

## RAPPORT DE PROJET PORTANT SUR LES INEGALITES EN MATIERE D'EDUCATION

Réalisé par :

**Amadou DIALLO**

**Ibrahima KABA**

**Abdellatif MESSAOUDI**

**Moulaye Abdoule Hady HAIDARA**

Encadrante :

**Prof. Mehdia AJANA**

Date de Soumission :

**20/12/2025**

Année Universitaire : 2025/2026

## Introduction

Dans un contexte marqué par des inégalités persistantes dans les systèmes éducatifs, l'analyse des données constitue un outil essentiel pour mieux comprendre les facteurs influençant la réussite scolaire. Dans le cadre du module *Compétences Numériques*, ce projet vise à appliquer les bases de la programmation Python à l'analyse et à la visualisation de données réelles.

Le dataset choisi porte sur les **inégalités éducatives** et regroupe plusieurs indicateurs relatifs aux établissements scolaires, tels que le financement par élève, les performances académiques, le taux de décrochage, l'accès à Internet et le type d'école.

Ce choix permet d'explorer de manière objective les relations entre ressources, contexte socio-économique et résultats éducatifs.

Les objectifs du projet sont :

- ✓ manipuler et analyser un dataset réel à l'aide de **Pandas** ;
- ✓ produire des visualisations pertinentes avec **Matplotlib** ;
- ✓ interpréter les résultats de manière critique ;
- ✓ travailler efficacement en équipe en utilisant **Trello** comme outil de gestion de projet.

## Méthodologie

### Présentation du dataset

Le dataset utilisé provient de la plateforme **Kaggle** et est intitulé *Education Inequality Data*. Ce dataset ne contenait pas de valeurs manquantes , des valeurs manquantes ont été ajoutées et le dataset a été renommé « `projet_python` » afin de répondre aux exigences de l'étude.

Il est accessible via le lien suivant :

☞ <https://www.kaggle.com/datasets/shamimhasan8/education-inequality-data>

Ce fichier, fourni au format CSV, contient 12 colonnes et 1000 lignes correspondant à des établissements scolaires. Les principales variables incluent :

- ✓ le type d'école (Public, Private, Charter),
- ✓ le financement par élève,
- ✓ le score moyen aux tests,

- ✓ le pourcentage d'élèves à faible revenu,
- ✓ le taux de décrochage scolaire,
- ✓ le niveau scolaire,
- ✓ le taux d'accès à Internet.

## Analyse et prétraitement des données avec Pandas

La première étape a consisté à inspecter la structure du dataset afin d'identifier les types de variables et la présence de valeurs manquantes.

Le dataset initial (*projet\_python.csv*) présentait plusieurs anomalies, notamment des incohérences de types, des valeurs manquantes et des doublons.

Un nettoyage rigoureux a été effectué afin d'obtenir un fichier final fiable (*projet\_python\_corrigé.csv*). Les colonnes catégorielles ont été filtrées pour éliminer les valeurs incohérentes, tandis que les colonnes numériques ont été corrigées en remplaçant les valeurs invalides par des valeurs représentatives. Les doublons ont également été supprimés afin d'éviter toute surreprésentation d'un établissement.

Des statistiques descriptives de base (moyenne, minimum, maximum) ont ensuite été calculées pour les variables numériques. Enfin, des opérations de filtrage et de regroupement ont permis d'analyser certains sous-ensembles du dataset, notamment selon le type d'école ou le niveau scolaire.

## Organisation du code Python

Le code Python a été structuré de manière claire et modulaire, en adoptant une approche orientée objet.

Les principales fonctionnalités incluent :

- ✓ le chargement du dataset,
- ✓ l'analyse statistique des données,
- ✓ la génération des visualisations à l'aide de **Matplotlib**.

Cette organisation améliore la lisibilité du code et facilite sa maintenance.

## Visualisation des données et résultats

### Répartition des types d'écoles

Un **diagramme circulaire** a été utilisé pour représenter la répartition des établissements selon leur statut.

Les résultats montrent une distribution équilibrée :

- ✓ **35,6 %** d'écoles privées,
- ✓ **32,5 %** d'écoles publiques,
- ✓ **31,9 %** d'écoles à charte.

