

NAME: Krishna Kushwah

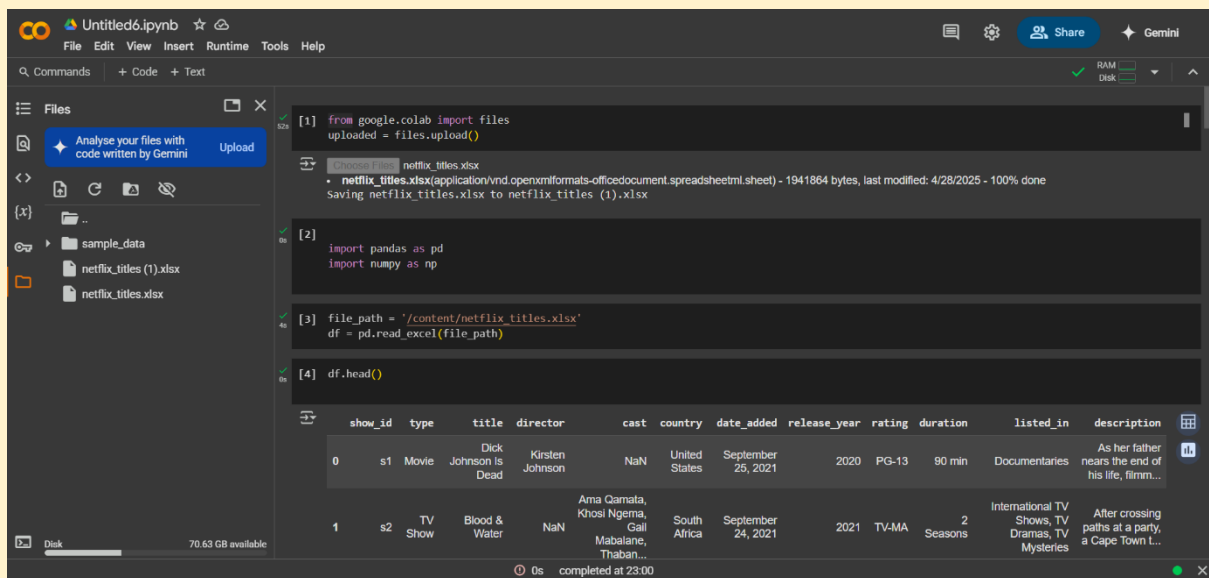
Div :CS7

ROLL NO :CS7_56

PRN:202401110039

Output On Colab link:

https://colab.research.google.com/drive/1uUoOFvmm9LMeeulCvttYLfSlSQ_49hek?usp=sharing



The screenshot shows a Google Colab notebook titled 'Untitled6.ipynb'. The interface includes a menu bar (File, Edit, View, Insert, Runtime, Tools, Help), a search bar, and a 'Share' button. On the left, a 'Files' sidebar shows an uploaded file 'netflix_titles.xlsx' (1941864 bytes) and a folder 'sample_data'. The main code area contains four cells:

```
[1] from google.colab import files
    uploaded = files.upload()

netflix_titles.xlsx (application/vnd.openxmlformats-officedocument.spreadsheetml.sheet) - 1941864 bytes, last modified: 4/28/2025 - 100% done
Saving netflix_titles.xlsx to netflix_titles (1).xlsx

[2] import pandas as pd
    import numpy as np

[3] file_path = '/content/netflix_titles.xlsx'
    df = pd.read_excel(file_path)

[4] df.head()
```

The output of the fourth cell is a table with 12 columns: show_id, type, title, director, cast, country, date_added, release_year, rating, duration, listed_in, and description. The first two rows of data are visible:

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 Camata, Khosi Ngema, Gail Mababalane, 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 L...

The bottom status bar indicates '0s completed at 23:00'.

0s [4] 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 Mablane, 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...

Next steps: [Generate code with df](#) [View recommended plots](#) [New interactive sheet](#)

English (United States)
English (India)

0s [7] # Find the total number of Movies and TV Shows.
df['type'].value_counts()

type	count
Movie	6131
TV Show	2676

dtype: int64

0s #Find the number of unique countries represented.
df['country'].nunique()

748

0s [9] #Find the top 5 countries with the most content on Netflix.
df['country'].value_counts().head(5)

country	count
United States	2818
India	972
United Kingdom	419
Japan	245
South Korea	199

dtype: int64

0s # List all the unique ratings available.
df['rating'].unique()

array(['PG-13', 'TV-MA', 'PG', 'TV-14', 'TV-PG', 'TV-Y', 'TV-Y7', 'R', 'TV-G', 'G', 'NC-17', '74 min', '84 min', '66 min', 'NR', nan, 'TV-Y7-FV', 'UR'], dtype=object)

```

0s [10] #Find the number of missing values in each column.
df.isnull().sum()

0
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

```

```

0s [12] #Find the most common director on Netflix.
df['director'].value_counts().idxmax()

'Rajiv Chilaka'

0s [13] #Find the number of titles with no listed director.
df['director'].isnull().sum()

np.int64(2634)

0s [14] #. Find the average duration of movies (in minutes).
movies = df[df['type'] == 'Movie']
movies['duration_min'] = movies['duration'].str.replace(' min', '').astype(float)
movies['duration_min'].mean()


<ipython-input-14-88c49ab49811>:3: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
movies['duration_min'] = movies['duration'].str.replace(' min', '').astype(float)
np.float64(99.57718668407311)


```

0s  #Find the distribution of TV show seasons (1 Season, 2 Seasons, etc.).


```
tv_shows = df[df['type'] == 'TV Show']
tv_shows['duration'].value_counts()
```

 count

duration	count
1 Season	1793
2 Seasons	425
3 Seasons	199
4 Seasons	95
5 Seasons	65
6 Seasons	33
7 Seasons	23
8 Seasons	17
9 Seasons	9
10 Seasons	7
13 Seasons	3
12 Seasons	2
15 Seasons	2
11 Seasons	2

0s [16]  #Find the number of shows in each genre (listed_in).

```
genre_count = df['listed_in'].str.split(' ', expand=True).stack().value_counts()
genre_count
```

 count

International Movies	2752
Dramas	2427
Comedies	1674
International TV Shows	1351
Documentaries	869
Action & Adventure	859
TV Dramas	763
Independent Movies	756
Children & Family Movies	641
Romantic Movies	616
TV Comedies	581
Thrillers	577
Crime TV Shows	470
Kids' TV	451
Docuseries	395

✓ [19] #Find the top 5 actors who appeared the most.
actors = df['cast'].dropna().str.split(', ', expand=True).stack().value_counts()
actors.head(5)

↕

	count
Anupam Kher	43
Shah Rukh Khan	35
Julie Tejwani	33
Takahiro Sakurai	32
Naseeruddin Shah	32

dtype: int64

✓ [21] #Find all movies/shows directed by 'Christopher Nolan'
df[df['director'] == 'Christopher Nolan'][['title', 'type']]

↕

title	type
340 Inception	Movie

✓ [22] #Find the number of movies released after 2015
df[(df['type'] == 'Movie') & (df['release_year'] > 2015)].shape[0]

↕ 3619

✓ [23] #Find the show with the longest title (in terms of characters)
df['title'].loc[df['title'].str.len().idxmax()]

↕ 'Jim & Andy: The Great Beyond - Featuring a Very Special, Contractually Obligated Mention of Tony Clifton'

✓ [24] #Find the most common year when content was released.
df['release_year'].value_counts().idxmax()

↕ np.int64(2018)

+ Code + Text

✓ [25] #Find how many Movies and TV Shows are available for each Rating (e.g., PG-13, TV-MA).
df.groupby(['rating', 'type']).size().unstack(fill_value=0)

↕

rating	Movie	TV Show
66 min	1	0
74 min	1	0
84 min	1	0
G	41	0
NC-17	3	0
NR	75	5