KOVAI.CO ASSESSMENT

Task 1: Prime Dataset

Exploratory Data Analysis:

1. Reading the dataset using pandas

```
import pandas as pd
import numpy as np
```

```
[2] data = pd.read_csv('/content/prime.csv')
```

2. Examining the dataset

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8807 entries, 0 to 8806
Data columns (total 12 columns):
 #   Column        Non-Null Count  Dtype 
---  ------        --------------  ----- 
 0   show_id       8807 non-null   object
 1   type          8807 non-null   object
 2   title         8807 non-null   object
 3   director      8807 non-null   object
 4   cast          8807 non-null   object
 5   country       8807 non-null   object
 6   date_added    8807 non-null   object
 7   release_year  8807 non-null   int64 
 8   rating        8807 non-null   object
 9   duration      8807 non-null   object
 10  listed_in     8807 non-null   object
 11  description   8807 non-null   object
dtypes: int64(1), object(11)
memory usage: 825.8+ KB
```

3. Checking for null values and replacing it with ‘unknown’ or ‘NA’

```
data.isnull().sum()
```

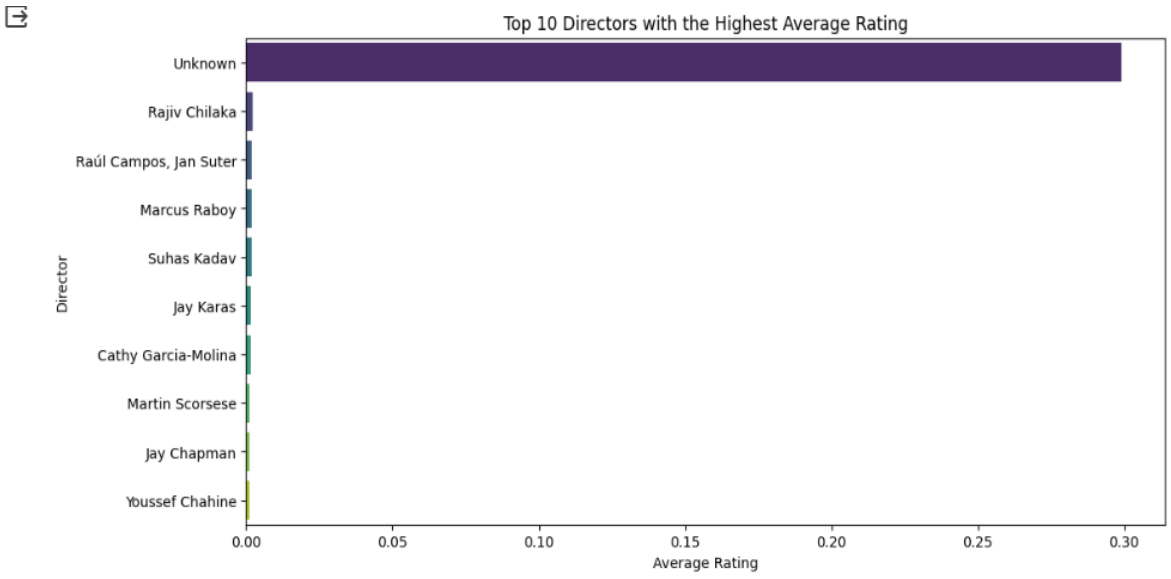
```
show_id          0
type             0
title            0
director         0
cast           825
country        831
date_added      10
release_year     0
rating           4
duration         3
listed_in        0
description      0
```

```
[6] data["director"].fillna("Unknown", inplace = True)
```

```
[8] data["cast"].fillna("Unknown", inplace = True)
```

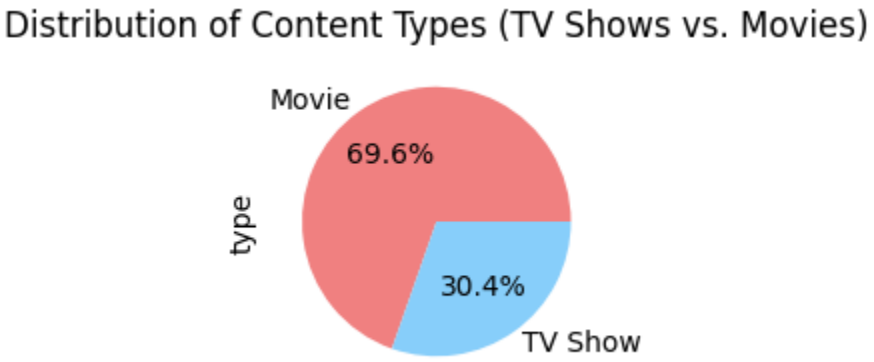
INSIGHTS:

1. Rating vs Director:



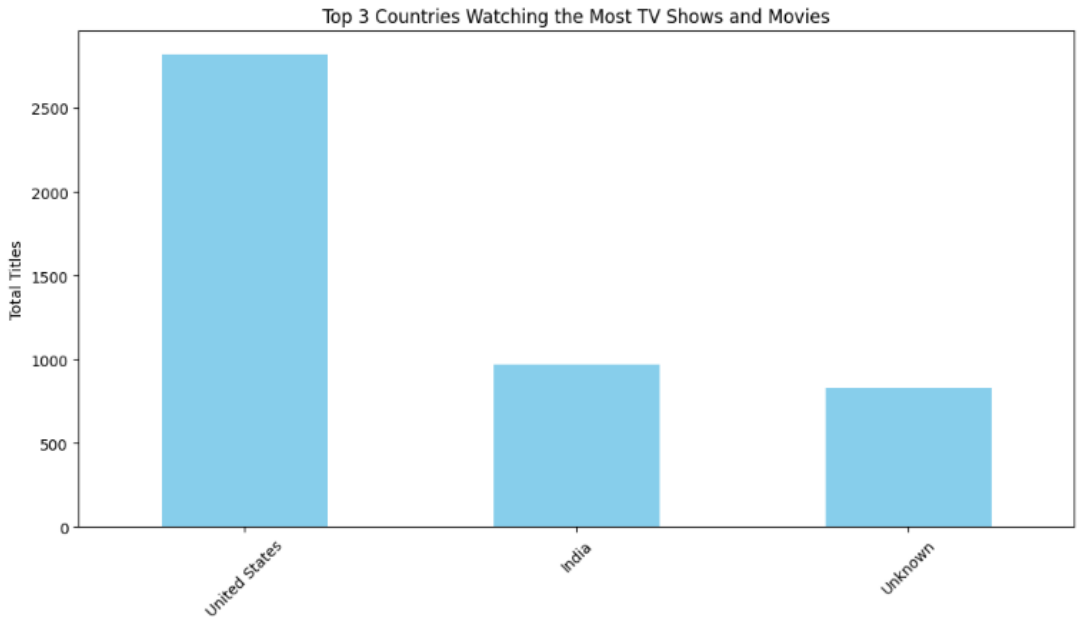
In above diagram we listed top 10 directors whose movies have highest rating. we can infer that the top rating movie's director names are unknown.

2. **Distribution of Content Types (TV Shows vs. Movies)**



The provided code creates a pie chart to visualize the distribution of content types, specifically TV Shows and Movies, in your dataset. From the pie chart we can infer that people watch mostly movies

3. Top three countries that watches TV shows and movies most



From the above figure we can infer that US watch most of TV shows and movies followed by INDIA