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 seu catalogo alcançaram o top 10 mais assistidos da plataforma para fins de estudos e aprendizado da biblioteca "pandas".

In [ ]:
```
# Instalando a biblioteca de analise de dados de 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_sen
ai\venv\lib\site-packages (from pandas) (2.1.1)
Requirement already satisfied: python-dateutil>=2.8.2 in d:\users\aluno\desktop\p
ython_senai\venv\lib\site-packages (from pandas) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in d:\users\aluno\desktop\python_sena
i\venv\lib\site-packages (from pandas) (2024.1)
Requirement already satisfied: tzdata>=2022.7 in d:\users\aluno\desktop\python_se
nai\venv\lib\site-packages (from pandas) (2024.1)
Requirement already satisfied: six>=1.5 in d:\users\aluno\desktop\python_senai\ve
nv\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

import pandas as pd
```

In [ ]:
```
# Carregando dataframe na memoria

df_netflix = pd.read_csv("netflix daily top 10.csv")
```

In [ ]:
```
# Visualizando as primeiras linhas do dataframe
df_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 [ ]:
```
# Analisando as ultimas linhas do dataframe
df_netflix.tail(15)
```

Out[ ]:

	As of	Rank	Year to Date Rank	Last Week Rank	Title	Type	Netflix Exclusive	Netflix Release Date	Days In Top 10	Viewersh Sco
7085	2022-03-10	6	5	5	Love is Blind	TV Show	Yes	Feb 13, 2020	45	3
7086	2022-03-10	7	6	2	Vikings: Valhalla	TV Show	Yes	Feb 25, 2022	13	
7087	2022-03-10	8	10	-	Shooter	Movie	NaN	Aug 1, 2014	2	
7088	2022-03-10	9	7	7	Shrek 2	Movie	NaN	Mar 1, 2022	9	
7089	2022-03-10	10	8	-	Shrek	Movie	NaN	May 1, 2018	6	
7090	2022-03-11	1	2	-	The Last Kingdom	TV Show	NaN	Jul 10, 2016	13	
7091	2022-03-11	2	1	-	Pieces of Her	TV Show	Yes	Mar 4, 2022	7	
7092	2022-03-11	3	3	-	Good Girls	TV Show	NaN	Jan 1, 2019	26	1
7093	2022-03-11	4	4	3	Inventing Anna	TV Show	Yes	Feb 11, 2022	28	2
7094	2022-03-11	5	6	6	Love is Blind	TV Show	Yes	Feb 13, 2020	46	3
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_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_netflix.shape
```

Out[ ]:
```
(7100, 10)
```

In [ ]:
```
# Informações individuais de cada coluna com base em seu tipo, nome, e quantidad

df_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_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 [ ]:
```
df_netflix["Netflix Exclusive"].isnull().sum()
```

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

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

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

In [ ]:
```
# biblioteca para plotagem de graficos

%pip install matplotlib
```

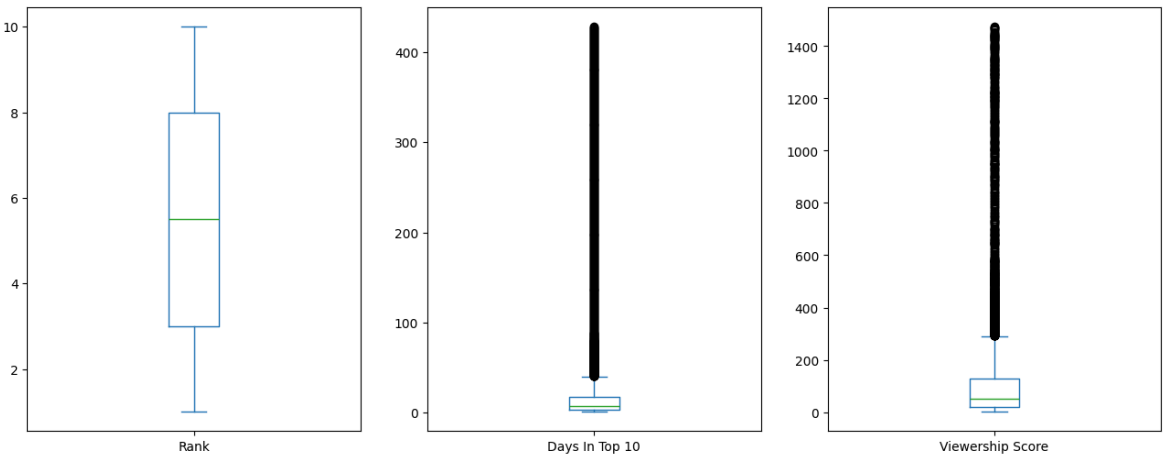
```
Requirement already satisfied: matplotlib in d:\users\aluno\desktop\python_senai
\venv\lib\site-packages (3.9.2)
Requirement already satisfied: contourpy>=1.0.1 in d:\users\aluno\desktop\python_
senai\venv\lib\site-packages (from matplotlib) (1.3.0)
Requirement already satisfied: cycler>=0.10 in d:\users\aluno\desktop\python_sena
i\venv\lib\site-packages (from matplotlib) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in d:\users\aluno\desktop\python
_senai\venv\lib\site-packages (from matplotlib) (4.53.1)
Requirement already satisfied: kiwisolver>=1.3.1 in d:\users\aluno\desktop\python
_senai\venv\lib\site-packages (from matplotlib) (1.4.7)
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_s
enai\venv\lib\site-packages (from matplotlib) (24.1)
Requirement already satisfied: pillow>=8 in d:\users\aluno\desktop\python_senai\v
env\lib\site-packages (from matplotlib) (10.4.0)
Requirement already satisfied: pyparsing>=2.3.1 in d:\users\aluno\desktop\python_
senai\venv\lib\site-packages (from matplotlib) (3.1.4)
Requirement already satisfied: python-dateutil>=2.7 in d:\users\aluno\desktop\pyt
hon_senai\venv\lib\site-packages (from matplotlib) (2.9.0.post0)
Requirement already satisfied: six>=1.5 in d:\users\aluno\desktop\python_senai\ve
nv\lib\site-packages (from python-dateutil>=2.7->matplotlib) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

