**A place to store notes related to this project**

* Only a place to store quick text based notes, everything related to code and whatnot should be used in the playground.ipynb file

Pybaseball documentation: <https://github.com/jldbc/pybaseball/tree/master>

* Check the docs folder of the repository for specific functions

General Data Science Project Flow

1. Get the data
2. Clean the data
3. EDA of data
4. Feature Engineering
5. Modeling
6. Model Evaluation
7. Done?

So I got the data from fangraphs and was doing a bunch of EDA on it until I realized that I kinda needed to do more research into what factors actually are a good predictor of future baseball performance, rather than just looking at what fangraphs has.

* I am going to restart the EDA using data also from statcast

I am going to separate this project into several parts, the first part is going to be an EDA heavy part using things like PCA.

However, since I don’t have a lot of time, I am going to take a subset of the dataset, and use that to train a neural network using pytorch,

* In the future, I would like to use things like the PCA results to train a more fully fledged machine learning model, but I don’t know if I have the time right now and want to prioritize learning things like pytorch.