Inferential Analysis

1)Replace the NaN values with correct value. And justify why you have chosen the same.

We want to replace the Nan values I choosen **median** to find the correct values.The middle value in the dataset it provides a better representation of the central value.

2)How many of them are not placed?

```
]: dataset["status"].value_counts()

]: status
   Placed        148
   Not Placed     67
   Name: count, dtype: int64
```

There are 67 people not placed .I used value_counts () ,it shows the placement and non placed status.

3)Find the reason for non-placement from the dataset?

The reason for non-placement candidate means they don't have previous experience in working environment this is the one of the reason they don't placed in any companies.

4) What kind of relation between salary and mba_p?

The correlation between the mba_p and salary it is directly proportional only 14%.if mba_p mark get increased salary also get increased but the increments ratio is only 14%.

5) Which specialization is getting minimum salary?

Code: dataset["salary"].min()

200000 marketing and finance

6)How many of them getting above 500000 salaries?

```
salaries_above_500000=dataset[dataset["salary"]>500000]
```

```
print(salaries_above_500000["salary"])
```

```
119    940000.0
150    690000.0
177    650000.0
Name: salary, dtype: float64
```

Three of them are getting above 500000 salaries.

7)Test the Analysis of Variance between etest_p and mba_p at signifance level 5%. (Make decision using Hypothesis Testing)

```
import scipy.stats as stats
stats.f_oneway(dataset['etest_p'],dataset['mba_p'])
```

```
F_onewayResult(statistic=98.64487057324706, pvalue=4.672547689133573e-21)
```

p<5% means reject null hypothesis accept alternate hypothesis

here we get probability value is 46% . 46%<5%

so that we accept null hypothesis.

8) Test the similarity between the degree_t (Sci&Tech) and specialisation (Mkt&HR) with respect to salary at significance level of 5%.(Make decision using Hypothesis Testing)

Different condition and same group

Different conditions are degree_t(sci&Tech) and specialisation(mkt&HR)

Same group is salary

```
#Independent sample -unpaired T Test

#same Group(salary) but different condition(degree_t(Sci&Tech),specialisation(Mkt&HR))


from scipy.stats import ttest_ind
dataset=dataset.dropna()
degree_t=dataset[dataset['degree_t']=='Sci&Tech']['salary']
specialisation=dataset[dataset['specialisation']=='Mkt&HR']['salary']
t_stat,p_value=ttest_ind(degree_t,specialisation)
print(f"T-statistic:{t_stat},P-value:{p_value}")
```

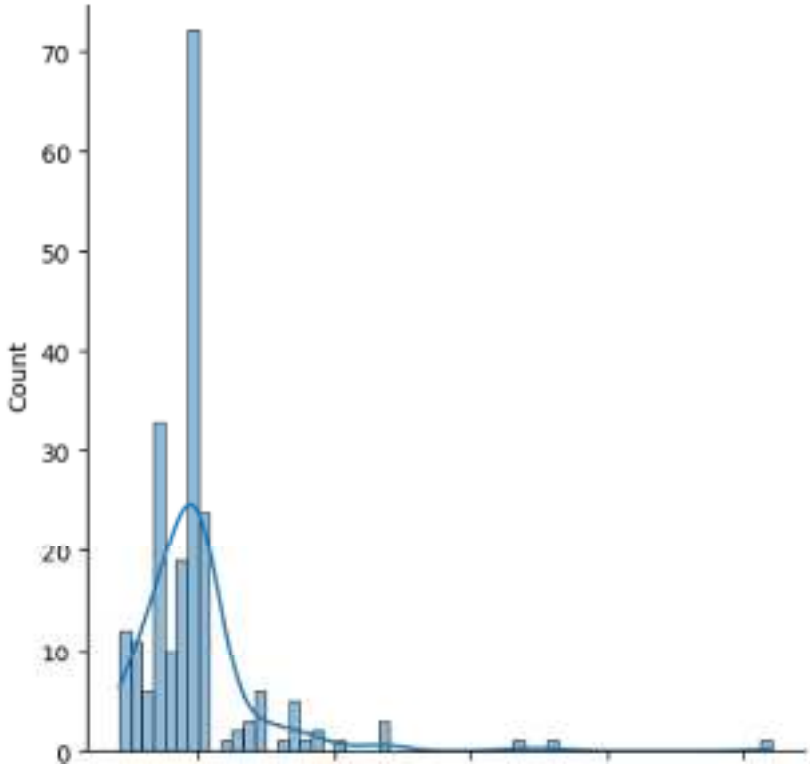
```
T-statistic:2.692041243555374,P-value:0.007897969943471179
```

