



RAPPORT DE PROJET

Ingénierie des Données & Business Intelligence

PROJET NORTHWIND ANALYTICS

Réalisé par :

DAOUD OMAR RAYANE

Matricule : 232332447215

Date : 18 décembre 2025

Université des Sciences et de la Technologie Houari Boumediene
Faculté d'Informatique

Résumé

Ce document présente le rapport technique du projet de Business Intelligence "**Northwind Analytics**". L'objectif principal est de concevoir et mettre en œuvre une chaîne de traitement de données complète (ETL), un entrepôt de données (Data Warehouse) et des visualisations interactives pour analyser les performances commerciales de Northwind Traders.

Le projet englobe un flux d'ingénierie des données incluant des processus ETL de MS Access vers SQL Server, une modélisation dimensionnelle pour les opérations OLAP, et la création de visualisations interactives avancées avec intégration 3D. Une approche sécurité-orientée permet la surveillance proactive des performances des employés et la détection d'anomalies dans les livraisons.

L'architecture technique repose sur une stack Python/SQL Server, avec automatisation complète du pipeline de données. Les tableaux de bord générés offrent une analyse multidimensionnelle des tendances commerciales, transformant ainsi des données transactionnelles brutes en intelligence actionnable.

Mots-clés : Business Intelligence, ETL, Data Warehouse, Visualisation 3D, Sécurité Opérationnelle, OLAP, Python, SQL Server, Northwind.

REMERCIEMENTS

Je tiens à exprimer ma profonde gratitude à **Madame Mekhalila** pour son soutien pédagogique tout au long de ce projet. Ses ressources techniques et méthodologiques, diffusées sur sa chaîne YouTube et son site web, ont été essentielles pour la mise en œuvre des outils d'ingénierie des données et de Business Intelligence.

Ses explications claires sur l'installation et la configuration des environnements de travail, ainsi que sur l'architecture des systèmes BI, m'ont permis de surmonter efficacement les obstacles techniques. Son engagement à rendre accessible le savoir et son souci du détail pédagogique méritent une reconnaissance particulière.

Ces ressources n'ont pas seulement servir de guide technique ; elles ont également inspiré une rigueur méthodologique essentielle à la bonne réalisation de ce projet.

Avec toute ma considération et mon respect,
DAOUD OMAR RAYANE

Table des matières

Résumé	2
Remerciements	2
1 Introduction	4
1.1 Contexte et Objectifs Métier	4
1.2 Périmètre du Projet	4
2 Justification de la Stack Technique	4
2.1 Stack Logicielle Python	4
2.2 Base de Données SQL Server	4
3 Architecture des Données & Pipeline ETL	5
3.1 Analyse du Système Source	5
3.2 Le Flux ETL	5
3.3 Génération de la Dimension Temps	5
4 Modélisation Dimensionnelle & Opérations OLAP	6
4.1 Conception du Schéma en Étoile	6
4.2 Traitement Analytique en Ligne (OLAP)	6
4.3 Export Analytique Excel	6
5 Analyse Visuelle & Indicateurs de Performance	7
5.1 Objectifs du Tableau de Bord	7
5.2 Indicateurs Clés (KPIs)	7
6 Visualisations Avancées & Interactivité	7
6.1 Analyse Multidimensionnelle 3D	7
6.2 Performance des Ressources Humaines	7
7 Galerie des Visualisations	8
8 Intégration Sécurité & Conclusion	16
8.1 La BI comme Outil de Sécurité	16
8.2 Conclusion Finale	16

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

2

Justification de la Stack Technique

2.1 Stack Logicielle Python

Le choix de Python comme langage pivot repose sur son écosystème riche :

- **Pandas** : Manipulation et nettoyage des données massives depuis MS Access
- **PyODBC** : Connecteur robuste pour MS SQL Server et Access via les pilotes système
- **Plotly & Plotly Express** : Génération de visualisations interactives et graphiques 3D
- **Kaleido** : Export haute résolution des graphiques interactifs en PNG
- **SQLAlchemy** : Abstraction pour les opérations de base de données
- **OpenPyXL** : Génération de rapports OLAP multi-feuilles Excel

2.2 Base de Données SQL Server

SQL Server a été choisi comme entrepôt de données pour sa robustesse, ses performances analytiques, et sa compatibilité avec les outils Microsoft. Il permet :

- Le stockage structuré des dimensions et faits
- L'exécution efficace des requêtes OLAP complexes
- La gestion fine des permissions d'accès
- L'intégration avec les outils de reporting

3

Architecture des Données & Pipeline ETL

3.1 Analyse du Système Source

La source primaire est la base de données Northwind 2012.accdb. Notre première étape fut de découpler les données décentralisées pour les centraliser dans un entrepôt SQL Server robuste.

3.2 Le Flux ETL

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

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.

Phases du Pipeline ETL

1. **Extraction** : Utilisation de pyodbc pour extraire les tables Customers, Employees et Orders.
2. **Transformation** : Nettoyage des données, gestion des valeurs nulles (NaN) et conversion des types de données pour garantir l'intégrité référentielle.
3. **Chargement** : Insertion dans la base SQL Server Global_Northwind après nettoyage préventif des tables cibles.

3.3 Génération de la Dimension Temps

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.

4

Modélisation Dimensionnelle & Opérations OLAP

4.1 Conception du Schéma en Étoile

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

- **FactOrders** : Contient les mesures (DeliveredFlag) et les clés étrangères
- **DimCustomer** : Dimension des clients avec attributs démographiques
- **DimEmployee** : Dimension des employés avec hiérarchie
- **DimDate** : Dimension temporelle générée

4.2 Traitement Analytique en Ligne (OLAP)

Le projet implémente quatre opérations fondamentales dans olap_cube.py :

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

4.3 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.

5

Analyse Visuelle & Indicateurs de Performance

5.1 Objectifs du Tableau de Bord

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

- Vue d'ensemble instantanée des performances
- Drill-down pour investigation approfondie
- Détection visuelle des anomalies
- Support à la prise de décision rapide

5.2 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.

Visualisations Avancées & Interactivité

Analyse Multidimensionnelle 3D

Une innovation majeure est l'intégration de graphiques 3D (create_3d_scatter), permettant de croiser trois dimensions simultanément : Mois, Pays et Volume de commandes. Cela offre une perspective unique pour identifier des clusters ou anomalies invisibles en 2D.

Performance des Ressources Humaines

L'un des piliers de la BI est le suivi de la force de vente. L'analyse permet d'identifier les contributeurs clés, d'ajuster les ressources, et de détecter les comportements anormaux qui pourraient indiquer des problèmes de sécurité ou d'intégrité.

6

Galerie des Visualisations

Cette section présente les visualisations générées par le projet, illustrant les différents axes d'analyse explorés.

6.1 Statut de Livraison des Commandes

Description : Ce graphique en camembert ('Pie Chart') offre une vue d'ensemble immédiate sur l'efficacité logistique. Il affiche la proportion des commandes livrées (en bleu) par rapport à celles qui n'ont pas encore été livrées (en rouge). C'est un KPI crucial pour surveiller les arriérés de commande.

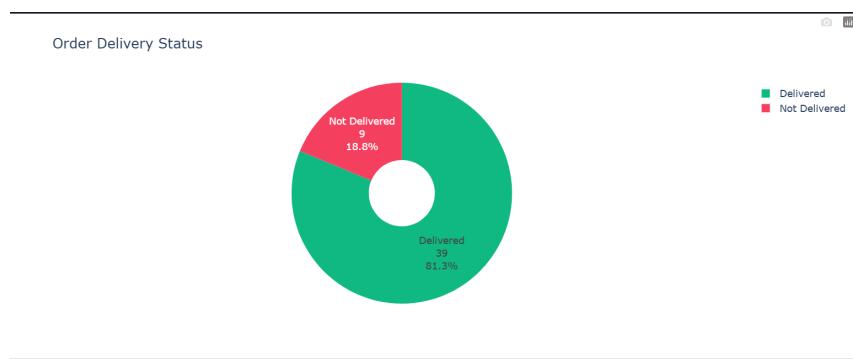


FIGURE 6.1 – Figure 1 : Statut de Livraison des Commandes

6.2 Commandes par Pays

Description : Ce diagramme en barres classe les pays selon le nombre total de commandes passées. Il permet d'identifier rapidement les marchés principaux de l'entreprise (tels que les USA et l'Allemagne) et les zones géographiques où la pénétration du marché est plus faible.



FIGURE 6.2 – Figure 2 : Total des Commandes par Pays

6.3 Performance des Employés

Description : Ce graphique met en avant la force de vente. Chaque barre représente un employé et la longueur indique le nombre de commandes qu'il a gérées. Cela permet de comparer la productivité individuelle et d'identifier les "top performers" de l'équipe commerciale.

