In [1]:
```
pip install pandas
```

```
Collecting pandas
  Downloading pandas-1.4.2-cp38-cp38-win32.whl (9.4 MB)
Requirement already satisfied: numpy>=1.18.5 in c:\users\mahima-pc\appdata\local\prog
rams\python\python38-32\lib\site-packages (from pandas) (1.20.3)
Requirement already satisfied: python-dateutil>=2.8.1 in c:\users\mahima-pc\appdata\l
ocal\programs\python\python38-32\lib\site-packages (from pandas) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in c:\users\mahima-pc\appdata\local\progr
ams\python\python38-32\lib\site-packages (from pandas) (2020.5)
Requirement already satisfied: six>=1.5 in c:\users\mahima-pc\appdata\roaming\python
\python38\site-packages (from python-dateutil>=2.8.1->pandas) (1.14.0)
Installing collected packages: pandas
Successfully installed pandas-1.4.2
Note: you may need to restart the kernel to use updated packages.
WARNING: You are using pip version 21.1.2; however, version 22.0.4 is available.
You should consider upgrading via the 'c:\users\mahima-pc\appdata\local\programs\pyth
on\python38-32\python.exe -m pip install --upgrade pip' command.
```

In [4]:
```
import pandas as pd
url="emails.csv"
df = pd.read_csv(url)
print(df)
```

```
       Email No.  the  to  ect  and  for  of    a  you  hou  ...  connevey  \
0         Email 1    0   0    1    0    0   0    2    0    0  ...         0
1         Email 2    8  13   24    6    6   2  102    1   27  ...         0
2         Email 3    0   0    1    0    0   0    8    0    0  ...         0
3         Email 4    0   5   22    0    5   1   51    2   10  ...         0
4         Email 5    7   6   17    1    5   2   57    0    9  ...         0
...           ...  ...  ..  ...  ...  ...  ..  ...  ...  ...  ...       ...
5167  Email 5168    2   2    2    3    0   0   32    0    0  ...         0
5168  Email 5169   35  27   11    2    6   5  151    4    3  ...         0
5169  Email 5170    0   0    1    1    0   0   11    0    0  ...         0
5170  Email 5171    2   7    1    0    2   1   28    2    0  ...         0
5171  Email 5172   22  24    5    1    6   5  148    8    2  ...         0

      jay  valued  lay  infrastructure  military  allowing  ff  dry  \
0       0       0    0               0         0         0   0    0
1       0       0    0               0         0         0   1    0
2       0       0    0               0         0         0   0    0
3       0       0    0               0         0         0   0    0
4       0       0    0               0         0         0   1    0
...   ...     ...  ...             ...       ...       ...  ..  ...
5167    0       0    0               0         0         0   0    0
5168    0       0    0               0         0         0   1    0
5169    0       0    0               0         0         0   0    0
5170    0       0    0               0         0         0   1    0
5171    0       0    0               0         0         0   0    0

      Prediction
0              0
1              0
2              0
3              0
4              0
...          ...
5167           0
5168           0
5169           1
5170           1
5171           0

[5172 rows x 3002 columns]
```

