

SNT - DONNÉES STRUCTURÉES (CSV)

Sur le site data.gouv.fr (plateforme d'*open data* de l'État français), j'ai récupéré le jeu de données¹ suivant :

Parcoursup 2020 - vœux de poursuite d'études et de réorientation dans l'enseignement supérieur et réponses des établissements

Ce jeu de données provient d'un service public certifié

Ce 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 – hors apprentissage – à la fin du processus d'affectation de la plateforme Parcoursup pour la session 2020 (du 22 janvier au 24 septembre 2020). Les données sont tirées de la plateforme Parcoursup 2020 et prennent en compte l'ensemble de la campagne jusqu'à la date du 24 septembre 2020. Elles sont disponibles sous forme de base de données à télécharger et sous forme de visualisation graphique « dataviz ».

Ce jeu de données est bâti à partir des données Parcoursup de la campagne 2020. L'application Parcoursup est la plateforme nationale de préinscription mise en place par le Ministère de l'Enseignement supérieur, de la recherche et de l'innovation permettant aux élèves de candidater à l'entrée dans l'enseignement supérieur.

Il couvre l'ensemble des candidats ayant au moins un vœu d'orientation validé en phase principale et/ou complémentaire, ce parmi les 12 760 formations proposées hors apprentissage. Il couvre ainsi 985 538 candidats. Un sous ensemble de données porte spécifiquement sur les 663 859 néo-bacheliers parmi ces candidats.

Ce jeu de données est complété par une « dataviz » proposant une visualisation des principaux indicateurs. Ainsi, sont représentés graphiquement pour chaque formation détaillée, la capacité d'accueil, le nombre de vœux faits par les candidats dans une formation, le nombre de propositions d'admission faites par l'établissement sur cette formation, les rangs du dernier appelé en phase principale, et le nombre de candidats qui ont accepté cette proposition dans l'ensemble des phases de la procédure.

En bas de cette page, on trouve (entre autres) :

Ressources

Export au format CSV

Session: session[[text](#)] Statut de l'établissement de la filière de formation (public, privé...): contrat_etab[[text](#)] Code UAI de l'établissement: cod_uai[[text](#)] Établissement:...

csv 635 Disponible

PRÉVISUALISER

TÉLÉCHARGER



Export au format JSON

Session: session[[text](#)] Statut de l'établissement de la filière de formation (public, privé...): contrat_etab[[text](#)] Code UAI de l'établissement: cod_uai[[text](#)] Établissement:...

json 1 Disponible

PRÉVISUALISER

TÉLÉCHARGER



Nous allons télécharger les données au format CSV afin de trouver certaines informations.

1. Télécharger le fichier sur le site officiel² et l'enregistrer dans vos documents avec le nom de fichier suivant : `parcoursup2020_opendata.csv`.

Télécharger également le fichier PDF qui explique la méthodologie utilisée pour produire ce fichier (il explique, entre autres, la signification de chaque colonne).

2. Ouvrir le fichier CSV avec un tableur comme LibreOffice Calc, sachant qu'on peut lire sur le site la phrase suivante : *Le CSV utilise le point-virgule (;) comme séparateur.*

Enregistrer le fichier au format du tableur utilisé (.ods pour LibreOffice Calc).

¹ <https://www.data.gouv.fr/fr/datasets/parcoursup-2020-voeux-de-poursuite-detudes-et-de-reorientation-dans-lenseignement-superieur-et-reponses-des-etablissements/>

² Si ce fichier n'est pas disponible, vous pouvez le télécharger sur mon site : snt.mathemathieu.fr/fr-esr-parcoursup.csv (≈ 10 Mo).

3. Déterminer l'établissement dont le code UAI est 0811293R.
4. Combien de formations sont enregistrées dans cette base de données ?
5. Créer des filtres pour toutes les colonnes : Données → AutoFiltre.
- a. Déterminer la liste des établissements du Tarn qui sont privés sous contrat d'association.
- b. On souhaite déterminer combien d'établissements sont des IFSI.
Commencez par utiliser le filtre pour n'afficher que ces établissements.
En bas de la colonne correspondante, utiliser la formule suivante qui permet de calculer le nombre de cellules non vides dans une plage qui contient éventuellement des lignes masquées (ce que fait le filtre) :
- $$=SOUS.TOTAL(3;plage).$$
- c. Déterminer le nombre d'IFSI qui sont des établissements publics.
6. Je suis né un 12 novembre. Par curiosité, j'aimerais savoir si au moins un établissement a un code UAI qui contient le nombre 1211. Utiliser les filtres : quels sont les établissements correspondants ?
7. a. Déterminer les coordonnées géographiques (au format DD) de l'unique formation sélective du Tarn qui est une CPGE MPSI.
- b. En déduire le nom de la rivière qui est située juste à côté de cet établissement.
8. Dans le fichier PDF qui décrit la méthodologie, trouver à quoi correspond la colonne `capa_fin`.
Trouver les formations dont l'établissement est public, situé dans le Tarn et proposant plus de 100 places.
9. a. Déterminer la colonne donnant le pourcentage d'admis néo bacheliers avec mention Très Bien au bac.
- b. Déterminer les formations de l'Université Toulouse 1 Capitole qui ont le pourcentage le plus élevé d'admis néo bacheliers avec mention Très Bien au bac.
10. Déterminer les trois BTS de l'académie de Lyon qui ont plus de 90 % d'admis néo bacheliers sans mention au bac.
11. Combien de candidats ont fait le vœu d'intégrer la Licence de Mathématiques proposée dans le Tarn ? Parmi eux, combien y avait-il de candidates ? Combien de places propose cette formation ?