

## Untitled-1

```

1 # %% [markdown]
2 # <div style="display: flex; justify-content: center; align-items: center; min-height: 100px;
  min-width: 550px; background-color: #334445; border-radius: 30px; box-shadow: 0px 0px 10px
  rgba(0, 0, 0, 0.5); border: 5px solid transparent; border-image: linear-gradient(45deg,
  #FF9950, #42E695); border-image-slice: 20; text-align: center;">
3 #
4 #
5 # <h3 style="color: #FF9950; font-family: 'Arial', sans-serif; font-size: 2em; margin-top:
  20px;">Tittle || Visualizing the Reel World 📺: <span style="color: #FFD700;">A Comprehensive
  EDA of </span>Netflix's Content</h3>
6 # </div>
7
8 # %% [markdown]
9 # ![image](netflix.png)
10
11 # %% [markdown]
12 # <div style="display: flex; justify-content: center; align-items: center; min-height: 50px;
  min-width: 700px; background-color: #334445; border-radius: 20px; box-shadow: 0px 0px 10px
  rgba(0, 0, 0, 0.5);">
13 #   <h1 style="color: #FF1140; font-family: 'Arial Black', Arial, sans-serif; font-size: 2em;
    text-transform: uppercase; letter-spacing: 4px; text-shadow: 2px 2px 4px rgba(0, 0, 0, 0.7);">
    📄 Table of Contents 📄 </h1><br>
14 # </div>
15 #
16
17 # %% [markdown]
18 # <div style="border: 4px solid #FF5733; padding: 10px; max-width: 500px; margin: 0 auto;
  border-radius: 20px;">
19 #   <p style="color: #4CAF50;">
20 #       I'm Umar Mehmood, a passionate Data Scientist and Analyst, specializing in data
    exploration, analysis, and visualization. My expertise lies in unraveling insights from complex
    datasets and presenting them in compelling visualizations. Let's dive deep into the world of
    data together! 📊🔍💡
21 #   </p>
22 # </div>
23 #
24
25 # %% [markdown]
26 # <div style="color:white;display:fill;border-radius:8px;font-size:200%; letter-
  spacing:1.0px;"><p style="padding: 5px;color:white;text-align:left;"><b><span
  style='color:#FF1140'>Social Media Links</span></b></p></div>
27 #
28 # <div style="text-align: left;">
29 #   <table>
30 #       <tr>
31 #           <td><a href="https://www.linkedin.com/in/umar-mehmood-147224294/"></a></td>
32 #           <td><a href="mailto:umarmehmood885@gmail.com"></a></td>
33 #         </tr>
34 #         <tr>
35 #             <td><a href="https://github.com/Umar885"></a></td>
36 #             <td><a href="https://twitter.com/UmarMeh12214006"></a></td>
37 #         </tr>
38 #         <tr>
39 #             <td><a href="https://www.kaggle.com/umarmehmood"></a></td>
40 #             <td><a href="https://medium.com/@umarmehmood885"></a></td>
41 #         </tr>
42 #     </table>
43 # </div>
44 #
45
46 # %% [markdown]
47 # <a id="1"></a>
48 # <div style="display: flex; justify-content: center; align-items: center; min-height: 50px;
min-width: 500px; background-color: #334445; border-radius: 20px; box-shadow: 0px 0px 10px
rgba(0, 0, 0, 0.5);">
49 #     <h1 style="color: #FF1140; font-family: 'Arial Black', Arial, sans-serif; font-size: 2em;
text-transform: uppercase; letter-spacing: 4px; text-shadow: 2px 2px 4px rgba(0, 0, 0, 0.7);">
🌟 Introduction 🌟 </h1><br>
50 # </div>
51
52 # %% [markdown]
53 #
54 # ***📽️📺 Lights, camera, data! Welcome to the captivating realm of Netflix dataset
exploration, where every datapoint holds the key to unlocking the secrets of entertainment's
digital frontier. 🌟📈
55 # In this exhilarating expedition, we'll delve deep into the vast expanse of Netflix's content
universe. From critically acclaimed series to blockbuster movies, we'll dissect metadata,
scrutinize viewer preferences, and uncover the trends shaping the future of streaming.***
56
57 # %% [markdown]
58 # <a id="2"></a>
59 # <div style="display: flex; justify-content: center; align-items: center; min-height: 50px;
min-width: 500px; background-color: #334445; border-radius: 20px; box-shadow: 0px 0px 10px
rgba(0, 0, 0, 0.5);">
60 #     <h1 style="color: #FF1140; font-family: 'Arial Black', Arial, sans-serif; font-size: 2em;
text-transform: uppercase; letter-spacing: 4px; text-shadow: 2px 2px 4px rgba(0, 0, 0, 0.7);">
📽️ About Netflix 📽️
61 #     </h1><br>
62 # </div>
63

```

```

64 # %% [markdown]
65 # ***📺 Netflix is one of the most popular media and video streaming platforms. They have over
8000 movies or TV shows available on their platform. As of mid-2021, they have over 200M
subscribers globally. 🌐 This tabular dataset consists of listings of all the movies and TV
shows available on Netflix, along with details such as cast, directors, ratings, release year,
duration, etc. 🎬📊***
66 # **The following links allowed users to obtain the Data Set from Kaggle** [link]
(https://www.kaggle.com/datasets/shivamb/netflix-shows)
67
68 # %% [markdown]
69 # <div style="text-align: center;">
70 #     <h1 style="color: #FF1140;">Kernel Version Used:</h1>
71 #     <ul style="list-style-type: none; padding: 0;">
72 #         <li style="color: #333; font-size: 18px; background-color: #FFEB3B; padding: 10px
20px; border-radius: 5px; display: inline-block; margin: 5px;">Python 3.11.8</li>
73 #     </ul>
74 # </div>
75 #
76
77 # %% [markdown]
78 # <div style="display: flex; justify-content: center; align-items: center; min-height: 50px;
min-width: 500px; background-color: #334445; border-radius: 20px; box-shadow: 0px 0px 10px
rgba(0, 0, 0, 0.5);">
79 #     <h1 style="color: #FF1140; font-family: 'Arial Black', Arial, sans-serif; font-size: 2em;
text-transform: uppercase; letter-spacing: 4px; text-shadow: 2px 2px 4px rgba(0, 0, 0, 0.7);">
📖 Import Librarie 📖 </h1><br>
80 # </div>
81
82 # %%
83 import pandas as pd
84 import numpy as np
85 import matplotlib.pyplot as plt
86 import seaborn as sns
87 import plotly.express as px
88 %matplotlib inline
89
90 # %% [markdown]
91 # <div style="display: flex; justify-content: center; align-items: center; min-height: 50px;
min-width: 500px; background-color: #334445; border-radius: 20px; box-shadow: 0px 0px 10px
rgba(0, 0, 0, 0.5);">
92 #     <h1 style="color: #FF1140; font-family: 'Arial Black', Arial, sans-serif; font-size: 2em;
text-transform: uppercase; letter-spacing: 4px; text-shadow: 2px 2px 4px rgba(0, 0, 0, 0.7);">
📁 Import Dataset 📁 </h1><br>
93 # </div>
94
95
96 # %%
97 df = pd.read_csv("netflix_titles.csv")
98
99 # %% [markdown]
100 # <p style="color: #FF6347; font-family: 'Helvetica Neue', Helvetica, Arial, sans-serif; font-
size: 24px; font-weight: bold; text-transform: uppercase; letter-spacing: 1px; text-shadow: 2px
2px 4px rgba(0, 0, 0, 0.3);">Show the head of the dataset</p>

```

```

101
102 # %%
103 df.head()
104
105 # %% [markdown]
106 # <div style="display: flex; justify-content: center; align-items: center; min-height: 50px;
min-width: 500px; background-color: #334445; border-radius: 20px; box-shadow: 0px 0px 10px
rgba(0, 0, 0, 0.5);">
107 #   <h1 style="color: #FF1140; font-family: 'Arial Black', Arial, sans-serif; font-size: 2em;
text-transform: uppercase; letter-spacing: 4px; text-shadow: 2px 2px 4px rgba(0, 0, 0, 0.7);">
Basic Exploration 🔍 </h1><br>
108 # </div>
109
110 # %% [markdown]
111 # <p style="color: #FF6347; font-family: 'Helvetica Neue', Helvetica, Arial, sans-serif; font-
size: 24px; font-weight: bold; text-transform: uppercase; letter-spacing: 1px; text-shadow: 2px
2px 4px rgba(0, 0, 0, 0.3);">Let us take a brief look at the dataset's shape</p>
112 #
113
114 # %%
115 print("Number of rows in the dataset:", df.shape[0])
116 print("Number of columns in the dataset:", df.shape[1])
117
118 # %% [markdown]
119 # <p style="color: #FF6347; font-family: 'Helvetica Neue', Helvetica, Arial, sans-serif; font-
