

Data Analyst – Internship

Task 1: Netflix Movies and TV Shows Analysis

```
Python Console x +
NameError: name 'pd' is not defined
In [3]: import numpy as np
In [4]: import pandas as pd
In [5]: import matplotlib.pyplot as plt
In [6]: import seaborn as sns
Backend tkagg is interactive backend. Turning interactive mode on.
In [7]: df = pd.read_csv("/kaggle/input/netflix-shows/netflix_titles.csv")

NameError: name 'df' is not defined
In [9]: df = pd.read_csv(r"C:\Users\lenovo\Downloads\netflix_titles.csv")
Traceback (most recent call last): @ Explain with AI
File "C:\Users\lenovo\PyCharmMiscProject\.venv\Lib\site-packages\IPython\core\interactiveshell.py", line 3701, in
In [18]:
```

For Resolving Error:

I have used ChatGPT for help, and then the issue was resolved. I was using pycharm Tool for completing task, and the dataset that I was mapping with was on cloud dataset. Then I've downloaded the dataset from Kaggle. The dataset then got extracted.

```
Python Console x +
C:\Users\lenovo\PyCharmMiscProject\.venv\Scripts\python.exe "C:/Program Files/Jet
import sys; print('Python %s on %s' % (sys.version, sys.platform))
sys.path.extend(['C:\\Users\\lenovo\\PyCharmMiscProject'])

Python Console
...: import pandas as pd
...: import matplotlib.pyplot as plt
...: from matplotlib.pyplot import ylabel
...: from matplotlib.pyplot import title
...: from matplotlib.pyplot import xlabel
...: from matplotlib.pyplot import show
...: import seaborn as sns
Backend tkagg is interactive backend. Turning interactive mode on.
```

```
Python Console x +
show_id ... description
0 s1 ... As her father nears the end of his life, filmm...
1 s2 ... After crossing paths at a party, a Cape Town t...
2 s3 ... To protect his family from a powerful drug lor...
3 s4 ... Feuds, flirtations and toilet talk go down amo...
4 s5 ... In a city of coaching centers known to train I...

In [18]:
```

df.info()

```
Python Console x +
In [13]: df.info
Out[13]:
<bound method DataFrame.info of      show_id  ...
0      s1  ...  As her father nears the end of his life, filmm...
1      s2  ...  After crossing paths at a party, a Cape Town t...
2      s3  ...  To protect his family from a powerful drug lor...
3      s4  ...  Feuds, flirtations and toilet talk go down amo...
In [18]:
```

```
Python Console x +
[8807 rows x 12 columns]>
In [14]: df.shape
Out[14]: (8807, 12)
In [15]: df.columns
Out[15]:
Index(['show_id', 'type', 'title', 'director', 'cast', 'country', 'date_added',
      'release_year', 'rating', 'duration', 'listed_in', 'description'],
      dtype='str')
In [16]: df.sample(10)
Out[16]:
      show_id  ...      description
8049  s8050  ...  This biographical drama about Indonesia's firs...
5976  s5977  ...  Vignettes of the summer holidays follow vacati...
7134  s7135  ...  From his star turn at Woodstock to his battles...
In [18]:
```

```
Python Console x +
2796  s2797  ...  An aging soccer fanatic faces down the reality...
[10 rows x 12 columns]
In [17]: df.head()
Out[17]:
      show_id  ...      description
0      s1  ...  As her father nears the end of his life, filmm...
1      s2  ...  After crossing paths at a party, a Cape Town t...
2      s3  ...  To protect his family from a powerful drug lor...
3      s4  ...  Feuds, flirtations and toilet talk go down amo...
4      s5  ...  In a city of coaching centers known to train I...
[5 rows x 12 columns]
```

```
Python Console x +
4 s5 ... In a city of coaching centers known to train I...
[5 rows x 12 columns]
In [18]: df.tail()
Out[18]:
>>>      show_id  ...      description
8802  s8803  ...  A political cartoonist, a crime reporter and a...
8803  s8804  ...  While living alone in a spooky town, a young g...
8804  s8805  ...  Looking to survive in a world taken over by zo...
8805  s8806  ...  Dragged from civilian life, a former superhero...
8806  s8807  ...  A scrappy but poor boy worms his way into a ty...

[5 rows x 12 columns]
```

```
Python Console x +
[5 rows x 12 columns]
In [19]: df.describe()
Out[19]:
>>>      release_year
count      8807.000000
mean       2014.180198
std         8.819312
min        1925.000000
25%        2013.000000
50%        2017.000000
75%        2019.000000
max         2021.000000
```

```
Python Console x +
In [24]: df.describe(include="all")
Out[24]:
```

	show_id	...	description
count	8807	...	8807
unique	8807	...	8775
top	s1	...	Paranormal activity at a lush, abandoned prope...
freq	1	...	4
mean	NaN	...	NaN
std	NaN	...	NaN
min	NaN	...	NaN
25%	NaN	...	NaN
50%	NaN	...	NaN
75%	NaN	...	NaN
max	NaN	...	NaN

```
In [25]:
```

```
Python Console x +
In [25]: df["director"].value_counts().head(10)
Out[25]:
```

director	count
Rajiv Chilaka	19
Raúl Campos, Jan Suter	18
Suhas Kadav	16
Marcus Raboy	16
Jay Karas	14
Cathy Garcia-Molina	13
Martin Scorsese	12
Youssef Chahine	12
Jay Chapman	12
Steven Spielberg	11

```
Name: count, dtype: int64
In [26]:
```

```
Jay Chapman 12
Steven Spielberg 11
Name: count, dtype: int64
In [26]: df["director"].value_counts().sum()
Out[26]: np.int64(6173)
In [27]: |
```

```

In [27]: df["date_added"] = pd.to_datetime(df["date_added"], dayfirst=True, errors="coerce")
In [28]: df.info()
<class 'pandas.DataFrame'>
RangeIndex: 8807 entries, 0 to 8806
>>> Data columns (total 12 columns):
#   Column      Non-Null Count  Dtype
---  -
0   show_id     8807 non-null   str
1   type        8807 non-null   str
2   title       8807 non-null   str
3   director    6173 non-null   str
4   cast        7982 non-null   str
5   country     7976 non-null   str

```

```

6   date_added  8709 non-null  datetime64[us]
7   release_year 8807 non-null  int64
8   rating      8803 non-null  str
9   duration    8804 non-null  str
10  listed_in   8807 non-null  str
11  description  8807 non-null  str
dtypes: datetime64[us](1), int64(1), str(10)
memory usage: 825.8 KB

```

Python Console × +

```

In [29]: df
Out[29]:
  show_id  ... description
0      s1  ... As her father nears the end of his life, filmm...
1      s2  ... After crossing paths at a party, a Cape Town t...
2      s3  ... To protect his family from a powerful drug lor...
3      s4  ... Feuds, flirtations and toilet talk go down amo...
4      s5  ... In a city of coaching centers known to train I...
...      ...
8802  s8803  ... A political cartoonist, a crime reporter and a...
8803  s8804  ... While living alone in a spooky town, a young g...
8804  s8805  ... Looking to survive in a world taken over by zo...
8805  s8806  ... Dragged from civilian life, a former superhero...
8806  s8807  ... A scrappy but poor boy worms his way into a ty...
In [30]:

```

```
Python Console x +
[8807 rows x 12 columns]
In [30]: df.isna().sum()
Out[30]:
show_id      0
type         0
title        0
director    2634
cast        825
country     831
date_added   98
release_year  0
rating       4
duration     3
listed_in    0
In [31]: |
```

```
Python Console x +
In [31]: df[["title", "cast"]].sort_values(by="cast",ascending=False).head(10)
Out[31]:
      title                                     cast
1775  His House  Şopé Dirisù, Wunmi Mosaku, Matt Smith, Malaik...
8749  Winter Sun  Şükrü Özyıldız, Aslı Enver, Şenay Gürler, Başa...
2024  Night of Knots  Şahin Irmak, İrem Sak, Gonca Vuslateri, Emre K...
5395  Intersection  İbrahim Çelikkol, Belçim Bilgin, Alican Yüceso...
2618  Küçük Esnaf  İbrahim Büyükak, Zeynep Koçak, Gupse Özay, Cen...
4124  The Tree of Blood  Úrsula Corberó, Álvaro Cervantes, Najwa Nimri,...
109   La casa de papel  Úrsula Corberó, Itziar Ituño, Álvaro Morte, Pa...
1708  Ethos  Öykü Karayel, Fatih Artman, Funda Eryiğit, Def...
6204  Babam ve Oğlum  Çetin Tekindor, Fikret Kuşkan, Hümeyra, Ege Ta...
2428  Overnight Republic  Çağlar Çorumlu, Büşra Pekin, Ersin Korkut, Gül...
```



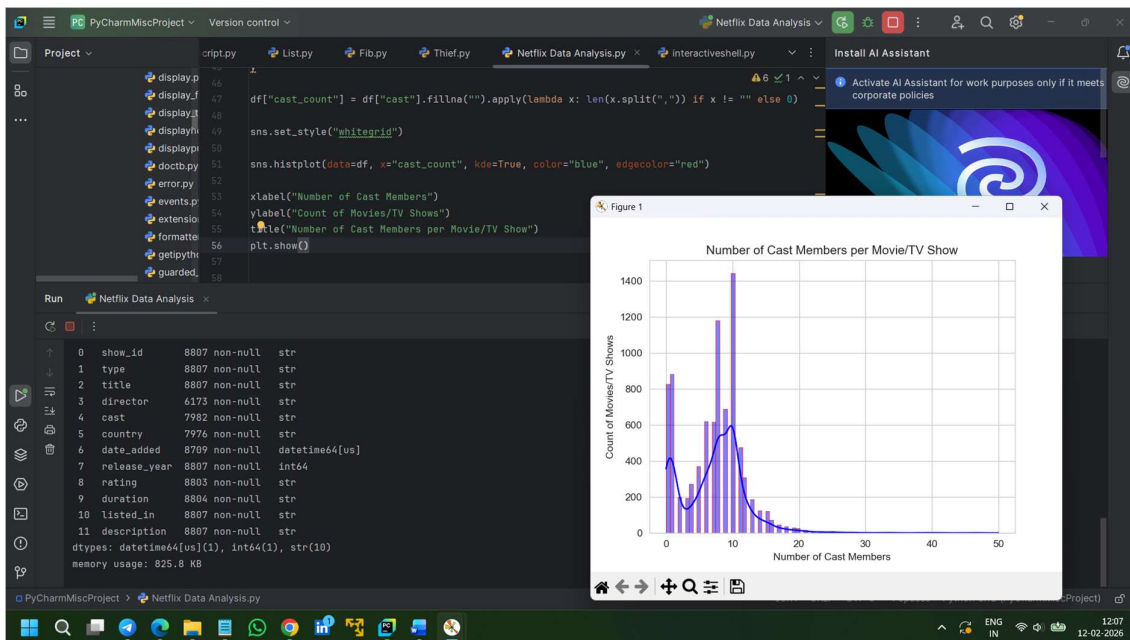
```
Python Console x +
In [32]: df[["cast", "director"]].sort_values(by="cast",ascending=False).head(10)
Out[32]:
```

	cast	director
1775	Şopé Dirisù, Wunmi Mosaku, Matt Smith, Malaik...	Remi Weekes
8749	Şükrü Özyıldız, Aslı Enver, Şenay Gürler, Başa...	NaN
2024	Şahin Irmak, İrem Sak, Gonca Vuslateri, Emre K...	Hakan Algül
5395	İbrahim Çelikkol, Belçim Bilgin, Alican Yüceso...	NaN
2618	İbrahim Büyükek, Zeynep Koçak, Gupse Özay, Cen...	Bedran Güzel
4124	Úrsula Corberó, Álvaro Cervantes, Najwa Nimri,...	Julio Medem
109	Úrsula Corberó, Itziar Ituño, Álvaro Morte, Pa...	NaN
1708	Öykü Karayel, Fatih Artman, Funda Eryiğit, Def...	NaN
6204	Çetin Tekindor, Fikret Kuşkan, Hümeysra, Ege Ta...	Çagan Irmak
2428	Çağlar Çorumlu, Büşra Pekin, Ersin Korkut, Gül...	Murat Kepez

```
In [33]: y=df.pivot_table(index='type',columns='release_year',values='date_added',aggfunc=["count"])
...: y
Out[33]:
```

	count										
release_year	1925	1942	1943	1944	1945	...	2017	2018	2019	2020	2021
type	...										
Movie	NaN	2.0	3.0	3.0	3.0	...	767.0	767.0	633.0	517.0	277.0
TV Show	1.0	NaN	NaN	NaN	1.0	...	249.0	373.0	397.0	436.0	315.0

[2 rows x 74 columns]

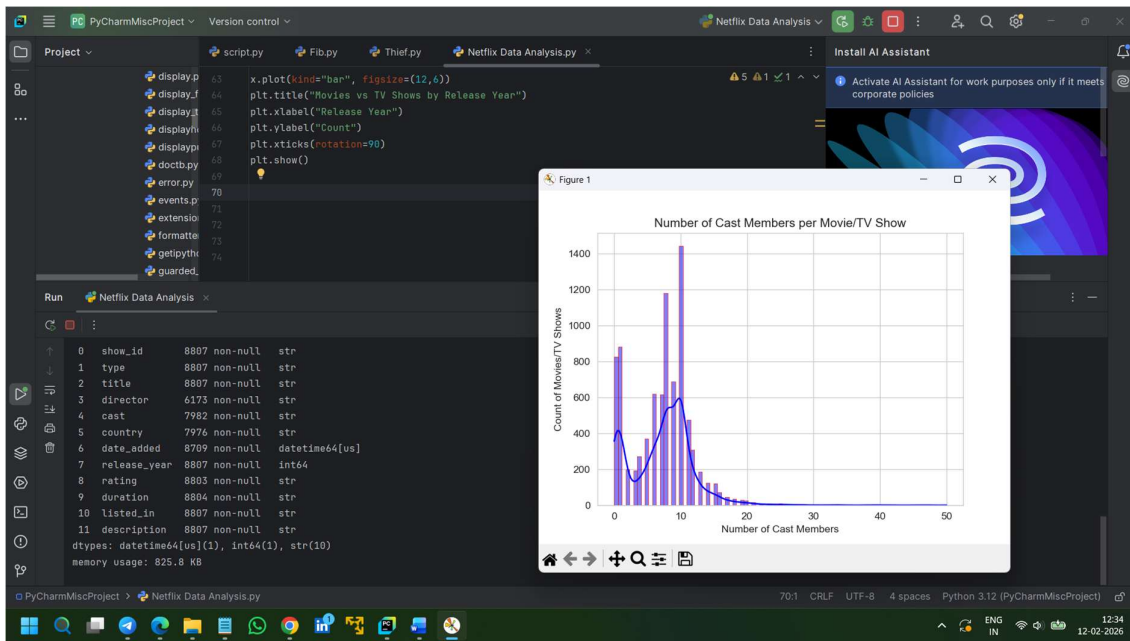


```
Python Console x +
8805 s8806 ... Dragged from civilian life, a former superhero...
8806 s8807 ... A scrappy but poor boy worms his way into a ty...
[8807 rows x 12 columns]
In [5]: df.groupby("release_year")["title"].count().head()
Out[5]:
release_year
1925      1
1942      2
1943      3
1944      3
1945      4
Name: title, dtype: int64
```

```
Python Console x +
Name: title, dtype: int64
In [6]: x=df.pivot_table(index="release_year",columns="type",values="title",aggfunc="count")
...: x
Out[6]:
type      Movie  TV Show
release_year
1925      NaN    1.0
1942      2.0    NaN
1943      3.0    NaN
1944      3.0    NaN
1945      3.0    1.0
...      ...    ...
2017     767.0   265.0
2018     767.0   380.0
In [7]: |
```

2019	633.0	397.0
2020	517.0	436.0
2021	277.0	315.0

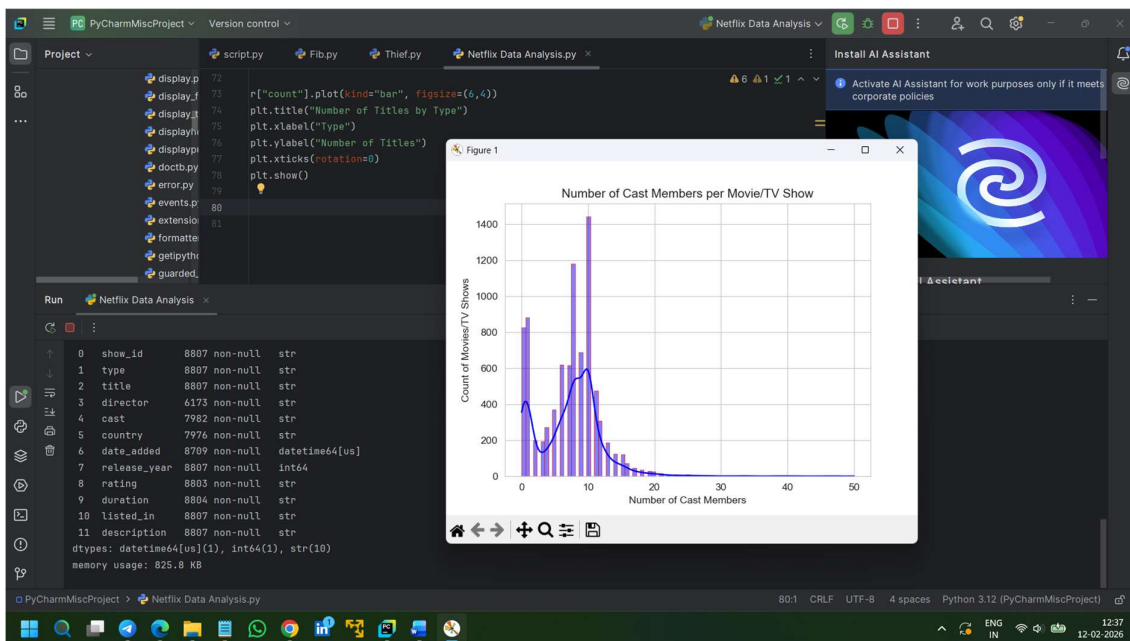
[74 rows x 2 columns]



```

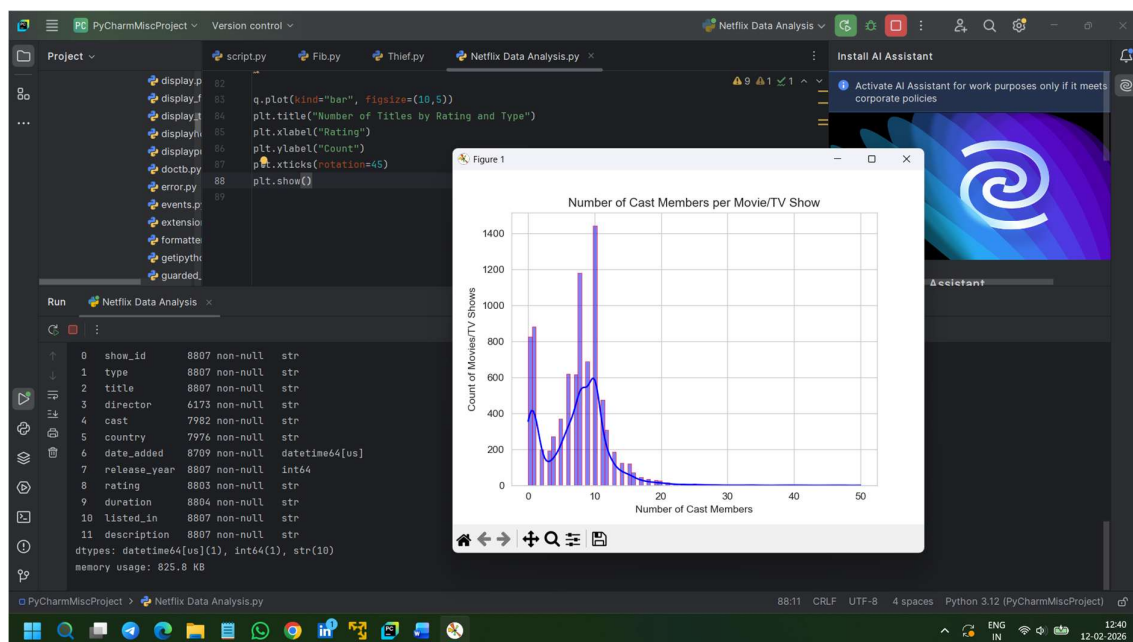
>>> In [6]: r=df.pivot_table(index="type",values="cast",aggfunc=["count"])
...: r
Out[6]:
count
cast
type
Movie    5656
TV Show   2326

