

# **Assignment – Simple Data Extraction**

Report by Riina Kikkas

## Introduction

This analysis examines the 200 highest-grossing movies of all time.

The dataset was extracted from the Box Office Mojo website, which tracks and reports movie box office revenue.

The goal of this analysis was to investigate how movie grosses have changed since the year 2010.

## Methodology

- The dataset contained 200 movies, including their release year and total lifetime box office gross.
- The data was imported into Google Sheets and cleaned.
- Two categorical variables were created:
  - **Release year:** before 2010 vs. 2010 and later
  - **Total gross:** under \$1 billion vs. at least \$1 billion
- A 2×2 contingency table was constructed.
- Row sums, column sums, and the grand total were calculated.
- Proportions by rows, columns, and across the entire dataset were analyzed.
- Ratios of the original values, ratios of the proportions, and ratios of those ratios were computed.

## Links

Box Office Mojo website: [Box Office Mojo](#)

Google sheet: [Top lifetime grosses](#)

## Screenshots

	A	B	C	D	E	F	G
1	Movie	Lifetime gross	Year	Before 2010	From 2010	Under 1 billion	At least 1 billion
2	Avatar	2,923,710,708.00	2009	1	0	0	1
3	Avengers: Endgame	2,799,439,100.00	2019	0	1	0	1
4	Avatar: The Way of Water	2,320,250,281.00	2022	0	1	0	1
5	Titanic	2,264,812,968.00	1997	1	0	0	1
6	Ne Zha 2	2,150,000,000.00	2025	0	1	0	1
7	Star Wars: Episode VII - The Force Awakens	2,071,310,218.00	2015	0	1	0	1
8	Avengers: Infinity War	2,052,415,039.00	2018	0	1	0	1
9	Spider-Man: No Way Home	1,921,426,073.00	2021	0	1	0	1
10	Inside Out 2	1,698,863,816.00	2024	0	1	0	1
11	Jurassic World	1,671,537,444.00	2015	0	1	0	1
12	The Lion King	1,662,020,819.00	2019	0	1	0	1
13	The Avengers	1,520,538,536.00	2012	0	1	0	1
14	Furious 7	1,515,342,457.00	2015	0	1	0	1
15	Top Gun: Maverick	1,495,696,292.00	2022	0	1	0	1
16	Frozen II	1,453,683,476.00	2019	0	1	0	1
17	Barbie	1,447,038,421.00	2023	0	1	0	1
18	Avengers: Age of Ultron	1,405,018,048.00	2015	0	1	0	1
19	The Super Mario Bros. Movie	1,360,847,665.00	2023	0	1	0	1
20	Black Panther	1,349,926,083.00	2018	0	1	0	1

Madagascar: Escape 2 Africa	603,900,354.00	2008	1	0	1	0
No of movies	200.00	Total percentage out of 200				
Movies made before 2010	51.00	25.5				
Movies made from 2010	149	74.5				
Lifetime gross under 1 billion	142.00	71				
Lifetime gross at least 1 billion	58.00	29				

	Under 1 billion gross	At least 1 billion gross	Sum (before/from 2010)
Movies before 2010	43.00	8	51.00
Movies from 2010	99.00	50	149.00
Sum (under/at least 1 billion)	142.00	58	200

Propotions by rows		Propotions by columns		Rations of the original data	
Movies before 2010		Movies under 1 billion gross		By rows	
Under 1 B percentage	84.31	Before 2010		Before 2010	5.4 : 1
At least 1 B percentage	15.69	From 2010		From 2010	2 : 1
Movies from 2010		Movies from 1 billion gross		By columns	
Under 1 B percentage	66.44	Before 2010		Under 1 billion	0.43 : 1
At least 1 B percentage	33.56	From 2010		At least 1 billion	0.16 : 1

Among movies released from 2010, 66.44% earned under 1 billion and 33.56% earned at least 1 billion.  
 Among movies released before 2010, 84.31% earned under 1 billion and 15.69% earned at least 1 billion.  
 Of all movies that grossed from 1 billion, 13.79% were released before 2010 and 86.21% movies were released from 2010.  
 Of all movies that grossed under 1 billion, 30.28% were released before 2010 and 69.72% movies were released from 2010.  
 Before 2010, movies were more than five times as likely to gross under 1B than over 1B.  
 After 2010, movies were about twice as likely to gross under 1B as over 1B.  
 Among under 1B movies, there were about 0.43 older movies for every newer one.  
 Among billion-dollar movies, there were only 0.16 older movies for every newer one.

Total percentage out of 200	Ratios of the ratios
Movies made before 2010 & under 1 billion gross	Row ratios 21.50
Movies made before 2010 & at least 1 billion gross	Column ratios 4.00
Movies made from 2010 & under 1 billion gross	2.7
Movies made from 2010 & at least 1 billion gross	2.69
<b>Ratio = 0.843 : 15.7 ≈ 5.4 : 1</b>	The under/at least 1B ratio was 2.7 times higher for movies before 2010 than for movies from 2010.
Movies made before 2010 under 1B	The old/new ratio among under 1B movies was 2.7 times higher than among billion-dollar movies.
Movies made before 2010 at least 1B	Billion-dollar movies became much more common after 2010.
<b>Ratio = 66.4 : 33.56 ≈ 2 : 1</b>	Among movies released before 2010, the share of under-1-billion grosses was about 5.4 times higher than the share of billion-dollar grosses.
Movies made from 2010 under 1B	Among movies released from 2010 onwards, the share of under-1-billion grosses was about twice as high as the share of billion-dollar grosses.
Movies made from 2010 at least 1B	Among under-1-billion movies, there were only 0.43 older movies for every newer movie.
<b>Ratio = 0.303 : 0.697 ≈ 0.43 : 1</b>	Among billion-dollar movies, there were only 0.16 older movies for every newer movie.
Gross under 1B before 2010	
Gross under 1B from 2010	
<b>Ratio = 0.138 : 0.862 ≈ 0.16 : 1</b>	