**Capstone Project**

**Automatic Ticket Assignment**

**REPORT**

Contents

[ **Summary:** 2](#_Toc74129094)

[ **Process Details - Incident Management** 2](#_Toc74129095)

[ **Project Objective:** 4](#_Toc74129096)

[ **EDA (Exploratory Data Analysis)** 4](#_Toc74129097)

[ **Data Pre-Processing steps** 8](#_Toc74129098)

# **Summary:**

One of the key activities of any IT function is to “Keep the lights on” to ensure there is no impact to the Business operations. IT leverages Incident Management process to achieve the above Objective. An incident is something that is unplanned interruption to an IT service or reduction in the quality of an IT service that affects the Users and the Business. The main goal of Incident Management process is to provide a quick fix / workarounds or solutions that resolves the interruption and restores the service to its full capacity to ensure no business impact. In most of the organizations, incidents are created by various Business and IT Users, End Users/ Vendors if they have access to ticketing systems, and from the integrated monitoring systems and tools. Assigning the incidents to the appropriate person or unit in the support team has critical importance to provide improved user satisfaction while ensuring better allocation of support resources.

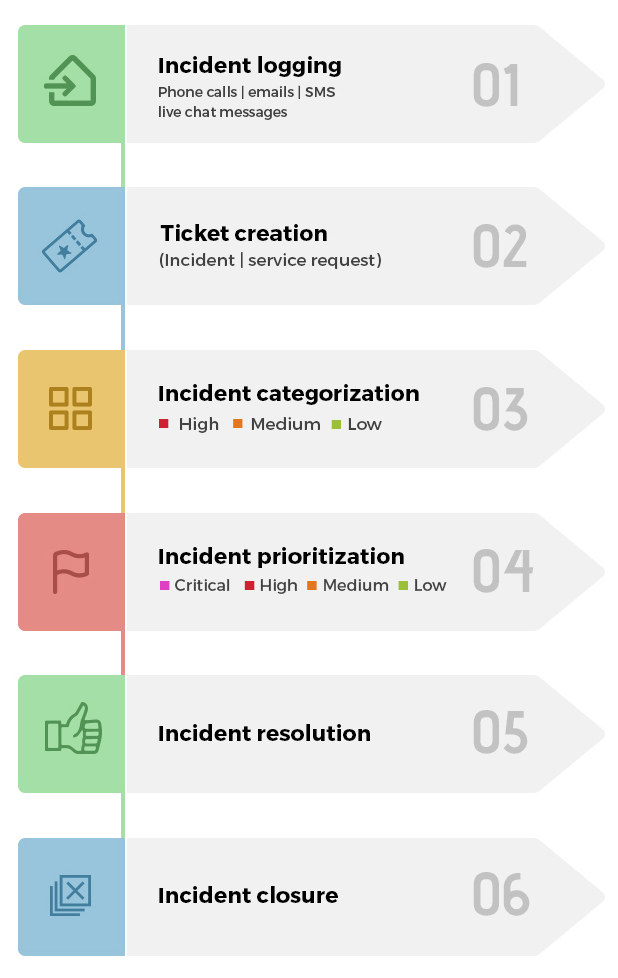
The assignment of incidents to appropriate IT groups is still a manual process in many of the IT organizations. Manual assignment of incidents is time consuming and requires human efforts. There may be mistakes due to human errors and resource consumption is carried out ineffectively because of the misaddressing. On the other hand, manual assignment increases the response and resolution times which result in user satisfaction deterioration / poor customer service.

# **Process Details - Incident Management**

An incident ticket is created by various groups of people within the organization based on data gathered through phone calls, emails, messages to resolve an issue as quickly as possible based on its severity. Whenever an incident is created, it reaches the Service desk team and then it gets assigned to the respective teams to work on the incident.

The Service Desk team (L1/L2) will perform basic analysis on the user's requirement, identify the issue based on given descriptions and assign it to the respective teams and further ticket can be solved and taken to its closure.

