Omar Zaher

Will Hughes

Japhet Mwamba

Gage Anderson

***Texas Beer Recommender Model***

**Introduction:**

Beer is one of the most popular alcoholic beverages served across the world, it can range in different varying styles, looks, and taste. There are so many beers out there it can all the different options one can consume can be very overwhelming. Not only that but one could try a random beer and realize quickly not all beers are made the same. As the craft beer industry continues to grow, consumers are faced with an overwhelming number of options when it comes to choosing a beer. This has led to the development of various beer recommendation systems that use data analytics and machine learning techniques to provide personalized beer recommendations based on the user's taste preferences. Being that we are in Texas, our group built a website that uses three different KNN models to recommend an array of Texas made beers based off of “Beer Name”, “Beer Style”, and “Alcohol Content /Bitterness”.

**Inspiration:**

* Concept: We were inspired to build this recommender model because one of our group member’s father, is going to open up a brewery here in Texas in the near future. Beer also, when consumed responsibly, is a great compliment for a good time and can be a means of loosening up and enjoying the moment for many. Another reason we were inspired to build this model is because there is such a large variety of beers with different tastes and looks, we were hoping we could create a tool that can help expose consumers to a new Texas made beer.
* Tableau: After viewing some tableau graphs, we were inspired to include a map of all the breweries. We attempted to make a beer lollipop chart but had to settle for a regular lollipop chart.

1. <https://public.tableau.com/app/profile/ladataviz/viz/BeerInEurope/BeerInEurope>
2. https://public.tableau.com/app/profile/victor.d.pr./viz/BeerConsumptionintheUS/Dashboard1

* Color Scheme:

For our color scheme we decided to use Untapped’s color scheme since it is the “twitter” of sorts for different type of beer recommendations and evaluations.

1. <https://untappd.com/>

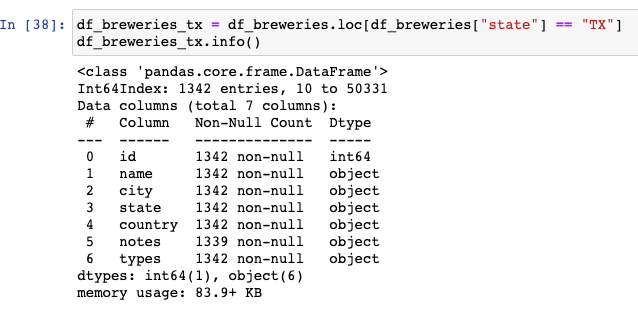
**Data Source:**

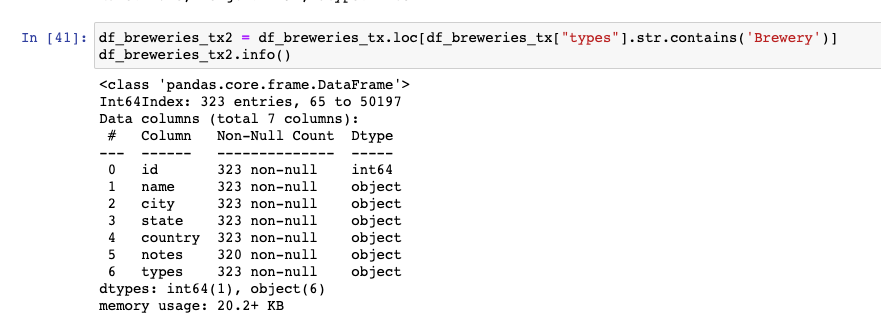
Since we were inspired by the Untappd color scheme we initially came across some information showing that Untappd, at one point, had an API that we could have used and would have been perfect for us but unfortunately, they shut the API down. Instead, we found a data set on Kaggle that included 3 different CSV’s; “breweries”, “beer”, and” reviews”. Combined, the csv’s gave us more than enough data to get us a working recommendation model. However, one of our csv’s contained over 9million rows we had to open a jupyter notebook to be able to look at our data and make the necessary data frames. Our first goal with each csv was to make a data frame of only the data in Texas that we could merge in to one data frame to use for our model.

**Data Cleaning:**

Breweries CSV-

In our initial data frame we had over 50,000 rows so we dropped all rows that were not in Texas using the “state” column bringing our rows down to 1342.

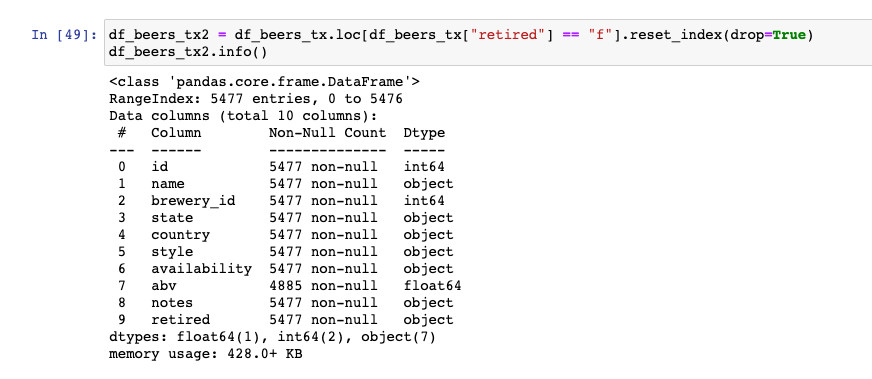
The first thing we noticed about this data frame we made was that we had a column named types with many different categories like “Brewery, Store, Gas-station, etc.” so we made sure to drop all rows that did not include a type that had the string “brewery” in it.



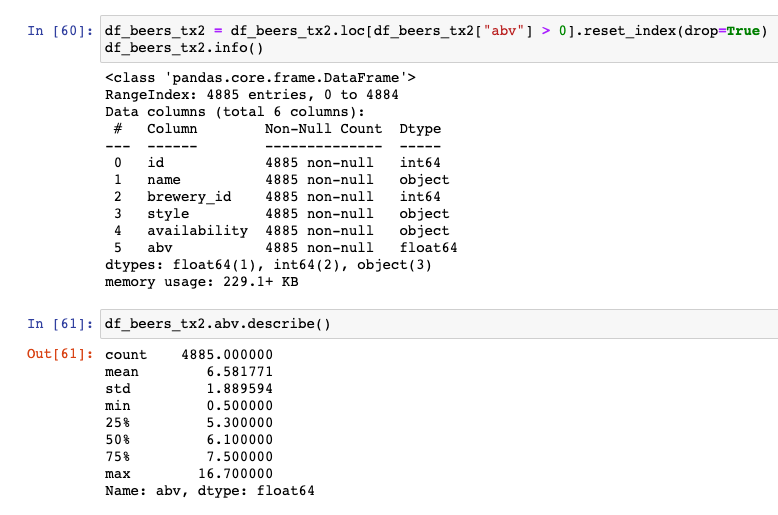
We took a closer look at our “notes” column and 296 of our rows were “no notes at this time”. We dropped the notes column and now had a clean breweries data frame.

Beers CSV-

With this data set we created a data frame of all the Texas beers but realized that there are some beers whose availability is not year-round, was only made once, or is retired. So, we continued to drop all rows where the value of retired was false so our model would not recommend a beer that is not made anymore.



Another thing that caught our eyes is that we had some beers with a 0 abv and realized this would not help with our recommendation model so we dropped all rows that had a abv less than 0. We also dropped any unnecessary columns leaving us with a clean beer data frame.



Reviews CSV- This csv had over 9 million rows so we dropped any rows not in our “beer\_ids” list. We also dropped any unnecessary columns like “notes, user\_name, date, text, and score”.

Now that we had cleaned our data, we merged our data frames in to one to use for our machine learning model.

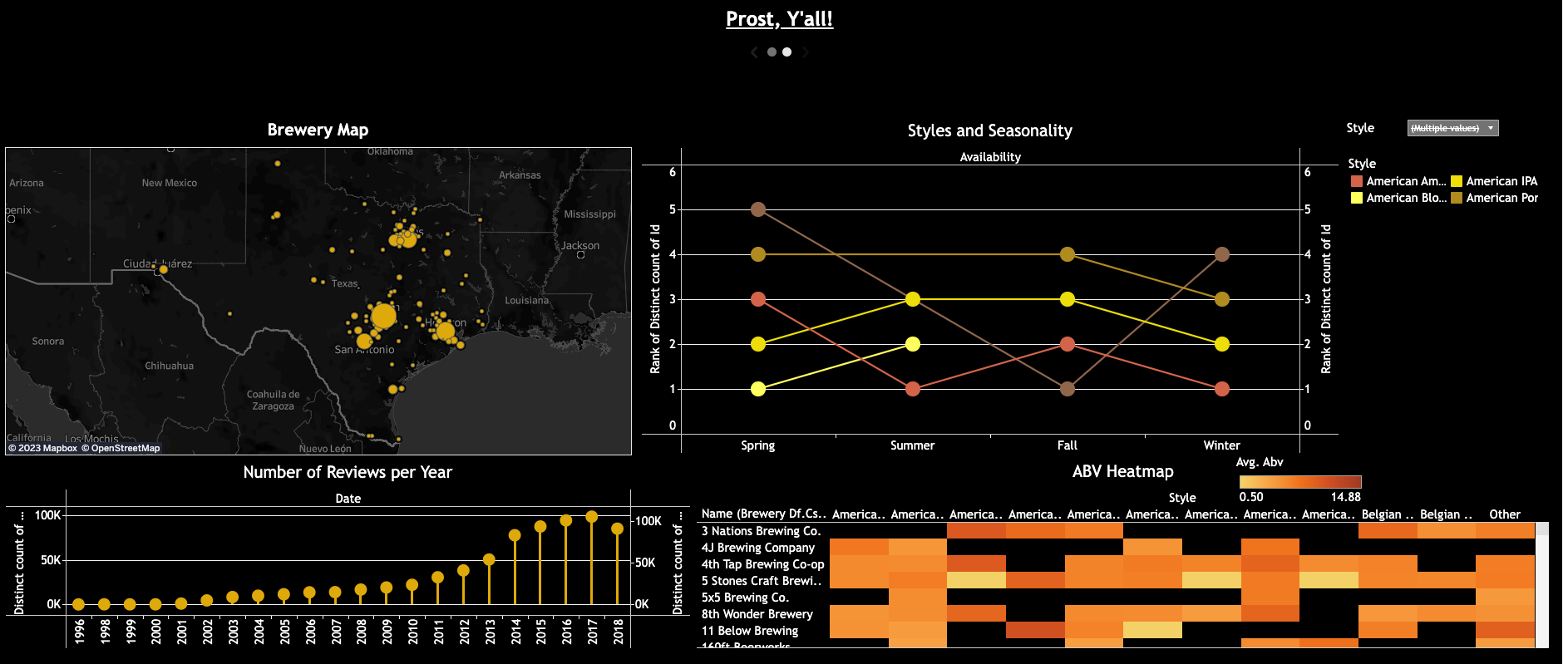
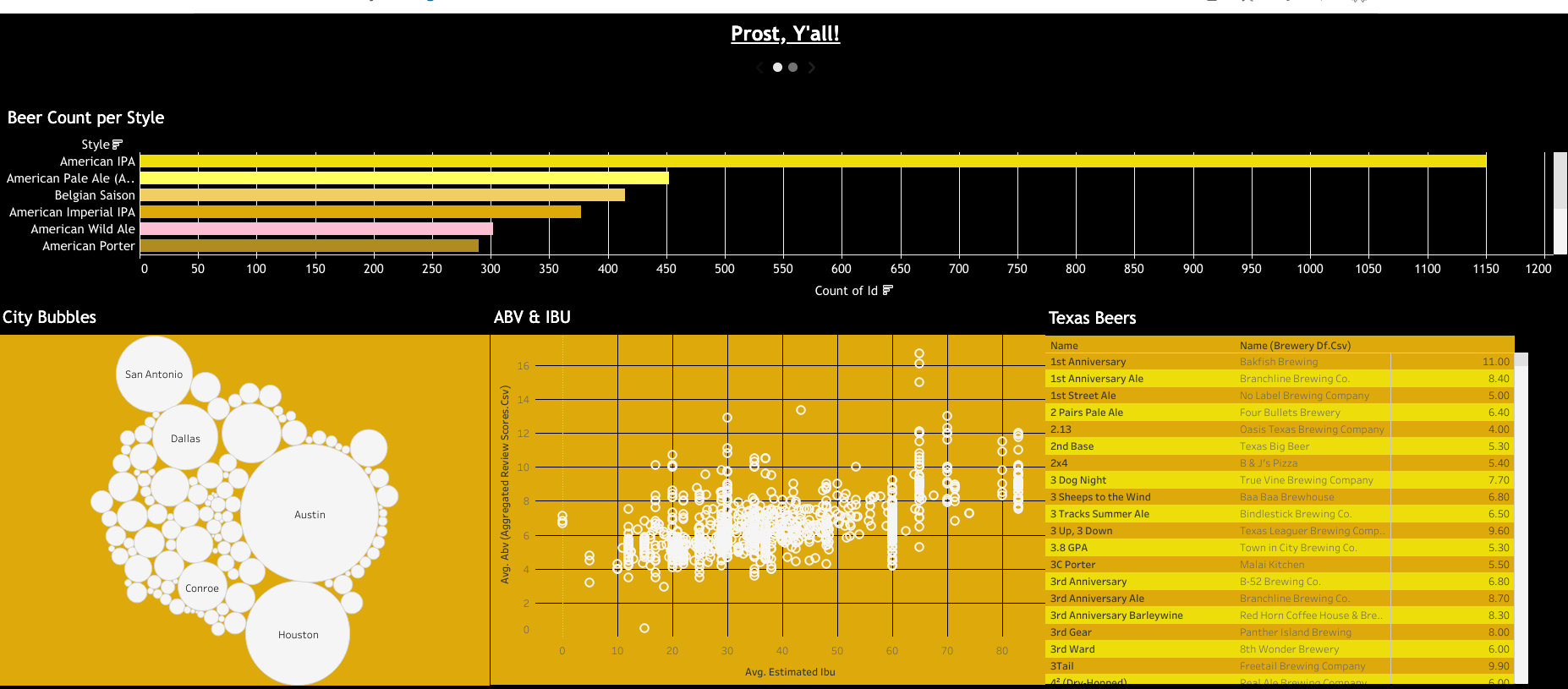
**Website Design:**

We wanted our website to be user friendly and easy enough that it can be used even after a couple beers or so. Our interactive web dashboard was built using Flask that allows users to input their preferences like beer name, style, or alcohol content/ bitterness and based off the KNN algorithm the user will then be presented with recommendations that the model believes or predicts will be the best fit based off rating. The site also displays necessary map visuals of breweries all across Texas so the user can visit their local brewery and enjoy their favorite beer or travel to try a highly rated one they’ve never tried before.

**Tableau:**

For the Dashboard design we wanted to provide several visualizations that would help the user learn about the breweries, beers, styles, ABV and IBU. We tried to make the visuals both appealing and informative. Aesthetically, we drew on inspirations from Untappd and from beer itself. One of the first things we wanted to visualize was simply which beer style was the most popular. In this visualization we selected colors for each of the styles. One of those colors had a nice warm orange color reminiscent of Community Mosaic IPA and became one of the primary colors throughout our dashboards. Two other visualizations featured circles, which, in combination with the orange background, we made to resemble the white foam on top of beer.

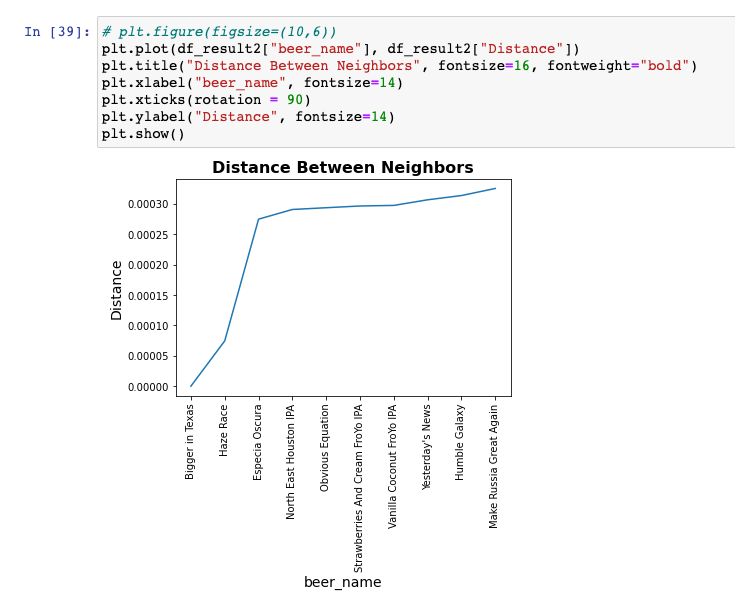
Other visuals include a Bump Chart, to demonstrate the rankings in seasonal beer styles, a Heatmap, to show the range of the ABV’s in different beer styles, a Lollipop chart showing the increase in reviews leading up to 2015, and a map, showing the density of unique breweries in Texas.



**Machine Learning:**

Since our data was cleaned and merged, we were able to call the get dummies function and one hot encode our features to use for our KNN model since it takes in numeric values to make its recommendations. The KNN is a widely used algorithm for recommendation systems the algorithm predicts the most likely rating for a new instance based on the K-Nearest Neighbor from the data set. In the case of our beer recommendation model the model for example will take in to consideration the beer name, beer style, ABV, IBU, and ratings, then based off the distance of the nearest neighbor (beer) will take an average rating/instance of the feature and make its recommendation.





**Limitations & Future Work:**

* Limitations-Some limitations that we faced are that some of the column labels weren’t exactly clear. For example, in our reviews csv we had a column named overall. We had to do some real deep digging to find out where the original csv data came from and we were able to finally find the site and based off the site description we were able to determine it was best to drop that column. Another limitation we faced is that Untappd shut down their API, that would have been an amazing resource to have for our model considering.
* Future Work- In regards to our machine learning model and algorithm we could have and plan to use a Collaborative filtering technique. Being that it would recommend a beer by identifying other users with similar taste so essentially it uses their opinions to recommend a beer to the user. We also plan to possibly look at scaling our features to see if that may help improve our model.

Resources:

* https://www.kaggle.com/datasets/ehallmar/beers-breweries-and-beer- reviews?select=breweries.csv
* https://public.tableau.com/app/profile/victor.d.pr./viz/BeerConsumptionintheUS/Dashboard1
* https://www.kaggle.com/code/ruancmoral/beer-analysis
* **https://www.kaggle.com/code/joonim/beer-reviews-asking-questions**