

```
In [1]: import pandas as pd
import numpy as np
# read the text data
train = pd.read_csv('train_E6oV3lV.csv')
```

```
In [2]: # remove punctuations and space
train['tweet'] = train['tweet'].str.replace('[^\w\s]', '')
```

```
In [3]: # 10 most frequently occurring words in the text data
freq = pd.Series(' '.join(train['tweet']).split()).value_counts()[:10]
freq
```

```
Out[3]: user      17473
the         10139
to          9825
a           6416
i           5745
you         5407
and         4871
in          4638
for         4479
is          4176
dtype: int64
```

```
In [4]: # get the alphabetical word-count pairs
freq.sort_index()
```

```
Out[4]: a           6416
and          4871
for          4479
i           5745
in           4638
is           4176
the          10139
to           9825
user         17473
you          5407
dtype: int64
```

In []:

In []:

In []: