

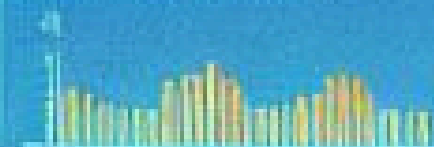


NETFIX

LLII —
TEII —

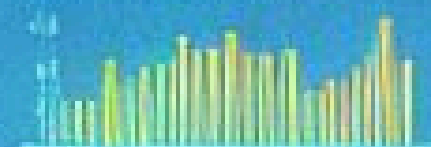
CONTENTS

www.netflix.com



SHOWS

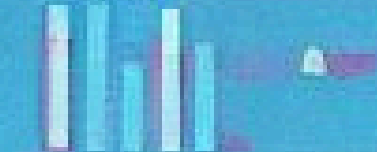
www.netflix.com



MOVIES

DOCUMENTARIES

Netflix



COTENATIONS

2010

ON BOARD

Netflix
100%
0.5 to 11

01

22 30

95%

30%

60%

N

NETFLIX DATA ANALYSIS

Exploring Content Trends and Patterns on Netflix



Siddhi Nirmal

N

Netflix Dataset Overview

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description	duration1
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As her father nears the end of his life, filmm...	90.0
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...	2.0
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...	To protect his family from a powerful drug lor...	1.0
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV	Feuds, flirtations and toilet talk go down amo...	1.0
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	In a city of coaching centers known to train I...	2.0

N

Dataset Information and Missing Values

```
netflix.info()  
missing_columns = netflix.columns[netflix.isnull().any()  
print("Columns with missing values:", missing_columns)
```

DATASET STRUCTURE:

THE DATASET HAS 8807 ENTRIES AND 12 COLUMNS.

DATA TYPES: 11 OBJECT COLUMNS AND 1 INTEGER COLUMN.

COLUMNS WITH MISSING VALUES:

- DIRECTOR (2634 MISSING)
- CAST (825 MISSING)
- COUNTRY (831 MISSING)
- DATE_ADDED (10 MISSING)
- RATING (4 MISSING)
- DURATION (3 MISSING)

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 8807 entries, 0 to 8806  
Data columns (total 12 columns):  
#      Column      Non-Null Count  Dtype  
---  -  
0     show_id      8807 non-null    object  
1     type         8807 non-null    object  
2     title        8807 non-null    object  
3     director     6173 non-null    object  
4     cast         7982 non-null    object  
5     country      7976 non-null    object  
6     date_added   8797 non-null    object  
7     release_year 8807 non-null    int64  
8     rating       8803 non-null    object  
9     duration     8804 non-null    object  
10    listed_in    8807 non-null    object  
11    description  8807 non-null    object
```


N

Top 10 Countries Contributing Content.

```
top_countries = netflix['country'].value_counts().head(10)
print(top_countries)
```

An analysis of the countries contributing the most content to Netflix. Example insight: "The United States leads in content contribution, followed by other major countries."

```
country
United States    2818
India            972
United Kingdom   419
Japan            245
South Korea      199
Canada           181
Spain            145
France           124
Mexico           110
Egypt            106
Name: count, dtype: int64
```

N

Top 5 Most Popular Genres

```
] popular_genres = netflix['listed_in'].value_counts().head(5)
print(popular_genres)
```

OVERVIEW OF POPULAR GENRES SUCH AS "DRAMA" AND "COMEDY," WITH INSIGHTS ON TRENDS.

```
listed_in
Dramas, International Movies      362
Documentaries                    359
Stand-Up Comedy                  334
Comedies, Dramas, International Movies  274
Dramas, Independent Movies, International Movies  252
Name: count, dtype: int64
```



Categorizing Content as TV Shows and Movies

This slide compares the characteristics of Netflix content, split by Movies and TV Shows. The tables below provide a quick look at sample rows, showing differences in fields like duration, release year, genre, and country. This distinction helps highlight Netflix's variety in content formats.

Movies:

show_id	type	title	director	country	release_year	rating	duration	listed_in
s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	United States	2020	PG-13	90 min	Documentaries
s7	Movie	My Little Pony: A New Gen	Robert Cullen, José	NaN	2021	PG	91 min	Family Movies

TV Shows:

show_id	type	title	director	country	release_year	rating	duration
s2	TV Show	Blood & Water	NaN	South Africa	2021	TV-MA	2 Seasons
s3	TV Show	Ganglands	Julien Leclercq	NaN	2021	TV-MA	1 Season

```
movies = netflix[netflix['type'] == 'Movie'].head(2)
shows = netflix[netflix['type'] == 'TV Show'].head(2)
print("Movies:\n", movies)
print("TV Shows:\n", shows)
```

N

Oldest Movies on Netflix

The oldest movies available on Netflix are from the 1940s, specifically during the years 1942–1943.

```
] oldest_movies = netflix[netflix['type'] == 'Movie'].sort_values(by='release_year')
print(oldest_movies[['title', 'release_year']])
```

	title	release_year
7790	Prelude to War	1942
8205	The Battle of Midway	1942
8660	Undercover: How to Operate Behind Enemy Lines	1943
8763	WWII: Report from the Aleutians	1943
8739	Why We Fight: The Battle of Russia	1943

N

Duration Column Value Counts

```
duration_counts = netflix['duration'].value_counts()  
print(duration_counts)
```

Netflix offers content
with varying durations,
from short episodes to
full-length movies:

```
duration  
1 Season      1793  
2 Seasons     425  
3 Seasons     199  
90 min        152  
94 min        146  
...  
16 min         1  
186 min        1  
193 min        1  
189 min        1  
191 min        1  
Name: count, Length: 220, dtype: int64
```

N

Duration - Measures of Central Tendency and Dispersion

```
netflix['duration_num'] = netflix['duration'].str.extract('(\d+)').astype(float)
print("Mean:", netflix['duration_num'].mean())
print("Median:", netflix['duration_num'].median())
print("Standard Deviation:", netflix['duration_num'].std())
```

This analysis shows that while the average content length is around 70 minutes, there's significant variability, with some content much shorter or longer than the median of 88 minutes.

```
Mean: 69.84688777828259
Median: 88.0
Standard Deviation: 50.81482778918891
```

N

Longest Movie on Netflix

Black Mirror: Bandersnatch is an interactive movie that stands out as the longest content on Netflix, with a runtime of over 5 hours.

```
netflix_movies = netflix.dropna(subset=['duration'])
netflix_movies = netflix_movies[netflix_movies['duration'].str.contains("min")]
netflix_movies['duration_num'] = netflix_movies['duration'].str.extract('(\d+)').astype(float)
longest_movie = netflix_movies.loc[netflix_movies['duration_num'].idxmax()]
print(longest_movie[['title', 'country', 'duration']])
```

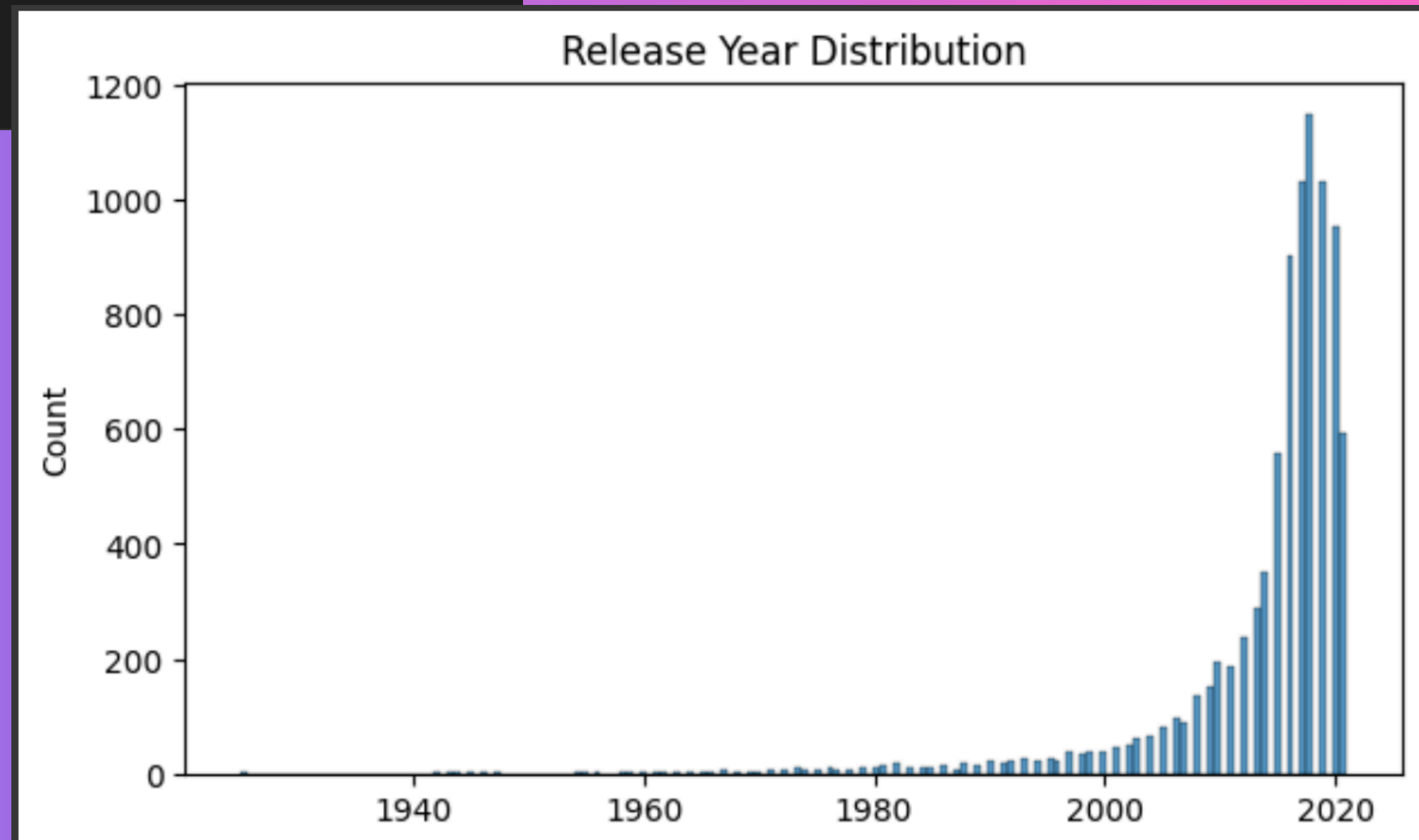
```
title          Black Mirror: Bandersnatch
country                United States
duration                312 min
Name: 4253, dtype: object
```



Content Release Year Distribution

```
plt.figure(figsize=(10,6))
sns.histplot(x=netflix["release_year"])
plt.title('Release Year Distribution')
plt.xlabel("Release Year")
plt.ylabel("Count")
plt.show()
```

- Majority of content was released in recent years, especially after 2000, with a significant peak in releases around 2018-2020.
- Netflix's content library has expanded rapidly in recent years, indicating a growing focus on producing or acquiring more recent titles.



N

Popularity Distribution by Content Category

```
plt.figure(figsize=(7, 4))
sns.countplot(data=Loading..., y="listed_in", order=netflix["listed_in"].value_counts().index[0:20])
plt.title("Popularity Distribution by Content Category")
plt.show()
```

- "Dramas, International Movies" is the most popular content category, followed by "Documentaries" and "Stand-Up Comedy."
- This distribution reflects Netflix's diverse library with a high demand for drama and international movies, as well as specialized categories like documentaries and stand-up comedy.

