Precentation me project

Sanctions Risk: Assessment Model for Data Analysis

Dataset Content

This Sanctions Risk project contains UK Sanctions, OFAC SDN, FATF datasets. These are datasets is specific for identification and compliance purposes, that contain identifying information about entities (individuals, organizations, ship/vessels and jurisdictions) that are subject to restrictions, asset freezes, or enhanced monitoring due to activities such as terrorism, proliferation, or money laundering. Most of this data is presented as text categories.
To understand what data datasets includes, the basic data described below:

Full Name / Entity Name - Textual - official name individuals or organizations.

Aliases / AKA (Also Known As) - Textual - all known alternative names or spellings.

Date & Place of Birth - Textual/Date - biographical information for identification.

Identification Numbers - Textual - passport #, national ID, reg. Number (ships/aircraft).

Address - Textual - address, country of residence, or registration location.

Sanctions Regime - Textual - rules under which sanctions are applied

Entity Type - Categorical - Individual, Entity, Ship/Vessel, Aircraft).

Jurisdiction (FATF Specific) - Categorical - Name of the country identified as black or grey

   list

Tonnage and Length of ship – Textual/Numerical - description of characteristics

Data download sources

(https://sanctionslist.ofac.treas.gov/Home/SdnList)

(https://www.gov.uk/government/publications/the-uk-sanctions-list)

FATF_list.xlsx was created from (https://www.fatf-gafi.org/en/countries/black-and-grey-lists.html)
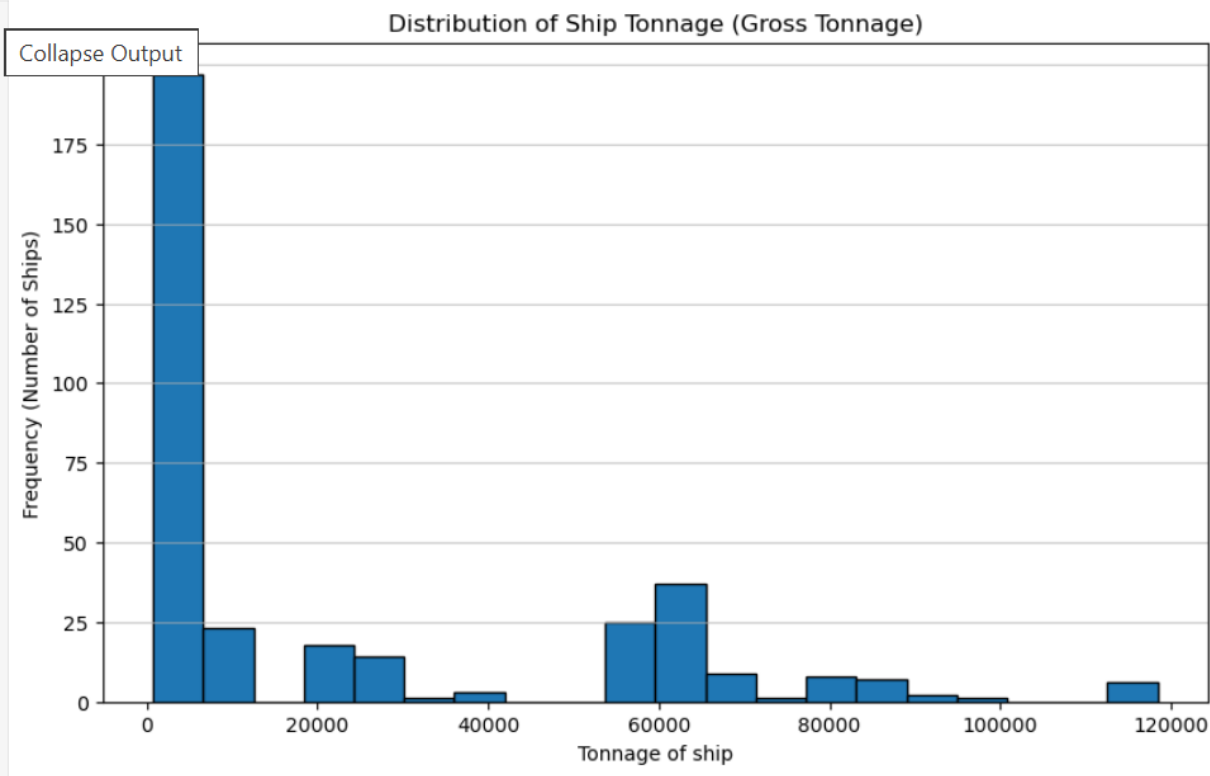
Business Requirements

   To understand which characteristics of data from the UK sanctions list, the OFAC SDN list are most effective for identifying and assessing risks, taking into account the

largest number of sanctions in a regional context, analyze the breakdown of sanctions by entity type and compare the OFAC and UK lists
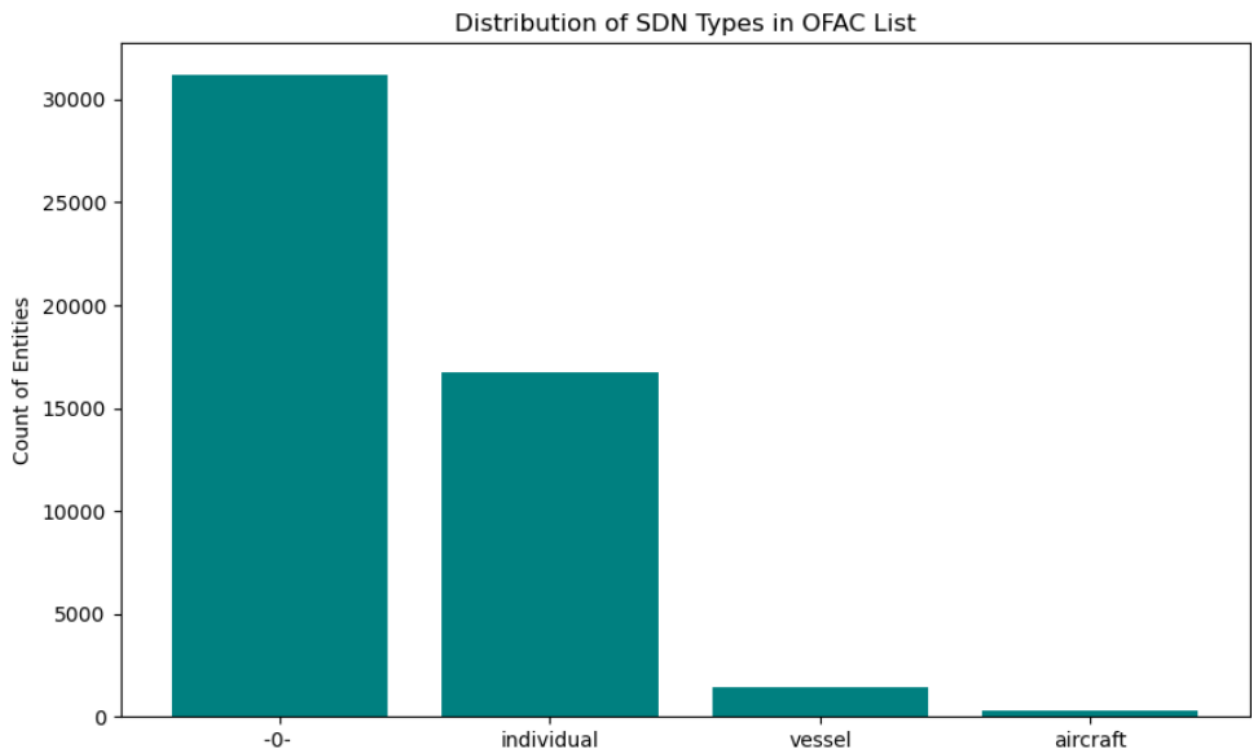
Hypotheses and Approaches

Distribution of Ship Tonnage (Gross Tonnage) to describe the characteristics of the ship in the sanctions list
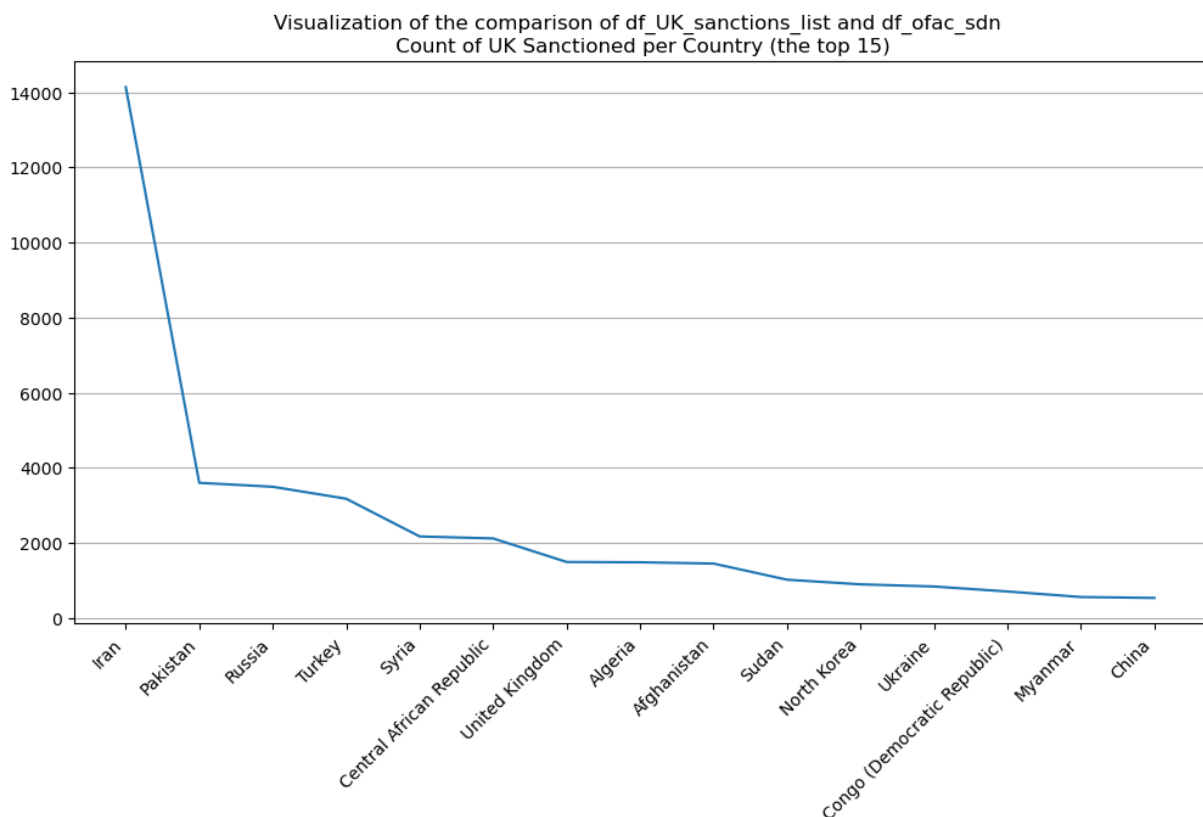

Pic 1: Distribution of Ship Tonnage (Gross Tonnage) (histogram)

The distribution of SDN types on the OFAC list shows the distribution of sanctions by entity type
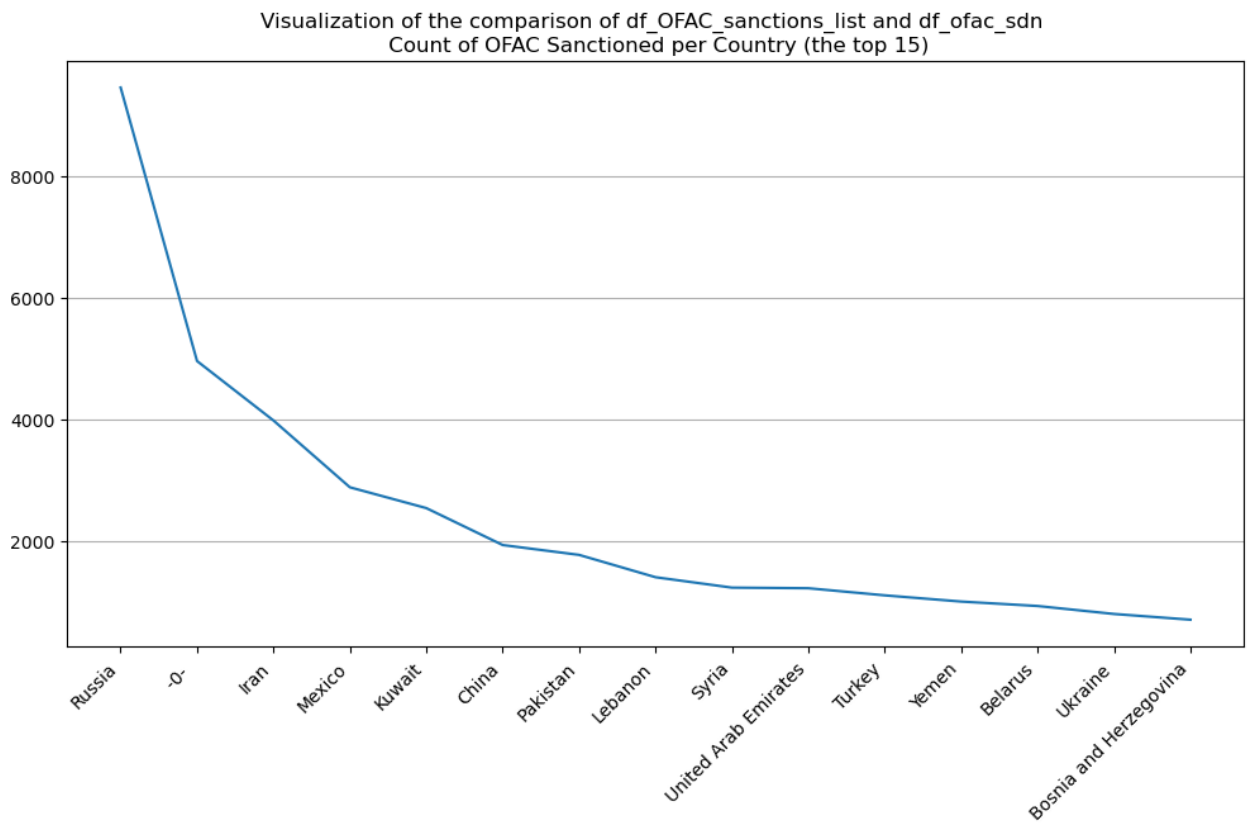
Pic 2: Distribution of SDN Types in OFAC List (bar chart)

