```
1 #handling missing values using imputation
2 #importing the libraries
3 import pandas as pd
4 import matplotlib.pyplot as plt
5 import seaborn as sns
```

```
1 # loading the dataset to pandas dataframe
2 dataset=pd.read_csv('/content/Placement_Dataset.csv')
```

```
1 dataset.head()
```

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

Next steps:   **Generate code with** `dataset`   **View recommended plots**   **New interactive sheet**

```
1 dataset.shape
```

(215, 15)

```
1 dataset.isnull().sum()
```

|  | 0 |
|---|---|
| **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** | 67 |

**dtype:** int64

```
1 # above there are 67 null values in salary column
2 #analyse the distribution of data in the salary
3 fig,ax=plt.subplots(figsize=(5,5))
4 sns.distplot(dataset['salary'])
5 plt.show()
```

⇥  `<ipython-input-6-07781a21be3f>`:4: UserWarning:

   `distplot` is a deprecated function and will be removed in seaborn v0.14.0.

   Please adapt your code to use either `displot` (a figure-level function with
   similar flexibility) or `histplot` (an axes-level function for histograms).

   For a guide to updating your code to use the new functions, please see
   https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

   sns.distplot(dataset['salary'])
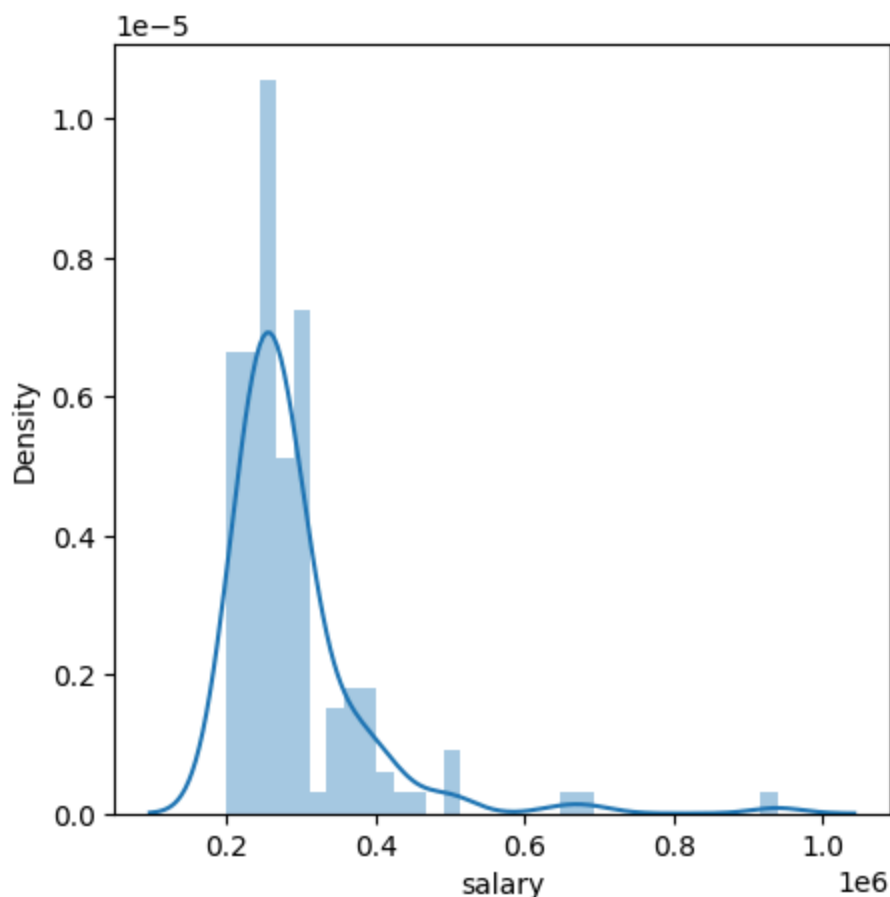


we cannot use mean values in the place og graph. it is skew-distribution so we use median,mode

```
1 #replace the missing values with median value
2 dataset['salary'].fillna(dataset['salary'].median(),inplace=True)
```

⇥  `<ipython-input-7-36758c657fc5>`:2: FutureWarning: A value is trying to be set on a copy o
   The behavior will change in pandas 3.0. This inplace method will never work because the

   For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col

   dataset['salary'].fillna(dataset['salary'].median(),inplace=True)

```
1 dataset.isnull().sum()
```

|  | **0** |
|---|---|
| **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

now null values in salaries become 0 above one is imputation technique.

```
1 #dropping technique
2 #dropping all the rows which contains the missing values
3 salary_dataset=pd.read_csv('/content/Placement_Dataset.csv')
```

```
1 salary_dataset.shape
```

(215, 15)

```
1 salary_dataset.isnull().sum()
```

|  | 0 |
|---|---|
| **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** | 67 |

**dtype:** int64

```
1 salary_dataset=salary_dataset.dropna(how='any')
```

```
1 salary_dataset.isnull().sum()
```

|        | 0 |
|--------|---|
| sl_no  | 0 |