Rapport PI ERP-BI

Supply Chain Cosmetic

Spécialité : ERP-BI

Réalisé par : Data Wave

Proposé par : Esprit

Encadré par : ....

Année universitaire 2024/2025

Table des matières

1 Analysis and Specification **5**

1.1 Introduction . . . 5
1.2 Products and Services . . . 5
1.3 State of the Art . . . 5
1.3.1 Existing System Analysis . . . 5
1.3.2 Critique of the Existing System . . . 6
1.3.3 Project Presentation . . . 6
1.3.3.1 Proposed Solution . . . 6
1.3.3.2 Project Goal . . . 7
1.3.3.3 Business Objectives . . . 7
1.4 SDG 12: Responsible Consumption and Production . . . 8
1.4.1 1. Optimization of Stock and Waste Reduction . . . 8
1.4.2 2. Reducing the Carbon Footprint of the Supply Chain . . . 8
1.4.3 3. Improving the Traceability of Raw Materials . . . 8
1.5 Data Discovery . . . 9
1.5.1 Excel Files: Structured Data . . . 9
1.5.2 SSMS File (.bak): Structured Data . . . 10
1.5.3 JSON Files: Semi-Structured Data . . . 11
1.5.4 PDF Files: Semi-structured Data . . . 12
1.5.5 Data Discovery externe . . . 12
1.6 Agile BI Backlog . . . 14
1.6.1 Epic 1: Data Integration and Governance . . . 14
1.6.2 Epic 2: Data Modeling and Storage . . . 14
1.6.3 Epic 3: Visualization and Reporting . . . 15
1.6.4 Epic 4: Security and Performance . . . 15
1.7 Case Study: . . . 15
1.7.1 Context and Problem Statement . . . 15
1.7.2 Identified Problems . . . 15
1.7.3 Study Objectives . . . 16
1.7.4 Implemented Solution: . . . 16

1.7.4.1 Data Used 16
1.7.4.2 Key Performance Indicators (KPIs) 16
1.7.5 Obtained Results 16
1.7.6 Conclusion and Recommendations 17
1.8 System architecture diagram 17
1.9 Choice of Business Intelligence tools 18
1.9.1 Tools for creating the Data Warehouse 18
1.9.2 Tools for ETL (Extract, Transform, Load) 18
1.9.3 Tools for AI frameworks 19
1.9.4 Tools for data visualization 20
1.9.5 Programming languages and frameworks 20

Table des figures

1.1 Equivalence tree 8
1.2 System architecture Diagram 17
1.3 Logo SQL Server Management Studio 18
1.4 Logo Talend 19
1.5 Logo TensorFlow 19
1.6 Logo PyTorch 19
1.7 Logo Scikit-learn 20
1.8 Logo Power BI 20
1.9 Logo Flask 20
1.10 Logo Angular 21

Introduction

The cosmetics supply chain is a complex ecosystem involving numerous stakeholders, including suppliers, manufacturers, distributors, and end customers. Managing these interconnected processes efficiently is critical to ensuring product availability while controlling costs and maintaining customer satisfaction. However, inefficiencies in sourcing, production, and distribution can lead to challenges such as elevated operational costs, stockouts, and dissatisfied customers, impacting overall performance and competitiveness.

This project seeks to address these challenges by introducing advanced Business Intelligence (BI) and Artificial Intelligence (AI) solutions tailored to the cosmetics supply chain. By leveraging data from diverse sources such as inventory records, supplier performance metrics, production schedules, and sales, the project aims to provide actionable insights that enhance decision-making at all levels. These insights will enable stakeholders to optimize processes, reduce costs, prevent stockouts, and improve overall efficiency.

Through the integration of BI dashboards and AI-driven predictive models, the proposed solution will empower the supply chain to adapt dynamically to fluctuating market demands and operational constraints. This approach not only minimizes inefficiencies but also creates a foundation for proactive and informed management across the entire supply chain network.

Chapter 1 Analysis and Specification

1.1 Introduction

This chapter represents Sprint 0 in the methodology followed for the project. During this sprint, we specify the functional and non-functional requirements of our decision-making system to ensure a clear and precise design.
Additionally, we outline the BI tools selected and utilized in the project to achieve the defined objectives.

