

Netflix Data Analysis Project (Python)

Netflix is a leading global streaming platform offering a vast library of movies, TV shows, documentaries, and original content across various genres. With millions of subscribers worldwide, it continues to shape the entertainment industry through its vast data on user preferences and viewing habits.

This document presents an analysis of Netflix's streaming data using Python in Google Colab, incorporating sentiment analysis. The analysis aims to uncover insights into viewing trends, user preferences, and content performance, providing a data-driven approach to understanding audience sentiment and engagement.

Google Colab Code and Its Result

Importing Necessary Libraries

```
#Importing the Necessary Libraries

import numpy as np #For Algebra Operations
import pandas as pd #For Data Preparation
import plotly.express as px #For Data Visualization
from textblob import TextBlob #For Sentiment Analysis
```

Importing the Dataset

```
#Importing The Dataset

df = pd.read_csv('netflix_titles.csv')
```

Previewing the Dataset

```
[8] #Previewing the Dataset  
df.shape
```

```
(8807, 12)
```

```
df.head()
```


	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
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...
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	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...
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...
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 l...

```
df.info()
```

```
<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  
dtypes: int64(1), object(11)  
memory usage: 825.8+ KB
```


Creating a Group of Ratings to use it in Pie Chart

```
[17] x = df.groupby(['rating']).size().reset_index(name='counts')
      print (x)
```



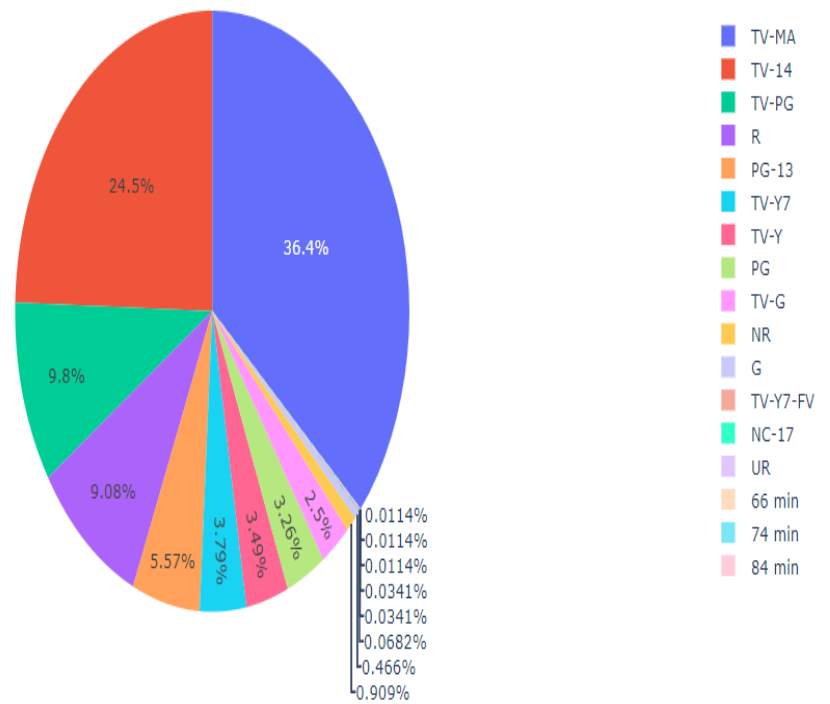
	rating	counts
0	66 min	1
1	74 min	1
2	84 min	1
3	G	41
4	NC-17	3
5	NR	80
6	PG	287
7	PG-13	490
8	R	799
9	TV-14	2160
10	TV-G	220
11	TV-MA	3207
12	TV-PG	863
13	TV-Y	307
14	TV-Y7	334
15	TV-Y7-FV	6
16	UR	3

Creating the Pie Chart



```
piechart = px.pie(x, values="counts", names="rating", title= "Distribution of Content Ratings on Netflix")
piechart.show()
```

Distribution of Content Ratings on Netflix



Replacing the Null Values in the Director's Column

```
df['director'] = df['director'].fillna('Director Not Specified')
df.head()
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
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...
1	s2	TV Show	Blood & Water	Director Not Specified	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	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...
3	s4	TV Show	Jailbirds New Orleans	Director Not Specified	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV	Feuds, flirtations and toilet talk go down amo...
4	s5	TV Show	Kota Factory	Director Not Specified	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 L...

Creating Director List

```

directors_list = pd.DataFrame
directors_list = df['director'].str.split(',', expand=True).stack()
directors_list = directors_list.to_frame()
print(directors_list)

```

```

0      0      Kirsten Johnson
1      0  Director Not Specified
2      0      Julien Leclercq
3      0  Director Not Specified
4      0  Director Not Specified
...
8802  0      David Fincher
8803  0  Director Not Specified
8804  0      Ruben Fleischer
8805  0      Peter Hewitt
8806  0      Moez Singh

[9612 rows x 1 columns]

```

Counting the no. of Times one director has appeared by grouping them

```

directors = directors_list.groupby(['Director']).size().reset_index(name='Total Count')
print(directors)

```

```

0      Director  Total Count
1      Aaron Moorhead      2
2      Aaron Woolf      1
3  Abbas Alibhai Burmawalla      1
4  Abdullah Al Noor      1
5  Abhinav Shiv Tiwari      1
...
5116      Cagan Irmak      1
5117  Ísöld Uggadóttir      1
5118  Óskar Thór Axelsson      1
5119  Ömer Faruk Sorak      2
5120  Şenol Şönmez      2

[5121 rows x 2 columns]

```

Removing Director not specified value

```
directors = directors[directors.Director != 'Director Not Specified']  
print(directors)
```

```
Director  Total Count  
0      Aaron Moorhead      2  
1      Aaron Woolf         1  
2      Abbas Alibhai Burmawalla  1  
3      Abdullah Al Noor      1  
4      Abhinav Shiv Tiwari    1  
...      ...              ...  
5116     Çagan Irmak        1  
5117     Ísold Uggadóttir    1  
5118     Óskar Thór Axelsson  1  
5119     Ömer Faruk Sorak    2  
5120     Şenol Sönmez       2  
[5120 rows x 2 columns]
```

Sorting Directors in Descending Order by their Total Count of Shows/Movies

```
directors = directors.sort_values(by=['Total Count'], ascending=False)  
print(directors)
```

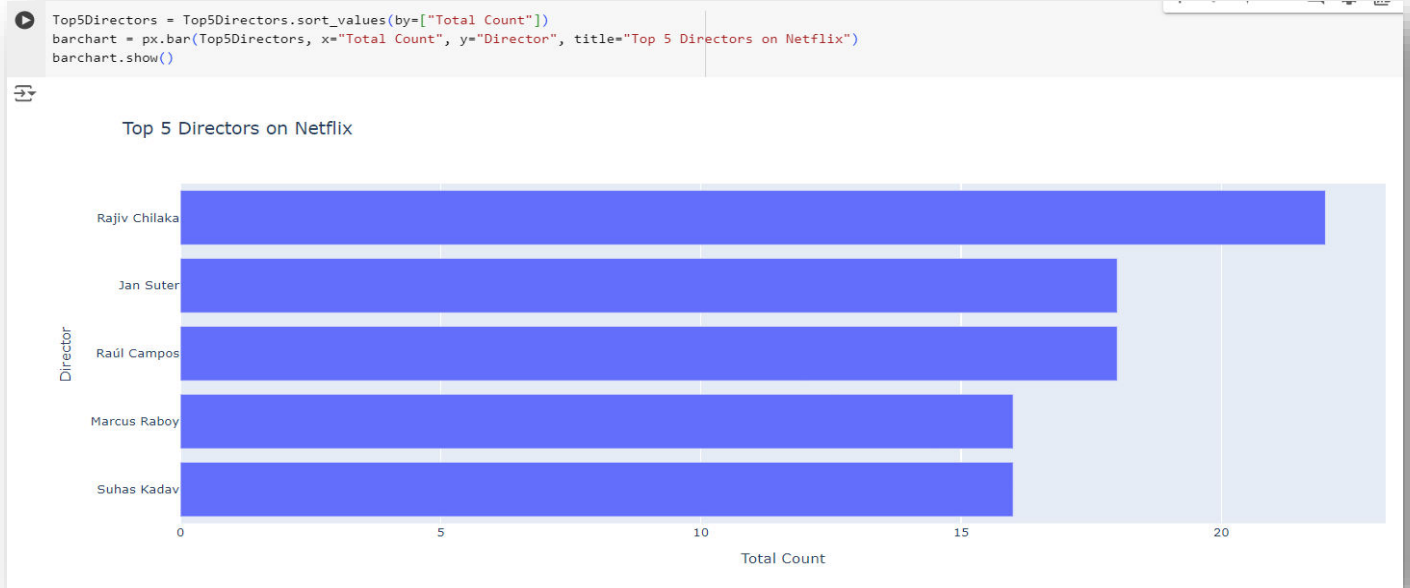
```
Director  Total Count  
4021  Rajiv Chilaka      22  
4068  Raúl Campos       18  
261   Jan Suter         18  
4652  Suhas Kadav       16  
3236  Marcus Raboy      16  
...   ...              ...  
2341   J. Davis          1  
2342  J. Lee Thompson    1  
2343  J. Michael Long    1  
609   Smriti Keshari     1  
2561  Joaquín Mazón      1  
[5120 rows x 2 columns]
```

Sorting out Top 5 Directors

```
Top5Directors = directors.head(5)  
print(Top5Directors)
```

```
Director  Total Count  
4021  Rajiv Chilaka      22  
4068  Raúl Campos       18  
261   Jan Suter         18  
4652  Suhas Kadav       16  
3236  Marcus Raboy      16
```

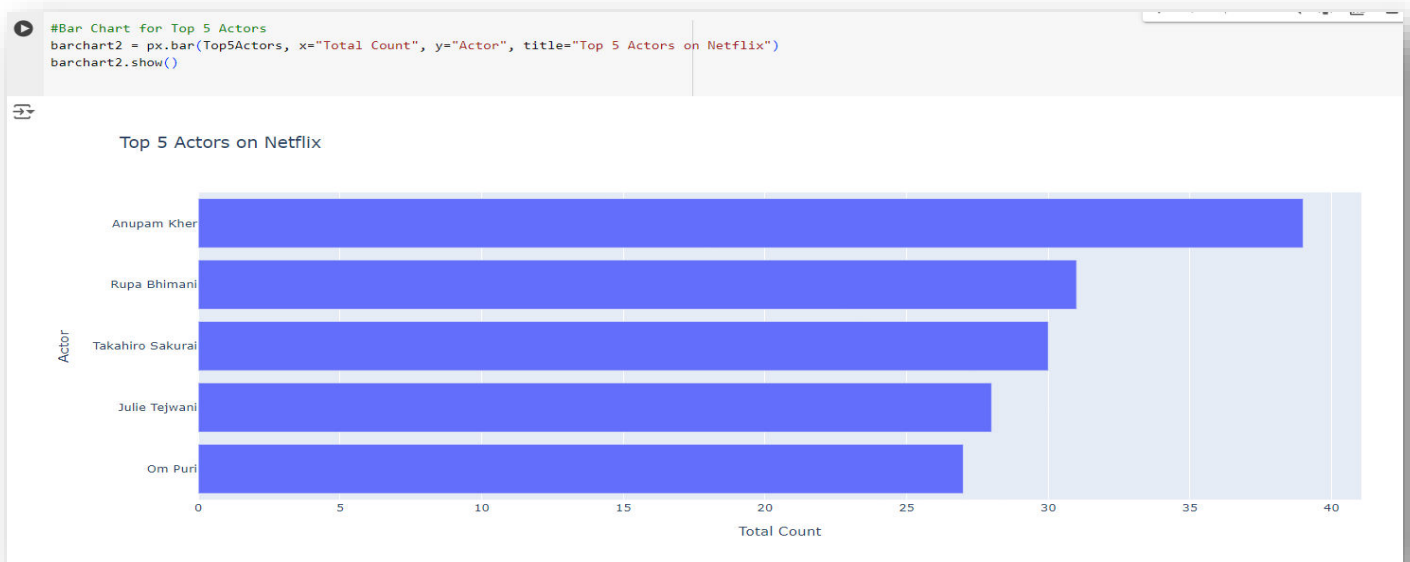
Bar Chart for Top 5 Directors



Replacing Null Values from Cast, grouping them by Content and Then Sorting out the Top 5 Actors

