**Idea 1: Model Clarification Efficiency of Beer Recipes**

Goal: Find what features contribute to clarification efficiency in beer.

Dataset: This will come from historical clarification data of over 200 beer recipes in a production facility. I will need to compile this data from the PostgreSQL production database.

Outline: Targeting the final packaged volume of beer in a brewery production facility is critical to delivering the volume requested by sales while also not overshooting volume that results in wasted product. For new brands and beer recipes, the approach is to essentially go by beer style and hope for the best. This model would look at clarification efficiencies across more than 200 beer recipes, some of which have been brewed dozens of times. This would include beer recipe features such as fraction of base malt, lbs/BBL of hot side and dry hop, and use of any clarification agents. Process features would include rate of clarification, fermenter temperature at time of clarification, and time between start of chilling in the fermenter and clarification.

**Problem Statement Formation**

**Context**

**Criteria for Success**

**Scope of Solution Space**

**Constraints**

**Stakeholders**

**Data Sources**

[Data Source 1](https://github.com/trentleslie/Springboard/blob/main/7.1.2%20Capstone%20Two%20-%20Project%20Proposal/clar_eff_by_batch.txt): Clarification Efficiencies by Batch (1,333 records, many of which are the same **Recipe ID** multiple times). Also includes some process and batch-specific recipe parameters.

[Data Source 2](https://github.com/trentleslie/Springboard/blob/main/7.1.2%20Capstone%20Two%20-%20Project%20Proposal/rcp_malt.txt): Malt Recipe (299 unique **Recipe IDs**). This provides a percentage of base malt for each recipe. Many times, a lower percentage of base malt indicates a wheat malt – or higher protein content – malt has a relatively large fraction of the malt bill. This could contribute to clarification efficiencies.

[Data Source 3](https://github.com/trentleslie/Springboard/blob/main/7.1.2%20Capstone%20Two%20-%20Project%20Proposal/rcp_additives.txt): Additive Recipes (646 records indicating how much hops, Fruit and Honey, and/or Sugar & Syrups are added at various process steps - or locations - of the beer **Recipe IDs**). The focus of this will be hops (a type) added in the boil kettle, whirlpool, and fermenter (locations).

**Hypothesis:** *Create a Hypothesis with an emphasis on SMART principles.* ***(S – Specific, M – Measurable, A – Achievable, R – Realistic, T – Timebound).*** If you cannot do this, you **do not** have a good grasp on the business problem.

**Context:** With context, we have **clearly identified the problem at hand** and have elucidated on how our initiative may solve this problem, alongside the commercial implications this will have on the business.

**Criteria for Success**: Clearly defining the criteria for success ensures that the scope of your work is clearly defined and understood. Otherwise, if this isn’t defined – your work will never end which will result in mismatched expectations.

**Scope of Solution Space:** Scoping out the solution space ensures that the business initiative is SPECIFIC for a certain segment or area. This prevents solutions that have been developed being scaled and applied for all other business units that the solution may not be responsible or scalable for.

**Constraints within Solution Space:** Looking forward, what are the foreseeable problems we are likely to encounter? Could this be stakeholder resistance? Could this be we don’t have access to the right data?

**Stakeholders to provide key insight:** Who are the people I need to speak to, to get the answers I need for my data analysis?

**What key data sources are required**?

Based off my discussions with the key stakeholders – can we clearly list out all the data sources we need so we can make a highly targeted request as opposed to a scatter-gun approach where we ask for a bit of everything?