

Analyse des admissions sur Parcoursup en 2023 en fonction des caractéristiques des établissements et du profil des candidats

Blaste MULAMBA, Basma MORABIT, Hamza CHERKAOUI, Manyl TIDJANI

2024-04-02

Contents

Introduction	1
Contexte et description de l'ensemble de données	1
Objectif de l'étude	2
Méthodologie	2
Procédures de nettoyage et de renommage des données	2
Affichage du jeu de données avant nettoyage et renommage	2
Affichage du jeu de données après nettoyage et renommage	2
Renommage des variables retenues pour notre étude	3
Analyse des caractéristiques des établissements et des candidats	4
Lien entre le nombre d'établissements et le nombre de formations par région :	4
Analyse de la sélectivité des formations et du taux de remplissage :	7
Relation entre le type de diplôme et la répartition des candidats :	9
Impact du genre sur le choix de la filière en Licence	11
Conclusion	12
Références	12

Introduction

Contexte et description de l'ensemble de données

Notre jeu de données présente les vœux de poursuite d'études et de réorientation dans l'enseignement supérieur ainsi que les propositions des établissements pour chaque formation à la fin du processus d'affectation de la plateforme Parcoursup pour la session 2023. Il est produit par le Ministère de l'Enseignement Supérieur et de la Recherche et comprend 118 variables et 13 869 enregistrements. Ces données sont disponibles sous forme de csv à télécharger et sous forme de visualisation graphique "dataviz".

Objectif de l'étude

Le but de notre étude est de comprendre si les caractéristiques des établissements tels que leur statut, leur capacité d'accueil et leur région influent sur les candidatures et les admissions. De plus, nous chercherons à comprendre si des facteurs tels que le sexe ou type de bac du candidat ont également une influence dessus.

Méthodologie

Procédures de nettoyage et de renommage des données

Un nettoyage de données a été effectué d'abord en éliminant des colonnes redondantes, c'est-à-dire celles qui regroupaient des informations contenues dans d'autres colonnes. Ensuite un choix a été porté sur 25 variables jugées pertinentes pour notre étude, parmi les 118 variables existantes. Ces colonnes sélectionnées ont enfin été renommées pour garantir une meilleure lisibilité et donc une meilleure compréhension lors des manipulations.

Bien évidemment, nous avons fait attention à la conformité des colonnes par rapport aux colonnes exclues.

Voici un aperçu avant/après de notre jeu de données à la suite des procédures de nettoyage et renommage.

Affichage du jeu de données avant nettoyage et renommage

```
## # A tibble: 5 x 118
##   session contrat_etab cod_uai   g_ea_lib_vx      dep  dep_lib region_etab_aff
##   <dbl> <chr>          <chr>   <chr>      <chr> <chr>   <chr>
## 1  2023 Public      0333269Z La Prépa des INP ~ 33  Gironde Nouvelle-Aquit~
## 2  2023 Public      0383098L La Prépa des INP ~ 38  Isère   Auvergne-Rhône~
## 3  2023 Public      0542391F La Prépa des INP ~ 54  Meurth~ Grand Est
## 4  2023 Public      0212108C ESIREM Dijon      21  Côte-d~ Bourgogne-Fran~
## 5  2023 Public      0371610Z Polytech Tours   37  Indre-- Centre-Val de ~
## # i 111 more variables: acad_mies <chr>, ville_etab <chr>,
## #   lib_for_voe_ins <chr>, select_form <chr>, fili <chr>,
## #   lib_comp_voe_ins <chr>, form_lib_voe_acc <chr>, fil_lib_voe_acc <chr>,
## #   detail_forma <chr>, g_olocalisation_des_formationen <chr>, capa_fin <dbl>,
## #   voe_tot <dbl>, voe_tot_f <dbl>, nb_voe_pp <dbl>, nb_voe_pp_internat <dbl>,
## #   nb_voe_pp_bg <dbl>, nb_voe_pp_bg_brs <dbl>, nb_voe_pp_bt <dbl>,
## #   nb_voe_pp_bt_brs <dbl>, nb_voe_pp_bp <dbl>, nb_voe_pp_bp_brs <dbl>, ...
```

Affichage du jeu de données après nettoyage et renommage

```
## # A tibble: 5 x 25
##   id      statut `nom établissement`  région ville selectivité diplôme filière
##   <chr>   <chr>   <chr>                <chr> <chr> <chr>      <chr>   <chr>
## 1 0333269Z Public La Prépa des INP - G~ Nouve~ Bord~ formation ~ Ecole ~ Format~
## 2 0383098L Public La Prépa des INP - G~ Auver~ Gren~ formation ~ Ecole ~ Format~
## 3 0542391F Public La Prépa des INP - G~ Grand~ Nanc~ formation ~ Ecole ~ Format~
## 4 0212108C Public ESIREM Dijon          Bourg~ Dijon formation ~ Ecole ~ Format~
## 5 0371610Z Public Polytech Tours          Centr~ Orlé~ formation ~ Ecole ~ Format~
## # i 17 more variables: `nom de la formation` <chr>, `capacité accueil` <dbl>,
## #   candidatures <dbl>, `candidats filles` <dbl>,
## #   `candidats phase principale` <dbl>, `candidats phase pr bac gen` <dbl>,
## #   `candidats phase pr bac tech` <dbl>, `candidats phase pr bac pro` <dbl>,
```

```
## # `candidats phase pr bac autres spé` <dbl>,
## # `candidats phase complémentaire` <dbl>, `candidats phase comp gen` <dbl>,
## # `candidats phase comp tech` <dbl>, `candidats phase comp pro` <dbl>, ...
```

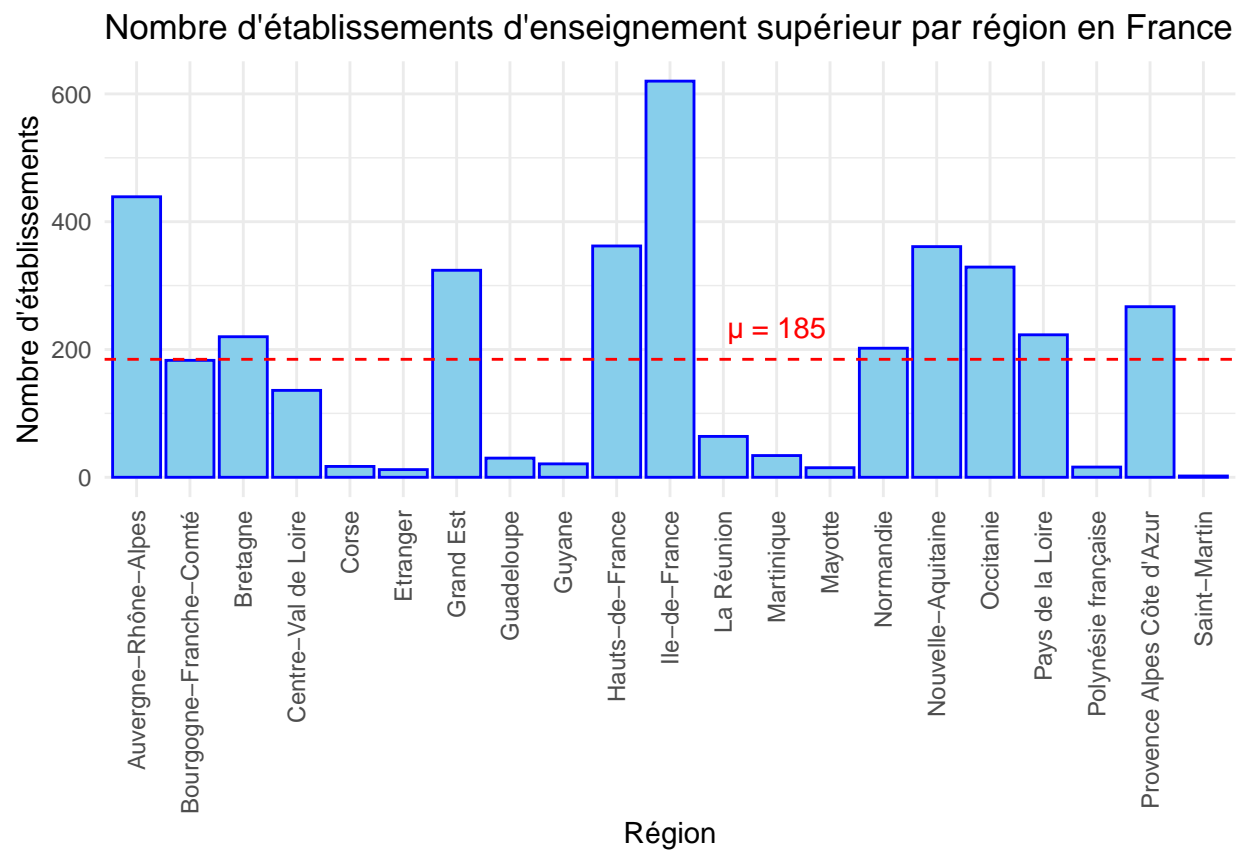
Renommage des variables retenues pour notre étude

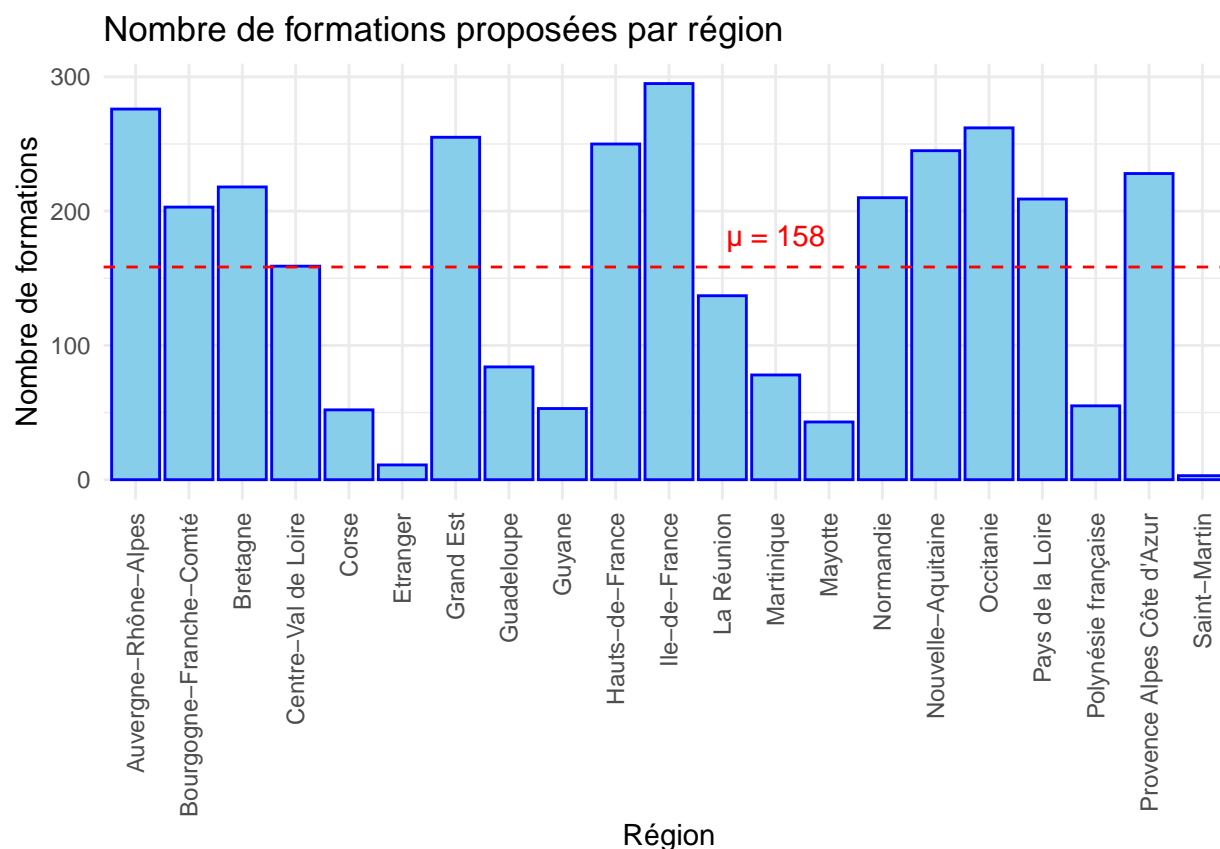
Voici en détail le renommage des variables pour question de lisibilité et d'aisance dans la manipulation.

