Hey Professor,

For this week, I enjoyed looking into the process of determining which machine learning algorithm you should utilize when it comes to your data and business goals as a team or as an individual. Throughout the process, I felt that it kept being iterated that we always have to ensure that our data is in the best position possible before we can build a model off of it. In my mind though, I think we would complete this step no matter what in cleaning the data and grooming it all together, by looking for missing values or values that are not consistent throughout the entire set. With looking into the hyper parameters that are involved in machine learning, it definitely took me a little bit to understand its role and the impact that it has on the predictability of any model that we choose for us. At the time, it reminds me of how we were choosing the best features for our projects and which variables will give us the best output to answer the questions we have for the project. When looking at the model evaluation steps from the discussion boards, it is interesting to see the various steps of evaluating a model to ensure that it is the best fit for your current data.

With the tutorial aspect, I appreciate seeing it all slowly come together by using the same data set and attempting to answer the question. Even though it may differ slightly since we all have different data sets, the process for the most part is similar and we have some wiggle room to apply different models or algorithms. As I continue working on the original case study, I know one worry that I have is that I will not be able to apply the same methods that were used within the tutorial that we have been working on these past few weeks such as last week. I know for the feature extraction aspect I used statistical methods that allow you to choose how many features you would want to use instead of directly selecting the best ones. For that process, I decided to choose three features and the method would determine the best three features would work the popularity feature that I am focusing on for the case study.

-Gabe