code

January 21, 2024

```python
[ ]:  # Importing pandas and matplotlib
      import pandas as pd
      import matplotlib.pyplot as plt
      from prettytable import PrettyTable
      import seaborn as sns

      # Reading csv
      netflix_data = pd.read_csv("netflix_data.csv")
      netflix_data
```

```
[ ]:       show_id      type                                 title  \
      0         s1   TV Show                                    3%
      1         s2     Movie                                  7:19
      2         s3     Movie                                 23:59
      3         s4     Movie                                     9
      4         s5     Movie                                    21
      …        …       …                                        …
      7782   s7783     Movie                                  Zozo
      7783   s7784     Movie                                Zubaan
      7784   s7785     Movie                     Zulu Man in Japan
      7785   s7786   TV Show               Zumbo's Just Desserts
      7786   s7787     Movie   ZZ TOP: THAT LITTLE OL' BAND FROM TEXAS

                    director                                         cast  \
      0                  NaN   João Miguel, Bianca Comparato, Michel Gomes, R…
      1     Jorge Michel Grau   Demián Bichir, Héctor Bonilla, Oscar Serrano, …
      2         Gilbert Chan   Tedd Chan, Stella Chung, Henley Hii, Lawrence …
      3          Shane Acker   Elijah Wood, John C. Reilly, Jennifer Connelly…
      4        Robert Luketic   Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar…
      …                  …                                              …
      7782        Josef Fares   Imad Creidi, Antoinette Turk, Elias Gergi, Car…
      7783        Mozez Singh   Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan…
      7784                NaN                                       Nasty C
      7785                NaN                     Adriano Zumbo, Rachel Khoo
      7786          Sam Dunn                                           NaN

                country          date_added   release_year   duration  \
      0          Brazil     August 14, 2020           2020          4
```

```
1              Mexico    December 23, 2016           2016          93
2           Singapore    December 20, 2018           2011          78
3       United States    November 16, 2017           2009          80
4       United States     January 1, 2020           2008         123
...               ...                 ...       ...        ...
7782           Sweden     October 19, 2020           2005          99
7783            India        March 2, 2019           2015         111
7784              NaN  September 25, 2020           2019          44
7785        Australia     October 31, 2020           2019           1
7786   United Kingdom        March 1, 2020           2019          90

                                            description                genre
0        In a future where the elite inhabit an island …  International TV
1        After a devastating earthquake hits Mexico Cit…           Dramas
2        When an army recruit is found dead, his fellow…    Horror Movies
3        In a postapocalyptic world, rag-doll robots hi…          Action
4        A brilliant group of students become card-coun…          Dramas
...                                                   …                …
7782  When Lebanon's Civil War deprives Zozo of his …           Dramas
7783  A scrappy but poor boy worms his way into a ty…           Dramas
7784  In this documentary, South African rapper Nast…    Documentaries
7785  Dessert wizard Adriano Zumbo looks for the nex…  International TV
7786  This documentary delves into the mystique behi…    Documentaries

[7787 rows x 11 columns]
```

```python
# Checking the structure of the table
print(netflix_data.info()) #structure info

print(netflix_data.describe()) #summary stats

# Checking nulls
missing_values = netflix_data.isnull().sum()
print("Missing values:", missing_values)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7787 entries, 0 to 7786
Data columns (total 11 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   show_id       7787 non-null   object
 1   type          7787 non-null   object
 2   title         7787 non-null   object
 3   director      5398 non-null   object
 4   cast          7069 non-null   object
 5   country       7280 non-null   object
 6   date_added    7777 non-null   object
 7   release_year  7787 non-null   int64
```

```
8   duration      7787 non-null   int64
9   description   7787 non-null   object
10  genre         7787 non-null   object
dtypes: int64(2), object(9)
memory usage: 669.3+ KB
None
       release_year      duration
count  7787.000000   7787.000000
mean   2013.932580     69.122769
std       8.757395     50.950743
min    1925.000000      1.000000
25%    2013.000000      2.000000
50%    2017.000000     88.000000
75%    2018.000000    106.000000
max    2021.000000    312.000000
Missing values: show_id          0
type              0
title             0
director       2389
cast            718
country         507
date_added       10
release_year      0
duration          0
description       0
genre             0
dtype: int64
```

```python
# Groupings and aggregations
grouped_data = netflix_data.groupby("type")["duration"].mean()
print("Grouped data:", grouped_data)
```

```
Grouped data: type
Movie     99.307978
TV Show    1.775934
Name: duration, dtype: float64
```

```python
# Exploring the column values
table = PrettyTable(["Column", "Unique Values"])

for column in netflix_data.columns:
    unique_values = netflix_data[column].unique()
    table.add_row([column, unique_values])

print(table)
```

```
+--------------+--------------------------------------------------------------
--------------------------------------------------------------------------------
----------------------------+
```

```
|    Column    |                                                    Unique Values                                                    |
+-------------+-------------------------------------------------------------------------------------------------------------------+
|   show_id   | ['s1' 's2' 's3' … 's7785' 's7786' 's7787']                                                                          |
|    type     | ['TV Show' 'Movie']                                                                                                 |
|    title    |                                                                                                          ['3%'      |
|             | '7:19' '23:59' … 'Zulu Man in Japan' "Zumbo's Just Desserts"                                                         |
|             | "ZZ TOP: THAT LITTLE OL' BAND FROM TEXAS"]                                                                           |
|   director  |                                                                                                  [nan 'Jorge        |
|             | Michel Grau' 'Gilbert Chan' … 'Josef Fares' 'Mozez Singh'                                                           |
|             | 'Sam Dunn']                                                                                                          |
|    cast     | ['João Miguel, Bianca Comparato, Michel Gomes, Rodolfo Valente, Vaneza Oliveira, Rafael Lozano, Viviane Porto, Mel Fronckowiak, Sergio Mamberti, Zezé Motta, Celso Frateschi' |
|             |                                                                  'Demián Bichir, Héctor Bonilla, Oscar Serrano, Azalia Ortiz, Octavio Michel, Carmen Beato'                  |
|             |                                                        'Tedd Chan, Stella Chung, Henley Hii, Lawrence Koh, Tommy Kuan, Josh Lai, Mark Lee, Susan Leong, Benjamin Lim'        |
|             | …                                                                                                                   |
|             |                                                        'Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanana, Manish Chaudhary, Meghna Malik, Malkeet Rauni, Anita Shabdish, Chittaranjan Tripathy' |
|             | 'Nasty C' 'Adriano Zumbo, Rachel Khoo']                                                                             |
|   country   |                                                                                              ['Brazil'              |
|             | 'Mexico' 'Singapore' 'United States' 'Turkey' 'Egypt' 'India'                                                       |
|             |                                                                                                  'Poland'           |
|             | 'Thailand' 'Nigeria' nan 'Norway' 'United Kingdom' 'Japan'                                                          |
```

4

| | 'South
Korea' 'Italy' 'Canada' 'Indonesia' 'Romania' 'Spain' 'Iceland'
|
| | 'South
Africa' 'France' 'Portugal' 'Hong Kong' 'Argentina' 'Germany'
|
| | 'Denmark'
'Kenya' 'New Zealand' 'Pakistan' 'Australia' 'China' 'Taiwan'
|
| |
'Netherlands' 'Philippines' 'United Arab' 'Iran' 'Israel' 'Uruguay'
|
| |
'Bulgaria' 'Mauritius' 'Lebanon' 'Colombia' 'Soviet Union' 'Sweden'
|
| |
'Malaysia' 'Ireland' 'Serbia' 'Finland' 'Peru' 'Chile' 'Ghana'
|
| | 'Saudi
Arabia' 'Namibia' 'Russia' 'Belgium' 'Vietnam' 'Kuwait'
|
| | 'Czech
Republic' 'Zimbabwe' 'Hungary' 'Venezuela' 'Cambodia' 'Austria'
|
| | 'West
Germany' 'Slovenia' 'Switzerland' 'Jordan' 'Bangladesh' 'Georgia'
|
| | 'Puerto
Rico' 'Senegal' 'Guatemala' 'Ukraine' 'Jamaica' 'Belarus'
|
| |
'Cyprus' 'Greece' 'Paraguay' 'Croatia' 'Somalia']
|
| date_added |
['August 14, 2020' 'December 23, 2016' 'December 20, 2018' …
|
| |
'December 6, 2018' 'March 9, 2016' 'January 11, 2020']
|
| release_year | [2020 2016
2011 2009 2008 2019 1997 2017 2018 2014 2015 2010 2005 2013
|
| | 2012 2006
2000 2003 1989 2004 1977 1971 1993 2007 1998 1988 1996 2002
|
| | 1986 1990
1979 1980 1974 2001 1999 1966 1991 1982 1994 1995 1978 1992
|

| | | 1972 1975
2021 1973 1967 1960 1981 1958 1976 1968 1956 1985 1965 1984
|
| | 1983 1987
1945 1946 1962 1925 1942 1969 1955 1963 1964 1954 1970 1944
|
| |
1947 1959 1943]
|
| duration | [4 93 78
80 123   1 95 119 118 143 103 89 91 149 144 124 87 110
|
| | 128 117 100
 2 84 99 90 102 104 105 56 125 81 97 106 107 109 44
|
| | 75 101   3
37 113 114 130 94 140 135 82 70 121 92 164 53 83 116
|
| | 86 120 96
126 129 77 137 148 28 122 176 85 22 68 111 29 142 168
|
| | 21 59 20
98 108 76 26 156 30 57 150 133 115 154 127 146 136 88
|
| | 131 24 112
74 63 38 25 174 60 153 158 151 162 54 51 69 64 147
|
| | 42 79   5
40 45 172 10 163   9 55 72 61 71 160 171 48 139 157
|
| | 15 65 134
161 62   8 186 49 73 58 165 166 138 159 141 132 52 67
|
| | 34 66 312
180 47   6 155 14 177 11 46 145   7 13 12 43 50 23
|
| | 185 200 169
27 170 196 33 181 204 32 35 167 16 179 193 214 17 173
|
| | 192 209 187
41 182 224 233 189 152 19 208 237 31 178 230 194 228 195
|
| |
18 205 190 36 201 253 203 191]
|
| description | ['In a future where the elite inhabit an
island paradise far from the crowded slums, you get one chance to join the 3%
saved from squalor.' |

```
|              |                    'After a devastating earthquake hits Mexico City,
trapped survivors from all walks of life wait to be rescued while trying
desperately to stay alive.'           |
|              |                    "When an army recruit is found dead, his fellow
soldiers are forced to confront a terrifying secret that's haunting their jungle
island training camp."           |
|              |
…
|
|              |                    "In this documentary, South African rapper Nasty C
hits the stage and streets of Tokyo, introducing himself to the city's sights,
sounds and culture."           |
|              |                    'Dessert wizard Adriano Zumbo looks for the next
"Willy Wonka" in this tense competition that finds skilled amateurs competing
for a $100,000 prize.'           |
|              |                    'This documentary delves into the mystique
behind the blues-rock trio and explores how the enigmatic band created their
iconic look and sound.']           |
|    genre     |
['International TV' 'Dramas' 'Horror Movies' 'Action' 'Crime TV'
|
|              |
'Documentaries' 'Independent Movies' 'Comedies' 'Sports Movies'
|
|              |                                        'Anime
Series' 'Reality TV' 'TV Comedies' 'Docuseries' 'Uncategorized'
|
|              |                                        'British TV'
'International Movies' 'Sci-Fi' 'Kids' 'Children' 'TV Shows'
|
|              |                                        'Classic
Movies' 'Thrillers' 'Stand-Up' 'Anime Features' 'Music'
|
|              |                                        'TV Dramas'
'TV Horror' 'TV Action' 'Classic' 'Romantic TV' 'Cult Movies'
|
|              |                                          'TV
Sci' 'Romantic Movies' 'Spanish-Language' 'LGBTQ Movies']
|
+--------------+----------------------------------------------------------------
--------------------------------------------------------------------------------
----------------------------+
```

```python
# Filtering all TV Shows
netflix_subset = netflix_data[netflix_data["type"] != "TV Show"]

# Subsetting the dataset
```

```
netflix_movies = netflix_subset[["title", "country", "genre", "release_year",␣
  ↪"duration"]]
```

```
# Finding movies shorter than 60 minutes
short_movies = netflix_movies[netflix_movies["duration"] < 60]
display(short_movies)
```

```
                                                    title        country  \
35                                               #Rucker50  United States
55                        100 Things to do Before High School  United States
67      13TH: A Conversation with Oprah Winfrey & Ava …            NaN
101                                        3 Seconds Divorce         Canada
146                                             A 3 Minute Hug         Mexico
...                                                     ...            ...