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

Out[ ]:

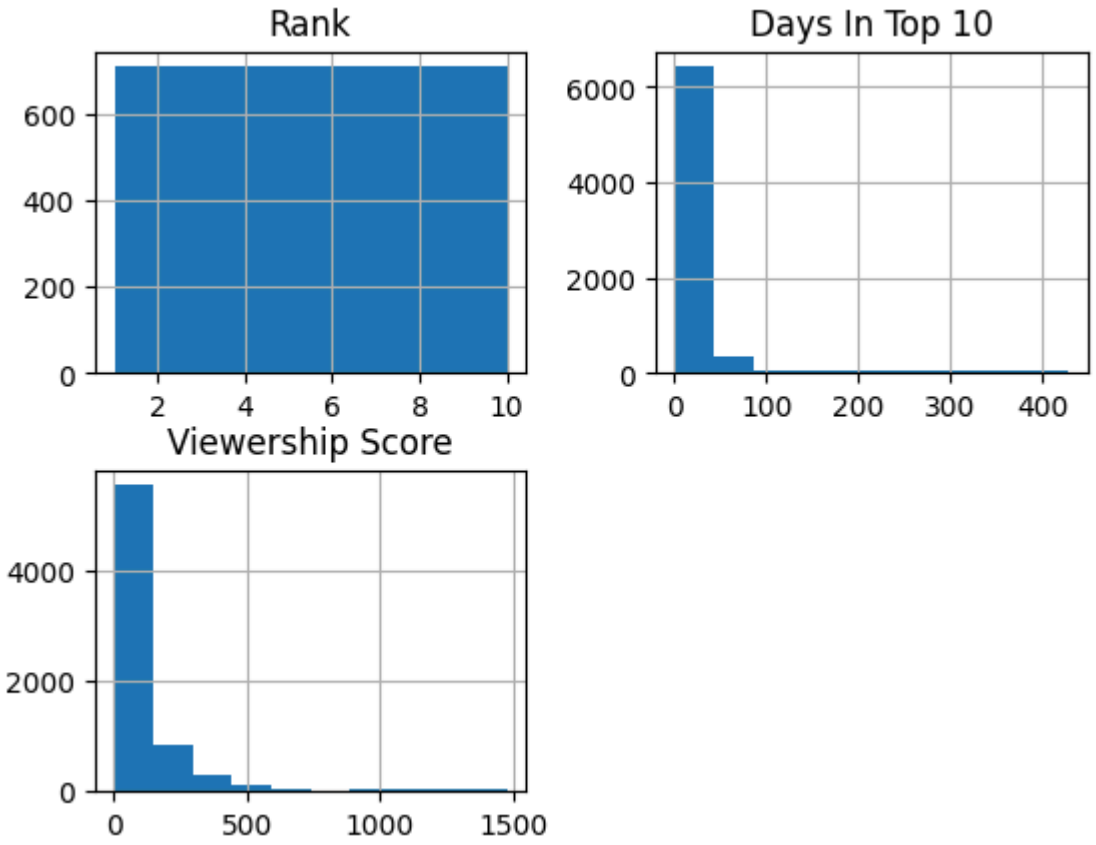
	Rank	Days In Top 10	Viewership Score
count	7100.000000	7100.000000	7100.000000
mean	5.500000	24.123662	122.790141
std	2.872484	58.473789	213.861642
min	1.000000	1.000000	1.000000
25%	3.000000	3.000000	19.000000
50%	5.500000	7.000000	50.000000
75%	8.000000	18.000000	128.000000
max	10.000000	428.000000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

In [ ]:
```
df_netflix.plot(kind="box", figsize=(16,6), subplots=True);
```



In [ ]:
```
df_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_netflix["Netflix Exclusive"] = df_netflix["Netflix Exclusive"].apply(aplica_n
```

In [ ]:
```
display(df_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_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

In [ ]: