



UNIVERSITÉ  
CHEIKH ANTA DIOP  
DE DAKAR

FACULTÉ DES  
SCIENCES ET  
TECHNIQUES  
**FST**



---

## MÉMOIRE DE RECHERCHE DE II CYCLE

**Analyse de l'impact des récentes réformes du  
baccalauréat au Sénégal sur le taux de réussite au bac  
et le parcours universitaire des bacheliers**

*Présenté et soutenu par :*

**ABDOUROIHIM Attoumane**

Pour l'obtention du double diplôme de **Master 2**,

**D'Ingénierie Mathématique et Numérique**

Sous la direction de : **Pr Bamba GUEYE & Pr Mountaga LAM**

*Présenté le 5 juillet 2025, devant le jury composé de :*

Dr Souley KANE

Maître de conférence

Pr Papa NGOM

Professeur Titulaire (President)

Dr Aissatou SOW

Maître de conférence

---

# DÉDICACE

**Je dédie ce travail à ma famille :**

À mes parents, M. **ATTOUMANE Abdallah** et Mme **ZALIHATAH Abdallah**, qui ont été un soutien inconditionnel tout au long de mon parcours académique, tant sur le plan financier que moral. Merci pour vos conseils dans mes moments de doute et pour votre présence constante dans tous mes choix et dans toute ma vie.

À ma femme, Mme **ADELIA Soiffaouiddine**, dont la présence à mes côtés a été essentielle pour affronter cette vie ensemble. Sans elle, ce travail n'aurait tout simplement pas existé.

À mes deux grandes sœurs, Mme **ITFAOU Attoumane** et Mme **ROIHMA Attoumane**, pour leur affection et leur soutien inconditionnel.

À ma belle-famille, en particulier Dr **SOIFFAOUIDDINE Sidi**, le père de ma femme, pour son aide précieuse, tant financière que morale, et pour la confiance qu'il m'a accordée en acceptant de me donner sa fille la plus précieuse.

---

# REMERCIEMENTS

**Alhamdulillah Rabbi al-‘Alamin.**

Ya **ALLAH**, merci de m’avoir permis de réaliser ce travail, de m’avoir donné la force et le courage d’affronter toutes les difficultés que j’ai pu rencontrer.

Je tiens à témoigner toute ma reconnaissance et mes remerciements les plus sincères au **Professeur Mountaga LAM**, mon directeur à la DES et encadrant professionnel de ce travail.

Merci, tout d’abord, de m’avoir permis d’obtenir mon tout premier travail professionnel et salaire de ma vie en m’acceptant pour travailler sous votre direction à la DES. Merci pour votre patience, vos encouragements, votre disponibilité, votre confiance et vos nombreux conseils précieux.  
*C’est un immense honneur d’avoir pu travailler avec vous.*

Je remercie également le **Professeur Bamba GUEYE**, directeur de l’Office du Bac et encadrant académique de ce travail, pour ce sujet particulièrement stimulant, pour sa disponibilité et ses conseils précieux tout au long de ce travail.

Je tiens à remercier tout particulièrement **Dr Souley KANE**, mon directeur de formation, qui m’a permis d’obtenir ce stage et qui a profondément changé ma vie professionnelle en m’ouvrant les portes d’un monde d’opportunités que je n’aurais jamais imaginé intégrer.

Je n’oublierai pas mes collègues à la DES, en particulier Madame **MBAYE** et Madame **PAYE** qui, par leur hospitalité et leur gentillesse, ont été comme des mères pour moi. Merci à vous et ainsi qu’à tous mes autres collègues de travail pour votre aide précieuse tout au long de mon stage.

Je remercie également les membres du jury pour l’attention accordée à ce travail, ainsi que pour leurs remarques pertinentes et enrichissantes.

Je remercie tous les camarades que j’ai rencontrés au fil de mon parcours pour leur fraternité et leur présence.

Ma profonde gratitude va à l’endroit de tous les enseignants qui m’ont transmis leur savoir avec passion. Grâce à vous, je suis devenu la personne que je suis aujourd’hui.

*Merci infiniment à tous, c’était un honneur.*

---

## **RÉSUMÉ**

Ce mémoire présente une analyse approfondie de l'impact des réformes du baccalauréat au Sénégal sur le taux de réussite et l'insertion universitaire. Nous avons utilisé des méthodes statistiques avancées pour évaluer ces effets.

## **ABSTRACT**

This thesis presents a comprehensive analysis of the impact of the baccalaureate reforms in Senegal on the success rate and university integration. We used advanced statistical methods to evaluate these effects.

# Table des matières

<b>Introduction Générale</b>	<b>9</b>
Contexte et justification . . . . .	9
Problématique et Objectifs de l'étude . . . . .	9
Méthodologie et Organisation du document . . . . .	10
<b>1 État de l'art</b>	<b>12</b>
1.1 Introduction . . . . .	12
1.2 Historique et rôle du baccalauréat . . . . .	12
1.3 Organisation du baccalauréat : structure et acteurs . . . . .	13
1.4 Réformes du baccalauréat au Sénégal . . . . .	13
1.4.1 Réformes de 1995 (décret n° 95-947) . . . . .	13
1.4.2 Réformes de 2000 (décret n° 2000-585) . . . . .	14
1.4.3 Réformes de 2013 (décret n° 2013-913) . . . . .	15
1.4.4 Réformes de 2019(décret n° 2019-645) . . . . .	15
1.5 Travaux antérieures . . . . .	15
1.6 Approche méthodologique et innovante de l'étude . . . . .	16
1.7 Conclusion . . . . .	16
<b>2 Données et outils utilisés</b>	<b>18</b>
2.1 Introduction . . . . .	18
2.2 Sources et Description des données . . . . .	18
2.2.1 Sources de données . . . . .	18
2.2.2 Description des données . . . . .	19
2.3 Outils et Technologies utilisés . . . . .	20
2.3.1 Python et ses bibliothèques . . . . .	20
2.3.2 Power BI . . . . .	21
2.4 Conclusion . . . . .	22
<b>3 Nettoyage et préparation des données</b>	<b>23</b>
3.1 Introduction . . . . .	23
3.2 Préparation de données des résultats du baccalauréat . . . . .	23
3.2.1 Doublons dans la clé primaire numero_table . . . . .	24

3.2.2	Les valeurs manquantes dans les colonnes . . . . .	25
3.2.3	Correction du type de la colonne <code>moy_finale</code> . . . . .	27
3.2.4	Création des colonnes <code>admis</code> et <code>session</code> . . . . .	28
3.3	Fusion des données des Inscrits et des Résultats de l'UCAD . . . . .	28
3.3.1	Données des inscriptions à l'UCAD . . . . .	28
3.3.2	Données des résultats de l'UCAD . . . . .	29
3.3.3	Fusion des données d'inscription et de résultats . . . . .	30
3.4	Filtage des données . . . . .	31
3.4.1	Nationalité . . . . .	31
3.4.2	Séries du baccalauréat . . . . .	31
3.4.3	Niveau universitaire . . . . .	31
3.4.4	DataFrame final . . . . .	32
3.5	Conclusion . . . . .	32
<b>4</b>	<b>Analyse du taux de réussite au bac</b>	<b>33</b>
4.1	Introduction . . . . .	33
4.2	Analyse de globale . . . . .	33
4.2.1	Évolution des effectifs . . . . .	34
4.2.2	Évolution du taux de réussite . . . . .	35
4.3	Analyse de la transition entre les séries G et STEG . . . . .	36
4.3.1	Évolution du nombre d'inscrits dans les séries G et STEG . . . . .	36
4.3.2	Évolution du taux de réussite dans les séries G et STEG . . . . .	37
4.4	analyse des séries Arabes et Franco-Arabes . . . . .	38
4.4.1	Série LA et LAR . . . . .	38
4.4.2	Série S2A . . . . .	40
4.4.3	Série S1A . . . . .	41
4.5	prédiction du taux de réussite . . . . .	42
4.6	Conclusion . . . . .	42
<b>5</b>	<b>Analyse du parcours universitaire des bacheliers (UCAD)</b>	<b>43</b>
5.1	Introduction . . . . .	43
5.2	Évolution des inscriptions à l'UCAD . . . . .	44
5.2.1	Les inscrits des série STEG et G . . . . .	44
5.2.2	Les inscrits des séries Arabes et Franco-Arabes . . . . .	45
5.3	Répartition des Inscrits par Établissement et Département . . . . .	47
5.3.1	Série STEG et G . . . . .	47
5.3.2	Série Arabes et Franco-Arabes . . . . .	49
5.4	Analyse du parcours universitaire des bacheliers (suivi des cohortes) . . . . .	56
5.4.1	Série Arabes et Franco-Arabes . . . . .	56
5.4.2	Série de reference pour les séries Arabes et Franco-Arabes . . . . .	60

5.4.3	Série STEG et G . . . . .	63
5.5	Analyse des performances académiques . . . . .	65
5.6	Conclusion . . . . .	65
<b>6</b>	<b>Restitution interactive et visualisation</b>	<b>66</b>
6.1	Introduction . . . . .	66
6.2	objectifs du dashboard . . . . .	66
6.3	Présentation des indicateurs suivis . . . . .	66
6.4	Construction du dashboard Power BI . . . . .	68
6.4.1	Première version : Résultats de 2024 uniquement . . . . .	68
6.4.2	Deuxième version : Résultats de 2006 à 2024 . . . . .	69
6.4.3	Utilisation de Power Query et création de mesures . . . . .	70
6.5	Conclusion . . . . .	70
	<b>Conclusion Générale et Recommandations</b>	<b>71</b>

# Table des figures

4.1	Évolution du nombre d'inscrits, présents et admis au bac (2006-2024) . . . . .	34
4.2	Évolution du taux de réussite au bac (2006-2024) . . . . .	35
4.3	Évolution du nombre d'inscrits dans les séries G et STEG (2006-2024) . . . . .	36
4.4	Évolution du taux de réussite dans les séries G et STEG (2006-2024) . . . . .	37
4.5	Évolution du nombre d'inscrits et du taux de réussite dans les séries LA et LAR (2006-2024) . . . . .	38
4.6	Évolution du nombre d'inscrits et du taux de réussite dans la série S2A (2006-2024)	40
4.7	Évolution du nombre d'inscrits et du taux de réussite dans la série S2A (2006-2024)	41
5.1	Évolution des inscriptions à l'UCAD pour les séries STEG et G . . . . .	44
5.2	Évolution des inscriptions à l'UCAD pour les séries LA et LAR . . . . .	45
5.3	Évolution des inscriptions à l'UCAD pour les séries S2A et S1A . . . . .	46
5.4	Top 5 des établissements avec le plus d'inscrits (G, 2018-2019) . . . . .	47
5.5	Top 5 des établissements avec le plus d'inscrits (STEG, 2023-2024) . . . . .	47
5.6	Top 5 des départements avec le plus d'inscrits (G, 2018-2019) . . . . .	48
5.7	Top 5 des départements avec le plus d'inscrits (STEG, 2023-2024) . . . . .	48
5.8	Top 5 des établissements avec le plus d'inscrits (LA, 2023-2024) . . . . .	49
5.9	Top 5 des départements avec le plus d'inscrits (LA, 2023-2024) . . . . .	50
5.10	Top 5 des établissements avec le plus d'inscrits (LAR, 2023-2024) . . . . .	50
5.11	Top 5 des départements avec le plus d'inscrits (LAR, 2023-2024) . . . . .	51
5.12	Top 5 des établissements avec le plus d'inscrits (S2A, 2023-2024) . . . . .	52
5.13	Top 5 des départements avec le plus d'inscrits (S2A, 2023-2024) . . . . .	52
5.14	Top 5 des établissements avec le plus d'inscrits (S1A, 2023-2024) . . . . .	54
5.15	Top 5 des départements avec le plus d'inscrits (S1A, 2023-2024) . . . . .	54
5.16	Graphe suivi de la cohorte 2018 (LA) . . . . .	56
5.17	Graphe suivi de la cohorte 2018 (LAR) . . . . .	57
5.18	Graphe suivi de la cohorte 2018 (S2A) . . . . .	58
5.19	Graphe suivi de la cohorte 2018 (S1A) . . . . .	59
5.20	Graphe suivi de la cohorte 2018 (L'1) . . . . .	60
5.21	Graphe suivi de la cohorte 2018 (S2) . . . . .	61
5.22	Graphe suivi de la cohorte 2018 (S1) . . . . .	62
5.23	Graphe suivi de la cohorte 2019 (STEG) . . . . .	63



5.24	Graphe suivi de la cohorte 2018 (G)	64
6.1	Tableau de bord Power BI - Résultats du baccalauréat 2024	68
6.2	Tableau de bord Power BI - Résultats du baccalauréat 2006 à 2024	69

## Table des codes

3.1	Informations général du DataFrame	24
3.2	Nombre de numero_table en doublon par année	25
3.3	Correction du type de la colonne moy_finale	27
3.4	Création de nouvelles colonnes	28
3.5	Info global du data des inscriptions	29
3.6	Info global du data des résultats	29
3.7	Jointure des données d'inscription et de résultats	30
3.8	Info global du data des inscriptions et résultats	30
3.9	Filtrage des données	32

## Liste des tableaux

3.1	lignes avec les numéros en doublon en 2012	25
3.2	Valeurs manquantes dans les colonnes du DataFrame des résultats du bac	26
3.3	Valeurs manquantes dans la colonne moy_finale	26

# Introduction Générale

## Contexte et justification

Le système éducatif sénégalais, hérité du modèle français, repose sur une structuration en trois cycles : primaire, moyen et secondaire. La fin du secondaire est sanctionnée par le baccalauréat, diplôme pivot marquant l'accès à l'enseignement supérieur et considéré comme le premier grade universitaire. Il joue un rôle central, non seulement en tant que certificat de fin d'études, mais aussi comme indicateur de performance du système éducatif.

Malgré les réformes successives visant à adapter ce diplôme aux réalités nationales et aux enjeux contemporains, le baccalauréat au Sénégal reste confronté à un taux de réussite relativement faible, oscillant autour de 40%, loin des standards observés dans des pays comme la France ou le Canada où il s'oscille au tour de 99% [1]. Ce paradoxe entre le nombre croissant de candidats et la stagnation du taux de réussite interroge sur l'efficacité des politiques éducatives mises en œuvre.

Dans ce contexte, plusieurs réformes majeures ont été introduites : l'instauration du baccalauréat arabo-français en 2000, la création du bac arabe en 2013, et la transformation de la série G en série STEG (Sciences et Techniques de la Gestion) en 2019. Ces réformes traduisent la volonté des autorités d'élargir les opportunités d'accès à l'enseignement supérieur, de diversifier les profils de bacheliers, et d'améliorer l'adéquation entre formation et marché du travail. Ce mémoire se propose d'évaluer l'impact réel de ces transformations sur le taux de réussite au baccalauréat et sur le parcours universitaire des bacheliers, notamment à l'UCAD.

## Problématique et Objectifs de l'étude

Les réformes successives du baccalauréat sénégalais, notamment l'introduction du bac arabe en 2013 et de la série STEG en 2019, ont été motivées par la volonté de moderniser le système éducatif, de mieux adapter l'offre de formation aux besoins socio-économiques du pays et de favoriser l'inclusion. Toutefois, leur impact réel reste encore peu documenté de manière rigoureuse.

Une question centrale se pose :

**Ces réformes ont-elles réellement contribué à améliorer les performances au baccalauréat et à faciliter le parcours universitaire des nouveaux bacheliers ?**

Plusieurs interrogations découlent de cette problématique :

- Les réformes ont-elles eu un effet mesurable sur les taux de réussite au baccalauréat, globalement et par série ?
- Comment les nouvelles séries, comme le bac arabe ou la série STEG, se positionnent-elles en termes de performances au bac ?
- Comment les bacheliers issus des séries arabes évoluent-ils dans l'enseignement supérieur, notamment à l'UCAD ?
- Existe-t-il des disparités significatives entre les différentes séries du baccalauréat en termes de réussite et d'insertion à l'UCAD ?

À travers une analyse des données du baccalauréat (2006–2024) et des inscriptions à l'UCAD (2002–2024), ce travail poursuit plusieurs objectifs :

- **Analyser l'évolution des taux de réussite au bac**, avec une attention particulière aux périodes post-réformes.
- **Modéliser et prédire les taux de réussite** pour estimer ce qu'ils auraient été sans réformes (approche contrefactuelle).
- **Étudier le parcours universitaire des bacheliers (suivi de cohortes)**, en identifiant les départements, filières et performances selon les séries d'origine.
- Fournir des **recommandations basées sur les données** pour orienter les politiques éducatives futures.

Ce travail vise ainsi à combler un manque crucial d'évaluation quantitative des réformes éducatives, en mobilisant des méthodes statistiques, des modèles prédictifs et des visualisations, pour offrir une lecture factuelle et visuelle des effets des réformes, afin d'éclairer la prise de décision dans le secteur éducatif.

## Méthodologie et Organisation du document

Cette étude repose sur une démarche méthodologique structurée, combinant analyse statistique, modélisation prédictive et visualisation avancée. L'objectif est d'évaluer de manière quantitative et visuelle l'impact des réformes du baccalauréat sur la réussite scolaire et le parcours universitaire.

La méthodologie adoptée comprend :

- Une **revue des réformes éducatives** ayant marqué le baccalauréat sénégalais.
- L'exploitation de trois grandes bases de données : les **résultats du baccalauréat** (2006–2024), les **inscriptions universitaires à l'UCAD** (2002–2024) et les **résultats universitaires à l'UCAD** (2011–2024).
- Une **analyse descriptive et prédictive**, appuyée par des représentations visuelles riches afin de faciliter la compréhension et l'interprétation des résultats.
- Un **suivi longitudinal des cohortes** de bacheliers à l'université, avec une attention particulière portée aux différences selon les séries d'origine.
- Une **restitution interactive** à l'aide de tableaux de bord dynamiques réalisés avec Power BI.

Le document est organisé en six chapitres principaux, suivis d'une section dédiée à la conclusion et aux recommandations, suivant la logique de cette démarche :

1. **État de l'art** : Présentation des réformes du baccalauréat et revue des études antérieures sur le sujet.
2. **Données et outils utilisés** : Description des sources de données et des outils mobilisés.
3. **Nettoyage et préparation des données** : Traitement, fusion et structuration des données pour les rendre exploitables.
4. **Analyse du taux de réussite au bac** : lecture visuelle et modélisation des taux avec simulations.
5. **Analyse du parcours universitaire et suivi de cohortes** : visualisation du cheminement des bacheliers à l'UCAD.
6. **Restitution interactive et visualisation** : Présentation des dashboard Power BI, incluant les filtres dynamiques, mesures, colonnes conditionnelles et visualisations interactives.

**Conclusion et recommandations** : Bilan général de l'étude, limites rencontrées et propositions pour renforcer l'efficacité des politiques éducatives.

# Chapitre 1

## État de l’art

### 1.1 Introduction

Depuis sa création, comme énoncé précédemment, le baccalauréat sénégalais a connu de nombreuses évolutions, tant dans sa structure que dans ses objectifs. Initialement conçu comme un simple diplôme de fin d’études secondaires, il est aujourd’hui devenu un véritable levier d’accès à l’enseignement supérieur et un indicateur clé de performance du système éducatif national.

Ce chapitre propose une revue synthétique des principales réformes ayant marqué l’histoire du baccalauréat au Sénégal, notamment celles de 1995, 2000, 2013 (introduction du bac arabe) et 2019 (création de la série STEG). Il s’agira d’examiner le contexte de leur mise en œuvre, les motivations qui les ont justifiées, ainsi que leurs ambitions en termes de modernisation, d’inclusion et de diversification des profils d’élèves.

Cette mise en perspective historique permet de mieux situer l’organisation de notre étude, qui vise à évaluer l’impact réel de ces réformes sur les performances scolaires (taux de réussite) et les parcours universitaires (suivi de cohortes à l’UCAD). Elle constitue ainsi une étape essentielle pour comprendre les enjeux de notre analyse et justifier le recours à une approche fondée sur les données et la visualisation interactive.

### 1.2 Historique et rôle du baccalauréat

Le baccalauréat a été inventé en France au XIX<sup>e</sup> siècle par un décret de l’empereur Napoléon Ier en 1808. À l’origine appelé le « **Bachot** », il tire son étymologie du latin médiéval "*bacca laurea*", il désigne en latin médiéval **la couronne de laurier** remise aux vainqueurs [3].

Dans le contexte sénégalais, il a été progressivement intégré au système éducatif colonial, puis nationalisé après l'indépendance. Il sanctionne la fin du cycle secondaire et donne accès à l'université, devenant ainsi un indicateur majeur de performance du système éducatif.

Aujourd'hui, le baccalauréat incarne à la fois une étape symbolique vers l'âge adulte et un enjeu stratégique pour le développement du capital humain, d'où l'importance des réformes visant à le rendre plus inclusif et mieux adapté aux réalités sociales et économiques du pays.

### 1.3 Organisation du baccalauréat : structure et acteurs

Le baccalauréat Sénégalais est organisé par l'**Office du baccalauréat**, rattaché à l'Université Cheikh Anta Diop de Dakar (UCAD), ce qui renforce l'idée qu'il constitue le premier diplôme universitaire. Chaque année, son organisation mobilise plusieurs milliers d'acteurs : enseignants, surveillants, correcteurs et présidents de jury.

L'examen est structuré autour de différentes séries, réparties en trois grandes filières : **littéraire**, **scientifique** et **tertiaire**. Cette diversification vise à mieux prendre en compte la pluralité des profils et des parcours de formation des élèves. Les épreuves s'appuient sur des programmes nationaux fixés par le ministère de l'Éducation, et sont organisées de manière centralisée, tant dans leur déroulement que dans leur correction.

Ce travail est réalisé en étroite collaboration avec l'Office du Baccalauréat, ce qui permet un accès privilégié à certaines données et une meilleure compréhension des mécanismes internes liés à l'organisation et à l'évolution de l'examen.

### 1.4 Réformes du baccalauréat au Sénégal

Depuis les années 1990, plusieurs réformes ont été introduites pour adapter le baccalauréat aux réalités sociolinguistiques, économiques et pédagogiques du pays.

#### 1.4.1 Réformes de 1995 (décret n° 95-947)

Certainement la réforme la plus marquante, elle s'inscrit dans le cadre de la concertation nationale sur l'enseignement supérieur tenu le 9 décembre 1995, dont sont issues les propositions de la commission qui avait été chargée de réfléchir sur l'examen du baccalauréat [7].

Elle a introduit les changements majeurs dans la structure des séries telles qu'on les connaît aujourd'hui, notamment par l'article 7 du décret établissant les choix des séries que les candidats devront choisir au moment de leurs inscriptions :

- Série L1 : Langues et Civilisations
- Série L2 : Sciences sociales et humaines
- Série G : Techniques quantitatives d'économie et de gestion
- Série S1 : Sciences exactes (Mathématiques et Physique)
- Série S2 : Sciences expérimentales
- Série S3 : Sciences et Techniques
- Série T1 : Fabrication mécanique
- Série T2 : Électrotechnique-Électronique

C'est pour modifier et compléter cette article 7 du décret n°1995-947 que toutes les réformes étudiées dans ce mémoire ont été introduites.

### 1.4.2 Réformes de 2000 (décret n° 2000-585)

Apparue très tôt dans l'élémentaire, la langue arabe sera reconnue comme langue vivante étrangère au Sénégal, laissée au choix de l'élève dans l'enseignement moyen, puis secondaire.

Parallèlement à cette évolution dans l'enseignement public, l'initiative privée, portée par une demande socio-culturelle, a donné naissance à un système d'enseignement arabe encore embryonnaire. Ainsi, avec la multiplication des établissements exclusivement arabe ou bilingue franco-arabe, la nécessité d'encadrer ce phénomène a conduit l'État sénégalais à mettre en place un référentiel de diplôme : c'est la création du **Certificat arabe** et du **BFEM arabe** [4].

La création du collège public franco-arabe **Mouhamadou Fadilou Mbacké** de Dakar en 1963 procède de cette volonté de répondre à cette demande sociale. Un second cycle y a été mis en place, offrant les même perspectives que celles proposées aux autres collégiens. Ainsi, le décret n°2000-585 du 6 juillet 2000 a introduit le baccalauréat option arabe qui modifie l'article 7 du décret n° 95-947 en un article 7 bis en intégrant les baccalauréats option arabe suivants :

- Langues et Sciences sociales : LA
- Sciences fondamentales : S1A
- Sciences appliquées : S2A

### 1.4.3 Réformes de 2013 (décret n° 2013-913)

*"Le baccalauréat organisé (matérialisé) par le décret n°2000-586 du 20 juillet 2000, n'a pas atteint tous ses objectifs car des écoles privées franco-arabes et des instituts islamiques se multiplient entraînant un manque de contrôle sur les programmes enseignés ainsi qu'une prolifération de diplômes qui empêche l'existence d'un standard commun à toutes les études secondaires du pays"[5].*

C'est pour encadrer ce phénomène que, tout comme les études secondaires franco-arabes, les études secondaires arabes seront sanctionnées par les baccalauréats arabes. Ces derniers sont définis dans un article 7 ter, qui complète l'article 7 bis du décret n° 2000-585, en intégrant les séries suivantes :

- Littératures et Civilisations arabes : L-AR
- Mathématiques et Sciences Physique : S1-AR
- Sciences expérimentales : S2-AR

**Remarque importante :** Il convient de souligner que, bien que le décret n° 2013-913 ait introduit les séries **S1-AR** et **S2-AR**, les données disponibles et la pratique sur le terrain indiquent que ces séries scientifiques n'ont **jamais été concrètement mises en place ni organisées** depuis la promulgation de ce décret. Seule la série L-AR est actuellement opérationnelle.

### 1.4.4 Réformes de 2019(décret n° 2019-645)

À la différence des précédentes réformes, celle-ci a pour but de renforcer la série G aux exigences des Formations Professionnelles et Techniques(FPT) et des Instituts Supérieurs d'Enseignement Professionnel (ISEP).

*"C'est dans ce cadre qu'il est proposé que la série G (économie et gestion) soit transformée en série technologique (sciences et technologique de l'économie et de la gestion : STEG) visant essentiellement à installer chez les élèves les compétences en associant la culture générale et la technologie"[6].*

## 1.5 Travaux antérieures

Plusieurs études ont porté sur la performance du système éducatif sénégalais et l'évolution du baccalauréat.



L'article de Diagne (2023) analyse l'évolution du taux de réussite au bac sur deux décennies (2001–2022). Il met en évidence l'instabilité chronique du système, marquée par une réussite oscillant autour de 40 %, bien inférieure aux standards internationaux. L'auteur identifie plusieurs facteurs influents : interruptions pédagogiques, surcharge des programmes, conditions socio-économiques, etc. Malgré les nombreuses réformes (EPT 1990, PDEF 2003, PAQUET-ET 2013), les performances restent en deçà des attentes [1].

Par ailleurs, le rapport de la DES de l'UCAD (2021) adopte une approche longitudinale de suivi de cohorte pour évaluer l'efficacité interne des formations universitaires. À partir de la cohorte 2013-2014, il retrace les parcours des étudiants selon leur série de bac et leur établissement d'accueil à l'UCAD. Cette étude montre des taux importants de redoublement et d'abandon dès les premières années universitaires, avec des écarts marqués selon les séries d'origine [2].

Ces travaux révèlent un manque d'articulation entre les réformes du bac et les trajectoires universitaires, d'où l'importance d'une évaluation intégrée comme celle proposée ici.

### 1.6 Approche méthodologique et innovante de l'étude

Ce travail s'inscrit dans le cadre d'un stage à la Direction des Études et des Statistiques (DES) de l'UCAD, en collaboration avec l'Office du Baccalauréat. Il combine une analyse data-driven et des méthodes de machine learning pour évaluer l'impact des réformes sur deux axes :

1. **performance académique** : Analyse de l'évolution du taux de réussite au bac de 2006 à 2024 globale et des série concerner, via des modèles de régression.
2. **performance universitaire** : Une suivit des cohortes ici des séries concerner du bac, avec une attention particulière à leur insertion et leur progression à l'UCAD.

Un **tableau de bord interactif** sera mis en place pour visualiser les résultats de l'analyse et permettre une exploration dynamique des données, en vue de faciliter la prise de décision et la formulation de recommandations éclairées.

### 1.7 Conclusion

Ce premier chapitre a permis de retracer l'évolution historique et structurelle du baccalauréat au Sénégal, en mettant en évidence les principales réformes qui l'ont façonné depuis les années 1990. Motivées par des enjeux pédagogiques, socioculturels et économiques, ces réformes ont conduit à une transformation en profondeur du système : modification des séries existantes, introduction de nouvelles séries, et adaptation progressive aux réalités nationales.

Elles témoignent de la volonté des autorités de moderniser et de diversifier l'offre éducative afin de mieux répondre aux besoins de la société sénégalaise. Cette contextualisation est indispensable pour comprendre les fondements de notre étude, qui vise à évaluer l'impact concret de ces réformes à travers une approche analytique et prédictive, fondée sur l'exploitation rigoureuse de données réelles.

# Chapitre 2

## Données et outils utilisés

### 2.1 Introduction

Ce chapitre introduit les données exploitées ainsi que les outils utilisés pour mener à bien notre analyse. Il constitue une étape essentielle dans la construction de l'étude, car la qualité et la structuration des données conditionnent la pertinence des résultats.

Les données utilisées couvrent une période suffisamment large pour permettre des comparaisons temporelles et une évaluation rigoureuse des effets des réformes. Elles ont été préparées et organisées en vue de faciliter les analyses statistiques, les visualisations, ainsi que les modélisations prédictives.

Les outils mobilisés permettent à la fois le traitement, l'analyse et la visuelle claire et dynamique des résultats.

### 2.2 Sources et Description des données

#### 2.2.1 Sources de données

Ce projet s'appuie sur trois principales sources de données fournies par deux institutions officielles :

- **l'Office du Baccalauréat du Sénégal**, chargé de l'organisation de l'examen et de la production des résultats officiels du bac sur l'ensemble du territoire national ;
- **la Direction de l'Informatique et des Systèmes d'Information (DISI)** de l'UCAD, responsable de la gestion des données académiques et administratives à l'université ;

### 2.2.2 Description des données

L'étude repose sur trois ensembles de données complémentaires, couvrant le parcours des apprenants depuis l'obtention du baccalauréat jusqu'à leur progression dans l'enseignement supérieur à l'UCAD.

#### Résultats du Baccalauréat (2006–2024)

Ces données ont été fournies par l'Office du Baccalauréat du Sénégal. Elles contiennent les informations relatives aux candidats (année, série, résultat, mention, session, etc.). Elles permettent de mesurer les taux de réussite globaux et par série, et de suivre l'impact des réformes.

#### Inscriptions universitaires à l'UCAD (2002–2024)

Issues de la Direction de la DISI, ces données renseignent sur le profil des étudiants inscrits à l'UCAD : année d'inscription, série d'origine au bac, établissement d'accueil, etc. Elles offrent une vision globale de l'évolution des flux d'entrée à l'université selon les caractéristiques des bacheliers.

#### Résultats académiques à l'UCAD (2010–2024)

Ces données, également fournies par la DISI, détaillent la performance universitaire des étudiants : moyennes annuelles, crédits, session, mention, résultats. Elles permettent d'apprécier la réussite dans le supérieur en fonction du parcours scolaire initial.

La combinaison de ces trois bases – résultats au bac, inscriptions et performances universitaires – permet de reconstruire des trajectoires individuelles et de réaliser un suivi de cohorte. Grâce aux variables communes (année, série du bac, résultat, etc.), il est possible de lier le profil d'entrée des étudiants à leur évolution à l'université, et d'évaluer ainsi l'impact des séries du baccalauréat sur leur réussite postscolaire.

## 2.3 Outils et Technologies utilisés

### 2.3.1 Python et ses bibliothèques



Python est un langage de programmation open source, interprété, simple à apprendre et largement utilisé dans le domaine scientifique et technique. Python s'est imposé comme l'un des outils les plus puissants pour l'analyse de données et le développement de modèles d'intelligence artificielle, notamment en raison de la richesse de ses bibliothèques spécialisées [11].

Dans le cadre de cette étude, Python a été utilisé à la fois pour le traitement, l'exploration et la visualisation des données, ainsi que pour le développement de modèles de machine learning.

#### bibliothèques utilisées



Pandas est une bibliothèque Python spécialisée dans la manipulation et l'analyse de données. Elle offre des structures de données flexibles et efficaces, notamment les **DataFrame**, qui permettent de gérer facilement des tableaux de données similaires à ceux d'Excel ou d'une base de données[9].

C'est sans doute le package le plus utilisé en science des données. Grâce à ses nombreuses fonctionnalités, il permet de charger, filtrer, nettoyer et transformer les jeux de données de manière rapide et intuitive.



Matplotlib est une bibliothèque Python dédiée à la visualisation de données. Elle permet de créer une grande variété de graphiques statiques. Son interface simple et sa compatibilité avec les structures de données comme les DataFrame de pandas en font un outil de choix pour représenter visuellement les résultats d'une analyse [8].



Scikit-learn est une bibliothèque Python dédiée au machine learning. Elle offre des outils simples et efficaces pour appliquer des modèles de classification, de régression et de clustering [12].

Dans mon projet, je l'ai utilisée pour prédire le taux de réussite au baccalauréat. Elle m'a aussi permis d'évaluer les performances des modèles grâce à des métriques comme (RMSE).



Statsmodels est une bibliothèque Python conçue pour l'estimation de modèles statistiques. Elle est particulièrement utilisée pour les analyses de régression, les séries temporelles et les tests statistiques [13].

Dans mon projet, Statsmodels m'a servi à réaliser des régressions linéaires et à effectuer des tests d'hypothèses. Elle m'a permis d'interpréter les relations entre les variables, grâce à des résultats détaillés incluant les coefficients, les p-values et les intervalles de confiance.

### 2.3.2 Power BI



Power BI est un outil de visualisation et d'analyse de données développé par Microsoft. Il permet de créer des tableaux de bord interactifs et dynamiques à partir de diverses sources de données, facilitant ainsi l'exploration visuelle et la prise de décision basée sur les données [10].

Dans le cadre de mon projet, j'ai utilisé Power BI pour concevoir un tableau de bord interactif regroupant les statistiques du baccalauréat de 2006 à 2024. Ce tableau de bord, présenté dans le chapitre 6, est destiné à l'Office du Baccalauréat. Il a pour objectif de faciliter l'analyse des données historiques du bac et de servir de support à toute nouvelle étude portant sur l'évolution du système éducatif.

## 2.4 Conclusion

Ce chapitre a permis de présenter les données exploitées ainsi que les outils mobilisés pour leur traitement, leur analyse et leur visualisation. Les trois jeux de données, provenant de sources officielles telles que l'Office du Baccalauréat et la DISI de l'UCAD, offrent une base solide pour évaluer l'impact des réformes éducatives sur les résultats au bac et l'insertion universitaire. Par ailleurs, l'utilisation d'outils puissants comme Python et ses bibliothèques spécialisées, ainsi que Power BI pour la visualisation, garantit une analyse rigoureuse, reproductible et accessible. Ces ressources forment ainsi le socle méthodologique sur lequel reposent les analyses menées dans les chapitres suivants.

# Chapitre 3

## Nettoyage et préparation des données

### 3.1 Introduction

Avant d’entamer toute analyse statistique, une étape cruciale consiste à examiner, nettoyer et structurer les données. Bien que les jeux de données utilisés dans ce projet proviennent de sources officielles telles que l’Office du Bac et la DISI/UCAD, et présentent globalement une structure cohérente, un travail de nettoyage s’est avéré nécessaire.

Contrairement à l’hypothèse initiale selon laquelle les données ne nécessiteraient qu’un traitement minimal, plusieurs opérations classiques de nettoyage ont finalement été réalisées. Cela inclut la détection et le traitement de valeurs aberrantes, la gestion des doublons, ainsi que le traitement ciblé des valeurs manquantes. Par exemple, dans les résultats universitaires, certaines valeurs nulles traduisent l’absence à une évaluation, mais d’autres relevaient d’incohérences ou de saisies incomplètes qu’il a fallu corriger ou exclure selon le contexte.

En parallèle, des opérations de préparation ont été menées pour structurer les données en vue de l’analyse. Celles-ci comprennent la fusion de bases de données complémentaires (Les Inscrits et Résultats de UCAD), la standardisation des formats de certaines variables, ainsi que la création de nouvelles colonnes dérivées utiles à l’étude.

### 3.2 Préparation de données des résultats du baccalauréat

Les données relatives aux résultats du baccalauréat ont été consolidées dans une table unique après concaténation de fichiers annuels couvrant la période de 2006 à 2024. Cette base contient plus de deux millions d’enregistrements, représentant une volumétrie significative pour l’analyse.



Le code ci-dessous (code 3.1) donne un aperçu général de la structure du DataFrame après concaténation des fichiers annuels.

Code 3.1 – Informations général du DataFrame

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2236490 entries, 0 to 2236489
Data columns (total 15 columns):
#   Column                Dtype
---  -
0   nom                   object
1   prenom                object
2   numero_table          object
3   serie                 object
4   sexe                 object
5   age                  object
6   etablissement         object
7   type_candidat        object
8   resultat             object
9   acad_provenance      object
10  moy_finale            object
11  mention               object
12  abs                   object
13  exclusion             object
14  year                  object
dtypes: object(15)
memory usage: 255.9+ MB
```

Cette base comprend des informations variées telles que le nom, le prénom, le numéro de table, la série, le sexe, l'âge, l'établissement d'origine, le type de candidat, les résultats obtenus, la mention, la moyenne finale, ainsi que l'année de passage. Elle constitue une source clé pour l'analyse du taux de réussite et l'évaluation de l'impact des réformes introduites dans certaines séries.

### 3.2.1 Doublons dans la clé primaire numero\_table

L'identifiant principal des candidats étant le `numero_table`, une vérification a été effectuée afin d'identifier d'éventuels doublons par année.

Code 3.2 – Nombre de numero\_table en doublon par année