7679                      WWII: Report from the Aleutians  United States
7692    Ya no estoy aquí: Una conversación entre Guill…            NaN
7718                        Yoo Byung Jae: Discomfort Zone    South Korea
7771                                                  Zion  United States
7784                                     Zulu Man in Japan            NaN

             genre  release_year  duration
35     Documentaries          2016        56
55     Uncategorized         2014        44
67     Uncategorized         2017        37
101    Documentaries          2018        53
146    Documentaries          2019        28
...              ...           ...       ...
7679   Documentaries          1943        45
7692   Documentaries          2020        15
7718         Stand-Up          2018        54
7771   Documentaries          2018        12
7784   Documentaries          2019        44

[420 rows x 5 columns]
```

```
# Plotting histograms
fig, axes = plt.subplots(3, 1, figsize=(12, 15))

# Histogram for 'genre'
axes[0].hist(short_movies['genre'], bins=30, color='skyblue', edgecolor='black')
axes[0].set_title('Histogram of Genre')
axes[0].set_xlabel('Genre')
axes[0].set_ylabel('Count')
axes[0].tick_params(axis='x', rotation=45)

# Histogram for 'country'
```
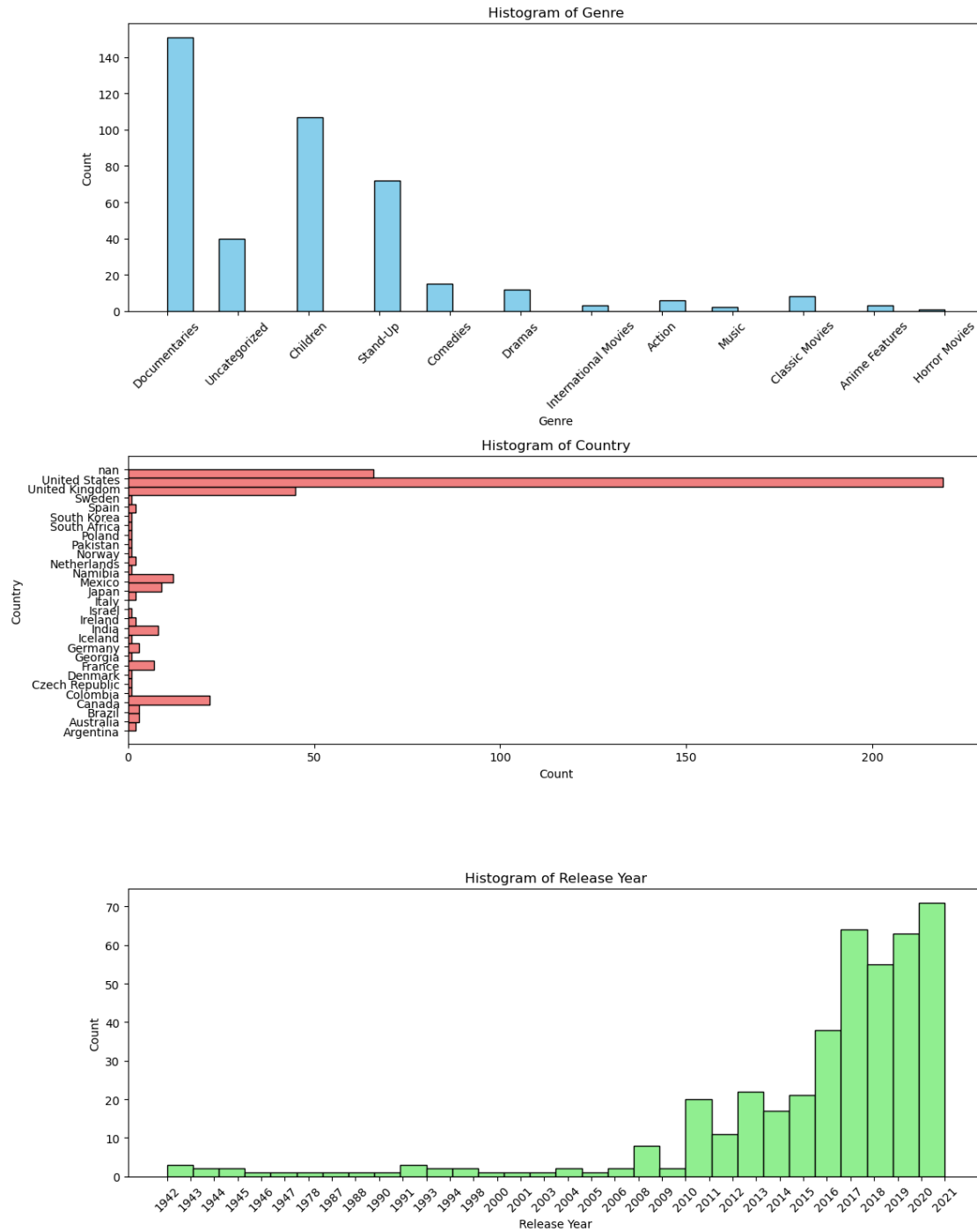
```
axes[1].hist(short_movies['country'].astype(str).sort_values(), bins=30,␣
 ↪color='lightcoral', edgecolor='black', orientation='horizontal')
axes[1].set_title('Histogram of Country')
axes[1].set_xlabel('Count')
axes[1].set_ylabel('Country')

# Histogram for 'release_year'
axes[2].hist(short_movies['release_year'].astype(str).sort_values(), bins=30,␣
 ↪color='lightgreen', edgecolor='black')
axes[2].set_title('Histogram of Release Year')
axes[2].set_xlabel('Release Year')
axes[2].set_ylabel('Count')
axes[2].tick_params(axis='x', rotation=45)

# Adjust layout
plt.tight_layout()

# Show the plots
plt.show()
```

Histogram of Genre

Histogram of Country

Histogram of Release Year

```python
# Iterate over the genres to assign colors
colors = []

for index, rows in netflix_movies.iterrows():
    genre = rows["genre"]
```

```python
    if genre == "Children":
        color = "Red"
    elif genre == "Documentaries":
        color = "Black"
    elif genre == "Stand-up":
        color = "Blue"
    else:
        color = "Purple"

    colors.append(color)

#Plot by colors
release_year = netflix_movies['release_year']
duration = netflix_movies['duration']

# Plot by colors with enhanced settings
plt.figure(figsize=(10, 6))
scatter = plt.scatter(release_year, duration, c=colors, alpha=0.6,␣
 ↪edgecolors='k', linewidths=0.5)

# Add labels and title
plt.xlabel('Release Year')
plt.ylabel('Duration (min)')
plt.title('Movie Duration by Year of Release')

# Add legend
legend_labels = {'Red': 'Children', 'Black': 'Documentaries', 'Blue':␣
 ↪'Stand-up', 'White': 'Other'}
legend_elements = [plt.Line2D([0], [0], marker='o', color='w',␣
 ↪markerfacecolor=color, markersize=10, label=label) for color, label in␣
 ↪legend_labels.items()]
plt.legend(handles=legend_elements, title='Genre', loc='upper right')

# Customize the colorbar
cbar = plt.colorbar(scatter)
cbar.set_ticks([])
cbar.set_ticklabels([])

# Add a general trendline for the entire plot
sns.regplot(x=release_year, y=duration, scatter=False, color='gray',␣
 ↪line_kws={'label': 'Trendline'})

# Show the plot
plt.show()
```

Movie Duration by Year of Release