



# Data Analysis of a Movie Dataset

Presented by Team Data Science and Analyst G1

## Members

- Obaid Sayyed
- Nand Joshi
- Ashritha Mudam
- Satyam Gadhave
- Vikash Kumar
- Rishabh

# Introduction & Objectives

Our objective was to extract meaningful insights from a comprehensive movie dataset to understand industry trends and audience preferences.



## Identify Key Trends

Uncover popular genres and rating trends over the years.



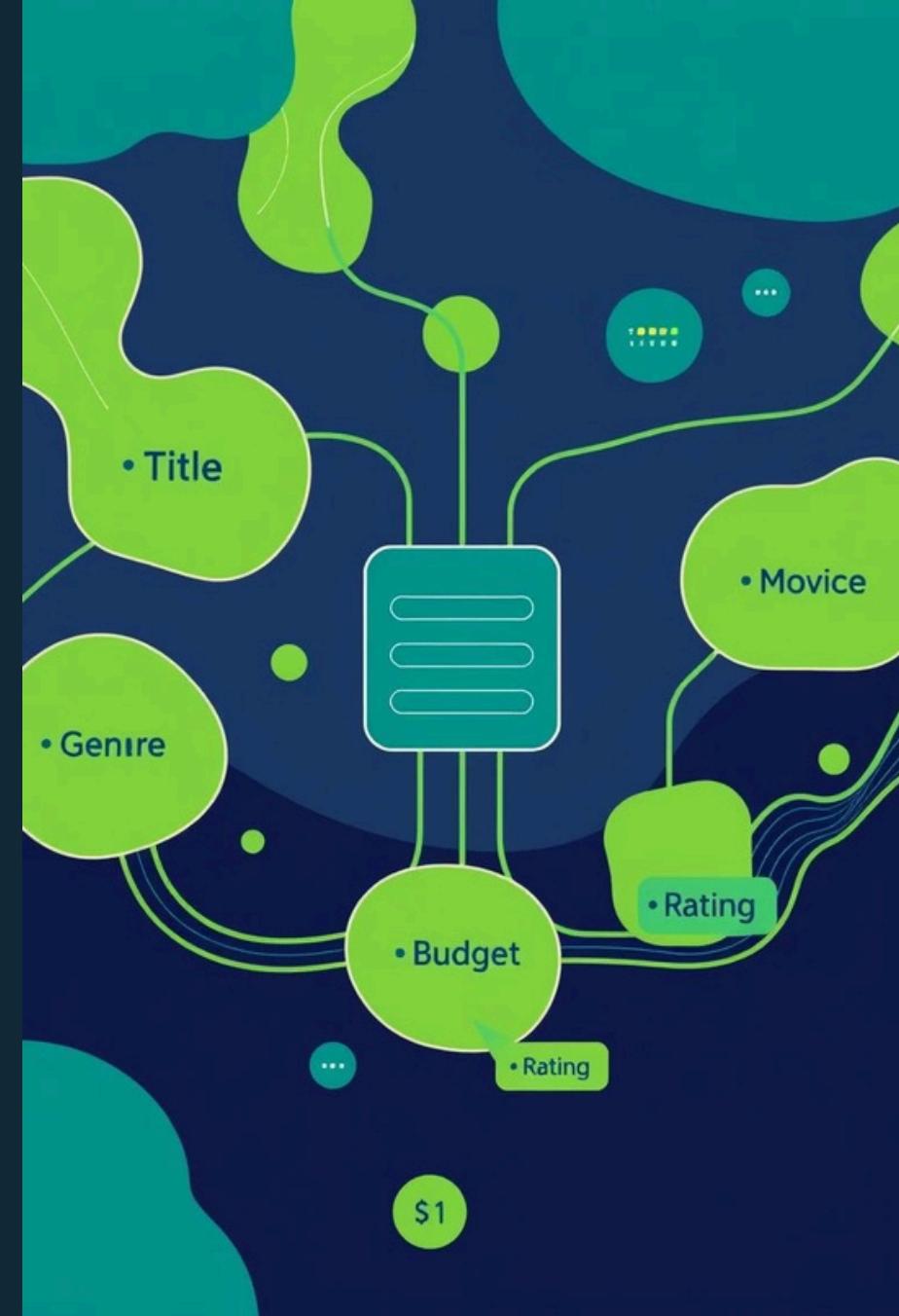
## Audience Insights

Top ratings, top movies and no of votes.

# Raw Dataset Description

We utilized a publicly available movie dataset from Kaggle to fuel our analysis.

- **Source:** Kaggle
- **Number of Records:** 63,249 films(rows) x 23 fields (columns)
- **Time Period:** Spanning from early cinema to 2025
- **Key Attributes:**
  - Movie Title
  - Release Year
  - Genres
  - Votes
  - Ratings



# Data Cleaning & Preprocessing

Ensuring data quality was crucial for accurate insights.

## Handling Missing and Duplicate Values

Handled around 34,500 null and duplicate values to gain perfect insights out of the data.

## Data Type Conversion

Converted columns like 'release year' to datetime objects and numeric strings like votes and ratings to integers/floats.

# Glimpse of Dashboard

Visualizing the data revealed immediate patterns and distributions.



# Key Insights & Findings

Our analysis uncovered several significant trends in the movie industry.

## Genre Popularity

Drama was the most popular genre not just having the highest votes but highest ratings as well.