```
[47] #Replacing the Null Values from cast, Grouping the cast by Content and Sorting top 5 Actors
df['cast'] = df['cast'].fillna('Cast Not Specified')
cast_df = pd.DataFrame()
cast_df = df['cast'].str.split(',', expand=True).stack()
cast_df = cast_df.to_frame()
cast_df.columns = ['Actor']
actors = cast_df.groupby(['Actor']).size().reset_index(name='Total Count')
actors = actors[actors.Actor != 'Cast Not Specified']
actors = actors.sort_values(by=['Total Count'], ascending=False)
Top5Actors = actors.head(5)
Top5Actors = Top5Actors.sort_values(by=["Total Count"])
```

Bar Chart for Top 5 Actors



Creating a data frame for Content produced Every Year on Netflix

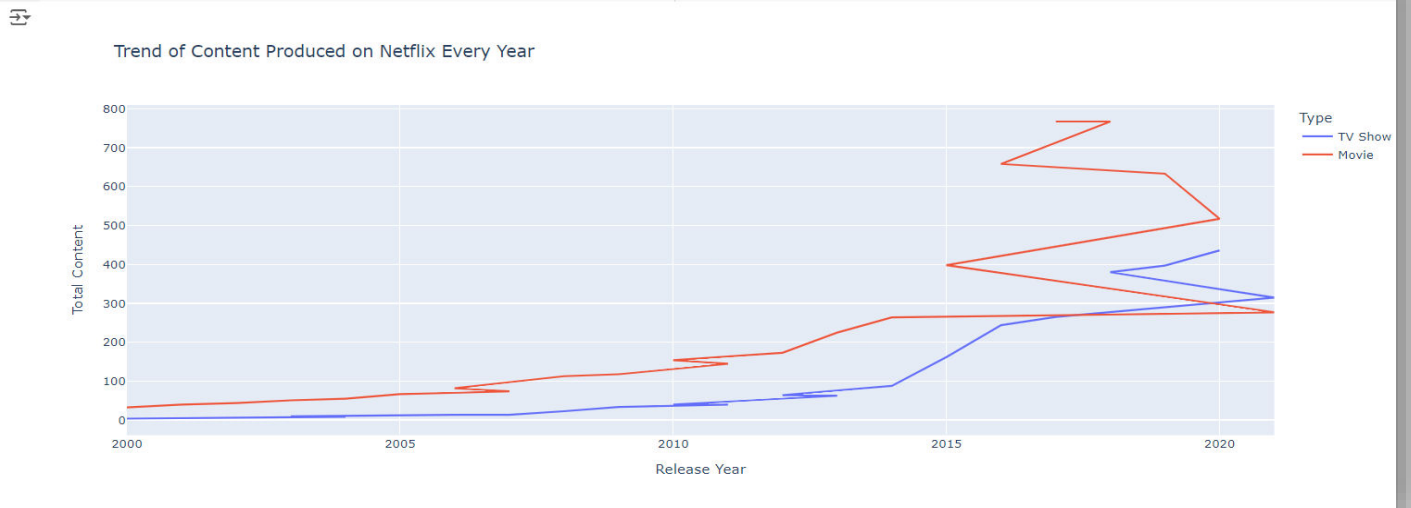
```
[55] #Analyzing the Content Produced on Netflix Based on Years
df1 = df[['type', 'release_year']]
df1 = df1.rename(columns={"release_year": "Release Year", "type": "Type"})
df2 = df1.groupby(['Release Year', 'Type']).size().reset_index(name='Total Content')
df2 = df2.sort_values(by=['Total Content'], ascending=True)
print(df2)
```

	Release Year	Type	Total Content
0	1925	TV Show	1
54	1989	TV Show	1
18	1963	TV Show	1
17	1963	Movie	1
29	1972	TV Show	1
..
115	2020	Movie	517
113	2019	Movie	633
107	2016	Movie	658
111	2018	Movie	767
109	2017	Movie	767

[119 rows x 3 columns]

Creating line Chart for the Type of Content released from Year 2000


```
df2 = df2[df2['Release Year'] >= 2000]
graph = px.line(df2, x="Release Year", y="Total Content", color="Type", title="Trend of Content Produced on Netflix Every Year")
graph.show()
```



Grouping the Countries and Finding out Total Releases by them

```
[63] # Grouping by country and counting the total content releases
df_country_content = df[['country', 'type']]
df_country_content = df_country_content.rename(columns={"country": "Country", "type": "Type"})

df_country_total = df_country_content.groupby('Country').size().reset_index(name='Total Content')

df_country_total = df_country_total.sort_values(by='Total Content', ascending=False)
print(df_country_total)
```

	Country	Total Content
603	United States	2818
251	India	972
506	United Kingdom	419
319	Japan	245
435	South Korea	199
...
281	Ireland, Canada, Luxembourg, United States, Un...	1
282	Ireland, Canada, United Kingdom, United States	1
283	Ireland, Canada, United States, United Kingdom	1
284	Ireland, France, Iceland, United States, Mexic...	1
747	Zimbabwe	1

[748 rows x 2 columns]

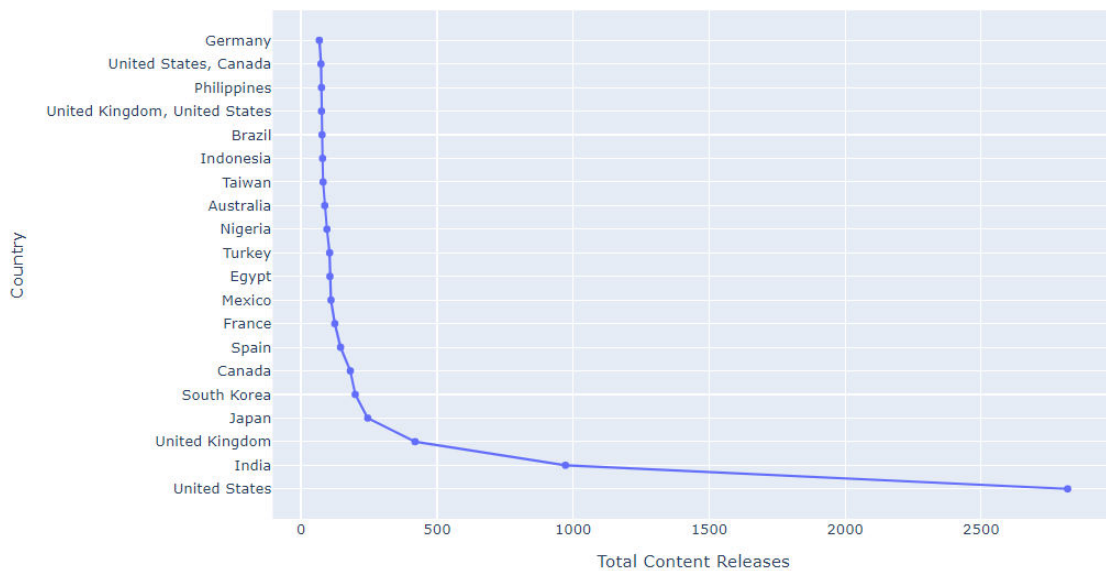
Creating a Line Chart for Top 20 Countries in releasing content

```
#Top 20 countries with the highest content releases
top_20_countries = df_country_total.head(20)
line_chart = px.line(top_20_countries,
                     x='Total Content',
                     y='Country',
                     title='Top 20 Countries by Content Releases',
                     labels={'Total Content': 'Number of Content Releases', 'Country': 'Country'},
                     markers=True)

line_chart.update_layout(
    xaxis_title='Total Content Releases',
    yaxis_title='Country',
    template='plotly',
    height=600,
    width=1000,
    showlegend=False
)
line_chart.show()
```



Top 20 Countries by Content Releases



Sentiment Analysis of Netflix Content

Creating Data frame and categorizing Content Into neutral Positive and Negative and Creating the BAR Graph

```
#Sentiment Analysis of Netflix Content
#Creating Dataframe and categorizing Content Into neutral Positive and Negative and Creating the BAR Graph
df3 = df[['release_year', 'description']] # Select relevant columns
df3 = df3.rename(columns={"release_year": "Release year", "description": "Description"}) # Rename for clarity

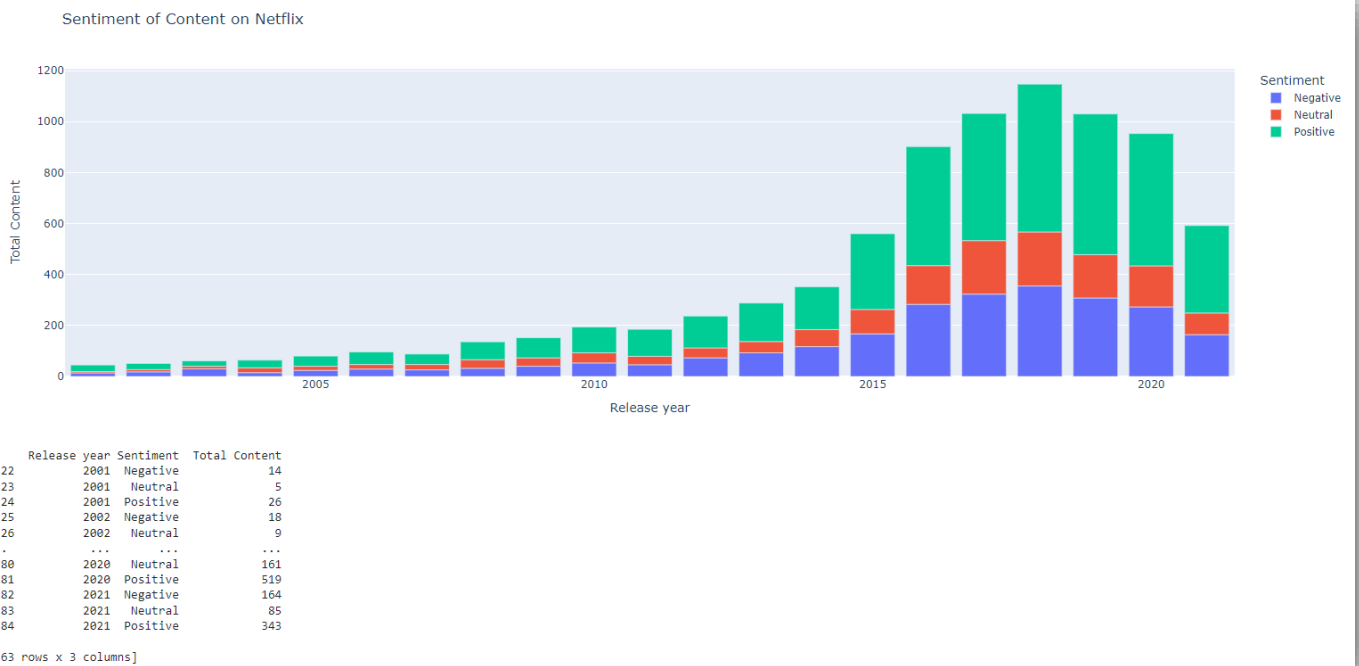
sentiments = []
for index, row in df3.iterrows():
    description = row['Description']
    testimonial = TextBlob(description)
    polarity = testimonial.sentiment.polarity
    if polarity == 0:
        sentiment = "Neutral"
    elif polarity < 0:
        sentiment = "Negative"
    else:
        sentiment = "Positive"

    sentiments.append(sentiment)

df3['Sentiment'] = sentiments
df3_grouped = df3.groupby(['Release year', 'Sentiment']).size().reset_index(name='Total Content')
df3_grouped = df3_grouped[df3_grouped['Release year'] > 2000] # Filter for years greater than 2000

BarGraph = px.bar(df3_grouped, x="Release year", y="Total Content", color="Sentiment", title="Sentiment of Content on Netflix")

BarGraph.show()
print(df3_grouped)
```



Netflix Data Analysis Project (SQL)

This part of the document explores SQL-based data analysis using PostgreSQL. It addresses 15 diverse queries, providing the corresponding SQL code and detailed results for each. The project aims to demonstrate practical problem-solving and data manipulation techniques in a relational database.

PostgreSQL Queries and Results

1. Count the no. of Movies and TV Shows

Query

```
-- Problem 1 Count the Number of Movies and TV shows
select
    type,
    count (*) as total_content
from Netflix
group by type
```

Result

	type character varying (10) 🔒	total_content bigint 🔒
1	Movie	6131
2	TV Show	2676

2. Result Find the Most common rating for Movies and TV Shows

Query

```
-- Problem 2 Find the most common rating for Movies and TV Shows
SELECT
    type,
    rating
FROM
    (SELECT
        type,
        rating,
        COUNT(*) AS count,
        RANK() OVER (PARTITION BY type ORDER BY COUNT(*) DESC) AS ranking
    FROM Netflix
    GROUP BY type, rating
    ) as t1
where
    ranking = 1
```

Result

	type character varying (10)	rating character varying (10)
1	Movie	TV-MA
2	TV Show	TV-MA

3. List all Movies released in the year 2020

Query

```
-- Problem 3 List all the Movies released in 2020
select * from Netflix
where
    type = 'Movie'
    and
    release_year = 2020
```

Result

	show_id character varying (6)	type character varying (10)	title character varying (150)	director character varying (250)	casts character varying (1000)
1	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	[null]
2	s17	Movie	Europe's Most Dangerous Man: Otto Skorzeny in Spain	Pedro de Echave García, Pablo Azorín Williams	[null]
3	s79	Movie	Tughlaq Durbar	Delhiprasad Deenadayalan	Vijay Sethupathi, Parthiban, Raashi Khanna
4	s85	Movie	Omo Ghetto: the Saga	JJC Skillz, Funke Akindele	Funke Akindele, Ayo Makun, Chioma Chukwuka Akpotha, Yemi Eberech
5	s104	Movie	Shadow Parties	Yemi Amodu	Jide Kosoko, Omotola Jalade-Ekeinde, Yemi Blaq, Sola Sobowale, Ken
6	s120	Movie	Here and There	JP Hahar	Ianina Gutierrez, JC Santos, Victor Anastasin, Vesh Ruirra, I ntint De L z

4. Find the Top 5 Countries with Most Content on Netflix

Query

```
SELECT
    UNNEST(string_to_array(country, ',')) AS new_country,
    COUNT(show_id) AS total_content
FROM Netflix
GROUP BY new_country
order by total_content desc
limit 5
```

Result

	new_country text	total_content bigint
1	United States	3211
2	India	1008
3	United Kingdo...	628
4	United States	479
5	Canada	271

5. Identify the Longest Movie

Query

```
-- Problem 5 Identify the longest Movie
SELECT *
FROM Netflix
WHERE
    type = 'Movie'
    AND
    CAST(SPLIT_PART(duration, ' ', 1) AS INTEGER) = (
        SELECT MAX(CAST(SPLIT_PART(duration, ' ', 1) AS INTEGER))
        FROM Netflix
        WHERE type = 'Movie'
    );
```

Result

title	director	casts	country	date_added	release_year	rating	duration
character varying (150)	character varying (250)	character varying (1000)	character varying (150)	character varying (50)	integer	character varying (10)	character varying (15)
Black Mirror: Bandersnatch	[null]	Fionn Whitehead, Will Poulter, Craig Parkinson, Alice Lowe, Asim Chaud...	United States	December 28, 2018	2018	TV-MA	312 min

6. Find the content added in the last 5 years

Query

```
-- Problem 6 Find the content added in the last 5 years
select * from Netflix
where
    TO_DATE(date_added, 'Month dd, YYYY') >= CURRENT_DATE - INTERVAL '5 years'
```

Result

	show_id character varying (6)	type character varying (10)	title character varying (150)
9	s9	TV Show	The Great British Baking Show
10	s10	Movie	The Starling
11	s11	TV Show	Vendetta: Truth, Lies and The Mafia
12	s12	TV Show	Bangkok Breaking
13	s13	Movie	Je Suis Karl
14	s14	Movie	Confessions of an Invisible Girl

date_added character varying (50)	release_year integer	rating character varying (10)	duration character varying (15)
September 24, 2021	2021	TV-14	9 Seasons
September 24, 2021	2021	PG-13	104 min
September 24, 2021	2021	TV-MA	1 Season
September 23, 2021	2021	TV-MA	1 Season
September 23, 2021	2021	TV-MA	127 min
September 22, 2021	2021	TV-PG	91 min

7. Find all the Movies/TV Shows By Director "Rajiv Chilaka"

Query

```
-- Problem 7 Find all the Movie/TV Shows by Director 'Rajiv Chilaka'
select * from Netflix
where director LIKE '%Rajiv Chilaka%'
```

Result

	show_id character varying (6)	type character varying (10)	title character varying (150)	director character varying (250)
1	s407	Movie	Chhota Bheem - Neeli Pahaadi	Rajiv Chilaka
2	s408	Movie	Chhota Bheem & Ganesh	Rajiv Chilaka
3	s409	Movie	Chhota Bheem & Krishna: Mayanagari	Rajiv Chilaka
4	s410	Movie	Chhota Bheem & Krishna: Pataliputra- City of the De...	Rajiv Chilaka
5	s411	Movie	Chhota Bheem And The Broken Amulet	Rajiv Chilaka
6	s412	Movie	Chhota Bheem And The Crown of Valhalla	Rajiv Chilaka
7	s413	Movie	Chhota Bheem and the Incan Adventure	Rajiv Chilaka
8	s414	Movie	Chhota Bheem and The ShiNobi Secret	Rajiv Chilaka
9	s415	Movie	Chhota Bheem Aur Hanuman	Rajiv Chilaka
10	s416	Movie	Chhota Bheem aur Krishna	Rajiv Chilaka
11	s417	Movie	Chhota Bheem aur Krishna vs Zimbara	Rajiv Chilaka
12	s420	Movie	Chhota Bheem: Bheem vs Aliens	Rajiv Chilaka
13	s421	Movie	Chhota Bheem: Dholakpur to Kathmandu	Rajiv Chilaka
14	s422	Movie	Chhota Bheem: Dus Pe Dus	Rajiv Chilaka, Owl Mina
15	s423	Movie	Chhota Bheem: Journey to Petra	Rajiv Chilaka

8. List all TV Shows with more than 5 Seasons

Query

```
-- Problem 8 List all TV Shows with more than 5 seasons
select * from Netflix
where
    type = 'TV Show'
    and
    SPLIT_PART(duration, ' ', 1)::numeric > 5
```

Result

	show_id character varying (6)	type character varying (10)	title character varying (150)
1	s9	TV Show	The Great British Baking Show
2	s56	TV Show	Nailed It
3	s66	TV Show	Numberblocks
4	s68	TV Show	Saved by the Bell
5	s83	TV Show	Lucifer
6	s275	TV Show	Grace and Frankie
7	s324	TV Show	30 Rock
8	s339	TV Show	Hunter X Hunter (2011)
9	s381	TV Show	The Flash
10	s391	TV Show	The Walking Dead
11	s526	TV Show	My Little Pony: Friendship Is Magic
12	s527	TV Show	Orange Is the New Black
13	s538	TV Show	Terrace House: Opening New Doo...
14	s549	TV Show	Grey's Anatomy

duration character varying (15)
9 Seasons
6 Seasons
6 Seasons
9 Seasons
6 Seasons
7 Seasons
7 Seasons
6 Seasons
7 Seasons
10 Seasons
8 Seasons
7 Seasons
6 Seasons
17 Seasons

9. Count the number of content items in each genre

Query

```
-- Problem 9 Count the number of Content items in each genre
select
  unnest (string_to_array(listed_in, ',')) as genre,
  count(show_id) as total_content
from Netflix
group by genre
```

Result

	genre text	total_content bigint
1	International TV Shows	774
2	Stand-Up Comedy	9
3	Spanish-Language TV Shows	172
4	Romantic Movies	613
5	Anime Features	21
6	TV Horror	11
7	Stand-Up Comedy & Talk Shows	34
8	Kids' TV	63
9	Docuseries	221
10	Children & Family Movies	605
11	Movies	57
12	Horror Movies	82
13	International Movies	2624
14	Music & Musicals	18
15	Reality TV	120

10. Find each year and the average number of content release by india on netflix return top 5 year with highest avg content release

Query

```
-- Problem 10. Find each year and the average number of content release by india on netflix
-- return top 5 year with highest avg content release
SELECT
    country,
    release_year,
    COUNT(show_id) AS total_release,
    ROUND(
        COUNT(show_id)::numeric /
        (SELECT COUNT(show_id) FROM netflix WHERE country = 'India')::numeric * 100, 2
    ) AS avg_release
FROM netflix
WHERE country = 'India'
GROUP BY country, release_year
ORDER BY avg_release DESC
LIMIT 5;
```

Result

	country character varying (150)	release_year integer	total_release bigint	avg_release numeric
1	India	2017	101	10.39
2	India	2018	94	9.67
3	India	2019	87	8.95
4	India	2020	75	7.72
5	India	2016	73	7.51

11. List all the content which is a Documentary

Query

```
-- Problem 11 List all the Content which is a documentary
SELECT *
FROM netflix
WHERE listed_in LIKE '%Documentaries'
```

Result

	show_id character varying (6)	type character varying (10)	title character varying (150)	director character varying (250)	casts character varying (1000)
1	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	[null]
2	s46	Movie	My Heroes Were Cowboys	Tyler Greco	[null]
3	s118	Movie	Final Account	Luke Holland	[null]
4	s7603	Movie	NOVA: Chinese Chariot Revealed	Giulia Clark	Eric Meyers
5	s222	Movie	Bob Ross: Happy Accidents, Betrayal & Greed	Joshua Rofé	Bob Ross
6	s305	Movie	Lady Boss: The Jackie Collins Story	Laura Fairlie	Jackie Collins
7	s7604	Movie	NOVA: Day the Dinosaurs Died	Sarah Holt	Craig Sechler
8	s379	Movie	Fantastic Funni	Louie Schwartzberg	[null]

12. Find all the content with no director

Query

```
-- Problem 12 Find all the content without Director
SELECT *
FROM netflix
WHERE director IS NULL
```

Result

	show_id character varying (6) 🔒	type character varying (10) 🔒	title character varying (150) 🔒	director character varying (250) 🔒
1	s2	TV Show	Blood & Water	[null]
2	s4	TV Show	Jailbirds New Orleans	[null]
3	s5	TV Show	Kota Factory	[null]
4	s11	TV Show	Vendetta: Truth, Lies and The Mafia	[null]
5	s15	TV Show	Crime Stories: India Detectives	[null]
6	s16	TV Show	Dear White People	[null]
7	s369	TV Show	Khawatir	[null]
8	s18	TV Show	Falsa identidad	[null]

13. Find out in how many movies salman khan appeared in last 10 years

Query

```
-- Problem 13 Find out in how many movies salman khan appeared in last 10 years
SELECT *
FROM netflix
WHERE casts LIKE '%Salman Khan%'
AND release_year > EXTRACT(YEAR FROM CURRENT_DATE) - 10
```

Result

	show_id character varying (6)	type character varying (10)	title character varying (150)	director character varying (250)	casts character varying (1000)
1	s2340	Movie	Prem Ratan Dhan Payo	Sooraj R. Barjatya	Salman Khan, Sonam Kapoor, Anupam Kher, Neil Nitin Mukesh, Sanjay Mishra, Swara Bhaskar, Deepak Dobriyal, Arman Kohli, Manoj Joshi, Samaira Rao, Karun
2	s2757	Movie	Paharganj	Rakesh Ranjan Kumar	Lorena Franco, Bijesh Jayarajan, Neet Chowdhary, Rajeev Gaursingh, Karan Jeet, Rajesh Sharma, Gautam Kurup, Poonam Mathur, Salman Khan, Manish Devku

14. Find the Top 10 Actors Who Have Appeared in the Highest Number of Movies Produced in India

Query

```
-- Problem 14 Find the Top 10 Actors Who Have Appeared in the Highest Number of Movies Produced in India
SELECT
    UNNEST(STRING_TO_ARRAY(casts, ',')) AS actor,
    COUNT(*)
FROM netflix
WHERE country = 'India'
GROUP BY actor
ORDER BY COUNT(*) DESC
LIMIT 10;
```

Result

	actor text	count bigint
1	Anupam Kher	36
2	Paresh Rawal	24
3	Shah Rukh Khan	24
4	Om Puri	23
5	Akshay Kumar	22
6	Boman Irani	21
7	Amitabh Bachchan	20
8	Kareena Kapoor	19
9	Naseeruddin Shah	19
10	Asrani	17

15. Categorize content as Parental Guidance Content when there is Presence of keywords like 'Kill' and 'Violence'

Query

```
-- Problem 15 Categorize content as Parental Guidance Content when there is Presence of keywords like 'Kill' and 'Violence'
SELECT
  category,
  COUNT(*) AS content_count
FROM (
  SELECT
    CASE
      WHEN description ILIKE '%kill%' OR description ILIKE '%violence%' THEN 'Parental Guidance Content'
      ELSE 'Normal'
    END AS category
  FROM netflix
) AS categorized_content
GROUP BY category;
```

Result

	category text	content_count bigint
1	Normal	8465
2	Parental Guidance Content	342

