Data Science Union - Summer 2022

Project Lead: Vince Front

**Project Name: MLB Playoff Predictions**

**Overview**

Which baseball statistics are the greatest influencers on determining a team’s playoff odds? Which metrics are less important than they generally get credit for? Which metrics are more important than they generally get credit for? These are a few questions that we will be exploring throughout the summer on the MLB playoff predictions project!

This project is for anyone who is passionate about improving their Python skills, working together in a fun team environment, and/or the MLB or sports in general. We will be exploring which baseball metrics have the largest impact on a team’s playoff chances. Work will include webscraping and cleaning data, conducting exploratory data analysis (EDA), creating a machine learning (ML) algorithm that predicts a team’s playoff odds, and creating an interactive component (i.e. a web app). In addition, a big part of this project will be to improve Python skills, so all of this will be done exclusively in Python! This might sound a bit daunting at first, but there will be a lot of help and it will be a fun experience for all of us in the end!

**Goals**

* Find data and determine which baseball metrics are the best for predicting which MLB teams make the playoff and which don’t using EDA and ML
* Build an interactive component that will allow users to select/input metrics and get an output about the team’s playoff odds
* To strengthen Python skills including webscraping, data cleaning, data visualization, ML, and interactive components
* To learn how to share work on GitHub
* To work in a collaborative and fun team environment toward a common goal

**Example Potential Dataset:** [**2021 MLB Team Standard Pitching Statistics**](https://www.baseball-reference.com/leagues/majors/2021-standard-pitching.shtml)

**Week By Week Timeline:**

* Last Week of June: First meeting, finding data, and determining which metrics we want
* July Week 1: Webscraping and Cleaning data
* July Week 2: EDA
* July Week 3: ML
* July Week 4: ML
* August Week 1: Interactive Component
* August Week 2: Interactive Component
* August Week 3: Wrap up, presentation and write Up

**Skills**

Basic knowledge of Python (completing curriculum is sufficient)

**Specifications:**

* A commitment of 2-3 hours weekly including weekly check-in meetings
* Members are expected to participate and communicate throughout the duration of the project
* Deadlines will be strict and are expected to be met on time
* Code is expected to be neat, readable, and in Python
* Individual roles and responsibilities will be determined at our first meeting