Untitled-1

```
# %% [markdown]
# <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;">
#
#
# <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>
# </div>

# %% [markdown]
# ![image](netflix.png)

# %% [markdown]
# <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);">
#   <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>
# </div>
#

# %% [markdown]
# <div style="border: 4px solid #FF5733; padding: 10px; max-width: 500px; margin: 0 auto; border-radius: 20px;">
#     <p style="color: #4CAF50;">
#         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! 📊🔍🌟
#     </p>
# </div>
#

# %% [markdown]
# <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>
#
# <div style="text-align: left;">
#     <table>
#         <tr>
#             <td><a href="https://www.linkedin.com/in/umar-mehmood-147224294/"><img src="https://img.shields.io/badge/LinkedIn-Profile-blue?style=for-the-badge&logo=linkedin" alt="LinkedIn"/></a></td>
#             <td><a href="mailto:umarmehmood885@gmail.com"><img src="https://img.shields.io/badge/Gmail-Profile-red?style=for-the-badge&logo=gmail"
```

```
alt="Gmail"/></a></td>
#         </tr>
#         <tr>
#             <td><a href="https://github.com/Umar885"><img src="https://img.shields.io/badge/GitHub-Profile-green?style=for-the-badge&logo=github" alt="GitHub"/></a></td>
#             <td><a href="https://twitter.com/UmarMeh12214006"><img src="https://img.shields.io/badge/Twitter-Profile-blue?style=for-the-badge&logo=twitter" alt="Twitter"/></a></td>
#         </tr>
#         <tr>
#             <td><a href="https://www.kaggle.com/umarmehmood"><img src="https://img.shields.io/badge/Kaggle-Profile-orange?style=for-the-badge&logo=kaggle" alt="Kaggle"/></a></td>
#             <td><a href="https://medium.com/@umarmehmood885"><img src="https://img.shields.io/badge/Medium-Profile-black?style=for-the-badge&logo=medium" alt="Medium"/></a></td>
#         </tr>
#     </table>
# </div>
#


# %% [markdown]
# <a id="1"></a>
# <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);">
#   <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>
# </div>

# %% [markdown]
#
# ***🎬📊 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. 🌟📈
# 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.***

# %% [markdown]
# <a id="2"></a>
# <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);">
#   <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 🎬
#  </h1><br>
# </div>

```

```
# %% [markdown]
# ***📺 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. 🎬📊***
# **The following links allowed users to obtain the Data Set from Kaggle** [link](https://www.kaggle.com/datasets/shivamb/netflix-shows)

# %% [markdown]
# <div style="text-align: center;">
#     <h1 style="color: #FF1140;">Kernel Version Used:</h1>
#     <ul style="list-style-type: none; padding: 0;">
#         <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>
#     </ul>
# </div>
#

# %% [markdown]
# <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);">
#   <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>
# </div>

# %%
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
%matplotlib inline

# %% [markdown]
# <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);">
#   <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>
# </div>

# %%
df = pd.read_csv("netflix_titles.csv")

# %% [markdown]
# <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>
# </ul>
#

# %% [markdown]
# <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>

# %%
df.describe()

# %% [markdown]
# <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);">
#   <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>
# </div>

# %% [markdown]
# <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>

# %%
df.isnull().sum()

# %% [markdown]
# <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>

# %%
df.isnull().sum()/len(df)*100

# %% [markdown]
# <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);">
#   <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);">
# 📊 Data Visualization 📊 </h1><br>
# </div>

# %% [markdown]
# <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>

```

```
# %%
# Create a plot show the missing values
sns.heatmap(df.isnull())
plt.show()

# %% [markdown]
# <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);">
#     <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>
#     <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>
#     <ul style="list-style-type: none; padding-left: 0; margin-top: 10px;">
#         <li style="margin-bottom: 10px; background-color: #fff; padding: 10px; border-radius: 5px; box-shadow: 0px 2px 4px rgba(0, 0, 0, 0.1);">
#             <strong>1. Remove Rows:</strong><br>
#             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.
#         </li>
#         <li style="margin-bottom: 10px; background-color: #fff; padding: 10px; border-radius: 5px; box-shadow: 0px 2px 4px rgba(0, 0, 0, 0.1);">
#             <strong>2. Mode Imputation:</strong><br>
#             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.
#         </li>
#     </ul>
# </div>
#

# %%
df.dropna(subset= ["date_added", "rating", "duration"], inplace=True)

# %%
df["director"].fillna(df["director"].mode()[0], inplace=True)
df["cast"].fillna(df["cast"].mode()[0], inplace=True)
df["country"].fillna(df["country"].mode()[0], inplace=True)

# %%
df.isnull().sum()

# %% [markdown]
# <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>

# %%
# check this duplicates in dataset
df.duplicated().sum()

# %%
```

```
df["show_id"].unique()

# %%
df["show_id"].duplicated().sum()

# %% [markdown]
# <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);">
#     <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>
# </div>

# %% [markdown]
# <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);">
#   <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>
# </div>

# %% [markdown]
# <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);">
#     <ul style="list-style-type: none; padding-left: 0;">
#         <li style="font-size: 2em; color: black; line-height: 1.6;">Question 1. What is the distribution of content types on Netflix?</li>
#     </ul>
# </div>

# %%
df["type"].value_counts()

# %%
fig = px.bar(x=df["type"].value_counts().index,
             y=df["type"].value_counts().values,
             color=df["type"].value_counts().index)

# Update layout with customized title, axis labels, and background colors
fig.update_layout(
    title={
        'text': "Distribution of Content Type",
        'font': {'color': 'white'}
    },
    xaxis_title={
        'text': "Content Type",
        'font': {'color': 'white'}
    },
    yaxis_title={
```

```
        'text': "Count",
        'font': {'color': 'white'}
    },
    plot_bgcolor='rgba(0,0,0,0)',
    paper_bgcolor='black',
    font=dict(color='white')
)

fig.show()


# %%
import plotly.express as px

fig = px.pie(df, values=df["type"].value_counts().values, names=df["type"].value_counts().index)

fig.update_layout(title="Distribution of Content Type", plot_bgcolor='rgba(0,0,0,0)', paper_bgcolor='black', font=dict(color='white'))

fig.show()


# %% [markdown]
# <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);">
#     <ul style="list-style-type: none; padding-left: 0;">
#         <li style="font-size: 1.1em; color: black; line-height: 1.6;">
#             <strong>Answer:</strong><br>
#             The dataset contains two main content types: Movies and TV Shows. The distribution is as follows:<br>
#             1. Movies: 69.7%<br>
#             2. TV Show: 30.3%
#         </li>
#     </ul>
# </div>
#

# %% [markdown]
# <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);">
#     <ul style="list-style-type: none; padding-left: 0;">
#         <li style="font-size: 2em; color: black; line-height: 1.6;">Question 2. What are the highest release years for TV shows and movies in the Netflix dataset?</li>
#     </ul>
# </div>

# %%
df["release_year"].unique()

# %%
```

```
df["release_year"].value_counts()

# %%

fig = px.bar(x=df["release_year"].value_counts().index,
             y=df["release_year"].value_counts().values,
             color=df["release_year"].value_counts().index)

# Update layout with customized title, axis labels, and background colors

fig.update_layout(
    title={
        'text': "Highest Release Years for TV Shows and Movies",
        'font': {'color': 'white'}
    },
    xaxis_title={
        'text': "Release Year",
        'font': {'color': 'white'}
    },
    yaxis_title={
        'text': "Count",
        'font': {'color': 'white'}
    },
    plot_bgcolor='rgba(0,0,0,0)',
    paper_bgcolor='black',
    font=dict(color='white')
)

# %% [markdown]
# <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);">
#     <ul style="list-style-type: none; padding-left: 0;">
#         <li style="font-size: 1.1em; color: black; line-height: 1.6;">
#             <strong>Answer:</strong><br>
#             the highest release year for both TV shows and movies is 2018. No data exceeds
that year.
#         </li>
#     </ul>
# </div>

# %%
grouped_data = df.groupby(['release_year', 'type']).size().reset_index(name='count')

fig = px.bar(grouped_data, x='release_year', y='count', color='type',
             title="Highest Release Years for TV Shows and Movies",
             labels={'release_year': "Release Year", 'count': "Count", 'type': "Type"},
             color_discrete_map={'Movie': 'blue', 'TV Show': 'green'})
fig.update_layout(
    plot_bgcolor='rgba(0,0,0,0)',
    paper_bgcolor='black',
```

```
    font=dict(color='white')
)
highest_release_year = grouped_data[grouped_data["count"] == grouped_data["count"].max()]["release_year"].iloc[0]
highest_type = grouped_data[grouped_data["count"] == grouped_data["count"].max()]["type"].iloc[0]
print(f"The highest type on the release year {highest_release_year} is {highest_type}.")
fig.show()

# %% [markdown]
# <!DOCTYPE html>
# <html lang="en">
# <head>
#     <meta charset="UTF-8">
#     <meta name="viewport" content="width=device-width, initial-scale=1.0">
#     <title>Kernel Version</title>
#     <style>
#         body {
#             background-color: #e0f7fa; /* Changed background color */
#             font-family: 'Arial', sans-serif;
#             display: flex;
#             justify-content: center;
#             align-items: center;
#             height: 100vh;
#             margin: 0;
#         }
#         .container {
#             text-align: center;
#             padding: 20px 30px; /* Reduced padding */
#             border-radius: 15px;
#             box-shadow: 0 4px 8px rgba(0, 0, 0, 0.2);
#             max-width: 700px; /* Restricted width */
#             width: 100%;
#         }
#         h1 {
#             color: black; /* Changed text color */
#             font-size: 1.5em; /* Reduced font size */
#             margin: 0;
#             text-align: left; /* Align heading to the left */
#         }
#         p {
#             font-size: 1em; /* Reduced font size */
#             color: #000; /* Changed text color */
#             background: rgba(0, 0, 0, 0.1);
#             padding: 10px;
#             border-radius: 10px;
#             display: inline-block;
#             margin-top: 10px;
#         }
#     </style>
```

```
# </head>
# <body>
#     <div class="container">
#         <h1>Observations</h1> <!-- Move the h1 tag inside the p tag -->
#         <ul>
#             <p>From the visualization of movie and TV show releases by year, it's evident that the count of movie releases is consistently higher than TV shows across multiple years. This suggests a stronger focus on producing movies within the Netflix dataset.</p>
#         </ul>
#     </div>
# </body>
# </html>
#

# %% [markdown]
# <!DOCTYPE html>
# <html lang="en">
# <head>
#     <meta charset="UTF-8">
#     <meta name="viewport" content="width=device-width, initial-scale=1.0">
#     <title>Kernel Version</title>
#     <style>
#         body {
#             background-color: #e0f7fa; /* Changed background color */
#             font-family: 'Arial', sans-serif;
#             display: flex;
#             justify-content: center;
#             align-items: center;
#             height: 100vh;
#             margin: 0;
#         }
#         .container {
#             padding: 20px 30px; /* Reduced padding */
#             border-radius: 15px;
#             box-shadow: 0 4px 8px rgba(0, 0, 0, 0.2);
#             max-width: 700px; /* Restricted width */
#             width: 100%;
#         }
#         p {
#             font-size: 1em; /* Reduced font size */
#             color: #000; /* Changed text color */
#             background: rgba(0, 0, 0, 0.1);
#             padding: 10px;
#             border-radius: 10px;
#             display: inline-block;
#             margin-top: 10px;
#             text-align: left; /* Align text to the left */
#         }
#     </style>
# </head>
```

```
# <body>
#     <div class="container">
#         <ul>
#             <p>
# Question 3. Which country produces the most content on Netflix?</p>
#         </ul>
#     </div>
# </body>
# </html>

# %%
grouped_data = df.groupby(['type', 'country']).size().reset_index(name='count')
grouped_data

# %%
grouped_data = df.groupby(['type', 'country']).size().reset_index(name='count')
grouped_data = grouped_data.sort_values(by="count", ascending=False)
grouped_data = grouped_data.head(10)
fig = px.bar(grouped_data, x='country', y='count', color='country',
             title="Highest Countries for Movies&TV Show",
             labels={'country': "Country", 'count': "Count"},
             color_discrete_map={'United States': 'blue', 'India': 'green'})
fig.update_layout(
     plot_bgcolor='rgba(0,0,0,0)',
     paper_bgcolor='black',
     font=dict(color='white')
 )
fig.show()

# %% [markdown]
# <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);">
#     <ul style="list-style-type: none; padding-left: 0;">
#         <li style="font-size: 1.1em; color: black; line-height: 1.6;">
#             <strong>Answer:</strong><br>
#             The United States is the leading content producer on Netflix, followed by India,
the United Kingdom, and Japan.
#         </li>
#     </ul>
# </div>

# %% [markdown]
# <!DOCTYPE html>
# <html lang="en">
# <head>
#     <meta charset="UTF-8">
#     <meta name="viewport" content="width=device-width, initial-scale=1.0">
#     <title>Kernel Version</title>
#     <style>
#         body {
```

```
#             background-color: #e0f7fa; /* Changed background color */
#             font-family: 'Arial', sans-serif;
#             display: flex;
#             justify-content: center;
#             align-items: center;
#             height: 100vh;
#             margin: 0;
#         }
#         .container {
#             padding: 20px 30px; /* Reduced padding */
#             border-radius: 15px;
#             box-shadow: 0 4px 8px rgba(0, 0, 0, 0.2);
#             max-width: 700px; /* Restricted width */
#             width: 100%;
#         }
#         p {
#             font-size: 1em; /* Reduced font size */
#             color: #000; /* Changed text color */
#             background: rgba(0, 0, 0, 0.1);
#             padding: 10px;
#             border-radius: 10px;
#             display: inline-block;
#             margin-top: 10px;
#             text-align: left; /* Align text to the left */
#         }
#     </style>
# </head>
# <body>
#     <div class="container">
#         <ul>
#             <p>Question 4. What is the average duration of Movies and TV Shows?</p>
#         </ul>
#     </div>
# </body>
# </html>

# %%
grouped_data = df.groupby(['type', 'duration']).size().reset_index(name='count')

grouped_data = grouped_data.sort_values(by="count", ascending=False)
grouped_data

# %%
# Create a bar plot

fig = px.bar(grouped_data, x='duration', y='count', color='type',
             title="Duration of Movies&TV Show",
             labels={'duration': "Duration", 'count': "Count"},
             color_discrete_map={'Short': 'blue', 'Long': 'green'})
fig.update_layout(
```

```
    plot_bgcolor='rgba(0,0,0,0)',
    paper_bgcolor='black',
    font=dict(color='white')
 )

# %% [markdown]
# <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);">
#     <ul style="list-style-type: none; padding-left: 0;">
#         <li style="font-size: 1.1em; color: black; line-height: 1.6;">
#             <strong>Answer:</strong><br>
#             1. Movies: The average duration is approximately 90 minutes.<br>
# 2. TV Shows: The average number of seasons per show is 1.
#         </li>
#     </ul>
# </div>

# %%
grouped_data = df.groupby(['director']).size().reset_index(name='count')
grouped_data.head(10)

# %%
# create a bar plot show the top director

fig = px.bar(grouped_data.head(10), x='director', y='count',
                title="Top 10 directors",
                labels={'director': "Director", 'count': "Count"},
                color_discrete_map={'United States': 'blue', 'India': 'green'})
fig.update_layout(
        plot_bgcolor='rgba(0,0,0,0)',
        paper_bgcolor='black',
        font=dict(color='white')
    )
fig.show()

# %% [markdown]
# <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);">
#     <ul style="list-style-type: none; padding-left: 0;">
#         <li style="font-size: 1.1em; color: black; line-height: 1.6;">
#             <strong>Answer:</strong><br>
#             The top 10 directors in the Netflix dataset include A. L. Vijay, A.R. Murugadoss, and Aamir Khan among others, showcasing a diverse range.
#         </li>
#     </ul>
# </div>

# %% [markdown]
# <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);">
```

```
#   <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>
# </div>

# %% [markdown]
# <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);">
#     <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>
# </div>

# %% [markdown]
# <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);">
#   <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>
# </div>

# %% [markdown]
# <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);">
#     <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>
# </div>

# %% [markdown]
# <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);">
#   <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>
# </div>
```