

**VIETNAM NATIONAL UNIVERSITY HOCHIMINH CITY  
UNIVERSITY OF ECONOMICS AND LAW**



**FINAL REPORT:**

**BUILDING BUSINESS INTELLIGENCE  
SOLUTIONS FOR SALES DEPARTMENT OF  
ADVENTURE WORKS COMPANY**

**DATA ANALYSIS IN BUSINESS**

**Class: 232MI1701**

**Lecturer: Ph.D. Le Ba Thien**

**Ho Chi Minh City, January 4th, 2024**

**GROUP 2**

<b>NO.</b>	<b>Full name</b>	<b>Student Id</b>	<b>Role</b>
1	Vu Thi Phuong Anh	K204110558	Team leader
2	Dinh Thi Thuy An	K204111762	Member
3	Nguyen Thi Thanh Binh	K204111767	Member
4	Le Thi Thuy Linh	K204111780	Member
5	Mai Le Ngoc Tram	K204110587	Member

## **ACKNOWLEDGEMENTS**

Group 2 would like to express our sincere gratitude to Mr. Le Ba Thien for providing a strong knowledge base, countless comments and suggestions to help the team finish the project effectively, and solutions to challenging issues as they came up, supporting us in completing the final project of the course "Data Analysis in Business

Despite the project's short timeline, our team made an effort to brainstorm and do our best to meet the initial criteria. Nevertheless, challenges remained to be addressed, and errors were not averted. So that we can grow even more and learn from our mistakes, we hope that our professors will be able to review and comment on our project.

Once again, We would like to express our gratitude to Mr. Thien who has helped us during the course and for the knowledge and skills we have acquired.

Sincerely,  
Group 2

## **COMMITMENT**

The team guarantees that all regulations will be adhered to during the project's implementation and that the data and outcomes included in the report are accurate. All references from the Internet, books, and teaching materials will be specifically cited.

The team is willing to take full responsibility for any mistakes made and to embrace all forms of sanctions.

Sincerely,

Group 2

## LIST OF ACRONYMS

<b>BI</b>	Business Intelligence
<b>DW</b>	Data Warehouse
<b>ETL</b>	Extract - Transform - Load
<b>ELT</b>	Extract - Load - Transform
<b>KPI</b>	Key Performance Indicator
<b>OTLP</b>	Online Transaction Processing
<b>OLAP</b>	Online Analysis Processing
<b>SQL</b>	Structured Query Language
<b>SSAS</b>	SQL Server Analysis Services
<b>SSIS</b>	SQL Server Integration Services

## LIST OF FIGURES

Figure 2-1. Data governance layer in BI architecture .....	6
Figure 2-2. Overview of ETL Process.....	8
Figure 2-3. Star schema.....	12
Figure 2-4. Snowflake Schema.....	13
Figure 2-5. Galaxy Schema .....	14
Figure 3-1. BI Solution.....	22
Figure 4-1. Data warehouse diagram.....	32
Figure 4-2. Data Flow Task for DimGeography .....	35
Figure 4-3. Data Flow Task for DimTerritory.....	36
Figure 4-4. Data Flow Task for DimProduct.....	37
Figure 4-5. Data Flow Task for DimCustomer .....	38
Figure 4-6. Data Flow Task for DimPromotion .....	39
Figure 4-7. Data Flow Task for DimSalesReason.....	39
Figure 4-8. Data Flow Task for DimDate .....	40
Figure 4-9. Script to create “Date” column .....	40
Figure 4-10. Derive new columns for DimDate.....	41
Figure 4-11. Data Flow Task for FactSales.....	41
Figure 4-12. Overview of the Sales Performance Dashboard .....	42
Figure 4-13. Overview of the Product Detail Dashboard.....	45
Figure 4-14. Total Revenue and Total Orders by Month in 2013 .....	46
Figure 4-14. The top best-selling products by Region .....	47
Figure 4-15. Customer Detail Dashboard.....	48

## LIST OF TABLES

Table 4-1. Bus Matrix.....	23
Table 4-2. The master data .....	24
Table 4-3. The transaction data .....	25
Table 4-4. Fact and Dim Table Overview .....	25
Table 4-5. Fact Table.....	26
Table 4-6. Dimension Tables .....	27
Table 4-7. Description of key relationship in SalesDW .....	32

## TABLE OF CONTENTS

ACKNOWLEDGEMENTS .....	1
COMMITMENT .....	2
LIST OF ACRONYMS .....	3
LIST OF FIGURES .....	4
LIST OF TABLES .....	5
CHAPTER 1: INTRODUCTION .....	7
<b>1.1. Business Case</b> .....	7
<b>1.2. Objectives</b> .....	8
<b>1.3. Research objects</b> .....	8
<b>1.4. Project scope</b> .....	8
<b>1.5. Research process</b> .....	9
<b>1.6. Structure of the report</b> .....	9
CHAPTER 2: THEORETICAL BACKGROUND .....	11
<b>2.1. Overview about BI</b> .....	11
<b>2.2. ETL Process Overview</b> .....	13
<b>2.3. Data Warehouse and Data Mart</b> .....	15
<b>2.4. Schema Types</b> .....	16
<b>2.5. KPIs</b> .....	19
CHAPTER 3. REQUIREMENTS ANALYSIS AND EXPERIMENTAL MODELING ...	20
<b>3.1. Business Context of Sales Department</b> .....	20
<b>3.2. Business Challenges</b> .....	22
<b>3.3. Business Requirements</b> .....	23
<b>3.4. Proposing BI Solution</b> .....	25
CHAPTER 4. EXPERIMENTAL RESULTS AND ANALYSIS .....	27
<b>4.1. Result Structure</b> .....	27
<b>4.2. ETL Process</b> .....	38
<b>4.3. Data Analytics</b> .....	44
CHAPTER 5. CONCLUSION .....	54
<b>5.1. Results</b> .....	54
<b>5.2. Future work</b> .....	54
REFERENCE .....	56
PROJECT MEMBERS EVALUATION .....	57



## **CHAPTER 1: INTRODUCTION**

### **1.1. Business Case**

Adventure Work (AW) specializes in the world of bicycles. They make and sell a wide variety of bikes, with over 97 brands across three main types: mountain, road, and touring. AW takes pride in building some of their own components, but also partners with trusted vendors for additional parts and accessories. Whether you're looking for a thrilling mountain trail adventure, a smooth road ride, or a comfortable journey across scenic landscapes, AW has everything you need to get rolling, including clothing and gear.

Adventure Work is a global leader in the world of bicycles, reaching customers across the United States, Canada, Australia, the United Kingdom, France, and Germany through the power of the internet. They cater to both individual riders and bicycle retailers, ensuring everyone has access to their extensive range. To achieve this, Adventure Work utilizes a team of dedicated professionals working across various departments, including sales, production, purchasing, engineering, finance, technology, marketing, logistics, and research & development.

In this day of advanced technology assisting in tracking and making business decisions, AW wants to build a system to get the insight from their daily observed operating data, especially sale activities. They have recorded its sale activities via a transactional database, but it is raw and needs to be processed before being used for analytics. And they have to process these data before each time using and process. They also preferred automatic reports instead of their current basic statistical reports.

They want the system to automatically get and process data into a dashboard that shows the sale performance and some common sale metrics. AW expects this system to save time and labor, and be easy to control and use. Any tool is acceptable within the planned budget and time for getting ready.

The demo system should be completed by the beginning of January 2023 for testing and evaluation before being officially implemented in 2024.

## **1.2. Objectives**

### **1.2.1. General objectives**

Our team will build a project on a Business Intelligence solution to support the sales department of Adventure Work through sales activities. In particular, we will build a system that provides this organization with a big-picture view of their business through many key performance indicators, helps the organization identify their business problems and from there, easily make informed decisions.

### **1.2.2. Specific objectives**

- To research the Adventure Work database, then select necessary data serving for sales analysis.
- To identify the business problems and establish the Key Performance Indicators (KPIs) to measure achieved results.
- To build a Business Intelligence system with a data warehouse and ETL process.
- To implement ETL Process to preprocess data before loading it into a data warehouse.
- To use Power BI tools to create dashboards for effective data representation. From there, identify these challenges and propose suitable solutions for business.

## **1.3. Research objects**

- AdventureWorks2022 is one of the sample databases used in Microsoft SQL Server and Microsoft Azure SQL Database. The AdventureWorks sample database includes a variety of objects such as products, customers, orders, employees, and other data to illustrate real-life scenarios in a business environment.
- Sales business process
- KPI
- ETL process
- Visual studios
- Power BI

## **1.4. Project scope**

The scope of the project's research is to comprehend Adventure's sales model and develop a business intelligence solution for it. Work with sales income, profit, and

agent data on a worldwide basis. A global corporation that produces and promotes bicycles composed of metal and composite materials in North America, Europe, and Asia.

General sales report such as:

- Report on Sales Performance
- Report on Sales Products Detail
- Report on Customer Detail

Tool used:

- Power BI
- SQL Server
- Microsoft Visual Studio (SSIS)

## **1.5. Research process**

The research process for this project began with (1) an introduction to the topic, establishing the rationale for choosing Business Intelligence (BI) and setting clear objectives and expected results. This was followed by (2) an in-depth exploration of the theoretical foundations of BI, including the ETL process, data warehousing, schema types, and KPIs. The project then proceeded to (3) a detailed analysis of business and IT requirements, utilizing a comparative approach to select appropriate BI and data visualization tools. The practical phase involved (4) building a data warehouse, integrating various data types, and focusing on the ETL process. The project (5) culminated in the analysis and visualization of data, using sophisticated reporting and dashboard systems. The conclusion (6) summarized the project's achievements and limitations, offering insights into potential future research directions. This process seamlessly integrated theoretical understanding with practical application, ensuring a comprehensive exploration of BI in a sales system context.

## **1.6. Structure of the report**

### ***Chapter 1: Introduction***

This chapter introduces the project, explaining the rationale behind choosing the topic, outlining the primary objectives, and setting expectations for the results. It provides a

glimpse into the overall structure and scope of the project, setting the stage for a detailed exploration of the subject matter.

### ***Chapter 2: Theoretical Background***

In this chapter, foundational theories and concepts related to Business Intelligence (BI) are presented. It covers an overview of BI, the ETL (Extraction, Transformation, Loading) process, definitions and distinctions between data warehouses and data marts, an introduction to Schema types, and a discussion on Key Performance Indicators (KPIs).

### ***Chapter 3: Requirements Analysis and Experimental Modeling***

This chapter focuses on analyzing the business processes related to the Sales system and the data sources utilized. It delves into a detailed analysis of business requirements, including goals, inputs, outputs, and KPIs. The chapter also includes a comparative analysis of various BI and data visualization tools, setting the stage for experimental modeling.

### ***Chapter 4: Experimental Results and Analysis***

Chapter 4 is dedicated to the practical aspects of building a data warehouse and integrating data. It discusses master data and transaction data, explains the structure and relationships within the data warehouse model, and provides an overview of the ETL process. The chapter culminates with the analysis and visualization of the data, highlighting the use of reporting and dashboard systems.

### ***Chapter 5: Conclusion***

The final chapter summarizes the project's results, including the achievements and any limitations or constraints faced. It reflects on the project's impact and effectiveness in meeting its objectives and suggests potential areas for future research and development to extend the scope or address any shortcomings of the current project.

## CHAPTER 2: THEORETICAL BACKGROUND

### 2.1. Overview about BI

#### 2.1.1. What is BI?

Business intelligence (BI) is a collection of techniques and tools used by businesses to evaluate data and turn it into insights that can be used to make tactical and strategic choices. BI technologies give users comprehensive information about the status of the business by gaining access to, analyzing, and presenting data sets and analytical results in reports, summaries, dashboards, graphs, charts, and maps.

Insights into the present state of an organization, derived from existing data, can be quickly and easily accessed with a variety of technologies that are together referred to as business intelligence.

#### 2.1.2. BI Architecture

Business intelligence (BI) architecture is the infrastructure that a business sets up to support every step of the business intelligence (BI) process, from gathering and cleaning data to organizing, storing, and analyzing it to delivering reports and dashboards and operationalizing insights.

BI architecture is a component of enterprise architecture, which specifies the business processes (how the firm operates) and IT environment (its structure) in order to help the organization meet its tactical and strategic objectives.

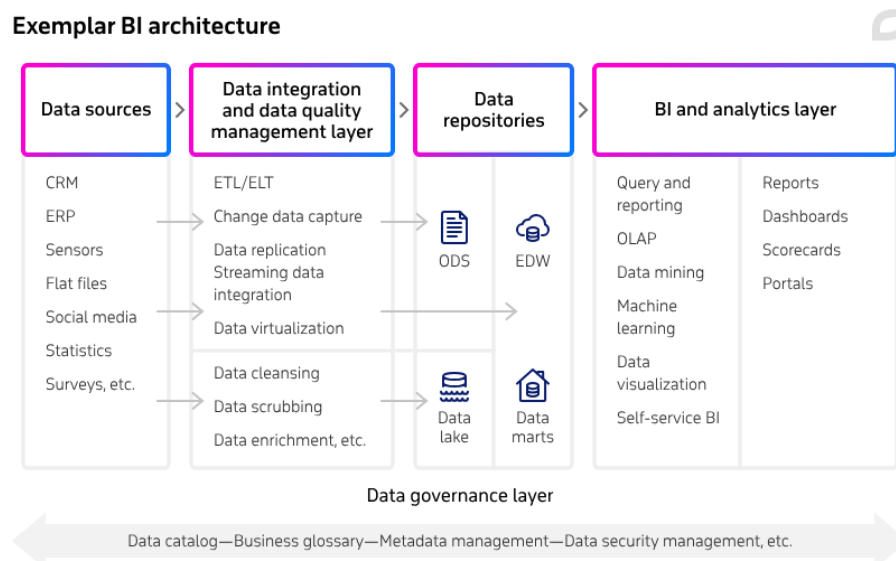


Figure 2-1. Data governance layer in BI architecture

(Source: itransition.com)

*Data Source:* A data source is anything that generates the digital information that business intelligence systems need to function. There are two types of data sources: internal and external. Internal data sources store and capture information generated within an organization.

*Data integration and data quality management layer:* Consolidating datasets from many sources for a single picture is the goal of the second phase in the business intelligence process. This makes the information usable for analytics and operational needs. The kind, format, and volume of the information, as well as the intended use (operational reporting, business analysis, machine learning use cases, etc.), will determine which of the numerous data integration techniques is used.

*Data repositories:* This part consists of different repositories that organize and hold data in preparation for additional processing. There are two main categories of data repositories:

- + Enterprise data warehouse – a single repository with aggregated, cleansed data. Different kinds of databases are used by businesses for this purpose: multidimensional (stores data in a format similar to a data cube), relational (stores data in rows), and columnar (stores data in columns)
- + Data marts – storage repositories designed with certain user groups' (departments, divisions, etc.) analytics and reporting requirements in mind. Depending on how a data warehouse is constructed, they could serve as self-consistent analytics repository or be a supplement to an EDW.

*BI and analytics layer:* Solutions for gaining access to and utilizing data are included in this layer, which is intended for business users, data scientists, and analysts. This layer naturally reflects the organization's goals for data analytics and BI maturity: while some businesses can get by with just descriptive and diagnostic analytics, others require more in-depth studies supported by ML and AI through a self-service user interface.

*Data governance layer:* This component is inextricably linked to the other four because its primary goal is to monitor and manage the entire BI process. By implementing data governance rules and regulations, an organization can regulate who has access to the information, how it is accessed, whether the data is properly protected, and other factors. A data management program consisting of all these standards and policies can be automated using data governance tools.

## 2.2. ETL Process Overview

### 2.2.1. What is ETL?

Besides ETL, ELT is also a common technique for this work. ELT was developed later than ETL, and ETL is the base of ELT. Some differences between ETL and ELT make us choose ETL for this project:

- ETL is suitable for all businesses using SaaS solutions, while ELT is not suitable for small and medium businesses.
- Through ETL, the time to transform data is dependent on the size of data, while ELT makes it longer to wait for large sizes of data.
- ETL is best for unstructured and non-relational data within large volumes, while ELT is best for relational and structured data within small to medium amounts of data.
- As the AW sample dataset is quite clean and in the table format, ETL would be more suitable than ELT.

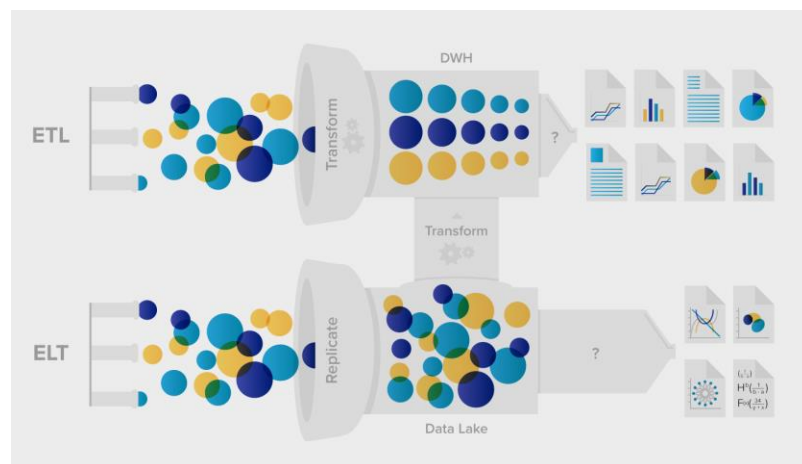


Figure 2-2. Overview of ETL Process

(Source: Datacamp)

### 2.2.2. ETL Process

The ETL (Extract, Transform, Load) process can be described in details:

#### a) Extract

- Purpose: In this initial step, data is extracted from various source systems such as databases, APIs, flat files, or web services.
- Methods:
  - + Full Extraction: All data is pulled from the source system.
  - + Incremental Extraction: Only new or modified data since the last extraction is retrieved.
  - + Change Data Capture (CDC): Captures changes (inserts, updates, deletes) in real-time.
- Challenges:
  - + Ensuring data consistency and accuracy during extraction.
  - + Handling different data formats and structures.

#### b) Transform

- Purpose: Data undergoes transformation to make it suitable for analysis and reporting.
- Tasks:
  - + Cleansing: Removing duplicates, correcting errors, and handling missing values.
  - + Aggregation: Summarizing data (e.g., calculating totals, averages).
  - + Enrichment: Adding calculated fields or merging data from multiple sources.
  - + Data Quality Checks: Verifying data integrity.
- Tools: ETL tools (e.g., Informatica, Talend, SSIS) facilitate transformations.
- Challenges:
  - + Complex business rules and logic.
  - + Performance optimization.

#### c) Load



- Purpose: Transformed data is loaded into the target data warehouse.
- Types:
  - + Staging Area: Temporary storage for transformed data.
  - + Data Warehouse: Central repository for structured data.
  - + Data Marts: Subset of data warehouse for specific business areas.
- Methods:
  - + Bulk Load: Loading large volumes efficiently.
  - + Incremental Load: Updating only changed records.
  - + Real-time Load: Immediate data availability.
- Challenges:
  - + Ensuring data consistency across tables.
  - + Handling data concurrency and locking.
- Post-Load Activities:
  - + Indexing: Creating indexes for faster querying.
  - + Partitioning: Dividing tables into smaller segments.
  - + Metadata Management: Documenting data lineage and definitions.
  - + Monitoring and Maintenance: Regular checks for data quality and performance.

## **2.3. Data Warehouse and Data Mart**

### **2.3.1. Definitions**

A data warehouse is a central repository of integrated data from one or more disparate sources. It stores current and historical data and is used for creating analytical reports for knowledge workers throughout the enterprise. The primary goal of a data warehouse is to provide a single source of truth for an organization's data, allowing for informed decision-making and strategic planning.

On the other hand, a data mart is a subset of a data warehouse that is designed to serve a specific business line or department. Data Marts are smaller in scope and focus on providing data for a particular set of users, such as sales, marketing, or finance teams.

They are often created to address the specific needs of a particular group within an organization, allowing for more targeted and efficient access to relevant data.

### **2.3.2. Approaches**

When it comes to implementing a data warehouse, there are two main approaches: the top-down approach and the bottom-up approach. The top-down approach involves designing and building a single, centralized data warehouse that serves the entire organization. This approach requires significant upfront planning and investment but can provide a comprehensive and consistent view of the organization's data. Inmon's approach to building a data warehouse is a top-down approach that designs centralized storage first and then creates data marts from the summarized data warehouse data and metadata. This approach focuses on integrating data from various sources into a single, consistent data model.

Conversely, the bottom-up approach involves building individual data marts for specific business units and then integrating them into a larger data warehouse as needed. This approach allows for more agile and incremental development, as well as the ability to quickly address the specific needs of different departments. However, it may lead to inconsistencies and redundancies if not carefully managed. Kimball's approach to building a data warehouse is a bottom-up approach that focuses on creating specialized data marts first and then developing a data warehouse database incrementally from independent data marts. This approach focuses on providing understandable and flexible information to business users.

Whether taking a top-down or bottom-up approach, the ultimate goal is to provide reliable, accessible, and actionable data for driving business success.

