

VIETNAM NATIONAL UNIVERSITY HOCHIMINH CITY
UNIVERSITY OF ECONOMICS AND LAW



BUSINESS INTELLIGENCE
AND DECISION SUPPORT SYSTEM

Topic:

**APPLYING BUSINESS INTELLIGENCE TECHNOLOGY AND
AUTOMATION PROCESSES TO MAKE MORE ACCURATE
DECISIONS FOR SALES OPERATIONS**

Lecturer: Nguyen Van Ho, M.S

Le Ba Thien, M.S

Group: BoKho

Ho Chi Minh, May 20

MEMBER OF GROUP BO KHO

| No. | Student ID | Class | Full Name | Role |
|-----|------------|---------|------------------------|--------|
| 1. | K204061440 | K20406T | Tran Nhat Nguyen | Leader |
| 2. | K204061446 | K20406C | Man Dac Sang | Member |
| 3. | K204061411 | K20406T | Nguyen Thai Ngoc Suong | Member |
| 4. | K204060310 | K20406C | Thai Thien Truc | Member |

ACKNOWLEDGEMENTS

During the limited time we had, we put our best effort to create a Business Intelligence solution that met all of the necessary requirements. Despite the challenges we faced, we were able to overcome them thanks to the assistance and enthusiastic guidance provided by M.S. Nguyen Van Ho. His contributions played a critical role in ensuring the success of this project. We would also like to express our deep appreciation to M.S. Le Ba Thien for his vast knowledge and experience in his career path. His teachings on PowerBI were incredibly helpful, and we are grateful for the time he took to assist us.

Throughout the implementation of this project, we received valuable advice and ideas from our teachers, which greatly aided us in overcoming any stumbling blocks. We are truly grateful for their support and guidance.

We hope that this topic will generate meaningful discussion and provide knowledge. It is our sincere desire that this project will continue to develop, and we recognize that the assessments from our teachers will be crucial in helping us achieve and improve this goal.

Ultimately, we are committed to using the feedback we receive to become better and to continue improving the quality of our work. We are cheerful for this opportunity and look forward to developing our skills and knowledge about building Business Intelligence solutions.

COMMITMENT

To ensure that we produce the highest quality work possible, we will be using a variety of reference sources, including books and the internet.

In order to maintain the highest standards of academic integrity, all citations and references will be provided in a clear and concise manner. We recognize that it is essential to properly acknowledge any sources we utilize throughout our work.

We believe that by maintaining a strong sense of discipline and adhering to rigorous academic standards, we will be able to achieve the best possible outcome for this project.

Table of Contents

| | |
|--|-------------|
| MEMBER OF GROUP BO KHO | I |
| ACKNOWLEDGEMENTS..... | II |
| COMMITMENT | III |
| LIST OF FIGURES | VIII |
| LIST OF TABLES | X |
| LIST OF ACRONYMS..... | XI |
| CHAPTER 1: INTRODUCTION | 1 |
| 1.1. Business case for the project | 1 |
| 1.2. Objectives of the project | 1 |
| 1.2.1. General Objective..... | 1 |
| 1.2.2 Specific Objectives..... | 2 |
| 1.3. Research Objects | 2 |
| 1.4. Scope of the project | 2 |
| 1.5 Value and desired outcome of the project..... | 3 |
| 1.6 Structure of project..... | 3 |
| CHAPTER 2: THEORETICAL BASIS..... | 4 |
| 2.1 Overview about BI | 4 |
| 2.1.1 What is BI? | 4 |
| 2.1.2 BI Architecture | 4 |
| 2.1.3 Advantage of BI in enterprises | 5 |
| 2.1.4 BI Strategy for Business..... | 6 |
| 2.2 ETL Process..... | 7 |
| 2.2.1 What is ETL?..... | 7 |
| 2.2.2 Why do we need ETL? | 7 |
| 2.2.3 ETL Process..... | 7 |

| | | |
|---|--|-----------|
| 2.3 | Data warehouse and Data mart..... | 8 |
| 2.3.1 | What are Data warehouse and Data mart? | 8 |
| 2.3.2 | Who needs Data warehouse and Data mart? | 9 |
| 2.3.3 | Advantages and disadvantages of Data warehouse | 9 |
| 2.3.4 | Snowflake and Star schemas | 9 |
| 2.4 | KPIs..... | 10 |
| 2.4.1 | KPIs Definition..... | 10 |
| 2.4.2 | The advantages and disadvantages of KPIs..... | 10 |
| 2.4.3 | Categories of KPIs..... | 10 |
| 2.5 | MDX language for analyzing multidimensional data and OLAP | 11 |
| 2.6 | Microsoft Azure..... | 12 |
| 2.7 | Power Automate | 13 |
| CHAPTER 3: REQUIREMENTS ANALYTICS AND INTRODUCTION TO BI SOLUTION..... | | 14 |
| 3.1. | Sales Business processes..... | 14 |
| 3.1.1 | Sales department..... | 14 |
| 3.1.2 | The purpose of Sales | 14 |
| 3.1.3 | Sales process..... | 15 |
| 3.2 | Data source and challenges: | 16 |
| 3.2.1 | Data source | 16 |
| 3.2.2 | Challenges | 17 |
| 3.3 | Business Requirements Analysis..... | 18 |
| 3.3.1 | How is the sales performance of the business..... | 18 |
| 3.3.2 | Identify the customer segments of the business | 18 |
| 3.3.3 | What is the tendency of customers' purchasing behaviors | 18 |
| 3.3.4 | Research objectives | 18 |

| | |
|---|-----------|
| 3.4 IT Requirements Analysis | 21 |
| 3.5 Comparative Analysis of BI and Data Visualization Tools..... | 22 |
| 3.5.1 Microsoft Power BI..... | 22 |
| 3.5.2 Tableau | 23 |
| 3.6 Proposal BI solution | 23 |
| CHAPTER 4: BUILDING DATA WAREHOUSE AND INTEGRATING DATA | 25 |
| 4.1 Designing Data Warehouse | 25 |
| 4.1.1 Bus Matrix..... | 25 |
| 4.1.2 Master Data | 26 |
| 4.1.3 Transaction Data | 26 |
| 4.1.4 ETL mapping..... | 27 |
| 4.1.5 Fact and dimension tables | 36 |
| 4.1.6 Data Warehouse model | 39 |
| 4.2. ETL processes | 41 |
| 4.2.1 ETL pipeline..... | 41 |
| 4.2.2 Staging Area's ETL Process | 42 |
| 4.2.3 Dimension Table's ETL Process..... | 43 |
| 4.2.4 Fact Table's ETL Process | 47 |
| 4.2.5 Truncate Table in Staging Area | 52 |
| CHAPTER 5: RESULTS – DATA ANALYTICS AND VISUALIZATION..... | 54 |
| 5.1 Data analytics with SSAS technology | 54 |
| 5.2 Report and dashboard systems | 56 |
| 5.3 Incremental refresh using Power Automate..... | 57 |
| 5.4 Data analysis with Power BI | 58 |
| 5.4.1 RFM model for customer segmentation..... | 58 |

| | |
|--|-----------|
| 5.4.2 Delivery performance..... | 60 |
| 5.4.3 Sales performance | 63 |
| 5.4.4 Product sales analysis..... | 66 |
| 5.4 Recommend business strategies | 68 |
| 5.4.1 Marketing strategies for 3 main segments | 68 |
| 5.4.2 Optimize delivery performance..... | 68 |
| 5.4.3 Enhance sales performance | 68 |
| 5.4.4 <i>Product-based strategy</i> | 69 |
| CHAPTER 6: CONCLUSION AND FUTURE WORKS | 70 |
| 6.1. Results..... | 70 |
| 6.2. Limitations | 70 |
| 6.3. Future works..... | 70 |
| REFERENCES | 71 |
| APPENDIX | 73 |
| TASK ASSIGNMENT TABLE..... | 74 |
| TURNITIN CHECKING..... | 75 |

LIST OF FIGURES

| | |
|---|----|
| Figure 2- 1 Benefits of BI followed by a survey of 2600 users by BI-Survey.com [3]... | 5 |
| Figure 2- 2 ETL process [8] | 8 |
| Figure 3- 1 Sales Module | 17 |
| Figure 3- 2 BI Solution (Source: Author's proposal)..... | 23 |
| Figure 4- 1 DimCustomer table..... | 36 |
| Figure 4- 2 DimEmployee table | 36 |
| Figure 4- 3 DimProduct table..... | 37 |
| Figure 4- 4 DimStore table..... | 37 |
| Figure 4- 5 DimGeography table | 37 |
| Figure 4- 6 DimShipMethod table | 38 |
| Figure 4- 7 DimChannel table..... | 38 |
| Figure 4- 8 DimDate table..... | 38 |
| Figure 4- 9 FactSales table | 39 |
| Figure 4- 10 Data Warehouse Star Schema | 40 |
| Figure 4- 11 ETL pipeline | 41 |
| Figure 4- 12 Load data from Source to Staging Area | 42 |
| Figure 4- 13 Load data from Staging Area from Dimension Tables | 43 |
| Figure 4- 14 Load data to DimCustomer | 44 |
| Figure 4- 15 Load data to DimEmployee..... | 44 |
| Figure 4- 16 Load data to DimProduct..... | 45 |
| Figure 4- 17 Load data to DimGeography | 45 |
| Figure 4- 18 Load data to DimStore..... | 46 |
| Figure 4- 19 Load data to DimShipMethod | 46 |
| Figure 4- 20 Load data to DimChannel..... | 47 |
| Figure 4- 21 Create DimDate | 47 |
| Figure 4- 22 Load data from Staging Area to Fact Table | 47 |
| Figure 4- 23 Lookup DimCustomer & DimEmployee | 48 |

| | |
|---|----|
| Figure 4- 24 Lookup DimProduct & DimShipMethod | 49 |
| Figure 4- 25 Lookup DimChannel & DimGeography | 50 |
| Figure 4- 26 Lookup DimStore & DimDate | 51 |
| Figure 4- 27 Lookup FactSales and SCD FactSales_Updated..... | 52 |
| Figure 4- 28 Truncate all tables in Staging Area | 52 |
| Figure 4- 29 Schedule jobs configure the step truncate | 53 |
| Figure 4- 30 Job schedule properties configure the job Truncate..... | 53 |
| Figure 5- 1 Data warehouse schema (SSAS) | 54 |
| Figure 5- 2 The flow of automated refresh a dataset | 57 |
| Figure 5- 3 General RFM model for customer segmentation | 58 |
| Figure 5- 4 Left side of general RFM model for customer segmentation..... | 59 |
| Figure 5- 5 Deliver performance dashboard - Germany (2013) | 60 |
| Figure 5- 6 Total order value by channel (delivery methods) - Germany (2013)..... | 61 |
| Figure 5- 7 Trend between number of orders and AVG Freight - Germany (2013)..... | 62 |
| Figure 5- 8 Fluctuating trend of weekly deliveries - Germany (2013) | 62 |
| Figure 5- 9 AVG Freight by Region among years - Germany (2013)..... | 62 |
| Figure 5- 10 General sales performance | 63 |
| Figure 5- 11 The top customer had high sales with online channel..... | 65 |
| Figure 5- 12 The top store had high sales with offline channel | 65 |
| Figure 5- 13 General product sales analysis | 66 |
| Figure 5- 14 Filter of General product sales analysis | 67 |
| Figure 5- 15 ABC class of dashboard Product Sales Insight | 67 |
| Figure o- 1 Result of Turnitin Checking with 4% match overview | 75 |

LIST OF TABLES

| | |
|---|----|
| Table 4- 1 Bus matrix..... | 25 |
| Table 4- 2 Master data..... | 26 |
| Table 4- 3 Transaction da..... | 26 |
| Table 4- 4 ETL mapping | 35 |
| Table 4- 5 Staging Area's table structure..... | 42 |
| Table 4- 6 Data warehouse's dimension table structure | 43 |
| Table 4- 7 Data warehouse's facttable structure | 48 |
| Table 5- 1 The detail of customer segmentation..... | 59 |
| Table o- 1 Task Assignment..... | 74 |

LIST OF ACRONYMS

| Abbreviation | Explanation |
|--------------|---------------------------------|
| BI | Business Intelligence |
| DW | Data Warehouse |
| ETL | Extract Transform Load |
| OLTP | Online Transaction Processing |
| OLAP | Online Analysis Processing |
| SQL | Structured Query Language |
| SSIS | SQL Server Integration Services |
| SSAS | SQL Server Analysis Services |
| MDX | Multi-Dimensional Expression |
| RFM | Recenvy – Frequency – Monetary |
| KPI | Key Performance Indicator |

CHAPTER 1: INTRODUCTION

1.1. Business case for the project

Decision-making refers to using data and analytical tools to make informed decisions that support business goals and objectives. It enables organizations to make data-driven decisions that can improve operational efficiency, increase profitability, and gain a competitive advantage in the market. This project focuses on building Business Intelligence for Sales modules to get a better understanding of the features of this department that enhance decision-making for managers.

Sales performance: Analyzing sales performance can help businesses understand their revenue streams, sales trends, and customer behavior. By knowing which products or services are performing well, businesses can allocate their resources and efforts more effectively to maximize profits.

Customer segmentation: Customer segmentation involves grouping customers based on their shared characteristics such as demographics, behavior, and needs. By understanding customer segments, businesses can tailor their products, services, and marketing strategies to meet the specific needs of each segment.

Delivery performance: Delivery performance refers to how well a business delivers its products or services to customers. By monitoring delivery performance, businesses can ensure that customers receive their orders on time and in good condition.

1.2. Objectives of the project

1.2.1. General Objective

The general objective is to suggest a business intelligence that helps to analyze data and enhance the decision-making process.

- Identify the customer segments of the business.
- How is the sales performance of the business?
- What is the tendency of customers' purchasing behaviors?

1.2.2 Specific Objectives

Based on the general objective, the specific objectives are identified as follows:

1. Identify customer segments using RFM model.
2. Evaluate sales performance by analyzing delivery performance in different regions
3. Acknowledge sales performance across online and offline channels.
4. Discover sales performance by regions
5. Analyze customer trends in purchasing behavior for specific products using cross-selling
6. Categorize products items based on their importance by ABC classification.

1.3. Research Objects

The AdventureWorks dataset is a sample database that was originally created by Microsoft to demonstrate the functionality of its SQL Server database software. It simulates a bicycle company and contains data on sales, customers, products, and employees. It contains a wide range of tables, views, stored procedures, and other database objects that represent different aspects of the business. We will utilize the most recent version of the database as of 2021 that is AdventureWorks2019.

1.4. Scope of the project

The Sales and Human Resources modules in the AdventureWorks dataset contain data related to sales and salesperson information. The Sales module includes tables such as SalesOrderHeader, SalesOrderDetail, Customer, Store, SalesPerson, and SalesTerritory. The Human Resources module includes tables such as Employee, Person, Address, Department, and EmployeeDepartmentHistory.

Together, these modules provide a comprehensive view of sales and salesperson data within an organization. They can be used for various analytical purposes such as identifying areas for improvement in the sales process and personnel performance.

1.5 Value and desired outcome of the project

Business Intelligence tools enable organizations to collect and analyze data from multiple sources, enabling informed decision-making. By identifying patterns, trends, and correlations, organizations can make strategic decisions. These tools also help optimize operations by providing insights into delivery performance, customer segmentation, and personnel performance.

Our project goal is to create a comprehensive Business Intelligence solution that covers data collection, storage, processing, analysis, and presentation. It's important to identify user requirements for data, reports, and dashboards, and maintain data accuracy throughout the process. Automating the process from data source to end-user saves time, reduces errors, and increases productivity.

1.6 Structure of project

Chapter 1: Introduction: This chapter covers the introduction of the project including objectives with the scope, desired outcome, and structure of the project.

Chapter 2: Theoretical basis: This chapter will provide all theoretical background about the project such as knowledge about Business Intelligence, ETL Processes, Data warehouse, Data mart, KPIs, Microsoft Azure, and Automate tools.

Chapter 3: Requirements Analytics and Introduction to BI Solution: This chapter includes all of the requirements of the project from general ones to specific ones. It also reveals challenges when doing this project.

Chapter 4: Building Data Warehouse and integrating data: This chapter will tell about the steps involved in creating a data warehouse that meets the requirements given above. It also describes clearly the ETL process.

Chapter 5: Results - Data Analytics and Visualization: This chapter provides all the results after implementation. Dashboards are included in this chapter which will give insights or useful information about the business.

Chapter 6: Conclusion and future works: This chapter covers the results and limitations when conducting this project and proposes future works to improve this project.

CHAPTER 2: THEORETICAL BASIS

2.1 Overview about BI

In general, BI gives businesses a method to understand their operations better and to make decisions that are better supported by data. Businesses may boost productivity, become more competitive, and stimulate development by utilizing BI.

2.1.1 *What is BI?*

Business intelligence (BI) is a collection of techniques and tools that businesses employ to analyze corporate data and turn it into insights that can be used to make strategic and tactical business decisions [1]. In order to give users in-depth knowledge about the condition of the business, BI tools access and analyze data sets and show analytical results in reports, summaries, dashboards, graphs, charts, and maps.

The phrase "business intelligence" is frequently used to describe a variety of technologies that offer rapid, simple-to-understand access to insights about the present status of a company based on accessible data.

2.1.2 *BI Architecture*

BI architecture is the technology infrastructure that supports BI applications and tools [2]. A typical BI architecture consists of several layers:

1. Data sources: This layer consists of all of the many data sources, including databases, spreadsheets, and other applications, that are utilized to provide BI insights.

2. Data integration is the second layer, which entails combining data from several sources into one repository or data warehouse. Processes like data extraction, transformation, and loading (ETL) are frequently used for this.

3. Data storage: This layer entails keeping the combined data in a manner that BI tools can quickly access and evaluate. This can entail utilizing a specialist data warehousing platform or a conventional relational database management system.

4. BI tools: The programs and tools used to analyze and display the data, including dashboards, data mining are included in this layer.

5. Presentation layer: This layer entails providing end users with insights produced by BI technologies in an easy-to-understand and actionable way.

6. Security layer: This layer makes sure that data is safe and secure and that only authorized people have access to sensitive data.

2.1.3 Advantage of BI in enterprises

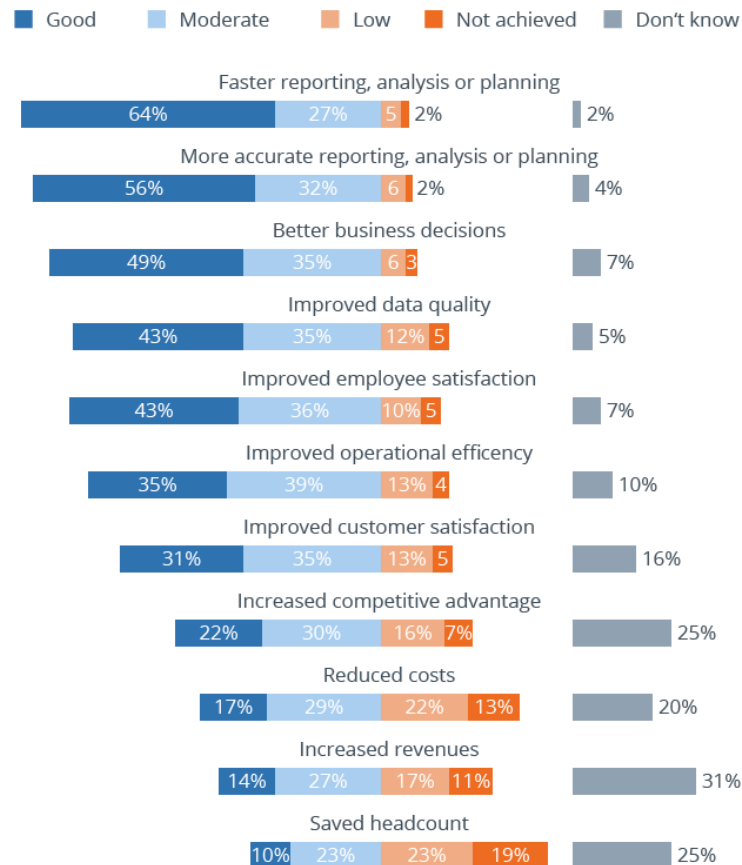


Figure 2- 1 Benefits of BI followed by a survey of 2600 users by BI-Survey.com [3]

According to a survey of 2600 business intelligence users of Figure 2- 1 above, the following are the top five benefits:

- Faster reporting, analysis or planning
- More accurate reporting, analysis or planning
- Better business decision
- Improved data quality
- Improved employee satisfaction

2.1.4 BI Strategy for Business

The process of deploying a BI system in your firm is referred to as a business intelligence strategy [4]. This entails identifying the key players, analyzing the issue, setting goals, and identifying the KPIs that will track your progress toward achieving those goals.

The implementation of business intelligence is a difficult undertaking since it involves a lot of upfront planning, a large cast of characters, and costs. However, the benefits far outweigh the costs, and it is well known that business intelligence ROI is real even if it is occasionally difficult to measure. In the long run, especially, the consequences of not adopting it are more detrimental.

Step 1: Identify your organization's goals

What does your organization want to achieve with Business Intelligence? Once these goals are identified, it is important to understand the data needed to achieve them. This data can come from either internal or external third-party sources.

Step 2: Select the right BI solution

There is a wide range of BI tools available on the market, so it's important to choose the ones that will best meet your organization's needs. Some must-have features to look out for:

Interactive data visualization – This feature allows users to quickly and easily see patterns, trends, and outliers in data sets to make data-driven decisions at the moment of impact.

Reporting – This feature allows users to generate reports that can be used to track progress, measure success, and identify areas for improvement.

Data mining – This feature allows users to extract valuable information from large data sets so they can uncover hidden trends and insights.

Predictive analytics – This feature helps users to make predictions about future events based on past data so they can make informed planning decisions.

Other criteria to look for in a BI solution:

Collaboration – Can users easily share data, reports, and insights with others?

Data security – How is the vendor keeping your data safe and secure so that unauthorized users can't access or misuse it?

Flexibility – How easily can you customize views or processes in the solution to meet your specific needs?

Step 3: Turn your data into insights

After collecting data, it's time to start creating insights with your BI solution.

Step 4: Put your data insights into action

2.2 ETL Process

2.2.1 What is ETL?

ETL stands for Extract, Transform, Load, which is the process of extracting data from a source, transforming the data to ensure its independence and efficiency, and then loading the transformed data into a target database [5]. ETL is an important process in integrating data from various sources to create a comprehensive database and is widely used in data analysis and reporting projects. ETL technology simplifies the data integration process and improves the efficiency of data analysts and data managers.

2.2.2 Why do we need ETL?

ETL is essential for improving data quality and reducing errors in decision-making [6]. It achieves this by removing inaccurate, duplicate, or inconsistent data and converting data into formats that can be easily analyzed and reported on.

Implementing ETL provides several benefits to organizations, including improved data integrity and consistency, enhanced system performance, support for data analysis, and increased data accuracy.

2.2.3 ETL Process

ETL (Extract, Transform, Load) is the process of extracting data from various sources, transforming it to meet predefined rules, and loading it into a destination database [7].

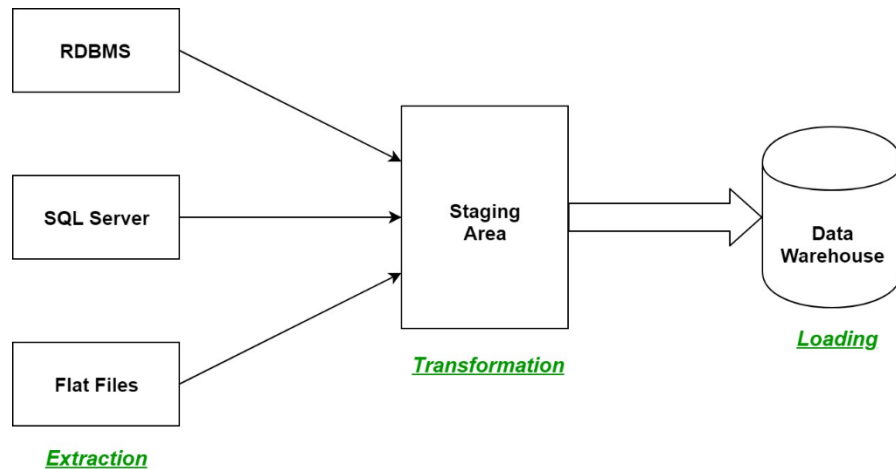


Figure 2- 2 ETL process [8]

The ETL process consists of three basic stages [7] in Figure 2- 2:

- The ETL process consists of three basic stages: extract, transform, and load. In the extract stage, data is extracted from different sources such as databases, data files, CRM systems, or ERP systems. This data may be in different formats and may need to be cleaned or filtered before it can be used in the ETL process.
- The transform stage involves converting the extracted data into compatible formats for the destination database, while removing inaccurate, duplicate, or inconsistent data. This stage may also involve data mapping, data validation, and data enrichment.
- The final stage of the ETL process is the load stage, where the transformed data is loaded into the destination database. This stage involves inserting the data into the appropriate tables and ensuring that the data is properly indexed and optimized for querying.

2.3 Data warehouse and Data mart

2.3.1 What are Data warehouse and Data mart?

Data warehouse: Stores current and historical data from one or more systems in a predefined and fixed schema, which allows business analysts and data scientists to easily analyze the data [9].

Data mart: is a subset of a data warehouse, typically consisting of a single subject area (e.g., marketing, operations) [9].

2.3.2 *Who needs Data warehouse and Data mart?*

Organizations that require efficient management and analysis of massive data volumes can derive significant advantages from utilizing data warehouses. Data warehouses are versatile and can be employed in various sectors, including but not limited to retail, healthcare, finance, and manufacturing. Furthermore, employees who need to access and analyze specific subsets of data within an organization can benefit from data marts, which are a smaller subset of data warehouses. These data marts are often utilized in specific departments such as sales, marketing, finance, and operations, among others.

2.3.3 *Advantages and disadvantages of Data warehouse*

Advantages: Firstly, it facilitates decision-makers' access to the necessary data for informed decision-making and analysis. Secondly, it is capable of handling vast amounts of data and complicated queries, resulting in fast and efficient data analysis. Thirdly, it consolidates data from multiple sources into a single source. Lastly, it stores historical data, enabling users to analyze past trends and patterns and make future predictions.

Disadvantages: Setting up and maintaining data warehouses can be costly as it involves investments in hardware, software, and human resources. It can also be complicated to design and implement, requiring specialized skills and expertise. Building a data warehouse can be a time-consuming process that may take years to complete. Additionally, ensuring data quality can be challenging due to inconsistencies in data from multiple sources. Since data warehouses contain confidential data, security is a crucial concern, and implementing appropriate security measures can be both difficult and time-consuming.

2.3.4 *Snowflake and Star schemas*

The snowflake schema [9] is a type of organization for tables in a multidimensional database that creates a shape resembling a snowflake in its entity-relationship diagram.

The star schema [9], which is also known as the star join schema, is the most popular and straightforward approach to dimensional modeling.

2.4 KPIs

2.4.1 KPIs Definition

A KPI [9], or key performance indicator, is a metric that assesses progress toward achieving a strategic objective by measuring performance against a goal.

2.4.2 The advantages and disadvantages of KPIs

Advantages of KPIs: Key performance indicators (KPIs) provide measurable outcomes that aid in gauging progress and pinpointing areas that require improvement. They align objectives and goals with strategies, ensuring that all stakeholders are working towards the same objectives. Additionally, KPIs offer data that enable managers to make informed decisions about resource allocation and performance management. Moreover, KPIs help pinpoint areas for enhancement and encourage efforts for continuous improvement.

Disadvantages of KPIs: The misuse or misinterpretation of KPIs can result in incorrect comprehension and actions. KPIs are only effective when high-quality data is used to measure them; otherwise, the results may be incorrect. KPIs may not be flexible and may not adjust well to changes in the business environment or unforeseen events. Concentrating excessively on particular KPIs can lead to unintended consequences, such as employees engaging in unethical activities to achieve KPI targets.

2.4.3 Categories of KPIs

Sales KPI to track [10]:

Total Sales: refers to the total amount of revenue generated by a business or organization over a specific period of time, typically a day, week, month, quarter, or year.

Total Profit: refers to the amount of money a business or organization has earned after deducting all of its expenses from its total revenue.

Sales YTD: stands for "sales year-to-date" and refers to the total sales revenue generated by a business or organization from the beginning of the current calendar year up to the current date.

Cumulative sales: refer to the total sales revenue generated by a business or organization over a specific period of time, which is the sum of all sales revenue up to a certain point in time.

Sales growth: refers to the rate at which a business or organization's sales revenue is increasing or decreasing over a specific period of time, typically a year.

Average purchase value: refers to the average amount of money a customer spends per transaction when purchasing goods or services from a business or organization.

Product performance: refers to how well a product is selling or performing in the market, based on various metrics such as sales revenue, and market share.

2.5 MDX language for analyzing multidimensional data and OLAP

MDX [11], or Multidimensional Expressions, is a language used in OLAP (Online Analytical Processing) systems to query and work with multidimensional data. It is an effective tool that gives users access to and multidimensional data analysis capabilities, allowing for a deeper comprehension of corporate performance.

Although MDX was created specifically for OLAP systems, which store data in multidimensional cubes, it is based on SQL (Structured Query Language). Users of MDX may construct intricate queries that delve deeply into particular cube dimensions like time, location, product, or customer. Additionally, a wide range of operators and functions are available for data manipulation, including aggregation, filtering, sorting, and formatting.

MDX is used by business analysts, data scientists, and other professionals who need to analyze multidimensional data in real-time. It is often used in conjunction with data visualization tools, such as dashboards and reports, to provide insights into key performance indicators (KPIs) and other metrics.

There are several common functions and operators used in MDX:

1. Aggregate functions: used to calculate aggregate values for a data set including SUM, MIN, MAX, and COUNT.
2. Filter functions: used to extract a subset of data from a larger set based on specific criteria including WHERE, FILTER, and EXISTS.
3. Aggregation functions: used to create and manipulate data sets including UNION, INTERSECT, and EXCEPT.
4. Time function: used to navigate time-based dimensions in a cube including PERIODS-TO-DATE, YEAR-TO-DATE, and PREVIOUSMBER.
5. Math operators: used to perform mathematical operations on data including +, -, *, and /.
6. Comparison operator: used to compare two values to determine if they are equal, greater than, or less than each other, including =, <, >, and <=.
7. Logical operators: used to combine multiple expressions to create more complex criteria including AND, OR, and NOT.

By using this combination of functions and operators, users can create complex MDX queries for multidimensional data analysis in OLAP systems.

2.6 Microsoft Azure

Azure is a powerful and flexible platform for developing, deploying, and managing cloud applications and services [12]. It provides a variety of services for developing, deploying, and managing applications and services via Microsoft-managed data centers, including virtual machines, networking, storage, internet of things (IoT), and more.

Azure Analysis Services is a fully managed platform as a service (PaaS) that offers cloud-based enterprise-grade data modeling. One of the primary advantages of Azure Analysis Services is its capacity to manage enormous amounts of data. It can handle billions of rows of data and scale up and down as needed. Furthermore, it employs innovative compression techniques to reduce storage costs and increase query performance.

Azure Analysis Services also integrates with other Azure services, such as Azure Data Factory, Azure Stream Analytics, and Azure Event Hubs, to provide real-time data analysis capabilities.

Some key features of Azure Analysis Services:

- Tabular modeling: enables create data models using tables and relationships.
- Integration with other Azure services: Power BI, Azure Databricks, and Azure Machine Learning.
- Real-time data analysis: Azure Stream Analytics and Azure Event Hubs, to provide real-time data analysis capabilities.
- Advanced compression: minimize storage costs and improve query performance.
- User-friendly interface: provides a user-friendly interface for creating and managing data models.

2.7 Power Automate

Power Automate is one of Microsoft's powerful business applications, formerly known as Microsoft Flow [13]. It allows users to automate manual processes such as data entry, reporting, invoicing, etc. to help businesses optimize operations. Power Automate provides different triggers and action models, enabling users to easily customize their own automated processes and integrate different applications and services.

The benefits of Power Automate include improved efficiency, increased productivity, reduced costs, and improved data quality. In addition, Power Automate offers advanced features such as the integration of AI models into processes, RPA automation, and AI generators.

In Power Automate, users can set up 3 types of processes:

- Automated Workflows: This type of workflow is triggered based on some other action.
- Scheduled Workflows: These workflows are scheduled to execute at specific times of the day, week, or month.
- Button Workflows: These processes are activated by pressing a button.

CHAPTER 3: REQUIREMENTS ANALYTICS AND INTRODUCTION TO BI SOLUTION

3.1. Sales Business processes

3.1.1 Sales department

The Sales Department is a crucial division of any organization that is responsible for generating revenue through the sale of goods or services. Its main objective is to promote and sell the company's products or services to potential customers.

The Sales Department is responsible for a range of functions, including:

- Developing sales strategies and plans
- Identifying new sales opportunities
- Building and maintaining relationships with customers
- Conducting market research to understand customer needs and preferences
- Providing customer service and support
- Negotiating deals and contracts with customers
- Meeting sales targets and goals
- Providing feedback to other departments to improve products or services

3.1.2 The purpose of Sales

Business Growth

The purpose of sales in business is to generate revenue and drive growth. Sales are the lifeblood of any business, as they bring in the money that allows the company to operate, invest in new products or services, and expand into new markets. By selling more products or services, a company can increase its revenues and profits, which can then be reinvested back into the business to fuel further growth.

Customer Retention

Sales also play an important role in customer retention. While the primary goal of sales is to generate revenue, it is equally important to ensure that customers are satisfied with their purchases and continue to do business with the company over the long term.

Effective sales practices can help to build strong relationships with customers by providing them with personalized attention, demonstrating a deep understanding of their needs and preferences, and offering ongoing support and service. By building trust and rapport with customers, sales teams can create a sense of loyalty and encourage repeat business.

Generate new leads

Another important purpose of sales is to generate new leads for the business. Lead generation refers to the process of identifying and cultivating potential customers who have shown some level of interest in the company's products or services.

It may attract new leads by creating awareness of the company's offerings, demonstrating the value of its products or services, and building trust and credibility with potential customers. This can involve a range of tactics, such as targeted marketing campaigns, attending trade shows and events, and leveraging social media and other digital channels to reach potential customers.

3.1.3 Sales process

The sales process typically involves several steps, which may vary depending on the industry, company, or product/service being sold. However, the following are some common steps in the sales process [14]:

1. **Prospecting:** This involves identifying and qualifying potential customers who may be interested in the product or service. You may identify your ideal customer and screen them based on many factors such as their monthly salary... This is to narrow your customer pool and try to convert them into current customers.
2. **Pre-approach:** This step involves gathering information about the prospect to understand customer's needs, preferences, and pain points.
3. **Approach:** This step involves making initial contact with the prospect, either in person, over the phone, or through email, to introduce the product or service and establish a relationship. You can ask questions to get customers to join the conversation, or you can also give them a good gift to try before investing more in

the products. This helps affirm the customer's confidence in the product or service you are selling.

4. **Presentation:** This step involves presenting the product or service to the prospect in a way that highlights its benefits and addresses the prospect's needs and concerns.
5. **Handling objections:** This step involves addressing any objections or concerns the prospect may have and providing additional information or reassurance if necessary. This is a normal and important part of the selling process. View objections as an opportunity to learn more about your prospect. When you research and prepare appropriately, you'll have all the information needed to overcome objections.
6. **Closing:** This step involves asking for the sale and finalizing the transaction.
7. **Follow-up:** This step involves maintaining contact with the customer after the sale to ensure satisfaction and build a long-term relationship.

3.2 Data source and challenges:

3.2.1 Data source

AdventureWorks Database is a sample online transaction processing (OLTP) database provided by Microsoft for free [15]. The AdventureWorks Database supports a multinational manufacturing company called Adventure Works Cycles. This project will utilize the AdventureWorks Database version 2019 and concentrates on module Sales only. Data sources include Manufacturing, Sales, Purchasing, Product Management, Contact Management, and Human Resources. Moreover, they also comprehensively comprise all table structures, including column name, column data type, column constraints, primary key, foreign key, and relationships between tables. The data sources have already been integrated from various sources to use.

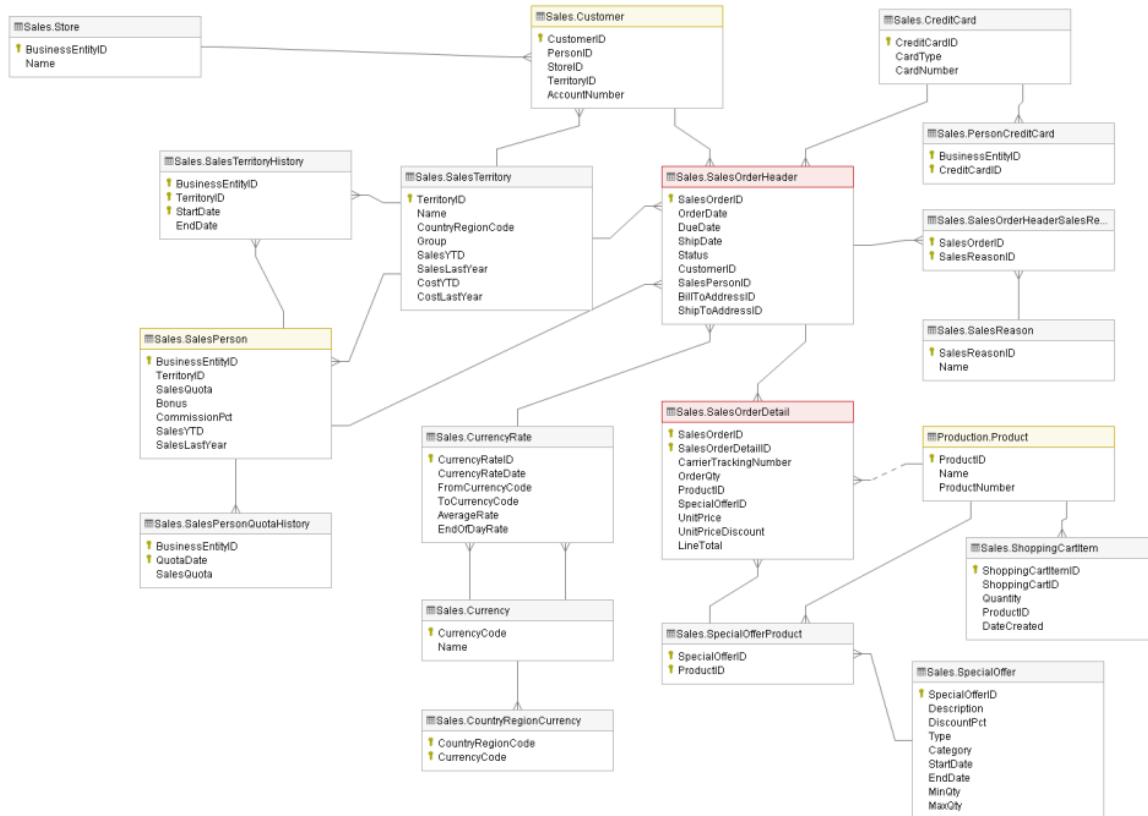


Figure 3- 1 Sales Module [15]

3.2.2 Challenges

- Not enough information needed: Although the data sources are combined from many sources, AdventureWork database can not provide other features that need to be acknowledged to produce reports.
- Low-quality data: There are still duplicated or null values in these tables which lead to more effort to handle data.
- Confusing attributes: There are some attributes that are unable to be understood that can be obstacles in the analysis process.
- End-user adaptation: Training and encouragement to end-users so that they can easily make use of BI tools and BI projects.

3.3 Business Requirements Analysis

3.3.1 How is the sales performance of the business

Sales performance is one of the most significant concerns because it mainly contributes to the revenue of the company. Business needs to acknowledge various aspects of sales operation such as delivery performance, sales operation in terms of sales channels, sales areas,...

3.3.2 Identify the customer segments of the business

Customer segmentation plays an important role in marketing and customer services. Instead of launching shared marketing campaigns for all customers, it is necessary to cluster customers into segments so that business can customize the campaigns and implement toward target segments. This will enhance the efficiency of the marketing strategy and reduce the overheads.

3.3.3 What is the tendency of customers' purchasing behaviors

The tendency of customers' purchasing behaviors is a must-know, especially regarding product categories. Acknowledging the best-seller products will help the managers to allocate effectively resources to stimulate sales and reduce cost for the company.

3.3.4 Research objectives

Based on these analysis above, we suppose to answer the following questions when completing the BI solution:

1. Discover sales performance by regions

Analyzing sales data by region can help businesses acknowledge the regions that are performing well and those that are not generating desired revenue. This information enables businesses to direct resources and investments to regions that are generating more sales and propose solutions for underperforming regions. Examining sales performance by region also provides businesses with useful insights into customer behaviors and preferences in different places. This knowledge allows businesses to adapt their marketing

and sales strategies to better align with the demands and preferences of customers in specific regions.

2. Acknowledge sales performance across online and offline channels

Examining sales performance through various channels can offer useful information about customer engagement. Evaluating sales data by channel enables companies to identify which channels generate the most revenue, which channels are performing poorly, and where resources should be allocated to improve sales performance overall. Additionally, analyzing sales performance by channel can reveal general patterns and trends in customer behavior, such as the favorable channels among particular demographic groups or the most effective channels for promoting specific product items. This knowledge can then be leveraged to launch sales and marketing strategies.

3. Analyzing delivery performance in different regions

Delivery performance is a performance measure that plays an essential role in the success of an organization that fulfills order requests and provides products or services to customers. The satisfaction of customers and future purchasing decisions are decided by freight and delivery time. If the delivery is timely, accurate, and cost-effective, it can enhance the customer experience and build customer loyalty. With the analysis of these problems, the decision-maker can easily track status, optimize transportation costs, and make decisions that improve customer satisfaction, such as on-time delivery rate, the number of completed deliveries, the average time per delivery, and the highest/lowest performing carrier.

To analyze, we use these columns in 3 tables - DimShipMode, DimDate, FactSales:

DimShipMode: ShipModeName

FactSales: SalesOrderID, Status, Freight, OrderDate, ShipDate, DueDate

From those columns, measures will be calculated:

On-time delivery rate = $1 - (\text{The number of order delayed} / \text{The number of orders})$

The average time per delivery = $(\text{The total (ShipDate - OrderDate)}) / \text{The number of orders}$

The average freight value per delivery = The total of freight value / The number of orders

4. Identify customer segments using RFM model

By segmenting customers into smaller groups based on their distinct behaviors, businesses can customize their marketing efforts for each group. This not only enables businesses to target the right audience more accurately, but it also facilitates the creation of more personalized experiences for individual customers, which can result in increased customer loyalty and satisfaction [16].

The time-honored model used to analyze the buying behavior of an RFM customer is based on three factors: the most recent purchase, purchase frequency, and total amount spent. This report utilizes historical transaction data and customer segmentation using the RFM model as a tool to assist marketing managers in making decisions about future marketing campaigns. This helps reduce costs for businesses while increasing the effectiveness of their marketing campaigns. For example, for customers with good R, F, M scores, businesses need to have special care policies. In addition, for customers with high R but good F, M, businesses need to explore opinions and rebuild trust.

To analyze, we use these columns in 3 tables – DimCustomer, DimDate and FactSales:

DimCustomer: CustomerID, Gender, BirthDate, YearlyIncome, LastName

FactSale : SalesOrderID, OrderQty, UnitPrice, ChannelKey, GeographyKey

Recency = DimDate.Date - OrderDate

Frequency = The total (OrderQty)

Monetary = The total (OrderQty * UnitPrice)

5. Analyze customer trends in purchasing behavior for specific products using cross-selling

By using cross-selling analytical, businesses can gain useful knowledge about customer preferences and purchasing behavior. This data allows businesses to create bundled packages or promotions that encourage customers to buy frequently bought-together products. In addition, by offering complementary products, businesses can obtain

a deeper understanding of customer needs and preferences. Analyzing cross-selling data helps businesses identify the products that customers tend to purchase together, which can provide valuable insight into what customers are looking for.

6. Categorize products items based on their importance by ABC classification

Businesses can use ABC classification to categorize items or products based on their significance, which enables them to allocate their resources and efforts more efficiently. By prioritizing management of the most valuable and important products (A-items), businesses can achieve desired sales and profits. By using ABC classification, businesses can gain insights into customer demand and behavior, which can help them modify their sales and marketing strategies to enhance growth and increase revenue. Additionally, businesses can reduce the costs associated with managing less important products and identify opportunities to optimize their supply chain and minimize inventory management costs.

Class A: account for 70% of the Total Sales (Best-Selling)

Class B: making up the next 20% of the Total Sales

Class C: representing the last 10% of the Total Sales

3.4 IT Requirements Analysis

To ensure that the project achieves its objectives and meets the business requirements, the following features must be incorporated into the technology solution.

Accuracy in ETL (Extract, Transform, Load)

For a business, data inaccuracies can lead to significant consequences. Therefore, the ETL data process must ensure accuracy. Various methods are used to verify the similarity of source data and the data warehouse.

High Availability

An organization can be more successful if decisions are made in a timely and prompt manner. Therefore, the project must ensure the system is readily available. From a technical standpoint, this requires having a backup data warehouse, cloud nodes, or other means, as well as a system that operates smoothly and only pauses when necessary.

Intuitive visualization

To ensure that dashboards and other charts are presented in a clear and comprehensible manner that meets the requirements, it is necessary to use intuitive visualization techniques. Moreover, it is essential that charts are updated automatically when changes occur in the source data. To achieve these goals, it is recommended to use dedicated tools such as Power BI and Qlik, which are designed specifically for data visualization and analysis.

Scalability

A developing business must continuously update its technology and improve its systems. Data will be collected from a diverse array of sources, with varying structures and formats. Therefore, the project must ensure that the system is easily scalable in both depth and breadth, allowing for future expansion and growth.

3.5 Comparative Analysis of BI and Data Visualization Tools

3.5.1 Microsoft Power BI

Microsoft Power BI is a top business intelligence tool that blends an intuitive user experience with industry-leading sophisticated analytics [17]. Microsoft Power BI is a set of software services, applications, and interfaces that work together to transform disparate data sources into coherent, graphically immersive, and dynamic insights.

Some of the key features of Microsoft Power BI are:

- Data connectivity: connect both on-premise and cloud-based data sources.
- Mobile support: view dashboards, reports and data visualizations, and can also receive alerts and notifications.
- Interact and share dashboards, reports in both on-premise and cloud.
- No memory and speed constraints
- Extract business information quickly and accurately
- Balance between simplicity and performance
- Support for Advanced Data services
- Incorporating multiple AI-based features

3.5.2 Tableau

Tableau is a powerful and well-known data visualization application. It enables you to explore and interrogate your data by simply clicking, dragging, and dropping data components [18]. From connection through collaboration, Tableau is the most powerful, secure, and flexible end-to-end analytics platform. Besides, Tableau makes it simple to create stunning, dynamic visualizations from raw data. Tableau also helps people and organizations be more data-driven and faster to discover and share insights that can change businesses and the world.

Some of the key features of Tableau are:

- Handle large volumes of data with better performance.
- Power real-time data exploration through integration with third-party apps: Salesforce, Google Analytics, and Amazon Redshift,...
- Get faster, deeper insights
- Drive smarter decisions
- Enhance security
- Visualizations that are both intuitive and graphically rich
- Combine and clean your data without writing code.

3.6 Proposal BI solution

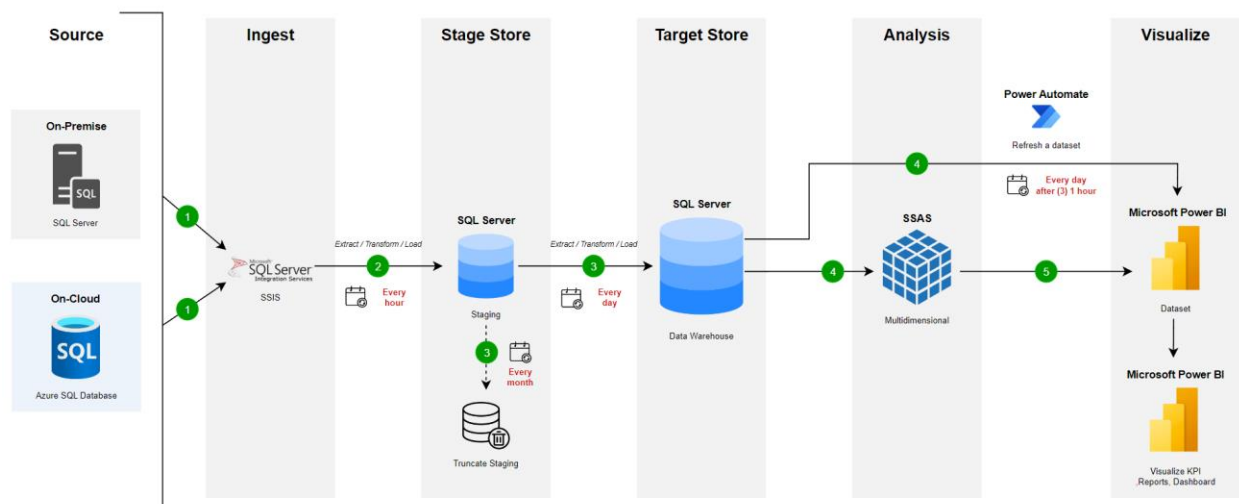


Figure 3- 2 BI Solution (Source: Author's proposal)

Figure 3- 2 above is the BI process proposed by the team to solve the problems raised:

Step 1: Data is slowly loaded from On-Premise and On-Cloud into SQL Server Integration Services (SSIS) tool from Source via Ingest Phase.

Step 2: The data will be extracted, transformed, and loaded into the Staging Area, which is a database in SQL Server with the schema name Integration. This process is created a job to set up to automatically run the package once every hour.

Step 3: Data is extracted, converted, and loaded into the Data warehouse that has been designed with the available Star schema model, which is another database in SQL Server with the schema name DW.{dim/fact_table}. Between the two Stage Store and Target Store Phase are created to run the package automatically once a day. At the same time, every month, the data from the Staging Area will automatically Truncate all the tables in the database.

Step 4: Analysis Phase is divided into 2 branches. The first branch will be automatically loaded/refreshed by the Power Automate tool to load/refresh data from the Data warehouse into Microsoft Power BI Dataset to perform analysis report generation with time once a day and after Step 3 one hour to make sure the data in this step has been pushed through the Target Store Phase to complete. In the second branch, the data will be pushed into the SQL Server Analysis Services (SSAS) tool to create cubes (a special structure for storing multidimensional data) for data analysis and then pushed through Microsoft Power BI Dataset and perform report generation.

CHAPTER 4: BUILDING DATA WAREHOUSE AND INTEGRATING DATA

This chapter will cover the creation of a data warehouse that meets specific requirements. It provides an overview of the process from start to finish, including details about the ETL process and how its various stages work together to produce an effective data warehouse.

4.1 Designing Data Warehouse

4.1.1 Bus Matrix

The Bus matrix visually displays the relationships between business processes and dimensions. Each dimension is listed on one axis and each business process on the other. The intersection of a dimension and a business process represents the data being tracked. It helps create a standardized set of measures for each dimension, aiding in the design and development of a data warehouse. This allows for supporting a wide range of business processes and easy updates as new requirements arise.

| | Employee | Product | Customer | Store | Channel | Geography | ShipMethod | Date |
|-------------------------------|----------|---------|----------|-------|---------|-----------|------------|------|
| Customer segmentation | | | x | | | | | x |
| Delivery performance tracking | | | | | x | x | x | x |
| Sales performance by channel | x | | x | x | x | | | x |
| Sales performance by region | | | | | | x | | x |
| Purchasing behavior | | x | x | | x | x | | x |
| ABC Classification | | x | x | | x | x | | x |

Table 4- 1 Bus matrix

4.1.2 Master Data

Master data of the sales department is a collection of basic and important information about customers, products, and services that the company's sales department is providing. This information includes descriptions of products or services, pricing, customer information, terms and conditions of sales contracts, and other information related to the sales process.

For our case, master data is the information related to many objects in the Table 4-2 below:

| Object | Description |
|-------------|--|
| Customer | Information about Customers (Name,YearlyIncome, Gender..) |
| Employee | Information about Employee(Name, Birth Date, Gender..) |
| Product | Information about Products (SubCategory, Category, Name..) |
| Store | Information about Stores (Name, Year Opened..) |
| Geography | Information about Postal Code, Region.. |
| Ship Method | Information about Ship Methods |

Table 4- 2 Master data

4.1.3 Transaction Data

Transaction data of the sales department is information about sales transactions of the sales department, including customer information, products or services sold, quantity, price, time, and location of the transaction.

Companies and organizations often use transaction data of the sales department to evaluate the effectiveness of sales campaigns, analyze trends and buying behaviors of customers, optimize business processes, and forecast future revenue.

| Object | Description |
|-------------------------------|--|
| Order Header Transaction Data | Information about SubTotal, TaxAmt, Freight,.. |
| Order Detail Transaction Data | Information about ProductID, Order Quantity,.. |

Table 4- 3 Transaction da

4.1.4 ETL mapping

| DataWarehouse | | | DataSource | | | | |
|--------------------|--------------|--------|---------------------------|-------------------|--------------|--------|---------------------------------------|
| AttributeName | DataType | isNULL | Table (Source) | AttributeName | DataType | isNULL | Rule |
| DimCustomer | | | | | | | |
| CustomerKey | int | | | | | | Auto generated |
| CustomerID | int | | Sales.Customer | CustomerID | int | | From source |
| FirstName | nvarchar(50) | x | Person.Person | FirstName | nvarchar(50) | x | From source |
| MiddleName | nvarchar(50) | x | Person.Person | MiddleName | nvarchar(50) | x | From source |
| LastName | nvarchar(50) | x | Person.Person | LastName | nvarchar(50) | x | From source |
| Gender | varchar(1) | x | Sales.vIndividualCustomer | Gender | | x | From source (F = Female; M = Male) |
| BirthDate | varchar(10) | x | Sales.vIndividualCustomer | BirthDate | | x | From source |
| YearlyIncone | varchar(50) | x | Sales.vIndividualCustomer | YearlyIncone | | x | From source |
| DateFirstPurchase | varchar(10) | x | Sales.vIndividualCustomer | DateFirstPurchase | | x | From source |
| ActiveFrom | datetime | | | | | | Auto generated |

| | | | | | | | |
|---------------------|--------------|---|---------------------------|------------------|--------------|---|----------------|
| ActiveTo | datetime | | | | | x | Auto generated |
| Dim Employee | | | | | | | |
| EmployeeKey | int | | | | | | Auto generated |
| EmployeeID | int | | HumanResources.Employee | BusinessEntityID | int | | From source |
| FirstName | nvarchar(50) | | Person.Person | FirstName | nvarchar(50) | | From source |
| MiddleName | nvarchar(50) | x | Person.Person | MiddleName | nvarchar(50) | x | From source |
| LastName | nvarchar(50) | | Person.Person | LastName | nvarchar(50) | | From source |
| Gender | nchar(1) | | HumanResources.Employee | Gender | nchar(1) | | From source |
| BirthDate | date | | HumanResources.Employee | BirthDate | date | | From source |
| PersonType | nchar(2) | | Person.Person | PersonType | nchar(2) | | From source |
| JobTitle | nvarchar(50) | | HumanResources.Employee | JobTitle | nvarchar(50) | | From source |
| HireDate | date | | HumanResources.Employee | HireDate | date | | From source |
| Current Department | nvarchar(50) | | HumanResources.Department | Name | nvarchar(50) | | From source |

| | | | | | | | |
|------------------------|--------------|---|-------------------------------|----------------------|--------------|---|----------------|
| ActiveFrom | datetime | | | | | | Auto generated |
| ActiveTo | datetime | | | | | x | Auto generated |
| DimProduct | | | | | | | |
| ProductKey | int | | | | | | Auto generated |
| ProductID | int | | Production.Product | ProductID | int | | From source |
| ProductSubcategoryID | int | x | Production.ProductSubcategory | ProductSubcategoryID | int | x | From source |
| ProductCategoryID | int | x | Production.ProductCategory | ProductCategoryID | int | x | From source |
| ProductName | nvarchar(50) | | Production.Product | Name | nvarchar(50) | | From source |
| ProductSubcategoryName | nvarchar(50) | x | Production.ProductSubcategory | Name | nvarchar(50) | x | From source |
| ProductCategoryName | nvarchar(50) | x | Production.ProductCategory | Name | nvarchar(50) | x | From source |
| StandardCost | money | | Production.Product | StandardCost | money | | From source |
| List Price | money | | Production.Product | List Price | money | | From source |

| | | | | | | | |
|---------------------|--------------|--|----------------|--------------|--------------|---|---|
| ActiveFrom | datetime | | | | | | Auto generated |
| ActiveTo | datetime | | | | | x | Auto generated |
| DimStore | | | | | | | |
| StoreKey | int | | | | | | Auto generated |
| GeographyKey | int | | DimGeography | | | | Look up from [DimGeography].[Geograp hyKey] |
| StoreID | int | | Sales.Store | StoreID | | | From source |
| StoreName | nvarchar(50) | | Sales.Store | Name | | | From source |
| BusinessType | int | | Sales.Store | Demographics | | | From source |
| YearOpened | Int | | Sales.Store | Demographics | | | From source |
| NumberEmployees | int | | Sales.Store | Demographics | | | From source |
| DimGeography | | | | | | | |
| GeographyKey | int | | | | | | Auto generated |
| PostalCode | nvarchar(15) | | Person.Address | PostalCode | nvarchar(15) | | From source |
| City | nvarchar(30) | | Person.Address | City | nvarchar(30) | | From source |

| | | | | | | | |
|-----------------------|--------------|--|-----------------------|----------------|--------------|--|---|
| SalseProvinceName | nvarchar(50) | | Person.StateProvince | Name | nvarchar(50) | | Look up from [Person.StateProvince].[StateProvinceID] |
| SalesTerritoryCountry | nvarchar(50) | | Person.CountryRegion | Name | nvarchar(50) | | Look up from [Person.CountryRegion].[CountryRegionCode] |
| SalesTerritoryRegion | nvarchar(50) | | Sales.SalesTerritory | Name | nvarchar(50) | | Look up from [Sales.SalesTerritory].[TerritoryID] |
| SalesTerritoryGroup | nvarchar(50) | | Sales.SalesTerritory | Group | nvarchar(50) | | Look up from [Sales.SalesTerritory].[TerritoryID] |
| TerritoryID | int | | Person.StateProvince | TerritoryID | int | | Look up from [Person.StateProvince].[StateProvinceID] |
| DimShipMethod | | | | | | | |
| ShipMethodKey | int | | | | | | Auto generated |
| ShipMethodID | int | | Purchasing.ShipMethod | ShipMethodID | int | | From source |
| ShipMethodName | nvarchar(50) | | Purchasing.ShipMethod | ShipMethodName | nvarchar(50) | | From source |
| DimChannel | | | | | | | |

| | | | | | | | |
|-----------------|------------|---|------------------------|-----------------|------------|--|--|
| ChannelKey | int | | | | | | Auto generated |
| ChannelID | int | | Sales.SalesOrderHeader | OnlineOrderFlag | bit | | 0 (False) = 0; 1 (True) = 1 |
| ChannelName | varchar(7) | | Sales.SalesOrderHeader | OnlineOrderFlag | varchar(7) | | ChannelID = 0 --> Offline; ChannelID = 1 --> Online |
| DimDate | | | | | | | |
| DateKey | int | | | | | | Auto generated |
| Date | datetime | | | | | | Auto generated |
| FullDate | char(10) | | | | | | Auto generated |
| DayOfMonth | varchar(2) | x | | | | | Auto generated |
| DayName | varchar(9) | x | | | | | Auto generated |
| DayOfWeek | char(1) | x | | | | | Auto generated |
| Month | varchar(2) | x | | | | | Auto generated |
| MonthName | varchar(9) | x | | | | | Auto generated |
| Quarter | char(1) | x | | | | | Auto generated |
| Year | char(4) | x | | | | | Auto generated |
| MonthYear | char(10) | x | | | | | Auto generated |
| MMYYYY | char(6) | x | | | | | Auto generated |
| FirstDayOfMonth | char(6) | x | | | | | Auto generated |
| LastDayOfMonth | date | x | | | | | Auto generated |

| | | | | | | | |
|-------------------|-------------|---|---------------|---------------|-----|--|--|
| FirstDayOfQuarter | date | x | | | | | Auto generated |
| LastDayOfQuarter | date | x | | | | | Auto generated |
| FirstDayOfYear | date | x | | | | | Auto generated |
| LastDayOfYear | date | x | | | | | Auto generated |
| IsWeekend | bit | x | | | | | Auto generated |
| IsHoliday | bit | x | | | | | Auto generated |
| HolidayName | varchar(50) | x | | | | | Auto generated |
| FactSales | | | | | | | |
| CustomerKey | int | | DimCustomer | CustomerKey | int | | Look up from [DimCustomer].[CustomerKey] |
| EmployeeKey | int | x | DimEmployee | EmployeeKey | int | | Look up from [DimEmployee].[EmployeeKey] |
| ProductKey | int | | DimProduct | ProductKey | int | | Look up from [DimProduct].[ProductKey] |
| ShipMethodKey | int | | DimShipMethod | ShipMethodKey | int | | Look up from [DimShipMethod].[ShipMethodKey] |

| | | | | | | | |
|--------------------|----------|---|------------------------|--------------------|----------|---|--|
| GeographyKey | int | | DimGeography | GeographyKey | int | | Look up from [DimGeography].[GeographyKey] |
| DateKey | int | | DimDate | DateKey | int | | Look up from [DimDate].[DateKey] |
| StoreKey | int | x | DimStore | StoreKey | int | | Look up from [DimStore].[StoreKey] |
| ChannelKey | int | | DimChannel | ChannelKey | int | | Look up from [DimChannel].[ChannelKey] |
| SalesOrderID | int | | Sales.SalesOrderHeader | SalesOrderID | int | | From source |
| SalesOrderDetailID | int | | Sales.SalesOrderDetail | SalesOrderDetailID | Int | | From source |
| OrderDate | datetime | | Sales.SalesOrderHeader | OrderDate | datetime | | From source |
| DueDate | datetime | | Sales.SalesOrderHeader | DueDate | datetime | | From source |
| ShipDate | datetime | x | Sales.SalesOrderHeader | ShipDate | datetime | x | From source |
| OrderStatus | tinyint | | Sales.SalesOrderHeader | Status | tinyint | | From source |

| | | | | | | | |
|-----------|----------|--|------------------------|-----------|----------|--|-------------|
| OrderQty | smallint | | Sales.SalesOrderDetail | OrderQty | smallint | | From source |
| UnitPrice | money | | Sales.SalesOrderDetail | UnitPrice | money | | From source |
| SubTotal | money | | Sales.SalesOrderHeader | SubTotal | money | | From source |
| TaxAmt | money | | Sales.SalesOrderHeader | TaxAmt | money | | From source |
| Freight | money | | Sales.SalesOrderHeader | Freight | money | | From source |

Table 4- 4 ETL mapping

4.1.5 Fact and dimension tables

DimCustomer table: Store the features related to Customers, including their primary key, natural key, and attributes of the customer's information.

| | Column Name | Data Type | Allow Nulls |
|---|-------------------|--------------|-------------------------------------|
| ▶ | CustomerKey | int | <input type="checkbox"/> |
| | CustomerID | int | <input type="checkbox"/> |
| | FirstName | nvarchar(50) | <input checked="" type="checkbox"/> |
| | MiddleName | nvarchar(50) | <input checked="" type="checkbox"/> |
| | LastName | nvarchar(50) | <input checked="" type="checkbox"/> |
| | Gender | varchar(1) | <input checked="" type="checkbox"/> |
| | BirthDate | varchar(10) | <input checked="" type="checkbox"/> |
| | YearlyIncome | varchar(50) | <input checked="" type="checkbox"/> |
| | DateFirstPurchase | varchar(10) | <input checked="" type="checkbox"/> |
| | ActiveFrom | datetime | <input type="checkbox"/> |
| | ActiveTo | datetime | <input checked="" type="checkbox"/> |

Figure 4- 1 DimCustomer table

DimEmployee table: Store the features related to Employees, including their primary key, natural key, and attributes of the employee's information.

| | Column Name | Data Type | Allow Nulls |
|---|-------------------|--------------|-------------------------------------|
| ▶ | EmployeeKey | int | <input type="checkbox"/> |
| | EmployeeID | int | <input type="checkbox"/> |
| | FirstName | nvarchar(50) | <input type="checkbox"/> |
| | MiddleName | nvarchar(50) | <input checked="" type="checkbox"/> |
| | LastName | nvarchar(50) | <input type="checkbox"/> |
| | Gender | nvarchar(1) | <input type="checkbox"/> |
| | BirthDate | date | <input type="checkbox"/> |
| | PersonType | nvarchar(2) | <input type="checkbox"/> |
| | JobTitle | nvarchar(50) | <input type="checkbox"/> |
| | HireDate | date | <input type="checkbox"/> |
| | CurrentDepartment | nvarchar(50) | <input type="checkbox"/> |
| | ActiveFrom | datetime | <input type="checkbox"/> |
| | ActiveTo | datetime | <input checked="" type="checkbox"/> |

Figure 4- 2 DimEmployee table

DimProduct table: Store the features related to Products, including their primary key, natural key, and attributes of the product's information.


| | Column Name | Data Type | Allow Nulls |
|---|------------------------|--------------|-------------------------------------|
|  | ProductKey | int | <input type="checkbox"/> |
| | ProductID | int | <input type="checkbox"/> |
| | ProductSubcategoryID | int | <input checked="" type="checkbox"/> |
| | ProductCategoryID | int | <input checked="" type="checkbox"/> |
| | ProductName | nvarchar(50) | <input type="checkbox"/> |
| | ProductSubCategoryName | nvarchar(50) | <input checked="" type="checkbox"/> |
| | ProductCategoryName | nvarchar(50) | <input checked="" type="checkbox"/> |
| | StandardCost | money | <input type="checkbox"/> |
| | ListPrice | money | <input type="checkbox"/> |
| | ActiveFrom | datetime | <input type="checkbox"/> |
| | ActiveTo | datetime | <input checked="" type="checkbox"/> |

Figure 4- 3 DimProduct table

DimStore table: Store the features related to Stores, including their primary key, natural key, and attributes of the store's information.


| | Column Name | Data Type | Allow Nulls |
|---|-----------------|--------------|--------------------------|
|  | StoreKey | int | <input type="checkbox"/> |
| | GeographyKey | int | <input type="checkbox"/> |
| | StoreID | int | <input type="checkbox"/> |
| | StoreName | nvarchar(50) | <input type="checkbox"/> |
| | BusinessType | varchar(20) | <input type="checkbox"/> |
| | YearOpened | int | <input type="checkbox"/> |
| | NumberEmployees | int | <input type="checkbox"/> |

Figure 4- 4 DimStore table

DimGeography table: Store detailed Geography information, including their primary key, and attributes.


| | Column Name | Data Type | Allow Nulls |
|---|-----------------------|--------------|--------------------------|
|  | GeographyKey | int | <input type="checkbox"/> |
| | PostalCode | nvarchar(15) | <input type="checkbox"/> |
| | City | nvarchar(30) | <input type="checkbox"/> |
| | StateProvinceName | nvarchar(50) | <input type="checkbox"/> |
| | SalesTerritoryCountry | nvarchar(50) | <input type="checkbox"/> |
| | SalesTerritoryRegion | nvarchar(50) | <input type="checkbox"/> |
| | SalesTerritoryGroup | nvarchar(50) | <input type="checkbox"/> |
| | TerritoryID | int | <input type="checkbox"/> |

Figure 4- 5 DimGeography table

DimShipMethod table: Store the features related to ship method, including their primary key, natural key, and attributes of ship method's information.

| | Column Name | Data Type | Allow Nulls |
|---|----------------|--------------|--------------------------|
| ▶ | ShipMethodKey | int | <input type="checkbox"/> |
| | ShipMethodID | int | <input type="checkbox"/> |
| | ShipMethodName | nvarchar(50) | <input type="checkbox"/> |

Figure 4- 6 DimShipMethod table

DimChannel table: Store the features related to Channel, including their primary key, natural key, and attributes of channel's information.

| | Column Name | Data Type | Allow Nulls |
|---|-------------|------------|--------------------------|
| ▶ | ChannelKey | int | <input type="checkbox"/> |
| | ChannelID | int | <input type="checkbox"/> |
| | Channel | varchar(7) | <input type="checkbox"/> |

Figure 4- 7 DimChannel table

DimDate table: Store primary key, and attributes of date information.

| | Column Name | Data Type | Allow Nulls |
|---|-------------------|-------------|-------------------------------------|
| ▶ | DateKey | int | <input type="checkbox"/> |
| | Date | datetime | <input type="checkbox"/> |
| | FullDate | char(10) | <input type="checkbox"/> |
| | DayOfMonth | varchar(2) | <input checked="" type="checkbox"/> |
| | DayName | varchar(9) | <input checked="" type="checkbox"/> |
| | DayOfWeek | char(1) | <input checked="" type="checkbox"/> |
| | Month | varchar(2) | <input checked="" type="checkbox"/> |
| | MonthName | varchar(9) | <input checked="" type="checkbox"/> |
| | Quarter | char(1) | <input checked="" type="checkbox"/> |
| | Year | char(4) | <input checked="" type="checkbox"/> |
| | MonthYear | char(10) | <input checked="" type="checkbox"/> |
| | MMYYYY | char(6) | <input checked="" type="checkbox"/> |
| | FirstDayOfMonth | date | <input checked="" type="checkbox"/> |
| | LastDayOfMonth | date | <input checked="" type="checkbox"/> |
| | FirstDayOfQuarter | date | <input checked="" type="checkbox"/> |
| | LastDayOfQuarter | date | <input checked="" type="checkbox"/> |
| | FirstDayOfYear | date | <input checked="" type="checkbox"/> |
| | LastDayOfYear | date | <input checked="" type="checkbox"/> |
| | IsWeekend | bit | <input type="checkbox"/> |
| | IsHoliday | bit | <input type="checkbox"/> |
| | HolidayName | varchar(50) | <input checked="" type="checkbox"/> |
| | | | <input type="checkbox"/> |

Figure 4- 8 DimDate table

FactSales table: Store the characteristics that are used for examination, including foreign key referenced to every Dimension table, along with the measure attributes of every transaction.

| | Column Name | Data Type | Allow Nulls |
|---|--------------------|-----------|-------------------------------------|
| ► | DateKey | int | <input type="checkbox"/> |
| | CustomerKey | int | <input type="checkbox"/> |
| | EmployeeKey | int | <input checked="" type="checkbox"/> |
| | ProductKey | int | <input type="checkbox"/> |
| | GeographyKey | int | <input type="checkbox"/> |
| | ShipMethodKey | int | <input type="checkbox"/> |
| | StoreKey | int | <input checked="" type="checkbox"/> |
| | ChannelKey | int | <input type="checkbox"/> |
| | SalesOrderID | int | <input type="checkbox"/> |
| | SalesOrderDetailID | int | <input type="checkbox"/> |
| | OrderDate | datetime | <input type="checkbox"/> |
| | DueDate | datetime | <input type="checkbox"/> |
| | ShipDate | datetime | <input checked="" type="checkbox"/> |
| | OrderStatus | tinyint | <input type="checkbox"/> |
| | OrderQty | smallint | <input type="checkbox"/> |
| | UnitPrice | money | <input type="checkbox"/> |
| | SubTotal | money | <input type="checkbox"/> |
| | TaxAmt | money | <input type="checkbox"/> |
| | Freight | money | <input type="checkbox"/> |

Figure 4- 9 FactSales table

4.1.6 Data Warehouse model

Constructing a schema is one of the most important steps in building a Data Warehouse system. The model is built based on data obtained from the business and customers. In this project, we designed the simplest model used in DWH, which is the Star Schema. The fact table is positioned at the center of the model, while the dimension tables are arranged around it. Moreover, the simplicity of the structure also reduces the time required to load large batches of records into the database and execute query results efficiently and consistently. This helps businesses improve their performance and easily expand the model in the future.

We propose a data warehouse model as shown in Figure 4- 10 below with 8 dimension tables and 1 fact table. The diagram below illustrates the fundamental conceptual diagram of the proposed data warehouse in Star format.

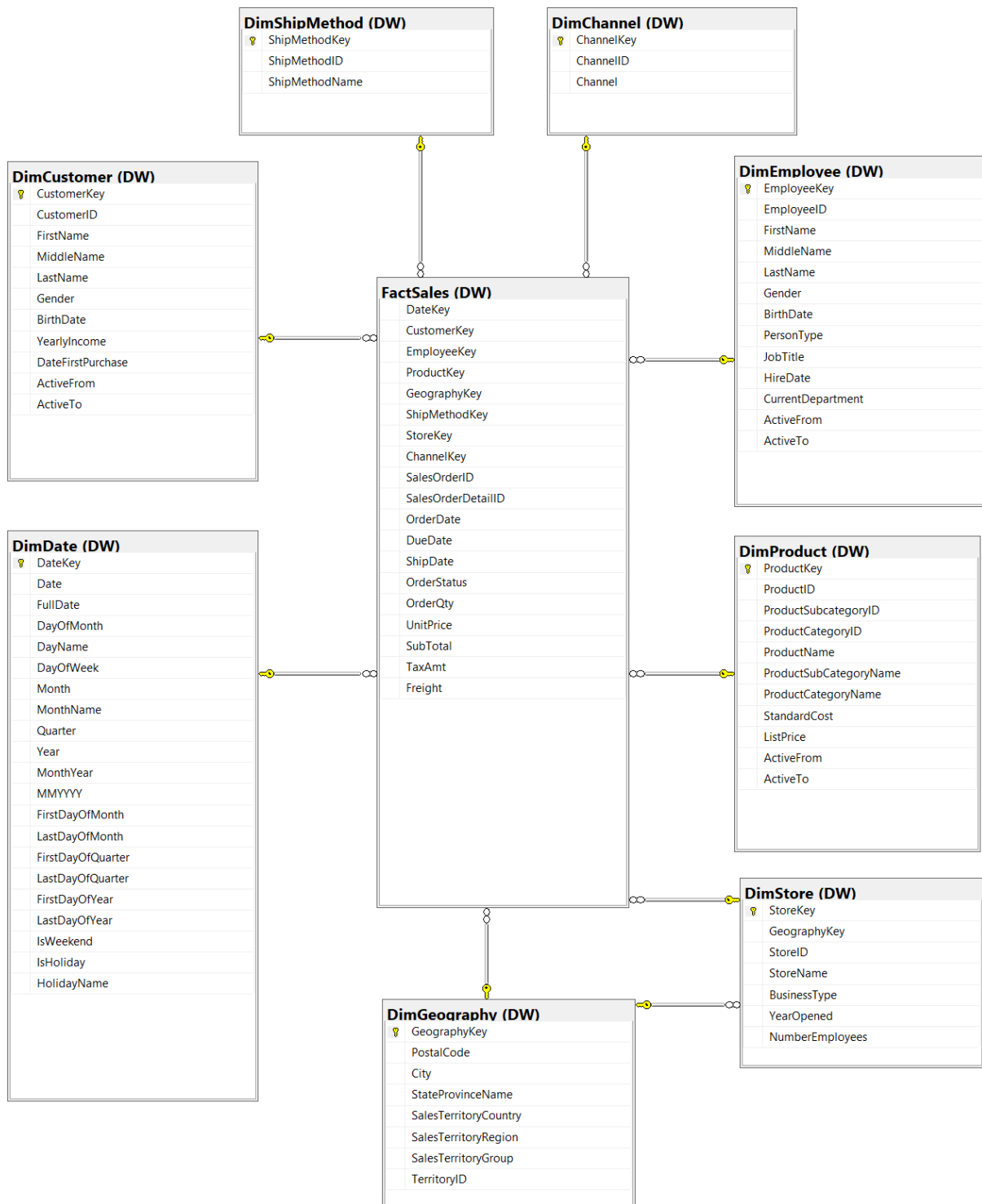


Figure 4- 10 Data Warehouse Star Schema

4.2. ETL processes

4.2.1 ETL pipeline

The ETL pipeline is a process of extracting data from various sources, transforming the data, and loading it into a suitable storage location. We use the SSIS tool to perform ETL processes. This process is carefully designed and refined to ensure the consistency and accuracy of the input and output data. The proposed ETL pipeline is illustrated below:



Figure 4- 11 ETL pipeline

Based on the pipeline shown in Figure 4- 11 above, it is divided into 3 phases:

Phase 1: Load data from Source to Staging Area

Phase 2: Load data from Staging to Dimension Tables

Phase 3: Load data from Staging to Fact Table

4.2.2 Staging Area's ETL Process

The Staging Area is a temporary storage and management location for data until it is processed by ETL into a Data warehouse. The tables in the Staging Area are designed independently of each other and do not have any relationships. In this phase, data is extracted from the source, transformed, and loaded into the Integration schema in the Staging Area that is shown as Figure 4- 12 below.

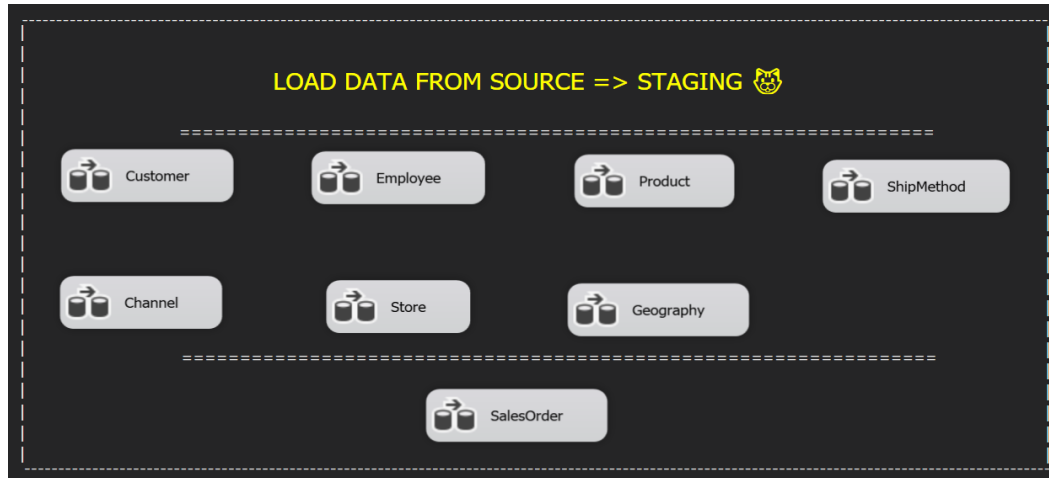


Figure 4- 12 Load data from Source to Staging Area

The Table 4- 5 is structure of the Staging Area with the schema name Integration is represented:

| Database Name | Table Name |
|--------------------------|--------------------------------|
| BI_Staging_BoKho_K20406C | Integration.Customer_Staging |
| | Integration.Employee_Staging |
| | Integration.Product_Staging |
| | Integration.Store_Staging |
| | Integration.Geography_Staging |
| | Integration.ShipMethod_Staging |
| | Integration.Channel_Staging |
| | Integration.SalesOrder_Staging |

Table 4- 5 Staging Area's table structure

4.2.3 Dimension Table's ETL Process

In this phase, we have used Sequence Container as a component in SSIS to group activities and create an order of execution in an ETL package. Grouping activities together facilitates the management and establishes a predetermined order of execution. Following the Figure 4- 13 below for our designed implementation:

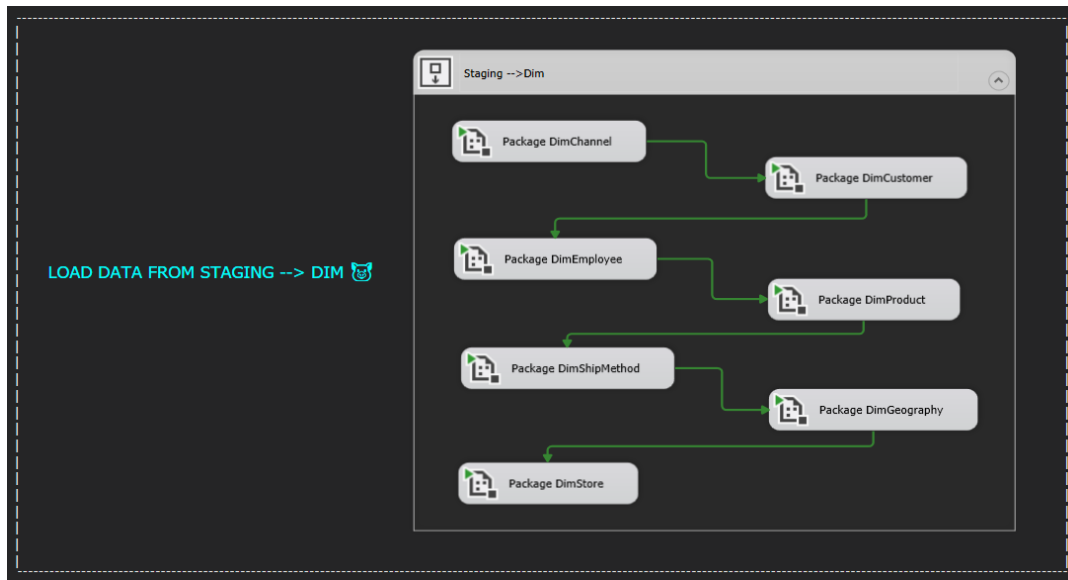


Figure 4- 13 Load data from Staging Area from Dimension Tables

The Table 4- 6 is structure of the Data Warehouse with the schema name DW is represented:

| Database Name | Table Name |
|---------------------|------------------|
| BI_DW_BoKho_K20406C | DW.DimCustomer |
| | DW.DimEmployee |
| | DW.DimProduct |
| | DW.DimStore |
| | DW.DimGeography |
| | DW.DimShipMethod |
| | DW.DimChannel |
| | DW.DimDate |

Table 4- 6 Data warehouse's dimension table structure

DimCustomer: Load data from Integration.Customer_Staging to DimCustomer

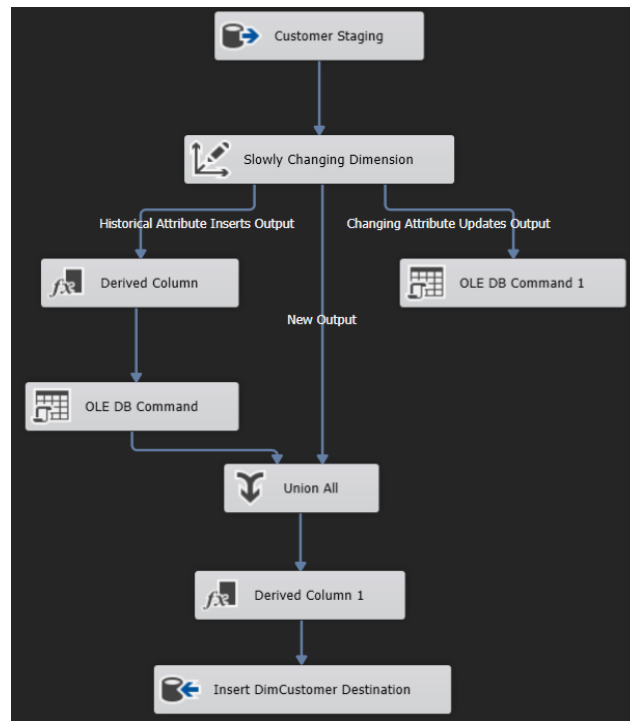


Figure 4- 14 Load data to DimCustomer

DimEmployee: Load data from Integration.Employee_Staging to DimEmployee

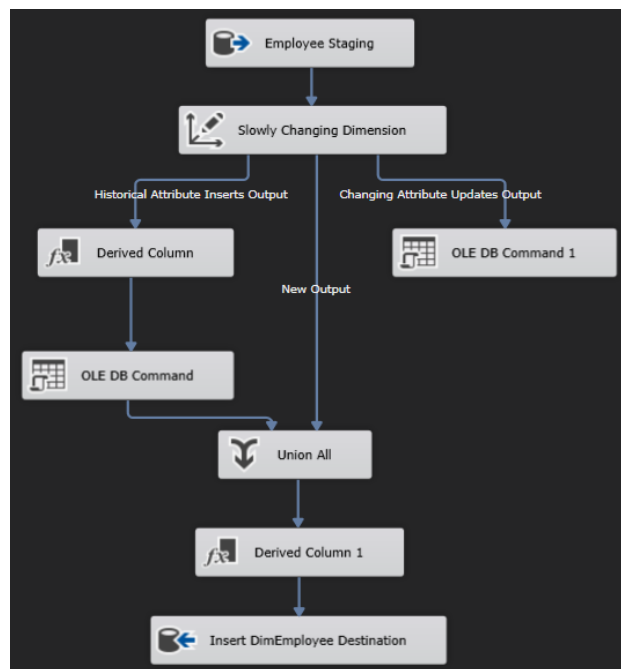


Figure 4- 15 Load data to DimEmployee

DimProduct: Load data from Integration.Product_Staging to DimProduct

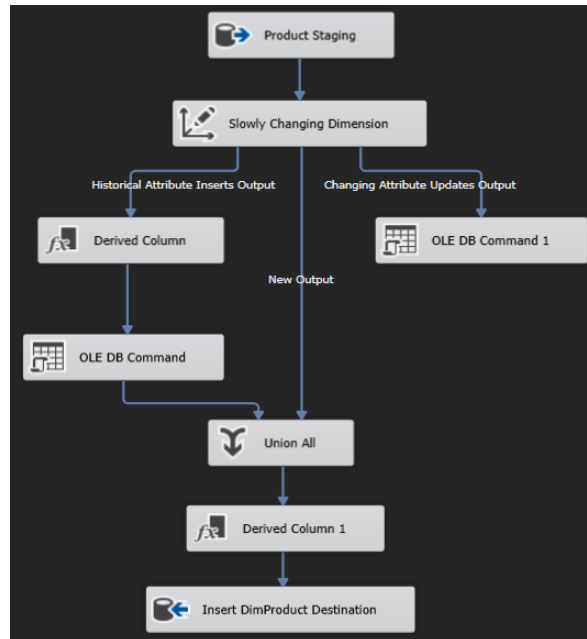


Figure 4- 16 Load data to DimProduct

DimGeography: Load data from Integration.Geography_Staging to DimGeography

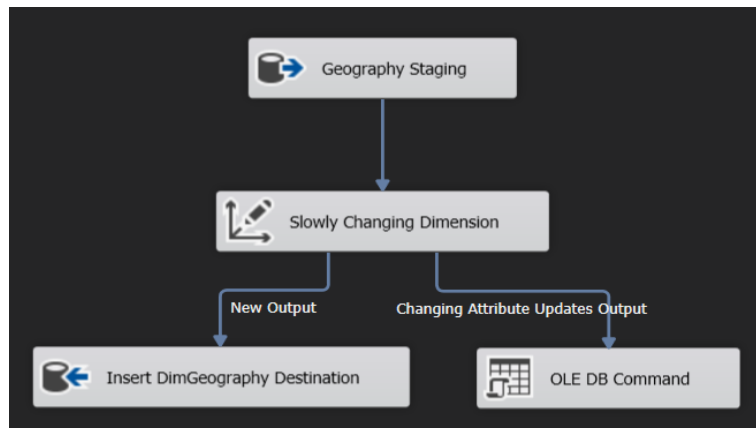


Figure 4- 17 Load data to DimGeography

DimStore: Load data from Integration.Store_Staging to DimStore

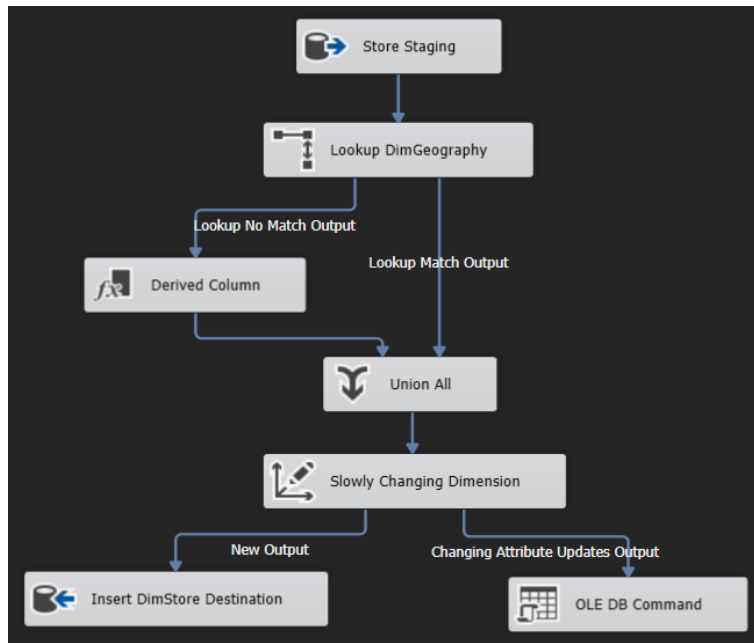


Figure 4- 18 Load data to DimStore

DimShipMethod: Load data from Integration.ShipMethod_Staging to DimShipMethod

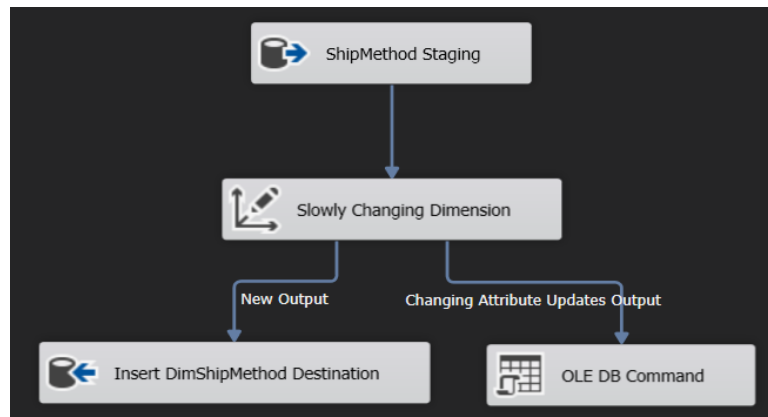


Figure 4- 19 Load data to DimShipMethod

DimChannel: Load data from Integration.Channel_Staging to DimChannel

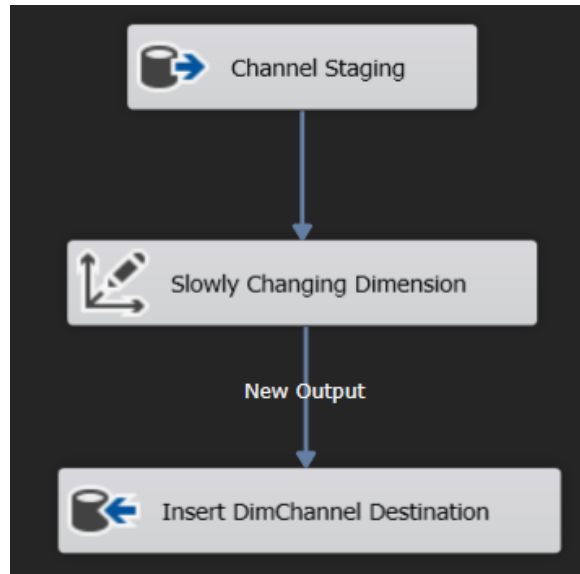


Figure 4- 20 Load data to DimChannel

DimDate: Create DimDate by using Execute SQL Task

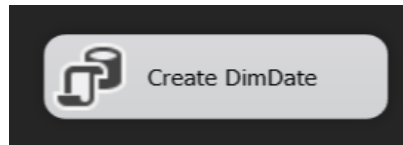


Figure 4- 21 Create DimDate

4.2.4 Fact Table's ETL Process

We have only one fact table in our data warehouse, which is FactSales. This table represents sales transactions over time that is indicated as Figure 4- 22 below.

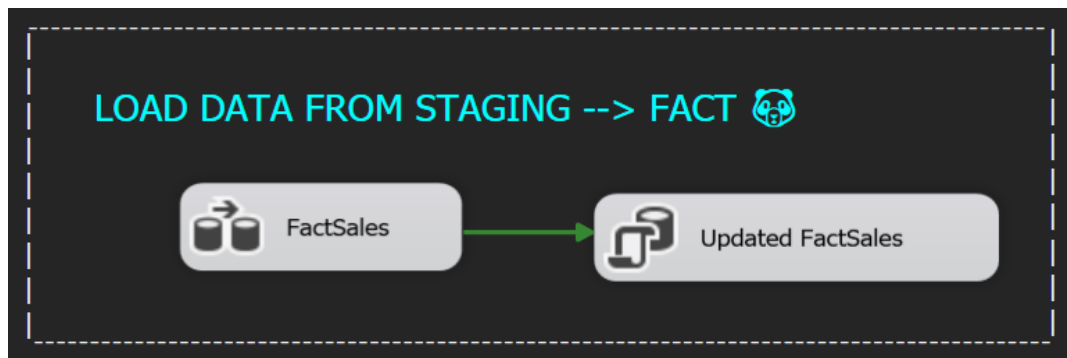


Figure 4- 22 Load data from Staging Area to Fact Table

| Database Name | Table Name |
|---------------------|----------------------|
| BI_DW_BoKho_K20406C | DW.FactSales |
| | DW.FactSales_Updated |

Table 4- 7 Data warehouse's facttable structure

To perform a lookup on customer and employee information, we need to access and use the corresponding dimension tables, including the DimCustomer and DimEmployee tables.

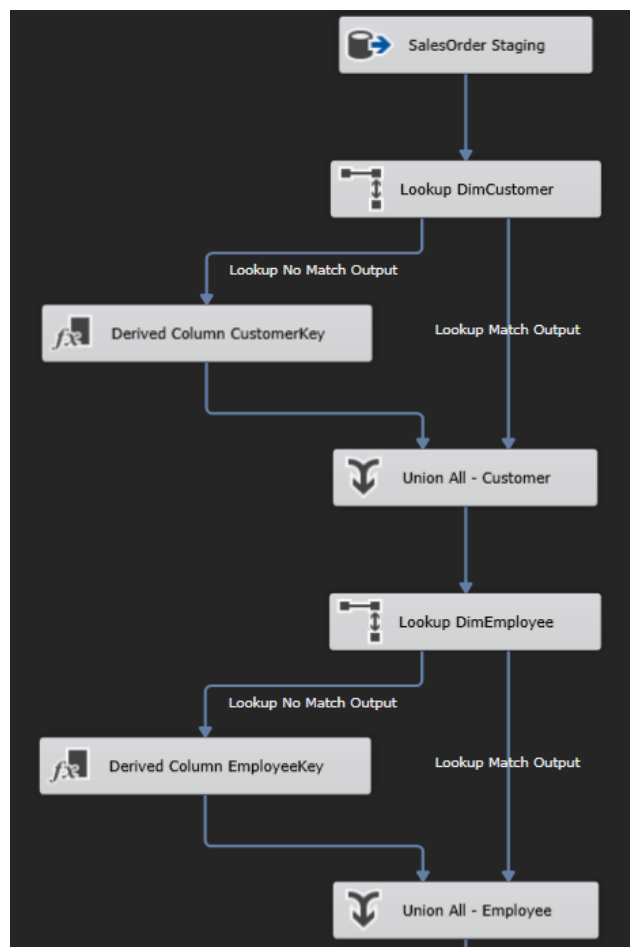


Figure 4- 23 Lookup DimCustomer & DimEmployee

To perform a lookup on product and ship method information, we need to access and use the corresponding dimension tables, including the DimProduct and DimShipMethod tables.

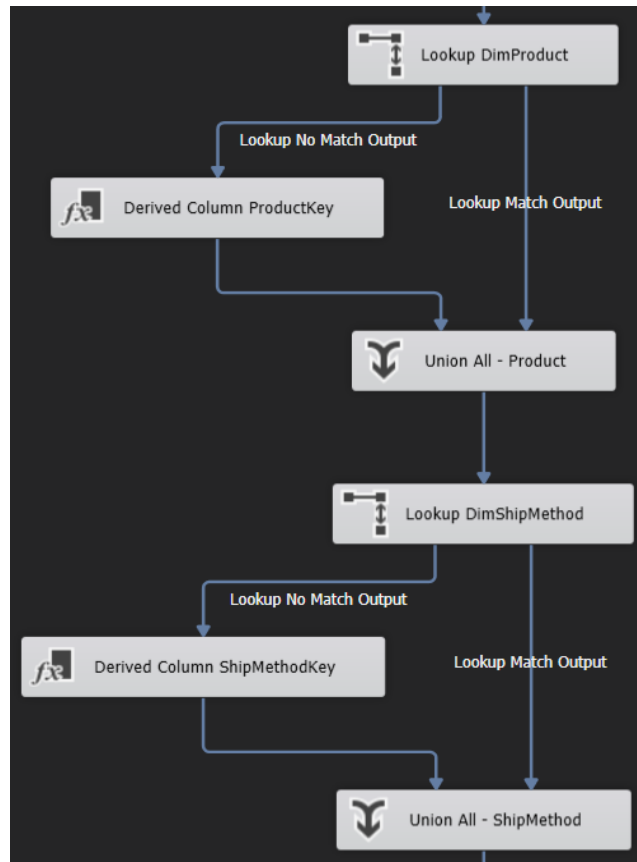


Figure 4- 24 Lookup DimProduct & DimShipMethod

To perform a lookup on channel and geography information, we need to access and use the corresponding dimension tables, including the DimChannel and DimGeography tables.

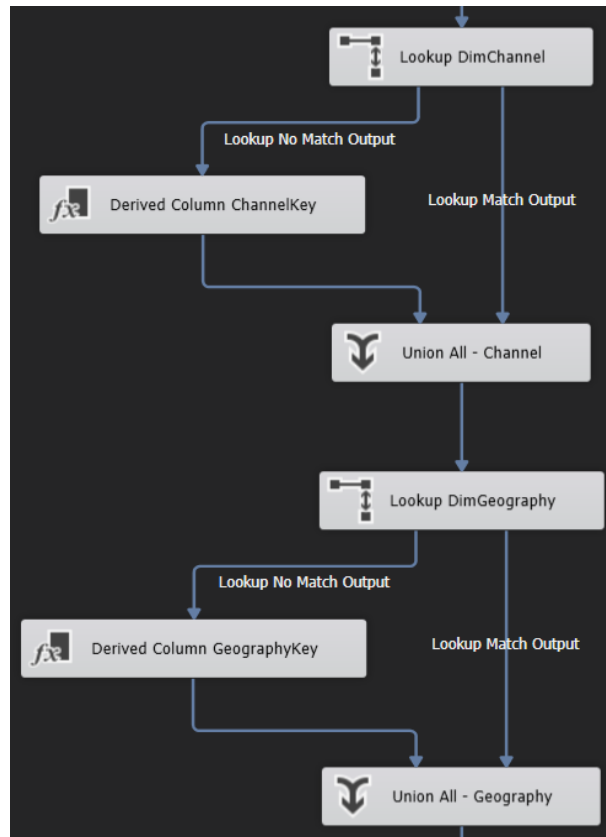


Figure 4- 25 Lookup DimChannel & DimGeography

To perform a Lookup on store and date information, we need to access and use the corresponding dimension tables, including the DimStore and DimDate tables.

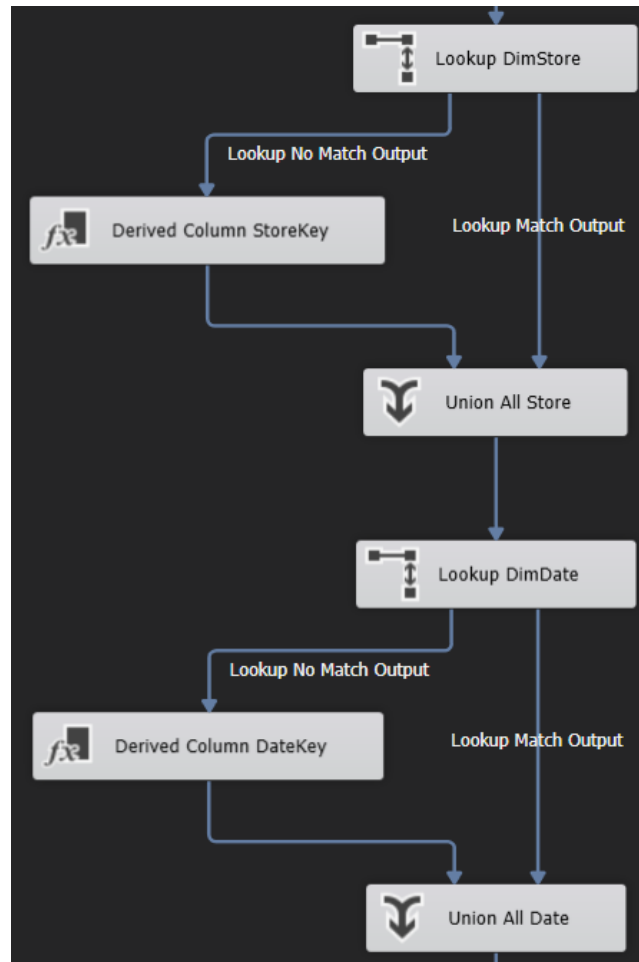


Figure 4- 26 Lookup DimStore & DimDate

Final, we perform a Look up on the FactSales table using the SalesOrderID and SalesOrderDetailID columns. The Lookup No Match Output will add new rows to the FactSales table. The Lookup Match Output will then go through an SCD process to check if any attributes in the data set have changed, and if so, they will be added and updated in the FactSales_Updated table. Then, we execute an SQL Task to update the data back into the FactSales table. This ensures data integrity and avoids overwriting old data in case of errors during the update process.

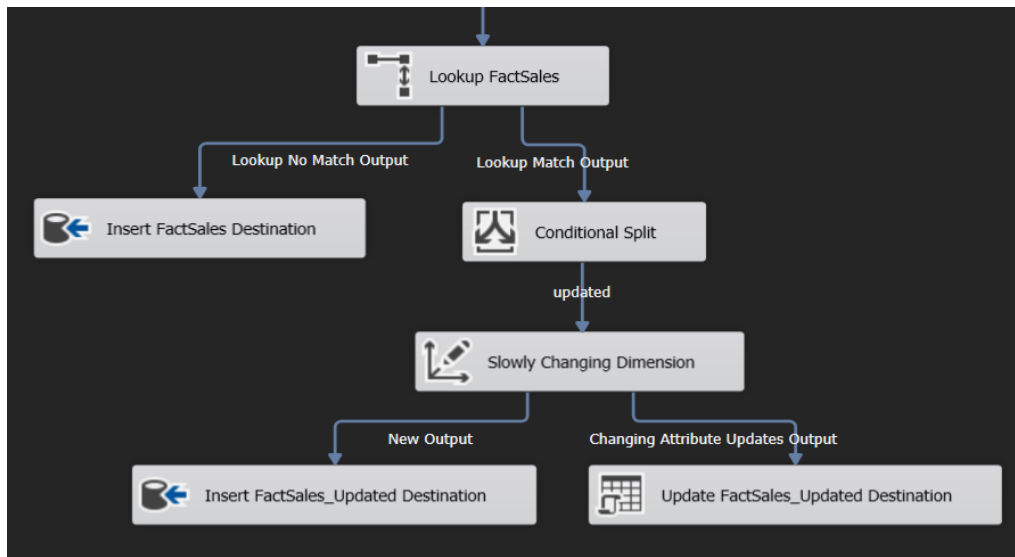


Figure 4- 27 Lookup FactSales and SCD FactSales_Updated

4.2.5 Truncate Table in Staging Area

At this phase, we perform a full truncate of all tables in the Staging Area using an automated job that runs once a month. This is done to clean up old data, improve system performance, and prepare for receiving new data.



Figure 4- 28 Truncate all tables in Staging Area

Configure the ETL steps for Truncate all tables in Staging Area, the schedule jobs described in the Figure 4- 29.

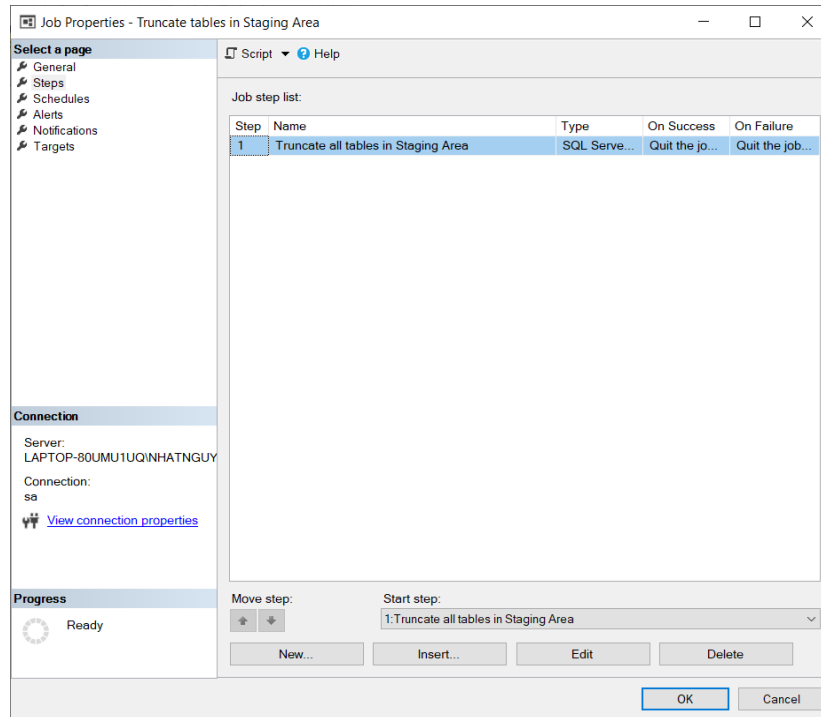


Figure 4- 29 Schedule jobs configure the step truncate

Determine the suitable time for this iteration based on our business needs. In this situation, the optimal time to extract, transform, and load all our daily transactions is at midnight. As a result, the solution will execute every midnight as indicated in Figure 4- 30

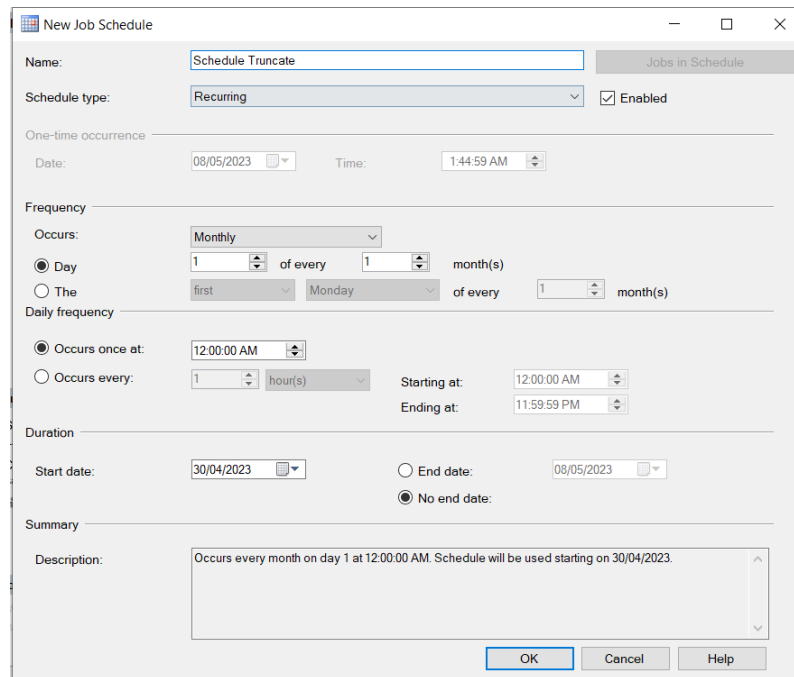


Figure 4- 30 Job schedule properties configure the job Truncate

CHAPTER 5: RESULTS – DATA ANALYTICS AND VISUALIZATION

In this chapter, we will be presenting the results of our implementation efforts in detail. Our focus will be on providing an overview of the insights and useful information that we have been able to gather about the business through the use of dashboards. Moreover, this chapter also proposes some suggestions for analyzed results.

5.1 Data analytics with SSAS technology

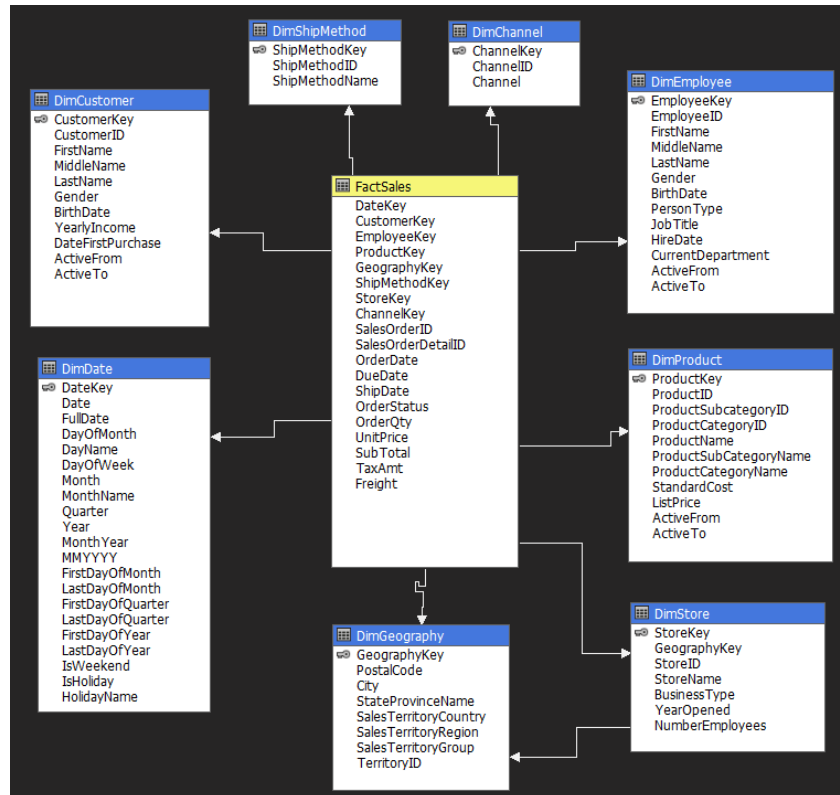


Figure 5- 1 Data warehouse schema (SSAS)

The data warehouse consists of 8 Dimension tables as shown in Figure 5- 1: DimCustomer, DimShipMethod, DimChannel, DimEmployee, DimProduct, DimStore, DimGeography, DimDate, and 1 Fact table: FactSales. The Fact table is connected to the Dimension tables via foreign keys, as follows:

The DimCustomer table includes CustomerKey (primary key) and essential customer information: CustomerID, FullName, Gender, BirthDate, YearlyIncome, and DateFirstPurchase. Additionally, the table has two columns, ActiveFrom and ActiveTo, to

distinguish between active and inactive data. This table is linked to the FactSales table on CustomerKey.

The DimShipMethod table includes ShipMethodKey (primary key) and information about the ship method: ShipMethodID and ShipMethodName. This table is linked to the FactSales table on the ShipMethodKey.

The DimChannel table includes ChannelKey (primary key) and information about the channel: ChannelID, Channel (name). This table is linked to the FactSales table on ChannelKey.

The DimEmployee table includes EmployeeKey (primary key) and information about the employee: EmployeeID, FullName, Gender, BirthDate, PersonType, JobTitle, HireDate, and CurrentDepartment. Additionally, the table has two columns, ActiveFrom and ActiveTo, to distinguish between active and inactive data. This table is linked to the FactSales table on EmployeeKey.

The DimProduct table includes ProductKey (primary key) and information about the product: ProductID, ProductSubcategoryID, ProductCategoryID, ProductName, ProductSubCategoryName, ProductCategoryName, StandardCost, and ListPrice. Additionally, the table has two columns, ActiveFrom and ActiveTo, to distinguish between active and inactive data. This table is linked to the FactSales table on ProductKey.

The DimStore table includes StoreKey (primary key) and information about the store: StoreID, GeographyKey (foreign key linked to the DimGeography table), StoreName, BusinessType, YearOpened, and NumberEmployees. This table is linked to the FactSales table on the StoreKey

The DimGeography table includes GeographyKey (primary key) and information about the geography: PostalCode, City, StateProvinceName, SalesTerritoryCountry, SalesTerritoryRegion, SalesTerritoryGroup, and TerritoryID. This table is linked to the FactSales table on the GeographyKey.

The DimDate table includes DateKey (primary key) and information about the date and time: Date, FullDate, DayOfMonth, DayName, DayOfWeek, Month, MonthName, Quarter, Year, MonthYear, MMYYYY, FirstDayOfMonth, LastDayOfMonth,

FirstDayOfQuarter, LastDayOfQuarter, FirstDayOfYear, LastDayOfYear, IsWeekend, IsHoliday, and HolidayName. This table is linked to the FactSales table on DateKey.

The FactSales table includes all foreign keys from the Dimension tables and detailed information about the sales transaction: SaleOrderID, OrderDate, DueDate, ShipDate, OrderStatus, OrderQty, UnitPrice, SubTotal, TaxAmt, and Freight.

Advantages of SSAS Analysis:

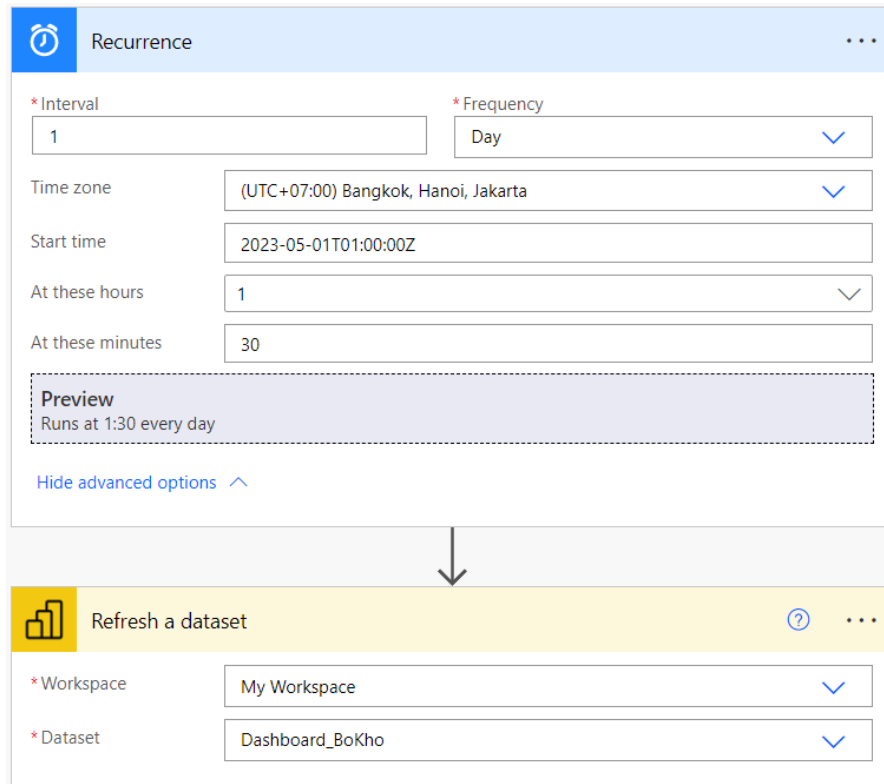
- Using SSAS can help prevent conflicts with the source system's resources.
- SSAS is well-suited for numerical analysis tasks.
- The data mining capabilities of SSAS can reveal hidden patterns that might not be immediately obvious.
- SSAS provides a comprehensive view of business data, including reporting, KPI analysis, scorecards, and data mining.
- With SSAS, you can perform OLAP analysis on data from various sources
- SSAS enables users to analyze data using a range of tools.

5.2 Report and dashboard systems

The BI tool, particularly PowerBI, can present a comprehensive overview of a business's performance, highlighting its advantages and disadvantages, and areas of concern or improvement, which decision-makers can easily understand. The next section will showcase the dashboard, focusing on six business requirements:

- Use the RFM model to determine customer segments.
- Assess sales performance by examining delivery efficiency in various regions.
- Recognize sales performance across both online and offline channels.
- Uncover sales performance patterns in different regions.
- Analyze customer purchasing trends for particular products using cross-selling techniques.
- Classify product items based on their significance using the ABC classification method.

5.3 Incremental refresh using Power Automate



The image shows two steps from a Power Automate flow. The first step, 'Recurrence', is configured with the following settings: Interval of 1, Frequency of Day, Time zone of (UTC+07:00) Bangkok, Hanoi, Jakarta, Start time of 2023-05-01T01:00:00Z, At these hours of 1, and At these minutes of 30. A preview box indicates it runs at 1:30 every day. The second step, 'Refresh a dataset', is configured with Workspace set to 'My Workspace' and Dataset set to 'Dashboard_BoKho'. A downward arrow connects the two steps.

| Recurrence | |
|---------------------------------------|-------------------------------------|
| * Interval | 1 |
| * Frequency | Day |
| Time zone | (UTC+07:00) Bangkok, Hanoi, Jakarta |
| Start time | 2023-05-01T01:00:00Z |
| At these hours | 1 |
| At these minutes | 30 |
| Preview Runs at 1:30 every day | |
| Hide advanced options | |

| Refresh a dataset | |
|-------------------|-----------------|
| * Workspace | My Workspace |
| * Dataset | Dashboard_BoKho |

Figure 5- 2 The flow of automated refresh a dataset

We have set up an automated process to refresh the Dashboard_BoKho dataset triggered at a specific time each day as Figure 5- 2 above. This process ensures that the data in the dataset is always up-to-date and displayed on the dashboard. This helps to minimize the time and effort required for repeated manual refresh tasks, thereby increasing efficiency and productivity.

5.4 Data analysis with Power BI

5.4.1 RFM model for customer segmentation

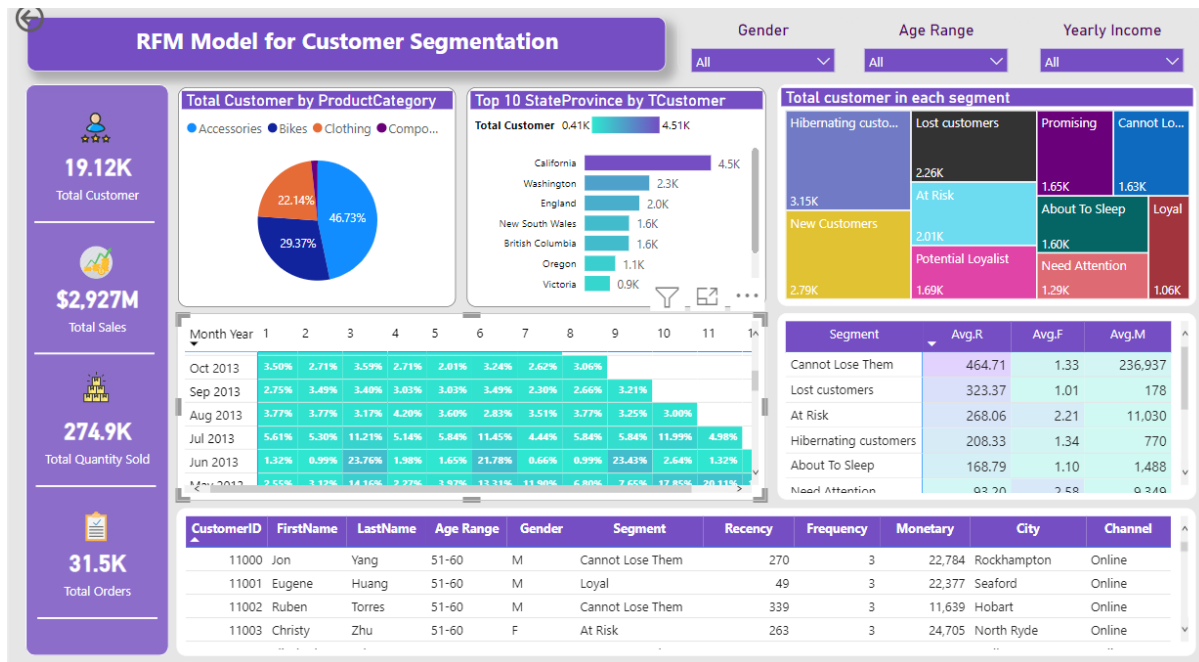


Figure 5- 3 General RFM model for customer segmentation

The dashboard in Figure 5- 3 illustrates customer segmentation by RFM model including 10 segments. On the left side of dashboard is KPI about Total customers, Total Sales, Total Quantity Sold and Total Orders of all customers. On the top right corner, they filter for this dashboard based on Gender, Age Range and Yearly Income of customers.

This project focuses on customers of each cluster which are divided in Table 5- 1 based on Recency, Frequency and Monetary Score and our business rules:

| Segment | Activity |
|--------------------|--|
| Loyal | Orders regularly. Responsive to promotions. |
| Potential Loyalist | Recent customers and spent a good amount. |
| New Customers | Bought most recently. |
| Promising | Potential loyalist a few months ago. Spends frequently and a good amount. But the last purchase was several weeks ago. |

| | |
|-----------------|---|
| Needs attention | Core customers whose last purchase happened more than one month ago. |
| About to sleep | They are existing customers who have not made any purchases for a long time. They have below-average recency, frequency, and monetary values. |
| Can't lose them | Made the largest orders, and often. But I haven't returned for a long time. |
| At Risk | Similar to "Can't lose them" but with smaller monetary and frequency value. |
| Hibernating | They haven't been back in a long time, and they haven't bought anything recently. |
| Lost | Made last purchase a long time ago and didn't engage at all in the last 4 weeks. |

Table 5- 1 The detail of customer segmentation

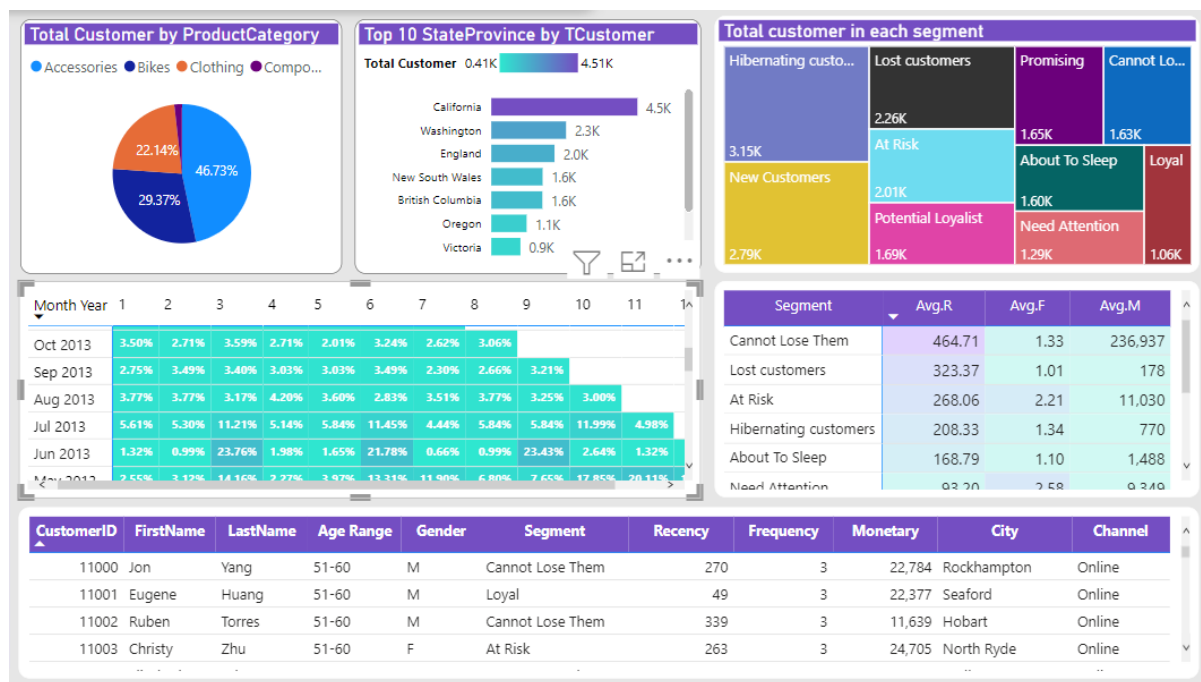


Figure 5- 4 Left side of general RFM model for customer segmentation

On the top left side, it can be seen that there are numbers of customers according to categories of products. Next to it is the top 10 state/province that have the most number of

2014. Our team will choose a specific country, in this case, Germany, because this country has some differences compared to others.

Our team will select fields to filter the data by country, year, and order type (channel).

At the top of the chart, there are 5 cards that show basic data such as:

- AVG Time Delivery (Days): The average delivery time of an order.
- Shipments: The total number of shipments per year.
- AVG Freight: The average delivery cost of an order.
- Freight/Order: The ratio of delivery cost to the value of an order. In this case, delivery costs will account for approximately 2.5% of the total value of an order.
- OnTime Delivery: The percentage of orders delivered early and on time.

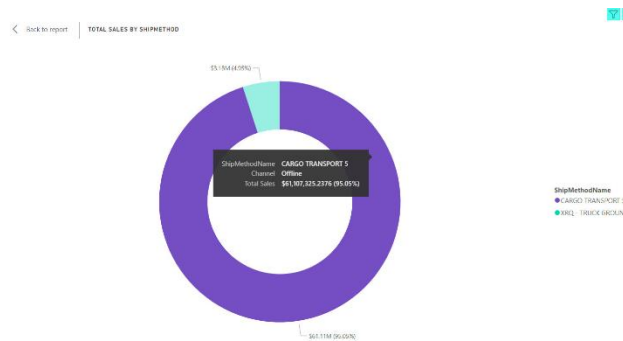


Figure 5- 6 Total order value by channel (delivery methods) - Germany (2013)

The Figure 5- 6 shows the most commonly selected delivery method. Overall, all 6 countries only choose 2 out of 5 delivery methods, which are Cargo Transport 5 and XRQ - Truck Ground. However, the majority of orders will choose to purchase offline (with the Cargo Transport 5 delivery method), and in the case of Germany, it accounts for up to 95.05% of the total value of orders in 2013. Meanwhile, online purchases with the XRQ - Truck Ground delivery method only account for 4.95%.

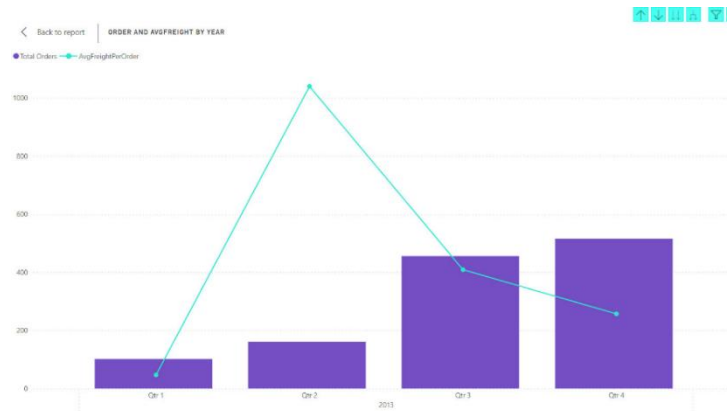


Figure 5- 7 Trend between number of orders and AVG Freight - Germany (2013)

It can be observed that when the delivery cost tends to increase, the trend of customer orders decreases significantly, and reversed. However, this may not always be the case, as this difference may be due to the timing of when customers place their orders, such as during their country's holidays, family members' birthdays, or the end of the year.

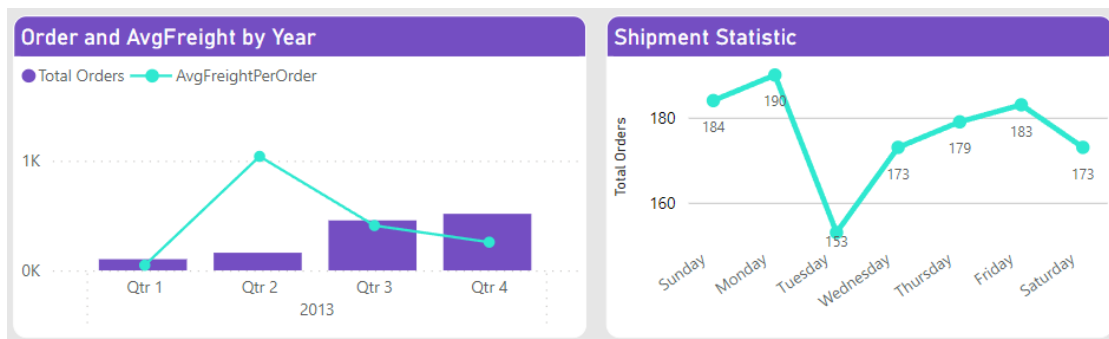


Figure 5- 8 Fluctuating trend of weekly deliveries - Germany (2013)

We also analyzed the total number of shipments per week for each country. Overall, it can be observed that the delivery demand is highest on weekdays and decreases towards the end of the week, similar to the case of Germany (2013) as mentioned.

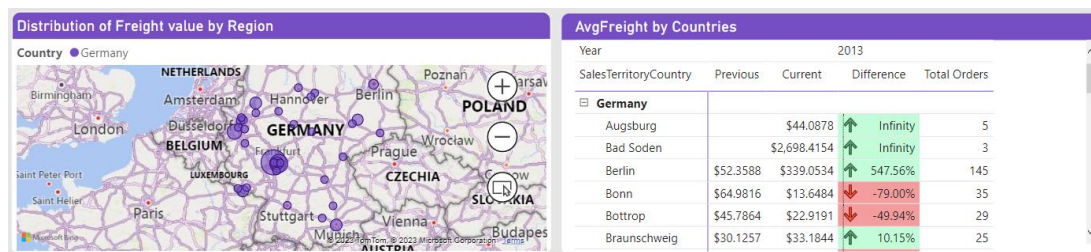


Figure 5- 9 AVG Freight by Region among years - Germany (2013)

In the last row of charts, the first chart displays the average delivery cost for each city in each country, and the remaining chart compares the average delivery costs of each country to the previous year. These charts can help businesses identify consumer behavior and track changes in customer purchasing due to delivery services.

5.4.3 Sales performance

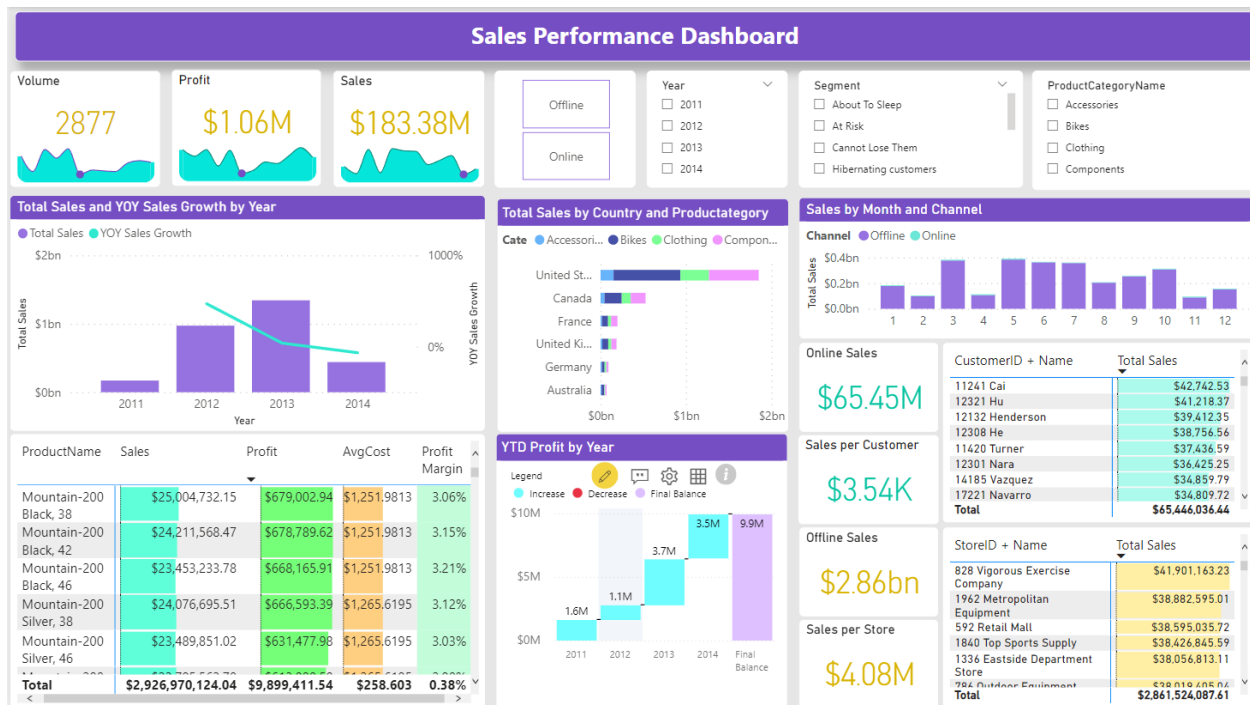


Figure 5- 10 General sales performance

This Figure 5- 10 will illustrate the sales performance of countries from 2011 to 2014 through the correlation between metrics such as sales revenue, profit...

Our group will select the following fields to filter the data: Year, Channel, Segment, ProductCategoryName.

As can be seen in the top row, we have 3 cards representing, respectively:

- Volume: Represents total orders.
- Profit: Represents the company's sales profit.
- Sales: Represents the company's sales revenue.

The first chart will display the total sales and year-over-year sales growth of countries for each year. This chart will help businesses track sales performance, identify the years with the lowest and highest sales, and determine the trend in sales growth.

As seen in the example above, total sales increased gradually from 2011 and reached its peak in 2013, followed by a sharp decline in 2014. Therefore, in 2014, the YOY sales growth value is < 0 .

The next chart compares the total sales of countries for each product category. Through this, businesses can evaluate which products are most popular and which countries prefer them the most. This information is useful in marketing and sales strategies for businesses, such as providing incentives and pricing to maintain competitiveness.

It can be seen that there is a significant difference between countries and the United States. The United States plays an important role in the company's total revenue for all four product categories, while the market in other countries is still relatively low.

The sales by month and channel chart compares the sales performance of a business between two channels (offline and online) for each month. Overall, the business achieved the highest sales revenue in the months of May, June, and July (summer season) through offline orders. This information can help the business focus on building an appropriate marketing strategy to increase sales revenue for both channels.

The next chart compares the total order value, average cost, and profit margin of each product. Through this comparison, businesses can determine which specific products are generating the highest profits, the amount of costs required to produce each product, and which products may need to be adjusted in their business plan.

In the specific example above:

- Top revenue generator: Mountain-200 Black, 38 product.
- Top profit generator: AWC Logo Cap product.
- Top cost generator: Road-150 Red, 44 product.

The YTD profit by years chart shows the profit growth of a business for each year compared to the previous year. The chart provides businesses with information about the financial balance of the company during each period. In the chart above, it can be seen that

the profit has been increasing each year. This information can help businesses understand the situation and make adjustments to meet their financial goals.

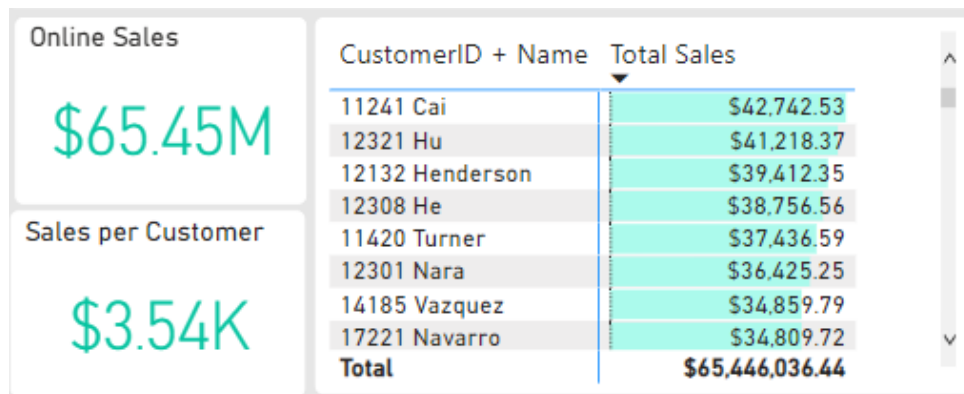


Figure 5- 11 The top customer had high sales with online channel

The Figure 5- 11 will compare the total sales of each customer, from which the business will know which customer is generating the most revenue for the company and which customers need to receive more special offers.

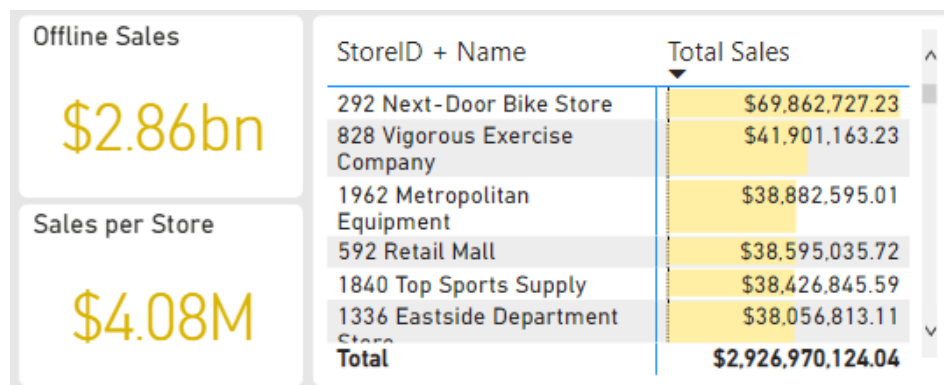


Figure 5- 12 The top store had high sales with offline channel

Similarly, this Figure 5- 12 will show which stores are bringing in the most revenue for the business. Through this, businesses can make informed decisions about which stores to prioritize and how to allocate resources to optimize their sales process.

In the example above, the 292 Next-Door Bike Store in the United States is ranked first with the highest total sales of over \$69, while the 1904 Mobile Outlet store ranks last with total sales of only roughly \$1.5.

5.4.4 Product sales analysis

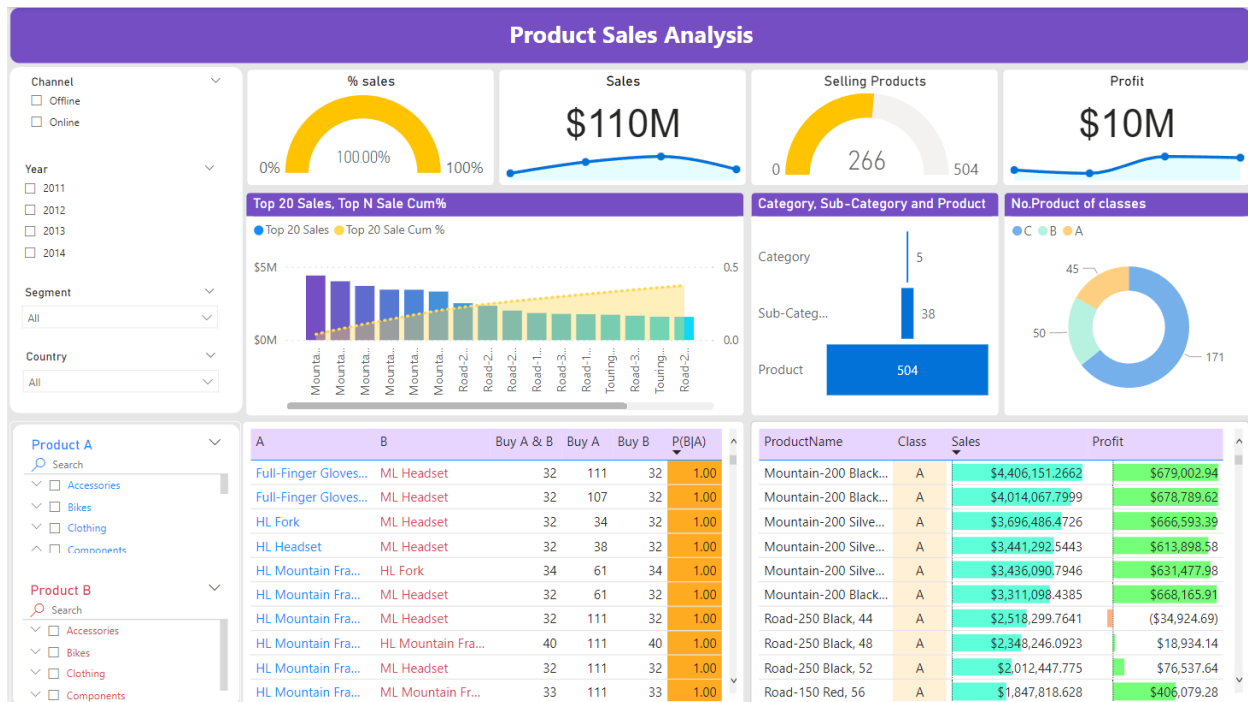


Figure 5- 13 General product sales analysis

The "Product Sales Insight" dashboard provides an overview of the business performance by product for the enterprise that is shown as Figure 5- 13 above.

We utilize 4 KPIs, which are:

- “% Sale” : that reflect the contribution rate to revenue from sales of products sold.
- Sale: Revenue from selling products
- Selling Products: The number of product types that have been sold
- Profit : Profit from selling products

Channel
☐ Offline
☐ Online

Year
☐ 2011
☐ 2012
☐ 2013
☐ 2014

Segment
All

Country
All

Figure 5- 14 Filter of General product sales analysis

To obtain a more detailed view, we have designed 4 filters: Channel, Year, Customer Segmentation, and Region..

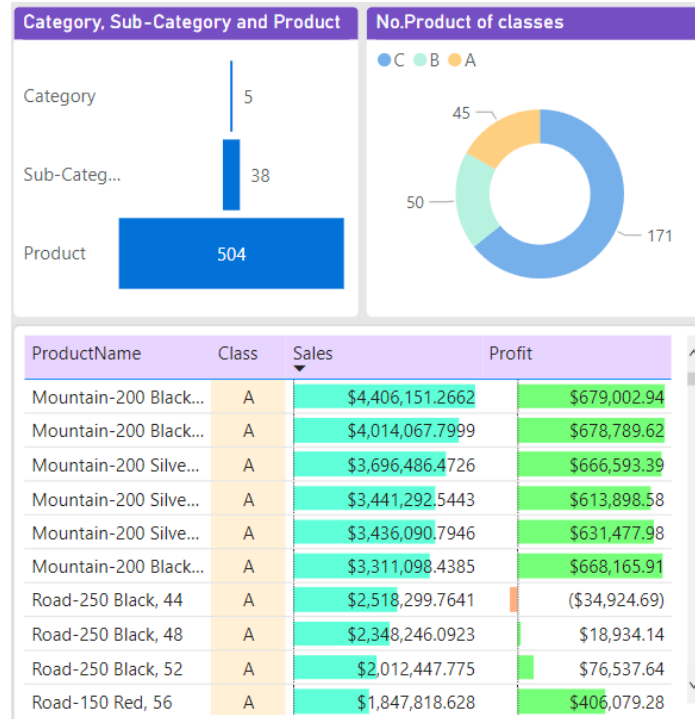


Figure 5- 15 ABC class of dashboard Product Sales Insight

We also use the ABC classification method to demonstrate the revenue contribution level of products. Specifically, the products are divided as follows:

- Class A: The products with the highest revenue and account for 70% of total revenue.
- Class B: The products with medium revenue and account for 20% of total revenue.
- Class B: The products that contribute low revenue and account for 10% of total revenue.

Two bar charts are used in the dashboard to respectively show the difference in revenue between products and categories. Furthermore, to answer the question of which products are often purchased together, we use the correlation method to calculate the probability of customers buying pairs of products together, thereby serving the purpose of cross-selling.

5.4 Recommend business strategies

5.4.1 Marketing strategies for 3 main segments

Loyal: To retain loyal customers, motivate them to purchase higher-value items and request them to leave feedback. Give them a customized onboarding experience to help them familiarize themselves with goods or services, such as tutorials, demos, and other useful resources. Grant them early access to new products or services to make them feel appreciated and unique.

New customers: Assist them through the onboarding process and foster a connection with them early on. Introduce a loyalty program or rewards system to encourage repeat business. Personalized product recommendations based on their browsing or purchase history to demonstrate that their needs are understood. Encourage them to leave reviews or feedback to enhance social proof and attract potential customers. Upsell higher-value products or services to increase their long-term value to the business.

Hibernating customers: To win back hibernating customers, the business can use various tactics, such as providing them with incentives that motivate them to come back. Another approach is to offer personalized recommendations based on their past purchase behavior. This can show them that the company values their business and understands their preferences, which can help rekindle their interest in products or services.

5.4.2 Optimize delivery performance

Review and streamline the entire delivery process from order fulfillment to delivery tracking. Look for ways to reduce delivery times and optimize the route taken by delivery personnel so that the average delivery time will decline and enhance customer satisfaction. Implement delivery management software to automate and optimize the delivery process. This will help reduce human error and speed up the entire delivery process.

5.4.3 Enhance sales performance

The sales figure of online sales channels cause disappointment compared to offline sales channels. Therefore, making use of social media channels like Facebook, Instagram, Twitter, and LinkedIn to market products and interact with prospective clients. Share

attractive content like product images, reviews, and other engaging material. Also, taking advantage of targeted advertising on social media platforms and search engines to reach potential customers who are more likely to be interested in products. This approach will help improve the efficiency of advertising efforts. Additionally, most of the sales revenue comes from the United States and Canada while products are sold in many countries. Perform comprehensive market research to determine the reasons for the low sales in other countries. This will allow us to comprehend the local customers' preferences and needs as well as the competition in the area. Providing discounts and promotions may be an effective way to entice customers who may be hesitant to make a purchase. This could include discounts on specific products or services, free shipping, or special promotions.

5.4.4 Product-based strategy

A-class items generate the highest revenue, it is advisable to give priority to marketing efforts toward them. Identifying supplementary products that complement A-class items, businesses can leverage cross-selling tactics to promote them sold with A-class items. Cross selling strategy not only helps the company with sales increase, but customer loyalty will also improve. For example, if a customer has purchased a high-end Bike - Mountain Bike, the business could promote complementary items such as Component - Mountain Frames (an A-class item). Despite having the most number of products, the C-class generates the least revenue for the company, so the company focuses only on promoting the most important products within this class.

CHAPTER 6: CONCLUSION AND FUTURE WORKS

6.1. Results

The project team was able to acknowledge the process of implementing a Business Intelligence solution and successfully building it for the project. The team produced several deliverables, including a data warehouse, dashboards using PowerBI. The data warehouse was built to meet the sales team's requirements, and dashboards are utilized to gain deeper and informative insights about delivery, personnel performance and customer segmentation . Eventually, they help to visualize the company's performance and support decision-making for managers. Moreover, the team also manages to make ETL process automatically from sources to data warehouse and from data warehouse to visualization tools.

6.2. Limitations

However, there are some limitations during the process of conducting the project, including the fact that missing data made the ETL process time-consuming and inefficient. The team also struggles with Power BI applications because we do not use it frequently in other subjects. Additionally, the team does not have previous experience in implementing a Business Intelligence solution before, so we have to cope with challenges about knowledge, time and skills. Luckily, this is also an opportunity to improve the team's soft skills and academic knowledge.

6.3. Future works

In the future, we will deal with other business requirements and broaden the data warehouse structure that collects data from more sources. This project also integrates with Python language to produce more models about forecasting and propose useful insights for business. Additionally, we are going to utilize various visualization tools such as Tableau, Qlik to have diverse forms of dashboards.

REFERENCES

- [1] T. O. Josh Fruhlinger, "What is business intelligence? transforming data into business insights," 20 Jan 2023. [Online]. Available: <https://www.cio.com/article/272364/business-intelligence-definition-and-solutions.html>.
- [2] Davidiseminger, "BI solution architecture in the Center of Excellence - Power BI," 27 Feb 2023. [Online]. Available: <https://learn.microsoft.com/en-us/power-bi/guidance/center-of-excellence-business-intelligence-solution-architecture>.
- [3] "The Benefits of Business Intelligence... Why Do Organizations Need BI?," 08 Feb 2021. [Online]. Available: <https://bi-survey.com/benefits-business-intelligence>.
- [4] B. Calzon, "Bi roadmap to build a thriving business intelligence strategy," 02 Sep 2022. [Online]. Available: <https://www.datapine.com/blog/roadmap-to-a-successful-business-intelligence-strategy/>.
- [5] "What is ETL (extract, transform, load)?," [Online]. Available: <https://www.ibm.com/topics/etl>.
- [6] Stephen, "What is ETL & Why is it important for business decision making?," 31 Jan 2022. [Online]. Available: <https://blog.hubspot.com/website/what-is-etl>.
- [7] "What is ETL (extract, transform, load)? definition, process, and Tools," [Online]. Available: <https://www.talend.com/resources/what-is-etl/>.
- [8] "ETL process in Data Warehouse," 02 Feb 2023. [Online]. Available: <https://www.geeksforgeeks.org/etl-process-in-data-warehouse/>.
- [9] R. a. D. D. a. T. E. Sharda, "Business intelligence," *Analytics, and Data Science*, 2018.
- [10] Caldwell, "Top sales kpis to track," 27 Jan 2023. [Online]. Available: <https://www.netsuite.com/portal/resource/articles/accounting/sales-kpis.shtml>.

- [11] Jyothisuri, "Overview of OLAP cubes for advanced analytics | Microsoft Learn," 16 Mar 2023. [Online]. Available: <https://learn.microsoft.com/en-us/system-center/scsm/olap-cubes-overview>.
- [12] "Cloud computing services: Microsoft Azure," [Online]. Available: <https://azure.microsoft.com/en-us>.
- [13] "Take care of what's important. automate the rest," [Online]. Available: <https://powerautomate.microsoft.com/en-us/>.
- [14] O. M. Editorial, "Importance of sales in any organization," 02 Jan 2023. [Online]. Available: <https://www.onlinemanipal.com/blogs/importance-of-sales-in-organization>.
- [15] "Data dictionary AdventureWorks," [Online]. Available: https://dataedo.com/samples/html/AdventureWorks/doc/AdventureWorks_2/home.html.
- [16] L. P. L. a. K. R. a. S. S. a. J. K. a. B. G. a. C. M. K. a. R. R. a. R. S. S. Cavaliere, "The Impact of customer relationship management on customer satisfaction and retention: The mediation of service quality," *Turkish Journal of Physiotherapy and Rehabilitation*, vol. 32, no. 3, pp. 22107--22121, 2021.
- [17] "Data Visualization: Microsoft power bi," [Online]. Available: <https://powerbi.microsoft.com/en-us/>.
- [18] "What is tableau?," [Online]. Available: <https://www.tableau.com/why-tableau/what-is-tableau>.

APPENDIX

1. Source Code: https://github.com/trannhatnguyen2/K20406C_BoKho

TASK ASSIGNMENT TABLE

| Full name | Task | Completion rate | % Contribution |
|-------------------------------|---|------------------------|-----------------------|
| Tran Nhat Nguyen | <ul style="list-style-type: none"> - Assignment and progress monitoring of project implementation. - Experimenting with loading data using SSIS, SSAS. - Set up a flow that automates the entire process. - Support completion of report content: all chapters. - Aggregating content from Word, dashboard, source code (SSIS, SSAS, SQL). | Good | 25% |
| Man Dac Sang | <ul style="list-style-type: none"> - Experimenting with loading data using SSIS, SSAS. - Support completion of report content: all chapters. - Visualize product sales analysis. | Good | 25% |
| Nguyen Thai Ngoc Suong | <ul style="list-style-type: none"> - Support completion of report content: all chapters. - Visualize customer segmentation. | Good | 25% |
| Thai Thien Truc | <ul style="list-style-type: none"> - Support completion of report content: all chapters. - Visualize sales performance, delivery performance. - Slide design. | Good | 25% |

Table o- 1 Task Assignment

TURNITIN CHECKING

The screenshot displays the Turnitin interface for a document check. The document is titled "BUSINESS INTELLIGENCE AND DECISION SUPPORT SYSTEM" and is from the "VIETNAM NATIONAL UNIVERSITY HOCHIMINH CITY UNIVERSITY OF ECONOMICS AND LAWS". The topic is "APPLYING BUSINESS INTELLIGENCE TECHNOLOGY AND AUTOMATION PROCESSES TO MAKE MORE ACCURATE DECISIONS FOR SALES OPERATIONS". The lecturers are Nguyen Van Ho, M.S. and Le Ba Thien, M.S., and the group is BoKho. The document is dated Ho Chi Minh, May 20. The Turnitin logo is in the top left, and the user name "Nguyễn Trần Nhật" and document ID "K20406C_BoKho_Check" are in the top right. A "Match Overview" panel on the right shows a total match of 4% and a list of four matches: 1. www.coursehero.com (3%), 2. Submitted to Universid... (1%), 3. Submitted to University... (1%), and 4. Submitted to Florida In... (1%). The bottom status bar shows "Page: 1 of 87", "Word Count: 14071", and "Text-Only Report" and "High Resolution" options.

turnitin

Nguyễn Trần Nhật K20406C_BoKho_Check

Match Overview

4%

Match 1 of 5

| | | | |
|---|----------------------------|----|---|
| 1 | www.coursehero.com | 3% | > |
| 2 | Submitted to Universid... | 1% | > |
| 3 | Submitted to University... | 1% | > |
| 4 | Submitted to Florida In... | 1% | > |

Page: 1 of 87 Word Count: 14071 Text-Only Report High Resolution On

Figure o- 1 Result of Turnitin Checking with 4% match overview