SINGAPORE UNIVERSITY OF SOCIAL SCIENCES

ANL503
Data Wrangling

Tutor-Marked Assignment

July 2024 Semester

TUTOR-MARKED ASSIGNMENT (TMA)

This assignment is worth 30% of the final mark for ANL503 Data Wrangling.

The cut-off date for this assignment is **Wednesday 23 October 2024 at 2355hrs**.

Note to Students:

Compose your report using Microsoft Office Word, and save either as **.docx.**

You are to include the following particulars in your submission: Course Code, Title of the TMA, SUSS PI No., Your Name, and Submission Date.

Use of Generative AI Tools (Allowed)

The use of generative AI tools is allowed for this assignment.

- You are expected to provide proper attribution if you use generative AI tools while completing the assignment, including appropriate and discipline-specific citation, a table detailing the name of the AI tool used, the approach to using the tool (e.g. what prompts were used), the full output provided by the tool, and which part of the output was adapted for the assignment;

- To take note of section 3, paragraph 3.2 and section 5.2, paragraph 2A.1 (Viva Voce) of the Student Handbook;

- The University has the right to exercise the viva voce option to determine the authorship of a student's submission should there be reasonable grounds to suspect that the submission may not be fully the student's own work.

- For more details on academic integrity and guidance on responsible use of generative AI tools in assignments, please refer to the TLC website for more details;

- The University will continue to review the use of generative AI tools based on feedback and in light of developments in AI and related technologies.

Question 1

You are an analytics consultant for a recently launched company. Your company is on a hiring spree and has been working with three different recruitment firms to fill the available positions. The breakdown of the performance of each recruitment firm across three company locations (i.e., Singapore, Hong Kong, Tokyo) is detailed in Figure 1 below.

			Number of Offers Made				Number of Offers Accepted			
Location	Department	Headcount Available	Recruitment Firm 1	Recruitment Firm 2	Recruitment Firm 3	Total	Recruitment Firm 1	Recruitment Firm 2	Recruitment Firm 3	Total
Singapore	IT Systems	135	183|1|9	32|2|0	39|0|0	**254|3|9**	67|0|4	22|1|0	20|0|0	**109|1|4**
Singapore	Corporate Services	130	205|6|7	51|0|0	43|1|0	**299|7|7**	119|2|2	40|1|0	33|2|0	**192|5|2**
Singapore	Customer Service	170	235|6|7	26|0|0	14|1|0	**275|2|2**	121|2|1	24|1|0	6|2|0	**151|5|1**
Singapore	Operations	118	99|8|2	27|2|0	14|0|0	**140|10|2**	65|3|1	23|1|0	9|0|0	**97|4|1**
Singapore	Customer Support	169	167|4|2	19|1|0	19|1|0	**205|6|2**	115|2|1	17|1|0	14|0|0	**146|3|1**
Singapore	*Total*	**722**	**889|20|22**	**155|5|0**	**129|3|0**	**1173|28|21**	**487|9|9**	**126|5|0**	**82|4|0**	**695|18|9**
Hong Kong	IT Systems	125	139|5|1	0|0|0	2|1|0	**141|6|1**	107|5|2	0|0|0	2|1|0	**109|6|2**
Hong Kong	Corporate Services	125	122|4|2	12|0|0	6|0|0	**140|4|2**	66|2|1	11|0|0	4|1|0	**81|3|1**
Hong Kong	Customer Service	120	131|35|2	17|1|0	10|2|0	**158|38|2**	87|26|1	17|1|0	7|2|0	**111|29|1**
Hong Kong	Operations	90	125|5|0	10|0|0	5|1|0	**140|6|0**	76|3|0	8|0|0	4|0|0	**88|3|0**
Hong Kong	Customer Support	100	58|1|8	4|0|0	3|0|0	**65|1|8**	37|0|5	3|0|0	2|0|0	**42|0|5**
Hong Kong	*Total*	**560**	**575|50|13**	**43|1|0**	**26|4|0**	**644|55|13**	**373|36|9**	**39|1|0**	**19|4|0**	**431|41|10**
Tokyo	Customer Service	170	269|7|5	42|0|0	29|1|0	**340|8|5**	129|5|2	30|0|0	12|1|0	**171|6|2**
Tokyo	Customer Support	110	149|6|7	39|5|0	38|0|0	**226|11|7**	57|2|4	21|1|0	9|0|0	**87|3|4**
Tokyo	*Total*	**280**	**418|13|12**	**81|5|0**	**67|1|0**	**566|19|12**	**186|7|6**	**51|1|0**	**21|1|0**	**258|9|6**
Overall	*Total*	**1562**	**1882**	**279**	**222**	**2383**	**1046**	**216**	**122**	**1384**

Figure 1: *Performance of the recruitment firms.*

Note: entries with the pattern "xx\|yy\|zz" are interpreted as follows: xx = total number of offers made/accepted; yy = number of offers for senior positions made/accepted; zz = number of offers for inclusive positions made/accepted.

The dataset without the "Total" rows is available as a MySQL dump file (`TMA_data.sql`).

(a) Import the dataset (`TMA_data.sql`) into your MySQL server. Check the resulting `TMA_data` MySQL table and evaluate the suitability of the data type for each field. Use the `ALTER TABLE` command to change them (where appropriate) to save storage space while not losing any information, and/or to change them to a data type more suited to their purpose. If, in your assessment, no change is required to the data types, please state this as your answer to this part and justify why.

(20 marks)

(b) Using appropriate MySQL statements, recompute all the totals columns in the `TMA_data` table, and update the table with these recomputed values.

(20 marks)

(c) Using appropriate MySQL statements, recreate Figure 1 as a MySQL table named `fig1` by using the updated `TMA_data` table from Q1(b).

(20 marks)

(d) Using appropriate MySQL statements, create a new MySQL table named `easy_data` based on the updated `TMA_data` table from Q1(b). This new `easy_data` table should contain all the information available in the `TMA_data` table and should be structured in a way that makes it easy to use for further analysis and visualisation. In 250 words or less, justify the structure of your `easy_data` table.

(40 marks)

---- END OF ASSIGNMENT ----