Following image describes the incident management process in general.



# **Project Objective:**

In the support process, incoming incidents are analyzed and assessed by organization’s support teams to fulfill the request. In many organizations, better allocation and effective usage of the valuable support resources will directly result in substantial cost savings.

However, successful closure is not the only thing which matters in customer satisfaction. SLA management of incident must be managed effectively. Generally, manually assigning the incident has some challenges as follows:

❖ More resource usage and expenses.

❖ Human errors - Incidents get assigned to incorrect groups

❖ Delay in assigning the incidents

❖ More resolution times

❖ If an incident takes more time in analysis, other productive tasks get affected for the Service Desk

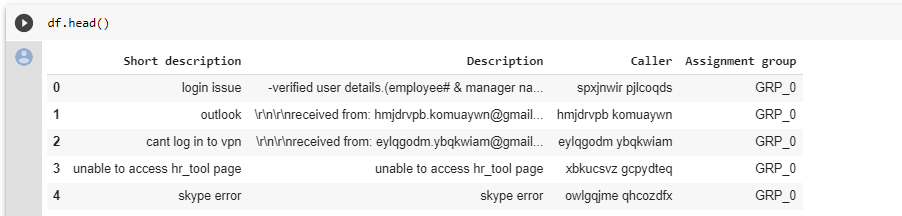
In this capstone project, the goal is to build a classifier that can classify the tickets by analyzing text and minimize the challenges mentioned above.

Our objective here is to build a classifier model to assign the incidents to right functional groups by analyzing the given description with maximum accuracy.

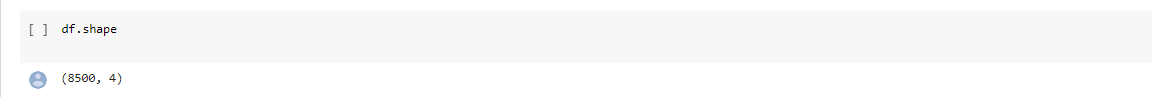
* Learned how to use different classification models.
* Used transfer learning to use pre-built models.
* Learned to set the optimizers, loss functions, epochs, learning rate, batch size, checkpointing, early stopping etc.

# **EDA (Exploratory Data Analysis)**

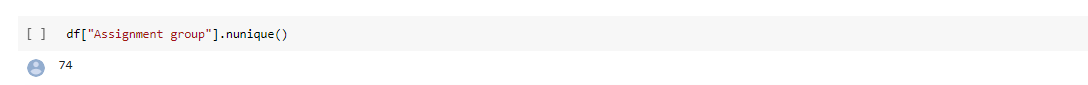
* There are Four columns – Short Description, Description, Caller and Assignment group.



* Total 8500 entries



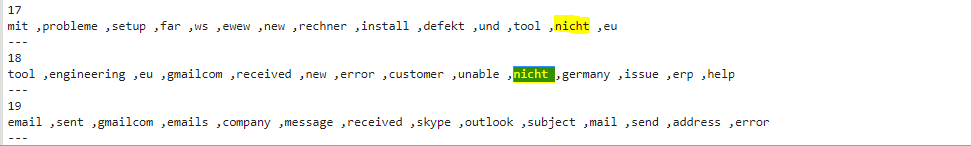
* Total Unique 74 Assignment groups found - Target classes



