

# DATA ANALYSIS SUMMER INTERNSHIP - 2023



## A DATA-DRIVEN ANALYSIS OF IMDB RATINGS

---

### MENTORS:

- K. NARMADA MANI, APSSDC
- K. MEENAKSHI, APSSDC

### COLLEGE:

GODAVARI INSTITUTE OF ENGINEERING AND TECHNOLOGY, AP

### BATCH: 90

MD. NUSRATH SHARIFF	20551A4231
U. DHARANI	20551A4252
K. MAHESH	20551A4229
AISHA MARYAM	21551A4202

# CONTENT

- 1 PROBLEM STATEMENT
- 2 SYSTEM REQUIREMENTS
- 3 ARCHITECTURE OF THE PROJECT
- 4 ADVANTAGES
- 5 ANALYSIS & VISUALIZATION
- 6 CONCLUSION

# PROBLEM STATEMENT

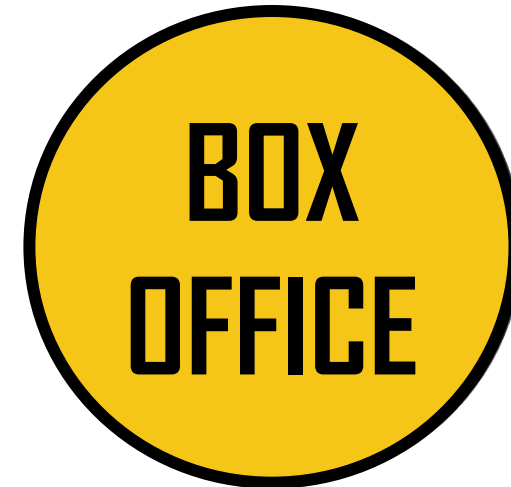
# STATEMENT:

---

This project aims to explore the intricate **relationship between IMDb ratings and movie gross**, unraveling the role of critical acclaim versus content quality in shaping the commercial triumph of films. Through comprehensive data analysis, we seek to provide valuable insights that guide decision-making and contribute to the evolving narrative of the film industry.



*VS*



# IMDb

- ❑ IMDb is a prominent online database for films and TV shows, offering comprehensive details and user ratings.
- ❑ IMDb features user-generated ratings and reviews, aiding audiences in making informed viewing choices.

# BOX OFFICE

- ❑ Box office refers to the revenue generated by a movie's theatrical release, reflecting its commercial success.
- ❑ The box office performance indicates a film's popularity and financial impact within the entertainment industry.

# SYSTEM REQUIREMENTS

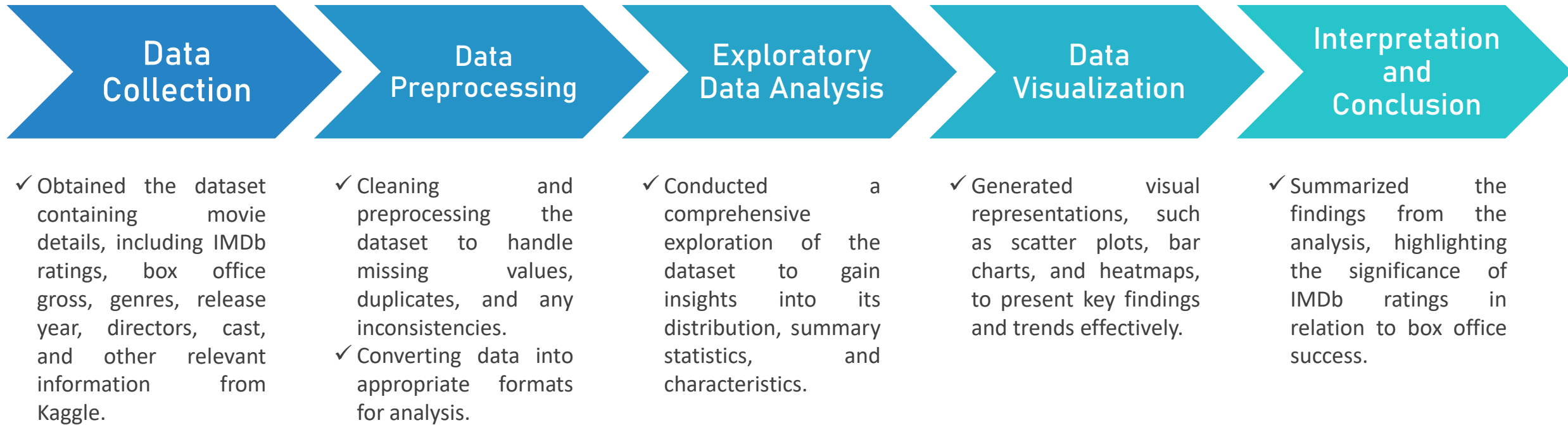
# HARDWARE & SOFTWARE REQUIREMENTS



OPERATING SYSTEM	Microsoft Windows 11 Home Single Language
CPU	Intel(R) Core(TM) i5-10300H CPU @ 2.50GHz, 2496 Mhz, 4 Core(s), 8 Logical Processor(s)
RAM	16GB
HARD DRIVE	250GB SSD
SOFTWARE	Jupyter Notebook, Microsoft Excel
PROGRAMMING LANGUAGE	Python

# ARCHITECTURE OF THE PROJECT





# LIBRARIES USED...

**NumPy**

It is used to efficiently process and manipulate numeric data, laying the groundwork for further analysis.

**pandas**

Pandas is essential for loading, cleaning, and transforming the movie dataset into a structured format, making it a vital tool for exploratory data analysis.

*matplotlib*

Matplotlib.pyplot is a widely used library for data visualization in Python.

**seaborn**

Seaborn is a statistical data visualization library built on top of Matplotlib.

# ADVANTAGES

# KEY ADVANTAGES

## Understanding Movie Success Factors:

- The project sheds light on the complex dynamics that influence a movie's commercial trajectory.

## Resource Optimization:

- Studios can allocate resources more efficiently based on the analysis, focusing on projects that align with audience preferences and have higher chances of commercial success.

## Engaging Content:

- Correlation aids in creating resonant content.
- Unveiling genre preferences and content attributes that drive success provides a competitive edge in content creation.

## Competitive Advantage:

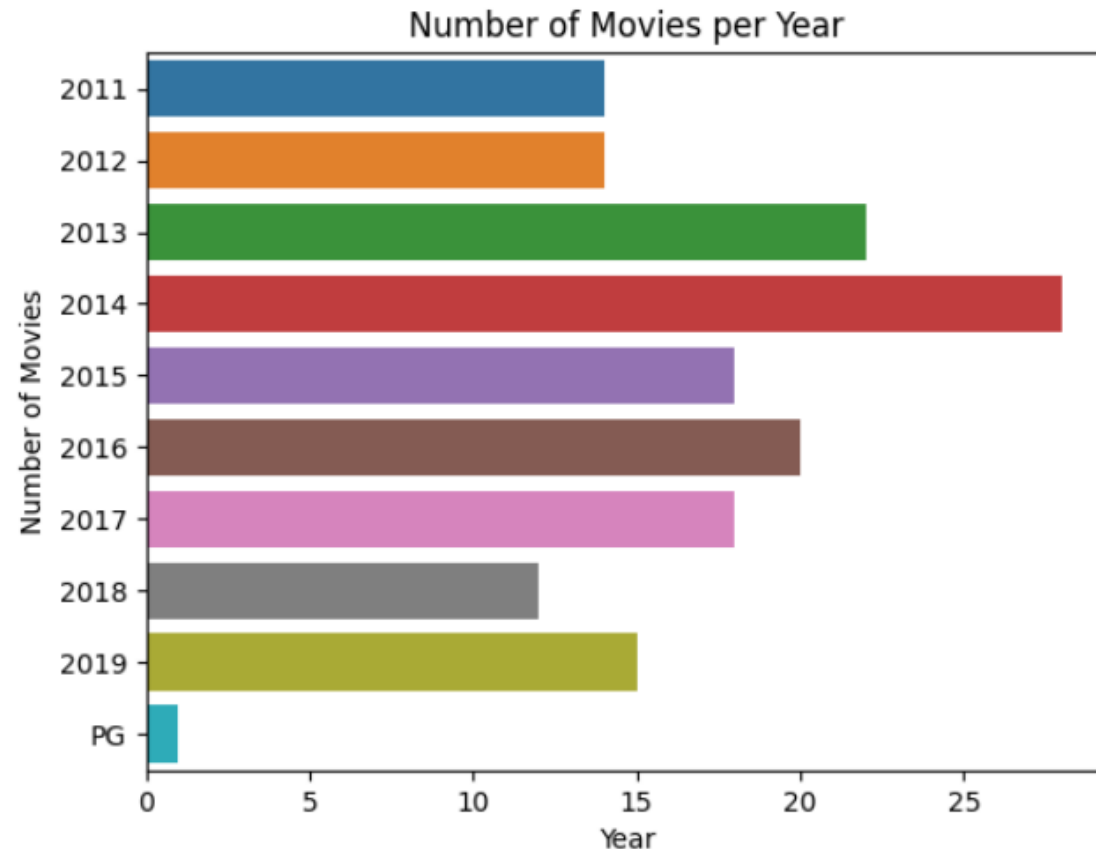
- Studios and filmmakers who use data to understand audience sentiments can adapt to changing preferences, stay ahead of trends, and create content that resonates with viewers.

## Valuable Industry Insights:

- Sharing insights with industry professionals fosters collaboration and continuous improvement in the pursuit of cinematic excellence.

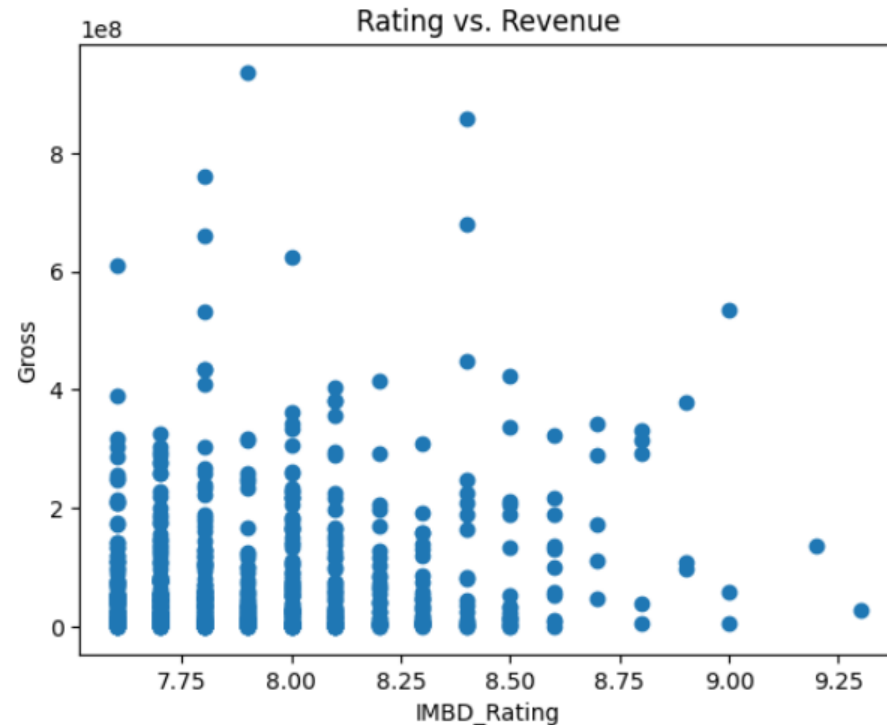
# ANALYSIS & VISUALIZATION

# Displaying number of movies released in each year for the last 10 years i.e., (2011-2019).



- ✓ The number of movies released per year shows an upward trend from 2011 to 2014.
- ✓ The year 2014 had the highest number of movie releases, with a count of 28 movies.
- ✓ Sudden decrease in movies in 2015 is due to people exposure to new all kinds of cinema and decreases the remake rate.
- ✓ The data suggests that the movie industry experienced growth and increased production over the years, with a peak in 2016.

# DOES RATING EFFECT THE REVENUE?



← SCATTER PLOT

*correlation*

0.1259915773854415

↑  
CORRELATION VALUE

A correlation coefficient of **0.125** suggests a positive but **weak correlation** between the movie's rating and its revenue. Hence, we can say that **rating does not greatly affect the revenue of the film.**

# CONCLUSION



# Interpretation and Conclusion...

❑ As we have taken the IMDb dataset for analyzing top grosser films over past 100 Years.

- In summary, our analysis indicates that reviews have **limited direct impact** on box office outcomes.
- Instead, our findings emphasize the crucial role of captivating **content in attracting audiences**.
- Compelling genres, narratives, and performances appear to be more influential in driving box office success.

**“This underscores the significance of engaging storytelling and content quality in shaping audience preferences and cinematic triumph”**

**THANK YOU !!!**