size: 24px; font-weight: bold; text-transform: uppercase; letter-spacing: 1px; text-shadow: 2px
2px 4px rgba(0, 0, 0, 0.3);">Now Verifying the Column Names</p>
120 #
121
122 # %%
123 df.columns
124
125 # %% [markdown]
126 # <p style="color: #FF6347; font-family: 'Helvetica Neue', Helvetica, Arial, sans-serif; font-
size: 24px; font-weight: bold; text-transform: uppercase; letter-spacing: 1px; text-shadow: 2px
2px 4px rgba(0, 0, 0, 0.3);">Exploring the Netflix Dataset: Uncovering Insights with the info()
Function</p>
127
128 # %%
129 df.info()
130
131 # %% [markdown]
132 # <h1 style="color: #FF1140; font-family: 'Arial Black', Arial, sans-serif; font-size: 2em;
text-transform: uppercase; letter-spacing: 4px; text-shadow: 2px 2px 4px rgba(0, 0, 0,
0.7);">Observations</h1>
133 # <ul style="list-style-type: none; padding-left: 0;">
134 #   <li style="color: #FF9900; font-weight: bold; margin-bottom: 10px;">In our analysis of
the Netflix dataset, we uncovered 8807 entries spread across 12 columns. Notably, certain
columns like 'director', 'cast', 'country', 'rating', and 'duration' exhibit missing values,
prompting further investigation. Moreover, while one column primarily comprises integer values,
the remaining 11 columns are represented as object data types, indicating a predominance of

```

```
textual information. This diverse mix of data types presents an intriguing opportunity for
exploration, promising valuable insights into Netflix's extensive content catalog.</li>
135 # </ul>
136 #
137
138 # %% [markdown]
139 # <p style="color: #FF6347; font-family: 'Helvetica Neue', Helvetica, Arial, sans-serif; font-
size: 24px; font-weight: bold; text-transform: uppercase; letter-spacing: 1px; text-shadow: 2px
2px 4px rgba(0, 0, 0, 0.3);">An Overview of Statistics</p>
140
141 # %%
142 df.describe()
143
144 # %% [markdown]
145 # <div style="display: flex; justify-content: center; align-items: center; min-height: 50px;
min-width: 500px; background-color: #334445; border-radius: 20px; box-shadow: 0px 0px 10px
rgba(0, 0, 0, 0.5);">
146 #   <h1 style="color: #FF1140; font-family: 'Arial Black', Arial, sans-serif; font-size: 2em;
text-transform: uppercase; letter-spacing: 4px; text-shadow: 2px 2px 4px rgba(0, 0, 0, 0.7);">
🔍 Dealing with Missing Values 🔍 </h1><br>
147 # </div>
148
149 # %% [markdown]
150 # <p style="color: #FF6347; font-family: 'Helvetica Neue', Helvetica, Arial, sans-serif; font-
size: 24px; font-weight: bold; text-transform: uppercase; letter-spacing: 1px; text-shadow: 2px
2px 4px rgba(0, 0, 0, 0.3);">Finding the Missing Values</p>
151
152 # %%
153 df.isnull().sum()
154
155 # %% [markdown]
156 # <p style="color: #FF6347; font-family: 'Helvetica Neue', Helvetica, Arial, sans-serif; font-
size: 24px; font-weight: bold; text-transform: uppercase; letter-spacing: 1px; text-shadow: 2px
2px 4px rgba(0, 0, 0, 0.3);">Checking the percentage of missing value</p>
157
158 # %%
159 df.isnull().sum()/len(df)*100
160
161 # %% [markdown]
162 # <div style="display: flex; justify-content: center; align-items: center; min-height: 50px;
min-width: 500px; background-color: #334445; border-radius: 20px; box-shadow: 0px 0px 10px
rgba(0, 0, 0, 0.5);">
163 #   <h1 style="color: #FF1140; font-family: 'Arial Black', Arial, sans-serif; font-size: 2em;
text-transform: uppercase; letter-spacing: 4px; text-shadow: 2px 2px 4px rgba(0, 0, 0, 0.7);">
164 # 📊 Data Visualization 📊 </h1><br>
165 # </div>
166
167 # %% [markdown]
168 # <p style="color: #FF6347; font-family: 'Helvetica Neue', Helvetica, Arial, sans-serif; font-
size: 24px; font-weight: bold; text-transform: uppercase; letter-spacing: 1px; text-shadow: 2px
2px 4px rgba(0, 0, 0, 0.3);">Visualizing Missing Values in the Dataset: A Plotting Guide</p>
169
```

```
170 # %%
171 # Create a plot show the missing values
172 sns.heatmap(df.isnull())
173 plt.show()
174
175 # %% [markdown]
176 # <div class="container" style="font-family: Arial, sans-serif; background-color: #f0f0f0;
padding: 20px; border-radius: 10px; box-shadow: 0px 0px 10px rgba(0, 0, 0, 0.1);">
177 #     <h1 style="color: #FF1140; font-size: 2em; text-transform: uppercase; letter-spacing:
2px; text-shadow: 2px 2px 4px rgba(0, 0, 0, 0.2);">Observations</h1>
178 #     <p style="font-size: 1.1em; color: #FF9900;">To ensure the accuracy of our analysis, we
need to address missing data. There are several approaches we can take:</p>
179 #     <ul style="list-style-type: none; padding-left: 0; margin-top: 10px;">
180 #         <li style="margin-bottom: 10px; background-color: #fff; padding: 10px; border-radius:
5px; box-shadow: 0px 2px 4px rgba(0, 0, 0, 0.1);">
181 #             <strong>1. Remove Rows:</strong><br>
182 #             The missing values in the columns date_added, rating, and duration are 0.113546%,
0.045418%, and 0.034064%, respectively. Due to the low percentage, rows with missing data in
these columns were dropped.
183 #         </li>
184 #         <li style="margin-bottom: 10px; background-color: #fff; padding: 10px; border-radius:
5px; box-shadow: 0px 2px 4px rgba(0, 0, 0, 0.1);">
185 #             <strong>2. Mode Imputation:</strong><br>
186 #             Considering the high percentage of missing values in the director, cast, and
country columns, we decided to impute these missing values using the mode method.
187 #         </li>
188 #     </ul>
189 # </div>
190 #
191
192 # %%
193 df.dropna(subset= ["date_added", "rating", "duration"], inplace=True)
194
195 # %%
196 df["director"].fillna(df["director"].mode()[0], inplace=True)
197 df["cast"].fillna(df["cast"].mode()[0], inplace=True)
198 df["country"].fillna(df["country"].mode()[0], inplace=True)
199
200 # %%
201 df.isnull().sum()
202
203 # %% [markdown]
204 # <p style="color: #FF6347; font-family: 'Helvetica Neue', Helvetica, Arial, sans-serif; font-
size: 24px; font-weight: bold; text-transform: uppercase; letter-spacing: 1px; text-shadow: 2px
2px 4px rgba(0, 0, 0, 0.3);">Checking for Duplicates in Your Dataset Effectively</p>
205
206 # %%
207 # check this duplicates in dataset
208 df.duplicated().sum()
209
210 # %%
```

```

211 df["show_id"].unique()
212
213 # %%
214 df["show_id"].duplicated().sum()
215
216 # %% [markdown]
217 # <div class="container" style="font-family: Arial, sans-serif; background-color: #f0f0f0;
padding: 20px; border-radius: 10px; box-shadow: 0px 0px 10px rgba(0, 0, 0, 0.1);">
218 #     <p style="font-size: 1.1em; color: black; line-height: 1.6;"> To verify that there are
no duplicates in the dataset, I checked the Show ID column. This column should contain unique
values, ensuring there are no repeated entries. By using this method, I confirmed that the
dataset has zero duplicates in the Show ID column.</p>
219 # </div>
220
221 # %% [markdown]
222 # <div style="display: flex; justify-content: center; align-items: center; min-height: 50px;
min-width: 500px; background-color: #334445; border-radius: 20px; box-shadow: 0px 0px 10px
rgba(0, 0, 0, 0.5);">
223 #     <h1 style="color: #FF1140; font-family: 'Arial Black', Arial, sans-serif; font-size: 2em;
text-transform: uppercase; letter-spacing: 4px; text-shadow: 2px 2px 4px rgba(0, 0, 0, 0.7);">
Questions Answered ? </h1><br>
224 # </div>
225
226 # %% [markdown]
227 # <div class="container" style="font-family: Arial, sans-serif; background-color: #f0f0f0;
padding: 10px; border-radius: 10px; box-shadow: 0px 0px 10px rgba(0, 0, 0, 0.1);">
228 #     <ul style="list-style-type: none; padding-left: 0;">
229 #         <li style="font-size: 2em; color: black; line-height: 1.6;">Question 1. What is the