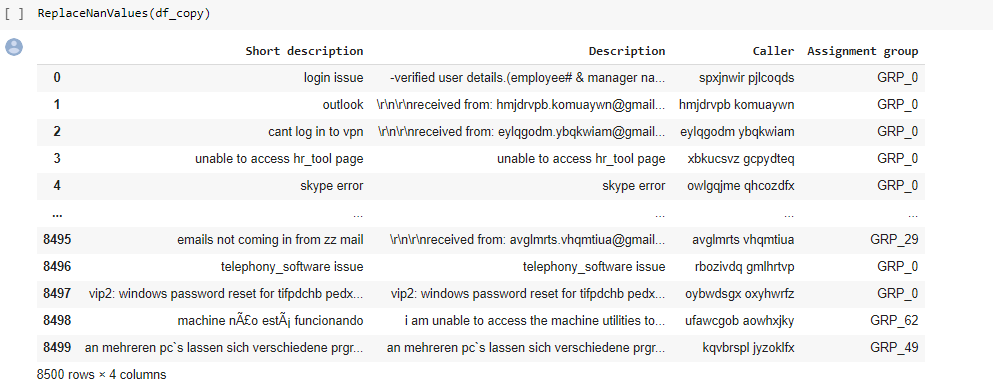
* Multiple languages found (eg: English, German, etc.)

Ref: Below shown is the German word from the dataset along with English words. ~82 % are English words.

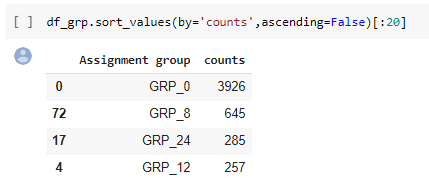
As per language detect library, there are total 29 languages in the Description column of the given dataset.

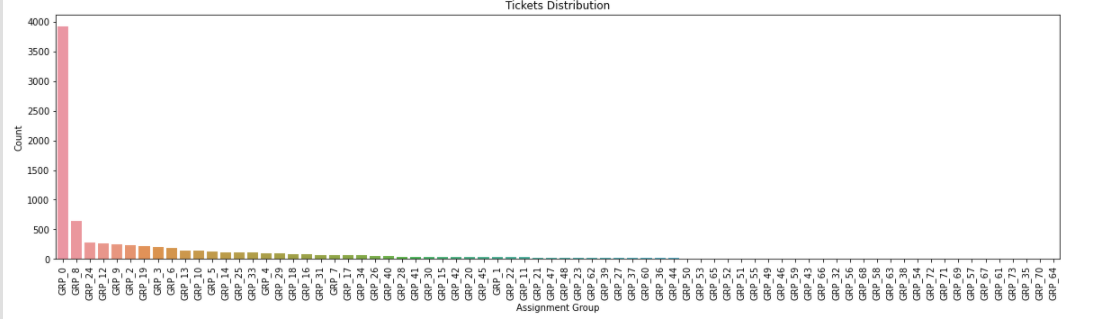


* Description column observation:
  + Email/chat format in description
  + Symbols & other characters in the description
  + Hyperlinks, URLS & few image data found in the description
  + Blanks found either in the short description or description field
  + Few descriptions same as the short description
  + Few words were combined
  + Duplicate data observed.



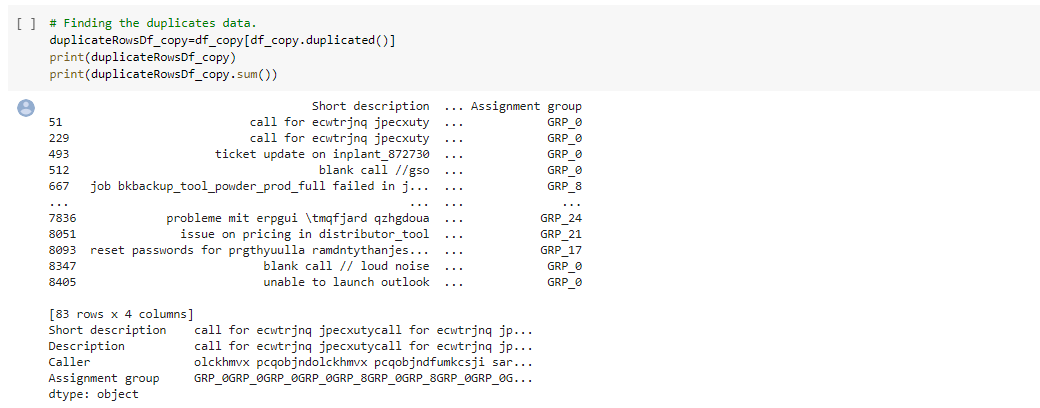
* Large no of tickets is assigned to GRP\_0 (amounting to 3926) which account for ~50% of the ticket assignment.



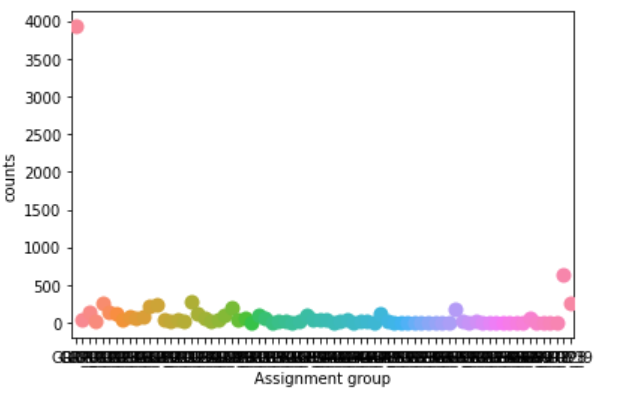


* As seen in the fig above, the Target class distribution is extremely skewed

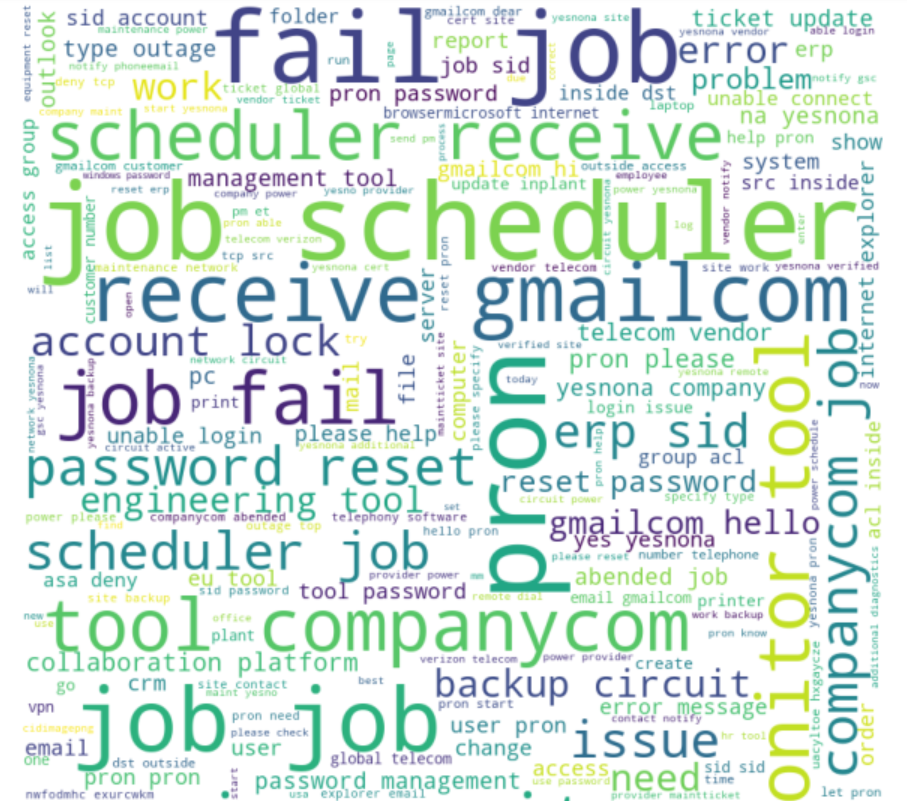
* Presence of duplicate data: There are total 83 duplicate entries in the given dataset. We have combined short description and description column which helped in getting duplicate entries based on combined data.