1.2 Products and Services

The cosmetics supply chain involves delivering a wide range of products and services to meet diverse market needs. These include sourcing raw materials, managing production processes, optimizing inventory levels, ensuring timely distribution to retailers, and providing efficient customer service. Each step of the supply chain requires careful coordination to minimize costs, reduce inefficiencies, and enhance customer satisfaction.

1.3 State of the Art

This section is essential to identify existing challenges and demonstrate the importance of the proposed solution for optimizing the cosmetics supply chain.

1.3.1 Existing System Analysis

Before initiating this project, it is necessary to analyze the current state of the supply chain. The primary processes and tools currently in use include:

- Static reports generated for supply chain monitoring.
- Data filtering and visualization performed using Excel.
- Separate systems used as data sources, leading to fragmented insights.
- Use of raw, unstructured data for decision-making.

1.3.2 Critique of the Existing System

The main limitations of the current system can be summarized as follows:

- Dependence on traditional tools like Excel for data visualization and filtering, which lack interactivity and scalability.
- Absence of advanced visualization tools that offer dynamic and rich graphical representations.
- Use of raw and uncleaned data, which affects the accuracy and reliability of insights.
- Lack of an ETL (Extract-Transform-Load) tool for efficient data integration and cleaning.
- Discrepancies in data quality between different source tables.
- Need for centralized data integration to consolidate insights across the supply chain.

1.3.3 Project Presentation

To address the challenges identified, this project aims to develop a dynamic and efficient Business Intelligence (BI) solution to optimize the cosmetics supply chain.

1.3.3.1 Proposed Solution

The proposed solution focuses on integrating advanced BI and AI technologies to optimize the cosmetics supply chain.

By leveraging data from various sources such as inventory records, supplier performance metrics, production schedules, and sales, the solution will provide real-time insights to improve decision-making.

BI dashboards will offer stakeholders a comprehensive view of the entire supply chain, enabling them to make informed decisions at every level. AI-driven predictive

models will be employed to forecast demand, optimize inventory levels, and prevent stockouts, ensuring product availability while reducing costs.

1.3.3.2 Project Goal

The objectives of the proposed project are:

- Define and monitor key performance indicators (KPIs) to measure supply chain performance.
- Design a centralized system for monitoring and reporting supply chain activities.
- Develop automated tools for generating supply chain reports, enabling continuous improvement and decision-making.

1.3.3.3 Business Objectives

Our solution is a decision support system designed to enhance the efficiency of the cosmetic supply chain by providing comprehensive insights beyond traditional inventory and logistics reports. It aims to identify hidden optimization opportunities and streamline operations across all levels, from suppliers to distributors and decision-makers.

Our mission is to establish a structured decision-making process by collecting, analyzing, and leveraging both internal and external data to extract valuable insights. This is achieved through the implementation of key monitoring indicators that enable better forecasting, inventory management, and supplier selection.

The core objectives of our project are structured in a hierarchical format to ensure clarity, operational efficiency, and strategic alignment.

