



# Rapport de Projet

## *Exploration et Analyse de AdventureWorks*

### Informations :

Nom et Prénom : ELOUAFI Hajar

Nom et Prénom : DIRANE Hind

Module : Pilotage de la performance et tableaux de bord

Enseignant : Pr. BE.ELBAGHAZAOUI

Année universitaire : 2024/2025

# 1. Introduction

## Contexte

Ce projet a pour objectif de nettoyer, analyser, et visualiser les données relatives à la production et à la gestion des stocks dans le cadre de l'entreprise fictive *AdventureWorks*. L'objectif principal est d'améliorer la prise de décision stratégique à travers la gestion efficace des données de production et de stock. L'équipe 2 se concentre spécifiquement sur l'analyse des tables liées aux produits et à la production, telles que `Product`, `ProductCategory`, et `ProductionOrder`, afin d'identifier les anomalies, effectuer un nettoyage des données, et produire des visualisations pertinentes pour aider à la prise de décision.

## Objectifs du rapport

Ce rapport couvre les différentes étapes suivies par l'équipe pour mener à bien la mission qui lui a été assignée. Il documente l'état initial des données, décrit les étapes de nettoyage et de transformation, et fournit un résumé des phases du projet ainsi que des livrables associés.

Les objectifs spécifiques de ce rapport sont les suivants :

- Documenter l'état initial des données : Une présentation de l'état brut des données, avec les problèmes identifiés (doublons, valeurs manquantes, formats incorrects, etc.).
- Décrire les étapes du nettoyage et des transformations : Les méthodes et outils utilisés pour nettoyer les données et effectuer les transformations nécessaires pour les rendre exploitables.
- Fournir un résumé des phases et des livrables : Un aperçu des différentes étapes du projet et des résultats obtenus, y compris les visualisations finales créées avec Power BI ou Tableau.

# Innehåll

<b>1</b>	<b>État des Données</b>	<b>4</b>
<b>2</b>	<b>Nettoyage des Données</b>	<b>19</b>
2.1	Méthodologie utilisée . . . . .	19
2.2	Étapes spécifiques . . . . .	20
<b>3</b>	<b>Exportation des Données Nettoyées</b>	<b>24</b>
3.1	Choix du Formats . . . . .	24
3.2	Étapes Principales . . . . .	24
<b>4</b>	<b>Visualisation des Données</b>	<b>24</b>
4.1	Dashboard 1 : Page d'Accueil . . . . .	24
4.2	Dashboard 2 : Production et Performance . . . . .	25
4.2.1	Détails des KPI . . . . .	26
4.2.2	Détails des Diagrammes . . . . .	27
4.3	Dashboard 3 : Gestion des Stocks . . . . .	27
4.3.1	Détails des KPI . . . . .	28
4.3.2	Détails des Diagrammes . . . . .	29
4.4	Dashboard 4 : KPI Financiers . . . . .	30
4.4.1	Détails des KPI . . . . .	30
4.4.2	Détails des Diagrammes . . . . .	32
4.5	Dashboard 5 : Feedback . . . . .	32
4.5.1	Détails des KPI . . . . .	33
4.5.2	Détails des Diagrammes . . . . .	34
4.6	Dashboard 6 : Prédictions . . . . .	35
<b>5</b>	<b>Phases du projet</b>	<b>35</b>
5.1	Sprint 1 : Exploration des données . . . . .	36
5.2	Sprint 2: Nettoyage des données . . . . .	38
5.3	Sprint 3 : Création de visualisations . . . . .	39
5.4	Sprint 4 : Revue et Finilisation . . . . .	40
<b>6</b>	<b>Recommandations futures pour le projet de gestion de la production</b>	<b>41</b>
<b>7</b>	<b>Conclusion</b>	<b>43</b>
<b>8</b>	<b>Annexe</b>	<b>44</b>

# 1 État des Données

## Description des données

Dans le cadre de ce projet, l'équipe 2 s'est concentrée sur l'analyse de la table `BillOfMaterials`, extraite de la base de données *AdventureWorks2019*. Cette table contient 504 lignes et 9 colonnes, et elle joue un rôle central dans la gestion des matériaux nécessaires à la fabrication des produits.

Caractéristiques principales de la table `BillOfMaterials`:

- Nom de la table : `BillOfMaterials`
- Nombre de lignes : 2679
- Nombre de colonnes : 9
- Description des colonnes :

	Nom de la colonne	Type de données
1	<code>BillOfMaterialsID</code>	int
2	<code>ProductAssemblyID</code>	int
3	<code>ComponentID</code>	int
4	<code>StartDate</code>	datetime
5	<code>EndDate</code>	datetime
6	<code>UnitMeasureCode</code>	nchar
7	<code>BOMLevel</code>	smallint
8	<code>PerAssemblyQty</code>	decimal

Figur 1: Les colonnes de `BillOfMaterials`

Caractéristiques principales de la table `Culture`:

- Nom de la table : `Culture`
- Nombre de lignes : 8
- Nombre de colonnes : 3
- Description des colonnes :

	Nom de la colonne	Type de données
1	CultureID	nchar
2	Name	nvarchar
3	ModifiedDate	datetime

Figur 2: Les colonnes de Culture

Caractéristiques principales de la table Document:

- Nom de la table : Document
- Nombre de lignes : 13
- Nombre de colonnes : 14
- Description des colonnes :

	Nom de la colonne	Type de données
1	DocumentNode	hierarchyid
2	DocumentLevel	smallint
3	Title	nvarchar
4	Owner	int
5	FolderFlag	bit
6	FileName	nvarchar
7	FileExtension	nvarchar
8	Revision	nchar
9	ChangeNumber	int
10	Status	tinyint

Figur 3: Les colonnes de Document

Caractéristiques principales de la table Illustration:

- Nom de la table : Illustration
- Nombre de lignes : 5
- Nombre de colonnes : 3
- Description des colonnes :

	Nom de la colonne	Type de données
1	IllustrationID	int
2	Diagram	xml
3	ModifiedDate	datetime

Figur 4: Les colonnes de Illustration

#### Caractéristiques principales de la table Location:

- Nom de la table : Location
- Nombre de lignes : 14
- Nombre de colonnes : 5
- Description des colonnes :

	Nom de la colonne	Type de données
1	LocationID	smallint
2	Name	nvarchar
3	CostRate	smallmoney
4	Availability	decimal
5	ModifiedDate	datetime

Figur 5: Les colonnes de Location

#### Caractéristiques principales de la table Product:

- Nom de la table : Product
- Nombre de lignes : 504
- Nombre de colonnes : 25
- Description des colonnes :

	Nom de la colonne	Type de données
1	ProductID	int
2	Name	nvarchar
3	ProductNumber	nvarchar
4	MakeFlag	bit
5	FinishedGoodsFlag	bit
6	Color	nvarchar
7	SafetyStockLevel	smallint
8	ReorderPoint	smallint
9	StandardCost	money
10	ListPrice	money
11	Size	nvarchar
12	SizeUnitMeasureCode	nchar
13	WeightUnitMeasureCode	nchar
14	Weight	decimal
15	DaysToManufacture	int
16	ProductLine	nvarchar
17	Class	nvarchar
18	Style	nvarchar
19	ProductSubcategoryID	int
20	ProductModelID	int
21	SellStartDate	datetime
22	SellEndDate	datetime
23	DiscontinuedDate	datetime
24	rowguid	uniqueidentifier
25	ModifiedDate	datetime

Figur 6: Les colonnes de Product

Caractéristiques principales de la table ProductCategory:

- Nom de la table : ProductCategory
- Nombre de lignes : 4
- Nombre de colonnes : 4
- Description des colonnes :

	Nom de la colonne	Type de données
1	ProductCategoryID	int
2	Name	nvarchar
3	rowguid	uniqueidentifier
4	ModifiedDate	datetime

Figur 7: Les colonnes de ProductCategory

\*

#### Caractéristiques principales de la table ProductCostHistory:

- Nom de la table : ProductCostHistory
- Nombre de lignes : 395
- Nombre de colonnes : 5
- Description des colonnes :

	Nom de la colonne	Type de données
1	ProductID	int
2	StartDate	datetime
3	EndDate	datetime
4	StandardCost	money
5	ModifiedDate	datetime

Figur 8: Les colonnes de ProductCostHistory

#### Caractéristiques principales de la table ProductDescription:

- Nom de la table : ProductDescription
- Nombre de lignes : 762
- Nombre de colonnes : 4
- Description des colonnes :

	Nom de la colonne	Type de données
1	ProductDescriptionID	int
2	Description	nvarchar
3	rowguid	uniqueidentifier
4	ModifiedDate	datetime

Figur 9: Les colonnes de ProductDescription



#### Caractéristiques principales de la table ProductDocument:

- Nom de la table : ProductDocument
- Nombre de lignes : 32
- Nombre de colonnes : 3
- Description des colonnes :

	Nom de la colonne	Type de données
1	ProductID	int
2	DocumentNode	hierarchyid
3	ModifiedDate	datetime

Figur 10: Les colonnes de ProductDocument

#### Caractéristiques principales de la table ProductInventory:

- Nom de la table : ProductInventory
- Nombre de lignes : 1069
- Nombre de colonnes : 7
- Description des colonnes :

	Nom de la colonne	Type de données
1	ProductID	int
2	LocationID	smallint
3	Shelf	nvarchar
4	Bin	tinyint
5	Quantity	smallint
6	rowguid	uniqueidentifier
7	ModifiedDate	datetime

Figur 11: Les colonnes de ProductInventory

#### Caractéristiques principales de la table ProductListPriceHistory:

- Nom de la table : ProductListPriceHistory
- Nombre de lignes : 395
- Nombre de colonnes : 5
- Description des colonnes :

	Nom de la colonne	Type de données
1	ProductID	int
2	StartDate	datetime
3	EndDate	datetime
4	ListPrice	money
5	ModifiedDate	datetime

Figur 12: Les colonnes de ProductListPriceHistory

#### Caractéristiques principales de la table ProductModel:

- Nom de la table : ProductModel
- Nombre de lignes : 128
- Nombre de colonnes : 6
- Description des colonnes :

	Nom de la colonne	Type de données
1	ProductModelID	int
2	Name	nvarchar
3	CatalogDescription	xml
4	Instructions	xml
5	rowguid	uniqueidentifier
6	ModifiedDate	datetime

Figur 13: Les colonnes de ProductModel

#### Caractéristiques principales de la table ProductModelIllustration:

- Nom de la table : ProductModelIllustration
- Nombre de lignes : 7
- Nombre de colonnes : 3
- Description des colonnes :

	Nom de la colonne	Type de données
1	ProductModelID	int
2	IllustrationID	int
3	ModifiedDate	datetime

Figur 14: Les colonnes de ProductModelIllustration

**Caractéristiques principales de la table ProductModelProductDescriptionCulture:**

- Nom de la table : ProductModelProductDescriptionCulture
- Nombre de lignes : 762
- Nombre de colonnes : 4
- Description des colonnes :

	Nom de la colonne	Type de données
1	ProductModelID	int
2	ProductDescriptionID	int
3	CultureID	nchar
4	ModifiedDate	datetime

Figur 15: Les colonnes de ProductModelProductDescriptionCulture

**Caractéristiques principales de la table ProductPhoto:**

- Nom de la table : ProductPhoto
- Nombre de lignes : 101
- Nombre de colonnes : 6
- Description des colonnes :

	Nom de la colonne	Type de données
1	ProductPhotoID	int
2	ThumbNailPhoto	varbinary
3	ThumbnailPhoto...	nvarchar
4	LargePhoto	varbinary
5	LargePhotoFileN...	nvarchar
6	ModifiedDate	datetime

Figur 16: Les colonnes de ProductPhoto

Caractéristiques principales de la table ProductProductPhoto:

- Nom de la table : ProductProductPhoto
- Nombre de lignes : 504
- Nombre de colonnes : 4
- Description des colonnes :

	Nom de la colonne	Type de données
1	ProductID	int
2	ProductPhotoID	int
3	Primary	bit
4	ModifiedDate	datetime

Figur 17: Les colonnes de ProductProductPhoto

Caractéristiques principales de la table ProductReview:

- Nom de la table : ProductReview
- Nombre de lignes : 4
- Nombre de colonnes : 8
- Description des colonnes :

	Nom de la colonne	Type de données
1	ProductReviewID	int
2	ProductID	int
3	ReviewerName	nvarchar
4	ReviewDate	datetime
5	EmailAddress	nvarchar
6	Rating	int
7	Comments	nvarchar
8	ModifiedDate	datetime

Figur 18: Les colonnes de ProductReview

Caractéristiques principales de la table ProductSubCategory:

- Nom de la table : ProductSubCategory
- Nombre de lignes : 37
- Nombre de colonnes : 5
- Description des colonnes :

	Nom de la colonne	Type de données
1	ProductSubcategoryID	int
2	ProductCategoryID	int
3	Name	nvarchar
4	rowguid	uniqueidentifier
5	ModifiedDate	datetime

Figur 19: Les colonnes de ProductSubCategory

Caractéristiques principales de la table ScrapReason:

- Nom de la table : ScrapReason
- Nombre de lignes : 101
- Nombre de colonnes : 6
- Description des colonnes :

	Nom de la colonne	Type de données
1	ScrapReasonID	smallint
2	Name	nvarchar
3	ModifiedDate	datetime

Figur 20: Les colonnes de ScrapReason

**Caractéristiques principales de la table TransactionHistory:**

- Nom de la table : TransactionHistory
- Nombre de lignes : 113443
- Nombre de colonnes : 9
- Description des colonnes :

	Nom de la colonne	Type de données
1	TransactionID	int
2	ProductID	int
3	ReferenceOrderID	int
4	ReferenceOrder...	int
5	TransactionDate	datetime
6	TransactionType	nchar
7	Quantity	int
8	ActualCost	money

Figur 21: Les colonnes de TransactionHistory

**Caractéristiques principales de la table TransactionHistoryArchive:**

- Nom de la table : TransactionHistoryArchive
- Nombre de lignes : 89253
- Nombre de colonnes : 9
- Description des colonnes :

	Nom de la colonne	Type de données
2	ProductID	int
3	ReferenceOrderID	int
4	ReferenceOrder...	int
5	TransactionDate	datetime
6	TransactionType	nchar
7	Quantity	int
8	ActualCost	money
9	ModifiedDate	datetime

Figur 22: Les colonnes de TransactionHistoryArchive

#### Caractéristiques principales de la table UnitMeasure:

- Nom de la table : UnitMeasure
- Nombre de lignes : 39
- Nombre de colonnes : 3
- Description des colonnes :

	Nom de la colonne	Type de données
1	UnitMeasureCode	nchar
2	Name	nvarchar
3	ModifiedDate	datetime

Figur 23: Les colonnes de UnitMeasure

#### Caractéristiques principales de la table vProductAndDescription:

- Nom de la table : vProductAndDescription
- Nombre de lignes : 1764
- Nombre de colonnes : 5
- Description des colonnes :

	Nom de la colonne	Type de données
1	ProductID	int
2	Name	nvarchar
3	ProductModel	nvarchar
4	CultureID	nchar
5	Description	nvarchar

Figur 24: Les colonnes de vProductAndDescription

Caractéristiques principales de la table vProductModelCatalogDescription:

- Nom de la table : vProductModelCatalogDescription
- Nombre de lignes : 6
- Nombre de colonnes : 5
- Description des colonnes :



	Nom de la colonne	Type de données
1	ProductModelID	int
2	Name	nvarchar
3	Summary	nvarchar
4	Manufacturer	nvarchar
5	Copyright	nvarchar
6	ProductURL	nvarchar
7	WarrantyPeriod	nvarchar
8	WarrantyDescrip...	nvarchar
9	NoOfYears	nvarchar
10	MaintenanceDes...	nvarchar
11	Wheel	nvarchar
12	Saddle	nvarchar
13	Pedal	nvarchar
14	BikeFrame	nvarchar
15	Crankset	nvarchar
16	PictureAngle	nvarchar
17	PictureSize	nvarchar
18	ProductPhotoID	nvarchar
19	Material	nvarchar
20	Color	nvarchar
21	ProductLine	nvarchar
22	Style	nvarchar
23	RiderExperience	nvarchar
24	rowguid	uniqueidentifier
25	ModifiedDate	datetime

Figur 25: Les colonnes de vProductModelCatalogDescription

Caractéristiques principales de la table vProductModelInstruction:

- Nom de la table : vProductModelInstruction
- Nombre de lignes : 131
- Nombre de colonnes : 11
- Description des colonnes :

	Nom de la colonne	Type de données
1	ProductModelID	int
2	Name	nvarchar
3	Instructions	nvarchar
4	LocationID	int
5	SetupHours	decimal
6	MachineHours	decimal
7	LaborHours	decimal
8	LotSize	int
9	Step	nvarchar
10	rowguid	uniqueidentifier
11	ModifiedDate	datetime

Figur 26: Les colonnes de vProductModelInstruction

#### Caractéristiques principales de la table WorkOrder:

- Nom de la table : WorkOrder
- Nombre de lignes : 72591
- Nombre de colonnes : 10
- Description des colonnes :

	Nom de la colonne	Type de données
1	WorkOrderID	int
2	ProductID	int
3	OrderQty	int
4	StockedQty	int
5	ScrappedQty	smallint
6	StartDate	datetime
7	EndDate	datetime
8	DueDate	datetime
9	ScrapReasonID	smallint
10	ModifiedDate	datetime

Figur 27: Les colonnes de WorkOrder

### Caractéristiques principales de la table WorkOrderRouting:

- Nom de la table : WorkOrderRouting
- Nombre de lignes : 67131
- Nombre de colonnes : 12
- Description des colonnes :

	Nom de la colonne	Type de données
1	WorkOrderID	int
2	ProductID	int
3	OperationSeque...	smallint
4	LocationID	smallint
5	ScheduledStart...	datetime
6	ScheduledEndD...	datetime
7	ActualStartDate	datetime
8	ActualEndDate	datetime
9	ActualResource...	decimal
10	PlannedCost	money
11	ActualCost	money
12	ModifiedDate	datetime

Figur 28: Les colonnes de WorkOrderRouting

## 2 Nettoyage des Données

Le nettoyage des données est une étape cruciale dans tout processus d'analyse de données. Dans cette section, nous détaillons les méthodes et les techniques utilisées pour préparer les données avant leur analyse. Le nettoyage a pour objectif d'améliorer la qualité des données en éliminant les incohérences et en assurant leur conformité. Les principales techniques utilisées sont la suppression des doublons, l'imputation des valeurs manquantes et la transformation ou normalisation des données. Ces étapes permettent d'optimiser la précision des analyses futures.

### 2.1 Méthodologie utilisée

Le nettoyage des données s'est effectué en plusieurs étapes clés :

- **Suppression des doublons** : Les doublons ont été identifiés et supprimés pour éviter des biais dans les résultats.
- **Imputation des valeurs manquantes** : Les valeurs manquantes ont été imputées en utilisant des méthodes appropriées pour chaque type de donnée.

## 2.2 Étapes spécifiques

Les étapes suivantes ont été suivies lors du nettoyage des données :

	ProductID	Name	ProductNumber	MakeFlag	FinishedGoodsFlag	Color	SafetyStockLevel	ReorderPoint	StandardCost	ListPrice	Size	SizeL
1	1	Adjustable Race	AR-5381	0	0	NULL	1000	750	0,00	0,00	NULL	NULL
2	2	Bearing Ball	BA-8327	0	0	NULL	1000	750	0,00	0,00	NULL	NULL
3	3	BB Ball Bearing	BE-2349	1	0	NULL	800	600	0,00	0,00	NULL	NULL
4	4	Headset Ball Bearings	BE-2908	0	0	NULL	800	600	0,00	0,00	NULL	NULL
5	316	Blade	BL-2036	1	0	NULL	800	600	0,00	0,00	NULL	NULL
6	317	LL Crankarm	CA-5965	0	0	Black	500	375	0,00	0,00	NULL	NULL
7	318	ML Crankarm	CA-6738	0	0	Black	500	375	0,00	0,00	NULL	NULL
8	319	HL Crankarm	CA-7457	0	0	Black	500	375	0,00	0,00	NULL	NULL

Figur 29: La table Production Product avant nettoyage

Le processus de nettoyage des données a débuté par l'identification et le traitement des valeurs nulles dans les colonnes clés du jeu de données. En particulier, les colonnes **Color**, **Weight** et **Size** ont été ciblées. Afin de préserver l'intégrité et la cohérence des données, des valeurs par défaut ont été attribuées aux enregistrements où des valeurs manquantes ont été détectées :

- **Mise à jour des valeurs nulles dans la colonne Color** : Les valeurs manquantes ont été remplacées par la valeur par défaut 'Unknown' pour maintenir la cohérence des informations.
- **Mise à jour des valeurs nulles dans la colonne Weight** : Les valeurs nulles ont été remplacées par 1, une valeur par défaut indiquant l'absence de données spécifiques.
- **Mise à jour des valeurs nulles dans la colonne Size** : Les valeurs nulles ont été remplacées par 'N/A', un standard permettant d'indiquer que la taille n'a pas été précisée.

Ces mises à jour ont permis de traiter efficacement les valeurs manquantes et d'assurer que les analyses suivantes soient basées sur un jeu de données complet et cohérent.

	TotalRows	NullColors	NullWeights	NullSizes
1	504	248	299	293

Figur 30: Colors/Weight/Sizes avant nettoyage

	TotalRows	NullColors	NullWeights	NullSizes
1	504	0	0	0

Figur 31: Colors/Weight/Sizes après nettoyage

Dans le cadre du nettoyage des données, nous avons également effectué deux analyses distinctes pour détecter et traiter les doublons ainsi que les incohérences dans les dates.

- **Détection des doublons** : Une analyse approfondie a permis d'identifier les doublons dans les colonnes `Name` et `ProductNumber`, ce qui peut fausser les calculs ou les rapports. Les doublons ont été supprimés afin d'assurer l'unicité des produits dans le jeu de données.
- **Vérification des incohérences dans les dates** : Une vérification a été réalisée sur les colonnes `SellStartDate` et `SellEndDate` pour repérer les enregistrements où la date de début de vente était après la date de fin, ou bien lorsque la date de début était manquante. Ces incohérences ont été corrigées afin de garantir la validité des données temporelles.

Name	ProductNumber	DuplicateCount
ProductID	SellStartDate	SellEndDate

Figur 32: Résultats des doublons et des incohérences des dates

Dans le cadre du nettoyage et de la validation des données, plusieurs autres analyses ont été réalisées pour identifier et résoudre des problèmes potentiels :

- **Vérification des sous-catégories manquantes** : Une analyse a permis d'identifier les produits sans sous-catégorie associée. Ces cas ont été étudiés pour garantir l'intégrité des relations entre les produits et leurs catégories.
- **Validation des coûts et prix de vente** : Les enregistrements où le coût standard d'un produit excédait son prix de vente ont été détectés et corrigés, ou signalés pour garantir la cohérence des informations financières.
- **Analyse des valeurs par défaut et des données manquantes** :
  - Le nombre de produits ayant une couleur par défaut (indiquée comme Inconnu ) a été calculé pour évaluer l'impact de l'imputation des valeurs manquantes dans la colonne `Color`.
  - Le pourcentage de lignes ayant une date de début de vente manquante (`SellStartDate`) a été mesuré pour quantifier l'étendue des données incomplètes dans ce champ spécifique.