* Swarm plot – From the below plotted graph it can be ascertained that  Group-0 has highest count in terms of assigned tickets.



* Word Cloud – Word cloud Function is used to visualize most frequent words present in data set after processing data through various cleaning functions like, stop word removal, punctuation removal e.t.c



# **Data Pre-Processing steps**

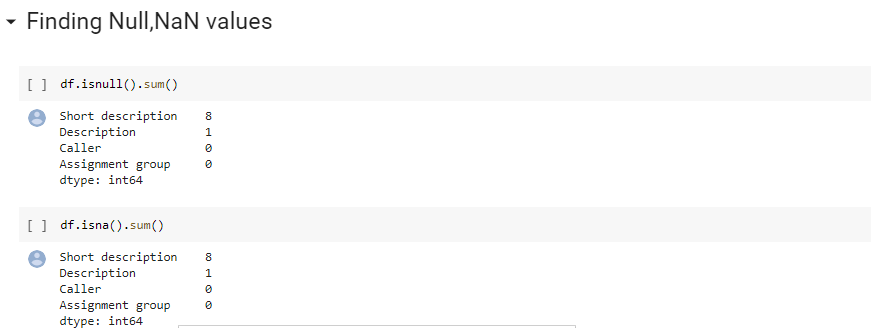
* Cleansing dataset as follows:
* Removed duplicates data using functions.



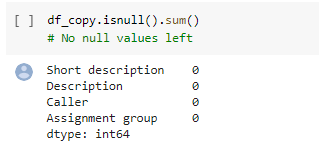
* Merging Short Desc and Description



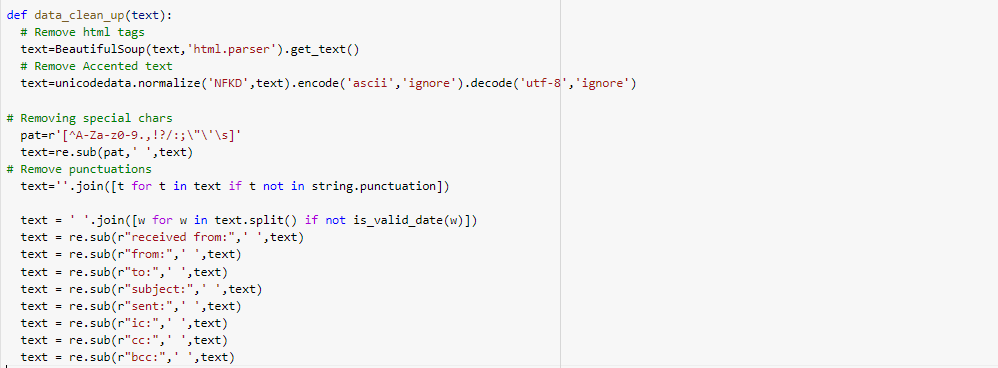
* Removing null values:
  + Before replacing null values:



* + After replacing null values:

