**The National College Jayanagar (Autonomous)**

7th Block, Jayanagar, Bengaluru - 560070



Big Data Analytics project

Harshitha K R

(V Sem BCA Data science)

U03PA21S0112

Submitted to

Associate Professor

Ambarish Nath

Table of contents

1.ACKNOWLEDGEMENT

2.INTRODUCTION

3.ABSTRACT

4.WORKING

5.OBJECTIVE

6.SOURCE CODE

7.GRAPHS

8.OUTPUTS

9.CONCLUSION

10.REFERENCES

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to all those who contributed to the completion of this Big Data Analytics project on Netflix data analysis.

I extend my deepest gratitude to Associate Professor Ambarish Nath for his guidance, invaluable insights, and unwavering support throughout this project. Their expertise and encouragement significantly enhanced the quality and depth of this analysis.

Furthermore, I am grateful to The National College, Jayanagar for providing the opportunities essential for the successful execution of this project.

Lastly, I acknowledge the numerous individuals and resources, including scholarly works, articles, and databases, whose contributions formed the foundation of this analysis.

Thank you all for your invaluable contributions and support.

Sincerely,

Harshitha K R

INTRODUCTION

In the digital age of entertainment, few platforms have made as profound an impact as Netflix. With its extensive library of shows and movies, Netflix has not only revolutionized how we consume media but has also become a quintessential source of global entertainment. Understanding the intricate tapestry of content types, audience preferences, and the evolution of trends within Netflix's repository is not just insightful but pivotal in deciphering the platform's strategy and appeal.

The Netflix Data Analysis Project employing Tableau as its cornerstone presents an immersive journey into the depths of Netflix's vast reservoir of information. This project is designed to delve into the rich dataset encompassing show specifics, user ratings, production details, and more, allowing for a comprehensive exploration of Netflix's content landscape.

Tableau is used to create graphs for the data and for the further analysis of the data python is also used.

Title: Netflix Data Analysis Project Using Tableau

Abstract:

The Netflix Data Analysis Project using Tableau is a comprehensive exploration of one of the world's leading streaming platforms, Netflix, through the lens of data visualization and analysis. This project aims to uncover insights and patterns within Netflix's vast catalog and user behavior, utilizing Tableau, a powerful data visualization tool.

The dataset contains information on show identifiers, titles, content types, descriptions, release years, age certifications, runtimes, genres, production countries, season counts, IMDb and TMDb scores, as well as viewer ratings and votes. This dataset provides deep insights on Netflix shows.

OBJECTIVES

\*The dataset help understand the distribution of content types (movies or TV shows), genres, and their evolution over time. This provides an idea of how Netflix curates its content to a global and diverse audience.

\* The IMDb and TMDb scores, with user votes, we gain insights into how viewers rate Netflix shows.

\* The dataset includes information on the production countries of shows.

Tableau, is a primary data visualization tool, used to create interactive and insightful dashboards. These dashboards provide a user-friendly interface for stakeholders, to explore the data and gain actionable insights.

Python is also used to analyse the data using jupyter notebook.

WORKING

The Netflix Data Analysis Project revolves around examining the provided dataset to gain valuable insights into the content and user behavior on the Netflix platform. Let's outline a theoretical workflow for this project:

1. Dataset Understanding and Preprocessing

Objective:

Understand the structure, content, and quality of the dataset to prepare it for analysis.

Data Loading: Import the dataset (CSV, Excel, etc.) into a Pandas DataFrame in Python.

Initial Exploration: Review basic statistics, column names, data types, and examine a sample of the data to understand its structure.

Data Cleaning: Handle missing values, inconsistencies, duplicates, and format inconsistencies.

2. Exploratory Data Analysis (EDA)

Objectives:

Uncover patterns, distributions, and correlations within the data.

Content Distribution: Explore the distribution of movies versus TV shows, genres, release years, etc.

User Engagement: Analyze IMDb and TMDb scores,runtime.

Geographical Insights: Investigate production countries and their contribution to Netflix content.

3. Data Visualization using Tableau

Objective:

Create interactive and informative visualizations for stakeholders.

Connect Dataset to Tableau: Import the preprocessed dataset into Tableau.

Create Visualizations: Design various charts, graphs, and dashboards showcasing key insights like trends over time, audience preferences, etc.

OBJECTIVE

The objective of your Netflix Data Analysis Project is to extract meaningful insights and patterns from the provided dataset about Netflix content.

1.Understanding Content Distribution

Content Types: Analyze the distribution of movies and TV shows to understand the ratio and evolution of content over time.

Genres: Explore the most prevalent genres and their popularity trends.

