Fermentation and brewing beer have been a fixture of the home since ancient times as a means of preparing shelf stable food. As we have developed more technology, we have improved upon the fermentation process. One website, www.brewersfriend.com has assisted the modern homebrewer in organizing and standardizing a myriad of beer recipes that novices to masters can brew at their home. However, homebrewing beer can be considered both an art and a science, oftentimes with projects that have many complex factors beyond the brewer's control.

This project is designed to take the data generated in the brewing process (e.g. original, boil, and final gravity measurements, color, IBU, beer style, etc.) and compare them to the database of beers within www.brewersfriend.com. Using this data, the project intends to predict how well a given beer would rate against the database on a scale of one to five stars. We also intend to use this model to estimate the number of carbs and calories. This project is aimed to help the novice homebrewer to improve upon their homebrews by taking care to improve technique and data gathering while simultaneously providing information on similar styles of beers other homebrewers have made.

The starting data set that will be used is found on kaggle: https://www.kaggle.com/jtrofe/beer-recipes. This dataset contains information on 70,000 different homebrewed recipes from www.brewersfriend.com. In addition to this data, I'll use the API from brewersfriend.com to pull data on the ratings, how many reviews, calories per 12 oz, and carbs per 12 oz. This content will be merged with the initial data set and be used to build the primary model.

While the initial goal is to just use the data described above, a next step of this project would be to utilize the ingredients used in each recipe and include this data in the model as well. Using the ingredient data, further suggestions (e.g. boil times, malt compositions, hops, steeping times, etc.) could be inferred based on higher rated beers that were similar to the style of the desired beer. This data will be in addition to the previously described data set.