**<Data Science Toolbox: Python Programming>**

**PROJECT REPORT**

(Project Semester January-April 2025)

***(Movie Ratings)***

Submitted by

(Prema Sai Kimmi)

Registration No 12306000

Programme and Section – B. Tech CSE, K23GD

Course Code INT-375

Under the Guidance of

**Name of faculty:** Baljinder Kaur

**UID:** 28968

**Discipline of CSE/IT**

**Lovely School of Computer Science**

**Lovely Professional University, Phagwara**

**CERTIFICATE**

This is to certify that Prema Sai Kimmi bearing Registration no. 12306000 has completed INT 375 project titled, **“Movie Ratings”** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

**Signature and Name of the Supervisor**

**Designation of the Supervisor**

**School of …………………………………………….**

Lovely Professional University

Phagwara, Punjab.

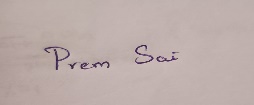
Date:

**DECLARATION**

I, Prema Sai Kimmi, student of Computer Science under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: Signature

Registration No. 12306000 Prema Sai Kimmi



1. **Problem Statement:**
2. Analyse rating trends over time to see how movie scores have changed by year or decade.
3. Use bar graphs to visualize top genres, studios, directors, and rating certifications.
4. Create pie charts to show the percentage distribution of genres, tomato meter status, and movie ratings.
5. Identify top 10 movies by critic and audience scores.
6. Find correlations between critic and audience scores, and between runtime and ratings.
7. Create histograms to show the distribution of critic ratings, audience ratings, and runtimes.
8. Calculate average critic, audience ratings, and runtime across all movies.
9. **Data set link**: <https://mavenanalytics.io/data-playground?page=12&pageSize=5>
10. **Implementation: -**

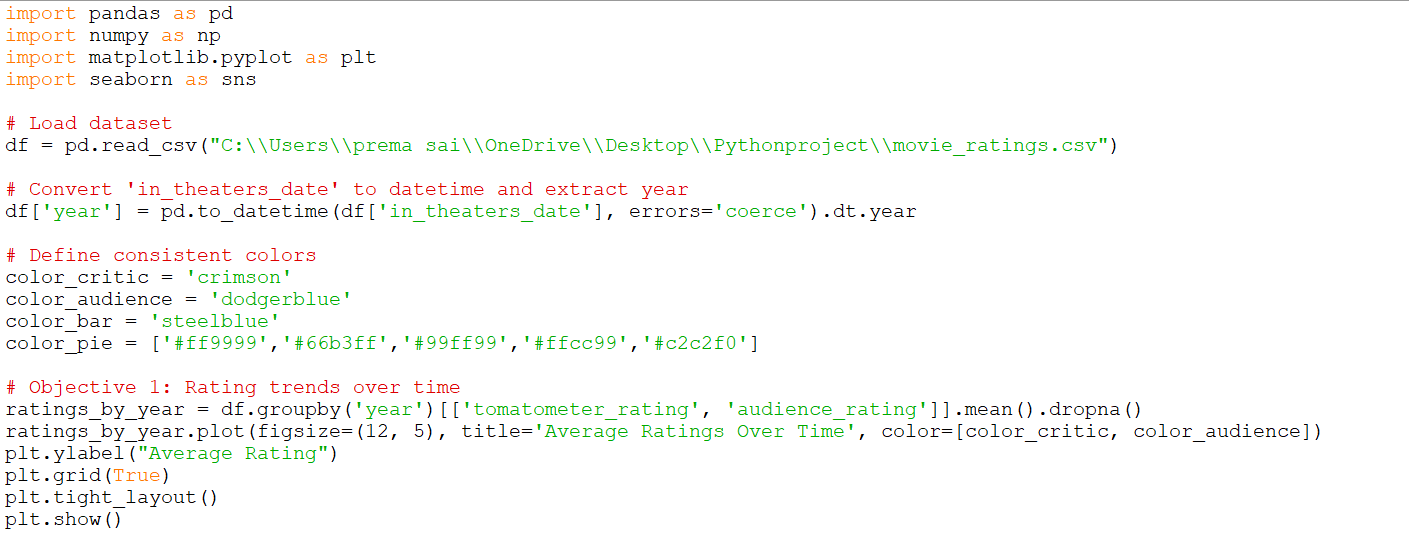
**Objective 1:** Analyse rating trends over time to see how movie scores have changed by year or decade.

**Purpose:** To explore how average critic and audience scores have changed across release years and decades.

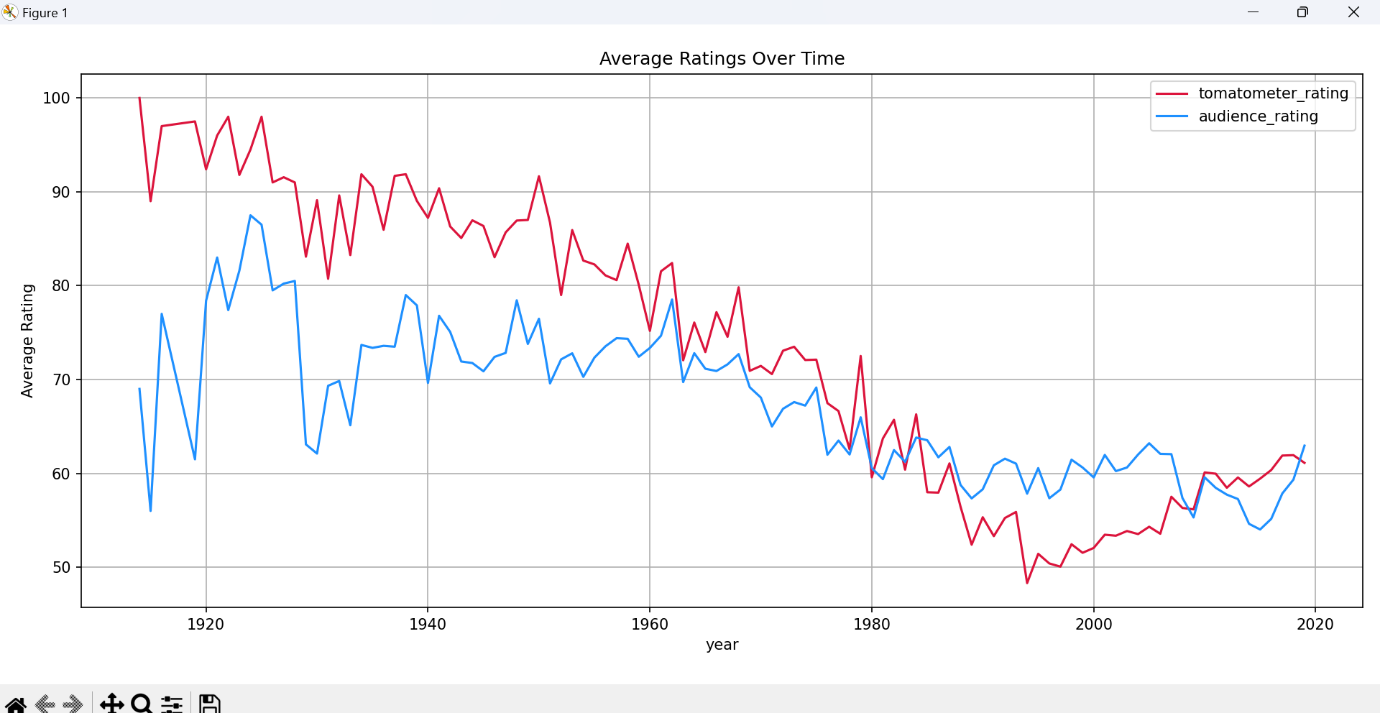
**Approach**: Group movies by year and decade, calculate average ratings, and visualize them using bar graphs.

