Scenario Based Question

Topic: SQL

Q1. At St. Xavier's College, a Faculty has the following data in My SQL in database named as Class having table student related to Semester Examination

Enrollment No.	Student Name	Section	Subject Id	Marks
1	Tim	A	1	70
2	Jim	A	2	75
3	Kim	B	3	65
4	Tom	B	4	77
5	John	C	5	60
6	Joe	C	1	82
7	James	B	2	76
8	Henry	C	5	68
9	Matt	B	3	71
10	Paul	A	4	79

The faculty needs a section-wise Number of candidates who have secured more than or equal to 75 marks in the Semester Exam.

Note: Enrollment No. is declared as Primary Key

Output Table

Section	No. of Candidate greater than or equal to 75 marks
A	3
B	4
C	3

Can you suggest how this can be achieved? Write steps along with output screenshot

Ans - Code snippets for the same in MySQL can be written as:

$Select\ Section, count(*)\ as\ 'No.of\ Candidate\ greater\ than\ or\ equal\ to\ 75\ Marks'\ from\ Student$

$where\ Marks\ >=\ 75$

$group\ by\ Section;$

	Section	No. of Candidate greater than or equal to 75 Marks
▶	A	2
	B	2
	C	1

Figure 1: Output Table SQL

Topic: Tableau

Arun has the following data of Employees in CSV format

```
Emp_name, Id,Salary
Ravish,10,1000
Suresh,101,20000
Priya,1010,50000
Neha,10101,70000
Nitin,1101,15000
```

Arun's Client wants to make all Employee Id (Id) a 7-digit number in Tableau.

For Example, the updated Employee Id of Priya should be 0001010. Can you suggest any way how Arun can achieve this for all Employees in Tableau? Write steps along with output screenshot

Ans - We can use left function where we want to add leading 0's. We can create calculated field with formula - $LEFT('0000000', (7 - LEN(STR([Id]))))) + STR([Id])$

Emp name	Employee ID	Salary
Neha	0010101	70,000
Nitin	0001101	15,000
Priya	0001010	50,000
Ravish	0000010	1,000
Suresh	0000101	20,000

Figure 2: Output Table Tableau

Topic: Excel

Sneha is an MIS executive her boss provided her a list of customers in Excel as shown below

Customer Name	Duplicate Name
Kapil khatri	
Arti Ahuja	
Eshank sharma	
Amit kumar	
Kapil khatri	
Raj Sharma	
Sunil Yadav	
Eshank sharma	
Swati Singh	
Animesh verma	
Mohit Jain	
Arti Ahuja	
Ashutosh Mahajan	
Akshay Rathod	
Harmeet kaur	
Amit kumar	

Now her boss wants to populate the **Duplicate name** field for names that occurs more than once. The required table is shown below

Customer Name	Duplicate Name
Kapil khatri	Kapil khatri
Arti Ahuja	Arti Ahuja
Eshank sharma	Eshank sharma
Amit kumar	Amit kumar
Kapil khatri	Kapil khatri
Raj Sharma	
Sunil Yadav	
Eshank sharma	Eshank sharma
Swati Singh	
Animesh verma	
Mohit Jain	
Arti Ahuja	Arti Ahuja
Ashutosh Mahajan	
Akshay Rathod	
Harmeet kaur	
Amit kumar	Amit kumar

Kindly suggest any solution for this using Microsoft Excel. You can write steps and logic

Step 1. We can find whether name is duplicated or not using $= COUNTIF(A:A, A2) > 1$

Step 2. Select the previous column and check if it is true or false.

   a. If True, mention the same name
   b. Or mention it blank

The formula used for the same is $= If(countif(A:A, A2) > 1 = True, A2, "")$

Topic: Machine Learning

Description

You need to classify fetal health to avoid the abnormalities for the child and mother while giving birth. You will be provided with a CSV consisting of 22 columns out of which the first 21 columns are the factors based on which you need to classify the fetal health which is the last column . You need to classify the health status into the following:

```
1.0-Normal
2.0-Suspect
3.0-Pathological
```

The columns in the CSV are:

a) baseline value (Baseline Fetal Heart Rate)
b) accelerations (Number of accelerations per second)
c) fetal_movement (Number of fetal movements per second)
d) uterine_contractions (Number of uterine contractions per second)
e) light_decelerations (Number of LDs per second)
f) severe_decelerations (Number of SDs per second)
g) prolongued_decelerations (Number of PDs per second)
h) abnormal_short_term_variability (Percentage of time with abnormal short term variability)

i) mean_value_of_short_term_variability
j) percentage_of_time_with_abnormal_long_term_variability
k) mean_value_of_long_term_variability
l) histogram_width (Width of the histogram made using all values from a record)
m) histogram_min
n) histogram_max
o) histogram_number_of_peaks
p) histogram_number_of_zeroes
q) histogram_mode
r) histogram_mean
s) histogram_median
t) histogram_variance
u) histogram_tendency
v) fetal_health

Training data set:
This data set will be available to the developers for training their model.
https://edubridgeindiain-my.sharepoint.com/:u:/g/personal/gaurav_hajela_edubridgeindia_in/EV0nE49gERFCkHLLudlWPlYBirh37eXNifwzNsC05PknwQ?e=SPtLoT

Validation data set:
This data set will be available to the developers for validating their model before they submit for evaluation.
https://edubridgeindiain-my.sharepoint.com/:u:/g/personal/gaurav_hajela_edubridgeindia_in/ETJfxPmlLr9OhyBt2AErfNMBrg0eS_gH_DBioV_Gu8mg5g?e=ZDwJlk

Test data set:
https://edubridgeindiain-my.sharepoint.com/:u:/g/personal/gaurav_hajela_edubridgeindia_in/ETglAwPr5HVAgHuyhPPxT5kBPPQzmYmT5p1KORTHnH2pcw?e=506dvr

Sample Output:
The output is given to show the format in which the output is expected (with header) and has no relation with above.

```
fetal_health
1.0
1.0
2.0
1.0
3.0
```

Create a CSV file with the header as “fetal_health”.

Evaluation:
Your solution will be accepted only if the accuracy of the prediction is greater than 75%.