The story of the PGA Championship:

An exploration of PGA Championship 2024 results

A research project by Matt Asquith

Much more than just the results…

Name	Country	Position	Total score	R1	R2	R3	R4	Total strokes	Status
Xander Schauffele	USA	1	-21	62	68	68	65	263	R4 Completed
Bryson DeChambeau	USA	2	-20	68	65	67	64	264	R4 Completed
Viktor Hovland	NOR	3	-18	68	66	66	66	266	R4 Completed
Thomas Detry	BEL	T4	-15	66	67	70	66	269	R4 Completed
Collin Morikawa	USA	T4	-15	66	65	67	71	269	R4 Completed
...	...	...	...	...	...	...	...	...	...
Jeff Kellen	USA	CUT	+17	87	72	-	-	159	R2 Completed
Rich Beem	USA	CUT	+20	79	83	-	-	162	R2 Completed
John Daly	USA	WD	-	82	-	-	-	82	Withdrew
Ben Griffin	USA	WD	-	73	65	-	-	138	Withdrew
K.H. Lee	KOR	WD	-	74	66	-	-	140	Withdrew

Xander Schauffele - Winner

263 strokes

21 under par

Quick primer

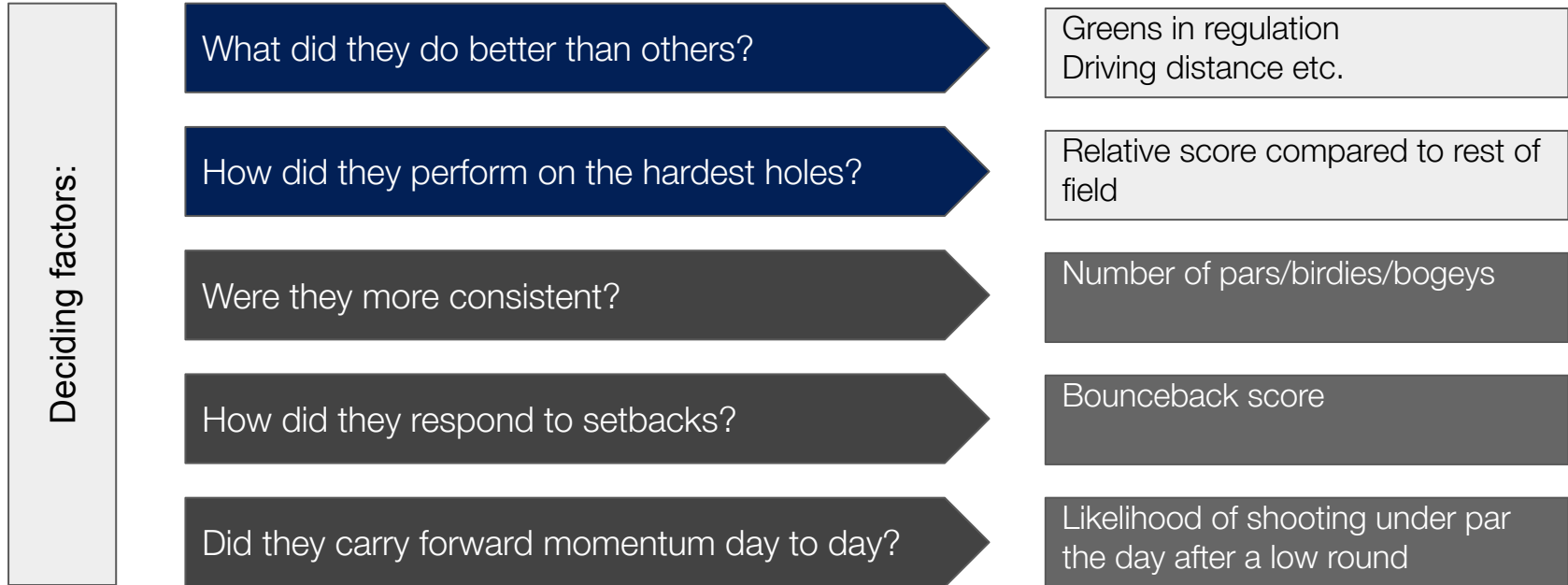**What are the players trying to achieve?**

- Go round the course 4 times in the lowest number of shots possible

What are the funny terms?

- **Par** is the score that the professionals should score on each hole
- **Birdie** is scoring one under par or (-1) on a hole
- **Eagle** is scoring two under par or (-2) on a hole
- **Bogey** is scoring one over par or (+1) on a hole

But how did the top players distinguish themselves?

Deciding factors:

What did they do better than others?
> Greens in regulation
> Driving distance etc.

How did they perform on the hardest holes?
> Relative score compared to rest of field

Were they more consistent?
> Number of pars/birdies/bogeys

How did they respond to setbacks?
> Bounceback score

Did they carry forward momentum day to day?
> Likelihood of shooting under par the day after a low round

How did the top 5 stand out?

Name	SG: Off The Tee	SG: Approach to Green	SG: Around The Green	SG: Putting	SG: Total	Driving Accuracy	Driving Distance	Longest Drive	Greens in Regulation	Sand Saves	Scrambling	Putts per GIR	Feet of Putts Made
Xander Schauffele	5.227	7.811	0.424	4.704	18.165	73.21% (41/56)	310.10 yds	330.00 yds	83.33% (60/72)	100.00% (5/5)	83.33% (10/12)	1.63	95.0
Bryson DeChambeau	5.890	2.261	5.006	4.009	17.165	62.50% (35/56)	330.50 yds	360.00 yds	72.22% (52/72)	100.00% (6/6)	90.00% (18/20)	1.64	72.0
Viktor Hovland	4.575	6.522	1.468	2.601	15.165	67.86% (38/56)	303.30 yds	328.00 yds	77.78% (56/72)	66.67% (6/9)	81.25% (13/16)	1.64	88.0
Thomas Detry	1.361	0.703	5.989	4.113	12.165	66.07% (37/56)	305.80 yds	329.00 yds	65.28% (47/72)	83.33% (10/12)	84.00% (21/25)	1.64	78.0
Collin Morikawa	3.674	3.177	6.078	-0.763	12.165	78.57% (44/56)	294.10 yds	328.00 yds	70.83% (51/72)	50.00% (3/6)	80.95% (17/21)	1.71	70.0

Schauffele's winning formula was:
- 3rd for SG off the tee
- 2nd overall for approach to green
- 1st for Greens in regulation

An exploration of PGA Championship 2024 results

Hole 6 - Graveyard of champions?

Which was the most difficult hole to score on?

Hole 6 had an average score of **0.228 over par**, the highest average among the 18 holes.

And how did the top players perform on it?

Whole field averages:

R1: 4.254 | R2: 4.320 | R3: 4.202 | R4: 4.013

Name	6	6	6	6
Xander Schauffele	4	4	4	4
Bryson DeChambeau	4	4	4	4
Viktor Hovland	4	4	5	3
Thomas Detry	4	3	4	3
Collin Morikawa	4	4	5	4
Justin Rose	3	3	4	4
Shane Lowry	4	4	4	5
Billy Horschel	4	4	4	4
Scottie Scheffler	4	4	4	4
Justin Thomas	4	4	4	4

Key observation: Top 10 players accounted for only 3/111 bogeys on this hole and scored 5/38 birdies.

Methodology

Web scraping

- **Static:** html scraping with BeautifulSoup in JSON format and converted to py dict to generate main results table
- **Dynamic:** Scraping with Selenium for dynamic tables to generate hole by hole scoring data

Data wrangling

- **Using concat()** to add the par scores to the DataFrames for context
- **Subtracting pandas Series** of par scores from individual rounds to generate tables for pars/birdies/eagles
- **Nested list comprehension** to generate hole-by-hole data

Outputs/Packages

Web scraping	BeautifulSoup Selenium, By json
Data wrangling & cleaning	Numpy, Pandas

Challenges

Web scraping

- Dynamic tables

Missing data

- Clean dataset
- Gaps by virtue of the structure of the competition
- Casting large amounts of string data to integers

Tournament structure

| R1 | R2 | R3 | R4 |

Full field

156 players

Top 78 players

CUT

Thank you!