**Outcome:** Clear identification of eras with peak audience or critic approval, aiding strategic content planning.

**Code:**

****

**Output:**

****

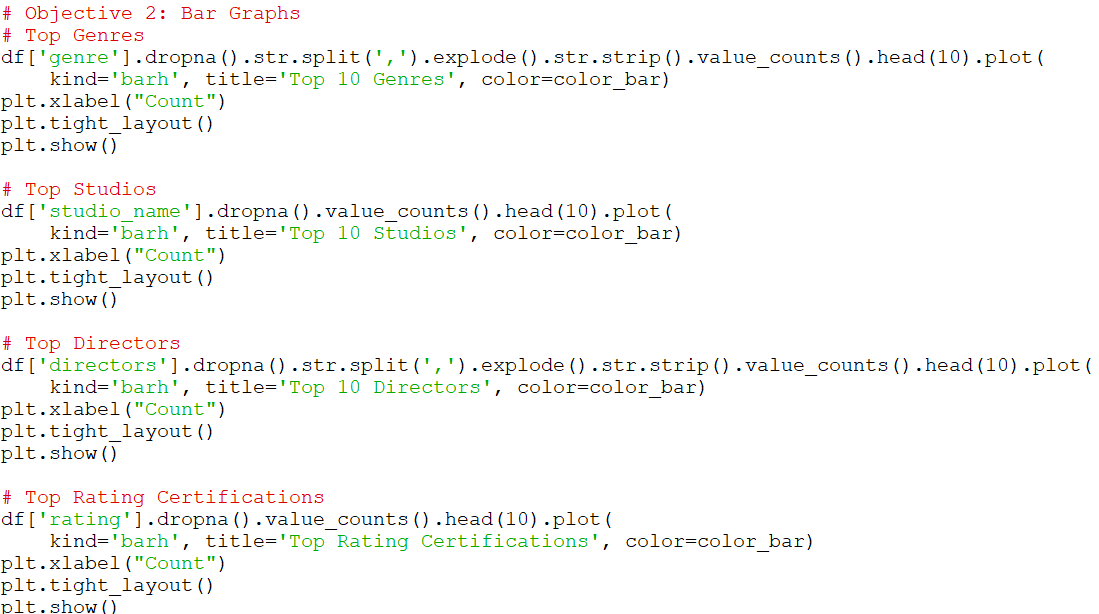
**Objective 2:** Use bar graphs to visualize top genres, studios, directors, and rating certifications.

**Purpose:** To identify which movie genres, production studios, and directors are most common or dominant.

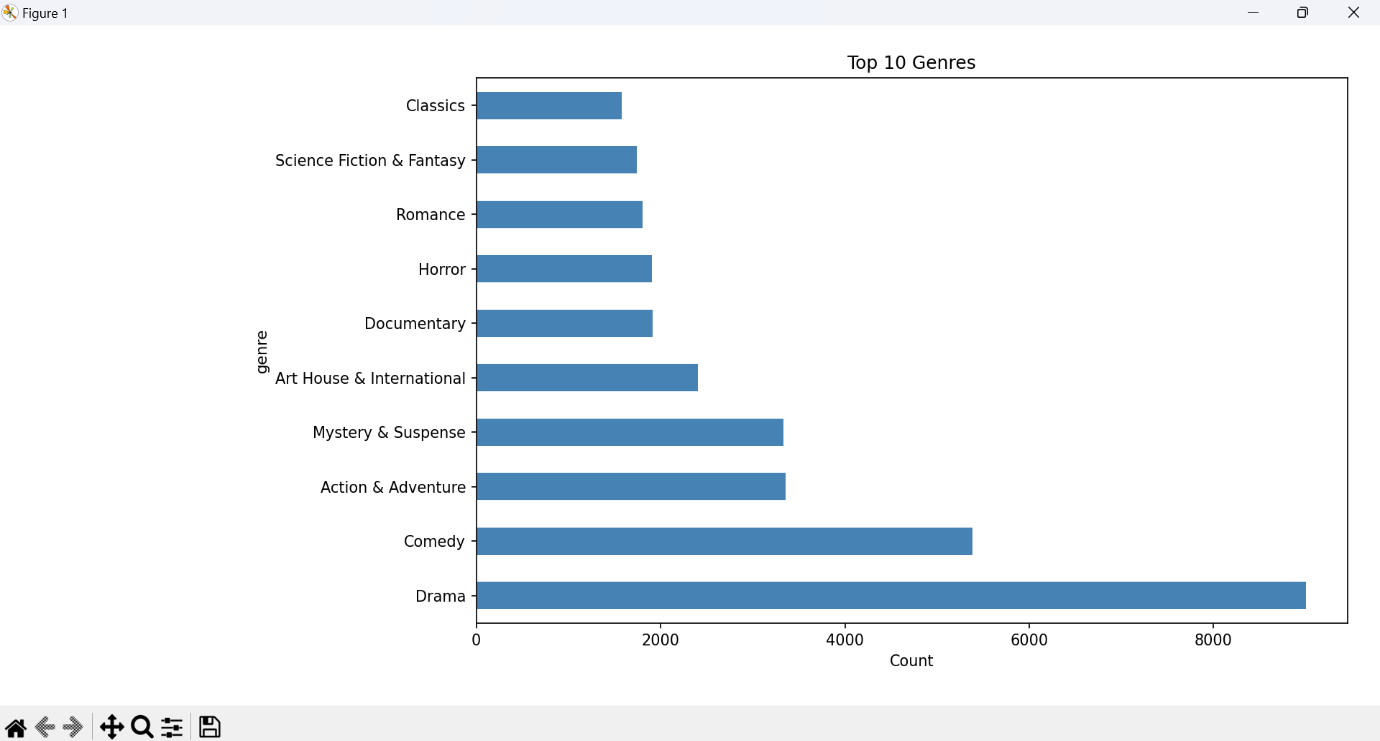
**Approach:** Count occurrences and ratings, visualize top 10 in each category using bar charts.

**Outcome:** Understanding dominant creators and types of content in the dataset.

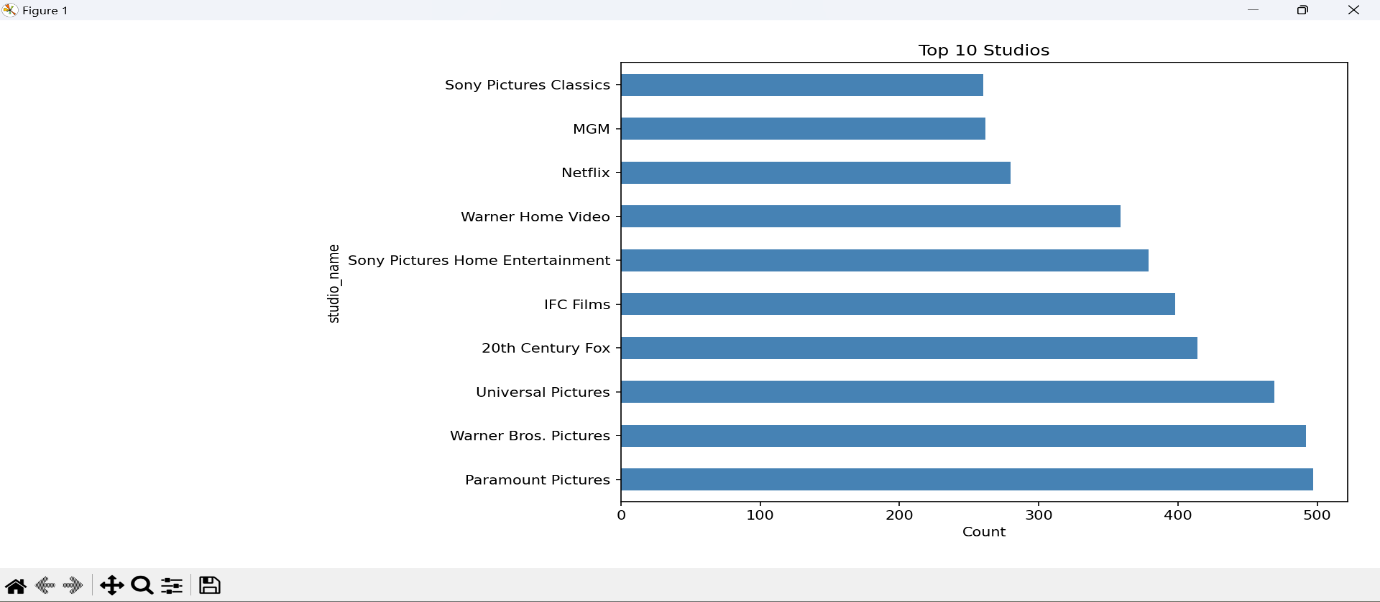
**Code:**

**Output:**

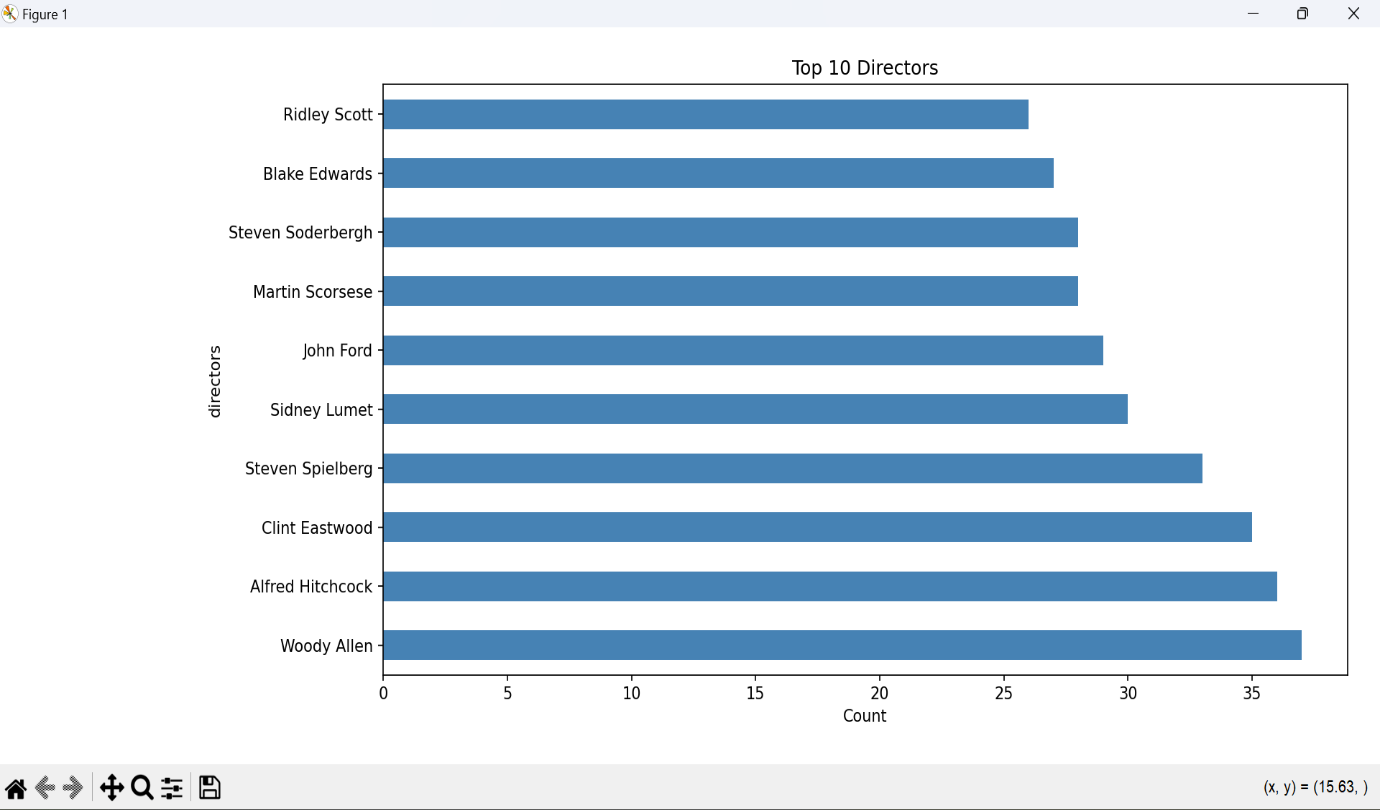
* Top 10 Genres



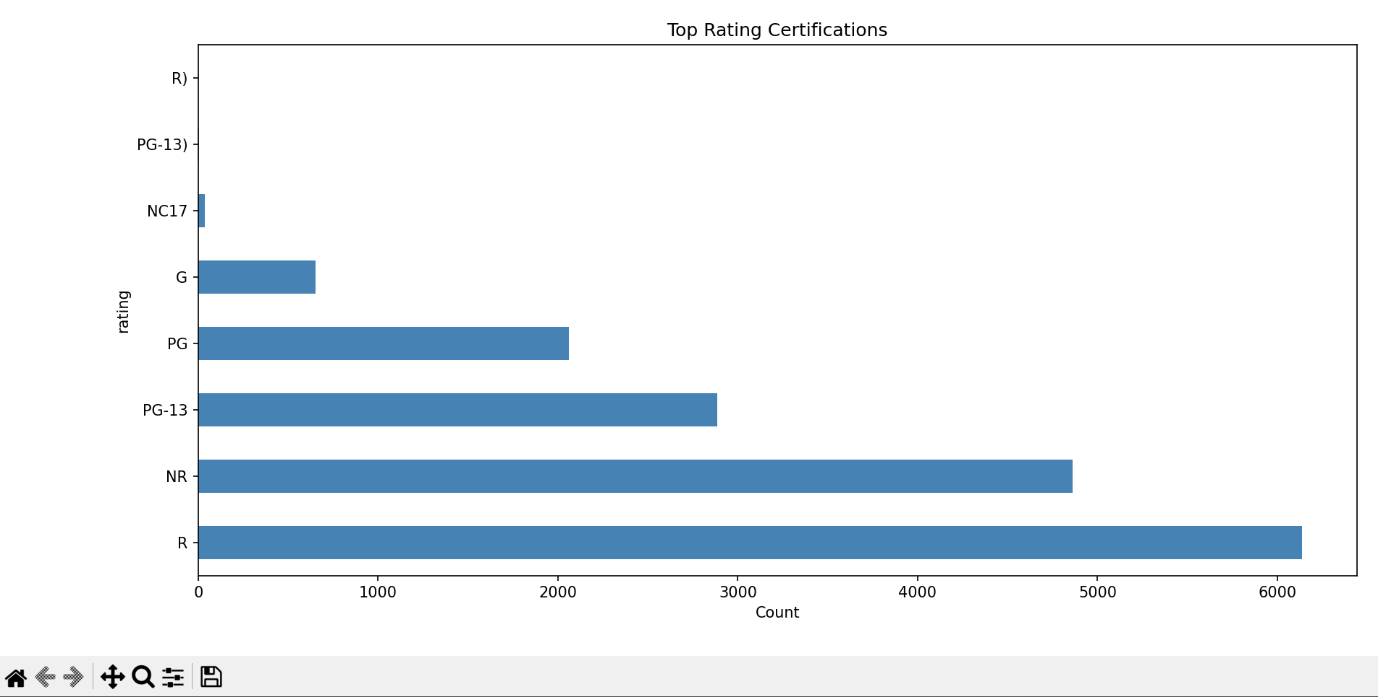
* **Top 10 Studios**



* **Top 10 Directors**

****

* **Top Rating Certifications**

****

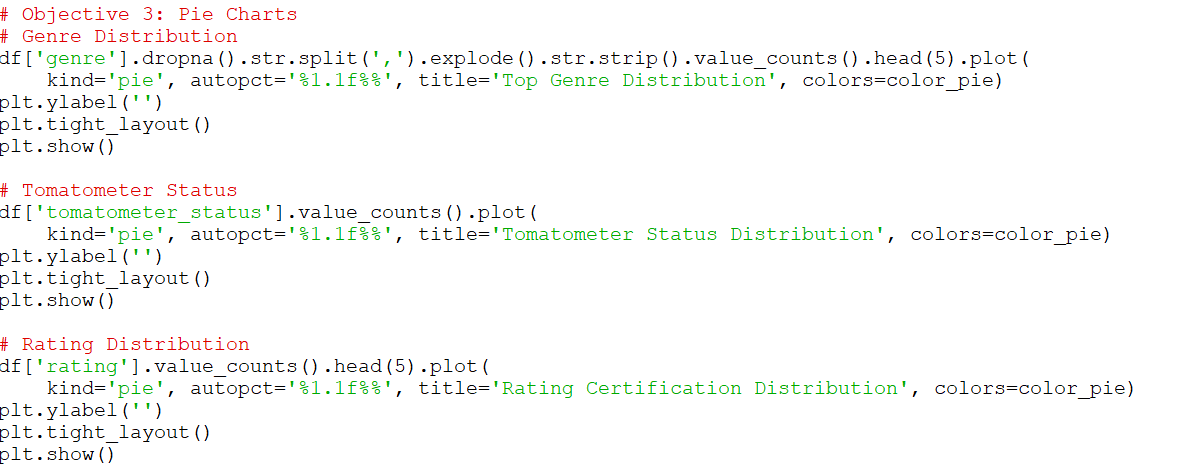
**Objective 3:** Create pie charts to show the percentage distribution of genres, tomato meter status, and movie ratings

**Purpose:** To understand the proportional representation of various movie characteristics.

**Approach:** Use pie charts to show the percentage breakdowns.

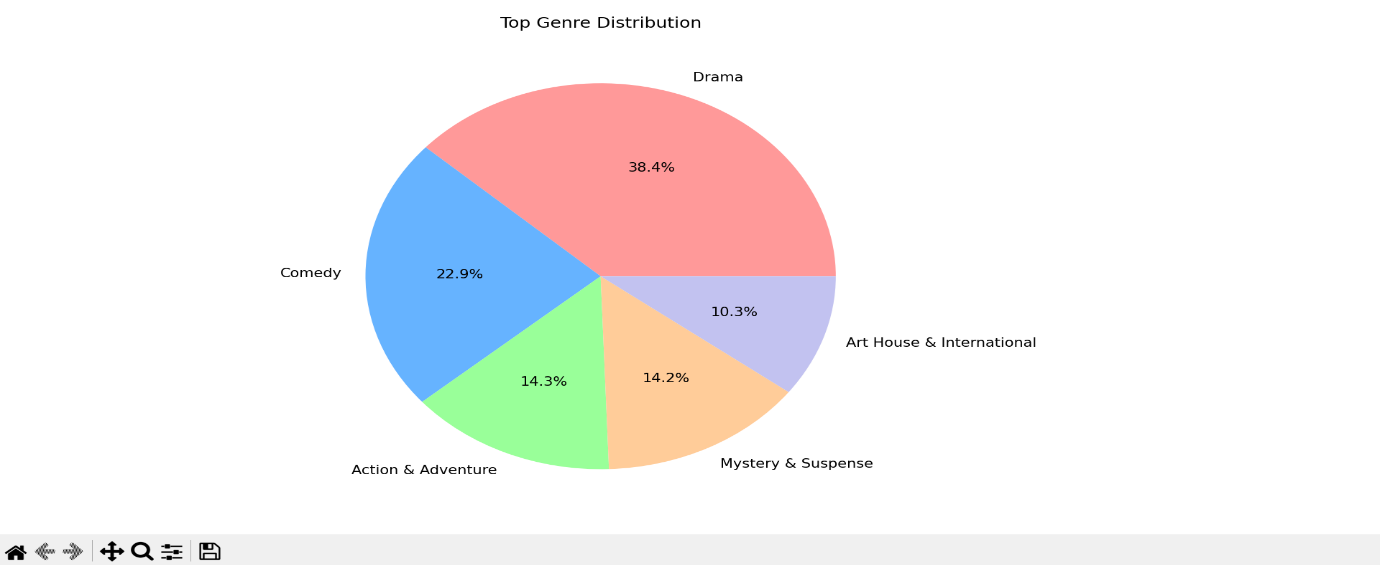
**Outcome:** Quick visual summary of genre spread, tomato status (Fresh/Rotten), and rating certifications

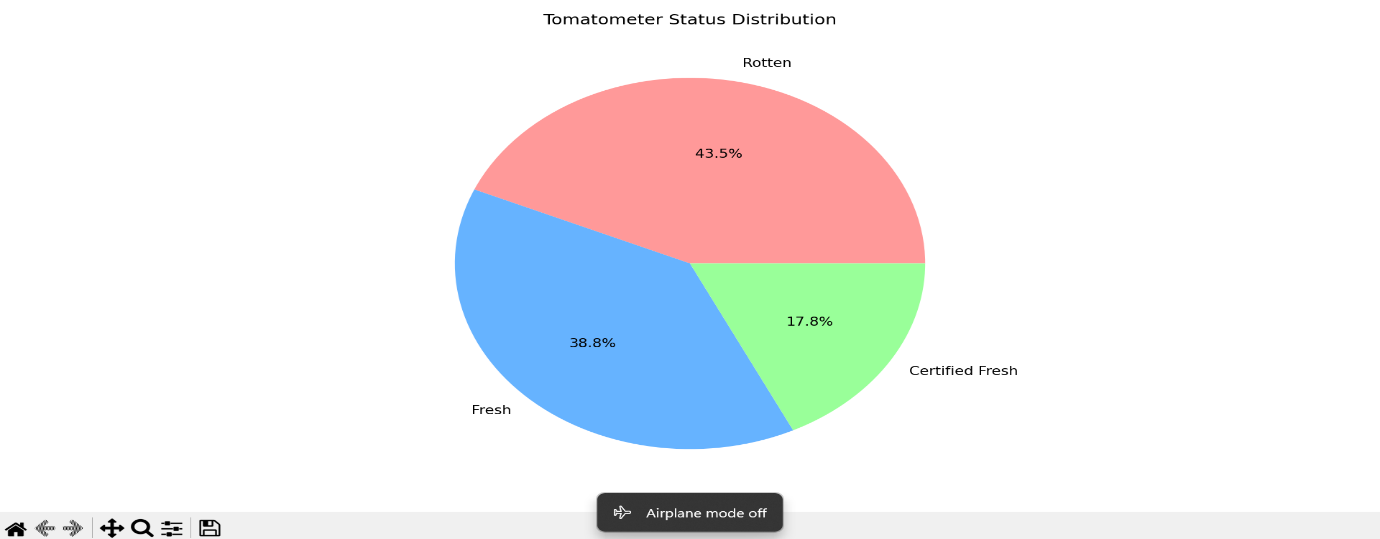
**Code:**

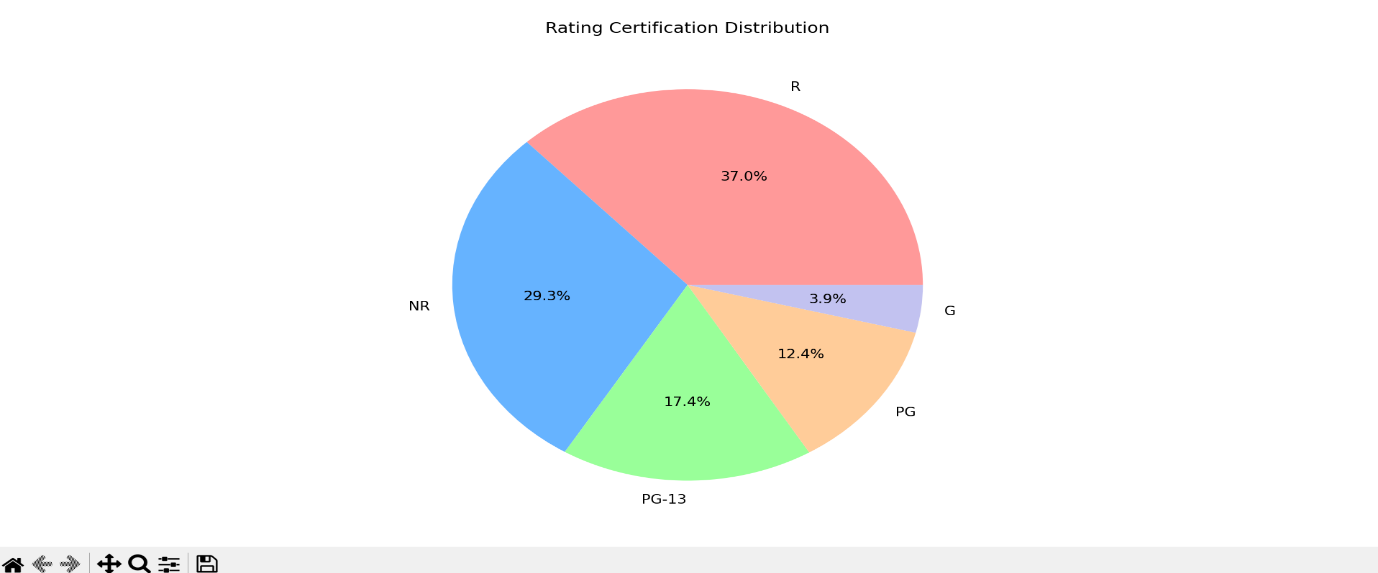


**Output:**

* **Genre Distribution**

****

* **Tomato meter Status Distribution**
* ****
* **Rating Certification Distribution**

****

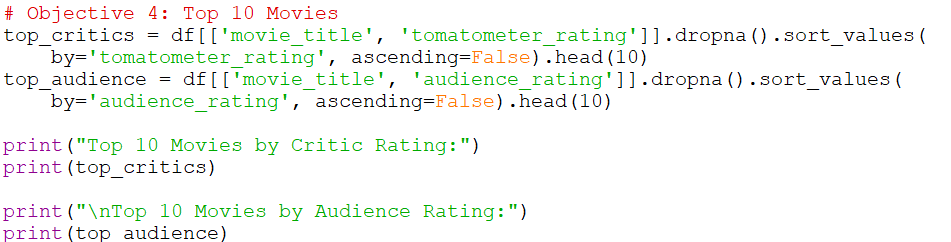
**Objective 4:** Identify top 10 movies by critic and audience scores

**Purpose:** To highlight the highest-rated movies from critics’ and viewers’ perspectives.

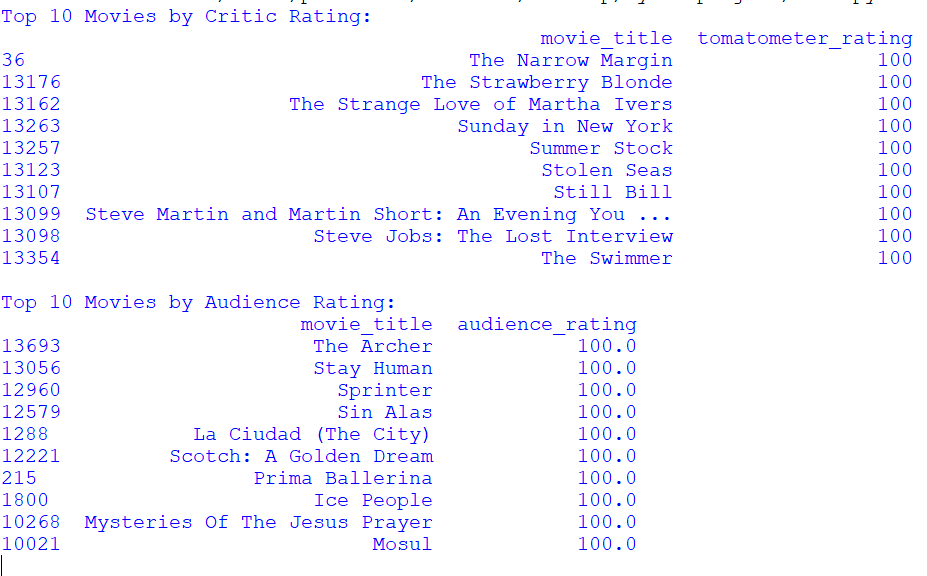
**Approach:** Sort dataset by tomatometer\_rating and audience\_rating and extract the top 10.

**Outcome:** Insight into critically acclaimed and most beloved movies

**Code:**

****

**Output:**



**Objective 5:** Find correlations between critic and audience scores, and between runtime and ratings

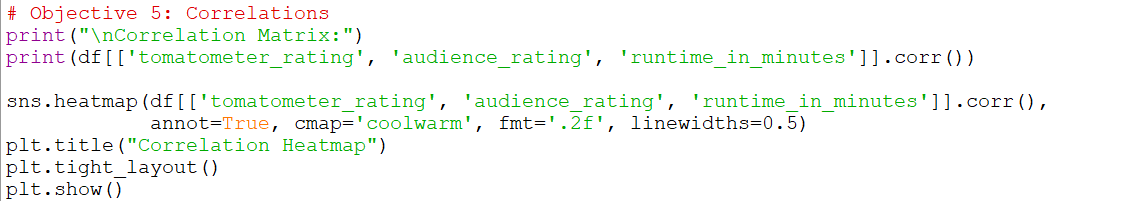
**Purpose:** To detect relationships between:

* Critic and Audience Ratings
* Runtime and Ratings

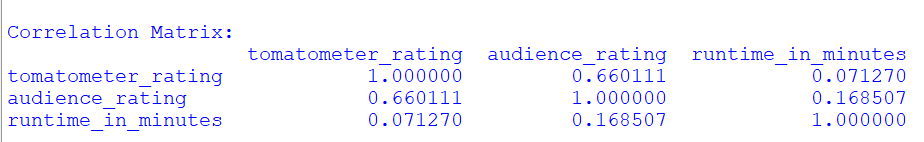
**Approach:** Use correlation coefficients and scatter plots.

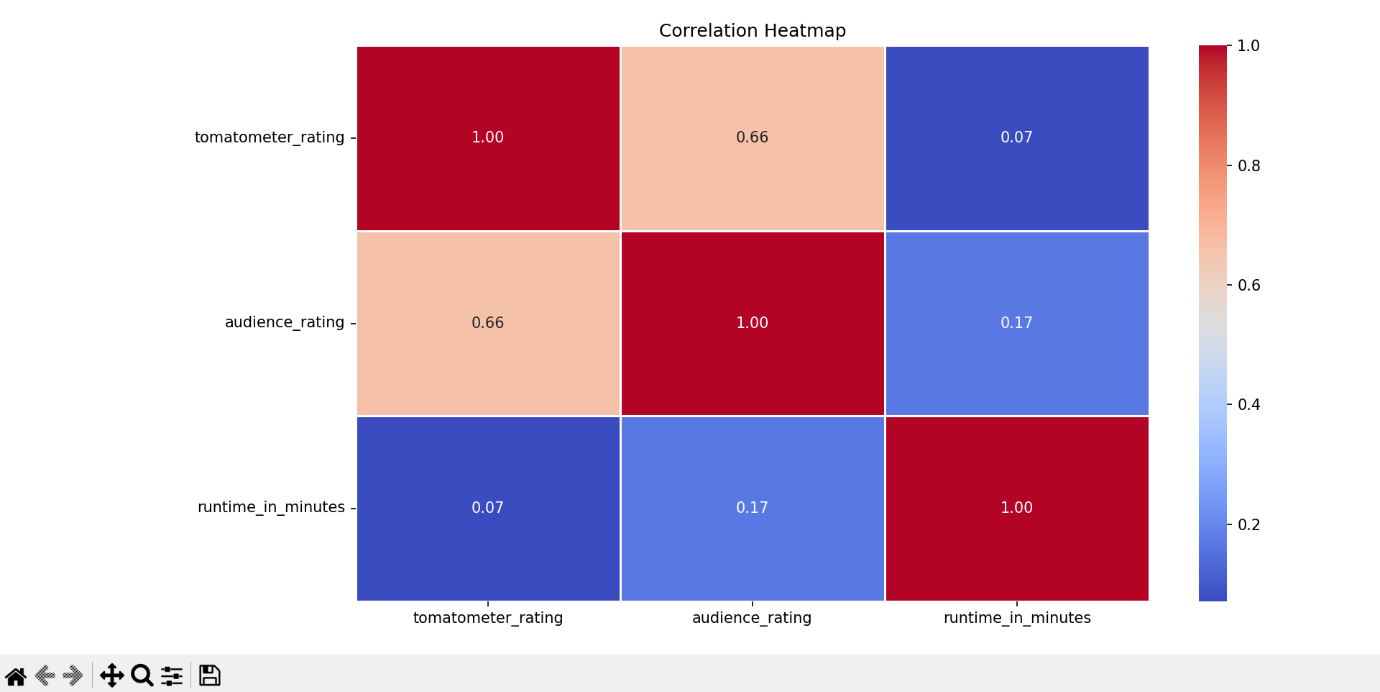
**Outcome:** Evaluate if longer movies or higher critic scores influence viewer satisfaction.

**Code:**

****

**Output:**

****

****

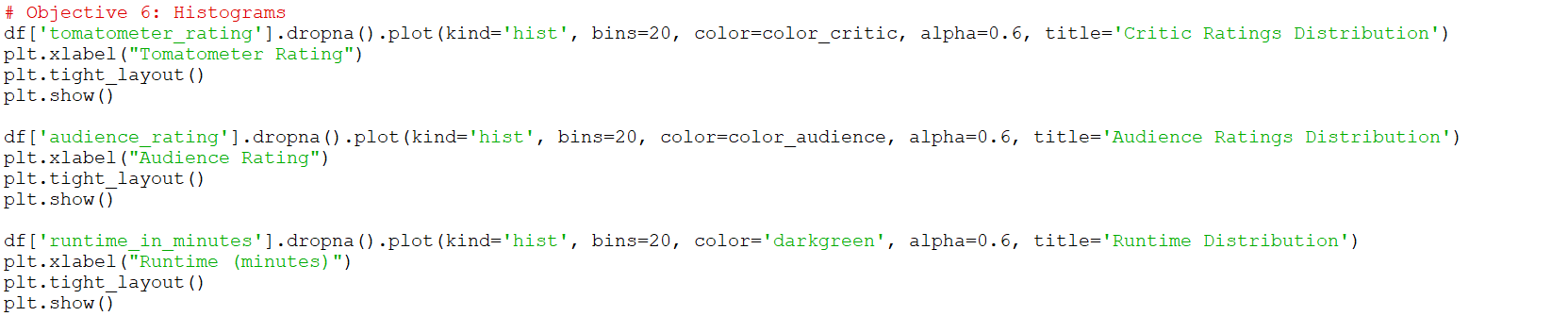
**Objective 6:** Create histograms to show the distribution of critic ratings, audience ratings, and runtimes

**Purpose**: To show how values like critic score, audience rating, and runtime are distributed.

**Approach**: Use histograms to display the frequency of each score and runtime range.

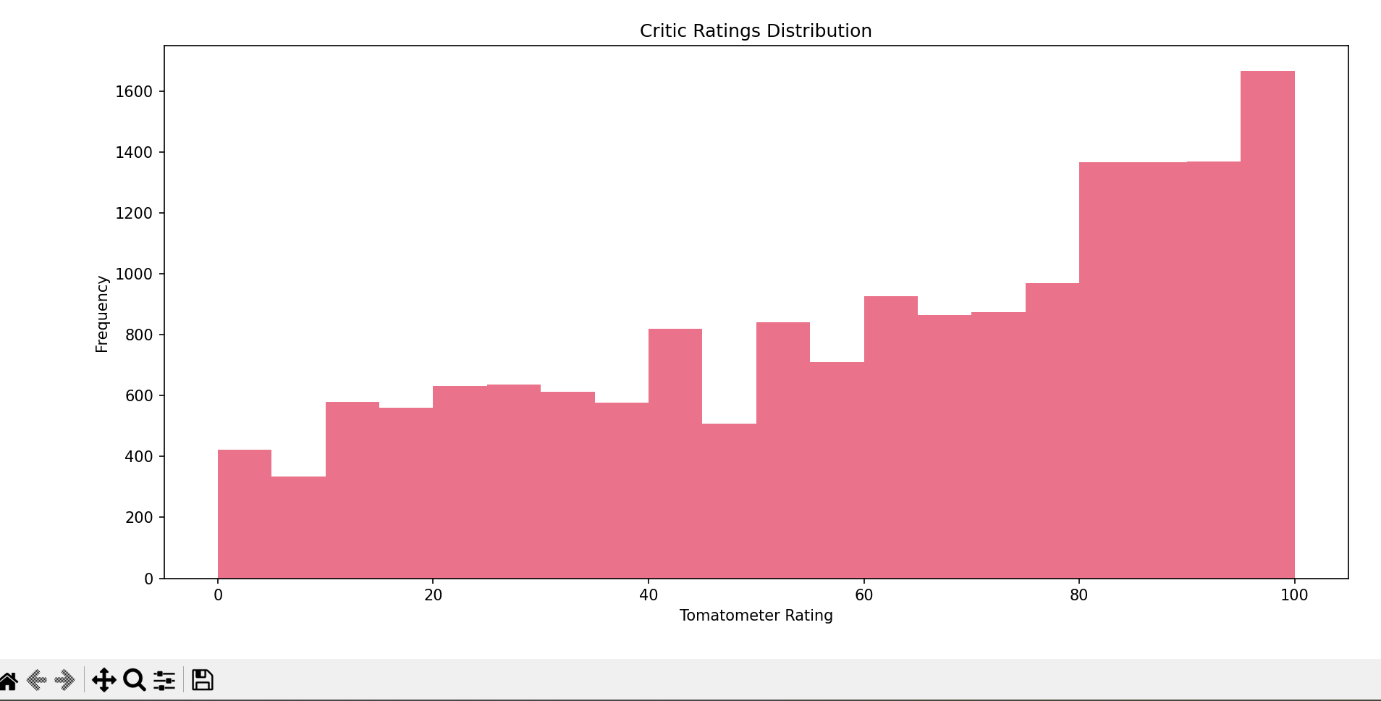
**Outcome**: Identify central tendencies and outliers in scores and movie lengths.

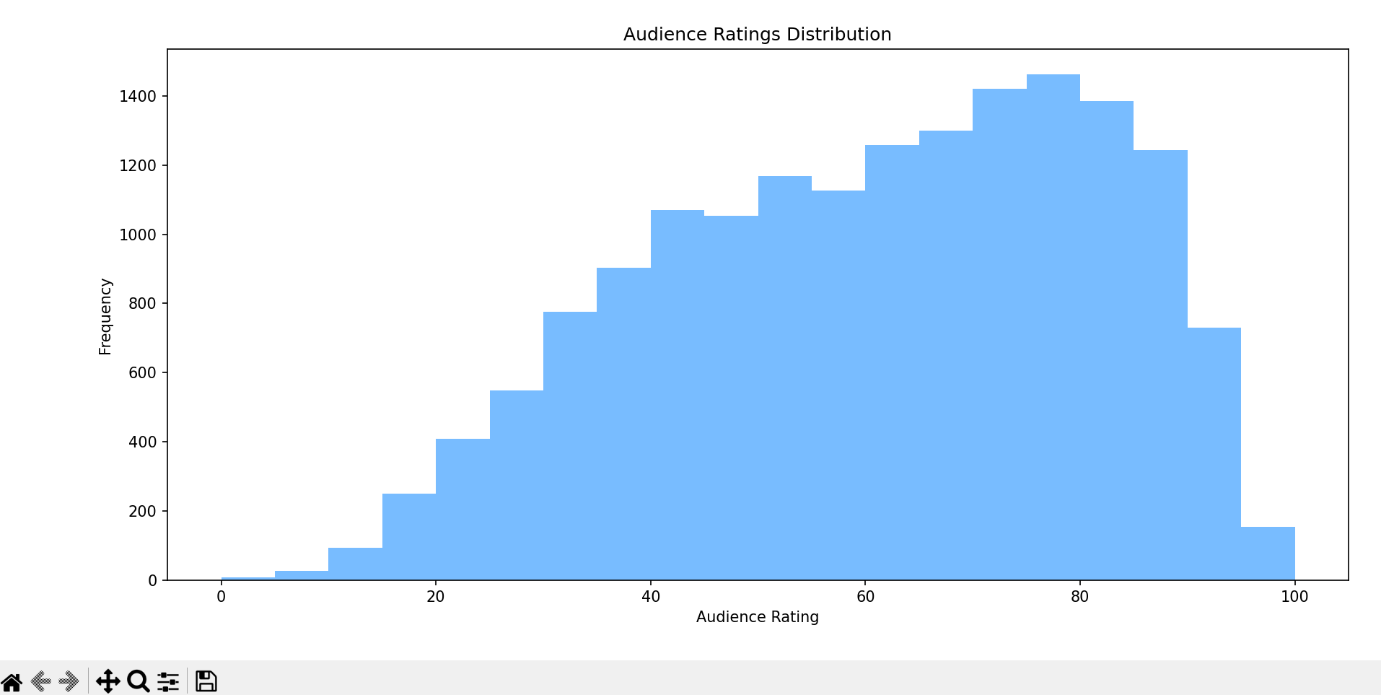
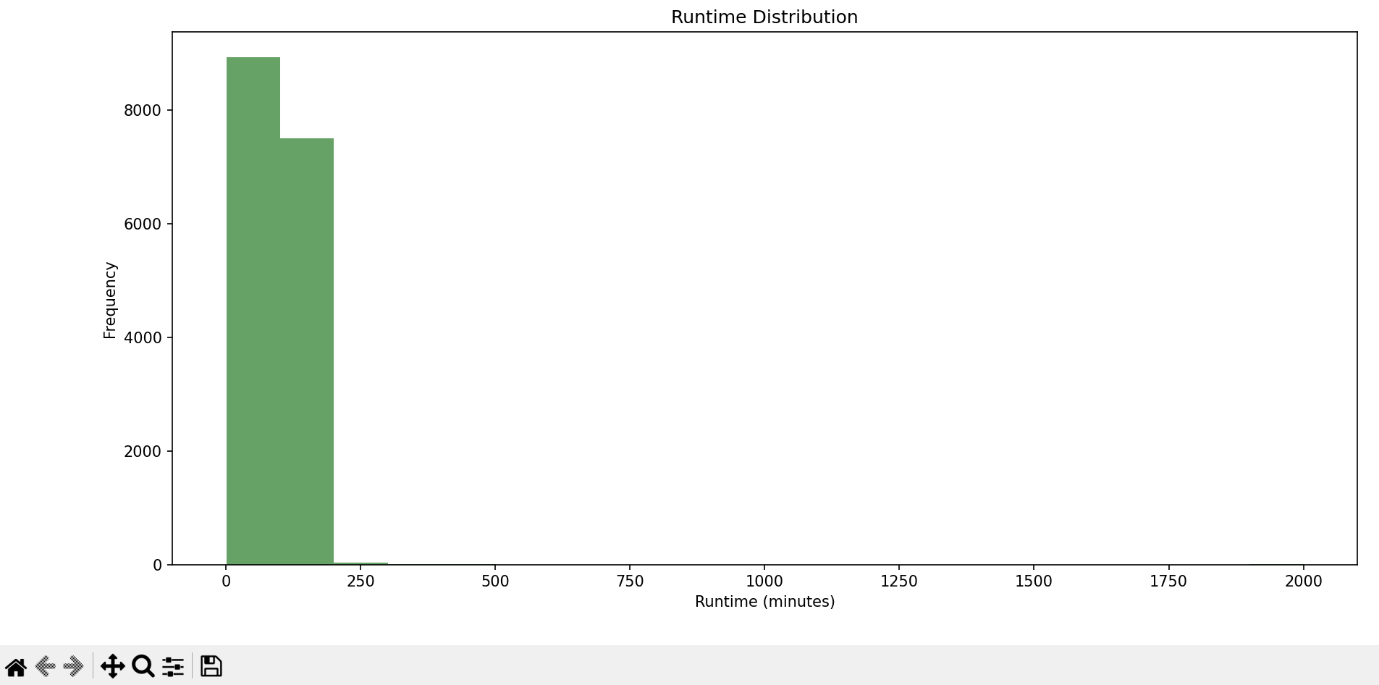
**Code:**

****

**Output:**

* **Critic Ratings Distribution**

****

* **Audience Rating Distributions**
* **Runtime Distribution**

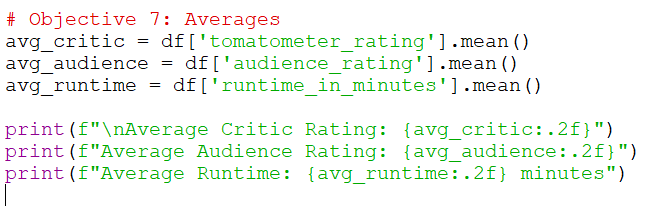
**Objective 7:** Calculate average critic, audience ratings, and runtime across all movies

**Purpose:** To determine the overall average critic score, audience rating, and runtime.

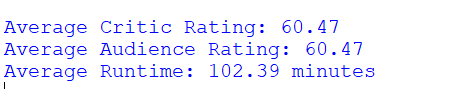
**Approach:** Use Pandas .mean() function on relevant columns.

**Outcome:** Quantitative summary of the dataset for comparative or trend analysis.

**Code:**

****

**Output:**

****

LinkedIn: <https://www.linkedin.com/posts/prema-sai_python-datascience-datascience-activity-7317198672989958144-PMIc?utm_source=share&utm_medium=member_desktop&rcm=ACoAAEcgeHQBxGnHuXFDPI9H0EikSzmR_Iu_qQc>