(Full marks: 100)

Answer All Questions (100 marks)
Answer all questions in this section.

Question 1

You are an analyst for a market research firm and have been given access to a CSV file ("`ECA_data_raw.csv`") containing a dataset. This dataset contains 6,095 responses to a survey conducted on behalf of a new shopping mall that opened at the start of 2024.

Question 1a

Create a Python program to access the dataset in the CSV file and store it as a MySQL table named `ECA_data_raw` in your MySQL database.
Your imported table should look similar to the screenshot in Figure 1. A description of each field is contained in Table 1.

doi	satis	confirm	ideal	comp	handle	nocomp
2024-04-08	4	3	3	0	NULL	2
2024-03-07	7	8	9	0	NULL	1
40	9	10	9	0	NULL	1
1	7	8	8	0	NULL	1
249	8	8	8	0	NULL	1
2024-05-27	8	8	8	0	NULL	1
356	8	8	8	0	NULL	1
2024-08-30	8	5	5	0	NULL	1
145	7	3	2	0	NULL	1
2024-12-30	10	9	9	0	NULL	1
290	7	6	8	0	NULL	1
2024-12-24	7	7	8	0	NULL	1
2024-07-11	9	8	7	0	NULL	1
250	6	6	5	1	5	NULL
311	9	8	8	0	NULL	1

Figure 1: A screenshot of the first few rows of the `ECA_data_raw` table.

Column	Description
`doi`	Date of interview recorded in two different formats, (1) calendar date (YYYY-MM-DD); or (2) Number of days since 1 Jan 2024.
`satis`	The respondent's overall satisfaction with the mall on a 1 to 10 scale, where 1 means 'very dissatisfied' and 10 means 'very satisfied'.
`confirm`	How well the respondent's expectations were met on a 1 to 10 scale, where 1 means 'falls short of their expectations' and 10 means 'exceeds their expectations'.

`ideal`	How similar the mall is to the respondent's imagined ideal mall on a 1 to 10 scale, where 1 means 'not very close to their ideal' and 10 means 'very close to their ideal'.
`comp`	Whether the respondent lodged a complaint to the mall in the last 3 months, where 1=Yes and 0=No.
`handle`	For respondents who complained, handle records the respondent's assessment of how well their complaint was handled on a 1 to 10 scale, where 1 means 'handled very poorly' and 10 means 'handled very well'. For respondents who did not complain, handle will be set to NULL.
`nocomp`	For respondents who did not complain, nocomp records the respondent's reason for not complaining. 1=No reason; 2=Too difficult; 3=No point; 4=Other reason. For respondents who complained, nocomp will be set to NULL.

Table 1: A description of each field in the `ECA_data_raw` table.

(10 marks)

Question 1b

Using a single MySQL statement, create a new table named `ECA_data` that contains all the columns from `ECA_data_raw` except that the `doi` column values have been updated to consistently use the `YYYY-MM-DD` format for all rows.

(15 marks)

Question 1c

Assess the data types used in the `ECA_data` table, and use a single `ALTER TABLE` statement to change them (where appropriate) to save storage space while not losing any information. If, in your assessment, no change is required to the data type of any of the columns, please state this (with appropriate justification) as your answer to this part.

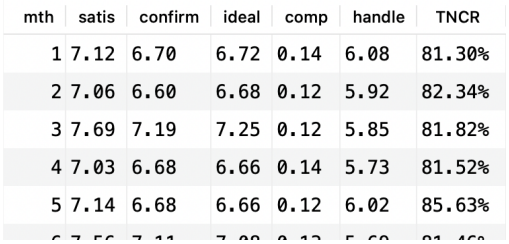
(5 marks)

Question 1d

Using a single MySQL statement, create a new table named `ECA_summary` that contains 12 rows, one for each month of the year, and 7 columns, namely `mth`, `satis`, `confirm`, `ideal`, `comp`, `handle`, `TNCR`. The `mth` column contains an integer representing the month of the year. The `satis`, `confirm`, `ideal`, `comp`, `handle` columns contain the corresponding month average values for those variables. The True Non-Complaint Rate column, `TNCR`, contains the percentage of respondents

in that month who did not complain because there was no reason to complain.

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



mth	satis	confirm	ideal	comp	handle	TNCR
1	7.12	6.70	6.72	0.14	6.08	81.30%
2	7.06	6.60	6.68	0.12	5.92	82.34%
3	7.69	7.19	7.25	0.12	5.85	81.82%
4	7.03	6.68	6.66	0.14	5.73	81.52%
5	7.14	6.68	6.66	0.12	6.02	85.63%

Figure 2: A screenshot of the first few rows of the `ECA_summary` table.

(40 marks)

Question 1e

Create an R program to read in the `ECA_summary` MySQL table and store it as an R data.frame. Create a single visual that would make it easy for the customer experience manager of the mall to understand the information in the table. (Assume that the manager is already familiar with the definitions of the variables.)

In addition to the code you used to create the visual, you should also include a screenshot of the visual as well as a writeup (in 250 words or less) to justify your visualisation decisions.

(30 marks)

----- END OF ECA PAPER -----