



# IMDB MOVIES

(delving into *Hollywood*)

BY Carlos Cutillas AND  
Yusuf Olodo

# Data Acquisition

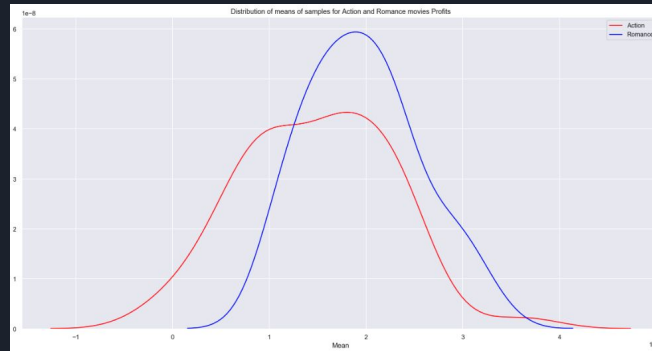
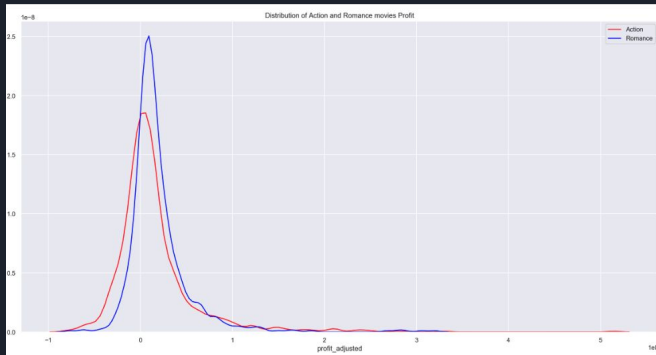
- Web Scrapping (IMdb movies website)
- Kaggle movies dataset
- Inflation data to adjust budget and gross revenue

	all_titles	all_years	all_scores	all_meta	all_votes	all_gross	all_genres	id	decade	budget	imdb_id	genre_bis	profit	Index_all_years
0	Star Wars: Episode VII - The Force Awakens	2015	79	81	799424	936662225	Action, Adventure, Sci-Fi	tt2488496	10s	245000000	tt2488496	Action	691662225	237.000000
1	Avatar	2009	78	83	1065708	760507625	Action, Adventure, Fantasy	tt0499549	00s	237000000	tt0499549	Action	523507625	214.537000
2	Titanic	1997	78	75	979853	659325379	Drama, Romance	tt0120338	90s	200000000	tt0120338	Romance	459325379	160.516667
3	Jurassic World	2015	70	59	549843	652270625	Action, Adventure, Sci-Fi	tt0369610	10s	150000000	tt0369610	Action	502270625	237.000000

# Hypothesis 1 - Difference between profit margins

GROUPS:

- Action movies
- Romance movies
  - $H_0 = \text{Mean}(\text{Action movies profit}) = \text{Mean}(\text{Romance movies profit})$
  - $H_1 = \text{Mean}(\text{Action movies profit}) \neq \text{Mean}(\text{Romance movies profit})$



T-statistic -2.365  
p-value 0.0202565

Cohen's d: 0.13669400069075582



# CONCLUSION FOR HYPOTHESIS 1

The conclusion for Hypothesis 1 is that although we got a P-value less than 0.05, there is no practical significance to justify rejecting the Null hypothesis

# HYPOTHESIS 2

Null Hypothesis - The Null Hypothesis states that there is no difference in the romance movie budget and the Action movie budget.

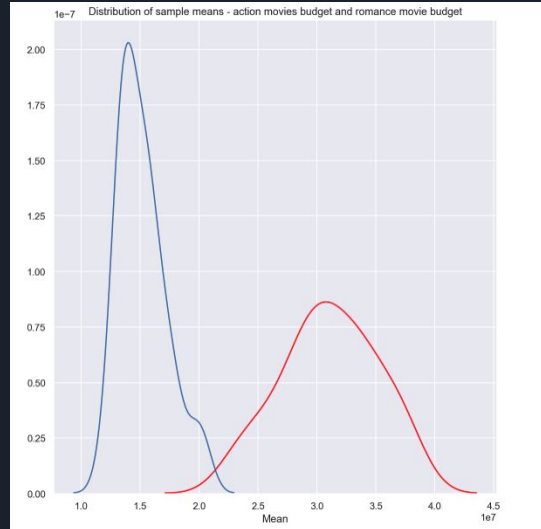
Alternative Hypothesis states that there is a difference between the two independent variables

Two samples of 393 samples

p-value: 0.0

T-statistic (Welch): 19.4

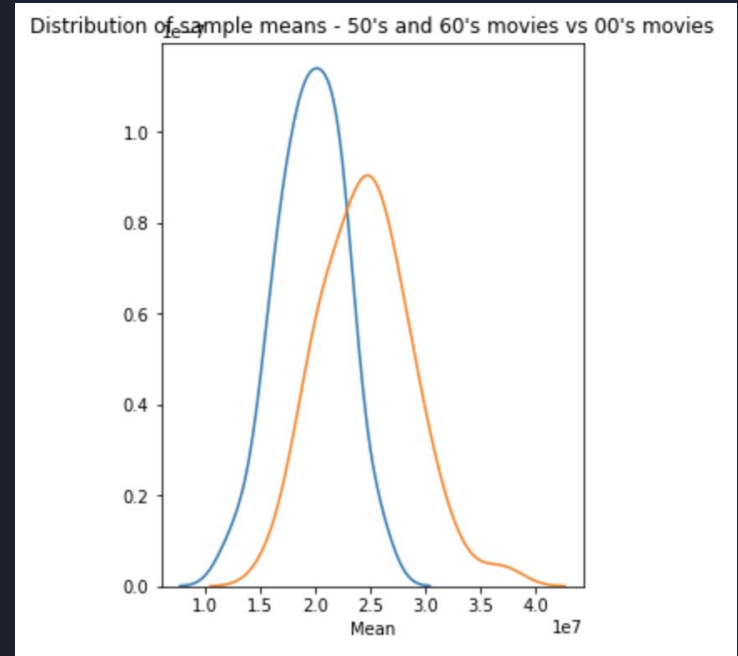
Cohen d-test: 0.88



# Hypothesis 3 - Movies nowadays have higher budget than movies on 50's and 60's

- Two samples of 197 movies
- p-value: 0.02
- T-statistic (Welch): 2.02
- Cohen d-test: 0.2

**REJECTED NULL HYPOTHESIS**



## Hypothesis 4 - Movies on 50's and 60's made more profit than movies nowadays (00's and 10's)

- Two samples of 191 movies
- T-statistic (Welch): 10.97
- p-value= ~0.0
- Cohen's d: 1.1

**REJECTED NULL HYPOTHESIS**

