Practical 1 DSABDA

May 14, 2023

```
[2]: import numpy as np
     import pandas as pd
     from sklearn import preprocessing
```

```
[4]: #Importing the file
     df=pd.read_csv("D:\Data science practical\spambase_csv.csv")
```

```
[6]: df
```

```
[6]:       word_freq_make  word_freq_address  word_freq_all  word_freq_3d  \
     0              0.00               0.64           0.64           0.0
     1              0.21               0.28           0.50           0.0
     2              0.06               0.00           0.71           0.0
     3              0.00               0.00           0.00           0.0
     4              0.00               0.00           0.00           0.0
     ...             ...                ...            ...           ...
     4596           0.31               0.00           0.62           0.0
     4597           0.00               0.00           0.00           0.0
     4598           0.30               0.00           0.30           0.0
     4599           0.96               0.00           0.00           0.0
     4600           0.00               0.00           0.65           0.0

           word_freq_our  word_freq_over  word_freq_remove  word_freq_internet  \
     0              0.32            0.00              0.00                0.00
     1              0.14            0.28              0.21                0.07
     2              1.23            0.19              0.19                0.12
     3              0.63            0.00              0.31                0.63
     4              0.63            0.00              0.31                0.63
     ...             ...             ...               ...                 ...
     4596           0.00            0.31              0.00                0.00
     4597           0.00            0.00              0.00                0.00
     4598           0.00            0.00              0.00                0.00
     4599           0.32            0.00              0.00                0.00
     4600           0.00            0.00              0.00                0.00

           word_freq_order  word_freq_mail  …  char_freq_%3B  char_freq_%28  \
     0                 0.00            0.00  …          0.000          0.000
     1                 0.00            0.94  …          0.000          0.132
```

1

2	0.64	0.25	…	0.010	0.143
3	0.31	0.63	…	0.000	0.137
4	0.31	0.63	…	0.000	0.135
…	…	…	…	…	…
4596	0.00	0.00	…	0.000	0.232
4597	0.00	0.00	…	0.000	0.000
4598	0.00	0.00	…	0.102	0.718
4599	0.00	0.00	…	0.000	0.057
4600	0.00	0.00	…	0.000	0.000

	char_freq_%5B	char_freq_%21	char_freq_%24	char_freq_%23	\
0	0.0	0.778	0.000	0.000	
1	0.0	0.372	0.180	0.048	
2	0.0	0.276	0.184	0.010	
3	0.0	0.137	0.000	0.000	
4	0.0	0.135	0.000	0.000	
…	…	…	…	…	
4596	0.0	0.000	0.000	0.000	
4597	0.0	0.353	0.000	0.000	
4598	0.0	0.000	0.000	0.000	
4599	0.0	0.000	0.000	0.000	
4600	0.0	0.125	0.000	0.000	

	capital_run_length_average	capital_run_length_longest	\
0	3.756	61	
1	5.114	101	
2	9.821	485	
3	3.537	40	
4	3.537	40	
…	…	…	
4596	1.142	3	
4597	1.555	4	
4598	1.404	6	
4599	1.147	5	
4600	1.250	5	

