# PostGreSQL Questions - Classification - Caitlin Data

(Use sub queries or views wherever necessary)

1. Create a basic ERD of the main tables/tables you will be using from your database - including lookup tables which you would access using the foreign key and a JOIN

2. Create SQL queries selecting the columns you want to work with in items , customers, buys tables

3. Select all the data from table `buys` for those columns

5. Use the \_alter table\_ command to add a new column to any of your tables. You can hardcode a fixed (dummy) variable or drop and re add a field (optional)

6. Use sql query to find how many rows of data you have in one of your tables

7. Now we will try to find the unique values in some of the categorical columns:

- how many unique values in the column 'postal code'?

- What are the unique values in the column `gender`?

- how many unique values in the column `id` in buys table?

- What are the unique values in the column `points\_return`?

- What are the unique values in the column `material\_id`?

8. Arrange the data in buys table by decreasing order of the `retail price` of the buys. Return only the `customer\_id` of the top 50 customers with the highest `retail price` in your data.

9. What is the average retail price of all the customers in your buys table?

10. In this exercise we will use a simple group by query to check the properties of some of the categorical variables in your item table. Note wherever `offer value` is asked, please take the average of the column `offer value':

- What is the average offer value of the buys? The returned result should have only two columns, offer value and buy id

- What is the average offer value of each brand? The returned result should have only two columns, offer value and brand, with as many rows as brands

- What is the average offer value by customerid? The returned result should have only two columns, customerid (not name) and offer value

- what is the average offer value by month? the returned result will have as many rows as you have months and years in your dataset.

- Is there any obvious correlation re offer value between the columns `material\_id` and `pattern\_id`? You can analyse this by grouping the data by one of the variables and then aggregating the results of the other column. Visually check if there is a positive correlation or negative correlation or no correlation between the variables.

11. Your analysis is only focused on the customers with the following properties:

- have accepted only cash in the past

- have been a customer on more than one occasion

For the rest of the customer profiles, you are not too concerned. Write a simple query that captures such data with columns which may be interesting, in your perspective, to understand the behaviour of such customers.

12. You want to find a list of customers whose average points sum is higher than the average points sum of all the customers in the database. Write a query to show the list of such customers. You might need to use a subquery for this problem.

13. Since this is something that you are regularly interested in, create a view of the same query.

14. What is the number of customers or buys who accepted the points vs number of people who did not?

15. Provide the postcode and gender (not name) details of a customer that is the 15th highest `points\_sum` in your database.

16. Provide the postcode and gender (not name) details of a customer that is the 15th highest `retail\_price` in your database.