CS4048 – Data Science
Assignment 3
BSE7A – BCS7C

Submission Guidelines

- Create a PDF report on all the questions given below with query and result set.
- Strictly follow the SQL formatting guidelines shared with you to avoid negative marking.
- Turn in the solution to google classroom.
- Keep the naming convention of report and query solutions as f20xxxx_sec_A3.pdf/sql.
- Submit all solution queries to just one .sql file.
- Naming conventions are required to be followed strictly to avoid any inconvenience.
- **Deadline: Friday, September 22, 2023, 4:00 Sharp.**

You are already familiar with the Parch Posey paper company's data. In this assignment you will be working on this data to answer the following questions.

Question 1

Use the web_events table to find all information regarding individuals who were contacted via the organic or adwords channels, and started their account at any point in 2016, sorted from newest to oldest.

Question 2

Provide the name for each region for every order, as well as the account name and the unit price they paid (total_amt_usd/total) for the order. Your final table should have 3 columns: region name, account name, and unit price. A few accounts have 0 for total, so divide it by (total + 0.01) to assure not dividing by zero.

Question 3

Provide a table that provides the region for each sales_rep along with their associated accounts. This time only for the Midwest region. Your final table should include three columns: the region name, the sales rep name, and the account name. Sort the accounts alphabetically (A-Z) according to account name.

Question 4

Provide a table that provides the region for each sales_rep along with their associated accounts. This time only for accounts where the sales rep has a first name starting with S and in the Midwest region. Your final table should include three columns: the region name, the sales rep name, and the account name. Sort the accounts alphabetically (A-Z) according to account name.

Question 5

Provide a table that provides the region for each sales_rep along with their associated accounts. This time only for accounts where the sales rep has a last name starting with K and in the Midwest region. Your final table should include three columns: the region name, the sales rep name, and the account name. Sort the accounts alphabetically (A-Z) according to account name.

Question 6

Provide the name for each region for every order, as well as the account name and the unit price they paid (total_amt_usd/total) for the order. However, you should only provide the results if the standard order quantity exceeds 100. Your final table should have 3 columns: region name, account name, and unit price. In order to avoid a division by zero error, adding .01 to the denominator here is helpful total_amt_usd/(total+0.01).

Question 7

Provide the name for each region for every order, as well as the account name and the unit price they paid (total_amt_usd/total) for the order. However, you should only provide the results if the standard order quantity exceeds 100 and the poster order quantity exceeds 50. Your final table should have 3 columns: region name, account name, and unit price. Sort for the smallest unit price first.

Question 8

What are the different channels used by account id 1001? Your final table should have only 2 columns: account name and the different channels.

Question 9

Find all the orders that occurred in 2015. Your final table should have 4 columns: occurred_at, account name, order total, and order total_amt_usd.

Question 10

Find the total number of events happened for web for all the accounts. Your result set should have three columns account name, channel, and # of events.

Question 11

Determine the number of times a particular channel was used in the web_events table for each sales rep. Your final table should have three columns - the name of the sales rep, the channel, and the number of occurrences. Order your table with the highest number of occurrences first.

Question 12

Using Having clause with aggregations

a. How many of the sales reps have more than 5 accounts that they manage?
b. Which account has the most orders?
c. Which accounts spent more than 30,000 usd total across all orders?
d.Which account has spent the most with us?
e.Which account has spent the least with us?
f.Which accounts used facebook as a channel to contact customers more than 6 times?
g.Which account used facebook most as a channel?

Question 13

Write a query to display for each order, the account ID, total amount of the order, and the level of the order - ‘Large’ or ’Small’ - depending on if the order is $300 or more, or smaller than $3000.

Question 14
Write a query to display the number of orders in each of three categories, based on the total number of items in each order. The three categories are: 'At Least 2000', 'Between 1000 and 2000' and 'Less than 1000'

Question 15
Count the orders based on the categories you defined in previous question. Your table must have tow column, category and count of orders for each category

Question 16
We would now like to perform a similar calculation to the first, but we want to obtain the total amount spent by customers only in 2016 and 2017. Keep the same levels as in the previous question. Order with the top spending customers listed first.

Question 17
We would like to identify top performing sales reps, which are sales reps associated with more than 200 orders. Create a table with the sales rep name, the total number of orders, and a column with top or not depending on if they have more than 200 orders. Place the top sales people first in your final table.

Question 18
Find the number of events that happened each day for each channel. You query must return event_day, channel, and count in the result set.

Question 19
For the region with the largest sales total_amt_usd, how many total orders were placed?

Question 20
For the region with the largest sales total_amt_usd, how many total orders were placed? Write query using CTE for this same problem as previous.