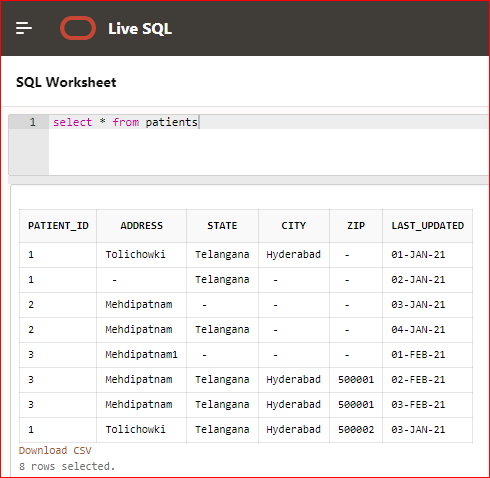
**Problem Statement:**

* In a Patients table which has duplicate patient ids and contains the addresses columns of the patients like
  + Address, state, city, zip.
* The problem here is each patient id may have some of the address columns as nulls
  + For example:
  + For patient\_id =1 address column is filled but state is empty but city and zip column is filled
  + For the same patient\_id =1 one of the other columns of the address may be null or fully filled.
* Query Task:

1. Get each patient\_id which has most no of address columns filled.
2. If there a tie between the any patient\_id that has most columns filled then we need to get the address columns with the most recent last\_updated date.

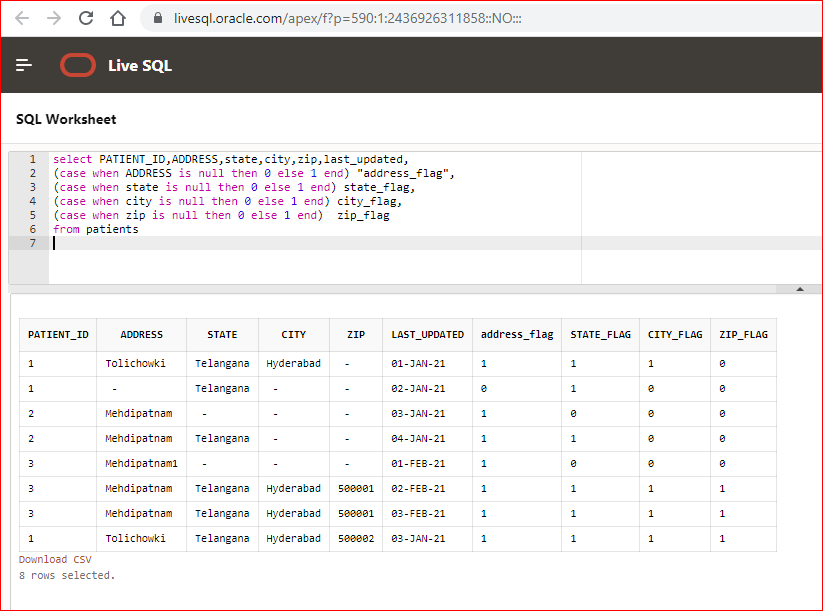
**Sample Data:**

Attached the query

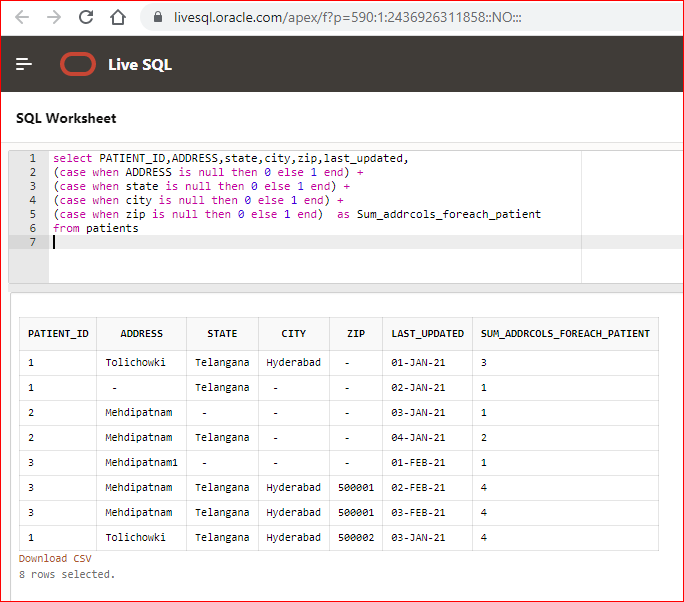


**Solution:** To solve this

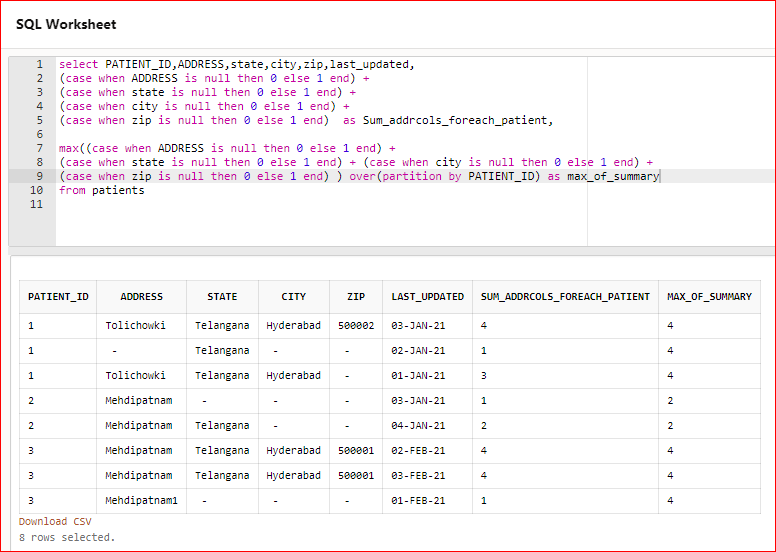
* First flagging each address column with the case statement into ‘0’ or ‘1’



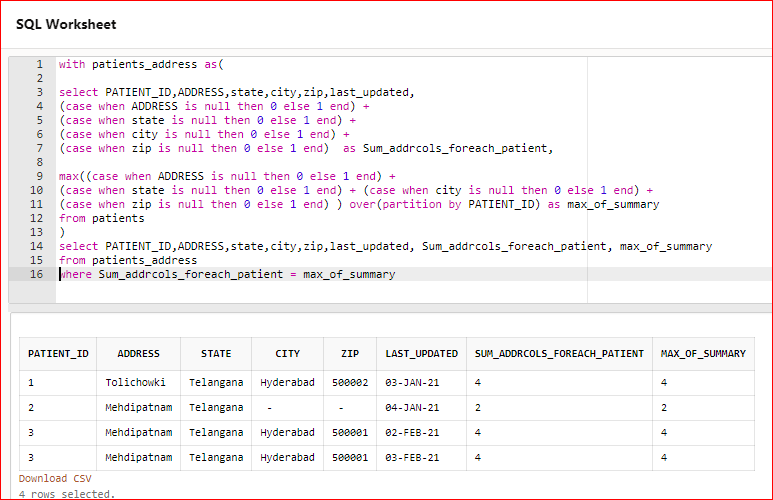
* Then I’m adding all the address column flags to make a single column.
* This will give me the number on how many address columns are filled for each row.



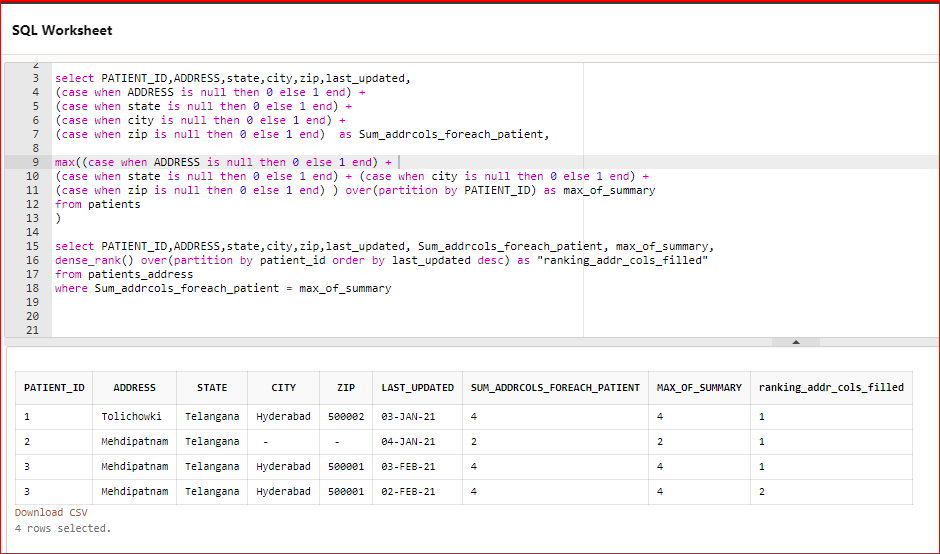
* Then I’m creating a window function to get the maximum of count of summary values for each patient\_id.
* For this I’m writing a window function this will calculate max for each patient\_id.



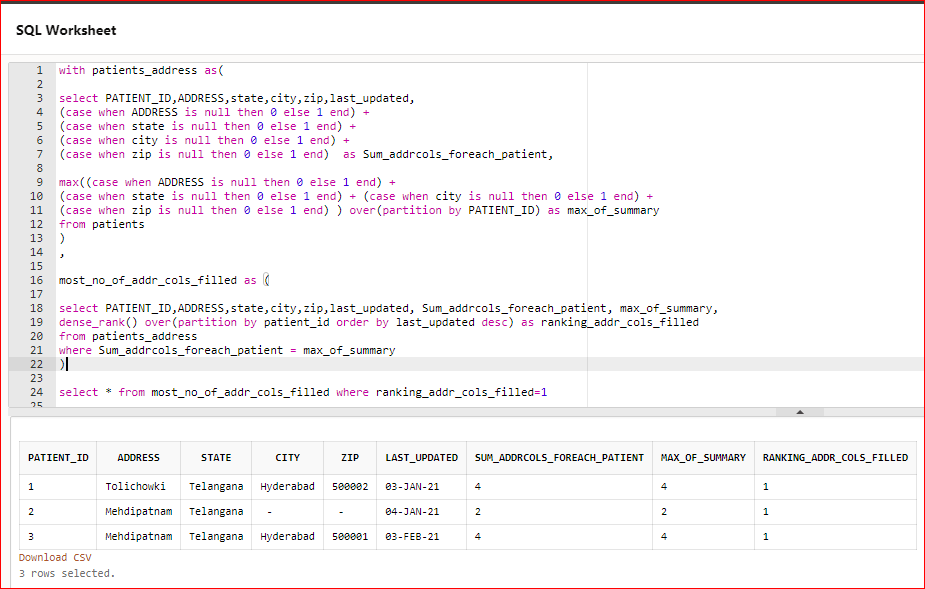
* Then I’m creating a common table expression using the sql statement that I wrote above and then query the common table expression with and include where condition that will give me the patient details with only the most no of columns filled.
* Condition is sum\_addrcols\_foreach\_patient = max\_of\_summary
* Look at the screenshot above based on the condition that I’m going to apply on my query I should get only four records in my result.



* This solves our first task – getting the patient\_id with most no of address columns filled.
* Now, coming to the second task.
* As we can see that there is a tie between the patient\_id “3”
* Now we need to get the patient data for patient\_id “3” with most recent last\_updated date.
* Now I will need to create another calculation to rank our data based on patient\_id and last\_updated date



* Now, I will create another common table expression with the above query so that we have two common table expression and then I will filter my final query to get the data where rank =1



I have solved this over the weekend. I feel happy to solve this and share it with you. 😊