Progress Report 1

Ammon Roberts 10793687 Connor Paulson 10684469

**Report:**

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 ( FG%, 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 is a library that will work for what we are trying to accomplish.

We also have continued to set up what we need for the project by setting up a repository on Github to manage the code and scripts we write between the two of us.

**Tasks accomplished:**

We were able to gather the data needed in order to get started with beginning the computational portion of the project. All of this data is to be stored on the GitHub repository that was created for the purpose of keeping track of this project.

So far, only the framework for the Python script has been created. As progress is made towards the second milestone, we will continue to fill it out and prepare to use it to make a prediction as we planned.

**Tasks pending:**

Currently, there are still a small number of tasks that are being worked on. Among these are the compiling of all of the needed data for the predictions (we have only started with the first 5 games of the season and intend to expand further, to more games). As stated before, the Python script is still in progress and will be added to as time goes on.

**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. We need to finalize some decisions on what data we will count as an outlier in our confidence interval, along with a small number of other things relating to the data. We also need to further discuss the details of the method which we will use in order for the script to import and calculate the data to make a prediction.

Documentation has been started, but only so far as the README.txt file, explaining some quick basics of running the script, so this will need to be greatly expanded on in the future.

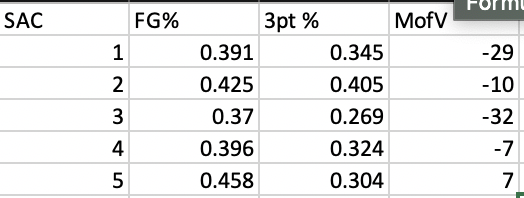
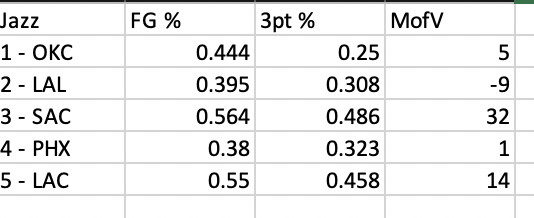
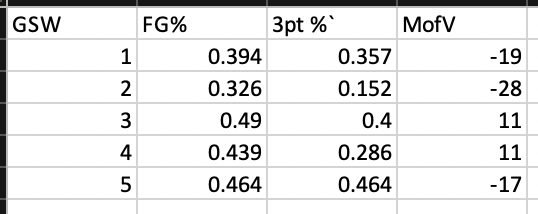
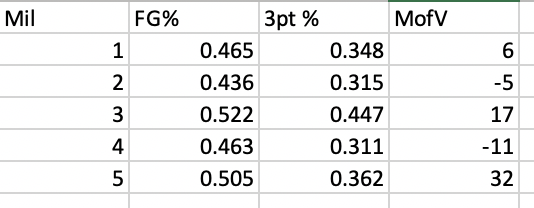
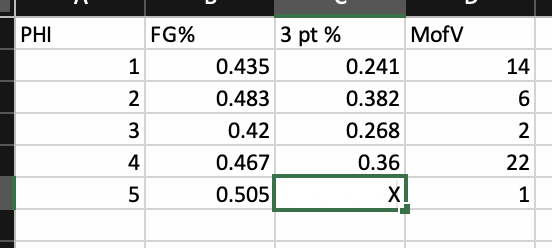
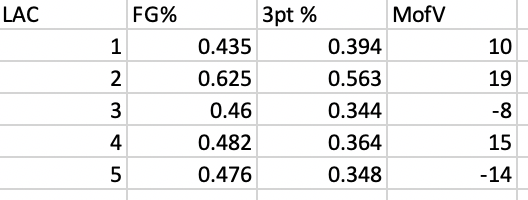
Once we have all of the sample data compiled and we have the Python script in a working state, we will be able to test the data and ensure its accuracy.

During this process, we will decide if we wish to expand the project further, using more data from more teams. Perhaps once the Python script is working as planned, we will include data that we haven’t run through the script, so that a prediction can be made based off of data that we have compiled, but not proven.

**Project tracking:**

So far, the project is nearly on track. We planned to have the data accumulated, with some testing done, but due to very slight changes in the scope of the project, some parts of this milestone are slightly delayed. Despite this, once we are able to get caught up in these areas, we will be back on track and anticipate that we will complete the project on time.

**Sample data:**

First 5 game stats for mian team(Jazz) and 5 of their opponents (6 total teams data used) 

**Source Code:**

The following is a sample of source code for calculating confidence interval using the scipy

**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*

This source code will provide a useful starting point for our Python script.