



# Groupe G3

Data Mining et Analyse de Patterns

Projet SDID 2025–2026

Matricules :

23618 – 23605 – 23635

26 janvier 2026

## Table des matières

<b>1</b>	<b>Objectif du groupe G3</b>	<b>2</b>
<b>2</b>	<b>Pré-requis</b>	<b>2</b>
<b>3</b>	<b>Structure du projet G3</b>	<b>2</b>
<b>4</b>	<b>Description fonctionnelle</b>	<b>2</b>
4.1	Accès aux données . . . . .	2
4.2	Prétraitement et normalisation . . . . .	3
4.3	Réduction de dimension . . . . .	3
4.4	Clustering DBSCAN . . . . .	3
<b>5</b>	<b>Visualisation</b>	<b>3</b>
<b>6</b>	<b>Artefacts générés</b>	<b>3</b>
<b>7</b>	<b>Exécution du pipeline</b>	<b>3</b>
<b>8</b>	<b>Conclusion</b>	<b>4</b>

## 1 Objectif du groupe G3

Le groupe G3 a pour objectif d'identifier les **comportements normaux de consommation électrique** à partir des données historiques stockées dans PostgreSQL, afin de fournir une base fiable pour la détection d'anomalies réalisée par le groupe G4.

Le travail repose sur des techniques de **data mining non supervisé**, notamment la réduction de dimension et le clustering.

## 2 Pré-requis

Les éléments suivants doivent être disponibles avant l'exécution du pipeline G3.

- Base PostgreSQL active (pipeline G2)
- Données présentes dans la table `power_consumption`
- Python 3.9 ou supérieur
- Environnement virtuel Python recommandé

## 3 Structure du projet G3

Le travail du groupe G3 est organisé selon la structure suivante.

```
G3_data_mining/  
    data_access/  
    preprocessing/  
    modeling/  
    visualization/  
    artifacts/  
    main.py  
    README.md
```

## 4 Description fonctionnelle

### 4.1 Accès aux données

Ce module assure la lecture des données historiques depuis la base PostgreSQL.

- Connexion à la base de données
- Exécution d'une requête SQL filtrée
- Chargement des données dans Pandas

## 4.2 Prétraitement et normalisation

Les données sont normalisées afin de rendre les variables comparables.

- Utilisation de `RobustScaler`
- Sauvegarde des paramètres de normalisation

## 4.3 Réduction de dimension

L'ACP permet de projeter les données dans un espace réduit et exploitable.

## 4.4 Clustering DBSCAN

DBSCAN permet d'identifier automatiquement les comportements normaux et les points atypiques.

# 5 Visualisation

Une visualisation ACP combinée à DBSCAN est générée afin d'analyser les profils de consommation.

*(Insérer ici l'image du clustering PCA + DBSCAN)*

# 6 Artefacts générés

Les artefacts suivants sont produits automatiquement.

```
scaler.pkl  
pca.pkl  
dbscan_params.json  
clusters.json
```

# 7 Exécution du pipeline

Commande d'exécution du pipeline G3.

```
python -m G3_data_mining.main
```

## 8 Conclusion

Le groupe G3 fournit une modélisation claire et robuste des comportements normaux de consommation électrique.