Original slack conversation – Oct 2, 2018 ish

* So... Baseball demo. I have built a project, deployed it and have three years of new data to score.
* Tristan gave me a script that will take those three years, break it into 14 parts and score it over the course of two weeks.
* Would you be interested in taking the data, setting up a script that scores it by day over time (ignoring those days outside the regular season). So each day in the regular season is scored on one day.
* Also, this was built on very old data, so we may want to rebuild it on new data and set up a second scoring on each day as it happens.
* This is the file that scores the dataset over the course of 14 days. We don't want to do this as we want to pull each day out and score it. This means there will be a long running process to make this work.
  + pred\_pitches.py
* So the day we start the script is day 1. Score the first day's data from the datafile (which will be sometime in 2016) on day 1. Then the second day from the datafile on day 2, etc.
* I take that back. Let's score the first week of data in the file on day 1, the second week of data on day2, etc. That way we can get through the datafile in a reasonably short timeframe.
* Then, once we've scored the data weekly from the past, we can begin scoring in real time.
* As soon as possible. At least to push one month through each day until we get all the data scored.
* Then, while that is running, work with James to get the ongoing scoring working.
* File to scrape data from MLB.com. Forgive the lousy python, but I borrowed this from someone else who had no idea how to use beautifulsoup and instead of rewriting it completely just forced it to work. I was not expecting to put this into production.
  + mlb\_pull\_year.py

Baseball Prediction Notes

Real time streaming