

Chapter 2

December 23, 2017

create a new data frame

```
In [1]: import numpy as np
import pandas as pd
```

create a new data frame of hosts & high vuln counts

```
In [3]: assets_df = pd.DataFrame( {
    "name" : ["danube","gander","ganges","mekong","orinoco" ],
    "os" : [ "W2K8","RHEL5","W2K8","RHEL5","RHEL5" ],
    "highvulns" : [ 1,0,2,0,0 ]
} )
```

take a look at the data frame structure & contents

```
In [5]: print(assets_df.dtypes)
assets_df.head()
```

```
highvulns    int64
name         object
os           object
dtype: object
```

```
Out[5]:
```

	highvulns	name	os
0	1	danube	W2K8
1	0	gander	RHEL5
2	2	ganges	W2K8
3	0	mekong	RHEL5
4	0	orinoco	RHEL5

show a "slice" just the operating systmes

```
In [7]: assets_df.os.head()
```

```
Out[7]:
```

0	W2K8
1	RHEL5
2	W2K8
3	RHEL5
4	RHEL5

Name: os, dtype: object

add a new column

```
In [8]: assets_df['ip'] = [ "192.168.1.5", "10.2.7.5", "192.168.1.7",  
                           "10.2.7.6", "10.2.7.7" ]
```

show only nodes with more than one high vulnerability

```
In [9]: assets_df[assets_df.highvulns>1].head()
```

```
Out[9]:
```

	highvulns	name	os	ip
2	2	ganges	W2K8	192.168.1.7

divide nodes into network 'zones' based on IP address

```
In [10]: assets_df['zones'] = np.where(  
        assets_df.ip.str.startswith("192"), "Zone1", "Zone2")
```

get one final view

```
In [11]: assets_df.head()
```

```
Out[11]:
```

	highvulns	name	os	ip	zones
0	1	danube	W2K8	192.168.1.5	Zone1
1	0	gander	RHEL5	10.2.7.5	Zone2
2	2	ganges	W2K8	192.168.1.7	Zone1
3	0	mekong	RHEL5	10.2.7.6	Zone2
4	0	orinoco	RHEL5	10.2.7.7	Zone2