**Codebook for Beers and Breweries midterm Project**

Datasets Overview

The two Dataset Names used for this project are:

* The Beers dataset which contains a list of 2410 US craft beers
* Breweries dataset which contains 558 US breweries.

Beer Dataset Variable Information

1. Variable: Name

- Description: Name of the Beer

- Data Type: Character

- Missing Values: None

2. Variable: Beer\_ID

- Description: Unique Identifier of the beer

- Data Type: Numeric

- Missing Values: None

3. Variable: ABV

- Description: Alcohol volume of the beer

- Measurement: Continuous

- Data Type: Numeric

- Range: 0.0001 to 0.51

- Missing Values: Yes (Blanks replaced with median values for each State)

4. Variable: IBU

- Description: International Bitterness Units of the beer

- Measurement: Continuous

- Data Type: Numeric

- Range: 4 to 138

- Missing Values: Yes (Blanks replaced with median values for each State)

5. Variable: Brewery\_ID

- Description: Brewery ID associated with the beer

- Range: 4 to 138

- Missing Values: None

6. Variable: Style

- Description: Style of Beer

- Measurement: Discrete

- Data Type: Factor

- Missing Values: Yes (Blanks replaced with style values found online)

7. Variable: Ounces

- Description: International Bitterness Units of the beer

- Measurement: Discrete

- Data Type: Numeric

- Range: 8.4 to 32

- Missing Values: None

Brewery Dataset Variable Information

1. Variable: Brew\_ID

- Description: Unique Identifier of the brewery

- Data Type: Numeric

- Missing Values: None

2. Variable: Name

- Description: Name of the brewery

- Data Type: Character

- Missing Values: None

3. Variable: City

- Description: City where brewery is located.

- Data Type: Character

- Missing Values: None

3. Variable: State

- Description: US State where brewery is located.

- Data Type: Character

- Missing Values: None

\*\* Data Cleaning and Transformation \*\*

- Missing Values: Following missing values were found:

Beer Dataset: IBU (1005 records), ABV (62 records), Style (4 records)

- Missing values treatment:

We do not know why there are gaps in the data for the IBU and ABV. So, we assumed that the data was missing at random (MAR) and would be treated as such.

Missing data imputation - first by seeking open-source information on the values and secondly by taking the median of each state and replacing the remaining missing values with those median values by State.

- Outliers: No Significant outliers were detected in either dataset

- Data Transformation: No data transformation has been applied to the original data.

\*\* Additional Notes \*\*

Additional Datasets were sourced online and were used to supplement and enhance our analysis. These datasets are optional and not an initial requirement for the analysis.

US\_locations\_Data – Used for our geographical map visualization to enhance the story telling

US Cities Data – Sourced online from <https://simplemaps.com/data/us-cities> and used in the geographical map visualization as well.

Beer Consumption by State – Used to provide further insights on how much consumption by state

- Codebook created by: Emmanuel Opoku and Adam Erganbrack

- Date: October 25, 2023