

## PoolPal Initial EDA Report

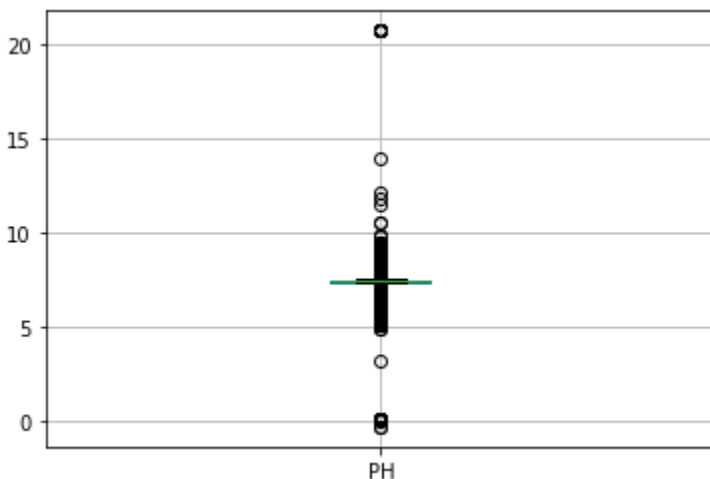
The Springboard Data Science team got together after doing some initial exploration on the data, there were a few components of the data which we wanted to dig deeper into. We hope you can provide us with some more information. The following report will contain the three factors we wished to delve deeper into: (1) collected data, (2) future/additional data, and (3) business metrics.

### Collected Data

Firstly, we had some questions about the collected data. Some of the major concerns relate to data quality and our ability to connect the data points with each other. The goal is to be able to use collected data to generate recommendations for treatment (pH stabilization through chemical treatment) and anticipate those needs (based on other pool factors). As a note, the time series data shifted due to the daylight savings time leading to some duplicate data.

The team had the following major assessments based on the initial data given:

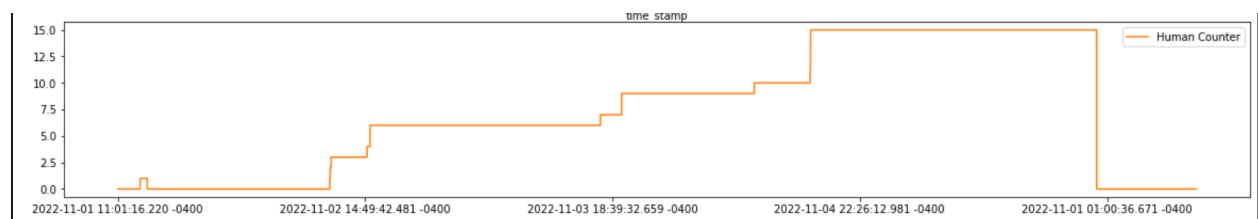
- We noted some columns were missing data entirely, could we get some clarification?
- pH itself has a large range from 0-20, were there some sensor issues? (**Figure 1**)
- Salinity, ORD, TDS, and some other measures seemed to greatly fluctuate in the beginning of the data. Was this due to sensor or weather issues? (**Figure 2**)
- Can we get clarification on the human counter? It seemed to grow over time and stay the same for long periods. Is this keeping track of weekly usage? (**Figure 3**)
- There are some terms such as pHfeedtime, pump current, orp feedtime, shw msg, cppm and others that are unclear. Could we get a breakdown of what the terms mean?



*Figure 1. pH range*



*Figure 2. Strange activity at start of data*



*Figure 3. Human Counter*

We believe by resolving the current data issues and the gap in understanding, we can better understand what approaches may be appropriate for the data. At present, our target seems to be pH. However, we do not seem to have any clear indicators of what impact pH and how changes in pH can be resolved since the pH measurement itself seems quite distraught. Clarification on sensors and how the data is collected/measured would be vastly helpful.

### **Future/Additional Data**

Secondly, we wanted to ask if we could get some more data that might help us understand the relationship between the data some more:

- Is there any information on chemical usage? Such as dosage, dosage time, and length of release?
- Is there information on chemical costs?
- Is there any information on energy data? Similarly, do we have a way of knowing when certain devices were turned on/off and how that impacts energy usage?

We believe that having this additional data is key to being able to create the recommendations for our consumers. As we build out the model our majors hopes are to be able to know when certain targets will go outside of acceptable constraints (i.e. predicting pH changes due to changes in salinity, human occupancy, etc;) and making recommendations (i.e. dosage of chemicals) to combat these changes in as efficient a way as possible.

### **Business Constraints/Thoughts**

Thirdly, we wanted to discuss the business constraints and goals.

- We found a case study [here](#) that discusses potential saving for a business. Would you consider that a reasonable estimate?
- We would like to understand the current situation of the business insofar chemical and energy consumption with relation to costs.
- May we be sent some guidelines regarding ideal ranges? We know that temperature wise, we are aiming for 24C and a pH of 7.2-7.8 per guidelines.

Understanding the business metrics and targets which we need to abide by gives us some metrics for success. We can use these metrics in comparison to “traditional” treatments. This will also allow us to help build projections for savings for customers which will allow PoolPal to better communicate the benefits of the system to customers.

Thank you for taking the time to look over this. Please let us know if you have any questions or further thoughts.

For some further details regarding the data, please refer to the following Pandas profiling file: [Pandas Profile HTML Download](#)