BI System Specification Documentation

Tal Shani

Version 5

12/2/2024

Contents

Document versions - change tracking:

- 24/01/2024 – Initial content – V1
- 29/01/2024 – Functional characterization update, ETL processes chapter and DM tables – V2
- 30/01/2024 – The final ERD update + PBI visualization indicators chapter – V3
- 05/02/2024 - Visualization chapter – V4
- 07/02/2024 – The Power BI Service process update – distribution, scheduling, application – V5

1.   General

1.1.   Project objectives

The purpose of the project is to provide a Full-Scale BI Solution Creation from PriorityERP Database for Zara, a Spanish company that is a well-known international Fashion retailer.

This project aims to establish a comprehensive BI solution leveraging data from the PriorityERP system for Zara. The solution will encompass summarized data tables, with a focus on sales data, alongside customer information, employee records, product details, stores, dates, and more. The BI solution includes dashboards and reports to assist management, department heads, and sales managers in gaining insights into customer preferences, behaviors, and loyalty. The goal is to optimize marketing campaigns and reduce costs through targeted strategies.

1.2. Project Contents

In this project, we will build a Data Mart that will contain information about sales data.

1. Data Cleaning and Preparation: Prior to analysis, we will need to perform thorough data cleaning and preparation to ensure their quality and consistency.
2. Main summary tables to be built for the company's needs.

- **FactSales –** Information about all the orders, which product in which quantity. Data loading process for this table will be incremental.
- **Dim_Products** – Information about the products divided by categories and subcategories.
- **Dim_Customers** – Information about the company's customers.
- **Dim_Employees** – Information about the company's employees.
- **Dim_Stores** – Information about the company's stores.

History Management Table:

- **Transfertable** - Information about all the updates of the tables.
- **Dim_Products_History** - Information about the historical record of changes to product information. The product history table will be included to track changes in products over time using Slowly Changing Dimensions (SCD) Type 4.

The ERD of the tables shown in the attached link: ERD

Source To Target document in this link: S2T

3. The project will contain reports & dashboard in power bi that will contribute to the achievement of the project's goal:

- **Dashboard:**
  In the dashboard, you'll find detailed insights about the company. This information is tailored for senior management, providing a comprehensive overview of the company's performances. The dashboard highlights key metrics, including Resell Percentage and average basket size, allowing us to assess performance effectively.
  The dashboard will display the following information:

  - General KPI data.
  - Resell Percentage by Continents.
  - AVG Basket Size by Country.
  - Physical Stores VS Online Store.
  - YTD Sales VS LY Sales.
  - YTD Units VS LY Units.
  - YTD Orders VS LY Orders.
  - Top 5 Selling Products.
  - Top 5 Employees by sales.

- **Sales Department Analysis:**
  The Sales Department will conduct a thorough analysis of sales data, seller performance, and revenue to evaluate market trends, refine sales strategies, and make year-on-year comparisons. The primary focus will be on identifying top-performing products and recognizing the best employees based on sales performance.
  The report will display the following information:

  - General KPI data.
  - Employees sales performance - the total sales YTD and comparison to the same period last year, along with a percentage change, KPI, and sparkline for sales amount, units' amount, and orders amount.
  - The top 5 best-selling products and their quantities.
  - The top 5 best employees by sales.

- **Sales Representative Analysis:**
  The Customer Sales Analysis report examines vital data and trends to provide insights into our business performance, with the goal of improving sales strategies in each country. Additionally, the report includes a comparative analysis between physical store sales and online sales.
  The report will display the following information:

  - General KPI data
  - Top 10 State Details

- Physical Stores VS Online Store by Country
- Total Sales by Country
- Monthly percentage Change in Units Sold.
- Monthly percentage Change in Total Sales
- Monthly percentage Change in Order Count
- Monthly percentage Change in ARPU (Average Revenue Per User)

2. Gantt

❖ Gantt_link

3. Technical Specification

3.1. Prerequisites

- SQL Server: ERP system in the operational DB (PriorityERP)- tables, data (SQL files)
- SSIS: ETL processes using SSIS in Visual Studio
- Data refresh processes through the definition of JOBS in SSMS
- Power BI: Creating reports and dashboards using Power BI

3.2. Solution Architecture

HLD:

Data collection and exploration from the ERP system will be performed in SQL Server. The data will undergo an ETL process for organization and arrangement into a Data Mart using SSIS. Finally, the presentation of measures in reports and visuals will be presented in Power BI.
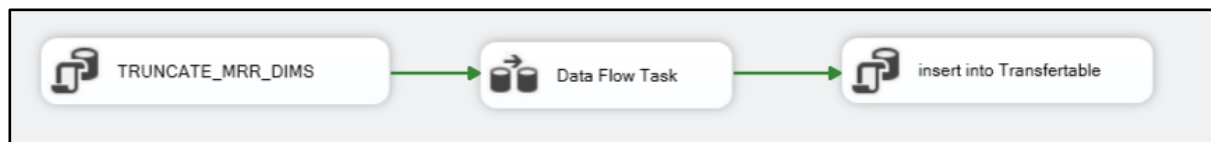
4. Functional Specification

4.1. ETL processes.

1. MRR_Dim package:

Into the Execute SQL Task - truncate_mrr:

This stored procedure, named truncate_mrr_tables, is designed to truncate (delete all rows) from multiple tables.

Finally, we insert the values to the transfer table. The TransferTable serves as a comprehensive log, meticulously capturing every update and insertion step as data moves through the stages from the database (DB) to the Data Mart.

In the Data Flow- dim_mrr - We will transfer information from OLTP tables to MRR tables
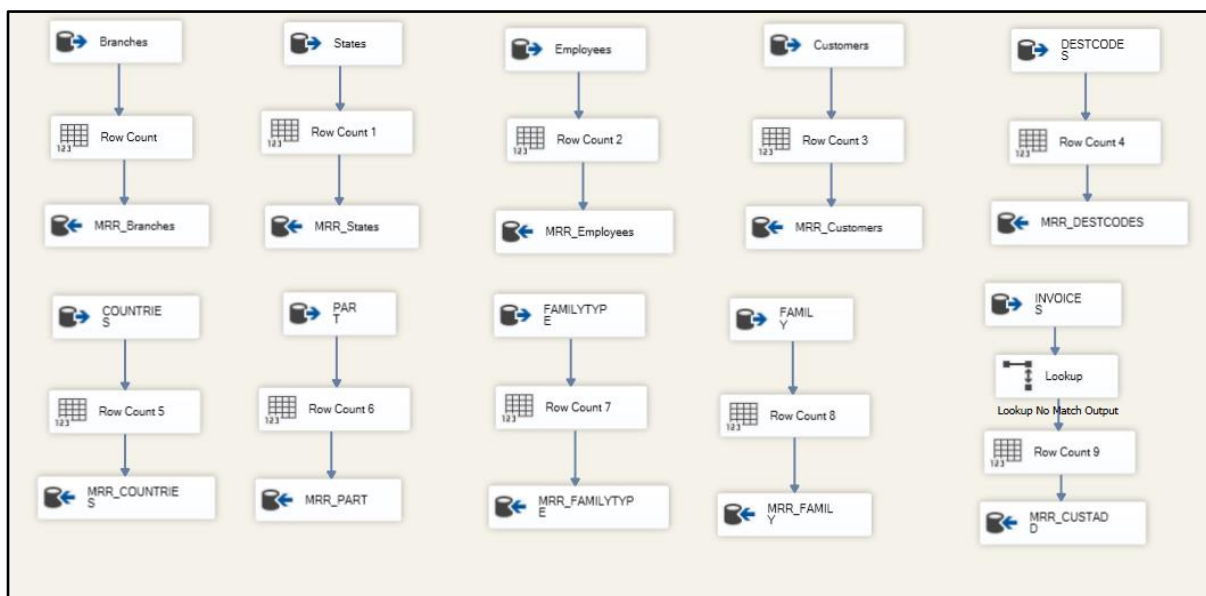

Initially, we migrated all customer data to our "DIM" (Dimension) table. Subsequently, we established a business rule stipulating that only customers who made purchases within the last three years would be considered. This rule was implemented to enable the identification and display of customers who have not engaged in any transactions within the last three years, marking them as inactive. This approach ensures that we are working with the most current and updated data, allowing us to make informed decisions based on the latest trends and developments.

We use a lookup in the Invoices table because it serves as our transaction table, utilized in both the MRR dimension package and MRR fact package.
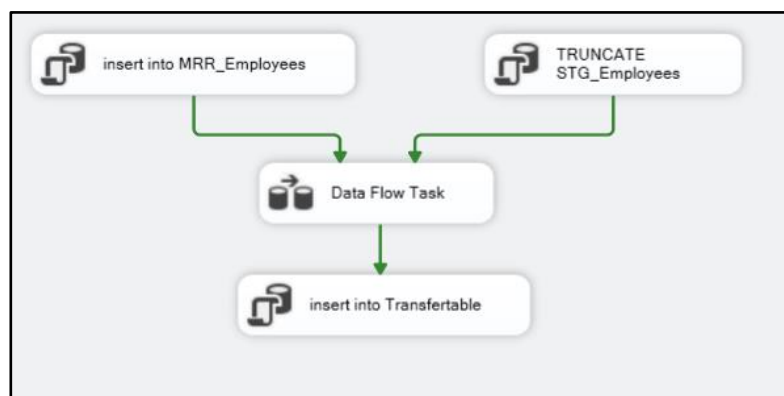
2. MRR_Fact package:

Into the Lookup - Invoices & InvoiceItems :
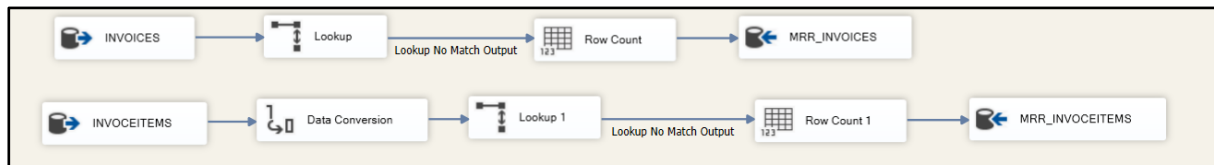
We used the LOOKUP transformation to load only the new rows to the MRR table. The target was to identify the gap (only the new rows from the operational DB that don't exist in the fact sales)

3. STG_Dim_Employees package:

In the Control Flow:

We updated the MRR_Employees table manually with a SQL command that adds a new employee that is associated with the Employee ID "77777". We want to make sure that orders placed online are properly associated with the employee we manually added to the MRR_Employee table. We want to maintain the connection between this employee and the online store that contains the order information.

In the Data Flow:
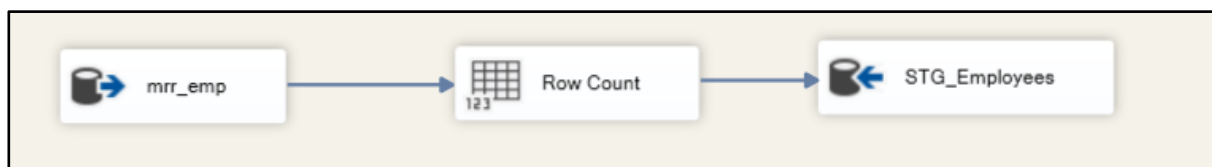
We collected information about employees from the employees' table. The selected columns include details such as Emp_ID, First_Name, Last_Name, Job_Title, Hire_Date, Phone_Number, Email_Address etc. Additionally, we have selectively retrieved specific columns from the 'MRR_Employees' table, focusing on entries where the 'skill' column contains the term 'sales.' Finally, we will insert the information into the stg_Employees table.

4. STG_Products package
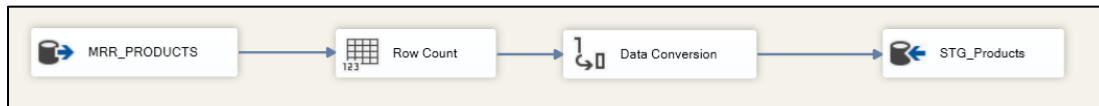
We will collect information about products from three tables: MRR_Part, MRR_Familly and MRR_FamillyType. The selected columns include details: ProductID, Product_Name, Category_Name and Sub_Category_Name.In the Data Conversion phase, we will meticulously refine data types that necessitate modification. Finally, we will insert all the information into the stg_products table.

5. STG CUSTOMERS package

In the Control Flow:

We updated the MRR_Customers table manually with SQL command, that every customer who has an order in the invoices table, and does not associated with any store id, will receive the store id "99999" as a sign for "Online Store". We made this in order to maintain the association between the orders and the stores through the customers.

In the Data Flow:
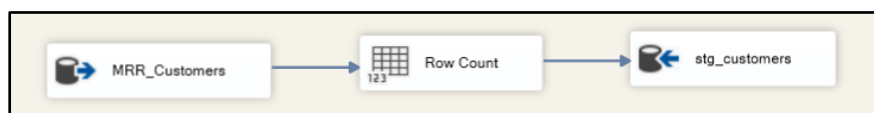
We collected information about Customers from the Customers, Destcode, State and Country tables. The selected columns include details such as Customer ID, Name, store ID, Address, City, State, Country.
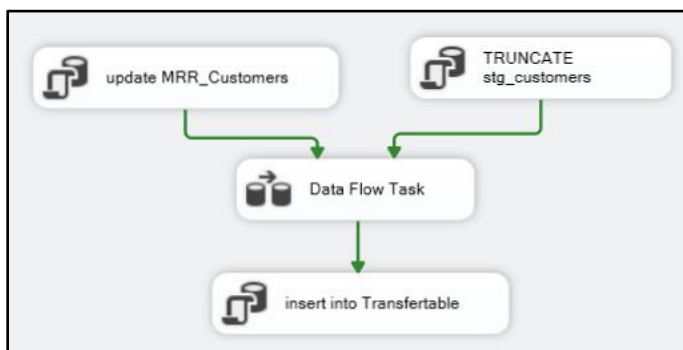
6. STG STORES

In the Control Flow:
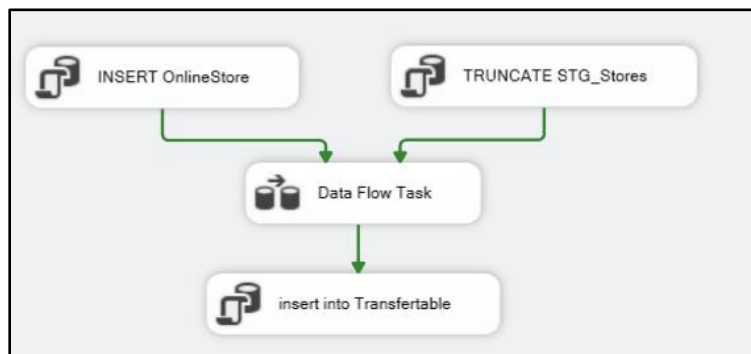
We updated the MRR_Branches table manually with SQL command with a new Online Store that is associated with the Store ID "99999".

In the Data Flow:
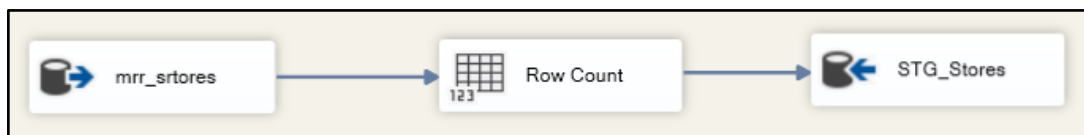
We will collect information about stores from five tables: MRR_Branches,MRR_Customers, MRR_CustAdd , MRR_Destcodes and MRR_States. The selected columns include details such as StoreId, StoreName and State. Additionally, we add to our join query a statement that if the branch ID is 99999, it's labeled as 'OnlineStore'; otherwise, it takes the actual state name from the 'MRR_States' table based on the relationship between the tables established in the LEFT JOIN operations.

7. STG SALES

In the Data Flow:
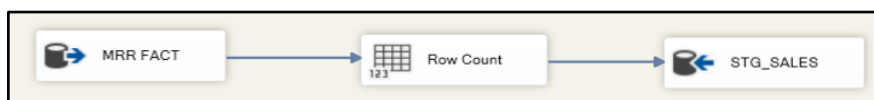
We will collect sales order information, including order details and product specific details, connecting the two tables MRR_INVOICES and MRR_INVOCEITEMS. Based on IV.
The selected columns include the details: OrderID, OrderDate, CustomerID, EmpID, ZoneID, ProductID, Qty, Price and Discount.

8. DW EMPLOYEES
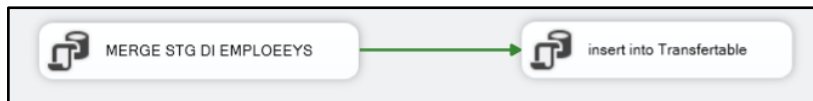
Into the Execute SQL Task – MERGE:

Using to synchronize data between the Dim_Employees and stg_employees tables based on EMP_ID. It performs the following actions:

- Insert: If there is a record in stg_employees that does not match Dim_Employees, insert a new record.
- Update: If there is a match and certain columns have changed, update the corresponding columns in Dim_Employees with values from stg_employees.
- Update (IsActive): If there is a record in Dim_Employees that does not exist in stg_Employees, set IsActive to 0 in Dim_Employees.

