## **Dataset and motivation slide (1 points)**

### - *How/why the dataset was collected and a description of the metadata of your dataset.*

The hop plant Humulus Lupulus is a member of the Cannabaceae family of flowering plants. The flower of this plant is called a hop, and can also be referred to as seed cones or strobiles. Hops are primarily used as a bittering, flavoring, and stabilizing agent in beer. In recent years, hops have become the center of attention in the beer industry, as hop-forward beers have become one of the most popular styles of beer. Hops varieties are developed and grown in moderate climates around the world. Every branded hop variety has a unique flavor and aroma profile.

This makes for an exciting and delicious reason to explore what flavor and aroma a hop can offer on its own, or together with other hops to bring waves and layers of of flavor and aromas. The flavors and aromas can be reminiscent of a wide variety of fruits, florals, citrus, and more. Sometimes with multiple waves and layers of flavors and aromas depending on the malts and hops used in the beer.

As brand positioning is especially important in this industry, a study that can discern hops geographically could potentially help in the decision-making process for any microbrewery aiming to craft distinct flavors with particular hops, to set itself apart.

Having several years of experience in the brewing industry, hop conversations are frequent and exciting. Every hop variery

## **Actual task definition/research question (2 points)**

### *- What real-world problem are you trying to solve?*

Are the aroma and flavor of hops from the same region/continent more similar to each other more so than hops from other regions/continents? What is the relationship between the brewing values and oil concentrations of hops and the region/continent they are from?

### *- What are the input and output of your analysis?*

Need to answer.

## **Literature review (2 points)**

### *What other work has been done in this area, and how is your work novel compared to others?*

While there is some research on hops, much of it is focused on breeding, genetics, and chemistry. In an attempt to find similar work, we came across this Github repository. This person scraped data on hops from sources different than the one used for this project, and created some cool visuals.

<https://github.com/vieuxsinge/hops-datasets>

In these two references, researchers used classification techniques of hop photos:

<https://www.kaggle.com/scruggzilla/hops-classification>

<https://www.researchgate.net/publication/354093321_Dataset_for_Hop_Varieties_Classification>

Here are two notable institutes that carry out research on beer hops:

<https://en.wikipedia.org/wiki/Hop_Research_Center_H%C3%BCll>

<https://www.hopresearchcouncil.org/>

## **Quality of cleaning (6 points, 2 points each)**

### *- Data cleaning and type conversion activity. Please share anything unusual you faced during this activity.*

Could we say something here?

### *- What did you do about missing values and why? Handling missing values properly is very important.*

Some brew values and oil concentrations are missing from some hops. For the purpose of this project, hops with missing values were dropped from the data frame.

### *- New feature/attribute creation and data summary statistics and interpretation.*

Flavor and aroma tags were scraped as one attribute. Instead of listing them in the cell, they were converted to new Boolean attributes. Each flavor/aroma tag is now represented by a Boolean column.

## **Visualization (8 points, 2 points each)**

### - *Data visualization activity (box plot, bar plot, violin plot, and pairplot to see relationships and distribution, etc.).*

Please see code.

### *- Describe anything you find in the data after each visualization.*

Please see code.

### *- What data visualization helped you understand about data distribution.*

Need to answer.

### *- What you did about possible outlier as per data distribution visualization. (Did you confirm with your client whether it is actually an outlier or put a disclosure statement in your notebook if you decided to remove it?)*

Need to answer.