</talentlabs>

Chapter 2 - Hands-on Assignment

Instructions

1. Answer the below question in the boxes provided.
2. Please submit the assignment through TalentLabs Learning System.

Question 1
Give **two** examples of how Big Data is used in your everyday life? (1 mark)

1. Social media interactions 2. Photos

Question 2
What data could you collect about your everyday life? For example, how your wakeup time changes from day to day or the number of calories you burn. Think of **six** other examples of data that you could collect about your day life. It can be collected by writing it down, making a spreadsheet, using a smart watch, phone data or anything else you can think of! State these six examples below and the methods of collecting them:
(6 marks)

Example: Daily wakeup time – smart watch 1. Expenses – TnG & banking app 2. expenses category – personal finance app & Maybank banking app 3. screen time – phone app 4. daily internet data usage – phone 5. heartbeat rate – smart watch 6. Mood & emotions – writing diary

Question 3

Using your answers to Question 2.2, state what type of data you would need to collect (datetime, string, etc) for each one of your examples and whether the data would be qualitative or quantitative: (2 marks)

Example: Daily wakeup time – datetime – quantitative
1. Expenses – float - quantitative
2. expenses category – string - qualitative
3. screen time – float – quantitative
4. daily internet data usage – float - quantitative
5. heartbeat rate – float - quantitative
6. Mood & emotions - string - qualitative

Question 4

Explain how you would store your data from Question 2.2 and explain the reason for your decision: (2 marks)

I would choose csv format (or xlsx format if I've subscription for MS Excel) to record the data I collected, because of its friendly interface to anyone (with the help of apps like MS Excel & Google Spreadsheet) to add and update the row records into it.

I would not choose to have a database to record these data, as the data I collected is the data size is relatively small (probably in few MBs or GBs), unless I wish to create an automation to collect all these data together.

Question 5

Throughout Questions 2.2 – 2.4 you have been building up information that could go into a metadata file. Explain the benefits of making a metadata file: (2 marks)

Improve data understanding. Metadata helps analysts to understand data contents and data purpose without needing to analyze the raw data itself.

Improve third-party trust in data reliability. Metadata records how and when the data is collected. It helps to showcase the data reliability (e.g., check if the data is from reliable sources, and if the data is up to date) and improve the trust of others in the data.

Improve Data governance. Metadata can document data ownership and usage permissions, which helps to ensure the data is used appropriately.

Question 6

Using your answers from Questions 2.2-2.4 you can now make a metadata file. In the box below please write up the information that would go into the file. Make sure to include all the necessary parts that would normally be found in metadata files (such as the what, the when, the why, etc.). Use **both** a **description** of the data collection process as well as a **table** to describe the columns of your data. (6 marks)

Datetime	Expense	Expense Categories	Screen time	Internet data usage	Heartbeat rate	Mood & Emotions
2023-1-1	40	Food	90	1.9	80	Happy
2023-1-2	30	Food, Miscellaneous	180	1.3	87	Sad
2023-1-3	200	Food, Debt Payments	140	2.3	76	Happy
2023-1-4	40	Food	150	2.1	80	Sad
2023-1-5	530	Rental, Food	130	1.4	89	Happy

Dataset containing personal lifestyle-related information. It provides insights into the person's daily expenses, digital behavior (screen time and internet data usage), physical well-being (heartbeat rate), and emotional state.

The data was collected by Yong Sheng on the 6/1/2023.

The data was collected via manual recording from the phone app, smart watch, and diary.

Column	Data Type	Description
Date	Datetime	Date, in the format of YYYY-MM-DD.
Expense	Float	The amount of money spent of various expenses, in unit RM
Expense Categories	List of String – Categorical	The categories or types of expenses incurred. It can include multiple categories separated by commas.
Screen time	Float	The amount of time spent looking at screen, in minutes
Internet data usage	Float	The amount of internet data used, in GBs
Heartbeat rate	Float	Heartbeat rate, in beats per

		minute
Mood & emotions	String - Categorical	The emotional condition, documented as either “Happy” or “Sad”