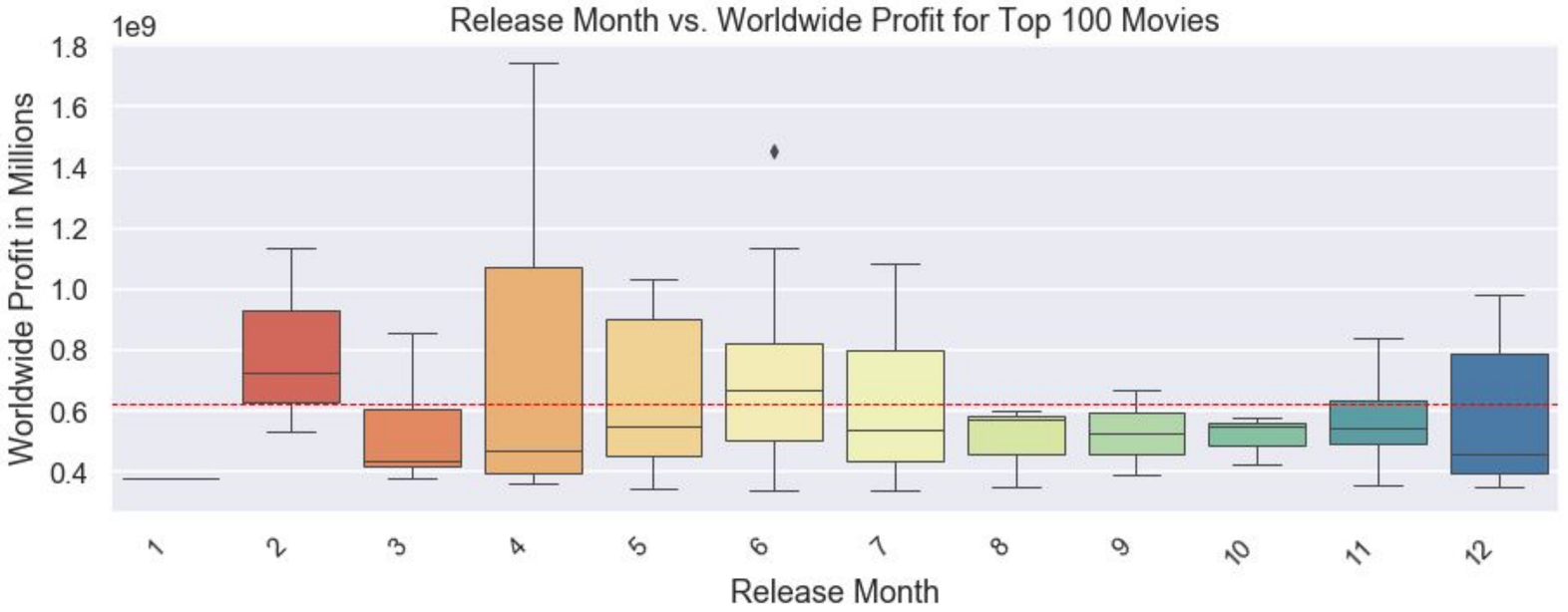


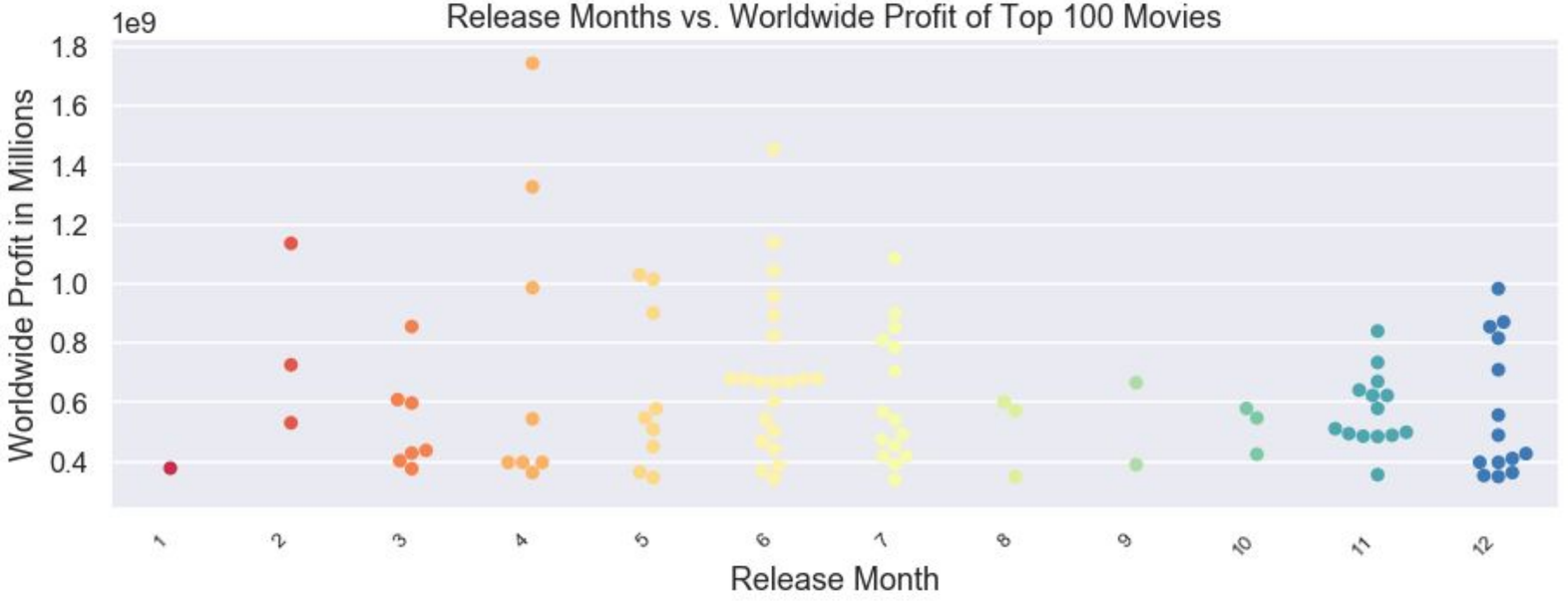
- Most release = December, June
- Most profit = June





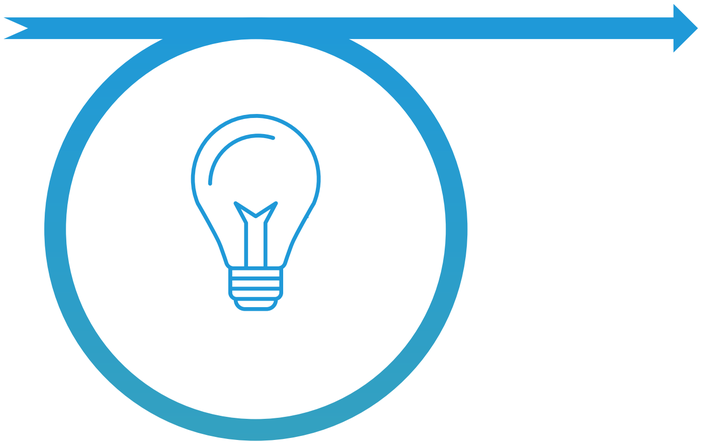
- Most release = June
- Most profit = Febuary, April but not much


Release Month vs. Worldwide Profit for Top 100 Movies
1e9
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
Worldwide Profit in Millions
1
2
3
4
5
6
7
8
9
10
11
12
Release Month


Release Months vs. Worldwide Profit of Top 100 Movies
1e9
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
Worldwide Profit in Millions
1
2
3
4
5
6
7
8
9
10
11
12
Release Month

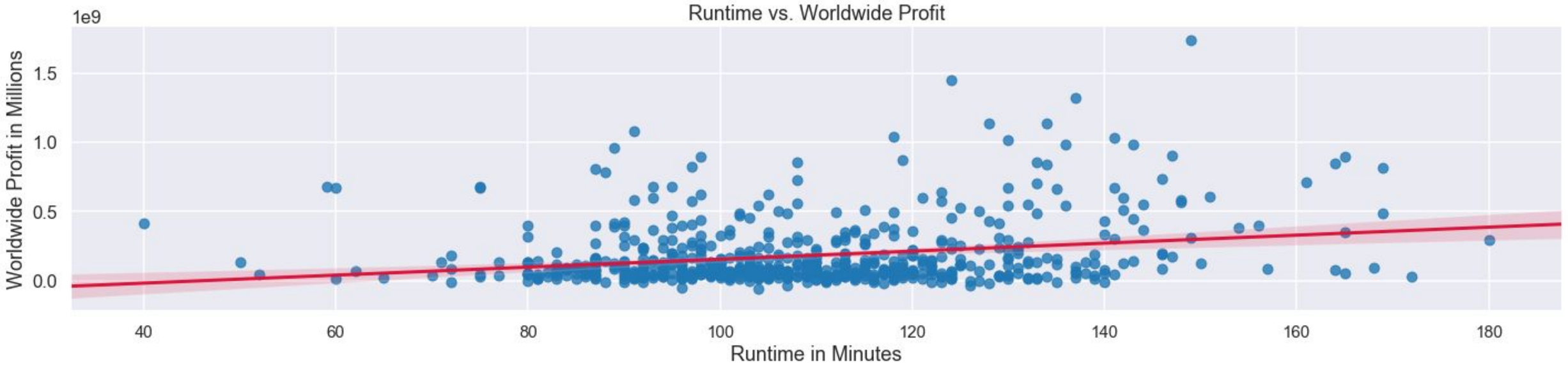
Conclusion

June

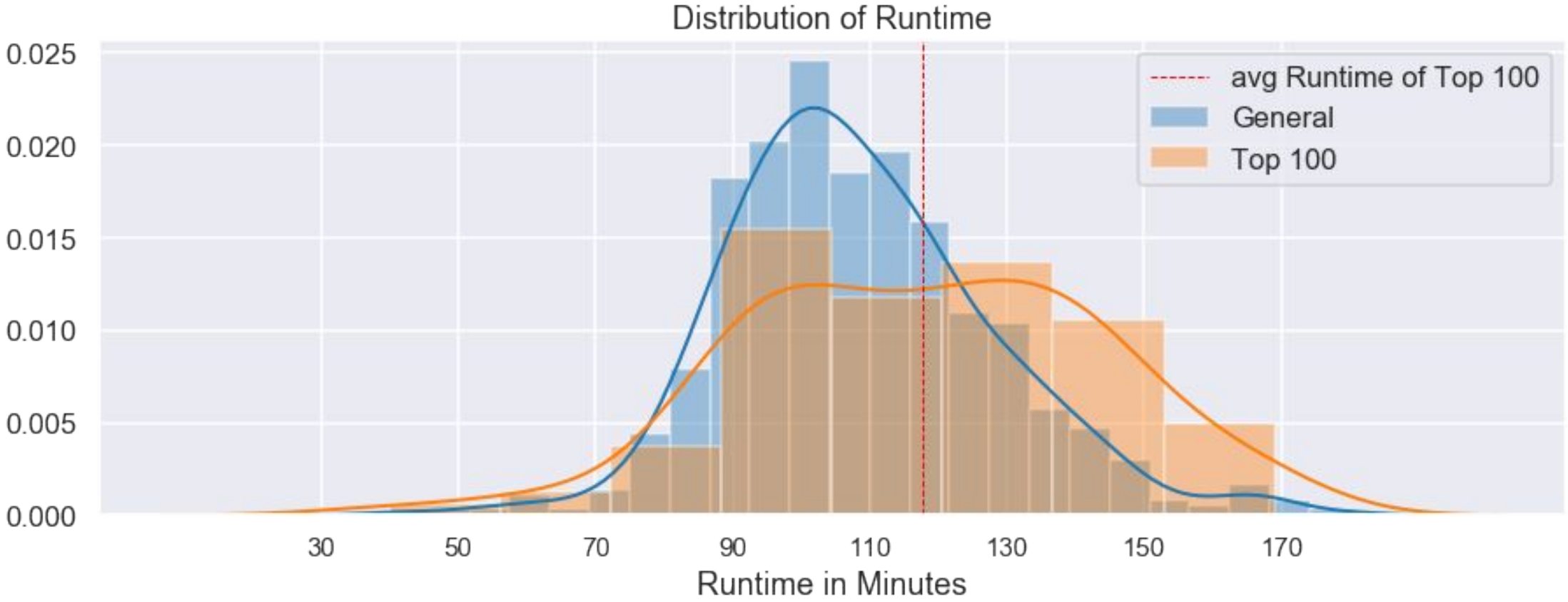


Business Solution

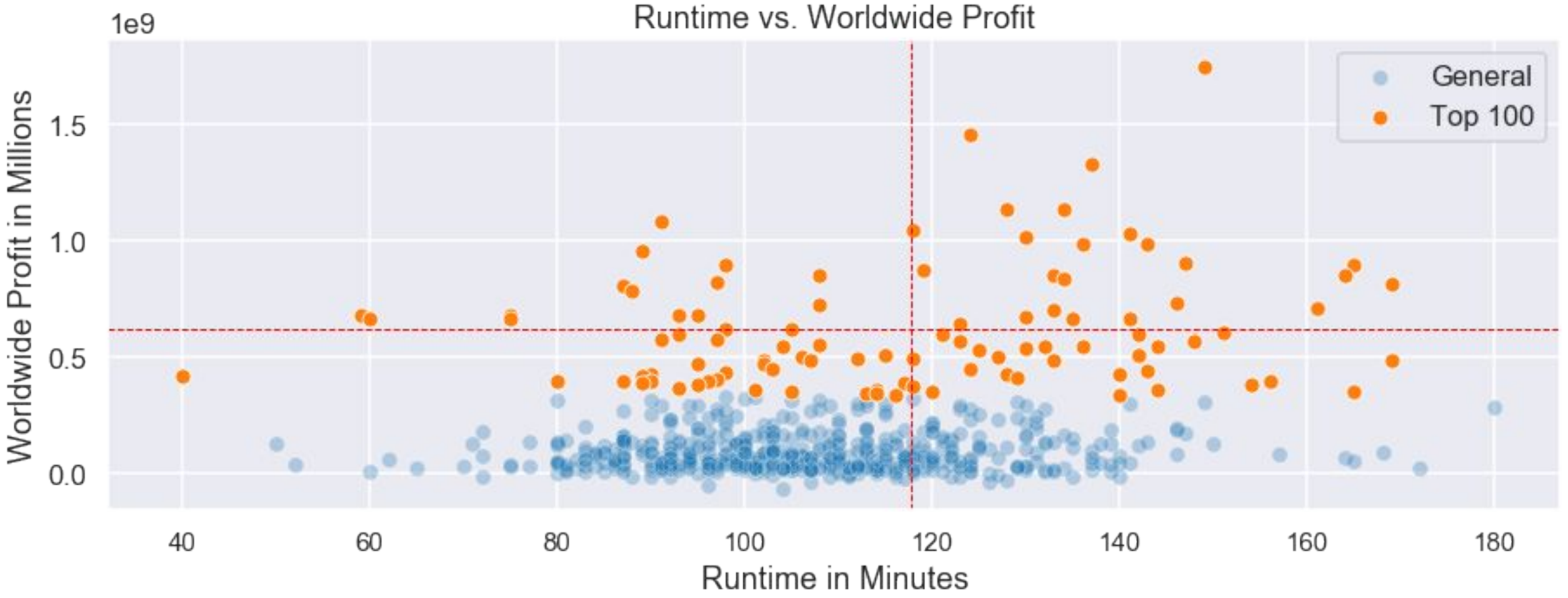
Runtime


Runtime vs. Worldwide Profit
1e9
Worldwide Profit in Millions
1.5
1.0
0.5
0.0
40
60
80
100
120
140
160
180
Runtime in Minutes

$r^2 = 0.058$

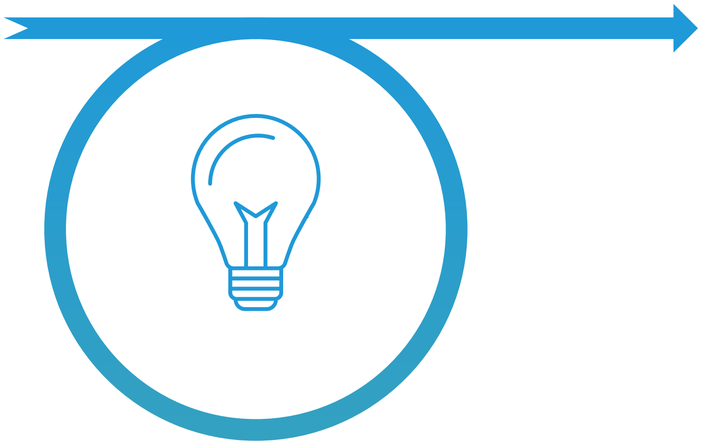

Distribution of Runtime
0.025
0.020
0.015
0.010
0.005
0.000
avg Runtime of Top 100
General
Top 100
30
50
70
90
110
130
150
170
Runtime in Minutes

120 MINS


Runtime vs. Worldwide Profit
1e9
Worldwide Profit in Millions
1.5
1.0
0.5
0.0
General
Top 100
40
60
80
100
120
140
160
180
Runtime in Minutes

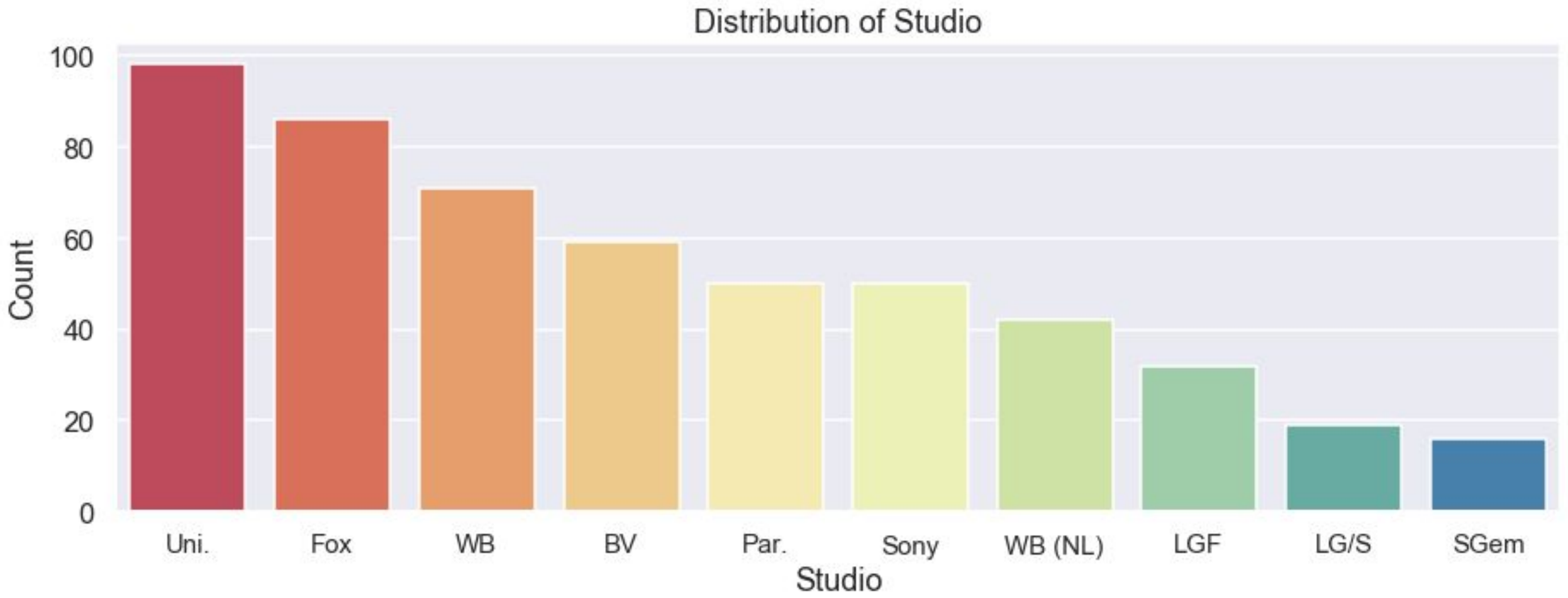
Conclusion

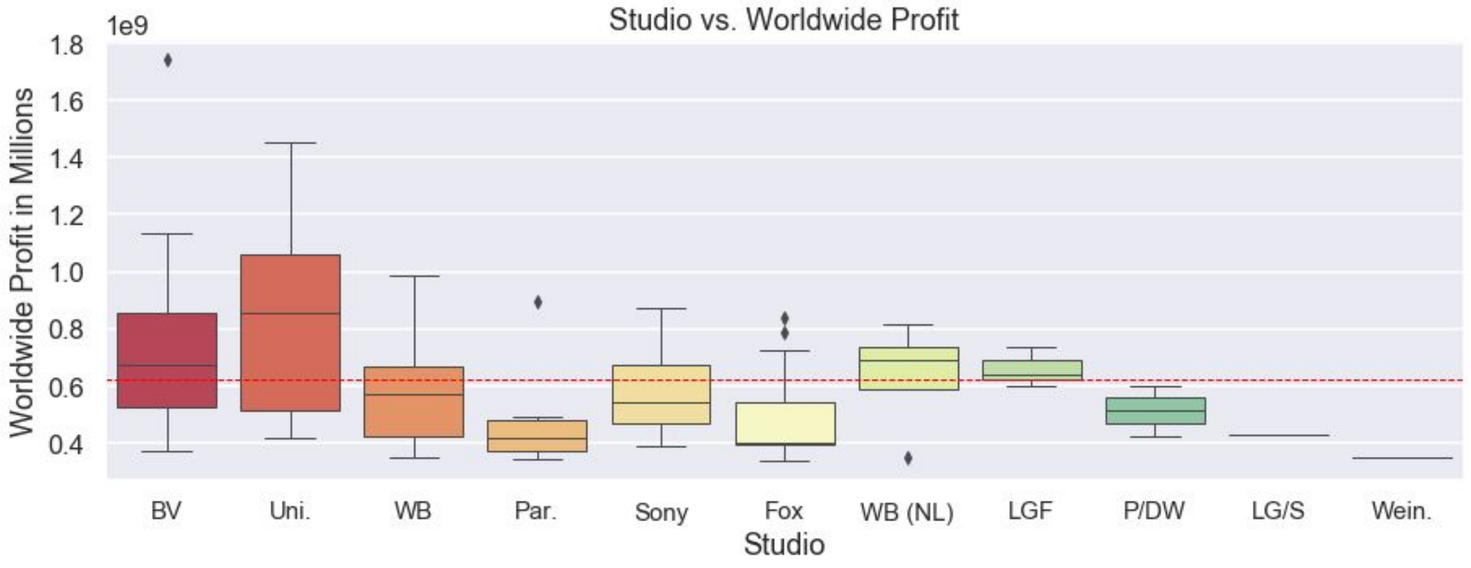
120+ MINS

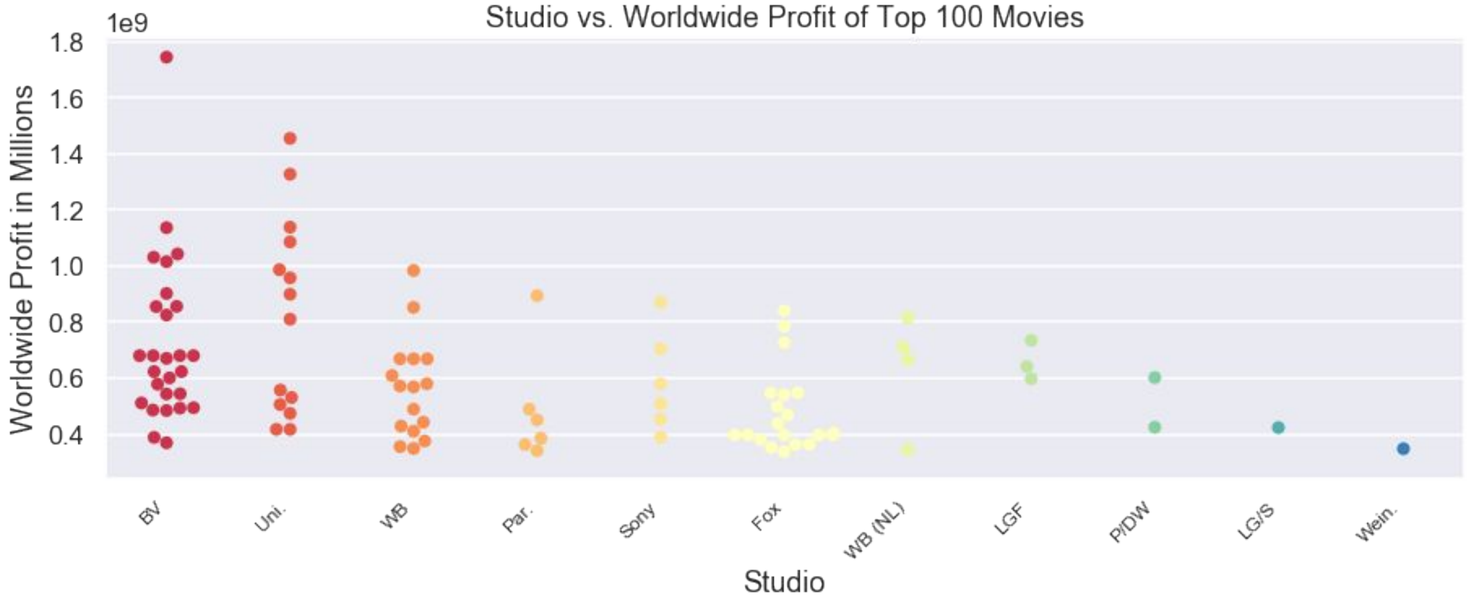


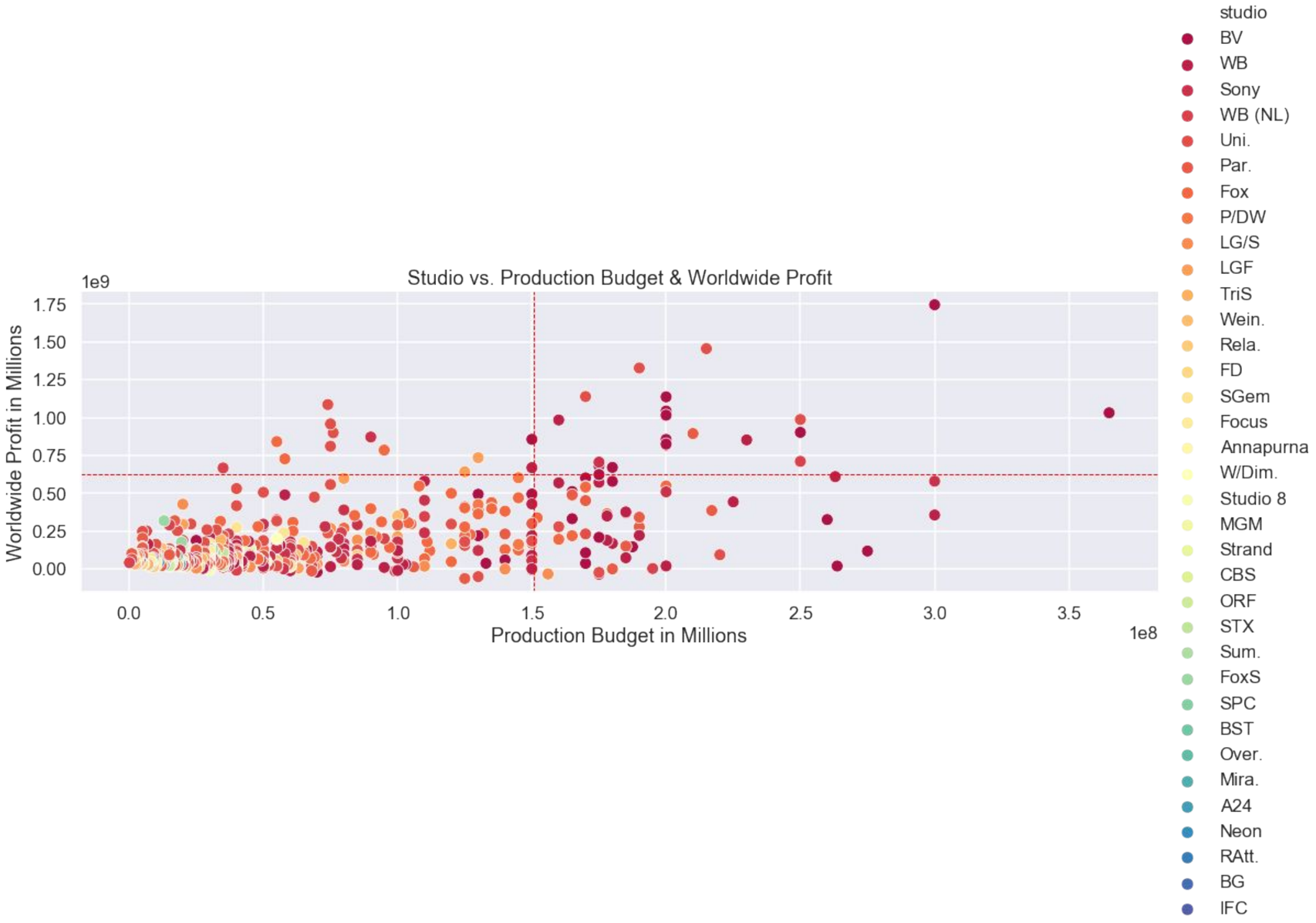
Business Solution

Studio


Distribution of Studio
Count
100
80
60
40
20
0
Uni.
Fox
WB
BV
Par.
Sony
WB (NL)
LGF
LG/S
SGem
Studio


Studio vs. Worldwide Profit
1e9
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
Worldwide Profit in Millions
BV
Uni.
WB
Par.
Sony
Fox
WB (NL)
LGF
P/DW
LG/S
Wein.
Studio


Studio vs. Worldwide Profit of Top 100 Movies
1e9
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
Worldwide Profit in Millions
BV
Uni.
WB
Par.
Sony
Fox
WB (NL)
LGF
P/DW
LG/S
Wein.
Studio


