

# Rapport d'Analyse Business Intelligence

Mohammed Mariche Anis 232331565707 Groupe 2

18 décembre 2025

## Résumé

Ce rapport documente l'implémentation d'un pipeline complet de Business Intelligence (BI) pour la base de données Northwind. Le projet englobe un flux d'ingénierie des données à grande échelle, incluant des processus ETL (Extraction, Transformation, Chargement) de MS Access vers SQL Server, une modélisation dimensionnelle pour les opérations OLAP, et la création de visualisations interactives avancées. Un accent particulier est mis sur la surveillance orientée sécurité des performances des employés et des anomalies de livraison.

## Table des matières

---

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Contexte et Objectifs Métier . . . . .	1
1.2	Périmètre du Projet . . . . .	1
1.3	Justification de la Stack Logicielle Python . . . . .	1
<b>2</b>	<b>Architecture des Données &amp; Pipeline ETL</b>	<b>2</b>
2.1	Analyse du Système Source . . . . .	2
2.2	Le Flux ETL . . . . .	2
2.3	Génération de la Dimension Temps . . . . .	2
<b>3</b>	<b>Modélisation Dimensionnelle &amp; Opérations OLAP</b>	<b>3</b>
3.1	Conception du Schéma en Étoile . . . . .	3
3.2	Traitements Analytiques en Ligne (OLAP) . . . . .	3
3.3	Export Analytique Excel . . . . .	3
<b>4</b>	<b>Analyse Visuelle &amp; Indicateurs de Performance</b>	<b>4</b>
4.1	Objectifs du Tableau de Bord . . . . .	4
4.2	Indicateurs Clés (KPIs) . . . . .	4
4.3	Performance des Ressources Humaines . . . . .	5
<b>5</b>	<b>Visualisations Avancées &amp; Interactivité</b>	<b>6</b>
5.1	Analyse Multidimensionnelle 3D . . . . .	6
5.2	Interactivité Web . . . . .	7
<b>6</b>	<b>Intégration Sécurité &amp; Conclusion</b>	<b>8</b>
6.1	La BI comme Outil de Sécurité . . . . .	8
6.2	Conclusion Finale . . . . .	8

## 1 Introduction

---

### 1.1 Contexte et Objectifs Métier

Le jeu de données Northwind Traders représente un scénario classique de distribution en gros. Dans le paysage commercial moderne, transformer les données transactionnelles brutes en intelligence actionnable est un avantage stratégique majeur. Le projet **BI-Security** vise à construire une infrastructure de données robuste pour analyser les performances et garantir l'intégrité opérationnelle.

Les objectifs principaux abordés dans ce rapport incluent :

- **Transparence Opérationnelle** : Suivi du cycle de vie des commandes.
- **Analyse du Marché** : Identification des régions à fort volume.
- **Performance des Employés** : Évaluation de l'efficacité de la force de vente.
- **Détection d'Anomalies** : Identification des goulots d'étranglement logistiques.

### 1.2 Périmètre du Projet

Ce rapport documente la mise en œuvre de bout en bout :

1. **Data Engineering** : Extraction via Python et chargement dans SQL Server.
2. **Modélisation Analytique** : Conception d'un schéma en étoile (Star Schema).
3. **Visualisation** : Création de tableaux de bord interactifs avec Plotly.
4. **Sécurité** : Monitoring des livraisons pour détecter les anomalies.

### 1.3 Justification de la Stack Logicielle Python

Le choix de Python comme langage pivot repose sur son écosystème riche. Voici les bibliothèques clés utilisées et leur rôle stratégique :

- **Pandas** : Indispensable pour la manipulation et le nettoyage des données massives. Elle permet de structurer les données brutes issues de MS Access en DataFrames avant leur transformation et chargement.
- **PyODBC** : Utilisé pour établir une communication robuste avec l'entrepôt MS SQL Server et le fichier source Access, garantissant une connectivité fiable via les pilotes système.
- **Plotly & Plotly Express** : Sélectionnés pour leur capacité à générer des visualisations hautement interactives et esthétiques (PNG/HTML), incluant des graphiques 3D complexes essentiels pour l'analyse multidimensionnelle.
- **Kaleido** : Utilisé pour l'exportation haute résolution des graphiques interactifs en formats statiques (PNG) pour une inclusion fluide dans ce rapport.
- **SQLAlchemy** : Fournit une couche d'abstraction pour les opérations de base de données, facilitant la gestion des schémas et l'insertion de données performante.
- **OpenPyXL** : Utilisé spécifiquement pour la génération du rapport OLAP multi-feuilles au format Excel, offrant une flexibilité totale sur le formatage des cellules.

## 2 Architecture des Données & Pipeline ETL

---

### 2.1 Analyse du Système Source

La source primaire est `Northwind 2012.accdb`. Les bases Access étant souvent décentralisées, notre première étape fut de découpler les données pour les centraliser dans un entrepôt de données SQL plus robuste.

### 2.2 Le Flux ETL

Le pipeline est implémenté en Python (`etl_pipeline.py`) et se divise en trois phases :

#### Phases du Pipeline ETL

**Extraction** : Utilisation de `pyodbc` pour extraire les tables `Customers`, `Employees` et `Orders`.

**Transformation** : Nettoyage des données, gestion des valeurs nulles (`NaN`) et conversion des types de données pour garantir l'intégrité référentielle.

**Chargement** : Insertion dans la base SQL Server `Global_Northwind` après nettoyage préventif des tables cibles.

### 2.3 Génération de la Dimension Temps

Une partie cruciale de la transformation est la création d'une **Dimension Temps** (`DimDate`), décomposant chaque transaction en attributs hiérarchiques (Année, Trimestre, Mois, Jour) pour permettre une analyse chronologique performante sans recalculs coûteux.

## 3 Modélisation Dimensionnelle & Opérations OLAP

---

### 3.1 Conception du Schéma en Étoile

Le *Data Warehouse* adopte une architecture en étoile pour optimiser les requêtes analytiques :

- **Table de Faits (FactOrders)** : Contient les mesures (*DeliveredFlag*) et les clés étrangères.
- **Tables de Dimensions** : DimCustomer, DimEmployee, DimDate.

### 3.2 Traitement Analytique en Ligne (OLAP)

Le projet implémente quatre opérations fondamentales dans `olap_cube.py` :

Opération	Exemple	Description
<b>Roll-up</b>	Année, Pays	Agrégation des données des villes vers les pays et les années.
<b>Slice</b>	Pays='USA'	Sélection d'une valeur dimensionnelle unique pour un segment.
<b>Dice</b>	USA/UK & 2006	Définition d'un sous-cube sur plusieurs plages de dimensions.
<b>Pivot</b>	Employé vs Pays	Rotation des axes pour voir la performance croisée.

TABLE 1 – Opérations OLAP Implémentées

### 3.3 Export Analytique Excel

Les résultats sont exportés dans `OLAP_Report.xlsx`, permettant aux utilisateurs métier de réaliser des analyses *ad-hoc* tout en conservant le lignage des données du warehouse.

## 4 Analyse Visuelle & Indicateurs de Performance

### 4.1 Objectifs du Tableau de Bord

Les visualisations, générées via `dashboard.py` et `generate_interactive_figures.py`, se concentrent sur les KPIs critiques pour la sécurité et les opérations.