Afin d'assurer la cohérence et la complétude des données, plusieurs colonnes contenant des valeurs manquantes ont été mises à jour selon des méthodes spécifiques. Les étapes suivantes ont été réalisées :

- **Valeurs manquantes pour les niveaux de sécurité (`SafetyStockLevel`) et points de réapprovisionnement (`ReorderPoint`)** : Ces colonnes ont été mises à jour avec une valeur par défaut de 0, garantissant ainsi une base fiable pour les analyses liées aux stocks.
- **Imputation pour les coûts standards (`StandardCost`) et prix de liste (`ListPrice`)** : Les valeurs manquantes ont été remplacées par les moyennes calculées au sein des sous-catégories respectives, assurant une estimation réaliste pour ces champs.

	ProductID	ProductSubcategoryID
1	1	NULL
2	2	NULL
3	3	NULL
4	4	NULL
5	316	NULL
6	317	NULL
7	318	NULL
8	319	NULL

	ProductID	StandardCost	ListPrice
1	248	504	

	DefaultColors	TotalRowsAfterCleaning
1	248	504

	TotalRows	NullSellStartDate
1	504	0

Figur 33: Résultats des vérifications

- **Mise à jour des unités de mesure (SizeUnitMeasureCode et WeightUnitMeasureCode)** : Les valeurs nulles ont été remplacées par une valeur par défaut (N/A ), et une entrée spécifique a été ajoutée dans la table des unités de mesure pour garantir la compatibilité des données.
- **Calcul des jours de fabrication (DaysToManufacture)** : Les valeurs manquantes ont été imputées en utilisant la moyenne calculée par sous-catégorie, permettant de combler les lacunes tout en respectant les spécificités de chaque catégorie.
- **Validation des chaînes de caractères** : Les colonnes contenant des chaînes ont été examinées pour corriger toute troncation ou anomalie détectée.

Les analyses des contraintes de vérification associées à la table **Product** ont également été réalisées pour garantir l'intégrité des données. Voici les actions entreprises :

- **Colonne Style** : La contrainte appliquée à la colonne **Style** a été identifiée, permettant de s'assurer que les valeurs respectent les règles de validation pour cette colonne.
- **Colonne ProductLine** : La contrainte pour la colonne **ProductLine** a été vérifiée pour s'assurer que seules les valeurs valides sont acceptées.
- **Colonne Class** : La contrainte pour la colonne **Class** a été inspectée pour garantir que les valeurs suivent les règles de la base de données.

Ces étapes ont permis d'assurer que les données respectent les règles métiers et les contraintes de validation mises en place.

Résultats				
ConstraintName	TableName	ColumnName	ConstraintDefinition	
1	CK_Product_ProductLine	Product	ProductLine	(upper([ProductLine])='R' OR upper([ProductLine])...
ConstraintName	TableName	ColumnName	ConstraintDefinition	
1	CK_Product_Class	Product	Class	(upper([Class])='H' OR upper([Class])='M' OR upp...
ConstraintName	TableName	ColumnName	ConstraintDefinition	
1	CK_Product_Style	Product	Style	(upper([Style])='U' OR upper([Style])='M' OR upp...

Figur 34: Résultat d'analyse des contraintes

Afin d'améliorer la flexibilité et l'intégrité des données dans la table **Product**, les colonnes clés ont été modifiées et leurs contraintes de vérification mises à jour. Voici les modifications apportées pour chaque colonne concernée :

### 1. Colonne ProductLine :

- La contrainte existante CK\_Product\_ProductLine a été supprimée.
- La colonne ProductLine a été modifiée pour accepter des valeurs de type NVARCHAR(3).
- Une nouvelle contrainte a été ajoutée pour garantir que les valeurs autorisées sont parmi ('R', 'M', 'T', 'S', 'N').

### 2. Colonne Class :

- La contrainte existante CK\_Product\_Class a été supprimée.
- La colonne Class a été modifiée pour accepter des valeurs de type NVARCHAR(3).
- Une nouvelle contrainte a été ajoutée pour garantir que les valeurs autorisées sont parmi ('H', 'M', 'L', 'N').

### 3. Colonne Style :

- La contrainte existante CK\_Product\_Style a été supprimée.
- La colonne Style a été modifiée pour accepter des valeurs de type NVARCHAR(3).
- Une nouvelle contrainte a été ajoutée pour garantir que les valeurs autorisées sont parmi ('U', 'M', 'W', 'N').

	ost	ListPrice	Size	SizeUnitMeasureCode	WeightUnitMeasureCode	Weight	DaysToManufacture	ProductLine	Class	Style	ProductSubcategoryID	ProductMod
1		0,00	N/A	NA	NA	1.00	0	N	N	N	NULL	NULL
2		0,00	N/A	NA	NA	1.00	0	N	N	N	NULL	NULL
3		0,00	N/A	NA	NA	1.00	1	N	N	N	NULL	NULL
4		0,00	N/A	NA	NA	1.00	0	N	N	N	NULL	NULL
5		0,00	N/A	NA	NA	1.00	1	N	N	N	NULL	NULL
6		0,00	N/A	NA	NA	1.00	0	N	L	N	NULL	NULL
7		0,00	N/A	NA	NA	1.00	0	N	M	N	NULL	NULL

Figur 35: La table Production Product après nettoyage

## 3 Exportation des Données Nettoyées

### 3.1 Choix du Formats

#### Chargement Direct depuis SQL Server

Pourquoi SQL Server ?

SQL Server est conçu pour stocker, gérer et interroger de grandes quantités de données de manière efficace. L'intégration avec Power BI est native, ce qui permet une connexion directe et dynamique. Les données dans SQL Server peuvent être structurées, sécurisées, et maintenues via des vues et des procédures stockées, ce qui garantit leur conformité avant l'analyse.

### 3.2 Étapes Principales

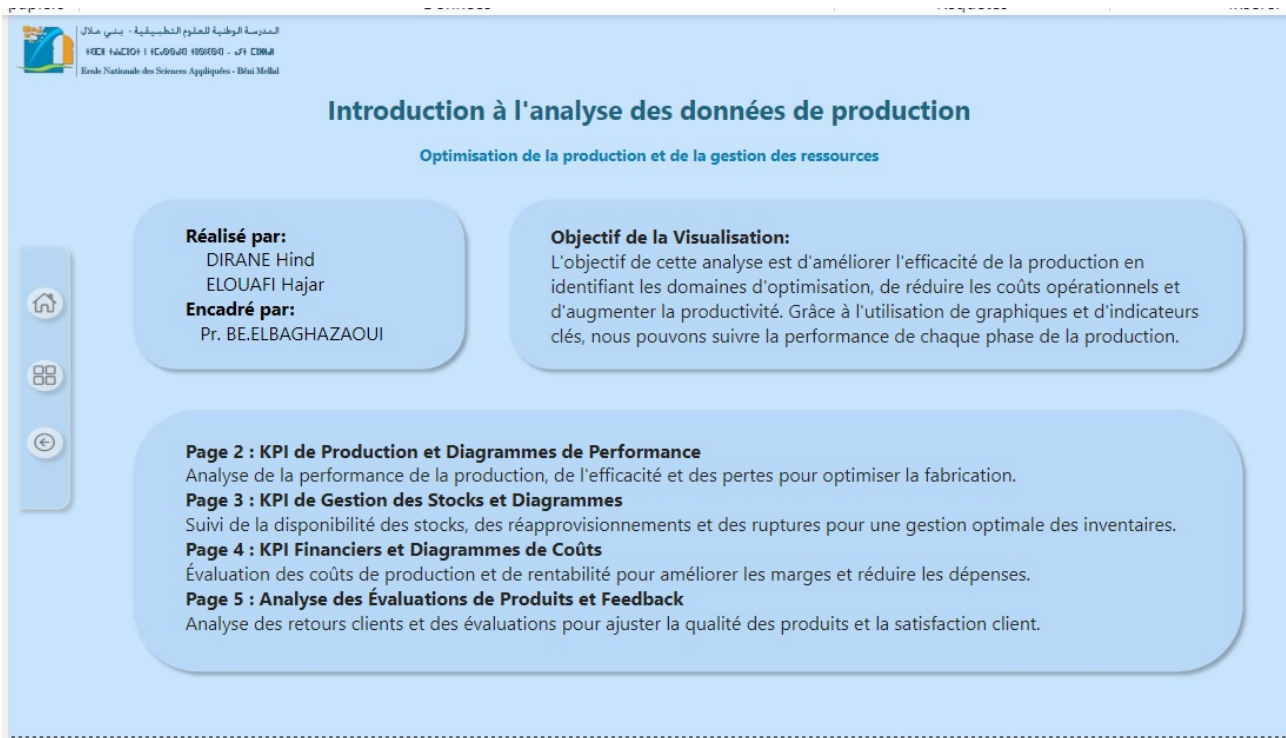
- **Connexion de Power BI à SQL Server :**
  - Ouvrir Power BI et cliquer sur l'option *Obtenir des données*.
  - Sélectionner le type de source *SQL Server*.
- **Fourniture des informations de connexion :**
  - Saisir les informations nécessaires telles que l'adresse du serveur SQL, le nom de la base de données, et les identifiants (nom d'utilisateur et mot de passe).
- **Sélection des tables ou vues :**
  - Choisir les tables ou les vues contenant les données nettoyées à importer dans Power BI.

## 4 Visualisation des Données

### 4.1 Dashboard 1 : Page d'Accueil

Le premier dashboard présente les informations générales de l'entreprise ainsi que des indicateurs clés sur l'état actuel. Il sert de point d'entrée pour les autres sections du rapport. Ce dashboard met en avant une vue d'ensemble des données essentielles, avec des éléments visuels comme des graphiques et des tableaux résumant les performances globales.

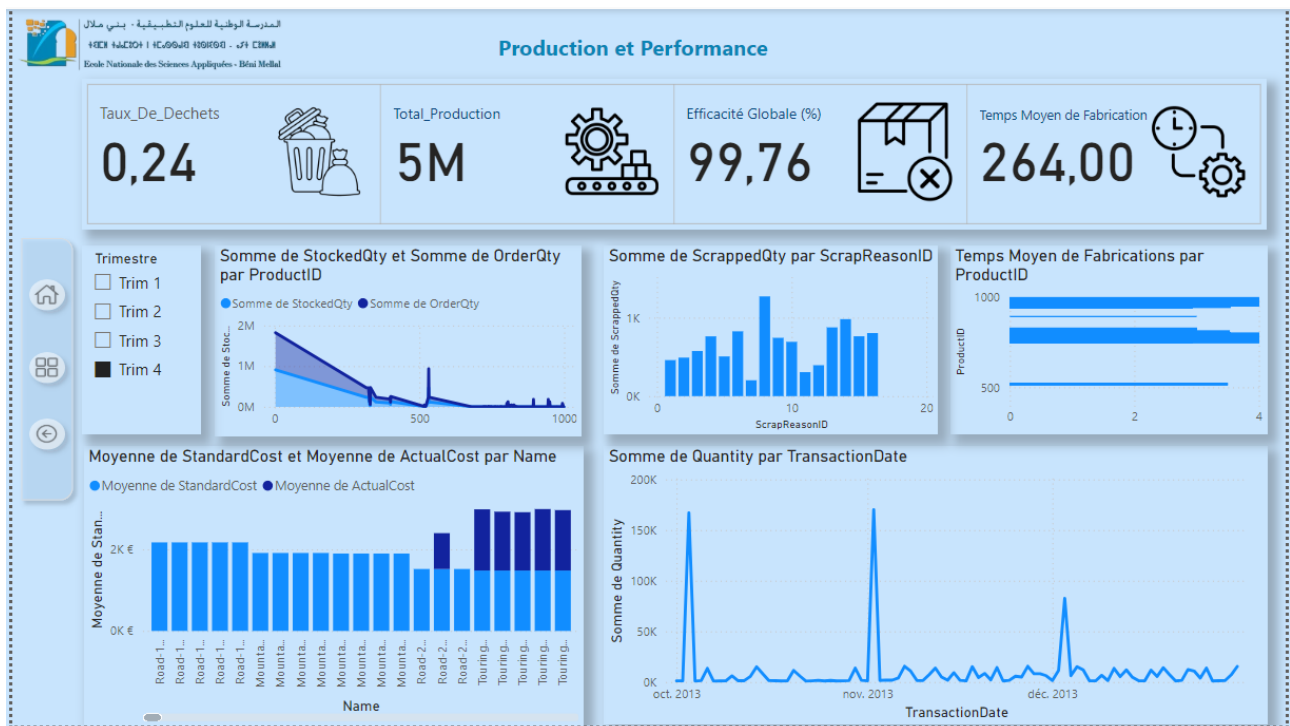




Figur 36: Dashboard de la Page d'Accueil

## 4.2 Dashboard 2 : Production et Performance

Ce dashboard permet de visualiser la performance de la production. Il met en avant les taux de production, les rendements, ainsi que les éventuelles erreurs de production ou non-conformités.



