MDSAA

Master Degree Program in
Data Science and Advanced Analytics

Business Intelligence

RETAIL4ALL - Final Delivery

Catarina Silva
Moeko Mitani
Oumayma Ben Hfaiedh
Sarah Leuthner

Group 4

Fabric Workspace: “2025 BI Project - Group 4”

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

INDEX

1. INTRODUCTION

In today's competitive retail landscape, data-driven decision-making is essential for business success. As a dynamic retail company, Retail4All specializes in sourcing and distributing a wide range of high-quality products, including electronics, clothing, sports equipment, and home essentials. To enhance operational efficiency and sales performance, the company recently upgraded its selling system with online-friendly services.

This report presents the second phase of the Business Intelligence project. Building on the foundation established in the first delivery, the initial phase focused on achieving a solid business understanding, identifying key business problems, and formulating targeted business questions related to sales analysis. It also addressed early-stage challenges and proposed solutions, laying the groundwork for effective analytical exploration. A thorough data understanding process led to the implementation of the Kimball Lifecycle Methodology, culminating in a star schema dimensional model tailored for efficient sales performance analysis.

In this second delivery, the focus shifts toward the practical implementation of the data architecture and analytics pipeline. Using the same Microsoft Fabric Workspace 2025 BI Project - Group 4 from the first phase, a Data Warehouse was developed, following a successful ETL process. The orchestration of this pipeline was achieved through a dedicated pipeline component, while dataflows were employed to carry out critical transformations and data loading operations.

Advanced analytics were incorporated through the use of a Notebook, which applied predictive techniques to generate new insights in the form of additional columns or tables now integrated into the Data Warehouse. The ETL processes reflected a diverse set of transformation techniques, such as Merge, Conditional Column, and Group By, demonstrating the practical application of skills acquired during laboratory sessions.

Together, these components represent a significant step forward in Retail4All's journey toward a scalable, insight-driven data infrastructure, designed to support strategic growth and operational excellence.

2. BUSINESS NEEDS AND REQUIRED OUTCOME

2.1 Business Understanding

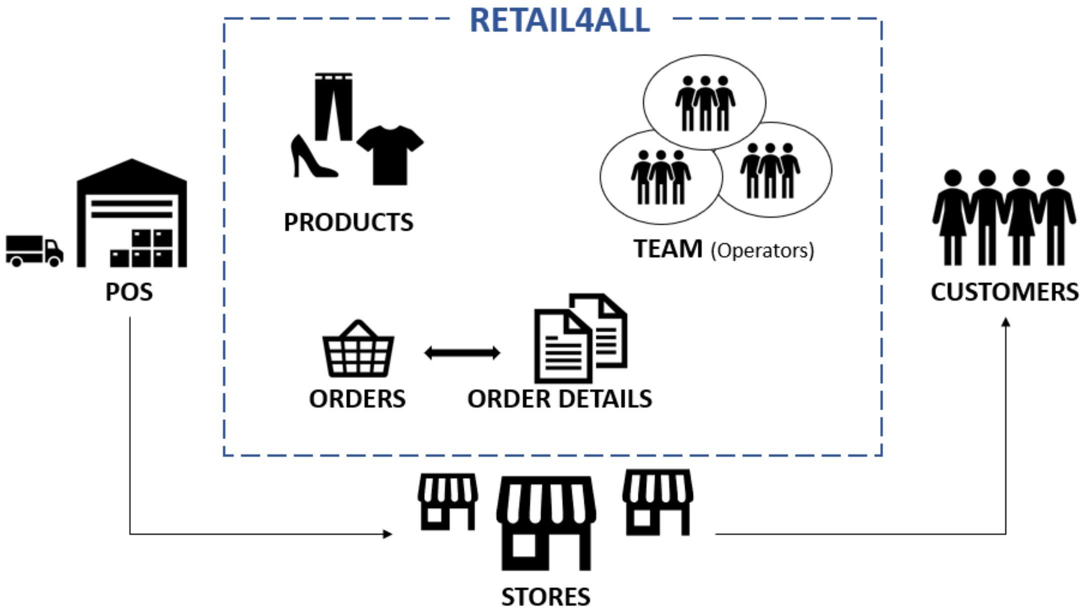


Figure 1 – Business Schema

Retail4All is a Portuguese retail company specializing in the sale of a diverse range of products, including electronics, clothing, fitness equipment, and home essentials. Rather than manufacturing its own products, the company strategically partners with suppliers to source high-quality goods, which are then distributed from the Point of Supply (POS) to third-party retailers. Customers can buy the products online or in one of many stores in Portugal.

The company operates with a lean and efficient structure, consisting of three specialized teams and a total of 15 employees, responsible for overseeing sourcing, logistics, processing, and product distribution to customers.

Committed to innovation and operational excellence, Retail4All continuously seeks to leverage the latest advancements in retail technology and supply chain management, having recently upgraded their selling system in order to include online-friendly services. The organization’s primary objective is to enhance its analytical capabilities, enabling data-driven decision-making to optimize sales performance, streamline operations, and drive business growth.

2.2 Business Problem

To improve decision-making process, Retail4all is looking for six solutions that enable the analysis of:

1. **Limited Sales Visibility**: The company lacks a centralized system to track and analyze sales trends across different time periods (monthly, quarterly, yearly). This makes it difficult to identify patterns and seasonality in sales performance.

2. **Inefficient Demand Forecasting**: Without accurate sales quantity analysis, Retail4All may struggle to predict future demand, leading to stock shortages or excess inventory.
3. **Lack of Comparative Performance Insights**: The company needs to compare current-year sales with previous years to understand growth patterns, identify potential declines, and adjust strategies accordingly.
4. **Unstructured Customer Data**: Retail4All does not have a structured approach to profiling its customers, which limits personalized marketing, targeted promotions, and customer retention strategies.
5. **Limited Transaction-Level Insights**: Without detailed transaction-level analysis, it is difficult to understand purchasing behavior, identify high-value customers, and detect anomalies such as fraudulent transactions as well as performance metrics per operator.
6. **Suboptimal Product and Store Performance Analysis**: The company lacks a system for aggregating sales by store, category, and subcategory, making it challenging to evaluate which locations or product lines are performing well and which need improvement.

2.3 Business Questions

Business Needs / Questions	Metric	Entities
What are the accumulated sales across different time periods (month, quarter, yearly)?	• Sales amount	• Time period
What are the sales quantities across different time periods (month, quarter, yearly)?	• Sales quantity	• Time period
What are the sales amount and quantities in comparison to the previous year?	• Sales amount • Sales quantity • Customer number	• Year
What are the top 3 products bought by store and location across different time periods (month, quarter, yearly)?	• Product amount	• Store ID • Location • Product ID • Product (sub-) category • Time period
What are the top 3 sellers (operator) by year?	• Sales amount	• Operator • Year
What are the aggregated sales by product category and product subcategory?	• Sales amount	• Product Category • Product subcategory
NOTE: Another objective is the security aspect of the report: Report accessibility by location and user later in Power BI in the following project.		

Table 1 – Business Questions

3. DATA UNDERSTANDING

The retail database consists of **seven** interconnected tables, each storing essential information about various aspects of retail operations in Portugal. These tables include **Locations, Points of Supply, Operators, Products, Stores (A & B), and Sales.** Below is a detailed description of each table:

3.1 LOCATIONS TABLE

The **Locations** table stores geographical information about various locations in Portugal. This table helps in associating sales and stores with specific locations. It consists of **four columns:**

- **Location ID** *(integer, primary key)* – A unique identifier for each location, assigned incrementally.
- **City** *(string)* – The name of the city.
- **District** *(string)* – The district where the city is located.
- **Country** *(string)* – The country, which in this case is Portugal.

3.2 POINT OF SUPPLY TABLE

The **Point of Supply** table contains details about inventory vendors supplying products to stores. This table is linked to sales transactions to track which vendors supplied the sold products. It includes **three columns:**

- **POS ID** *(integer, primary key)* – A unique identifier for each point of supply, assigned incrementally.
- **Name** *(string)* – The vendor's name.
- **Email** *(string)* – The vendor's contact email.

3.3 OPERATORS TABLE

The **Operators** table stores information about the staff members managing retail operations. This table enables tracking of sales and operations by linking transactions to specific operators. It consists of **seven columns:**

- **Operator ID** *(integer, primary key)* – A unique identifier for each operator, assigned incrementally.
- **First Name** *(string)* – The first name of the operator.
- **Last Name** *(string)* – The last name of the operator.
- **Email** *(string)* – The email address of the operator.
- **Gender** *(string)* – The gender of the operator.
- **Role** *(string)* – The job title or function of the operator.
- **Team Number** *(integer)* – The team ID to which the operator is assigned.

3.4 PRODUCTS TABLE

The **Products** table contains details about the different products sold in stores. It consists of **four columns:**

- **SKU (Stock Keeping Unit)** *(string, primary key)* – A unique identifier for each product.
- **Product Name** *(string)* – The name of the product.
- **Category** *(string)* – The category of the product (e.g., electronics, clothing, groceries).
- **Subcategory** *(string)* – A more specific classification under the main category.
  This table ensures proper product classification and tracking of sales based on product types.

3.5 STORES TABLES (STORE A & STORE B)

The **Stores** tables (Store A and Store B) store information about physical and online retail locations. These tables help distinguish between physical and online retail operations. Each table consists of **four columns:**

- **Store ID** *(integer, primary key)* – A unique identifier for each store, assigned incrementally.
- **Store Name** *(string)* – The name of the store.
- **Location ID** *(integer, foreign key from Locations table)* – The geographical location of the store.
- **Online** *(boolean)* – Indicates whether the store operates online (TRUE) or is a physical store (FALSE).

3.6 SALES TABLE

The **Sales** table captures transaction details for every purchase made. This table serves as the central point for analyzing sales performance, tracking revenue, and monitoring inventory. It consists of **ten columns:**

- **Sale ID** *(integer, primary key)* – A unique identifier for each sale, assigned incrementally.
- **Datetime** *(datetime)* – The timestamp of the transaction.
- **SKU** *(string, foreign key from Products table)* – Identifies the product sold.
- **Store ID** *(integer, foreign key from Stores table)* – The store where the sale occurred.
- **POS ID** *(integer, foreign key from Point of Supply table)* – The point of supply linked to the transaction.
- **Location ID** *(integer, foreign key from Locations table)* – The location associated with the sale.
- **Quantity (QTY)** *(integer)* – The number of units sold in the transaction.
- **Amount** *(decimal)* – The total amount paid by the customer for the transaction.
- **Operator ID** *(integer, foreign key from Operators table)* – The staff member responsible for the sale.

3.7 Obstacles and Solutions

Obstacle	Solution
Creating new tables produced errors	Column names are invalid, spaces and non UTF-8 symbols were manually removed for uploading in the Lakehouse
Creating dimensional tables for location and store would lead to a snowflake-schema due to dependencies	A star-schema is presevered by merging columns. In the dimensional table for the stores, an additional colums for the location is created

Table 2 – Obstacles and Solutions (Delivery 1)

4. METHODOLOGY – KIMBALL LIFECYCLE

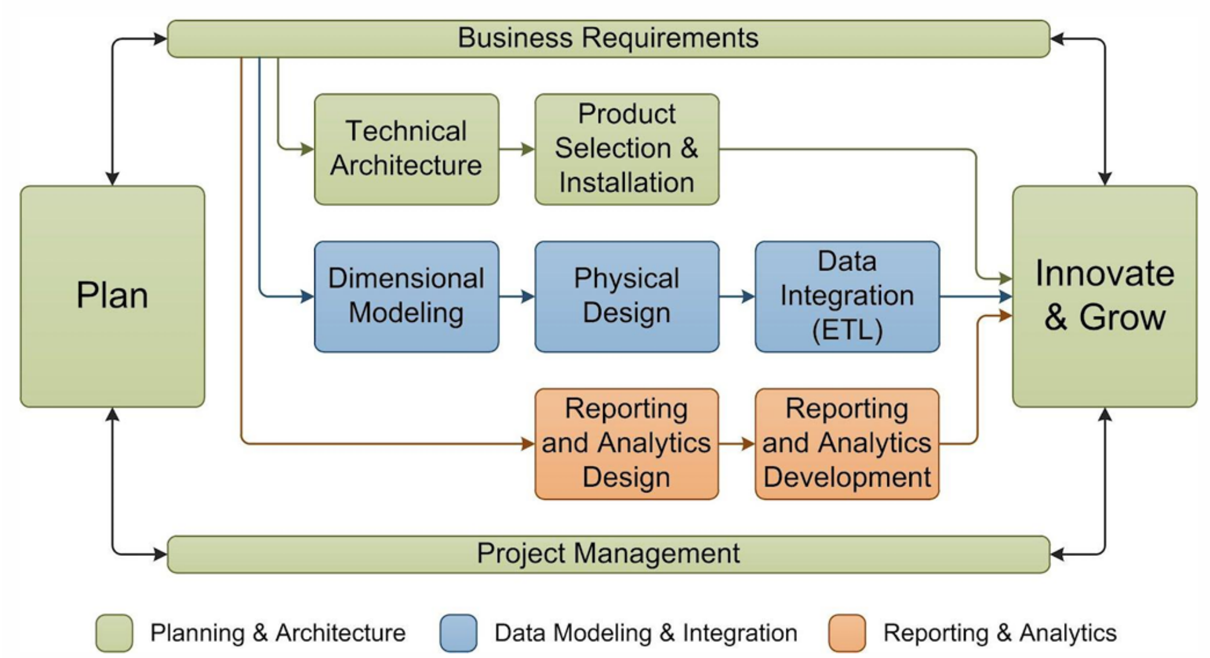


Figure 2 – Kimball Lifecycle

In this project and report the focus will be on the “Data Modeling & Integration phase” (blue). Dimensional modeling is a database design technique conceived specifically for data warehousing, enabling an efficient retrieval-based system that supports high-volume query access. Unlike transactional databases (OLTP), which focus on updates and complex relationships, dimensional modeling focuses on readability and performance for analytical queries. Introduced by Ralph Kimball in 1996, this approach structures data into facts (measurable events like sales) and dimensions (contextual attributes like product, store, and date). By contextualizing facts with relevant dimensions, it boosts business intelligence, making it easier for users to analyze trends and take informed insights.

5. DIMENSIONAL MODEL

The dimensional model developed in this project for Retail4All follows the Kimball Dimensional Modeling Methodology. The key characteristics of the Kimball Methodology reflected in this model include the use of a star schema, with a central fact table that stores sales transactions (*FACT_Sales*) surrounded by five dimensions that provide descriptive attributes for analysis (*DIM_Operator, DIM_Product, DIM_POS, DIM_Date, DIM_Stores*). Since Retail4All stores three years of data, the Kimball approach is suitable as it allows trend analysis over time. Additionally, the model is designed to facilitate the business requirements, such as analyzing sales performance per month, quarter, and year, comparing current and previous years, and aggregating sales by store, product category, and subcategory.

The influence of Moody & Kortink's Dimensional Modeling Methodology is also noticeable in this dimensional model. Hierarchies are incorporated across all dimensions to facilitate multi-level analysis and data aggregation. The *Date* dimension follows a natural hierarchy of *Year* → *Month* → *Day*, allowing time-based analysis at different granularities. The *Product* dimension is structured with a *Category* → *Subcategory* → *Product* hierarchy, facilitating product-level and category-level insights. Similarly, the *Store* dimension includes a geographic hierarchy of *Country* → *District* → *City*, for a better regional sales analysis, while the *Operator* dimension captures organizational structure through *Team* → *Role* → *Operator*. In terms of granularity, meaning the lowest level of detail, the following dimensions are defined: Sales * Stores * POS (Point of Supply) * Product * Date * Operator

This dimensional model is designed to empower decision-making by offering a well-structured, easy-to-query Data Warehouse that supports historical analysis and business intelligence needs. The Kimball and Moody & Kortink's Methodologies ensure that Retail4All can efficiently overview sales performance, optimize business operations, and gain meaningful insights into customer purchasing behaviors.

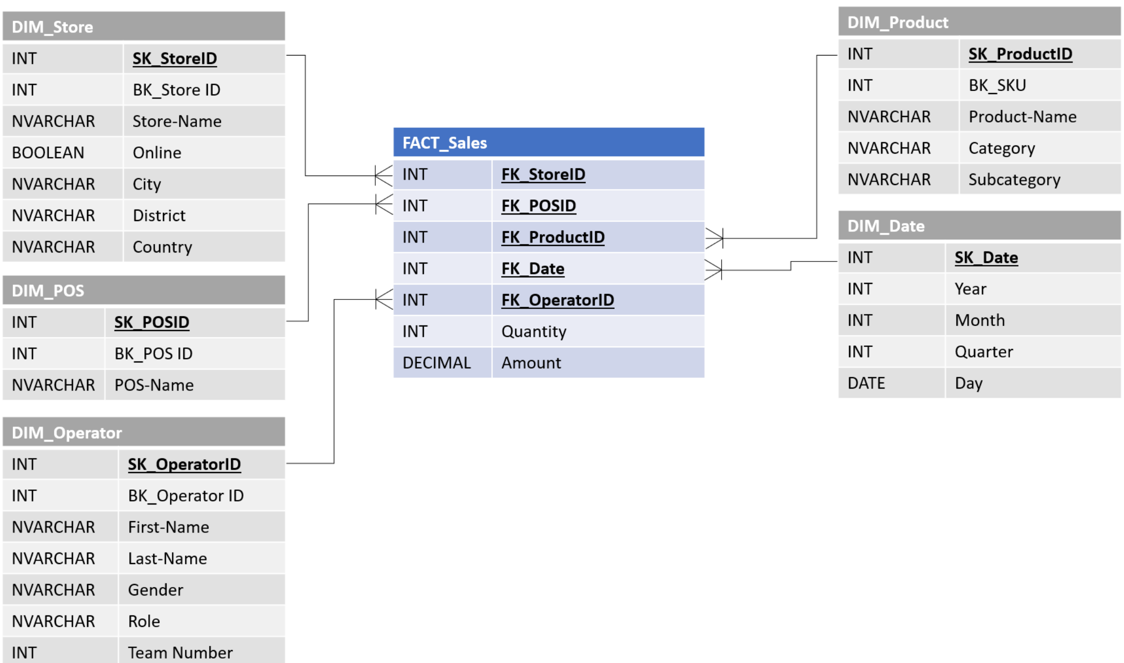


Figure 3 – Dimensional Model

Fact Table (2 measures)	Dimensions (5 dimensions)	Hierarchies (3 hierarchies, 3 levels of depth)
FACT_Sales (quantity, amount)	*DIM_Operator* *DIM_Product* *DIM_POS* *DIM_Date* *DIM_Stores*	Year > Month > Day Category > Subcategory > Product Name Country > District > City > Store

Table 3 – Summary of the Properties of the Fact Table

The *FACT_Sales* table serves as the core of the model, capturing key sales metrics like quantity and amount. It is linked to various dimension tables by foreign keys, facilitating a multi-perspective analysis of sales transactions. The *DIM_Operator* table provides details about Retail4All's sales operators, including their names, emails, roles, and teams, allowing performance tracking both at an individual and team level. The *DIM_Product* table contains product-related attributes such as product name, category, and subcategory, enabling category-level sales analysis. The *DIM_POS* table holds information about points of sale, crucial for the analysis of sales performance across different sales channels. The *DIM_Date* table provides a structured approach for time-based analysis, allowing businesses to assess sales trends on a daily, monthly, quarterly, and yearly basis. Lastly, the *DIM_Stores* table contains store-specific attributes, including store name, online presence, city, district, and country, making it possible to compare sales performance between different locations and online-versus-offline stores.

6. FIRST DELIVERY CONCLUSION

The present project aims to demonstrate the power of data-driven decision-making in Retail4All's operations. By leveraging a structured dimensional model based on the Kimball methodology, a more efficient framework for analyzing sales performance was established, demand forecasting, customer profiling, and store-level information. Key business challenges were identified, transforming raw transactional data into meaningful insights that drive strategic decision-making. Through the integration of well-defined fact and dimension tables, Retail4All will be more easily able to identify sales trends across time, compare historical data, and optimize inventory management. The hierarchical structures incorporated into the model ensure an informed and layered analysis, facilitating executive and operational decisions.

In the next phase of the project, we will focus on implementing a thorough ETL pipeline to automate data loading and transformation, followed by advanced reporting and analytics to further enhance Retail4All's decision-making process. Our ultimate goal is to provide an instructive Power BI report that gives stakeholders important insights, always bearing in mind the improvement of sales performance and customer satisfaction.

7. ETL (EXTRACT, TRANSFORM, AND LOAD)

7.1 Introduction

This second delivery of our Business Intelligence project focuses on the real development and deployment of Retail4All's Data Warehouse. Building upon the foundations established in the first delivery, our team designed and implemented a robust ETL process using Microsoft Fabric tools - more specifically, Pipelines, Dataflows, and Notebooks. Our main goal was to create a scalable and analysis-ready Data Warehouse, following the Kimball methodology, while integrating all previous learnings into a coherent system that allows for effective reporting and predictive analytics.

In this part of the report, we present the full structure of Retail4All's Data Warehouse, a detailed explanation of our ETL setup, a thorough description of each transformation step applied, and our use of advanced analytics techniques to enhance forecasting accuracy. Additionally, we explored more complex ETL techniques not covered in class to demonstrate initiative and expand the project's analytical capabilities.

7.2 Data Warehouse Description

As previously mentioned, our Data Warehouse was designed using the Kimball methodology, serving as a central repository that integrates, cleanses, and stores historical sales data from Retail4All. It follows a star schema model, with a central Fact Table (Fact_Sales) surrounded by 5 Dimension Tables that provide context for the business.

Table Name	Description
Fact_Sales	Sales amount.
Dim_Date	Contains time-based records, including information about the specific date of the purchase (e.g., Is_Special, Weekday_Type, etc.).
Dim_Product	Holds attributes like product name, category, and subcategory, allowing product-level insights and aggregation for analysis.
Dim_Operator	Data about the sales operators responsible for the transactions at each POS. Useful for performance tracking and auditing.
Dim_POS	Describes the physical or online point of sale, enabling analysis by store, region, or online vs. offline sales.
Dim_Stores	Contains location-wise information about each Retail4All store.

Table 4 – Metadata of the Fact and Dimension Tables

7.3 ETL Setup (Pipelines, Dataflows, and Notebooks)

To populate the Data Warehouse, a hybrid ETL process was implemented using Microsoft Fabric tools, namely Pipelines, Dataflows, and Notebooks. Pipelines were used to coordinate and automate the data ingestion process. They permitted the loading of raw CSV files into the Lakehouse and subsequently into the previously created Warehouse tables.

We decided to use a full load strategy to ingest data, meaning the entire dataset was reloaded with each ETL execution. This approach was chosen given the dataset's manageable size and allowed us to ensure data consistency during development. Additionally, we aimed to apply the concepts learned in class. While not optimized for large-scale or real-time systems, the simplicity of full load made it ideal for the project.

Dataflows were also employed, to carry out the bulk of the transformation logic. These low-code tools allowed us to clean, format, and enrich data using Power Query. At this stage, tasks such as removing duplicates, merging datasets, and renaming columns for consistency were handled. This ensured that the data arriving at the Data Warehouse was already cleaned and structured.

7.4 Explanation and Justification of The Transformation Steps Applied

Several transformation steps were applied in order to ensure data quality, consistency, and utility. All tables had some common transformation steps such as putting the first line as the header, getting rid of duplicates, correcting data types but in all tables several, specific problems were detected that raised the need to treat them individually and thoroughly. After creating the six dataflows:

Table	Description
Dim_Date	A complete calendar was generated by listing all necessary dates and converting the list into a table. Columns such as Day, Day Name, Weekday, Month, Month Name, Quarter, and Year were added, along with abbreviations for Day and Month Names. Conditional columns were used to define Semester, its full name and prefix, as well as Quarter prefix. A column was created to distinguish between weekdays and weekends. Special days like holidays were marked, and a full date string was created using Day Name, Day, Month Name, and Year. A surrogate key (SK) in the format YYYYMMDD was created from examples. Data types were checked and updated as needed.
Dim_Product	The table was created from the LH products data. After correcting the header, duplicates were removed. Multiple index columns were added and removed during the cleaning process, with the final index starting from 1 serving as the surrogate key (SK_Product), which was moved to the first column. Column names were updated to match SQL conventions. This process ensured a clean and deduplicated product dimension.
Dim_POS	Based on the LH point of supply table, this dimension required minimal cleaning. After setting headers and adjusting data types (POS_id as Integer and Name as Text), a copy of the POS_id column was created to act as the surrogate key (sk_POSID), which was moved to the first column. Column names were updated accordingly. The table was preserved in its near-original state, with a note for possible future cleanup of unused columns.
Dim_Operator	Sourced from the LH operators table, the headers were promoted and data types were updated (Operator_id as Integer, First_Name as Text). An index starting from 1 was added to create the surrogate key (SK_OperatorID) and moved to the first column. Columns were renamed to align with SQL query structure. This resulted in a clean and fully keyed operator dimension.
Dim_Store	Data from LH sales (retail), store_a, and store_b tables were combined. A location table was prepared by removing duplicates and renaming columns. In store_a, symbols were removed from the Store_ID, which was then split into two columns. Data types were set (Store_ID.1 as Integer, Store_ID.2 as Text). Similar cleaning was done on Location and Online columns—replacing "S" with "Y" (Yes) and filling blanks with "N" (No). A Store_Type column was added with value "A" for merging purposes. After cleaning, store_a and

	store_b were merged, Location was set to Integer, and a left join was performed to finalize the store dimension. Moreover, new columns with conditional formatting from Store_Name, City, District are created for further analysis, because some values contained special characters such as “ó” and “á”. The data types of those columns were changed.
Fact_Sales	Sales transaction data from the Data Lakehouse was merged with dimension tables using surrogate keys. A date column was extracted from the Datetime field. Each dimension was joined using appropriate keys, and only the SK columns were retained from merges. Data types were standardized, invalid rows (e.g., with zero or negative quantities) were removed, and calculated columns such as total sales value were added. This ensured a clean, accurate, and analysis-ready fact table that supported Retail4All’s reporting needs like sales tracking and year-over-year comparisons. We also removed rows that presented null values in the SKU column, since they had no product description.

Table 5 – Data Preparation and Cleaning Steps Taken for The Fact and Dimension Tables

7.5 Advanced Analytics

In this phase, we implemented advanced analytics and predictive modeling methods to generate sales forecasts and integrate them into the Data Warehouse. Two notebooks were developed to perform the time series forecasting and the models aimed to predict future sales based on historical data. The resulting predictions were used to create a new fact table in the Data Warehouse.

The entire workflow was executed within Fabric, using our Lakehouse as the data source. Python was selected as the programming language. Historical sales data stored in CSV format was imported and analyzed using two time series forecasting models, ARIMA and SARIMAX. These models were trained and tested to determine which would provide more reliable forecasts.

SARIMAX achieved a lower Mean Absolute Error (MAE), suggesting a better performance on average, while ARIMA produced a slightly lower Root Mean Squared Error (RMSE), indicating it made fewer extreme prediction errors. Although ARIMA was expected to handle outliers better, the difference in RMSE was not substantial, and SARIMAX’s consistency made it the more suitable choice for our purposes, so we chose it as our final option. The selected model was, then, saved in Fabric as an ML artifact, enabling reuse across different workspaces if needed.

Following model selection, a new SQL fact table named FACT_*Forecast_Quantity* was created in the Data Warehouse to store the forecasted sales data. This table contains two columns: FK_*date*, the foreign key that links it to the date dimension, and *forecast_quantity*, which stores the predicted values. The data pipeline was then updated to include a “Clear Fact Forecast Quantity” step, ensuring the forecast table is refreshed before new predictions are loaded.

A final notebook, titled “Forecast Sales”, was added to the pipeline to execute the model and generate the forecast. The predictions were exported to a CSV file and stored in the Lakehouse. A new dataflow was then created to read this CSV file, join the data with the date dimension for foreign key mapping, and load the merged data into the FACT_*Forecast_Quantity* table in the Data Warehouse.

All predictions revealed to be quite accurate with the real values. Advanced analytics is a crucial step, supporting decision-making by incorporating machine learning forecasts into existing data models. The choice of SARIMAX as the final model reflects a preference for stable and consistent forecasts, which aligns well with Retail4All's planning needs.

7.6 Complex ETL Developments (OPTIONAL)

We also aimed to optimize our ETL processes beyond the expected coursework, so we implemented several advanced techniques, both in pipeline management and data transformation. These improvements ensured enhanced workflows and allowed us to take deeper insights of the predictive analytics conducted through data integration.

7.6.1 Incremental Load Implementation

In order to optimize performance, we decided to implement incremental loading in our ETL process. This approach ensures that only new or updated records are loaded during each ETL run instead of reloading the entire dataset (full load approach). Compared to the latter, incremental loading reduces processing time, system load, and storage use. It also supports more frequent updates and makes error recovery easier, since only a subset of data is affected.

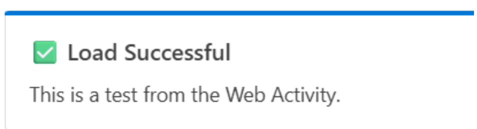
This mechanism was implemented by creating staging tables, which serve as exact replicas of our dimension and fact tables. Initially, data is loaded into these staging tables. Once loaded, SQL scripts are executed to compare the contents of the staging tables with the original tables, identifying any changes or new records. This comparison ensures that only new or modified data is processed during each run. By adopting this selective approach, we maintain data integrity while optimizing processing efficiency.

This enhancement improved both the scalability and reliability of our ETL pipeline, making it better suited for real-time insights and recurrent data integration needs.

7.6.2 Pipeline Activities (Custom Failure Handling and Notification System)

We started by implementing a failure handling mechanism in our pipeline. Using a web activity and connecting it through failure paths we introduced a dynamic error management system that improves monitoring and response to pipeline execution issues.

Specifically, for problems in data loading, we configured failure paths to trigger webhook calls to a Microsoft Teams channel of our BI team. This ensures that pipeline failures do not go unnoticed and allows for immediate intervention. These webhooks give real-time alerts containing a failure message and even though this level of automation with collaboration tools is not usually covered in standard ETL training, it has proved to be valuable in building data-heavy pipelines. Both the Fail and Success paths were tested in order to confirm their success. Below, screenshots of these said messages are provided as evidence of the system functioning.



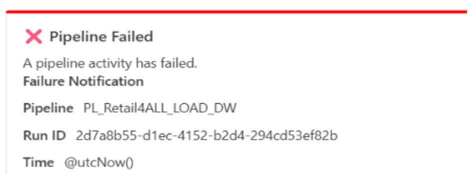


Figure 4 – Successful Loading Message

Figure 5 – Unsuccessful Loading Message

7.6.3 Advanced Transformations (Weather Data Integration and Predictive Modeling)

Another advanced component of our project was the integration of external weather data to analyze its potential correlation with sales performance. We ingested the dataset into the Lakehouse, already cleaned, and then we temporally and regionally aligned the weather conditions with the existing sales data. We analysed the relationship between Rain, Temperature and Daily Sales using the pearson correlation coeficient and, afterwards, a Random Forest model was trained with the new weather data to evaluate its impact on forecast accuracy.

While our main dataset includes sales data across multiple districts in Portugal, due to time and data constraints, we limited our weather data integration to the Bragança district. Focusing on this specific district still allowed us to explore the potential impact of weather on sales predictions, while keeping the results interpretable. In future iterations, we plan on expanding this analysis by incorporating broader geographic coverage.

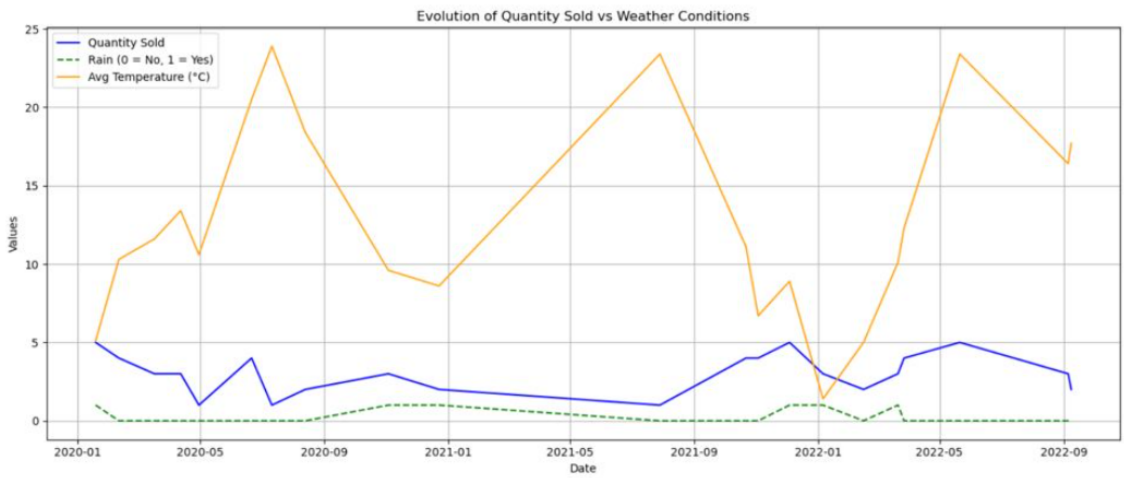


Figure 6 – Comparison between Sales and Weather Conditions

The visualization above reveals, has predicted, potential correlations worth noting. Periods with higher average temperatures often align with peaks in sales (e.g., july of 2020), suggesting a positive relationship between warmer weather and increased consumer demand, possibly because this facilitates and encourages consumers to go to supermarkets and physical shops.

Inversely, the presence of rain appears to correspond to lower sales in several instances. Although the correlation is not perfectly consistent throughout the timeline (especially by the end of the timeline, there's both a peak in raining and in sales; this could be easily explained by sales periods after Christmas and New Year), the patterns indicate that both temperature and precipitation may influence purchasing behavior. These findings reinforce the value of incorporating weather data into forecasting models.

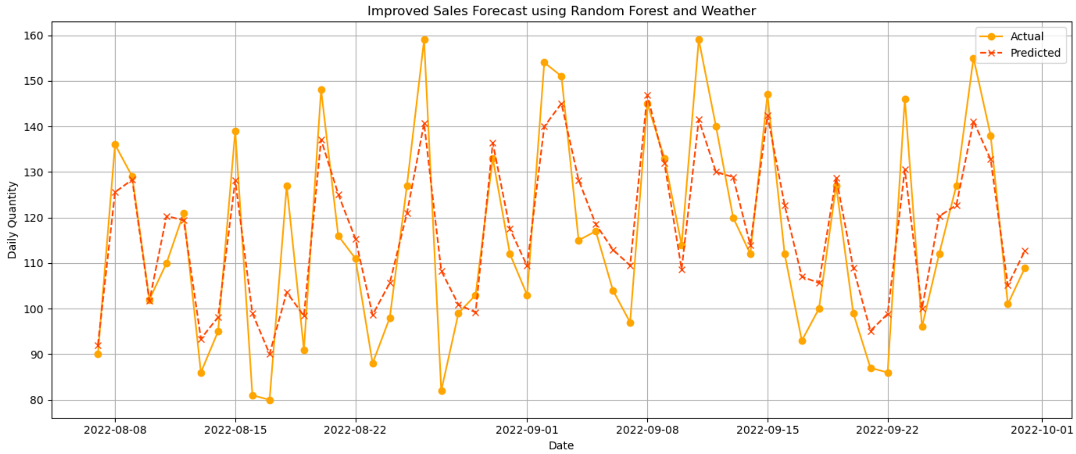


Figure 7 – Comparison between Predicted Sales Using Weather Data and Actual Sales

Sustaining our previous conclusions, the graph above shows a strong alignment between predicted sales using weather data and the real values of Retail4All sales, with only slight deviations. These small errors are an expected outcome when using a Random Forest model. We chose this algorithm because it succeeds at handling complex, non-linear relationships and is less prone to overfitting, being well-suited for variables like weather. However, its interpretability is quite limited and may not anticipate sudden sales spikes driven by external factors. Nevertheless, the overall accuracy confirms the value of integrating weather data in forecasting.

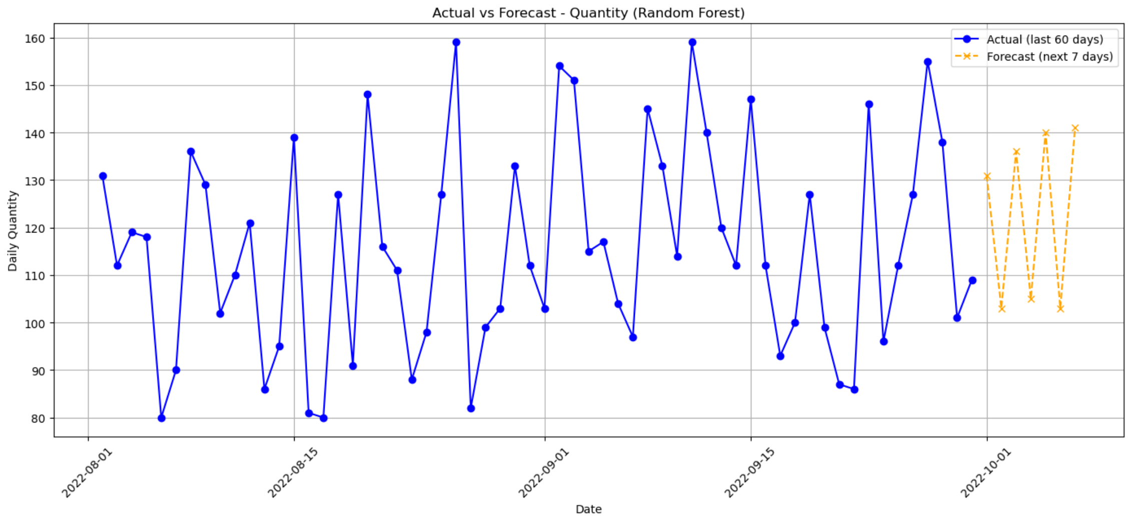


Figure 8 – Sales Forecast for The Next Seven Days

Our seven-day forecast shows some evident fluctuations, but it seems to closely follow the overall trend of sales, so even though short-term precision may not be high, in the long term the model captures the high-and-low pattern of each sales season.

8. SECOND DELIVERY CONCLUSION

This second part of the project strengthens the data system behind Retail4All, turning raw data into a clean and actually useful resource. Through the implementation of a well-orchestrated ETL process, we succeeded in integrating external data sources, applying meaningful transformations, and constructing a star-schema Data Warehouse tailored for performance and insight generation. Taking advantage of multiple Microsoft Fabric tools, we not only implemented successfully the entire pipeline but also incorporated advanced elements such as predictive modeling, external weather data integration, and custom error-handling mechanisms, exploring beyond the classroom scope. This delivery lays a strong technical foundation for the final reporting phase, ensuring that Retail4All's managers have access to clean data that enables them to make educated decisions.

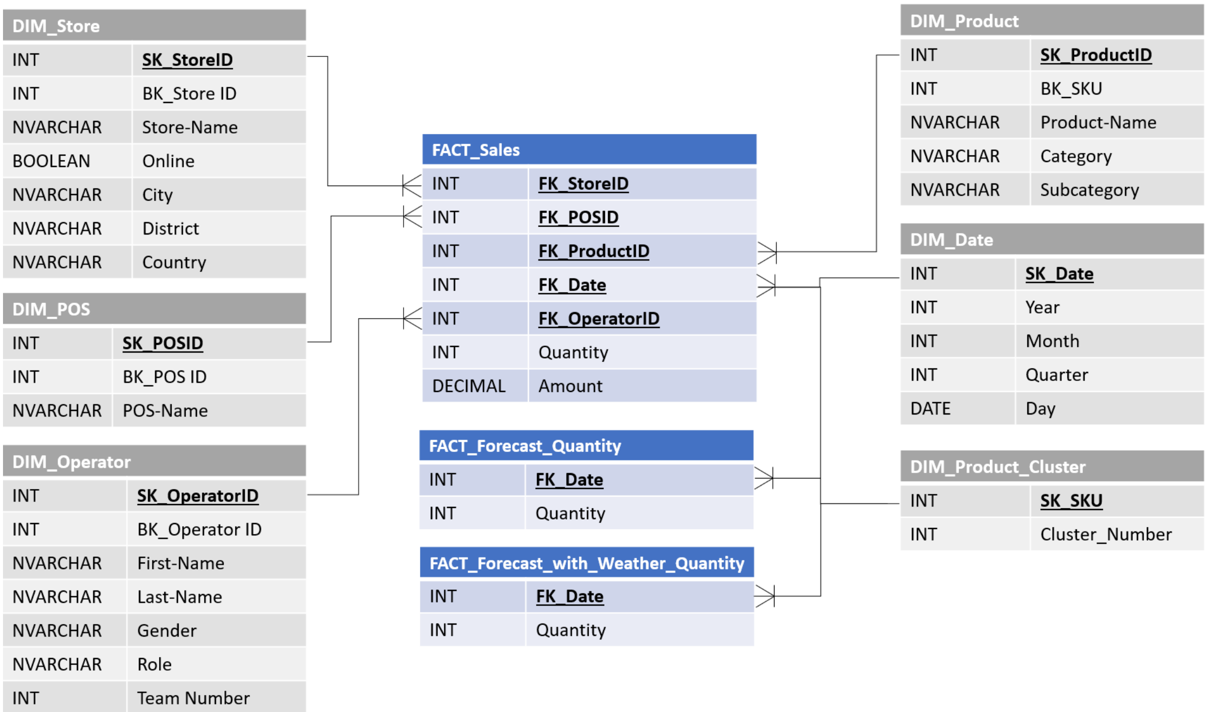


Figure 9 – Dimensional Model with added implementations

9. REPORTING

9.1 INTRODUCTION

The third and last delivery of the present project focuses on reporting and knowledge generation. Sustained by the developed dimensional model and the robust ETL processes previously applied, we aim to transform and organize the cleansed data into business-centered dashboards developed in Power BI. Our main objective for this phase is to provide meaningful and clear insights that meet Retail4All's both operational and strategic needs.

Keeping these goals in mind, in this section we will present the structure and description of the semantic model, key DAX measures, design decisions for the report pages, and a discussion of how the developed reports answer the original business questions from the first delivery.

9.2 SEMANTIC MODEL

In Power BI, the semantic model serves as an intermediary layer that connects the data stored in the data warehouse to the reporting interface. It is a crucial tool in any BI project, for it enables people without data science expertise to comprehend the raw data and the impact it has on the business.

Table Name	Columns	Hidden Columns/SKs/BKs/FKs	Hierarchy
DIM_Datez	Date Hierarchy	#Day, #Month, #Quarter, #Semester, #Weekday, Full Date, Holidays, Month Name, Month Name Short, Proper Date, Quarter Name, Quarter Name Short, Semester Name, Semester Name Short, SK Date, Weekday Name, Weekday Name Short, Year	Year > Semester Name Short > Quarter Name Short > Month Name Short > Proper Date
DIM_Operator	#Team, First Name, Gender, Last Name, Role	BK Operator ID, SK Operator ID	–
DIM_POS	POS Name	BK POS ID, SK POS ID	–
DIM_Product	Category Hierarchy	BK SKU, Category, SK Product ID, Subcategory, Product Name	Category > Subcategory > Product Name
DIM_Store	Online, Store Name, Store Type, Country Hierarchy	BK Store ID, City, Country, District, SK Store ID	Country > District > City

DIM_Product_Clusters	#Cluster	SK Product ID	–
FACT_Sales	Σ Amount, Σ Quantity **Measures**: Gross Sales Value, Gross Sales Volume, Sales Amount LY, Sales Quantity LY, Sales Amount YoY Diff, Sales Quantity YoY Diff, Gross Sales Volume, Product Amount MTD, Product Amount QTD, Product Amount YTD	FK Date, FK Operator ID, FK POS ID, FK Product ID, FK Store ID	–
FACT_Forecast_Quantity	Σ Forecast Quantity	FK Date	–
FACT_Forecast_with_Weather_Quantity	Σ Forecast Quantity	FK Date	–

Table 6 – Semantic Model Tables, Columns (Visible and Hidden) and Hierarchies

- Relationships:
  - FACT_Sales[FK Date] ← DIM_Date[SK Date]
  - FACT_Sales[FK Operator ID] ← DIM_Operator[SK OperatorID]
  - FACT_Sales[FK POS ID] ← DIM_POS[SK POS ID]
  - FACT_Sales[FK Product ID] ← DIM_Product[SK Product ID]
  - FACT_Sales[FK Store ID] ← DIM_Store[SK_StoreID]
  - FACT_Forecast_Quantity[FK Date] ← DIM_Date[SK Date]
  - FACT_Forecast_with_Weather_Quantity[FK Date] ← DIM_Date[SK Date]

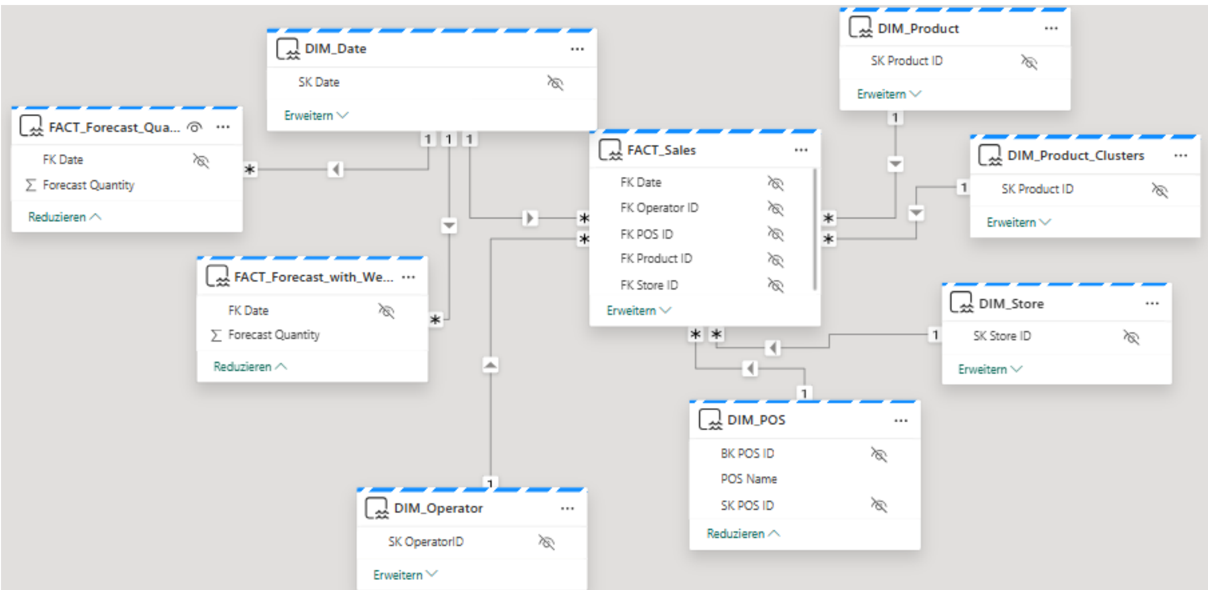


Figure 10 – Print of Semantic Model in Microsoft Fabric

As stated previously, our semantic model follows a star schema structure. At the center, there is the *FACT_Sales* table, containing the core transactional sales data. It is connected to all dimension tables in a one-to-many relationship. Additionally, *DIM_Date* is also linked to two additional fact tables, *FACT_Forecast_Quantity* and *FACT_Forecast_with_Weather_Quantity*, integrating predictive analytics and machine learning insights into our reporting layer.

Before starting constructing the semantic model itself, several important corrections were made to ensure its integrity. Firstly, all stage tables were removed. Originally used for intermediate transformations, these were no longer necessary once the ETL process was finished and keeping them would only add clutter to the model. *DIM_Weather* was also excluded for the same reasons. Lastly, the column previously and incorrectly labeled as "FK SKU" was renamed to "SK SKU".

Moving on to the construction of the model itself, the process began by launching a new semantic model within Microsoft Fabric. We started by cleaning and standardizing the data (names of columns were adjusted, and their data types and formats were corrected to ensure proper behavior in calculations and visualizations. All changes can be seen in the Appendix, Tables 8-17). Hierarchies such as Year > Semester Name Short > Quarter Name Short > Month Name Short in the *DIM_Date* table, and Category > Subcategory > Product Name in *DIM_Product* were then created to support informed and structured drilldowns. We also hid all the columns that were not necessary for report users, namely features outside of the hierarchies, as well as all business, foreign and surrogate key fields, to make the future reporting process clearer and more efficient (Table 6).

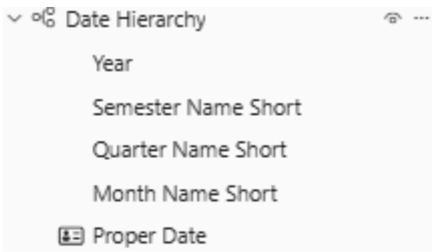


Figure 11 – Print of A Hierarchy Example: "Date Hierarchy"

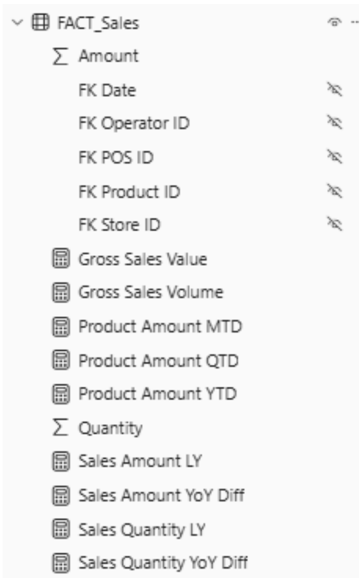


Figure 12 – Print of Measures Created in The Semantic Model in Microsoft Fabric

Unnecessary tables (leftover staging tables or support structures) were removed, and we then created DAX measures within the FACT tables to support the business needs outlined in the first delivery. These included time intelligence measures, year-over-year comparisons, and cumulative aggregations that will be more thoroughly explained in the next section of the report. Finally, all relevant tables were connected to the *FACT_Sales* table using proper one-to-many relationships, following a star schema design.

Additional adjustments include the declaration of the *DIM_Date* table as a proper date table and adjustment of fields such as Month Name, Semester Name, and all other time descriptors to sort correctly by their respective numeric columns (e.g., Month Number, Semester Number, etc.). These changes were made to guarantee that visualizations display timelines in the correct chronological order, rather than alphabetically (Power BI's default behavior). Furthermore, for each feature and metric the properties, if necessary, were adjusted to define the data type, the format, the unit, and the summarization, which is very important for the next steps.

9.3 Measures

Question	Measure	Report Page
What are the accumulated sales across different time periods (month, quarter, yearly)?	1. **Gross Sales Value*** = SUMX('FACT_Sales', 'FACT_Sales'[Amount]*'FACT_Sales'[Quantity])	**1st page Overview**
What are the sales quantities across different time periods (month, quarter, yearly)?	1. **Gross Sales Volume*¹** = SUM('FACT_Sales'[Quantity])	**1st page Overview**
What are the sales amount and quantities in comparison to the previous year?	1. **Sales Amount LY*²** = CALCULATE([Gross Sales Value], SAMEPERIODLASTYEAR('DIM_Date'[Proper Date]) 2. **Sales Quantity LY*²** = CALCULATE([Gross Sales Volume], SAMEPERIODLASTYEAR('DIM_Date'[Proper Date]) 3. **Sales Amount YoY Diff*²** = [Gross Sales Value] - [Sales Amount LY] 4. **Sales Quantity YoY Diff*²** = [Gross Sales Volume] - [Sales Quantity LY]	**6th page Advanced Analytics**
What are the top 3 products bought by store and location across different time periods (month, quarter, yearly)?	1. **Gross Sales Volume*¹** 2. **Product Amount MTD** = TOTALMTD([Gross Sales Volume], 'DIM_Date'[Proper Date]) 3. **Product Amount QTD** = TOTALQTD([Gross Sales Volume], 'DIM_Date'[Proper Date]) 4. **Product Amount YTD** = TOTALYTD([Gross Sales Volume], 'DIM_Date'[Proper Date])	**2nd page Sales Performance**
What are the top 3 sellers (operator) by year?	1. **Gross Sales Value***	**3rd page Employee Performance**
What are the aggregated sales by product category and product subcategory?	1. **Gross Sales Value***	**2nd page Sales Performance**

Table 7 – Measures and Respective Business Questions

(*² Calculated Measures using complex DAX formulas: 3. Iv, (OPTONAL) Complex Reporting developments)

The measures created present crucial roles in answering Reatil4All's business needs by providing insightful and time-based sales analysis. The *Gross Sales Value* measure calculates revenue by `'FACT_Sales'[Amount]*'FACT_Sales'[Quantity]` across Fact Sales. Analyzing the total revenue across different time periods such as monthly, quarterly or yearly, is essential to help the business understand the overall performance and adapt to seasonal trends. Similarly, *Gross Sales Volume* sums up the total quantity of items sold, offering a clear view of product movement over time. This is especially valuable for inventory planning and demand forecasting.

To assess performance changes over time, the *Sales Amount LY* and *Sales Quantity LY* measures calculate last year's sales amount and quantity, using the SAMEPERIODLASTYEAR function. These enable yearly comparisons and help identify growth or decline. Afterwards, the *Sales Amount YoY Diff* and *Sales Quantity YoY Diff* measures subtract last year's figures from the current year's, quantifying the difference and allowing us to assess performance improvement, which is of major importance for decision-making (e.g., adjusting marketing efforts or product focus).

In terms of product insights, the *Product Amount* measure aggregates the total sales amount for each product, and the *Rank* measure ranks them within a selected context (e.g., store, location, or time). This targets one of our most important business goals, namely the identification of the *top 3 products* by store and location, giving visibility into which items drive the most revenue in different markets. Additional measures like *Product Amount MTD, QTD,* and *YTD* use time-intelligent functions such as TOTALMTD, TOTALQTD, and TOTALYTD and evaluate the performance of each product during the current month, quarter, or year, and support product strategy adjustments.

The *Total Sales Amount* and *Operator Rank by Year* measures facilitate performance analysis of sales operators. By summing sales and ranking operators (based on yearly performance), we can identify the top sellers and implement reward systems, identify areas in need of training, or territory planning. Finally, aggregating sales by product category and subcategory using *Total Sales Amount* empowers a more robust segmentation and product portfolio analysis. This helps Reatil4All understand which categories or subcategories are performing as expected or not, guiding decisions on product development, marketing strategies, or supply chain optimization.

9.4 Reports

The developed Power BI report is composed of six different pages, each answering a specific goal that is aligned with Retail4All's needs. The first page focuses on a strategic level, consisting of an executive dashboard with high-level KPIs, sales trends, and a geographic performance map. The second and third page provide insights into sales performance by store, product, operator, and point of sale with the intent of supporting mid-level tactical decisions. The fourth page that helps track and compare store-level sales performance. The fifth page presents a product detailed drill-down with a decomposition tree for analyzing Gross Sales Volume drivers and a heatmap showing Gross Sales Value by Category. Lastly, the sixth page focuses on advanced analysis that will be helpful for understanding the customer base and future sales planning.

9.4.1 Visualization Techniques

Page 1: Cover Page & Executive Dashboard (Strategic Level)

This initial page was developed to serve as an overview for executive stakeholders. It emphasizes clear KPIs, seasonal sales trends and spatial analysis to support decision-making.

- **KPI Cards** display Gross Sales Value, Gross Sales Volume, and Sales Amount YoY Difference. These indicators are calculated using the mentioned DAX measures from the *Sales Fact* table, allowing users to assess business performance through clear and interactive visuals.

- The **Line Chart** represents Gross Sales Value and Gross Sales Volume over time. This trend analysis helps users detect seasonal patterns and track growth or decline; it uses a date hierarchy field sourced from the *DIM Date* table and is influenced by a Date slicer.

- A **Map Visualization** was included to distinguish Gross Sales Value by District. This geographic visualization is built using fields from *DIM Store* and provides a spatial overview of sales distribution across the country, allowing executives to identify high-performing or underperforming regions at a glance.

- A **Bar Chart** illustrates Gross Sales Volume per Month Name and Product Category. This enables users to compare monthly sales performance across different product categories, helping decision-makers quickly identify which categories perform best during specific times of the year.

- Lastly, **Slicers** helps to filter all visuals by Year, POS and Product Category, helping compare performance across different periods, business sectors and geographic references.

This page aims to enable decision-makers to monitor business performance and answer strategic questions like “What are the total sales this period?”, “How are we performing compared to last year?”, and “Where are the most profitable regions?”

Page 2: Sales Performance Analysis (Tactical Level)

The second report page is meant for store managers, regional leaders, and product specialists. It presents sales performance across stores, products and POS channels in detail, helping stakeholders to take decisions that optimize operations.

- **KPI Cards** display Product Amount by YTD, QTD and MTD. These indicators permit users to monitor product performance over different time frames, supporting both short-term and long-term decision-making. The temporal slicer allows for a more thorough exploration of these values.

- A **Matrix Table** displays Gross Sales Value by Product and Store, drawing from *DIM Store* and *DIM Product* while summarizing values from *FACT Sales*. This format allows users to easily compare combinations and spot under- or over-performing products.

- The **Bar Charts** on the right-hand side of the page demonstrates Gross Sales Value per Month and Year to highlight seasonality and year-over-year growth patterns, while the one at the bottom compares Gross Sales Value across POS to offer insights into which selling environments are most effective.

- **Slicers** for Date Hierarchy, Product, and Store were also added for a more thorough exploration. These filters permit users to narrow the analysis on specific time frames, products, or locations, increasing the analytical depth and flexibility of the page.

This page helps middle managers and analysts take informed tactical actions that and to answer tactical questions such as “Which products perform best in each store?”, “Which operators are most effective?”, and “What are the top POS channels driving our revenue?”.

Page 3: Employee Performance (Operational Level)

Our third page was created with the goal of supporting operational decision-makers such as team leaders and store supervisors, monitoring and evaluating both individual and collective performance across different time periods.

- A **Matrix Table** displays Gross Sales Value by Product Category and Team. This allows for a clear comparison between each team’s performance and highlights the product categories they sell better, helping to identify strengths, weaknesses, and potential areas for targeted improvement.

- At the top right corner, a **Pie Chart** of Gross Sales Volume per Operator provides a good visual representation of each Operator’s contribution to sales while, at the bottom right corner, a **Bar Chart** gives a more quantitative notion of the same concept.

- A **Heatmap** of Gross Sales Value by Team and Year helps identify the top- and under-performing teams throughout time.

- **Slicers** for Date Hierarchy, Product Category, Team and Store were added to allow an accurate filtering by relevant variables to employee performance.

- **Bookmarks** were created to support specific use cases for the Store Bits & Bytes, providing views tailored to the store managers’ needs: the first viewpoint filters the data to compare Electronics (top-selling category) sales in 2022 for this store, enabling quick analysis of this category; the second viewpoint consists of two bookmarks on the Employee Performance page that compare Team 3’s performance in 2022 and 2023, allowing direct year-over-year team analysis.

This page intends to answer operational questions like “Are employees maintaining performance consistently over time?”, “How do store-level results vary month by month?”, and “Where might additional support or training be needed?”

Page 4: Stores Performance

The fourth page was designed for store managers and regional leaders to monitor and compare sales performance across stores in detail.

- A **KPI Card** displays Gross Sales Value, offering a quick overview of overall store revenue. We have applied thresholds to which different colors are applied (see Appendix Figures Y):

- Gross Sales Value (per store per year)
  - 0 < 320000 → Red
  - 320000 <= Max → Green

- **Slicers** for Store and Date focus the analysis on specific locations and time periods.
- A **Clustered Bar Chart** shows Gross Sales Value by Store, allowing to quickly identify top and underperforming locations.
- An **Area Chart** presents Gross Sales Value by Month and Store, highlighting sales trends and seasonality.
- A **Matrix Table** breaks down Gross Sales Value by Store and Product Hierarchy, providing detailed insights into which products drive sales in each specific store.

This page helps managers answer key questions like "Which stores are performing best?", "How do sales vary over time by location?", and "What products contribute most to store revenue?"

Page 5: Product Details

The fifth page continues at the strategic level, offering executives a deeper understanding of sales performance drivers and specific product highlights.

- A **Decomposition Tree** is used to break down Gross Sales Volume in order to identify key contributing factors such as Product Category, Store, or POS (controlled by slicers) in a dynamic, drill-down format. Based on a strong root cause analysis approach, this visualization enables business users to take fast insights into what contributes to sales.

- A **Heatmap** displays Gross Sales Value by Product Category, enabling a quick visual comparison of different categories to sales performance and helps identify which ones require more attention and investment.

- **Slicers** for Date Hierarchy, POS, and Team are included to deepen the analysis based on time period or workforce structure.

This page is designed to support strategic questions like "What's contributing most to our volume?", "Which categories are contributing the most to overall revenue?", and "How do key drivers vary by team or POS?"

Page 6: Advanced Analytics (Predictive Insights Level)

The last page of our report focuses on predictive analytics and segmentation to enable managers to take future-focused insights and get a deeper business understanding.

- Two **KPI Cards** show Sales Amount YoY Difference and Sales Quantity YoY Difference, monitoring overall sales growth from both a value and volume perspective. These cards were color-filtered to be green in case of a positive difference and red in case of a negative one.

- A **Line Chart** compares Forecast Quantity and Sales Quantity over time, allowing users to assess forecast accuracy and plan according to the predicted demand.

- A **Pie Chart** presents Gross Sales Volume by Product Cluster, enabling quick identification of the clusters that are generating the highest sales volumes.

- A **Stacked Area Chart** breaks down Gross Sales Value by Month Name, Store, and Product Cluster. This visualization offers insight into how each segment contributes to total revenue over time.

This page answers questions like “How accurate are our forecasts?”, “Which products are driving growth?”, and “What patterns are emerging across stores and product groups?”

To ensure intuitive navigation across the report, on the right side of each page, four buttons were also included. The Home button, represented by a house icon, redirects users to the Cover Page, providing a quick return to the report's starting point. The Back and Forward buttons, symbolized by arrow icons, allow users to move back or forward. Additionally, the Reset button, depicted as an eraser, clears all active slicer selections, resetting filters and viewing the full, clean dataset.

9.5 Final Analysis

The executive dashboard highlights several business keys relevant to Retail4All’s strategic decision-making. Starting with our cover page, overall, the company has generated a Gross Sales Volume of approximately 3 million units and a Gross Sales Value of €47.54 million, reflecting solid sales performance with a general gain of 7.59 M€ since the beginning of the dataset. Geographically, Guarda emerges as the strongest performing district, followed by Braga and Santarém, indicating that the company’s presence is most impactful in the north and center of the country (see Appendix Figure D). Temporally speaking, there is a clear seasonal trend, with sales peaking during the early months of the year, but a noticeable decline after May. This pattern may suggest the influence of specific seasonal campaigns or consumer trends (e.g., special promotions after Christmas). Electronics stands as the most purchased category (see Appendix Figure E), followed by Clothing & Accessories, underlining the importance of these segments in driving overall revenue.

The second report page offers a detailed look into periodic performance and points of sale. In terms of distribution channels, Lazzy is the most impactful point of sale, exceeding €10 million in Gross Sales Value. It is followed closely by Kwimbee, reaching approximately this same value, and RealCube, slightly trailing just below that 10 M mark (see Appendix Figure E). More resources should be allocated to these locations to maximize revenue. On other hand, Gross Sales Value reached its peak during the first half of 2023 but experienced a sharp decline in June, with no recorded sales activity from July onward - an anomaly that may suggest missing data or a disruption in operations. In contrast, the years 2020 through 2022, as well as early 2023, demonstrate a relatively stable performance with a consistent year-over-year growth in Gross Sales Value (see Appendix Figure F).

The third page highlights individual operators and team-based sales contributions. Among all employees, Stern stands out as the top-performing operator, generating 11.88% of the total Gross Sales Value, followed closely by Sibyl (11.47%) and West (10.9%). These figures highlight the significant contributions of these individuals to overall revenue (more than 30% is generated by them three) and may serve as benchmarks for performance evaluations, team coaching and performance prizes (see Appendix Figure H). When analyzing team contributions, Team 3 consistently outperformed all others across all years, achieving its highest sales in 2022 with €7.3 million. Team 1 remained the lowest performer throughout the entire period. These insights may indicate the need for targeted support for underperforming teams, and further investigation into the operational gap observed mid-2023.

The fourth page reveals key insights into store-level performance. It is worth noticing that the stores Azulejo Tech and Digital Digs show no recorded sales data, which could suggest they are either recently opened stores yet to generate revenue or underperforming locations that may require evaluation for inevitable closure. Among the active stores, Tech Wizards stands out as the top contributor, generating approximately €4.3 million in Gross Sales Value (more than double that of an average store). In contrast, most other stores demonstrate similar contributions, each around €2 million (see Appendix Figure M). This balanced distribution suggests a stable store network, but the significant gap between Tech Wizards and the others might present an opportunity to address disparities and guide resource allocation, store support, and restructuring within the retail network.

On the fifth page, the heatmap highlights Electronics as the consistently most purchased category, peaking at €3.4 million. Next is Clothing & Accessories, with a peak of €839.4K (see Appendix Figure P). Within Electronics, Smartwatches and Gaming Accessories stand out as the top-selling products, as revealed by the decomposition tree. In the Clothing category, Jeans and Boots lead the category. These visualizations sustain previous findings showing higher sales in the initial months of the year, with a notable decline after May. Additionally, these insights help pinpoint key products and seasonal trends that help develop more targeted marketing strategies and inventory decisions.

Finally, the last page focuses on customer segmentation and forecasting accuracy. Cluster 3 represents the most valuable customer group, contributing approximately €963K (30.65%) to total Gross Sales Volume. This is followed by Cluster 4 with €890K (28.33%) and Cluster 0 with 840K (26.73%), indicating these segments as the most engaged and revenue-driving buyers. Cluster 1, by contrast, contributes minimally to sales and appears to consist of occasional or disengaged customers, possibly requiring tailored marketing strategies to improve retention (see Appendix Figure R). On the forecasting side, although the predicted sales values significantly overshoot the actual sales (nearly triple the real value), the overall trend remains directionally aligned, suggesting that while the forecast captures seasonality and directional movement, it may benefit from refinement. Potential limitations such as limited historical data, external variables, or time constraints may explain the variance, highlighting the need for ongoing model tuning and validation to enhance predictive accuracy in future iterations (see Appendix Figure T).

9.6 Optional Work - Complex Reporting Developments

9.6.1 Complex Interactions

Throughout the Power BI report, we tactfully inserted slicers to heighten user experience, refine the analysis, and tailor insights to each decision-making level. These interactive filters allow for a dynamic and intuitive exploration of the data, viewing it from different angles and uncovering patterns that would remain hidden in static reporting.

Page 1 – Executive Dashboard (Strategic Level)

On the cover page, the Date Hierarchy and Product Category Slicers enable executives to analyze performance over different time periods, business sectors and compare trends year-over-year, quarter-by-quarter, etc. Filtering by year or even month proved to shift the entire perspective of sales KPIs, charts, and geographic maps. Executives can, therefore, allocate resources, plan strategically and evaluate periodically performance according to their key business goals.

Page 2 – Sales Performance Deep Dive (Tactical Level)

Slicers for Date, Product, POS and Store allow mid-level managers to conduct focused evaluations. For example, they are able to isolate a particular product line within a specific store for a certain quarter and make tactical decisions like product repositioning or localized promotions based on the results of their analysis.

Page 3 – Employee Performance (Operational Level)

The combination of Date, Store, Product, and Team Slicers makes it possible to assess individual and team contributions across different dimensions. For example, filtering by team and month allows supervisors to identify performance trends, outliers, or areas that need training.

Page 4 – Sales Performance

Slicers for Store and Date empower targeted analysis of sales trends over time on a store-level, e.g., filter a specific store's monthly performance or compare multiple stores within a given period. This helps identifying underperforming stores and giving them the support they need or recognizing top-performing locations to replicate successful strategies across the network.

Page 5 – Product Detail Analysis

On this strategic page, slicers for POS, Team, and Date refine the Decomposition Tree and Heatmap visualizations, segmenting the data in ways that reveal hidden influencers. This helps business users to dig deeper into what's driving sales volume and value, and check if these factors are influenced by seasonal factors.

9.6.2 Original Visuals

We dedicated the whole fifth page (Product Details) of the report to the creation of original visuals that provide unique insights beyond the standard ones learned in class. The **Decomposition Tree** permits analysts to get a detailed view of Gross Sales Volume by drilling down to its contributing factors. This interactive visual helps managers quickly identify key sales drivers and understand how different variables influence overall performance, standing out as a powerful tool for root cause analysis and decision support.

Additionally, the **Heatmap** visualization, used both in the third (Employee Performance) and fifth pages (Product Details) of the report, offers an easy and clear way to analyze Gross Sales Value (by yearly Team and by category, respectively), revealing patterns and highlights.

Both these original visuals enrich the report, presenting complex information in a visually appealing and user-friendly format.

9.6.3 Calculated Measures

Lastly, we created a set of four calculated measures using complex DAX formulas, centered especially on time intelligence and going beyond simple aggregations. They were specifically designed to answer the question "*What are the sales amount and quantities in comparison to the previous year?",* providing insights into the Retail4All's year-over-year performance.

These measures make use of CALCULATE, an impactful DAX function that modifies the context in which data is evaluated, and SAMEPERIODLASTYEAR, a time-intelligence function. The created measures are considered complex DAX measures because they incorporate dynamic filtering, context transition, and time-aware logic, therefore enhancing the analytical capability of the report.

1. **Sales Amount LY**: *CALCULATE ([Gross Sales Value] ,SAMEPERIODLASTYEAR('DIM_Date'[Proper Date]))*.

   This measure calculates the Gross Sales Value for the same period last year, using the SAMEPERIODLASTYEAR time intelligence function within a CALCULATE context. It enables side-by-side comparisons of current performance against historical benchmarks, which is crucial for trend analysis.

2. **Sales Quantity LY**: *CALCULATE ([Gross Sales Volume], SAMEPERIODLASTYEAR('DIM_Date'[Proper Date]))*

   Similar in logic to the previous one, this measure tracks the Gross Sales Volume for the same period one year earlier. This is key to identifying volume-based growth or decline patterns.

3. **Sales Amount YoY Difference**: *[Gross Sales Value] - [Sales Amount LY]*

   This measure represents the absolute difference in sales value between the current period and the same period last year. It highlights whether sales have increased or decreased over time in monetary terms.

4. **Sales Quantity YoY Difference**: *[Gross Sales Volume] - [Sales Quantity LY]*

   This measure calculates the change in sales volume, helping managers understand shifts in quantity sold rather than just revenue.

9.7 Visual Techniques

9.7.1 Design Consistency and Visual Identity

Throughout the report, the visual design was carefully orchestrated to boost readability, ensure visual distinction between pages, and maintain a consistent identity. Each report page adopts a unique but noticeable color palette. For instance, the Cover Page uses a blue-green-yellow–orange palette, symbolizing clarity, growth, and strategic overview. The Sales Performance Analysis page uses pink-red-salmon tones to reflect action-oriented insights and clearly demonstrate performance intensity. The Employee Performance Dashboard follows a yellow-brown-red-orange scheme, reinforcing temporal variation and heat-based evaluation. The Store Performance Page stands out with a green-blue palette, to facilitate the comparison between visuals and store entities. The Product Detail Page employs a blue schema, for analytical neutrality and balance.

Finally, The Advanced Analytics Page uses a purple palette, aligning with forward-looking insights and forecast-driven thinking.

To ensure legibility and stylistic coherence, all titles across the report were also formatted using Segoe UI, size 24, a modern font choice that enhances clarity and uniformity.

9.7.2 Branding and Credibility

Branding elements were added to maintain institutional credibility. At the top-left corner of each page, the university logo and the Business Intelligence & Analytics department logo were consistently displayed, reinforcing the academic purpose of the present study. On the top-right corner, the BI4ALL logo appears as recognition of the data source provider, giving the due credit to the partnership that allowed for the report's development.

10. THIRD DELIVERY CONCLUSION

In this final delivery, we integrate all the work conducted throughout the project and transform it into an efficient, actionable reporting solution. Building upon the semantic model and ETL processes developed in earlier stages, we used Power BI to successfully translate the cleansed data into meaningful business insights. The semantic model ensures simplified data exploration and allows business users without technical knowledge to understand and take action based on the data. With solid hierarchies, relationships, and meaningful measures, all visualizations were planned and created to meet Retail4All's strategic, tactical, and operational goals.

Our Power BI report was structured into six pages, each focused on a specific business level. Strategic KPIs and sales trends were presented to executives on the first page. Product, store, and employee performance were analyzed in depth on the second and third pages, respectively, for tactical and operational use. Finally, the sixth page incorporated advanced analytics, offering forward-looking insights that support future planning. With thoughtful DAX measures, interactive visuals, and intuitive filters, the report not only targets key business questions but also boosts Retail4All's ability to monitor and respond to the ever-changing sales dynamics.

In summary, this report marks the end of a three-phase journey. In the first delivery, we conducted an initial analysis and designed the first draft of our data model, aiming to align with Retail4All's strategic objectives and needs. The second delivery focused on building a robust ETL pipeline, with external data integration and predictive forecasting capabilities that added depth and foresight to the dataset. In the third delivery, we have combined these previous efforts through the development of a dynamic and user-friendly reporting environment. Upon review, the following key takeaways stand out:

- The developed **semantic model** and its star-schema architecture enable clean, intuitive analysis and guarantee high performance for large datasets.
- The **ETL pipeline**, enriched with external weather sources, and the **predictive modeling** techniques applied added forecasting capacity, supporting planning and decision-making.

- The **Power BI dashboards** were tailored to each business need, supporting decision-makers across all levels of the organization.

This project has laid a solid BI foundation for Retail4All. It empowers the company not only to understand what has happened, but also to predict what will happen, and ultimately, to act more informedly and swiftly in a competitive retail landscape.

10.APPENDIX

10.1 ETL APPENDICES

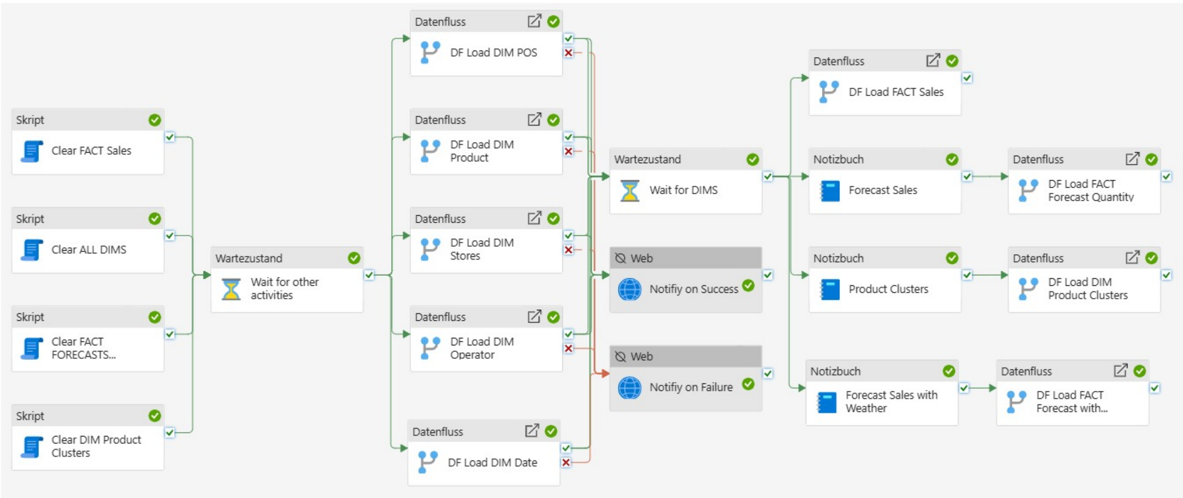


Figure A – Successful ETL Run Using Full Load*

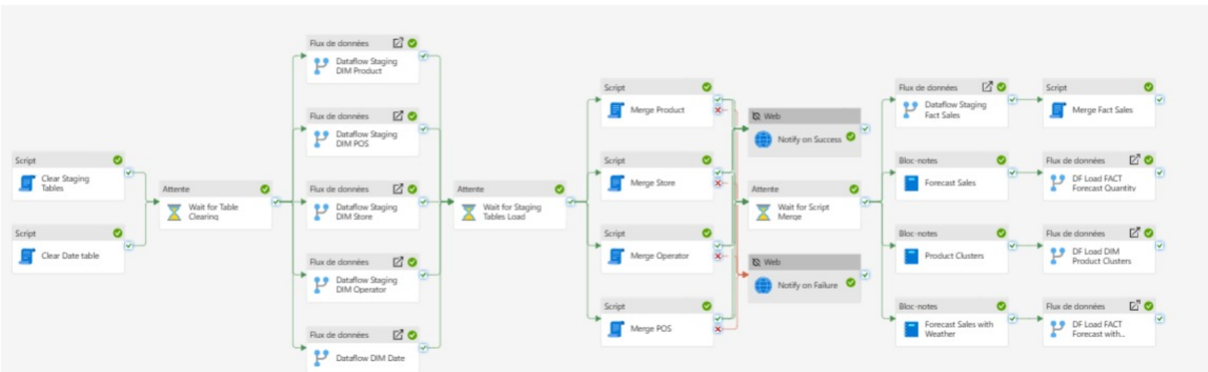


Figure B – Successful ETL Run Using Incremental Load*

*Web Activities are deactivated to reduce the number of messages in the Teams Channel during the development phase

10.2 Reporting Appendices

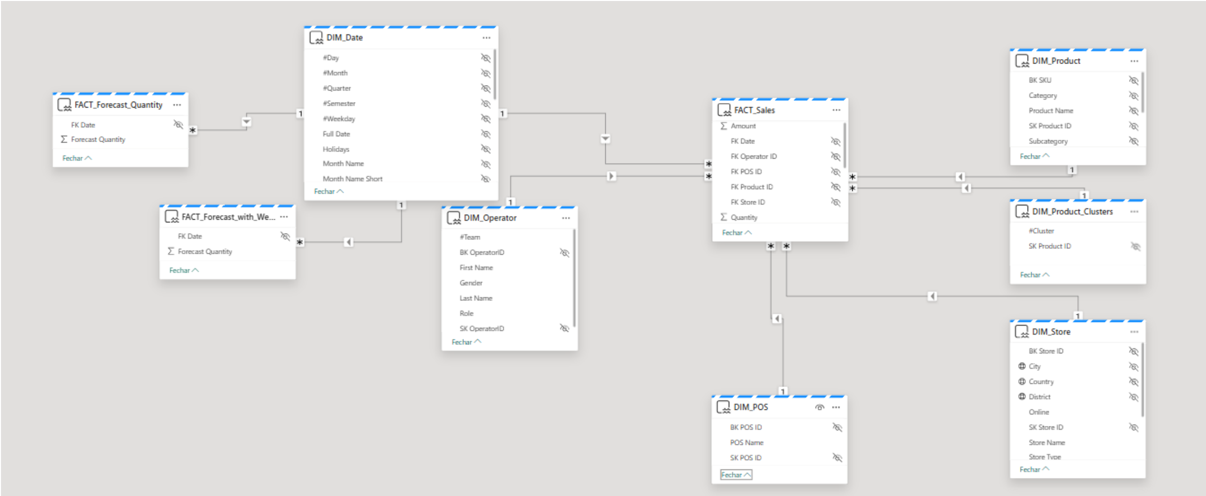


Figure C – Semantic Model

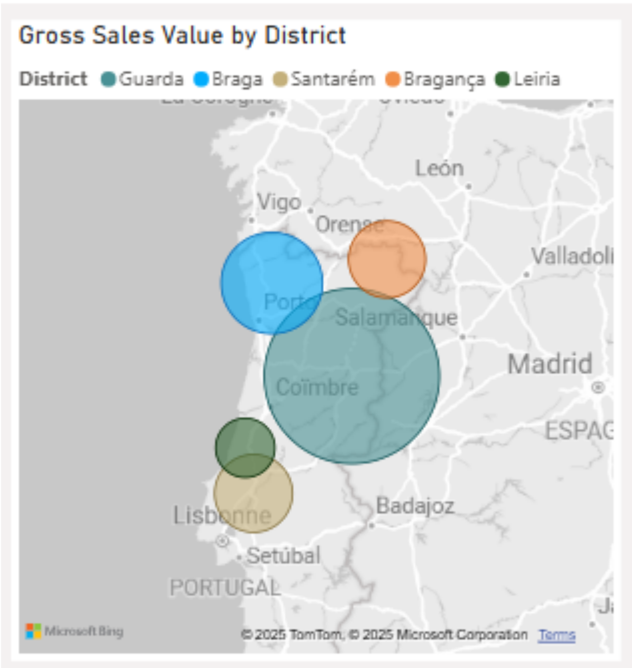


Figure D – Gross Sales Value by District

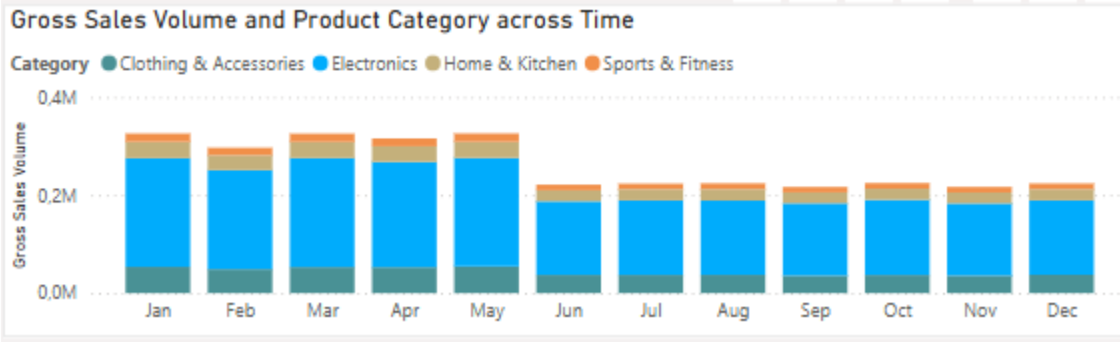


Figure E – Gross Sales Volume per Month Name and Category

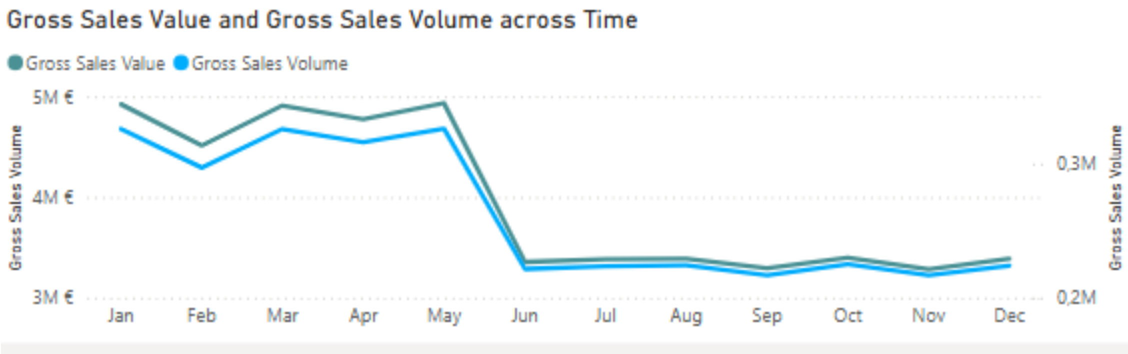


Figure F – Gross Sales Value and Volume across time

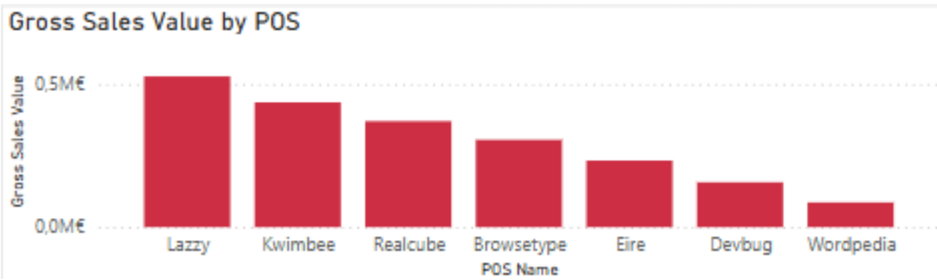


Figure E – Gross Sales Value per POS

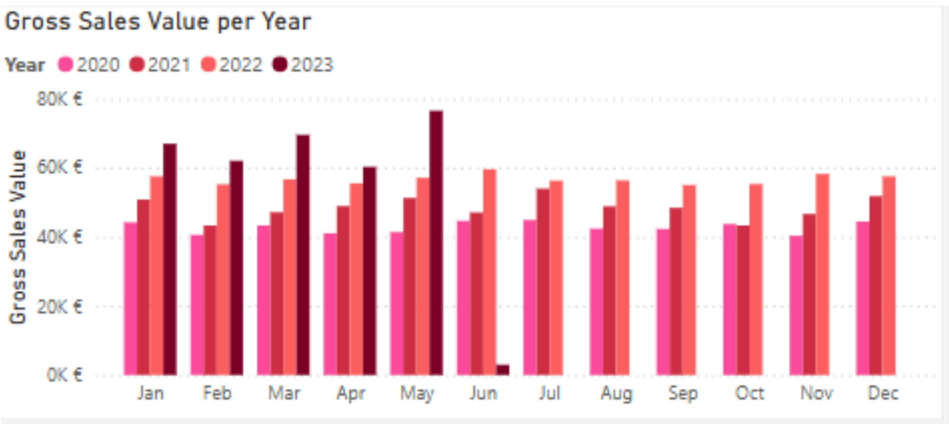


Figure F – Gross Sales Value per Year

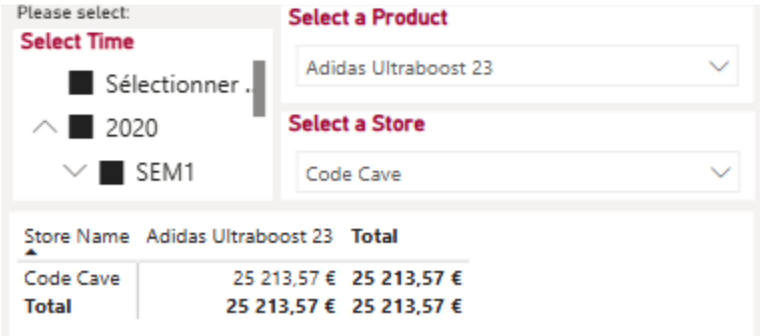


Figure G – Matrix of Gross Sales per product and store

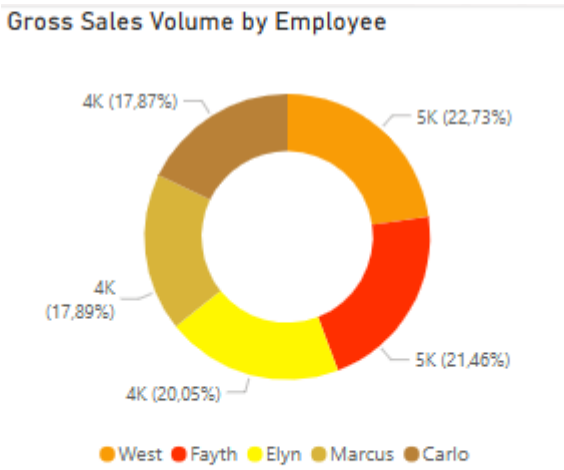


Figure H – Gross Sales Volume by employee

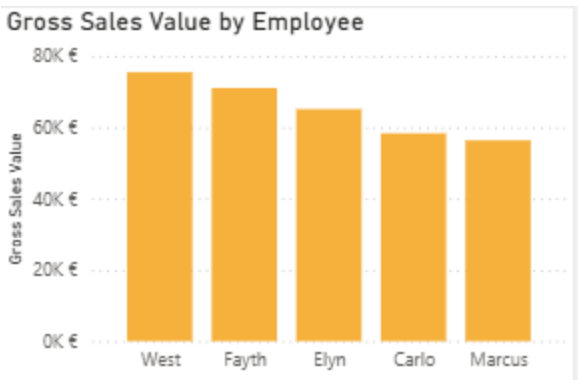


Figure J – Gross Sales Value per Month and Year

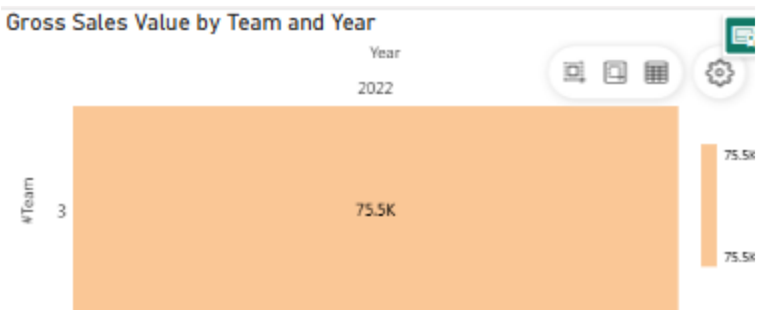


Figure K – Gross Sales Value by team and year

Category	3	Total
Clothing & Accessories	**52 696,51 €**	**52 696,51 €**
Boots	12 573,27 €	**12 573,27 €**
Bottoms	7 823,64 €	**7 823,64 €**
Chinos	4 019,19 €	**4 019,19 €**
Jeans	16 653,81 €	**16 653,81 €**
Sneakers	7 675,94 €	**7 675,94 €**
Tops	3 950,66 €	**3 950,66 €**
Electronics	**225 833,02 €**	**225 833,02 €**
2-in-1 Chromebooks	4 697,83 €	**4 697,83 €**
2-in-1 Laptops	8 987,63 €	**8 987,63 €**
Business Laptops	3 660,12 €	**3 660,12 €**
Chromebooks	8 654,46 €	**8 654,46 €**
Earphones	11 933,07 €	**11 933,07 €**
[illegible]	[illegible]	[illegible]
Total	**326 401,93 €**	**326 401,93 €**

Figure L – Gross Sales Value by product and team member

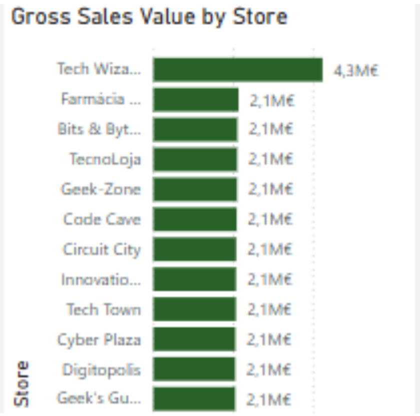


Figure M – Gross Sales Value by Store

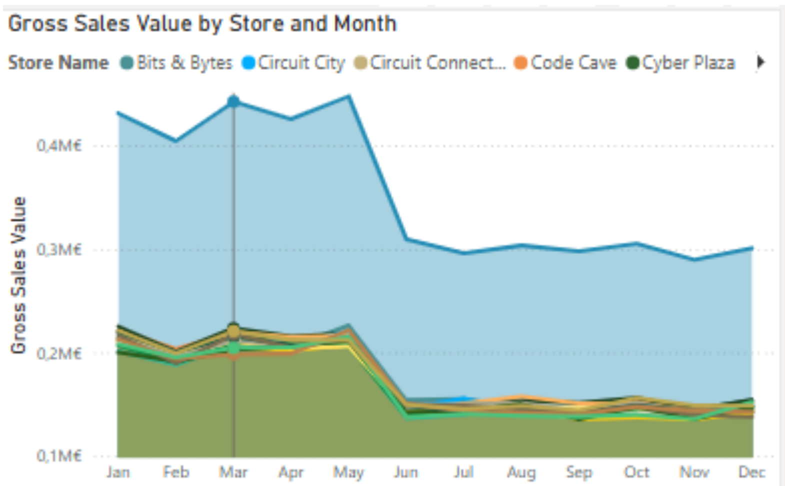
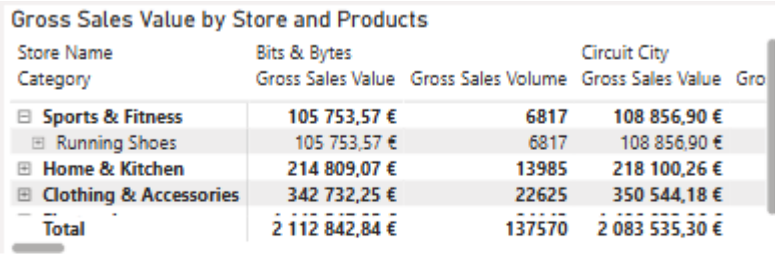


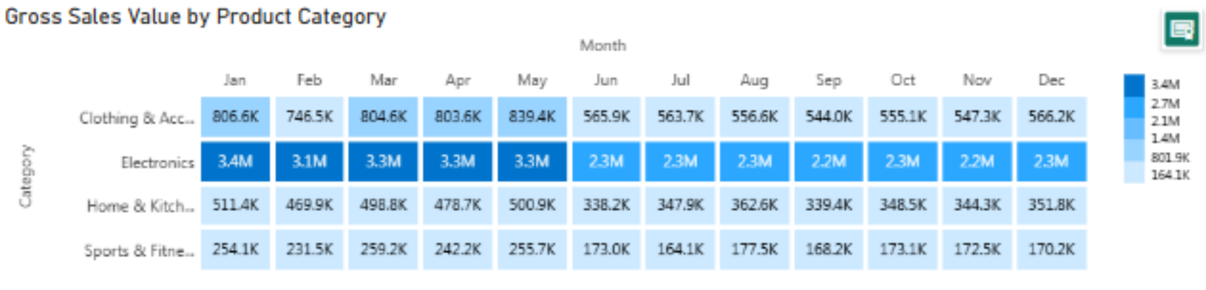
Figure N – Gross Sales Value by Store and Months



Gross Sales Value by Store and Products

Store Name / Category	Bits & Bytes Gross Sales Value	Bits & Bytes Gross Sales Volume	Circuit City Gross Sales Value	Gro
⊟ Sports & Fitness	105 753,57 €	6817	108 856,90 €	
⊞ Running Shoes	105 753,57 €	6817	108 856,90 €	
⊞ Home & Kitchen	214 809,07 €	13985	218 100,26 €	
⊞ Clothing & Accessories	342 732,25 €	22625	350 544,18 €	
Total	2 112 842,84 €	137570	2 083 535,30 €	

Figure O – Gross Sales Value by Store and Product



Gross Sales Value by Product Category

Category	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Clothing & Acc...	806.6K	746.5K	804.6K	803.6K	839.4K	565.9K	563.7K	556.6K	544.0K	555.1K	547.3K	566.2K
Electronics	3.4M	3.1M	3.3M	3.3M	3.3M	2.3M	2.3M	2.3M	2.2M	2.3M	2.2M	2.3M
Home & Kitch...	511.4K	469.9K	498.8K	478.7K	500.9K	338.2K	347.9K	362.6K	339.4K	348.5K	344.3K	351.8K
Sports & Fitne...	254.1K	231.5K	259.2K	242.2K	255.7K	173.0K	164.1K	177.5K	168.2K	173.1K	172.5K	170.2K

Figure P – Gross Sales Value by Product Category

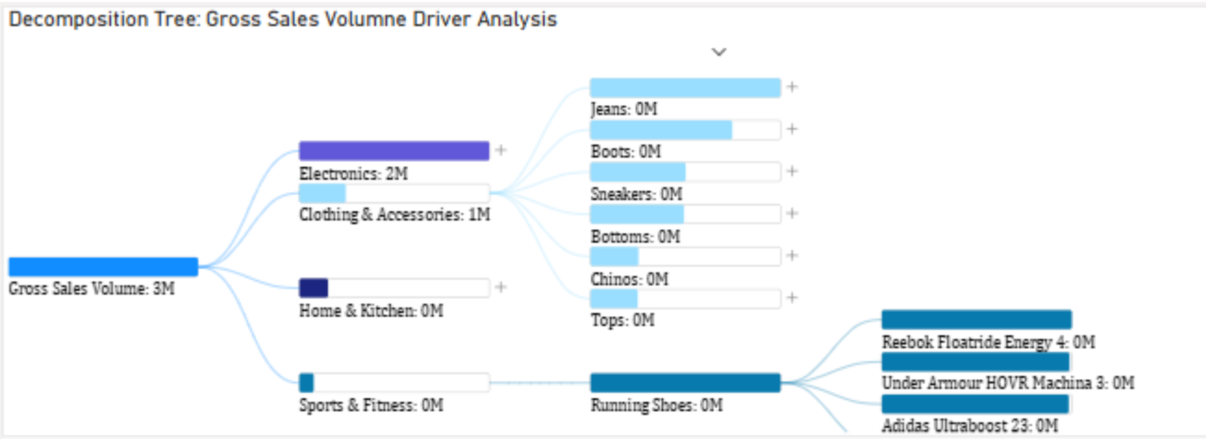


Figure Q – Decomposition Tree: Gross Sales Volume Driver Analysis

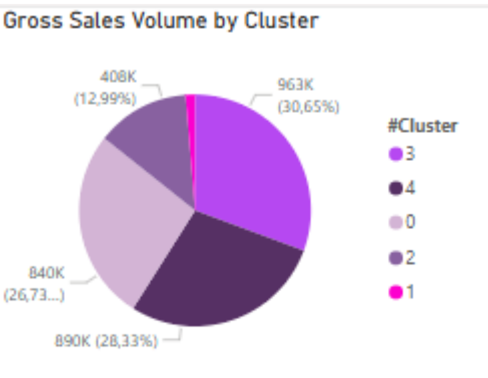


Figure R – Gross Sales Volume per Cluster

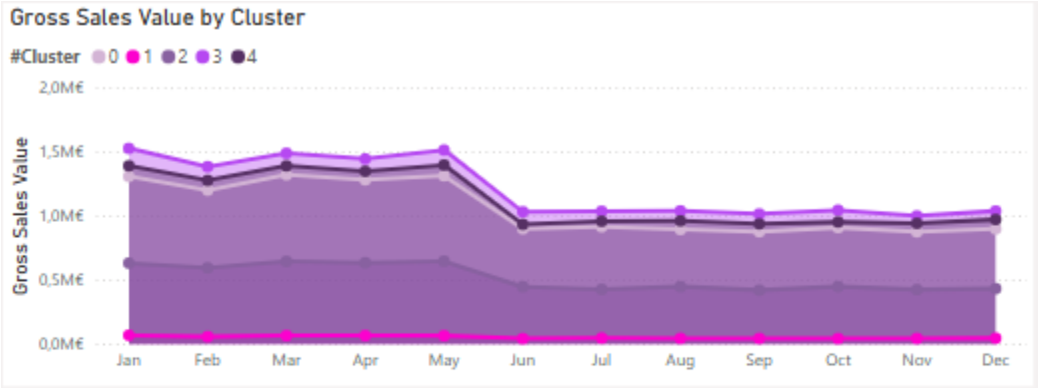


Figure S – Gross Sales Value per Cluster

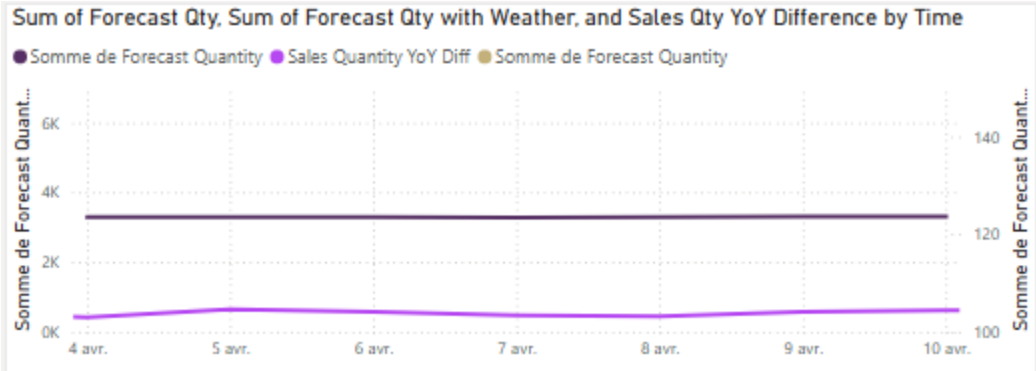


Figure T – Sum of Forecast Qty, Sum of Forecasr Qty with Weather, and Sales Qty YoY Difference by Time

Time

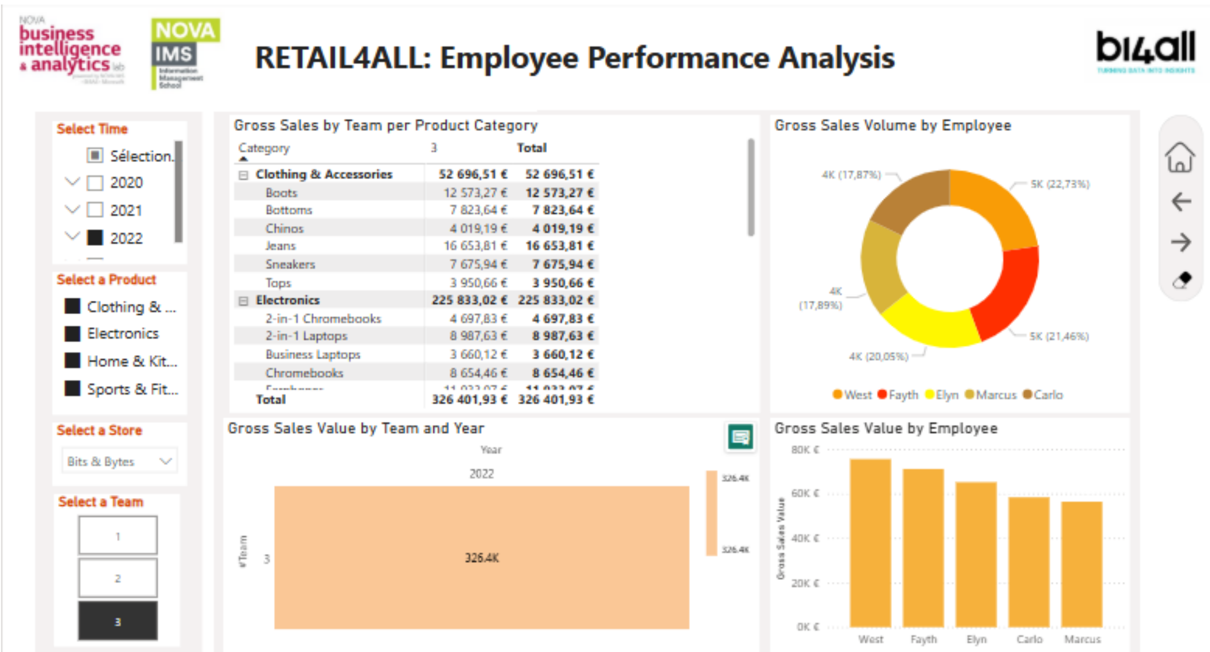


Figure U – Bookmark 2022 – B&B – 2022

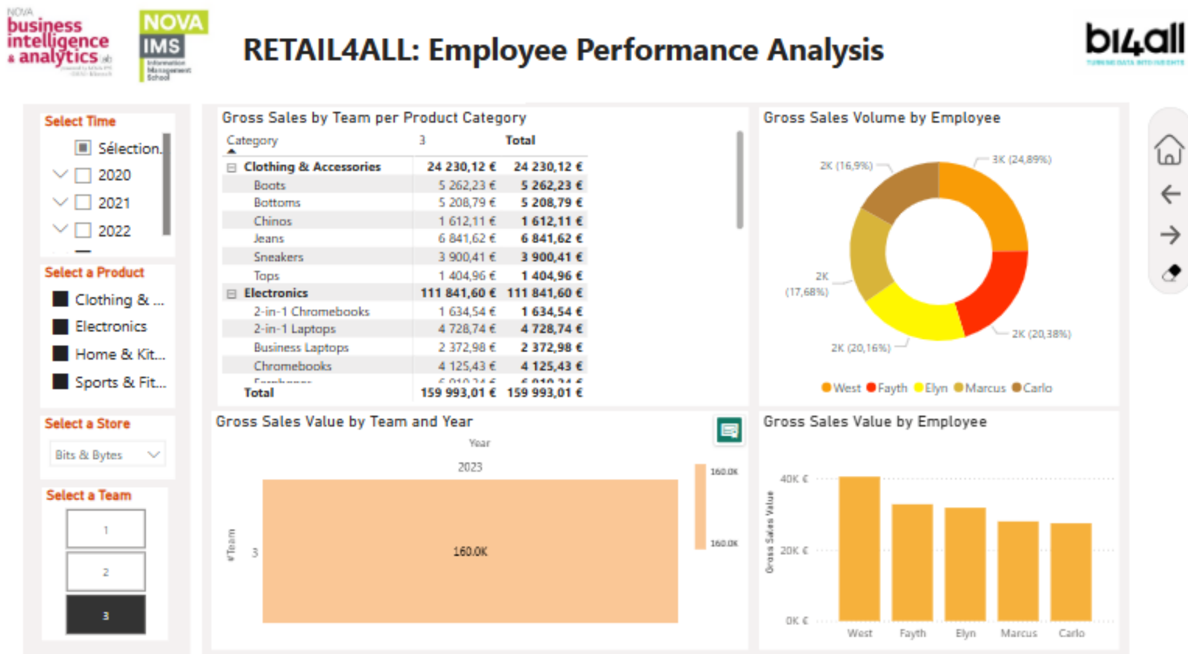


Figure V – Bookmark 2022 – B&B – 2023

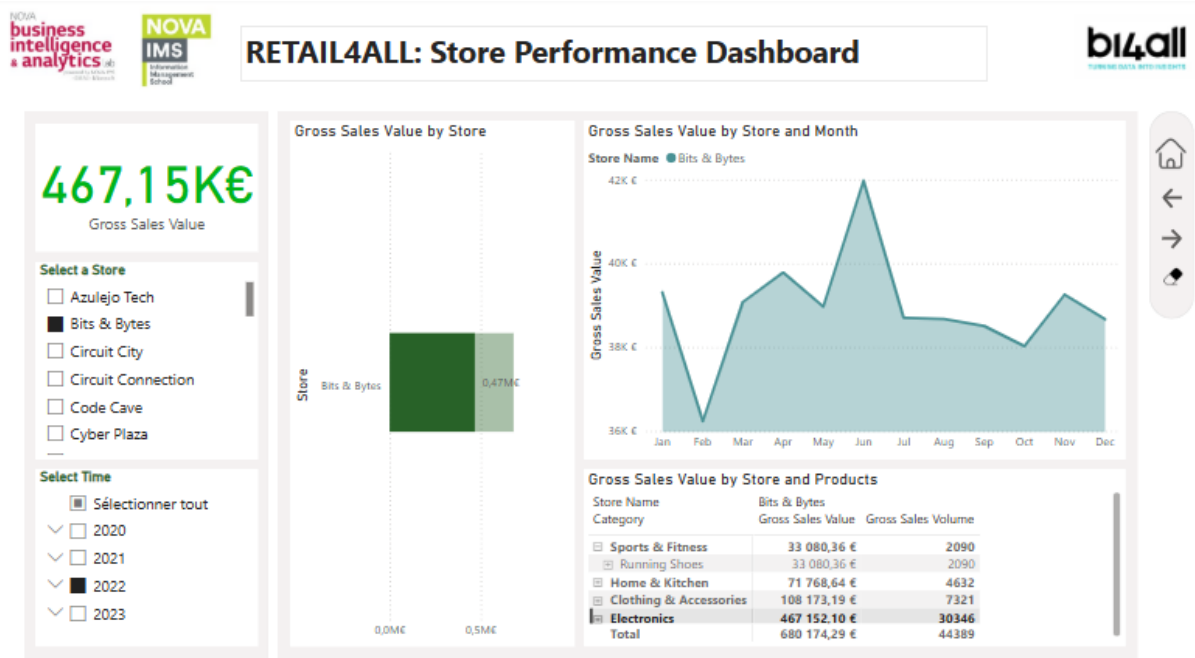


Figure W – Bookmark 2022 – B&B – Electronics

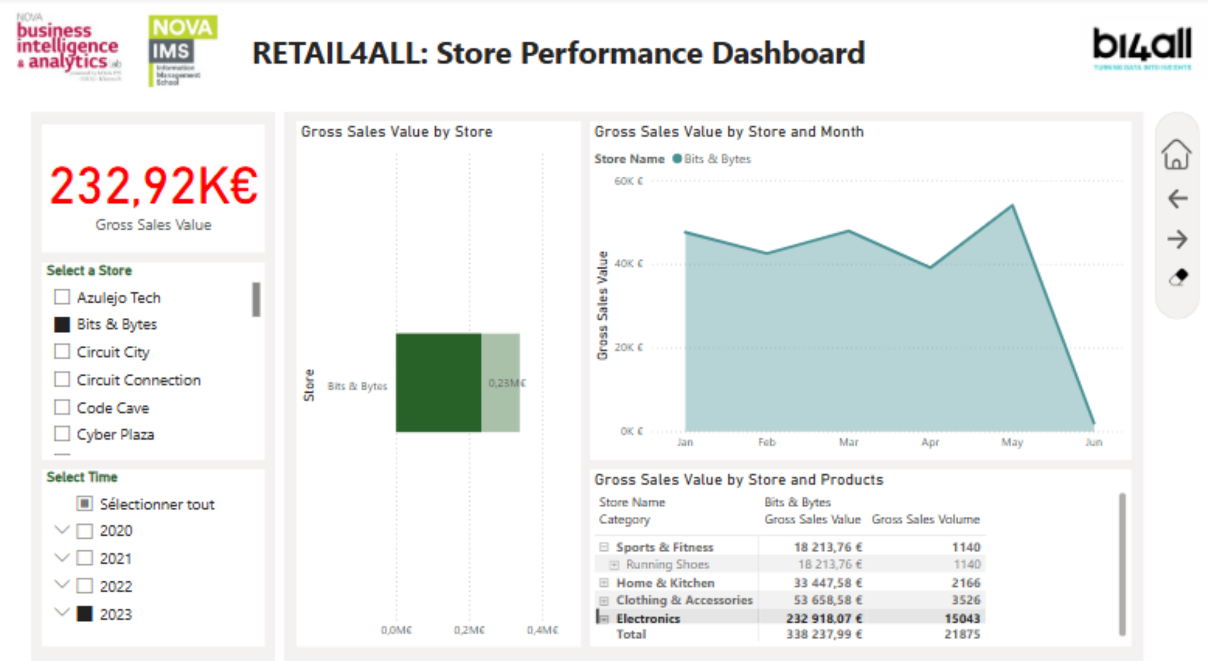
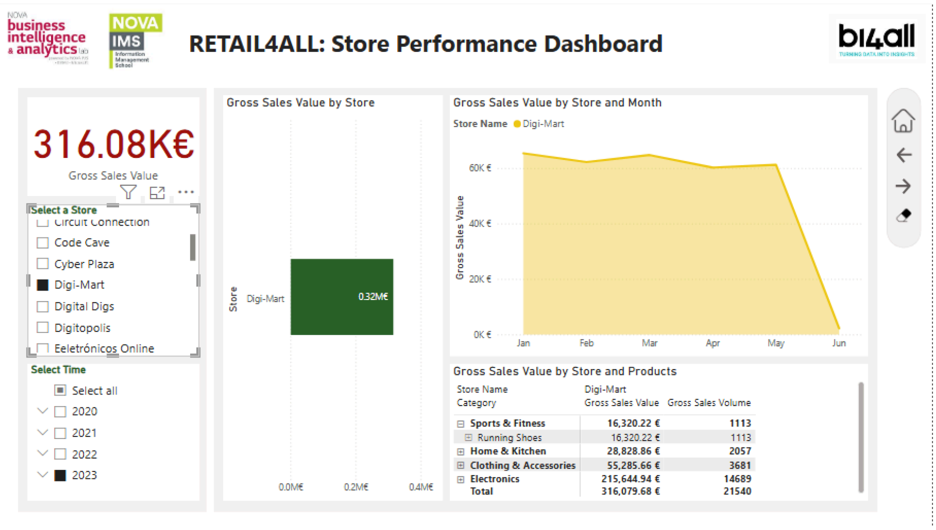
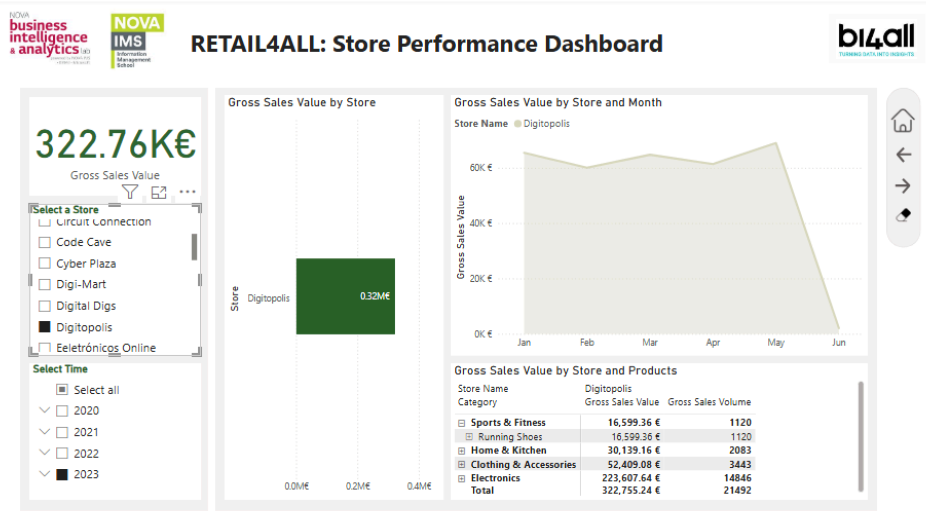


Figure X – Bookmark 2023 – B&B – Electronics





Figures Y – Threshold

Old Name	New Name	Data Type Change	Format Change
Is_Special	Holidays	-	-
Day_Number	#Day	-	-
Full_Date	Full Date	-	-
Month_Name	Month Name	-	-
Month_Name_Short	Month Name Short	-	-
Month_Number	#Month	-	-
Proper_Date	Proper Date	-	*3/14/2001(m/d/yyyy)
Quarter_Name	Quarter Name	-	-

Quarter_Name_Short	Quarter Name Short	-	-
Quarter_Number	#Quarter	-	-
Semester_Name	Semester Name	-	-
Semester_Name_Short	Semester Name Short	-	-
Semester_Number	#Semester	-	-
SK_Date	SK Date	-	-
Weekday_Name	Weekday Name	-	-
Weekday_Name_Short	Weekday Name Short	-	-
Weekday_Number	#Weekday	-	-
Weekday_Type	Weekday Type	-	-
Year	-	-	-

Table A – DIM_Date format changes

Old Name	New Name	Data Type Change	Format Change
BK_OperatorID	BK OperatorID	-	-
First_Name	First Name	-	-
Gender	-	-	-
Last_Name	Last Name	-	-
Role	-	-	-
SK_OperatorID	SK OperatorID	-	-
Team_Number	#Team	-	-

Table B – DIM_Operator format changes

Old Name	New Name	Data Type Change	Format Change
BK_POSID	BK POS ID	-	-

POS_Name	POS Name	-	-
SK_POSID	SK POS ID	-	-

Table C – DIM_POS format changes

Old Name	New Name	Data Type Change	Format Change
BK_SKU	BK SKU	-	-
Category	Category	-	-
Product_Name	Product Name	-	-
SK_ProductID	SK Product ID	-	-
Subcategory	Subcategory	-	-

Table D – DIM_Product format changes

Old Name	New Name	Data Type Change	Format Change
Date	-	Date	*3/14/2001(m/d/yyyy)
District	-	-	-
Rain	-	-	-
SK Weather	SK Weather	-	-
Temperature	Average Temperature	Need to check	

Table E – DIM_Weather format changes

Old Name	New Name	Data Type Change	Format Change
Cluster_Number	#Cluster	-	-
FK_SKU	FK SKU	Text	-

Table F – DIM_Product_Cluster format changes

Old Name	New Name	Data Type Change	Format Change
FK_Date	FK Date	-	-
forecast_quantity	Forecast Quantity	-	-

Table G – DIM_Forecast_Weather_Quantity format changes

Old Name	New Name	Data Type Change	Format Change
Amount	-	-	Decimal Number
FK_Date	FK Date	-	-
FK_OperatorID	FK Operator ID	-	-
FK_POSID	FK POS ID	-	-
FK_ProductID	FK Product ID	-	-
FK_StoreID	FK Store ID	-	-
Quantity	Quantity	-	-

Table H – FACT_Sales format changes

Old Name	New Name	Data Type Change	Format Change
FK_Date	FK Date	-	-
forecast_quantity	Forecast Quantity	-	-

Table I – FACT_Forecast_Quantity format changes

Old Name	New Name	Data Type Change	Format Change
Cluster_Number	#Cluster	-	-
FK_SKU	FK Product ID	Whole number	-

Table J – FACT_Product_Cluster format changes

Old Name	New Name	Data Type Change	Format Change
BK_StoreID	BK Store ID	-	-
City	-	-	-
Country	-	-	-
District	-	-	-
Online	-	-	-
SK_StoreID	SK Store ID	-	-
Store_Name	Store Name	-	-
Store_Type	Store Type	-	-

Table K – DIM_Store format changes