**Objective**:

To assess your knowledge on

1. How to connect to Snowflake from dbt
2. Creating models in dbt
3. Configuring sources in dbt
4. SQL and Analytical Skills

**Data for this hands-on:**

The data that we are going to use for this hands-on is look ecommerce data. A synthetic data provided by google. <https://www.kaggle.com/datasets/daichiuchigashima/thelook-ecommerce>

You will download 7 tables, but we will be using only below 4 tables

1. Users
2. Products
3. Orders
4. Order Items.

**Step 1: Data Loading**

Create a database ELOOK\_COMMERCE and create a Schema RAW

Create 4 tables Orders, Order\_Items, Users and Products by loading the data into snowflake by using **Add Data** option in snowflake. Look into below link to know more about how to load data into snowflake.

<https://docs.snowflake.com/en/user-guide/data-load-web-ui#create-a-table-when-loading-a-file>

Note: Don’t do anything in Table Definition Dialog, just click **load**.

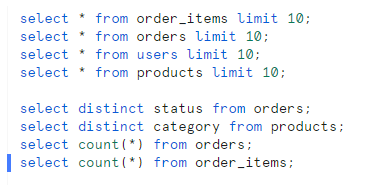
**Step 2: Understanding the data**

Order\_Items is related to Products, Orders and Users by Product\_Id, Order\_Id and User\_Id respectively.

Similarly, Orders is connected to Users by User\_Id column.

For this hands-on Let us assume below primary keys

1. Users – Id
2. Products – product\_id
3. Orders – order\_id
4. Order\_Items – order\_id, product\_id

Understand the data by using following queries, below is just an example.

**Step 3: Configuring DBT**

Create a project in dbt,

1. Name your project in the format <your\_first\_name>\_<last\_name>\_hands\_on\_1
2. Configure Snowflake Connection. Configure the project to create your models in Database and schema of your wish. (Make it outside of ELOOK\_COMMERCE database)
3. Configure the github connection
4. Configure the sources for above 4 tables which we created in above step.

**Step 4: Staging Layer**

Staging Layer: 4 models

1. Stg\_Orders
2. Stg\_Order\_Items
3. Stg\_Products
4. Stg\_Users

Select All the columns from all the table except **Inventory\_Item\_id** from **order\_Items** table in this staging layer.

Also, **convert all the timestamp columns** like created\_at, shipped\_at from varchar to timestamp.

***Note***: You Might need to use some string functions and type casting functions. Get help from your friend google how to do this conversion if you want.

**Step 5: Cleaning Layer**

You are allowed to create any number of layers / models between Staging Layer and Cleaning Layer

Cleaning Layer: 1 model

1. Cleaned\_Order\_Items

One Order\_Id and Product\_Id combination should have only one record in Order\_Items table. If there are multiple records available for same Order\_Id and Product\_Id, you should take only one record with latest Created\_At date.

Materialize this model as table.

Extra Trial: Try to materialize stg\_order\_items as

***Note***: You might need to use rank function for this issue.

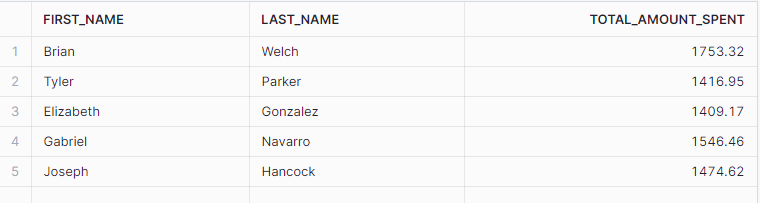
Tip: Break Stg\_Order\_Items into two models, one with no duplicate records and another with duplicate records, Clean the duplicate records and combine both the models.

**Step 6: Final Layer**

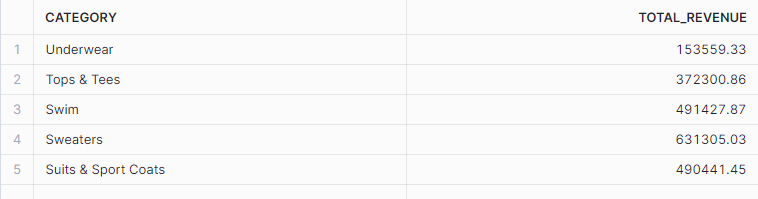
Again, you are allowed to create any number of models between Cleaning Layer and Final Layer

Final Layer: 2 models

1. Top\_5\_Users:
   1. Find the top 5 users who spent more money (consider Sale\_Price from order\_items table) in our company. **Orders should not be in Cancelled or Returned status.**
   2. Three columns should be there in this model. FirstName, LastName, TotalAmountSpent



1. Top\_5\_Category:
   1. Find the top 5 product category which get us more money (consider Sale\_Price from order\_items table). Again, **Orders should not be in Cancelled or Returned state.**
   2. Two columns should be there in this model. Category, TotalRevenue.



Output:

You have to provide a word document with below screenshots.

1. All 4 stage models
2. One Cleaned Model
3. Two Final Models
4. Success screenshot of dbt run command
5. Screenshot of DAG after running dbt docs generate command.
6. Two screenshots with the output of Snowflake tables **Top\_5\_Users, Top\_5\_Category.**

\*\*\*Make sure all the models screenshot have project name on its left side\*\*\*.