EDA on Top 200 PGA tour pros

Dominic M. Barnett

Dominican University of California

ACS 2510 Data Analysis & Visualization

Hani Jandali

Feb. 29, 2024

EDA on Top 200 PGA tour pros

When you are trying to get better at something you often look at people who are better than you. In this aspect today We will be exploring the data set PGA Tour Top 200 Player Data (2015-2019). This data set tracks the stats of the top 200 PGA Tour data. The capabilities of these players are the top 1% of golfers; they are the pinnacle of the sport so the numbers will reflect that. This data will not be a good comparison to look at for any average golfer; it is like comparing a local pond to the ocean. In this analysis we will explore what sets apart the top 10 from the top 100 and from the top 200. What can be the deciding factor on where a player falls in these rankings.

First we will look at cleaning up the data that way we are only looking at necessary information. This dataset contains 1497 entries of these top 200 players, these entries are also separated by the year of the season. There are also 29 different features which include; 'PLAYER NAME', 'Ball Speed', 'Driving Distance', 'Approaches from > 100 yards', 'Eagles (Holes per)', 'Putts Per Round', 'Birdie Average', 'Proximity to Hole', 'FedexCup Regular Season Points', 'Average Distance of Putts made', 'GIR Percentage from Fairway', 'Official Money', 'Bogey Average', 'Par 3 Scoring Average', 'Club Head Speed', 'Par 4 Scoring Average', 'Par 5 Scoring Average', 'Driving Accuracy Percentage', 'Scoring Average', 'Total Birdies', 'Total Eagles', 'Spin Rate', 'Top 10 Finishes', 'Sand Save Percentage', 'Scrambling', 'SG: Total', 'Smash Factor', 'Country', 'College’. There are only a few features that we are going to remove. We don’t need the features; Eagles (holes per), spin rate, country, and college. In this EDA we care about what is setting players apart within their scoring, not where they’re from or loosely tracked metrics.

Once the data is cleaned to your liking then we start asking questions. What are these top 200 pros putting out with their driving distances? With a simple histogram we see that a majority of them fall within the 280-310 yard range with a handful of drives being shorter and longer than those numbers. This sparks my imagination a little bit, if the entire goal of golf is to put the ball in the hole in the least amount of strokes. Does driving distance correlate to more top 10 finishes? This time I used a scatterplot to look at this comparison. The scatterplot did show that there were plenty of top 10 finishers that drove the ball more than 310 yards but a majority of the pack still fell in the 280-310 range. In that respect then there has to be a direct correlation between birdie average and top 10 finishers. I used a scatterplot for this as well, I found this to be true. The scatterplot shows an increasing trend of appearances in the top 10 the more often you could expect birdies out of players. Of course if it's true for birdie average it must be true for scoring average. I found that this is also true with the help of a line plot. You can see that the increase in appearances in the top 10 occurs when players' scoring averages start to dip below 72 strokes. From here I decided to work backwards from the course perspective. If a better birdie average leads to a better scoring average allowing players to appear in the top 10 more often. Is there a place that players can start from that allows them a better shot for birdie. Of course there is, while comparing GIR from the fairway I found a positive correlation to birdie averages. The higher the percentage of GIR from the fairway a player had the better birdie averages that player had. These finding allowed me to follow this thought path. When a Pro starts their approach in the fairway they tend to have a better chance at birdie. The more birdies a player makes the better their scoring average. As shown on a graph above there is a positive relationship to having a better scoring average and finishing in the top 10. The final question is does finishing in the top

10 lead to more official money. This is fairly simply to account for, of course the top 10 finishers would get more money. The #1 player gets the most and then it trickles down to everyone else

In summary we looked at the top 200 PGA tour players. We asked a couple questions regarding how those top players separated themselves from the pack and lowered their scoring averages. We found that the more often a player hit the fairway their approach shot was more likely to hit the green. That gives them a better proximity to the hole and gives them a better birdie average. The players with better birdie averages tended to have better scoring averages. This is where the pack separates because those with the best scoring averages tend to have a much better chance at the top 10 finishers.

References

TrackMan. (2014, June 6). TrackMan Average Tour Stats. TrackMan Golf. https://blog.trackmangolf.com/trackman-average-tour-stats/

Golf Stat and Records | PGA TOUR. (n.d.). PGATour. https://www.pgatour.com/stats

PGA Tour. (2020, March 18). Wikipedia. https://en.wikipedia.org/wiki/PGA_Tour