

Investigation Report

❖ Introduction

- In the purpose of applying data analysis techniques we will use movies dataset to do a wrangling process to discover some insights and analytics regarding movies ratings

❖ Process

- Gather
 - We have kaggle dataset for movies
- Assess
 - Assessing the data visually and programmatically to define the issues that needs to be fixed in the data
- Cleaning the data
 - Clean each issue discover by the assessing step in order to get the right results

❖ Quality issues :

- We have duplicate `imdb_id` so we need to drop duplicates with same id
- We have nan values we replaced them with mean values
-

❖ Tidiness Issues:

- To use the genre , cast and production_companies columns we need to split them
- Unwanted columns we dropped them

❖ Conclusion

- The movies with more votes gets higher rates
- The more popular movies have more Gross profit
- Drama is the most common genre in the movies
- Director Woody Allen directed the highest number of movies
- Universal Pictures production company have the highest number of movies
- Actor Robert De niro made the highest number of movies
- Number of movies increased rapidly through the years