The Practice Notebook - Part 1 - Distribution Analysis using Histograms

August 30, 2025

0.1 # Welcome To The Practice Notebook

Importing the pandas library

[4]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

Import the json Imdb Dataset

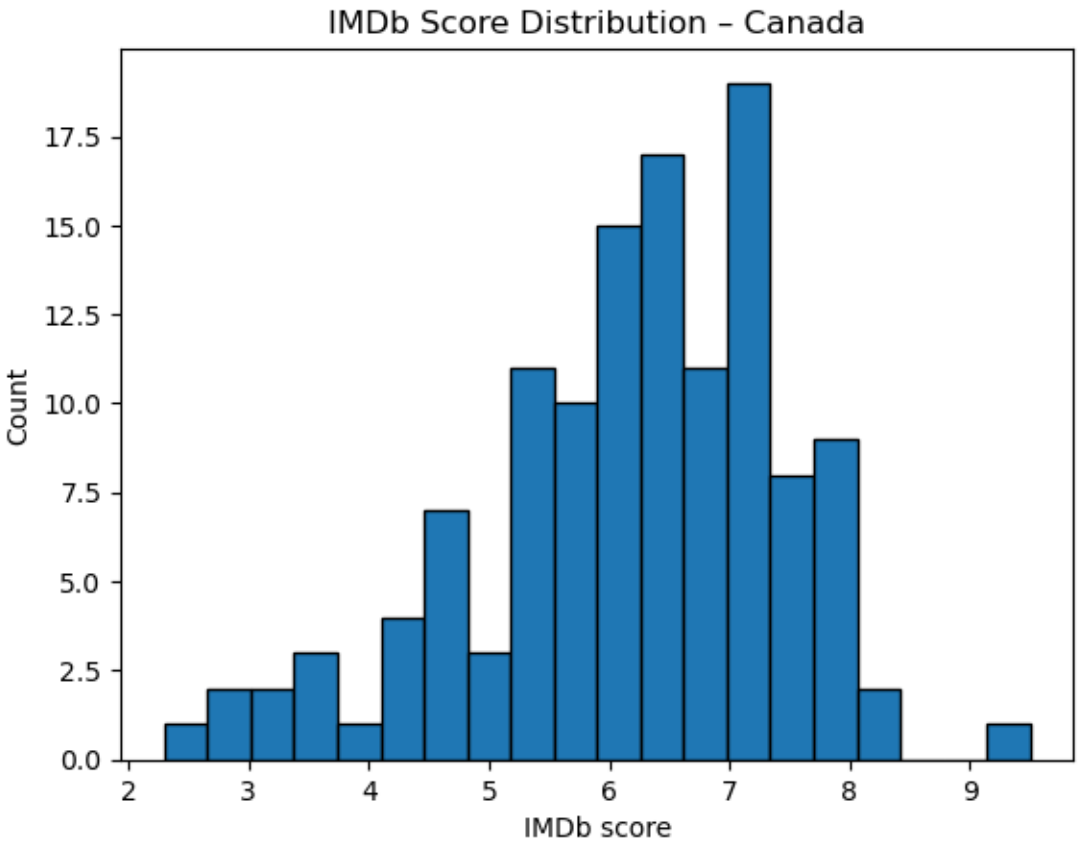
[7]:
```
df = pd.read_json('movies_dataset.json', lines=False)
df.columns = [c.strip().lower().replace(' ', '_') for c in df.columns]
```

Answer to the following Analytical questions:

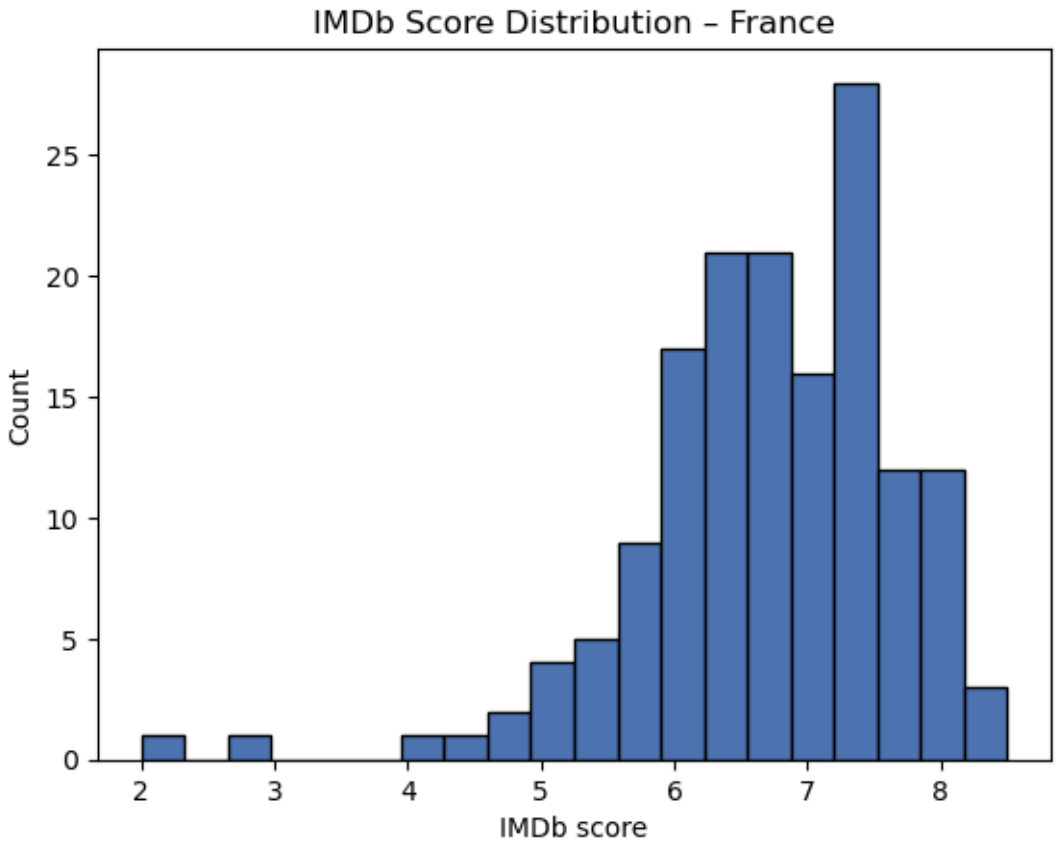
Question 1: Use Histograms to compare the disribution of movie `imdb_score` column related to movies produced in `France` and `Canada`.

While creating the histgoram: - Set the number of bins to 20 - Set the xAxis Label - Set the yAxis Label - Remove the grid lines

[8]:
```
# Create the histogram for Canada here
subset_canada = df[df['country'].str.contains('Canada', case=False, na=False)]
ax = subset_canada['imdb_score'].plot(kind='hist', bins=20, edgecolor='black')
ax.set_xlabel('IMDb score')
ax.set_ylabel('Count')
ax.grid(False)
plt.title('IMDb Score Distribution - Canada')
plt.show()
```



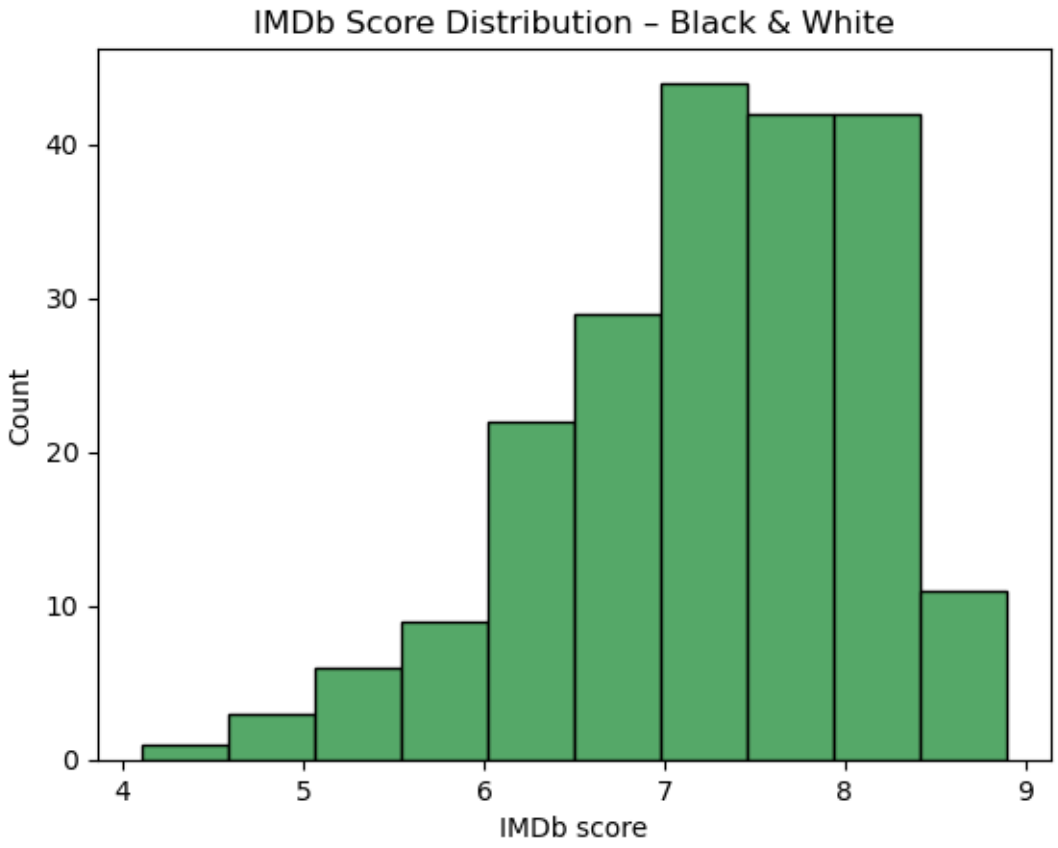
[9]:
```
# Create the histogram for France here
subset_france = df[df['country'].str.contains('France', case=False, na=False)]
ax = subset_france['imdb_score'].plot(kind='hist', bins=20, edgecolor='black',␣
 ↪color='#4C72B0')
ax.set_xlabel('IMDb score')
ax.set_ylabel('Count')
ax.grid(False)
plt.title('IMDb Score Distribution - France')
plt.show()
```



Question 2: Create a histogram to check the distribution of `imdb_score` spent to produce Black&White movies.

While creating the histgoram: - Set the number of bins to 10 - Set the xAxis Label - Set the yAxis Label - Remove the grid lines

[11]:
```
#To go back to the tree, click on the Jupyter icon
mask_bw = df['color'].str.contains('black', case=False, na=False)
subset_bw = df[mask_bw]
ax = subset_bw['imdb_score'].plot(kind='hist', bins=10, edgecolor='black',␣
 ↪color='#55A868')
ax.set_xlabel('IMDb score')
ax.set_ylabel('Count')
ax.grid(False)
plt.title('IMDb Score Distribution - Black & White')
plt.show()
```


IMDb Score Distribution – Black & White
Count
IMDb score
0
10
20
30
40
4
5
6
7
8
9

[ ]: