In [2]:
```python
#TASK1
#Name: YEO ZHENG XU ISAAC
```

In [3]:
```python
import pandas as  pd
from time import time
col_names= ["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","error_rate"
            ,"srv_error_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_error_rate","dst_host_srv_error_rate","dst_host_rerror_rate","dst_host_srv_rerror_rate","label"]
reviews = pd.read_csv("C:/Users/Isaac Yeo/Downloads/kddcup.data_10_percent",header= None,names = col_names)
```

In [5]:
```python
reviews
```

Out[5]:

	duration	protocol_type	service	flag	src_bytes	dst_bytes	land	wrong_fragment	urgent	h
0	0	tcp	http	SF	181	5450	0	0	0	
	0	tcp	http	SF	239	486	0	0	0	
	0	tcp	http	SF	235	1337	0	0	0	
	0	tcp	http	SF	219	1337	0	0	0	
	0	tcp	http	SF	217	2032	0	0	0	
...	...	...	...	...	...	...	...	...	...	...
	0	tcp	http	SF	310	1881	0	0	0	
	0	tcp	http	SF	282	2286	0	0	0	
	0	tcp	http	SF	203	1200	0	0	0	
	0	tcp	http	SF	291	1200	0	0	0	
	0	tcp	http	SF	219	1234	0	0	0	

494021 rows × 41 columns

```
In [9]: reviews[["srv_error_rate","dst_host_srv_error_rate"]].corr()
```

Out[9]:

	srv_error_rate	dst_host_srv_error_rate
srv_error_rate	1.000000	0.999304
dst_host_srv_error_rate	0.999304	1.000000

```
In [8]: reviews.mean()
```

```
Out[8]: flag                              3025.610296
        src_bytes                          868.532425
        dst_bytes                            0.000045
        land                                 0.006433
        wrong_fragment                       0.000014
        urgent                               0.034519
        hot                                  0.000152
        num_failed_logins                    0.148247
        logged_in                            0.010212
        num_compromisedroot_shell            0.000111
        su_attempted                         0.000036
        num_root                             0.011352
        num_file_creations                   0.001083
        num_shells                           0.000109
        num_access_files                     0.001008
        num_outbound_cmds                    0.000000
        is_host_login                        0.000000
        is_guest_login                       0.001387
        count                              332.285690
        srv_count                          292.906557
        serror_rate                          0.176687
        srv_serror_rate                      0.176609
        rerror_rate                          0.057433
        srv_rerror_rate                      0.057719
        same_srv_rate                        0.791547
        diff_srv_rate                        0.020982
        srv_diff_host_rate                   0.028997
        dst_host_count                     232.470778
        dst_host_srv_count                 188.665670
        dst_host_same_srv_rate               0.753780
        dst_host_diff_srv_rate               0.030906
        dst_host_same_src_port_rate          0.601935
        dst_host_srv_diff_host_rate          0.006684
        dst_host_serror_rate                 0.176754
        dst_host_srv_serror_rate             0.176443
        dst_host_rerror_rate                 0.058118
        dst_host_srv_rerror_rate             0.057412
        dtype: float64
```

```
In [9]: reviews.mean(axis=1)
```

```
Out[9]: 0      153.192162
        0       21.136486
        0       44.568378
        0       44.568378
        0       63.838378
                  ...
        0       68.769459
        0       76.871081
        0       45.990000
        0       48.467838
        0       47.523514
        Length: 494021, dtype: float64
```

In [10]: `reviews.min()`

Out[10]:
```
duration                            icmp
protocol_type                        IRC
service                              OTH
flag                                   0
src_bytes                              0
dst_bytes                              0
land                                   0
wrong_fragment                         0
urgent                                 0
hot                                    0
num_failed_logins                      0
logged_in                              0
num_compromisedroot_shell              0
su_attempted                           0
num_root                               0
num_file_creations                     0
num_shells                             0
num_access_files                       0
num_outbound_cmds                      0
is_host_login                          0
is_guest_login                         0
count                                  0
srv_count                              0
serror_rate                            0
srv_serror_rate                        0
rerror_rate                            0
srv_rerror_rate                        0
same_srv_rate                          0
diff_srv_rate                          0
srv_diff_host_rate                     0
dst_host_count                         0
dst_host_srv_count                     0
dst_host_same_srv_rate                 0
dst_host_diff_srv_rate                 0
dst_host_same_src_port_rate            0
dst_host_srv_diff_host_rate            0
dst_host_serror_rate                   0
dst_host_srv_serror_rate               0
dst_host_rerror_rate                   0
dst_host_srv_rerror_rate               0
label                              back.
dtype: object
```

In [11]: `reviews.max()`

Out[11]:
```
duration                          udp
protocol_type                   whois
service                            SH
flag                        693375640
src_bytes                     5155468
dst_bytes                           1
land                                3
wrong_fragment                      3
urgent                             30
hot                                 5
num_failed_logins                   1
logged_in                         884
num_compromisedroot_shell           1
su_attempted                        2
num_root                          993
num_file_creations                 28
num_shells                          2
num_access_files                    8
num_outbound_cmds                   0
is_host_login                       0
is_guest_login                      1
count                             511
srv_count                         511
serror_rate                         1
srv_serror_rate                     1
rerror_rate                         1
srv_rerror_rate                     1
same_srv_rate                       1
diff_srv_rate                       1
srv_diff_host_rate                  1
dst_host_count                    255
dst_host_srv_count                255
dst_host_same_srv_rate              1
dst_host_diff_srv_rate              1
dst_host_same_src_port_rate         1
dst_host_srv_diff_host_rate         1
dst_host_serror_rate                1
dst_host_srv_serror_rate            1
dst_host_rerror_rate                1
dst_host_srv_rerror_rate            1
label                       warezmaster.
dtype: object
```

In [12]: `reviews.std()`

Out[12]:
```
flag                          988218.101045
src_bytes                      33040.001252
dst_bytes                          0.006673
land                               0.134805
wrong_fragment                     0.005510
urgent                             0.782103
hot                                0.015520
num_failed_logins                  0.355345
logged_in                          1.798326
num_compromisedroot_shell          0.010551
su_attempted                       0.007793
num_root                           2.012718
num_file_creations                 0.096416
num_shells                         0.011020
num_access_files                   0.036482
num_outbound_cmds                  0.000000
is_host_login                      0.000000
is_guest_login                     0.037211
count                            213.147412
srv_count                        246.322817
serror_rate                        0.380717
srv_serror_rate                    0.381017
rerror_rate                        0.231623
srv_rerror_rate                    0.232147
same_srv_rate                      0.388189
diff_srv_rate                      0.082205
srv_diff_host_rate                 0.142397
dst_host_count                    64.745380
dst_host_srv_count               106.040437
dst_host_same_srv_rate             0.410781
dst_host_diff_srv_rate             0.109259
dst_host_same_src_port_rate        0.481309
dst_host_srv_diff_host_rate        0.042133
dst_host_serror_rate               0.380593
dst_host_srv_serror_rate           0.380919
dst_host_rerror_rate               0.230590
dst_host_srv_rerror_rate           0.230140
dtype: float64
```

In [13]: `reviews.median()`

Out[13]:
```
flag                           520.0
src_bytes                        0.0
dst_bytes                        0.0
land                             0.0
wrong_fragment                   0.0
urgent                           0.0
hot                              0.0
num_failed_logins                0.0
logged_in                        0.0
num_compromisedroot_shell        0.0
su_attempted                     0.0
num_root                         0.0
num_file_creations               0.0
num_shells                       0.0
num_access_files                 0.0
num_outbound_cmds                0.0
is_host_login                    0.0
is_guest_login                   0.0
count                          510.0
srv_count                      510.0
serror_rate                      0.0
srv_serror_rate                  0.0
rerror_rate                      0.0
srv_rerror_rate                  0.0
same_srv_rate                    1.0
diff_srv_rate                    0.0
srv_diff_host_rate               0.0
dst_host_count                 255.0
dst_host_srv_count             255.0
dst_host_same_srv_rate           1.0
dst_host_diff_srv_rate           0.0
dst_host_same_src_port_rate      1.0
dst_host_srv_diff_host_rate      0.0
dst_host_serror_rate             0.0
dst_host_srv_serror_rate         0.0
dst_host_rerror_rate             0.0
dst_host_srv_rerror_rate         0.0
dtype: float64
```

```
In [14]:  reviews.mean()
```

Out[14]: flag   3025.610296
          src_bytes   868.532425
          dst_bytes   0.000045
          land   0.006433
          wrong_fragment   0.000014
          urgent   0.034519
          hot   0.000152
          num_failed_logins   0.148247
          logged_in   0.010212
          num_compromisedroot_shell   0.000111
          su_attempted   0.000036
          num_root   0.011352
          num_file_creations   0.001083
          num_shells   0.000109
          num_access_files   0.001008
          num_outbound_cmds   0.000000
          is_host_login   0.000000
          is_guest_login   0.001387
          count   332.285690
          srv_count   292.906557
          serror_rate   0.176687
          srv_serror_rate   0.176609
          rerror_rate   0.057433
          srv_rerror_rate   0.057719
          same_srv_rate   0.791547
          diff_srv_rate   0.020982
          srv_diff_host_rate   0.028997
          dst_host_count   232.470778
          dst_host_srv_count   188.665670
          dst_host_same_srv_rate   0.753780
          dst_host_diff_srv_rate   0.030906
          dst_host_same_src_port_rate   0.601935
          dst_host_srv_diff_host_rate   0.006684
          dst_host_serror_rate   0.176754
          dst_host_srv_serror_rate   0.176443
          dst_host_rerror_rate   0.058118
          dst_host_srv_rerror_rate   0.057412
          dtype: float64
```

In [21]: ```
reviews.corr()
```

Out[21]:

	flag	src_bytes	dst_bytes	land	wrong_fragment	urgen
flag	1.000000	-0.000002	-0.000020	-0.000139	-0.000005	0.00448
src_bytes	-0.000002	1.000000	-0.000175	-0.001254	0.016288	0.00436
dst_bytes	-0.000020	-0.000175	1.000000	-0.000318	-0.000017	-0.00029
land	-0.000139	-0.001254	-0.000318	1.000000	-0.000123	-0.00210
wrong_fragment	-0.000005	0.016288	-0.000017	-0.000123	1.000000	0.00035
urgent	0.004483	0.004365	-0.000295	-0.002106	0.000356	1.00000
hot	-0.000027	0.049330	-0.000065	-0.000467	0.141996	0.00874
num_failed_logins	0.001701	0.047814	-0.002784	-0.019908	0.006164	0.10530
logged_in	0.000119	0.023298	-0.000038	-0.000271	0.014285	0.00734
num_compromisedroot_shell	-0.000022	0.031680	-0.000070	-0.000504	0.034790	0.02406
su_attempted	-0.000010	0.075656	-0.000031	-0.000223	-0.000012	-0.00020
num_root	-0.000010	0.020746	-0.000038	-0.000269	0.009476	0.00099
num_file_creations	0.000013	0.004958	-0.000075	-0.000536	0.015211	0.02524
num_shells	0.000005	0.000144	-0.000066	-0.000473	-0.000026	0.00637
num_access_files	-0.000052	0.008746	-0.000184	-0.001319	0.020068	0.00190
num_outbound_cmds	NaN	NaN	NaN	NaN	NaN	NaI
is_host_login	NaN	NaN	NaN	NaN	NaN	NaI
is_guest_login	-0.000082	0.001289	-0.000249	-0.001778	-0.000096	0.84357
count	-0.003098	-0.040373	-0.010260	-0.061934	-0.003997	-0.06845
srv_count	-0.002501	-0.030544	-0.007886	-0.047789	-0.003047	-0.05216
serror_rate	0.001558	-0.011908	0.013898	-0.013969	-0.001193	-0.02026
srv_serror_rate	0.001114	-0.011930	0.014422	-0.022119	-0.001192	-0.02021
rerror_rate	0.000591	-0.006166	-0.000777	-0.011529	-0.000638	-0.00830
srv_rerror_rate	0.001379	-0.005808	-0.001659	-0.011865	-0.000639	-0.00582
same_srv_rate	-0.001860	0.014002	0.002286	0.017416	0.001381	0.02269
diff_srv_rate	0.006207	-0.005702	0.002282	-0.007077	-0.000656	-0.00268
srv_diff_host_rate	-0.000015	0.008135	0.036985	0.000153	-0.000524	0.00197
dst_host_count	-0.001743	-0.048869	-0.023671	-0.005191	-0.007139	-0.02636
dst_host_srv_count	-0.003212	-0.005850	-0.011587	-0.058624	-0.004540	-0.03873
dst_host_same_srv_rate	-0.002052	0.007058	0.001984	-0.054903	-0.003279	-0.02911
dst_host_diff_srv_rate	0.000578	-0.005314	-0.000333	0.071857	0.010536	0.00131
dst_host_same_src_port_rate	-0.000724	-0.020143	0.003799	-0.031803	-0.002002	-0.05292
dst_host_srv_diff_host_rate	0.001186	0.008707	0.083320	0.012092	-0.000408	-0.00446
dst_host_serror_rate	-0.000718	-0.011334	0.012658	-0.019091	-0.001194	-0.01949
dst_host_srv_serror_rate	0.001122	-0.011235	0.007795	-0.022104	-0.001191	-0.02020

	flag	src_bytes	dst_bytes	land	wrong_fragment	urgen
dst_host_rerror_rate	-0.000393	-0.005000	-0.001511	0.029774	-0.000648	-0.00654
dst_host_srv_rerror_rate	0.001328	-0.005471	-0.001665	-0.011904	-0.000641	-0.00774

37 rows × 37 columns

In [16]:
```
reviews.describe()
```

Out[16]:

	flag	src_bytes	dst_bytes	land	wrong_fragment	urger
count	4.940210e+05	4.940210e+05	494021.000000	494021.000000	494021.000000	494021.00000
mean	3.025610e+03	8.685324e+02	0.000045	0.006433	0.000014	0.03451
std	9.882181e+05	3.304000e+04	0.006673	0.134805	0.005510	0.78210
min	0.000000e+00	0.000000e+00	0.000000	0.000000	0.000000	0.00000
25%	4.500000e+01	0.000000e+00	0.000000	0.000000	0.000000	0.00000
50%	5.200000e+02	0.000000e+00	0.000000	0.000000	0.000000	0.00000
75%	1.032000e+03	0.000000e+00	0.000000	0.000000	0.000000	0.00000
max	6.933756e+08	5.155468e+06	1.000000	3.000000	3.000000	30.00000

8 rows × 37 columns

In [12]:
```
df = pd.DataFrame(reviews.corr())
df.boxplot();
```
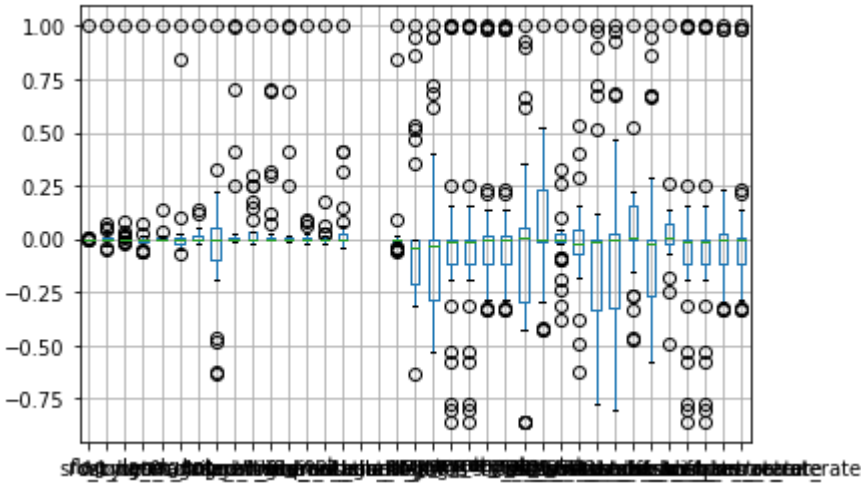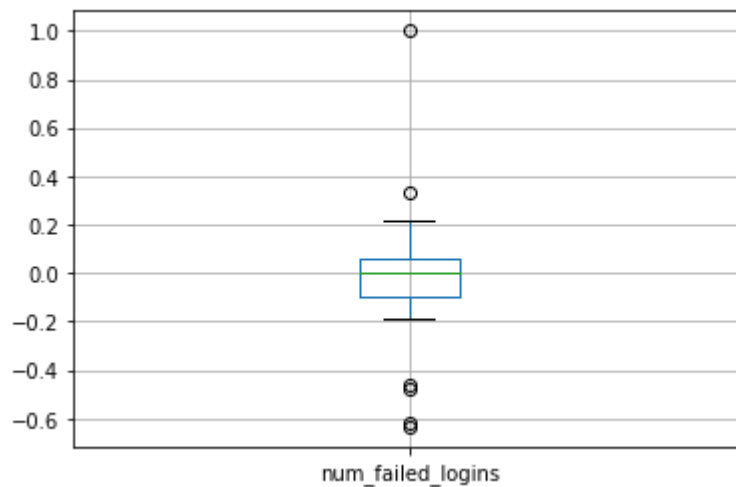
```
In [13]: df = pd.DataFrame(reviews.corr(),columns=["rerror_rate"])
         df.boxplot();
```
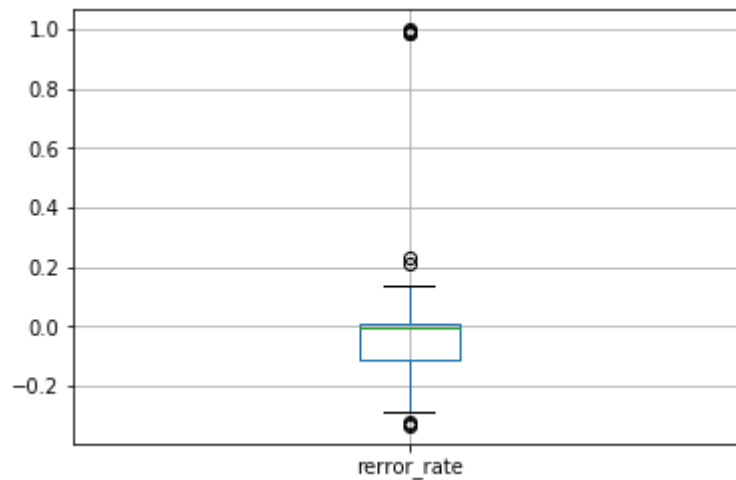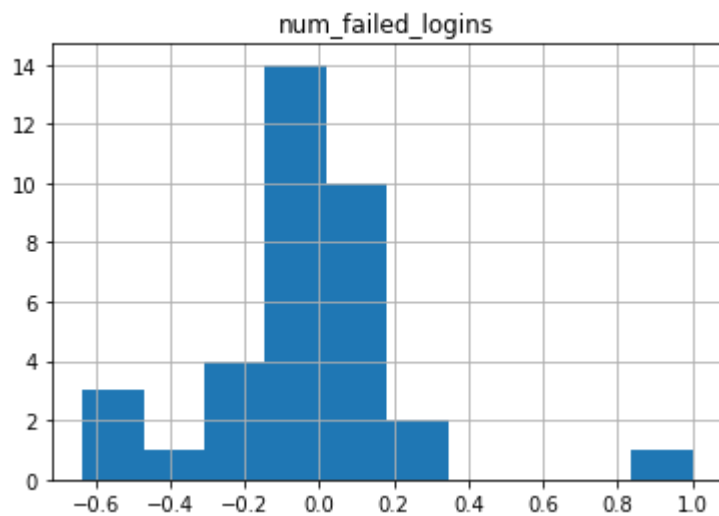


```
In [14]: df = pd.DataFrame(reviews.corr(), columns=["rerror_rate"])
         df.hist();
```
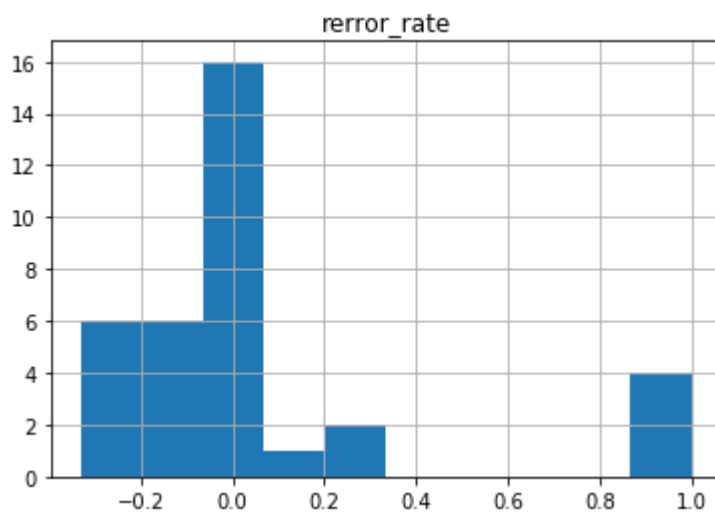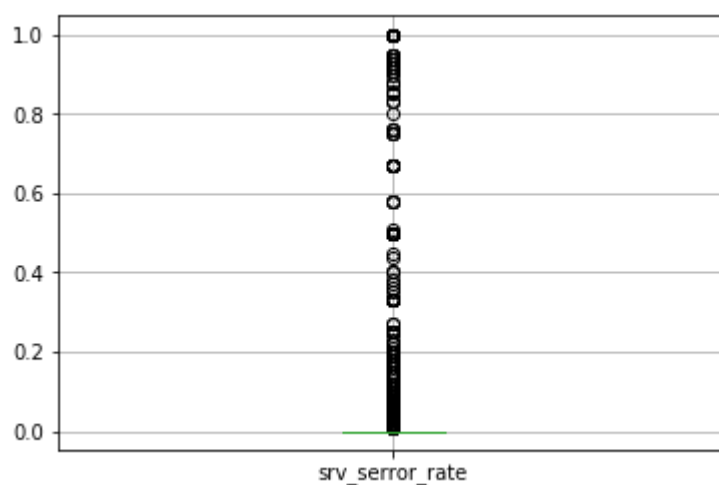


```
In [15]: df = pd.DataFrame(reviews.corr(),columns=["num_failed_logins"])
         df.boxplot();
```
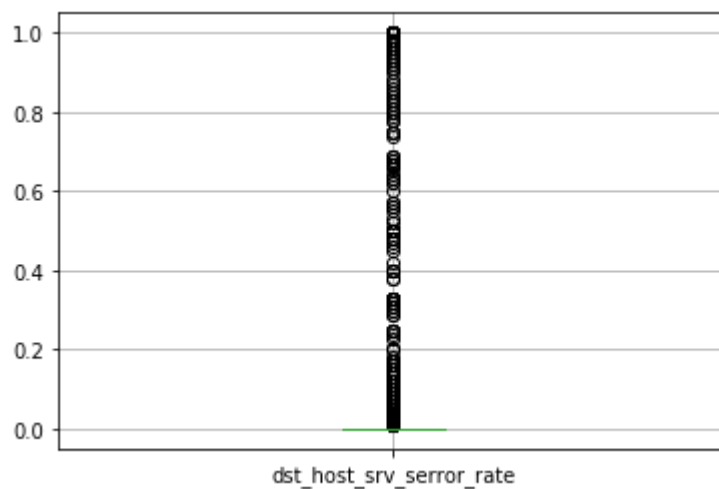
In [16]:
```python
df = pd.DataFrame(reviews.corr(),columns=["num_failed_logins"])
df.hist();
```


In [24]:
```python
df = pd.DataFrame(reviews, columns=['srv_serror_rate'])
df.boxplot();
```


In [26]:
```python
df = pd.DataFrame(reviews, columns=['dst_host_srv_serror_rate'])
df.boxplot();
```

In [25]: `reviews.boxplot()`

Out[25]: `<matplotlib.axes._subplots.AxesSubplot at 0x23b2ab0ba88>`