

1:Dataset Processing

```
from google.colab import drive
drive.mount('/content/drive')
```

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

Finding Missing values

```
import numpy as np

data = np.genfromtxt("/content/drive/MyDrive/data/data.csv",
delimiter=",", skip_header=1 , filling_values=np.nan)
print(data)

[[160.  60.  25.]
 [170.  70.  30.]
 [180.  nan  35.]
 [175.  80.  28.]]
```

Masking Values

```
print("Masked values:\n",np.isnan(data))

Masked values:
[[False False False]
 [False False False]
 [False  True False]
 [False False False]]
```

Find mean of columns

```
col_means = np.nanmean(data, axis=0)
print("Mean of the columns:\n",col_means)

Mean of the columns:
[171.25  70.   29.5 ]
```

Replacing NaN with mean value

```
index = np.where(np.isnan(data))
data[index] = np.take(col_means, index[1])
print("After replacing with new data:\n",data)
```

After replacing with new data:

```
[[160.  60.  25.]
 [170.  70.  30.]
 [180.  70.  35.]
 [175.  80.  28.]]
```

Normalization

```
col_min = np.nanmin(data,axis=0)
col_max= np.nanmax(data, axis=0)

col_range=col_max - col_min
col_range[col_range==0]=1.0
normalized=(data-col_min)/col_range

print("Normalized data:\n",normalized)
print("Min per column:\n",normalized.min(axis=0))
print("Max per column:\n",normalized.max(axis=0))
```

Normalized data:

```
[[0.  0.  0. ]
 [0.5  0.5  0.5 ]
 [1.  0.5  1. ]
 [0.75 1.  0.3 ]]
```

Min per column:

```
[0. 0. 0.]
```

Max per column:

```
[1. 1. 1.]
```