

# **The Masters Golf Tournament Prediction and Equipment Analysis**

## **Progress Report**

David Yang

### **Project Scope Update**

Nothing significant has changed, everything is going according to plan. I started modeling the Masters Predictions and having a notebook so I do not need to run an entire script over and over, all while having legible markdown comments explaining what I am doing.

### **Data Sources**

The data that I am using currently is downloaded from [datagolf.com](https://datagolf.com) as CSV's. They are the past Masters Tournament results and statistics of all players, from 2021 to 2024.

Instead of using an API, I decided to scrape web pages using BeautifulSoup + Selenium. In [pgaclubtracker.com](https://pgaclubtracker.com), each page has a table of a golfer's WITB (What's In the Bag) which I am scraping and currently storing as a dictionary. Later, I will store them in a MySQL database so I can run queries on the data and even potentially connect with the Masters Prediction Model.

### **Issues / difficulties**

A potential issue might be the difference in naming schemes for the WITB portion, where players may have very similar or even the same clubs from the same brand, but are named differently in [pgaclubtracker.com](https://pgaclubtracker.com) due to it getting the sources from different third-party websites. If that issue is big enough, I will have to come up with a logic to group similar clubs together.