



UNIVERSITÉ
CHEIKH ANTA DIOP
DE DAKAR

FACULTÉ DES
SCIENCES ET
TECHNIQUES
FST



MÉMOIRE DE RECHERCHE DE II CYCLE

Analyse de l'impact des dernières réformes du baccalauréat au Sénégal sur le taux de réussite au Bac et l'insertion universitaire

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 à 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

Introduction Générale	11
Contexte et justification	11
Problématique et Objectifs de l'étude	12
Méthodologie et Organisation du document	12
1 État de l'art	13
1.1 Introduction	13
1.2 Historique et rôle du baccalauréat	13
1.3 Organisation du baccalauréat : structure et acteurs	14
1.4 Réformes du baccalauréat au Sénégal	14
1.4.1 Réformes de 1995 (décret n° 95-947)	14
1.4.2 Réformes de 2000 (décret n° 2000-585)	15
1.4.3 Réformes de 2013 (décret n° 2013-913)	15
1.4.4 Réformes de 2019(décret n° 2019-645)	16
1.5 Travaux antérieures	16
1.6 Approche méthodologique et innovante de l'étude	16
1.7 Conclusion	17
2 Données et outils utilisés	18
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
3 Nettoyage et préparation des données	23
3.1 Introduction	23
3.2 Préparation de données des résultats du baccalauréat	24

3.2.1	Doublons dans la clé primaire <code>numero_table</code>	25
3.2.2	Les valeurs manquantes dans les colonnes	27
3.2.3	Correction du type de la colonne <code>moy_finale</code>	28
3.2.4	Création des colonnes <code>admis</code> et <code>session</code>	29
3.2.5	creation du <code>dataFrame</code> récapitulatif	29
3.3	Fusion des données des Inscrits et des Résultats de l'UCAD	30
3.3.1	Données des inscriptions à l'UCAD	30
3.3.2	Données des résultats de l'UCAD	31
3.3.3	Fusion des données d'inscription et de résultats	32
3.4	Filtage des données	33
3.4.1	Nationalité	33
3.4.2	Séries du baccalauréat	33
3.4.3	Niveau universitaire	33
3.5	Conclusion	35
4	Analyse du taux de réussite au bac	36
4.1	Introduction	36
4.2	Analyse de globale	36
4.2.1	Évolution du nombre d'inscrits, présents et admis	36
4.3	Conclusion	36
5	Analyse du parcours universitaire des bacheliers (UCAD)	37
5.1	Introduction	37
5.2	Analyse des résultats du baccalauréat	37
5.2.1	Analyse descriptive du taux de réussite	37
5.2.2	Étude comparative avant et après les réformes	37
5.2.3	Modélisation prédictive du taux de réussite	37
5.3	Analyse du parcours universitaire des bacheliers	37
5.4	Conclusion	37
6	Restitution interactive et visualisation	38
6.1	Introduction	38
6.2	objectifs du dashboard	38
6.3	Présentation des indicateurs suivis	39
6.4	Construction du dashboard Power BI	40
6.4.1	Première version : Résultats de 2024 uniquement	40
6.4.2	Deuxième version : Résultats de 2006 à 2024	41
6.4.3	Utilisation de Power Query et création de mesures	42
6.5	Conclusion	42

TABLE DES MATIÈRES

7	Discussion et Recomandation	43
7.1	Introduction	43
7.2	Conclusion	43
	Conclusion Générale	44

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	28
3.4	Création de nouvelles colonnes	29
3.5	Info global du data des inscriptions	30
3.6	Info global du data des résultats	31
3.7	Jointure des données d’inscription et de résultats	32
3.8	Info global du data des inscriptions et résultats	32
3.9	Filtrage des données	35

Table des figures

6.1	Tableau de bord Power BI - Résultats du baccalauréat 2024	40
6.2	Tableau de bord Power BI - Résultats du baccalauréat 2006 à 2024	41

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 .	27
3.3	Valeurs manquantes dans la colonne moy_finale	27

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 l'insertion universitaire des bacheliers, notamment à l'UCAD.

Problématique et Objectifs de l'étude

Les réformes éducatives ont-elles permis de redynamiser le baccalauréat sénégalais et d'améliorer l'accès à l'enseignement supérieur? Malgré l'intention affichée de moderniser le système et de le rendre plus inclusif, plusieurs interrogations subsistent :

- Ces réformes ont-elles eu un impact significatif sur les taux de réussite?
- 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?

C'est à l'analyse de ces questions que s'attelle cette étude. Ce travail vise ainsi à combler un manque crucial d'évaluation quantitative des réformes éducatives, tout en proposant des outils analytiques pour guider les futures politiques publiques. En croisant l'exploitation de données massives (résultats du bac, inscriptions à l'UCAD) et des méthodes avancées de data science, il offre une vision factuelle pour optimiser l'efficacité du système éducatif sénégalais.

Méthodologie et Organisation du document

1. Le premier chapitre présente l'état de l'art sur les réformes du baccalauréat au Sénégal,...
2. Le deuxième chapitre présente les données et outils utilisés, détaillant les outils statistiques et analytiques utilisés pour évaluer l'impact des réformes.
3. Le troisième chapitre traite l'exploration et la préparation des données,...
4. Le quatrième chapitre présente l'analyse statistique des impacts des réformes,...
5. Le cinquième chapitre présente le Dashboard Power Bi de visualisation des résultats au baccalauréat de 2006 à 2024 à fin de faciliter toute nouvelle étude.
6. Enfin, mes recommandations et conclusions sont présentées, soulignant les implications de cette étude pour les politiques éducatives futures.

Chapitre 1

État de l'art

1.1 Introduction

Depuis sa création, comme énoncé précédemment, le baccalauréat a connu de nombreuses évolutions, tant dans sa structure que dans sa finalité. Ce chapitre propose une revue des principales réformes ayant marqué le système du baccalauréat au Sénégal, en mettant en lumière leur contexte, leur portée et leurs objectifs. Il vise également à situer l'organisation de notre étude dans ce cadre évolutif.

1.2 Historique et rôle du baccalauréat

Le baccalauréat a été inventé en France au XIX^e siècle par un décret de l'empereur Napoléon I^{er} en 1808. Au commencement le baccalauréat s'appelait le **Bachot**. Son étymologie "*bacca laurea*" désigne en latin médiéval, **la couronne de laurier** remise aux vainqueurs [2].

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.

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. Il mobilise chaque année plusieurs milliers d'enseignants, surveillants, correcteurs et examinateurs.

Il se décline en différentes séries réparties en trois filières distinctes : **littéraire**, **scientifique** et **tertiaire**, afin de répondre à la diversité des profils et des parcours de formation.

L'examen repose sur des programmes définis par le ministère en charge de l'éducation nationale, avec une centralisation de l'organisation et des corrections.

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 [6].

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** [3].

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 mêmes 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"[4].

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

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"[5].

1.5 Travaux antérieurs

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 partenariat 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 séries concernées, via des modèles de régression.
2. **performance universitaire** : Une suite des cohortes ici des séries concernées du bac

Un **tableau de bord interactif** sera mis en place pour visualiser les résultats de l'analyse et permettre une exploration dynamique des données.

1.7 Conclusion

Ce premier chapitre a permis de retracer l'évolution historique et structurelle du baccalauréat au Sénégal, en mettant en lumière les principales réformes qui l'ont façonné depuis les années 1990. Ces réformes, souvent motivées par des enjeux pédagogiques, socioculturels ou économiques, ont profondément modifié les séries existantes, introduit de nouvelles filières, et adapté l'examen aux réalités locales. Elles traduisent la volonté constante des autorités de moderniser et de diversifier le système éducatif pour mieux répondre aux besoins de la société sénégalaise. Cette contextualisation est essentielle pour comprendre la portée de notre étude, qui se propose d'évaluer l'impact de ces réformes à travers une approche analytique et prédictive, fondée sur l'exploitation des données réelle.

Chapitre 2

Données et outils utilisés

2.1 Introduction

Ce chapitre présente l’approche méthodologique adoptée dans le cadre de ce projet. Nous y détaillons les différentes sources de données exploitées, les étapes de Nettoyage et de préparation des données, ainsi que les outils utilisés pour leur traitement, leur analyse et leur visualisation.

