PROJECT: INVESTIGATING NETFLIX MOVIES

Netflix! What started in 1997 as a DVD rental service has since exploded into one of the largest entertainment and media companies.

Given the large number of movies and series available on the platform, it is a perfect opportunity to flex your exploratory data analysis skills and dive into the entertainment industry.

You work for a production company that specializes in nostalgic styles. You want to do some research on movies released in the 1990's. You'll delve into Netflix data and perform exploratory data analysis to better understand this awesome movie decade!

You have been supplied with the dataset `netflix_data.csv`, along with the following table detailing the column names and descriptions. Feel free to experiment further after submitting!

The data

netflix_data.csv

Column	Description
`show_id`	The ID of the show
`type`	Type of show
`title`	Title of the show
`director`	Director of the show
`cast`	Cast of the show
`country`	Country of origin
`date_added`	Date added to Netflix
`release_year`	Year of Netflix release
`duration`	Duration of the show in minutes
`description`	Description of the show
`genre`	Show genre

```
# Importing pandas and matplotlib
import pandas as pd
import matplotlib.pyplot as plt

# Read in the Netflix CSV as a DataFrame
netflix_df = pd.read_csv("netflix_data.csv")
```

```
# Start coding here! Use as many cells as you like
```

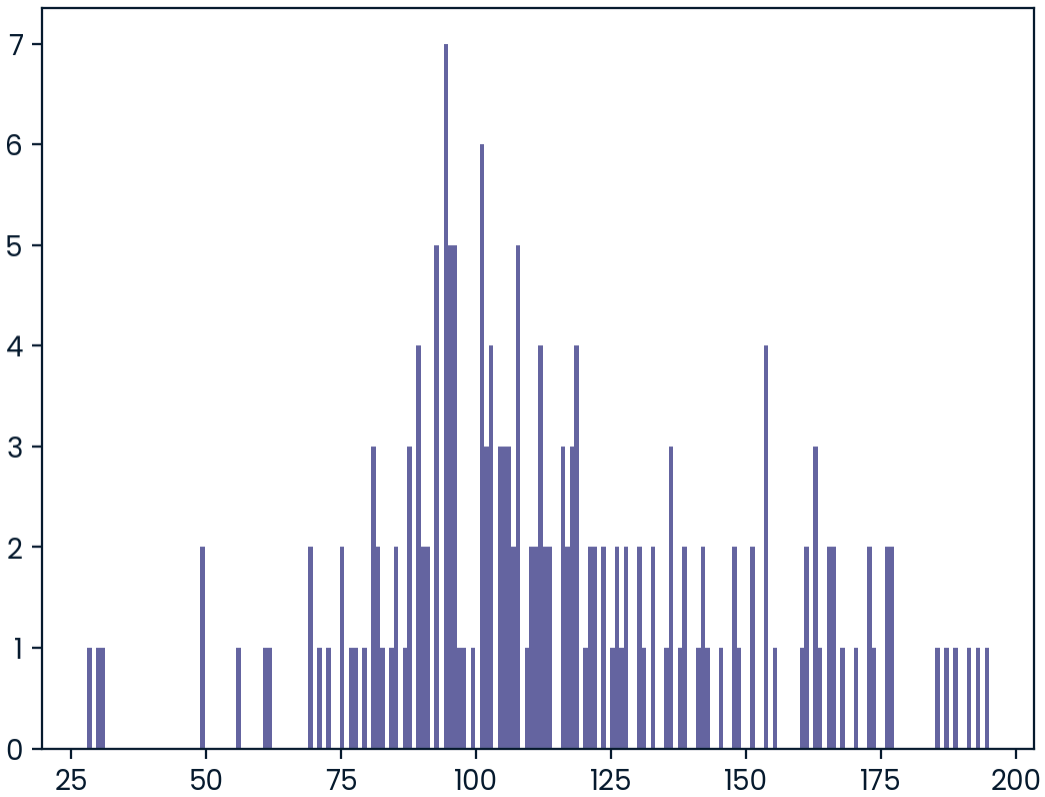
```
movies = netflix_df[netflix_df['type']=='Movie']
print(movies)
years90s = movies[(movies['release_year']>1989) & (movies['release_year']<2000)]
```

```
     show_id  ...          genre
0         s2  ...         Dramas
1         s3  ...  Horror Movies
2         s4  ...         Action
3         s5  ...         Dramas
5         s7  ...  Horror Movies
...      ...  ...            ...
4807   s7779  ...       Comedies
4808   s7781  ...         Dramas
4809   s7782  ...       Children
4810   s7783  ...         Dramas
4811   s7784  ...         Dramas

[4677 rows x 11 columns]
```

```
import matplotlib.pyplot as plt

plt.hist(years90s['duration'], bins=200)
```

```
(array([1., 0., 1., 1., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
        0., 0., 0., 0., 0., 0., 0., 0., 2., 0., 0., 0., 0., 0., 0., 1.,
        0., 0., 0., 0., 0., 1., 1., 0., 0., 0., 0., 0., 0., 0., 0., 2., 0.,
        1., 0., 1., 0., 0., 2., 0., 1., 1., 0., 1., 0., 3., 2., 1., 0., 1.,
        2., 0., 1., 3., 0., 4., 2., 2., 0., 5., 0., 7., 5., 5., 1., 1., 0.,
        1., 0., 6., 3., 4., 0., 3., 3., 3., 2., 5., 0., 1., 2., 2., 4., 2.,
        2., 0., 0., 3., 2., 3., 4., 0., 1., 2., 2., 0., 2., 0., 1., 2., 1.,
        2., 0., 0., 2., 1., 0., 2., 0., 0., 1., 3., 0., 1., 2., 0., 0., 1.,
        2., 1., 0., 0., 1., 0., 0., 2., 1., 0., 0., 2., 0., 0., 4., 0., 1.,
        0., 0., 0., 0., 0., 1., 2., 0., 3., 1., 0., 2., 2., 0., 1., 0., 0.,
        1., 0., 0., 2., 1., 0., 0., 2., 2., 0., 0., 0., 0., 0., 0., 0., 0.,
        0., 1., 0., 1., 0., 1., 0., 0., 1., 0., 1., 0., 1.]),
 array([ 28.   ,  28.835,  29.67 ,  30.505,  31.34 ,  32.175,  33.01 ,
         33.845,  34.68 ,  35.515,  36.35 ,  37.185,  38.02 ,  38.855,
         39.69 ,  40.525,  41.36 ,  42.195,  43.03 ,  43.865,  44.7  ,
         45.535,  46.37 ,  47.205,  48.04 ,  48.875,  49.71 ,  50.545,
         51.38 ,  52.215,  53.05 ,  53.885,  54.72 ,  55.555,  56.39 ,
         57.225,  58.06 ,  58.895,  59.73 ,  60.565,  61.4  ,  62.235,
         63.07 ,  63.905,  64.74 ,  65.575,  66.41 ,  67.245,  68.08 ,
         68.915,  69.75 ,  70.585,  71.42 ,  72.255,  73.09 ,  73.925,
         74.76 ,  75.595,  76.43 ,  77.265,  78.1  ,  78.935,  79.77 ,
         80.605,  81.44 ,  82.275,  83.11 ,  83.945,  84.78 ,  85.615,
         86.45 ,  87.285,  88.12 ,  88.955,  89.79 ,  90.625,  91.46 ,
         92.295,  93.13 ,  93.965,  94.8  ,  95.635,  96.47 ,  97.305,
         98.14 ,  98.975,  99.81 , 100.645, 101.48 , 102.315, 103.15 ,
        103.985, 104.82 , 105.655, 106.49 , 107.325, 108.16 , 108.995,
        109.83 , 110.665, 111.5  , 112.335, 113.17 , 114.005, 114.84 ,
        115.675, 116.51 , 117.345, 118.18 , 119.015, 119.85 , 120.685,
        121.52 , 122.355, 123.19 , 124.025, 124.86 , 125.695, 126.53 ,
        127.365, 128.2  , 129.035, 129.87 , 130.705, 131.54 , 132.375,
        133.21 , 134.045, 134.88 , 135.715, 136.55 , 137.385, 138.22 ,
        139.055, 139.89 , 140.725, 141.56 , 142.395, 143.23 , 144.065,
        144.9  , 145.735, 146.57 , 147.405, 148.24 , 149.075, 149.91 ,
        150.745, 151.58 , 152.415, 153.25 , 154.085, 154.92 , 155.755,
        156.59 , 157.425, 158.26 , 159.095, 159.93 , 160.765, 161.6  ,
        162.435, 163.27 , 164.105, 164.94 , 165.775, 166.61 , 167.445,
        168.28 , 169.115, 169.95 , 170.785, 171.62 , 172.455, 173.29 ,
        174.125, 174.96 , 175.795, 176.63 , 177.465, 178.3  , 179.135,
        179.97 , 180.805, 181.64 , 182.475, 183.31 , 184.145, 184.98 ,
        185.815, 186.65 , 187.485, 188.32 , 189.155, 189.99 , 190.825,
```

```
       191.66 , 192.495, 193.33 , 194.165, 195.   ]),
<BarContainer object of 200 artists>)
```



DataFrames and CSVs DataFrame as

```
select title, duration
from years90s
group by title, duration
order by count(title) desc
limit 1
```

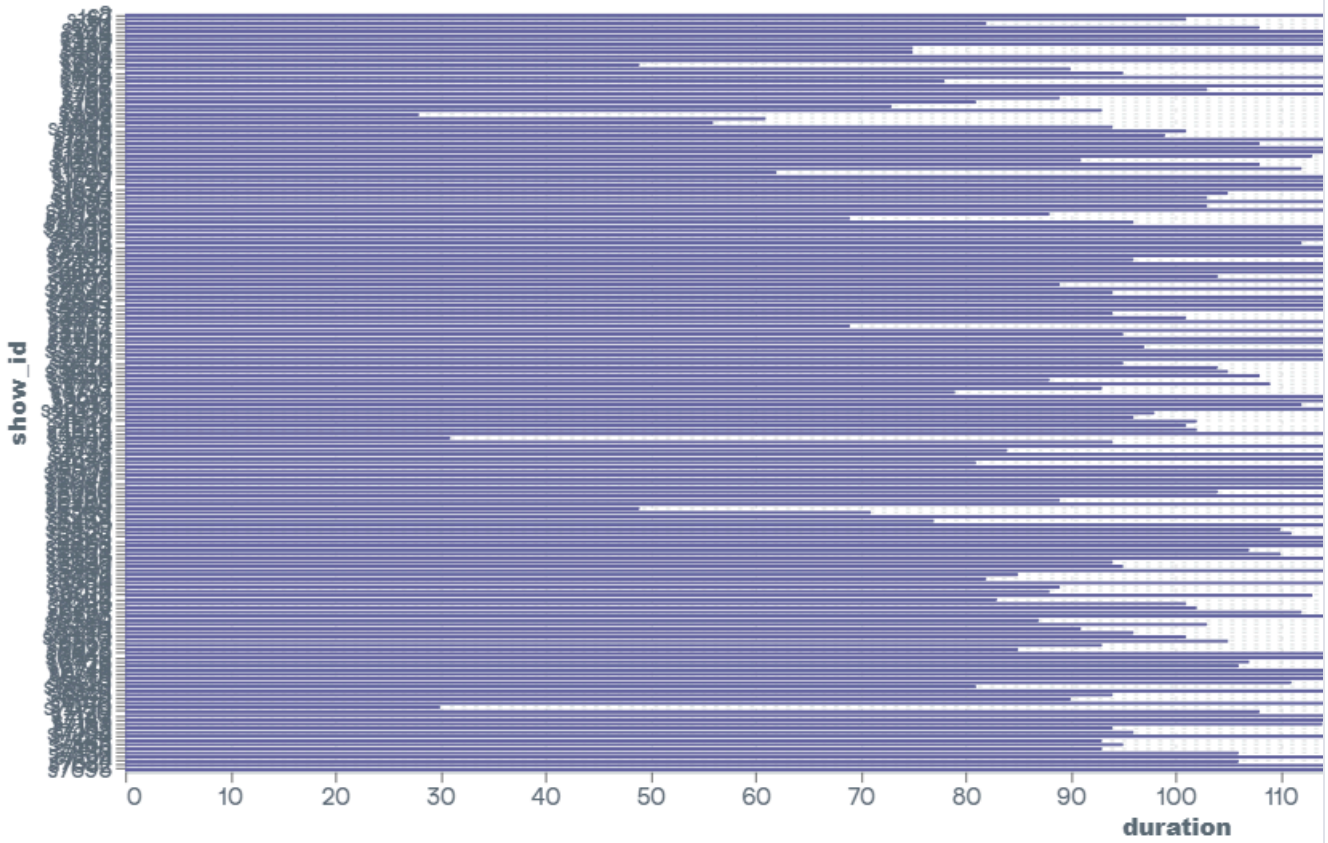
...	...	...
0	187	119

Rows: 1 Expand

```
years90s['duration']
```

...	...
6	119
118	101
145	82
167	108
194	154
240	174
315	122
320	119
333	75
334	75
352	160
410	139
428	49
430	90
431	95
468	119

Rows: 183 Expand




```
duration = 94

action = years90s[years90s['genre']=='Action']
for i in action["duration"]:
    i=int(i)

# if action["duration"] < 90:
    # print(action["duration"])
# counter = 0
# for i in action.itertuples():
#     if i.duration < 90:
#         counter += 1
# print(counter)
```

```
short_movie_count = 0
for movie in action:
    if isinstance(movie, dict) and "duration" in movie:
        if movie["duration"] < 90:
            short_movie_count += 1
print(short_movie_count)
```

```
0
```

```
short_movie_count = 0
for i in action["duration"]:
    i=int(i)
    if i < 90:
        short_movie_count += 1
        print(i)
    else:
        short_movie_count += 0
print(short_movie_count)
```

```
69
89
88
84
89
83
87
7
```