**Assignment 8**

**Project Proposal**

**Topic Enhancement**

Our team has decided to add a column which will tell how many times an article is cited by others currently and also; we plan to scrape h-index and i10-index from google scholar to check author credibility and popularity.

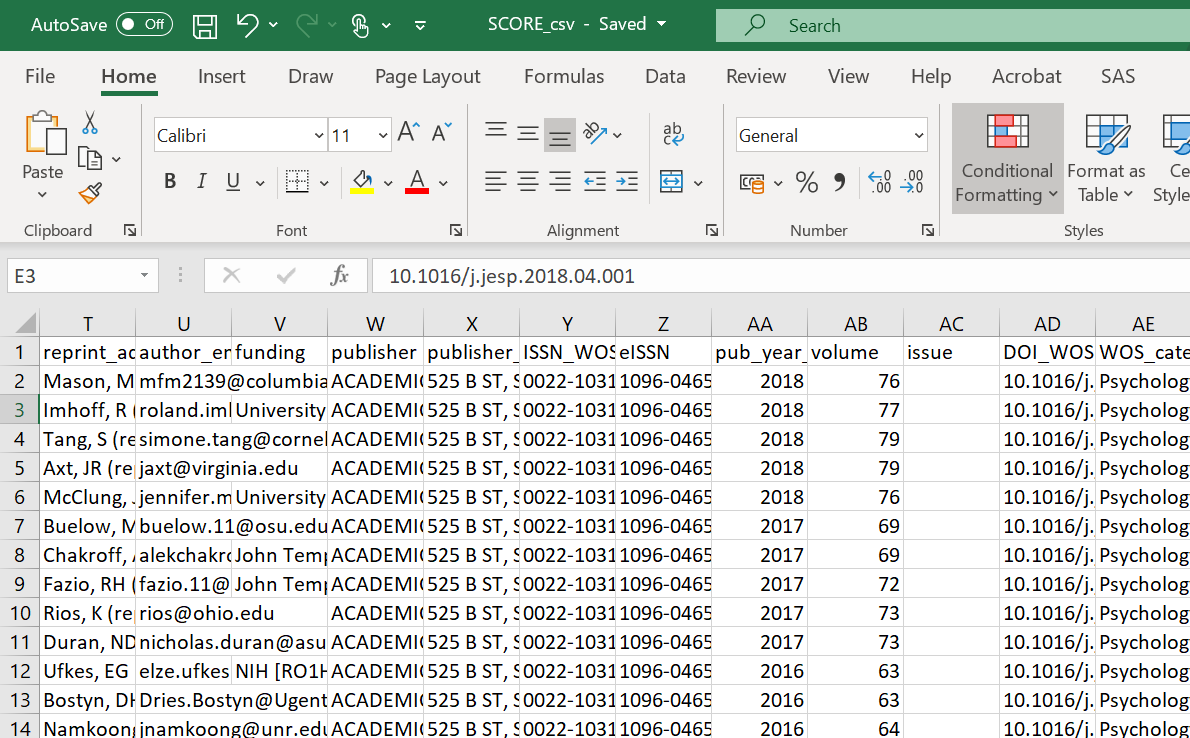
**Plan**

From our point of view, we decided to perform crawling algorithm on google scholar to scrape required citation, h-index, i10- index measures of particular author. Our data will be obtained and stored in numerical format to easily identify and enhance dataset.

**Dataset:**

<https://osf.io/um7kc/?view_only=1c79efb782e54b05acd3b1aa2dd375fd>

**Score\_csv.csv**



**Project phases**

Week 1

Exploring ideas: In total we gathered 40 ideas.

Week 2

Research and exploration for final idea: among which we finalized the idea to find h-index, i10-index and citations for authors on google scholar focusing on column 5 of the dataset.

h-index: The h-index is defined as the maximum value of h such that the given author/journal has published h papers that have each been cited at least h times. The index is designed to improve upon simpler measures such as the total number of citations or publications.

i10-index: The i10-index is the newest in the line of journal metrics and was introduced by Google Scholar in 2011. It is a simple and straightforward indexing measure found by tallying a journal’s total number of published papers with at least 10 citations

Citations: A citation is a reference to a published or unpublished source that you consulted and obtained information from while writing your research paper.

Week 3

Implementation: Since it’s a very big dataset we will distribute in 4 parts among all four team members and will implement the enhancement by using the crawling algorithm to compute the h-index, i10-index and citations. After finding these three columns we will add the column which is most useful in finding the most popular or authentic authors.

Week 4

Testing: We will test and re-test the accuracy of our findings.

Week 5: Presentation

In the end phase of the project, we will present the new dataset with h-index, I-index and citation score and enrich the dataset with it. This will help to identify the credibility of the author and how he is ranked among other authors and how much of his work is referenced by other people.