Importing Libraries

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
```

File Loading

```
df=pd.read_csv(r'C:\Users\Kalpesh\Downloads\netflix1.csv' ,
encoding='unicode_escape')
```

Data Cleaning

```
df.shape
```

```
(8790, 10)
```

```
df.head(10)
```

```
  show_id     type                                title
director  \
0      s1    Movie               Dick Johnson Is Dead      Kirsten
Johnson
1      s3  TV Show                          Ganglands      Julien
Leclercq
2      s6  TV Show                      Midnight Mass        Mike
Flanagan
3     s14    Movie  Confessions of an Invisible Girl        Bruno
Garotti
4      s8    Movie                            Sankofa         Haile
Gerima
5      s9  TV Show     The Great British Baking Show      Andy
Devonshire
6     s10    Movie                        The Starling       Theodore
Melfi
7    s939    Movie   Motu Patlu in the Game of Zones          Suhas
Kadav
8     s13    Movie                        Je Suis Karl  Christian
Schwochow
9    s940    Movie          Motu Patlu in Wonderland          Suhas
Kadav

           country date_added  release_year rating   duration  \
0   United States  9/25/2021          2020  PG-13     90 min
1          France  9/24/2021          2021  TV-MA   1 Season
2   United States  9/24/2021          2021  TV-MA   1 Season
3          Brazil  9/22/2021          2021  TV-PG     91 min
```

```
4   United States  9/24/2021          1993  TV-MA    125 min
5  United Kingdom  9/24/2021          2021  TV-14  9 Seasons
6   United States  9/24/2021          2021  PG-13    104 min
7           India   5/1/2021          2019  TV-Y7     87 min
8         Germany  9/23/2021          2021  TV-MA    127 min
9           India   5/1/2021          2013  TV-Y7     76 min

                                              listed_in
0                                          Documentaries
1  Crime TV Shows, International TV Shows, TV Act...
2                TV Dramas, TV Horror, TV Mysteries
3                Children & Family Movies, Comedies
4   Dramas, Independent Movies, International Movies
5                      British TV Shows, Reality TV
6                                   Comedies, Dramas
7  Children & Family Movies, Comedies, Music & Mu...
8                      Dramas, International Movies
9         Children & Family Movies, Music & Musicals
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8790 entries, 0 to 8789
Data columns (total 10 columns):
 #   Column        Non-Null Count  Dtype 
---  ------        --------------  ----- 
 0   show_id       8790 non-null   object
 1   type          8790 non-null   object
 2   title         8790 non-null   object
 3   director      8790 non-null   object
 4   country       8790 non-null   object
 5   date_added    8790 non-null   object
 6   release_year  8790 non-null   int64 
 7   rating        8790 non-null   object
 8   duration      8790 non-null   object
 9   listed_in     8790 non-null   object
dtypes: int64(1), object(9)
memory usage: 686.8+ KB
```

```
df=df.drop_duplicates()
```

```
df['type'].value_counts()
```

```
type
Movie      6126
TV Show    2664
Name: count, dtype: int64
```

```
pd.isnull(df)
```

```
      show_id   type  title  director  country  date_added
release_year  \
0       False  False  False     False    False       False
False
1       False  False  False     False    False       False
False
2       False  False  False     False    False       False
False
3       False  False  False     False    False       False
False
4       False  False  False     False    False       False
False
...       ...    ...    ...       ...      ...         ...            .
..
8785    False  False  False     False    False       False
False
8786    False  False  False     False    False       False
False
8787    False  False  False     False    False       False
False
8788    False  False  False     False    False       False
False
8789    False  False  False     False    False       False
False

      rating  duration  listed_in
0      False     False      False
1      False     False      False
2      False     False      False
3      False     False      False
4      False     False      False
...      ...       ...        ...
8785   False     False      False
8786   False     False      False
8787   False     False      False
8788   False     False      False
8789   False     False      False

[8790 rows x 10 columns]
```

```
pd.isnull(df).sum()
```

```
show_id         0
type            0
title           0
director        0
country         0
date_added      0
release_year    0
rating          0
```

```
duration        0
listed_in       0
dtype: int64
```

```
df.dropna(inplace=True)
```

```
df.dropna(subset=['director','title','country' ],inplace=True)
```

```
 df['date_added'] = df['date_added'].astype('datetime64[ns]')
```

```
df['director'] = df['director'].fillna('Unspecified')
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8790 entries, 0 to 8789
Data columns (total 10 columns):
 #   Column        Non-Null Count  Dtype         
---  ------        --------------  -----         
 0   show_id       8790 non-null   object        
 1   type          8790 non-null   object        
 2   title         8790 non-null   object        
 3   director      8790 non-null   object        
 4   country       8790 non-null   object        
 5   date_added    8790 non-null   datetime64[ns]
 6   release_year  8790 non-null   int64         
 7   rating        8790 non-null   object        
 8   duration      8790 non-null   object        
 9   listed_in     8790 non-null   object        
dtypes: datetime64[ns](1), int64(1), object(8)
memory usage: 686.8+ KB
```

```
df['date_added'].dtypes
```

```
dtype('<M8[ns]')
```

```
df.columns
```

```
Index(['show_id', 'type', 'title', 'director', 'country',
'date_added',
       'release_year', 'rating', 'duration', 'listed_in'],
      dtype='object')
```

```
df.rename(columns={'date_added':'added_date'})
```

```
     show_id     type                                title
director  \
0         s1    Movie             Dick Johnson Is Dead  Kirsten
Johnson
1         s3  TV Show                         Ganglands  Julien
Leclercq
2         s6  TV Show                     Midnight Mass    Mike
```

```
Flanagan
3        s14    Movie  Confessions of an Invisible Girl    Bruno
Garotti
4         s8    Movie                          Sankofa     Haile
Gerima
...      ...      ...                             ...               .
..
8785   s8797  TV Show                       Yunus Emre        Not
Given
8786   s8798  TV Show                        Zak Storm        Not
Given
8787   s8801  TV Show                Zindagi Gulzar Hai        Not
Given
8788   s8784  TV Show                             Yoko        Not
Given
8789   s8786  TV Show                              YOM        Not
Given

            country added_date  release_year rating   duration  \
0     United States 2021-09-25          2020  PG-13     90 min
1            France 2021-09-24          2021  TV-MA   1 Season
2     United States 2021-09-24          2021  TV-MA   1 Season
3            Brazil 2021-09-22          2021  TV-PG     91 min
4     United States 2021-09-24          1993  TV-MA    125 min
...             ...        ...           ...    ...        ...
8785         Turkey 2017-01-17          2016  TV-PG  2 Seasons
8786  United States 2018-09-13          2016  TV-Y7  3 Seasons
8787       Pakistan 2016-12-15          2012  TV-PG   1 Season
8788       Pakistan 2018-06-23          2016   TV-Y   1 Season
8789       Pakistan 2018-06-07          2016  TV-Y7   1 Season

                                              listed_in
0                                        Documentaries
1     Crime TV Shows, International TV Shows, TV Act...
2                    TV Dramas, TV Horror, TV Mysteries
3                    Children & Family Movies, Comedies
4      Dramas, Independent Movies, International Movies
...                                                 ...
8785                  International TV Shows, TV Dramas
8786                                       Kids' TV
8787  International TV Shows, Romantic TV Shows, TV ...
8788                                       Kids' TV
8789                                       Kids' TV

[8790 rows x 10 columns]
```

```
df.rename(columns={'date_added':'added_date'},inplace=True)

df_drop= df.dropna(subset=['director'],inplace=True)
```

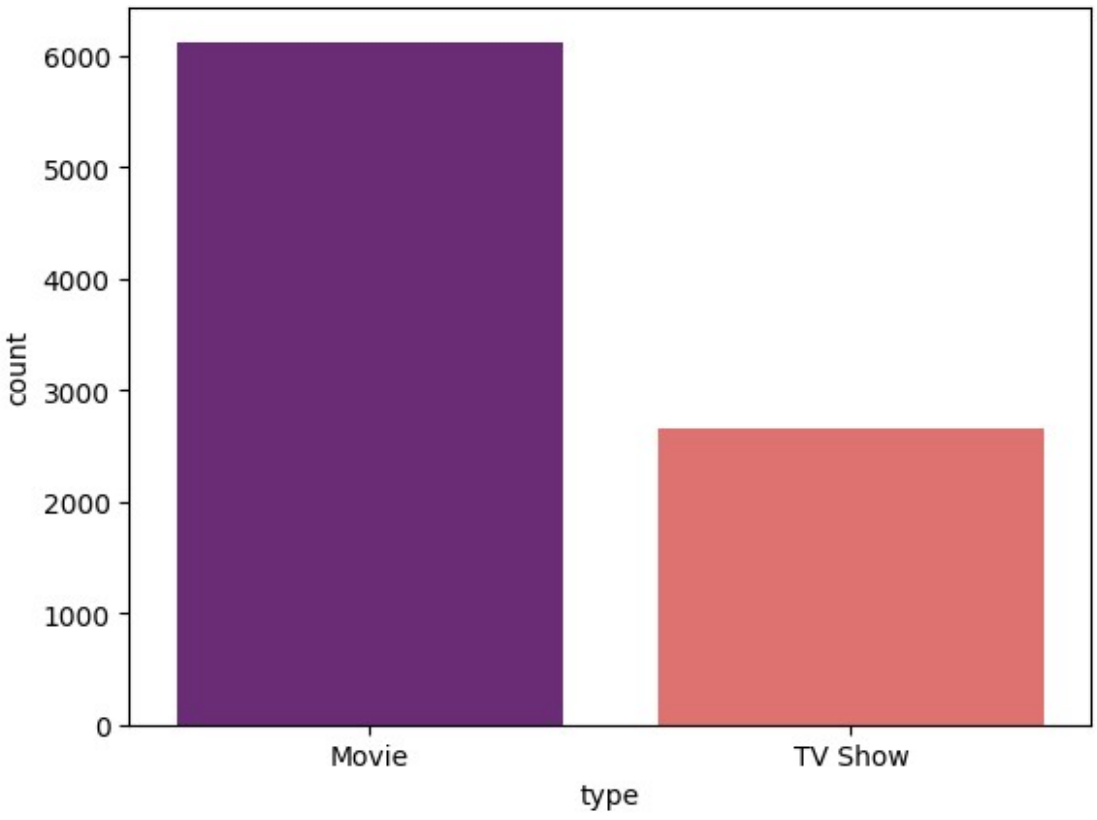
```
df

     show_id     type                                title
director  \
0         s1    Movie              Dick Johnson Is Dead  Kirsten
Johnson
1         s3  TV Show                          Ganglands  Julien
Leclercq
2         s6  TV Show                      Midnight Mass    Mike
Flanagan
3        s14    Movie  Confessions of an Invisible Girl    Bruno
Garotti
4         s8    Movie                            Sankofa     Haile
Gerima
...      ...      ...                               ...               .
..
8785   s8797  TV Show                        Yunus Emre        Not
Given
8786   s8798  TV Show                         Zak Storm        Not
Given
8787   s8801  TV Show                Zindagi Gulzar Hai        Not
Given
8788   s8784  TV Show                              Yoko        Not
Given
8789   s8786  TV Show                               YOM        Not
Given

            country added_date  release_year rating   duration  \
0     United States 2021-09-25          2020  PG-13     90 min
1            France 2021-09-24          2021  TV-MA   1 Season
2     United States 2021-09-24          2021  TV-MA   1 Season
3            Brazil 2021-09-22          2021  TV-PG     91 min
4     United States 2021-09-24          1993  TV-MA    125 min
...             ...        ...           ...    ...        ...
8785         Turkey 2017-01-17          2016  TV-PG  2 Seasons
8786  United States 2018-09-13          2016  TV-Y7  3 Seasons
8787       Pakistan 2016-12-15          2012  TV-PG   1 Season
8788       Pakistan 2018-06-23          2016   TV-Y   1 Season
8789       Pakistan 2018-06-07          2016  TV-Y7   1 Season

                                              listed_in
0                                         Documentaries
1     Crime TV Shows, International TV Shows, TV Act...
2                    TV Dramas, TV Horror, TV Mysteries
3                    Children & Family Movies, Comedies
4      Dramas, Independent Movies, International Movies
...                                                 ...
8785                  International TV Shows, TV Dramas
8786                                          Kids' TV
8787  International TV Shows, Romantic TV Shows, TV ...
```

```
8788                                                  Kids' TV
8789                                                  Kids' TV

[8790 rows x 10 columns]
```

Exploratory Data Analysis

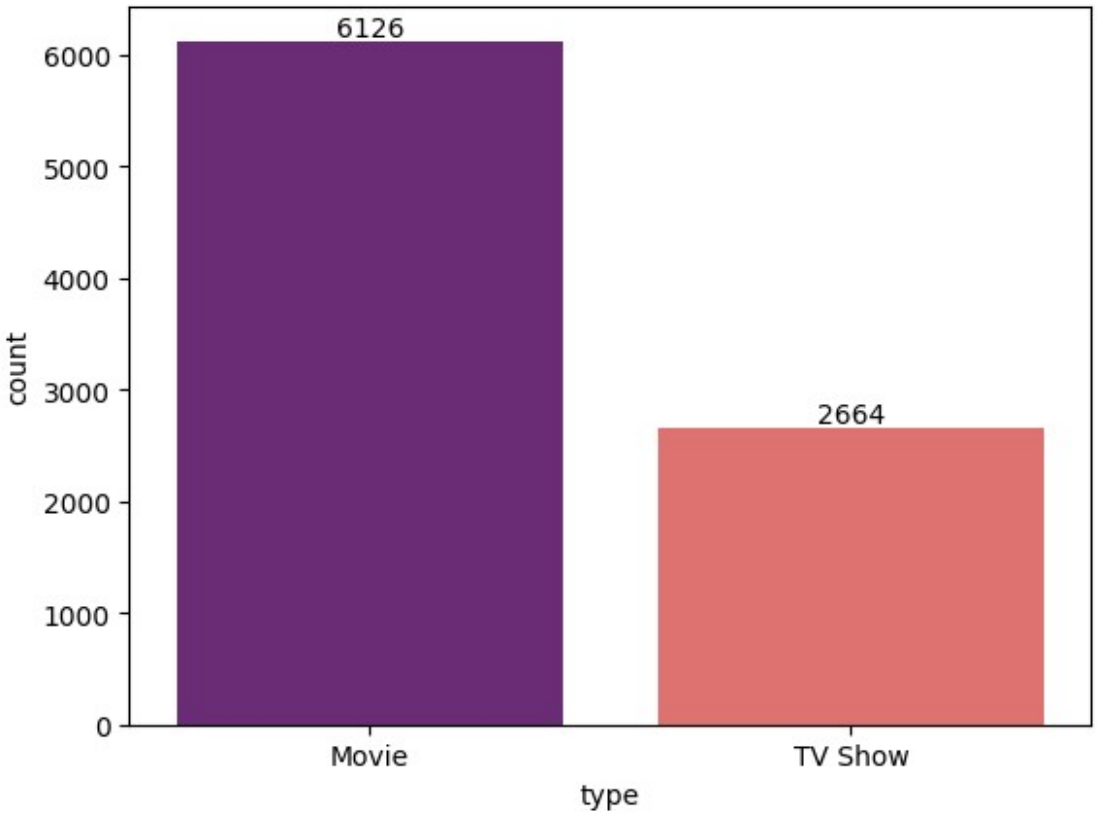
Content Type Distribution(TV Shows VS Movies)

```
df['type'].value_counts()

type
Movie      6126
TV Show    2664
Name: count, dtype: int64

type_counts=df['type'].value_counts()

sns.countplot(x='type',hue='type',data=df,palette='magma')
plt.show()
```



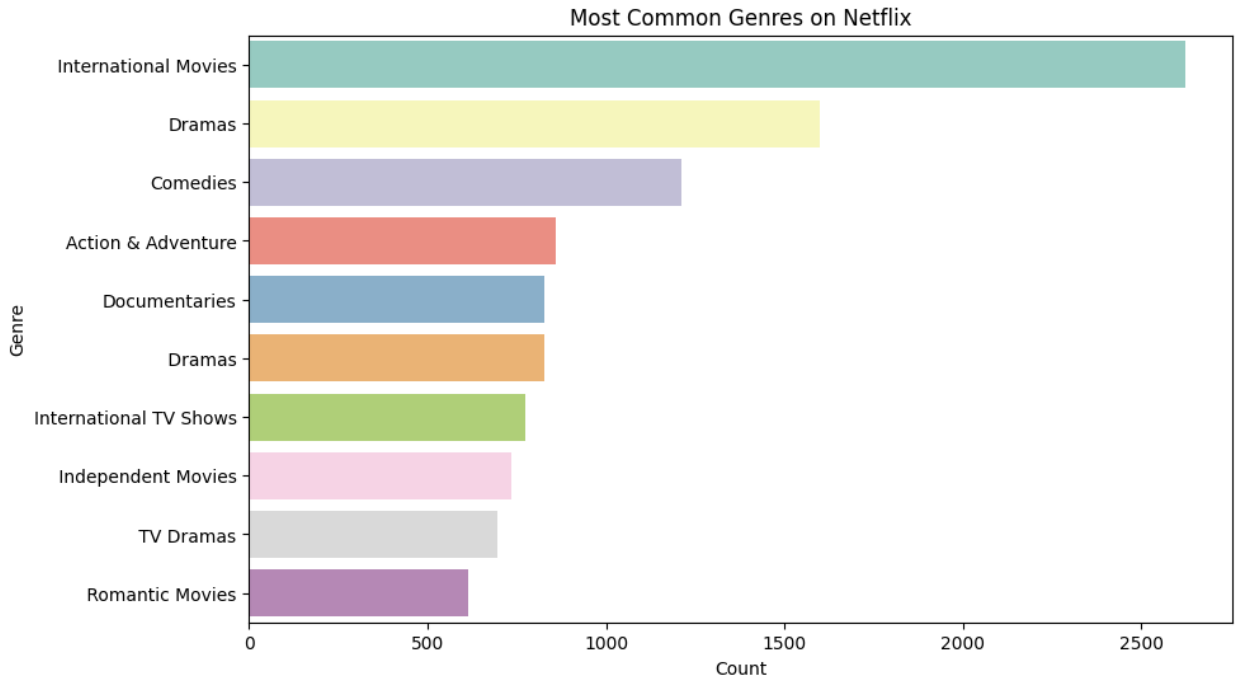

