Title Slide

* Project Name
* Team members

Motivation and Summary 1

* Golfers want to hit the ball far
* Golf club manufacturers aggressively develop and price drivers, promising longer distance and more accurate tee shots
* Technology has played a big part in driver development, e.g.
  + Material science
  + Aerodynamics
  + Artificial Intelligence
* Has driver performance really improved over the years?
* If so, has driver technology left the rest of the game behind?
* We want to know if driver technology has left the short-game and putting relatively untouched
* We’ll do this by looking at the PGA Tour, from 2010-2018, and analyze driving performance vs. short-game and putting performance

Motivation and Summary 2

* THESIS:
  + Driving performance is not related to short game or putting performance
* 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 and putting performance have remained the same
  + Any increase in short-game and/or putting performance are by random chance
* We were able to answer these questions and we found that:
  + We failed to reject the null hypothesis
  + Driving performance increased, without a corresponding increase in short-game or putting performance

Data Cleanup & Exploration 1

* We used a large dataset from Kaggle with PGA historical statistical data for 2010-2018
* We used another dataset from Kaggle with PGA player scoring averages from 1980-2019, but only to determine years of playing experience
* For data cleaning, we:
  + Calculated years of experience and merged into the main dataframe
  + Removed any unneeded statistics
  + Transformed dataset from vertical to horizontal so our 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
* (Screenshot of head output)
* (Screenshot of column datatypes)
* (Screenshot of row counts and # players per season)

Data Analysis 1

* We focused on 5 PGA Tour statistics:
  + 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.
  + 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.
    - Measure of player ability as well as club fit
  + Short Game Rating
    - An aggregate score based on several 'Around the Green' statistics used to measure a player's performance 100 yards and in (not including putting). The statistics used to compute this are: Scrambling, Proximity to the Hole from sand, PTH from Rough, PTH from fringe, PTH from less than 100 yards and Greens in Regulation from less than 100 yards.
    - PTH = Proximity to Hole
  + 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. Only those shots determined by a laser will be included. (Note: ‘Around the green’ indicates the player is within 30 yards of the edge of the green).
  + Putting Average
    - The average number of putts per green in regulation. By using greens hit in regulation, we are able to eliminate the effects of chipping close and one-putting in the computation.

Data Analysis 2

* For each of our 5 PGA Tour statistics:
  + We created a table of summary stats, including tests for normality
    - (screenshot example)
  + Boxplots and Density plots to analyze distribution of data across the seasons
  + One-way ANOVA tests and Tukey's range test to compare all possible pairs of seasons' means
* We also looked for correlations and trends by using:
  + Scatter matrix
  + Correlation table

Data Analysis 3 - Driving Performance - Driving Distance

* Discuss boxplot
* Discuss ANOVA and Tukey
* Discuss Density Plot
* Conclusions:

Data Analysis 4 - Driving Performance - Smash Factor

* Discuss boxplot
* Discuss ANOVA and Tukey
* Discuss Density Plot
* Conclusions:

Data Analysis 5 - Short Game Performance - Rating

* Discuss boxplot
* Discuss ANOVA and Tukey
* Discuss Density Plot
* Conclusions:

Data Analysis 6 - Short Game Performance - Proximity to Hole (ARG)

* Discuss boxplot
* Discuss ANOVA and Tukey
* Discuss Density Plot
* Conclusions:

Data Analysis 7 - Putting Average

* Discuss boxplot
* Discuss ANOVA and Tukey
* Discuss Density Plot
* Conclusions:

Data Analysis 8 - Correlations

* Discuss scatter matrix
* Discuss correlation table

Data Analysis 9 - Extra: Player-Specific Performance

* Discuss Phil vs J.J. vs Dustin performance

Data Analysis 9 - Extra: Money Performance

* Discuss average money per season by top/bottom 10

Discussion 1

Discuss your findings. Did you find what you expected to find? If not, why not? What inferences or general conclusions can you draw from your analysis?

Post Mortem 1

\* Discuss any difficulties that arose, and how you dealt with them

\* 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?

Questions 1