

Expérience scientifique en SQL

Vous trouverez ci-dessous toutes les informations qui vous sont nécessaires pendant cette heure de test.

Vous devez la garder ouverte tout au long de l'exercice, pour avoir rapidement accès aux informations qui vous seront nécessaires.

En cas de questions, n'hésitez pas à nous solliciter.

Ce test n'a pas pour but de vous évaluer vous : par conséquent, ne vous inquiétez pas sur vos compétences. Nous vous demandons simplement de respecter les consignes (voir feuille papier) et de répondre du mieux que vous pouvez. L'identifiant demandé à chaque question est simplement là pour nous aider à regrouper les résultats.

Scénario

Vous venez d'arriver dans un bureau de poste, pour être en charge de la gestion des colis. En dehors des heures d'ouverture au public, vous avez accès aux données enregistrées sur les envois de colis de votre bureau de poste. Pour simplifier le contexte, on ne s'intéresse qu'aux colis qui partent du bureau de poste.

Information sur les données :

Voici les ordres ayant permis la création de la base de données :

```
CREATE TABLE Villes(  
    id_ville DECIMAL,  
    distance DECIMAL,  
    PRIMARY KEY (id_ville)  
)
```

```
CREATE TABLE Colis(  
    id_colis DECIMAL,  
    destination DECIMAL,  
    longueur DECIMAL,  
    largeur DECIMAL,  
    hauteur DECIMAL,  
    poids DECIMAL,  
    prix DECIMAL,  
    PRIMARY KEY (id_colis)  
    FOREIGN KEY (destination) references Villes(id_ville)  
)
```

La table Colis comporte une ligne par colis étant parti de votre bureau de poste. A partir de la destination renseignée, vous pouvez savoir à quelle distance il a été envoyé du bureau de poste, en faisant la jointure entre l'attribut destination de la table Colis, et id_ville de la table Villes.

Taille des tables :

- Colis : 11 000 tuples
- Villes : 30 tuples

L'unité des données :

- La distance est donnée en kilomètres
- La longueur, la largeur et la hauteur sont données en centimètres
- Le poids est donné en grammes
- Le prix est donné en euros

Quelques statistiques :

	Moyenne	Minimum	Maximum
Distance	1960.7km	5km	9749km
Longueur	63.70cm	8.00cm	143.51cm
Largeur	36.55cm	8.00cm	139.04cm
Hauteur	23.85cm	8.00cm	141cm
Poids	4266.72g	413.35g	9540.33g
Prix	60.03€	8.03€	156.96€

(Toutes ces statistiques peuvent être calculées avec SQL, et vous pouvez d'ailleurs en calculer d'autres si vous le jugez nécessaire.)

Les questions sont classées par ordre de difficulté : par conséquent, il vous est demandé d'y répondre dans l'ordre. La première partie concerne des requêtes simples. Dans la seconde partie, les énoncés sont volontairement plus complexes, et trouver la requête voulue peut nécessiter de tâtonner.

Première partie

Question 1

Cette première question n'est là que pour vous aider à vous familiariser avec les outils.

Testez les deux outils à votre disposition (l'appli et ce questionnaire) avec la requête suivante, qui permet de faire la jointure entre les deux tables :

Select *

From Colis, Villes

Where Colis.destination = Villes.id_ville

Taille du résultat attendu : 10 999 tuples

Vous pouvez utiliser cette requête pour tester l'application à votre disposition. Lorsque vous avez terminé d'explorer, vous pouvez la copier dans le champs de réponse, et soumettre ce premier questionnaire.

Question 2

La limite maximale de poids autorisé d'un colis est de 9000 grammes. Cependant, certains échappent à la vigilance des employés. Donnez la référence des colis dont le poids excède strictement cette limite.

Taille du résultat attendu : 73 tuples

Question 3

Quelle est la requête qui permet d'obtenir la longueur moyenne des colis envoyés à moins de 100 km ?

Taille du résultat attendu : 1 tuple

Deuxième partie

Question 4

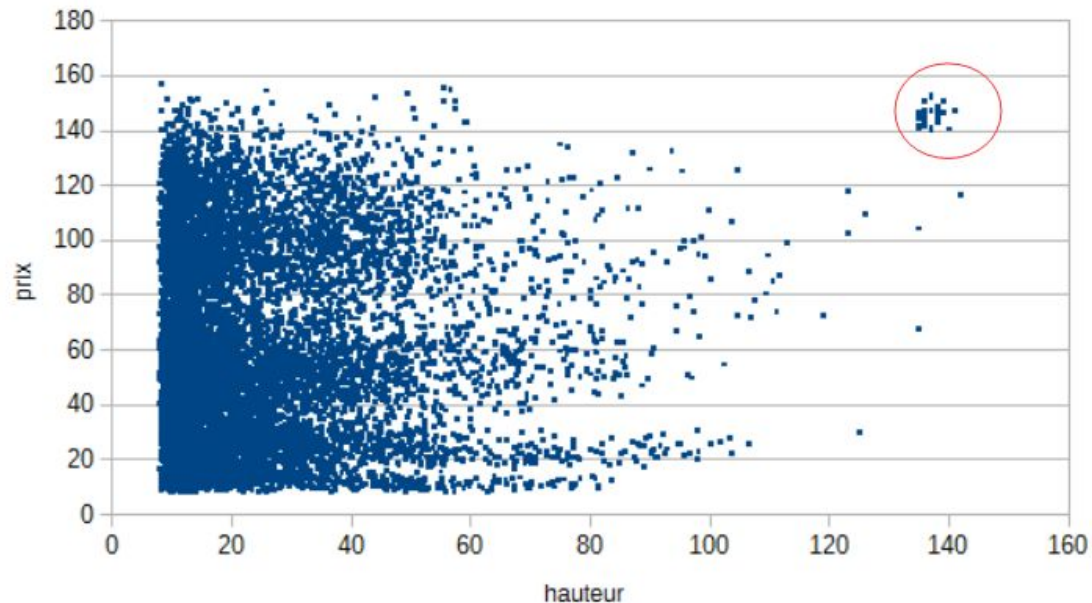
Galvanisé par le succès de ces requêtes de routine, vous en parlez à la prochaine réunion d'équipe, pour savoir si vos collègues auraient des idées de requêtes qui pourraient être intéressantes pour leur métier. Peu familier du SQL, vos collègues ont néanmoins des remarques intéressantes.

Un peu intéressé par les données, l'un.e de vos collègues avait, avec un tableur, essayé de tracer quelques courbes pour visualiser les données de la base.

En traçant la courbe du prix en fonction de la hauteur, il/elle a notamment remarqué un petit groupe de colis qui se détache très nettement des autres, comme on peut le voir dans l'image ci-dessous, entouré en rouge.

Pour pouvoir investiguer sur l'origine de cette nette différence, pouvez-vous identifier la requête qui renvoie les colis (et toutes leurs caractéristiques) appartenant à ce groupe ?

Taille du résultat attendu : 33 tuples



Question 5

D'après des collègues, en poste depuis plusieurs années, les colis les plus lourds partent pour les destinations lointaines. L'intuition serait que pour compenser le prix plus élevé des destinations lointaines, les clients chargent plus leurs colis.

Pouvez-vous identifier les colis qui ne satisfont pas cette remarque, à savoir qui ne sont pas lourds mais partent pour des destinations lointaines ?

Taille du résultat attendu : 13 tuples

Question 6

Une fois au centre de tri, tous les colis passent dans une machine permettant de les trier automatiquement, qui peut parfois être défectueuse. Ainsi, lorsqu'un colis pèse moins de 480g, la machine ne parvient pas toujours à détecter le passage de ce colis et un opérateur devra donc le récupérer et le prendre en charge manuellement.

Ce phénomène reste marginal, mais est d'autant plus probable que les dimensions du colis sont petites (longueur ou largeur principalement). Sur les colis enregistrés dans votre bases, douze ont posé soucis, et d'après la personne qui les a pris en charge, c'était à cause de leur largeur.

Pouvez-vous identifier ces 12 colis ?

Question 7

Certains colis sont envoyés dans une ville très proche de votre bureau de poste, à moins de 10km. Plus étonnant, certains sont assez légers (moins de 550g), et vous vous demandez pourquoi certaines personnes préfèrent donc les faire transporter par la poste plutôt que de les emmener eux-mêmes. Une collègue a une hypothèse : certains de ces colis sont peut-être encombrants car très larges et donc compliqués à transporter. Pouvez-vous trouver des colis validant cette hypothèse ?

Nombre de tuples attendus : 8 tuples

Question 8

Un client arrive au bureau de poste car il a perdu la référence de son colis, et en a besoin pour surveiller le suivi. Il n'a pour vous aider à la retrouver que quelques informations :

- Le colis était très léger, son contenu pesant moins de 450g
- Ses dimensions (surtout longueur et largeur) étaient étonnamment grandes pour son faible poids

Pouvez-vous donner la requête permettant d'identifier ce colis ?

Taille du résultat : 1 tuple

Question 9

L'un de vos collègues a commis une erreur sur quatre colis qu'il a été chargé d'enregistrer. Par chance, il s'en rappelle bien car ils mesuraient plus de 140cm de longueur, et il a appliqué une tarification spéciale, puisqu'ils étaient difficile à transporter de par leur taille. Mais il a appliqué une mauvaise tarification, et les colis ont donc un prix anormalement élevé. Pouvez-vous identifier ces colis ?

Nombre de tuples attendu : 4

Question 10

Vous avez montré à la question 2 que la base a enregistré 73 colis dont le poids dépasse la limite de poids autorisée, ce qui reste une minorité. Vos collègues chargés du stockage des colis dans l'entrepôt sont formels : un tiers des colis en moyenne leur paraît vraiment lourd et nécessite de travailler à deux pour ne pas se faire mal au dos. Pouvez-vous modifier la requête trouvée à la question 2 pour identifier ces colis "lourds" ?

Taille du résultat attendu : 3073 tuples