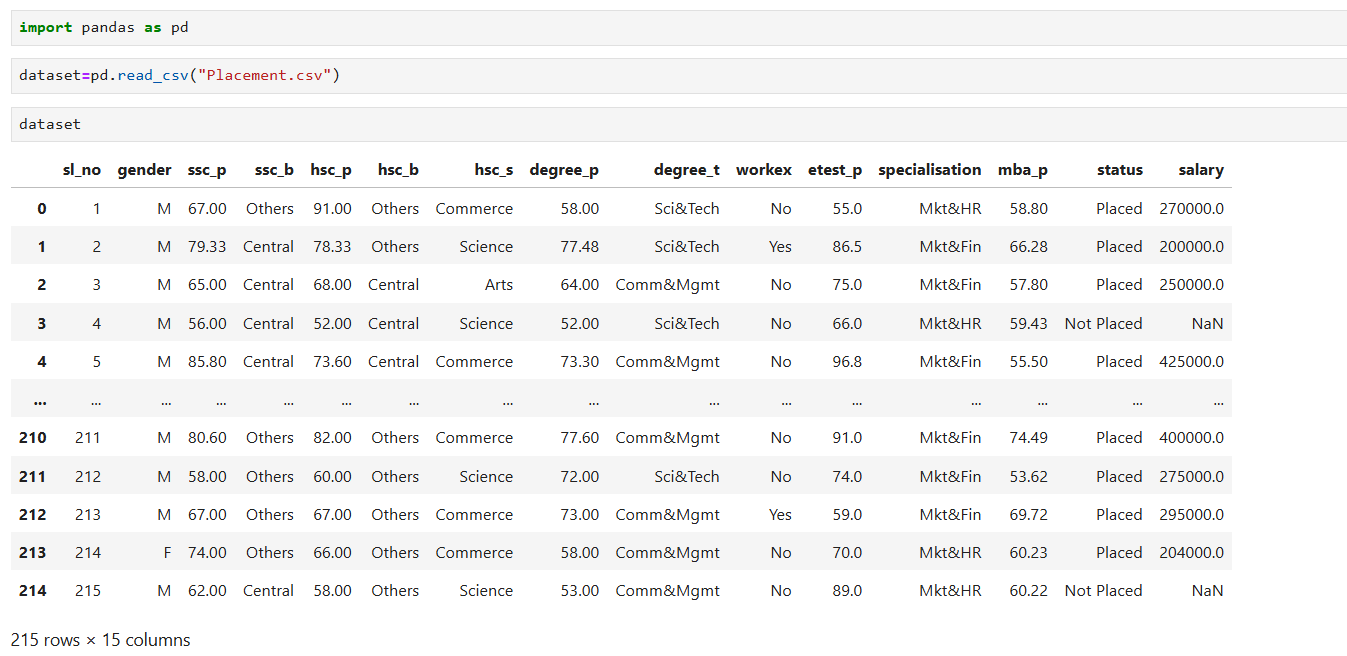
**Analysis of placement dataset of students:**

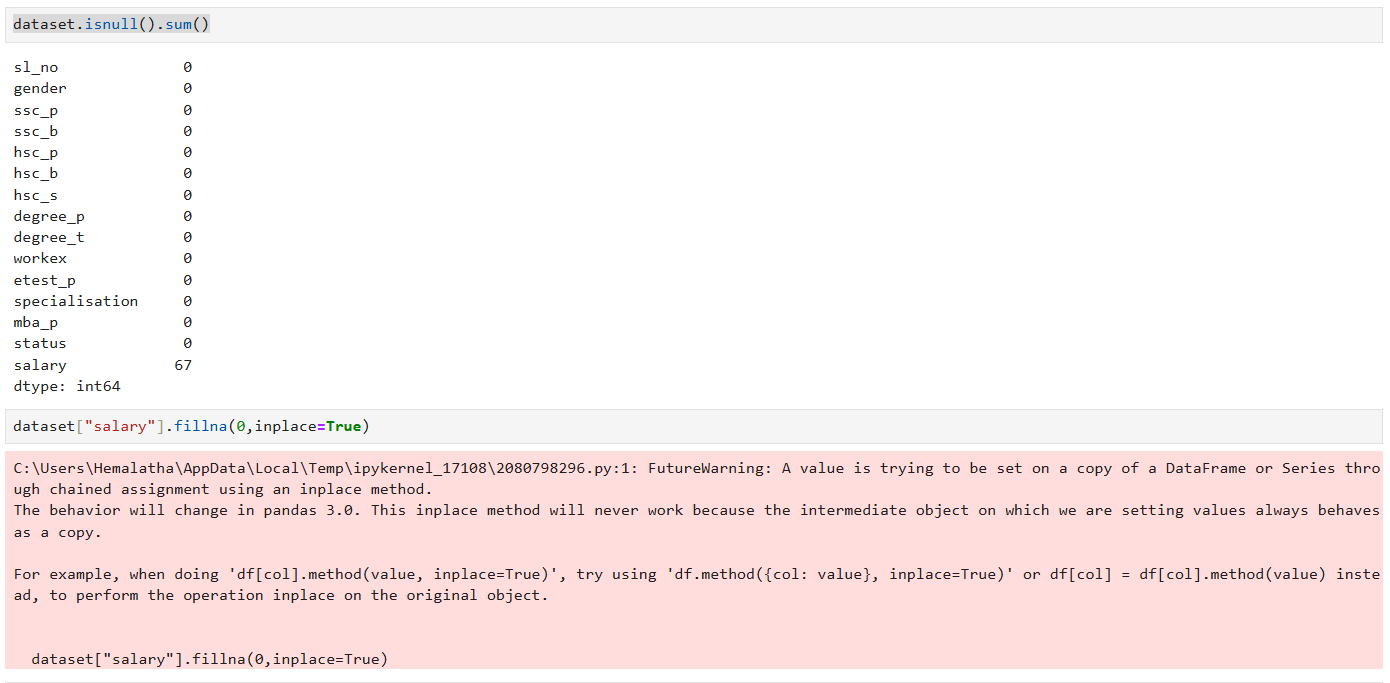
1. **Replace the NaN values with the correct value. And justify why you have chosen the same.**

Step1: Import necessary library

Step2: Read the dataset and assign it to a variable

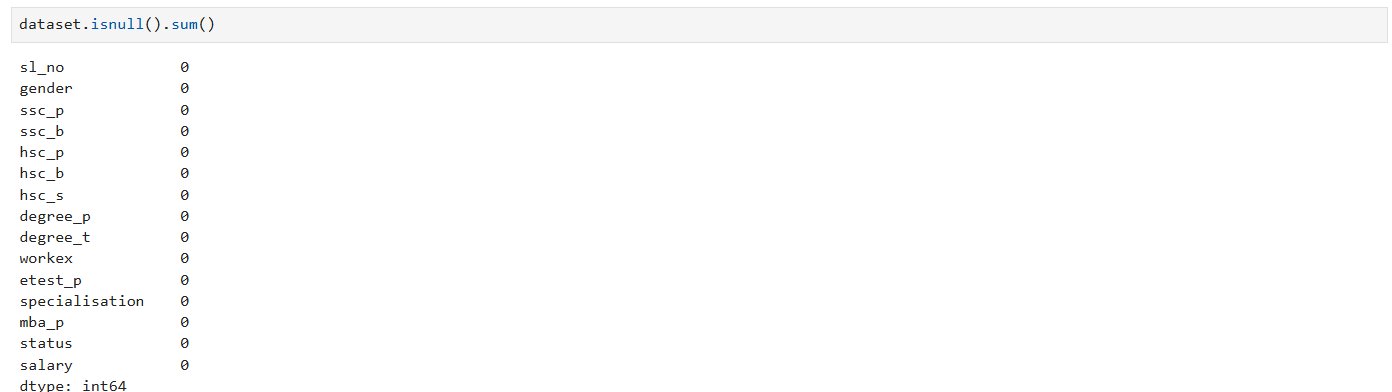
****

Step3: Check for Nan values and replace values based on the columns

****

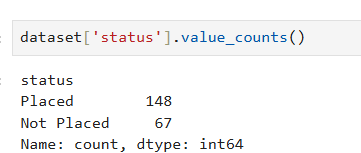
Step4: Based on this dataset, if salary column is replaced with mean/median/mode or any other value then it means the person gets salary when his status is Not Placed, then the logic gets affected so we can replace with 0 values. So that logic will remain same and ‘Nan’ will be replaced with 0 value.

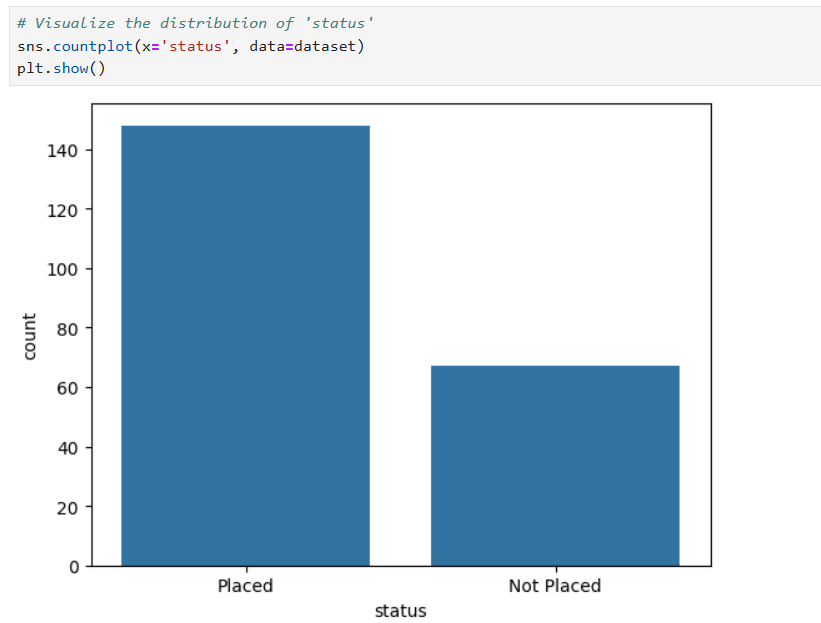
Step5: dataset.isnull().sum() using this code we can cross check whether the dataset has any null values or not.

****

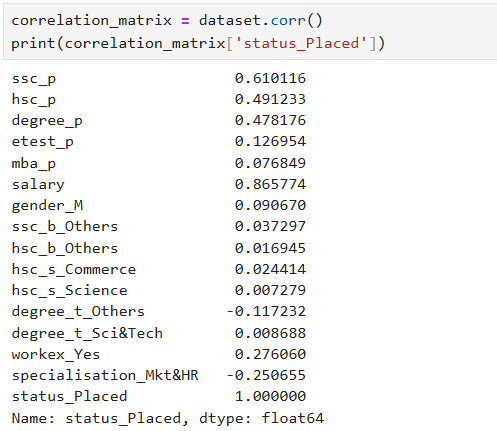
1. **How many of them are not placed?**

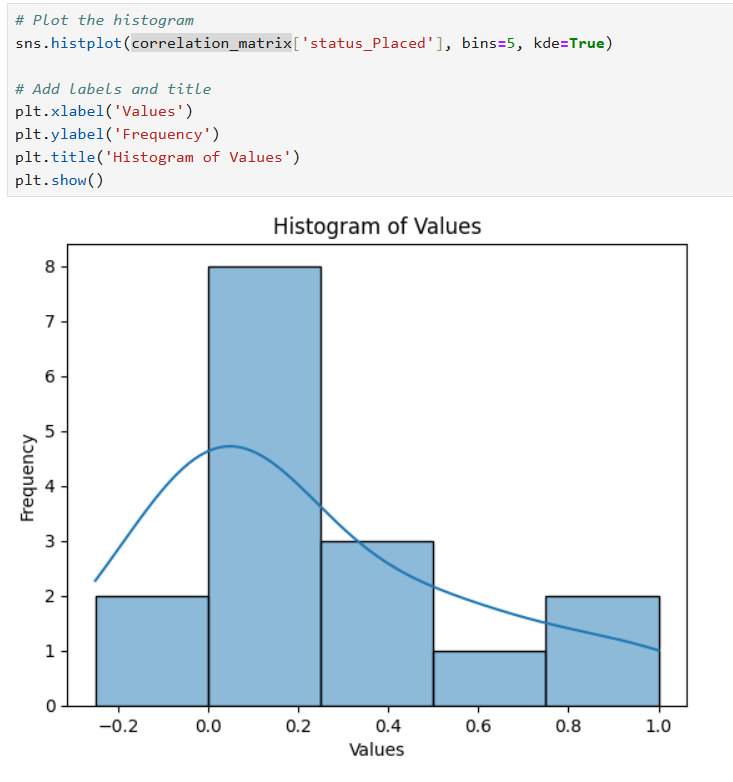
In this dataset 67 members out of 215 didn’t get Placed.

****

****

1. **Find the reason for non placement from the dataset?**

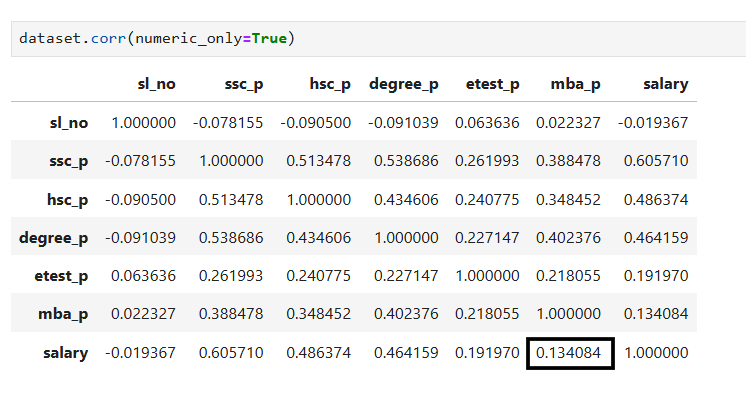
****

****

****

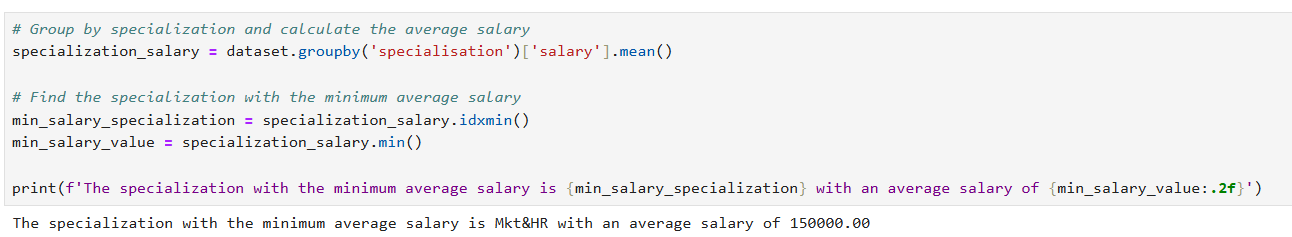
On the above graph we can see that the members who scored less marks in ssc\_p, hsc\_p and degree\_p didn’t get placed.

1. **What kind of relation between salary and mba\_p**

****

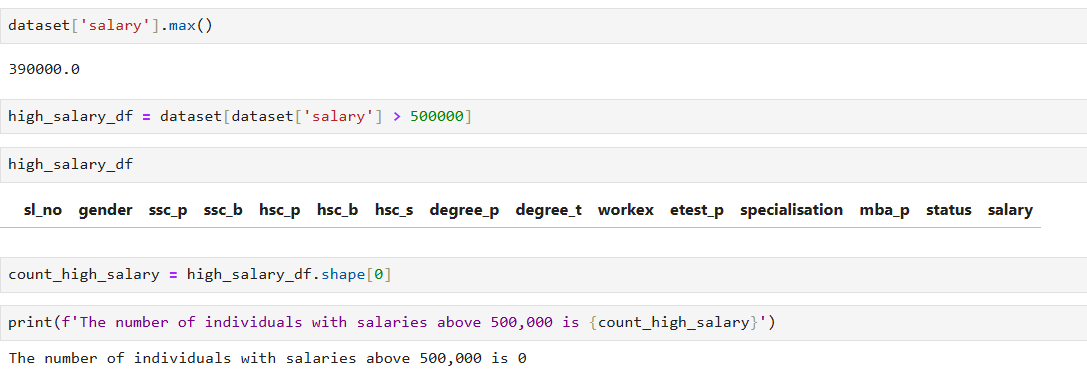
The MBA pass mark and Salary of the student are only 13% correlated. The given value of (r = 0.134048) indicates a weak positive linear correlation between the two variables.

1. **Which specialization is getting a minimum salary?**

****

Specialisation under Mkt & Hr is getting minimum salary of 150000.00

1. **How many of them are getting above 500,000 salaries?**

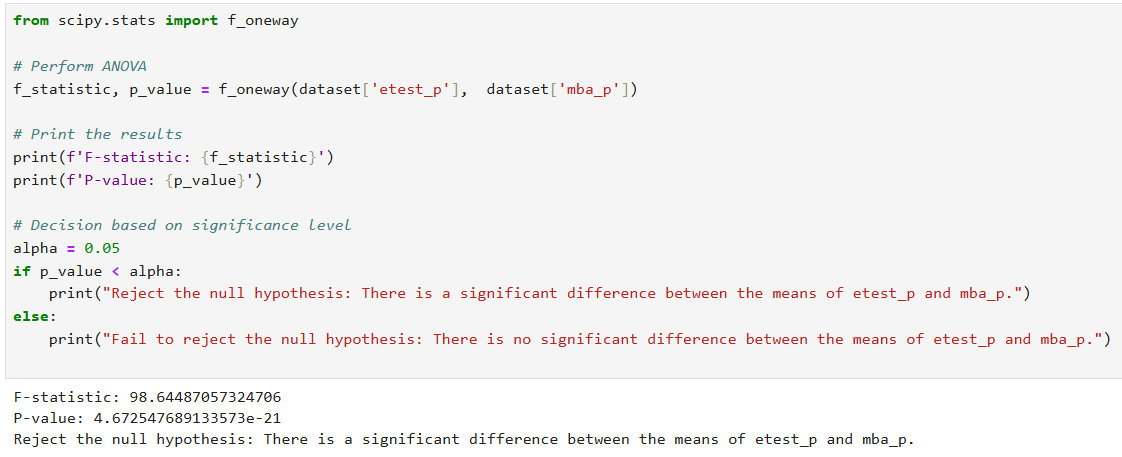
****

There are 0 individuals getting salary above 500000, because we

removed outliers in the dataset.****

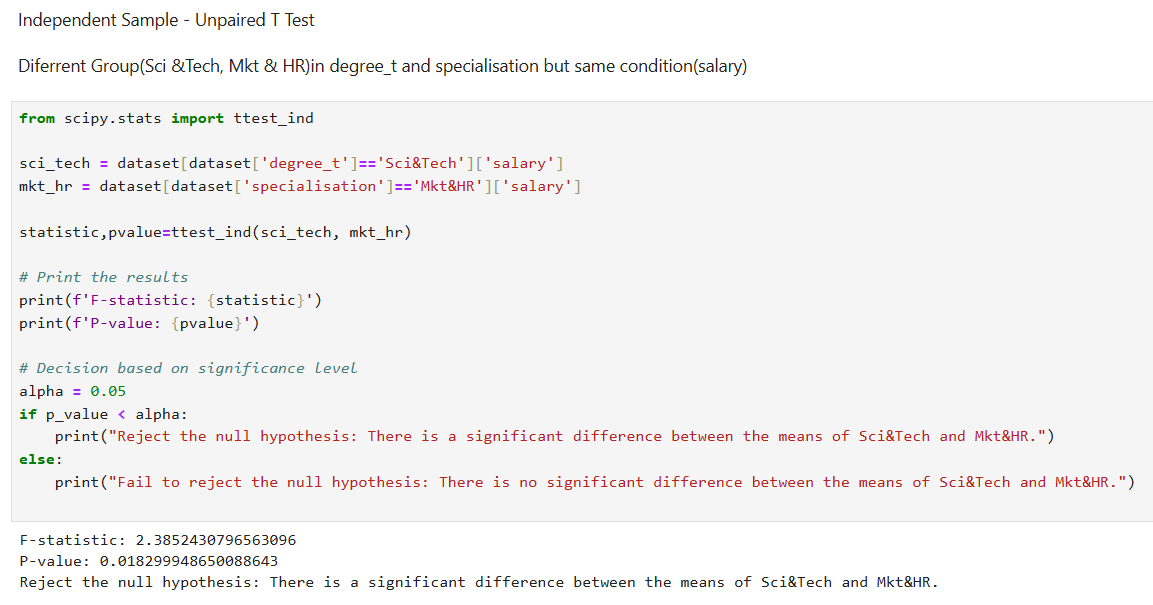
Before removing outliers we had 3 individuals getting salary above 500000.

1. **Test the Analysis of Variance between etest\_p and mba\_p at significance level 5%.(Make decisions using Hypothesis Testing).**

****

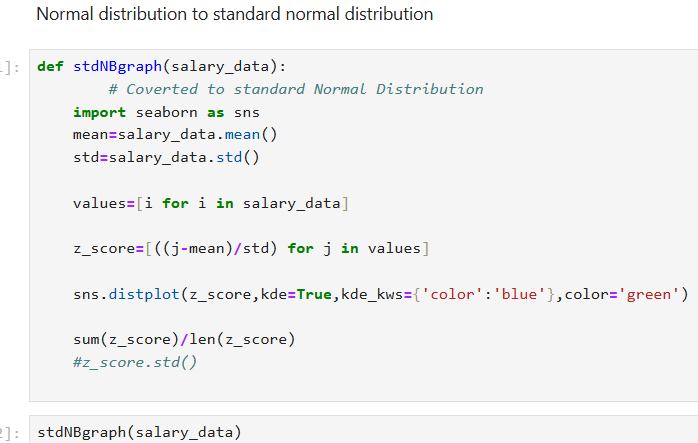
Since **4.672547689133573e-21** is much smaller than **0.05**, this indicates that the result is statistically significant. In other words, you can reject the null hypothesis with a high degree of confidence.

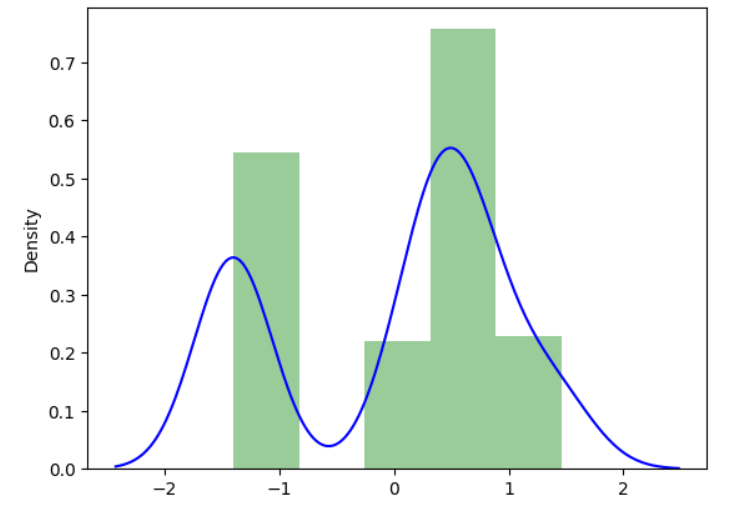
1. **Test the similarity between the degree\_t(Sci &Tech) and specialization(Mkt & HR) with respect to salary at a significance level of 5%.(Make decisions using Hypothesis Testing).**

****

From the calculation p-value is less than 0.05, so that we reject the null hypothesis. There is significance between the degree\_t(Sci &Tech) and specialisation(Mkt & HR) with respect to salary.

1. **Convert the normal distribution to standard normal distribution for the salary column.**

****

****

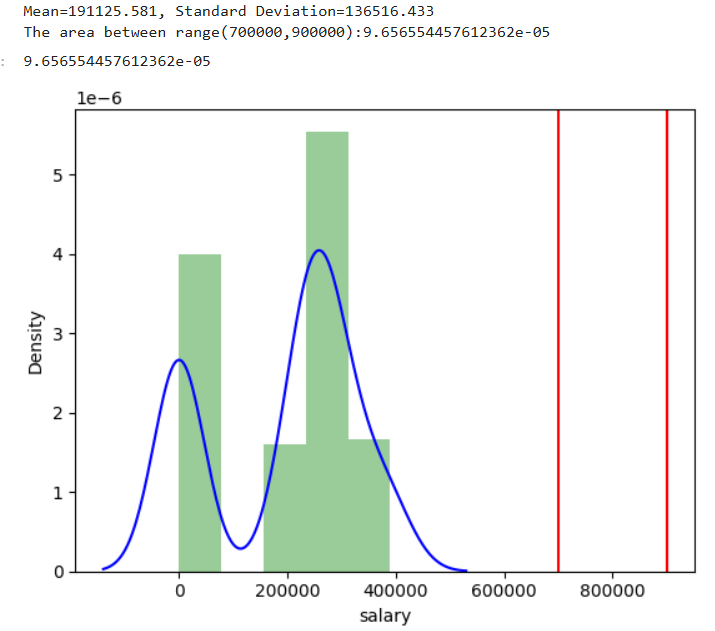
Normal distribution is converted into standard normal distribution for salary column.

𝑁(µ, σ . 2 )￫𝑁(0, 1).

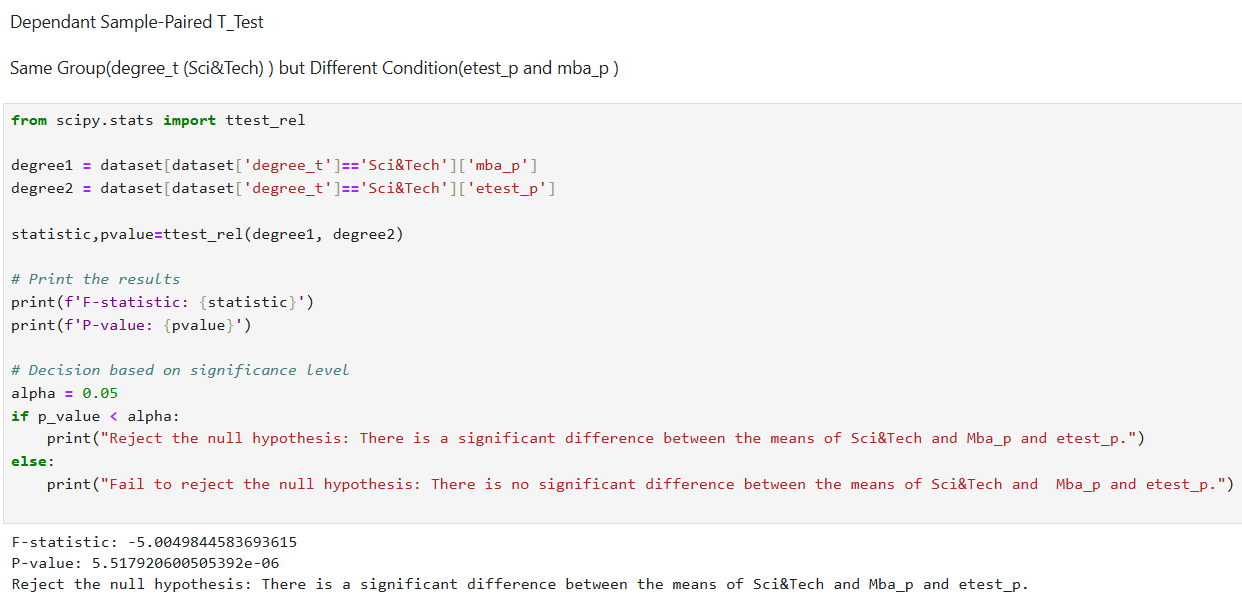
1. **What is the probability Density Function of the salary range from 700000 to 900000?**

No one is receiving the salary between the range 700000 to 900000.

****

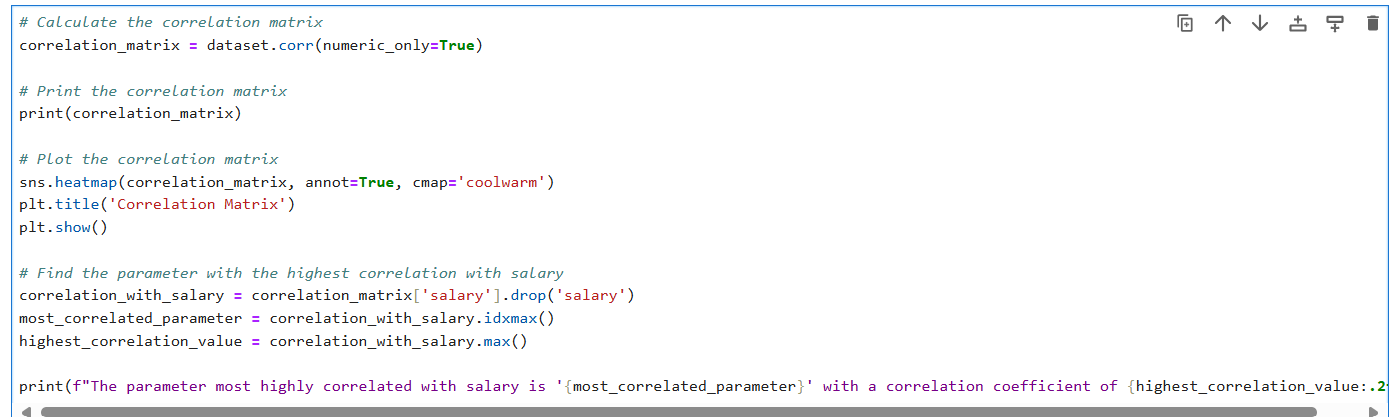
****

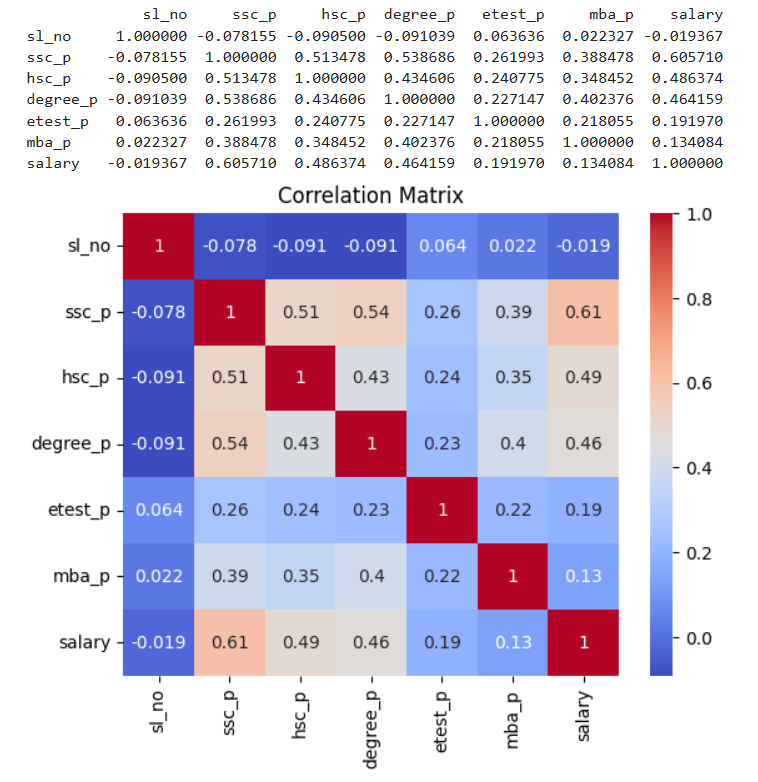
1. **Test the similarity between the degree\_t (Sci&Tech) with respect to etest\_p and mba\_p at significance level of 5%.(Make decisions using Hypothesis Testing).**

****

From the calculation p-value is less than 0.05, so that we reject the null hypothesis. There is a significant difference between the degree\_t (Sci&Tech) with respect to etest\_p and mba\_p.

1. **Which parameter is highly correlated with salary?**



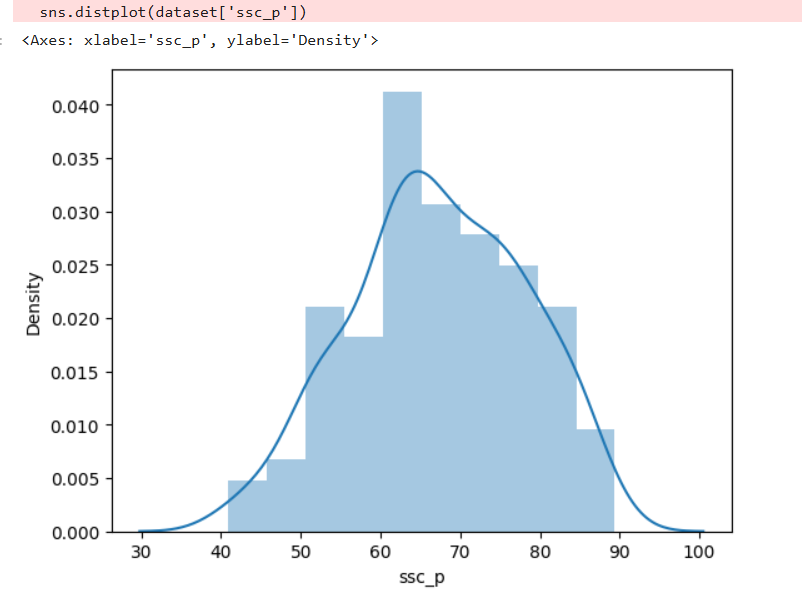
****

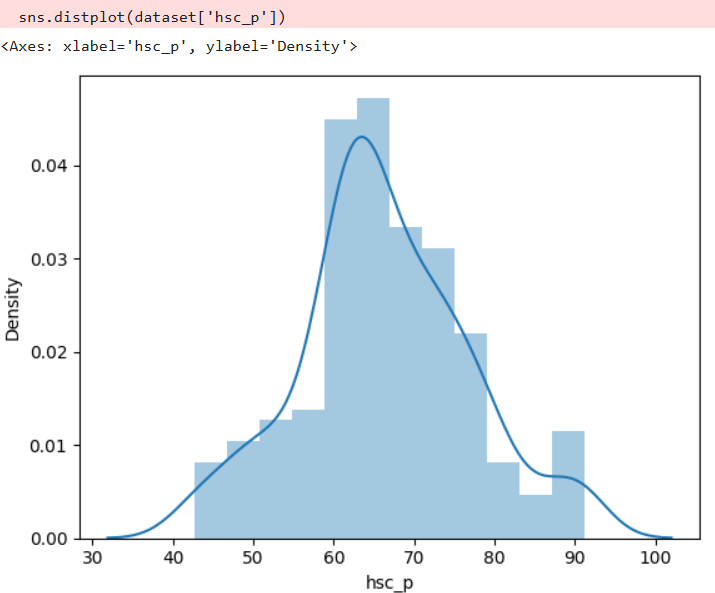
****

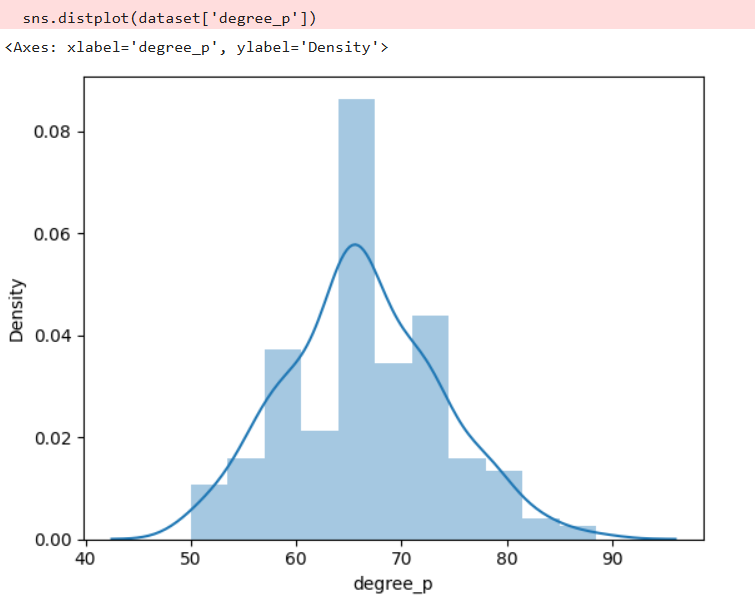
All parameters are highly correlated with salary. But the variable ssc\_p and salary have a high correlation.

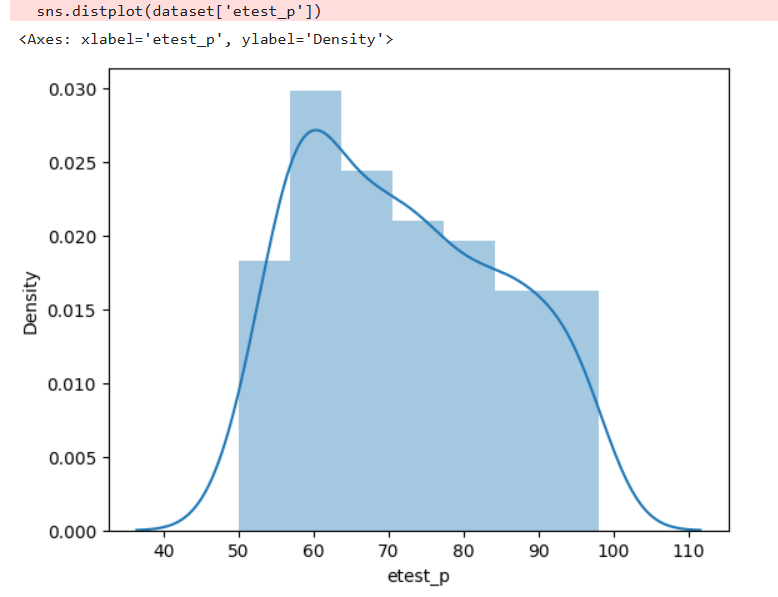
1. **plot any useful graph and explain it.**

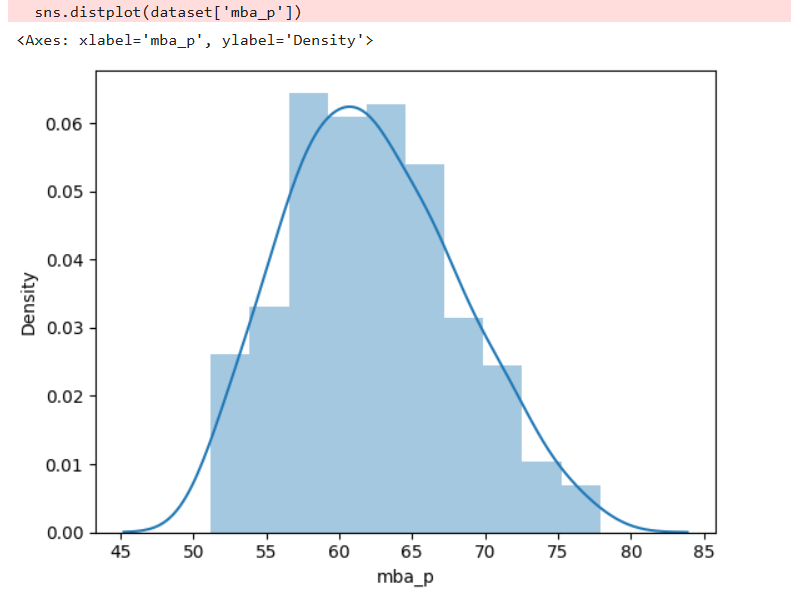
****

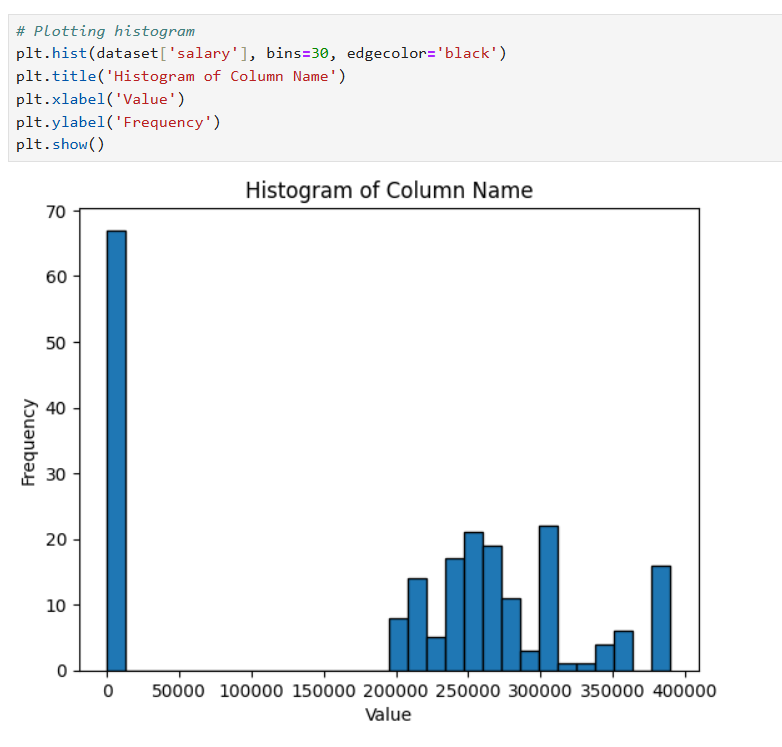
****

****

****

****

****

****