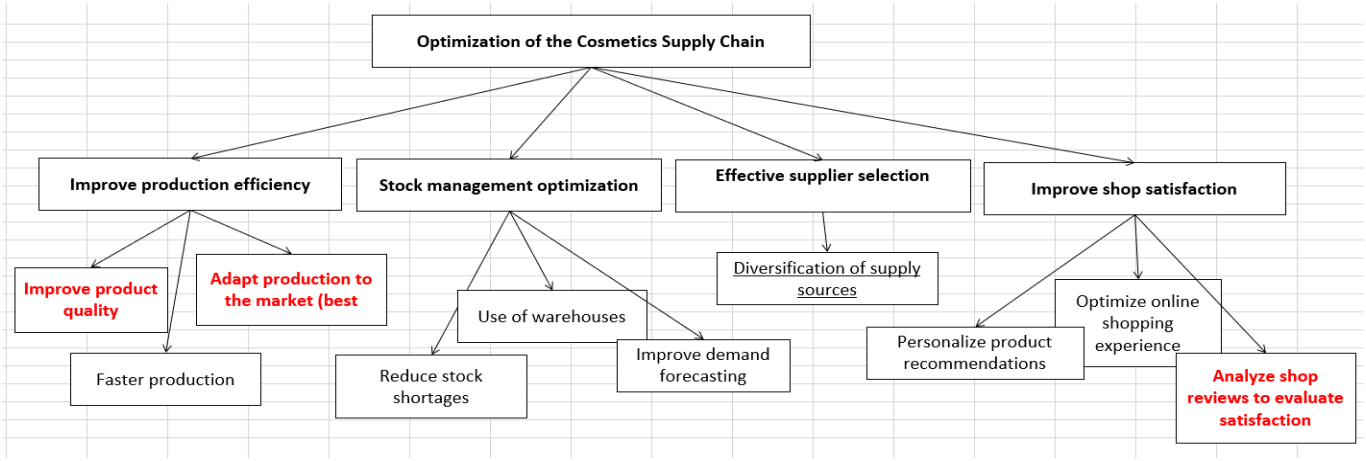

Figure 1.1: Equivalence tree

1.4 SDG 12: Responsible Consumption and Production

Sustainable Development Goal (SDG) 12 is particularly relevant as it aims at optimizing resources, reducing waste, and improving supply chain efficiency.

1.4.1 1. Optimization of Stock and Waste Reduction

- **KPI:** Stock turnover rate, level of unsold products, percentage of expired products.
- **BI Action:** Trend analysis to anticipate overstock and shortages.

1.4.2 2. Reducing the Carbon Footprint of the Supply Chain

- **KPI:** CO2 emissions from transport, percentage of local suppliers.
- **BI Action:** Identification of short supply chains and more ecological logistics alternatives.

1.4.3 3. Improving the Traceability of Raw Materials

- **KPI:** Percentage of recycled/recyclable materials used.
- **BI Action:** Monitoring of supply sources and identification of responsible partners.

1.5 Data Discovery

In the data discovery phase, we analyze and categorize the data sources that will be used to populate the data warehouse. These sources include both structured and semi-structured data, which come from various formats such as Excel files, JSON files, PDF documents, and SSMS backups. Below, we categorize and describe the data sources in detail.

1.5.1 Excel Files: Structured Data

- **Cosmetic Products Table**: This table contains information about the cosmetic products, including the base materials used in their production.

Field	Description
product_id	A unique identifier for the product.
product_name	The name of the product.
category	The category the product belongs to.
brand_name	The name of the brand.
material_id	A unique identifier for the material.
material_name	The name of the material.
material_category	The category the material belongs to.
dosage	The amount of material used in the product.

Table 1.1: Cosmetic Products Table

- **Inventory Table**: This table contains details about the stock levels of products in warehouses, including quantities available.

Field	Description
warehouse_id	A unique identifier for the warehouse.
warehouse_name	The name of the warehouse.
location	The location of the warehouse.
product_id	A unique identifier for the product.
quantity	The quantity of the product in the warehouse.

Table 1.2: Inventory Table

- **Production Table**: This table records the production dates of the cosmetic products, along with their IDs and names.

Field	Description
product_id	A unique identifier for the product.
product_name	The name of the product.
production_start_time	The start time of production.
production_end_time	The end time of production.

Table 1.3: Production Table

- **Base Materials Table with Units**: This table includes information about the materials used, along with their respective units of measurement.

Field	Description
material_id	A unique identifier for the material.
material_name	The name of the material.
material_category	The category the material belongs to.
unit	The unit of measurement for the material.

Table 1.4: Base Materials Table with Units

1.5.2 SSMS File (.bak): Structured Data

- **Warehouses Table**: This table contains information about the warehouses, including their names, locations, and storage capacities.

Field	Description
warehouse_name	The name of the warehouse.
location	The location of the warehouse.
capacity	The capacity of the warehouse.

Table 1.5: Warehouses Table

- **Suppliers Table**: This table contains information about the suppliers, such as their names, addresses, contacts, and emails.

Field	Description
supplier_id	A unique identifier for the supplier.
supplier_name	The name of the supplier.
contact	The contact person at the supplier.
email	The email address of the supplier.
address	The address of the supplier.
city	The city where the supplier is located.
country	The country where the supplier is located.

Table 1.6: Suppliers Table

- **Shops Table**: This table includes information about the shops, including their names and contact details.

Field	Description
shop_id	A unique identifier for the shop.
shop_name	The name of the shop.
contact	The contact person at the shop.
email	The email address of the shop.
address	The address of the shop.
city	The city where the shop is located.
country	The country where the shop is located.

Table 1.7: Shops Table

1.5.3 JSON Files: Semi-Structured Data

- **Cosmetic Products Table (JSON)**: This table contains information about the cosmetic products, including unit price, manufacture date, and expiration date.

Field	Description
product_id	A unique identifier for the product.
product_name	The name of the product.
category	The category the product belongs to.
brand_name	The name of the brand.
unit_price	The price per unit of the product.
manufacture_date	The date the product was manufactured.
expire_date	The date the product expires.
base_material	The base materials used in the product.

Table 1.8: Cosmetic Products Table (JSON)

- **Orders Table (JSON)**: This table includes data on orders, detailing the products ordered, their quantities, and the shipping address.

Field	Description
order_id	A unique identifier for the order.
order_date	The date the order was placed.
shop_id	A unique identifier for the shop.
shipping_address	The address to which the order is shipped.
products	A list of products in the order, each with:
product_name	The name of the product.
quantity	The quantity of the product ordered.
product_id	A unique identifier for the product.

Table 1.9: Orders Table (JSON)

1.5.4 PDF Files: Semi-structured Data

- **Factures**: The invoices, initially unstructured PDF files, were processed with OCR tools to create structured tables containing details about purchased products, quantities, unit prices, and totals.

Field	Description
ref_facture	The reference number of the invoice.
date	The date of the invoice.
magasin	The name of the shop.
fournisseur	The supplier.
tableau	A table containing:
produit	The product.
quantité	The quantity of the product.
prix_unitaire	The unit price of the product.
total_général	The total amount.

Table 1.10: Factures

1.5.5 Data Discovery externe

- **Supplier**: This table contains information about cosmetic product suppliers, extracted via web scraping from **Xometry**. It includes details such as supplier name, location, and estimated revenue.

Field	Description
Supplier Title	Name of the supplier company.
Location	Country or region where the supplier is based.
Revenue	Estimated revenue of the supplier.

Table 1.11: Supplier

- **Comments**: This table contains user reviews collected via web scraping from **Sephora** and **Sheglam**, as well as additional customer feedback data extracted from **Kaggle**. It includes customer comments on cosmetic products, ratings, and publication dates.

Field	Description
userName	Name of the user who posted the comment.
comment	Review written by a user.
rating	Rating given by the user (1 to 5 stars).
comment_date	Date when the comment was posted.

Table 1.12: Comments

- **Best Sellers**: This table contains top-selling cosmetic products from **Sephora**, collected via web scraping. It identifies the most popular products based on the number of sales for the current season.

Field	Description
Brands	Product's brand.
Products	Name of the cosmetic product.
prices	Price of the product in dollars/euros.
reviews	Number of reviews received by the product.
rating	Average rating of the product based on customer reviews.

Table 1.13: Best Sellers

- **Most Used Cosmetic Products**: This table includes the most-used cosmetic products based on data from **Kaggle**. It ranks items by their frequency of use in the available dataset.

Field	Description
Category	Type of cosmetic product.
Usage_frequency	Number of times the product is used in the dataset.
Gender_Target	Gender for which the product is primarily marketed (Male, Female, Unisex).
Packaging_Type	Type of packaging used (bottle, jar, tube).
Main_Ingredient	Key ingredient in the product (hyaluronic acid, retinol).
Cruelty_Free	Indicates if the product is cruelty-free (Yes/No).

Table 1.14: Most Used Cosmetic Products

1.6 Agile BI Backlog

1.6.1 Epic 1: Data Integration and Governance

ID	User Story	Priority	Acceptance Criteria
BI-1	As a BI Analyst, I want to collect and clean raw data through an automated ETL pipeline to ensure data quality.	High	Data is extracted from multiple sources (ERP, CRM, API). Deduplication and handling of null values.
BI-2	As a BI Administrator, I want to enforce data governance rules to ensure compliance and traceability.	High	Implementation of roles and permissions. Logging of modifications and audits.