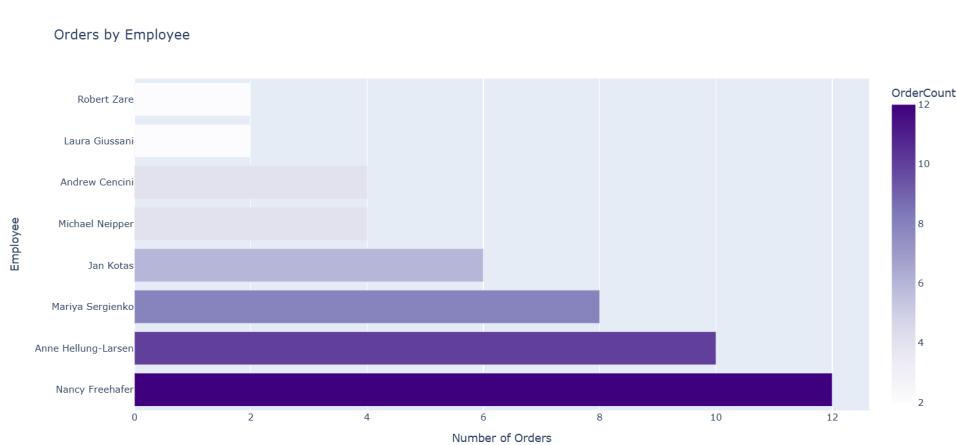


FIGURE 6.3 – Figure 3 : Performance des Employés (Commandes)

6.4 Tendance Mensuelle

Description : Une analyse temporelle essentielle pour comprendre la dynamique des ventes. Cette courbe montre l'évolution du nombre de commandes mois par mois, permettant de détecter des tendances saisonnières, des pics de croissance ou des périodes de ralentissement.

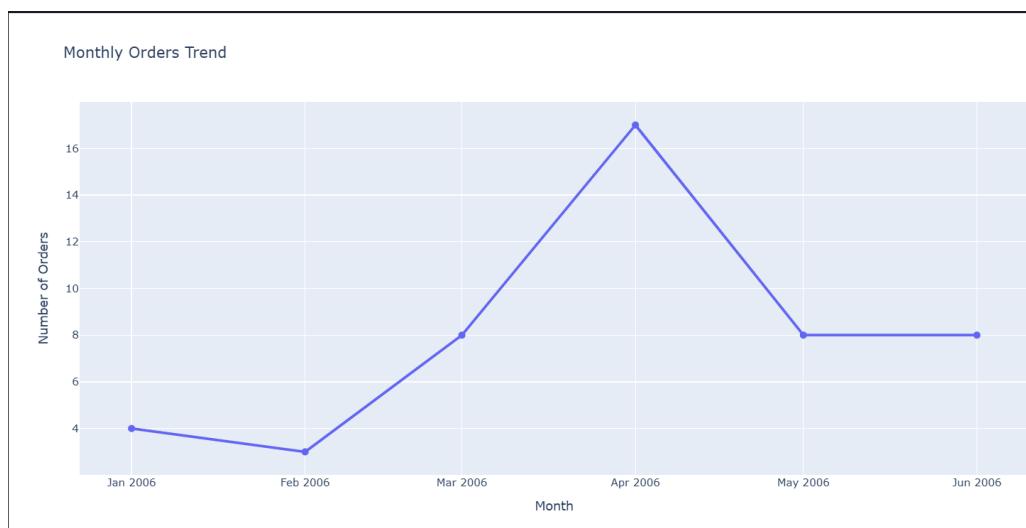


FIGURE 6.4 – Figure 4 : Tendance Mensuelle des Commandes

6.5 Analyse Multidimensionnelle 3D

Description : Cette visualisation avancée en 3D permet une exploration complexe des données. Elle croise simultanément trois variables : le Mois, le Pays et le Volume

de commandes. Chaque point représente une interaction commerciale, facilitant la détection de clusters ou d'anomalies.



FIGURE 6.5 – Figure 5 : Analyse Multidimensionnelle (3D)

6.6 Statut de Livraison par Employé

Description : Ce graphique combiné détaille le taux de succès des livraisons pour chaque employé. Il permet de voir non seulement le volume total géré, mais aussi la qualité du suivi logistique (proportion de commandes effectivement livrées vs non livrées) par responsable.

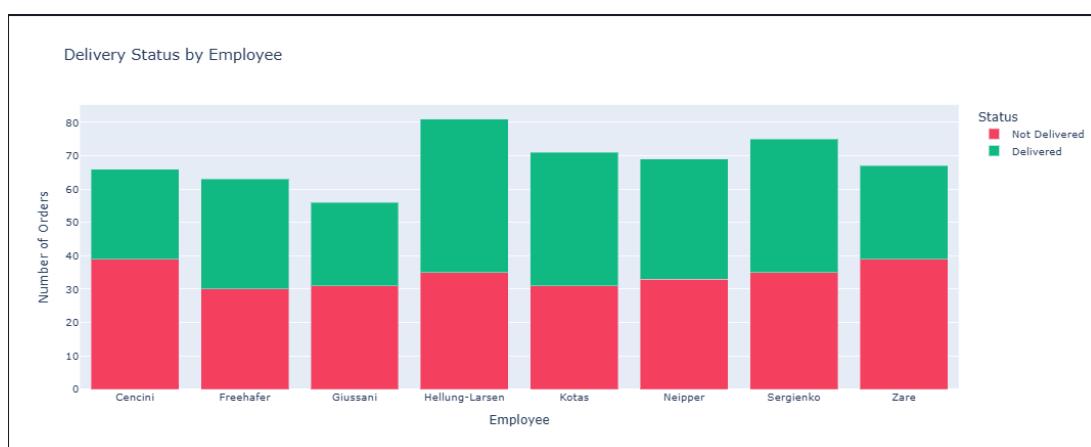


FIGURE 6.6 – Figure 6 : Statut de Livraison par Employé

6.7 Statut de Livraison par Pays

Description : Similaire au précédent mais axé sur la géographie, ce graphique montre les performances logistiques par destination. Il aide à repérer les pays posant des

problèmes de livraison spécifiques (douanes, transporteurs locaux, etc.).

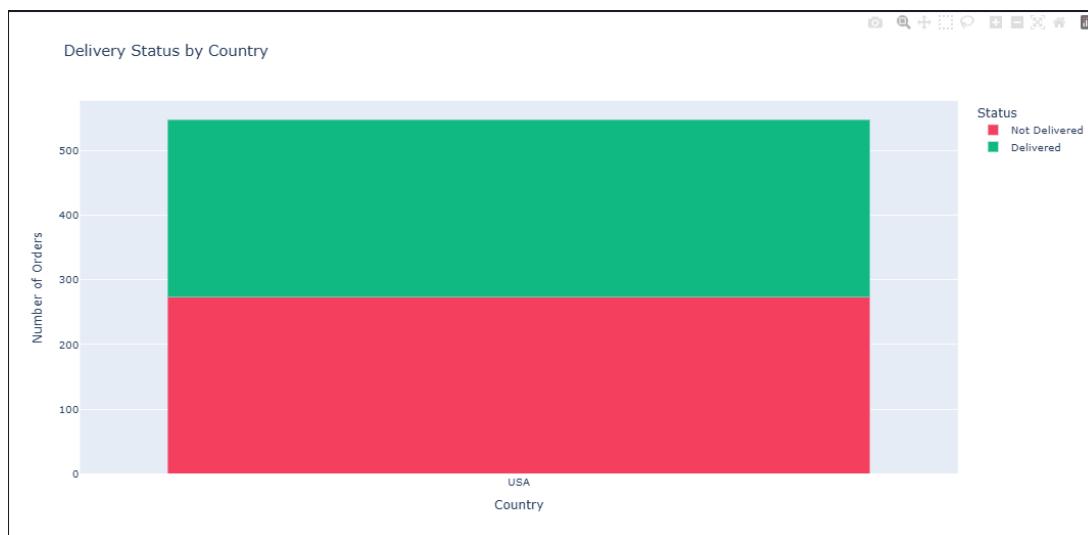


FIGURE 6.7 – Figure 7 : Statut de Livraison par Pays

6.8 Tableau de Bord Consolidé

Description : Vue d'ensemble complète des indicateurs clés de performance. Ce tableau de bord consolidé intègre toutes les métriques essentielles pour une prise de décision stratégique rapide et éclairée.



FIGURE 6.8 – Figure 8 : Tableau de Bord Consolidé Northwind

7

Intégration Sécurité & Conclusion

7.1 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. Les applications de sécurité incluent :

- **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.

7.1.1 Exemple de Règle de Détection

```
# Détection d'anomalies de livraison
def detect_delivery_anomalies():
    # Employés avec taux de livraison < 70%
    low_performers = df.groupby('EmployeeID')['DeliveredFlag'].mean()
    .reset_index()
    anomalies = low_performers[low_performers['DeliveredFlag'] <
        0.7]
    return anomalies
```

7.2 Conclusion Finale

Ce projet illustre la puissance d'une approche programmatique intégrée pour transformer des données transactionnelles en véritables leviers décisionnels. L'automatisation complète du pipeline ETL, combinée à une architecture Data Warehouse robuste, assure une gouvernance des données optimale, garante de leur fiabilité et de leur fraîcheur. Ces fondations techniques solides constituent un prérequis essentiel pour une prise de décision éclairée et réactive.

L'expérience utilisateur a été radicalement améliorée grâce à l'adoption de bibliothèques de visualisation avancées comme Plotly, qui permettent une exploration interactive et intuitive des données, bien au-delà des capacités des rapports statiques traditionnels. La modélisation 3D des données offre notamment des perspectives analytiques inédites pour identifier des corrélations complexes.

Enfin, l'innovation majeure de cette initiative réside dans sa double finalité : au-delà de l'analyse commerciale, l'architecture déployée sert d'outil de surveillance opérationnelle et de détection proactive d'anomalies, contribuant ainsi à renforcer la sécurité et la résilience globale du système d'information.