

Brett Mele
7/2/2023
MSDS498

Business Case

Overview

Professional golf data has historically been difficult to leverage outside of the gaming industry. Publicly available data is sparse (though increasing in recent years), and advanced shot-by-shot tracking data via ShotLink is proprietary to the PGA tour. Despite the lack of data advanced player analysis has gained traction in recent years with the popularization of strokes gained, thanks in large part to the 2014 book *Every Shot Counts* written by Mark Broadie. The PGA tour now lists strokes gained on their professional golf leaderboards, and websites like datagolf.com use to to produce tournament projections, player rankings and betting tools.

These advances in professional golf analysis still leave some to be desired for individual professional golfers, aspiring-professional amateurs and sponsors. Available sources currently provide analysis and monitoring of performance via player rankings and likely performance in upcoming tournaments. But prediction of future performance and earnings, and subsequent inference, is a largely unexplored area of the sport. The goal of this project is to develop an end to end architecture and working prototype complete with projected performance, future earnings, and inference tools that could be offered as a product to individual golfers looking for an objective self-evaluation or companies looking for lucrative sponsorship deals.

Limitations

The primary limitation will be data availability. The most granular data available can be sourced from the datagolf.com api, which provides round by round totals for all professional golf tours and major tournaments. Individual shot and course data is not publicly available. Additionally, the available data only dates back to 2017, so the ability to forecast well into the future, or forecast amateur players, will be difficult.

Benefits

Individual Clients

1. Analysis of strengths and weaknesses
2. Prescriptive assessment & optimization of playing style
3. Understanding of potential future earnings

Enterprise Clients

1. Quantify cost of sponsorship and projected return on investment
2. Identify investment opportunities

Project Proposal

Objective

The goal of this project is to develop an end to end application for PGA Tour player projections that can be sold to prospective clients. This includes an ETL/ELT pipeline to collect data, cloud data warehouse, forecast models and a prototype application.

Deliverables

1. Data architecture
 - a. ETL/ELT pipeline to collect data from web API
 - b. Cloud data warehouse to store data
2. Performance & earnings projections
 - a. Skill forecasting models (off-the-tee, approach, around the green, putting)
 - b. Overall forecasting model
 - i. Using either (a) strokes gained or (b) some measure of tournament finish
 - c. Projected major tournament wins (or finishes)
 - d. Projected performance to earnings conversion
 - i. Using historical money from pga.com
 - e. Earnings to sponsorship cost and ROI*
 - i. Requires multiple assumptions including but not limited to typical sponsorship cost relative to career earnings, typical return on investment, type of those relationships (linear vs non-linear)
3. Prototype application
 - a. Player Dashboard
 - i. Rshiny or Flask app
 - b. Comparable players and similarity scores*

* if possible

Data and Technologies

Historical performance data and event information back to 2017 is available via the datagolf.com API. Player earnings will be scraped from pgatour.com.

Data will be stored in a cloud data warehouse (BigQuery or Postgres) hosted on Google Cloud Platform. Go and Docker will be used to collect data from the datagolf API and ingest into the warehouse. Python will be used for web scraping and modeling. R or python will be used to develop a prototype web application. All code will be hosted in a public Github repository.