1.6.2 Epic 2: Data Modeling and Storage

ID	User Story	Priority	Acceptance Criteria
BI-3	As a BI Analyst, I want to structure a dimensional data model (star schema) to optimize analytical queries.	Medium	Creation of facts and dimensions. Indexing for performance improvement.
BI-4	As a BI Administrator, I want to ensure the scalability of the Data Warehouse by optimizing partitions and indexes.	Medium	Query response time < 2 seconds for 1M+ records.

1.6.3 Epic 3: Visualization and Reporting

ID	User Story	Priority	Acceptance Criteria
BI-5	As a BI Analyst, I want to create interactive dashboards in Power BI to track KPIs in real time.	High	Integration of interactive charts. Automatic data updates.
BI-6	As a Finance Manager, I want a sales forecast report based on AI to better plan budgets.	Medium	Machine learning algorithm applied to sales history.

1.6.4 Epic 4: Security and Performance

ID	User Story	Priority	Acceptance Criteria
BI-7	As a BI Administrator, I want to enable multi-factor authentication to secure data access.	High	Implementation of MFA (Google Auth, OTP, etc.).
BI-8	As a BI Analyst, I want to optimize dashboard loading times for a better user experience.	Medium	Reduce loading time to < 3 seconds.

1.7 Case Study:

Supply Chain Optimization in the Cosmetic Industry with BI

1.7.1 Context and Problem Statement

Radiance Cosmetics, a company specializing in beauty products, faces challenges in its supply chain. Issues in inventory management and raw material supply lead to frequent stock shortages, long delivery times, and increased logistics costs.

1.7.2 Identified Problems

1. Poor demand visibility: insufficient or excessive stock levels.
2. Unreliable suppliers: variable delivery times.
3. High storage costs: inefficient product turnover.
4. Environmental impact: high carbon footprint due to inefficient logistics.

1.7.3 Study Objectives

- Reduce stock shortages through predictive analytics.
- Optimize supplier selection using performance KPIs.
- Improve inventory management to lower costs.
- Automate demand forecasting with BI and Machine Learning.

1.7.4 Implemented Solution:

Business Intelligence for Supply Chain

1.7.4.1 Data Used

- Sales history (3 years)
- Average supplier delivery time
- Real-time stock levels
- Logistics costs and carbon footprint

1.7.4.2 Key Performance Indicators (KPIs)

- Supplier reliability (delays, delivery compliance)
- Inventory turnover rate
- Reduction in storage and transportation costs
- Optimization of replenishment (optimal stock level)

1.7.5 Obtained Results

- 35% reduction in stock shortages thanks to BI forecasts.
- 25% improvement in supplier reliability with precise KPIs.
- 20% reduction in storage costs due to optimized orders.
- 15% reduction in carbon footprint by optimizing logistics routes.

1.7.6 Conclusion and Recommendations

- Generalization of BI to optimize the entire supply chain.
- Advanced automation of replenishment based on sales trends.
- AI integration for more precise predictive demand analysis.

1.8 System architecture diagram

Business Intelligence (BI) is the set of tools and methods for transmitting relevant information to business managers. Its goal is to help decision-makers understand their environment and support them in making good decisions

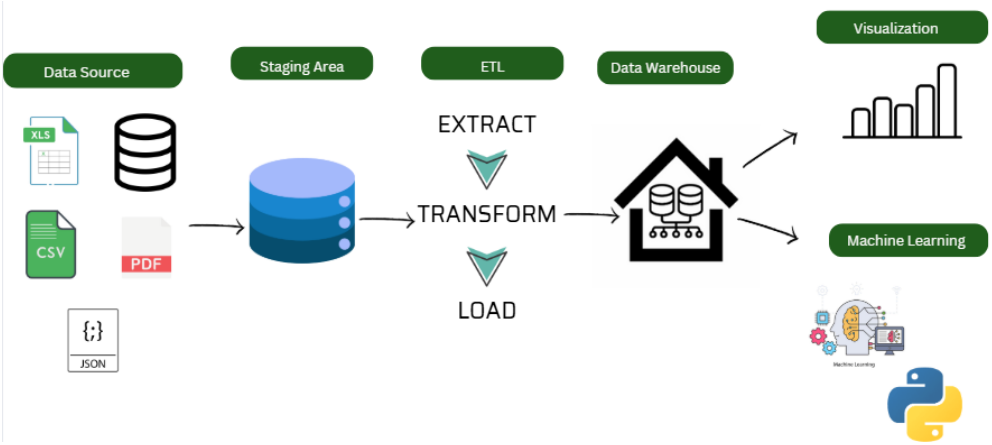

Figure 1.2: System architecture Diagram

- Collect, clean and consolidate data, extract data from production and adapt them to decision making
- Analysis: identifying patterns and establishing relationships in a group of data
- Distribute or rather facilitate the accessibility of information according to the functions and types of use.
- Reporting: the process of accessing data, formatting it and delivering it inside and outside the organization

1.9 Choice of Business Intelligence tools

Existing companies face a significant accumulation of data, which becomes challenging to process and understand. Here, the role of BI software is highlighted, as well as the reasons why businesses adopt it. BI represents the strategy and technology utilized by companies to analyze commercial data and information.

ETL tools allow the generation of clean, easily accessible data that can be effectively used for analytics and decision-making purposes.

In this section, we will outline the various BI tools selected for our project. These tools are divided into five categories:

- Tools for creating the Data Warehouse
- Tools for ETL (Extract, Transform, Load)
- Tools for AI frameworks
- Tools for data visualization
- Programming languages and frameworks

1.9.1 Tools for creating the Data Warehouse

SQL Server Management Studio (SSMS) :is an integrated development environment (IDE) for managing SQL Server databases. It was developed by Microsoft and provides a graphical interface for creating, modifying, and executing SQL queries.The figure 1.3 provides an overview of the SQL Server Integration Services logo.

Figure 1.3: Logo SQL Server Management Studio

1.9.2 Tools for ETL (Extract, Transform, Load)

Talend : is an integrated development environment (IDE) for data integration, offering a graphical interface for designing ETL processes. It simplifies the extraction, transformation, and loading (ETL) of data from various sources, with support for multiple connectors. The figure 1.4 provides an overview of the Talend logo.

Figure 1.4: Logo Talend

1.9.3 Tools for AI frameworks

TensorFlow : is an open-source library for machine learning and artificial intelligence, developed by Google. It provides a comprehensive ecosystem for building and deploying machine learning models, particularly deep learning models. The figure 1.5 provides an overview of the TensorFlow logo.

Figure 1.5: Logo TensorFlow

BeautifulSoup : A Python library designed for web scraping by efficiently parsing and navigating HTML and XML documents. It provides simple methods to extract, search, and manipulate web data, making it useful for tasks like data mining and content extraction from static web pages

Figure 1.6: Logo PyTorch

Selenium :A powerful automation framework primarily used for testing web applications but also widely utilized for web scraping. Unlike static parsers, Selenium can interact with dynamic web pages, execute JavaScript, handle user inputs, and navigate websites, making it ideal for scraping data from sites that require interactions like clicking, scrolling, or filling forms. The figure 1.7 provides an overview of the Selenium logo.

Figure 1.7: Logo Scikit-learn

1.9.4 Tools for data visualization

Power BI : is a business analytics tool developed by Microsoft, enabling users to visualize and share insights from their data. It provides a range of interactive reports and dashboards, with seamless integration to various data sources. The figure 1.8 provides an overview of the Power BI logo.

Figure 1.8: Logo Power BI

1.9.5 Programming languages and frameworks

Flask : is a lightweight web framework for Python, designed for building web applications quickly and with minimal code. It offers flexibility and simplicity, allowing developers to scale projects as needed. The figure 1.9 provides an overview of the Flask logo.

Figure 1.9: Logo Flask

Angular : is a platform and framework for building single-page applications using HTML and TypeScript. Developed by Google, it offers powerful tools for building dynamic and responsive web applications. The figure 1.10 provides an overview of the Angular logo.

Figure 1.10: Logo Angular

Conclusion Générale

the end

BIBLIOGRAPHIE

[1] link