



EduTrack

# cahier des charges

---

projet d'analyse

2024

THE IDEA BOOK

# Introduction

## CONTEXTE DU PROJET

Le projet EduTrack a pour objectif principal d'analyser les données sur la performance académique des élèves et des enseignants. L'idée est de comprendre les différents facteurs qui influencent les résultats scolaires et de proposer des recommandations concrètes pour les améliorer. Nous allons nous concentrer sur les relations entre les matières, les départements, l'expérience des enseignants, et les choix des élèves.



### Portée du projet (Scope) :

Le projet EduTrack couvrira l'analyse des données issues de plusieurs semestres et matières. Nous allons examiner les performances des élèves dans différents départements, tout en tenant compte des dates d'entrée des élèves et des enseignants, ainsi que de leur expérience. L'analyse inclura aussi les choix des élèves concernant les départements qu'ils préfèrent. Le but est de simplifier les données pour mieux comprendre comment ces différents éléments influencent les résultats scolaires.



# Sommaire



INTRODUCTION	02
OBJECTIFS	04
LES DONNÉES	06
MÉTHODOLOGIE	08
RESSOURCES	11
LIVRABLES	13
PLANNING	14
RISQUES	15



# 2. Objectifs

---

## QUESTION PRINCIPAL DU PROJET

Quels sont les facteurs qui influencent le plus la performance des étudiants et des enseignants à l'institut, et comment pouvons-nous utiliser ces facteurs pour améliorer leurs performances académiques ?

**Les objectifs principaux de ce projet « EduTrack » sont les suivants :**

1. Comprendre les facteurs influençant la performance académique :
  - Identifier les éléments clés qui influencent le succès ou l'échec des étudiants dans leurs études.
  - Analyser comment l'expérience et la qualification des enseignants impactent les résultats des étudiants.
2. Analyser la relation entre les départements et les performances des étudiants :
  - Comparer les performances académiques des étudiants en fonction du département où ils sont inscrits.
  - Étudier les différences de résultats entre les départements pour identifier les forces et les faiblesses de chaque programme.



# Les Données

## RÉSUMÉ DES DONNÉES



**1. Contexte :** Ces données proviennent d'un institut fictif, et elles sont utilisées pour analyser la performance des étudiants et des professeurs. Elles enregistrent le parcours des étudiants depuis leur admission jusqu'à la fin de leur formation. Cette version est la première, et d'autres tables seront ajoutées à l'avenir.

**2. Contenu des données :** Les données contiennent des informations sur les départements, les professeurs, l'orientation des étudiants, les cours proposés, les choix d'admission des étudiants, ainsi que leurs performances aux examens.

# Les Données

## DESCRIPTION DES TABLES :

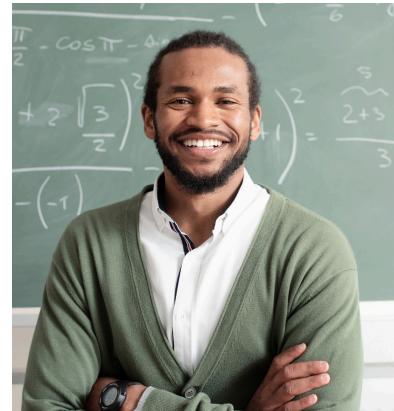
Department\_Information



- **Department\_ID** : Identifiant unique du département.
- **Department\_Name** : Nom du département (ex : "Aerospace Engineering").
- **DOE (Date of Establishment)** : Date de création du département.

- **Employee ID** : Identifiant unique de l'employé.
- **DOB (Date of Birth)** : Date de naissance de l'employé.
- **DOJ (Date of Joining)** : Date d'entrée en fonction de l'employé.
- **Department\_ID** : Identifiant du département auquel l'employé est affecté.

Employee\_Information



Student\_Information



- **Student\_ID** : Identifiant unique de l'étudiant.
- **DOA (Date of Admission)** : Date d'admission de l'étudiant.
- **DOB (Date of Birth)** : Date de naissance de l'étudiant.
- **Department\_CHOICES** : Département(s) choisi(s) par l'étudiant.
- **Department\_Admission** : Département dans lequel l'étudiant a été admis.

# Les Données

## DESCRIPTION DES TABLES :

Student\_Performance



- **Student\_ID** : Identifiant unique de l'étudiant.
- **DOA (Date of Admission)** : Date d'admission de l'étudiant.
- **DOB (Date of Birth)** : Date de naissance de l'étudiant.
- **Department\_CHOICES** : Département(s) choisi(s) par l'étudiant.
- **Department\_ADMISSION** : Département dans lequel l'étudiant a été admis.

- **Student\_ID** : Identifiant unique de l'étudiant.
- **Semester\_Name** : Nom du semestre.
- **Marks for Each Paper** : Colonnes contenant les notes pour chaque matière.

Consolidated\_Results



## NETTOYAGE DES DONNÉES

Les données pourraient comporter certaines incohérences qui nécessitent un nettoyage pour garantir l'intégrité des analyses. Les étapes de nettoyage incluent :

- Correction des valeurs manquantes ou erronées.
- Harmonisation des formats de date.
- Identification des duplicitas dans les tables.
- Validation de l'exactitude des relations entre les différentes tables (ex : correspondance entre les départements des étudiants et ceux des employés).





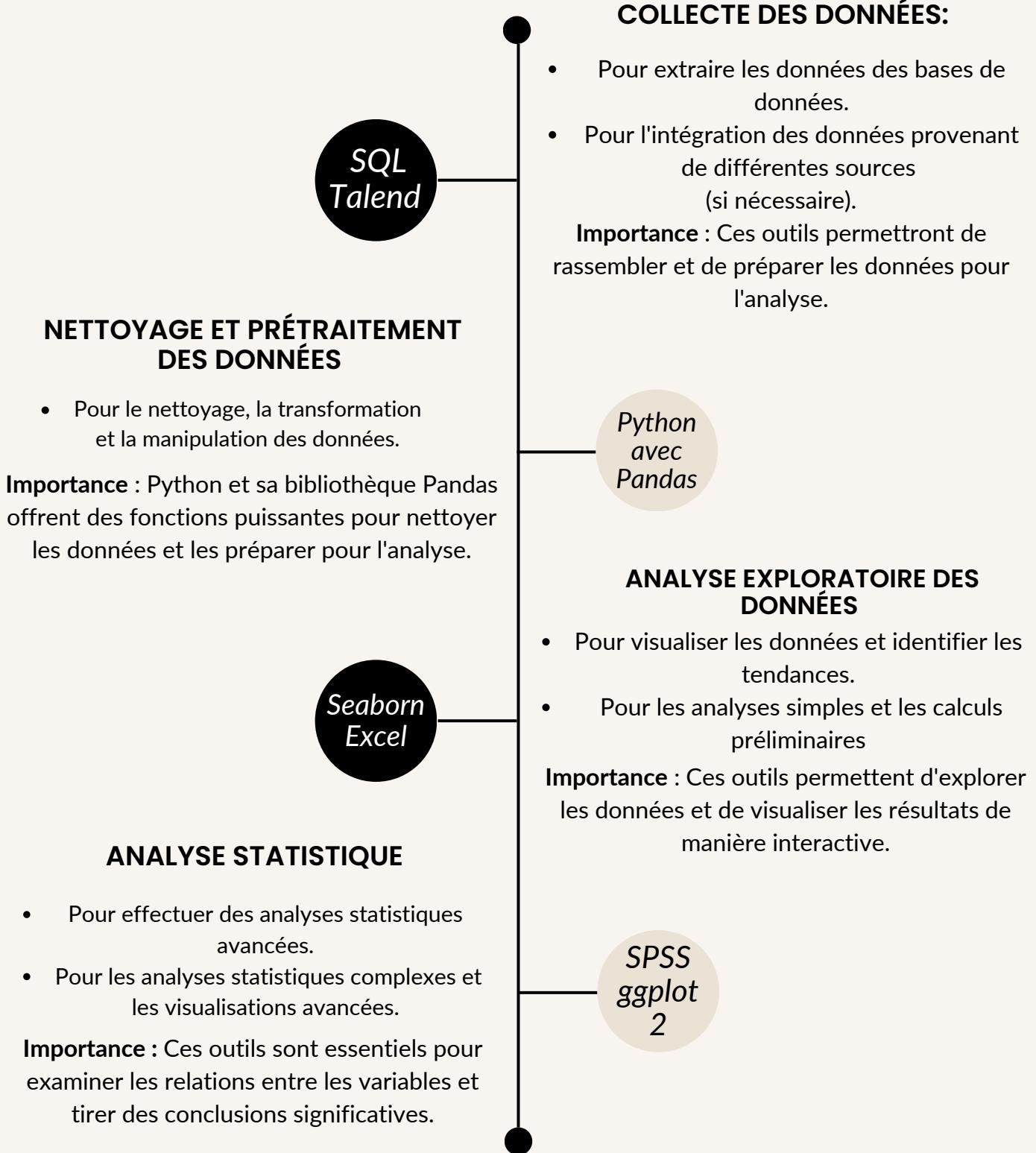
# Méthodologie

---

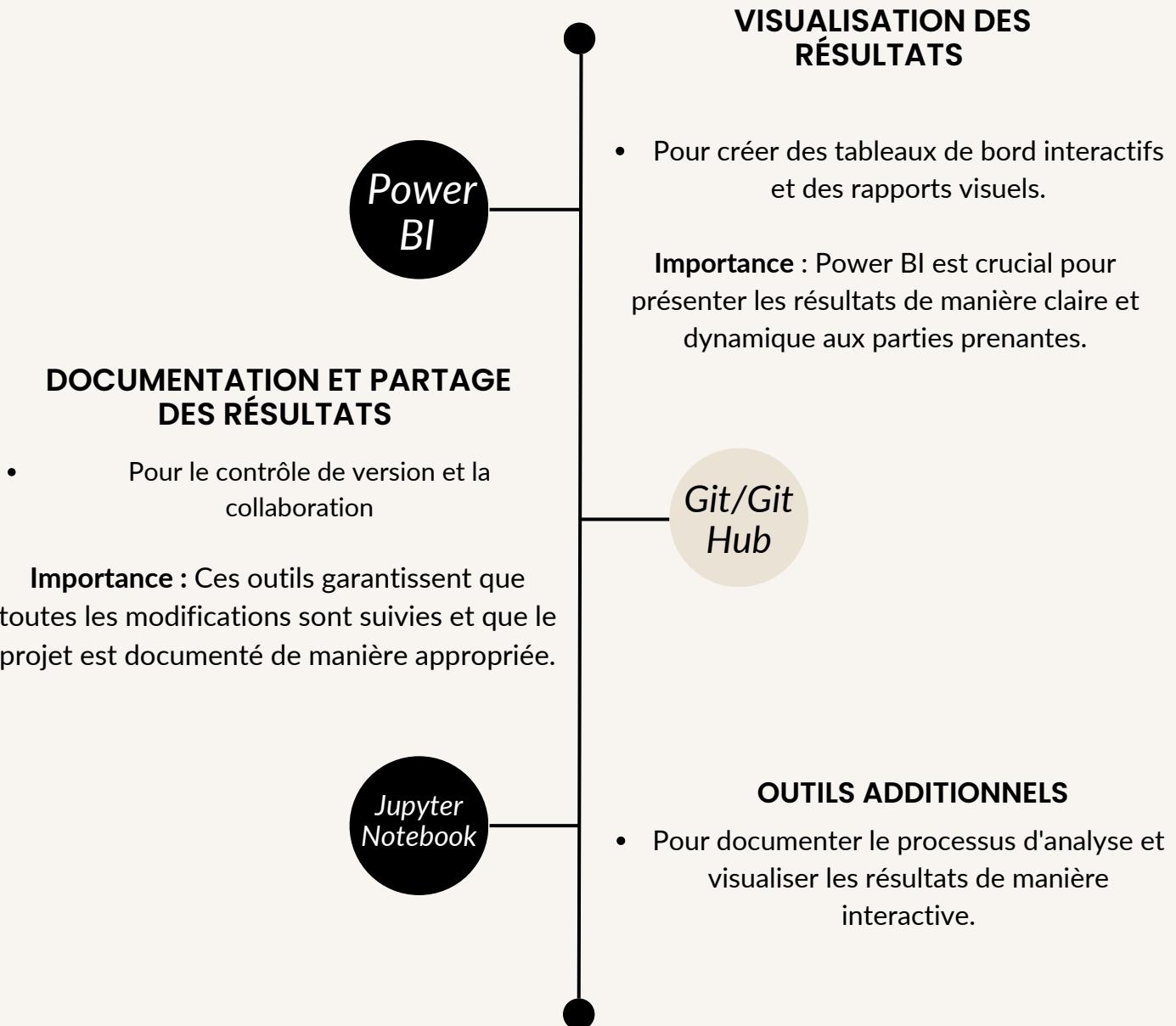
Pour le projet « EduTrack », nous allons organiser notre méthodologie en étapes successives, en utilisant les outils les plus adaptés à chaque phase. Voici l'ordre de priorité et les outils recommandés, en tenant compte de la séquence du projet :

- ✓ **Sources de données :** Nous allons rassembler des données sur les performances académiques des étudiants, les informations sur les enseignants, les matières, les départements et les semestres
  
- ✓ **Structuration des données :** Organisation des données sous un format qui facilite les analyses, en classant les informations par étudiants, départements, matières, et périodes de temps.

# Méthodologie



# Méthodologie



## Outil Principal

L'outil central tout au long du projet sera **Python**, principalement avec la bibliothèque Pandas pour la manipulation des données et Seaborn pour la visualisation. Cela nous permettra de mener une analyse complète et approfondie des performances académiques, tout en étant flexible et adaptable aux besoins spécifiques du projet.

# Ressources Nécessaires

Pour assurer le bon déroulement du projet « EduTrack », plusieurs ressources seront nécessaires, regroupées en différentes catégories :

## Ressources Humaines :

- **Analystes de données** : Pour l'extraction, le nettoyage et l'analyse des données.
- **Statisticiens** : Pour effectuer des analyses statistiques avancées et interpréter les résultats.
- **Développeurs** : Pour automatiser les processus d'extraction et de transformation des données.
- **Experts en visualisation** : Pour créer des rapports et des tableaux de bord interactifs.



## Ressources Techniques :

- Matériel informatique :
  - Ordinateurs avec des spécifications adéquates (processeur puissant, RAM suffisante) pour exécuter des analyses lourdes.

## Logiciels :

- **Systèmes de gestion de base de données (DBMS)** : SQL Server, MySQL ou PostgreSQL.
- **Outils d'analyse et de visualisation** : SPSS, R, Power BI, Talend, Excel.
- **Environnement de développement** : Anaconda pour Python, Jupyter Notebook pour la documentation et la visualisation interactive.
- **Plateforme de gestion de version** : Git et GitHub pour le suivi des modifications et la collaboration.

# Ressources Nécessaires

## Ressources de Données :

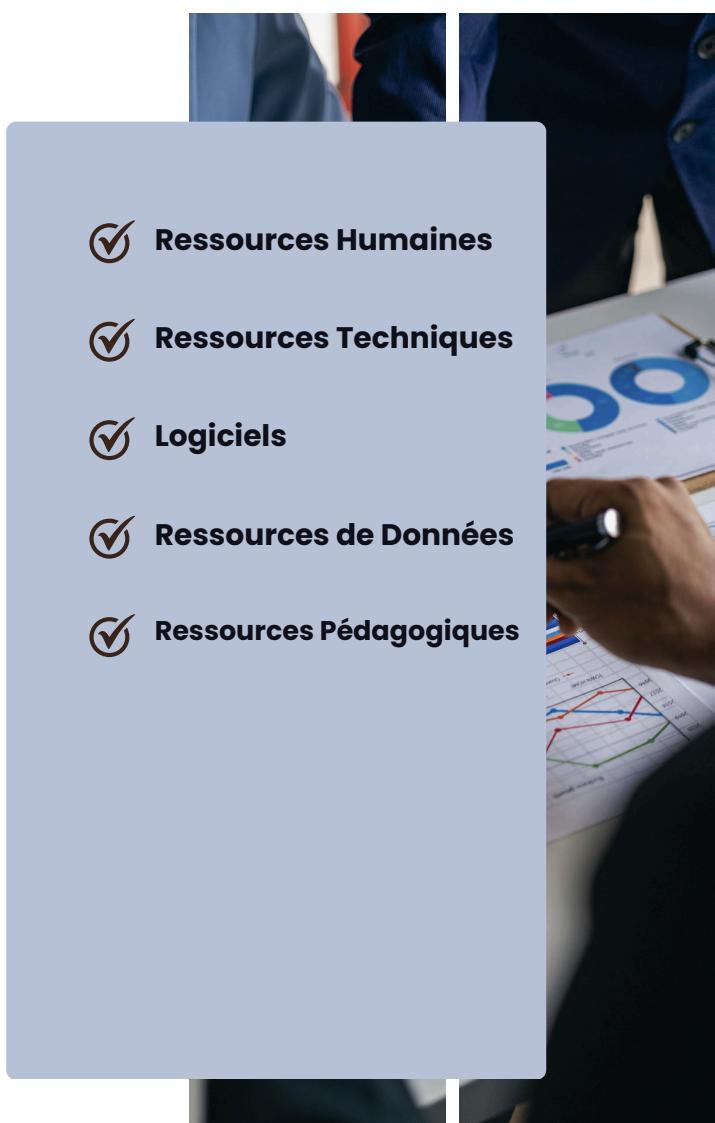
- **Accès aux bases de données académiques** : Données sur les performances des étudiants, les informations sur les enseignants, les dates d'inscription, etc.
- **Sources externes** : Si nécessaire, accès à des données externes pour enrichir l'analyse, comme des études de marché ou des recherches académiques.

## Ressources Pédagogiques :

- **Documentation** : Manuels et guides sur les outils utilisés (Python, SPSS, Power BI).
- **Formations** : Sessions de formation pour l'équipe sur l'utilisation des outils d'analyse et de visualisation, ainsi que sur les méthodes d'analyse statistique.

## Conclusion

La mise à disposition de ces ressources est essentielle pour garantir le succès du projet « EduTrack ». Cela permettra de réaliser une analyse approfondie des performances académiques et de proposer des recommandations pratiques pour l'amélioration continue.

- 
- Ressources Humaines
  - Ressources Techniques
  - Logiciels
  - Ressources de Données
  - Ressources Pédagogiques

# Livrables

LES PRINCIPAUX LIVRABLES DU PROJET « EDUTRACK » SONT :



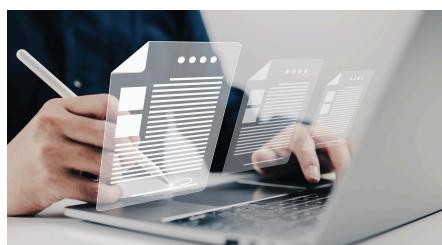
**Rapport d'analyse des données :** Incluant les résultats, les conclusions, et des recommandations pour améliorer les performances académiques.



**Tableaux de bord interactifs :** Crées avec Power BI et Excel pour visualiser les données et suivre les performances des étudiants et enseignants.



**Modèles statistiques et prédictifs :** Analyses réalisées dans SPSS et R pour identifier les facteurs clés et prédire les tendances.



**Documentation technique :** Code commenté, guides d'utilisation, et documentation pour faciliter la maintenance et l'utilisation des outils.



**Présentation finale :** Résumé visuel des résultats et recommandations, avec une discussion des conclusions.

# Planning

VOICI UNE PROPOSITION SIMPLIFIÉE DU PLANNING POUR LE PROJET « EDUTRACK »

**1**

## PRÉPARATION (JOUR 1-3)

- Définir les objectifs et créer le cahier des charges.*
- Collecter et préparer les données.*

**2**

## ANALYSE DES DONNÉES (JOUR 4-10)

- Nettoyage et exploration des données avec Python et Talend.*
- Analyse statistique de base avec Excel et SPSS.*

**3**

## MODÉLISATION ET VISUALISATION (JOUR 11-18)

- Analyse approfondie avec R et création de modèles statistiques.*
- Visualisation des résultats avec Power BI.*

**4**

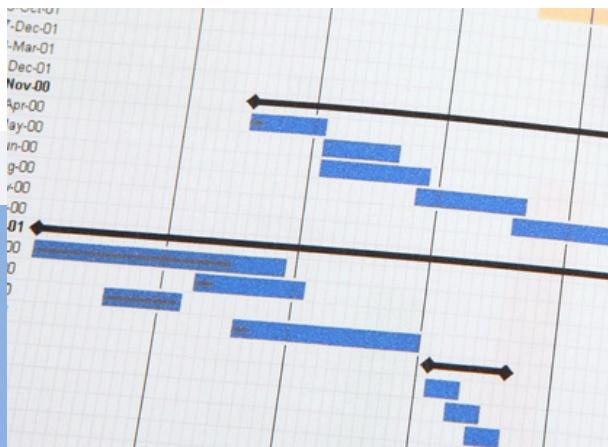
## VALIDATION ET DOCUMENTATION (JOUR 19-22)

- Documenter les analyses et valider les résultats.*
- Préparer les livrables finaux et les rapports.*

**5**

## PRÉSENTATION FINALE (JOUR 23-25)

- Présenter les conclusions et les recommandations.*



# Risques

## RISQUES POTENTIELS



### QUALITÉ DES DONNÉES

*Données incomplètes ou erronées*

**Solution :** Nettoyage et validation anticipée.



### COMPLEXITÉ DE L'ANALYSE

*Analyse complexe*

**Solution :** Utiliser des outils adaptés aux compétences de l'équipe.



### RESSOURCES TECHNIQUES

*Outils non disponibles*

**Solution :** Prévoir l'accès aux outils ou utiliser des alternatives.



### GESTION DU TEMPS

*Retards*

**Solution :** Suivi rigoureux avec des jalons clairs.



### COMMUNICATION

*Mauvaise collaboration*

**Solution :** Utiliser des plateformes de gestion de projet et communication.



### DONNÉES HISTORIQUES

*Données biaisées*

**Solution :** Analyser la validité des conclusions avec des données alternatives.



CE PROJET **EDUTRACK**, RÉALISÉ PAR **AYMEN RAMI**, VISE À ANALYSER LES PERFORMANCES ACADEMIQUES DES ÉTUDIANTS ET DES ENSEIGNANTS POUR IDENTIFIER LES FACTEURS INFLUENÇANT CES PERFORMANCES. L'OBJECTIF PRINCIPAL EST DE PERMETTRE AUX PARTICIPANTS D'APPLIQUER DES COMPÉTENCES PRATIQUES, DE DÉVELOPPER LEURS CAPACITÉS DE COMMUNICATION ET DE RENFORCER LE TRAVAIL EN ÉQUIPE.

JE TIENS À EXPRIMER MES REMERCIEMENTS À TOUS CEUX QUI PARTICIPERONT À LA CONSTRUCTION DE CE PROJET, CAR LEUR CONTRIBUTION EST ESSENTIELLE À NOTRE SUCCÈS COLLECTIF.

# Merci.

*Theramithkarface*