

```
import numpy as np
import pandas as pd
```

```
df = pd.read_csv('/content/Placement.csv')
```

```
df.describe()
```

	sl_no	ssc_p	hsc_p	degree_p	etest_p	mba_p	si
count	215.000000	213.000000	214.000000	213.000000	213.000000	213.000000	148.000000
mean	108.000000	67.372911	66.344065	66.388404	72.219390	62.266103	288655.40
std	62.209324	10.811464	10.921885	7.256840	13.255118	5.837303	93457.40
min	1.000000	40.890000	37.000000	50.200000	50.000000	51.210000	200000.00
25%	54.500000	60.800000	60.850000	61.000000	60.000000	57.900000	240000.00
50%	108.000000	67.000000	65.000000	66.000000	71.000000	62.000000	265000.00
75%	161.500000	76.000000	73.000000	72.000000	84.000000	66.230000	300000.00
max	215.000000	89.400000	97.700000	91.000000	98.000000	77.890000	940000.00

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 215 entries, 0 to 214
Data columns (total 15 columns):
#   Column                Non-Null Count  Dtype
---  -
0   sl_no                  215 non-null   int64
1   gender                 213 non-null   object
2   ssc_p                  213 non-null   float64
3   ssc_b                  214 non-null   object
4   hsc_p                  214 non-null   float64
5   hsc_b                  213 non-null   object
6   hsc_s                  213 non-null   object
7   degree_p               213 non-null   float64
8   degree_t               214 non-null   object
9   workex                 213 non-null   object
10  etest_p                213 non-null   float64
11  specialisation         213 non-null   object
12  mba_p                  213 non-null   float64
13  status                 215 non-null   object
14  salary                 148 non-null   float64
dtypes: float64(6), int64(1), object(8)
memory usage: 25.3+ KB
```

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

```
sl_no      0
gender      2
ssc_p       2
ssc_b       1
hsc_p       1
hsc_b       2
hsc_s       2
degree_p    2
degree_t    1
workex      2
etest_p     2
specialisation  2
```

```
mba_p      2
status     0
salary     67
dtype: int64
```

```
print("Information of Datasets:\n", df.info())
print("Shape of Dataset (row x column):\n", df.shape)
print("Columns Name:\n", df.columns)
print("Total elements in dataset:\n", df.size)
print("Datatype of attributes (columns):\n", df.dtypes)
print("First five rows:\n", df.head)
print("Last five row:\n", df.tail)
print("Any five rows:\n", df.sample(5))
```

▲

3

```

12 mba_p          213 non-null    float64
13 status        215 non-null    object
14 salary        148 non-null    float64
dtypes: float64(6), int64(1), object(8)
memory usage: 25.3+ KB

```

```

df['gender'].fillna(df['gender'].mode()[0], inplace = True)
df['ssc_p'].fillna(df['ssc_p'].mean(), inplace = True)
print("Mode of ssc_b: ", df['ssc_b'].mode())
df['ssc_b'].fillna(df['ssc_b'].mode()[0], inplace = True)

print("Total number of missing values in Dataset: ", df.isna().sum())

```

```

Mode of ssc_b: 0    Central
Name: ssc_b, dtype: object
Total number of missing values in Dataset:  sl_no    0
gender    0
ssc_p    0
ssc_b    0
hsc_p    1
hsc_b    2
hsc_s    2
degree_p  2
degree_t  1
workex    2
etest_p   2
specialisation  2
mba_p    2
status    0
salary    67
dtype: int64

```

```

df['hsc_p'].fillna(df['hsc_p'].mean(), inplace = True)
df['hsc_b'].fillna(df['hsc_b'].mode()[0], inplace = True)
df['hsc_s'].fillna(df['hsc_s'].mode()[0], inplace = True)
df['degree_p'].fillna(df['degree_p'].mean(), inplace = True)
df['degree_t'].fillna(df['degree_t'].mode()[0], inplace = True)
df['workex'].fillna(df['workex'].mode()[0], inplace = True)
df['etest_p'].fillna(df['etest_p'].mean(), inplace = True)
df['specialisation'].fillna(df['specialisation'].mode()[0], inplace = True)
df['mba_p'].fillna(df['mba_p'].mean(), inplace = True)
df['salary'].fillna(df['salary'].mean(), inplace = True)

```

```
df.isna().sum()
```

```

sl_no    0
gender    0
ssc_p    0
ssc_b    0
hsc_p    0
hsc_b    0
hsc_s    0
degree_p  0
degree_t  0
workex    0
etest_p   0
specialisation  0
mba_p    0
status    0
salary    0
dtype: int64

```

Changing Datatypes of Columns

