Part A – California environmental agency

The California environmental agency are interested in monitoring the condition of the sea off its coastline to protect its local marine life. One factor they want to monitor, in particular, is how many large shipping vessels pass through their shores. Currently, they accomplish this by examining daily satellite imagery and manually counting the number of large shipping vessels in and around their coastline. Ideally, they would like to automate this task.

They have provided a dataset that includes 4000 80x80 RGB images labelled either "ship" or "no-ship". The satellite images have a pixel ratio of 3 meters. Please see the attached file "shipdata_2025.zip"

Inside the shipata_2025.zip file is a directory named `cropped_ship_dataset`, which contains the entire dataset as PNG images. Each individual image filename follows a specific format: `{label}__{scene id}__{longitude}_{latitude}.png`, where:
- `label`: Valued 1 or 0, representing the "ship" class and "no-ship" class, respectively.
- `scene id`: The unique identifier of the PlanetScope visual scene the image was extracted from.
- `longitude_latitude`: The longitude and latitude coordinates of the image centre point, with values separated by a single underscore.

There is a further directory, `scenes`, which contains some examples of full-satellite images. These are not the same images as the 80x80 cropped images.

Your task is to deliver a report aimed at a technical audience in a business-friendly format that will be send to the authority. The report should contain:
- Exploratory data analysis of the given dataset.
- Advice/pipeline suggestions on how to automate the counting of the number of ships in the full scene satellite images.
- Any limitations to the approaches suggested and/or the dataset.
- Anything else that you think may be relevant for the agency (e.g. deployment or real-time monitoring).

The report should be a maximum of 500 words and 3 figures. You are free to use whichever methods you see fit and may submit the report in any format you wish, ensuring that the interview panel can assess your file(s). Please also submit any relevant code you used to produce the report (e.g. figures, tables). We do not expect you to spend more than four hours on this task.

Finally, please be prepared to answer questions about your report during the interview.