

Analise exploratoria de top 10 do catalogo Netflix

Analisaremos a seguir uma base de dados do catalogo Netflix, com intuito de entender quais produtos do se catalogo alcançaram o top 10 mais assistidos da plataforma para fins de estudo e aprendizado da biblioteca "pandas".

```
In [ ]: # Instalando e Importando a biblioteca de analise de dados pandas.  
%pip install pandas
```

```
Requirement already satisfied: pandas in d:\users\aluno\desktop\python_senai\venv\lib\site-packages (2.2.2)  
Requirement already satisfied: numpy>=1.26.0 in d:\users\aluno\desktop\python_senai\venv\lib\site-packages (from pandas) (2.1.1)  
Requirement already satisfied: python-dateutil>=2.8.2 in d:\users\aluno\desktop\python_senai\venv\lib\site-packages (from pandas) (2.9.0.post0)  
Requirement already satisfied: pytz>=2020.1 in d:\users\aluno\desktop\python_senai\venv\lib\site-packages (from pandas) (2024.1)  
Requirement already satisfied: tzdata>=2022.7 in d:\users\aluno\desktop\python_senai\venv\lib\site-packages (from pandas) (2024.1)  
Requirement already satisfied: six>=1.5 in d:\users\aluno\desktop\python_senai\venv\lib\site-packages (from python-dateutil>=2.8.2->pandas) (1.16.0)  
Note: you may need to restart the kernel to use updated packages.
```

```
In [ ]: #importando a biblioteca pandas apos sua instalação  
import pandas as pd
```

```
In [ ]: df_base.netflix = pd.read_csv("netflix_dataframe.csv")
```

```
In [ ]: # Carregando as primeiras Linhas do dataframe  
df_base.netflix.head()
```

Out[]:

	As of	Rank	Year to Date Rank	Last Week Rank	Title	Type	Netflix Exclusive	Netflix Release Date	Days In Top 10	Viewership Score
0	2020-04-01	1	1	1	Tiger King: Murder, Mayhem ...	TV Show	Yes	Mar 20, 2020	9	90
1	2020-04-01	2	2	-	Ozark	TV Show	Yes	Jul 21, 2017	5	45
2	2020-04-01	3	3	2	All American	TV Show	NaN	Mar 28, 2019	9	76
3	2020-04-01	4	4	-	Blood Father	Movie	NaN	Mar 26, 2020	5	30
4	2020-04-01	5	5	4	The Platform	Movie	Yes	Mar 20, 2020	9	55

In []: df_base_netflix.tail()

Out[]:

	As of	Rank	Year to Date Rank	Last Week Rank	Title	Type	Netflix Exclusive	Netflix Release Date	Days In Top 10	Viewersh Scc
7095	2022-03-11	6	5	1	Worst Roommate Ever	TV Show	Yes	Mar 1, 2022	10	
7096	2022-03-11	7	7	2	Vikings: Valhalla	TV Show	Yes	Feb 25, 2022	14	1
7097	2022-03-11	8	8	-	Shooter	Movie	NaN	Aug 1, 2014	3	
7098	2022-03-11	9	9	7	Shrek 2	Movie	NaN	Mar 1, 2022	10	
7099	2022-03-11	10	10	-	Shrek	Movie	NaN	May 1, 2018	7	

In []: display(df_base_netflix)

	As of	Rank	Year to Date Rank	Last Week Rank	Title	Type	Netflix Exclusive	Netflix Release Date	Days In Top 10	Viewership Score
0	2020-04-01	1	1	1	Tiger King: Murder, Mayhem ...	TV Show	Yes	Mar 20, 2020	9	90
1	2020-04-01	2	2	-	Ozark	TV Show	Yes	Jul 21, 2017	5	45
2	2020-04-01	3	3	2	All American	TV Show	NaN	Mar 28, 2019	9	76
3	2020-04-01	4	4	-	Blood Father	Movie	NaN	Mar 26, 2020	5	30
4	2020-04-01	5	5	4	The Platform	Movie	Yes	Mar 20, 2020	9	55
...
7095	2022-03-11	6	5	1	Worst Roommate Ever	TV Show	Yes	Mar 1, 2022	10	81
7096	2022-03-11	7	7	2	Vikings: Valhalla	TV Show	Yes	Feb 25, 2022	14	100
7097	2022-03-11	8	8	-	Shooter	Movie	NaN	Aug 1, 2014	3	7
7098	2022-03-11	9	9	7	Shrek 2	Movie	NaN	Mar 1, 2022	10	33
7099	2022-03-11	10	10	-	Shrek	Movie	NaN	May 1, 2018	7	12

7100 rows × 10 columns



In []: # Verificando tamanho do dataframe

df_base_netflix.shape

Out[]: (7100, 10)

In []: # Informações individuais de cada coluna do dataframe, com base em seu tipo, nome df_base_netflix.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7100 entries, 0 to 7099
Data columns (total 10 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   As of            7100 non-null    object  
 1   Rank             7100 non-null    int64  
 2   Year to Date Rank 7100 non-null    object  
 3   Last Week Rank  7100 non-null    object  
 4   Title            7100 non-null    object  
 5   Type             7100 non-null    object  
 6   Netflix Exclusive 4599 non-null    object  
 7   Netflix Release Date 7100 non-null    object  
 8   Days In Top 10   7100 non-null    int64  
 9   Viewership Score 7100 non-null    int64  
dtypes: int64(3), object(7)
memory usage: 554.8+ KB
```

```
In [ ]: df_base_netflix.dtypes
```

```
Out[ ]: As of          object
Rank           int64
Year to Date Rank  object
Last Week Rank  object
Title          object
Type           object
Netflix Exclusive  object
Netflix Release Date  object
Days In Top 10   int64
Viewership Score  int64
dtype: object
```

```
In [ ]: # Verificamos uma coluna específica de um dataframe e analisamos a quantidade de
df_base_netflix["Netflix Exclusive"].isnull().sum()
```

```
Out[ ]: np.int64(2501)
```

```
In [ ]: df_base_netflix["Netflix Exclusive"].value_counts()
```

```
Out[ ]: Netflix Exclusive
Yes      4599
Name: count, dtype: int64
```

```
In [ ]: df_base_netflix.describe()
```

Out[]:

	Rank	Days In Top 10	Viewership Score
count	7100.000000	7100.000000	7100.000000
mean	5.500000	24.123662	122.790141
std	2.872484	58.473789	213.861642
min	1.000000	1.000000	1.000000
25%	3.000000	3.000000	19.000000
50%	5.500000	7.000000	50.000000
75%	8.000000	18.000000	128.000000
max	10.000000	428.000000	1474.000000

In []:

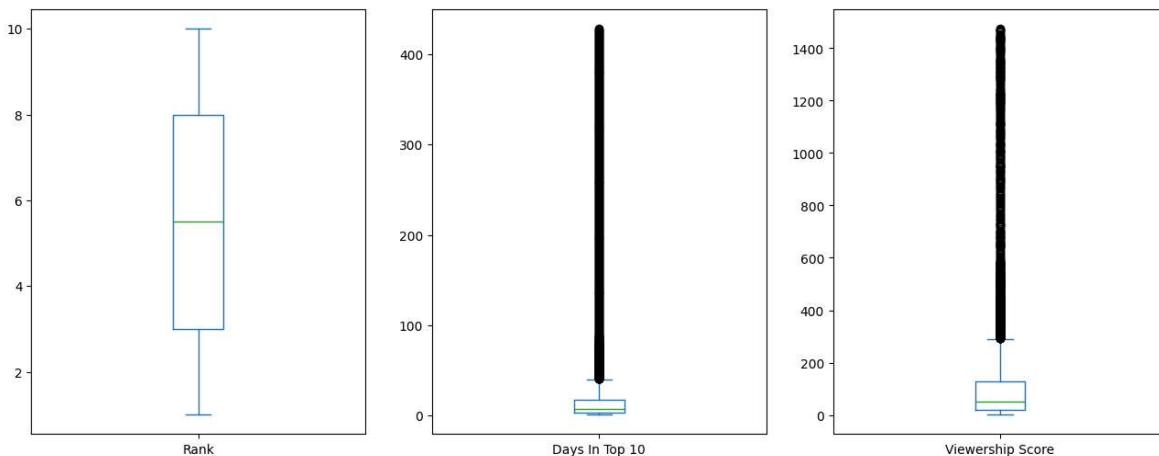
biblioteca para plotagem de graficos

%pip install matplotlib

```
Collecting matplotlib
  Downloading matplotlib-3.9.2-cp312-cp312-win_amd64.whl.metadata (11 kB)
Collecting contourpy>=1.0.1 (from matplotlib)
  Downloading contourpy-1.3.0-cp312-cp312-win_amd64.whl.metadata (5.4 kB)
Collecting cycler>=0.10 (from matplotlib)
  Downloading cycler-0.12.1-py3-none-any.whl.metadata (3.8 kB)
Collecting fonttools>=4.22.0 (from matplotlib)
  Downloading fonttools-4.53.1-cp312-cp312-win_amd64.whl.metadata (165 kB)
Collecting kiwisolver>=1.3.1 (from matplotlib)
  Downloading kiwisolver-1.4.7-cp312-cp312-win_amd64.whl.metadata (6.4 kB)
Requirement already satisfied: numpy>=1.23 in d:\users\aluno\desktop\python_senai\venv\lib\site-packages (from matplotlib) (2.1.1)
Requirement already satisfied: packaging>=20.0 in d:\users\aluno\desktop\python_senai\venv\lib\site-packages (from matplotlib) (24.1)
Collecting pillow>=8 (from matplotlib)
  Downloading pillow-10.4.0-cp312-cp312-win_amd64.whl.metadata (9.3 kB)
Collecting pyparsing>=2.3.1 (from matplotlib)
  Downloading pyparsing-3.1.4-py3-none-any.whl.metadata (5.1 kB)
Requirement already satisfied: python-dateutil>=2.7 in d:\users\aluno\desktop\python_senai\venv\lib\site-packages (from matplotlib) (2.9.0.post0)
Requirement already satisfied: six>=1.5 in d:\users\aluno\desktop\python_senai\venv\lib\site-packages (from python-dateutil>=2.7->matplotlib) (1.16.0)
Downloading matplotlib-3.9.2-cp312-cp312-win_amd64.whl (7.8 MB)
----- 0.0/7.8 MB ? eta -:-:-
----- 1.3/7.8 MB 6.1 MB/s eta 0:00:02
----- 1.6/7.8 MB 7.0 MB/s eta 0:00:01
----- 2.9/7.8 MB 4.5 MB/s eta 0:00:02
----- 3.7/7.8 MB 4.8 MB/s eta 0:00:01
----- 3.7/7.8 MB 4.8 MB/s eta 0:00:01
----- 4.7/7.8 MB 3.8 MB/s eta 0:00:01
----- 5.5/7.8 MB 3.9 MB/s eta 0:00:01
----- 5.8/7.8 MB 3.9 MB/s eta 0:00:01
----- 6.3/7.8 MB 3.4 MB/s eta 0:00:01
----- 6.8/7.8 MB 3.2 MB/s eta 0:00:01
----- 6.8/7.8 MB 3.2 MB/s eta 0:00:01
----- 7.1/7.8 MB 2.8 MB/s eta 0:00:01
----- 7.1/7.8 MB 2.8 MB/s eta 0:00:01
----- 7.3/7.8 MB 2.5 MB/s eta 0:00:01
----- 7.6/7.8 MB 2.4 MB/s eta 0:00:01
----- 7.6/7.8 MB 2.4 MB/s eta 0:00:01
----- 7.8/7.8 MB 2.3 MB/s eta 0:00:00
Downloading contourpy-1.3.0-cp312-cp312-win_amd64.whl (218 kB)
Downloading cycler-0.12.1-py3-none-any.whl (8.3 kB)
Downloading fonttools-4.53.1-cp312-cp312-win_amd64.whl (2.2 MB)
----- 0.0/2.2 MB ? eta -:-:-
----- 0.3/2.2 MB ? eta -:-:-
----- 0.8/2.2 MB 2.4 MB/s eta 0:00:01
----- 1.6/2.2 MB 2.7 MB/s eta 0:00:01
----- 2.1/2.2 MB 2.9 MB/s eta 0:00:01
----- 2.2/2.2 MB 2.6 MB/s eta 0:00:00
Downloading kiwisolver-1.4.7-cp312-cp312-win_amd64.whl (55 kB)
Downloading pillow-10.4.0-cp312-cp312-win_amd64.whl (2.6 MB)
----- 0.0/2.6 MB ? eta -:-:-
----- 0.8/2.6 MB 3.7 MB/s eta 0:00:01
----- 1.6/2.6 MB 4.0 MB/s eta 0:00:01
----- 2.4/2.6 MB 4.2 MB/s eta 0:00:01
----- 2.6/2.6 MB 3.5 MB/s eta 0:00:00
Downloading pyparsing-3.1.4-py3-none-any.whl (104 kB)
Installing collected packages: pyparsing, pillow, kiwisolver, fonttools, cycler, contourpy, matplotlib
```

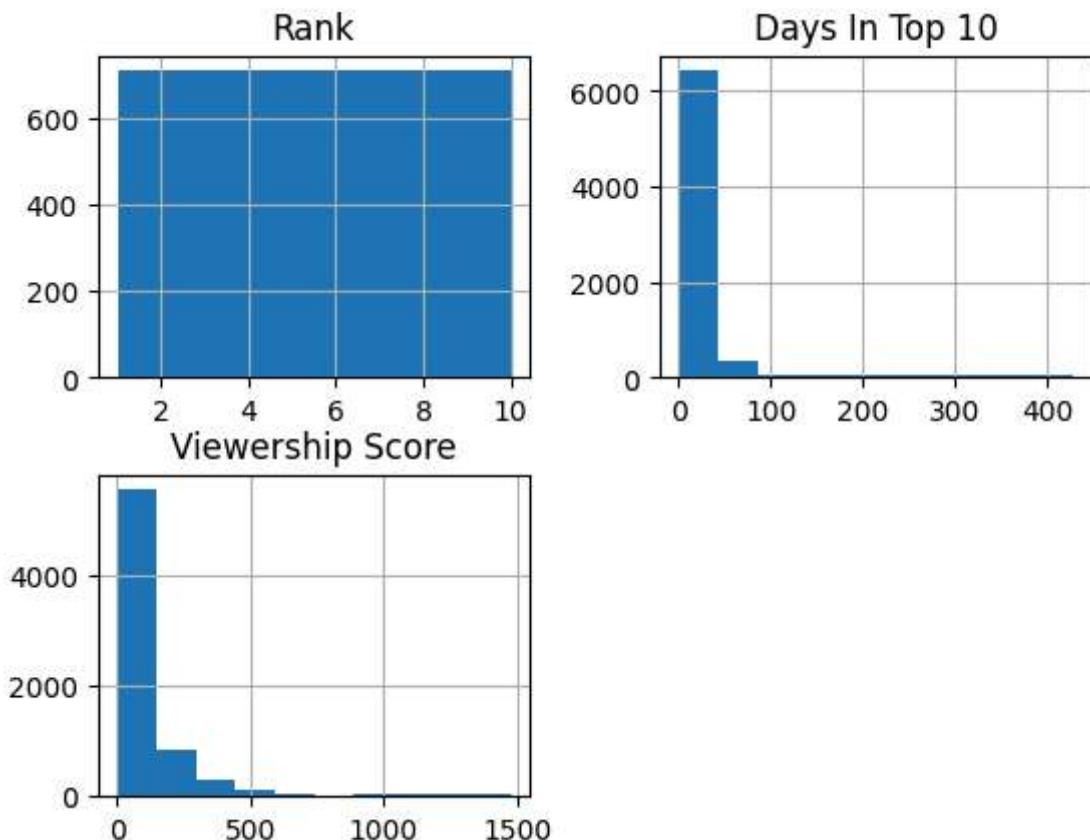
Successfully installed contourpy-1.3.0 cycler-0.12.1 fonttools-4.53.1 kiwisolver-1.4.7 matplotlib-3.9.2 pillow-10.4.0 pyparsing-3.1.4
Note: you may need to restart the kernel to use updated packages.

```
In [ ]: df_base.netflix.plot(kind='box', figsize=(16, 6), subplots=True );
```



```
In [ ]: df_base.netflix.hist()
```

```
Out[ ]: array([[[<Axes: title={'center': 'Rank'}>,
   <Axes: title={'center': 'Days In Top 10'}>],
  [<Axes: title={'center': 'Viewership Score'}>, <Axes: >]],
 dtype=object)
```



```
In [ ]: def aplica_no(exclusive):
    if exclusive != "Yes":
        return "No"
    return exclusive
```

```
In [ ]: df_base.netflix["Netflix Exclusive"] = df_base.netflix["Netflix Exclusive"].apply(aplica_no)
```

In []: `display(df_base_netflix)`

	As of	Rank	Year to Date Rank	Last Week Rank	Title	Type	Netflix Exclusive	Netflix Release Date	Days In Top 10	Viewership Score
0	2020-04-01	1	1	1	Tiger King: Murder, Mayhem ...	TV Show	Yes	Mar 20, 2020	9	90
1	2020-04-01	2	2	-	Ozark	TV Show	Yes	Jul 21, 2017	5	45
2	2020-04-01	3	3	2	All American	TV Show	No	Mar 28, 2019	9	76
3	2020-04-01	4	4	-	Blood Father	Movie	No	Mar 26, 2020	5	30
4	2020-04-01	5	5	4	The Platform	Movie	Yes	Mar 20, 2020	9	55
...
7095	2022-03-11	6	5	1	Worst Roommate Ever	TV Show	Yes	Mar 1, 2022	10	81
7096	2022-03-11	7	7	2	Vikings: Valhalla	TV Show	Yes	Feb 25, 2022	14	100
7097	2022-03-11	8	8	-	Shooter	Movie	No	Aug 1, 2014	3	7
7098	2022-03-11	9	9	7	Shrek 2	Movie	No	Mar 1, 2022	10	33
7099	2022-03-11	10	10	-	Shrek	Movie	No	May 1, 2018	7	12

7100 rows × 10 columns



In []: `df_base_netflix.drop("Last Week Rank", axis=1)`

Out[]:

	As of	Rank	Year to Date Rank	Title	Type	Netflix Exclusive	Netflix Release Date	Days In Top 10	Viewership Score
0	2020-04-01	1	1	Tiger King: Murder, Mayhem ...	TV Show	Yes	Mar 20, 2020	9	90
1	2020-04-01	2	2	Ozark	TV Show	Yes	Jul 21, 2017	5	45
2	2020-04-01	3	3	All American	TV Show	No	Mar 28, 2019	9	76
3	2020-04-01	4	4	Blood Father	Movie	No	Mar 26, 2020	5	30
4	2020-04-01	5	5	The Platform	Movie	Yes	Mar 20, 2020	9	55
...
7095	2022-03-11	6	5	Worst Roommate Ever	TV Show	Yes	Mar 1, 2022	10	81
7096	2022-03-11	7	7	Vikings: Valhalla	TV Show	Yes	Feb 25, 2022	14	100
7097	2022-03-11	8	8	Shooter	Movie	No	Aug 1, 2014	3	7
7098	2022-03-11	9	9	Shrek 2	Movie	No	Mar 1, 2022	10	33
7099	2022-03-11	10	10	Shrek	Movie	No	May 1, 2018	7	12

7100 rows × 9 columns