Visualization of the comparison of the UK sanctions list and the OFAC VDN list shows the top 15 countries on whose entities sanctions have been imposed in terms of quantitative indicators
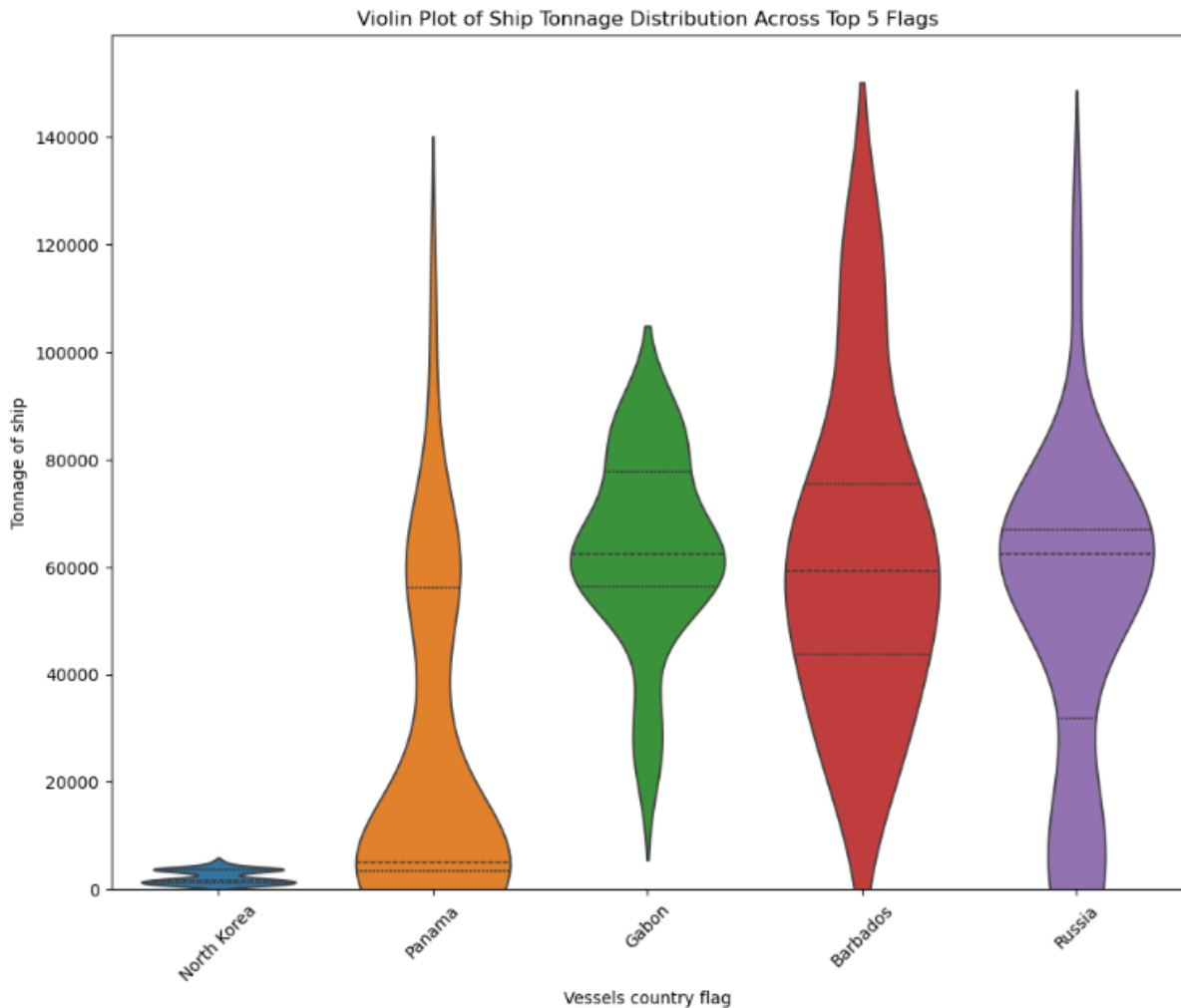

Pic 3.1: Visualization of the comparison of
df_UK_sanctions_list and df_ofac_sdn Count of UK Sanctioned per Country (the top 15) (line plot)

Visualization of the comparison of df_OFAC_sanctions_list and df_ofac_sdn
Count of OFAC Sanctioned per Country (the top 15)

pic 3.2: Visualization of the comparison of df_OFAC_sanctions_list and df_ofac_sdn \n Count of OFAC Sanctioned per Country (the top 15) (line plot)

H4: The distribution of tonnage by the 5 most common flags of countries provides an understanding of which vessels from which countries violate the sanctions regime

pic 4: Tonnage distribution by top 5 most common flags (Violin Plot)

Project Plan

Datasets for the project will be sourced directly from official primary sources to ensure accuracy and consistency. Key sources include the US Treasury Department's Office of Foreign Assets Control (OFAC), the UK Sanctions List via GOV.UK, and the Financial Action Task Force (FATF) website. The data will be processed and stored for future use. The data will be used to create visualizations that will provide insight into the relationships between variables and help answer hypotheses.

Rationale to map the business requirements to the Data Visualizations and ML tasks

Data visualization helped visualize and analyze sanctions lists by geographic region to understand which countries have the most sanctions imposed and reflect this in terms of quantitative indicators of imposed sanctions and reflect this on the example of sanction lists of ships; give quantitative characteristics in terms of entity type.

The data was started in Python via Jupyter Notebook.

Sanctions Risk: Assessment Model for Data Analysis

my first project

```
[1]:  # Import everything we need
      import pandas as pd
      import numpy as np
      import matplotlib.pyplot as plt
      import seaborn as sns
      import plotly.express as px
```

```
[2]:  # Load our data (for furure work in postgreSQL)
      df_sdn = pd.read_csv('sdn.csv')
      df_add = pd.read_csv('add.csv')
      df_alt = pd.read_csv('alt.csv')
```

```
[3]:  # Load data from postgreSQL (OFAC_SDN.sql, combine sdn.csv, add.csv, alt.csv)
      df_ofac_sdn = pd.read_csv('cons_data_ofac.csv')
```

```
[4]:  df_ofac_sdn.head(3)
```

[4]:

	ent_num	sdn_name	sdn_type	program_	title	call_sign	vess_type	tonnage	grt	vess_flag	...	add_num	address	city_state_province_postalcode	c
0	36	AEROCARIBBEAN AIRLINES	-0-	CUBA	-0-	-0-	-0-	-0-	-0-	-0-	...	25	-0-	Havana	
1	173	ANGLO-CARIBBEAN CO., LTD.	-0-	CUBA	-0-	-0-	-0-	-0-	-0-	-0-	...	129	Ibex House, The Minories	London EC3N 1DY	Ki
2	306	BANCO NACIONAL DE CUBA	-0-	CUBA	-0-	-0-	-0-	-0-	-0-	-0-	...	199	Zweierstrasse 35	Zurich CH-8022	Swit

3 rows × 23 columns

```
[5]:  # Load data UK_Sanctions_List
      # Columns (2, 50, 54, 55, 56, 57) have mixed types, so low_memory=False was set during import
      df_UK_sanctions_list=pd.read_csv('UK_sanctions_list.csv', low_memory=False)
```

```
[6]:  df_UK_sanctions_list.head(3)
```

[6]:

	[Report for: 07-Nov-2025]	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnamed: 6	Unnamed: 7	Unnamed: 8	Unnamed: 9	...	Unnamed: 48	Unnamed: 49	Unnamed: 5(
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN

The next step was to process the files and create consolidated data according to the OFAC lists in PostgreSQL.

