Checkpoint 5

Data 6560 Sport Analytics
Ryan Tobin
12/1/25

Risk vs. Reward: Analyzing Aggressive vs. Conservative Decision-Making in Golf

Question Snapshot

This project analyzes how aggressive vs conservative decision making influences performance outcomes in golf. Success means identifying measurable situations where aggressive play (high risk tee shots, attacking approaches, scoring volatility) results in higher or lower strokes gained outcomes compared to safer alternatives. The goal is to detect patterns that can inform optimal strategy for different player types and course decisions.

Data Used

The analysis uses the cleaned dataset created in checkpoint 4 (https://github.com/rtobin2344/Golf-Decision-Making-Model/blob/main/ASA_All_PGA-Tourn_Level_Derived.csv), generated from OpenDataBay's round level PGA tour performance data up to 2022. Each row represents a player's round in a tournament, including strokes gained metrics (sg_ott, app, arg, putt, t2g, total), finishing positions, par/score information, and derived performance indicators. No ShotLink shot by shot data was available to use although initial proposals indicated this data being a part of the analysis.

Univariate Distributions

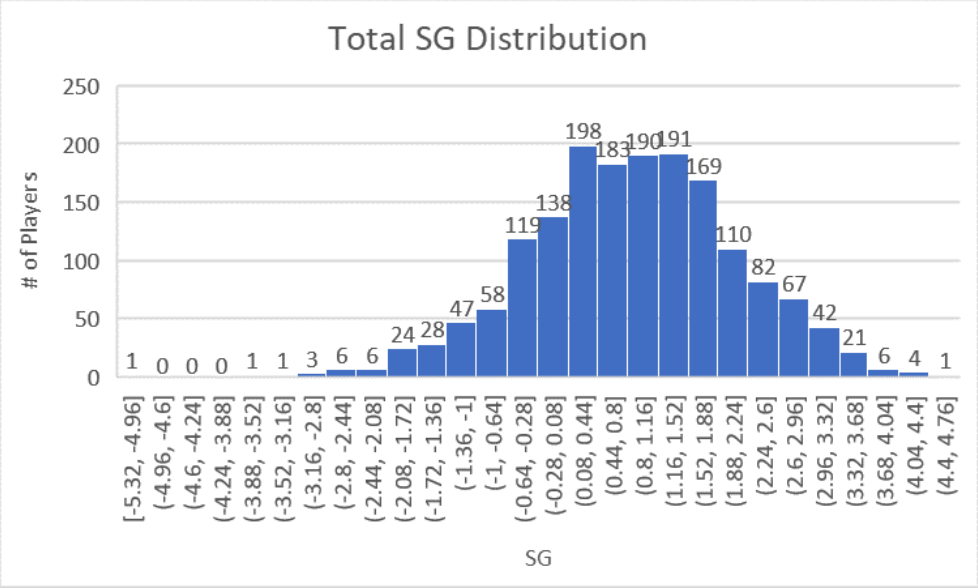

Fig. 1 – distribution of strokes gained (sg_total) [28 points, 0.36 SG between markers)
Takeaway: most rounds fall within -0.64 to 2.24 SG total, with the right tail representing elite performances (above 3 strokes gained)

