**Lab Exercise 11– dbt Seed to Load and Transform from Multiple CSV Files**

This exercise will guide you through the steps to:

1. Load multiple CSV files as seeds in **dbt**.
2. Use these seed files in a **dbt model** for data transformation.

**Scenario**

Let's assume you have two CSV files:

* **customers.csv**: Contains customer information.
* **orders.csv**: Contains order details associated with customers.

The goal is to:

* Load these two CSV files as seeds.
* Transform the data by joining customers with orders to create a new model that shows customer details along with their order information.

**Steps:**

**Step 1: Prepare Your CSV Files**

Create the following CSV files in seed folder in dbt project.

**1. customers.csv:**

This file contains information about customers.

**customers.csv**:

customer\_id,name,email,country\_code

1,John Doe,john.doe@example.com,US

2,Jane Smith,jane.smith@example.com,CA

3,Alice Johnson,alice.johnson@example.com,MX

4,Bob Lee,bob.lee@example.com,FR

5,Charlie Brown,charlie.brown@example.com,DE

**2. orders.csv:**

This file contains order information, and links to customers via customer\_id.

**orders.csv**:

order\_id,customer\_id,order\_date,amount

101,1,2024-01-15,100.50

102,2,2024-02-17,250.75

103,3,2024-03-01,300.00

105,5,2024-04-05,80.00

106,6,2024-04-05,80.00

107,7,2024-04-05,80.00

**Step 3: Configure dbt\_project.yml File**

Ensure your dbt\_project.yml file is correctly configured to look for the CSV files in the **data** folder.

Example dbt\_project.yml:

name: 'my\_dbt\_project'

version: '1.0'

profile: 'my\_snowflake\_profile'

# The path to the folder containing seed files

data-paths: ["data"]

# Seed configurations

seeds:

my\_dbt\_project:

customers:

+materialized: table

orders:

+materialized: table

This ensures that dbt will load both customers.csv and orders.csv into Snowflake as tables.

**Step 3: Run dbt seed**

Now, run the dbt seed command to load the data into Snowflake.

1. In **dbt Cloud**, go to the **Seeds** tab.
2. You should see both customers.csv and orders.csv listed.
3. Click **Run** next to each seed to load the data into Snowflake.

Alternatively, you can run the following command in the **dbt Cloud Development Environment**:

dbt seed

This will create two tables in your Snowflake schema:

* customers (from customers.csv)
* orders (from orders.csv)

**Step 5: Create the Model to Transform Data**

Now, let’s create a **dbt model** that joins customers and orders to get a detailed view of the customers' orders.

1. In the **models** folder, create a new file named customer\_order\_summary.sql.
2. Add the following SQL code to the ***customer\_order\_summary.sql*** model:

-- models/customer\_order\_summary.sql

SELECT

c.customer\_id,

c.name AS customer\_name,

c.email,

c.country\_code,

o.order\_id,

o.order\_date,

o.amount

FROM {{ ref('customers') }} AS c

LEFT JOIN {{ ref('orders') }} AS o

ON c.customer\_id = o.customer\_id

ORDER BY o.order\_date;

In this SQL model:

* We're using {{ ref('customers') }} and {{ ref('orders') }} to reference the seed tables we loaded.
* We're joining the customers table with the orders table on the customer\_id column.
* The result will contain customer details along with the order information, sorted by order\_date.

**Step 6: Run the Model**

Once the model is created, run it to create the new transformed table.

1. Go to the **Run** tab in **dbt Cloud**.
2. Click **Run** to execute the model and create the transformed customer\_order\_summary table or view.

Alternatively, run the following command from the **dbt Cloud Development Environment** or **local terminal**:

dbt run -m customer\_order\_summary

This will create the customer\_order\_summary table or view in Snowflake.

**Step 7: Verify the Results**

After running the model, query the resulting table or view in Snowflake to verify the transformed data.

SELECT \* FROM my\_schema.customer\_order\_summary;

You should see the following output:

| **customer\_id** | **customer\_name** | **email** | **country\_code** | **order\_id** | **order\_date** | **amount** |
| --- | --- | --- | --- | --- | --- | --- |
| 1 | John Doe | john.doe@example.com | US | 101 | 2024-01-15 | 100.50 |
| 2 | Jane Smith | jane.smith@example.com | CA | 102 | 2024-02-17 | 250.75 |
| 3 | Alice Johnson | alice.johnson@example.com | MX | 103 | 2024-03-01 | 300.00 |
| 4 | Bob Lee | bob.lee@example.com | FR | 104 | 2024-03-18 | 150.25 |
| 5 | Charlie Brown | charlie.brown@example.com | DE | 105 | 2024-04-05 | 80.00 |