Use the python client that was give to you to generate a directory of 10K files with a max link count of 250. Copy the created files into a directory under a storage bucket on cloud google storage. Write a program in Python that opens the bucket, reads the list of files, reads each file and computes the following properties:

- Average, Median, Max, Min and Quintiles of incoming and outgoing links across all the files.
- Construct a graph of the pages, and compute their pagerank (PR) and output the top 5 pages by their pagerank score.
- Code the original iterative pagerank algorithm where PR(A) = 0.15 + 0.85 (PR(T1)/C(T1) + … +PR(Tn)/C(Tn)) where PR(X) is the pagerank of a page X, T1..Tn are all the pages pointing to page X, and C(X) is the number of outgoing links that page X has. Iterate until the sum of pageranks across all pages does not change by more than .5% across iterations.

Make sure your storage bucket is in a datacenter close to you and you can run your program either on your laptop directly or on your google cloud shell instance. Make sure your bucket is world readable (give the right ACL permissions) so your program can be tested against your bucket by the TFs. In your report include your total spend on cloud resources for the development and execution of this program.

What to turn in: A pdf file describing what you did including all relevant instructions on how to run and verify your code (i.e. your project name, your bucket name, the way to run your program, the meaning of any parameters it accepts etc). The program itself should be provided as a link to github containing your code. Make sure your github repository is world accessible and the link you provide works with git clone.