**Level 1:**

**My findings and thoughts:**

1. There needs to be a clear relationship between response and score.
2. The response column contains multiple values separated by commas. Each value represents a key-value pair, where the key corresponds to an option number and the value corresponds to the response given. Eg: A,B,C ,"1=>(EMPTY), 1=>2, 2=>(EMPTY), 2=>5, 3=>3, 3=>1, 4=>4, 5=>(EMPTY), 6=>6" etc.
3. When looking at the full data(Before applying the SQL function on the data):

For the score: 2.000 and question\_id: 12595, the most repeated value for the `response` column is `borrow=>[Receiving money with an agreement to repay it in the future, usually with interest charged.], lend=>[Giving money to a person or a business with the expectation that it will be repaid.], trade=>[The transfer of goods and services, often in exchange for money from one individual or organization to another.]`. It is repeated 24664 times

1. For the score: 0.000 and question\_id: 12563, the most repeated value is "D". It is repeated 24282.
2. For the score: 1.000 and question\_id: 12563, the most repeated value is "A". It is repeated 16913.
3. When I analyzed further the for score:2.00 and question\_id 12563, It showed multiple responses such as B, B; A, "1=>(EMPTY), 2=>(EMPTY), A=>1, B=>(EMPTY), C=>(EMPTY), D=>(EMPTY), E=>(EMPTY), F=>(EMPTY), H=>2, I=>(EMPTY), J=>(EMPTY)"

Now coming back to the SQL query:

``` sql

select sqr.taqr\_id, sqr.response, sqr.score

from 20220813\_tw\_40.SubmittedQuestionResponses sqr

where sqr.question\_id = 17404

```

**The output of the query should contain only rows with question\_id = 17404**

Number of data rows in the query: 26217

1. For the score:2.000, only one response which is "1=>(EMPTY), 1=>2, 2=>(EMPTY), 2=>5, 3=>3, 3=>1, 4=>4, 5=>(EMPTY), 6=>6"
2. The number of unique score values is "4" and "2" being the highest score and "0" being the lowest
3. For the score value: "2" there is only one unique response. From there I assumed that the "2" is the highest score for that question and that the response value is the correct answer for that question.
4. To find the correct score according to the response The formula used is the number of correct answers/ Total of questions (9)\*2
5. After finding the difference of the new\_score value between the original and new values of the scores are compared to find the anomalous scores
6. A list of changes to be done is found.
7. The Dataframe anomalous\_scores shows the anomalous rows
8. The changes are to be made for the values in the data frame changes\_needed
9. Saved both these dataframes to CSV files “anomalous\_data.csv” and “changes\_needed.csv”

**Level 2:**

**My findings and thoughts:**

1. When I checked the difference between the response row of satisfied and notsatisfied JSON files
2. Under root "3" 'isFilled' value changes to True from False
3. Under root "3" 'isKeyIdUsed' value is added False
4. Under root "3" 'isCorrect' value changes to True
5. Under root "3" 'isResponded' value changes to True
6. Under root "3" 'isStarted' value changes to True
7. Simplifiedstatetargets are added which have id:'Answer'
8. My finding is that when both conditions (is\_nr == 1 and isResponded is False) are not satisfied means the assessment is filled and the student responds to the assessment
9. Saving the discrepant data in the alignment of `is\_nr=1` and none(response\_raw[entryId].isResponded=true) into csv file named “discrepant\_data.csv”

**Level 3:**

**My findings and thoughts:**

1. When comparing questions 15 and 17
2. Both questions 15 and 17 had two questions with different id
3. Every question is there in two languages English and French
4. Question 15 is 'Did you find the following features of the online assessment useful?
5. Question 17 is 'Please tell us how easy or difficult you found the assessment.'
6. When I tried to join the question\_df and response\_df like the SQL query I noticed that the response only contain question 1 they do not contain questions 15 and 17
7. So I couldn't find anomalous data about questions 15 and 17