**Our analysis ideas:-**

1. Eg. matplotlib community vs ggplot2 community (subset bigger questions dataset based on tags).

* Idea lead - Agrim
* Sub tasks:-  
  → Choose a topic (maybe just go with matplotlib vs ggplot2)  
    
  → Query dataset by relevant tags  
  → EDA on this newly made dataset

1. Weekdays vs weekends, is there a difference in the quality of questions being asked?

* Idea lead - Darren
* Sub tasks:

→ Figure out start and end dates of data we want to analyze

→ Figure out whether to sum up questions asked per day, have running average per day, or something else

→ Define what “good” questions are by length (?) and other qualities

→ Find number of good questions per 1000 (?) questions asked to standardize

→ Create barplot where x-axis are days of week and y-axis is good q’s/1000 q’s

→ Consider running t-tests comparing number of good q’s on weekdays vs good q’s on weekends to see if they are statistically significant

1. Comparing popularity of syntax based questions over different months.

* Idea lead - Agrim
* Sub tasks:-  
  → Filter dataset by syntax tag (firstly check if even possible)  
  → Plot monthly count histogram  
  → Write down subjective analysis of histogram

1. Percentage of questions being answered/percentage of high-quality questions over the years

* Idea lead - Yuan
* Sub tasks:
  + Percentage of questions being answered

→ Figure out a topic (compare between tags or ?)

→ Plot a histogram where x-axis are the tag and y-axis are the percentage

→ Make conclusion about

* + Percentage of high-quality questions

→ Criteria for high-quality questions, possibly:

* Number of views
* Number of answers (and votes?)
* Number of upvotes - downvotes
* How long it has been active (compare the last activity date and the creation date)