For the initial round of Datathon 2019, you need to showcase the proficiency and experience of your team in Data Science. This document consists of seven questions that are related to various aspects of Data Science. You need to answer all the questions.

You have to submit your answers and respective codes to the same mail on or before Sunday 06th 2019 12:00 midnight.

SLIIT CODEFEST 2019

ZONE24x7 INSPIRING EMOTIONS

Question 1

You are given to paint a floor area of size A. There will be 12 paint buckets from 4 primary colors with each having 3 shades (i.e. total 12 buckets). The bucket size, or more specifically, the amount of area you can paint from each shade is given in the following table/matrix/arrays. The different shades of the same primary color are shown in the same row.

- [12, 23, 14]
- [10, 30, 15]
- [16, 22, 35]
- [14, 24, 20]

Problem & Constraints

You need to select 4 shades to paint the floor area such that;

1. Entire floor area should be painted; also no overlaps of shades are allowed
2. "One and only one" shade from each primary color has been selected for the final painting
3. Amount of wastage is minimized (i.e. assume once we open and use a bucket, any remainings will be discarded)

Implement a python program to answer the following problems;

Q1. The color shades (or buckets) satisfying the above constraints (if A = 100)

Q2. The amount of wastage in the above scenario

Q3. What will be the solution for Q1 and Q2 if A = 90?

Note: You may use the below notation to reference each shade in the above map.

R - row index

C - column index

(r,c) - shade in (r+1)th row and (c+1)th column

e.g. (0,0) -> 12, (0,1) -> 23, (1,2) -> 15, etc.

With this, the answer for Q1 can be given in the format [(0,1), (1,2), (2,0), (3,2)]

In addition to the answers to the above, please share the working code project as a zip file OR the location of the Git repository OR the link to an online coding platform.(e.g. https://repl.it/languages/python3)

Question 2

The “sample_twitter_personal_data.tsv” file contains personal data of few twitter users. It includes the following list of data,

"twitter_id"	"age"
"dob_day"	"dob_year"
"dob_month"	"gender"
"followers_count"	“initiated_to_follow"
"heart"	"heart_received"
"mobile_app_heart"	"mobile_app_heart_received"
"web_heart"	"web_heart_received"

Read and study the data carefully and implement R Program to answer the following questions;

Q1. List all the headers in the dataset.

Q2. Order (ascending) the headers and assign numbers for the ordered headers.

Q3. List all Twitter users whose "followers_count" is greater (>) than 100.

Q4. List all the twitter MALE users whose "followers_count" is greater (>) than 100.

Q5. It is a saying,

“*Males always initiate to follow another Twitter account (female/male account) FIRST than Female*”

To prove the above statement, analyze the data and provide some evidence.

Q6.Your manager has requested study the data and asked you to provide a graphical representation for the following use cases,

1.HISTOGRAMS for day column in DOB (eg: dob_day)

2.FREQUENCY PLOT for the followers count

3.FREQUENCY PLOT for the age of the users over the years in the sample dataset

4.Plot BAR CHART to indicate and show the total hearts given by male, female users

5.Analyze and visualize in CHART to confirm that people use a mobile/web interface to experience twitter

Q7. Based on your critical analysis write down in points your thoughts and suggestions to improve the twitter features.

Question 3

Head of Innovation in a restaurant chain is keen on identifying customer behavior during the stay in the restaurant. They are planning to get the streams of videos from four cameras overlooking the customers. There are 500 restaurants located all over the country. Video data is aggregated every hour through a video analytics solution. Data consist of the following.

"FrameID"	"StoreID"
"Date"	"Time"
"Customer Count"	"Customer Male Count"
"Customer Female Count"	"Kids Count"
"Teen Count"	"Adult Count"
"Couples"	"Individuals"
"Groups"	"Happy Customers"
"Other Customers"	

Innovation team does not have much experience with pattern identification and how customer behavioral insights can be derived from the data. You have been consulted to derive customer behavioral insights and predictions.

Answer the following

Q1.You have been asked to identify anomalous behavior in customer counts

a. Describe the three different anomaly types
b. What are the Data Science methods to detect those three different anomaly types
c. Explain how one method can be applied to detect anomalies
d. You have seen that there is a pattern that every Friday evening has a higher customer count, explain how you are going to deal with the seasonal patterns when detecting anomalies

Q2.a.You have been asked to predict the customer counts for a given day. Describe Data Science methods that you can apply to perform the customer count predictions for a given day.

b.How would you change your transform your training methodology when the data gets larger

Q3.You have also been given access to the customer transaction details. You have given a task to predict whether a customer is going to visit another day or not.

From the past two years of data, it was identified that there were only 98,457 customers who have visited only once. There were 9,987,125 customers who have visited more than once.

a. Describe what are the possible performance metrics that can be used to evaluate the predictive model
b. Suppose the mode that you have built has a high recall rate but the precision is lower. Explain the approach you will take to increase the precision of the prediction.

Question 4

The global retailer chain has the vision to increase their average order value of an online purchase. Average order value(AOV) tracks the average dollar amount spent by a customer each time when they place an online order. To increase the AOV, they are planning to utilize the power of data and AI.

Answer the following

Q1.Mention the algorithms that can be used for this scenario considering the browsing and purchasing habits of customers?

Q2.Explain in detail how one algorithm can be utilized to achieve the target. Your answer should consist of the following but not limiting

i. How does the algorithm work?
ii. How to use the algorithm?
iii. How to measure the effectiveness of the algorithm for the given problem?

Question 5

Visitors to the search page will look for different products in different regions of the country based on their interests. For example, in January, people searching for shoes in the New England area will perhaps look for heavier footwear than people from San Diego at the same time of the year. The algorithm should be able to capture such signals so that, search results can be improved for the products based on locations from which the search activity happened.

Answer the following

Q1.Mention the algorithms that can be used for this scenario considering the browsing and purchasing habits of customers?

Q2.Explain in detail how one algorithm can be utilized to achieve the target. Your answer should consist of the following but not limiting

a. How does the algorithm work?
b. How to use the algorithm?
c. How to measure the effectiveness of the algorithm for the given problem?

Question 6

A recommendation platform provides travel recommendations to users based on the images posted on social media by the user and places the user checked-in.

Q1. Suggest methods to extract the data posted by the user.
Q2. What are the most suitable algorithms to identify the features of the image and classify the images based on the type of location(ex:- beach, restaurant, etc.)

Question 7

Q1. Write a linux command to display files in a directory?

Q2. How will you find the file that was updated last in a directory?

Q3. Let's say that you have a file with special characters such as "&$?;" in its file name, how safely will you delete that file? Elaborate your answer.

Q4. There is a file with multiple rows containing numeric values, propose a simple shell command to calculate the sum of all the numbers.

Q5. Let's say you are a standard Linux user (you are using CentOS 7.2 x64) but you are unable to change your password using "passwd" command in the terminal. What went wrong? Elaborate your answer (assume you have root password if needed).

-The End-