AIML_2024 IMDB DATA SCRAPING AND VISUALIZATIONS

Project Documentation

March 2025

Prepared by: Bhuvanesh Thangaraj BSc (CS)

Mini Project Given By GUVI

WELCOME

Table of Contents

1. Abstract

This document outlines the development of an interactive data analysis application using

Streamlit. The project provides real-time visualizations for movie datasets, including ratings,

voting trends, and genre-based insights.

2. Introduction

With the growth of digital entertainment, analyzing movie trends has become essential. This

project leverages Python and Streamlit to visualize patterns in movie ratings, voting behavior,

and correlations, offering an intuitive interface for data exploration.

3. Project Objectives

The key objectives of this project are:

• Develop an interactive dashboard for movie data analysis.

• Enable dynamic filtering for enhanced user insights.

• Ensure efficient processing of large datasets.

• Maintain a modular and scalable codebase.

• Enhance user experience through intuitive design.

4. Technology Stack

Python, Streamlit, Pandas, Matplotlib & Seaborn, SQLite, Plotly

5. Data Sources & Preprocessing

The dataset is derived from multiple sources, including IMDb. Data preprocessing includes

handling missing values, normalizing formats, and structuring tables for efficient querying.

6. Features & Functionalities

• Top 10 Movies by Rating & Voting Counts

• Genre Distribution Analysis

• Voting Trends by Genre

• Interactive Filtering

• Heatmap for Ratings Correlation

7. Implementation Details

The project follows a modular approach with separate components for data processing,

visualization, and database interactions. Optimized SQL queries enhance performance.

8. Results & Findings

The platform successfully identifies key trends in movie data, such as rating distributions, genre

popularity, and voting behaviors. The interactive filtering allows users to extract meaningful

insights dynamically.

9. Challenges & Solutions

• **Handling Missing Data:** Implemented imputation techniques for missing values.

• **Performance Optimization:** Used indexing and optimized SQL queries for large datasets.

• **User Experience:** Designed a responsive UI for seamless interaction.

10. Future Scope

• Integration with real-time movie databases.

• Advanced ML-based rating predictions.

• Additional user-customized reports and insights.

11. Conclusion

The project successfully delivers an interactive and insightful movie data analysis tool. Future

iterations may introduce real-time data integration and predictive analytics.

12. References

• IMDb dataset

• Streamlit documentation

• Python data analysis libraries

END OF REPORT

THANKYOU