We (John O’Donnell and Oscar So) will be working with a “Cereals” Dataset given to us by CDS insights team during interviews and working on feature engineering. Our plan is to create multiple visuals of the data before we feature engineer it to discover any traits and correlations between the different nutrition labels. One major correlation we can focus on is calorie count vs protein. Or the amount of sugar/healthiness with manufacturers. We will try to provide some interactive visuals, possibly coded in either python or javascript. Then, we will use KNN/KMeans and also linear regression to classify some of the features. We are expecting an inverse relation between protein and calorie, an inverse relation between protein and sugar, and a direct proportional relationship between calories and sugar (carbs). We will divide the work evenly, one will work on linear regression while the other will work on the KNN side of the project. We will work together in coming up with the visual representations of our dataset before we feature engineer it to understand the dataset together and move on without leaving each other behind.