Freight Data Programming Challenge: Vessel Analysis

Overview

You are provided with a zipped CSV file with the following:

- `vessel_positions.csv` - containing tracking data of vessels for a period of 5 days. Each row represents a vessel's observation at a specific point in time.
- `vessel_characteristics.csv` - containing vessel characteristics

Your task is to performs a series of analyses and summary results.

This challenge involves creativity; you have the freedom to process and analyse the data in whichever way you prefer to achieve the required objective.

Requirements

- Use only standard data libraries (pandas, numpy, matplotlib, plotly, etc.) (Bonus points if you use `virtualenv` or uv)
- Your final delivery should include:
    - A link to a github repository which should contain:
        - A python notebook or python file with the implemented logic, analysis, and charts.
        - A readme file with a description of the repository and steps to run your code

      The repository should have a series of commits for each major step of the tasks.

Judgment Criteria

- Code Clarity and Structure: clean readable code with appropriate use of functions.
- Documentation and Reasoning: clear comments and markdown cells explaining assumptions made.
- Visualization and Insightfulness: effective use of charts.

Tasks

1. Data Loading and Cleaning

Load the attached dataset using pandas and clean the data. Explain (within the notebook) what techniques you used to clean the data and how any erroneous data points were resolved (if any).

2. Exploratory Analysis (Creative)

Provide an initial exploratory analysis of the data linking the following dimensions:

- vessel_operational_status
- ais_speed
- ais_draft

3. Add New Dimension

In the shipping industry, vessels are classified into categories based on their DWT.

vessel_class	vessel_dwt Range
Handymax	10,000 – 40,000
Supramax	40,000 – 59,999
Panamax	60,000 – 99,999
Capesize	≥ 100,000

Create a new column called vessel_class that maps each vessel to one of the above size categories based on its vessel_dwt value.

Summarize the data:

- How many vessels fall into each category?
- Provide a bar chart showing the number of unique vessels per class.

4. Cargo Analysis

Using the newly created attribute, analyse the vessel class in the context of the vessel commodity group (“vessel_commodity_group_onboard” attribute). Is there a link? Detail your observations.

5. Geographical Analysis

Define a polygon (approximate set of longitude and latitude values) that encompasses Southeast Asian countries. Generate a map with the polygon you've defined. Which vessels fall within this polygon?

6. Putting It All Together

Provide an analysis of vessels that loaded cargo in Southeast Asia between August 20, 2024 and August 24, 2024.

- What cargo was loaded onto these vessels?
- What are the top 2 exports from each country in Southeast Asia?
- Where is the cargo headed?
- Combined with the data in `vessel_characteristics.csv`, who are the most active operators in the region during this time period?
- The draft of a ship is a determined depth of the vessel below the waterline, measured vertically to its hull's lowest—its propellers, or keel, or other reference point. Draft varies according to the loaded condition of the ship. A deeper draft means the ship will have greater vertical depth below the waterline.
    - What's the relationship between the draft reported via AIS (`ais_draft`) compared with the vessel characteristic draft (`draft`)
    - How can we derive if a vessel is finished loading cargo based on draft levels?

Bonus points for additional insights.