

# DTSC660: Data and Database Management with SQL

## Module 7

### Assignment 6

#### Purpose

For this assignment you have been asked to look into some consumer sales trends and to perform some analysis. For this assignment you will rely on the skills obtained from Modules 5 and 7 to successfully create the banking database and query it to retrieve the appropriate responses. Note that many, but not necessarily all, of the tools you learned in Modules 5 and 7 will be applied in this assignment. To complete this assignment, download and import the dataset and then create queries that respond to each prompt. Please make sure that you only use postgres language conventions. **Each question is all or nothing. Graders will not attempt to correct or interpret malformed SQL queries.** You will be responsible for testing your code on the provided data set before submission. The question will be graded based on whether or not it generates the correct output and addresses all requirements specified in the question. Extraneous columns will not count against you as long as correct results are obtained.

#### Submission

You will submit a total of **1** sql files to CodeGrade. Files should be named appropriately and be in .sql format. Each file must use the postgres standards taught in the course. Use of other SQL languages such as T-SQL will result in an automatic 0 for the assignment. **Each question is all or nothing. Ensure your file runs in its entirety in pgAdmin. This means ensuring each query ends in a semicolon ( ; ). Graders will not attempt to correct or interpret malformed SQL queries.** You will have one submission attempt for this assignment.

- **File 1:** You must submit a SQL document called <LastName>\_Assignment6. This document must include ALL queries requested in the instructions below.
- You will submit the file to the Assignment 6 link to CodeGrade.

#### Instructions

As in the practice assignments you will be querying a large dataset to gather insights about that data. You have been provided a SQL file with all the necessary data. Make sure you take a moment to familiarize yourself with the columns, their data types, and any other pertinent information. I would recommend printing out a query with the first few records using the following command to help you get started. It is expected that if you have questions or difficulties with any portion of this assignment that you utilize the assignment discussion board or email the GAs to gain clarity (dtsc\_ga\_660@eastern.edu).

## *PART 1 Creating the Table and Importing the Data*

1. In the assignment 6 folder, download the **customer\_spending.csv** file
2. Place this file in a public folder on your computer
3. Take note of the path to this file (copy the path)
4. Utilize the COPY command you learned in Module 2 (revisit this module if necessary) to import the data.
5. Run a basic select statement that verifies the data is present and matches what is in the csv file.

## *Part 2 (Queries)*

To complete this part, download the assignment\_6\_template.sql file from the assignment 6 folder. Rename this file using the naming convention: <LastName>\_Assignment6. Complete each query in the identified space in this document. Once you are done, submit the document to the Assignment 6 folder.

Please note that some questions are intentionally vague. Remember that you will not be penalized for extraneous columns, but you will be for extraneous rows (data). There may be multiple correct ways to solve some queries. Note that a request for a list of values always expects no duplicate values unless specified otherwise.

1. Write a query that returns the top selling (revenue) categories for the year 2016. Include the category name and total revenue and organize alphabetically by category name.
2. Write a query that returns a list of sub categories and their average unit price, average unit cost, and the difference between these values called margin for the year 2015. Organize the results alphabetically by sub category.
3. Write a query that returns the total number of female buyers who purchased clothes.
4. Write a query that provides the average quantity and average cost of products purchased by each age and by sub category. Organize the columns with age, sub category, average quantity, and average cost. Make the data organized by age oldest to youngest and then by sub category alphabetically.
5. Write a query that returns a list of countries that had more than 30 transactions to customers between the ages of 18-25 in the entire data set.
6. Write a query that returns a list of sub categories, the average quantity and average cost both rounded to 2 decimal places. Only include sub categories that have at least 10 records in the data set. Organize the data by sub category alphabetically.

\*\*\*\*\*GRADING RUBRIC ON NEXT PAGE\*\*\*\*\*

This assignment will be graded on the following rubric. Remember that questions are ALL OR NOTHING. Incorrect syntax, extraneous results, or incorrectly addressing all question requirements will result in loss of points for that question. Graders will NOT attempt to correct malformed sql code. :

Question Number	Points
Creating Table	5
Importing Data	5
1	10
2	10
3	15
4	15
5	20
6	20
<b>Total</b>	100