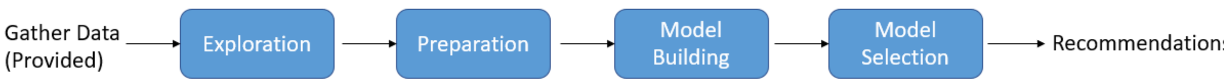
ABC BEVERAGE NON-TECHNICAL REPORT

Background

The purpose of this project was to explore and provide insight regarding the beverage-making process. To ultimately derive a predictive model for our beverages pH level and provide guidance as to how we might best control and manipulate beverage pH levels.

To arrive at the recommendations provided in this report, we took the following approach:



We explored and prepared the data we were provided to consider a broader array of models. Of the nine models we considered, we selected the best performing, easiest-to-use model and proceeded to cast predictions and determine most impactful variables (on pH levels).

Executive Summary

The sections that follow are succinctly summarized below:

1. We've elected a Cubist model for its predictive performance and ease of use.
2. We'd recommend controlling beverage pH levels by varying Density, Alch.Rel, Air.Pressurer, Temperature, and/or Carb.Rel for different levels of Minimum Night Flow.

Predictive Model

We considered nine different models of linear and non-linear regression type in addition to those of a tree-based type. Below, we've highlighted the performance statistics for our strongest linear, non-linear, and tree-based models:

	RMSE	Rsquared	MAE
Partial Least Squares	0.1437415	0.3499468	0.1102135
Support Vector Machine	0.1280496	0.4854961	0.0921707
Cubist	0.1032232	0.6655173	0.0753867

The Cubist (tree-based) model *by far* outperforms the strongest linear and nonlinear models.

We selected the Cubist model first based on performance and second based on ease-of-use (setup). The Cubist model performed the best on test data with regard to R-squared and mean absolute error (MAE), requires less data pre-processing than our linear and nonlinear regression models and provides the best "out of the box" accuracy (no variable tweaking).

Cubist models are a powerful, rule-based model that balance the call for predictive accuracy with that of intelligibility. For this reason we're confident in its selection.

Recommendations

In order to provide recommendations, we first consider variable usage within our model:

```
Attribute usage:
  Conds  Model

   86%    53%    Mnf.Flow
   43%    76%    Alch.Rel
   33%    55%    Pressure.Vacuum
   29%    83%    Density
   24%    67%    Temperature
   22%    43%    Filler.Level
   20%    59%    Carb.Rel
   19%    42%    Usage.cont
   18%    48%    Oxygen.Filler
   13%    48%    Carb.Flow
   11%    25%    Filler.Speed
    8%    69%    Air.Pressurer
    4%    60%    Hyd.Pressure2
    2%    56%    Carb.Pressure1
    2%    24%    Fill.Pressure
    1%    25%    Pressure.Setpoint
    1%    27%    Carb.Volume
          13%    PSC
          16%    MFR
          36%    Carb.Temp
          23%    Hyd.Pressure4
          25%    PC.Volume
          35%    Carb.Pressure
           8%    Fill.Ounces
           8%    PSC.Fill
           7%    PSC.CO2
```

We see two columns above, one for conditions and one for the model itself. While some variables were important in making rules (for downstream variables), others were important for the model and predictive accuracy itself. From this we extend:

- Minimum Night Flow is important but should primarily be used as a threshold / flag variable. The Cubist model used Mnf. Flow in 86% of rule criteria. We could use flows above/below a certain threshold to help us predict pH levels.
- When we next consider the three most impactful variables, outside of rule criteria, we see that Density (83%), Alch.Rel (76%), Air.Pressurer (69%), Temperature (67%), and Carb.Rel (59%) all play roles in accurately predicting pH levels.

Controlling PH levels would thus be done by varying Density, Alch.Rel, Air.Pressurer, Temperature, and/or Carb.Rel at different levels of Minimum Night Flow.