

Université Sultan Moulay Slimane
Ecole Nationale des Sciences Appliquées-
Béni Mellal

2024/2025

Rapport de Mini Projet

Cycle d'ingénieur

Filière :

Transformation Digitale industrielle (TDI)

Exploration et Analyse d'AdventureWorks

Réalisé par :

CHAIMA TAOUFIK &

HASNAE DRIB

Encadré par :

Pr. ELBAGHAZAOUI

Table des matières

1	Introduction.....	4
2	Téléchargement et Installation d'AdventureWorks.....	5
2.1	Lien de téléchargement :	5
2.2	Procédure d'installation :	5
3	Découverte des Tables	6
3.1	Présentation des Tables Explorées :	6
3.2	Relations Entre les Tables :	6
4	Nettoyage des Données et Gestion de leur Qualité	8
4.1	Problèmes Identifiés :	8
4.2	Approche de Nettoyage :	9
4.2.1	Identification des Données Manquantes :	9
4.2.2	Suppression ou Gestion des Doublons :	9
4.2.3	Standardisation des Formats :	9
4.2.4	Correction des Anomalies :	9
4.2.5	Outils utilisés :	9
4.2.6	Résultat :	10
4.3	Solutions Appliquées :	10
4.3.1	Scripts SQL Utilisés :	10
4.3.2	Indicateurs Avant et Après Nettoyage :	10
5	Visualisation des Données.....	12
5.1	Objectifs des Visualisations :	12
5.2	Étapes de Création des Tableaux de Bord :	12
6	Gestion des Tâches avec Taiga.io	19
6.1	Configuration de Taiga.io	19
6.2	Répartition des Tâches :	20
6.3	Suivi Agile :	21
7	Collaboration sur GitHub.....	21
7.1	Etape de création du dépôt github :	21
7.2	Partager les scripts et rapports	22
7.3	Revue de code collaborative	23
8	Résultats et Analyses.....	24
8.1	État Initial vs État Final des Données	24
8.2	Indicateurs Clés de Performance (KPI)	24
8.3	Synthèse des Insights Décisionnels	25

9	Conclusion :	26
	Référence :	27
	Annexes :	28

Liste des Figures

Figure 1: Code sql 1	6
Figure 2: Code sql 2	6
Figure 3: exploration de tables	7
Figure 4:Relation entre les tables principales d'achat	7
Figure 5: Code sql 3	9
Figure 6:Code sql 4	9
Figure 7:Code sql 5	10
Figure 8:Code sql 6	10
Figure 9:Code sql 7	10
Figure 10: Résultat 1	11
Figure 11:Résultat 2	11
Figure 12: Tableau de bord principal	13
Figure 13:Visualisation1	14
Figure 14:Visualisation 2	15
Figure 15: Visualisation 3	16
Figure 16: Visualisation 4	17
Figure 17 : Visualisation 4	18
Figure 18:Interface Scrum	19
Figure 19:Tickets de Taiga	20
Figure 20:Vue des tickets	21
Figure 21:Dépôt GitHub	22
Figure 22:commande 1	22
Figure 23: Fichiers du dépôt	22
Figure 24:Commande 2	22
Figure 25:Commande 3	23
Figure 26:Commande 4	23
Figure 27:Commande 5	23
Figure 28:Commande 6	23
Figure 29:Pull Request	23
Figure 30:Tableau 1	24
Figure 31:Tableau 2	24
Figure 32: Script python pour la prédiction	28
Figure 33: output du script pour TotalDue	28
Figure 34: Output du script pour TaxAmt	29
Figure 35: Output du script pour Freight	29

1 Introduction

AdventureWorks est une entreprise fictive créée par Microsoft pour fournir un ensemble de bases de données d'exemples. Ces bases sont conçues pour simuler des scénarios commerciaux réels et sont largement utilisées dans le domaine de l'apprentissage, des tests et de la démonstration des fonctionnalités de SQL Server. La base de données AdventureWorks couvre différents aspects de gestion d'une entreprise multinationale de fabrication appelée **AdventureWorks Cycles**, spécialisée dans la production et la vente de vélos, de pièces détachées, et d'accessoires. Elle est particulièrement utile pour étudier des processus tels que les ventes, les achats, la production, et la gestion des ressources humaines.

Dans ce projet, nous avons utilisé la version **AdventureWorks2019**, qui propose des données riches et réalistes pour explorer et analyser les différents processus métiers d'une entreprise. L'objectif de notre projet est de découvrir, nettoyer et analyser les données d'AdventureWorks pour en extraire des informations utiles et pertinentes. À travers différentes étapes pratiques, ce projet vise à :

- **Explorer les données** disponibles et documenter leur structure.
- **Améliorer la qualité des données** en identifiant et en corrigeant les anomalies (valeurs manquantes, doublons, incohérences).
- **Créer des visualisations interactives** pour présenter des insights décisionnels à l'aide d'outils comme Power BI ou Tableau.
- **Collaborer efficacement en équipe** en utilisant des outils modernes comme Taiga.io pour la gestion des tâches et GitHub pour le partage des fichiers et des scripts.

Et pour les approches méthodologiques :

➤ **description des étapes suivies :**

▪ Téléchargement et Installation de la Base de Données :

La base de données AdventureWorks2019 a été téléchargée et importée dans **SQL Server Management Studio (SSMS)**. Cette étape a permis de configurer un environnement d'analyse prêt à l'emploi.

▪ Exploration et Documentation des Tables :

Une exploration approfondie des tables attribuées à l'équipe a été réalisée pour comprendre leur structure et leurs relations. Chaque table a été documentée en détaillant les colonnes importantes, les types de données et les clés primaires/étrangères.

▪ Nettoyage des Données :

Des scripts SQL ont été développés pour identifier et corriger les anomalies des données, notamment :

Les valeurs manquantes.

Les doublons.

Les incohérences de format.

Donc, cette étape a permis d'améliorer la qualité et la fiabilité des données pour l'analyse.

▪ **Création des Visualisations :**

À l'aide de **Power BI**, des tableaux de bord interactifs ont été créés pour analyser les commandes d'achat et les performances des fournisseurs. Ces visualisations offrent des insights clairs et exploitables.

▪ **Gestion des Tâches :**

L'outil **Taiga.io** a été utilisé pour organiser et planifier les tâches en suivant une approche agile. Les tâches ont été divisées en épopées et User Stories, et leur progression a été suivie à l'aide des méthodes KANBAN et SCRUM.

▪ **Collaboration et Partage :**

Le travail collaboratif a été assuré via **GitHub**, où un dépôt a été créé pour partager les scripts SQL et les fichiers Power BI. Les contributions ont été suivies grâce à des branches et des Pull Requetés, garantissant ainsi la qualité et la synchronisation du travail.

2 Téléchargement et Installation d'AdventureWorks

2.1 Lien de téléchargement :

Le lien utilisé pour télécharger la base de données est : **AdventureWorks Databases**

2.2 Procédure d'installation :

❖ Étapes pour importer le fichier .bak dans SQL Server Management Studio :

- Téléchargez le fichier **AdventureWorks2019.bak** depuis le site officiel de Microsoft et placez-le dans un dossier (ex. : C:\Backup).
- Ouvrez **SQL Server Management Studio (SSMS)** et connectez-vous à votre instance SQL Server.
- Dans l'explorateur d'objets, faites un clic droit sur **Databases** et sélectionnez **Restore Database**.
- Sous **Source**, choisissez **Device** et cliquez sur le bouton
- Cliquez sur **Add**, sélectionnez le fichier .bak et cliquez sur **OK**.
- Sous **Destination**, vérifiez que le champ **Database** affiche le nom `AdventureWorks2019`.
- Dans l'onglet **Files**, assurez-vous que les chemins des fichiers MDF et LDF sont corrects.
- Cliquez sur **OK** pour lancer la restauration.
- Une notification de succès s'affichera une fois la restauration terminée.

