Zealthlife Technologies
Interview Round - 1
Time - 30 mins

Question 1
[Civics]
At any given session of the Rajya Sabha, there are about 21 new questions that are brought up for debate and approximately 18 old questions that are pending for discussion. Each question is contained in a file which is marked with a numerical “level of importance” tag. So, a question which is more important takes priority in discussion over the ones that are less important. If the importance level of 2 questions is the same, the older one gets picked up first for discussion. The prioritization of these files is carried out using the concept of queues. What is the minimum number of queues that are required to carry out this process of prioritization?
(a) 1
(b) 2
(c) 3
(d) 4

ANS – (b) 2

Question 2
[Literature]
White Queen: Can you do addition? What's one and one and one and one and one and one and
one and one and one?
Alice (after about 10secs): Umm.. nine.
White Queen: Good. And what's one and one and one and one and one and one and one and
one and one .. and one?
Alice (immediately): Ten!
White Queen: That was fast, I see you like recursion.
Alice: Recursion? How silly! It was actually_____.
(a) Iteration
(b) Binary Search
(c) Dynamic Programming
(d) A guess

ANS – (d) A guess

Question 3
[Economics]
Transportation economics is one of most focussed area for logistic and supply chain lines of businesses. One of the primary optimizations in transportation economics is to choose the most optimal path of transportation across k locations. When a computer scientist was tasked with this problem, (s)he chose to solve this by creating a Minimum Spanning Tree (MST) of a graph connecting these k locations. The Chief Operations Officer of a supply chain company wishes to know if by using Minimum Spanning Tree approach, will we have a sub-graph that gives the shortest distance between any two locations in this map of k locations?

(a) Yes, this is the core property of MST
(b) Yes, but since obtaining MST of a graph is NP hard it's not practical
(c) Nope. This won't give us the shortest distance between any two locations
(d) I think the solution to this is same as the solution to the Travelling Salesman Problem

ANS- (c) Nope. This won't give us the shortest distance between any two locations

Question 4
[Finance]
Mukesh maintains a diary to track his finances. The diary has a page for each day of the year in chronological order, and each month's pages are colourcoded differently on the long edge of the diary. In each day's page, he enters 2 numbers: the total amount he spent on that day and the total amount he's spent since the beginning of the year (including the current day). His wife Nita does not believe this is an efficient system and to make Mukesh realise this fact, asks him a few questions related to the spend at the end of the year. She believes Mukesh will need to scan every page of his diary to answer each question. Which of the following questions of Nita can Mukesh definitely answer (even in the worst case) without scanning every page of his diary? **Please give your reasons for each answer.**

(a) On which day of the year did the total spend go beyond ₹100000 for the first time?
(b) Which month of the year had the highest spend?
(c) Which day of the year had the highest spend?
(d) On how many days of the year were spend 0?

ANS – (a) & (b) are the questions which mukesh can answer without scanning every page of his diary.
Reason for (a) – As mukesh enters the total spend int the start of every year in each day's page he can find the first day when spend is greater than 100000.
Reason for (b) – Mukesh can find the colour coded month and find the highest spend of the month, as pages of each month are differently colour coded.

Question 5
[Chemistry]
One of the earlier techniques of classifying a chemical compound was based on their molecular weight. Scientist Alan Wood proposed a method of classifying the compounds into 9 classes (0 through 8) using the hash function C(k) = Mk mod 9 , (where C(k) is the class of the compound k and M is the molecular weight of the compound). If there is a hash collision, he proposed, that k they shall be resolved by chaining.)
If we were to classify the following 9 compounds with molecular weights 5, 28, 19, 15, 20, 33,12, 17 and 10 using Alan Wood's proposal. The maximum, minimum, and average chain lengths in the hash table, respectively, would be
(a) 3, 0, and 1
(b) 3, 3, and 3
(c) 4, 0, and 1
(d) 3, 0, and 2

ANS –(a) 3,0 and 1

Question 6
[Discrete Math]
Consider a simple undirected graph G with n vertices and no edges (yet). What is the minimum number of edges that must be added to G in order to ensure that it is connected?

(a) (n+1)
(b) (n-1)
(c) ((n1)*(n2))/2
(d) (n*(n1)*(n2))/6

ANS – (b) n-1

Consider a simple undirected graph G with n vertices and no edges (yet). What is the maximum number of edges that can be added to G ?

(a) (n+1)
(b) (n-1)
(c) ((n)*(n-1))/2
(d) (n*(n-1)*(n-2))/6

ANS – (c) ((n)*(n-1))/2

Question 7
[Computer Science]

We have a file in our server named “temp.csv”, the contents are :
T, 1, 2
X, 3, 4
Y, 5, 6
Z, 7, 8

What would be the output of the following bash command?

cat temp.csv | sort -R | head -1 | cut -d, -f2,3

(a) 7, 8
(b) 1, 2
(c) Z, 7, 8
(d) None of the above

ANS – (a) 7,8

Question 8
[eCommerce]
DealSnap is an eCommerce website and they’ve recently redesigned their userexperience on the website. You are asked to run an experiment to see whether there is a difference in length of visit between customers viewing the old website and new website. One way of confirming this is:
a. Calculating the difference in mean length of visit
b. Performing a ttest on the mean length of visit for the two groups
c. Performing linear regression on the observations of length of visit, with group membership as the dependent variable
d. Performing a chi square test on the mean length of visit for the two groups

ANS – b. Performing a test on the mean length of visit for the two groups

Question 9
[Statistics]
Regularization, in mathematics and statistics and particularly in the fields of machine learning and inverse problems, refers to a process of introducing additional information in order to solve an ill-posed problem or to prevent overfitting. Can you tell us what is the difference between L1 and L2 regularization is:
a. L1 regularization is useful for performing variable selection prior to running a model, whereas L2 regularization is not
b. L1 regularization cannot be combined with other forms of model tuning, whereas L2 can

c. L1 regularization uses the absolute value of the model parameters, while L2 regularization uses the squared distance of the parameter vector

d. L1 regularization uses the squared distance of the parameter vector, while L2 regularization uses the absolute value of the model parameters

ANS – c. L1 regularization uses the absolute value of the model parameters, while L2 regularization uses the squared distance of the parameter vector

Question 10

[Politics]

With the ongoing United States' presidential primary elections you are attempting to model the number of delegations each candidate would secure at a state based on their age, education and tenure in politics. Along the way, you crunch in a lot of historical data and train your model. However, you find that your model has trained well, but subsequently does poorly on new data.

This could be because:

a. Human behaviour isn't a good subject for modelling
b. Your model is overfitting
c. Your model is underfitting
d. None of the above

ANS – b. Your model is overfitting

Question 11

[Marketing]

The customer churn analysis feature helps you identify and focus on higher value customers,determine what actions typically precede a lost customer or sale, and better understand what factors influence customer spending. When you improve customer retention, you substantially improve the bottom line. The marketing department of your company is looking for a way to call customers who are likely to churn and persuade them to stop. When you are building a model to do this, one of the best ways to assess its usefulness is:

a. The recall score, as this gives the best indication of the strike rate per call
b. The precision score, as this gives the best indication of the whether the model is successfully identifying customers about to churn
c. The precision score, as this gives the best indication of the strike rate per call
d. None of the above

ANS – b The precision score, as this gives the best indication of the whether the model is successfully identifying customers about to churn