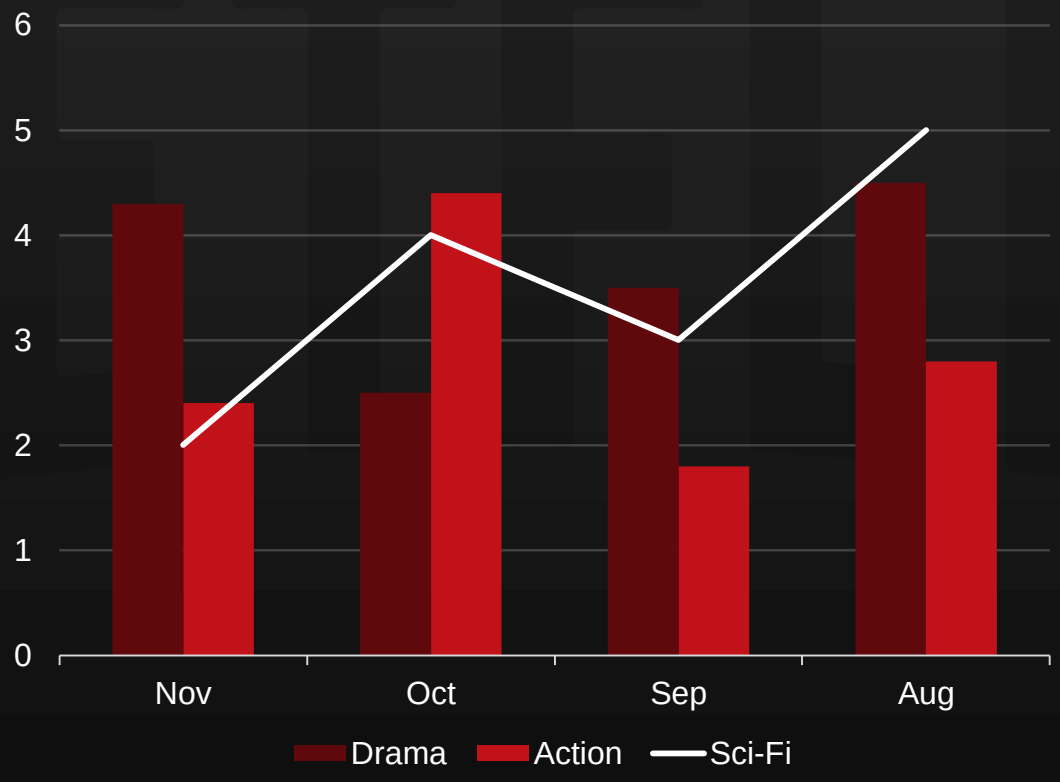
NETFLIX

NETFLIX Data Analysis Using Python & Power BI



Presented by: Bruna Augustin

NETFLIX

Introduction

Netflix is a global **OTT streaming giant** that thrives on **data-driven decisions** to enhance user experience and business growth.

Why Data Analysis?

- Identifying popular content trends (Movies vs. TV Show)
- Understanding year-wise content growth for better production planning.
- Analyzing genre & country-wise distribution to optimize regional content strategies.

Project Focus

- Data Cleaning & Preprocessing (Removing inconsistencies, handling missing values).
- Movies vs. TV Shows Analysis (Comparing content distribution).
- Year-wise Trend Analysis (Observing content growth patterns over the years).
- Country-wise Content Distribution (Identifying regional content availability).
- Genre Analysis (Finding the most common content types).
- Power BI Dashboard Creation (Interactive visualizations for better insights).

NETFLIX

Dataset Overview

Dataset Name – Netflix
Source – Unified Mentor (Internship Project)
Total Records – 8,790
Columns Included – show_id, type, title, director, country, date_added, release_year, rating, duration, listed_in
Purpose – Analysing content trends, distribution, and user preferences

"This dataset, provided by Unified Mentor as part of the internship project, contains 8,790 records of movies and TV shows available on Netflix. It includes key details like genre, country of origin, release year, and type. The analysis focuses on understanding content trends, distribution, and audience preferences using Python and Power BI."

NETFLIX

Data Cleaning & Preprocessing with Python

Raw Dataset Preview (Original Data)

```python
#Importing library and Loading the Netflix dataset from a CSV file
import pandas as pd
netflix = pd.read_csv("netflix1.csv")
netflix.head()
```

Output

	show_id	type	title	director	country	date_added	release_year	rating	duration	listed_in
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	United States	9/25/2021	2020	PG-13	90 min	Documentaries
1	s3	TV Show	Ganglands	Julien Leclercq	France	9/24/2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...
2	s6	TV Show	Midnight Mass	Mike Flanagan	United States	9/24/2021	2021	TV-MA	1 Season	TV Dramas, TV Horror, TV Mysteries
3	s14	Movie	Confessions of an Invisible Girl	Bruno Garotti	Brazil	9/22/2021	2021	TV-PG	91 min	Children & Family Movies, Comedies
4	s8	Movie	Sankofa	Haile Gerima	United States	9/24/2021	1993	TV-MA	125 min	Dramas, Independent Movies, International Movies

"This is the original dataset before any cleaning. It contains Netflix content details such as title, genre, release year, and country."

NETFLIX

Checking Data Types & Structure

```
#Shows column names, non-null counts, and data types
netflix.info()
```

Output

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8790 entries, 0 to 8789
Data columns (total 10 columns):
 #   Column        Non-Null Count  Dtype 
---  ------        --------------  ----- 
 0   show_id       8790 non-null   object
 1   type          8790 non-null   object
 2   title         8790 non-null   object
 3   director      8790 non-null   object
 4   country       8790 non-null   object
 5   date_added    8790 non-null   object
 6   release_year  8790 non-null   int64 
 7   rating        8790 non-null   object
 8   duration      8790 non-null   object
 9   listed_in     8790 non-null   object
dtypes: int64(1), object(9)
memory usage: 686.8+ KB
```

"The dataset contains structured information with no missing values in column types."

NETFLIX

Checking Missing Values

```
#check missing values
netflix.isnull().sum()
```

Output

```
show_id         0
type            0
title           0
director        0
country         0
date_added      0
release_year    0
rating          0
duration        0
listed_in       0
dtype: int64
```

"No missing values were found in the dataset, ensuring data completeness."

NETFLIX

Checking Duplicates Values

```
#check duplicate data
netflix.duplicated().sum()
```

Output

```
np.int64(0)
```

"No duplicate records were found, confirming data uniqueness."

NETFLIX

Data Visualization with Python

"After verifying data quality, we proceeded with exploratory data analysis (EDA) using Python. Visualization helps in uncovering patterns, trends, and insights from the dataset. Using Matplotlib and Seaborn, we created multiple charts to analyze Netflix's content distribution, genre popularity, and trends over time."

"Exploratory Data Analysis (EDA) in Python“

"Unveiling Insights with Data Visualization“

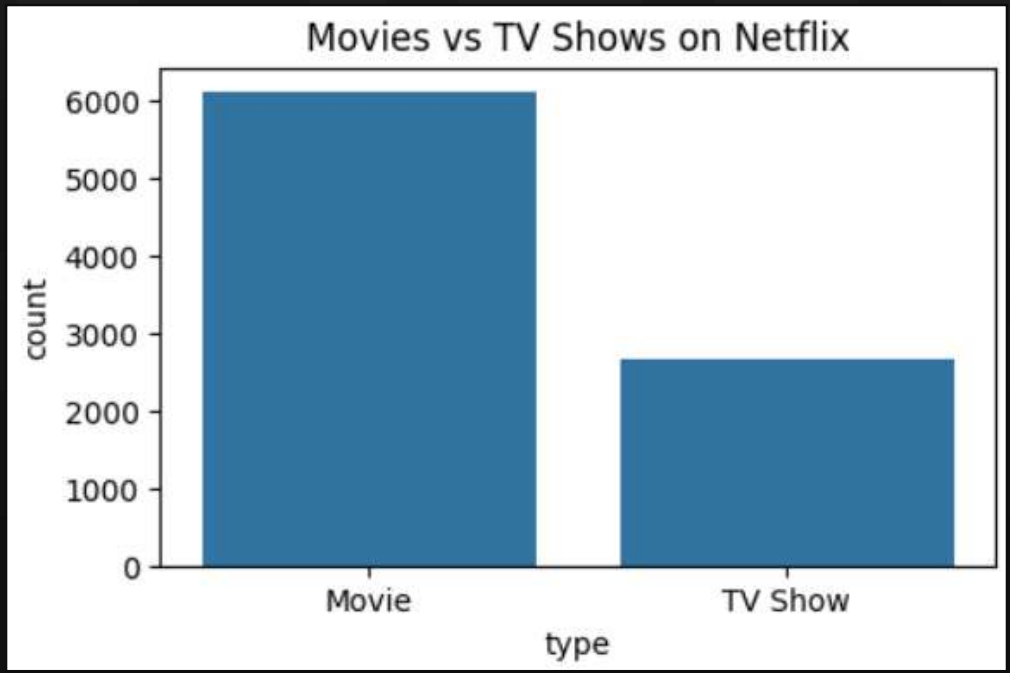
"Visualizing Netflix Data Trends"

NETFLIX

Visualizations (Matplotlib & Seaborn)

"Distribution of Movies & TV Shows on Netflix"

```
#Movies vs TV Shows Count
plt.figure(figsize=(6,4))
sns.countplot(x='type', data=netflix)
plt.title("Movies vs TV Shows on Netflix")
plt.show()
```

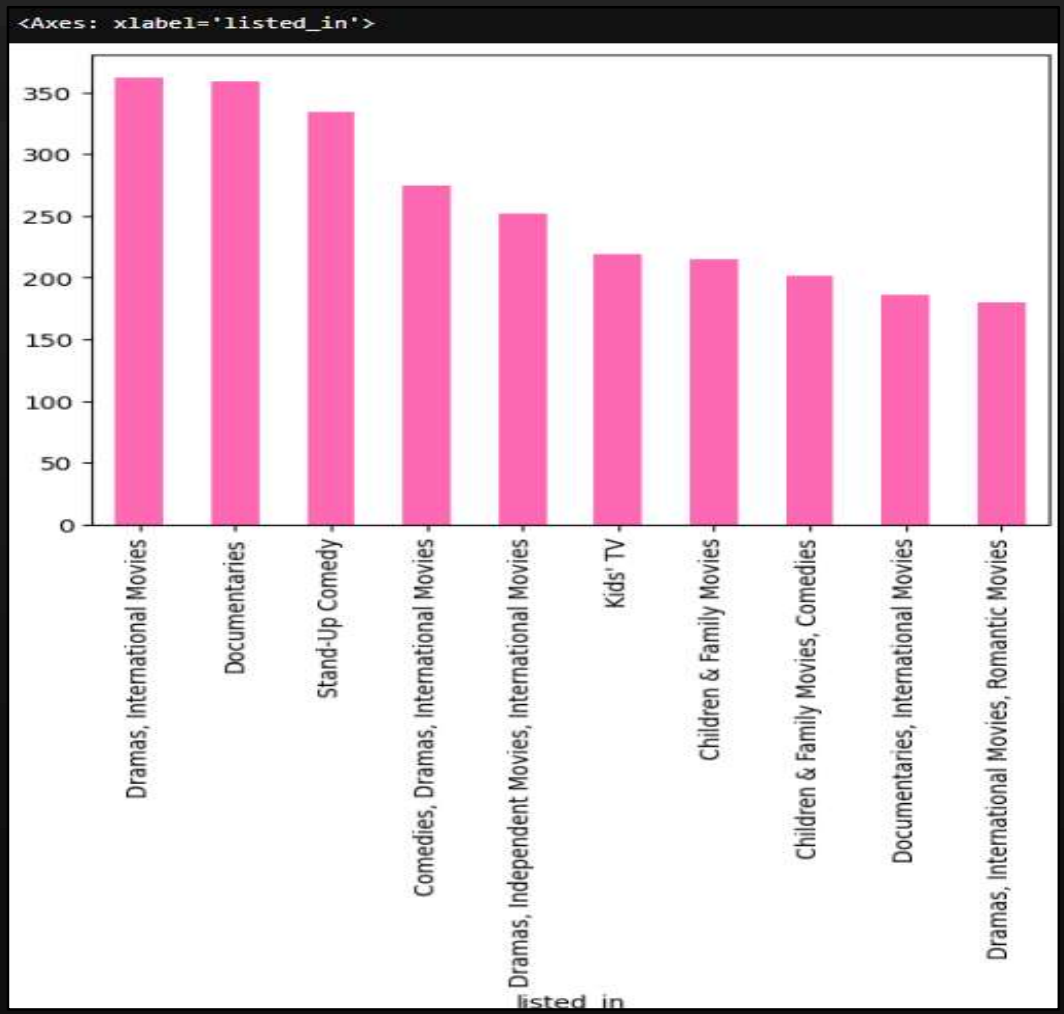


"This chart shows the count of Movies and TV Shows available on Netflix. The platform has a higher number of movies compared to TV shows."

NETFLIX

Top 10 Most Popular Netflix Genres

```
#Top 10 Most Popular Netflix Genres
netflix['listed_in'].value_counts().head(10).plot(kind='bar', color='hotpink')
```



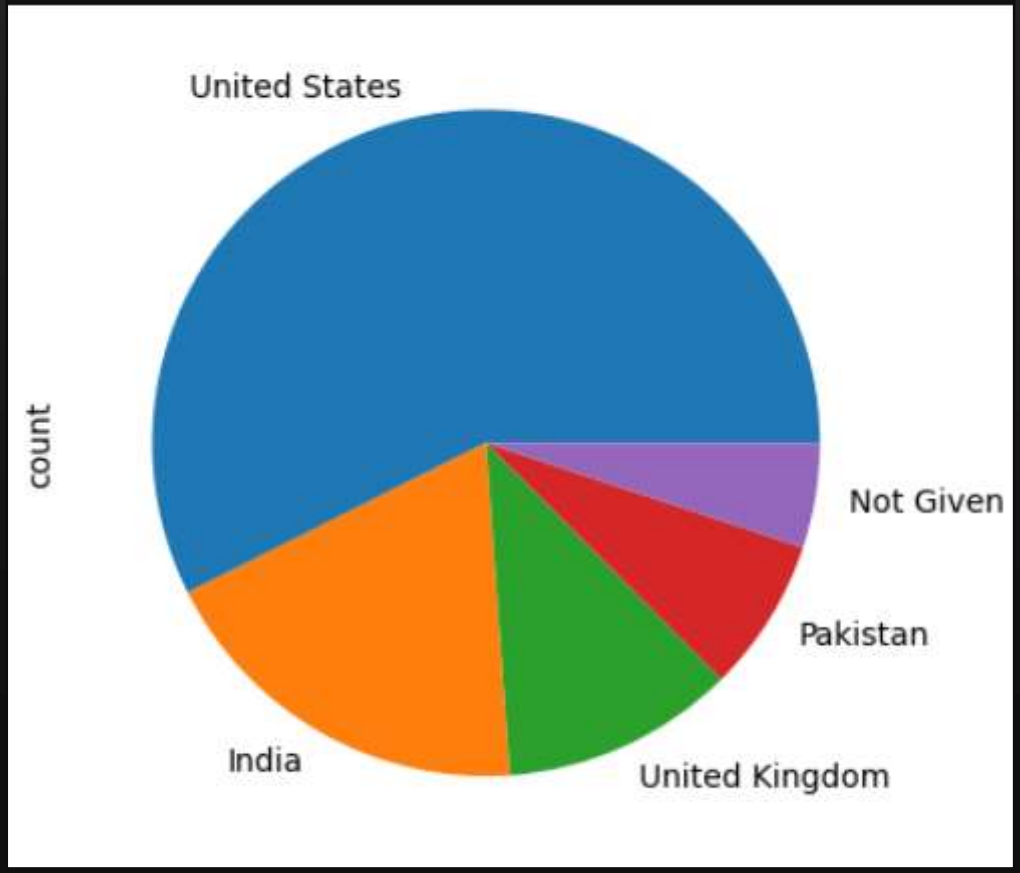
"This bar chart displays the top 10 most common genres available on Netflix, highlighting user preferences."

NETFLIX

Netflix Content Distribution by Country

```
#Country-wise Content Count
netflix['country'].value_counts().head(5).plot(kind='pie')
```

<Axes: ylabel='count'>

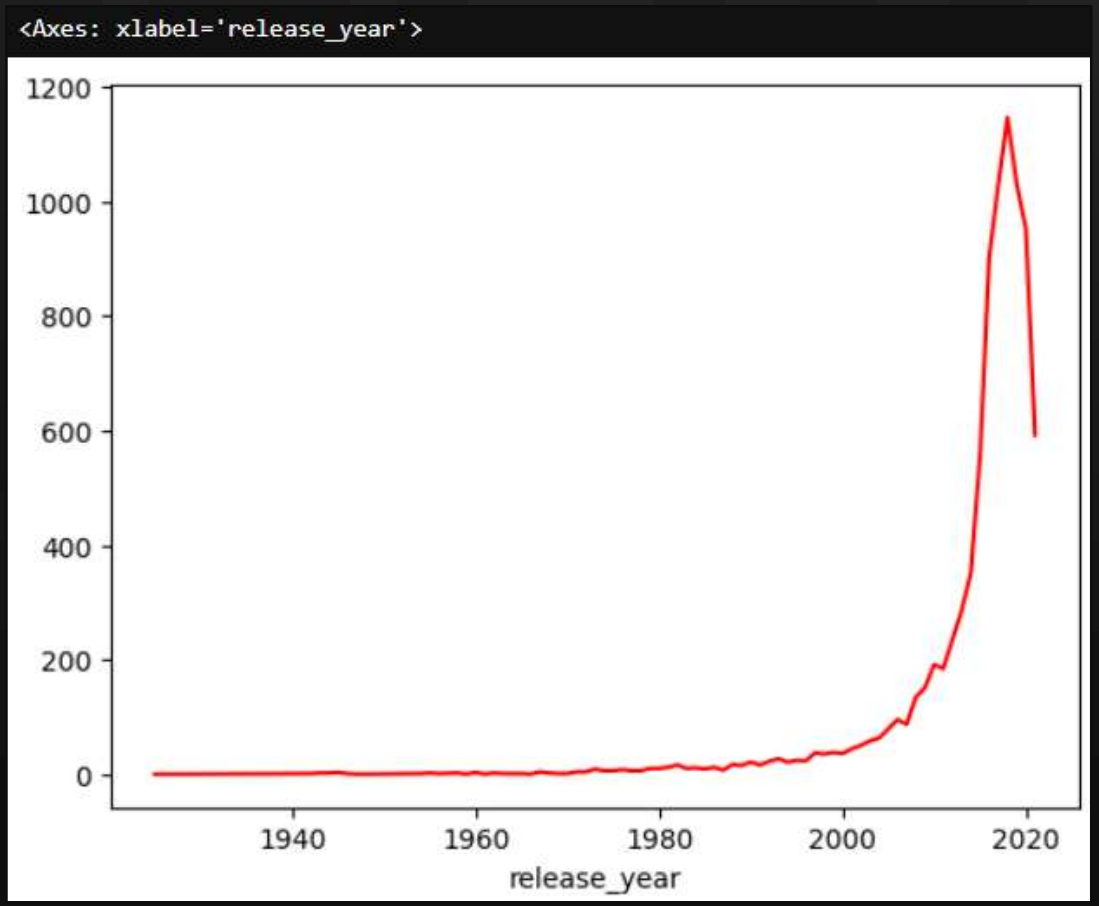


"The pie chart illustrates the proportion of content available in different countries on Netflix."

NETFLIX

Trend of Netflix Content Over the Years

```
#Year-wise Content Trend
netflix['release_year'].value_counts().sort_index().plot(kind='line')
```



"This line chart shows the trend of content added to Netflix over the years, indicating a sharp rise in recent years."

NETFLIX

Power BI Cleaning & Dashboard

After performing initial data checks in Python, we used Power BI for further data transformation and visualization. Power BI allows us to:

✅ Clean and Transform Data – We fixed formatting issues like date format corrections.
✅ Create Interactive Dashboards – To analyze Netflix content trends dynamically.
✅ Visualize Key Insights – Using bar charts, pie charts, and line charts for better understanding.

Using Power BI, we transformed raw data into an interactive and visually appealing dashboard that provides deep insights into Netflix's content distribution, genre trends, and yearly content addition patterns. 🚀

NETFLIX

Handling Date Format Issues in Power BI

During the data cleaning process in Power BI, we encountered errors in the "date_added" column. The values were not recognized as proper date formats, which could affect time-based analysis.

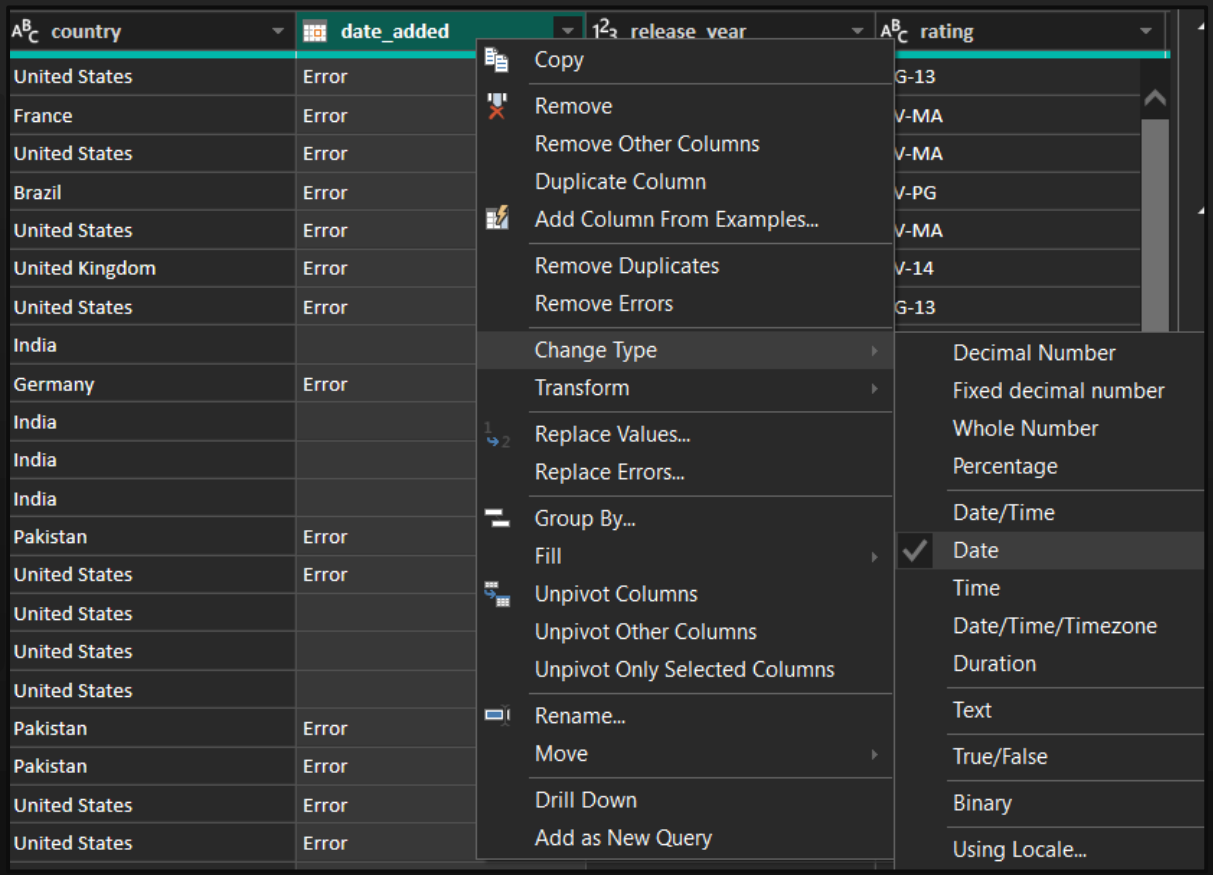
To resolve this issue, we performed the following steps:

- Identified the error in the "date_added" column.
- Changed the data type from "Error" to "Date" using the Transform Data option.
- Ensured that all values were correctly formatted, making the dataset ready for analysis.

After fixing this issue, the dataset was clean and ready for visualization.

NETFLIX

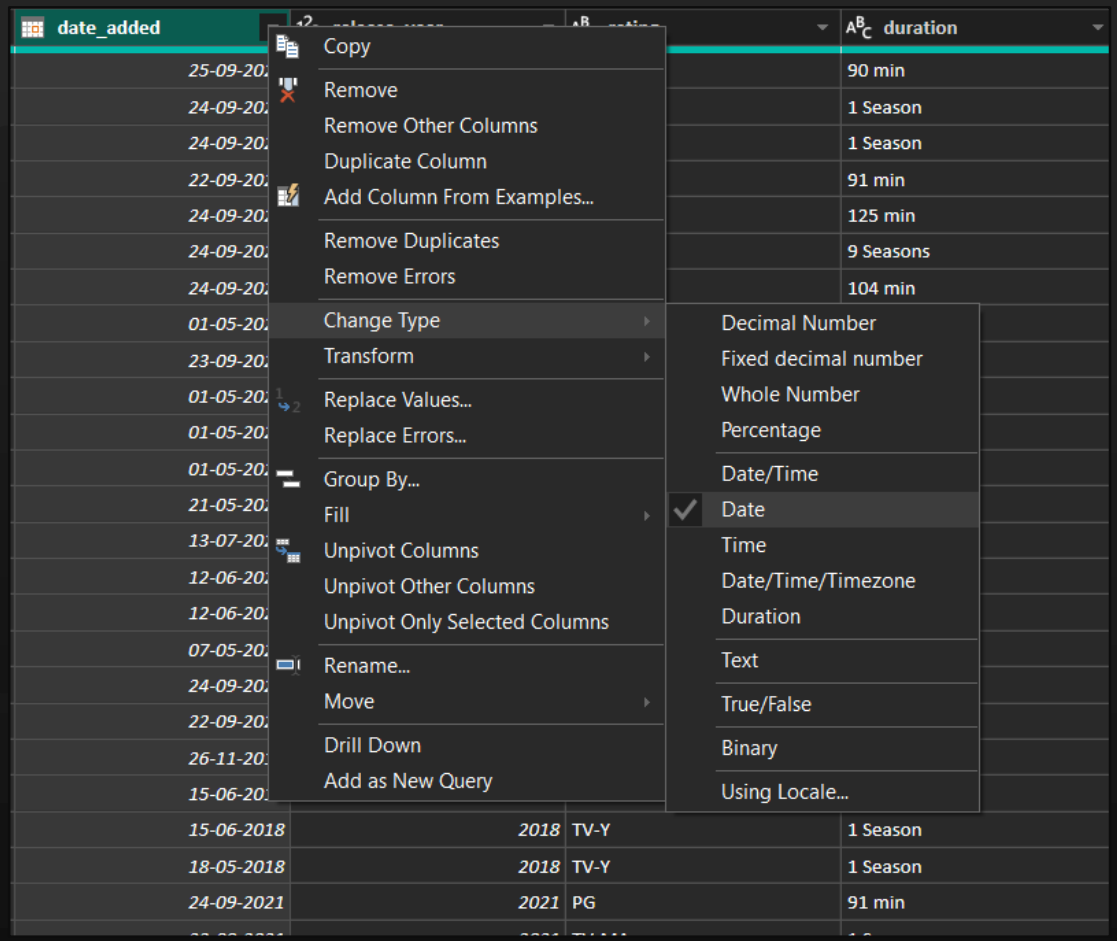
Date Format Error



The **"date_added"** column contained errors due to incorrect data formatting. This issue needed to be resolved for accurate time-based analysis.

NETFLIX

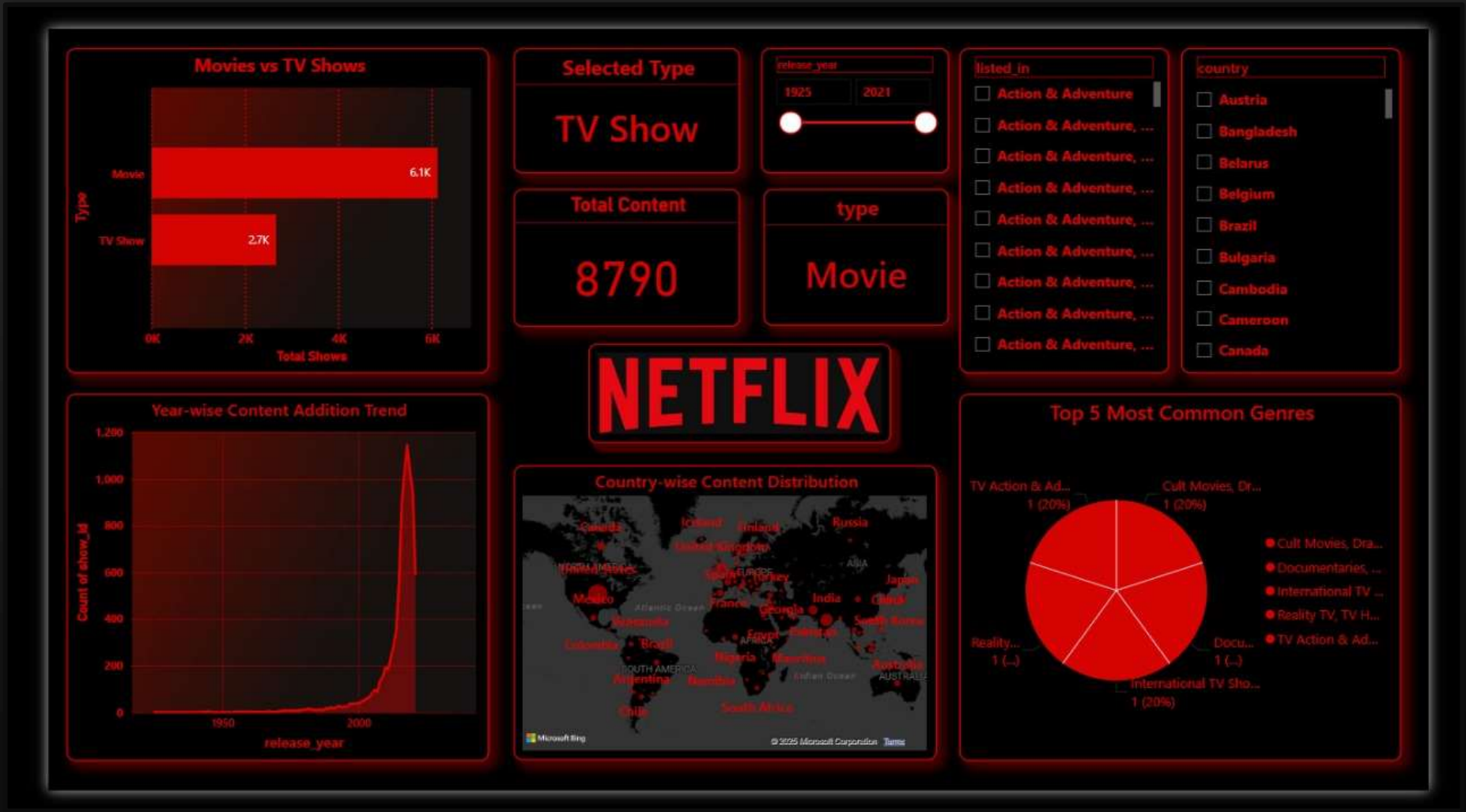
Fixed Date Format



The data type of "date_added" was successfully changed to Date, resolving the errors and ensuring proper formatting for analysis.

NETFLIX

Netflix Data Analysis Dashboard



NETFLIX

Challenges Faced

✓ Date Format Issue: Power BI had incorrect date formatting, which was fixed during data cleaning.

✓ Inconsistent Data: Some values needed formatting adjustments for proper analysis.

✓ Data Understanding: Identifying key insights required detailed exploration of the dataset.

✓ Visualization Choices: Selecting the right charts to represent data effectively.

NETFLIX

Conclusion & Key Takeaways

✓ Clean Data: No missing values or duplicates found; fixed date format errors in Power BI.

✓ Movies Dominate: Netflix has more movies than TV shows, with most content from the U.S.

✓ Popular Genres: Drama & Comedy lead the charts.

✓ Rising Trend: Content production has surged in recent years.

✓ Power of Visuals: Python & Power BI helped uncover key trends.

✓ Clean data: Well-structured data ensures meaningful and accurate insights.

NETFLIX

THANK YOU