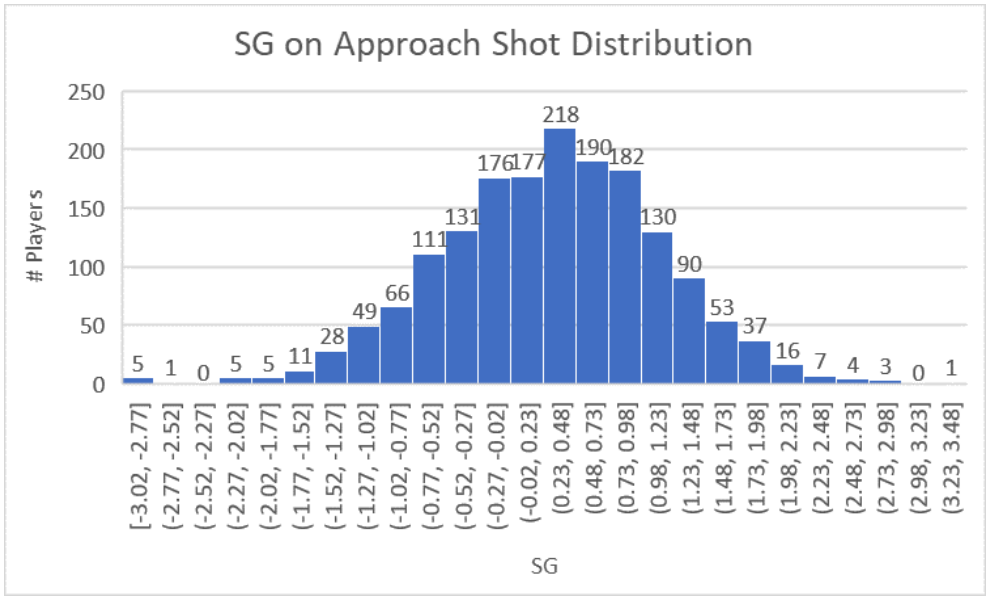

Fig. 2 – distribution of sg_app (approach play)
Takeaway: sg_app has a very normal looking distribution, with most rounds coming in at -0.77 to 1.48 SG on approach shots.

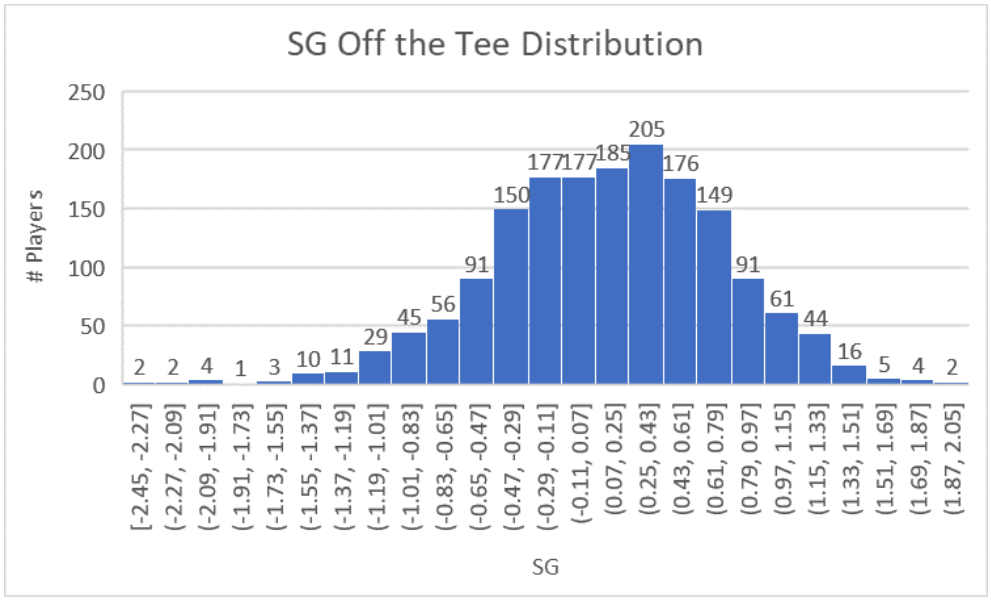

Fig. 3 – distribution of sg_ott (off the tee)
Takeaway: sg_ott is also normally distributed, with most rounds falling between -0.65 to 0.97 SG off the tee. The SG here are notably slightly less than the first 2 figures, likely because a lot can happen after the tee shot and a good/bad tee ball doesn't determine SG outcome as heavily as approach for example.

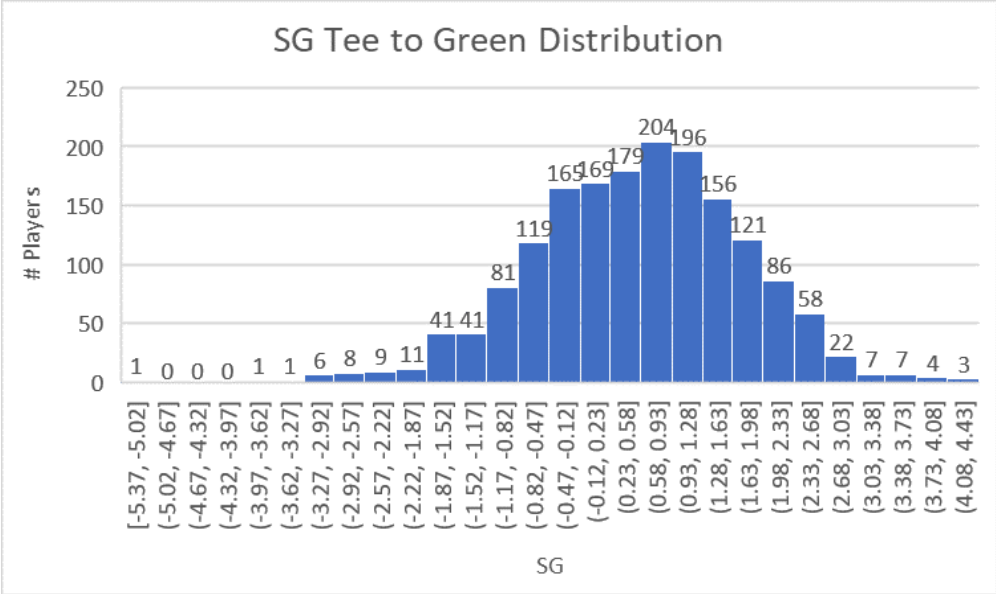

Fig. 4 – Distribution of aggressiveness proxy (sg_T2G volatility)
Takeaway: SG from tee to green is a similar metric to total SG, minus putting. This might be the most accurate measure for aggressiveness, as putting is straightforward with less decision making (and more directly skill based vs decision based). The data falls heavily between -1.17 to 2.33 SG.

Relationships Between Key Variables

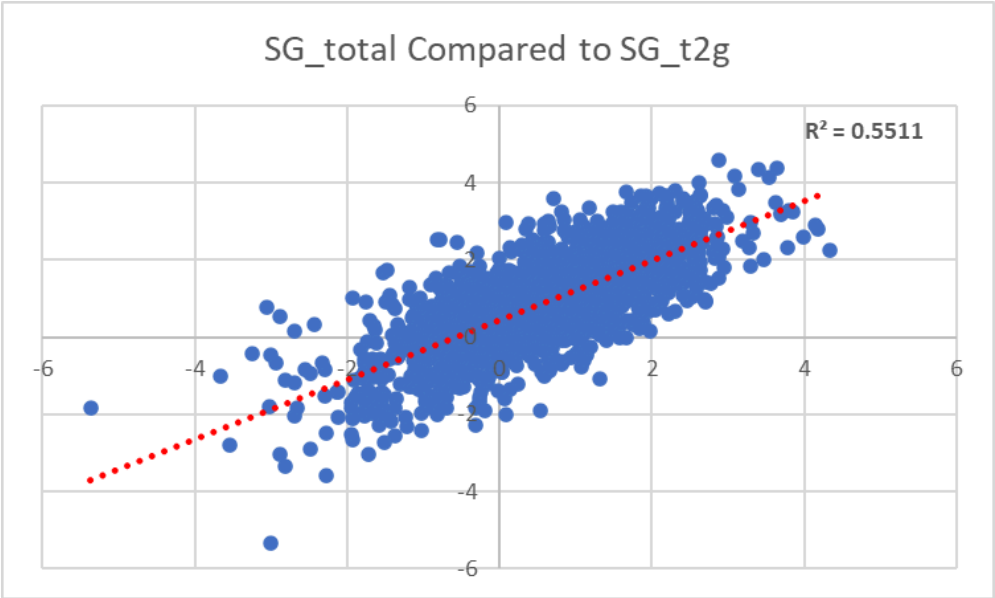

Fig. 5 – Aggressiveness Proxy vs SG_Total
Takeaway: There is a moderate, positive linear relationship between total strokes gained and strokes gained from tee to green. The R squared coefficient of 0.551 demonstrates this relationship. These variables move together as expected, as more strokes gained getting from

the tee to the green would put a player in a better position to score with their putt compared to the field,

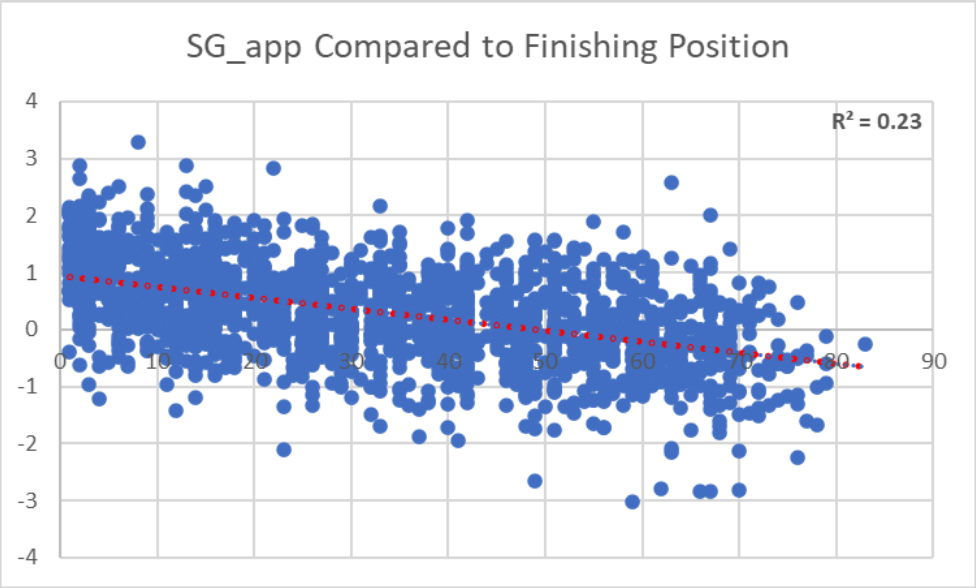

Fig. 6 – SG_APP vs Final Position

Takeaway: The scatterplot of SG on approach shots vs finishing position shows a relatively weak, negative linear relationship. A negative relationship in this scenario is a good thing, as higher numbered finishing position is worse (i.e. finishing #1 is better than #10, even though 10 is 'higher'). The R squared value of 0.23 demonstrates this. I thought there would be a stronger correlation here with this surface level analysis.

