

Import required Packages

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

Read Datasets

```
In [2]: df=pd.read_csv("eeg-eye-state_csv.csv")
```

Print the datasets

```
In [3]: df.head(10)
```

Out[3]:

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11
0	4329.23	4009.23	4289.23	4148.21	4350.26	4586.15	4096.92	4641.03	4222.05	4238.46	4211.2
1	4324.62	4004.62	4293.85	4148.72	4342.05	4586.67	4097.44	4638.97	4210.77	4226.67	4207.6
2	4327.69	4006.67	4295.38	4156.41	4336.92	4583.59	4096.92	4630.26	4207.69	4222.05	4206.6
3	4328.72	4011.79	4296.41	4155.90	4343.59	4582.56	4097.44	4630.77	4217.44	4235.38	4210.7
4	4326.15	4011.79	4292.31	4151.28	4347.69	4586.67	4095.90	4627.69	4210.77	4244.10	4212.8
5	4321.03	4004.62	4284.10	4153.33	4345.64	4587.18	4093.33	4616.92	4202.56	4232.82	4209.7
6	4319.49	4001.03	4280.51	4151.79	4343.59	4584.62	4089.74	4615.90	4212.31	4226.67	4201.0
7	4325.64	4006.67	4278.46	4143.08	4344.10	4583.08	4087.18	4614.87	4205.64	4230.26	4195.9
8	4326.15	4010.77	4276.41	4139.49	4345.13	4584.10	4091.28	4608.21	4187.69	4229.74	4202.0
9	4326.15	4011.28	4276.92	4142.05	4344.10	4582.56	4092.82	4608.72	4194.36	4228.72	4212.8

Print the shape of Datasets

```
In [4]: df.shape
```

Out[4]: (14980, 15)

Here we can see total 14980 data instances are there in the dataset. It is also having 15 cloumns, in which one column is of target variable(Label/class) and remaining 14 are the features.

Print Unique Classes

```
In [5]: df.Class.unique()
```

```
Out[5]: array([1, 2], dtype=int64)
```

There are 2 different classes in this dataset i.e 1 and 2

Count instances each class having

```
In [6]: df.Class.value_counts()
```

```
Out[6]: 1    8257  
       2    6723  
       Name: Class, dtype: int64
```

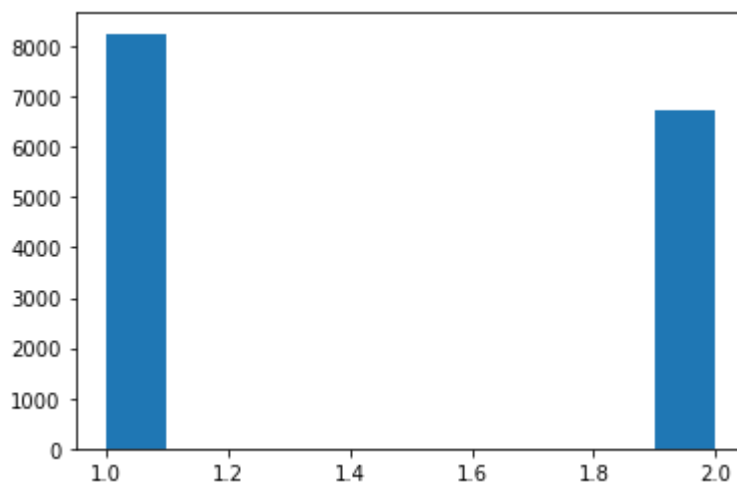
Print Percentage of each class

```
In [8]: round ((df.Class.value_counts()/df.shape[0])*100,2)
```

```
Out[8]: 1    55.12  
       2    44.88  
       Name: Class, dtype: float64
```

Plot Histogram

```
In [13]: plot =plt.hist(df['Class'])
```



This Dataset has 14980 instances of class 1 which is 55.12% of total instances. And class 2 having 6723 instances which is 44.88% of total.

Print datatypes of each column

```
In [23]: df.dtypes
```

```
Out[23]: V1         float64
V2         float64
V3         float64
V4         float64
V5         float64
V6         float64
V7         float64
V8         float64
V9         float64
V10        float64
V11        float64
V12        float64
V13        float64
V14        float64
Class      int64
dtype: object
```

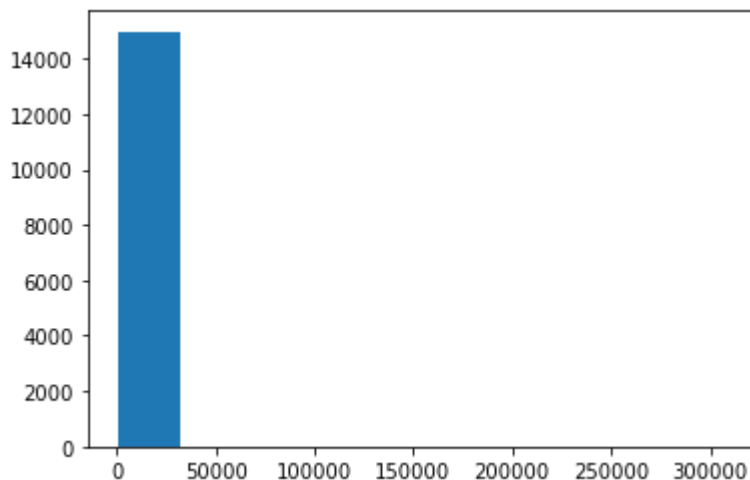
Here we can see class having datatype as int64(i.e int64 1 and 2) . Then V1,V2,V3,V4,V5,V6,V7,V8,V9,V10...etc have float64 datatype values.

Plot hist for different features

```
In [14]: print(df.V1.describe())

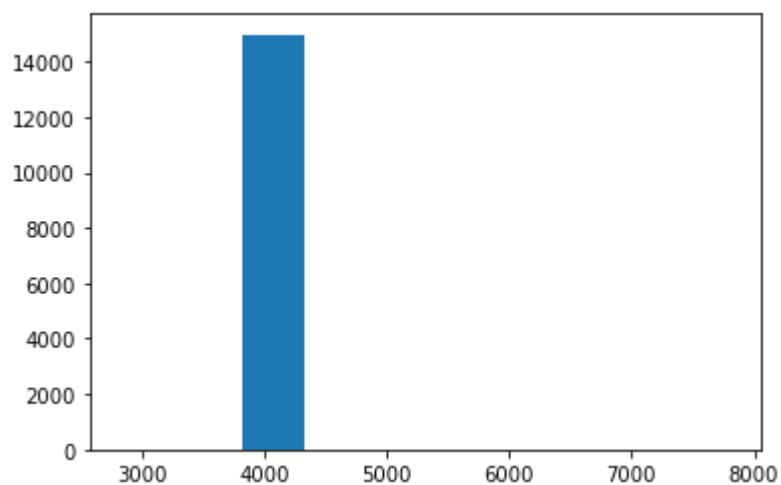
plot = plt.hist(df['V1'])
```

```
count    14980.000000
mean      4321.917777
std       2492.072174
min       1030.770000
25%       4280.510000
50%       4294.360000
75%       4311.790000
max      309231.000000
Name: V1, dtype: float64
```



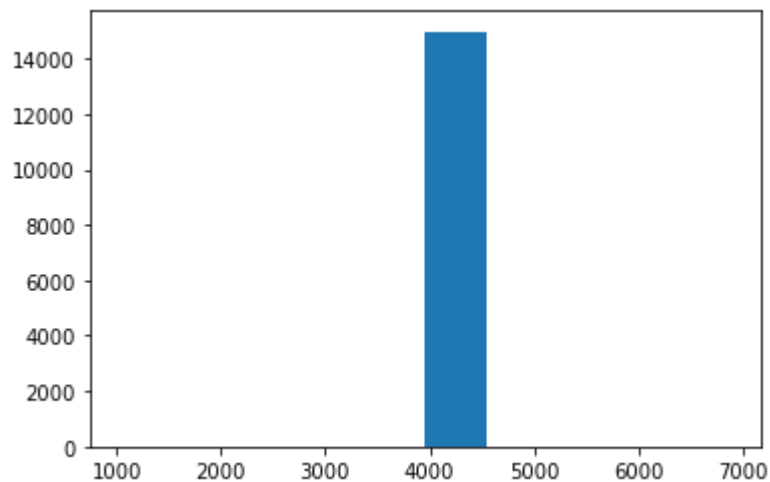
```
In [15]: print(df.V2.describe())  
  
plot = plt.hist(df['V2'])
```

```
count    14980.000000  
mean      4009.767694  
std        45.941672  
min       2830.770000  
25%       3990.770000  
50%       4005.640000  
75%       4023.080000  
max       7804.620000  
Name: V2, dtype: float64
```



```
In [16]: print(df.V3.describe())  
  
plot = plt.hist(df['V3'])
```

```
count    14980.000000  
mean      4264.022433  
std        44.428052  
min       1040.000000  
25%       4250.260000  
50%       4262.560000  
75%       4270.770000  
max       6880.510000  
Name: V3, dtype: float64
```



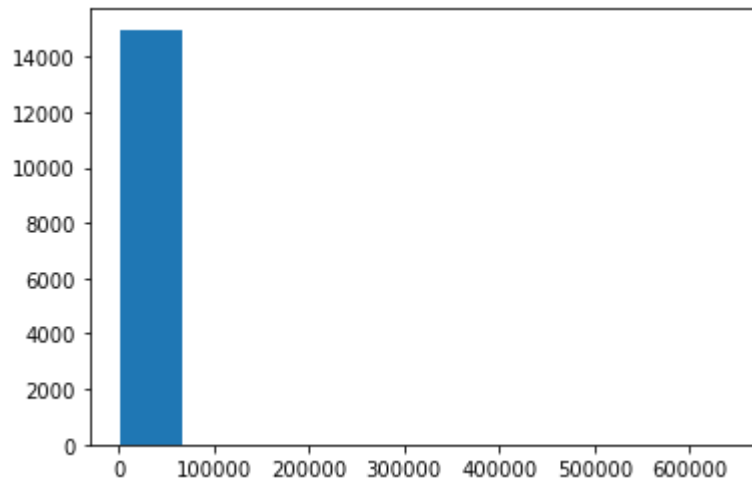
```
In [18]: df.V2.equals(df.V3)
```

```
Out[18]: False
```

Here V2 and V3 not having same values and it is not equally distributed over the range.

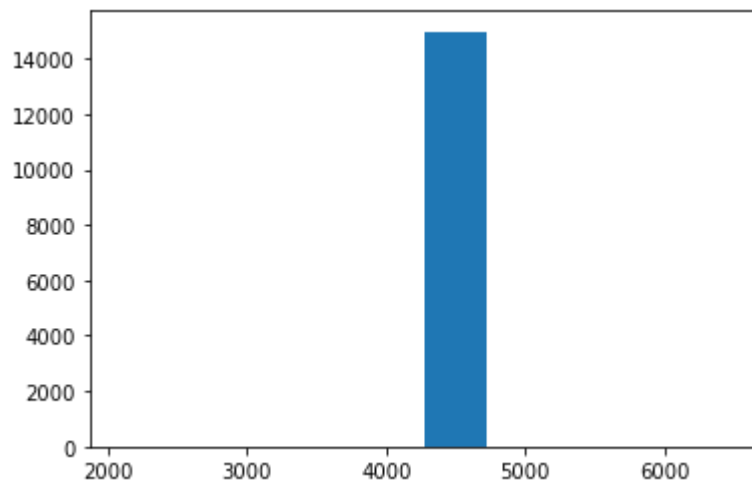
```
In [19]: print(df.V4.describe())  
  
plot = plt.hist(df['V4'])
```

```
count      14980.000000  
mean       4164.946326  
std        5216.404632  
min        2453.330000  
25%        4108.210000  
50%        4120.510000  
75%        4132.310000  
max        642564.000000  
Name: V4, dtype: float64
```



```
In [20]: print(df.V5.describe())  
  
plot = plt.hist(df['V5'])
```

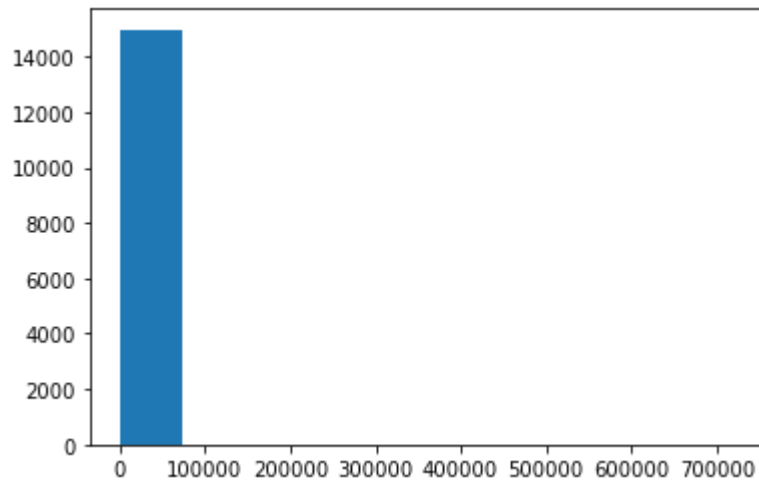
```
count    14980.000000  
mean      4341.741075  
std        34.738821  
min       2089.740000  
25%       4331.790000  
50%       4338.970000  
75%       4347.180000  
max       6474.360000  
Name: V5, dtype: float64
```



In V4 and V5, max values are different and also means are different. From plots of both we can conclude there distribution is also almost different.

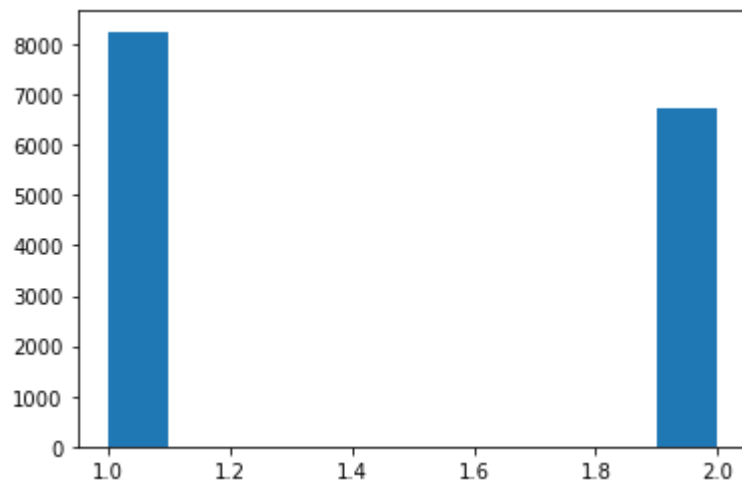
```
In [35]: print(df.V14.describe())  
  
plot = plt.hist(df['V14'])
```

```
count      14980.000000  
mean       4416.435832  
std        5891.285043  
min        1366.150000  
25%        4342.050000  
50%        4354.870000  
75%        4372.820000  
max        715897.000000  
Name: V14, dtype: float64
```




```
In [36]: print(df.Class.describe())  
  
plot = plt.hist(df['Class'])
```

```
count    14980.000000  
mean       1.448798  
std        0.497388  
min        1.000000  
25%        1.000000  
50%        1.000000  
75%        2.000000  
max        2.000000  
Name: Class, dtype: float64
```



Here values from 1.0 to 2.0 having frequency of atleast 6000. we can see from 0 to 1.1 the frequency of occurrence is high.