**4.** **How many annotated examples do you think you will be able to collect? Provide a justification for this number (e.g. note how many instances of unlabeled data are in your source repository and guess how many you will have after cleaning and filtering. Then, if your team members will do the annotation, calculate how many items they will need to annotate each week in order to hit your target).**

The dataset contains around 10877  samples out of which 5027 are not repeated. Additionally, out of these samples, 2130 are categorized as *'Relevant'* by the original dataset. As per our definitions of *'Actionable'* and *'Non Actionable'*, only *'Relevant'* data can be *'Actionable'*. On further inspection, we found that the 2130 samples are largely skewed towards *'Non Actionable'* data.
Thus, we have decided to augment the dataset with tweets pulled from the Twitter API so we have a total of 1000 samples to annotate with a decent split of labels. Divided equally, annotation would need 1 week at 250 tweets per group member.'

**5.** **What are the names and usc email addresses of your team members?**

| Name | Email |
|---|---|
| Shreyas Kolpe | kolpe@usc.edu |
| Shravya Gorur Sheshadri | gorurshe@usc.edu |
| Sanjana Srinivasa | srin317@usc.edu |
| Shyam Venkatesh | shyamgav@usc.edu |