Comprendre la nature des données est essentiel pour interpréter correctement les résultats obtenus. Par ailleurs, le choix des outils d’analyse joue un rôle déterminant dans la rigueur et la reproductibilité des analyses menées.

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 des institutions officielles :

- **l’Office du Baccalauréat du Sénégal ;**
- **la Direction de l’Informatique et des Systèmes d’Information (DISI) de l’UCAD ;**

Ces sources couvrent l’ensemble du parcours des étudiants, depuis leur réussite au bac jusqu’à leur évolution dans l’enseignement supérieur.

2.2.2 Description des données

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 sont utiles pour suivre l’évolution du taux de réussite au bac et évaluer l’effet des réformes introduites dans certaines séries.

Inscriptions universitaires à l’UCAD (2002–2024)

Ces données proviennent de la Direction de l’Informatique et des Systèmes d’Information (DISI) de l’UCAD. Elles incluent les informations sur les étudiants inscrits : année d’inscription, série d’origine, sexe, âge, établissement, etc. Elles permettent d’analyser les flux d’entrée à l’université selon le profil des bacheliers.

Résultats académiques à l’UCAD (2010–2023)

Également issues de la DISI de l’UCAD, ces données donnent des détails sur la performance académique des étudiants : moyenne annuelle, crédits validés, session, mention et résultat (passage, redoublement). Elles servent à évaluer la réussite universitaire en fonction des parcours scolaires initiaux.

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 [10].

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[8].

C'est sans doute le package le plus utilisé en science des données. Grâce à ses nombreuses fonctionnalités, il me 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 [7].



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 [11].

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 [12].

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 [9].

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.

Les sous-sections suivantes détaillent les étapes spécifiques effectuées lors de ce processus de nettoyage et de préparation.

3.2 Préparation de données des résultats du baccalauréat

Les données relatives aux résultats du baccalauréat ont été centralisées dans une table unique contenant plus de deux millions d'enregistrements, couvrant la période de 2006 à 2024. Le code ci-dessous (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
```

On y retrouve des informations telles que le prénom, le nom, le numéro de table, la série, le sexe, l'âge, l'établissement, les résultats, la moyenne finale, la mention, etc. On observe un total de 14 colonnes et 2236490 lignes.

3.2.1 Doublons dans la clé primaire `numero_table`

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. Cela peut s'expliquer par des erreurs de saisie ou des anomalies administratives.

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 un exemple concret pour l'année 2012, où deux élèves de centres différents ont le même numéro de table mais des informations distinctes. Ces cas ont été traités avec prudence dans la suite des analyses.

Remarque :

On remarque qu'il ne s'agit pas de doublons exacts, mais de personnes différentes à qui le

même numéro de table a été attribué.

3.2.2 Les valeurs manquantes dans les colonnes

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

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

Valeurs manquantes dans la colonne resultat

Les valeurs possibles dans la colonne resultat

3.2.3 Correction du type de la colonne `moy_finale`

Comme on peut le constater dans le code présenté (listing 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 chaine de caracteres  
    .str.replace('-', '', regex=False) # Supprimer les tirets  
    .str.replace(',', '.', regex=False) # les virgules en points  
    .str.strip() # Supprimer les espaces au debut et fin de chaine  
    .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.

Ces étapes garantissent la cohérence de la variable `moy_finale` pour les analyses statistiques à venir.

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

Pour faciliter les analyses, deux nouvelles colonnes ont été dérivées de la variable `resultat` :

- La colonne `admis`, indiquant si un candidat est admis ou non. Les codes '111' et '101' ont été interprétés comme signifiant « admis ».
- La colonne `session`, précisant s'il s'agit du premier ou du second tour, en se basant sur les mêmes codes.

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 ''))
```

Ce pré-traitement permet d'améliorer la lisibilité des résultats, notamment dans les agrégations et visualisations.

3.2.5 creation du dataframe récapitulatif

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 2023.

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 20 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 2010 à 2024, et concerne uniquement les établissements ayant effectué leurs délibérations sur la plateforme institutionnelle de la DISI. On y retrouve un total de 753 828 enregistrements répartis sur 27 colonnes.

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 fusionner les bases de données d'inscription et de résultats, les colonnes *NUMERO* et *ANNEE UNIVERSITAIRE* ont été utilisées comme clés de jointure.

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' # garde toutes les lignes de df_inscrit
                    )
```

La fusion a été réalisée à l'aide de l'option `how='left'` pour conserver l'ensemble des inscrits, y compris ceux pour lesquels aucun résultat académique n'est disponible.

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
```

Cela explique la présence d'un grand nombre de valeurs manquantes dans la colonne *RESULTAT*, notamment pour les étudiants dont les résultats n'ont pas encore été délibérés ou publiés sur la plateforme.

3.4 Filtrage des données

Afin de concentrer l'analyse sur la population cible et sur les éléments pertinents pour notre étude, une étape de filtrage rigoureuse des données a été nécessaire. Cette étape vise à exclure les individus et les séries qui ne rentrent pas dans le périmètre de l'étude.

3.4.1 Nationalité

La variable *NATIONALITE* a été filtrée pour ne conserver que les étudiants de nationalité sénégalaise. Seuls les candidats de nationalité sénégalaise ont été retenus. Cela permet de mieux étudier l'impact des réformes du baccalauréat sur les élèves du système éducatif sénégalais.

3.4.2 Séries du baccalauréat

La variable *SERIE_BACC* a été filtrée pour ne conserver que les séries pertinentes pour l'analyse.

Nous avons sélectionné uniquement les séries concernées par les réformes étudiées et pertinentes dans le contexte de l'enseignement général ou technique, telles que :

- L1 (Littéraire 1)
- L2 (Littéraire 2)
- S1 (Scientifique 1)
- S2 (Scientifique 2)
- T1 (Technique 1)
- T2 (Technique 2)
- T3 (Technique 3)
- T4 (Technique 4)

3.4.3 Niveau universitaire

Seuls les étudiants inscrits en L1, L2 ou L3, correspondant au premier cycle universitaire, ont été retenus. Cela permet d'observer directement les effets des réformes du baccalauréat

sur les premières années d'insertion 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

Chapitre 4

Analyse du taux de réussite au bac

4.1 Introduction

4.2 Analyse de globale

4.2.1 Évolution du nombre d'inscrits, présents et admis

4.3 Conclusion

Chapitre 5

Analyse du parcours universitaire des bacheliers (UCAD)

5.1 Introduction

5.2 Analyse des résultats du baccalauréat

5.2.1 Analyse descriptive du taux de réussite

5.2.2 Étude comparative avant et après les réformes

5.2.3 Modélisation prédictive du taux de réussite

5.3 Analyse du parcours universitaire des bacheliers

5.4 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 :

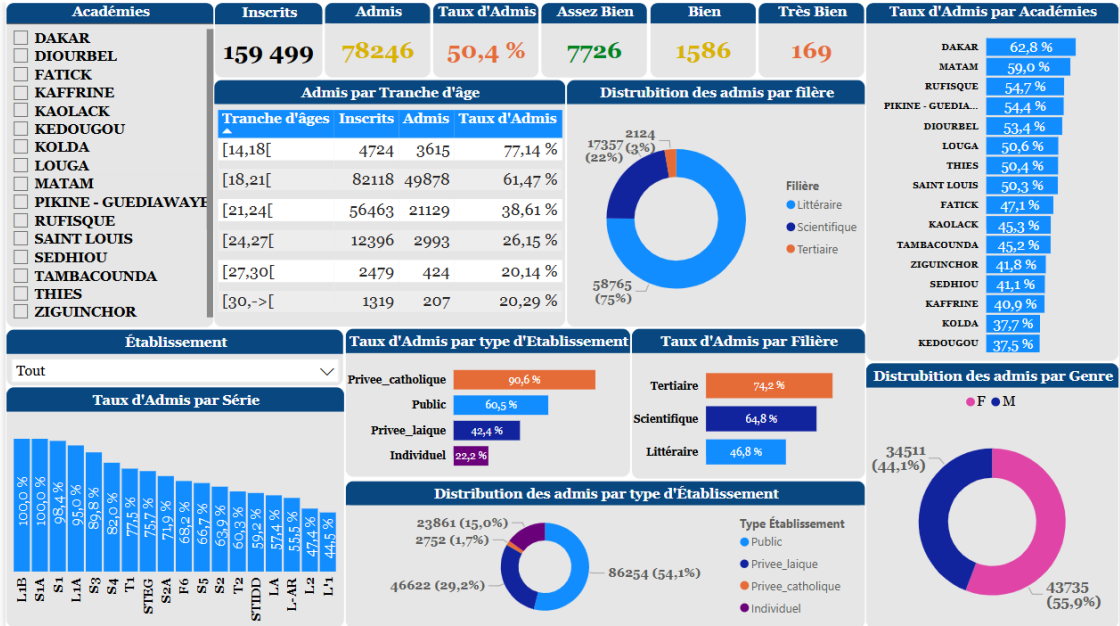
- Le taux de réussite global par année ;
- Le taux de réussite par série ;
- Le nombre total d'admis et d'inscrits ;
- La répartition par session (1er ou 2e tour) ;
- La répartition par mention obtenue.

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.

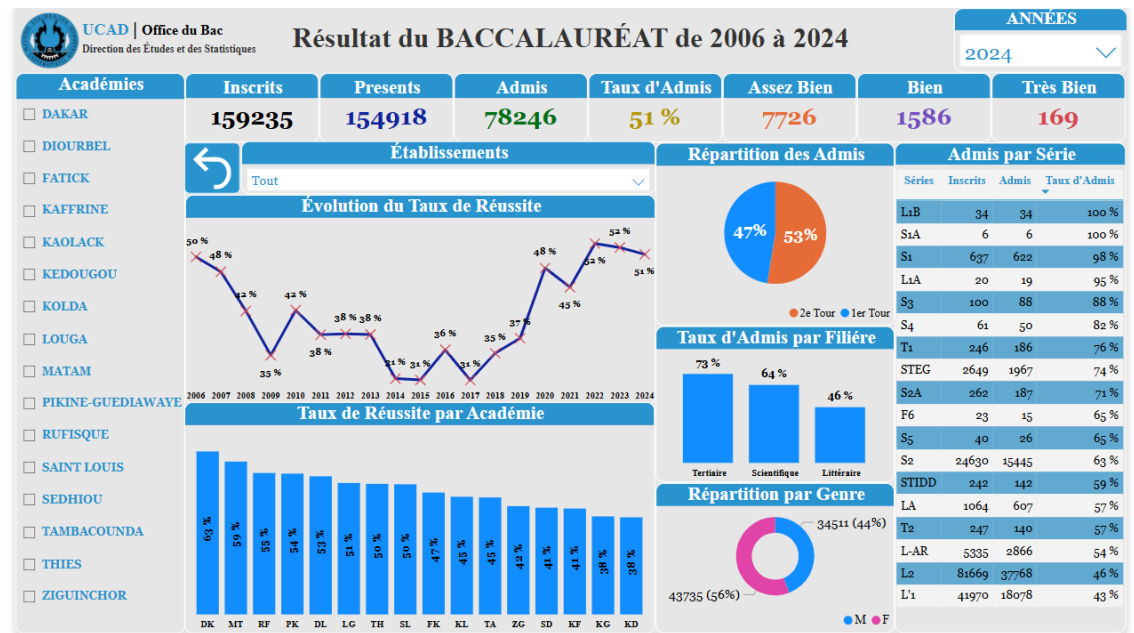
FIGURE 6.1 – Tableau de bord Power BI - Résultats du baccalauréat 2024



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.

Chapitre 7

Discussion et Recomandation

7.1 Introduction

7.2 Conclusion

Conclusion Générale

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.
- [2] LIVREAUSENEGAL. *Origine du baccalaureat*. 2024. URL : <https://www.instagram.com/livresausenegal/p/C8hKo-ksWP3/?hl=fr>.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] 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%20C3%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..
- [7] *site officiel de Matplotlib*. URL : <https://matplotlib.org/>.
- [8] *site officiel de Pandas*. URL : <https://pandas.pydata.org/>.
- [9] *site officiel de Power BI*. URL : <https://powerbi.microsoft.com/fr-fr/>.

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