SQL(20 Marks)

Topic - Condition + String matching + logical operator)

1. Give me list of Customers whose First Name starts with N Or they Live in city xyz.  (Return me Customer Id, First name, Last Name, City) (2 Marks)
2. Give me list of Tracks where unit prize range between x and y and whose composer name does not contain 'T' (Return me Track Id, Name, Composer, Unit Price) (2 Marks)

Topic - Group By, Order By

1. Calculate the sum of total Invoice bill and average of total Invoice bill for each day of month (June 2020). (Return me Invoice Date, Sum of Total bill for that day, Average of Total Bill for that day) (3 Marks)
2. Total Quantity of Each Track ID purchased so far. Sort the Track IDs in Descending order of it’s total Count (Return me Track ID and its total Quantity) (3 Marks)

 Topic - Join/Subqueries

1. Give me list of Artist IDs,  whose none of the tracks are present in any Playlist. (Return Artist IDs and their names) (5 Marks)
2. Give me a list of Album ID whose Tracks are present in more than 10 playlist(Return me Album ID, it's Title and the count of playlists wherein it's tracks are present) {Use Join Statement}  (5 Marks)

**Python/Data Visualization/Feature Engineering (20 Marks)**

<Prerequisite - Local Python setup + one csv file>

1. Load csv file in notebook using python code
2. Explore Data using python code (2)

* Total records, Null count
* Mean, median, std-dev for each numerical column

1. Plot charts and explain them in words(min 2 charts) (3)
2. Check if “BMI” column has null values? what is the percentage of null values ? use data engineering method to fill the null values. (3)
3. Remove outliers from any two numerical columns using any appropriate ML technique(3)
4. Perform Scaling on all numerical columns(3)
5. Perform Encoding on all categorical columns(3)
6. Generate at least 2 new features using existing features of your choice(3).

**Machine Learning (30 marks)**

<Prerequisite - Local Python setup + one csv file , Libs : Pandas, Sklearn>

1. Select Important Features and create a dataframe (4)

1. Split Data to train, test, validate (3)
2. Train Model (5)
3. Test Model (3)
4. Evaluate Model using evaluation matrices (5)

1. Hyper Parameter tuning to improve performance (4)
2. Save Model to a pickle file (2)
3. Predict values for validate set using pickle file(2)
4. Write validate set along with it's predicted values to csv file (2)