	capital_run_length_total	class
0	278	1
1	1028	1
2	2259	1
3	191	1
4	191	1
…	…	…
4596	88	0
4597	14	0
4598	118	0
4599	78	0

```
4600                          40       0

[4601 rows x 58 columns]
```

```
[7]: #data visualization
     df.head(n=10)
```

```
[7]:    word_freq_make  word_freq_address  word_freq_all  word_freq_3d  \
     0            0.00               0.64           0.64           0.0
     1            0.21               0.28           0.50           0.0
     2            0.06               0.00           0.71           0.0
     3            0.00               0.00           0.00           0.0
     4            0.00               0.00           0.00           0.0
     5            0.00               0.00           0.00           0.0
     6            0.00               0.00           0.00           0.0
     7            0.00               0.00           0.00           0.0
     8            0.15               0.00           0.46           0.0
     9            0.06               0.12           0.77           0.0

        word_freq_our  word_freq_over  word_freq_remove  word_freq_internet  \
     0           0.32            0.00              0.00                0.00
     1           0.14            0.28              0.21                0.07
     2           1.23            0.19              0.19                0.12
     3           0.63            0.00              0.31                0.63
     4           0.63            0.00              0.31                0.63
     5           1.85            0.00              0.00                1.85
     6           1.92            0.00              0.00                0.00
     7           1.88            0.00              0.00                1.88
     8           0.61            0.00              0.30                0.00
     9           0.19            0.32              0.38                0.00

        word_freq_order  word_freq_mail  …  char_freq_%3B  char_freq_%28  \
     0             0.00            0.00  …           0.00          0.000
     1             0.00            0.94  …           0.00          0.132
     2             0.64            0.25  …           0.01          0.143
     3             0.31            0.63  …           0.00          0.137
     4             0.31            0.63  …           0.00          0.135
     5             0.00            0.00  …           0.00          0.223
     6             0.00            0.64  …           0.00          0.054
     7             0.00            0.00  …           0.00          0.206
     8             0.92            0.76  …           0.00          0.271
     9             0.06            0.00  …           0.04          0.030

        char_freq_%5B  char_freq_%21  char_freq_%24  char_freq_%23  \
     0            0.0          0.778          0.000          0.000
     1            0.0          0.372          0.180          0.048
     2            0.0          0.276          0.184          0.010
```

```
3              0.0          0.137              0.000              0.000
4              0.0          0.135              0.000              0.000
5              0.0          0.000              0.000              0.000
6              0.0          0.164              0.054              0.000
7              0.0          0.000              0.000              0.000
8              0.0          0.181              0.203              0.022
9              0.0          0.244              0.081              0.000

   capital_run_length_average  capital_run_length_longest  \
0                       3.756                          61
1                       5.114                         101
2                       9.821                         485
3                       3.537                          40
4                       3.537                          40
5                       3.000                          15
6                       1.671                           4
7                       2.450                          11
8                       9.744                         445
9                       1.729                          43

   capital_run_length_total  class
0                       278      1
1                      1028      1
2                      2259      1
3                       191      1
4                       191      1
5                        54      1
6                       112      1
7                        49      1
8                      1257      1
9                       749      1

[10 rows x 58 columns]
```

[8]: #Indexing
     df.index

[8]: RangeIndex(start=0, stop=4601, step=1)

[9]: # TO get no of rows and columns
     df.shape

[9]: (4601, 58)

[10]: #To check whether the target attribute is binary or not
      np.unique(df['class'])

```
[10]: array([0, 1], dtype=int64)
```

```
[11]: df.columns.values
```

```
[11]: array(['word_freq_make', 'word_freq_address', 'word_freq_all',
             'word_freq_3d', 'word_freq_our', 'word_freq_over',
             'word_freq_remove', 'word_freq_internet', 'word_freq_order',
             'word_freq_mail', 'word_freq_receive', 'word_freq_will',
             'word_freq_people', 'word_freq_report', 'word_freq_addresses',
             'word_freq_free', 'word_freq_business', 'word_freq_email',
             'word_freq_you', 'word_freq_credit', 'word_freq_your',
             'word_freq_font', 'word_freq_000', 'word_freq_money',
             'word_freq_hp', 'word_freq_hpl', 'word_freq_george',
             'word_freq_650', 'word_freq_lab', 'word_freq_labs',
             'word_freq_telnet', 'word_freq_857', 'word_freq_data',
             'word_freq_415', 'word_freq_85', 'word_freq_technology',
             'word_freq_1999', 'word_freq_parts', 'word_freq_pm',
             'word_freq_direct', 'word_freq_cs', 'word_freq_meeting',
             'word_freq_original', 'word_freq_project', 'word_freq_re',
             'word_freq_edu', 'word_freq_table', 'word_freq_conference',
             'char_freq_%3B', 'char_freq_%28', 'char_freq_%5B', 'char_freq_%21',
             'char_freq_%24', 'char_freq_%23', 'capital_run_length_average',
             'capital_run_length_longest', 'capital_run_length_total', 'class'],
            dtype=object)
```

```
[12]: df['word_freq_address']
```

```
[12]: 0        0.64
      1        0.28
      2        0.00
      3        0.00
      4        0.00
               ...
      4596     0.00
      4597     0.00
      4598     0.00
      4599     0.00
      4600     0.00
      Name: word_freq_address, Length: 4601, dtype: float64
```

```
[13]: df.isnull()
```

```
[13]:        word_freq_make  word_freq_address  word_freq_all  word_freq_3d  \
      0               False              False          False         False
      1               False              False          False         False
      2               False              False          False         False
      3               False              False          False         False
      4               False              False          False         False
```

```
...          ...                ...             ...             ...
4596         False              False           False           False
4597         False              False           False           False
4598         False              False           False           False
4599         False              False           False           False
4600         False              False           False           False
```

	word_freq_our	word_freq_over	word_freq_remove	word_freq_internet \
0	False	False	False	False
1	False	False	False	False
2	False	False	False	False
3	False	False	False	False
4	False	False	False	False
...	...	...	...	...
4596	False	False	False	False
4597	False	False	False	False
4598	False	False	False	False
4599	False	False	False	False
4600	False	False	False	False

	word_freq_order	word_freq_mail	...	char_freq_%3B	char_freq_%28 \
0	False	False	...	False	False
1	False	False	...	False	False
2	False	False	...	False	False
3	False	False	...	False	False
4	False	False	...	False	False
...	...	... ...	...	...	...
4596	False	False	...	False	False
4597	False	False	...	False	False
4598	False	False	...	False	False
4599	False	False	...	False	False
4600	False	False	...	False	False

	char_freq_%5B	char_freq_%21	char_freq_%24	char_freq_%23 \
0	False	False	False	False
1	False	False	False	False
2	False	False	False	False
3	False	False	False	False
4	False	False	False	False
...	...	...	...	...
4596	False	False	False	False
4597	False	False	False	False
4598	False	False	False	False
4599	False	False	False	False
4600	False	False	False	False

```
        capital_run_length_average  capital_run_length_longest  \
```

```
0                              False                   False
1                              False                   False
2                              False                   False
3                              False                   False
4                              False                   False
...                              ...                     ...
4596                           False                   False
4597                           False                   False
4598                           False                   False
4599                           False                   False
4600                           False                   False

      capital_run_length_total  class
0                        False  False
1                        False  False
2                        False  False
3                        False  False
4                        False  False
...                        ...    ...
4596                     False  False
4597                     False  False
4598                     False  False
4599                     False  False
4600                     False  False

[4601 rows x 58 columns]
```

[14]: `df.iloc`

[14]: <pandas.core.indexing._iLocIndexer at 0x27920800a40>

[15]: `df.isnull().ne`

[15]: <bound method flex_comp_method_FRAME.<locals>.f of   word_freq_make
```
      word_freq_address  word_freq_all  word_freq_3d  \
0                 False          False         False          False
1                 False          False         False          False
2                 False          False         False          False
3                 False          False         False          False
4                 False          False         False          False
...                 ...            ...           ...            ...
4596              False          False         False          False
4597              False          False         False          False
4598              False          False         False          False
4599              False          False         False          False
4600              False          False         False          False
```

	word_freq_our	word_freq_over	word_freq_remove	word_freq_internet \
0	False	False	False	False
1	False	False	False	False
2	False	False	False	False
3	False	False	False	False
4	False	False	False	False
...	...	...	...	...
4596	False	False	False	False
4597	False	False	False	False
4598	False	False	False	False
4599	False	False	False	False
4600	False	False	False	False

	word_freq_order	word_freq_mail	...	char_freq_%3B	char_freq_%28 \
0	False	False	...	False	False
1	False	False	...	False	False
2	False	False	...	False	False
3	False	False	...	False	False
4	False	False	...	False	False
...	...	...	...	...	...
4596	False	False	...	False	False
4597	False	False	...	False	False
4598	False	False	...	False	False
4599	False	False	...	False	False
4600	False	False	...	False	False

	char_freq_%5B	char_freq_%21	char_freq_%24	char_freq_%23 \
0	False	False	False	False
1	False	False	False	False
2	False	False	False	False
3	False	False	False	False
4	False	False	False	False
...	...	...	...	...
4596	False	False	False	False
4597	False	False	False	False
4598	False	False	False	False
4599	False	False	False	False
4600	False	False	False	False

	capital_run_length_average	capital_run_length_longest \
0	False	False
1	False	False
2	False	False
3	False	False
4	False	False
...	...	...
4596	False	False

```
4597                                    False                              False
4598                                    False                              False
4599                                    False                              False
4600                                    False                              False

        capital_run_length_total  class
0                          False  False
1                          False  False
2                          False  False
3                          False  False
4                          False  False
...                          ...    ...
4596                       False  False
4597                       False  False
4598                       False  False
4599                       False  False
4600                       False  False

[4601 rows x 58 columns]>
```

[16]: `df.isnull().sum().sum()`

[16]: 0

[17]: `df.isnull().sum(axis=1)`

[17]:
```
0       0
1       0
2       0
3       0
4       0
       ..
4596    0
4597    0
4598    0
4599    0
4600    0
Length: 4601, dtype: int64
```

[18]: `df.word_freq_remove.isnull().sum()`

[18]: 0

[19]: `df.groupby(['word_freq_address'])['word_freq_make'].apply(lambda x:x.isnull().sum())`

```
[19]: word_freq_address
      0.00      0
      0.01      0
      0.02      0
      0.03      0
      0.04      0
              ..
      5.47      0
      6.34      0
      7.01      0
      9.52      0
      14.28     0
      Name: word_freq_make, Length: 171, dtype: int64
```

```
[20]: df.describe(include='all')
```

```
[20]:          word_freq_make   word_freq_address   word_freq_all   word_freq_3d  \
      count      4601.000000         4601.000000     4601.000000    4601.000000
      mean          0.104553            0.213015        0.280656       0.065425
      std           0.305358            1.290575        0.504143       1.395151
      min           0.000000            0.000000        0.000000       0.000000
      25%           0.000000            0.000000        0.000000       0.000000
      50%           0.000000            0.000000        0.000000       0.000000
      75%           0.000000            0.000000        0.420000       0.000000
      max           4.540000           14.280000        5.100000      42.810000

               word_freq_our   word_freq_over   word_freq_remove   word_freq_internet  \
      count      4601.000000      4601.000000        4601.000000          4601.000000
      mean          0.312223         0.095901           0.114208             0.105295
      std           0.672513         0.273824           0.391441             0.401071
      min           0.000000         0.000000           0.000000             0.000000
      25%           0.000000         0.000000           0.000000             0.000000
      50%           0.000000         0.000000           0.000000             0.000000
      75%           0.380000         0.000000           0.000000             0.000000
      max          10.000000         5.880000           7.270000            11.110000

               word_freq_order   word_freq_mail   …   char_freq_%3B   char_freq_%28  \
      count       4601.000000      4601.000000     …     4601.000000     4601.000000
      mean           0.090067         0.239413     …        0.038575        0.139030
      std            0.278616         0.644755     …        0.243471        0.270355
      min            0.000000         0.000000     …        0.000000        0.000000
      25%            0.000000         0.000000     …        0.000000        0.000000
      50%            0.000000         0.000000     …        0.000000        0.065000
      75%            0.000000         0.160000     …        0.000000        0.188000
      max            5.260000        18.180000     …        4.385000        9.752000

               char_freq_%5B   char_freq_%21   char_freq_%24   char_freq_%23  \
```

```
count   4601.000000   4601.000000   4601.000000   4601.000000
mean       0.016976      0.269071      0.075811      0.044238
std        0.109394      0.815672      0.245882      0.429342
min        0.000000      0.000000      0.000000      0.000000
25%        0.000000      0.000000      0.000000      0.000000
50%        0.000000      0.000000      0.000000      0.000000
75%        0.000000      0.315000      0.052000      0.000000
max        4.081000     32.478000      6.003000     19.829000

       capital_run_length_average  capital_run_length_longest  \
count                 4601.000000                 4601.000000
mean                     5.191515                   52.172789
std                     31.729449                  194.891310
min                      1.000000                    1.000000
25%                      1.588000                    6.000000
50%                      2.276000                   15.000000
75%                      3.706000                   43.000000
max                   1102.500000                 9989.000000

       capital_run_length_total        class
count               4601.000000  4601.000000
mean                 283.289285     0.394045
std                  606.347851     0.488698
min                    1.000000     0.000000
25%                   35.000000     0.000000
50%                   95.000000     0.000000
75%                  266.000000     1.000000
max                15841.000000     1.000000

[8 rows x 58 columns]
```

```python
[21]: df.sort_index(axis=1,ascending=False)
```

```
[21]:       word_freq_your  word_freq_you  word_freq_will  word_freq_telnet  \
      0               0.96           1.93            0.64               0.0
      1               1.59           3.47            0.79               0.0
      2               0.51           1.36            0.45               0.0
      3               0.31           3.18            0.31               0.0
      4               0.31           3.18            0.31               0.0
      ...              ...            ...             ...               ...
      4596            0.00           0.62            1.88               0.0
      4597            2.00           6.00            0.00               0.0
      4598            0.30           1.50            1.80               0.0
      4599            0.32           1.93            0.32               0.0
      4600            0.65           4.60            0.00               0.0

            word_freq_technology  word_freq_table  word_freq_report  \
```

0	0.0	0.0	0.00
1	0.0	0.0	0.21
2	0.0	0.0	0.00
3	0.0	0.0	0.00
4	0.0	0.0	0.00
…	…	…	…
4596	0.0	0.0	0.00
4597	0.0	0.0	0.00
4598	0.0	0.0	0.00
4599	0.0	0.0	0.00
4600	0.0	0.0	0.00

	word_freq_remove	word_freq_receive	word_freq_re	…	class \
0	0.00	0.00	0.00	…	1
1	0.21	0.21	0.00	…	1
2	0.19	0.38	0.06	…	1
3	0.31	0.31	0.00	…	1
4	0.31	0.31	0.00	…	1
…	…	…	… …	…	…
4596	0.00	0.00	0.31	…	0
4597	0.00	0.00	0.00	…	0
4598	0.00	0.00	0.00	…	0
4599	0.00	0.00	0.00	…	0
4600	0.00	0.00	1.97	…	0

	char_freq_%5B	char_freq_%3B	char_freq_%28	char_freq_%24 \
0	0.0	0.000	0.000	0.000
1	0.0	0.000	0.132	0.180
2	0.0	0.010	0.143	0.184
3	0.0	0.000	0.137	0.000
4	0.0	0.000	0.135	0.000
…	…	…	…	…
4596	0.0	0.000	0.232	0.000
4597	0.0	0.000	0.000	0.000
4598	0.0	0.102	0.718	0.000
4599	0.0	0.000	0.057	0.000
4600	0.0	0.000	0.000	0.000

	char_freq_%23	char_freq_%21	capital_run_length_total \
0	0.000	0.778	278
1	0.048	0.372	1028
2	0.010	0.276	2259
3	0.000	0.137	191
4	0.000	0.135	191
…	…	…	…
4596	0.000	0.000	88
4597	0.000	0.353	14

```
4598              0.000            0.000                        118
4599              0.000            0.000                         78
4600              0.000            0.125                         40

      capital_run_length_longest  capital_run_length_average
0                             61                       3.756
1                            101                       5.114
2                            485                       9.821
3                             40                       3.537
4                             40                       3.537
...                          ...                         ...
4596                           3                       1.142
4597                           4                       1.555
4598                           6                       1.404
4599                           5                       1.147
4600                           5                       1.250

[4601 rows x 58 columns]
```

[22]: `df.sort_values(by="word_freq_address")`

[22]:
```
      word_freq_make  word_freq_address  word_freq_all  word_freq_3d  \
2300            0.00               0.00           0.57           0.0
2967            0.86               0.00           0.00           0.0
2968            0.00               0.00           0.24           0.0
2969            0.00               0.00           0.00           0.0
2970            0.00               0.00           0.00           0.0
...              ...                ...            ...           ...
3128            0.00              14.28           0.00           0.0
3127            0.00              14.28           0.00           0.0
3126            0.00              14.28           0.00           0.0
3121            0.00              14.28           0.00           0.0
3143            0.00              14.28           0.00           0.0

      word_freq_our  word_freq_over  word_freq_remove  word_freq_internet  \
2300           0.00             0.0               0.0                 0.0
2967           0.00             0.0               0.0                 0.0
2968           0.00             0.0               0.0                 0.0
2969           0.00             0.0               0.0                 0.0
2970           1.19             0.0               0.0                 0.0
...             ...             ...               ...                 ...
3128           0.00             0.0               0.0                 0.0
3127           0.00             0.0               0.0                 0.0
3126           0.00             0.0               0.0                 0.0
3121           0.00             0.0               0.0                 0.0
3143           0.00             0.0               0.0                 0.0
```

	word_freq_order	word_freq_mail	…	char_freq_%3B	char_freq_%28 \
2300	0.0	0.00	…	0.000	0.093
2967	0.0	0.00	…	0.561	0.140
2968	0.0	0.48	…	0.000	0.035
2969	0.0	0.00	…	0.000	0.000
2970	0.0	0.00	…	0.000	0.000
…	…	…	…	…	…
3128	0.0	0.00	…	0.000	0.000
3127	0.0	0.00	…	0.000	0.000
3126	0.0	0.00	…	0.000	0.000
3121	0.0	0.00	…	0.000	0.000
3143	0.0	0.00	…	0.000	0.000

	char_freq_%5B	char_freq_%21	char_freq_%24	char_freq_%23 \
2300	0.0	0.0	0.0	0.0
2967	0.0	0.0	0.0	0.0
2968	0.0	0.0	0.0	0.0
2969	0.0	0.0	0.0	0.0
2970	0.0	0.0	0.0	0.0
…	…	…	…	…
3128	0.0	0.0	0.0	0.0
3127	0.0	0.0	0.0	0.0
3126	0.0	0.0	0.0	0.0
3121	0.0	0.0	0.0	0.0
3143	0.0	0.0	0.0	0.0

	capital_run_length_average	capital_run_length_longest \
2300	1.136	3
2967	1.352	6
2968	3.179	87
2969	2.607	11
2970	1.733	14
…	…	…
3128	1.800	5
3127	1.800	5
3126	1.800	5
3121	1.800	5
3143	1.800	5

	capital_run_length_total	class
2300	25	0
2967	23	0
2968	283	0
2969	219	0
2970	52	0
…	…	…
3128	9	0

```
3127                              9        0
3126                              9        0
3121                              9        0
3143                              9        0

[4601 rows x 58 columns]
```

[23]: `df.iloc[5]`

[23]:
```
word_freq_make            0.000
word_freq_address         0.000
word_freq_all             0.000
word_freq_3d              0.000
word_freq_our             1.850
word_freq_over            0.000
word_freq_remove          0.000
word_freq_internet        1.850
word_freq_order           0.000
word_freq_mail            0.000
word_freq_receive         0.000
word_freq_will            0.000
word_freq_people          0.000
word_freq_report          0.000
word_freq_addresses       0.000
word_freq_free            0.000
word_freq_business        0.000
word_freq_email           0.000
word_freq_you             0.000
word_freq_credit          0.000
word_freq_your            0.000
word_freq_font            0.000
word_freq_000             0.000
word_freq_money           0.000
word_freq_hp              0.000
word_freq_hpl             0.000
word_freq_george          0.000
word_freq_650             0.000
word_freq_lab             0.000
word_freq_labs            0.000
word_freq_telnet          0.000
word_freq_857             0.000
word_freq_data            0.000
word_freq_415             0.000
word_freq_85              0.000
word_freq_technology      0.000
word_freq_1999            0.000
word_freq_parts           0.000
```

```
word_freq_pm                   0.000
word_freq_direct               0.000
word_freq_cs                   0.000
word_freq_meeting              0.000
word_freq_original             0.000
word_freq_project              0.000
word_freq_re                   0.000
word_freq_edu                  0.000
word_freq_table                0.000
word_freq_conference           0.000
char_freq_%3B                  0.000
char_freq_%28                  0.223
char_freq_%5B                  0.000
char_freq_%21                  0.000
char_freq_%24                  0.000
char_freq_%23                  0.000
capital_run_length_average     3.000
capital_run_length_longest    15.000
capital_run_length_total      54.000
class                          1.000
Name: 5, dtype: float64
```

[24]: `df[0:3]`

[24]:
```
   word_freq_make  word_freq_address  word_freq_all  word_freq_3d  \
0            0.00               0.64           0.64           0.0
1            0.21               0.28           0.50           0.0
2            0.06               0.00           0.71           0.0

   word_freq_our  word_freq_over  word_freq_remove  word_freq_internet  \
0           0.32            0.00              0.00                0.00
1           0.14            0.28              0.21                0.07
2           1.23            0.19              0.19                0.12

   word_freq_order  word_freq_mail  …  char_freq_%3B  char_freq_%28  \
0             0.00            0.00  …           0.00          0.000
1             0.00            0.94  …           0.00          0.132
2             0.64            0.25  …           0.01          0.143

   char_freq_%5B  char_freq_%21  char_freq_%24  char_freq_%23  \
0            0.0          0.778          0.000          0.000
1            0.0          0.372          0.180          0.048
2            0.0          0.276          0.184          0.010

   capital_run_length_average  capital_run_length_longest  \
0                       3.756                          61
1                       5.114                         101
```

```
2                      9.821                           485

    capital_run_length_total  class
0                       278      1
1                      1028      1
2                      2259      1

[3 rows x 58 columns]
```

[25]: `df.loc[:,["word_freq_make","word_freq_address"]]`

```
[25]:       word_freq_make  word_freq_address
0                 0.00               0.64
1                 0.21               0.28
2                 0.06               0.00
3                 0.00               0.00
4                 0.00               0.00
...                ...                ...
4596              0.31               0.00
4597              0.00               0.00
4598              0.30               0.00
4599              0.96               0.00
4600              0.00               0.00

[4601 rows x 2 columns]
```

[26]: `df.iloc[:4,:]`

```
[26]:    word_freq_make  word_freq_address  word_freq_all  word_freq_3d  \
0                 0.00               0.64           0.64           0.0
1                 0.21               0.28           0.50           0.0
2                 0.06               0.00           0.71           0.0
3                 0.00               0.00           0.00           0.0

    word_freq_our  word_freq_over  word_freq_remove  word_freq_internet  \
0            0.32            0.00              0.00                0.00
1            0.14            0.28              0.21                0.07
2            1.23            0.19              0.19                0.12
3            0.63            0.00              0.31                0.63

    word_freq_order  word_freq_mail  …  char_freq_%3B  char_freq_%28  \
0              0.00            0.00  …           0.00          0.000
1              0.00            0.94  …           0.00          0.132
2              0.64            0.25  …           0.01          0.143
3              0.31            0.63  …           0.00          0.137

    char_freq_%5B  char_freq_%21  char_freq_%24  char_freq_%23  \
```

```
0        0.0        0.778        0.000        0.000
1        0.0        0.372        0.180        0.048
2        0.0        0.276        0.184        0.010
3        0.0        0.137        0.000        0.000

   capital_run_length_average  capital_run_length_longest  \
0                       3.756                          61
1                       5.114                         101
2                       9.821                         485
3                       3.537                          40

   capital_run_length_total  class
0                       278      1
1                      1028      1
2                      2259      1
3                       191      1

[4 rows x 58 columns]
```

[27]: `df.iloc[:, :4]`

[27]:
```
      word_freq_make  word_freq_address  word_freq_all  word_freq_3d
0               0.00               0.64           0.64           0.0
1               0.21               0.28           0.50           0.0
2               0.06               0.00           0.71           0.0
3               0.00               0.00           0.00           0.0
4               0.00               0.00           0.00           0.0
...              ...                ...            ...           ...
4596            0.31               0.00           0.62           0.0
4597            0.00               0.00           0.00           0.0
4598            0.30               0.00           0.30           0.0
4599            0.96               0.00           0.00           0.0
4600            0.00               0.00           0.65           0.0

[4601 rows x 4 columns]
```

[28]: `df.iloc[:4, :6]`

[28]:
```
    word_freq_make  word_freq_address  word_freq_all  word_freq_3d  \
0             0.00               0.64           0.64           0.0
1             0.21               0.28           0.50           0.0
2             0.06               0.00           0.71           0.0
3             0.00               0.00           0.00           0.0

   word_freq_our  word_freq_over
0           0.32            0.00
1           0.14            0.28
```

```
2              1.23              0.19
3              0.63              0.00
```

[29]: `df.iloc[3:5, 0:2]`

[29]:
```
   word_freq_make  word_freq_address
3             0.0                0.0
4             0.0                0.0
```

[30]: `df.iloc[[1,2,4],[0,2]]`

[30]:
```
   word_freq_make  word_freq_all
1            0.21           0.50
2            0.06           0.71
4            0.00           0.00
```

[31]: `df.iloc[3:5,:]`

[31]:
```
   word_freq_make  word_freq_address  word_freq_all  word_freq_3d  \
3             0.0                0.0            0.0           0.0
4             0.0                0.0            0.0           0.0

   word_freq_our  word_freq_over  word_freq_remove  word_freq_internet  \
3           0.63             0.0              0.31                0.63
4           0.63             0.0              0.31                0.63

   word_freq_order  word_freq_mail  …  char_freq_%3B  char_freq_%28  \
3             0.31            0.63  …            0.0          0.137
4             0.31            0.63  …            0.0          0.135

   char_freq_%5B  char_freq_%21  char_freq_%24  char_freq_%23  \
3            0.0          0.137            0.0            0.0
4            0.0          0.135            0.0            0.0

   capital_run_length_average  capital_run_length_longest  \
3                       3.537                          40
4                       3.537                          40

   capital_run_length_total  class
3                       191      1
4                       191      1

[2 rows x 58 columns]
```

[32]: `df.iloc[:, 1:3]`
```

```
[32]:        word_freq_address  word_freq_all
    0                   0.64           0.64
    1                   0.28           0.50
    2                   0.00           0.71
    3                   0.00           0.00
    4                   0.00           0.00
    ...                  ...            ...
    4596                0.00           0.62
    4597                0.00           0.00
    4598                0.00           0.30
    4599                0.00           0.00
    4600                0.00           0.65

    [4601 rows x 2 columns]
```

```
[33]: df.iloc[1:1]
```

```
[33]: Empty DataFrame
    Columns: [word_freq_make, word_freq_address, word_freq_all, word_freq_3d,
    word_freq_our, word_freq_over, word_freq_remove, word_freq_internet,
    word_freq_order, word_freq_mail, word_freq_receive, word_freq_will,
    word_freq_people, word_freq_report, word_freq_addresses, word_freq_free,
    word_freq_business, word_freq_email, word_freq_you, word_freq_credit,
    word_freq_your, word_freq_font, word_freq_000, word_freq_money, word_freq_hp,
    word_freq_hpl, word_freq_george, word_freq_650, word_freq_lab, word_freq_labs,
    word_freq_telnet, word_freq_857, word_freq_data, word_freq_415, word_freq_85,
    word_freq_technology, word_freq_1999, word_freq_parts, word_freq_pm,
    word_freq_direct, word_freq_cs, word_freq_meeting, word_freq_original,
    word_freq_project, word_freq_re, word_freq_edu, word_freq_table,
    word_freq_conference, char_freq_%3B, char_freq_%28, char_freq_%5B,
    char_freq_%21, char_freq_%24, char_freq_%23, capital_run_length_average,
    capital_run_length_longest, capital_run_length_total, class]
    Index: []

