Capstone Project

By Faiz Khan

Lets Perform Some Data Analytis On Netflix's DataSet

First we have to import some major liabraries

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import plotly.express as px
import warnings
warnings.filterwarnings("ignore")
```

Now we have to load DataSet

```
In [2]: df=pd.read_csv("netflix_titles.csv")

In [3]: df
```

Out[3]:

	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 I...
...	...	...	...	...	...	...	...	...	...	...	...	...
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007	R	158 min	Cult Movies, Dramas, Thrillers	A political cartoonist, a crime reporter and a...
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	July 1, 2019	2018	TV-Y7	2 Seasons	Kids' TV, Korean TV Shows, TV Comedies	While living alone in a spooky town, a young g...
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009	R	88 min	Comedies, Horror Movies	Looking to survive in a world taken over by zo...
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006	PG	88 min	Children & Family Movies, Comedies	Dragged from civilian life, a former superhero...
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies,	A scrappy but poor boy worms

Lets check null values ,drop duplicates and fill null values

```
In [6]: df.isna().sum()

Out[6]: show_id         0
        type            0
        title           0
        director     2634
        cast          825
        country       831
        date_added     10
        release_year    0
        rating          4
        duration        3
        listed_in       0
        description     0
        dtype: int64

In [7]: df.duplicated().sum()

Out[7]: 0
```

Now we have to fill missing values

```
In [8]: df.director=df.director.fillna("unknown")
        df.cast=df.cast.fillna("unknown")
        df.duration=df.duration.ffill()
        df.date_added=df.date_added.fillna("unknown")
        df.country=df.country.fillna("unknown")
        df=df.dropna()

In [9]: df.isna().sum()

Out[9]: show_id         0
        type            0
        title           0
        director        0
        cast            0
        country         0
        date_added      0
        release_year    0
        rating          0
        duration        0
        listed_in       0
        description     0
        dtype: int64
```

Now we have to remove outliers

We can't remove outliers because whole data is of object type

Lets perform some visualisation

```
In [11]: df.director.value_counts()

Out[11]: director
         unknown                                2631
         Rajiv Chilaka                            19
         Raúl Campos, Jan Suter                   18
         Suhas Kadav                              16
         Marcus Raboy                             16
                                                 ...
         Raymie Muzquiz, Stu Livingston            1
         Joe Menendez                              1
         Eric Bross                                1
         Will Eisenberg                            1
         Mozez Singh                               1
         Name: count, Length: 4528, dtype: int64
```

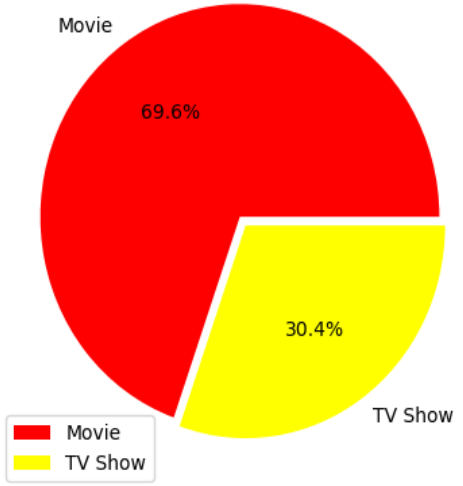
Percentage of content on Netflix

- This code is using Matplotlib, a popular Python plotting library, to create a pie chart based on the counts of different types in a DataFrame (assumed to be named df). Let's break it down:

- plt.pie(df.type.value_counts(), labels=df.type.value_counts().index, explode=[0.05, 0], colors=["red", "yellow"], autopct="%1.1f%%"): This line creates the pie chart. Here's what each argument does:
- df.type.value_counts(): This calculates the frequency of each unique value in the 'type' column of the DataFrame.
- labels=df.type.value_counts().index: This sets the labels of the pie chart slices to be the unique values of the 'type' column.
- explode=[0.05, 0]: This specifies the fraction of the radius with which to offset each wedge. It's used to explode one or more slices from the rest of the pie. Here, the first slice (corresponding to the first unique value of 'type') is slightly exploded.
- colors=["red", "yellow"]: This sets the colors of the pie chart slices. The first color "red" corresponds to the first unique value of 'type', and the second color "yellow" corresponds to the second unique value.
- autopct="%1.1f%%": This formats the percentage numbers shown in each slice. It displays one decimal place.
- plt.legend(): This line adds a legend to the plot. Since labels are already provided in the plt.pie() function, Matplotlib uses those labels to create the legend automatically

```
In [13]: plt.pie(df.type.value_counts(),labels=df.type.value_counts().index,explode=[0.05,0],colors=["red","yellow"],autopct="%1.1f%%")
         plt.legend()

Out[13]: <matplotlib.legend.Legend at 0x18eeed62610>
```

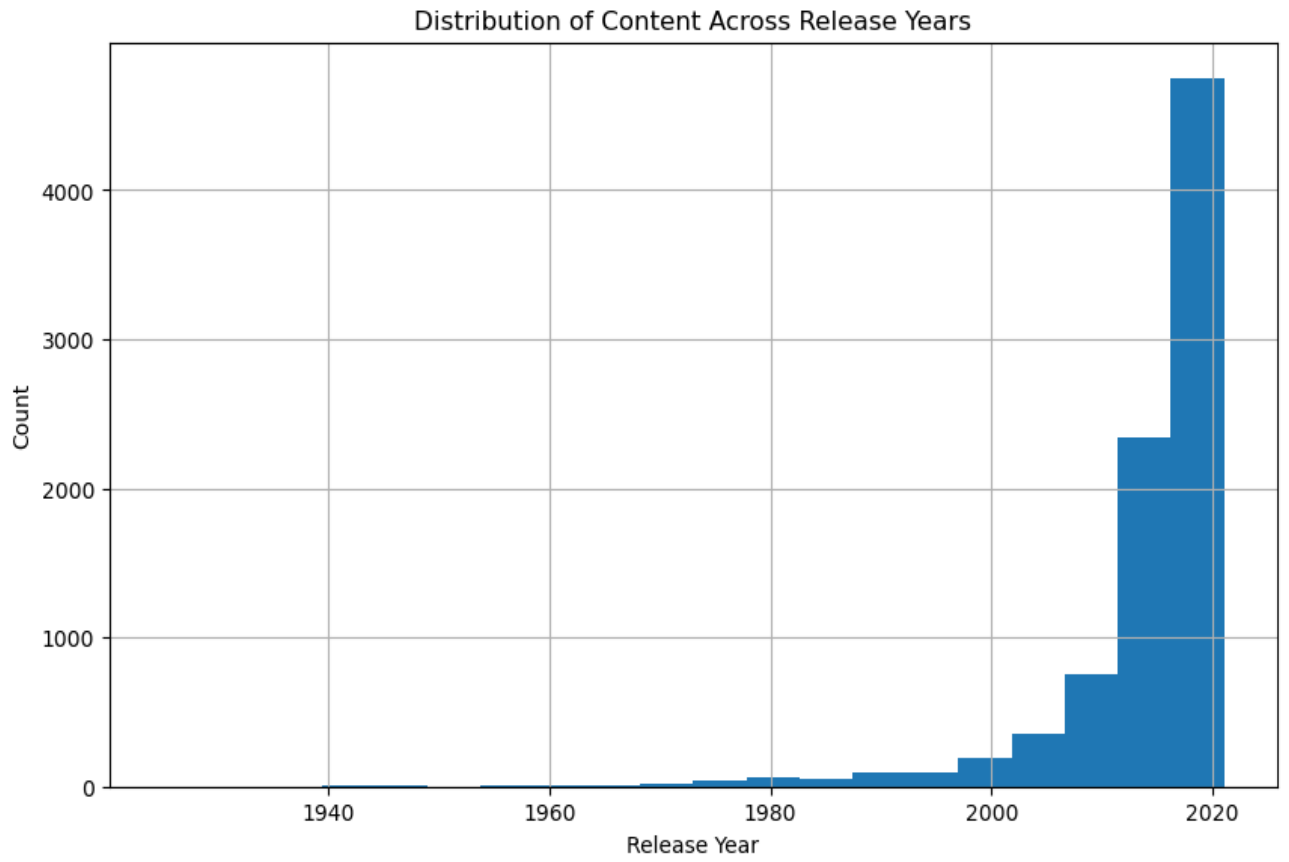


From the Pie Chart we can clearly see that Netflix have more number of movies as campare to TV shows

Distribution of Context Across the Year

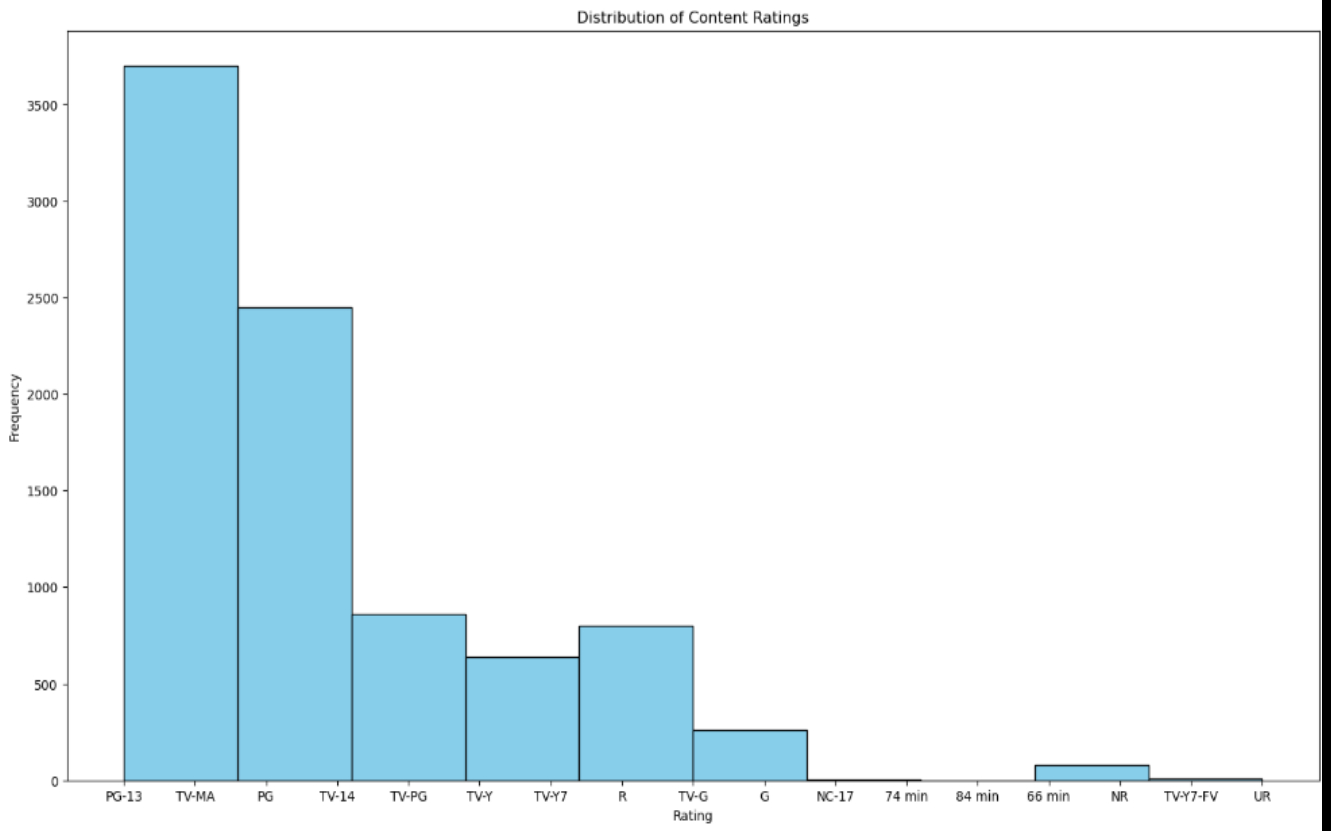
The provided code is used to create a histogram of the 'release_year' column in a DataFrame (df). The histogram is displayed with a title and labels for the x and y axes. Here's a breakdown of the code:
plt.figure(figsize=(10, 6)):
This line creates a new figure with a specified size. The size is set to 10 inches wide and 6 inches tall.
df['release_year'].hist(bins=20):
This line generates a histogram of the 'release_year' column in the DataFrame (df). The bins parameter is set to 20, which means the data will be divided into 20 equal-sized bins.
plt.xlabel('Release Year'):
This line sets the label for the x-axis of the histogram. The label is 'Release Year'.
plt.ylabel('Count'):
This line sets the label for the y-axis of the histogram. The label is 'Count'.
plt.title('Distribution of Content Across Release Years'):
This line sets the title of the histogram. The title is 'Distribution of Content Across Release Years'.
plt.show():
This line displays the histogram.

In [24]:
```
plt.figure(figsize=(10, 6))
df['release_year'].hist(bins=20)
plt.xlabel('Release Year')
plt.ylabel('Count')
plt.title('Distribution of Content Across Release Years')
plt.show()
```



We can see that as time is passing content rate start increasing rapidly

```
plt.hist(df['rating'], color='skyblue', edgecolor='black')
plt.xlabel('Rating')
plt.ylabel('Frequency')
plt.title('Distribution of Content Ratings')
plt.show()
```



Content that has TV-MA rating is most on netflix

Distribution of Content Ratings

- he provided code is used to create a histogram of the 'rating' column in a DataFrame (df). The histogram is displayed with a title and labels for the x and y axes. Here's a breakdown of the code:
- plt.figure(figsize=(18, 10)):
- This line creates a new figure with a specified size. The size is set to 18 inches wide and 10 inches tall.
- plt.hist(df['rating'], color='skyblue', edgecolor='black'):
- This line generates a histogram of the 'rating' column in the DataFrame (df). The color parameter is set to 'skyblue', which means the histogram bars will be filled with this color. The edgecolor parameter is set to 'black', which means the edges of the bars will be black.
- plt.xlabel('Rating'):
- This line sets the label for the x-axis of the histogram. The label is 'Rating'.
- plt.ylabel('Frequency'):
- This line sets the label for the y-axis of the histogram. The label is 'Frequency'.
- plt.title('Distribution of Content Ratings'):
- This line sets the title of the histogram. The title is 'Distribution of Content Ratings'.
- plt.show():
- This line displays the histogram.

Genre Liked by People as the Time Passes

```
In [135]: a=df[df["release_year"]>2000]

In [136]: a
```

Out[136]:

	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	unknown	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	unknown	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...	unknown	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	unknown	unknown	unknown	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	unknown	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...
...	...	...	...	...	...	...	...	...	...	...	...	...
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007	R	158 min	Cult Movies, Dramas, Thrillers	A political cartoonist, a crime reporter and a...

```
38]: b=df[df["release_year"]<2000]

39]: b
```

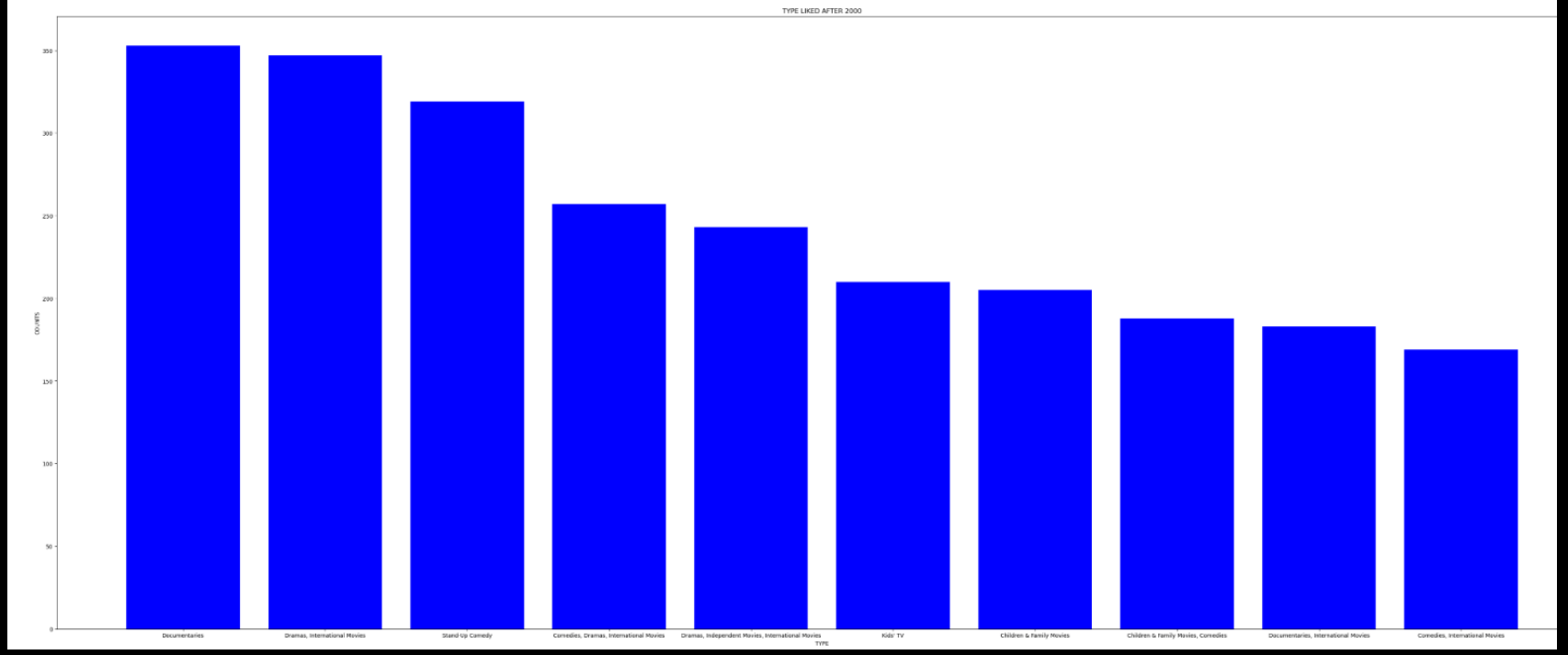
39]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
7	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...	United States, Ghana, Burkina Faso, United Kin...	September 24, 2021	1993	TV-MA	125 min	Dramas, Independent Movies, International Movies	On a photo shoot in Ghana, an American model s...
22	s23	Movie	Avvai Shanmughi	K.S. Ravikumar	Kamal Hassan, Meena, Gemini Ganesan, Heera Raj...	unknown	September 21, 2021	1996	TV-PG	161 min	Comedies, International Movies	Newly divorced and denied visitation rights wi...
24	s25	Movie	Jeans	S. Shankar	Prashanth, Aishwarya Rai Bachchan, Sri Lakshmi...	India	September 21, 2021	1998	TV-14	166 min	Comedies, International Movies, Romantic Movies	When the father of the man she loves insists t...
26	s27	Movie	Minsara Kanavu	Rajiv Menon	Arvind Swamy, Kajol, Prabhu Deva, Nassar, S.P....	unknown	September 21, 2021	1997	TV-PG	147 min	Comedies, International Movies, Music & Musicals	A tangled love triangle ensues when a man fall...
41	s42	Movie	Jaws	Steven Spielberg	Roy Scheider, Robert Shaw, Richard Dreyfuss, L...	United States	September 16, 2021	1975	PG	124 min	Action & Adventure, Classic Movies, Dramas	When an insatiable great white shark terrorize...
...	...	...	...	...	...	...	...	...	...	...	...	...
8745	s8746	Movie	Willy Wonka & the Chocolate Factory	Mel Stuart	Gene Wilder, Jack Albertson, Peter Ostrum, Roy...	United States, East Germany, West Germany	January 1, 2020	1971	G	100 min	Children & Family Movies, Classic Movies, Come...	Zany Willy Wonka causes a stir when he announc...
8748	s8749	Movie	Winter of Our Dreams	John Duigan	Judy Davis, Bryan Brown, Cathy Downes, Baz Luh...	Australia	November 1, 2016	1981	NR	86 min	Classic Movies, Dramas	After the death of a long-ago lover, married p...

Type Liked After 2000

```
plt.figure(figsize=(50,20))
plt.bar(a.listed_in.value_counts().head(10).index,a.listed_in.value_counts().head(10),color="blue")
plt.xlabel("TYPE")
plt.ylabel("COUNTS")
plt.title("TYPE LIKED AFTER 2000")
```

```
Text(0.5, 1.0, 'TYPE LIKED AFTER 2000')
```



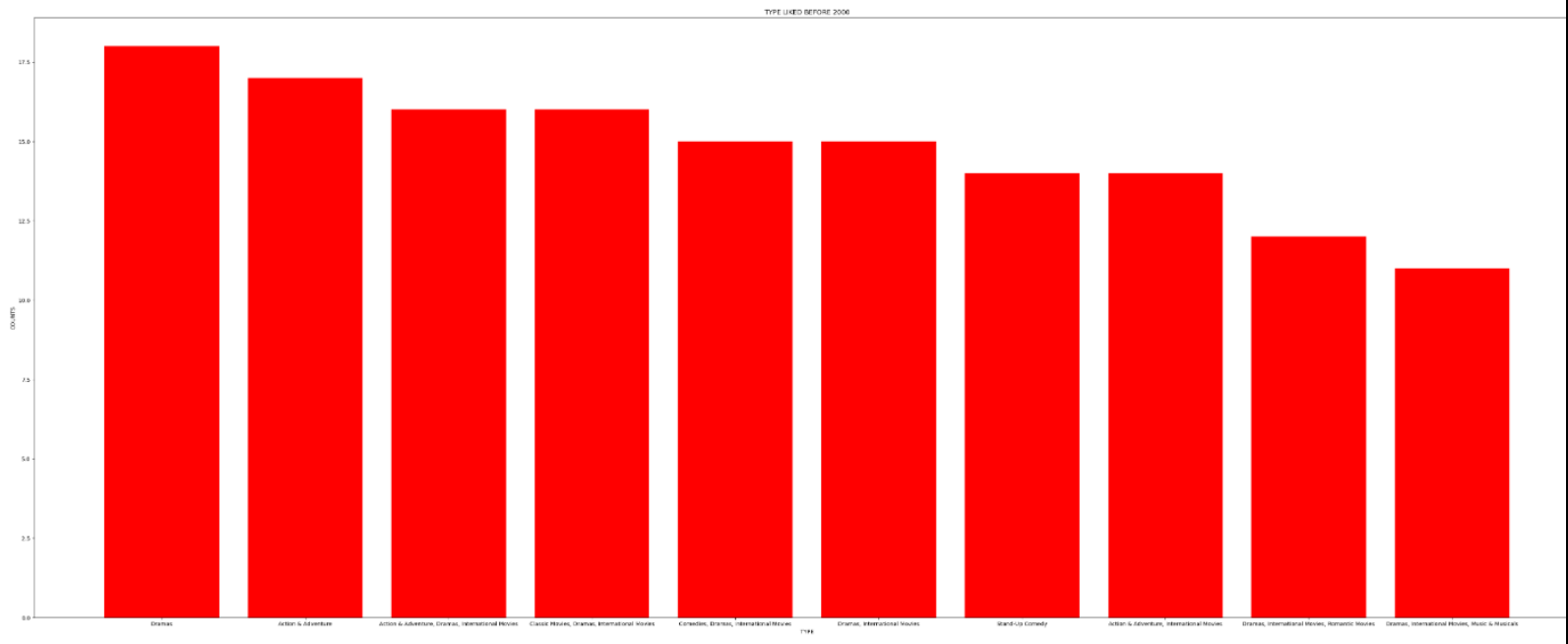
- This code is used to create a bar chart that displays the top 10 types of content liked by people after the year 2000. Here's a breakdown of the code:
- plt.figure(figsize=(50,20)):
- This line sets the figure size to 50 inches wide and 20 inches tall.
- plt.bar(a.listed_in.value_counts().head(10).index, a.listed_in.value_counts().head(10), color="blue"):
- This line creates a bar chart. The x-axis represents the types of content, and the y-axis represents the counts. The value_counts() method is used to count the occurrences of each type. The head(10) method is used to get the top 10 types. The index attribute is used to get the labels for the x-axis. The color parameter is set to "blue" to color the bars blue.
- plt.xlabel("TYPE"):
- This line sets the label for the x-axis to "TYPE".
- plt.ylabel("COUNTS"):
- This line sets the label for the y-axis to "COUNTS".
- plt.title("TYPE LIKED AFTER 2000"):
- This line sets the title of the chart to "TYPE LIKED AFTER 2000".
- plt.show():
- This line displays the chart.
- This code is used to visualize the distribution of types of content liked by people after the year 2000, with the top 10 types displayed on the chart

Type Liked Before 2000

- The provided code is used to create a bar chart that displays the top 10 types of content liked by people before the year 2000. Here's a breakdown of the code:
- plt.figure(figsize=(50,20)):
- This line sets the figure size to 50 inches wide and 20 inches tall.
- plt.bar(b.listed_in.value_counts().head(10).index, b.listed_in.value_counts().head(10), color="red"):
- This line creates a bar chart. The x-axis represents the types of content, and the y-axis represents the counts. The value_counts() method is used to count the occurrences of each type. The head(10) method is used to get the top 10 types. The index attribute is used to get the labels for the x-axis. The color parameter is set to "red" to color the bars red.
- plt.xlabel("TYPE"):
- This line sets the label for the x-axis to "TYPE".
- plt.ylabel("COUNTS"):
- This line sets the label for the y-axis to "COUNTS".
- plt.title("TYPE LIKED BEFORE 2000"):
- This line sets the title of the chart to "TYPE LIKED BEFORE 2000".
- plt.show():
- This line displays the chart.
- This code is used to visualize the distribution of types of content liked by people before the year 2000, with the top 10 types displayed on the chart.

```
]: plt.figure(figsize=(50,20))
   plt.bar(b.listed_in.value_counts().head(10).index,b.listed_in.value_counts().head(10),color="red")
   plt.xlabel("TYPE")
   plt.ylabel("COUNTS")
   plt.title("TYPE LIKED BEFORE 2000")
```

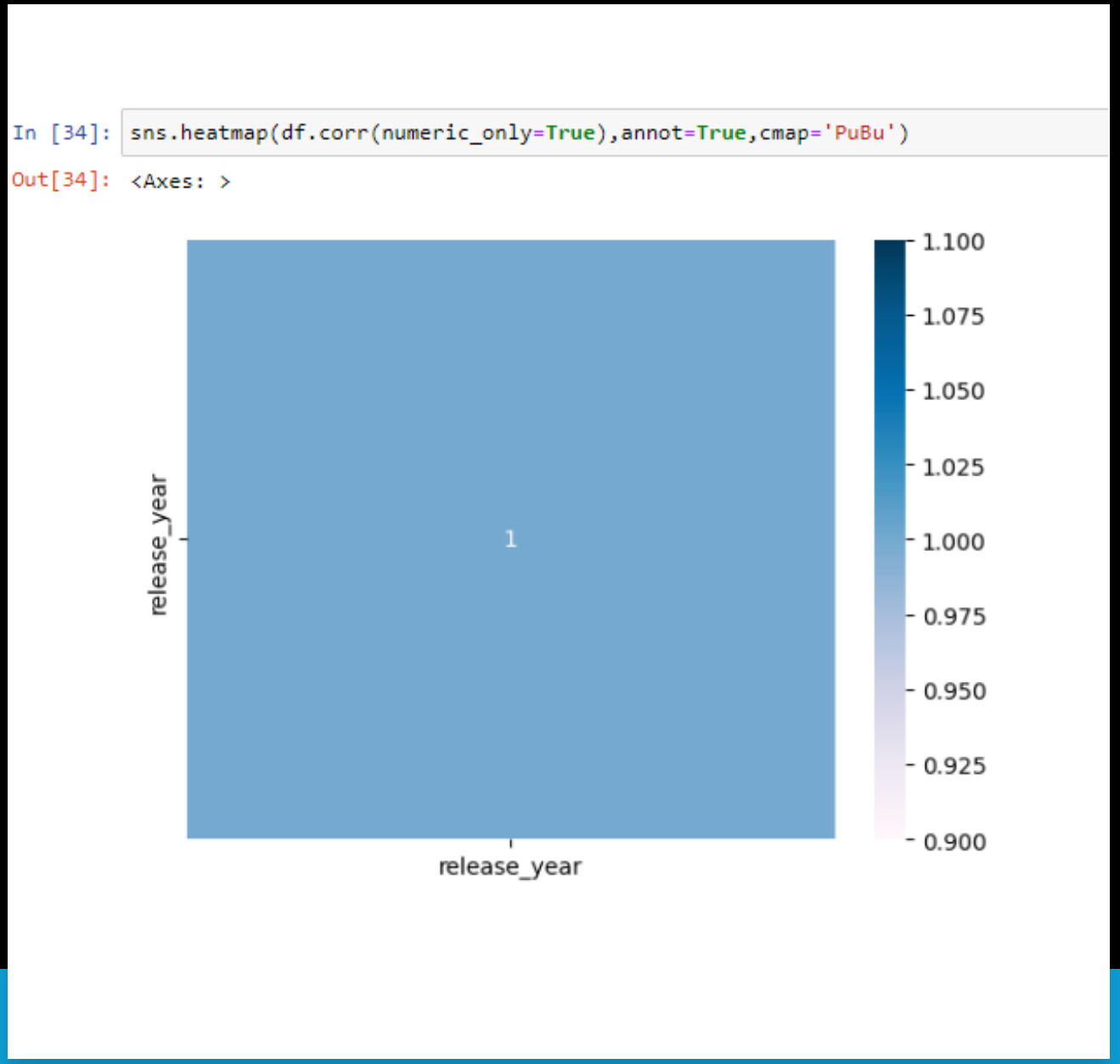
```
]: Text(0.5, 1.0, 'TYPE LIKED BEFORE 2000')
```



As the time passes people start liking Documentries

Correlation

- The provided code is used to create a heatmap using the sns.heatmap() function from the Seaborn library. Here's a breakdown of the code:
- sns.heatmap(df.corr(numeric_only=True), annot=True, cmap='PuBu'):
- This line creates a heatmap of the correlation matrix of the DataFrame df. The corr() method is used to calculate the correlation matrix. The numeric_only=True parameter ensures that only numeric columns are considered.
- The annot=True parameter is used to display the correlation values in each cell of the heatmap.
- The cmap='PuBu' parameter specifies the color map to use for the heatmap. 'PuBu' is a diverging color map that is suitable for visualizing correlation matrices.
- plt.show():
- This line displays the heatmap.
- This code is used to visualize the correlation between the columns of the DataFrame df. The heatmap shows the correlation values in each cell, with darker colors indicating stronger correlations.

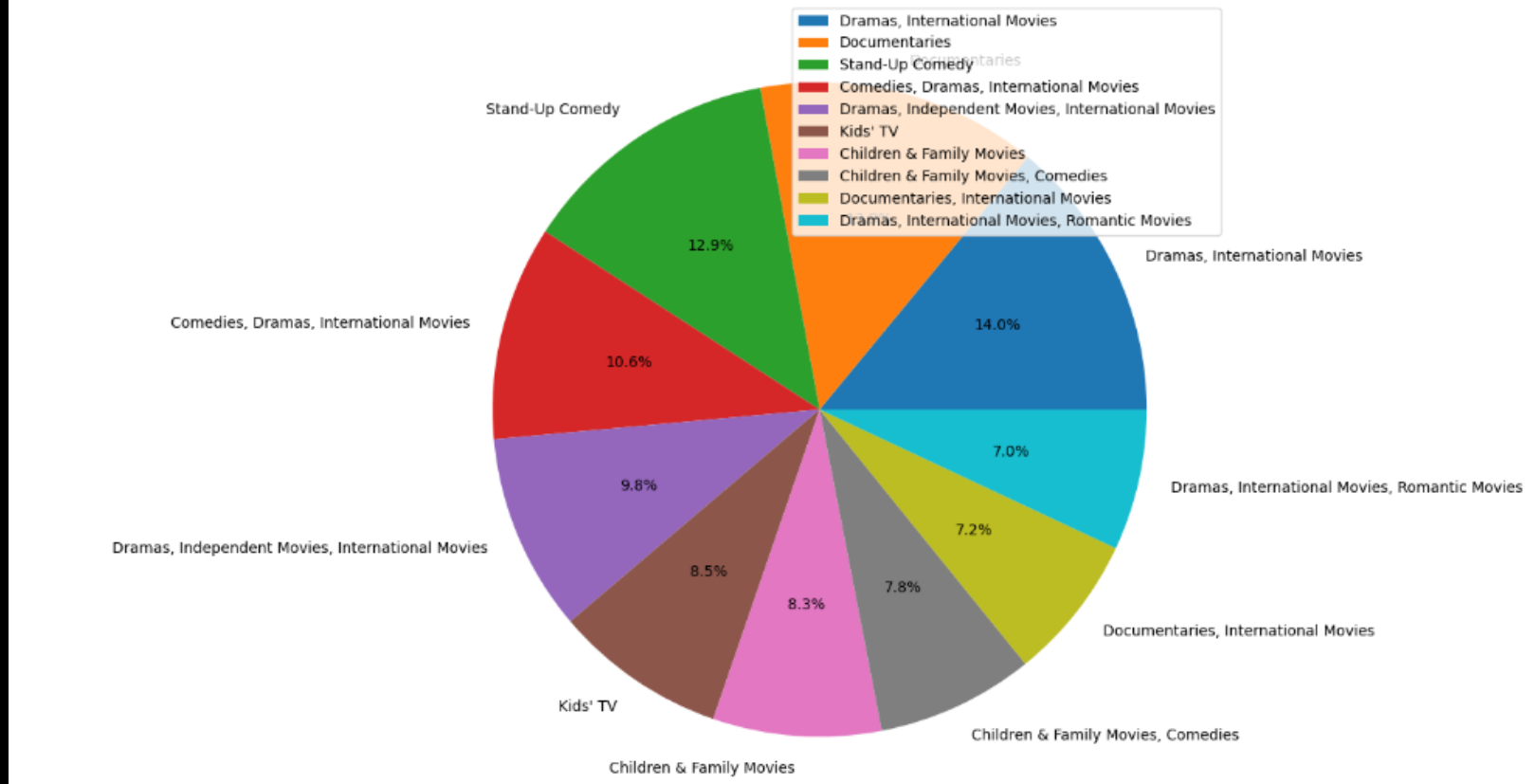


Most Loved Genre

- The provided code is used to create a pie chart using the plt.pie() function from the Matplotlib library. Here's a breakdown of the code:
- plt.figure(figsize=(20,10)):
- This line sets the figure size to 20 inches wide and 10 inches tall.
- plt.pie(df.listed_in.value_counts().head(10), labels=df.listed_in.value_counts().head(10).index, autopct="%1.1f%%"):
- This line creates a pie chart. The value_counts() method is used to count the occurrences of each type in the listed_in column. The head(10) method is used to get the top 10 types. The index attribute is used to get the labels for the pie chart. The autopct parameter is used to format the percentage values displayed on the pie chart. The format string "%1.1f%%" means that the percentage values will be displayed with one decimal place and a percent sign.
- plt.legend():
- This line adds a legend to the pie chart.
- plt.show():
- This line displays the pie chart.
- This code is used to visualize the distribution of the top 10 types in the listed_in column of the DataFrame df. The pie chart shows the proportion of each type, with the percentage values displayed on the chart.

```
n [36]: plt.figure(figsize=(20,10))
        plt.pie(df.listed_in.value_counts().head(10),labels=df.listed_in.value_counts().head(10).index,autopct="%1.1f%%")
        plt.legend()

ut[36]: <matplotlib.legend.Legend at 0x18ef0d62390>
```



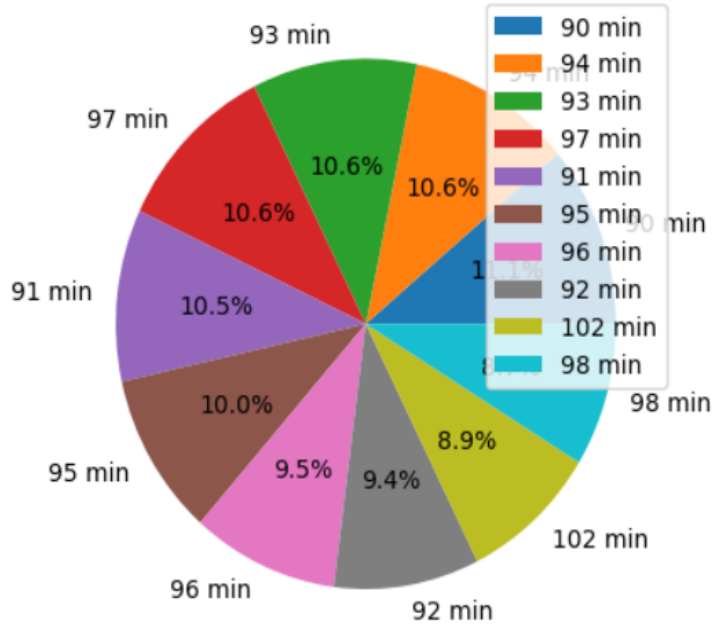
Peple like Dramas ,International movies the most

Duration of Movies

- The provided code is used to create a pie chart using the plt.pie() function from the Matplotlib library. Here's a breakdown of the code:
- plt.pie(movies.duration.value_counts().head(10), labels=movies.duration.value_counts().head(10).index, autopct="%1.1f%%"):
- This line creates a pie chart. The value_counts() method is used to count the occurrences of each duration in the duration column. The head(10) method is used to get the top 10 durations. The index attribute is used to get the labels for the pie chart. The autopct parameter is used to format the percentage values displayed on the pie chart. The format string "%1.1f%%" means that the percentage values will be displayed with one decimal place and a percent sign.
- plt.legend():
- This line adds a legend to the pie chart.
- plt.show():
- This line displays the pie chart.
- This code is used to visualize the distribution of the top 10 durations in the duration column of the DataFrame movies. The pie chart shows the proportion of each duration, with the percentage values displayed on the chart.

```
10]: plt.pie(movies.duration.value_counts().head(10),labels=movies.duration.value_counts().head(10).index,autopct="%1.1f%%")
     plt.legend()

10]: <matplotlib.legend.Legend at 0x18ef0d9c210>
```



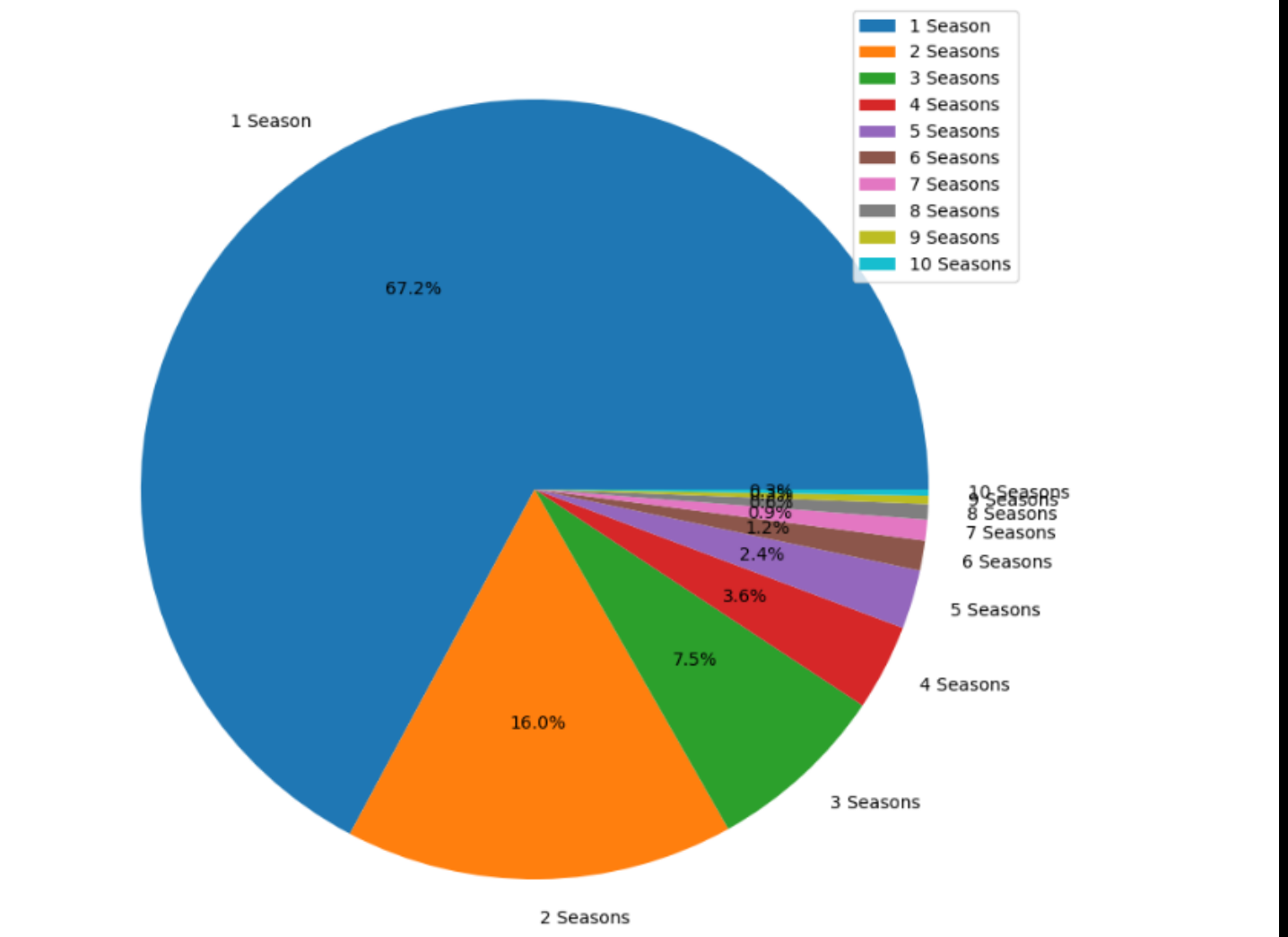
most of the people like to watch movies that is of 90 to 97 mins in duration

Number of Seasons of TV Shows people liked to watch

- The provided code is used to create a pie chart using the plt.pie() function from the Matplotlib library. Here's a breakdown of the code:
- plt.figure(figsize=(20,10)):
- This line sets the figure size to 20 inches wide and 10 inches tall.
- plt.pie(tv_shows.duration.value_counts().head(10), labels=tv_shows.duration.value_counts().head(10).index, autopct="%1.1f%%"):
- This line creates a pie chart. The value_counts() method is used to count the occurrences of each duration in the duration column of the DataFrame tv_shows. The head(10) method is used to get the top 10 durations. The index attribute is used to get the labels for the pie chart. The autopct parameter is used to format the percentage values displayed on the pie chart. The format string "%1.1f%%" means that the percentage values will be displayed with one decimal place and a percent sign.
- plt.legend():
- This line adds a legend to the pie chart.
- plt.show():
- This line displays the pie chart.
- This code is used to visualize the distribution of the top 10 durations in the duration column of the DataFrame tv_shows. The pie chart shows the proportion of each duration, with the percentage values displayed on the chart.

```
plt.figure(figsize=(20,10))
plt.pie(tv_shows.duration.value_counts().head(10),labels=tv_shows.duration.value_counts().head(10).index,autopct="%1.1f%%")
plt.legend()
```

```
<matplotlib.legend.Legend at 0x18ef0b1fc10>
```



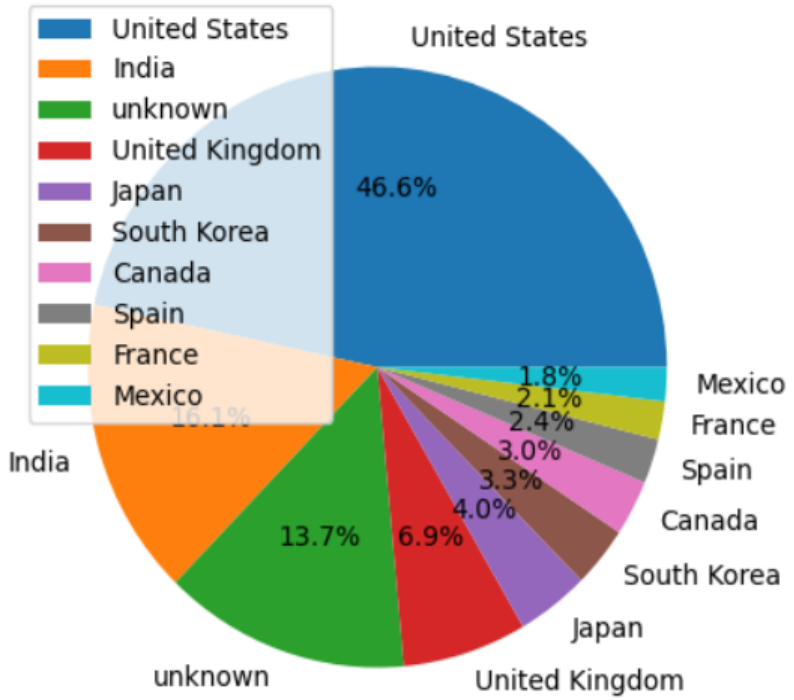
Most of the people like to watch TV Shows that have only 1 season

Country with most Number of Movies and TV shows

- The provided code is used to create a pie chart using the plt.pie() function from the Matplotlib library. Here's a breakdown of the code:
- plt.figure(figsize=(10,5)):
- This line sets the figure size to 10 inches wide and 5 inches tall.
- plt.pie(df.country.value_counts().head(10), labels=df.country.value_counts().head(10).index, autopct="%1.1f%%"):
- This line creates a pie chart. The value_counts() method is used to count the occurrences of each country in the country column of the DataFrame df. The head(10) method is used to get the top 10 countries. The index attribute is used to get the labels for the pie chart. The autopct parameter is used to format the percentage values displayed on the pie chart. The format string "%1.1f%%" means that the percentage values will be displayed with one decimal place and a percent sign.
- plt.legend():
- This line adds a legend to the pie chart.
- plt.show():
- This line displays the pie chart.
- This code is used to visualize the distribution of the top 10 countries in the country column of the DataFrame df. The pie chart shows the proportion of each country, with the percentage values displayed on the chart.

```
plt.figure(figsize=(10,5))
plt.pie(df.country.value_counts().head(10),labels=df.country.value_counts().head(10).index,autopct="%1.1f%%")
plt.legend()
```

```
<matplotlib.legend.Legend at 0x18ef0c3bdd0>
```



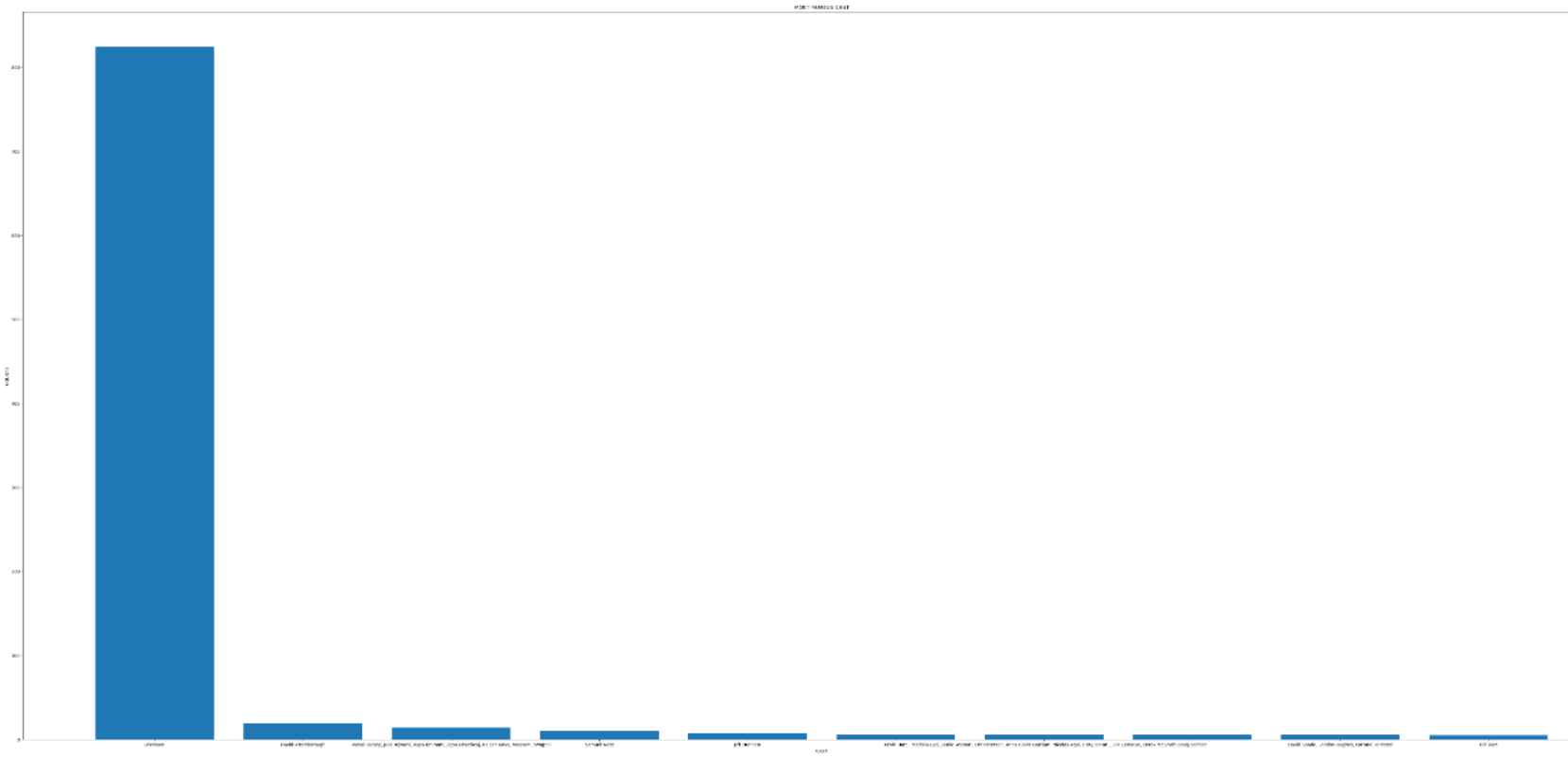
Movies and TV Shows made by the United States are the most on Netflix

Most Famous Cast

- The provided code is used to create a bar chart using the plt.bar() function from the Matplotlib library. Here's a breakdown of the code:
- plt.figure(figsize=(65,30)):
- This line sets the figure size to 65 inches wide and 30 inches tall.
- plt.bar(df.cast.value_counts().head(10).index, df.cast.value_counts().head(10)):
- This line creates a bar chart. The value_counts() method is used to count the occurrences of each cast in the cast column of the DataFrame df. The head(10) method is used to get the top 10 casts. The index attribute is used to get the labels for the x-axis. The y-axis represents the counts.
- plt.xlabel("CAST"):
- This line sets the label for the x-axis to "CAST".
- plt.ylabel("COUNTS"):
- This line sets the label for the y-axis to "COUNTS".
- plt.title("MOST FAMOUS CAST"):
- This line sets the title of the chart to "MOST FAMOUS CAST".
- plt.show():
- This line displays the bar chart.
- This code is used to visualize the distribution of the top 10 casts in the cast column of the DataFrame df. The bar chart shows the count of each cast, with the x-axis representing the casts and the y-axis representing the counts.

```
]: plt.figure(figsize=(65,30))
   plt.bar(df.cast.value_counts().head(10).index,df.cast.value_counts().head(10))
   plt.xlabel("CAST")
   plt.ylabel("COUNTS")
   plt.title("MOST FAMOUS CAST")

]: Text(0.5, 1.0, 'MOST FAMOUS CAST')
```



David Attenborough is most popular cast on netflix as he has most number of Tv Shows and Movies on netflix

Specific Country Data Analysis

UNITED STATES

United States Data Analysis

```
[47]: united_states=df[df.country=="United States"]
```

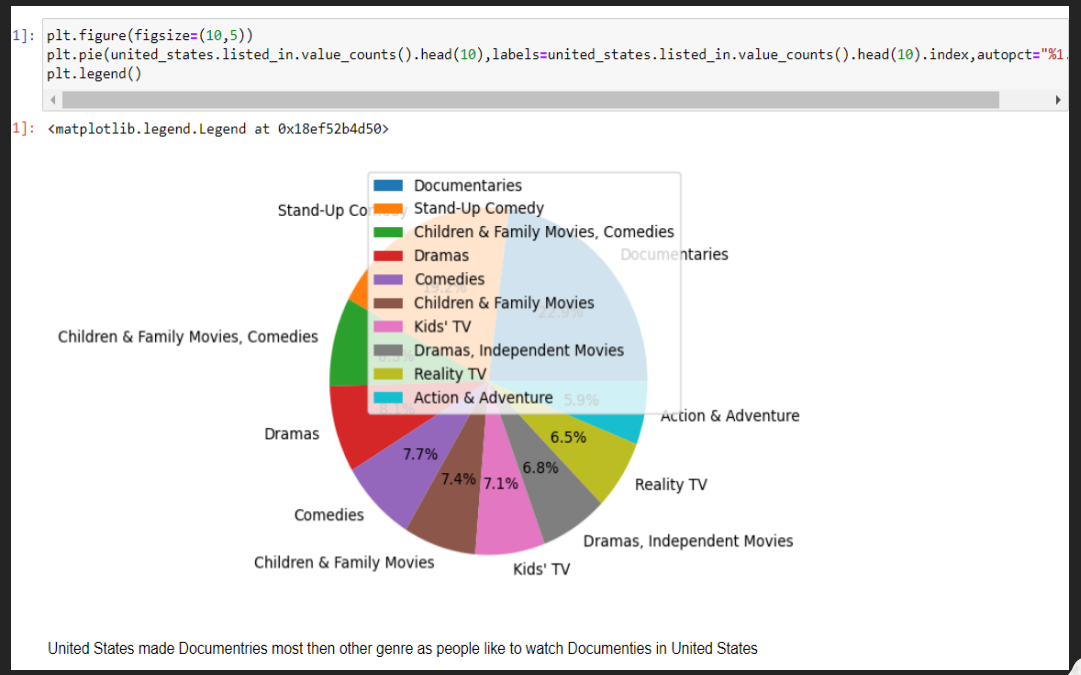
```
[48]: united_states
```

[48]:

	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	unknown	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As her father nears the end of his life, filmm...
9	s10	Movie	The Starling	Theodore Melfi	Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...	United States	September 24, 2021	2021	PG-13	104 min	Comedies, Dramas	A woman adjusting to life after a loss contend...
15	s16	TV Show	Dear White People	unknown	Logan Browning, Brandon P. Bell, DeRon Horton,...	United States	September 22, 2021	2021	TV-MA	4 Seasons	TV Comedies, TV Dramas	Students of color navigate the daily slights a...
27	s28	Movie	Grown Ups	Dennis Dugan	Adam Sandler, Kevin James, Chris Rock, David S...	United States	September 20, 2021	2010	PG-13	103 min	Comedies	Mourning the loss of their beloved junior high...
28	s29	Movie	Dark Skies	Scott Stewart	Keri Russell, Josh Hamilton, J.K. Simmons, Dak...	United States	September 19, 2021	2013	PG-13	97 min	Horror Movies, Sci-Fi & Fantasy	A family's idyllic suburban life shatters when...
...	...	...	...	...	...	...	...	...	...	...	...	...
8791	s8792	Movie	Young Adult	Jason Reitman	Charlize Theron, Patton Oswalt, Patrick Wilson...	United States	November 20, 2019	2011	R	94 min	Comedies, Dramas, Independent Movies	When a divorced writer gets a letter from an o...
8793	s8794	Movie	Yours, Mine and Ours	Raja Gosnell	Dennis Quaid, Rene Russo, Sean Faris, Katija P...	United States	November 20, 2019	2005	PG	88 min	Children & Family Movies, Comedies	When a father of eight and a mother of 10 prep...
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007	R	158 min	Cult Movies, Dramas, Thrillers	A political cartoonist, a crime reporter and a...
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009	R	88 min	Comedies, Horror Movies	Looking to survive in a world taken over by zo...

GENRE

- The provided code is used to create a pie chart using the plt.pie() function from the Matplotlib library. Here's a breakdown of the code:
- plt.figure(figsize=(10,5)):
- This line sets the figure size to 10 inches wide and 5 inches tall.
- plt.pie(united_states.listed_in.value_counts().head(10), labels=united_states.listed_in.value_counts().head(10).index, autopct="%1.1f%%"):
- This line creates a pie chart. The value_counts() method is used to count the occurrences of each listed_in in the listed_in column of the DataFrame united_states. The head(10) method is used to get the top 10 listed_in. The index attribute is used to get the labels for the pie chart. The autopct parameter is used to format the percentage values displayed on the pie chart. The format string "%1.1f%%" means that the percentage values will be displayed with one decimal place and a percent sign.
- plt.legend():
- This line adds a legend to the pie chart.
- plt.show():
- This line displays the pie chart.
- This code is used to visualize the distribution of the top 10 listed_in in the listed_in column of the DataFrame united_states. The pie chart shows the proportion of each listed_in, with the percentage values displayed on the chart.

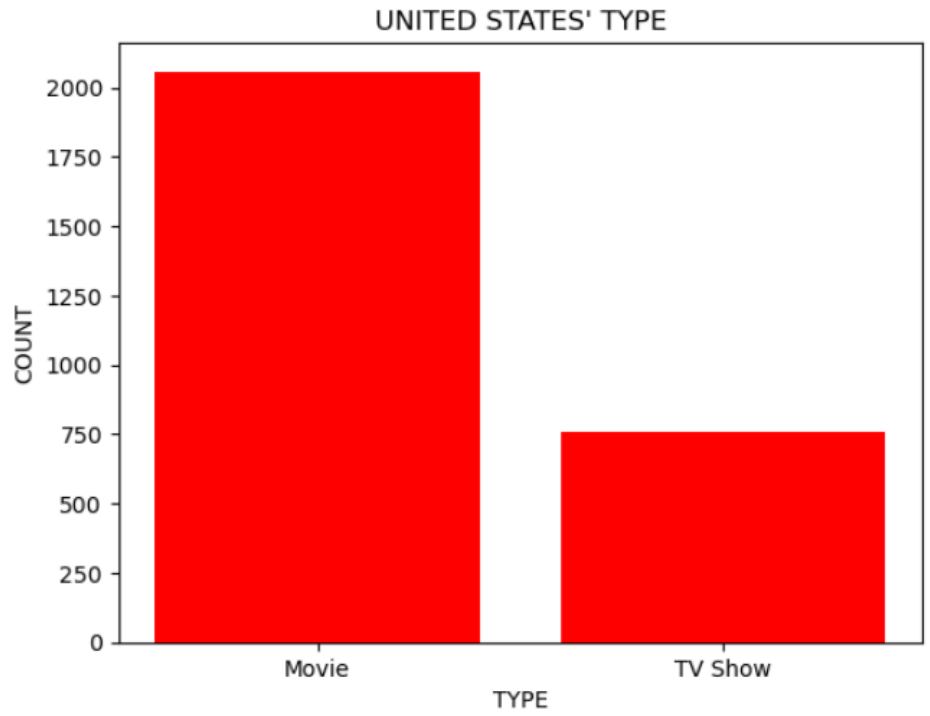


United State'Type

- The provided code is used to create a bar chart using the plt.bar() function from the Matplotlib library. Here's a breakdown of the code:
- plt.bar(united_states.type.value_counts().head(10).index, united_states.type.value_counts().head(10), color="red"):
- This line creates a bar chart. The value_counts() method is used to count the occurrences of each type in the type column of the DataFrame united_states. The head(10) method is used to get the top 10 types. The index attribute is used to get the labels for the x-axis. The y-axis represents the counts. The color parameter is set to "red" to color the bars red.
- plt.xlabel("TYPE"):
- This line sets the label for the x-axis to "TYPE".
- plt.ylabel("COUNT"):
- This line sets the label for the y-axis to "COUNT".
- plt.title("UNITED STATES' TYPE"):
- This line sets the title of the chart to "UNITED STATES' TYPE".
- plt.show():
- This line displays the bar chart.
- This code is used to visualize the distribution of the top 10 types in the type column of the DataFrame united_states. The bar chart shows the count of each type, with the x-axis representing the types and the y-axis representing the counts.
- More on this

```
plt.bar(united_states.type.value_counts().head(10).index,united_states.type.value_counts().head(10),color="red")
plt.xlabel("TYPE")
plt.ylabel("COUNT")
plt.title("UNITED STATES' TYPE")
```

```
Text(0.5, 1.0, "UNITED STATES' TYPE")
```



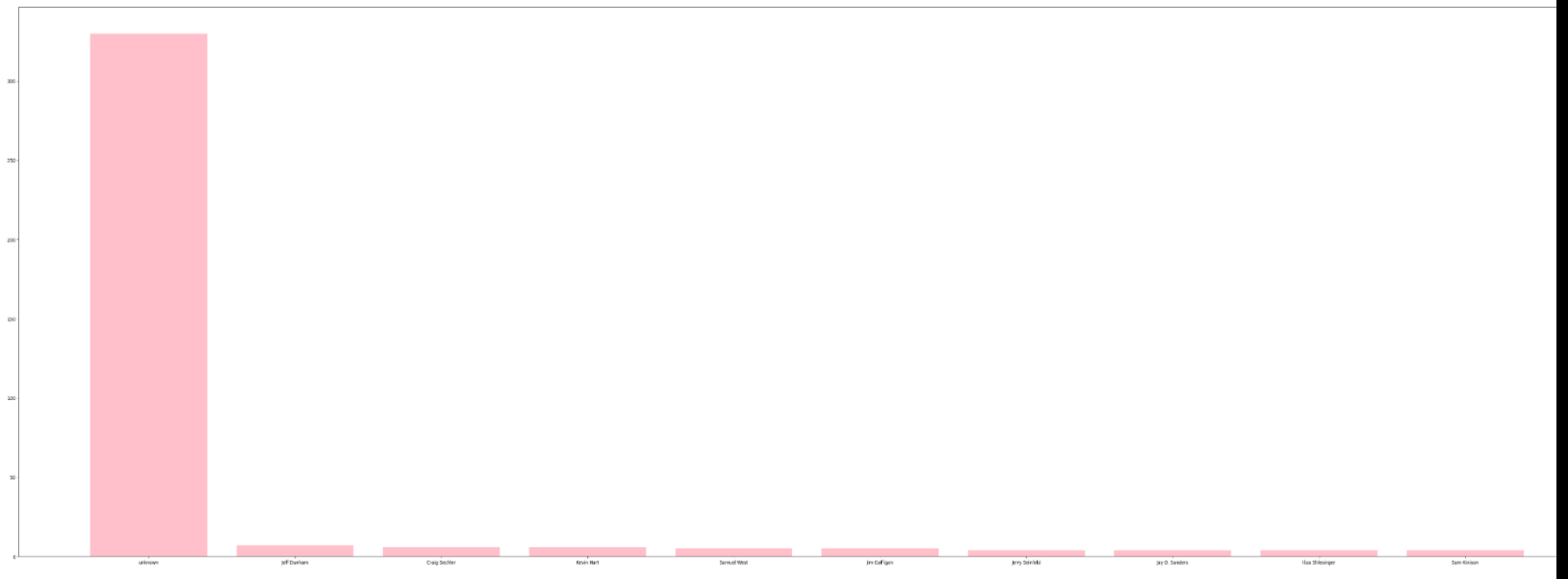
Movies made in United States are more on Netflix than the Tv Shows made by United STates

Most Famous Cast in United States

- The provided code is used to create a bar chart using the plt.bar() function from the Matplotlib library. Here's a breakdown of the code:
- plt.figure(figsize=(65,20)):
- This line sets the figure size to 65 inches wide and 20 inches tall.
- plt.bar(united_states.cast.value_counts().head(10).index, united_states.cast.value_counts().head(10), color="pink"):
- This line creates a bar chart. The value_counts() method is used to count the occurrences of each cast in the cast column of the DataFrame united_states. The head(10) method is used to get the top 10 casts. The index attribute is used to get the labels for the x-axis. The y-axis represents the counts. The color parameter is set to "pink" to color the bars pink.
- plt.show():
- This line displays the bar chart.
- This code is used to visualize the distribution of the top 10 casts in the cast column of the DataFrame united_states. The bar chart shows the count of each cast, with the x-axis representing the casts and the y-axis representing the counts.

```
plt.figure(figsize=(65,20))
plt.bar(united_states.cast.value_counts().head(10).index,united_states.cast.value_counts().head(10),color="pink")
```

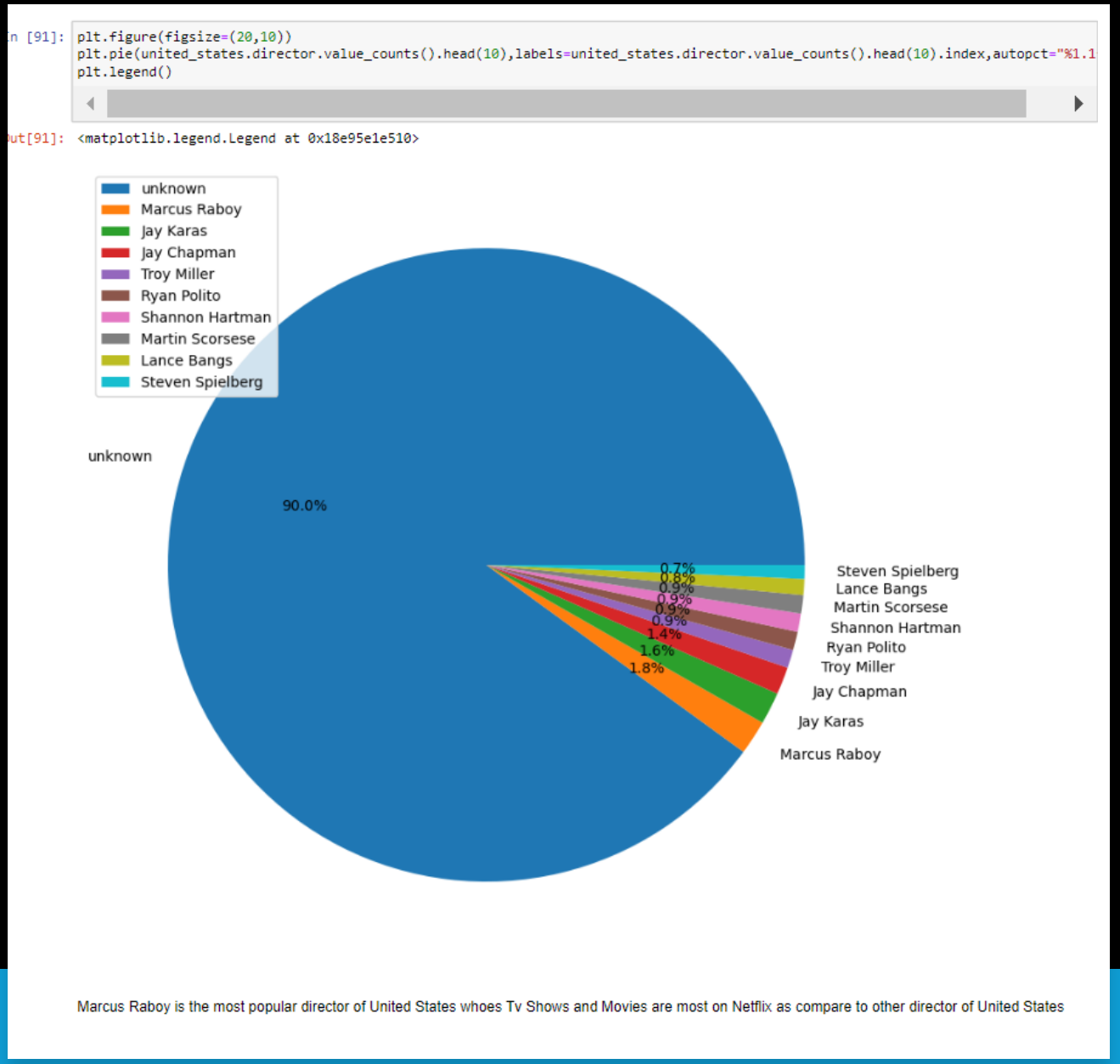
```
<BarContainer object of 10 artists>
```



Jeff Dunham is most popular cast in United States

Most Famous Director In United States

- The provided code is used to create a pie chart using the plt.pie() function from the Matplotlib library. Here's a breakdown of the code:
- plt.figure(figsize=(20,10)):
- This line sets the figure size to 20 inches wide and 10 inches tall.
- plt.pie(united_states.director.value_counts().head(10), labels=united_states.director.value_counts().head(10).index, autopct="%1.1f%%"):
- This line creates a pie chart. The value_counts() method is used to count the occurrences of each director in the director column of the DataFrame united_states. The head(10) method is used to get the top 10 directors. The index attribute is used to get the labels for the pie chart. The autopct parameter is used to format the percentage values displayed on the pie chart. The format string "%1.1f%%" means that the percentage values will be displayed with one decimal place and a percent sign.
- plt.legend():
- This line adds a legend to the pie chart.
- plt.show():
- This line displays the pie chart.
- This code is used to visualize the distribution of the top 10 directors in the director column of the DataFrame united_states. The pie chart shows the proportion of each director, with the percentage values displayed on the chart.



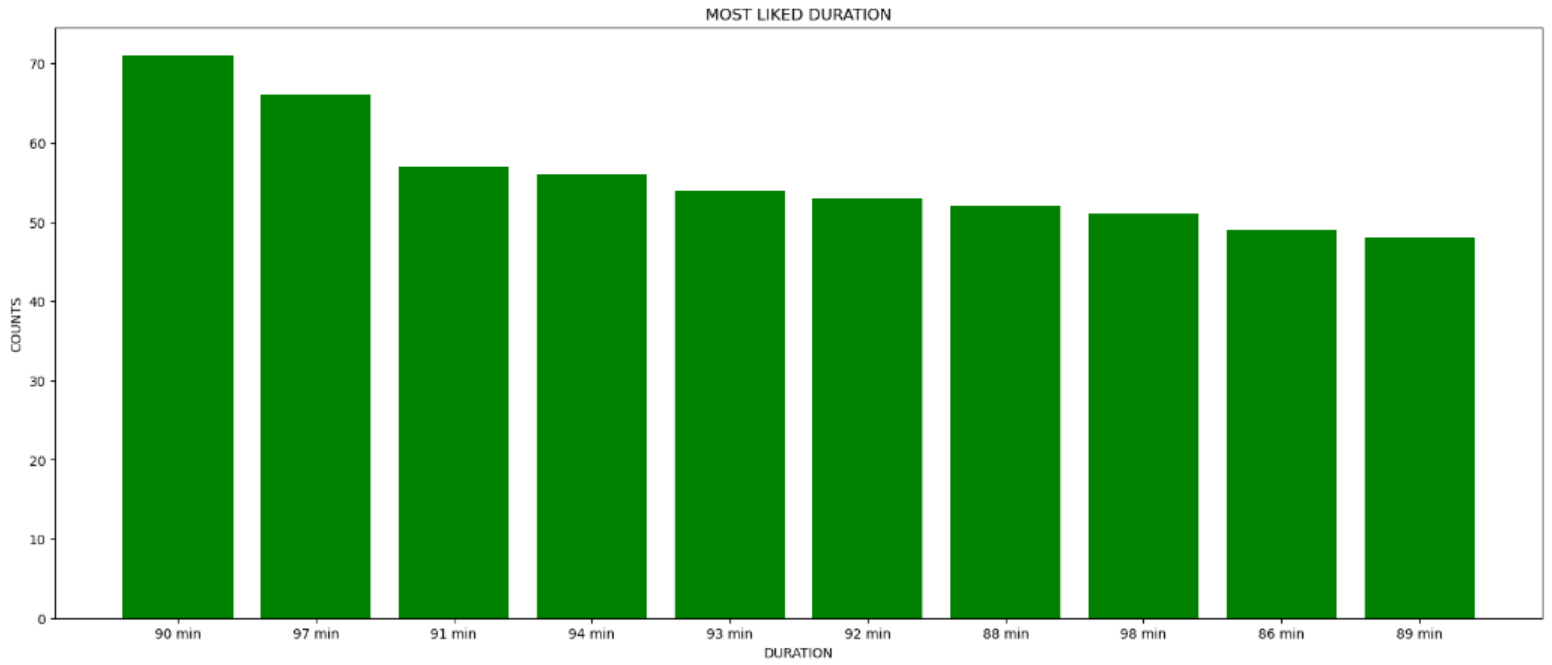
Marcus Raboy is the most popular director of United States whoes Tv Shows and Movies are most on Netflix as compare to other director of United States

Most Liked Duration by the people of United States

- Answer
- The provided code is used to create a bar chart using the plt.bar() function from the Matplotlib library. Here's a breakdown of the code:
- plt.figure(figsize=(20,8)):
- This line sets the figure size to 20 inches wide and 8 inches tall.
- plt.bar(movies_of_united_states.duration.value_counts().head(10).index, movies_of_united_states.duration.value_counts().head(10), color="green"):
- This line creates a bar chart. The value_counts() method is used to count the occurrences of each duration in the duration column of the DataFrame movies_of_united_states. The head(10) method is used to get the top 10 durations. The index attribute is used to get the labels for the x-axis. The y-axis represents the counts. The color parameter is set to "green" to color the bars green.
- plt.xlabel("DURATION"):
- This line sets the label for the x-axis to "DURATION".
- plt.ylabel("COUNTS"):
- This line sets the label for the y-axis to "COUNTS".
- plt.title("MOST LIKED DURATION"):
- This line sets the title of the chart to "MOST LIKED DURATION".
- plt.show():
- This line displays the bar chart.
- This code is used to visualize the distribution of the top 10 durations in the duration column of the DataFrame movies_of_united_states. The bar chart shows the count of each duration, with the x-axis representing the durations and the y-axis representing the counts.

```
plt.figure(figsize=(20,8))
plt.bar(movies_of_united_states.duration.value_counts().head(10).index,movies_of_united_states.duration.value_counts().head(10),
plt.xlabel("DURATION")
plt.ylabel("COUNTS")
plt.title("MOST LIKED DURATION")
```

```
Text(0.5, 1.0, 'MOST LIKED DURATION')
```

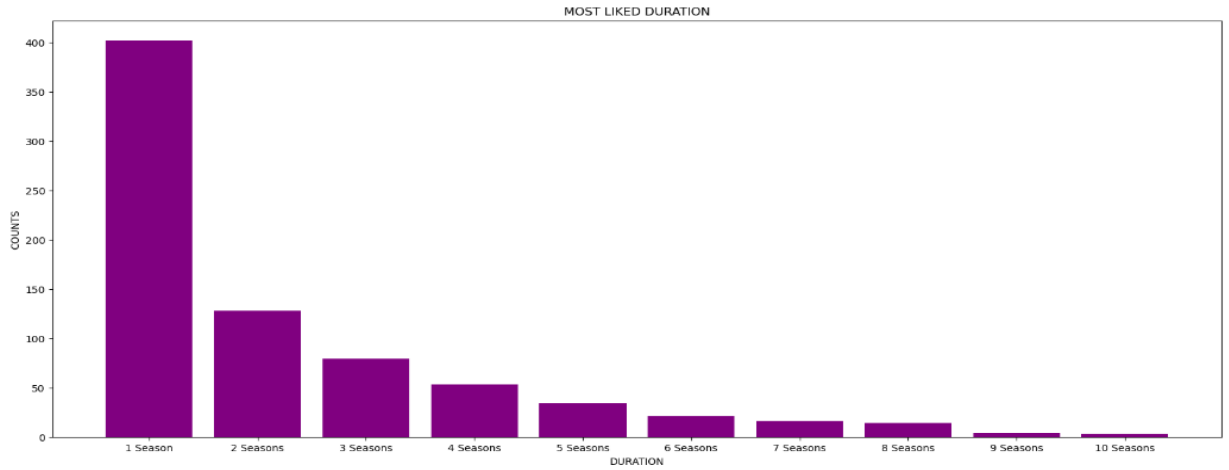


Movies that is of 90-97 minutes are most liked by the people of United States

Number of Seasons Liked to watch by the People of United States

```
plt.figure(figsize=(20,8))
plt.bar(tv_shows_of_united_states.duration.value_counts().head(10).index,tv_shows_of_united_states.duration.value_counts().head(
plt.xlabel("DURATION")
plt.ylabel("COUNTS")
plt.title("MOST LIKED DURATION")
```

```
Text(0.5, 1.0, 'MOST LIKED DURATION')
```



TV Shows that have only 1 season are most liked by the people of United States

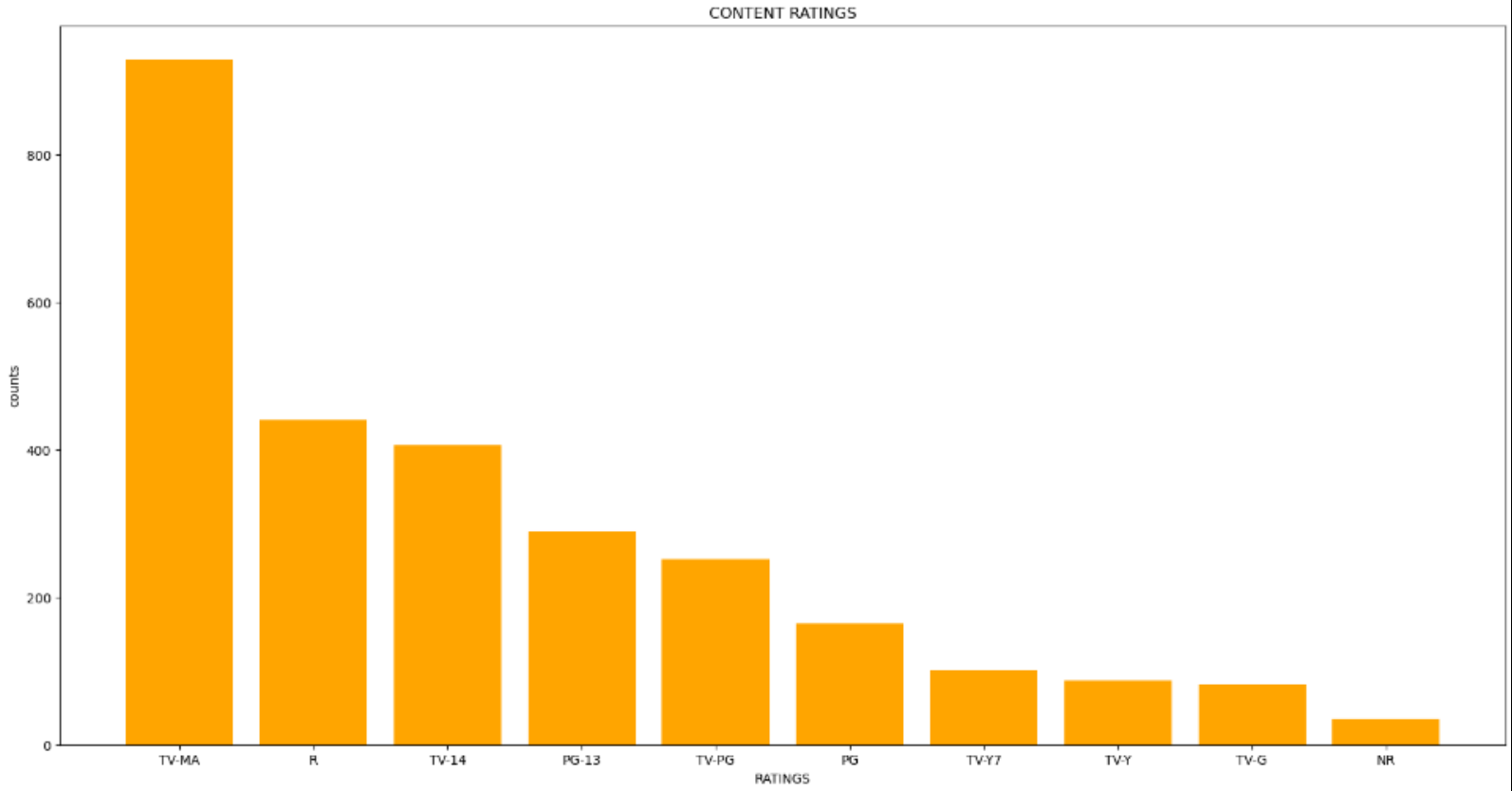
- The provided code is used to create a bar chart using the plt.bar() function from the Matplotlib library. Here's a breakdown of the code:
- plt.figure(figsize=(20,8)):
- This line sets the figure size to 20 inches wide and 8 inches tall.
- plt.bar(tv_shows_of_united_states.duration.value_counts().head(10).index, tv_shows_of_united_states.duration.value_counts().head(10), color="purple"):
- This line creates a bar chart. The value_counts() method is used to count the occurrences of each duration in the duration column of theDataFrame tv_shows_of_united_states. The head(10) method is used to get the top 10 durations. The index attribute is used to get the labels for the x-axis. The y-axis represents the counts. The color parameter is set to "purple" to color the bars purple.
- plt.xlabel("DURATION"):
- This line sets the label for the x-axis to "DURATION".
- plt.ylabel("COUNTS"):
- This line sets the label for the y-axis to "COUNTS".
- plt.title("MOST LIKED DURATION"):
- This line sets the title of the chart to "MOST LIKED DURATION".
- plt.show():
- This line displays the bar chart.
- This code is used to visualize the distribution of the top 10 durations in the duration column of the DataFrame tv_shows_of_united_states. The bar chart shows the count of each duration, with the x-axis representing the

Ratings

- The provided code is used to create a bar chart using the plt.bar() function from the Matplotlib library. Here's a breakdown of the code:
- plt.figure(figsize=(20,10)):
- This line sets the figure size to 20 inches wide and 10 inches tall.
- plt.bar(united_states.rating.value_counts().head(10).index, united_states.rating.value_counts().head(10), color="orange"):
- This line creates a bar chart. The value_counts() method is used to count the occurrences of each rating in the rating column of the DataFrame united_states. The head(10) method is used to get the top 10 ratings. The index attribute is used to get the labels for the x-axis. The y-axis represents the counts. The color parameter is set to "orange" to color the bars orange.
- plt.xlabel("RATINGS"):
- This line sets the label for the x-axis to "RATINGS".
- plt.ylabel("counts"):
- This line sets the label for the y-axis to "counts".
- plt.title("CONTENT RATINGS"):
- This line sets the title of the chart to "CONTENT RATINGS".
- plt.show():
- This line displays the bar chart.
- This code is used to visualize the distribution of the top 10 ratings in the rating column of the DataFrame united_states. The bar chart shows the count of each rating, with the x-axis representing the ratings and the y-axis representing the counts.

```
]: plt.figure(figsize=(20,10))
   plt.bar(united_states.rating.value_counts().head(10).index,united_states.rating.value_counts().head(10),color="orange")
   plt.xlabel("RATINGS")
   plt.ylabel("counts")
   plt.title("CONTENT RATINGS")
```

```
]: Text(0.5, 1.0, 'CONTENT RATINGS')
```



Most of the movies which are made by United States are for mature audience only

Genre Liked by the People as Time Passes

```
158]: b=united_states[united_states["release_year"]<2000]

159]: b
```

159]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
41	s42	Movie	Jaws	Steven Spielberg	Roy Scheider, Robert Shaw, Richard Dreyfuss, L...	United States	September 16, 2021	1975	PG	124 min	Action & Adventure, Classic Movies, Dramas	When an insatiable great white shark terrorize...
42	s43	Movie	Jaws 2	Jeannot Szwarc	Roy Scheider, Lorraine Gary, Murray Hamilton, ...	United States	September 16, 2021	1978	PG	116 min	Dramas, Horror Movies, Thrillers	Four years after the last deadly shark attacks...
43	s44	Movie	Jaws 3	Joe Alves	Dennis Quaid, Bess Armstrong, Simon MacCorkind...	United States	September 16, 2021	1983	PG	98 min	Action & Adventure, Horror Movies, Thrillers	After the staff of a marine theme park try to ...
44	s45	Movie	Jaws: The Revenge	Joseph Sargent	Lorraine Gary, Lance Guest, Mario Van Peebles,...	United States	September 16, 2021	1987	PG-13	91 min	Action & Adventure, Horror Movies, Thrillers	After another deadly shark attack, Ellen Brody...
67	s68	TV Show	Saved by the Bell	unknown	Mark-Paul Gosselaar, Tiffani Thiessen, Mario L...	United States	September 15, 2021	1994	TV-PG	9 Seasons	Kids' TV, TV Comedies	From middle school to college, best friends Za...
...	...	...	...	...	...	...	...	...	...	...	...	...
8735	s8736	Movie	Who's That Knocking at My Door?	Martin Scorsese	Zina Bethune, Harvey Keitel, Anne Collette, Le...	United States	July 1, 2019	1967	R	90 min	Classic Movies, Dramas, Independent Movies	A woman's revelation that she was once raped s...
8739	s8740	Movie	Why We Fight: The Battle of Russia	Frank Capra, Anatole Litvak	unknown	United States	March 31, 2017	1943	TV-PG	82 min	Documentaries	This installment of Frank Capra's acclaimed do...
8742	s8743	Movie	Wild Wild West	Barry Sonnenfeld	Will Smith, Kevin Kline, Kenneth Branagh, Salm...	United States	January 1, 2020	1999	PG-13	106 min	Action & Adventure, Comedies, Sci-Fi & Fantasy	Armed with an ingenious arsenal, two top-notch...
8763	s8764	Movie	WWII: Report from the Aleutians	John Huston	unknown	United States	March 31, 2017	1943	TV-PG	45 min	Documentaries	Filmmaker John Huston narrates this Oscar-nomi...
					Kevin Costner,							Legendary

```
46]: a=united_states[united_states["release_year"]>2000]

47]: a
```

47]:

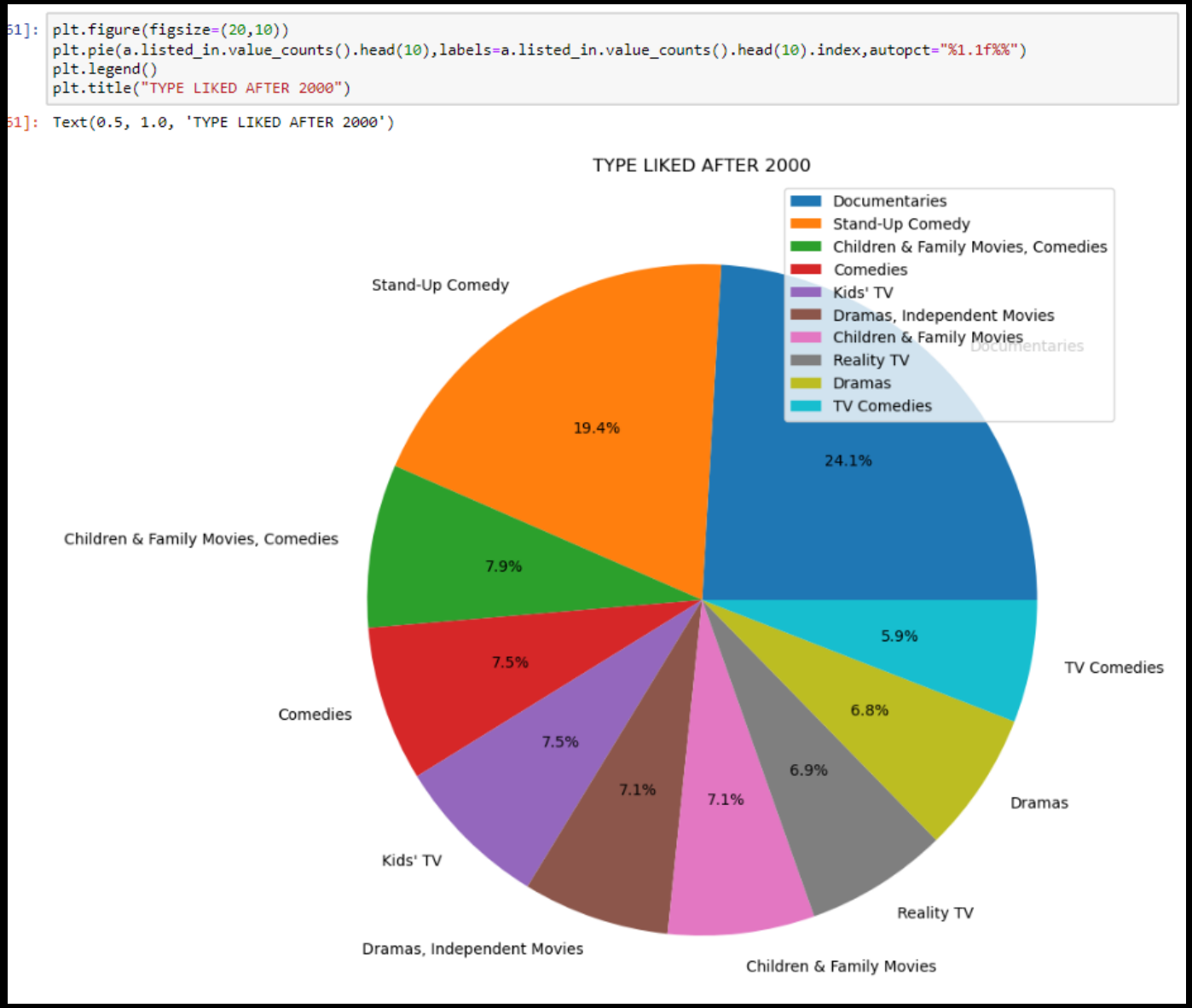
	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	unknown	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As her father nears the end of his life, filmm...
9	s10	Movie	The Starling	Theodore Melfi	Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...	United States	September 24, 2021	2021	PG-13	104 min	Comedies, Dramas	A woman adjusting to life after a loss contend...
15	s16	TV Show	Dear White People	unknown	Logan Browning, Brandon P. Bell, DeRon Horton,...	United States	September 22, 2021	2021	TV-MA	4 Seasons	TV Comedies, TV Dramas	Students of color navigate the daily slights a...
27	s28	Movie	Grown Ups	Dennis Dugan	Adam Sandler, Kevin James, Chris Rock, David S...	United States	September 20, 2021	2010	PG-13	103 min	Comedies	Mourning the loss of their beloved junior high...
28	s29	Movie	Dark Skies	Scott Stewart	Keri Russell, Josh Hamilton, J.K. Simmons, Dak...	United States	September 19, 2021	2013	PG-13	97 min	Horror Movies, Sci-Fi & Fantasy	A family's idyllic suburban life shatters when...
...	...	...	...	...	...	...	...	...	...	...	...	...
8791	s8792	Movie	Young Adult	Jason Reitman	Charlize Theron, Patton Oswalt, Patrick Wilson...	United States	November 20, 2019	2011	R	94 min	Comedies, Dramas, Independent Movies	When a divorced writer gets a letter from an o...
8793	s8794	Movie	Yours, Mine and Ours	Raja Gosnell	Dennis Quaid, Rene Russo, Sean Faris, Katija P...	United States	November 20, 2019	2005	PG	88 min	Children & Family Movies, Comedies	When a father of eight and a mother of 10 prep...
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007	R	158 min	Cult Movies, Dramas, Thrillers	A political cartoonist, a crime reporter and a...
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009	R	88 min	Comedies, Horror Movies	Looking to survive in a world taken over by zo...

Type Liked After 2000

- he provided code is used to create a pie chart using the plt.pie() function from the Matplotlib library. Here's a breakdown of the code:
- plt.figure(figsize=(20,10)):
- This line sets the figure size to 20 inches wide and 10 inches tall.
- plt.pie(a.listed_in.value_counts().head(10), labels=a.listed_in.value_counts().head(10).index, autopct="%1.1f%%"):
- This line creates a pie chart. The value_counts() method is used to count the occurrences of each type in the listed_in column of the DataFrame a. The head(10) method is used to get the top 10 types. The index attribute is used to get the labels for the pie chart. The autopct parameter is used to format the percentage values displayed on the pie chart. The format string "%1.1f%%" means that the percentage values will be displayed with one decimal place and a percent sign.
- plt.legend():
- This line adds a legend to the pie chart.
- plt.title("TYPE LIKED AFTER 2000"):
- This line sets the title of the chart to "TYPE LIKED AFTER 2000".
- plt.show():
- This line displays the pie chart.
- This code is used to visualize the distribution of the top 10 types in the listed_in column of the DataFrame a, with the pie chart showing the proportion of each type, and the percentage values displayed on the chart.

```
51]: plt.figure(figsize=(20,10))
     plt.pie(a.listed_in.value_counts().head(10),labels=a.listed_in.value_counts().head(10).index,autopct="%1.1f%%")
     plt.legend()
     plt.title("TYPE LIKED AFTER 2000")

51]: Text(0.5, 1.0, 'TYPE LIKED AFTER 2000')
```

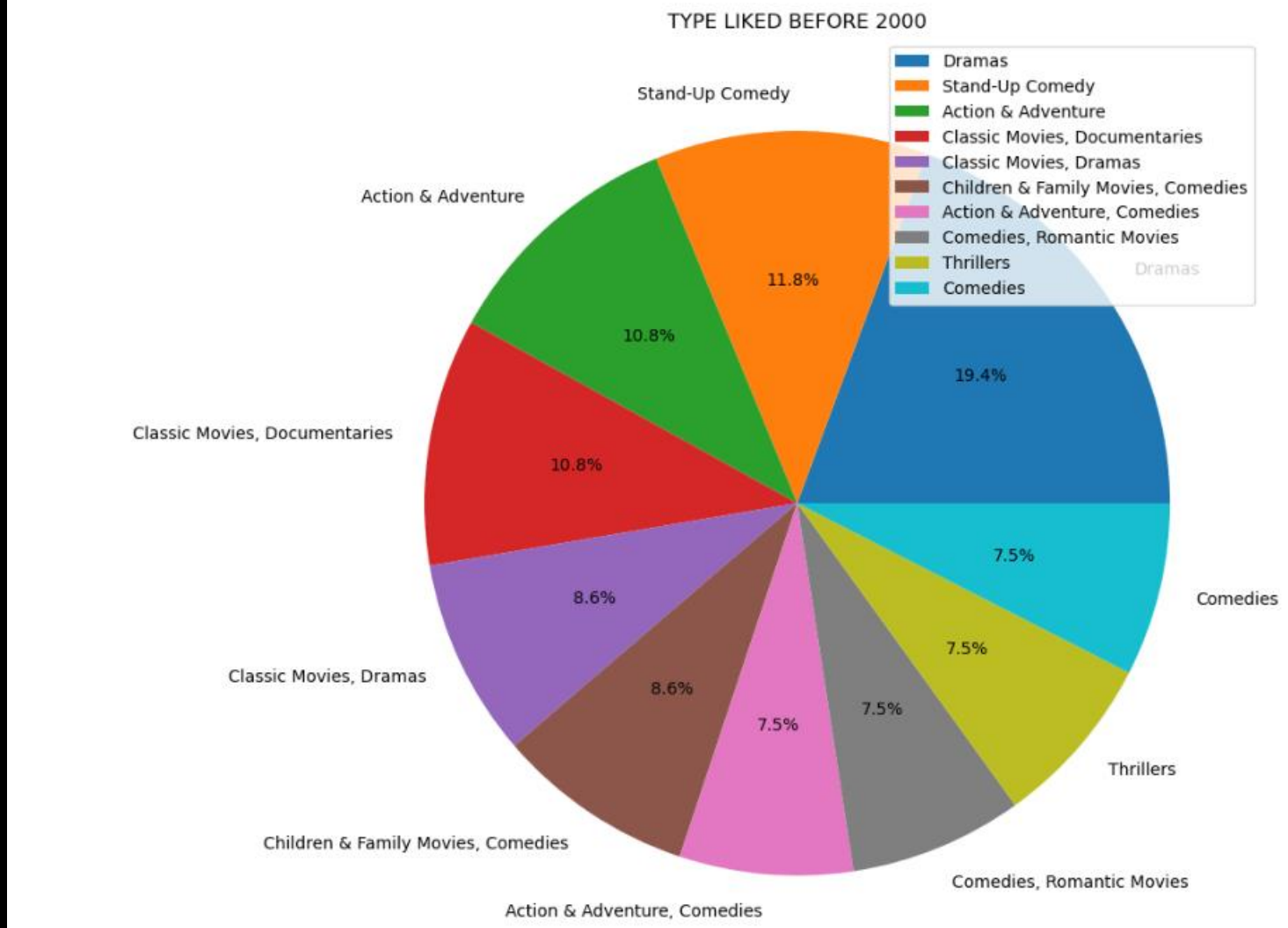


Type liked before 2000

- The provided code is used to create a pie chart using the plt.pie() function from the Matplotlib library. Here's a breakdown of the code:
- plt.figure(figsize=(20,10)):
- This line sets the figure size to 20 inches wide and 10 inches tall.
- plt.pie(b.listed_in.value_counts().head(10), labels=b.listed_in.value_counts().head(10).index, autopct="%1.1f%%"):
- This line creates a pie chart. The value_counts() method is used to count the occurrences of each type in the listed_in column of the DataFrame b. The head(10) method is used to get the top 10 types. The index attribute is used to get the labels for the pie chart. The autopct parameter is used to format the percentage values displayed on the pie chart. The format string "%1.1f%%" means that the percentage values will be displayed with one decimal place and a percent sign.
- plt.legend():
- This line adds a legend to the pie chart.
- plt.title("TYPE LIKED BEFORE 2000"):
- This line sets the title of the chart to "TYPE LIKED BEFORE 2000".
- plt.show():
- This line displays the pie chart.
- This code is used to visualize the distribution of the top 10 types in the listed_in column of the DataFrame b, with the pie chart showing the proportion of each type, and the percentage values displayed on the chart.

```
In [163]: plt.figure(figsize=(20,10))
          plt.pie(b.listed_in.value_counts().head(10),labels=b.listed_in.value_counts().head(10).index,autopct="%1.1f%%")
          plt.legend()
          plt.title("TYPE LIKED BEFORE 2000")

Out[163]: Text(0.5, 1.0, 'TYPE LIKED BEFORE 2000')
```



From the above graphs we can clearly see that before 2000 peple like dramas but after 2000 people of United States like to watch Documentries

Conclusion

- Rajiv Chilaka is the popular director as he has most number of TV Shows and Movies on Netflix.
- Netflix have more number of movies as campare to TV shows.
- Content that has TV-MA rating is most on netflix.
- As the time passes people start liking Documentries.
- Peple like Dramas ,International movies the most all over the world.
- Most of the people like to watch movies that is of 90 to 97 mins in duration.
- Most of the people like to watch TV Shows that have only 1 season.
- Movies and TV Shows made by the United States are the most on Netflix.
- David Attenborough is most popular cast on netflix as he has most number of Tv Shows and Movies on netflix.

Conclusion on United States Data Analysis

- United States made Documentries the most then other genre as people like to watch Documenties in United States.
- Movies made in United States are more on Netflix than the Tv Shows made by United States.
- Marcus Raboy is the most popular director of United States whoes Tv Shows and Movies are most on Netflix as compare to other director of United States.
- Movies that is of 90-97 minutes are most liked by the people of United States.
- TV Shows that have only 1 season are most liked by the people of United States.
- Most of the movies which are made by United States are for mature audience only(TV-MA).
- before 2000 peple like dramas but after 2000 people of United States like to watch Documentries

This project structure provides a comprehensive framework for understanding and improving content on Netflix. Directors and Netflix users can take benefit from these recommendations to enhance their overall experience and engagement within the platform.