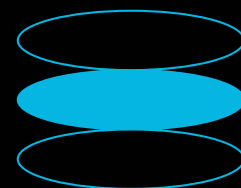


ANALYSE DES RÉSULTATS DU BACCALAURÉAT ET DE L'IPS DES LYCÉES PARISIENS (2016–2019)

Bases de Données Avancées

INESS IRBOUH & IMENE ZEBIRI



Bienvenue dans cet exercice de base de données conçu pour vous aider à maîtriser les concepts fondamentaux. Ce projet s'inscrit dans le cadre du cours de **Bases de Données Avancées**. Il vise à concevoir, implémenter et interroger une base de données relationnelle afin d'analyser les **résultats au baccalauréat des lycées parisiens** et de les mettre en relation avec un indicateur socio-économique : l'**Indice de Position Sociale (IPS)**.

La problématique centrale du projet est la suivante : ***Dans quelle mesure le niveau socio-économique des lycées parisiens, mesuré par l'IPS, influence-t-il leurs performances au baccalauréat entre 2016 et 2019 ?***

L'analyse se concentre exclusivement sur les lycées situés à Paris, identifiés par des codes de communes compris entre **75101 et 75120**, correspondant aux vingt arrondissements parisiens.

Suivez attentivement les instructions et n'hésitez pas à demander de l'aide si vous rencontrez des difficultés.

Objectifs

À la fin de cet exercice, vous serez capable de :

- Comprendre les concepts de base des bases de données relationnelles.
- Concevoir et créer une base de données simple.
- Insérer, modifier et supprimer des données.
- Effectuer des requêtes pour extraire des informations.
- Initiation à l'utilisation conjointe d'un langage de programmation (Python) et de votre base de données SQL.

Sources

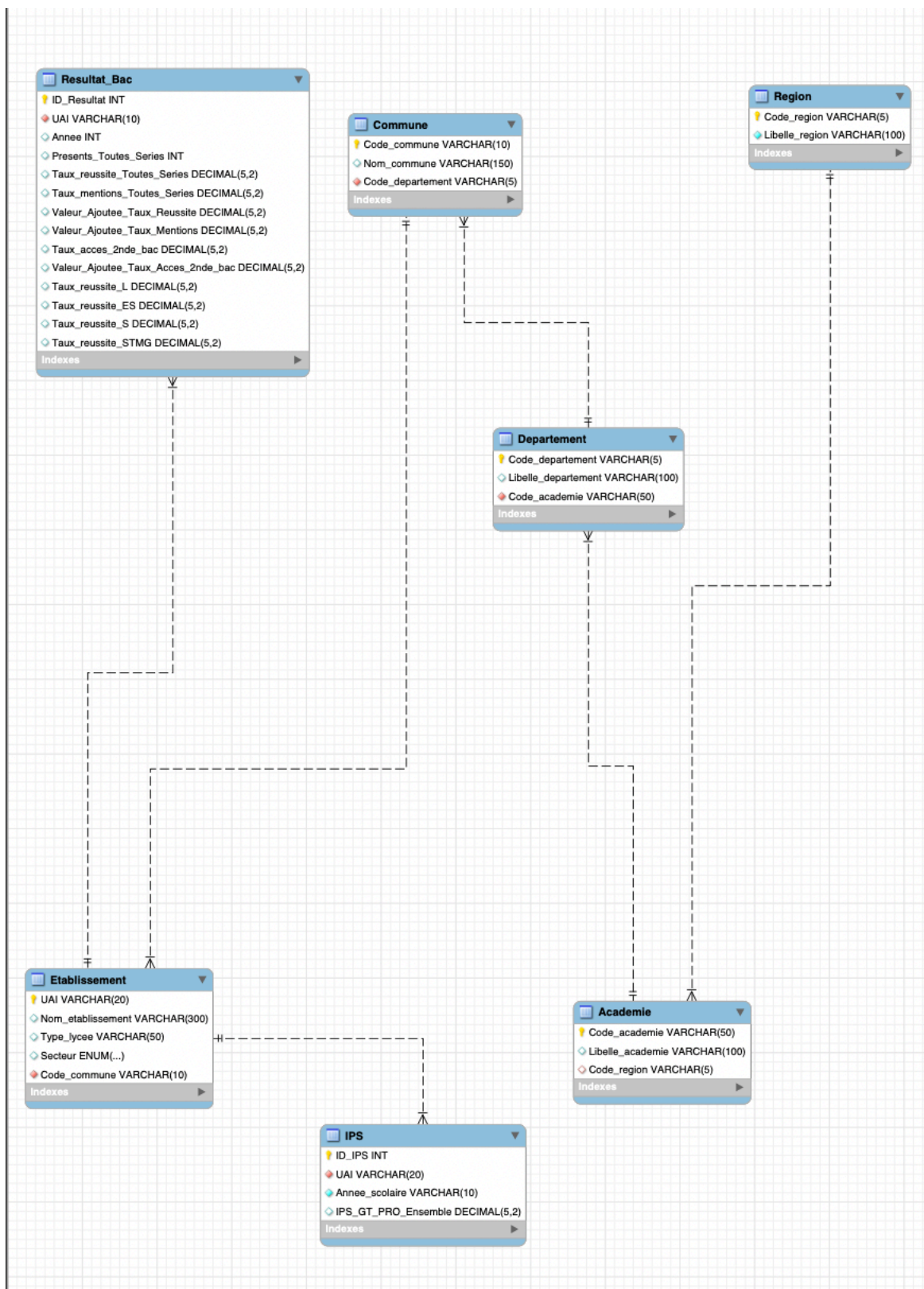
Les données data.gouv utilisées proviennent de jeux de données institutionnels portant sur :

- les résultats au baccalauréat : [Indicateurs de valeur ajoutée des lycées d'enseignement général et technologique \(ancien\)](#).
- l'indice de position sociale (IPS) : [Indices de position sociale des lycées \(2016-2021\)](#).
- la localisation administrative des établissements : [Annuaire de l'éducation](#)

Le périmètre a été volontairement restreint aux **lycées parisiens**, afin de garantir une homogénéité territoriale et de faciliter les comparaisons entre établissements situés dans un même environnement urbain de 2016 à 2019, avant l'épidémie du Covid-19 qui peuvent faussés l'analyse au vu des résultats exceptionnels.

Étapes

1. **Conception de la base de données** : Créer la Database Projet_BD_Resultats_Bac, Définir les tables, les colonnes et les types de données comme le schéma fourni.
2. **Nettoyage sur Excel** des colonnes, choix de la période qu'on souhaite couvrir ainsi que du périmètre des données.
3. **Création de la base de données** : Utiliser un SGBD (Système de Gestion de Base de Données) comme SQLite, MySQL ou PostgreSQL pour créer la base de données et les tables.
4. **Insertion de données** : Ajouter les données des bases de données que nous vous avons fourni au dessus grâce a INSERT INTO... values.
5. **Requêtes** : Effectuer des requêtes pour extraire des informations spécifiques.



Tables en détails

Tables	Colonnes	Clé primaire	Clé(s) externe(s)
Region	Code_region, Libelle_region	Code_region	/
Departement	Code_departement, Libelle_departement, Code_academie	Code_departement	Code_academie → Academie(Code_academie)
Commune	Code_commune, Nom_commune, Code_departement	Code_commune	Code_departement → Departement(Code_departement)
Academie	Code_academie, Libelle_academie, Code_region	Code_academie	Code_region → Region(Code_region)
Etablissement	UAI, Nom_etablissement, Type_lycee, Secteur, Code_commune	UAI	Code_commune → Commune(Code_commune)
IPS	ID_IPS, UAI, Annee_scolaire, IPS_GT_PRO_Ensemble	ID_IPS	UAI → Etablissement(UAI)
Resultat_Bac	ID_Resultat, UAI, Annee, Presents_Toutes_Series, Taux_reussite_Toutes_Series, Taux_mentions_Toutes_Series, Valeur_Ajoutée_Taux_Reussite, Valeur_Ajoutée_Taux_Mentions, Taux_acces_2nde_bac	ID_Resultat	UAI → Etablissement(UAI)

Vous devrez définir des relations similaires entre les tables..

Outils utilisés

- **MySQL Workbench** pour la création, la gestion de la base de données, rédaction des requêtes d'analyse, et enfin exécuter et déboguer.
- **Visual Studio Code** pour le code Python.
- **Canva** pour créer notre énoncé.

Difficultés rencontrées

Plusieurs difficultés techniques ont été rencontrées lors de la préparation des données :

- **Difficulté d'importation des fichiers Excel**
Nous avons rencontré des problèmes lors de la tentative d'importation directe des tableaux Excel dans MySQL. Les

outils d'import automatique de MySQL Workbench ne permettaient pas d'insérer correctement les données, en raison de problèmes de format et d'erreurs lors de la reconnaissance des colonnes.

- **Contraintes liées à l'environnement macOS**

Iness Irbouh travaillant sur Mac, la version spécifique de MySQL Workbench présentait certaines limitations. Les schémas créés automatiquement n'étaient pas visibles dans l'interface graphique, ce qui rendait leur lecture et leur vérification difficiles. De plus, les fonctionnalités d'export wizard et d'import wizard n'étaient pas utilisables, compliquant la gestion des données.

- **Choix d'une insertion manuelle des données**

Pour contourner ces difficultés, nous avons choisi d'insérer les données directement à l'aide de requêtes SQL de type INSERT INTO ... VALUES. Cette méthode nous a permis de garder le contrôle sur les valeurs insérées et de respecter précisément la structure de la base.

- **Problèmes liés aux clés primaires et aux clés étrangères**

Lors de l'insertion des données, plusieurs erreurs sont apparues en lien avec les clés primaires et les clés étrangères. Certaines clés primaires étaient dupliquées, tandis que certaines clés étrangères faisaient référence à des enregistrements inexistantes dans les tables parentes. Ces problèmes ont été résolus en vérifiant l'unicité des identifiants et en respectant l'ordre logique d'insertion des tables mais nous avons tout de même des warnings.

- **Difficultés de liaison entre Python et MySQL**

La connexion entre Python et MySQL a également posé problème au départ, notamment lors de la configuration des bibliothèques et des paramètres de connexion. Après plusieurs essais, nous avons réussi à établir une connexion fonctionnelle, permettant l'exploitation de la base depuis Python.

- **Apports de ces difficultés**

Ces différentes difficultés ont été formatrices et ont permis de mieux comprendre le rôle des clés primaires et des clés étrangères, l'importance de l'intégrité référentielle, ainsi que les contraintes pratiques liées à l'utilisation d'outils de gestion de bases de données sur différents systèmes.

Requêtes

Partie 1 : Définition du périmètre d'étude et premiers constats sur les lycées parisiens

Requête 1.1 Liste des lycées parisiens (UAI + nom)

- **But** : Construire notre population d'étude : uniquement les établissements situés à Paris.
- **Idée** : Paris = Code_commune qui commence par **75** (arrondissements : 75101 → 75120).
- **Résultat attendu** : la liste des **UAI + noms** des lycées concernés.

Requête 1.2 Privé d'abord, public ensuite

- **But** : Séparer les lycées parisiens selon leur **secteur** (privé / public).
- **Pourquoi ?** : Parce qu'on veut tester l'hypothèse "le secteur influence-t-il IPS et réussite ?".
- **Résultat attendu** : liste des lycées parisiens triée avec :
 - **privé en premier**,
 - puis **public**.

Requête 1.3 Quels arrondissements ont au moins un lycée ?

- **But** : Vérifier la répartition territoriale : est-ce que chaque arrondissement possède au moins un lycée ?
- **Important** : Les codes arrondissements parisiens sont : **75101, 75102, ..., 75120**.
- **Résultat attendu** : la liste distincte des arrondissements où il existe au moins un établissement présent dans la table Etablissement.

Requête 1.4 Repérer les années “surprenantes” : réussite > ce que l’IPS laisse prévoir

- **But** : Identifier les années où Paris “surperforme” :
 - IPS moyen (attendu) vs réussite moyenne (observée).
- **Résultat attendu** : pour chaque année :
 - IPS moyen parisien
 - réussite moyenne parisienne
 - écart (réussite – IPS)

Requête 1.5A Le “meilleur” lycée parisien

- **But** : Trouver l'établissement ayant le **taux de réussite le plus élevé**.
- **Résultat attendu** :
 - UAI, nom
 - arrondissement
 - année
 - taux réussite, mentions
 - valeur ajoutée réussite & mentions

Requête 1.5B Le “pire” lycée parisien

- **But** : Trouver l'établissement ayant le **taux de réussite le plus faible**.
- **Résultat attendu** : mêmes informations que 1.5A.

Requête 1.6 IPS d'un lycée précis sur une année précise

- **But** : Consulter l'IPS d'un lycée spécifique pour faire un zoom qualitatif.
- **Exemple demandé** : lycée d'Alembert en **2016-2017** (UAI = 0750650Z).
- **Résultat attendu** : UAI, nom, année scolaire, IPS.

Partie 2 : Cartographier Paris “Chaque lycée dans son arrondissement”

Requête 2.1 Associer chaque lycée à son arrondissement

- **But** : Lier Etablissement à Commune pour obtenir le **nom de l'arrondissement**.
- **Résultat attendu** : UAI, nom du lycée, arrondissement.

Requête 2.2 IPS + Résultats Bac réunis : la rencontre des deux mondes

- **But** : Mettre dans la même table logique :
 - IPS (contexte socio-éco)
 - Résultats bac (performance)
- **Résultat attendu** : pour chaque lycée parisien :
 - UAI, nom
 - année (bac)
 - année scolaire (IPS)
 - IPS
 - taux réussite + taux mentions

Ici, on commence vraiment à tester : **IPS <-> performance**.

Requête 2.3 Comparaison par arrondissement : combien de lycées, IPS moyen, réussite moyenne

- **But** : Mesurer les inégalités territoriales dans Paris.
- **Résultat attendu** : par arrondissement :
 - nombre de lycées
 - IPS moyen
 - réussite moyenne

Requête 2.4 Afficher le taux de réussite moyen par arrondissement, même ceux où on a pas d'information sur leur lycée

- **But** : Lister les 20 arrondissements avec leur taux de réussite n'est pas indiquer pour certains lycée (donc moyenne = NULL).
- **Résultat attendu** : 20 lignes (Paris 1 → Paris 20).
- **Remarque** : NULL possible (ex. si pas d'information sur les lycées de l'arrondissement).
SQL est honnête : s'il n'y a rien, il affiche "NULL".

Requête 2.5 Lycées parisiens sans IPS

- **But** : Identifier les établissements qui n'ont pas d'information socio-économique.
- **Résultat attendu** : liste des lycées parisiens dont IPS est NULL (avec arrondissement).
Très utile pour éviter de fausser les analyses (si pas d'IPS, on ne compare).

Partie 3 : Comparaisons "intelligentes" au-dessus, en dessous, plus fort que prévu

Requête 3.1 Lycées avec IPS inférieur à la moyenne parisienne

- **But** : Repérer les lycées "moins favorisés" que la moyenne.
- **Moyenne parisienne donnée** : **117,81**.
- **Résultat attendu** : liste des lycées avec IPS < moyenne, triés par IPS croissant.
Ici on prépare la question : "est-ce qu'ils réussissent quand même ?"

