Craft Beer Segmentation

Making the beers your customers want

Rory Breslin

March 2021

Business Problem

Craft beer is a rapidly growing industry that stood at $89bn in 2019 and is expected to grow at 10.4% annually (CAGR) to reach a market size of $161bn by 2027 (taking into account the impact of COVID-19 on the industry).[1] The growth in the market has also coincided with a large increase in the number breweries competing for this revenue, with the US seeing 8.9% increase in the number of breweries between 2018 and 2019 alone.[2] This expected increase in revenue and competition is making it more important than ever for craft breweries to make the right choices regarding the beer they produce – making sure it meets the needs of the craft beer consumer to boost sales and allowing them to grow their business.

For this project, I am working with a small craft brewery that is looking to do just that and expand its beer offering. The brewery has made its name to date on producing high-quality beers but has focused mainly on a small number of niche beers. To further expand its business, it is looking to produce a beer that continues to maintain the high quality that is synonymous with its brewery but that also appeals to a wider audience. For this reason, the brewery is looking to identify a beer type that is considered high-quality but also has mass appeal to allow it appease its existing customer base (who expect high quality) but also tap into the wider craft beer market. The brewery would like to have this beer on sale in-time for the peak summer sale period (July 2021). It estimates that it will take 3 months between decision on which beer to produce and being able to officially actual launch it due the various steps of the brewing and distribution process (although this could vary slightly depending on the beer type selected). With this in mind, the brewery is looking to have a recommendation on the beer to produce by the 23rd March.

Data

The analysis and modelling of this solution will based on customer reviews left on the BeerAdvocate website (and available on Kaggle) as well as additional information about the beers and breweries (also available on Kaggle). These dataset contains information relating to the brewery, the beer (name, style, alcohol level), the reviewer, and the review scores (overall, aroma, taste, appearance, palate).

Anticipated Data Science Approach

The end goal is to be able to identify a beer that the brewery can produce that is both high in quality and appeals to many craft beer enthusiasts. To do this, we will need to complete several steps:

1) Data Wrangling
    - Merge datasets together to provide as much detail as possible on the beer and breweries being reviewed
    - Manipulated and roll-up the dataset as appropriate to give us a view from reviewer, beer or brewery perspective
    - Deal with NaN values as appropriate
2) Exploratory Data Analysis (EDA)
    - Review dataset to understand the quality (review score) metric – how it varies by categorical variables and the relationship it has with numerical variables
    - Apply a clustering algorithm (kMeans or Mean Shift) to the dataset to segment reviewers based on the beers they rate and how often they rate them

---

[1] "Craft Beer Market Size, Share & COVID-19 Impact Analysis, By Type (Ale, Lager, Pilsner, and Others), Distribution Channel (On-Trade and Off-Trade), and Regional Forecast, 2020-2027". Fortune Business Insights. https://www.fortunebusinessinsights.com/industry-reports/craft-beer-market-100736

[2] "National Beer Sales and Production Data". Brewers Association. https://www.brewersassociation.org/statistics-and-data/national-beer-stats/

3) Pre-Processing and Feature Engineering
    - Remove or tailor variables that may create dimensionality issues, and create dummy variables for any categorical variables that we will look to include
    - Identify your target variable – depending on the result of your cluster analysis, this could be a customer segment or metric developed to capture both quality and review share
    - Split model between training and test set
    - Scale and engineer your features to ensure only most relevant features included in model
4) Modelling
    - Assumption is that this will be a classification problem where we want to identify if a beer type matches with a particular customer segment
    - Based on this, we will apply several different types of classification models to understand which performs best. This will include a combination of Logistic Regression, SGD Classifier, LinearSVC, Gradient Boosting, and Random Forest models.

Deliverables

At the end of this project, I aim to deliver the following:

- Datasets used for analysis and modelling
- Juypter notebooks outlining the code for each stage of the project (Data Wrangling, EDA, Pre-processing and Feature Engineering, Modelling)
- Final report – detailing process for each stage of the project
- Presentation – recommending the beer the brewery should produce based on the analysis and modelling undertaken
- Dashboard of interesting visualisations