

## Data Report - PGA Tour

Group Members: **Hari, Lala, Daniel, Daniel, and Skyler**

Dataset Link: [Kaggle PGA Tour Link](#)

Slides Link: [PGA Tour Project](#)

# *PGA Performance Impacts on Prize Money*

## Abstract/Introduction

- This study will explore the relationships between a player's performance metrics and prize money on the PGA Tour. Performance plays a key role in professional sports as the better you play and climb up the leaderboards, the higher seed you'll be obtaining the main trophy as well as the prize money. We believe that this overall topic of an individual's performance will impact their prize money, as the better you play in your sport, the better your chances of being at the top of the leaderboard. We estimated that prize money would increase over time and there would be at least one main predictor for money earned each year. Overall through our research, we confirmed that money earned has been gradually growing over time. We also found that the strongest individual predictor for money earned is from points. The possible next step would be to combine all of our findings to create a prediction for money from points earned, GIR, putting average, and rounds played

## Data & Methods

- *Data Dictionary*
  - **Rounds Played:** Rounds of golf played in the year
  - **Year:** year of data
  - **Fairway Percentage:** Percentage the drive off the tee box lands in the fairway
  - **Average Drive Distance:** Average distance when hitting a driver off the tee
  - **Greens In Regulation(GIR):** being on the green with a putt for birdie
  - **Average Putts:** Average number of putts per round
  - **Average Scrambling Percentage:** percent scoring Par or better after missing green in regulation
  - **Average Score:** Average score of all rounds

- **Points:** FedEx Cup points, each player is given a certain amount of points per round depending on where they place, higher points mean better placements overall through the year.

- **Data Cleaning**

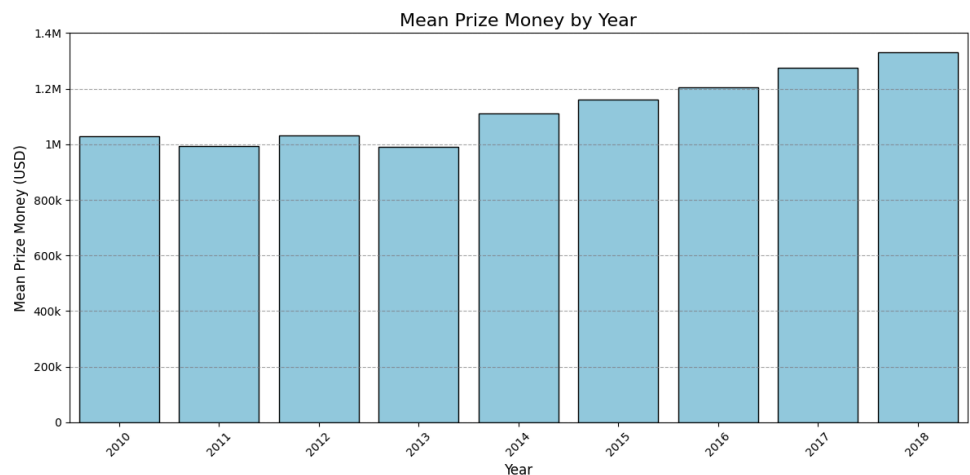
- We had to remove the “\$” sign in the prize money column as the computers believed it as a String type so we removed it through some code to redirect it into an integer value for our charts.

```
# Removes the dollar sign off.
pgaTour_df.dropna(subset=['Money'], inplace=True)
pgaTour_df['Money'] =
pgaTour_df['Money'].str.replace('[\$,]', '',
regex=True).astype(int)
```

## Results

- **Slide 1**

- Graph



- Question

- What is the average annual prize money earned by an individual on the PGA Tour?

- Hypothesis

- My hypothesis is that the mean annual prize money will increase gradually overtime due to inflation or simply the popularity of the sport as more companies will want to sponsor the event.

