Data Science Assignment

You are given the dataset *'github_comments.tsv'* that carries 4000 comments that were published on pull requests on Github by developer teams.

Here is an explanation of the table columns:

- ***Comment***: the comment made by a developer on the pull request.
- ***Comment_date****:* date at which the comment was published
- ***Is_merged***: shows whether the pull request on which the comment was made has been accepted (therefore merged) or rejected.
- ***Merged_at***: date at which the pull request was merged (if accepted).
- ***Request_changes****:* each comment is labelled either 1 or 0: if it's labelled as 1 if the comment is a request for change in the code. If not, it's labelled as 0.

The goal of this assignment is to dig deeper into the nature of blockers and analyze the requests for change. If possible, try to answer the following questions:

- What are the most common problems that appear in these comments?
- Can we cluster the problems by topic/problem type?
- How long is the resolution time after a change was requested?

These questions are meant to be open ended. The goal is to see how you approach a dataset that contains unstructured data combined with structured data.