

Mark Riley Final Project Problem 3

DS 730 – Fall 2019

Dataset Source

PGA Tour Golf Data

Every statistic recorded on the PGA Tour from 2010 - Nov 3, 2019

By Brad Klassen

<https://www.kaggle.com/bradklassen/pga-tour-20102018-data>

I used two of the datasets from the source. The first is annual player statistics from the 2010 – 2018 Professional Golf Association (PGA) Tour seasons. The second is the weekly cumulative statistics from the 2019 PGA tour which began January 24, 2019 and ended at The Tour Championship on August 25, 2019. I did not include the 2020 data because it represents only the early part of the season.

I joined the two datasets together after filtering the 2019 data for the final tournament of the year (the Tour Championship) because the 2019 stats represent cumulative data. I also filtered down to a number of statistics I thought would be interesting. These include:

- Driving Distance - (ROUNDS)
- Driving Distance - (AVG.)
- Driving Accuracy Percentage - (%)
- Greens in Regulation Percentage - (%)
- Putting Average - (AVG)
- Scoring Average (Actual) - (AVG)
- Official Money - (YTD VICTORIES)
- Sand Save Percentage - (%)
- Scoring Average - (AVG)
- Scrambling - (%)
- Top 10 Finishes - (TOP 10)
- Top 10 Finishes - (1ST)
- Putting from Inside 5' - (%)
- Putting from 5-10' - (%)
- Putting from - 10-15' - (% MADE)
- Putting from - 15-20' - (% MADE)
- Putting from - 20-25' - (% MADE)
- Putting from - > 25' - (% MADE)
- Club Head Speed - (AVG.)
- Ball Speed - (AVG.)

Mark Riley Final Project Problem 3

DS 730 – Fall 2019

Questions

I used Hive to answer the following questions.

Q1. Who was the worst scrambler (scramble = score of par or better without a green in regulation = ball on green in two strokes less than par) who won a tournament in each season?

Cameron Champ	52.14	2019
Satoshi Kodaira	51.02	2018
Rod Pampling	52.83	2017
Bubba Watson	52.74	2016
Ben Martin	54.04	2015
Russell Henley	52.37	2014
Derek Ernst	49.34	2013
Scott Stallings	52.33	2012
Bubba Watson	50.82	2011
Robert Garrigus	51.98	2010

Q2. Which player had the biggest difference between percent of putts made outside 25 feet minus percent of putts made inside 5 feet in the same season?

Mark Calcavecchia 2010

Q3. Who are the top five players with at least 10 wins with the highest percentage of their top 10 finishes are tournament wins?

Player Name	Pct. Top 10 Wins	Rank
Steven Bowditch	0.83	1
Martin Kaymer	0.79	2
Russell Knox	0.67	3
J.J. Henry	0.62	4
Geoff Ogilvy	0.62	4
Francesco Molinari	0.59	5

Q4. Which of the following statistics is most negatively correlated to Scoring Avg across all players and seasons? Avg. Driving Distance, Driving Accuracy Percentage, Greens in Regulation Percentage, or Scrambling Percentage? (Negative is better in golf because lowest score wins).

Driving Distance	Driving Accuracy	Greens in Regulation	Scrambling
-0.1436	-0.0263	-0.1738	-0.2179