## Rating Trends Over The Years

Rating trend has seen a major dip over the years with initially climbing up but falling down since 1975. The highest average rating was in the years 1947 and 1956 with 8.6 whereas the lowest can be seen in the year 2025(current) with 5.95.

## Critical Reception Impact

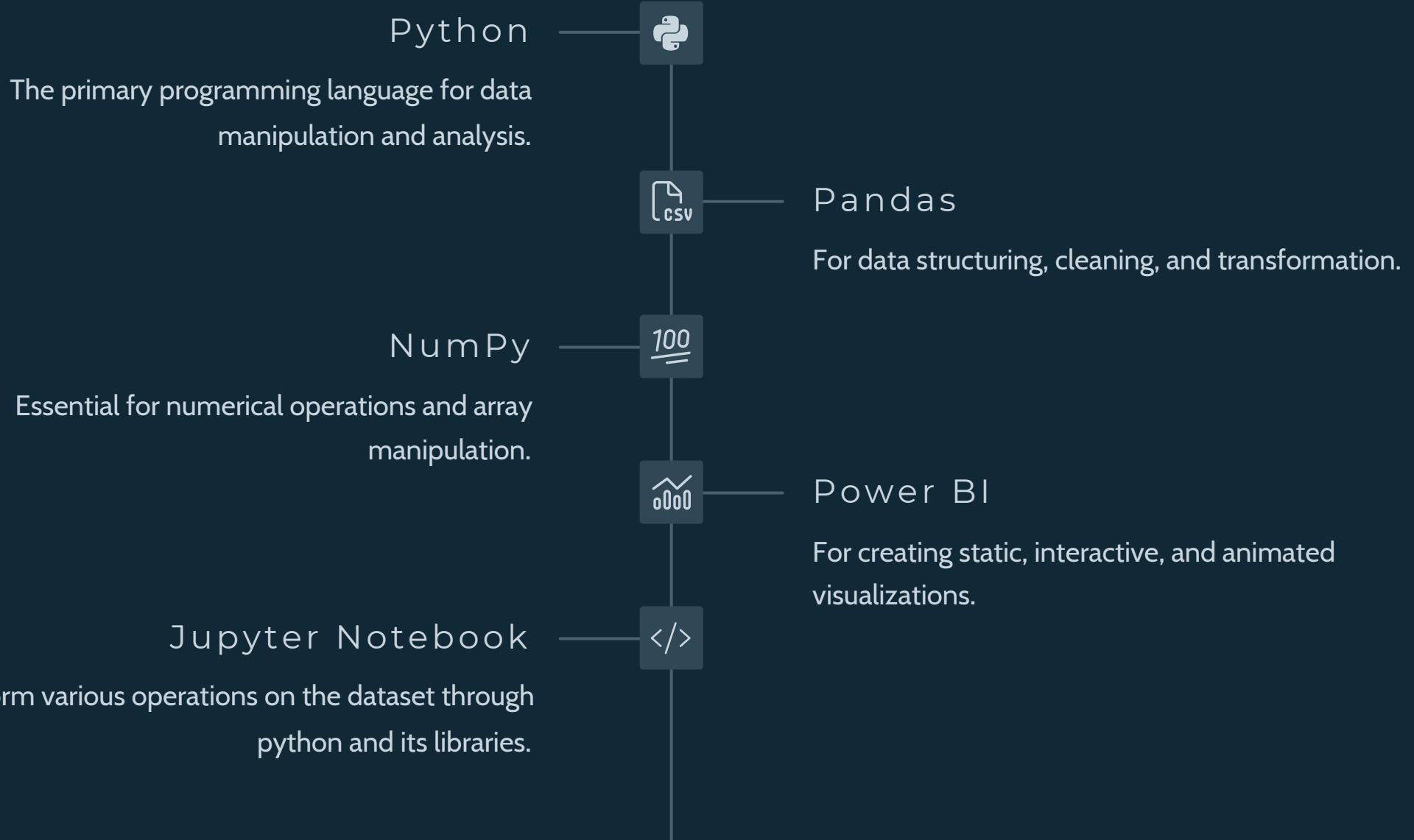
Films with higher average vote scores also tended to have a higher count of votes to their names.

## Top10Movies

The top 10 movies ranged from a rating of 9.3 till 8.8

# Tools & Techniques Used

Leveraging powerful data science tools for efficient analysis.



# Team Contributions

A collaborative effort made this project a success.

## Obaid Sayyed (Lead)

- Visualizations
- Dashboard furnishing
- Final Project

## Nand Joshi

- Dataset selection
- Dataset cleaning
- Helped in EDA

## Aashritha Mudam

- Exploratory Data Analysis (EDA)
- Quality assurance for code
- Dashboard testing

## Rishabh

- Dataset selection
- Fields selection

## Vikash Kumar

- Dashboard making
- Dashboard testing
- KPIs

## Satyam Gadhav

- Dashboard making
- Dashboard testing

# Challenges & Solutions

Overcoming hurdles to ensure project success.

## Inconsistent Data Formats

**Solution:** Developed robust parsing functions using Python's string manipulation and regular expressions.

## Large Dataset Size

**Solution:** Optimized Pandas operations and utilized chunking for memory efficient processing.



# Thank You!

"Thank you Launchspring  
for giving us this  
opportunity to work as  
interns."

We appreciate your support and  
guidance throughout this project.