

In [1]:

```
import numpy as np
import pandas as pd
```

In [2]:

```
df = pd.read_csv("netflix_titles.csv", sep=',')
```

In [3]:

```
df.head()
```

Out[3]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration
0	s1	TV Show	3%	NaN	João Miguel, Bianca Comparato, Michel Gomes, R...	Brazil	August 14, 2020	2020	TV-MA	Se
1	s2	Movie	7:19	Jorge Michel Grau	Demián Bichir, Héctor Bonilla, Oscar Serrano, ...	Mexico	December 23, 2016	2016	TV-MA	5
2	s3	Movie	23:59	Gilbert Chan	Tedd Chan, Stella Chung, Henley Hii, Lawrence ...	Singapore	December 20, 2018	2011	R	1
3	s4	Movie	9	Shane Acker	Elijah Wood, John C. Reilly, Jennifer Connelly...	United States	November 16, 2017	2009	PG-13	8
4	s5	Movie	21	Robert Luketic	Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar...	United States	January 1, 2020	2008	PG-13	12

In [4]:

```
del df['show_id']
```

In [5]:

```
df.head()
```

Out[5]:

	type	title	director	cast	country	date_added	release_year	rating	duration	
0	TV Show	3%	NaN	João Miguel, Bianca Comparato, Michel Gomes, R...	Brazil	August 14, 2020	2020	TV-MA	4 Seasons	Inte T Tv
1	Movie	7:19	Jorge Michel Grau	Demián Bichir, Héctor Bonilla, Oscar Serrano, ...	Mexico	December 23, 2016	2016	TV-MA	93 min	Inte
2	Movie	23:59	Gilbert Chan	Tedd Chan, Stella Chung, Henley Hii, Lawrence ...	Singapore	December 20, 2018	2011	R	78 min	Inte
3	Movie	9	Shane Acker	Elijah Wood, John C. Reilly, Jennifer Connelly...	United States	November 16, 2017	2009	PG-13	80 min	A Ind Mc
4	Movie	21	Robert Luketic	Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar...	United States	January 1, 2020	2008	PG-13	123 min	

In [6]:

```
df.describe()
```

Out[6]:

	release_year
count	7787.000000
mean	2013.932580
std	8.757395
min	1925.000000
25%	2013.000000
50%	2017.000000
75%	2018.000000
max	2021.000000

In [7]:

df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7787 entries, 0 to 7786
Data columns (total 11 columns):
 #   Column          Non-Null Count  Dtype
---  -
 0   type            7787 non-null   object
 1   title           7787 non-null   object
 2   director        5398 non-null   object
 3   cast            7069 non-null   object
 4   country         7280 non-null   object
 5   date_added      7777 non-null   object
 6   release_year    7787 non-null   int64
 7   rating          7780 non-null   object
 8   duration        7787 non-null   object
 9   listed_in       7787 non-null   object
10  description     7787 non-null   object
dtypes: int64(1), object(10)
memory usage: 669.3+ KB
```

In [8]:

print(df.head(20).to\_html())

```
<table border="1" class="dataframe">
  <thead>
    <tr style="text-align: right;">
      <th></th>
      <th>type</th>
      <th>title</th>
      <th>director</th>
      <th>cast</th>
      <th>country</th>
      <th>date_added</th>
      <th>release_year</th>
      <th>rating</th>
      <th>duration</th>
      <th>listed_in</th>
      <th>description</th>
    </tr>
  </thead>
  <tbody>
    <tr>
```

In [9]:

df[df['country']=='Brazil'].head(20).to\_html('netflix.html')

In [10]:

```
!pip install sqlalchemy
!pip install pymysql
```

Collecting sqlalchemy

Downloading SQLAlchemy-1.4.23-cp39-cp39-win\_amd64.whl (1.5 MB)

Collecting greenlet!=0.4.17

Downloading greenlet-1.1.1-cp39-cp39-win\_amd64.whl (96 kB)

Installing collected packages: greenlet, sqlalchemy

Successfully installed greenlet-1.1.1 sqlalchemy-1.4.23

WARNING: You are using pip version 21.2.3; however, version 21.2.4 is available.

You should consider upgrading via the 'D:\curso-data-science\venv\Scripts\python.exe -m pip install --upgrade pip' command.

Collecting pymysql

Downloading PyMySQL-1.0.2-py3-none-any.whl (43 kB)

Installing collected packages: pymysql

Successfully installed pymysql-1.0.2

WARNING: You are using pip version 21.2.3; however, version 21.2.4 is available.

You should consider upgrading via the 'D:\curso-data-science\venv\Scripts\python.exe -m pip install --upgrade pip' command.

In [11]:

```
from sqlalchemy import create_engine
import pymysql
```

In [12]:

```
db_connection = create_engine("mysql+pymysql://root:@localhost:3306/wef")
df = pd.read_sql('select * from posts', con=db_connection)
df[df['Id']==42].head()
```

```
-----
-
ConnectionRefusedError                                Traceback (most recent call last)
D:\curso-data-science\venv\lib\site-packages\pymysql\connections.py in connect(self, sock)
    612
--> 613             try:
    614                 sock = socket.create_connection(
                                (self.host, self.port), self.conne
ct_timeout, **kwargs

~\AppData\Local\Programs\Python\Python39\lib\socket.py in create_connection
n(address, timeout, source_address)
    843             try:
--> 844                 raise err
    845             finally:

~\AppData\Local\Programs\Python\Python39\lib\socket.py in create_connectio
n(address, timeout, source_address)
    841         sock = socket._create_connection(address,
                                timeout, source_address,
                                socket_options=socket_options)
```

In [13]:

!pip install openpyxl

Collecting openpyxl

WARNING: You are using pip version 21.2.3; however, version 21.2.4 is available.

You should consider upgrading via the 'D:\curso-data-science\venv\Scripts\python.exe -m pip install --upgrade pip' command.

Downloading openpyxl-3.0.7-py2.py3-none-any.whl (243 kB)

Collecting et-xmlfile

Downloading et\_xmlfile-1.1.0-py3-none-any.whl (4.7 kB)

Installing collected packages: et-xmlfile, openpyxl

Successfully installed et-xmlfile-1.1.0 openpyxl-3.0.7

In [14]:

df.head()

Out[14]:

	type	title	director	cast	country	date_added	release_year	rating	duration	
0	TV Show	3%	NaN	João Miguel, Bianca Comparato, Michel Gomes, R...	Brazil	August 14, 2020	2020	TV-MA	4 Seasons	Inte T TV
1	Movie	7:19	Jorge Michel Grau	Demián Bichir, Héctor Bonilla, Oscar Serrano, ...	Mexico	December 23, 2016	2016	TV-MA	93 min	Inte
2	Movie	23:59	Gilbert Chan	Tedd Chan, Stella Chung, Henley Hii, Lawrence ...	Singapore	December 20, 2018	2011	R	78 min	Inte
3	Movie	9	Shane Acker	Elijah Wood, John C. Reilly, Jennifer Connelly...	United States	November 16, 2017	2009	PG-13	80 min	A Ind Mc
4	Movie	21	Robert Luketic	Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar...	United States	January 1, 2020	2008	PG-13	123 min	

In [15]:

df

Out[15]:

	type	title	director	cast	country	date_added	release_year	rating	duratic
0	TV Show	3%	NaN	João Miguel, Bianca Comparato, Michel Gomes, R...	Brazil	August 14, 2020	2020	TV-MA	Seasor
1	Movie	7:19	Jorge Michel Grau	Demián Bichir, Héctor Bonilla, Oscar Serrano, ...	Mexico	December 23, 2016	2016	TV-MA	93 m
2	Movie	23:59	Gilbert Chan	Tedd Chan, Stella Chung, Henley Hii, Lawrence ...	Singapore	December 20, 2018	2011	R	78 m
3	Movie	9	Shane Acker	Elijah Wood, John C. Reilly, Jennifer Connelly...	United States	November 16, 2017	2009	PG-13	80 m
4	Movie	21	Robert Luketic	Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar...	United States	January 1, 2020	2008	PG-13	123 m
...	...	...	...	...	...	...	...	...	...
7782	Movie	Zozo	Josef Fares	Imad Creidi, Antoinette Turk, Elias Gergi, Car...	Sweden, Czech Republic, United Kingdom, Denmar...	October 19, 2020	2005	TV-MA	99 m
7783	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14	111 m
7784	Movie	Zulu Man in Japan	NaN	Nasty C	NaN	September 25, 2020	2019	TV-MA	44 m
7785	TV Show	Zumbo's Just Desserts	NaN	Adriano Zumbo, Rachel Khoo	Australia	October 31, 2020	2019	TV-PG	Seasc

	type	title	director	cast	country	date_added	release_year	rating	duration
7786	Movie	ZZ TOP: THAT LITTLE OL' BAND FROM TEXAS	Sam Dunn	NaN	United Kingdom, Canada, United States	March 1, 2020	2019	TV- MA	90 m

7787 rows × 11 columns



In [16]:

```
df.to_excel('db.xlsx')
```

In [17]:

```
!pip install lxml
```

Collecting lxml

Downloading lxml-4.6.3-cp39-cp39-win\_amd64.whl (3.5 MB)

Installing collected packages: lxml

Successfully installed lxml-4.6.3

WARNING: You are using pip version 21.2.3; however, version 21.2.4 is available.

You should consider upgrading via the 'D:\curso-data-science\venv\Scripts\python.exe -m pip install --upgrade pip' command.

In [18]:

```
df = pd.read_html("https://www.w3schools.com/html/html_tables.asp")
```

In [19]:

```
df
```

Out[19]:

	Company	Contact	Country
0	Alfreds Futterkiste	Maria Anders	Germany
1	Centro comercial Moctezuma	Francisco Chang	Mexico
2	Ernst Handel	Roland Mendel	Austria
3	Island Trading	Helen Bennett	UK
4	Laughing Bacchus Winecellars	Yoshi Tannamuri	Canada
5	Magazzini Alimentari Riuniti	Giovanni Rovelli	Italy,
	Tag	Description	
0	<table>	Defines a table	
1	<th>	Defines a header cell in a table	
2	<tr>	Defines a row in a table	
3	<td>	Defines a cell in a table	
4	<caption>	Defines a table caption	
5	<colgroup>	Specifies a group of one or more columns in a ...	
6	<col>	Specifies column properties for each column wi...	
7	<thead>	Groups the header content in a table	
8	<tbody>	Groups the body content in a table	
9	<tfoot>	Groups the footer content in a table]	

In [20]:

```
df[0]
```

Out[20]:

	Company	Contact	Country
0	Alfreds Futterkiste	Maria Anders	Germany
1	Centro comercial Moctezuma	Francisco Chang	Mexico
2	Ernst Handel	Roland Mendel	Austria
3	Island Trading	Helen Bennett	UK
4	Laughing Bacchus Winecellars	Yoshi Tannamuri	Canada
5	Magazzini Alimentari Riuniti	Giovanni Rovelli	Italy

In [ ]: