



المدرسة الوطنية للذكاء الاصطناعي والرقمية - بركان  
ÉCOLE NATIONALE DE L'INTELLIGENCE ARTIFICIELLE ET DU DIGITAL - BERKANE  
الكلمة المحمدية لـ إيكالا - مراكش - المغرب



جامعة محمد الخامس وجدة  
UNIVERSITE MOHAMMED PREMIER OUJDA  
جامعة محمد الخامس وجدة



---

# Développement d'un système BI pour le suivi des performances académiques des étudiants au maroc

---

Auteurs:

- **LAMKADDEM Ayoub**
- **BOUABDALLAH Aymane**
- **RHOUATI Mohammed**
- **LYAGOUBI Ziad**

Encadré par :

**BOUKIL Naoual**

<b>I. Données.....</b>	<b>5</b>
1. A. Description des données :.....	5
2. Selection des données :.....	5
3. Nettoyage et pré-traitement des données :.....	6
<b>II. Conception du Data Warehouse:.....</b>	<b>8</b>
1. Définition.....	8
a. Modélisation multidimensionnelle en étoile :.....	8
b. Table de faits :.....	8
c. Table de dimensions :.....	9
d. Data Warehouse :.....	9
2. Identification de la table des faits.....	9
3. Identification des dimensions.....	10
4. Schéma du Data Warehouse :.....	11
5. Implémentation dans MySQL.....	11
<b>III. Processus ETL : Extraction, Transformation et Chargement des Données :.....</b>	<b>12</b>
1. Définition.....	12
a. ETL : Extraction, Transformation, Chargement.....	12
b. Talend Open Studio:.....	12
2. Extraction des données.....	13
3. Transformation des données.....	13
4. Chargement des données.....	14
5. Résultat :.....	15
<b>IV- Visualisation des données sous Power BI.....</b>	<b>16</b>
1. Introduction : Pourquoi utiliser Power BI ?.....	16
2. Analyse du première dashboard : Analyse des performances académiques des élèves.....	18
2.1 Analyse graphique.....	19
2.2 Tendance générale.....	19
2.3 Interprétation.....	20
3. Analyse du deuxième dashboard : Analyse Comportement d'apprentissage des élèves.....	21
3.1 Vue globale.....	21
3.2 Analyse par Graphique.....	21
4. Analyse du troisième dashboard : Analyse Comportement d'apprentissage des élèves.....	22
4.1 Vue globale.....	23
4.2 Analyse graphique.....	23
4.3 Tendance générale.....	23
Interprétation.....	24
<b>V- Les trois décisions à l'issue de l'analyse.....</b>	<b>24</b>

## Résumé

Dans ce projet, nous avons conçu un **système de Business Intelligence (BI)** pour le suivi des performances académiques des étudiants au Maroc.

Notre travail s'est appuyé sur la base de données xAPI-Edu-Data, enrichie et restructurée pour alimenter un **entrepôt de données**.

Nous avons développé un **modèle en étoile**, dont la **table des faits faits\_performance** est liée à plusieurs **dimensions** : Étudiant, Matière, Scolarité et Parents. Cette table collecte des indicateurs clés tels que :

- La réussite académique des étudiants,
- Leur nombre de jours d'absence,
- Le degré de satisfaction des parents.

Le processus **ETL**, réalisé avec **Talend Open Studio**, nous a permis :

- D'extraire les données brutes depuis un fichier CSV,
- De les transformer (nettoyage, normalisation des valeurs, création d'identifiants),
- Puis de les charger dans une base **MySQL**.

Enfin, avec **Power BI**, nous avons construit des **tableaux de bord dynamiques** pour :

- Analyser l'impact du statut social et du niveau parental sur les performances,
- Suivre les absences et la satisfaction parentale,
- Visualiser la réussite académique selon plusieurs axes.

## ***Introduction***

Dans un contexte où l'amélioration de l'éducation est une priorité nationale, disposer d'outils permettant d'analyser et de comprendre les performances académiques des étudiants devient essentiel.

Le projet que nous avons réalisé s'inscrit dans cette dynamique : il vise à développer un **système de Business Intelligence (BI)** permettant d'offrir une **vision globale et détaillée** de la réussite scolaire au Maroc.

En nous appuyant sur un jeu de données provenant de **Kaggle** (xAPI-Edu-Data), nous avons mis en œuvre l'ensemble du processus de traitement de données :

- **Extraction, Transformation et Chargement (ETL)** des données avec **Talend Open Studio**,
- **Création d'un entrepôt de données** dans **MySQL**, en suivant une **modélisation en étoile**,
- **Construction de tableaux de bord dynamiques** avec **Power BI** pour faciliter l'analyse des résultats.

Notre base de données, organisée autour d'une table des faits `faits_performance` et de dimensions pertinentes (étudiant, matière, scolarité, parents), permet de suivre non seulement les performances académiques mais aussi des indicateurs tels que les absences ou la satisfaction des parents.

Ce projet fournit ainsi aux décideurs un outil précieux pour orienter les stratégies éducatives sur des bases factuelles.

# I. Données

## Source des données :

Le projet s'appuie sur un jeu de données public intitulé **Xapi-edu-Data**, disponible sur la plateforme **Kaggle**. Ce dataset a été créé pour explorer et prédire les performances académiques des étudiants à travers différentes variables socio-démographiques, comportementales et académiques.

- Lien vers les données : [Xapi-edu-data](#).

### 1. A. Description des données :

Le fichier **xAPI-Edu-Data.csv** contient **17 colonnes** représentant divers attributs relatifs aux étudiants :

- **Informations personnelles** : gender, NationalITY, PlaceofBirth.
- **Informations scolaires** : StageID (niveau d'études), GradeID (classe), SectionID, Topic (matière étudiée), Semester.
- **Relations familiales** : Relation (lien du parent), ParentAnsweringSurvey, ParentschoolSatisfaction.
- **Comportements académiques** : raisedhands (nombre de mains levées), VisITEDResources (consultation des ressources), AnnouncementsView (vues des annonces), Discussion (participation aux discussions).
- **Données d'absences** : StudentAbsenceDays (nombre de jours d'absence).
- **Performance finale** : Class (niveau de réussite de l'étudiant : High, Middle, Low).

Les données sont bien structurées, mais nécessitent des étapes de préparation avant l'intégration dans un entrepôt de données.

### 2. Selection des données :

Pour le projet, nous avons sélectionné les variables essentielles à l'analyse de la performance académique :

- **Variables explicatives** : gender, NationalITY, StageID, Topic, Relation, ParentAnsweringSurvey, ParentschoolSatisfaction, StudentAbsenceDays.
- **Variables comportementales** : raisedhands, VisITEDResources, AnnouncementsView, Discussion.
- **Variable cible** : Class, qui définit le niveau final de performance.

Certaines colonnes telles que PlaceofBirth ou SectionID, moins pertinentes pour notre objectif d'analyse globale, ont été volontairement écartées lors de la phase ETL afin d'alléger la structure et d'améliorer la lisibilité des analyses.

### 3. Nettoyage et pré-traitement des données :

Avant l'intégration des données dans notre entrepôt, plusieurs étapes de **nettoyage** et de **prétraitement** ont été réalisées pour garantir la qualité et la cohérence des informations :

#### □ Vérification de la complétude :

Nous avons contrôlé l'absence de valeurs manquantes dans le fichier **xAPI-Edu-Data.csv**. Le dataset était complet, ne nécessitant aucune opération d'imputation.

#### □ Correction des incohérences :

Quelques valeurs textuelles ont été uniformisées pour éviter les doublons dus aux variations d'écriture (par exemple, normalisation de la casse des textes en minuscules ou majuscules constantes si nécessaire).

#### □ Transformation des types de données :

Certaines colonnes ont été préparées pour être intégrées efficacement dans MySQL :

- Les colonnes catégorielles (`gender`, `NationalITY`, `Topic`, `StageID`, etc.) ont été transformées en **entiers** via des **clés substituts** (ID) pour normaliser les dimensions dans la modélisation en étoile.
- Les colonnes numériques (`raisedhands`, `VisITEDResources`, `AnnouncementsView`, `Discussion`) ont été converties en types numériques appropriés (entier).

#### □ Création des dimensions et de la table des faits :

À partir du dataset d'origine, nous avons extrait des tables distinctes pour :

- **Dimension Étudiant** : `gender`, `NationalITY`, `StageID`.
- **Dimension Matière** : `Topic`, `Semester`.
- **Dimension Parent** : `Relation`, `ParentAnsweringSurvey`, `ParentschoolSatisfaction`.
- **Dimension Scolarité** : `StudentAbsenceDays`.
- **Fait Performance** : `raisedhands`, `VisITEDResources`, `AnnouncementsView`, `Discussion`, `Class` (niveau de performance), `absencedays`, `parentsatisfaction`.

#### □ Codage des valeurs catégorielles :

Certaines valeurs textuelles ont été traduites en formats standards, par exemple :

- `StudentAbsenceDays` transformé en valeurs binaires (0 pour "Under-7", 1 pour "Above-7").
- `ParentschoolSatisfaction` converti en indicateurs (1 pour "Good", 0 pour "Bad").

classEtudiant	ParentschoolSatisfaction	StudentAbsenceDays
1	1	0
1	1	0
0	0	1
0	0	1
1	0	1
1	0	1
0	0	1
1	1	0
1	1	0
1	1	0
2	1	0
1	1	0
0	0	1
0	0	1
2	0	1
1	1	0
1	0	1
1	0	1
1	1	0
2	1	0
1	0	1
1	0	0
1	0	0

□ **Suppression de colonnes inutiles :**

Les colonnes jugées non pertinentes pour l'analyse décisionnelle telles que PlaceofBirth et SectionID ont été supprimées pour simplifier la base de données.

## II. Conception du Data Warehouse:

### 1. Définition

La conception de notre Data Warehouse repose sur une **modélisation multidimensionnelle en étoile**, permettant d'analyser efficacement les performances académiques des étudiants.

#### a. Modélisation multidimensionnelle en étoile :

La **modélisation multidimensionnelle en étoile** (ou **Star Schema**) est une méthode de conception d'entrepôt de données qui organise l'information de manière intuitive et optimisée pour l'analyse.

Elle repose sur une **table des faits** centrale, contenant les mesures quantitatives (par exemple : notes, performances, ventes), qui est reliée directement à plusieurs **tables de dimensions** décrivant le contexte de ces faits (par exemple : étudiant, matière, parent).

Dans un schéma en étoile :

- La **table des faits** est au centre et contient principalement des **clés étrangères** vers les dimensions et des **mesures**.
- Les **tables de dimensions** entourent la table des faits, comme les branches d'une étoile, et contiennent des données descriptives riches (par exemple : le nom de l'étudiant, le genre, la matière étudiée, etc.).

Ce type de modélisation est particulièrement adapté aux systèmes d'aide à la décision (BI) car :

- Il facilite l'écriture de requêtes analytiques simples,
- Il optimise les temps de réponse pour les analyses,
- Il améliore la lisibilité du modèle de données pour les utilisateurs métiers.

#### b. Table de faits :

La **table de faits** est l'élément central du modèle en étoile.

Elle contient les **mesures quantitatives** ou **événements** que l'on souhaite analyser.

Chaque enregistrement de la table de faits est généralement lié à des enregistrements spécifiques dans les tables de dimensions par l'intermédiaire de **clés étrangères**.

Dans notre projet, la table de faits **faits\_performance** contient par exemple :

- Les notes finales des étudiants,
- Les performances académiques,
- Les identifiants liés aux étudiants, aux matières, aux parents, etc.

La table de faits est principalement composée de :

- **Mesures** : données numériques à analyser (exemple : note finale),
- **Clés étrangères** : reliant la table aux dimensions associées.

### c. Table de dimensions :

Les **tables de dimensions** contiennent les **attributs descriptifs** associés aux faits. Elles fournissent le contexte nécessaire pour interpréter les mesures présentes dans la table de faits.

Chaque dimension représente un axe d'analyse possible :

- **Dimension Étudiant** : fournit des informations telles que l'âge, le genre, le lieu de résidence...
- **Dimension Matière** : donne des détails sur les matières enseignées.
- **Dimension Parent** : décrit les réponses des parents à certains sondages.
- **Dimension Scolarité** : décrit le niveau scolaire de l'étudiant.

Les tables de dimensions sont caractérisées par :

- Une **clé primaire** unique pour chaque enregistrement,
- Plusieurs **attributs descriptifs** facilitant l'analyse et les regroupements dans les tableaux de bord.

### d. Data Warehouse :

Un **Data Warehouse** (ou **entrepôt de données**) est une base de données spécialisée conçue pour **centraliser, stocker et analyser** de grandes quantités de données provenant de sources diverses.

Contrairement aux bases de données transactionnelles classiques (qui servent aux opérations quotidiennes), un Data Warehouse est optimisé pour :

- **L'analyse historique**,
- **La prise de décision stratégique**,
- **L'exploration de tendances et le reporting**.

Caractéristiques principales d'un Data Warehouse :

- **Intégration** : collecte des données issues de différentes sources (bases opérationnelles, fichiers CSV, API, etc.).
- **Transformation** : harmonisation et nettoyage des données pour assurer leur cohérence.
- **Historisation** : conservation des données sur une longue période pour permettre des analyses temporelles.
- **Optimisation pour l'analyse** : structure adaptée pour exécuter rapidement des requêtes complexes.

Dans notre projet, nous avons construit un Data Warehouse basé sur un modèle en étoile, permettant de suivre et d'analyser les **performances académiques des étudiants** au Maroc.

## 2. Identification de la table des faits

Nous avons créé une **table des faits** appelée **faits\_performance**, qui contient les indicateurs numériques mesurant l'activité et la réussite académique des étudiants :

- **raisedhands** : Nombre de fois que l'étudiant a levé la main en classe.
- **VisITEDResources** : Nombre de ressources visitées.
- **AnnouncementsView** : Nombre d'annonces consultées.
- **Discussion** : Nombre de participations aux discussions.
- **Class** : Résultat académique de l'étudiant (High, Middle, Low).

Chaque enregistrement dans la table des faits est lié aux dimensions via des **clés étrangères**.

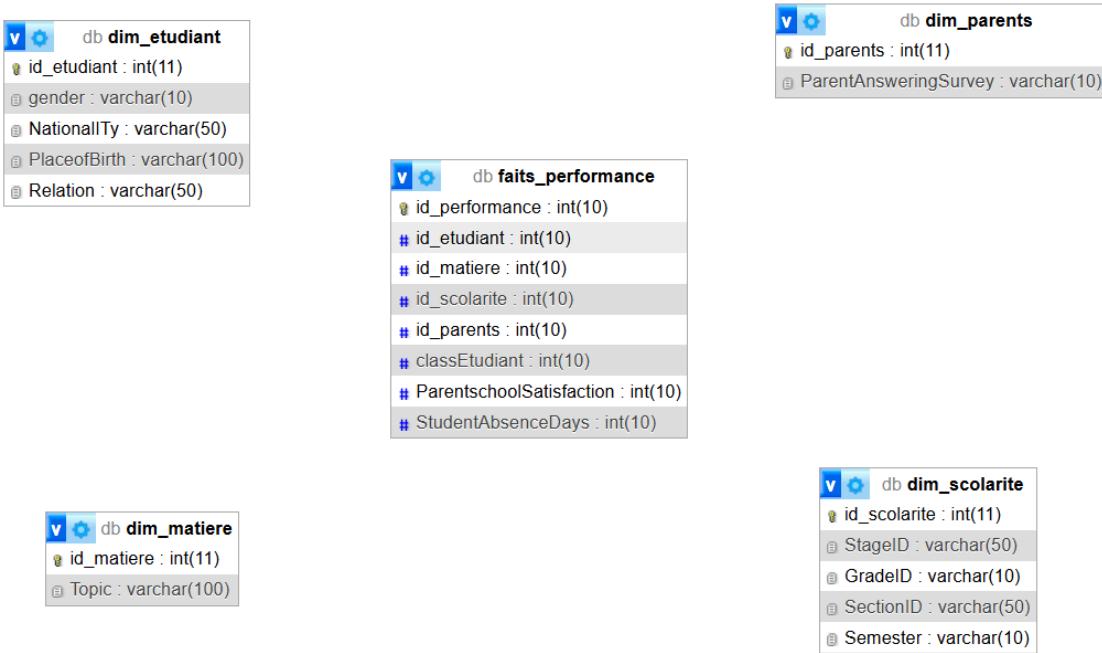
## 3. Identification des dimensions

Pour enrichir l'analyse, nous avons défini plusieurs **tables de dimensions** :

- **Dimension Étudiant** :
  - gender (sexe de l'étudiant)
  - NationalITY (nationalité)
  - StageID (niveau scolaire)
- **Dimension Matière** :
  - Topic (matière étudiée)
  - Semester (semestre)
- **Dimension Parent** :
  - Relation (relation parentale avec l'étudiant)
  - ParentAnsweringSurvey (le parent a répondu à l'enquête ?)
  - ParentschoolSatisfaction (satisfaction des parents vis-à-vis de l'école)
- **Dimension Scolarité** :
  - StudentAbsenceDays (jours d'absence de l'étudiant)

Chaque dimension possède une **clé primaire** (par exemple `id_etudiant`, `id_matiere`, etc.) pour lier efficacement les données avec la table des faits.

#### 4. Schéma du Data Warehouse :



Ce modèle permet :

- Des jointures simples pour faciliter les requêtes analytiques.
- Une lecture rapide des mesures en fonction de plusieurs axes (matière, étudiant, parent, scolarité).

#### 5. Implémentation dans MySQL

Les tables ont été créées dans MySQL en respectant :

- **Clé primaire** pour chaque table de dimension.
- **Clés étrangères** dans la table des faits pointant vers les clés primaires des dimensions.
- **Types de données adaptés :**
  - VARCHAR pour les textes.
  - INT pour les identifiants et les mesures quantitatives.

### III. Processus ETL : Extraction, Transformation et Chargement des Données :

#### 1. Définition

Afin d'alimenter notre entrepôt de données (Data Warehouse) à partir de la source initiale, nous avons mis en œuvre un **processus ETL (Extract, Transform, Load)** en utilisant **Talend Open Studio**.

##### a. ETL : Extraction, Transformation, Chargement

Le processus ETL (pour **Extraction, Transformation, Loading**) est une étape essentielle dans la construction d'un Data Warehouse.

Il permet de **transférer et préparer** les données issues de sources variées pour les rendre exploitables à des fins d'analyse.

Le processus ETL se divise en trois étapes principales :

- **Extraction (Extract)**

Cette étape consiste à **récupérer les données** depuis une ou plusieurs sources différentes : bases de données, fichiers plats (CSV, Excel), applications externes, etc. L'objectif est de collecter les données brutes de manière fiable et rapide.

- **Transformation (Transform)**

Une fois extraites, les données sont **nettoyées, filtrées, normalisées et préparées** pour répondre aux exigences du modèle de données du Data Warehouse.

Cela peut inclure :

- Le nettoyage des doublons,
- La correction des incohérences,
- L'encodage de certaines valeurs,
- La création de nouvelles colonnes,
- L'adaptation des formats de données (date, numérique, texte, etc.).

- **Chargement (Load)**

Dans cette dernière étape, les données transformées sont **insérées** dans les tables du Data Warehouse (tables de faits et tables de dimensions).

Le chargement peut se faire en mode **complet** (réécriture totale) ou **incrémental** (ajout ou mise à jour des nouvelles données).

##### b. Talend Open Studio:

**Talend** est un outil d'intégration de données puissant et open-source, principalement utilisé pour concevoir et exécuter des **processus ETL** (Extraction, Transformation, Chargement). Il permet de **connecter, transformer et déplacer** efficacement les données entre différentes sources et cibles (bases de données, fichiers plats, systèmes cloud, etc.).

Caractéristiques principales de Talend :

- **Interface graphique intuitive** basée sur le glisser-déposer de composants,
- **Large bibliothèque de connecteurs** pour divers types de bases de données et de fichiers,
- **Automatisation des flux de données complexes**,

- Gestion des erreurs et contrôle de la qualité des données,
- Capacité à manipuler de gros volumes de données.

Dans notre projet, nous avons utilisé **Talend Open Studio for Data Integration** pour :

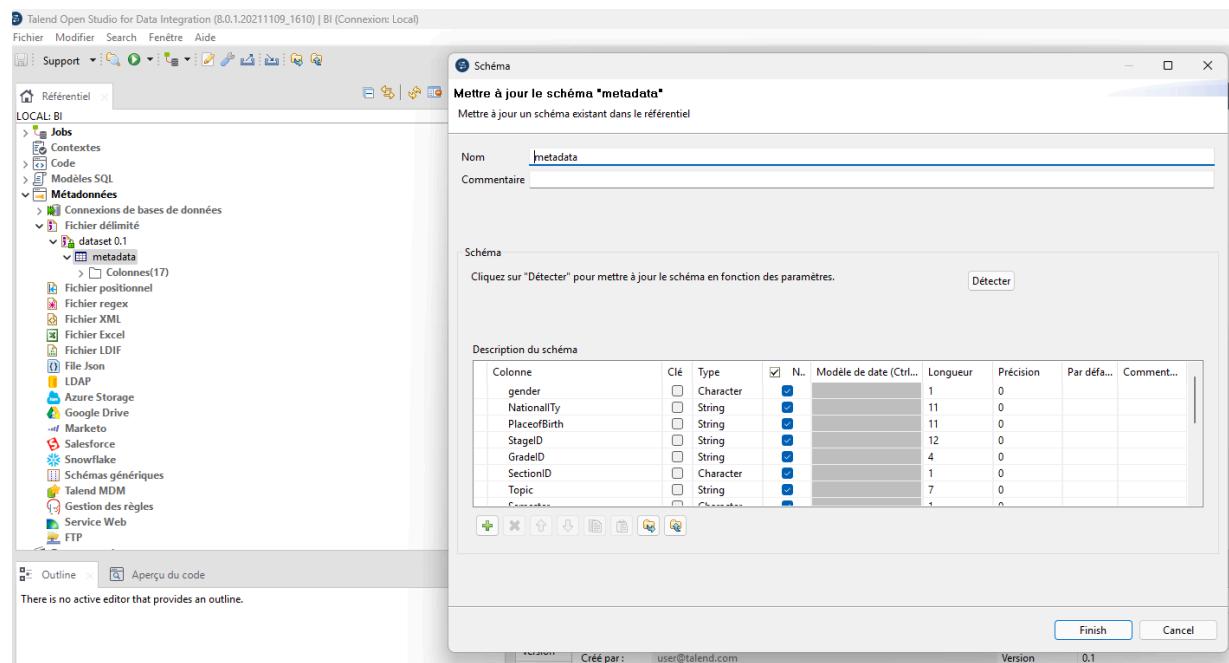
- Extraire les données du fichier CSV de Kaggle,
- Nettoyer, transformer et encoder les données selon notre modèle de Data Warehouse,
- Charger les tables de dimensions et la table de faits dans une base de données **MySQL**.

Grâce à Talend, nous avons pu construire un processus ETL complet de manière simple, rapide et professionnelle.

## 2. Extraction des données

Nous avons utilisé Talend pour lire le fichier source **xAPI-Edu-Data.csv** qui contient toutes les informations académiques des étudiants.

- **Composant utilisé** : `tFileInputDelimited`
- **Description** : Ce composant permet d'importer les données depuis un fichier CSV en configurant correctement le séparateur, l'encodage et les colonnes.



## 3. Transformation des données

Après l'extraction, plusieurs étapes de transformation ont été réalisées pour organiser les données sous forme de tables dimensionnelles et de faits, conformément au modèle en étoile :

- Séparation des informations :

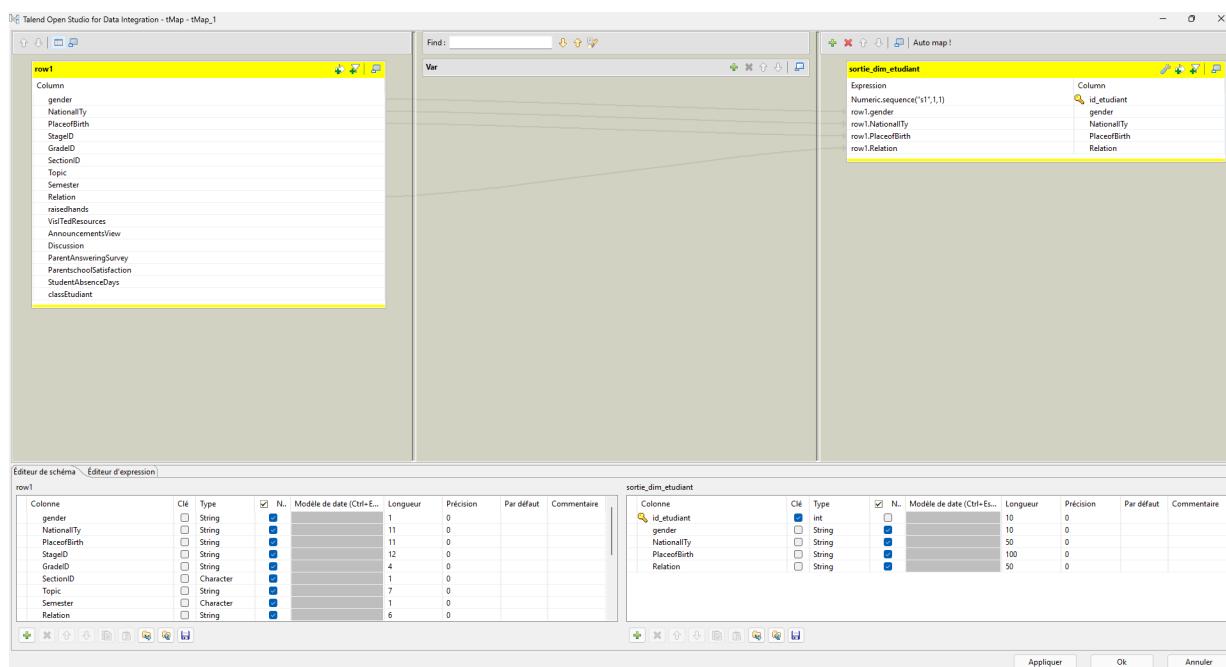
- Utilisation d'un composant **tMap** pour répartir les données extraites dans différentes structures.
- Création d'un fichier de données pour chaque dimension (**dimension\_etudiant**, **dimension\_matiere**, **dimension\_parent**, **dimension\_scolarite**) à partir des colonnes correspondantes du CSV.
- Création d'un fichier pour la table des faits (**faits\_performance**) contenant les mesures de performance.

- Nettoyage et harmonisation :

- Suppression des doublons dans les dimensions pour éviter la redondance des enregistrements.
- Attribution d'un identifiant unique (ID) pour chaque ligne de dimension afin d'assurer les correspondances avec la table des faits.

- Préparation des fichiers MySQL :

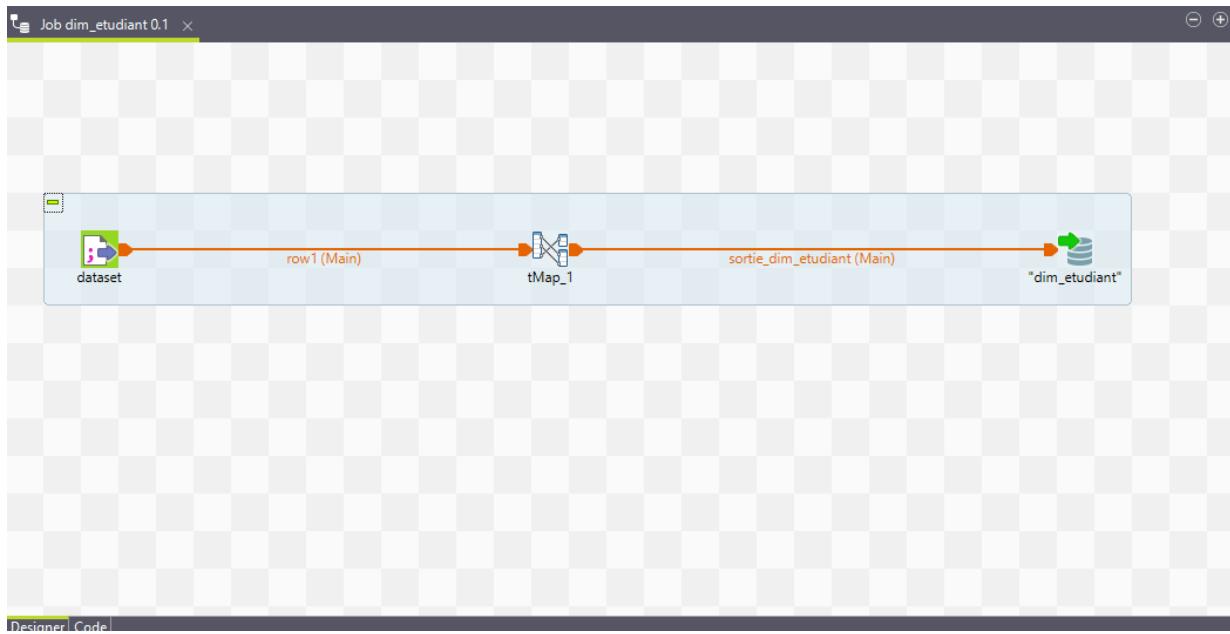
- Chaque table (dimension ou faits) a été exportée sous forme d'un fichier MySQL (fichier **.sql** ou insertion directe dans la base) prêt à être chargé.



## 4. Chargement des données

Après transformation, les données nettoyées ont été chargées dans la base de données MySQL :

- **Composant utilisé** : **tMySQLOutput**
- **Opérations réalisées** :
  - Connexion à la base de données MySQL.
  - Insertion des lignes dans les **tables de dimensions** d'abord, puis dans la **table des faits** pour respecter les clés étrangères.



## 5. Résultat :

À la fin du processus ETL :

- Chaque dimension est correctement remplie avec des identifiants uniques.
- La table des faits référence correctement les identifiants des différentes dimensions.
- La base de données est prête pour être utilisée dans des analyses BI avec Power BI.

	<a href="#">Edit</a>	<a href="#">Copy</a>	<a href="#">Delete</a>	<a href="#">id_etudiant</a>	<a href="#">gender</a>	<a href="#">Nationality</a>	<a href="#">PlaceofBirth</a>	<a href="#">Relation</a>
<input type="checkbox"/>	<a href="#">Edit</a>	<a href="#">Copy</a>	<a href="#">Delete</a>	1	M	KW	Kuwait	Father
<input type="checkbox"/>	<a href="#">Edit</a>	<a href="#">Copy</a>	<a href="#">Delete</a>	1	M	KW	Kuwait	Father
<input type="checkbox"/>	<a href="#">Edit</a>	<a href="#">Copy</a>	<a href="#">Delete</a>	1	M	KW	Kuwait	Father
<input type="checkbox"/>	<a href="#">Edit</a>	<a href="#">Copy</a>	<a href="#">Delete</a>	1	M	KW	Kuwait	Father
<input type="checkbox"/>	<a href="#">Edit</a>	<a href="#">Copy</a>	<a href="#">Delete</a>	1	M	KW	Kuwait	Father
<input type="checkbox"/>	<a href="#">Edit</a>	<a href="#">Copy</a>	<a href="#">Delete</a>	1	M	KW	Kuwait	Father
<input type="checkbox"/>	<a href="#">Edit</a>	<a href="#">Copy</a>	<a href="#">Delete</a>	1	M	KW	Kuwait	Father
<input type="checkbox"/>	<a href="#">Edit</a>	<a href="#">Copy</a>	<a href="#">Delete</a>	1	M	KW	Kuwait	Father
<input type="checkbox"/>	<a href="#">Edit</a>	<a href="#">Copy</a>	<a href="#">Delete</a>	1	M	KW	Kuwait	Father
<input type="checkbox"/>	<a href="#">Edit</a>	<a href="#">Copy</a>	<a href="#">Delete</a>	1	M	KW	Kuwait	Father
<input type="checkbox"/>	<a href="#">Edit</a>	<a href="#">Copy</a>	<a href="#">Delete</a>	1	M	KW	Kuwait	Father
<input type="checkbox"/>	<a href="#">Edit</a>	<a href="#">Copy</a>	<a href="#">Delete</a>	1	M	KW	Kuwait	Father
<input type="checkbox"/>	<a href="#">Edit</a>	<a href="#">Copy</a>	<a href="#">Delete</a>	1	M	KW	Kuwait	Father
<input type="checkbox"/>	<a href="#">Edit</a>	<a href="#">Copy</a>	<a href="#">Delete</a>	1	M	KW	Kuwait	Father

## IV- Visualisation des données sous Power BI

### 1. Introduction : Pourquoi utiliser Power BI ?

Dans un contexte où les décideurs publics et les responsables pédagogiques doivent exploiter des volumes croissants de données (résultats d'examens, taux d'assiduité, indicateurs socio-démographiques...), **Power BI** se distingue comme une solution **self-service BI** puissante et accessible :

- **Rapidité de prise en main** : interface intuitive “glisser-déposer” pour créer tableaux et graphiques en quelques clics.
- **Connectivité native** : connexion en direct à MySQL, CSV, Excel, web API (ex. Kaggle), Data Warehouses...
- **Interactivité** : filtres, slicers, drill-down et “hover-over” facilitent l'exploration instantanée des tendances.
- **Mise à jour automatisée** : actualisation planifiée des données pour des rapports toujours à jour.
- **Partage et collaboration** : publication sur Power BI Service, export PDF/PPT ou intégration dans SharePoint/Teams pour diffusion aux parties prenantes.

### 2. Introduction globale du projet

Dans un contexte où le suivi des performances académiques des étudiants revêt une importance stratégique pour les établissements et les décideurs publics, ce projet vise à mettre en place un **système BI** fondé sur **Power BI**. À partir du jeu de données xAPI-Edu-Data (480 élèves, informations démographiques et indicateurs d'engagement), l'objectif est de fournir un outil interactif capable de :

1. **Mesurer et comparer** les résultats selon diverses dimensions (matière, niveau, genre...),
2. **Explorer** les comportements d'apprentissage (ressources consultées, interactions en classe...),
3. **Anticiper** les risques d'échec et proposer des scénarios d'intervention.

Le dashboard se décline en **trois pages** complémentaires :

- **Page 1 – Vue d'ensemble des performances**

Présentation des indicateurs clés (nombre d'élèves, taux d'absentéisme) et analyses segmentées : moyenne par matière, par niveau scolaire et par genre, ainsi que répartition des élèves selon leur performance. Cette première page sert de point d'ancrage pour évaluer rapidement la santé académique globale.

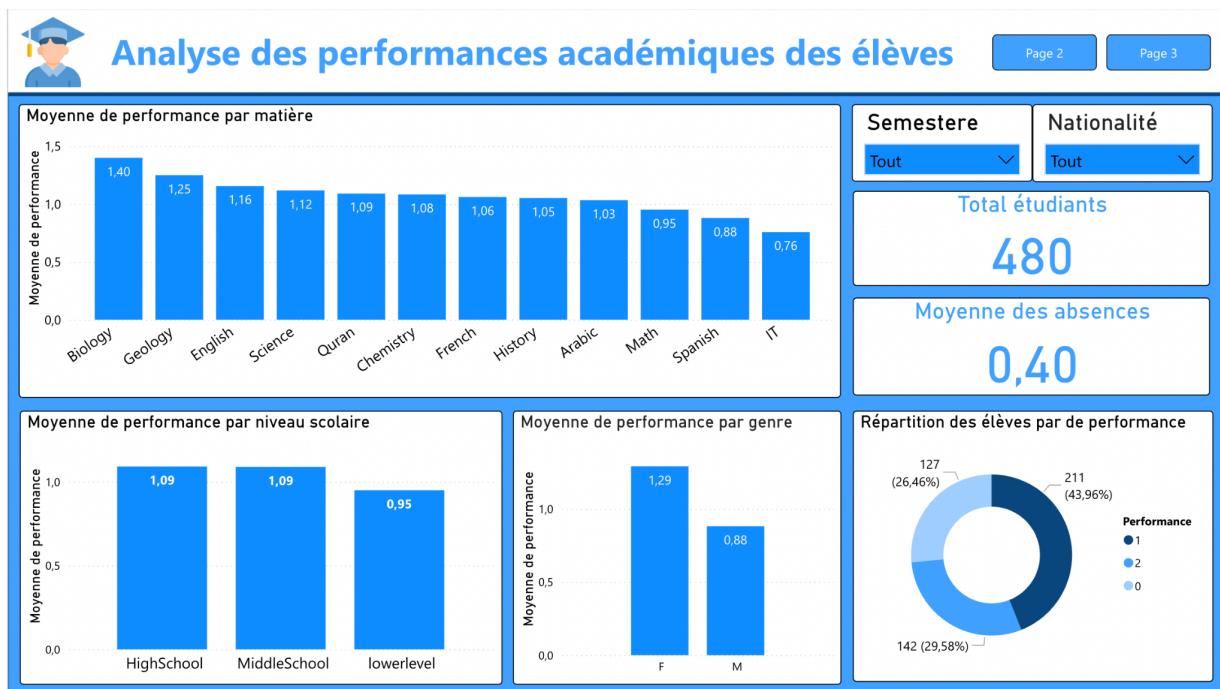
- **Page 2 – Analyse de l'engagement et du comportement**

Exploration des données d'interaction : nombre de fois où l'élève a levé la main, consultation des ressources, participation aux forums et visionnage d'annonces. Des graphiques de type histogramme et heatmap mettent en évidence les corrélations entre engagement et réussite.

- **Page 3 – Modélisation prédictive et scénarios**

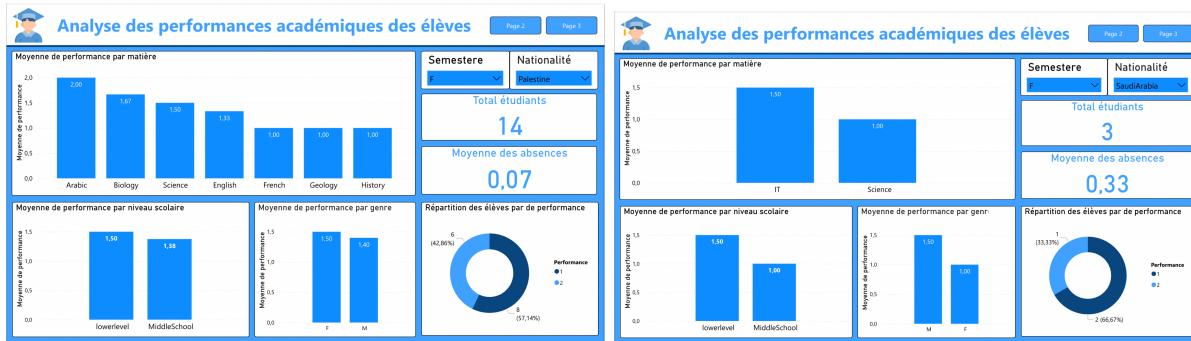
Ce dashboard compare l'impact de l'absentéisme et de l'implication parentale sur la performance scolaire. Il présente des graphiques sur la moyenne de performance selon les réponses des parents, leur satisfaction et le niveau scolaire des élèves. Les données suggèrent un lien entre l'engagement parental et de meilleurs résultats, tandis que l'absentéisme reste stable selon les niveaux. Les visualisations permettent une comparaison rapide entre ces différents facteurs.

## 2. Analyse du première dashboard : Analyse des performances académiques des élèves



Cette première page du dashboard a pour objectif de fournir une vue d'ensemble immédiate et synthétique des performances académiques des élèves, en combinant à la fois les indicateurs globaux et les analyses segmentées. Elle permet d'abord de mesurer la taille de l'échantillon étudié (480 élèves) et d'évaluer le taux moyen d'absentéisme (0,40), afin de situer la fiabilité et la représentativité des résultats. Grâce aux filtres « Semestre » et « Nationalité », l'utilisateur peut ensuite affiner son analyse pour observer l'impact du contexte temporel ou culturel sur les scores. Enfin, les graphiques associés – moyenne par matière, par niveau scolaire et par genre, ainsi que la répartition des élèves selon leur performance – mettent en lumière les points forts et les zones de vigilance : quelles matières nécessitent un soutien renforcé, comment évolue le rendement d'un cycle à l'autre et si des disparités apparaissent entre filles et garçons. Cette page sert donc de point de départ pour orienter les recommandations pédagogiques et guider les explorations plus fines dans les onglets suivants.

Lorsque l'on modifie les filtres « Semestre » ou « Nationalité », l'ensemble des graphiques (moyennes par matière, par niveau scolaire, par genre, indicateurs clés et répartition des performances) se recalculent et s'actualisent instantanément pour ne présenter que les données correspondant aux critères sélectionnés ; comme le montrent les captures ci-dessous, on obtient des graphiques distincts pour l'Arabie Saoudite et la Palestine et pour deux semestres différents.



## 2.1 Analyse graphique

### Moyenne de performance par matière

Ce graphique montre que la Biologie (1,40) et la Géologie (1,25) sont les matières où les élèves obtiennent les meilleures performances. L'Anglais (1,16) complète le podium. À l'inverse, l'Informatique (IT) présente la moyenne la plus basse (0,76), suivie de l'Espagnol (0,88) et des Mathématiques (0,95). Les autres matières (Sciences, Quran, Chimie, Français, Histoire et Arabe) ont des performances moyennes comprises entre 1,03 et 1,12.

### Moyenne de performance par niveau scolaire

Les élèves de lycée (HighSchool) et de collège (MiddleSchool) ont exactement la même moyenne de performance (1,09), tandis que les élèves du niveau inférieur (lowerlevel) présentent une performance plus faible (0,95). Cette différence de performance entre les niveaux est significative mais pas extrême.

### Moyenne de performance par genre

Le graphique révèle un écart notable entre les performances selon le genre : les élèves de sexe féminin (F) ont une moyenne de 1,29, alors que les élèves masculins (M) ont une moyenne de 0,88, soit une différence de 0,41 point.

### Répartition des élèves par performance

Ce diagramme circulaire montre que la plus grande proportion d'élèves (43,96%, soit 211 élèves) obtient un niveau de performance de 1. 29,58% (142 élèves) atteignent un niveau de performance de 2, tandis que 26,46% (127 élèves) sont au niveau 0. La majorité des élèves (73,54%) obtient donc un résultat positif (niveau 1 ou 2).

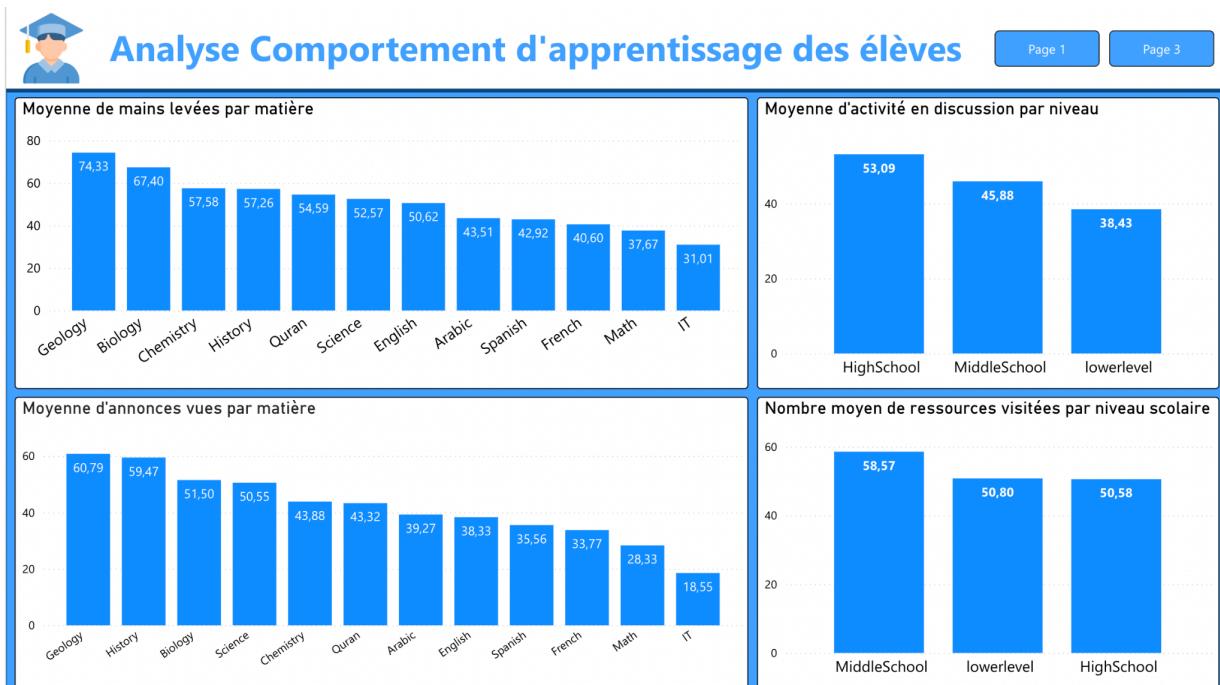
## 2.2 Tendance générale

L'analyse révèle des disparités significatives dans les performances selon les matières et le genre, avec des sciences naturelles (Biologie, Géologie) qui obtiennent les meilleurs résultats, tandis que l'Informatique et l'Espagnol présentent les scores les plus faibles. Les filles surpassent nettement les garçons en termes de performance. La répartition par niveau scolaire montre une progression positive entre le niveau inférieur et les niveaux supérieurs.

## 2.3 Interprétation

- Disparité entre matières** : Les sciences naturelles (Biologie, Géologie) semblent mieux enseignées ou mieux assimilées que les matières techniques (IT) ou certaines langues (Espagnol). Cela pourrait refléter la qualité de l'enseignement, l'intérêt des élèves pour ces matières, ou les ressources disponibles.
- Progression par niveau scolaire** : L'amélioration des performances entre le niveau inférieur et les niveaux supérieurs suggère une adaptation progressive au système éducatif et une consolidation des connaissances au fil du parcours scolaire.
- Écart de genre significatif** : La différence importante de performance entre filles et garçons (1,29 contre 0,88) mérite une attention particulière. Elle pourrait résulter de différences dans les méthodes d'apprentissage, l'engagement scolaire, ou refléter d'autres facteurs sociaux ou éducatifs.
- Distribution des performances** : Avec près de 74% des élèves obtenant des résultats positifs (niveaux 1 et 2), l'établissement semble globalement efficace. Cependant, plus d'un quart des élèves (26,46%) obtiennent un niveau 0, ce qui indique un besoin d'intervention pour ce groupe.
- Taux d'absentéisme** : La moyenne d'absences de 0,40 est modérée et correspond aux données de la première page, où l'on voyait que le taux d'absentéisme variait entre 0,33 et 0,47 selon le niveau scolaire.

### 3. Analyse du deuxième dashboard : Analyse Comportement d'apprentissage des élèves



#### 3.1 Vue globale

Ce dashboard met en lumière des disparités significatives dans l'engagement des élèves selon les matières et les niveaux. Les sciences (hors mathématiques) stimulent davantage la participation orale, tandis que les plus jeunes sont plus réactifs aux annonces et ressources. Ces insights pourraient guider des ajustements pédagogiques pour renforcer l'implication dans les matières moins attractives ou auprès des publics moins actifs.

#### 3.2 Analyse par Graphique

##### 1. Moyenne de mains levées par matière

- Tendance générale :** La participation active (mains levées) est la plus élevée en "Geology" (74,33) et en "Biology" (67,40), tandis que les matières comme "IT" (31,01) et "Math" (37,67) enregistrent les taux les plus bas.
- Interprétation :** Les matières scientifiques (hors mathématiques) suscitent plus d'interactions, contrairement aux disciplines techniques ou abstraites (IT, Math). Cela pourrait refléter un intérêt accru pour les sciences ou une difficulté perçue dans les matières exactes.

##### 2. Moyenne d'activité en discussion par niveau

- Tendance générale :** Les élèves de HighSchool sont les plus actifs dans les discussions, suivis de MiddleSchool et enfin Lowerlevel.

- **Interprétation :** L'engagement dans les échanges augmente avec l'âge, ce qui pourrait s'expliquer par une meilleure aisance à s'exprimer ou par des méthodes pédagogiques adaptées aux niveaux supérieurs.

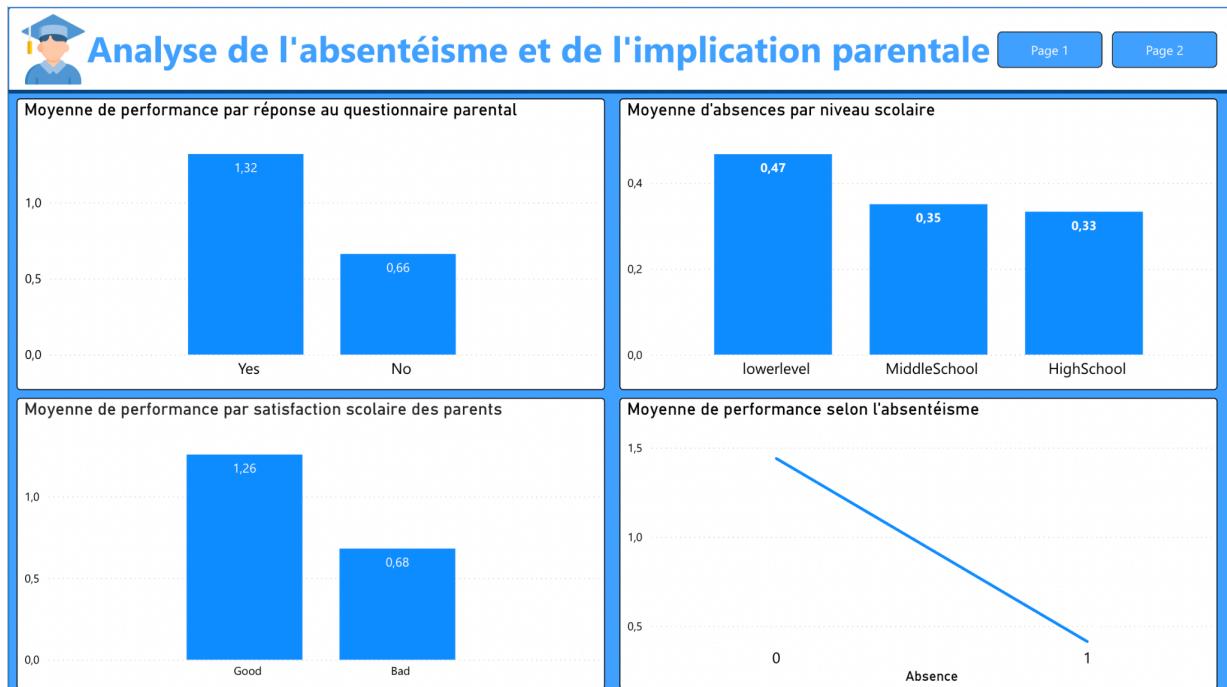
### 3. Moyenne d'annonces vues par matière

- **Tendance générale :** Les annonces sont le plus consultées en MiddleSchool (60,79), devant Lowerlevel (59,47) et HighSchool (51,50).
- **Interprétation :** Les élèves plus jeunes (MiddleSchool et Lowerlevel) semblent plus attentifs aux communications officielles, tandis que les lycéens (HighSchool) pourraient privilégier d'autres canaux d'information ou être moins réactifs.

### 4. Nombre moyen de ressources visitées par niveau scolaire

- **Tendance générale :** Les MiddleSchool (58,57) utilisent légèrement plus de ressources que les Lowerlevel (50,80) et HighSchool (50,58).
- **Interprétation :** La légère supériorité des MiddleSchool pourrait indiquer une utilisation plus systématique des supports pédagogiques, tandis que les HighSchool, bien que autonomes, pourraient se reposer sur d'autres méthodes (révisions ciblées, travaux de groupe).

## 4. Analyse du troisième dashboard : Analyse Comportement d'apprentissage des élèves



## **4.1 Vue globale**

Ces données présentent l'analyse des relations entre l'implication parentale, l'absentéisme scolaire et la performance des élèves. Les graphiques montrent comment les performances des élèves varient selon la participation des parents (questionnaire), leur satisfaction envers l'école, et comment l'absentéisme se distribue par niveau scolaire et affecte les performances.

## **4.2 Analyse graphique**

### **Moyenne de performance par réponse au questionnaire parental**

Ce graphique montre une nette différence de performance entre les élèves dont les parents ont répondu "Yes" au questionnaire (1,32) et ceux qui ont répondu "No" (0,66). Les élèves dont les parents sont impliqués (réponse "Yes") ont une performance deux fois plus élevée.

### **Moyenne d'absences par niveau scolaire**

Le graphique révèle que l'absentéisme diminue progressivement avec le niveau scolaire. Les élèves du niveau inférieur (lowerlevel) ont le taux d'absence le plus élevé (0,47), suivi des élèves de collège (MiddleSchool) avec 0,35, puis des lycéens (HighSchool) avec le taux le plus bas (0,33).

### **Moyenne de performance par satisfaction scolaire des parents**

Les données indiquent que lorsque les parents sont satisfaits de l'école (Good), la performance moyenne des élèves est significativement plus élevée (1,26) que lorsque les parents expriment une insatisfaction (Bad - 0,68).

### **Moyenne de performance selon l'absentéisme**

Ce graphique présente une corrélation négative claire entre l'absentéisme et la performance scolaire. La ligne descendante indique que plus l'absentéisme augmente (de 0 à 1), plus la performance diminue, passant d'environ 1,4 à moins de 0,5.

## **4.3 Tendance générale**

Les données suggèrent fortement que l'implication et la satisfaction parentales sont positivement corrélées avec les performances scolaires des élèves, tandis que l'absentéisme est négativement corrélé avec ces performances. De plus,

l'absentéisme semble diminuer à mesure que les élèves progressent dans les niveaux scolaires.

## Interprétation

1. **Impact de l'implication parentale** : Les élèves dont les parents s'impliquent activement (répondent aux questionnaires) obtiennent des résultats nettement supérieurs. Cela suggère que l'engagement parental est un facteur déterminant dans la réussite scolaire.
2. **Effet de la satisfaction parentale** : La satisfaction des parents envers l'institution scolaire influence positivement la performance des élèves, possiblement en créant un environnement familial plus favorable à l'apprentissage.
3. **Évolution de l'absentéisme** : La diminution de l'absentéisme à mesure que les élèves avancent dans leur parcours scolaire pourrait indiquer une prise de conscience progressive de l'importance de l'assiduité ou une meilleure adaptation au système scolaire.
4. **Relation absentéisme-performance** : La forte corrélation négative entre absentéisme et performance confirme l'importance de la présence régulière en classe pour maintenir un bon niveau académique.

## V- Les trois décisions à l'issue de l'analyse

### Décision à partir du dashboard 1 (Performances académiques)

Les matières « IT » et « Spanish » enregistrent les moyennes les plus basses (0,76 et 0,88), tandis que les sciences naturelles affichent de solides résultats. **Nous décidons de lancer un programme de soutien ciblé** : des modules d'accompagnement et tutorats en petits groupes pour les étudiants en IT et Spanish, afin de remonter rapidement leurs performances au niveau des autres disciplines.

### Décision à partir du dashboard 2 (Comportement d'apprentissage)

On observe un taux de mains levées très faible en IT (31 vs 74 en Géologie) et une moindre consultation de ressources numériques. **Nous décidons d'enrichir les cours avec des activités interactives** : ateliers pratiques, quiz en temps réel et travaux collaboratifs, pour stimuler l'engagement des élèves dans ces matières et augmenter à la fois leur participation orale et leur usage des supports pédagogiques.

### **Décision à partir du dashboard 3 (Absentéisme & implication parentale)**

Les élèves dont les parents n'ont pas répondu au sondage présentent une moyenne de 0,66 contre 1,32 pour ceux dont les parents sont impliqués, et l'absentéisme est le plus élevé en lowerlevel (0,47). **Nous décidons de mettre en place une campagne de relance parentale** : suivi individualisé des familles n'ayant pas répondu, ateliers de sensibilisation à l'importance de l'assiduité et de l'implication, ainsi que des points hebdomadaires pour réduire l'absentéisme dès les premiers niveaux scolaires.

# Conclusion

Au terme de ce projet, nous disposons désormais d'un système de Business Intelligence solide et opérationnel, capable de transformer un simple jeu de données en un outil décisionnel stratégique. Dès l'analyse initiale, il est apparu que le suivi des performances académiques des étudiants au Maroc manquait de rigueur et de réactivité. Notre approche a consisté à centraliser l'ensemble des informations issues du jeu xAPI-Edu-Data, puis à les structurer de manière à répondre précisément aux besoins des décideurs publics.

Le cœur technique de notre solution repose sur Talend Open Studio, qui orchestre l'intégralité du processus ETL. Nous avons d'abord extrait les données brutes depuis le fichier CSV, en configurant un composant d'entrée dédié à la prise en compte des séparateurs et de l'encodage. Puis, à l'aide de mappings, chaque attribut a été acheminé vers la table de faits ou vers l'une des dimensions : Étudiant, Matière, Parent ou Scolarité. Cette étape a été l'occasion de nettoyer les doublons, d'uniformiser la casse et de traduire les valeurs textuelles en identifiants numériques, garantissant ainsi la cohérence et la performance des requêtes ultérieures.

Une fois les jeux de données préparés, Talend a chargé les tables de dimensions dans MySQL avant de peupler la table faits\_performance, en respectant strictement l'ordre des dépendances. Nous avons mis en place un mode de chargement incrémental, permettant d'actualiser quotidiennement le Data Warehouse sans pour autant perdre l'historique des données. Ce dispositif assure un temps d'exécution maîtrisé et une disponibilité permanente des indicateurs pour les analyses.

La couche de visualisation, réalisée avec Power BI, offre un accès simple et interactif aux décisions. Trois pages de rapports permettent d'explorer tour à tour la vue d'ensemble des performances, l'engagement des élèves dans leurs activités et l'impact de l'implication parentale sur la réussite. Grâce aux filtres dynamiques et aux fonctionnalités de drill-down, l'utilisateur peut isoler un semestre, une matière ou une zone géographique pour affiner ses diagnostics en quelques clics.

Enfin, ce projet ouvre la voie à de nombreuses évolutions : l'intégration de nouveaux jeux de données (profil socio-économique, ressources pédagogiques externes), l'enrichissement des rapports par des modèles de machine learning pour prédire les risques d'échec, ou encore l'automatisation complète des mises à jour et des alertes via Power BI Service. En fournissant un cadre technique et

méthodologique robuste, cette solution BI contribue dès aujourd’hui à une meilleure compréhension de la dynamique scolaire et prépare l’avenir de la prise de décision éducative au Maroc.