```
ax=sns.countplot(x='type',hue='type',data=df,palette='magma')
```

```
for i in ax.containers:
    ax.bar_label(i)
```



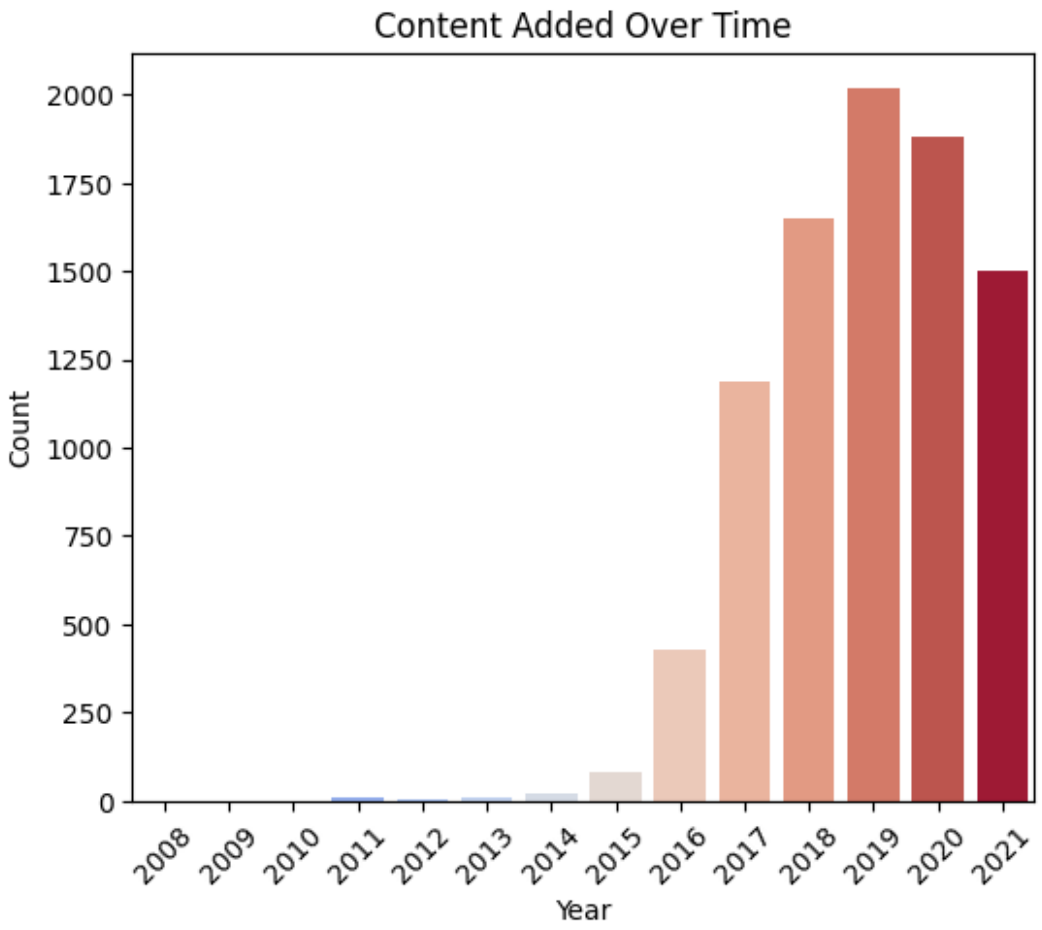
Most Common Genres On Netflix

```
df['genres'] = df['listed_in'].apply(lambda x: x.split(','))
all_genres = sum(df['genres'], [])
genre_counts = pd.Series(all_genres).value_counts().head(10)

plt.figure(figsize=(10, 6))
sns.barplot(x=genre_counts.values,y=genre_counts.index,hue=genre_count
s.index ,palette='Set3',legend=False,dodge=False)
plt.title('Most Common Genres on Netflix')
plt.xlabel('Count')
plt.ylabel('Genre')
plt.show()
```



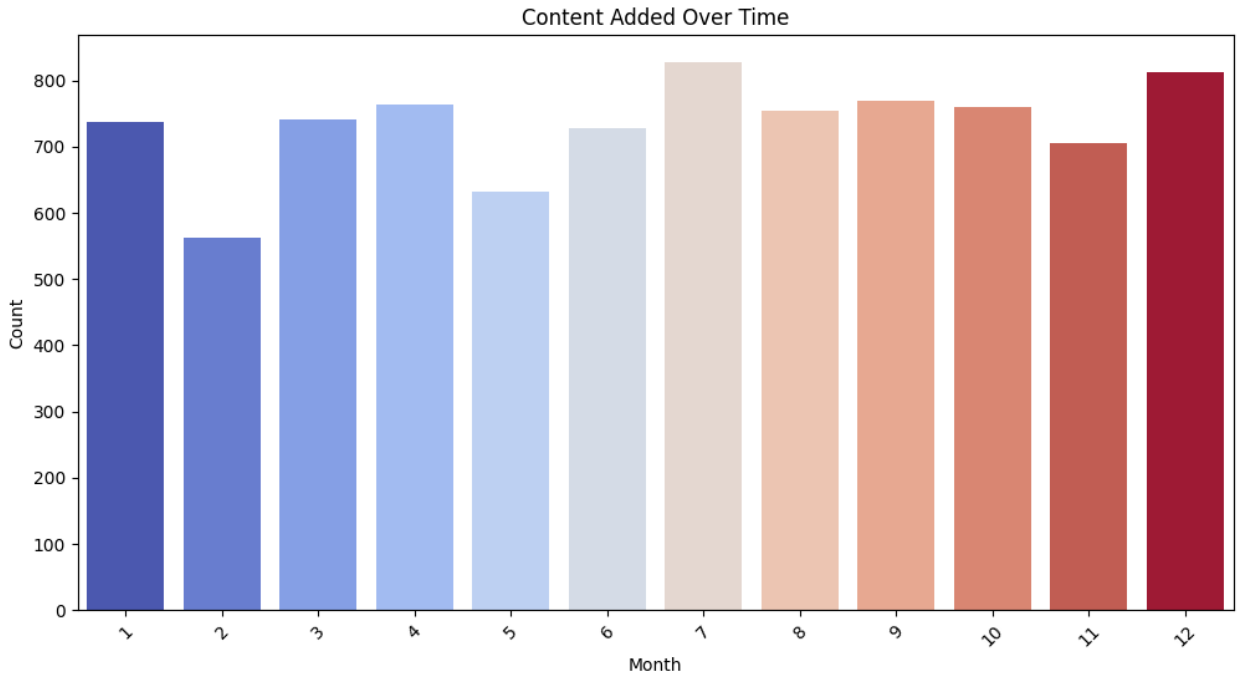
Content Added Over Time Yearly

```
df['year_added'] = df['added_date'].dt.year
df['month_added'] = df['added_date'].dt.month

plt.figure(figsize=(6, 5))
sns.countplot(x='year_added',hue='year_added',data=df,
palette='coolwarm',legend=False,dodge=False)
plt.title('Content Added Over Time')
plt.xlabel('Year')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.show()
```



Content Added Over Time Monthly

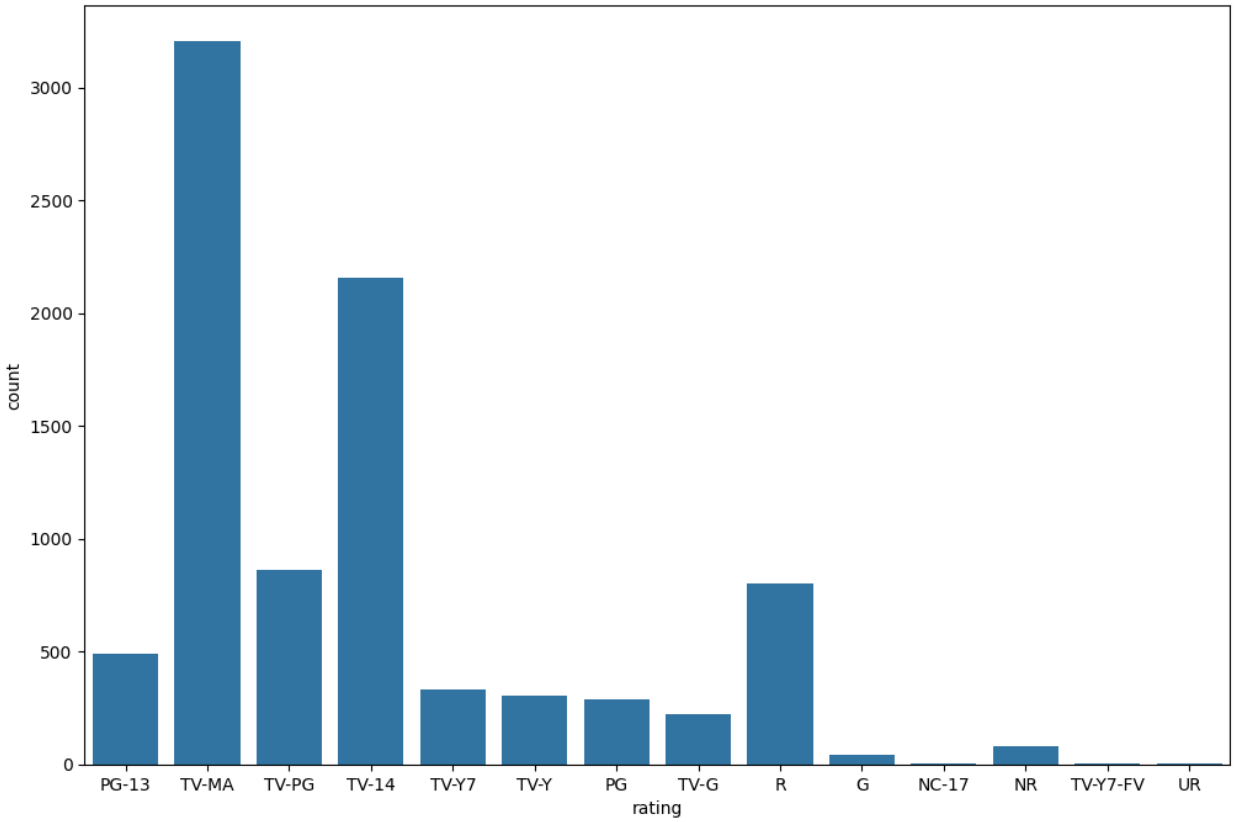
```
plt.figure(figsize=(12,6))
sns.countplot(x='month_added',hue='month_added',data=df,
palette='coolwarm',legend=False,dodge=False)
plt.title('Content Added Over Time')
plt.xlabel('Month')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.show()
```



Rating Count

```
plt.figure(figsize = (12,8))
sns.countplot(x='rating',data = df)
```

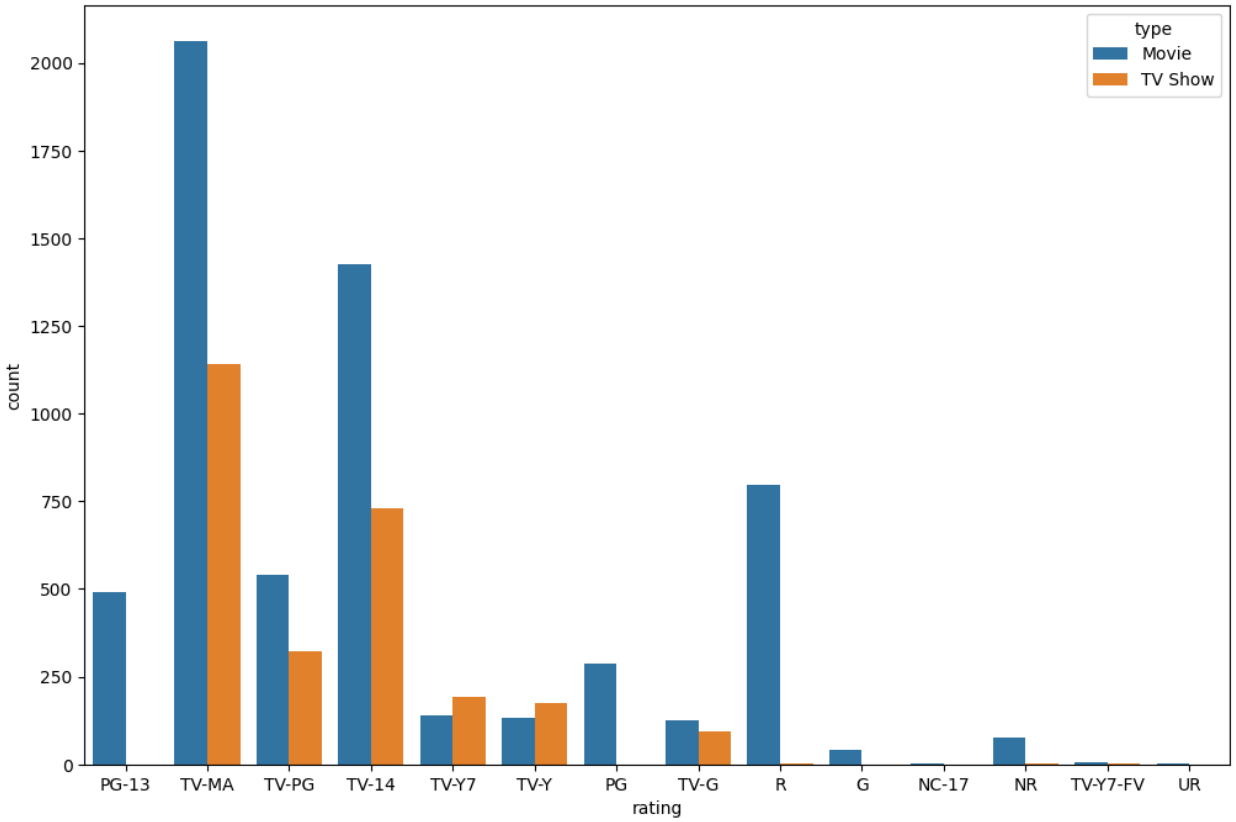
```
<Axes: xlabel='rating', ylabel='count'>
```



Rating Comparison Between TV Shows vs Movies

