📊 Netflix Data Visualization Project

This project performs exploratory data analysis and visualization on Netflix's content dataset using Python. It aims to reveal patterns in content types, release years, countries, ratings, and durations.

# 📁 Dataset

- Source: Kaggle - Netflix Shows Dataset (https://www.kaggle.com/datasets/shivamb/netflix-shows)  
- Data Includes: Title, Type, Country, Date Added, Release Year, Duration, Rating, Genre

# 🧰 Tools & Technologies

- Python 🐍  
- Pandas  
- Matplotlib  
- Jupyter Notebook

# data clean

df = df.dropna(subset=['type','release\_year','rating','country','duration'])

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

plt.figure(figsize=(6,4))

plt.bar(type\_counts.index,type\_counts.values, color=['skyblue','orange'])

plt.title('Number of Movies Vs Tv Shows on Netflix')

plt.xlabel('Type')

plt.ylabel('Count')

plt.tight\_layout()

plt.savefig('Moives\_vs\_tvshows.png')

plt.show()

rating\_counts= df['rating'].value\_counts()

plt.figure(figsize=(8,6))

plt.pie(rating\_counts, labels=rating\_counts.index,

autopct='%1.1f%%', startangle=90)

plt.title('Percentage of Content Ratings')

plt.tight\_layout()

plt.savefig('content\_Ratings\_pie.png')

plt.show()

movie\_df = df[df['type'] == 'Movie'].copy()

movie\_df['duration\_int'] = movie\_df['duration'].str.replace('min','').astype(int)

plt.figure(figsize=(8,6))

plt.hist(movie\_df['duration\_int'], bins=30, color='purple',edgecolor='black')

plt.title("Distributaion of Movie Duration")

plt.xlabel('Duration (minutes)')

plt.ylabel('Number of Movies')

plt.tight\_layout()

plt.savefig('movie\_duration\_histogram.png')

plt.show()

release\_counts = df['release\_year'].value\_counts().sort\_index()

plt.figure(figsize=(10,6))

plt.scatter(release\_counts.index, release\_counts.values, color='red')

plt.title("Release Year Vs Number of Shows")

plt.xlabel('Release Year')

plt.ylabel('Number of Shows')

plt.tight\_layout()

plt.savefig('release\_year\_Scatter.png')

plt.show()

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

plt.figure(figsize=(8,6))

plt.barh(country\_counts.index, country\_counts.values, color='teal')

plt.title('Top 10 Countries by Number of Shows')

plt.xlabel('Number of Shows')

plt.ylabel('Country')

plt.tight\_layout()

plt.savefig('Top10\_countries.png')

plt.show()

content\_by\_year = df.groupby(['release\_year','type']).size().unstack().fillna(0)

fig, ax = plt.subplots(1,2, figsize=(12,5))

#first Subplot:movies

ax[0].plot(content\_by\_year.index, content\_by\_year['Movie'], color='blue')

ax[0].set\_title('Movies Released Per year')

ax[0].set\_xlabel('Year')

ax[0].set\_ylabel('Number of Movies')

#Second Subplot:movies

ax[0].plot(content\_by\_year.index, content\_by\_year['TV Show'], color='orange')

ax[0].set\_title('TV Shows Released Per year')

ax[0].set\_xlabel('Year')

ax[0].set\_ylabel('Number of Movies')

fig.suptitle('Comparison of Movies and TV shows Released Over Years')

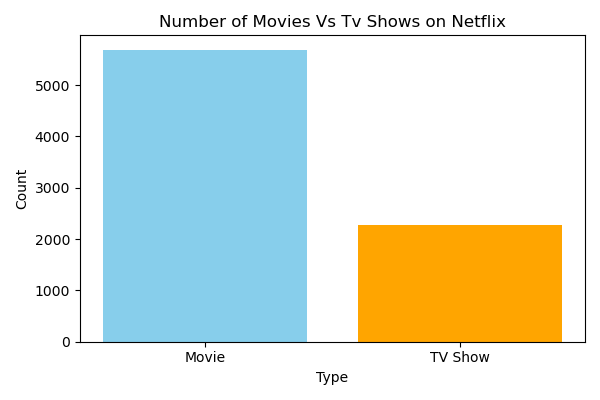
plt.tight\_layout()

plt.savefig('movies\_tv\_shows\_comparison.png')

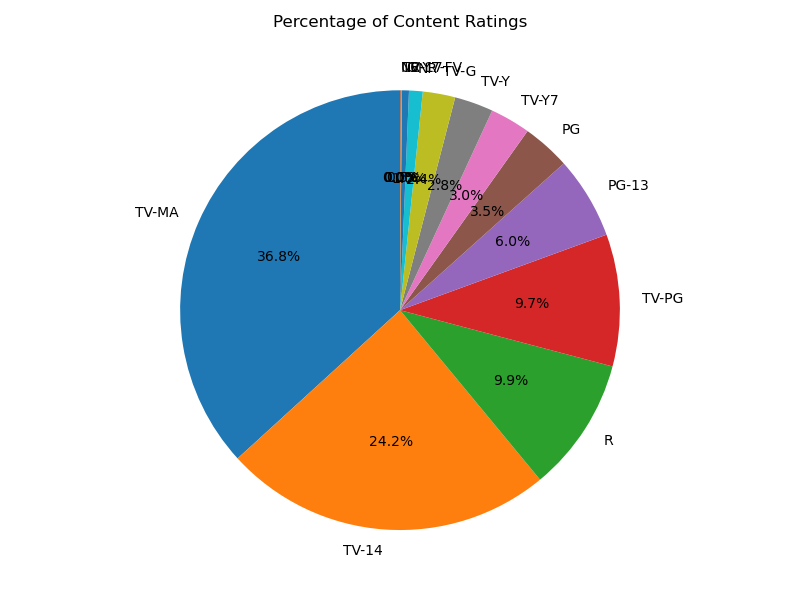
plt.show()

# 📌 Key Visualizations

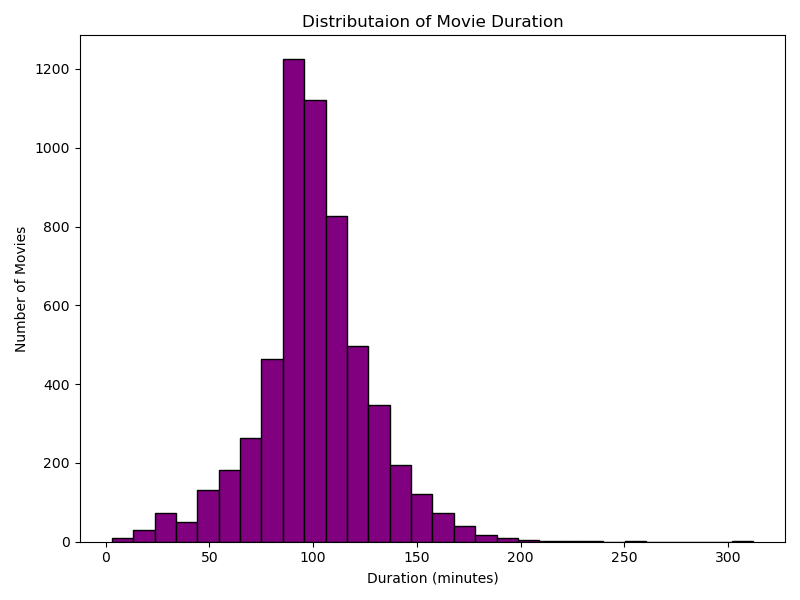
## 📉 1. Number of Movies vs TV Shows



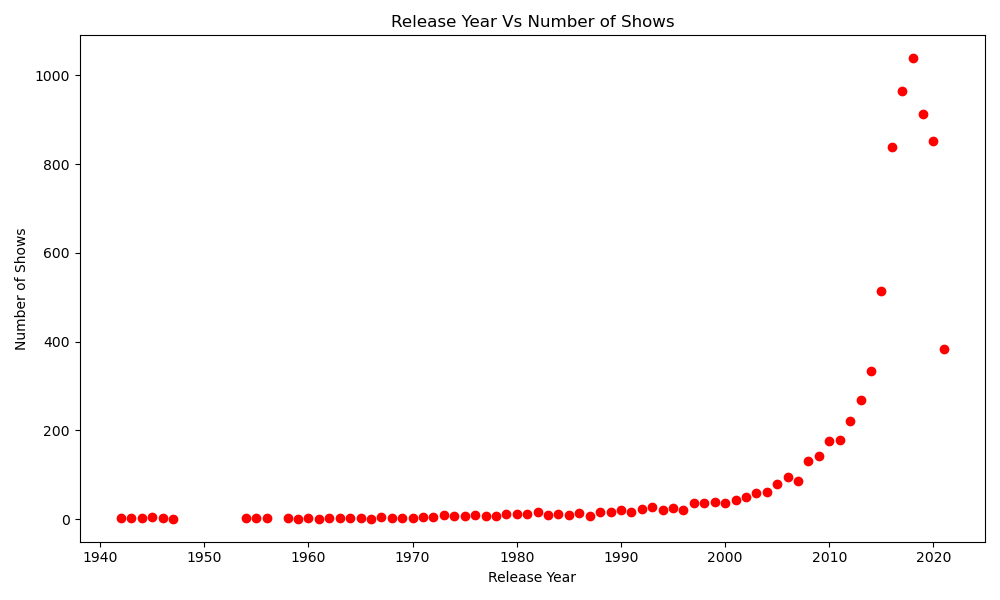
## 🧁 2. Percentage of Content Ratings



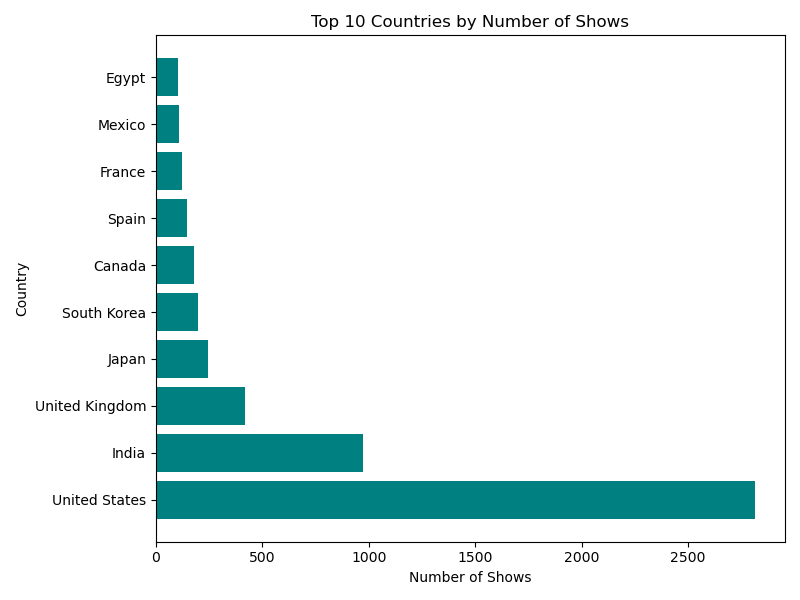
## ⏱ 3. Distribution of Movie Durations



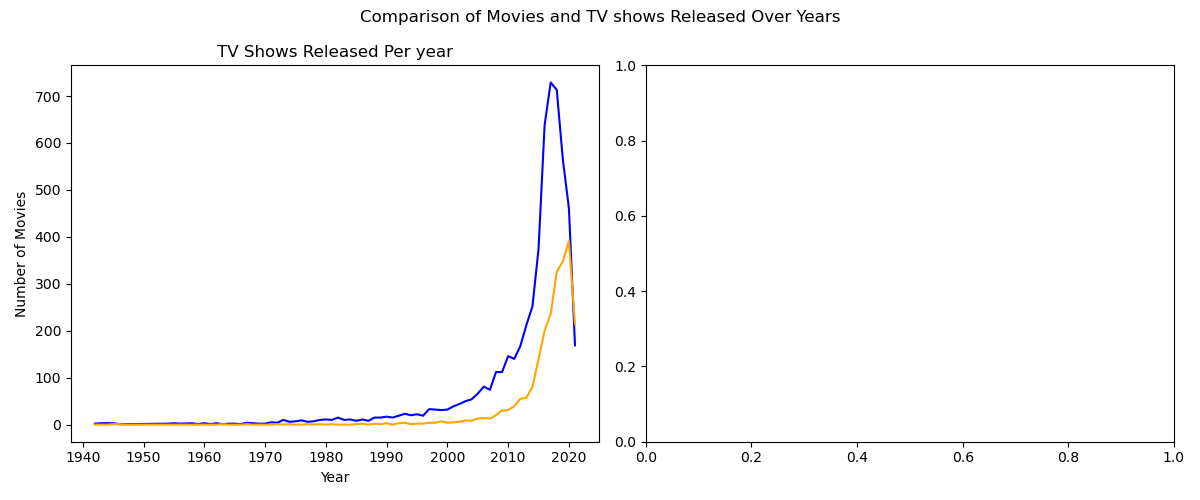
## 📆 4. Release Year vs Number of Shows



## 🌍 5. Top 10 Countries by Number of Shows



## 🔄 6. Comparison of Movies and TV Shows Over the Years



# 🔍 Key Insights

- TV-MA and TV-14 are the most frequent ratings.  
- Netflix has more Movies than TV Shows.  
- The platform’s content library grew massively after 2015.  
- USA and India are the leading producers of Netflix content.  
- Movie durations are mostly between 80–120 minutes.

# 🧠 What You'll Learn

- Real-world data cleaning using dropna, .str.replace(), .groupby(), etc.  
- Creating different types of visualizations using Matplotlib and Seaborn.  
- Drawing business insights from data.  
- Presenting a clean, visual story through charts.

# ▶️ How to Run

1. Clone the repository:  
 git clone https://github.com/your-username/netflix-data-visualisation.git  
  
2. Navigate to the folder:  
 cd netflix-data-visualisation  
  
3. Launch Jupyter Notebook:  
 jupyter notebook Netflix\_Data\_visualisation.ipynb  
  
4. Run all cells to view the analysis, visuals, and insights.

# 🙋‍♂️ Author

Jaideep Patil  
📧 jaideeppatil100@gmail.com  
🔗 LinkedIn: https://www.linkedin.com/in/jaideep-patil-42797a25a  
💻 GitHub: https://github.com/Jaideep0710

# 🙌 Acknowledgements

- Dataset: Netflix on Kaggle   
- Libraries: Python, Pandas, Matplotlib