Progress Report 2

Ammon Roberts 10793687 Connor Paulson 10684469

**Report:**

With the last milestone, we focused on getting our hands on the data we would use for our project. We included with that all the research into statistics of the current 2019-2020 NBA season and our decision for how much data we were going to use. We decided to start with the first 5 games of the Utah Jazz’ season and compute the confidence interval of certain stats to be able to predict their record over their next 5 games and opponents.

For this milestone, we have worked more on our source code for this project. After the end of the last milestone, we had an extremely basic Python script, which could read in a CSV and print the data therein. We expanded on that script, which now is able to compute the data from CSVs and calculate a confidence interval based on the data retrieved from the CSV. We organized the data from the Utah Jazz’ games and their next 5 opponents' first 5 games into 6 separate CSV files in preparation to be used in our Python script. The Python script is still unfinished, as we still have added functionality that needs to be implemented and tested to produce the results we want.

**Tasks accomplished:**

As mentioned previously, we have made significant progress on the Python script we will run on our various data sets. The Python script starts, and imports the necessary libraries. At which point, it loads each dataset from the respective CSV files, one at a time, and makes the calculation based on the parsed data, then outputs the results.

The output of the script is generated as the script progresses. Some basic formatting was added to make the results a little more readable.

Comments were also added to the Python script to increase readability.

**Tasks Remaining:**

We are at a point where we still need to finish the Python script and enable it to be able to find the confidence interval of the team data and compare them against each other to be able to predict the Utah Jazz’ record over their next 5 games.

We have not attempted to input data from more than 5 games currently, but are planning on looking more into what will need to be done in order to expand the project to be able to handle such an increase.

**Project tracking:**

According to our timeline from our project proposal, we are still on a good track to finish our project. We have already found our data samples, planned how the Python script we have now created will read in the data sets we have gathered and produce a data structure (with parsed data) that we can run the script against. We are actively creating and testing out the Python script with the data that we have. We still have other things to add to the script and test, but, so far, we are on track with the building and testing with our project.

**Source Code:**

**import scipy as s**

**import numpy as n**

**import matplotlib as m**

**import csv**

**import warnings**

**import scipy.stats**

**# Ignore deprecation warnings**

**warnings.filterwarnings("ignore")**

**# data object containing lists for each major stat category**

**data = {'fg': [], 'threePt': [], 'margin': []}**

**# Declare some constants: confidence level, array size (number of games)**

**confidence = 0.95**

**n = 5**

**# Calculates the interval based on the array that was given**

**def calculateConfidence(input):**

**m = s.mean(input)**

**std\_err = s.stats.sem(input)**

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

**start = m - h**

**end = m + h**

**print(f'Start: {start}\nEnd: {end}\n\n')**

**# Grabs each of these fields from each row of the**

**# CSV and adds it to its corresponding list**

**def getData(r):**

**data['fg'].append(float(r['FG%']))**

**data['threePt'].append(float(r['3pt%']))**

**data['margin'].append(float(r['MofV']))**

**print('~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~\n')**

**print('UTA\n')**

**with open('data/UTA5.csv', newline='') as csvfile:**

**reader = csv.DictReader(csvfile, delimiter=',', quotechar='|')**

**for row in reader:**

**getData(row)**

**print(row)**

**print(f'Field Goal Confidence Interval:')**

**calculateConfidence(data["fg"])**

**print(f'Three Point Confidence Interval:')**

**calculateConfidence(data["threePt"])**

**print(f'Margin of Victory Confidence Interval:')**

**calculateConfidence(data["margin"])**

**data = {'fg': [], 'threePt': [], 'margin': []}**