



Golf Technology Impact on Driving vs. Short Game

Group 4 - Amit Patel, Steven Lee, Roopa Patel, Samuel Parks, and Sean Jokhai

Motivation and Summary



- Golf club manufacturers know that all golfers want to hit the ball long and straight
- They aggressively develop and price drivers, promising longer distance and more accurate tee shots
- Driver is the longest club, with the biggest clubhead, designed to hit the ball the farthest off the tee
- Technology has been integral in driver development, utilizing among others:
 - Material science
 - Aerodynamics
 - Artificial Intelligence
- We want to know if driver performance has really improved over the years
- And if so, has technology also improved the performance of short-game clubs (used around the green, e.g. pitching, chipping and putting)

Motivation and Summary



We'll do this by looking at PGA Tour player stats from 2010-2018 and analyze their driving performance vs short-game performance

THESIS

- We believe driving performance is not related to short game performance. In other words, technology has mainly improved the driving aspect of golf

HYPOTHESIS

- Driving performance has improved over the years, with an accompanying increase in short-game performance

NULL HYPOTHESIS

- Driving performance has improved over the years, while short-game performance has remained the same
- Any increase in short-game performance is not due to driving performance

Motivation and Summary



HYPOTHESIS

- Driving performance has improved over the years, with an accompanying increase in short-game and putting performance

NULL HYPOTHESIS

- Driving performance has improved over the years, while short-game performance has remained the same
- Any increase in short-game performance is not due to driving performance

We were able to answer these questions and determined:

- We failed to reject the null hypothesis
- Driving performance increased, without a corresponding increase in short-game performance
- Any improvements in short-game performance were by random chance or a different factor



Data Cleanup & Exploration

- Input data sources:
 - Kaggle: PGA Historical Statistical Data for 2010-2018
 - 2,740,403 rows x 5 columns
 - Kaggle: PGA Player Scoring Averages for 1980-2019
 - 126,284 rows x 5 columns
- Data cleanup included:
 - Filtering for only the statistics we needed
 - Pivoting our dataset so the statistics would be columns
 - Dropped NaNs
 - Converted distance measurements like 8' 9" to inches
 - Cleaned currency values to remove \$s and commas
 - Converted data to correct datatypes

Data Cleanup & Exploration

- Our cleaned dataframe:

	Player Name	Season	Driving Distance - (AVG.)	Smash Factor - (AVG.)	Short Game Rating - (RATING)	Proximity to Hole (ARG) - (AVG DTP) (IN)	Putting Average - (AVG)	Total Money (Official and Unofficial) - (MONEY)	Years_on_Tour
0	Robert Garrigus	2010	315.5	1.475	5.5	103.0	1.786	1583275.0	13
1	Bubba Watson	2010	309.8	1.477	5.6	97.0	1.763	3655665.0	13
2	Dustin Johnson	2010	308.5	1.479	5.8	94.0	1.767	5280289.0	11
3	Brett Wetterich	2010	307.3	1.477	6.0	97.0	1.801	510672.0	9
4	J.B. Holmes	2010	307.2	1.480	6.2	84.0	1.770	2746248.0	13

- 1,654 rows, 430 unique players
- Player counts by Season:





Data Analysis - PGA Tour Statistics

We focused on 5 PGA Tour statistics. Lots more available at pgatour.com/stats

1. Driving Distance Average

- The average number of yards per measured drive. These drives are measured on two holes per round. Care is taken to select two holes which face in opposite directions to counteract the effect of wind. Drives are measured to the point at which they come to rest regardless of whether they are in the fairway or not

2. Smash Factor Average

- Ratio of Ball speed to Club head speed (Ball speed / Club speed) on Par 4 and Par 5 tee shots where a valid radar measurement was taken
- Player aspect: player's ability to generate clubhead speed
- Club aspect: transfer energy and convert clubhead speed to ball speed



Data Analysis - PGA Tour Statistics

3. Short Game Rating

- An aggregate score based on several statistics used to measure a player's performance 100 yards and in
- Does not include putting
- Includes: Scrambling, Proximity to the Hole from sand, from Rough, from fringe, from less than 100 yards, and Greens in Regulation from less than 100 yards

4. Proximity to Hole (Around the Green), Average Distance to Hole

- The average distance to the hole (in feet) after hitting the ball onto the putting surface from around the green (within 30 yards of the edge of the green)

5. Putting Average

- The average number of putts per green in regulation.



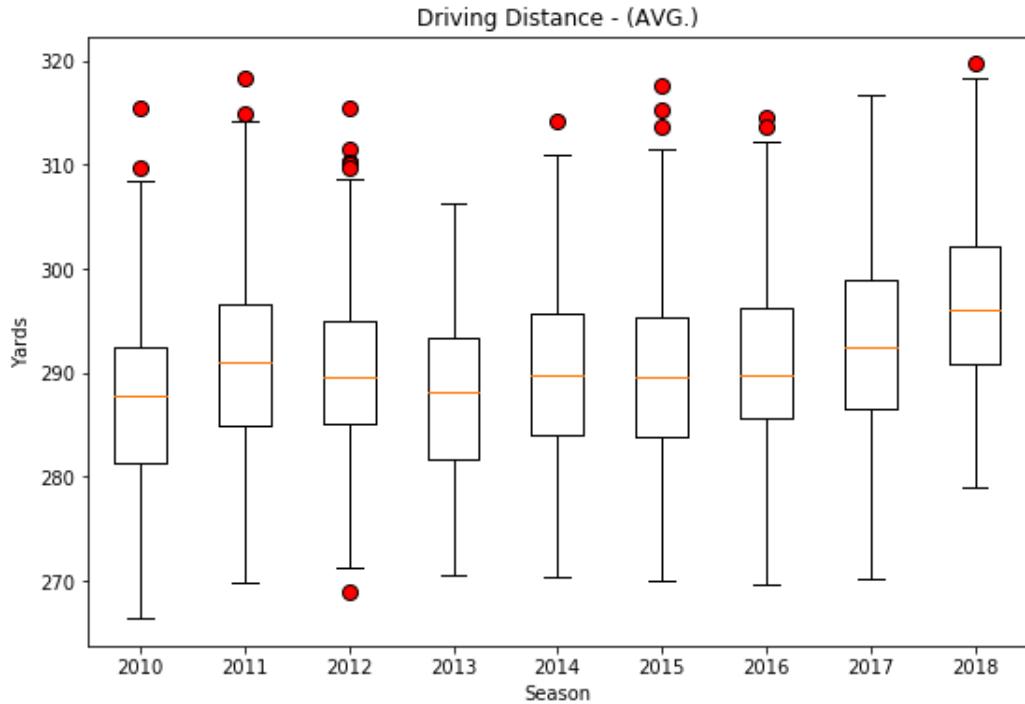
Data Analysis - Our Approach

- We first analyzed each statistic separately to determine if there was any improvement from 2010-2018
- We then looked for any correlations between the statistics
- We examined some individual players and groupings of players to gain more insight
- Our analysis utilized:
 - Summary stat tables (mean, median, mode, variance, std dev, SEM, normaltest)
 - Boxplots
 - Density plots
 - One-way ANOVA
 - Tukey's HSD (honestly significant difference) test
 - Scatter plots and scatter matrix
 - Correlation table

Driving Distance



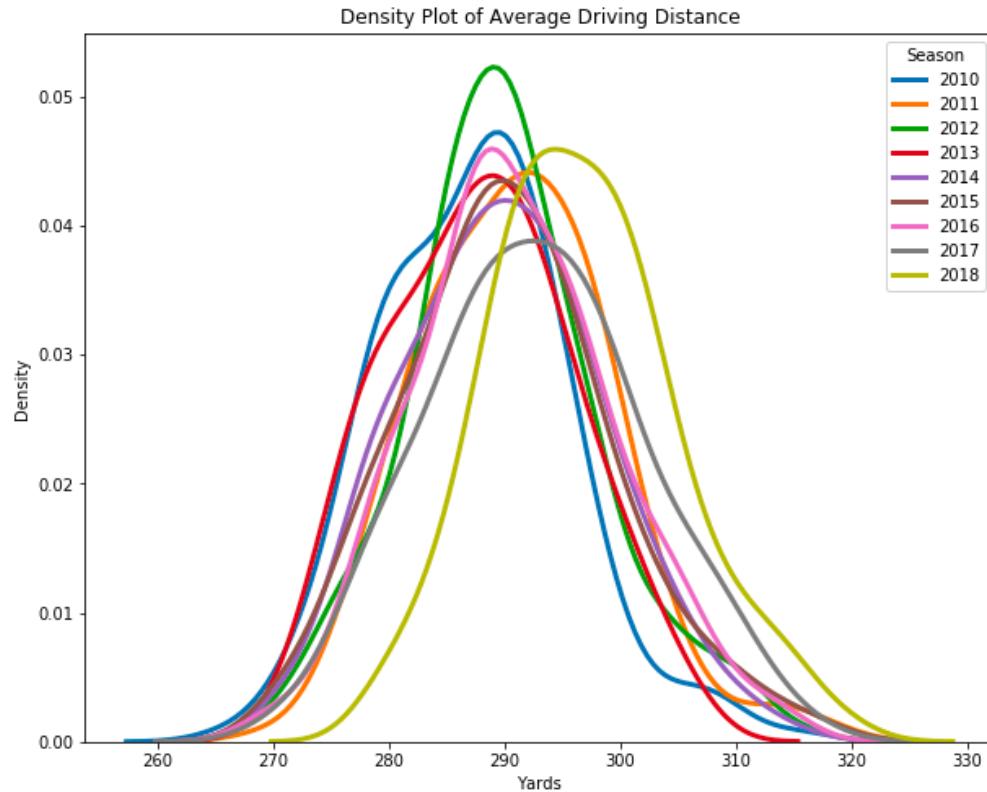
- What can we conclude off of this boxplot?
- Where do we see the improvement?



Driving distance



- Why did we use a density plot over a histogram?
- What information did we gather from this particular plot?



Driving Distance



- Tukey's HSD Test

Multiple Comparison of Means – Tukey HSD, FWER=0.05

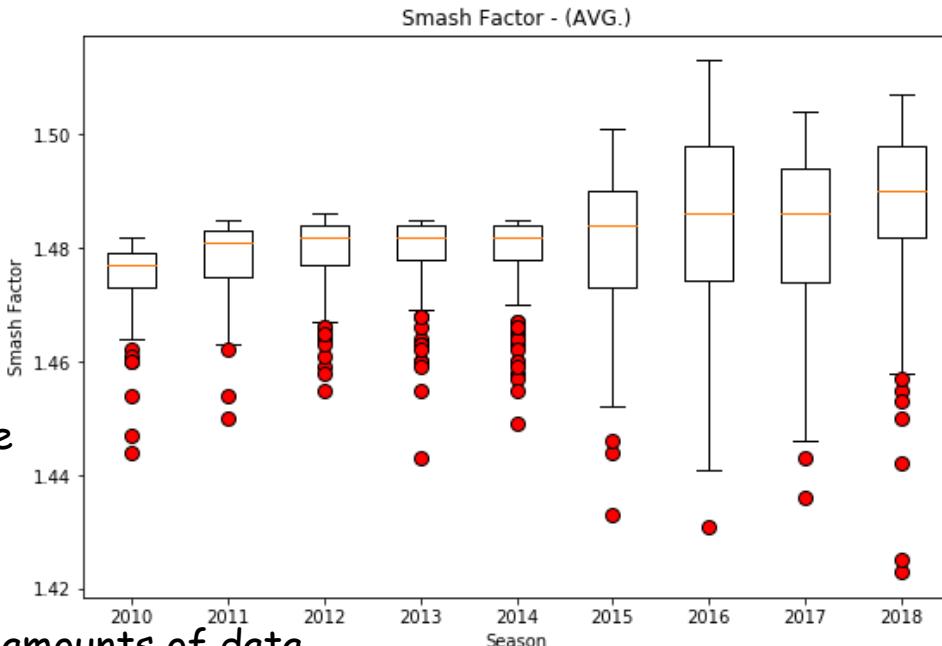
group1	group2	meandiff	p-adj	lower	upper	reject
2010	2011	3.6557	0.0013	0.9037	6.4077	True
2010	2012	2.6116	0.0739	-0.1182	5.3414	False
2010	2013	0.4871	0.9	-2.2884	3.2625	False
2010	2014	2.5159	0.1167	-0.276	5.3078	False
2010	2015	2.8673	0.0356	0.0998	5.6348	True
2010	2016	3.4864	0.0029	0.7268	6.2461	True
2010	2017	5.3003	0.001	2.5521	8.0485	True
2010	2018	9.3278	0.001	6.6015	12.054	True
2011	2012	-1.0441	0.9	-3.7961	1.7079	False
2011	2013	-3.1686	0.0132	-5.9659	-0.3714	True
2011	2014	-1.1398	0.9	-3.9534	1.6737	False
2011	2015	-0.7884	0.9	-3.5777	2.0009	False
2011	2016	-0.1693	0.9	-2.9508	2.6123	False
2011	2017	1.6446	0.6318	-1.1256	4.4148	False
2011	2018	5.6721	0.001	2.9237	8.4205	True
2012	2013	-2.1245	0.2972	-4.9	0.6509	False
2012	2014	-0.0957	0.9	-2.8876	2.6962	False
2012	2015	0.2557	0.9	-2.5117	3.0232	False
2012	2016	0.8749	0.9	-1.8848	3.6345	False
2012	2017	2.6887	0.061	-0.0595	5.4369	False
2012	2018	6.7162	0.001	3.99	9.4425	True
2013	2014	2.0288	0.3944	-0.8077	4.8653	False
2013	2015	2.3802	0.1752	-0.4323	5.1927	False
2013	2016	2.9994	0.0256	0.1946	5.8042	True
2013	2017	4.8132	0.001	2.0197	7.6068	True
2013	2018	8.8407	0.001	6.0688	11.6126	True
2014	2015	0.3514	0.9	-2.4773	3.1801	False
2014	2016	0.9706	0.9	-1.8505	3.7916	False
2014	2017	2.7844	0.0544	-0.0254	5.5943	False
2014	2018	6.8119	0.001	4.0235	9.6003	True
2015	2016	0.6191	0.9	-2.1778	3.416	False
2015	2017	2.433	0.1438	-0.3526	5.2186	False
2015	2018	6.4605	0.001	3.6965	9.2244	True
2016	2017	1.8139	0.5202	-0.964	4.5917	False
2016	2018	5.8413	0.001	3.0852	8.5975	True
2017	2018	4.0275	0.001	1.2828	6.7721	True

Smash Factor



- Calculation of club head speed and ball speed
 - Smash Factor = Ball Speed/ Club Speed
- The higher the ratio is the better
- 1.5 gold standard

Smash Factor helps us understand the amount of energy transferred from the club head to the ball at impact

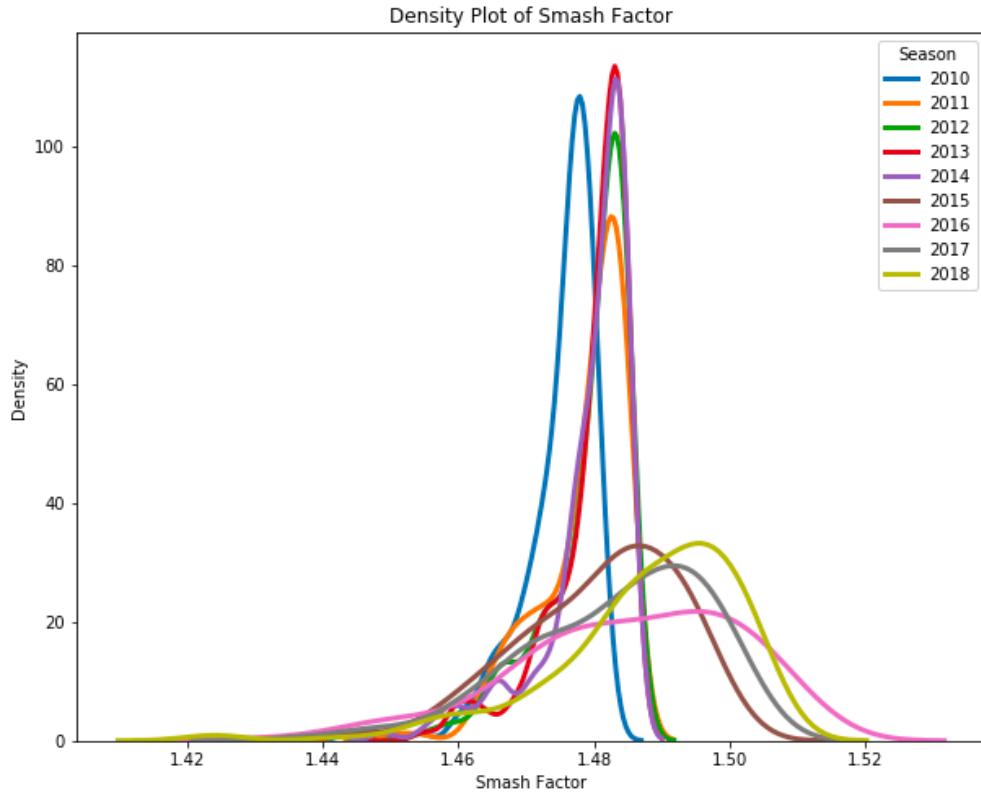


- Provide a clear summary of large amounts of data
- 2010 to 2014 static years with gradual improvement
- 2015 to 2018 less steep slope but improvement in key number

Smash Factor



- The density plot visualizes the data over a continuous variable of time period
- Variation of a histogram
- 2013-2014 Crowded mean
- 2015 Breakout improvement



Smash Factor



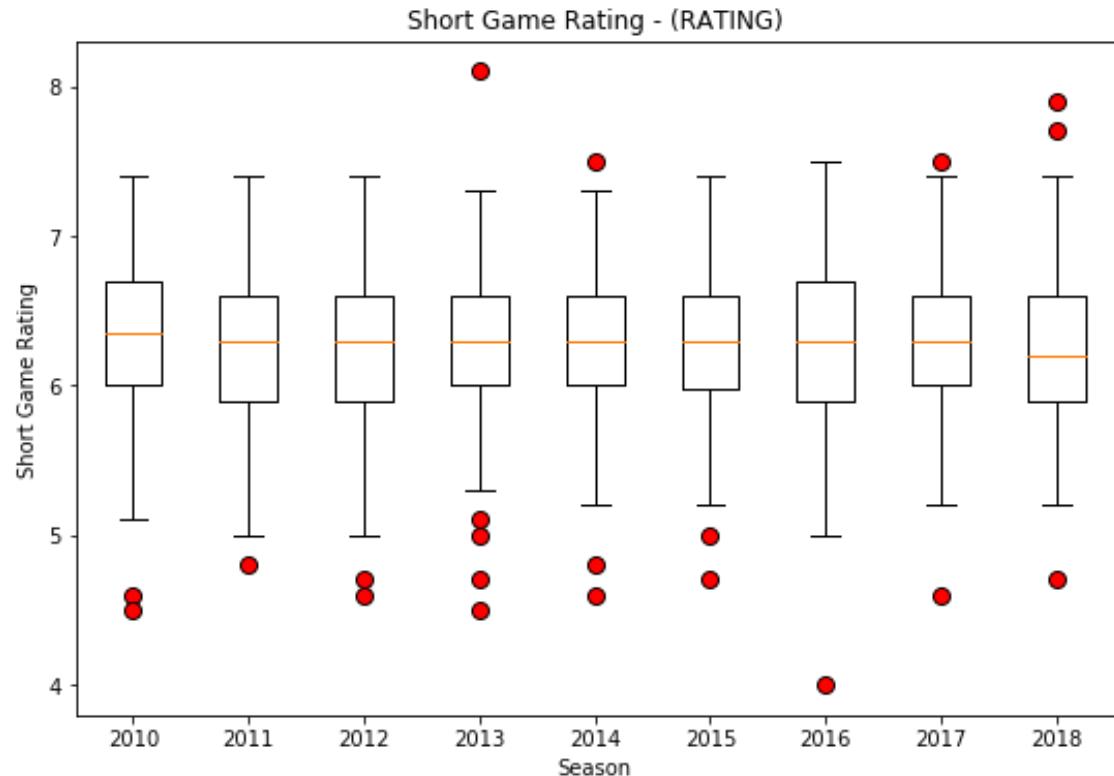
- Tukey's HSD Test

Multiple Comparison of Means – Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
2010	2011	0.0035	0.0394	0.0001	0.0069	True
2010	2012	0.0042	0.0034	0.0008	0.0076	True
2010	2013	0.0044	0.003	0.0009	0.0078	True
2010	2014	0.0041	0.0069	0.0007	0.0076	True
2010	2015	0.0059	0.001	0.0025	0.0093	True
2010	2016	0.0103	0.001	0.0069	0.0138	True
2010	2017	0.0084	0.001	0.005	0.0118	True
2010	2018	0.0128	0.001	0.0094	0.0162	True
2011	2012	0.0007	0.9	-0.0027	0.0042	False
2011	2013	0.0008	0.9	-0.0026	0.0043	False
2011	2014	0.0006	0.9	-0.0029	0.0041	False
2011	2015	0.0024	0.4552	-0.0011	0.0059	False
2011	2016	0.0068	0.001	0.0033	0.0103	True
2011	2017	0.0049	0.001	0.0014	0.0083	True
2011	2018	0.0093	0.001	0.0059	0.0127	True
2012	2013	0.0001	0.9	-0.0033	0.0036	False
2012	2014	-0.0001	0.9	-0.0036	0.0034	False
2012	2015	0.0016	0.8481	-0.0018	0.0051	False
2012	2016	0.0061	0.001	0.0026	0.0095	True
2012	2017	0.0041	0.0054	0.0007	0.0076	True
2012	2018	0.0086	0.001	0.0052	0.012	True
2013	2014	-0.0002	0.9	-0.0038	0.0033	False
2013	2015	0.0015	0.9	-0.002	0.005	False
2013	2016	0.006	0.001	0.0025	0.0095	True
2013	2017	0.004	0.0099	0.0006	0.0075	True
2013	2018	0.0085	0.001	0.005	0.0119	True
2014	2015	0.0018	0.8094	-0.0018	0.0053	False
2014	2016	0.0062	0.001	0.0027	0.0097	True
2014	2017	0.0043	0.0051	0.0008	0.0078	True
2014	2018	0.0087	0.001	0.0052	0.0122	True
2015	2016	0.0044	0.0027	0.0009	0.0079	True
2015	2017	0.0025	0.385	-0.001	0.006	False
2015	2018	0.0069	0.001	0.0035	0.0104	True
2016	2017	-0.0019	0.6996	-0.0054	0.0015	False
2016	2018	0.0025	0.3634	-0.0009	0.0059	False
2017	2018	0.0044	0.0019	0.001	0.0079	True

Short Game Rating

- Not much change across the years



Short Game Rating

- Based on these statistics we see little difference

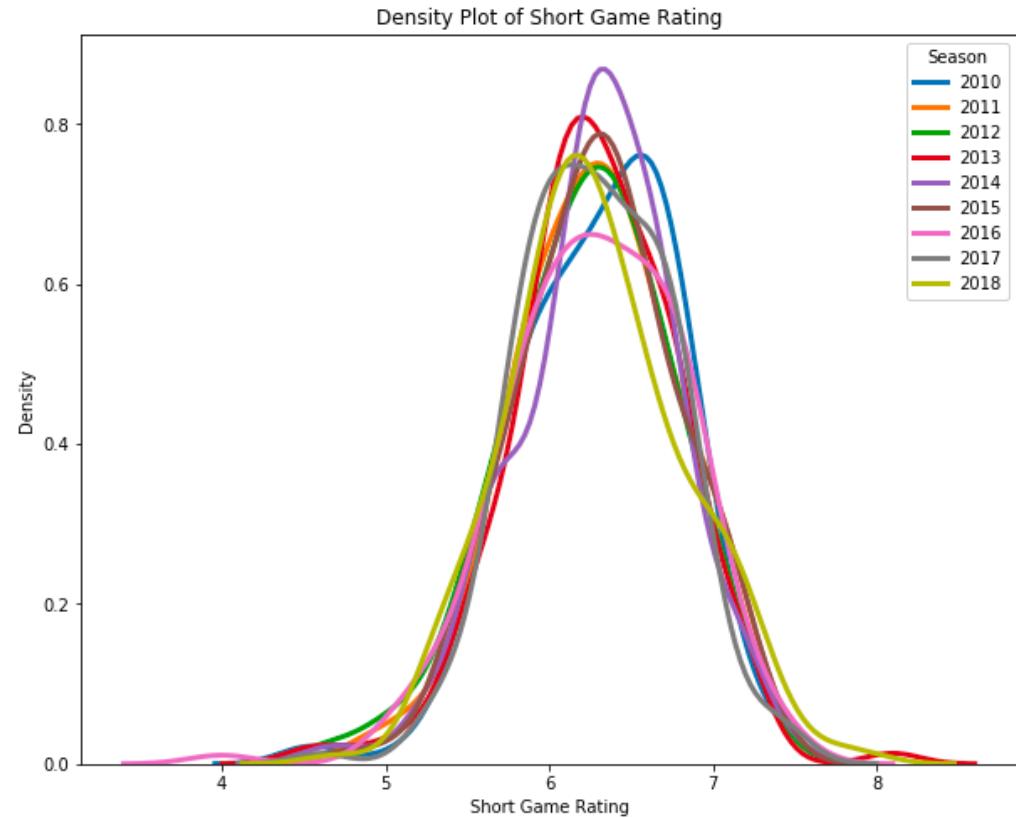
Season	Mean Short Game Rating	Median Short Game Rating	Variance Short Game Rating	Std Dev Short Game Rating	SEM Short Game Rating	Normality (p) Short Game Rating
2010	6.313158	6.35	0.236616	0.486432	0.035383	0.002933
2011	6.288587	6.30	0.244381	0.494349	0.036543	0.532774
2012	6.240000	6.30	0.263558	0.513379	0.037343	0.073532
2013	6.286517	6.30	0.251728	0.501725	0.037712	0.017887
2014	6.294253	6.30	0.235714	0.485504	0.036912	0.026615
2015	6.294444	6.30	0.242302	0.492242	0.036792	0.616054
2016	6.279121	6.30	0.285059	0.533909	0.039685	0.002451
2017	6.285405	6.30	0.213571	0.462137	0.034069	0.522961
2018	6.279581	6.20	0.287175	0.535887	0.038877	0.500227



Short Game Rating



- Not much change over the years



.



