```
Start coding or generate with AI.
```

```
from google.colab import files
uploaded = files.upload()
```

```
Choose Files df_file.csv
• df_file.csv(text/csv) - 5097048 bytes, last modified: 12/4/2023 - 100% done
Saving df_file.csv to df_file (2).csv
```

```
import pandas as pd

# Load the dataset
df = pd.read_csv('/content/df_file.csv')

# Display the first 5 rows to verify
df.head()
```

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()
```

2. Finding the Shape (Rows, Columns) of the Dataset

```
df.shape
```

3. Displaying the Column Names of the Dataset

```
df.columns
```

4. Checking the Data Types of Each Column

```
df.dtypes
```

5. Finding Missing (Null) Values

```
df.isnull().sum()
```

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

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

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

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

8. Calculating the Average Length of Text in the Dataset

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

9. Finding the Minimum and Maximum Text Length

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

10. Finding the Standard Deviation of Text Lengths

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

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

```
np.mean(df['word_count'])
```

15. Displaying the Texts with More Than 100 Words

```
df[df['word_count'] > 100]
```

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

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

17. Replacing Missing Text Values with "No Text"

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

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

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