2.User Engagement and Preferences

Ratings and Scores: Investigate IMDb and TMDb scores along with user votes to gauge audience reception and preferences.

3. Geographical and Production Insights

Production Countries: Examine the contribution of different countries to Netflix's content library.

SOURCE CODE

import numpy as np

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

df = pd.read\_excel(r'C:\Users\Handi\Desktop\ML and BDA notes\Tableau\titles.xlsx')

df.head()

df.info()

#checking for duplicate values

dup\_df = df.duplicated().any()

print("Any duplicate values?", dup\_df)

#cleaning missing values, by removing them.

print("Any missing value?", df.isnull().values.any())

df.dropna(axis = 0, inplace= True)

print()

print(df.isnull().sum())

mask\_movies = df.type == "MOVIE"

df\_movies= df[mask\_movies]

mask\_shows = df.type == "SHOW"

df\_shows= df[mask\_shows]

corr\_movies = df\_movies.corr()

print(corr\_movies)

movie\_pop = df\_movies.groupby("genres").agg({"imdb\_votes":"mean", "imdb\_votes":"count"})

movie\_pop.sort\_values("imdb\_votes", ascending = False)

print(movie\_pop)

df\_movies.describe()

df\_shows.describe()

# Distribution of content types (movies vs TV shows)

content\_distribution = df['type'].value\_counts()

print(content\_distribution)

# Explore genres

genres\_count = df['genres'].value\_counts().head(10)

print(genres\_count)

# Analyze IMDb and TMDb scores

imdb\_scores = df['imdb\_score']

tmdb\_scores = df['tmdb\_score']

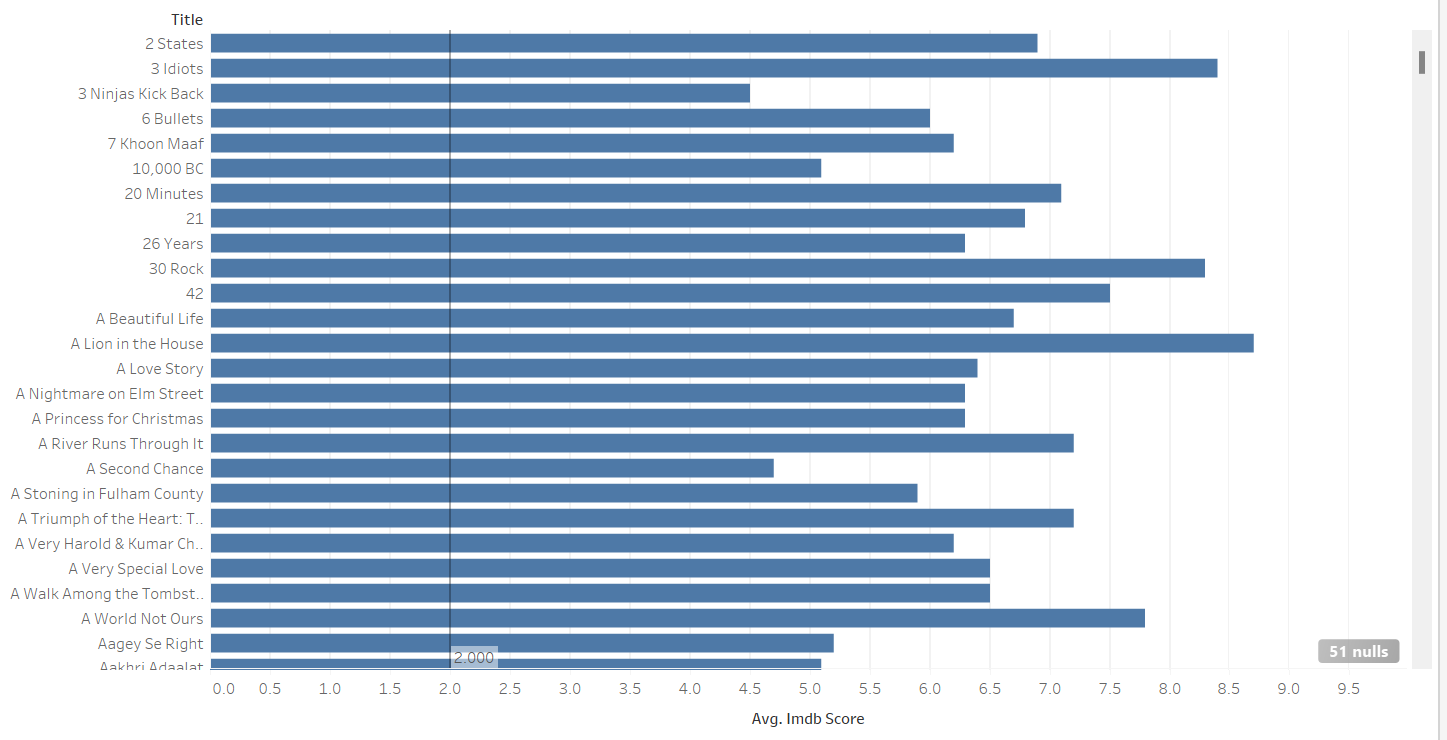
# Distribution of runtime

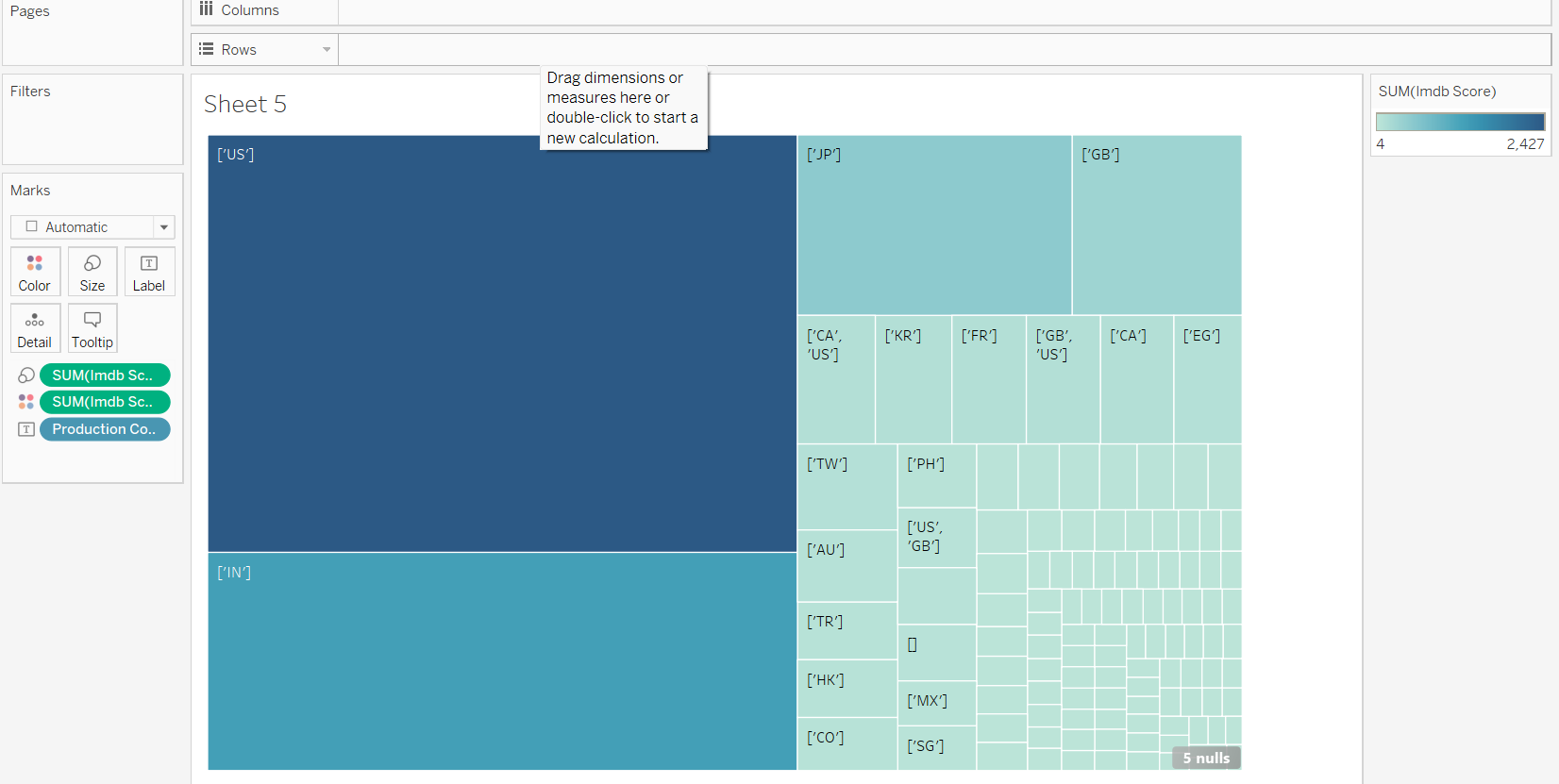
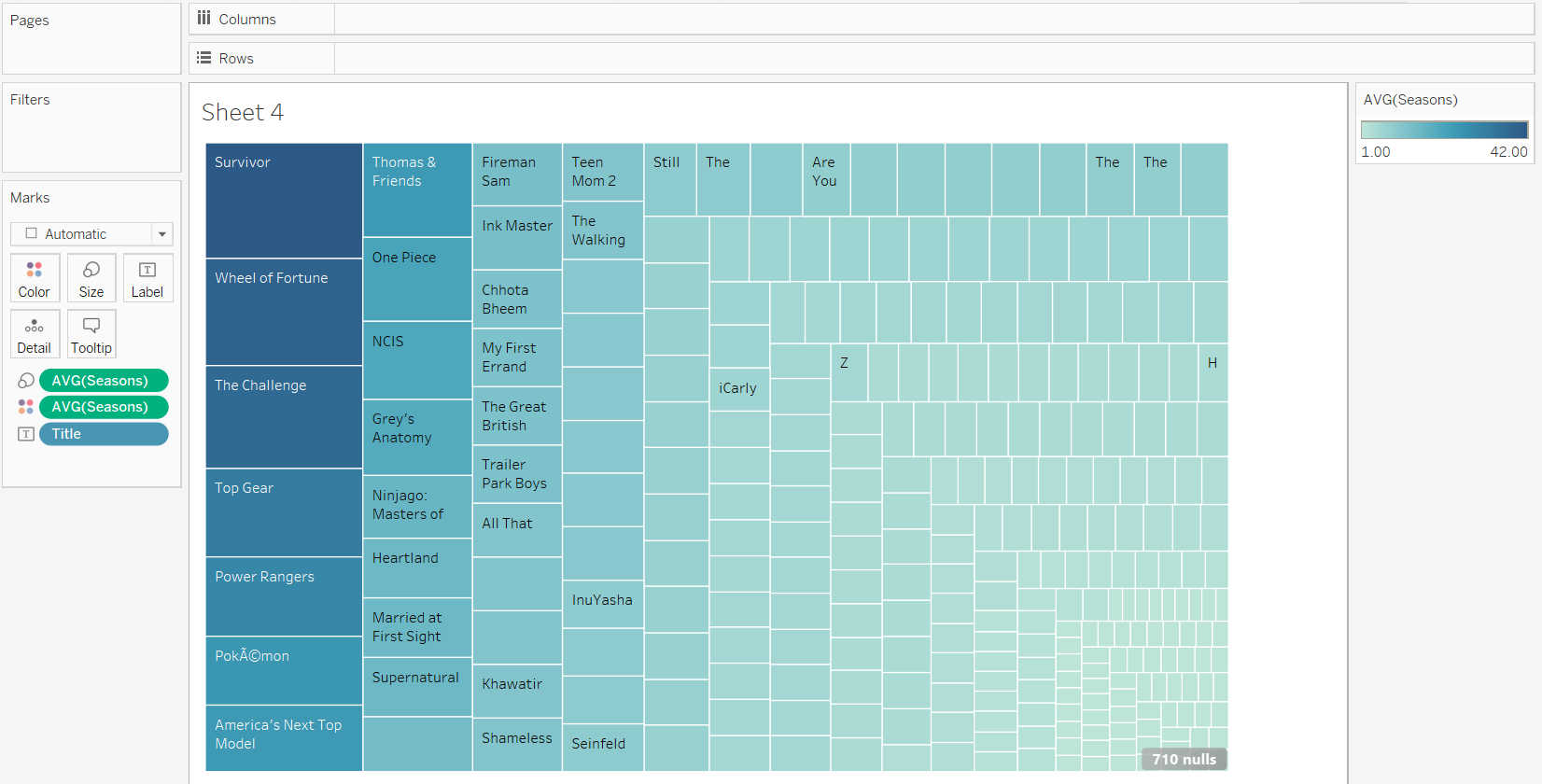
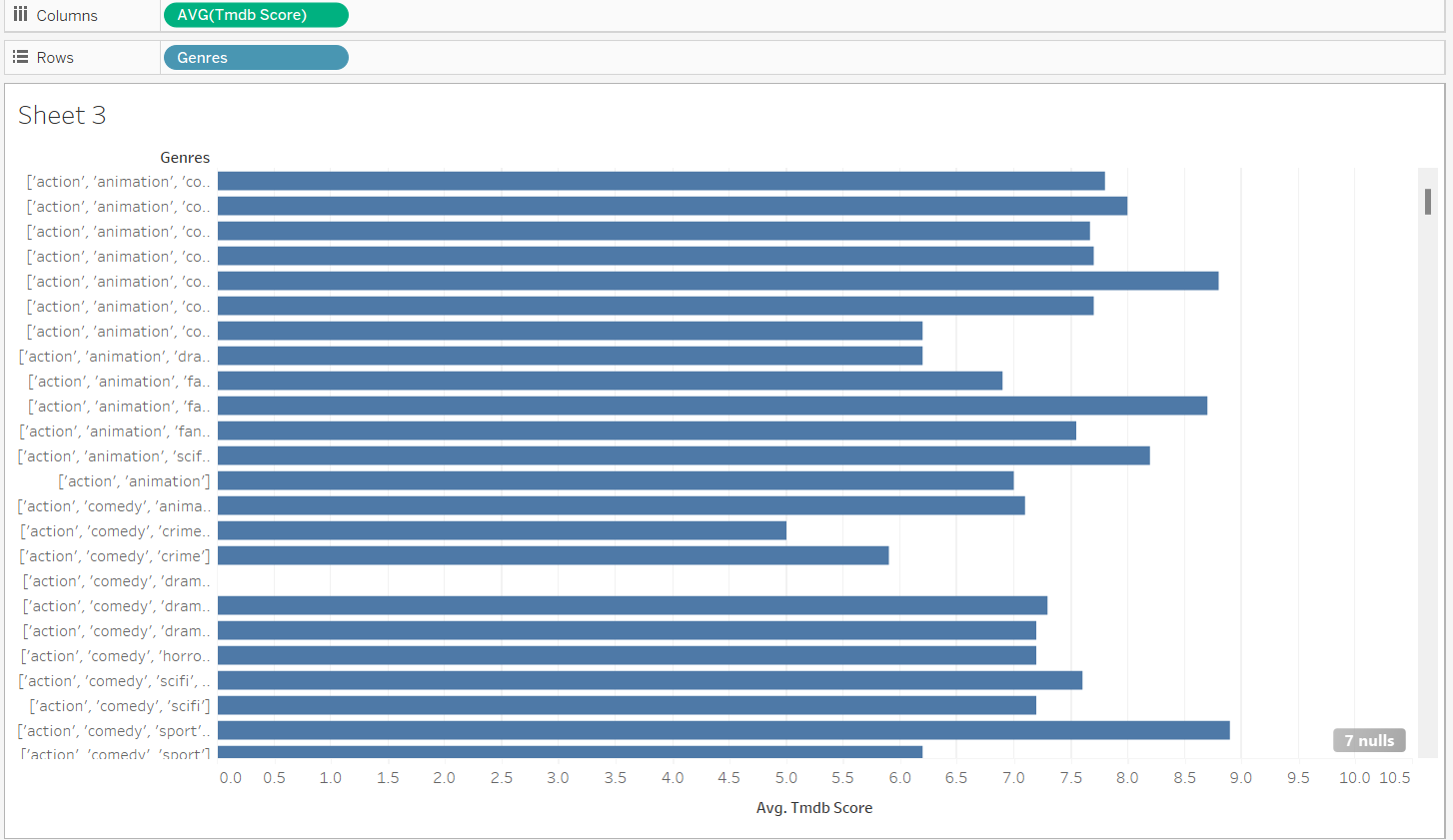
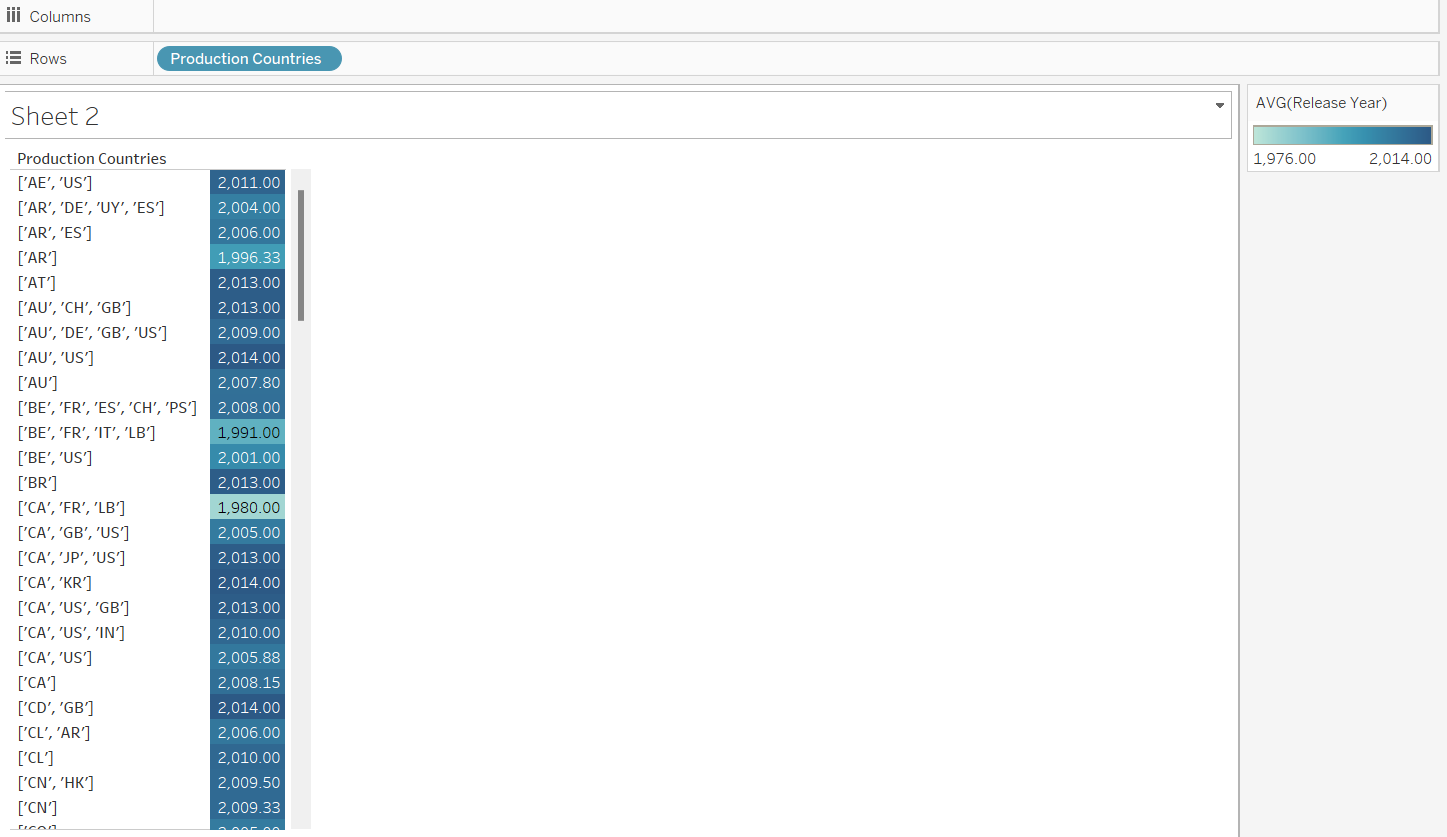
runtime\_distribution = df['runtime'].value\_counts()

# Explore production countries

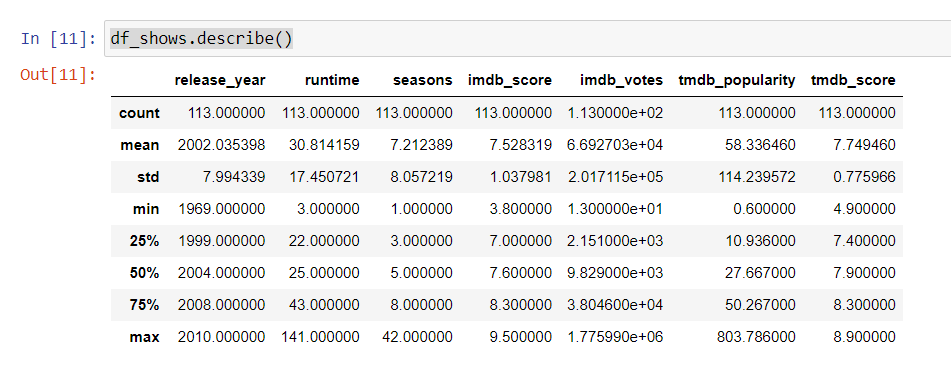
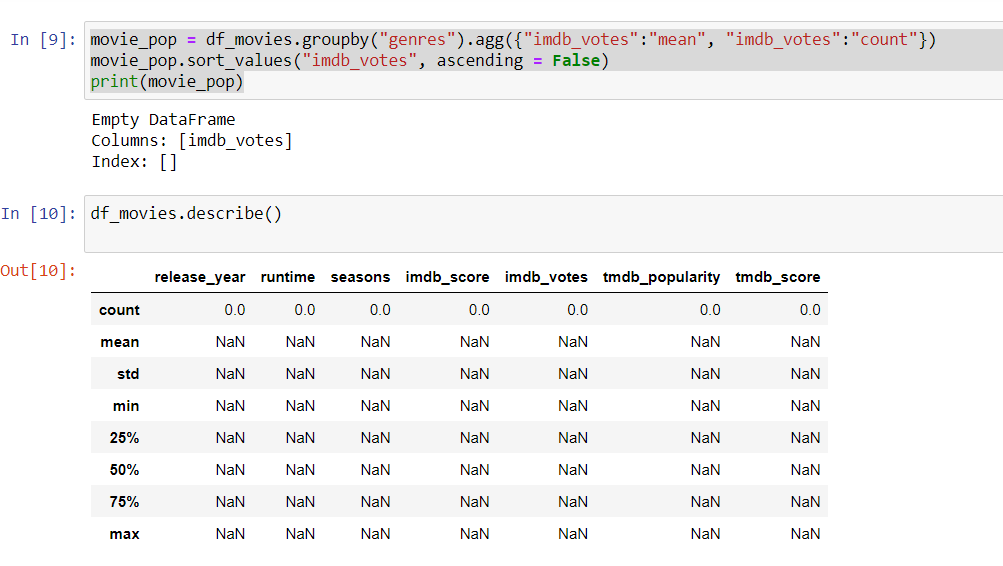
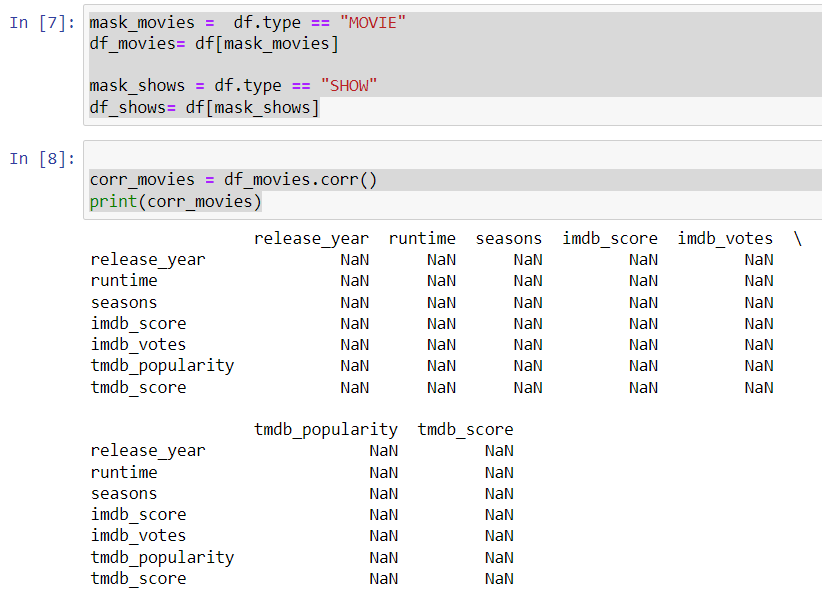
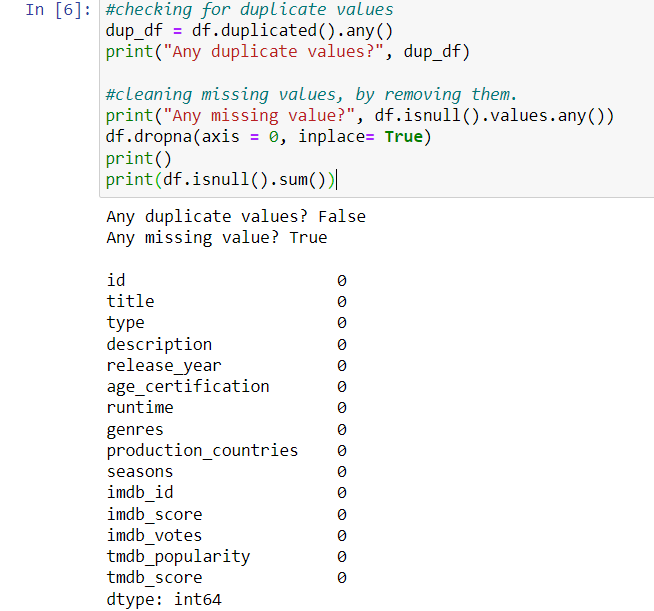
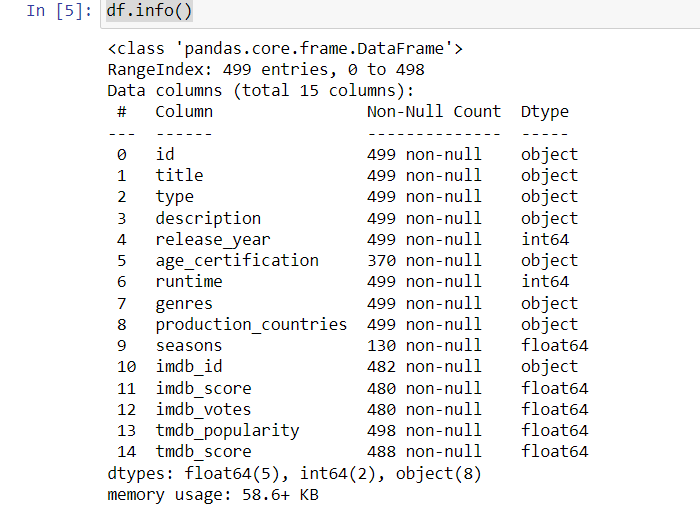
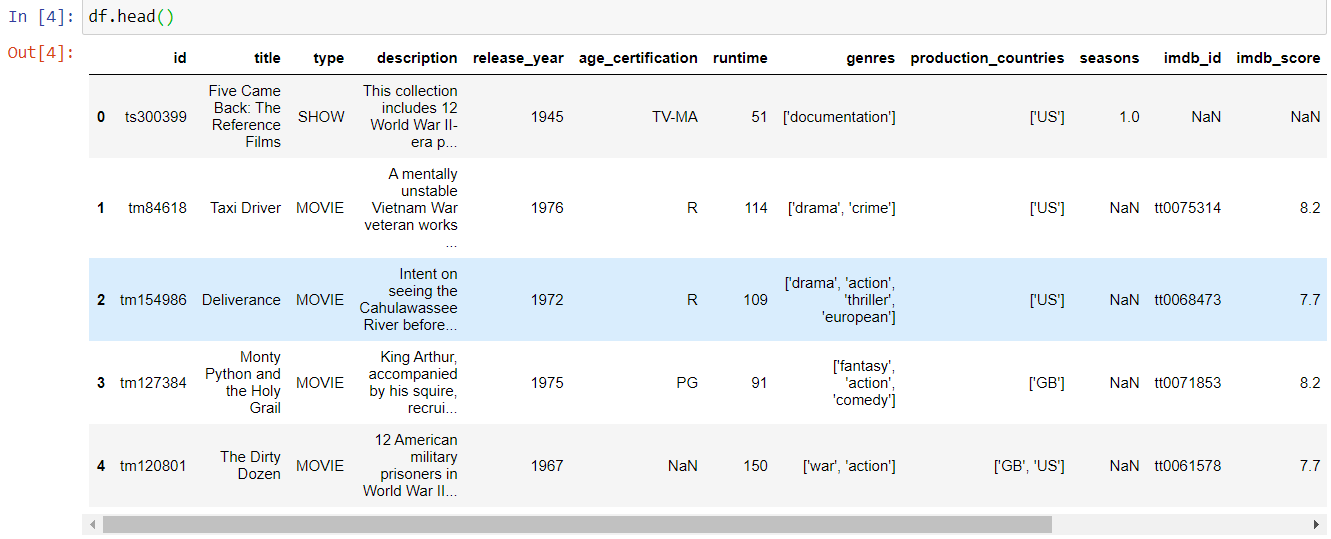
production\_countries\_count = df['production\_countries'].value\_counts().head(10)

print(production\_countries\_count)

GRAPHS



OUTPUTS



CONCLUSION

The Netflix Data Analysis Project has been a captivating exploration into the intricate landscape of one of the world's leading streaming platforms. Through meticulous data analysis and visualization, this project has unveiled a tapestry of insights and trends that offer a profound understanding of Netflix's content ecosystem and user engagement dynamics.

Content Insights:

The examination of content distribution between movies and TV shows revealed a shifting trend in Netflix's offerings over time. Understanding the pulse of popular genres showcased the dynamic preferences of the audience, with certain genres reigning supreme while others emerged as hidden gems.

User Engagement and Preferences:

The analysis of IMDb and TMDb scores, coupled with viewer ratings and votes, provided a holistic view of how audiences perceive and engage with Netflix content. These metrics not only reflected audience preferences but also hinted at the correlation between ratings and viewership.

Geographical and Production Nuances:

Insights into the contribution of different production countries shed light on the global nature of Netflix's content catalog. Observing regional trends highlighted unique preferences across diverse audiences, emphasizing the importance of localized content strategies.

Visualization

The integration of Tableau facilitated the creation of interactive graphs that served as a gateway to a wealth of insights.

REFERENCES

[Netflix\_Data\_Analysis | Kaggle](https://www.kaggle.com/code/chukwukaokeke/netflix-data-analysis)

[Netflix TV Shows and Movies (kaggle.com)](https://www.kaggle.com/datasets/victorsoeiro/netflix-tv-shows-and-movies/data)

[Netflix Movies Dataset (kaggle.com)](https://www.kaggle.com/datasets/muhammadanasmahmood/netflix-movies-dataset)