```



```
Python Console x +
In [7]: q=df.pivot_table(index="rating",columns="type",values="title",aggfunc="count")
...: q
Out[7]:
type      Movie  TV Show
rating
66 min      1.0      NaN
74 min      1.0      NaN
84 min      1.0      NaN
G           41.0      NaN
NC-17       3.0      NaN
NR          75.0      5.0
PG         287.0      NaN
PG-13      490.0      NaN
R          797.0      2.0

>>> TV-14      1427.0    733.0
TV-G           126.0     94.0
TV-MA          2062.0   1145.0
TV-PG          540.0    323.0
TV-Y           131.0    176.0
TV-Y7          139.0    195.0
TV-Y7-FV        5.0      1.0
UR              3.0      NaN
```



```

Python Console x +
UR 3.0 NaN
In [8]: df.pivot_table(index="rating",columns="type",values="cast",aggfunc=["max","min"])
Out[8]:
           max ... min
type
Movie ... TV Show
>>> rating
66 min Louis C.K. ... NaN
74 min Louis C.K. ... NaN
84 min Louis C.K. ... NaN
G Zach Braff, Joan Cusack, Patrick Stewart, Stev... ... NaN
NC-17 Mike Smith, John Paul Tremblay, Robb Wells, Pa... ... NaN
NR Zachary Levi ... Burnie Burns, Jason Saldaña, Gustavo Sorola, G...
PG YaYa Gosselin, Pedro Pascal, Priyanka Chopra, ... ... NaN
PG-13 Zoey Deutch, Lucy Fry, Danila Kozlovsky, Gabri... ... NaN
In [9]:

R Zoe Lister-Jones, Adam Pally, Fred Armisen, Su... ... Aksel Hennie, Tuva Novotny, Anders Danielsen L...
TV-14 Sópé Dirisù, Wunmi Mosaku, Matt Smith, Malaik... ... Aakarshan Singh, Vikramjeet Virk, Reena Aggarw...
TV-6 Woody Harrelson ... 'Najite Dede, Jude Chukwuka, Taiwo Arimoro, Od...
>>> TV-MA Şahin Irmak, İrem Sak, Gonca Vuslateri, Emre K... ... Aaditi Pohankar, Vijay Varma, Vishwas Kini, Ki...
TV-P6 Zion Clark ... Adriano Zumbo, Rachel Khoo
TV-Y Zendaya, Chanelle Peloso, Spencer Boldman, Emi... ... Abby Bergman, Anna Cooke, Aston Droomer, Jamil...
TV-Y7 Yara Shahidi, Miles Robbins, Miguel J. Pimente... ... Adrian Petriw, Ashleigh Ball, Connor Parnall, ...
TV-Y7-FV Saurav Chakrabarty, Vinay Pathak ... Al Mukadam, Dan Chameroy, Seán Cullen, Stacey ...
UR Lise Danvers, Fabrice Luchini, Charlotte Alexa... ... NaN

[17 rows x 4 columns]

```

