```
Start coding or generate with AI.
                                                                                                                   EDS Activity-1
from google.colab import files
                                                                                                                   Name: Srushti Satish
uploaded = files.upload()
                                                                                                                   Bhivare
     Choose Files df_file.csv
                                                                                                                   Class:ET2
       df_file.csv(text/csv) - 5097048 bytes, last modified: 12/4/2023 - 100% done
                                                                                                                   Roll No.; ET2-54
     Saving df_file.csv to df_file (2).csv
                                                                                                                   PRN-202401070179
import pandas as pd
# Load the dataset
df = pd.read_csv('/content/df_file.csv')
# Display the first 5 rows to verify
df.head()
→
                                                              畾
                                               Text Label
      0 Budget to set scene for election\n \n Gordon B...
                                                         0
                                                              ıl.
           Army chiefs in regiments decision\n \n Militar...
                                                         0
      2 Howard denies split over ID cards\n \n Michael...
                                                         0
         Observers to monitor UK election\n \n Minister...
                                                         0
          Kilroy names election seat target\n \n Ex-chat...
                                                         0
 Next steps: ( Generate code with df
                                                                    New interactive sheet

    View recommended plots

   1. Loading the Dataset and Displaying the First 5 Records
import pandas as pd
import numpy as np
# Load the dataset
df = pd.read_csv('df_file.csv')
# Display the first 5 records
df.head()
₹
                                                              扁
                                               Text Label
      0 Budget to set scene for election\n \n Gordon B...
                                                         0
                                                              ıl.
           Army chiefs in regiments decision\n \n Militar...
                                                         0
      2 Howard denies split over ID cards\n \n Michael...
                                                         0
        Observers to monitor UK election\n \n Minister...
                                                         n
          Kilroy names election seat target\n \n Ex-chat...
                                                         0
 Next steps: (
              Generate code with df
                                     View recommended plots
                                                                    New interactive sheet
   2. Finding the Shape (Rows, Columns) of the Dataset
df.shape
→ (2225, 2)
   3. Displaying the Column Names of the Dataset
df.columns
Index(['Text', 'Label'], dtype='object')
   4. Checking the Data Types of Each Column
```



dtype: object

5. Finding Missing (Null) Values

df.isnull().sum()



6. Finding the Number of Unique Labels in the 'Label' Column

```
df['Label'].nunique()
```

→ 5

7. Displaying the Count of Each Label (0s and 1s)

df['Label'].value_counts()



4 5100 417

2 401

3 386

dtype: int64

8. Calculating the Average Length of Text in the Dataset

```
df['text_length'] = df['Text'].apply(len)
np.mean(df['text_length'])
```

np.float64(2275.363595505618)

9. Finding the Minimum and Maximum Text Length

```
print("Minimum length:", df['text_length'].min())
print("Maximum length:", df['text_length'].max())
```

Minimum length: 507
Maximum length: 25597

10. Finding the Standard Deviation of Text Lengths

```
np.std(df['text_length'])
```

→ 1370.4745873382267

11. Finding the Row with the Longest Text

df.loc[df['text_length'].idxmax()]



12. Finding the Row with the Shortest Text

df.loc[df['text_length'].idxmin()]



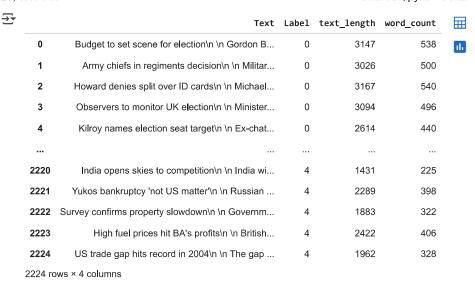
13. Adding a New Column Showing the Number of Words in Each Text

df['word_count'] = df['Text'].apply(lambda x: len(str(x).split()))
df[['Text', 'word_count']].head()



14. Finding the Average Number of Words in the Texts

15. Displaying the Texts with More Than 100 Words



16. Sorting the Dataset Based on Word Count (Highest First)

df.sort_values(by='word_count', ascending=False)

→ *		Text	Label	text_length	word_count	
	63	Terror powers expose 'tyranny'\n \n The Lord C	0	25597	4432	ıl.
	1421	Scissor Sisters triumph at Brits\n \n US band	3	19288	3482	
	412	Minimum wage increased to £5.05\n \n The mini	0	18497	3295	
	1078	Losing yourself in online gaming\n \n Online r	2	16249	2969	
	299	Kilroy launches 'Veritas' party\n \n Ex-BBC ch	0	13921	2393	
	568	Yelling takes Cardiff hat-trick\n \n European	1	810	122	
	753	Solskjaer raises hopes of return\n \n Manchest	1	725	120	
	866	Worcester v Sale (Fri)\n \n Sixways\n \n Frida	1	829	115	
	872	Tottenham bid £8m for Forest duo\n \n Not	1	738	114	
	174	Blunkett hints at election call\n \n Ex-Home S	0	507	89	
	2225 rows × 4 columns					

17. Replacing Missing Text Values with "No Text"

df['Text'].fillna('No Text', inplace=True)

<ipython-input-23-8e8ebcf65de6>:1: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series through chained assignm The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting value. For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[col] = df[col].method('No Text', inplace=True)

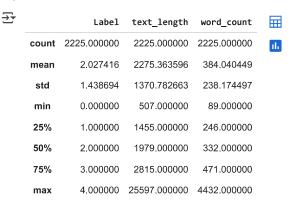
18. Finding How Many Texts Are Exactly Empty (Length 0)

len(df[df['text_length'] == 0])

→ 0

19. Checking Basic Statistics of Numerical Columns

df.describe()



20. Creating a Pivot Table Showing Average Text Length for Each Label

pd.pivot_table(df, values='text_length', index='Label', aggfunc=np.mean)

	text_length	
Label		ıl.
0	2695.824940	
1	1906.545988	
2	2987.690773	
3	1938.230570	
4	1996.194118	