In order to project future performance, given this data set, it is important to first look over the data and figure out the imperfections or trends of each measurement that is being considered. The primary variable that we were told to consider is the exit velocity. We are told that system A is trusted more than system B, therefore it is preferable to use measurements from system A. However, it is important to first note the differences between the two systems, as there may be a potential skew in our results if one system is chosen over the other. By creating histograms for the overall data, as well as the data for each specific hit type, a few differences between the systems became noticeable. Primarily, system B generated lower exit velocities across all hit types, with the ground balls having the greatest difference between the two systems. While there is no overall trend for the difference between launch angle measurements, there are some differences regarding fly balls and line drives. For fly balls, the measurements for system B tend to be higher, while for line drives the measurements for system A tend to be higher.

The next issue to look at centered around the imperfect data. Many of the batted balls had missing data for one or both of the systems. I determined that the best projections, given the data that is provided, would be based off of system A, as we are told that system A is much more accurate. However, not all batted balls had measurements for system A. In order to fix this, if a measurement was missing for system A, the corresponding measurements for system B would be used. This may skew our data a bit, particularly our exit velocities, since we know that system B tends to generate lower values. As a result, projections for players with many missing values for system A may be lower than they should be.

I then took the data and calculated average exit velocities and launch angles for each player. I first did this with the total data set, which was not broken down by hit type. I calculated each player's average exit velocities across each of the four hit type subsets. I then calculated each player's average launch angle for just fly balls and line drives, as these are the most desirable outcomes for each batted ball. With enough batted balls this will give us accurate estimates for future performance, but for many players in this data set, more information or measurements are required to determine the validity of these projections. I also generated tables that show the counts of batted balls for each player, both in total and for each type of batted ball. By cross referencing the batter identification numbers in the tables with those in the calculated averages, one can look at the sample sizes to determine how accurate the estimate truly is. If I was not told to calculate estimates for every batter, I would have only displayed estimates in which the sample size was large enough for the prediction to be accurate, but since I was told to project the true average speed-off-bat for each batter I left each projection. In addition, to predict the overall hitter performance, one should consider both average velocity and launch angle, rather than only looking at one or the other. For example, a player may have a high average exit velocity, but if every hit type is a pop up or grounder he will not be very useful. On the other hand, if a player has a slightly lower exit velocity but hits mostly line drives and fly balls he will be more valuable. To help solve this problem I would recommend comparing exit velocities for each hit type rather than overall exit velocities. Given more time, I would have also computed the percentage of line drives among the batted balls hit for each player.

Furthermore, in order to project the total hitter performance, many more factors must be considered that were not given to us, such as the park factors, running speed of the batter, and quality/velocity of the pitchers faced. Only after all of these factors are implemented can we truly be able to predict with confidence a player's overall future performance. Exit velocity and launch angle are very important in determining a player's value, so the measurements that I calculated should give a solid foundation for determining a player's ability, but for more precise measurements I would love the ability to dig deeper and find the smaller factors that play a role.