##	ancien_nom	nouveau_nom
## 1	cod_uai	id
## 2	contrat_etab	statut
## 3	g_ea_lib_vx	nom établissement
## 4	region_etab_aff	région
## 5	acad_mies	ville
## 6	select_form	selectivité
## 7	fili	diplôme
## 8	form_lib_voe_acc	filière
## 9	fil_lib_voe_acc	nom de la form-ation
## 10	capa_fin	capacité accueil
## 11	voe_tot	candidats
## 12	voe_tot_f	candidats filles
## 13	nb_voe_pp	candidats phase principale
## 14	nb_voe_pp_bg	candidats phase pr bac gen
## 15	nb_voe_pp_bt	candidats phase pr bac tech
## 16	nb_voe_pp_bp	candidats phase pr bac pro
## 17	nb_voe_pp_at	candidats phase pr bac autres spé
## 18	nb_voe_pc	candidats phase complémentaire
## 19	nb_voe_pc_bg	candidats phase comp gen
## 20	nb_voe_pc_bt	candidats phase comp tech
## 21	nb_voe_pc_bp	candidats phase comp pro
## 22	nb_voe_pc_at	candidats phase comp bac autres spé
## 23	prop_tot	candidats préselectionnés
## 24	acc_tot	candidats admis
## 25	candidates_admises	candidates admises

Analyse des caractéristiques des établissements et des candidats

Lien entre le nombre d'établissements et le nombre de formations par région :

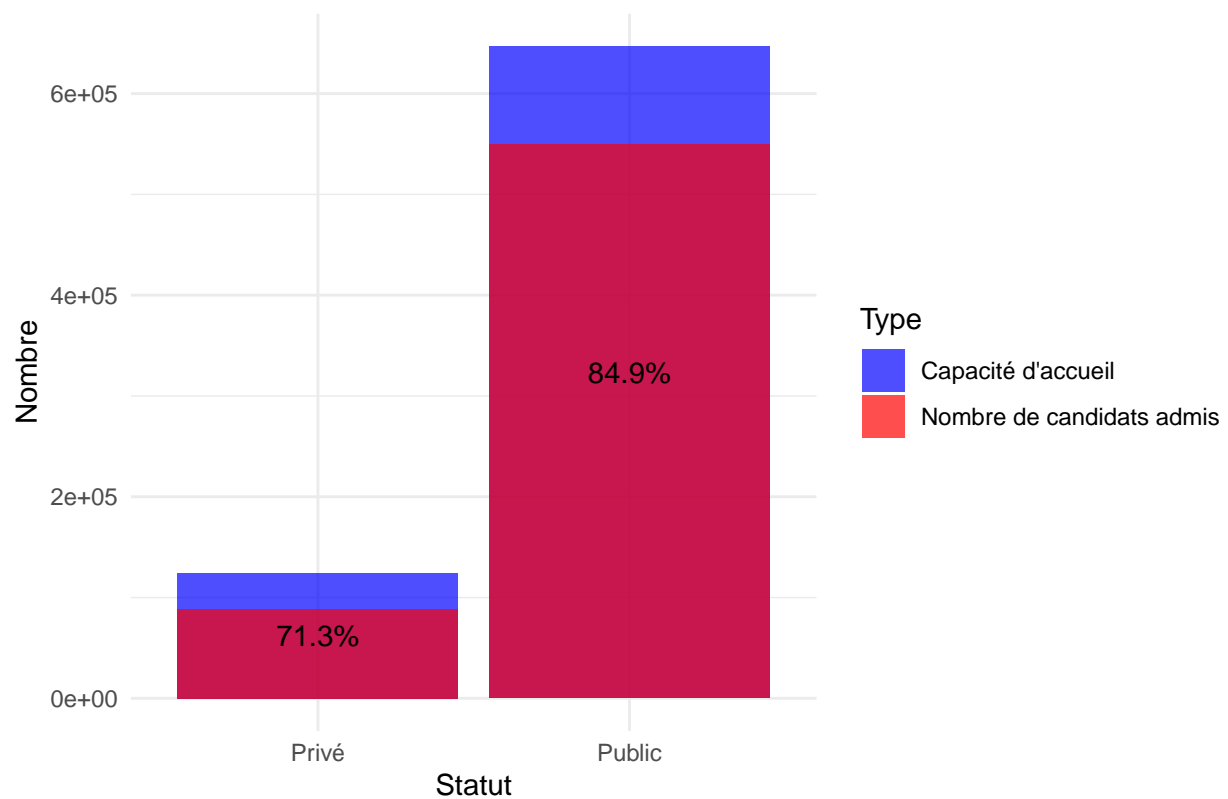




La taille de la population de la région a une influence directe sur le nombre d'établissements présents dans cette région. En moyenne, en France métropolitaine, le nombre d'établissements est supérieur à la moyenne avec en tête de classement la région d'Ile de France et celle d'Auvergne-Rhône-Alpes. Par contre, dans les territoires d'outre-mer, le nombre d'établissement est significativement inférieur à la moyenne.

Par contre, la région n'a pas d'influence directe sur le nombre de formations qui y sont proposées. Il y a moins de disparités entre territoires en moyenne. Par exemple, en Centre-Val de Loire, il y a très peu d'établissements par rapport à la moyenne mais il y a un bon taux de couverture des formations proposées par rapport à la moyenne en France.

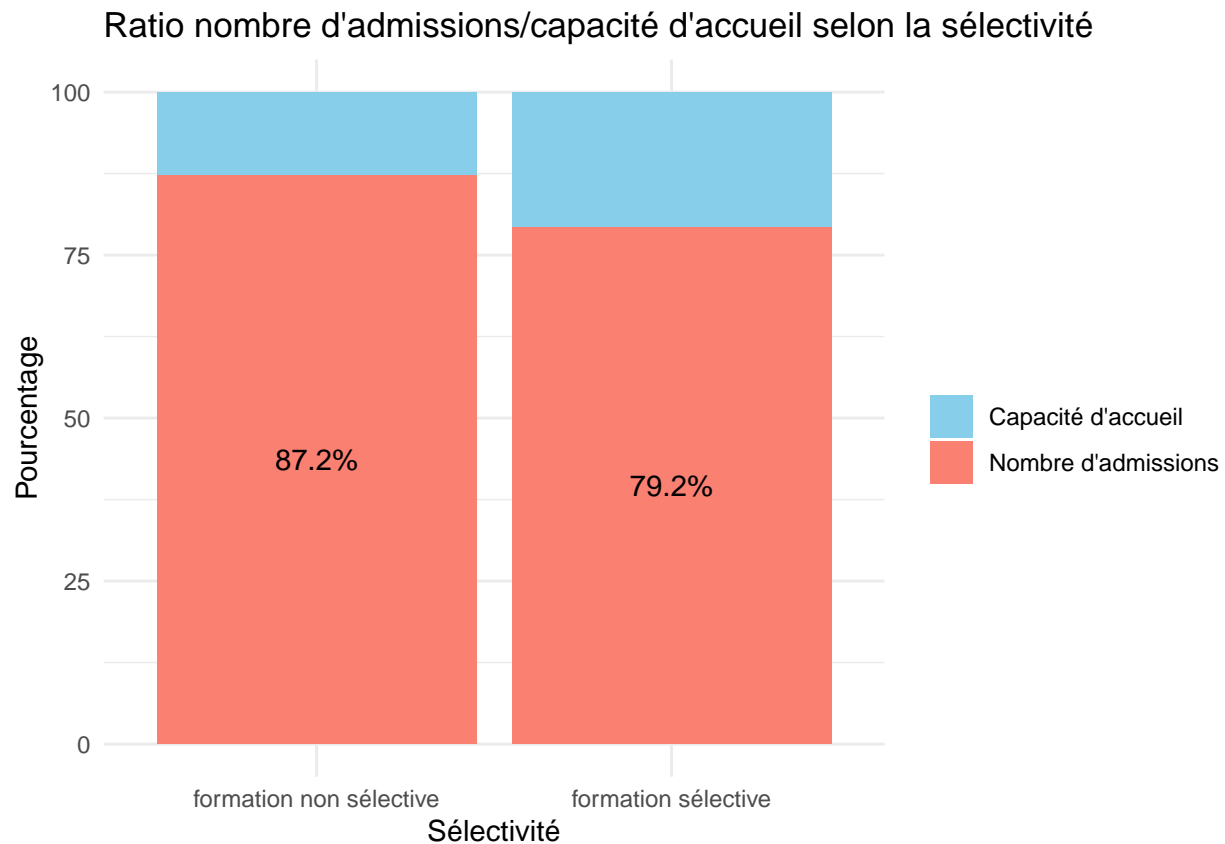
Capacité d'accueil versus nombre de candidats admis par statut



Le constat c'est que les établissements publics ou privés n'atteignent jamais le maximum de leur capacité d'accueil. Donc le statut de l'établissement n'a pas d'influence directe sur le taux de remplissage de l'établissement.

En conclusion, bien que dans l'ensemble les établissements privés présentent une capacité d'accueil inférieure à celle des établissements publics, ils parviennent néanmoins à admettre un nombre considérable de candidats.

Analyse de la sélectivité des formations et du taux de remplissage :



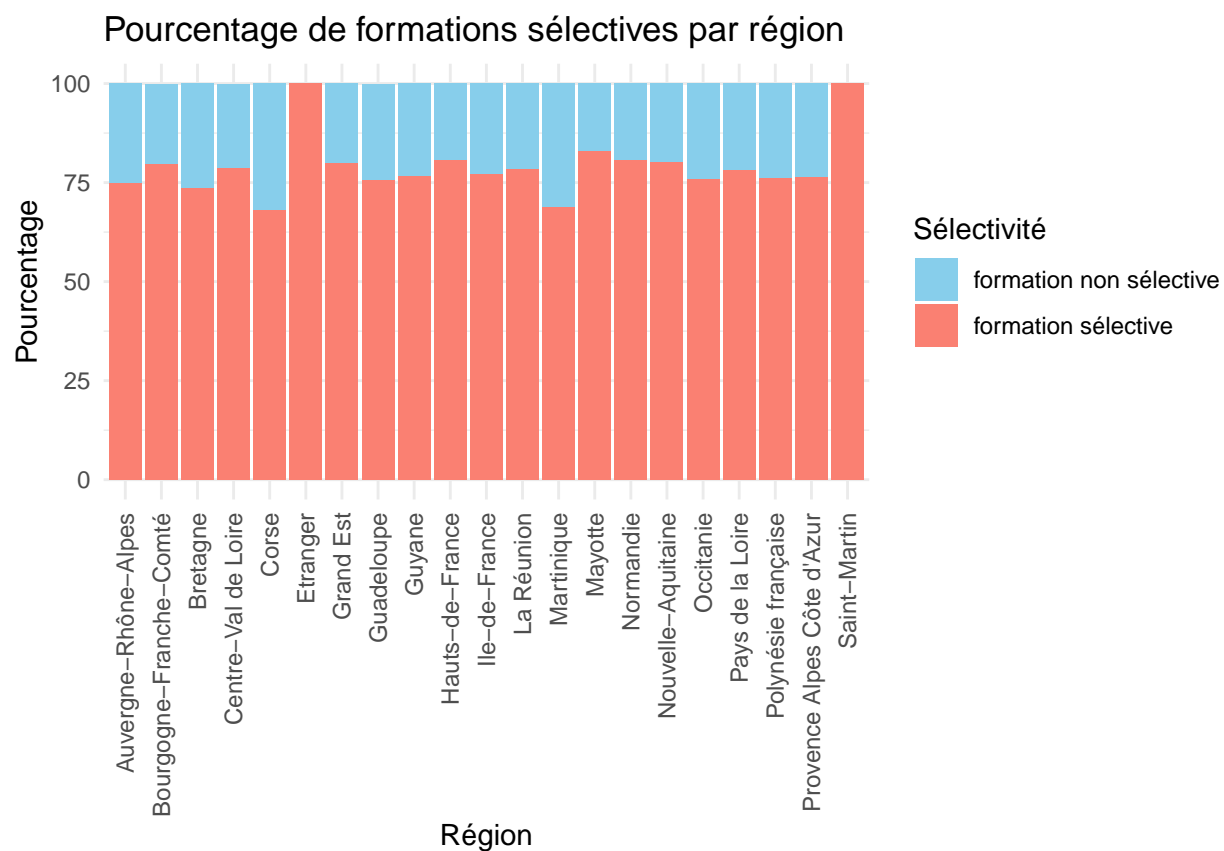
A première vue, les formations sélectives sont censées remplir leur capacité d'accueil puisqu'elles reçoivent beaucoup de candidatures et qu'elles sont censées attirer les candidats. Donc les candidats devraient normalement accepter les propositions des formations sélectives.

Cependant, nous remarquons que les formations sélectives ne remplissent pas leur capacité d'accueil, et même elles ont un taux de remplissage moins élevée par rapport aux formations sélectives.

En conclusion, la sélectivité n'a pas une influence significative sur le taux de remplissage de la formation.

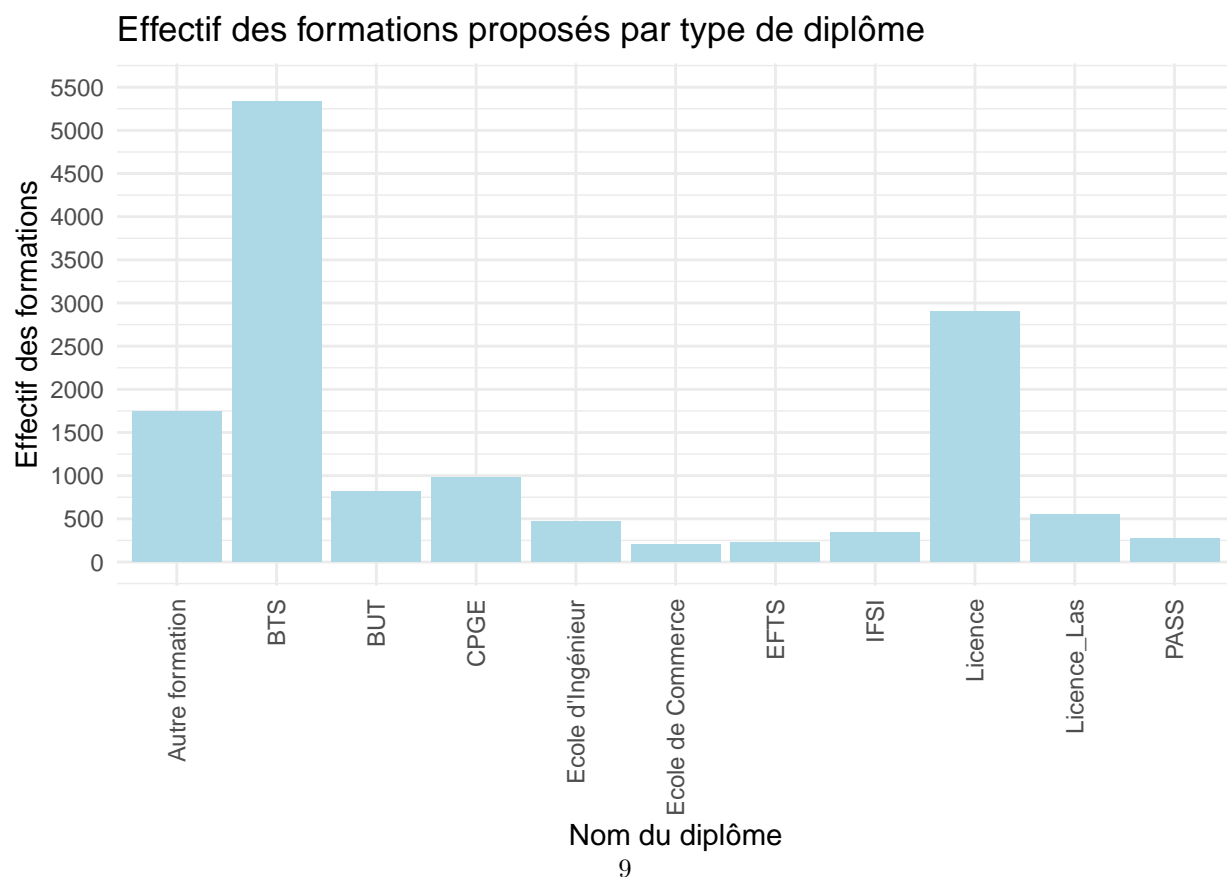
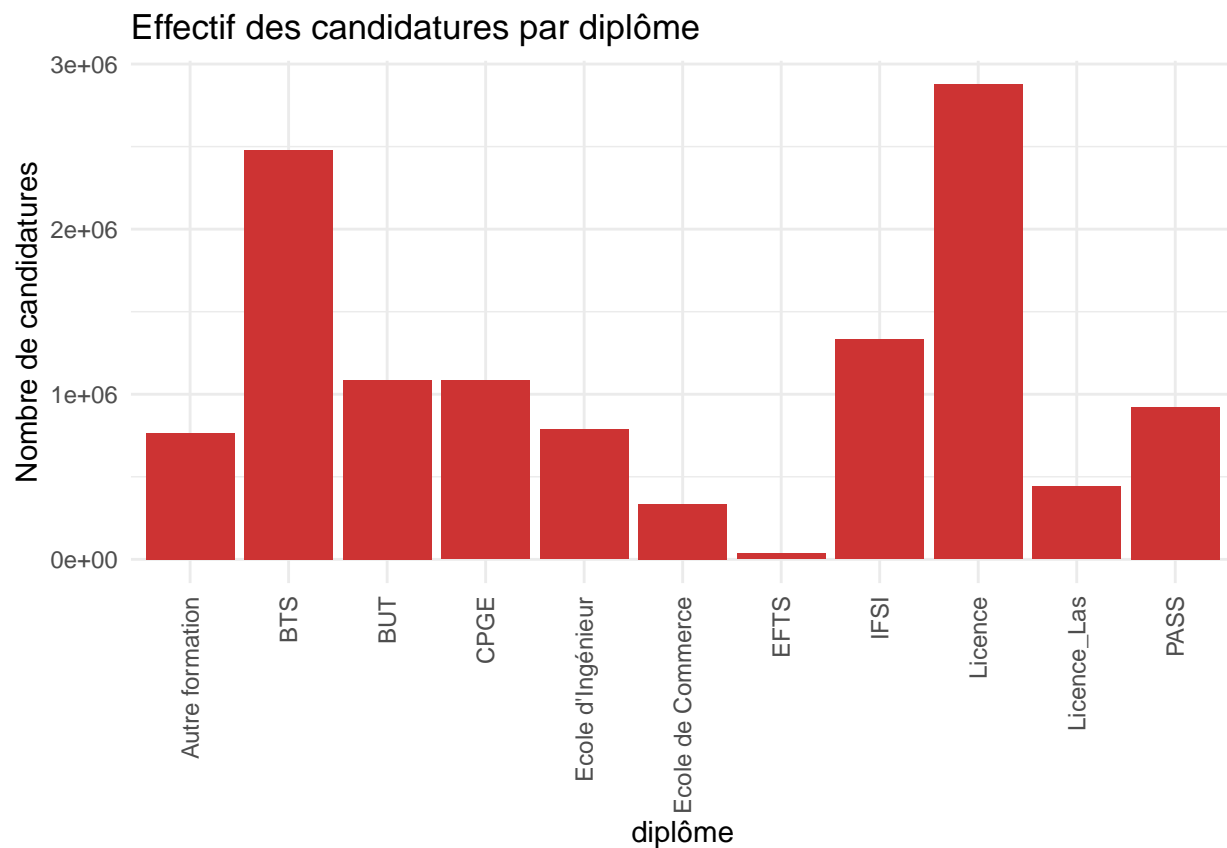
```
## # A tibble: 2 x 3
##   selectivité      nb_occurences pourcentage
##   <chr>          <int>         <dbl>
## 1 formation non sélective      3104         22.4
## 2 formation sélective      10765         77.6

## `summarise()` has grouped output by 'selectivité'. You can override using the
## `.groups` argument.
```



On remarque que les formations sélectives représentent 75% des formations en France. On constate que la région n'a pas d'impact sur la sélectivité car par exemple en Ile de France, le taux de formations sélectives est le même qu'en Guyane ou en Guadeloupe.

Relation entre le type de diplôme et la répartition des candidats :

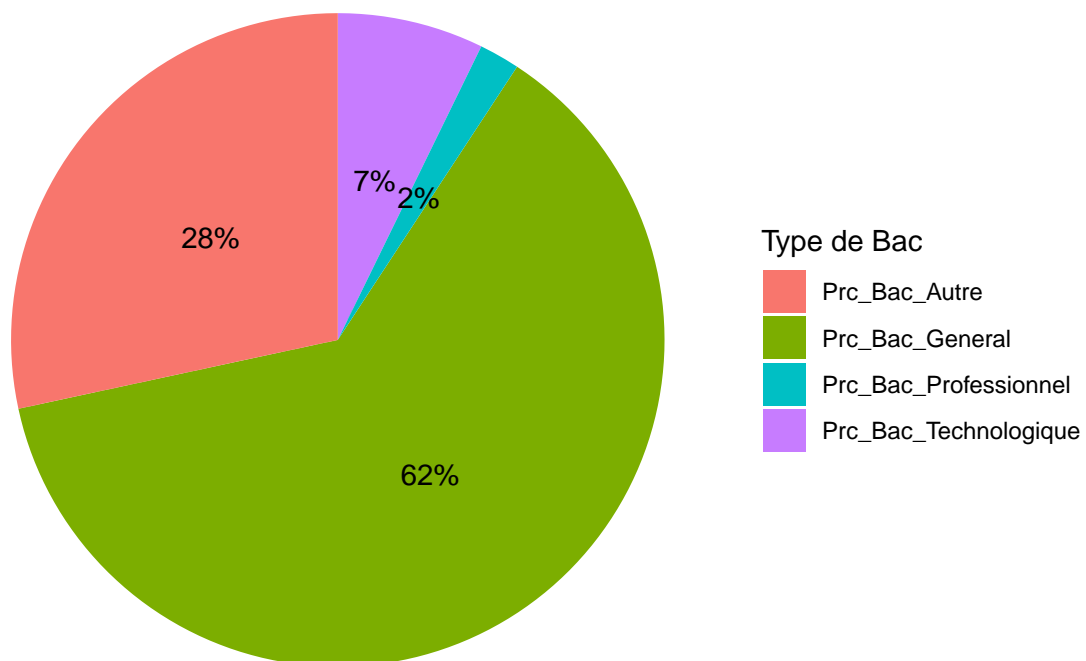


```
## # A tibble: 2 x 2
##   `(statut)` formationsBTSLicence
##   <chr>                <int>
## 1 Privé                1633
## 2 Public               7437
```

