



Rapport détaillé



Analyse des données
Éducation, Emploi et Migration
en Tunisie

2025

Préparé par:

Salim Mgaidi A05 – Sami Ferchichi A07 – Iheb Mokhtar Ben Chebil A05 – Farah Ouni A02
Raslen Korbosli A02 – Islem Belarbi A02 – Yacine Ayari A05 – Moetaz Marzouki A06
– Firas Bessioud A01 – Emna Ajmi A03

1. Introduction

Ce projet vise à fournir à l'association un **tableau de bord analytique et interactif** permettant de suivre les indicateurs clés dans les domaines de l'éducation, de l'emploi et de la migration en Tunisie.

Les données sont extraites depuis le **portail officiel de l'INS** et transformées via un processus ETL avant d'être chargées dans un entrepôt de données modélisé selon la **méthodologie Kimball** (star schéma).

Objectifs :

- Identifier les zones géographiques et populations nécessitant un soutien.
- Visualiser les indicateurs clés (KPI) par gouvernorat, district, genre, milieu et âge.
- Fournir un outil décisionnel interactif via **Power BI**.

- ***Besoins non fonctionnels du système :***

Le système décisionnel doit respecter les exigences suivantes :

- 1- Sécurité : contrôle d'accès (utilisateurs internes/externes), anonymisation si données individuelles utilisées.
- 2- Performance : réponses interactives (< 2s pour filtres usuels) pour visualisations groupées sur un entrepôt dimensionnel optimisé.
- 3- Fiabilité : assurer la cohérence et la qualité des informations affichées.
- 4- Évolutivité : permettre l'ajout de nouvelles données ou indicateurs sans modification majeure.
- 5- Accessibilité : offrir une interface claire, intuitive et accessible sur plusieurs supports.
- 6- Mise à jour : permettre l'actualisation régulière des données du portail INS.
- 7- Qualité : veiller à la précision, la cohérence et la propreté des données.
- 8- Traçabilité : conserver un historique des opérations d'intégration et des mises à jour.
- 9- Interopérabilité : utilisation de SSMS/SSIS/Power BI ; formats standards (CSV, JSON, XML) pour l'ingestion.
- 10- Portabilité & Disponibilité : sauvegardes régulières de la base DW et du staging.

11- Conformité temporelle : horodatage et versioning des données (date d'extraction, période statistique).

12- Robustesse ETL : reprise sur erreur, logs détaillés, surveillance des jobs SSIS.

- ***Besoins fonctionnels du système :***

Les besoins fonctionnels décrivent ce que le système doit permettre :

1-Visualiser les taux de scolarisation par gouvernorat, délégation/district (Nord/Centre/Sud), genre, milieu (urbain/rural) et niveau d'instruction (primaire/secondaire/supérieur).

2-Identifier les zones et populations à risque : délégations/âges/genres/secteurs avec taux de chômage élevé.

3-Cartographier la migration : nombre de migrants par délégation, motifs (emploi, études, mariage...), âge, genre, milieu.

