## **Introduction / Motivation**

Beer is one of the oldest and most preferred beverages by humans on earth. Although it has been brewed and consumed for centuries, one of it’s main ingredients has changed through the years. It wasn’t until the 16th century when hops became more widely used and replaced the use of gruit, an herb mixture that was used to bitter and flavor beer. 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 cone or strobile. Hops are used as a bittering, flavoring, and stabilizing agent in beer. They consist of oils and other compounds that are largely responsible for the unique flavors and aromas present in each hop variety.

## **Relevant Info**

The standards for aroma identification were initially formed in 2011 by the BarthHass Group (BHG) with the help of Frank Rittler, a world-renowned and experienced perfumer based in Düsseldorf, Germany. In 2018, the BarthHass Group (BHG) finally introduced this set of standards with the name HOPSESSED®, based on the initial 12 aroma categories noted by Rittler.

In order to measure the oil composition in a hop, a gas chromatography is performed in a lab. The following are four of the most prominent oils in hops.

**Myrcene (β-myrcene)** – This oil is mostly associated with citrusy and resinous-pine aromas. It is the most abundant of all the oils in hops and is the most potent since it has the lowest odor threshold (13 ppb). Hops with high myrcene include Amarillo, Citra, Simcoe and Cascade. The evaporation temperature of myrcene is 147 °F (63.9 °C) (this is important attribute that relates to the brewing process).

**Humulene (α-humulene and α-caryophyllene)** – This oil is mostly associated with woody and pine aromas; however, hops with higher amounts tend to be more floral, herbal, and black pepper in character. Hops with high caryophyllene percentage include Vanguard, Perle (GR) and East Kent Goldings. The evaporation temperature of humulene is 210 °F (99 °C).

**Caryophyllene (β-caryophyllene)** – This oil is mostly associated with black pepper, spiciness, and herbal aromas. The evaporation temperature of caryophyllene is 262 °F (129 °C).

**Farnesene (α-farnesene and β-farnesene)** – This oil is mostly associated with floral aromas, and slightly associate with notes of woods and citrus. Hops high in farnesene include Tettnanger, Sterling and Saaz. The evaporation temperature of farnesene is 203-257 °F (95-125 °C).

The 12 categories of HOPSESSED® are described below.

Source: <https://beermaverick.com/the-science-behind-identifying-hop-aromas/>

1. **Floral**

**Aromas**: elderflower, chamomile blossom, lily of the valley, jasmine, apple blossom, rose, geranium, carnation, lily, lilac, lavender, osmanthus.

Compounds Responsible: rose oxide, geraniol, geraniol acetate, citronellol, neral

1. **Citrus**

**Aromas**: grapefruit, orange, lime, lemon, bergamot, lemon grass, ginger, tangerine, pomelo

Compounds Responsible: alpha-terpineol, limonene, linalool, citral, decanal

1. **Tropical/Sweet Fruits**

Aromas: banana, watermelon, honeydew melon, peach, apricot, passion fruit, lychee, dried fruit, plum, pineapple, cherry, kiwi, mango, guava

Compounds Responsible: 2-methylpropyl hexanoate, ethyl 2-methylpropanoate, sec-amyl acetate, ethyl caproante, ethyl 3-methylbutanoate

1. **Stone/Green Fruits**

Aromas: pear, apple, quince, gooseberry, white wine grapes

Compounds Responsible: decanal, cis-3-dexenal, d-3-carene, 2-dodecanone, hexyl 2-methyl-propanoate

1. **Berries & Currant**

Aromas: cassis, blueberry, raspberry, blackberry, strawberry, red currant, black currant, wild strawberry, cranberry, mulberry

Compounds Responsible: beta ionone, 4-mercapto-4-methylpentan-2-one, ethyl 3-methylbutanoate, raspberry ketone, p-metha-8-thiol-3-one

1. **Cream & Caramel**

Aromas: butter, chocolate, yogurt, honey, cream, caramel, toffee, coffee, tonka bean, vanilla, coconut

Compounds Responsible: methyl decanoate, Y-nonalactone, vanillin, phenylacetic acid

1. **Woody Aromatic**

Aromas: tobacco, cognac, barrique, leather, woodruff, incense, myrrh, resin, cedar, pine, earthy

Compounds Responsible: humulene, alpha-pinene, beta-pinene, farnesene, carvacrol, beta-caryophyllene

1. **Menthol**

Aromas: mint, lemon balm, sage, camphor, menthol, wine yeast, eucalyptus

Compounds Responsible: carvone, terpinen-4-ol, camphene

1. **Herbal**

Aromas: marjoram, tarragon, dill, parsley, basil, fennel, cilantro, rosemary, thyme, green tea, black tea, mate tea, oregano

Compounds Responsible: myrcene, humulene, epoxide, p-cymene, cis-b-ocimene, thymol

1. **Spicy**

Aromas: lovage, pepper, chili, curry, juniper, aniseed, licorice, fennel seeds, clove, cinnamon, gingerbread, coriander seeds, nutmeg

Compounds Responsible: beta-caryophyllene, eugenol, 2-isopropyl-3-methoxypyrazine, beta-eudesmol

1. **Grassy**

Aromas: fresh cut grass, hay, tomato leaves, green pepper, nettle, cucumber, bamboo leaves

Compounds Responsible: E, Z-2, 6-nonadienal, cis-3-hexenol, trans-2-hexenal

1. **Vegetal**

Aromas: celery, leek, onion, artichoke, garlic, wild garlic, radish

Compounds Responsible: diallyl sulphide, dimethyl disulfide, s-methylthiohexanoate

## **Research Question**

Having worked in the brewing industry for several years, one finds themselves in conversations about the aromas and flavors of different styles of beer from around the world. Beers with hops from New Zealand have a unique profile that seems very different than hops from Europe. There’s something about South African hops profiles that aren’t found in hops from the United States. This leads to the questions:

* *What is the relationship between the brewing values and oil concentrations of hops from the same region of origin?*
* *Do hops from the same region have more similarities between them than hops from other regions?*

## **Literature Review**

While there is some research on hops, much of it is focused on breeding, genetics, and chemistry. In the research to find similar work, this Github repository was discovered. The hop data were scraped from sources different than the one used for this project, and a couple extensive visuals are presented.

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

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

* <https://www.kaggle.com/scruggzilla/hops-classification>
* <https://www.researchgate.net/publication/354093321_Dataset_for_Hop_Varieties_Classification>

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

* <https://en.wikipedia.org/wiki/Hop_Research_Center_H%C3%BCll>
* <https://www.hopresearchcouncil.org/>
* <https://www.barthhaas.com/en/world-of-flavor/hopsessed>

## **Dataset**

The source of the data: <https://beermaverick.com/hops/>

Beer Maverick has complied the largest and most extensive database of current beer hop varieties. Much of this data were sourced from hop farms, hop breeders, and hop sellers.

Unfortunately, many times these sources provided information that was different or blatantly contradictory for the same hop. In many cases (like for alpha acids, oils) we simply expanded the range of that particular value we provide to you. This should create the most accurate average found anywhere online.

## **Data Scraping**

Please refer to file: step1\_scraper.ipynb.

The data were web-scrape directly from Beer Maverick’s website using the library **BeautifulSoup**. Once it is confirmed with **urllib.robotparser** that the website would allow the scrape, a connection is establish using the **requests** library.

1. The hop names are scraped into a list from: <https://beermaverick.com/hops/>
2. This is used to build a link for each individual hop. For example: <https://beermaverick.com/hop/citra/>
3. From there, there are three points of interest:
   1. Hop profile table – first table on the page
   2. Flavor & aroma tags – center of the page
   3. Brewing values – table at the bottom of the page

A scraper generator function is created to handle the scrape that will yield a dictionary with data for each hop. Each dictionary represents a row in the dataframe. The dataframe is created from all the dictionaries using **pandas** library and saved to a csv file (raw\_hops\_main.csv). Some reference materials containing meta-data were scraped and saved into csv files. They are described below:

* The Hop Substitution table from this page: <https://beermaverick.com/hops/hop-substitutions-chart/>, and saved as csv (raw\_ref\_hops\_substitutions.csv)
* Flavor & aroma meta-data scraped from this page: <https://beermaverick.com/the-science-behind-identifying-hop-aromas/>, and saved as csv (raw\_ref\_aroma\_types.csv)
* Brewing values meta-data scrape from this page: <https://beermaverick.com/hop/newport/>, and saved as csv (raw\_ref\_brew\_values.csv)

## **Data Cleaning**

Please refer to file: step2\_cleaner.ipynb

Use of **pandas** dataframe methods and processes to clean the raw data that were scraped.

raw\_hops\_main.csv

* Removed columns: 'Unnamed: 0', 'Scraping Status', 'Cultivar/Brand ID:', 'International Code:', 'Ownership:', 'Hop Storage Index (HSI)'
* Cleaned characters of no use.
* ‘Hope Name’ as the index
* Make ‘Country’ and ‘Purpose’ categorical data.
* The brewing value columns have min/max/avg in there. Created new columns for each of those values.
* Fill empty values with np.nan from **Numpy** library
  + Cleaned file: cln\_hops\_brewvalues.csv
* Convert the column ‘Flavor & Aroma Profile’, which consists of lists of tags, to Boolean columns for each tag.
  + Cleaned file: cln\_hops\_profile.csv

raw\_ref\_hops\_substitutions.csv

* ‘Hop Name’ made into index.
* ‘Substitutions’ converted from comma separated strings, to list of strings.
* Cleaned file: cln\_ref\_hops\_substitutions.csv

raw\_ref\_aroma\_types.csv

* ‘Aroma Type’ made into index.
* ‘Aromas’ and ‘Compounds’ converted from comma separated strings, to list of strings.
* Cleaned file: cln\_ref\_aroma\_types.csv

raw\_ref\_brew\_values.csv

* Cleaned characters of no use.
* Cleaned file: cln\_ref\_brew\_values.csv

Missing values in the data were changed to np.nan values. These rows will be dropped for visualization and analysis.

## **Exploratory Data Analysis**

It is important to explore and understand the data prior to the implementation of analytics. A few functions that can help with this are df.info() and df.describe().

* df.info() returns general information about the dataframe size, column datatype, etc.
* df.describe() performs basic analysis and returns the count, mean, std, min, max, and quartiles of each column.

The rest of the exploratory analysis consists of visualizations. Their descriptions, presentations, and conclusions are presented below.

## **Data Visualization­­­­**

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

- 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?)

Box and Whisker Plot

Violin Plot

Pair Plot

Heatmap

Hops per Country with Purpose

Most Popular Flavor & Aroma Tags

Most Popular Flavor & Aroma Tags per Region – N.America, Europe, AUS, Africa

World Map – Most Popular Aromas per Country

Circular, Stacked Bar Plot of Aromas per Hop

Aroma Intensity per Hop

**Conclusions**