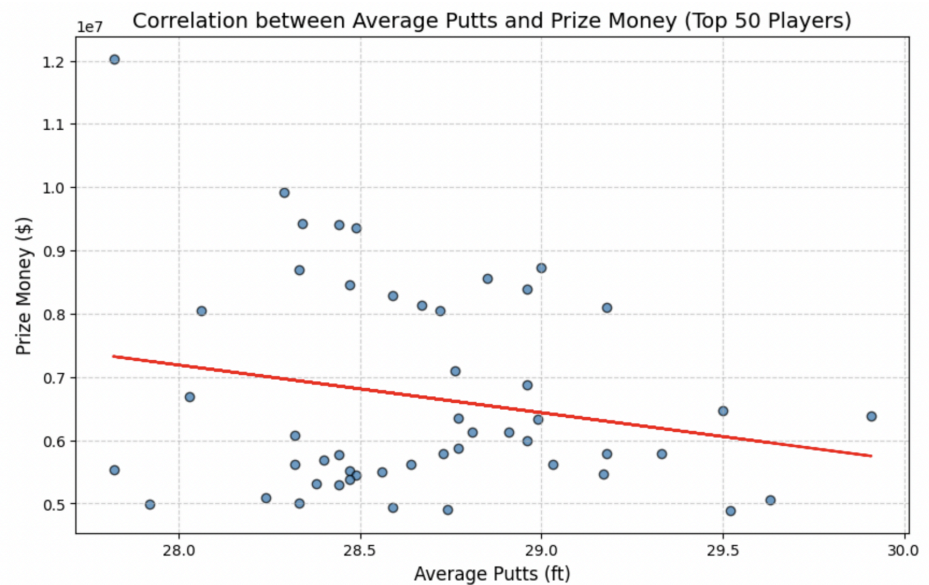
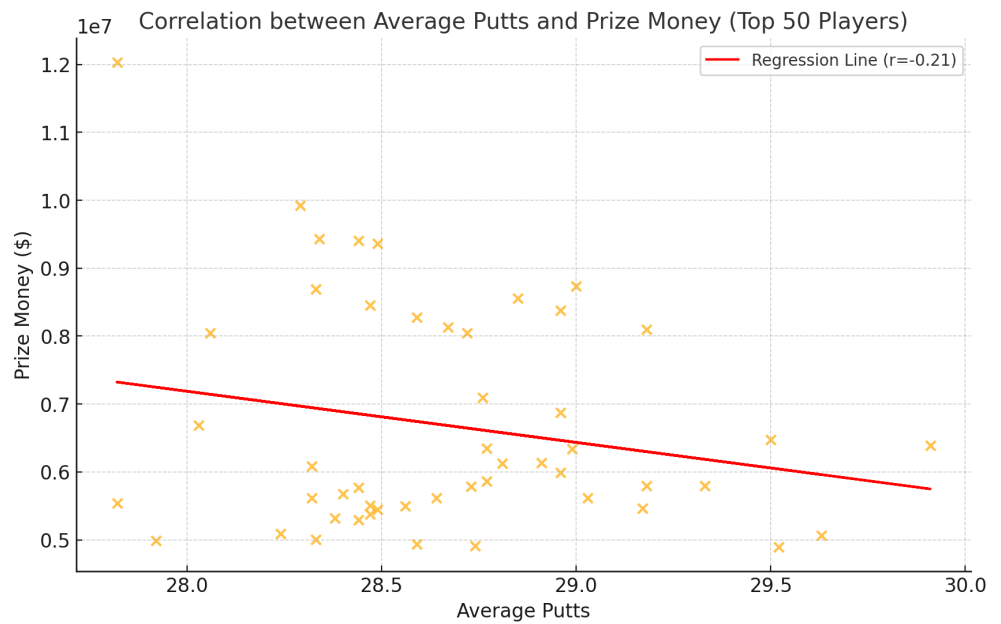
- Analysis/Reflections

- Prize money has remained stable from 2010-2013 and slowly began to increase 1by 00k per year, from 2014-2018.

- One variable that could factor in the growth subtle increase in prize money can be through their revenue streams.
  - For example, in 2014-2018 the popular golfer, Tiger Woods, helped boost back sponsorship and viewership which most likely led to an increase in the total prize pool.

- **Slide 2**

- Graph

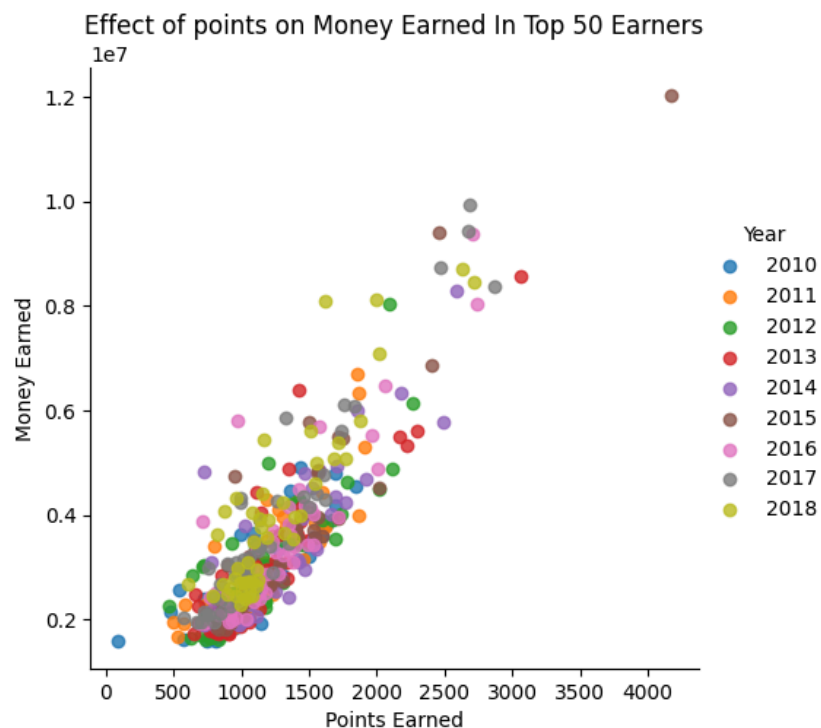


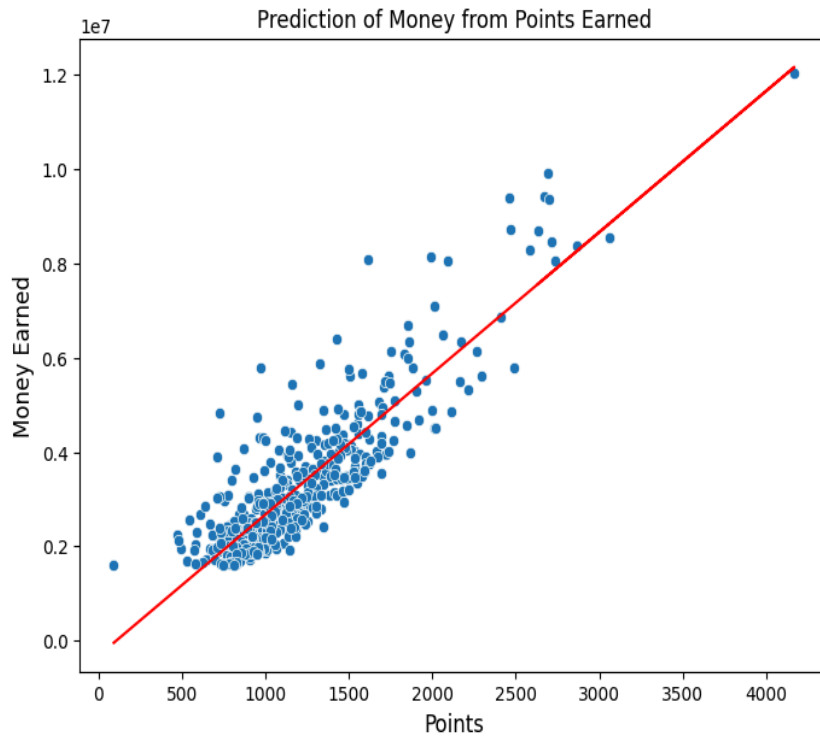
- Question: Is there a correlation between putting average and prize money among the Top 50 players ?
- Hypothesis: There will be a negative correlation between putting average and prize money among the top 50 players.
- Analysis/Reflections :

- The regression line shows a weak negative correlation with a coefficient  $r = -0.21$ .
- There is a slight trend showing that players with lower putting averages (better putting performance) tend to earn more prize money.
- The correlation is weak, indicating that putting average alone does not strongly predict prize money.
- Other factors that could help predict prize money strongly include: driving accuracy, overall performance consistency, tournament difficulty or # of times player has played the course.

### ● Slide 3

#### ○ Graphs



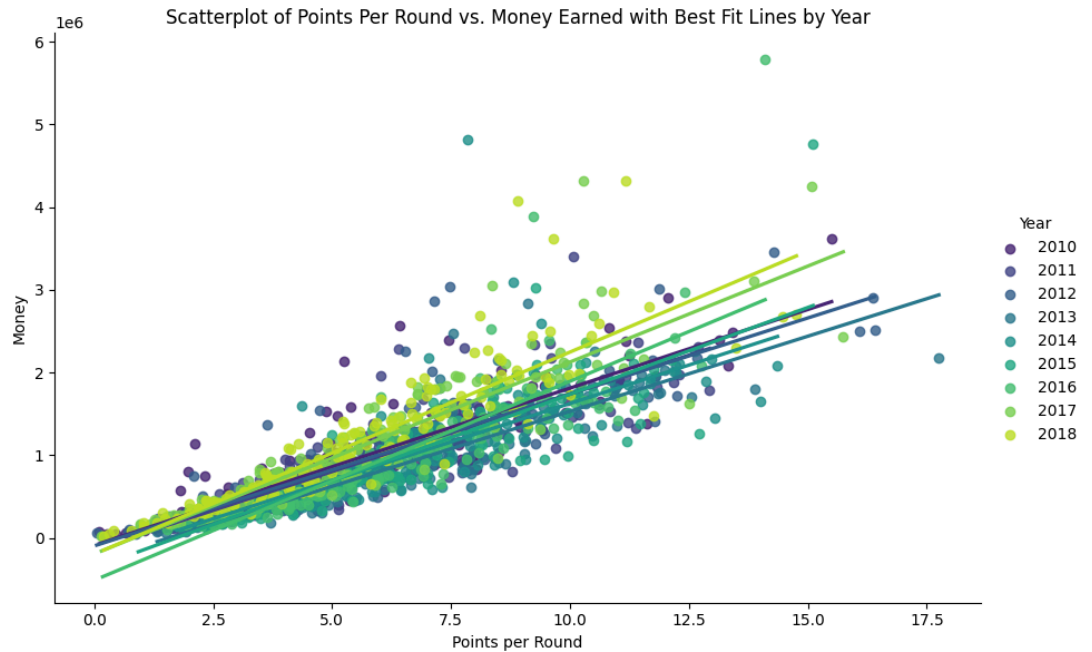


■

- Question
  - Predicting Money from Points
- Hypothesis
  - There will be a strong positive relationship between points and money earned.
- Analysis/Reflections
  - There is a moderately strong positive relationship between the two variables.
  - The variation in Money Earned can be explained by 76% of the variation in Points Earned
  - From this data, I was able to learn that while points earned were a good predictor, it wasn't a great one. To make a better prediction model we would need to plug in more predictors to gain a more accurate prediction.

● **Slide 4**

- Graphs



○

○ Question

- How strong is the correlation between points per round and prize money and how does this change over years

○ Hypothesis

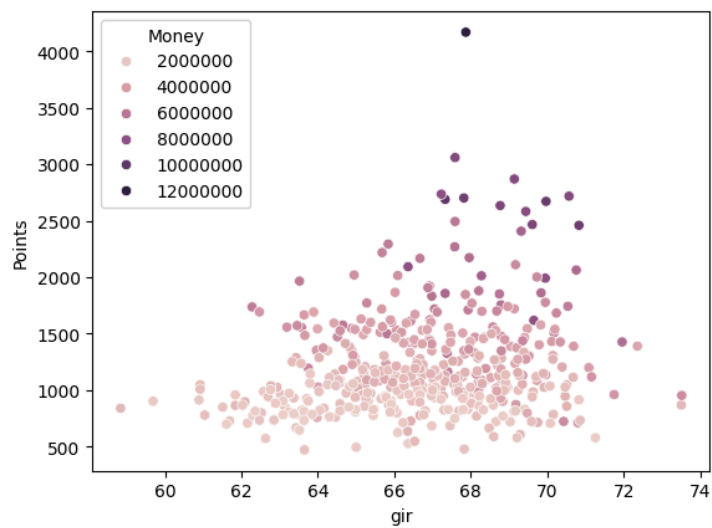
- The correlation will be very strong and positive and the correlation will get stronger year by year

○ Analysis/Reflections

- The overall  $r$  value of 0.84 showed that the correlation was very strong and positive
- By using the graph information with the lines of best fit we see that the correlation gets more positive and stronger through the years

## ● Slide 5

○ Graph



■

- Question:
  - Does GIR (greens in regulation) correlate to points/money earned
- Hypothesis
  - GIR will correlate very well to points/money earned
- Analysis/Reflections
  - There was little to no correlation between points/money earned and GIR. This surprised me at first but the more I think about it the more it makes sense. Opposed to the average golfer, who's chipping skills aren't refined enough to get the ball close when a green in regulation is not achieved, the professional golfers short game is so refined a long putt is often as close as a chip (this is, of course, another hypothesis, but one I've reached after reflecting on my data).

### **Group Colab Links:**

- <https://colab.research.google.com/drive/1ZjeRNNxnYyE5l6DUTwjlqDCJ3nV4E0YA?usp=sharing>
- [Google Colab - Hari](#)
- <https://colab.research.google.com/drive/1LfP0yQhh4SQAnpmvpYjV2X-ECztnJoEm>
- [PGA - Colab](#)
- [PGA Putting Data](#)