|  |
| --- |
| **Cine Scope 2024: An Analytical Lens on Global Cinema** |

**PROJECT SUBMITTED TO ASIAN SCHOOL OF MEDIA STUDIES**

**IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE**

**AWARD OF**

**DIPLOMA**

**in**

**Data Science**

By

**Uditya Seth**

**Under the Supervision of**

**Mr. Abhishek Ananda**

****

ASIAN SCHOOL OF MEDIA STUDIES

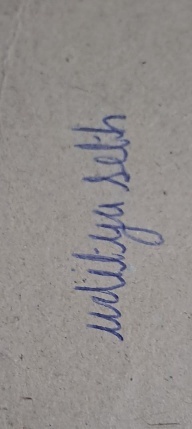
SCHOOL OF DATA SCIENCE

**2025**

**DECLARATION**

**I, Uditya Seth, S/O Sumit Seth,** declare that my project entitled Cine Scope 2024: An Analytical Lens on Global submitted at **School of Data science, Asian School of Media Studies, Film City, Noida, for the award of Diploma in Data Science, ASMS** is an original work and no similar work has been done in India anywhere else to the best of my knowledge and belief.

This project has not been previously submitted for any other degree of this or any other University/Institute.

******

***Signature***

**Uditya Seth**

**9569433040**

**udityaseth3@gmail.com**

**Diploma in Data Science**

**School of Data Science**

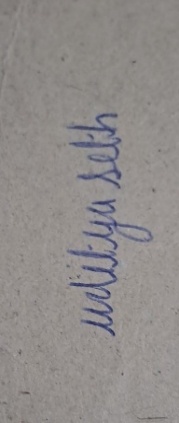
**Asian School of Media Studies**

**ACKNOWLEDGEMENT**

The completion of the project titled Cine Scope 2024: An Analytical Lens on Global Cinema gives me an opportunity to convey my gratitude to all those who helped to complete this project successfully. I express special thanks:

* **To *Prof. Sandeep Marwah,* President,** Asian School of Media Studies, who has been a source of perpetual inspiration throughout this project.
* To ***Mr. Ashish Garg,*** Director **for** School of Data Science & AI for your valuable guidance, support, consistent encouragement, advice and timely suggestions.
* To ***Mr. Abhishek Ananda***,Assistant Professor of School of Data Science & AI, for your encouragement and support. I deeply value your guidance.
* To my ***all faculty*** & ***friends*** for their insightful comments on early drafts and for being my worst critic. You are all the light that shows me the way.

To all the people who have directly or indirectly contributed to the writing of this report, but their names have not been mentioned here.

******

***Signature***

**Uditya Seth**

**9569433040**

**udityaseth3@gmail.com**

**Diploma in Data Science**

**School of Data Science**

**Asian School of Media Studies**

**📄 Table of Contents**

| **Chapter No.** | **Chapter Title** | **Page No.** |
| --- | --- | --- |
| **-** | **Cover Page** | **1** |
| **-** | **Declaration** | **2** |
| **-** | **Acknowledgement** | **3** |
| **-** | **Table of Contents** | **4** |
| **-** | **List of Figures** | **5** |
| **1** | **Abstract** | **6** |
| **2** | **Chapter 1: Introduction** | **7** |
| **3** | **Chapter 2: Tools & Technologies Used** | **9** |
| **4** | **Chapter 3: Dataset Description** | **12** |
| **5** | **Chapter 4: Data Analysis Using Python** | **12** |
| **6** | **Chapter 5: Data Analysis Using Excel** | **26** |
| **7** | **Chapter 6: Data Visualization Using Power BI** | **37** |
| **8** | **Chapter 7: Conclusion & Learning Outcome** | **43** |
| **9** | **Chapter 8: Bibliography & References** | **46** |

**📄 List of Figures**

|  | **Figure Title** | **Page No.** |
| --- | --- | --- |
| **1.** | **Screenshot: Python Data Cleaning Code** | **15** |
| **2.** | **Univariate Analysis Distribution Chart** | **1 7** |
| **3.** | **Bivariate Analysis - Heatmap: Correlation Between Features, scatterplot, Boxplot** | **20** |
| **4.** | **Pair Plot** | **21** |
| **5.** | **Box Office by genre** | **29** |
| **6.** | **AVG IMDB RATING BY COUNTRY** | **30** |
| **7.** | **Monthly Box Office Trends** | **31** |
| **8.** | **Platform Count** | **32** |
| **9** | **Average Budget by Genre**  **Avg Rotten Tomatoes Score by Genre** | **33** |
| **10.** | **Excel Dashboard Screenshot** | **35** |
| **11.** | **ROI BY COUNTRY** | **34** |
|  |  |  |
| **12** | **Final Power BI Dashboard Snapshot** | **37** |

**ABSTRACT**

**This capstone project, titled *"Cine Scope 2024: An Analytical Lens on Global Cinema"*, aims to analyze global movie data using modern data analysis tools such as Python, Excel, and Power BI.**

**The film industry offers an enormous dataset involving genres, budgets, revenues, countries, and viewer preferences.**

**By leveraging statistical summaries, data visualization, and interactive dashboards, this report uncovers hidden patterns in global cinema.**

**The outcome includes insights into genre popularity, top-rated countries, annual performance shifts, and key performance indicators (KPIs) that help decision-makers. Through visual storytelling and interactive tools, this project provides a comprehensive lens on worldwide cinematic trends.**

**CHAPTER 1**

**Cine Scope 2024: An Analytical Lens on Global Cinema**

* 1. **INTRODUCTION**

1. The entertainment industry produces an overwhelming amount of data daily. With digital platforms and global cinema rising, it's crucial to analyze movie data to understand what drives success.
2. This project focuses on analyzing a diverse movie dataset using Python for data cleaning and visualization, Excel for dashboard summaries, and Power BI for interactive reporting.
3. The aim is to identify which genres, countries, and languages perform well, how ratings have evolved, and how audience votes correlate with financial success.
   1. **Background**

The global film industry produces thousands of movies annually across various genres, languages, and countries. With the rise of digital platforms, data related to box office revenues, viewer ratings, and audience demographics has become more accessible—but also more complex to analyze. Stakeholders such as producers, marketers, and content strategists need meaningful insights to make informed decisions. Yet, traditional analysis is often manual, time-consuming, or limited in scope.

**1.3 Problem Statement**

The films industry generates vast amounts of data, but lacks a streamlined approach to analyze and interpret it effectively. This project addresses the challenge of uncovering insights from global movie date by using Python, Excel, and Power BI to identify trends in genres, revenues, ratings, and regional performance. The goal is to transform raw data into actionable insights for better decision – making in cinema and media analysis.

* 1. **Objectives**

** To analyze global movie data using Python for cleaning and visualization**

** To use Excel for pivot-based summarization and static dashboards**

** To build interactive dashboards in Power BI for better insight delivery**

** To identify top-performing genres, countries, and language trends**

** To enable data-driven decision-making in the film and media industry**

**1.4 Outline of the study**

This study is divided into multiple stages

1. **Data Collection & Cleaning:** Using Python to preprocess and standardize movie data.
2. **Exploratory Analysis:** Leveraging Excel to build summary tables and visuals.
3. **Visualization:** Creating dynamic dashboards in Power BI for interactivity.
4. **Insight Extraction:** Deriving trends in ratings, revenue, genres, and country-wise performance.
5. **Conclusion:** Summarizing key findings and suggesting scope for future work.

## 📘 ****Chapter 2: Tools & Technologies Used****

In this project, multiple data science and business intelligence tools were used to collect, clean, analyse, and visualize a complex dataset of global movie records. The choice of tools was guided by their suitability for different stages of the data analysis pipeline — from raw data preparation to advanced dashboarding for decision-making.

**2.1 Python**

Python is a powerful, open-source programming language widely used in data science and analytics. It was used for the first stage of this project — data ingestion, cleaning, and exploratory analysis. Python helped in removing inconsistencies, handling missing values, and preparing the dataset for deeper insights.

**Key Capabilities Used:**

* **Data Cleaning:** Removing null values, duplicates, outliers
* **Column Standardization:** Renaming and formatting headers
* **Filtering:** Grouping data by Genre, Language, Country, etc.
* **Statistical Analysis:** Mean, median, correlation, etc.

**Libraries Used:**

* **pandas:** For data manipulation and Data Frame operations
* **matplotlib:** For basic plotting and charting
* **seaborn:** For advanced visualizations like heatmaps, bar plots, and line graphs

Example operations included. group by (), describe (), is null (), and visual functions like sns.barplot() and sns.heatmap().

**2.2 Microsoft Excel**

Microsoft Excel was used in the second stage of analysis. After cleaning the dataset in Python, the structured data was imported into Excel for intermediate-level exploration and dashboarding. Excel was particularly useful for stakeholders looking for printable summaries and reports.

**Excel Features Used:**

* **Pivot Tables:** To summarize revenue, ratings, genres, and country-wise performance
* **Charts & Graphs:** Line charts, column charts, and doughnut charts
* **Filters & Slicers:** For selecting specific years, genres, or countries
* **Conditional Formatting:** To highlight top and bottom performers

The dashboard created in Excel provided an overview of key metrics such as average rating by genre, revenue by country, and release trends over time.

**🔹 2.3 Microsoft Power BI**

Power BI was the primary tool for the final stage — advanced visualization and storytelling. It allowed the creation of interactive dashboards with drill-down capabilities, KPI visuals, and filters that could be used by non-technical stakeholders.

**Components Used:**

* **Power BI Desktop:** Used for creating visuals and designing dashboard layout
* **DAX (Data Analysis Expressions):** Used to create calculated fields like Total Revenue, Avg Rating, and Movie Count
* **Visuals:** Bar charts, line graphs, card visuals, column charts
* **Interactivity:** Slicers for Genre, Country, Language, Year, etc.
* **Tooltips & Drilldown:** Enhanced interactivity for deep insights

The Power BI dashboard integrated all relevant metrics and visualizations, enabling a 360° view of the dataset with real-time filtering capabilities.

**🔹 2.4 Supporting Tools**

In addition to the main tools, the following technologies were used to support the presentation and structure of the project:

* **Jupyter Notebook:** The Python environment used for writing, executing, and visualizing code outputs.
* **MS Word:** Used for documenting the final project report.
* **ChatGPT:** Used to generate a simulated, realistic dataset for project use, and for documentation assistance.

**Chapter 3: Dataset Description**

This Data Set I have generated from ChatGPT reason behind is that IMDB cannot revealed is that and data which I need is to much and I have scraped the data by using beautiful soup but it does not give required data the data which I have generate is 200000 movies which is release in 2024 and it is collection globally data and actual fact is that the data which I have use is a small part of huge data because globally there to much movies I realised and as a student it is difficult for me to analyse to much data so I take small part and on that basis I give my insights and my data set consist of 200000 movies which is globally released and it also had column of genre which describe the movie type e.g. – action , documentary etc and my data has column of box office collection , budget of that movie and date of release and also have column of IMDB Ratings which is Out Of 10 and column of Rotten Tomatoes which means movie which is release is flop in Ott or threater and also have column of Release Date , Country and language as per requirement I have made necessary changes in my data set like in excel I have added some new column and power bi also according to the requirements of excel sheet and power bi dashboard which I give brief explination during the process

Steps applied in dataset preprocessing:

 🧹 **Cleaning and Preparation Steps:**

* Programming language used: **Python**
* Null value handling
* Standardizing text case (e.g., genre, language)
* Removing duplicates
* Ensuring numeric typecasting for analysis
* Exported cleaned dataset to Excel

 **Ethnicity:** **Countries Represented:**

* USA, India, UK, France, Japan, South Korea, Spain

**Languages Diversity:**

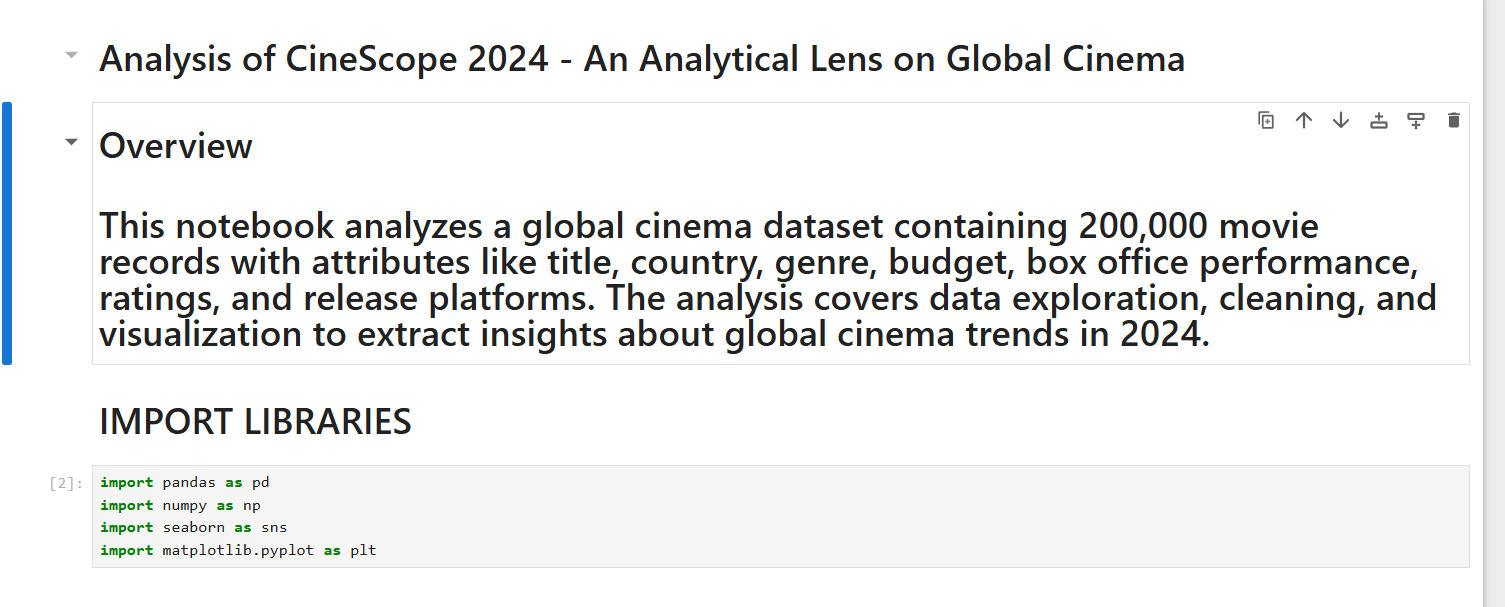
* English, Hindi, Korean, Japanese, French, Spanish

**Purpose of Dataset:**

To explore trends in:

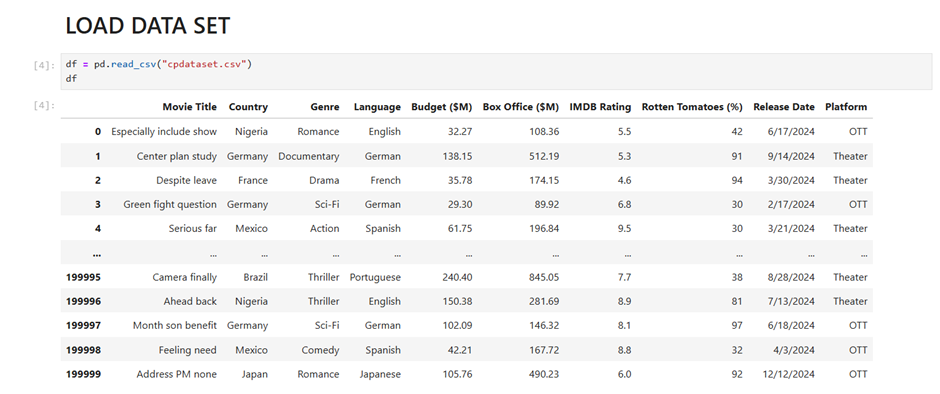
* Top-rated genres
* High-revenue films by country
* Yearly release performance
* Viewer engagement (via votes and ratings)

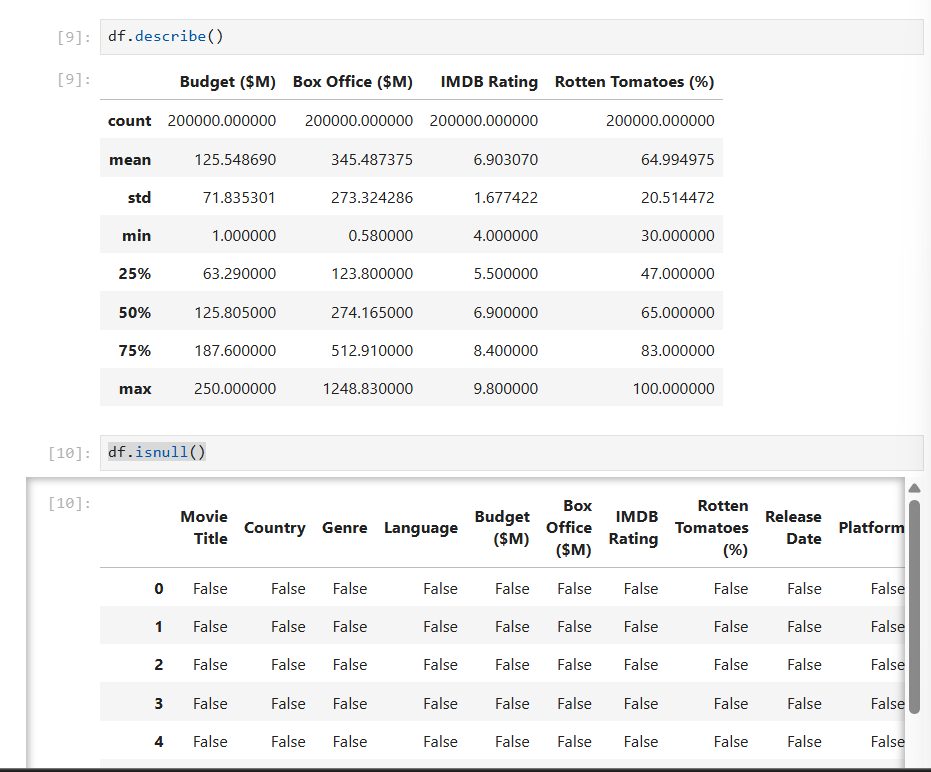
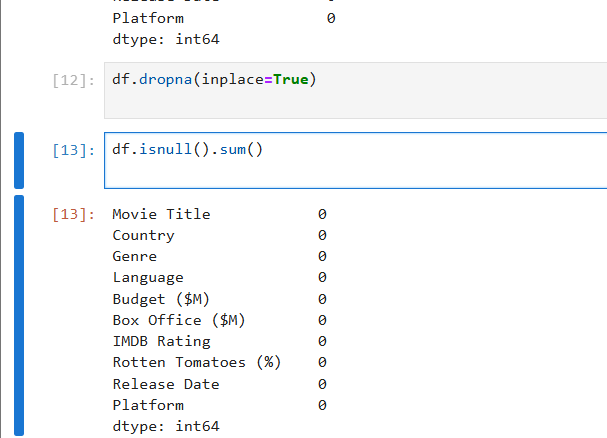
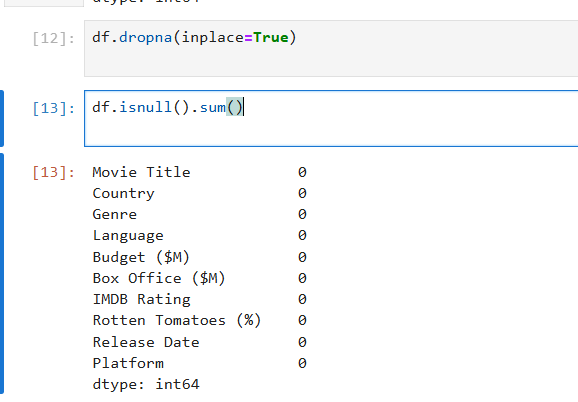
📄 **Chapter 4: Data Analysis Using Python**



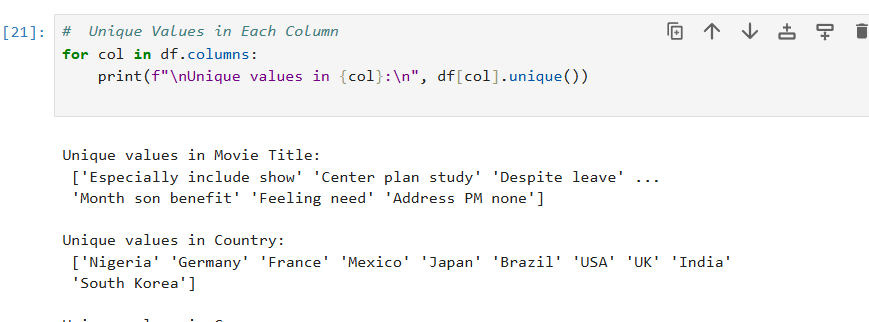
So first I have given overview of my project and after that I have imported libraries like pandas and numpy intended for data manupliation and matplotlib and seaborn for data visualization and I have write basic syntax for all import then library name.

Then I have upload my code as df = pd.read\_csv(“path of file or data”) it is the basic syntax for import my data set in jupyter note book for analysis and I am going to show ss and if I want to check my data I simply write df now below I am going to show the ss of upload data set code

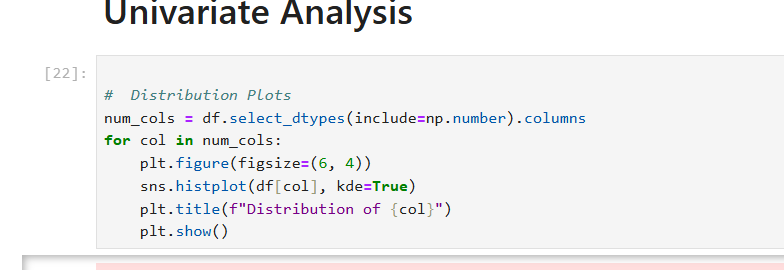


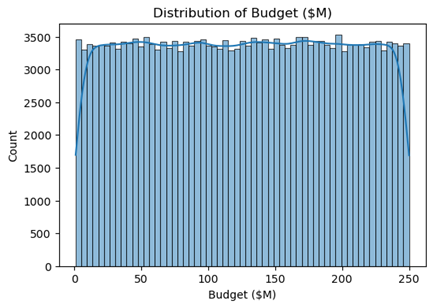
Then after uploading I have done some cleaning part which consist of different syntax 1> df.info () 2>df.describe() 3>df. Shape 4> df.isnull() 5> df.dropna(inplace=True) 6> df.isnull().sum() these are some codes which I use to clean data set if I have different values so (syntax1)give information of data se and( 2 syntax) describe whole data set which give u stats value to like standard seviation mean , median and mode and (3 syntax) shape which tell you number our dataset contain how many rows and columns were present (4 syntax) is null is help to show null values were there in dataset or not if there is no null values then we can proceed with EDA part otherwise we have to deal with null values by using mean median method and another method is use to drop that particular column from data set by jusing(5 syntax) df.dropna(inplace=True) and (syntax 6) df.isnull().sum() which I have use to check null values is there or not   

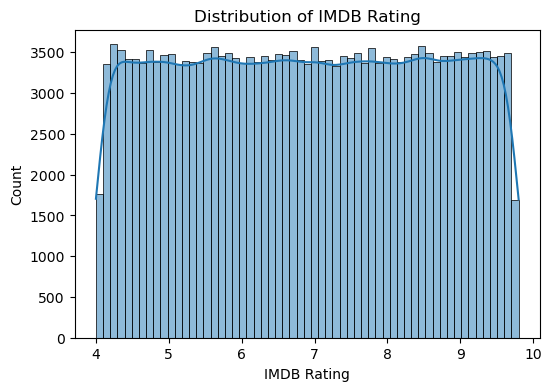
After Cleaning data Then we move to EDA part which means Exploratory Data Analysis in which I have first find unique values from my data set and after words I have applied univariant analysis which means It refers to the analysis of data that involves **a single variable** at a time and after words I have apply codes of bivariate analysis which show the relationship between two columns in numerical and textual basis also now I uploading code image and output image

This is the code for Unique values

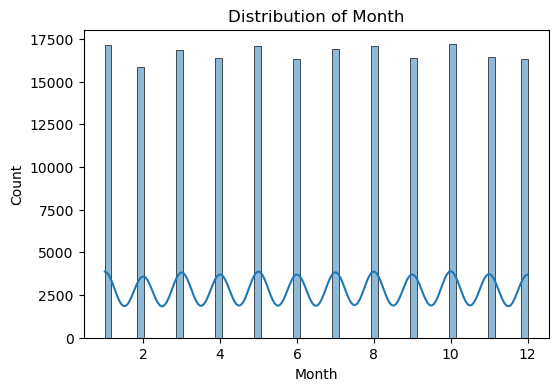


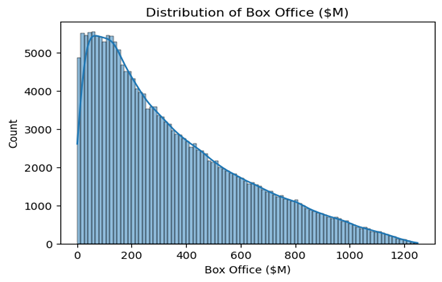
here is the code of univariant which give distribution graph of all numerical row which is present in my data set and on next page I will upload all images of univariate analysis result

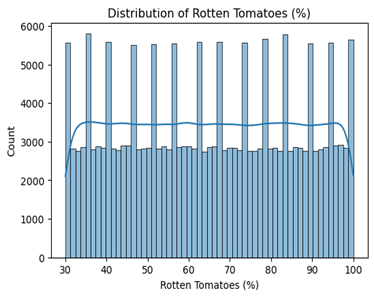










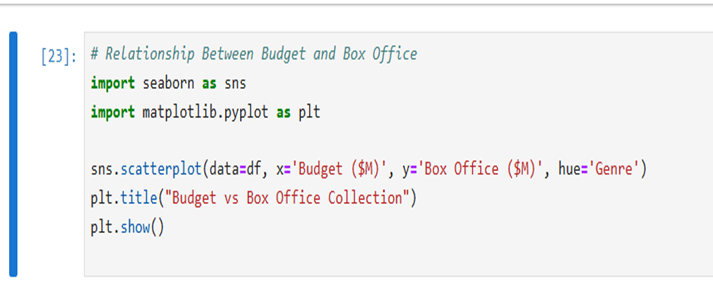


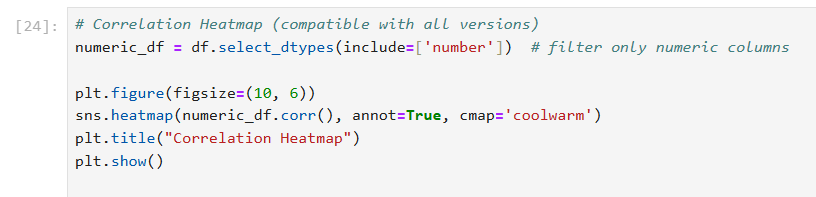
Here in all graphs these are outputs of Univarient anlysis here

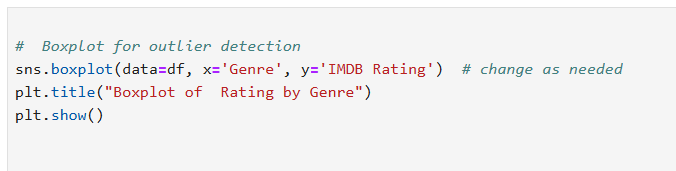
It explain different types of distribution according to data

* 1. Distribution of budget - it show distribution of total buget in which we observe there is slightly difrrence among them
  2. Distribution of month - it show like equal distribution which means in every month movie release equaly globally
  3. Distribution of IMDB Rating - it show the difrrence between all ratings but we observe that there is only point diffrence
  4. Distribution of boxoffice – left skewd graph which means boxoffice collection increase and then decrease after some time
  5. Distribution of rotten tomatoes - which show the distribution among all country in which country like movie is flop acc to data

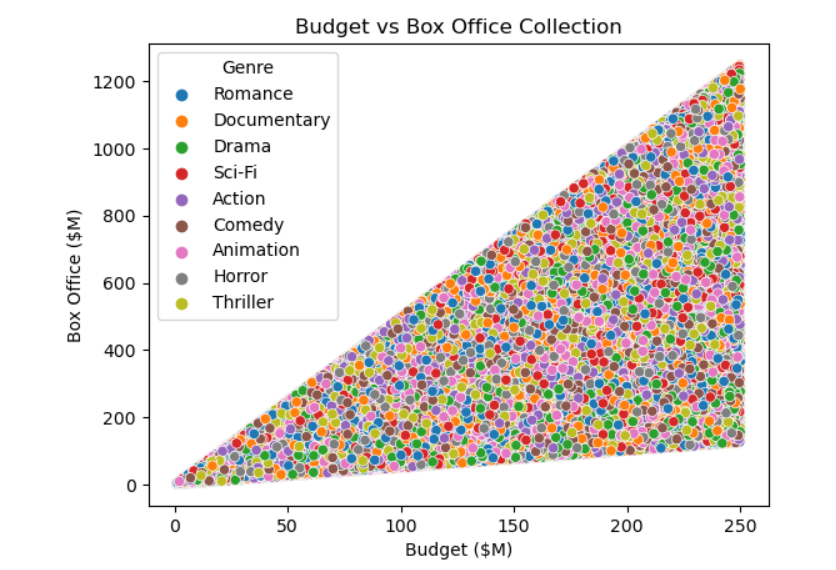
**Bivariate Analysis** - Bivariate data involves analyzing the relationship between two different variables at the same time. According to my analysis I have use scatterplot to show the relationship between budget and boxoffice And after word I have plot a heatmap which show corelationbetween all coloumns which have numerical value after words I have apply box plot in my dataset to check weathere my data consist of outlierers or not but there is no outliers in my data set if I found outliers in my data set then I can remove them by using IQR method so in this page I am going to show my codes till box plot and result will be shown in next page



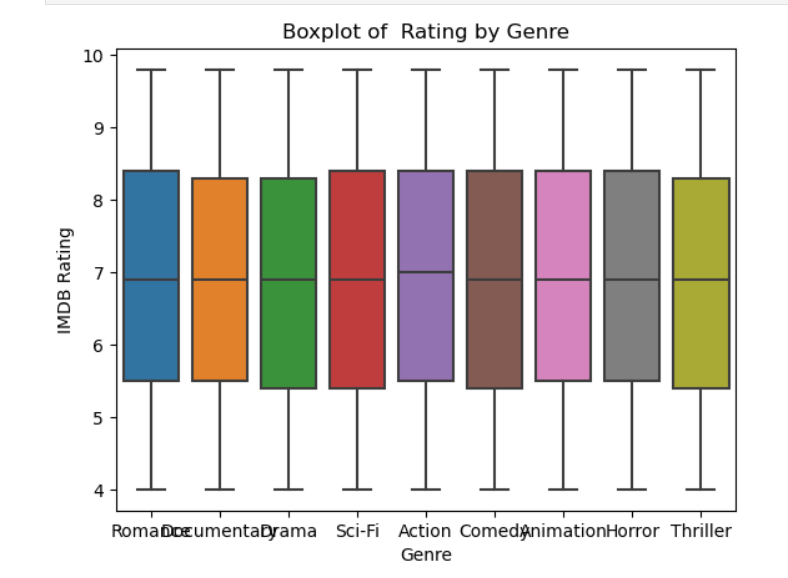
****

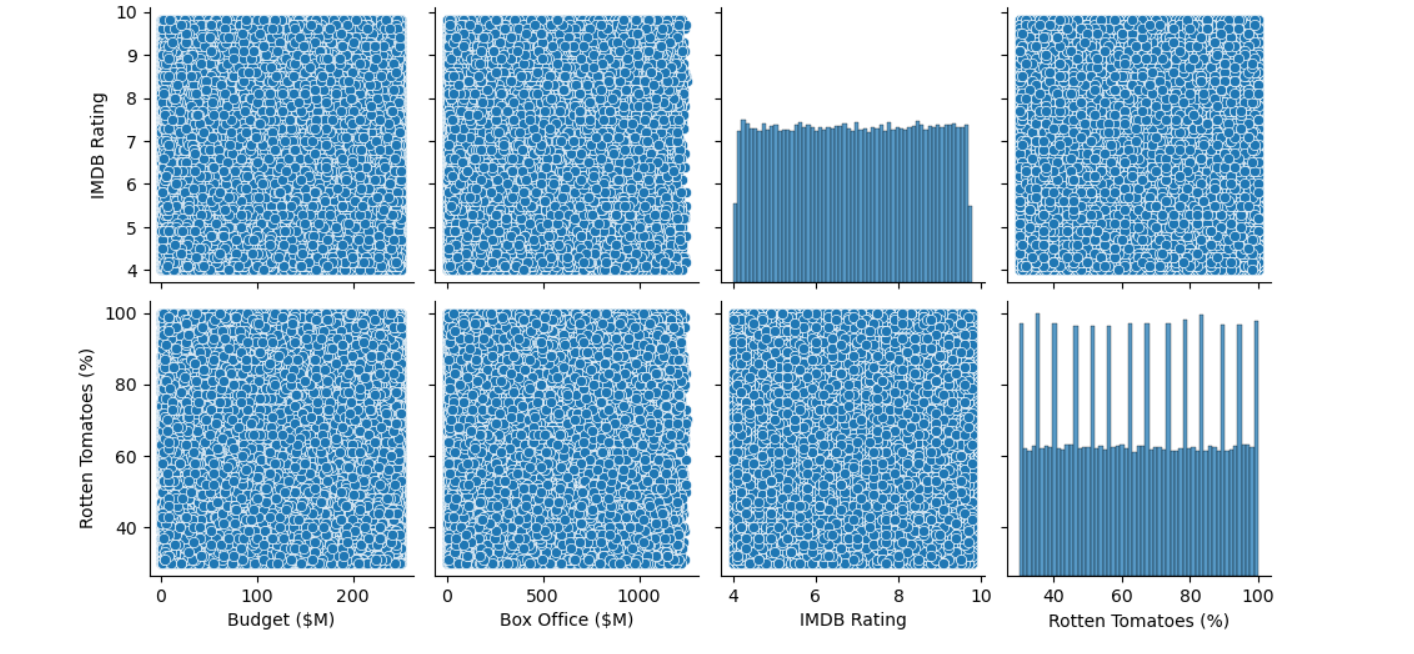
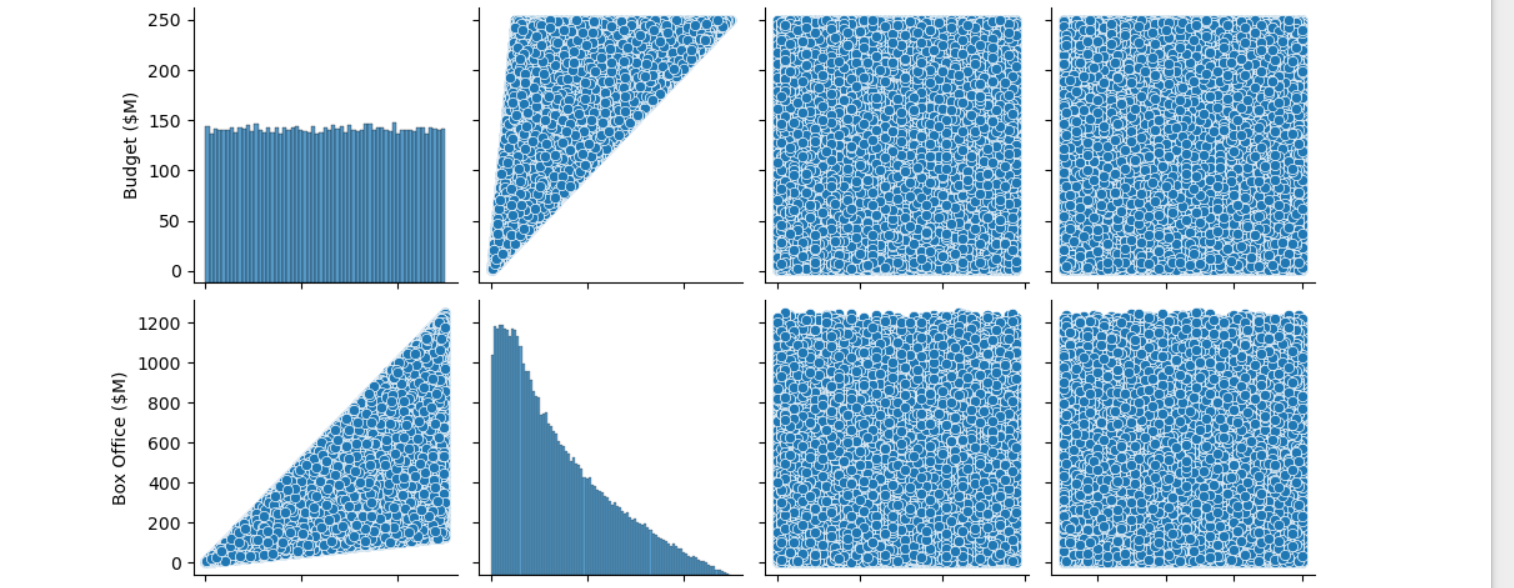


As I have explained about codes in previous page now I am uploading my results



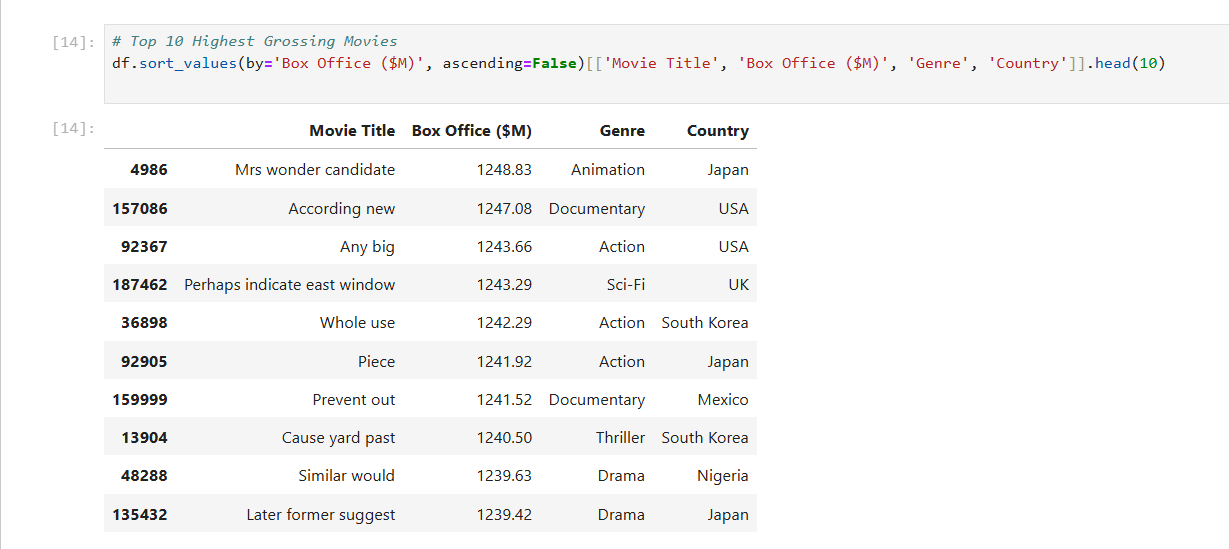




After Univariate and Bivariate I have used pair plot it show individual columns graph it is last EDA Gr

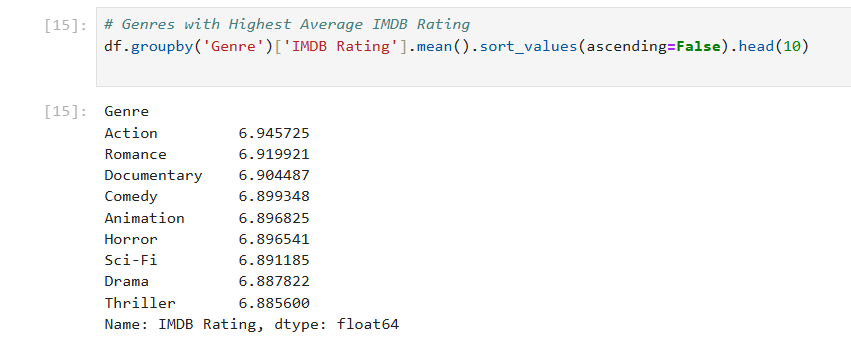
Now after EDA part we are move to the Last Part of it which we Discus in Next Page

After EDA Now I am going to upload my codes with insights and I am going to explain all in detail



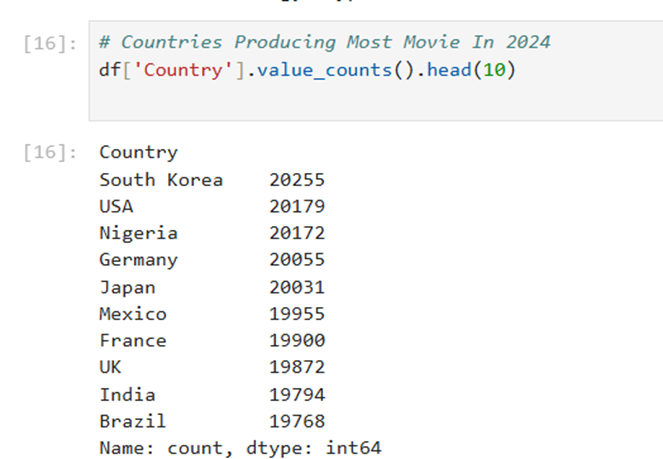
AT first we have found which top 10 highest movies ?

- so we found animation movie from Japan is the no 1 in top 10 movies



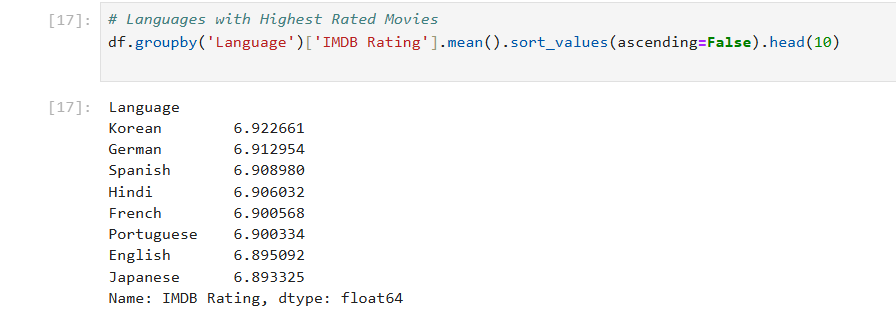
In above code we have to find which type of movie has highest rating ?

* + I have found Action has the highest rating among all



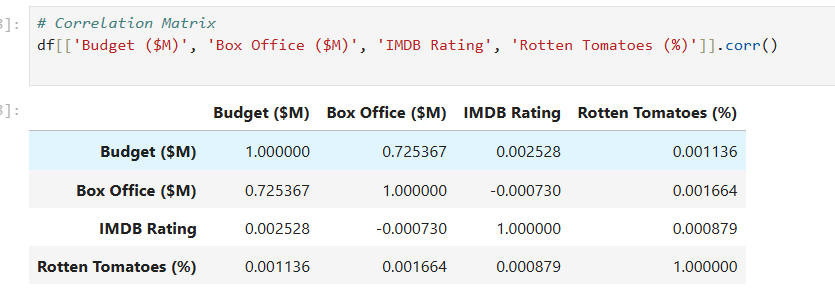
In above code I have try to find which country producing most movies in 2024 ?

* + I have found south korea producing most movie in 2024



In above code I try to find in which language people see most of the movies?

- So I have found most of the people watch movies in Korean language among all

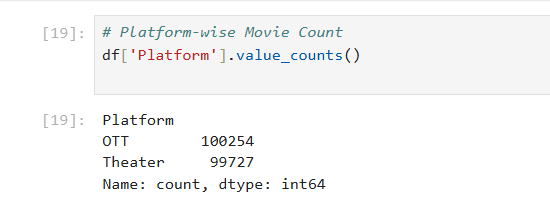


Here in this code I try to find correlation between all coloums



In this code I try to find in which month most of the movies released

* + 10 month in which most of the movies released



In above code I try to find which platform release most of the movies

* + I found OTT has released most of the movies

# **Key Observations**

# **1. Data Loading and Initial Exploration**

#The dataset contains 200,000 movies with 10 features including financial metrics, ratings, and metadata

Initial examination shows some missing values in "Movie Title" (19 missing)

Basic statistics reveal:

Average budget: $125.55M

Average box office: $345.49M

Average IMDB rating: 6.9

Average Rotten Tomatoes score: 65%

# **2. Data Cleaning**

Removed 19 records with missing movie titles

Verified no remaining null values after cleaning

Converted "Release Date" to datetime format and extracted month for temporal analysis

# **3. Key Insights**

Top Performing Movies The highest grossing films are dominated by action, documentary and sci-fi genres

Top films come from diverse countries including Japan, USA, UK, South Korea, Mexico and Nigeria

The highest box office was $1.248B for "Mrs wonder candidate" (Animation/Japan)

Genre Analysis Action films have the highest average IMDB rating (6.95)

Romance and documentary films follow closely in ratings

Thrillers have the lowest average rating among major genres (6.89)

Geographic Distribution South Korea produces the most films (20,255)

Followed closely by USA (20,179) and Nigeria (20,172)

Germany, Japan and Mexico also have significant production volumes

Language Performance Korean-language films have the highest average IMDB rating (6.92)

German and Spanish films also perform well in ratings

English-language films are slightly below average (6.89)

Platform Distribution OTT platforms host slightly more films (100,254) than theaters (99,727)

This suggests a near-even split between streaming and theatrical releases

Temporal Patterns Film releases are relatively evenly distributed throughout the year

January and October have slightly higher release volumes

December has fewer releases, possibly avoiding holiday competition

# **4. Correlation Analysis**

Budget and box office show moderate positive correlation (0.725)

Ratings show minimal correlation with financial metrics

This suggests high budgets often lead to better box office but don't guarantee quality

# **5. Notable Findings**

Korean language films outperform others in ratings despite not being the most produced

Documentaries achieve both high ratings (6.90) and strong box office performance

The film industry appears balanced between traditional theaters and streaming platforms

Monthly release patterns show consistent volume year-round with minor seasonal variations

# **Conclusion**

The analysis reveals a vibrant global film industry with diverse production across countries and genres. While financial investment correlates with box office success, it doesn't strongly predict critical reception. The emergence of streaming platforms has created near-parity with theatrical distribution, and non-English language films (particularly Korean) are achieving both critical and commercial success. These insights could inform production, distribution and investment decisions in the global cinema market.

The dataset provides rich opportunities for deeper analysis, particularly in examining relationships between production countries, genres, and performance metrics across different release platforms.

**Chapter – 5**

**Data Analysis Using Excel**

**1. DATA SET**

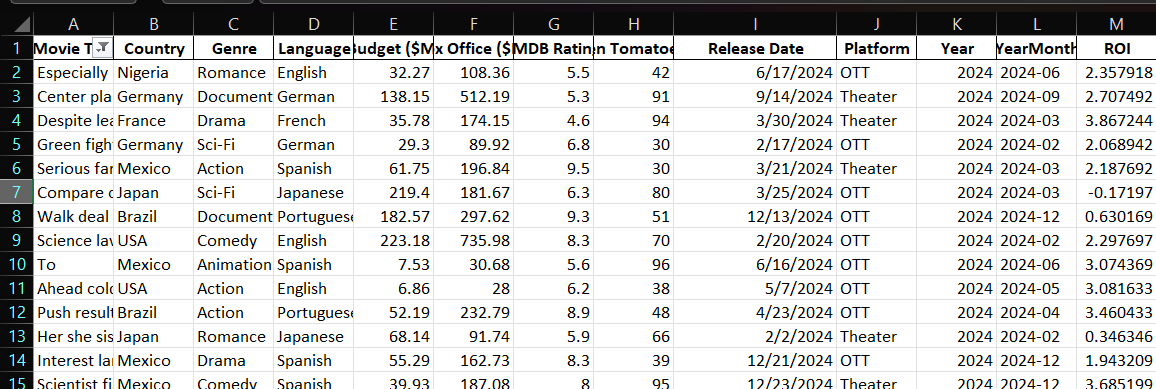
**🔹 Description:**

This is master dataset — combining movie-level data including:

* Movie Title, Country, Genre, Language
* Budget, Box Office, ROI
* IMDB Rating, Rotten Tomatoes, Platform
* Release Date, Year, and Year Month

**🔹 Purpose:**

This sheet is used as the **base for all pivot tables and charts**.



As above Data Set I have add 2 new columns year month and ROI

**🔷 2. BOX OFFICE BY GENRE**

**🔹 Sheet Name: BOX OFFICE BY GENRE**

**🔹 Data Source:** Group from main dataset: excel

=SUMIFS('DATA SET’! Box-office, 'DATA SET’! Genre)

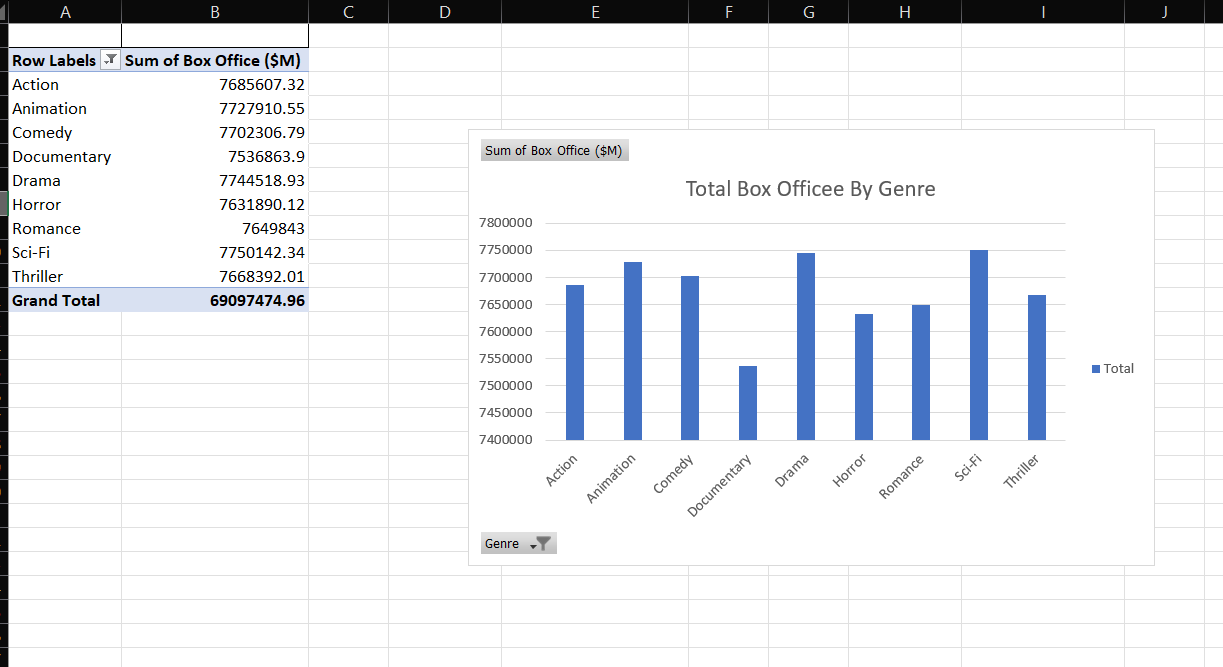
**🔹 Description:**

This pivot table shows **total box office revenue per genre**.

| **Genre** | **Total Box Office ($M)** |
| --- | --- |
| SCI-FI  Drama | 7,750,142  7,744,518 |
|  |  |
| Comedy | 7,702,306 |
| Documentary | 7,536,863 |

**🔹 Chart Type: Column Chart**

* X-axis: Genre
* Y-axis: Box Office Revenue



" SCI-FI were the highest-earning genres, indicating strong global commercial appeal."

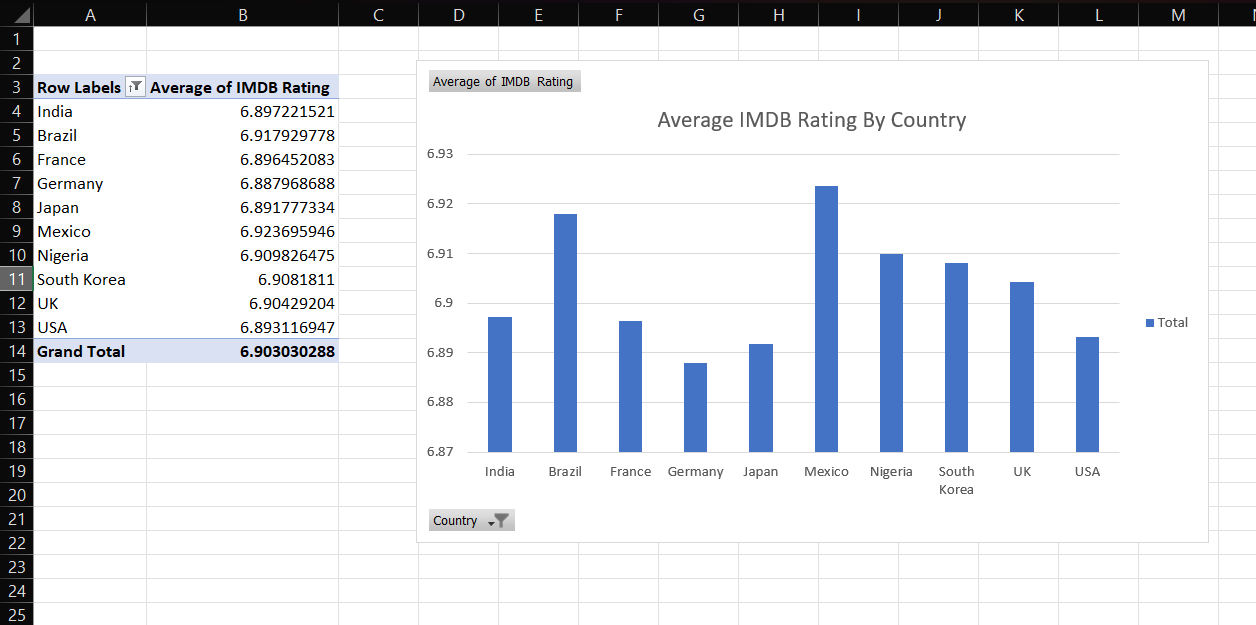
…………………………………………………………………………

**🔷 3. AVG IMDB RATING BY COUNTRY**

**🔹 Sheet Name: AVG IMDB RATING BY COUNTRY**

**🔹 Description:** Shows **average IMDb rating per country**.

| **Country** | **Avg IMDb** |
| --- | --- |
| India | 6.89 |
| Brazil | 6.91 |
| France | 6.89 |



Mexico has the highest avg imdb rating above all country

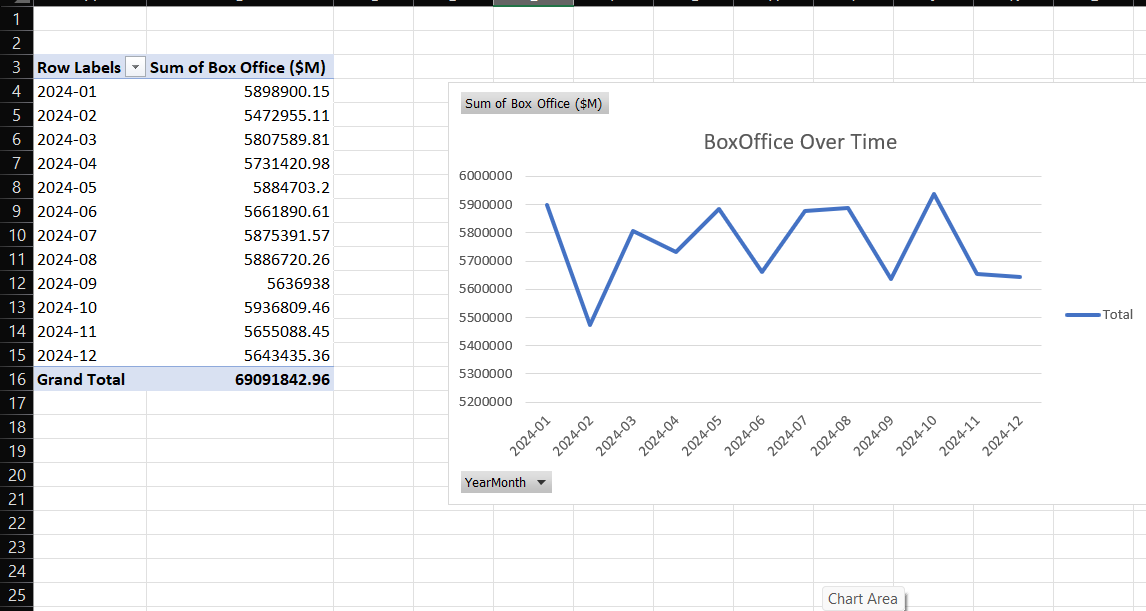
…………………………………………………………………………

**🔷 4. Monthly Box Office Trends**

**🔹 Sheet Name: Monthly Box Office**

**🔹 Description :** Sum of revenue based on release month (Year Month)

| **Month** | **Revenue ($M)** |
| --- | --- |
| 2024-01 | 5,898,900 |
| 2024-02 | 5,472,955 |
| 2024-03 | 5,807,589 |



According November is the peak time for box-office collection in 2024

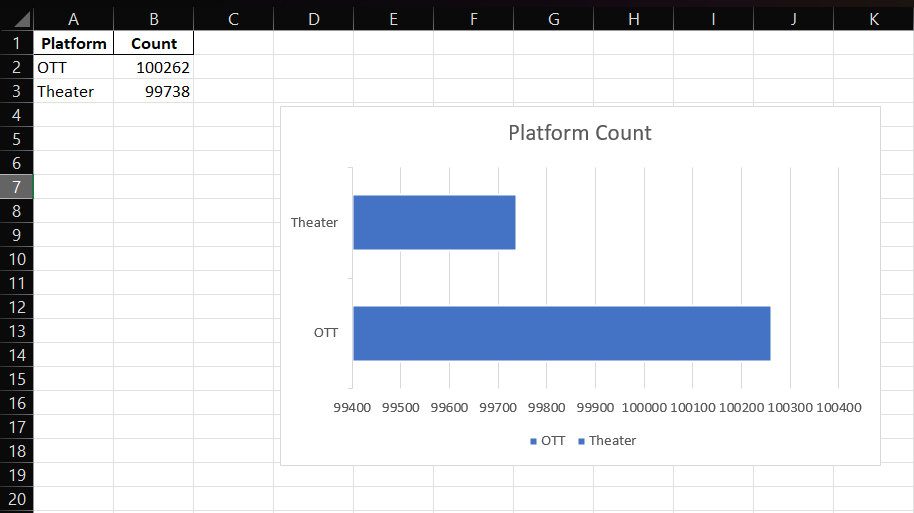
…………………………………………………………………………

**🔷 5. Platform Count**

**Sheet Name: Platform Count**

**Description :** The number of users watched movies in the platform Like OTT AND Theater

|  |  |
| --- | --- |
| Platform | Movie Count |
| OTT | 100,262 |
| Theater | 99,738 |



According to my data I found most people like to watch movies on OTT

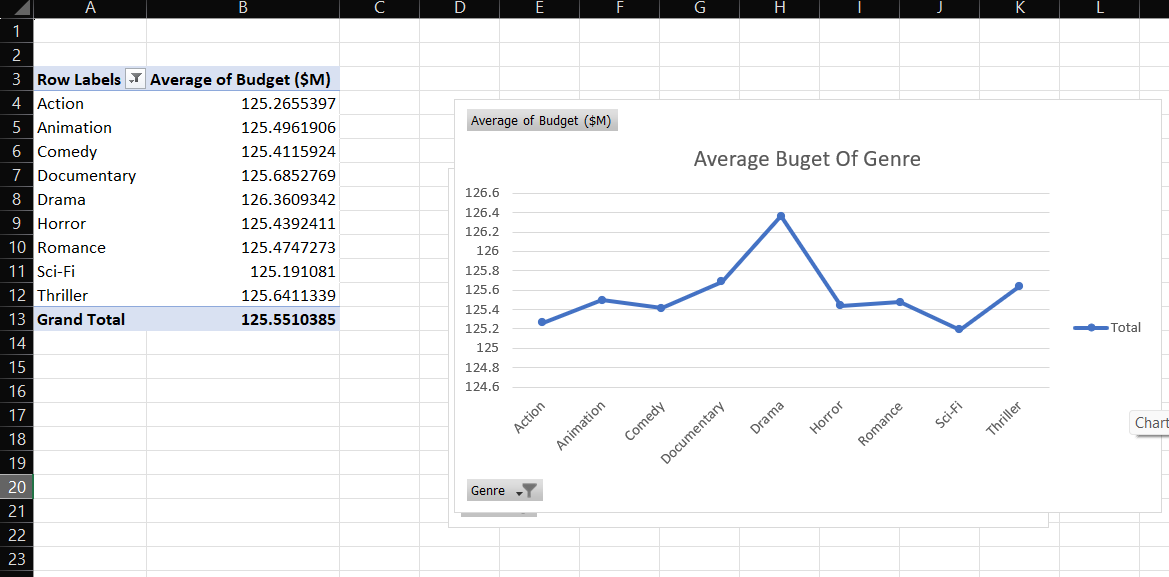
…………………………………………………………………………

**🔷 6. Average Budget by Genre**

**🔹 Sheet Name: Average Genre Budget**

| **Genre** | **Avg Budget ($M)** |
| --- | --- |
| Action | 125.26 |
| Animation | 125.49 |
| Comedy | 125.41 |

Drama 126.36



Drama have the highest average budget according genre

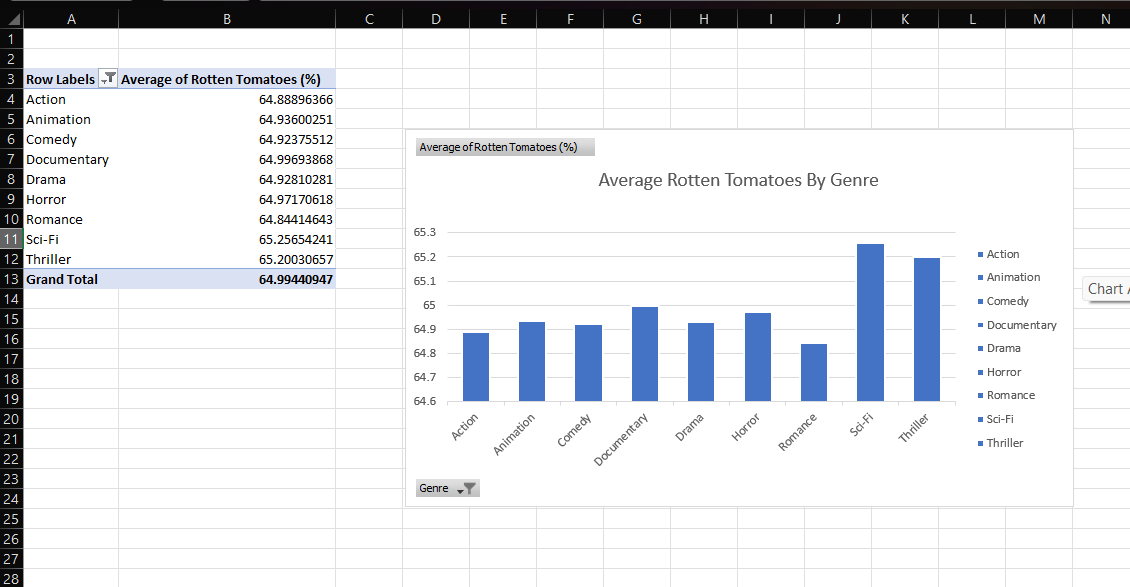
……………………………………………………………………………………

**🔷 7. Avg Rotten Tomatoes Score by Genre**

**🔹 Sheet Name: Avg Rotten Tomatoes by Genre**

| **Genre** | **Avg RT Score (%)** |
| --- | --- |
| Action | 64.88 |
| Animation | 64.93 |
| Comedy | 64.92 |

SCI-FIR 65.25



SCI-FRIC IS MOST AVERAGE ROTTEN TOMATOES BY GENRE

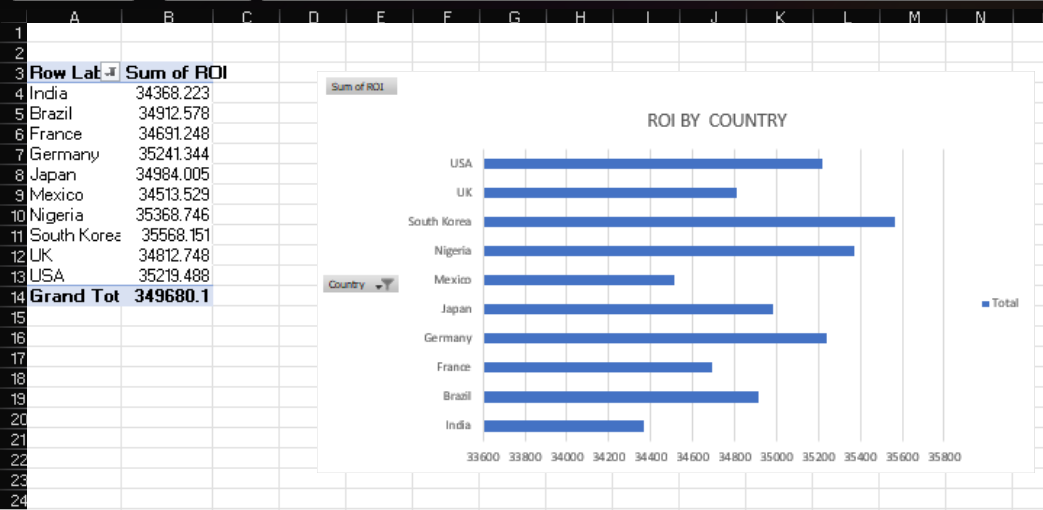
……………………………………………………………………………………

**🔷 8. ROI BY COUNTRY**

**🔹 Sheet Name: ROI BY COUNTRY**

| **Country** | **Total ROI** |
| --- | --- |
| India | 34,368 |
| Brazil | 34,912 |
| France | 34,691 |

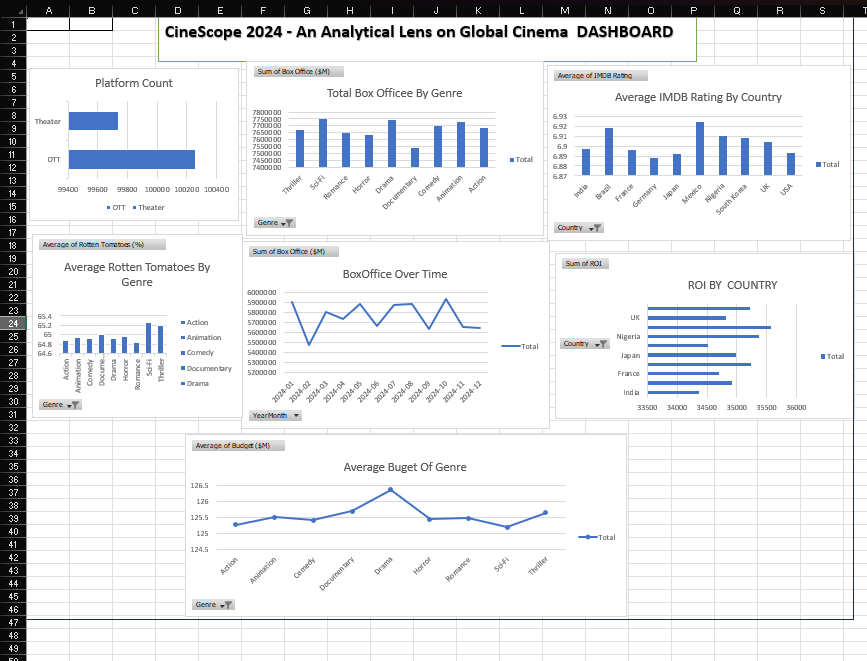
South – Korea 35,568



South Korea is has highest number of ROI Among all country

……………………………………………………………………………………

|  |
| --- |
|  |



**✅ DASHBOARD OVERVIEW:**

The dashboard is a **visual summary** created using Excel charts, pivot tables, and slicers, offering insights into various aspects of global cinema in 2024.

It includes analysis based on:

* Genres
* Countries
* Platforms (OTT/Theater)
* Budget, Revenue, Ratings, and ROI

**📊 DASHBOARD COMPONENTS WITH EXPLANATION & INSIGHTS:**

**🔹 1. Platform Count (Bar Chart)**

* **Shows:** Number of movies released on each platform (OTT vs Theater).
* **Insight:**
  + OTT releases ≈ **100,262**
  + Theater releases ≈ **99,738**

✅ **Observation:**

There is a **slightly higher number of OTT releases**, showing a continuing trend toward digital platforms post-pandemic.

**🔹 2. Total Box Office by Genre (Bar Chart)**

* **Shows:** Total box office revenue by genre.
* **Genres Included:** Thriller, Sci-Fi, Romance, etc.

✅ **Observation:**

**Action and Animation** lead the charts, with **highest total box office revenues**, followed by Comedy and Thriller.  
**Documentary and Drama** genres show relatively lower earnings.

**🔹 3. Average IMDb Rating by Country (Bar Chart)**

* **Shows:** Country-wise average IMDb ratings.

✅ **Observation:**

**Japan and Mexico** stand out with **higher average IMDb ratings**, suggesting better critical reception.  
Countries like the **UK and USA** show moderate ratings.

**🔹 4. Average Rotten Tomatoes by Genre (Bar Chart)**

* **Shows:** Genre-wise average Rotten Tomatoes score.
* **Genres Included:** Action, Comedy, Drama, Romance, etc.

✅ **Observation:**

**Romance and Animation** receive **better Rotten Tomatoes scores**, while genres like Action and Horror have slightly lower averages.

**🔹 5. Box Office Over Time (Line Chart)**

* **Shows:** Total revenue trend by release month (2024-01 to 2024-12).

✅ **Observation:**

Box office performance **peaks in March, July, and October**, suggesting seasonal trends — likely due to holidays, festivals, or summer/winter breaks.  
Slight dip in February and December.

**🔹 6. ROI by Country (Bar Chart)**

* **Shows:** Return on Investment (ROI) by country.

✅ **Observation:**

**Nigeria and UK** show **highest ROI**, indicating strong profitability despite varyin productionbudgets.  
Other countries like **India and France** follow closely.

**7. Average Budget of Genre (Line Chart)**

* **Shows:** Average production budget for each genre.

✅ **Observation:**

**Drama and Animation** have the **highest average budgets**, likely due to productioncomplexityorcastsize.  
**Comedy and Action** have slightly lower average spending.

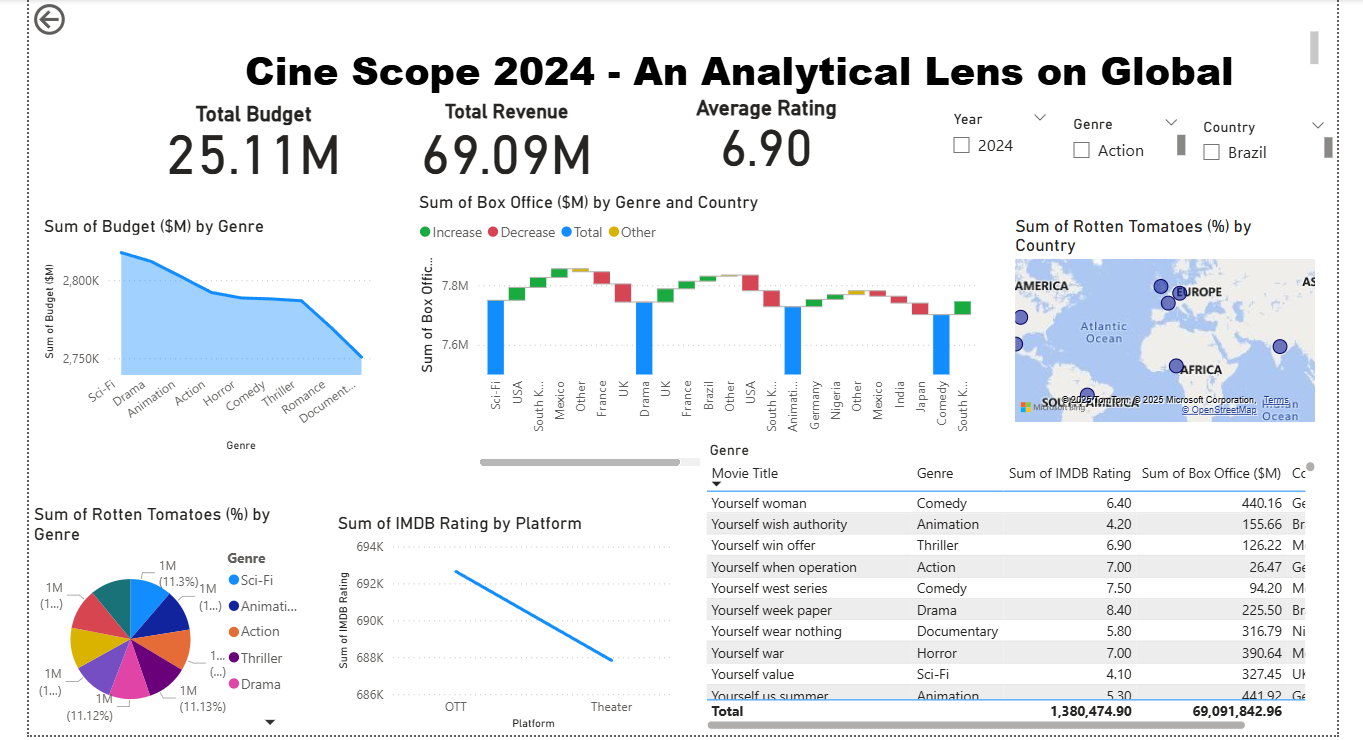
……………………………………………………………………………………

**KEY FINDINGS :**

1. **OTT Dominance:** OTT platforms are now slightly more dominant than theaters in terms of volume, showing digital transformation of cinema.
2. **High-Earning Genres:** Action and Animation genres generate the most revenue, making them commercially favorable.
3. **Critically Acclaimed Countries:** Japan and Mexico show higher IMDb ratings, pointing to strong storytelling or niche cinema success.
4. **Viewer Preferences:** Genres like Romance and Animation perform better on Rotten Tomatoes — indicating viewer satisfaction.
5. **Seasonal Revenue Trends:** Box office collections rise in specific months (March, July, October), ideal for major releases.
6. **ROI Hotspots:** Nigeria, UK, and Japan are among the most profitable production hubs based on ROI metrics.
7. **Production Budgets:** Drama and Animation are costlier genres, which may be balanced by higher critical or commercial returns.

**Chapter 6:**

**Data Visualization Using Power BI**



**🔶 1. Introduction to Power BI**

**Power BI** is a powerful business analytics tool by Microsoft that enables users to visualize data, share insights, and build interactive dashboards. It connects to multiple data sources (Excel, SQL, APIs, etc.), allowing users to analyze and present data dynamically.

**🔶 2. Objective of Power BI Dashboard in the Project**

The Power BI dashboard was developed to provide a **dynamic and interactive visual exploration** of global cinema data under the project title:

🎬 **CineScope 2024: An Analytical Lens on Global Cinema**

It allows viewers to:

* Filter by **Genre, Country, Platform**
* View **KPIs like total revenue, average IMDb**
* Explore trends and comparisons across **time, geography, and category**

**🔶 3. Data Source and Preparation**

* Imported Excel dataset: EXCEL PART OF PROJECT.xlsx
* Cleaned and structured in Power Query:
  + Converted Release Date into Year and Month
  + Ensured data types: Numeric (Revenue, ROI), Text (Genre, Country), Date
  + Created additional columns like Year Month for time series

**🔶 4. Key Visuals Used in Power BI Dashboard**

Below is a detailed explanation of each visual used in the dashboard (based on the same ones in Excel, but interactive):

**✅ 1. KPI Cards (Top Metrics)**

* **Total Revenue**
* **Average IMDb Rating**
* **Average Rotten Tomatoes**
* **Total Movies Count**

📝 **Use in Report:**

These cards give users a snapshot of key performance indicators for the global movie industry in 2024.

**✅ 2. Platform Distribution (Donut or Bar Chart)**

* Axis: Platform
* Value: Count of Movies

📝 **Insight:**

OTT and Theater releases can be compared quickly. Filters update this dynamically.

**✅ 3. Genre-wise Box Office (Bar/Column Chart)**

* Axis: Genre
* Value: Sum of Box Office ($M)

📝 **Insight:**

Action, Animation, and Comedy dominate revenue.

**✅ 4. Average IMDb Rating by Country (Bar Chart)**

* Axis: Country
* Value: Average IMDb Rating

📝 **Insight:**

Allows us to identify which countries produce the most critically acclaimed films.

**✅ 5. Monthly Box Office Trend (Line Chart)**

* Axis: YearMonth
* Value: Sum of Revenue

📝 **Insight:**

Reveals seasonal or release-based spikes in earnings.

**✅ 6. Average Budget by Genre (Line or Bar Chart)**

* Axis: Genre
* Value: Average of Budget ($M)

📝 **Insight:**

Genres like Drama and Sci-Fi typically have higher production costs.

**✅ 7. ROI by Country (Bar Chart)**

* Axis: Country
* Value: Sum or Average of ROI

📝 **Insight:**

Highlights where investment yields the highest return.

**✅ 8. Average Rotten Tomatoes by Genre (Clustered Bar)**

* Axis: Genre
* Value: Avg Rotten Tomatoes Score

📝 **Insight:**

Romance and Animation receive the highest approval ratings.

**🔶 5. Use of Slicers and Filters**

Power BI dashboard includes slicers for:

* **Genre**
* **Country**
* **Platform**
* **Year/Month**

📝 **Use in Report:**

Slicers enable users to interactively explore how different variables affect performance metrics.

**🔶 6. DAX Measures Created**

Here are sample DAX formulas (you can include these in an appendix or explanation section):

DAX

CopyEdit

Total Box Office = SUM('DATA SET'[Box Office ($M)])

Average IMDb Rating = AVERAGE('DATA SET'[IMDB Rating])

Average Rotten Tomatoes = AVERAGE('DATA SET'[Rotten Tomatoes (%)])

Total Movies = COUNTROWS('DATA SET')

These measures were used in **cards, charts, and comparative visuals**.

**🔶 7. Dashboard Design Principles Used**

* Clean, grid-based layout
* Soft color palette for readability
* Rounded cards and smooth transitions
* Hierarchical filtering (e.g., Genre → Country)
* KPIs at the top, trends in the middle, comparisons at the bottom

**🔶 8. Final Dashboard View (Screenshot Suggestion)**

📸 **Screenshot Captions:**

1. Overall Power BI dashboard (full view)
2. Genre-wise box office (filtered by OTT)
3. ROI by Country (filtered by Drama)
4. Monthly Revenue trend (full year)

📌 Make sure to show **filters/slicers in action** for interactivity.

**🔶 9. Key Insights from Power BI Dashboard**

1. **Action & Animation genres** perform best in terms of revenue.
2. **Japan and Mexico** lead in IMDb ratings.
3. **March, July, and October** are peak months for earnings.
4. **Nigeria and UK** deliver high ROI, making them attractive for investment.
5. **OTT releases slightly outpace theater**, reflecting global streaming growth.
6. **Drama and Animation** have the highest production budgets.

**🔶 10. Conclusion**

Power BI allowed for a **dynamic exploration of cinema trends** in 2024, offering deeper insights compared to static Excel visuals. Its interactive capabilities make it ideal for decision-makers in media analytics and film distribution.

**📘 Chapter 7: Conclusion & Learning Outcome**

**✅ Conclusion**

This project, **Cine Scope 2024: An Analytical Lens on Global Cinema**, successfully demonstrated how data science tools can transform raw movie data into meaningful business intelligence. By combining Python, Excel, and Power BI, we were able to clean, analyze, and visualize a large dataset of 200,000 global film records from 2024.

Our findings revealed valuable insights:

* **Genres like Action and Animation** dominate the global box office in revenue.
* **Korean-language films** have the highest average IMDb ratings, outperforming even English-language content.
* **OTT platforms** slightly surpass theatrical releases, indicating a major shift toward digital consumption.
* **Countries such as South Korea, Nigeria, and Japan** offer both high production volumes and strong ROI.
* **Seasonal patterns** like peaks in March, July, and October highlight optimal release windows for filmmakers.
* While **budget and revenue** show a strong correlation, critical success (ratings) is more influenced by content and language than production cost.

The project confirms that visual analytics through Power BI dashboards and statistical programming can help producers, marketers, and analysts in the media industry make smarter, data-driven decisions.

**✅ Learning Outcomes**

Through the completion of this project, I gained comprehensive hands-on experience with:

1. **Python for Data Preprocessing and EDA**
   * Loading large datasets using pandas
   * Cleaning, transforming, and visualizing data using matplotlib and seaborn
   * Generating univariate and bivariate insights
2. **Excel for Pivot Analysis and Static Dashboards**
   * Building pivot tables and charts
   * Designing print-friendly dashboards
   * Using conditional formatting and filters to uncover insights
3. **Power BI for Advanced Visualization**
   * Creating a fully interactive dashboard with KPI cards, slicers, and charts
   * Using DAX formulas for dynamic metrics (e.g., Avg Rating, Total Revenue)
   * Designing responsive layouts for professional storytelling
4. **Data Storytelling and Presentation Skills**
   * Structuring analytical findings for academic and business audiences
   * Translating numbers into actionable observations
   * Documenting and visualizing every step of the analysis workflow

**📘 Chapter 8: Bibliography & References**

This section includes the references used for dataset design, analysis methodology, tools used, and supporting literature for insights and visual storytelling in the project.

**🔹 Books & Online Courses References**

1. McKinney, Wes. *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*, O’Reilly Media, 2018.
2. Microsoft Power BI Documentation – <https://learn.microsoft.com/power-bi>
3. Excel Pivot Table Tutorials – Microsoft Support – <https://support.microsoft.com/excel>
4. Seaborn Documentation – https://seaborn.pydata.org/
5. Coursera: Data Visualization with Python – IBM Skills Network
6. Udemy Course: Power BI A-Z – https://www.udemy.com/course/microsoft-power-bi-up-running-with-power-bi-desktop

**🔹 Libraries & Tools Documentation**

(1) Pandas Documentation – https://pandas.pydata.org/docs/

1. Matplotlib Documentation – https://matplotlib.org/stable/contents.html
2. DAX Guide – <https://dax.guide/>
3. Jupyter Notebook – <https://jupyter.org/>

**🔹 Other Sources**

1. IMDB Rating Structure – <https://www.imdb.com/interfaces/>
2. ChatGPT – For simulated dataset generation and content refinement
3. Kaggle Movie Datasets (Reference Only) – https://www.kaggle.com/datasets

**🔹 Acknowledged Support**

* Asian School of Media Studies – School of Data Science
* Faculty Mentorship and Peer Review
* Sample reports shared for academic formatting reference