

Projet BDD : Analyse des Sections Internationales et BFI

1. Présentation du sujet et des données

Ce projet a pour objectif de modéliser et analyser l'offre de formation en sections internationales et classes BFI (Baccalauréat Français International), proposées par les établissements français en France et à l'étranger.

Les données sont issues de l'Éducation Nationale, disponibles sur :

<https://data.education.gouv.fr/explore/dataset/fr-en-sections-internationales>

Cette base de données permet de centraliser et structurer l'information pour répondre à plusieurs besoins : analyse de la répartition géographique, comparaison des langues proposées, soutien à l'orientation des familles expatriées, et accompagnement des politiques éducatives.

2. Modèle Conceptuel de Données (MCD) final

Le modèle conceptuel a évolué par rapport à la première version. Les principales modifications sont :

- Ajout d'une entité Niveau (école, collège, 2nde et BFI, BFI)
- Ajout d'une entité Ville, Académie et Pays pour gérer la localisation des établissements (pour éviter la redondance)
- Normalisation des langues (ex : AMÉRICAINNE, BRITANNIQUE → ANGLAIS)
- Gestion correcte de l'année scolaire avec les effectifs en fonction des langues proposées par les sections

3. Volumes concernés

Données du fichier source :

- Nombre de lignes dans le fichier initial : environ 1000
- Nombre d'onglets (fichiers CSV générés) : 8 (pays, académie, ville, établissement, section_internationale, langue, annee_scolaire, niveau)

Dans la base de données (MySQL) :

- Nombre de tables : 8

- Nombre de lignes par table : variable selon les entités (par exemple : plusieurs centaines d'établissements)
- Nombre de variables / attributs retenus : entre 1 et 3 par table, après simplification et normalisation

4. Méthode et retour d'expérience

Outils et traitements effectués :

- Python / pandas : nettoyage des colonnes (suppression des caractères spéciaux, harmonisation des noms), génération des fichiers CSV
- MySQL via phpMyAdmin : création de la base de données, des tables, importation des CSV
- Scripts SQL : automatisation de la création des tables

Retour d'expérience :

- Gestion complexe des caractères spéciaux dans les titres de colonnes et les données
- Nettoyage des dates incohérentes (par exemple 'nov -26' remplacé par '2026')
- Correction manuelle de l'association pays/académie
- Normalisation des langues (fusionner les variantes : AMÉRICAIN → ANGLAIS)
- Création d'identifiants uniques pour les villes sans code INSEE applicable internationalement

5. Résultats — Requêtes SQL et analyses

Voici plusieurs exemples de requêtes réalisées pour analyser les données :

Nombre d'établissements par académie (répartition géographique)

```
SELECT id_academie, COUNT(*) AS NombreEtablissements
FROM etablissement GROUP BY id_academie;
```

Interprétation : Cette requête permet de visualiser la répartition des établissements par académie et d'identifier les zones sous ou sur-équipées.

Répartition des sections par langue (analyse linguistique)

```
SELECT l.NomLangue, COUNT(DISTINCT s.id_section) AS
NombreSections FROM langue l JOIN section_internationale s
ON l.id_section = s.id_section GROUP BY l.NomLangue ORDER
BY NombreSections DESC;
```

Interprétation : Cette requête montre quelles langues sont les plus représentées dans l'offre de sections internationales.

Liste des établissements à l'étranger proposant des sections internationales

```
SELECT e.NomEtablissement, v.NomVille, p.NomPays FROM  
etablissement e JOIN ville v ON e.id_ville = v.id_ville JOIN pays p  
ON v.id_pays = p.id_pays WHERE p.NomPays != 'France';
```

Interprétation : Identifie l'offre internationale hors de France.

6. Conclusion

Ce projet a permis de mettre en place une base de données exploitable à partir de données publiques, de révéler des tendances géographiques et linguistiques, et de renforcer la maîtrise de l'importation, du nettoyage et de l'analyse de données. Ce travail peut servir d'outil d'aide à la décision pour les familles, les établissements et les décideurs du réseau éducatif français à l'international.