

# Analyzing Movie Trends: IMDb Ratings, Genres, and More

Hello! I'm Adon Banker, a passionate aspiring data scientist. You can connect with me on

[Click here](#)

and explore my projects on [GitHub](#). My goal is to leverage data science for understanding viewer preferences and movie industry trends, and this project is a step in that direction.

**A** by Adon banker





# Project Overview: Unveiling Movie Success Factors



## Project Objective

Investigate the impact of genres, runtime, and release year on IMDb ratings and votes.



## Data Source

Comprehensive movie dataset from IMDb (Kaggle open source).



## Tools Used

Python with Pandas, Matplotlib, Seaborn, and Scikit-learn.

petecset codd



Missing Values

Telet

☐ Miledl VF 2 desposstates med/redting

☐ Saredgisney2Loflecast or denfreding

Strovsirget Stures

Telet

☐ Splealt Statcooflecast - pred/redng

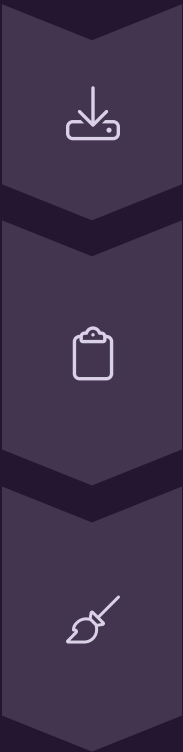
☐ Smeal Trutes

Teles

Structured Viue is &- webaly,  
Irtin JLoftedest or tenlirdding



# Data Collection and Cleaning



## IMDb Dataset

Over 5,000 movies with rich metadata.

## Key Variables

Title, genre, runtime, release year, IMDb rating, and number of votes.

## Data Cleaning

Addressed missing values and inconsistent formats for accuracy.

# Exploratory Data Analysis (EDA)

6.85

Average IMDb Rating

Standard deviation of 1.15 across all movies.

45K

Average Votes

Standard deviation of 85,000 indicates wide variance.

105

Average Runtime

Movies typically run about 105 minutes.

Correlation analysis revealed interesting relationships between numerical features.

Genre distribution highlights Drama (40%), Comedy (35%), and Action (25%) as dominant categories.

# Genre Impact on IMDb Ratings

## Highest Rated Genres

- Animation (7.5)
- Documentary (7.3)
- Biography (7.2)

## Lowest Rated Genres

- Horror (5.8)
- Action (6.2)
- Thriller (6.4)

Genre combinations, particularly Drama/Romance, frequently achieve higher scores. "The Shawshank Redemption" (Drama) exemplifies this with a 9.3 rating and 2.7M votes.

# Release Year & Runtime Influence

## Ratings Peak

IMDb ratings peaked in the 1990s.

## Statistical Significance

Pearson correlation of 0.45 ( $p < 0.05$ ).



## Declining Trend

A slight decline in ratings since the 90s.

## Runtime Correlation

Moderate positive correlation with ratings.

"Pulp Fiction" (1994) serves as a prime example, achieving an 8.9 rating with 2.1M votes.

# Predictive Modeling: Predicting IMDb Ratings

1

## Linear Regression Model

Genre, runtime, and release year as predictors.

2

## Model Performance

R-squared of 0.65, explaining 65% of variance.

3

## Feature Importance

Genre most influential, followed by runtime and release year.

Our model can predict an average rating of 7.8 for a movie with specific attributes.



## Conclusion: Trends and Insights

### Key Influencers

Genre, runtime, and release year significantly impact IMDb ratings.

### Strategic Decisions

Data-driven insights can inform movie production decisions.

### Future Work

Sentiment analysis of reviews and deeper genre dives.





## Background

I'm currently a **Grade 12 student in the International Baccalaureate Diploma Programme (IBDP)**.

As part of my curriculum, I study subjects like **Mathematics AA HL, Computer Science, and Physics**, which have helped me build strong analytical and problem-solving skills. I've also been actively exploring programming and real-world applications of data.



## Why I Want to Learn Data Science

I want to learn Data Science because I'm passionate about using data to **uncover insights and solve real-world problems**. Whether it's understanding consumer behavior, trends in media, or optimizing systems, I find it fascinating how data-driven decisions impact industries.

I especially enjoy working with tools like **Python, Pandas, and Matplotlib**, and I'm excited to continue building projects that explore meaningful datasets.



## Work Experience / Projects

While I don't have formal work experience yet, I've worked on several **self-initiated and academic projects**.

For example, in my **IMDb Movie Data Analysis project**, I scraped and cleaned real-world data, handled missing values and outliers, and performed exploratory data analysis (EDA) using Python. I visualized trends in ratings, runtimes, and genres to better understand what contributes to a movie's success.

I'm also involved in extracurriculars where I use digital tools to organize community projects and document local heritage, showing how data and tech can help in various fields.