```
import pandas as pd
import matplotlib.pyplot as plt
from wordcloud import WordCloud
```

```
from google.colab import drive
drive.mount('/content/drive')
```

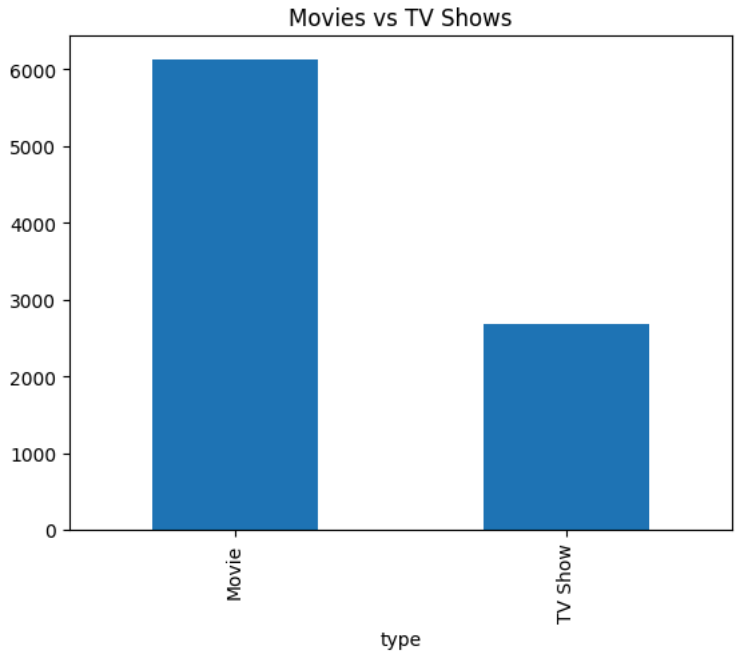
```
Mounted at /content/drive
```

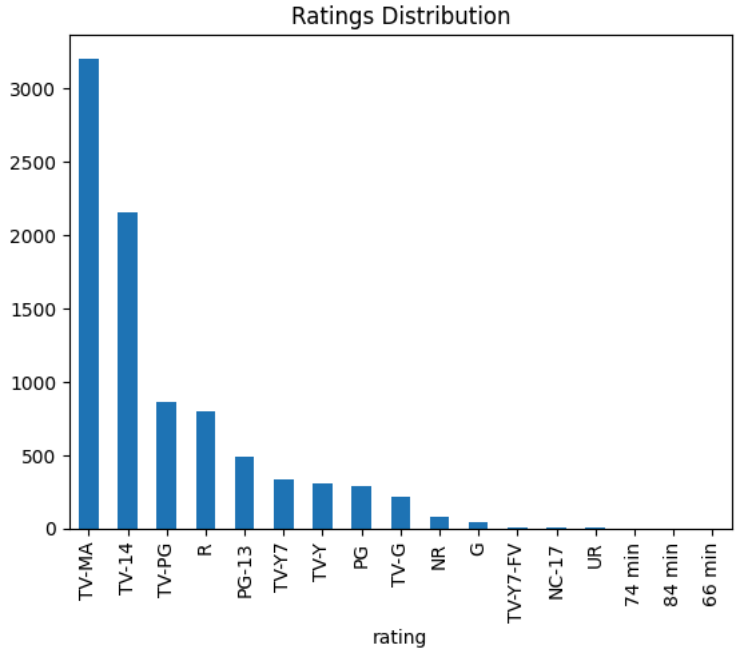
```
import pandas as pd
file_path = "/content/drive/MyDrive/netflix_titles.csv"
```

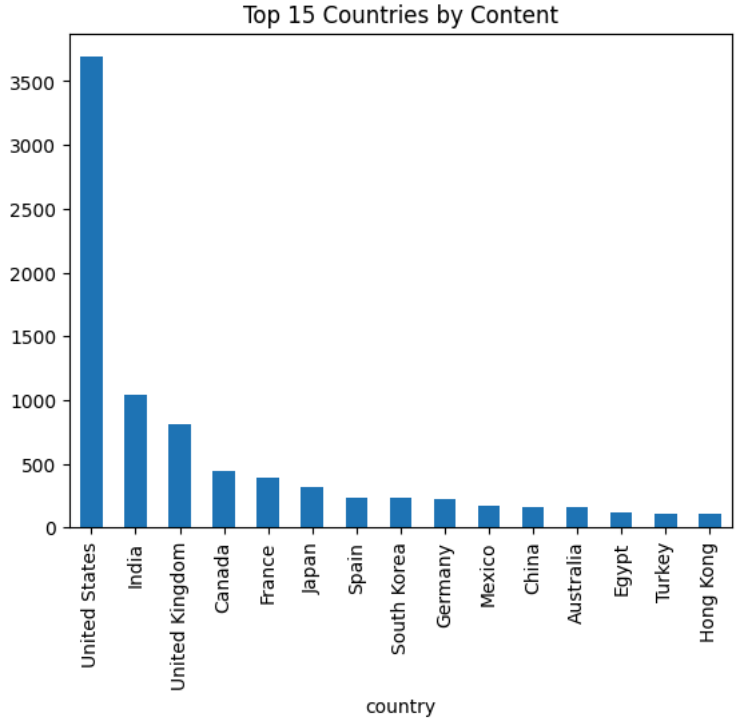
```
df = pd.read_csv(file_path)
df['date_added'] = pd.to_datetime(df['date_added'], errors='coerce')
df['release_year'] = pd.to_numeric(df['release_year'], errors='coerce')

print(df.shape)
print(df.head())
```

```
(8807, 12)
  show_id     type                  title         director  \
0      s1    Movie   Dick Johnson Is Dead  Kirsten Johnson
1      s2  TV Show           Blood & Water              NaN
2      s3  TV Show               Ganglands  Julien Leclercq
3      s4  TV Show  Jailbirds New Orleans              NaN
4      s5  TV Show           Kota Factory              NaN

                                          cast        country  \
0                                          NaN  United States
1  Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...   South Africa
2  Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...            NaN
3                                          NaN            NaN
4  Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...          India

  date_added  release_year rating   duration  \
0 2021-09-25          2020  PG-13     90 min
1 2021-09-24          2021  TV-MA  2 Seasons
2 2021-09-24          2021  TV-MA   1 Season
3 2021-09-24          2021  TV-MA   1 Season
4 2021-09-24          2021  TV-MA  2 Seasons

                                     listed_in  \
0                                Documentaries
1   International TV Shows, TV Dramas, TV Mysteries
2  Crime TV Shows, International TV Shows, TV Act...
3                            Docuseries, Reality TV
4  International TV Shows, Romantic TV Shows, TV ...

                                    description
0  As her father nears the end of his life, filmm...
1  After crossing paths at a party, a Cape Town t...
2  To protect his family from a powerful drug lor...
3  Feuds, flirtations and toilet talk go down amo...
4  In a city of coaching centers known to train I...
```

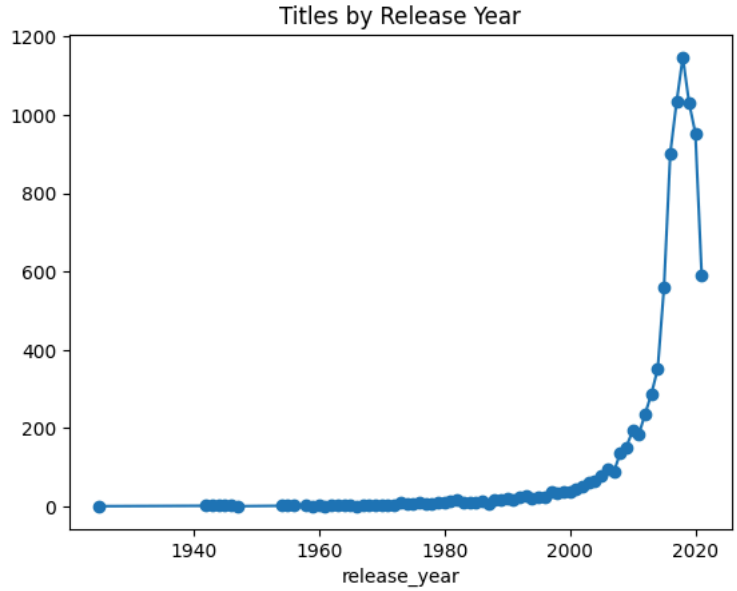
```
df['type'].value_counts().plot(kind='bar', title="Movies vs TV Shows")
plt.show()
df['rating'].value_counts().plot(kind='bar', title="Ratings Distribution")
plt.show()
countries = df['country'].dropna().str.split(',').explode().str.strip()
top_countries = countries.value_counts().head(15)
top_countries.plot(kind='bar', title="Top 15 Countries by Content")
plt.show()
```

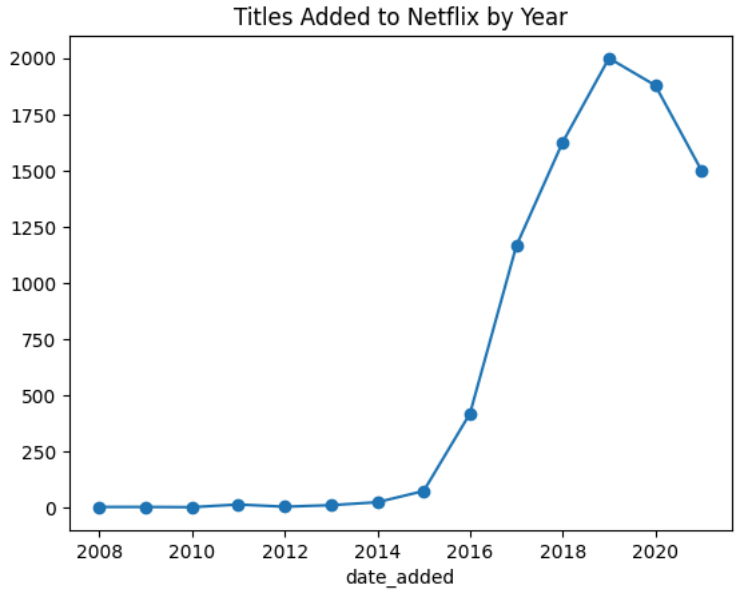

Movies vs TV Shows
6000
5000
4000
3000
2000
1000
0
Movie
TV Show
type

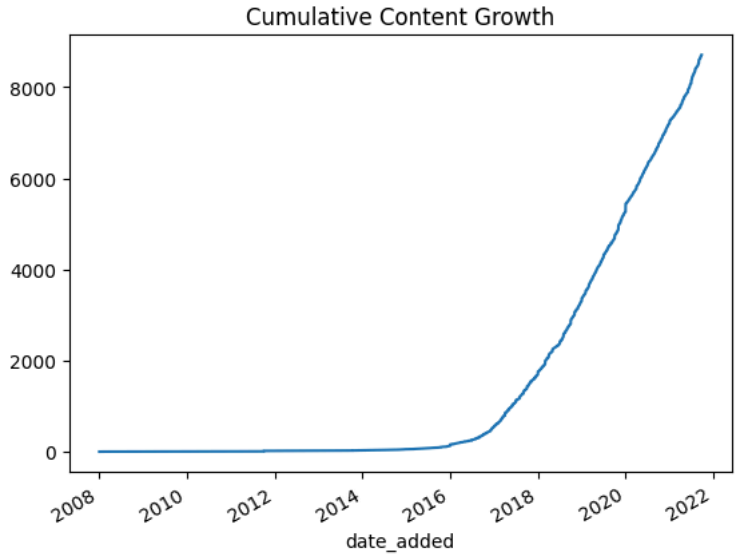

Ratings Distribution
3000
2500
2000
1500
1000
500
0
TV-MA
TV-14
TV-PG
R
PG-13
TV-Y7
TV-Y
PG
TV-G
NR
G
TV-Y7-FV
NC-17
UR
74 min
84 min
66 min
rating


Top 15 Countries by Content
3500
3000
2500
2000
1500
1000
500
0
United States
India
United Kingdom
Canada
France
Japan
Spain
South Korea
Germany
Mexico
China
Australia
Egypt
Turkey
Hong Kong
country

```
df['release_year'].value_counts().sort_index().plot(kind='line', marker='o')
plt.title("Titles by Release Year")
plt.show()
df['date_added'].dt.year.value_counts().sort_index().plot(kind='line', marker='o')
plt.title("Titles Added to Netflix by Year")
plt.show()
added_daily = df.sort_values('date_added')['date_added'].dropna()
growth = pd.Series(1, index=added_daily).cumsum()
growth.plot(title="Cumulative Content Growth")
plt.show()
```

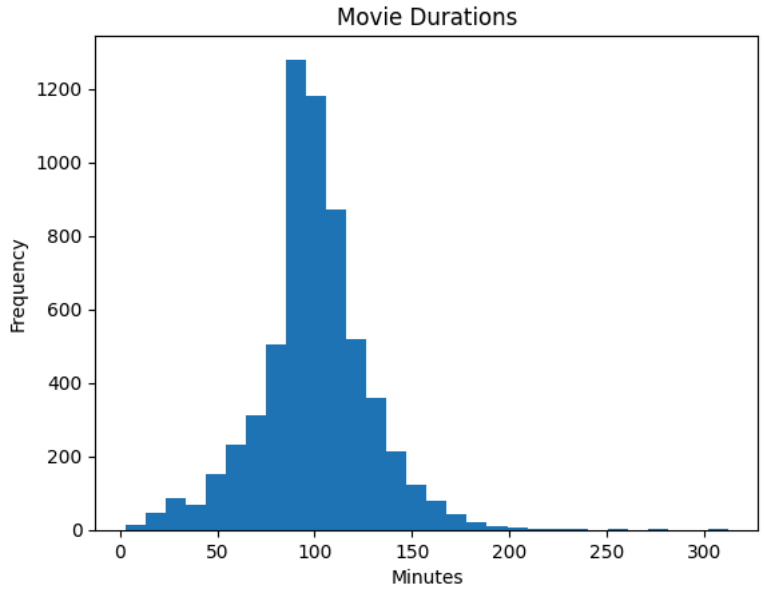




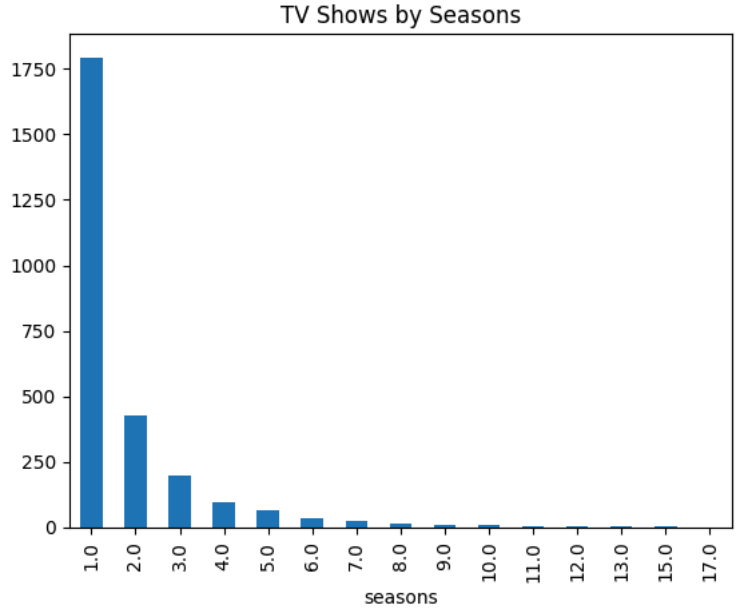



```
movies = df[df['type']=="Movie"]
movies['minutes'] = movies['duration'].str.extract(r'(\d+)').astype(float)
movies['minutes'].dropna().plot(kind='hist', bins=30, title="Movie Durations")
plt.xlabel("Minutes")
plt.show()
shows = df[df['type']=="TV Show"]
shows['seasons'] = shows['duration'].str.extract(r'(\d+)').astype(float)
shows['seasons'].value_counts().sort_index().plot(kind='bar', title="TV Shows by Seasons")
plt.show()
```

```
/tmp/ipython-input-3956297429.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-
  movies['minutes'] = movies['duration'].str.extract(r'(\d+)').astype(float)
```




```
/tmp/ipython-input-3956297429.py:7: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-
  shows['seasons'] = shows['duration'].str.extract(r'(\d+)').astype(float)
```




```
pairs = df[['director', 'cast']].dropna()
pairs = pairs.assign(cast=pairs['cast'].str.split(',')).explode('cast')
pairs['cast'] = pairs['cast'].str.strip()
pairs['director'] = pairs['director'].str.strip()

collab = pairs.groupby(['director','cast']).size().reset_index(name='count')
top_collab = collab.sort_values('count', ascending=False).head(10)
print(top_collab)
```

```
            director                cast  count
31530   Rajiv Chilaka         Rajesh Kava     17
31523   Rajiv Chilaka      Jigna Bhardwaj     17
31524   Rajiv Chilaka       Julie Tejwani     17
31531   Rajiv Chilaka        Rupa Bhimani     16
31534   Rajiv Chilaka        Vatsal Dubey     15
31533   Rajiv Chilaka             Swapnil     12
31527   Rajiv Chilaka              Mousam     12
38526     Suhas Kadav  Saurav Chakraborty      8
34562  S.S. Rajamouli             Prabhas      7
34561  S.S. Rajamouli              Nassar      7
```

```
genres = df['listed_in'].dropna().str.split(',').explode().str.strip()
genres.value_counts().head(10).plot(kind='bar', title="Top 10 Genres")
plt.show()
directors = df['director'].dropna().str.split(',').explode().str.strip()
directors.value_counts().head(10).plot(kind='bar', title="Top Directors")
plt.show()
actors = df['cast'].dropna().str.split(',').explode().str.strip()
actors.value_counts().head(10).plot(kind='bar', title="Top Actors")
plt.show()
descriptions = df['description'].dropna().str.lower().str.replace(r'[^a-z\s]', '', regex=True)
text = " ".join(descriptions.tolist())
wc = WordCloud(width=1000, height=500, background_color="white").generate(text)

plt.imshow(wc, interpolation="bilinear")
plt.axis("off")
plt.title("Word Cloud of Descriptions")
plt.show()
```