## **2.4. Schema Types**

### **2.4.1. Star Schema**

The Star Schema is a straightforward database schema characterized by a central fact table connected to several dimension tables. Each dimension table has a direct link to

the fact table. The schema's name, "star," stems from its diagrammatic resemblance to a star, where points radiate from a central point. The fact table houses key business metrics, such as sales figures or transaction counts. In contrast, the dimension tables generally contain descriptive attributes pertinent to the fact data. This configuration is engineered for enhanced query performance, primarily because it minimizes the necessity for joins, enabling quicker query execution. (Geetika & Bharat, 2014)

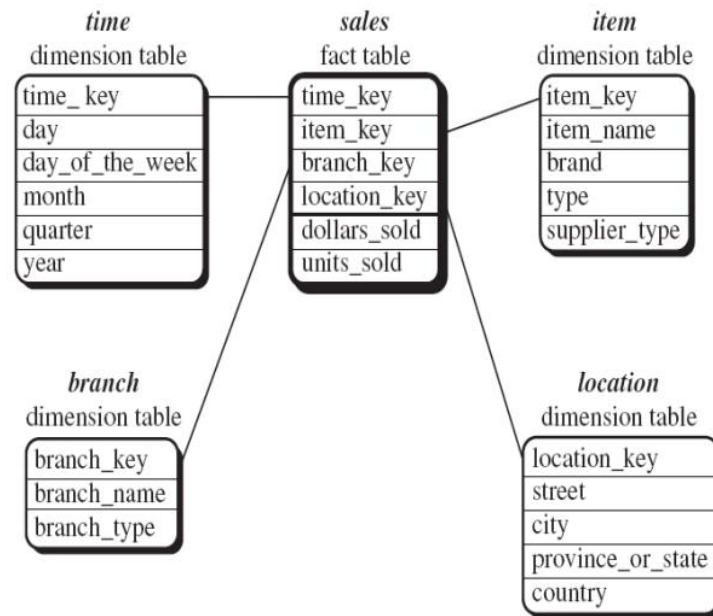


Figure 2-3. Star schema

(Source: Geetika & Bharat, 2014)

### 2.4.2. Snowflake Schema

The Snowflake Schema represents a more intricate variant of the Star Schema. In this design, the dimension tables undergo normalization, which entails dividing them into further tables. This process of normalization serves to diminish data redundancy and enhances the integrity of the data. However, a consequence of this is the increased complexity in querying, as it necessitates more joins to retrieve the desired data. Although this schema might contribute to more efficient data storage, it potentially compromises query performance when compared to the simpler Star Schema. (Geetika & Bharat, 2014)

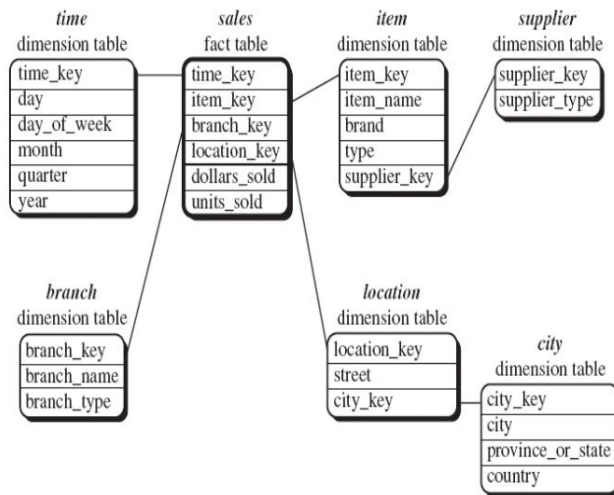


Figure 2-4. Snowflake Schema

(Source: Geetika & Bharat, 2014)

### 2.4.3. Galaxy Schema

Also known as the Fact Constellation Schema, the Galaxy Schema involves multiple fact tables that share many dimension tables. This schema is used for more complex data warehouse environments where businesses need to analyze data from multiple perspectives. Each fact table in the Galaxy Schema corresponds to a different business process, and shared dimensions allow for cross-fact analysis. This schema can handle more complex business scenarios but at the cost of increased complexity in database design and query execution. (Geetika & Bharat, 2014)

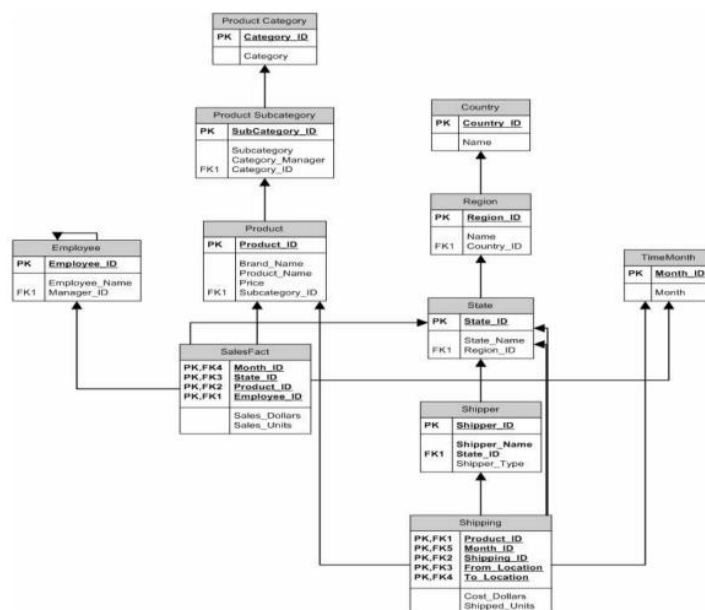


Figure 2-5. Galaxy Schema

(Source: Geetika & Bharat, 2014)

## **2.5. KPIs**

### **2.5.1. What is KPIs?**

Key Performance Indicators (KPIs) are quantifiable metrics that measure and evaluate the success of an organization, a department, a project or an individual in achieving specific business objectives. KPIs are important elements for assessing the performance of various aspects of a business. These indicators help organizations keep track of their progress and make informed decisions.

### **2.5.2. KPIs Types**

KPIs are divided into 4 types:

- *Strategic KPIs*: These big-picture indicators toward organization objectives. They provide a general view of how well an organization is progressing toward its strategic goals.

Examples: Revenue, Return on investment (ROI), Market share

- *Operational KPIs*: These indicators focus on shorter-term performance (typically day-to-day activities). They help organizations monitor the efficiency and the process of various operations.

Examples: Sales by region, Cost per acquisition (CPA)

- *Functional KPIs*: These indicators assess the performance of specific functions within an organization, such as sales, marketing or finance.

Examples: Return on assets (Finance), Time to resolution (IT)

- *Leading and Lagging KPIs*:

- + *Leading KPIs*: These predictive indicators that help organizations anticipate future outcomes and take measures to maintain or improve their current performance.

Examples: Website traffic, Customer engagement levels

- + *Lagging KPIs*: These indicators reflect historical performance and provide insights into it.

Examples: Customer churn rate, Revenue, Product defect rate

Organizations should use a combination of both leading and lagging KPIs to ensure a comprehensive understanding of performance.

## **CHAPTER 3. REQUIREMENTS ANALYSIS AND EXPERIMENTAL MODELING**

### **3.1. Business Context of Sales Department**

AdventureWorks functions as a multinational retail company specializing in providing bicycles and related accessories, operating primarily in the North American, European, Asian, and Australian markets. Adventure Works Cycles has two types of customers: individual consumers who buy products from online stores and retail or wholesale stores who buy products for resale from sales representatives.

The sales department is responsible for connecting the company's products with its customers. Their tasks extend beyond planning, introducing, and selling products to also include respecting and understanding the customers. They need to communicate attentively, monitor customer needs, persuade them, and propose solutions to satisfy those needs. The key is to maintain customers, especially those with high sales volume.

Sales department staff also have other responsibilities such as increasing market share by defining and serving potential customer groups, expanding the product portfolio through an external website, and reducing sales costs by reducing production costs.

#### ***a) Purpose of Sales***

Conversion of Sales:

The primary goal of the sales department is to increase the number of customers and sales volume. Their conversion rate needs to be improved to achieve this objective. This involves various metrics such as converting website visits into purchases/leads, converting potential customers attending conferences into paying customers, and converting store visitors into customers.

Customer Retention:

In addition to attracting customers, retaining them is equally important. This includes ensuring their satisfaction with the product and service. The sales department needs to monitor customers to ensure their continuous satisfaction and desires.

### Encouraging Business Growth:

All the sales department's efforts aim to increase sales and business growth. Skilled sales personnel always seek potential customer portfolios to convert into customers. Success in creating excellent sales services will build trust and open up growth opportunities.

### ***b) Sales Process***

#### Step 1: Identify potential customers

In this stage, the sales team needs to identify potential customers, determine their need for the company's bicycles, and assess their financial capability. This is an important part of the sales process that sales staff need to carry out on a daily or weekly basis.

#### Step 2: Preparation

Market research and gathering information about the product to be introduced to customers, target customers, the industry, and the unique value the brand offers. This information is used to create a presentation and tailor it to the specific needs of potential customers. Thorough preparation provides the foundation for the next steps in the sales process and helps in initial contact with potential customers.

#### Step 3: Approach

This is the first formal contact with the customer, which can be through a face-to-face meeting or a phone call. Some methods of approaching potential customers include giving gifts to create a positive impression at the first meeting, asking questions to engage their interest, and providing product samples for them to experience.

#### Step 4: Presentation

This is the stage where the salesperson demonstrates how the product meets the needs of potential customers. The sales department often organizes an official product presentation for potential customers. The use of PowerPoint or sales pitches may be applied, but it is also important to listen and respond appropriately to the customer's feedback.

#### Step 5: Handling objections

When potential customers raise objections, it requires timely preparation and response. This involves careful research and thorough preparation before the presentation; predicting and preparing to address potential objections. This is an opportunity to listen and address the concerns of potential customers, which forms the basis for adjusting the product to meet their needs.

#### Step 6: Finalize the deal and Sign the contract

When the customer agrees to make a purchase, finalize the deal, agree on the price, and sign a purchase contract.

#### Step 7: Follow-up

An important step to ensure customer satisfaction, maintain and seek new opportunities. Post-sales contact helps maintain and create opportunities to expand sales, as well as gain referrals from loyal customers. Maintaining relationships with current customers is also important for saving costs on attracting new customers.

### **3.2. Business Challenges**

AdventureWorks is a sample database for Microsoft SQL Server, designed to simulate a real-world business scenario for a fictitious company, Adventure Works Cycles. The AdventureWorks database features a complex schema designed to mimic a real-world business environment. It encompasses various modules including Human Resources, Product Information, Sales, Purchasing, and Production, with interconnected tables for a holistic business model. Key tables include Employees, Customers, Products, Sales Orders, and Vendors, allowing in-depth analysis across different business functions. This relational structure enables comprehensive data analysis, from sales trends to inventory management.

In this project, we will leverage the sales module data from the AdventureWorks database to analyze and provide insights for key business issues. These include evaluating Sales Performance, gaining a deeper understanding of Customer Behavior, and extracting valuable Product Insights as follows:



***Sales Performance Analysis:*** This challenge involves evaluating the effectiveness of sales strategies, identifying top-performing products, and understanding sales trends. By analyzing sales data, businesses can identify successful products and regions, compare performance over different periods, and make data-driven decisions to improve sales strategies.

***Customer Behavior Insights:*** Analyzing customer purchase history and demographics helps in understanding customer preferences and buying patterns. This insight is crucial for targeted marketing, product development, and improving customer satisfaction and loyalty.

***Product Insights:*** This involves analyzing product performance, understanding market demand, and identifying areas for improvement or innovation. By examining sales data, businesses can discover which products are successful, which need improvement, and potential gaps in the product line.

Through this data-driven approach, we aim to uncover patterns and trends that can inform strategic decisions, enhance customer engagement, and optimize product offerings in the competitive market landscape.

### **3.3. Business Requirements**

#### **How is Sales performance in each territory?**

**Sales Revenue:** This KPI measures the amount of revenue generated by each sales territory. It provides insight into the financial performance of each sales territory and insight into which territory is the most effective.

$\text{Sales Revenue} = \text{Number of Units Sold} \times \text{Price per Unit}$
---

#### **How does the company estimate future sales and profit?**

By analyzing the revenue and profit over time, we can see the trend by time (month, quarter, and year). It provides insight to estimate future sales. It helps businesses

anticipate future demand, identify potential problems or opportunities, and adjust their strategies accordingly. It can also help businesses optimize their inventory levels, production schedules, and staffing requirements. Understanding projected revenue is crucial for being able to plan for future growth.

### **How is the business strategy of the enterprise by time?**

Average Order Value (AOV) means the average value per order. This is a metric that measures the average amount of money customers spend on each order. We use this metric to evaluate the overall effectiveness of the business and marketing strategies.

$$\text{AOV} = \frac{\text{Revenue}}{\text{Total Order Quantity}}$$

If the AOV increases, it may be due to effective promotion strategies and growing customer demand. In contrast, if the AOV decreases, the enterprise may need to reconsider the pricing, marketing strategies and enhance the customer experience to encourage them to buy more products.

### **What are the best-selling products in the world by country?**

Determining the best-selling products in specific countries helps businesses understand the local resident's purchasing trends, allowing them to identify the strengths and weaknesses of their products. This information is crucial for optimizing marketing and advertising strategies in each region.

For products that are already popular, it indicates a high customer demand. Businesses can capitalize on this by intensifying efforts to boost sales in that particular region. Additionally, strategic approaches such as bundling low-performing products with best-sellers in combos can be implemented to stimulate overall sales.

### **What are the revenue of top best-selling products?**

Absolutely, tracking and gaining a deeper understanding of the business performance of top-selling products help the company make informed decisions regarding product maintenance, development, and strategic marketing. Additionally, the revenue generated from these best-selling products assists the business in making insightful decisions about marketing strategies.

### **Where is the most effective customer segment?**

Understanding the customer segments that a business can effectively target is crucial. This knowledge enables the business to make decisions to intensify marketing and advertising efforts in these segments while increasing sales through strategies aimed at fostering customer loyalty, such as birthday gifts and discounts for repeat purchases.

### **3.4. Proposing BI Solution**

To solve the problems, a complete BI solution would be suitable.

- Requirement Gathering: The first step is taking detailed requirements from Adventure Work Sale Manager for what is the data source, what is the outcome they want, such as factors to keep track or metrics to show in the dashboards.
- Planning: Then, the data team search for the data source and propriate tools for the whole solution. In this step, we will define data in detail (tables, attributes) for building dashboards and tools for ETL, data analysis, building data model.
- Data Gathering: Data sources usually come from internal data sources such as OLTP data from SAP system. In addition, we can gather data from external sources such as from third parties or public data sources. In this project, we use data from Adventure Work OLTP database and Adventure Work Sample Datawarehouse.
- ETL: Because data sources are Adventure Work, and this data is quite clean and in a relational table format. Then we decided to choose ETL technique to extract and load data to data warehouse. The sales analysis is conducted on data of individual customers during the period from 2005 to 2014.
- Building Data Warehouse: From requirements and data nature, the data warehouse would be built based on Star, Snowflake or Galaxy schema.

- Data analysis & visualization: Finally, data analysis and visualization to investigate into the data, gain the insights about the sale activities and performances, and suggest recommendations to stakeholders. In this very first stage, we focus on building overview about sales performance and two dashboards about products and customers if possible.

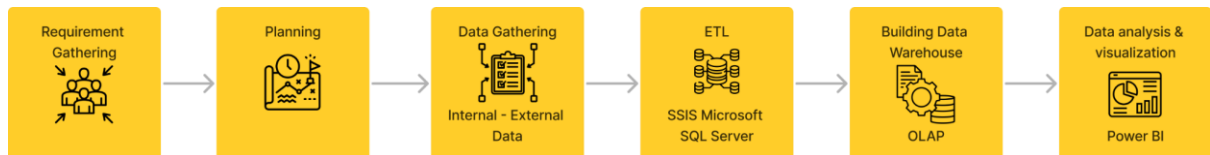


Figure 3-1. BI Solution

*(source: Authors)*

## CHAPTER 4. EXPERIMENTAL RESULTS AND ANALYSIS

### 4.1. Result Structure

#### 4.1.1. Bus Matrix

The Bus Architecture in Data Warehousing involves a configuration of fact and dimension tables, along with their interrelationships, that are aligned with business processes. This setup facilitates the creation of Data Marts. Establishing connections between fact and dimension tables is a critical step in developing a relational model, which aids administrators in efficient decision-making. There are two primary approaches to this process: the Top-down and Bottom-up methods.

In our team's project, we employ the Top-down approach. This method involves constructing the entire enterprise database in a centralized manner. The advantage of using the Data Warehouse Bus Architecture is that it allows for the integration of Data Warehouse components. This integration is achieved through the use of standardized dimensions that are not only appropriate and shared but also applicable across various processes.

In this project, columns will be the main Dimensions, namely: Customer, Product, Time, Promotion, Sales Territory, Sales Reason. Rows are business processes that are intended to fulfill the business requirements that the project team wants to achieve.

Table 4-1. Bus Matrix

*(Source: authors)*

	<b>Business Process</b>		
<b>Common Dimensions</b>	Sales performance	Product detail	Customer detail
DimCustomer	X		X
DimProduct	X	X	
DimPromotion	X		
DimGeography	X		X
DimSalesTerritory	X		X

DimSalesReason	X		
DimDate	X	X	X

#### 4.1.2. Master Data

Master Data is a specific kind of data known for its high stability and long-term validity. It constitutes the foundational base for all business processes within a company, and is utilized across various departments. Due to its crucial importance, master data often becomes the center of attention in data evaluations and is a key factor in statistical analyses. A notable aspect of master data is its relationship with transaction data, which relies heavily on it, highlighting its essential role in the overall functioning of a company. Examples of master data include product names, warehouse locations, and the raw materials used in manufacturing processes. Given its critical role in business operations, there is a need for maintaining and managing master data meticulously and over extended periods.

Table 4-2. The master data

*(Source: Authors)*

Object	Description
Product	Contains details about the products being sold, including product specifications, pricing, and other relevant data.
Customer	Holds information about the customers who purchase products, including customer profiles, purchase history, and preferences.
SalesTerritory	Encompasses information about sales territories, detailing the regions or areas where products are sold.
ProductCategory	This data gives details about the various types of products for sale, including the names and descriptions of each product category.

ProductSubcategory	This data includes details about product subcategories, providing their names and descriptions for a more detailed classification.
--------------------	--

#### 4.1.3. Transaction Data

Transactional data and master data are two different kinds of data used in an organization. Master data is the main data that includes important information like product details, customer details, and information about sales. It doesn't change much and is important for the long-term planning and structure of a company.

Transactional data, on the other hand, is data that changes often and is used for daily operations. It includes details about specific actions, like the number assigned to a purchase order. This data is constantly updated and is used mainly by certain departments for a short time.

Both types of data are important, but master data is key for the overall structure and long-term goals of a company, while transactional data helps with day-to-day activities.

Table 4-3. The transaction data

*(Source: Authors)*

Object	Description
Sale.SalesOrderDetail	This object includes data about individual products linked to a specific sales order. It provides detailed information on each product within an order.
Sale.SalesOrderHeader	This object contains comprehensive information about each sales order. Details such as customer information, order date, and the current status of the order are included.

#### 4.1.4. Fact and Dim Table

Table 4-4. Fact and Dim Table overview

*(Source: Authors)*

No	Type	Table
1	Fact	FactSales
2	Dim	DimCustomer
3		DimProduct
4		DimPromotion
5		DimGeography
6		DimTerritory
7		DimSalesReason
8		DimDate

Table 4-5. Fact Table

(Source: Authors)

Attribute Name	Data type	Length	Table (Source)	ETL rule
FactSales				
SalesOrderID (PK)	int	4	Sales.SalesOrderHeader	From Source
UnitPrice	money	8	Sales.SalesOrderDetail	
UnitPriceDiscount	money	8		
PromotionID	int	4		
SalesAmount	numeric	17	Sales.SalesOrderDetail (Total)	
OrderQuantity	smallint	2	Sales.SalesOrderDetail	
ProductID	int	4		
OrderDate	datetime	8	Sales.SalesOrderHeader	
DueDate	datetime	8		
ShipDate	datetime	8		



Freight	money	8		
TaxAmt	money	8		
SalesOrderNumber	nvarchar	50		
RevisionNumber	tinyint	1		
CustomerID	int	4		
TerritoryID	int	4		
CurrencyRateID	int	4		
SalesOrderDetailID	int	4	Sales.SalesOrderDetail	
CarrierTrackingNumber	nvarchar	50		
CustomerPONumber	nvarchar	50		
SalesReasonID	int	4	Sales.SalesOrderHeader SalesReason	
TotalProductCost	numeric	13		Derived column by multiplying OrderQuantity and ProductStandard Cost

Table 4-6. Dimension Tables

(Source: Authors)

Attribute Name	Data type	Length	Table (Source)	ETL Rule
<b>DimCustomer</b>				
CustomerID(PK)	int	4	Sales.Customer	From Source
PersonID	int	4		
TerritoryID	int	4		
AccountNumber	varchar	10		
PersonTitle	nvarchar	16	Person.Person (Title)	

FirstName	nvarchar	100	Person.Person	
MiddleName	nvarchar	100		
LastName	nvarchar	100		
GeographyKey	int	4	DimGeography	
AddressLine1	nvarchar	120	Person.Address	
AddressLine2	nvarchar	120		
BirthDate	nvarchar	20	DimCustomer (AWDW2022)	
MaritalStatus	nvarchar	2		
Gender	nvarchar	2		
EmailAddress	nvarchar	100		
YearlyIncome	money	8		
TotalChidren	tinyint	1		
NumberChildrenAtHome	tinyint	1		
EnglishEducation	nvarchar	80		
EnglishOccupation	nvarchar	200		
HouseOwnerFlag	nvarchar	2		
Phone	nvarchar	40		
DateFirstPurchase	nvarchar	20		
CommuteDistance	nvarchar	30		
ActiveFrom	datetime	8		
ActiveTo	datetime	8		
DimProduct				
ProductID (PK)	int	4	Production.Product	From Source
ProductName	nvarchar	100		

StandardCost	money	8		
ListPrice	money	8		
ProductLine	nvarchar	4		
SellStartDate	datetime	8		
SellEndDate	datetime	8		
ProductNumber	nvarchar	50		
SizeUnitMeasureCode	nvarchar	6		
WeightUnitMeasureCode	nvarchar	6		
FinishedGoodsFlag	bit	1		
Color	nvarchar	30		
SafetyStockLevel	smallint	2		
ReorderPoint	smallint	2		
Size	nvarchar	10		
Weight	numeric	5		
DaysToManufacture	int	4		
Class	nvarchar	4	Production.ProductCategory	
Style	nvarchar	4		
ProductSubcategoryName	nvarchar	100	Production.ProductCategory	
ProductCategoryName	nvarchar	100		
ActiveFrom	datetime	8		Slowly changing dimension
ActiveTo	datetime	8		

DimPromotion				
PromotionID (PK)	int	4	Sales.SpecialOffer	From Source
PromotionName	nvarchar	510		
DiscountPct	money	8		
PromotionType	nvarchar	100		
PromotionCategory	nvarchar	100		
StartDate	datetime	8		
EndDate	datetime	8		
MinQty	int	4		
MaxQty	int	4		
ActiveFrom	datetime	8		
ActiveTo	datetime	8		
DimGeography				
GeographyKey (PK)	int	4		Auto Generate
StateProvinceCode	nvarchar	6	Person.StateProvince	From Source
CountryRegionCode	nvarchar	6		
TerritoryID	int	4		
StateProvinceName	nvarchar	100	Person.StateProvince (Name)	
City	nvarchar	60	Person.Address	
PostalCode	nvarchar	30		
StateProvinceID	int	4		
DimTerritory				
TerritoryID (PK)	int	4	Sales.SalesTerritory	From Source
RegionName	nvarchar	100		

CountryRegionCode	nvarchar	6		
TerritoryGroupName	nvarchar	100		
TerritoryCountry	nvarchar	28		Derived Column based on CountryRegionCode
<b>DimSalesReason</b>				
SalesReasonID (PK)	int	4		
SalesReasonName	nvarchar	100	Sales.SalesReason	From Source
SalesReasonType	nvarchar	100		
<b>DimDate</b>				
DateKey (PK)	int	4		Auto Generate
Date	datetime	8	Component Script as Source	Date generated from 01/01/2005 to 12/31/2014
Year	int	4		Generated based on Date
Month	int	4		
Day	int	4		
Quarters	int	4		
WeekOfYear	int	4		
DayOfYear	int	4		
DayOfMonth	int	4		
DayOfWeek	int	4		
MonthName	nvarchar	18		

### 4.1.5. Data Warehouse Model

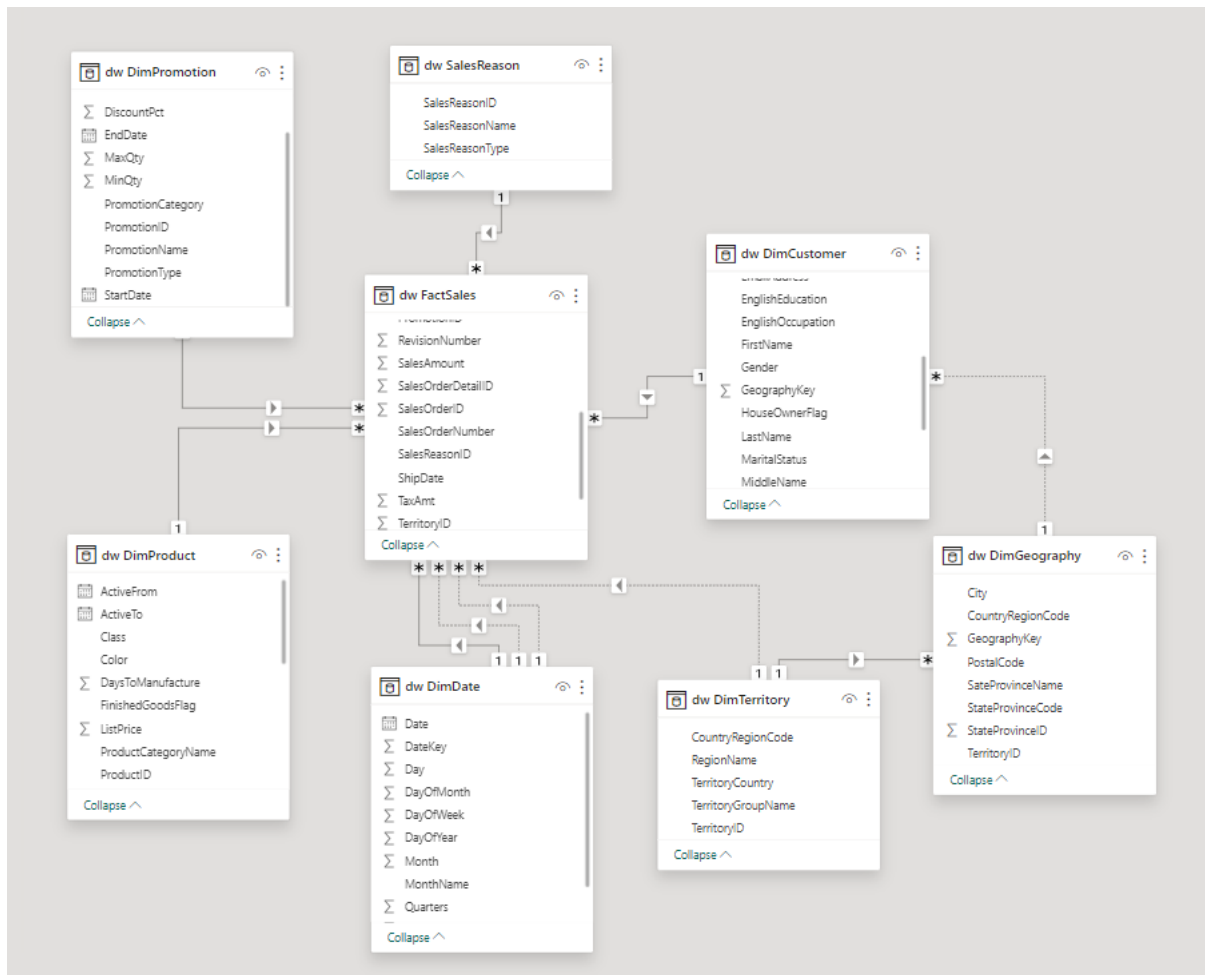


Figure 4-1. Data warehouse diagram

(Source: Authors)

Table 4-7. Description of key relationship in SalesDW

(Source: Authors)

No	Relationship	Type of relationship	Description
1	DimCustomer- FactSales	1-n	A customer can relate to one or more rows in FactSales, and each row in FactSales connects with only one customer.