```

Annee 2006 : 0 numeros apparaissent au moins deux fois
Annee 2007 : 0 numeros apparaissent au moins deux fois
Annee 2008 : 371 numeros apparaissent au moins deux fois
Annee 2009 : 0 numeros apparaissent au moins deux fois
Annee 2010 : 0 numeros apparaissent au moins deux fois
Annee 2011 : 1 numeros apparaissent au moins deux fois
Annee 2012 : 2 numeros apparaissent au moins deux fois
Annee 2013 : 0 numeros apparaissent au moins deux fois
Annee 2014 : 0 numeros apparaissent au moins deux fois
Annee 2015 : 0 numeros apparaissent au moins deux fois
Annee 2016 : 0 numeros apparaissent au moins deux fois
Annee 2017 : 0 numeros apparaissent au moins deux fois
Annee 2018 : 0 numeros apparaissent au moins deux fois
Annee 2019 : 0 numeros apparaissent au moins deux fois
Annee 2020 : 0 numeros apparaissent au moins deux fois
Annee 2021 : 0 numeros apparaissent au moins deux fois
Annee 2022 : 0 numeros apparaissent au moins deux fois
Annee 2023 : 0 numeros apparaissent au moins deux fois
Annee 2024 : 0 numeros apparaissent au moins deux fois

```

Les résultats révèlent que la majorité des années ne présentent aucun doublon. Toutefois, quelques années, comme 2008 (371 doublons), 2011 (1) et 2012 (2) comportent des cas où deux candidats partagent le même numéro de table dans la même année.

TABLE 3.1 – lignes avec les numéros en doublon en 2012

nom	numero_table	serie	sexe	age	etablissement	moy_finale
CISSE	56234	L'1	M	17	LYCEE CHARLES DEGAULLE	11,00
KANE	56234	L2	M	21	LYCEE DE ROSS - BETHIO	07,07
BOYE	56513	L'1	F	19	LYCEE EL HADJ OMAR FOUTIYOU TALL	05,58
CAMARA	56513	L'1	M	23	LYCEE EL HADJ OMAR TALL	07,17

Le tableau (table 3.1) illustre deux exemples de doublons survenus en 2012. On observe que, bien que les numéros de table soient identiques, les profils associés sont différents (noms, âges, établissements, moyennes, etc.). Ces doublons ne traduisent donc pas des répétitions, mais plutôt des erreurs de saisie ou des anomalies dans l'attribution des identifiants.

### 3.2.2 Les valeurs manquantes dans les colonnes

L'analyse du DataFrame révèle la présence de valeurs manquantes dans plusieurs variables. Le tableau (table 3.2) en présente la répartition par colonne.

TABLE 3.2 – Valeurs manquantes dans les colonnes du DataFrame des résultats du bac

Colonnes	nb_valeur_null
nom	1
prenom	1
numero_table	3
serie	5442
sexe	1
age	1
etablissement	1
type_candidat	1
resultat	3977
acad_provenance	1
moy_finale	11
mention	945542
abs	1
exclusion	1
year	0

**Valeurs manquantes dans la colonne moy\_finale**

La colonne `moy_finale`, qui contient la moyenne finale obtenue par les candidats, comporte 11 valeurs manquantes (table 3.3).

TABLE 3.3 – Valeurs manquantes dans la colonne moy\_finale

nom	numero_table	resultat	moy_finale
NaN	42379	NaN	NaN
XXXXXXXXXX	41803	NaN	NaN
XXXXXXXXXX	42043	NaN	NaN
XXXXXXXXXX	27911	NaN	NaN
XXXXXXXXXX	24635	NaN	NaN
XXXXXXXXXX	17167	NaN	NaN
XXXXXXXXXX	17669	NaN	NaN
XXXXXXXXXX	1054	NaN	NaN
NDIAYE	NaN	NaN	NaN
FAYE	NaN	NaN	NaN
KANE	NaN	NaN	NaN

Ces lignes, souvent associées à des candidats absents ou exclus, ont été supprimées afin d’éviter tout biais dans l’analyse statistique et les modélisations

### Valeurs manquantes dans la colonne `resultat`

La colonne `resultat` contient 3 977 valeurs manquantes (table 3.2). Après vérification croisée avec les colonnes `abs` et `exclusion`, il s'avère que ces absences de valeurs concernent exclusivement des candidats exclus ou absents. Par souci de cohérence, toutes les valeurs manquantes ont été remplacées par **0**, conformément à la signification officielle des résultats :

- **111** : Admis d'office avec mention ;
- **101** : Admis après le second tour ;
- **100** : Passé au second tour sans succès ;
- **10** : Échec direct (moyenne insuffisante dès le premier tour) ;
- **0** : Absent, exclu ou dossier incomplet.

### 3.2.3 Correction du type de la colonne `moy_finale`

Comme on peut le constater dans le code présenté (code 3.1), le type initial de la colonne `moy_finale` n'était pas exploitable tel quel. En effet, certaines valeurs contenaient des tirets, d'autres utilisaient des virgules comme séparateur décimal, et certaines étaient tout simplement vides ou non numériques.

Code 3.3 – Correction du type de la colonne `moy_finale`

```
all_data_filtre['moy_finale'] = (  
    all_data_filtre['moy_finale']  
    .astype(str)                                # Convertir en string  
    .str.replace('-', '', regex=False)          # Supprimer les tirets  
    .str.replace(',', '.', regex=False)         # les virgules en points  
    .str.strip()                                # Supprimer les espaces  
    .replace({'': None, 'nan': None})           # Supprimer les chaines vides  
    .astype(float)                             # Convertir en float  
)
```

Pour rendre cette colonne exploitable statistiquement, plusieurs opérations de nettoyage ont été effectuées :

- Suppression des tirets '- '.
- Remplacement des virgules ',' par des points '.'.
- Suppression des espaces superflus.
- Conversion des chaînes vides ou non valides ('', 'nan') en valeurs manquantes.
- Conversion finale de la colonne en type `float`.

### 3.2.4 Création des colonnes **admis** et **session**

Dans le but de simplifier l'analyse et d'améliorer la lisibilité des résultats, deux nouvelles colonnes ont été dérivées à partir de la variable `resultat` :

- **admis** : indique si un candidat a été admis ou non. Les codes 111 (admis d'office avec mention) et 101 (admis au second tour) sont considérés comme « admis », les autres comme « non admis ».
- **session** : précise la session de réussite du candidat. Le code 111 correspond au « 1er Tour », et le code 101 au « 2e Tour ».

Code 3.4 – Création de nouvelles colonnes

```
# Creation de la colonne 'admis' de 'resultat'
all_data['admis'] = all_data['resultat'].apply(lambda x: 'admis'
                                              if pd.notna(x) and x in ['111', '101']
                                              else 'non_admis'
                                              )

# Creation de la colonne 'session' de 'admis'
all_data['session'] = all_data['resultat'].apply(lambda x: '1er_Tour'
                                                  if pd.notna(x) and x == '111'
                                                  else ('2e_Tour'
                                                       if pd.notna(x) and x == '101'
                                                       else ''
                                                  ))
```

## 3.3 Fusion des données des Inscrits et des Résultats de l'UCAD

### 3.3.1 Données des inscriptions à l'UCAD

Les données d'inscription utilisées dans cette étude couvrent la période de 2002 à 2024. Cependant, pour garantir la cohérence avec les données de résultats (disponibles uniquement de 2011 à 2024), nous avons retenu uniquement les inscriptions allant de 2011 à 2024.

La sortie de code 3.5 présente un aperçu global de la base de données des inscriptions à l'UCAD pour les années universitaires allant de 2011 à 2023. Elle contient 1 141 120 enregistrements répartis sur 13 variables. Les variables *NUMERO* et *ANNEE UNIVERSITAIRE*, constituent les clés essentielles pour la fusion avec la base des résultats académiques.

Code 3.5 – Info global du data des inscriptions

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1141120 entries, 0 to 1141119
Data columns (total 14 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   NUMERO                                1141120 non-null object
1   SEXE                                  1141120 non-null object
2   ANNEE_BACC                            1136174 non-null float64
3   NATIONALITE                           1141120 non-null object
4   SERIE_BACC                            1109114 non-null object
5   ETABLISSEMENT_CODE                    1141120 non-null object
6   NIVEAU_SECTION                        1141120 non-null object
7   ANNEE_INSCRIPTION                     1141120 non-null int64
8   ANNEE_UNIVERSITAIRE                   1141120 non-null object
9   TYPE_FORMATION                        1141120 non-null object
10  CODE_NIVEAU                           1141120 non-null int64
11  NIVEAU LMD ET NON LMD                  1141120 non-null object
12  SYSTEME                                1141120 non-null object
13  DEPARTEMENT FORMATION                  1141120 non-null object
dtypes: float64(1), int64(2), object(11)
memory usage: 121.9+ MB
```

### 3.3.2 Données des résultats de l'UCAD

La sortie de code 3.6 présente les informations générales de la base de données contenant les résultats universitaires des étudiants de l'UCAD. Cette base couvre la période allant de 2011 à 2024, et concerne uniquement les établissements ayant effectué leurs délibérations sur la plateforme institutionnelle de la DISI. C'est pourquoi le fichier ne contient que 753 693 enregistrements, un total largement inférieur aux 1 141 120 inscrits.

Code 3.6 – Info global du data des résultats

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 753693 entries, 0 to 753692
Data columns (total 3 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   NUMERO                                753693 non-null object
1   ANNEE_UNIVERSITAIRE                   753693 non-null object
2   RESULTAT                              753562 non-null object
dtypes: object(3)
memory usage: 17.3+ MB
```

### 3.3.3 Fusion des données d'inscription et de résultats

Pour combiner les données d'inscription et les résultats académiques des étudiants de l'UCAD, une jointure a été effectuée en utilisant les colonnes NUMERO (identifiant étudiant) et ANNEE UNIVERSITAIRE comme clés.

Code 3.7 – Jointure des données d'inscription et de résultats

```
df_final = pd.merge(df_inscrit,                # Premier DataFrame
                    df_resultat,              # Deuxieme DataFrame
                    on=['NUMERO', 'ANNEE_UNIVERSITAIRE'], # Cles de jointure
                    how='left'                # le type de jointure
                    )
```

La jointure a été réalisée à l'aide de l'option `left` pour garantir la conservation de l'ensemble des inscriptions, y compris celles des étudiants dont les résultats n'ont pas été délibérés ou publiés sur la plateforme académique de la DISI

Code 3.8 – Info global du data des inscriptions et résultats

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1215837 entries, 0 to 1215836
Data columns (total 15 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   NUMERO                                1215837 non-null object
1   SEXE                                  1215837 non-null object
2   ANNEE_BACC                            1210891 non-null float64
3   NATIONALITE                           1215837 non-null object
4   SERIE_BACC                            1183525 non-null object
5   ETABLISSEMENT_CODE                    1215837 non-null object
6   NIVEAU_SECTION                        1215837 non-null object
7   ANNEE_INSCRIPTION                     1215837 non-null int64
8   ANNEE_UNIVERSITAIRE                   1215837 non-null object
9   TYPE_FORMATION                        1215837 non-null object
10  CODE_NIVEAU                           1215837 non-null int64
11  NIVEAU_LMD_ET_NON_LMD                  1215837 non-null object
12  SYSTEME                               1215837 non-null object
13  DEPARTEMENT_FORMATION                  1215837 non-null object
14  RESULTAT                              750198 non-null object
dtypes: float64(1), int64(2), object(12)
memory usage: 139.1+ MB
```

On observe que la colonne RESULTAT contient un nombre important de valeurs manquantes qui s'explique par le fait que certains établissements ne délibèrent via la plateforme de la DISI. C'est pourquoi, ils avaient moins de lignes dans la même période dans les données des résultats.

## 3.4 Filtrage des données

Afin de concentrer l'analyse sur la population cible et les séries directement concernées par les réformes du baccalauréat, une étape de filtrage rigoureuse a été appliquée aux bases de données. L'objectif est d'exclure les individus et les séries ne correspondant pas au périmètre de l'étude.

### 3.4.1 Nationalité

La variable NATIONALITE a été filtrée pour ne conserver que les étudiants de nationalité sénégalaise. Ce choix permet de se concentrer sur les élèves issus du système éducatif sénégalais, et d'évaluer plus précisément l'effet des réformes nationales sur leurs trajectoires.

### 3.4.2 Séries du baccalauréat

L'étude porte principalement sur les séries impactées par les réformes éducatives récentes, notamment l'introduction du baccalauréat arabe (2013) et de la série STEG (2019). Ainsi, seules les séries suivantes ont été retenues :

- **Séries arabes** : L1-AR, S1-AR, S2-AR;
- **Séries franco-arabes** : LA, S1A, S2A;
- **Série STEG** : STEG;
- **Série G** : G (anciennement série G, remplacée progressivement par la STEG);
- **Séries de référence (littéraire et scientifique)** :
  - L'1 : utilisée comme référence pour les séries littéraires non arabes ;
  - S1 et S2 : références pour les filières scientifiques.

Ce choix permet de comparer les performances et parcours entre les anciennes séries, les nouvelles séries réformées, et les séries de référence.

### 3.4.3 Niveau universitaire

L'analyse se limite aux étudiants inscrits dans le **premier cycle universitaire**, c'est-à-dire les niveaux L1, L2 et L3. Ce choix vise à observer les effets immédiats et à court terme des réformes du baccalauréat sur l'insertion et la progression dans les premières années à l'université.



### 3.4.4 DataFrame final

Après l'ensemble des étapes de filtrage, de nettoyage et de fusion, le DataFrame final contient un total de 579 039 enregistrements correspondant à des étudiants de nationalité sénégalaise, inscrits à l'UCAD entre 2011 et 2024, issus des séries ciblées, et se situant dans le premier cycle universitaire.

Code 3.9 – Filtrage des données

```
<class 'pandas.core.frame.DataFrame'>
Index: 579039 entries, 7 to 1215834
Data columns (total 15 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   NUMERO                                579039 non-null  object
1   SEXE                                  579039 non-null  object
2   ANNEE_BACC                            579039 non-null  float64
3   NATIONALITE                           579039 non-null  object
4   SERIE_BACC                            579039 non-null  object
5   ETABLISSEMENT_CODE                   579039 non-null  object
6   NIVEAU_SECTION                       579039 non-null  object
7   ANNEE_INSCRIPTION                   579039 non-null  int64
8   ANNEE_UNIVERSITAIRE                  579039 non-null  object
9   TYPE_FORMATION                       579039 non-null  object
10  CODE_NIVEAU                           579039 non-null  int64
11  NIVEAU_LMD_ET_NON_LMD                 579039 non-null  object
12  SYSTEME                               579039 non-null  object
13  DEPARTEMENT_FORMATION                 579039 non-null  object
14  RESULTAT                              400944 non-null  object
dtypes: float64(1), int64(2), object(12)
memory usage: 70.7+ MB
```

## 3.5 Conclusion

Ce chapitre a permis de mettre en place une base de données propre, cohérente et ciblée, indispensable pour mener les analyses ultérieures.

Le DataFrame final obtenu constitue le socle de notre étude. Il résulte de la liaison entre les données d'inscription à l'UCAD, qui contiennent les informations sur les séries et années du baccalauréat des étudiants, et les données de résultats universitaires. Cette intégration permet de relier les parcours académiques aux caractéristiques du bac.

# Chapitre 4

## Analyse du taux de réussite au bac

### 4.1 Introduction

Ce chapitre est consacré à l'analyse du taux de réussite au baccalauréat au Sénégal sur la période 2006–2024. L'objectif est de mettre en lumière les tendances globales et spécifiques aux séries concernées par les réformes, à travers une exploration visuelle des données.

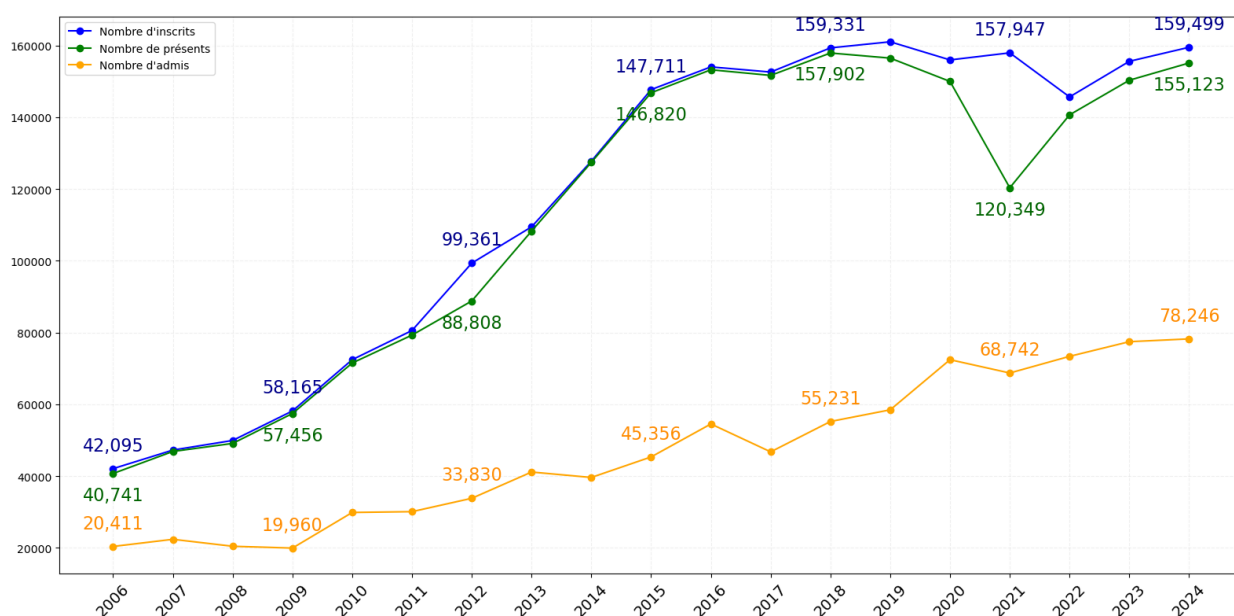
Nous débuterons par l'évolution du nombre d'inscrits et du taux de réussite global, avant de nous concentrer sur les séries spécifiques analysées dans ce travail, notamment les séries arabes, franco-arabes, STEG, G.

Des visualisations comparatives permettront d'identifier les ruptures, dynamiques et écarts entre les différentes filières. En complément, une modélisation prédictive sera réalisée afin d'estimer l'évolution théorique du taux de réussite en l'absence des réformes, offrant ainsi une base de comparaison pour mesurer leur impact.

### 4.2 Analyse de globale

### 4.2.1 Évolution des effectifs

FIGURE 4.1 – Évolution du nombre d’inscrits, présents et admis au bac (2006-2024)



La figure (Figure 4.1) représente l'évolution du nombre d'inscrits, de présents et d'admis au baccalauréat entre 2006 et 2024. L'observation met en évidence deux phases temporelles distinctes.

#### Phase 1 — Croissance rapide (2006–2015)

Durant cette période, le système éducatif sénégalais connaît une expansion notable du nombre de candidats au bac. Le nombre d'inscrits passe de 42 095 en 2006 à 147 711 en 2015, soit une croissance d'environ **250 %** en 9 ans.

Les admis augmentent également mais de manière moins rapide, passant de 20 411 en 2006 à 45 356 en 2015, soit une hausse d'environ **125 %**. Ce décalage entre l'augmentation des présents et celle des admis révèle un déséquilibre croissant sur les performances au baccalauréat.

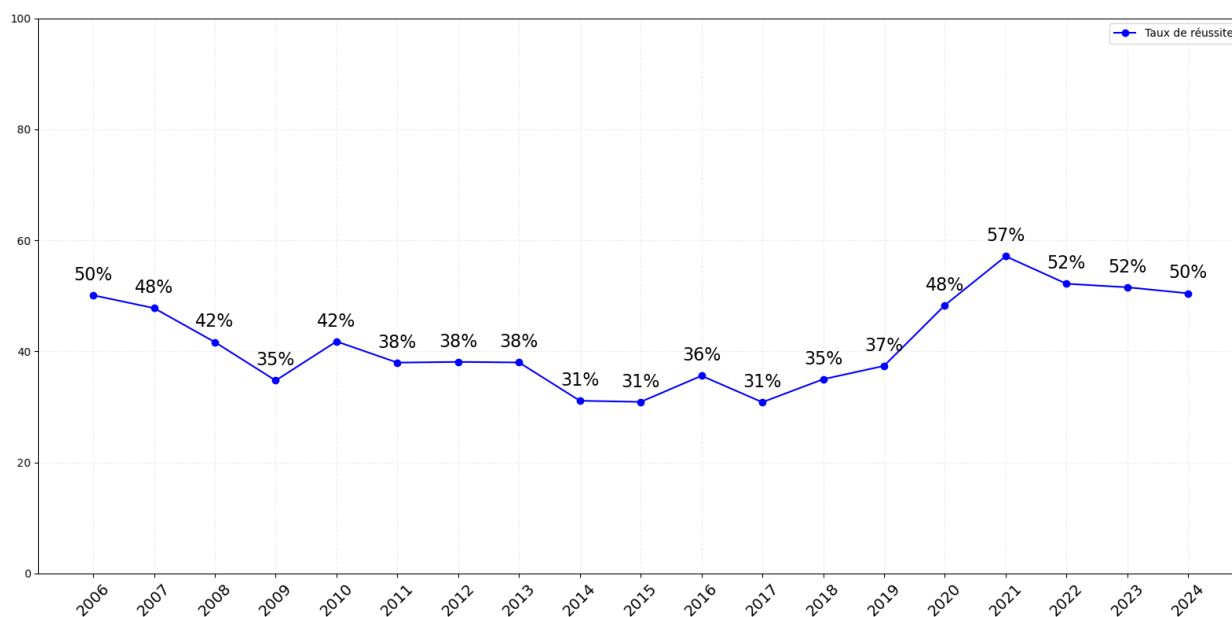
#### Phase 2 — Stabilisation relative (2015–2024)

À partir de 2015, le nombre d'inscrits se stabilise autour de 150 000 à 160 000 candidats par an. La croissance ralentit, marquant une saturation ou une stabilisation des flux en fin de cycle secondaire. Une chute brutale du nombre de présents est toutefois observée en 2021, ce qui peut être associé aux perturbations causées par la pandémie de COVID-19.

Les admissions au baccalauréat continuent d'afficher une croissance modérée, atteignant 78 246 en 2024. Toutefois, cette évolution reste notablement faible comparée au nombre de candidats inscrits et présents.

## 4.2.2 Évolution du taux de réussite

FIGURE 4.2 – Évolution du taux de réussite au bac (2006-2024)



La figure (Figure 4.2) illustre l'évolution du taux de réussite au baccalauréat sénégalais entre 2006 et 2024. L'analyse révèle des fluctuations marquées, avec des périodes de déclin et de reprise.

### Période 2006–2017 — Baisse significative

Durant cette phase, le taux de réussite connaît une chute alarmante, passant de 50% en 2006 à 31% en 2017. Cette diminution de près de 20 points en onze ans souligne des difficultés structurelles.

### Période 2017–2024 — Progression notable

À partir de 2017, le taux de réussite montre une nette amélioration, atteignant 50% en 202, avec un pic record 57% en 2021 (ce taux élevé peut s'expliquer par la chute brutale du nombre de présents liée à la pandémie, puisqu'il est calculé en fonction des candidats ayant effectivement passé l'examen). Toutefois, une légère baisse est observée entre 2021 et 2024, ce qui peut être source d'inquiétude.

Malgré une légère amélioration récente, le taux de réussite au baccalauréat demeure généralement trop faible pour un système éducatif aspirant à la performance.

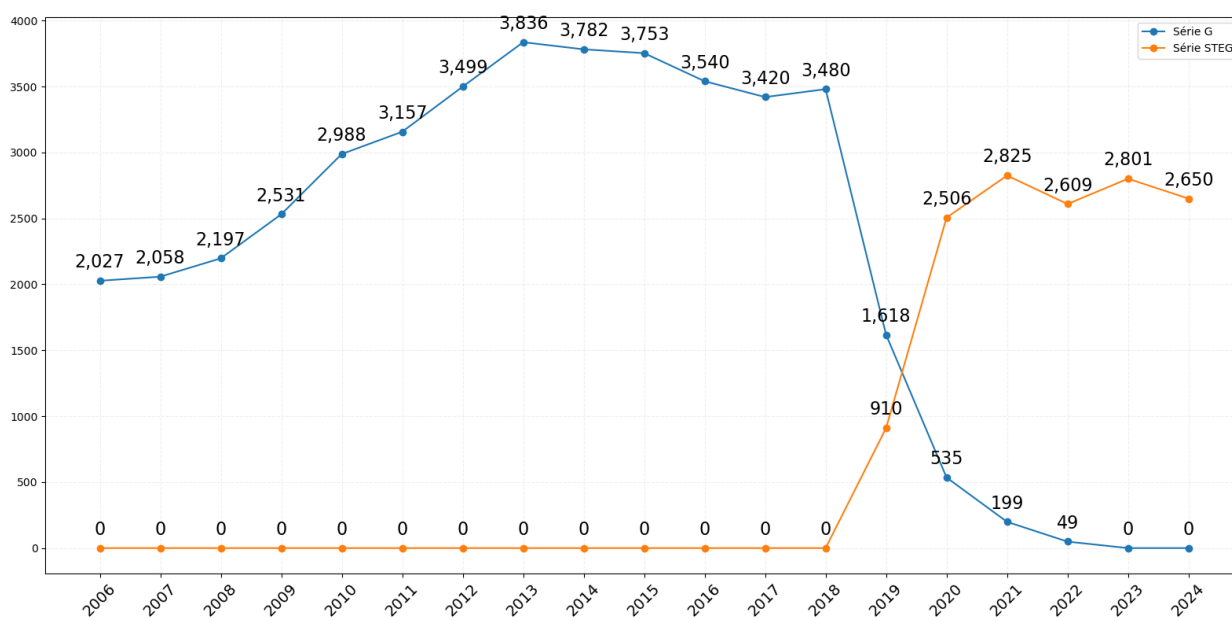
### 4.3 Analyse de la transition entre les séries G et STEG

La réforme du baccalauréat technique au Sénégal, concrétisée par le décret de 2019, marque un tournant important dans l'organisation de la série Techniques quantitatives d'économie et de gestion. Elle prévoit la **suppression progressive de la série G**, au profit de la nouvelle série **STEG**. Cette dernière vise à mieux articuler les enseignements généraux et technologiques afin de développer chez les élèves des compétences applicables dans l'enseignement supérieur et la vie professionnelle.

Conformément au décret, la série G a continué à être organisée jusqu'en 2022, date de sa dernière session. À partir de 2023, seule la série STEG est maintenue. Cette transition permet d'évaluer les effets de la réforme en comparant les performances (nombre d'inscrits et taux de réussite) de la série STEG à celles de son prédécesseur, la série G.

#### 4.3.1 Évolution du nombre d'inscrits dans les séries G et STEG

FIGURE 4.3 – Évolution du nombre d'inscrits dans les séries G et STEG (2006-2024)



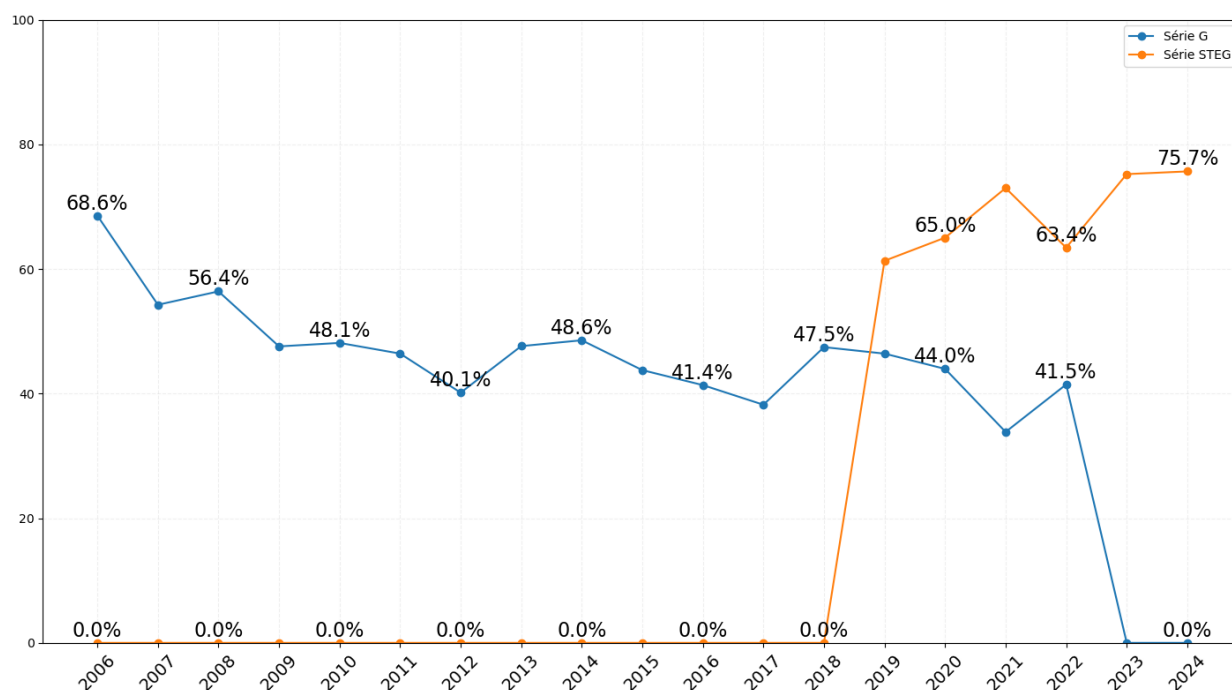
La figure (Figure 4.3) montre l'évolution du nombre d'inscrits au baccalauréat dans les séries G et STEG entre 2006 et 2024.

Jusqu'en 2018, seule la série G est présente, avec un pic d'inscription atteint entre 2012 et 2014. À partir de 2019, la série STEG est introduite et connaît une croissance rapide, tandis que les effectifs de la série G diminuent progressivement, jusqu'à leur extinction en 2023.

Ce croisement entre les deux courbes montre clairement une transition bien gérée sur le plan quantitatif, avec un transfert progressif des effectifs. Dès 2020, les inscriptions en STEG dépassent celles de la série G, traduisant une bonne adhésion des établissements et des élèves à la réforme.

### 4.3.2 Évolution du taux de réussite dans les séries G et STEG

FIGURE 4.4 – Évolution du taux de réussite dans les séries G et STEG (2006-2024)



La figure (Figure 4.4) met en perspective les performances des deux séries en termes de réussite. On observe une forte variabilité du taux de réussite en série G, fluctuant généralement entre **40%** et **50%**, avec une tendance légèrement décroissante.

En revanche, la série STEG, dès sa première session en 2019, affiche **des taux de réussite supérieurs**, allant de **60%** à plus de **75% en 2024**. Ce résultat semble valider l'objectif de la réforme qui est de renforcer l'efficacité du système en recentrant les contenus pédagogiques autour de compétences concrètes, professionnelles et transversales.

En somme, **la série STEG se distingue par de meilleures performances en matière de réussite**, tout en parvenant à capter un volume d'élèves au moins équivalent, voire supérieur, à celui de la série G à son apogée. Cela confirme la pertinence de la réforme dans le contexte de modernisation du système éducatif sénégalais.

## 4.4 analyse des séries Arabes et Franco-Arabes

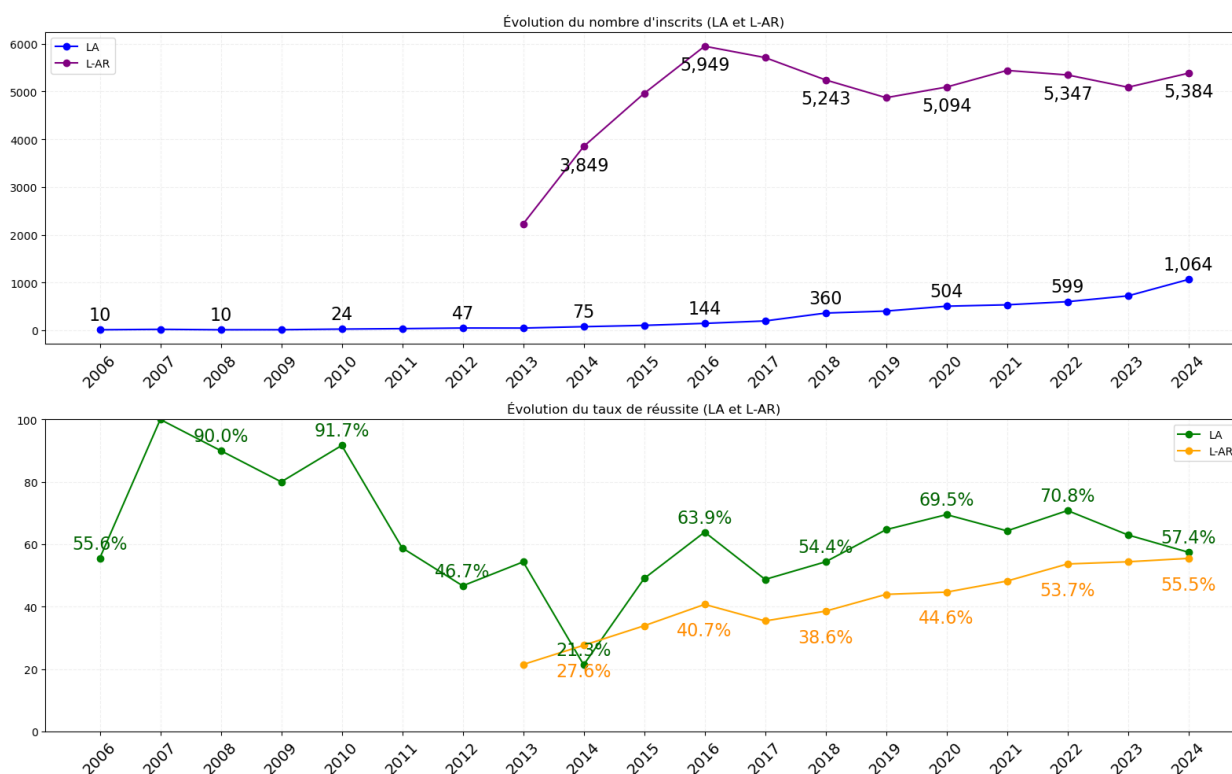
Le système éducatif sénégalais a progressivement intégré l'enseignement de l'arabe dans le secondaire, avec la création de séries Franco-arabes en 2000. Cependant, face à la prolifération des établissements d'enseignement exclusivement arabe et des instituts islamiques, le décret n°2013-057 a modifié et complété ces dispositions a fin de mieux répondre à la demande sociale suivantes :

- Littératures et Civilisations arabes (L-AR)
- Mathématiques et Sciences physiques (S1-AR)
- Sciences expérimentales (S2-AR)

Les études secondaires arabes sont désormais sanctionnées par des diplômes appelés **baccalauréats arabes**, tandis que les enseignements bilingues franco-arabe donnent lieu à des **baccalauréats franco-arabes**.

### 4.4.1 Série LA et LAR

FIGURE 4.5 – Évolution du nombre d'inscrits et du taux de réussite dans les séries LA et LAR (2006-2024)



La Figure 4.5 illustre l'évolution du nombre d'inscrits et des taux de réussite aux baccalauréats des séries Littératures et Sciences sociale (franco-arabe) LA et Littératures et Civilisations Arabes (L-AR) entre 2006 et 2024.

### Évolution du nombre d'inscrits

La série LA, présente depuis 2000, a toujours eu un nombre d'inscrits relativement faible avec une légère croissance, passant de 10 en 2006 à 1 064 en 2024. Cette augmentation, bien que modérée, témoigne d'un intérêt croissant et soutenu pour cette filière au fil des ans.

En revanche, la série L-AR, introduite en 2013, a connu une croissance fulgurante dès son apparition. Elle a attiré 2 849 inscrits en 2013, atteignant un pic de 5 949 en 2015. Bien qu'une légère diminution ait été observée par la suite, le nombre d'inscrits en L-AR s'est stabilisé autour de 5 000 à 5 300, avec 5 184 inscrits en 2024. La série L-AR a rapidement dominé en termes d'effectifs, surpassant de loin la série LA, ce qui suggère une forte adhésion à cette nouvelle filière et un transfert significatif des étudiants vers cette option réformée.

Cette dynamique met en évidence la réussite de la mise en place du baccalauréat arabe, notamment la série L-AR, qui a su capter un volume important d'étudiants, répondant ainsi à la demande sociale et aux objectifs de structuration des études arabes au Sénégal.

### Évolution du taux de réussite

La série LA a montré une grande variabilité de ses taux de réussite. Après un début à 55,6% en 2006, elle a connu des pics remarquables à 90% en 2008 et 91,7% en 2010, des chiffres à juger avec prudence car l'effectif était considérablement faible, avec une moyenne d'environ 15 candidats entre 2006 et 2010. Cependant, des baisses significatives ont suivi, avec des taux chutant à 46,7% en 2012 et un point bas à 27,5% en 2013. Par la suite, le taux de réussite en LA a montré une tendance à la reprise, atteignant 63,9% en 2015 avant de se stabiliser autour de 50% à 70% dans les années suivantes, pour finir à 57,4% en 2024.

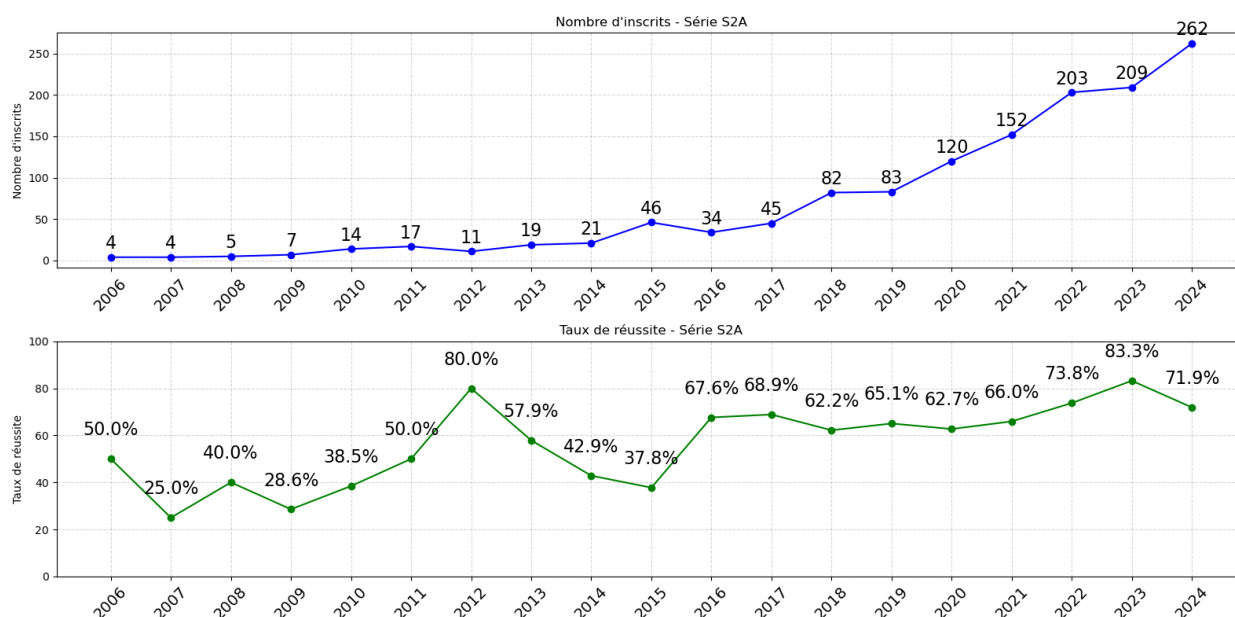
En comparaison, la série L-AR, bien que plus récente, a affiché des taux de réussite plus cohérents (explicables par un nombre de candidats assez élevé) avec une tendance croissante. À son introduction en 2013, le taux de réussite était d'environ 21%, montant progressivement pour atteindre 55,5% en 2024 qui se rapproche de celui de la série LA. Ce résultat est d'autant plus notable que la série L-AR a géré un volume d'inscrits considérablement plus important.

En conclusion, la série L-AR, malgré un volume d'effectifs beaucoup plus important, a globalement réussi à maintenir des taux de réussite compétitifs et plus stables que ceux de la série LA. Cette performance suggère que la réforme ayant introduit la série L-AR, avec son nouveau référentiel et ses objectifs d'harmonisation, a contribué à une meilleure efficacité du système pour les filières arabes au Sénégal.



### 4.4.2 Série S2A

FIGURE 4.6 – Évolution du nombre d'inscrits et du taux de réussite dans la série S2A (2006-2024)



La figure (Figure 4.6) montre l'évolution du nombre d'inscrits et du taux de réussite dans la série Sciences appliquées (S2A).

#### Évolution du nombre d'inscrits

De 2006 à 2016, la série S2A a connu une première phase de faible adhésion, avec une légère tendance à la hausse, passant de 4 inscrits en 2006 à 34 en 2016. À partir de 2017, la série S2A enregistre une croissance exponentielle de ses effectifs, atteignant 262 inscrits en 2024. La série parvient ainsi à capter un volume d'élèves en nette progression, traduisant une reconnaissance croissante de son importance.

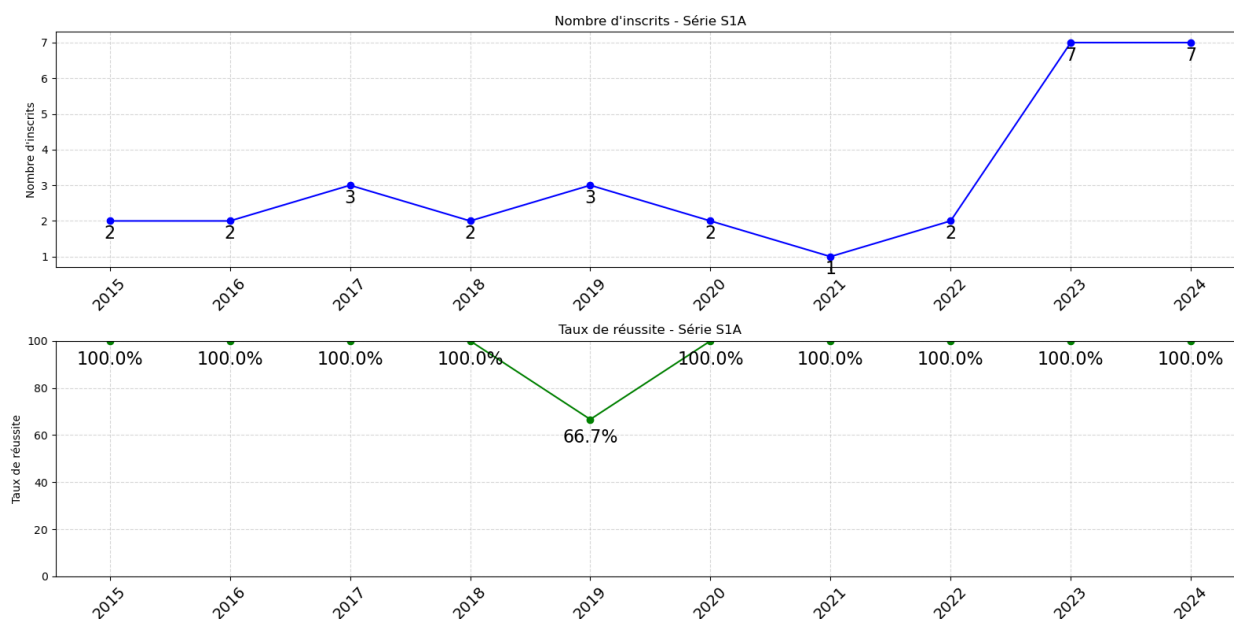
#### Évolution du taux de réussite :

Concernant le taux de réussite, la série S2A présente une forte variabilité au début de la période, en raison du faible nombre de candidats. Après un taux de 50% en 2006, il chute à 25% en 2007, puis atteint un pic exceptionnel de 80% en 2011. À partir de 2016, on observe une nette amélioration, avec des taux de réussite généralement supérieurs à 60%. La série atteint même 83,3% en 2023, avant de légèrement reculer à 71,9% en 2024.

Malgré des débuts irréguliers, la série S2A affiche de bonnes performances sur la dernière décennie, confirmant son potentiel de réussite croissant.

#### 4.4.3 Série S1A

FIGURE 4.7 – Évolution du nombre d'inscrits et du taux de réussite dans la série S2A (2006-2024)



La figure (Figure 4.7) met en perspective l'évolution du nombre d'inscrits et du taux de réussite dans la série Sciences fondamentales (S1A).

#### Évolution du nombre d'inscrits

La série S1A se caractérise par un nombre d'inscrits extrêmement faible tout au long de la période étudiée. Introduite comme option du baccalauréat secondaire arabe, elle débute avec 2 inscrits et ne dépasse jamais 7 candidats. Cette faible adhésion, avec des années où les effectifs descendent à 1 ou 2, suggère que la série S1A est une filière attirant un public très très limité.

#### Évolution du taux de réussite :

En revanche, les performances de la série S1A en termes de taux de réussite sont remarquables. Malgré un nombre très restreint de candidats, le taux de réussite est de 100% pour la grande majorité des années entre 2013 et 2024. La seule exception notable est l'année 2019, où le taux chute à 66.7%. Cette régularité peut s'expliquer par une sélection rigoureuse des candidats.

En somme, la série S1A se distingue par un effectif très réduit, mais affiche une excellence constante en matière de réussite.

**Remarque importante :** Les séries S1-AR (Mathématiques et Sciences physiques) et S2-AR (Sciences expérimentales), introduites par le décret de 2013 pour le baccalauréat arabe, n'ont jamais été concrètement mises en place ni organisées, comme mentionné dans le Chapitre 1.

### 4.5 prédiction du taux de réussite

### 4.6 Conclusion

## Chapitre 5

# Analyse du parcours universitaire des bacheliers (UCAD)

### 5.1 Introduction

Ce chapitre s'intéresse à la trajectoire des bacheliers, une fois admis à l'université, en particulier à l'UCAD. L'objectif est d'évaluer l'impact des différentes réformes du baccalauréat sur l'orientation et la réussite universitaire des étudiants issus des séries concernées.

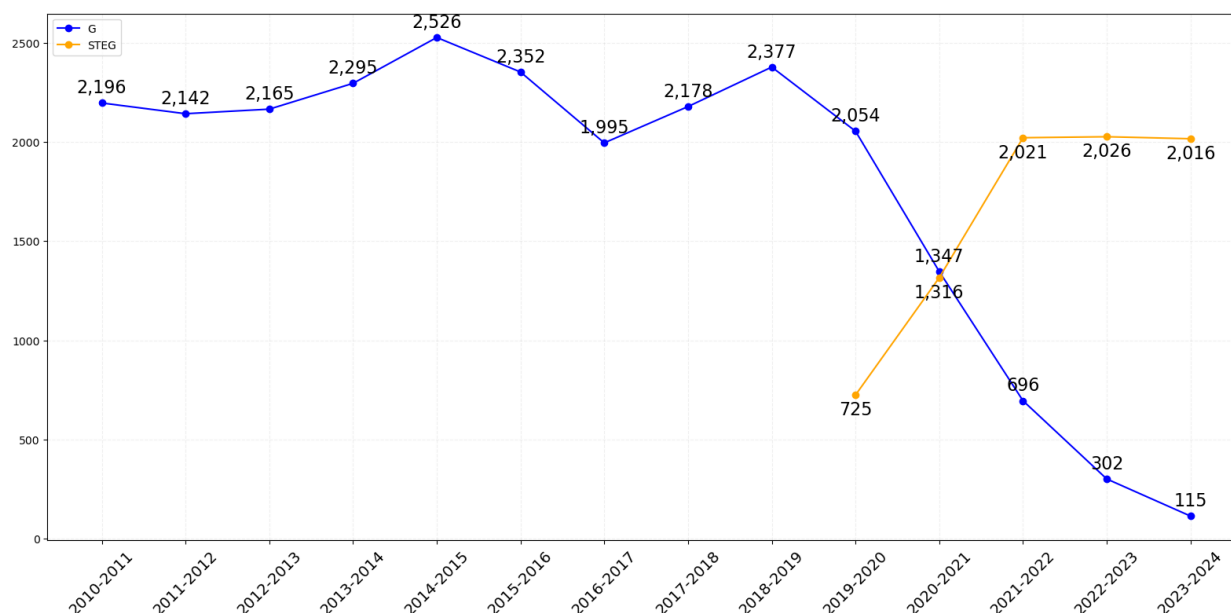
Dans un premier temps, nous analyserons l'évolution des effectifs inscrits à l'UCAD selon les séries de baccalauréat impactées par les réformes. Nous observerons ensuite les établissements et départements universitaires où ces bacheliers sont le plus souvent orientés, afin d'identifier les filières de destination privilégiées selon la série d'origine.

Enfin, une analyse de type **suivi de cohorte** permettra d'évaluer la progression et les performances académiques de ces étudiants dans le temps (réussite, redoublement, abandon), afin de mieux comprendre les effets des réformes sur la réussite universitaire.

## 5.2 Évolution des inscriptions à l'UCAD

### 5.2.1 Les inscrits des série STEG et G

FIGURE 5.1 – Évolution des inscriptions à l'UCAD pour les séries STEG et G



La figure (Figure 5.1) illustre l'évolution des inscriptions à l'Université Cheikh Anta Diop (UCAD) pour les bacheliers issus des séries G et STEG, sur la période 2010-2011 à 2023-2024.

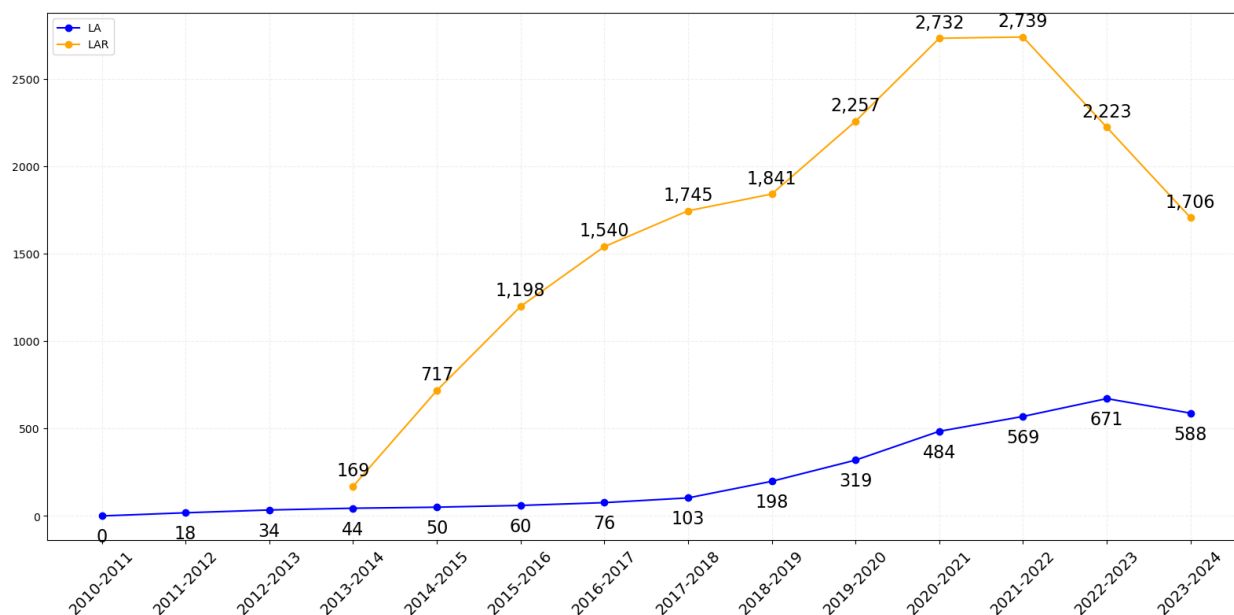
Jusqu'à l'année universitaire 2018-2019, seule la série G est représentée à l'UCAD, avec un nombre d'inscrits fluctuant autour de 2 000 à 2 500. Un pic est observé en 2014-2015 avec 2 526 inscrits. À partir de 2019-2020, la série STEG fait son apparition, avec 725 inscrits, marquant le début de la transition. Simultanément, les effectifs de la série G commencent à décliner fortement, passant de 2 054 en 2019-2020 à seulement 115 en 2023-2024. Ce déclin correspond à la suppression progressive de la série G au profit de la série STEG dans le système du baccalauréat.

Le croisement des courbes est particulièrement visible en 2020-2021, où le nombre d'inscrits en STEG (1 347) dépasse celui de la série G (1 316). Cette tendance se confirme les années suivantes, la série STEG affichant des effectifs croissants (2 021 en 2021-2022, 2 026 en 2022-2023, et 2 016 en 2023-2024), tandis que la série G continue sa chute. La série STEG maintient ainsi un volume d'inscriptions à l'UCAD comparable à celui que la série G connaissait avant sa suppression, démontrant une transition quantitativement réussie au niveau de l'entrée à l'université. Ce transfert des effectifs de la série G vers la série STEG à l'UCAD est un indicateur de l'efficacité de la réforme du baccalauréat dans l'orientation des étudiants vers la nouvelle filière.

## 5.2.2 Les inscrits des séries Arabes et Franco-Arabes

### Série LA et LAR

FIGURE 5.2 – Évolution des inscriptions à l'UCAD pour les séries LA et LAR



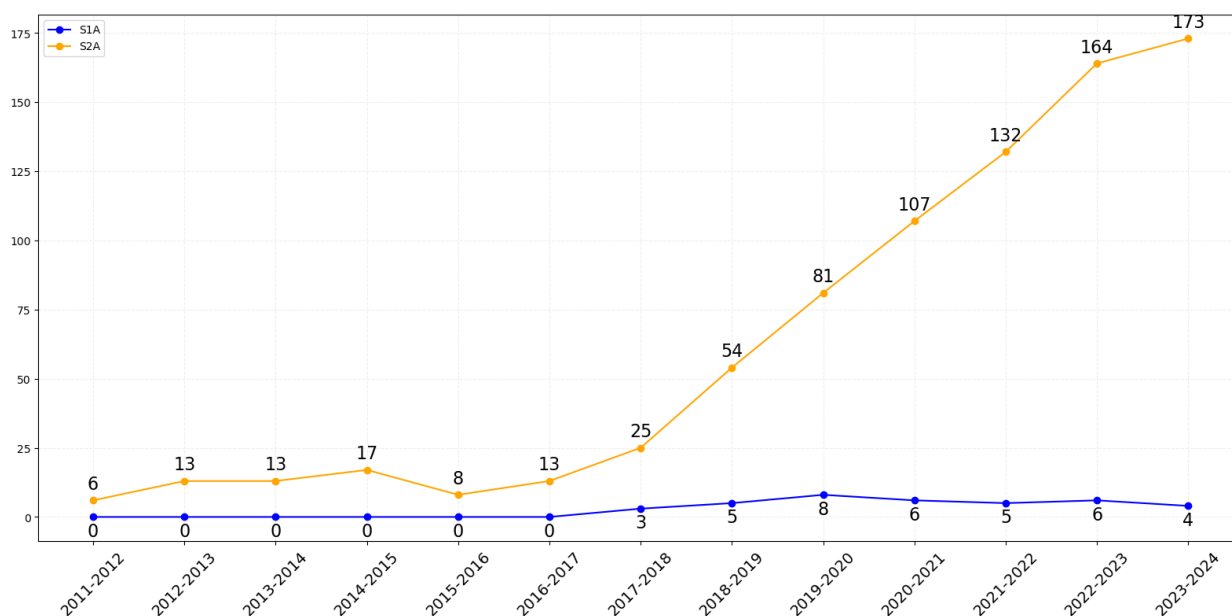
La figure (Figure 5.2) présente l'évolution des inscriptions à l'UCAD pour les bacheliers issus des séries Littératures Arabes (LA) et Littératures et Civilisations Arabes (L-AR) de 2010-2011 à 2023-2024.

La série LA, bien que présente depuis 2010-2011, a enregistré un nombre très faible d'inscrits à l'UCAD jusqu'en 2013-2014, oscillant entre 0 et 44. À partir de 2014-2015, le nombre d'inscrits en LA connaît une croissance progressive, passant de 50 à 671 en 2022-2023, avant de redescendre légèrement à 588 en 2023-2024.

La série L-AR, introduite à l'UCAD à partir de 2013-2014, a connu une croissance beaucoup plus rapide et significative. Elle débute avec 169 inscrits en 2013-2014 et grimpe rapidement pour atteindre un pic de 2 739 inscrits en 2021-2022. Bien qu'une légère diminution soit observée les années suivantes, avec 2 223 inscrits en 2022-2023 et 1 706 en 2023-2024, la série L-AR maintient un volume d'inscriptions considérablement plus élevé que la série LA. Ce phénomène s'aligne avec l'observation d'un transfert progressif des effectifs vers la série L-AR au niveau du baccalauréat lui-même, la série L-AR ayant été mise en place pour mieux encadrer et répondre à la demande sociale des filières arabes et franco-arabes.

## Série S2A et S1A

FIGURE 5.3 – Évolution des inscriptions à l'UCAD pour les séries S2A et S1A



La figure (Figure 5.3) retrace l'évolution des inscriptions à l'UCAD pour les bacheliers des séries Sciences fondamentales (S1A) et Sciences appliquées (S2A) entre 2011-2012 et 2023-2024.

La série S1A, bien que présente, affiche un nombre d'inscrits extrêmement faible à l'UCAD tout au long de la période. Les effectifs oscillent entre 0 et 8 (atteint en 2019-2020), avec seulement 4 inscrits en 2023-2024. Cette quasi-absence d'inscriptions à l'université confirme le caractère très marginal de cette filière, qui déjà au niveau du baccalauréat, n'attire qu'un nombre dérisoire de candidats. Cela suggère que la série S1A ne débouche que sur très peu d'orientations universitaires à l'UCAD.

En revanche, la série S2A présente une dynamique d'inscriptions beaucoup plus significative. Après des débuts modestes entre 6 et 17 inscrits de 2011-2012 à 2015-2016, le nombre d'étudiants en S2A à l'UCAD connaît une croissance exponentielle. On passe de 25 inscrits en 2016-2017 à 107 en 2020-2021, pour atteindre un pic de 173 inscrits en 2023-2024. Cette forte augmentation des inscriptions en S2A à l'UCAD reflète une reconnaissance croissante de cette filière scientifique parmi les bacheliers arabes et franco-arabes, leur offrant des perspectives universitaires concrètes.

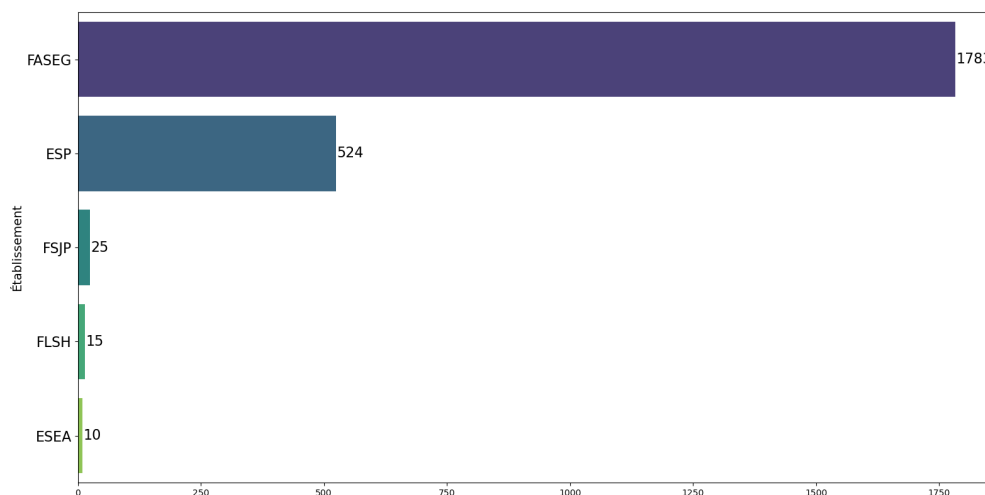
En somme, l'analyse des inscriptions à l'UCAD confirme les dynamiques observées au niveau du baccalauréat : la série S1A reste une filière confidentielle, tandis que la S2A gagne en importance et en attractivité pour les études supérieures, offrant ainsi une voie viable aux bacheliers issus de l'enseignement scientifique arabe et franco-arabe.

## 5.3 Répartition des Inscrits par Établissement et Département

### 5.3.1 Série STEG et G

#### Répartition par Établissements

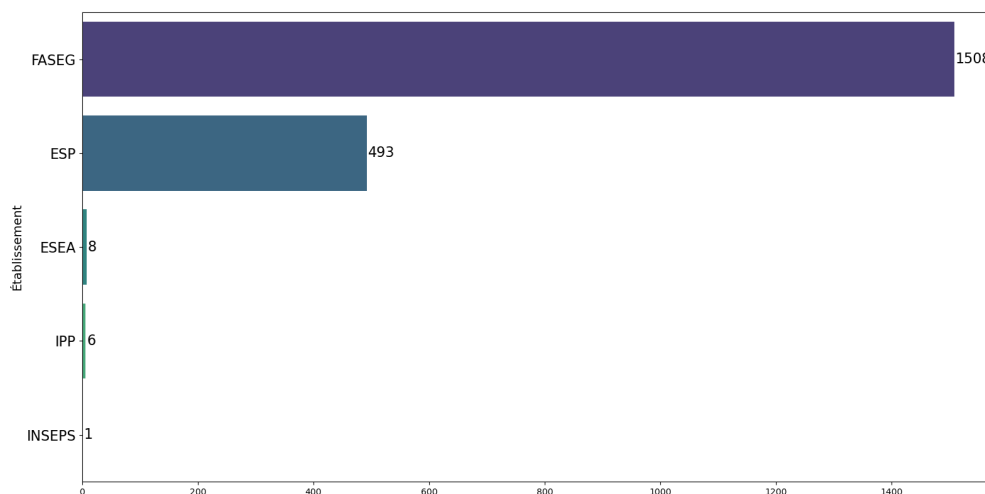
FIGURE 5.4 – Top 5 des établissements avec le plus d’inscrits (G, 2018-2019)



La Figure 5.4 présente la répartition des inscriptions des bacheliers G au sein des établissements de l'UCAD pour l'année universitaire 2018-2019.

Avant la réforme complète, en 2019, la Faculté des Sciences Économiques et de Gestion (FASEG) était déjà l'établissement accueillant la majorité des bacheliers de la série G, avec 1783 inscrits, suivie par l'École Supérieure Polytechnique (ESP) avec 524 inscrits.

FIGURE 5.5 – Top 5 des établissements avec le plus d’inscrits (STEG, 2023-2024)



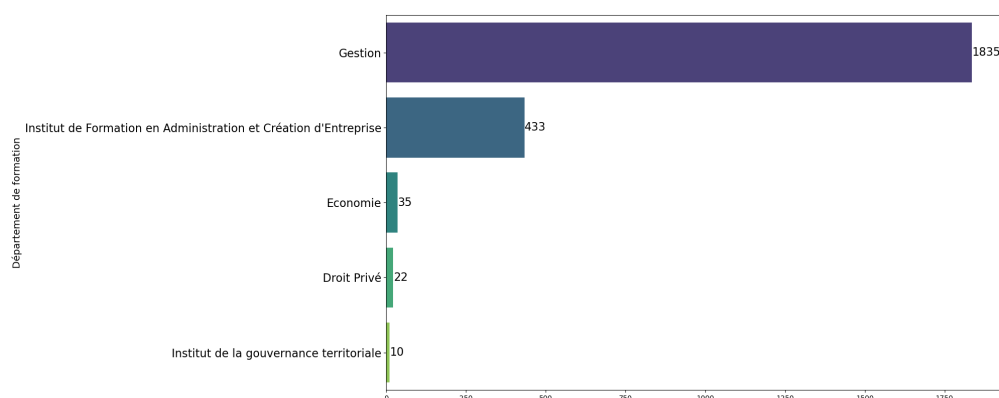
La Figure 5.5 présente la répartition des inscriptions des bacheliers STEG au sein des établissements de l'UCAD pour l'année universitaire 2023-2024.



Pour l'année universitaire 2023-2024, après la transformation de la série G en STEG, la FASEG continue de dominer très largement les inscriptions des bacheliers STEG à l'UCAD, accueillant 1508 étudiants. L'École Supérieure Polytechnique (ESP) maintient sa deuxième position avec 493 inscrits. Ces chiffres confirment que, malgré le changement de dénomination et d'orientation pédagogique, la FASEG demeure la principale destination pour les bacheliers de cette filière, en raison de la nature économique et de gestion de la série STEG.

### Répartition par Départements

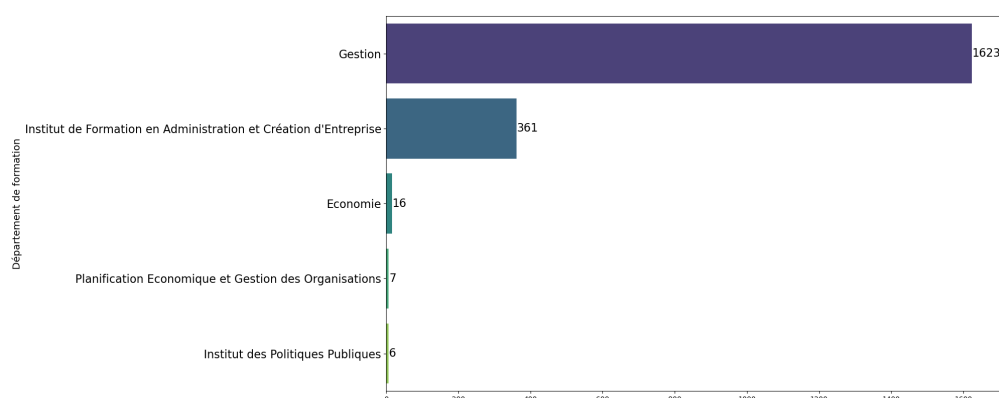
FIGURE 5.6 – Top 5 des départements avec le plus d'inscrits (G, 2018-2019)



La Figure 5.6 illustre la répartition des inscrits G par département de formation à l'UCAD pour l'année universitaire 2018-2019.

En 2019, le département de Gestion concentrait la grande majorité des inscrits de la série G, avec 1835 étudiants, suivi par l'Institut de Formation en Administration et Création d'Entreprise avec 433 inscrits. Le département d'Économie comptait 35 inscrits.

FIGURE 5.7 – Top 5 des départements avec le plus d'inscrits (STEG, 2023-2024)



La Figure 5.7 illustre la répartition des inscrits STEG par département de formation à l'UCAD pour l'année universitaire 2023-2024.

Pour l'année universitaire 2023-2024, le département de Gestion conserve sa position dominante

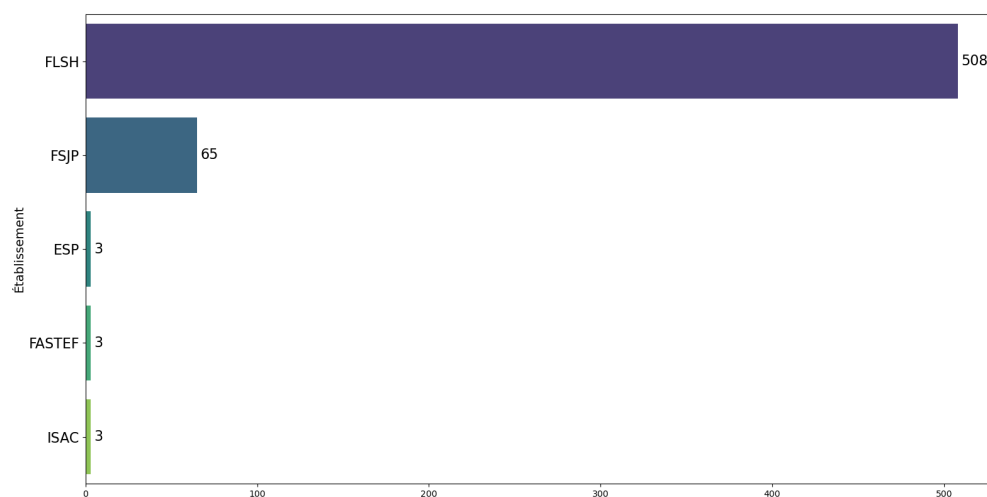
pour les inscrits de la série STEG, avec 1623 inscrits. L’Institut de Formation en Administration et Création d’Entreprise suit avec 361 inscrits. Le département d’Économie compte 16 inscrits. Cette répartition des inscrits par département, avant et après la réforme, confirme la forte vocation de cette série pour les études en gestion, avec un intérêt secondaire pour l’entrepreneuriat, et une présence marginale dans les autres branches de l’économie.

### 5.3.2 Série Arabes et Franco-Arabes

#### Série LA

##### Établissements

FIGURE 5.8 – Top 5 des établissements avec le plus d’inscrits (LA, 2023-2024)



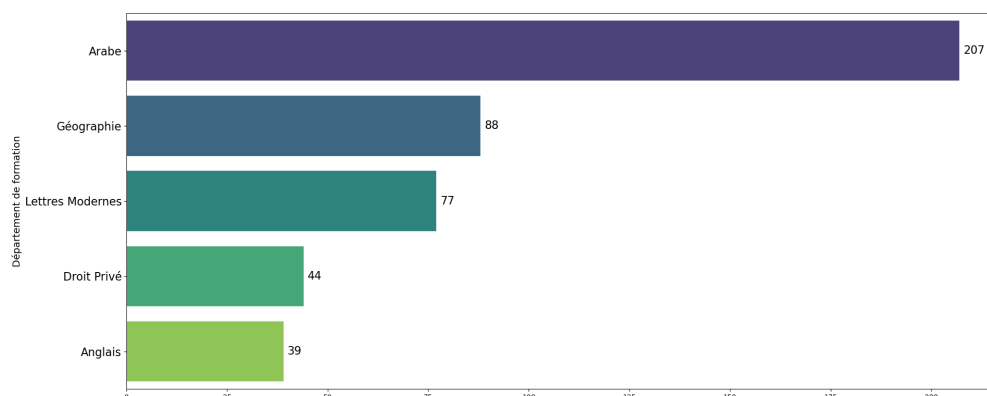
La Figure 5.8 montre la répartition des inscrits de la série LA par établissement à l’UCAD pour l’année universitaire 2023-2024.

La Faculté des Lettres et Sciences Humaines (FLSH) est l’établissement qui accueille de loin le plus grand nombre de bacheliers LA, avec 508 inscrits. Cette prédominance est entièrement cohérente avec la nature littéraire de la série LA, la FLSH étant traditionnellement le pôle d’excellence pour les études de lettres, de langues et de civilisations.

La Faculté des Sciences Juridiques et Politiques (FSJP) arrive en deuxième position avec 65 inscrits. Bien que significativement moins importante que la FLSH, la présence de bacheliers LA dans cette faculté peut s’expliquer par un intérêt pour le droit islamique.

## Départements

FIGURE 5.9 – Top 5 des départements avec le plus d’inscrits (LA, 2023-2024)



La Figure 5.9 présente la répartition des inscrits de la série LA par département de formation à l’UCAD pour l’année universitaire 2023-2024.

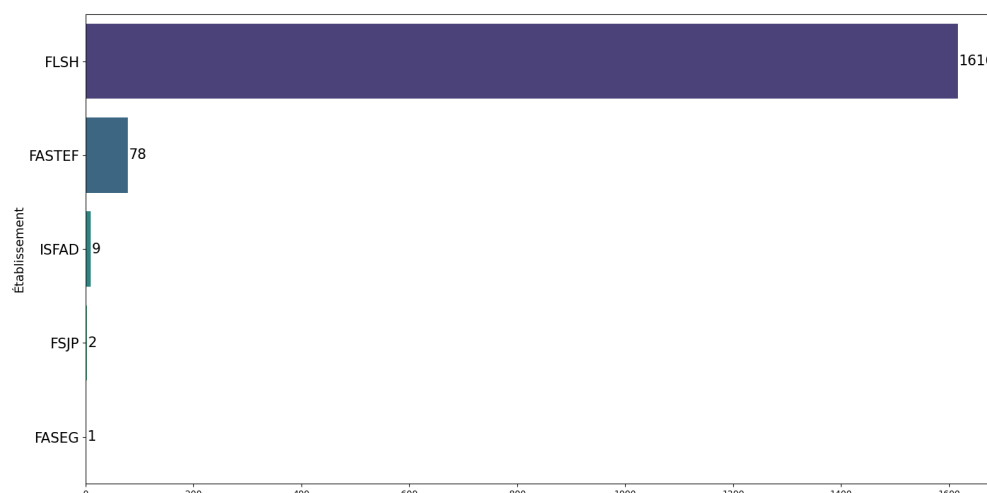
Le département d’Arabe domine avec 207 bacheliers LA, ce qui est attendu pour cette série littéraire axée sur la langue arabe. Cependant, d’autres départements comme la Géographie (88), les Lettres Modernes (77), le Droit Privé (44) et l’Anglais (39) attirent aussi des bacheliers LA.

Cette dispersion s’explique par le caractère franco-arabe de la série LA, qui offre plus de flexibilité que la série LAR, permettant aux étudiants de s’orienter vers un éventail plus large de filières en sciences humaines.

## Série LAR

### Établissements

FIGURE 5.10 – Top 5 des établissements avec le plus d’inscrits (LAR, 2023-2024)



La Figure 5.10 illustre la répartition des inscrits de la série LAR par établissement à l'UCAD pour l'année universitaire 2023-2024.

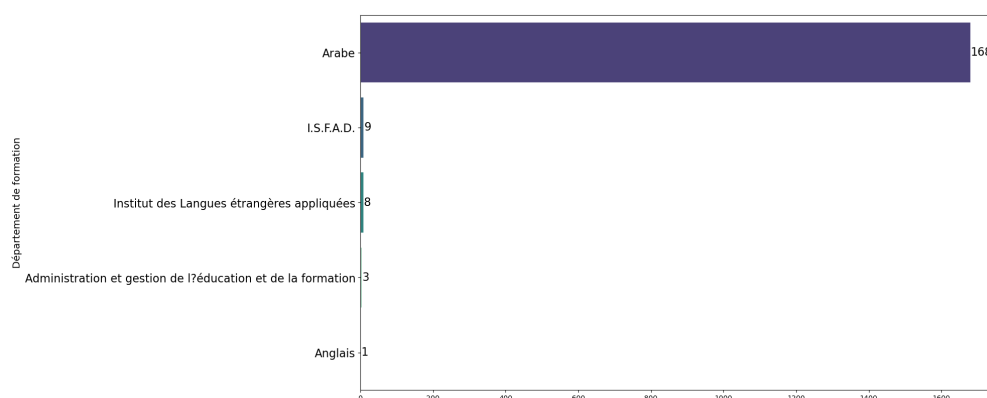
Comme pour la série LA, la Faculté des Lettres et Sciences Humaines (FLSH) est l'établissement le plus plébiscité par les bacheliers LAR, avec un total massif de 1 616 inscrits. Cette écrasante majorité confirme le rôle central de la FLSH dans la formation en langue, littérature et civilisation arabes, faisant de cette faculté la destination naturelle et presque exclusive des bacheliers issus de cette série.

En deuxième position, la FASTEF accueille 78 inscrits. Cette orientation s'explique par les débouchés dans l'enseignement, en particulier pour les matières liées à la langue arabe ou à l'éducation islamique, domaines en cohérence avec la formation reçue dans la série LAR.

À l'inverse, la Faculté des Sciences Juridiques et Politiques (FSJP) ne compte que 2 inscrits LAR, contre 65 pour la série LA. Cette faible représentation s'explique très probablement par une contrainte linguistique importante : les enseignements à la FSJP se déroulant majoritairement en français, les bacheliers LAR, dont la formation est exclusivement en arabe, peuvent être freinés dans leur orientation vers des filières francophones, contrairement aux bacheliers LA disposant d'une double compétence linguistique.

### Départements

FIGURE 5.11 – Top 5 des départements avec le plus d'inscrits (LAR, 2023-2024)



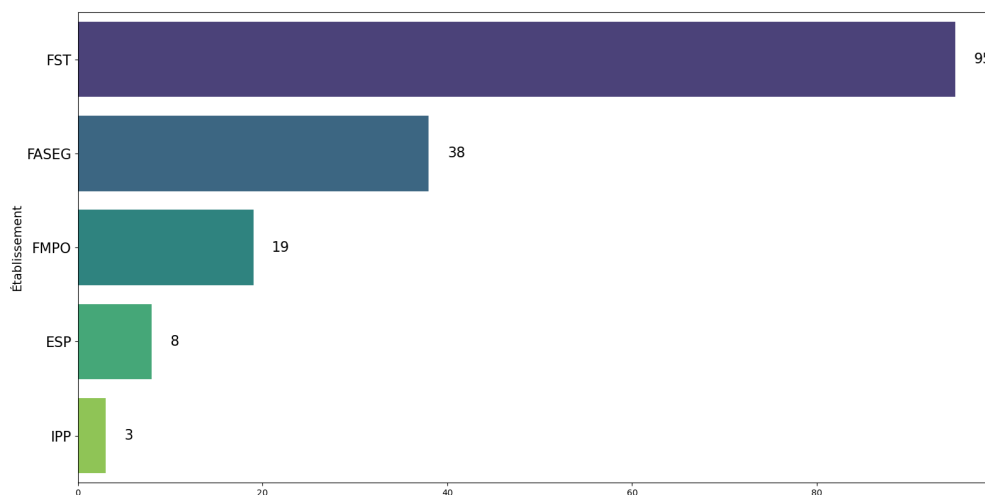
Le département d'Arabe domine de manière écrasante avec 1 680 inscrits, concentrant ainsi presque la totalité des bacheliers LAR. Cette prépondérance est tout à fait attendue, la série LAR (Littératures et Civilisations Arabes) étant spécifiquement conçue pour préparer les étudiants à des études approfondies en langue, littérature et culture arabes. Le département d'Arabe représente donc la destination la plus logique, naturelle et cohérente pour ces profils.

En conclusion, l'orientation des bacheliers LAR est quasiment exclusive vers le département d'Arabe, traduisant une spécialisation universitaire marquée, en parfaite continuité avec leur formation secondaire.

## Série S2A

## Établissements

FIGURE 5.12 – Top 5 des établissements avec le plus d'inscrits (S2A, 2023-2024)

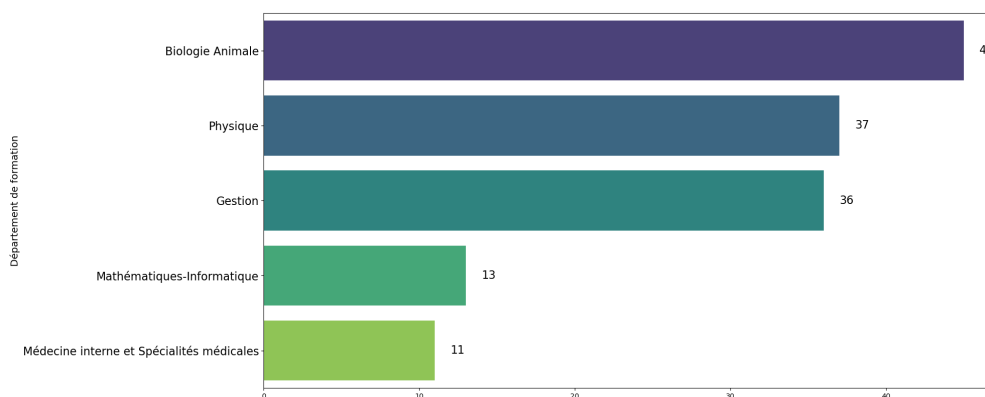


La Figure 5.12 présente la répartition des inscrits de la série S2A par établissement à l'UCAD pour l'année universitaire 2023-2024.

La FST domine avec 95 inscrits, reflet de l'orientation scientifique de la série S2A. D'autres établissements accueillent aussi ces étudiants, notamment la FASEG (38), la FMPO (19), ainsi que plus modestement l'ESP et l'IPP. Cette répartition illustre la polyvalence et l'adaptabilité des diplômés S2A, qui s'intègrent dans des filières variées allant des sciences à l'économie, la santé ou l'ingénierie.

## Départements

FIGURE 5.13 – Top 5 des départements avec le plus d'inscrits (S2A, 2023-2024)

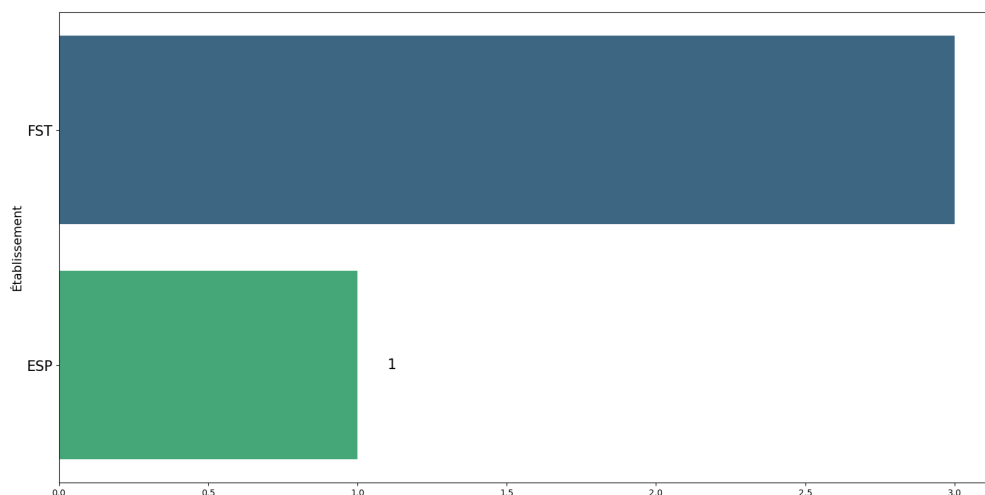


La Figure 5.13 détaille la répartition des inscrits de la série S2A par département de formation à l'UCAD pour l'année universitaire 2023-2024.

Le département de Biologie Animale accueille le plus grand nombre de bacheliers S2A (45 inscrits), reflétant leur intérêt pour les sciences du vivant. Les départements de Mathématiques-Informatique (13) et de Médecine (11) enregistrent des effectifs plus modestes, montrant une ouverture vers des filières techniques et médicales. Cette diversité souligne la capacité des bacheliers S2A à s'intégrer dans des parcours variés grâce à leur formation polyvalente.

**Série S1A****Établissements**

FIGURE 5.14 – Top 5 des établissements avec le plus d’inscrits (S1A, 2023-2024)



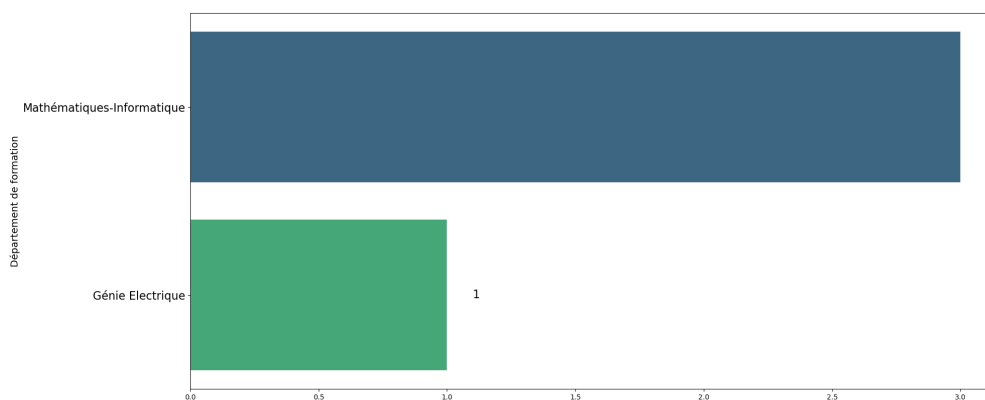
La Figure 5.14 présente la répartition des inscrits de la série S1A par établissement à l’UCAD pour l’année universitaire 2023-2024.

La Faculté des Sciences et Technologies (FST) est la principale destination avec 3 inscrits, ce qui correspond à la nature fondamentale de la série S1A. L’École Supérieure Polytechnique (ESP) accueille un seul inscrit, reflétant un intérêt marginal pour les formations techniques.

Cependant, le très faible nombre d’inscrits dans ces établissements souligne la rareté de cette série à l’UCAD. Ces données limitées ne permettent pas de tirer de conclusions robustes sur les choix d’orientation des bacheliers S1A.

**Départements**

FIGURE 5.15 – Top 5 des départements avec le plus d’inscrits (S1A, 2023-2024)



La Figure 5.15 illustre la répartition des inscrits de la série S1A par département de formation à l'UCAD pour l'année universitaire 2023-2024.

Le département de Mathématiques-Informatique compte 3 inscrits, ce qui correspond bien au profil scientifique fondamental de la série S1A. Le Génie Électrique accueille 1 étudiant, reflétant un intérêt pour des applications techniques. Ces effectifs limités confirment la faible présence des bacheliers S1A à l'UCAD, qui restent néanmoins orientés vers des filières scientifiques et d'ingénierie spécialisées.

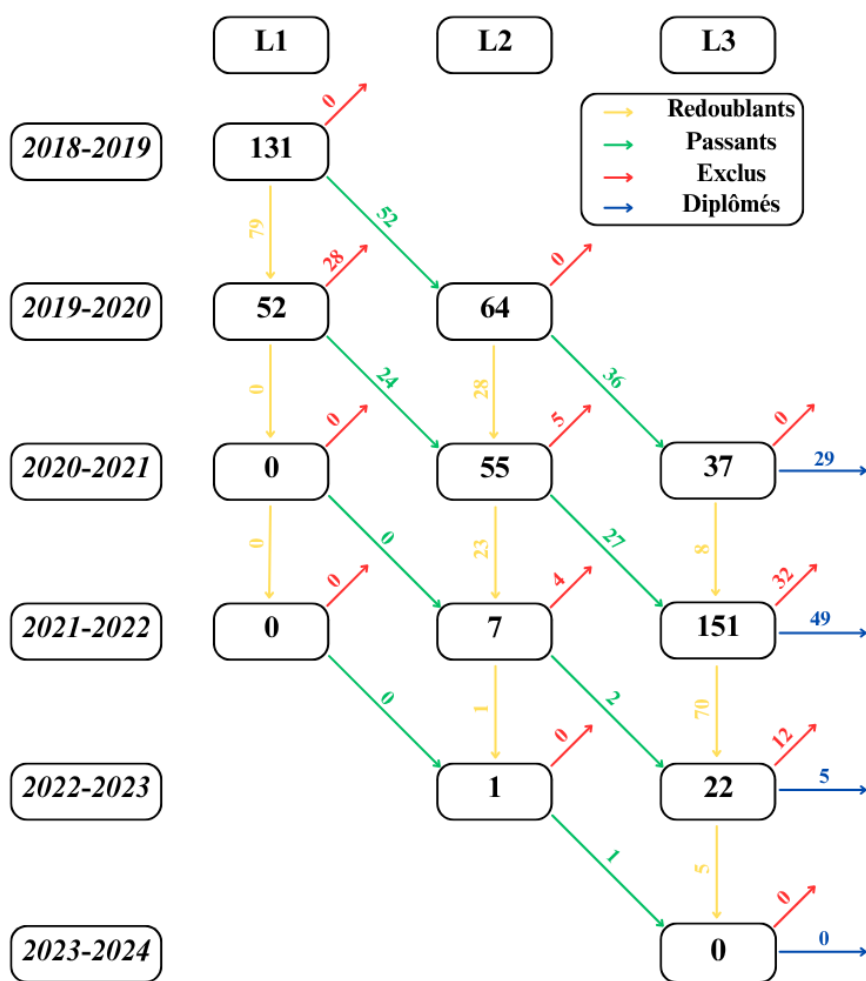


## 5.4 Analyse du parcours universitaire des bacheliers (suivi des cohortes)

### 5.4.1 Série Arabes et Franco-Arabes

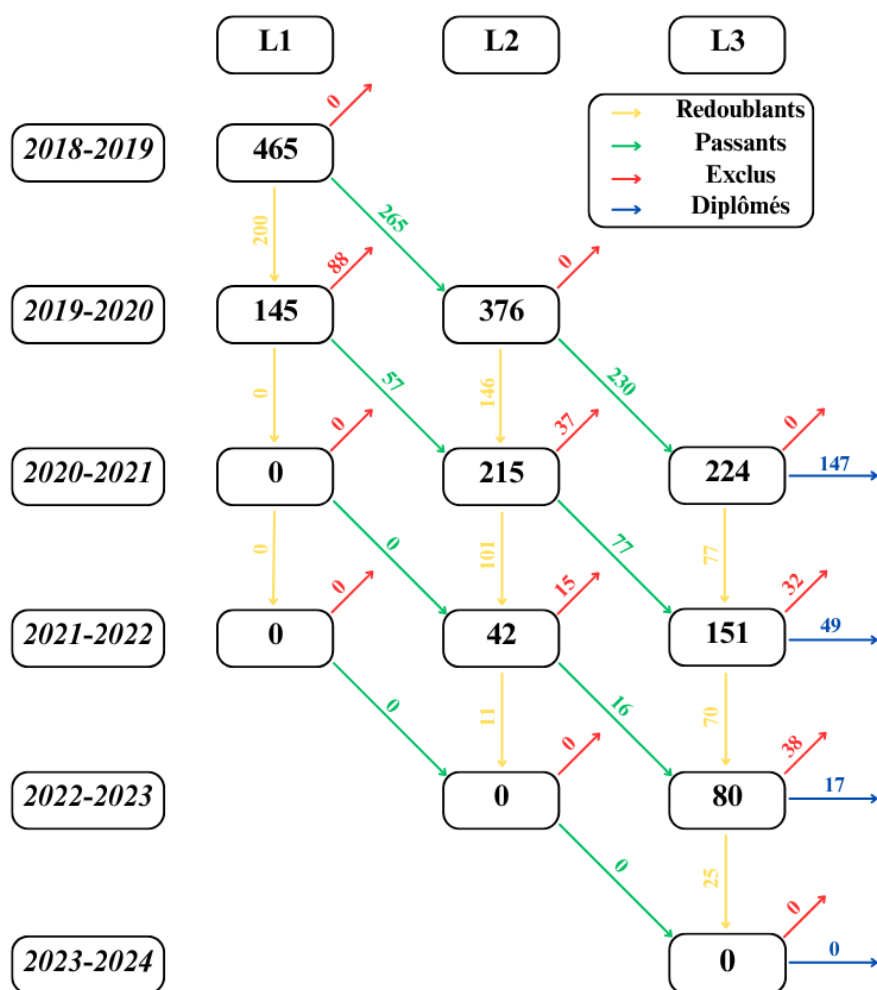
cohorte 2018(LA)

FIGURE 5.16 – Graphe suivi de la cohorte 2018 (LA)



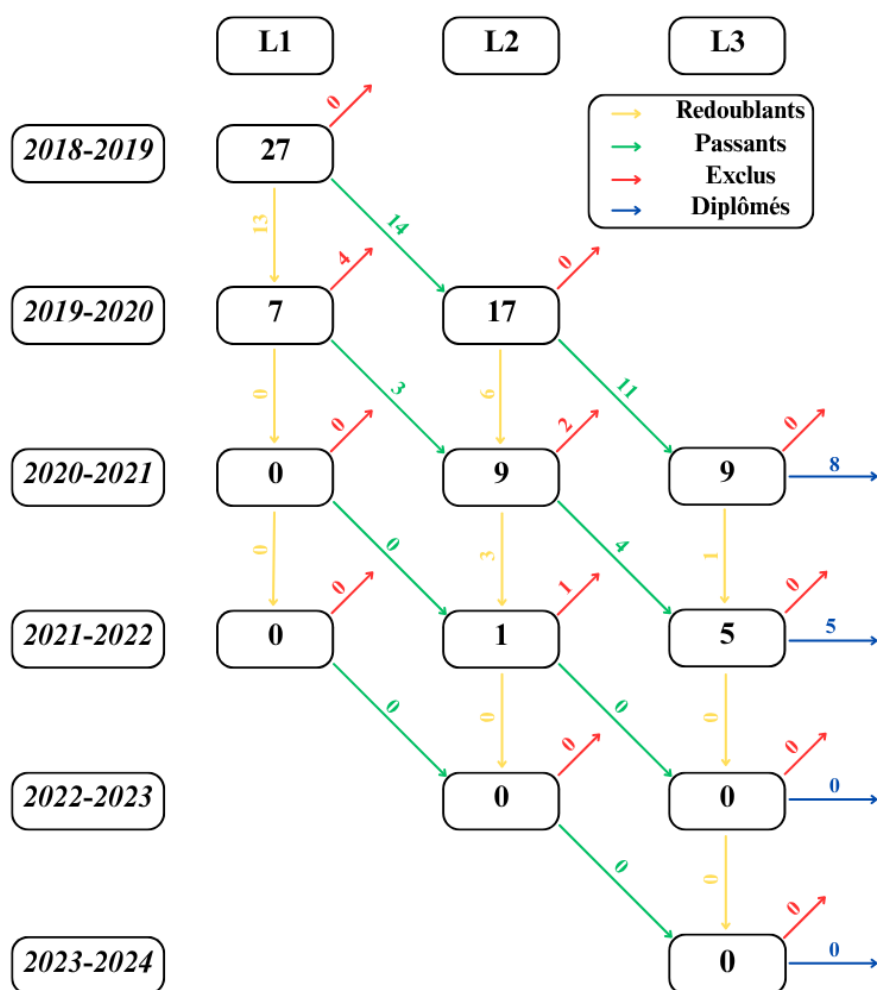
cohorte 2018(LAR)

FIGURE 5.17 – Graphe suivi de la cohorte 2018 (LAR)



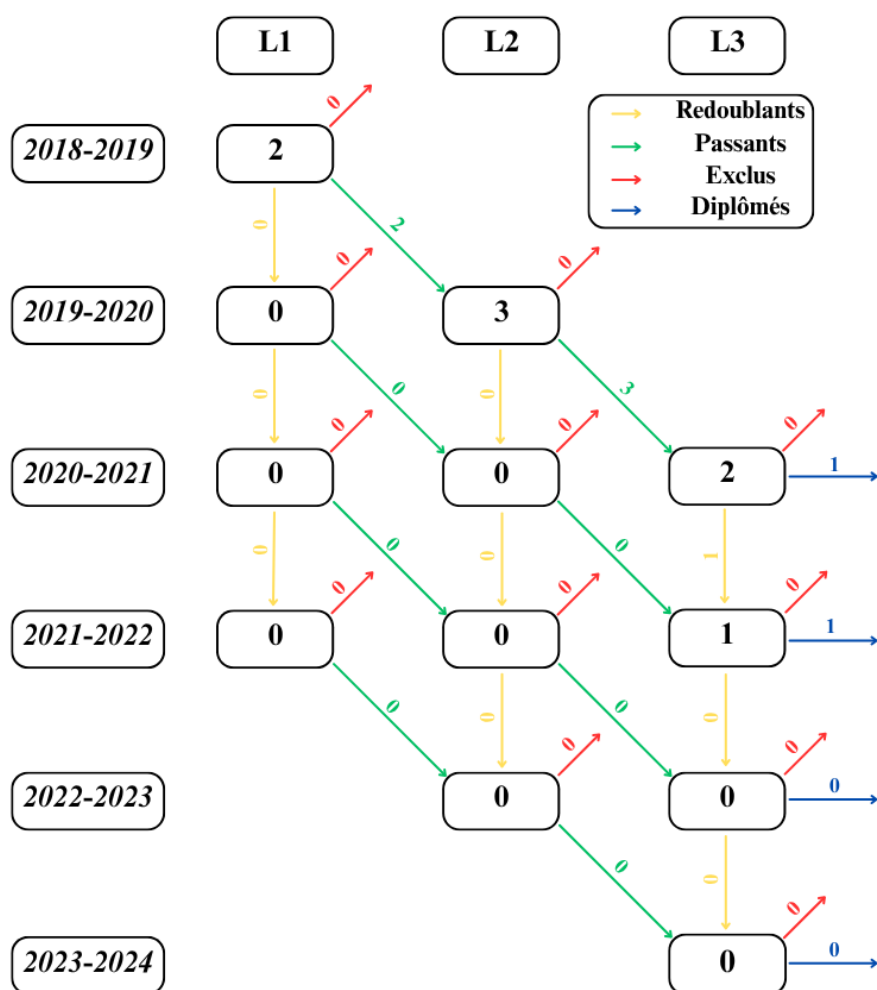
cohorte 2018(S2A)

FIGURE 5.18 – Graphe suivi de la cohorte 2018 (S2A)



cohorte 2018(S1A)

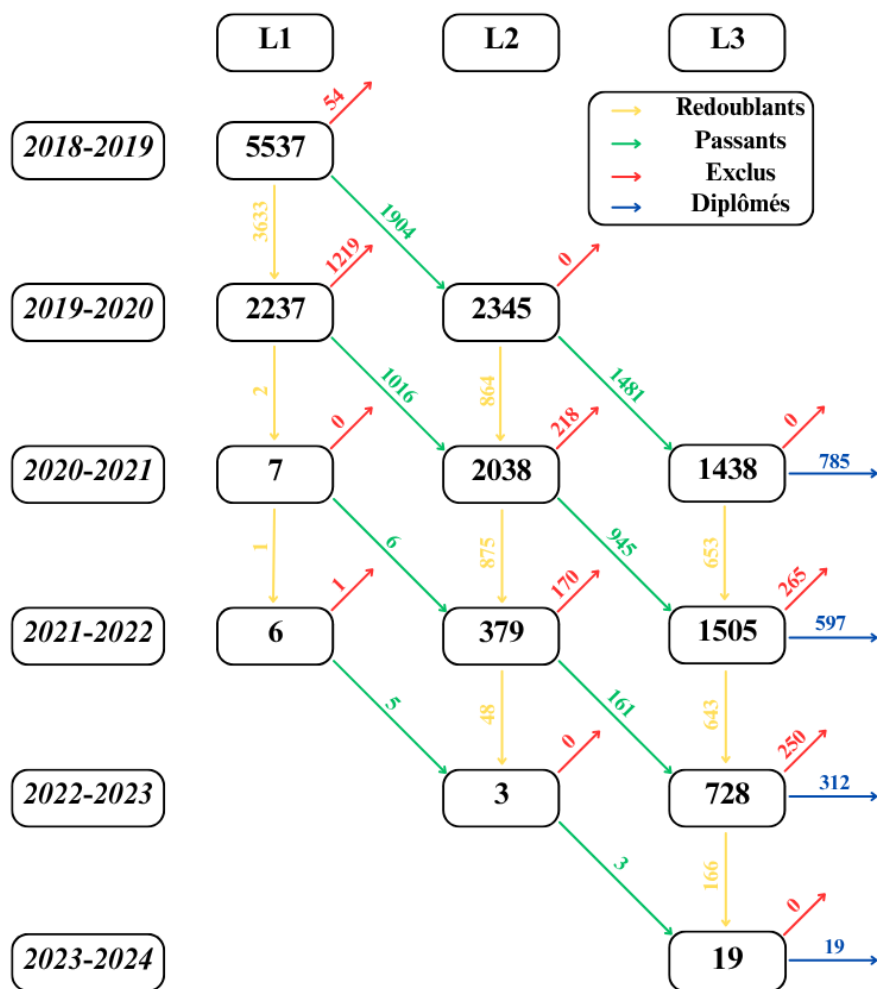
FIGURE 5.19 – Graphe suivi de la cohorte 2018 (S1A)



### 5.4.2 Série de reference pour les séries Arabes et Franco-Arabes

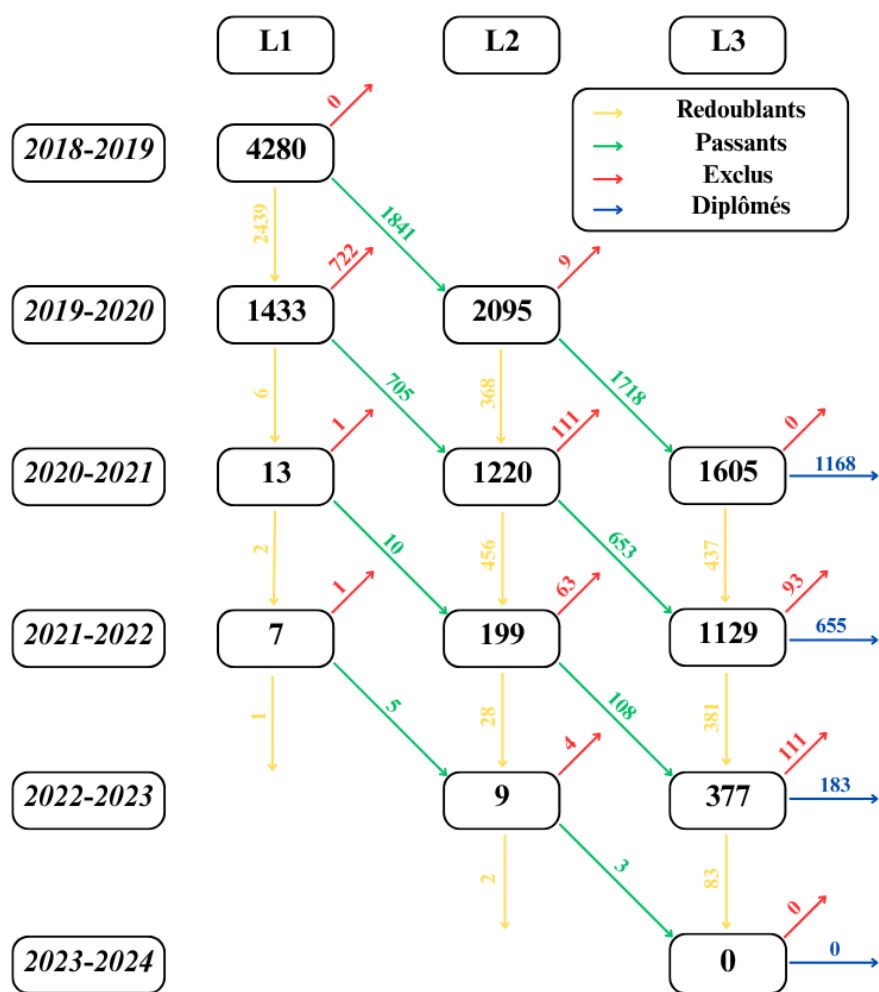
cohorte 2018(L'1)

FIGURE 5.20 – Graphe suivi de la cohorte 2018 (L'1)



cohorte 2018(S2)

FIGURE 5.21 – Graphe suivi de la cohorte 2018 (S2)



**Legend:**

- Yellow arrow: Redoublants
- Green arrow: Passants
- Red arrow: Exclues
- Blue arrow: Diplômés

Année	L1	L2	L3
2018-2019	187		
2019-2020	27	132	
2020-2021	1	31	97
2021-2022	1	0	46
2022-2023		0	4
2023-2024			0

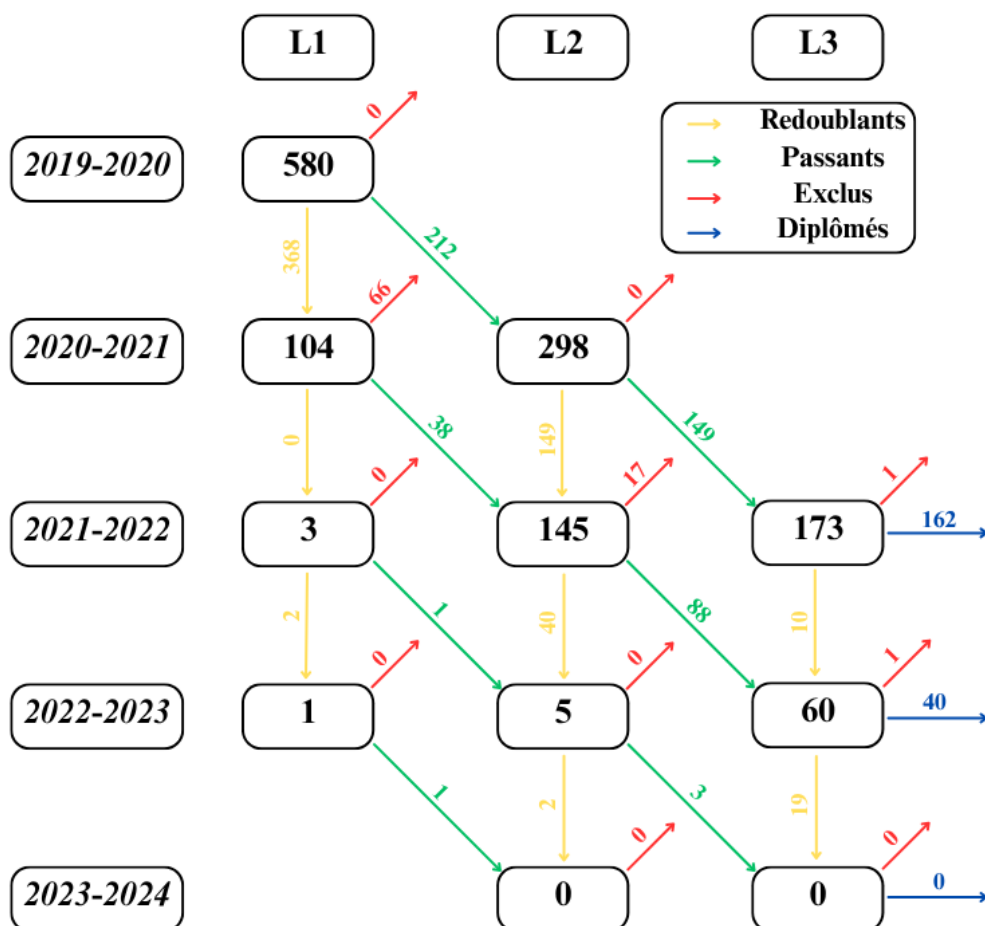
**Transitions (Arrows):**

- L1 to L2:**
  - 2018-2019 to 2019-2020: 58 (Yellow), 11 (Red)
  - 2019-2020 to 2020-2021: 0 (Yellow), 1 (Red)
  - 2020-2021 to 2021-2022: 0 (Yellow), 1 (Red)
  - 2021-2022 to 2022-2023: 0 (Yellow), 0 (Green)
- L2 to L3:**
  - 2018-2019 to 2019-2020: 129 (Green)
  - 2019-2020 to 2020-2021: 25 (Yellow), 5 (Red)
  - 2020-2021 to 2021-2022: 10 (Yellow), 0 (Red)
  - 2021-2022 to 2022-2023: 0 (Yellow), 0 (Green)
  - 2022-2023 to 2023-2024: 0 (Yellow), 0 (Green)
- L3 to Exit:**
  - 2020-2021 to 2021-2022: 61 (Blue)
  - 2021-2022 to 2022-2023: 24 (Blue)
  - 2022-2023 to 2023-2024: 1 (Blue)
  - 2023-2024 to Exit: 0 (Blue)

## 5.4.3 Série STEG et G

cohorte 2019(STEG)

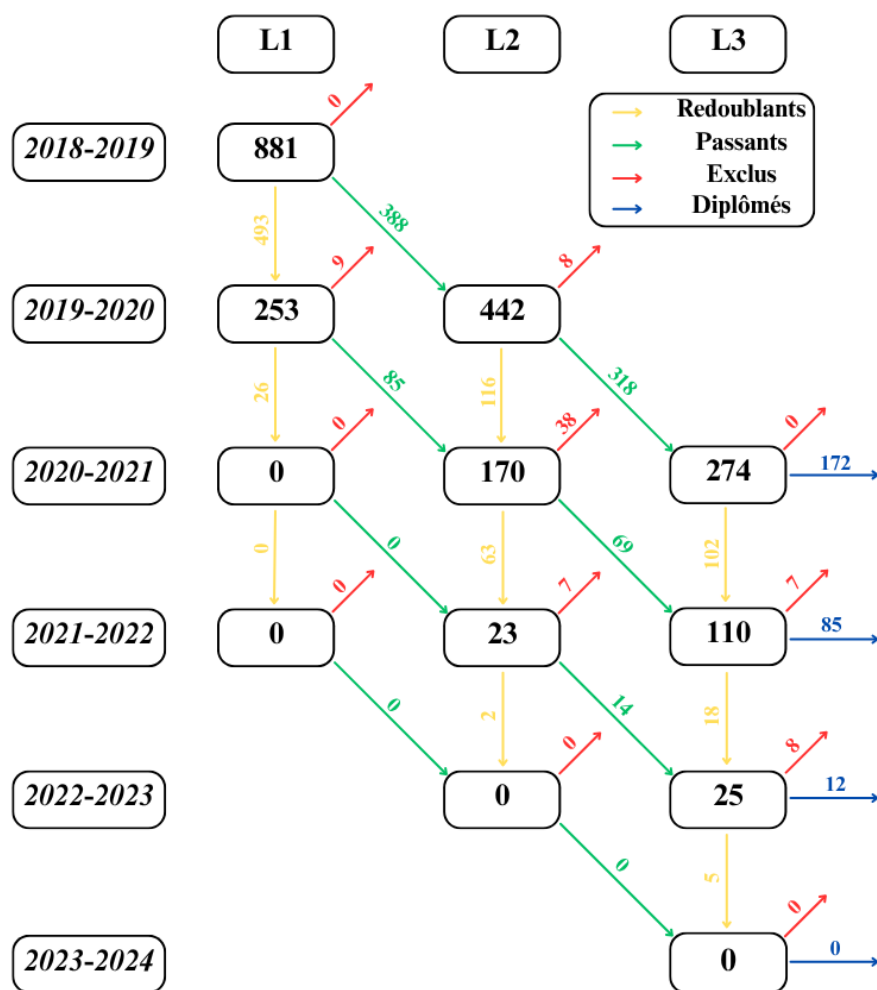
FIGURE 5.23 – Graphe suivi de la cohorte 2019 (STEG)





cohorte 2018(G)

FIGURE 5.24 – Graphe suivi de la cohorte 2018 (G)



## **5.5 Analyse des performances académiques**

## **5.6 Conclusion**

# Chapitre 6

## Restitution interactive et visualisation

### 6.1 Introduction

Ce chapitre présente la phase finale de l'analyse : la restitution interactive des résultats à travers un tableau de bord conçu avec Power BI. L'objectif est de permettre une exploration dynamique et intuitive des données du baccalauréat sénégalais de 2006 à 2024, à destination notamment de l'Office du Baccalauréat et de toute personne souhaitant approfondir une étude sur le sujet.

### 6.2 objectifs du dashboard

Le tableau de bord a pour but :

- de synthétiser les indicateurs clés liés au taux de réussite au baccalauréat par année, série et session ;
- d'observer l'évolution de ces indicateurs dans le temps ;
- de permettre un filtrage dynamique selon les besoins d'analyse ;
- et de soutenir la prise de décision à travers des visualisations claires et interactives.

### 6.3 Présentation des indicateurs suivis

Les indicateurs principaux intégrés dans le dashboard sont :

- Le taux de réussite global par année ;

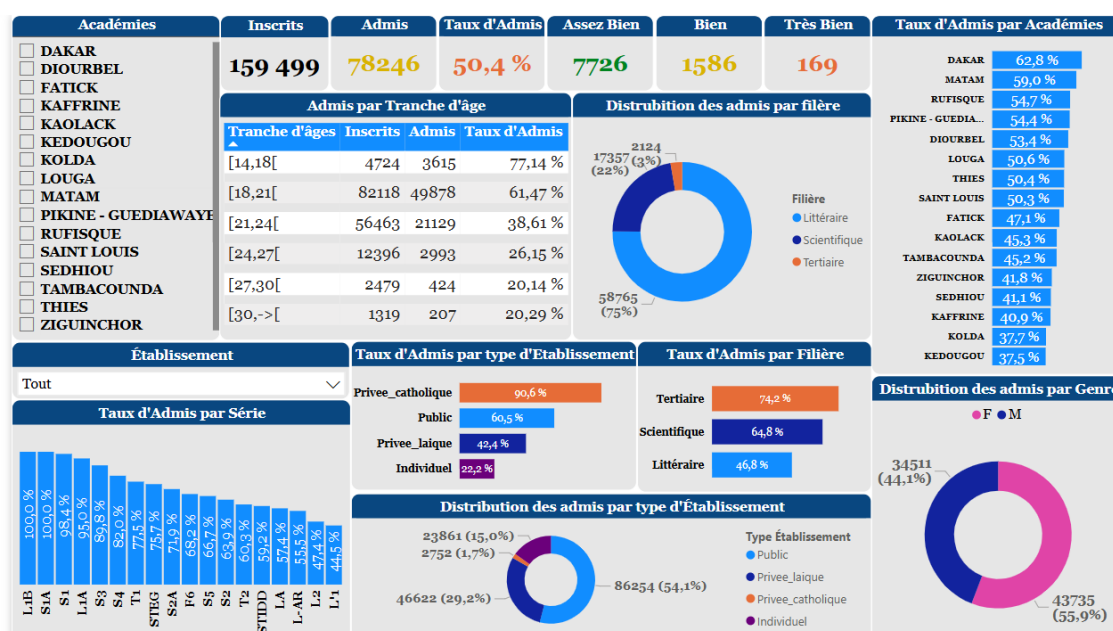
- Le taux de réussite par série ;
- Le nombre total d'admis et d'inscrits ;
- La répartition par session (1er ou 2e tour) ;
- La répartition par mention obtenue.

## 6.4 Construction du dashboard Power BI

### 6.4.1 Première version : Résultats de 2024 uniquement

Dans un premier temps, n'ayant pas encore accès à l'ensemble des données historiques, j'ai construit une première version du tableau de bord en me basant uniquement sur les résultats de l'année 2024. Cette version m'a permis de tester la structure du dashboard et de définir les visualisations pertinentes à suivre. Elle comportait déjà tous les indicateurs énumérés précédemment.

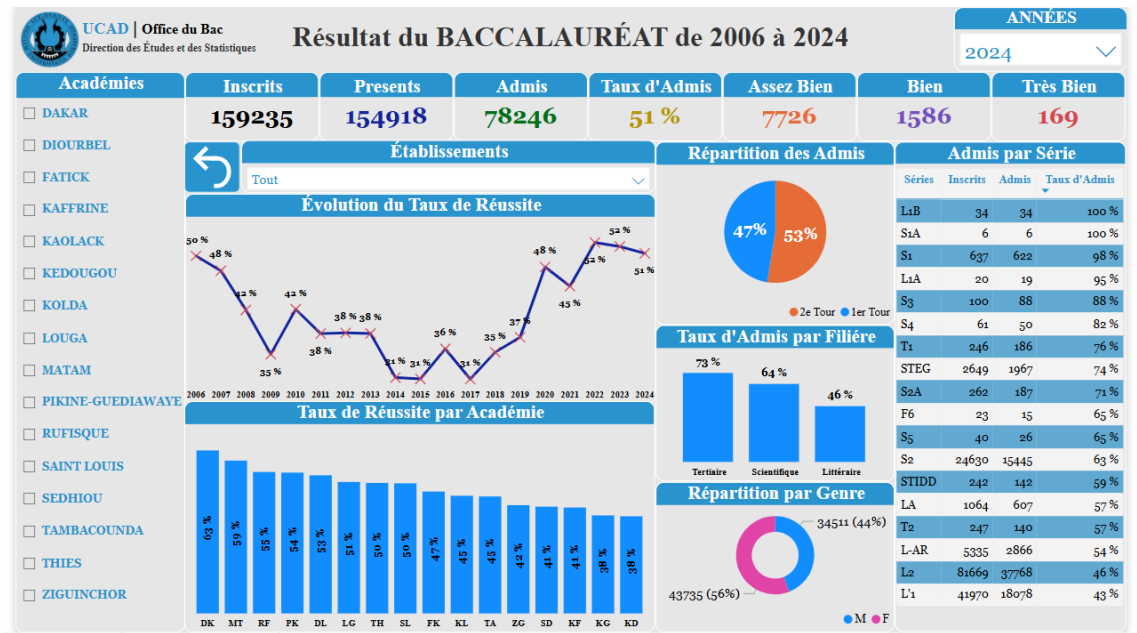
FIGURE 6.1 – Tableau de bord Power BI - Résultats du baccalauréat 2024



6.4.2 Deuxième version : Résultats de 2006 à 2024

Une fois l'accès aux données complètes obtenu, j'ai pu enrichir le dashboard avec les résultats du bac de 2006 à 2024. Cette version permet d'analyser l'évolution temporelle des indicateurs clés et de mieux comprendre l'impact des différentes réformes au fil des années.

FIGURE 6.2 – Tableau de bord Power BI - Résultats du baccalauréat 2006 à 2024



### 6.4.3 Utilisation de Power Query et création de mesures

Power Query a été utilisé pour nettoyer et transformer les données importées dans Power BI. J'y ai notamment créé des colonnes conditionnelles pour catégoriser les mentions ou les sessions, et j'ai défini plusieurs mesures DAX (Data Analysis Expressions) telles que le taux de réussite, le nombre total d'inscrits ou encore les ratios d'admis par série.

## 6.5 Conclusion

La restitution interactive via Power BI offre une approche moderne et intuitive de l'analyse des données du baccalauréat. Elle facilite l'interprétation des tendances et appuie la prise de décision à partir d'une lecture visuelle et dynamique des résultats. Ce tableau de bord constitue un apport concret à l'Office du Baccalauréat, pouvant soutenir les futures études sur l'évolution du système éducatif au Sénégal.

## **Conclusion Générale et Recommandations**



# Bibliographie

- [1] Mbaye DIAGNE. « ÉVOLUTION DU TAUX DE RÉUSSITE AU BACCALAURÉAT AU SÉNÉGAL DURANT CES 20 DERNIÈRES ANNÉES ET LES FACTEURS AGISSANTS ». In : (2023). URL : [https://djiboul.org/wp-content/uploads/2023/03/Tire-a-part\\_13.pdf](https://djiboul.org/wp-content/uploads/2023/03/Tire-a-part_13.pdf).
- [2] Direction des ÉTUDES ET DES STATISTIQUES. *Efficacité et efficience internes – Suivi de la cohorte 2013-2014*. Rapp. tech. Rapport interne. Dakar : Université Cheikh Anta Diop de Dakar, 2021.
- [3] LIVREAU SENEGAL. *Origine du baccalaureat*. 2024. URL : <https://www.instagram.com/livresausenegal/p/C8hKo-ksWP3/?hl=fr>.
- [4] République du SÉNÉGAL. *Decret n° 2000-586 modifiant le décret n°95-947 du 18 octobre 1995 portant sur organisation du baccalauréat*. Ministère de l'Éducation nationale. 2000.
- [5] République du SÉNÉGAL. *Decret n° 2013-913 modifiant et complétant certaines disposition du décret n°2000-586 du 20 juillet 2000 modifiant et complétant le décret n°95-947 du 18 octobre 1995 portant sur organisation du baccalauréat*. Ministère de l'Éducation nationale. 2001.
- [6] République du SÉNÉGAL. *Decret n° 2019-645 modifiant le décret n°95-947 du 18 octobre 1995 portant sur organisation du baccalauréat*. Ministère de l'Éducation nationale. 2019.
- [7] République du SÉNÉGAL. *Decret n°1995-947 portant sur organisation du baccalauréat*. Ministère de l'Éducation nationale. 1995. URL : [https://bo.senegal.services.sn/storage/texte\\_references/t-dec-portant-organisation-baccalaureat.pdf#:~:text=Les%20%C3%A9preuves%20du%20baccalaur%C3%A9at%20r%C3%A9gies%20par%20le,organis%C3%A9es%20pour%20la%20derni%C3%A8re%20fois%20en%201998.&text=En%20attendant%20la%20cr%C3%A9ation%20des%20brevets%20de,du%20decret%2087%2D914%20du%2011%20Juillet%201987..](https://bo.senegal.services.sn/storage/texte_references/t-dec-portant-organisation-baccalaureat.pdf#:~:text=Les%20%C3%A9preuves%20du%20baccalaur%C3%A9at%20r%C3%A9gies%20par%20le,organis%C3%A9es%20pour%20la%20derni%C3%A8re%20fois%20en%201998.&text=En%20attendant%20la%20cr%C3%A9ation%20des%20brevets%20de,du%20decret%2087%2D914%20du%2011%20Juillet%201987..)
- [8] *site officiel de Matplotlib*. URL : <https://matplotlib.org/>.
- [9] *site officiel de Pandas*. URL : <https://pandas.pydata.org/>.
- [10] *site officiel de Power BI*. URL : <https://powerbi.microsoft.com/fr-fr/>.

- [11] *site officiel de Python*. URL : <https://www.python.org/>.
- [12] *site officiel de Scikit-learn*. URL : <https://scikit-learn.org/stable/index.html>.
- [13] *site officiel de Statsmodels*. URL : <https://www.statsmodels.org/stable/index.html>.