📊 Netflix Data Cleaning, Analysis & Visualization Project Report

📌 Title Page

Project Title: Netflix Data Cleaning, Analysis & Visualization
Name:Neha

1. Abstract

This project focuses on **analyzing Netflix content data** to uncover insights into content trends, genre distribution, ratings, and global reach. Using **Python for data cleaning & analysis** and **Tableau for interactive visualizations**, we transform raw data into an insightful dashboard. The study identifies key trends in content types, country-wise distributions, and time-based additions to Netflix. This project serves as an analytical foundation for content recommendation, marketing strategies, and future research in streaming trends.

2. Introduction

2.1 Purpose of the Project

Netflix is one of the largest streaming platforms worldwide, hosting diverse content. However, understanding **content distribution patterns** requires data-driven insights. This project aims to analyze Netflix data to:

- Identify **trends in Movies vs TV Shows**
- Track **content additions over time**
- Examine **genre popularity**
- Study **content distribution by country**
- Create an **interactive Tableau dashboard** for easy exploration

2.2 Significance of the Study

Understanding Netflix's content trends is valuable for:

- **Data analysts & researchers** studying streaming platforms
- **Content creators & producers** for strategic decision-making
- **Marketing teams** to understand viewer preferences

2.3 Objectives

- Clean and preprocess the Netflix dataset
- Perform **exploratory data analysis (EDA)**

- Create **interactive visualizations using Tableau**
- Provide key insights into **content trends & distribution**

3. Dataset Description

3.1 Source

- **Dataset Name**: Netflix Movies and TV Shows
- **Source**: Kaggle (link)
- **No. of Entries**: 8,790 rows
- **No. of Columns**: 10

3.2 Key Columns

Column Name	Description
`show_id`	Unique ID for each title
`type`	Movie or TV Show
`title`	Name of the title
`director`	Director's name
`country`	Country where the content was produced
`date_added`	Date when it was added to Netflix
`release_year`	Year of release
`rating`	Content rating (e.g., PG-13, TV-MA)
`duration`	Duration in minutes/seasons
`listed_in`	Genres/categories of the title

4. Methodology

4.1 Tools & Technologies Used

- **Python (Pandas, NumPy, Matplotlib, Seaborn)** – Data cleaning & EDA
- **Tableau** – Interactive visualization dashboard
- **GitHub** – Version control & documentation

4.2 Data Cleaning Process

- Handled missing values (director, country, rating columns)
- Converted date_added to **datetime format**
- Extracted **year, month, and day** for analysis
- Removed duplicates & standardized data types

4.3 Exploratory Data Analysis (EDA)

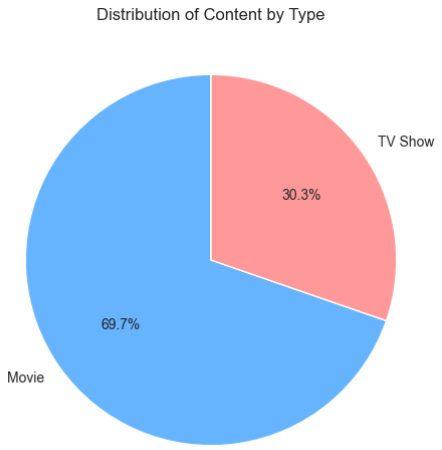
Performed **statistical analysis & visualization** on:

- **Movies vs TV Shows distribution**
- **Yearly content trends**
- **Top genres & ratings**
- **Country-wise Netflix content**

5. Exploratory Data Analysis (EDA) Results

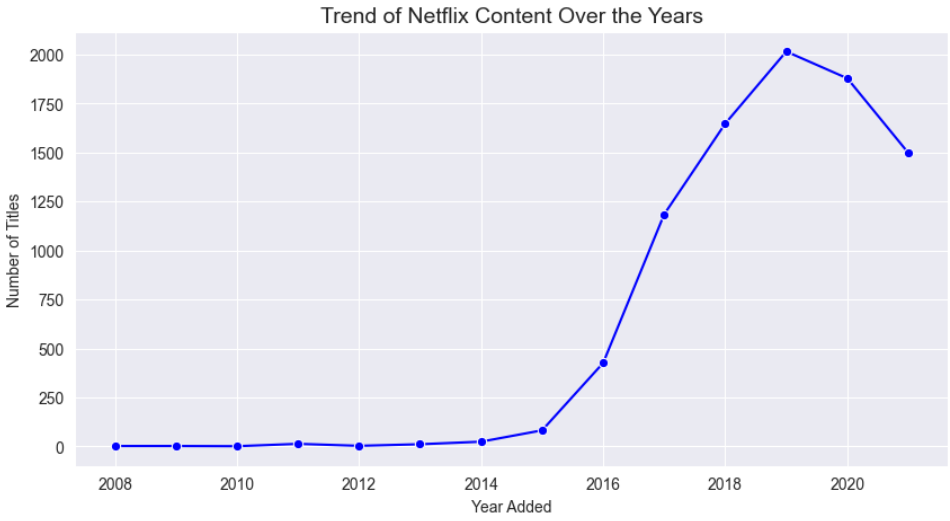
5.1 Movies vs TV Shows

- **Movies (70%) dominate over TV Shows (30%)**
- Netflix adds **more Movies per year** than TV Shows



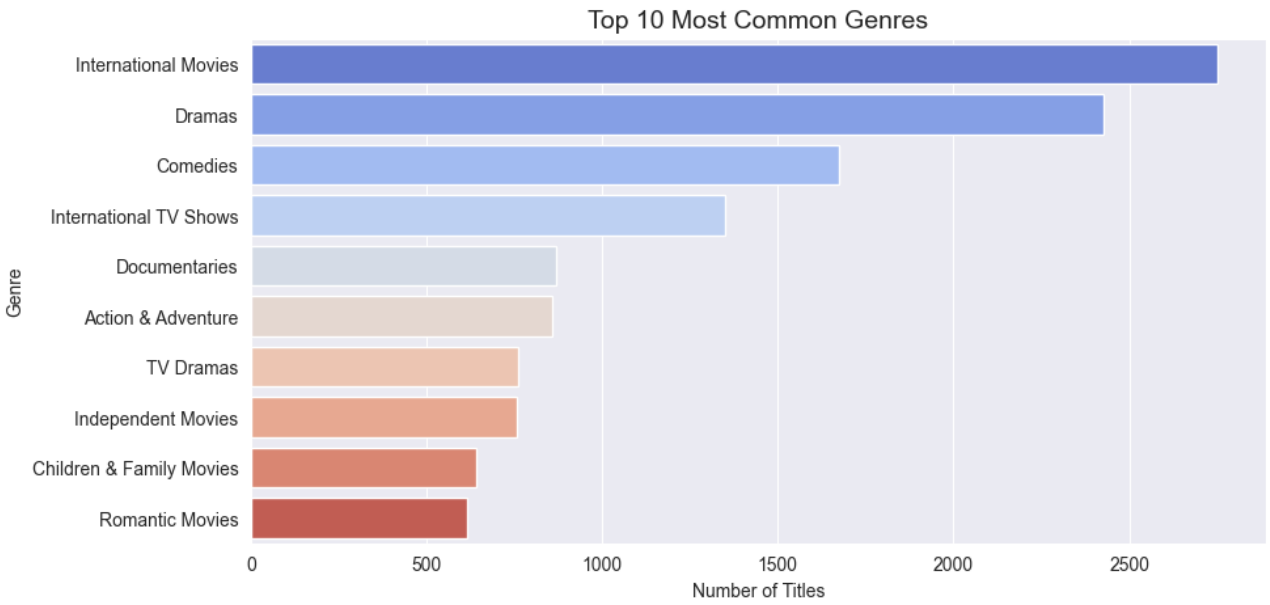
5.2 Yearly Content Trends

- **Peak additions** occurred in **2019-2020**
- TV Shows **grew steadily**, while Movie additions fluctuated



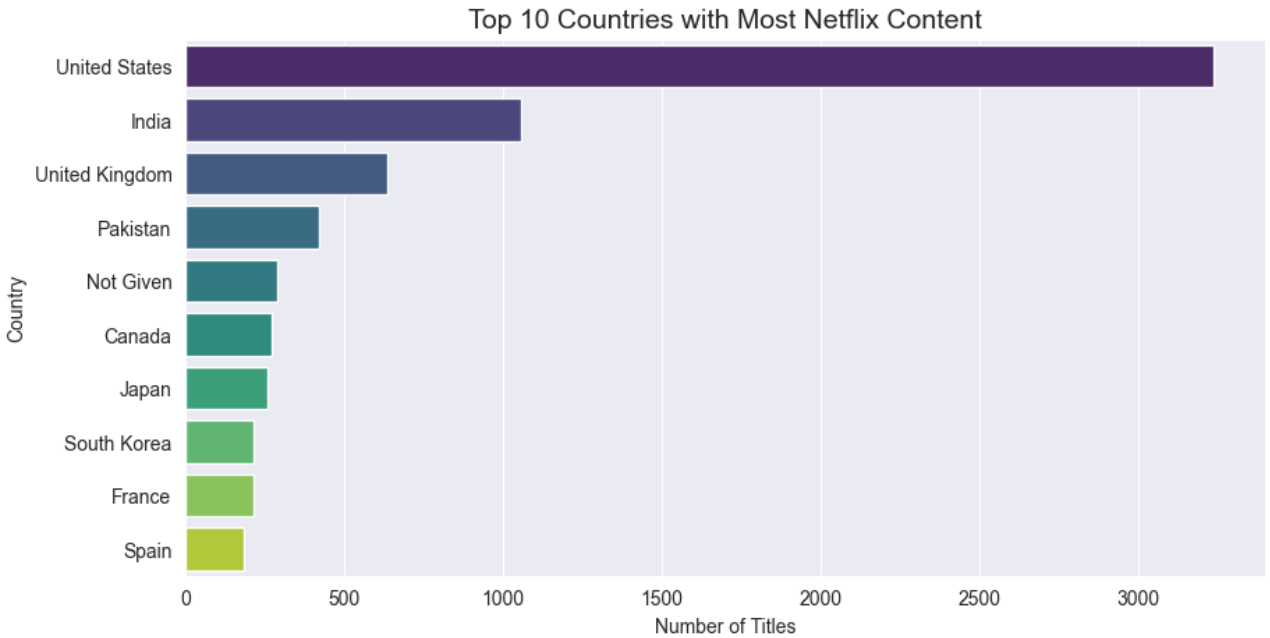
5.3 Genre Analysis

- **Top Genres:** International Movies, Dramas, Comedies
- Content is **diverse**, with multiple genres overlapping



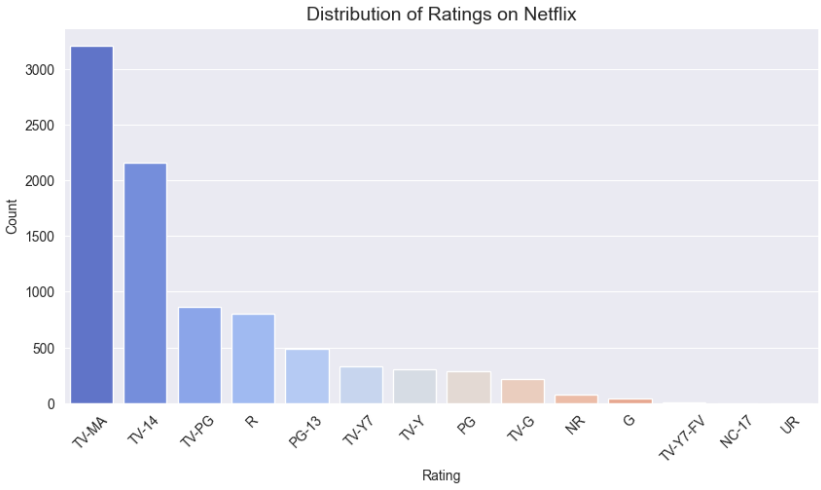
5.4 Content Distribution by Country

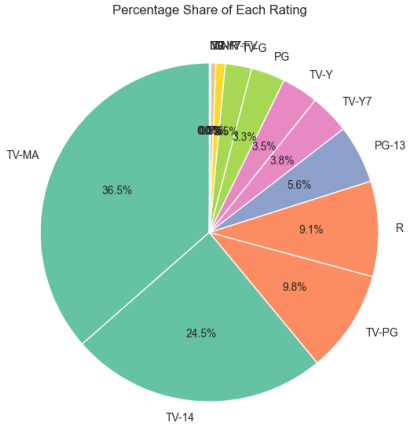
- **Top Producing Countries:** United States, India, United Kingdom
- **Emerging Markets:** South Korea, Japan, Canada



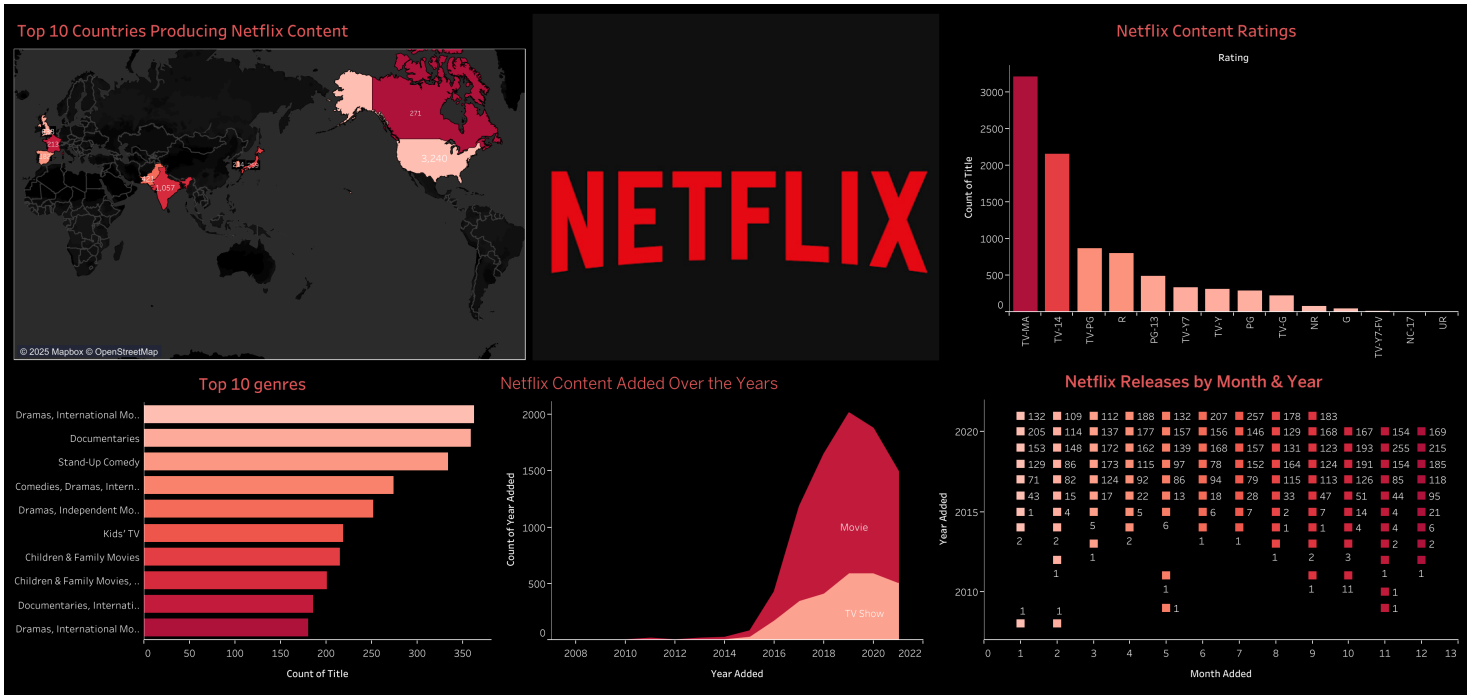
5.5 Ratings Distribution

- **TV-MA** (Mature Audience) is the most common rating
- **PG & PG-13 content** is lower compared to adult content





6. Tableau Dashboard Overview



6.1 Dashboard Elements

The interactive **Netflix Tableau Dashboard** includes:

- **Area Chart** – Movies vs TV Shows
- **Horizontal Bar Chart** – Yearly Content Trends
- **Gantt Chart** – Monthly Content Releases
- **Bar Chart** – Ratings Distribution
- **Geographical Map** – Country-wise Content

6.2 Key Insights from the Dashboard

- Netflix **focuses more on Movies**
- **2019-2020 saw the highest content additions**
- **International content (non-US) is growing**

7. Challenges & Solutions

Challenge	Solution
Missing values in `director`, `country`, `rating`	Filled with "Unknown" or used mode value
Dataset contained inconsistent date formats	Converted `date_added` to proper datetime format
Large dataset size for visualization	Used optimized `.csv` file format for Tableau
Tableau map showing incorrect country labels	Standardized country names before visualization

8. Conclusion & Future Scope

8.1 Conclusion

- Netflix's **library is growing**, with **Movies dominating**
- **Genres like Drama & Comedy are most common**
- **Peak content addition years: 2019-2020**
- **Global expansion** is evident from diverse country origins

8.2 Future Scope

- **Predictive Modeling:** Can we predict the next popular genre?
- **Sentiment Analysis:** Analyzing viewer ratings/reviews
- **Streamlit App:** Creating an interactive web-based dashboard

9. References

1. Netflix Dataset – Kaggle (link)
2. Python Libraries: Pandas, NumPy, Matplotlib, Seaborn
3. Tableau Public – Used for Interactive Dashboard