In [5]:
```
print(df.describe())
```

```
                the           to          ect          and          for  \
count  5172.000000  5172.000000  5172.000000  5172.000000  5172.000000
mean      6.640565     6.188128     5.143852     3.075599     3.124710
std      11.745009     9.534576    14.101142     6.045970     4.680522
min       0.000000     0.000000     1.000000     0.000000     0.000000
25%       0.000000     1.000000     1.000000     0.000000     1.000000
50%       3.000000     3.000000     1.000000     1.000000     2.000000
75%       8.000000     7.000000     4.000000     3.000000     4.000000
max     210.000000   132.000000   344.000000    89.000000    47.000000

                of            a          you          hou           in  ...  \
count  5172.000000  5172.000000  5172.000000  5172.000000  5172.000000  ...
mean      2.627030    55.517401     2.466551     2.024362    10.600155  ...
std       6.229845    87.574172     4.314444     6.967878    19.281892  ...
min       0.000000     0.000000     0.000000     0.000000     0.000000  ...
25%       0.000000    12.000000     0.000000     0.000000     1.000000  ...
50%       1.000000    28.000000     1.000000     0.000000     5.000000  ...
75%       2.000000    62.250000     3.000000     1.000000    12.000000  ...
max      77.000000  1898.000000    70.000000   167.000000   223.000000  ...

          connevey          jay       valued          lay  infrastructure  \
count  5172.000000  5172.000000  5172.000000  5172.000000     5172.000000
mean      0.005027     0.012568     0.010634     0.098028        0.004254
std       0.105788     0.199682     0.116693     0.569532        0.096252
min       0.000000     0.000000     0.000000     0.000000        0.000000
25%       0.000000     0.000000     0.000000     0.000000        0.000000
50%       0.000000     0.000000     0.000000     0.000000        0.000000
75%       0.000000     0.000000     0.000000     0.000000        0.000000
max       4.000000     7.000000     2.000000    12.000000        3.000000

          military     allowing           ff          dry   Prediction
count  5172.000000  5172.000000  5172.000000  5172.000000  5172.000000
mean      0.006574     0.004060     0.914733     0.006961     0.290023
std       0.138908     0.072145     2.780203     0.098086     0.453817
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       0.000000     0.000000     0.000000     0.000000     0.000000
50%       0.000000     0.000000     0.000000     0.000000     0.000000
75%       0.000000     0.000000     1.000000     0.000000     1.000000
max       4.000000     3.000000   114.000000     4.000000     1.000000

[8 rows x 3001 columns]
```

In [6]:
```
print(df.isnull().values.any())
```

```
False
```

In [7]:
```
print("Size", df.size)
print("Tuple Shape", df.shape)
print("Dimension", df.ndim)
```

```
Size 15526344
Tuple Shape (5172, 3002)
Dimension 2
```

In [8]:
```
print(df.head)
print(df.dtypes)
print("INFO",df.info())
```

```
<bound method NDFrame.head of        Email No.  the  to  ect  and  for  of    a  you
hou  ...  connevey  \
0        Email 1    0   0    1    0    0   0    2    0    0  ...         0
1        Email 2    8  13   24    6    6   2  102    1   27  ...         0
2        Email 3    0   0    1    0    0   0    8    0    0  ...         0
3        Email 4    0   5   22    0    5   1   51    2   10  ...         0
4        Email 5    7   6   17    1    5   2   57    0    9  ...         0
...          ...  ...  ..  ...  ...  ...  ..  ...  ...  ...  ...       ...
5167  Email 5168    2   2    2    3    0   0   32    0    0  ...         0
5168  Email 5169   35  27   11    2    6   5  151    4    3  ...         0
5169  Email 5170    0   0    1    1    0   0   11    0    0  ...         0
5170  Email 5171    2   7    1    0    2   1   28    2    0  ...         0
5171  Email 5172   22  24    5    1    6   5  148    8    2  ...         0

      jay  valued  lay  infrastructure  military  allowing  ff  dry  \
0       0       0    0               0         0         0   0    0
1       0       0    0               0         0         0   1    0
2       0       0    0               0         0         0   0    0
3       0       0    0               0         0         0   0    0
4       0       0    0               0         0         0   1    0
...   ...     ...  ...             ...       ...       ...  ..  ...
5167    0       0    0               0         0         0   0    0
5168    0       0    0               0         0         0   1    0
5169    0       0    0               0         0         0   0    0
5170    0       0    0               0         0         0   1    0
5171    0       0    0               0         0         0   0    0

      Prediction
0              0
1              0
2              0
3              0
4              0
...          ...
5167           0
5168           0
5169           1
5170           1
5171           0

[5172 rows x 3002 columns]>
Email No.     object
the            int64
to             int64
ect            int64
and            int64
               ...
military       int64
allowing       int64
ff             int64
dry            int64
Prediction     int64
Length: 3002, dtype: object
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5172 entries, 0 to 5171
Columns: 3002 entries, Email No. to Prediction
dtypes: int64(3001), object(1)
memory usage: 118.4+ MB
INFO None
```

In [ ]: