

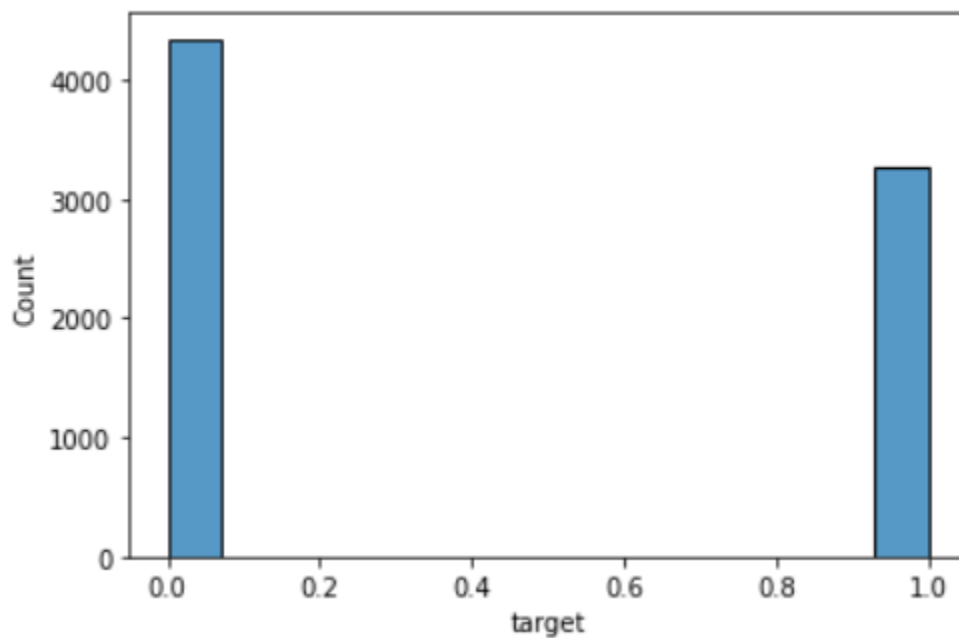
# LAB REPORT-5

**Mayank Raj (B19CSE053)**

B.Tech CSE

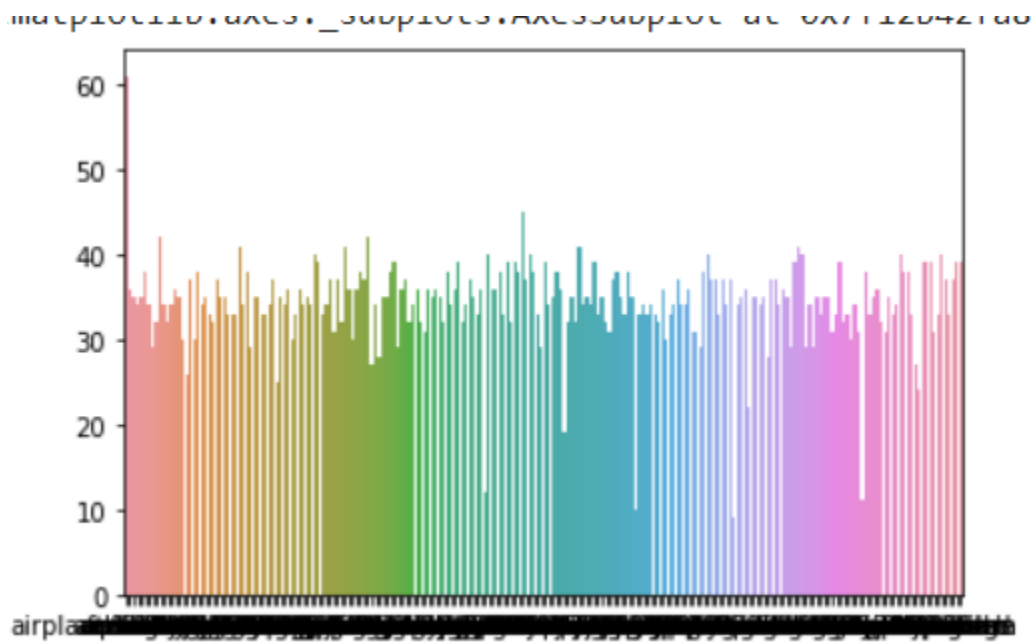
## 1)PREPROCESSING

1. Plotted count plot:



2. Printed unique words: 13566

3. Plotted bar plot of each keywords:

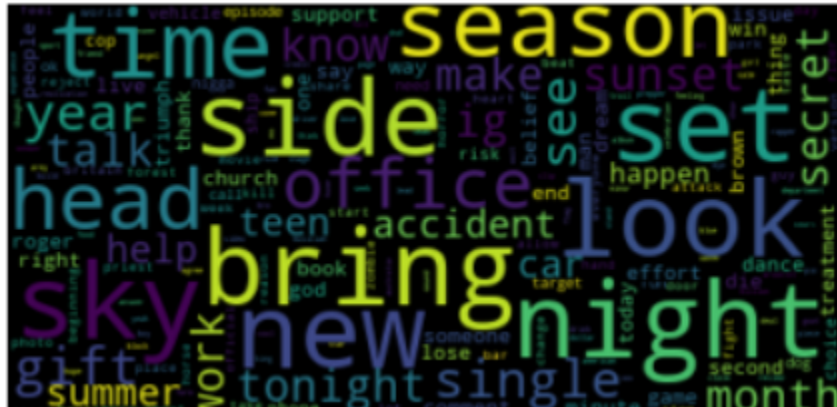


4. Correlation between length and target:



5. Refined the string using various functions described during the lab using the regex module.

## 6. Plotted word cloud of fake and real target



## 2)Term Document Matrix

Term document matrix was printed using countvectorizer.

[illegible]

in v 42622 column

### 3) Frequency of both the classes

```
frequency of class 1= 26011
frequency of class 2= 21570
```

### 4) Does the sum of the unique words in target 0 and 1 sum to the total number of unique words in the whole document? Why or why not? Explain in the report.

No, it may be present in both classes. So, the sum can be greater.

### 5) Probability of each class was calculated

Class 0: 0.568

Class 1: 0.432

### 6) Used bayes classification to predict the Class using posterior probability

### 7) Printed confusion matrix, precision, recall , f1 score for class 0 and class 1 respectively

```
Accuracy 0.8625
[[36  8]
 [ 3 33]]
```

➞ Precision: [0.92307692 0.80487805]  
Recall: [0.81818182 0.91666667]  
F1 score: [0.86746988 0.85714286]

