Quizzes : Module 1 Part 1

Data Analytic Thinking

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

What is the first step in the Data Science pipeline?

Answer Choice	Verdict	Explanation
Problem Formulation	Correct	Yup! You can't even start a Data Science process without having a well-formed problem.
Machine Learning	Wrong	Really? How do you know what to solve using Machine Learning? Think again.
Data Visualization	Wrong	Sometimes it good to visualize the data, but only when you don't even have a problem at hand. Think again.
Statistical Inference	Wrong	Nopes! You can't infer things from the data unless you know what is the problem at hand. Think again.

Reference Module 1 Topic 1 : What is Data Science? Slide 4 and Slide 10

Question 2

Which step in Data Science relies heavily on Algorithmic Optimization?

Answer Choice	Verdict	Explanation
Machine Learning	Correct	Yes, this is correct. In fact, most Machine Learning algorithms are clever optimizations with respect to a "cost function".
Data Collection	Wrong	Data collection is a huge part of Data Science, but we are still not in the zone of algorithms. Think again.
Data Visualization	Wrong	Sometimes, visualization requires some clever algorithms to handle big data. But this is not the main algorithmic sector.
Digital Storytelling	Wrong	Nopes! There's a lot of artistic elements (and programming) in storytelling, but algorithmic optimization is not a part of it.

Reference Module 1 Topic 1 : What is Data Science? Slide 7

Question 3

Match the Data Science problems with their respective Problem Types.

Problem Definition	Problem Type	Explanation
How many students will visit K-Cuts (the hair-cutting place) at North Spine this Friday?	Numeric Prediction	That's pretty intuitive. You are literally predicting the "number" of students.
Will K-Cuts (the hair-cutting place) at North Spine remain open at 11 am this Sunday?	Prediction of Class	The two classes are "YES" and "NO", and you are predicting which one.

Reference Module 1 Topic 2 : Data Science Problems Slide 3 and Slide 4

Question 4

Suppose you want to buy a Condominium in Singapore, and you want to find the best locality. What type of problem would this be in a Data Science context? Choose the most appropriate answer.

Answer Choice	Verdict	Explanation
Both Numeric and Class Prediction	Correct	That's correct! There are several factors associated with the problem, and you may encounter both numeric predictions (like estimating the price of the place) and class predictions (like choosing the best neighborhood in the city). So, both of them.
Prediction of a Numeric Variable	Wrong	Somewhat correct, as you will probably need to estimate/predict some numeric values, like the price of the place. However, there's more to a Condominium than just its price. Think again.
Prediction of a specific Class	Wrong	Somewhat correct, as you will probably need to estimate/predict some class values, like the best neighborhood in the city. However, there's more to a Condominium than just its neighborhood. Think again.
Neither Numeric nor Class Prediction	Wrong	Well, we do not know what else it could be. There are two major zones for prediction. Think again.

Reference Module 1 Topic 2 : Data Science Problems Slide 3 and Slide 4

Question 5

Match the Data Science problems with their respective Problem Types.

Problem Definition	Problem Type	Explanation
How many "types" of product does the Prime Supermarket at North Spine carry?	Detection of Structure in Data	As you do not know “what” the products are, you can’t perform classification. You can only look at the pattern of products, and detect a structure within, using clustering.
Is there a product in the Prime Supermarket of North Spine which does not belong there?	Detection of Anomaly in Data	As you do not always know the exact characterization of products, you can’t perform classification. More appropriate will be to find the “weird” product within all products, as an anomaly.

Reference Module 1 Topic 2 : Data Science Problems Slide 5 and Slide 6

Question 6

Suppose you want to design a Computer Game where the computer plays chess against a human player. What category of Data Science will the problem belong? Choose the most appropriate answer.

Answer Choice	Verdict	Explanation
Taking Adaptive Decisions	Correct	Yup! That's correct. We are in the zone of Artificial Intelligence in this case, where learning means adaptive learning.
Identifying Structure in Data	Wrong	There is of course some structure in any game. But this is not the main task for the computer to master in case of Chess. Think again.
Identifying Anomalies in Data	Wrong	Anomaly detection is probably not the most appropriate task in playing Chess. Think again.

Reference Module 1 Topic 2 : Data Science Problems Slide 7

Question 7

Prediction of Numeric Values (Regression) can be accomplished using …

Answer Choice	Verdict	Explanation
Any of the models mentioned below	Correct	That's right. Almost any model can do anything in practice.
Only the Linear Regression Models	Wrong	Models are not limited -- almost any model can do anything in practice. Think again.
Only the Tree-based Models	Wrong	Somewhat correct, as you will probably need to estimate/predict some class values, like the best neighborhood in the city. However, there's more to a Condominium than just its neighborhood. Think again.
Only Neural Network Models	Wrong	Well, we do not know what else it could be. There are two major zones for prediction. Think again.

Reference Module 1 Topic 3 : Data Science Solutions Slide 5

Question 8

Suppose you want to know if "that specific boy/girl" will go out with you. You model it as a Classification Problem, and use the data from all your previous attempts at dating to learn from. What will you learn at the end?

Answer Choice	Verdict	Explanation
The probability of "that specific boy/girl" going out with you.	Correct	That's correct. Any classification problem will produce a probability, and not just a class.
Whether "that specific boy/girl" will go out with you.	Wrong	Yes, that's the goal, right? But remember, classification generally returns probability. Think again.
Whether "that specific boy/girl" will go out with your arch-enemy.	Wrong	Well, you would want to know that too. But that's not the problem we are solving. Think again.
The probability of "that specific boy/girl" already having a girl/boyfriend.	Wrong	Hmm, tricky! Very hard to predict that from the data we have. Try again.

Reference Module 1 Topic 3 : Data Science Solutions Slide 7

Question 9

What is the most crucial concept in identifying Clusters in data?

Answer Choice	Verdict	Explanation
The notion of "distance" to identify close and far points.	Correct	Yes, indeed. The notion of distance is the most crucial aspect of identifying groups or clusters. We will see more of it later.
Techniques of visualizing the data to identify groups.	Wrong	Visualization is extremely important, no doubt, but in high dimensional datasets, you will not be able to do that effectively.
The notion of "anomalies" to weed out extreme points.	Wrong	Weeding out anomalies is a different ball-game altogether. Think again.
The expertise of a "domain expert" to identify clusters.	Wrong	You do need domain expertise, no doubt. But often, machine learning reveals more than an expert knows. Think again.

Reference Module 1 Topic 3 : Data Science Solutions Slide 10

Question 10

How would you spot the next "Joseph Schooling", that is, the next biggest star in Swimming from Singapore?

Answer Choice	Verdict	Explanation
Anomaly Detection across all Swimmers at the school-level.	Correct	This is probably the best approach, as of course, Joseph Schooling is an exception within swimmers from Singapore.
Clustering groups across all Athletes at the school-level.	Wrong	Clustering will surely help you with spotting the general structure in swimmers of Singapore. But Schooling is an exception, isn't it? Think again.
Predicting the best lap-times of all Swimmers in Singapore.	Wrong	You can surely do that. However, you will have to look for exceptions at the end of the day. Think again.
Predicting the School/JC/Poly to produce the best Swimmer.	Wrong	This classification can work, but it's not just the School/JC/Poly that matters. We will need to find these exceptional talent, as is Joseph Schooling. Think again.

Reference Module 1 Topic 3 : Data Science Solutions Slide 13