The tables were created according to the described file dat_spec.txt. Files sdn.csv, add.csv, alt.csv were imported into the created tables and consolidated as cons_data_ofac.csv

Name	Modified	File Size
☐ ▪ 🔴 project.ipynb	1 hour ago	5.2 MB
☐ ⊞ add.csv	2 days ago	1.5 MB
☐ ⊞ alt.csv	2 days ago	1 MB
☐ 🖼 bar_chart.png	1 hour ago	22.5 KB
☐ 🖼 comb_plot.png	1 hour ago	108.7 KB
☐ ⊞ cons_data_ofac.csv	2 days ago	21 MB
☐ 🗋 dat_spec.txt	2 days ago	14.8 KB
☐ ⊞ FATF_list_converted.csv	1 hour ago	1.9 KB
☐ 🗋 FATF_list.xlsx	2 days ago	11.2 KB
☐ 🖼 Heatmap.png	1 hour ago	28 KB
☐ 🖼 histogram.png	1 hour ago	26.6 KB
☐ 🖼 line_plot_ofac.png	1 hour ago	55.9 KB
☐ 🖼 line_plot_uk.png	1 hour ago	59.5 KB
☐ 🔧 OFAC_cons_str.pdf	7 hours ago	80 KB
☐ 🗋 OFAC_SDN.sql	2 days ago	1.9 KB
☐ 🖼 pair_plot.png	1 hour ago	29.7 KB
☐ ⊞ sdn_comments.csv	2 days ago	32.1 KB
☐ ⊞ sdn.csv	2 days ago	5.1 MB
☐ ⊞ UK_Sanctions_List.csv	3 days ago	44.3 MB
☐ 🖼 violin_plot.png	1 hour ago	82.1 KB

and also added an illustration file OFAC_cons_str.pdf.

All further data processing was carried out in Python in a Jupyter notebook. File OFAC_SDN.sql has been added. The FATF data file was downloaded from an Excel spreadsheet previously created from lists in official sources.

```python
[11]:  # Load data from Excel
       df_fatf_xlsx = pd.read_excel("FATF_list.xlsx", sheet_name="Sheet1")

[12]:  df_fatf_xlsx.head(3)
```

[12]:	List type	Country	list name description
0	NaN	NaN	NaN
1	black list	Democratic People's Republic of Korea	High-Risk Jurisdictions subject to a Call for ...
2	black list	Iran	High-Risk Jurisdictions subject to a Call for ...

```python
[13]:  #create csv file from df_fatf_xlsx
       df_fatf_xlsx.to_csv("fatf_list_converted.csv", index=False, encoding='utf-8')
       print("Successfully created CSV file: 'fatf_list_converted.csv'")

       Successfully created CSV file: 'fatf_list_converted.csv'
```

Preprocessing was performed using the Python pandas and NumPy modules. The data was adapted for the present work taking into account the specifics of the data use.

```
[7]:   # update column names. assigning the column name to the second row
       df_UK_sanctions_list.columns = df_UK_sanctions_list.iloc[1]
```

```
[8]:   df_UK_sanctions_list = df_UK_sanctions_list.iloc[2:] #delete two first rows
```

```
[18]:  # cleaning operations
       df_UK_sanctions_short.columns = df_UK_sanctions_short.columns.str.replace(' ', '_').str.strip()
       df_UK_sanctions_ships.columns = df_UK_sanctions_ships.columns.str.replace(' ', '_').str.strip()
       df_UK_sanctions_individual.columns = df_UK_sanctions_individual.columns.str.replace(' ', '_').str.strip()
```

```
[19]:  #Convert columns to numeric
       df_UK_sanctions_ships.loc[:, 'Tonnage_of_ship'] = pd.to_numeric(df_UK_sanctions_ships['Tonnage_of_ship'], errors='coerce')
       df_UK_sanctions_ships.loc[:, 'Length_of_ship'] = pd.to_numeric(df_UK_sanctions_ships['Length_of_ship'], errors='coerce')
       df_UK_sanctions_ships = df_UK_sanctions_ships[df_UK_sanctions_ships['Tonnage_of_ship'] >= 0]
       #Convert Date_Designated to datetime
       df_UK_sanctions_ships.loc[:, 'Date_Designated'] = pd.to_datetime(df_UK_sanctions_ships['Date_Designated'], dayfirst=True)
       df_UK_sanctions_individual.loc[:, 'Date_Designated'] = pd.to_datetime(df_UK_sanctions_individual['Date_Designated'], dayfirst=True)
```
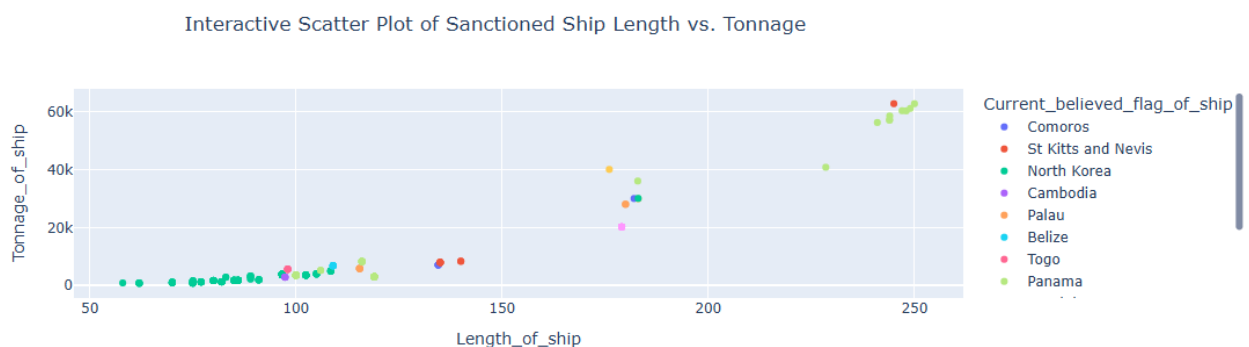
```
[16]:  # Filter to only keep rows where the 'Individual, Entity, Ship' column has the value 'Ship'
       df_ships_only = df_UK_sanctions_short[df_UK_sanctions_short['Individual, Entity, Ship'] == 'Ship'].copy()
       ships = ['Last Updated', 'Unique ID', 'OFSI Group ID', 'UN Reference Number', 'Name 6', 'Name type', 'Regime Name', 'Individual, Entity, Ship',
               'Designation source', 'Sanctions Imposed', 'UK Statement of Reasons', 'Address Country', 'Date Designated', 'IMO number',
               'Current believed flag of ship', 'Tonnage of ship', 'Length of ship']

       df_UK_sanctions_ships = df_ships_only[ships]
       df_UK_sanctions_ships.head(3)
```

```
[17]:  # Filter to only keep rows where the 'Individual, Entity, Ship' column has the value 'Individual'
       df_individual_only = df_UK_sanctions_list[df_UK_sanctions_list['Individual, Entity, Ship'] == 'Individual'].copy()
       individual = ['Last Updated', 'Unique ID', 'OFSI Group ID', 'UN Reference Number',
               'Name 6', 'Name type', 'Alias strength', 'Title', 'Regime Name',
               'Individual, Entity, Ship', 'Designation source', 'Sanctions Imposed',
               'UK Statement of Reasons','Address Country', 'Date Designated', 'D.O.B',
               'Nationality(/ies)', 'National Identifier number', 'Position', 'Gender', 'Country of birth', 'Type of entity',
               'Subsidiaries']

       df_UK_sanctions_individual = df_individual_only[individual]
       df_UK_sanctions_individual.head(3)
```
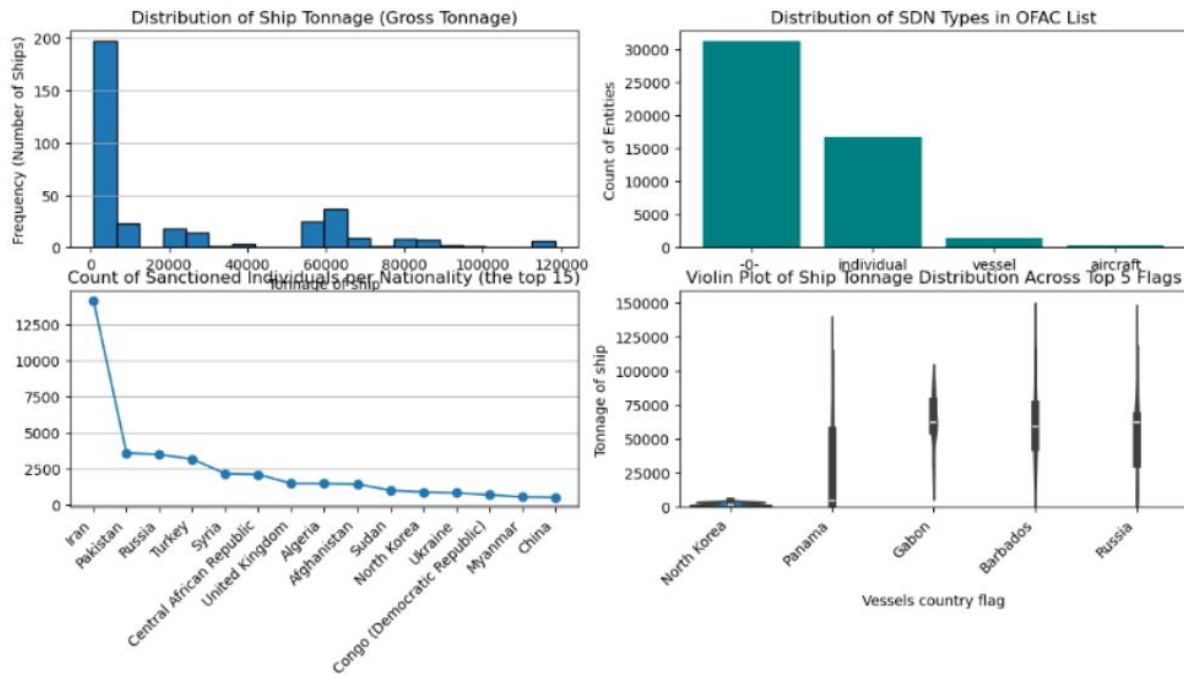
The EDA steps were performed by using the Python modules matplotlib, seaborn and plotly.

Interactive Scatter Plot of Sanctioned Ship Length vs. Tonnage

A combination of graphical models was developed that together visualize key characteristics of sanctioned vessels and individuals/entities on the lists, combining distributions, classifications, and geographic data.

Visualise the different models dashboards

KEY OBSERVATIONS:
Plot 1: Distribution of Ship Tonnage (Histogram)
Plot 2: Distribution of SDN Types (Bar Chart)
Plot 3: Count of Sanctioned Individuals per Nationality (Line/Bar Chart)
Plot 4: Violin Plot of Ship Tonnage Distribution Across Top 5 Flags
Summary - A combination of graphical models was developed that together visualize key characteristics of sanctioned vessels and individuals/entities on the lists, combining distributions, classifications, and geographic data

Together, histograms, line and pair plot, and bar charts provide an understanding of sanction lists, illustrating their content .

Some bug were fixes made thanks to hints when run the code. AI was used to fix some code errors.

Limitations: the data is somewhat specific, so its processing and visualization differs from standard ones. In this project, the data was analyzed and illustrated to give an overall picture of sanctions and show the areas of influence, including in a geographical context. In the future, I plan to develop this project to optimize the work.

Summary of Findings

We can draw the following conclusions from this project:

Analyzing transactions and clients to which countries should pay increased attention; It is necessary to additionally analyze the essence of the transaction and accompanying documents, as otherwise it may cause unjustified risks (for example, by analyzing the lists of top countries from the general analysis and analysis of data on vessels);
For correct work, it is necessary to take into account data from all lists, i.e. the data in them is different .

Suggestions for further work:

Data enrichment: Identify potential external data sources (e.g., ship registers, corporate registries) to enrich existing records, allowing for more complete identification and compliance checks.

Conduct network graph analysis using additional sources to visualize the connections between sanctioned individuals, organizations, and vessels, revealing complex ownership structures and hidden risks not apparent in simple tabular data.

Using machine learning methods to build predictive models based on historical data to optimize processes.

using machine learning methods to build predictive models based on historical data to optimize processes.

Ethical considerations

In accordance with GDPR, no personally identifiable information (PII) was included in the analyses. All data used in this project is public and available for use.

Unfixed bugs

After conducting a set of tests, I haven't found any bugs. Initial problems have been fixed. AI was used to fix some code errors.

Developement Roadmap

Among the difficulties I encountered were problems finding data that I could use in the project, which were resolved by searching the Internet. New skills and tools that I plan to learn further include deeper exploration of Python and SQL, visualization of spatiotemporal trends in a plot diagram using interactive functionality.

I also plan to learn how to apply machine learning methods to build predictive models (historical data), build forecasting models, network analysis such as building relationships between customers, accounts, and transactions, etc.

Main Data Analysis libraries

numpy
pandas
matplotlib
seaborn
plotly

Credits

The Code Institute course materials for the course "Data Analysis with Artificial Intelligence" were used as a template for the code in this project.

The code for feature development was adapted from (https://sanctionslist.ofac.treas.gov/Home/SdnList)

(https://www.gov.uk/government/publications/the-uk-sanctions-list)

FATF_list.xlsx was created from (https://www.fatf-gafi.org/en/countries/black-and-grey-lists.html)

All files including graphs were created and uploaded to GitHub