Buoys to Barrels

CSPB 4122 - Fall 2022

Ronald Durham
Applied Computer Science
University of Colorado Boulder
Boulder, Colorado, USA
rodu4835@colorado.edu

CS CONCEPTS

- Information Visualization
- Data Cleaning and Processing

REFERENCE FORMAT:

Ronald Durham. 2022. Buoys to Barrels. In CSPB 4122 - Information Visualization. Boulder, CO, USA, 9 pages.

INTRODUCTION

The objective of this project is to better understand ocean wave information collected from a stationary buoy to predict optimal surf conditions, “Barrels!”. While there are resources that already visualize similar datasets and provide users with surf conditions, the deeper understanding of what conditions cause each outcome is, for the majority, left out. The target user would be someone interested in how the raw data from the buoys can be used to determine wave conditions throughout time, created with the surfer in mind.

The main target user for this data visualization will be those interested in understanding how the data taken from a buoy translates into predictions for optimal waves. There are certain particular conditions that create waves that are great for surfing. With much research into the sport, consulting other surfers, and self practice, it has been seen that the best waves occur during the following conditions: in between low and high points during a rising tide, light off shore winds (blowing away from the shore), and differing conditions of wave period, wave height, and swell direction depending on the shape and direction of the beach, reef, sand bar, etc. Because this project focuses on the buoy data taken from just off shore of Leucadia State Beach (a.k.a. Beacon’s) in Leucadia, California, we will look to understand the metrics based on the type and direction of the break at Beacon’s.

The beach at Beacon’s faces West-South West and has 3 distinct breaks, south (deep farther out reef break), middle (sandy bottom beach break), and north (another reef but not as far out as south). These are key features to keep in mind when understanding the data visualization for this project. The visualization will look to highlight when conditions are in their prime for the spot, this will be what we have mentioned above as well as looking for longer wave periods, larger wave height, and swells coming from west / north west.

DATASET

The data is taken from the National Data Buoy Center, NDBC, for the Leucadia, California buoy 46274 from 8/31/2022 - 9/26/2022. Buoy 46274 is located at the following coordinates:

33.062 N 117.314 W (33°3'43" N 117°18'51" W)

The data was originally collected in raw text format with spaces as separators and null values filled in as "N/A" or "MM." As for the size of the data, there are 1249 entries in the set taken every half hour spanning back 27 days. The data was obtained from the reports hosted on the National Data Buoy Center website which is collected automatically from all of the registered data collection buoys across the nation. I have specifically selected buoy 46274 as it is just offshore from the beach that I surf most often and as far as I can tell, they use exceptionally accurate measuring devices installed on the buoys.

However, I did find that some data was missing. It turns out that not all of the buoys have all of the instruments installed on them. The specific buoy that I chose, 46274, does not have readings for wind speed, wind direction, gust winds, atmospheric pressure, pressure tendency, dew point temperature, salinity, visibility, or tide data. There are scattered buoys off the coast of California that have all of the instruments on them and weather services will report data based on a combination of data from the closest buoys to your area. For my test / sample set, I have removed the columns that are missing data entirely.

The data is actually acquired in two different reports from the same buoy, Standard Meteorological Data and Detailed Wave Summary, that I have merged into one. You can see the column title in the first row followed by the unit of measure in the second row, then the actual dataset beyond that. As for describing the data, each column ranges from being incredibly specific to pretty broad so I felt that the complete descriptions taken directly from the NDBC website would do best to help understand the dataset. The description of the columns are as follows, I have only included the columns that I ended up using:

- WVHT - Significant wave height (meters) is calculated as the average of the highest one-third of all of the wave heights during the 20-minute sampling period.
- SwH - Swell height is the vertical distance (meters) between any swell crest and the succeeding swell wave trough.
- SwP - Swell Period is the time (usually measured in seconds) that it takes successive swell wave crests or troughs to pass a fixed point.
- DPD - Dominant wave period (seconds) is the period with the maximum wave energy.
- WWH - Wind Wave Height is the vertical distance (meters) between any wind wave crest and the succeeding wind wave trough (independent of swell waves).
- WWP - Wind Wave Period is the time (in seconds) that it takes successive wind wave crests or troughs to pass a fixed point.

- SwD - The direction from which the swell waves at the swell wave period (SWPD) are coming. The units are degrees from true North, increasing clockwise, with North as 0 (zero) degrees and East as 90 degrees.
- WWD - The direction from which the wind waves at the wind wave period (WWPD) are coming. The units are degrees from true North, increasing clockwise, with North as 0 (zero) degrees and East as 90 degrees.
- STEEPNESS - Wave steepness is the ratio of wave height to wavelength and is an indicator of wave stability. When wave steepness exceeds a 1/7 ratio; the wave becomes unstable and begins to break.
- MWD - The direction from which the waves at the dominant period (DPD) are coming. The units are degrees from true North, increasing clockwise, with North as 0 (zero) degrees and East as 90 degrees.
- APD - Average Wave Period is the average period (seconds) of the highest one-third of the wave observed during a 20 minute sampling period.
- ATMP - Air temperature (Celsius).
- WTMP - Sea surface temperature (Celsius). For buoys the depth is referenced to the hull's waterline.

I first looked to address the format of the data by converting it into a csv or excel format and ensured that all of the column titles, data unit descriptions, etcetera, each accurately represents the raw data file.

I did find that one column was incomplete, the wave steepness column. In order to remedy this, I found the method that the NDBC uses to calculate wave steepness which included formulas, pseudocode, and a more detailed description for this data. I then wrote a python function that worked through the formulas and filled in the missing values for that row. [1]

I first created a separate csv file called test.csv that only contained the values within the attribute "STEEPNESS". I then pushed that file through the code to the right and compared it to the original data to verify that the two columns matched and the null values were filled in.

The algorithm follows:

```
val = exp(-3.3 * ln(fp))
if Hs > (val/250)
   steepness = 'very steep';
elseif Hs > (val/500)
   steepness = 'steep';
elseif Hs > (val/1000)
   steepness = 'average';
else
   if(SwH or WWH is missing)
      steepness = 'N/A';
   elseif SwH >= WWH
      steepness = 'swell';
   else
      steepness = 'average';
   endif
endif
```

```python
import pandas as pd
import numpy as np
import math
def steepness(fileName):
    filePath = './'+str(fileName)
    df = pd.read_csv(filePath)
    df["Steepness"] = ""
    for i in range(0, len(df)):
        fp = (1 / df["DPD"].iloc[i])
        val = math.exp(-3.3*np.log(fp))
        if (df["Hs"].iloc[i] > (val/250)):
            df.loc[i, "Steepness"] = "VERY_STEEP"
        elif (df["Hs"].iloc[i] > (val/500)):
            df.loc[i, "Steepness"] = "STEEP"
        elif (df["Hs"].iloc[i] > (val/1000)):
            df.loc[i, "Steepness"] = "AVERAGE"
        else:
            if (df["SwH"].iloc[i] >= df["WWH"].iloc[i]):
                df.loc[i, "Steepness"] = "SWELL"
            else:
                df.loc[i, "Steepness"] = "AVERAGE"
    return df.to_csv('steepness.csv')
```

```python
steepness('test.csv')
```

Below is a sample of my original cleaned dataset:

Year	Month	Day	Hour	Minute	WVHT	SwH	SwP	DPD	WWH	WWP	SwD	WWD	Steepness	MWD	APD	ATMP	WTMP
#yr	mo	dy	hr	mn	m	m	sec	sec	m	sec	-	degT	-	degT	sec	degC	degC
2022	9	26	5	26	1.1	0.5	15.4	15	0.9	5.9	SW	W	AVERAGE	232	5.7	20.9	21.5
2022	9	26	4	56	1.1	0.7	15.4	15	0.9	5.9	SW	W	AVERAGE	228	6.2	21.1	21.6
2022	9	26	4	26	1.1	0.6	16.7	17	1	5.9	SW	W	AVERAGE	233	5.8	21.1	21.6
2022	9	26	3	56	1.1	0.6	15.4	15	0.9	5	SW	W	AVERAGE	231	5.9	21.2	21.7
2022	9	26	3	26	0.9	0.4	16.7	17	0.8	5.9	WSW	W	AVERAGE	239	5.4	21.2	21.8
2022	9	26	2	56	1	0.6	15.4	15	0.8	5.3	SW	W	AVERAGE	232	5.7	21.2	21.9
2022	9	26	2	26	1	0.7	16.7	17	0.8	5.3	SW	W	AVERAGE	232	6.1	21.3	21.9
2022	9	26	1	56	1	0.5	16.7	17	0.8	5.6	SW	W	AVERAGE	231	5.4	21.5	21.9
2022	9	26	1	26	0.9	0.5	16.7	17	0.8	5.6	SW	W	AVERAGE	232	5.7	21.9	22
2022	9	26	0	56	0.9	0.5	16.7	17	0.8	5.3	SW	WNW	AVERAGE	228	5.3	22.3	22
2022	9	26	0	26	0.9	0.5	16.7	17	0.8	5.9	SW	W	AVERAGE	233	5.5	22.3	22
2022	9	25	23	56	0.9	0.5	16.7	17	0.7	6.2	SW	W	AVERAGE	232	5.8	22.4	22.1
2022	9	25	23	26	0.9	0.5	15.4	15	0.8	5.9	SW	W	AVERAGE	232	5.6	22.4	22.3

TASKS

Taking into account the data we have from this buoy plus the knowledge of what constitutes better wave conditions at the breaks at Beacons, I've determined that this project should be able to let the user accomplish the following tasks In order to better understand potential surf conditions:

- The user should be able to figure out trends in previous ocean conditions
- They should be able to determine if conditions are improving or diminishing
- They should be able to see what direction the waves have most commonly been coming from and how that affects the overall steepness of the waves at Beacons
- They should be able to consider water temperature data to be able to choose the correct wetsuit thickness for that time of year

In order to accomplish these tasks, the user will need to be familiar with the beginning of this paper where I discussed the conditions that create better waves at beacons. Reiterated as, longer wave periods, larger wave height, and swells coming from west / north west.

These specific tasks are important to being able to see trends in surf conditions. Some surf spots are better than others at different times of the year and if someone was so inclined, they could look at all of the historic data using a tool like this one to determine which spot to go to when. Mainly, the wave breaks follow a consistent pattern for each spot over the course of the year. Not accounting for shifting sand bars and sporadic storm activity over time. The direction of waves will help the user understand the overall ideas behind the waves at beacons. Ideally, someone could look at this data and know that, with these types of ground conditions, Beacons is normally a mellow lump of a wave, perfect for beginners, and easy to advance through more skill levels with its forgiving shape.

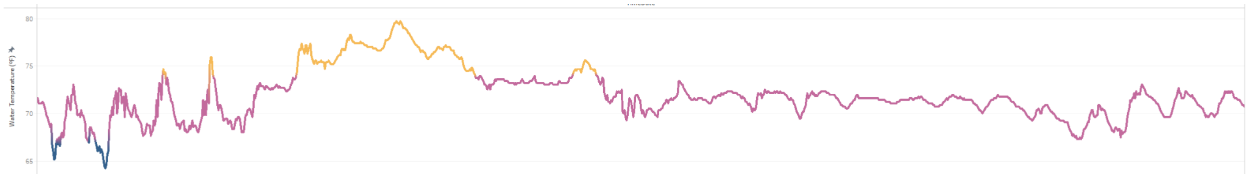
This visualization tool will also be able to help the surfer be better prepared with what type of board to bring by understanding the recent swell and wave conditions as well as what attire to bring in order to be prepared for the temperature of the water. As for boards, generally longer sloping waves would require longer boards and taller steep waves would require shorter boards. As for clothing, everyone has a different tolerance or preference for what type of gear they like to wear in each temperature water so having the pure temperature data is important. Generally in water temperatures from 58° - 68°F, I would wear booties and what is known as a 3/2, where your chest has 3mm of wetsuit material thickness and the sleeves and legs have 2mm material thickness. From 68° - 72° I would most likely wear a spring suit, 2mm short sleeves and legs, as long as the ambient air temperature is warm and the sun is not hiding behind clouds. Above 72°, generally trunks and a rashguard is all that is needed. This is my personal preference and others may wear different gear. It is important for the user to know their preference in order to understand and apply this visualization tool.

VISUAL DESIGNS

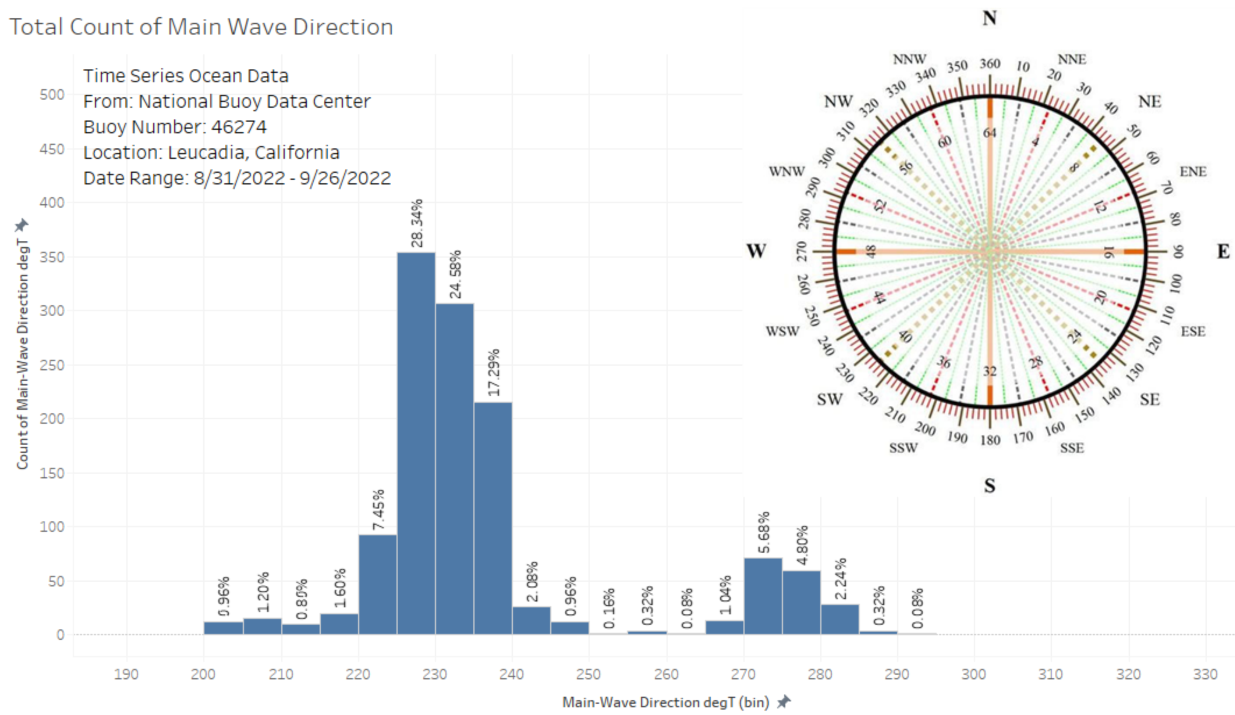
As mentioned, the major key ideas to look for to determine better waves for this specific break are longer wave periods, larger wave height, and swells coming from west / north west. In order to look further into these metrics and to be able to accomplish the tasks I previously determined, I've plotted Average Wave Period, Significant Wave Height, Swell Height, and Water Temperature over time. Additionally, I plotted a histogram of the wave direction during this time frame and a sort of stacked bar graph for wave steepness compared to the main wave direction. Hovering over any of the points on the graphs will show specific information for that day. The days are split up by hour and every 30 minutes but, in order to better understand each given day, I grouped the times for each day together and you are able to look at one specific day in detail by selecting the day group on the right side. I also converted the celsius and metric measures in the dataset to fahrenheit and standard as I am more familiar with those.

I decided to connect the color palette for swell height, wave height, and wave period which are worse when the color is red and better when the color is green as those three metrics are directly connected and important to our understanding of better waves. Choosing these color coded bar graphs worked out great, especially because the set had so many data points. When plotted over time, it is easy to see the connection between the attributes.

I also plotted a line graph along the same X axis showing change in water temperature over the time period. This one is color coded as well to show when the water was colder vs warmer, blue being cold, yellow being warm, and magenta in between. A line graph with stepping colors was perfect for this and makes understanding the temperature data quite clear.



Next, we can look at the histogram of the total count of wave directions during this time frame. A histogram is great for this as it quickly lets us see the general proportion of where the waves are coming from and then we can look at the main 4 charts to see how generally good the conditions were. I ended up including the image of the windrose for those that do not know exactly how the degrees measurement connects to cardinal directions.



Finally, in order to understand how the wave direction correlates with the steepness of the wave, I thought it was important to graph wave steepness vs main wave direction. This is a sort of stacked bar graph that plots the counts of each direction and how steep the wave is. It is also color coded for the direction bin which are grouped by every 5 degrees. This is great as you are able to see specifically how the direction compares to each type of wave steepness. For reference, the steepness measures are put in order from the most mellow, “swell”, to very steep. There could have been other ways of charting these attributes but I liked this version as it still allows you to see specifics about each group rather than further grouping and plotting on something like a density plot.

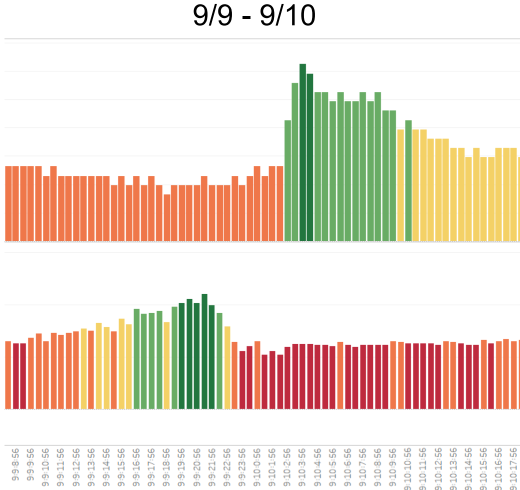
STEEPNESS

RESULTS

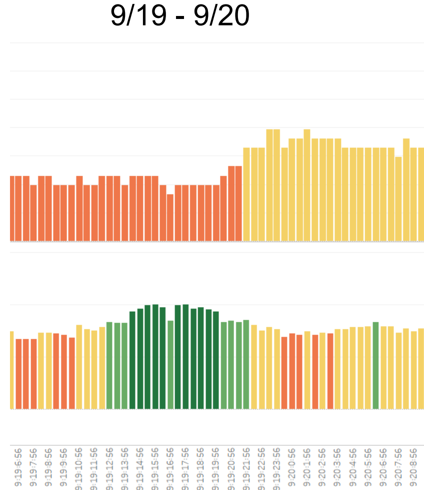
From first glance, it is easy to see how this tool can be used to look at historical trends. If the dataset was expanded over the course of a year and then compared that year's data with sets from years in the past you could definitely use this as a tool for historical analysis. This version of the tool shows historical trends but over the 45 days in the dataset. This satisfies the first task that was set out by this project.

Next, the tool can be used to understand whether or not conditions are improving or diminishing. Keeping in mind the ideal conditions for the break at Beacons, you would start by looking for longer wave periods. It is absolutely apparent that as the wave period increases so does the wave height as seen in these sections here:

9/9 - 9/10



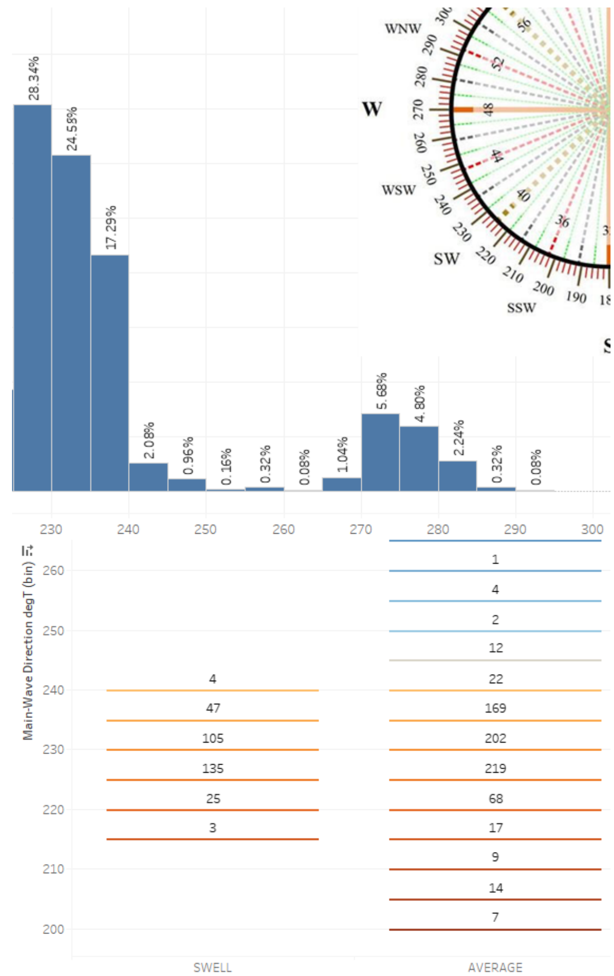
9/19 - 9/20



The next two graphs, the histogram and the wave steepness, are in response to task 3 and allow the user to see the resulting wave steepness compared to which direction the waves came from. In this dataset, roughly 70% of the waves came from between 225 and 235 degrees, or mainly south west. Knowing that the ideal conditions for this spot come from the west / north west, we can assume that these waves are not going to be the greatest but should not be too terrible, i.e. average.

Then we can look at the wave steepness graph, and as it turns out waves that come from the southwest in fact produce average wave steepness. You can also see that the waves that did actually come from the west / north west or roughly 270 or 280 degrees did produce waves that fell into the steep or very steep category, confirming that our ideal conditions are waves coming from that direction.

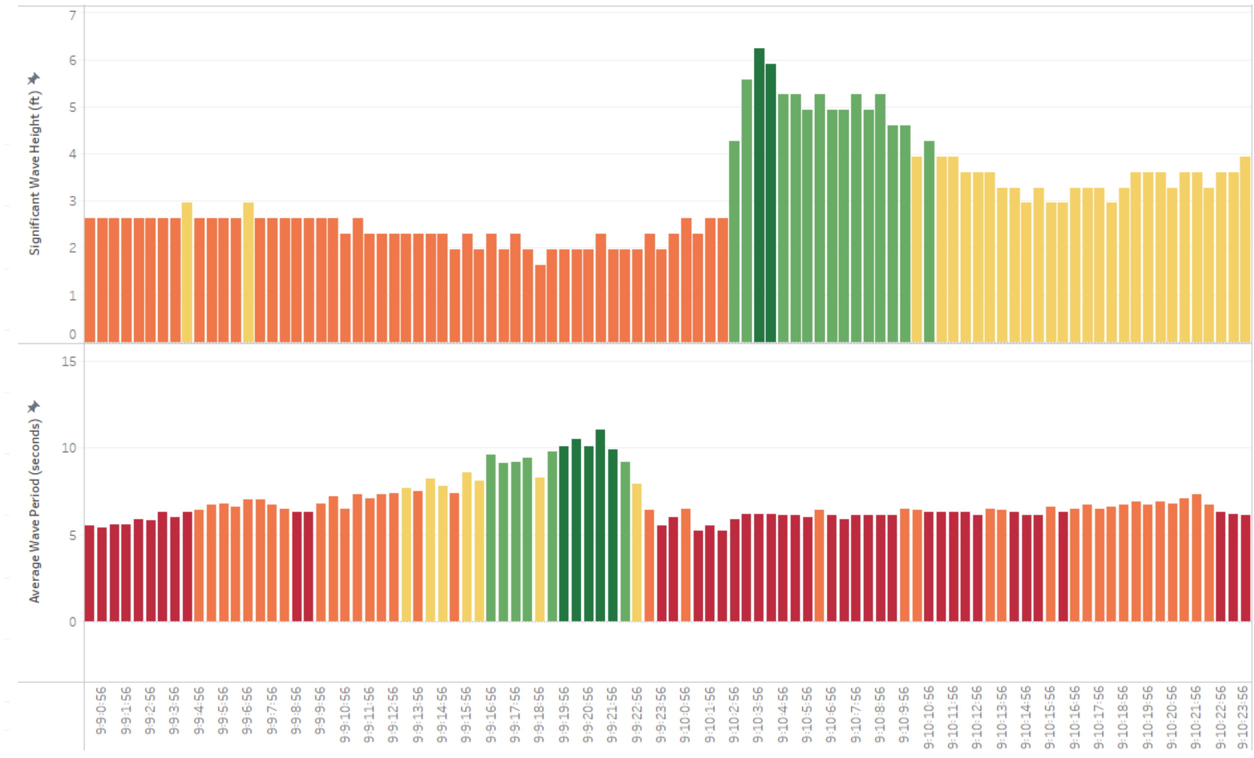
Future work could be to expand the dataset and incorporate multiple year's worth of values to compare. From here, it would be possible to further look into trends and really narrow down when the best time to visit beacons would be to potentially get barrels!



DESIGN PROCESS AND CHALLENGES

In order to complete this project, I mainly used Tableau, Python, and Excel. Excel and Python were great secondary tools alongside Tableau when performing my data cleaning. The simplicity of Excel when looking at individual sheets, columns, and rows, plus being able to easily integrate the files with Python was key in organizing my data and filling out incomplete columns. Tableau did have its difficulties and proved to be a challenge to get working as intended but in the end I'm happy with the results. First, I needed to bring together the columns for date and time. The fact that these were separated by year, month, day, hour, and minute made it necessary to have a date/time format that would retain the order but join these four columns into one. This was accomplished in tableau by defining and creating groups for these columns/rows.

From here, I was able to plot each of the main attributes in the stack of 4 charts against the DateTime variable. Once I had decided on which graphing chart would best represent the data, I began to set on the color coding aspect. This really brought the visualization to life and I wanted to take it a step further. All of the 4 stacked charts show connecting data in the hover tooltip. This way you are able to look at the specific data for any specific date and time. Once you select that portion the rest of the visualization fades and your selection is highlighted. Even better, if you use the Day Groups box to the right you are able to view the data from one specific day, as seen below for 9/9 - 9/10:



Next, I looked into making the windrose for the wave direction data. This was a major challenge and with some substantial guidance I believe it would be possible to plot the direction bins superimposed over a windrose in order to match up the data with the measurement tool, but I was unable to get it to do so. This is the next best version and possibly shows the curve of the histogram in a more understandable method.

Finally, I created the wave steepness graph which came together quickly as I had already created the day groups bins and the wave direction bins. I had originally set out to include a map of the area overlaid with wind data but as I was unable to get the data to plot directionally, and I did not have any information on wind speed I decided that this was a great way to work with what I had in my dataset.

REFERENCES

[1] Anon. 1996. How are significant wave height, dominant period, average period, and wave steepness calculated? (November 1996). Retrieved October 7, 2022 from https://www.ndbc.noaa.gov/wavecalc.shtml

[2] Anon. 1996. Station 46274 - Leucadia Nearshore, CA (262). (November 1996). Retrieved September 26, 2022 from https://www.ndbc.noaa.gov/station_page.php?station=46274