**Documentation du Projet : Visualisation des Données de Population avec Power BI**

**1. Contexte et Objectifs**

**Contexte**

Le projet consiste à transformer des données démographiques brutes issues d'une base de données MariaDB (population\_bronze) en une table exploitable (population\_gold) pour analyser et visualiser des tendances de population à l’aide de Power BI.

Les analyses se concentrent sur :

* Les hommes en 1990, filtrés par région.
* Une étude spécifique sur les femmes dans la région 01.
* Une comparaison combinée des hommes et des femmes par région et par année.

**Objectifs**

* Nettoyer et transformer les données brutes pour les rendre exploitables.
* Effectuer des analyses et visualisations interactives sur des indicateurs clés comme la répartition par sexe, région, et période.
* Offrir une interface conviviale pour l'exploration des données via Power BI.
* Faciliter la prise de décision à partir des insights générés.

**2. Architecture et Flux de Données**

**Étapes Principales**

1. **Table Bronze** : Import des données brutes (population\_bronze) contenant des champs variés comme le sexe, l’âge, la région, etc.
2. **Table Silver** : Nettoyage et transformation, incluant :
   * Suppression des valeurs inconnues (\_T) pour les champs AGE et SEX.
   * Filtrage sur les régions définies par des codes postaux français.
3. **Table Gold** : Agrégation des données pour produire des indicateurs exploitables comme la population totale (total\_population) et le pourcentage de population (population\_percent).
4. **Visualisation dans Power BI** : Création de rapports interactifs pour explorer les données.

**3. Explications Techniques par Étape**

**Étape 1 : Nettoyage et Transformation (Population Silver)**

* **Objectif** : Créer une version nettoyée des données brutes pour éliminer les anomalies et préparer une structure cohérente.
* **Actions** :
  + Filtrage des valeurs \_T pour AGE et SEX (valeurs manquantes ou inconnues).
  + Retrait des régions qui ne correspondent pas aux codes postaux français.
  + Structure finale de la table population\_silver :

sql

Copy code

CREATE TABLE population\_silver (

GEO VARCHAR(10) NOT NULL,

SEX CHAR(1) NOT NULL,

AGE VARCHAR(20) NOT NULL,

TIME\_PERIOD YEAR NOT NULL,

OBS\_VALUE INT NOT NULL,

POPULATION\_PERCENT FLOAT NOT NULL

);

**Étape 2 : Agrégation et Calculs (Population Gold)**

* **Objectif** : Résumer les données pour permettre des analyses rapides et pertinentes.
* **Actions** :
  + Agrégation par région (GEO), sexe (SEX), et année (TIME\_PERIOD).
  + Calcul de la population totale (total\_population) et du pourcentage de population (population\_percent).
  + Ajout de labels (sex\_label) pour une meilleure lisibilité.
  + Structure finale de la table population\_gold :

sql

Copy code

CREATE TABLE population\_gold (

GEO VARCHAR(10) NOT NULL,

SEX CHAR(1) NOT NULL,

TIME\_PERIOD YEAR NOT NULL,

total\_population INT NOT NULL,

population\_percent FLOAT NOT NULL,

sex\_label VARCHAR(10) NOT NULL

);

**Étape 3 : Visualisation dans Power BI**

* **Objectif** : Exploiter les données agrégées pour générer des visualisations interactives et des tableaux de bord.
* **Avantages de Power BI** :
  + Interface conviviale pour la création de rapports.
  + Capacités de filtrage et exploration intuitive.
  + Intégration directe avec MariaDB pour des mises à jour dynamiques.

**4. Création des Visualisations dans Power BI**

**Étapes pour la Visualisation**

1. **Connexion à la Base MariaDB** :
   * Ouvrir Power BI Desktop.
   * Aller dans **Obtenir des données > Base de données MySQL**.
   * Connecter la table population\_gold.
2. **Configuration des Champs** :
   * Vérifier les colonnes importées : GEO, SEX, TIME\_PERIOD, total\_population, population\_percent, sex\_label.
3. **Création des Visualisations** :
   * **Visualisation des Hommes en 1990 du 01**:
     + Type : **Graphique en barres**.
     + Axe X : GEO.
     + Axe Y : total\_population.
     + Filtres : TIME\_PERIOD = 1990, SEX = 'M' GEO=01.
   * **Étude sur les Femmes de la Région 01** :
     + Type : **Graphique en colonnes empilées**.
     + Axe X : TIME\_PERIOD.
     + Axe Y : total\_population.
     + Filtres : SEX = 'F', GEO = '01'.
4. **Ajout de Cartes Géographiques** :
   * Type : **Carte**.
   * Localisation : GEO.
   * Taille des bulles : total\_population.
5. **Tableaux Croisés Dynamiques** :
   * Créer des tables pour explorer les populations par :
     + Région (GEO).
     + Sexe (SEX).
     + Période (TIME\_PERIOD).

**5. Résultats Attendues**

1. **Rapport interactif** permettant d’explorer les données démographiques.
2. **Visualisations spécifiques** pour répondre aux cas d’usage décrits :
   * Analyse des hommes en 1990.
   * Étude des femmes dans une région particulière.
   * Comparaison globale des sexes par région et année.

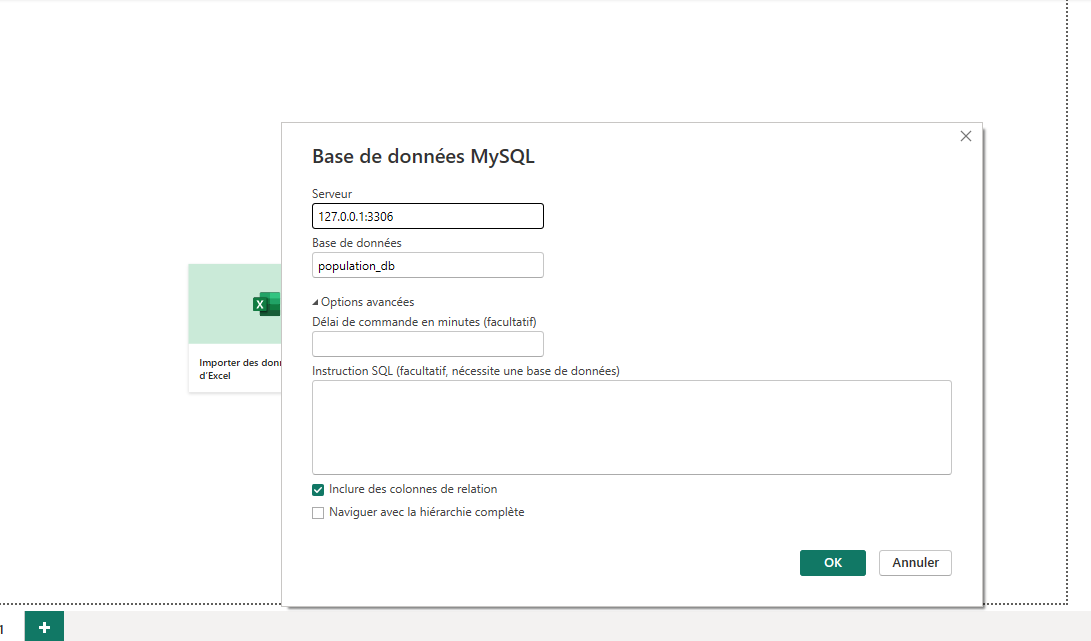
**6. Améliorations Futures**

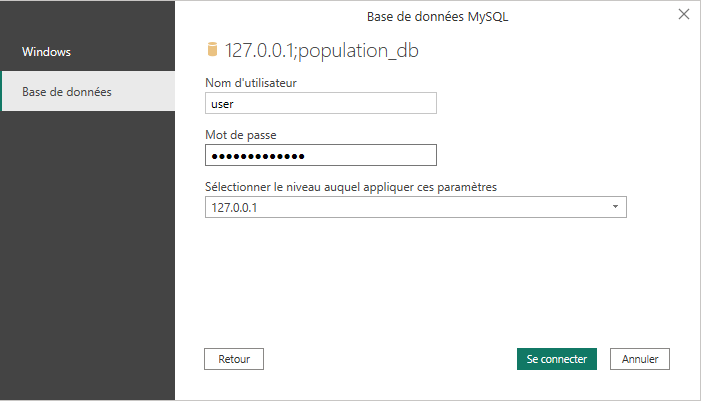
* Ajouter des prédictions basées sur les tendances historiques.
* Intégrer des données supplémentaires (économie, éducation, etc.).
* Automatiser l’actualisation des données via une API.

**Conclusion**

Ce projet permet d’exploiter efficacement les données démographiques pour fournir des insights précieux. Power BI offre une plateforme puissante pour visualiser ces informations de manière intuitive, aidant ainsi à la prise de décision basée sur les données.

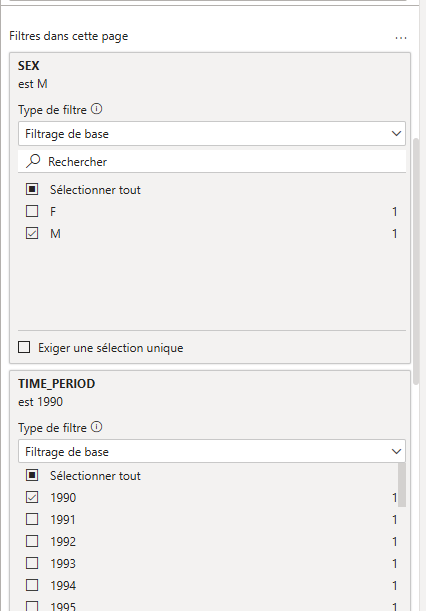
Visualisation dans power bi

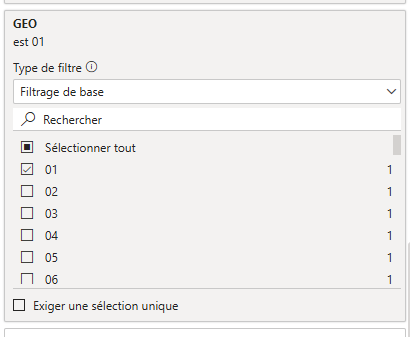


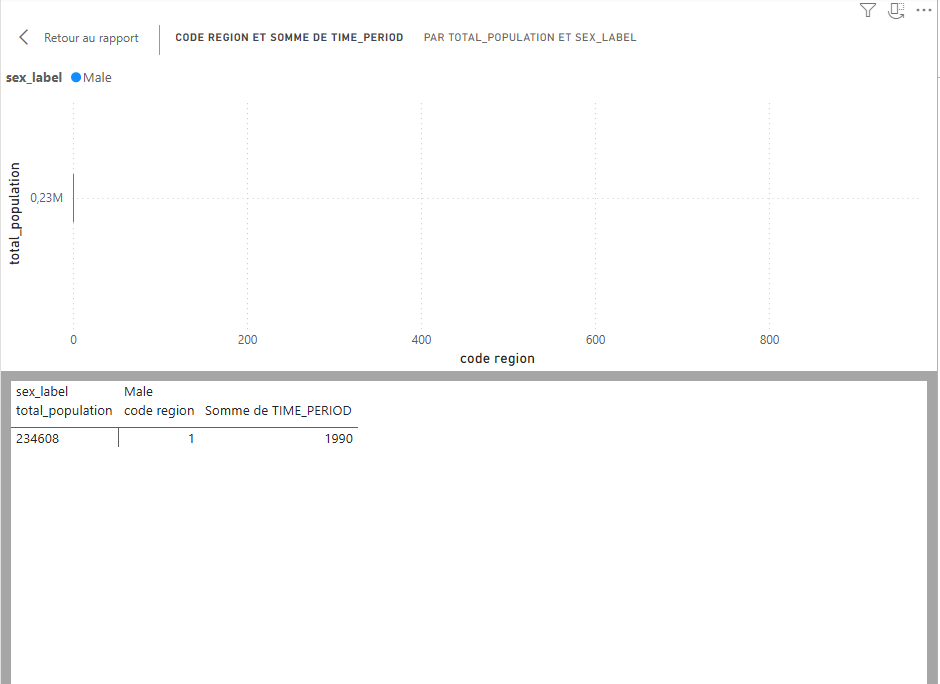


**Étude sur les hommes dans la région 01 en 1990**

* **Filtre** :
  + Ajoutez un **Segment** ou un **Filtre** pour GEO avec la valeur 01.
  + Filtrez sur SEX = 'F' (femmes).







**2. Étude sur les femmes dans la région 01**

* **Filtre** :
  + Ajoutez un **Segment** ou un **Filtre** pour GEO avec la valeur 01.
  + Filtrez sur SEX = 'F' (femmes).
* **Visualisation** :
  + Utilisez un **Graphique en barres empilées** ou un **Tableau croisé dynamique**.
  + Axe X : TIME\_PERIOD.
  + Axe Y : total\_population.

