

# ANALYSIS OF MOVIE DATA

## OVERVIEW

- The Microsoft company is looking to venture in the movie industry by establishing a new movie studio but they have no prior experience in making movies. They need to understand what types of films are doing well at the box office to make informed decisions on what type of films they should create. The objective is to identify the most successful types of films in the current market and provide actionable insights to help Microsoft make decisions that will increase their chances of success. To achieve this goal, data on box office performance and movie characteristics will need to be collected, cleaned

# DATA

- The movie data was collected from various sources and available in the folder zipped data. It includes datasets from box office mojo, IMDb, Rotten tomatoes, The movieDB, and The Numbers. The data files have different formats, including compressed CSV, TSV and SQLite database files.
- To make the analysis less challenging we will use the following data files :
  - 1.im.db.zip: This is a zipped SQLite database file that contains relevant data from IMDb. The movie \_basics and movie \_ratings tables are the most relevant in our analysis
  - 2.bom.movie\_gross.csv.gz. This is a compressed csv file containing data on movie box office gross earnings. It can be opened using `pd.read_csv` without expanding the file

# BUSINESS UNDERSTANDING

- Microsoft wants to enter a highly competitive movie industry and establish a new movie studio but the lack experience in creating movies . They need to understand what types of films are popular and successful currently on box office . This understanding will help Microsoft to make better decisions on what types of films they should create to increase their chances of success . The goal of the project is to provide insights on the current trends in the movie industry and identify movies doing well at box office . This analysis will enable Microsoft make data driven decisions on what type of movies they should make to increase their chances of being successful.

# DATA UNDERSTANDING

- The data understanding of this project involves Identifying and acquiring relevant data that can be used to gain insights into this movie industry . The data will be cleaned and analyzed to identify trends and patterns that can help Microsoft understand what types of films are performing well at the box office.
- The data required for this project may include box office data , movie characteristics such as genre , budget , release date , and ratings as well as data on marketing strategies and audience demographics .This data can be sourced from various publicly available datasets such as IMDb , Box office mojo and The numbers.
- The data will be cleaned to ensure that it is accurate ,complete and consistent . This may involve in removing duplicates ,correcting errors and filling missing values . Once the data is cleaned it can be analyzed using statistical techniques to identify trends and patterns . The insights provided from the data analysis will help Microsoft what type of films are successful in box office and inform the decision making process for the new movie studio

# DATA ANALYSIS

- To analyze this data and provide actionable insights for the new movie studio, we can use the following steps:

## **1. Data cleaning and processing:**

We need to clean and preprocess the data to remove any duplicate missing values or incorrect data. We can use pandas, numpy and SQL to clean and preprocess data

## **2.Data intergration:**

We need to intergrate data from the various sources and combine them into a single dataset. We merge data using tools like SQL or pandas

## **3.Exploratory data analysis**

# RECOMMENDATIONS

- Based on exploratory data analysis and statistical modelling , the following recommendations can be made :
- Focus on production action ,adventure and thriller movies as these genres tend to have higher box office revenues compared to others
- Hire popular directors and actors with a proven track record of success in these genres, as they are likely to attract a larger audience
- Plan to release movies during the summer and holiday seasons as these periods tend to have higher box office revenues
- consider producing franchises as this tend to perform well and have built in audiences
- Stay up to date with current movie trends and audience preferences by conducting regular market research and analyzing box office