Studio vs. Production Budget & Worldwide Profit
1e9
Worldwide Profit in Millions
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
1e8
Production Budget in Millions
studio
BV
WB
Sony
WB (NL)
Uni.
Par.
Fox
P/DW
LG/S
LGF
TriS
Wein.
Rela.
FD
SGem
Focus
Annapurna
W/Dim.
Studio 8
MGM
Strand
CBS
ORF
STX
Sum.
FoxS
SPC
BST
Over.
Mira.
A24
Neon
RAtt.
BG
IFC

Conclusion

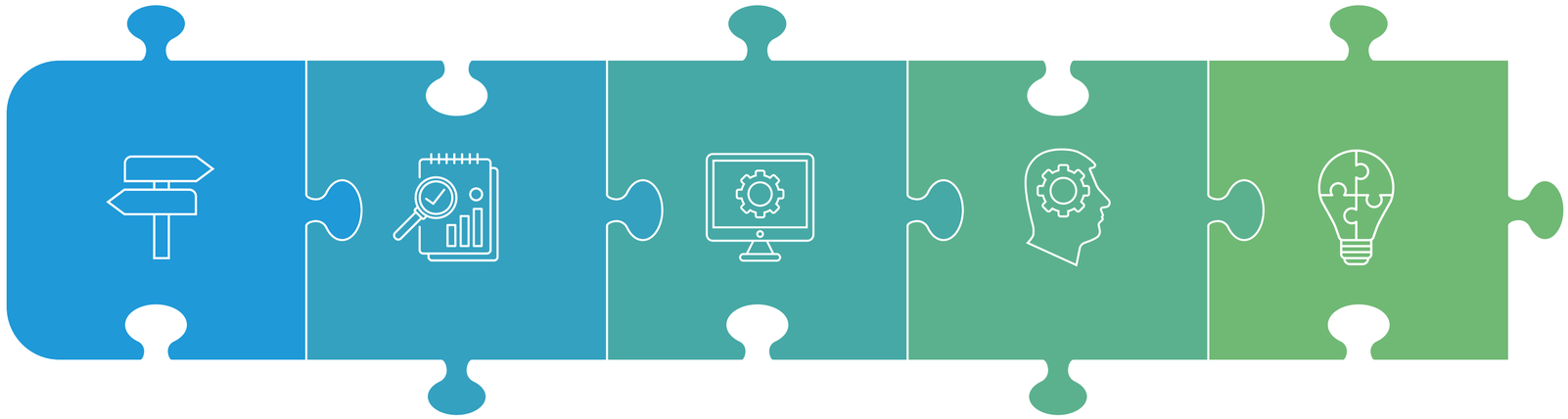
Buena Vista aka Disney

Research Process
Step 5

Business
Objectives

Data
Cleaning & Preparation

Actionable Insights

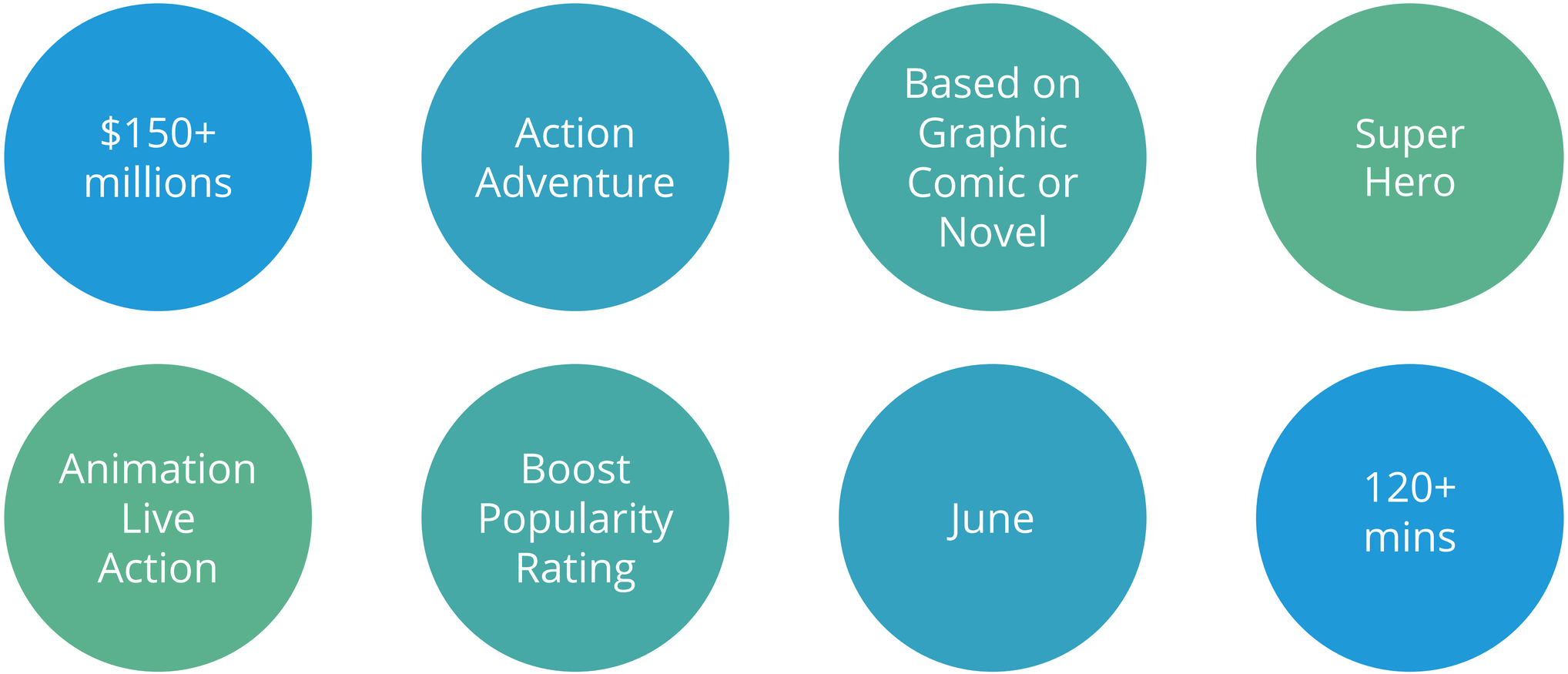


Data
Gathering

Data
Analysis & Visualization

Our **Recommendations**

What Makes a Movie Successful

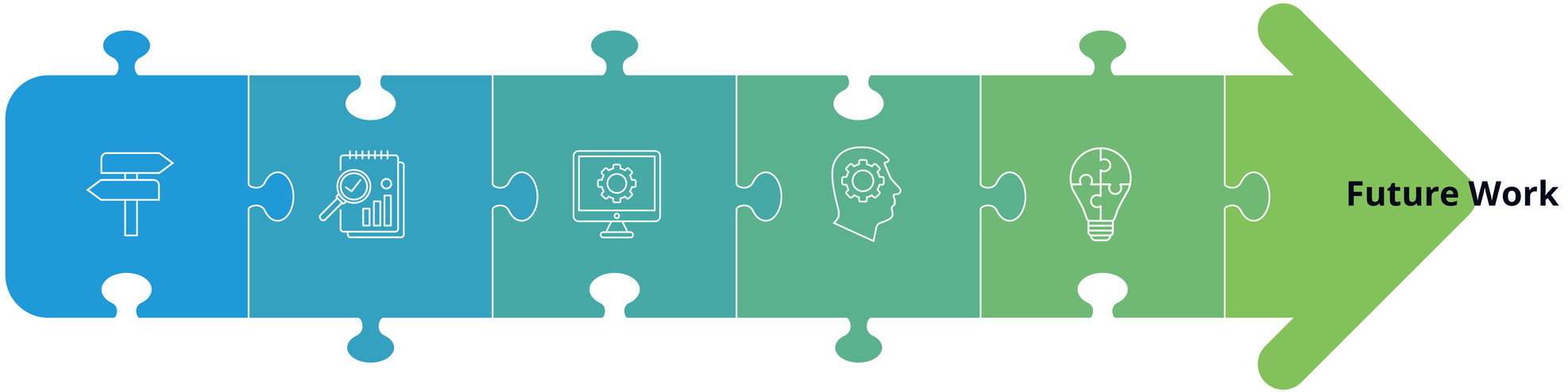


Research Process
Step 6

Business
Objectives

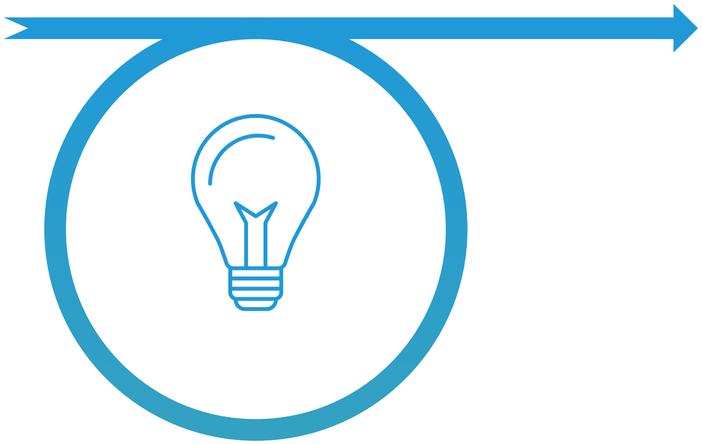
Data
Cleaning & Preparation

Actionable Insights

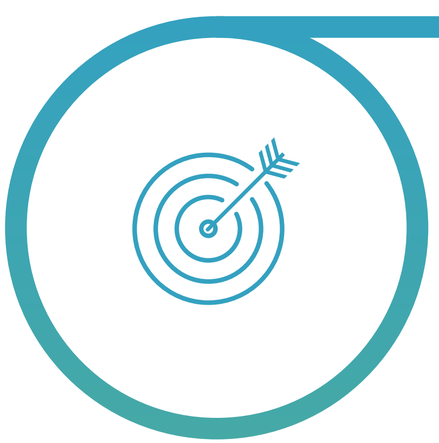


Data
Gathering

Data
Analysis & Visualization



Business Solution
Future Works



Future Works

1. ROI
2. Domestic market vs. International market
3. MPAA rating
4. Quality of stars
5. Quality of scripts
6. Special effect, sound design, music
7. Marketing campaign
8. Popularity of the preceding films
9. Competitions of movie & non-movie events
10. Directors
11. Critics
12. Diversity

Q & A

Thank You