Waimea Bay Waves

Waimea Bay, located on the North Shore of Oahu, is one of the original big wave surf destinations.

When Waimea receives a big swell (large waves!), it's an incredible — and dangerous — place for surfers to make a name for themselves. When the swell is small, there's nothing but a modest shore break for tourists to play in.

In this project, we'll imagine that the *World Surf League* (WSL) has contacted us about the waves at Waimea Bay. They would like to hold a new surf contest at Waimea Bay, but they are worried that they won't be able to anticipate when the waves will be large enough to surf (at least 3m in height!).

Can we help them forecast the wave conditions at Waimea Bay so that they are able to successfully run a big wave contest?

Data

To prepare for this project, we have collected data from 5 buoys located in the Pacific Ocean — three off the North Shore of Oahu, one located off the coast of Tokyo, and the other near Alaska. Each buoy collects *some or all* of these metrics at daily intervals:

- Air temperature (degrees Celsius)
- Wave height (metres)
- Average wave period (seconds) — interval between all subsequent waves
- Dominant wave period (seconds) — interval between significant (swell) waves

These data are located on Github:

1. **buoy-data.csv** - A record for each day from each buoy.
2. **wide.csv** - For convenience, the same data stretched into "wide" format — one record per day, with metrics from each buoy in each column. Empty columns have been removed.

Part 1: Open Questions

1. Framing the Problem

Before we dive in and begin working with the WSL, it's important that we consider their objectives, technical requirements, project feasibility, data sources, and other issues that can influence the success of the project.

What questions would you ask the WSL on a kick-off call before beginning on the project?

2. Enrichment

We have already collected and cleaned data from several buoys located in the Pacific. Nevertheless, we know that forecasts largely depend on the *quality* and *quantity* of the available data.

Before diving into the analysis and forecasting, which new data sources would you consider to complement to the buoy data? How would you consider using these data on an ongoing basis?

3. Data Integrity

If you haven't already done so, inspect the buoy data that we have prepared. As you'll notice, there are several datapoints that are missing.

How would you handle missing data in this project?

4. Approaches

There are numerous approaches to forecasting problems, none of which are perfect.

Which approaches would you consider for this project? Explain your recommendations as if you were communicating directly to the CEO of WSL.

5. Communication

Assuming we are able to generate a sufficiently-accurate forecast for the WSL, we need some means by which to communicate it so that they can schedule their Waimea Bay event.

How would you recommend that these results be communicated to the WSL on an ongoing basis?

Part 2: Proof of Concept

Using R and/or Python and the data that we have provided, develop a proof of concept approach to forecasting future wave heights at Waimea Bay.

Given the time constraints, this approach need neither be perfect nor comprehensive. Nevertheless, it should be a clear and justifiable “starting point” for future iteration on more complex, production-ready forecasting approaches.

After completing your proof of concept, briefly outline the next steps that you would take if you had an entire month to work on the problem.