**\*\*NW**

**Client:**

Untappd

* Geosocial networking and mobile phone
* Allows users to “check in” to beers they drink
* Share check-ins and location with friends, see what friends are drinking, share pictures
* Rate Beer
* Earn Badges for different things:
* Review tap lists
* Make suggestions

7 million users world-wide

Merged with Next Glass in January 2016 and have begun a strategic planning process to consider ways to keep growing.

**Business Challenges:** two primary questions they want to explore –

Why?

* Craft beer is growing in popularity and is gaining more market share. With millennials preferring craft beer over domestic, we want better ways to engage with them and support local businesses.

How can we expand online engagement?

* Have a lot of great options to engage community though app, i.e. engaging friends, making recommendations, earning badges for special events like Portsmouth Beer Week or Goose Island’s 3.12 day.
* Want to explore ways to deepen engagement around specific beer styles and location

How can we expand engagement offline?

* What would those experiences look like?
* How to beginning an exploratory planning process.

**Data Overview**

* Original data sets: two files were provided by Untappd
* First Set: beer data, which included 2410 observations, 7 variables – Name of Beer, Style of beer, measure of Alcohol by Volume, the International Bittering Units, serving size, plus several internal proprietary IDs to Untappd
* Second Set: brewery data, which included 558 United States-based breweries, along with location of City and State, plus internal IDs.
* Exploratory analysis process, using R statistical software package
  + First, looked at missing data. Understand where we were missing information.
  + Most data missing was IBU. While common for beers to have, only part of analysis and we were not concerned about missing values. Possible some were 0.
  + Other missing data were beer styles – given most beers fall into a style category, would be the next area to dig into further. Since this was high level exploratory analysis, team felt comfortable with missing values and proceeded.
* Duplicate Breweries – spent time here since it would specifically tie to location and began to think location may be key to recommendations

**\*\* JM**

**Duplicate Data Analysis/Resolution**

* Set up analysis to look at combined BreweryNameCityState and determined how many strings would match with a one-character change.
* The team then looked into these as duplicates. Corrections were made to obvious errors in both formatting – i.e. St. Paul vs St Paul, spelling errors, and erroneous states. The Untappd app search function was used to confirm these.
* Opted to not do additional clean up without consulting the Untappd team, recognizing we needed local expertise to determine if remaining breweries might have multiple locations.

**Data Snapshot**

* After this cursory clean up, the team merged the two data sets to create a working file that combined the beers with the corresponding breweries. As example, show first 6 and last 6.
* Explain full data set using these as examples – matched beer, location, style, IBU, ABV

**Summary of Findings – Bar Chart and Map of Breweries by State**

* Then we proceeded to learn more about the relationships between the beers, the breweries, the alcohol by volume, the bitterness rating, and how these factors might relate to location.
* Looked at distribution of Breweries by State
* More useful, looked at geographic representation to understand distribution of beers across country.
  + Colorado, Pacific Northwest and California, New England, Midwest, and MidAtlantic regions

**\*\* CK**

**Summary of Findings – Median ABV by State**

* Then looked at Median Alcohol Content by State
* Not surprising Utah had lowest, where I believe ABV is regulated by state law
* Other regions did not show significant trends toward high or low ABV
* Most hovered around the middle range

**Summary of Findings – ABV Summary \*\*MOVE SUMMARY HERE??**

**Summary of Findings – Median IBU by State**

* Notable that South Dakota did not measure any beers with an IBU rating.
* Should work with Untappd delve into beer styles and need more complete IBU ratings, need to continue to clean
* Since IBU is often related to preference – and there is a maximum threshold point at which our taste buds cannot detect – a complete data set here was not critical to high level analysis

**Summary of Findings – Top States**

* Highest ABV in Colorado: the Lee Hill Series
* Highest IBU in Oregon: the Bitter Bitch

**Recommendations**

Online: rich understanding of preferences, breweries, recommend expanding engagement program beyond badges and recommendations and sharing to capitalize on gaming craze of Pokemon.

\*\* Jonathan – We will create a machine learning algorithm that will learn the user’s beer preferences and suggest new or favorite beers in nearby areas or during travel.

Alert will pop-up on phone as other users log in new rare beers in the area. We will also offer data commercially, for a price, for beer data around the U.S. as to what beers work vs not by location to increase revenue.

Dataset will provide beginning features in regards to alcohol and IBU content and we will continue to get more data on beers and breweries to improve ML algorithm.

Offline: important to think about how to create a similar, social experience offline. Untappd is already highlighting key events around the country, recommend taking one step further and exploring full-fledged entry into beer tourism.

Untappd would make travel arrangements, create unique programming in partnership with breweries, offering content not available to general public. Possible launch markets include Pacific Northwest, New England…especially since there are seasonal travel opportunities in fall.