Assignment #2

Assignment Deliverables:

Submit your assignment as a single PDF document on Canvas that has the answers (SQL code and output) for the requirements listed below. Use the following file name format: **lastname_Assignment_2.pdf**.
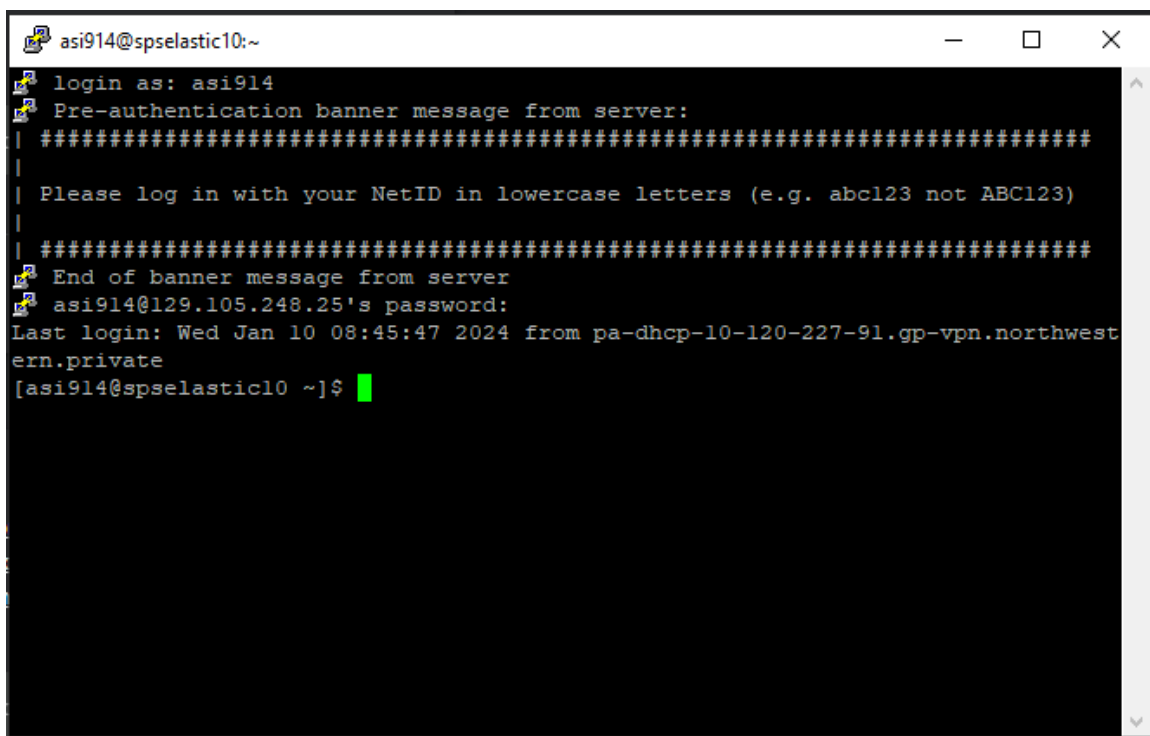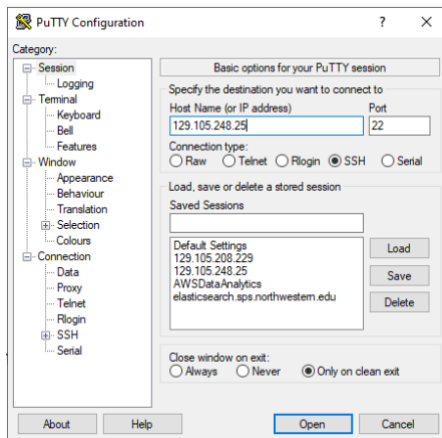
Assignment Objectives:

Use SQL to answer ad-hoc queries for the SaleCO data warehouse with the star schema represented by the fact table dwdaysalesfact and the dimension tables represented by: dwcustomer,dwproduct, dwregion, and dwtime.

Assignment Instructions:

1. Establish a connection with the SaleCO data warehouse by following these steps:
    a. Use GlobalProtect to connect to the NU VPN.
    b. Connect to the server (IP address: 129.105.248.25) using an SSH client such a PuTTY (for Windows) or terminal.app (for Mac).

    Connecting with PUTTy:
        a. Open PuTTY.
        b. In the "Host Name" field, enter: 129.105.248.25.
        c. Select the SSH protocol.
        d. Click "Open".
        e. A terminal window will open; enter your NetID and password when prompted.

Connecting with the Terminal:

a. Open the Terminal.

b. Enter the following command:

- ssh YourNetID@129.105.248.25

2. After successfully connecting to the server via SSH, log in to the PostgreSQL server by executing the command (Replacing "YourNetID" with your own NetID):

- psql -h 129.105.248.26 -U YourNetID -d saleco_dw

```
EdwardArroyo — eas141@spselastic10:~ — ssh -x 129.105.248.25 -l eas141 — 98×14
(base) →  ~ git:(master) ✗ !7274
(base) →  ~ git:(master) ✗ ssh -x 129.105.248.25 -l eas141
############################################################################

Please log in with your NetID in lowercase letters (e.g. abc123 not ABC123)

############################################################################
eas141@129.105.248.25's password:
Last login: Thu May 18 12:03:50 2023 from pa-dhcp-10-120-224-158.gp-vpn.northwestern.private
[eas141@spselastic10 ~]$  psql -h 129.105.248.26 -U eas141 -d saleco_dw
psql (10.14, server 10.15)
Type "help" for help.

saleco_dw=>
```

3. *You can list the available tables in the data warehouse by using the command "\dt".*

```
EdwardArroyo — eas141@spselastic10:~ — ssh -x 129.105.248.25 -l eas141 — 98×14
saleco_dw=> \dt
            List of relations
 Schema |      Name       | Type  | Owner
--------+-----------------+-------+--------
 public | dwcustomer      | table | ajb254
 public | dwdaysalesfact  | table | ajb254
 public | dwproduct       | table | ajb254
 public | dwregion        | table | ajb254
 public | dwtime          | table | ajb254
 public | dwvendor        | table | ajb254
(6 rows)

saleco_dw=>
```

Assignment Requirements:

Use the SaleCO data warehouse to answer the following queries. Submit a pdf document with filename of **lastname_Assignment_2.pdf**. Make sure to show the code and the output of every question.

1. List the total sales by region and customer. Your output should be sorted by region name and customer code.
2. Repeat #1 but produce the output using ROLLUP with region name and customer code.
3. Repeat #1 but product the output using CUBE with region name and customer code. (**2 pts**)
4. a) Explain the additional information/intelligence gained when using ROLLUP or CUBE (**5 pts**) b) Use the output from questions 1, 2 and 3 to explain what the data reveals. (**5 pts**)
5. List the total sales by customer code, month, and product code. Sort by customer code and month. (**5 pts**)
6. Show all purchases (total sales) in September to show which customer bought the most product in September. Show customer code, customer name and total sales; sort all output by total sales with the highest sales on top. (**5 pts**).
7. List the total sales by month and product category. Your output should be sorted by month and product category. (**10 pts**)
8. List the number of product sales (number of rows) and total sales by month. Your output should be sorted by month and should show one row per month. (**10 pts**)
9. Show product category, product code, product description and units sold (sum). Which product is the best seller based on units sold? a) Show units sold for September (**5 pts**), b) Show units sold for October (**5 pts**).
10. List the number of product sales (number of rows) and total sales by month, product category, and product. Your output should be sorted by month, product category and product. (**10 pts**)
11. List the top 5 vendors based on the total sales of their products. Show both the vendors' names and the total sales of their product. Sort by total sales. (**10 pts**)
12. List the products that have not been sold in the year 2015. Show the product code, the product description, and the product category. (**10 pts**)
13. Find the top-selling products in each region based on the number of units sold. Show the region names, product description sand total units sold. Order by region name and total units sold (from largest to smallest). **(10 pts)**

ERD Diagram: