To: Jeff Nieman - Director, Production
From: Donny Lofland, Dennis Pong, Charlie Rosemond - Data Science, Production
Date: 12/13/2020
RE: Understanding the ABC Beverage manufacturing process and product pH

Summary and Key Takeaways

This memo outlines the Production Data Science team's analysis of the ABC Beverage manufacturing process towards constructing a predictive model of product *pH*. Our team's model should help the company adapt its operational processes in a changing regulatory environment. Core analysis takeaways are as follows:

- Our Generalized Boosting (GBM) model identifies *Mnf Flow*, *Oxygen Filler*, and *Usage Content* as well as *Brand Code C* as manufacturing process factors that could be important in predicting the *pH* of ABC Beverage products.
- We recommend that Production prioritizes these factors when updating the company's manufacturing processes.
- We recommend refining the model iteratively as new process data becomes available.
- Lastly, we recommend follow-up split testing to identify actual *causal* relationships between the factors we identified and impacts on pH.

Motivation

New regulations have prompted ABC Beverage to understand better the relationship between its manufacturing process and its products' *pH*. Executive leadership tasked Production with analyzing the company's manufacturing process and the related factors that help predict *pH*. Our Production Data Science team applied machine learning approaches to historical company data to construct a predictive model of *pH*.

Methodology and Model

Our analysis began with exploring the dataset via descriptive statistics. This exploration guided preparation of the data ahead of modeling. We then constructed eight different models, each one employing a different approach to predicting pH. Ultimately, using common evaluation criteria, we selected GBM, which is known for performing well across prediction tasks. Our model identifies *Mnf Flow*, *Oxygen Filler*, and *Usage Content* as manufacturing process factors that are particularly important in predicting our products' *pH*. It also identifies *Brand Code C* as an important category.

Concerns

We note that: 1) the relatively small size of the dataset is sufficient but not ideal for modeling, and 2) the dataset suffers from strong between-factor relationships that could adversely affect model performance. Nevertheless, we are confident that our model accounts for these concerns to the extent possible.

Recommendations

Our model identified correlations between specific manufacturing steps and pH; however, we cannot say from a historic review that the relationship is causal. Per our model, we recommend that Production prioritizes *Mnf Flow*, *Oxygen Filler*, and *Usage Content* - with attention to *Brand Code C* - as manufacturing process factors that are important in predicting *pH*. However, it would be ideal if Production could plan follow-up split testing to identify if the relationships we found are causal, meaning changes to those steps actually lead to changes in *pH*. There is a chance we identified factors that appear related to *pH*, but in fact, won't change *pH* if adjusted.

If *pH* monitoring and calibration will be an ongoing regulatory need, we recommend setting up a nightly or weekly process to retrain this model with new data. Should model performance decrease more than a critical threshold, this automated job should send alerts to key personnel who can revisit model setup.