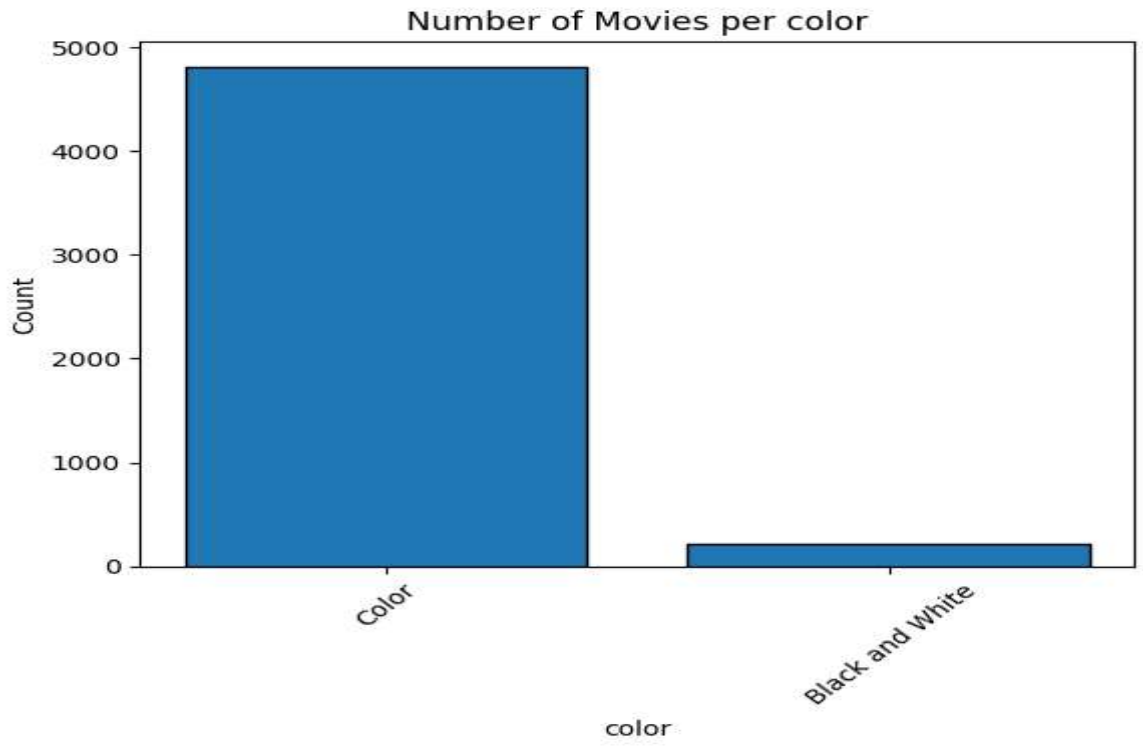
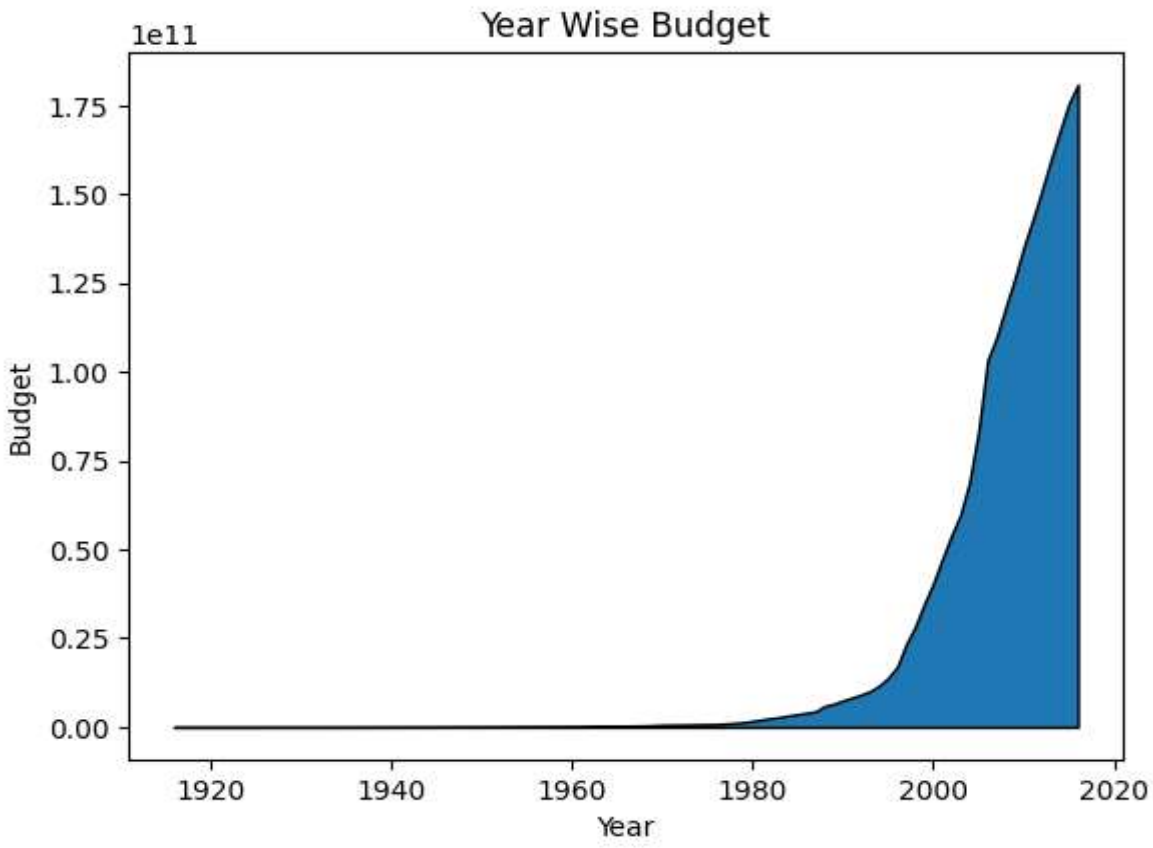
Assignment No 5 (Group No 5)

Ajay (207) , Snehal (206) , Sayali (214)

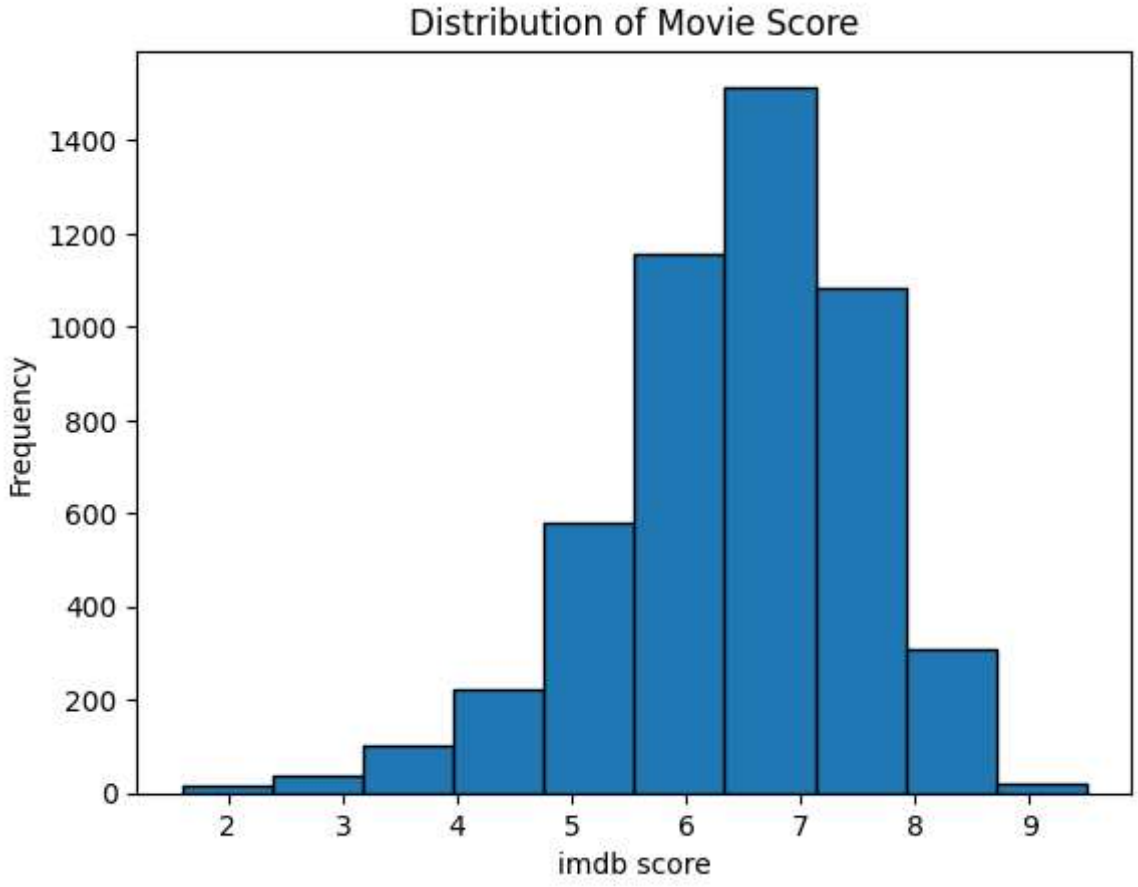
```
import matplotlib.pyplot as plt
import pandas as pd

# Read the CSV file
movies = pd.read_csv('MOVIES DATASET.csv')

# Assuming you have a DataFrame called 'movies' with a column 'color'
color_counts = movies['color'].value_counts()

plt.bar(color_counts.index, color_counts.values, edgecolor='black')
plt.xlabel('color')
plt.ylabel('Count')
plt.title('Number of Movies per color')
plt.xticks(rotation=45)
plt.show()
