The Port of Long Beach is one of the busiest ports in the United States and is currently in need for smarter yard management to keep up with the expected increase in cargo volume. One of the first steps to effectively managing yard resources is to understand the variance of demand at the port.

Datasets have been shared with you via a one-drive link, but you may also download them directly from:

AIS Data (**AIS\_2020\_01\_01.zip** and **AIS\_2020\_01\_02.zip**): <https://coast.noaa.gov/htdata/CMSP/AISDataHandler/2020/index.html>

Port coordinates: <https://msi.nga.mil/api/publications/download?type=view&key=16920959/SFH00000/UpdatedPub150.csv>

AIS data dictionary may be found in:

<https://coast.noaa.gov/data/marinecadastre/ais/data-dictionary.pdf>

You are given two data sets:

1. AIS\_2020\_01\_01.zip and AIS\_2020\_01\_02.zip, which contains vessel tracking data for the entire US continental region, for the 1st and 2nd of January 2020
2. UpdatedPub150.csv, which contains the coordinates of marine ports.

You are required to carry out the following tasks

1. Create a local Postgres server and import the data from AIS\_2020\_01\_01.zip, AIS\_2020\_01\_02.zip, and UpdatedPub150.csv. One table for the AIS data, and one for the port coordinates.
   1. The SQL code that you use to create the user role, table schema, and import the data, is to be included in your submission.
2. The data within AIS\_2020\_01\_\* is large, therefore, it would be best to analyze samples that are relevant for our task. One way to do this is to create a bounding box and filter out samples that are outside this region, as well as only consider **cargo vessels**.

A bounding box can be created by defining a center point, and a width and height. In this case the center point can be the location of Long Beach Port, and the width and height can be user defined. (Hint: Plot bounding box on a map to determine an appropriate size for your bounding box). Knowing this, write a python function that accepts as input a port un/locode (port-code), a width, and height in decimal degrees and outputs the list of relevant cargo vessel samples from the AIS table created in the previous step.

1. We want to analyze the number of unique **cargo vessels** within the port of Long Beach, throughout the day. Write a function that accept a port-code and outputs a table with two columns. A **date-time column** and the **number of unique vessels** within the port at a given block of time.
2. With appropriate plots, illustrate the times that the port experiences low and high demand. What is the best temporal resolution for illustrating this? Furthermore, is there any large variance with demand from the 1st of January and the 2nd of January? If so, why do you think this is?

Your final code (SQL scripts and python code) should be pushed to a personal GitHub repo. Any comments, or assumptions made should be included as part of your comments in the code. Illustrations and insights gathered from it should be included in your repos wiki. The final submission should be a link to the GitHub repo. Make sure that the visibility setting on your repo allows us to view it.

What we are looking for:

1. Clean code. Functions should be created using software engineering best practices including comments, docstrings, data types, descriptive variable and function names, modularity of code, as well as the following of best security practices when passing credentials to functions.
2. Clear instructions in your readme for how to get your project running.
3. Visualization that clearly illustrate your point.