In [1]:

```
import pandas as pd
```

In [2]:

```
df_sample = pd.DataFrame(index = ['Row1','Row2','Row3','Row4'],columns = ['UniqueId','Nam
```

In [3]:

```
df_sample
```

Out[3]:

	UniqueId	Name
Row1	NaN	NaN
Row2	NaN	NaN
Row3	NaN	NaN
Row4	NaN	NaN

In [4]:

```
df_sample['Mark'] = 0
```

In [5]:

```
df_sample
```

Out[5]:

	UniqueId	Name	Mark
Row1	NaN	NaN	0
Row2	NaN	NaN	0
Row3	NaN	NaN	0
Row4	NaN	NaN	0

In [17]:

```
df_sample.iloc[0,0] = "E0115478"
df_sample.iloc[1,0] = "E0245678"
df_sample.iloc[2,0] = "E0357813"
df_sample.iloc[3,0] = "E0478945"
```

In [13]:

```
df_sample.iloc[0,1] = "A"
df_sample.iloc[1,1] = "B"
df_sample.iloc[2,1] = "C"
df_sample.iloc[3,1] = "D"
```

In [14]:

```
df_sample.iloc[0,2] = 40
df_sample.iloc[1,2] = 30
df_sample.iloc[2,2] = 40
df_sample.iloc[3,2] = 50
```

In [18]:

```
df_sample
```

Out[18]:

	UniqueId	Name	Mark
Row1	E0115478	A	40
Row2	E0245678	B	30
Row3	E0357813	C	40
Row4	E0478945	D	50

In [16]:

```
df_sample.duplicated()
```

Out[16]:

```
Row1    False
Row2    False
Row3    False
Row4    False
dtype: bool
```

In [23]:

```
df_sample['stream'] = df_sample['UniqueId'].str[0:3]
```

In [24]:

```
1 df_sample
```

Out[24]:

	UniqueId	Name	Mark	stream
Row1	E0115478	A	40	E01
Row2	E0245678	B	30	E02
Row3	E0357813	C	40	E03
Row4	E0478945	D	50	E04

In [25]:

```
1 dic = {"E01":"AIML","E02":"Cyber","E03":"AIDA","E04":"Bio-Med"}
```

In [26]:

```
1 df_sample["Department"] = df_sample['stream'].map(dic)
```

In [27]:

```
1 df_sample
```

Out[27]:

	UniqueId	Name	Mark	stream	Department
Row1	E0115478	A	40	E01	AIML
Row2	E0245678	B	30	E02	Cyber
Row3	E0357813	C	40	E03	AIDA
Row4	E0478945	D	50	E04	Bio-Med

In [29]:

```
1 import numpy as np
```

In [66]:

```
1 arr = np.arange(1,21).reshape(4,5)
```

In [67]:

```
1 df_sample = pd.DataFrame(arr,index = ['ROW1','ROW2','ROW3','ROW4'],columns=['COL1','COL2'
```

In [68]:

```
df_sample
```

Out[68]:

	COL1	COL2	COL3	COL4	COL5
ROW1	1	2	3	4	5
ROW2	6	7	8	9	10
ROW3	11	12	13	14	15
ROW4	16	17	18	19	20

In [69]:

```
for column in df_sample.columns:
    df_sample[column] = df_sample[column]/df_sample[column].max()
```

In [70]:

```
df_sample
```

Out[70]:

	COL1	COL2	COL3	COL4	COL5
ROW1	0.0625	0.117647	0.166667	0.210526	0.25
ROW2	0.3750	0.411765	0.444444	0.473684	0.50
ROW3	0.6875	0.705882	0.722222	0.736842	0.75
ROW4	1.0000	1.000000	1.000000	1.000000	1.00

In [ ]:

```
Normalization
#done
value/max-value
#yet to be done
(min-value)/(max-min)
Standardization
(value-mean)/std
```