

Movie Trends and Financial Analysis 1995 - 2015

By Aaron Morgan, Kyoo Ha Cha, Shawn
Advani, Terrence Gray, Pedro Flores

SMU Boot Camp Project, 1 2019





Research Question and Hypothesis





Research Question and Hypothesis

Research Question:

- Do we have sufficient reliable data for our analysis?
- What genres were most popular between 1995 and 2015?
- What films had the biggest return on investment?
- Which content rating had the highest revenue on average?
- Which genres had the highest IMDb ratings and does it have a correlation with film duration?

Hypothesis:

- H1: Our data is reliable
- H2: G rated films had the highest revenue
- H3: Drama films had the greatest increase in profits over the last 20 years
- H4: Mean IMDb score has a difference per genre
- H5: Mean duration has a difference per genre



Data Background and Methodology





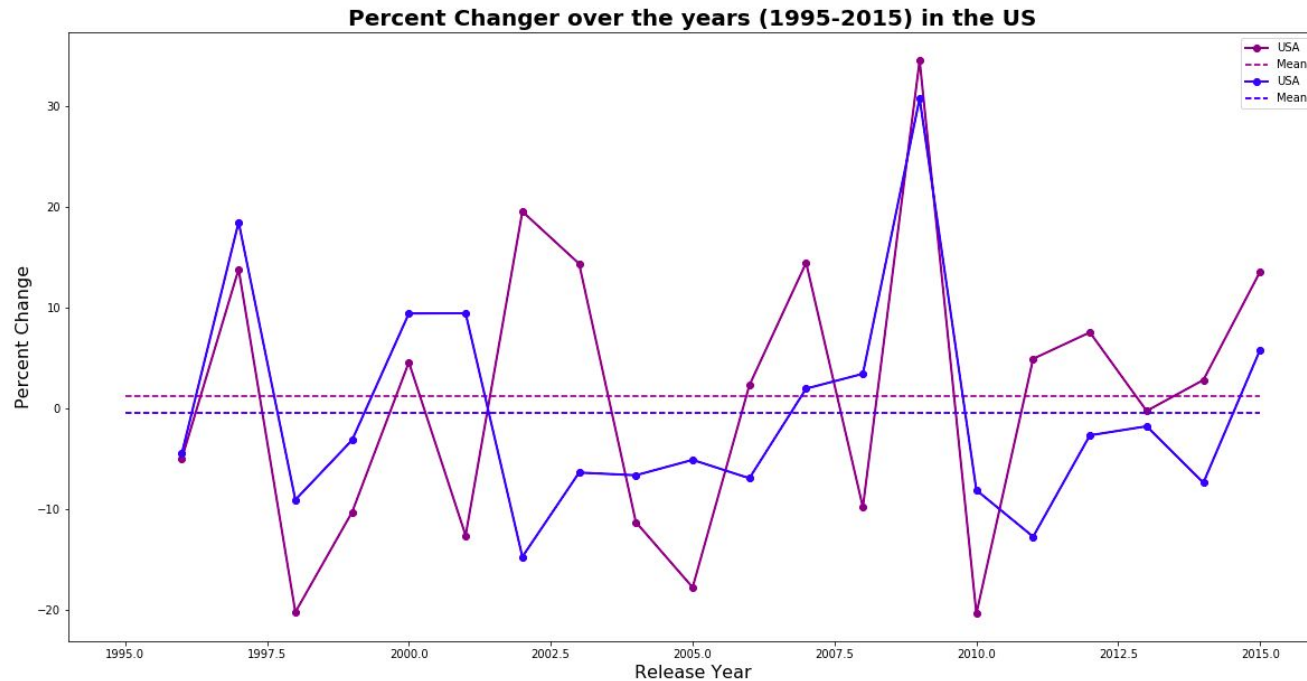
Data Background

- We found our data that we actually used through Kaggle
- The data is three CSV files, we did not use an API
- Merged the three CSV files on different columns
- Dropped N/A values and duplicates
- Deleted columns that were not necessary to our analysis
- Final data frame contained information including:
 - Movie Title
 - IMDb Score
 - Release Year
 - Film Budget
 - Gross Income
 - Genre
 - Film Language
 - Country of Origin
 - Most complete information for years 1995 and 2015



1) Data Quality





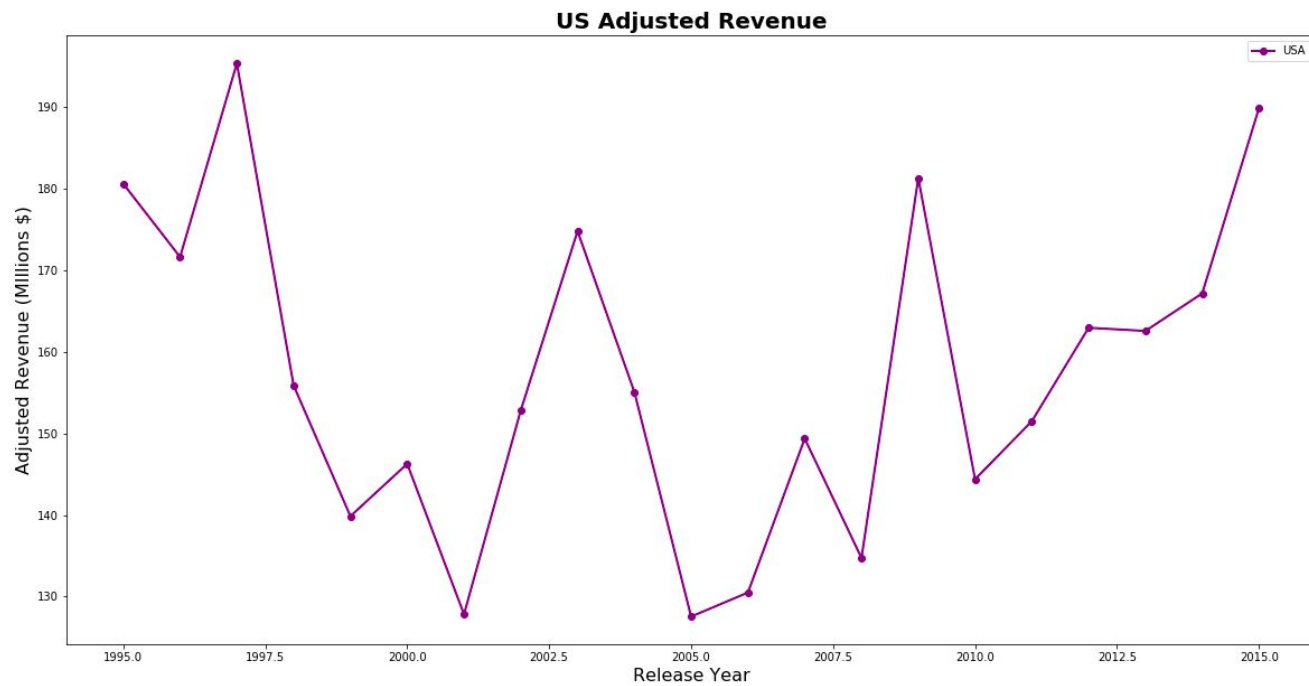
```

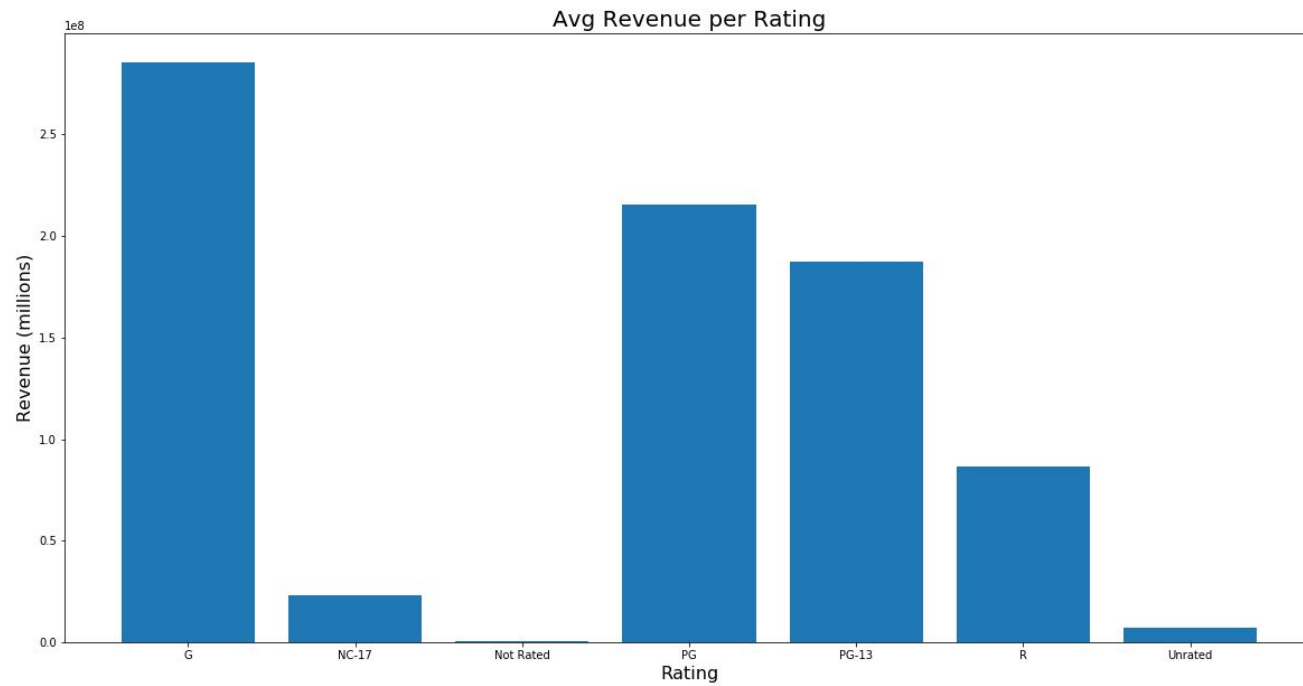
1  ## p value< .05 they are not the same
2  ## p value> .05 data is not not the same
3  from scipy.stats import ttest_ind
4
5  cat1 = subDF
6  #cat1 is the original dataset we started using (about 4.5 k items)
7  cat2 = newdf
8  #cat2 is the new relevant dataset (over 15k items)
9
10 ttest_ind(cat1["Adjusted Revenue"], cat2['lifetime_gross'])

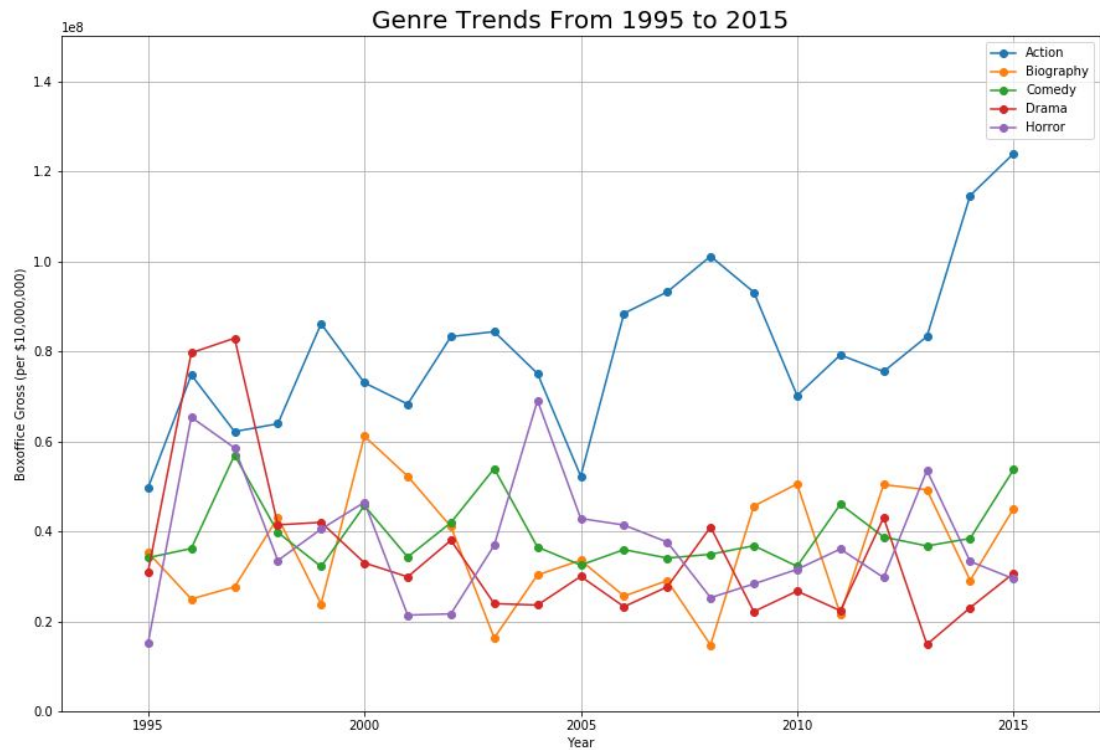
```

]: Ttest_indResult(statistic=-34.87137194448279, pvalue=1.4662380629909464e-31)

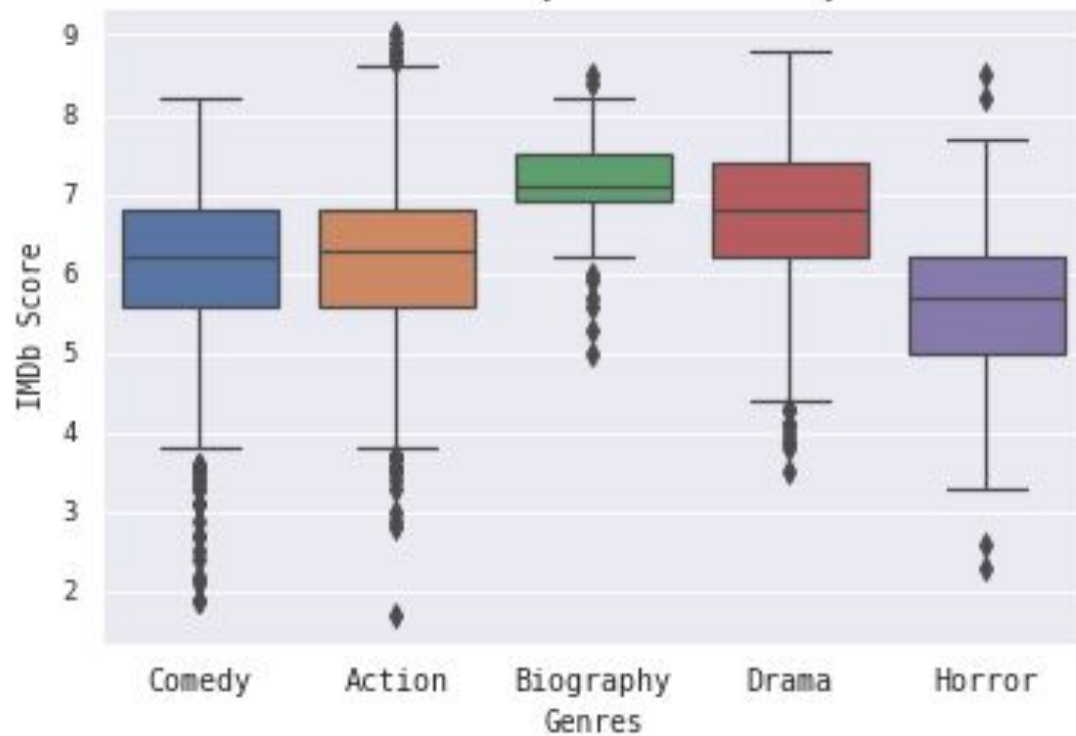
2) Results and Analyses



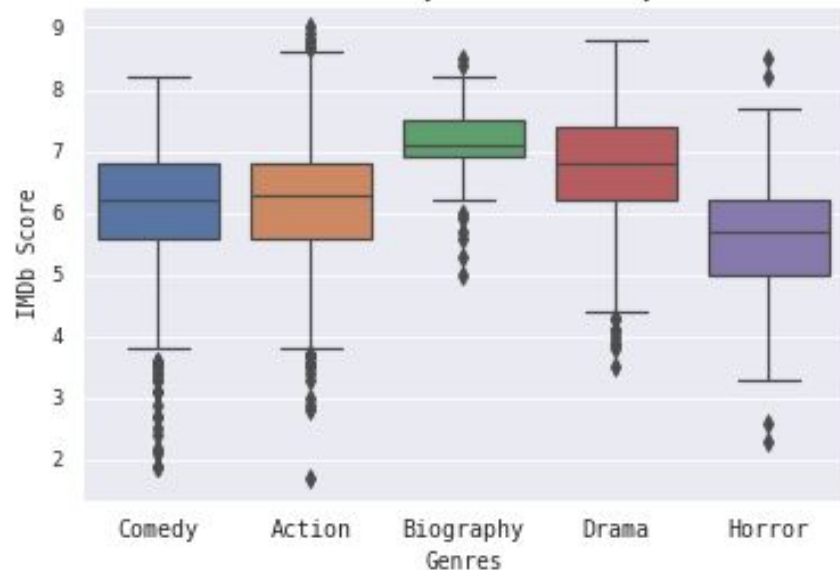




IMDb Score Comparison per Genre



IMDB Score Comparison per Genre



Check whether there is any significant difference between the means of five genres

Extract individual groups

group1 = genres_interested[genres_interested["Genres"] == "Comedy"]["IMDb Score"]

group2 = genres_interested[genres_interested["Genres"] == "Action"]["IMDb Score"]

group3 = genres_interested[genres_interested["Genres"] == "Biography"]["IMDb Score"]

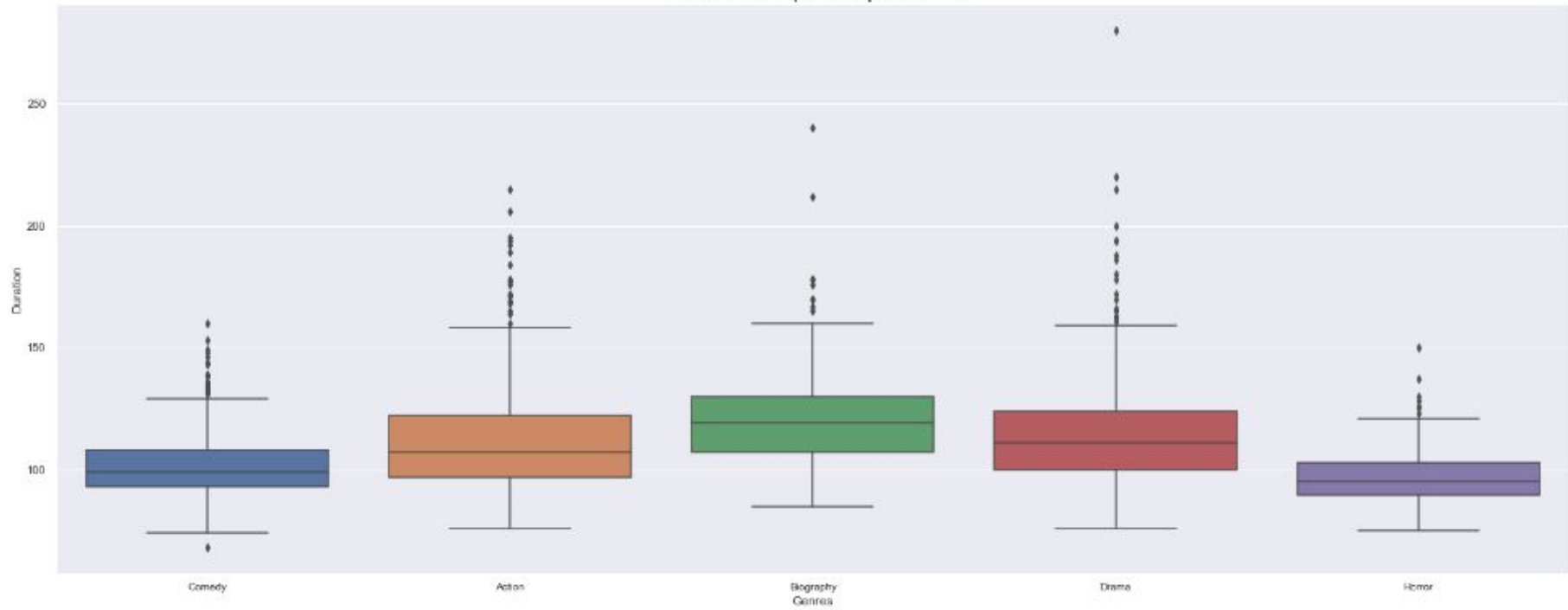
group4 = genres_interested[genres_interested["Genres"] == "Drama"]["IMDb Score"]

group5 = genres_interested[genres_interested["Genres"] == "Horror"]["IMDb Score"]

