



MovieLens Dataset Analysis Report

Data Analysis and Dashboard Report using Python

Exploratory Data Analysis (EDA)

Prepared by: Osolake Mariam Omotolani

1. Introduction

This report provides an analytical overview of the MovieLens dataset. It explores user ratings, movie genres, and overall trends within the dataset. The dataset was analyzed using Python for data extraction and visualization. The goal is to identify viewing patterns, rating behaviors, and other insightful relationships among movies, users, and categories.

2. Dataset Overview

The MovieLens dataset contains information about movies, users, genres, and ratings. It includes key columns such as:

- **userId** – Unique identifier for users
- **movieId** – Unique identifier for movies
- **rating** – User-assigned score (typically between 0.5 and 5.0)
- **timestamp** – When the rating was submitted
- **genre** – Movie category (e.g., Action, Comedy, Drama)

3. Python Analysis

The following code were executed to explore and analyze the dataset:

- To calculate the average rating per movie.
- To find the number of ratings per genre.
- To identify the top 5 highest-rated movies.
- To determine the number of unique customers who purchased or rated movies from each category.

4. Key Performance Indicators (KPIs)

The following KPIs were created in Power BI to monitor and measure performance metrics:

1. Total Movies Rated
2. Total Users
3. Average Rating
4. Number of Ratings per Genre

5. Unique Customers per Category

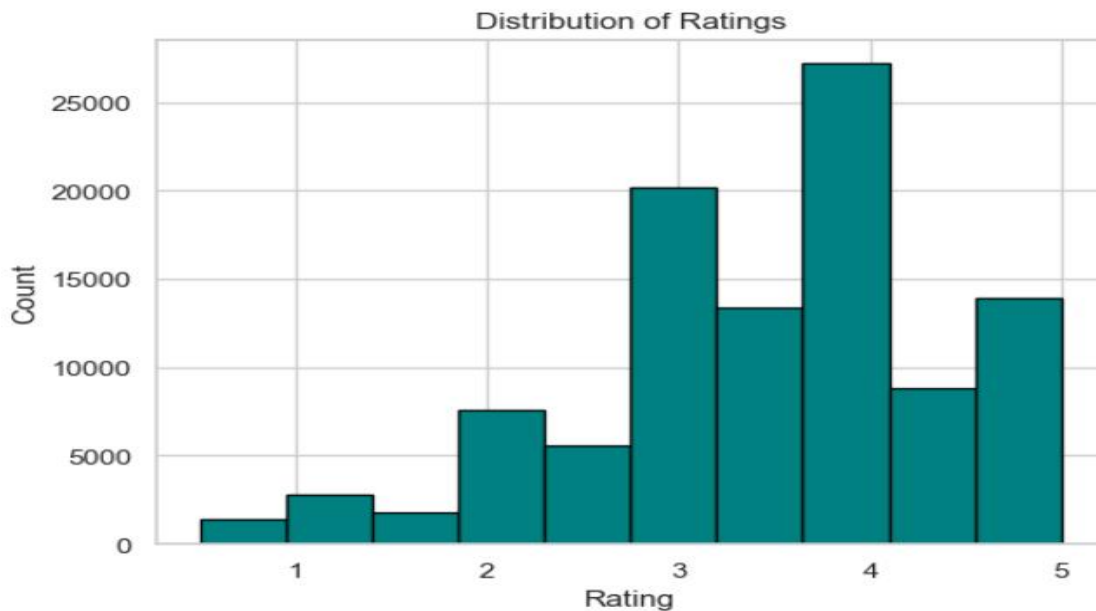
6. Top 5 Movies by Rating

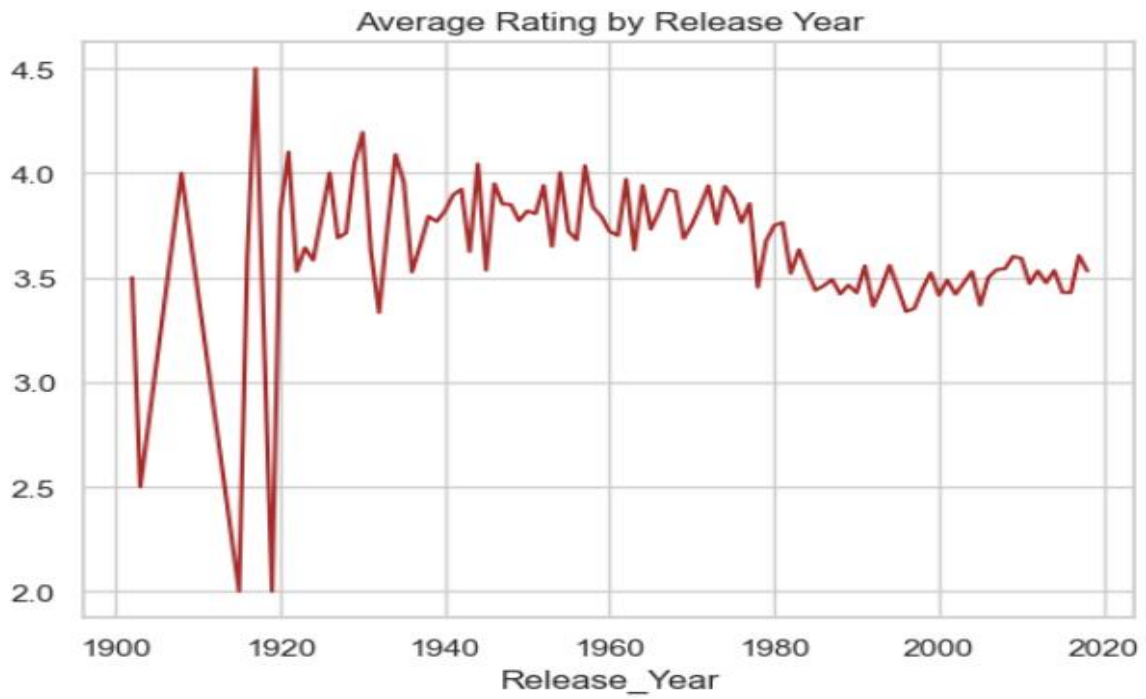
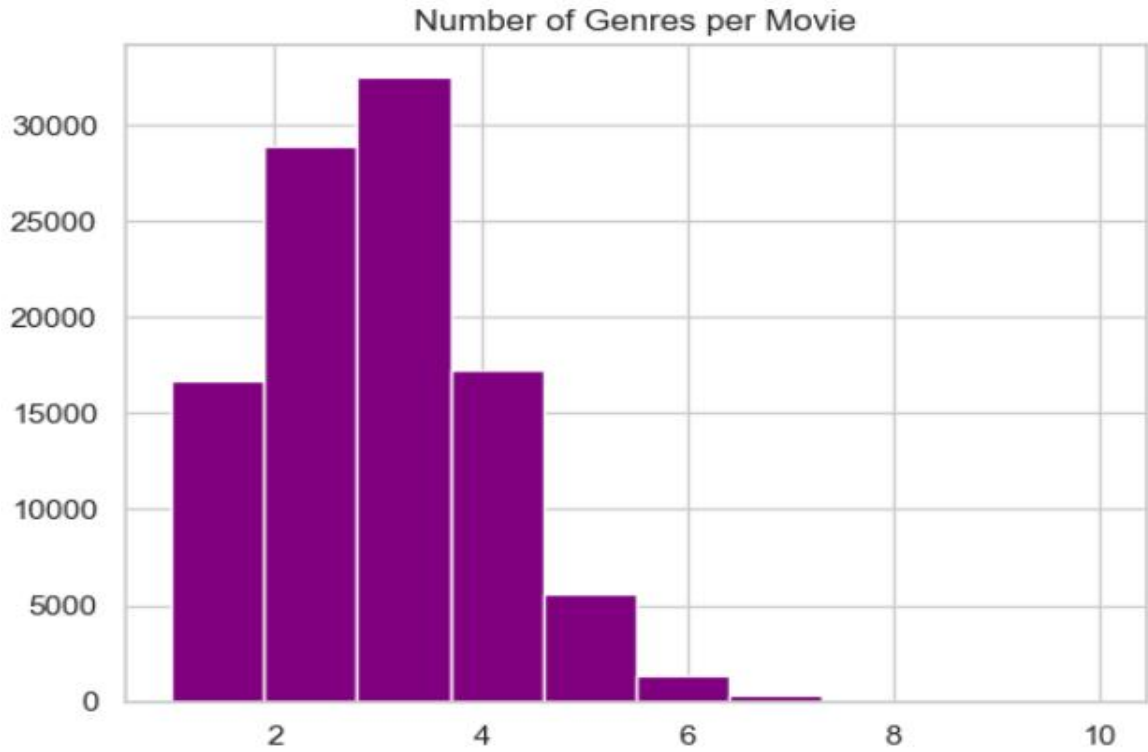
5. Visualizations and Dashboard

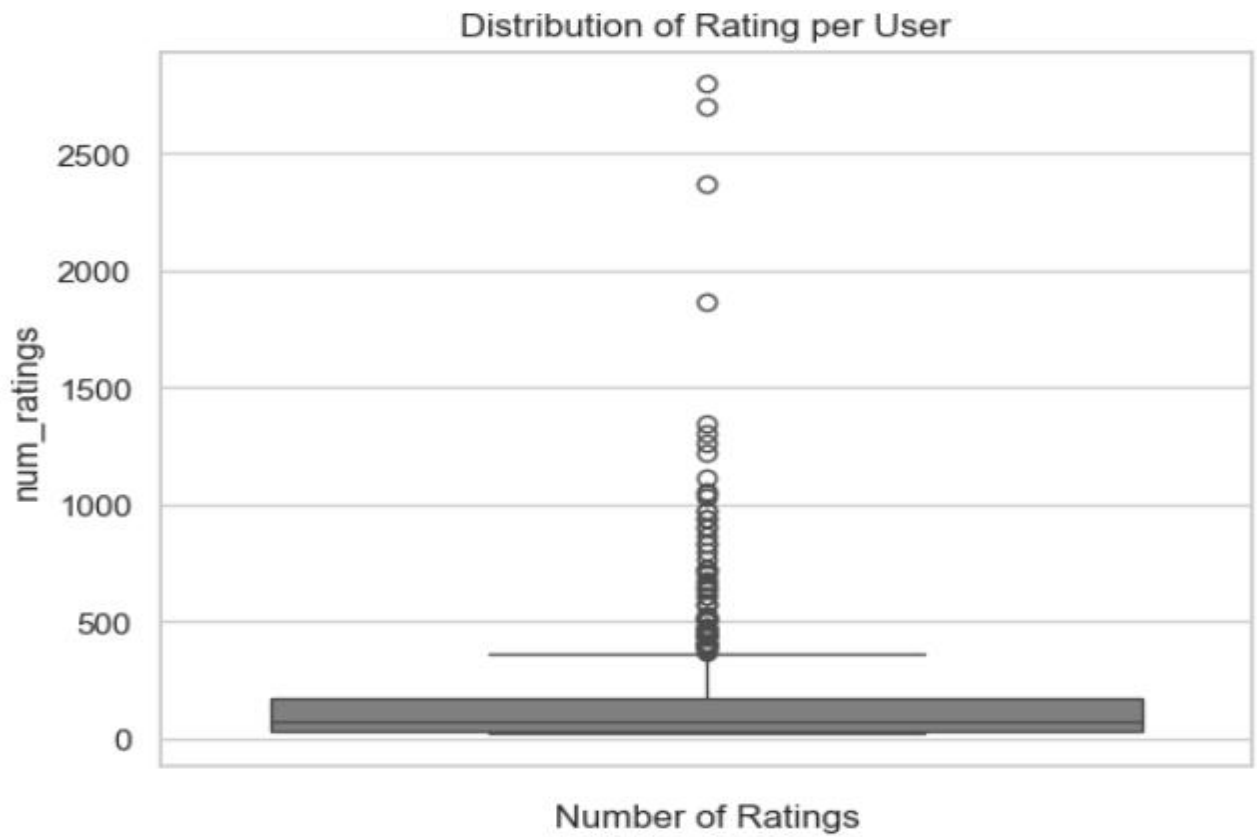
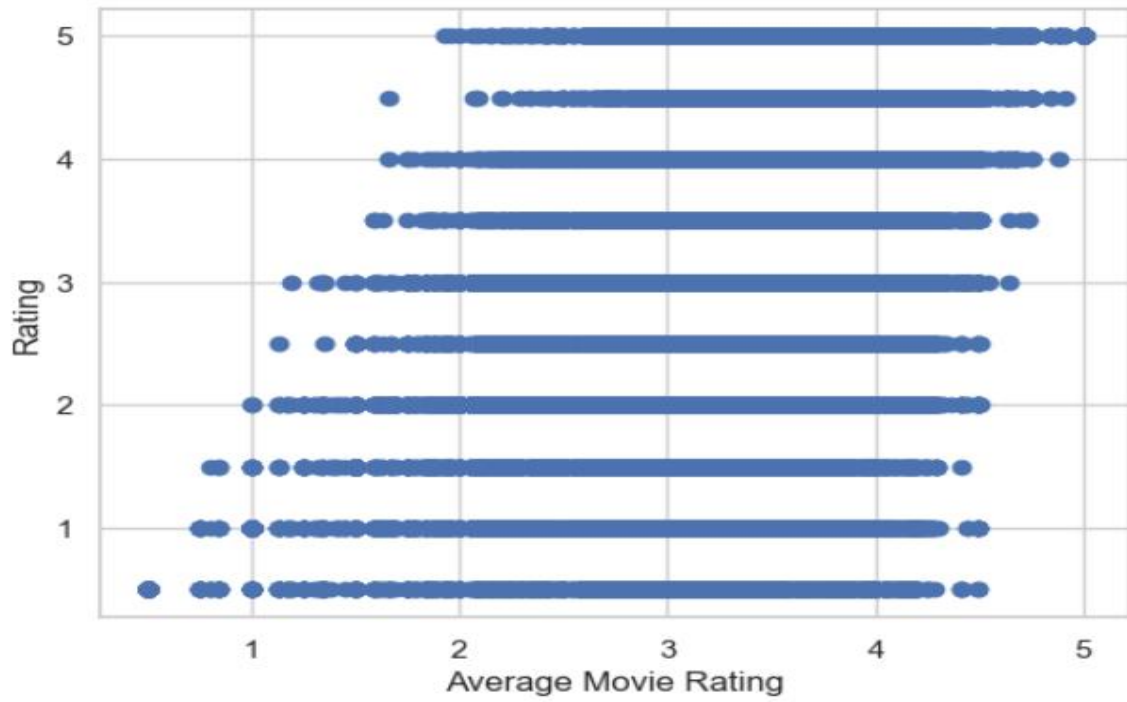
Six key visualizations were designed in Power BI to represent insights from the dataset:

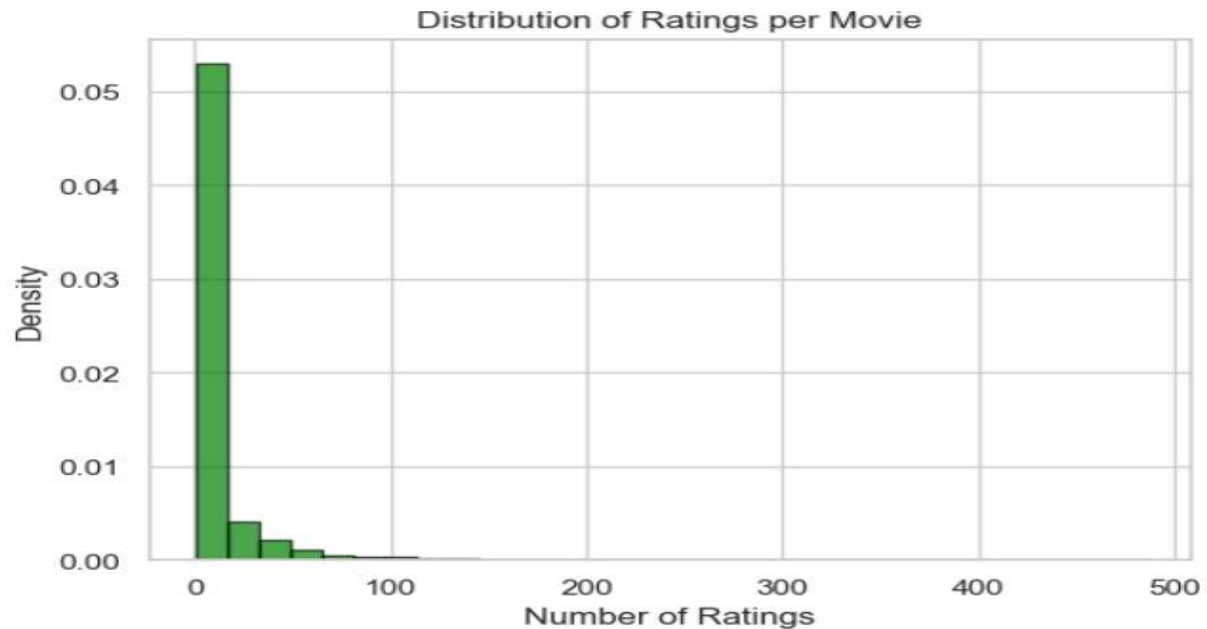
1. Rating Distribution (Histogram)
2. Number of Movies per Genre (Histogram)
3. Average Rating by Release Year (Line Chart)
4. Correlation Between Movie Count and Average Rating (Scatter Plot)
5. Distribution of Rating per User (Box Plot)
6. Distribution of Rating Across Movies (Histogram)

The following pages display the screenshots of the Python-generated charts.









6. Insights and Observations

Based on the analysis, several insights were derived:

- The majority of users rated movies within the 3.0–4.0 range.
- Drama and Comedy are the most popular genres with the highest rating counts.
- A few movies consistently received higher ratings, indicating strong viewer preference.
- Users tend to focus on specific genres rather than spreading ratings across all categories.

7. Conclusion

The MovieLens dataset provides valuable insights into user preferences and movie ratings. By integrating Python Script and visualizations, we successfully transformed raw data into meaningful analytics. The dashboard highlights trends, performance metrics, and user engagement patterns that can guide recommendations and content strategies.