Progress Report for Siddhant (2019111038)

- 5th October to 14th October
    - I was a part of team-Q
    - Spent ~10 hours on learning and reading about quantum computing
- 15th October - 24th October (First interaction)
    - Created fucntion to scrape a Wikipedia page for all it's links
    - Created a mock CLI version of the game with Aman
    - Read about various database technologies
    - Implemented rudimentary caching to prevent looking up for the same links more than once
    - Hours dedicated: 15
- 25th October to 1st November
    - Looking up various ways to get Wikipedia stored offline
    - Decided on using the official WikiDump of all english articles (~75GB uncompressed)
    - Learned how to work with large XML files, since they can't be loaded into memory at once
    - Tried multiple python libraries to work with the wikidump.xml file
    - Hours dedicated: 15
- 2nd November to 14th November
    - None of the pre-existing solutions were good enough for our use
    - Editted an open source library to support:
        - Multiprocessing so we could create upto 7 threads to process 5000 articles each, at a time.
        - Extracting all the links from an article correctly, the original implementation had several bugs.
    - Learned docker and hasting containerized servers
    - Saved all the scraping data to a MySQL DB running on docker
    - Stopped at 200k articles for the time being.
    - Hours dedicated: 25
- 15th November to 22nd November (Final Presentation)
    - The DB got unexpectedly large so we were unable to upload to a remote server. Due to this I setup my laptop as a makeshift server for other teammates to use, which wasn't shutdown (willingly) for the entire remaining duration of the project.
    - Created the "Peak" functionality (Backend and a part of frontend) for the game. This lets the user read any wikipedia article while playing the game
    - Added a filter to only show links for articles present in our Database (Which is now sitting at 600K articles)

- Stored all the paths a user takes in a database which was used by Adwait to show as a table on the frontend.
- Hours dedicated: 25

MISCELLANEOUS

- Endless discord calls to discuss changes and next steps
- Laptop crash reports:
  - Creating too many threads doing computationally intensive tasks is a bad idea.
  - Docker will use all the RAM available and WILL crash the laptop.
  - Learned how to fix overheating: