In [2]:
```
import numpy as np
```

In [3]:
```
import pandas as pd
```

In [46]:
```
import plotly.io as pio
pio.renderers.default ='notebook'
```

In [4]:
```
import plotly.express as px
```

In [5]:
```
!pip install textblob
```

```
Defaulting to user installation because normal site-packages is not writeable
Collecting textblob
  Downloading textblob-0.17.1-py2.py3-none-any.whl (636 kB)
     ------------------------------------ 636.8/636.8 kB 145.8 kB/s eta 0:00:00
Requirement already satisfied: nltk>=3.1 in c:\programdata\anaconda3\lib\site-packages (from textblob) (3.7)
Requirement already satisfied: tqdm in c:\programdata\anaconda3\lib\site-packages (from nltk>=3.1->textblob) (4.64.1)
Requirement already satisfied: joblib in c:\programdata\anaconda3\lib\site-packages (from nltk>=3.1->textblob) (1.1.0)
Requirement already satisfied: click in c:\programdata\anaconda3\lib\site-packages (from nltk>=3.1->textblob) (8.0.4)
Requirement already satisfied: regex>=2021.8.3 in c:\programdata\anaconda3\lib\site-packages (from nltk>=3.1->textblob) (2022.7.9)
Requirement already satisfied: colorama in c:\programdata\anaconda3\lib\site-packages (from click->nltk>=3.1->textblob) (0.4.5)
Installing collected packages: textblob
Successfully installed textblob-0.17.1
```

