

3<sup>ème</sup> année

## VISUALISATION DES DONNEES

## Data Visualization

# Projet 2020-2021

**Travail demandé :**

A partir des bases de données proposées, vous devez faire le travail suivant :

- 1- Décrire les données et la problématique qu'elles représentent
- 2- Construire des chaines de traitement (workflow) en python permettant de fournir des visualisations de ces données
- 3- Interpréter les résultats graphiques
- 4- Faire une présentation PowerPoint de 5 minutes
- 5- Fournir le rapport sous forme d'un seul fichier format : « Projet-X- VizData-Groupe-Y.ipynb »

# **Projet N° 1 : Classement thématique des sujets de journaux télévisés (janvier 2005 - septembre 2020)**

## **Description du jeu de données :**

Classement thématique des sujets diffusés sur les journaux télévisés du soir de six chaînes (TF1, France 2, France 3, Canal +, Arte, M6) pour la période janvier 2005 à septembre 2020.

L'Ina indexe depuis 1995 la totalité des sujets diffusés sur les journaux télévisés du soir des six chaînes dites « historiques » (TF1, France 2, France 3, Canal +, Arte, M6), par une addition de mots clés appartenant à un thésaurus. A partir de cette indexation, des algorithmes thématisent les sujets entre quatorze rubriques. Un sujet est classé sous une rubrique et une seule.

## **Lien de téléchargement :**

<https://www.data.gouv.fr/fr/datasets/classement-thematique-des-sujets-de-journaux-televises-janvier-2005-septembre-2020/>

# **Projet N° 2 : Données du portail d'information Trouver Mon Master TMM - Parcours de formation de master**

## **Description du jeu de données :**

Ce jeu de données présente les informations destinées au portail [Trouver Mon Master \(TMM\)](#) au titre de l'année universitaire 2019-2020.

Ces informations sont fournies par les établissements accrédités en vue de la délivrance de diplômes nationaux de master (DNM).

Le portail TMM a en effet pour vocation de recenser l'ensemble des DNM existants. Il ne recense ni les autres diplômes conférant grade de master, ni les autres diplômes du deuxième cycle universitaire.

Pour en savoir plus, [consultez la documentation dédiée aux trois jeux de données](#) relatifs aux informations destinées au portail Trouver Mon Master (TMM).

## **Lien de téléchargement :**

<https://www.data.gouv.fr/fr/datasets/donnees-du-portail-dinformation-trouver-mon-master-tmm-parcours-de-formation-de-master/>

## **Projet N° 3 : Open Food Facts - Produits alimentaires : ingrédients, nutrition, labels**

### **Description du jeu de données :**

[Open Food Facts](#) répertorie les informations sur les produits alimentaires : ingrédients, informations nutritionnelles, labels etc.

Les données proviennent majoritairement de la collecte citoyenne (crowdsourcing) des informations.

### **Lien de téléchargement :**

<https://www.data.gouv.fr/fr/datasets/open-food-facts-produits-alimentaires-ingredients-nutrition-labels/>

## **Projet N° 4 : Données hospitalières relatives à l'épidémie de COVID-19**

### **Description du jeu de données**

Le présent jeu de données renseigne sur la situation hospitalières concernant l'épidémie de COVID-19.

Cinq fichiers sont proposés :

- **Les données hospitalières relatives à l'épidémie du COVID-19 par département et sexe du patient** : nombre de patients hospitalisés, nombre de personnes actuellement en réanimation ou soins intensifs, nombre cumulé de personnes retournées à domicile, nombre cumulé de personnes décédées.
- **Les données hospitalières relatives à l'épidémie du COVID-19 par région, et classe d'âge du patient** : nombre de patients hospitalisés, nombre de personnes actuellement en réanimation ou soins intensifs, nombre cumulé de personnes retournées à domicile, nombre cumulé de personnes décédées.
- **Les données hospitalières quotidiennes relatives à l'épidémie du COVID-19 par département** : nombre quotidien de personnes nouvellement hospitalisées, nombre quotidien de nouvelles admissions en réanimation, nombre quotidien de personnes nouvellement décédées, nombre quotidien de nouveaux retours à domicile.
- **Les données relatives aux établissements hospitaliers par département** : nombre cumulé de services ayant déclaré au moins un cas.

- **Les données relatives à les nouvelles admissions en réanimation par région :**  
nombre de nouveaux patients admis en réanimation dans les 24 dernières heures.

**Lien de téléchargement :**

**<https://www.data.gouv.fr/fr/datasets/donnees-hospitalieres-relatives-a-lepidemie-de-covid-19/>**

<p><b>Projet N° 5 : Base de données de la pollution aérienne aux abords des écoles et crèches d’Île-de-France</b></p>
-----------------------------------------------------------------------------------------------------------------------

**Description du jeu de données :**

Cette base de données contient les concentrations modélisées en microgramme par mètre cube de moyenne journalière annuelle pour trois polluants aériens (NO<sub>2</sub>, PM<sub>10</sub> et PM<sub>2,5</sub>) aux abords des crèches et établissements scolaires d’Île-de-France pour les années de 2012 à 2017.

Ces données ont été compilées par le croisement des bases de données des établissements d'accueil du jeune enfant ([base EAJE](#)) et de [la base adresse des établissements d'enseignement du premier et second degrés](#) avec les fichiers de qualité de l'air modélisés par Airparif (à raison d'un fichier par an et par polluant).

Chaque mesure correspond au niveau de concentration relevé au point géolocalisé correspondant à l'adresse postale.

Ces données ne renseignent donc pas sur la qualité de l'air dans les cours des établissements ou à l'intérieur des bâtiments.

**Lien de téléchargement :**

**<https://www.data.gouv.fr/fr/datasets/r/cc16163c-aca0-4977-97da-8ce592f78de1>**

## **Projet N° 6 : Les étudiants étrangers dans les universités**

### **Description du jeu de données :**

Tableaux statistiques : évolution des effectifs d'étudiants de nationalité étrangère dans les universités ; répartition des étudiants de nationalité étrangère dans les universités (hors IUFRM) selon l'origine et la discipline.

### **Lien de téléchargement :**

<https://www.data.gouv.fr/fr/datasets/r/ff8c4cf4-cf52-432d-8a55-3c91d1a52b56>

## **Projet N° 7 : Données de scolarité de l'université Sorbonne Paris Nord (ex. Paris 13)**

### **Description du jeu de données :**

L'université Sorbonne Paris Nord a enregistré dans son système d'information (logiciel Apogée), des données sur l'inscription des étudiants pour chaque année universitaire entre 2006(-2007) et 2015(-2016). Ces données portent sur les diplômes préparés, les étapes pour y parvenir, le régime (s'il s'agit de formation initiale ou d'apprentissage), les composantes concernées (UFR, IUT, etc.), et l'origine des étudiants (type de Bac, académie d'origine, nationalité). Chaque entrée concerne l'inscription principale d'un étudiant à l'université pour une année.

### **Lien de téléchargement :**

<https://www.data.gouv.fr/fr/datasets/donnees-de-scolarite-de-luniversite-paris-13/>

## **Projet N° 8 : Parcoursup 2020 - vœux de poursuite d'études et de réorientation dans l'enseignement supérieur et réponses des établissements**

### **Description du jeu de données :**

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.

### **Lien de téléchargement :**

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

## **Projet N° 9 : Concentrations annuelles moyennes de polluants dans l'air ambiant issues du réseau permanent de mesures en région Île-de-France sur 5 ans glissants**

### **Description du jeu de données :**

Concentrations issues du réseau fixe de mesure des principaux polluants réglementés dans l'air : SO<sub>2</sub>, NO<sub>2</sub>, NO<sub>x</sub>, PM<sub>10</sub>, PM<sub>2.5</sub>, O<sub>3</sub>, CO sur la région Île-de-France en moyennes annuelles.

### **Lien de téléchargement :**

<https://www.data.gouv.fr/fr/datasets/concentrations-annuelles-moyennes-de-polluants-dans-lair-ambiant-issues-du-reseau-permanent-de-mesures-en-region-ile-de-france-sur-5-ans-glissants/>

## **Projet N° 10 : Visualisation de la Chirurgie du nerf médian au canal carpien guidée par l'imagerie**

La chirurgie du nerf médian au canal carpien correspond à l'ensemble des interventions réalisées pour libérer le nerf médian au niveau du poignet, ou à le réparer en cas de lésion. Il s'agit d'un nerf qui chemine depuis la racine du bras jusqu'à la main, en traversant la région du coude, de l'avant-bras et du poignet. Il se termine au niveau de la main, en se divisant en branches sensitives (qui donnent la sensibilité dans les doigts) et une branche motrice (qui permet de contracter les muscles à la base du pouce).

La compression du nerf médian survient lorsque la pression exercée sur le nerf à un endroit de son trajet est trop importante pour que le nerf fonctionne correctement. Il peut s'agir d'une compression aiguë, survenant lors d'un effort important, ou d'une compression chronique, évoluant sur plusieurs mois ou plusieurs années. A part, la compression dynamique, représentée par l'augmentation des pressions dans les loges musculaires lors d'efforts prolongés, comme chez certains sportifs et dans certains travaux manuels.

L'intervention chirurgicale réalisée pour soulager le syndrome du canal carpien répond à un cahier des charges précis. L'objectif principal est de libérer le nerf de la pression qui l'empêche de fonctionner normalement. Cet objectif est atteint en ouvrant le canal dans sa portion ligamentaire : c'est la section du ligament annulaire antérieur du carpe.

Ce projet consiste à implémenter un simulateur de chirurgie guidée par les images médicales et les graphiques pour illustrer l'intervention.