



Projet– BI Commerciale

Réalisé par : Ala eddine bessghaier

Amal Abid

Aziz Kallel

Ce projet a pour objectif la conception et la mise en œuvre d'une application d'aide à la décision permettant de suivre et d'évaluer les performances commerciales des principales branches d'activité : Carburants, Gaz, Lubrifiants et Aviation. Il s'agit de développer un système décisionnel intégrant, traitant et visualisant les données commerciales de manière claire et dynamique. Pour ce faire, le projet comprend la modélisation d'un datamart dédié, le développement de flux ETL sous Talend, ainsi que la création de tableaux de bord interactifs avec Power BI. L'objectif final est de fournir des indicateurs pertinents permettant aux responsables commerciaux de mieux comprendre les tendances, d'identifier les opportunités et d'optimiser les performances.

1. Création de l'environnement de travail

Pour isoler les dépendances du projet et garantir un environnement propre, un environnement virtuel Python a été créé. Voici les étapes suivies :

```
python -m venv bi
```

Cette commande permet de créer un environnement virtuel nommé **bi**.

Ensuite, l'environnement a été activé avec la commande suivante :

```
.\bi\Scripts\activate
```

🔍 Prétraitement des Données de Vente

Dans cette étape, nous analysons et nettoyons les données de vente de l'année 2015 afin de les préparer pour le chargement dans le datamart

📁 Chargement et Exploration Initiale :

Le jeu de données `data_vente_2015` est d'abord exploré à l'aide des commandes `head()` et `info()`. Cela permet d'avoir une vue d'ensemble sur la structure du fichier, les types de données et la présence éventuelle de valeurs manquantes.

✓ Vérification des Valeurs Manquantes La fonction `isnull().sum()` est utilisée pour identifier les colonnes contenant des valeurs nulles. Cette étape est cruciale pour anticiper les traitements de nettoyage nécessaires.

Nettoyage des Données

Les lignes contenant des valeurs nulles sont supprimées à l'aide de `dropna()`, produisant un nouveau DataFrame propre nommé `vente_2015_ac`. Une vérification complémentaire permet de confirmer l'absence de valeurs manquantes après ce traitement.

Détection de Doublons (Données 2016)

Sur les données de vente 2016 (`vente_2016_ac`), un contrôle des doublons est effectué via la méthode `duplicated()` appliquée à la colonne `NOMCLI`. Le résultat indique l'existence de **54 776 doublons**, ce qui suggère une possible redondance ou un besoin de traitement spécifique sur les identifiants clients.

```
print(data_vente_2015.head())
print(data_vente_2015.info())
print(data_vente_2015.isnull().sum())
vente_2015_ac = data_vente_2015.dropna()
print(vente_2015_ac.isnull().sum())
```

Statistiques Descriptives :

Enfin, un résumé statistique global est généré grâce à la méthode `describe()`, fournissant des informations essentielles sur la distribution des variables quantitatives dans `vente_2016_ac`:

```
print(vente_2016_ac.describe())
```

	CODAME	CODDPO	COPRDR	NUMCPT	MOIS	CODGVR	CODLOC	SUM(MNTHT)	SUM(QTEPRD)
count	56667.000000	56667.000000	56667.000000	56667.000000	56667.000000	56667.000000	56667.000000	5.666700e+04	56667.000000
mean	28343.582003	585.105423	15.148481	28736.712566	234528.829354	6.768754	10.852771	33.915983	2.443712e+04
std	16363.894582	182.219954	11.003035	47798.375774	126017.380022	3.438160	6.972490	24.773945	9.271131e+04
min	1.000000	100.000000	1.000000	0.000000	158300.000000	1.000000	1.000000	1.000000	-5.663426e+04
25%	14174.500000	510.000000	2.000000	6.000000	153900.000000	4.000000	4.000000	14.000000	3.111730e+02
50%	28343.000000	700.000000	20.000000	12200.000000	155550.000000	7.000000	11.000000	29.000000	1.150092e+03
75%	42515.500000	700.000000	26.000000	32100.000000	229361.000000	10.000000	17.000000	47.000000	1.218750e+04
max	56689.000000	999.000000	70.000000	214010.000000	676742.000000	12.000000	24.000000	99.000000	6.047634e+06

Dictionnaire des Données :

Le tableau suivant présente une description des principales colonnes présentes dans les fichiers de vente. Ces informations sont essentielles pour identifier les dimensions et les faits dans la modélisation du datamart.

Column	Meaning
CATPRD	Product category (e.g., LUBRIFIANT, CARBURANT)
CODAME	Code for the distributor (e.g., 700)
LIBAME	Distributor name (e.g., DISTRIBUTEUR)
CODDPO / NOMDPO	Code and name of the depot or sales location
CODPRD / CODEMB / LIBPRD	Product codes and product name
NUMCPT / NOMCLI	Client account number and name
MOIS	Month of transaction
CODGVR / LIBGVR	Code and name of the governorate (region)
CODLOC / LIBLOC	Code and name of the local area
TYPOPE	Type of operation (L or C)
SUM(MNTHT)	Total amount (in money) for the product sold
SUM(QTEPRD)	Total quantity of product sold

Cette table est utilisée comme base pour identifier les **dimensions** (clients, produits, dépôts, régions, temps...) et les **faits** (quantité et montant des ventes) lors de la phase de modélisation multidimensionnelle.

Analyse exploratoire des données (EDA) :

Une première analyse a été effectuée afin d'examiner les dimensions principales présentes dans les jeux de données de vente pour les années 2015 et 2016, ainsi que dans le fichier des stations.

```
print(vente_2016_ac['NOMCLI'].nunique()) : 1891  
print(vente_2016_ac['LIBPRD'].nunique()) : 307  
print(vente_2016_ac['LIBGVR'].nunique()) : 24  
print(vente_2016_ac['LIBLOC'].nunique()) : 420
```

EDA 2015

```
print(vente_2015_ac['LIBLOC'].nunique()) : 421  
print(vente_2015_ac['LIBPRD'].nunique()) : 286  
print(vente_2015_ac['NOMCLI'].nunique()) : 1913  
print(vente_2015_ac['LIBGVR'].nunique()) : 24
```

station

```
Python ▾  
Copier Légende ...  
print(data_station_ac['STATION'].nunique()) : 223
```

Données de vente 2016 :

Le nombre de valeurs distinctes par dimension est le suivant :

- **Clients (NOMCLI)** : 1 891 clients uniques
- **Produits (LIBPRD)** : 307 produits différents
- **Régions (LIBGVR)** : 24 gouvernorats
- **Localités (LIBLOC)** : 420 localités

Données de vente 2015 :

Pour comparaison, les données de 2015 présentent :

- **Localités (LIBLOC)** : 421 localités
- **Produits (LIBPRD)** : 286 produits
- **Clients (NOMCLI)** : 1 913 clients
- **Régions (LIBGVR)** : 24 gouvernorats

Ces résultats montrent une relative stabilité du nombre de clients et de régions d'une année à l'autre, avec une légère variation au niveau des produits et localités.

Données des stations :

- **Stations (STATION)** : 223 stations uniques sont référencées dans le fichier correspondant.

client

```
print(data_client_cod['Client'].nunique()): 3553
print(data_client_cod['Secteurs clients'].nunique()): 10
print(data_client_cod['AME'].nunique()): 71
```

Données clients :

Le fichier des clients fournit des informations détaillées sur plusieurs dimensions :

- **Clients (Client)** : 3 553 clients distincts sont recensés dans la base.
- **Secteurs clients (Secteurs clients)** : 10 secteurs d'activité différents sont représentés.
- **AME** : 71 entités AME distinctes ont été identifiées.

Ces éléments permettent une segmentation pertinente des clients selon leur secteur ou leur rattachement organisationnel (AME), ce qui est essentiel pour des analyses ciblées ou la modélisation d'un datamart.

Column	Meaning	Example Value
Compte	The client's account number or code	200091
Client	Full name and address of the client	CHARGUIA SERVICE CONSEIL 28 RUE 8601 Z.I.CHARGUIA B P 65
Secteurs clients	The business sector or activity of the client	AGRICULTURE
Consolidation Client	The main client name used for grouping (cleaned version of the "Client" field)	CHARGUIA SERVICE CONSEIL
GOUVERNORAT	The region or governorate where the client is located	(Empty in this example, but typically something like TUNIS)
AME	Business activity category of the client	ACTIVITES CONSEIL, REPARATION ET ENTRETIEN (Consulting, Repair & Maintenance Activities)

gaz

```
Python ▾ Copier Légende ...  
print(data_client_gaz['Compte'].nunique()) : 773  
print(data_client_gaz['Ancien Compte'].nunique()) : 654  
print(data_client_gaz['Secteurs Statistique'].nunique()) : 12  
print(data_client_gaz['Catégorie'].nunique()) : 8  
print(data_client_gaz['Secteurs clients'].nunique()) : 18  
print(data_client_gaz['Consolidation Client'].nunique()) : 635  
print(data_client_gaz['Gouvernorat'].nunique()) : 24  
print(data_client_gaz['Client'].nunique()) : 677
```

Le **nombre de valeurs uniques** dans plusieurs colonnes du DataFrame `data_client_gaz`

jet

```
print(data_client_jet['Client'].nunique()) : 101  
print(data_client_jet['Secteurs clients'].nunique()) : 11  
print(data_client_jet['Consolidation Client'].nunique()) : 82  
print(data_client_jet['Type Clients'].nunique()): 11
```

Le **nombre de valeurs uniques** dans plusieurs colonnes du DataFrame `data_client_jet` :

Ces statistiques donnent un aperçu de la diversité des clients, de leurs classifications sectorielles, et des regroupements pour les analyses.

Column	Meaning / Description
Compte	This is a unique account number assigned to each client. Think of it like a customer ID.
Client	The name and address of the client or company.
Secteurs clients	The type of market or sector the client belongs to (e.g., foreign or Tunisian clients).
Consolidation Client	A standardized name used for grouping clients (especially if a client has multiple accounts or branches).
Type Clients	A broader classification of clients, often used for reporting or analysis (e.g., foreign vs local).

Ce tableau présente différentes **colonnes de données** utilisées pour identifier, catégoriser et regrouper les clients dans un système de gestion. Chaque ligne du tableau décrit une colonne spécifique, accompagnée de sa signification ou de son rôle , es colonnes facilitent la **centralisation, l'analyse et le suivi efficace des clients**

🔍 Description des données – data_mart :

data mart

Column	Meaning
CATPRD	Product category (e.g., LUBRIFIANT, CARBURANT)
CODAME	Code for the distributor (e.g., 700)
LIBAME	Distributor name (e.g., DISTRIBUTEUR)
CODDPO / NOMDPO	Code and name of the depot or sales location
CODPRD / CODEMB / LIBPRD	Product codes and product name
NUMCPT / NOMCLI	Client account number and name
MOIS	Month of transaction
CODGVR / LIBGVR	Code and name of the governorate (region)
CODLOC / LIBLOC	Code and name of the local area
TYPOPE	Type of operation (L or C)
SUM(MNTHT)	Total amount (in money) for the product sold
SUM(QTEPRD)	Total quantity of product sold

Les colonnes de ce tableau concernent des **données transactionnelles détaillées**, orientées vers le suivi des ventes et de la distribution , es données de data_mart forment une **table de faits** orientée vers :

- **L'analyse des ventes** (quantité, montant, produits),
- **La distribution géographique** (gouvernorat, localité),
- **La segmentation client/distributeur**,
- **Le suivi temporel** (par mois),
- Et **le suivi opérationnel** (type d'opération).

<u>Column</u>	<u>Meaning</u>	<u>Example Value</u>
Compte	The client's account number or code	200091
Client	Full name and address of the client	CHARGUIA SERVICE CONSEIL 28 RUE 8601 Z.I.CHARGUIA B P 65
Secteurs clients	The business sector or activity of the client	AGRICULTURE
Consolidation Client	The main client name used for grouping (cleaned version of the "Client" field)	CHARGUIA SERVICE CONSEIL
GOUVERNORAT	The region or governorate where the client is located	(Empty in this example, but typically something like TUNIS)
AME	Business activity category of the client	ACTIVITES CONSEIL, REPARATION ET ENTRETIEN (Consulting, Repair & Maintenance Activities)

Structure de données client (fichier GAZ) : Le fichier **clients gaz** contient des colonnes permettant une analyse fine des types de clients, leur localisation, et leur regroupement logique.

Utilités et finalités de cette structure :

- Regroupement efficace des clients
- Segmentation intelligente
- les analyses BI
- Amélioration de la qualité des données
- Outil d'aide à la décision

Column	Meaning	Example Value
Ancien Compte	The client's old account number before system update	9412
Compte	The new account number (current system)	300231
Client	Full name of the client with location	S. T. I. R. ZARZOUNA
Secteurs Statistique	Client's sector used for statistical classification	INDUSTRIEL GAZ
Catégorie	General category the client belongs to	INDUSTRIEL GAZ
Secteurs clients	Client's business sector (possibly for internal classification)	INDUSTRIEL GAZ
Consolidation Client	Standardized or grouped name for analysis	S. T. I. R.
Gouvernorat	The region/governorate where the client is located	BIZERTE

Ce tableau fournit une version enrichie et plus détaillée du profil client, intégrant non seulement l'identité et les secteurs d'activité, mais aussi un **historique des identifiants** et une **double classification sectorielle**.

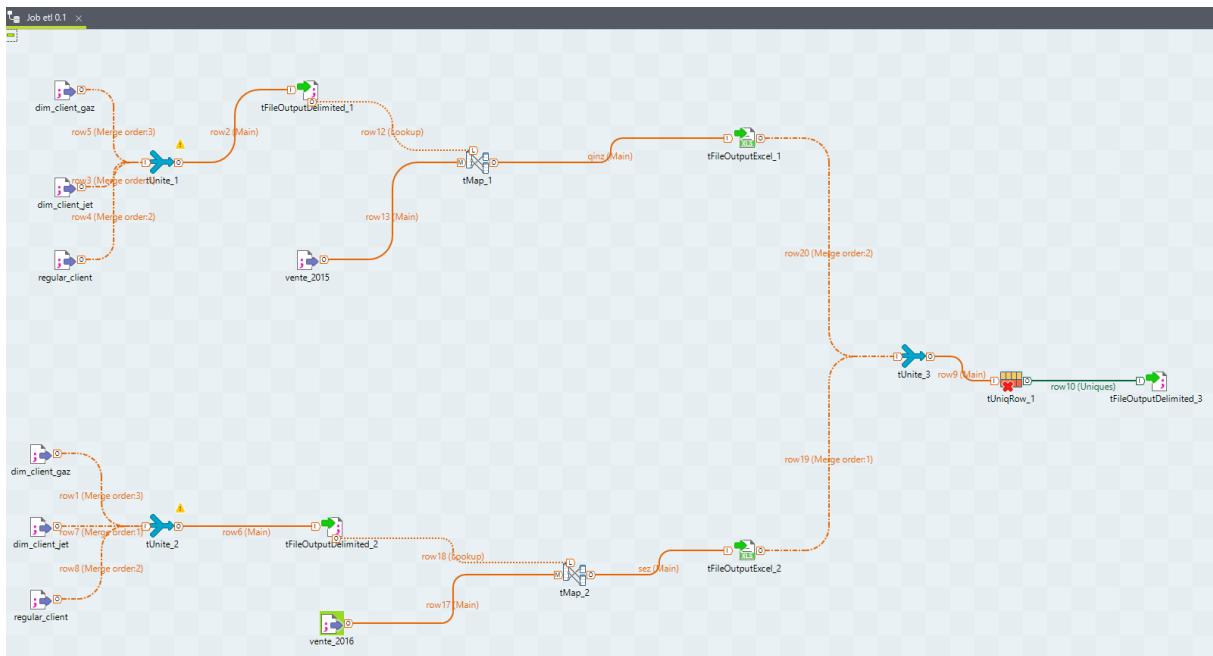
Column	Meaning / Description
Compte	This is a unique account number assigned to each client. Think of it like a customer ID.
Client	The name and address of the client or company.
Secteurs clients	The type of market or sector the client belongs to (e.g., foreign or Tunisian clients).
Consolidation Client	A standardized name used for grouping clients (especially if a client has multiple accounts or branches).
Type Clients	A broader classification of clients, often used for reporting or analysis (e.g., foreign vs local).

CODCAT	LIBCAT	+	
+	+		
CODGVR	CODLOC	LIBGVR	LIBLOC

- Jobs ETL :

Client :

- Phase 1 : Nous traitons les clients réguliers et les clients des stations de manière équivalente. L'objectif est de capturer les clients sur deux saisons, puis de fusionner les données et d'en extraire uniquement les clients uniques.
- Phase 2 : Nous extrayons spécifiquement les clients des stations à partir des deux saisons, puis nous les combinons avec les autres clients.



Phase 1 – Saisir et fusionner les données clients (Saison 2015)

1. **Sources de données clients** : dim_client_gaz, dim_client_jet, regular_client : trois sources clients différentes.
2. **Composant tUnite_1 :**
 - Combine les trois sources en un seul flux (merge).
 - Les flux sont ensuite sauvegardés avec tFileOutputDelimited_1.
3. **Composant tMap_1 :**
 - Effectue une jointure entre les clients (flux fusionné) et les ventes de 2015 (vente_2015) via une recherche (lookup).
 - Applique potentiellement des règles de transformation ou de filtrage.
4. **Sortie :**
 - Le résultat est exporté sous format Excel avec tFileOutputExcel_1.

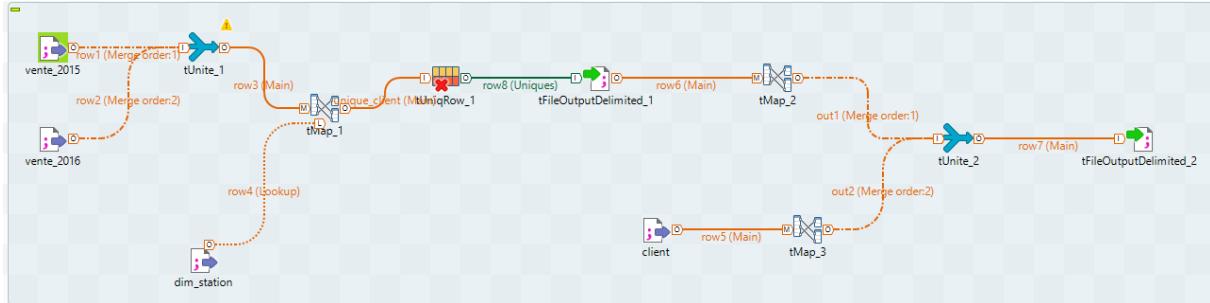
Phase 2 – Saisir et fusionner les données clients (Saison 2016) :

1. **Sources identiques :**
 - Même structure que pour 2015 : dim_client_gaz, dim_client_jet, regular_client.
2. **Composant tUnite_2 :**
 - Combine les sources en un flux unique.
 - Enregistre dans un fichier avec tFileOutputDelimited_2.
3. **Composant tMap_2 :**
 - Joint les données clients avec les ventes de 2016 (vente_2016).
 - Résultat exporté via tFileOutputExcel_2.

Phase finale – Regroupement et dé-duplication

1. **Composant tUnite_3 :**
 - Combine les résultats des deux saisons (lignes 19 et 20).

2. **Composant tUniqRow_1 :**
 - o Supprime les doublons pour obtenir uniquement les **clients uniques**.
3. **Sortie finale :**
 - o Le résultat est écrit dans un fichier **tFileOutputDelimited_3**.



Ce job ETL vise à :

- Rassembler les clients de différentes sources pour deux saisons.
- Joindre avec les ventes par saison.
- Fusionner les résultats pour obtenir une **liste unique de clients**.

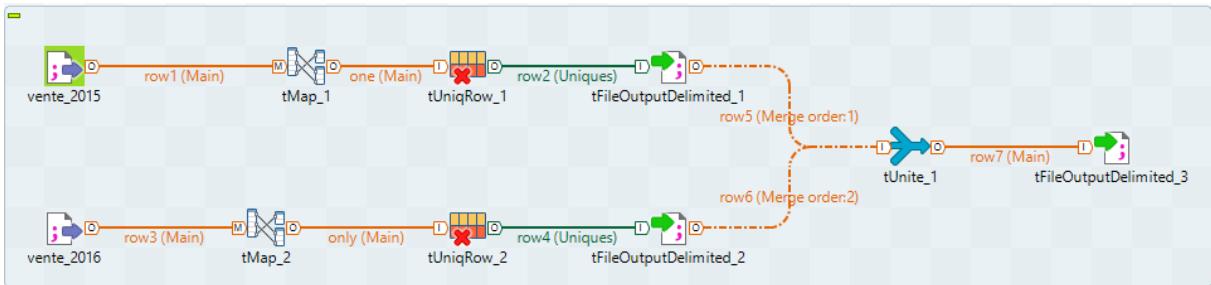
◆ Traitement des ventes 2015 et 2016 :

1. **Source de données :**
 - o Contenant les ventes de l'année 2015/2016
2. **TMap_1 :**
 - o Transformation ou sélection de certaines colonnes du dataset
 - o Sortie principale nommée one
3. **TUniqRow_1 :**
 - o Supprime les doublons pour ne garder que les ventes uniques de 2015/2016
4. **TFileOutputDelimited_1 :**
 - o Sauvegarde les ventes uniques de 2015/2016 dans un fichier

◆ Fusion finale :

1. **TUnite_1 :**
 - o Fusionne les deux fichiers de ventes uniques (2015 et 2016).
 - o Les ordres de fusion sont définis par Merge order :1 et Merge order :2
2. **TFileOutputDelimited_3 :**
 - o Enregistre le fichier final contenant toutes les ventes uniques de 2015 et 2016 combinées.

Date :



Ce job permet de :

- Nettoyer les données de ventes (éliminer les doublons),
- Transformer les données si nécessaire,
- Fusionner les résultats dans un fichier final consolidé.

Product :

Ce job ETL représente un **flux de consolidation des données de ventes des années 2015 et 2016**, avec un accent sur la **fusion, transformation et déduplication**

1. Sources de données :

- Vente_2016 et vente_2015 : fichiers sources contenant les ventes de chaque année.

2. tUnite_1 :

- **But** : Fusionner les deux flux de ventes.
- **Remarques :**
 - Merge order:1 pour vente_2016
 - Merge order:2 pour vente_2015

3. tMap_1 :

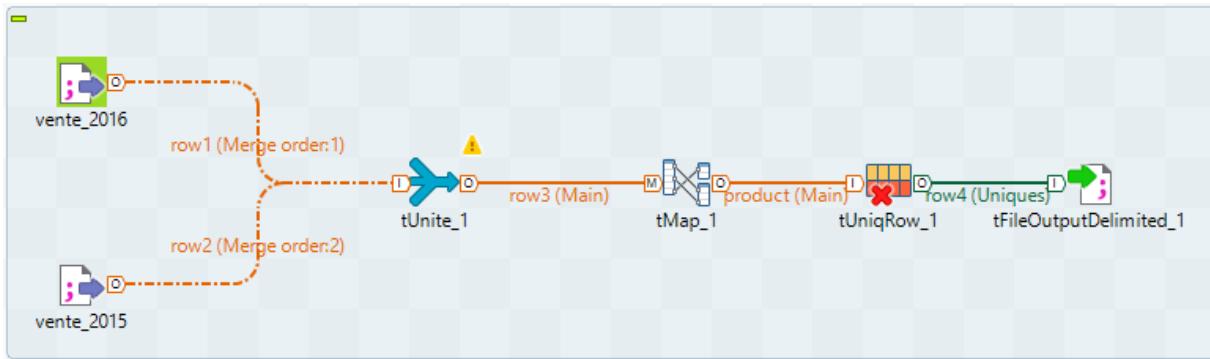
- **But** : Transformation des données (filtrage, renommage de colonnes, calculs ou enrichissement).
- **Sortie nommée** : product (suggérant qu'on se concentre ici sur un champ "produit").

4. tUniqRow_1 :

- **But** : Éliminer les doublons dans les données issues de tMap_1.

5. tFileOutputDelimited_1 :

- **But** : Sauvegarde du résultat final (ventes uniques après transformation) dans un fichier au format texte délimité (CSV, TSV...).

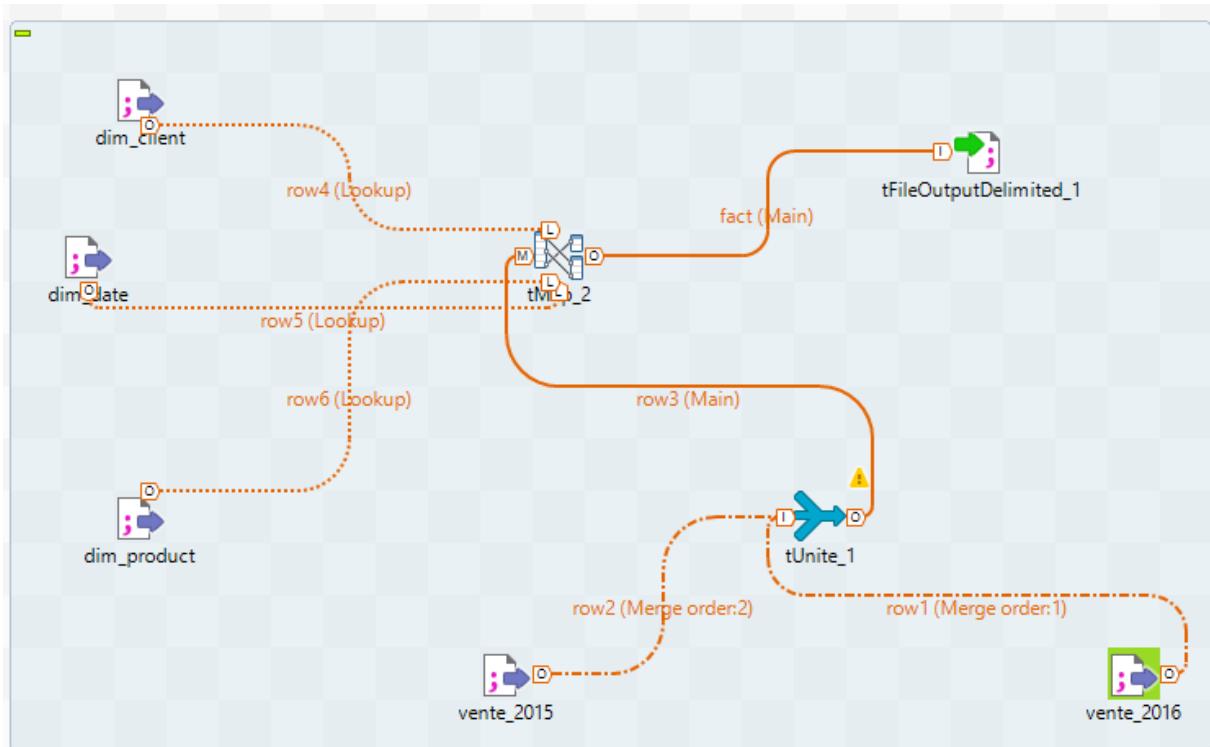


Ce job est **plus optimisé** que les précédents, car :

- Il **fusionne directement** les deux années avant traitement.
- Il **centralise la transformation et le filtrage** dans un seul composant (tMap_1).
- Il évite la duplication d'étapes entre les deux années.

Fact table :

Ce job ETL représente un **flux de construction de la table de faits** à partir des données de ventes et de dimensions (client, date, produit)



1. Sources des ventes :

- Vente_2016
 - Vente_2015
- ➡ Ces deux jeux de données sont fusionnés avec tUnite_1 (ordre 1 et 2 respectivement).

2. tUnite_1 :

- **But** : Fusionner les ventes des deux années dans un seul flux (row3).

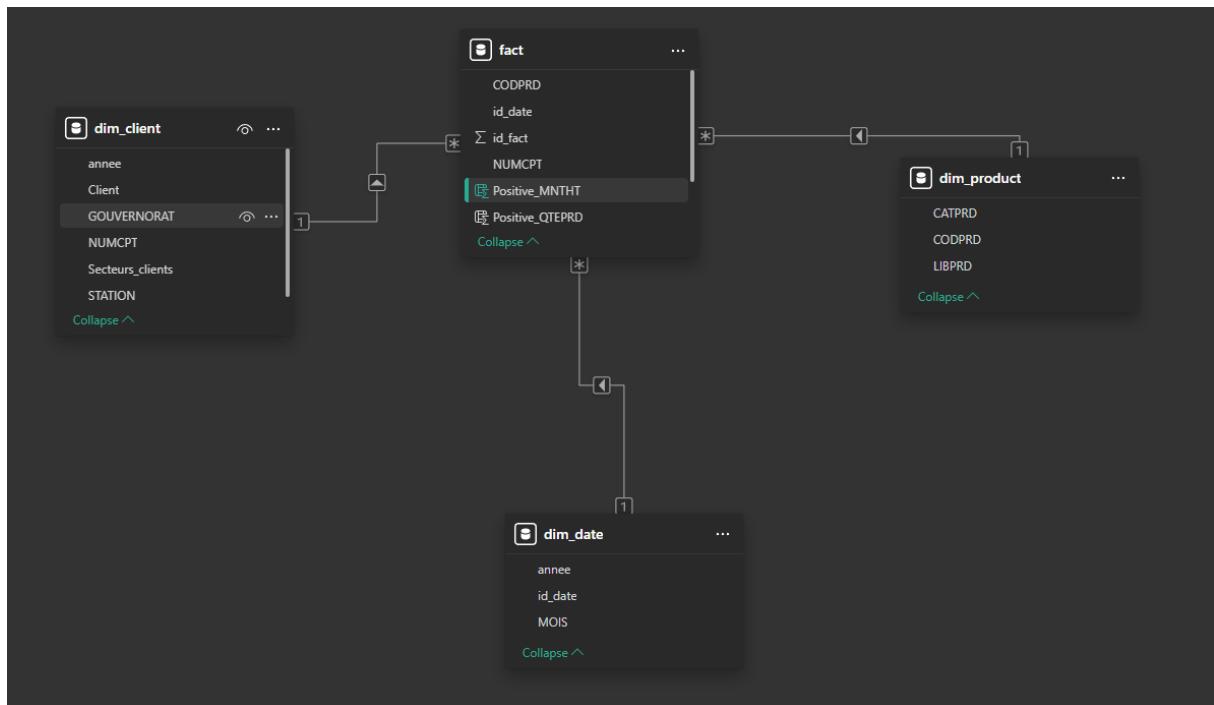
3. tMap_2 :

- **But : Transformation et enrichissement** via jointures :
 - Row 4: jointure avec la dimension **client** (dim_client)
 - Row 5 : jointure avec la dimension **date** (dim_date)
 - Row 6: jointure avec la dimension **produit** (dim_product)
- Ces jointures permettent d'ajouter des **clés substituts ou des attributs** issus des dimensions, nécessaires pour construire la table de faits.

4. Sortie principale : fact :

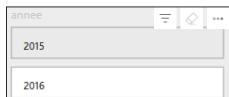
- Les données enrichies sont envoyées vers un fichier délimité via **tFileOutputDelimited_1**, représentant la **table de faits finale**.

★ Modèle en étoile dans Power BI :



La table de faits centrale (fact) liée aux dimensions dim_client, dim_product et dim_date. Il permet d'analyser les ventes par client, produit et période grâce à des relations un-à-plusieurs, facilitant les analyses décisionnelles.

AGIL SALES REPORT Store Sales Overview



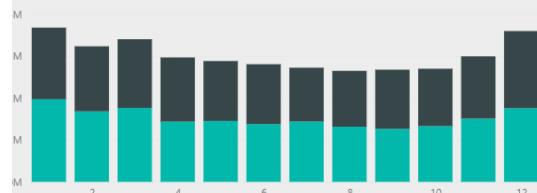
ChiffreAffaires by MOIS and annee

annee ● 2015 ● 2016

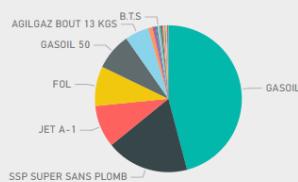


QuantiteVendue by MOIS and annee

annee ● 2015 ● 2016



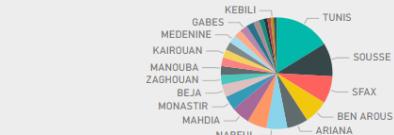
Total Sales by LIBPRD



Total Sales and Total Sales by Client



Total Sales by GOVERNORAT



AGIL SALES REPORT Store Sales Overview



ChiffreAffaires by MOIS and annee

annee ● 2015



QuantiteVendue by MOIS and annee

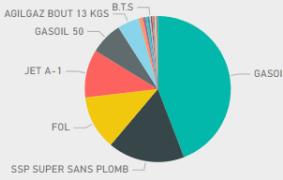
annee ● 2015



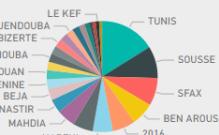
Total Sales and Total Sales by Client

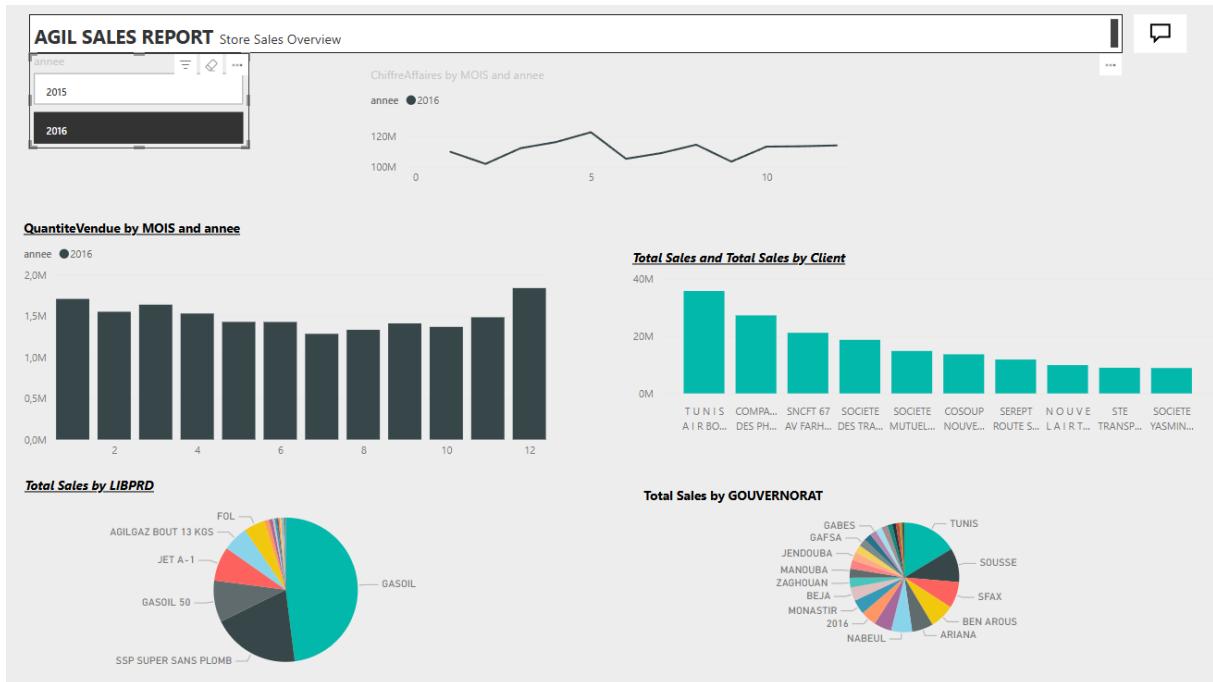


Total Sales by LIBPRD



Total Sales by GOVERNORAT





Ce tableau de bord présente une vue d'ensemble des ventes par année, mois, client, produit et gouvernorat. Il permet de comparer les performances de 2015 et 2016, d'identifier les clients et produits les plus rentables, ainsi que de visualiser la répartition des ventes par région.

Conclusion :

En conclusion, le rapport met en évidence une répartition détaillée des ventes en 2015, avec une performance marquée des produits comme le gasoil, une concentration des ventes dans certains gouvernorats tels que Tunis et Sousse, et des clients majeurs comme la STEG. Ces insights permettent d'orienter les décisions stratégiques pour optimiser les ventes futures.