Short Game Rating

- ANOVA test
 - P-Value was .97
- Tukey's Test

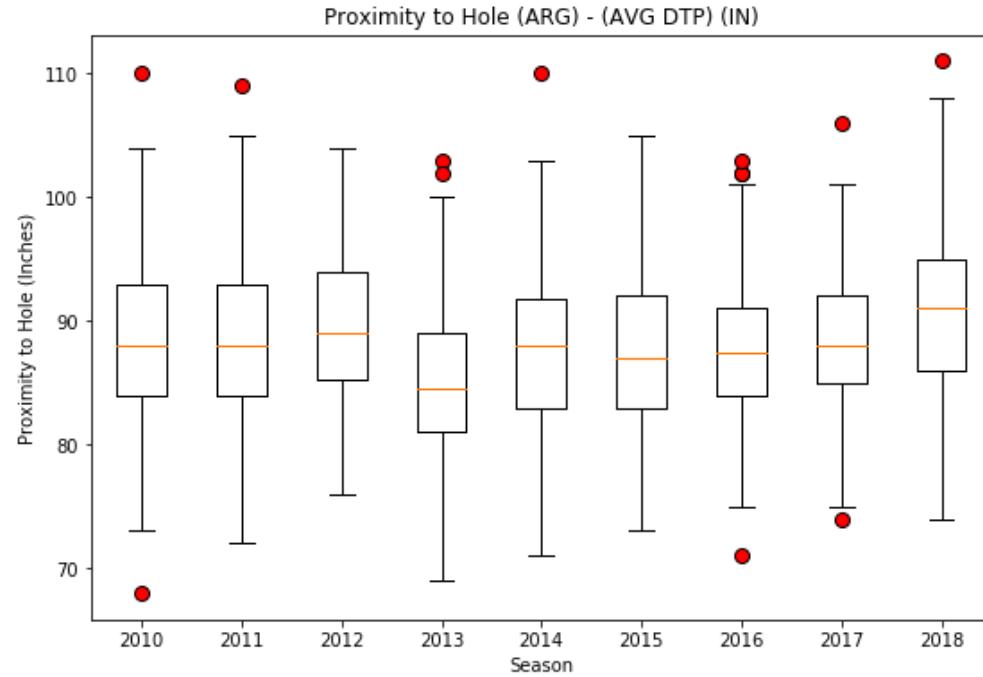
Multiple Comparison of Means - Tukey HSD, FWER=0.05						
group1	group2	meandiff	p-adj	lower	upper	reject
2010	2011	-0.0246	0.9	-0.1861	0.1369	False
2010	2012	-0.0732	0.8895	-0.2333	0.087	False
2010	2013	-0.0266	0.9	-0.1895	0.1362	False
2010	2014	-0.0189	0.9	-0.1827	0.1449	False
2010	2015	-0.0187	0.9	-0.1811	0.1437	False
2010	2016	-0.034	0.9	-0.196	0.1279	False
2010	2017	-0.0278	0.9	-0.189	0.1335	False
2010	2018	-0.0336	0.9	-0.1936	0.1264	False
2011	2012	-0.0486	0.9	-0.2101	0.1129	False
2011	2013	-0.0021	0.9	-0.1662	0.1621	False
2011	2014	0.0057	0.9	-0.1594	0.1708	False
2011	2015	0.0059	0.9	-0.1578	0.1695	False
2011	2016	-0.0095	0.9	-0.1727	0.1538	False
2011	2017	-0.0032	0.9	-0.1657	0.1594	False
2011	2018	-0.009	0.9	-0.1703	0.1523	False



Short Game Rating Conclusion

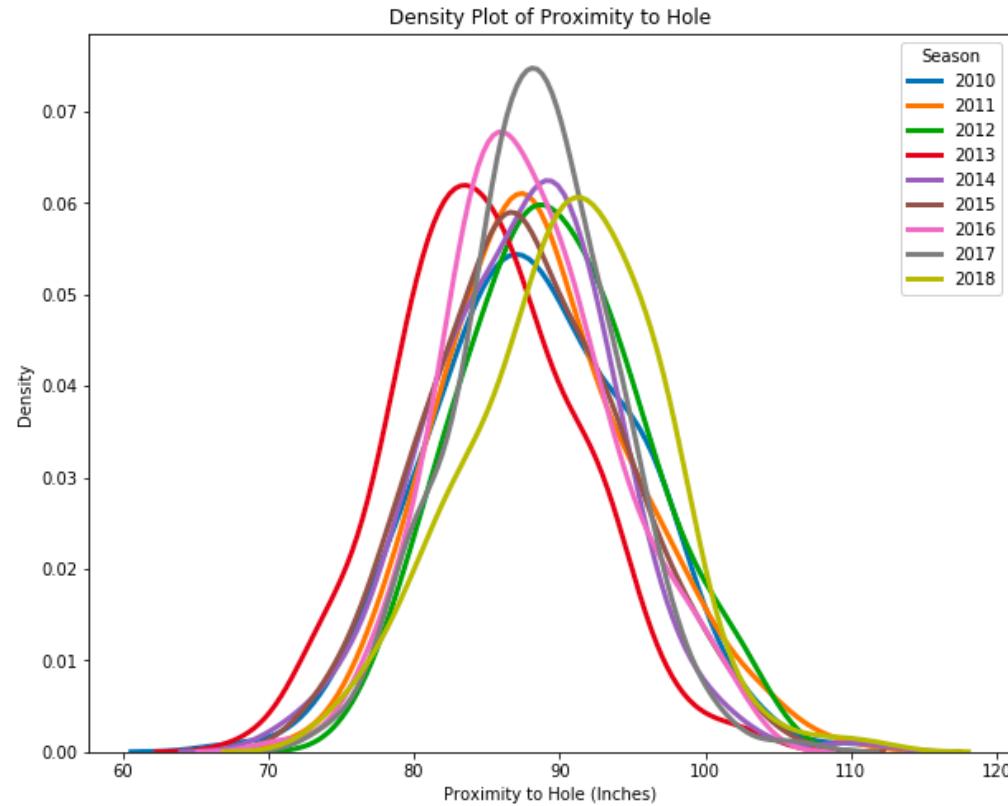
- The One-Way ANOVA and Tukey's tests show that there are no significant differences among the means
- Therefore, the Short Game Rating has not improved

Proximity to Hole (ARG)



Proximity to Hole (ARG)

- Actually got worse in 2018?





Proximity to Hole (ARG)

- ANOVA test
 - P-Value was $4.25e-15$
- Tukey's Test

Multiple Comparison of Means - Tukey HSD, FWER=0.05

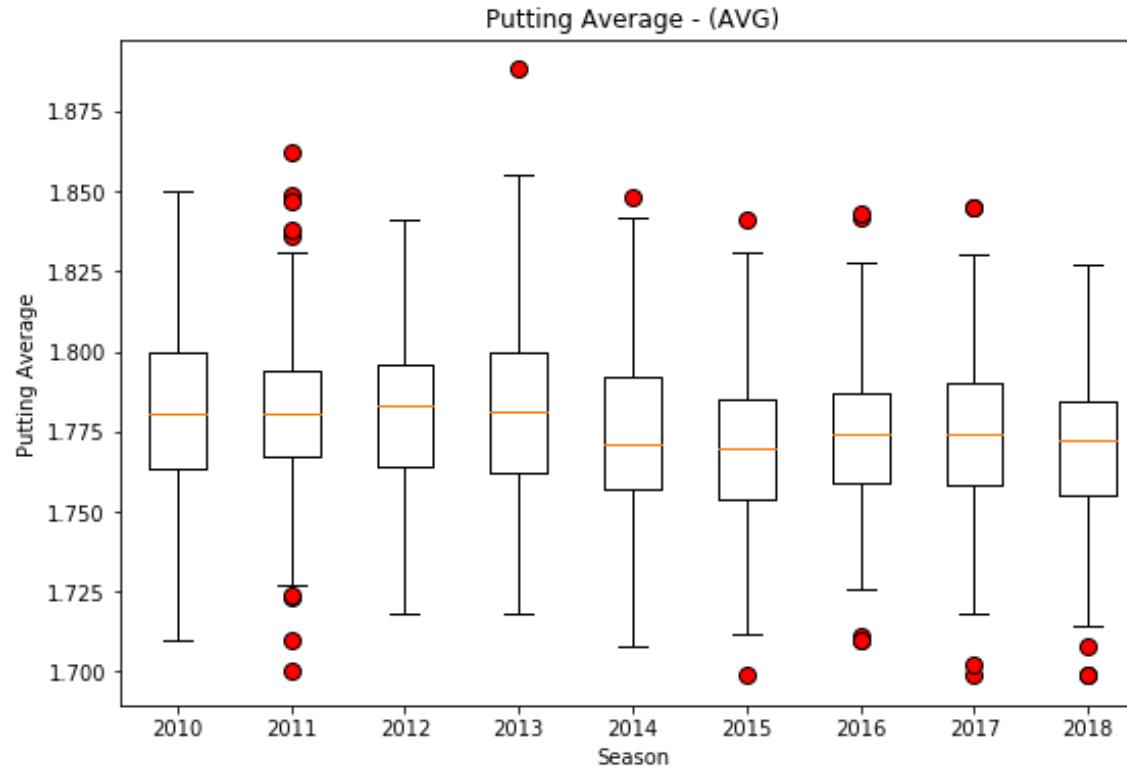
group1	group2	meandiff	p-adj	lower	upper	reject
2010	2011	0.3892	0.9	-1.6145	2.3929	False
2010	2012	1.3842	0.4336	-0.6033	3.3717	False
2010	2013	-3.3583	0.001	-5.3791	-1.3376	True
2010	2014	-0.8087	0.9	-2.8414	1.2241	False
2010	2015	-0.6073	0.9	-2.6223	1.4076	False
2010	2016	-0.3069	0.9	-2.3161	1.7024	False
2010	2017	0.0721	0.9	-1.9288	2.073	False
2010	2018	2.0667	0.0339	0.0817	4.0516	True
2011	2012	0.995	0.8145	-1.0086	2.9987	False
2011	2013	-3.7475	0.001	-5.7841	-1.7109	True
2011	2014	-1.1978	0.6486	-3.2463	0.8506	False
2011	2015	-0.9965	0.8256	-3.0274	1.0344	False
2011	2016	-0.6961	0.9	-2.7213	1.3291	False
2011	2017	-0.3171	0.9	-2.334	1.6999	False
2011	2018	1.6775	0.1855	-0.3236	3.6786	False
2012	2013	-4.7425	0.001	-6.7633	-2.7218	True
2012	2014	-2.1929	0.0233	-4.2256	-0.1602	True



