**MEETING 1 NOTES**

**General Direction for the blog:-**

1. Our readership should be people who ask questions, we could help them figure out how to ask better questions.

**Datasets:-**

1. <https://www.kaggle.com/stackoverflow/stackoverflow?select=stackoverflow_posts>
2. <https://www.kaggle.com/stackoverflow/stacklite?select=questions.csv> (merge question and question\_id based on id)
3. Try to build our own version of - <https://www.kaggle.com/aishu200023/stackindex>
4. Dataset with question and comment text - <https://www.kaggle.com/stackoverflow/stackoverflow?select=stackoverflow_posts>
5. <https://data.stackexchange.com/stackoverflow/queries>

**Ideas for Analysis:-**

1. Eg. matplotlib community vs ggplot2 community (subset bigger questions dataset based on tags).
2. Weekdays vs weekends, is there a difference in the quality of questions being asked?
3. Comparing popularity of syntax based questions over different months.
4. Percentage of questions being answered/percentage of high-quality questions over the years

**Ideas on ML stuff:-**

1. NLP on comments to help us figure out the quality of questions asked.
2. Predicting whether the question was answered based (we could use 10% of the entire dataset to train our model).