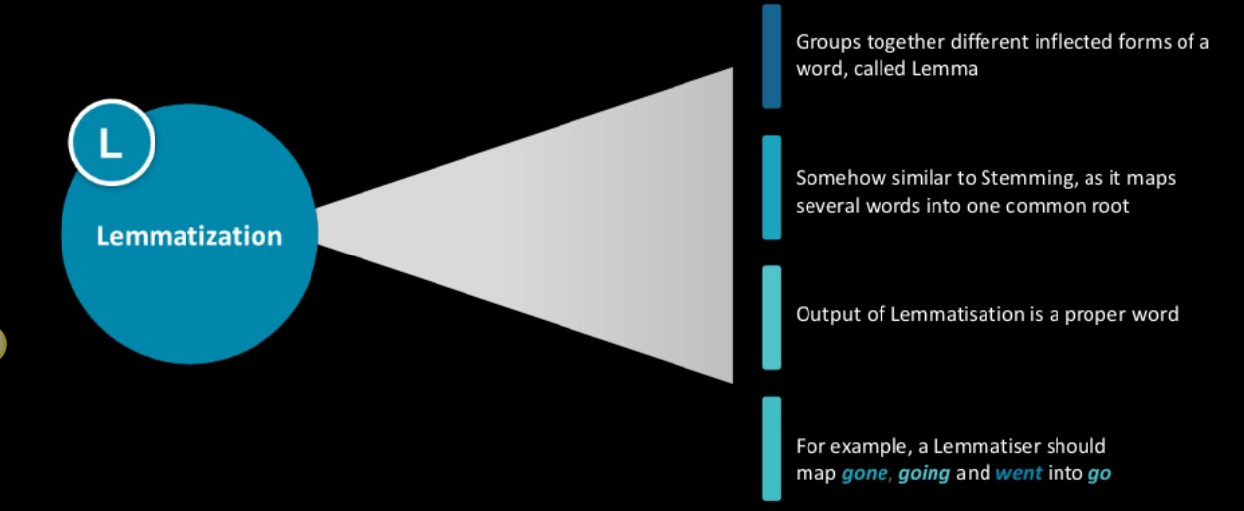


* Remove html tags
* Removing special chars
* Remove punctuations

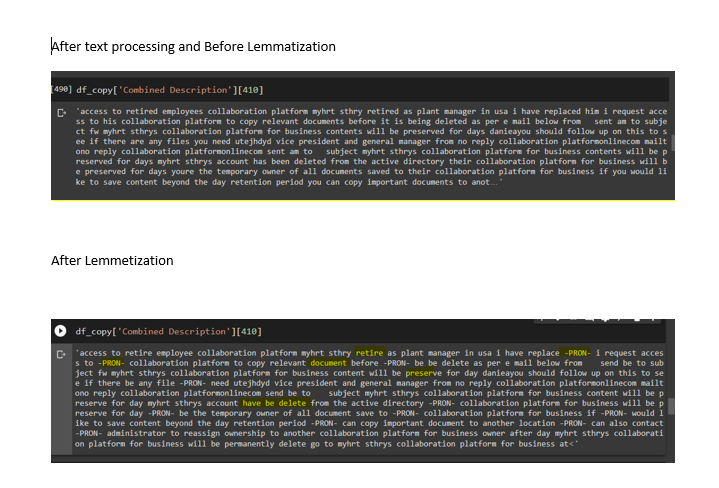


Based on above function, it removes html tags, punctuations and special characters.

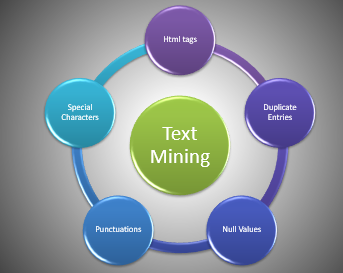
* + Lemmatization



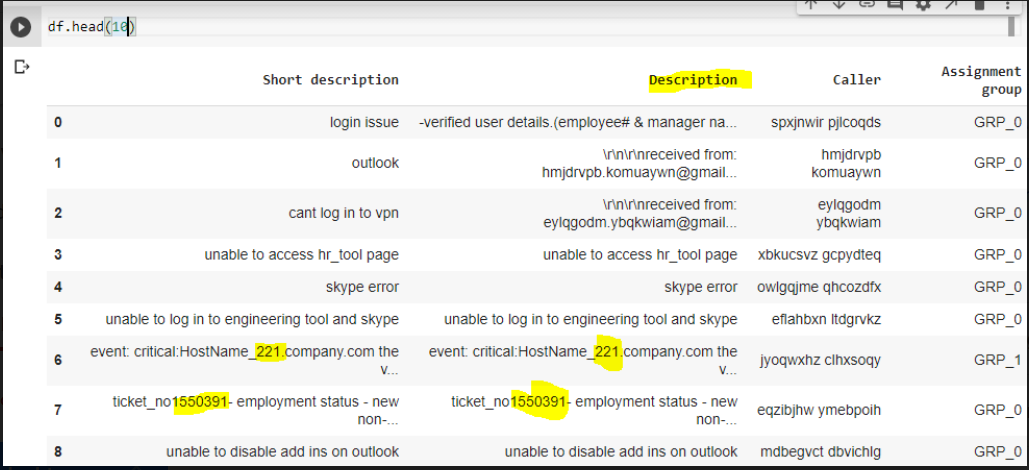
Lemmatization in linguistics is the process of grouping together the inflected forms of a word so they can be analyzed as a single item, identified by the word's lemma, or dictionary form. The examples are shown above. In lemmatization spelling is not impacted as compared to stemming.



