Quizzes : Module 1 Part 2

The Data Pipeline

In the LAMS Sequence, you have learned the theory behind this module. It is also expected that you have attempted the quizzes embedded within the LAMS Sequence, and have used the “unlimited attempts” opportunity to score 100%. Here are the quiz questions, consolidated with their answers and corresponding feedback. This is for your after-LAMS revision.

Question 1

Which of the variables (out of the four options) in this dataset are "Categorical" (with Factors/Levels) in nature? Choose all answers that seem correct to you.

Year	Sex	Level of school	Age	No of vice principals (numbers)
2019	MF	PRIMARY	30 - 34	1
2019	F	PRIMARY	30 - 34	1
2019	MF	SECONDARY	30 - 34	2
2019	F	SECONDARY	30 - 34	0
2019	MF	PRE-UNIVERSITY	30 - 34	1
2019	F	PRE-UNIVERSITY	30 - 34	1
2019	MF	PRIMARY	35 - 39	23
2019	F	PRIMARY	35 - 39	19
2019	MF	SECONDARY	35 - 39	37
2019	F	SECONDARY	35 - 39	20
2019	MF	PRE-UNIVERSITY	35 - 39	7
2019	F	PRE-UNIVERSITY	35 - 39	3

Answer Choice	Verdict	Explanation
Sex	Correct	Yes, of course. This is a categorical variable.
Level of School	Correct	Yes. It has levels. Naturally a categorical variable.
Age	Correct	Age is generally numeric, but in this case, it is broken up in classes or bands. Hence, categorical.
No of Vice Principals	Wrong	This is numeric for sure.

Reference Module 1 Topic 4 : Structured Data in Practice Slide 3 and Slide 4 and Slide 5

Question 2

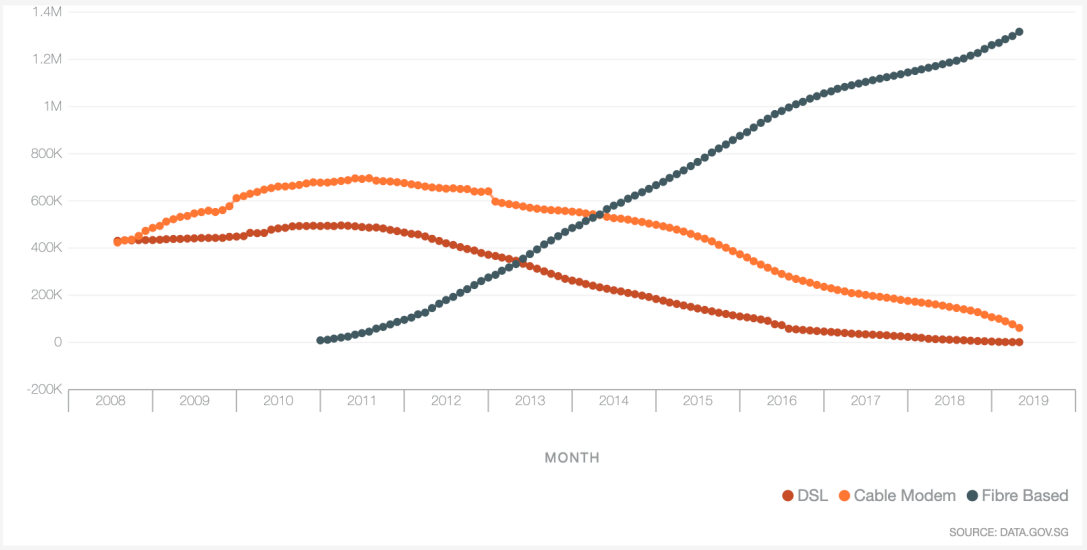
Suppose you find a variable (column) in a structured dataset with ten values : [1, 0, 1, 1, 1, 0, 0, 1, 0, 0]. Which one of the following conclusions can you draw from this information? Choose all answers that seem correct to you.

Answer Choice	Verdict	Explanation
Half of the values (the 1's) are different from the other half (the 0's).	Correct	This is correct, as the levels for categorical can also be "different", as is true for numeric values. But remember, you can't assume some variable is numeric if their values look like numbers. Be careful.
The average value (or mean) of the variable is 5/10 = 1/2 = 0.5.	Wrong	Are you assuming that the variable is "Numeric", as the values are 0 and 1. What if it's gender, with 0 = Male and 1 = Female? Would Mean make sense in that case? Think again.
Half of the values (the 1's) are greater than the other half (the 0's).	Wrong	Are you assuming that the variable is "Numeric", as the values are 0 and 1. What if it's gender, with 0 = Male and 1 = Female? Would "greater" or "lower" make sense in that case? Think again.
The maximum value of the variable is 1, and the minimum value is 0.	Wrong	Are you assuming that the variable is "Numeric", as the values are 0 and 1. What if it's gender, with 0 = Male and 1 = Female? Would "maximum" or "minimum" make sense in that case? Think again.

Reference Module 1 Topic 4 : Structured Data in Practice Slide 3 and Slide 4 and Slide 5

Question 3

The following Time Series presents the Monthly number of Broadband Subscriptions across Singapore. If we take every datapoint in this dataset, and subtract the previous datapoint from it, the resulting dataset will be:



Answer Choice	Verdict	Explanation
Time Series, with different interpretation and labels for the time axis	Correct	Yes, correct. The data format is still a time series. Just the time axis is now "difference" in time.
Time Series, with the same interpretation and labels for the time axis	Wrong	It is a time series, no doubt, but the time axis has changed. Think again.
Structured Data of type "Numeric", but NOT a Time Series any more	Wrong	Think again. Time series means a numeric value varying with time. Isn't it the same?
Structured Data of type "Mixed", but NOT a Time Series any more	Wrong	Think again. Time series means a numeric value varying with time. Isn't it the same?

Reference Module 1 Topic 4 : Structured Data in Practice Slide 6

Question 4

In which of the following ways can you interpret Facebook data as a Network (or Graph)? Note : Multiple answers may be correct. Choose all answers that seem correct to you.

Answer Choice	Verdict	Explanation
Facebook Profiles as nodes and Facebook Friendships (yes/no) as links	Correct	Yes, right. This is of course one interpretation. Two nodes (profiles) are connected if they are "friends" on Facebook.
Facebook Profiles as nodes and Number of Common Interests as links	Correct	Yes, right. This is of course one interpretation. Two nodes (profiles) are connected if they have "common interests" on Facebook, and the number of common interests become the weight of the edge in the network/graph.
Facebook Profiles as nodes and Number of Individual Friends as links	Wrong	Think again. Number of individual friends is a property of the nodes independent of one another. How does it affect any sort of connection or edge between them? For graphs, we need something to define an edge.
Facebook Profiles as nodes and Number of Photos Uploaded as links	Wrong	Think again. Number of individual photos is a property of the nodes independent of one another. How does it affect any sort of connection or edge between them? For graphs, we need something to define an edge.

Reference Module 1 Topic 4 : Structured Data in Practice Slide 7

Question 5

Suppose you are designing a Survey Questionaire to collect the feedback of your fellow students about this course. If you want to obtain a Structured Dataset of Mixed type as a result of this survey, what type of questions can you include in the Questionaire?

Answer Choice	Verdict	Explanation
Multiple Choice Questions with fixed number of options	Correct	Yes of course. The responses will be Categorical. But you need Numeric too.
Rating for the Course, within a specific range (say 0 to 10)	Correct	Yes of course. The responses will be Numeric. But you need Categorical too.
Detailed feedback/comments on how to improve the course	Wrong	Not quite. The responses will be Unstructured Text. Think again.

Reference Module 1 Topic 5 : Unstructured Data in Practice Slide 2