**Problem Statement or Requirement:**

**Note: Inference or Conclusion needed for every question**

**Download dataset from** [**here**](https://drive.google.com/file/d/1hiR9XbTjMsNk_uA025w5nrw5LTBUDCEC/view?usp=sharing)

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

Ans: In the given data set, the column denoted as “Salary” has 67 counts of missing values. The most obvious reason is that those 67 candidates did not receive any placements at the time the dataset was recorded. Hence, the Python code was created to replace the missing value with an integer, i.e., “0”. Besides this, there were two redundant rows. Since these two rows did not have any numerical values. They were removed completely from the data set.

2)How many of them are not placed?

Ans: There are 67 candidates who did not get any placement

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

Ans: It is plausible that the students who did not secure a placement made one of the following choices.

* Applied for higher education
* Opted to become an entrepreneur.

4)What kind of relation between salary and mba\_p?

Ans: It is noted that the correlation between MBA pass marks and salary is 0.139823 (a positive value). Though this correlation is positive, it is closer to zero. This suggests that most students who cleared the MBA were able to secure a decent salary.

5)Which specialization is getting minimum salary?

Ans: Using Python code to determine the frequency distribution of salary, it was noted that ₹. 200000 is the lowest drawn salary. However, there are six candidates who are drawing that salary. In this case, the students having specialization of both “Mkt&HR” and “Mkt&Fin” are drawing this minimum salary.

6)How many of them getting above 500000 salaries?

Ams: The percentile distribution of the variables in the data set was calculated using the Python code. The resulting table revealed that only 2 % of the students received ₹. 500000 and more as the salary.

7)Test the Analysis of Variance between etest\_p and mba\_p at signifance level 5%. (Make decision using Hypothesis Testing)

Ans: Before carrying out the ANOVA, outliers were removed from the dataset. Then, the Python code for one-way ANOVA was used. The result revealed p-value = 4.67 X 10-6. In this case, p<0.05. Hence, we can declare that the independent variables, i.e., “etest\_p” and “mba\_p”, exhibit no significant difference. As a result, the end result is reliable for the considered dataset.

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)

Ans: Python code for the T-test was used. The result revealed that the p-value = 0.7%. In this case, p<0.05. Hence, we can accept the null hypothesis and reject the alternative hypothesis. This means that the observed results are not likely to have occurred by chance alone. In other words, there is no statistically significant difference between the two independent variables in the dataset.

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

Ans: Python code was written to convert normal distribution to standard normal distribution for salary column. The resulting graph is shown below.

A graph with a blue line

AI-generated content may be incorrect.

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

Ans: The probability density function for the salary ranges from 700000 to 900000 is 0.000413.

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)

Ans: Python code for the T-test was used. The result revealed that the p-value = 5.51 X 10-6. In this case, p<0.05. Hence, we can accept the null hypothesis and reject the alternative hypothesis. This means that the observed results are not likely to have occurred by chance alone. In other words, there is no statistically significant difference between the two independent variables in the dataset.

12)Which parameter is highly correlated with salary?

Ans: By using Python code for calculating the correlation values, it is revealed that the SSLC pass marks have high correlation with the salary. In this care, the value is 0.538

13) Plot any useful graph and explain it.

Ans: Pair plots for the variables provided in the dataset were plotted using Python code from Seaborn. The obtained plots are shown below.

A screenshot of a graph

AI-generated content may be incorrect.

**submit as "YOURNAME\_InferentialAnalysis.ipynb" to admin@hopelearning.net**

Communication is important (How you are representing the document.)

Kindly mail to [admin@hopelearning.net](mailto:admin@hopelearning.net)