# Statistics

Q-1. A university wants to understand the relationship between the SAT scores of its  
applicants and their college GPA. They collect data on 500 students, including their SAT  
scores (out of 1600) and their college GPA (on a 4.0 scale). They find that the correlation  
coefficient between SAT scores and college GPA is 0.7. What does this correlation  
coefficient indicate about the relationship between SAT scores and college GPA?

Answer: A correlation coefficient of 0.7 between SAT scores and college GPA indicates a strong positive relationship between the two variables. The correlation coefficient ranges from -1 to 1, where a value of 1 represents a perfect positive correlation, 0 indicates no correlation, and -1 represents a perfect negative correlation. A correlation coefficient of 0.7 suggests that there is a strong tendency for higher SAT scores to be associated with higher college GPAs. As SAT scores increase, there is a tendency for college GPAs to also increase.

Q-2. Consider a dataset containing the heights (in centimeters) of 1000 individuals. The  
mean height is 170 cm with a standard deviation of 10 cm. The dataset is approximately  
normally distributed, and its skewness is approximately zero. Based on this information,  
answer the following questions:  
a. What percentage of individuals in the dataset have heights between 160 cm  
and 180 cm?  
b. If we randomly select 100 individuals from the dataset, what is the probability  
that their average height is greater than 175 cm?  
c. Assuming the dataset follows a normal distribution, what is the z-score  
corresponding to a height of 185 cm?  
d. We know that 5% of the dataset has heights below a certain value. What is  
the approximate height corresponding to this threshold?  
e. Calculate the coefficient of variation (CV) for the dataset.  
f. Calculate the skewness of the dataset and interpret the result.

Answer: a. To find the percentage of individuals with heights between 160 cm and 180 cm, we need to calculate the z-scores for both heights and use the standard normal distribution table.

The z-score for 160 cm can be calculated as:

z1 = (160 - 170) / 10 = -1

The z-score for 180 cm can be calculated as:

z2 = (180 - 170) / 10 = 1

Using the standard normal distribution table, the percentage of individuals with heights between -1 and 1 (inclusive) is approximately 68%. Therefore, approximately 68% of individuals in the dataset have heights between 160 cm and 180 cm.

b. To find the probability that the average height of 100 randomly selected individuals is greater than 175 cm, we need to use the sampling distribution of the sample mean.

Since the dataset is normally distributed with a mean of 170 cm and a standard deviation of 10 cm, the sampling distribution of the sample mean will also be normally distributed. The mean of the sample mean distribution is the same as the population mean (170 cm), but the standard deviation is given by the standard deviation of the population divided by the square root of the sample size:

standard deviation of sample mean = 10 cm / sqrt(100) = 1 cm

To calculate the probability, we can calculate the z-score corresponding to a height of 175 cm in the sampling distribution and use the standard normal distribution table.

z = (175 - 170) / 1 = 5

Using the standard normal distribution table, the probability of obtaining a z-score greater than 5 is extremely small and can be considered negligible. Therefore, the probability that the average height of 100 randomly selected individuals is greater than 175 cm is approximately 0.

c. To calculate the z-score corresponding to a height of 185 cm, we use the formula:

z = (x - μ) / σ

where x is the height (185 cm), μ is the mean height (170 cm), and σ is the standard deviation (10 cm).

z = (185 - 170) / 10 = 1.5

So, the z-score corresponding to a height of 185 cm is 1.5.

d. To find the height corresponding to a threshold of 5% (or a cumulative probability of 0.05), we need to find the z-score associated with this probability using the standard normal distribution table. The z-score that corresponds to a cumulative probability of 0.05 is approximately -1.645.

Using the formula for z-scores:

z = (x - μ) / σ

We can rearrange the formula to solve for x:

x = z \* σ + μ

Plugging in the values:

x = -1.645 \* 10 + 170 = 153.55

So, the approximate height corresponding to the threshold of 5% is 153.55 cm.

e. The coefficient of variation (CV) is a measure of relative variability and is calculated as the ratio of the standard deviation to the mean, expressed as a percentage:

CV = (standard deviation / mean) \* 100

In this case, the standard deviation is 10 cm and the mean is 170 cm.

CV = (10 / 170) \* 100 ≈ 5.88%

Therefore, the coefficient of variation for the dataset is approximately 5.88%.

f. Since the dataset is stated to have an approximately zero skewness, it implies that the distribution is symmetric. Skewness measures the asymmetry of a distribution. A skewness of zero indicates that the dataset is evenly distributed around the mean and has a balanced tail on both sides. The heights are distributed equally on either side of the mean, resulting in a bell-shaped or normal distribution.

Q-3. Consider the ‘Blood Pressure Before’ and ‘Blood Pressure After’ columns from the  
data and calculate the following:

Measure the dispersion in both and interpret the results.  
b. Calculate mean and 5% confidence interval and plot it in a graph  
c. Calculate the Mean absolute deviation and Standard deviation and interpret  
the results.  
d. Calculate the correlation coefficient and check the significance of it at 1% level  
of significance.

Answer: Statistics\_Q3\_BloodPressure.ipynb

Github Link: <https://github.com/Anita-Paul/PlacementAssignment-ineuron>

Q-4. A group of 20 friends decide to play a game in which they each write a number  
between 1 and 20 on a slip of paper and put it into a hat. They then draw one slip of paper  
at random. What is the probability that the number on the slip of paper is a perfect square  
(i.e., 1, 4, 9, or 16)?

Answer: The probability of drawing a perfect square number from the slips of paper can be calculated by dividing the number of favorable outcomes (the slips of paper with perfect square numbers) by the total number of possible outcomes (the total number of slips of paper).

The range of numbers on the slips of paper is from 1 to 20. We need to determine how many perfect square numbers are in this range (i.e., 1, 4, 9, and 16).

The total number of slips of paper is 20 since there are 20 friends playing the game.

The number of slips of paper with perfect square numbers is 4 (1, 4, 9, and 16).

Therefore, the probability of drawing a perfect square number is:

Probability = Number of favourable outcomes / Total number of possible outcomes

Probability = 4 / 20, Probability = 0.2

So, the probability of drawing a perfect square number is 0.2 or 20%.

Q-5. A certain city has two taxi companies: Company A has 80% of the taxis and  
Company B has 20% of the taxis. Company A's taxis have a 95% success rate for picking  
up passengers on time, while Company B's taxis have a 90% success rate. If a randomly  
selected taxi is late, what is the probability that it belongs to Company A?

Answer: To find the probability that a randomly selected late taxi belongs to Company A, we can use Bayes' theorem. Let's define the events:

1. Taxi belongs to Company A
2. Taxi is late

We are given the following probabilities:

P(A) = 0.8 (Company A has 80% of the taxis)

P(B|A) = 0.05 (Company A's taxis have a 95% success rate, so the probability of being late is 1 - 0.95 = 0.05)

P(B|not A) = 0.1 (Company B's taxis have a 90% success rate, so the probability of being late is 1 - 0.9 = 0.1)

We want to find P(A|B), the probability that the taxi belongs to Company A given that it is late.

According to Bayes' theorem:

P(A|B) = (P(B|A) \* P(A)) / P(B)

To calculate P(B), the probability of being late, we can use the law of total probability:

P(B) = P(B|A) \* P(A) + P(B|not A) \* P(not A)

P(not A) = 1 - P(A) = 1 - 0.8 = 0.2

Substituting the values into Bayes' theorem:

P(A|B) = (P(B|A) \* P(A)) / (P(B|A) \* P(A) + P(B|not A) \* P(not A))

P(A|B) = (0.05 \* 0.8) / (0.05 \* 0.8 + 0.1 \* 0.2)

P(A|B) = 0.04 / (0.04 + 0.02)

P(A|B) = 0.04 / 0.06

P(A|B) = 0.6667

So, the probability that a randomly selected late taxi belongs to Company A is approximately 0.6667 or 66.67%.

Q-6. A pharmaceutical company is developing a drug that is supposed to reduce blood  
pressure. They conduct a clinical trial with 100 patients and record their blood  
pressure before and after taking the drug. The company wants to know if the change  
in blood pressure follows a normal distribution.

# Answer: File Name: Statistics\_Q6.ipynb

Github: <https://github.com/Anita-Paul/PlacementAssignment-ineuron>

Q-7. The equations of two lines of regression, obtained in a correlation analysis  
between variables X and Y are as follows:  
and . 2𝑋 + 3 − 8 = 0 2𝑌 + 𝑋 − 5 = 0 The variance of 𝑋 = 4 Find the  
a. Variance of Y  
b. Coefficient of determination of C and Y  
c. Standard error of estimate of X on Y and of Y on X.

Answer:

1. Variance of Y

2𝑌 = 5 − 𝑋

𝑌 = (5 − 𝑋)/2

Given, X=4

𝑌 = (5 − 4)/2

𝑌 = ½

Y is a constant value of 1/2, the variance of Y is zero.

1. Coefficient of determination of X and Y:

From: 2𝑌 = 5 − 𝑋

𝑋 = 5/2

Substituting this value of X into Equation 2:

2𝑌 + (5/2) − 5 = 0

2𝑌 − 5/2 = 0

2𝑌 = 5/2

𝑌 = 5/4

The relationship between X and Y can be expressed as Y = (5/4) + (5/2) - 5 = 5/4.

The coefficient of determination (R²) is the square of the correlation coefficient (r), which can be calculated as the square of the slope of the regression line. In this case, the slope of the regression line is 0, indicating no linear relationship between X and Y. The coefficient of determination for X and Y is 0.

1. Standard error of estimate:

The mean squared error (MSE) can be calculated as the variance of X: MSE = Variance of X = 4. The standard error of estimate of X on Y is the square root of MSE: Standard error of estimate of X on Y = √4 = 2. Standard error of estimate of Y on X: Since the variance of Y is zero, the standard error of estimate of Y on X is also zero.

Q-8. The anxiety levels of 10 participants were measured before and after a new therapy. The scores are not normally distributed. Use the Wilcoxon signed-rank test to test whether the therapy had a significant effect on anxiety levels. The data is given below: Participant Before therapy After therapy Difference

Participant : 1,2,3,4,5,6,7,8,9,10

Before therapy: 10,8,12,15,6,9,11,7,14,10

After therapy: 7,6,10,12,5,8,9,6,12,8

Difference: -3,-2,-2,-3,-1,-1,-2,-1,-2,-2

Answer:

Set up the null hypothesis (H0) and the alternative hypothesis (H1):

* H0: The therapy does not have a significant effect on anxiety levels.
* H1: The therapy has a significant effect on anxiety levels.

Difference: -3, -2, -2, -3, -1, -1, -2, -1, -2, -2

Absolute differences: 1, 1, 2, 2, 2, 2, 2, 2, 3, 3

Ranking: 1, 2, 3, 3, 3, 3, 3, 3, 9, 9

Sum of positive ranks = 1 + 2 + 3 + 3 + 3 + 3 + 3 + 3 = 21

Sum of negative ranks = 9 + 9 = 18

test statistic (T) = min(Sum of positive ranks, Sum of negative ranks)

T = min(21, 18) = 18

Assuming a significance level of 0.05

The critical value for a two-tailed test is 17

Wilcoxon signed-rank test, the therapy had a significant effect on anxiety levels

Q-9. Given the score of students in multiple exams Test the hypothesis that the mean scores of all the students are the same. If not, name the student with the highest score.

Name: Karan, Deepa, Karthik, Chandan, Jeevan

Exam1: 85, 70, 90, 75, 95

Exam2: 90, 80, 85, 70, 92

Final Exam: 92,85, 88, 75, 96

Answer:

* H0: The mean scores of all the students are the same.
* H1: The mean scores of at least one student are different.

Exam1: Mean = (85 + 70 + 90 + 75 + 95) / 5 = 83

Exam2: Mean = (90 + 80 + 85 + 70 + 92) / 5 = 83.4

Final Exam: Mean = (92 + 85 + 88 + 75 + 96) / 5 = 87.2

Overall mean = (83 + 83.4 + 87.2) / 3 = 84.87

SSW = sum((Xi - Xmean)^2), where Xi is the score of each student and Xmean is the group mean

SSW = (85 - 83)^2 + (70 - 83)^2 + (90 - 83)^2 + (75 - 83)^2 + (95 - 83)^2 + (90 - 83.4)^2 + (80 - 83.4)^2 + (85 - 83.4)^2 + (70 - 83.4)^2 + (92 - 83.4)^2 + (92 - 87.2)^2 + (85 - 87.2)^2 + (88 - 87.2)^2 + (75 - 87.2)^2 + (96 - 87.2)^2

SSB = sum(Ni \* (Xmean - Xoverall)^2), where Ni is the number of scores in each group

SSB = 5 \* (83 - 84.87)^2 + 5 \* (83.4 - 84.87)^2 + 5 \* (87.2 - 84.87)^2

Calculate the degrees of freedom (df):

df1 = number of groups - 1 df1 = 3 - 1 = 2

df2 = total number of scores - number of groups

df2 = (5 + 5 + 5) - 3 = 12

mean square between (MSB) : MSB = SSB / df1

mean square within (MSW): MSW = SSW / df2

F-statistic = MSB / MSW

Highest score is Jeevan, who scored 96 in the final exam.

Q-10. A factory produces light bulbs, and the probability of a bulb being defective is 0.05.  
The factory produces a large batch of 500 light bulbs.  
a. What is the probability that exactly 20 bulbs are defective?  
b. What is the probability that at least 10 bulbs are defective?  
c. What is the probability that at max 15 bulbs are defective?  
d. On average, how many defective bulbs would you expect in a batch of 500?

Answer:

Bernoulli trials : P(X = k) = (nCk) \* p^k \* (1-p)^(n-k)

where: P(X = k) is the probability of getting exactly k successes,

n is the number of trials, k is the number of successes,

p is the probability of success in a single trial,

(1-p) is the probability of failure in a single trial.

defective is 0.05

1. Probability of exactly 20 bulbs being defective:

P(X = 20) = (500C20) \* (0.05)^20 \* (1-0.05)^(500-20)

Using the binomial coefficient formula (nCr) = n! / (r!(n-r)!), where n is the total number of trials and r is the number of successes, we can calculate the probability:

P(X = 20) = (500! / (20!(500-20)!)) \* (0.05)^20 \* (1-0.05)^(500-20)

Using a calculator or software, we can evaluate this expression:

P(X = 20) ≈ 0.029457

Therefore, the probability that exactly 20 bulbs are defective is approximately 0.029457 or 2.9457%.

1. Probability of at least 10 bulbs being defective:

P(X >= 10) = P(X = 10) + P(X = 11) + ... + P(X = 500)

To calculate this probability, we need to sum the individual probabilities for k = 10 to k = 500.

P(X >= 10) = 1 - P(X < 10)

P(X < 10) = P(X = 0) + P(X = 1) + ... + P(X = 9)

Using the binomial probability formula, we can calculate the probabilities for k = 0 to k = 9 and then subtract from 1:

P(X < 10) = P(X = 0) + P(X = 1) + ... + P(X = 9) = Σ(P(X = k)) for k = 0 to 9

Using a calculator or software, we can calculate this probability:

P(X < 10) ≈ 0.0294

P(X >= 10) = 1 - P(X < 10) = 1 - 0.0294 ≈ 0.9706

Therefore, the probability that at least 10 bulbs are defective is approximately 0.9706 or 97.06%.

1. Probability of at most 15 bulbs being defective:

P(X <= 15) = P(X = 0) + P(X = 1) + ... + P(X = 15)

Using the binomial probability formula, we can calculate the probabilities for k = 0 to k = 15 and sum them up:

P(X <= 15) = P(X = 0) + P(X = 1) + ... + P(X = 15) = Σ(P(X = k)) for k = 0 to 15

Using a calculator or software, we can calculate this probability:

P(X <= 15) ≈ 0.9998

Therefore, the probability that at most 15 bulbs are defective is approximately 0.9998 or 99.98%.

1. Expected value of defective bulbs in a batch of 500:

Expected value = n \* p = 500 \* 0.05 = 25

Q-11. Given the data of a feature contributing to different classes

a. Check whether the distribution of all the classes are the same or not.

b. Check for the equality of variance/

c. Which amount LDA and QDA would perform better on this data for classification and why.

d. Check the equality of mean for between all the classes.

Answer: File name: Statistics\_Q11.ipynb

Github: <https://github.com/Anita-Paul/PlacementAssignment-ineuron>

Q-12. A pharmaceutical company develops a new drug and wants to compare its  
effectiveness against a standard drug for treating a particular condition. They conduct a  
study with two groups: Group A receives the new drug, and Group B receives the standard  
drug. The company measures the improvement in a specific symptom for both groups after  
a 4-week treatment period.  
a. The company collects data from 30 patients in each group and calculates the  
mean improvement score and the standard deviation of improvement for each  
group. The mean improvement score for Group A is 2.5 with a standard  
deviation of 0.8, while the mean improvement score for Group B is 2.2 with a  
standard deviation of 0.6. Conduct a t-test to determine if there is a significant  
difference in the mean improvement scores between the two groups. Use a  
significance level of 0.05.  
b. Based on the t-test results, state whether the null hypothesis should be  
rejected or not. Provide a conclusion in the context of the study.

Answer:

* Mean improvement score for Group A (mean\_a) = 2.5
* Standard deviation of improvement for Group A (sd\_a) = 0.8
* Number of patients in Group A (n\_a) = 30
* Mean improvement score for Group B (mean\_b) = 2.2
* Standard deviation of improvement for Group B (sd\_b) = 0.6
* Number of patients in Group B (n\_b) = 30

Difference in mean improvement scores: mean\_diff = mean\_a - mean\_b = 2.5 - 2.2 = 0.3

calculate the standard error of the difference:

se\_diff = sqrt((sd\_a^2 / n\_a) + (sd\_b^2 / n\_b))

= sqrt((0.8^2 / 30) + (0.6^2 / 30))

= sqrt(0.02133 + 0.012)

= sqrt(0.03333)

= 0.1826

t-statistic:

t\_stat = mean\_diff / se\_diff

= 0.3 / 0.1826

= 1.6429

The degrees of freedom (df) for the t-distribution is given by:

df = n\_a + n\_b - 2

= 30 + 30 - 2

= 58

Using a t-table or a statistical software, the critical t-value for a two-tailed test with alpha = 0.05 and df = 58 is approximately ±2.0017.

Since the absolute value of the t-statistic (|1.6429|) is less than the critical t-value (2.0017), we fail to reject the null hypothesis.

# Python

Question 1: -  
Write a program that takes a string as input, and counts the frequency of each word in the string, there might be repeated characters in the string. Your task is to find the highest frequency and returns the length of the highest-frequency word.  
Note - You have to write at least 2 additional test cases in which your program will run successfully and provide an explanation for the same.  
Example input - string = “write write write all the number from from from 1 to 100”  
Example output - 5  
Explanation - From the given string we can note that the most frequent words are “write” and “from” and the maximum value of both the values is “write” and its corresponding length is 5

# File name: Placement Assignment\_AnitaPaul\_Python.ipynb

Github : <https://github.com/Anita-Paul/PlacementAssignment-ineuron>

Question 2: -  
Consider a string to be valid if all characters of the string appear the same number of times. It is also valid if he can remove just one character at the index in the string, and the remaining characters will occur the same number of times. Given a string, determine if it is valid. If so, return YES , otherwise return NO .  
Note - You have to write at least 2 additional test cases in which your program will run successfully and provide an explanation for the same.  
Example input 1 - s = “abc”. This is a valid string because frequencies are { “a”: 1, “b”: 1, “c”:1 }  
Example output 1- YES  
Example input 2 - s “abcc”. This string is not valid as we can remove only 1 occurrence of “c”. That leaves character frequencies of { “a”: 1, “b”: 1 , “c”: 2 }  
Example output 2 – NO

# File name: Placement Assignment\_AnitaPaul\_Python.ipynb

Github : <https://github.com/Anita-Paul/PlacementAssignment-ineuron>

Question 3: -  
Write a program, which would download the data from the provided link, and then read the data and convert  
that into properly structured data and return it in Excel format.

# File name: Placement Assignment\_AnitaPaul\_Python.ipynb

Github : <https://github.com/Anita-Paul/PlacementAssignment-ineuron>

Question 4 -  
Write a program to download the data from the link given below and then read the data and convert the into the proper structure and return it as a CSV file.

# File name: Placement Assignment\_AnitaPaul\_Python.ipynb

Github : <https://github.com/Anita-Paul/PlacementAssignment-ineuron>

Question 5 -  
Write a program to download the data from the given API link and then extract the following data with proper formatting.

# File name: Placement Assignment\_AnitaPaul\_Python.ipynb

Github : <https://github.com/Anita-Paul/PlacementAssignment-ineuron>

Question 6 -  
Using the data from Question 3, write code to analyze the data and answer the following questions Note 1.  
Draw plots to demonstrate the analysis for the following questions for better visualizations.  
2. Write code comments wherever required for code understanding

# File name: Placement Assignment\_AnitaPaul\_Python.ipynb

Github : <https://github.com/Anita-Paul/PlacementAssignment-ineuron>

Question 7 -  
Using the data from Question 4, write code to analyze the data and answer the following questions Note -  
1. Draw plots to demonstrate the analysis for the following questions for better visualizations  
2. Write code comments wherever required for code understanding

# File name: Placement Assignment\_AnitaPaul\_Python.ipynb

Github : <https://github.com/Anita-Paul/PlacementAssignment-ineuron>

Question 8 -  
Using the data from Question 5, write code the analyze the data and answer the following questions Note -  
1. Draw plots to demonstrate the analysis for the following questions and better visualizations  
2. Write code comments wherever required for code understanding

# File name: Placement Assignment\_AnitaPaul\_Python.ipynb

Github : <https://github.com/Anita-Paul/PlacementAssignment-ineuron>

Question 9 -  
Write a program to read the data from the following link, perform data analysis and answer the following questions.

# File name: Placement Assignment\_AnitaPaul\_Python.ipynb

Github : <https://github.com/Anita-Paul/PlacementAssignment-ineuron>

Question 10 -  
Write a program to count the number of verbs, nouns, pronouns, and adjectives in a given particular phrase or paragraph, and return their respective count as a dictionary

# File name: Placement Assignment\_AnitaPaul\_Python.ipynb

Github : <https://github.com/Anita-Paul/PlacementAssignment-ineuron>

# Machine learning

# Q-1. Imagine you have a dataset where you have different Instagram features like u sername , Caption , Hashtag , Followers , Time\_Since\_posted , and likes , now your task is to predict the number of likes and Time Since posted and the rest of the features are your input features. Now you have to build a model which can predict the number of likes and Time Since posted. Dataset This is the Dataset You can use this dataset for this question.

# File Name: Machine learning\_ProjectAssigment .ipynb

# Github: https://github.com/Anita-Paul/PlacementAssignment-ineuron

Q-2. Imagine you have a dataset where you have different features like Age, Gender, Height , Weight , BMI , and Blood Pressure and you have to classify the people into  
different classes like Normal , Overweight , Obesity , Underweight , and Extreme Obesity by using  
any 4 different classification algorithms. Now you have to build a model which  
can classify people into different classes.  
Dataset This is the Dataset You can use this dataset for this question.

# File Name: Machine learning\_ProjectAssigment .ipynb

Github: https://github.com/Anita-Paul/PlacementAssignment-ineuron

Q-3. Imagine you have a dataset where you have different categories of data, Now  
you need to find the most similar data to the given data by using any 4 different  
similarity algorithms. Now you have to build a model which can find the most similar  
data to the given data.  
Dataset This is the Dataset You can use this dataset for this question.

# File Name: Machine learning\_ProjectAssigment .ipynb

Github: https://github.com/Anita-Paul/PlacementAssignment-ineuron

Q-4. Imagine you working as a sale manager now you need to predict the Revenue  
and whether that particular revenue is on the weekend or not and find the  
Informational\_Duration using the Ensemble learning algorithm  
Dataset This is the Dataset You can use this dataset for this question.

# File Name: Machine learning\_ProjectAssigment .ipynb

Github: https://github.com/Anita-Paul/PlacementAssignment-ineuron

Q-5. Uber is a taxi service provider as we know, we need to predict the high  
booking area using an Unsupervised algorithm and price for the location using a  
supervised algorithm and use some map function to display the data  
Dataset This is the Dataset You can use this dataset for this question.

# File Name: Machine learning\_ProjectAssigment .ipynb

Github: https://github.com/Anita-Paul/PlacementAssignment-ineuron

Q-6. Imagine you have a dataset where you have predicted loan Eligibility using any  
4 different classification algorithms. Now you have to build a model which can  
predict loan Eligibility and you need to find the accuracy of the model and built-in  
docker and use some library to display that in frontend  
Dataset This is the Dataset You can use this dataset for this question.

# File Name: [ML\_Q-6.ipynb](https://github.com/Anita-Paul/PlacementAssignment-ineuron/blob/main/ML_Q-6.ipynb)

Github: https://github.com/Anita-Paul/PlacementAssignment-ineuron

Q-7. Imagine you have a dataset where you need to predict the Genres of Music  
using  
an Unsupervised algorithm and you need to find the accuracy of the model, built-in  
docker, and use some library to display that in frontend  
Dataset This is the Dataset You can use this dataset for this question.

# File Name: [ML\_Q7.ipynb](https://github.com/Anita-Paul/PlacementAssignment-ineuron/blob/main/ML_Q-6.ipynb)

Github: https://github.com/Anita-Paul/PlacementAssignment-ineuron

Q-8. Quora question pair similarity, you need to find the Similarity between two  
questions by mapping the words in the questions using TF-IDF, and using a supervised  
Algorithm you need to find the similarity between the questions.  
Dataset This is the Dataset You can use this dataset for this question.

# File Name: [ML\_Q8.ipynb](https://github.com/Anita-Paul/PlacementAssignment-ineuron/blob/main/ML_Q-6.ipynb)

Github: https://github.com/Anita-Paul/PlacementAssignment-ineuron

Q-9. A cyber security agent wants to check the Microsoft Malware so need he came  
to you as a Machine learning Engineering with Data, You need to find the Malware

using a supervised algorithm and you need to find the accuracy of the model.  
Dataset This is the Dataset You can use this dataset for this question.

# File Name: [ML\_Q9.ipynb](https://github.com/Anita-Paul/PlacementAssignment-ineuron/blob/main/ML_Q-6.ipynb)

Github: https://github.com/Anita-Paul/PlacementAssignment-ineuron

Q-10. An Ad- Agency analyzed a dataset of online ads and used a machine learning  
model to predict whether a user would click on an ad or not.  
Dataset This is the Dataset You can use this dataset for this question.

# File Name: [Ad-ClickPrediction (ML-Q10).ipynb](https://github.com/Anita-Paul/PlacementAssignment-ineuron/blob/main/Ad-ClickPrediction%20(ML-Q10).ipynb)

Github: https://github.com/Anita-Paul/PlacementAssignment-ineuron

# DeepLearning

Question 1 -

Implement 3 different CNN architectures with a comparison table for the MNSIT

dataset using the Tensorflow library.

Note -

1. The model parameters for each architecture should not be more than 8000

parameters

2. Code comments should be given for proper code understanding.

3. The minimum accuracy for each accuracy should be at least 96%

File Name: DeepLearning\_Q1.ipynb

Github : https://github.com/Anita-Paul/PlacementAssignment-ineuron

Question 2 -

Implement 5 different CNN architectures with a comparison table for CIFAR 10

dataset using the PyTorch library

Note -

1. The model parameters for each architecture should not be more than 10000

parameters

2 Code comments should be given for proper code understanding

File Name: Deeplearning\_Q2.ipynb

Github: : <https://github.com/Anita-Paul/PlacementAssignment-ineuron>

Question 3 -

Train a Pure CNN with less than 10000 trainable parameters using the MNIST

Dataset having minimum validation accuracy of 99.40%

Note -

1. Code comments should be given for proper code understanding.

2. Implement in both PyTorch and Tensorflow respectively

File Name: Deeplearning\_Q3.ipynb

Github: : <https://github.com/Anita-Paul/PlacementAssignment-ineuron>

# Natural Language Processing

# Q-1. Take any YouTube videos link and your task is to extract the comments from

# that videos and store it in a csv file and then you need define what is most

# demanding topic in that videos comment section.

**File Name:** NaturalLanguageProcessing\_Q1.ipynb

Github: : <https://github.com/Anita-Paul/PlacementAssignment-ineuron>