```



```
import matplotlib.pyplot as plt
import pandas as pd

movies = pd.read_csv('MOVIES DATASET.csv')

# Assuming you have a DataFrame called 'movies' with columns 'year' and
'budget'
cumulative_budget_by_year =
movies.groupby('title_year')['budget'].sum().cumsum()

plt.fill_between(cumulative_budget_by_year.index,
cumulative_budget_by_year.values, edgecolor='black')
plt.xlabel('Year')
plt.ylabel('Budget')
plt.title('Year Wise Budget')
plt.show()
```

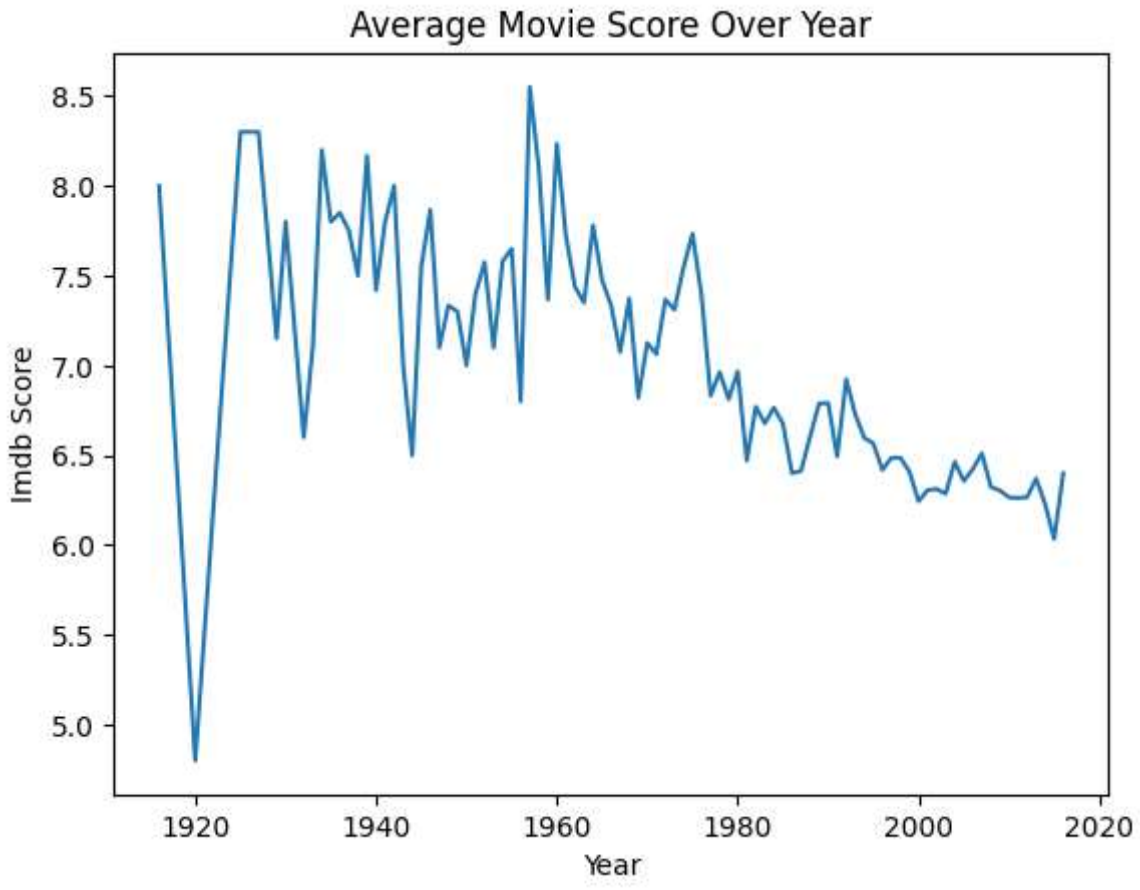


```
import matplotlib.pyplot as plt
import pandas as pd

movies = pd.read_csv('MOVIES DATASET.csv')

# Assuming you have a DataFrame called 'movies' with a column 'imdb_score'
plt.hist(movies['imdb_score'], bins=10, edgecolor='black')
plt.xlabel('imdb score')
plt.ylabel('Frequency')
plt.title('Distribution of Movie Score')
plt.show()
```



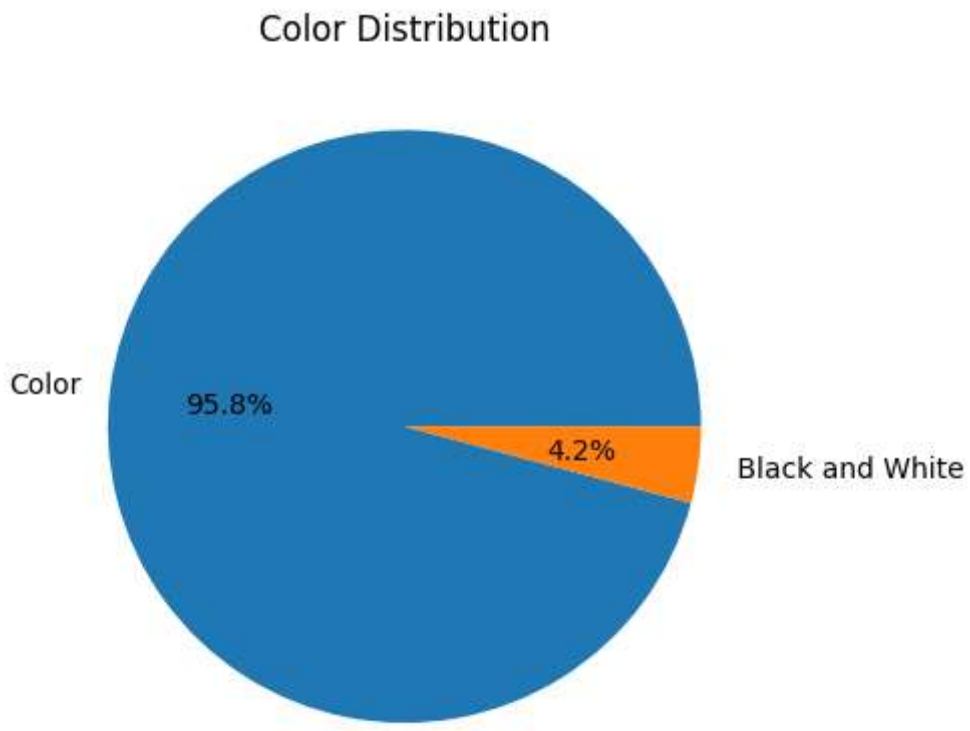

```
import matplotlib.pyplot as plt
import pandas as pd

movies = pd.read_csv('MOVIES DATASET.csv')
```

```
# Assuming you have a DataFrame called 'movies' with columns 'year' and
'imdb_score'
average_imdb_score_by_year =
movies.groupby('title_year')['imdb_score'].mean()

plt.plot(average_imdb_score_by_year.index,
average_imdb_score_by_year.values)
plt.xlabel('Year')
plt.ylabel('Imdb Score')
plt.title('Average Movie Score Over Year')
plt.show()
```



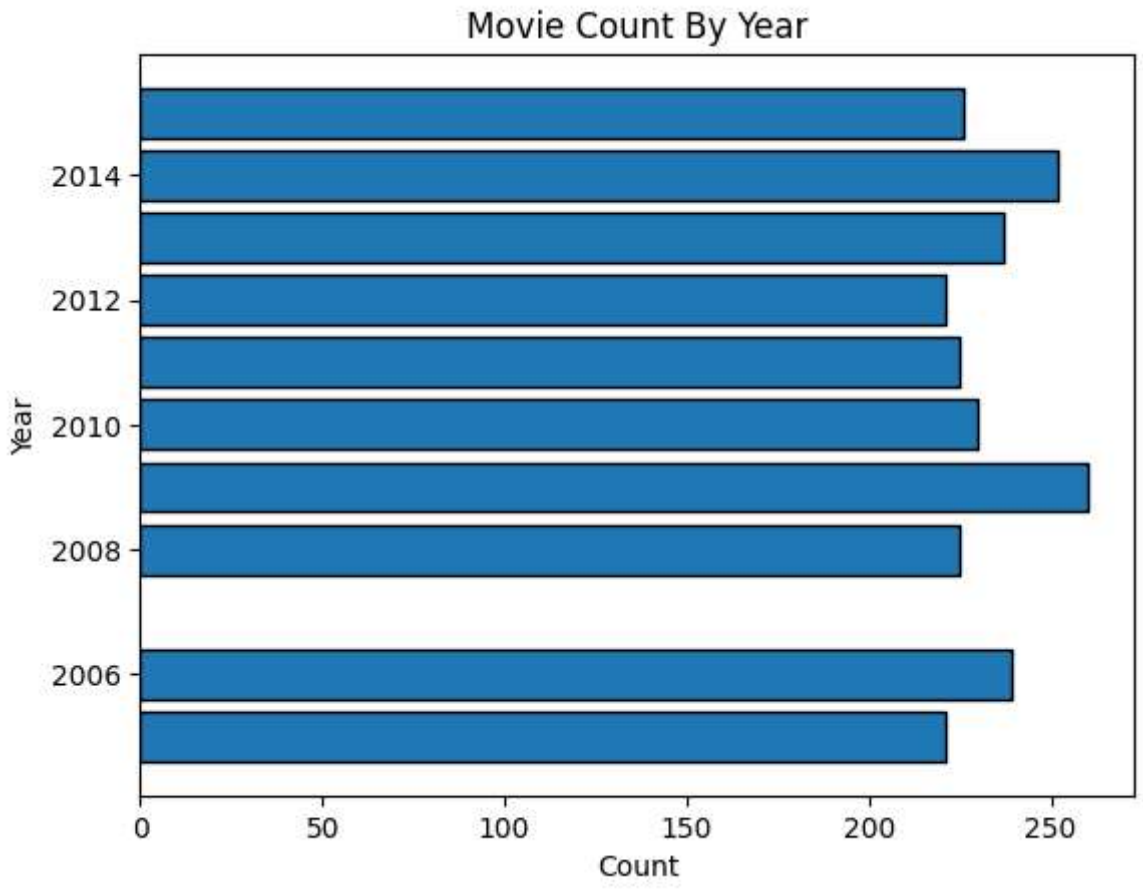

```
import matplotlib.pyplot as plt
import pandas as pd
```

```
movies = pd.read_csv('MOVIES DATASET.csv')

# Assuming you have a DataFrame called 'movies' with a column 'color'
color_counts = movies['color'].value_counts()

plt.pie(color_counts.values, labels=color_counts.index, autopct='%1.1f%%')
plt.title('Color Distribution')
plt.show()
```



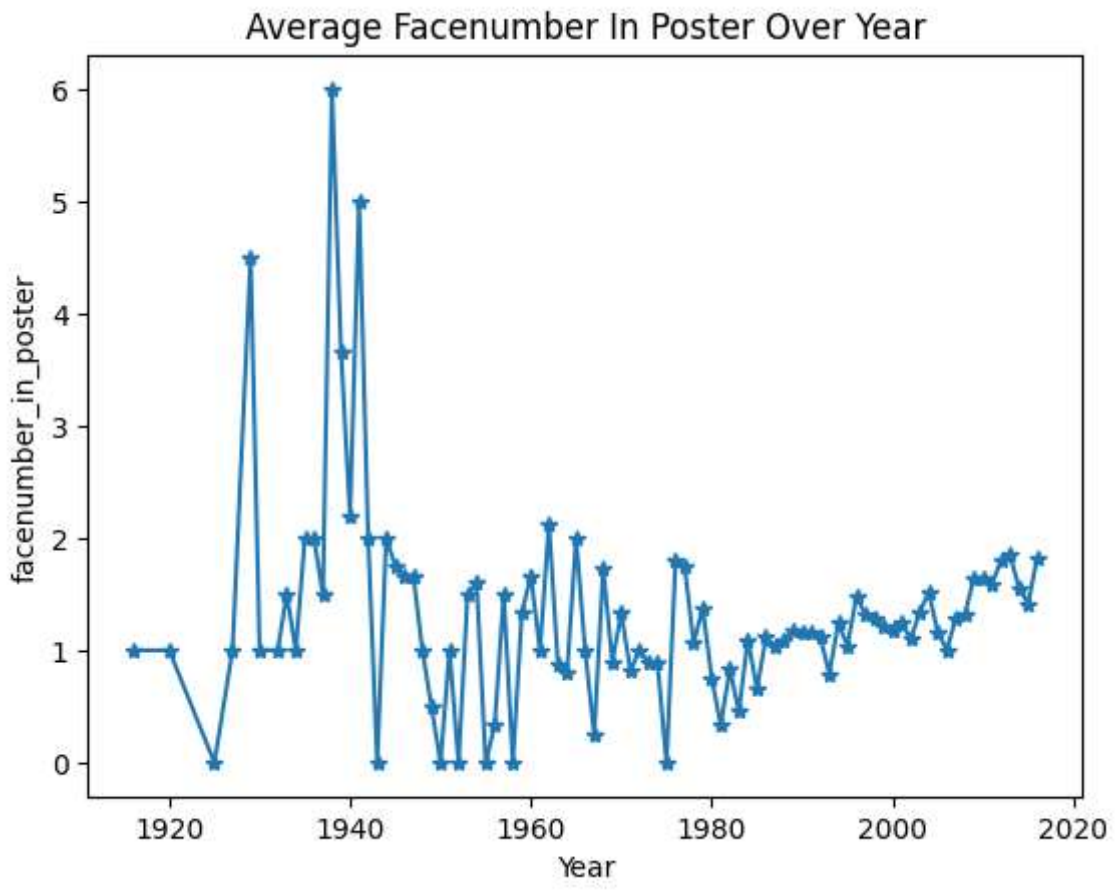

```
import pandas as pd
import matplotlib.pyplot as plt

# Read the CSV file
movies = pd.read_csv('MOVIES DATASET.csv')

# Calculate the count of movies per year
title_year_counts = movies['title_year'].value_counts().nlargest(10)
```

```
# Plot the horizontal bar chart
plt.barh(title_year_counts.index, title_year_counts.values,
edgecolor='black')
plt.xlabel('Count')
plt.ylabel('Year')
plt.title('Movie Count By Year')
plt.show()
```



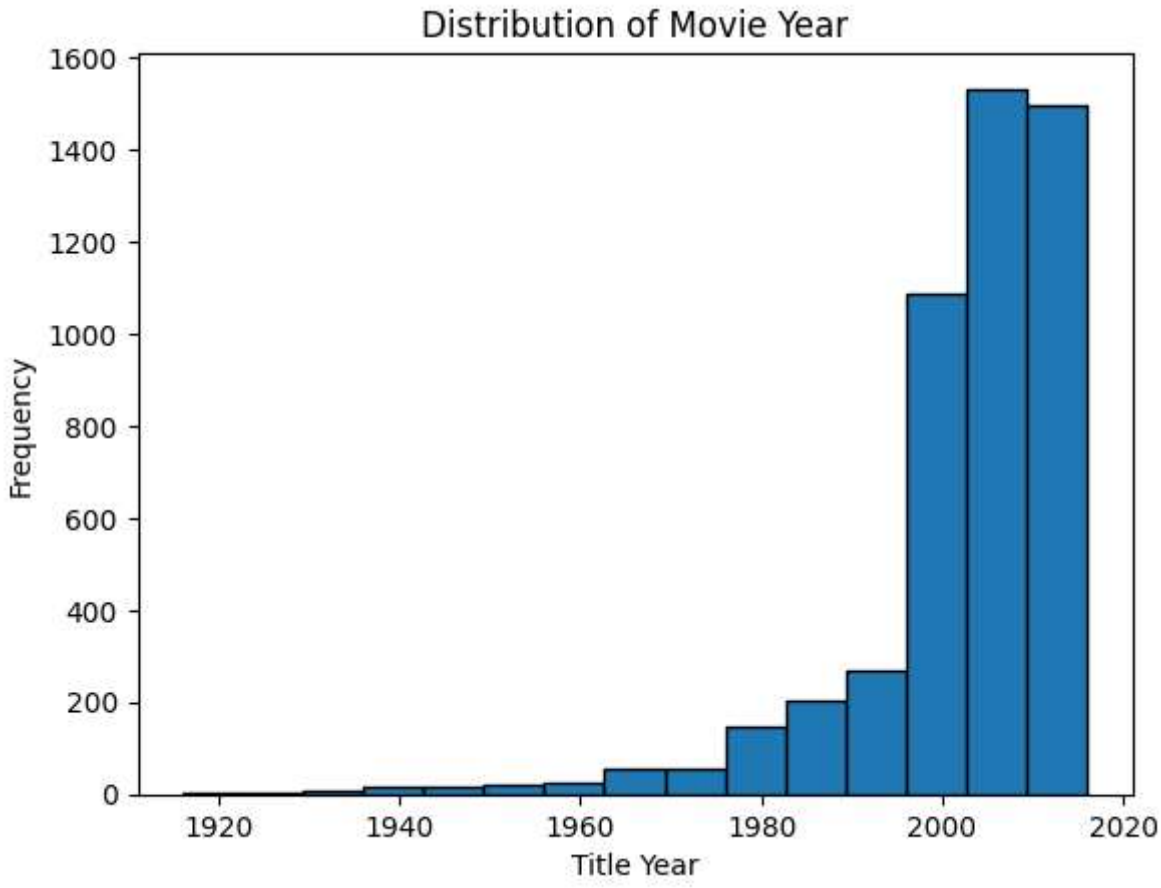

```
import pandas as pd
import matplotlib.pyplot as plt

# Read the CSV file
movies = pd.read_csv('MOVIES DATASET.csv')

# Convert the 'title_year' column to year format
movies['title_year'] = pd.to_datetime(movies['title_year'], format='%Y')
```

```
# Group by title year and calculate average facenumber_in_poster
average_facenumber_in_poster_by_year =
movies.groupby(movies['title_year'].dt.year)['facenumber_in_poster'].mean(
)

# Plot the line graph
plt.plot(average_facenumber_in_poster_by_year.index,
average_facenumber_in_poster_by_year.values,marker='*')
plt.xlabel('Year')
plt.ylabel('facenumber_in_poster')
plt.title('Average Facenumber In Poster Over Year')
plt.show()
```



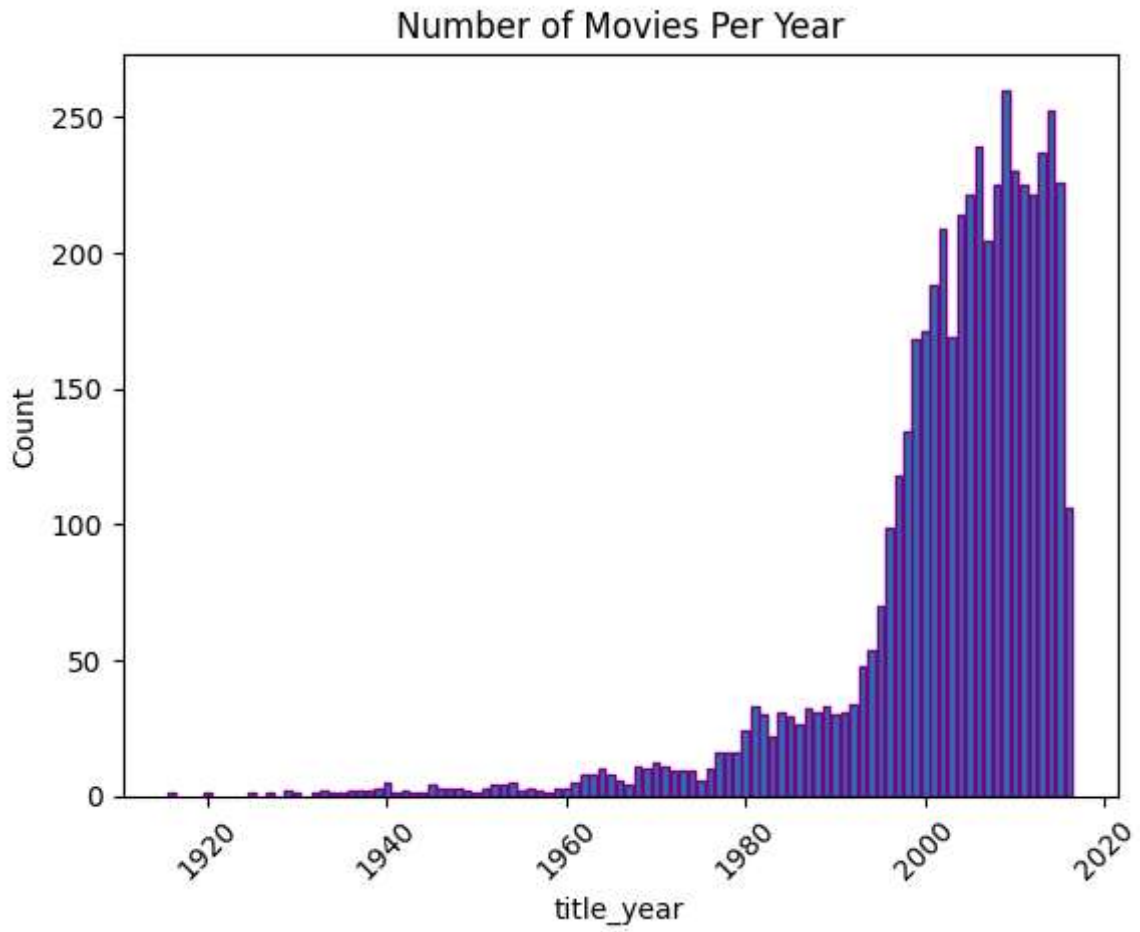

```
import pandas as pd
import matplotlib.pyplot as plt

# Read the CSV file
```

```
movies = pd.read_csv('MOVIES DATASET.csv')

# Extract the 'TITLE_YEAR' column for the histogram
ratings = movies['title_year']
# Plot the histogram
plt.hist(ratings,bins=15,edgecolor='black')
plt.xlabel('Title Year')
plt.ylabel('Frequency')
plt.title('Distribution of Movie Year')
plt.show()
```




```
import matplotlib.pyplot as plt
import pandas as pd

# Read the CSV file
movies = pd.read_csv('MOVIES DATASET.csv')
```

```
# Assuming you have a DataFrame called 'movies' with a column 'title_year'
title_year_counts = movies['title_year'].value_counts()
plt.bar(title_year_counts.index,
title_year_counts.values,edgecolor='purple')
plt.xlabel('title_year')
plt.ylabel('Count')
plt.title('Number of Movies Per Year')
plt.xticks(rotation=45)
plt.show()
```




```
import pandas as pd
import matplotlib.pyplot as plt

# Read the CSV file
movies = pd.read_csv('MOVIES DATASET.csv')
```

```python
# Calculate the count of movies by language
language_counts = movies['language'].value_counts().nlargest(10)

# Plot the horizontal bar chart
plt.barh(language_counts.index,language_counts.values, edgecolor='black')
plt.xlabel('Count')
plt.ylabel('Language')
plt.title('Movie Count By Languages')
plt.show()
```