**SET- A**

1. Why do we need to go for ANOVA when we have T-test to compare the means?
2. How logistic regression is different from linear regression?
3. Why do we need to go for Standard Deviation when Variance is able to explain the variation ?
4. What is Post-hoc analysis, How does it helps in real-time?
5. What is the difference between Average and Mean/Median/Mode?
6. Can you share the real time examples of Mode and Median ?
7. Which type of error is dangerous when it comes to business and why explain with example?

Type1 - False positive

Type2 - False negative

1. What % of observations fall within 1 standard deviation of mean in normal distribution?
2. What is OOB error in random forest?
3. What does Imbalanced Data/Uneven class mean?
4. Explain Variance / Bias / Noise ?
5. What is non-linearity ?
6. What is Random is random-forest ?
7. How to identify whether Logistic or Discriminant analysis is good?
8. Write a pyton code given a number 'N', print all the numbers 0 to N, if the number is divisible by 2 print 'FUZZ', is divisible by 3 print 'BUZZ' if divisible by both then print 'FUZZ-BUZZ".
9. Parse people from the input people and get below output, write a pyton code

People = [{"name": "Alice", "age": 30, "occupation":"engineer"},

{"name": "Bob", "age": 28, "occupation":"doctor"},

{"name": "Charlie", "age": 35, "occupation":"engineer"},

{"name": "David", "age": 22, "occupation":"doctor"},

{"name": "Eve", "age": 29, "occupation":"artist"}]

Expected output: {'engineer': ['Alice', 'Charlie'], 'doctor': ['Bob', 'David'],'artist': ['Eve']}

1. Write python code to extract domain name from below emails

['abd@gmail.com','she@yahoo.com','am@mydomain.com','your@yourdomain.com','them@aol.in','Theyare@aol.in','he@gmail.com']

get only unique domains as a LIST output

1. Write a python code from the nested list, create a list with unique values without

spaces in it

[['laptop', 'desktop'], ['laptop'], ['laptop', 'desktop', ' mouse'], ['laptop',

'desktop'], ['laptop'], ['laptop', 'desktop'], ['pen', ' laptop'], ['pen', ' laptop'],

['laptop', ' desktop', ' mouse', 'pen', ' keyboard'], ['pen', ' laptop'], ['laptop', '

desktop', ' mouse', 'pen', ' keyboard'], ['laptop', ' keyboard'], ['laptop'], ['laptop',

' desktop', ' mouse', 'pen', ' keyboard'], ['laptop', 'desktop'], ['laptop', ' desktop',

' mouse', 'pen', ' keyboard'], ['laptop', 'desktop'], ['laptop', 'desktop'], ['laptop',

' desktop', ' mouse', 'pen', ' keyboard']]

1. Write a SQL code to join both table A and B, print the resulting output with all fileds in both the tables

Table A:

Key A1 A2 A3

101 0.23 KA 1998

102 1.24 PB 1999

103 3.21 RJ 2023

NULL 3.4 ER 2345

Table B:

Key B1 B2 B3

101 2 RJ 298

105 1 TN 100

102 1 PB 1999

102 2 AP 19

105 1 TN 100

**SET- B**

1. How does the p values of each feature is calculated in linear regression?
2. What is the difference between Mode and Median ? State the case where the median is a better measure when compared to the mean.
3. What is the difference between Variance, Co-efficient of Variation and Covariance ?
4. How can we figure out if precision or recall is more important? Can you share a few examples to illustrate?
5. What is the The mean of Z-score in standard normal distribution?
6. How does a tree deicide where to split or which variable to use in root node?
7. What Are the Main Measures of Variability? Explain them
8. What is the Difference Between Probability and Likelihood?
9. How to make sure you are not overfitting while training a model?
10. What is the difference between type I vs type II errors?
11. What are quantitative data and qualitative data?
12. How to convert normal distribution to standard normal distribution?
13. How are the trees trained in random forest?
14. What is bootstrap sample?
15. What is the relationship between standard deviation and standard variance?
16. You have a DataFrame employee\_df containing information about employees. It has columns 'Name', 'Salary', and 'Experience'. You want to give a 10% raise to employees who have more than 5 years of experience and update their salaries accordingly using a pandas function.

data = {'Name': ['Alice', 'Bob', 'Charlie', 'David', 'Eva'],'Salary': [60000, 75000, 80000, 55000, 70000], 'Experience': [3, 7, 4, 9, 2]}

1. Write python code to extract domain extension from below emails

['abd@gmail.com','she@yahoo.com','am@mydomain.com','your@yourdomain.com','them@aol.in','Theyare@aol.in','he@gmail.com']

get only unique extension as a LIST output

1. In a class, there is a dictionary containing student names and their corresponding scores. Write a Python expression that generates a new dictionary by filtering out students who scored less than 30 and includes only the names of those students along with their scores.

student\_data = {

"Alice": 85, "Bob": 92, "Charlie": 78, "David": 95,

"Ella": 45, "Frank": 60, "Grace": 20, "Henry": 75,

"Isabel": 90, "Jack": 28, "Kate": 62, "Liam": 42,

"Mia": 75, "Nathan": 15, "Olivia": 80, "Peter": 38,

"Quincy": 50, "Rachel": 70, "Samuel": 85, "Tara": 10,

"Uma": 65, "Victor": 48, "Wendy": 55, "Xander": 70, "Yara": 30

}

1. Write the results of Inner join, Left join for below tables?

TABLE A TABLE B

ID ID

1 1

1. 1

1

1. Parse people from the input 'people' and get the following output, write a Python code:

People = [

{"name": "Alice", "age": 30, "department": "engineering"},

{"name": "Bob", "age": 28, "department": "medical"},

{"name": "Charlie", "age": 35, "department": "engineering"},

{"name": "David", "age": 22, "department": "medical"},

{"name": "Eve", "age": 29, "department": "arts"}]

Expected output: { engineering: ['Alice', 'Charlie'], medical: ['Bob', 'David'], arts: ['Eve']}