

Analyze the dataset and create graphs using seaborn and matplotlib. Dataset :-
<https://docs.google.com/spreadsheets/d/e/2PACX-1vTSS-TcErkXNk8KB0AlijhitwetxeHD2M3R0HJI2QPMayFq0fxFX4PFKnzAWLDnratlz67DNL6GsZnV/pub?output=csv>

```
import random
```

```
import pandas as pd
```

```
import seaborn as sns
```

```
from matplotlib import pyplot as plt
```

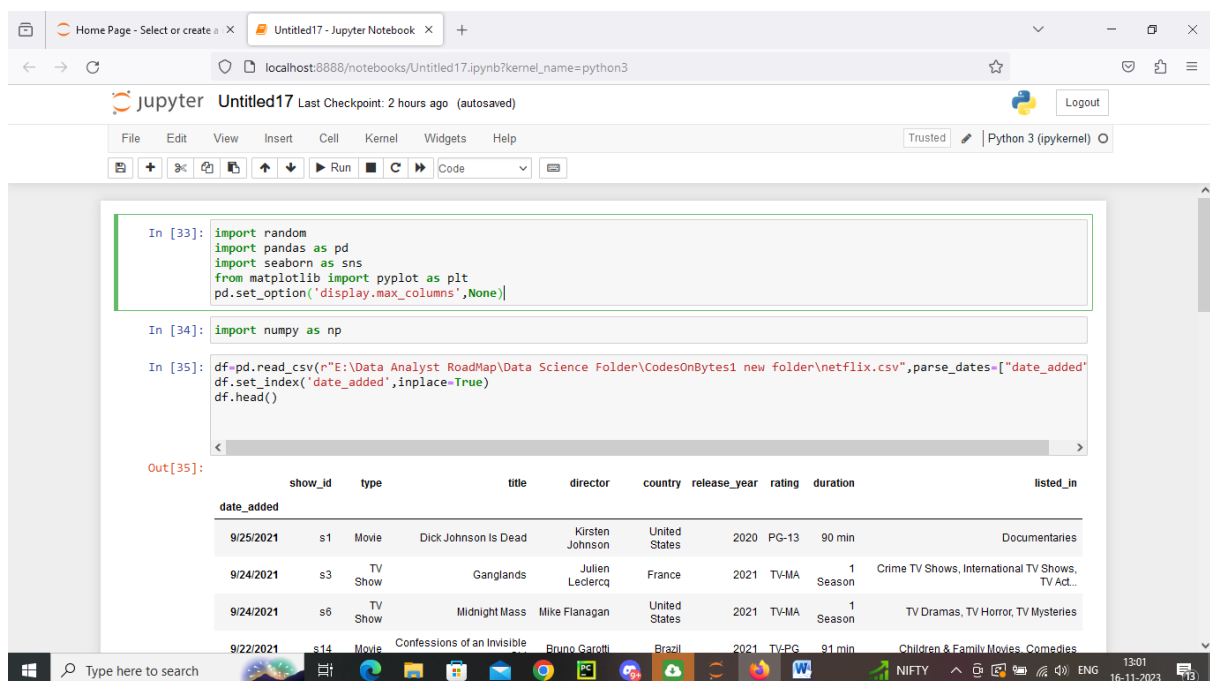
```
pd.set_option('display.max_columns',None)
```

```
import numpy as np
```

```
df=pd.read_csv(r"E:\Data Analyst RoadMap\Data Science  
Folder\CodesOnBytes1 new folder\netflix.csv",parse_dates=["date_added"])
```

```
df.set_index('date_added',inplace=True)
```

```
df.head()
```



The screenshot displays a Jupyter Notebook window titled 'Untitled17 - Jupyter Notebook'. The code in the notebook is as follows:

```
In [33]: import random
import pandas as pd
import seaborn as sns
from matplotlib import pyplot as plt
pd.set_option('display.max_columns',None)

In [34]: import numpy as np

In [35]: df=pd.read_csv(r"E:\Data Analyst RoadMap\Data Science Folder\CodesOnBytes1 new folder\netflix.csv",parse_dates=["date_added"])
df.set_index('date_added',inplace=True)
df.head()
```

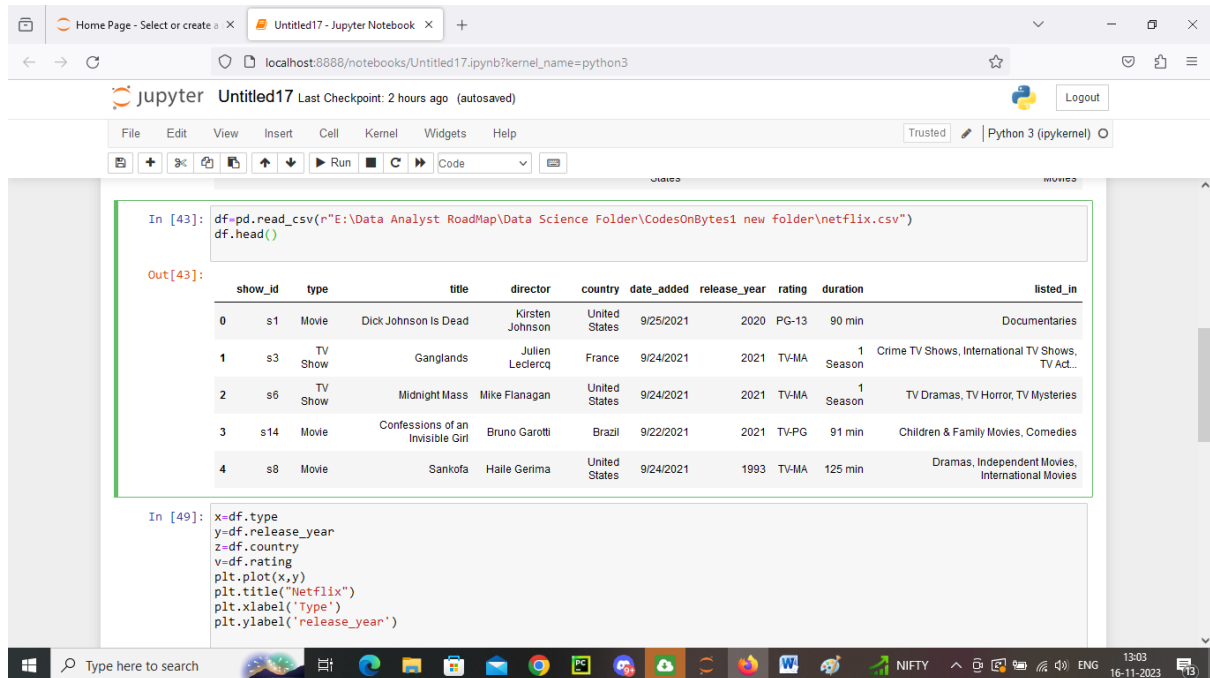
The output of the code is displayed below the cells:

```
Out[35]:
```

date_added	show_id	type	title	director	country	release_year	rating	duration	listed_in
9/25/2021	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	United States	2020	PG-13	90 min	Documentaries
9/24/2021	s3	TV Show	Ganglands	Julien Leclercq	France	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...
9/24/2021	s6	TV Show	Midnight Mass	Mike Flanagan	United States	2021	TV-MA	1 Season	TV Dramas, TV Horror, TV Mysteries
9/22/2021	s14	Movie	Confessions of an Invisible	Bruno Garotti	Brazil	2021	TV-PG	91 min	Children & Family Movies, Comedies

```
df=pd.read_csv(r"E:\Data Analyst RoadMap\Data Science  
Folder\CodesOnBytes1 new folder\netflix.csv")
```

```
df.head()
```



The screenshot shows a Jupyter Notebook window titled 'Untitled17' with a Python 3 kernel. The first code cell (In [43]) reads a CSV file and displays the first five rows of the DataFrame. The second code cell (In [49]) creates a scatter plot of release year versus type.

Out[43]:

	show_id	type	title	director	country	date_added	release_year	rating	duration	listed_in
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	United States	9/25/2021	2020	PG-13	90 min	Documentaries
1	s3	TV Show	Ganglands	Julien Lederq	France	9/24/2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...
2	s6	TV Show	Midnight Mass	Mike Flanagan	United States	9/24/2021	2021	TV-MA	1 Season	TV Dramas, TV Horror, TV Mysteries
3	s14	Movie	Confessions of an Invisible Girl	Bruno Garotti	Brazil	9/22/2021	2021	TV-PG	91 min	Children & Family Movies, Comedies
4	s8	Movie	Sankofa	Haile Gerima	United States	9/24/2021	1993	TV-MA	125 min	Dramas, Independent Movies, International Movies

In [49]:

```
x=df.type  
y=df.release_year  
z=df.country  
v=df.rating  
plt.plot(x,y)  
plt.title("Netflix")  
plt.xlabel('Type')  
plt.ylabel('release_year')
```

```
x=df.type
```

```
y=df.release_year
```

```
z=df.country
```

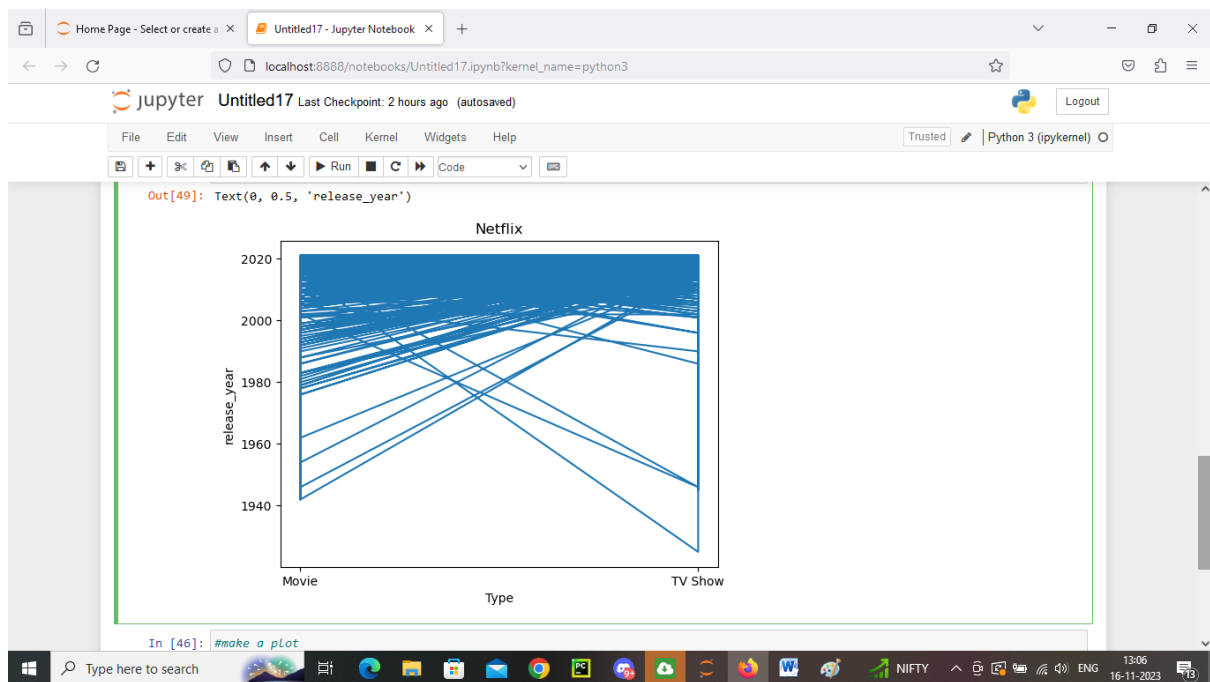
```
v=df.rating
```

```
plt.plot(x,y)
```

```
plt.title("Netflix")
```

```
plt.xlabel('Type')
```

```
plt.ylabel('release_year')
```



```
import pandas as pd
```

```
import matplotlib.pyplot as plt
```

```
df=pd.read_csv(r"E:\Data Analyst RoadMap\Data Science  
Folder\CodesOnBytes1 new folder\netflix.csv")
```

```
newdf=df[['type','release_year','duration']]
```

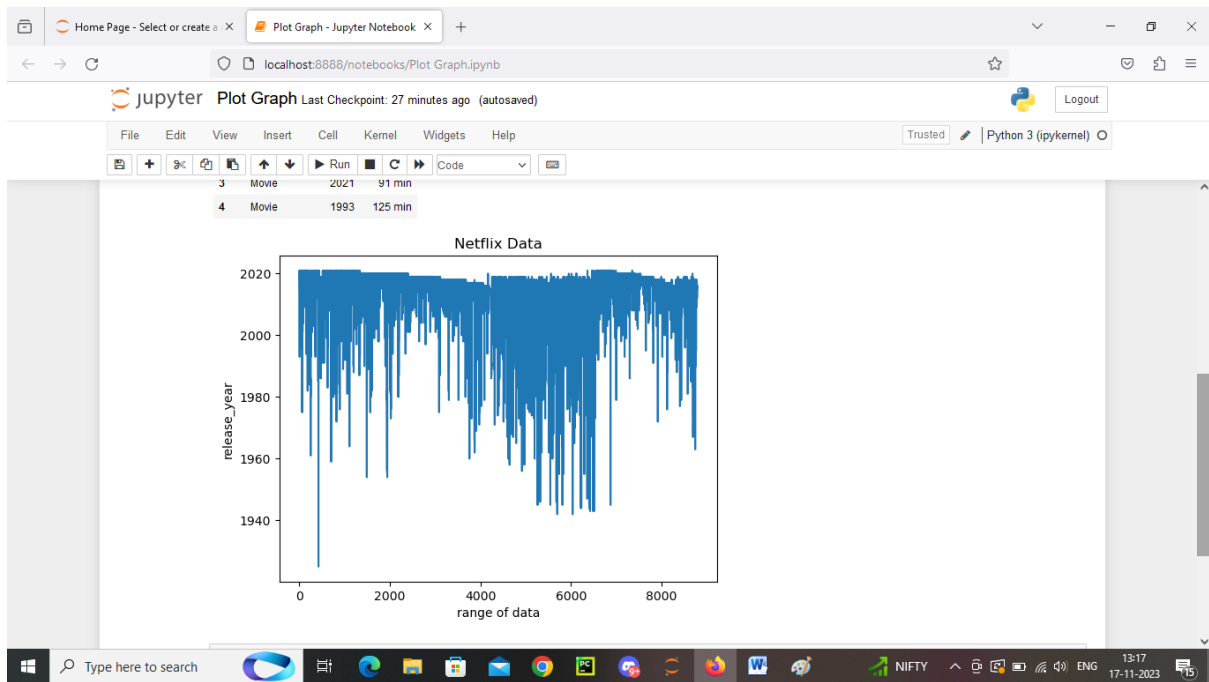
```
plt.plot(newdf['release_year'])
```

```
plt.title('Netflix Data')
```

```
plt.xlabel('range of data')
```

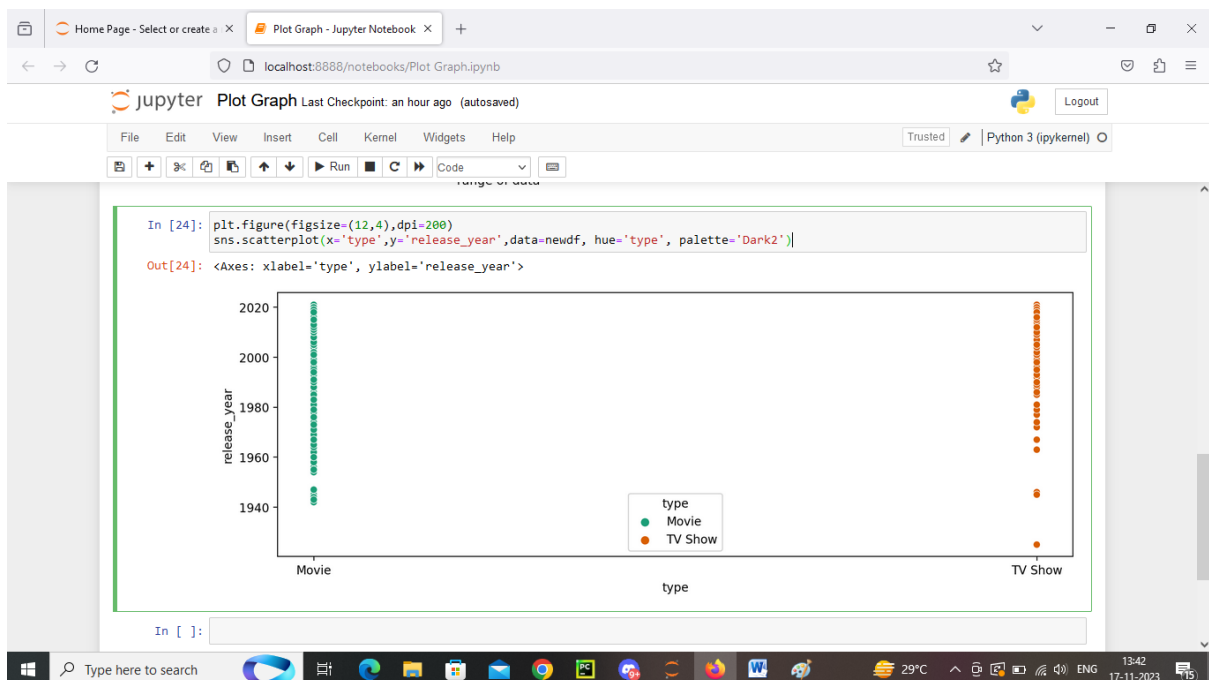
```
plt.ylabel('release_year')
```

```
newdf.head()
```



```
plt.figure(figsize=(12,4),dpi=200)
```

```
sns.scatterplot(x='type',y='release_year',data=newdf, hue='type',  
palette='Dark2')
```



```
plt.figure(figsize=(12,4),dpi=200)
```

```
sns.scatterplot(x='type',y='release_year',data=newdf,s=200, hue='type',  
style='type')
```