    [0 rows x 58 columns]
```

```
[34]: col_1_4=df.columns[1:4]
```

```
[35]: df[df.columns[2:4]].iloc[5:10]
```

```
[35]:    word_freq_all  word_freq_3d
    5            0.00           0.0
    6            0.00           0.0
    7            0.00           0.0
    8            0.46           0.0
    9            0.77           0.0
```

```
[36]: min_max_scaler=preprocessing.MinMaxScaler()
```

```
[37]: x=df.iloc[:,:4]
```

```
[38]: x_scaled=min_max_scaler.fit_transform(x)
```

```
[39]: df_normalized= pd.DataFrame(x_scaled)
```

```
[40]: df_normalized
```

```
[40]:               0         1         2    3
      0      0.000000  0.044818  0.125490  0.0
      1      0.046256  0.019608  0.098039  0.0
      2      0.013216  0.000000  0.139216  0.0
      3      0.000000  0.000000  0.000000  0.0
      4      0.000000  0.000000  0.000000  0.0
      ...         ...       ...       ...  ...
      4596   0.068282  0.000000  0.121569  0.0
      4597   0.000000  0.000000  0.000000  0.0
      4598   0.066079  0.000000  0.058824  0.0
      4599   0.211454  0.000000  0.000000  0.0
      4600   0.000000  0.000000  0.127451  0.0

      [4601 rows x 4 columns]
```

```
[41]: df.head
```

```
[41]: <bound method NDFrame.head of        word_freq_make  word_freq_address
      word_freq_all  word_freq_3d  \
      0              0.00              0.64              0.64              0.0
      1              0.21              0.28              0.50              0.0
      2              0.06              0.00              0.71              0.0
      3              0.00              0.00              0.00              0.0
      4              0.00              0.00              0.00              0.0
      ...             ...               ...               ...               ...
      4596           0.31              0.00              0.62              0.0
      4597           0.00              0.00              0.00              0.0
      4598           0.30              0.00              0.30              0.0
      4599           0.96              0.00              0.00              0.0
      4600           0.00              0.00              0.65              0.0

              word_freq_our  word_freq_over  word_freq_remove  word_freq_internet  \
      0              0.32            0.00              0.00              0.00
      1              0.14            0.28              0.21              0.07
      2              1.23            0.19              0.19              0.12
      3              0.63            0.00              0.31              0.63
      4              0.63            0.00              0.31              0.63
      ...             ...             ...               ...               ...
      4596           0.00            0.31              0.00              0.00
```

```
4597          0.00           0.00             0.00                   0.00
4598          0.00           0.00             0.00                   0.00
4599          0.32           0.00             0.00                   0.00
4600          0.00           0.00             0.00                   0.00

      word_freq_order  word_freq_mail  …  char_freq_%3B  char_freq_%28  \
0                0.00            0.00  …          0.000          0.000
1                0.00            0.94  …          0.000          0.132
2                0.64            0.25  …          0.010          0.143
3                0.31            0.63  …          0.000          0.137
4                0.31            0.63  …          0.000          0.135
…                 …              …    …            …              …
4596             0.00            0.00  …          0.000          0.232
4597             0.00            0.00  …          0.000          0.000
4598             0.00            0.00  …          0.102          0.718
4599             0.00            0.00  …          0.000          0.057
4600             0.00            0.00  …          0.000          0.000

      char_freq_%5B  char_freq_%21  char_freq_%24  char_freq_%23  \
0               0.0          0.778          0.000          0.000
1               0.0          0.372          0.180          0.048
2               0.0          0.276          0.184          0.010
3               0.0          0.137          0.000          0.000
4               0.0          0.135          0.000          0.000
…                …            …              …              …
4596            0.0          0.000          0.000          0.000
4597            0.0          0.353          0.000          0.000
4598            0.0          0.000          0.000          0.000
4599            0.0          0.000          0.000          0.000
4600            0.0          0.125          0.000          0.000

      capital_run_length_average  capital_run_length_longest  \
0                          3.756                          61
1                          5.114                         101
2                          9.821                         485
3                          3.537                          40
4                          3.537                          40
…                            …                             …
4596                       1.142                           3
4597                       1.555                           4
4598                       1.404                           6
4599                       1.147                           5
4600                       1.250                           5

      capital_run_length_total  class
0                          278      1
1                         1028      1
```

```
2                            2259        1
3                             191        1
4                             191        1
...                           ...       ...
4596                           88        0
4597                           14        0
4598                          118        0
4599                           78        0
4600                           40        0

[4601 rows x 58 columns]>
```

[ ]: