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.

```
[7]: rows,columns=data.shape

[8]: print(rows)
     150

[9]: print(columns)
     9
```

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

```
[6]: print(data["Soldiers_Deployed"].mean())

     2645342.993333333
```

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

```
[12]: data['Casualties']. max()

[12]: 1991546

[17]: data.loc[data['Casualties'].idxmax(), 'Country']

[17]: 'USSR'
```

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

```
[20]: #q4
      print(data['Victory_Rate'].mean())
      0.5272666666666668

[22]: print(data['Victory_Rate'].std())
      0.2093354671516661
```

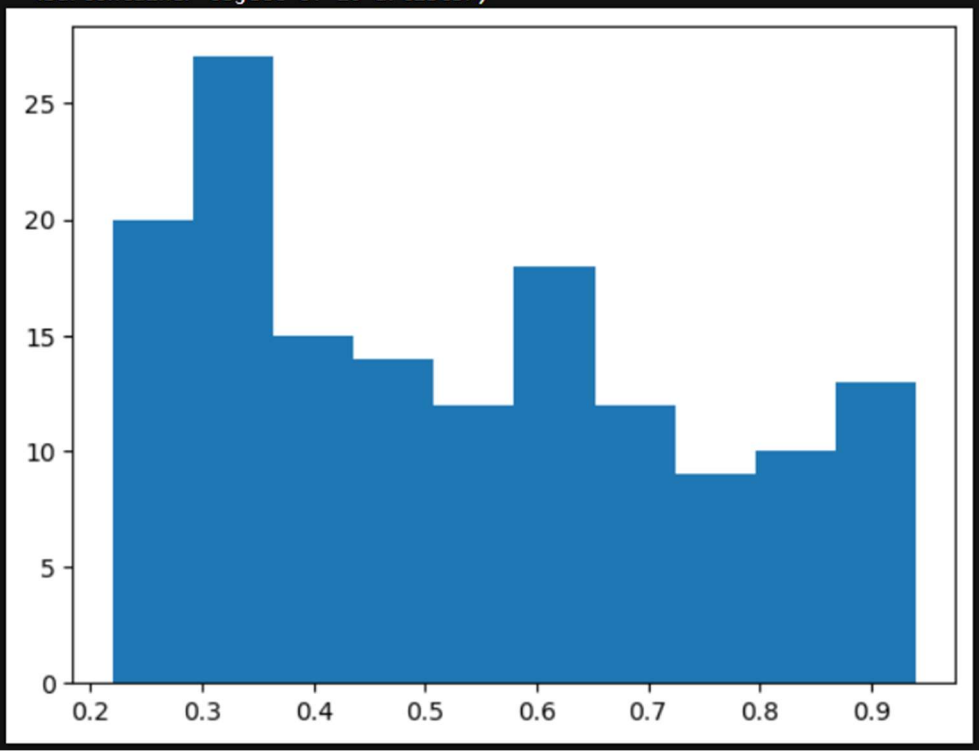
5. 5. Determine the total military spending by each country throughout the war period.

```
[23]: data.groupby('Country')['Military_Spending_USD_Million'].sum()

[23]: 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)

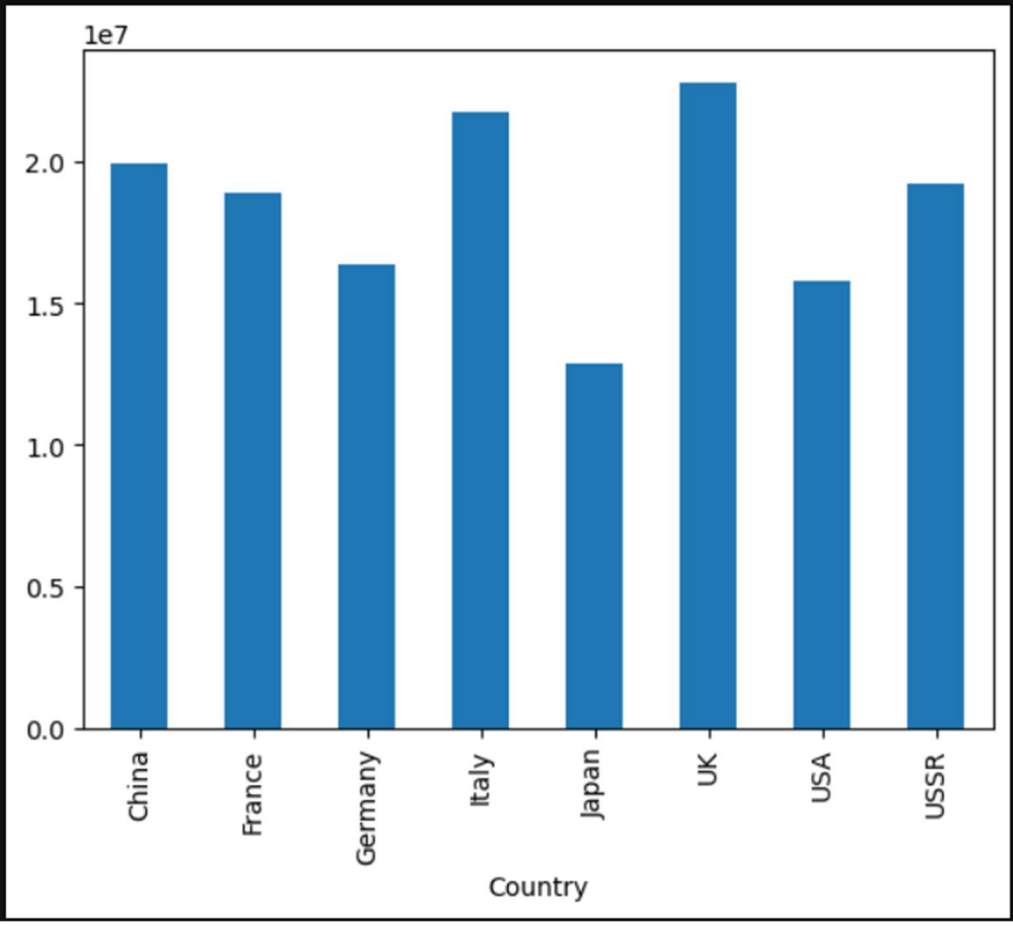
6. 6. Plot a histogram of Victory_Rate values to visualize the distribution.

```
[27]: plt.hist(data['Victory_Rate'],bins=10)

[27]: (array([20., 27., 15., 14., 12., 18., 12.,  9., 10., 13.]),
       array([0.22 , 0.292, 0.364, 0.436, 0.508, 0.58 , 0.652, 0.724, 0.796,
              0.868, 0.94 ]),
       <BarContainer object of 10 artists>)
```



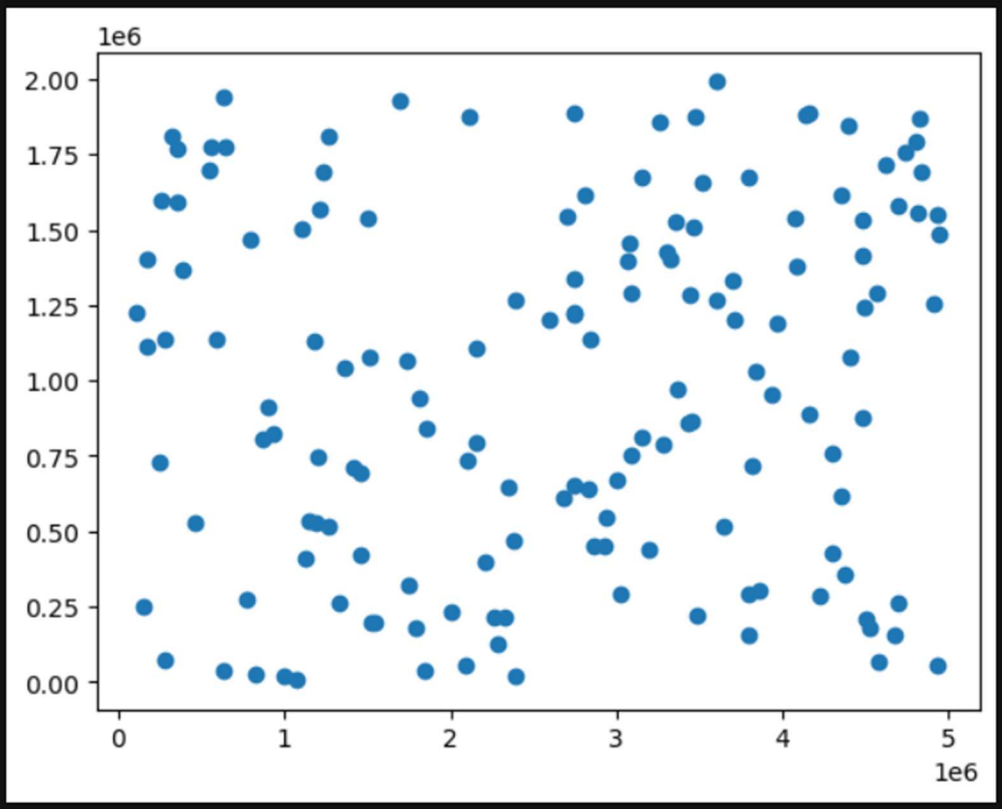
7. 7. Create a bar chart showing total casualties for each country.

```
[28]: Casualties = data.groupby('Country')['Casualties'].sum()
      Casualties.plot(kind='bar')

[28]: <Axes: xlabel='Country'>
```



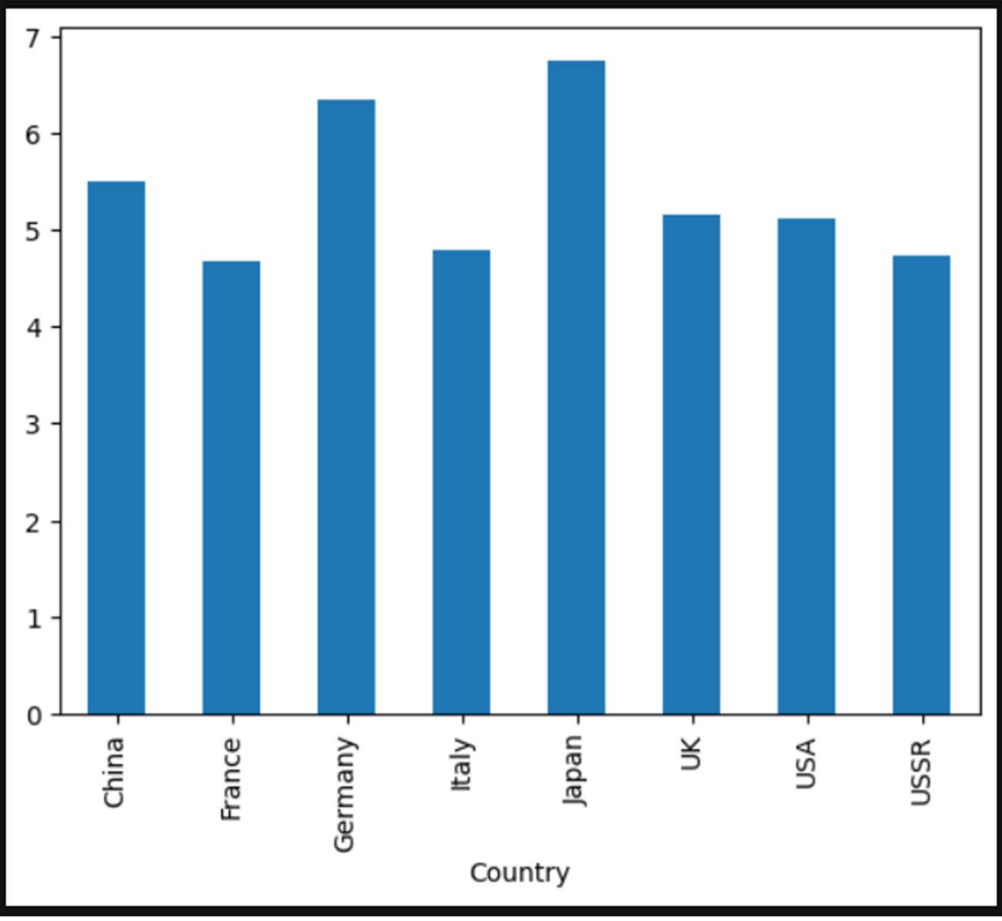
8. 8. Plot a scatter plot between Soldiers_Deployed and Casualties to explore their relationship.

```
[29]: plt.scatter(data["Soldiers_Deployed"], data["Casualties"])
      plt.show()
```



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

[30]:
```
data.groupby("Country")["Allied_Support_Index"].mean().plot(kind="bar")
plt.show()
```



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

```
[31]: data.groupby("Year")["Military_Spending_USD_Million"].mean().plot(kind="line")
      plt.show()
```