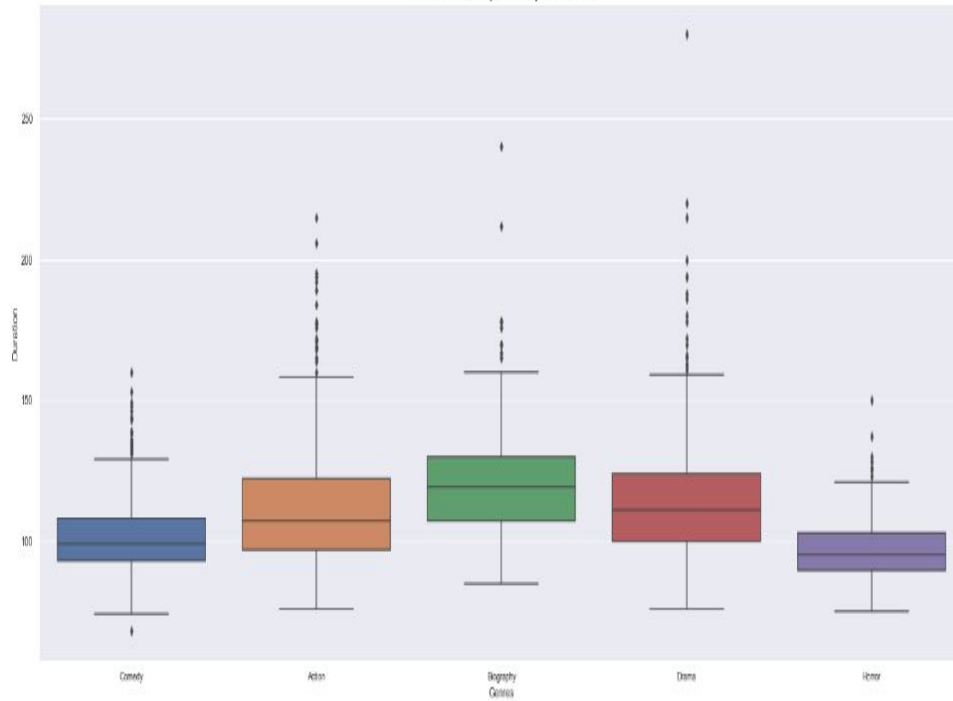
F_onewayResult(statistic=97.71610842367964, pvalue=3.1213217742449244e-77)



Duration Comparison per Genre



Duration Comparison per Genre



Check whether there is any significant difference between the mean duration of five genres

Extract individual groups

`group1 = genres_interested[genres_interested["Genres"] == "Comedy"]["Duration"]`

`group2 = genres_interested[genres_interested["Genres"] == "Action"]["Duration"]`

`group3 = genres_interested[genres_interested["Genres"] == "Biography"]["Duration"]`

`group4 = genres_interested[genres_interested["Genres"] == "Drama"]["Duration"]`

`group5 = genres_interested[genres_interested["Genres"] == "Horror"]["Duration"]`

`F_onewayResult(statistic=90.06549299404065, pvalue=1.674900414488818e-71)`



	Title	IMDB Score	Gross	Budget	Return of Investment
1961	Paranormal Activity	6.3	\$107,917,283.00	\$15,000.00	719,248.55%
2653	The Blair Witch Project	6.4	\$140,530,114.00	\$60,000.00	234,016.86%
2814	The Gallows	4.2	\$22,757,819.00	\$100,000.00	22,557.82%
2517	Super Size Me	7.3	\$11,529,368.00	\$65,000.00	17,537.49%
1352	In the Company of Men	7.3	\$2,856,622.00	\$25,000.00	11,226.49%

	Title	IMDb Score	Genres
2743	The Dark Knight	9.0	Action
4	12 Angry Men	8.9	Crime
2992	The Lord of the Rings: The Return of the King	8.9	Action
2991	The Lord of the Rings: The Fellowship of the Ring	8.8	Action
969	Fight Club	8.8	Drama





Conclusion





Conclusion

- H1: Data is considered reliable
- H2: G rated films had highest adjusted revenue
- H3: Drama films had the greatest decrease in revenue, action films had the greatest increase in revenue
- H4: There is a significant difference in IMDb scores per genre
- H5: There is a significant difference in durations per genre