Figur 37: Dashboard sur la Production et la Performance

#### 4.2.1 Détails des KPI

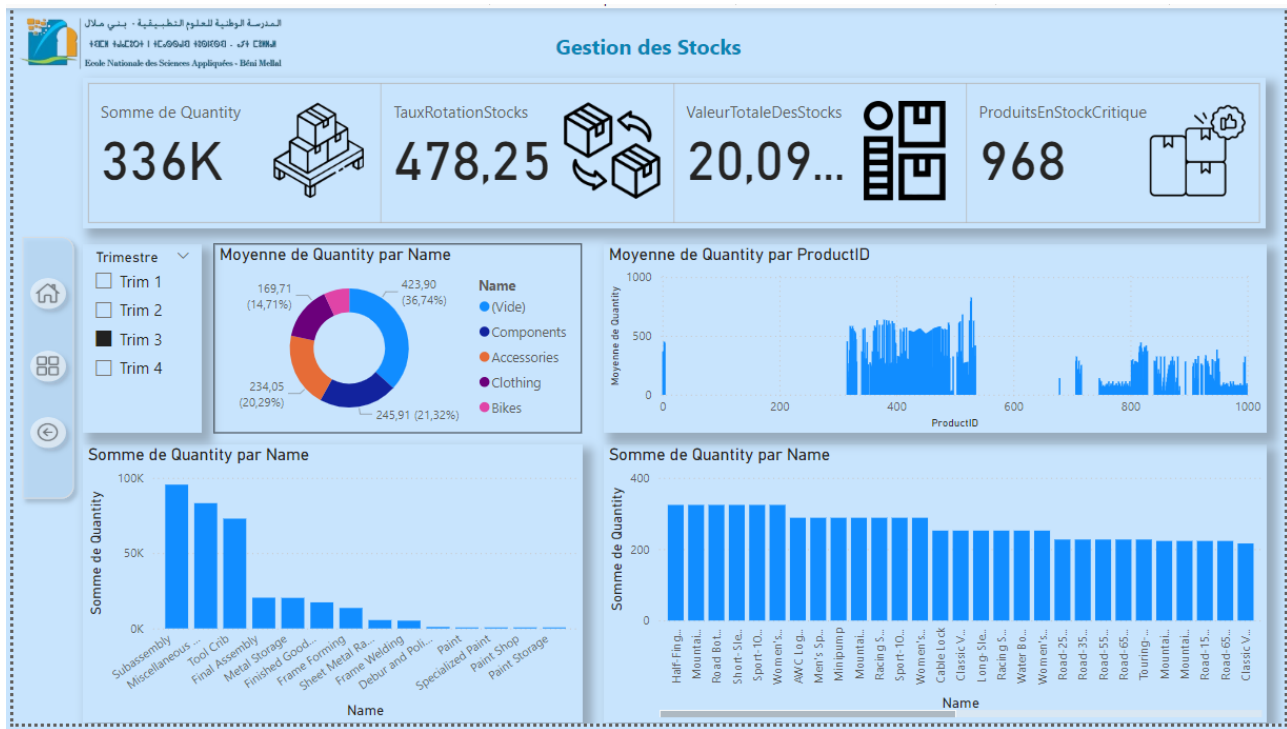
- **Taux de Déchet** : Le taux de déchet mesure la quantité de production qui ne répond pas aux normes de qualité ou qui a été rejetée. Cet indicateur est essentiel pour évaluer l'efficacité du processus de production et identifier les zones où des améliorations peuvent être apportées.
  - **Description** : Le taux de déchet est exprimé en pourcentage et montre la proportion de produits non conformes par rapport à la production totale.
  - **Visualisation** : Ce KPI est visualisé dans une carte indiquant le pourcentage du taux de déchet par rapport aux objectifs fixés.
- **Total de Production** : Le total de production donne une vue d'ensemble du volume total de production effectué sur une période donnée. Ce KPI est crucial pour évaluer la capacité de production et identifier des tendances de performance.
  - **Description** : Ce KPI montre le volume total des produits fabriqués, qu'ils soient conformes ou non.
  - **Visualisation** : Le total de production est visualisé sous forme d'un graphique ou d'une carte qui affiche les quantités produites au cours d'une période donnée.
- **Efficacité Globale** : L'EG mesure l'efficacité combinée de la production en prenant en compte les disponibilités, les performances et la qualité. Ce KPI est un indicateur clé pour évaluer la productivité du processus de fabrication.
  - **Description** : L'EG est calculé en multipliant le taux de disponibilité, le taux de performance et le taux de qualité. Il fournit un pourcentage qui reflète l'efficacité globale des équipements de production.
  - **Visualisation** : Ce KPI est visualisé sous forme de carte ou de graphique à barres montrant le pourcentage de l'efficacité par rapport à un objectif prédéfini.
- **Temps Moyen de Fabrication** : Le temps moyen de fabrication mesure le temps moyen nécessaire pour produire une unité ou un lot de produits. Cet indicateur est essentiel pour identifier les goulots d'étranglement dans le processus de production et optimiser les processus.
  - **Description** : Le temps moyen de fabrication est calculé en fonction du temps total passé sur la production divisé par le nombre d'unités fabriquées. Un temps de fabrication plus court signifie une production plus rapide et plus efficace.
  - **Visualisation** : Ce KPI est présenté sous forme de graphique ou de carte, montrant la durée moyenne nécessaire pour produire un lot d'articles.

### 4.2.2 Détails des Diagrammes

- **Filtrage par Trimestre** : Les graphiques ci-dessus peuvent être associés à un filtre par trimestre, permettant d'analyser les données par période (trimestre) afin de mieux comprendre les tendances saisonnières et l'impact des stratégies à court terme sur la performance globale.
- **Diagramme 1 : Somme de StockedQty et OrderQty par ProductID** : Un graphique de zone empilée montrant l'évolution de la somme de 'stockedquantity' et 'orderquantity' pour chaque 'productid' au fil du temps. Ce graphique permet de visualiser les variations des niveaux de stock et des commandes en parallèle, afin d'évaluer la gestion des stocks.
- **Diagramme 2 : Somme de ScrappedQty par ScrapReasonID** : Un histogramme groupé qui montre la somme de 'scrappedqty' par 'scrapreasonid'. Cela permet de visualiser la distribution des rejets de produits en fonction des raisons spécifiques identifiées dans le processus de production.
- **Diagramme 3 : Temps Moyen de Fabrication par ProductID** : Un graphique à barres groupées montrant le temps moyen de fabrication pour chaque 'productid'. Ce graphique permet de comparer les performances de fabrication entre différents produits et d'identifier ceux qui nécessitent plus de temps pour leur production.
- **Diagramme 4 : Moyenne de StandardCost et ActualCost par Name** : Un histogramme empilé montrant la moyenne de 'standardcost' et de 'actualcost' pour chaque 'name'. Ce graphique permet de visualiser les écarts entre les coûts standard et réels, et d'analyser l'efficacité des coûts de production.
- **Diagramme 5 : Somme de Quantity par TransactionDate** : Un graphique en groupe montrant la somme de 'quantity' par 'transactiondate'. Ce graphique permet de visualiser les tendances de la production et des ventes sur une période donnée, offrant ainsi une vue d'ensemble de l'évolution de l'activité.

### 4.3 Dashboard 3 : Gestion des Stocks

Le dashboard de gestion des stocks fournit un aperçu des niveaux de stock actuels, des tendances d'approvisionnement, et des produits en rupture de stock ou à faible disponibilité.



Figur 38: dashboard sur la gestion des stocks

#### 4.3.1 Détails des KPI

- **Somme de Quantity de Stock** : Ce KPI mesure la quantité totale de stock disponible pour tous les produits dans l'inventaire. Il est essentiel pour avoir une vue d'ensemble sur le niveau de stock global et pour prévoir la gestion future du stock.
  - **Description** : La somme de la quantité de stock représente la quantité totale de produits stockés dans l'entrepôt, sans distinction entre les produits finis, en cours ou les matières premières.
  - **Visualisation** : Ce KPI est visualisé dans une carte indiquant le volume total des stocks disponibles. Cela peut être affiché sous forme de chiffres ou d'un graphique à barres.
- **Taux de Rotation de Stock** : Le taux de rotation de stock mesure la rapidité avec laquelle les stocks sont utilisés et renouvelés au sein de l'entreprise. Ce KPI est crucial pour évaluer l'efficacité de la gestion des stocks et minimiser le capital immobilisé.
  - **Description** : Le taux de rotation de stock est calculé en divisant le coût des ventes par le stock moyen sur une période donnée. Un taux de rotation élevé indique que les produits sont vendus rapidement et que le stock ne reste pas trop longtemps en entrepôt.
  - **Visualisation** : Ce KPI peut être visualisé sous forme de graphique en ligne ou de barre, montrant le taux de rotation sur une période donnée.

- **Valeur de Stock (Prix) :** Ce KPI indique la valeur totale du stock en termes monétaires. Il permet de suivre l'investissement total dans les stocks et d'analyser l'impact financier de la gestion des stocks.
  - **Description :** La valeur de stock représente le prix total des articles stockés. Cela inclut tous les produits finis, les matières premières, ainsi que les produits en cours de fabrication.
  - **Visualisation :** Ce KPI est visualisé sous forme de carte ou de graphique à barres montrant la valeur totale des stocks en termes de coût d'achat ou de coût de production.
- **Produits en Stock Critique :** Ce KPI identifie les produits qui sont en rupture de stock ou proches de l'épuisement. Il est essentiel pour anticiper les besoins en réapprovisionnement et éviter les ruptures de stock.
  - **Description :** Ce KPI montre le nombre de produits dont le niveau de stock est inférieur à un seuil critique. Cela permet de prendre des mesures proactives pour réapprovisionner les produits avant qu'ils ne soient complètement épuisés.
  - **Visualisation :** Ce KPI est présenté sous forme d'un tableau ou d'un graphique indiquant le nombre de produits en dessous du niveau de stock minimum. Les produits en rupture sont mis en évidence.

#### 4.3.2 Détails des Diagrammes

- **Filtrage par Trimestre :** Les graphiques ci-dessus peuvent être associés à un filtre par trimestre, permettant de visualiser les données sur une base trimestrielle. Cela permet d'évaluer la gestion des stocks et la répartition des produits à différentes périodes de l'année, en tenant compte des variations saisonnières ou des tendances de consommation.
- **Diagramme 1 : Moyenne de Quantity par Name (Table: Production.ProductInventory, Table: Production.ProductCategory) :**

Un graphique de type anneau montrant la moyenne de 'quantity' par 'name' à partir des tables 'Production.ProductInventory' et 'Production.ProductCategory'. Ce graphique permet d'analyser la répartition des quantités stockées par produit, tout en prenant en compte les catégories de produits pour mieux comprendre les tendances des stocks.

- **Diagramme 2 : Moyenne de Quantity par ProductID (Table: Production.ProductInventory) :**

Un histogramme empilé qui montre la moyenne de 'quantity' par 'productid' dans la table 'Production.ProductInventory'. Ce graphique permet de comparer les niveaux de stock moyens par produit, ce qui est utile pour évaluer les performances de gestion des stocks et détecter des anomalies ou des besoins en réapprovisionnement.

- **Diagramme 3 : Somme de Quantity par Name et Location (Table: Production.ProductInventory, Table: Production.Location) :**

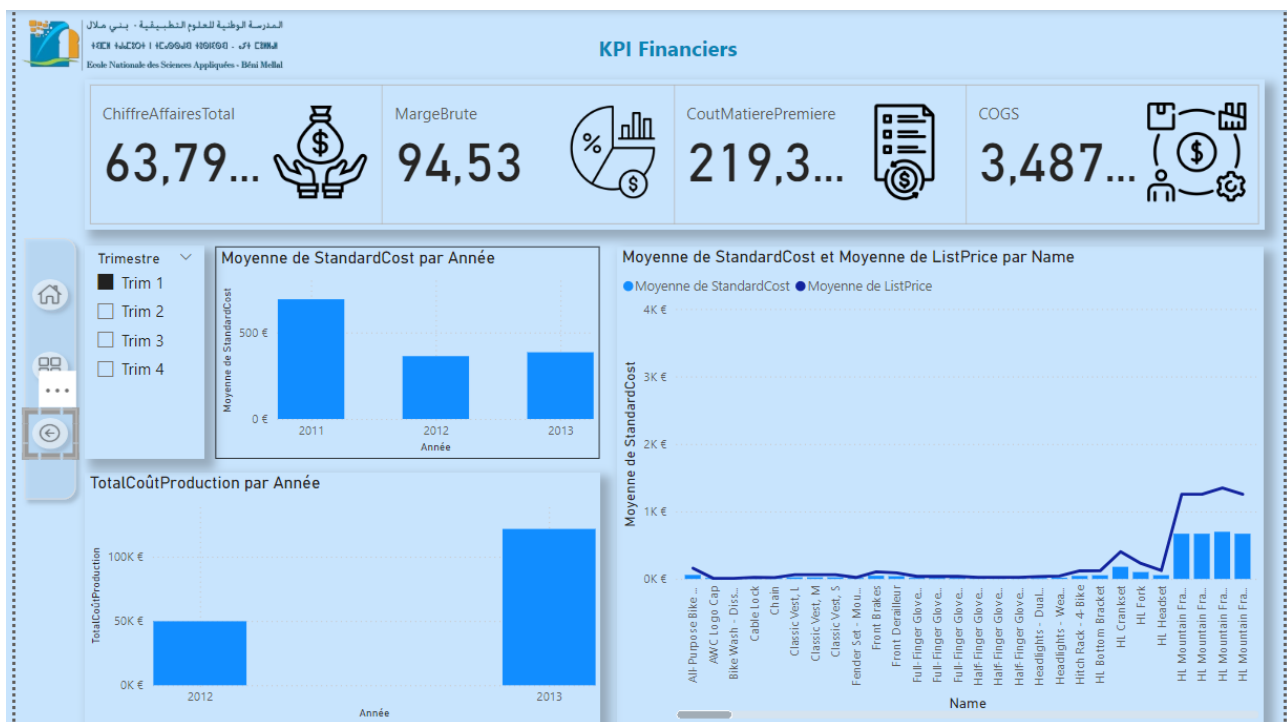
Un histogramme empilé montrant la somme de 'quantity' par 'name' et 'location'. Ce graphique permet de visualiser la distribution des quantités stockées en fonction des lieux de stockage et des produits, facilitant ainsi la gestion des stocks à travers différents sites ou entrepôts.

- **Diagramme 4 : Somme de Quantity par Name et Product (Table: Production.ProductInventory, Table: Production.Product) :**

Un histogramme groupé montrant la somme de 'quantity' par 'name' et 'product' dans les tables 'Production.ProductInventory' et 'Production.Product'. Ce graphique permet de visualiser la quantité totale de produits stockés pour chaque type de produit et d'analyser les niveaux de stock de manière détaillée par produit.

## 4.4 Dashboard 4 : KPI Financiers

Ce dashboard permet de visualiser les principaux indicateurs financiers, tels que le chiffre d'affaires, la rentabilité, et le flux de trésorerie.



Figur 39: Dashboard sur les KPI Financiers

### 4.4.1 Détails des KPI

- **Chiffre d'Affaires Total :** Ce KPI mesure le revenu total généré par les ventes de produits ou services sur une période donnée. Il est essentiel pour évaluer la performance financière globale de l'entreprise.

- **Description** : Le chiffre d'affaires total représente la somme de toutes les ventes réalisées, qu'elles soient en numéraire ou à crédit. Il est un indicateur clé pour évaluer la capacité de l'entreprise à générer des revenus.
- **Visualisation** : Ce KPI peut être visualisé sous forme d'une carte indiquant le chiffre d'affaires total réalisé, ou sous forme de graphique montrant l'évolution du chiffre d'affaires au fil du temps.
- **Marge Brute** : Ce KPI mesure la différence entre le chiffre d'affaires et le coût des biens ou services vendus, ce qui reflète la rentabilité directe des ventes avant les frais fixes et variables.

- **Description** : La marge brute est calculée comme suit :

$$MargeBrute = Chiffre d'Affaires - Coût des Ventes$$

Ce KPI est un indicateur clé pour analyser la rentabilité directe de l'entreprise avant les charges fixes (frais généraux, salaires, etc.).

- **Visualisation** : Ce KPI est souvent représenté sous forme d'une carte ou d'un graphique à barres pour montrer l'évolution de la marge brute dans le temps.
- **Coût des Matières Premières** : Ce KPI représente le coût des matières premières utilisées dans la production de biens ou services. C'est un indicateur crucial pour suivre les coûts variables associés à la fabrication des produits.
  - **Description** : Le coût des matières premières représente la somme des dépenses pour les matières premières consommées durant une période donnée. Ce KPI permet de surveiller l'efficacité de l'approvisionnement et de la gestion des coûts de production.
  - **Visualisation** : Ce KPI est visualisé sous forme de graphique ou de carte montrant le montant total des dépenses pour les matières premières pendant une période spécifique.
- **COGS (Coût des Biens Vendus)** : Ce KPI mesure le coût total des biens ou services vendus durant une période donnée. Il comprend les coûts directs associés à la production des biens ou services.
  - **Description** : Le COGS inclut tous les coûts directs de production, comme les matières premières, la main-d'œuvre directe, et les coûts variables associés à la production des biens ou services.
  - **Visualisation** : Le COGS peut être visualisé sous forme d'une carte ou d'un graphique à barres indiquant le coût total des biens ou services produits durant la période, ou d'une courbe montrant l'évolution du COGS dans le temps.

#### 4.4.2 Détails des Diagrammes

- **Filtrage par Trimestre** : Les graphiques ci-dessus peuvent être associés à un filtre par trimestre, permettant de visualiser les données sur une base trimestrielle. Cela permet d'évaluer la gestion des stocks et la répartition des produits à différentes périodes de l'année, en tenant compte des variations saisonnières ou des tendances de consommation.

- **Diagramme 1 : Moyenne de Quantity par Name (Table: Production.ProductInventory, Table: Production.ProductCategory)** :

Un graphique de type anneau montrant la moyenne de 'quantity' par 'name' à partir des tables 'Production.ProductInventory' et 'Production.ProductCategory'. Ce graphique permet d'analyser la répartition des quantités stockées par produit, tout en prenant en compte les catégories de produits pour mieux comprendre les tendances des stocks.

- **Diagramme 2 : Moyenne de Quantity par ProductID (Table: Production.ProductInventory)** :

Un histogramme empilé qui montre la moyenne de 'quantity' par 'productid' dans la table 'Production.ProductInventory'. Ce graphique permet de comparer les niveaux de stock moyens par produit, ce qui est utile pour évaluer les performances de gestion des stocks et détecter des anomalies ou des besoins en réapprovisionnement.

- **Diagramme 3 : Somme de Quantity par Name et Location (Table: Production.ProductInventory, Table: Production.Location)** :

Un histogramme empilé montrant la somme de 'quantity' par 'name' et 'location'. Ce graphique permet de visualiser la distribution des quantités stockées en fonction des lieux de stockage et des produits, facilitant ainsi la gestion des stocks à travers différents sites ou entrepôts.

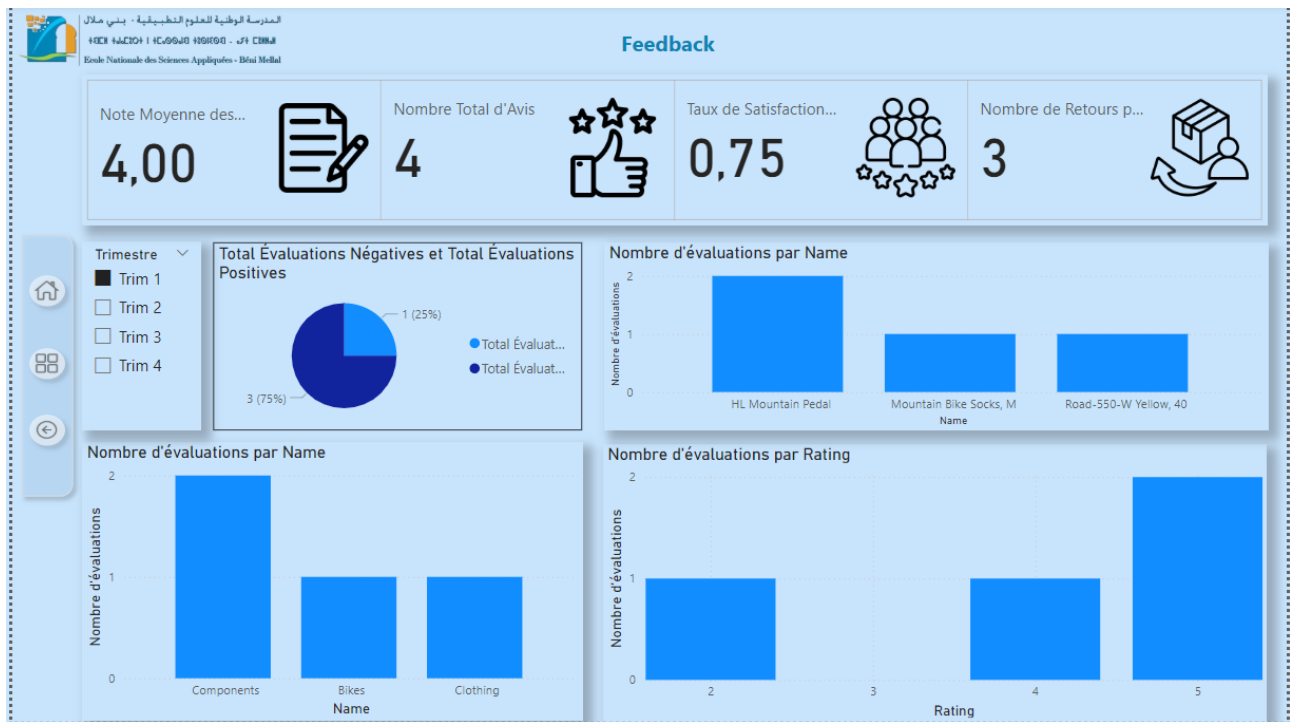
- **Diagramme 4 : Somme de Quantity par Name et Product (Table: Production.ProductInventory, Table: Production.Product)** :

Un histogramme groupé montrant la somme de 'quantity' par 'name' et 'product' dans les tables 'Production.ProductInventory' et 'Production.Product'. Ce graphique permet de visualiser la quantité totale de produits stockés pour chaque type de produit et d'analyser les niveaux de stock de manière détaillée par produit.

#### 4.5 Dashboard 5 : Feedback

Le dashboard de feedback permet de recueillir et d'analyser les retours des clients ou des employés, en mettant l'accent sur la satisfaction et la fidélité.





Figur 40: Dashboard sur FeedBack

#### 4.5.1 Détails des KPI

- Note Moyenne des Produits :** Ce KPI mesure la moyenne des évaluations données par les clients sur les produits. Il est essentiel pour évaluer la qualité perçue des produits et identifier les domaines nécessitant des améliorations.
  - Description :** La note moyenne des produits est calculée en faisant la moyenne des évaluations attribuées par les clients, généralement sur une échelle de 1 à 5 ou 1 à 10. Une note élevée indique que les clients sont généralement satisfaits des produits.
  - Visualisation :** Ce KPI peut être visualisé sous forme de carte indiquant la note moyenne obtenue, ou sous forme de graphique représentant l'évolution de la note moyenne au fil du temps.
- Nombre Total d'Avis :** Ce KPI mesure le nombre total d'avis reçus pour les produits sur une période donnée. Il est essentiel pour évaluer la participation des clients et l'engagement avec les produits.
  - Description :** Le nombre total d'avis représente le total des évaluations soumises par les clients pour les produits. Il permet de mesurer l'intérêt des clients et la fréquence de l'interaction avec les produits.
  - Visualisation :** Ce KPI peut être visualisé sous forme d'un graphique à barres ou d'une carte affichant le nombre total d'avis reçus pendant une période spécifiée.

- **Taux de Satisfaction Clients** : Ce KPI mesure le pourcentage de clients satisfaits par les produits ou services offerts. C'est un indicateur clé pour évaluer la fidélité des clients et la qualité perçue des produits.
  - **Description** : Le taux de satisfaction est calculé en fonction des réponses des clients (par exemple, des enquêtes ou des évaluations), exprimées en pourcentage. Un taux élevé reflète une forte satisfaction des clients.
  - **Visualisation** : Ce KPI est souvent représenté sous forme de pourcentage sur une carte ou un graphique circulaire, montrant la proportion de clients satisfaits par rapport au nombre total de clients interrogés.
- **Nombre de Retours Produits** : Ce KPI mesure le nombre de produits retournés par les clients. Un nombre élevé de retours peut indiquer des problèmes avec la qualité ou l'adéquation des produits.
  - **Description** : Le nombre de retours produits représente la quantité de produits retournés par les clients pour diverses raisons (défaut, insatisfaction, etc.). Un taux élevé peut signaler la nécessité de revoir les caractéristiques des produits ou la politique de qualité.
  - **Visualisation** : Ce KPI peut être visualisé sous forme de graphique à barres ou d'une carte indiquant le nombre de produits retournés, ou sous forme d'un graphique à courbes montrant les tendances de retour de produits au fil du temps.

#### 4.5.2 Détails des Diagrammes

- **Diagramme 1 : Total d'Évaluation Négative et Positive** :  
 Un graphique en secteur montrant la répartition du 'total d'évaluation négative' et du 'total d'évaluation positive' (mesures calculées). Ce graphique permet de visualiser la répartition des évaluations positives et négatives des produits, facilitant ainsi l'analyse de la satisfaction des clients.
- **Diagramme 2 : Nombre d'Évaluations par Name (Table: Production.Product)**:  
 Un histogramme empilé montrant le 'nombre d'évaluations' par 'name' à partir de la table 'Production.Product'. Ce graphique permet d'analyser le nombre d'évaluations reçues pour chaque produit. Il est utile pour identifier les produits les plus évalués et analyser leur performance globale.
- **Diagramme 3 : Nombre d'Évaluations par Name (Table: Production.Product Category)**:  
 Un histogramme empilé montrant le 'nombre d'évaluations' par 'name' à partir de la table 'Production.ProductCategory'. Ce graphique permet de visualiser la distribution des évaluations par catégorie de produit, facilitant l'analyse de la performance des différentes catégories.

- **Diagramme 4 : Nombre d'Évaluations par Rating (Table: Production.Product Review):**

Un histogramme montrant le 'nombre d'évaluations' par 'rating' à partir de la table 'Production.ProductReview'. Ce graphique permet d'analyser les évaluations reçues en fonction de la note attribuée, offrant ainsi une vue d'ensemble de la satisfaction des clients pour chaque produit.

## 4.6 Dashboard 6 : Prédiction

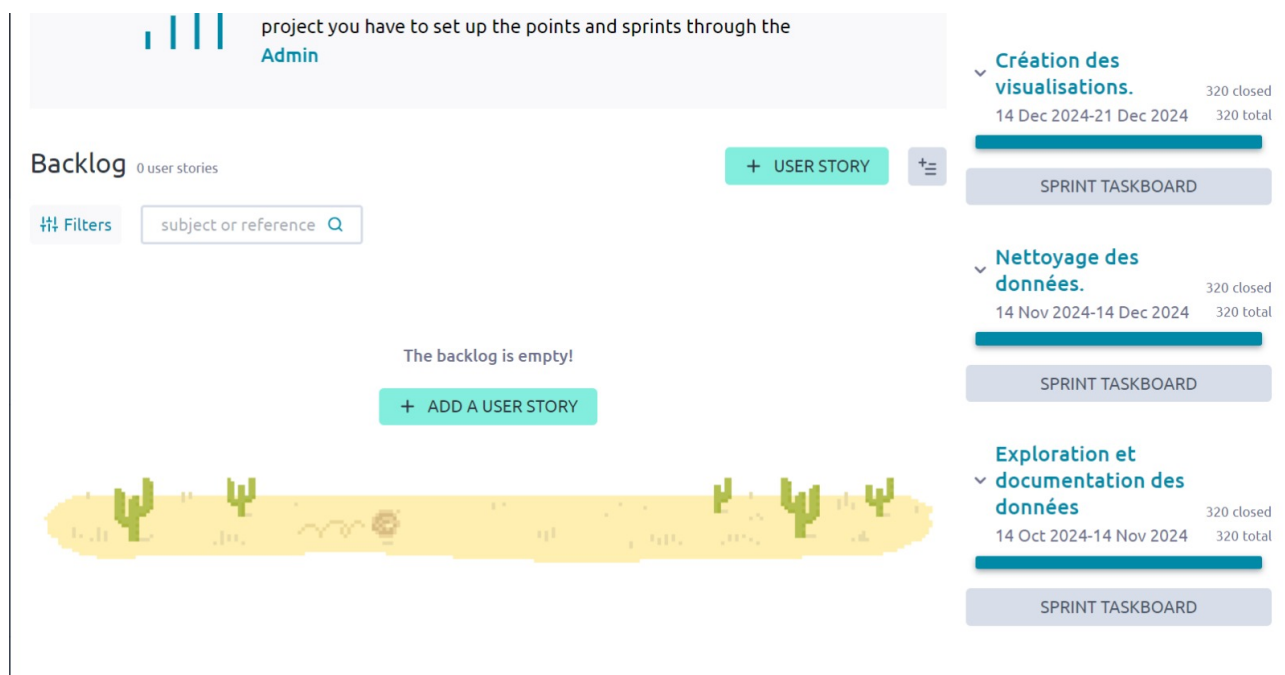
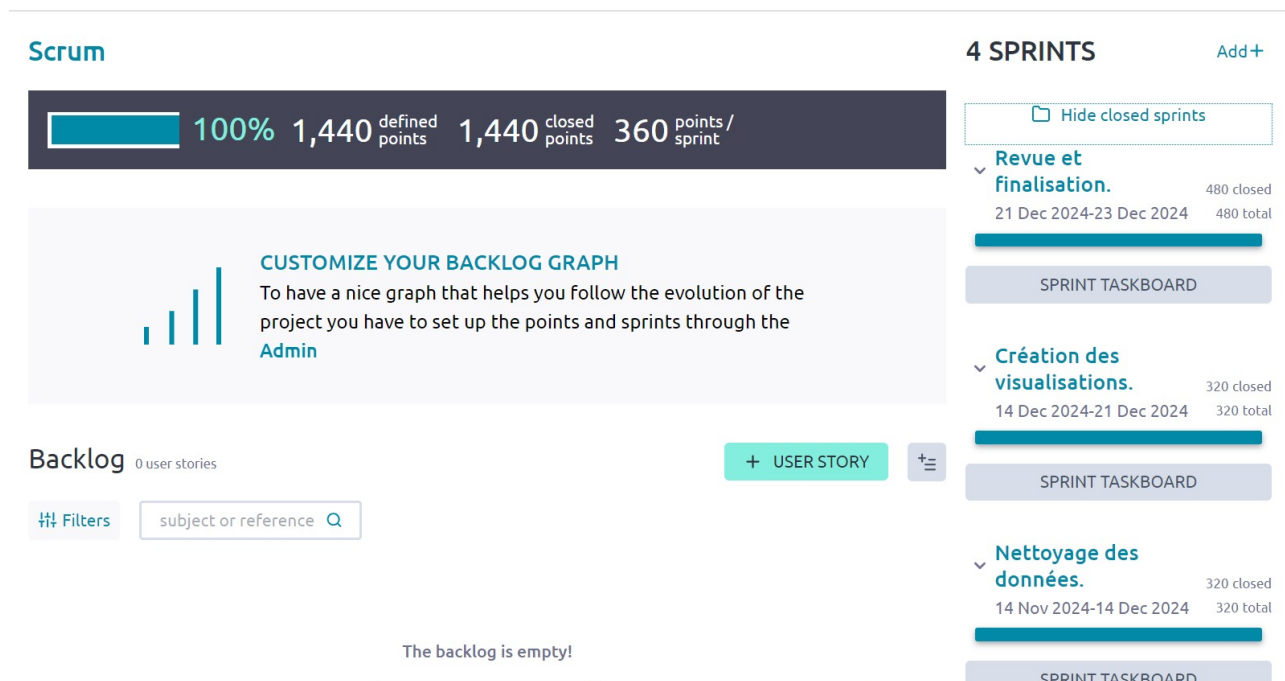
Ce dashboard est dédié aux prévisions basées sur les données historiques et les modèles d'analyse. Il permet de prédire les ventes futures, la demande, et d'autres éléments importants pour la planification.



Figur 41: Dashboard de la Prédiction

## 5 Phases du projet

Dans le cadre de notre projet, nous avons adopté le cadre méthodologique Scrum en utilisant Taiga.io pour gérer nos tâches et suivre l'avancement des travaux. Scrum nous a permis d'organiser le projet en sprints, chacun correspondant à une phase spécifique, comme l'exploration des données, le nettoyage, ou la création de visualisations. À chaque début de sprint, nous avons planifié les tâches à accomplir, définissant des priorités et attribuant les responsabilités à chaque membre de l'équipe. Des réunions quotidiennes (daily scrums) ont été réalisées pour partager les progrès, identifier les obstacles et ajuster le plan si nécessaire. À la fin de chaque sprint, une revue a été menée pour évaluer les livrables, suivie d'une rétrospective pour améliorer nos processus. Cette approche a permis une collaboration structurée et un respect rigoureux des délais.



Figur 42: Ces deux images montrent le Backlog.

## 5.1 Sprint 1 : Exploration des données

L'objectif de ce sprint était d'explorer les données disponibles afin de comprendre leur structure, identifier les tables pertinentes, et documenter les informations nécessaires pour les étapes suivantes. Nous avons commencé par analyser les schémas des bases de données, notamment les tables de production, pour repérer les relations entre elles. Cette étape a permis de dresser une liste des champs utiles pour les KPI et visualisations futurs. Les défis rencontrés incluaient l'identification des doublons, les données manquantes, et l'identification des champs calculables. Nous avons également créé un tableau de documentation pour chaque table, décrivant le nombre



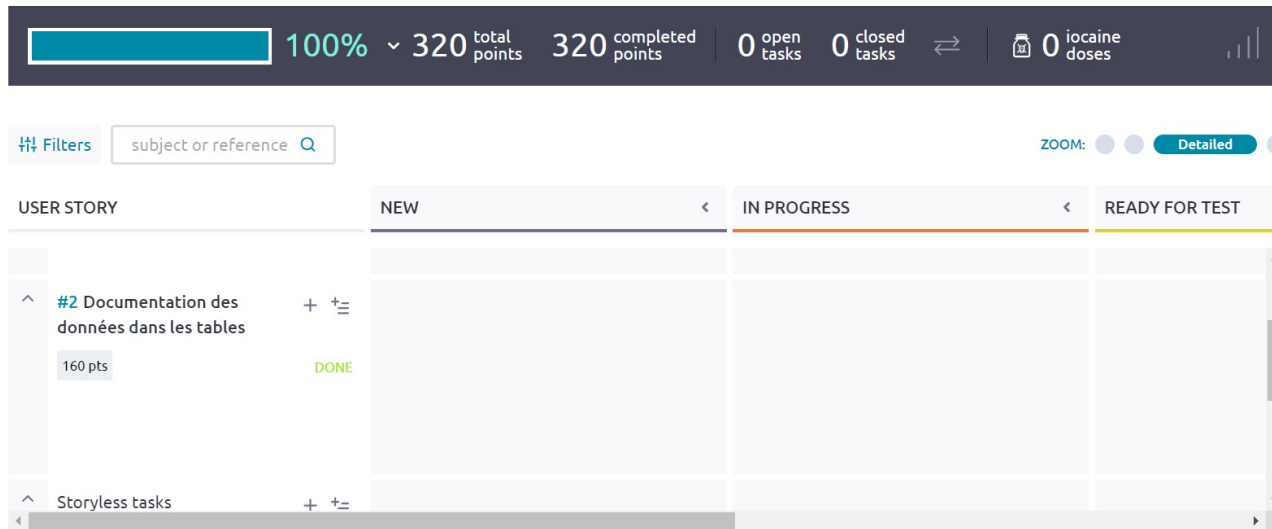
Figur 43: Ces deux images montrent les Sprints.

de colonnes, le type de données et le nombre d'enregistrements.



Figur 44: Exploration des données

## Exploration et documentation des données 14 Oct 2024 to 14 Nov 2024

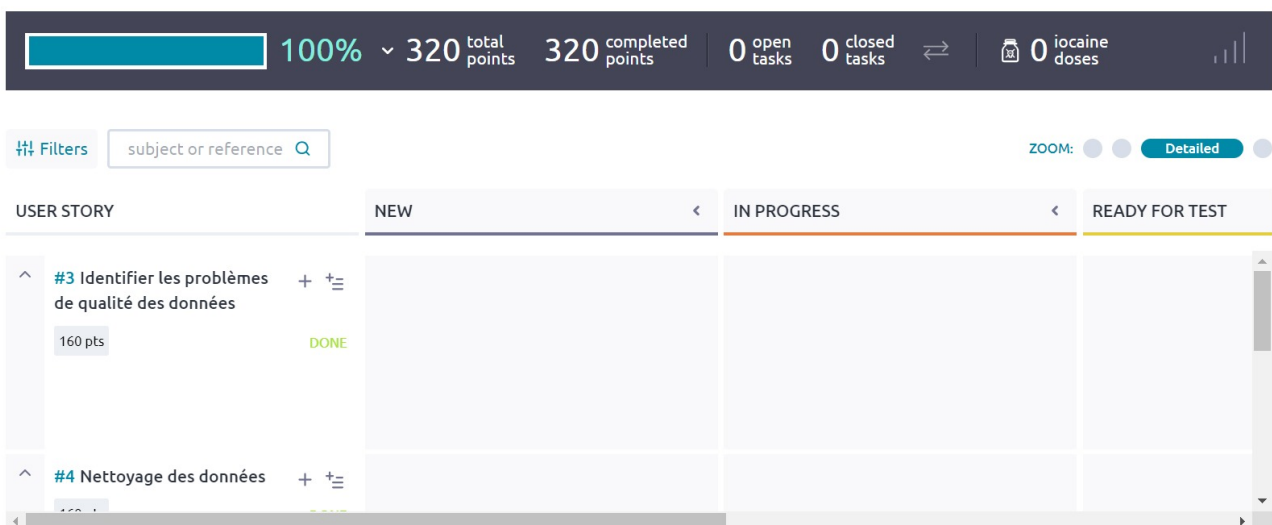


Figur 45: Exploration des données

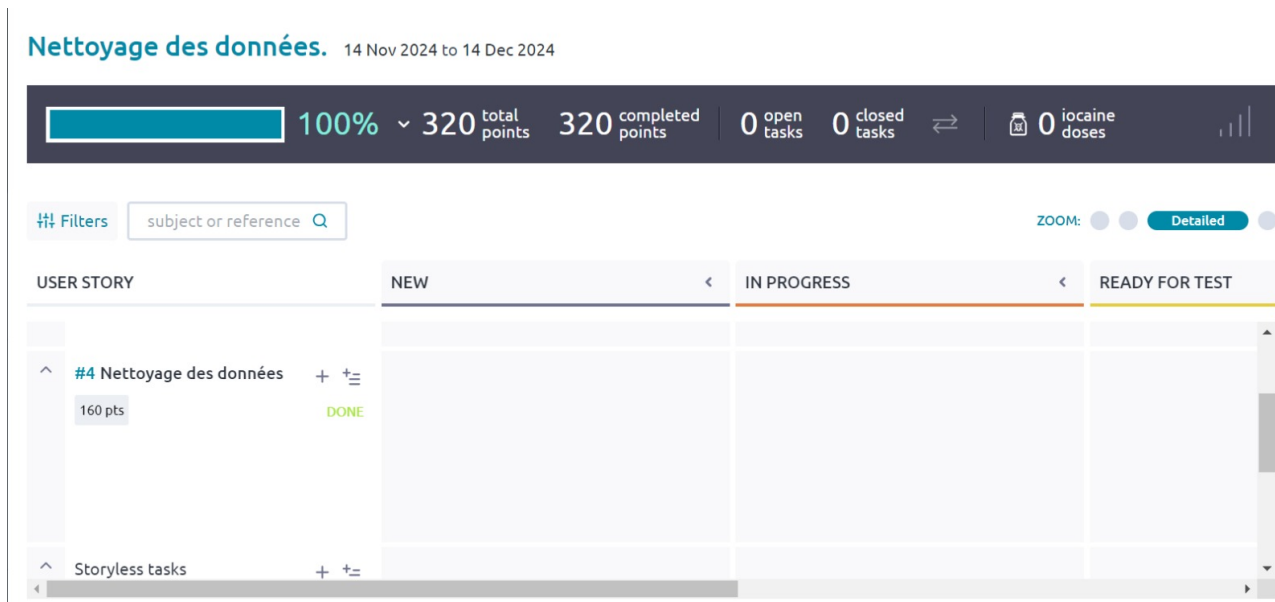
## 5.2 Sprint 2: Nettoyage des données

Ce sprint s'est concentré sur le nettoyage et la préparation des données pour garantir leur cohérence, leur complétude, et leur qualité. Nous avons utilisé des méthodes de suppression des doublons, d'imputation des valeurs manquantes, et de vérification des contraintes d'intégrité référentielles entre les tables. Les règles métier ont été appliquées pour standardiser les formats (ex. : dates, devises, codes produits). Cette étape a nécessité des scripts SQL personnalisés pour transformer les données brutes en un format exploitable. À la fin de ce sprint, un rapport a été généré pour comparer les données avant et après nettoyage, mettant en avant les transformations effectuées et leur impact sur la qualité des données.

## Nettoyage des données. 14 Nov 2024 to 14 Dec 2024



Figur 46: Nettoyage des données



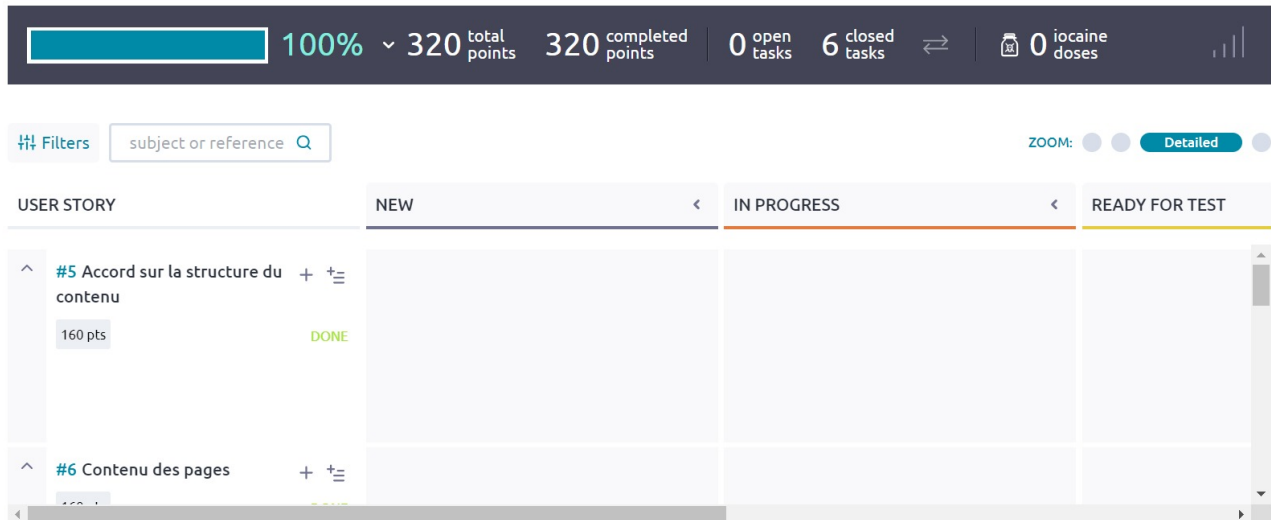
Figur 47: Nettoyage des données

### 5.3 Sprint 3 : Création de visualisations

L'objectif de ce sprint était de transformer les données nettoyées en visualisations pertinentes pour répondre aux besoins des parties prenantes. Nous avons défini la structure des six pages principales :

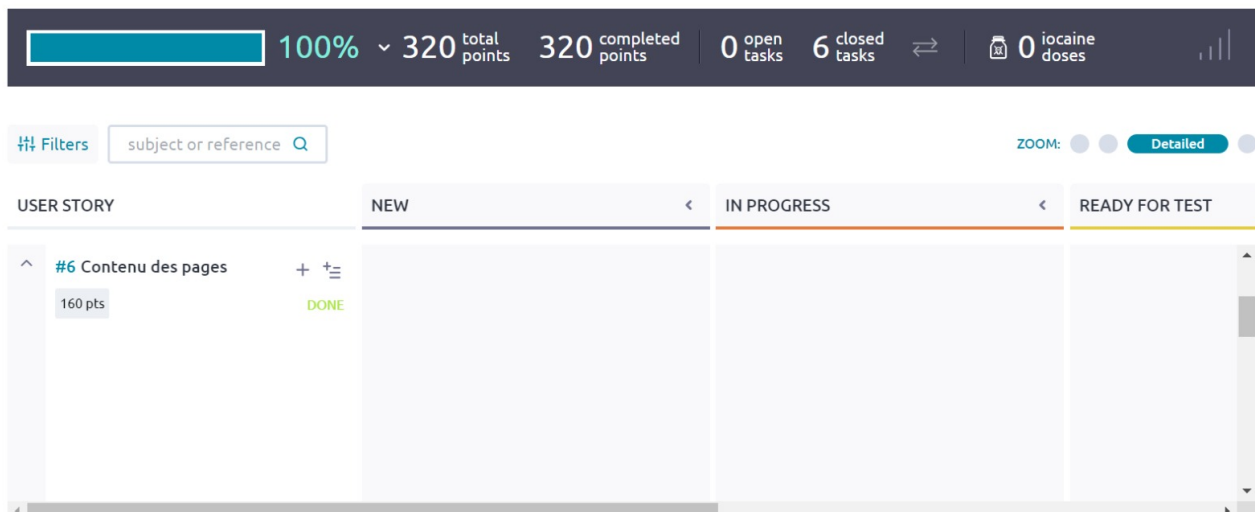
Introduction et informations générales. KPI de production et diagrammes de performance. KPI de gestion des stocks et diagrammes. KPI financiers et diagrammes de coûts. Analyse des évaluations de produits et feedback. Prédictions et attributs à surveiller. Les outils utilisés incluaient Power BI et Tableau, avec des graphiques comme des barres, lignes, et diagrammes en camembert. Une attention particulière a été portée à la sélection des KPI pour qu'ils reflètent les objectifs du projet. Les premiers retours ont permis d'affiner les visualisations pour maximiser leur lisibilité et leur impact.

## Création des visualisations. 14 Dec 2024 to 21 Dec 2024



Figur 48: Création de visualisations

## Création des visualisations. 14 Dec 2024 to 21 Dec 2024

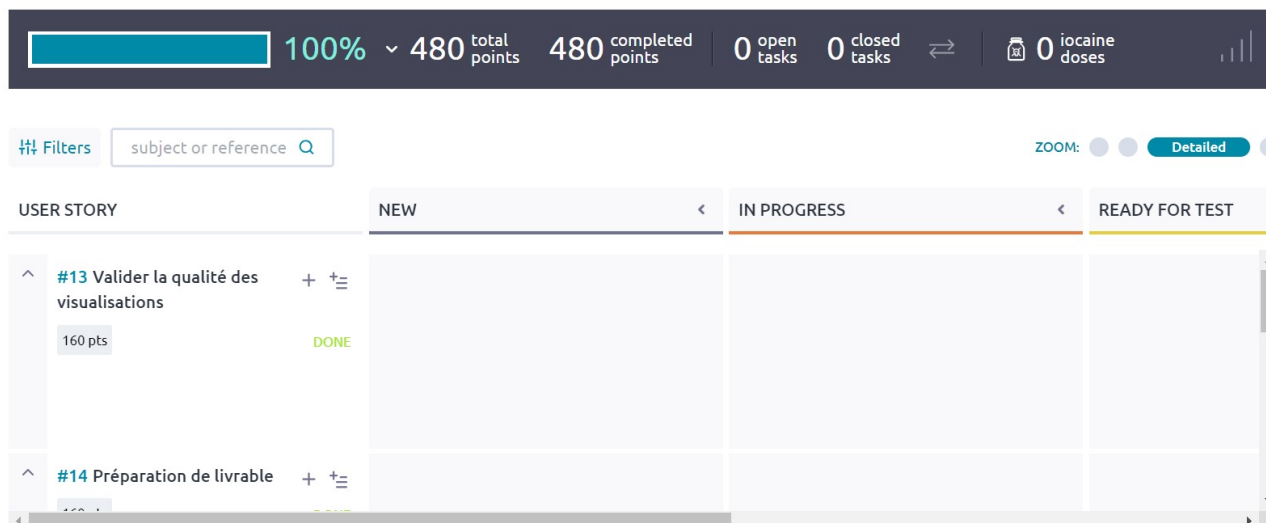


Figur 49: Création de visualisations

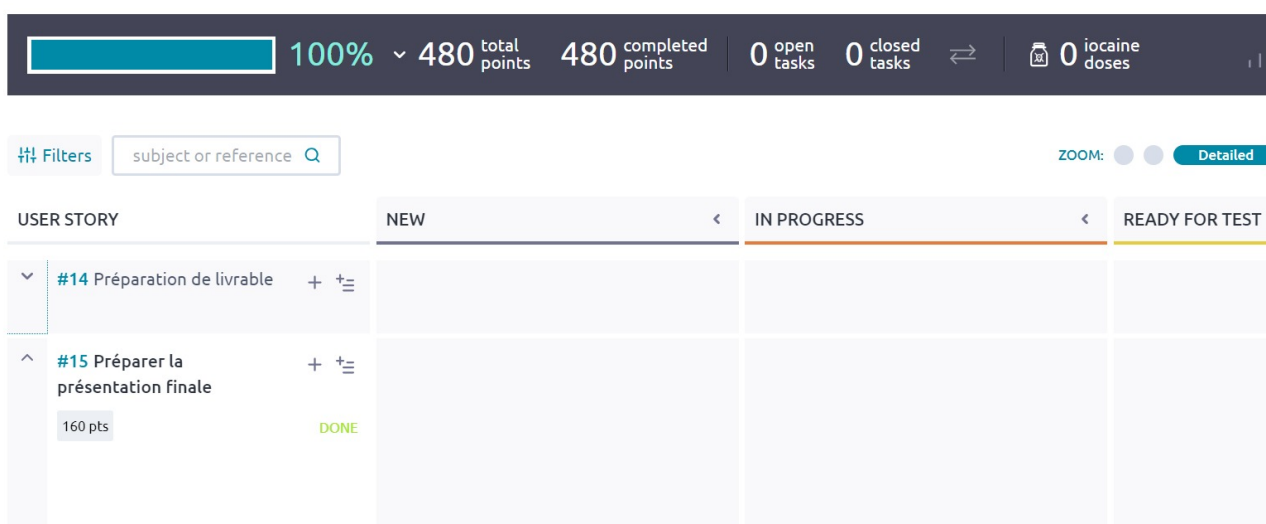
### 5.4 Sprint 4 : Revue et Finilisation

Lors de ce sprint final, nous avons réalisé une revue complète des livrables pour nous assurer qu'ils répondaient aux attentes. Les visualisations ont été testées pour garantir leur exactitude et leur pertinence. Nous avons également validé les rapports documentés, incluant les processus suivis pour l'exploration et le nettoyage des données, ainsi que les résultats obtenus à chaque étape. Une attention particulière a été accordée à l'expérience utilisateur, avec des ajustements sur les graphiques pour les rendre plus intuitifs. Enfin, nous avons présenté les résultats au client et intégré leurs retours pour finaliser le projet.





Figur 50: Revue et Finilisation



Figur 51: Revue et Finilisation

## 6 Recommandations futures pour le projet de gestion de la production

- **Amélioration de la qualité des données :**

Bien que le nettoyage des données ait permis d'éliminer de nombreuses incohérences et erreurs, il est recommandé d'implémenter un processus continu de validation des données à chaque étape de leur collecte. Cela permettra de détecter automatiquement les anomalies et d'assurer la fiabilité des données sur le long terme. Un suivi régulier des données permettrait également d'éviter les ruptures ou lacunes dans les informations disponibles.

- **Automatisation des rapports :**

Pour améliorer l'efficacité et réduire les erreurs humaines, il serait bénéfique d'automatiser la génération de rapports de production et de performance. L'utilisation de solutions comme Power BI ou Tableau, en combinaison avec des scripts Python ou des processus ETL, pourrait permettre de générer des rapports en temps réel, directement extraits de la base de données, et d'envoyer des alertes automatiques en cas d'écarts significatifs par rapport aux objectifs.

- **Intégration des prévisions de production :**

Il est recommandé de développer des modèles prédictifs pour estimer la production future basée sur les données historiques. Cela permettrait de mieux planifier les ressources et d'optimiser la production en fonction des fluctuations attendues. L'intégration d'outils d'intelligence artificielle (IA) ou d'analyse prédictive aiderait à anticiper les besoins en matériel, en personnel et à réduire les coûts opérationnels.

- **Optimisation de la gestion des stocks :**

Une gestion plus fine des stocks serait bénéfique pour le projet. Il est recommandé d'implémenter des outils de suivi en temps réel des inventaires, avec des alertes automatiques lorsque les niveaux de stocks tombent sous un seuil critique. L'intégration avec des systèmes ERP pourrait améliorer cette gestion, permettant une synchronisation des informations entre la production et l'approvisionnement.

- **Mise en place d'un tableau de bord dynamique :**

L'implémentation d'un tableau de bord dynamique, avec des indicateurs clés de performance (KPI) et des alertes visuelles pour suivre l'évolution en temps réel des différentes métriques de production, offrirait une vue d'ensemble efficace pour les responsables et les équipes. Ce tableau de bord pourrait être mis à jour automatiquement pour fournir des informations précises sur les performances actuelles, la gestion des stocks, et les coûts de production.

- **Formation continue des équipes :**

Bien que le projet ait avancé avec succès, une attention particulière doit être portée à la formation continue des équipes de production. Le développement de sessions de formation sur l'utilisation des outils d'analyse de données, des systèmes de gestion de production, et des meilleures pratiques de maintenance des équipements permettra de garantir l'efficacité du processus sur le long terme.

- **Extension de l'analyse des coûts :**

Une analyse plus approfondie des coûts de production, y compris des coûts indirects comme l'entretien des équipements, la gestion des ressources humaines, et les coûts énergétiques, serait bénéfique pour optimiser les marges. Il serait utile de mettre en place un suivi détaillé des coûts à chaque étape du processus de production pour identifier les leviers d'optimisation.

- **Feedback des utilisateurs :**

Enfin, il serait précieux d'intégrer un système de feedback régulier de la part des utilisateurs (opérateurs, responsables de production) pour identifier des problèmes non détectés lors de la phase de tests. Une collecte continue de ce feedback aidera à améliorer l'interface utilisateur et l'ergonomie des outils de suivi, rendant ainsi l'outil plus efficace et plus adapté aux besoins des équipes.

## 7 Conclusion

Le projet de gestion de la production a permis d'atteindre plusieurs objectifs clés en matière de traitement et d'analyse des données. Grâce à l'exploration approfondie des tables de production et au nettoyage des données, nous avons pu fournir des bases solides pour la création de visualisations et de rapports de performance. Ce travail a été réalisé de manière itérative en suivant les principes de la méthode Agile, avec une planification par sprints qui a facilité l'avancement du projet tout en répondant aux besoins spécifiques de chaque étape.

Les visualisations créées, notamment les graphiques de performance de production et de gestion des stocks, ont fourni des aperçus précieux sur les opérations et les indicateurs clés de performance. Elles ont contribué à une meilleure compréhension des processus de production, tout en identifiant des zones d'amélioration possibles.

Cependant, malgré ces réalisations, plusieurs domaines nécessitent encore des améliorations pour optimiser davantage les processus. La mise en place d'outils de prévision, d'automatisation des rapports et une gestion améliorée des stocks représentent des pistes essentielles pour la réussite future du projet. En outre, l'analyse continue des données et des coûts de production permettra d'identifier de nouvelles opportunités d'optimisation et de réduire les dépenses opérationnelles.

En somme, ce projet a permis de jeter les bases d'une gestion de production plus efficace, mais il reste encore un potentiel important pour automatiser et améliorer la prise de décision, tout en garantissant un meilleur contrôle de la qualité des données.

## 8 Annexe

Vous pouvez consulter le projet sur le lien suivant :

- **Github:** <https://github.com/hajar0000/AdventureWorks-.git>
- **Taigo:** <https://tree.taiga.io/project/hajar0000-adventureworks-dataset/backlog>