1. Import all python libraries
2. Load Dataset (.csv file)
3. perform describe(), info(), isnull() functions
4. Remove all Null values (Data Cleaning)
5. Type conversion
6. Turn catagorical variable into quantitative variable

```
import pandas as pd  #To perform the operations on tabular dataset
import numpy as np #To perform mathematical operations. ex: Operations on Array
```

```
df = pd.read_csv("/content/Academic-Student.csv")
df
```

	rollno	marks	gender	age	phd
0	1	22.0	0	22.0	yes
1	2	45.0	0	23.0	no
2	3	66.0	0	NaN	no
3	4	77.0	0	34.0	NaN
4	5	33.0	0	22.0	no
...	...	...	...	...	...
95	96	15.0	0	23.0	yes
96	97	20.0	1	22.0	no
97	98	30.0	1	21.0	NaN
98	99	30.0	1	21.0	no
99	100	20.0	1	22.0	yes

100 rows × 5 columns

```
df.head()
```

	rollno	marks	gender	age	phd
0	1	22.0	0	22.0	yes
1	2	45.0	0	23.0	no
2	3	66.0	0	NaN	no
3	4	77.0	0	34.0	NaN
4	5	33.0	0	22.0	no

```
df.tail()
```

	rollno	marks	gender	age	phd
95	96	15.0	0	23.0	yes
96	97	20.0	1	22.0	no
97	98	30.0	1	21.0	NaN
98	99	30.0	1	21.0	no
99	100	20.0	1	22.0	yes

```
df.describe()
```

	rollno	marks	gender	age

```
df.isnull()
```

	rollno	marks	gender	age	phd
0	False	False	False	False	False
1	False	False	False	False	False
2	False	False	False	True	False
3	False	False	False	False	True
4	False	False	False	False	False
...	...	...	...	...	...
95	False	False	False	False	False
96	False	False	False	False	False
97	False	False	False	False	True
98	False	False	False	False	False
99	False	False	False	False	False

100 rows × 5 columns

```
df.isnull().sum()
```

```
rollno    0
marks     0
gender    0
age       5
phd       8
dtype: int64
```

```
df.dropna(inplace=True)
```

```
df.isnull().sum()
```

```
rollno    0
marks     0
gender    0
age       0
phd       0
dtype: int64
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 87 entries, 0 to 99
Data columns (total 5 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   rollno  87 non-null     int64
 1   marks   87 non-null     float64
 2   gender  87 non-null     int64
 3   age     87 non-null     float64
 4   phd     87 non-null     object
dtypes: float64(2), int64(2), object(1)
memory usage: 4.1+ KB
```

```
df['marks'] = df['marks'].astype('int')
df['marks'].dtypes
```

```
dtype('int64')
```

```
df['age'] = df['age'].astype('int')
df['age'].dtypes
```

```
dtype('int64')
```

```
df["phd"] = df["phd"].replace('yes', 1)
df['phd'] = df['phd'].replace('no', 0)

print(df)
```

```
    rollno  marks  gender  age  phd
0        1     22       0   22    1
1        2     45       0   23    0
4        5     33       0   22    0
5        6     22      10   21    0
```

```
7        8     63       0   24    0
..     ...    ...     ...  ...  ...
93      94     17       0   25    1
95      96     15       0   23    1
96      97     20       1   22    0
98      99     30       1   21    0
99     100     20       1   22    1

[87 rows x 5 columns]
```

```
df['phd'] = df['phd'].astype(int)
df['phd'].dtypes
```

```
dtype('int64')
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 87 entries, 0 to 99
Data columns (total 5 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   rollno  87 non-null     int64
 1   marks   87 non-null     int64
 2   gender  87 non-null     int64
 3   age     87 non-null     int64
 4   phd     87 non-null     int64
dtypes: int64(5)
memory usage: 4.1 KB
```