Subgroup Comparison

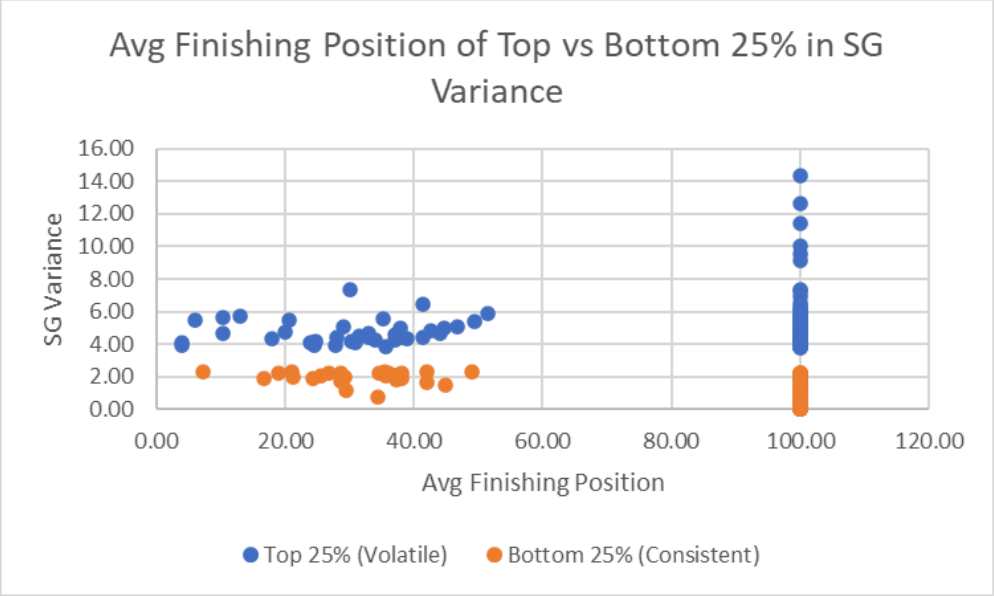

Fig. 7 – SG_Total comparison; high vs low volatility players

Takeaway: This chart is fairly complicated and is good as a quick visual (i.e. not something included in a final presentation). Blue is the top quartile of players in total SG variance. These are volatile players that have a high SG variance tournament to tournament. Orange is the opposite, the bottom quartile of players in total SG variance. The average 2022 finishing position is compared for these two groups. Notably, a finishing position of '100' is used to indicate the player did not make the tournament cut after 2 rounds all of 2022 (no data for finishing position after 4 rounds). There are 82 players like this in the high volatility group, and 94 in the low volatility group. There are 6 players in a top 15 avg finishing position among the volatile group, and only 1 player from the consistent group. These players are on the left most side of the chart. I think this preliminary visual analysis is indicating that aggressiveness (measured here by SG volatility) is leading to more success than playing conservatively.

Time Sequence Trend

There is no time sequence data to analyze with the data I have at my disposal. If there were shotlink data publicly available, I could analyze round 1 through round 4 data, seeing if aggressiveness changes from the start to the end of a tournament.

Outliers & Anomalies

There were two instances where a player had above a 4.5 SG and below a -4.5 SG. However, these appear legitimate, especially since each golfer can play incredibly poorly or well compared to the field in any given tournament. These outliers will be flagged during modeling, but look to be real data points.

Missing SG metrics occur in rounds where players withdrew or where partial SG stats were not available. I won't use these rounds for modeling, and there are very few of them since the raw data was already well processed/cleaned by the provider

Some tournaments had unusually easy or difficult scoring conditions; these tournaments will potentially be ignored, or used as a sub category (is it better to be aggressive in easy vs hard course conditions)

Missingness & Coverage

- Missingness clusters in withdrawn rounds and early season events
- SG components are fully available for almost all rounds
- No significant missingness patterns seen, don't threaten subgroup comparisons

Early Takeaways

- SG_T2G may be the strongest single predictor of finishing position other than SG_total
- Aggressiveness in metrics of volatility show some seemingly meaningful differences in ceiling and floor of round performance
- High volatility players have slightly higher peak SG_total, possibly suggesting aggressive strategy may pay off
- Professional golfers care most about finishing position – making cuts gets them paychecks, and better finishing positions means more money and tour status. Early signs are showing aggressive play may help accomplish this more than being conservative.

Next Steps

1. Formalize and Aggressiveness metric (SG Volatility, SG_OTT + SG_APP weighted ‘risk index’, Frequency of high/low performance rounds)
2. Build first regression model (predict SG_Total using SG components) [careful of overfitting model]
3. Add aggressiveness metric into model to test linear effects, nonlinear effects, and interaction w/ player skill tier and/or course difficulty
4. Evaluate effect sizes and model interpretation for strategic recommendations