```
# Assignment 1
import pandas as pd
import numpy as np


# Task 1 (Create a pandas dataframe (DataFrame name as 'df') (10 observation and 5 features))
df = pd.DataFrame([['Sow',20,'M',75000,'CIRS'],['Vag',20,'M',75000,'CIRS'],['Pav',25,'M',80000,'GG'],
['Sed',21,'M',np.nan,'TIPS'],['Vib',np.nan,'M',55000,'EVPS'],['Sar',25,'M',40000,'TIPS'],
['Muk',19,'F',60000,'Jutes'],['San',20,'M',65000,'EVPS'],['Sam',20,'M',80000,np.nan],['Ara',20,'F',35000,'EVPS']])
df.columns=['Name','Age','Gender','Salary','School']
df.index=[1,2,3,4,5,6,7,8,9,10]
df
```

	Name	Age	Gender	Salary	School
1	Sow	20.0	M	75000.0	CIRS
2	Vag	20.0	M	75000.0	CIRS
3	Pav	25.0	M	80000.0	GG
4	Sed	21.0	M	NaN	TIPS
5	Vib	NaN	M	55000.0	EVPS
6	Sar	25.0	M	40000.0	TIPS
7	Muk	19.0	F	60000.0	Jutes
8	San	20.0	M	65000.0	EVPS
9	Sam	20.0	M	80000.0	NaN
10	Ara	20.0	F	35000.0	EVPS

```
# Task 2 (Check the info of 'df')
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 10 entries, 1 to 10
Data columns (total 5 columns):
 #   Column  Non-Null Count  Dtype  
---  ------  --------------  -----  
 0   Name    10 non-null     object 
 1   Age     9 non-null      float64
 2   Gender  10 non-null     object 
 3   Salary  9 non-null      float64
 4   School  9 non-null      object 
dtypes: float64(2), object(3)
memory usage: 480.0+ bytes
```

```
# Task 3 (Check the descriptive statistics of 'df')
df.describe()
```

	Age	Salary
count	9.000000	9.000000
mean	21.111111	62777.777778
std	2.260777	16791.201400
min	19.000000	35000.000000
25%	20.000000	55000.000000
50%	20.000000	65000.000000
75%	21.000000	75000.000000
max	25.000000	80000.000000

```
# Task 4 (check the 4th index observation with 'loc' slicing operator.)
# index = 4 implies that the serial number = 5, since index starts from 0
df.loc[5]
```

```
Name          Vib
Age           NaN
Gender          M
Salary    55000.0
School       EVPS
Name: 5, dtype: object
```

```
# or we can use iloc
df.iloc[4]
```

```
Name          Vib
Age           NaN
Gender          M
Salary    55000.0
School       EVPS
Name: 5, dtype: object
```

```
# Task 5 (Check the null values in your 'df')
df.isnull().any()
```

```
Name      False
Age        True
Gender    False
Salary     True
School     True
dtype: bool
```

```
# Remove nan
df.Age = df.Age.fillna(df.Age.median())
df.Salary = df.Salary.fillna(df.Salary.median())
df.School = df.School.fillna(df.School.mode().iloc[0])
df
```

	Name	Age	Gender	Salary	School
1	Sow	20.0	M	75000.0	CIRS
2	Vag	20.0	M	75000.0	CIRS
3	Pav	25.0	M	80000.0	GG
4	Sed	21.0	M	65000.0	TIPS
5	Vib	20.0	M	55000.0	EVPS
6	Sar	25.0	M	40000.0	TIPS
7	Muk	19.0	F	60000.0	Jutes
8	San	20.0	M	65000.0	EVPS
9	Sam	20.0	M	80000.0	EVPS
10	Ara	20.0	F	35000.0	EVPS