AIS-based Analysis of Vessel Trajectories

<u>Assignment Objective</u>:

This assignment aims to evaluate your ability to analyse the provided data in pursuit of meaningful context-specific insights. In particular, you will focus on identifying various patterns in vessel movements—such as transits through open waters, stops at ports, anchoring outside port areas, waiting in port queues, instances of not serviced port visits, deviations from typical routes, multiple port visits, and other spatial and temporal patterns of relevance. The objective is to develop a set of functionality that enables better understanding and context-specific categorization of vessel trajectories.

<u>Introduction to AIS Data</u>:

The **Automatic Identification System (AIS)** is a maritime communication system that automatically broadcasts a ship's location and other relevant information to nearby vessels and coastal authorities. Originally developed for collision avoidance, AIS has become an essential tool for maritime safety, traffic management, and research. AIS records include information such as the vessel's position, speed, heading, and time of transmission.

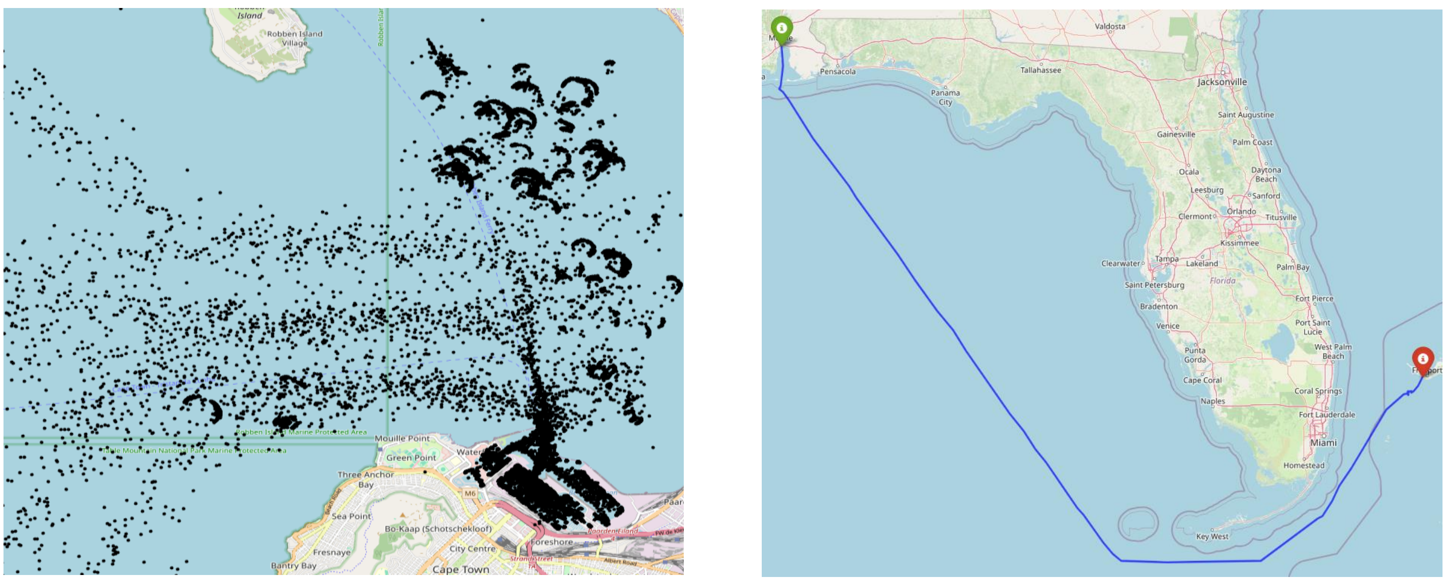

Figure 1: Left - Raw AIS data over a month over the port of Cape Town. **Right** – An example of a port-to-port trajectory.

To make sense of this vast amount of information, we need to preprocess and clean the data. Once the data is cleaned, we can reconstruct vessel trajectories—the paths that vessels take over time, represented by a series of geographic coordinates (see Figure 1). Analysing these trajectories in combination with other information (e.g. ports' locations) allows us to identify patterns in vessel movements, such as transits through open waters, stops at ports, anchoring outside port areas, waiting in port queues, and deviations from typical routes. By examining these patterns, we can extract meaningful, context-specific insights that are valuable for understanding maritime operations and behaviours. For further details on AIS and its applications, refer to the papers by Tu et al. (2017)[1] and Yang et al. (2019)[2].

[1] Tu, E., Zhang, G., Rachmawati, L., Rajabally, E., & Huang, G. B. (2017). Exploiting AIS data for intelligent maritime navigation: A comprehensive survey from data to methodology. *IEEE Transactions on Intelligent Transportation Systems*, *19*(5), 1559-1582.

[2] Yang, D., Wu, L., Wang, S., Jia, H., & Li, K. X. (2019). How big data enriches maritime research–a critical review of Automatic Identification System (AIS) data applications. *Transport Reviews*, *39*(6), 755-773.

Data Description:

The provided data set contains raw AIS data, containing 22,813,717 records across six columns. Details about these six columns are given in Table 1.

Columns	Data Type	Example	Info
anonymized_mmsi	string	deb97e3414	Vessel identifier
timestamp	object	2023-03-02 00:00:00	Timestamp of the AIS record
latitude	float64	51.346148	Latitude in degrees
longitude	float64	3.200460	Longitude in degrees
course	int64	274	Angle of the vessel with respect to the north
speed	float64	12.2	Speed of the vessel in knots

Table 1: Information on the columns of the provided AIS data.

Furthermore, the following files are provided:

1. **"global_port_info.pkl":** A list of dictionaries that holds information on almost all ports in the world.[3]
2. **"port_gdf.pkl":** A GeoPandas dataframe with polygons representing the areas of most ports in the world.
3. **"data_cleaning.py"**: A python file for basic preprocessing of the raw data.
4. **"helper_utils.py":** A python file with useful functions such as loading, saving and simple visualisation.

Deliverables:

Candidates are expected to submit a report explaining their approach and methods applied to complete the assignment (see Sections below). Additionally, the code should be shared through a GitHub repository.

[3] National Geospatial-Intelligence Agency. (n.d.). World Port Index. Maritime Safety Information. Retrieved November 7, 2024, from https://msi.nga.mil/Publications/WPI

Section 1: Data Preprocessing

Candidates are expected to clean and prepare the raw AIS data to ensure it is ready for analysis in the next section. They have the flexibility to use, modify, or expand upon the provided preprocessing code (file "data_cleaning.py") as they see fit – any approach should be justified and presented in the report. The functionality provided includes:

1. **Remove records with N/A values:**
   The raw dataset contains missing values. Basic functionality that removes records with missing values is provided in the function "remove_NA".
2. **Remove invalid latitude and longitude records:**
   Latitude values must be within the range of -90 to 90, and longitude values must fall between -180 and 180. Basic functionality that removes records not adhering to these ranges is provided in the function "remove_invalid_lat_and_long".
3. **Remove invalid speed records:**
   Negative speed values and unrealistically high speeds in the data should be removed. This is done in the function "remove_invalid_speed".

Section 2: Analysis of Vessel Trajectories

Candidates are encouraged to structure their analysis into the following levels, progressing from direct observations to actionable insights. These levels are provided for your convenience to help structure your analysis; however, you are free to diverge from them or focus on specific aspects according to your interests, without any penalty. By progressing through the different levels of analysis to the extent you find appropriate, you will collect a variety of features and insights that you can use to cluster and categorize vessel trajectories.

Level 1: Direct Observations

These are straightforward conditions identified directly from the data using simple thresholds or criteria. They serve as the foundation for further analysis.
Examples: ● **Speed-based Conditions:** Identify instances where a vessel's speed drops below a defined threshold (e.g., 0.5 knots), indicating potential anchoring or idle states. ● **Port Visits:** Detect when a vessel enters a port area based on its geographical coordinates.

Level 2: Derived Behaviours

At this level, you interpret patterns from the direct observations to infer more complex vessel behaviours. Examples: ● **Waiting in Port Queues or Berths:** Determine if a vessel remains stationary in a specific area for a minimum duration, suggesting it is waiting in a queue or at a berth. ● **Unserved Port Visits:** Analyse cases where a vessel visits a port anchorage but leaves without entering the port or receiving services, based on the time spent and movement patterns.

Level 3: Aggregated Patterns

This level involves summarizing and analysing data across multiple vessels or time periods to identify broader trends and patterns.
Examples: ● **Spatial Mappings:** Create maps that display e.g. all trajectories for a specific vessel or trajectories that meet certain criteria, such as traveling above 0.5 knots in a particular region. ● **Temporal Patterns:** Calculate metrics like the average duration of stays at specific locations or identify peak times for port entries. ● **Spatial Metrics:** Identify frequently visited areas, such as popular anchorage zones or congested berths, to reveal activity hotspots.

Level 4: Strategic Insights

Here, you derive practical insights that can inform decision-making or operational strategies, based on your analysis from the previous levels.
Examples: ● **Operational Efficiency:** Use duration and spatial metrics to pinpoint inefficiencies in port operations or shipping routes, potentially suggesting areas for improvement. ● **Behavioural Trends:** Examine patterns to understand common vessel behaviours, such as typical waiting times in port queues or frequent unserved port visits, which could indicate industry-wide issues. ● **Traffic Predictions:** Utilize aggregated data to forecast future vessel traffic to specific ports or regions, aiding in resource planning and management.