PGA Golfer Statistical Analysis

Graeme Galloway

December 20, 2023

Introduction

In the dynamic world of professional golf, it is important to understand the factors that contribute to a golfer's success. This project delves into the statistical analysis of PGA golfer performance, leveraging advanced data science techniques to extract meaningful insights. Using Python and key libraries such as NumPy, Pandas, and Matplotlib, this analysis explores the intricate relationships between various performance metrics and success on the golf course at the professional level.

The primary objectives include conducting an in-depth examination of driving distance (in yards), driving accuracy (%), greens in regulation (GIR), and putts per hole. By applying statistical methods such as ANOVA and correlation analysis, the aim is to identify which of these metrics exhibit the strongest correlations with the number of cuts made, a crucial indicator of a golfer's success.

This report not only presents the statistical findings but also offers visual representations, including side-by-side bar plots comparing golfers and scatter plots illustrating the performance trends of individual players across multiple seasons. The comprehensive analysis aims to provide actionable insights for both enthusiasts and industry professionals, shedding light on the multifaceted nature of success in professional golf.

Data Exploration

The original dataset for this project was sourced from ESPN. The seasons that were used in the analysis were 2020-21, 2021-22, 2022-23. The top 100 golfers from each of the three seasons were taken from the website along with their corresponding metrics and transferred over into an excel file.

The analysis focused on several key variables that play a role in evaluating golfer performance. These variables include Driving Distance (DDIS), Driving Accuracy (DACC), Greens in Regulation (GIR), Putts per Hole (PUTTS) and Cuts Made (CUTTS).

Part One: Identifying Crucial Factors for Tournament Success

To assess the factors influencing golfer success, Analysis of Variance (ANOVA) was the main statistical tool that was utilized. ANOVA is particularly useful in determining whether there are any statistically significant differences between the means of multiple groups, making it well-suited for the investigation into the impact of various golf performance metrics.

The primary focus of the analysis was on the F-statistic and p-value derived from ANOVA. The F-statistic gauges the ratio of variance between groups to variance within groups, providing a measure of how well the group means explain the variance in the data. A higher F-statistic suggests a more significant difference between group means. The p-value helps determine the statistical significance of the observed results. A low p-value indicates that the observed differences are unlikely to be due to random chance alone.

With the highest average F-statistic suggesting a strong relationship between variables, and the lowest average p-value among the variables considered, indicating that this relationship was the most statistically significant, one can conclude that in Greens Regulation (GIR) had the strongest correlation to Cuts made (CUTS) on average over the three seasons. The interpretation aligns with the idea that maintaining a high Greens in Regulation (GIR) is associated with higher success in terms of making cuts in golf tournaments.

Key Statistics from ANOVA Tables:

<u>Driving Distance (DDIS) and CUTS:</u>
2022-2023 Season:
F-Statistic: 13.2328
p-value: 0.000441

2021-2022 Season:
F-Statistic: 18.378464
p-value: 0.000042

2020-2021 Season:
F-Statistic: 15.723608
p-value: 0.000139

Average Statistics:
F-Statistic: 15.77
p-value: 0.000207

Driving Accuracy (DACC) and CUTS:
2022-2023 Season:
F-Statistic: 6.357132
p-value: 0.013304

2021-2022 Season:
F-Statistic: 28.456101
p-value: 6.169681e-07

2020-2021 Season:
F-Statistic: 18.600733
p-value: 0.000038

Average Statistics:
F-Statistic: 17.804
p-value: 0.00444

Greens in Regulation (GIR) and CUTS:
2022-2023 Season:
F-Statistic: 18.247835
p-value: 0.000045

2021-2022 Season:
F-Statistic: 27.239962
p-value: 0.000001

2020-2021 Season:
F-Statistic: 28.484608
p-value: 6.099727e-07

Average Statistics:
F-Statistic: 24.6574
p-value: 0.0000155

Putts per Round (PUTTS) and CUTS:
2022-2023 Season:
F-Statistic: 11.19056
p-value: 0.001166

2021-2022 Season:
F-Statistic: 20.315482
p-value: 0.000018

2020-2021 Season:
F-Statistic: 12.165462
p-value: 0.000731

Average Statistics:
F-Statistic: 14.5571
p-value: 0.000638

Part Two: Performance Comparison of Scottie Scheffler (1st Rank) vs Nate Lashley (100th Rank) in the 2022-2023 Season

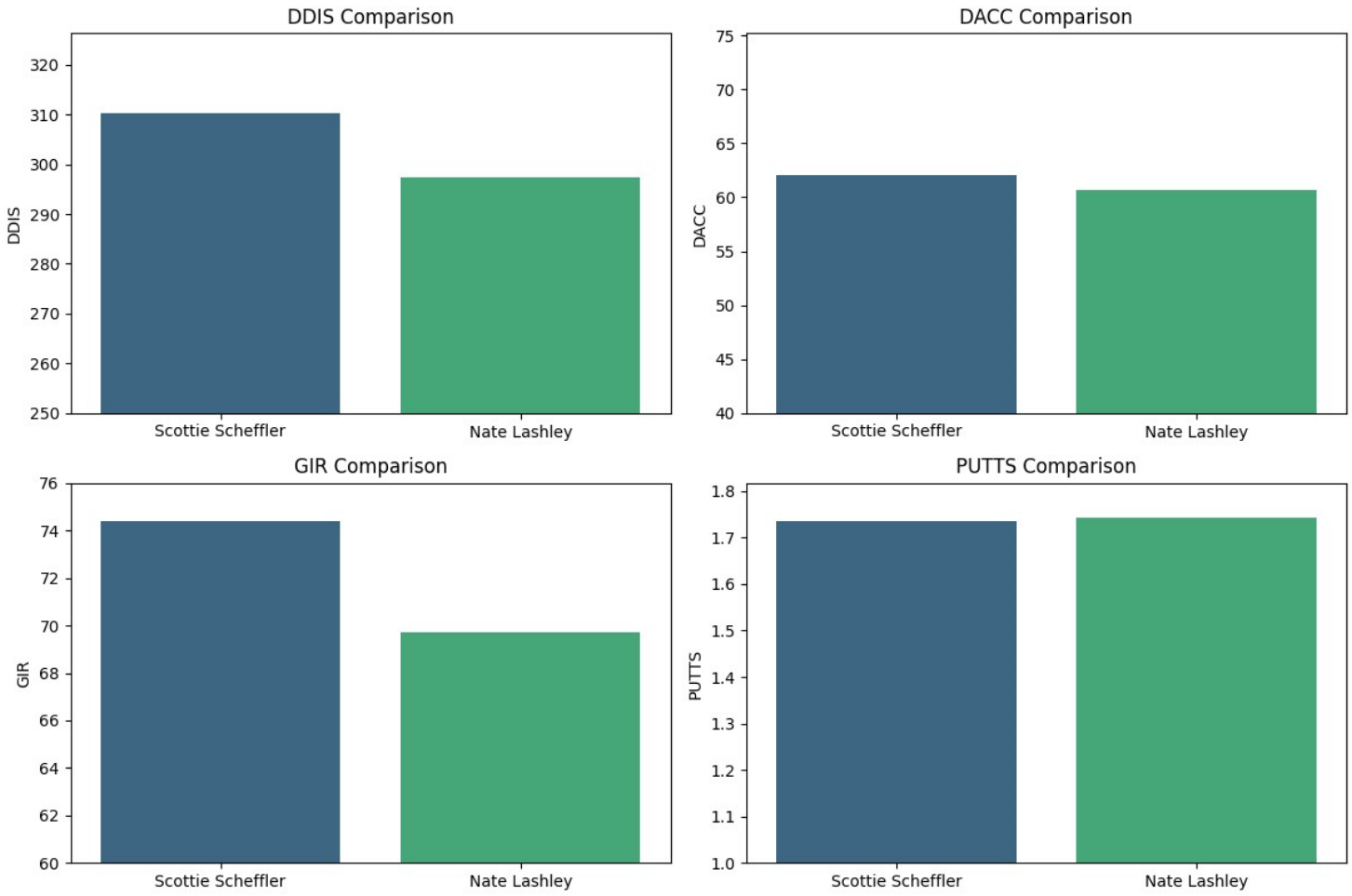
Motivated by curiosity about the distinctions between top-ranked and lower-ranked golfers, this section involves conducting a detailed analysis between Scottie Scheffler (1st Rank) and Nate Lashley (100th Rank) during the 2022-2023 season. The aim is to identify key differences in their performance metrics and shed light on potential areas for improvement.

Through comprehensive bar plots, four critical variables – Driving Distance (DDIS), Driving Accuracy (DACC), Greens in Regulation (GIR), and Putts per Hole (PUTTS) – are compared between Scottie Scheffler and Nate Lashley.

Clear disparities emerge, particularly in GIR, aligning with the earlier conclusion that GIR strongly correlates with making more cuts and better performance. The bar plots also highlight interesting similarities in putting performance, offering strategic considerations for each player's training focus. For Nate Lashley, the data suggests a potential focus on approach shots to enhance GIR, leveraging his existing putting proficiency. Conversely, Scottie Scheffler might benefit from dedicating additional time to putting.

One of the fascinating conclusions from this analysis is the extremely thin margins between certain metrics, even when comparing the extremes of the leaderboard. For example, although Nate Lashley averages 0.007 more putts per hole than Scottie Scheffler, this slight discrepancy is not nearly enough to make him a top golfer or outperform his competitors. These graphs outline the fact that at such a high level of golf, every decimal point holds significance, and what distinguishes the top golfers from the rest are fractional improvements in each metric.

[Link to the Photo]



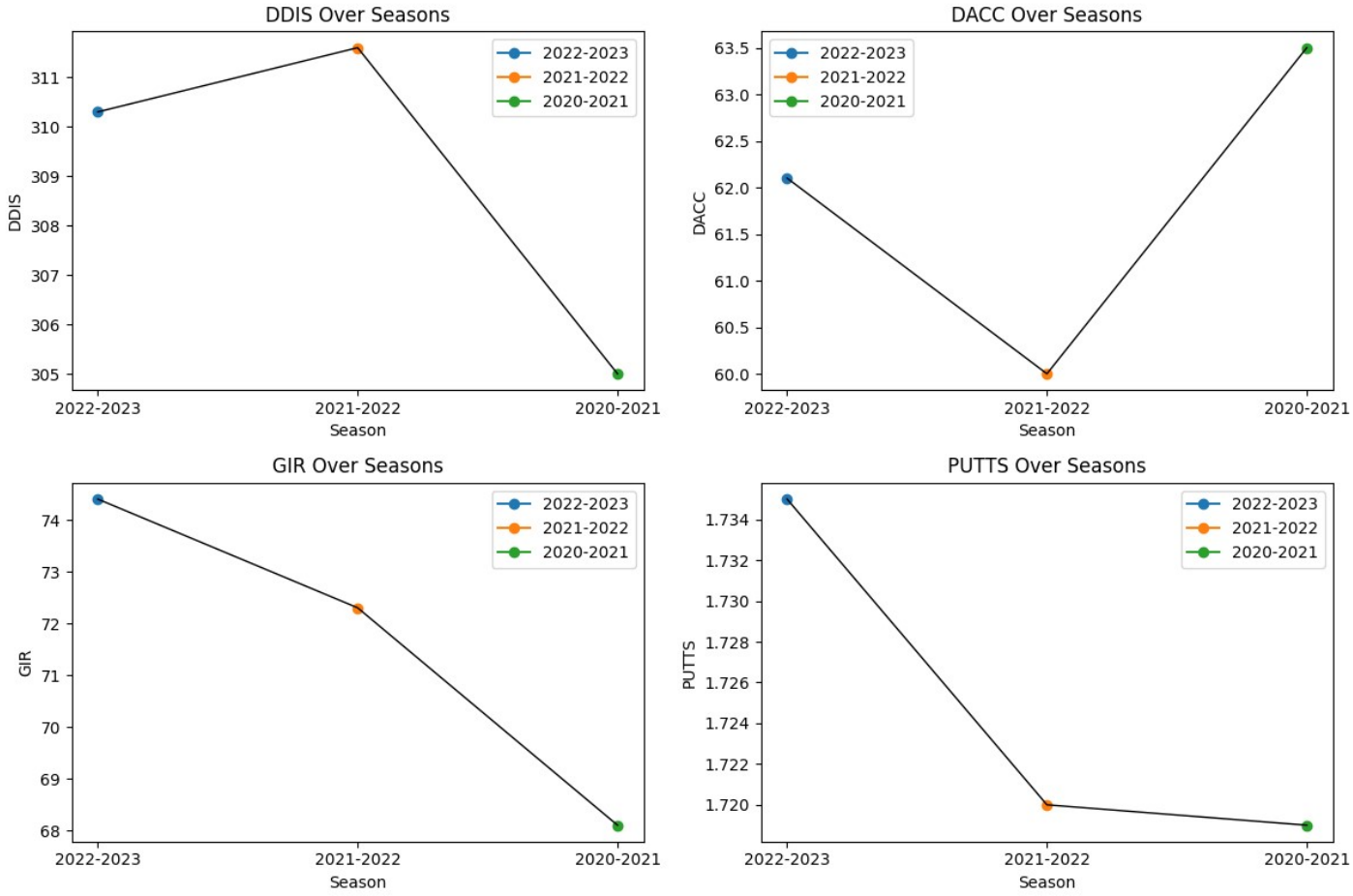
Part Three: Scottie Scheffler’s Performance Trends over the Seasons

After evaluating Scottie Scheffler's performance across the three most recent PGA tour seasons, noteworthy trends have emerged. Notably, Scottie transitioned from the 19th rank in the 2020-2021 season to achieving the top rank in both the 2021-2022 and 2022-2023 seasons. The analysis of Scottie's metrics over this period reveals intriguing patterns of progression and regression.

One striking observation is the consistency in the positive increase of Greens in Regulation (GIR) over the three seasons. This aligns with the earlier analysis, which identified GIR as the strongest correlator to a golfer's success in terms of cuts made. The consistent upward trend in GIR suggests its crucial role in Scottie Scheffler's rise to the top.

Examining Scottie's putting statistics, a subtle shift is noticeable, particularly in the 2022-2023 season, where putts per hole increased by 0.015 compared to the previous season. While the ideal scenario involves improvement in all metrics each season, this fluctuation reminds us of the nature of any sport, and how constant progression is not attainable. In essence, the main takeaway underscores how improving Greens in Regulation (GIR) has been a key factor in Scottie Scheffler's success over the years. This trend not only mirrors his steady climb to the top but also supports the conclusions drawn in the earlier sections.

Link to the Photo



Conclusion

In conclusion, this project focused on the statistical analysis of PGA golfer performance, revealing key insights that outlined the factors influencing success on tour. The application of ANOVA, focusing on F-statistics and p-values, highlighted Greens in Regulation (GIR) as a consistent and influential metric in determining the number of cuts made. The comparative analysis of golfers Scottie Scheffler and Nate Lashley provided a unique understanding of performance variations across key metrics. Lastly, the examination of Scottie Scheffler's metrics over three seasons emphasized the key role that the progression of GIR had in career. These findings not only contribute to the understanding of golf performance dynamics but also showcase the significance of strategic data analysis in gaining actionable insights from complex datasets.