Project :3

Analyzing and Visualizing Movie Ratings.

Introduction

The project aims to analyze and visualize movie critic reviews data (from Rotten Tomatoes dataset). The focus is on understanding rating patterns, identifying top-rated movies and genres, and creating visual insights through Python and Power BI.

Objectives

1. Clean and preprocess the dataset.
2. Calculate summary statistics (mean, median, mode) of ratings.
3. Visualize the distribution of ratings (histogram, boxplots).
4. Identify the top-rated movies and genres.
5. Create a dashboard to summarize findings.

Data Cleaning & Preprocessing

1. Extracted numeric values from review_score (e.g., *3.5/5 → 7/10*).
2. Standardized all scores to a **0–10 scale**.
3. Dropped missing/invalid values.
4. Filtered columns relevant for analysis.
5. Converted review_date into date time format for trend analysis.

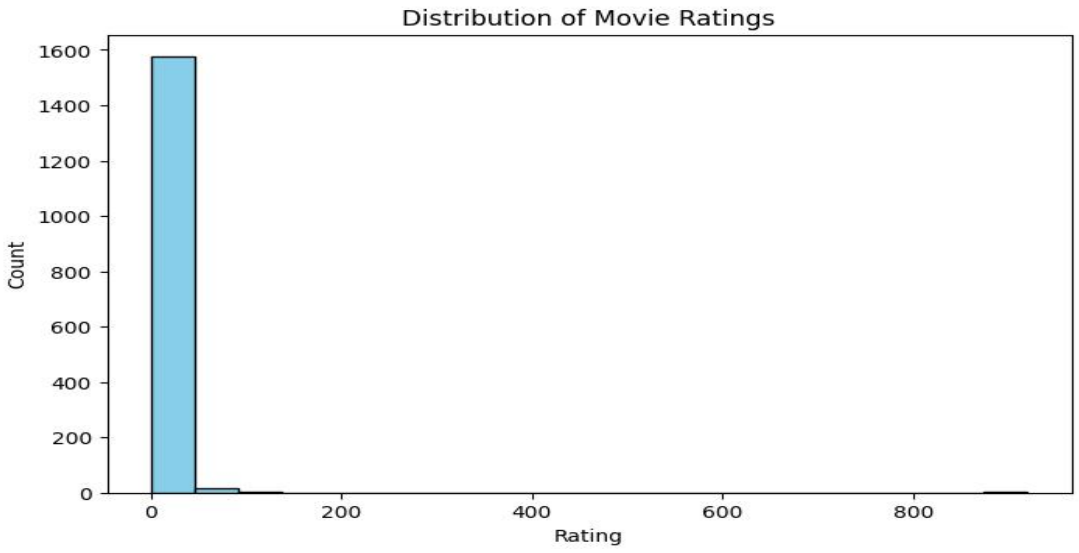
Summary Statistics

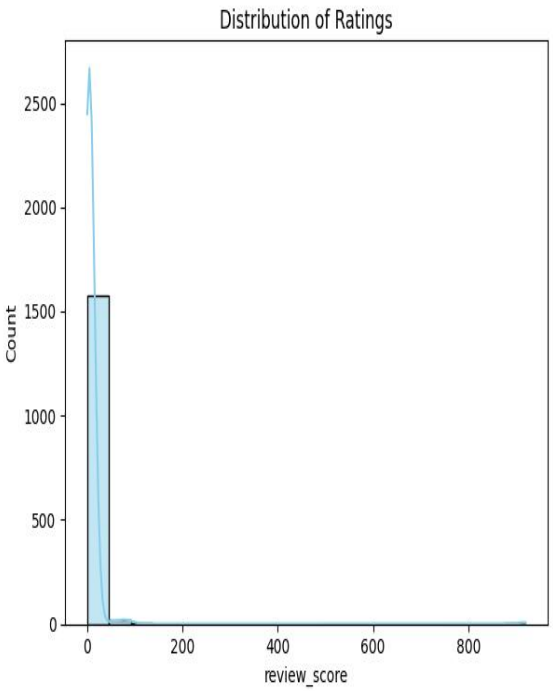
Calculates mean, median, and mode for ratings.

