DRIVE FOR SHOW, PUTT FOR DOUGH

Kyle Murphy, Nathan Kidder, Donnie Canchela, Will Girardi

WHAT DATA DID WE USE?

- The dataset for this activity was obtained from the PGA statistics website.
- Cases include all golfers who made the cut in each of 19 PGA tournaments in 2022. The dataset includes variables for driving ability, putting ability, and measuring success in the tournament.
- Dataset only includes players that made the CUT. That is the major limitation to this dataset.

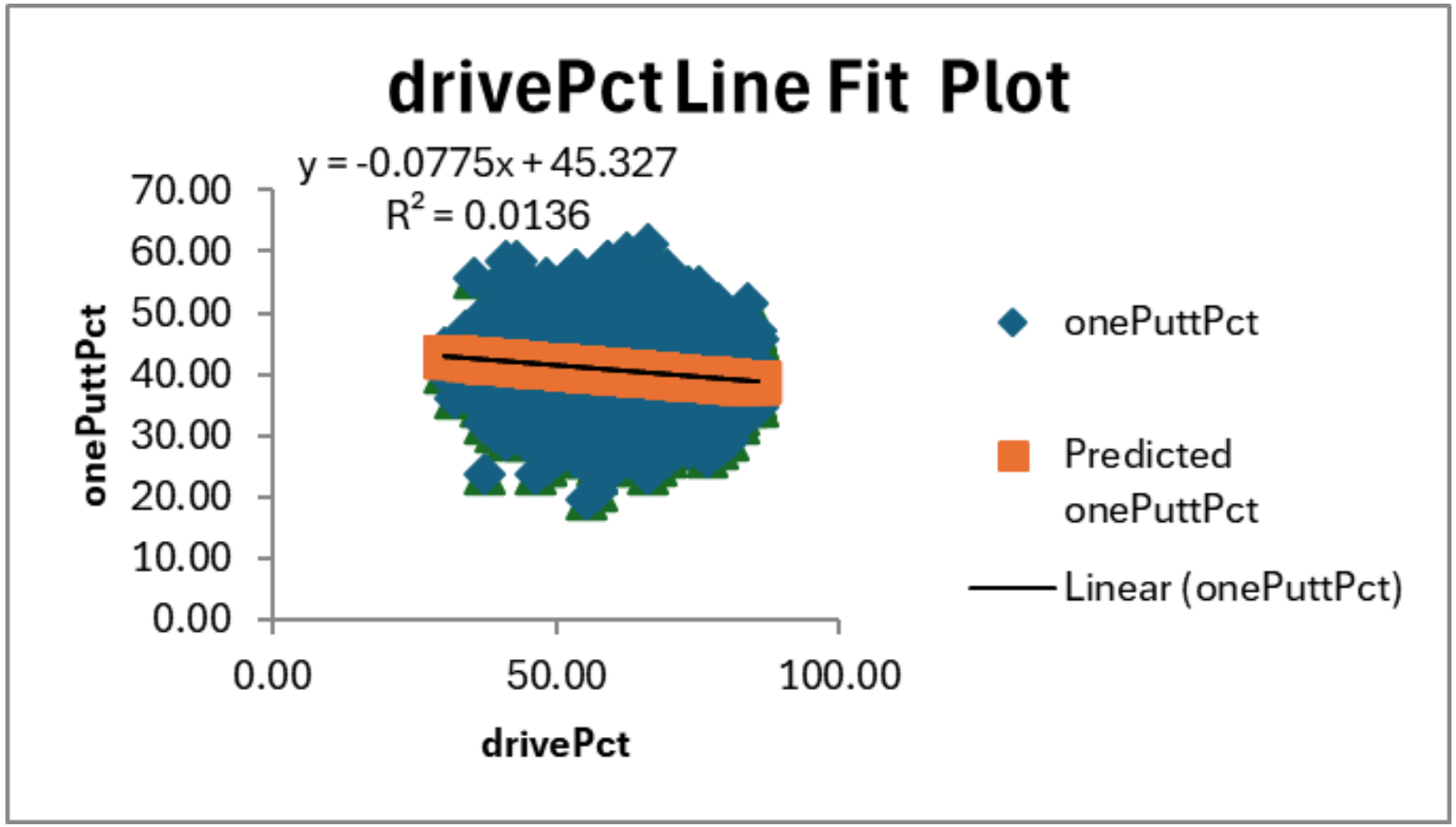
WHAT VARIABLES DOES OUR DATASET HAVE?

- The “driving” variables include average driving distance (avgDriveDist), driving accuracy percentage (drivePct), and strokes gained off the tee (driveSG).
- The “putting” variables are average putts per round (avgPuttsPerRound), one putt percentage (onePuttPct), and strokes gained putting (puttsSG).
- The variables to measure success are scoring average (avgScore), official money won (Money), and Fedex Cup points (Points).

QUESTION 1

- Is there a direct correlation between drivePct (% of fairways landed on during weekend) and the onePuttPct (% of time it took one put to make it in the cup)? What is the magnitude of the scatterplot and what is the R^2 value of the scatterplot created?
- Analyzing this correlation is important because it helps determine if there's a relationship between a golfer’s accuracy off the tee (drivePct) and their ability to one-putt (onePuttPct).
- High correlation: Suggests more fairways hit could lead to better positions for easier putts.
- Low correlation: Implies that putting performance might be independent of driving accuracy, focusing attention more on the short game.

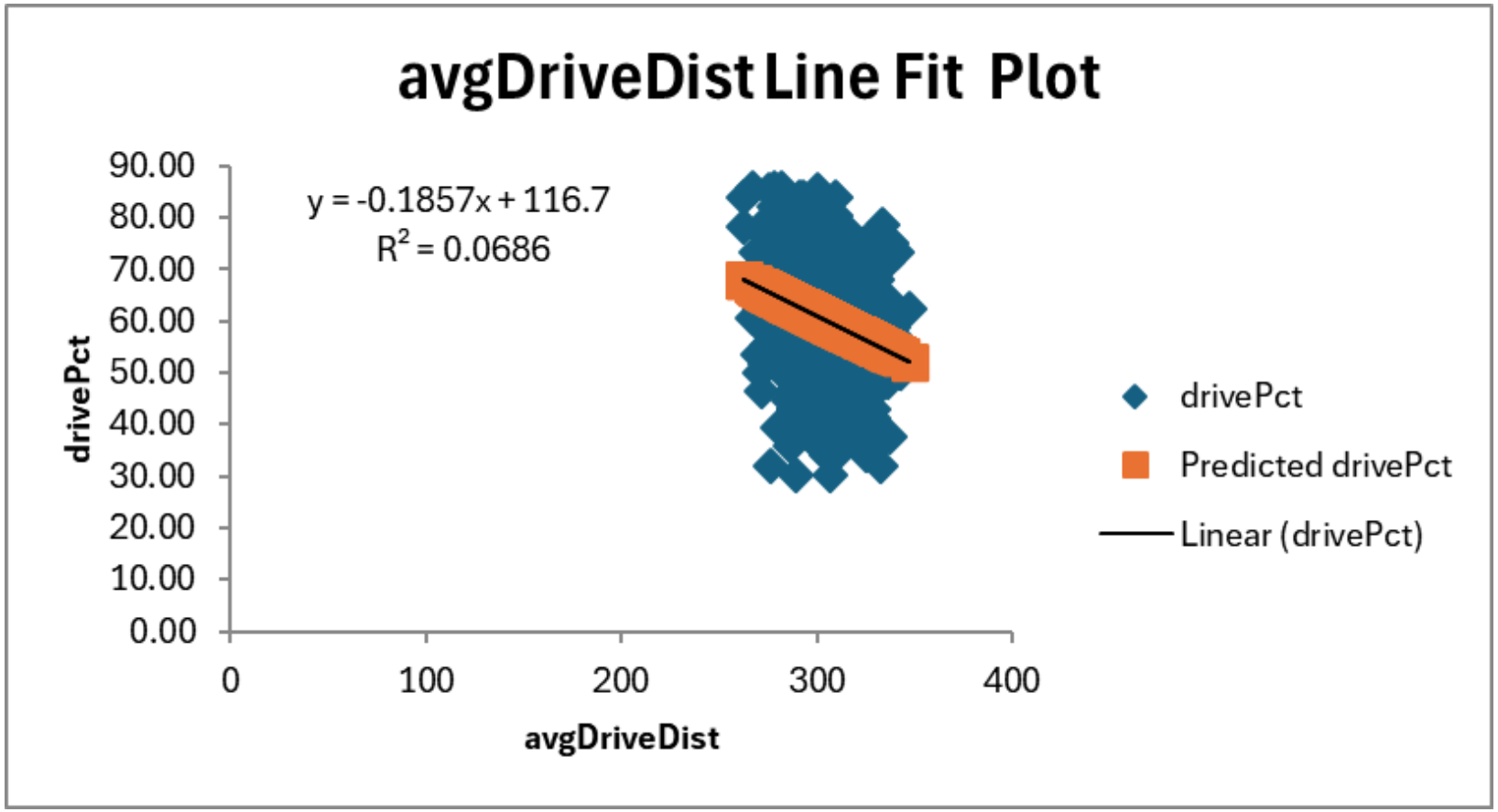
QUESTION 1


drivePct Line Fit Plot
y = -0.0775x + 45.327
R² = 0.0136
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
onePuttPct
0.00
50.00
100.00
drivePct
onePuttPct
Predicted onePuttPct
Linear (onePuttPct)

QUESTION 2

- As the players average distance per drive increases do their chances of hitting the fairway increase or decrease?
- It is important to analyze this relationship because it helps determine whether players are sacrificing accuracy for distance. While longer drives can give players a better chance of reaching the green in fewer strokes, they may also be more likely to land off the fairway.

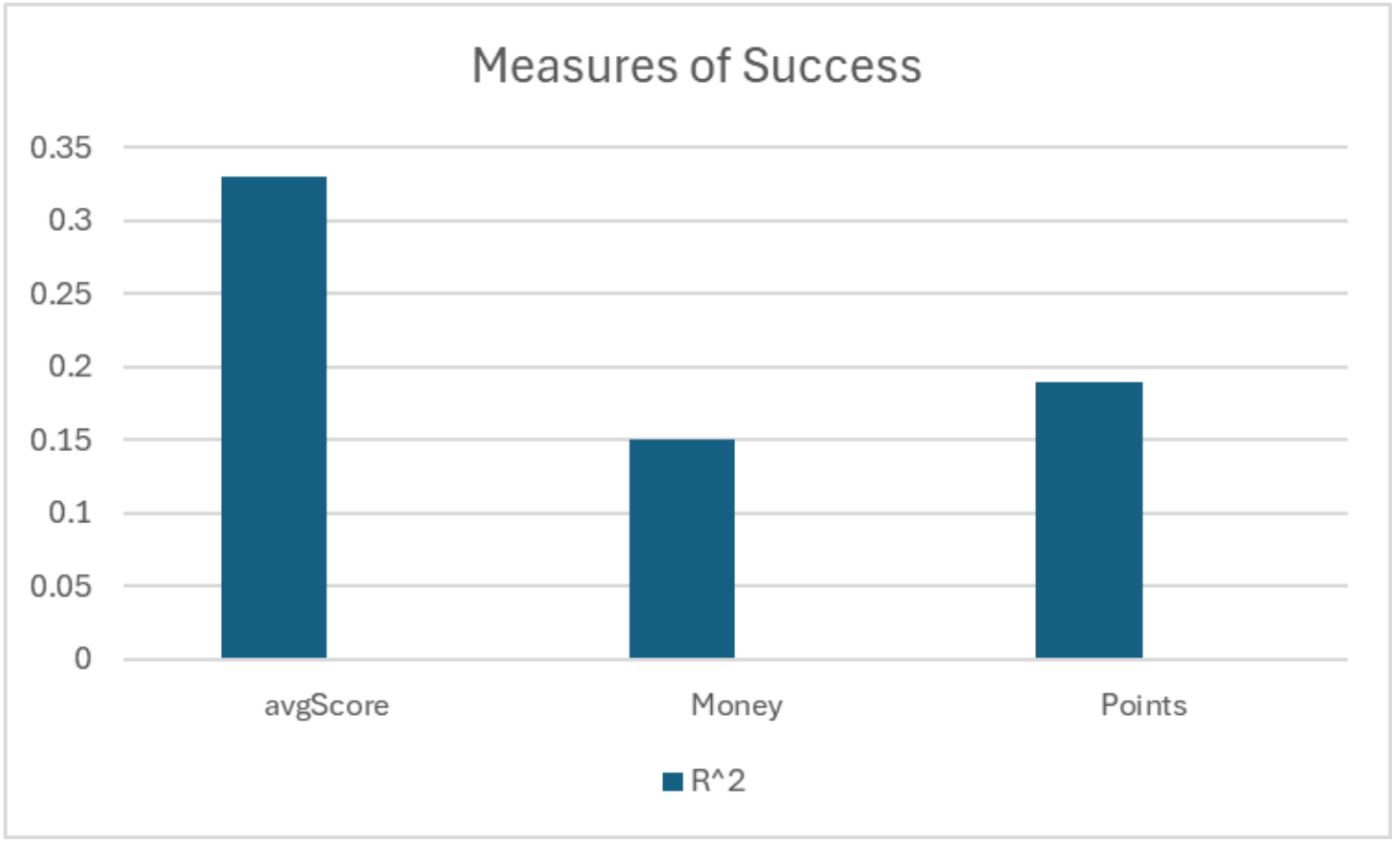
QUESTION 2


avgDriveDist Line Fit Plot
y = -0.1857x + 116.7
R² = 0.0686
drivePct
90.00
80.00
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
0
100
200
300
400
avgDriveDist
drivePct
Predicted drivePct
Linear (drivePct)

QUESTION 3

- Which measure of success (avgScore, Money, Points) appears to be easiest to predict with both driving and putting information?
- We will use driving and putting information to predict players score, money earned, and points gained on a specific weekend.
- This helps identify which part of a golfer's game, driving or putting, has the strongest impact on their performance and results. Knowing whether these stats better predict score, earnings, or points can guide training, strategy, and recruitment on the PGA Tour.
- Combining both usually improves predictions of success. If the R^2 value is high, it means driving and putting together explain much of the variation in avgScore, Money, or Points. If it's low, other factors likely play a bigger role in performance.

QUESTION 3


Measures of Success
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
avgScore
Money
Points
R^2

QUESTION 3
SIGNIFICANT PREDICTORS FOR AVGSCORE

	Coefficients	P-Value
avgDriveDist	-0.015994311	1.72945E-08
drivePct	-0.021961234	1.23953E-07
onePuttPct	0.058506274	6.6939E-07
avgPuttsPerRound	0.092830562	0.089911669
driveSG	-0.667408204	6.5233E-23
puttsSG	-0.852170544	1.49E-69

QUESTION 3
SIGNIFICANT PREDICTORS MONEY

	Coefficients	P-Value
avgDriveDist	-274.6846848	0.644098489
drivePct	-363.3995611	0.676508326
onePuttPct	-887.695053	0.719307012
avgPuttsPerRound	-17638.26968	0.12628391
driveSG	120422.4891	2.33331E-17
puttsSG	93847.69033	8.34071E-22

DRIVE FOR SHOW, PUTT FOR DOUGH, OR IS IT?

- Given the data from the many regression run our group thinks that driveSG has the largest impact on money earned.
- It has the highest positive coefficient (120,422), meaning better strokes gained from driving results in much more money earned.
- It is statistically significant with a very low p-value (2.33e-17), indicating a strong and reliable relationship.

CAN WE PREDICT IF A GOLFER MAKES THE CUT?

- Dataset: PGA Tour player metrics — strokes gained from:
  - Putting
  - Around the Green
  - Approach Shots
  - Off the Tee
  - Tee to Green
- We will use a Logistical Regression Model (Offers Strong Predictive Power)
- Output: Ln (Odds), Estimated Odds, P(BankApproves), and Model Decision, Accuracy

FIRST 15 GOLFERS ON LIST

- 93.33% ACCURACY IN PREDICTING IF A PLAYER MAKES THE CUT (FIRST 15). 85% ACCURACY WITH WHOLE LIST OF NAMES.
- STRONGEST PREDICTORS ARE TEE TO GREEN AND APPROACH SHOTS
-TO EXPAND THIS ANALYSIS, WE COULD SECTION OFF INTO DIFFERENT TOURNAMENTS TO SEE HOW THE MODEL DIFFERS.

Player Name	strokes gained from putting	strokes gained around the green	strokes gained on approach shots	strokes gained off the tee	strokes gained tee to green	cut	Ln(Odds)	Estimated Odds	P(BankApproves)	Model Decision	Accuracy
A. Ancer	0.2	-0.13	-0.08	0.86	0.65	1	0.42232957	1.525511205	0.604040561	1	Correct
A. Hadwin	0.36	0.75	0.31	0.18	1.24	1	0.460398109	1.584704745	0.613108614	1	Correct
A. Lahiri	-0.56	0.74	-1.09	0.37	0.02	0	0.385056645	1.469697571	0.595092123	0	Correct
A. Long	-1.46	-1.86	-0.02	0.8	-1.08	0	0.366332146	1.442434262	0.590572399	0	Correct
A. Noren	0.53	-0.36	-1.39	0.19	-1.56	0	0.391864614	1.479737361	0.596731486	0	Correct
A. Put0m	-0.97	0.14	-2.02	0.31	-1.56	0	0.339291696	1.403952813	0.584018457	0	Correct
A. Rai	2.05	0.74	-1.32	-0.12	-0.7	1	0.448142159	1.565401216	0.610197425	1	Correct
A. Schenk	-0.96	-0.01	1.84	0.48	2.31	1	0.469775434	1.599634929	0.615330603	1	Correct
A. Scott	-0.82	-1.79	2	-1.04	-0.83	1	0.483743629	1.622135724	0.618631488	1	Correct
A. Smalley	-1.89	-0.71	0.71	-0.65	-0.65	0	0.417276318	1.517821856	0.602831313	1	Wrong
A. Svensson	0.28	-0.02	-0.18	0.26	0.07	1	0.434268028	1.543832603	0.606892372	1	Correct
A. Wise	1.16	0.05	1.96	0.18	2.19	1	0.531033616	1.700689261	0.629724154	1	Correct
S. Theegala	0.06	0.4	1.48	0.65	2.54	1	0.486361663	1.626388094	0.619248959	1	Correct
B. Hagy	-0.94	-1.58	-1.41	-0.61	-3.6	0	0.354887762	1.426020592	0.587802345	0	Correct
B. Harman	0.12	0.66	0.17	0.65	1.48	1	0.441149778	1.554493514	0.608532966	1	Correct
											93.33%

QUESTIONS?