Completed Tasks:

---

1. I have identified the issue specific twitter handles to be - Pampers, Tide and Gillette (@pampers, @tide and @Gillette) as specific brands to work with as they have a lot of customer interaction typically about their complaints regarding the product that they are using.
2. I have explored different twitter scraping tools such as taspinar/twitterscraper, bisguzar/twitter-scraper and TWINT. I found that TWINT suited my need the best as it was accessible through the command line interface and provided all the necessary flexibility.
3. I am working on cleaning the scraped tweets so that I can eliminate spam and inappropriate tweets

Pending Tasks:

1. Once the data is cleaned, I need to split the data into train and test components.
2. Evaluate the best deep learning vector space embedding UKPLab/sentence-transformers
3. I need to index the cleaned data into annoy (fast indexer for retrieving the nearest neighbour in vector space) using the vector space embedding chosen in step 2
4. Report the train and test accuracy of the system.

Challenges:

1. Twitter response api changes affect the tool and cause it to break. I need to figure out a mechanism to scrape the data in bulk and store it in the system.
2. Choosing the correct vector space embedding optimization might require time than initially estimated.