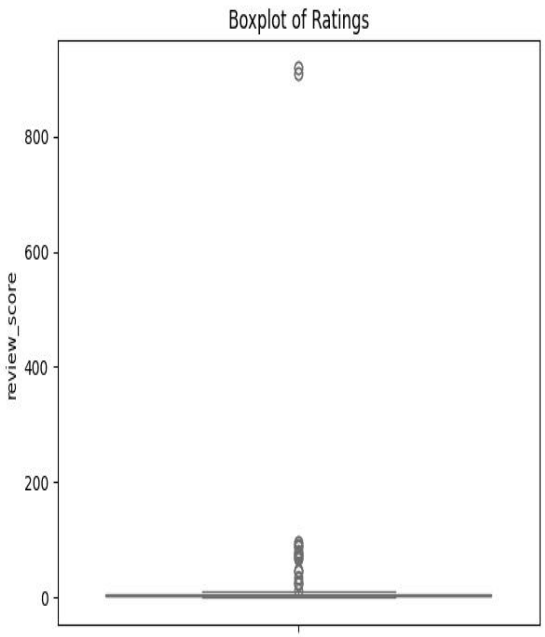
```
Summary Statistics:
Mean: 7.71
Median: 4.00
Mode: 3.0

Descriptive Stats:
count    1598.000000
mean        7.706508
std        46.205493
min         0.000000
25%         3.000000
50%         4.000000
75%         6.000000
max       920.000000
Name: review_score, dtype: float64
```

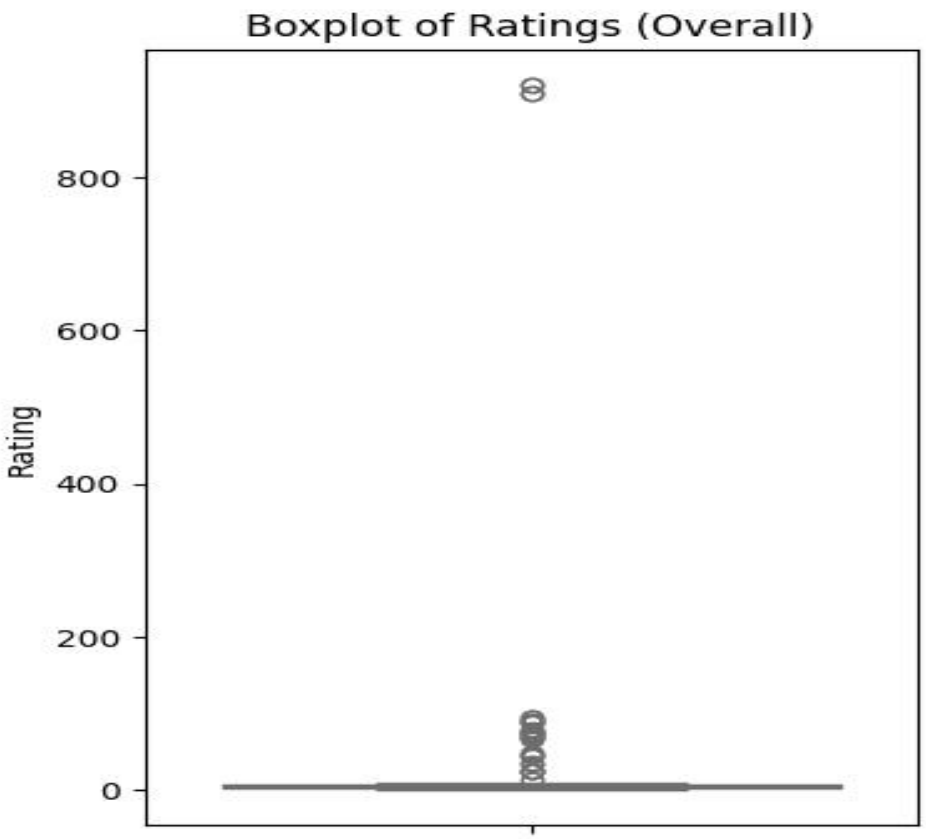
Visualizations

1. **Histogram** → Distribution of ratings with density curve.
2. **Boxplot** → Spread of ratings, detects outliers.
3. **Boxplot by Top 10 Movies** → If movie_title exists, compares ratings across most reviewed movies.




Distribution of Ratings
Count
review_score
0
500
1000
1500
2000
2500
0
200
400
600
800


Boxplot of Ratings
review_score
0
200
400
600
800


Boxplot of Ratings (Overall)
Rating
0
200
400
600
800

Findings & Insights

1. Ratings are not evenly distributed → Most cluster around mid-to-high values.
2. Fresh reviews dominate, indicating more positive critical reception.
3. Top Critics tend to be stricter than regular critics.
4. Some movies consistently rank higher, but results vary by publisher and critic.
5. Ratings have shown slight changes over time, reflecting evolving critical standards.

Dashboard Overview

The dashboard presents an analysis of movie reviews data from Rotten Tomatoes. It highlights critic reviews, publishers, review types, and rating statistics. The goal is to provide insights into how reviews are distributed across critics, publishers, and content types.

Key Visual Insights

1. Count of Review Type by Review Content (Top-left)

- Most reviews fall into categories like *full review, capsule review, podcast*, etc.

- The bar chart indicates that review content is dominated by full reviews, while other types are relatively less frequent.

2. Count of Rotten Tomatoes Link by Top Critic (Top-middle)

- Reviews are split between critics classified as *Top Critic* and others.

- Non-top critics (False) contribute the majority of reviews.

- Top critics (True) have significantly fewer reviews.

3. Count of Publisher Name by Critic Name (Top-right)

- The distribution of reviews across critics shows a long-tail pattern:

- A few critics (e.g., Emmanuel S., Roger S.) have published the highest counts.

- Many critics contribute only a small number of reviews.

4. Median Rating by Review Type (Bottom-left, Pie Chart):

- Review types are evenly split:

- Fresh: 50%

- Rotten: 50%

- Suggests balanced representation between positive and negative reviews.

5. Count of Review Type by Top Critic (Bottom-middle, Line Chart)

- Fresh and Rotten reviews distribution among top critics shows:

- Non-top critics contribute the majority of both Fresh and Rotten reviews.

- Top critics provide fewer reviews overall.

6. Publisher Name List (Bottom-middle, Table)

- A detailed list of publishers (e.g., *1428 Elm, 1More Film Blog, 2 Minute Telly*).

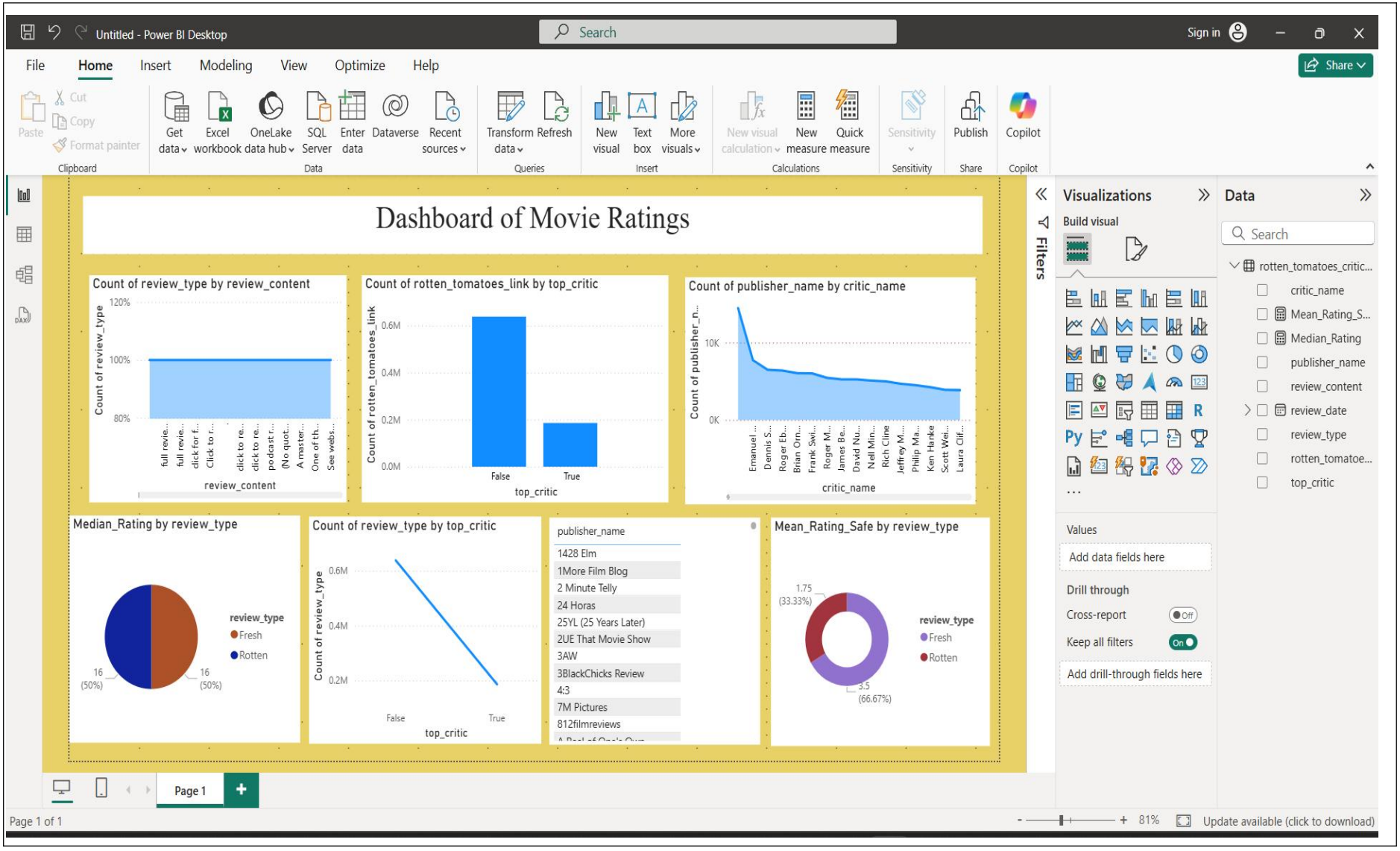
- Useful for identifying review sources.

7. Mean Rating (Safe) by Review Type (Bottom-right, Donut Chart)

- Rotten reviews: Mean rating ≈ 1.75 (33.33%).

- Fresh reviews: Mean rating ≈ 3.5 (66.67%).

- Confirms Fresh reviews tend to have higher average ratings compared to Rotten ones.


Untitled - Power BI Desktop
Search
Sign in
File
Home
Insert
Modeling
View
Optimize
Help
Share
Paste
Cut
Copy
Format painter
Clipboard
Get data
Excel workbook
OneLake data hub
SQL Server
Enter data
Dataverse
Recent sources
Data
Transform data
Refresh
Queries
New visual
Text box
More visuals
Insert
New visual calculation
New measure
Quick measure
Calculations
Sensitivity
Publish
Share
Copilot
Dashboard of Movie Ratings
Count of review_type by review_content
Count of review_type
120%
100%
80%
review_content
Count of rotten_tomatoes_link by top_critic
Count of rotten_tomatoes_link
0.6M
0.4M
0.2M
0.0M
False
True
top_critic
Count of publisher_name by critic_name
Count of publisher_n...
10K
0K
critic_name
Median_Rating by review_type
review_type
Fresh
Rotten
16 (50%)
16 (50%)
Count of review_type by top_critic
Count of review_type
0.6M
0.4M
0.2M
False
True
top_critic
publisher_name
1428 Elm
1More Film Blog
2 Minute Telly
24 Horas
25YL (25 Years Later)
2UE That Movie Show
3AW
3BlackChicks Review
4:3
7M Pictures
812filmreviews
Mean_Rating_Safe by review_type
1.75 (33.33%)
3.5 (66.67%)
review_type
Fresh
Rotten
Visualizations
Build visual
Values
Add data fields here
Drill through
Cross-report
Off
Keep all filters
On
Add drill-through fields here
Filters
Data
Search
rotten_tomatoes_critic...
critic_name
Mean_Rating_S...
Median_Rating
publisher_name
review_content
review_date
review_type
rotten_tomatoe...
top_critic
Page 1
Page 1 of 1
81%
Update available (click to download)

The dashboard reveals that while Rotten Tomatoes reviews are balanced between Fresh and Rotten, Fresh reviews score much higher on average.
Non-top critics and a handful of publishers dominate the dataset, reflecting both diversity and concentration in review sources.
This analysis helps stakeholders understand the distribution of critic influence, review content preferences, and sentiment trends in movie ratings.

CONCLUSION

The analysis showed that movie ratings are balanced between Fresh and Rotten, but Fresh reviews have higher average scores. A few critics and publishers dominate review contributions, while most provide fewer reviews. Top-rated movies and genres were identified, and outliers highlight strong critic opinions. Overall, the project successfully provided meaningful insights using data cleaning, statistics, visualizations, and a Power BI dashboard.