

```
In [23]: import pandas as pd #reading behaviour of data using paandas
```

```
In [24]: pd.__version__
```

```
Out[24]: '2.2.2'
```

```
In [25]: #pip install --upgrade openpyxl(for reading excel)
```

```
In [39]: emp = pd.read_excel(r"C:\Users\Ramya\Downloads\Rawdata.xlsx")
```

```
In [40]: emp #this is raw data, i need to clean it
```

```
Out[40]:
```

	Name	Domain	Age	Location	Salary	Exp
0	Mike	Datascience#\$	34 years	Mumbai	5^00#0	2+
1	Teddy^	Testing	45' yr	Bangalore	10%%000	<3
2	Uma#r	Dataanalyst^^#	NaN	NaN	1\$5%000	4> yrs
3	Jane	Ana^^lytics	NaN	Hyderbad	2000^0	NaN
4	Uttam*	Statistics	67-yr	NaN	30000-	5+ year
5	Kim	NLP	55yr	Delhi	6000^\$0	10+

```
In [41]: id(emp) #address of memory
```

```
Out[41]: 2341216167808
```

```
In [42]: emp.columns
```

```
Out[42]: Index(['Name', 'Domain', 'Age', 'Location', 'Salary', 'Exp'], dtype='object')
```

```
In [43]: emp.shape
```

```
Out[43]: (6, 6)
```

```
In [44]: emp.head # 1st 5 rows
```

```
Out[44]: <bound method NDFrame.head of
```

	Name	Domain	Age	Location	Salary	Exp
0	Mike	Datascience#\$	34 years	Mumbai	5^00#0	2+
1	Teddy^	Testing	45' yr	Bangalore	10%%000	<3
2	Uma#r	Dataanalyst^^#	NaN	NaN	1\$5%000	4> yrs
3	Jane	Ana^^lytics	NaN	Hyderbad	2000^0	NaN
4	Uttam*	Statistics	67-yr	NaN	30000-	5+ year
5	Kim	NLP	55yr	Delhi	6000^\$0	10+>

```
In [45]: emp.tail # bottom 5 rows
```

```
Out[45]: <bound method NDFrame.tail of
```

	Name	Domain	Age	Location
0	Mike	Datascience##	34 years	Mumbai
1	Teddy^	Testing	45' yr	Bangalore
2	Uma#r	Dataanalyst^^#	NaN	NaN
3	Jane	Ana^^lytics	NaN	Hyderabad
4	Uttam*	Statistics	67-yr	NaN
5	Kim	NLP	55yr	Delhi

```
In [46]: emp.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6 entries, 0 to 5
Data columns (total 6 columns):
#   Column      Non-Null Count  Dtype
---  -
0   Name        6 non-null      object
1   Domain      6 non-null      object
2   Age         4 non-null      object
3   Location    4 non-null      object
4   Salary      6 non-null      object
5   Exp         5 non-null      object
dtypes: object(6)
memory usage: 420.0+ bytes
```

```
In [47]: emp.isnull() #missing value
```

```
Out[47]:
```

	Name	Domain	Age	Location	Salary	Exp
0	False	False	False	False	False	False
1	False	False	False	False	False	False
2	False	False	True	True	False	False
3	False	False	True	False	False	True
4	False	False	False	True	False	False
5	False	False	False	False	False	False

```
In [48]: emp.isna() # same to check null
```

```
Out[48]:
```

	Name	Domain	Age	Location	Salary	Exp
0	False	False	False	False	False	False
1	False	False	False	False	False	False
2	False	False	True	True	False	False
3	False	False	True	False	False	True
4	False	False	False	True	False	False
5	False	False	False	False	False	False

```
In [49]: emp.isnull().sum
```

```
Out[49]: <bound method DataFrame.sum of
0  False  False  False  False  False  False
1  False  False  False  False  False  False
2  False  False  True   True   False  False
3  False  False  True   False  False  True
4  False  False  False  True   False  False
5  False  False  False  False  False  False>
```

```
In [ ]: # Data
```