Lifestyle Investments, the investment arm of Premium Venture Capital, contracted Century Data Solutions (CDS) to assist them with a conducting preliminary data gathering and analysis for one of their planned investment funds (Guilty Pleasures Fund).

The Guilty Pleasures Investment Fund’s (GPIF) strategy is focused on investing in companies/firms that cater to what are often considered ‘guilty pleasures’. A guilty pleasure is something that one enjoys despite feeling that it is not generally held in high regard by society. Guilty pleasure can be alcoholic beverage consumption, cigarette smoking, gambling, foods (i.e. chocolate, sweets, sugary beverages, ..), clothing, etc. It is known that individuals will spend money on their guilty pleasure even when the economy and personal finances do not support these non-essential purchases. This makes this market a very stable and robust market to invest in.

GPIF currently is focused on investing in the alcohol beverage market within the United States. While two major international conglomerates recently joined forces AB InBev - SAB Miller assuming control of nearly 47% of the beer sales in the US. AB InBev's sales have fallen in the US as consumers shift away from American lagers (Budweiser, Bud Light, Miller) toward smaller craft beers. GPIF’s strategy is to invest in existing breweries and/or invest in new breweries located in underserved markets, pulling market share away from the mega breweries and capitalizing on the shift to more boutique brews.

The purpose of this consultancy agreement was to gather, clean and summarize data sets that was presented to CDS by Lifestyle Investments, which contained Brewery and Beer data from US breweries, to assist GPIF in developing strategies and market plans focused on investing in craft breweries within the US.

GPIF presented CDS with 2 initial data sets relating to the is project. The original source of the datasets is unknown. The original dataset was obtained by GPIF from Openbeerdb.com <https://openbeerdb.com/> . The databases were made available under the Open Database License agreement. Note: The original databases were altered by GPIF prior to submission to CDS.

Beers dataset:

The Beers dataset contains a list of 2410 US craft beers currently being produced in the US, along with other pertinent data (Beer ID, Alcohol content (ABV), International Bitterness Units (IBU), Brewery ID, Style, and Ounces).

Breweries dataset:

The Breweries dataset contains a list of 558 craft beer Breweries in the US, along with other pertinent data (Brew\_ID, City and State of headquarters).

The project CODEBOOK contains a more detailed descriptive listing of the variables used.

Project: MicBrew

For reproducibility across a wide variety of computer operating platforms, CDS chose R as the preferred statistical software environment to gather, clean and summarize data sets that was presented to CDS by Lifestyle Investments.

The datasets were received in csv format. A preliminary review of the files showed that the data was well organized and rather complete.

First, we used the Breweries dataset and developed a summary table listing the number of breweries in each state. As seen in the Table 1.1 Breweries by State, there is a wide range of number of breweries by state. Colorado leads the way with 47 breweries with California a distant 2nd with 39. Several states only had 1 brewery.

Visualization of the data introduces another dimension to the data. Figure 1.1 Breweries by State is a color-coded map, displaying the number of breweries in each state by varying shades of coloring. As you can see from the map, the West Coast, and Northeast US, have the most craft breweries, with Colorado and Texas being outliers in the Western and Southern regions. Further breakdowns, by US regions are available if requested.

CDS merged the datasets together, to get a more robust dataset, which includes information about brewery, location, beers produced, alcohol content (ABV), bitterness (IBU), style, size (ounces). For quality control purposes, the first six rows (Table 1.2a) of and the last six rows (Table 1.2b) of the combined dataset are displayed below.

A check of the dataset’s robustness/completeness was conducted. Table 1.3 list each variable and the number of NA (Not Available/missing) data points is listed. Most of the variables have no missing data. ABV (alcohol content) has 62 missing values while IBU (bitterness units) has 1005 missing values. Bitterness units is not a typically used criteria in the US, which might explain the large number of missing values. Any analysis using IBU must be viewed with caution.

Consumers taste in beer vary by state-by-state and region-by-region. Some people enjoy bitter stout/IPA brews while others prefer a softer/pale brew. To investigate look for trends, CDS developed a dual bar graph comparing median alcohol content (ABV) vs median bitterness units (IBU) by state Figure 1.4. Visual inspection of the data does not appear to show any correlation between alcohol content and bitterness units. Also, there is no visual evidence to suggest that different regions with the US have a propensity to a certain level of alcohol and/or bitterness in their craft beers. Note: there were many IBU observations missing (1005) from the dataset. Please use caution when interpreting these results.

Many states regulate the maximum alcohol content in beers brewed and sold in their state. The state with the highest alcohol craft brew is Colorado [Lee Hill Series Vol.5 – Belgian Style Quadrupel Ale] with an ABV of 12.8%.

The state with the most bitter (IBU) craft brew is Oregon [Bitter Bitch Imperial IPA] with a IBU of 138.

Need to fix code to display state, Name, Brewery, ABV value

Need to fix code to display state, Name, Brewery, IBU value

The level of alcohol varies from brew to brew, but the ABV level of most commercial brews averages 4-12%, compared to 15-20% for wine, and 35-50% for vodka. A summary statistic was performed on the ABV content for all craft brews in our dataset (Table 1.6). The mean ABV is 5.97% by volume with and interquartile range of [5.0%, 6.7%].

Do we want to create a boxplot for this data?

We wanted to investigate if there is an apparent relationship between the bitterness of a beer and its alcohol content (Fig 1.7). A visual inspection of a scatter plot of ABV vs IBU does appear to have a positive correlation between IBU and ABV. Higher content alcohol craft brews tend to have higher bitterness ratings. Note: there were many IBU observations missing (1005) from the dataset. Please use caution when interpreting these results.

Need to add title to Graph

Guilty Pleasure Investment Fund consulted CDS to gather, clean and summarize data sets that were presented to CDS by Lifestyle Investments, containing Brewery and Beer data from US breweries, to assist GPIF in developing strategies and market plans focused on investing in craft breweries within the US.

Using tidy data techniques and statistical software, CDI combined, cleaned and summarize the datasets, using statistical measures, summaries, and data visualization techniques. The data analysis was only a preliminary review, to assist GPIF in identifying the next step(s) in their process of identifying and selecting possible investment strategies. The preliminary findings suggest West Coast and Northeast US tend to have a higher concentration of craft breweries, while Colorado has the most breweries for any state. With a high concentration of craft breweries and increasing population, these regions and the state of Colorado warrant additional, more in-depth investigations as a first step in finding investment grade opportunities in the craft brew market.

In RMarkdown, we need to turn off warnings so they don’t show up in the report

warning = FALSE