In [0]:

```
import pandas as pd
import numpy as np
```

In [0]:

```
list_d = [
'duration',
'protocol_type',
'service',
'flag',
'src_bytes',
'dst_bytes',
'land',
'wrong_fragment',
'urgent',
'hot',
'num_failed_logins',
'logged_in',
'num_compromised',
'root_shell',
'su_attempted',
'num_root',
'num_file_creations',
'num_shells',
'num_access_files',
'num_outbound_cmds',
'is_host_login',
'is_guest_login',
'count',
'srv_count',
'serror_rate',
'srv_serror_rate',
'rerror_rate',
'srv_rerror_rate',
'same_srv_rate',
'diff_srv_rate',
'srv_diff_host_rate',
'dst_host_count',
'dst_host_srv_count',
'dst_host_same_srv_rate',
'dst_host_diff_srv_rate',
'dst_host_same_src_port_rate',
'dst_host_srv_diff_host_rate',
'dst_host_serror_rate',
'dst_host_srv_serror_rate',
'dst_host_rerror_rate',
'dst_host_srv_rerror_rate',
'outcome'
]
list_d
```

Out[0]:

```
['duration',
 'protocol_type',
 'service',
 'flag',
 'src_bytes',
 'dst_bytes',
 'land',
 'wrong_fragment',
 'urgent',
 'hot',
 'num_failed_logins',
 'logged_in',
 'num_compromised',
 'root_shell',
 'su_attempted',
 'num_root',
 'num_file_creations',
 'num_shells',
 'num_access_files',
 'num_outbound_cmds',
 'is_host_login',
 'is_guest_login',
 'count',
 'srv_count',
 'serror_rate',
 'srv_serror_rate',
 'rerror_rate',
 'srv_rerror_rate',
 'same_srv_rate',
 'diff_srv_rate',
 'srv_diff_host_rate',
 'dst_host_count',
```

```
 'dst_host_srv_count',
 'dst_host_same_srv_rate',
 'dst_host_diff_srv_rate',
 'dst_host_same_src_port_rate',
 'dst_host_srv_diff_host_rate',
 'dst_host_serror_rate',
 'dst_host_srv_serror_rate',
 'dst_host_rerror_rate',
 'dst_host_srv_rerror_rate',
 'outcome']
```

In [0]:

```
df = pd.read_csv("./Downloads/testdata_10percent.csv.tar.gz")
df
```

Out[0]:

	testdata_10percent.csv	protocol_type	service	flag	src_bytes	dst_bytes	land	wrong_fragment	urgent	hot	...	dst_host_srv_count	dst_host_same_sr
0	14.0	udp	private	SF	105.0	146.0	0.0	0.0	0.0	0.0	...	241.0	0.95
1	0.0	udp	private	SF	105.0	146.0	0.0	0.0	0.0	0.0	...	241.0	0.95
2	0.0	udp	private	SF	105.0	146.0	0.0	0.0	0.0	0.0	...	241.0	0.95
3	0.0	udp	private	SF	105.0	0.0	0.0	0.0	0.0	0.0	...	241.0	0.95
4	0.0	udp	private	SF	105.0	0.0	0.0	0.0	0.0	0.0	...	241.0	0.95
5	0.0	udp	private	SF	105.0	145.0	0.0	0.0	0.0	0.0	...	241.0	0.95
6	0.0	udp	private	SF	105.0	146.0	0.0	0.0	0.0	0.0	...	241.0	0.95
7	0.0	udp	private	SF	105.0	146.0	0.0	0.0	0.0	0.0	...	241.0	0.95
8	0.0	udp	private	SF	105.0	146.0	0.0	0.0	0.0	0.0	...	241.0	0.95
9	0.0	udp	private	SF	105.0	146.0	0.0	0.0	0.0	0.0	...	241.0	0.95
10	5.0	udp	private	SF	105.0	146.0	0.0	0.0	0.0	0.0	...	241.0	0.95
11	0.0	udp	private	SF	105.0	146.0	0.0	0.0	0.0	0.0	...	241.0	0.95
12	0.0	udp	private	SF	105.0	0.0	0.0	0.0	0.0	0.0	...	241.0	0.95
13	0.0	udp	private	SF	105.0	0.0	0.0	0.0	0.0	0.0	...	241.0	0.95
14	14.0	udp	private	SF	105.0	146.0	0.0	0.0	0.0	0.0	...	241.0	0.95
15	0.0	udp	private	SF	105.0	0.0	0.0	0.0	0.0	0.0	...	242.0	0.95
16	5.0	udp	private	SF	105.0	146.0	0.0	0.0	0.0	0.0	...	243.0	0.95
17	0.0	icmp	urp_i	SF	182.0	0.0	0.0	0.0	0.0	0.0	...	10.0	0.04
18	0.0	icmp	urp_i	SF	182.0	0.0	0.0	0.0	0.0	0.0	...	11.0	0.04
19	0.0	icmp	urp_i	SF	182.0	0.0	0.0	0.0	0.0	0.0	...	12.0	0.05
20	0.0	icmp	urp_i	SF	182.0	0.0	0.0	0.0	0.0	0.0	...	13.0	0.05
21	0.0	icmp	urp_i	SF	182.0	0.0	0.0	0.0	0.0	0.0	...	14.0	0.05
22	0.0	icmp	urp_i	SF	182.0	0.0	0.0	0.0	0.0	0.0	...	15.0	0.06
23	0.0	udp	private	SF	105.0	146.0	0.0	0.0	0.0	0.0	...	238.0	0.93
24	0.0	udp	private	SF	105.0	146.0	0.0	0.0	0.0	0.0	...	238.0	0.93
25	0.0	udp	private	SF	105.0	146.0	0.0	0.0	0.0	0.0	...	238.0	0.93
26	0.0	udp	private	SF	105.0	146.0	0.0	0.0	0.0	0.0	...	238.0	0.93
27	0.0	udp	private	SF	105.0	146.0	0.0	0.0	0.0	0.0	...	238.0	0.93
28	0.0	icmp	eco_i	SF	30.0	0.0	0.0	0.0	0.0	0.0	...	3.0	0.01
29	0.0	udp	private	SF	105.0	146.0	0.0	0.0	0.0	0.0	...	237.0	0.93
...	...	...	...	...	...	...	...	...	...	...	...	...	...
399970	0.0	tcp	http	SF	289.0	1096.0	0.0	0.0	0.0	0.0	...	255.0	1.00
399971	0.0	tcp	http	SF	289.0	1862.0	0.0	0.0	0.0	0.0	...	255.0	1.00
399972	0.0	tcp	http	SF	203.0	242.0	0.0	0.0	0.0	0.0	...	255.0	1.00
399973	0.0	tcp	http	S1	196.0	0.0	0.0	0.0	0.0	0.0	...	255.0	1.00
399974	0.0	tcp	http	SF	143.0	17463.0	0.0	0.0	0.0	0.0	...	255.0	1.00
399975	0.0	tcp	http	SF	202.0	4017.0	0.0	0.0	0.0	0.0	...	255.0	1.00
399976	0.0	tcp	http	SF	0.0	234.0	0.0	0.0	0.0	0.0	...	255.0	1.00
399977	0.0	tcp	http	SF	203.0	1200.0	0.0	0.0	0.0	0.0	...	255.0	1.00
399978	0.0	tcp	http	SF	198.0	2169.0	0.0	0.0	0.0	0.0	...	255.0	1.00
399979	0.0	tcp	http	SF	197.0	466.0	0.0	0.0	0.0	0.0	...	255.0	1.00
399980	0.0	tcp	http	SF	203.0	1862.0	0.0	0.0	0.0	0.0	...	255.0	1.00
399981	0.0	tcp	http	SF	203.0	1096.0	0.0	0.0	0.0	0.0	...	255.0	1.00
399982	0.0	tcp	http	SF	284.0	2286.0	0.0	0.0	0.0	0.0	...	255.0	1.00
399983	0.0	tcp	http	SF	290.0	4017.0	0.0	0.0	0.0	0.0	...	255.0	1.00
399984	0.0	tcp	http	SF	291.0	234.0	0.0	0.0	0.0	0.0	...	255.0	1.00
399985	0.0	tcp	http	SF	291.0	242.0	0.0	0.0	0.0	0.0	...	255.0	1.00
399986	0.0	tcp	http	SF	231.0	17463.0	0.0	0.0	0.0	0.0	...	255.0	1.00
399987	0.0	tcp	http	SF	291.0	1200.0	0.0	0.0	0.0	0.0	...	255.0	1.00
399988	0.0	tcp	http	SF	285.0	466.0	0.0	0.0	0.0	0.0	...	255.0	1.00
399989	0.0	tcp	http	SF	286.0	2169.0	0.0	0.0	0.0	0.0	...	255.0	1.00
399990	0.0	tcp	http	SF	291.0	1096.0	0.0	0.0	0.0	0.0	...	255.0	1.00

	testdata_10percent.csv	protocol_type	service	flag	src_bytes	dst_bytes	land	wrong_fragment	urgent	hot	...	dst_host_srv_count	dst_host_same_sr
399991	0.0	tcp	http	SF	291.0	1862.0	0.0	0.0	0.0	0.0	...	255.0	1.00
399992	0.0	tcp	http	SF	159.0	15808.0	0.0	0.0	0.0	0.0	...	255.0	1.00
399993	0.0	tcp	http	SF	219.0	244.0	0.0	0.0	0.0	0.0	...	255.0	1.00
399994	0.0	tcp	http	SF	212.0	2288.0	0.0	0.0	0.0	0.0	...	255.0	1.00
399995	0.0	tcp	http	SF	219.0	236.0	0.0	0.0	0.0	0.0	...	255.0	1.00
399996	0.0	tcp	http	SF	218.0	3610.0	0.0	0.0	0.0	0.0	...	255.0	1.00
399997	0.0	tcp	http	SF	219.0	1234.0	0.0	0.0	0.0	0.0	...	255.0	1.00
399998	0.0	tcp	http	SF	219.0	1098.0	0.0	0.0	0.0	0.0	...	255.0	1.00
399999	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN

400000 rows Ã 42 columns

In [0]:

```
newdf = df.sample(frac=0.1,replace=False)
newdf
```

Out[0]:

	testdata_10percent.csv	protocol_type	service	flag	src_bytes	dst_bytes	land	wrong_fragment	urgent	hot	...	dst_host_srv_count	dst_host_same
206058	0.0	tcp	exec	RSTO	0.0	0.0	0.0	0.0	0.0	0.0	...	15.0	0.06
351157	0.0	icmp	ecr_i	SF	1032.0	0.0	0.0	0.0	0.0	0.0	...	255.0	1.00
109951	0.0	tcp	private	REJ	0.0	0.0	0.0	0.0	0.0	0.0	...	15.0	0.06
147199	0.0	tcp	private	REJ	0.0	0.0	0.0	0.0	0.0	0.0	...	20.0	0.08
38080	6804.0	udp	other	SF	147.0	105.0	0.0	0.0	0.0	0.0	...	2.0	0.01
234394	0.0	tcp	private	REJ	0.0	0.0	0.0	0.0	0.0	0.0	...	10.0	0.04
59714	0.0	tcp	smtp	SF	1066.0	385.0	0.0	0.0	0.0	0.0	...	195.0	0.81
234084	0.0	tcp	private	REJ	0.0	0.0	0.0	0.0	0.0	0.0	...	5.0	0.02
227598	0.0	tcp	private	REJ	0.0	0.0	0.0	0.0	0.0	0.0	...	4.0	0.02
135738	0.0	tcp	private	REJ	0.0	0.0	0.0	0.0	0.0	0.0	...	20.0	0.08
314367	0.0	tcp	private	S0	0.0	0.0	0.0	0.0	0.0	0.0	...	20.0	0.08
117776	0.0	tcp	private	REJ	0.0	0.0	0.0	0.0	0.0	0.0	...	11.0	0.04
277444	0.0	udp	domain_u	SF	37.0	90.0	0.0	0.0	0.0	0.0	...	44.0	0.42
383163	0.0	tcp	http	SF	222.0	579.0	0.0	0.0	0.0	0.0	...	255.0	1.00
235708	0.0	tcp	private	REJ	0.0	0.0	0.0	0.0	0.0	0.0	...	19.0	0.07
84029	0.0	tcp	private	REJ	0.0	0.0	0.0	0.0	0.0	0.0	...	8.0	0.03
165750	0.0	tcp	uucp_path	REJ	0.0	0.0	0.0	0.0	0.0	0.0	...	12.0	0.05
282080	9423.0	udp	other	SF	147.0	105.0	0.0	0.0	0.0	0.0	...	1.0	0.00
188489	0.0	tcp	private	REJ	0.0	0.0	0.0	0.0	0.0	0.0	...	5.0	0.02
63506	0.0	tcp	netstat	REJ	0.0	0.0	0.0	0.0	0.0	0.0	...	9.0	0.04
375492	0.0	udp	other	SF	147.0	0.0	0.0	0.0	0.0	0.0	...	2.0	0.01
189082	0.0	tcp	private	REJ	0.0	0.0	0.0	0.0	0.0	0.0	...	16.0	0.06
16417	0.0	tcp	http	REJ	0.0	0.0	0.0	0.0	0.0	0.0	...	255.0	1.00
331521	0.0	tcp	http	SF	258.0	6418.0	0.0	0.0	0.0	0.0	...	253.0	0.99
6647	22260.0	udp	other	SF	146.0	105.0	0.0	0.0	0.0	0.0	...	2.0	0.01
117229	0.0	tcp	private	REJ	0.0	0.0	0.0	0.0	0.0	0.0	...	13.0	0.05
204729	0.0	tcp	private	REJ	0.0	0.0	0.0	0.0	0.0	0.0	...	19.0	0.07
244176	0.0	tcp	private	REJ	0.0	0.0	0.0	0.0	0.0	0.0	...	15.0	0.06
376241	0.0	tcp	smtp	SF	526.0	332.0	0.0	0.0	0.0	0.0	...	56.0	0.76
179929	0.0	tcp	private	REJ	0.0	0.0	0.0	0.0	0.0	0.0	...	7.0	0.03
...	...	...	...	...	...	...	...	...	...	...	...	...	...
145542	0.0	tcp	private	REJ	0.0	0.0	0.0	0.0	0.0	0.0	...	19.0	0.07
310049	0.0	icmp	eco_i	SF	18.0	0.0	0.0	0.0	0.0	0.0	...	180.0	1.00
142243	0.0	tcp	private	REJ	0.0	0.0	0.0	0.0	0.0	0.0	...	6.0	0.02
202350	0.0	tcp	private	REJ	0.0	0.0	0.0	0.0	0.0	0.0	...	18.0	0.07
250318	0.0	tcp	private	REJ	0.0	0.0	0.0	0.0	0.0	0.0	...	2.0	0.01
348350	0.0	icmp	ecr_i	SF	1032.0	0.0	0.0	0.0	0.0	0.0	...	255.0	1.00
247282	0.0	tcp	private	REJ	0.0	0.0	0.0	0.0	0.0	0.0	...	8.0	0.03
91734	0.0	tcp	private	REJ	0.0	0.0	0.0	0.0	0.0	0.0	...	9.0	0.04
173309	0.0	tcp	private	REJ	0.0	0.0	0.0	0.0	0.0	0.0	...	4.0	0.02
43942	2450.0	udp	other	SF	147.0	105.0	0.0	0.0	0.0	0.0	...	1.0	0.00
345758	0.0	icmp	ecr_i	SF	1032.0	0.0	0.0	0.0	0.0	0.0	...	255.0	1.00
274682	16.0	udp	private	SF	105.0	147.0	0.0	0.0	0.0	0.0	...	204.0	0.80
163318	0.0	tcp	private	REJ	0.0	0.0	0.0	0.0	0.0	0.0	...	3.0	0.01
243435	0.0	tcp	private	REJ	0.0	0.0	0.0	0.0	0.0	0.0	...	19.0	0.07
198666	0.0	tcp	private	REJ	0.0	0.0	0.0	0.0	0.0	0.0	...	3.0	0.01
57980	0.0	tcp	http	SF	318.0	2444.0	0.0	0.0	0.0	0.0	...	255.0	1.00
361214	0.0	icmp	ecr_i	SF	1032.0	0.0	0.0	0.0	0.0	0.0	...	255.0	1.00

	testdata_10percent.csv	protocol_type	service	flag	src_bytes	dst_bytes	land	wrong_fragment	urgent	hot	...	dst_host_srv_count	dst_host_same
84767	0.0	tcp	private	REJ	0.0	0.0	0.0	0.0	0.0	0.0	...	18.0	0.07
285213	1988.0	udp	other	SF	147.0	105.0	0.0	0.0	0.0	0.0	...	2.0	0.01
244746	0.0	tcp	private	REJ	0.0	0.0	0.0	0.0	0.0	0.0	...	3.0	0.01
177986	0.0	tcp	private	REJ	0.0	0.0	0.0	0.0	0.0	0.0	...	15.0	0.06
61355	0.0	tcp	private	REJ	0.0	0.0	0.0	0.0	0.0	0.0	...	7.0	0.03
296264	0.0	tcp	http	SF	206.0	278.0	0.0	0.0	0.0	0.0	...	255.0	1.00
26321	0.0	udp	private	SF	105.0	147.0	0.0	0.0	0.0	0.0	...	245.0	0.96
151339	0.0	tcp	private	REJ	0.0	0.0	0.0	0.0	0.0	0.0	...	1.0	0.00
54218	0.0	tcp	http	SF	220.0	3313.0	0.0	0.0	0.0	0.0	...	191.0	1.00
47995	0.0	tcp	http	SF	205.0	2750.0	0.0	0.0	0.0	0.0	...	255.0	1.00
97618	0.0	tcp	private	REJ	0.0	0.0	0.0	0.0	0.0	0.0	...	18.0	0.07
324809	0.0	tcp	http	SF	219.0	502.0	0.0	0.0	0.0	0.0	...	255.0	1.00
118703	0.0	tcp	private	REJ	0.0	0.0	0.0	0.0	0.0	0.0	...	9.0	0.04

40000 rows Ã 42 columns

In [0]:

```
X_train = newdf.select_dtypes(exclude='object')
X_train
y_train = newdf.label
y_train.value_counts()
```

Out[0]:

```
neptune.      21136
normal.       14995
smurf.         3526
portsweep.      182
ipsweep.        137
back.            12
teardrop.        11
pod.              1
Name: label, dtype: int64
```

In [0]:

```
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

In [0]:

```
def correlation_heatmap(train):
    correlations = train.corr()

    fig, ax = plt.subplots(figsize=(10,10))
    sns.heatmap(correlations, vmax=1.0, center=0, fmt='.2f',
                square=True, linewidths=.5, annot=True, cbar_kws={"shrink": .70})
    plt.show()

correlation_heatmap(X_train)
```

In [0]:

```
#Using Pearson Correlation

from sklearn.preprocessing import LabelEncoder
labelencoder = LabelEncoder()
newdf['label']= labelencoder.fit_transform(newdf['label'])

plt.figure(figsize=(12,10))
cor = df.corr()
sns.heatmap(cor)
plt.show()
```

In [0]:

```
from sklearn.decomposition import PCA
pca = PCA(n_components=2)
principalComponents = pca.fit_transform(X_train)
principalDf = pd.DataFrame(data = principalComponents
             , columns = ['principal component 1', 'principal component 2'])
principalDf
```

Out[0]:

	principal component 1	principal component 2
0	-575.427355	-912.145516
1	456.565784	-913.551117

	principal component 1	principal component 2
2	-575.427606	-912.150883
3	-575.434588	-912.299788
4	-428.834836	-814.755207
5	-575.434042	-912.288242
6	491.002121	-527.891170
7	-575.433727	-912.281574
8	-575.426956	-912.137109
9	-575.434881	-912.306050
10	-575.432825	-912.262203
11	-575.434272	-912.293154
12	-538.315790	-821.968377
13	-352.801885	-333.061780
14	-575.434692	-912.302033
15	-575.427627	-912.151389
16	-575.428508	-912.170142
17	-429.036083	-817.663398
18	-575.432637	-912.258315
19	-575.427187	-912.141988
20	-428.422822	-912.201625
21	-575.432659	-912.258671
22	-575.413914	-911.795019
23	-310.649647	5505.855033
24	-431.022590	-831.915388
25	-575.433266	-912.271667
26	-575.427942	-912.158007
27	-575.428487	-912.169669
28	-49.056418	-580.444301
29	-575.427963	-912.159291
...	...	...
39970	-575.434692	-912.302033
39971	-557.402684	-911.787901
39972	-575.432784	-912.261438
39973	-575.434671	-912.301594
39974	-575.426914	-912.136231
39975	456.565784	-913.551117
39976	-575.427417	-912.146916
39977	-575.434146	-912.290487
39978	-575.427752	-912.154106
39979	-428.500273	-809.922420
39980	456.565784	-913.551117
39981	-470.268788	-765.026025
39982	-575.432050	-912.245808
39983	-575.428529	-912.170531
39984	-575.431253	-912.228811
39985	-254.843632	1531.793695
39986	456.565784	-913.551117
39987	-575.428256	-912.164724
39988	-428.464878	-809.411767
39989	-575.428234	-912.164402
39990	-575.427690	-912.152672
39991	-575.427061	-912.139320
39992	-369.129445	-634.084524
39993	-470.267510	-764.976528
39994	-575.426934	-912.136687
39995	-351.918946	2400.896985
39996	-367.500152	1838.058577
39997	-575.432826	-912.262233
39998	-355.873508	-409.954736
39999	-575.431547	-912.235023

40000 rows Ã 2 columns

In [0]:

```
df2 = pd.read_csv("./Downloads/testdata_10percent.csv (1).tar.gz",header=None,)
df2.columns = list_d
```

```
df2
```

```
/home/spit/anaconda3/lib/python3.7/site-packages/IPython/core/interactiveshell.py:3057: DtypeWarning: Columns (0,4,5,6,7,8,9,16
  interactivity=interactivity, compiler=compiler, result=result)
```

Out[0]:

	duration	protocol_type	service	flag	src_bytes	dst_bytes	land	wrong_fragment	urgent	hot	...	dst_host_srv_count	dst_host_same_srv
0	testdata_10percent.csv	protocol_type	service	flag	src_bytes	dst_bytes	land	wrong_fragment	urgent	hot	...	dst_host_srv_count	dst_host_same_srv_
1	14	udp	private	SF	105	146	0	0	0	0	...	241	0.95
2	0	udp	private	SF	105	146	0	0	0	0	...	241	0.95
3	0	udp	private	SF	105	146	0	0	0	0	...	241	0.95
4	0	udp	private	SF	105	0	0	0	0	0	...	241	0.95
5	0	udp	private	SF	105	0	0	0	0	0	...	241	0.95
6	0	udp	private	SF	105	145	0	0	0	0	...	241	0.95
7	0	udp	private	SF	105	146	0	0	0	0	...	241	0.95
8	0	udp	private	SF	105	146	0	0	0	0	...	241	0.95
9	0	udp	private	SF	105	146	0	0	0	0	...	241	0.95
10	0	udp	private	SF	105	146	0	0	0	0	...	241	0.95
11	5	udp	private	SF	105	146	0	0	0	0	...	241	0.95
12	0	udp	private	SF	105	146	0	0	0	0	...	241	0.95
13	0	udp	private	SF	105	0	0	0	0	0	...	241	0.95
14	0	udp	private	SF	105	0	0	0	0	0	...	241	0.95
15	14	udp	private	SF	105	146	0	0	0	0	...	241	0.95
16	0	udp	private	SF	105	0	0	0	0	0	...	242	0.95
17	5	udp	private	SF	105	146	0	0	0	0	...	243	0.95
18	0	icmp	urp_i	SF	182	0	0	0	0	0	...	10	0.04
19	0	icmp	urp_i	SF	182	0	0	0	0	0	...	11	0.04
20	0	icmp	urp_i	SF	182	0	0	0	0	0	...	12	0.05
21	0	icmp	urp_i	SF	182	0	0	0	0	0	...	13	0.05
22	0	icmp	urp_i	SF	182	0	0	0	0	0	...	14	0.05
23	0	icmp	urp_i	SF	182	0	0	0	0	0	...	15	0.06
24	0	udp	private	SF	105	146	0	0	0	0	...	238	0.93
25	0	udp	private	SF	105	146	0	0	0	0	...	238	0.93
26	0	udp	private	SF	105	146	0	0	0	0	...	238	0.93
27	0	udp	private	SF	105	146	0	0	0	0	...	238	0.93
28	0	udp	private	SF	105	146	0	0	0	0	...	238	0.93
29	0	icmp	eco_i	SF	30	0	0	0	0	0	...	3	0.01
...	...	...	...	...	...	...	...	...	...	...	...	...	...
399971	0	tcp	http	SF	289	1096	0	0	0	0	...	255	1
399972	0	tcp	http	SF	289	1862	0	0	0	0	...	255	1
399973	0	tcp	http	SF	203	242	0	0	0	0	...	255	1
399974	0	tcp	http	S1	196	0	0	0	0	0	...	255	1
399975	0	tcp	http	SF	143	17463	0	0	0	0	...	255	1
399976	0	tcp	http	SF	202	4017	0	0	0	0	...	255	1
399977	0	tcp	http	SF	0	234	0	0	0	0	...	255	1
399978	0	tcp	http	SF	203	1200	0	0	0	0	...	255	1
399979	0	tcp	http	SF	198	2169	0	0	0	0	...	255	1
399980	0	tcp	http	SF	197	466	0	0	0	0	...	255	1
399981	0	tcp	http	SF	203	1862	0	0	0	0	...	255	1
399982	0	tcp	http	SF	203	1096	0	0	0	0	...	255	1
399983	0	tcp	http	SF	284	2286	0	0	0	0	...	255	1
399984	0	tcp	http	SF	290	4017	0	0	0	0	...	255	1
399985	0	tcp	http	SF	291	234	0	0	0	0	...	255	1
399986	0	tcp	http	SF	291	242	0	0	0	0	...	255	1
399987	0	tcp	http	SF	231	17463	0	0	0	0	...	255	1
399988	0	tcp	http	SF	291	1200	0	0	0	0	...	255	1
399989	0	tcp	http	SF	285	466	0	0	0	0	...	255	1
399990	0	tcp	http	SF	286	2169	0	0	0	0	...	255	1
399991	0	tcp	http	SF	291	1096	0	0	0	0	...	255	1
399992	0	tcp	http	SF	291	1862	0	0	0	0	...	255	1
399993	0	tcp	http	SF	159	15808	0	0	0	0	...	255	1
399994	0	tcp	http	SF	219	244	0	0	0	0	...	255	1
399995	0	tcp	http	SF	212	2288	0	0	0	0	...	255	1
399996	0	tcp	http	SF	219	236	0	0	0	0	...	255	1
399997	0	tcp	http	SF	218	3610	0	0	0	0	...	255	1

	duration	protocol_type	service	flag	src_bytes	dst_bytes	land	wrong_fragment	urgent	hot	...	dst_host_srv_count	dst_host_same_srv
399998	0	tcp	http	SF	219	1234	0	0	0	0	...	255	1
399999	0	tcp	http	SF	219	1098	0	0	0	0	...	255	1
400000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN

400001 rows Ã 42 columns

In [0]:

```
from sklearn.preprocessing import LabelEncoder
labelencoder = LabelEncoder()
df2['outcome']= labelencoder.fit_transform(df2['outcome'])
X_test = df2.select_dtypes(exclude='object')
X_test
y_test = df2.outcome
y_test.value_counts()
```

```
---------------------------------------------------------------------------
TypeError                                 Traceback (most recent call last)
~/anaconda3/lib/python3.7/site-packages/sklearn/preprocessing/label.py in _encode(values, uniques, encode)
    104         try:
--> 105             res = _encode_python(values, uniques, encode)
    106         except TypeError:

~/anaconda3/lib/python3.7/site-packages/sklearn/preprocessing/label.py in _encode_python(values, uniques, encode)
     58     if uniques is None:
---> 59         uniques = sorted(set(values))
     60         uniques = np.array(uniques, dtype=values.dtype)

TypeError: '<' not supported between instances of 'float' and 'str'

During handling of the above exception, another exception occurred:

TypeError                                 Traceback (most recent call last)
<ipython-input-21-3cee50609a2c> in <module>
      1 from sklearn.preprocessing import LabelEncoder
      2 labelencoder = LabelEncoder()
----> 3 df2['outcome']= labelencoder.fit_transform(df2['outcome'])
      4 X_test = df2.select_dtypes(exclude='object')
      5 X_test

~/anaconda3/lib/python3.7/site-packages/sklearn/preprocessing/label.py in fit_transform(self, y)
    234         """
    235         y = column_or_1d(y, warn=True)
--> 236         self.classes_, y = _encode(y, encode=True)
    237         return y
    238

~/anaconda3/lib/python3.7/site-packages/sklearn/preprocessing/label.py in _encode(values, uniques, encode)
    105             res = _encode_python(values, uniques, encode)
    106         except TypeError:
--> 107             raise TypeError("argument must be a string or number")
    108         return res
    109     else:

TypeError: argument must be a string or number
```

In [0]:

```
X_test
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-22-0c1cdefd54ea> in <module>
----> 1 X_test

NameError: name 'X_test' is not defined
```

In [0]:

```
print("Hi")
```

```
Hi
```

In [0]:

In [0]:

```
import time
t0 = time.time()
from sklearn.ensemble import RandomForestClassifier
```

```
clf = RandomForestClassifier()
clf.fit(X_train[bxx], y_train)

t1=time.time()
print(t1-t0)
```

```
0.14283084869384766
```

```
/home/spit/anaconda3/lib/python3.7/site-packages/sklearn/ensemble/forest.py:245: FutureWarning: The default value of n_estimatc
  "10 in version 0.20 to 100 in 0.22.", FutureWarning)
```

In [0]:

```
t0 = time.time()
y_pred = clf.predict(X_test[bxx])
t1=time.time()
print(t1-t0)
```

```
0.815310001373291
```

In [0]:

```
from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score
from sklearn.metrics import classification_report


results = confusion_matrix(y_test, y_pred)
print ('Confusion Matrix :')
print(results)
print ('Accuracy Score :',accuracy_score(y_test, y_pred) )
print ('Report : ')
print(classification_report(y_test, y_pred))
```

```
Confusion Matrix :
[[    52      0      3      0      0      0      0      0      0      0
      44      0      0      0      0      0      0      0      1      0]
 [     0      3      0      0      0      0      0      0      0      0
       6      0      0      0      0      0      0      0      0      0]
 [     0      0      0      0    572      0      0      0      1     40
     608      0     10      0      0      0      0      0     83      0]
 [     0      0      0      0      0      0      0      0      0      0
       3      0      0      0      0      0      0      0      0      0]
 [     0      9      0      0      0      0      0      0 210689      0
     200      0      0      0     11      0      0      1      0      0]
 [   248      0     45      0    300      0      9      0    177     66
  147078      5     51    487      7    245      4     70    414      2]
 [     0      0      0      0      0      0      0      0      0      0
       4      0     10      0      0      0      0      7      1      0]
 [     0      0      0      0      0      0      0      0      0      0
    1617      0      0    187      0      0      0      0      0      0]
 [     0      0      0      0      0      0      0      0      0      0
       1      0      0      0      0      0      0      0      0      0]
 [     0      0      0      0      0      0      0      0      1      0
     103      0     11      0      0  36324      0     90      0      0]
 [     0      0      0      0      0      0      0      0      1      0
      31      0      0      0      0      0      0     68      0      0]
 [     0      0      0      0      0      0      0      0      0      0
       0      0      0      0      0      0      0      0      0      0]
 [     0      0      0      0      0      0      0      0      0      0
       0      0      0      0      0      0      0      0      0      0]
 [     0      0      0      0      0      0      0      0      0      0
       0      0      0      0      0      0      0      0      0      0]
 [     0      0      0      0      0      0      0      0      0      0
       0      0      0      0      0      0      0      0      0      0]
 [     0      0      0      0      0      0      0      0      0      0
       0      0      0      0      0      0      0      0      0      0]
 [     0      0      0      0      0      0      0      0      0      0
       0      0      0      0      0      0      0      0      0      0]
 [     0      0      0      0      0      0      0      0      0      0
       0      0      0      0      0      0      0      0      0      0]
 [     0      0      0      0      0      0      0      0      0      0
       0      0      0      0      0      0      0      0      0      0]
 [     0      0      0      0      0      0      0      0      0      0
       0      0      0      0      0      0      0      0      0      0]]
Accuracy Score : 0.000215
Report :
              precision    recall  f1-score   support

           0       0.17      0.52      0.26       100
```

```
           1       0.25      0.33      0.29         9
           2       0.00      0.00      0.00      1314
           3       0.00      0.00      0.00         3
           4       0.00      0.00      0.00    210910
           5       0.00      0.00      0.00    149208
           6       0.00      0.00      0.00        22
           7       0.00      0.00      0.00      1804
           8       0.00      0.00      0.00         1
           9       0.00      0.00      0.00     36529
          10       0.00      0.31      0.00       100
          11       0.00      0.00      0.00         0
          12       0.00      0.00      0.00         0
          13       0.00      0.00      0.00         0
          14       0.00      0.00      0.00         0
          15       0.00      0.00      0.00         0
          16       0.00      0.00      0.00         0
          17       0.00      0.00      0.00         0
          18       0.00      0.00      0.00         0
          19       0.00      0.00      0.00         0

    accuracy                           0.00    400000
   macro avg       0.02      0.06      0.03    400000
weighted avg       0.00      0.00      0.00    400000
```

```
/home/spit/anaconda3/lib/python3.7/site-packages/sklearn/metrics/classification.py:1437: UndefinedMetricWarning: Precision and
  'precision', 'predicted', average, warn_for)
/home/spit/anaconda3/lib/python3.7/site-packages/sklearn/metrics/classification.py:1439: UndefinedMetricWarning: Recall and F-s
  'recall', 'true', average, warn_for)
```

In [0]:

```
from sklearn.externals import joblib

# Save the model as a pickle in a file
joblib.dump(clf, 'LDA.pkl')

# Load the model from the file
knn_from_joblib = joblib.load('LDA.pkl')

# Use the loaded model to make predictions
knn_from_joblib.predict(X_test)
```

Out[0]:

```
array(['normal.', 'normal.', 'normal.', ..., 'normal.', 'normal.',
       'normal.'], dtype='<U16')
```

In [0]: