**Project : Automatic Ticket Assignment**

**Milestone 1: Pre-Processing, Data Visualisation and EDA**

**Problem Statement**: Incident management system in an organisation. Incidents are generally interruptions in the normal process which must be reported. These are reported by generating tickets. The organisation has different groups to solve different types of issues. Therefore, based upon the ticket content we need to categorise which group should be assigned the ticket.

**Data**: Consists of an excel file it has four fields, short description, Description ,Caller name, Assignment group.

**EDA:**

EDA, problem statement and conclusion is present in the following files attached

**INTERIM\_REPORT\_PROJECT\_NLP\_MILESTONE\_1.ipynb**

**INTERIM\_REPORT\_PROJECT\_NLP\_MILESTONE\_1.html**

**Data Preprocessing Pipeline:**

Code for preprocessing data is present in the **Preprocessing.py** file attached.

It includes code for :

1. Stripping html tags
2. Removing accented characters
3. Remove email ids
4. Extend stopword list
5. Get name initials
6. Expand contractions
7. Lemmatize text
8. Simple stemming
9. Remove stopwords
10. Remove custom words
11. Remove special characters
12. Clean corpus
13. Removal of irrelevant punctuations, hyperlinks
14. Detected foreign language words