**DECLERCQ Tifaine**

**GOLLIOT Camille**

SAE 2.04

Exploitation d’une base de données

Partie STATISTIQUES 

Année 2023 - 2024

SOMMAIRE

[**PREAMBULE**](#_2hip6lqy9lw7) **2**

[**PARTIE 1**](#_1cjqwnaxngu9) **3**

[Sujet :](#_qc28yyvo9wkg) 3

[Région choisie :](#_lagb3kd4b3oo) 3

[1- Les filles n’aiment pas les études scientifiques.](#_mw93fvp3z6fi) 4

[2- Les bacs technologiques ont plus de chances que les bacs généraux d'être acceptés en filières sélectives.](#_mvcwq9mjf5nj) 7

[3- Les boursiers préfèrent les études courtes.](#_qtvntdewhyy1) 9

[**PARTIE 2**](#_55a72lwgogk5) **12**

[Sujet :](#_lpxsou27ylwr) 12

[Établissements choisit :](#_m07kmw4h5yyt) 13

[1. Taux de filles qui ont candidaté & taux d’admis avec mention](#_9wx5p1fx7ifo) 13

[2. Du taux de BTN, BG et autres classés](#_qqmdtimhm6r7) 17

[3. Tout autre critère qui vous semble pertinent](#_4hsjpogwk33b) 21

# 

# PREAMBULE

Ce rapport concentre toutes les requêtes et analyses de la SAE 2.04 de Base de Données du côté STATS.

Rapport, analyses et recherches faites par DECLERCQ Tifaine et GOLLIOT Camille.

Nous savions que les requêtes étaient à faire avec les tables ventilées, mais ayant eu plusieurs problèmes avec lesdites tables, nous avons préféré faire les requêtes mais avec la table import pour pouvoir rendre notre rapport à l'heure et complet.

# PARTIE 1

## Sujet :

À l’aide des données de votre BDD, vous discuterez des trois assertions suivantes :

1- Les filles n’aiment pas les études scientifiques.

2- Les bacs technologiques ont plus de chances que les bacs généraux d'être acceptés en filières sélectives.

3- Les boursiers préfèrent les études courtes.

## Région choisie :

Hauts-de-France

--------------------------------------------------------------------

## 1- Les filles n’aiment pas les études scientifiques.

Pour déterminer si les filles aiment les études scientifiques, nous avons décidé de calculer le pourcentage de candidates dans les filières scientifiques.

Tout d'abord, nous avons dû identifier les colonnes dont nous avions besoins:

la colonne 7 : “Région de l’établissement”

la colonne 20 : “Dont effectif des candidates pour une formation”

la colonne 19 : “Effectif total des candidats pour une formation”

la colonne 12 : “Filière de formation très agrégée”

la colonne 14 : “Filière de formation”

la colonne 15 : “Filière de formation détaillée bis”

Ensuite, nous avons dû sélectionner seulement la région Hauts-de-France grâce à la colonne 7 (Région de l’établissement) que nous avons limité à “Hauts-de-France”.

Puis nous avons sélectionné des critères bien précis pour déterminer si une formation était scientifique ou non.

Nous avons déterminé qu'une filière était scientifique s’il y avait des maths, de la physique, de l'informatique, de la chimie ou de la biologie. Et nous avons limité au BTS, BUT, licences et classes préparatoires.

Ce qui nous donne :

Pour la colonne 14 (Filière de formation), après avoir étudié plusieurs lignes, nous avons trouvé que tous les BTS et BUT qui étaient scientifiques, étaient catégorisés comme “BUT - Production” et “BTS - Production”. Ensuite, les classes préparatoires sont toutes scientifiques, nous avons également sélectionné la licence scientifique “Licence - Sciences - technologies - santé”. Nous avons donc sélectionné seulement les lignes où apparaissent ces critères.

Ensuite, dans la colonne 15 (Filière de formation détaillée bis), nous avons sélectionné les énoncés de formation qui contiennent les parties de mot :

“math” pour mathématique

“phy” pour physique

“info” pour informatique

“chim” pour chimie

“bio” pour biologie

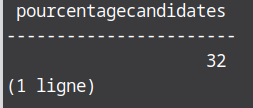
Afin de sélectionner seulement les filières où les titres sont scientifiques.

Et pour finir nous avons calculé le pourcentage de filles avec la colonne 20 (dont effectif des candidates pour une formation) que nous avons multiplié par 100 pour ensuite diviser le résultat par la colonne 19 (Effectif total des candidats pour une formation).

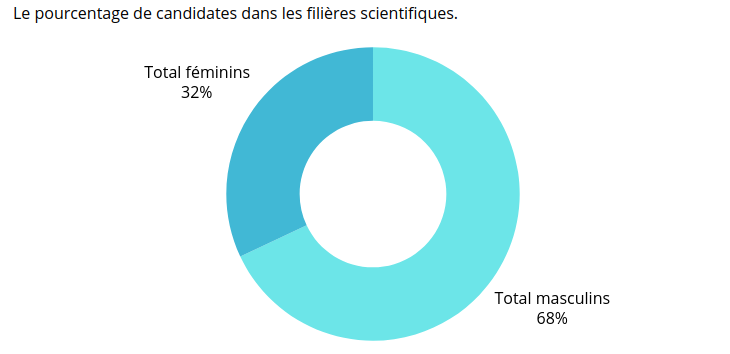
Ce qui nous donne cette requête :

| **SELECT ROUND(SUM(n20)\*100/SUM(n19)) as PourcentageCandidates FROM import WHERE n7 = 'Hauts-de-France' AND (n12 IN ('BTS', 'BUT', 'Licence', 'CPGE') OR n12 LIKE 'Ecole d%Ingénieur') AND n14 IN ('BTS - Production', 'BUT - Production', 'Classe préparatoire scientifique', 'Licence - Sciences - technologies - santé') AND (LOWER(n15) LIKE '%math%' OR LOWER(n15) LIKE '%phy%' OR LOWER(n15) LIKE '%info%' OR LOWER(n15) LIKE '%chim%' OR LOWER(n15) LIKE '%bio%');** |
| --- |

Il y a donc 32% du total de candidats qui ont postulé en filière scientifique qui sont des filles, c’est-à-dire une personnes sur 3 environ.



Nous pouvons donc dire que les filles ne sont pas forcément attirées par les études scientifiques mais qu'elles ne sont pas non plus inexistantes dans ces filières.



Nous ne pouvons donc ni affirmer, ni réfuter la véracité de cette assertion car elle dépend de comment chacun interprète ce taux.

## 2- Les bacs technologiques ont plus de chances que les bacs généraux d'être acceptés en filières sélectives.

Pour déterminer si les BAC technologiques ont plus de chance d'être acceptés dans les filières sélectives que les BAC généraux. On a décidé de calculer le taux d'acceptation dans les filières sélectives du BAC général et du BAC technologique.

Tout d'abord, nous avons dû identifier les colonnes dont nous avions besoins:

la colonne 7 : “Région de l’établissement”

la colonne 11 : “Sélectivité”

la colonne 57 : “Effectif des admis néo bacheliers généraux”

la colonne 58 : “Effectif des admis néo bacheliers technologiques”

la colonne 47 : “Effectif total des candidats ayant accepté la proposition de l’établissement (admis)”

Ensuite nous avons dû sélectionner seulement la région Hauts-de-France grâce à la colonne 7 (Région de l’établissement) que nous avons limité à “Hauts-de-France”.

Puis nous avons gardé à l’aide de la colonne 11 (Sélectivité) seulement les formations sélectives.

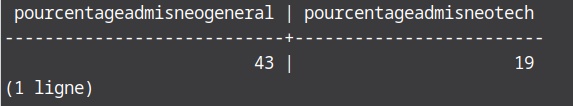
Puis nous avons calculé le pourcentage d'élèves néo bacheliers sortant de BAC général en multipliant la colonne 57 (Effectif des admis néo bacheliers généraux) par 100 puis en divisant le résultat par la colonne 47 (Effectif total des candidats ayant accepté la proposition de l’établissement (admis)).

Nous avons fait pareil pour les néo bacheliers sortant d’un BAC technologique mais en multipliant cette fois la colonne 58 (Effectif des admis néo bacheliers technologiques) par 100 puis en divisant le résultat par la colonne 47 (Effectif total des candidats ayant accepté la proposition de l’établissement (admis)).

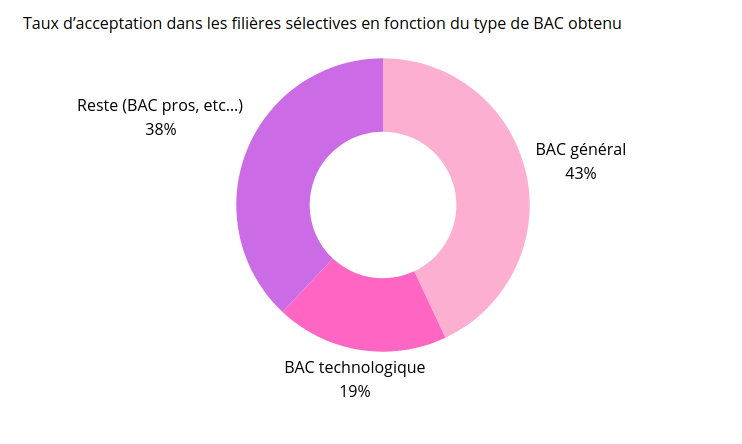
Ce qui nous donne la requête suivante :

| **SELECT ROUND((SUM(n57)\*100)/SUM(n47)) AS pourcentageAdmisNeoGeneral, ROUND(SUM(n58)\*100/SUM(n47)) AS pourcentageAdmisNeoTech FROM import WHERE n7 = 'Hauts-de-France' AND n11 = 'formation sélective';** |
| --- |

Nous obtenons ce résultat :



Nous pouvons donc constater que les élèves sortant d' un BAC technologique ont plus de mal à être acceptés par les filières sélectives que les élèves sortant d'un BAC général.



Nous pouvons donc réfuter la véracité de cette assertion.

## 3- Les boursiers préfèrent les études courtes.

Pour déterminer si les boursiers préfèrent les études courtes ou les études longues, nous avons décidé de calculer le nombre de postulant boursiers dans chaque catégorie d’études (longues ou courtes). Nous nous sommes concentrés seulement sur la phase principale car nous pensons que cette phase est la plus représentative de la poursuite d'étude voulue.

Tout d'abord, nous avons dû sélectionner des critères pour déterminer si des études étaient longues ou courtes :

Les études courtes sont considérées comme des études prenant 2 ans ou moins d’enseignements et les études longues sont considérées comme des études prenant 3 ans ou plus d’enseignements.

| Études courtes | Études longues |
| --- | --- |
| BTS  CPGE  PASS  IFSI | BUT  Écoles de Commerce  Écoles d'Ingénieurs  Licences |

Ensuite, nous avons dû identifier les colonnes dont nous avions besoins:

la colonne 7 : “Région de l’établissement”

la colonne 12 : “Filière de formation très agrégée”

la colonne 24 : “Effectif des candidats néo bacheliers généraux en phase principale”

la colonne 26 : “Dont effectif des candidats boursiers néo bacheliers technologiques en phase principale”

la colonne 28 : “Dont effectif des candidats boursiers néo bacheliers professionnels en phase principale”

Ensuite, nous avons dû sélectionner seulement la région Hauts-de-France grâce à la colonne 7 (Région de l’établissement) que nous avons limité à “Hauts-de-France”.

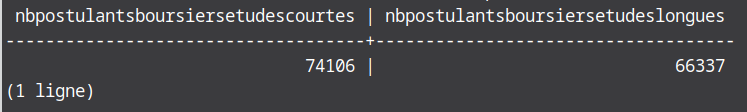
Puis nous avons sélectionné d’abord les études courtes à l'aide de la colonne 12 (Filière de formation très agrégée) en sélectionnant les lignes où apparaissent les critères “BTS, CPGE, PASS, IFSI”. Puis nous avons additionné les candidats boursiers sortant de chaque BAC (général, technologique, professionnel). À l’aide des colonne 24 (Effectif des candidats néo bacheliers généraux en phase principale), colonne 26 (dont effectif des candidats boursiers néo bacheliers technologiques en phase principale) et 28 (dont effectif des candidats boursiers néo bacheliers professionnels en phase principale).

