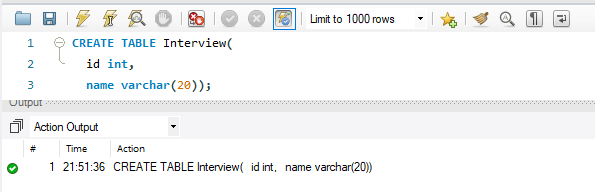
**Pre-Screen Technical Questions**

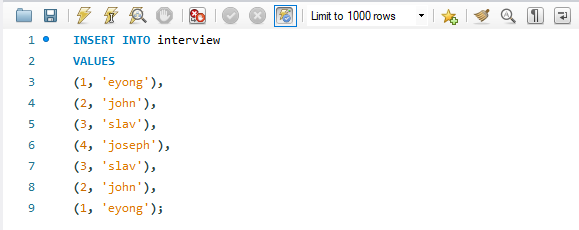
**Q1) Tell me a situation when you collected, interpreted, and analyzed data from variety of sources. How did you managed to combine these datasets into single source to make a decision or provide a recommendation? Describe in detail all the steps taken.**

|  |
| --- |
| This is a complete Data Analysis Pipeline; like from data load to insights; I did these kind of processing using following;   * Playing with SQL for data collection and some pre-processing * Everything in Python (**Recommended**)   The very first step is like what is the problem by nature; for example the problem is Customer segmentation. My thought process started with following questions;   * Where is the data (**need data sources**)   + Customer data with company (**transactional data**)   + Customer demographic data (**some other sources**)   + Customer life style data (**social profile etc.**)   + Data on any platform * Who are the stakeholders?   + Business People for making decisions   + Marketing team for making strategies   **Step 1: Data Collection**  I can easily connect to different sources such as MongoDB, SQL, Firebase or any data storage on cloud using available open source Python Libraries. The data from different sources has been loaded as Pandas DataFrame (this provide easy way to process data and very fastly).    **Step 2: Data Transformation**  Data in different sources might have different representations and forms. In order to merge different data sources together, we need data in standard form. For this purpose we perform data transformations. This include basic pre-processing steps. Processes in this steps depends on the project types and number of sources. But on general we might have following steps;   * Data Pre-processing: Removing unwanted data, inconsistent values etc. * Aggregated values calculations * Making new columns and measures * Merging/ Joining of difference sources (if possible)   **Step 3: Data Loading**  Once we have transformed and process data, we save that in some other storage that is might be a cloud storage, local SQL etc. storage or even file system to avoid repetition of transformation step if possible.  **Step 4: Data Analysis**  Now depends on the type of system it is, if the system is like some prediction system etc. then we can perform step 2, 3 and 4 one time. For the current system i.e. Customer Segmentation, we start data analysis steps; which are as follows:-   * Feature Engineering * Exploratory Data Analysis (EDA) * Machine Learning/ Statistical Modeling * Model Validation and Testing   **Step 5: Data Interpretations and Story Telling**  Once we are confident with Machine learning or statistical models, then it’s time to deploy the model for actual usage. The results from model are now will be shown to users in the form of dashboards/ and other interactive tool.    Example Dashboard 1    Example Dashboard 2 |

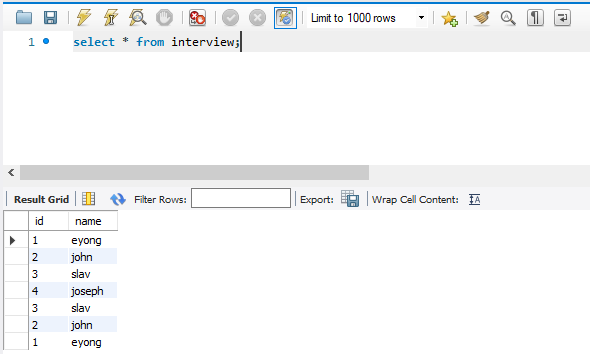
**For the rest of the questions, I have created table “interview”**



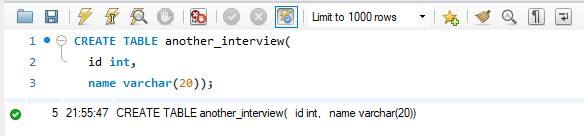
**Insert 7 records in table (duplicate records)**



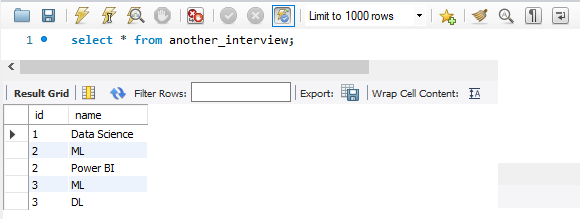
**Show all records**



**Create another table**



**Show all records**



**Q2) What does Intersect function do in SQL. How is it different from INNER JOIN?**

|  |
| --- |
| SQL INTERSECT operator combines two select statements and returns only the dataset that is common in both the statements. To put it simply, it acts as a mathematical intersection.  The INNER JOIN keyword selects all rows from both tables as long as there is a match between the columns.  Difference(s):   * INTERSECT removes duplicates; while INNER JOIN will returns duplicates if ID is duplicated in either of the tables * INTERSECT can return NULL; while INNER JOIN never return NULL * Technically speaking, INTERSECT is an OPERATOR; while INNER JOIN is set-based operator or JOIN * To compare the rows of two or more Oracle SELECT statements, the Oracle INTERSECT operator is used. After the comparing process, the INTERSECT operator returns the common or intersecting records from the corresponding columns of the selected expressions.     Result Returns by INNER JOIN |

**Q3) How to get distinct records from the table without using distinct keyword.**

|  |
| --- |
| We have many options to do that depends on what are you using, such as MSSQL, MySQL, SQL Server etc. Following are some of the ways to achieve:-   * GROUP BY      * UNION      * INTERSECT   The INTERSECT operator can be used to query for distinct rows in a table:  SELECT id, name FROM interview  INTERSECT  SELECT id, name FROM interview;   * CTE   CTE stands for Common Table Expressions. It can also be used to query for distinct rows in a table with the row\_number() function as shown below: |

**Q4) How do you get last 50% of rows from a data table?**

|  |
| --- |
|  |

**NOTE: For all the above questions in SQL, I can also use (Preferred) to use Python and then do same operations easily and quickly.**