

Project Proposal: Tapping into the Data

A Pint-by-Pint Analysis of Craft Beer Styles by Characteristics and Region

Aidan Sawyer

October 2, 2016

1. Introduction

In light of the craft beer explosion, there are a number of guides and resources available about how best to brew, classify, and describe beers and beer styles. Moreover, with the huge supply of beers and styles out there, it would be helpful to have a quantitative guide to what you might like.

I intend to investigate the correlation between beer styles, location, style, ABV, IBU, Original Final and Gravity, and Standard Reference Method.

I'd like to see whether a systematic analysis of actual beer statistics would breed results that would align with the published style guides. I'd like, if possible, to construct 1) a style guide capable of grouping and relating styles on quantitative values, and 2) guessing the style of a beer by its measureable attributes.

2. Focus

Instead of focusing on the beers themselves, I want to use them as a representative sample to look at the styles. I think that this will allow me to minimize errors and get more meaningful and useful data about the entire craft of brewing beer instead of a selective view of particular breweries or regions of the US.

3. Challenges

This project will no doubt be affected by the recurring problems of missing data, unformatted data, and duplication. Additional challenges will include coordinating and integrating datasets from distinct sources, determining “word-of-god” standard for what is taken to be the “correct” values for the classification training, and organizing/standardizing a hierarchical listing of distinct types and states into a flat and inclusive list of overall styles and regions.

4. Data Sources

There are multiple resources available from which to mine this data. RateBeer, Untappd, BreweryDB, and others have databases of rich data accessible through APIs, but at least Untappd has a Terms of Service that directly prohibits data mining.

I've found bulk data through openbeer and craftcans that will allow me to get a representative and substantial sample for me to test. The CraftCans data in particular has 958 unique entries of beer styles.

5. Evaluation

Given these differing sources of data and the number of guides, I hope to cross-reference my findings to test for quality of representation and to coordinate any disagreements between how a beer should be classified.

6. Motivation

Craft and Homebrewing has become a booming industry in both popularity and marketshare over the past 5-10 years [<https://www.brewersassociation.org/statistics/national-beer-sales-production-data/>]. This explosion of interest has brought with it enumerate recipes and style guides as well as a wealth of data to potentially test and refine them with.

I hope that in the course of this analysis I can gain a better understanding of what goes in to making beer to use in the