Nous pouvons constater que les deux formations les plus proposées sont le BTS et la licence , ces dernières sont pour la majorité des formations se trouvant dans des établissements publics. Dans la figure 2 , on remarque aussi que c'est en BTS et en Licence que les candidats postulent le plus. C'est donc pour cela que les établissements publics attirent plus d'étudiants que les établissements privés.

Par ailleurs, les étudiants semblent privilégier la licence au BTS, marquant ainsi leur intérêt pour la poursuite d'études, comme un master, dans un cadre plus académique et théorique. Contrairement au BTS, qui est une formation professionnalisante visant une insertion rapide sur le marché du travail après l'obtention du diplôme, la licence semble ouvrir la voie vers des études supérieures plus approfondies.

Répartition des candidats en Licence par Type de Bac

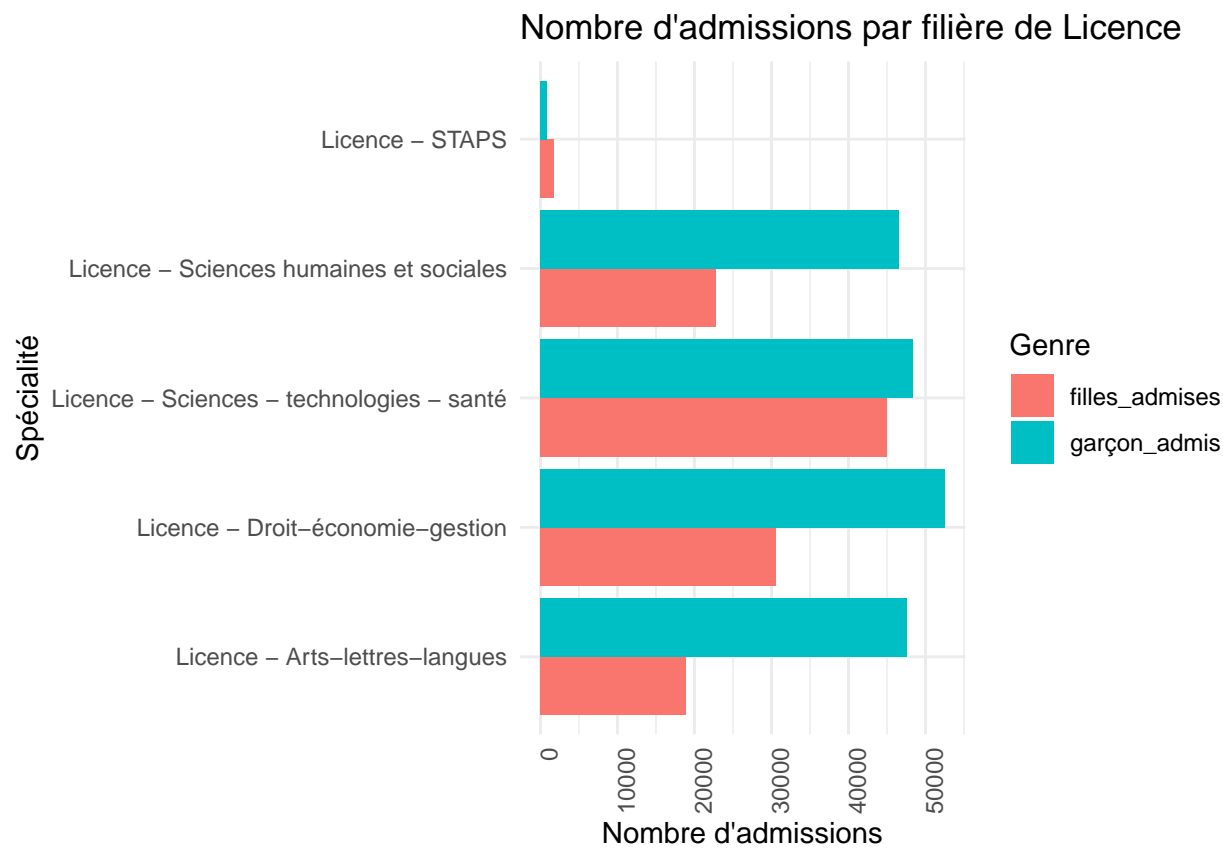
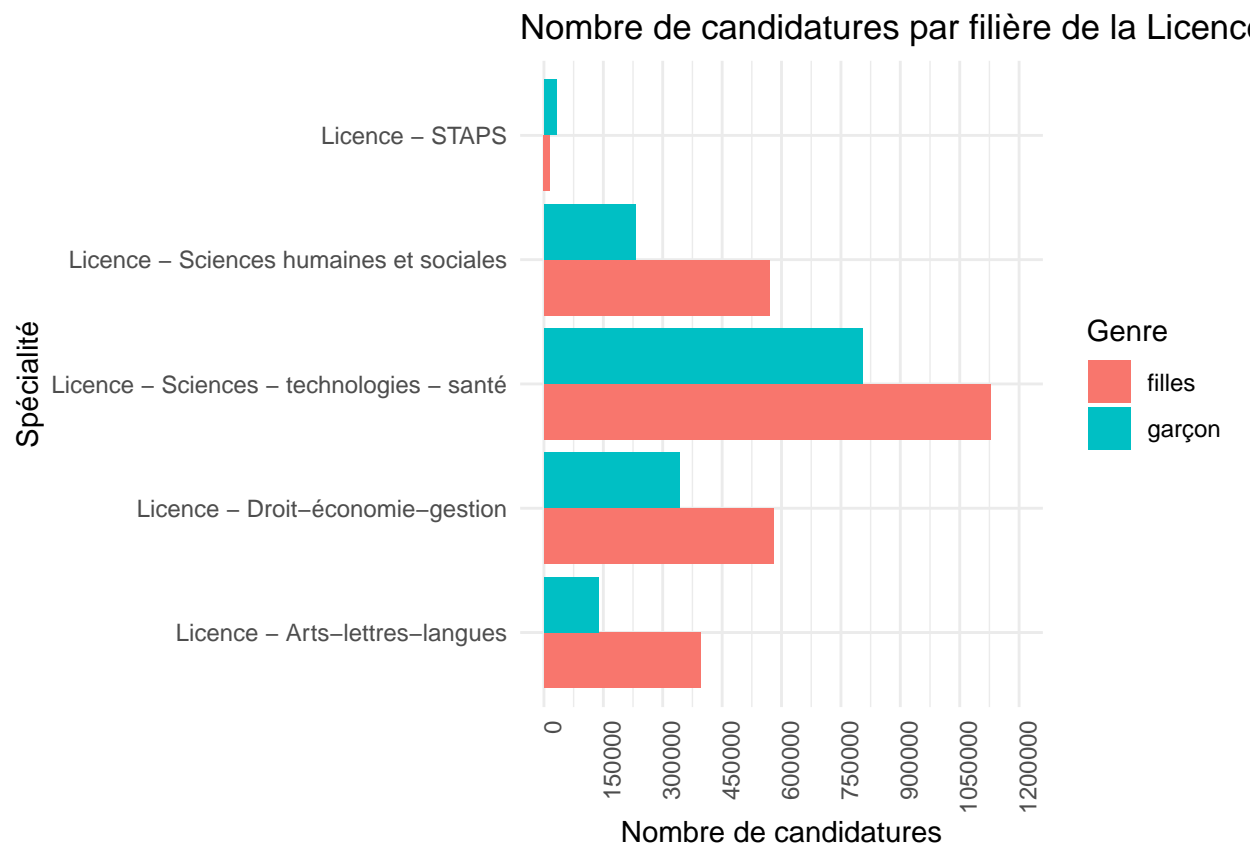


On s'intéresse particulièrement au diplôme le plus demandé par les candidats, qui est la licence.

Dans notre diagramme, nous constatons que les candidats titulaires d'un Bac Général postulent le plus pour la Licence, cela s'explique par le fait que le baccalauréat général offre une formation plutôt générale avec le choix de plusieurs spécialités (littérature, histoire, science ...). De plus, le Bac Pro et le Bac Technologique est plus technique ou pratique que théorique donc les bacheliers issues de ces Bacs sont moins attirés par la Licence.

Donc le type de bac a un grand impact sur le choix de la formation par les candidats.

Impact du genre sur le choix de la filière en Licence



Le premier constat c'est que la filière la plus demandée est Sciences—technologies—santé indépendamment du genre. Cela peut s'expliquer par le fait que de nos jours la plupart des étudiants ont une appétence pour le domaine de la science.

A l'inverse, la filière STAPS attire extrêmement moins de candidats, cela peut s'expliquer par le fait que la plupart de jeunes ne se projettent pas dans le domaine du sport pour gagner leur vie.

On constate par ailleurs que malgré que la filière Sciences—technologies—santé est très demandée par rapport aux autres (par exemple Droit-Economie-gestion), on remarque qu'il y a à peu près le même nombre d'admissions. Donc la filière Sciences—technologies—santé est plus sélective par rapport aux autres.

Concernant le genre, on remarque que dans toutes les filières (à l'exception de STAPS), c'est les filles qui candidatent en majorité. Cependant les garçons sont les plus admis.

A partir de notre jeu de données, on ne peut donc pas déterminer si le genre a un impact sur le nombre d'admissions.

Conclusion

Les conclusions qu'on peut tirer de cette étude sont les suivantes :

- La taille de la population d'une région semble avoir une influence directe sur le nombre d'établissements présents dans cette région, mais pas sur le nombre de formations proposées par région.
- Le statut de l'établissement (public ou privé) n'a pas d'influence directe sur le taux de remplissage de l'établissement.
- La sélectivité de la formation n'a pas une influence significative sur son taux de remplissage. D'ailleurs, même les formations sélectives ne remplissent pas toujours leur capacité d'accueil; de plus la sélectivité elle-même ne semble pas être influencée par la région.
- Le type de bac a un impact sur le choix de la formation des candidats en licence.
- Le genre n'a pas d'impact sur les admissions ni sur les candidatures en licence.

Références

<https://www.studyrama.com/formations/diplomes/bac>

<https://www.data.gouv.fr/fr/datasets/r/1d916b7c-bd4c-4951-845a-70f7ad7c17db>

<https://www.data.gouv.fr/fr/datasets/r/98a7a1e5-64e8-4098-9180-ee4adc451896>