

Lab_4_Andrade

February 4, 2024

7 February 2024 #

Lab 4 Assignment - CS 4315

Doug Andrade

1. Load the armada.csv file into a Pandas data frame.

```
[1]: # Import the Pandas module for data frame operations
import pandas as pd

# Set the csv file name to a object, to avoid hard-coding
csv_file = 'armada.csv'

# Read-in the csv file as a Pandas data frame, as an object to be operated on
↳ later
armada_df = pd.read_csv(filepath_or_buffer = csv_file)

# Preview the top 5 rows of the data frame
armada_df.head()
```

```
[1]:
```

	Battle	Year	Portuguese Ships	Dutch Ships	English Ships	\
0	Bantam	1601	6	3	0	
1	Malacca Strait	1606	14	11	0	
2	Ilha das Naus	1606	6	9	0	
3	Pulo Butum	1606	7	9	0	
4	Surrat	1615	6	0	4	

	Ratio of Portuguese Ships to Dutch/British Ships	Spanish Involvement	\
0	2.000	No	
1	1.273	No	
2	0.667	No	
3	0.778	No	
4	1.500	No	

	Portuguese Outcome
0	Draw
1	Draw
2	Defeat

```

3          Victory
4          Draw

```

2. Return a data frame that contains the rows associated with a Portuguese victory.

```

[2]: # Set the column name and filter category
col_name = 'Portuguese Outcome'
col_val_filter = 'Victory'

# Select the applicable rows (Portuguese victories) by using the series
↳ indexing (.loc for the row value)
armada_df.loc[armada_df[col_name] == col_val_filter]

```

```

[2]:      Battle  Year  Portuguese Ships  Dutch Ships  English Ships  \
3    Pulo Butum  1606                7            9              0
12      Goa     1638                6            8              0
13   Colombo   1654                5            3              0
17    Bahia    1625               35           20              0
26    Recife   1653               14            5              0

      Ratio of Portuguese Ships to Dutch/British Ships  Spanish Involvement  \
3                                                    0.778                  No
12                                                    0.750                  No
13                                                    1.667                  No
17                                                    1.750                  Yes
26                                                    2.800                  Yes

      Portuguese Outcome
3          Victory
12         Victory
13         Victory
17         Victory
26         Victory

```

3. Return the average number of Portuguese ships present in the victories.

```

[3]: # Set the column name and filter category
col_name = 'Portuguese Outcome'
col_val_filter = 'Victory'

# Add an index for the "Portuguese Ships" column with the .mean() function to
↳ calculate the average
PT_vict_ships_avg = armada_df.loc[armada_df[col_name] ==
↳ col_val_filter]['Portuguese Ships'].mean()

# Extract the length of the total ships present in Portuguese victories

```

```
PT_num_vict = len(armada_df.loc[armada_df[col_name] ==
    ↪col_val_filter]['Portuguese Ships'])

# Print the average number of Portuguese ships present in victories using
    ↪string formatting
print('Portugal had an average of %.2f ships in their %d recorded victories.' %
    ↪(PT_vict_ships_avg, PT_num_vict))
```

Portugal had an average of 13.40 ships in their 5 recorded victories.

4. Create a bar plot with title and axis labels that visualizes the number of Portuguese ships in each battle (each battle is identified by the location and the year), whether a victory or not.

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

# Apply R's ggplot grid and background styling
plt.style.use('ggplot')

# Group by Battle and Portuguese Outcome, then sum the Portuguese Ships for
    ↪each group
grouped_df = armada_df.groupby(['Battle', 'Portuguese Outcome'])['Portuguese
    ↪Ships'].sum().unstack()

# Create a stacked bar plot to visualize the portion portugese ships that
    ↪contributed to a victory, defeat, or draw
fig, ax = plt.subplots(figsize = (12, 6))
grouped_df.plot(kind = 'bar', stacked = True, ax = ax)

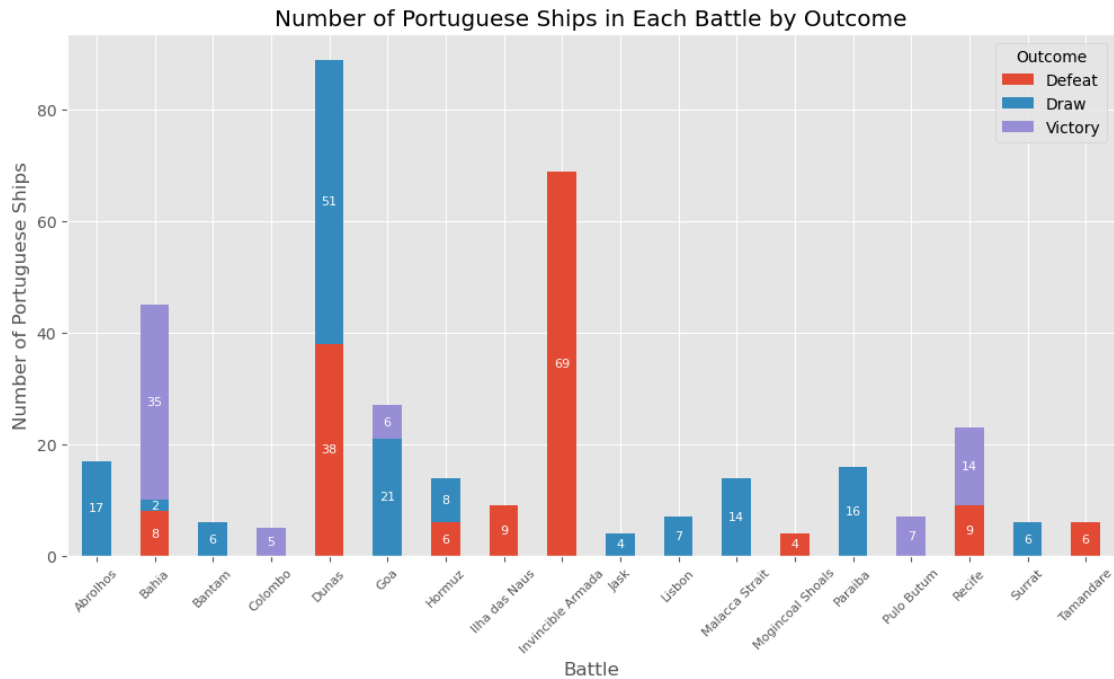
# Color ratio of bars with the number of ships for each battle outcome
for p in ax.patches:
    width = p.get_width()
    height = p.get_height()
    x, y = p.get_xy()
    if height > 0: # Add ships count only for non-zero "heights"
        ax.annotate(f'{height:.0f}', (x + width / 2, y + height / 2),
            ha = 'center', va = 'center', fontsize = 8, color = 'white')

# Customize the title, tick labeling, and legend
plt.title('Number of Portuguese Ships in Each Battle by Outcome')
plt.xlabel('Battle')
plt.xticks(rotation = 45, size = 8)
plt.ylabel('Number of Portuguese Ships')
plt.legend(title = 'Outcome')

# Remove the right and top spines for visualization
```

```
plt.gca().spines['right'].set_visible(False)
plt.gca().spines['top'].set_visible(False)

plt.show()
```



5. Return a data frame that contains the rows where the Portuguese outnumbered their opponent.

```
[5]: # Select the applicable rows (Ratio > 1) by using the series indexing (.loc for ↪
      ↪the row value)
      armada_df.loc[armada_df['Ratio of Portuguese Ships to Dutch/British Ships'] > 1.
      ↪000]
```

```
[5]:
```

	Battle	Year	Portuguese Ships	Dutch Ships	English Ships	\
0	Bantam	1601	6	3	0	
1	Malacca Strait	1606	14	11	0	
4	Surrat	1615	6	0	4	
7	Hormuz	1622	6	0	5	
10	Goa	1636	6	4	0	
13	Colombo	1654	5	3	0	
15	Invincible Armada	1588	69	0	31	
17	Bahia	1625	35	20	0	
20	Abrolhos	1631	17	16	0	
22	Dunas	1639	51	11	0	
26	Recife	1653	14	5	0	

	Ratio of Portuguese Ships to Dutch/British Ships	Spanish Involvement \
0	2.000	No
1	1.273	No
4	1.500	No
7	1.200	No
10	1.500	No
13	1.667	No
15	2.226	Yes
17	1.750	Yes
20	1.063	Yes
22	4.636	Yes
26	2.800	Yes

	Portuguese Outcome
0	Draw
1	Draw
4	Draw
7	Defeat
10	Draw
13	Victory
15	Defeat
17	Victory
20	Draw
22	Draw
26	Victory