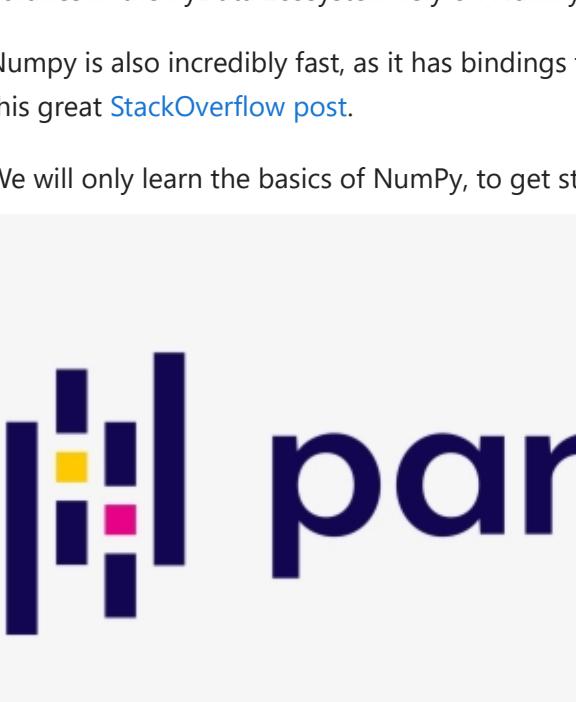


EDA on Netflix Data

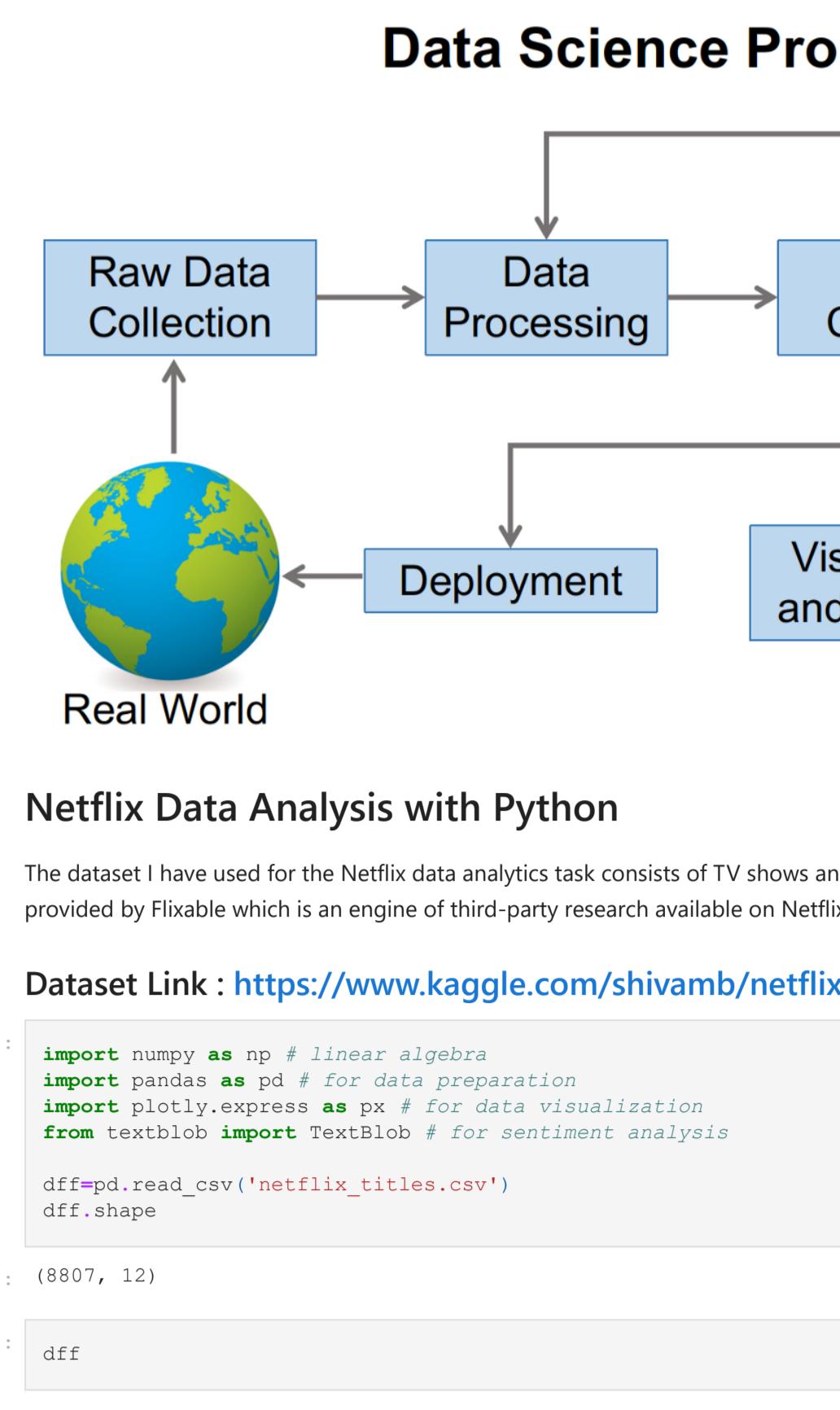


NumPy

NumPy (Numpy) is a linear Algebra Library for Python, the reason it is so important for Data Science with Python is that almost all of the libraries in the PyData Ecosystem rely on NumPy as one of their main building blocks.

Numpy is also incredibly fast, as it has bindings to C libraries. For more info on why you would want to use Arrays instead of lists, check out this great [StackOverflow post](#).

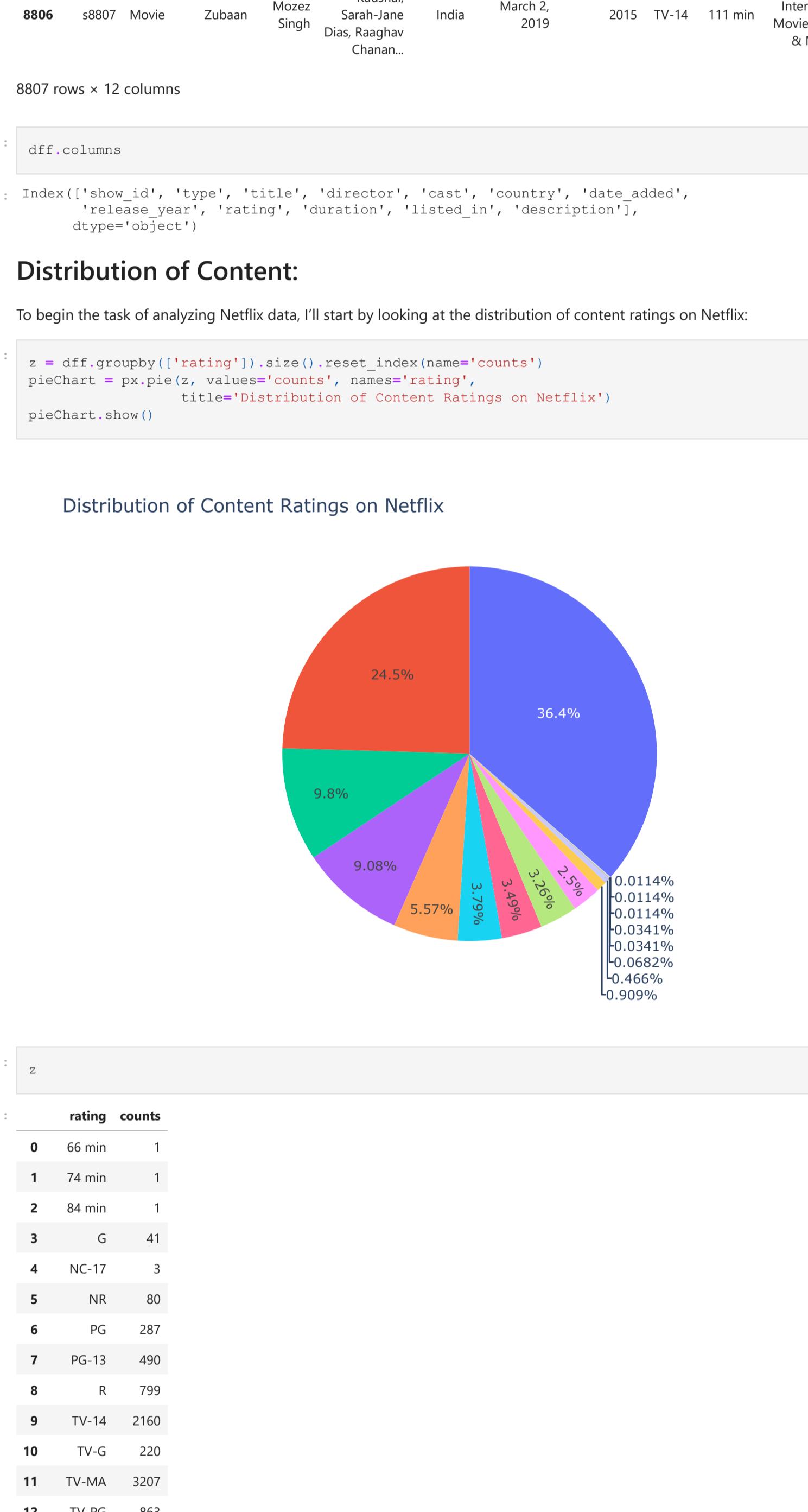
We will only learn the basics of NumPy, to get started we need to install it!



Introduction to Pandas

In this section of the course we will learn how to use pandas for data analysis. You can think of pandas as an extremely powerful version of Excel, with a lot more features. In this section of the course you should go through the notebooks in this order:

- Introduction to Pandas
- Series
- DataFrames
- Missing Data
- GroupBy
- Merging,Joining, and Concatenating
- Operations



plotly

plotly.py is an interactive, open-source, and browser-based graphing library for Python 🌐

Built on top of plotly.js, plotly.py is a high-level, declarative charting library. plotly.js ships with over 30 chart types, including scientific charts, 3D graphs, statistical charts, SVG maps, financial charts, and more.

plotly.py is MIT Licensed. Plotly graphs can be viewed in Jupyter notebooks, standalone HTML files, or hosted online using Chart Studio Cloud.

Data Science Process



Netflix Data Analysis with Python

The dataset I have used for the Netflix data analytics task consists of TV shows and movies streamed on Netflix as of 2021. The dataset is provided by Flixible which is an engine of third-party research available on Netflix.

Dataset Link : <https://www.kaggle.com/shivamb/netflix-shows>

```
In [1]: import numpy as np # linear algebra
import pandas as pd # for data preparation
import plotly.express as px # for data visualization
from textblob import Textblob # for sentiment analysis

df=pd.read_csv('netflix_titles.csv')
df.shape
```

```
Out[1]: (8807, 12)
```

```
In [2]: df
```

```
Out[2]: show_id type title director cast country date_added release_year rating duration listed_in description
0 s1 Movie Dick Johnson, Kristen Dead NaN United States September 25, 2021 2020 PG-13 90 min Documentaries father ends his life, film...
1 s2 TV Show Blood & Water Ama Quimata, Khosi Ngeuna, Makalane, Sall Thabane... 2021 TV-MA 2 Seasons International TV Dramas, TV Mysteries After crossing paths at a party, a Cape Town t...
2 s3 TV Show Ganglands Julien Leclercq Sam Bougila, Tracy Gobos, Samuel Jouy, Nabil... 2021 TV-MA 1 Season Crime TV Shows, International TV Shows, TV Act... To protect his family from a powerful drug lord...
3 s4 TV Show Jailbirds New Orleans NaN NaN NaN September 24, 2021 2021 TV-MA 1 Season Docuseries, Reality TV Feuds, flirtations and toilet talk go down...
4 s5 TV Show Factory Kota Kumar, Jitendra Ranjan Raj, Alam K. 2021 TV-MA 2 Seasons International TV Shows, Romantic TV Shows, TV ... In a city of coaching centers known to train...
... ...
8802 s8803 Movie Zodiac David Finch... Jake Gyllenhaal, Robert Downey J... 2007 R 158 min Cult Movies, Dramas, Thrillers A political cartoonist, a crime reporter and...
8803 s8804 TV Show Zombie Dumb NaN NaN NaN July 1, 2019 2018 TV-Y7 2 Seasons Kids' TV Shows, TV Comedies, Horror Movies While living alone in a spooky town, a...
8804 s8805 Movie Zombieland Ruben Fleischer Jesse Eisenberg, Woody Harrelson, Emma Stone, ... 2009 R 88 min Comedies, Horror Movies Looking to survive in a world taken over by zo...
8805 s8806 Movie Zoom Peter Hewitt Tim Allen, Courtney Cox, Chevy Chase, Kate... 2006 PG 88 min Family Movies, Comedies Dragged from civilian life, a former superhero...
8806 s8807 Movie Zubaan Mozez Singh Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan... 2015 TV-14 111 min Dramas, International Movies, Music & Musicals A scrappy but poor boy works his way into a...
8807 rows x 12 columns
```

```
In [3]: df.columns
```

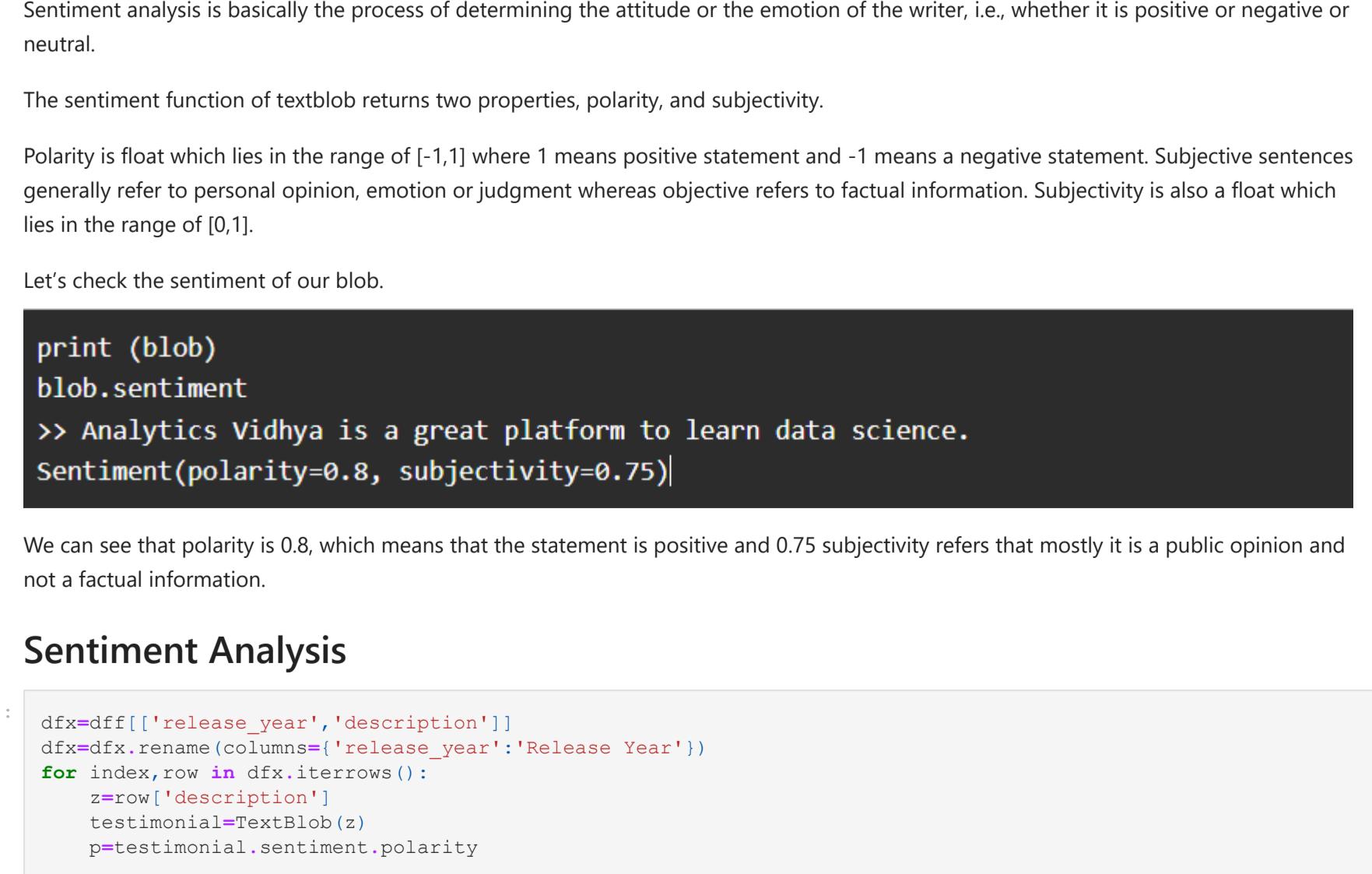
```
Out[3]: Index(['show_id', 'type', 'title', 'director', 'cast', 'country', 'date_added', 'release_year', 'rating', 'duration', 'listed_in', 'description'], dtype='object')
```

Distribution of Content:

To begin the task of analyzing Netflix data, I'll start by looking at the distribution of content ratings on Netflix.

```
In [4]: z = df.groupby(['rating']).size().reset_index(name='counts')
pieChart = px.pie(z, values='counts', names='rating', title='Distribution of Content Ratings on Netflix')
pieChart.show()
```

Distribution of Content Ratings on Netflix



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 intended for viewing by mature and adult audiences.

Top 5 Actors and Directors:

Now let's see the top 5 successful directors on this platform:

```
In [6]: df['director'].dropna().fillna('No Director Specified')
filteredDirectors=pd.DataFrame()
filteredDirectors=df['director'].str.split(',',expand=True).stack()
filteredDirectors=filteredDirectors.to_frame()
filteredDirectors.columns=['Director']
directors=df[['director']].groupby(['Director']).size().reset_index(name='Total Content')
directors=directors[directors['Director']!='No Director Specified']
directors=directors.sort_values(by='Total Content',ascending=False)
directorsTop5=directors.head(5).sort_values(by='Total Content')
directorsTop5=directorsTop5[['Director','Total Content']]
fig=px.bar(directorsTop5,x='Total Content',y='Director',title='Top 5 Directors on Netflix')
fig.show()
```

Top 5 Directors on Netflix

From the above graph it is derived that the top 5 directors on this platform are:

- Raul Campos
- Jan Suter
- Jay Karas
- Marcus Raboy
- Jay Chapman

Now let's have a look at the top 5 successful actors on this platform:

```
In [7]: df['cast'].dropna().fillna('No Cast Specified')
filteredCast=df['cast'].str.split(',',expand=True).stack()
filteredCast=filteredCast.to_frame()
filteredCast.columns=['Actor']
actors=df[['actor']].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(5).sort_values(by='Total Content')
actorsTop5=actorsTop5[['Actor','Total Content']]
fig=px.bar(actorsTop5,x='Total Content',y='Actor',title='Top 5 Actors on Netflix')
fig.show()
```

Top 5 Actors on Netflix

From the above plot, it is derived that the top 5 actors on Netflix are:

- Anupam Kher
- Om Puri
- Shah Rukh Khan
- Takhiro Sakurai
- Boman Irani

Analyzing Content on Netflix:

The next thing to analyze from this data is the trend of production over the years on Netflix:

```
In [23]: df=df[['type','release_year']]
df1=df.rename(columns={'release_year':'Release Year'})
df2=df1.groupby(['Release Year']).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()
```

Trend of content produced over the years on Netflix

The above line graph shows that there has been a decline in the production of the content for both movies and other shows since 2018.

At last, to conclude our analysis, I will analyze the sentiment of content on Netflix:

TextBlob

Natural Language Processing for Beginners: Using TextBlob

1. About TextBlob?
2. Setting up the System
3. Having a go at NLP tasks using TextBlob

1. Tokenization
2. Noun phrase extraction
3. POS-Tagging
4. Words inflection and lemmatization
5. N-grams
6. Sentiment Analysis

4. Other cool things to do with TextBlob

1. Spelling correction
2. Creating a short summary of a text
3. Translation and language detection

5. Text classification using TextBlob

6. Pros and Cons

About TextBlob?

TextBlob is a python library and offers a simple API to access its methods and perform basic NLP tasks.

A good thing about TextBlob is that they are just like python strings. So, you can transform and play with it same like we did in python. Below, I have shown you below some basic tasks. Don't worry about the syntax, it is just to give you an intuition about how much-related TextBlob is to Python strings.

6 - Sentiment Analysis

Sentiment analysis is basically the process of determining the attitude or the emotion of the writer, i.e., whether it is positive or negative or neutral.

The sentiment function of textblob returns two properties, polarity, and subjectivity.

Polarity is float which lies in the range of [-1,1] where 1 means positive statement and -1 means a negative statement. Subjective sentences generally refer to personal opinion, emotion or judgment whereas objective refers to factual information. Subjectivity is also a float which lies in the range of [0,1].

Let's check the sentiment of our blob.

```
print(blob)
blob.sentiment
>> Analytics Vidhya is a great platform to learn data science.
Sentiment(polarity=0.8, subjectivity=0.75)
```

We can see that polarity is 0.8, which means that the statement is positive and 0.75 subjectivity refers that mostly it is a public opinion and not a factual information.

Now let's have a look at the top 5 successful actors on this platform:

- Anupam Kher
- Om Puri
- Shah Rukh Khan
- Takhiro Sakurai
- Boman Irani

From the above plot, it is derived that the top 5 actors on Netflix are:

- Raul Campos
- Jan Suter
- Jay Karas
- Marcus Raboy
- Jay Chapman

Now let's have a look at the top 5 successful actors on this platform:

1. Anupam Kher
2. Om Puri
3. Shah Rukh Khan
4. Takhiro Sakurai
5. Boman Irani

From the above plot, it is derived that the top 5 actors on Netflix are:

- Anupam Kher
- Om Puri
- Shah Rukh Khan
- Takhiro Sakurai
- Boman Irani

Analyzing Content on Netflix:

The next thing to analyze from this data is the trend of production over the years on Netflix:

```
In [23]: df=df[['type','release_year']]
df1=df.rename(columns={'release_year':'Release Year'})
df2=df1.groupby(['Release Year']).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()
```

Trend of content produced over the years on Netflix

The above line graph shows that there has been a decline in the production of the content for both movies and other shows since 2018.

At last, to conclude our analysis, I will analyze the sentiment of content on Netflix:

TextBlob

Natural Language Processing for Beginners: Using TextBlob

1. About TextBlob?
2. Setting up the System
3. Having a go at NLP tasks using TextBlob

1. Tokenization
2. Noun phrase extraction
3. POS-Tagging
4. Words inflection and lemmatization
5. N-grams
6. Sentiment Analysis

4. Other cool things to do with TextBlob

1. Spelling correction
2. Creating a short summary of a text
3. Translation and language detection

5. Text classification using TextBlob

6. Pros and Cons

About TextBlob?

TextBlob is a python library and offers a simple API to access its methods and perform basic NLP tasks.

A good thing about TextBlob is that they are just like python strings. So, you can transform and play with it same like we did in python. Below, I have shown you below some basic tasks. Don't worry about the syntax, it is just to give you an intuition about how much-related TextBlob is to Python strings.

6 - Sentiment Analysis

Sentiment analysis is basically the process of determining the attitude or the emotion of the writer, i.e., whether it is positive or negative or neutral.

The sentiment function of textblob returns two properties, polarity, and subjectivity.

Polarity is float which lies in the range of [-1,1] where 1 means positive statement and -1 means a negative statement. Subjective sentences generally refer to personal opinion, emotion or judgment whereas objective refers to factual information. Subjectivity is also a float which lies in the range of [0,1].

Let's check the sentiment of our blob.

```
print(blob)
blob.sentiment
>> Analytics Vidhya is a great platform to learn data science.
Sentiment(polarity=0.8, subjectivity=0.75)
```

We can see that polarity is 0.8, which means that the statement is positive and 0.75 subjectivity refers that mostly it is a public opinion and not a factual information.

Now let's have a look at the top 5 successful actors on this platform:

- Anupam Kher
- Om Puri
- Shah Rukh Khan
- Takhiro Sakurai
- Boman Irani

From the above plot, it is derived that the top 5 actors on Netflix are:

- Raul Campos
- Jan Suter
- Jay Karas
- Marcus Raboy
- Jay Chapman

Now let's have a look at the top 5 successful actors on this platform:

1. Anupam Kher
2. Om Puri
3. Shah Ruk