Requête 3.2 Lycées qui réussissent mieux que la moyenne parisienne

- **But** : Repérer les établissements qui dépassent la performance moyenne.
- **Résultat attendu** : lycées avec taux réussite > moyenne de Paris.
On filtre les "bons élèves" de Paris.

Requête 3.3 Lycée meilleur que TOUS ceux qui ont un IPS similaire (ALL)

- **But** : Trouver les établissements qui dominent totalement leur "catégorie sociale" (± 2 points d'IPS).
- **Résultat attendu** : lycées dont la réussite est > à tous ceux ayant IPS proche.

Requête 3.4 Lycées défavorisés meilleurs qu'au moins un lycée favorisé (ANY)

- **But** : Identifier des lycées à IPS faible (<100) qui font mieux que certains lycées très favorisés (>120).
- **Résultat attendu** : liste des lycées qui surprennent.

Requête 3.5 Valeur ajoutée supérieure à la moyenne

- **But** : Repérer les lycées qui font mieux que prévu selon leur contexte (VA > moyenne).
- **Résultat attendu** : liste triée par valeur ajoutée.
Ici on cherche ceux qui “créent de la performance” au-delà du contexte.

Requête 3.6 Arrondissement le plus favorisé

- **But** : Trouver l'arrondissement où l'IPS moyen est maximal.
- **Résultat attendu** : nom de l'arrondissement + IPS moyen.

Requête 3.7 Lycées avec taux de mentions > moyenne parisienne

- **But** : Mesurer la réussite “qualitative” (mentions).
- **Résultat attendu** : liste des lycées dépassant la moyenne des mentions.

Requête 3.8 Lycées qui dépassent la moyenne de leur arrondissement

- **But** : Comparer un lycée non pas à tout Paris, mais à son environnement direct.
- **Résultat attendu** : lycées dont la réussite > moyenne de leur arrondissement.

Partie 4 : Opérations ensemblistes UNION / INTERSECTION / EXCEPT

Requête 4.1 UNION : lycées favorisés OU très performants

- **But** : Obtenir la liste des lycées :
 - IPS > 120 **ou** réussite > 95%
- **Résultat attendu** : ensemble combiné sans doublons.

Requête 4.2 INTERSECTION : favorisés ET très performants

- **But** : Trouver les lycées qui cochent les deux cases :
 - IPS élevé
 - réussite exceptionnelle
- **Résultat attendu** : liste des lycées cumulant les deux.

Requête 4.3 EXCEPT simulé : très performants MAIS pas favorisés

- **But** : Repérer les lycées qui réussissent très bien (>95%) tout en n'étant pas favorisés (>120).
- **Résultat attendu** : “performants mais non favorisés”.
On cherche les “vrais méritants” selon ce critère.

Partie 5 : Enrichissement des données → inventer sans mentir

Requête 5.1 Ajouter un lycée fictif crédible

- **But** : Tester la cohérence de la base avec un nouvel établissement réaliste.
- **Résultat attendu** : insertion d'un lycée parisien cohérent.

Requête 5.2 Ajouter son IPS

- **But** : Compléter la table IPS pour que ce lycée soit utilisable dans les analyses.
- **Résultat attendu** : insertion d'une ligne IPS (2018-2019).

Partie 6 : Corrections parce que les bases parfaites n'existent pas

Requête 6.1 Correction d'un IPS erroné

- **But** : Corriger une valeur d'IPS incorrecte.
- **Résultat attendu** : IPS mis à jour.

Requête 6.2 Mise à jour d'une valeur ajoutée

- **But** : Simuler une réévaluation en augmentant la VA sur une année donnée.
- **Résultat attendu** : VA modifiée.

Partie 7 : Indicateurs globaux, Paris en résumé

Requête 7.1 IPS moyen global (2016–2019)

But : Obtenir un point de repère global du niveau socio-économique des lycées parisiens, pour avoir “la moyenne de référence” de Paris.

Résultat attendu : Une seule valeur : l'IPS moyen global calculé sur toutes les observations IPS des lycées dont le Code_commune commence par 75.

Requête 7.2 IPS moyen par année scolaire (2016–2019)

But : Observer l'évolution de l'IPS moyen au fil des années scolaires : est-ce que Paris devient plus favorisé, moins favorisé, ou stable ?

Résultat attendu : Un tableau avec une ligne par année scolaire, et l'IPS moyen correspondant.

Requête 7.3 Taux de réussite moyen par année (2016–2019)

But : Mesurer si les performances au bac progressent, stagnent ou régressent à Paris entre 2016 et 2019.

Résultat attendu : Un tableau avec une ligne par année et le taux de réussite moyen toutes séries confondues.

Requête 7.4 Taux de réussite moyen par arrondissement

But : Comparer les performances scolaires selon le territoire (Paris n'est pas un bloc uniforme : 75101 à 75120).

Résultat attendu : Un tableau listant chaque arrondissement avec le taux de réussite moyen des lycées situés dedans.

Requête 7.5 Écart moyen IPS ↔ réussite

But : Mesurer la cohérence globale entre contexte socio-économique (IPS) et performance (réussite).

Résultat attendu : Une seule valeur : la moyenne de $|IPS - \text{taux de réussite}|$ sur l'ensemble des lycées parisiens. Plus l'écart est faible, plus “réussite et IPS marchent ensemble main dans la main”.

Requête 7.6 Top 10 des lycées selon la meilleure valeur ajoutée (VA maximale)

But : Repérer les lycées qui ont déjà atteint des pics de surperformance : ceux qui “font mieux que prévu” au moins une fois.

Résultat attendu : Un top 10 des lycées parisiens avec leur VA maximale, leur nom et leur arrondissement.

Requêtes 7.7a → 7.7d Répartition des lycées par tranches d'IPS

But : Compter combien d'observations IPS parisiennes se trouvent dans chaque tranche sociale (du très défavorisé au très favorisé).

Résultat attendu : 4 compteurs :

- **7.7a :** nombre d'observations avec $IPS < 80$
- **7.7b :** nombre d'observations avec $80 \leq IPS < 100$
- **7.7c :** nombre d'observations avec $100 \leq IPS < 120$
- **7.7d :** nombre d'observations avec $IPS \geq 120$

Attention : comme l'année n'est pas fixée, un même lycée peut être compté plusieurs fois s'il a plusieurs années IPS.

Requête 7.8 Nombre de lycées publics vs privés à Paris

But : Décrire la structure du système éducatif parisien : combien de lycées publics et combien de lycées privés dans notre périmètre.

Résultat attendu : Un tableau avec 2 lignes (public / privé) et le nombre d'établissements dans chaque catégorie.

Requête 7.9 Nombre total de présents au bac par année

But : Étudier l'évolution de la population candidate au bac à Paris : plus d'élèves, moins d'élèves, stabilité ?

Résultat attendu : Un tableau avec une ligne par année et la somme des présents toutes séries confondues.

Requête 7.10 Score de performance synthétique (classement global)

But : Construire un indicateur maison pour classer les lycées de manière globale en combinant contexte social et performance.

Résultat attendu : Un classement des lycées parisiens selon un score calculé à partir de :

- IPS minimal (IPS_min)
- VA maximale (VA_max)
- réussite maximale (Reussite_max)

Un "score composite" façon recette : un peu d'IPS, un peu de VA, un peu de réussite... et on regarde qui ressort en tête.

Partie 8 Vues : pour travailler plus vite (et souffrir moins)

Requête 8.1 Vue "Performance_globale_lycee_Paris"

But : Créer une vue résumé qui regroupe, pour chaque lycée parisien, les meilleurs indicateurs utiles (IPS min, VA max, réussite max).

Résultat attendu : Une vue avec une ligne par lycée : nom, arrondissement, IPS_min, VA_max, Reussite_max.

Requête 8.2 Vue "Evolution_2016_2019_Paris"

But : Suivre l'évolution par lycée et par année : IPS de l'année + réussite de l'année correspondante.

Résultat attendu : Une vue "long format" avec une ligne par lycée et par année (2016–2019).

Partie 9 : Requêtes analytiques avec CTE, le SQL "intelligent"

Requête 9.1 Classement par écart IPS ↔ réussite

But : Repérer les lycées les plus “atypiques” : ceux dont la réussite est très éloignée de ce que l'IPS laisserait imaginer.

Résultat attendu : Un tableau trié du plus grand écart au plus petit, avec l'écart calculé pour chaque lycée.

Requête 9.2 Lycées “meilleurs que prévu” ($VA > 0$)

But : Identifier les lycées qui surperforment réellement, c'est-à-dire ceux dont la valeur ajoutée maximale est positive.

Résultat attendu : La liste des lycées parisiens ayant $VA_{max} > 0$, triée par VA décroissante.

Partie 10 L'analyse territoriale finale

Requête 10.1 Arrondissements où l'écart IPS ↔ réussite est le plus anormal

But : Trouver les arrondissements où réussite et IPS ne “collent” pas : zones où la réussite est trop haute ou trop basse par rapport au contexte social.

Résultat attendu : Un classement des arrondissements du plus anormal au plus cohérent.

Requête 10.2 Résultats bac présents mais IPS absent

But : Repérer les établissements pour lesquels on a des résultats au bac, mais aucune donnée IPS disponible.

Résultat attendu : La liste des lycées parisiens concernés (UAI, nom, arrondissement).

Partie 11 : Optimisation, quand la base commence à transpirer

Requête 11.1 EXPLAIN : comprendre le plan d'exécution

But : Observer comment MySQL exécute une requête lourde (multi-jointures + group by) afin d'identifier les étapes coûteuses.

Résultat attendu : Le tableau EXPLAIN indiquant le type de jointure, les index utilisés (ou pas), et le volume de lignes parcourues.

Requête 11.2 Index : accélérer les jointures et filtres

But : Réduire les temps de réponse en indexant les colonnes les plus sollicitées : UAI, Code_commune, Année.

Résultat attendu : Des index permettant d'accélérer les jointures et les filtres Paris/année.

Requête 11.3 Réécriture plus efficace (remplacer une sous-requête corrélée)

But : Produire le même résultat que la requête 3.8 mais en évitant de recalculer la moyenne arrondissement à chaque ligne.

Résultat attendu : La liste des lycées “au-dessus de la moyenne de leur arrondissement”, mais avec une exécution plus efficace.

Même résultat, mais MySQL souffre moins (et le PC aussi).

Visualisation et analyse des performances (Bonus)

On commence par établir une **connexion entre Python et la base MySQL**, permettant d'exécuter directement des requêtes SQL depuis Python et de récupérer les résultats sous forme de tableaux exploitables.

Les données sont ensuite traitées et visualisées à l'aide de bibliothèques adaptées :

- manipulation des données,
- production de graphiques statistiques,
- génération automatique d'un rapport visuel au format PDF.

Un **style graphique homogène** est appliqué à l'ensemble des figures afin de garantir cohérence et lisibilité.

Partie 1 : Évolution du contexte socio-économique à Paris

Visualisation 1 : Évolution de l'IPS moyen des lycées parisiens (2016–2019)

Objectif

Analyser l'évolution du niveau socio-économique moyen des lycées parisiens sur la période étudiée.

Principe

Pour chaque année scolaire, on calcule l'IPS moyen de l'ensemble des lycées situés à Paris.

Intérêt analytique

Cette visualisation permet de déterminer si le contexte social global :

- progresse,
- se dégrade,
- ou reste stable dans le temps.

Elle constitue un point de référence indispensable pour interpréter les performances scolaires observées ensuite.

Partie 2 : Évolution des performances au baccalauréat

Visualisation 2 : Évolution du taux de réussite moyen au bac

Objectif

Étudier la dynamique des performances scolaires globales des lycées parisiens entre 2016 et 2019.

Principe

Pour chaque année, on calcule le taux de réussite moyen au baccalauréat pour l'ensemble des établissements parisiens.

Intérêt analytique

Cette visualisation permet de comparer l'évolution de la réussite avec celle de l'IPS moyen, afin d'identifier une éventuelle corrélation entre contexte socio-économique et performance globale.

Partie 3 : Lien entre IPS et réussite scolaire

Visualisation 3 : Distribution du taux de réussite selon les tranches d'IPS

Objectif

Comparer les performances scolaires des lycées selon leur niveau socio-économique.

Principe

Les lycées sont regroupés en **quatre tranches d'IPS** :

- $IPS < 80$
- $80 \leq IPS < 100$
- $100 \leq IPS < 120$
- $IPS \geq 120$

Pour chaque tranche, on observe la distribution des taux de réussite.

Intérêt analytique

Cette visualisation permet de vérifier si un IPS élevé garantit systématiquement de meilleures performances, tout en mettant en évidence la dispersion des résultats à l'intérieur d'un même groupe social.

Visualisation 4 : Répartition des lycées par tranche d'IPS

Objectif

Décrire la structure socio-économique globale des lycées parisiens.

Principe

Un diagramme circulaire représente la part des lycées appartenant à chaque tranche d'IPS.

Intérêt analytique

Cette visualisation contextualise les analyses précédentes en montrant le poids relatif de chaque catégorie socio-économique dans le paysage éducatif parisien.

Partie 4 : Analyse territoriale : disparités entre arrondissements

Visualisation 5 : IPS moyen par arrondissement parisien

Objectif

Comparer le niveau socio-économique moyen des lycées selon leur arrondissement.

Principe

Pour chaque arrondissement parisien, on calcule l'IPS moyen des lycées qui y sont implantés.

Intérêt analytique

Cette analyse met en évidence les inégalités territoriales à l'intérieur de Paris et montre que la ville ne constitue pas un ensemble homogène du point de vue socio-économique.

Partie 5 : Identifier les lycées surperformants

Visualisation 6 : Top 10 des lycées parisiens selon la valeur ajoutée

Objectif

Identifier les établissements qui obtiennent des résultats supérieurs à ceux attendus compte tenu de leur contexte.

Principe

Pour chaque lycée, on retient la **valeur ajoutée maximale** observée sur la période, puis on classe les 10 établissements

les plus performants selon cet indicateur.

Intérêt analytique

Cette visualisation permet de repérer les lycées qui “créent” de la performance, indépendamment du niveau socio-économique de leurs élèves.

Partie 6 : Croisement IPS et réussite à l'échelle territoriale

Visualisation 7 : Carte thermique IPS moyen et réussite moyenne par arrondissement

Objectif

Croiser simultanément le niveau socio-économique et les performances scolaires par arrondissement.

Principe

Pour chaque arrondissement parisien, on représente :

- l'IPS moyen,
- le taux de réussite moyen au bac, sous forme de **carte thermique**.

Intérêt analytique

Cette visualisation permet d'identifier rapidement :

- les arrondissements favorisés et performants,
- ceux en difficulté,
- et surtout les situations atypiques où IPS et réussite ne suivent pas la même logique.

Maintenant, c'est à vous de commencer à concevoir votre base de données. Identifiez les tables, les colonnes et les types de données pour chaque table. Ensuite, choisissez un SGBD et débutez la création de votre base de données. L'objectif est d'analyser chaque résultat afin de répondre à la problématique. Vous trouverez ci-dessous les codes, une vidéo de correction de ce mini projet ainsi que l'analyse qui répondra à la problématique. Bon courage !

[Lien vers le projet](#)