You've collected your data, cleaned it up, wrangled it into shape and explored it. Now it's time to perform some in-depth data analysis using machine learning. This step depends on you and your mentor, but here are some suggestions to get you going.

1. How do you frame your main question as a machine learning problem? Is it a supervised or unsupervised problem? If it is supervised, is it a regression or a classification?
2. What are the main features (also called independent variables or predictors) that you'll use?
3. Which machine learning technique will you use?
4. How will you evaluate the success of your machine learning technique? What metric will you use?

I believe I can frame the question of my project in a way to say, can you determine world series champions after regular season ends and playoffs begin. I can say that you should be able to narrow it down due to the variables I will be using, such as Payroll Rank, WAR , and WAR rank. These 3 seem to be about the same for the teams who have won the championship that past almost 25 years. I believe my best method of machine learning with be logical regression due to In logistic regression, the outcome (dependent variable) has only a limited number of possible values. Meaning “can you win the world series?” based on the data I have is either a yes or no question not an infinite set of possibilities as it would be if I were to use linear regression.