amazon_prime

April 19, 2024

```python
[1]: import numpy as np
     import pandas as pd
     import matplotlib.pyplot as plt
     import seaborn as sns
     import datetime as dt
```

```python
[6]: data = pd.read_csv('amazon_prime_titles.csv')
```

```python
[7]: data.head()
```

```
[7]:   show_id   type                 title          director  \
     0      s1  Movie   The Grand Seduction     Don McKellar
     1      s2  Movie   Take Care Good Night    Girish Joshi
     2      s3  Movie   Secrets of Deception     Josh Webber
     3      s4  Movie     Pink: Staying True  Sonia Anderson
     4      s5  Movie          Monster Maker    Giles Foster

                                                    cast         country  \
     0     Brendan Gleeson, Taylor Kitsch, Gordon Pinsent          Canada
     1   Mahesh Manjrekar, Abhay Mahajan, Sachin Khedekar          India
     2  Tom Sizemore, Lorenzo Lamas, Robert LaSardo, R…   United States
     3  Interviews with: Pink, Adele, Beyoncé, Britney…   United States
     4  Harry Dean Stanton, Kieran O'Brien, George Cos…  United Kingdom

            date_added  release_year rating duration                 listed_in  \
     0  March 30, 2021          2014    NaN  113 min            Comedy, Drama
     1  March 30, 2021          2018    13+  110 min      Drama, International
     2  March 30, 2021          2017    NaN   74 min  Action, Drama, Suspense
     3  March 30, 2021          2014    NaN   69 min              Documentary
     4  March 30, 2021          1989    NaN   45 min           Drama, Fantasy

                                             description
     0  A small fishing village must procure a local d…
     1  A Metro Family decides to fight a Cyber Crimin…
     2  After a man discovers his wife is cheating on …
     3  Pink breaks the mold once again, bringing her …
     4  Teenage Matt Banting wants to work with a famo…
```

```
[9]: data.shape
```

```
[9]: (9668, 12)
```

```
[10]: data.columns
```

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

```
[11]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9668 entries, 0 to 9667
Data columns (total 12 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   show_id       9668 non-null   object
 1   type          9668 non-null   object
 2   title         9668 non-null   object
 3   director      7585 non-null   object
 4   cast          8435 non-null   object
 5   country       672 non-null    object
 6   date_added    155 non-null    object
 7   release_year  9668 non-null   int64
 8   rating        9331 non-null   object
 9   duration      9668 non-null   object
 10  listed_in     9668 non-null   object
 11  description   9668 non-null   object
dtypes: int64(1), object(11)
memory usage: 906.5+ KB
```

0.1 Clean the data

```
[12]: data.duplicated()
```

```
[12]: 0       False
      1       False
      2       False
      3       False
      4       False
              …
      9663    False
      9664    False
      9665    False
      9666    False
      9667    False
      Length: 9668, dtype: bool
```

```
[13]: data.isnull().sum()
```

```
[13]: show_id           0
      type              0
      title             0
      director       2083
      cast           1233
      country        8996
      date_added     9513
      release_year      0
      rating          337
      duration          0
      listed_in         0
      description        0
      dtype: int64
```

```
[14]: data['director']
```

```
[14]: 0          Don McKellar
      1          Girish Joshi
      2           Josh Webber
      3        Sonia Anderson
      4           Giles Foster
                    …
      9663    Joseph H. Lewis
      9664                NaN
      9665       Steve Barker
      9666                NaN
      9667      Daniel Barber
      Name: director, Length: 9668, dtype: object
```

```
[16]: data = data.dropna(subset=['director'])
      data = data.dropna(subset=['cast'])
      data = data.dropna(subset=['rating'])
```

```
[17]: data.isnull().sum()
```

```
[17]: show_id           0
      type              0
      title             0
      director          0
      cast              0
      country        6107
      date_added     6644
      release_year      0
      rating            0
      duration          0
```

```
listed_in           0
description         0
dtype: int64
```

[18]:
```python
columns_to_remove = ['cast']
data = data.drop(columns=columns_to_remove)
```

[19]:
```python
data.isnull().sum()
```

[19]:
```
show_id            0
type               0
title              0
director           0
country         6107
date_added      6644
release_year       0
rating             0
duration           0
listed_in          0
description        0
dtype: int64
```

0.2 Analyzing data

[22]:
```python
data.describe()
```

[22]:
```
       release_year
count   6645.000000
mean    2005.725658
std       21.141680
min     1920.000000
25%     2002.000000
50%     2015.000000
75%     2019.000000
max     2021.000000
```

[20]:
```python
data['director'].unique()
```

[20]:
```
array(['Girish Joshi', 'Glenn Miller', 'Drake Doremus', …,
       'John-Paul Davidson, Stephen Warbeck', 'Emily Skye',
       'Steve Barker'], dtype=object)
```

[21]:
```python
data['director'].value_counts()
```

[21]:
```
director
Mark Knight        38
Jay Chapman        23
Manny Rodriguez    22
```

```
John English          20
Cannis Holder         19
                      ..
Herve Bodilis          1
Arthur Joffe           1
Jorge Gaggero          1
Laurent Chouchan       1
Steve Barker           1
Name: count, Length: 5268, dtype: int64
```

0.3 En quelle année le plus grand nombre d'émissions de télévision et de films ont été diffusés ?

```
[23]: data['release_year'].value_counts()
```

```
[23]: release_year
      2021    985
      2020    595
      2019    545
      2018    356
      2016    337
              ...
      1926      2
      1922      2
      1923      1
      1924      1
      1927      1
      Name: count, Length: 100, dtype: int64
```

0.4 Top 3 des réalisateurs qui ont offert le plus de films et d'émissions de télévision à Amazon Prime

```
[24]: data['director'].value_counts().head(3)
```
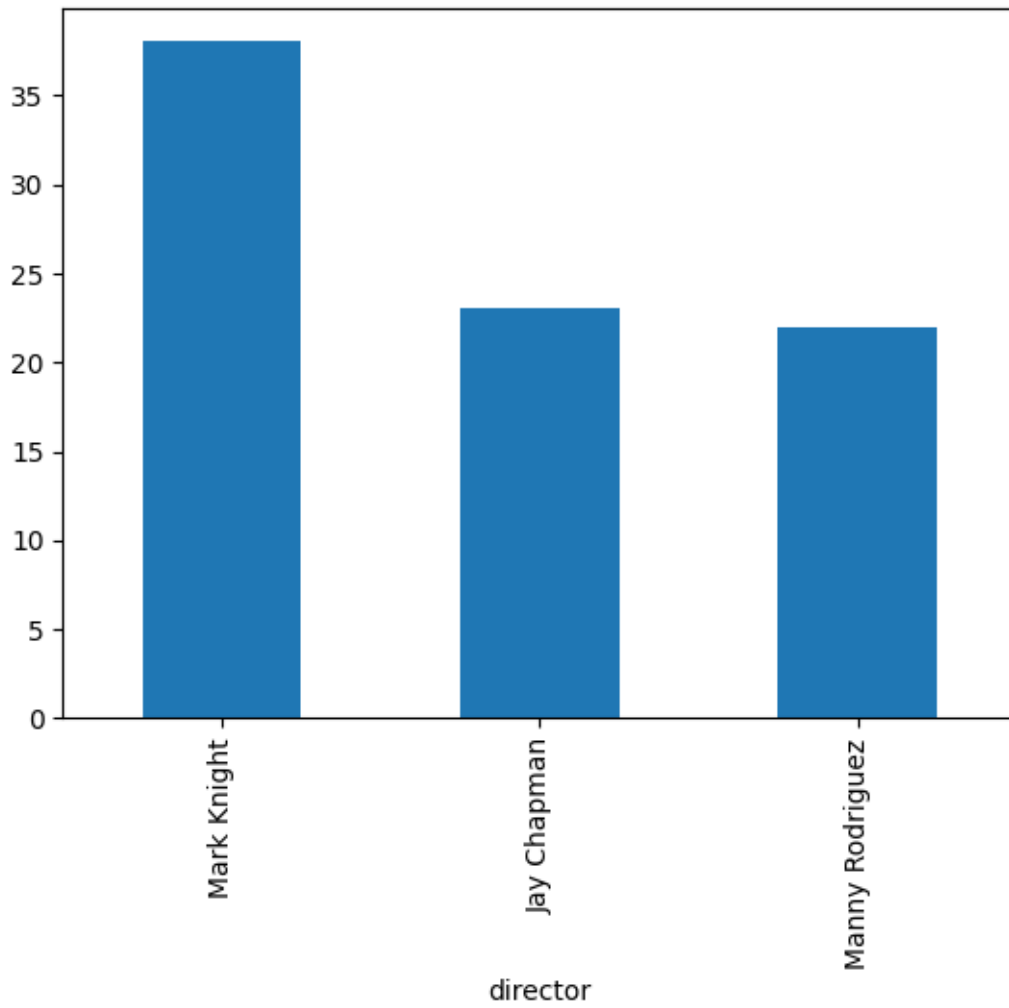
```
[24]: director
      Mark Knight        38
      Jay Chapman        23
      Manny Rodriguez    22
      Name: count, dtype: int64
```

```
[26]: data['director'].value_counts().head(3).plot(kind='bar')
```

```
[26]: <Axes: xlabel='director'>
```

director

[ ]: