

May 4 , 2022

DS Unit 1 Build Week Portfolio Project

Is Piracy On The Rise?

Number of Pirate Attacks Worldwide 1993-2020



Introduction

Pirate attacks. Although the term “pirate” may conjure up images of bearded men with eye patches, wooden legs and parrots who were convicted and buried centuries ago, pirate attacks are indeed posing a threat to today’s shipping lines all over the world.

Is the rise real?

As an advent traveler amongst the seas, I curiously ponder whether Pirates and their attacks on modern vessels are increasing in frequency. To put my mind at ease I will formulate a hypothesis that will shed light on this presumed plight to settle the score.

Hypothesis

H_0 : There **IS** a rise in piracy around the world from 1993 to 2020

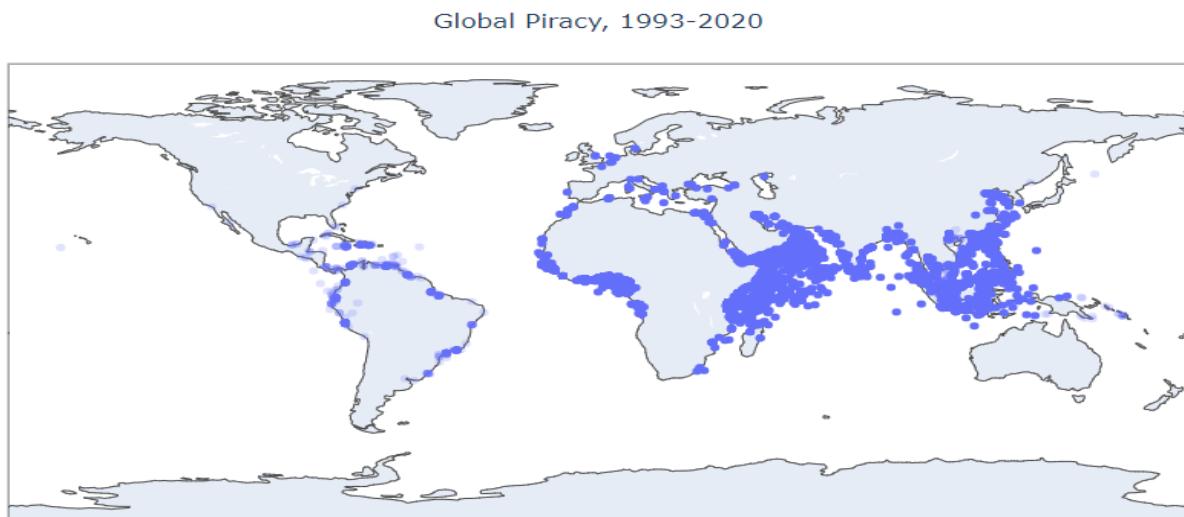
H_a : There **IS NOT** a rise in piracy around the world from 1993 to 2020

Process

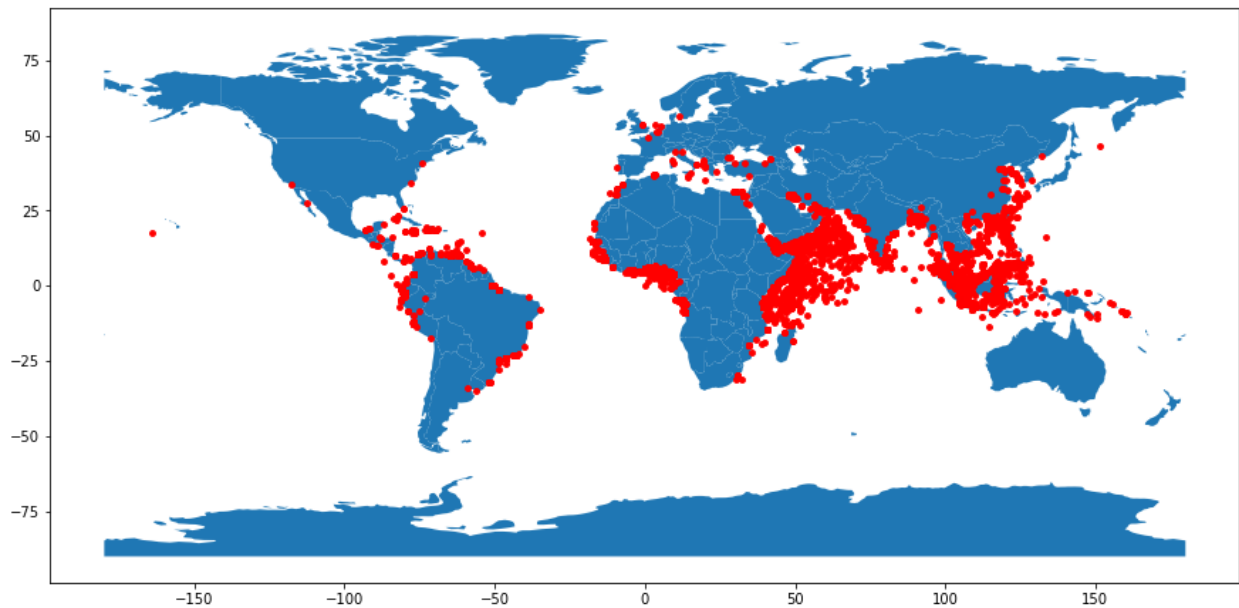
First, I'm going to load my dataset and import the required packages. The dataset that I chose is, Crime at Sea: A Global Database of Maritime Pirate Attacks (1993 - 2020). This dataset includes data on over 7,500 maritime pirate incidents that occurred between January 1993 and December 2020, as well as nation indicator data for the same period. The data on pirate attacks was gathered from the International Maritime Bureau (IMB), tidied, and geospatially enhanced. The data for the country indicators came from a variety of places, mainly The World Bank.

I'm going to import pandas, geopandas, datetime, shapely and matplotlib. I'm going to view and clean my DataFrame, dropping unnecessary columns. I will change the date column into DateTime. I will fill NaN values with zero using fillna(0). I will reset my index and make my changes True. I've created three plots for data visualization using Python, and one plot from Tableau.

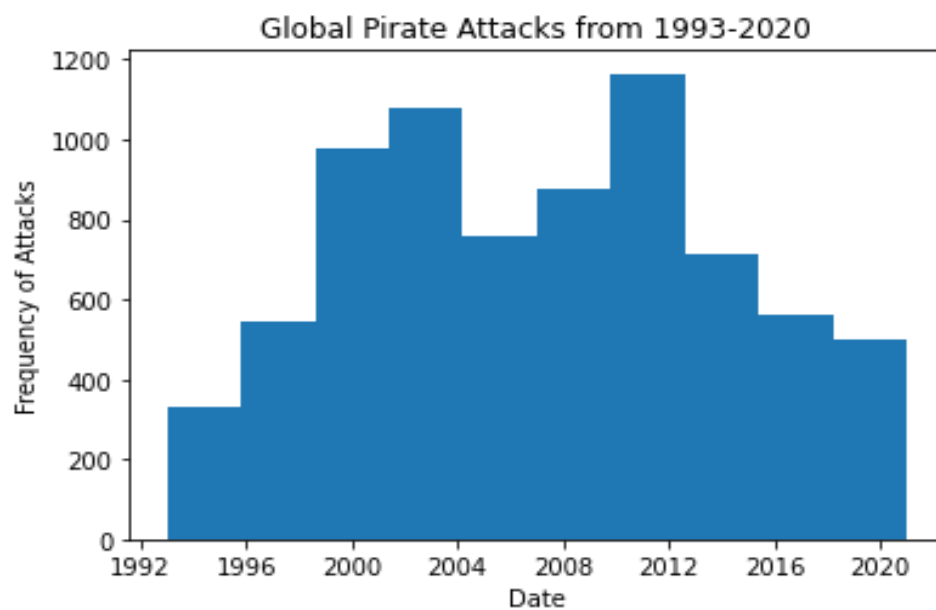
1. An interactive zoomable scatter plot using violet and gray:



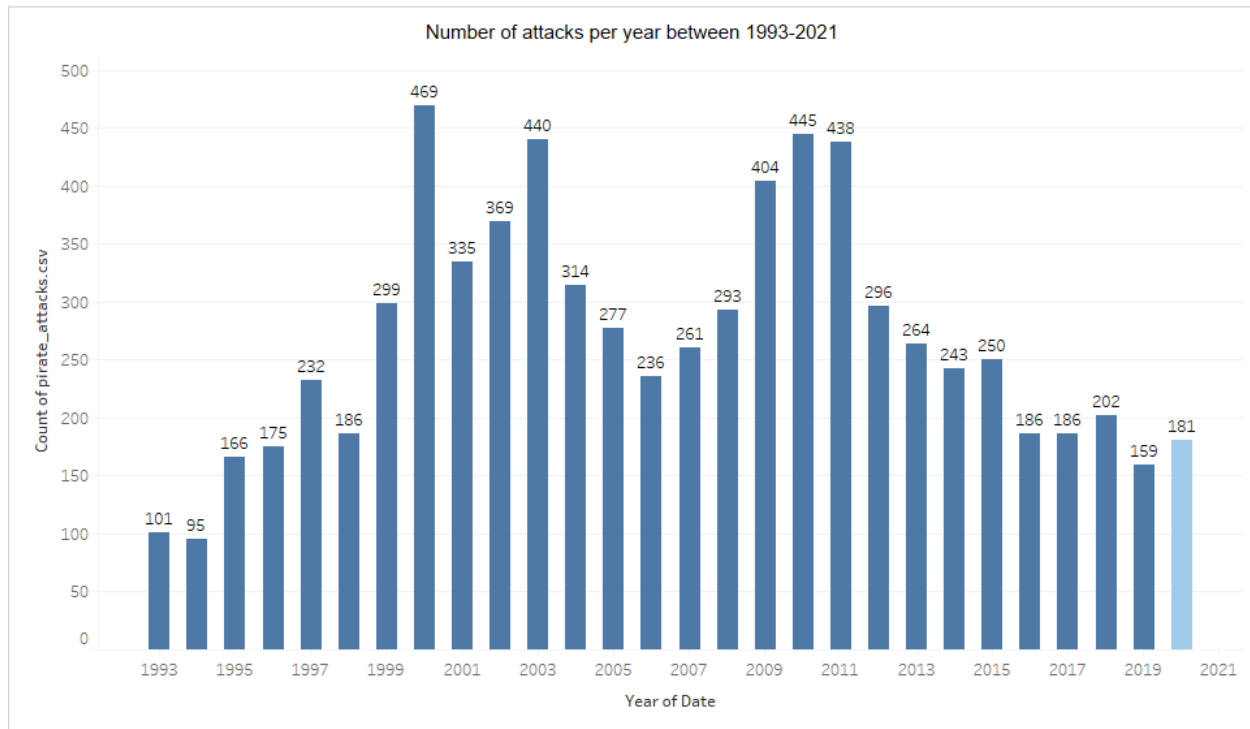
2. A geometry plot with the longitude and latitude geospatially combined on a world map using red and blue:



3. A histogram showing the distribution of attacks over time:



4. A bar chart showing the number of attacks per year over time with predictive result



Conclusion

Piracy on the high seas peaked in 2010, with approximately 445 instances reported. Indonesia, Malaysia, and Nigeria are among the countries most likely to be targeted by pirates. Pirates are drawn to these countries because of the richness of natural resources within them or nearby. Maritime crime has made strategic oil transport channels a renowned target. So it seems It is often the crew and the pirates themselves who pay most dearly for maritime crime. As more piracy reports were generated and handed over to law enforcement, there was a correlation in piracy arrests, this with a skew left descending trend, it is safe to say that we fail to reject the null hypothesis and conclude that piracy **IS NOT** on a rise between the years of 1993 to 2020. I lastly used a built-in feature of Tableau to predict the current year's pirate attack volume. The last page is the code read out for my process in Python.

```

#packages
pip install geopandas
import pandas as pd
import geopandas as gpd
from geopandas import GeoDataFrame
from datetime import datetime
from shapely import geometry
from shapely.geometry import Point
import matplotlib.pyplot as plt

#load data, tidy data, removed columns & NaN values, reset index.
pirate = pd.read_csv('/content/drive/MyDrive/pirate_attacks.csv')

pirate_list = list(pirate)
pirate = pirate.drop(['time', 'location_description',
                    'attack_type', 'nearest_country', 'eez_country', 'data_source',
                    'shore_distance', 'shore_longitude', 'shore_latitude',
                    'attack_description', 'vessel_name', 'vessel_type', 'vessel_status'],
                    axis=1)

pirate["date"] = pd.to_datetime(pirate["date"])

pirate.fillna(0, inplace=True)
pirate.reset_index(drop=True, inplace=True)

[146] import plotly.express as px

fig = px.scatter_geo(pirate, lat='latitude', lon='longitude', hover_name="date")
fig.update_layout(title = 'Global Piracy, 1993-2020', title_x=0.5)
fig.show()

[147] geometry = [Point(xy) for xy in zip(pirate["longitude"], pirate["latitude"])]
gdf = GeoDataFrame(pirate, geometry=geometry)

world = gpd.read_file(gpd.datasets.get_path('naturalearth_lowres'))
gdf.plot(ax=world.plot(figsize=(15, 15)), marker='o', color='red', markersize=15);

pirate = pirate.drop(['longitude', 'latitude'], axis=1)

[148] fig, ax = plt.subplots()

# Plot a histogram of Parasleep from the Sleep DataFrame
ax.hist(pirate['date'])

# Specify the axis labels and plot title
ax.set_xlabel('Date')
ax.set_ylabel('Frequency of Attacks')
ax.set_title('Global Pirate Attacks from 1993-2020')

plt.show()

```