



Rapport de mini Projet.

*Exploration et Analyse de
AdventureWorks*

Réalisé Par :

- BHIHI Hajar
- ELANSARI Salma

Encadré Par :

- Pr.ELBAGHAZAOUI
Bahaa Eddine

Sommaire.

Rapport de mini Projet

01 **Introduction.**

02 **Méthodologie**

03 **Analyse et Visualisation**

04 **Gestion de Projet et Collaboration**

05 **Conclusion et Perspectives**

01 Introduction.

Contexte :

AdventureWorks est une base de données fictive développée par Microsoft, utilisée à des fins éducatives et de démonstration. Elle reproduit les activités d'une entreprise multinationale de fabrication, AdventureWorks Cycles, spécialisée dans la production et la vente de vélos, pièces et accessoires. Grâce à ses données réalistes, elle constitue une ressource précieuse pour apprendre et pratiquer des techniques avancées en gestion et analyse des données.

Objectif du projet

L'objectif principal de ce projet est de mener une analyse approfondie des données de vente d'AdventureWorks afin d'extraire des insights décisionnels. Ces analyses permettront de mieux comprendre les performances commerciales, d'identifier les zones géographiques et les produits les plus performants, et de visualiser les tendances des ventes au fil du temps.



Approche générale

Pour atteindre cet objectif, nous avons suivi plusieurs étapes :

1. **Exploration des données** : Analyse des tables et identification des relations.
2. **Nettoyage des données** : Suppression des anomalies et amélioration de la qualité des données.
3. **Visualisation des données** : Création de tableaux de bord interactifs dans Power BI.
4. **Gestion de projet et collaboration** : Utilisation d'outils modernes comme Taiga.io et GitLab pour organiser les tâches et faciliter le travail collaboratif.

02 Méthodologie

Téléchargement et Installation

La base de données AdventureWorks2019 a été téléchargée depuis le site officiel de Microsoft et installée sur SQL Server Management Studio (SSMS). Une fois la base importée, nous avons configuré l'environnement pour interagir avec les tables et exécuter des requêtes SQL. Par la suite, les tables pertinentes ont été exportées au format CSV afin de faciliter leur manipulation dans des outils externes comme KNIME pour les étapes de nettoyage et d'analyse.

Exploration des données

Nous avons exploré les tables liées aux ventes, notamment :

- SalesOrderDetail : Cette table contient les détails spécifiques de chaque commande, tels que les produits commandés, les quantités, et les prix unitaires. Elle permet d'analyser les performances des produits et les tendances d'achat des clients.
- SalesOrderHeader : Cette table regroupe les informations générales des commandes, comme la date de commande, le montant total, et le statut de la commande. Elle est essentielle pour suivre les ventes globales et calculer les revenus.
- SalesTerritory : Cette table fournit des informations sur les territoires de vente, comme la région géographique et le groupe de vente associé. Elle aide à identifier les zones géographiques les plus performantes.
- Store : Cette table contient les détails sur les magasins, tels que leurs noms et leurs localisations. Elle est utile pour analyser les performances des magasins et leur contribution aux ventes.
- Customer : Cette table regroupe les informations sur les clients, comme leurs noms, leurs adresses, et leurs affiliations à un magasin. Elle est cruciale pour segmenter les clients et personnaliser les stratégies marketing.

- SalesPerson : Cette table contient des données sur les représentants commerciaux, comme leurs objectifs de vente et leurs réalisations. Elle permet d'évaluer la performance des commerciaux et leur impact sur les ventes.
- SalesReason : Cette table liste les raisons pour lesquelles des ventes ont été réalisées (par exemple, promotions ou recommandations). Elle permet d'analyser les facteurs de motivation des clients.
- SpecialOffer : Cette table contient des informations sur les offres spéciales, telles que les réductions applicables et les périodes de validité. Elle aide à évaluer l'efficacité des promotions sur les ventes.
- SalesTerritoryHistory : Cette table retrace l'historique des changements des territoires de vente, incluant les dates de début et de fin. Elle est utile pour analyser l'évolution des territoires et leurs impacts sur les performances.

Chaque table a joué un rôle clé dans la structuration et la compréhension des données liées aux ventes, permettant une analyse approfondie et des visualisations ciblées.

L'exploration a permis d'identifier les relations entre ces tables et de détecter des anomalies comme :

- Données manquantes ou incorrectes.
- Doublons.
- Incohérences dans les formats de données.

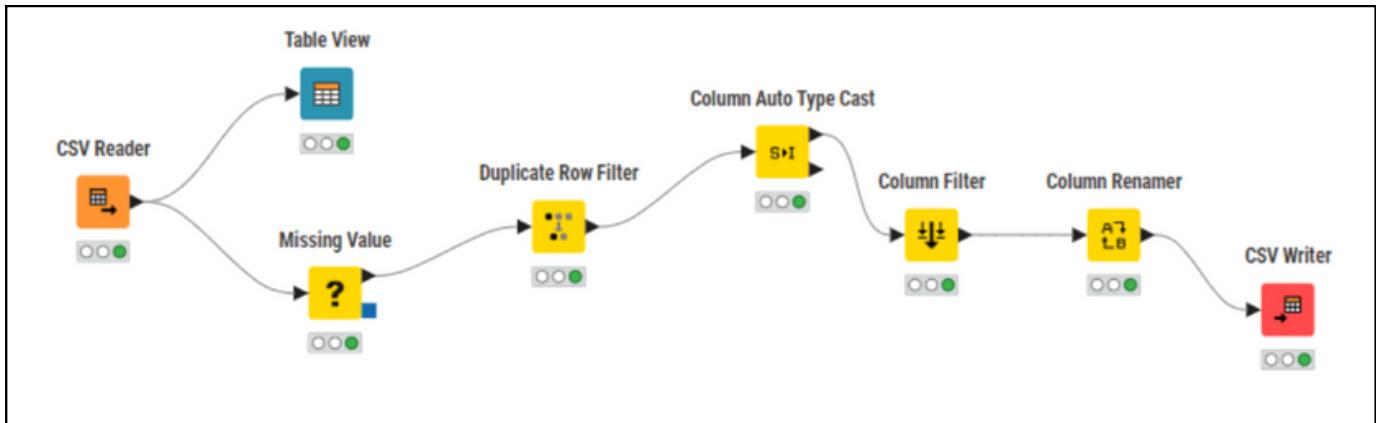
Nettoyage des données

Nous avons utilisé KNIME, une plateforme de data science, pour nettoyer les données.

Introduction à KNIME

KNIME (Konstanz Information Miner) est une plateforme open-source de data science qui permet d'analyser, transformer et modéliser des données de manière visuelle. Elle offre une interface utilisateur intuitive basée sur des workflows, où chaque étape du traitement des données est représentée par un nœud connecté à d'autres. KNIME prend en charge une large gamme d'extensions et d'intégrations, ce qui en fait un outil puissant pour le nettoyage, l'analyse, et la visualisation des données, sans nécessiter de compétences avancées en programmation.

Dans ce projet, nous avons utilisé KNIME pour effectuer le nettoyage des données des différentes tables exportées de la base AdventureWorks. Ce processus a été appliqué à toutes les tables en suivant le même schéma de nettoyage, comme illustré ci-dessus.



Importance de chaque nœud dans le workflow

1. CSV Reader

- Utilité : Ce nœud est utilisé pour importer les fichiers CSV contenant les données brutes de chaque table. Il configure les colonnes et types de données automatiquement pour une manipulation immédiate.

2. Table View

- Utilité : Ce nœud permet de visualiser les données importées avant de commencer les étapes de nettoyage. Il aide à identifier les anomalies, comme les valeurs manquantes ou incohérentes.

3. Missing Value

- Utilité : Ce nœud traite les valeurs nulles ou manquantes. Nous avons remplacé les valeurs manquantes par des valeurs par défaut ou des moyennes, selon la nature de la colonne, pour garantir la complétude des données.

4. Duplicate Row Filter

- Utilité : Ce nœud identifie et supprime les lignes dupliquées dans les données, garantissant que chaque enregistrement est unique. Cela est essentiel pour éviter les biais dans les analyses futures.

5.Column Auto Type Cast

- Utilité : Ce nœud ajuste automatiquement les types de données (par exemple, convertir une chaîne de caractères en un entier ou une date) pour correspondre au format attendu. Cela garantit la cohérence et la compatibilité avec les outils d'analyse.

6.Column Filter

- Utilité : Ce nœud permet de sélectionner les colonnes pertinentes pour l'analyse et d'exclure celles qui sont inutiles ou redondantes, simplifiant ainsi les datasets.

7.Column Renamer

- Utilité : Ce nœud standardise et renomme les colonnes pour les rendre plus compréhensibles et cohérentes à travers toutes les tables. Cela facilite l'intégration avec des outils comme Power BI.

8.CSV Writer

- Utilité : Ce nœud exporte les données nettoyées dans un fichier CSV prêt pour l'analyse. Cela permet une intégration directe avec des outils comme Power BI ou Tableau.

Pour SalesOrderDetail :

CSV Reader

- Le fichier CSV contenant les données brutes de la table SalesOrderDetail a été importé à l'aide de ce nœud.
- Résultat : Les données initiales ont été chargées avec toutes les colonnes disponibles, prêtes pour le nettoyage.

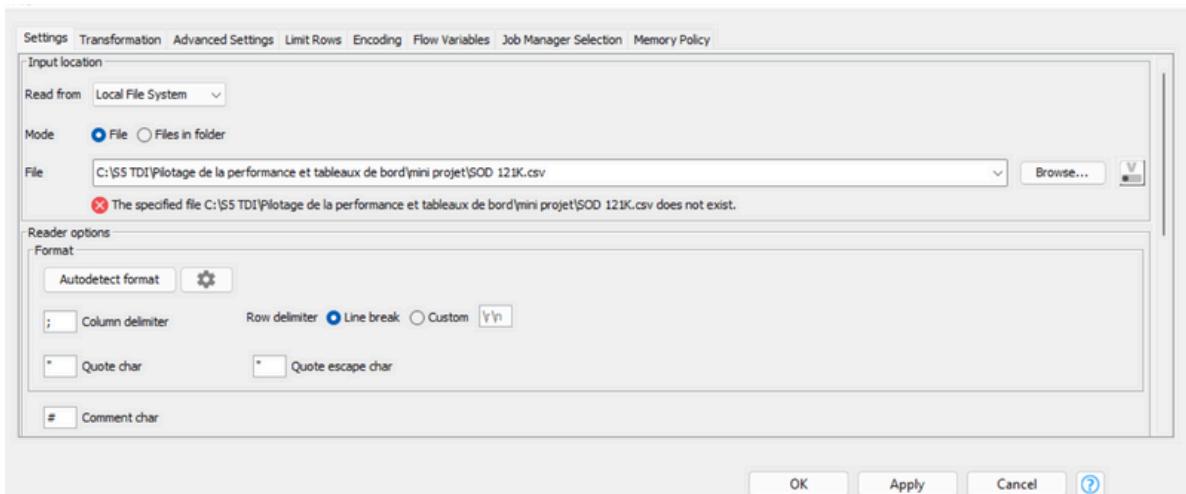


Table View

- Une visualisation préliminaire des données a été effectuée pour identifier les anomalies.
- Résultat :
 - Quelques lignes contenaient des valeurs nulles dans des colonnes spécifiques.
 - Des incohérences ont été constatées dans les types de données (par exemple, des dates mal formatées).

The screenshot shows the 'Table View' dialog box. On the left is a preview grid of 12 rows of data with 11 columns. The columns are labeled: RowID, SalesOrderID, SalesOrderDetailID, CarrierTrackingNumber, OrderQty, ProductID, SpecialOfferID, UnitPrice, and two unnamed columns. The preview shows various values like '43659' and '1'. To the right of the preview are several configuration sections:

- Data** section: Includes tabs for 'Manual', 'Wildcard', 'Regex', and 'Type'. The 'Type' tab is selected. It also contains a search bar and a list of included columns: SalesOrderID, SalesOrderDetailID, CarrierTrackingNumber, OrderQty, ProductID, SpecialOfferID, and UnitPrice.
- Displayed columns**: A list of columns: SalesOrderID, SalesOrderDetailID, CarrierTrackingNumber, OrderQty, ProductID, SpecialOfferID, and UnitPrice.
- Excludes**: A list box containing the message "No columns in this list".
- Includes**: A list box containing the same list of columns as the displayed columns.
- View** section: Contains options for 'Show row numbers' (unchecked) and 'Show RowIDs' (checked).
- Title**: A text input field containing "Table View".
- Show table size**: A checked checkbox.
- Show column data types in header**: A checked checkbox.
- Buttons**: 'Cancel' and 'Ok' buttons.

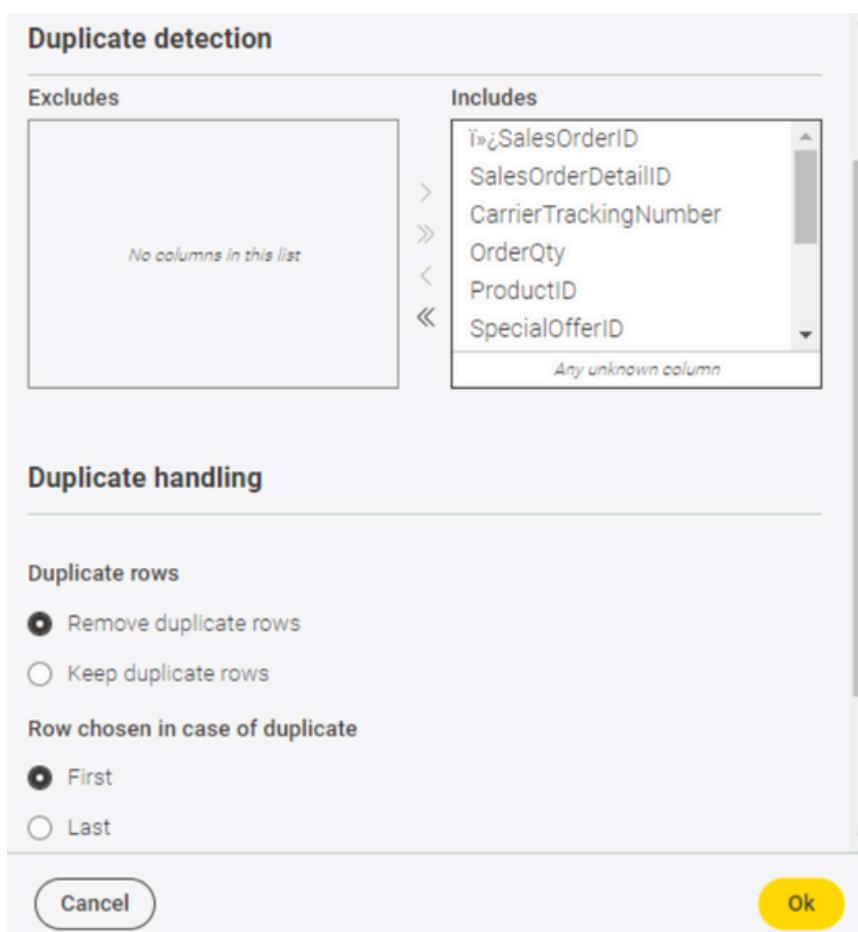
Missing Value

- Les valeurs manquantes ont été remplacées pour éviter les lacunes dans les analyses futures.
- Détails :
 - Pour les colonnes numériques (par exemple, UnitPrice), les valeurs manquantes ont été remplacées par la moyenne des valeurs non nulles.
 - Pour les colonnes catégoriques (par exemple, ModifiedDate), une valeur par défaut ou la valeur la plus fréquente a été utilisée.

Number (integer)	Mean
String	Remove Row*
Number (double)	Median

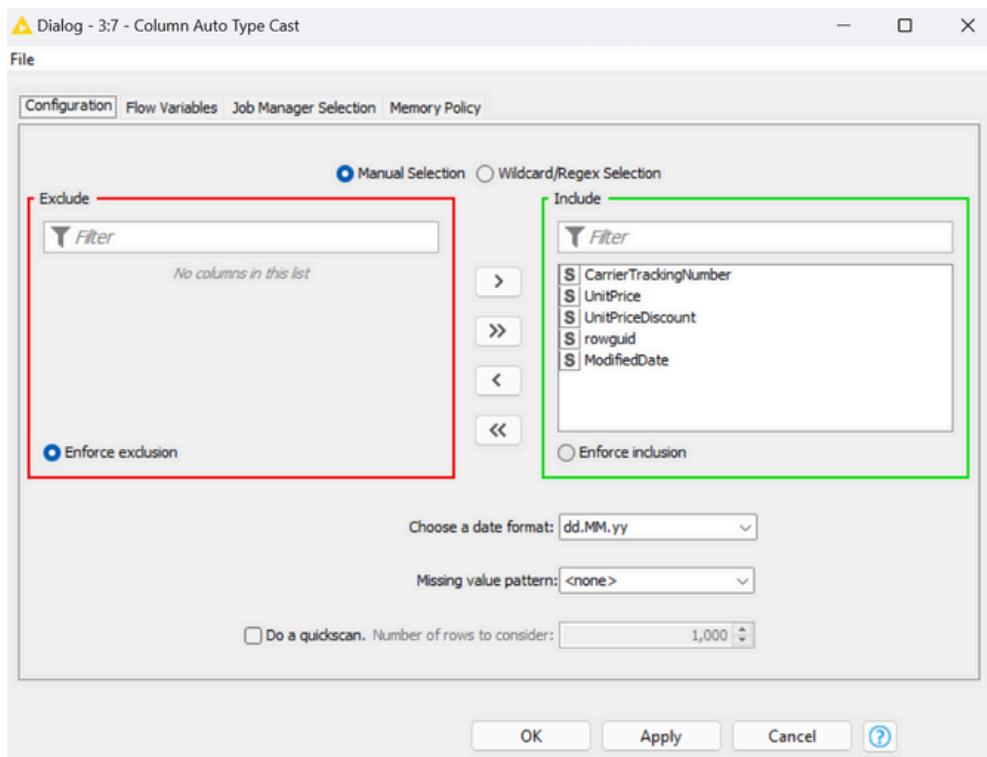
Duplicate Row Filter

- Les lignes dupliquées ont été supprimées pour garantir l'unicité des enregistrements.
- Résultat :
 - Plusieurs lignes redondantes, issues de doublons dans les données d'origine, ont été éliminées.



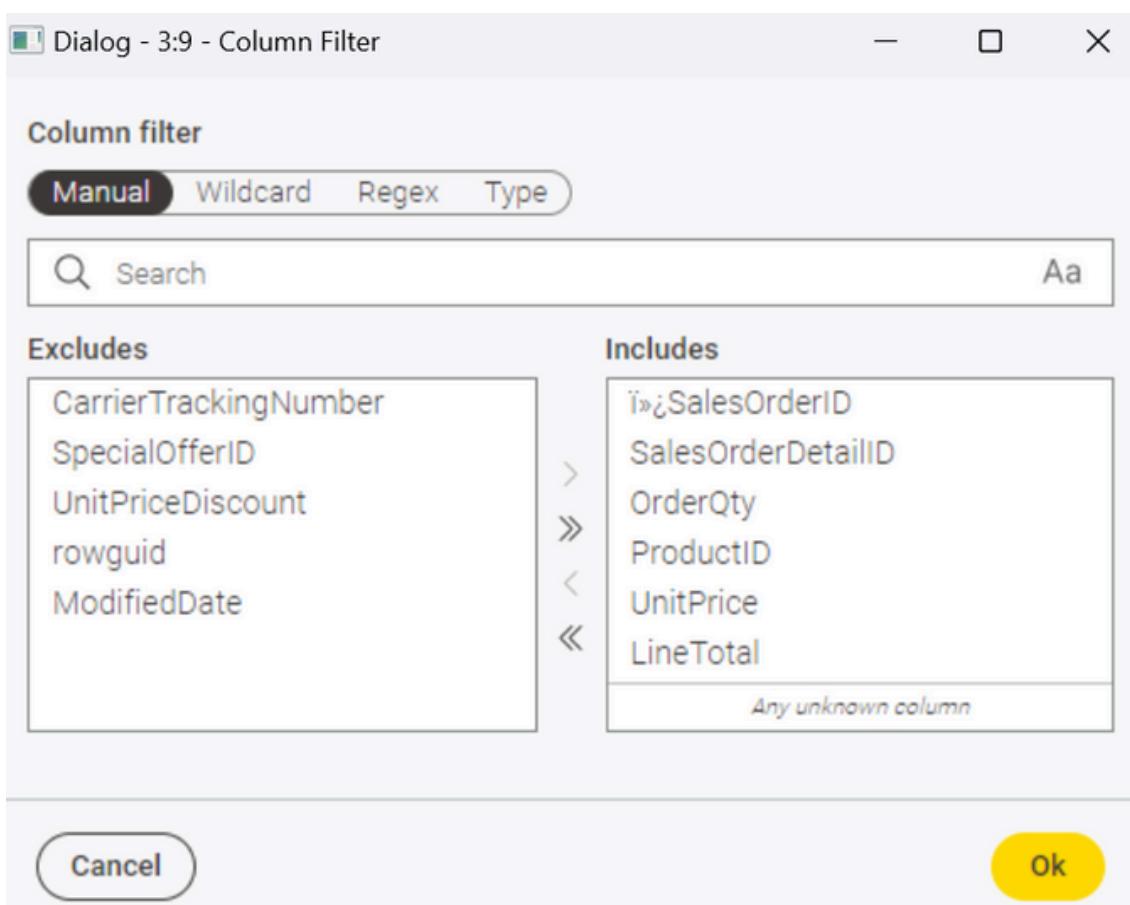
Column Auto Type Cast

- Les types de données ont été automatiquement ajustés pour correspondre à leur nature réelle.
- Détails :
 - Les colonnes numériques comme OrderQty et LineTotal ont été converties en type float.
 - Les dates dans ModifiedDate ont été formatées comme des types datetime standard.



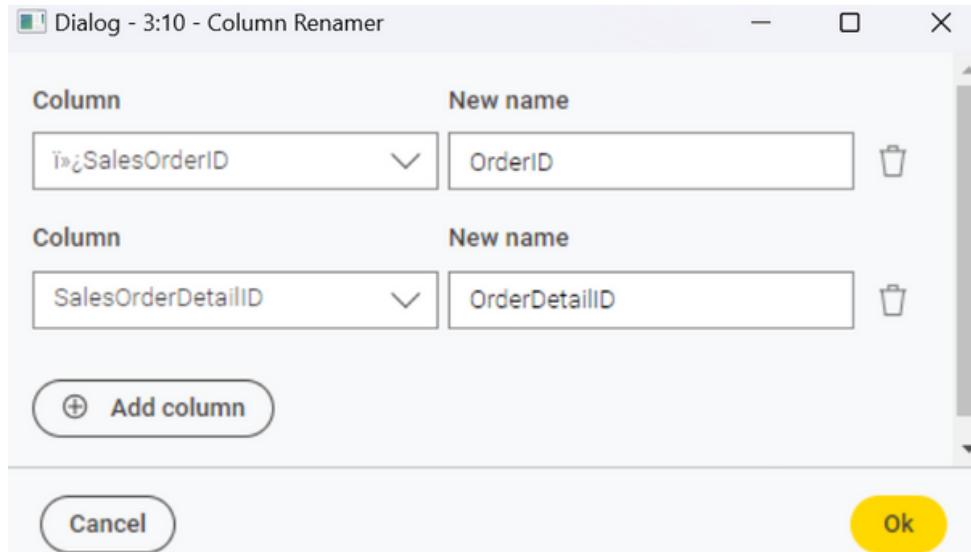
Column Filter

- Certaines colonnes inutiles ou redondantes ont été exclues pour simplifier le jeu de données.
- Détails :
 - Les colonnes qui n'ajoutaient pas de valeur analytique ont été filtrées. Par exemple, les identifiants techniques inutiles dans l'analyse finale.



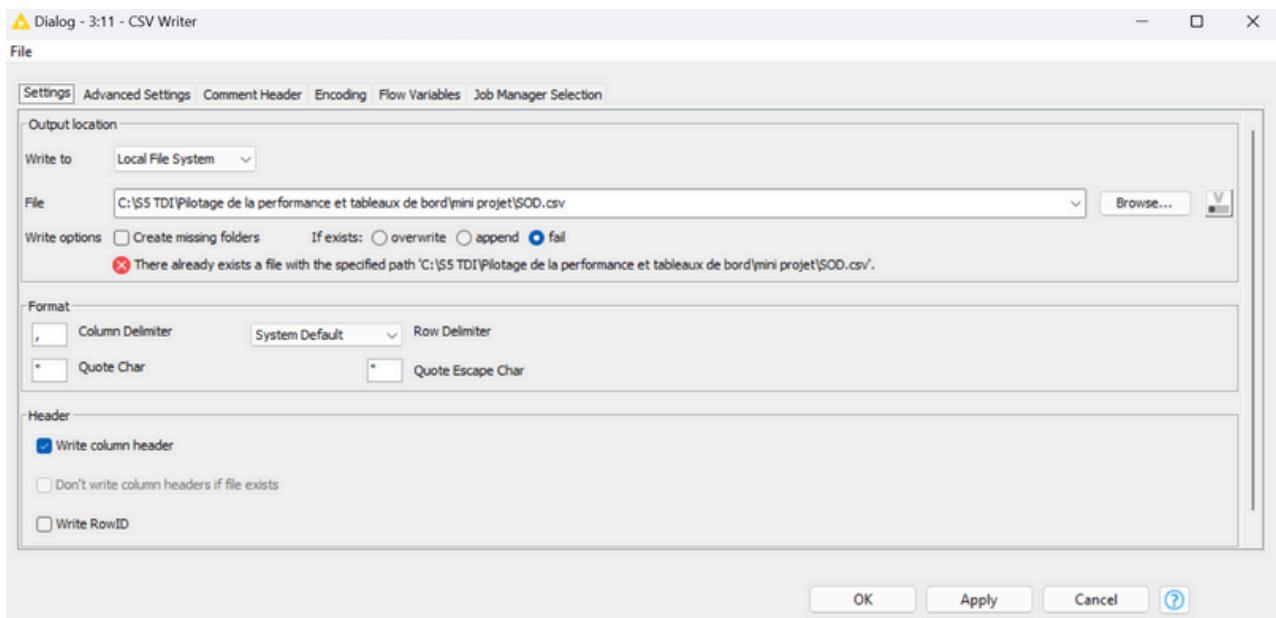
Column Renamer

- Les noms des colonnes ont été standardisés pour améliorer leur clarté.



CSV Writer

- Utilité : Les données nettoyées ont été exportées sous un format CSV pour une utilisation ultérieure dans des outils de visualisation.
- Résultat :
 - Un fichier CSV propre et prêt pour l'analyse a été généré, intégrant toutes les transformations réalisées.



Pour SalesStore :

CSV Reader

- Importation des données brutes de la table Store.

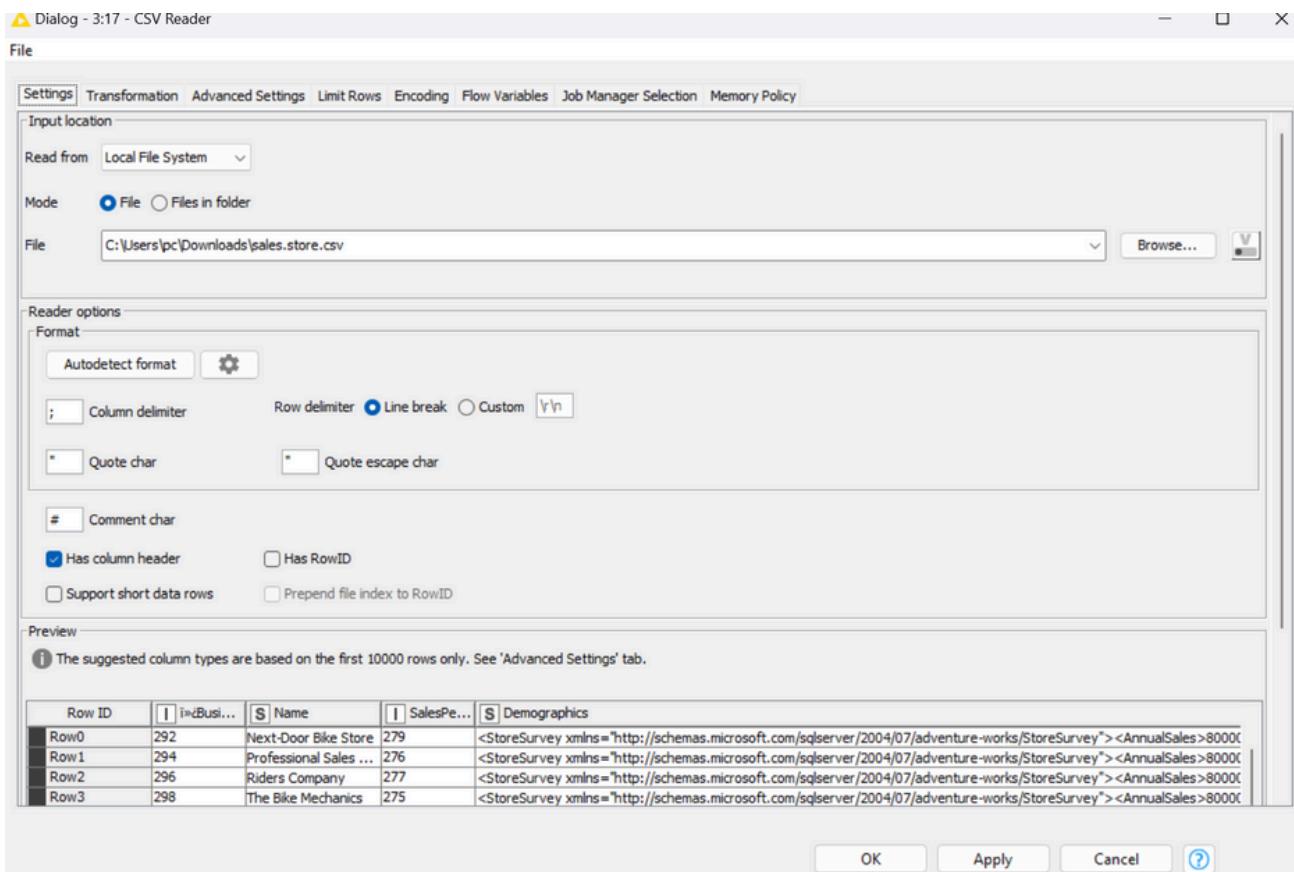


Table View

- Visualisation des données pour identifier les anomalies : quelques valeurs nulles détectées dans des colonnes .

Table View						
Rows: 701	Columns: 6					
RowID	BusinessEntityID	Name	SalesPersonID	Demographic	rowguid	ModifiedDate
Row0	292	Next-Door Bike Store	279	<StoreSurvey xmlns="http://schemas.microsoft.com/sqlserver/2004/07/adventure-works/StoreSurvey"><AnnualSales>8000</AnnualSales>	A22517E3-84...	2014-09-01
Row1	294	Professional Sales ...	276	<StoreSurvey xmlns="http://schemas.microsoft.com/sqlserver/2004/07/adventure-works/StoreSurvey"><AnnualSales>8000</AnnualSales>	B50CA50B-C...	2014-09-01
Row2	296	Riders Company	277	<StoreSurvey xmlns="http://schemas.microsoft.com/sqlserver/2004/07/adventure-works/StoreSurvey"><AnnualSales>8000</AnnualSales>	337C3688-13...	2014-09-01
Row3	298	The Bike Mechanics	275	<StoreSurvey xmlns="http://schemas.microsoft.com/sqlserver/2004/07/adventure-works/StoreSurvey"><AnnualSales>8000</AnnualSales>	7894F278-F0...	2014-09-01
Row4	300	Nationwide S...	286	<StoreSurvey xmlns="http://schemas.microsoft.com/sqlserver/2004/07/adventure-works/StoreSurvey"><AnnualSales>8000</AnnualSales>	C3FC9705-A8...	2014-09-01
Row5	302	Area Bike Acc...	281	<StoreSurvey xmlns="http://schemas.microsoft.com/sqlserver/2004/07/adventure-works/StoreSurvey"><AnnualSales>8000</AnnualSales>	368BE6DD-30...	2014-09-01
Row6	304	Bicycle Accesso...	283	<StoreSurvey xmlns="http://schemas.microsoft.com/sqlserver/2004/07/adventure-works/StoreSurvey"><AnnualSales>8000</AnnualSales>	35F40636-51...	2014-09-01
Row7	306	Clamps & Bra...	275	<StoreSurvey xmlns="http://schemas.microsoft.com/sqlserver/2004/07/adventure-works/StoreSurvey"><AnnualSales>8000</AnnualSales>	64D06BFC-D...	2014-09-01
Row8	308	Valley Bicycle...	277	<StoreSurvey xmlns="http://schemas.microsoft.com/sqlserver/2004/07/adventure-works/StoreSurvey"><AnnualSales>8000</AnnualSales>	59386B0C-65...	2014-09-01
Row9	310	New Bikes Co...	279	<StoreSurvey xmlns="http://schemas.microsoft.com/sqlserver/2004/07/adventure-works/StoreSurvey"><AnnualSales>8000</AnnualSales>	47E4B6BD-5...	2014-09-01
Row...	312	Vinyl and Pla...	282	<StoreSurvey xmlns="http://schemas.microsoft.com/sqlserver/2004/07/adventure-works/StoreSurvey"><AnnualSales>8000</AnnualSales>	DC610525-E3...	2014-09-01
Row...	314	Top of the Lin...	288	<StoreSurvey xmlns="http://schemas.microsoft.com/sqlserver/2004/07/adventure-works/StoreSurvey"><AnnualSales>8000</AnnualSales>	E290E93F-A9...	2014-09-01
Row...	316	Fun Toys and ...	281	<StoreSurvey xmlns="http://schemas.microsoft.com/sqlserver/2004/07/adventure-works/StoreSurvey"><AnnualSales>8000</AnnualSales>	6CDCF941-41...	2014-09-01
Row...	318	Great Bikes	283	<StoreSurvey xmlns="http://schemas.microsoft.com/sqlserver/2004/07/adventure-works/StoreSurvey"><AnnualSales>8000</AnnualSales>	956FBC35-5E...	2014-09-01

Data

Displayed columns: Manual, Wildcard, Regex, Type

Search: Aa

Excludes: No columns in this list

Includes: BusinessEntityID, Name, SalesPersonID, Demographics, rowguid, ModifiedDate, Any unknown column

Show row numbers:

Show RowIDs:

View

Title: Table View

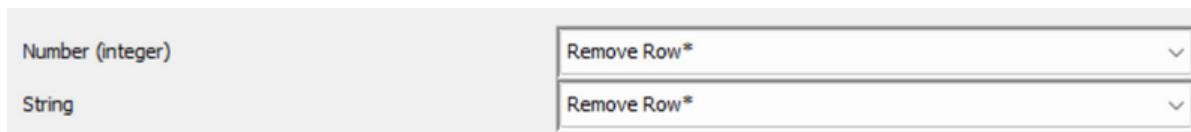
Show table size:

Show column data types in header:

Cancel Ok

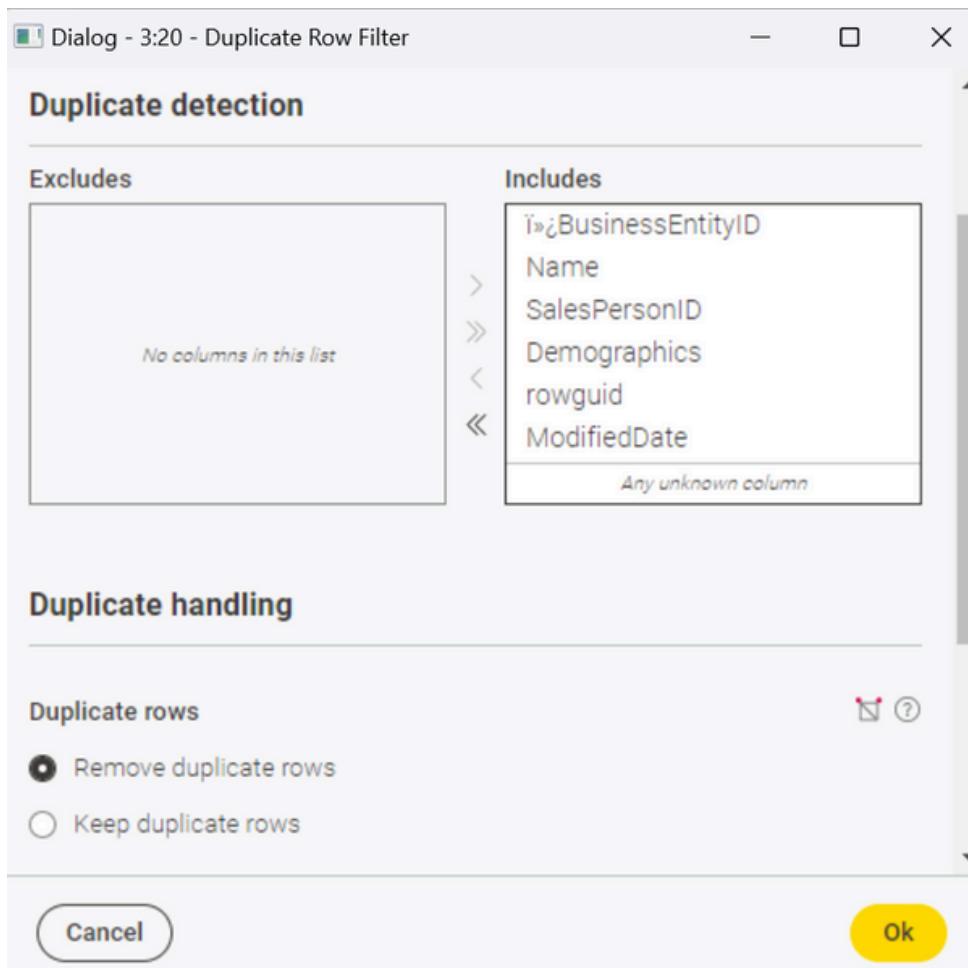
Missing Value

- Traitement des valeurs nulles :
 - Remplacement des champs textuels manquants par une valeur par défaut.



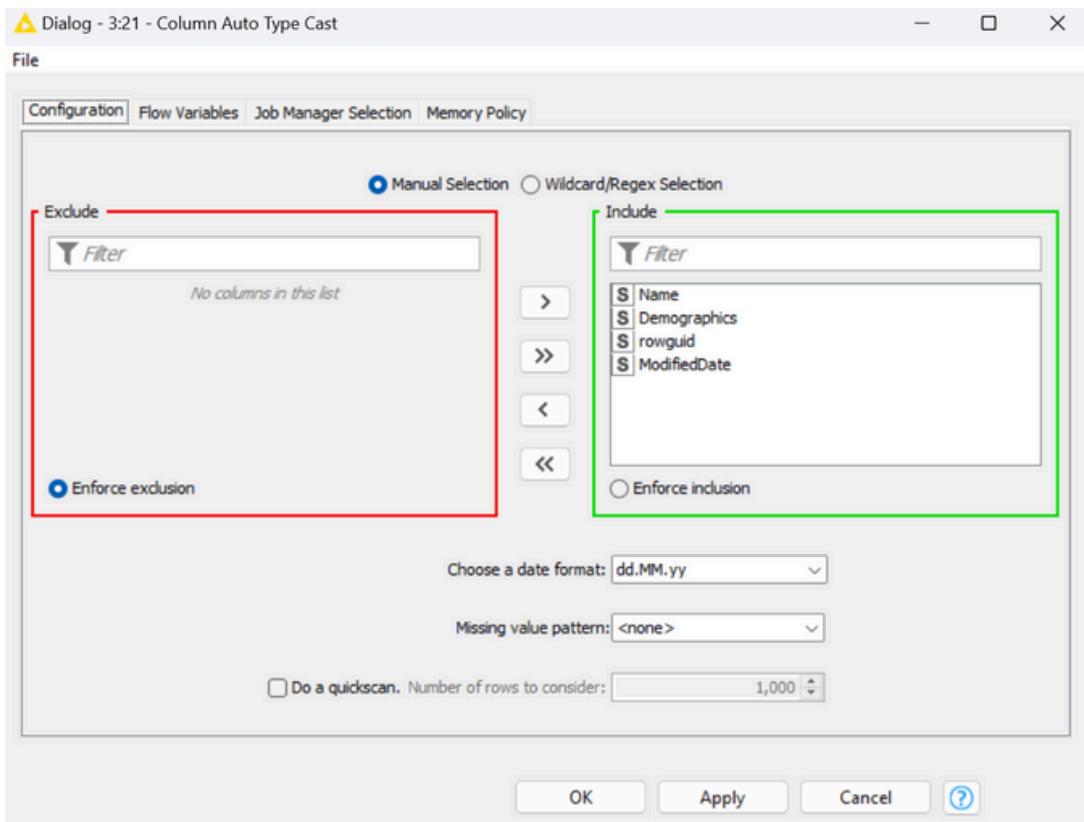
Duplicate Row Filter

- Suppression des lignes dupliquées pour garantir l'unicité des enregistrements.



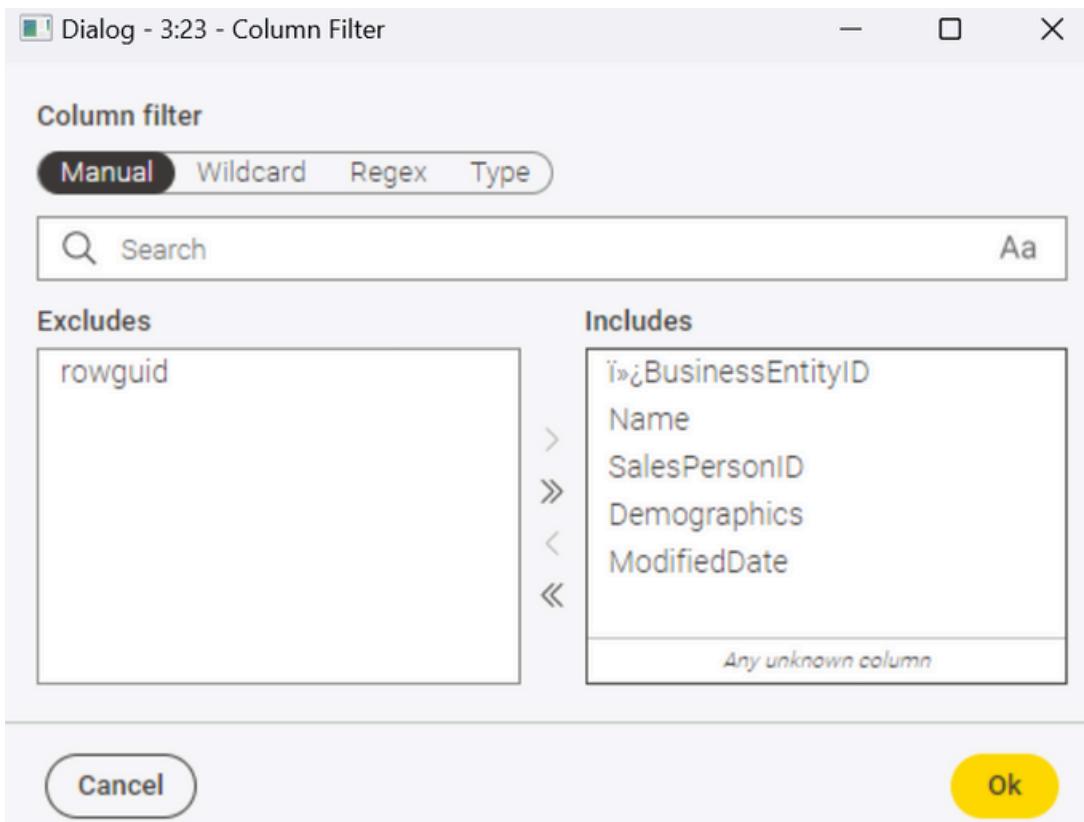
Column Auto Type Cast

- Ajustement automatique des types :
 - Conversion des identifiants en type entier.
 - Correction des formats de date dans ModifiedDate.



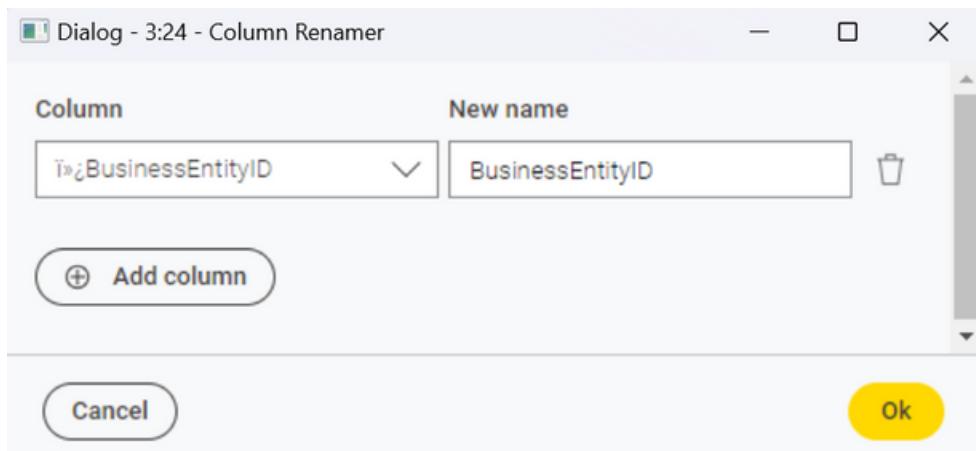
Column Filter

- Suppression des colonnes inutiles comme les métadonnées techniques non nécessaires à l'analyse.



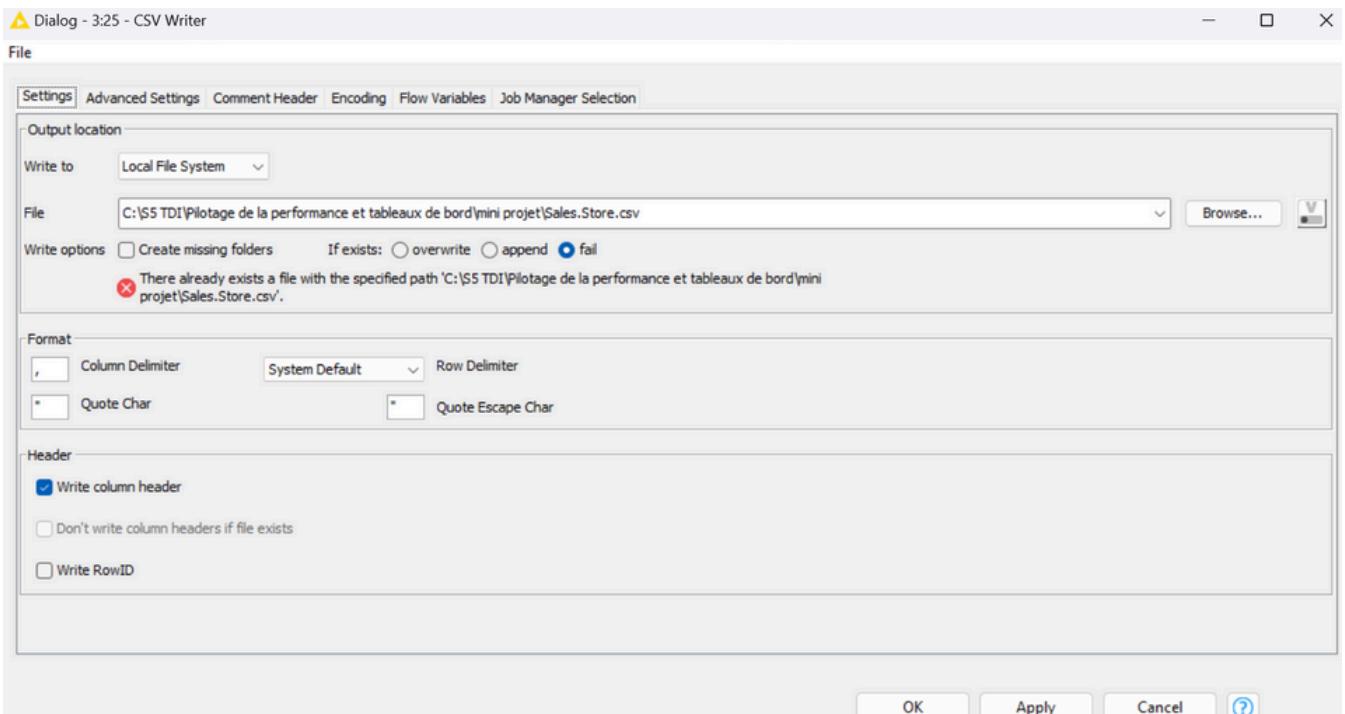
Column Renamer

- Standardisation des noms de colonnes



[. CSV Writer](#)

- Exportation des données nettoyées au format CSV, prêtes pour l'analyse dans Power BI.



Pour Customer [CSV Reader](#)

- Importation des données brutes de la table Customer.

Dialog - 3:27 - CSV Reader

File

Settings Transformation Advanced Settings Limit Rows Encoding Flow Variables Job Manager Selection Memory Policy

Input location

Read from Local File System

Mode File Files in folder

File C:\Users\pc\Downloads\sales.customer.csv

Reader options

Format

Autodetect format

Column delimiter ; Row delimiter Line break Custom \r\n

Quote char " Quote escape char "

Comment char #

Has column header Has RowID

Support short data rows Prepend file index to RowID

Preview

The suggested column types are based on the first 10000 rows only. See 'Advanced Settings' tab.

Row ID	TerritoryID	CustomerID	PersonID	StoreID	AccountNumber	rowguid	ModifiedDate
Row0	1	NULL	934	1	AW00000001	3F5AE95E-B87D-4AED-95B4-C3797AFCB74F	2014-09-12 11:15:07.263
Row1	2	NULL	1028	1	AW00000002	E552F657-A9AF-4A7D-A645-C429D6E02491	2014-09-12 11:15:07.263
Row2	3	NULL	642	4	AW00000003	130774B1-D821-4EF3-98C8-C104BCD6ED6D	2014-09-12 11:15:07.263
Row3	4	NULL	932	4	AW00000004	FF862851-1DAA-4044-BE7C-3E85583C054D	2014-09-12 11:15:07.263

OK Apply Cancel

Table View

- Visualisation des données pour identifier les anomalies : quelques valeurs nulles détectées dans des colonnes .

Dialog - 3:26 - Table View

Table View

Rows: 19820 | Columns: 7

RowID	CustomerID	PersonID	StoreID	TerritoryID	AccountNumber	rowguid
Row0	1	NULL	934	1	AW00000001	3F5AE95E-B87D-4AED-95B4-C3797AFCB74F
Row1	2	NULL	1028	1	AW00000002	E552F657-A9AF-4A7D-A645-C429D6E02491
Row2	3	NULL	642	4	AW00000003	130774B1-D821-4EF3-98C8-C104BCD6ED6D
Row3	4	NULL	932	4	AW00000004	FF862851-1DAA-4044-BE7C-3E85583C054D
Row4	5	NULL	1026	4	AW00000005	83905BD
Row5	6	NULL	644	4	AW00000006	1A92DF8
Row6	7	NULL	930	1	AW00000007	03E9273E
Row7	8	NULL	1024	5	AW00000008	801368B
Row8	9	NULL	620	5	AW00000009	B900BB7
Row9	10	NULL	928	6	AW00000010	CDB6698
Row...	11	NULL	1022	6	AW00000011	750F349F
Row...	12	NULL	622	6	AW00000012	947BCAF
Row...	13	NULL	434	7	AW00000013	B0FA585
Row...	14	NULL	1020	8	AW00000014	2F96BED

Data

Displayed columns

Manual Wildcard Regex Type

Search Aa

Excludes Includes

No columns in this list

Includes

CustomerID PersonID StoreID TerritoryID AccountNumber rowguid Any unknown column

Show row numbers

Show RowIDs

View

Title Table View

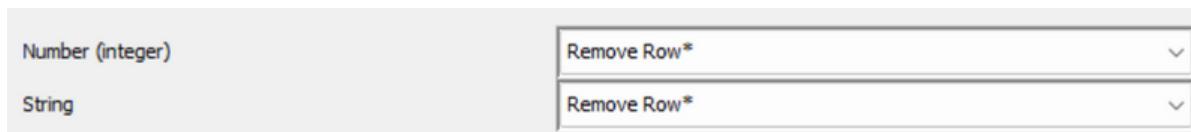
Show table size

Show column data types in header

Cancel Ok

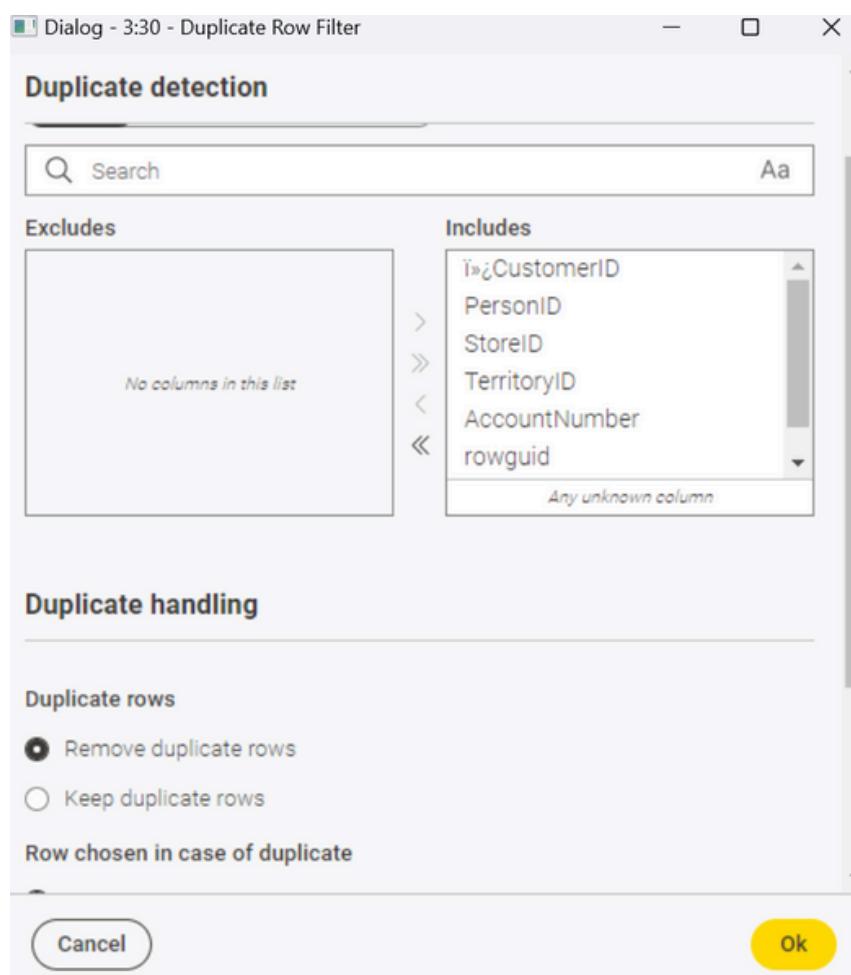
Missing Value

- Traitement des valeurs nulles :
 - Remplacement des champs textuels manquants par une valeur par défaut.



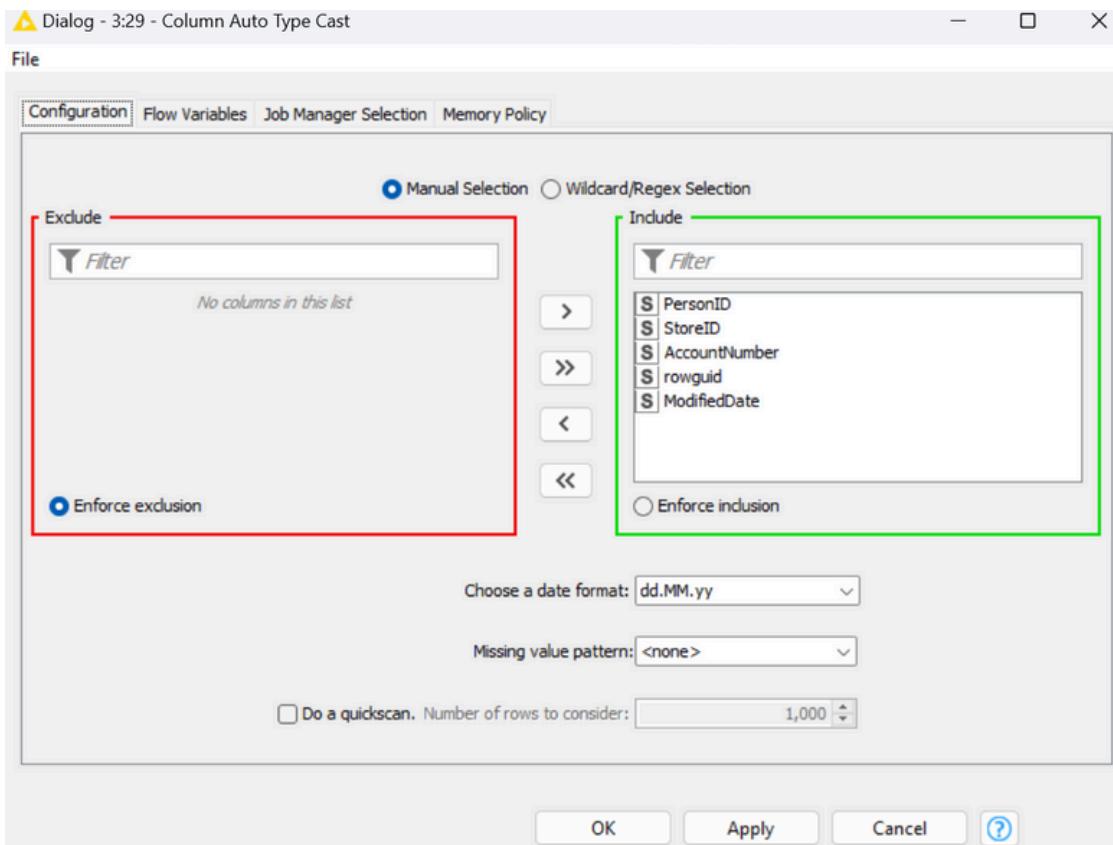
Duplicate Row Filter

- Suppression des lignes dupliquées pour garantir l'unicité des enregistrements.



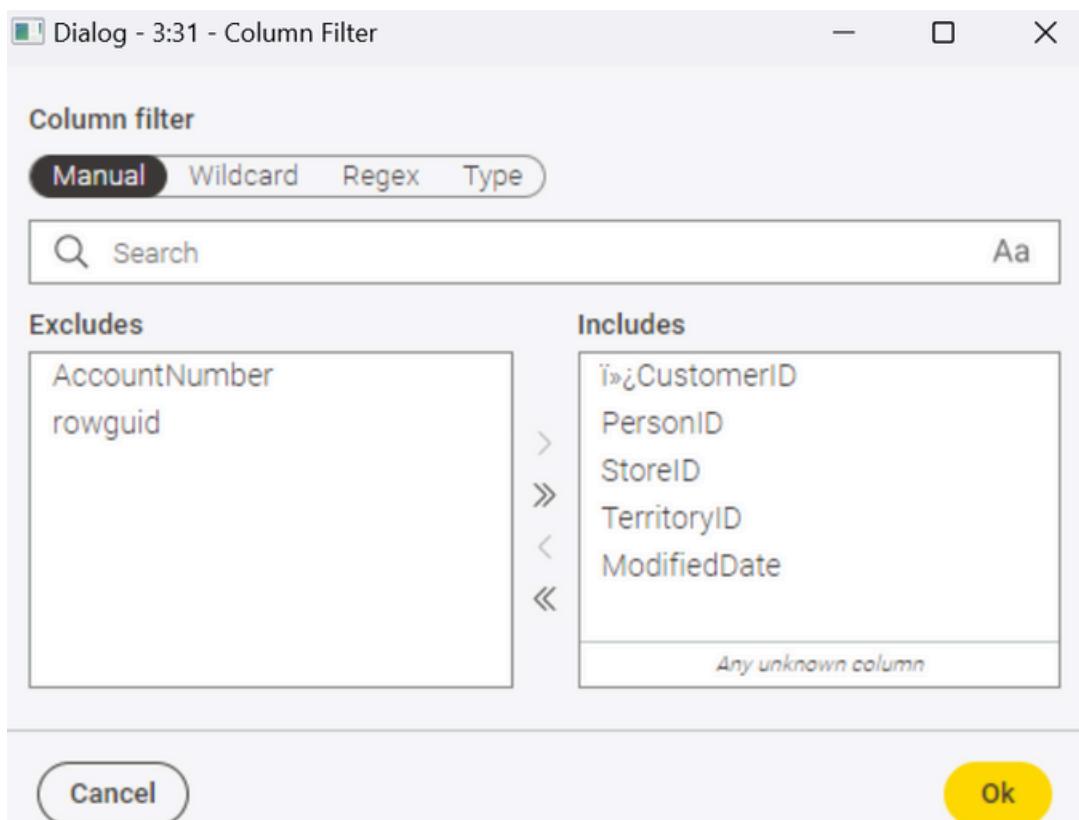
Column Auto Type Cast

- Ajustement automatique des types :
 - Conversion des identifiants en type entier.
 - Correction des formats de date dans ModifiedDate.



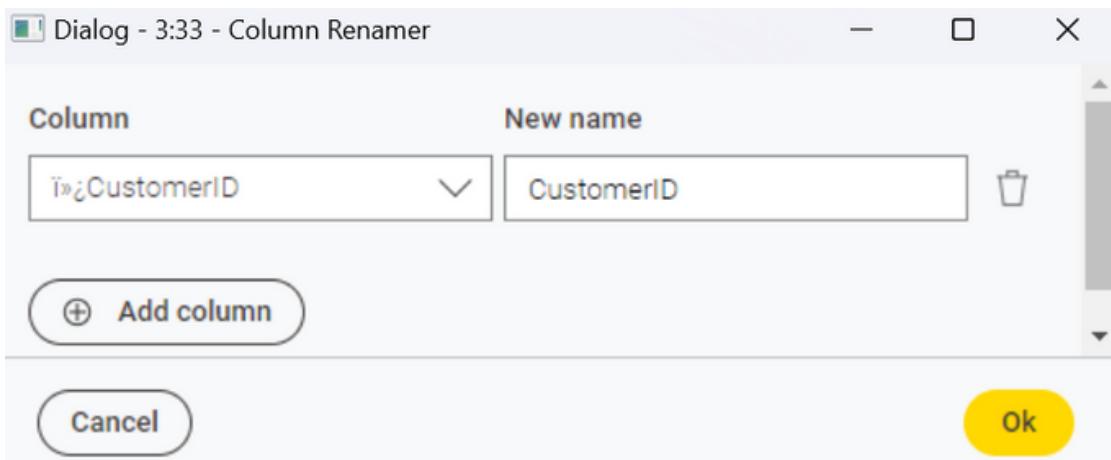
Column Filter

- Suppression des colonnes inutiles comme les métadonnées techniques non nécessaires à l'analyse.



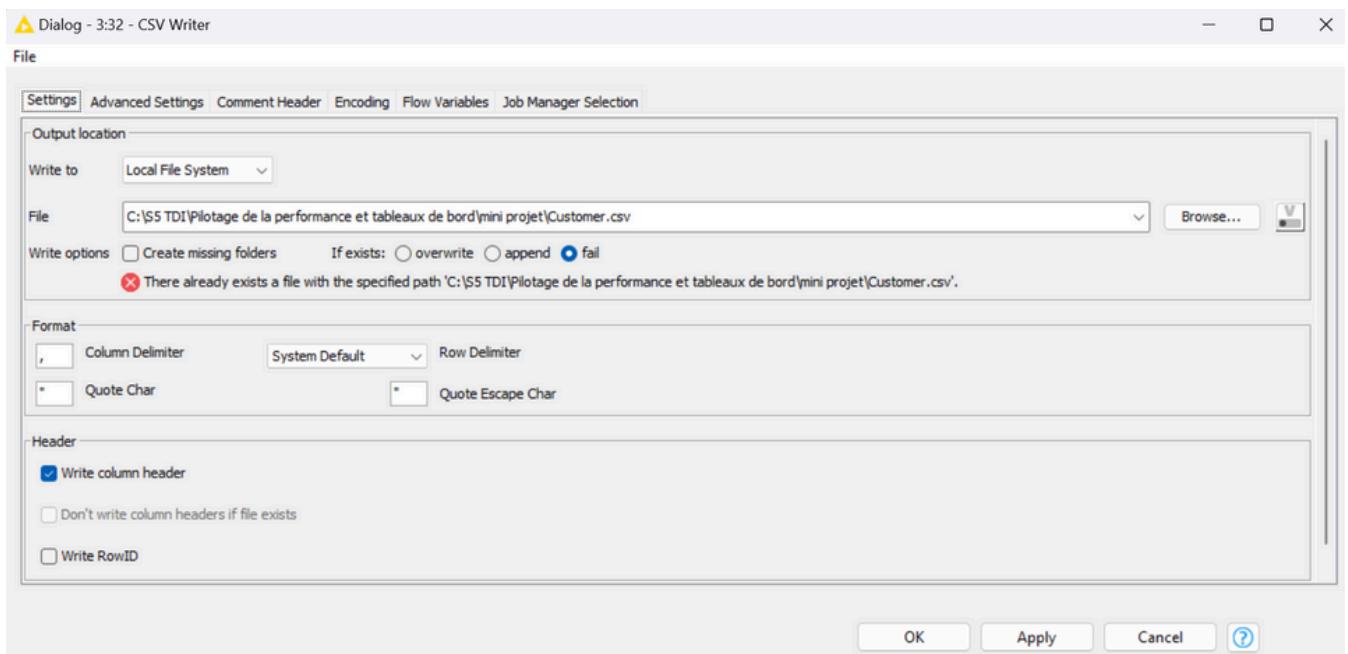
Column Renamer

- Standardisation des noms de colonnes



. CSV Writer

- Exportation des données nettoyées au format CSV, prêtes pour l'analyse dans Power BI.



Pour SalesPerson : CSV Reader

- Importation des données brutes de la table SalesPerson.

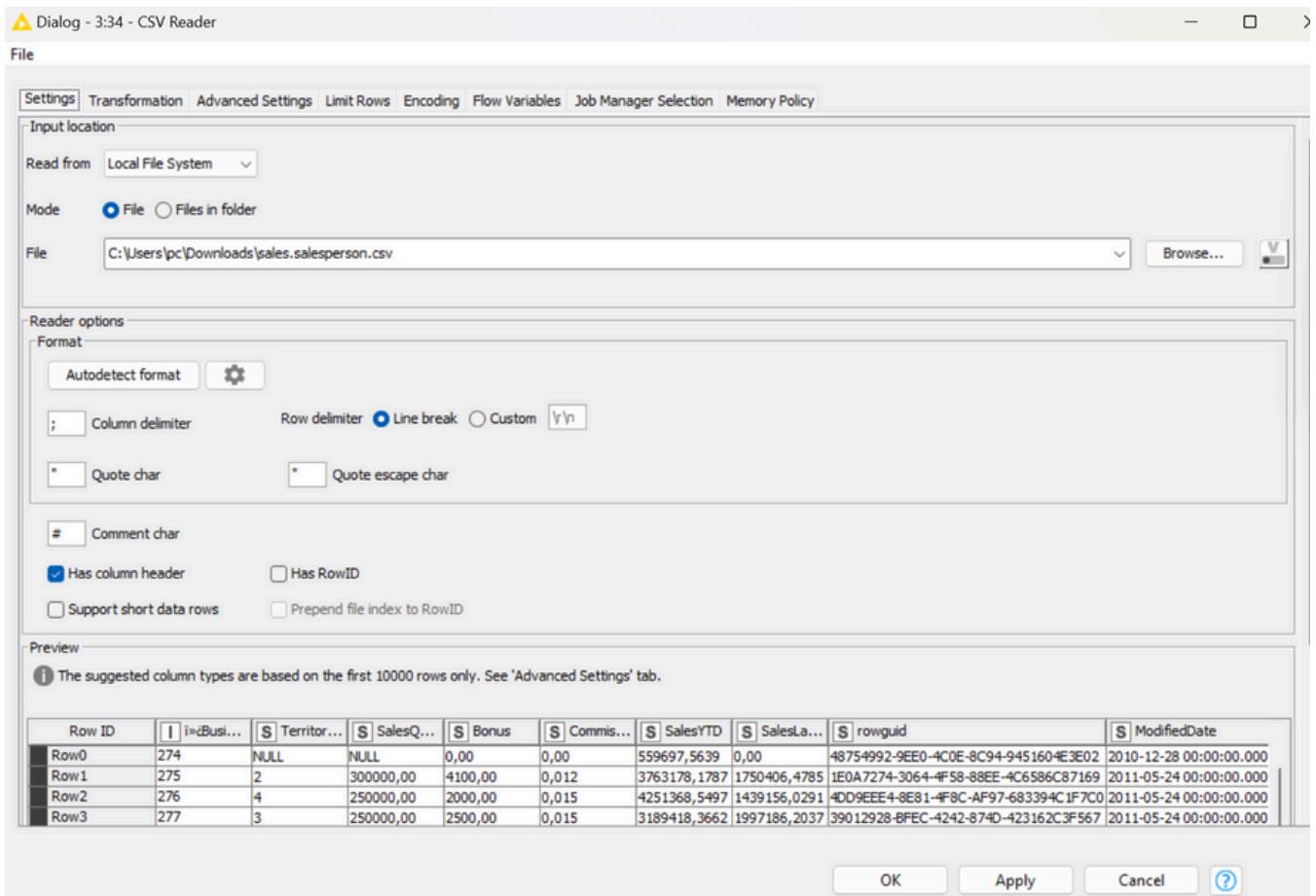


Table View

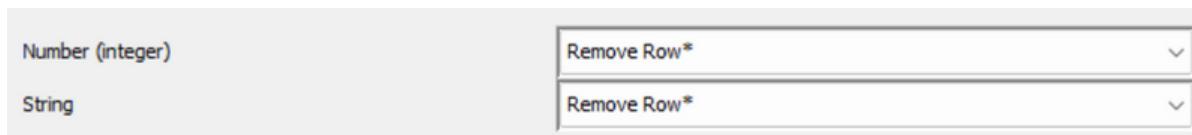
- Visualisation des données pour identifier les anomalies : quelques valeurs nulles détectées dans des colonnes .

The screenshot shows the 'Table View' dialog box with the following configuration:

- Data:**
 - Displayed columns: Manual
 - Search bar: Aa
 - Excludes: No columns in this list
 - Includes:
 - IndBusinessNumber (integer)
 - TerritoryID
 - SalesQuota
 - Bonus
 - CommissionRate
 - SalesYTD
 - Any unknown column
 - Show row numbers: unchecked
 - Show RowIDs: checked
- View:**
 - Title: Table View
 - Show table size: checked
 - Show column data types in header: checked

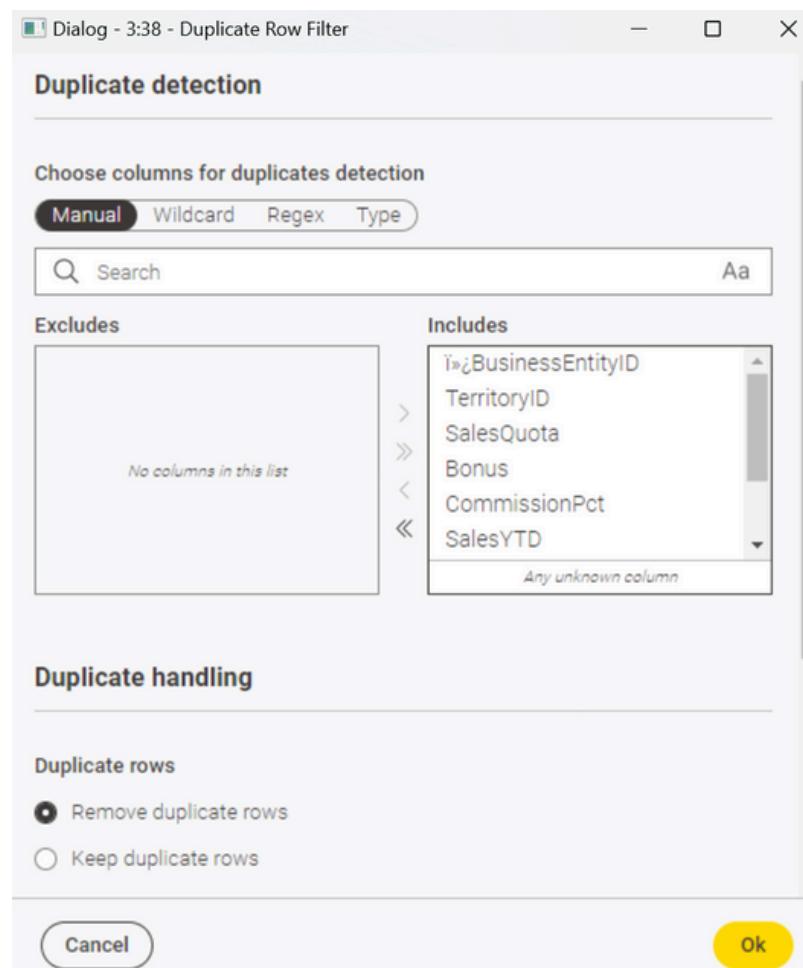
Missing Value

- Traitement des valeurs nulles :
 - Remplacement des champs textuels manquants par une valeur par défaut.



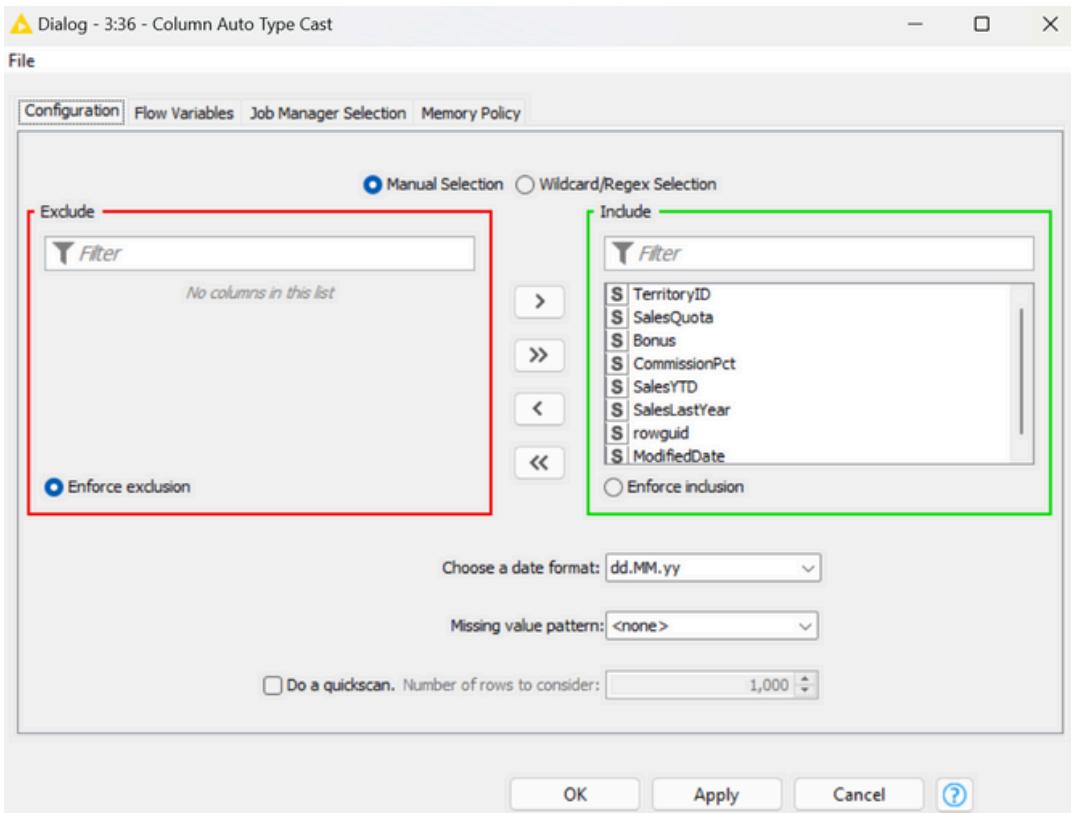
Duplicate Row Filter

- Suppression des lignes dupliquées pour garantir l'unicité des enregistrements.



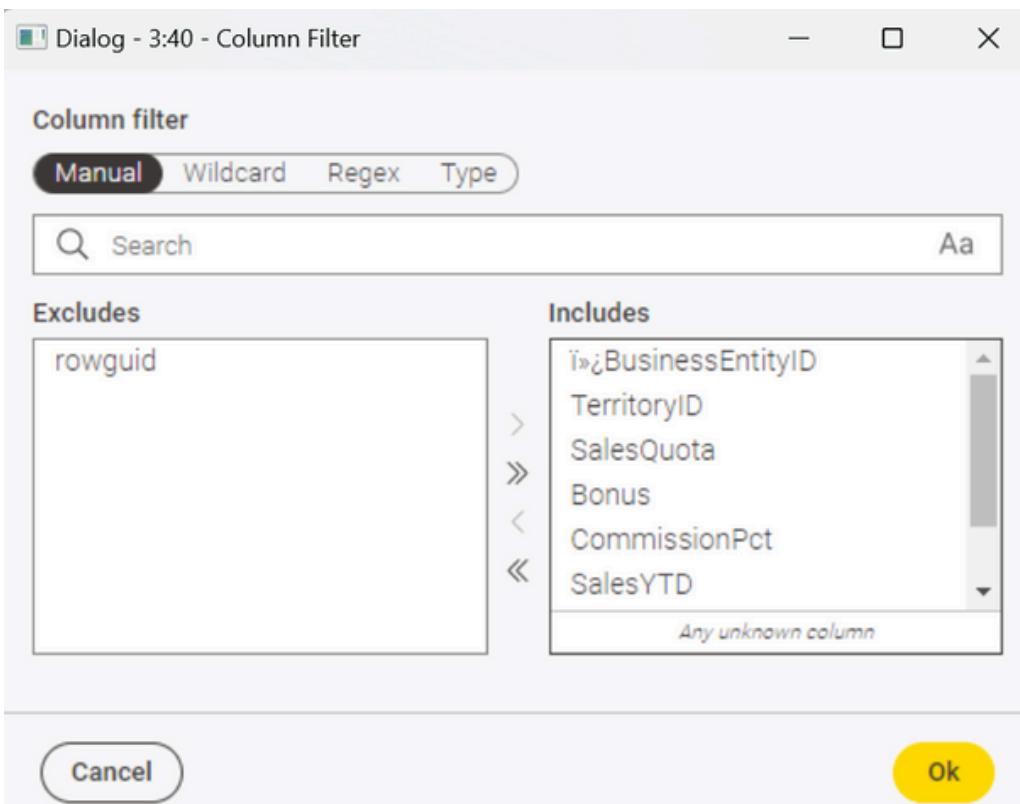
Column Auto Type Cast

- Ajustement automatique des types :
 - Conversion des identifiants en type entier.
 - Correction des formats de date dans ModifiedDate.



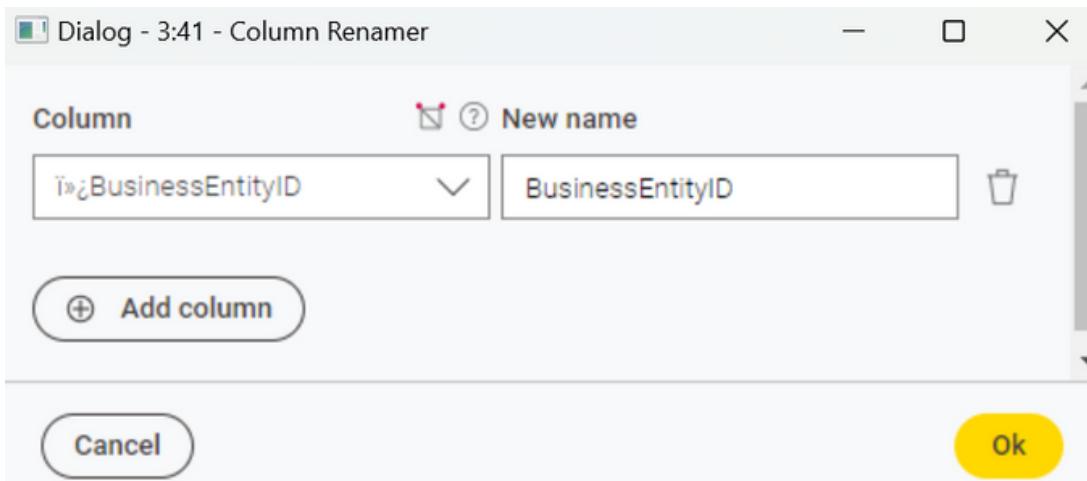
Column Filter

- Suppression des colonnes inutiles comme les métadonnées techniques non nécessaires à l'analyse.



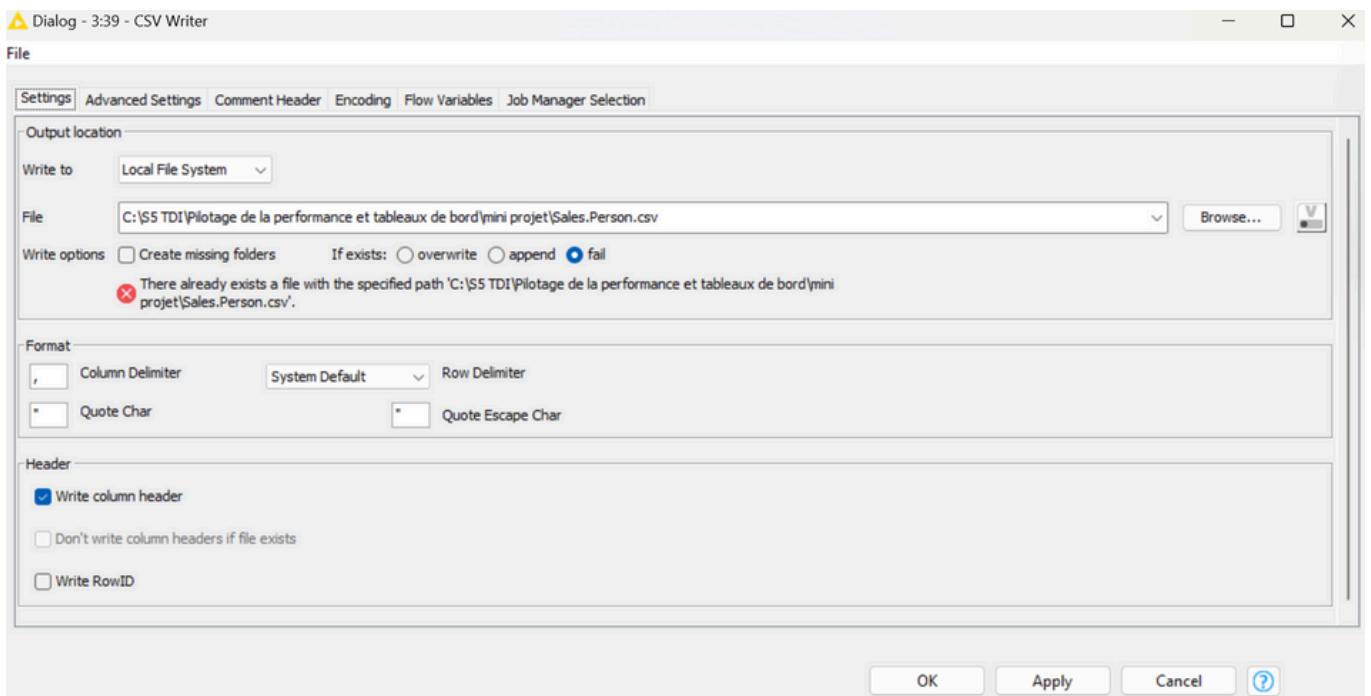
Column Renamer

- Standardisation des noms de colonnes



[. CSV Writer](#)

- Exportation des données nettoyées au format CSV, prêtes pour l'analyse dans Power BI.



Pour SalesReason :

[CSV Reader](#)

- Importation des données brutes de la table SalesReason.

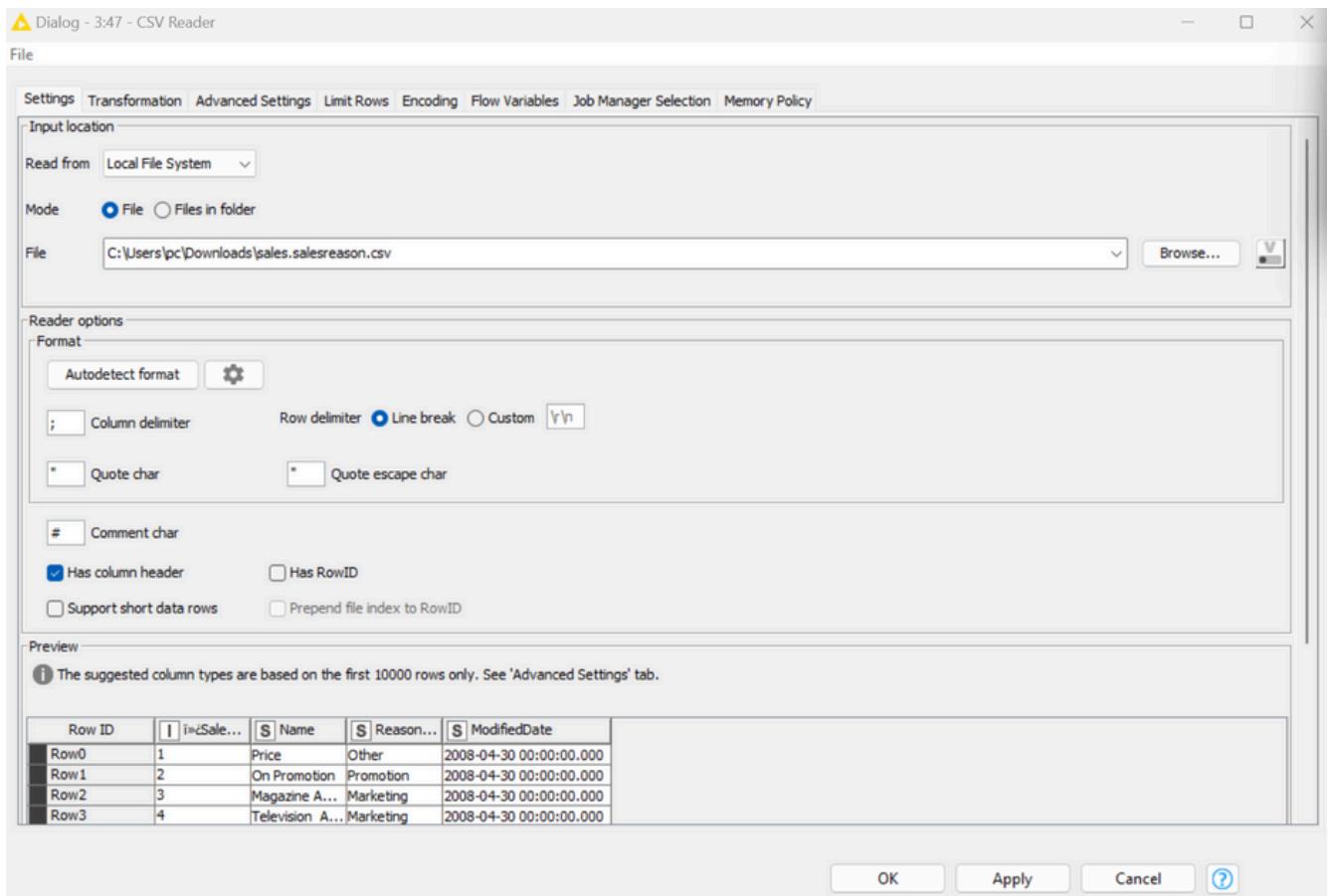
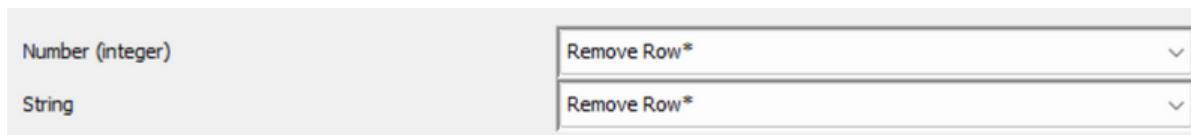


Table View

- Visualisation des données pour identifier les anomalies : quelques valeurs nulles détectées dans des colonnes .

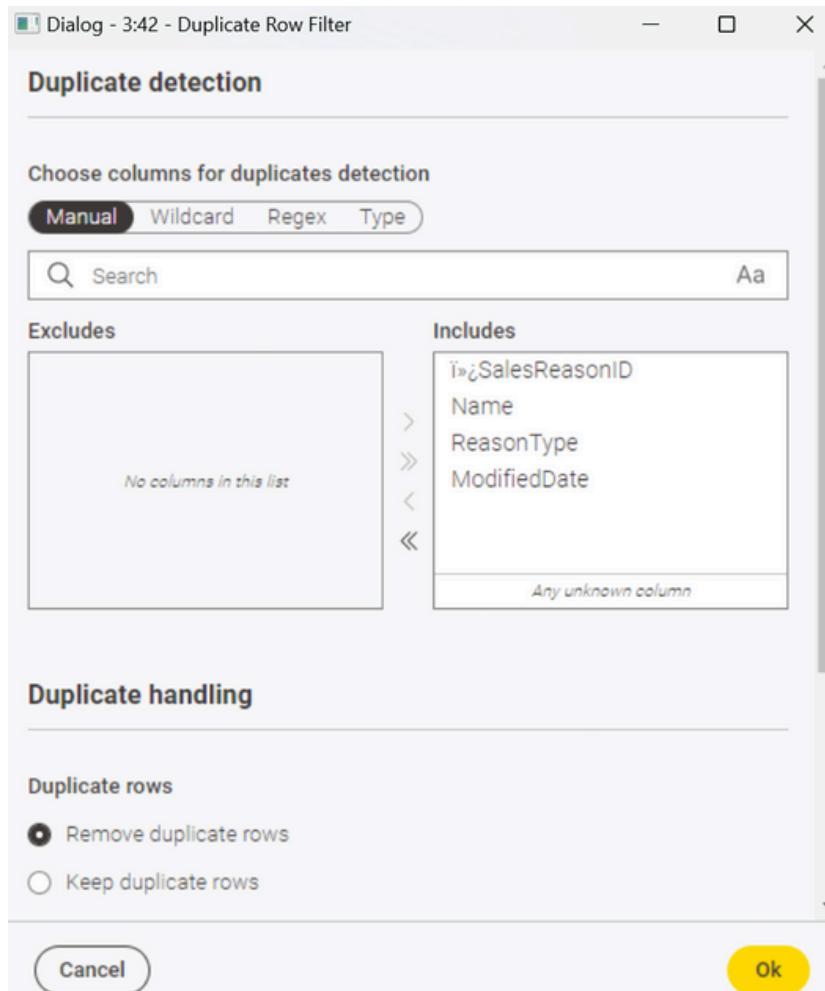
Missing Value

- Traitement des valeurs nulles :
 - Remplacement des champs textuels manquants par une valeur par défaut.



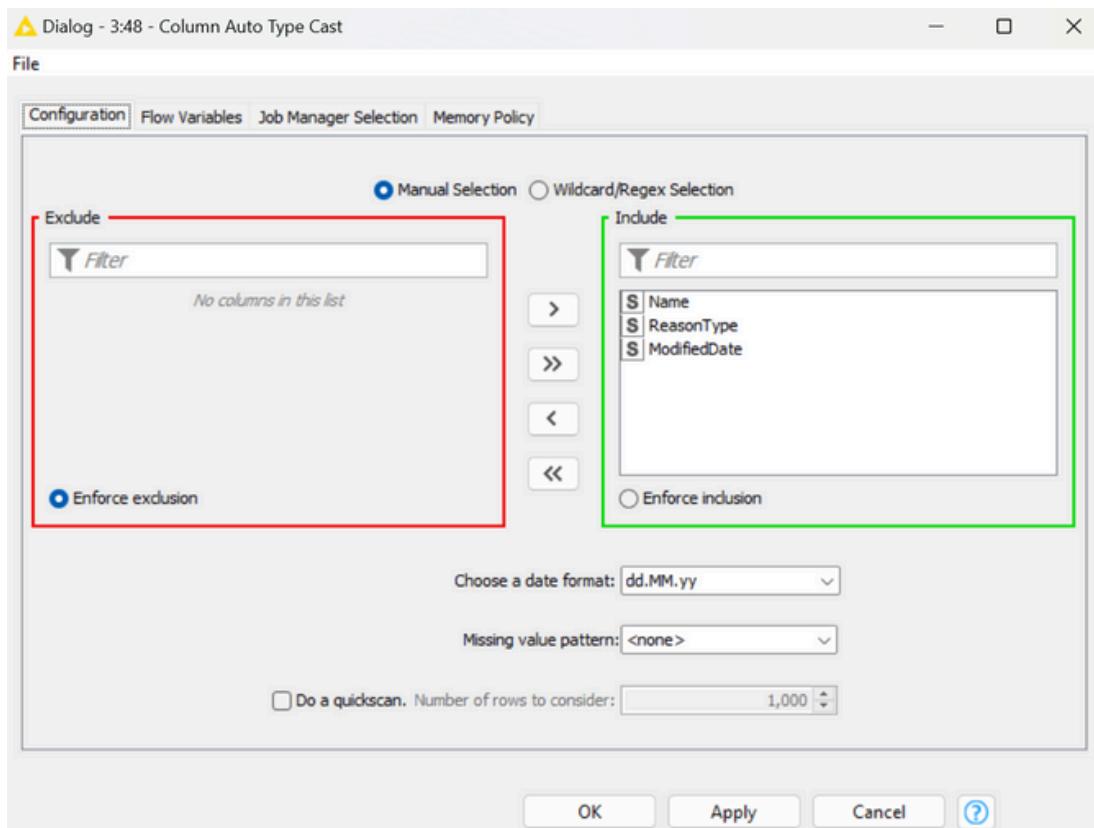
Duplicate Row Filter

- Suppression des lignes dupliquées pour garantir l'unicité des enregistrements.



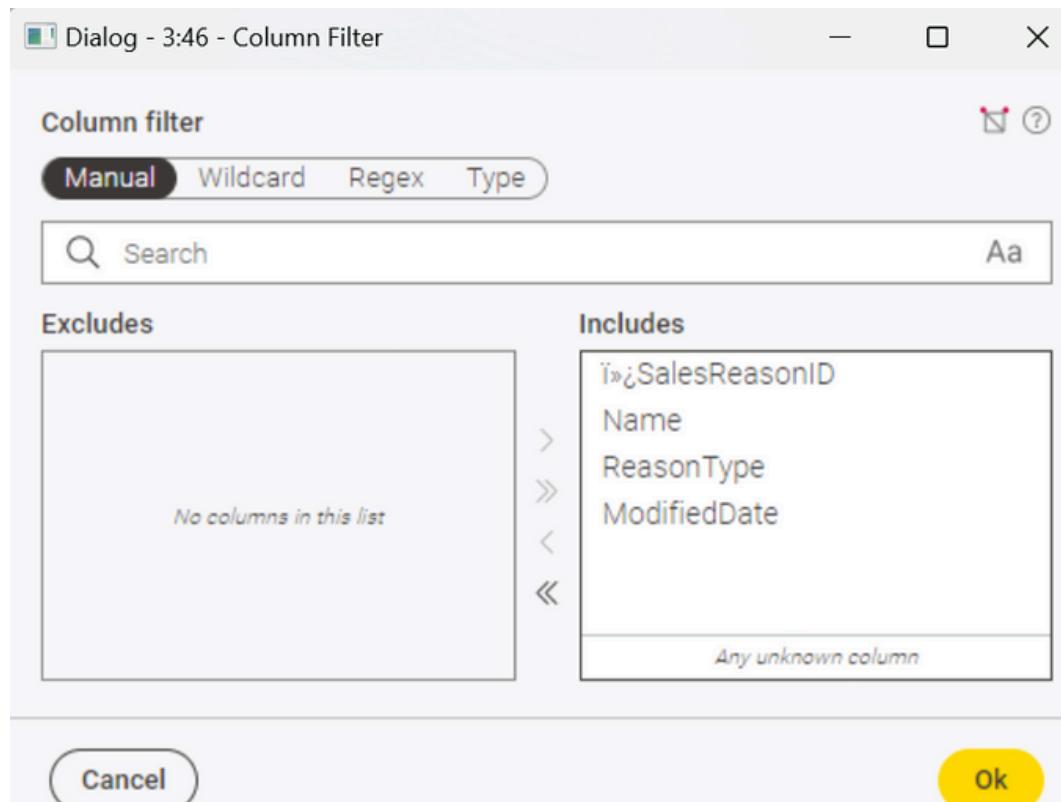
Column Auto Type Cast

- Ajustement automatique des types :
 - Conversion des identifiants en type entier.
 - Correction des formats de date dans ModifiedDate.



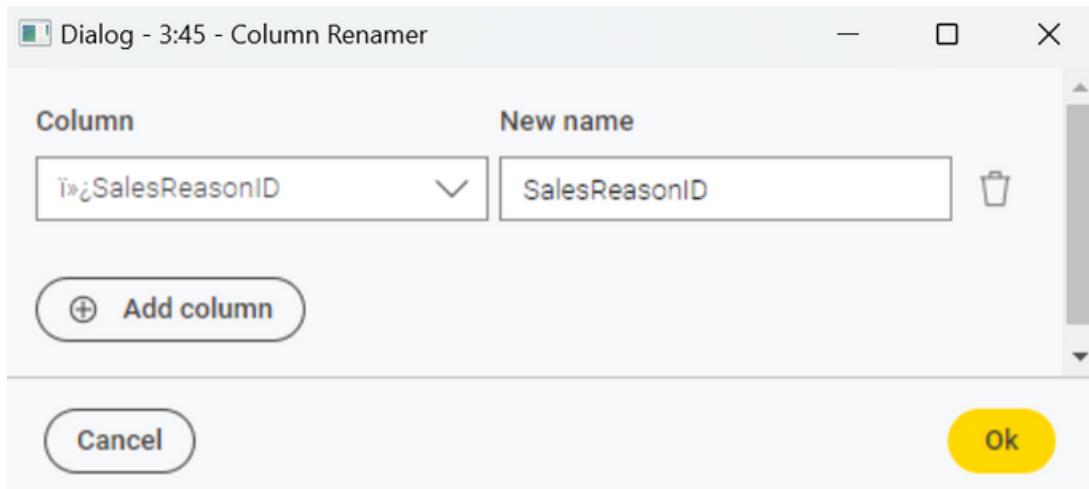
Column Filter

- Suppression des colonnes inutiles comme les métadonnées techniques non nécessaires à l'analyse.



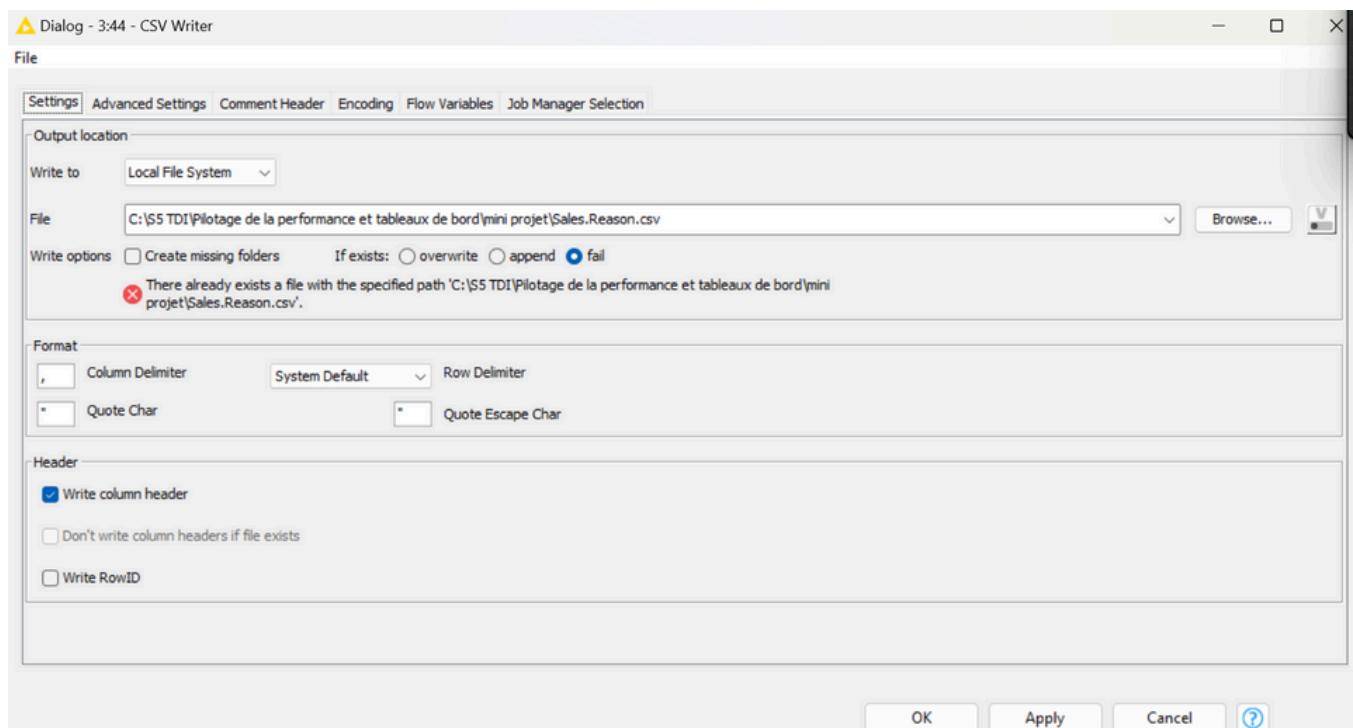
Column Renamer

- Standardisation des noms de colonnes



. CSV Writer

- Exportation des données nettoyées au format CSV, prêtes pour l'analyse dans Power BI.

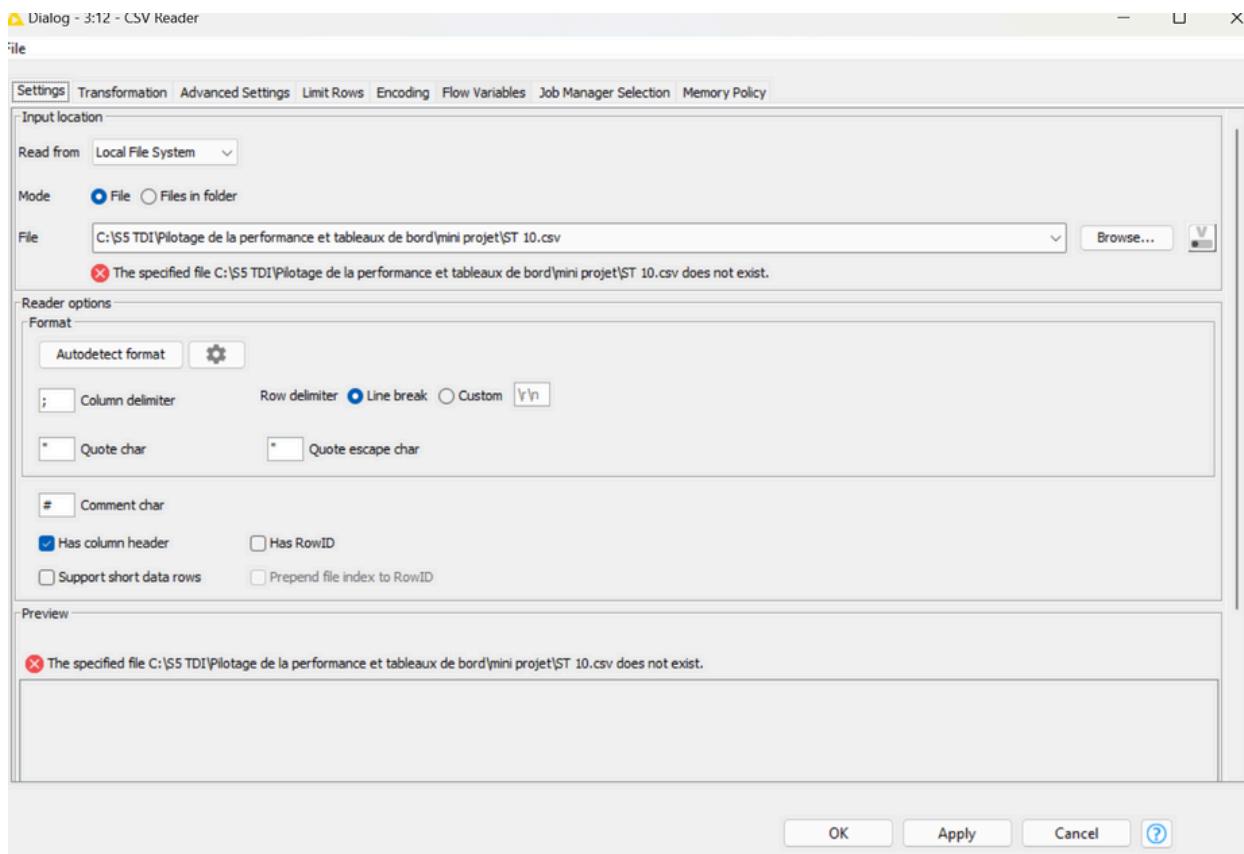


Pour SalesTerritory :



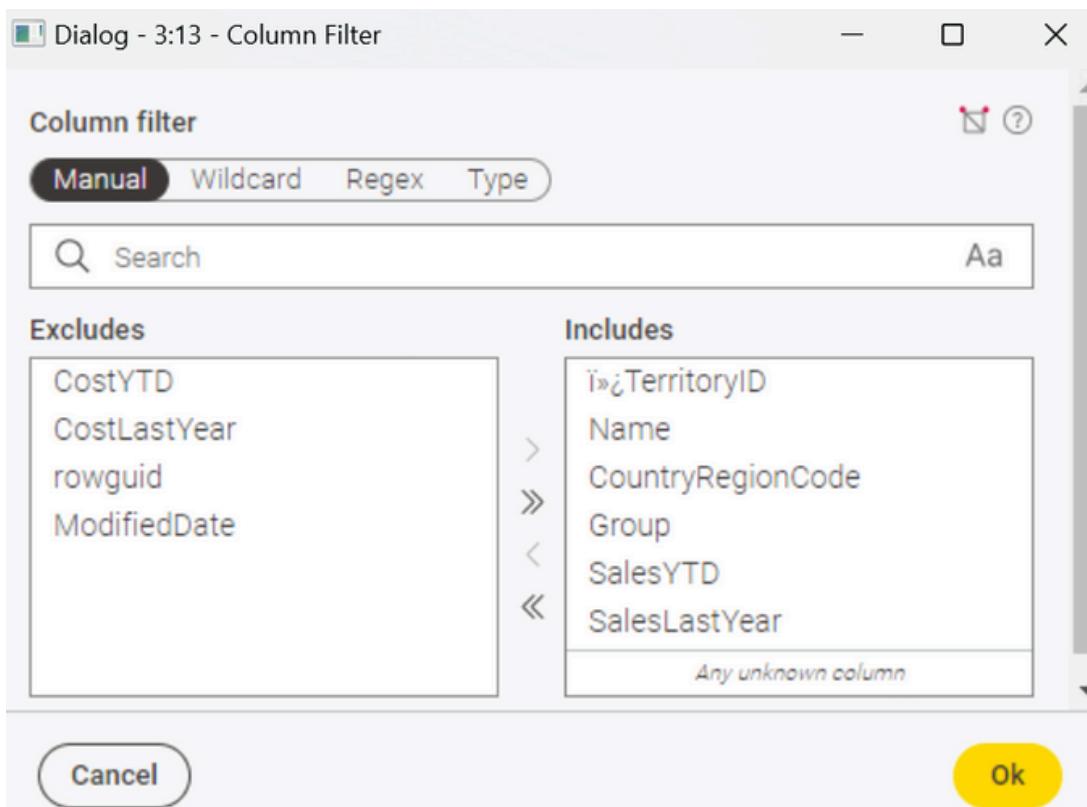
CSV Reader

- Importation des données brutes de la table SalesReason.



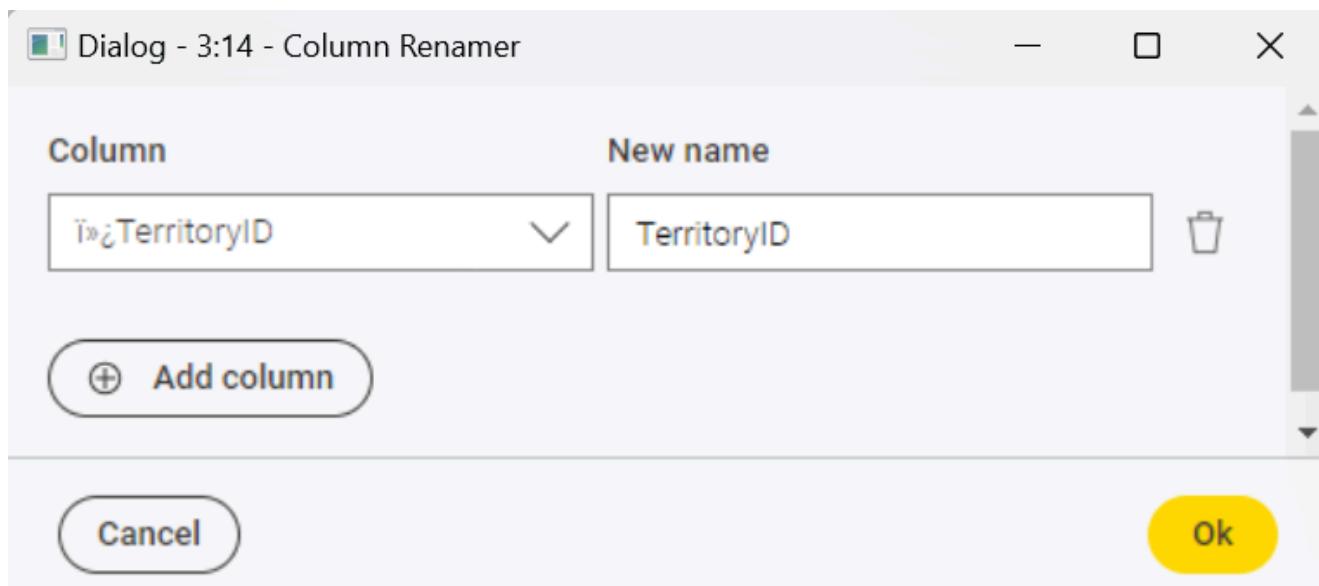
Column Filter

- Suppression des colonnes inutiles comme les métadonnées techniques non nécessaires à l'analyse.



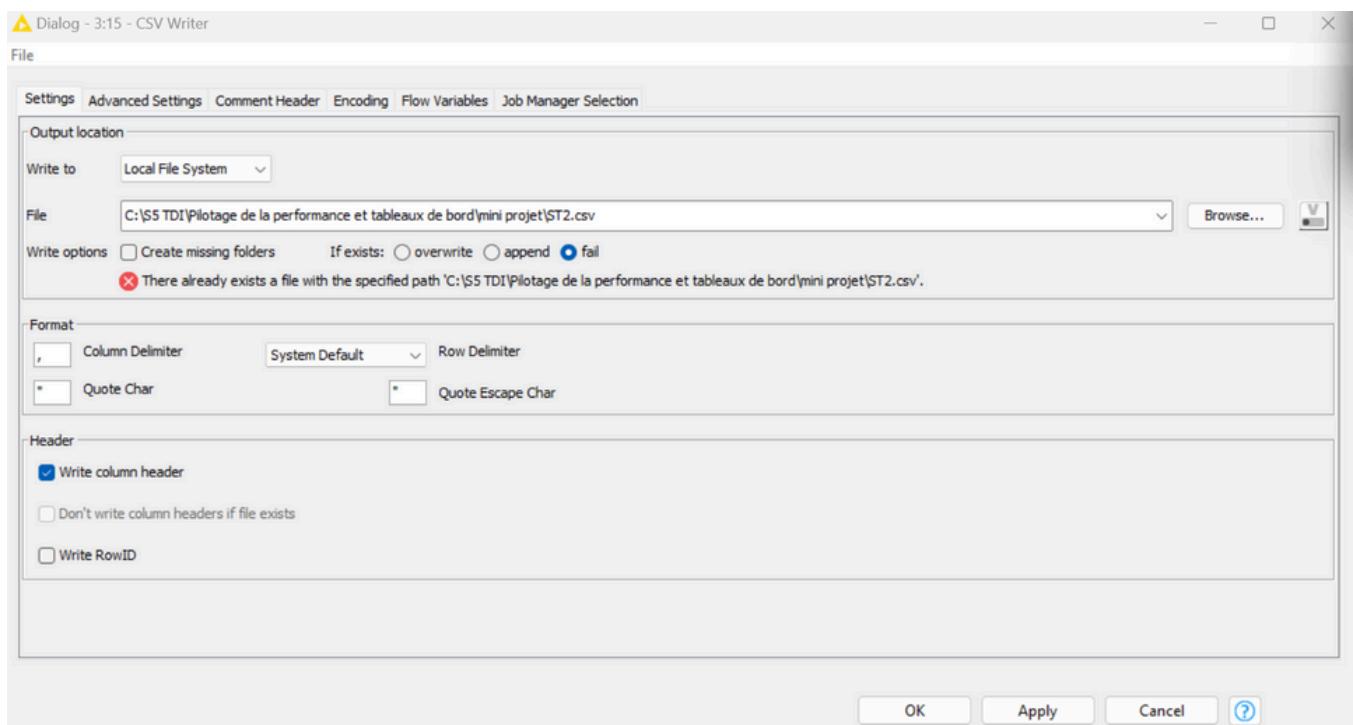
Column Renamer

- Standardisation des noms de colonnes



. CSV Writer

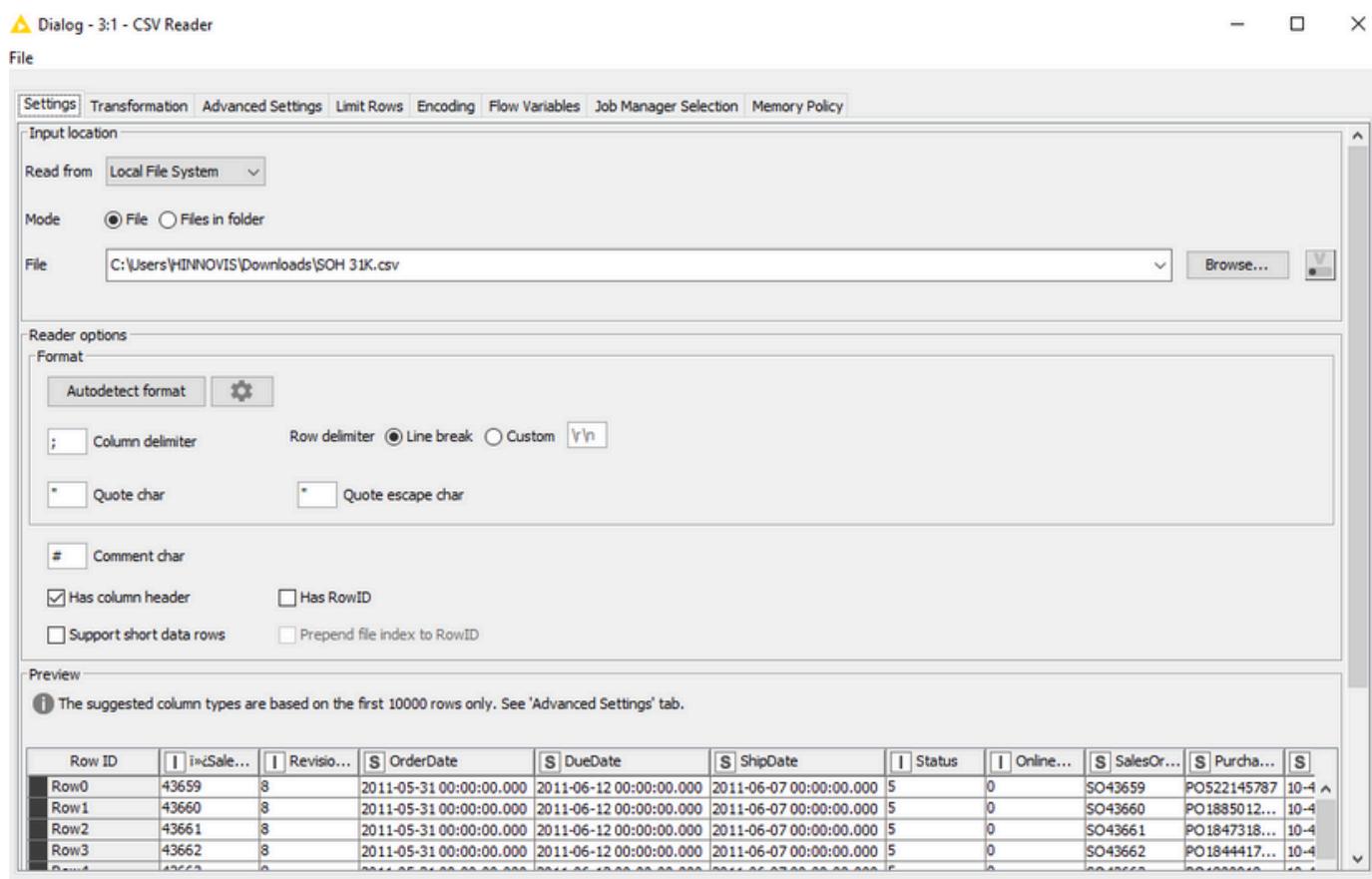
- Exportation des données nettoyées au format CSV, prêtes pour l'analyse dans Power BI.



Pour SalesOrderheader :

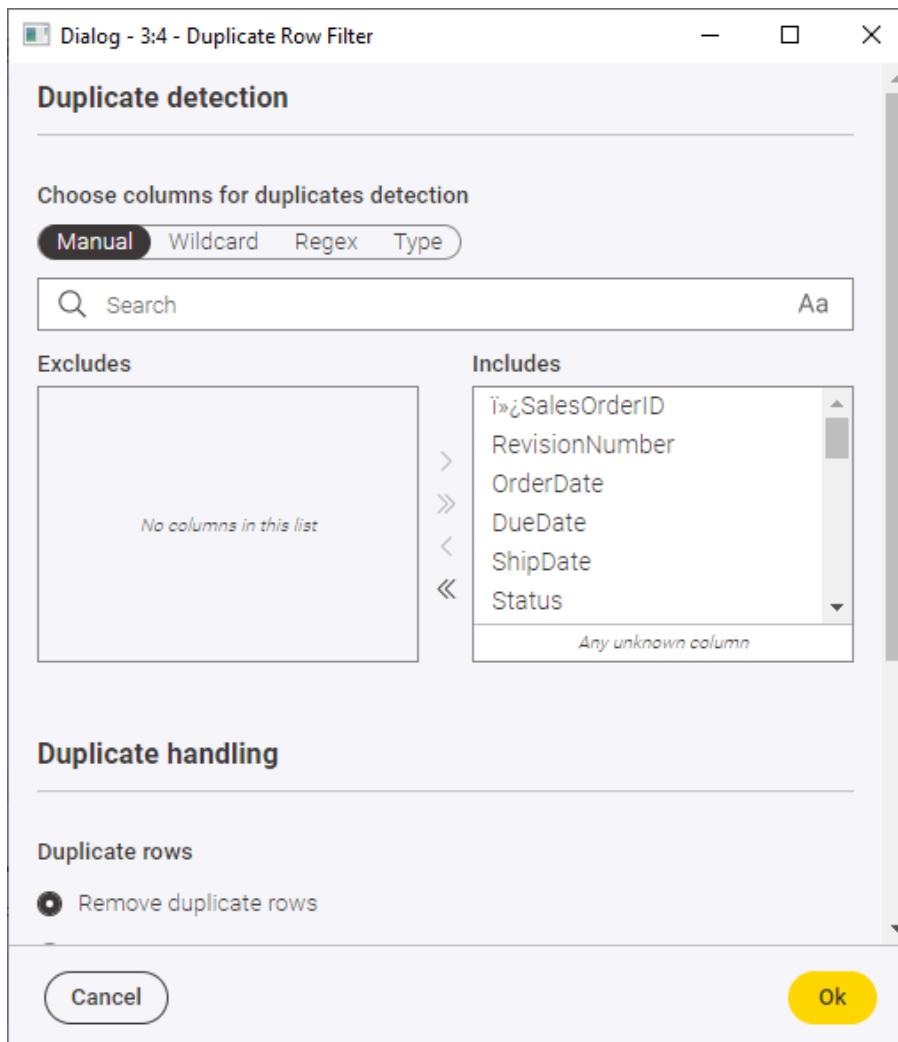
- **Lecture des Données (CSV Reader):**

Ce nœud permet de lire un fichier CSV contenant les données brutes. C'est une étape essentielle, car elle constitue le point de départ du workflow, permettant d'importer les données pour les étapes suivantes de nettoyage et de transformation.



- **Duplicate Row Filter :**

Ce nœud est utilisé pour identifier et supprimer les lignes en double dans le fichier. En éliminant les doublons, il garantit que chaque enregistrement est unique, ce qui est important pour éviter les biais ou erreurs dans les analyses futures.



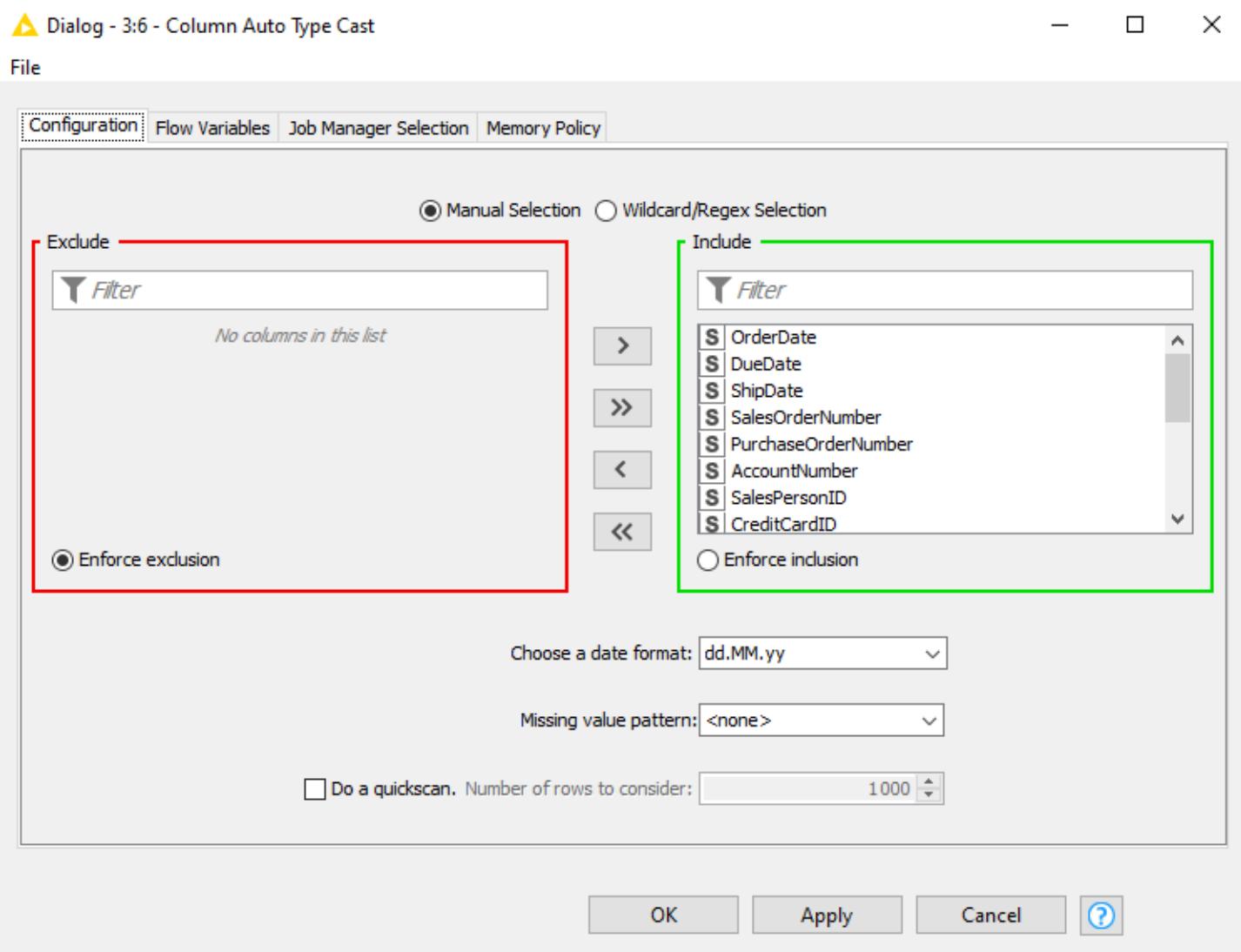
- **Gestion des Valeurs Manquantes (Missing Value) :**

Ce nœud permet de traiter les valeurs manquantes dans les données. Il peut soit remplir les valeurs manquantes avec des données calculées ou par défaut, soit supprimer les lignes ou colonnes affectées. Cette étape est cruciale pour assurer la complétude des données et éviter les erreurs ou biais qui pourraient influencer les résultats de l'analyse.

The screenshot shows the configuration of the 'Missing Value Imputation' node. For 'Number (integer)', the imputation method is set to 'Median'. For 'String', the imputation method is set to 'Most Frequent Value'.

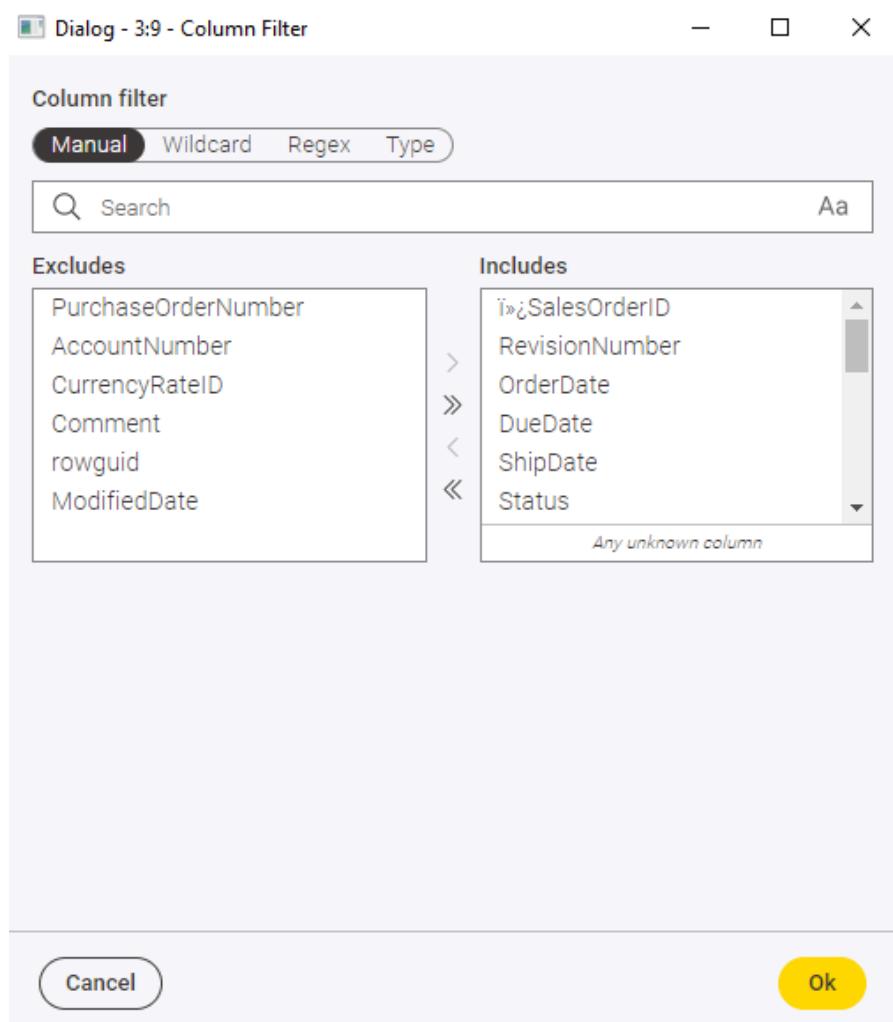
- **Column Auto Type Cast :**

Ce nœud harmonise automatiquement les types de données des colonnes (par exemple, texte, nombres, ou dates) pour qu'ils soient cohérents avec leur contenu. Cela évite les erreurs de typage qui pourraient rendre certaines opérations ou analyses impossibles.



- **Column Filter :**

Cette étape permet de sélectionner ou supprimer certaines colonnes spécifiques dans les données. Cela permet de garder uniquement les informations pertinentes pour l'analyse et d'éliminer les colonnes inutiles, rendant ainsi le jeu de données plus léger et structuré.



- **Column Renamer :**

Ce nœud est utilisé pour renommer les colonnes afin de les rendre plus compréhensibles ou conformes à un standard. Des noms de colonnes clairs et adaptés réduisent le risque d'erreurs dans les étapes suivantes et facilitent la lecture et l'utilisation des données par les utilisateurs.

The screenshot shows a 'Column Renamer' dialog box. It has two main sections: 'Column' and 'New name'. In the 'Column' section, there is a dropdown menu containing the text 'SalesOrderID'. To its right is a dropdown arrow. In the 'New name' section, the text 'Order ID' is entered into a text input field. To the right of this field is a small trash can icon for deleting the new name. Below these sections is a button labeled '+ Add column'.

- CSV Writer :

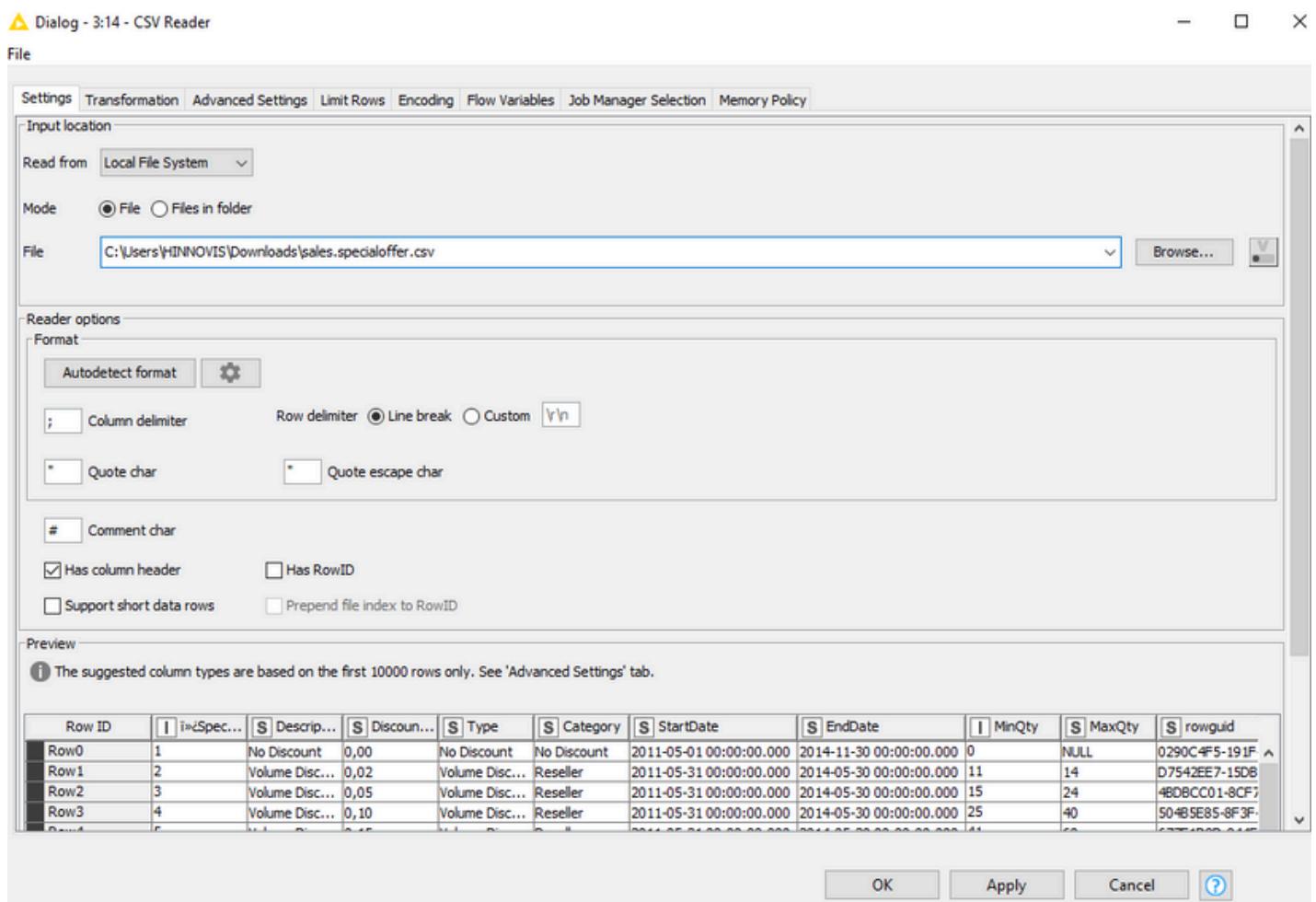
Ce nœud final exporte les données nettoyées et préparées dans un nouveau fichier CSV. Après transformation, ce fichier final peut être utilisé pour d'autres analyses ou partagé avec d'autres outils ou utilisateurs. Cette étape clôture le workflow en produisant un jeu de données fiable et organisé.

The screenshot shows the 'CSV Writer' dialog box with the 'Settings' tab selected. The 'Output location' section indicates 'Write to Local File System' and specifies the file path as 'C:\Users\INNOVIS\Desktop\knime\data1.csv'. A warning message at the bottom of this section states: 'There already exists a file with the specified path 'C:\Users\INNOVIS\Desktop\knime\data1.csv''. The 'Format' section includes options for 'Column Delimiter' (set to comma), 'Row Delimiter' (set to 'System Default'), 'Quote Char' (set to double quote), and 'Quote Escape Char'. The 'Header' section contains three checked checkboxes: 'Write column header', 'Don't write column headers if file exists', and 'Write RowID'. At the bottom of the dialog are buttons for 'OK', 'Apply', 'Cancel', and a question mark icon.

Pour specialoffer :

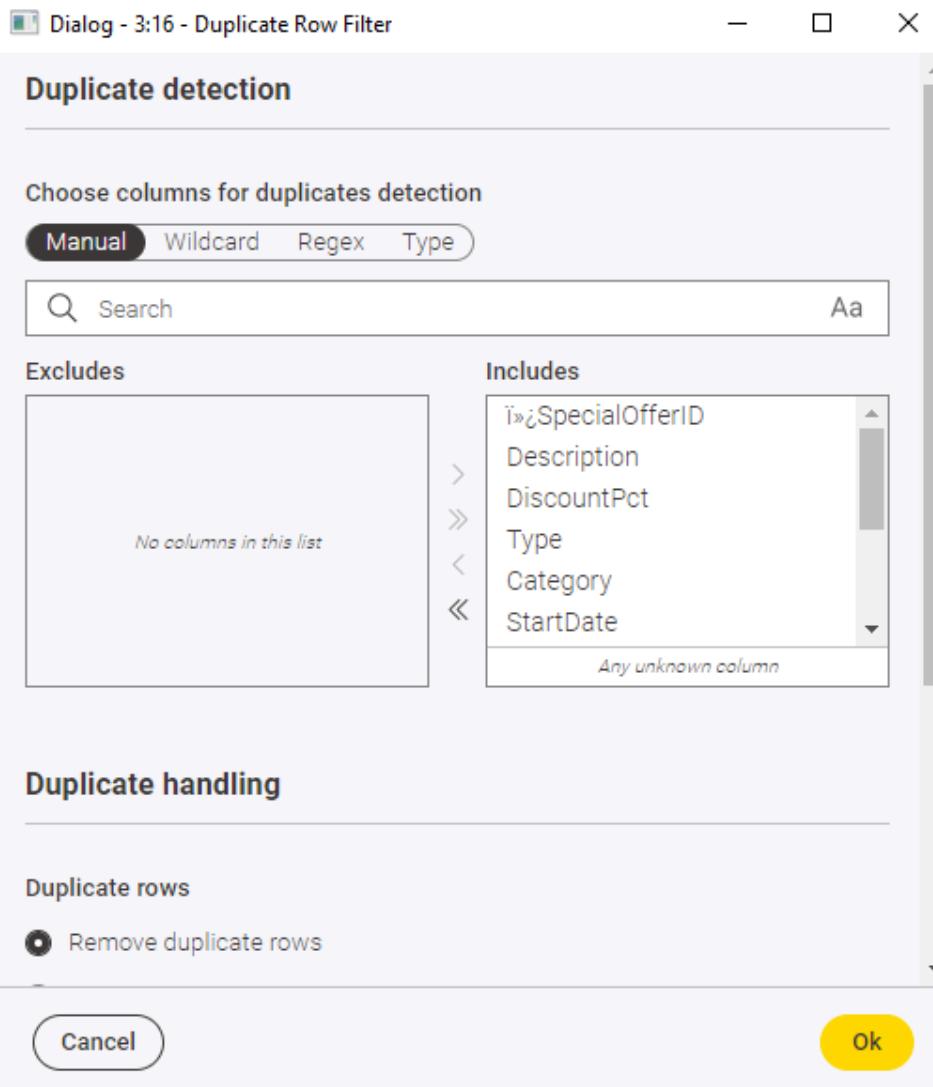
- **Lecture des Données (CSV Reader):**

Ce nœud permet de lire un fichier CSV contenant les données brutes. C'est une étape essentielle, car elle constitue le point de départ du workflow, permettant d'importer les données pour les étapes suivantes de nettoyage et de transformation.



- **Duplicate Row Filter :**

Ce nœud est utilisé pour identifier et supprimer les lignes en double dans le fichier. En éliminant les doublons, il garantit que chaque enregistrement est unique, ce qui est important pour éviter les biais ou erreurs dans les analyses futures.



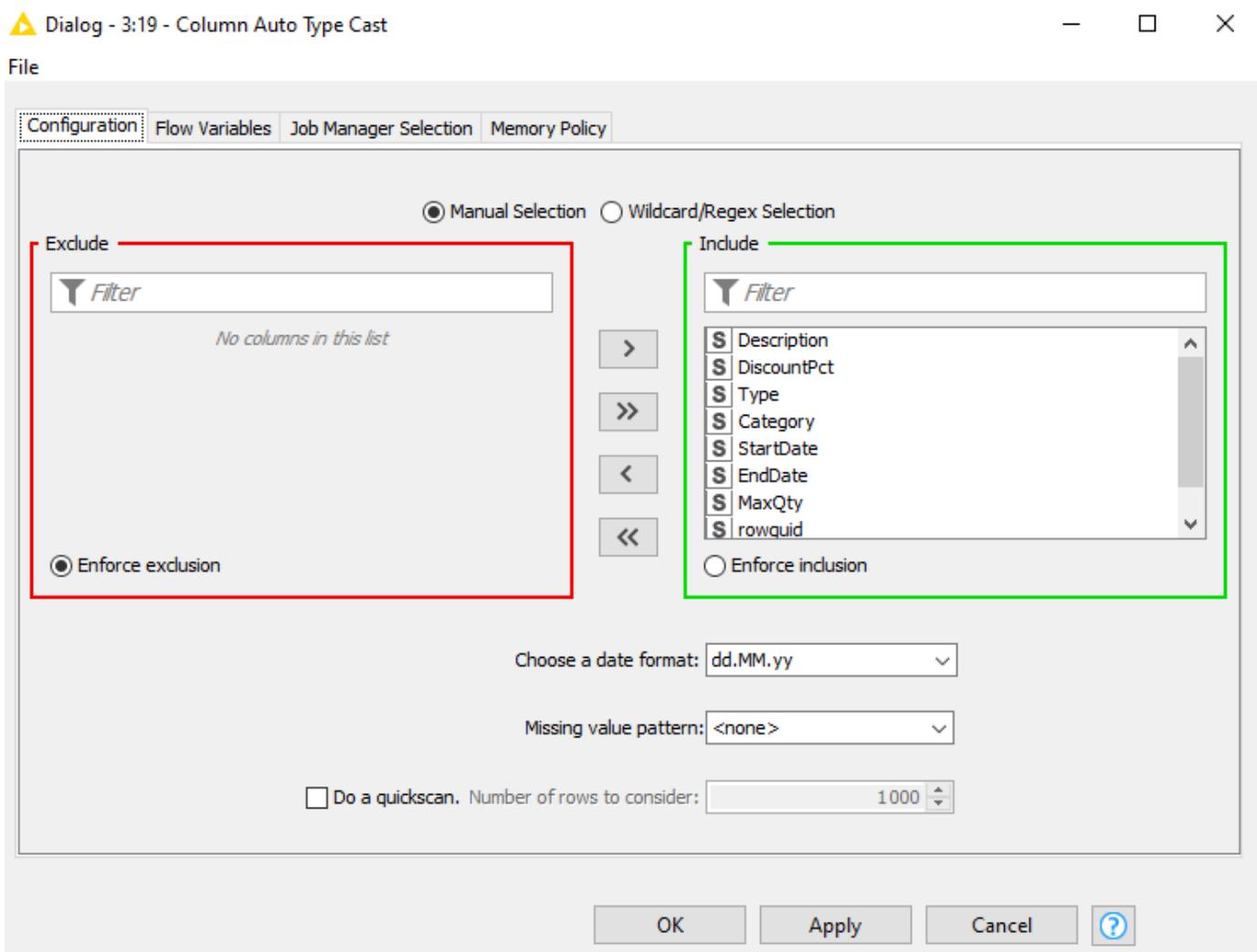
- **Gestion des Valeurs Manquantes (Missing Value) :**

Ce nœud permet de traiter les valeurs manquantes dans les données. Il peut soit remplir les valeurs manquantes avec des données calculées ou par défaut, soit supprimer les lignes ou colonnes affectées. Cette étape est cruciale pour assurer la complétude des données et éviter les erreurs ou biais qui pourraient influencer les résultats de l'analyse.

This screenshot shows the 'Missing Value Imputation' dialog box. On the left, there are two sections: 'Number (integer)' and 'String'. To the right, there are two dropdown menus. The first dropdown for 'Number (integer)' has 'Median' selected. The second dropdown for 'String' has 'Most Frequent Value' selected.

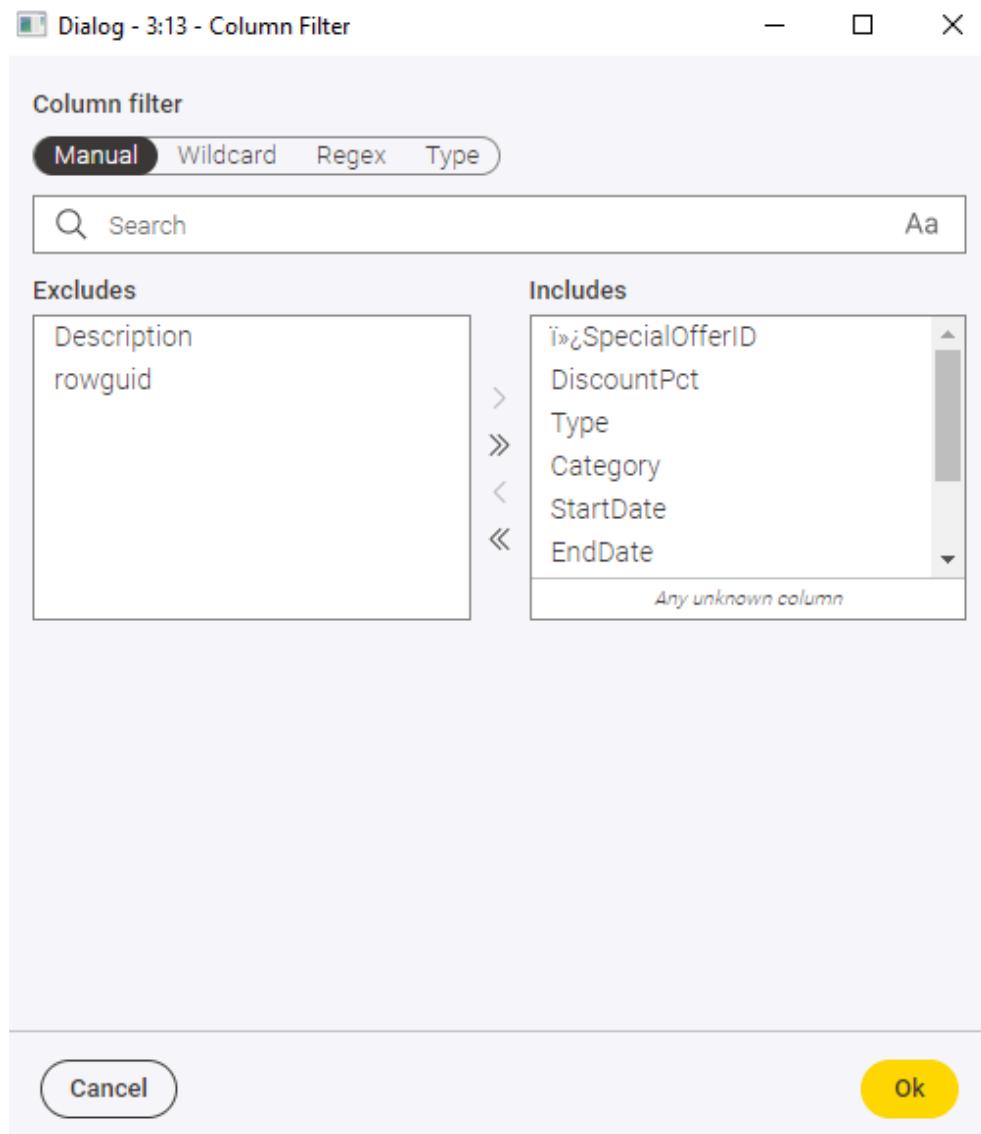
- **Column Auto Type Cast :**

Ce nœud harmonise automatiquement les types de données des colonnes (par exemple, texte, nombres, ou dates) pour qu'ils soient cohérents avec leur contenu. Cela évite les erreurs de typage qui pourraient rendre certaines opérations ou analyses impossibles.



- **Column Filter :**

Cette étape permet de sélectionner ou supprimer certaines colonnes spécifiques dans les données. Cela permet de garder uniquement les informations pertinentes pour l'analyse et d'éliminer les colonnes inutiles, rendant ainsi le jeu de données plus léger et structuré.



- **Column Renamer :**

Ce nœud est utilisé pour renommer les colonnes afin de les rendre plus compréhensibles ou conformes à un standard. Des noms de colonnes clairs et adaptés réduisent le risque d'erreurs dans les étapes suivantes et facilitent la lecture et l'utilisation des données par les utilisateurs.

Dialog - 3:12 - Column Renamer

Column	New name
SpecialOfferID	SpecialOfferID

Add column

- CSV Writer :

Ce nœud final exporte les données nettoyées et préparées dans un nouveau fichier CSV. Après transformation, ce fichier final peut être utilisé pour d'autres analyses ou partagé avec d'autres outils ou utilisateurs. Cette étape clôture le workflow en produisant un jeu de données fiable et organisé.

Dialog - 3:18 - CSV Writer

File

Settings Advanced Settings Comment Header Encoding Flow Variables Job Manager Selection

Output location

Write to Local File System

File C:\Users\HINNOVIS\Downloads\Cspecialoffer.csv

Write options Create missing folders If exists: overwrite append fail
There already exists a file with the specified path 'C:\Users\HINNOVIS\Downloads\Cspecialoffer.csv'.

Format

Column Delimiter , Row Delimiter System Default
Quote Char Quote Escape Char

Header

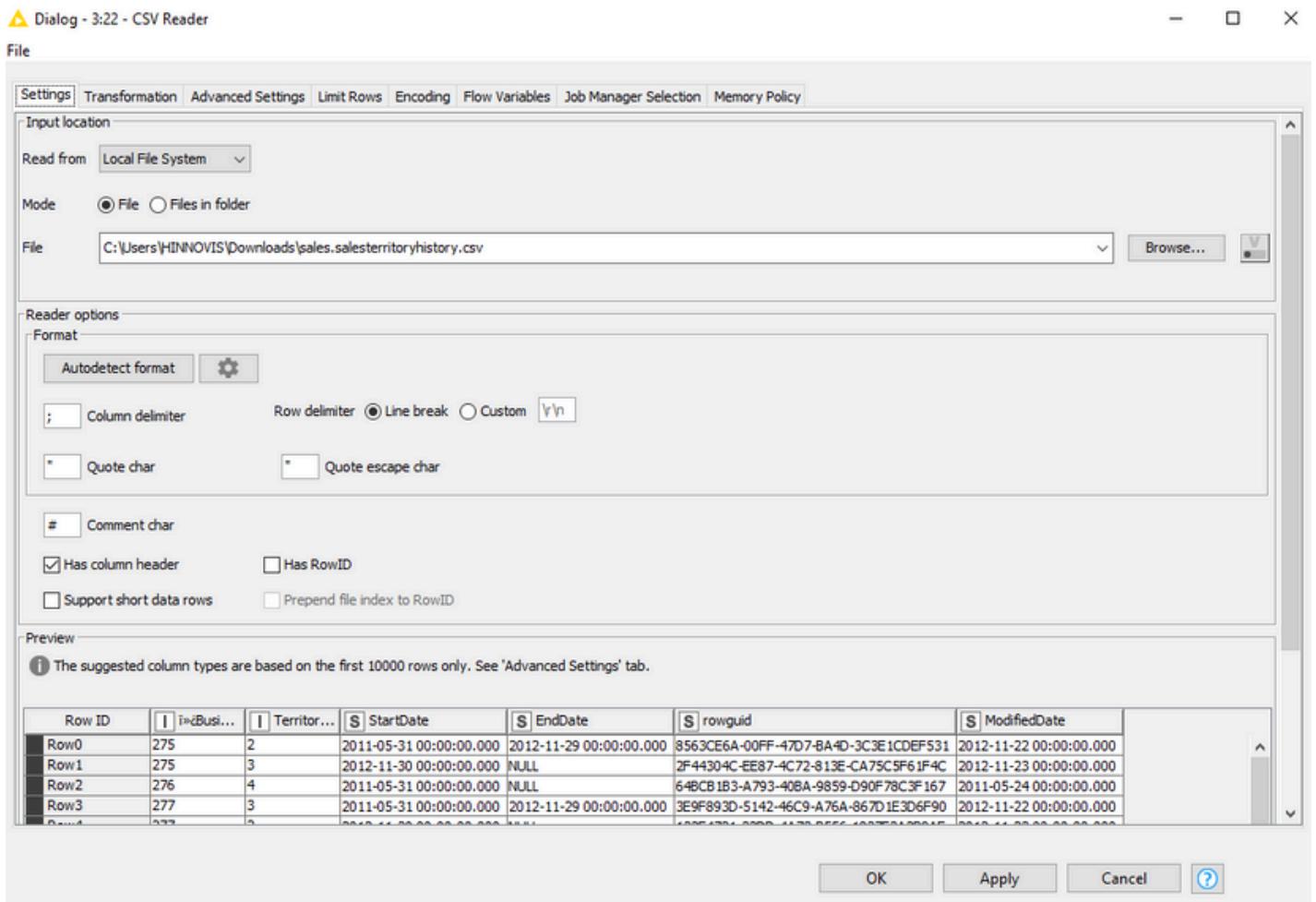
Write column header
 Don't write column headers if file exists
 Write RowID

OK Apply Cancel ?

Pour salesterritoryhistory :

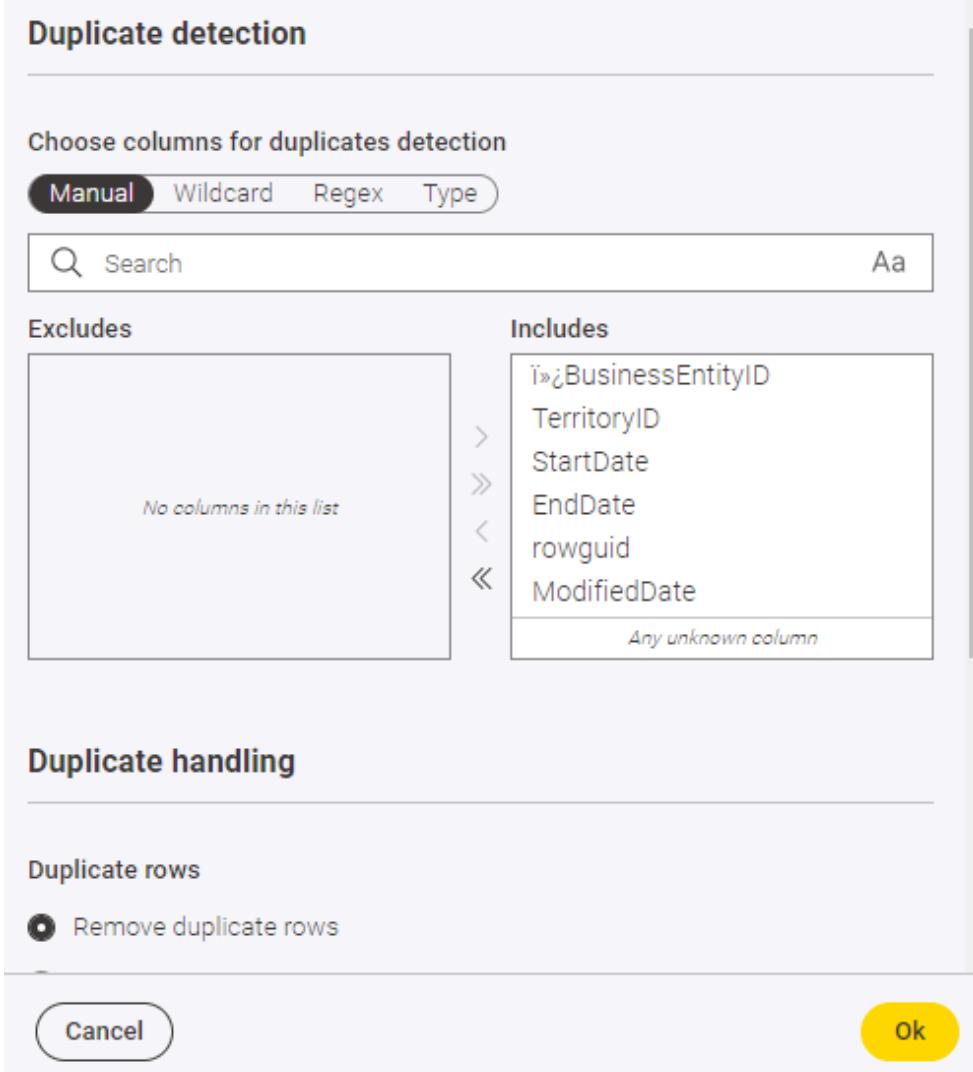
- **Lecture des Données (CSV Reader):**

Ce nœud permet de lire un fichier CSV contenant les données brutes. C'est une étape essentielle, car elle constitue le point de départ du workflow, permettant d'importer les données pour les étapes suivantes de nettoyage et de transformation.



- **Duplicate Row Filter :**

Ce nœud est utilisé pour identifier et supprimer les lignes en double dans le fichier. En éliminant les doublons, il garantit que chaque enregistrement est unique, ce qui est important pour éviter les biais ou erreurs dans les analyses futures.



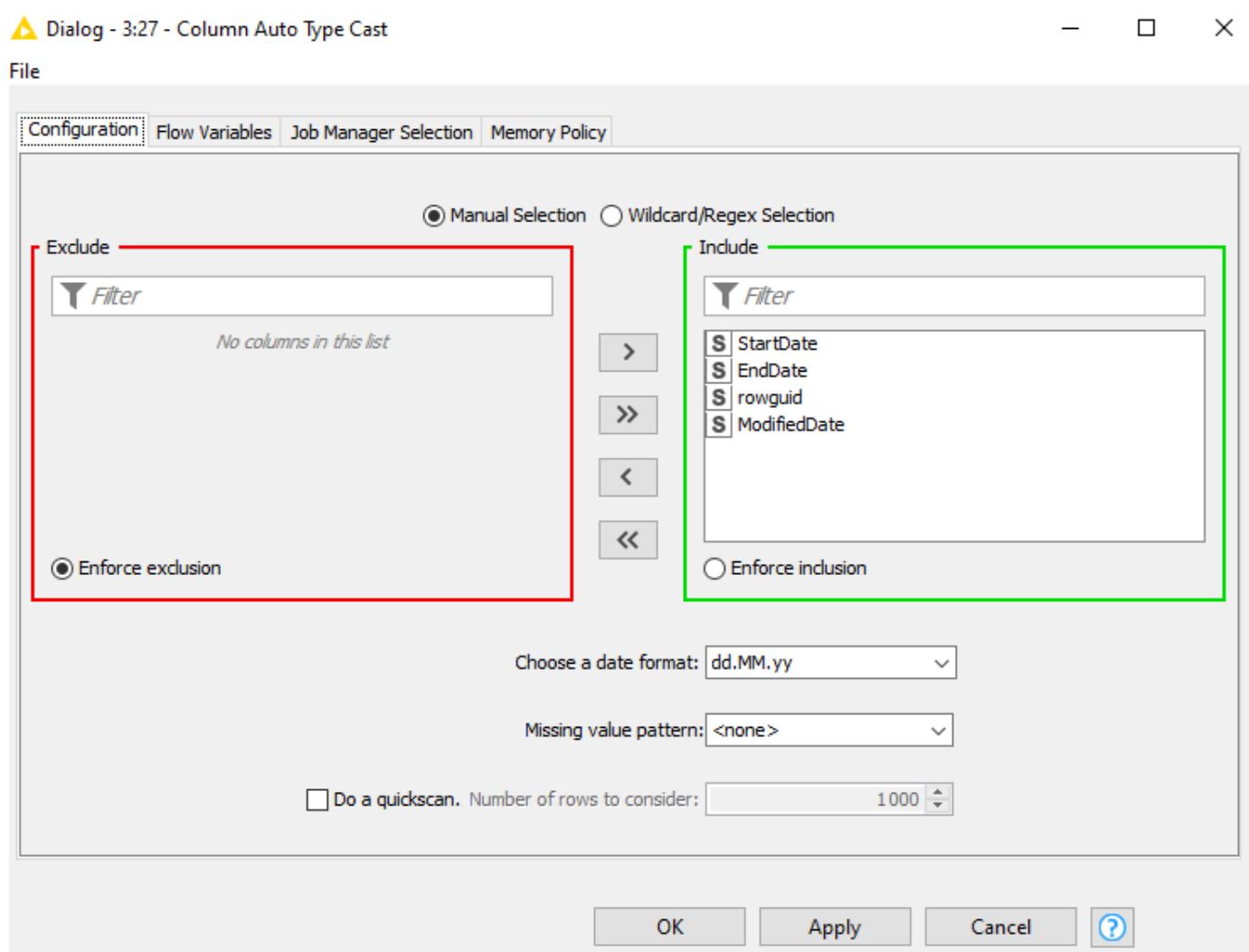
- **Gestion des Valeurs Manquantes (Missing Value) :**

Ce nœud permet de traiter les valeurs manquantes dans les données. Il peut soit remplir les valeurs manquantes avec des données calculées ou par défaut, soit supprimer les lignes ou colonnes affectées. Cette étape est cruciale pour assurer la complétude des données et éviter les erreurs ou biais qui pourraient influencer les résultats de l'analyse.

Number (integer)	Median
String	Most Frequent Value

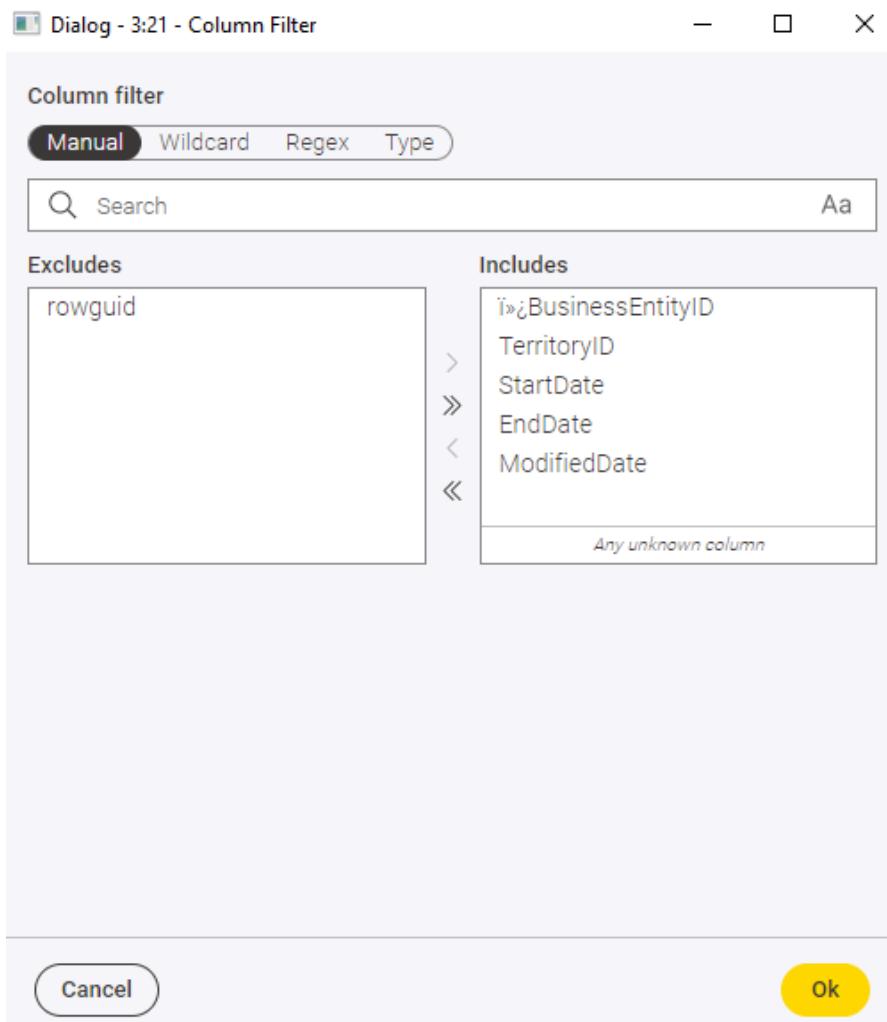
- **Column Auto Type Cast :**

Ce nœud harmonise automatiquement les types de données des colonnes (par exemple, texte, nombres, ou dates) pour qu'ils soient cohérents avec leur contenu. Cela évite les erreurs de typage qui pourraient rendre certaines opérations ou analyses impossibles.



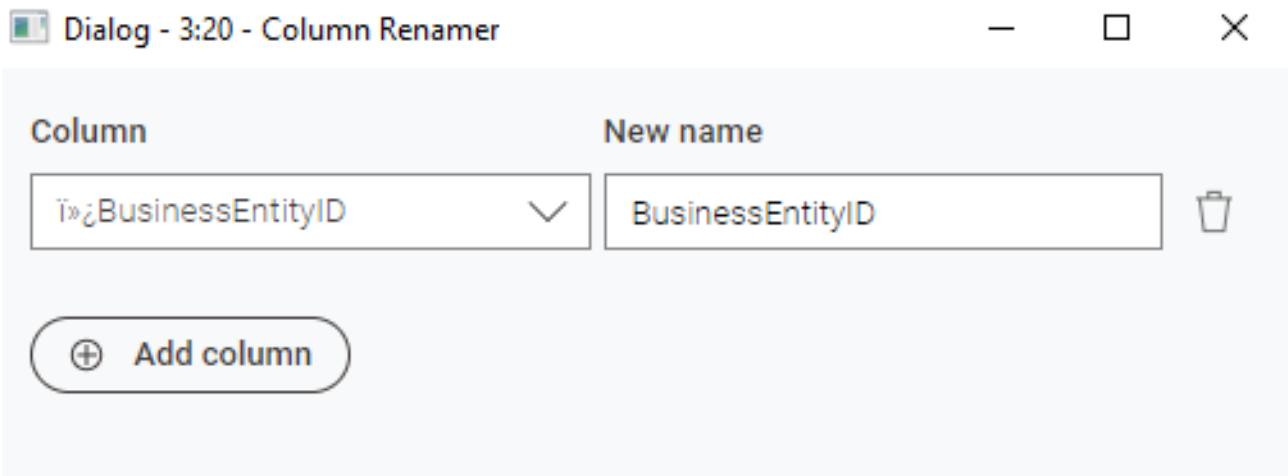
- **Column Filter :**

Cette étape permet de sélectionner ou supprimer certaines colonnes spécifiques dans les données. Cela permet de garder uniquement les informations pertinentes pour l'analyse et d'éliminer les colonnes inutiles, rendant ainsi le jeu de données plus léger et structuré.



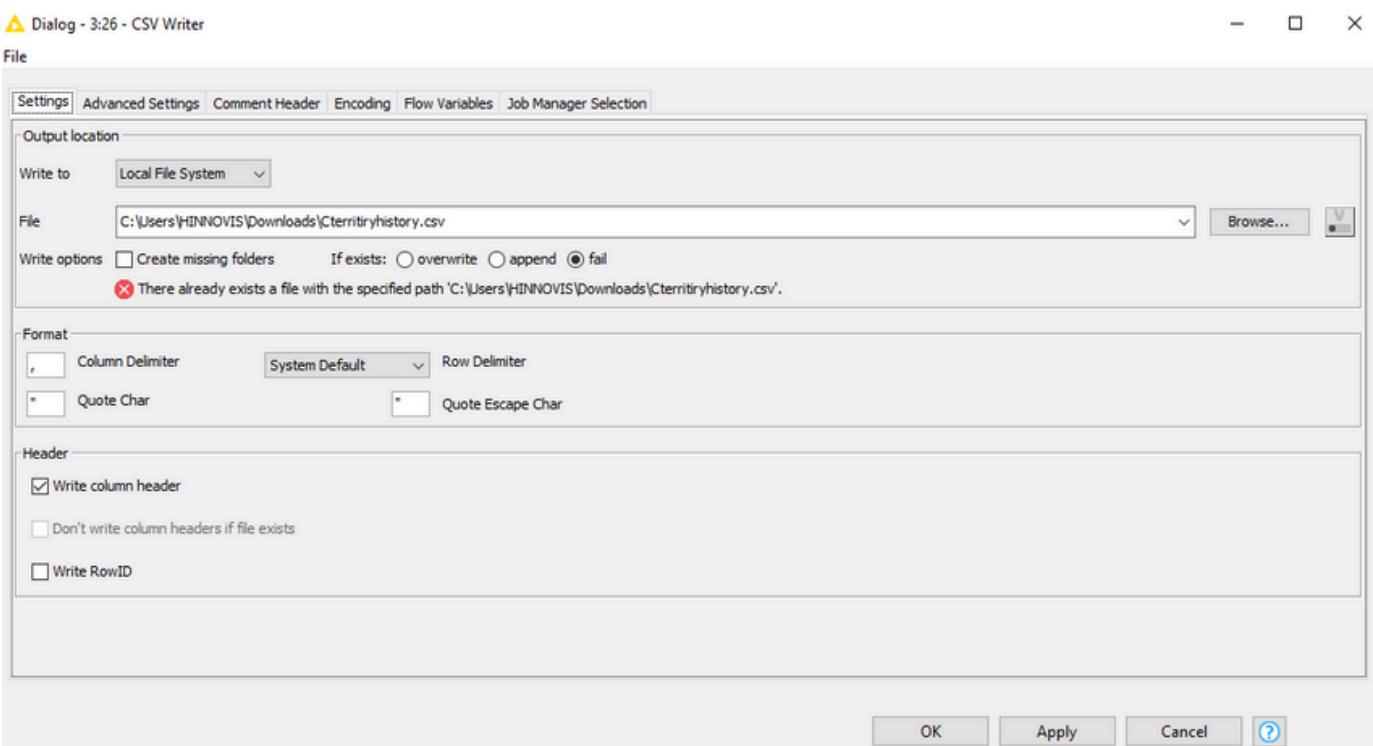
- **Column Renamer :**

Ce nœud est utilisé pour renommer les colonnes afin de les rendre plus compréhensibles ou conformes à un standard. Des noms de colonnes clairs et adaptés réduisent le risque d'erreurs dans les étapes suivantes et facilitent la lecture et l'utilisation des données par les utilisateurs.



- CSV Writer :

Ce nœud final exporte les données nettoyées et préparées dans un nouveau fichier CSV. Après transformation, ce fichier final peut être utilisé pour d'autres analyses ou partagé avec d'autres outils ou utilisateurs. Cette étape clôture le workflow en produisant un jeu de données fiable et organisé.



03 Analyse et Visualisation

Introduction à la visualisation des données sur Power BI

Dans le cadre de ce projet, nous avons utilisé Power BI comme outil principal pour la visualisation et l'analyse des données. Power BI, développé par Microsoft, est une solution de Business Intelligence puissante et intuitive qui permet de transformer des données brutes en visualisations interactives et en tableaux de bord dynamiques.

L'objectif principal de cette étape était de donner du sens aux données collectées et de les rendre facilement compréhensibles à travers des graphiques pertinents. Cette approche visuelle permet d'identifier rapidement les tendances, les points forts et les axes d'amélioration liés aux performances de l'organisation.

Pourquoi Power BI ?

Power BI a été choisi pour ses nombreux avantages, notamment :

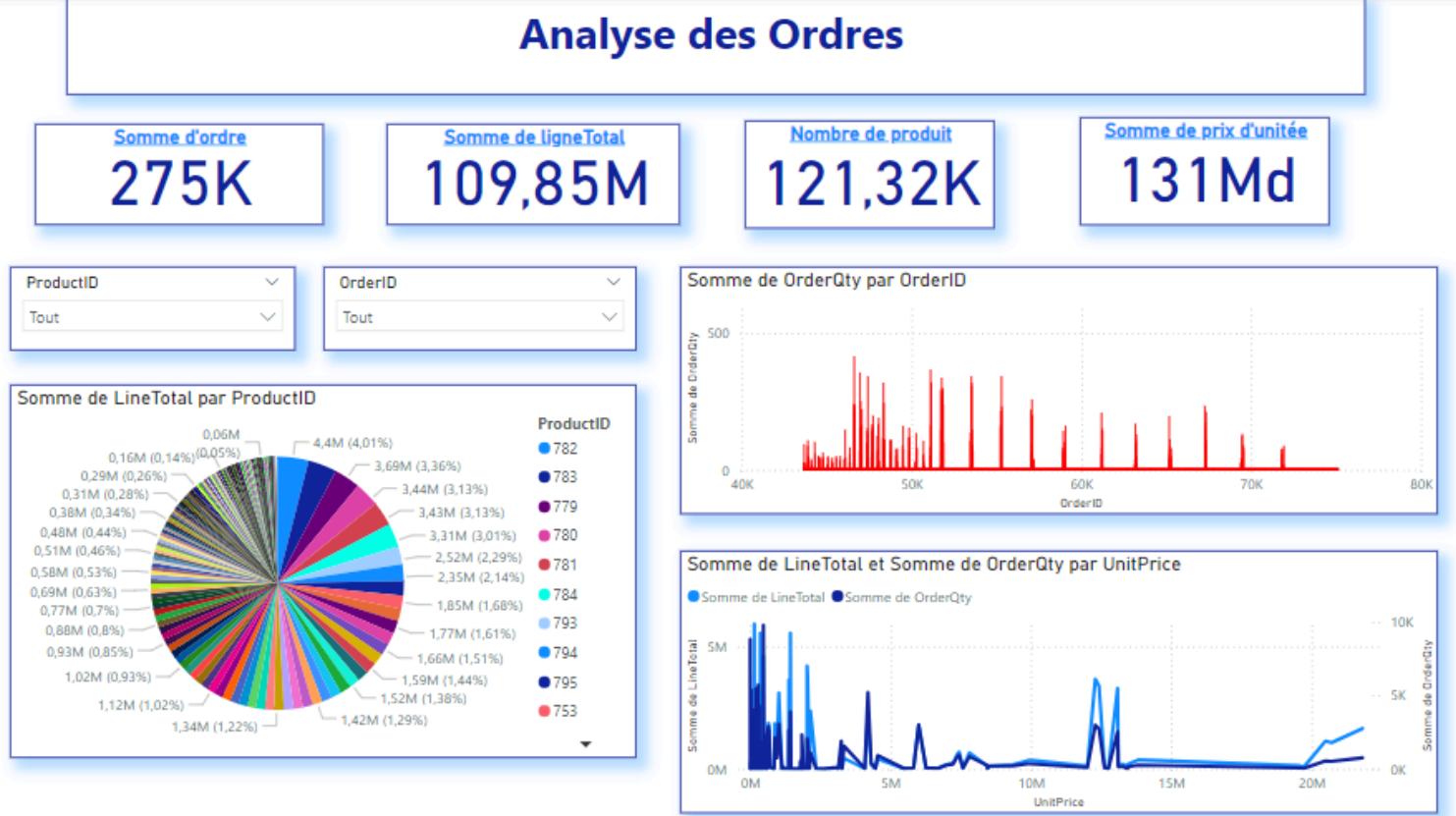
- Interface intuitive : Sa simplicité d'utilisation permet de concevoir des tableaux de bord interactifs en un temps réduit.
- Connectivité étendue : Il offre la possibilité de se connecter à différentes sources de données (Excel, bases de données SQL, Google Sheets, etc.).
- Visualisations dynamiques : Les graphiques interactifs permettent de filtrer, explorer et analyser les données en temps réel.
- Facilité de partage : Les rapports peuvent être publiés en ligne ou partagés avec les parties prenantes pour une prise de décision collaborative

Objectifs de la visualisation

- Fournir une vue d'ensemble des indicateurs clés de performance (KPI) tels que les ventes, les bonus, les commandes, etc.
- Identifier les relations entre les différentes dimensions, comme les entités commerciales, les commerciaux, les territoires et les produits.
- Simplifier l'analyse des données complexes pour faciliter la prise de décision stratégique.
- Permettre un suivi continu des performances grâce à des rapports interactifs et personnalisables.

Présentation du tableau de bord

Analyse des Ordres



Indicateurs clés (en haut)

- Somme d'ordre (275K) :
 - Représente le nombre total d'ordres (commandes) enregistrés.
 - L'unité "K" signifie "milliers".
- Somme de LineTotal (109,85M) :
 - Montant total généré par les commandes, exprimé en millions (M).
- Nombre de produit (121,32K) :
 - Indique le nombre total d'unités de produits vendus ou commandés.
- Somme de prix d'unité (131Md) :
 - Montant total des prix unitaires des produits, exprimé en milliards (Md).

Filtres (en haut à gauche)

- ProductID : Permet de sélectionner ou filtrer les données par identifiant de produit spécifique.
- OrderID : Permet de filtrer par identifiant de commande.

Visualisations

- Diagramme circulaire (Somme de LineTotal par ProductID) :
 - Représente la répartition des montants totaux générés (LineTotal) par produit, chaque segment représentant un ProductID.
 - Les pourcentages indiquent la contribution de chaque produit au total.
- Graphique en barres (Somme de OrderQty par OrderID) :
 - Visualisation des quantités commandées (OrderQty) pour chaque commande (OrderID).
 - Les barres rouges montrent la distribution et les variations par commande.
- Graphique combiné (Somme de LineTotal et OrderQty par UnitPrice) :
 - Montre la corrélation entre le prix unitaire (UnitPrice), le montant total des lignes (LineTotal), et les quantités commandées (OrderQty).
 - Ligne bleue : Montants totaux (LineTotal).
 - Barres noires : Quantités commandées (OrderQty).



En-tête

- "Somme Sales" (337Md) : Indique le total des ventes réalisées, exprimé en milliards (Md).
- "Somme Sales Last Year" (220Md) : Total des ventes effectuées l'année précédente.
- "Business Entity" (701) : Nombre total d'entités commerciales ou d'organisations impliquées dans les ventes.
- "Somme de Bonus" (5M) : Total des bonus attribués aux commerciaux, exprimé en millions (M).

Filtres

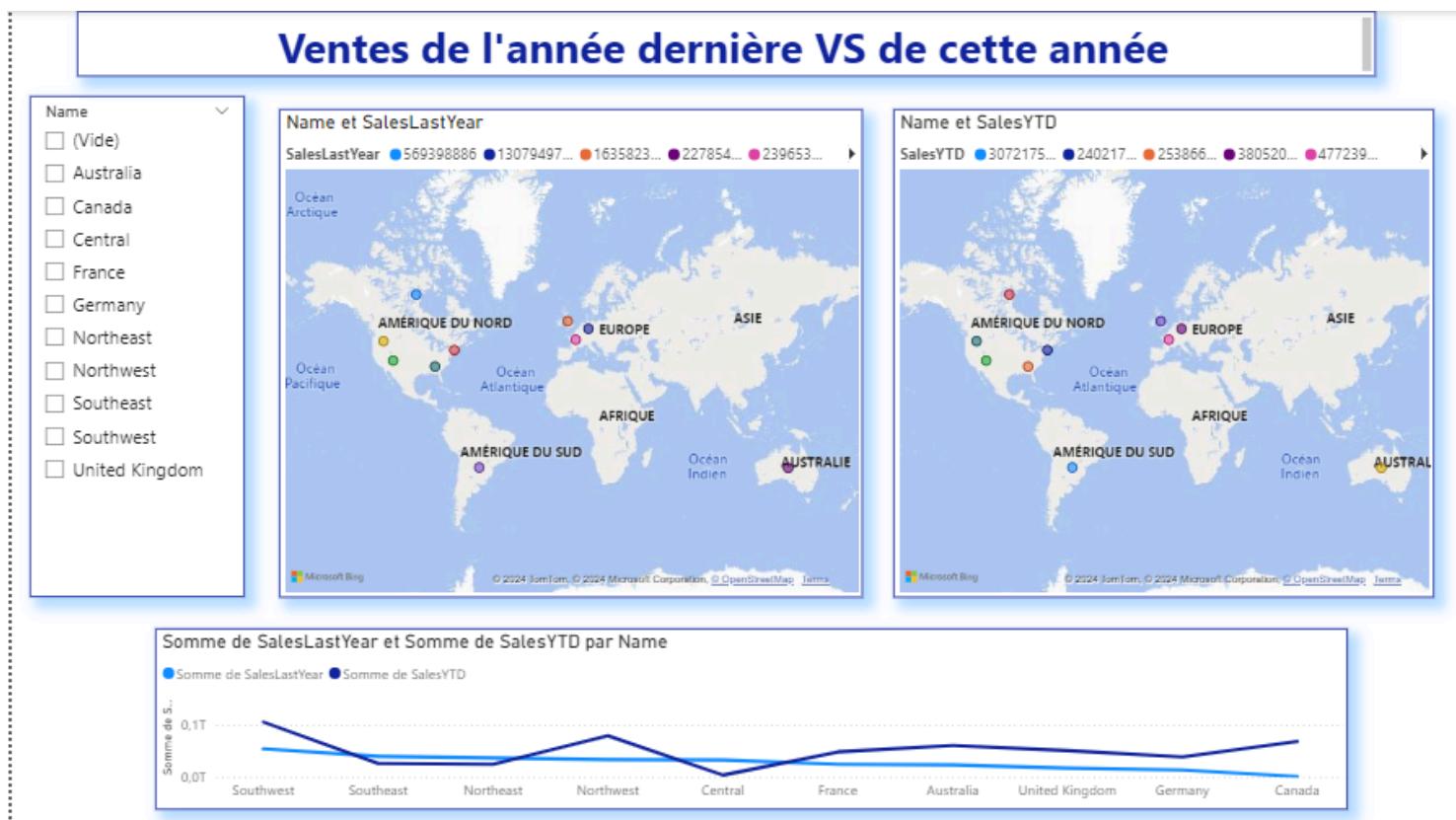
- **SalesPersonID** : Permet de filtrer les données par identifiant du commercial.
- **BusinessEntityID** : Permet de filtrer par identifiant de l'entité commerciale.

Visualisations

- Graphique en barres horizontales (Somme de Bonus par BusinessEntityID) :
 - Montre la répartition des bonus attribués à chaque entité commerciale (BusinessEntityID).
 - Les barres représentent les montants des bonus, permettant de visualiser les entités les mieux récompensées.
- Graphique linéaire (Nombre de BusinessEntityID par SalesPersonID) :
 - Affiche la relation entre les commerciaux (SalesPersonID) et le nombre d'entités commerciales avec lesquelles ils ont travaillé.
 - Les variations permettent d'identifier les commerciaux les plus actifs ou impliqués.
- Graphique à barres empilées (Somme de SalesYTD par SalesLastYear) :
 - Visualise les ventes cumulées pour l'année en cours (SalesYTD) par rapport à celles de l'année précédente (SalesLastYear).
 - Couleurs :
 - Vert clair : Augmentation des ventes.
 - Rouge : Diminution des ventes.
 - Vert foncé : Total des ventes.
- Graphique à barres groupées (Somme de CommissionPct par BusinessEntityID et TerritoryID) :
 - Représente les pourcentages de commission attribués par entité commerciale (BusinessEntityID) et territoire (TerritoryID).
 - Permet de visualiser les différences de commissions entre les régions ou entités.
- Diagramme circulaire (Nombre de ReasonType par Name) :
 - Indique les motifs ou raisons liés aux ventes ou actions commerciales, répartis par catégorie.
 - Chaque segment représente une catégorie (ex. "Demo Event", "Quality", "On Promotion").
 - Les pourcentages montrent la part relative de chaque type.

Interprétations possibles

- **Performance des ventes** : Comparaison entre les ventes totales (337Md) et celles de l'année précédente (220Md) pour évaluer la progression.
 - **Répartition des bonus** : Identifier les entités commerciales ou commerciaux les mieux récompensés.
 - **Activité des commerciaux** : Évaluer les commerciaux les plus actifs en termes de collaborations avec les entités commerciales.
 - **Analyse territoriale** : Comprendre la distribution des commissions par région (TerritoryID).
 - **Catégories d'actions commerciales** : Identifier les motifs les plus fréquents pour améliorer les stratégies commerciales.



Deux cartes géographiques (Maps)

- Carte des ventes de l'année dernière :
 - Cette carte illustre les ventes réalisées durant l'année précédente.
 - Chaque point ou zone géographique est coloré en fonction de l'intensité des ventes (plus la couleur est foncée, plus les ventes sont élevées).
 - Les noms des territoires ou régions de vente sont souvent visibles, permettant une analyse géographique détaillée.
- Carte des ventes de l'année en cours :
 - Cette carte montre les performances commerciales pour l'année en cours, avec une présentation similaire à celle de l'année dernière.
 - Permet de repérer rapidement les zones où les ventes ont augmenté, diminué ou stagné.

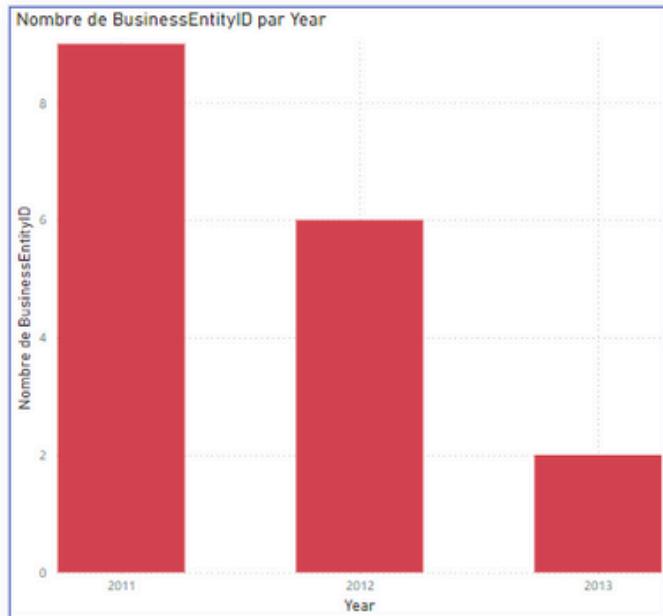
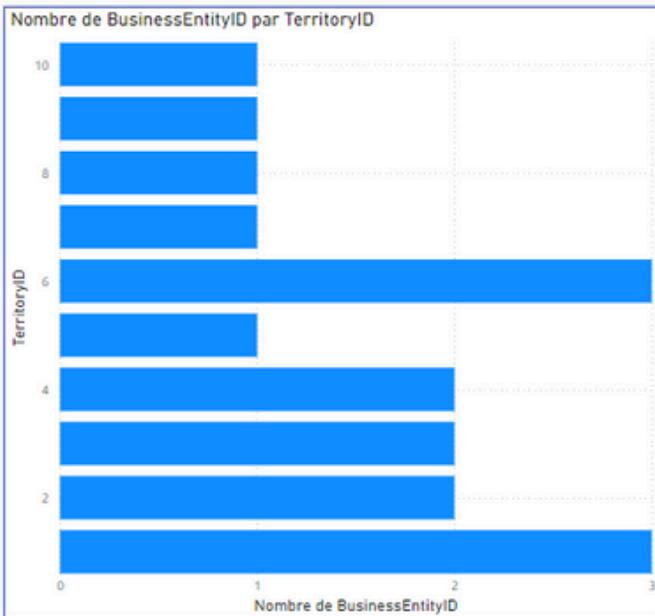
Filtre par nom de territoire

- Ce filtre interactif permet de sélectionner un ou plusieurs territoires spécifiques pour focaliser l'analyse sur ces régions.
- Les cartes et le graphique linéaire s'adaptent dynamiquement en fonction du ou des territoires sélectionnés.
- Cela facilite une analyse ciblée des performances géographiques.

Diagramme de Ligne

Le diagramme de ligne dans ce tableau de bord compare les ventes de l'année dernière à celles de l'année en cours en fonction des noms des pays. L'axe horizontal représente les différents pays, tandis que l'axe vertical illustre les montants des ventes. Deux lignes distinctes permettent d'observer les performances : l'une pour l'année dernière et l'autre pour l'année en cours. Ce graphique offre une vision claire des variations des ventes entre les pays, en identifiant les régions qui ont amélioré leurs performances, celles qui sont restées stables et celles où les ventes ont diminué. Avec le filtre de pays, cette visualisation devient interactive, permettant une analyse approfondie et ciblée pour orienter les stratégies commerciales en fonction des régions géographiques

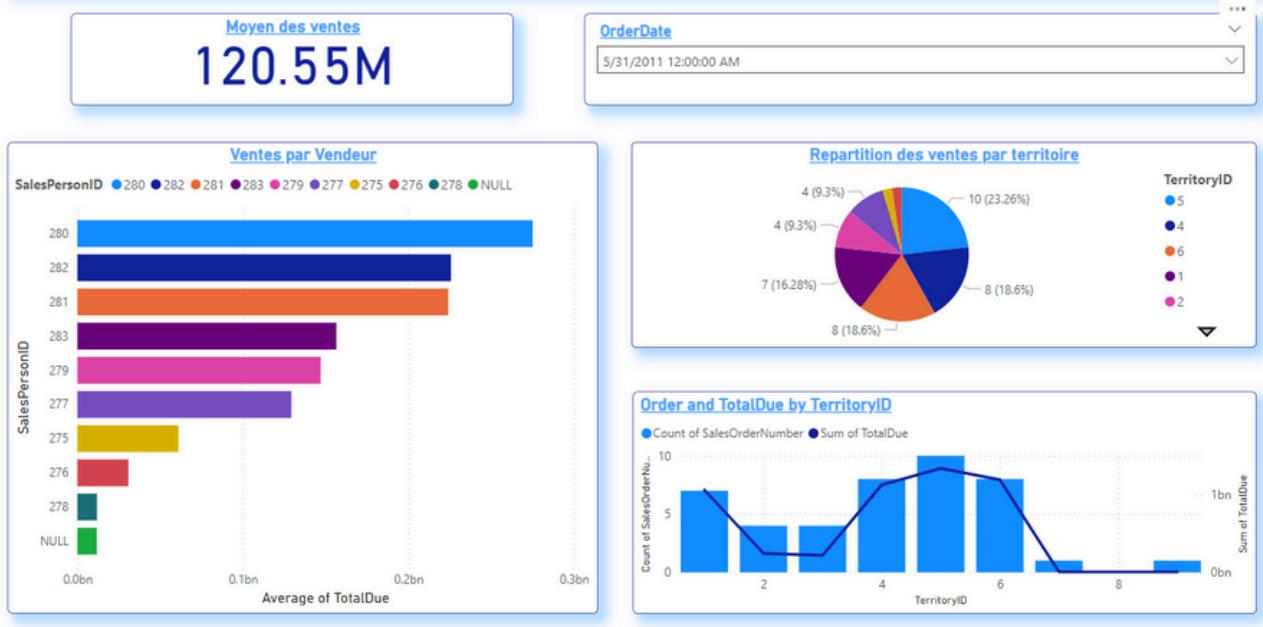
Entité par territoire et par année



Le tableau de bord présente une analyse de la répartition géographique et temporelle d'entités. Le premier graphique met en évidence des disparités significatives entre les territoires, certains concentrant une activité bien plus importante que d'autres. Le second graphique révèle une tendance à la baisse du nombre total d'entités au fil des années. Ces données suggèrent une concentration géographique de l'activité et une possible contraction du marché ou des changements dans les méthodes de comptabilisation.

Rapport des ventes mensuelles

Filters



Titre principal : "Rapport des ventes mensuelles"

- **But :** Fournir une vue d'ensemble des performances de vente pour une période donnée, avec un accent sur les ventes par territoire et par vendeur.

Visualisation 1 : Moyenne des ventes (Card)

- Indicateur affiché : 120.55M (Moyenne des ventes)
- Signification : Cette carte met en évidence la moyenne totale des ventes réalisées sur la période sélectionnée.
- Interprétation : Une valeur élevée indique une bonne performance globale des ventes. Elle sert de référence rapide pour évaluer les résultats moyens.

Visualisation 2 : Filtre par date (Slicer)

- Contenu : Une sélection dynamique pour choisir la période ou la date des données à analyser.
- Signification : Permet de visualiser les tendances ou les performances sur des plages temporelles spécifiques.

Visualisation 3 : Ventes par vendeur (Bar Chart)

- **Axes :**

X : Moyenne des ventes (Average of TotalDue).

Y : Identifiant des vendeurs (SalesPersonID).

- **But :** Comparer la performance des vendeurs sur la période donnée.
- **Interprétation :**

Les vendeurs avec des barres plus longues (comme 280 et 282) génèrent le plus de ventes.

Les vendeurs avec des performances faibles peuvent nécessiter un soutien ou une révision de leur stratégie.

Visualisation 4 : Répartition des ventes par territoire (Pie Chart)

- **Contenu :** Part de marché des ventes pour chaque territoire (TerritoryID).
- **Signification :**

Le graphique montre que le territoire 10 domine avec 23.26% des ventes.

Les territoires 8 et 7 sont également performants avec respectivement 18.6% et 16.28%.

Les territoires moins performants peuvent nécessiter des efforts marketing ou un renforcement des équipes.

Visualisation 5 : Commandes et TotalDue par territoire (Bar et Line Chart combinés)

- **Axes :**

X : Identifiants des territoires (TerritoryID).

Y gauche (Barres) : Nombre total de commandes (Count of SalesOrderNumber).

Y droite (Ligne) : Montant total des ventes (Sum of TotalDue).

- **Signification :**

Les barres montrent le volume de commandes par territoire.

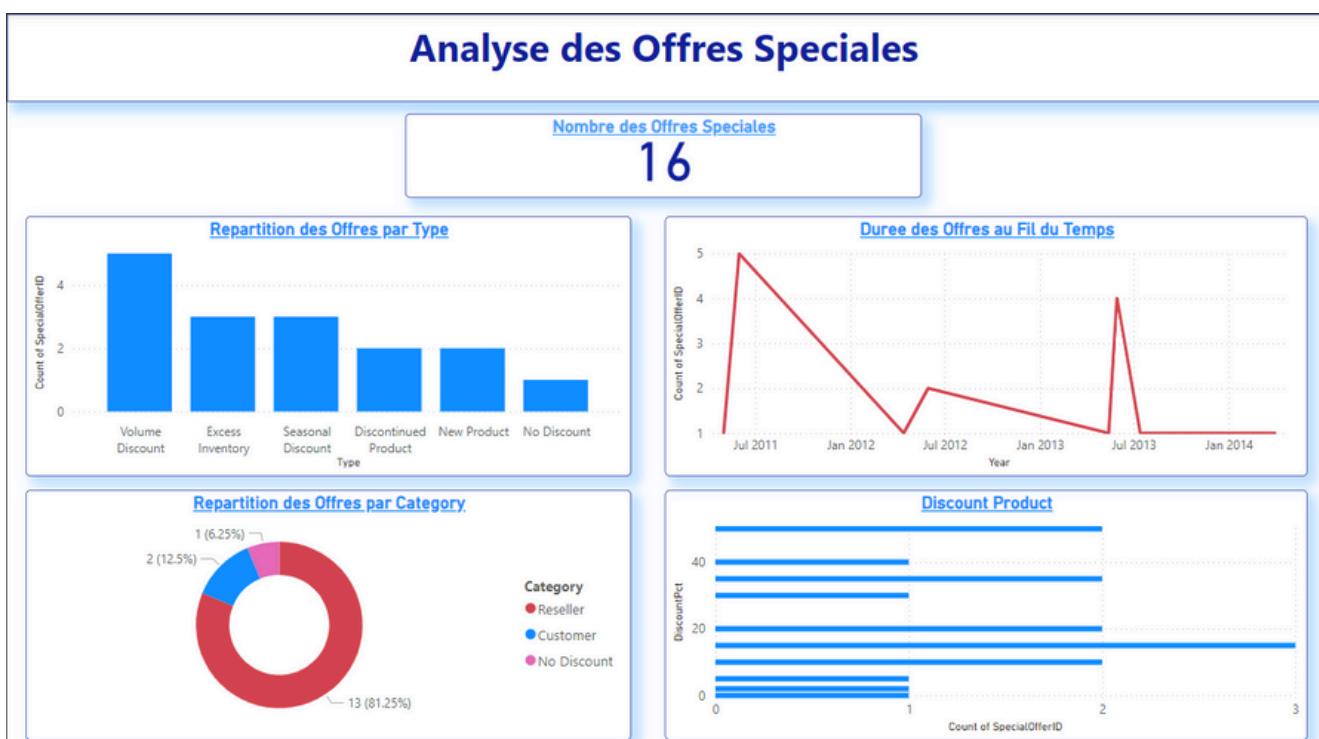
La ligne montre les montants totaux correspondants.

Une corrélation entre les deux permet d'identifier les territoires générant des ventes élevées avec peu de commandes (ventes à valeur élevée) ou l'inverse.

Interactivité avec les slicers et filtres

- Les slicers permettent de choisir des périodes ou de filtrer les territoires pour une analyse ciblée.
- Impact des filtres : Les visualisations se mettent à jour dynamiquement, permettant d'identifier des tendances spécifiques.

Ce tableau de bord offre une vue complète des performances de vente, en mettant en lumière les meilleurs vendeurs, territoires, et tendances temporelles. L'approche visuelle simplifie l'identification des opportunités de croissance et des zones nécessitant une attention particulière.



Titre Principal : "Analyse des Offres Spéciales"

- But : Présenter une vue d'ensemble sur les types, la répartition, la durée, et l'impact des offres spéciales.

Visualisation 1 : Nombre des Offres Spéciales (Card)

- Indicateur affiché : 16
- Signification : Le nombre total d'offres spéciales enregistrées dans la base de données.
- Interprétation : Ce chiffre indique la diversité ou la fréquence des promotions mises en place.

Visualisation 2 : Répartition des Offres par Type (Bar Chart)

- Axes :

X : Types d'offres spéciales (Type).

Y : Nombre d'offres (Count of SpecialOfferID).

- Signification : Le graphique montre la distribution des offres en fonction de leur type, comme "Volume Discount", "Excess Inventory", etc.
- Interprétation :
 - Le type "Volume Discount" est le plus fréquent, ce qui reflète une stratégie axée sur des remises en volume.
 - Les types comme "No Discount" sont rares, ce qui peut indiquer des promotions moins traditionnelles.

Visualisation 3 : Durée des Offres au Fil du Temps (Line Chart)

- Axes :

X : Année (Year).

Y : Nombre d'offres (Count of SpecialOfferID).

- Signification :
 - Ce graphique suit l'évolution du nombre d'offres par année.

- Interprétation :
 - Une baisse marquée en 2012 suivie d'un pic en 2013 pourrait indiquer une stratégie révisée ou une campagne spéciale.
 - Les périodes de faible activité nécessitent une analyse pour identifier des opportunités.

Visualisation 4 : Répartition des Offres par Catégorie (Donut Chart)

- Contenu : Catégories d'offres (Category) avec leur part respective dans le total.
- Signification :
 - La majorité des offres (81.25%) concernent la catégorie "No Discount".
 - Les catégories "Reseller" (12.5%) et "Customer" (6.25%) sont moins représentées.

- Interprétation :
 - Une prédominance de "No Discount" peut refléter un manque de ciblage pour les revendeurs ou les clients finaux.

Visualisation 5 : Discount Product (Bar Chart horizontal)

- Axes :

X : Nombre d'offres (Count of SpecialOfferID).

Y : Taux de remise (DiscountPct).

- Signification :

Ce graphique illustre la fréquence des offres en fonction du pourcentage de réduction.

- Interprétation :

Les remises les plus courantes sont dans des plages spécifiques.

Une concentration sur certains taux de remise peut refléter une stratégie uniforme ou des préférences client.

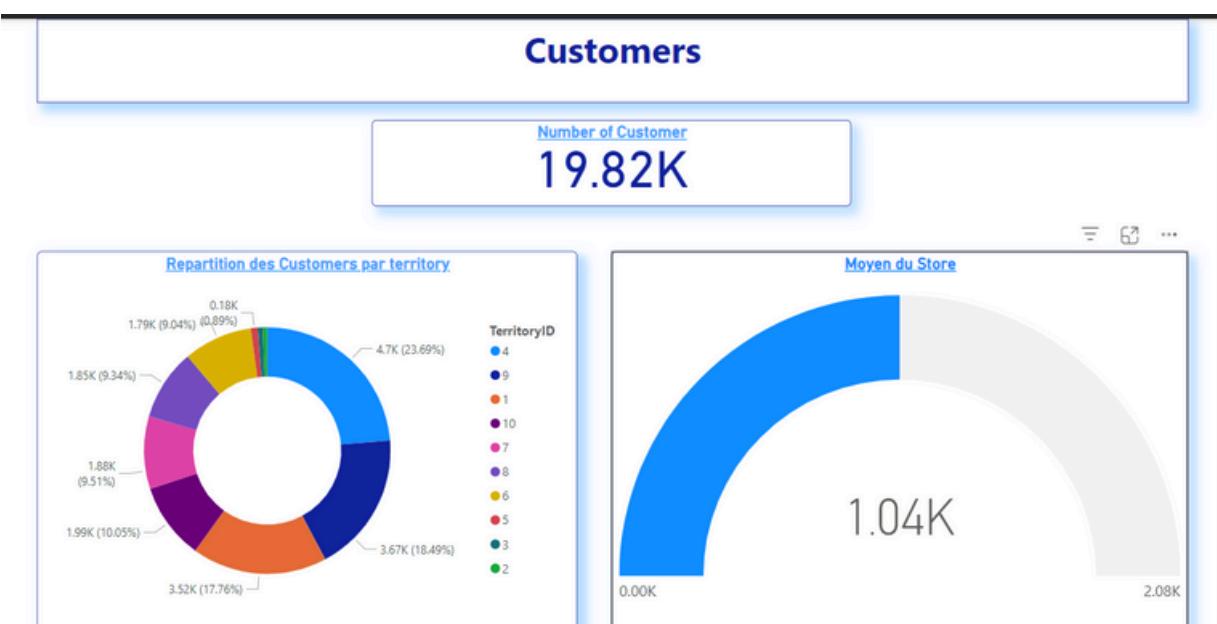
Interactivité et Filtres

- Filtres dynamiques :

Permettent de se concentrer sur des périodes, des types ou des catégories spécifiques.

Une interactivité avec les slicers ajuste les graphiques pour des analyses précises.

Ce tableau de bord offre une analyse claire des promotions spéciales, identifiant les types et catégories les plus fréquents, ainsi que leurs impacts dans le temps. Il permet aux décideurs d'évaluer l'efficacité des stratégies actuelles et d'explorer des opportunités d'amélioration.



Titre Principal : "Customers"

- **But** : Analyser la répartition géographique des clients et mesurer la moyenne des données par magasin.

Visualisation 1 : Nombre Total de Clients (Card)

- **Indicateur affiché** : 19.82K
- **Signification** :

Représente le total des clients enregistrés dans la base de données.

- **Interprétation** :

Ce nombre montre l'étendue de la clientèle actuelle et permet une estimation globale des performances des territoires.

Visualisation 2 : Répartition des Clients par Territoire (Donut Chart)

- **Contenu** : Répartition des clients en fonction de leur territoire d'origine (TerritoryID).
- **Signification** : Chaque segment représente un territoire avec la proportion des clients associés.
- **Données clés** :

Territory 4 : 4.7K clients (23.69%), le plus grand groupe.

Territory 9 : 3.67K clients (18.49%).

Autres territoires varient entre 0.18K (0.89%) et 3.52K (17.76%).

- **Interprétation** :

Les territoires 4 et 9 concentrent une majorité de clients, indiquant une importance stratégique.

Les territoires avec moins de 1K clients pourraient nécessiter des efforts de développement commercial.

Visualisation 3 : Moyenne des Données par Store (Gauge Chart)

- **Indicateur affiché** : 1.04K
- **Signification** : Représente la moyenne des clients par magasin ou par zone de distribution.
- **Interprétation** :

Une moyenne stable reflète une distribution équilibrée.

Si cette moyenne est éloignée du potentiel maximal (2.08K), cela pourrait indiquer des opportunités de croissance dans certains magasins.

Interactivité et Filtres

- Filtres dynamiques possibles :

Par TerritoryID : Focaliser sur un territoire spécifique pour analyser ses performances détaillées.

Par période ou type de client (non illustré ici).

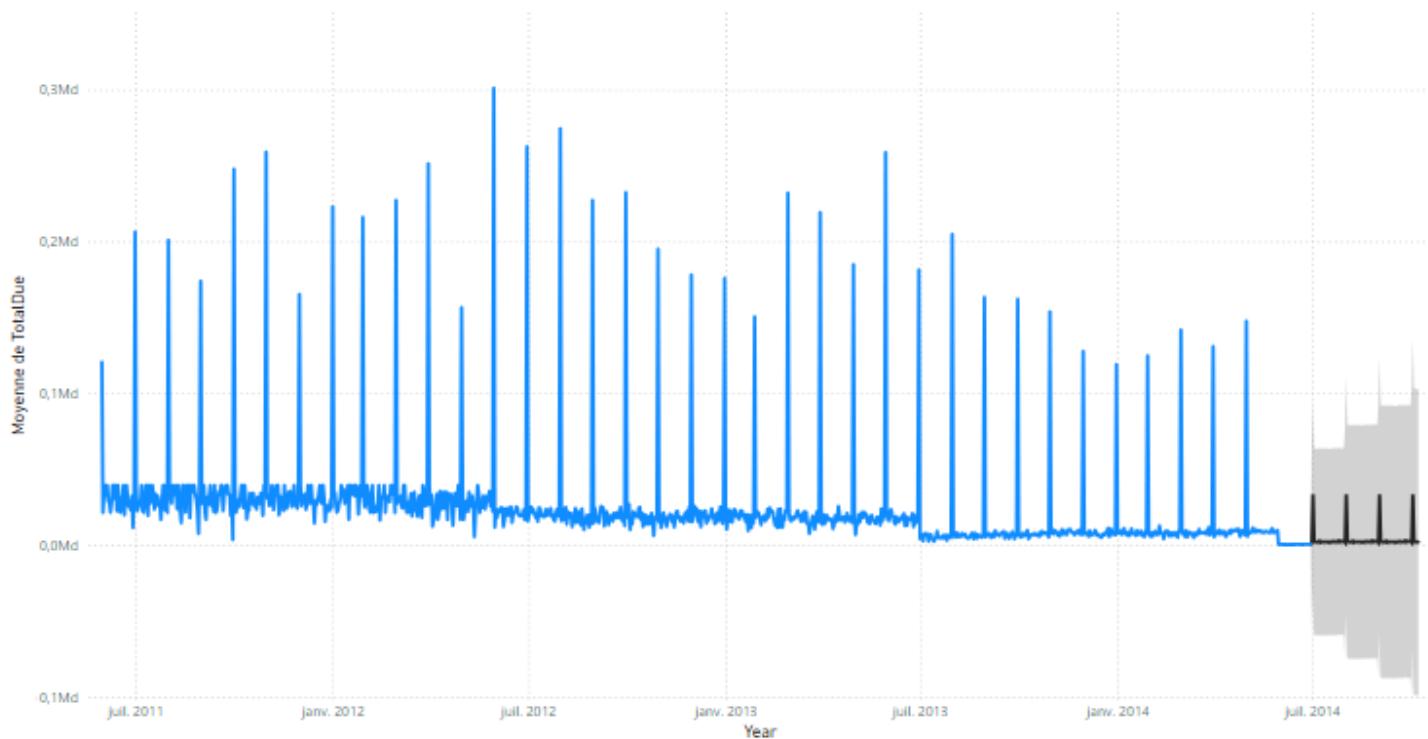
- Actions possibles :

Comparer la moyenne des clients par territoire.

Identifier des corrélations entre les territoires et les catégories de clients (résidentiel, commercial, etc.).

Ce tableau de bord fournit des insights précieux pour évaluer la répartition géographique des clients et les performances des magasins. Les territoires à fort volume (comme le 4) méritent une attention particulière pour optimiser les stratégies de fidélisation, tandis que les territoires sous-représentés nécessitent des plans pour accroître leur part de marché.

Moyenne de TotalDue par Year, Quarter, Month et Day



Prédiction de Moyenne de TotalDue par Année, Trimestre, Mois et Jour

Ce tableau de bord présente l'évolution historique et les prévisions de la moyenne de TotalDue, une métrique qui semble être liée à des transactions ou des activités financières.

Éléments clés :

- **Partie bleue (données historiques) :**

- La période couverte s'étend de juillet 2011 à juillet 2014.
- On observe une forte volatilité avec des pics et des creux marqués dans les valeurs de TotalDue.
- Des tendances saisonnières apparaissent, avec des valeurs plus élevées généralement autour de juillet-août chaque année.
- Il y a également une variabilité importante au niveau journalier, avec des fluctuations quotidiennes significatives.

- **Partie grise (prévisions) :**

- À partir de juillet 2014, les données passent en mode de prévision ou de forecasting.
- La partie grise montre les projections futures de la moyenne de TotalDue.
- Ces prévisions semblent indiquer une tendance à la baisse progressive des valeurs de TotalDue dans les années à venir.

Cette visualisation permet d'avoir une compréhension globale de l'évolution historique et des perspectives futures de cette métrique financière complexe. Les données réelles et les prévisions sont clairement identifiées, offrant une perspective complète sur les tendances et la dynamique de TotalDue.

04 Gestion de Projet et Collaboration

utilisation de Taiga.io avec répartition des tâches par sprint

l'utilisation de Taiga.io pour la gestion du projet lié à l'analyse des données AdventureWorks. Les différentes étapes du projet ont été structurées en tâches et organisées par sprint pour assurer un suivi efficace.

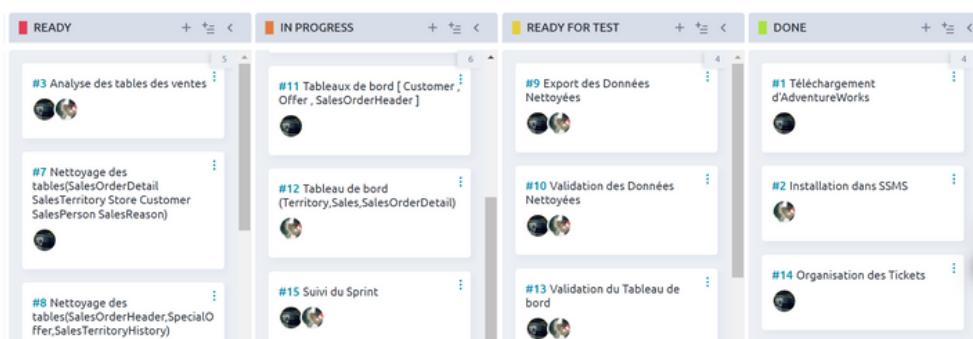
Structure du Tableau Kanban

Le tableau Kanban est organisé en cinq colonnes :

	New	Tâches nouvellement créées et non encore planifiées.
	Ready	Tâches prêtes à être commencées.
	In Progress	Tâches en cours d'exécution.
	Ready for Test	Tâches terminées mais en attente de validation.
	Done	Tâches finalisées et validées.

Sprints

Le projet a été divisé en trois sprints principaux :



Sprint 1 : Mise en place et exploration des données

Objectifs :

1. Installer et configurer la base de données AdventureWorks.
2. Explorer les tables pertinentes (ventes).

Tâches :



New

Création d'un projet sur Taiga.io pour la gestion des tâches



Ready

Télécharger AdventureWorks2019.

Importer le fichier .bak dans SQL Server Management Studio (SSMS).

Analyser les tables suivantes :

- SalesOrderHeader
- SalesOrderDetail
- SalesTerritory
- Store
- Customer
- SalesPerson
- SalesReason
- SpecialOffer
- SalesTerritoryHistory



In Progress

- Documenter les champs de chaque table.
- Créer des diagrammes ERD pour représenter les relations entre toutes les tables.



Ready for Test

- Vérifier la cohérence des relations identifiées dans les tables.



Done

- Base de données installée avec succès.
- Documentation complète des tables et relations finalisée.

Sprint 2 : Nettoyage des données avec KNIME

Objectifs :

1. Identifier et nettoyer les anomalies dans les données de toutes les tables liées aux ventes.
2. Exporter les données nettoyées pour la visualisation.

Tâches :



New

Définir les critères de nettoyage pour chaque table.



Ready

Créer un workflow KNIME pour le nettoyage des tables.

Nettoyer les données des tables suivantes :

- SalesOrderHeader
- SalesOrderDetail
- SalesTerritory
- Store
- Customer
- SalesPerson
- SalesReason
- SpecialOffer
- SalesTerritoryHistory



In Progress

- Tester le workflow KNIME avec des échantillons de données.
- Documenter le processus de nettoyage pour chaque table.



Ready for Test

- Valider les données nettoyées (cohérence, absence de doublons).
- Exporter les tables nettoyées dans des fichiers CSV.



Done

- Workflow KNIME publié sur GitHub.
- Données nettoyées et prêtes pour la visualisation.

Sprint 3 : Visualisation des données avec Power BI

Objectifs :

1. Créer des visualisations pertinentes pour les données de ventes.
2. Intégrer les visualisations dans un tableau de bord interactif.

Tâches :

New

Identifier les KPIs à visualiser, tels que :

- Volume des ventes
- Ventes par produit
- Ventes par territoire
- Performance par employé
- Raisons des ventes (SalesReason)

Ready

Créer des graphiques pour les ventes par :

- Territoire
- Produit
- Employé (SalesPerson)
- Magasin (Store)

Créer un graphique pour les tendances mensuelles des ventes.

In Progress

- Intégrer les graphiques dans un tableau de bord Power BI.
- Ajouter des filtres interactifs pour permettre une analyse dynamique.

Ready for Test

- Tester le tableau de bord pour s'assurer qu'il répond aux besoins.
- Corriger les éventuels problèmes d'affichage ou d'interaction.

Done

- Tableau de bord finalisé et prêt à être présenté.

Résultats

- Sprint 1 : Toutes les tables liées aux ventes ont été explorées et documentées, y compris leurs relations.
- Sprint 2 : Les données ont été nettoyées avec succès via KNIME, avec un workflow publié sur GitHub.
- Sprint 3 : Un tableau de bord interactif Power BI contenant des visualisations pertinentes pour toutes les tables a été livré.

Répartition détaillée des tâches par membre de l'équipe

MEMBRE 1 : RESPONSABLE DE L'EXPLORATION ET DU NETTOYAGE DES TABLES PRINCIPALES (SALESORDERHEADER, SALESORDERDETAIL, SALESTERRITORY) :

- Importation des données et documentation des tables.
- Création et documentation du workflow KNIME.
- Vérification et exportation des données nettoyées.

MEMBRE 2 : RESPONSABLE DE L'ANALYSE DES TABLES ANNEXES (STORE, CUSTOMER, SALESPERSON, SALESREASON, SPECIALOFFER, SALESTERRITORYHISTORY) ET DES VISUALISATIONS :

- Création des diagrammes ERD pour toutes les tables.
- Développement des visualisations Power BI.
- Intégration des KPIs et finalisation du tableau de bord.

L'utilisation de Taiga.io a permis une gestion efficace et transparente du projet en suivant les principes des méthodes agiles. Grâce à une organisation claire des sprints et des tâches, le projet a livré des résultats de qualité, incluant des données nettoyées, un workflow KNIME publié, et un tableau de bord Power BI interactif prêt pour une utilisation professionnelle.

05 Conclusion et Perspectives

L'analyse des données de vente d'AdventureWorks a permis d'identifier les performances globales, les zones géographiques et produits les plus rentables, ainsi que les tendances saisonnières des ventes. Les régions performantes et les produits phares offrent des opportunités de croissance, tandis que les zones et articles moins performants nécessitent des ajustements stratégiques.

Grâce à ces insights, AdventureWorks peut optimiser ses campagnes marketing, sa gestion des stocks et ses prévisions de ventes, renforçant ainsi sa compétitivité et sa capacité à saisir de nouvelles opportunités sur le marché.

PERSPECTIVES

Pour prolonger et enrichir cette analyse, plusieurs axes peuvent être explorés :

- Segmentation Client Avancée : Identifier les profils clients les plus rentables pour personnaliser les campagnes marketing et fidéliser davantage.
- Analyse Prédictive : Intégrer des modèles de machine learning pour prévoir les tendances de vente à plus long terme.
- Analyse Concurrentielle : Comparer les performances d'AdventureWorks avec celles de concurrents pour ajuster les stratégies commerciales.
- Optimisation Logistique : Utiliser les données géographiques pour améliorer la gestion des stocks et réduire les délais de livraison.
- Automatisation des Rapports : Mettre en place des tableaux de bord interactifs en temps réel pour un suivi continu des performances.
- Élargissement des Données : Intégrer d'autres sources de données, comme les retours clients ou les données du marché, pour une vision plus complète.

Ces perspectives permettront à AdventureWorks de s'adapter aux évolutions du marché et de consolider ses performances commerciales sur le long terme.