Probability is 0.007 so that we reject null hypothesis and accept alternate hypothesis

9)Convert the normal distribution to standard normal distribution for salary column

Code:

stdNBgraph(dataset["salary"])

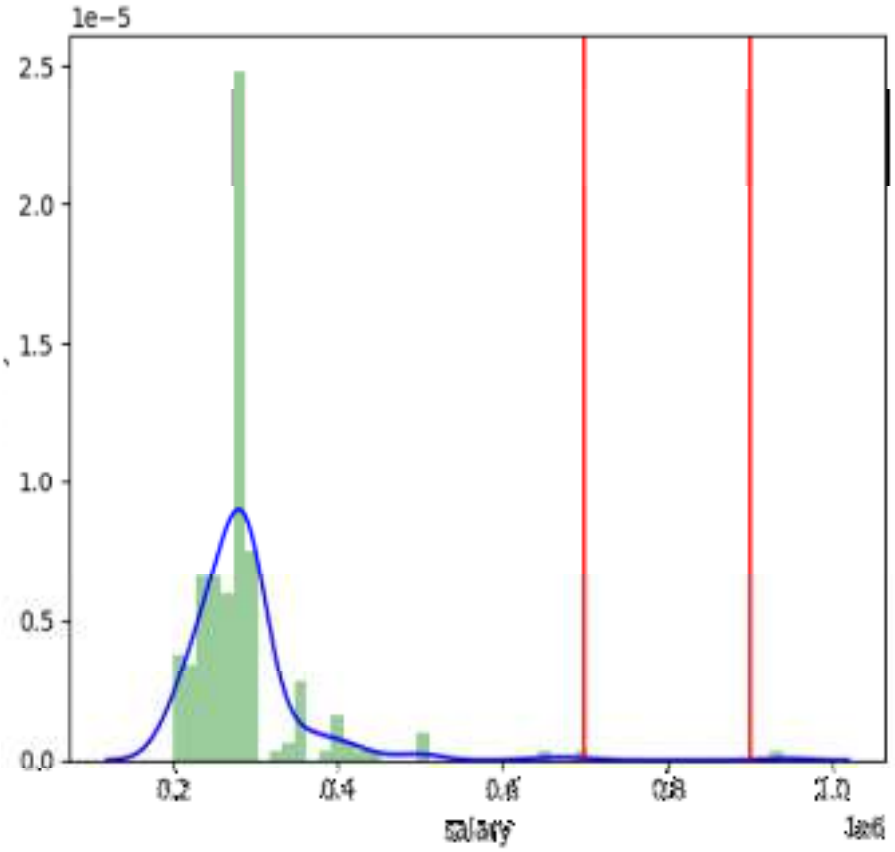


10)What is the probability Density Function of the salary range from 700000 to 900000?

```
get_pdf_probability(dataset["salary"],700000,900000)
```

```
C:\Users\RRDIL\AppData\Local\Temp\ipykernel_12096\3135787118.py:6: 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

  ax=sns.distplot(dataset,kde=True,kde_kws={'color':'blue'},color='Green')
Mean=288655.405,Standard Deviation=77457.900
The area between range(700000,900000:5.4647277868994836e-08
5.4647277868994836e-08
```

The probability range between the 700000 to 900000 is 5.46



11) Test the similarity between the degree_t(Sci&Tech)with respect to etest_p and mba_p at significance level of 5%.(Make decision using Hypothesis Testing)

```
#dependent sample -paired T Test

#same Group degree_t(Sci&Tech) but different condition(etest_p and mba_p)

from scipy.stats import ttest_ind
dataset=dataset.dropna()
etest_p=dataset[dataset['degree_t']=='Sci&Tech']['etest_p']
mba_p=dataset[dataset['degree_t']=='Sci&Tech']['mba_p']

ttest_ind(etest_p,mba_p)
```

```
TtestResult(statistic=4.532000225151251, pvalue=1.4289217003775636e-05, df=116.0)
```

Probability value is 1.42 .so that we reject alternate hypothesis and the we used null hypothesis.

12) Which parameter is highly correlated with salary?

```
dataset.corr(numeric_only=True)
```

	sl_no	ssc_p	hsc_p	degree_p	etest_p	mba_p	salary
sl_no	1.000000	-0.078155	-0.085711	-0.088281	0.063636	0.022327	0.051550
ssc_p	-0.078155	1.000000	0.511472	0.538404	0.261993	0.388478	0.023571
hsc_p	-0.085711	0.511472	1.000000	0.434206	0.245113	0.354823	0.054506
degree_p	-0.088281	0.538404	0.434206	1.000000	0.224470	0.402364	-0.014148
etest_p	0.063636	0.261993	0.245113	0.224470	1.000000	0.218055	0.152829
mba_p	0.022327	0.388478	0.354823	0.402364	0.218055	1.000000	0.146324
salary	0.051550	0.023571	0.054506	-0.014148	0.152829	0.146324	1.000000

Etest_p only have 15% so that this only highly correlated

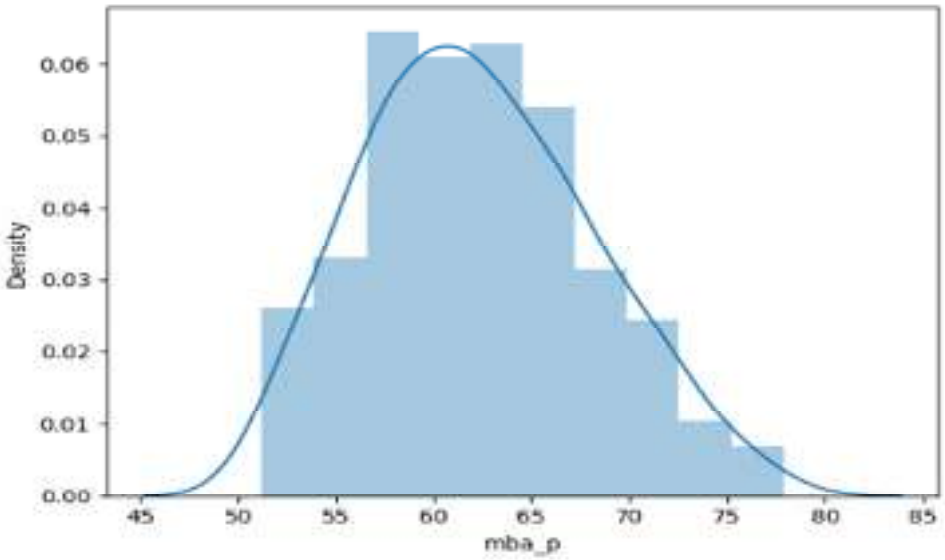
13) Plot any useful graph and explain it.

```
import seaborn as sns
```

```
sns.distplot(dataset["mba_p"])
```

```
C:\Users\RRDIL\AppData\Local\Temp\ipykernel_7844\3673285814.py:1: 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["mba_p"])
```

```
<Axes: xlabel='mba_p', ylabel='Density'>
```



I used to distplot for mba_p .This distplot represent both histogram and also probability density function starting range of mark is 48 and the density function also get started.high range of mark is in 60 to 65 range probability density value is 0.06 and then that graph get falls down the value of the probability density function also getting down next value is getting zero in range of 80 t0 85 marks.