Pour finir, nous avons fait la même chose avec les études longues, donc en prenant dans la colonne 12 les lignes où n’apparaissent pas les critères “BTS, CPGE, PASS, IFSI”. Nous avons additionné les candidats boursiers sortant de chaque BAC (général, technologique, professionnel). À l'aide des colonne 24 (Effectif des candidats néo bacheliers généraux en phase principale), colonne 26 (dont effectif des candidats boursiers néo bacheliers technologiques en phase principale) et 28 (dont effectif des candidats boursiers néo bacheliers professionnels en phase principale).

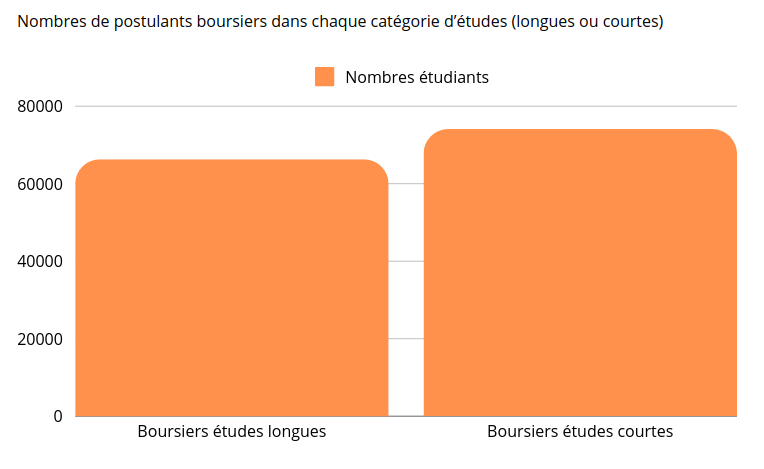
Nous obtenons donc cette requête :

| **SELECT (SELECT SUM(n24)+SUM(n26)+SUM(n28) FROM import WHERE n7 = 'Hauts-de-France' AND n12 IN ('BTS', 'CPGE', 'PASS', 'IFSI')) AS nbPostulantsBoursiersEtudesCourtes, (SELECT SUM(n24)+SUM(n26)+SUM(n28) FROM import WHERE n7 = 'Hauts-de-France' AND n12 NOT IN ('BTS', 'CPGE', 'PASS', 'IFSI')) AS nbPostulantsBoursiersEtudesLongues;** |
| --- |

et nous obtenons ce résultat.



Nous constatons qu’il y a plus de boursiers qui postulent en études courtes qu'en études longues mais l'écart est de moins de 8000 élèves.



Nous pouvons donc confirmer la véracité de cette assertion.

# PARTIE 2

## Sujet :

Vous choisirez 2 IUT, dans la région choisie pour la partie 1, ayant un BUT INFORMATIQUE et vous ferez une étude comparée :

- Du taux de filles qui ont candidaté

- Du taux d’admis avec mention

- Du taux de BTN, BG et autres classés

- Tout autre critère qui vous semble pertinent

## Établissements choisit :

- Institut universitaire de technologie de Lille - Université de Lille

- I.U.T. de Valenciennes - Site de Maubeuge

--------------------------------------------------------------------

## 1. Taux de filles qui ont candidaté & taux d’admis avec mention

Pour calculer le taux de fille qui ont candidaté nous avons dû identifier les colonnes dont nous avions besoins:

la colonne 3 : “Code UAI de l'établissement”

la colonne 7 : “Région de l’établissement”

la colonne 9 : “Commune de l’établissement”

la colonne 15 : “Filière de formation détaillée bis”

la colonne 19 : “Effectif total des candidats pour une formation”

la colonne 20 : “Dont effectif des candidates pour une formation”

Ensuite nous avons dû sélectionner seulement la région Hauts-de-France grâce à la colonne 7 (Région de l’établissement) que nous avons limité à “Hauts-de-France”.

Puis l'établissement, à l'aide de la colonne 3 (Code UAI de l'établissement) que nous avons limité à l'IUT de Maubeuge : 0596183A et à l’IUT de Villeneuve d'Ascq : 0597215X.

Nous avons dû aussi limiter la filière dans l'établissement à l'aide de la colonne 15 (Filière de formation détaillée bis) que nous avons limitée à “Informatique”.

Une fois que nous avons ces filières, nous avons calculé le taux de candidature féminine à l'aide de la colonne 20 (dont effectif des candidates pour une formation) que nous multiplions par 100.

Ensuite, on divise le résultat par la colonne 19 (Effectif total des candidats pour une formation).

Pour calculer le taux d’admis avec mention nous avons dû identifier les colonnes dont nous avions besoins:

la colonne 3 : “Code UAI de l'établissement”

la colonne 7 : “Région de l’établissement”

la colonne 9 : “Commune de l’établissement”

la colonne 15 : “Filière de formation détaillée bis”

la colonne 47 : “Effectif total des candidats ayant accepté la proposition de l’établissement (admis)”

la colonne 63 : “Dont effectif des admis néo bacheliers avec mention Assez Bien au bac”

la colonne 64 : “Dont effectif des admis néo bacheliers avec mention Bien au bac”

la colonne 65 : “Dont effectif des admis néo bacheliers avec mention Très Bien au bac”

la colonne 66 : “Dont effectif des admis néo bacheliers avec mention Très Bien avec félicitations au bac”

la colonne 90 : “Dont % d’admis avec mention (BG)”

la colonne 92 : “Dont % d’admis avec mention (BT)”

la colonne 94 : “Dont % d’admis avec mention (BP)”

Ensuite nous avons dû sélectionner seulement la région Hauts-de-France grâce à la colonne 7 (Région de l’établissement) que nous avons limité à “Hauts-de-France”.

Puis l'établissement, à l'aide de la colonne 3 (Code UAI de l'établissement) que nous avons limité à l'IUT de Maubeuge : 0596183A et à l’IUT de Villeneuve-d'Ascq : 0597215X.

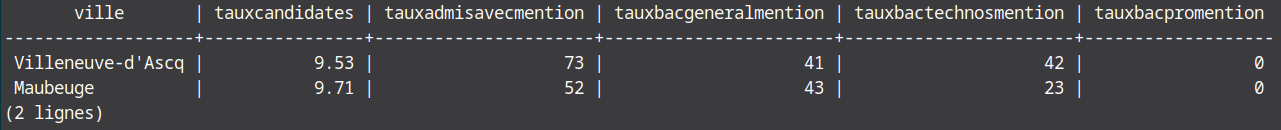
Nous avons dû aussi limiter la filière dans l'établissement à l'aide de la colonne 15 (Filière de formation détaillée bis) que nous avons limitée à “Informatique”.

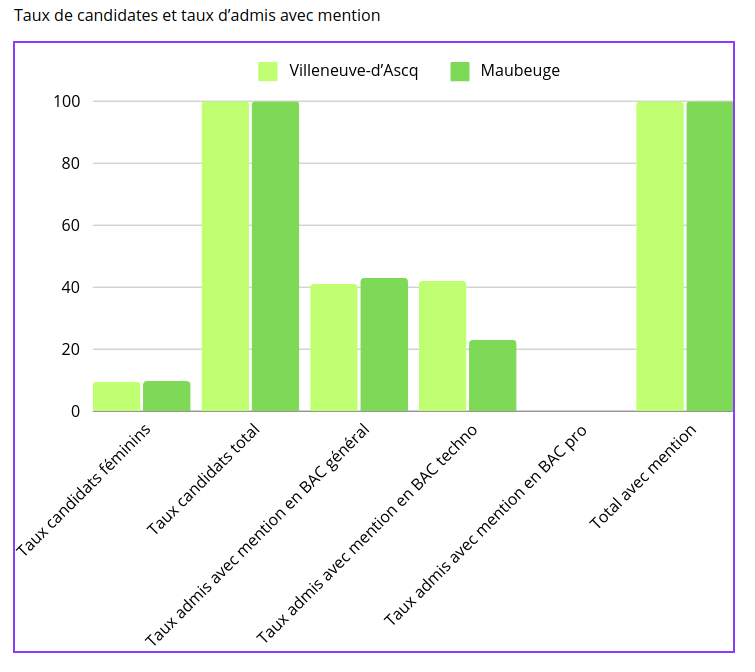
Une fois que nous avons ces filières, nous avons pu afficher le taux d’admis avec mention sans distinction de filière à l'aide de l’addition de la colonne 63 (dont effectif des admis néo bacheliers avec mention Assez Bien au BAC), la colonne 64 (dont effectif des admis néo bacheliers avec mention Bien au BAC), la colonne 65 (dont effectif des admis néo bacheliers avec mention Très Bien au BAC), et la colonne 66 (dont effectif des admis néo bacheliers avec mention Très Bien avec félicitations au BAC) que nous avons ensuite multiplié par 100 puis divisé par la colonne 47 (Effectif total des candidats ayant accepté la proposition de l’établissement (admis)).

En combinant ces deux calculs dans une seule requête nous obtenons ceci :

| **SELECT n9 AS ville, ROUND(n20\*100/n19, 2) AS tauxCandidates, ROUND((n63+n64+n65+n66)\*100/n47) AS tauxAdmisAvecMention, n90 AS tauxBacGeneralMention, n92 AS tauxBacTechnosMention, n94 AS tauxBacProMention FROM import WHERE n7 = 'Hauts-de-France' AND n3 IN ('0597215X', '0596183A') AND n15 = 'Informatique';** |
| --- |

Nous obtenons ce résultat :





Nous pouvons voir que l'IUT de Maubeuge accepte un peu plus de filles que l’IUT de Villeneuve-d'Ascq.

Nous pouvons voir également que l’IUT de Villeneuve-d’Ascq accepte plus le BAC avec mention que l’IUT de Maubeuge mais en regardant les détails l’IUT de Maubeuge accepte plus de BAC général avec mention que l’IUT de Villeneuve-d’Ascq. Nous remarquons aussi que l’IUT de Villeneuve-d’Ascq accepte plus de BAC technologique avec mention que l’IUT de Maubeuge et aucun des deux IUT accepte de BAC professionnel avec mention.

## 2. Du taux de BTN, BG et autres classés

Pour la catégorie “autres classés” nous avons décidé de rajouter les classés BAC pro.

Pour calculer le taux de classés, nous avons dû identifier les colonnes dont nous avions besoins:

la colonne 3 : “Code UAI de l'établissement”

la colonne 7 : “Région de l’établissement”

la colonne 15 : “Filière de formation détaillée”

la colonne 56 : “Effectif des admis néo bacheliers”

la colonne 57 : “Effectif des admis néo bacheliers généraux”

la colonne 58 : “Effectif des admis néo bacheliers technologiques”

la colonne 59 : “Effectif des admis néo bacheliers professionnels”

la colonne 60 : “Effectif des autres candidats admis”

Ensuite, nous avons dû sélectionner seulement la région Hauts-de-France grâce à la colonne 7 (Région de l’établissement) que nous avons limité à “Hauts-de-France”.

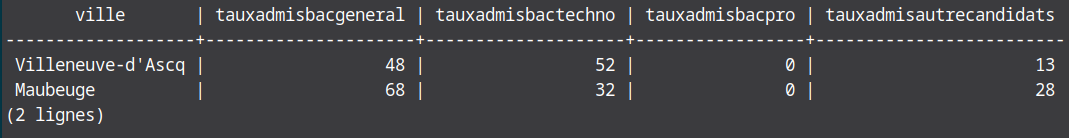
Puis, l'établissement à l'aide de la colonne 3 (Code UAI de l'établissement) que nous avons limité à l'IUT de Maubeuge : 0596183A et à l’IUT de Villeneuve-d'Ascq : 0597215X. Nous avons dû aussi limiter la filière dans l'établissement à l'aide de la colonne 15 (Filière de formation détaillée bis) que nous avons limitée à “Informatique”.

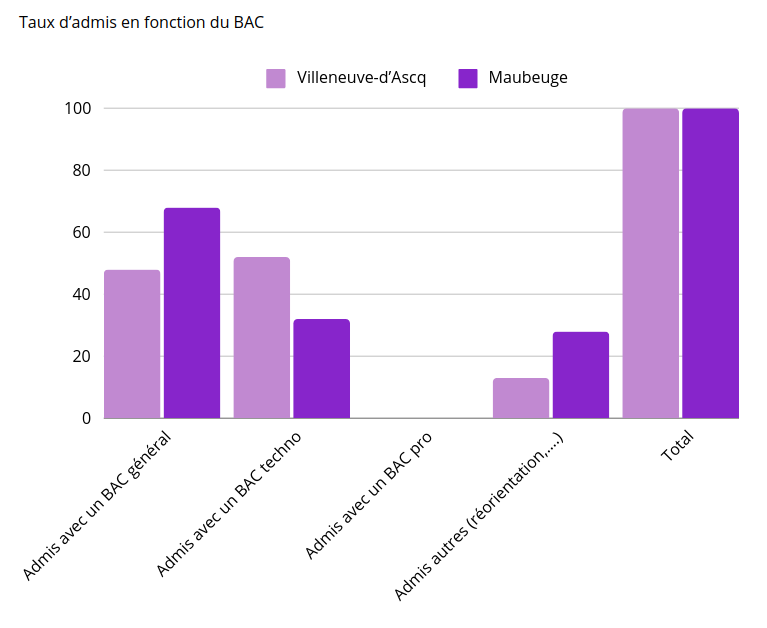
Ensuite, une fois ces filtres appliqués, nous avons pu calculer chaque taux d'admission par filière. Nous avons commencé par les BAC généraux avec la colonne 57 (Effectif des admis néo bacheliers généraux) que nous multiplions par 100 et nous divisons le résultat obtenu par la colonne 56 (Effectif des admis néo bacheliers). Ensuite, nous avons continué avec les admissions des BAC technologiques avec la colonne 58 (Effectif des admis néo bacheliers technologiques) que nous multiplions par 100 et nous divisions le résultat obtenu par la colonne 56 (Effectif des admis néo bacheliers). Puis, avec les BAC professionnels avec la colonne 59 (Effectif des admis néo bacheliers professionnels) que nous multiplions par 100 et nous divisons le résultat obtenu par la colonne 56 (Effectif des admis néo bacheliers). Et pour finir, les autres (c’est-à-dire les étudiants venant d’autres filières mais déjà en études supérieures ou bien des étudiants ayant fait une année de césure, etc… ) avec la colonne 60 (Effectif des autres candidats admis) que nous multiplions par 100 et nous divisons le résultat par la colonne 56 (Effectif des admis néo bacheliers).

Nous obtenons donc cette requête :

| **SELECT n9 AS ville, ROUND(n57\*100/n56) AS tauxAdmisBacGeneral, ROUND(n58\*100/n56) AS tauxAdmisBacTechno, ROUND(n59\*100/n56) AS tauxAdmisBacPro, ROUND(n60\*100/n56) AS tauxAdmisAutreCandidats FROM import WHERE n7 = 'Hauts-de-France' AND n3 IN ('0597215X', '0596183A') AND n15 = 'Informatique';** |
| --- |

et avec cette requête nous avons ce résultat :





Nous remarquons que l’IUT de Maubeuge accepte plus de BAC général que l’IUT de Villeneuve-d’Ascq et à contrario, l’IUT de Villeneuve-d’Ascq accepte plus de BAC technologiques que l’IUT de Maubeuge. Les deux IUT n'acceptent pas de BAC professionnels et pour les autres candidats l’IUT de Maubeuge en accepte plus que l’IUT de Villeneuve-d’Ascq.

## 3. Tout autre critère qui vous semble pertinent

Nous avons décidé de travailler sur le taux d'accès, la capacité d'accueil et le nombre de candidats pour le reste des critères.

Pour calculer tout cela, nous avons dû identifier les colonnes dont nous avions besoin de :

la colonne 3 : “Code UAI de l'établissement”

la colonne 7 : “Région de l’établissement”

la colonne 15 : “Filière de formation détaillée”

la colonne 113 : “Taux d’accès”

la colonne 18 : “Capacité de l’établissement par formation”

Ensuite nous avons dû sélectionner seulement la région Hauts-de-France grâce à la colonne 7 (Région de l’établissement) que nous avons limité à “Hauts-de-France”.

Puis, l'établissement à l' aide de la colonne 3 (Code UAI de l'établissement) que nous avons limité à l'IUT de Maubeuge : 0596183A et à l’IUT de Villeneuve-d'Ascq : 0597215X.

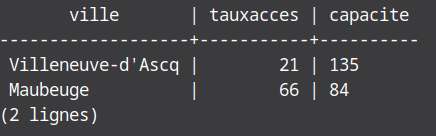
Nous avons dû aussi limiter la filière dans l'établissement à l'aide de la colonne 15 (Filière de formation détaillée bis) que nous avons limitée à “Informatique”.

Ensuite, nous avons simplement affiché le taux d'accès à l'aide de la colonne 113 (Taux d’accès) et la capacité de l'aide de la colonne 18 (Capacité de l’établissement par formation).

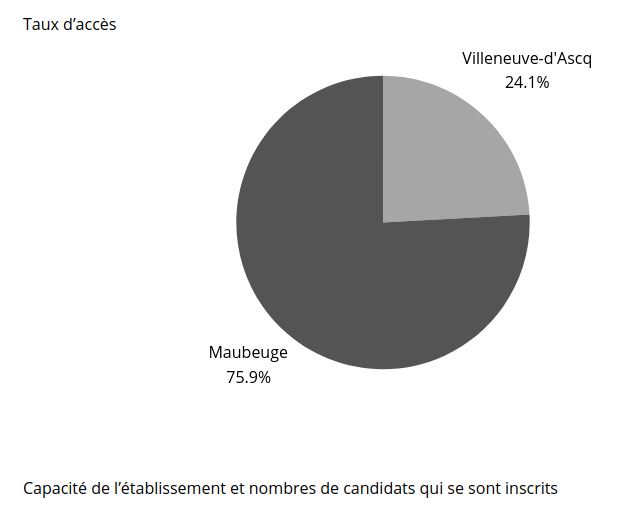
Nous obtenons donc cette requête :

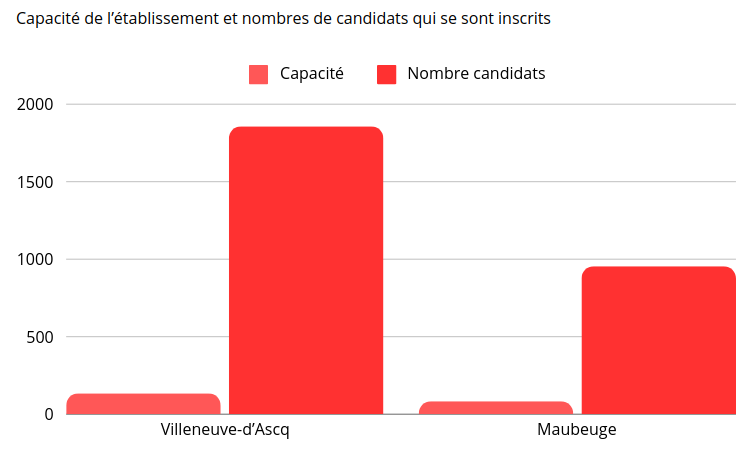
| **SELECT n9 AS ville, n113 AS tauxAcces, n18 AS capacite FROM import WHERE n7 = 'Hauts-de-France' AND n3 IN ('0597215X', '0596183A') AND n15 = 'Informatique';** |
| --- |

Qui nous donne ce résultat :



Nous pouvons voir que le taux d'accès à l’IUT de Maubeuge est trois fois plus haut que celui de l’IUT de Villeneuve-d’Ascq mais que sa capacité d'accueil est quasiment 1.5 fois plus petite.





Ensuite, d’après les graphiques on voit qu’il y a presque deux fois plus de candidats à l’IUT de Villeneuve-d’Ascq qu’à l’IUT de Maubeuge.

