# SQL Experimentation

You will find on this sheet on necessary information for the next hour of test.
You have to keep it during the whole exercise, to have easily access if you have any doubt.

If you have any question, don't hesitate to ask one of the organizers.

The objective of this test is not to evaluate you or your competencies : so don't worry about your abilities, just do your best. We only ask you to follow the rules and to answer the questions the best you can. You've been given an unique ID that you need to give whenever you answer a question : it is only here for post-analysis of the results so that we can group your answers.

# Scenario

You're a new member of a post office, in charge of packages. When you're not at the front desk taking care of customers, you have access to data recorded about the packages sent from your post office. For simplification, we will focus on the packages leaving the post office to other destinations.

Data information:
Here is how the database was created :

```
CREATE TABLE Cities(
   city_ID DECIMAL,
   distance DECIMAL,
   PRIMARY KEY (city_id )
)

CREATE TABLE Packages(
      package_ID DECIMAL,
      destination DECIMAL,
      length DECIMAL,
      width DECIMAL,
      height DECIMAL,
      weight DECIMAL,
      price DECIMAL,
      PRIMARY KEY (id_colis)
      FOREIGN KEY (destination) references Villes(id_ville)
)
```

Table *Packages* has one entry per package that left your post office. From the destination of a package, you can see how far it was sent, by joining tables *Packages* and *Cities* on attributes *destination* and *city_ID.*

Tables size :
- Packages: 11 000 tuples
- Cities: 30 tuples

Data units :
- Distance is in kilometers
- Length, width and height are in centimeters
- Weight is in grams
- Price is in euros

Some statistics:

| | Average | Minimum | Maximum |
|---|---|---|---|
| Distance | 1960.7km | 5km | 9749km |
| Length | 63.70cm | 8.00cm | 143.51cm |
| Width | 36.55cm | 8.00cm | 139.04cm |
| Height | 23.85cm | 8.00cm | 141cm |
| Weight | 4266.72g | 413.35g | 9540.33g |
| Price | 60.03€ | 8.03€ | 156.96€ |

(All those statistics can be computed using SQL, and you are encouraged to calculate others if you think it is necessary to answer a question).

Questions are ordered from easiest to hardest :you should therefore answer them in the given order. First part of the experimentation contains simple questions. The second part has voluntary more complex questions, and finding the required SQL query will require more exploration.

# First Part

## Question 1

This first question is here so that you can get familiar with the data and the tools at your disposal.

Please test the two tools (SQL software and online form for answers) with the following query, that is a join between the two tables :

```
Select *
  From Packages, Cities
  Where Packages.destination = Cities.city_ID
```

Expected result size : 10 999 tuples

## Question 2

Maximum size limit authorized for a package is 9000 grams. However, some exceed this limit without being detected. Give the query to obtain the ID of packages whose weight exceed this limit.

Expected result size :  73 tuples

## Question 3

What query can you write to obtain the average length of packages sent less than 100 kilometers from your post office ?

Expected result size : 1 tuple

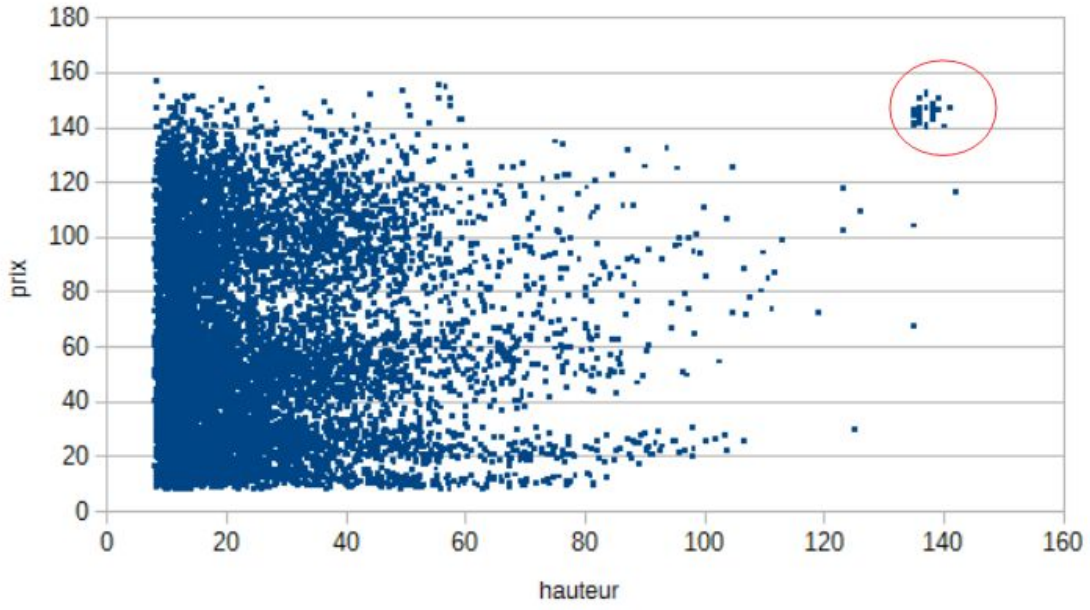
# Second part

Happy to see you've managed those first queries, you go to the weekly team meeting, and ask your colleagues if they have any idea of queries that could be helpful for the daily life of the post office. They are not familiar with SQL, but they have interesting remarks.

## Question 4

A little bit interested by data analysis, a colleague of yours had, with a spreadsheet, visualized some curves from the database.
By plotting packages prices against their height, he/she had noticed a group of packages very distinct and well separated from the others, which is presented on the figured below, and circled in red.
Can you find the query that returns all packages belonging to this group ?

Expected result size : 33 tuples




## Question 5

According to some colleagues who've been working here for years, heaviest packages are the one going to very distant destinations. The intuition behind this is that has sending a package far away is expansive, customers many things in one package to compensate.

Can you identify packages that do not comply with this, i.e that are not heavy but are sent far away ?

Expected result size: 13 tuples

## Question 6

Once at the regional sorting center, packages go through a machine that automatically sort them according to their destination. However, this machine is sometimes defective. Indeed, when a package is less than 480g, the machine does not always detect it, and a operator has to take it and process it manually.

This phenomenon is marginal, but more likely to happen if in addition to its light weight, the packages in small regarding its length and width. On all packages registered in your database, 12 have caused a problem.

Which query can identify those 12 packages ?

## Question 7

Some packages are sent to a city that is very close to your post office, less than 10km away. Moreover, some are very light (less than 550g), and you wonder why people pay the post office to transport them while they would quite easily do it themselves.
One of your colleagues has an hypothesis : maybe those packages are cumbersome and therefore hard to transport.

Can you identify packages validating this hypothesis ?

Expected result size : 8 tuples

## Question 8

A customer arrives at the post office, because he needs the ID of a package he had send, but isn't able to find. In order to help him, he gives you a few informations :

- The package was light, less than 450g
- Its dimensions (mainly length and width) were surprisingly big in regard to its weight

Can you give the query returning such a package ?

Expected result size : 1 tuple

## Question 9

When working at the front desk, one of your colleagues made a mistakes on four on the packages he registered. Luckily, he remembers their length was above 140cm, and he therefore applied a special tarification, as those kind of packages are more complicated to deliver due to their size. But he applied the wrong tarification, and those packages have therefore an abnormally elevated price

Can you identify those packages ?
Expected result size : 4

## Question 10

At question 2, you showed that 73 packages are above the weight limit. But your colleagues in charge of putting packages in the trucks say that a third of packages are really heavy, and require two employees to be lifted, in order to avoid back pains.
Can you modify the query for question 2 in order to identify those packages ?
Expected result size : 3073 tuples