**Project I Proposal – Golf Technology Impact on Drivers vs. Short-Game Clubs**

Group 4: Amit Patel

Roopa Patel

Samuel Parks

Sean Jokhai

Steven Lee

**Introduction**

Anyone who plays the game of golf wants to hit the ball far off the tee. Golf club manufacturers are always aggressively developing and marketing drivers that can hit the ball longer and straighter, while still within the limits set by governing bodies like the USGA (United States Golf Association) and R&A (Royal and Ancient Golf Club of St. Andrews). Our belief is that drivers are the most “improvable” golf club through technology, irons being second, and last being the clubs used inside of 100 yards. The clubs used for shots inside of 100 yards (let’s call these short game clubs) are typically wedges (gap wedges, sand wedges, lob wedges) and putters.

We want to know if the constant focus on driver technology has left the short-game relatively untouched. We’ll do this by looking at the stats of the best players in the world, the PGA Tour, across 2010-2018, and compare their performance off the tee with their short game performance.

**Hypothesis**

Driving distance and accuracy has improved over the years, with an accompanying increase in short-game performance (<100 yards including putting)

**Null Hypothesis**

Driving distance and accuracy has improved over the years, while short-game performance (<100 yards including putting) has remained the same

**Data Summary**

* Our data source is “PGA Tour Golf 2010-2018 Data” from Kaggle.com
  + The data is described as “Every statistic recorded on the PGA Tour from the 2010 to 2018 season”
  + URL: <https://www.kaggle.com/bradklassen/pga-tour-20102018-data/version/4>
* There are two versions of the data file
  + One version has the data vertically, while the other has the data horizontally
  + We will most likely use the vertical version
  + The vertical version has been loaded into our notebook and described in this proposal
* There are 2,740,403 rows of data
* The data represents 9 years of PGA Tour Golf statistical data

A screenshot of a cell phone

Description automatically generated

* There are 5 columns:
  + Player Name
  + Season
  + Statistic
  + Variable
  + Value
* The “Statistic” column represents a category of statistics, with “Variable” being a specific statistic within this category
  + The Value is the numerical value of the specific statistic, or text value where appropriate (e.g. player’s country, tournament name, etc…)
* There are at least 1,200 players per season

A screenshot of a cell phone

Description automatically generated

* The number of “Statistic” values available per season vary, but there are at least 383 statistical categories available each season:

A screenshot of a cell phone

Description automatically generated

* The number of specific stat “Variable”s available per season vary, but there are at least 1,505 available each season:

A screenshot of a cell phone

Description automatically generated

* For driving performance, we will most likely be focused on:
  + Statistic: Driving Distance
    - Variable: Driving Distance - (ROUNDS)
    - Variable: Driving Distance - (AVG.)
  + Statistic: Driving Accuracy Percentage
    - Variable: Driving Accuracy Percentage - (ROUNDS)
    - Variable: Driving Accuracy Percentage - (%)
* For short-game performance, we will most likely be focused on:
  + Statistic: Approaches from inside 100 yards. This is for approaches from the fairway
    - Variable: Approaches from inside 100 yards - (ROUNDS)
    - Variable: Approaches from inside 100 yards - (AVG)
  + Statistic: Approaches from inside 100 yards (Rgh). This is for approaches from the rough
    - Variable: Approaches from inside 100 yards (Rgh) - (ROUNDS)
    - Variable: Approaches from inside 100 yards (Rgh) - (AVG)
  + Statistic: Putting Average
    - Variable: Putting Average - (ROUNDS)
    - Variable: Putting Average - (AVG)