* + Remove extra white spaces
  + Lower case
  + Remove Email Id
  + Remove numbers



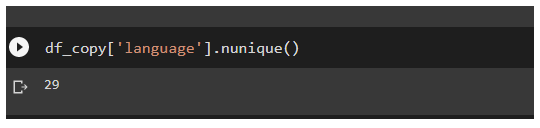
As seen below, we can see many numbers in the given data, which need to be removed.



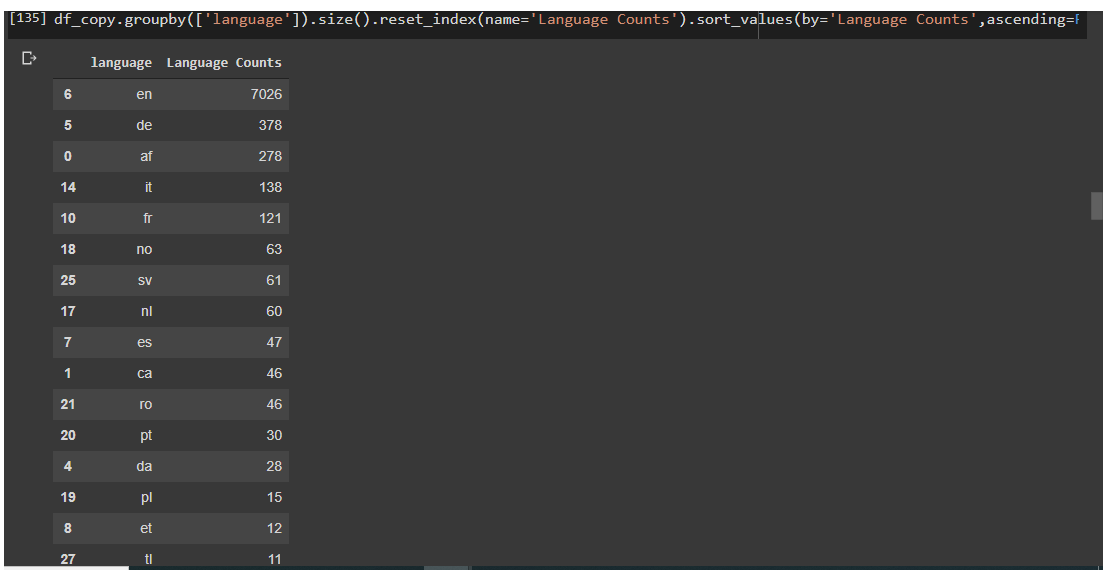
Using below , we are removing numbers, characters, extra white spaces, new line characters, hyperlinks, etc.



* Total 29 Unique languages in the dataset.



* Dataset contains the records from each language category as follows :



As per the above screenshot, the majority of the contents are ENGLISH. (82%)

* Word cloud – output as shown below:

Word Cloud is a data visualization technique used for representing text data in which the size of each word indicates its frequency or importance. Example shown below.



* fast Text – Semantic similarity between words.

