**Stage Three**: **Data Package**

Within this stage, we will ‘present’ the data into a consumable format. Here are my concerns between the transition of raw data into modelling that I wish to circumvent:

1. The data requires calculations. (i.e. Calculating OBP) If I want to make minor changes to the model itself, I do not want to recalculate everything. Therefore, the calculated data should be accessible.
2. The data should be stored in three different sections: pitchers, hitters, and defense. If I will be calculating the data prior to modelling, I need to separate the data since it may require different rows. It might be possible to combine the defense stats into pitchers and hitters.
3. I want the data to be tied to the specific game. Therefore, the game id should be associated with the data which would allow a seamless transition into the model.

When pre-calculating the data, I was unsure about how to efficiently target players. (i.e. When should I perform the calculation?) I decided that I would simply calculate everything at once. This ensures that the data is accessible.

I also must ensure that data is perfectly organised where the modelling stage does not require any further adjustment. I do not think this will be issue given the current architecture.