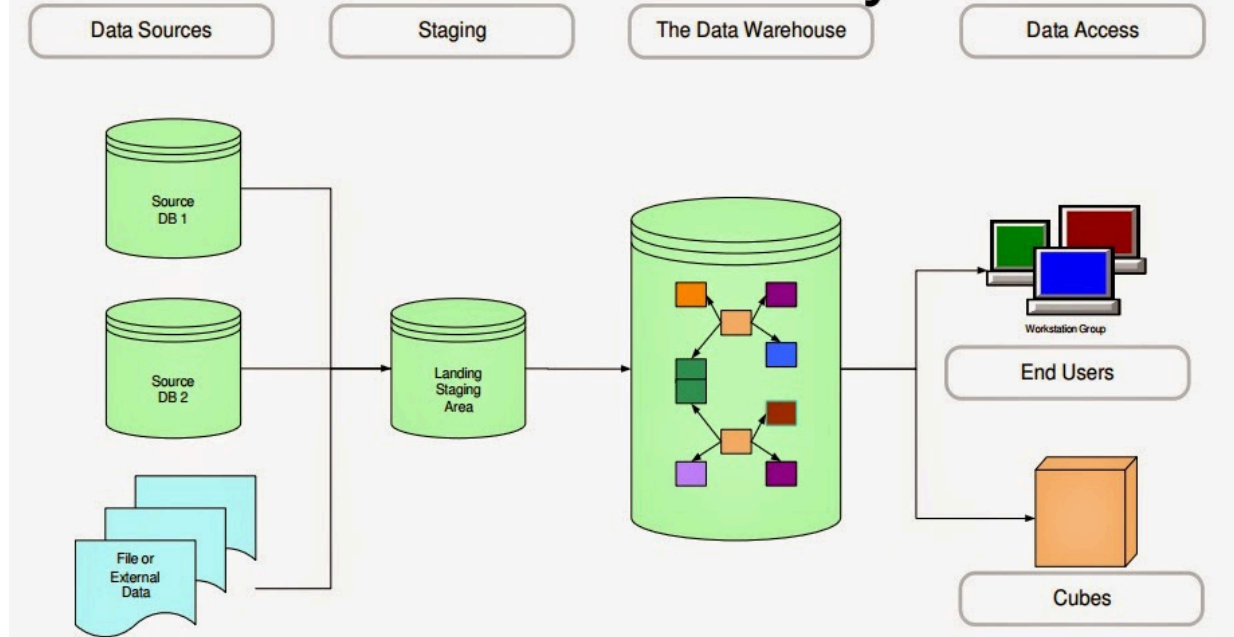
4-Filtrer dynamiquement : choix d'une période (année/trimestre), drill-down (gouvernorat → délégation → localité), et slices par genre, âge, milieu.

2. Modélisation multidimensionnelle – Méthodologie Kimball

Nous allons opter pour la méthodologie **Kimball** (Bottom-up).

- **Pourquoi :** Elle est idéale pour les projets BI itératifs et centrés sur les besoins utilisateurs (Data Marts).
- **Structure :** Constellation car nous avons plusieurs processus métiers : Éducation, Emploi, Migration qui partagent des dimensions communes.

The Kimball Data Lifecycle



2.1 Tables de faits et mesures

Nous allons définir **7 Tables de Faits**

Elles contiennent les métriques (chiffres) et les clés étrangères vers les dimensions.

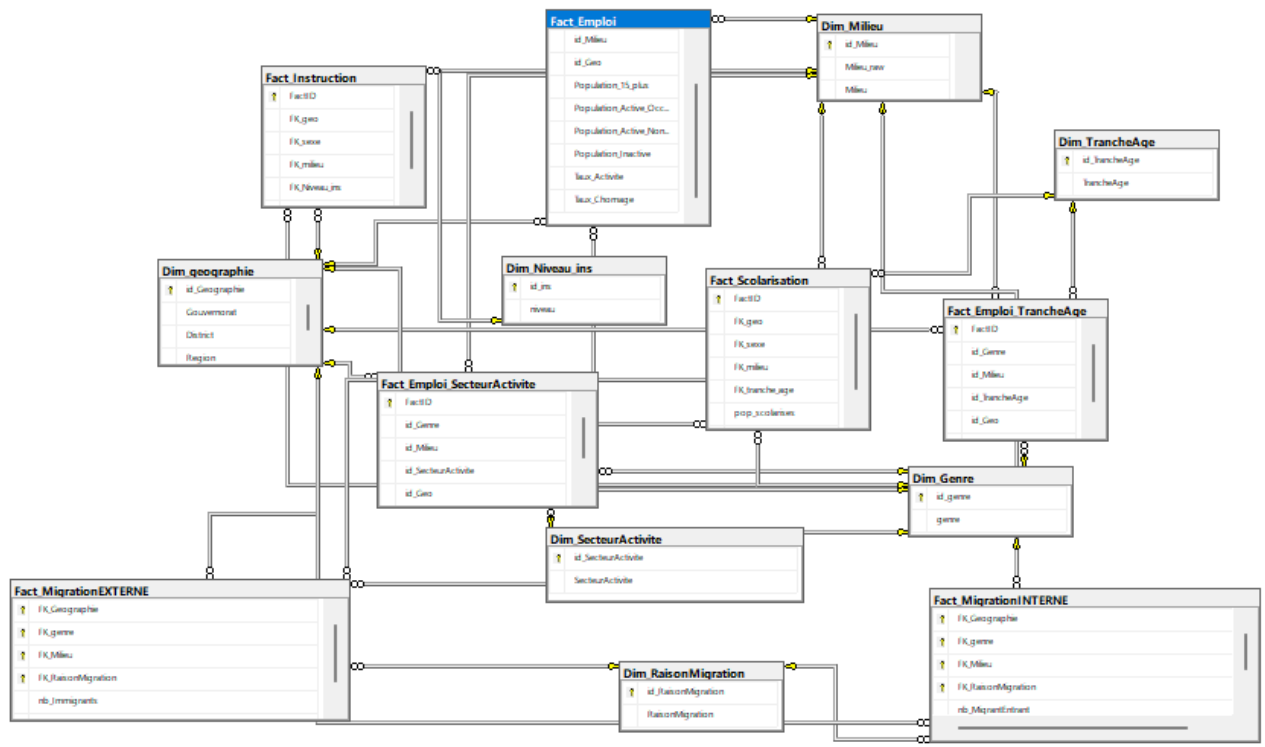
Table de faits	Clés étrangères (FK)	Mesures (chiffres)	Remarques / KPI Power BI
Fait_Education Fait_Instruction	Id_Geo, Id_sexe, Id_millieu,Id_tranche_age, ID_Niveau	effectifs (int) Population_scolarisée(int) population non scolarisée(int)	Taux_Scolarisation calculé dans Power BI : Nombre_Scolarises / Population_Totale_Age
Fait_Emploi Fait_Emploi_TrancheAge Fait_Emploi_SecteurActivite	ID_Geo, ID_sexe, ID_millieu, ID_Secteur,id_tranche_age	Nombre_Chomeurs (int) Population_Active (int)	Taux_Chomage calculé dans Power BI : Nombre_Chomeurs / Population_Active
Fait_MigrationInterne Fait_MigrationExterne	ID_Geo, ID_genre, ID_milieu, ID_Raison	nb_immigrants /nb_Emigrants nb_Entarnts / nb_Sortants	Nombre total de migrants Nombre Immigrants / Émigrants (pour la migration internationale). Nombre Entrants / Sortants (pour la migration interne). Migrants par gouvernorat Raison de migration (emploi/ études/mariage...)

2.2 Tables de dimensions et attributs

Ces tables contiennent le contexte de l'analyse (Qui, Où, Quand, Quoi).

Dimension	Clé primaire (PK)	Attributs / Description	Remarques
Dim_Geographie	ID_Geo (int)	<ul style="list-style-type: none"> - Gouvernorat (nvarchar) - District (nvarchar, ex: Nord/Centre/Sud) - Delegation (nvarchar) - Milieu (nvarchar, Urbain/Rural) 	Dimension partagée pour toutes les analyses spatiales
Dim_trancheAge	id_trancheAge	tranche age	
Dim_sexe	ID_sexe	masculin/feminin	
Dim_milieu	ID_milieu (int)	- milieu milieu (Urbain/Rural),	
Dim_Niveau_Instruction	ID_Niveau (int)	- Libelle_Niveau (nvarchar, Primaire, Secondaire, Supérieur)	Spécifique au Fait_Education
Dim_Secteur_Activite	ID_Secteur (int)	- Nom_Secteur (nvarchar)	Spécifique au Fait_Emplois
Dim_Raison_Migration	ID_Raison (int)	- Type_Raison (nvarchar, Études, Mariage, Emploi...)	Spécifique au Fait_Migration

2.3 Schéma en Constellation (Galaxy Schema)



Chaque table de faits contient les mesures quantitatives et les clés étrangères vers les dimensions correspondantes. Les dimensions fournissent les attributs analytiques pour filtrer et segmenter les données.

3. Conception ETL

3.1 Extraction

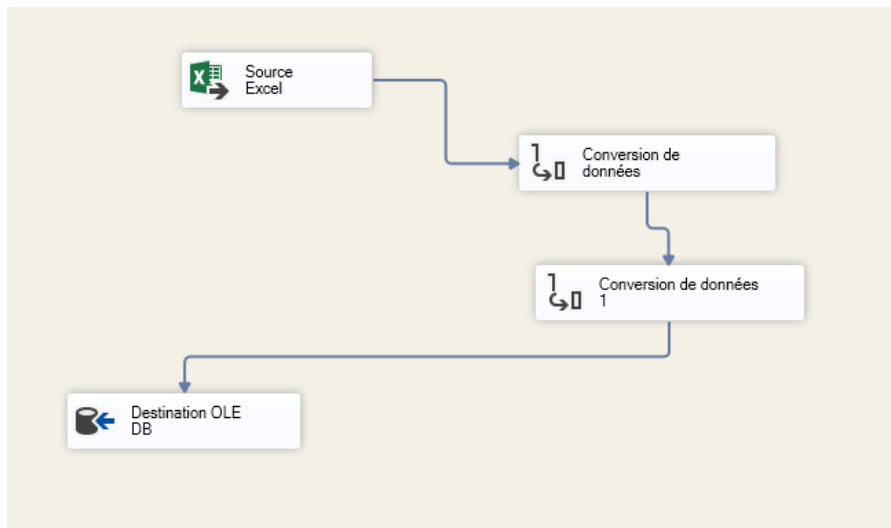
- Données provenant du portail INS ([INS Data Portal](#))
- Formats : CSV, Excel

3.2 Transformation

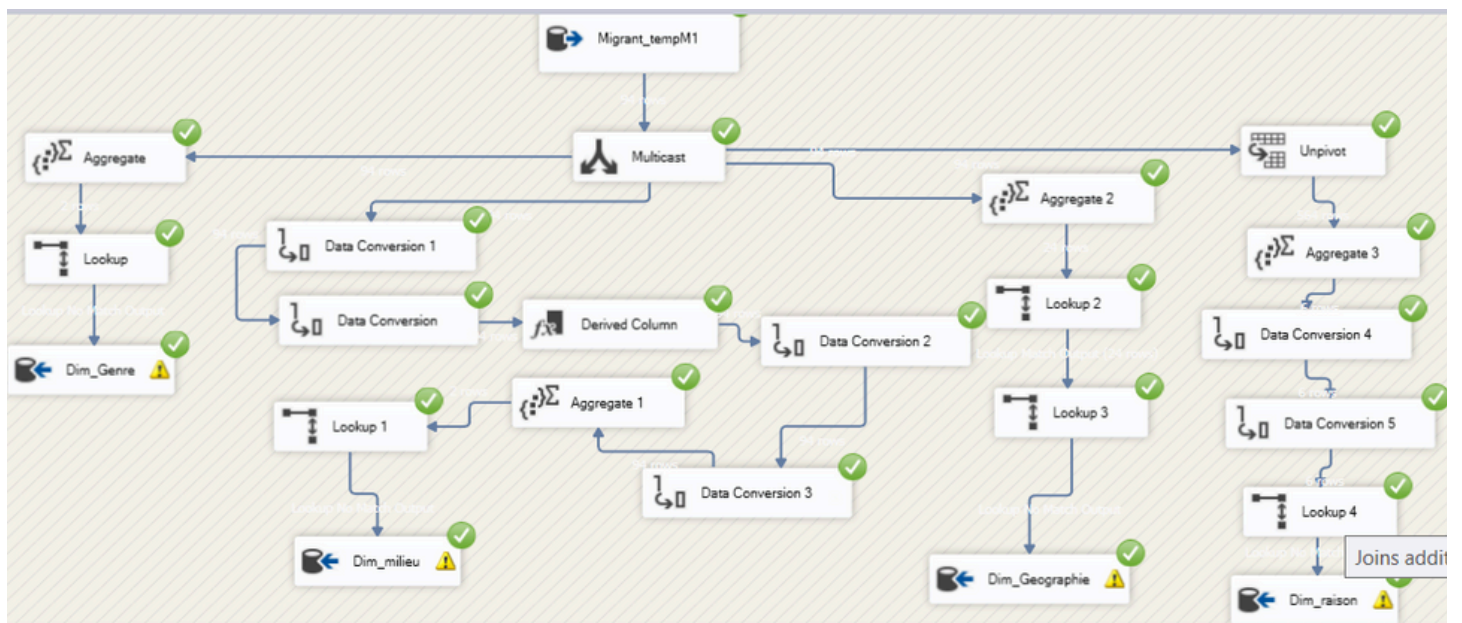
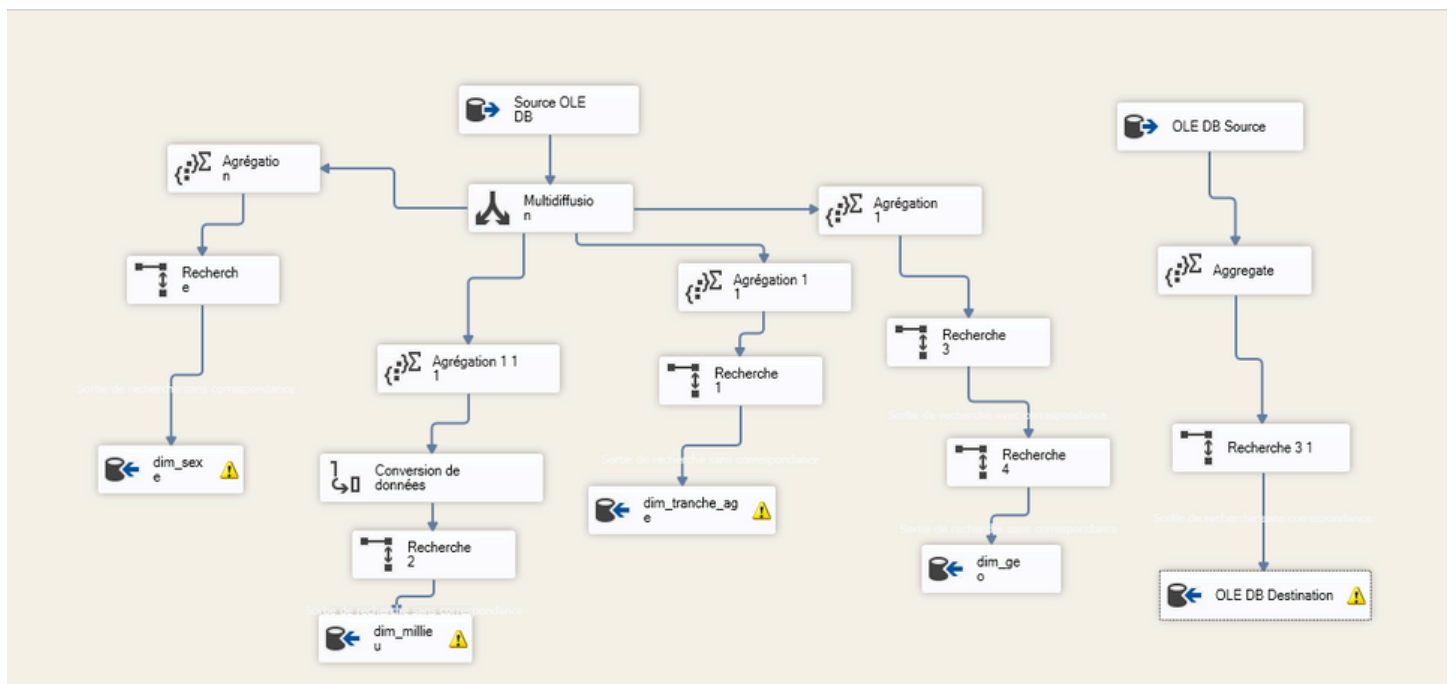
- Nettoyage des doublons et valeurs manquantes
- Harmonisation des formats (texte, codes)

3.3 Chargement

- **Staging area** : stockage temporaire des données brutes et transformées

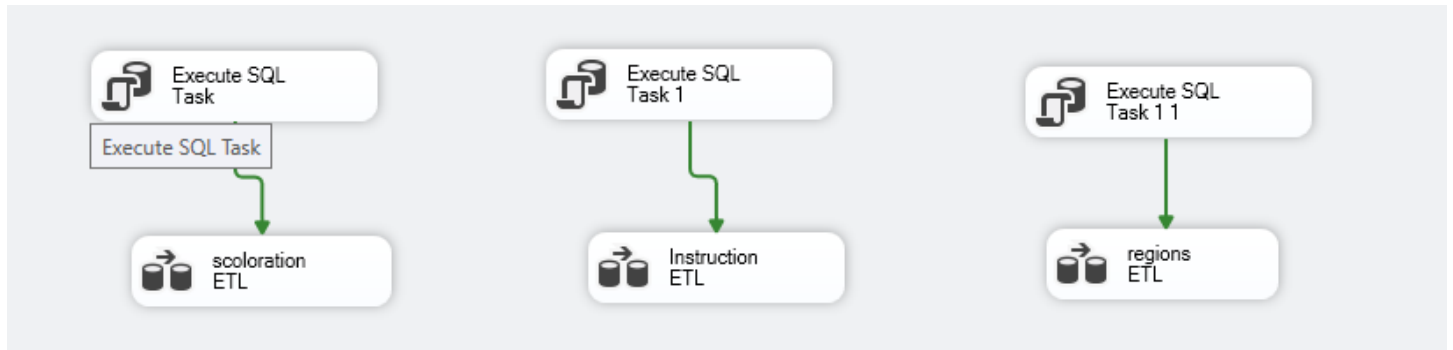


- **Entrepôt Kimball** : insertion dans les tables de faits et dimensions pour analyse multidimensionnelle



3.4 Automatisation avec SSIS

- Création de packages SSIS pour automatiser l'extraction, la transformation et le chargement
- Planification des flux ETL pour mise à jour régulière



4. Définition des KPI

KPI Education:

Mesurent la répartition de la population **selon les caractéristiques sociales** :

- Par **niveau d'instruction**
- Par **sexe**
- Par **milieu**
- Par **région / district / gouvernorat**

Mesurent la **scolarisation** de la population :

- **Taux de scolarisation**
- Population scolarisée par **âge**
- Par **sexe, milieu, région / district / gouvernorat**

KPI Emploi:

- Taux Chômage **par Milieu / par délégation/ par gouvernorat / par genre**
- Population Inactive par Gouvernorat
- Population Chômeur par Tranche Age

KPI Migration:

- Nombre total de migrants
- Nombre Immigrants / Émigrants (pour la migration internationale).
- Nombres Entrants / Sortants (pour la migration interne).
- Migrants par gouvernorat
- Raison de migration (emploi/études/mariage...)

5. Tableau de Bord Power BI

Définition:

La méthode GIMSI structure la création d'un tableau de bord en 4 étapes logiques pour garantir que la technique répond bien au besoin stratégique :

1. **Identification (Le Pourquoi)** : On analyse la stratégie de l'entreprise et les besoins des décideurs (ex : définir l'objectif de réduction du chômage).
2. **Conception (Le Quoi)** : On définit sur papier les indicateurs (KPIs) et les graphiques nécessaires pour suivre cet objectif.
3. **Mise en œuvre (Le Comment)** : C'est la réalisation technique : création de l'entrepôt de données (SQL), des flux (SSIS) et des interfaces (Power BI).
4. **Permanente (Le Suivi)** : On assure la vie du projet par la mise à jour automatique des données et l'évolution du tableau de bord.

En résumé : GIMSI permet de ne pas foncer tête baissée dans le code, mais de s'assurer d'abord que ce qu'on va développer sera vraiment utile pour la stratégie de l'association

5.1 Filtres interactifs

- Gouvernorat, district, genre, milieu (urbain/rural), tranche d'âge, niveau d'instruction

millieu

Communal

Non Communal

sexe

Féminin

Masculin

discret

Centre

Nord

Sud

region

Centre-Est

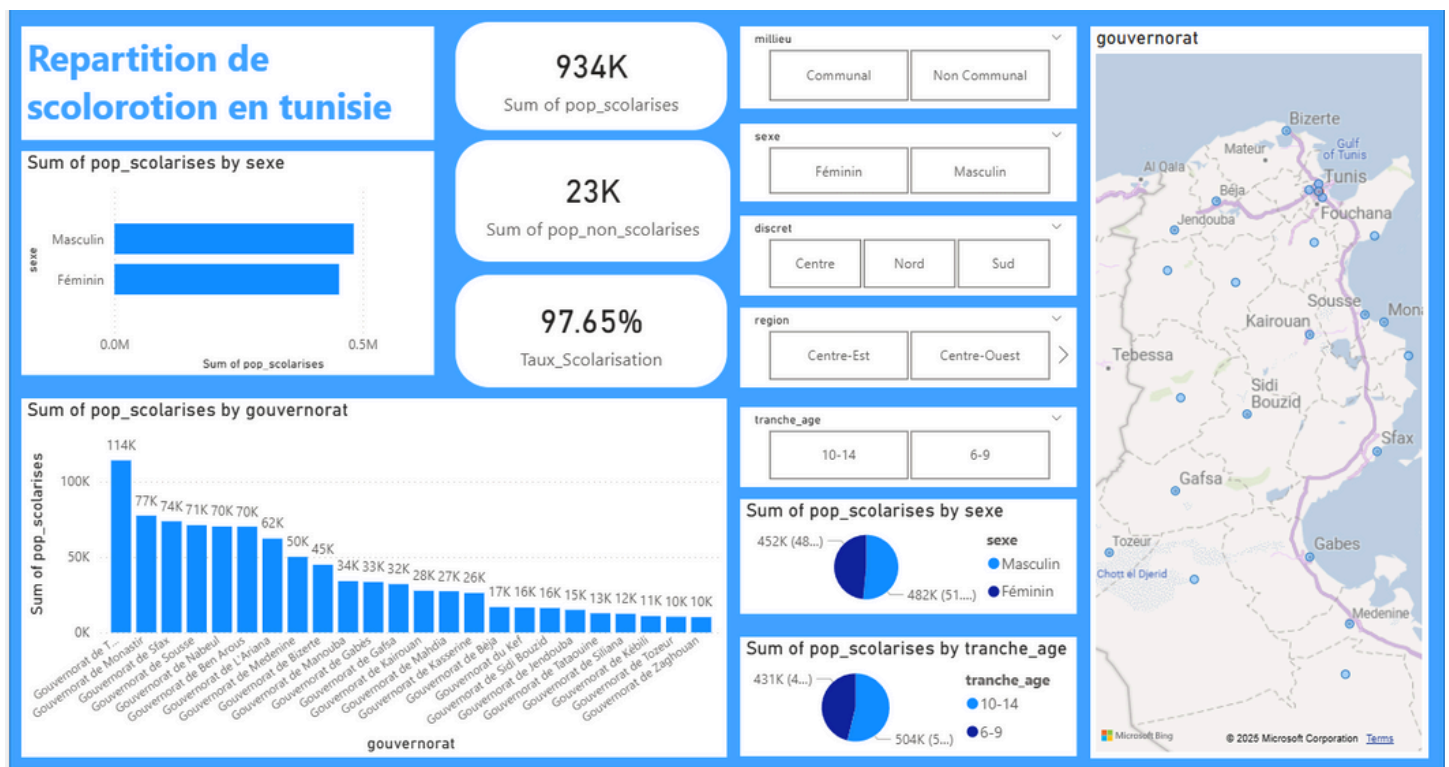
Centre-Ouest

tranche_age

10-14

6-9

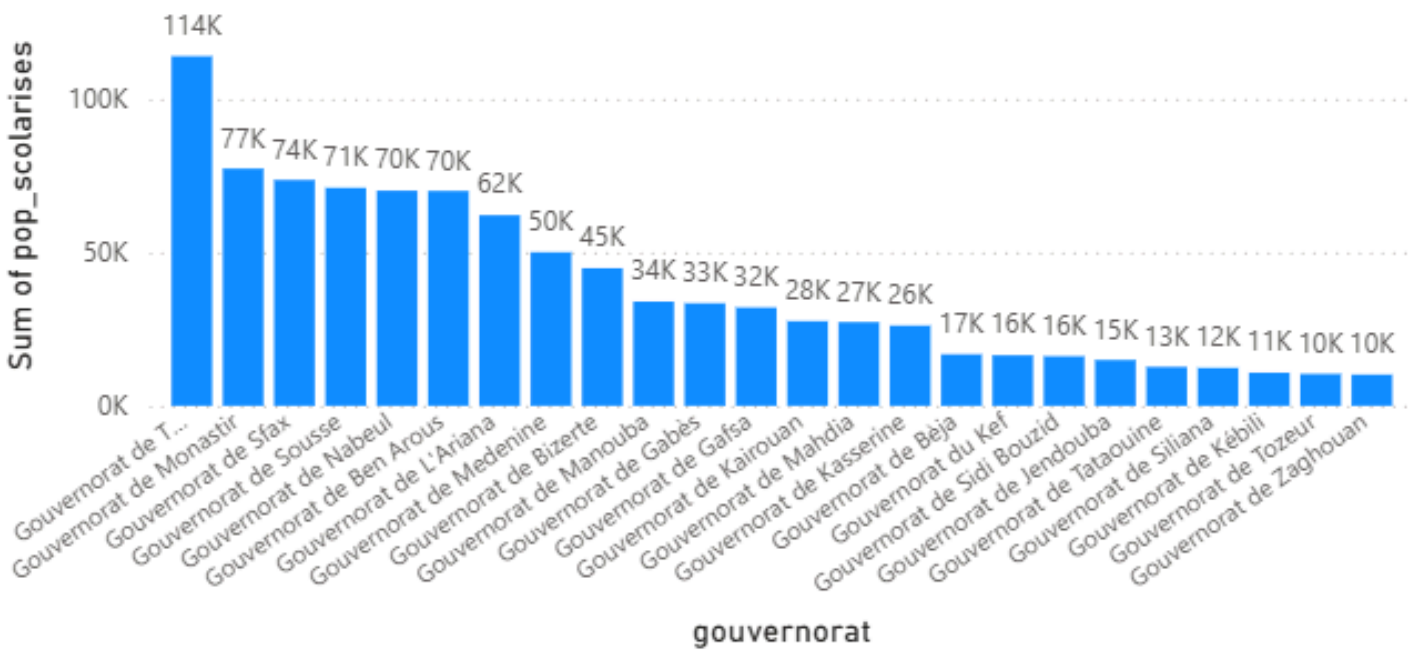
5.2 Visualisations:



- **Graphiques en barres** : comparaison des taux par gouvernorat.

- **Carte géographique** : répartition des indicateurs par région.
- **KPI Cards** : affichage des taux globaux et nombres totaux.

Sum of pop_scolarises by gouvernorat



5.39M

Sum of effectifs_ins

2M

Effectif_Par_Niveau

3M

Effectif_Par_Niveau s...

961K

Effectif_Par_Niveau_...

5.3 Fonctionnalités avancées

- Drill-down pour zoom sur le district depuis le gouvernorat
- Interactions entre visualisations pour analyse croisée
- Possibilité de filtrer par genre, tranche d'âge, niveau ou secteur

6. Conclusion

Grâce à la **méthodologie Kimball** :

- L'entrepôt est structuré et optimisé pour l'analyse multidimensionnelle.
- Les données de l'INS sont centralisées et nettoyées pour fiabilité.
- Le tableau de bord Power BI fournit un outil décisionnel interactif, permettant de suivre les KPI et d'orienter les actions de l'association en matière d'éducation, emploi et migration.