**Justification/Answer summary – Data Analysis Assignment.**

In the placement cell dataset, the columns except salary have taken to analyse the happenings.

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

67 entries are NaN values in salary column. These students are not get placed so their salary is NAN, it have to be replaced by 0 as they don’t get salary yet.

2) How many of them are not placed?

The status column of 67 entries are not placed.

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

Out of 67 students, 57 of them have no work experience, 42 of student who have specialization as Mkt&HR and 43 students who did comm&Mgt as degree and comparatively students who have taken greater percentages in their studies got placed. (All these analysis are in the python code)

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

As the mba percentage for placed student is 62 & no placed is 61, There is no much difference to identify the relationship.

5) Which specialization is getting minimum salary?

Students who have studied – Mkt&HR are getting the average minimum salary of Rs.270377

6) How many of them getting above 500000 salary?

3 students are getting salary above 500000

7) Test the Analysis of Variance between etest\_p and mba\_p at signifance

level 5%.(Make decision using Hypothesis Testing)

Anova is used to find one categorical and one continuous variable. Here etest\_p, mba\_p both are continuour/ numerical variables. ( please provide suggestion to handle it)

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)

The probability value is less than 5% so there is a significant difference in salary between Sci&Tech degree holders and Mkt&HR specialization.

9) Convert the normal distribution to standard normal distribution for

salary column

By using the mean & standard deviations and formula for standardization, it is calculated in program.

10) What is the probability Density Function of the salary range from

700000 to 900000?

Considering the salary is normally distributed, when there is a range, it is cumulative distribution function – the score is – 0.0006. for 7L salary, pdf is 0.00210443 & for 9L salary, the pdf is 0.00001390

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)

When comparing with etest\_p, the probability value got is more than 5 percent. So no significant difference identified.

But the probability value is less then 5 percent for mba\_p, so significant difference identified for MBA percent for science and technology students

12)Which parameter is highly correlated with salary?

Etest\_p-0.178307 is highly correlated with salary.

13) plot any useful graph and explain it.

Graph 1



1. **Boxes** represent the **interquartile range (IQR),** where the 50% of salaries comes in the middle ie.(2,60,000 approx for Mkt&HR & 2,75,000 for Mkt&Fin)
2. The **line inside the box** is the median salary.
3. Dot at the top represents the outliers.