Non-Technical Report

Dominika Markowska-Desvallons, John Mazon, Orli Khaimova

## **Background**

At ABC Beverage co., as a data science team , we were seeking to better understand our manufacturing processes due to new regulations. We were trying to comprehend the impact of our beverages' pH levels during production. We aimed to influence the senior management about our findings, and prepared inquiries/processes for the updated rules.

We explored and analyzed the given data with varieties of models. After investigation, we picked the best performing one which suggested the best forecasting in our case.

## **Data**

Two historical data sets were provided. “StudentData” was the data that was used to help build a model while “StudentEvaluation” was the data which we had to predict the pH levels.

## **Model Selection**

After creating and analyzing multiple models: linear, non-linear, and tree-based models, we selected the Random Forest tree model as the best performing model for our project. In our technical report, you can see our process and find more details about this.

Based on the RMSE and R2 values of all the models we ran, the model is the overall best performer. This is expected given what this model is. The R2 for the model tells us that it explains what percentage of the data variance which falls within an acceptable R2 value range. Based on this, we will proceed with the model as the best predictive model. It also requires less preprocessing than other models and is more computationally efficient.

## **Predictive Factors**

From our findings, we found that the brand code had the biggest effect on the pH levels. Those that were labeled “C”, tended to have the lowest pH level while those that were labeled “D” had the highest pH levels. “B” accounted for roughly a half of the data, and had the second highest pH levels on average.

The Minimum Night Flow (Mnf.Flow) has the second largest impact on pH levels. It is found to be negatively correlated with the pH levels, as well as the usage control (Usage.cont). We found that oxygen filler and pressure vacuum had a positive effect on the pH levels. The other factors had less of an impact on the pH levels.

There also seems to be a significant correlation between the Balling level and Carb. Rel, as well as the Carb. Rel and Alch. level. We also found that the oxygen filler and pressure vacuum had a negative relationship with the Minimum Night Flow, while the usage control had a positive effect on it.



## 

## **Conclusion**

With these new regulations in place, it is best to monitor the pH in order to be in compliance. The factors that we have mentioned in this report are the most important for the beverage manufacturing process and need to be monitored the most.

We have included our predictions of the pH levels in a separate Excel spreadsheet. Please let us know if there are any further questions and contact us as needed.