Ammon Roberts 10793687

Connor Paulson 10684469

**Tasks accomplished:**

With our project, We decided first after our proposal to talk more with each other about exactly what data we are using for this project, data that would be simple at first to actually accomplish what our goal for this project is, then to branch out with a little more complex data and a greater quantity of it.

We did some more digging and we decided not to do a simple random sample (SRS) of basketball games from an NBA season. We are instead going to grab a few major stats (2pt %, 3pt %, and margin of victory) from the first 5 games the Utah Jazz have played this 2019-2020 season, and then get the same stats from the first 5 games of the Jazz’s next 5 Opponents, and then based off those stats, we will compare each of their next opponents and predict what their record would be against those next 5 opponents. With the first 5 games, it will provide more consistent data for us to work with.

Along with making clear the kind of data we’re going to use, We then started to do a shallow dive into Python and we tried to investigate and figure out what libraries would be the best to use for what we are trying to accomplish. We figured the scipy library

* Decide what data to use
  + Decide to first 5 games instead of a random sample
  + This will provide a more consistent set of data to work with
* Decide how to pick what teams to use
  + Focus on the Utah Jazz’s recent game stats
  + See who they are matched up with in the future
  + Use the Jazz’s starting 5 game data compared with the first 5 starting games from the other 5 teams
* Compute some samples by hand or another method

**Tasks pending:**

* Compiling all data for analysis
* Python script

**Tasks Remaining:**

We have already been looking at examples of Python code to aid us in figuring out the confidence interval of a set of data.w

* Plan how the Python script will read the sample data, calculate, and predict the final results
  + Define parameters for how to determine the confidence interval (what counts as an outlier)
* Create Python script to analyze data and make prediction
  + Need to determine exact functions inside of which libraries to import
* Test and verify data against samples previously calculated
* Finalize any documentation and other tasks

**Project tracking:**

Source Code:

Sample source code for calculating confidence interval using the scipy Python library:

**from** scipy.stats **import** sem, t

**from** scipy **import** mean

confidence = 0.95

data = [1, 2, 3, 4, 5]

n = len(data)

m = mean(data)

std\_err = sem(data)

h = std\_err \* t.ppf((1 + confidence) / 2, n - 1)

start = m - h

**print** start

### OUTPUT

1.03675683852

end = m + h

**print** end

### OUTPUT

4.96324316148

Get 2 pt % 3pt% and margin of victory for Jazz first 5 games of this season (2019), and get the first 5 game stats for the Jazz’s next 5 opponents, and attempt to predict the Jazz’s W/L record for their next 5 games based off the confidence interval of the data.

Possible added complexity: get 15 games and predict the next 15 games, etc.