distribution of content types on Netflix?</li>
230 #     </ul>
231 # </div>
232
233 # %%
234 df["type"].value_counts()
235
236 # %%
237 fig = px.bar(x=df["type"].value_counts().index,
238             y=df["type"].value_counts().values,
239             color=df["type"].value_counts().index)
240
241 # Update layout with customized title, axis labels, and background colors
242 fig.update_layout(
243     title={
244         'text': "Distribution of Content Type",
245         'font': {'color': 'white'}
246     },
247     xaxis_title={
248         'text': "Content Type",
249         'font': {'color': 'white'}
250     },
251     yaxis_title={

```

```

252         'text': "Count",
253         'font': {'color': 'white'}
254     },
255     plot_bgcolor='rgba(0,0,0,0)',
256     paper_bgcolor='black',
257     font=dict(color='white')
258 )
259
260 fig.show()
261
262
263 # %%
264 import plotly.express as px
265
266 fig = px.pie(df, values=df["type"].value_counts().values,
267             names=df["type"].value_counts().index)
268
269 fig.update_layout(title="Distribution of Content Type", plot_bgcolor='rgba(0,0,0,0)',
270                 paper_bgcolor='black', font=dict(color='white'))
271
272
273 # %% [markdown]
274 # <div class="container" style="font-family: Arial, sans-serif; background-color: #f0f0f0;
275 # padding: 20px; border-radius: 10px; box-shadow: 0px 0px 10px rgba(0, 0, 0, 0.1);">
276 #     <ul style="list-style-type: none; padding-left: 0;">
277 #         <li style="font-size: 1.1em; color: black; line-height: 1.6;">
278 #             <strong>Answer:</strong><br>
279 #             The dataset contains two main content types: Movies and TV Shows. The
280 #             distribution is as follows:<br>
281 #             1. Movies: 69.7%<br>
282 #             2. TV Show: 30.3%
283 #         </li>
284 #     </ul>
285 # </div>
286 #
287
288 # %% [markdown]
289 # <div class="container" style="font-family: Arial, sans-serif; background-color: #f0f0f0;
290 # padding: 10px; border-radius: 10px; box-shadow: 0px 0px 10px rgba(0, 0, 0, 0.1);">
291 #     <ul style="list-style-type: none; padding-left: 0;">
292 #         <li style="font-size: 2em; color: black; line-height: 1.6;">Question 2. What are the
293 #             highest release years for TV shows and movies in the Netflix dataset?</li>
294 #     </ul>
295 # </div>
296 #
297
298 # %%
299 df["release_year"].unique()
300
301 # %%

```



```

297 df["release_year"].value_counts()
298
299 # %%
300
301 fig = px.bar(x=df["release_year"].value_counts().index,
302             y=df["release_year"].value_counts().values,
303             color=df["release_year"].value_counts().index)
304
305 # Update layout with customized title, axis labels, and background colors
306
307 fig.update_layout(
308     title={
309         'text': "Highest Release Years for TV Shows and Movies",
310         'font': {'color': 'white'}
311     },
312     xaxis_title={
313         'text': "Release Year",
314         'font': {'color': 'white'}
315     },
316     yaxis_title={
317         'text': "Count",
318         'font': {'color': 'white'}
319     },
320     plot_bgcolor='rgba(0,0,0,0)',
321     paper_bgcolor='black',
322     font=dict(color='white')
323 )
324
325 # %% [markdown]
326 # <div class="container" style="font-family: Arial, sans-serif; background-color: #f0f0f0;
padding: 20px; border-radius: 10px; box-shadow: 0px 0px 10px rgba(0, 0, 0, 0.1);">
327 #     <ul style="list-style-type: none; padding-left: 0;">
328 #         <li style="font-size: 1.1em; color: black; line-height: 1.6;">
329 #             <strong>Answer:</strong><br>
330 #             the highest release year for both TV shows and movies is 2018. No data exceeds
that year.
331 #         </li>
332 #     </ul>
333 # </div>
334
335 # %%
336 grouped_data = df.groupby(['release_year', 'type']).size().reset_index(name='count')
337
338 fig = px.bar(grouped_data, x='release_year', y='count', color='type',
339             title="Highest Release Years for TV Shows and Movies",
340             labels={'release_year': "Release Year", 'count': "Count", 'type': "Type"},
341             color_discrete_map={'Movie': 'blue', 'TV Show': 'green'})
342 fig.update_layout(
343     plot_bgcolor='rgba(0,0,0,0)',
344     paper_bgcolor='black',

```

```

345     font=dict(color='white')
346 )
347 highest_release_year = grouped_data[grouped_data["count"] == grouped_data["count"].max()]
348 ["release_year"].iloc[0]
349 highest_type = grouped_data[grouped_data["count"] == grouped_data["count"].max()]
350 ["type"].iloc[0]
351 print(f"The highest type on the release year {highest_release_year} is {highest_type}.")
352 fig.show()
353
354 # %% [markdown]
355 # <!DOCTYPE html>
356 # <html lang="en">
357 # <head>
358 #     <meta charset="UTF-8">
359 #     <meta name="viewport" content="width=device-width, initial-scale=1.0">
360 #     <title>Kernel Version</title>
361 #     <style>
362 #         body {
363 #             background-color: #e0f7fa; /* Changed background color */
364 #             font-family: 'Arial', sans-serif;
365 #             display: flex;
366 #             justify-content: center;
367 #             align-items: center;
368 #             height: 100vh;
369 #             margin: 0;
370 #         }
371 #         .container {
372 #             text-align: center;
373 #             padding: 20px 30px; /* Reduced padding */
374 #             border-radius: 15px;
375 #             box-shadow: 0 4px 8px rgba(0, 0, 0, 0.2);
376 #             max-width: 700px; /* Restricted width */
377 #             width: 100%;
378 #         }
379 #         h1 {
380 #             color: black; /* Changed text color */
381 #             font-size: 1.5em; /* Reduced font size */
382 #             margin: 0;
383 #             text-align: left; /* Align heading to the left */
384 #         }
385 #         p {
386 #             font-size: 1em; /* Reduced font size */
387 #             color: #000; /* Changed text color */
388 #             background: rgba(0, 0, 0, 0.1);
389 #             padding: 10px;
390 #             border-radius: 10px;
391 #             display: inline-block;
392 #             margin-top: 10px;
393 #         }
394 #     </style>

```

```
393 # </head>
394 # <body>
395 #     <div class="container">
396 #         <h1>Observations</h1> <!-- Move the h1 tag inside the p tag -->
397 #         <ul>
398 #             <p>From the visualization of movie and TV show releases by year, it's evident
399 #             that the count of movie releases is consistently higher than TV shows across multiple years.
400 #             This suggests a stronger focus on producing movies within the Netflix dataset.</p>
401 #         </ul>
402 #     </div>
403 # </body>
404 # </html>
405 # %% [markdown]
406 # <!DOCTYPE html>
407 # <html lang="en">
408 # <head>
409 #     <meta charset="UTF-8">
410 #     <meta name="viewport" content="width=device-width, initial-scale=1.0">
411 #     <title>Kernel Version</title>
412 #     <style>
413 #         body {
414 #             background-color: #e0f7fa; /* Changed background color */
415 #             font-family: 'Arial', sans-serif;
416 #             display: flex;
417 #             justify-content: center;
418 #             align-items: center;
419 #             height: 100vh;
420 #             margin: 0;
421 #         }
422 #         .container {
423 #             padding: 20px 30px; /* Reduced padding */
424 #             border-radius: 15px;
425 #             box-shadow: 0 4px 8px rgba(0, 0, 0, 0.2);
426 #             max-width: 700px; /* Restricted width */
427 #             width: 100%;
428 #         }
429 #         p {
430 #             font-size: 1em; /* Reduced font size */
431 #             color: #000; /* Changed text color */
432 #             background: rgba(0, 0, 0, 0.1);
433 #             padding: 10px;
434 #             border-radius: 10px;
435 #             display: inline-block;
436 #             margin-top: 10px;
437 #             text-align: left; /* Align text to the left */
438 #         }
439 #     </style>
440 # </head>
```

```

441 # <body>
442 #     <div class="container">
443 #         <ul>
444 #             <p>
445 # Question 3. Which country produces the most content on Netflix?</p>
446 #         </ul>
447 #     </div>
448 # </body>
449 # </html>
450
451 # %%
452 grouped_data = df.groupby(['type', 'country']).size().reset_index(name='count')
453 grouped_data
454
455 # %%
456 grouped_data = df.groupby(['type', 'country']).size().reset_index(name='count')
457 grouped_data = grouped_data.sort_values(by="count", ascending=False)
458 grouped_data = grouped_data.head(10)
459 fig = px.bar(grouped_data, x='country', y='count', color='country',
460             title="Highest Countries for Movies&TV Show",
461             labels={'country': "Country", 'count': "Count"},
462             color_discrete_map={'United States': 'blue', 'India': 'green'})
463 fig.update_layout(
464     plot_bgcolor='rgba(0,0,0,0)',
465     paper_bgcolor='black',
466     font=dict(color='white')
467 )
468 fig.show()
469
470 # %% [markdown]
471 # <div class="container" style="font-family: Arial, sans-serif; background-color: #f0f0f0;
padding: 20px; border-radius: 10px; box-shadow: 0px 0px 10px rgba(0, 0, 0, 0.1);">
472 #     <ul style="list-style-type: none; padding-left: 0;">
473 #         <li style="font-size: 1.1em; color: black; line-height: 1.6;">
474 #             <strong>Answer:</strong><br>
475 #             The United States is the leading content producer on Netflix, followed by India,
the United Kingdom, and Japan.
476 #         </li>
477 #     </ul>
478 # </div>
479
480 # %% [markdown]
481 # <!DOCTYPE html>
482 # <html lang="en">
483 # <head>
484 #     <meta charset="UTF-8">
485 #     <meta name="viewport" content="width=device-width, initial-scale=1.0">
486 #     <title>Kernel Version</title>
487 #     <style>
488 #         body {

```

```

489 #         background-color: #e0f7fa; /* Changed background color */
490 #         font-family: 'Arial', sans-serif;
491 #         display: flex;
492 #         justify-content: center;
493 #         align-items: center;
494 #         height: 100vh;
495 #         margin: 0;
496 #     }
497 #     .container {
498 #         padding: 20px 30px; /* Reduced padding */
499 #         border-radius: 15px;
500 #         box-shadow: 0 4px 8px rgba(0, 0, 0, 0.2);
501 #         max-width: 700px; /* Restricted width */
502 #         width: 100%;
503 #     }
504 #     p {
505 #         font-size: 1em; /* Reduced font size */
506 #         color: #000; /* Changed text color */
507 #         background: rgba(0, 0, 0, 0.1);
508 #         padding: 10px;
509 #         border-radius: 10px;
510 #         display: inline-block;
511 #         margin-top: 10px;
512 #         text-align: left; /* Align text to the left */
513 #     }
514 # </style>
515 # </head>
516 # <body>
517 #     <div class="container">
518 #         <ul>
519 #             <p>Question 4. What is the average duration of Movies and TV Shows?</p>
520 #         </ul>
521 #     </div>
522 # </body>
523 # </html>
524
525 # %%
526 grouped_data = df.groupby(['type', 'duration']).size().reset_index(name='count')
527
528 grouped_data = grouped_data.sort_values(by="count", ascending=False)
529 grouped_data
530
531 # %%
532 # Create a bar plot
533
534 fig = px.bar(grouped_data, x='duration', y='count', color='type',
535              title="Duration of Movies&TV Show",
536              labels={'duration': "Duration", 'count': "Count"},
537              color_discrete_map={'Short': 'blue', 'Long': 'green'})
538 fig.update_layout(

```

```

539     plot_bgcolor='rgba(0,0,0,0)',
540     paper_bgcolor='black',
541     font=dict(color='white')
542 )
543
544 # %% [markdown]
545 # <div class="container" style="font-family: Arial, sans-serif; background-color: #f0f0f0;
padding: 20px; border-radius: 10px; box-shadow: 0px 0px 10px rgba(0, 0, 0, 0.1);">
546 #     <ul style="list-style-type: none; padding-left: 0;">
547 #         <li style="font-size: 1.1em; color: black; line-height: 1.6;">
548 #             <strong>Answer:</strong><br>
549 #                 1. Movies: The average duration is approximately 90 minutes.<br>
550 # 2. TV Shows: The average number of seasons per show is 1.
