##1. Load the dataset and print the complete information of the dataset and name the columns that has missing values.

```
import pandas as pd

netflix=pd.read_csv("netflix.csv")
print(netflix.info())
print("columns that has missing values are director,cast,country")
```

```
<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
None
columns that has missing values are director,cast,country
```

##2.Load the dataset and print the top 10 countries contributing to the Content on Netflix

```
import pandas as pd

netflix=pd.read_csv("netflix.csv")
print(netflix["country"].value_counts().head(10))
```

```
country
United States     2818
India              972
United Kingdom     419
Japan              245
South Korea        199
Canada             181
Spain              145
France             124
Mexico             110
Egypt              106
Name: count, dtype: int64
```

##3.Load the dataset and list the 5 most popular types of genre on the platform.

```
import pandas as pd
netflix=pd.read_csv("netflix.csv")
genre_count=netflix["listed_in"].str.split(", ").explode().value_counts()
print(genre_count.head(5))
```

```
listed_in
International Movies      2752
Dramas                    2427
Comedies                  1674
International TV Shows    1351
Documentaries              869
Name: count, dtype: int64
```

##4.Load the Dataset and categorize into TV shows and Movies and display 2 rows for each.

```
import pandas as pd
netflix=pd.read_csv("netflix.csv")
movies=netflix[netflix["type"]=="Movie"]
Tvshows=netflix[netflix["type"]=="TV Show"]
print(movies.head(2))
print(Tvshows.head(2))
```

```
  show_id   type                              title  \
0      s1  Movie              Dick Johnson Is Dead
6      s7  Movie  My Little Pony: A New Generation

                        director  \
0               Kirsten Johnson
6  Robert Cullen, José Luis Ucha

                                              cast        country  \
0                                              NaN  United States
6  Vanessa Hudgens, Kimiko Glenn, James Marsden, ...            NaN

           date_added  release_year rating duration                listed_in  \
0  September 25, 2021          2020  PG-13   90 min             Documentaries
6  September 24, 2021          2021     PG   91 min  Children & Family Movies

                                        description
0  As her father nears the end of his life, filmm...
6  Equestria's divided. But a bright-eyed hero be...
  show_id     type          title         director  \
1      s2  TV Show  Blood & Water              NaN
2      s3  TV Show      Ganglands  Julien Leclercq

                                              cast       country  \
1  Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...  South Africa
2  Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...           NaN

           date_added  release_year rating   duration  \
1  September 24, 2021          2021  TV-MA  2 Seasons
2  September 24, 2021          2021  TV-MA   1 Season

                                         listed_in  \
1    International TV Shows, TV Dramas, TV Mysteries
2  Crime TV Shows, International TV Shows, TV Act...

                                        description
1  After crossing paths at a party, a Cape Town t...
2  To protect his family from a powerful drug lor...
```

##5. Load the Dataset and find the Oldest movies

```
import pandas as pd
netflix=pd.read_csv("netflix.csv")
movies=netflix[netflix["type"]=="Movie"]
print(movies[movies["release_year"]==movies["release_year"].min()]["title"])
```

```
7790          Prelude to War
8205    The Battle of Midway
Name: title, dtype: object
```

##6. Load the Dataset and get the count of values for duration column.

```
import pandas as pd
import numpy as np
netflix=pd.read_csv("netflix.csv")
print(netflix.duration.value_counts())
```

```
duration
1 Season     1793
2 Seasons     425
3 Seasons     199
90 min        152
94 min        146
             ...
16 min          1
186 min         1
193 min         1
189 min         1
```

```
191 min         1
Name: count, Length: 220, dtype: int64
```

##7. Load the Dataset and find the measures of central tendency and dispersion of duration column by extracting numerical part of the durat

```
# code goes here
import pandas as pd

netflix=pd.read_csv("netflix.csv")
netflix["duration1"]=netflix["duration"].str.extract("(\d+)",expand=False).astype("float")
print("The mean duration is",netflix["duration1"].mean())
print("The highest duration is",netflix["duration1"].max())
print("The minimum duration is",netflix["duration1"].min())
print("The variance of duration is",netflix["duration1"].var())
print("The standard deviation is",netflix["duration1"].std())
```

```
The mean duration is 69.84688777828259
The highest duration is 312.0
The minimum duration is 1.0
The variance of duration is 2582.146723244931
The standard deviation is 50.81482778918896
```

##8. Load the Dataset and find out to which country the highest duration movie belongs to by extracting the numerical part of the duration c

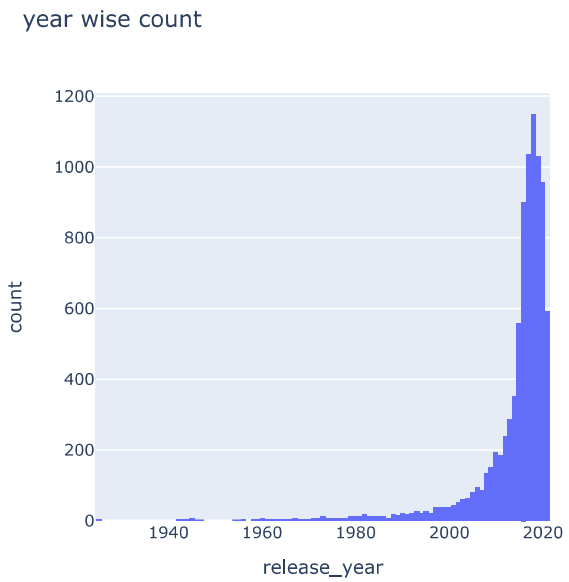
```
# print title and country column.

import pandas as pd
import numpy as np
netflix=pd.read_csv("netflix.csv")
netflix["duration1"]=netflix["duration"].str.extract("(\d+)",expand=False).astype("float")
print(netflix[netflix["duration1"]==netflix["duration1"].max()][["title","country"]])
```

```
                          title        country
4253  Black Mirror: Bandersnatch  United States
```

##9. Load the Dataset and plot the histogram on Netflix Content Release Year Distribution

```
import pandas as pd
import numpy as np
from matplotlib import pyplot as plt
import seaborn as sns
netflix=pd.read_csv("netflix.csv")
import plotly.express as px
px.histogram(netflix,x="release_year",title="year wise count",height=500,width=500)
```



```
##10. Load the Dataset and plot the Distribution of Popularity of different content categories using countplot

import pandas as pd
import numpy as np
from matplotlib import pyplot as plt
import seaborn as sns
netflix=pd.read_csv("netflix.csv")
ax=sns.countplot(data=netflix,y="listed_in", order=netflix["listed_in"].value_counts().index[0:20])
ax.bar_label(ax.containers[0]) # to get values next to graphs

plt.show()
```