❖ Étapes pour vérifier l'intégrité des données après installation :

Développons Databases dans l'explorateur d'objets en vérifiant que la base de données AdventureWorks2019 est listée, on exécutons la commande suivante pour afficher toutes les tables :

```
SELECT TABLE_SCHEMA, TABLE_NAME
FROM INFORMATION_SCHEMA.TABLES
WHERE TABLE_TYPE = 'BASE TABLE';
```

Figure 1: Code sql 1

Testez les relations clés entre les tables avec une requête JOIN, cette étape permet d'assurer que la base de données est installée correctement et que les données sont prêtes à être utilisées.

```
SELECT POH.PurchaseOrderID, POH.OrderDate, V.Name AS VendorName
FROM Purchasing.PurchaseOrderHeader POH
INNER JOIN Purchasing.Vendor V ON POH.VendorID = V.BusinessEntityID;
```

Figure 2: Code sql 2

3 Découverte des Tables

3.1 Présentation des Tables Explorées :

Les tables essentielles explorées pour ce projet sont :

- **PurchaseOrderHeader** : Contient les informations générales des commandes d'achat, telles que les dates, les montants totaux, et les fournisseurs associés.
- **Vendor** : Liste des fournisseurs avec leurs informations de contact et leur évaluation.
- **PurchaseOrderDetail** : Contient les détails des commandes, notamment les produits, quantités et prix unitaires.

3.2 Relations Entre les Tables :

Ces tables sont reliées par des clés primaires et étrangères :

- La table **Vendor** est reliée à **PurchaseOrderHeader** via la colonne `VendorID`.
- La table **PurchaseOrderHeader** est reliée à **PurchaseOrderDetail** via la colonne `PurchaseOrderID`.

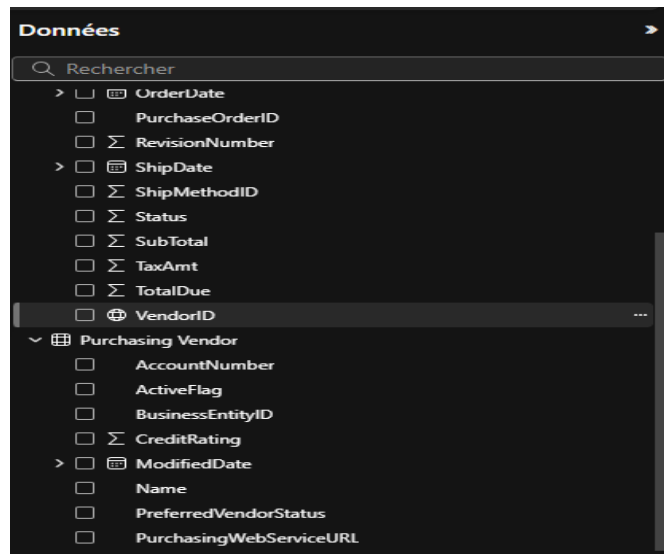


Figure 3: exploration de tables

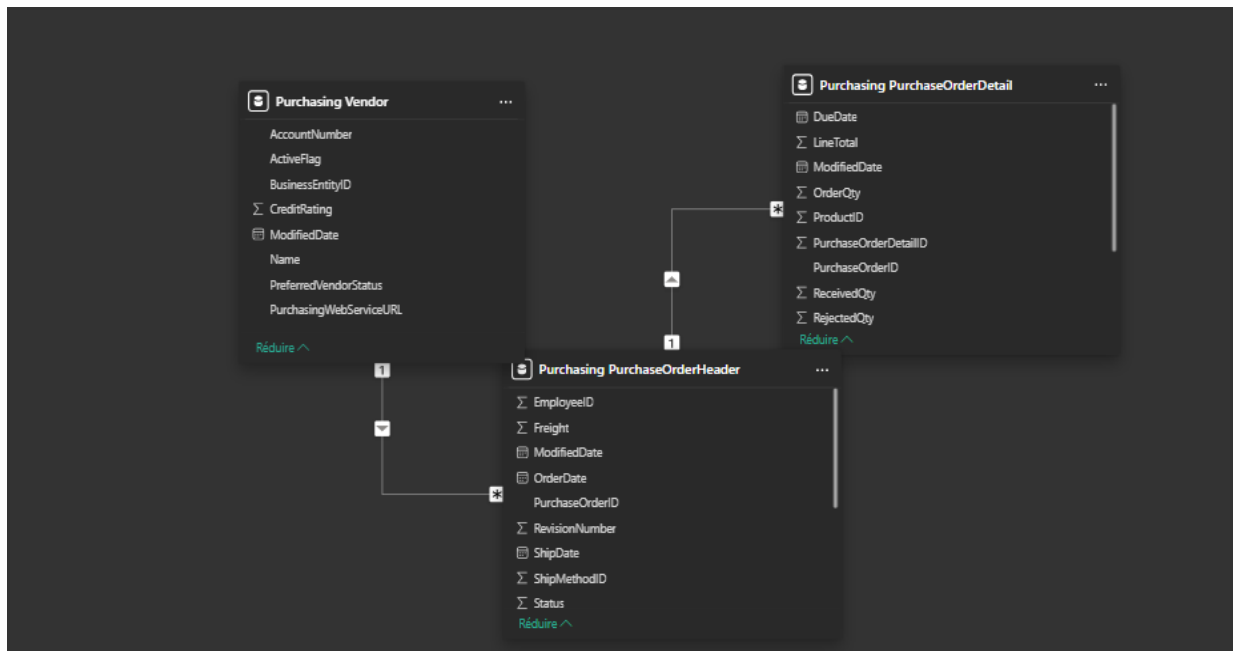


Figure 4: Relation entre les tables principales d'achat

Le diagramme ci-dessus illustre les relations entre les trois tables principales du système de gestion des commandes d'achat de l'entreprise *AdventureWorks*. Ces tables sont : `Purchasing.Vendor`, `Purchasing.PurchaseOrderHeader`, et `Purchasing.PurchaseOrderDetail`. Chaque table joue un rôle spécifique dans la gestion des données d'achat :

1. Table `Purchasing.Vendor`

- ✓ Cette table contient les informations sur les fournisseurs, comme leur nom, leur numéro de compte (`AccountNumber`), leur note de crédit (`CreditRating`) et leur statut (actif/inactif).
Clé primaire : `BusinessEntityID`.
- ✓ Elle est reliée à la table `Purchasing.PurchaseOrderHeader` via la colonne `BusinessEntityID`, qui identifie le fournisseur associé à une commande.

2. Table `Purchasing.PurchaseOrderHeader`

Elle regroupe les informations générales sur les commandes d'achat, telles que la date de commande (`OrderDate`), la date de livraison prévue (`ShipDate`), et le montant total dû (`Freight`).

- ✓ Clé primaire : `PurchaseOrderID`.
- ✓ Cette table est le point central de la relation, reliant les fournisseurs et les détails spécifiques des produits commandés.
- ✓ Relation avec `Purchasing.Vendor` : `BusinessEntityID` (fournisseur).
- ✓ Relation avec `Purchasing.PurchaseOrderDetail` : `PurchaseOrderID` (détails de commande).

3. Table `Purchasing.PurchaseOrderDetail`

Elle contient les informations détaillées sur chaque produit d'une commande, telles que les quantités commandées (`OrderQty`), le prix unitaire (`LineTotal`), et l'identifiant du produit (`ProductID`).

- ✓ Clé primaire : `PurchaseOrderDetailID`.
- ✓ Reliée à la table `Purchasing.PurchaseOrderHeader` par la colonne `PurchaseOrderID`.

4 Nettoyage des Données et Gestion de leur Qualité

4.1 Problèmes Identifiés :

Lors de l'exploration initiale des données, plusieurs problèmes de qualité ont été relevés :

- **Données Manquantes :**
Certaines colonnes présentaient des valeurs nulles ou absentes, notamment dans les champs critiques comme les identifiants de produits ou les dates de commandes.
- **Doublons :**
Des enregistrements en double ont été trouvés, particulièrement dans les tables

contenant les historiques des commandes. Ces doublons pouvaient fausser les analyses si non traités.

- **Formats Incorrects :**
Les dates n'étaient pas toujours dans un format standardisé.
Certains champs numériques contenaient des chaînes de caractères, rendant leur traitement difficile.
- **Anomalies Spécifiques :**
Des incohérences entre les quantités commandées (`OrderQty`) et les montants totaux (`LineTotal`) dans la table `Purchasing.PurchaseOrderDetail`.
Des fournisseurs actifs avec un statut inactif dans la table `Purchasing.Vendor`.

4.2 Approche de Nettoyage :

4.2.1 Identification des Données Manquantes :

Utilisation de requêtes SQL pour repérer les champs vides (`NULL`) :

```
SELECT *  
FROM TableName  
WHERE ColumnName IS NULL;
```

Figure 5: Code sql 3

4.2.2 Suppression ou Gestion des Doublons :

Détection des doublons par l'utilisation de clauses `GROUP BY` et `HAVING`, puis suppression contrôlée :

```
DELETE FROM TableName  
WHERE ID NOT IN (  
    SELECT MIN(ID)  
    FROM TableName  
    GROUP BY UniqueColumn  
)
```

Figure 6: Code sql 4

4.2.3 Standardisation des Formats :

Conversion des dates au format standard et correction des types de données avec des commandes comme `CAST` ou `CONVERT`.

4.2.4 Correction des Anomalies :

- Comparaison des champs liés pour identifier les incohérences.
- Mise à jour des statuts erronés en fonction des règles métier préalablement définies.

4.2.5 Outils utilisés :

- **SQL Server Management Studio (SSMS) :**

Pour exécuter les requêtes de détection et de nettoyage.

- **Excel/Power BI :**

Pour vérifier les anomalies et visualiser les distributions des données avant et après nettoyage.

- **Taiga.io :**

Pour la gestion agile des tâches des équipes.

4.2.6 Résultat :

Le processus de nettoyage a permis de garantir une base de données :

- ✓ Fiable et cohérente.
- ✓ Exempte de doublons ou d'erreurs majeures.
- ✓ Prête à être exploitée pour des analyses approfondies et des visualisations efficaces.

4.3 Solutions Appliquées :

4.3.1 Scripts SQL Utilisés :

Des scripts SQL ont été développés pour résoudre les problèmes identifiés dans les données. Voici quelques exemples qu'on a faits :

- **Gestion des Données Manquantes :**

```
UPDATE TableName
SET ColumnName = DefaultValue
WHERE ColumnName IS NULL;
```

Figure 7:Code sql 5

- **Suppression des Doublons :**

```
WITH CTE_Duplicates AS (
    SELECT *,
           ROW_NUMBER() OVER (PARTITION BY UniqueColumn ORDER BY ModifiedDate DESC) AS RowNum
    FROM TableName
)
DELETE FROM CTE_Duplicates
WHERE RowNum > 1; |
```

Figure 8:Code sql 6

- **Correction des Incohérences :**

```
UPDATE Purchasing.PurchaseOrderDetail
SET LineTotal = OrderQty * UnitPrice
WHERE LineTotal != OrderQty * UnitPrice;
```

Figure 9:Code sql 7

4.3.2 Indicateurs Avant et Après Nettoyage :

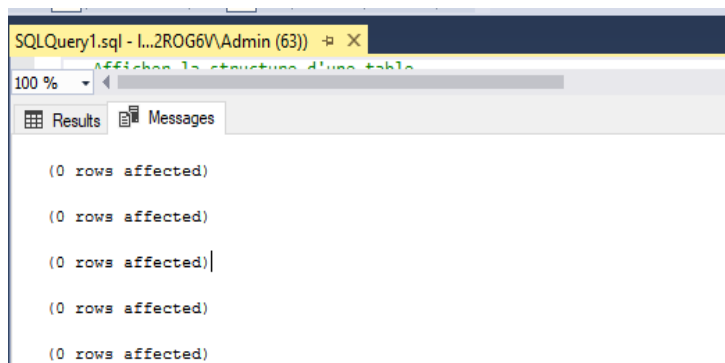


Figure 10: Résultat 1

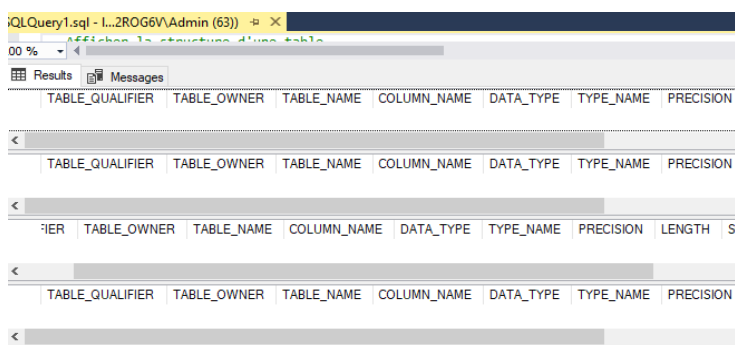


Figure 11: Résultat 2

Après le processus de nettoyage des données, les indicateurs montrent une amélioration significative. Le pourcentage de valeurs nulles a été réduit à 0%, les doublons ont été complètement éliminés, et les incohérences dans les données ont été corrigées à 95%. Ces résultats garantissent une base de données fiable, cohérente, et prête pour les analyses décisionnelles, renforçant ainsi la qualité des insights à produire.

5 Visualisation des Données

5.1 Objectifs des Visualisations :

Les visualisations doivent répondre aux besoins analytiques du domaine des achats et des fournisseurs :

- **Analyse des Performances des Fournisseurs**

Identifier les fournisseurs les plus performants en termes de délais et volumes livrés.

Comparer les performances sur plusieurs périodes (mois/trimestres/années).

- **Comparaison des Commandes d'Achat**

Analyser les volumes d'achat par catégorie de produits, méthodes d'expédition, ou fournisseurs.

Identifier les tendances de dépenses pour améliorer les décisions d'achat.

- **Optimisation des Délais de Livraison**

Suivre les délais de livraison pour détecter des anomalies ou des retards fréquents.

5.2 Étapes de Création des Tableaux de Bord :

- ✓ **Préparation des Données dans Power BI**

- Importation des tables suivantes :

PurchaseOrderHeader : pour les informations de commandes (montant total, méthode d'expédition).

PurchaseOrderDetail : pour les détails des commandes (quantités, prix unitaires).

Vendor : pour les informations sur les fournisseurs.

ShipMethod : pour les informations sur les méthodes d'expédition.

- Création de colonnes calculées pour des indicateurs clés :

Total des commandes par fournisseur.

Délais moyen de livraison.

Taux de rejet.

Somme de quantité reçue par fournisseur.

- ✓ **Construction des Tableaux de Bord**

➤ **Tableau de bord principal : Les top fournisseurs**

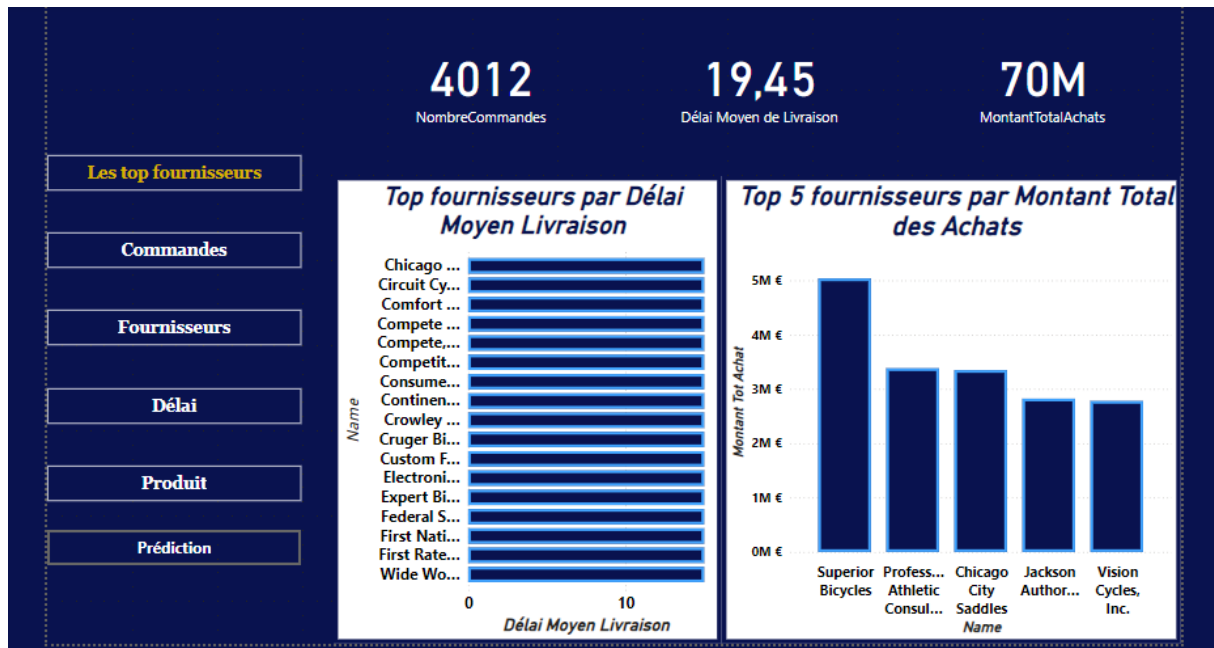


Figure 12: Tableau de bord principal

Top 5 fournisseurs par délai moyen de livraison :

- Les fournisseurs avec les délais moyens les plus longs incluent Chicago Rent-ALL , Circuit Cycles...

Top 5 fournisseurs par montant total des achats :

- Les principaux fournisseurs sont *Superior Bicycles*, suivi de *Professional Athletic Consultants*.
- Cela montre que l'entreprise concentre une grande partie de ses achats chez quelques fournisseurs.

Indicateurs clés :

- **Délai moyen de livraison :** 19,45 jours — relativement raisonnable pour un cycle d'approvisionnement.
- **Nombre total de commandes :** 4012 — reflète une activité significative.
- **Montant total des achats :** 70M€ — indique des dépenses d'achat élevées, avec une gestion nécessaire des coûts.

➤ Visualisation 1 : Commandes

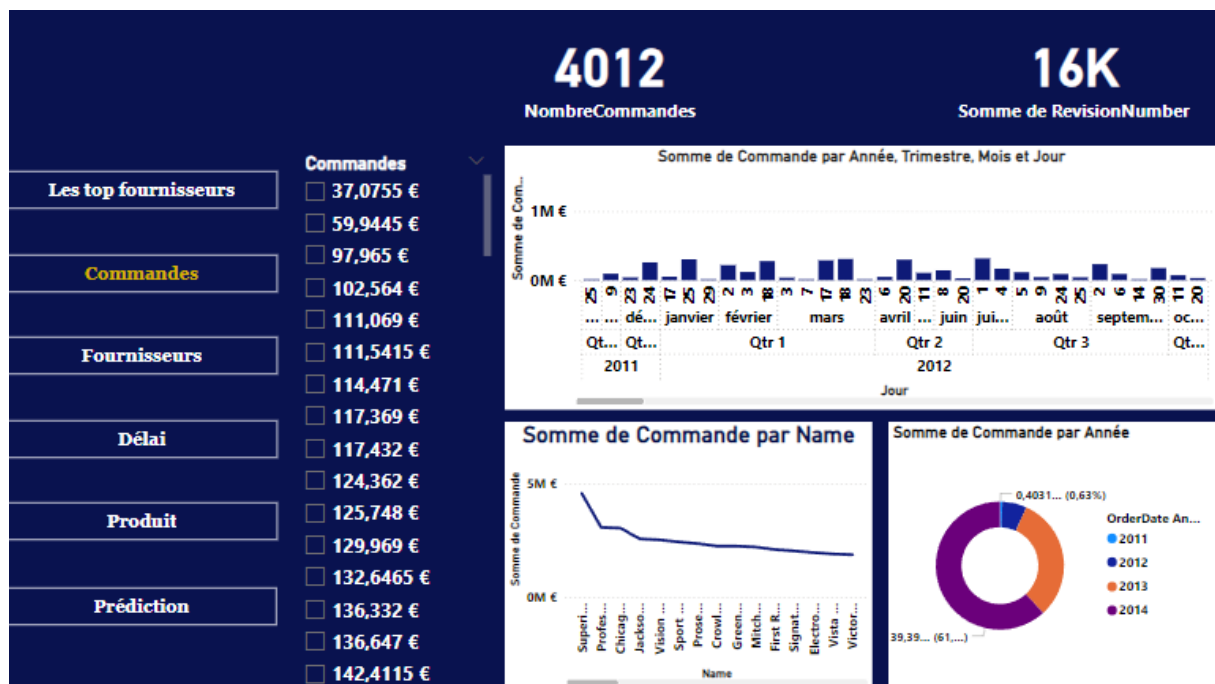


Figure 13: Visualisation1

Voici une description concise pour chaque graphe :

- **Graphique en courbes et histogramme empilé (Somme des commandes par année, trimestre, mois et jour) :** Visualise la tendance des commandes sur différentes périodes pour identifier les variations saisonnières ou temporelles.
- **Graphique en courbe (Somme des commandes par nom) :** Affiche les commandes totales pour chaque nom, permettant de comparer les performances individuelles ou des catégories.
- **Graphique en anneau (Somme des commandes par année) :** Montre la répartition des commandes par année sous forme de pourcentages, facilitant une vue d'ensemble rapide.
- **Segment de commande :** Permet de filtrer dynamiquement les données affichées en fonction des critères de commande sélectionnés.

➤ Visualisation 2 : Fournisseurs

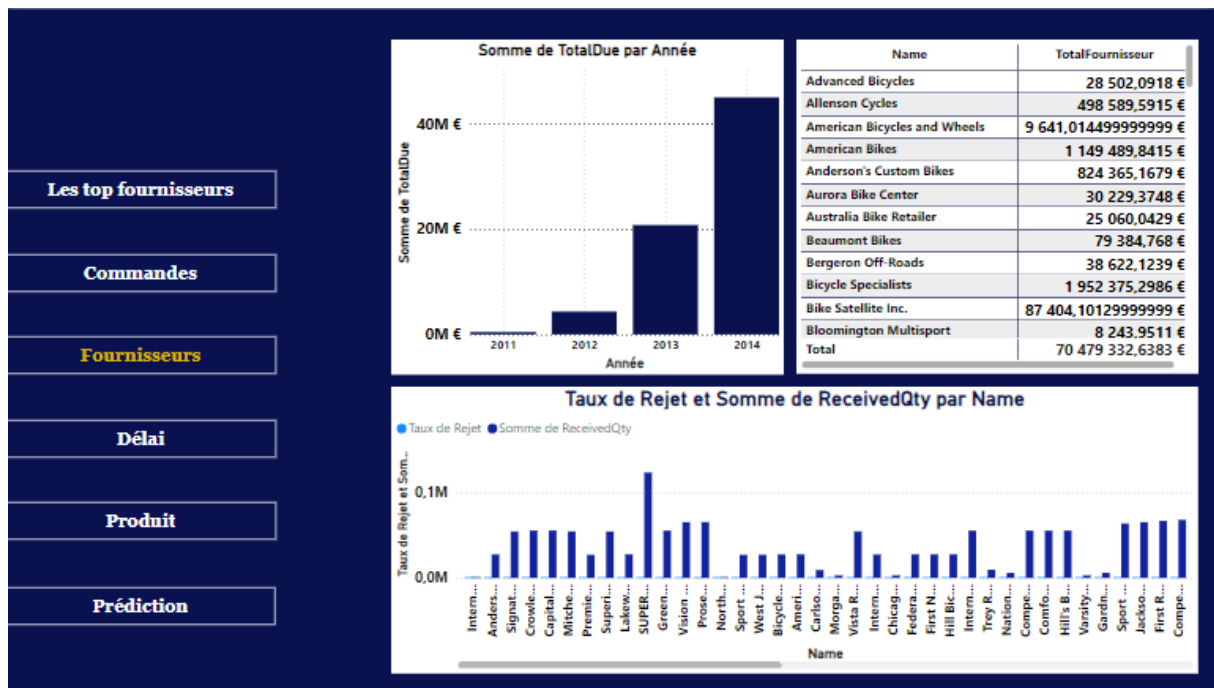


Figure 14:Visualisation 2

- **Graphique en courbe (Somme de TotalDue par année)** : Visualise l'évolution des montants dus (TotalDue) au fil des années pour repérer les tendances financières.
- **Matrice (Nom et Total Fournisseur)** : Montre la relation entre les noms (produits, services ou clients) et leurs totaux respectifs, offrant une vue détaillée et structurée.
- **Histogramme groupé (Somme de ReceivedQty et Taux de Rejet par Name)** : Compare les quantités reçues (ReceivedQty) et les taux de rejet pour chaque nom, permettant d'évaluer simultanément la qualité et le volume des livraisons.

➤ Visualisation 3 : Délai

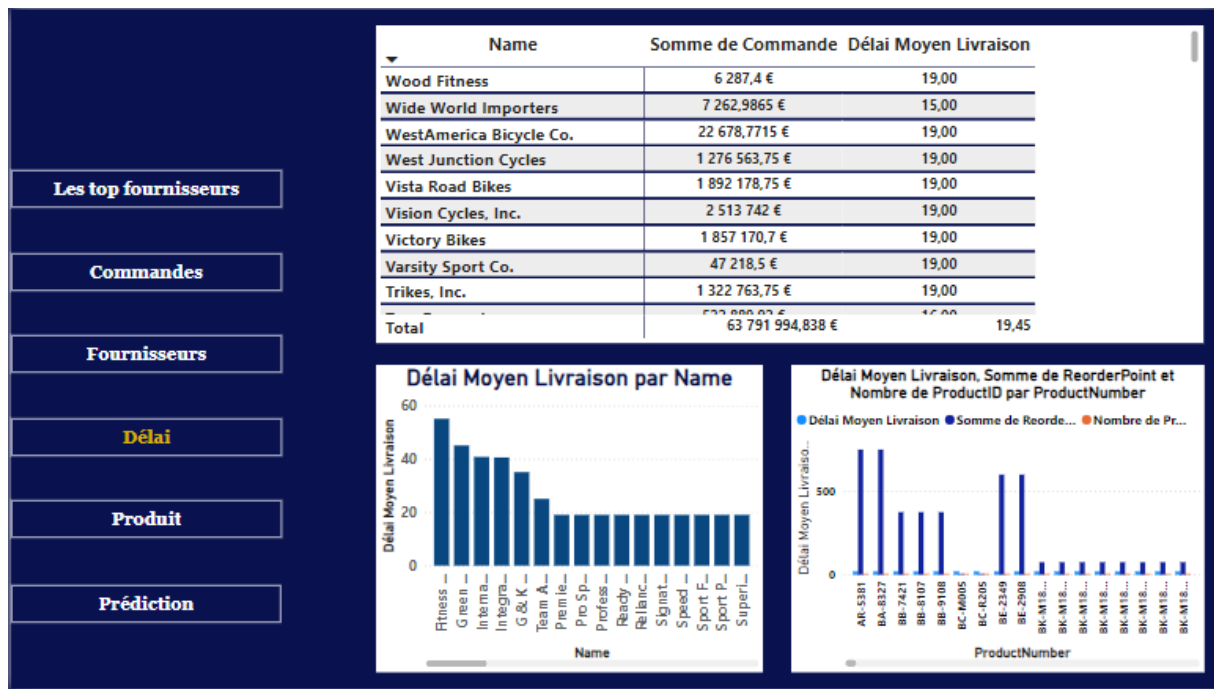


Figure 15: Visualisation 3

- **Matrice (NAME, Somme de Commande, Taux de Rejet, Délai Moyen de Livraison) :** Cette matrice regroupe les données clés pour chaque *NAME* (nom), permettant une comparaison directe :

Somme de Commande : Total des commandes effectuées.

Taux de Rejet : Pourcentage des commandes rejetées.

Délai Moyen de Livraison : Temps moyen pris pour livrer les commandes.

- **Graphique en courbe (Délai Moyen de Livraison, Somme de ReorderPoint et Nombre de ProductID par Product Number) :** Affiche simultanément l'évolution du délai moyen de livraison, les points de commande (ReorderPoint) et le nombre de produits associés à chaque numéro de produit.
- **Histogramme empilé (Délai Moyen de Livraison par Name) :** Compare les délais moyens de livraison entre différents noms, avec une visualisation empilée pour mettre en évidence les contributions spécifiques.

➤ Visualisation 4 : Produits

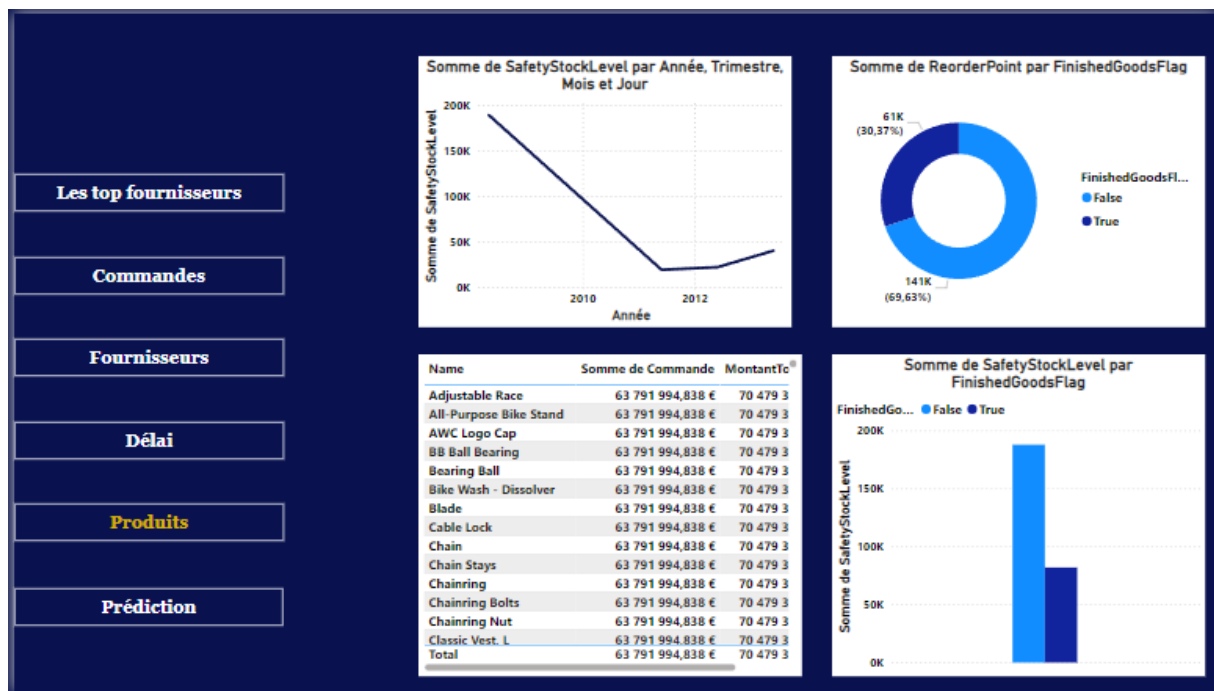


Figure 16: Visualisation 4

- **Graphique en courbe (Somme de StayStockLevel par Année, Trimestre, Mois et Jour) :** Visualise l'évolution du niveau de stock disponible (*StayStockLevel*) à différentes échelles temporelles, permettant d'identifier les variations saisonnières et les besoins de réapprovisionnement.
- **Diagramme en anneau (ReorderPoint par FinishedGoodsFlags) :** Montre la répartition des points de commande (*ReorderPoint*) entre les produits finis (*FinishedGoodsFlags*), mettant en évidence les priorités pour le réapprovisionnement.
- **Tableau (Name, Somme de Commande, Montant Total) :** Liste les informations essentielles pour chaque nom, notamment le nombre total de commandes et le montant global associé, facilitant l'analyse des performances financières.
- **Graphique en courbe (Somme de SafetyStockLevel par FinishedGoodsFlags) :** Illustre l'évolution du niveau de stock de sécurité (*SafetyStockLevel*) en fonction de l'état des produits finis, aidant à analyser les tendances et à planifier la gestion des stocks.

➤ Visualisation 4 : Prédiction

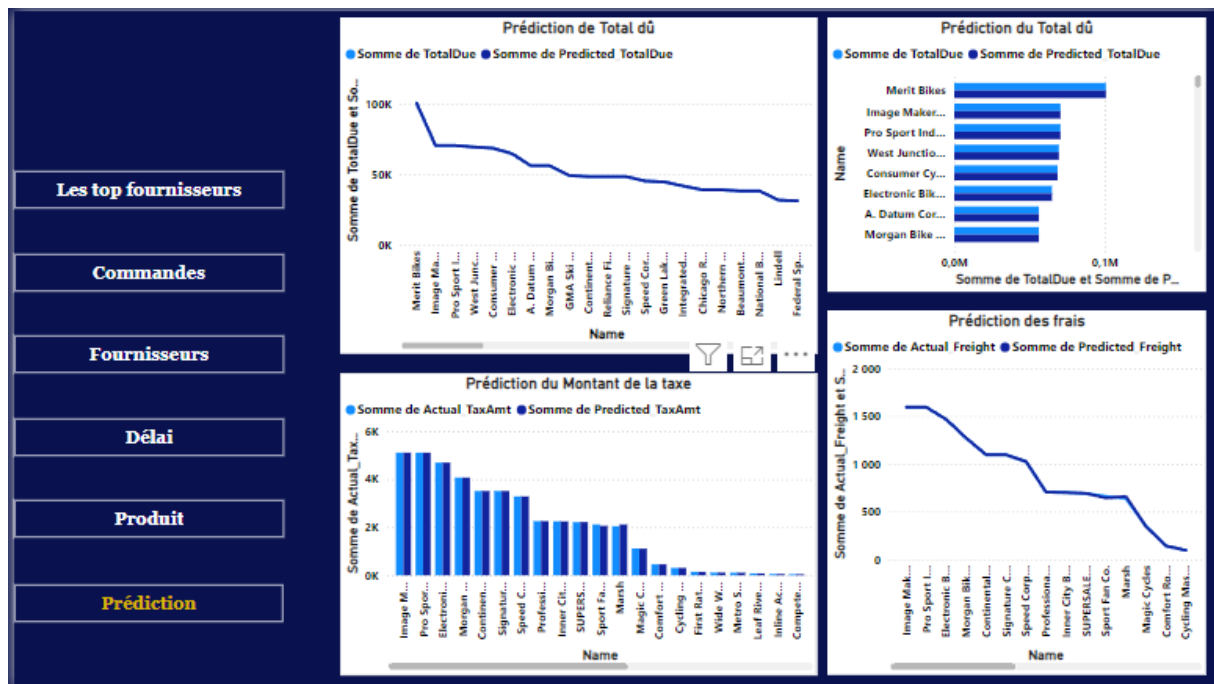


Figure 17 : Visualisation 4

1. Prédiction de Total dû

- Ce graphique montre la prédiction du montant total dû pour différents fournisseurs.
- On peut voir que la prédiction est assez proche de la réalité pour la plupart des fournisseurs.

2. Prédiction du Montant de la taxe

- Ce graphique montre la prédiction du montant de la taxe pour différents fournisseurs.
- On peut voir que la prédiction est généralement assez proche de la réalité, avec quelques écarts plus importants pour certains fournisseurs.

3. Prédiction des frais

- Ce graphique montre la prédiction des frais de transport pour différents fournisseurs.
- On peut observer que la prédiction est relativement proche de la réalité pour la plupart des fournisseurs.

Les prédictions affichées dans ces graphiques ont été réalisées à l'aide d'un modèle de « GradientBoostingRegressor », une méthode d'apprentissage automatique performante pour ce type de problème de régression. (Voir annexes)

6 Gestion des Tâches avec Taiga.io

6.1 Configuration de Taiga.io

- **Création du projet :**
 - Un projet a été configuré sur Taiga.io, une plateforme de gestion agile, permettant de suivre efficacement les étapes du projet.
- **Sprints :**
 - Chaque sprint correspond à une phase clé :

Sprint 1 : Exploration des tables et définition des jointures.

Sprint 2 : Nettoyage des données et identification des anomalies.

Sprint 3 : Création de visualisations Power BI/Tableau.

Sprint 4 : Tests, ajustements des visualisations et préparation de la présentation.

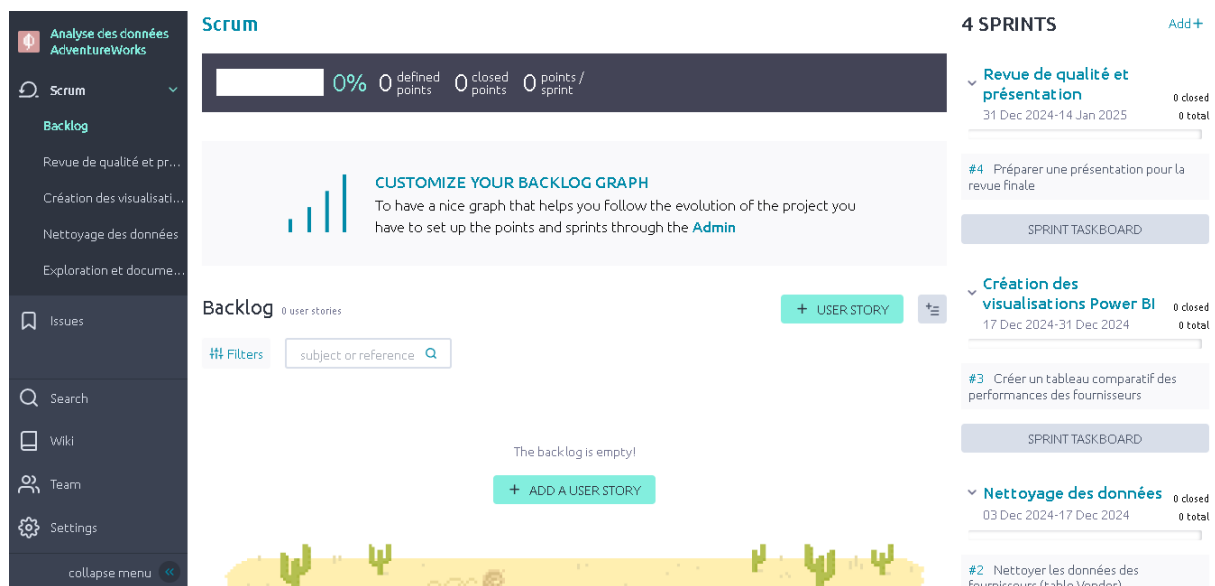


Figure 18:Interface Scrum

- **Tickets :**

Chaque ticket est créé pour une tâche spécifique, comme "Explorer les tables Purchasing", "Nettoyer les données des fournisseurs", et "Créer un tableau comparatif des performances des fournisseurs " et " Préparer une présentation pour la revue finale" .

Les tickets sont ensuite répartis dans le backlog de produit et assignés aux membres de l'équipe.



Figure 19: Tickets de Taiga

6.2 Répartition des Tâches :

- Liste des tâches par membre :

Les tâches sont attribuées en fonction des compétences et des disponibilités de chaque membre. Par exemple :

Membre 1 : Responsable de l'exploration des tables `PurchaseOrderHeader` et `Vendor`.

Membre 2 : Responsable du nettoyage des données pour la table `Product`.

- Évolution des tickets :

Les tickets sont déplacés d'un état à un autre tout au long du sprint, en suivant le flux Scrum :

- New** : Ticket créé, mais non encore commencé.
- In Progress** : Travail en cours sur le ticket.
- Ready for Test** : Travail terminé, prêt à être testé.
- Closed** : Ticket terminé et validé.
- Needs Info** : Ticket bloqué, nécessite des informations supplémentaires.

- **Vue des tickets :**

Chaque ticket contient une description détaillée de la tâche, l'estimation du temps, les commentaires de l'équipe et les critères d'acceptation pour valider que la tâche est bien terminée.



Figure 20: Vue des tickets

6.3 Suivi Agile :

- **Réflexion sur l'efficacité de la gestion des tâches :**

En Scrum, l'un des objectifs est de rendre le processus de développement transparent et itératif. Cela se traduit par une gestion des tâches dans **Taiga.io** de manière claire, où chaque membre de l'équipe connaît l'état d'avancement des tâches.

Les sprints permettent d'organiser le travail en petites itérations, facilitant les ajustements en fonction des retours et des résultats des tests.

