

Regression

Chris Byrnes Project Proposal

Due: Weds, April 7, 2021

Project Description:

I often watch golf on the weekends and with increasing frequency different statistics are highlighted during the coverage. Examples include driving distance and average putts per hole. My proposed project is to take a number of these statistics and develop a regression model to predict the amount of money a golfer will win in a given year.

Data:

The primary data source will be [espn.com](https://www.espn.com) and I will be scraping the data statistics. Most, if not all, data should be available here. Examples of potential fields used are:

- Year
- Golfer
- Driving Distance
- Total Prize Money
- Rounds Played
- Etc.

Also, in order to get enough data points I will be using data from multiple years. Prize money will be adjusted to current year based on actual increase in prize money per year.

Tools:

I'm using Beautiful Soup to web scrape the necessary data. This should be sufficient for this step of the project. I'm hoping to use SQLite to initially scrub the data before using a Pandas data frame. For regression and visualization I plan on using the covered regression library tools and Matplotlib for initial plotting. Additionally I would like to incorporate Tableau into the visualization aspect.

MVP Goal:

I would like to submit an initial regression plot as my MVP.

