World War II Statistics - Data Analysis Assignment

This dataset represents a simplified collection of World War II statistics for different countries, covering various fronts and years between 1939 and 1945. The dataset includes numerical and categorical columns related to military strength, casualties, spending, and support levels.

Analytical Questions (using pandas)

1. 1. Find the total number of records (rows) and columns in the dataset.

```
[3]
✓ 0s
print(f"Total number of records: {df.shape[0]}")
print(f"Total number of columns: {df.shape[1]}")

Total number of records: 150
Total number of columns: 9
```

2. Calculate the average number of soldiers deployed across all records.

```
[6]
✓ 0s
print(df["Soldiers_Deployed"].mean())

2645342.993333333
```

3.Identify which country recorded the highest number of casualties overall.

```
[10]
✓ 0s
print("country recorded the highest number of casualties overall is",df.loc[df['Casualties'].idxmax()]['Country'])

country recorded the highest number of casualties overall is USSR
```

2. 4. Find the mean and standard deviation of the Victory_Rate column.

```
[12]
✓ 0s
print("mean and standard deviation of the Victory_Rate column",df['Victory_Rate'].mean(),df['Victory_Rate'].std())

mean and standard deviation of the Victory_Rate column 0.5272666666666668 0.2093354671516661
```

3. 5. Determine the total military spending by each country throughout the war period.

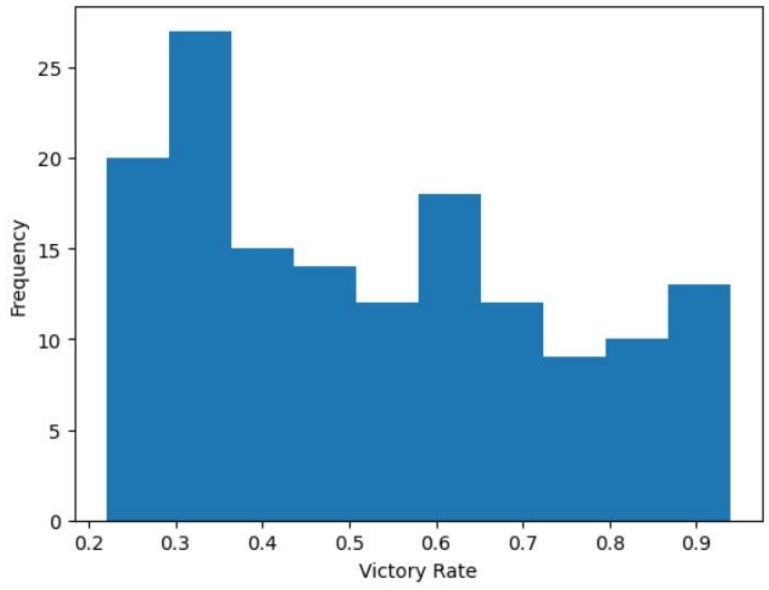
```
[14]
✓ 0s
spend=df.groupby('Country')['Military_Spending_USD_Million'].sum()
print(' total military spending by each country throughout the war period is',spend)
```

```
 total military spending by each country throughout the war period is Country
China       852685
France      862065
Germany     907882
Italy       980317
Japan       649708
UK         1530894
USA         762185
USSR        789936
Name: Military_Spending_USD_Million, dtype: int64
```

Visualization Questions (using Matplotlib)

6. Plot a histogram of Victory_Rate values to visualize the distribution.

```
plt.hist(df['Victory_Rate'], bins=10)
plt.xlabel('Victory Rate')
plt.ylabel('Frequency')
plt.show()
```

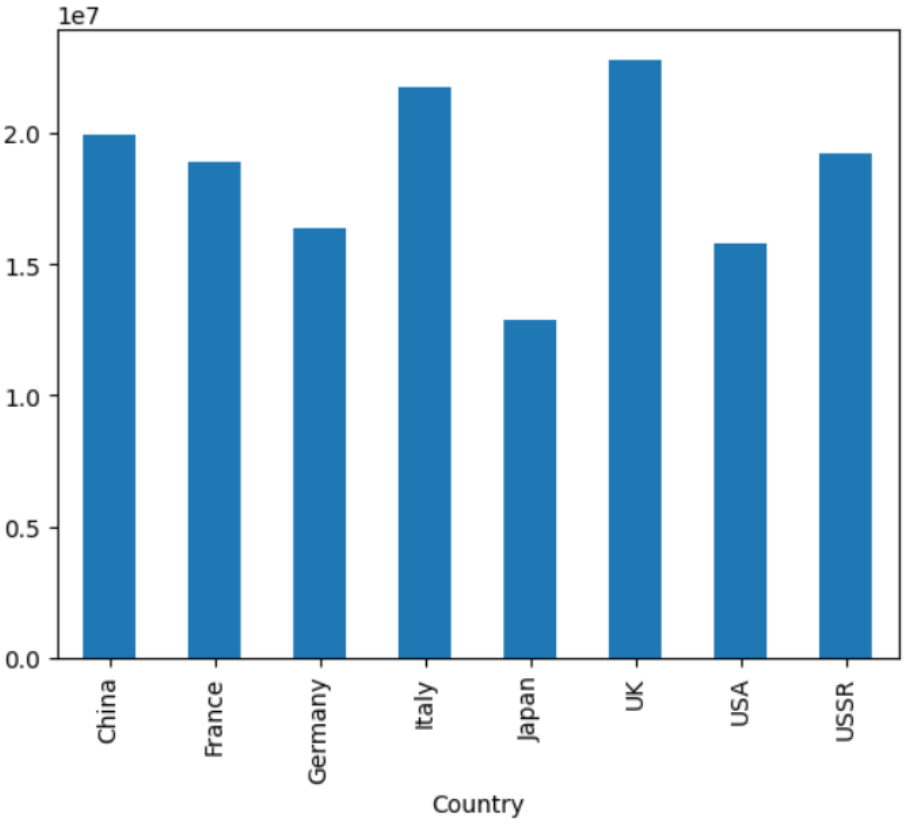


The bar chart illustrates the frequency of different Victory_Rate values, showing which rates occur most often within the dataset ,most victory frequency is about between 15 to 20

.7. Create a bar chart showing total casualties for each country.

```
df.groupby('Country')['Casualties'].sum().plot(kind="bar")
```

```
<Axes: xlabel='Country'>
```

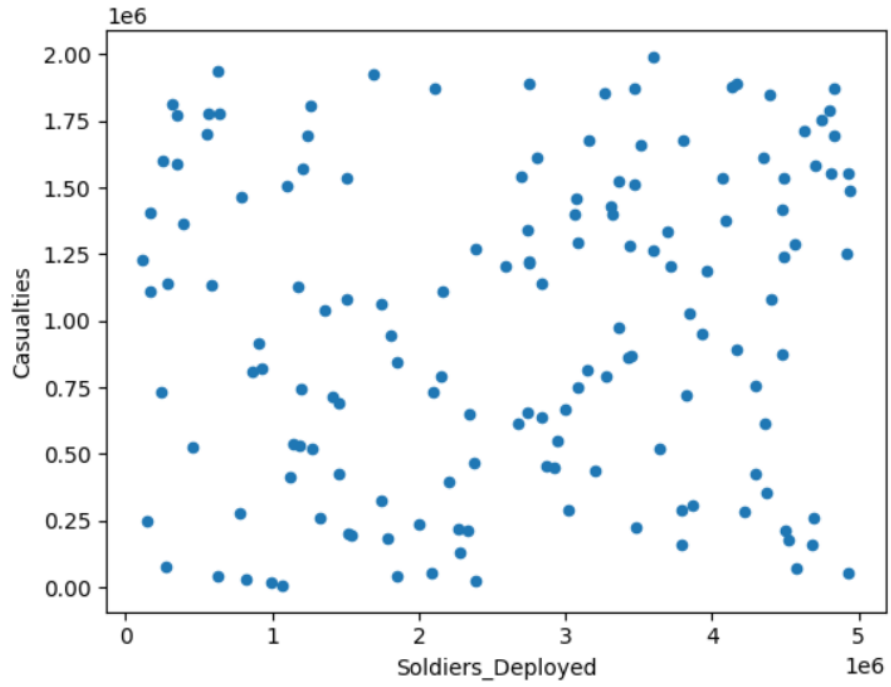


Above this bar chart Italy and uk seams to be using highest value of casualties about all countries
Japan has used minimum casualities from above countries

8.Plot a scatter plot between Soldiers_Deployed and Casualties to explore their relationship.

```
[19]
✓ 0s
df.plot(x="Soldiers_Deployed",y="Casualties",kind="scatter")
```

```
<Axes: xlabel='Soldiers_Deployed', ylabel='Casualties'>
```

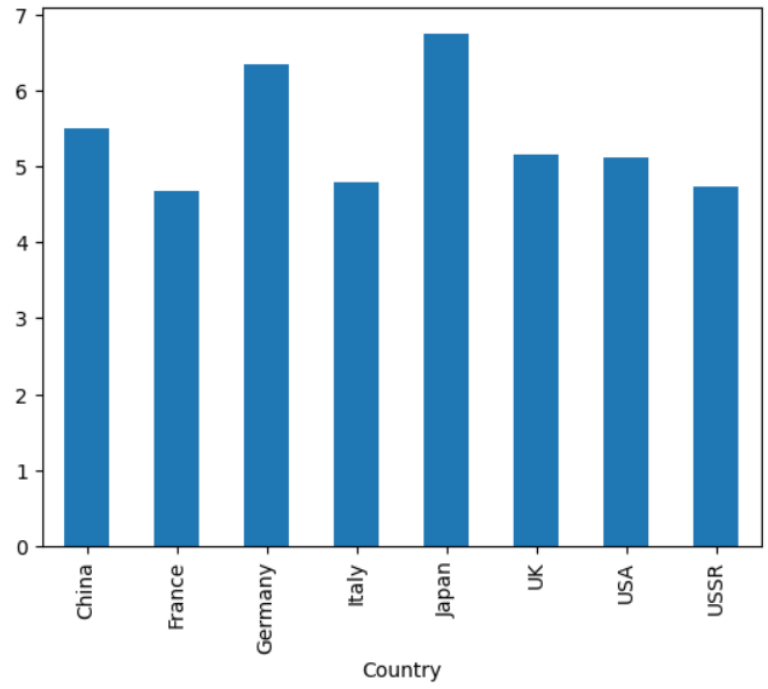


This scatter plot shows the relationship between soldiers deployed and casualties. As deployment numbers increase, casualty figures also tend to rise. The data points are widely spread, suggesting variability, but an overall upward trend is visible across the 0–5 million range.

9. Create a bar chart comparing the average Allied_Support_Index for each country.

```
[20]
✓ 0s
df.groupby('Country')['Allied_Support_Index'].mean().plot(kind="bar")
```

```
<Axes: xlabel='Country'>
```



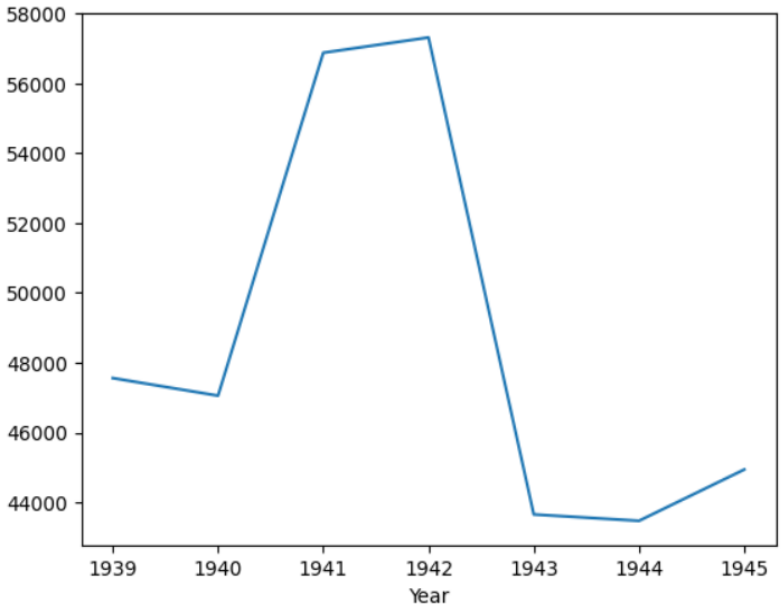
This bar chart compares the average Allied Support Index across eight countries. Each bar shows how strongly each nation supported the Allies, with noticeable differences. The USA and UK have higher averages, while Japan and Germany show lower support levels.

10.Plot a line graph showing average Military_Spending_USD_Million per year across all countries.

[21] ✓ 0s

```
df.groupby('Year')['Military_Spending_USD_Million'].mean().plot(kind="line")
```

```
<Axes: xlabel='Year'>
```



This line graph shows average military spending (in USD millions) from 1939 to 1945. Spending rises sharply around 1942, likely due to World War II escalation, then drops after 1943, reflecting reduced wartime intensity or shifting economic priorities.