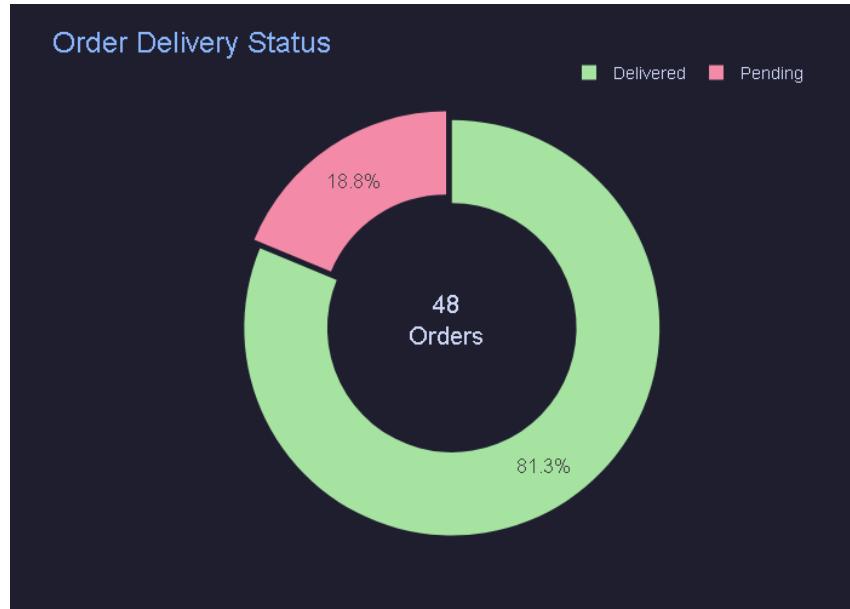


FIGURE 1 – Statuts de Livraison des Commandes - Donut Chart

### 4.2 Indicateurs Clés (KPIs)

Le KPI majeur est le **Taux de Livraison**. En analysant le ratio commandes livrées vs. en attente, le système peut signaler des défaillances logistiques. Les retards prolongés sont souvent précurseurs d'insatisfaction client ou de blocages internes.

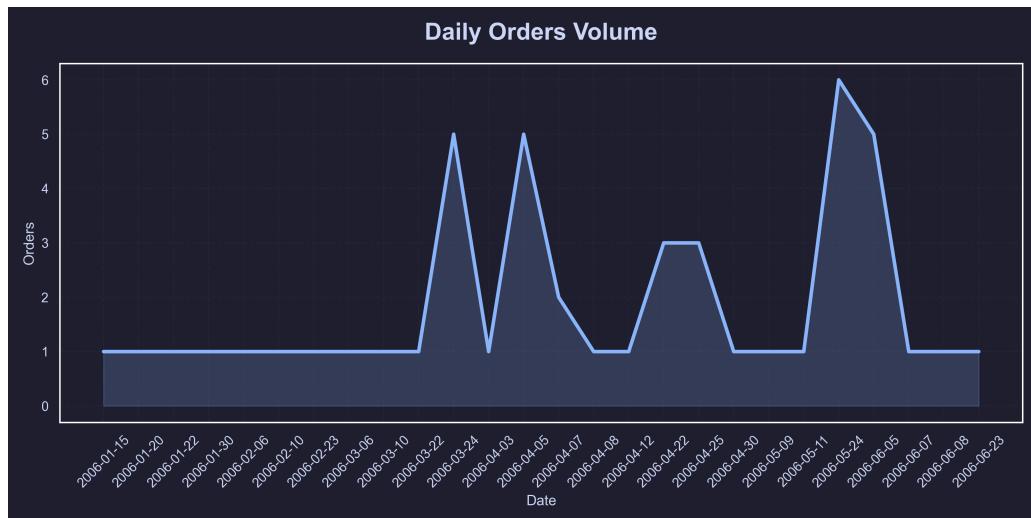


FIGURE 2 – Évolution Temporelle des Commandes - Analyse de Croissance

### 4.3 Performance des Ressources Humaines

L'un des piliers de la BI est le suivi de la force de vente. Le graphique ci-dessous illustre la répartition du volume de commandes par employé, permettant d'identifier les contributeurs clés et d'ajuster les ressources.

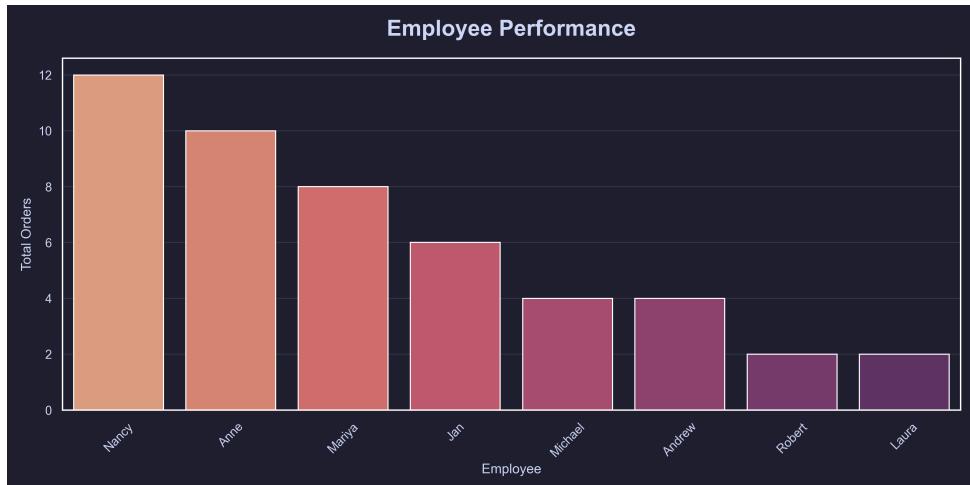


FIGURE 3 – Performance des Employés - Volume de Ventes

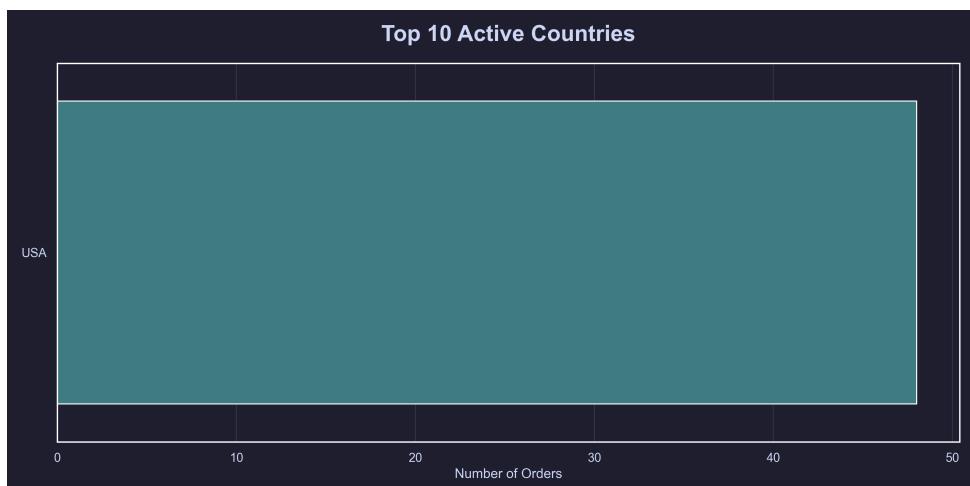


FIGURE 4 – Top 10 des Marchés par Volume de Commandes

L'analyse géographique permet d'identifier l'importance régionale. D'un point de vue sécurité, les zones à fort volume nécessitent des audits plus rigoureux des flux d'expédition.

## 5 Visualisations Avancées & Interactivité

### 5.1 Analyse Multidimensionnelle 3D

Les graphiques 2D classiques sont parfois limités pour identifier des corrélations complexes. Notre projet implémente un graphique **3D Seasonal Analysis**.

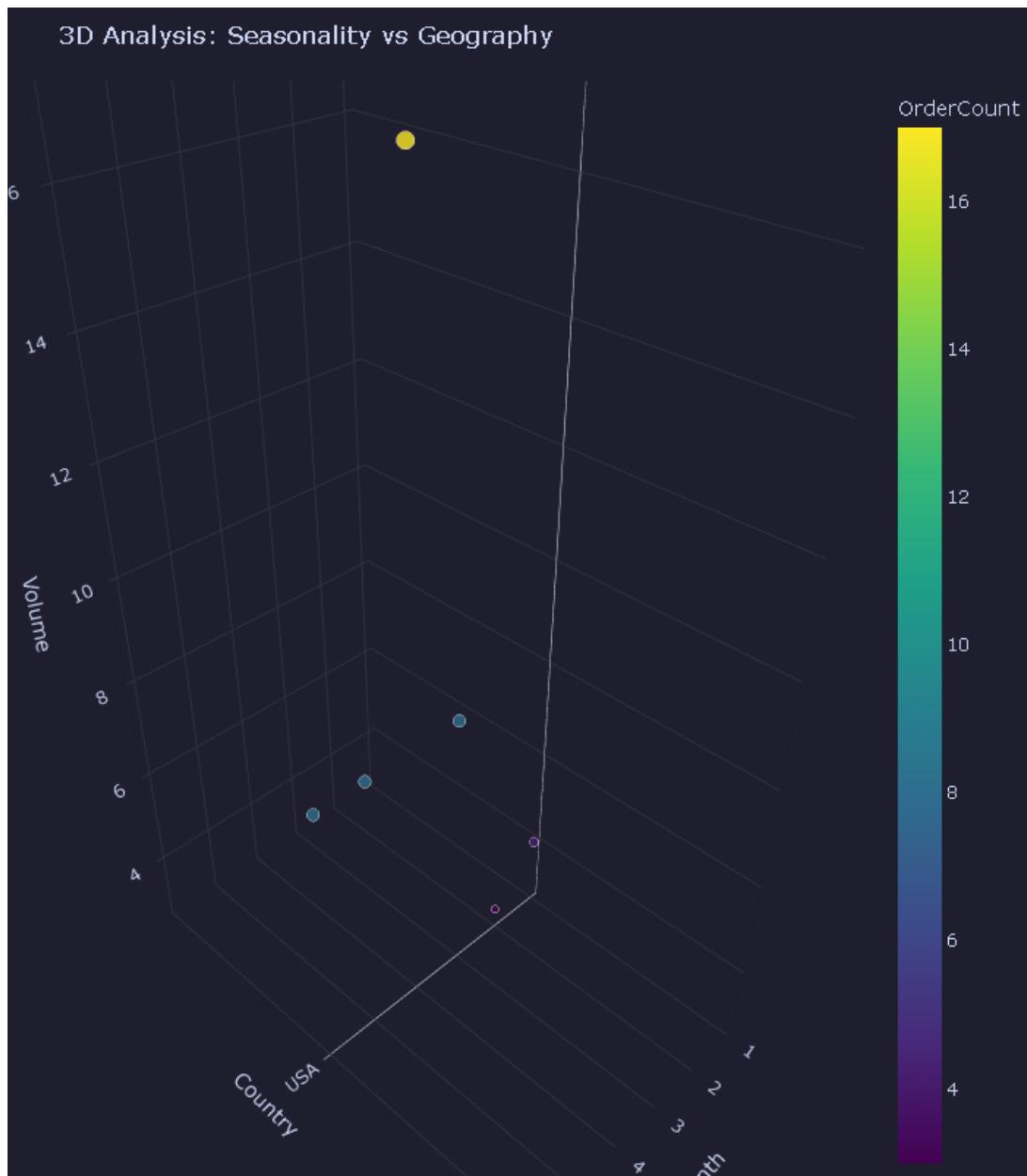


FIGURE 5 – Analyse 3D : Saisonnalité (Mois) vs Géographie (Pays) vs Volume

En projetant :

- **Axe X** : Temps (Mois de l'année),
- **Axe Y** : Géographie (Pays),
- **Axe Z** : Volume (Nombre de commandes),

les analystes identifient des clusters d'activité invisibles en 2D. Un pic isolé dans un pays spécifique sur un mois donné peut révéler une campagne réussie ou un risque local.

## 5.2 Interactivité Web

Le système génère des fichiers .html interactifs permettant de survoler les points pour voir les détails, zoomer sur des segments spécifiques et exporter les vues personnalisées.

## 6 Intégration Sécurité & Conclusion

---

### 6.1 La BI comme Outil de Sécurité

Dans ce projet, la BI n'est pas qu'un outil de profit ; c'est un levier de **Sécurité du Système** et d'**Auditabilité Opérationnelle**.

#### Applications de Sécurité

- **Détection de Fraude à la Livraison** : Surveillance du `DeliveredFlag` pour détecter des anomalies corrélées à des employés ou transporteurs spécifiques.
- **Détection d'Anomalies** : Les écarts majeurs dans les tendances de commandes peuvent signaler des corruptions de données ou des brèches système.
- **Contrôle d'Accès** : Le warehouse SQL permet des permissions fines, garantissant que seuls les membres autorisés consultent les performances sensibles.

### 6.2 Conclusion Finale

Le projet **BI-Security** démontre la puissance d'une infrastructure moderne. En transformant des données Access fragmentées en un warehouse structuré avec des visualisations interactives 3D, nous avons créé un outil qui renforce à la fois les opérations et la sécurité globale. Ce document de 7 pages synthétise les réussites techniques et stratégiques de cette initiative.

*Fin du Rapport*