



Un projet d'analyse multidimensionnelle pour la compréhension des facteurs liés au suicide

Data Warehouse

Réalisé par :

Khadija Ben Madane

Basma EL Barki

Jamaa Sahraoui

Encadré par :

Hassan Badir

10 Décembre 2024

Table des matières

1	Introduction	4
2	Problématique	4
3	Spécifications Techniques : Logiciels Utilisés	5
4	Spécifications fonctionnelles	5
4.1	Introduction au Data Warehouse (DW)	5
4.2	Objectif du DW dans ce projet	6
4.3	Sources de Données	6
4.3.1	Données sur les troubles de santé mentale (MHD.xlsx)	6
4.3.2	Données sur le bonheur (bonheur.json)	7
4.3.3	Données sur les suicides (suicide.csv)	9
4.3.4	Pré-traitement et Nettoyage des Données	9
4.4	Schéma Conceptuel du Data Warehouse	10
4.4.1	Table de Fait	10
4.4.2	Tables de Dimensions	10
4.4.3	Schéma Conceptuel	11
5	Modélisation et Processus ETL pour le Data Warehouse	12
5.1	Processus ETL	12
5.2	Modélisation de la Dimension Country	12
5.2.1	Étape 1 : Extraction	12
5.2.2	Étape 2 : Transformation	13
5.2.3	Étape 3 : Chargement	13
5.3	Modélisation de la Dimension Population	14
5.3.1	Étape 1 : Extraction	14
5.3.2	Étape 2 : Transformation	14
5.3.3	Étape 3 : Chargement	15
5.4	Modélisation de la Dimension Year	16
5.4.1	Étape 1 : Extraction	16
5.4.2	Étape 2 : Transformation	16
5.4.3	Étape 3 : Chargement	16
5.5	Modélisation de la Dimension MHD	17
5.5.1	Étape 1 : Extraction	17
5.5.2	Étape 2 : Transformation	17
5.5.3	Étape 3 : Chargement	18
5.6	Modélisation de la Table de Fait	19
5.6.1	Étape 1 : Extraction	19
5.6.2	Étape 2 : Transformation	19
5.6.3	Étape 3 : Chargement	20

6	Résultats (Réalisation)	21
6.1	Tableau de Bord	21
6.2	Objectifs du Tableau de Bord	21
6.3	Visualisations des Tableaux de Bord	21
6.4	Fonctionnalités du Tableau de Bord	22
6.5	Structure et Composantes du Tableau de Bord	24
6.6	Techniques de Visualisation	25
6.7	Résultats et Interprétations	26
6.8	Limites du Tableau de Bord	26
7	Conclusion Générale	26

Table des figures

1	Extrait des données sur les troubles de santé mentale (<code>MHD.xlsx</code>).	7
2	Extrait des données sur les indicateurs de bonheur (<code>bonheur.json</code>).	8
3	Extrait des données sur les cas de suicides (<code>suicide.csv</code>).	9
4	Schéma conceptuel du Data Warehouse.	11
5	Job ETL de la Dimension Country	13
6	tMap de la Dimension Country	13
7	Table country après chargement des données	14
8	Job ETL de la Dimension Population	14
9	tMap de la Dimension Population	15
10	Table population après chargement des données	15
11	Job ETL de la Dimension Year	16
12	tMap de la Dimension Year	16
13	Table Year après chargement des données	17
14	Job ETL de la Dimension MHD	17
15	tMap de la Dimension MHD	18
16	Table MHD après chargement des données	18
17	Job ETL de la Dimension MHD	19
18	Job ETL pour la création de la table de fait	20
19	Table de Fait après transformation et chargement	20
20	Tableau de Bord d'Analyse Globale des Taux de Suicide et Facteurs Associés	21
21	Tableau de Bord d'analyse du Bonheur et de son Impact sur les Taux de Suicide	22
22	Exemple de zoom sur un élément spécifique dans un graphique.	23
23	Exemple de tableau interactif permettant le tri et l'affichage des données. .	23
24	Exemple de Filtrage et Visualisation des Données par Critères Spécifiques .	24

1 Introduction

Dans un monde où les données jouent un rôle crucial dans la prise de décision, leur organisation et leur exploitation représentent des défis majeurs, notamment lorsqu'il s'agit de questions de santé publique comme le suicide. Ce phénomène complexe, touchant des millions de personnes chaque année, nécessite une analyse approfondie pour en comprendre les causes, identifier les groupes à risque, et élaborer des stratégies de prévention efficaces.

Un entrepôt de données (Data Warehouse) permet de centraliser et structurer de grandes quantités de données provenant de différentes sources. Dans le cadre de ce projet, un modèle de Data Warehouse a été conçu pour rassembler et analyser les informations sur les suicides à travers différentes dimensions (âge, sexe, région, facteurs économiques, etc.).

L'objectif est de fournir une base solide pour des analyses pertinentes à l'aide d'outils de visualisation tels que Power BI. Ces analyses offrent une vision claire des tendances, des corrélations potentielles, et des zones à prioriser pour les interventions.

Ce rapport détaille la conception et la mise en œuvre du Data Warehouse, en s'appuyant sur un processus structuré allant de la collecte des données à leur visualisation interactive, afin de répondre à des problématiques spécifiques liées à la compréhension des données sur le suicide.

2 Problématique

Malgré les efforts de collecte et d'analyse des données sur le suicide, plusieurs défis subsistent :

- **Fragmentation des données** : Les informations sur le suicide sont souvent réparties entre différentes bases de données, formats et structures, rendant leur analyse difficile.
- **Manque de structuration** : L'absence d'un modèle unifié et cohérent limite la capacité des analystes à identifier rapidement des tendances ou à corréler les facteurs de risque.
- **Analyse inefficace** : Les outils traditionnels d'analyse ne permettent pas toujours de transformer ces données brutes en informations exploitables, freinant la prise de décision rapide et précise.
- **Visualisation non intuitive** : Les décideurs manquent souvent d'outils visuels simples et interactifs pour interpréter les résultats de manière claire.

Comment concevoir et mettre en œuvre un modèle d'entrepôt de données permettant de :

- Centraliser les données relatives au suicide issues de diverses sources,
- Structurer ces données de manière à répondre efficacement aux besoins analytiques,
- Et offrir des visualisations interactives et exploitables via un outil comme Power BI?

Cette problématique met en lumière l'importance de concevoir une solution robuste et évolutive pour améliorer la compréhension et la prévention du suicide à l'aide des technologies modernes de gestion et d'analyse des données.

3 Spécifications Techniques : Logiciels Utilisés

Dans le cadre de ce projet, plusieurs logiciels ont été utilisés pour assurer la réalisation et l'analyse des données, chacun ayant un rôle spécifique dans la chaîne de traitement. Voici les principaux outils, leur description et leurs illustrations :

- **Power BI :**



Un outil de visualisation de données permettant de créer des tableaux de bord interactifs et des rapports analytiques. Il a été utilisé pour transformer les données en visualisations exploitables, facilitant ainsi l'analyse et la prise de décision.

- **Talend :**



Une plateforme d'intégration de données utilisée pour l'extraction, la transformation et le chargement (ETL). Elle a permis de centraliser et de structurer les données issues de différentes sources, assurant leur cohérence et leur qualité.

- **phpMyAdmin :**



Un outil de gestion de bases de données MySQL accessible via une interface web. Il a servi à gérer les données centralisées dans l'entrepôt de données, permettant d'exécuter des requêtes SQL et de maintenir la base.

- **LaTeX :**



Un système de composition de documents utilisé pour la rédaction de ce rapport. Il offre une mise en page professionnelle et des fonctionnalités avancées pour créer des documents structurés et esthétiques.

4 Spécifications fonctionnelles

4.1 Introduction au Data Warehouse (DW)

Un Data Warehouse (DW) est une base de données centralisée et optimisée pour le stockage et l'analyse de grandes quantités de données provenant de diverses sources. Dans ce projet, le DW a été conçu pour centraliser les données relatives au suicide, permettant

ainsi une meilleure analyse et des visualisations interactives via Power BI.

4.2 Objectif du DW dans ce projet

L'objectif du Data Warehouse (DW) est de fournir une structure centralisée et cohérente pour les données sur le suicide, facilitant leur accès, leur exploration et leur visualisation.

Ce DW permet :

- **Centralisation des données** : Les données provenant de plusieurs formats (CSV, Excel, JSON) sont réunies en un seul endroit.
- **Cohérence des données** : Grâce à un processus rigoureux de nettoyage et de structuration, les données sont homogènes et prêtes pour l'analyse.
- **Réponses aux besoins analytiques** : Le DW est conçu pour répondre efficacement aux besoins des décideurs en facilitant l'extraction d'informations pertinentes.

4.3 Sources de Données

Pour ce projet, trois sources principales de données ont été utilisées. Chacune d'elles a été sélectionnée en fonction de sa pertinence pour l'analyse et la modélisation des facteurs liés au suicide.

4.3.1 Données sur les troubles de santé mentale (MHD.xlsx)

Ce fichier contient des données sur les troubles de santé mentale (Mental Health Disorders) pour différents pays, réparties sur plusieurs années. Les principales colonnes de ce fichier sont :

- **ENTITY** : Nom du pays ou de la région.
- **CODE** : Code ISO du pays.
- **YEAR** : Année de l'enregistrement.
- **MENTAL_HEALTH_DISORDER** : Type de trouble mental (ex. *Bipolar Disorder*).
- **TAUX_MHD** : Taux de prévalence des troubles mentaux dans la population.

	A	B	C	D	E
	ENTITY	CODE	YEAR	MENTAL_HEALTH_DISORDER	TAUX_MHD
2	Afghanistan	AFG	2015	BIPOLAR_DISORDER	0,705313112
3	Afghanistan	AFG	2016	BIPOLAR_DISORDER	0,705688469
4	Afghanistan	AFG	2017	BIPOLAR_DISORDER	0,706086206
5	Afghanistan	AFG	2018	BIPOLAR_DISORDER	0,706471156
6	Afghanistan	AFG	2019	BIPOLAR_DISORDER	0,706875851
7	Afghanistan	AFG	2020	BIPOLAR_DISORDER	0,70727518
8	Afghanistan	AFG	2021	BIPOLAR_DISORDER	0,707685928
9	Afghanistan	AFG	2022	BIPOLAR_DISORDER	0,708088873
0	Albania	ALB	2015	BIPOLAR_DISORDER	0,703063617
1	Albania	ALB	2016	BIPOLAR_DISORDER	0,703313274
2	Albania	ALB	2017	BIPOLAR_DISORDER	0,703539474
3	Albania	ALB	2018	BIPOLAR_DISORDER	0,703759747
4	Albania	ALB	2019	BIPOLAR_DISORDER	0,703970319
5	Albania	ALB	2020	BIPOLAR_DISORDER	0,704160792
6	Albania	ALB	2021	BIPOLAR_DISORDER	0,704335235
7	Albania	ALB	2022	BIPOLAR_DISORDER	0,704479655
8	Algeria	DZA	2015	BIPOLAR_DISORDER	0,815265303
9	Algeria	DZA	2016	BIPOLAR_DISORDER	0,815808468
0	Algeria	DZA	2017	BIPOLAR_DISORDER	0,816334462
1	Algeria	DZA	2018	BIPOLAR_DISORDER	0,816873571

FIGURE 1 – Extrait des données sur les troubles de santé mentale (MHD.xlsx).

Ces données permettent d'analyser l'évolution des troubles mentaux dans différents pays au fil des années.

4.3.2 Données sur le bonheur (bonheur.json)

Ce fichier JSON contient des indicateurs de bonheur pour divers pays en 2022, incluant des métriques économiques et sociales. Les principales informations sont :

- **Country** : Nom du pays.
- **Regional Indicator** : Région géographique du pays.
- **Year** : Année de l'enregistrement.
- **GDP per capita** : Produit intérieur brut par habitant.
- **Social support** : Niveau de soutien social.
- **Healthy life expectancy** : Espérance de vie en bonne santé.
- **Freedom, Generosity, Perceptions of corruption** : Indicateurs relatifs à la liberté, la générosité et la perception de la corruption.


```
{} Bonheur.json X
C: > Users > khadija > Downloads > {} Bonheur.json > {} 2
2      {
4      "Regional_Indicator": "Western Europe",
5      "Year": 2022,
6      "Dystopia_residual": "2.518",
7      "GDP_per_capita": "1.892",
8      "Social_support": "1.258",
9      "Healthy_life_expectancy": "0.775",
10     "Freedom": "0.736",
11     "Generosity": "0.109",
12     "Perceptions_of_corruption": "0.534"
13 },
14 {
15     "Country": "Denmark",
16     "Regional_Indicator": "Western Europe",
17     "Year": 2022,
18     "Dystopia_residual": "2.226",
19     "GDP_per_capita": "1.953",
20     "Social_support": "1.243",
21     "Healthy_life_expectancy": "0.777",
22     "Freedom": "0.719",
23     "Generosity": "0.188",
24     "Perceptions_of_corruption": "0.532"
25 },
```

FIGURE 2 – Extrait des données sur les indicateurs de bonheur (bonheur.json).

Ces données sont utiles pour établir des corrélations entre le bonheur, les indicateurs économiques, et les troubles mentaux.

4.3.3 Données sur les suicides (suicide.csv)

Ce fichier CSV contient des informations détaillées sur les cas de suicides, avec des données démographiques et économiques. Les colonnes principales incluent :

- **Country** : Nom du pays.
- **Year** : Année de l'enregistrement.
- **Sex** : Genre (*male* ou *female*).
- **Age** : Groupe d'âge.
- **Suicides_no** : Nombre de suicides enregistrés.
- **Population** : Taille de la population correspondante.
- **Suicides/100k pop** : Taux de suicide pour 100 000 habitants.
- **GDP for year** : PIB du pays pour l'année donnée.
- **Generation** : Génération (par exemple, *Generation X*, *Boomers*).

	A	B	C	D	E
1	ENTITY	CODE	YEAR	MENTAL_HEALTH_DISORDER	TAUX_MHD
2	Afghanistan	AFG	2015	BIPOLAR_DISORDER	0,705313112
3	Afghanistan	AFG	2016	BIPOLAR_DISORDER	0,705688469
4	Afghanistan	AFG	2017	BIPOLAR_DISORDER	0,706086206
5	Afghanistan	AFG	2018	BIPOLAR_DISORDER	0,706471156
6	Afghanistan	AFG	2019	BIPOLAR_DISORDER	0,706875851
7	Afghanistan	AFG	2020	BIPOLAR_DISORDER	0,70727518
8	Afghanistan	AFG	2021	BIPOLAR_DISORDER	0,707685928
9	Afghanistan	AFG	2022	BIPOLAR_DISORDER	0,708088873
10	Albania	ALB	2015	BIPOLAR_DISORDER	0,703063617
11	Albania	ALB	2016	BIPOLAR_DISORDER	0,703313274
12	Albania	ALB	2017	BIPOLAR_DISORDER	0,703539474
13	Albania	ALB	2018	BIPOLAR_DISORDER	0,703759747
14	Albania	ALB	2019	BIPOLAR_DISORDER	0,703970319
15	Albania	ALB	2020	BIPOLAR_DISORDER	0,704160792
16	Albania	ALB	2021	BIPOLAR_DISORDER	0,704335235
17	Albania	ALB	2022	BIPOLAR_DISORDER	0,704479655
18	Algeria	DZA	2015	BIPOLAR_DISORDER	0,815265303
19	Algeria	DZA	2016	BIPOLAR_DISORDER	0,815808468
20	Algeria	DZA	2017	BIPOLAR_DISORDER	0,816334462
21	Algeria	DZA	2018	BIPOLAR DISORDER	0,816873571

FIGURE 3 – Extrait des données sur les cas de suicides (suicide.csv).

Ces données permettent une analyse approfondie des tendances de suicide selon les genres, les groupes d'âge, et les indicateurs économiques.

4.3.4 Pré-traitement et Nettoyage des Données

Avant leur intégration dans le Data Warehouse, les fichiers ont été standardisés :

- Les valeurs manquantes ont été imputées ou éliminées.
- Les formats de dates, de chiffres et de chaînes de caractères ont été harmonisés.

- Les données redondantes ou inconsistantes ont été corrigées.

Ces étapes garantissent une qualité optimale des données pour l'analyse et la visualisation.

4.4 Schéma Conceptuel du Data Warehouse

Le Data Warehouse a été conçu pour centraliser et structurer les données issues de plusieurs sources. Il permet de réaliser des analyses multidimensionnelles sur les troubles mentaux, les indicateurs de bonheur, les taux de suicide, et d'autres métriques socio-économiques. Voici la structure conceptuelle du Data Warehouse.

4.4.1 Table de Fait

La table de fait contient les mesures quantitatives nécessaires pour les analyses. Elle est connectée aux dimensions pour permettre des agrégations et des analyses contextuelles. Les attributs de la table de fait sont les suivants :

- **id_mh** : Identifiant unique de la mesure.
- **suicide_rate** : Taux de suicide pour 100 000 habitants.
- **taux_mhd** : Taux de prévalence des troubles mentaux.
- **hap_score** : Score de bonheur.
- **gdp_per_year** : Produit intérieur brut annuel.
- **gdp_per_capita** : PIB par habitant.
- **freedom** : Indicateur de liberté.
- **social_support** : Niveau de soutien social.
- **gdp_bonheur** : PIB corrigé par le bonheur.

4.4.2 Tables de Dimensions

Les dimensions sont les entités permettant de contextualiser et de segmenter les données de la table de fait. Voici les tables de dimensions utilisées :

- **year** : Fournit l'information temporelle pour les analyses.
 - **id_year** : Identifiant unique de l'année.
 - **year** : Année (ex : 2022).
- **population** : Contient les données démographiques.
 - **id_pop** : Identifiant unique de la population.
 - **age** : Groupe d'âge.
 - **sex** : Sexe (**male** ou **female**).
- **mhd** : Référence les différents troubles mentaux.
 - **id_mhd** : Identifiant unique du trouble mental.

- **name_mhd** : Nom du trouble mental (ex : DEPRESSION).
- **country** : Contient les informations géographiques.
 - **id_country** : Identifiant unique du pays.
 - **country_name** : Nom du pays.
 - **region** : Région géographique (ex : Europe).

4.4.3 Schéma Conceptuel

Le schéma conceptuel ci-dessous montre les relations entre la table de fait et les tables de dimensions. Il adopte une structure en étoile pour faciliter les analyses multidimensionnelles.

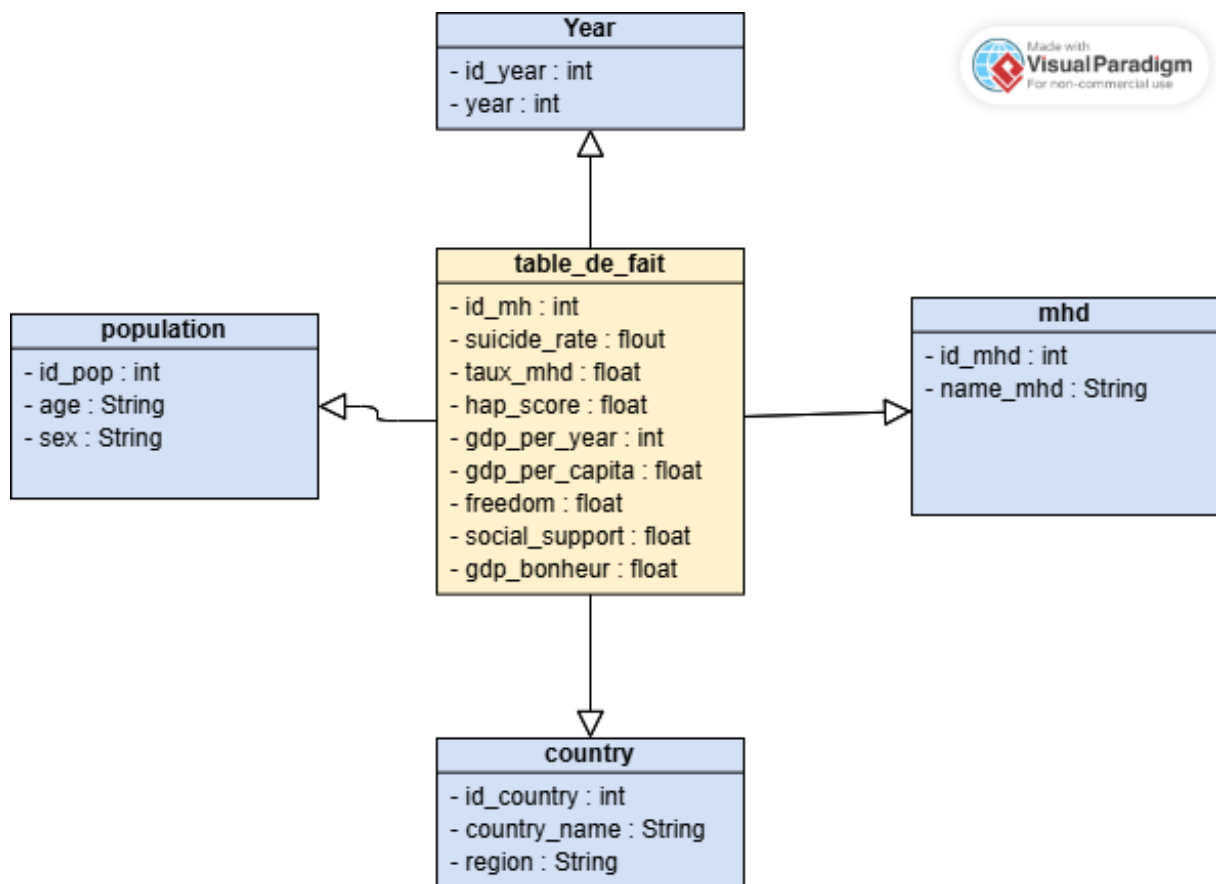


FIGURE 4 – Schéma conceptuel du Data Warehouse.

5 Modélisation et Processus ETL pour le Data Warehouse

5.1 Processus ETL

Le processus ETL (Extract, Transform, Load) est une étape clé pour construire un Data Warehouse. Ce processus est divisé en trois phases principales :

- **Extraction (Extract)** : Collecter les données brutes à partir de différentes sources.
- **Transformation (Transform)** : Nettoyer, formater et structurer les données pour les rendre compatibles avec le modèle du DW.
- **Chargement (Load)** : Insérer les données transformées dans le Data Warehouse.

Donc, comme on a déjà mentionné, on va travailler avec l'outil Talend pour extraire, transformer et alimenter notre Data Warehouse.

5.2 Modélisation de la Dimension Country

5.2.1 Étape 1 : Extraction

La table de dimension `country` permet de stocker des informations uniques et normalisées sur les pays. Les données proviennent de la première colonne du fichier `mhd.xlsx`, de la première colonne du fichier `suicide.csv` ainsi que de la colonne `Regional_Indicator` du fichier `bonheur.json`. Cette table joue un rôle essentiel dans le processus ETL. Elle permet notamment la **normalisation** des données en supprimant les doublons, la **structuration** en facilitant les jointures avec d'autres tables, la **réduction de la redondance** en évitant la duplication des informations pays et l'**amélioration des performances** en optimisant les filtres et les agrégations par pays. Ainsi, elle contribue à améliorer la qualité des données et l'efficacité globale du traitement.

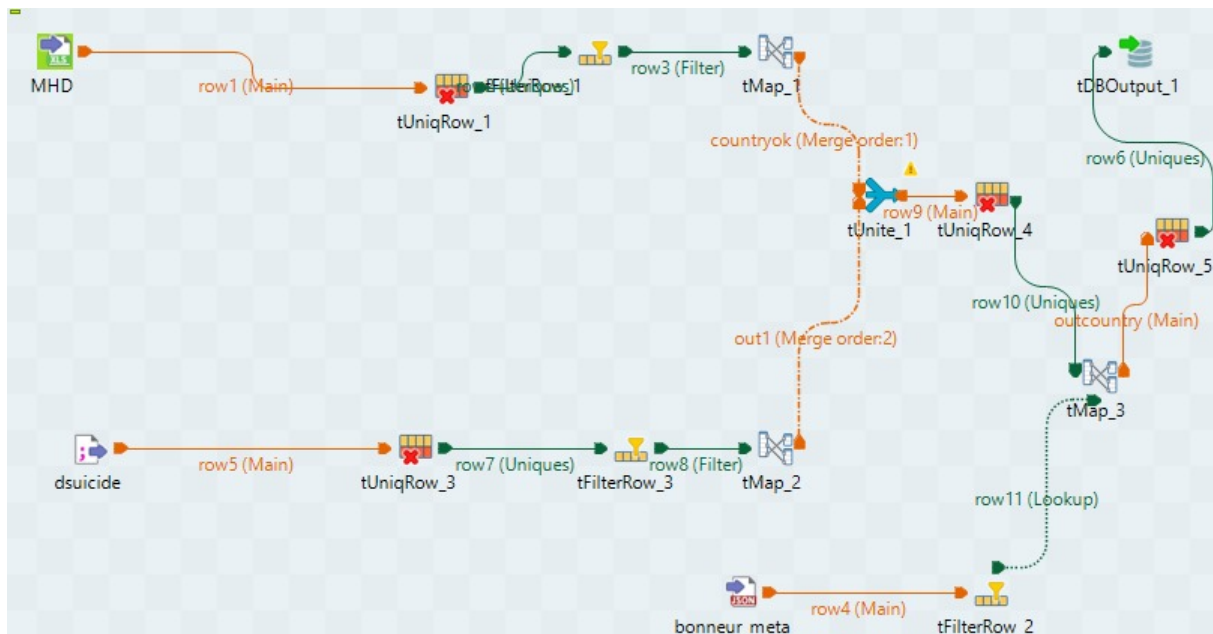


FIGURE 5 – Job ETL de la Dimension Country

5.2.2 Étape 2 : Transformation

Les étapes de transformation incluent le nettoyage des données et l'ajout d'identifiants uniques.

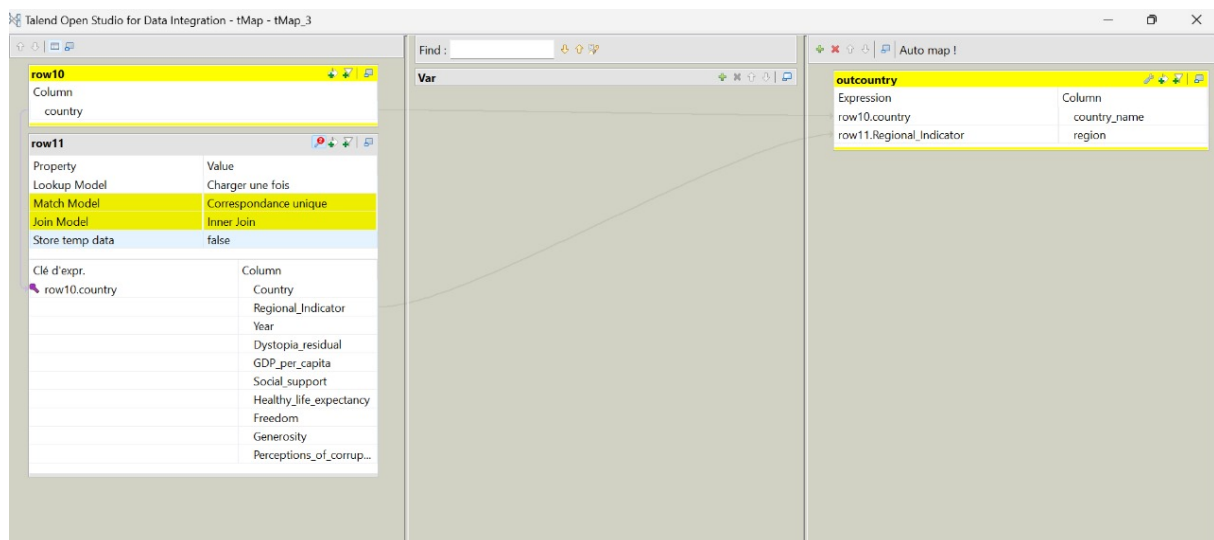


FIGURE 6 – tMap de la Dimension Country

5.2.3 Étape 3 : Chargement

Les données transformées sont ensuite chargées dans la base de données.

	id_country	country_name	region
<input type="checkbox"/> Éditer <input type="checkbox"/> Copier <input type="checkbox"/> Supprimer	1	Afghanistan	Southern Asia
<input type="checkbox"/> Éditer <input type="checkbox"/> Copier <input type="checkbox"/> Supprimer	2	Albania	Central and Eastern Europe
<input type="checkbox"/> Éditer <input type="checkbox"/> Copier <input type="checkbox"/> Supprimer	3	Algeria	Middle East and Northern Africa
<input type="checkbox"/> Éditer <input type="checkbox"/> Copier <input type="checkbox"/> Supprimer	4	Argentina	Latin America and Caribbean
<input type="checkbox"/> Éditer <input type="checkbox"/> Copier <input type="checkbox"/> Supprimer	5	Armenia	Central and Eastern Europe
<input type="checkbox"/> Éditer <input type="checkbox"/> Copier <input type="checkbox"/> Supprimer	6	Australia	Australia and New Zealand
<input type="checkbox"/> Éditer <input type="checkbox"/> Copier <input type="checkbox"/> Supprimer	7	Austria	Western Europe
<input type="checkbox"/> Éditer <input type="checkbox"/> Copier <input type="checkbox"/> Supprimer	8	Azerbaijan	Central and Eastern Europe
<input type="checkbox"/> Éditer <input type="checkbox"/> Copier <input type="checkbox"/> Supprimer	9	Bahrain	Middle East and Northern Africa
<input type="checkbox"/> Éditer <input type="checkbox"/> Copier <input type="checkbox"/> Supprimer	10	Bangladesh	Southern Asia
<input type="checkbox"/> Éditer <input type="checkbox"/> Copier <input type="checkbox"/> Supprimer	11	Belarus	Central and Eastern Europe
<input type="checkbox"/> Éditer <input type="checkbox"/> Copier <input type="checkbox"/> Supprimer	12	Belgium	Western Europe
<input type="checkbox"/> Éditer <input type="checkbox"/> Copier <input type="checkbox"/> Supprimer	13	Benin	Sub-Saharan Africa
<input type="checkbox"/> Éditer <input type="checkbox"/> Copier <input type="checkbox"/> Supprimer	14	Bolivia	Latin America and Caribbean
<input type="checkbox"/> Éditer <input type="checkbox"/> Copier <input type="checkbox"/> Supprimer	15	Bosnia and Herzegovina	Central and Eastern Europe
<input type="checkbox"/> Éditer <input type="checkbox"/> Copier <input type="checkbox"/> Supprimer	16	Botswana	Sub-Saharan Africa
<input type="checkbox"/> Éditer <input type="checkbox"/> Copier <input type="checkbox"/> Supprimer	17	Brazil	Latin America and Caribbean
<input type="checkbox"/> Éditer <input type="checkbox"/> Copier <input type="checkbox"/> Supprimer	18	Bulgaria	Central and Eastern Europe

FIGURE 7 – Table country après chargement des données

5.3 Modélisation de la Dimension Population

5.3.1 Étape 1 : Extraction

La table de dimension population permet de stocker des informations uniques et normalisées sur l'âge et le **sexe**. Les données proviennent du fichier `suicide.csv`.

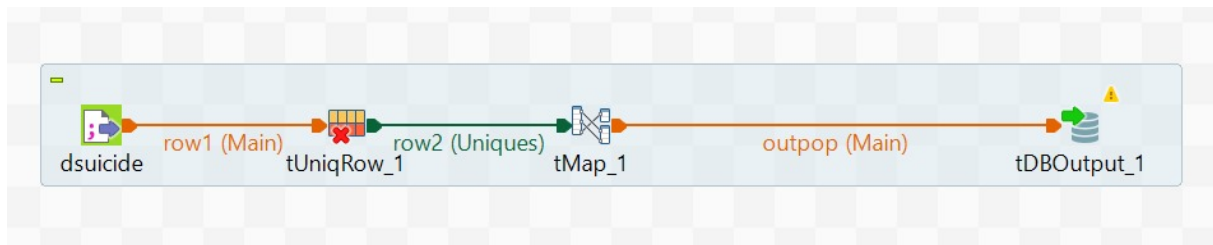


FIGURE 8 – Job ETL de la Dimension Population

5.3.2 Étape 2 : Transformation

Les étapes de transformation incluent le nettoyage des données et l'ajout d'identifiants uniques.

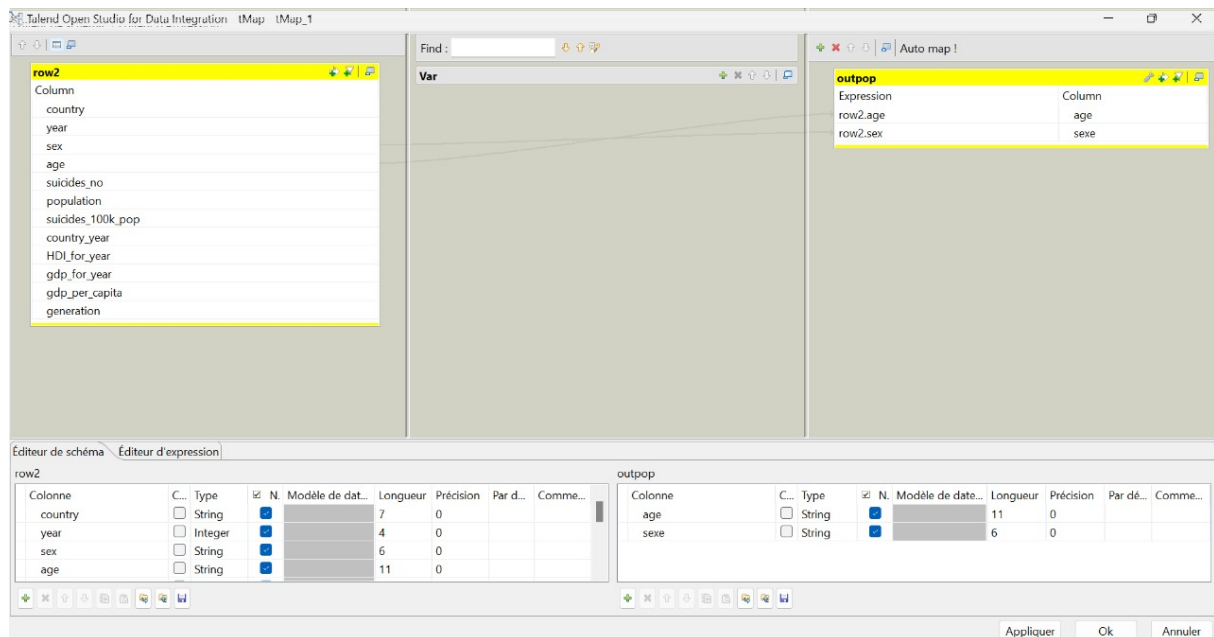


FIGURE 9 – tMap de la Dimension Population

5.3.3 Étape 3 : Chargement

Les données transformées sont ensuite chargées dans la base de données.

The screenshot shows the phpMyAdmin interface with the 'population' table selected in the 'mentale_health' database. The table structure is as follows:

id_pop	age	sexe
8701	15-24 years	male
8702	35-54 years	male
8703	15-24 years	female
8704	75+ years	male
8705	25-34 years	male
8706	75+ years	female
8707	35-54 years	female
8708	25-34 years	female
8709	55-74 years	male
8710	5-14 years	female
8711	55-74 years	female
8712	5-14 years	male
8713	75+ years	female
8714	15-24 years	male
8715	75+ years	male
8716	35-54 years	male
8717	55-74 years	male

FIGURE 10 – Table population après chargement des données

5.4 Modélisation de la Dimension Year

5.4.1 Étape 1 : Extraction

La table de dimension **Year** permet de stocker des informations uniques et normalisées sur l'**année**. Les données proviennent de la colonne **Year** du fichier **suicide.csv**.

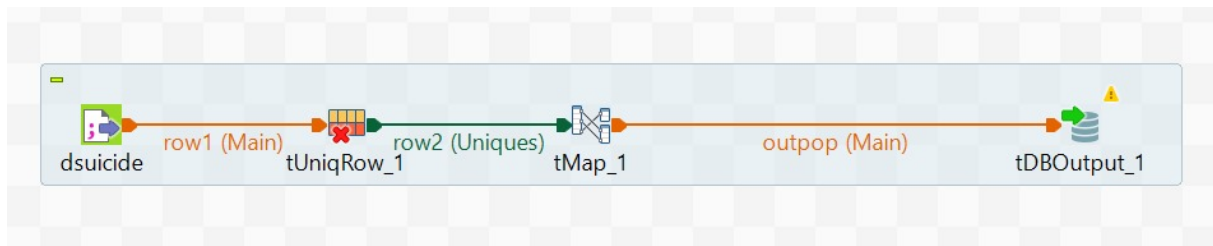


FIGURE 11 – Job ETL de la Dimension Year

5.4.2 Étape 2 : Transformation

Les étapes de transformation incluent le nettoyage des données et l'ajout d'identifiants uniques.

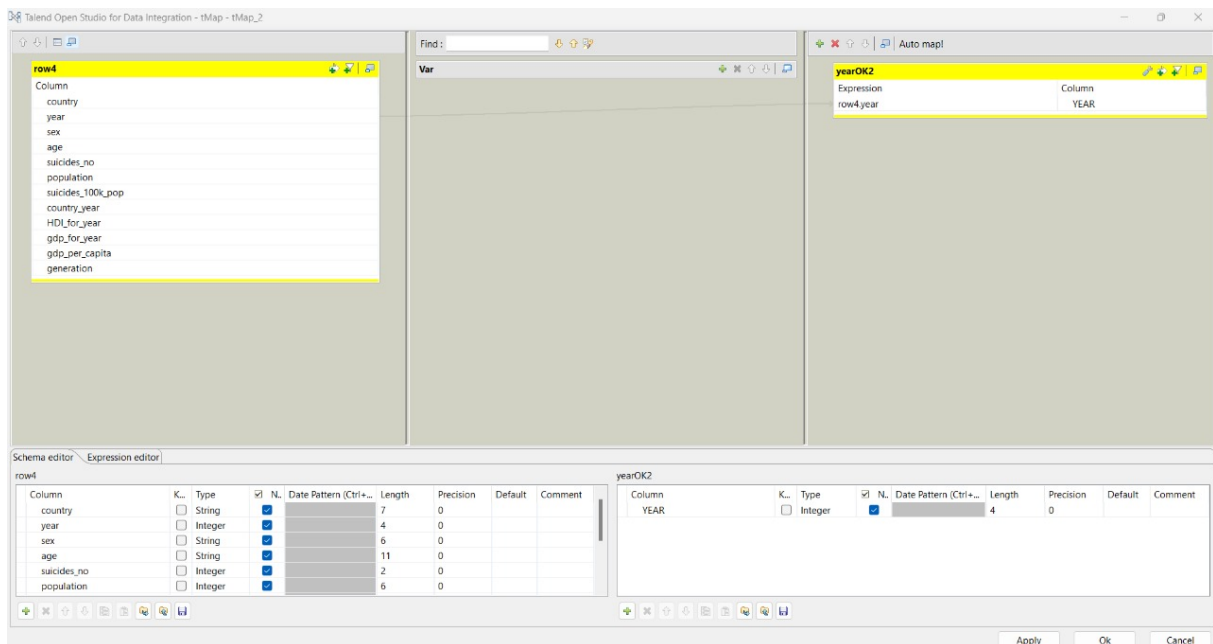


FIGURE 12 – tMap de la Dimension Year

5.4.3 Étape 3 : Chargement

Les données transformées sont ensuite chargées dans la base de données.

	id_year	annee
<input type="checkbox"/>	1	1987
<input type="checkbox"/>	13	1988
<input type="checkbox"/>	25	1989
<input type="checkbox"/>	37	1992
<input type="checkbox"/>	49	1993
<input type="checkbox"/>	61	1994
<input type="checkbox"/>	73	1995
<input type="checkbox"/>	85	1996
<input type="checkbox"/>	97	1997
<input type="checkbox"/>	109	1998
<input type="checkbox"/>	121	1999
<input type="checkbox"/>	133	2000
<input type="checkbox"/>	145	2001
<input type="checkbox"/>	157	2002
<input type="checkbox"/>	169	2003
<input type="checkbox"/>	181	2004
<input type="checkbox"/>	193	2005
<input type="checkbox"/>	205	2006

FIGURE 13 – Table Year après chargement des données

5.5 Modélisation de la Dimension MHD

5.5.1 Étape 1 : Extraction

La source des données est le fichier `mhd.xlsx`, contenant une liste de type de trouble mental. .

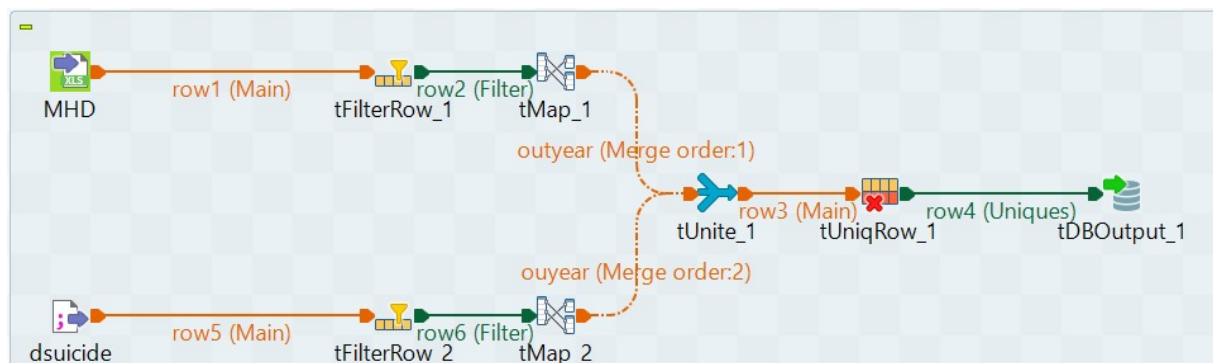


FIGURE 14 – Job ETL de la Dimension MHD

5.5.2 Étape 2 : Transformation

Les étapes de transformation incluent le nettoyage des données et l'ajout d'identifiants uniques.

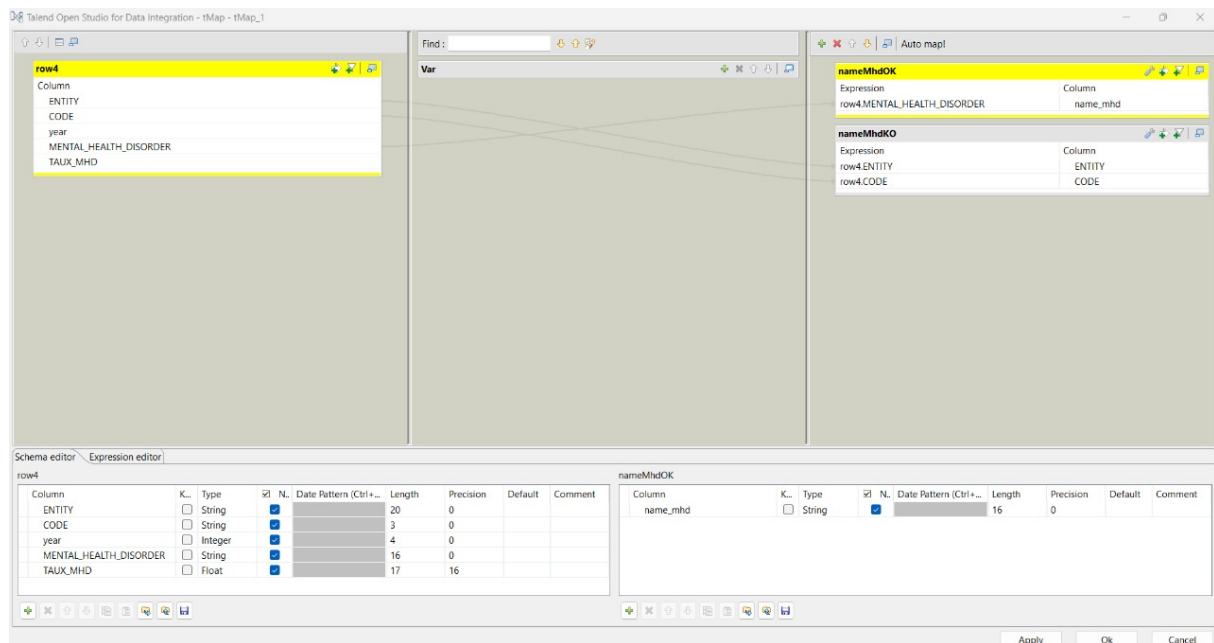


FIGURE 15 – tMap de la Dimension MHD

5.5.3 Étape 3 : Chargement

Les données transformées sont ensuite chargées dans la base de données.

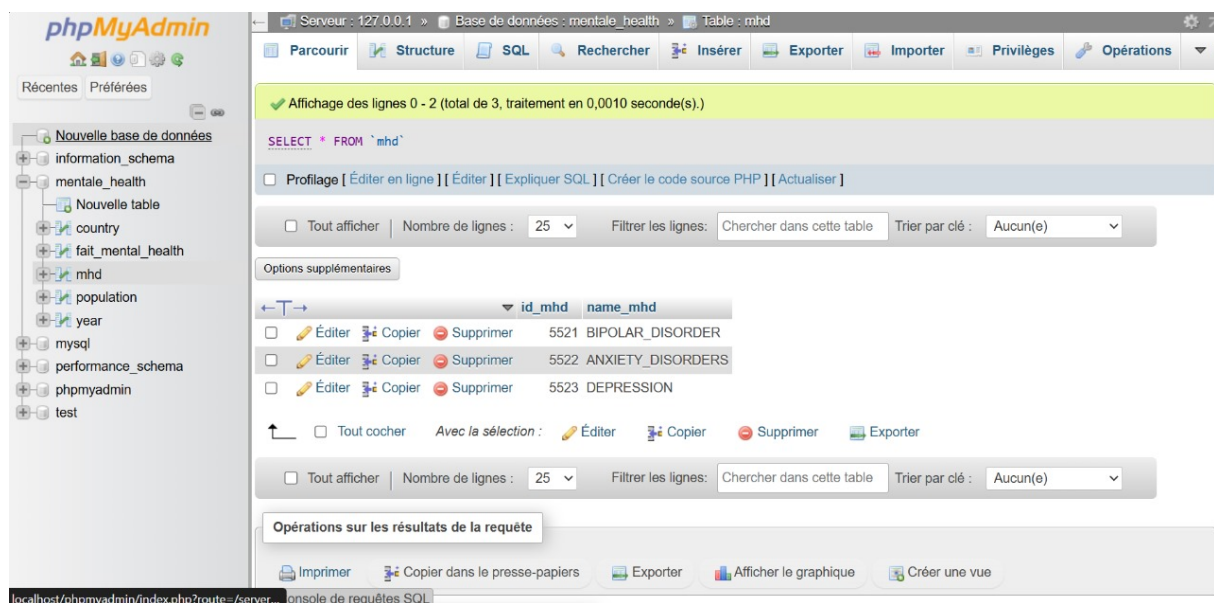


FIGURE 16 – Table MHD après chargement des données

5.6 Modélisation de la Table de Fait

5.6.1 Étape 1 : Extraction

La source des données est un fichier contenant les taux de suicide, le PIB, et d'autres mesures pour différents pays, années, et groupes démographiques.

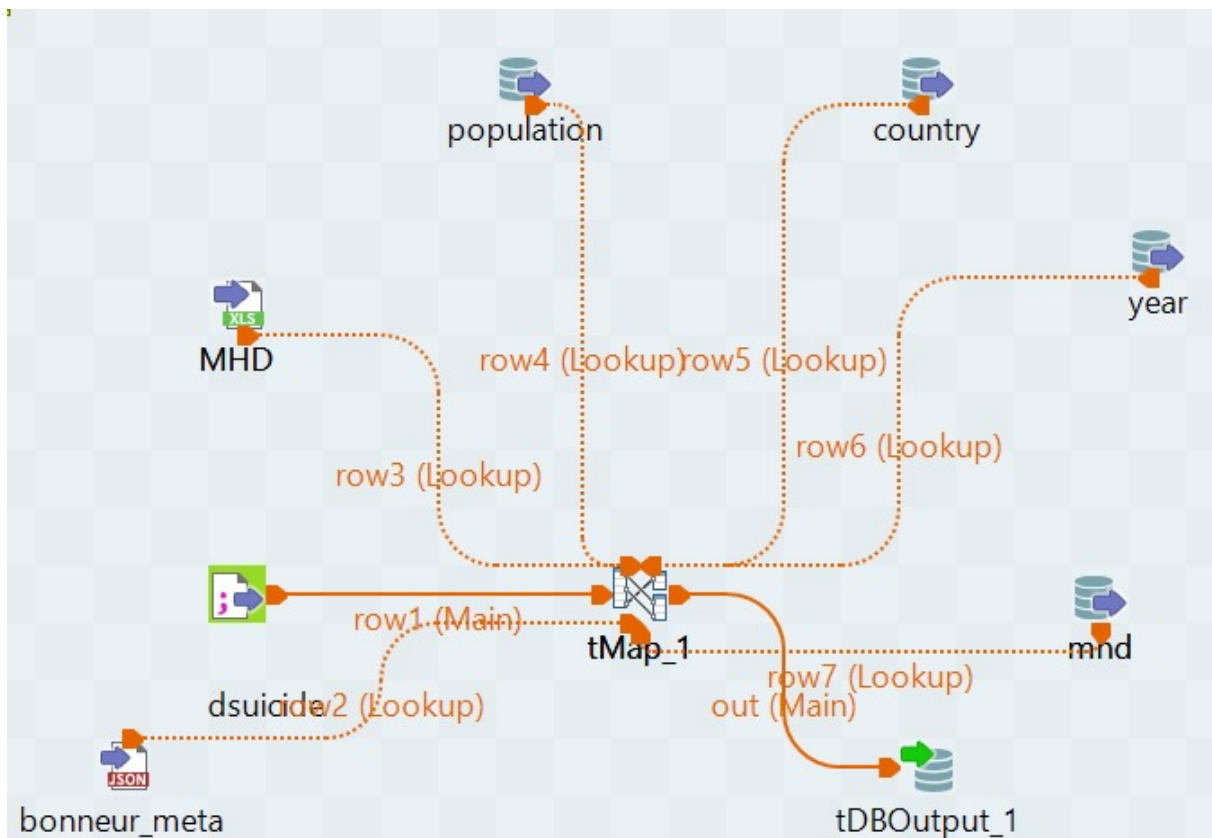


FIGURE 17 – Job ETL de la Dimension MHD

5.6.2 Étape 2 : Transformation

Les étapes de transformation incluent :

- Calcul des indicateurs nécessaires.
- Ajout des clés étrangères pour lier les dimensions.

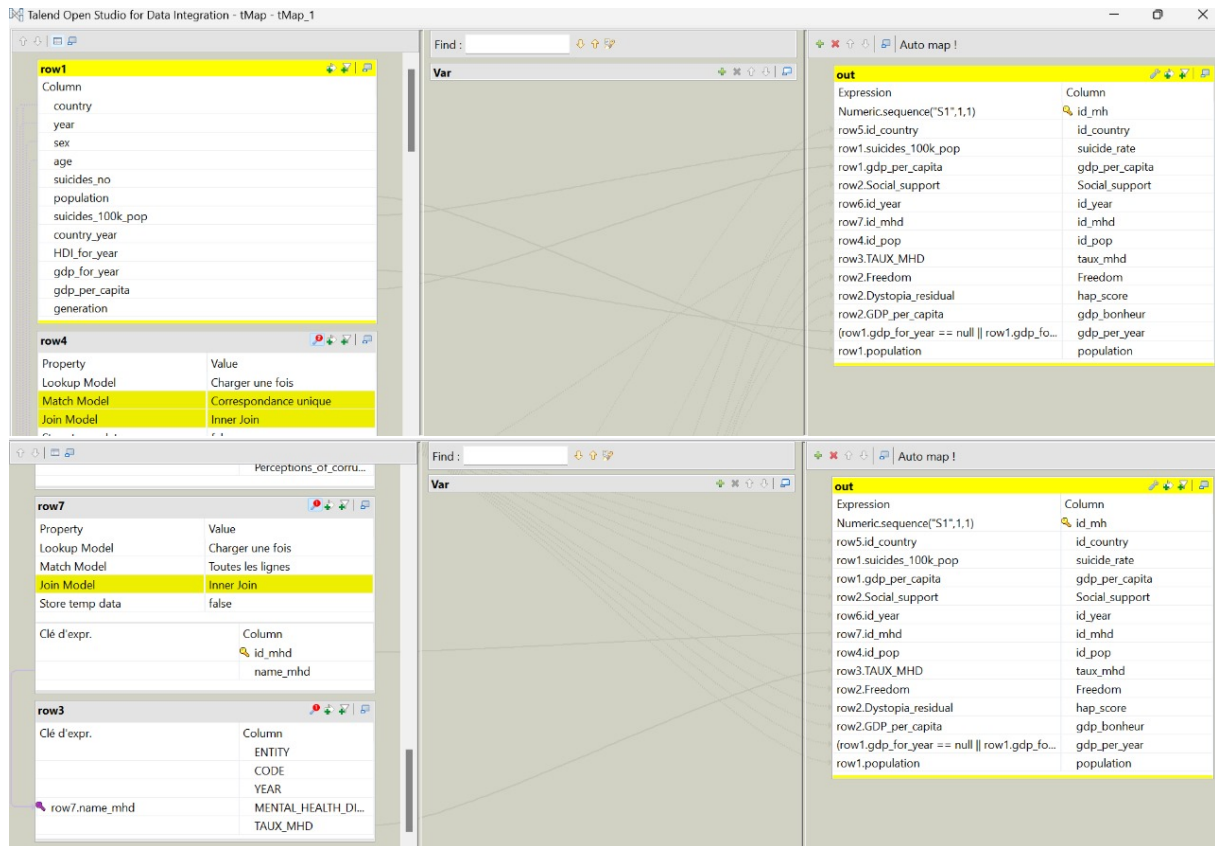


FIGURE 18 – Job ETL pour la création de la table de fait

5.6.3 Étape 3 : Chargement

Voici le résultat final après transformation et chargement dans la table de fait :

	id_mh	id_country	id_year	id_mhd	id_pop	taux_mhd	suicide_rate	hap_score	gdp_per_year	gdp_per_capita	freedom	social_support	gdp_bonheur	population
1	1	2	1	5521	36513	0.611242	6.71	1.89894	2156620000	796	0.35733	0.80434	0.87867	312900
2	2	2	1	5522	36513	3.11093	6.71	1.89894	2156620000	796	0.35733	0.80434	0.87867	312900
3	3	2	1	5523	36513	3.19279	6.71	1.89894	2156620000	796	0.35733	0.80434	0.87867	312900
4	4	2	1	5521	36509	0.611242	5.19	1.89894	2156620000	796	0.35733	0.80434	0.87867	308000
5	5	2	1	5522	36509	3.11093	5.19	1.89894	2156620000	796	0.35733	0.80434	0.87867	308000
6	6	2	1	5523	36509	3.19279	5.19	1.89894	2156620000	796	0.35733	0.80434	0.87867	308000
7	7	2	1	5521	36511	0.611242	4.83	1.89894	2156620000	796	0.35733	0.80434	0.87867	289700
8	8	2	1	5522	36511	3.11093	4.83	1.89894	2156620000	796	0.35733	0.80434	0.87867	289700
9	9	2	1	5523	36511	3.19279	4.83	1.89894	2156620000	796	0.35733	0.80434	0.87867	289700
10	10	2	1	5521	36514	0.611242	4.59	1.89894	2156620000	796	0.35733	0.80434	0.87867	21800
11	11	2	1	5522	36514	3.11093	4.59	1.89894	2156620000	796	0.35733	0.80434	0.87867	21800
12	12	2	1	5523	36514	3.19279	4.59	1.89894	2156620000	796	0.35733	0.80434	0.87867	21800
13	13	2	1	5521	36510	0.611242	3.28	1.89894	2156620000	796	0.35733	0.80434	0.87867	274300
14	14	2	1	5522	36510	3.11093	3.28	1.89894	2156620000	796	0.35733	0.80434	0.87867	274300
15	15	2	1	5523	36510	3.19279	3.28	1.89894	2156620000	796	0.35733	0.80434	0.87867	274300
16	16	2	1	5521	36517	0.611242	2.81	1.89894	2156620000	796	0.35733	0.80434	0.87867	35600
17	17	2	1	5522	36517	3.11093	2.81	1.89894	2156620000	796	0.35733	0.80434	0.87867	35600

FIGURE 19 – Table de Fait après transformation et chargement

6 Résultats (Réalisation)

6.1 Tableau de Bord

Dans le cadre de ce projet, un tableau de bord interactif a été développé pour analyser et visualiser les données relatives au suicide à travers le monde. Ce tableau de bord facilite l'exploration des données et fournit des informations clés sur les facteurs influençant les taux de suicide.

6.2 Objectifs du Tableau de Bord

L'objectif principal du tableau de bord est de fournir :

- **Une vue globale :** Présenter les tendances mondiales des taux de suicide en fonction des données disponibles.
- **Des analyses approfondies :** Explorer les corrélations entre divers facteurs, tels que le PIB, le score de bonheur et les données démographiques.
- **Un outil interactif :** Permettre aux utilisateurs de filtrer et d'explorer les données selon différents critères (pays, années, genre, etc.).

6.3 Visualisations des Tableaux de Bord



FIGURE 20 – Tableau de Bord d'Analyse Globale des Taux de Suicide et Facteurs Associés



FIGURE 21 – Tableau de Bord d'analyse du Bonheur et de son Impact sur les Taux de Suicide

6.4 Fonctionnalités du Tableau de Bord

Le tableau de bord offre plusieurs fonctionnalités interactives et intuitives, conçues pour faciliter l'analyse des données et la prise de décision. Les utilisateurs peuvent :

- **Zoomer sur un élément spécifique** pour examiner des détails précis dans les graphiques interactifs. La Figure 22 illustre cette fonctionnalité.
- **Lire les informations directement depuis un graphique** en pointant ou en sélectionnant une zone d'intérêt.
- **Trier et afficher les données dans un tableau** selon des critères croissants, décroissants ou personnalisés, permettant une exploration approfondie des informations. La Figure 23 montre cette fonctionnalité en action.
- **Rechercher automatiquement des clusters** pour identifier des regroupements significatifs dans les données.
- **Exporter les données** au format souhaité pour une utilisation ultérieure ou pour les partager avec d'autres parties prenantes.
- **Représenter les données sous forme de tableaux** pour une visualisation structurée.
- **Filtrer les données affichées** en fonction de divers critères pour répondre à des besoins d'analyse spécifiques.
- **Visualisation dynamique** Les utilisateurs peuvent choisir d'afficher les données selon des regroupements ou des métriques clés (par exemple, par pays ou par un score spécifique).

Ces fonctionnalités rendent le tableau de bord polyvalent et adapté à divers cas d'usage, allant de l'analyse exploratoire des données à la visualisation avancée.

Illustrations

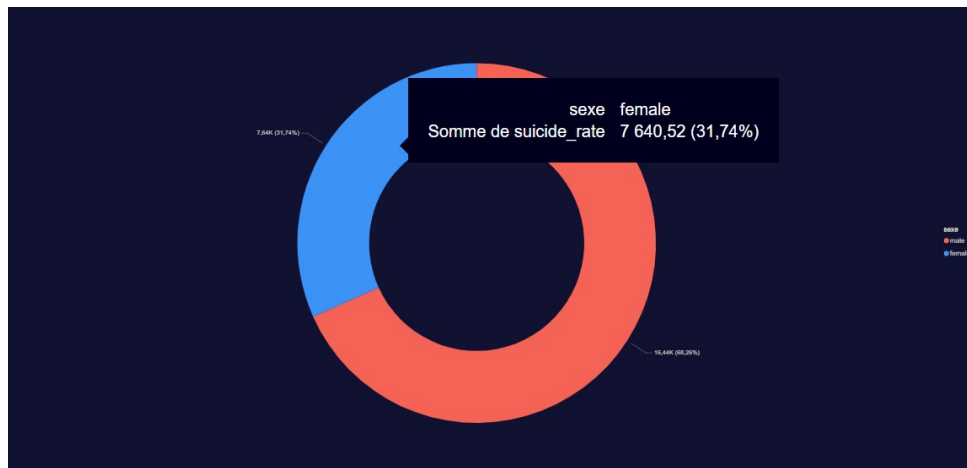


FIGURE 22 – Exemple de zoom sur un élément spécifique dans un graphique.

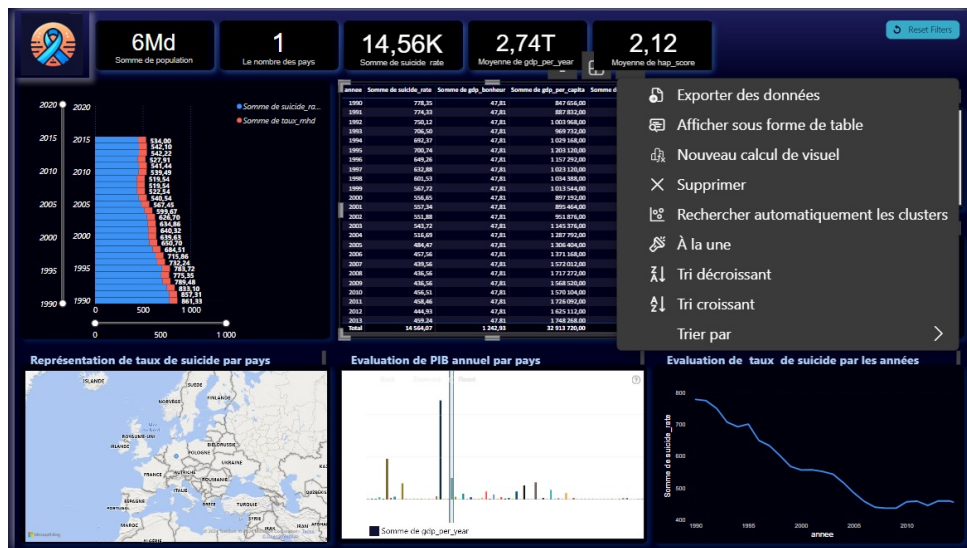


FIGURE 23 – Exemple de tableau interactif permettant le tri et l'affichage des données.



FIGURE 24 – Exemple de Filtrage et Visualisation des Données par Critères Spécifiques

6.5 Structure et Composantes du Tableau de Bord

Le tableau de bord est structuré en plusieurs sections, chacune dédiée à un aspect clé de l'analyse des données. Voici une description détaillée :

- **Indicateurs Clés :**
 - Population totale analysée
 - Nombre total de pays
 - Nombre total de suicides
 - PIB moyen par habitant (*GDP per capita*)
 - Score moyen de bonheur (*hap_score*)
- **Tendances Temporelles :**
 - Graphique des taux de suicide par décennie
 - Évolution des taux de suicide dans le temps
- **Analyse Démographique :**
 - Répartition des suicides par sexe
 - Analyse par tranche d'âge
- **Analyse Socio-Économique :**
 - Carte Géographique des taux de suicide
 - Corrélations PIB et Scores de Bonheur
- **Troubles Mentaux Associés :** Dépression, troubles anxieux, troubles bipolaires.
- **Analyse Spécifique par Pays :** Tendances au Mexique, Allemagne, Italie, et Royaume-Uni.

6.6 Techniques de Visualisation

Pour analyser et présenter les données de manière claire et intuitive, plusieurs types de visualisations ont été utilisés dans le tableau de bord :

Graphiques à Barres

- Ces graphiques permettent de comparer des catégories ou des périodes spécifiques, comme les **taux de suicide par tranche d'âge** ou les **variations des taux de suicide par décennie**.
- Ils sont particulièrement utiles pour identifier des tendances générales ou des écarts significatifs entre les groupes.

Graphiques Circulaires (Pie Charts)

- Utilisés pour montrer la répartition relative entre différentes catégories, comme la **proportion de suicides par sexe** (masculin vs féminin).
- Ces graphiques offrent une vue d'ensemble rapide sur la dominance ou la contribution relative de chaque catégorie.

Cartes Interactives

- Des **cartes géographiques** ont été intégrées pour visualiser les **taux de suicide par pays**.
- Ces cartes permettent d'explorer les variations régionales, de comparer les pays et de repérer les zones les plus touchées.
- L'interactivité permet aux utilisateurs de zoomer sur des régions spécifiques ou d'afficher des informations détaillées en survolant un pays.

Courbes Temporelles

- Les courbes sont utilisées pour illustrer l'**évolution des taux de suicide dans le temps**, offrant une perspective sur les tendances historiques.
- Elles permettent de détecter les années où les taux ont considérablement augmenté ou diminué.

Graphiques en Secteurs (Donut Charts ou Sunburst Charts)

- Utilisés pour analyser les **troubles mentaux associés au suicide**, tels que la dépression, les troubles anxieux ou bipolaires.
- Ces graphiques montrent la répartition des différents troubles en pourcentages, facilitant l'identification des principales causes.

Graphiques en Barres Horizontales

- Utilisés pour représenter les **taux de suicide par décennie**.
- Ce type de visualisation rend les comparaisons entre périodes plus claires, surtout lorsqu'il y a de nombreuses catégories.

En combinant ces techniques, le tableau de bord offre une analyse complète et interactive, permettant de mieux comprendre les facteurs influençant les taux de suicide à travers le monde.

6.7 Résultats et Interprétations

Les résultats obtenus à travers le tableau de bord sont résumés comme suit :

- **Différences par Sexe** : Les hommes représentent 78,72% des cas de suicide, tandis que les femmes représentent 21,28%.
- **Impact du PIB et du Score de Bonheur** : Les pays avec des scores de bonheur élevés tendent à avoir des taux de suicide plus faibles, bien que des exceptions existent.
- **Répartition par Âge** : Les groupes d'âge de 35 à 54 ans enregistrent les taux de suicide les plus élevés.
- **Corrélations avec les Troubles Mentaux** : La dépression est le principal facteur de risque (43,33%), suivie des troubles anxieux (33,33%) et bipolaires (23,34%).

6.8 Limites du Tableau de Bord

Bien que le tableau de bord fournisse des insights précieux, certaines limitations doivent être notées :

- **Disponibilité des Données** : Les données peuvent être incomplètes pour certains pays ou certaines années.
- **Interprétation des Corrélations** : Les relations entre les indicateurs économiques et sociaux avec les taux de suicide ne sont pas forcément des causalités.
- **Simplification des Données** : Certaines visualisations agrègent des informations complexes.

7 Conclusion Générale

Ce projet a permis de développer un tableau de bord interactif permettant d'analyser et de visualiser les données relatives aux suicides à travers le monde. L'objectif principal était de fournir un outil accessible et intuitif pour explorer les tendances des taux de suicide, d'analyser les corrélations avec des facteurs économiques, sociaux et démographiques, et de mettre en évidence les principaux facteurs de risque associés à ce phénomène tragique.

Grâce à des visualisations variées telles que des graphiques à barres, des courbes temporelles, des cartes géographiques et des graphiques de corrélation, le tableau de bord offre une vue d'ensemble détaillée et dynamique des données, tout en permettant une exploration en profondeur des variables clés influençant les taux de suicide. Il est également possible de filtrer et d'explorer les données par pays, année, sexe, ou autres critères, offrant ainsi une analyse personnalisée.

Les résultats obtenus ont révélé des tendances intéressantes, notamment l'impact du PIB et du score de bonheur sur les taux de suicide, ainsi que les différences marquées entre les sexes et les tranches d'âge. Cependant, certaines limitations subsistent, telles que la disponibilité inégale des données et la difficulté d'interpréter les corrélations comme des causalités.

En conclusion, ce tableau de bord constitue un outil précieux pour les chercheurs, les décideurs et les organismes de santé publique, en leur permettant de mieux comprendre les facteurs sous-jacents aux suicides et de guider les actions préventives. À l'avenir, des améliorations pourraient être apportées en ajoutant des fonctionnalités d'analyse prédictive et en intégrant davantage de données pour affiner les résultats et offrir une vision plus complète du phénomène.