memo

| Syracuse University |
| --- |

| To: | Dr. Debbie Landowski |
| --- | --- |
| From: | Nick Videtti |
| Date: | August 12th, 2022 |
| Re: | IST 652 Final Project Proposal |
| Topic:  Description  Data/Tasks:  Metrics of Interest: | Does Tweeting Make Kevin Durant Better at Basketball?  : Kevin Durant is a professional basketball player that is recognized as one of the most talented scorers in NBA history. Professional athletes are known by many people and inevitably some of those people are going to critique those athletes. The approach that the vast majority of athletes take vary from ignoring the critics, limiting reasons for people to critique them, or embracing the criticism so they can learn to live with it. Kevin Durant has taken a different approach and stands up for himself on social media, most famously on Twitter. While he has had a few characteristics and made some decisions that fans have critiqued, defending himself on Twitter has ironically enough become one of the top criticisms of Durant, but is it justified?    Data will be pulled from Basketball Reference, the NBA division of the prestigious Sports Reference online database. Here, Durant’s game logs can be found for all 14 seasons he has played thus far. Sports Reference allows exporting of data in CSV format, so 14 CSV files will be created, one for each season’s game logs. These will be loaded into Python and ultimately will be combined into a single pandas DataFrame. Durant’s tweets will then be pulled and aggregated to find the number of tweets by day. These data will be joined to the game logs DataFrame on the day of the game/tweets. An important step here is to make sure this is an inner join so that the results are only for days where Durant both played a game and sent a tweet. After that will be data cleanup (see Metrics of Interest section) and visualization, using PyPlot, of Durant’s game statistics compared to the number of tweets he sent on the same day.  For the Twitter data, many metrics can be pulled and/or calculated, but the only one of interest will be Tweets per day. The game logs data will have many different metrics provided by Basketball Reference as seen below.    The obvious metrics would be points, rebounds, assists, and shooting percentage, but there are a small enough amount of metrics here to create a separate plot for each. For the sake of feasibility, this project will focus on whether the game was won or lost, overall field goal percentage, free throw percentage, 3 point percentage, points, and turnovers.  This will leave the following columns in the data.   * Date * Won/Lost * FG% * FT% * 3P% * Points * Turnovers * Tweets |
|  |  |