In [7]:
```
from textblob import TextBlob
```

In [5]:
```
!pip install Pyppeteer
!pip install nbconvert
```

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: Pyppeteer in c:\users\karan\appdata\roaming\python\python39\site-packages (1.0.2)
Requirement already satisfied: urllib3<2.0.0,>=1.25.8 in c:\programdata\anaconda3\lib\site-packages (from Pyppeteer) (1.26.11)
Requirement already satisfied: importlib-metadata>=1.4 in c:\programdata\anaconda3\lib\site-packages (from Pyppeteer) (4.11.3)
Requirement already satisfied: websockets<11.0,>=10.0 in c:\users\karan\appdata\roaming\python\python39\site-packages (from Pyppeteer) (10.4)
Requirement already satisfied: appdirs<2.0.0,>=1.4.3 in c:\programdata\anaconda3\lib\site-packages (from Pyppeteer) (1.4.4)
Requirement already satisfied: tqdm<5.0.0,>=4.42.1 in c:\programdata\anaconda3\lib\site-packages (from Pyppeteer) (4.64.1)
Requirement already satisfied: pyee<9.0.0,>=8.1.0 in c:\users\karan\appdata\roaming\python\python39\site-packages (from Pyppeteer) (8.2.2)
Requirement already satisfied: certifi>=2021 in c:\programdata\anaconda3\lib\site-packages (from Pyppeteer) (2022.9.14)
Requirement already satisfied: zipp>=0.5 in c:\programdata\anaconda3\lib\site-packages (from importlib-metadata>=1.4->Pyppeteer) (3.8.0)
Requirement already satisfied: colorama in c:\programdata\anaconda3\lib\site-packages (from tqdm<5.0.0,>=4.42.1->Pyppeteer) (0.4.5)
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: nbconvert in c:\programdata\anaconda3\lib\site-packages (6.4.4)
Requirement already satisfied: jupyter-core in c:\programdata\anaconda3\lib\site-packages (from nbconvert) (4.11.1)
Requirement already satisfied: nbformat>=4.4 in c:\programdata\anaconda3\lib\site-packages (from nbconvert) (5.5.0)
Requirement already satisfied: jupyterlab-pygments in c:\programdata\anaconda3\lib\site-packages (from nbconvert) (0.1.2)
Requirement already satisfied: beautifulsoup4 in c:\programdata\anaconda3\lib\site-packages (from nbconvert) (4.11.1)
Requirement already satisfied: bleach in c:\programdata\anaconda3\lib\site-packages (from nbconvert) (4.1.0)
Requirement already satisfied: entrypoints>=0.2.2 in c:\programdata\anaconda3\lib\site-packages (from nbconvert) (0.4)
Requirement already satisfied: defusedxml in c:\programdata\anaconda3\lib\site-packages (from nbconvert) (0.7.1)
Requirement already satisfied: testpath in c:\programdata\anaconda3\lib\site-packages (from nbconvert) (0.6.0)
Requirement already satisfied: pygments>=2.4.1 in c:\programdata\anaconda3\lib\site-packages (from nbconvert) (2.11.2)
Requirement already satisfied: pandocfilters>=1.4.1 in c:\programdata\anaconda3\lib\site-packages (from nbconvert) (1.5.0)
Requirement already satisfied: jinja2>=2.4 in c:\programdata\anaconda3\lib\site-packages (from nbconvert) (2.11.3)
Requirement already satisfied: nbclient<0.6.0,>=0.5.0 in c:\programdata\anaconda3\lib\site-packages (from nbconvert) (0.5.13)
Requirement already satisfied: traitlets>=5.0 in c:\programdata\anaconda3\lib\site-packages (from nbconvert) (5.1.1)
Requirement already satisfied: mistune<2,>=0.8.1 in c:\programdata\anaconda3\lib\site-packages (from nbconvert) (0.8.4)
Requirement already satisfied: MarkupSafe>=0.23 in c:\programdata\anaconda3\lib\site-packages (from jinja2>=2.4->nbconvert) (2.0.1)
Requirement already satisfied: jupyter-client>=6.1.5 in c:\programdata\anaconda3\lib\site-packages (from nbclient<0.6.0,>=0.5.0->nbconvert) (7.3.4)
Requirement already satisfied: nest-asyncio in c:\programdata\anaconda3\lib\site-packages (from nbclient<0.6.0,>=0.5.0->nbconvert) (1.5.5)
Requirement already satisfied: fastjsonschema in c:\programdata\anaconda3\lib\site-packages (from nbformat>=4.4->nbconvert) (2.16.2)
Requirement already satisfied: jsonschema>=2.6 in c:\programdata\anaconda3\lib\site-packages (from nbformat>=4.4->nbconvert) (4.16.0)
Requirement already satisfied: soupsieve>1.2 in c:\programdata\anaconda3\lib\site-packages (from beautifulsoup4->nbconvert) (2.3.1)
Requirement already satisfied: webencodings in c:\programdata\anaconda3\lib\site-packages (from bleach->nbconvert) (0.5.1)
Requirement already satisfied: six>=1.9.0 in c:\programdata\anaconda3\lib\site-packages (from bleach->nbconvert) (1.16.0)
Requirement already satisfied: packaging in c:\programdata\anaconda3\lib\site-packages (from bleach->nbconvert) (21.3)
Requirement already satisfied: pywin32>=1.0 in c:\programdata\anaconda3\lib\site-packages (from jupyter-core->nbconvert) (302)
Requirement already satisfied: pyrsistent!=0.17.0,!=0.17.1,!=0.17.2,>=0.14.0 in c:\programdata\anaconda3\lib\site-packages (from jsonschema>=2.6->nbform
at>=4.4->nbconvert) (0.18.0)
Requirement already satisfied: attrs>=17.4.0 in c:\programdata\anaconda3\lib\site-packages (from jsonschema>=2.6->nbformat>=4.4->nbconvert) (21.4.0)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\programdata\anaconda3\lib\site-packages (from jupyter-client>=6.1.5->nbclient<0.6.0,>=0.5.0-
>nbconvert) (2.8.2)
Requirement already satisfied: tornado>=6.0 in c:\programdata\anaconda3\lib\site-packages (from jupyter-client>=6.1.5->nbclient<0.6.0,>=0.5.0->nbconver
t) (6.1)
Requirement already satisfied: pyzmq>=23.0 in c:\programdata\anaconda3\lib\site-packages (from jupyter-client>=6.1.5->nbclient<0.6.0,>=0.5.0->nbconvert)
(23.2.0)
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in c:\programdata\anaconda3\lib\site-packages (from packaging->bleach->nbconvert) (3.0.9)
```

In [8]:
```
dff=pd.read_csv("C:\\Users\\karan\\Data\\netflix_titles.csv")
```

In [9]:
```
dff.shape
```

Out[9]:
```
(8807, 12)
```

In [10]:
```
##there are 8807 rows, 12 columns
```

In [11]:
```
dff.columns
```

Out[11]:
```
Index(['show_id', 'type', 'title', 'director', 'cast', 'country', 'date_added',
       'release_year', 'rating', 'duration', 'listed_in', 'description'],
      dtype='object')
```

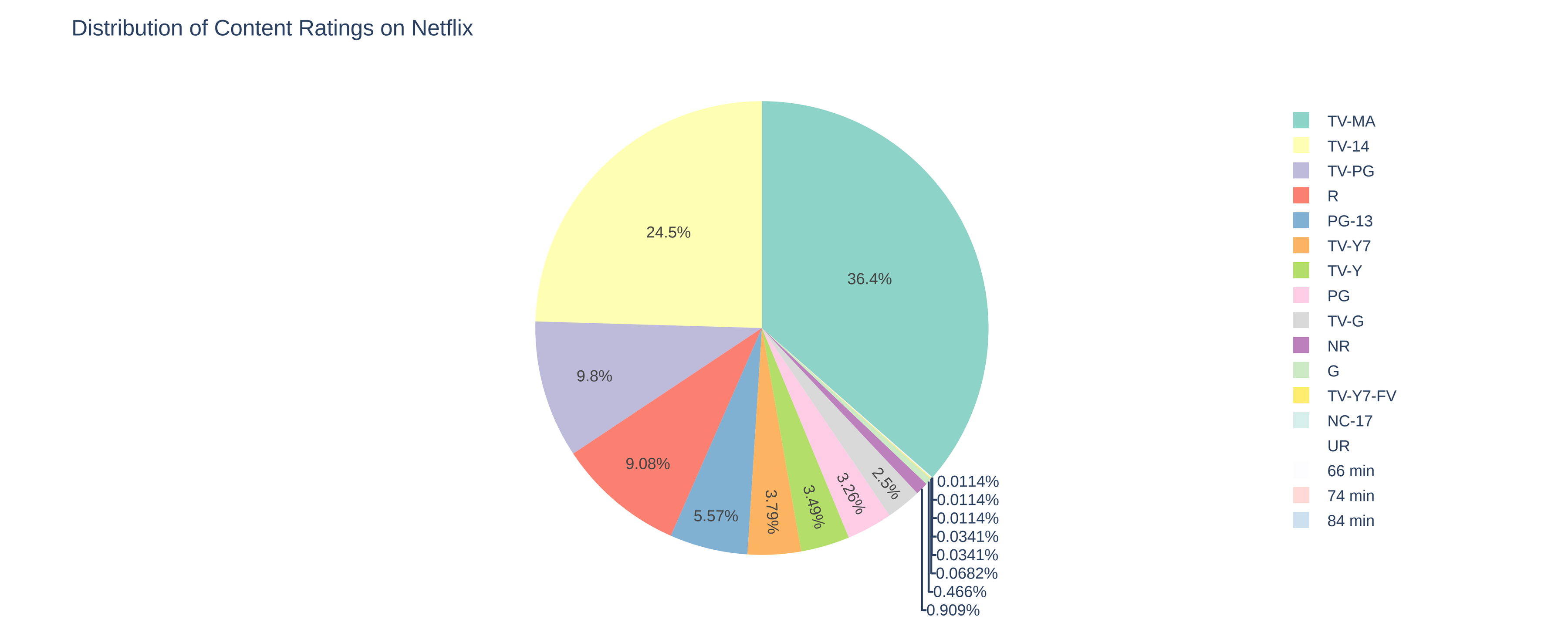
In [12]:
```
##let's begin with distribution of content rating
```

In [13]:
```
z = dff.groupby(['rating']).size().reset_index(name='counts')
```

In [14]:
```
print(z)
```

```
      rating  counts
0     66 min       1
1     74 min       1
2     84 min       1
3          G      41
4      NC-17       3
5         NR      80
6         PG     287
7      PG-13     490
8          R     799
9      TV-14    2160
10      TV-G     220
11     TV-MA    3207
12     TV-PG     863
13      TV-Y     307
14     TV-Y7     334
15  TV-Y7-FV       6
16        UR       3
```

In [15]:
```
pieChart = px.pie(z, values='counts', names='rating',
                  title='Distribution of Content Ratings on Netflix',
                  color_discrete_sequence=px.colors.qualitative.Set3)
pieChart.show()
```



In [16]:
```
##The graph above shows that the majority of content on Netflix is categorized as “TV-MA”, which means that most of the content available on Netflix is i
```

In [17]:
```
#Top 5 Actors and Directors:
##let’s see the top 5 successful directors on this platform:
```

In [18]:
```
dff['director']=dff['director'].fillna('No Director Specified')
filtered_directors=pd.DataFrame()
```

In [19]:
```
print(filtered_directors.head())
```

```
Empty DataFrame
Columns: []
Index: []
```

In [20]:
```
filtered_directors=dff['director'].str.split(',',expand=True).stack()
filtered_directors=filtered_directors.to_frame()
```

In [21]:
```
print(filtered_directors.head())
```

```
                        0
0 0       Kirsten Johnson
1 0  No Director Specified
2 0       Julien Leclercq
3 0  No Director Specified
4 0  No Director Specified
```

In [22]:
```
filtered_directors.columns=['Director']
```

In [23]:
```
print(filtered_directors.head())
```

```
                 Director
0 0       Kirsten Johnson
1 0  No Director Specified
2 0       Julien Leclercq
3 0  No Director Specified
4 0  No Director Specified
```

In [25]:
```
directors=filtered_directors.groupby(['Director']).size().reset_index(name='Total Content')
```

In [26]:
```
print(directors.head())
```

```
                  Director  Total Content
0           Aaron Moorhead              2
1              Aaron Woolf              1
2   Abbas Alibhai Burmawalla              1
3           Abdullah Al Noor              1
4        Abhinav Shiv Tiwari              1
```

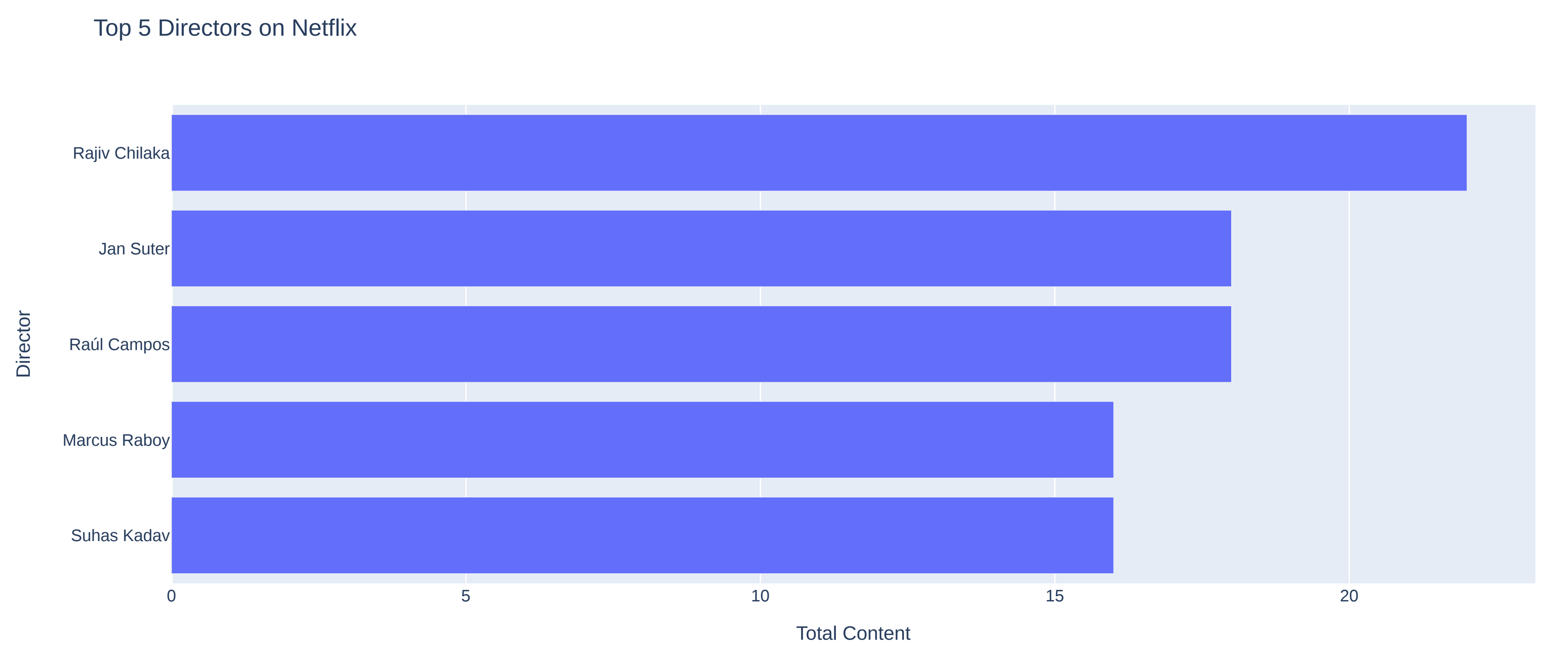
In [27]:
```
directors=directors[directors.Director !='No Director Specified']
```

In [28]:
```
directors=directors.sort_values(by=['Total Content'],ascending=False)
```

In [29]:
```
directorsTop5=directors.head()
```

In [30]:
```
directorsTop5=directorsTop5.sort_values(by=['Total Content'])
```

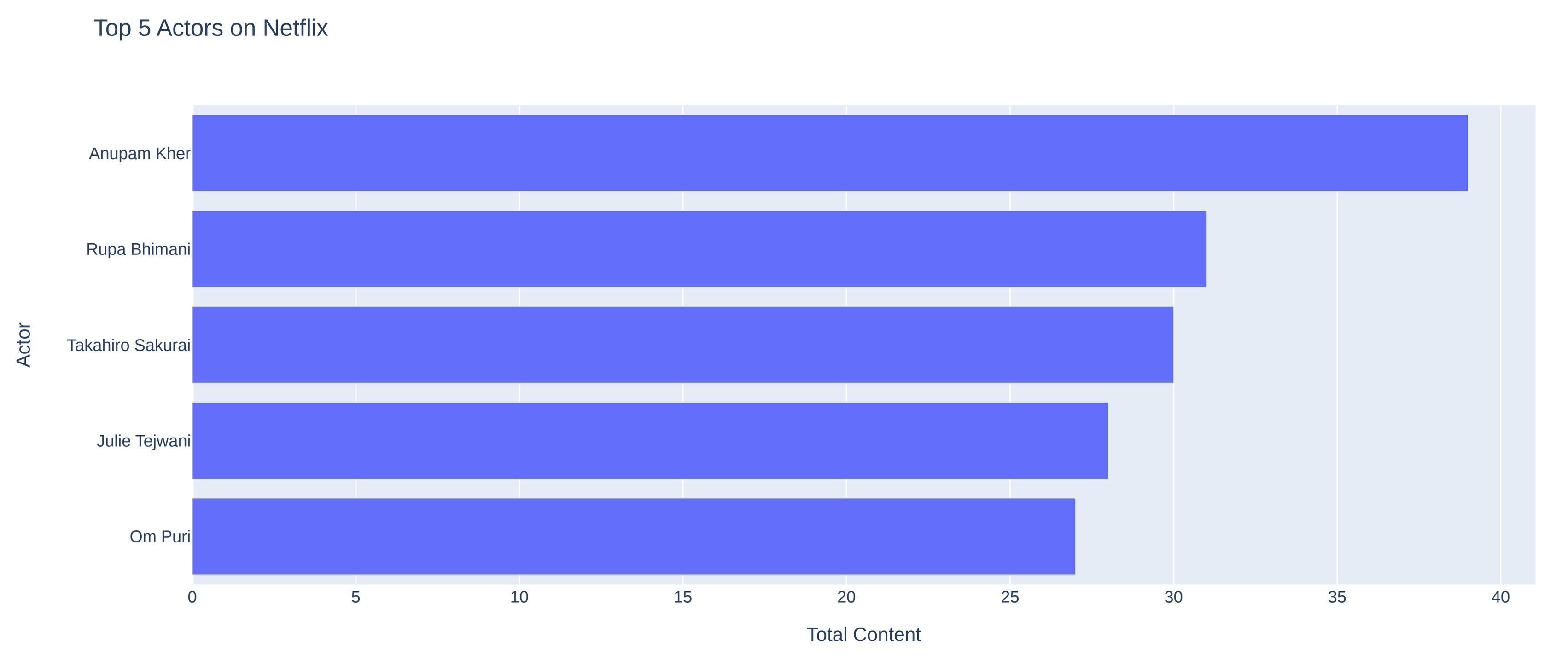
In [31]:
```
fig1=px.bar(directorsTop5,x='Total Content',y='Director',title='Top 5 Directors on Netflix')
fig1.show()
```



In [32]:
```
##Top 5 successful actors on this platform:
```

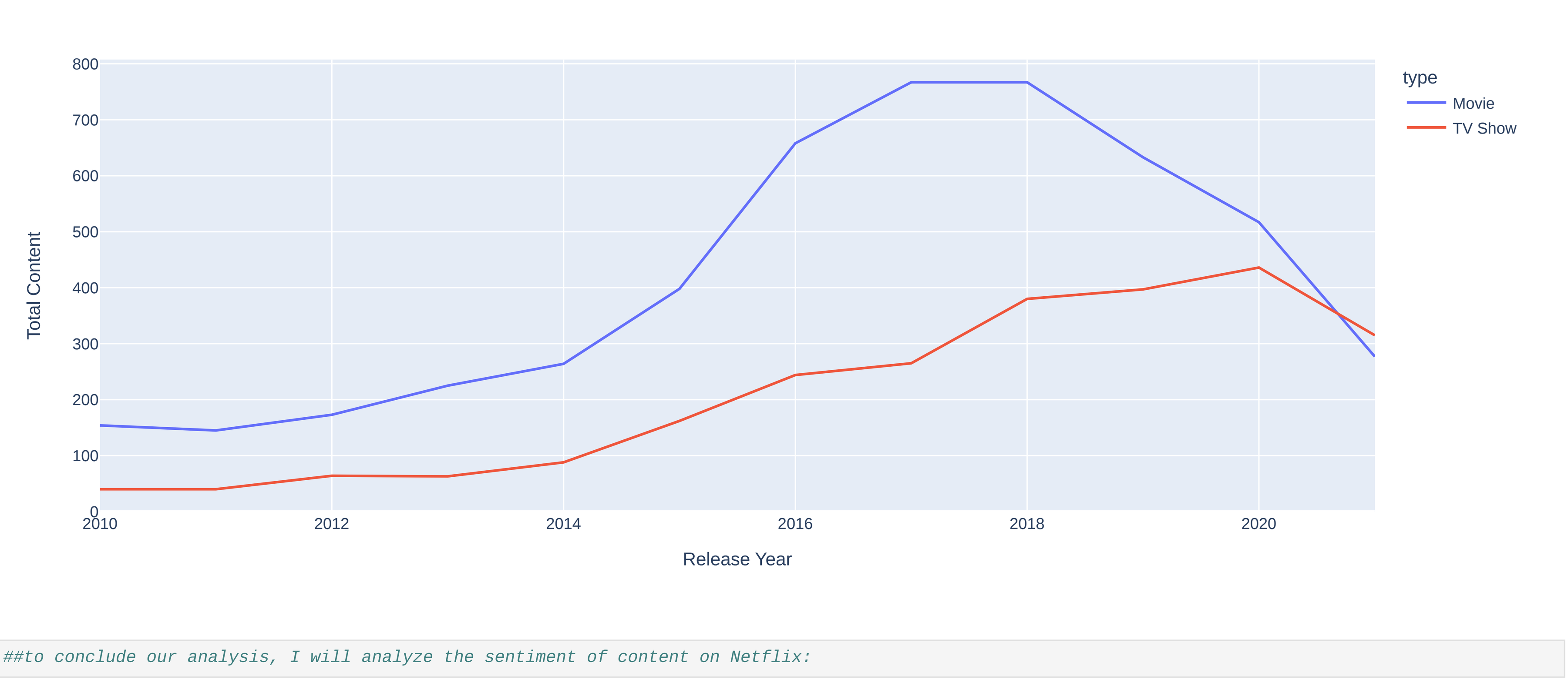
In [33]:
```
dff['cast']=dff['cast'].fillna('No Cast Specified')
```

In [34]:
```
filtered_cast=pd.DataFrame()
filtered_cast=dff['cast'].str.split(',',expand=True).stack()
filtered_cast=filtered_cast.to_frame()
filtered_cast.columns=['Actor']
actors=filtered_cast.groupby(['Actor']).size().reset_index(name='Total Content')
actors=actors[actors.Actor !='No Cast Specified']
actors=actors.sort_values(by=['Total Content'],ascending=False)
actorsTop5=actors.head()
actorsTop5=actorsTop5.sort_values(by=['Total Content'])
fig2=px.bar(actorsTop5,x='Total Content',y='Actor', title='Top 5 Actors on Netflix')
fig2.show()
```



In [35]:
```
##The next thing to analyze from this data is the trend of production over the years on Netflix:
```

In [36]:
```
df1=dff[['type','release_year']]
df1=df1.rename(columns={"release_year": "Release Year"})
df2=df1.groupby(['Release Year','type']).size().reset_index(name='Total Content')
df2=df2[df2['Release Year']>=2010]
fig3 = px.line(df2, x="Release Year", y="Total Content", color='type',title='Trend of content produced over the years on Netflix')
fig3.show()
```

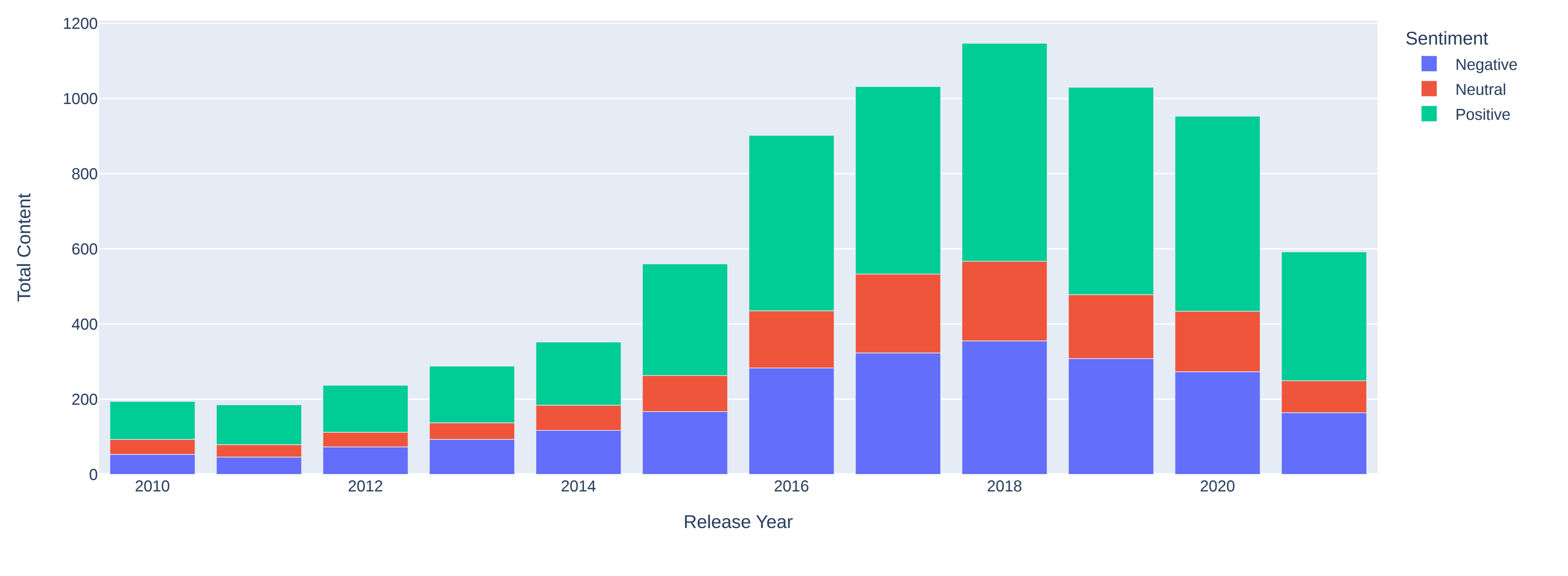


In [37]:
```
##to conclude our analysis, I will analyze the sentiment of content on Netflix:
```

In [45]:
```
dfx=dff[['release_year','description']]
dfx=dfx.rename(columns={'release_year':'Release Year'})
for index,row in dfx.iterrows():
    z=row['description']
    testimonial=TextBlob(z)
    p=testimonial.sentiment.polarity
    if p==0:
        sent='Neutral'
    elif p>0:
        sent='Positive'
    else:
        sent='Negative'
    dfx.loc[[index,2],'Sentiment']=sent


dfx=dfx.groupby(['Release Year','Sentiment']).size().reset_index(name='Total Content')

dfx=dfx[dfx['Release Year']>=2010]
fig4 = px.bar(dfx, x="Release Year", y="Total Content", color="Sentiment", title="Sentiment of content on Netflix")
fig4.show()
```



In [42]:
```
##So the above graph shows that the overall positive content is always greater than the neutral and negative content combined.
```

In [ ]:
```
```