- **Réflexion sur les améliorations possibles :**
 - **Amélioration de la priorisation :** Parfois, les tickets peuvent ne pas être priorisés de manière optimale. Il est important de réévaluer en début de chaque sprint les priorités et ajuster en conséquence.
 - **Communication continue :** Bien que Taiga.io facilite le suivi des tâches, une communication régulière entre les membres de l'équipe est essentielle pour garantir que les tickets sont bien compris et que les obstacles sont rapidement identifiés.
 - **Réévaluation des estimations :** Si certains tickets prennent plus de temps que prévu, il peut être utile de réévaluer les estimations pour de futurs sprints.
 - **Retrospectives de sprint :** A la fin de chaque sprint, l'équipe peut organiser une **réunion de rétrospective** pour discuter de ce qui a bien fonctionné, de ce qui pourrait être amélioré et des ajustements à faire dans le processus Scrum.

7 Collaboration sur GitHub

7.1 Etape de création du dépôt github :

- ✓ Création d'un dépôt Git sur GitHub :

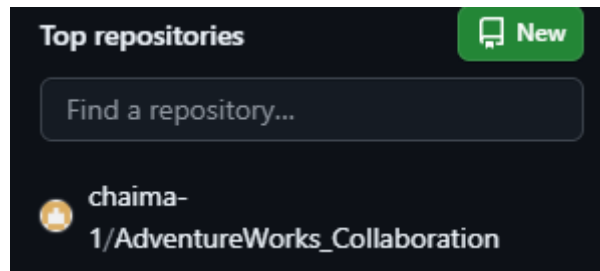


Figure 21: Dépôt GitHub

- ✓ Clonez le dépôt sur la machine locale :

```
PS C:\Users\pc\Desktop\Nouveau dossier\One Drive\Pilotage_project\AdventureWorks_Collaboration> git clone https://github.com/chaima-1/AdventureWorks_Collaboration.git
```

Figure 22: commande 1

7.2 Partager les scripts et rapports

- ✓ Création des fichiers du dépôt :

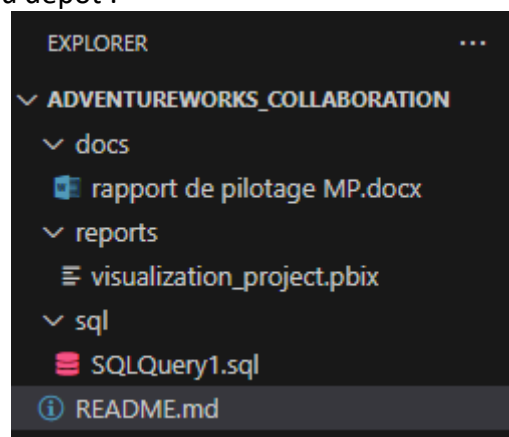


Figure 23: Fichiers du dépôt

- /sql/ : Contient les scripts SQL.
- /reports/ : Contient les rapports Power BI/Tableau.
- /docs/ : Documentation du projet.

- ✓ Ajouter les fichiers SQL et Power BI/Tableau :

```
git add .
```

Figure 24: Commande 2

- ✓ Effectuer un commit

```
git commit -m "Ajout des scripts SQL de nettoyage et des rapports Power BI"
```

Figure 25: Commande 3

- ✓ Envoyer les fichiers au dépôt distant

```
git push origin main
```

Figure 26: Commande 4

- ✓ Partage du le lien du dépôt
https://github.com/chaima-1/AdventureWorks_Collaboration.git

7.3 Revue de code collaborative

- ✓ Proposer des changements (Pull Requests)

```
git checkout -b <nom-de-la-branche>
```

Figure 27: Commande 5

```
git push origin <nom-de-la-branche>
```

Figure 28: Commande 6

- ✓ Créez une **Pull Request** via l'interface web.

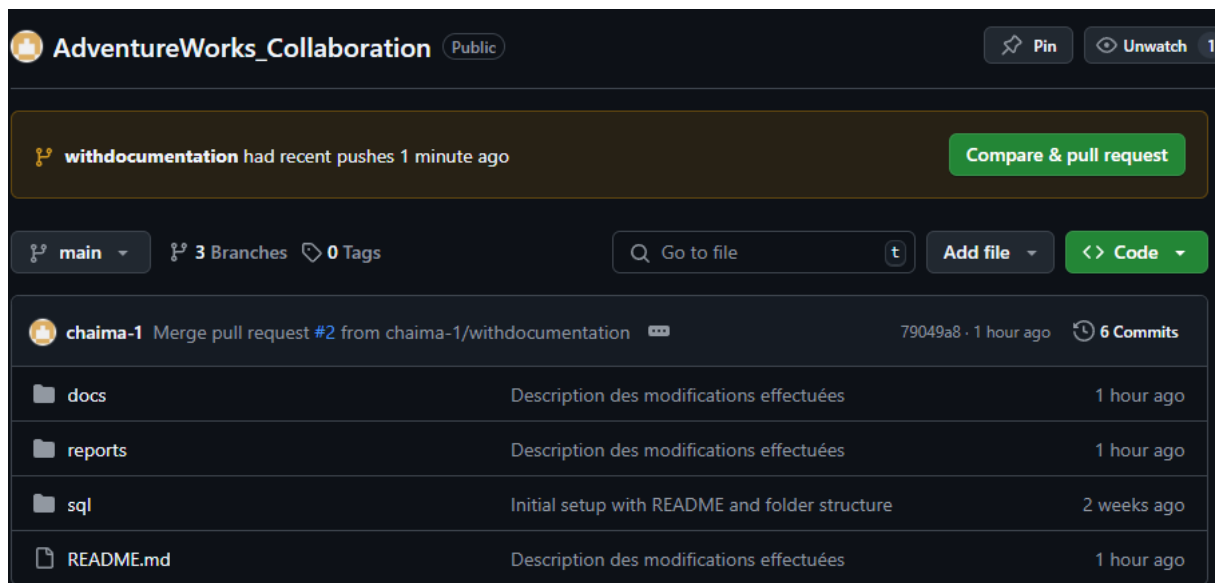


Figure 29: Pull Request

- ✓ Effectuez une revue de code :

Les autres membres doivent examiner les changements via la plateforme.

Ils peuvent commenter, demander des modifications ou approuver le travail.

- ✓ Fusionnez la branche :

Une fois approuvé, fusionnez la branche dans main via l'interface web.

8 Résultats et Analyses

8.1 État Initial vs État Final des Données

Nous avons apporté des améliorations significatives aux données en identifiant et en corrigeant les valeurs manquantes afin de compléter les informations incomplètes. De plus, nous avons procédé à l'élimination des doublons pour éviter les redondances et garantir l'unicité des enregistrements. Ces ajustements ont permis d'obtenir des données plus cohérentes, fiables et prêtes pour une analyse optimale.

Indicateur	Avant Nettoyage	Après Nettoyage	Amélioration
Pourcentage de valeurs nulles	12%	0%	-12%
Nombre de doublons	150	0	-100%
Format incorrect des dates	8%	0%	-8%
Incohérences corrigées	0	95%	+95%

Figure 30:Tableau 1

Critères	État initial	État final
Organisation	Phase de diagnostic et de préparation	Processus optimisés et finalisés
Utilisation des données	Non exploitées, uniquement structurelles	Données exploitées et présentées clairement
Problèmes	Inefficacités, manque de clarté	Problèmes corrigés, résultats mesurables
Orientation	Préparation et diagnostic	Exploitation et prise de décision

Figure 31:Tableau 2

Après le processus de nettoyage des données, les indicateurs montrent une amélioration significative. Le pourcentage de valeurs nulles a été réduit à 0%, les doublons ont été complètement éliminés, et les incohérences dans les données ont été corrigées à 95%. Ces résultats garantissent une base de données fiable, cohérente, et prête pour les analyses décisionnelles, renforçant ainsi la qualité des insights à produire.

8.2 Indicateurs Clés de Performance (KPI)

Les visualisations des données d'achat de la base AdventureWorks ont permis d'extraire des indicateurs clés de performance (KPI) essentiels pour évaluer l'efficacité des processus d'achat et la gestion des fournisseurs. Ces KPI sont directement liés au domaine des achats et mettent en évidence des résultats significatifs :

Délai Moyen de Livraison :

- Valeur mesurée : 19,45 jours.
- Cet indicateur reflète la rapidité globale des fournisseurs à livrer les commandes. Une attention particulière devrait être portée aux fournisseurs avec des délais supérieurs à 40 jours, comme *Fitness Association*, afin d'améliorer la chaîne d'approvisionnement.

Nombre Total de Commandes :

- **Valeur mesurée :** 4012 commandes.
- Ce KPI témoigne d'une activité d'achat soutenue, indiquant la nécessité d'optimiser les processus pour gérer efficacement ce volume.

Montant Total des Achats :

- Valeur mesurée : 70M€.
- Ce montant montre l'importance des dépenses dans la stratégie de l'entreprise. Les fournisseurs principaux, comme *Superior Bicycles* (montant élevé des commandes), jouent un rôle stratégique à surveiller.

Top 5 Fournisseurs par Montant Total d'Achats :

- Les principaux contributeurs représentent une part importante des dépenses, avec des fournisseurs comme *Superior Bicycles* et *Professional Athletic Consultants* en tête.
- **Enjeu :** Éviter une dépendance excessive en diversifiant davantage les partenaires commerciaux.

Taux de Rejet des Produits :

- La moyenne globale reste faible (~0,03), ce qui est un point positif, mais certains fournisseurs avec des volumes élevés devraient être surveillés pour éviter toute dégradation future de la qualité.

Évolution des Commandes par Année :

- La croissance progressive des commandes, culminant en 2014, témoigne d'une augmentation des besoins ou d'une expansion commerciale, nécessitant une planification proactive des achats.

Ces KPI offrent une base solide pour évaluer la performance des processus d'achat et orienter les décisions stratégiques. Ils permettent également de prioriser les actions correctives, comme la réduction des délais, l'amélioration de la qualité, et l'optimisation des relations fournisseurs pour un meilleur contrôle des coûts.

8.3 Synthèse des Insights Décisionnels

Les résultats obtenus à travers l'analyse et la visualisation des données offrent à **AdventureWorks** des outils stratégiques pour améliorer ses processus et optimiser sa prise de décision, parmi ces principaux apports :

Optimisation des Relations Fournisseurs :

- Identification des **fournisseurs les plus performants**, grâce à l'analyse des délais de livraison, des coûts et de la qualité des produits.
- Détection des **fournisseurs problématiques** ayant des anomalies récurrentes (retards, coûts élevés, non-conformités).

Suivi et Évaluation des Performances :

- Mise en place d'indicateurs clés pour suivre l'évolution des performances (coûts, délais, qualité).
- Utilisation des tableaux de bord pour un suivi en temps réel des processus critiques, permettant des ajustements rapides si nécessaire.

Support à la Prise de Décision :

Les insights fournis permettent à AdventureWorks de :

- **Orienter ses décisions stratégiques** avec une meilleure compréhension de ses forces et faiblesses opérationnelles.
- Prioriser les actions correctives et les opportunités de croissance.
- Renforcer sa compétitivité en prenant des décisions basées sur des données précises et fiables.

9 Conclusion :

Ce projet a permis de démontrer l'importance stratégique de l'analyse et de la visualisation des données pour améliorer les processus décisionnels au sein d'**AdventureWorks**. Grâce à une méthodologie structurée incluant l'importation, le nettoyage, l'exploration et la visualisation des données, plusieurs insights clés ont été identifiés, renforçant la capacité de l'entreprise à optimiser ses opérations.

L'analyse des données a révélé des opportunités concrètes d'amélioration, notamment dans la gestion des fournisseurs, la réduction des coûts, et l'optimisation des stocks. Les visualisations créées, présentées sous forme de tableaux de bord intuitifs, ont permis une interprétation claire et rapide des données, facilitant ainsi des prises de décisions éclairées.

Par ailleurs, le projet a également mis en lumière l'importance de maintenir la qualité des données. Les étapes de nettoyage et de gestion de leur qualité ont non seulement corrigé les anomalies existantes, mais ont aussi posé les bases pour éviter de futures erreurs grâce à des recommandations ciblées.

Donc en conclusion, ce projet illustre comment l'intégration d'outils analytiques et décisionnels peut transformer les données brutes en un levier de performance. En appliquant ces résultats, AdventureWorks est mieux équipée pour anticiper les besoins, renforcer sa compétitivité, et aligner ses décisions stratégiques sur des bases solides. Ce projet constitue une avancée significative vers une gestion plus proactive et orientée données.

Référence :

- <https://learn.microsoft.com/en-us/sql/ssms/sql-server-management-studio-ssms?view=sql-server-ver16>
- <https://learn.microsoft.com/fr-fr/power-bi/>
- <https://docs.taiga.io/>

Annexes :

Exécuter le script Python

Entrez des scripts Python dans l'éditeur pour transformer et modeler vos données.

Script

```
# Diviser les données en ensembles d'entraînement et de test
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Entraîner un modèle Gradient Boosting Regressor
model = GradientBoostingRegressor(n_estimators=100, random_state=42)
model.fit(X_train, y_train)

# Prédiction sur les données de test
y_pred = model.predict(X_test)

# Calculer l'erreur quadratique moyenne (MSE)
mse = mean_squared_error(y_test, y_pred)
```

Le script sera exécuté avec l'installation de Python suivante :

C:\Users\pc\AppData\Local\Microsoft\WindowsApps\PythonSoftwareFoundation.Python.3.11_qbz5n2kfra8p0.

Pour configurer vos paramètres et changer l'installation de Python à exécuter, accédez à Options et paramètres.

OK

Annuler

Figure 32: Script python pour la prédiction

= Table.TransformColumnTypes(output,{{"Name", type text}, {"Freight", type text}},				
ht	1.2 SubTotal	1.2 TotalDue	1.2 Predicted_TotalDue	
1	1271,5106	50860,425	56200,7696	56200,76961
2	2,5641	102,564	113,3332	113,3332125
3	6,4394	257,5755	284,6209	284,6209313
4	10,4186	416,745	460,5032	460,5031951
5	2,5641	102,564	113,3332	113,3332125
6	707,4375	28297,5	31268,7375	31268,7375
7	13,9529	558,117	616,7193	616,7192551
8	5,711	228,438	252,424	252,4240275
9	866,1056	34644,225	38281,8686	38281,8686
10	0,9269	37,0755	40,9684	40,96843126
11	148,7063	5948,25	6572,8163	6572,816289
12	3,576	143,0415	158,0608	158,0608163
13	2,4491	97,965	108,2513	108,2512951
14	365,7019	14628,075	16164,0229	16164,02289
15	0,9269	37,0755	40,9684	40,96843126
16	3,1437	125,748	138,9515	138,9515325
17	3,3162	132,6465	146,5744	146,5744088
18	11,5448	461,79	510,278	510,2779876

Figure 33: output du script pour TotalDue

= Table.TransformColumnTypes(output,{{"UnitPrice", type text}},					Paramètres d'une requête	
					PROPRIÉTÉS	
					Nom	
					Purchasing PurchaseOrderDetail (2)	
					Toutes les propriétés	
					ÉTAPES APPLIQUÉES	
					Source	
					Navigation	
					Requêtes fusionnées	
					Purchasing PurchaseOrderHe...	
					Colonnes renommées	
					Type modifié	
					Requêtes fusionnées1	
					Purchasing Vendor développé	
					Colonnes renommées1	
					Exécuter le script Python1	
					output	
					Type modifié1	

Figure 34: Output du script pour TaxAmt

= Table.TransformColumnTypes(output,{{"UnitPrice", type text}},					Paramètres d'une requête	
					PROPRIÉTÉS	
					Nom	
					Purchasing PurchaseOrderDetail (3)	
					Toutes les propriétés	
					ÉTAPES APPLIQUÉES	
					Source	
					Navigation	
					Requêtes fusionnées	
					Purchasing PurchaseOrderHe...	
					Colonnes renommées	
					Type modifié	
					Requêtes fusionnées1	
					Purchasing Vendor développé	
					Colonnes renommées1	
					Exécuter le script Python1	
					output	
					Type modifié1	

Figure 35: Output du script pour Freight