

RAPPORT DE PROJET BUSINESS INTELLIGENCE

Analyse des Opérations Import-Export de Northwind Traders

Projet : Solution BI pour Northwind Traders

Nom : Laraba

Prenom : Abdelaziz

Matricule : 232331444912

Year : 3ING Cyber security

Institution : USTHB – Faculté d’Informatique

Encadrant : Dr. MEKAHLIA Fatma Zohra

PLAN :

1-INTRODUCTION

2- PRÉSENTATION DU CONTEXTE

3-ARCHITECTURE DU PROJET

4- L' ETL

5-DÉVELOPPEMENT DU TABLEAU DE BORD

6-RÉSULTATS ET ANALYSES

7-PROBLÈMES RENCONTRÉS ET SOLUTIONS

8- CONCLUSION ET PERSPECTIVES

1. INTRODUCTION

1.1 Contexte du Projet

Northwind Traders est une société fictive spécialisée dans l'import-export de produits alimentaires.

Dans un environnement économique concurrentiel, l'entreprise souhaite améliorer ses capacités

d'analyse et de prise de décision. La mise en place d'une solution Business Intelligence permet

de centraliser les données, d'automatiser les traitements analytiques et de produire des indicateurs

fiables pour le pilotage stratégique.

1.2 Objectifs du Projet

Les objectifs de ce projet sont :

- Intégrer des données hétérogènes provenant de deux sources distinctes.
- Concevoir un Data Warehouse basé sur un modèle en étoile.
- Génération des cubes OLAP pour l'analyse multidimensionnelle.
- Développer un processus ETL automatisé en Python.
- Créer un tableau de bord interactif répondant aux besoins métiers.

1.3 Questions d'Analyse

Le système développé répond aux questions suivantes :

- Calcul du chiffre d'affaires commandé et réalisé par période, client, produit et catégorie.
- Application d'une taxe de 10 % sur les frais de transport supérieurs ou égaux à 500 \$.
- Analyse des commandes livrées et non livrées par client, employé et période.

1.4 Technologies Utilisées

Le projet utilise Python, Pandas, SQLAlchemy, Microsoft SQL Server, Streamlit, Plotly, Visual studio et Visual Studio Code.

2. PRÉSENTATION DU CONTEXTE

2.1 Sources de Données

Le projet repose sur deux sources principales :

- Base Northwind (20 tables CSV)
- Base SQL Server (13 tables CSV)

Ces sources présentent des différences de structure, de format et d'encodage.

2.2 Hétérogénéité des Données

Les principales différences concernent les noms de colonnes, les formats de date, la structure des champs et l'encodage des fichiers.