Cette homogénéité garantit que les analyses comparatives ultérieures ne sont pas biaisées par une surreprésentation d'un type d'établissement.

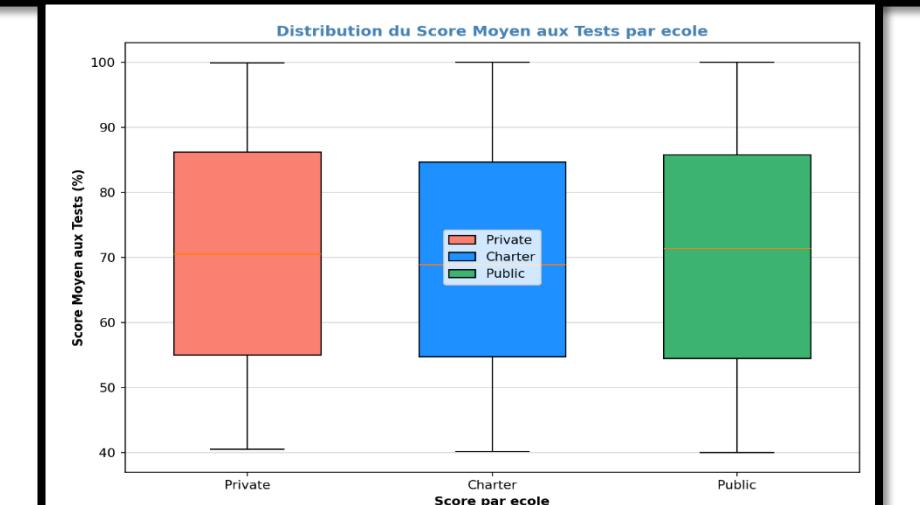
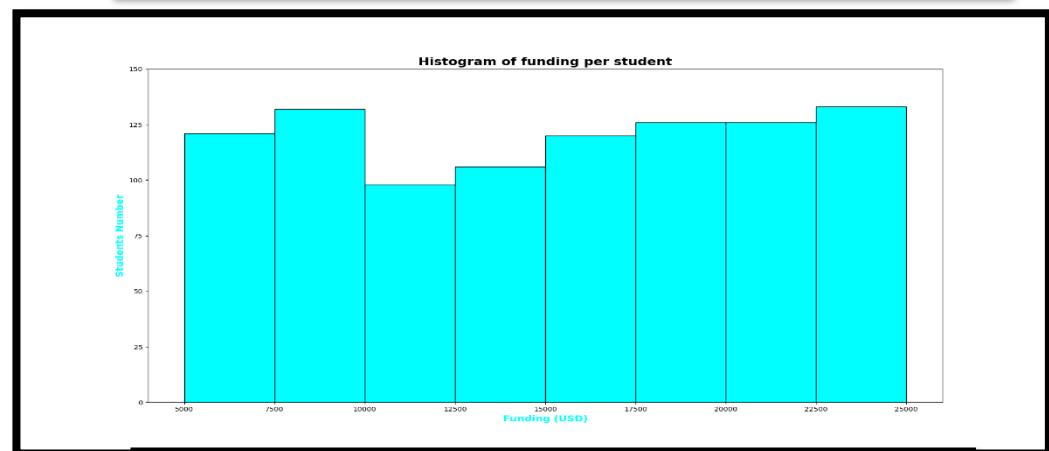
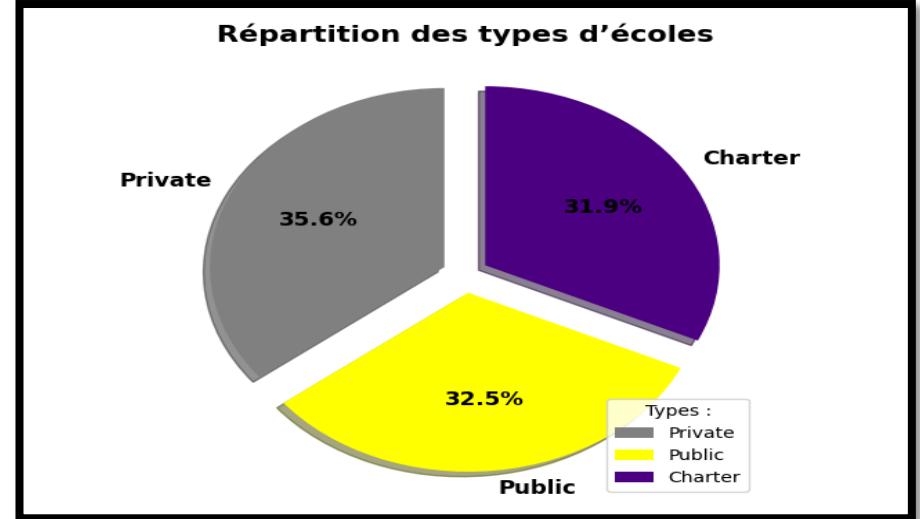
### Financement par élève

L'**histogramme du financement par élève** met en évidence une forte dispersion des budgets, allant de **5 000 USD à 25 000 USD**.

Deux pics de fréquence sont observés, indiquant l'absence d'un modèle de financement uniforme et révélant des écarts budgétaires importants entre établissements.

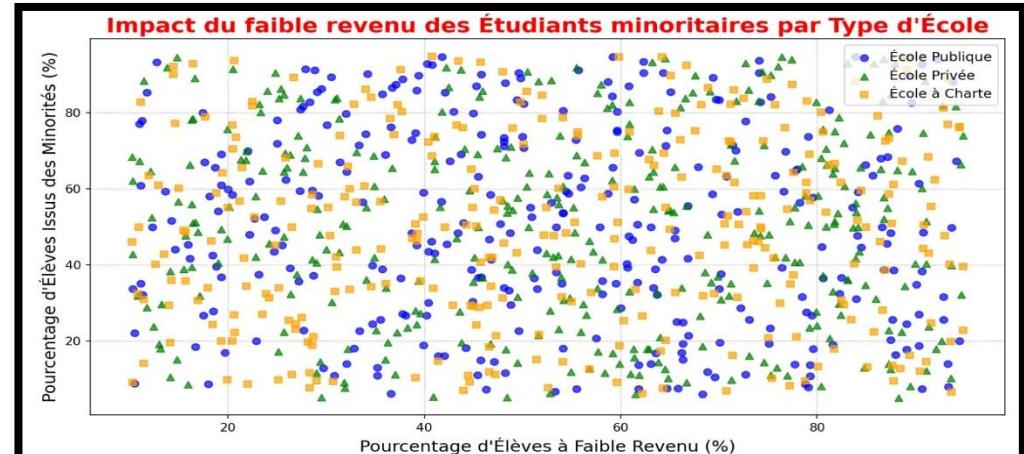
### Score moyen par type d'école

Le **diagramme en boîte** comparant les scores moyens selon le type d'école montre des médianes très proches, autour de **70 %** pour les trois catégories. Aucune supériorité académique nette n'est observée, suggérant que le statut administratif de l'établissement n'est pas un facteur déterminant de la performance dans ce dataset.



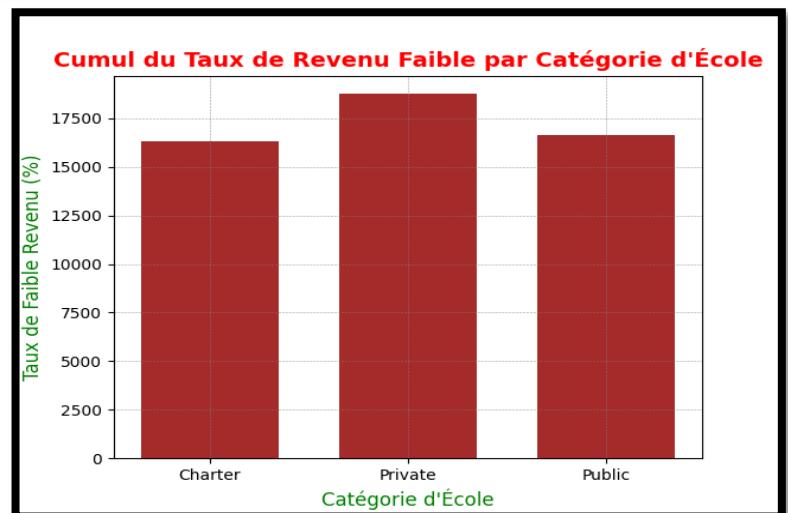
## Impact du faible revenu sur la performance

Le **nuage de points** croisant le pourcentage d'élèves à faible revenu et le score moyen aux tests révèle une corrélation très faible. Certaines écoles défavorisées obtiennent d'excellents résultats, tandis que d'autres établissements plus favorisés affichent des performances modestes. Cela indique que le niveau socio-économique, à lui seul, n'explique pas la réussite scolaire.



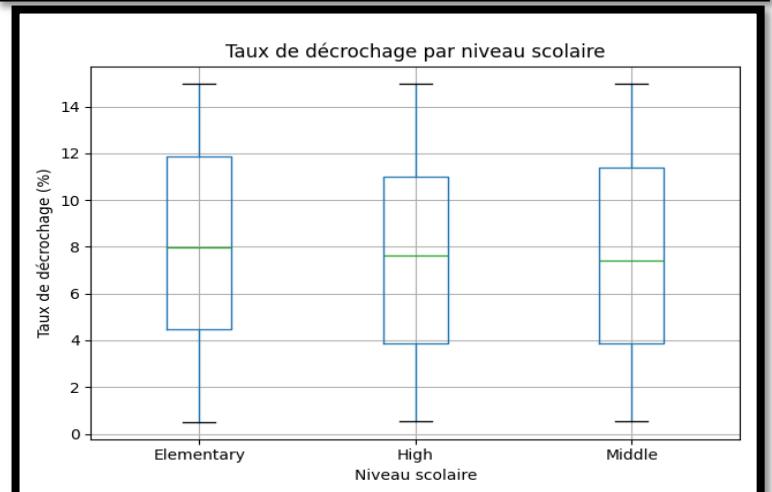
## Financement moyen par type d'école

Le **diagramme en barres** du financement moyen par type d'école montre que les établissements publics bénéficient légèrement de budgets plus élevés que les écoles privées et à charte. Cependant, en croisant ces résultats avec les performances académiques, il apparaît que des financements plus importants ne garantissent pas de meilleurs scores.



## Décrochage scolaire

Le **box plot du taux de décrochage par niveau scolaire** indique une médiane stable autour de **7,5 %**, avec une variabilité plus importante au niveau élémentaire. Cette visualisation montre que le décrochage n'est pas un problème exclusif au lycée. La tendance identifiée est que les difficultés menant à l'abandon scolaire commencent dès le plus jeune âge de manière significative dans certains établissements.



## Difficultés rencontrées

La principale difficulté a concerné le **nettoyage des données**, en raison de valeurs manquantes, d'incohérences de types et de doublons. Un traitement rigoureux a permis d'assurer la fiabilité du dataset avant la phase de visualisation.

Sur le plan organisationnel, la **gestion des tâches via Trello** a nécessité des ajustements initiaux. Une meilleure répartition des rôles et l'utilisation d'étiquettes ont permis d'améliorer la coordination et le respect des délais.

## Travail collaboratif et gestion du projet

Le projet a été mené de manière collaborative à l'aide de **Trello**, permettant une répartition claire des tâches et un suivi efficace de l'avancement.

Chaque membre du groupe a contribué activement aux différentes étapes du projet. Des captures d'écran du tableau Trello sont jointes afin d'attester du travail collectif.

The screenshot shows a Trello board titled "3. Visualization Tasks". A checklist under the title includes:

- Create some visualizations using Matplotlib (at least 3): For example:
  - A bar chart to compare categories.
  - A line plot to show trends and a scatter plot to show correlations.
- Add basic customizations for the plots like titles, axis labels, and legends.

Below the checklist, there's a section titled "Visualisation des tâches" with a comment from MOULAYE ABDOULE HADY HAIDARA:

Bon travail @abdelatifmessaoudi2 🌟 je viens de déposer ma partie, la dernière personne qui va déposer devra marquer la tâche comme achevée.  
Pour rappel le délai, c'est ce soir, avant mi-nuit.

Comments from other users:

- MOULAYE ABDOULE HADY HAIDARA (14 déc. 2025, 12:14): Bon travail @abdelatifmessaoudi2 🌟 je viens de déposer ma partie, la dernière personne qui va déposer devra marquer la tâche comme achevée.  
Pour rappel le délai, c'est ce soir, avant mi-nuit.
- ABDELLATIF MESSAOUDI (14 déc. 2025, 08:20): @abdule\_hady 🎉

The screenshot shows a Trello board titled "Group and Filter data". The task description is:

○ **Group and Filter data:** Perform simple grouping or filtering to focus on specific subsets of the dataset (e.g., filter students who scored above 15).

Below the task, there's a section titled "Regroupement et filtre des données" with a comment from ABDELLATIF MESSAOUDI:

I've finished. Please take a look.

Comments from other users:

- ABDELLATIF MESSAOUDI (10 déc. 2025, 22:38): I've finished. Please take a look.
- AMADOU DIALLO (10 déc. 2025, 22:33): @abdelatifmessaoudi2 Je viens de vous envoyez la dataset modifiee
- MOULAYE ABDOULE HADY HAIDARA (10 déc. 2025, 10:49): Vous devez finir le travail aujourd'hui pour rappel

Terminé

- Summarize data: Calculate basic statistics for numerical columns (e.g., average, maximum, minimum...).

Résumé des données

Calculer les statistiques de base pour les colonnes numériques (par exemple,moyenne,maximum,minimum...).

Summarize data: Calculate basic statistics for numerical columns (e.g., average, maximum, minimum...).

Vous trouverez le DataSet en pièce jointe @ibrahimakaba4

Pour toute remarque,n'hésitez pas à laisser un commentaire !

Pièces jointes Ajouter

Terminé

- Handle missing data: Explain and apply strategies such as removing or replacing missing values, dropping rows with missing values...

Gestion des données manquantes

Expliquer et appliquer des stratégies telles que la suppression ou le remplacement des valeurs manquantes, la suppression des lignes contenant des valeurs manquantes...

Handle missing data: Explain and apply strategies such as removing or replacing missing values, dropping rows with missing values...

Vous trouverez le DataSet en pièce jointe @am\_diallo

Pour toute remarque,n'hésitez pas à laisser un commentaire !

Pièces jointes Ajouter

Terminé

4. OOP Implementation:

- For e.g. create a Python class with the following simple methods:
  - load\_data() for importing the dataset.

+ Ajouter Checklist Pièce jointe

Membres Étiquettes IK Urgentement ! +

Dates 14 déc. - 15 déc., 23:59 Complété

Description Ton rôle est d'utiliser la Programmation orientée Objet pour créer la fonction "load\_data()" pour l'importation du jeu de données :

OOP Implementation

Egrégore

A faire: Commencez à utiliser Trello + Ajouter une carte

En cours: Rapport (15 déc. - 19 déc.) 1 AD

Python Code (18 déc. - 19 déc.) 1 IK

Terminé: plot\_data() for creating visualizations. POO - plot\_data() (14 déc. - 15 déc.) 4 AD

OOP Implementation: For e.g. create a Python class with the following simple methods:
 

- load\_data() for importing the dataset.

POO - load\_data() (14 déc. - 15 déc.) 3 IK

analyze\_data() for generating basic statistics. + Ajouter une carte

## Conclusion

Les résultats de ce projet montrent que les inégalités éducatives observées dans le dataset ne s'expliquent pas uniquement par des facteurs financiers ou technologiques.

La performance académique et le décrochage scolaire semblent davantage influencés par des facteurs humains, pédagogiques et organisationnels que par le niveau de financement ou l'accès aux infrastructures numériques.

Ce projet a permis de consolider nos compétences en **Pandas**, **Matplotlib** et en gestion de projet collaboratif. Il constitue une base solide pour des analyses plus avancées et met en évidence l'importance d'une approche critique dans l'interprétation des données.