**Description:** To learn how to use Google collab and read the data using Python Pandas library

**Objective:** Performing three data analysis tasks for the given data using pprint and cumulative frequency syntaxes.

**Design and code screen shots:**

**1.**

Text

Description automatically generated

In the first step we have imported Pandas, Nltk libraries, we don’t have to install these libraries separately as we can take directly for servers.

**2.**

Graphical user interface, text, application

Description automatically generated

In the second step, we uploaded the data file in the G drive then we have to mount G drive to Google collab.

**3.**

Graphical user interface, text, application

Description automatically generated

Then the data file was read from the G drive with the path "/content/drive/My Drive/Assignments/input\_file.txt”

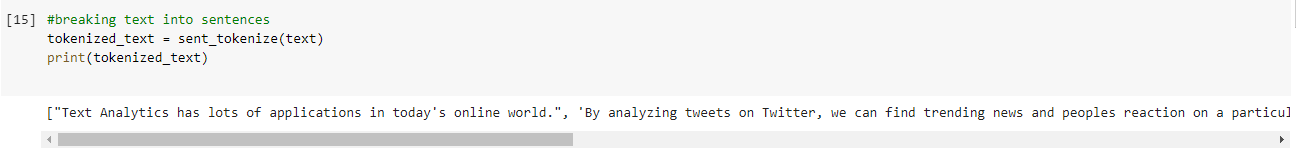
**4.**

Text, application

Description automatically generated with medium confidence

In the above syntax, the unwanted spaces are removed in the output.

**5.**



In this step, using the syntax we did tokenization of sentences where we converted text into sentences which became list.

**6.**

Graphical user interface, application

Description automatically generated with medium confidence

Using the above syntax in this step we did tokenization of words, where every word is separated.

**7.**

Graphical user interface, text, application, email

Description automatically generated

In this step, with the syntax, we found that there are a total of 65 words with 49 of them are repeating.

**8.**

Graphical user interface, text

Description automatically generated

Here using the above syntax we can determine how many times a word has repeated including comma and full stop.

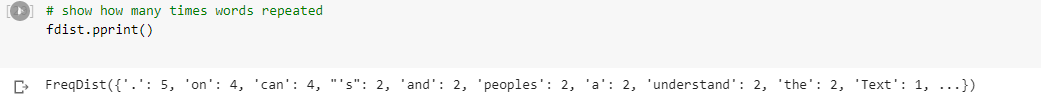
**9.**

Chart

Description automatically generated

Using the above syntax, we get output in the form of graphical representation where it shows, how many times a sample has repeated.

**10.**



In pprint syntax, the maximum number of items can be shown in a list

11.

Graphical user interface, application

Description automatically generated

In the above syntax we can find the memory word location of the indivdual word.

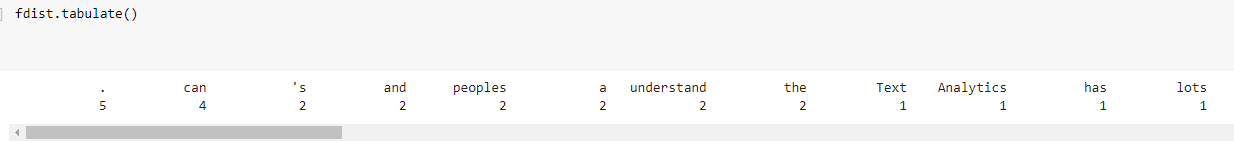
**12.**

A picture containing diagram

Description automatically generated

The number of observations in a data collection that are above (or below) a certain value is calculated using cumulative frequency syntax.

**13.**



**Conclusion:** In this ICP I have learned how to use Google collab and performed three data analysis tasks for the given data, one challenge I faced is while setting up the path for the data file in the Google collab, where I go to the file panel on the left side and select the path.