9. DW PRODUCTS

The query updates records in the dim_Products table in a database. The general description of the query is as follows:
The query iterates over each record in the dim_Products table and updates the columns UpdateDate and ISactive. The update conditions are complex and include:
The ProductID column should match the values returned by a subquery that retrieves Dim_ID from the join of the PART table in the PriorityERP database with the dim_Products table.
The ISactive column should be different from 0.
The record must be identified based on the criterion that its ProductID is not found in the PART table in the PriorityERP database.
After the update, the UpdateDate column is set to the current timestamp, and the ISactive column is set to 0.
We implemented this query as part of our Slowly Changing Dimension (SCD) strategy, which inherently manages new dates and updated data. Specifically, we created this query to address situations where records have been deleted. The query ensures that our 'dim_Products' table accurately reflects changes in external data, including the handling of deleted lines.
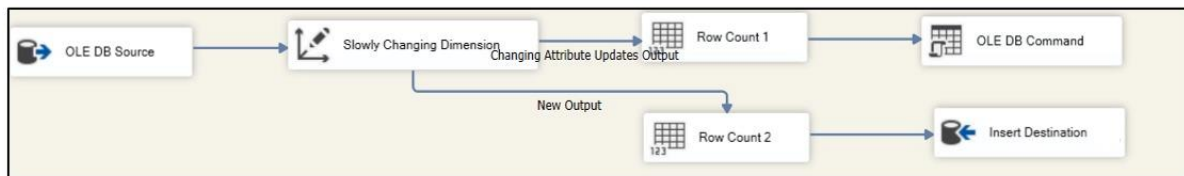
In the Slowly Changing Dimensions (SCD), the approach involves distinguishing between new lines and updated lines using the 'ProductID.' In the process of updating using the OLE DB Command, we focus on the changed fields. New lines, identified by a unique 'ProductID,' are handled separately to ensure proper management of evolving data. Meanwhile, for updated lines, the OLE DB Command's update operation selectively targets and updates only the fields that have changed.

10. DW CUSTOMERS

Using to synchronize data between the Dim_Customers and stg_Customers tables based on CustomersID. It performs the following actions:

- Insert: If there is a record in stg_Customers that does not match Dim_Customers, insert a new record.
- Update: If there is a match and certain columns have changed, update the corresponding columns in Dim_Customers with values from stg_Customers.
- Update (IsActive): If there is a record in Dim_Customers that does not exist in stg_Customers, set IsActive to 0 in Dim_Customers.

11. DW STORES
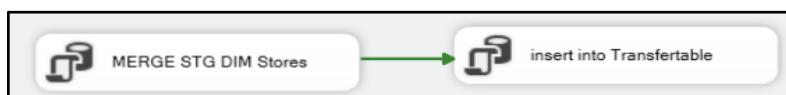
Using to synchronize data between the Dim_ Stores and stg_ Stores tables based on StoreID. It performs the following actions:

- Insert: If there is a record in stg_ Stores that does not match Dim_ Stores, insert a new record.
- Update: If there is a match and certain columns have changed, update the corresponding columns in Dim_ Stores with values from stg_Stores.
- Update (IsActive): If there is a record in Dim_ Stores that does not exist in stg_ Stores, set IsActive to 0 in Dim_ Stores.

12. DW FACT SALES

In the OLE DB Source - STG_SALES, we will channel comprehensive data into the FACT_Sales table. Subsequently, in the Data Conversion phase, we will meticulously refine data types that necessitate modification. Following this, within the Derived Column transformation, we will dynamically compute the Total column using the specified formula: (Qty * Price * (1 - Discount)). Finally, the meticulously transformed and calculated data will be inserted into the DWH_FACT_SALES table.

13. DW PRODUCTS HISTORY

The query updates records in the Dim_Products_History table. Here's a brief description:

The UPDATE statement modifies the EndDate column in the Dim_Products_History table. It sets the EndDate to the current timestamp for records that meet the following conditions:

- The ProductID is not found in the PART table of the PriorityERP database.
- The EndDate is currently NULL.

In summary, the query effectively marks records in the history table as ended for products that are no longer present in the PART table and have a NULL EndDate. This action helps differentiate between products actively selling in the company and those that are no longer part of the inventory history.
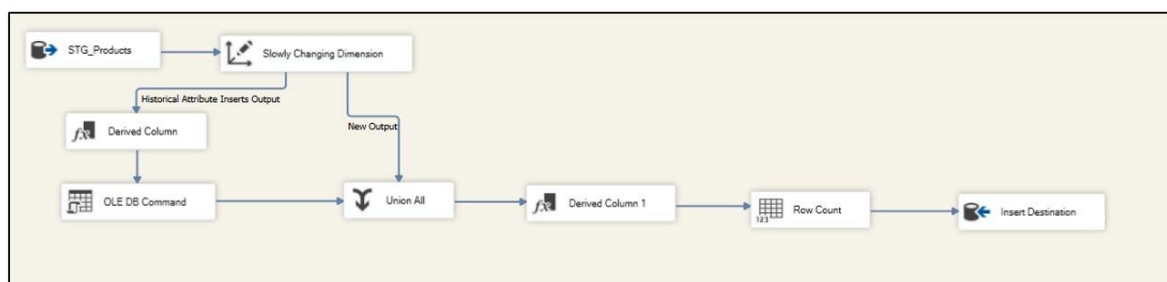
In the "dim_Products_History" package, when a record in the "dim_Products" table gets updated, we store its previous version in the "dim_Products_History" table. This historical table is structured like

"dim_Products" but includes extra date fields indicating when each version began and ended. This approach enables us to track changes over time, providing a clear timeline for the evolution of product information in the data model. It's a valuable strategy for maintaining a historical record and ensuring transparency in understanding how product data has changed.

14. DEV – DROP & CREATE DATABASE
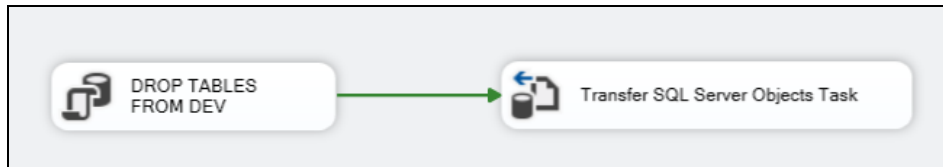
In this package, we've created a procedure in our development environment (using SSMS) that drops all the tables. Subsequently, we use an SQL task to execute this procedure. Afterward, we employ the Transfer SQL Server Objects Task to copy all tables from the production to the development environment. This entire process ensures that our development database is always up to date with what's in production.

4.2. Description of data tables in the Data Mart (DM):

- **Fact_Sales**

In the FactSales table, we have the following columns:

1. **OrderID**: The unique identifier for each order, serving as the primary key for the table.
2. **Order_Date:** The date when the order was placed, providing a timestamp for each transaction.
3. **CustomerID**: The identifier for the customer associated with the order, serving as a foreign key referencing the Dim_Customers table.
4. **EmpID**: The identifier for the employee responsible for making the order, serving as a foreign key referencing the Dim_Employees table.
5. **ZoneID**: The identifier for the geographical territory where the order was placed.
6. **ProductID**: The identifier for the products purchased in the order, serving as a foreign key referencing the Dim_Products table.
7. **Qty**: The quantity of each product purchased in the order.
8. **Price**: The unit price of each product, providing the cost per unit.
9. **Discount**: The discount applied to each unit, influencing the overall cost.
10. **Total**: The total cost for each product purchased, inclusive of taxes (calculated as: Price * Qty * (1 - Discount)). This represents the final amount to be paid for the specified quantity of each product, factoring in the discount.

OrderID	OrderDate	CustomerID	EmpID	ProductID	ZoneID	Price	Quantity	Discount	Total
50205	2013-03-30 00:00:00.000	29865	274	835	4	324.4527	1	0.00	324.45
55298	2013-08-30 00:00:00.000	29623	274	947	4	54.942	1	0.00	54.94
51786	2013-06-30 00:00:00.000	29666	274	965	6	334.0575	3	0.15	851.84
57111	2013-09-30 00:00:00.000	30024	274	868	1	41.994	5	0.00	209.97
50205	2013-03-30 00:00:00.000	29865	274	760	4	469.794	1	0.00	469.79
43849	2011-07-01 00:00:00.000	29818	274	776	1	2024.994	2	0.00	4049.98
51786	2013-06-30 00:00:00.000	29666	274	715	6	29.994	7	0.00	209.95
61200	2013-11-30 00:00:00.000	29987	274	972	4	728.91	4	0.00	2915.64
50205	2013-03-30 00:00:00.000	29865	274	761	4	469.794	1	0.00	469.79
55298	2013-08-30 00:00:00.000	29623	274	896	4	200.052	5	0.00	1000.26
51786	2013-06-30 00:00:00.000	29666	274	969	6	1430.442	3	0.00	4291.32
57111	2013-09-30 00:00:00.000	30024	274	783	1	1376.994	2	0.00	2753.98
50205	2013-03-30 00:00:00.000	29865	274	712	4	5.1865	3	0.00	15.55
43849	2011-07-01 00:00:00.000	29818	274	714	1	28.8404	1	0.00	28.84
51786	2013-06-30 00:00:00.000	29666	274	954	6	953.628	1	0.20	762.90
61200	2013-11-30 00:00:00.000	29987	274	865	4	38.10	1	0.00	38.10
50205	2013-03-30 00:00:00.000	29865	274	770	4	469.794	2	0.00	939.58
55298	2013-08-30 00:00:00.000	29623	274	979	4	445.41	2	0.00	890.82
51786	2013-06-30 00:00:00.000	29666	274	877	6	4.77	2	0.00	9.54
57111	2013-09-30 00:00:00.000	30024	274	782	1	1376.994	1	0.00	1376.99
50205	2013-03-30 00:00:00.000	29865	274	854	4	44.994	3	0.00	134.98
43849	2011-07-01 00:00:00.000	29818	274	777	1	2024.994	2	0.00	4049.98

- **Dim_Products:**

In the Dim_Products table, we have the following columns

1. **ProductID**: The unique identifier and primary key for each product, serving as the key reference in the table.
2. **ProductName**: The name of the product, providing a clear and concise label for each item.
3. **SubCategoryName**: The name of the subcategory to which the product belongs, offering additional categorization.
4. **CategoryName**: The name of the main category to which the product belongs, providing a higher-level classification.
5. **Update_Date:** The date of the last update for the product, indicating the most recent modification or revision to the product information.
6. **IsActive:** A binary indicator (True/False) denoting the current active status of the products within the organization.

	ProductID	Product_Name	Category_Name	Sub_Category_Name	UpdateDate	IsActive
1	753	Ultra Warm Hybrid Down Coat	Coats and jackets	Coat	2024-01-29 17:52:09.607	1
2	779	Ultra Light Down Jacket	Coats and jackets	Winter Jackets	2024-01-29 17:52:09.613	1
3	780	U Crew Neck T-Shirt	Shirts	Graphic Tees	2024-01-29 17:52:09.613	1
4	781	Ultra Stretch Color Jeans	Pants	Denim	2024-01-29 17:52:09.613	1
5	782	Flannel Checked Shirt	Shirts	Casual	2024-01-29 17:52:09.613	1
6	783	Wide-Fit Pleated Pants	Pants	Formal Trousers	2024-01-29 17:52:09.613	1
7	784	Cotton Crew Neck T-Shirt	Shirts	Casual	2024-01-29 17:52:09.613	1
8	793	Slim Black Pants	Pants	Formal Trousers	2024-01-29 17:52:09.617	1
9	794	Long-Sleeve T-Shirt	Shirts	Casual	2024-01-29 17:52:09.617	1
10	795	Fluffy Yarn Fleece Full-Zip Jacket	Coats and jackets	Winter Jackets	2024-01-29 17:52:09.617	1

- **Dim_Customer:**

In the Dim_Customer table, we have the following columns

1. **CustomerID**: The unique identifier and primary key for each customer, serving as a key reference within the table.
2. **Name**: The full name of the customer, providing comprehensive identification.
3. **StoreID:** The identifier for the store associated with the customer. This column establishes a relationship with the Store dimension, providing information about the specific store where the customer is associated.
4. **Address**: The customer's street address, offering specific location details.
5. **City**: The city where the customer resides, providing information about the customer's urban location.
6. **State**: The state within the country where the customer resides.
7. **Country**: The country of residence for the customer, indicating the nation from which the customer originates.
8. **Update_Date:** The date of the last update for customer information, indicating the most recent modification or revision to the customer details.
9. **IsActive**: A binary indicator (True/False) denoting the current active status of the products within the organization.

CustomerID	Name	storeid	Address	City	State	Country	UpdateDate	ISactive
29481	Ivan Suri	99999	Knaackstr 4	Hof	Bayern	Germany	2024-01-28 10:39:46.700	1
29482	Clayton Zhang	99999	1080, quai de Grenelle	Saint Ouen	Charente-Mariti...	France	2024-01-28 10:39:46.700	1
29483	Jésus Navarro	99999	244, rue de la Centenaire	Paris La Defe...	Hauts de Seine	France	2024-01-28 10:39:46.700	1
29484	Gustavo Achong	292	Mall Of Memphis	Memphis	Tennessee	United States	2024-01-28 10:39:46.700	1
29485	Catherine Abel	294	57251 Serene Blvd	Van Nuys	California	United States	2024-01-28 10:39:46.700	1
29486	Kim Abercrombie	296	Tanger Factory	Branch	Minnesota	United States	2024-01-28 10:39:46.700	1
29487	Humberto Aceve...	298	Johnny Appleseed Shop.c...	Mansfield	Ohio	United States	2024-01-28 10:39:46.700	1
29488	Pilar Ackerman	300	4250 Concord Road	Rhodes	New South Wales	Australia	2024-01-28 10:39:46.700	1
29489	Frances Adams	302	6900 Sisk Road	Modesto	California	United States	2024-01-28 10:39:46.700	1
29490	Margaret Smith	304	Lewiston Mall	Lewiston	Idaho	United States	2024-01-28 10:39:46.700	1
29491	Carla Adams	306	Leesburg Premium Outlet ...	Leesburg	Virginia	United States	2024-01-28 10:39:46.700	1
29492	Jay Adams	308	Blue Ridge Mall	Kansas City	Missouri	United States	2024-01-28 10:39:46.700	1
29493	Ronald Adina	310	Hilton Head Factory Outlet...	Bluffton	South Carolina	United States	2024-01-28 10:39:46.700	1
29494	Samuel Agcaoili	312	No. 25800-130 King Stree...	Toronto	Ontario	Canada	2024-01-28 10:39:46.700	1
29495	James Aguilar	314	Knaackstr 7	Paderborn	Hamburg	Germany	2024-01-28 10:39:46.700	1
29496	Robert Ahlering	316	6500 East Grant Road	Tucson	Arizona	United States	2024-01-28 10:39:46.700	1
29497	François Ferrier	318	Eastridge Mall	Casper	Wyoming	United States	2024-01-28 10:39:46.700	1
29498	Kim Akers	320	Granite State Marketplace	Hooksett	New Hampshire	United States	2024-01-28 10:39:46.700	1
29499	Amy Alberts	324	252851 Rowan Place	Richmond	British Columbia	Canada	2024-01-28 10:39:46.700	1
29500	Anna Albright	326	Flagler Park Plaza	Miami	Florida	United States	2024-01-28 10:39:46.700	1
29501	Milton Albury	328	Wrentham Village	Wrentham	Massachusetts	United States	2024-01-28 10:39:46.700	1
29502	Paul Alcorn	330	White Mountain Mall	Rock Springs	Wyoming	United States	2024-01-28 10:39:46.700	1
29503	Gregory Alderson	332	26910 Indela Road	Montreal	Quebec	Canada	2024-01-28 10:39:46.700	1
29504	J. Phillip Alexand...	334	Belz Factory Outlet	Pigeon Forge	Tennessee	United States	2024-01-28 10:39:46.700	1
29505	Michelle Alexan...	336	22589 West Craig Road	North Las Ve...	Nevada	United States	2024-01-28 10:39:46.700	1

- **Dim_Employees**:

In the Dim_Employees table we have the following column:

1. **EmployeeID**: The unique identifier and primary key for each employee, serving as a key reference within the table.
2. **First_Name:** The first name of the employee, offering personal identification.
3. **Last_Name**: The last name of the employee, completing the full name for accurate identification.
4. **Job_Title:** The job title or position held by the employee within the organization.
5. **Hire_Date**: The date when the employee was hired, providing insight into their tenure.
6. **Phone_Number:** The contact number associated with the employee, facilitating communication.
7. **Email_Address**: The email address of the employee, serving as a primary mode of professional communication.
8. **Zone:** The name of the zone associated with the employee's responsibilities.
9. **Update_Date:** The date of the last update for Employees information, indicating the most recent modification or revision to the Employees details.
10. **IsActive:** A binary indicator (True/False) denoting the current active status of the Employees within the organization.

EmpID	FirstName	LastName	JobTitle	HireDate	Phone_Number	Email_Address	Zone	UpdateDate	ISactive
273	Brian	Welcker	Vice President of Sales	2011-02-15	716-555-0127	brian3@adventure-works.com	NULL	2024-01-28 10:38:42.177	1
274	Stephen	Jiang	North American Sales Manager	2011-01-04	238-555-0197	stephen0@adventure-works.com	NULL	2024-01-28 10:38:42.177	1
275	Michael	Blythe	Sales Representative	2011-05-31	257-555-0154	michael9@adventure-works.com	Northeast	2024-01-28 10:38:42.177	1
276	Linda	Mitchell	Sales Representative	2011-05-31	883-555-0116	linda3@adventure-works.com	Southwest	2024-01-28 10:38:42.177	1
277	Jillian	Carson	Sales Representative	2011-05-31	517-555-0117	jillian0@adventure-works.com	Central	2024-01-28 10:38:42.177	1
278	Garrett	Vargas	Sales Representative	2011-05-31	922-555-0165	garrett1@adventure-works.com	Canada	2024-01-28 10:38:42.177	1
279	Tsvi	Reiter	Sales Representative	2011-05-31	664-555-0112	tsvi0@adventure-works.com	Southeast	2024-01-28 10:38:42.177	1
280	Pamela	Ansman-Wolfe	Sales Representative	2011-05-31	340-555-0193	pamela0@adventure-works.com	Northwest	2024-01-28 10:38:42.177	1
281	Shu	Ito	Sales Representative	2011-05-31	330-555-0120	shu0@adventure-works.com	Southwest	2024-01-28 10:38:42.177	1
282	Jos?	Saraiva	Sales Representative	2011-05-31	185-555-0169	jos?1@adventure-works.com	Canada	2024-01-28 10:38:42.177	1
283	David	Campbell	Sales Representative	2011-05-31	740-555-0182	david8@adventure-works.com	Northwest	2024-01-28 10:38:42.177	1
284	Tete	Mensa-Annan	Sales Representative	2012-09-30	615-555-0153	tete0@adventure-works.com	Northwest	2024-01-28 10:38:42.177	1
285	Syed	Abbas	Pacific Sales Manager	2013-03-14	926-555-0182	syed0@adventure-works.com	NULL	2024-01-28 10:38:42.177	1
286	Lynn	Tsoflias	Sales Representative	2013-05-30	1 (11) 500 555-0190	lynn0@adventure-works.com	Australia	2024-01-28 10:38:42.177	1
287	Amy	Alberts	European Sales Manager	2012-04-16	775-555-0164	amy0@adventure-works.com	NULL	2024-01-28 10:38:42.177	1
288	Rachel	Valdez	Sales Representative	2013-05-30	1 (11) 500 555-0140	rachel0@adventure-works.com	Germany	2024-01-28 10:38:42.177	1
289	Jae	Pak	Sales Representative	2012-05-30	1 (11) 500 555-0145	jae0@adventure-works.com	United Kingdom	2024-01-28 10:38:42.177	1
290	Ranjit	Varkey Chudukatil	Sales Representative	2012-05-30	1 (11) 500 555-0117	ranjit0@adventure-works.com	France	2024-01-28 10:38:42.177	1
77777	Online	EMP	Online sales	NULL	NULL	NULL	NULL	2024-01-28 10:38:42.177	1

- **DIM_Stores:**

In the Transfer table we have the following column:

1. **StoreID**: This column represents a unique identifier for each store in the DIM_Stores table.
2. **Store Name**: This column stores the name of the store.
3. **State**: The state or region in which the store is located. This column provides additional geographical information about the store's location within a specific state or region.
4. **UpdateDate**: The "UpdateDate" column records the date when information related to a specific store, employee, or their association was last updated.
5. **IsActive**: A binary indicator (True/False) denoting the current active status of the stores within the organization.

StoreId	StoreName	State	UpdateDate	ISactive
292	Next-Door Bike Store	Tennessee	2024-01-28 10:39:28.747	1
294	Professional Sales and Service	California	2024-01-28 10:39:28.747	1
296	Riders Company	Minnesota	2024-01-28 10:39:28.747	1
298	The Bike Mechanics	Ohio	2024-01-28 10:39:28.747	1
300	Nationwide Supply	New South Wales	2024-01-28 10:39:28.747	1
302	Area Bike Accessories	California	2024-01-28 10:39:28.747	1
304	Bicycle Accessories and Kits	Idaho	2024-01-28 10:39:28.747	1
306	Clamps & Brackets Co.	Virginia	2024-01-28 10:39:28.747	1
308	Valley Bicycle Specialists	Missouri	2024-01-28 10:39:28.747	1
310	New Bikes Company	South Carolina	2024-01-28 10:39:28.747	1
312	Vinyl and Plastic Goods Corporation	Ontario	2024-01-28 10:39:28.747	1
314	Top of the Line Bikes	Hamburg	2024-01-28 10:39:28.747	1
316	Fun Toys and Bikes	Arizona	2024-01-28 10:39:28.747	1
318	Great Bikes	Wyoming	2024-01-28 10:39:28.747	1
320	Metropolitan Sales and Rental	New Hampshire	2024-01-28 10:39:28.747	1
322	Irregulars Outlet	NULL	2024-01-28 10:39:28.747	1
324	Valley Toy Store	British Columbia	2024-01-28 10:39:28.747	1
326	Worthwhile Activity Store	Florida	2024-01-28 10:39:28.747	1
328	Purchase Mart	Massachusetts	2024-01-28 10:39:28.747	1
330	Major Sport Suppliers	Wyoming	2024-01-28 10:39:28.747	1
332	Family's Favorite Bike Shop	Quebec	2024-01-28 10:39:28.747	1
334	Global Plaza	Tennessee	2024-01-28 10:39:28.747	1
336	Imported and Domestic Cycles	Nevada	2024-01-28 10:39:28.747	1
338	Systematic Sales	Arizona	2024-01-28 10:39:28.747	1

- **Dim_Products_History**

The table Dim_Products_History contains the same columns as the Dim_Products table (without the Status column), along with the following additional columns:

1. **ProductID**: This is an integer column representing a unique identifier for each product.
2. **Product_Name**: Stores the name of the product.
3. **Sub_Category_Name**: It is intended to store the name of the sub-category to which the product belongs.
4. **Category_Name:** It is intended to store the name of the category to which the product belongs.
5. **InsertDate**: The date when we made any change to the product, such as inserting a new product, changing a characteristic of this product, or deleting the product.
6. **EndDate:** The date when a new change occurs. If it is null, that means it is the last update about the product.

ProductID	Product_Name	Category_Name	Sub_Category_Name	InsertDate	EndDate
680	HL Road Frame - Black, 58	Components	Road Frames	2024-01-24 13:07:55.000	2024-01-24 13:14:58.000
706	HL Road Frame - Red, 58	Components	Road Frames	2024-01-24 13:07:55.000	NULL
707	Sport-100 Helmet, Red	Accessories	Helmets	2024-01-24 13:07:55.000	NULL
708	Sport-100 Helmet, Black	Accessories	Helmets	2024-01-24 13:07:55.000	NULL
709	Mountain Bike Socks, M	Clothing	Socks	2024-01-24 13:07:55.000	NULL
710	Mountain Bike Socks, L	Clothing	Socks	2024-01-24 13:07:55.000	NULL
711	Sport-100 Helmet, Blue	Accessories	Helmets	2024-01-24 13:07:55.000	NULL
712	AWC Logo Cap	Clothing	Caps	2024-01-24 13:07:55.000	NULL
713	Long-Sleeve Logo Jersey, S	Clothing	Jerseys	2024-01-24 13:07:55.000	NULL
714	Long-Sleeve Logo Jersey, M	Clothing	Jerseys	2024-01-24 13:07:55.000	NULL
715	Long-Sleeve Logo Jersey, L	Clothing	Jerseys	2024-01-24 13:07:55.000	NULL
716	Long-Sleeve Logo Jersey, XL	Clothing	Jerseys	2024-01-24 13:07:55.000	NULL
717	HL Road Frame - Red, 62	Components	Road Frames	2024-01-24 13:07:55.000	NULL
718	HL Road Frame - Red, 44	Components	Road Frames	2024-01-24 13:07:55.000	NULL
719	HL Road Frame - Red, 48	Components	Road Frames	2024-01-24 13:07:55.000	NULL
720	HL Road Frame - Red, 52	Components	Road Frames	2024-01-24 13:07:55.000	NULL
721	HL Road Frame - Red, 56	Components	Road Frames	2024-01-24 13:07:55.000	NULL
722	LL Road Frame - Black, 58	Components	Road Frames	2024-01-24 13:07:55.000	NULL
723	LL Road Frame - Black, 60	Components	Road Frames	2024-01-24 13:07:55.000	NULL
724	LL Road Frame - Black, 62	Components	Road Frames	2024-01-24 13:07:55.000	NULL
725	LL Road Frame - Red, 44	Components	Road Frames	2024-01-24 13:07:55.000	NULL

- **Transfertable**:

In the Transfer table we have the following column:

1. **Package Name**: Introducing a new layer of detail, the Package Name column provides a descriptive marker for the overarching package under which these record updates are executed.
2. **Table Name**: This element identifies the specific table in which record updates have transpired, providing a clear and contextual reference to the data source.
3. **Count**: measures and expresses the enormity of records altered in each update. It provides valuable understanding of the vast scale of the data transfer operation.
4. **Start_Date**: The "Start_Date" column indicates the beginning date and time of the record update operation.
5. **End_Date**:The "End_Date" column signifies the completion date and time of the record update operation.

packagename	tablename	count	START_DATE	enddate
DW_FACT_Sales	FACT_Sales	121317	2024-1-25 12:12:58	2024-01-25 12:12:59.340
DW_Employess	DIM_Employess	1	2024-1-25 12:12:12	2024-01-25 12:12:12.900
STG_Employees	STG_Employees	19	2024-1-24 14:3:26	2024-01-25 12:12:07.793
DW_Employess	DIM_Employess	0	2024-1-25 12:9:26	2024-01-25 12:09:26.390
STG_Employees	STG_Employees	18	2024-1-24 14:3:26	2024-01-25 12:09:19.980
STG_SALES	STG_SALES	121317	2024-1-25 11:27:55	2024-01-25 11:31:49.900
MRR Fact	MRR Invoices	31465	2024-1-25 11:21:18	2024-01-25 11:21:20.280
MRR Fact	MRR InvoiceItems	121317	2024-1-25 11:21:18	2024-01-25 11:21:20.280
MRR Fact	MRR Invoices	31465	2024-1-25 11:18:18	2024-01-25 11:18:19.290
MRR Fact	MRR InvoiceItems	121317	2024-1-25 11:18:18	2024-01-25 11:18:19.290
DW_Customers	Dim_Customers	0	2024-1-25 10:6:36	2024-01-25 10:06:48.943
STG_customers	STG_customers	19119	2024-1-25 10:6:30	2024-01-25 10:06:30.733
DW_Customers	Dim_Customers	19119	2024-1-25 10:5:14	2024-01-25 10:05:14.513
STG_customers	STG_customers	19119	2024-1-25 9:52:54	2024-01-25 09:52:54.680
DW_Stores	DIM_Stores	0	2024-1-25 9:48:23	2024-01-25 09:48:23.630
STG_Stores	STG_Stores	636	2024-1-25 9:48:17	2024-01-25 09:48:18.093
STG_customers	STG_customers	19119	2024-1-25 9:48:6	2024-01-25 09:48:06.820
MRR_DIM	MRR_Branches	775	2024-1-25 9:47:26	2024-01-25 09:47:28.153
MRR_DIM	MRR_COUNTRIES	238	2024-1-25 9:47:26	2024-01-25 09:47:28.153
MRR_DIM	MRR_CUST_ADD	0	2024-1-25 9:47:26	2024-01-25 09:47:28.153
MRR_DIM	MRR_Customers	19119	2024-1-25 9:47:26	2024-01-25 09:47:28.153
MRR_DIM	MRR_DESTCODES	19614	2024-1-25 9:47:26	2024-01-25 09:47:28.153

4.3. Defining JOBS in SSIS:

To facilitate the daily refresh and loading process, a deploy was executed from SSIS to SSMS. Subsequently, a job was created to run on a daily schedule at a fixed time. This job comprises 13 steps, each representing a distinct SSIS package responsible for handling various phases of the project.

Error-handling rules have been defined to halt the process in case of an error, ensuring data integrity and reliability. Additionally, a success message is generated upon the successful completion of all 13 steps.

The entire process underwent testing by PQA, resulting in a successful validation without encountering any errors.

Step	Name	Type	On Success	On Failure
1	MRR DIM	SQL Server Integration Services Package	Go to the next step	Quit the job reporting failure
2	MRR FACT	SQL Server Integration Services Package	Go to the next step	Quit the job reporting failure
3	STG CUST	SQL Server Integration Services Package	Go to the next step	Quit the job reporting failure
4	STG EMP	SQL Server Integration Services Package	Go to the next step	Quit the job reporting failure
5	STG SRORES	SQL Server Integration Services Package	Go to the next step	Quit the job reporting failure
6	STG PRODUCTS	SQL Server Integration Services Package	Go to the next step	Quit the job reporting failure
7	STG FACT	SQL Server Integration Services Package	Go to the next step	Quit the job reporting failure
8	DIM STORES	SQL Server Integration Services Package	Go to the next step	Quit the job reporting failure
9	DIM CUST	SQL Server Integration Services Package	Go to the next step	Quit the job reporting failure
10	DIM EMP	SQL Server Integration Services Package	Go to the next step	Quit the job reporting failure
11	DIM PRODUCTS	SQL Server Integration Services Package	Go to the next step	Quit the job reporting failure
12	DIM Products History	SQL Server Integration Services Package	Go to the next step	Quit the job reporting failure
13	FACT SALES	SQL Server Integration Services Package	Quit the job reporting success	Quit the job reporting failure

It will automatically run each day at 02:00.

In addition, we've implemented a recurring job that executes this update process weekly, ensuring our development environment stays current with production.

Step	Name	Type	On Success	On Failure
1	UPDETES_Dev	SQL Server Integratio...	Quit the job reporting success	Quit the job reporting failure

It will automatically run each week on Sunday 06:00 am.

5. Power BI

In this chapter, I delve into the dynamic realm of Power BI, a pivotal component in my project's data analysis and visualization strategy. Using this tool, I have created 2 reports: customers and sales and also a dashboard. Those reports will help the senior management of the company to uncover trends, make informed decisions, and drive meaningful outcomes.

5.1. DAX Measures:

➢ In Power BI, I have created several measures:

1. Total Sales = CALCULATE(SUM(Fact_Sales[Total]))
2. YTD Sales = TOTALYTD([Total Sales], 'Dim_Date'[Date])
3. LY Sales = CALCULATE([Total Sales], SAMEPERIODLASTYEAR(datesytd(Dim_Date[Date])))
4. Sales Difference = [YTD Sales] - [LY Sales]
5. Sales Difference % = DIVIDE([YTD Sales] - [LY Sales], [LY Sales],0)
6. Average Revenue Per User = DIVIDE([Total Sales], COUNTROWS(SUMMARIZE(Fact_Sales, Fact_Sales[CustomerID])))
7. Best Sales Day = CALCULATE(MAXX(FILTER(VALUES(Dim_Date[Date]), YEAR(Dim_Date[Date]) = max(Dim_Date[YEAR])), [Total Sales]), ALLEXCEPT(Dim_Date, Dim_Date[date]))
8. Lowest Sales Day = CALCULATE(MINX(FILTER(Dim_Date,[Total Sales] <> BLANK() && YEAR(Dim_Date[Date]) = max(Dim_Date[YEAR])),[Total Sales]), ALLEXCEPT(Dim_Date, Dim_Date[Date]))
9. Total Orders = DISTINCTCOUNT(Fact_Sales[OrderID])
10. YTD Orders = TOTALYTD([Total Orders], 'Dim_Date'[Date])
11. LY Orders = CALCULATE([Total Orders], SAMEPERIODLASTYEAR(datesytd(Dim_Date[Date])))
12. Order Difference = [YTD Orders] - [LY Orders]
13. Order Difference % = DIVIDE([Order Difference], [LY Orders],0)
14. Best Sales Day Orders = CALCULATE(MAXX(FILTER(VALUES(Dim_Date[Date]), YEAR(Dim_Date[Date]) = max(Dim_Date[YEAR])), [Total Orders]), ALLEXCEPT(Dim_Date, Dim_Date[date]))
15. Lowest Sales Day Orders = CALCULATE(MINX(FILTER(Dim_Date,[Total Sales] <> BLANK() && YEAR(Dim_Date[Date]) = max(Dim_Date[YEAR])),[Total Orders]), ALLEXCEPT(Dim_Date, Dim_Date[Date]))
16. Total Units = CALCULATE (SUM(Fact_Sales[Quantity]))
17. YTD Units = TOTALYTD([Total Units], 'Dim_Date'[Date])
18. LY Units = CALCULATE([Total Units], SAMEPERIODLASTYEAR(datesytd(Dim_Date[Date])))
19. Units Difference = [YTD Units] - [LY Units]
20. Units Difference % = DIVIDE([YTD Units] - [LY Units], [LY Units],0)
21. Best Sales Day Units = CALCULATE(MAXX(FILTER(VALUES(Dim_Date[Date]), YEAR(Dim_Date[Date]) = max(Dim_Date[YEAR])), [Total Units]), ALLEXCEPT(Dim_Date, Dim_Date[date]))
22. Lowest Sales Day Units = CALCULATE(MINX(FILTER(Dim_Date,[Total Sales] <> BLANK() && YEAR(Dim_Date[Date]) = max(Dim_Date[YEAR])),[Total Units]), ALLEXCEPT(Dim_Date, Dim_Date[Date]))
23. online Sales = CALCULATE(SUM(Fact_Sales[Total]), Fact_Sales[EmpID]=77777)
24. Physical Store Sales = CALCULATE(SUM(Fact_Sales[Total]), Fact_Sales[EmpID]<>77777)
25. Total average per OrderID = AVERAGEX(KEEPFILTERS(VALUES('Fact_Sales'[OrderID])), CALCULATE(SUM('Fact_Sals'[Total])))

26. Average Basket Size = DIVIDE([Total Units], [Total Orders], 0)
27. Average Revenue Per User MoM % = VAR __PREV_MONTH = CALCULATE([Average Revenue Per User], DATEADD('Dim_Date'[Date], -1, MONTH))RETURN DIVIDE([Average Revenue Per User] - __PREV_MONTH, __PREV_MONTH)
28. Total Orders MoM % = VAR __PREV_MONTH = CALCULATE([Total Orders], DATEADD('Dim_Date'[Date], -1, MONTH))RETURN DIVIDE([Total Orders] -__PREV_MONTH, __PREV_MONTH)
29. Total Sales MoM % = VAR __PREV_MONTH = CALCULATE([Total Sales], DATEADD('Dim_Date'[Date], -1, MONTH))RETURN DIVIDE([Total Sales] -__PREV_MONTH, __PREV_MONTH)
30. Total Units MoM % = VAR __PREV_MONTH = CALCULATE([Total Units], DATEADD('Dim_Date'[Date], -1, MONTH))RETURN DIVIDE([Total Units] -__PREV_MONTH, __PREV_MONTH)
31. Number of Resold Units = CALCULATE(SUM(Fact_Sales[Quantity]),FILTER(VALUES(Fact_Sales[OrderID]), CALCULATE(COUNTROWS(Fact_Sales), Fact_Sales[OrderID] = EARLIER(Fact_Sales[OrderID])) > 1))
32. Resell Percentage = DIVIDE([Number of Resold Units], [Total Units], 0)

5.2. Sales Overview:

The dashboard presents a complete picture of the providers in the company, organized visually according to different filters, for example when selecting a specific year, the report will dynamically display key sales indicators such as total sales, total units sold, number of orders, average shopping basket size, repeat purchase percentage, etc.

The report also includes:

1. Repeat purchase percentage by continent.
2. Average shopping basket size by country.
3. The percentage of online sales VS sales in physical stores.
4. Top 5 Best-Selling Products.
5. Top 5 Best-Employees by Sales

With bookmarks, I've integrated three graphs offering a comprehensive year-to-date (YTD) overview, comparing current totals for sales, orders, and units with the same period from the previous year. Easily navigate between time periods, aggregate data by year and categories, and clear all slicers with intuitive buttons for a seamless user experience.
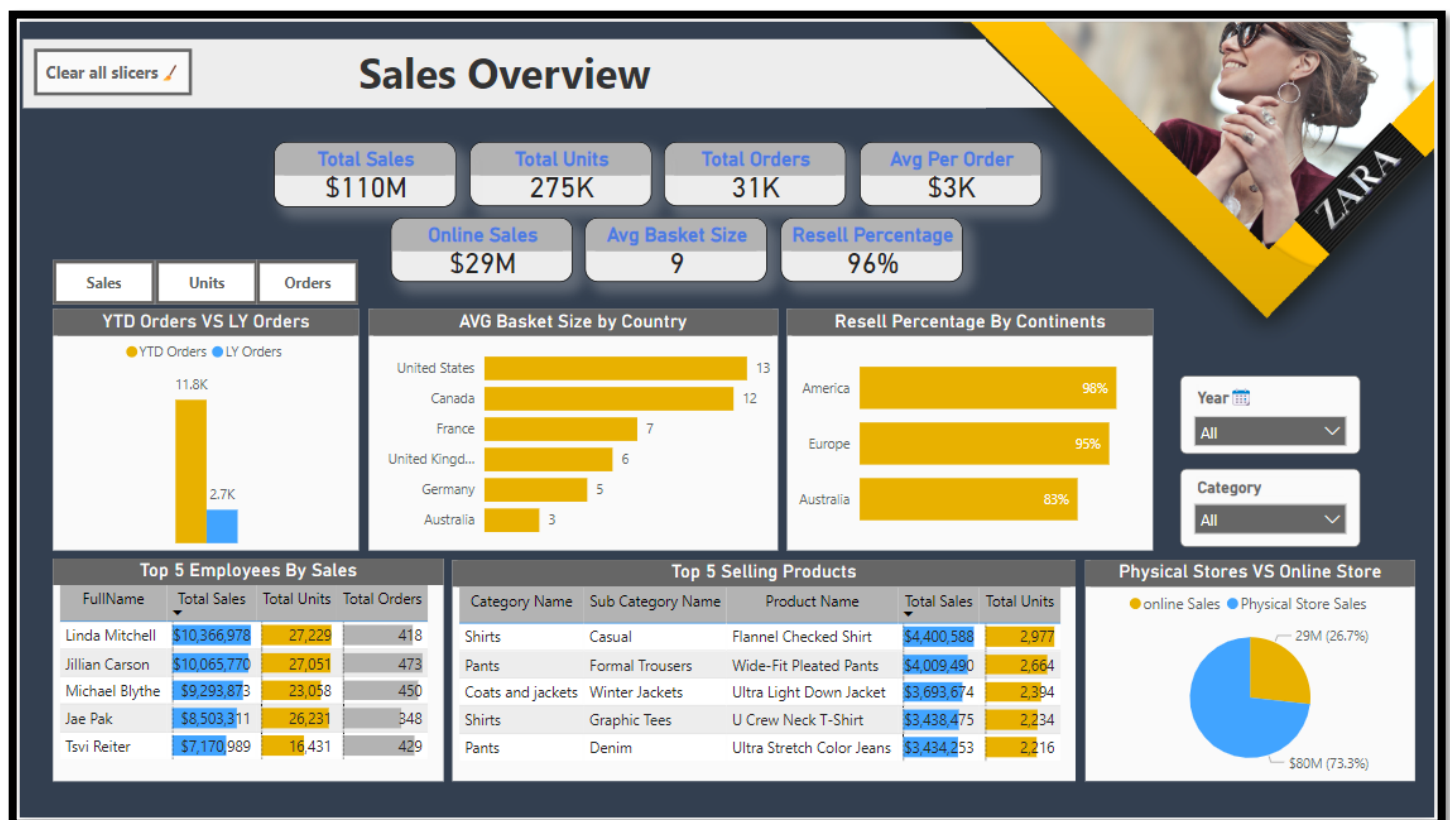
5.3. Customer Sales Analysis:

In the Customer Sales Report, comprehensive details about the company's customers are presented, organized visually through the display of country flags. When selecting a specific country, the report dynamically showcases key sales metrics, including total sales, total units, total orders, average per order, online sales, and average revenue per user.

The report also includes:

1. Detailed information on the top 5 states.
2. Total Sales breakdown by country.
3. A comparison of total sales between Physical Stores and the Online Store, categorized by country.

With bookmarks, I've incorporated four graphs that highlight monthly percentage changes in Order Count, Units Sold, Total Sales, and ARPU. Easily navigate through time periods, aggregate data by year or continent, and clear all slicers with intuitive buttons for a seamless user experience.
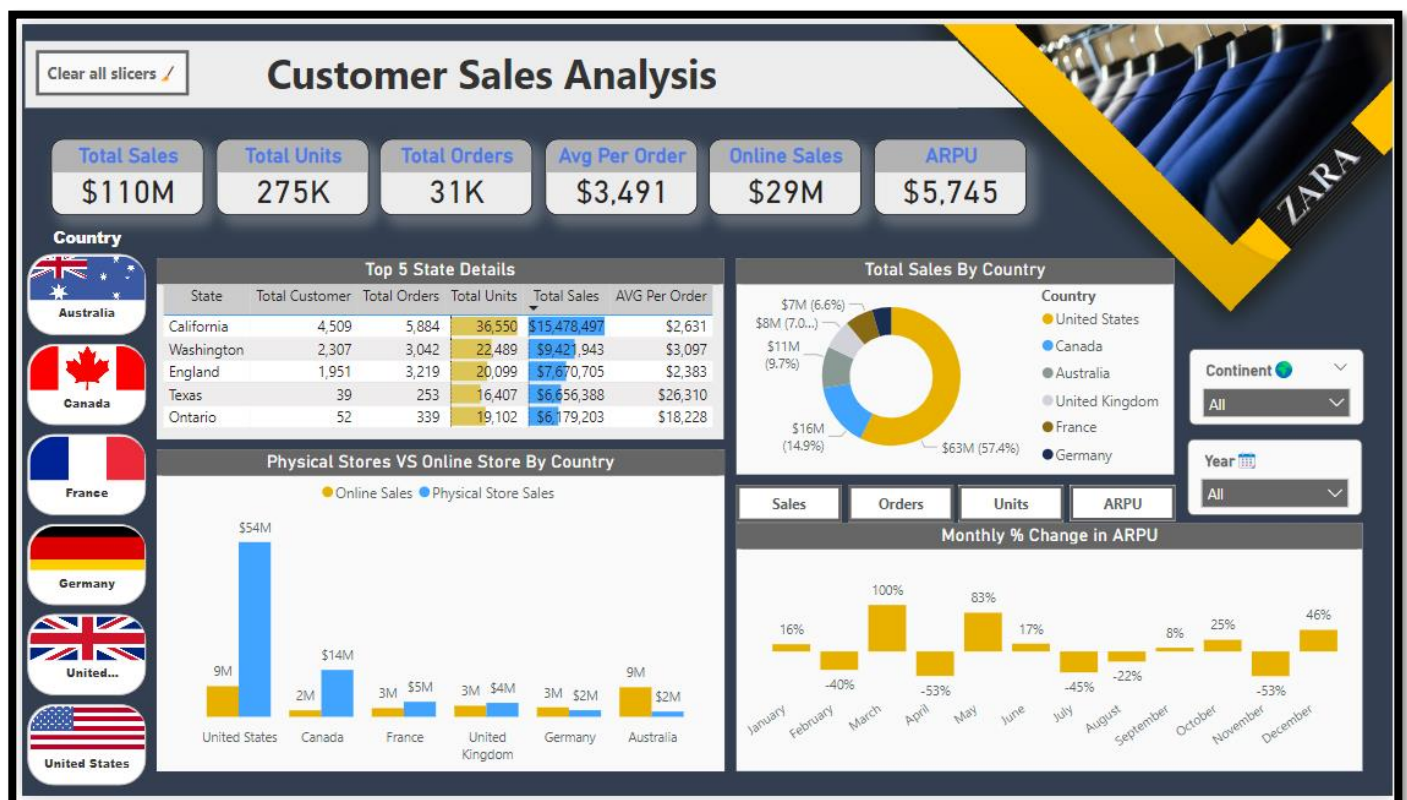
5.4. Sales Department Analysis:

In the Sales Department Report, you'll find essential insights and key performance indicators (KPIs) tailored for the sales team. The report includes:

1. General KPI Data: Highlights the Best and Lowest Sales Days along with corresponding unit metrics.
2. Employees Sales Performance: Provides employee rankings, year-to-date (YTD) total sales, a comparison to the same period last year, percentage change, KPIs, and sparklines for sales amount, units amount, and orders amount.
3. Top 5 Best-Selling Products: Identifies the top-performing products along with their quantities.
4. Top 5 Best-Employees by Sales: Showcases the top-performing employees based on sales.

The report is designed for easy navigation with buttons for filtering data by year, employee, category, and sub-category. Additionally, a 'Clear All Slicers' button ensures a seamless user experience.
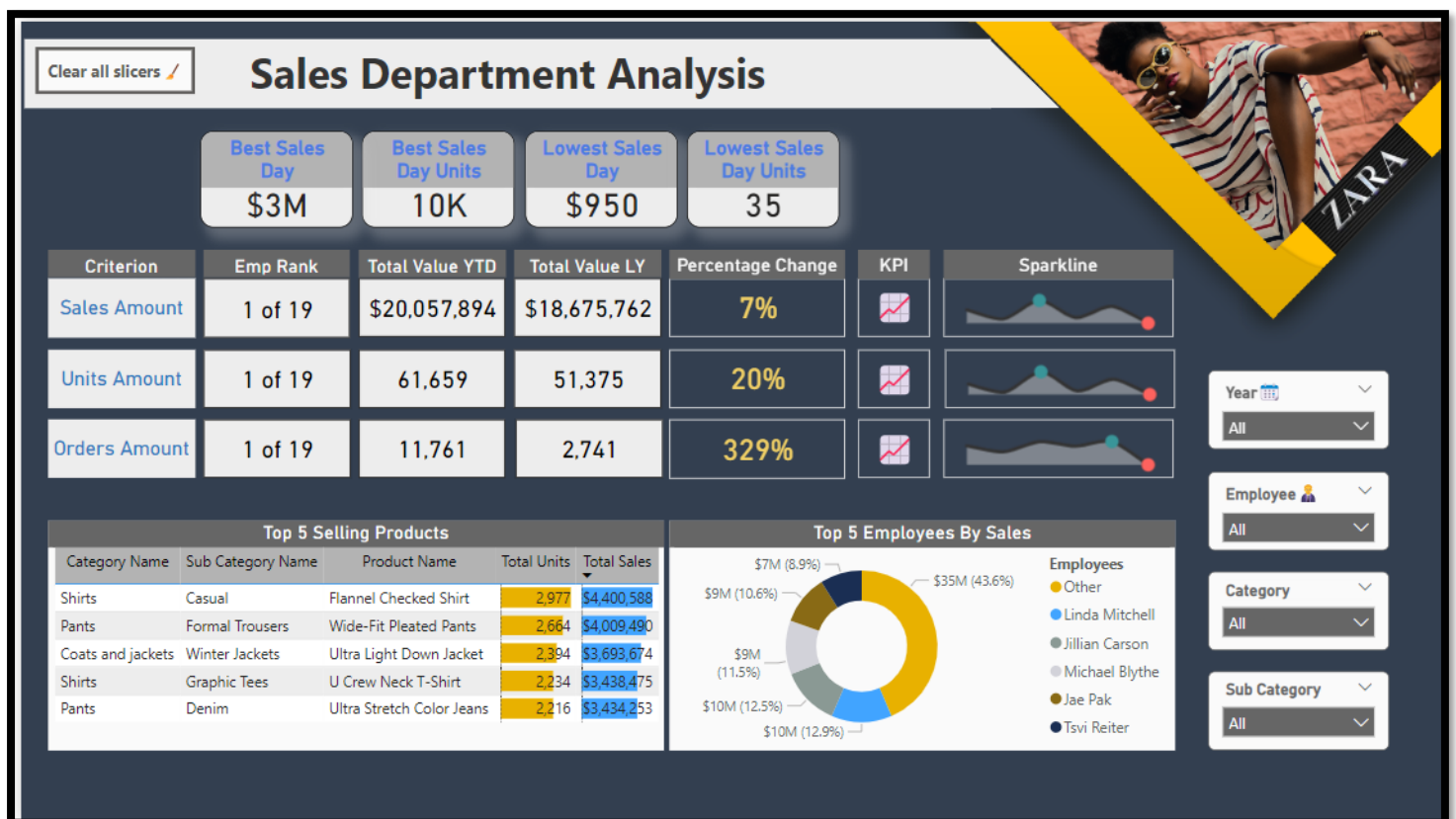
5.5. Refreshing reports and dashboards:

Similar to the data refresh procedure outlined in Chapter 4.3 for the Data Mart the settings for refreshing report and dashboard data are configured on a daily basis. This ensures complete consistency between the Data Mart and the Power BI dashboards. The information synchronization takes place through the Gateway to the ASUS-F15 computer, where the databases are stored.

The report is scheduled for daily refresh at 5:00 AM, following the completion of the data refresh procedure in the Data Mart. Alerts have also been configured to notify of any timing errors during this process.

5.6. Application:

To improve user accessibility and convenience, we've centralized all reports and dashboards into a user-friendly application named "Final Project." This consolidation ensures that users can easily navigate and access the necessary information from a single, organized platform. The Final Project application serves as a comprehensive hub for streamlined and efficient data utilization.

🗋	Name	Publisher	Published	App type
	Final Project Tal Shani Zara	tal	2/7/24, 1:12:34 PM	Org app

 Link: Final_Project_Tal_Shani_Zara

The application is accessible on mobile devices either by directly connecting to the provided link or by using the link within the POWER BI application. This flexibility allows users to access the application seamlessly, ensuring a convenient and responsive experience whether they prefer accessing it directly through a web link or through the dedicated POWER BI mobile application.