

|  |
| --- |
| Projet Data Science |
| Analyse de la qualité de l’air |
| 2023-2024 |



Martin Lucas & Chouchane Habiba FISA 25

Table des matières

[1. Contexte 1](#_Toc162350634)

[2. L’expérience 1](#_Toc162350635)

[3. Objectif du TP 2](#_Toc162350636)

[4. Méthodologie 2](#_Toc162350637)

[5. Taches 2](#_Toc162350638)

[1. Première tâche 2](#_Toc162350639)

[2. Deuxième tâche 3](#_Toc162350640)

[3. Troisième tâche 3](#_Toc162350641)

[4. Quatrième tâche 3](#_Toc162350642)

# Contexte

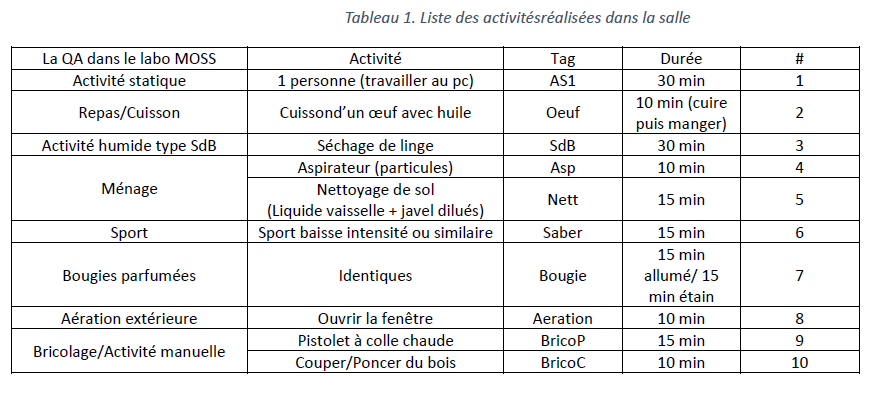
L’étude de la Qualité de l’Air Intérieur (QAI) est une préoccupation de santé publique, qui est notamment prise en compte dans le 4e plan national santé environnement. En effet, en calculant le temps passé au bureau, en classe ou à la maison, nous restons plus de 80 % de notre vie à l’intérieur. Les efforts récents pour améliorer l’efficacité énergétique des bâtiments font que le renouvellement de l’air dans les environnements intérieurs est limité et, par conséquent, la pollution générée dans ce type d’environnement reste concentrée, ce qui fait que la concentration des polluants est cinq fois plus élevée à l’intérieur qu’en extérieur.

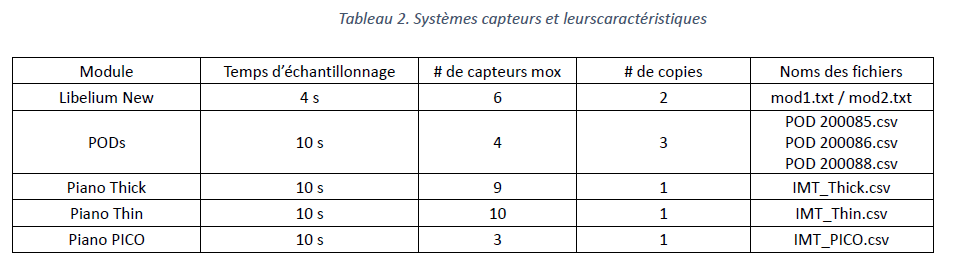
Les sources de pollution dans l’environnement intérieur sont multiples, mais les plus communs sont les émissions relatives à des travaux et des rénovations dans l’environnement (peintures, sol, isolations, …), à la présence de certains objets, d’animaux ou des émissions relatives à des activités humaines quotidiennes. Il est à noter que dès que les émissions causées par des travaux et rénovation sont stabilisées, la source principale de pollution dans l’environnement intérieur est l’être humain.

Des activités, comme la cuisson et le ménage, sont sources de plusieurs types de polluants, comme des particules fines (particulate matter en anglais, ou PM) et des composants organiques volatiles (COV), qui causent ou intensifient certaines maladies respiratoires comme l’asthme et des allergies. Quand les niveaux de pollution sont élevés pendant une longue période cela peut causer un ensemble de symptômes aigus (maux de tête, irritation des yeux, du nez ou de la gorge, toux sèche, peau sèche ou qui démange, vertiges et nausées, difficultés de concentration, fatigue et sensibilité aux odeurs) qui constituent le syndrome du bâtiment malsain (sick building syndrome en anglais, ou SBS). Ainsi, reconnaitre qu’une activité polluante est en cours, devient une étape essentielle pour développer des systèmes de ventilation automatiques et pour prévenir l’usager des dangers liés à cette activité.

# L’expérience

C’est dans ce contexte que nous avons fait l’expérience qui a généré les données utilisées pour ce TP. Cette expérience a été réalisée dans une salle de 13 m² (47 m³) avec des conditions environnementales (température et humidité) non contrôlées où 21 types de capteurs du type oxide métallique (MOX) ont été utilisés pour monitorer la QAI. Les capteurs sont distribués en 5 modules d’acquisition présentés dans le Tableau 2. Dans cette salle, 10 types d’activités quotidiennes ont été réalisées, elles sont listées dans le Tableau 1.





Les capteurs ont enregistré les émissions provenant de ces activités dans une base de données, et un calendrier a été rempli avec les moments de début et de fin de chaque activité (fichier activites.xlsx), constituant ainsi les labels de la base. En d'autres termes, nous savons que les mesures des capteurs entre le début et la fin d'une activité proviennent de cette activité (fichier « activites.xlsx »). L’ordre d’exécution des activités a été préétabli par le calendrier et elle a été construite à partir d’une chaine de Markov pour représenter la routine d’une maison dans les probabilités de transition entre des activités. Cette info peut être utilisée pour enrichir vos algorithmes. La matrice de transition de ces activités est présentée dans le fichier activites.xlsx.

# Objectif du TP

L’objectif de ce TP est de vous montrer toute la chaîne de travail d’un Data scientiste.

# Méthodologie

La méthodologie d’analyse est divisée en trois étapes :

1. Nettoyage/rangement des données dans une base propre (tache ½)
2. Extraction des caractéristiques au sein des données (tache 3)
3. Classification/prédiction et présentation des résultats mesurables (tache 4)

# Taches

## Première tâche

Enoncé :

Pendant des périodes d’expériences nous avons interrompu l’acquisition des données pour les sauvegarder. Par conséquent, nous avons des fichiers correspondent à des périodes différentes. La première tâche de ce TP est de grouper tous les fichiers qui ont le même nom dans un seul fichier organisé chronologiquement et sans échantillon dupliqué.

Travail :

## Deuxième tâche

Enoncé :

Chaque module d’acquisition a son propre temps d’échantillonnage. La deuxième tâche de ce TP est de synchroniser les données provenant de chaque module en utilisant le même temps d’échantillonnage et créer une base de données avec toutes les données disponibles dans un seul fichier .csv. Autrement dit, vous devez fusionner tous les fichiers en un seul qui contient toutes les colonnes des autres fichiers et sans lignes marqués comme NaN ou NULL. Ce fichier final sera nommé comme vous voulez et il sera la base des données sur laquelle la troisième tâche de ce TP sera faite.

Travail :

## Troisième tâche

Enoncé :

La troisième tâche de ce TP est de trouver la signature moyenne de chaque activité. Par signature, on considère la série temporelle (avec tous les capteurs) qui représente chaque activité en moyenne

Travail :

## Quatrième tâche

Enoncé :

La quatrième (et dernière) tâche de ce TP est de proposer une méthodologie pour faire la reconnaissance des activités en utilisant des données dans la base. Toutes les propositions de techniques doivent être justifiées. Vous devez également exécuter la méthodologie proposée et mesurer sa performance de reconnaissance d’activité.

Travail :