## Career Foundry

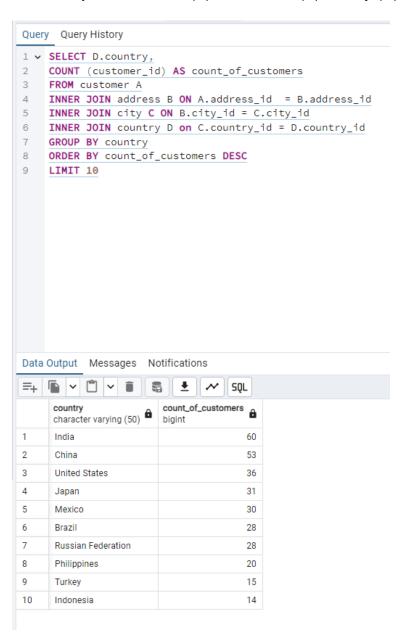
# **Data Analytics Immersion**

#### A3.E7

## Kendra Jackson

## Step 1:

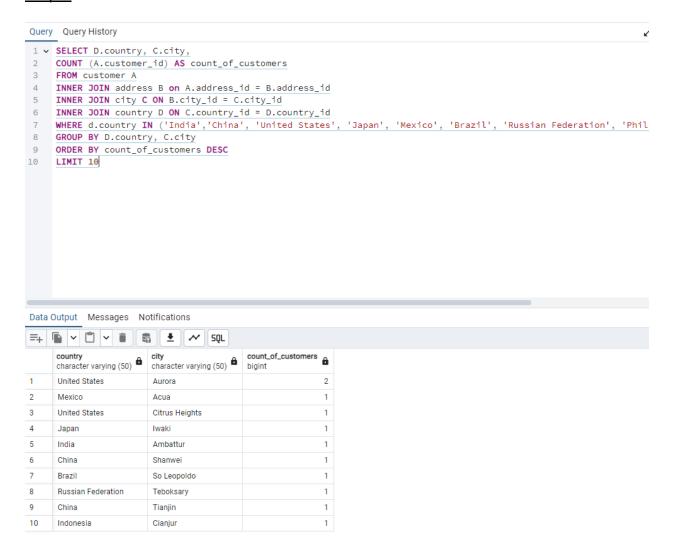
Tables to join: Customer (A)  $\rightarrow$  Address (B)  $\rightarrow$  City (C)  $\rightarrow$  Country (D)



#### Rationale:

Using the ERD document, in order to retrieve the information asked by management, you can see that in order to link a customer to a country, since there is no direct link, a multiple join must be used. In order to create the link a chain between the tables must be formed from customer to address to city then country. Management had asked for a information about which countries have the largest amount of customers. This means we need to count how many customers are in a country this is done by counting the customer\_id column which I then aliased to make the results more understandable. Then, we must group by country to ensure the count of the customers is grouped into its appropriate country. Additionally, they want the top 10 countries which means we must order the results descending (from largest to smallest) and then limit the results by 10. This gives us the results they were looking for.

#### Step 2:



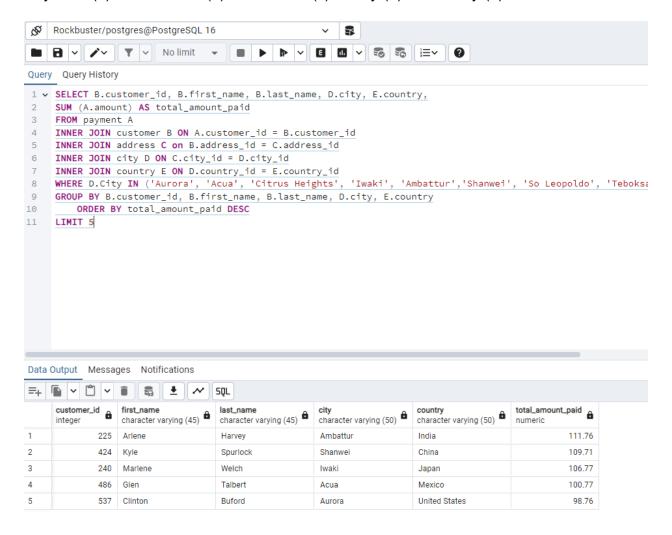
#### Rationale:

For this query, management is seeking to find the top 10 cities within the top 10 counties. As before, we will be using the same 4 tables customer, address, city, and country to form a link between the tables in order to find this information. Since the cities need to be within the top 10 counties, we can use our query from part 1 but need to add to it. We need to know the city, and country so these will appear in the select statement, and a count of customer\_id will occur as we need to know the amount of customers in each place to find which is the largest. After joining the aforementioned tables, a where clause is used to distinguish which counties we want to look for these cities in. The data is then grouped by the city and country and the ordering by the count from largest to

smallest, then limited by 10. With this we reach the conclusion of our top 10 cities with the largest at the top.

### Step 3:

Payment (a)  $\rightarrow$  customer (b)  $\rightarrow$  address (c)  $\rightarrow$  city (d)  $\rightarrow$  country (e)



#### Rationale:

Building on the previous queries, the payment table also needs to be added now. We need to know the customer.id, their first and last names as well as city and country so these will be in our SELECT statement. We also need to know who spends the most this means we will use the sum function on the amount column of the payment table to find this. Then we need to join all the tables by joining B to A, C to B so on and so forth. Since we are looking for the 5 highest paying customers in the top 10 cities, we can do

two things in which a WHERE statement will be needed – we can either write out a WHERE IN statement that has the top 10 cities listed or we can use the previous join statements from the prior exercise. Since join statements are expensive, it is better to write out the city results we found previously. Then we group by the columns in the select statement and order by DESC the aliased name of total\_amount\_paid. The top 5 customers should now appear.