```
df['sl_no'] = df['sl_no'].astype('int8')
```

```
df['sl_no'] = df['sl_no'].astype('int64')
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 215 entries, 0 to 214
Data columns (total 16 columns):
#   Column                Non-Null Count  Dtype
---  -
0   sl_no                  215 non-null    int8
1   gender                 215 non-null    object
2   ssc_p                  215 non-null    float64
3   ssc_b                  215 non-null    object
4   hsc_p                  215 non-null    float64
5   hsc_b                  215 non-null    object
6   hsc_s                  215 non-null    object
7   degree_p               215 non-null    float64
8   degree_t               215 non-null    object
9   workex                 215 non-null    object
10  etest_p                 215 non-null    float64
11  specialisation         215 non-null    object
12  mba_p                  215 non-null    float64
13  status                 215 non-null    object
14  salary                 215 non-null    float64
15  sl_no]                 215 non-null    int64
dtypes: float64(6), int64(1), int8(1), object(8)
memory usage: 25.5+ KB
```

Converting categorical (qualitative) variable to numeric (quantitative) variable

```
df['gender'].replace(['M', 'F'], [0, 1], inplace = True)
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 215 entries, 0 to 214
Data columns (total 16 columns):
#   Column                Non-Null Count  Dtype
---  -
0   sl_no                  215 non-null    int8
1   gender                 215 non-null    int64
2   ssc_p                  215 non-null    float64
3   ssc_b                  215 non-null    object
4   hsc_p                  215 non-null    float64
5   hsc_b                  215 non-null    object
6   hsc_s                  215 non-null    object
7   degree_p               215 non-null    float64
8   degree_t               215 non-null    object
9   workex                 215 non-null    object
10  etest_p                 215 non-null    float64
11  specialisation         215 non-null    object
12  mba_p                  215 non-null    float64
13  status                 215 non-null    object
14  salary                 215 non-null    float64
15  sl_no]                 215 non-null    int64
dtypes: float64(6), int64(2), int8(1), object(7)
memory usage: 25.5+ KB
```

```
df.head()
```

	hsc_s	degree_p	degree_t	workex	etest_p	specialisation	mba_p	status	
	Commerce	58.00	Sci&Tech	No	55.0	Mkt&HR	58.80	Placed	27000
	Science	77.48	Sci&Tech	Yes	86.5	Mkt&Fin	66.28	Placed	20000
	Arts	64.00	Comm&Mgmt	No	75.0	Mkt&Fin	57.80	Placed	25000
	Science	52.00	Sci&Tech	No	66.0	Mkt&HR	59.43	Not Placed	28865
	Commerce	73.30	Comm&Mgmt	No	96.8	Mkt&Fin	55.50	Placed	42500

Next steps:

Generate code with df

 View recommended plots

```
df['status'].replace(["Placed", "Not Placed"], [0, 1], inplace = True)
```

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 215 entries, 0 to 214
Data columns (total 16 columns):
#   Column                Non-Null Count  Dtype
---  -
0   sl_no                  215 non-null    int8
1   gender                 215 non-null    int64
2   ssc_p                  215 non-null    float64
3   ssc_b                  215 non-null    object
4   hsc_p                  215 non-null    float64
5   hsc_b                  215 non-null    object
6   hsc_s                  215 non-null    object
7   degree_p               215 non-null    float64
8   degree_t               215 non-null    object
9   workex                 215 non-null    object
10  etest_p                 215 non-null    float64
11  specialisation         215 non-null    object
12  mba_p                  215 non-null    float64
13  status                 215 non-null    int64
14  salary                 215 non-null    float64
15  sl_no]                 215 non-null    int64
dtypes: float64(6), int64(3), int8(1), object(6)
memory usage: 25.5+ KB
```

```
df['workex'].replace(['Yes','No'], [1, 0], inplace = True)
```

df.head()

	sl_no	gender	ssc_p	ssc_b	hsc_p	hsc_b	hsc_s	degree_p	degree_t	workex
0	1	0	67.00	Others	91.00	Others	Commerce	58.00	Sci&Tech	No
1	2	0	79.33	Central	78.33	Others	Science	77.48	Sci&Tech	Yes
2	3	0	65.00	Central	68.00	Central	Arts	64.00	Comm&Mgmt	No
3	4	0	56.00	Central	52.00	Central	Science	52.00	Sci&Tech	No
4	5	0	85.80	Central	73.60	Central	Commerce	73.30	Comm&Mgmt	No

Next steps:

Generate code with df

 View recommended plots

```
df['ssc_b'] = df['ssc_b'].astype('category')

df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 215 entries, 0 to 214
Data columns (total 16 columns):
#   Column                Non-Null Count  Dtype
---  -
0   sl_no                 215 non-null    int8
1   gender                215 non-null    int64
2   ssc_p                 215 non-null    float64
3   ssc_b                 215 non-null    category
4   hsc_p                 215 non-null    float64
5   hsc_b                 215 non-null    object
6   hsc_s                 215 non-null    object
7   degree_p              215 non-null    float64
8   degree_t              215 non-null    object
9   workex                215 non-null    int64
10  etest_p               215 non-null    float64
11  specialisation        215 non-null    object
12  mba_p                 215 non-null    float64
13  status                215 non-null    int64
14  salary                215 non-null    float64
15  sl_no]                215 non-null    int64
dtypes: category(1), float64(6), int64(4), int8(1), object(4)
memory usage: 24.2+ KB
```

df.head()

	hsc_s	degree_p	degree_t	workex	etest_p	specialisation	mba_p	status	
	Commerce	58.00	Sci&Tech	0	55.0	Mkt&HR	58.80	0	27000
	Science	77.48	Sci&Tech	1	86.5	Mkt&Fin	66.28	0	20000
	Arts	64.00	Comm&Mgmt	0	75.0	Mkt&Fin	57.80	0	25000
	Science	52.00	Sci&Tech	0	66.0	Mkt&HR	59.43	1	28865
	Commerce	73.30	Comm&Mgmt	0	96.8	Mkt&Fin	55.50	0	42500

Next steps:

Generate code with df

 View recommended plots

```
df['ssc_b'].replace(['Others', 'Central'], [0, 1], inplace = True)

df.head()
```

	sl_no	gender	ssc_p	ssc_b	hsc_p	hsc_b	hsc_s	degree_p	degree_t	work
0	1	0	67.00	0	91.00	Others	Commerce	58.00	Sci&Tech	
1	2	0	79.33	1	78.33	Others	Science	77.48	Sci&Tech	
2	3	0	65.00	1	68.00	Central	Arts	64.00	Comm&Mgmt	
3	4	0	56.00	1	52.00	Central	Science	52.00	Sci&Tech	
4	5	0	85.80	1	73.60	Central	Commerce	73.30	Comm&Mgmt	

Next steps:

[Generate code with df](#)

 [View recommended plots](#)

```
df['salary'] = (df['salary'] - df['salary'].min()) / (df['salary'].max() - df['salary'].min())
```

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 215 entries, 0 to 214
Data columns (total 16 columns):
#   Column                Non-Null Count  Dtype  
---  -
0   sl_no                 215 non-null   int8    
1   gender                215 non-null   int64   
2   ssc_p                 215 non-null   float64 
3   ssc_b                 215 non-null   category
4   hsc_p                 215 non-null   float64 
5   hsc_b                 215 non-null   object  
6   hsc_s                 215 non-null   object  
7   degree_p              215 non-null   float64 
8   degree_t              215 non-null   object  
9   workex                215 non-null   int64   
10  etest_p               215 non-null   float64 
11  specialisation         215 non-null   object  
12  mba_p                 215 non-null   float64 
13  status                 215 non-null   int64   
14  salary                 215 non-null   float64 
15  sl_no]                 215 non-null   int64   
dtypes: category(1), float64(6), int64(4), int8(1), object(4)
memory usage: 24.2+ KB
```

df.head()

	_b	hsc_s	degree_p	degree_t	workex	etest_p	specialisation	mba_p	status	salary	sl_n
	ers	Commerce	58.00	Sci&Tech	0	55.0	Mkt&HR	58.80	0	0.094595	
	ers	Science	77.48	Sci&Tech	1	86.5	Mkt&Fin	66.28	0	0.000000	
	tral	Arts	64.00	Comm&Mgmt	0	75.0	Mkt&Fin	57.80	0	0.067568	
	tral	Science	52.00	Sci&Tech	0	66.0	Mkt&HR	59.43	1	0.119805	
	tral	Commerce	73.30	Comm&Mgmt	0	96.8	Mkt&Fin	55.50	0	0.304054	

Next steps:

[Generate code with df](#)

 [View recommended plots](#)

Start coding or [generate](#) with AI.