551 #         </li>
552 #     </ul>
553 # </div>
554
555 # %%
556 grouped_data = df.groupby(['director']).size().reset_index(name='count')
557 grouped_data.head(10)
558
559 # %%
560 # create a bar plot show the top director
561
562 fig = px.bar(grouped_data.head(10), x='director', y='count',
563             title="Top 10 directors",
564             labels={'director': "Director", 'count': "Count"},
565             color_discrete_map={'United States': 'blue', 'India': 'green'})
566 fig.update_layout(
567     plot_bgcolor='rgba(0,0,0,0)',
568     paper_bgcolor='black',
569     font=dict(color='white')
570 )
571 fig.show()
572
573 # %% [markdown]
574 # <div class="container" style="font-family: Arial, sans-serif; background-color: #f0f0f0;
padding: 20px; border-radius: 10px; box-shadow: 0px 0px 10px rgba(0, 0, 0, 0.1);">
575 #     <ul style="list-style-type: none; padding-left: 0;">
576 #         <li style="font-size: 1.1em; color: black; line-height: 1.6;">
577 #             <strong>Answer:</strong><br>
578 #                 The top 10 directors in the Netflix dataset include A. L. Vijay, A.R. Murugadoss,
and Aamir Khan among others, showcasing a diverse range.
579 #         </li>
580 #     </ul>
581 # </div>
582
583 # %% [markdown]
584 # <div style="display: flex; justify-content: center; align-items: center; min-height: 50px;
min-width: 500px; background-color: #334445; border-radius: 20px; box-shadow: 0px 0px 10px
rgba(0, 0, 0, 0.5);">

```

```
585 # <h1 style="color: #FF1140; font-family: 'Arial Black', Arial, sans-serif; font-size: 2em;
text-transform: uppercase; letter-spacing: 4px; text-shadow: 2px 2px 4px rgba(0, 0, 0, 0.7);">
📄 Summary 📄 </h1><br>
586 # </div>
587
588 # %% [markdown]
589 # <div class="container" style="font-family: Arial, sans-serif; background-color: #f0f0f0;
padding: 20px; border-radius: 10px; box-shadow: 0px 0px 10px rgba(0, 0, 0, 0.1);">
590 # <p style="font-size: 1.1em; color: black; line-height: 1.6;">The Netflix dataset has 8807
entries: 69.7% movies and 30.3% TV shows. The peak release year is 2018. Top countries are the
USA, India, and the UK. Popular ratings are TV-MA and TV-14, with common genres being dramas,
comedies, and documentaries, showcasing Netflix's diverse and global content.</p>
591 # </div>
592
593 # %% [markdown]
594 # <div style="display: flex; justify-content: center; align-items: center; min-height: 50px;
min-width: 500px; background-color: #334445; border-radius: 20px; box-shadow: 0px 0px 10px
rgba(0, 0, 0, 0.5);">
595 # <h1 style="color: #FF1140; font-family: 'Arial Black', Arial, sans-serif; font-size: 2em;
text-transform: uppercase; letter-spacing: 4px; text-shadow: 2px 2px 4px rgba(0, 0, 0, 0.7);">
🎉 Conclusion 🎉 </h1><br>
596 # </div>
597
598 # %% [markdown]
599 # <div class="container" style="font-family: Arial, sans-serif; background-color: #f0f0f0;
padding: 20px; border-radius: 10px; box-shadow: 0px 0px 10px rgba(0, 0, 0, 0.1);">
600 # <p style="font-size: 1.1em; color: black; line-height: 1.6;">The Netflix dataset reveals
a dynamic and expansive content library, with a strong emphasis on movies, particularly in
recent years. The diversity in genres, ratings, and international representation highlights
Netflix's strategy to cater to a broad, global audience. This rich variety ensures that Netflix
remains a leading platform for entertainment, appealing to diverse tastes and preferences
worldwide.</p>
601 # </div>
602
603 # %% [markdown]
604 # <div style="display: flex; justify-content: center; align-items: center; min-height: 50px;
min-width: 500px; background-color: #334445; border-radius: 20px; box-shadow: 0px 0px 10px
rgba(0, 0, 0, 0.5);">
605 # <h1 style="color: #FF1140; font-family: 'Arial Black', Arial, sans-serif; font-size: 2em;
text-transform: uppercase; letter-spacing: 4px; text-shadow: 2px 2px 4px rgba(0, 0, 0, 0.7);">
🎉 The END 🎉 </h1><br>
606 # </div>
```