```
plt.figure(figsize = (12,8))
sns.countplot(x='rating',data = df,hue='type')
```

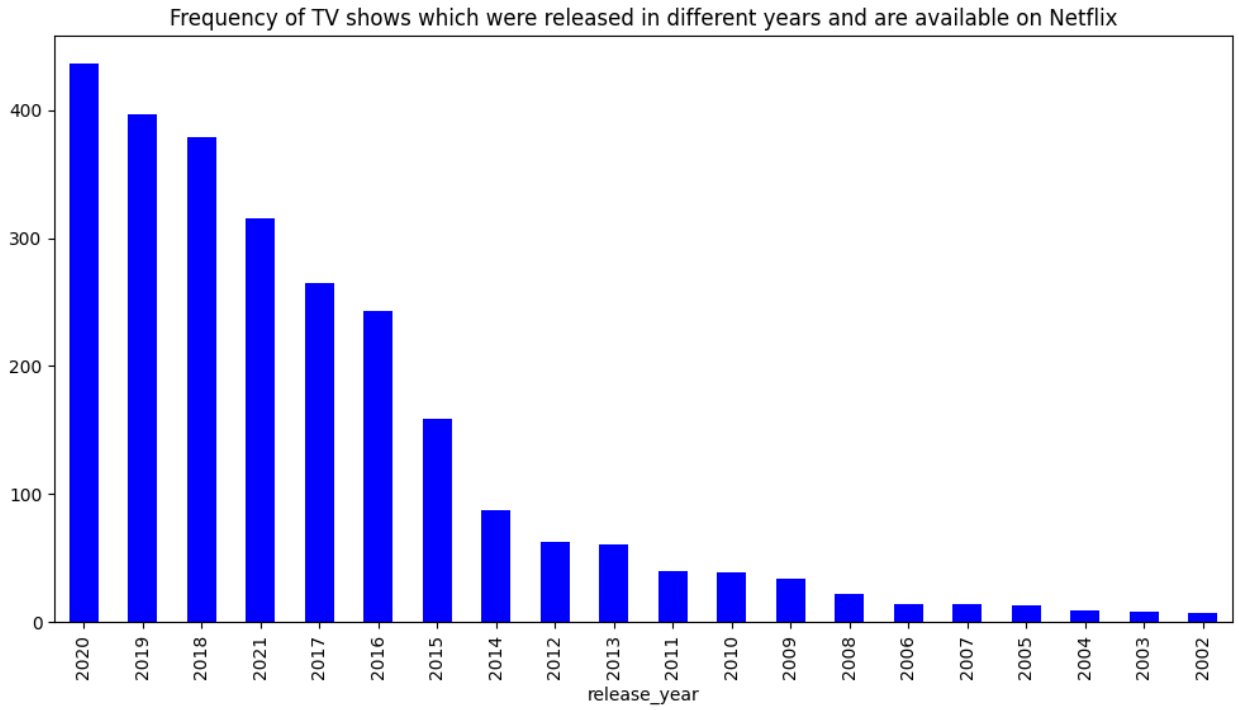
```
<Axes: xlabel='rating', ylabel='count'>
```



Frequency of TV Shows

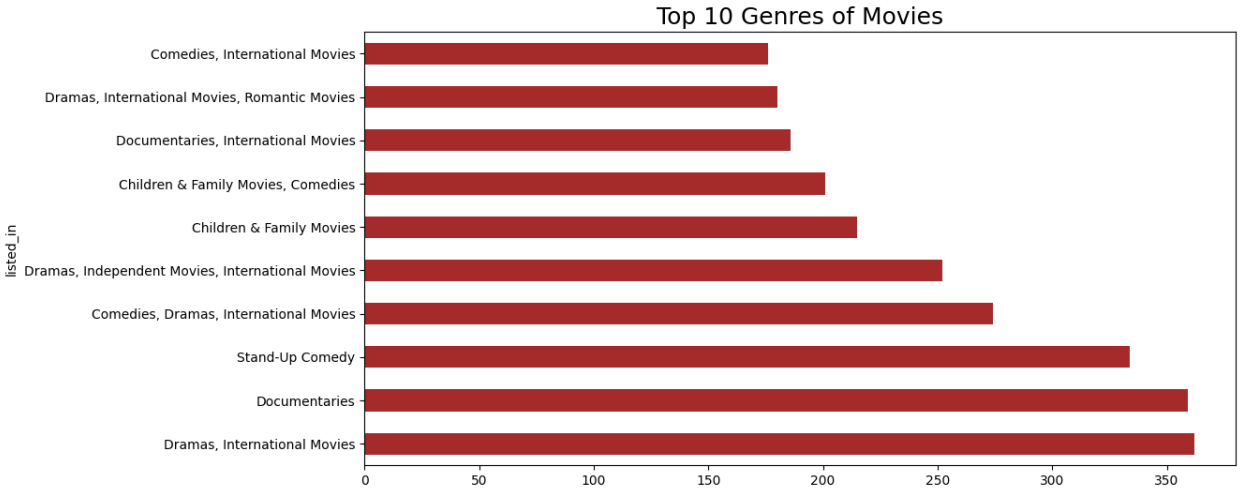
```
plt.figure(figsize=(12,6))
df[df["type"]=="TV Show"]["release_year"].value_counts()
[:20].plot(kind="bar",color="Blue")
plt.title("Frequency of TV shows which were released in different
years and are available on Netflix")
```

```
Text(0.5, 1.0, 'Frequency of TV shows which were released in different
years and are available on Netflix')
```



Top 10 Genres of Movies

```
plt.figure(figsize=(12,6))
df[df["type"]=="Movie"]["listed_in"].value_counts()
[:10].plot(kind="barh",color="brown")
plt.title("Top 10 Genres of Movies",size=18)

Text(0.5, 1.0, 'Top 10 Genres of Movies')
```



Top 10 Genres of TV Shows

plt.figure(figsize=(12,6)) df[df["type"]=="TV Show"]["listed_in"].value_counts()
[:10].plot(kind="barh",color="brown") plt.title("Top 10 Genres of TV Shows",size=18)

WordCloud

```
from wordcloud import WordCloud

movie_titles = df[df['type'] == 'Movie']['title']
wordcloud=WordCloud(width=1000,height=600,background_color='black').ge
nerate(' '.join(movie_titles))

plt.figure(figsize=(8, 5))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.show()
```

In conclusion, using Python libraries, we had taken valuable insights into the content on Netflix.These insights can help Netflix make better decisions about the content they add to their platform and how they market it to their users.Data analysis is a powerful tool that can be used in any industry to gain insights and make informed decisions. By using Python and data analysis techniques, we can gain a deeper understanding of any dataset and make data-driven decisions.