



# Analyse des Inégalités de Décrochage Scolaire en France

Régression logistique



# Description du Projet

## Source des données

Les données utilisées dans cette étude proviennent du **Ministère de l'Éducation nationale français** et sont accessibles via la plateforme officielle des données publiques :

 [data.education.gouv.fr](https://data.education.gouv.fr)

Le jeu de données porte sur les effectifs scolaires et inclut des informations détaillées sur :

- **Les effectifs des collèges** en France (publics et privés).
- **Le taux de décrochage scolaire** par établissement.
- **Le sexe des élèves** pour analyser les différences entre garçons et filles.
- **La taille des collèges** pour évaluer l'impact sur le décrochage scolaire.



# Méthodologie



## 1 - Description des variables et nettoyage

- Vérification de la structure du dataset
- Identification des valeurs manquantes
- Vérification des types de variables



## 2 - Méthodes statistiques utilisées



### Analyse exploratoire des données (EDA) :

- Histogrammes et boxplots pour analyser la distribution des effectifs et des taux de décrochage.
- Heatmaps pour identifier les **corrélations** entre variables



## 3. Tests statistiques :

- ANOVA pour comparer les taux de décrochage entre les collèges
- Comparaison



## Description des variables et nettoyage



541 Colonnes  
33K Lignes



4.6M ~ 25%  
Valeurs Manquantes



### Variables: Qualitatives & Quantitatives

Qualitatives

465

Quantitatives

76



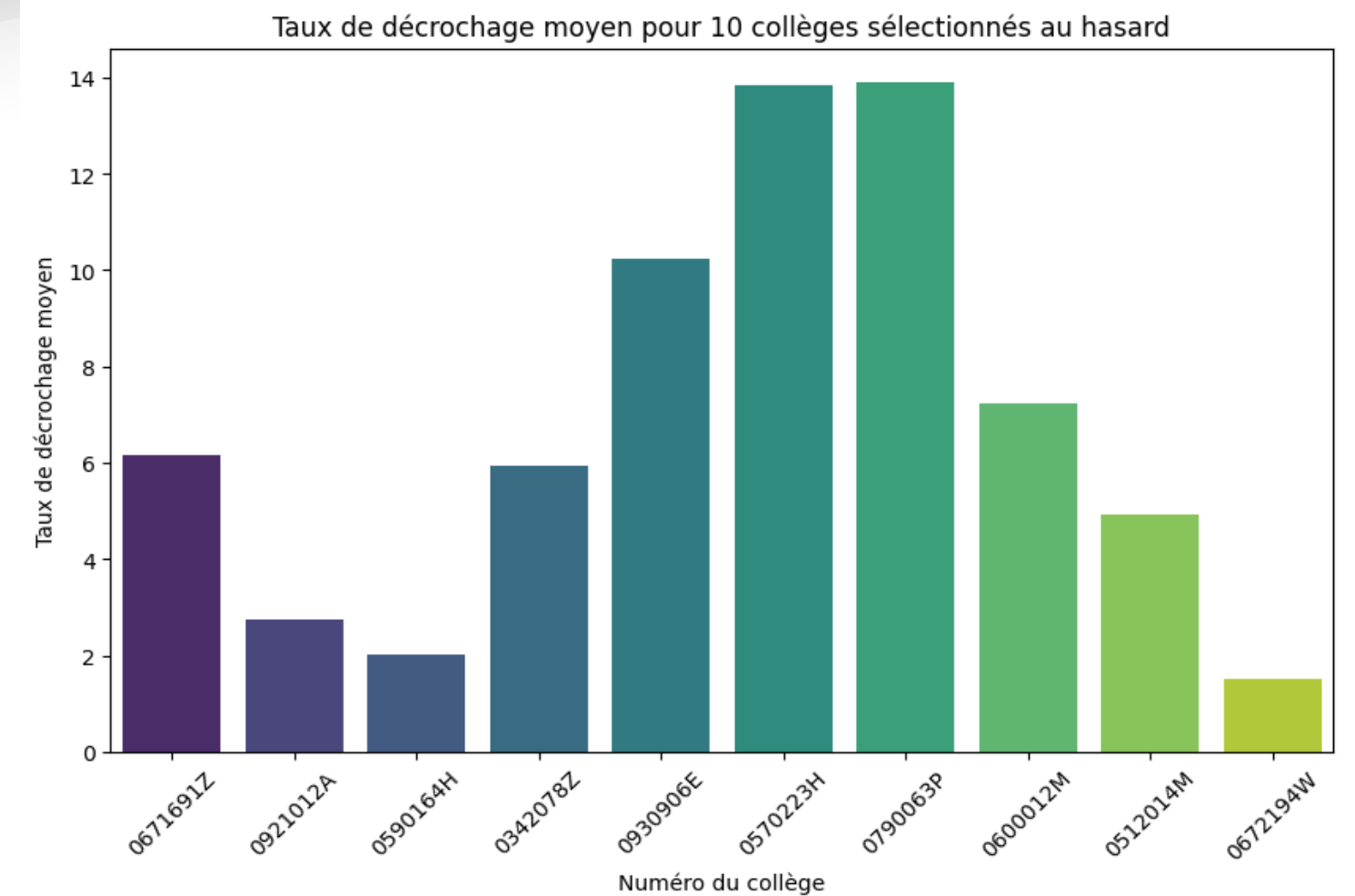
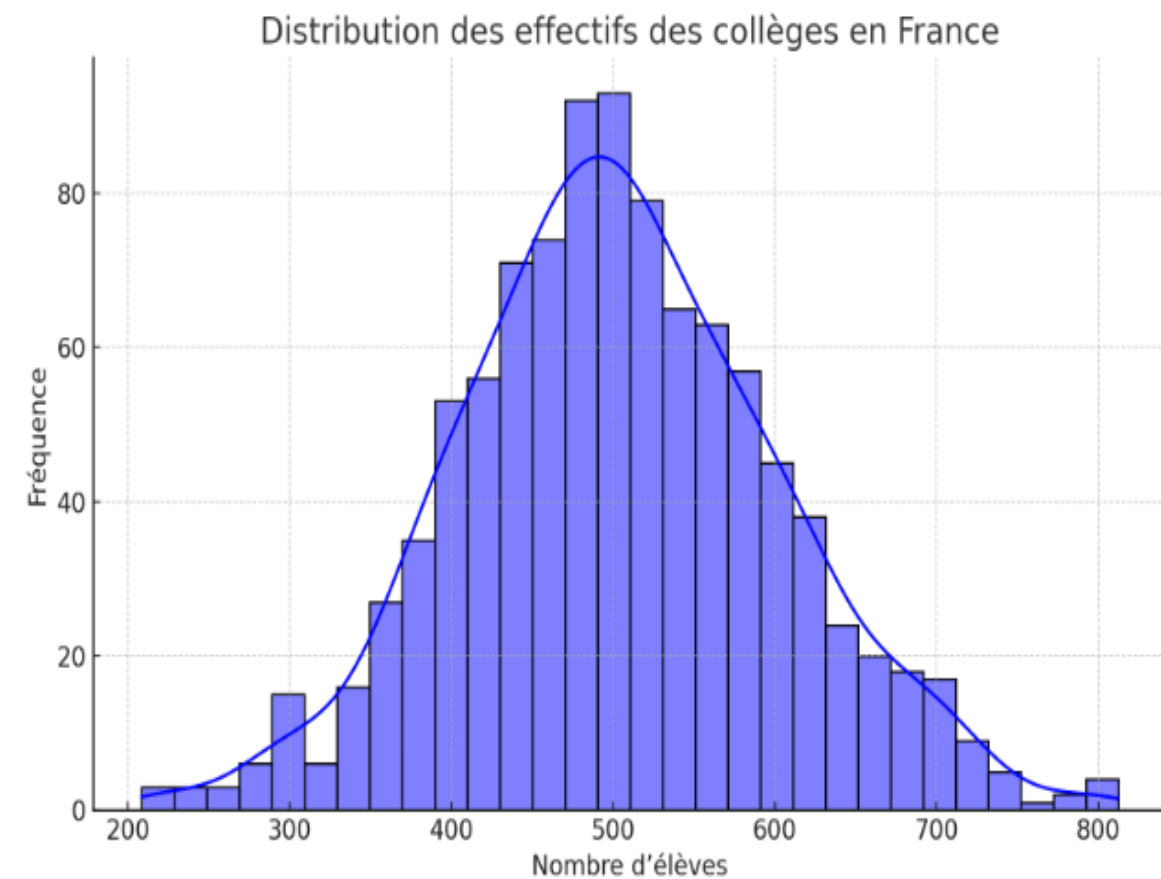
# Traitement des données

Étape	Description
1 Filtrage & Sélection	Sélection des variables essentielles pour l'étude
2 One Hot Encoding	Transformation des variables qualitatives en numériques
3 Gestion des Données Manquantes	Suppression ou imputation des valeurs manquantes

# Méthodes statistiques

## Variables Quantitatives

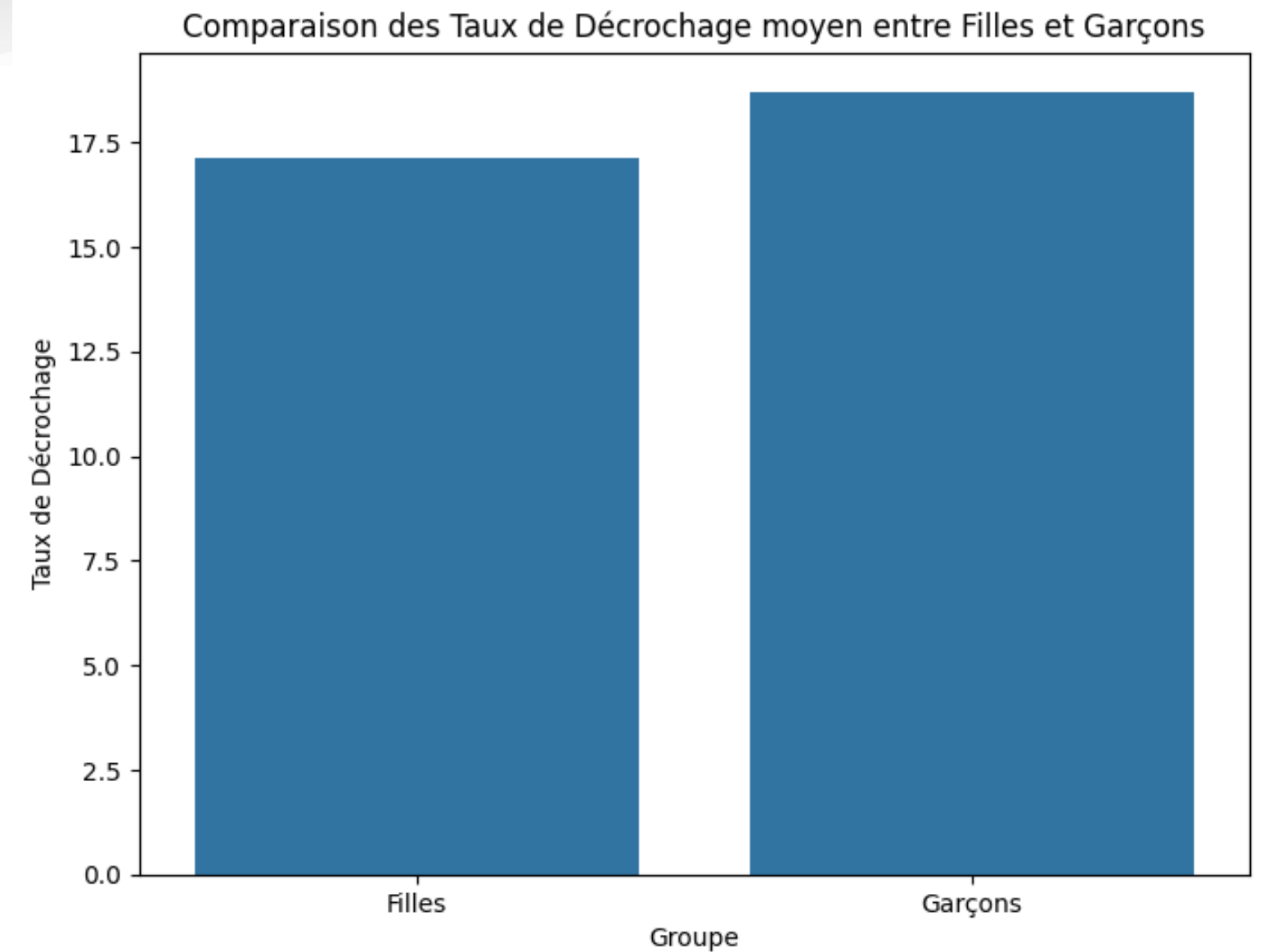
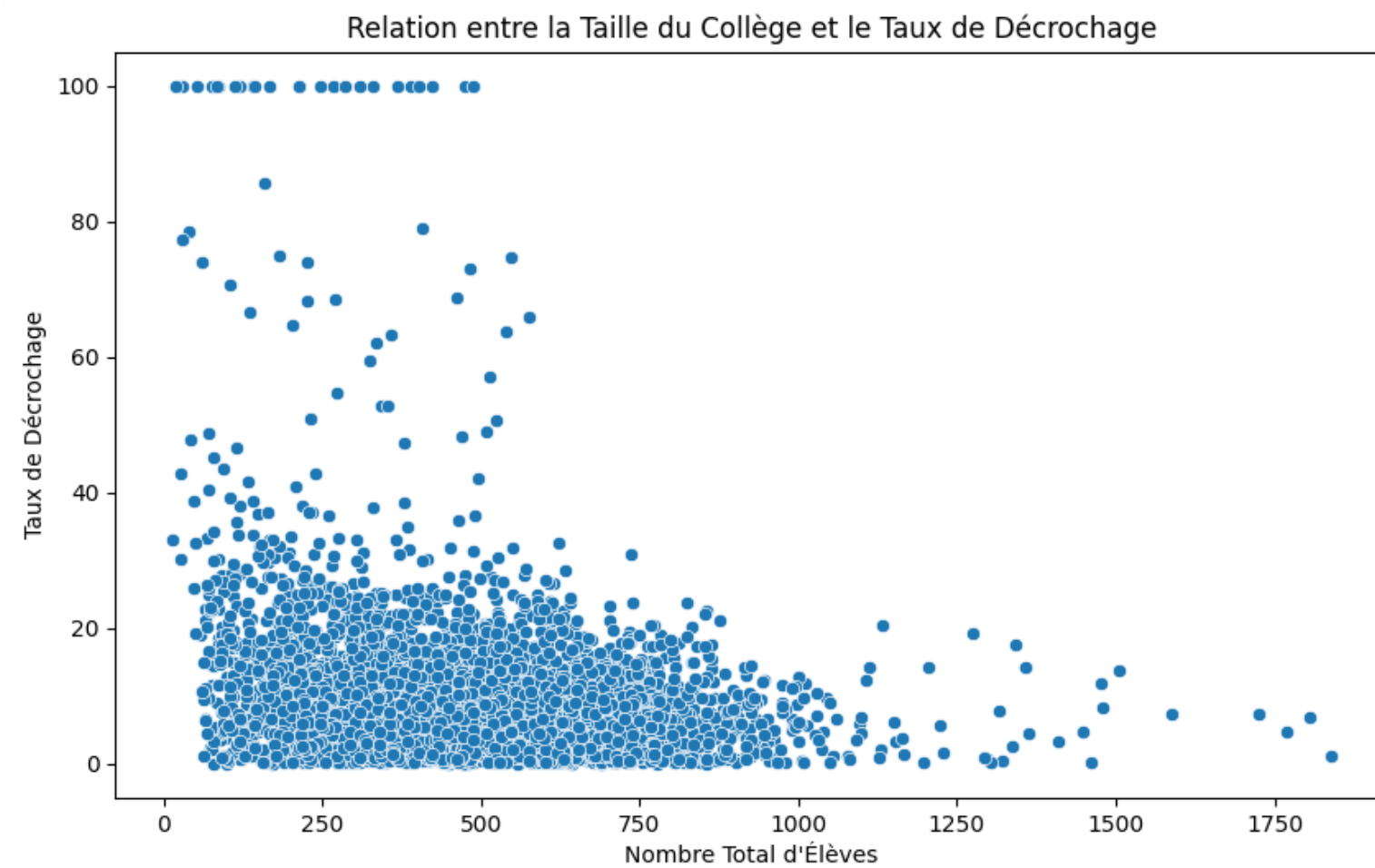
- Histogrammes et boxplots pour analyser la distribution des effectifs
- Taux de décrochage moyen



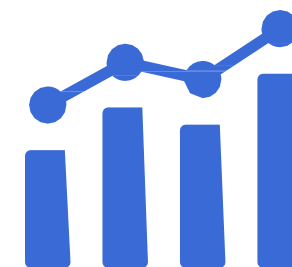
# Tests statistiques

## Variables Quantitatives

- ANOVA pour comparer les taux de décrochage entre les collèges
- Comparaison des Taux de Décrochage moyen entre Filles et Garçons



- 📌 **Hypothèse :** Les collèges de grande taille ont un taux de décrochage plus faible.
- 📌 **Hypothèse :** Les collèges privés ont un taux de décrochage plus faible que les collèges publics.
- 📌 **Hypothèse :** Certaines régions sont plus touchées par le décrochage scolaire que d'autres.
- 📌 **Hypothèse :** Le taux de décrochage scolaire varie selon le sexe des élèves.



## Analyse et résultats

Facteur	Observation	Conclusion
Taille du collège	Corrélation négative avec le taux de décrochage	Les grands collèges ont un taux de décrochage plus faible
Public vs Privé	Les collèges privés ont de meilleurs résultats	Meilleur encadrement et suivi pédagogique
Régions	Des écarts importants selon les territoires	Nécessité de politiques d'accompagnement régionalisées
Genre	Les garçons décrochent plus que les filles	Facteurs socioculturels et pédagogiques en jeu



# Conclusion

L'étude du **décrochage scolaire en France** a mis en évidence plusieurs facteurs déterminants influençant la réussite éducative des élèves. À travers une analyse approfondie des données, nous avons observé que **la taille de l'établissement, le statut public/privé, la localisation géographique et le genre** jouent un rôle clé dans le taux de décrochage.