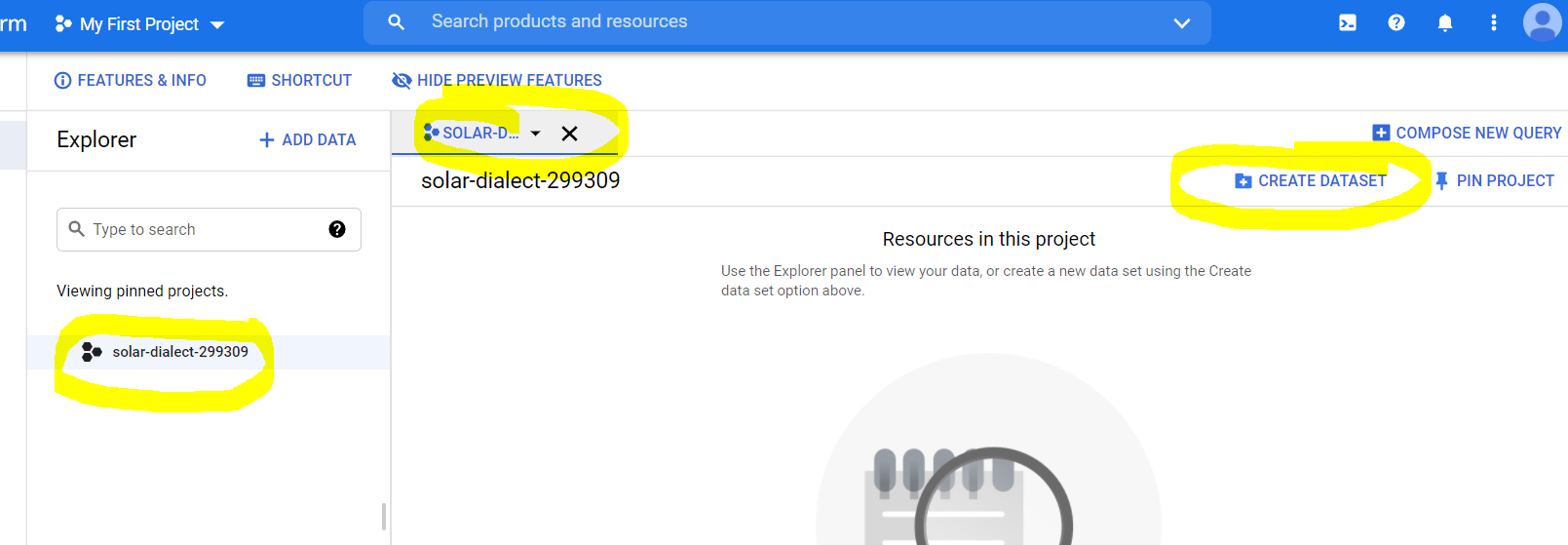
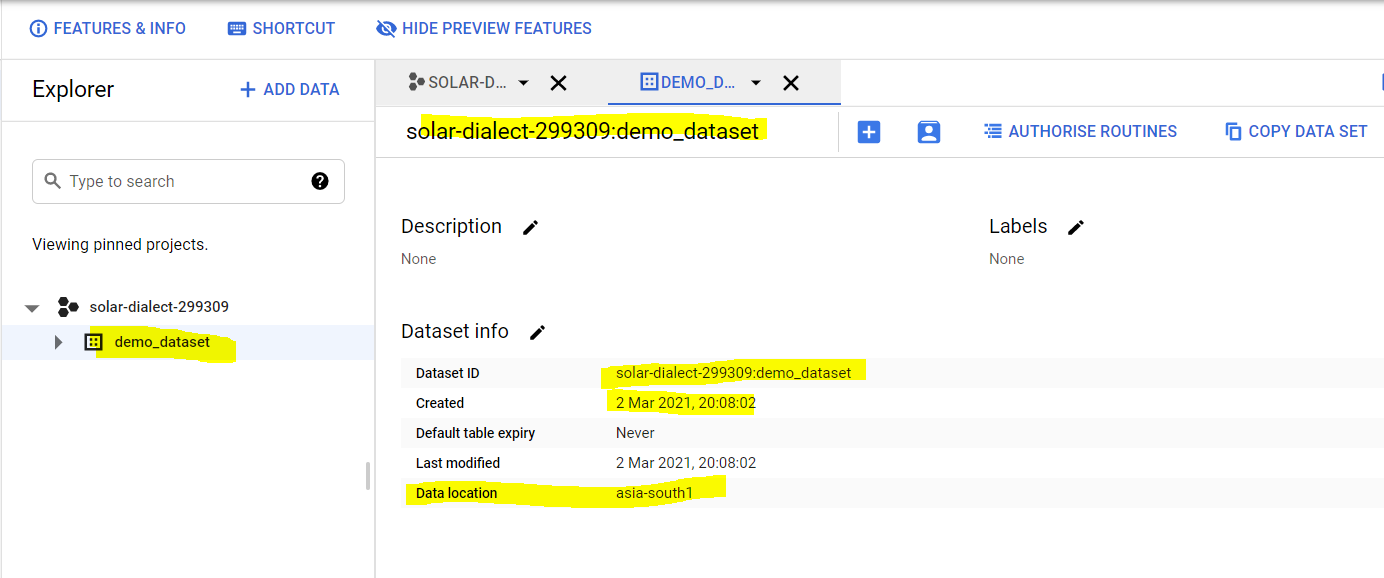
Dataset 🡪 One level above tables. Dataset can have any number of tables and metadata about table is stored in Dataset. Tables/views must belong to Dataset (I assumed Dataset is equivalent to SCHEMA in PostgresSQL)

**Create a Dataset**



Inside a project, you can create Dataset.

Region and Multi-region :: If you data is accessed by specific geographical region, then single region. If data is accessed by globally, have a multi-region. In multi-region, data is replicated to multiple data centres



.

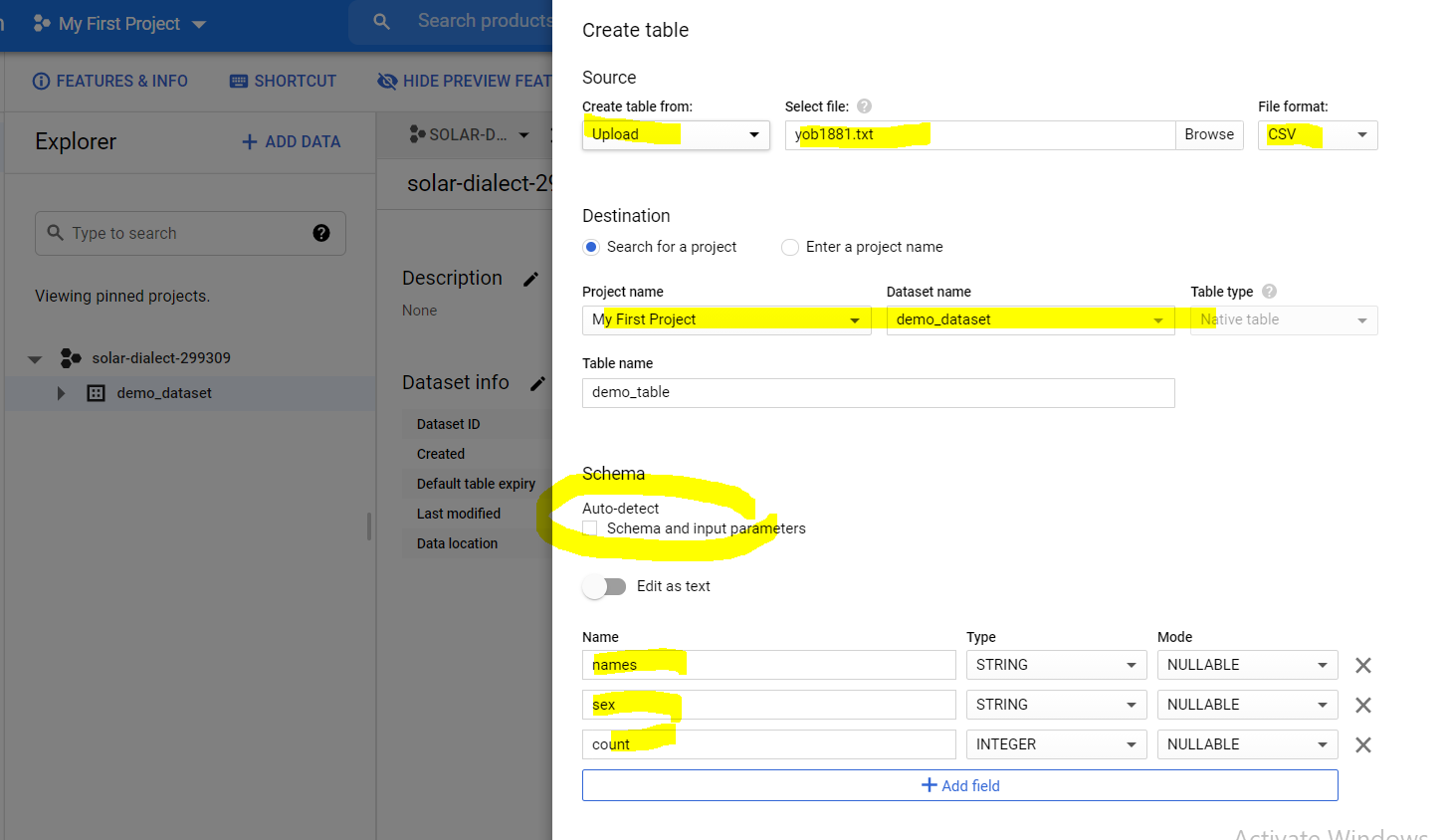
**Creating table inside dataset**

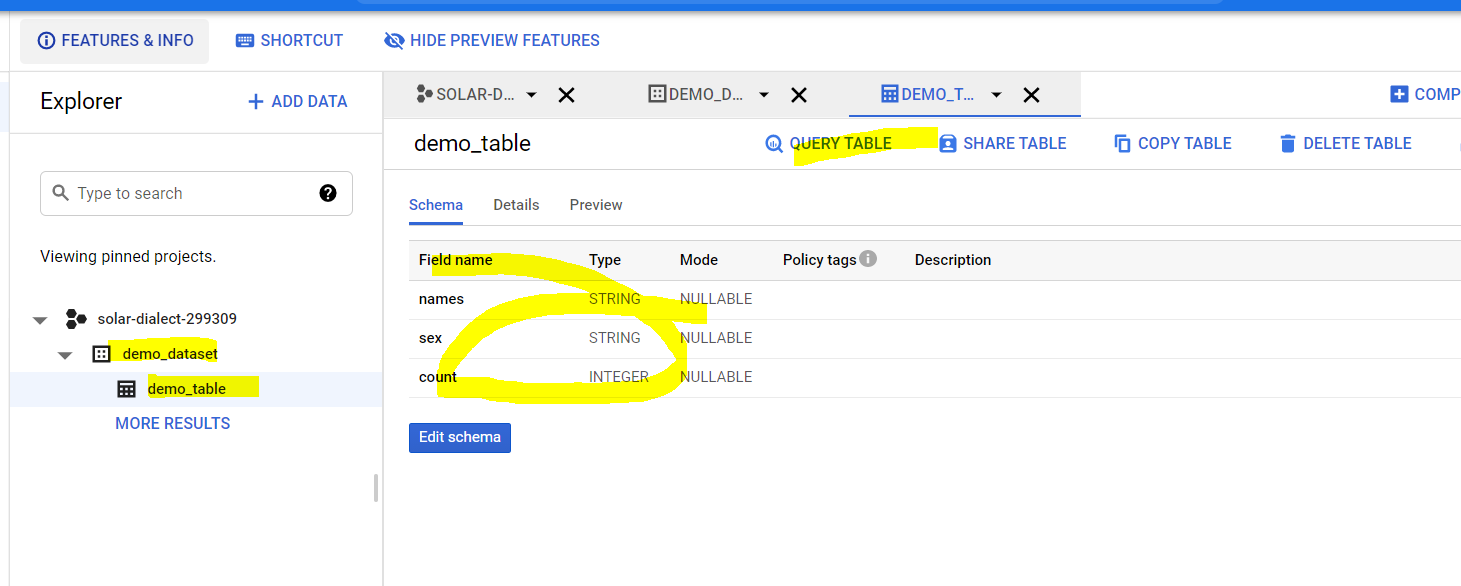
File available

****

You can upload file from local drive or from Google drive or from google cloud storage. Mention project name and dataset name. GCP can auto-detect columns OR you can give inputs by ‘Add filed’. You get advanced options to specify delimiter, header rows to skip etc.

Allow jagged rows- For Jagged rows, check Allow jagged rows to accept rows in CSV files that are missing trailing optional columns. The missing values are treated as nulls. If unchecked, records with missing trailing columns are treated as bad records, and if there are too many bad records, an invalid error is returned in the job result.





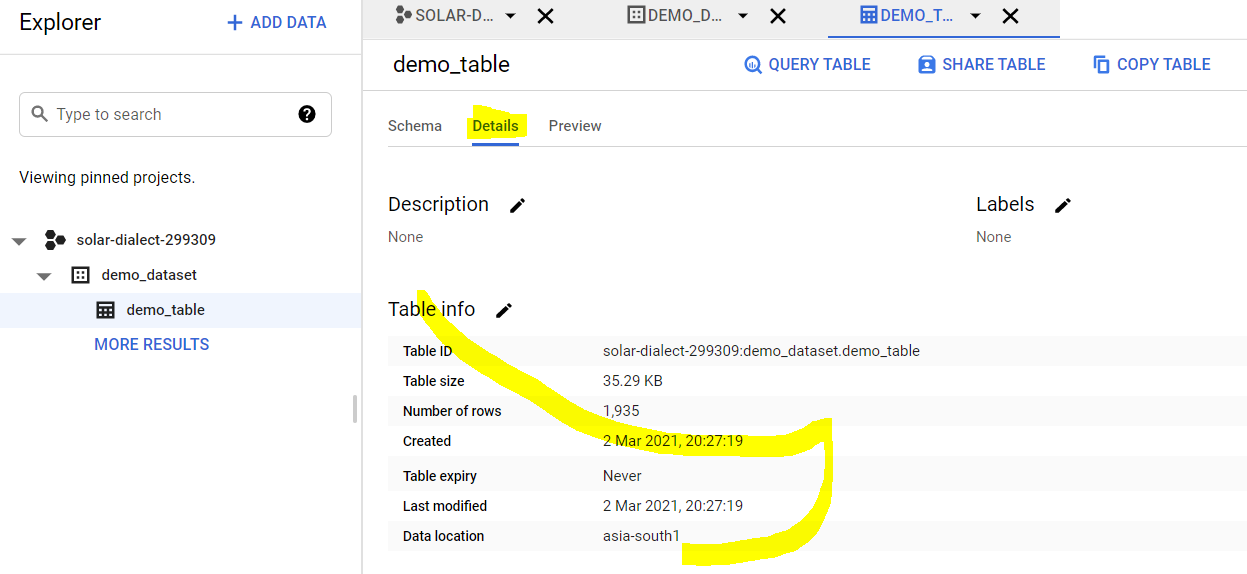


Table creation is free. Storage and querying the table is at cost in GCP.

**Assignment-**

Create a multi-region 'department' table with the attached data while choosing appropriate datatypes.

'department' table should expire after 7 days but other tables in the dataset should never expire.



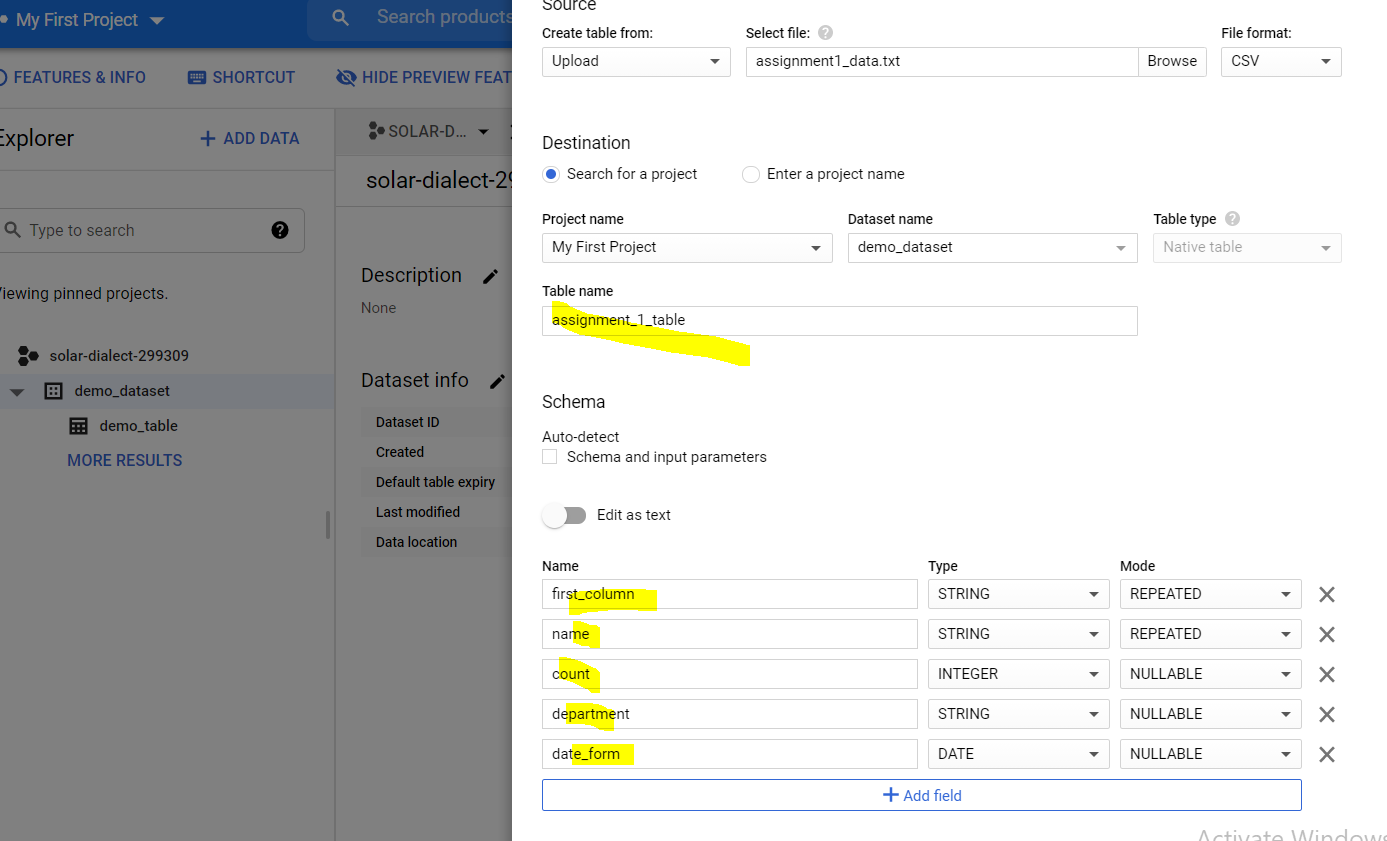
Bigquery should ignore 10 corrupted rows.

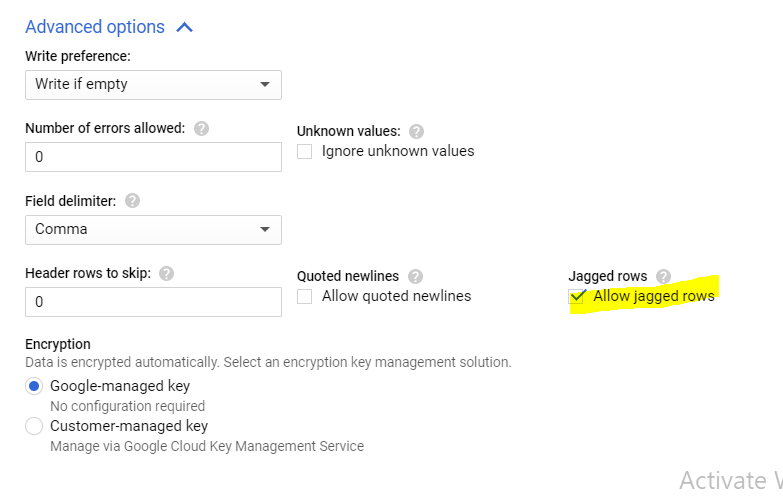
Questions for this assignment

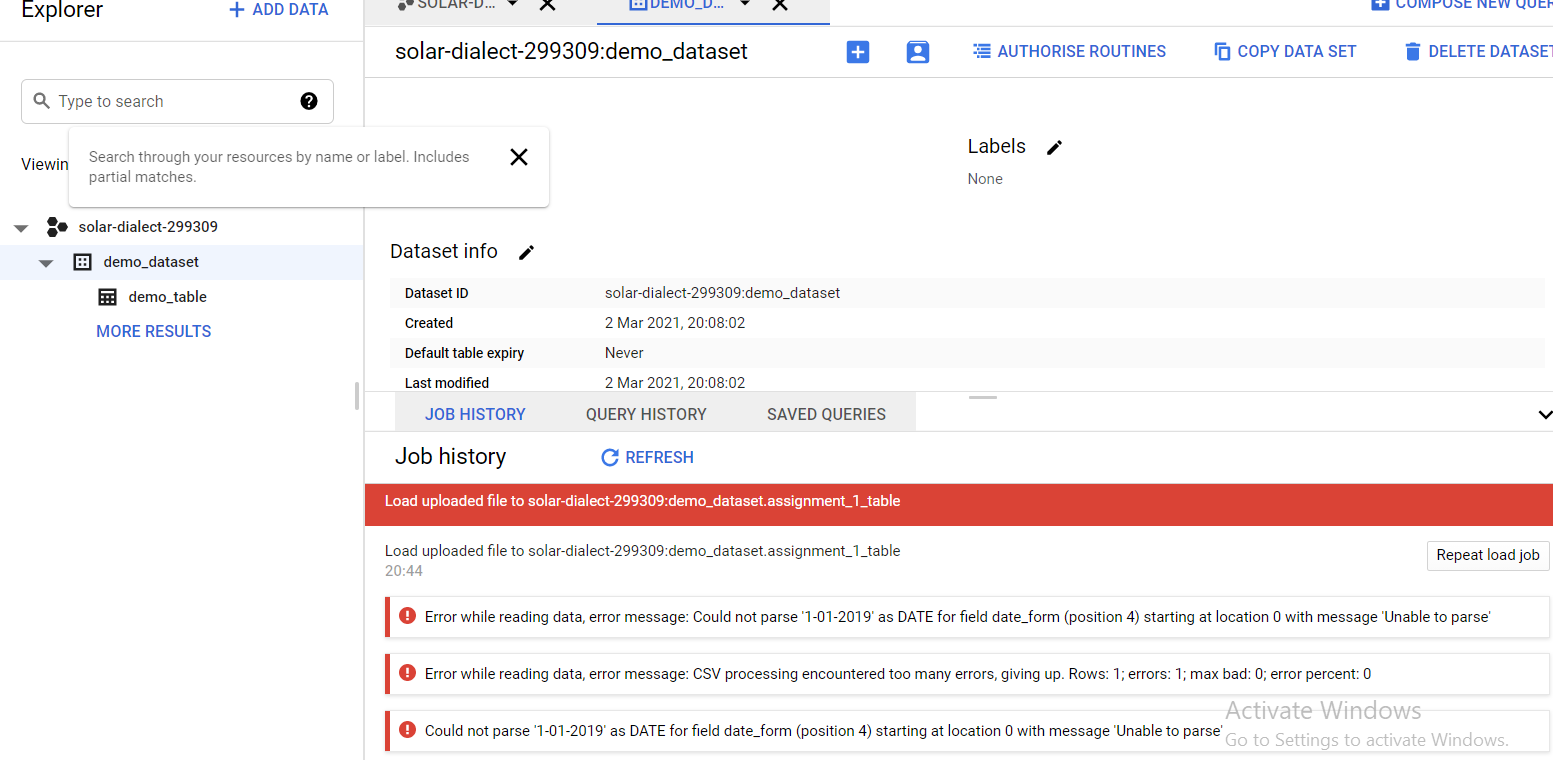
Where would you set the table expiration of 'department' table?

Can you provide 2 columns in the table with first name as "dept" and other column name as "DEPT"?

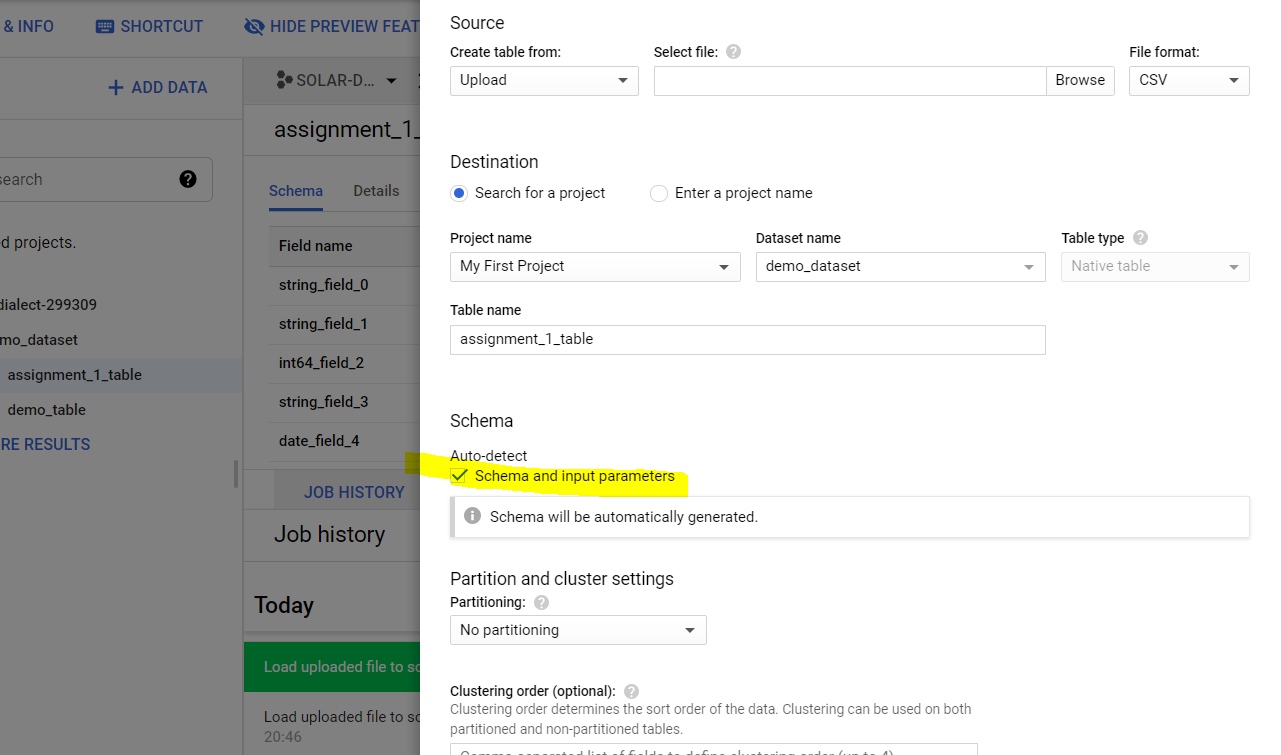
I manually gave column names and datatype and when table is created, job failed with error.

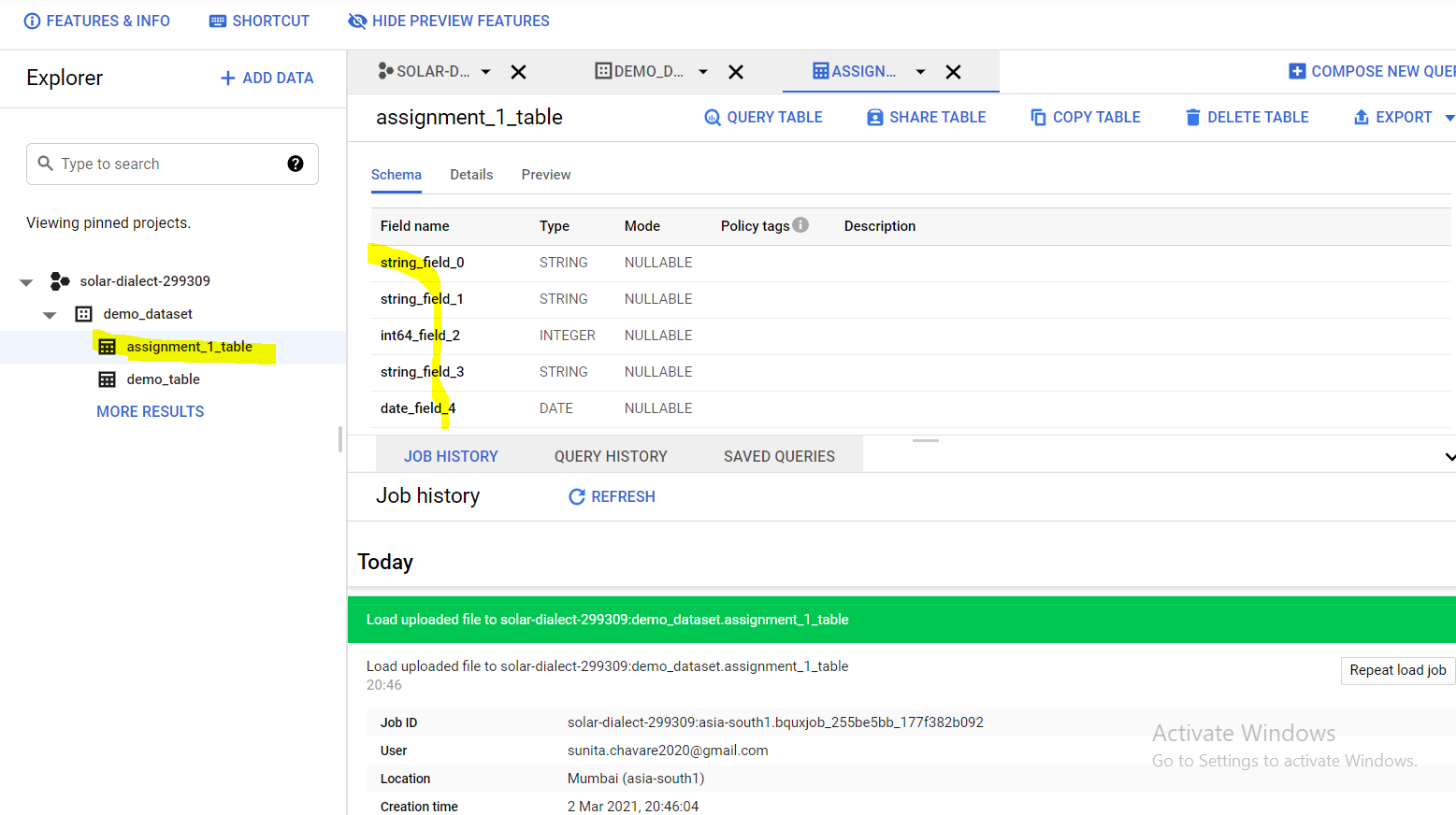






When I clicked on AUTO-DETECT, GCP automatically names columns and decided datatype and created table.





Why it was failing- GCP takes one particular date format. To load data, YYYY-MM-DD is supported. (https://sql.info/d/bigquery-load-error-could-not-parse-mm-dd-yyyy-as-date-for-field-shipped-from-csv-file)