

# SUCCESS OF DIFFERENT MOVIE TYPES ANALYSIS

AISHA HAMID

APRIL 16,2023



# OVERVIEW

- THIS PROJECT AIMS TO ANALYZE A MOVIE DATASET TO IDENTIFY THE TYPES OF MOVIES THAT ARE PERFORMING WELL IN TERMS OF PRODUCTION BUDGET, WORLDWIDE GROSS, GENRE, RUNTIME AND RATING\_Y TO IDENTIFY TRENDS, PATTERNS AND RELATIONSHIPS. WE WILL USE EXPLORATORY DATA ANALYSIS TO GENERATE INSIGHTS FOR A BUSINESS STAKEHOLDER.



# BUSINESS PROBLEM

- MICROSOFT SEES ALL THE BIG COMPANIES CREATING ORIGINAL VIDEO CONTENT AND THEY WANT TO GET IN ON THE FUN. THEY HAVE DECIDED TO CREATE A NEW MOVIE STUDIO, BUT THEY DON'T KNOW ANYTHING ABOUT CREATING MOVIES. YOU ARE CHARGED WITH EXPLORING WHAT TYPES OF FILMS ARE CURRENTLY DOING THE BEST AT THE BOX OFFICE. YOU MUST THEN TRANSLATE THOSE FINDINGS INTO ACTIONABLE INSIGHTS THAT THE HEAD OF MICROSOFT'S NEW MOVIE STUDIO CAN USE TO HELP DECIDE WHAT TYPE OF FILMS TO CREATE.
- THE KEY BUSINESS QUESTIONS ARE;
- 1. WHAT IS THE MOST PROFITABLE MOVIE GENRE?
- 2. WHAT IS THE IDEAL PRODUCTION BUDGET TO PRODUCE A SUCCESSFUL MOVIE?
- 3. WHICH TYPE OF MOVIE HAS THE HIGHEST RATINGS?
- 4. DOES THE MOVIE RUNTIME INFLUENCE THE SUCCESS OF A MOVIE?



# DATA UNDERSTANDING

- THE DATASET WAS OBTAINED FROM VARIOUS LOCATIONS; [IMDB](#) , [BOX OFFICE MOJO](#) , [ROTTEN TOMATOES](#) , [THEMOVIEDB](#) AND [THE NUMBERS](#) .THE DIFFERENT FILES HAVE DIFFERENT FORMATS, SOME ARE COMPRESSED CSV (COMMA-SEPARATED VALUES) OR TSV (TAB-SEPARATED VALUES) FILES AND CONTAINS INFORMATION ON OVER 10,000 MOVIES. THE DATA IS ORGANIZED INTO SEVERAL TABLES AND SEVERAL COLUMNS CONTAINING DIFFERENT INFORMATION ABOUT THE MOVIES E.G. THE MOVIE TITLE, PRODUCTION BUDGET, GENRE, RUNTIME, AND RATING\_Y; EACH MOVIE HAVING A UNIQUE ID. THE FILES USED ARE MOVIES, MOVIE BUDGET AND REVIEWS. SOME OF THE CHALLENGES ENCOUNTERED DURING DATA PREPARATION INCLUDED MISSING VALUES, OUTLIERS AND PLACEHOLDERS.



# DATA CLEANING

- WE BEGAN BY FIRST CLEANING THE THREE DATASETS ; FOR THE MOVIE\_INFO DATASET COLUMNS DROPPED ARE BOX OFFICE, CURRENCY AND STUDIO AS THEY HAD 70% MISSING DATA. AFTERWARDS NULL VALUES WERE DROPPED FROM THE MOVIE\_INFO DATASET.
- FOR THE REVIEW COLUMNS , THE COLUMN OF INTEREST IS RATING; THE DATATYPE WAS CHANGED TO NUMERIC AND NULL VALUES DROPPED.
- FOR THE MOVIE BUDGET HAD NO NULL VALUES, THE COLUMNS OF INTEREST WERE PRODUCTION\_BUDGET , DOMESTIC\_GROSS AND WORLDWIDEGLOSS WHICH HAD PLACE HOLDERS THAT WERE REMOVED. THE DOLLAR SIGN AND COMMAS WERE REMOVED AS THEY ALTERED THE DATATYPE OF THESE COLUMNS.
- THE THREE DATASETS WERE THEN MERGED ON THE ID AND CHECKED FOR THE OUTLIERS WHICH WERE PRESENT BUT WERE NOT REMOVED AS THEY REPRESENT THE REAL DATA AND DO NOT AFFECT OUR FINDINGS.



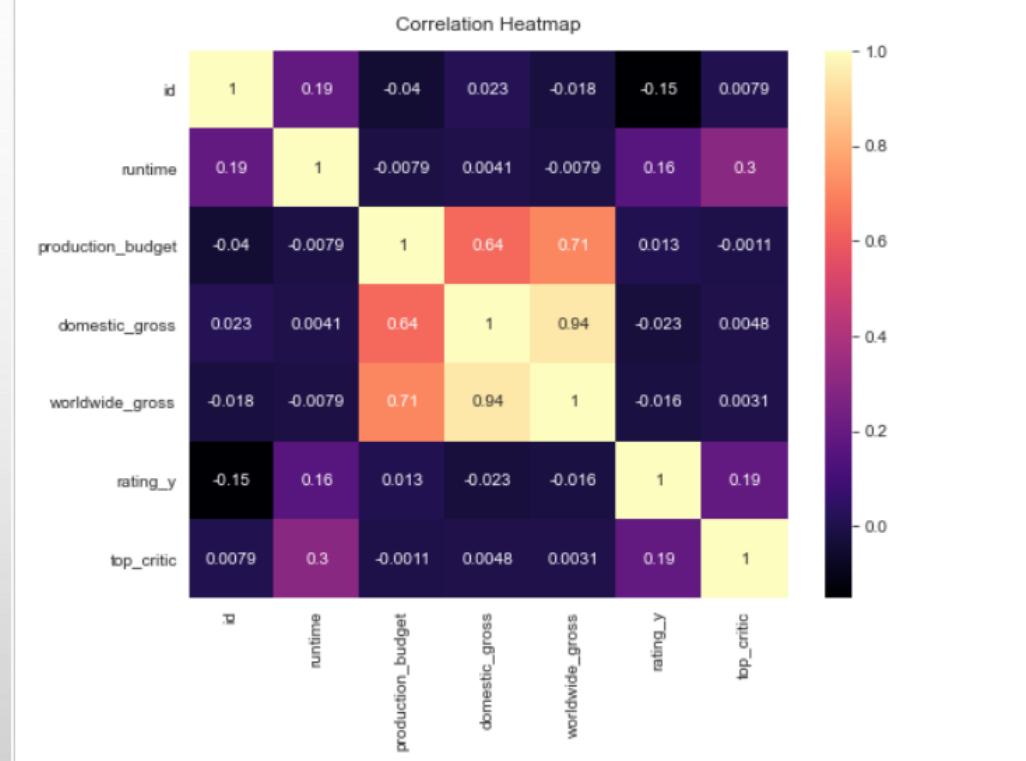
# DATA EXPLORATION

- THE **PRODUCTION BUDGET** OF MOVIES IN THE DATASET HAS THE FOLLOWING CHARACTERISTICS:
  - THE AVERAGE PRODUCTION BUDGET IS USD 33,931,150 WITH A VARIATION OF +/- USD 42,327,360.
  - THE LOWEST PRODUCTION BUDGET IS USD 1,100.
  - THE HIGHEST PRODUCTION BUDGET IS USD 350,000,000.
- **RUNTIME:** THE AVERAGE RUNTIME OF A MOVIE IS 111.8 MINUTES WITH A VARIATION OF +/- 18.2 MINUTES. THE HIGHEST RUNTIME OF A MOVIE IN THE DATASET IS 165 MINUTES, AND THE LOWEST RUNTIME IS 82 MINUTES.
- **GENRE:** THE MOST FREQUENT GENRE OF MOVIES IN THE DATASET IS ACTION AND ADVENTURE | MYSTERY AND SUSPENSE.
- **RATINGS:** THE AVERAGE RATING OF MOVIES IN THE DATASET IS 4.6



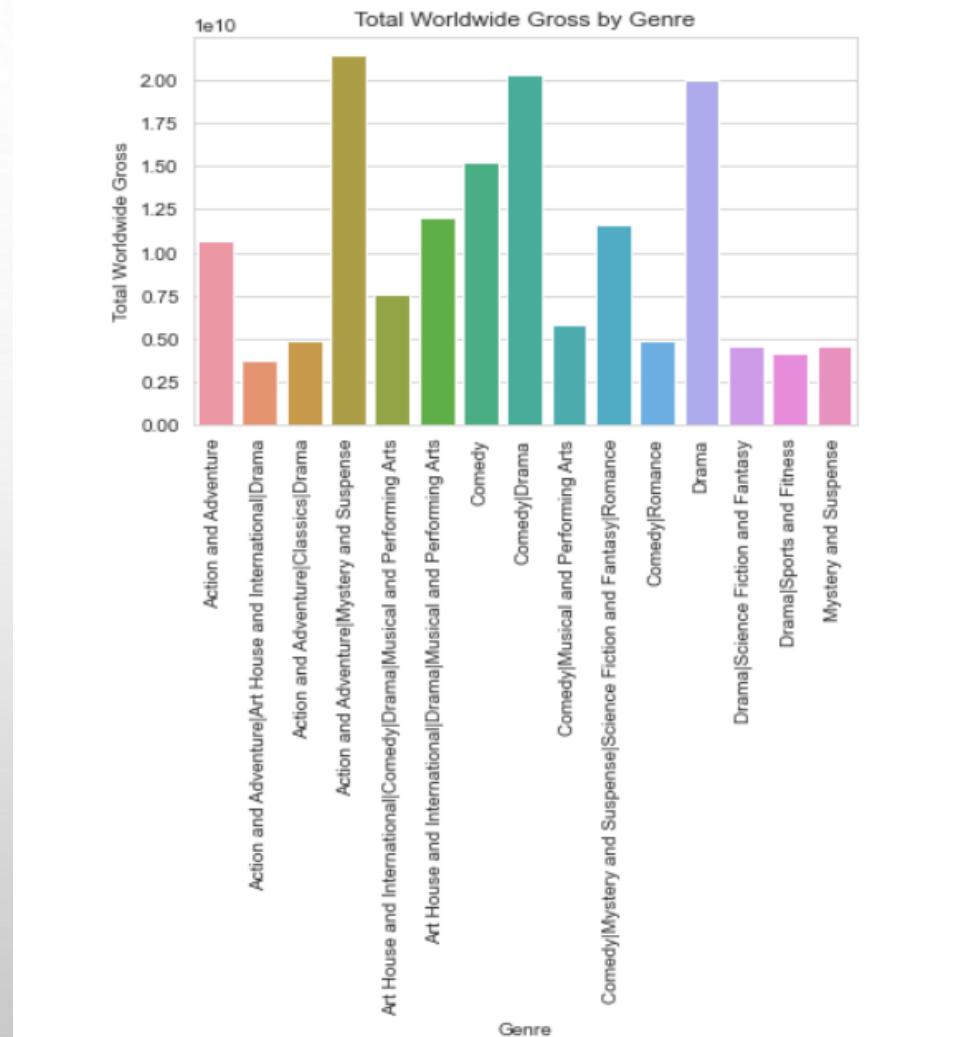
# CORRELATION ANALYSIS

- FROM THE CORRELATION HEATMAP;
- THERE IS A WEAK POSITIVE CORRELATION BETWEEN RUNTIME AND DOMESTIC\_GROSS (0.00405) AND A WEAK NEGATIVE CORRELATION BETWEEN RUNTIME AND WORLDWIDE\_GROSS (-0.0078).
- THERE IS A POSITIVE CORRELATION BETWEEN PRODUCTION\_BUDGET AND DOMESTIC\_GROSS (0.638) AND WORLDWIDE\_GROSS (0.710), MEANING THAT AS THE PRODUCTION BUDGET OF A MOVIE INCREASES, THE DOMESTIC AND WORLDWIDE GROSS ALSO TEND TO INCREASE.
- THERE IS A POSITIVE CORRELATION BETWEEN DOMESTIC\_GROSS AND WORLDWIDE\_GROSS (0.944), MEANING THAT AS THE DOMESTIC GROSS OF A MOVIE INCREASES, THE WORLDWIDE GROSS ALSO TENDS TO INCREASE.
- THERE IS A WEAK POSITIVE CORRELATION BETWEEN RATING\_Y AND TOP\_CRITIC (0.190), MEANING THAT MOVIES WITH HIGHER RATINGS TEND TO HAVE SLIGHTLY MORE TOP CRITICS.



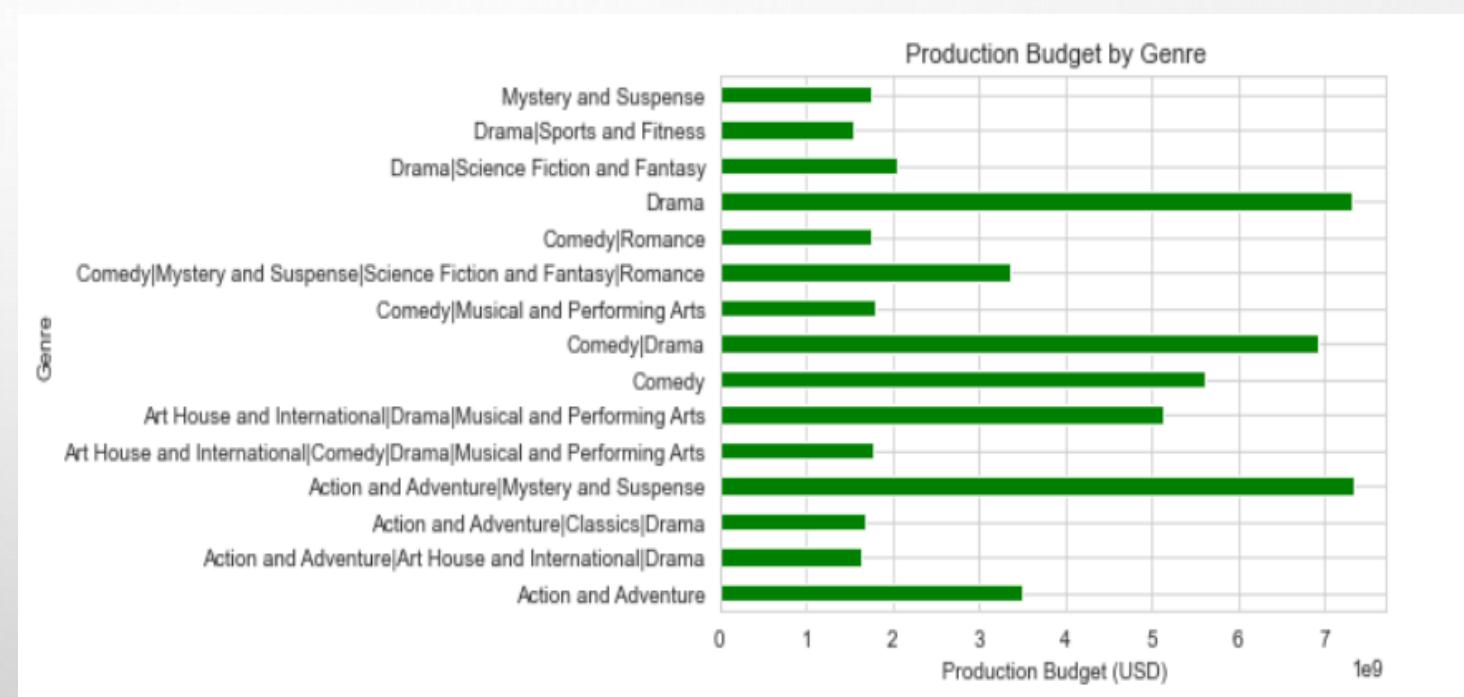
# GENRES AGAINST WORLDWIDE\_GROSS

- THE GENRE WITH THE HIGHEST WORLDWIDE\_GROSS IS ACTION AND ADVENTURE|MYSTERY AND SUSPENSE FOLLOWED BY DRAMA AND COMEDY DRAMA.
- THE GENRE WITH THE LEAST WORLDWIDE\_GROSS IS ACTION AND ADVENTURE|ART HOUSE AND INTERNATIONAL|DRAMA.



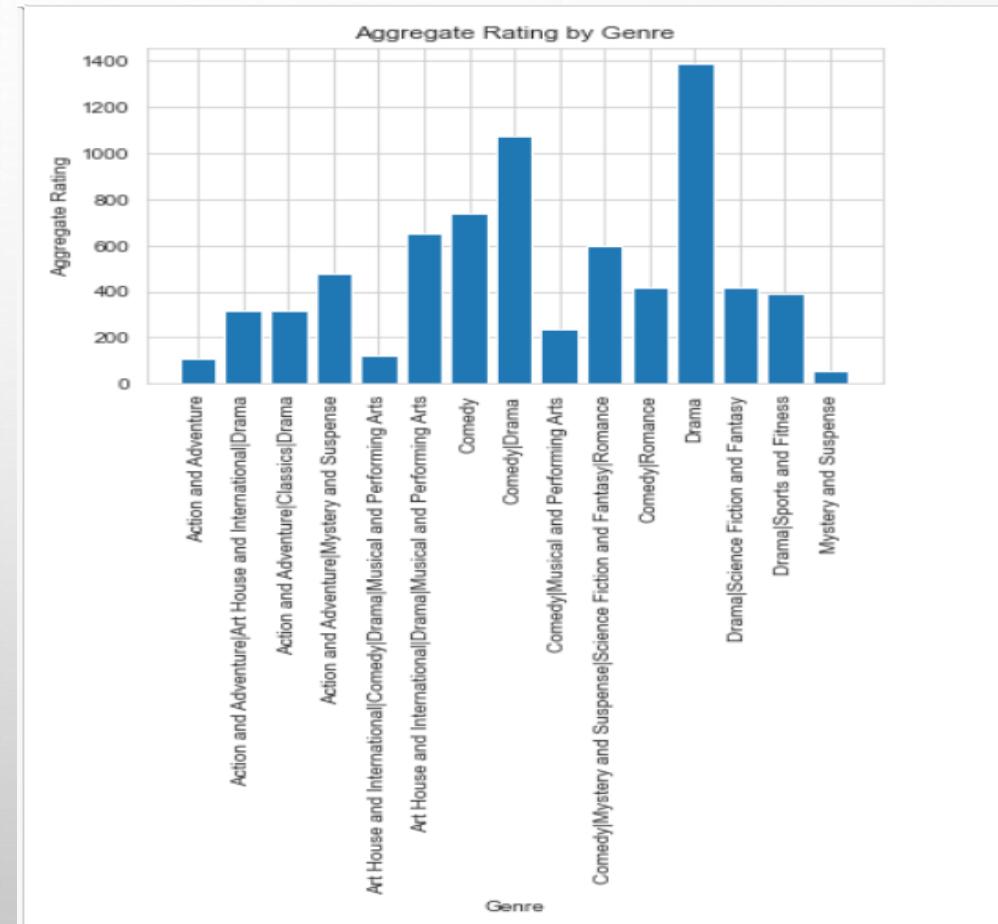
# GENRES AGAINST AGGREGATE PRODUCTION\_BUDGET

- THE GENRE WITH THE HIGHEST AGGREGATE PRODUCTION BUDGET IS DRAMA FOLLOWED BY ACTION AND ADVENTURE | MYSTERY AND SUSPENSE.
- THE GENRE WITH THE LOWEST MEAN PRODUCTION BUDGET IS ACTION AND ADVENTURE | ART HOUSE AND INTERNATIONAL | DRAMA.



# GENRE AGAINST AGGREGATE RATING\_Y

- MOVIE GENRE DRAMA HAS THE HIGHEST AVERAGE RATINGS.



# CONCLUSION

THIS ANALYSIS LED TO FOUR CONCLUSIONS;

- THE HIGHER THE PRODUCTION BUDGET THE HIGHER THE DOMESTIC AND WORLDWIDE GROSS OF A MOVIE
- THE HIGHER THE DOMESTIC GROSS THE HIGHER THE WORLDWIDE GROSS OF A MOVIE
- THE MOVIE GENRE THAT IS MOST PROFITABLE IS DRAMA AND ALSO HAS THE HIGHEST RATINGS.
- THE MOVIE RUNTIME DOES NOT INFLUENCE THE SUCCESS OF A MOVIE AS IT HAS A WEAK NEGATIVE CORRELATION BETWEEN BOTH WORLDWIDE GROSS AND PRODUCTION BUDGET OF -0.0078.



# RECOMMENDATIONS

OVERALL, IT IS IMPORTANT TO CONSIDER A VARIETY OF FACTORS WHEN MAKING DECISIONS ABOUT MOVIE PRODUCTION, MARKETING, AND DISTRIBUTION. WHILE THERE MAY BE SOME GENERAL TRENDS AND CORRELATIONS IN THE INDUSTRY, EACH MOVIE IS UNIQUE AND REQUIRES A TAILORED APPROACH TO MAXIMIZE ITS CHANCES OF SUCCESS.

- WHILE A HIGHER BUDGET MAY INCREASE THE CHANCES OF SUCCESS, IT IS NOT A GUARANTEE AND IT MAY BE MORE COST-EFFECTIVE TO FOCUS ON OTHER ASPECTS OF THE MOVIE'S RELEASE.
- WHILE DRAMA MAY BE A POPULAR AND SUCCESSFUL GENRE, IT IS IMPORTANT TO CONSIDER THE SPECIFIC STORY AND CHARACTERS OF EACH MOVIE AND DETERMINE THE MOST EFFECTIVE GENRE TO PRESENT THEM. OTHER GENRES SUCH AS ACTION, COMEDY AND HORROR MAY ALSO BE HIGHLY PROFITABLE AND SUCCESSFUL, DEPENDING ON THE AUDIENCE AND CRITICAL RECEPTION.
- WHILE MOVIE RUNTIME MAY NOT BE A SIGNIFICANT FACTOR IN A MOVIE'S SUCCESS, IT IS IMPORTANT TO CONSIDER THE SPECIFIC STORY AND PACING OF EACH MOVIE AND DETERMINE THE MOST EFFECTIVE LENGTH TO CONVEY THE DESIRED MESSAGE. A LONGER RUNTIME MAY BE EFFECTIVE FOR CERTAIN MOVIES, WHILE OTHERS MAY BENEFIT FROM A SHORTER, MORE FOCUSED APPROACH.
- WHILE DOMESTIC AND WORLDWIDE GROSSES MAY BE RELATED, IT IS IMPORTANT TO CONSIDER THE SPECIFIC TARGET AUDIENCE FOR A MOVIE AND TAILOR MARKETING AND DISTRIBUTION EFFORTS ACCORDINGLY. SOME MOVIES MAY PERFORM BETTER IN CERTAIN REGIONS OR COUNTRIES, AND IT MAY BE MORE EFFECTIVE TO FOCUS ON THESE MARKETS RATHER THAN TRYING TO APPEAL TO A BROADER AUDIENCE.

# NEXT STEPS

FURTHER ANALYSIS COULD YIELD ADDITIONAL INSIGHTS TO IMPROVE PREDICTIONS OF SUCCESSFUL MOVIES

- STAY UP-TO-DATE WITH INDUSTRY TRENDS AND CHANGES, AND ADJUST STRATEGIES AS NEEDED TO REMAIN COMPETITIVE. THE MOVIE INDUSTRY IS CONSTANTLY EVOLVING, AND IT IS IMPORTANT TO STAY INFORMED AND ADAPT TO NEW DEVELOPMENTS AND CHALLENGES AS THEY ARISE.
- CONDUCT FURTHER RESEARCH AND ANALYSIS TO BETTER UNDERSTAND THE SPECIFIC FACTORS THAT INFLUENCE THE SUCCESS OF MOVIES IN A TARGET MARKET E.G. LOOKING AT INDUSTRY DATA, CONDUCTING SURVEYS OR FOCUS GROUPS WITH TARGET AUDIENCES, OR ANALYZING CRITICAL REVIEWS AND SOCIAL MEDIA SENTIMENT.