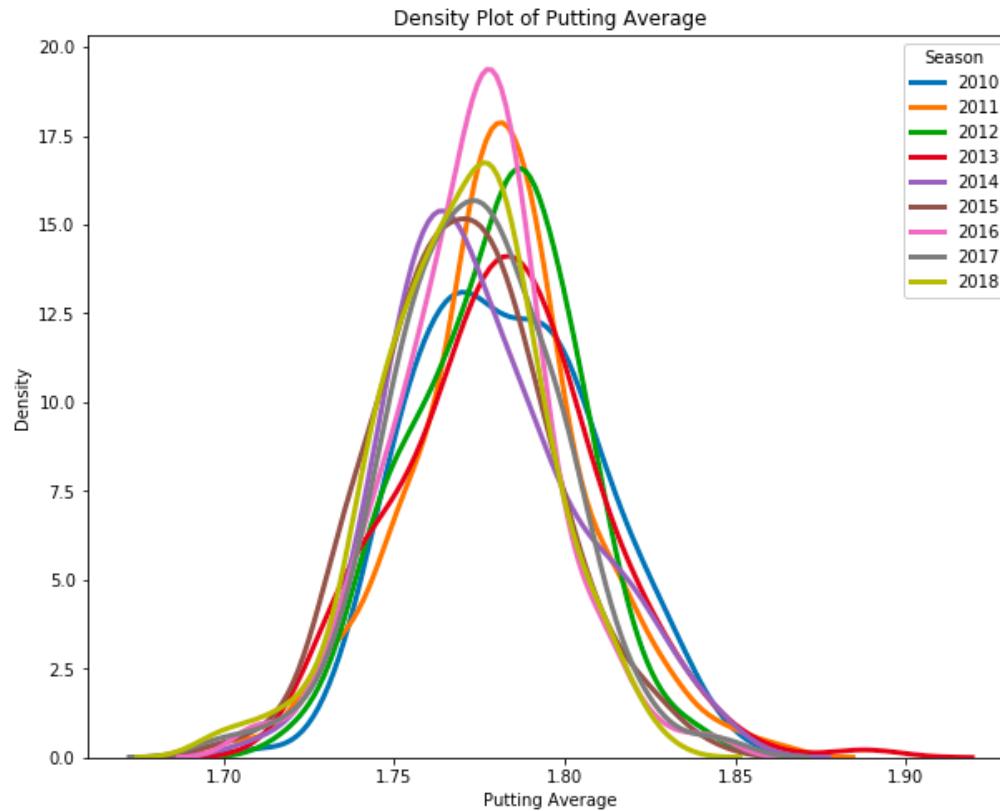
Proximity to Hole (ARG) Conclusion

- The One-Way ANOVA and Tukey's tests show that there are significant differences among the means
- This aspect of the short game has not improved

Putting Average - Boxplot



Putting Average - Density Plot





Putting Average - ANOVA and Tukey

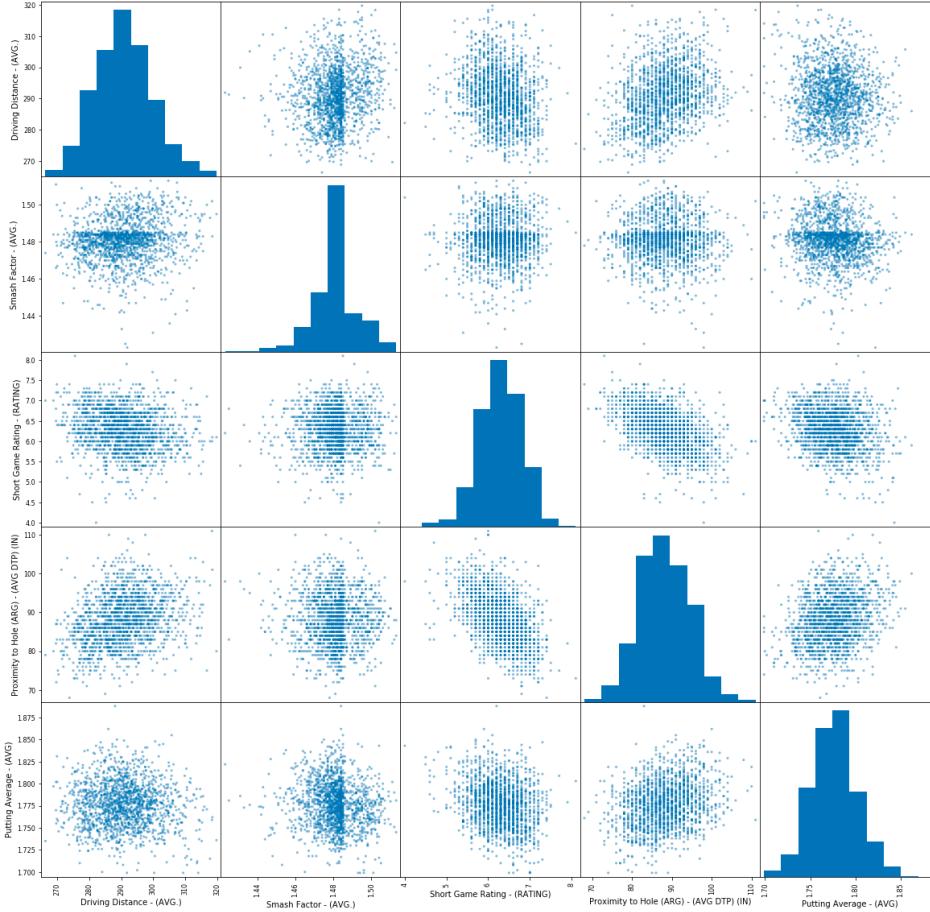
- One-Way ANOVA
 - F-Statistic = 6.991
 - P-Value = .00000004218
 - The P-value is less than 0.05. Therefore, there are significant differences among the means
- Tukey's Test
 - 10 of the 36 pairwise comparisons between seasons show that the P-value is less than 0.05. Therefore, there are significant differences between those pairs of means



Putting Average - Conclusion

- The One-Way ANOVA and Tukey show that there are significant differences among the averages
- The box plot and density plots show that there is no clear trend over the seasons
- Thus, there appears to be significant oscillations of the putting average between seasons, but no clear trend over time
- The putting element of the PGA Tour does not seem to be impacted by improved technology between the 2010 to 2018 seasons

Scatter Matrix - Looking for correlations and trends

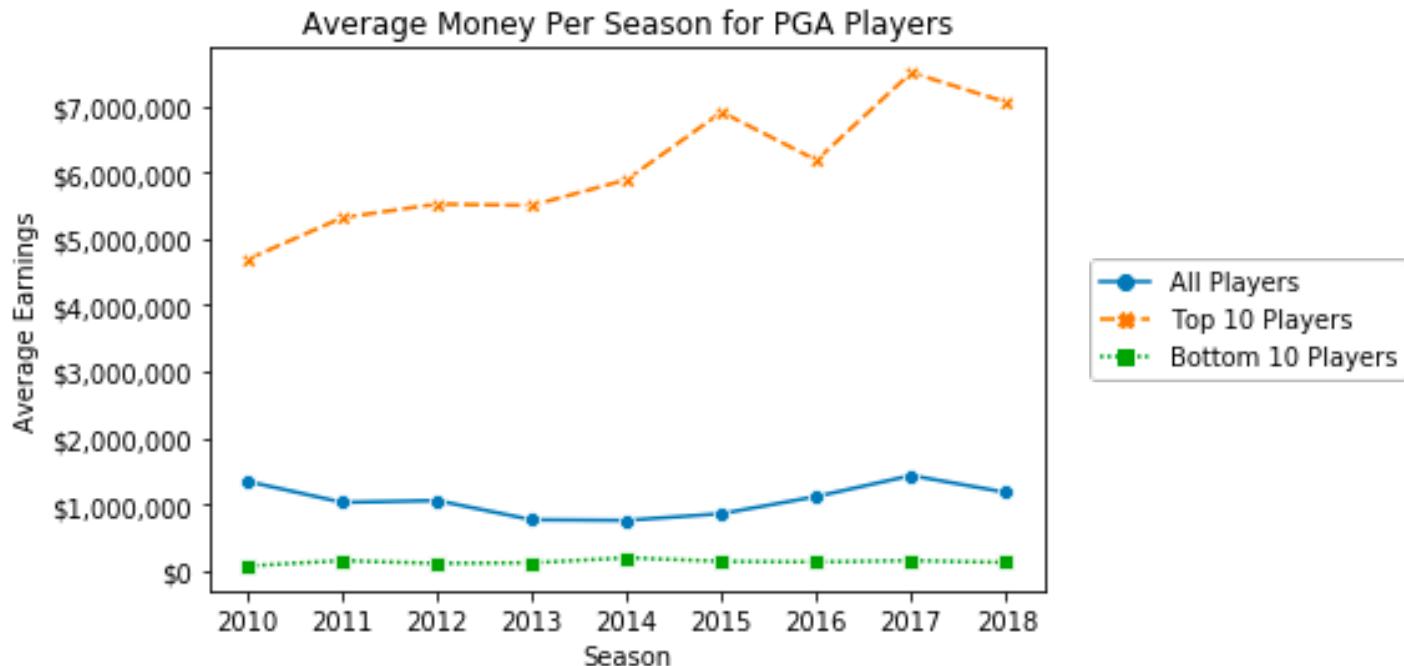


Correlation Table



	Driving Distance - (AVG.)	Smash Factor - (AVG.)	Short Game Rating - (RATING)	Proximity to Hole (ARG) - (AVG DTP) (IN)	Putting Average - (AVG)	Total Money (Official and Unofficial) - (MONEY)
Driving Distance - (AVG.)	1.00	0.12	-0.19	0.29	-0.05	0.33
Smash Factor - (AVG.)	0.12	1.00	0.03	-0.03	-0.11	0.06
Short Game Rating - (RATING)	-0.19	0.03	1.00	-0.51	-0.24	0.20
Proximity to Hole (ARG) - (AVG DTP) (IN)	0.29	-0.03	-0.51	1.00	0.26	-0.19
Putting Average - (AVG)	-0.05	-0.11	-0.24	0.26	1.00	-0.46
Total Money (Official and Unofficial) - (MONEY)	0.33	0.06	0.20	-0.19	-0.46	1.00

Money Earned

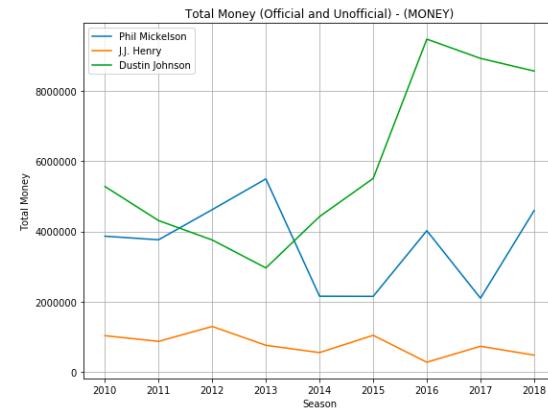
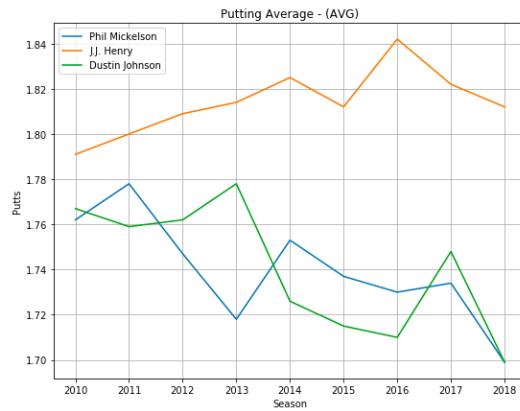
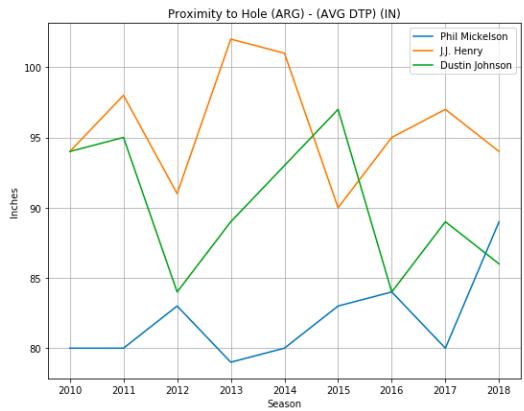
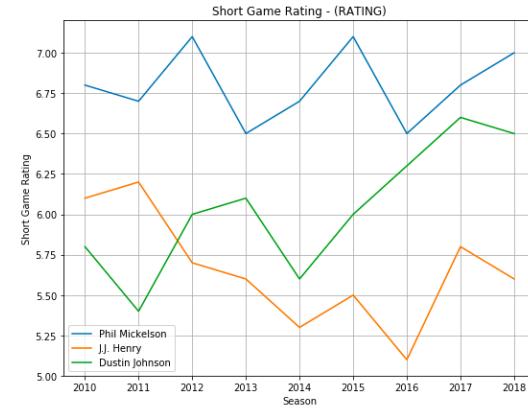
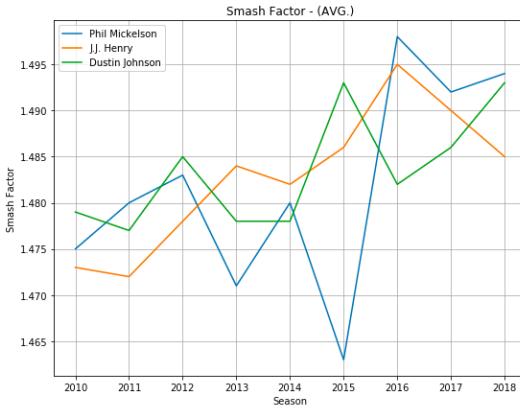
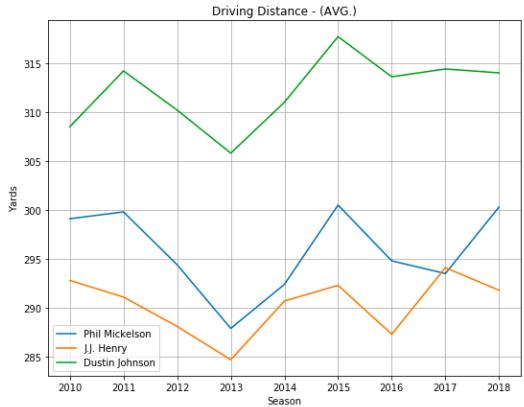


Money Earned and Performance



<u>Correlation</u>	Total Money (Official and Unofficial)
Driving Distance	0.33
Smash Factor	0.06
Short Game Rating	0.20
Proximity to Hole	-0.19
Putting Average	-0.46

Selected Players' Performance





Discussion of Findings

- We started the project with a belief in the null hypothesis, that driving improvements would not correlate to short-game improvements
- Our findings align with this initial belief
- The biggest improvements were in:
 - Driving distance, especially since 2016
 - Smash factor
 - Technology's biggest overall impact has been more ball speed
 - For the Tour players, this means getting the most performance out of their swings and clubs
- The short-game statistics did not show significant improvements, if any at all

Post Mortem

- Discuss any difficulties that arose, and how you dealt with them.



The image shows a white golf ball with blue and red markings resting on a patch of green grass. In the background, there's a light blue sky with some wispy clouds. The main focus is a screenshot of a project management interface divided into five horizontal sections: Ice Box, To do, In progress, Review, and Done.

- Ice Box (ideas that we can do)**: Contains one item: "Include what technological changes there were" (Added by Roopa-16).
- To do**: Contains two items: "Enter a note" (button: Add, Cancel) and "Project Requirements": "Create 6-8 visualizations" and "Create Presentation" (Added by Roopa-16).
- In progress**: Contains three items: "Create a detailed ReadMe file" (#24 opened by Roopa-16), "Do a Mock Presentation" (Added by Roopa-16), and "Presentation!!" (Added by Roopa-16).
- Review (by at least 2 others) then push**: Contains zero items.
- Done**: Contains nine items:
 - "Proposal!" (#10 opened by Roopa-16)
 - "Init files" (#22 opened by Roopa-16)
 - "Conversions for money and data type" (#21 opened by Roopa-16)
 - "Driving Distance" (#17 opened by Roopa-16)
 - "Short Game Rating Graphs" (#19 opened by Roopa-16)
 - "Project Requirements" (0 of 2, #12 opened by Roopa-16)
 - "Smash Factor Graphs" (#18 opened by Roopa-16)
 - "Putting average and Proximity to..."

Post Mortem

- Discuss any additional questions that came up, but which you didn't have time to answer: What would you research next, if you had two more weeks



TAYLORMADE NEW SHAPE OF PERFORMANCE WITH SIM METALWOODS

Dynamic Shaping and Unconventional Geometry Unlock the Next Level of Speed and Forgiveness



Callaway Mavrik metalwoods use artificial intelligence to push face and head designs to optimize specific player needs

By Mike Stachura

Share this story: [f](#) [t](#) [in](#)



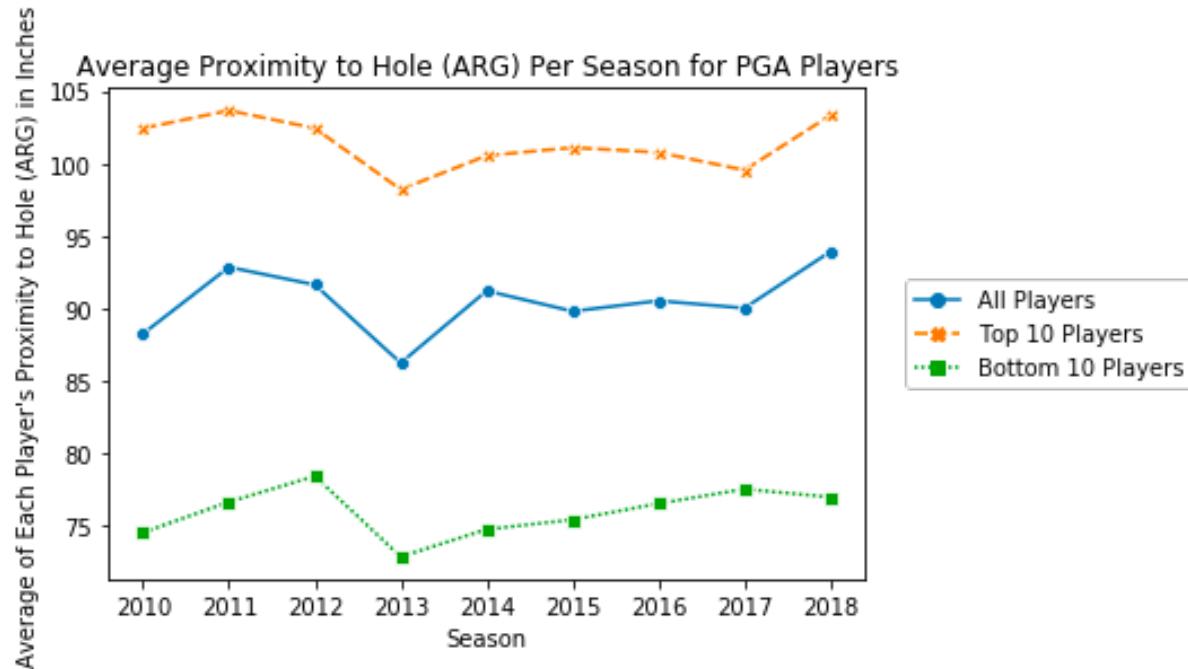
Chris Otsen

The Callaway Mavrik line of metalwoods certainly takes a cue from its name. The designs are a "maverick" kind of rebuke to past ideas about metalwood designs, fueled by an explosive expansion of the company's foray into artificial intelligence and machine learning. But just like the name removes unnecessary letters, these designs

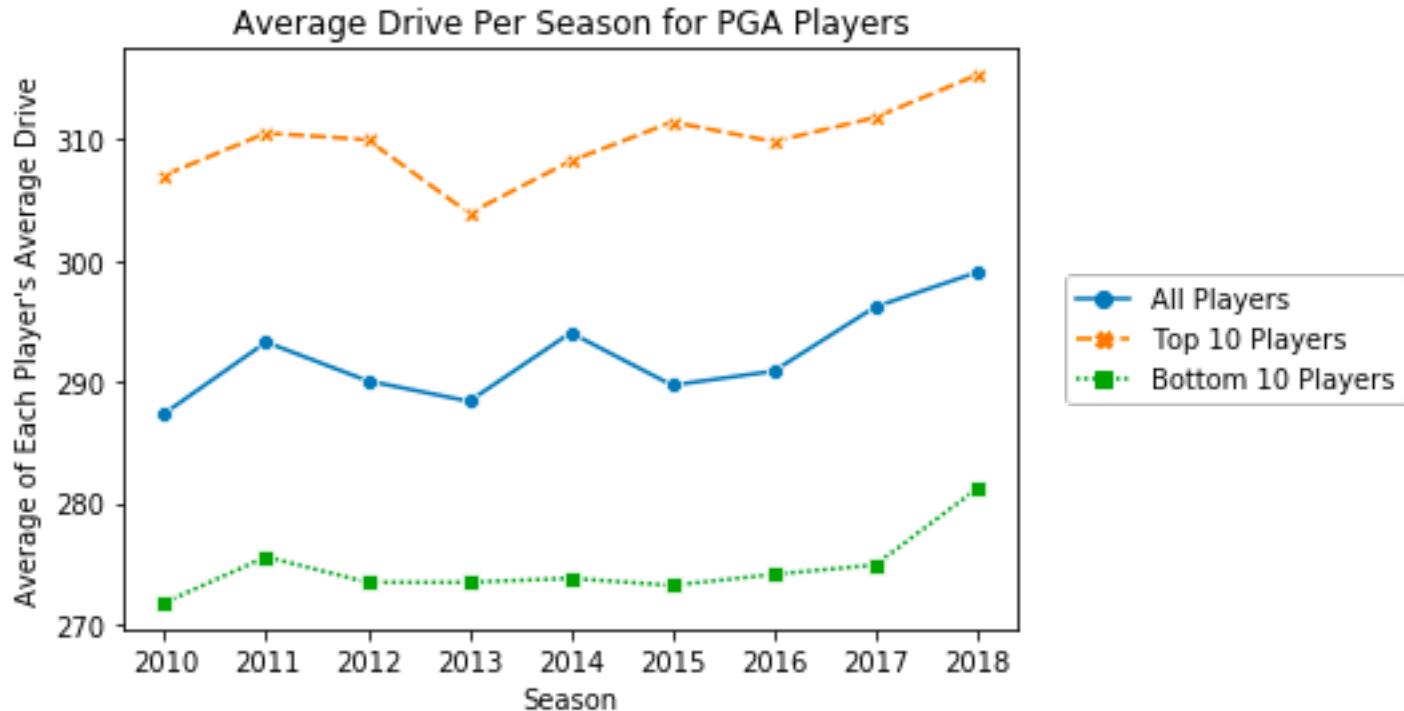
Questions?



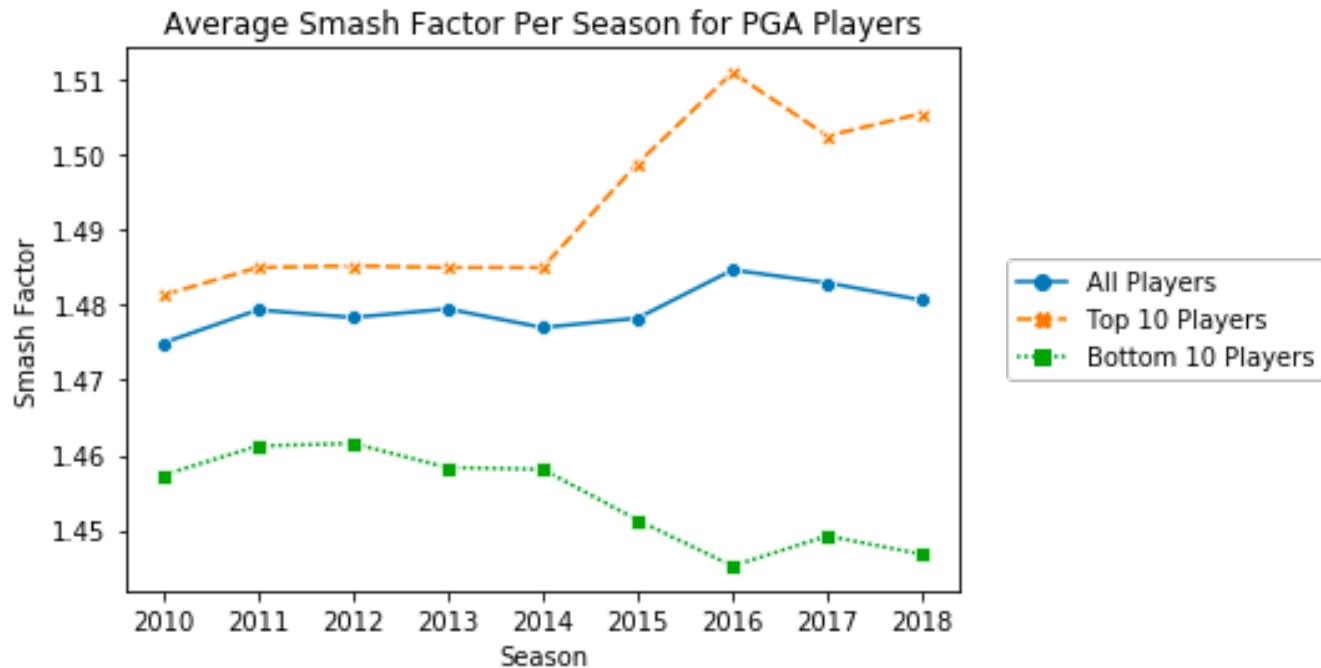
Appendix



Appendix



Appendix

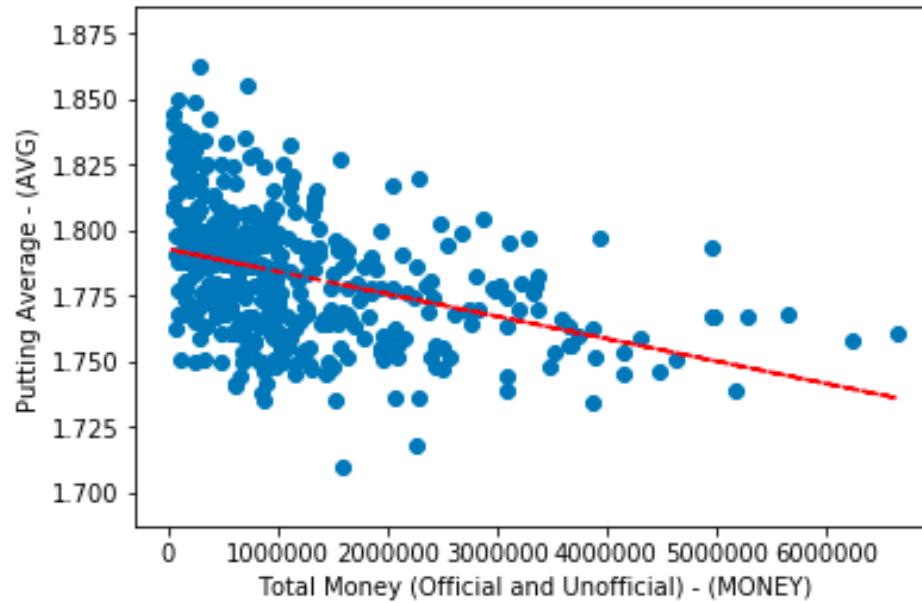


Appendix



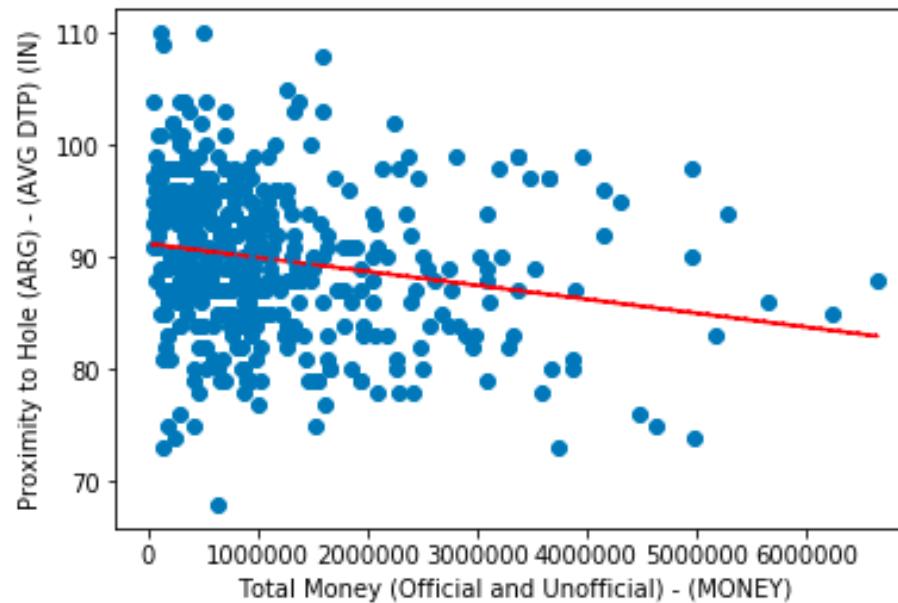
Appendix

Correlation of Money Earned versus Putting Average



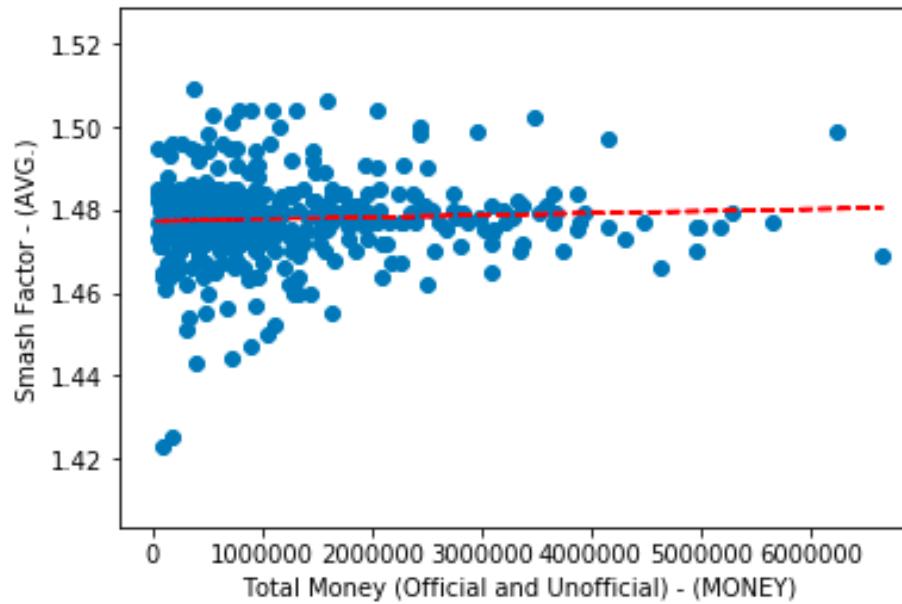
Appendix

Correlation of Money Earned versus Proximity to Hole (ARG)



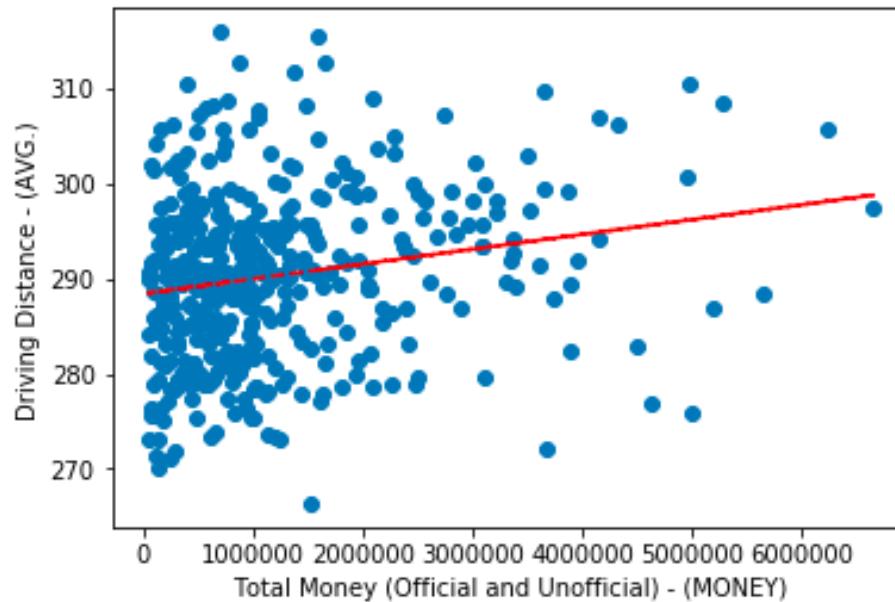
Appendix

Correlation of Money Earned versus Smash Factor



Appendix

Correlation of Money Earned versus Driving Distance





Credits

- *This presentation has been designed using resources from PoweredTemplate.com*