Question 1

You are an analyst in the customer experience department of a public agency that is funding four organisations. Each organisation offers three different services run by three different departments (i.e., each department is responsible for one service). Assume that department 1 in each organisation offers the same service. This makes comparing the relative performance of each organisation's "department 1" meaningful. This is true for the other two departments.

Your agency contracted an external market research firm to conduct a survey of the users of these three services across the four organisations. At the end of the data collection period, a total of 4,858 users completed the survey.

Figure 1 shows the first few rows of the dataset you received from the market research firm.

org	dept	urgency	subsidy	age	org_date	gender	coordinated	similarity_to_ideal	willingness_to_recommend
Org A	Dept 1	999	2	22	20689	1	4	8	2
Org A	Dept 1	999	1	68	20689	1	4	3	1
Org A	Dept 1	999	2	59	20689	1	4	8	4
Org A	Dept 1	999	2	51	20689	1	4	8	3
Org A	Dept 1	999	2	64	20689	2	4	8	3
Org A	Dept 1	999	1	26	20689	1	4	9	3
Org A	Dept 1	999	1	63	20689	2	3	7	3
Org A	Dept 1	999	1	26	20689	1	4	8	3
Org A	Dept 1	999	2	39	20689	1	4	8	3
Org A	Dept 1	999	2	50	20689	1	4	9	4

Figure 1: The first few rows of the raw dataset from the market research firm.

A data dictionary describing the values of each variable is shown in Table 1 below.

Variable Name	
`org`	**Description**: *Name of organisation* **Valid Values**: '`Org A`' … '`Org D`'
`dept`	**Description**: *Name of department* **Valid Values**: '`Dept 1`' … '`Dept 3`'
`urgency`	**Description**: *Priority level (P1 = highest, P4 = lowest, PX = special)* **Valid Values**: `1`=P1; `2`=P2; `3`=P3; `4`=P4; `5`=PX; `999`=NA
`subsidy`	**Description**: *Whether cost of service for a particular user was subsidised* **Valid Values**: `1`=Y; `2`=N
`age`	**Description**: *Age (in years)* **Valid Values**: Any integer greater than zero

`org_date`	**Description**: *Number of days since 9 Aug 1965* **Valid Values**: Any integer greater than zero
`gender`	**Description**: *Gender* **Valid Values**: `1`=M; `2`=F
`coordinated`	**Description**: *How coordinated the different parts of the organisation are* **Valid Values**: `1`=Never; `2`=Sometimes; `3`=Usually; `4`=Always; `5`=Unsure
`similarity_to_ideal`	**Description**: *How similar the user's experience of the department is to the user's imagined ideal* **Valid Values**: `0-10`; `11`=Unsure; `999`=NA
`willingness_to_recommend`	**Description**: *Would the user say positive things about the organisation* **Valid Values**: `1`=Definitely No; `2`=Probably No; `3`=Probably Yes; `4`=Definitely Yes; `5`=Unsure; `999`=NA

Table 1: A data dictionary describing the values for each variable in the dataset.

(a) Import the dataset (`TMA_data.sql`) into your MySQL server.

Create a new table named `TMA_data_labelled` that is based on the `TMA_data` table you imported, but with the numeric values replaced with their corresponding text labels.

In addition to replacing the numeric values with their corresponding text labels, you will also be required to convert the `org_date` variable into a calendar date variable named `caldate` that uses the MySQL DATE type (e.g., `2022-04-01` instead of `20689`). So in the new table, you should have a `caldate` column instead of the `org_date` column.

If necessary, use `ALTER TABLE` to modify the data type (and <u>not</u> the data values) for each field in the `TMA_data_labelled` table to save storage space while not losing any information. Include a screen capture of the result of running the "`DESCRIBE TMA_data_labelled;`" MySQL statement.

The first few rows of your new table should look similar to the screenshot in Figure 2.

(80 marks)

org	dept	urgency	subsidy	age	caldate	gender	coordinated	similarity_to_ideal	willingness_to_recommend
Org A	Dept 1	NULL	N	22	2022-04-01	M	Always	8	Probably No
Org A	Dept 1	NULL	Y	68	2022-04-01	M	Always	3	Definitely No
Org A	Dept 1	NULL	N	59	2022-04-01	M	Always	8	Definitely Yes
Org A	Dept 1	NULL	N	51	2022-04-01	M	Always	8	Probably Yes
Org A	Dept 1	NULL	N	64	2022-04-01	F	Always	8	Probably Yes
Org A	Dept 1	NULL	Y	26	2022-04-01	M	Always	9	Probably Yes
Org A	Dept 1	NULL	Y	63	2022-04-01	F	Usually	7	Probably Yes
Org A	Dept 1	NULL	Y	26	2022-04-01	M	Always	8	Probably Yes
Org A	Dept 1	NULL	N	39	2022-04-01	M	Always	8	Probably Yes
Org A	Dept 1	NULL	N	50	2022-04-01	M	Always	9	Definitely Yes

Figure 2: The first few rows of the TMA_data_labelled table.

(b) Use appropriate MySQL statements (e.g., `SELECT ... GROUP BY ...`) to generate two (2) summary tables that help explain the relationship amongst the variables in the survey. Each summary table should be accompanied with a short sentence that explains what you would like the reader to learn from the table.

(20 marks)

---- END OF ASSIGNMENT ----