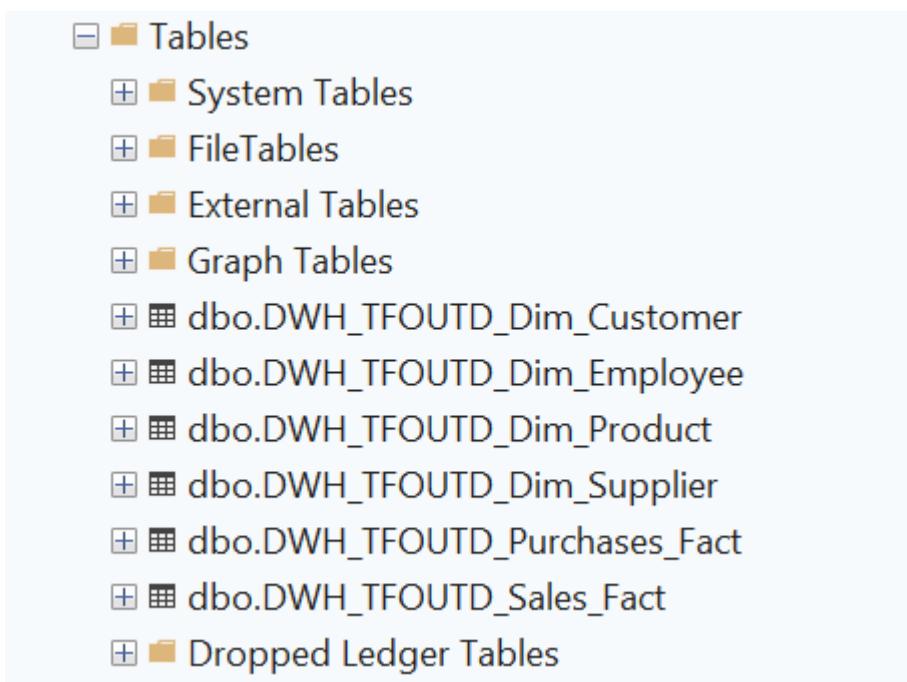
3. ARCHITECTURE DU PROJET

3.1 Architecture Globale

L'architecture adoptée suit une approche BI classique composée de quatre couches : sources de données, processus ETL, Data Warehouse et couche de visualisation.

3.2 Modèle en Étoile

Le Data Warehouse est organisé selon un schéma en étoile avec deux table de faits centrale et plusieurs tables de dimensions.



4. L' ETL :

PHASE 1 : EXTRACTION DES DONNÉES

L'extraction des données est réalisée à partir de fichiers CSV en utilisant Python.

Les données sont chargées dans des DataFrames Pandas en tenant compte de l'encodage et des valeurs manquantes.

```
def extract_csv(filename, source='northwind'):
    """
    Reads a CSV file from either data source.

    Args:
        filename: Name of the CSV file
        source: 'northwind' or 'sqlserver'
    """
    try:
        if source == 'northwind':
            path = os.path.join(NORTHWIND_PATH, filename)
        else:
            path = os.path.join(SQLSERVER_PATH, filename)

        df = pd.read_csv(path, encoding='latin1')

        if 'Unnamed: 0' in df.columns:
            df = df.drop(columns=['Unnamed: 0'])

        print(f"✓ [{source.upper()}] Extracted {len(df)} rows from {filename}")
        return df
    except Exception as e:
        print(f"✗ [{source.upper()}] Error reading {filename}: {e}")
        return None
```

```
=====
          NORTHWIND ETL PROCESS - DUAL SOURCE
          Source 1: Northwind DB (20 tables)
          Source 2: SQL Server DB (13 tables)
=====
```

PHASE 2 : TRANSFORMATION ET MODÉLISATION

Cette phase consiste à harmoniser les structures, résoudre les conflits d'identifiants, calculer les indicateurs métiers et construire les dimensions et tables de faits.

```
=====
ETL COMPLETE: 6 tables loaded successfully, 0 failed
=====

=====
DATA INTEGRATION SUMMARY
=====

📊 Final Table Counts:
• DWH_TFOUTD_Dim_Product: 122 records
• DWH_TFOUTD_Dim_Customer: 29 records
• DWH_TFOUTD_Dim_Employee: 28 records
• DWH_TFOUTD_Dim_Supplier: 39 records
• DWH_TFOUTD_Sales_Fact: 58 records
• DWH_TFOUTD_Purchases_Fact: 55 records
```

PHASE 3 : CHARGEMENT(LOADING) DANS LE DATA WAREHOUSE

Les données transformées sont chargées dans SQL Server.

L'intégrité référentielle est vérifiée avant l'insertion.

localhost (SQL Server 17.0.1000.7 - LAPTOP-2622QJAV\)

- ⊕ Databases
 - ⊕ System Databases
 - ⊕ Database Snapshots
 - ⊖ DWH_Northwind
 - ⊕ Database Diagrams
 - ⊖ Tables
 - ⊕ System Tables
 - ⊕ FileTables
 - ⊕ External Tables
 - ⊕ Graph Tables
 - ⊕ dbo.DWH_TFOUTD_Dim_Customer
 - ⊕ dbo.DWH_TFOUTD_Dim_Employee
 - ⊕ dbo.DWH_TFOUTD_Dim_Product
 - ⊕ dbo.DWH_TFOUTD_Dim_Supplier
 - ⊕ dbo.DWH_TFOUTD_Purchases_Fact
 - ⊕ dbo.DWH_TFOUTD_Sales_Fact
 - ⊕ Dropped Ledger Tables
 - ⊕ Views
 - ⊕ External Resources
 - ⊕ Synonyms
 - ⊕ Programmability
 - ⊕ Query Store
 - ⊕ Service Broker
 - ⊕ Storage

5. DÉVELOPPEMENT DU TABLEAU DE BORD

Un tableau de bord interactif est développé avec Streamlit et Plotly.

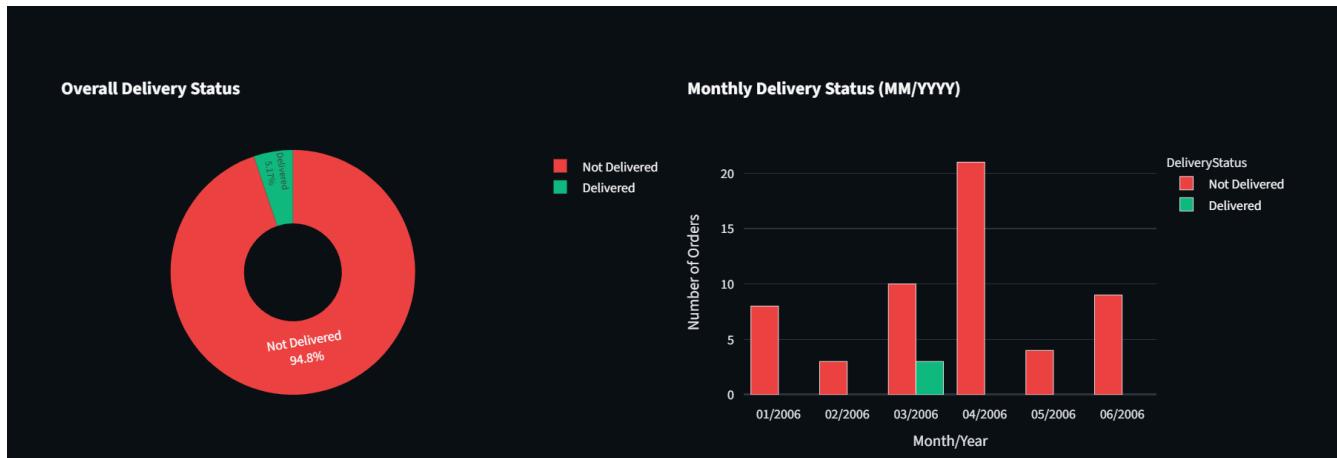
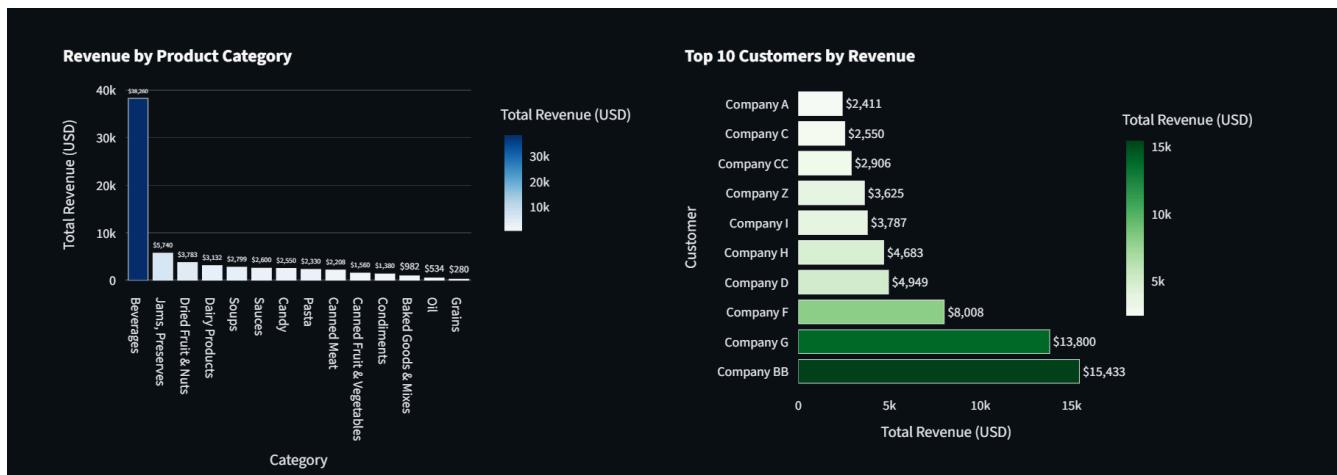
Il permet l'analyse des revenus, des commandes et des performances logistiques.

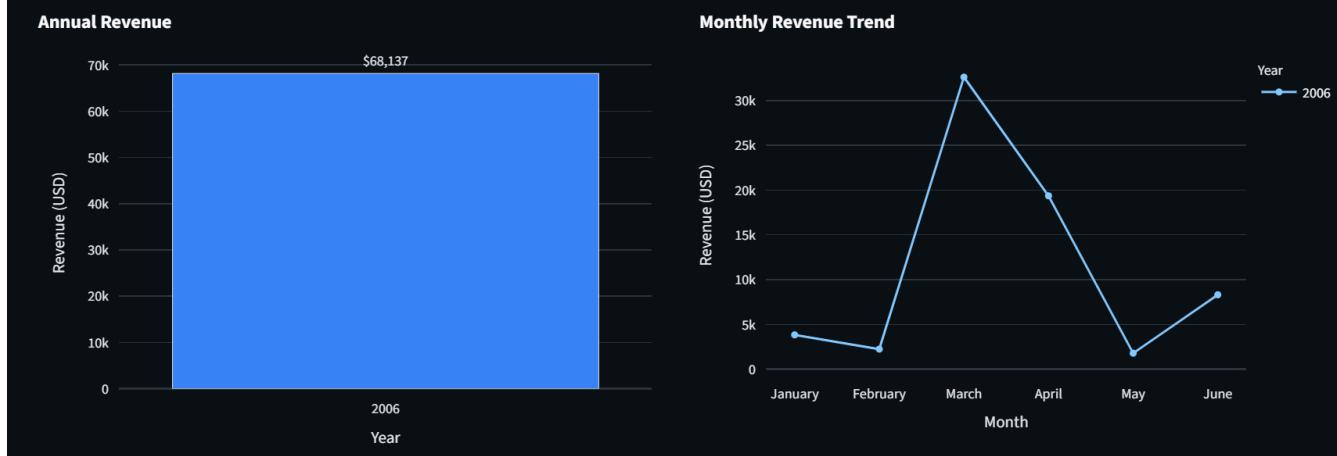
6. RÉSULTATS ET ANALYSES

Les résultats montrent une vision consolidée des ventes, une meilleure compréhension

des coûts logistiques et une distinction claire entre commandes livrées et non livrées.

Exemples :





7. PROBLÈMES RENCONTRÉS ET SOLUTIONS

1. INTÉGRATION DES DONNÉES MULTI-SOURCES

Problème : Données provenant de plusieurs sources (CSV, SQL Server, Excel) avec formats incohérents

Solution : Pipeline ETL modulaire avec connecteurs spécifiques et normalisation des formats

2. PERFORMANCES DES REQUÊTES

Problème : Requêtes lentes sur grandes volumétries, timeout du dashboard

Solution : Indexation stratégique, agrégations pré-calculées, mise en cache

3. SYNCHRONISATION DES RÉFÉRENTIELS

Problème : Codes produits/clients différents entre systèmes sources

Solution : Tables de correspondance et algorithmes de matching (fuzzy matching si nécessaire)

4. ÉVOLUTIVITÉ DU SCHÉMA

Problème : Changement de structure des données sources en production

Solution : Schéma flexible, versioning des tables, documentation des dépendances

5. CALCULS MÉTIER COMPLEXES

Problème : Règles de calcul du chiffre d'affaires, taxes, remises variables

Solution : Documentation formelle des règles, tests unitaires, validation par les métiers

6. MAINTENANCE DES DASHBOARDS

Problème : Demandes fréquentes de nouveaux KPIs et visualisations

Solution : Architecture modulaire, composants réutilisables, documentation

8. CONCLUSION ET PERSPECTIVES

Nous avons conçu et implémenté avec succès une solution Business Intelligence complète pour Northwind Traders, comprenant un pipeline ETL robuste, un data warehouse optimisé en schéma en étoile, et un tableau de bord interactif fournissant des insights actionnables sur les ventes, les livraisons et les opérations logistiques.