2	DimTerritory- Fact Sales	1-n	A territory can relate to one or more rows in FactSales, and each row in FactSales connects with only one territory.
3	DimProducts- FactSales	1-n	A product can relate to one or more rows in FactSales, and each row in FactSales connects with only one product.
4	DimDate- FactSales	1-n	An order date can relate to one or more rows in FactSales, and each row in FactSales connects with only one order date.
5	DimDate- FactSales	1-n	A ship date can relate to one or more rows in FactSales, and each row in FactSales connects with only one ship date.
6	DimDate- FactSales	1-n	A due date can relate to one or more rows in FactSales, and each row in FactSales connects with only one due date.
7	DimPromotion- FactSales	1-n	A promotion can relate to one or more rows in FactSales, and each row in FactSales connects with only one promotion.
8	DimSalesReason- FactSales	1-n	A sales reason can relate to one or more rows in FactSales, and each row in FactSales connects with only one sales reason.
9	DimTerritory- DimGeography	1-n	A territory can relate to one or more rows in DimGeography, and each row in

			DimGeography connects with only one territory.
10	DimGeography-DimCustomer	1-n	A geography can relate to one or more rows in DimCustomer, and each row in DimCustomer connects with only one geography.

## 4.2. ETL Process

### 4.2.1. Overall Process

The overall process undergoes three main steps:

- Extracting: We use the OLTP Adventure Works for most of the sources of dimension and fact tables. Only the Customer dimension table gets additional data from the DimCustomer table of the Adventure Works data warehouse. We extract columns that can contribute to future analysis work and align with business meaning.
- Transforming: Data is transformed by excluding NULL values and deriving new columns to support analytics. We do not implement type casting in this step due to the correct data type of the OLTP database.
- Loading: In this final step, we load all dimension and fact tables into the data warehouse named “ADVENTURE\_WORKS” in SQL Server.

The order to load is:

- + First Load: DimGeography.
- + Second Load: DimTerritory, DimProduct, DimCustomer, DimPromotion, DimSalesReason, DimDate.
- + Third Load: FactSales.

The tool used for this process is SSIS of Visual Studio 2022.

### 4.2.2. Detailed Processes

#### a) DimGeography



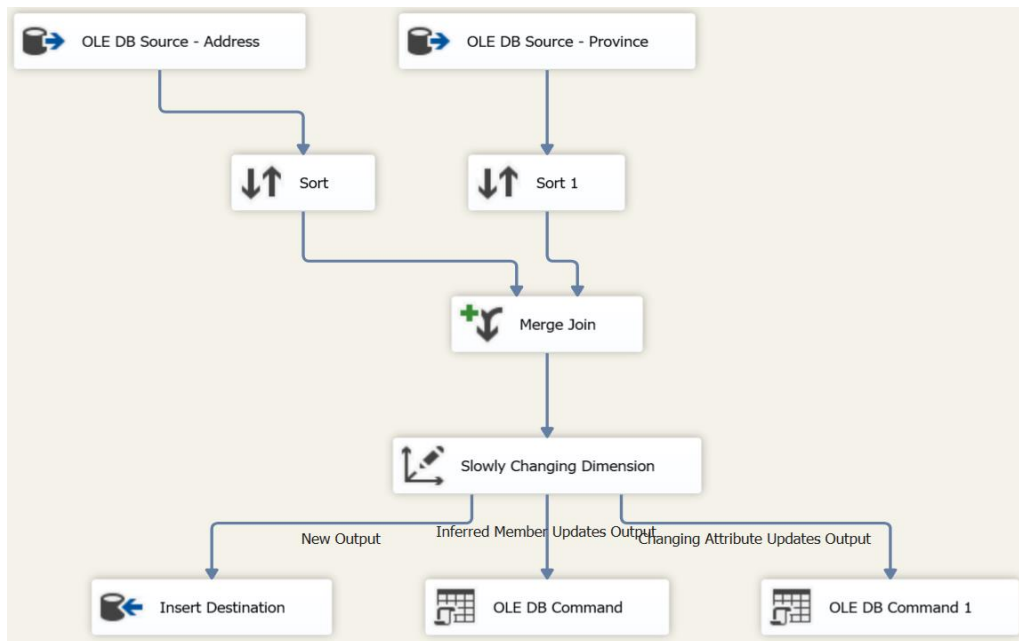


Figure 4-2. Data Flow Task for DimGeography  
(source: Authors)

Data from table Person.Address and Person.StateProvince is sorted by (1) StateProvince, (2) City and (3) PostalCode, then left outer merged. After this, these data are overwritten by Slowly Changing Dimension.

## b) DimTerritory

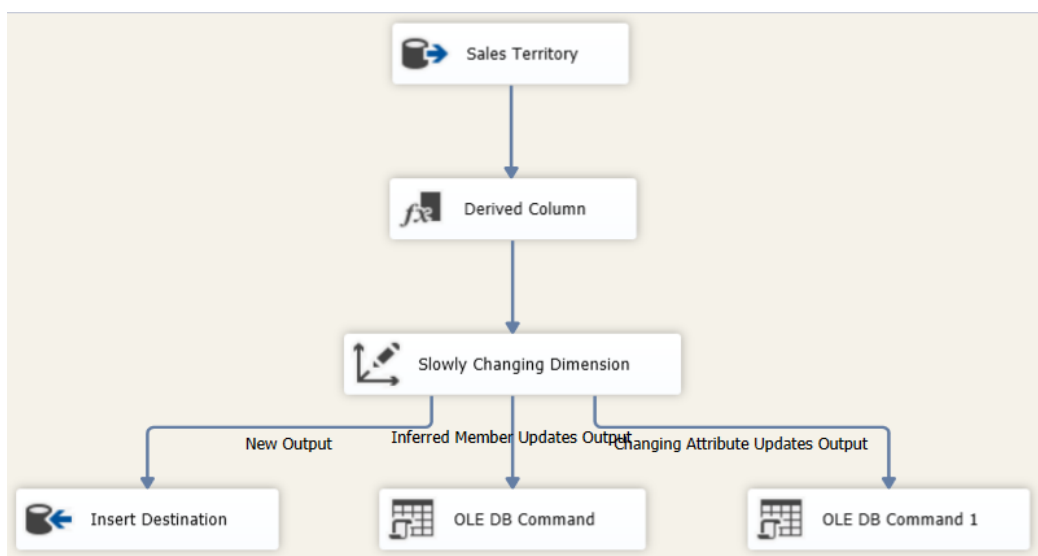


Figure 4-3. Data Flow Task for DimTerritory  
(source: Authors)

Data is retrieved from the Sales.SalesTerritory table. Subsequently, a new column named "TerritoryCountry" is generated using a Derived Column Component. The derivation rule is based on the Country Region Code.

After this, these data are overwritten by Slowly Changing Dimension Type.

### c) DimProduct

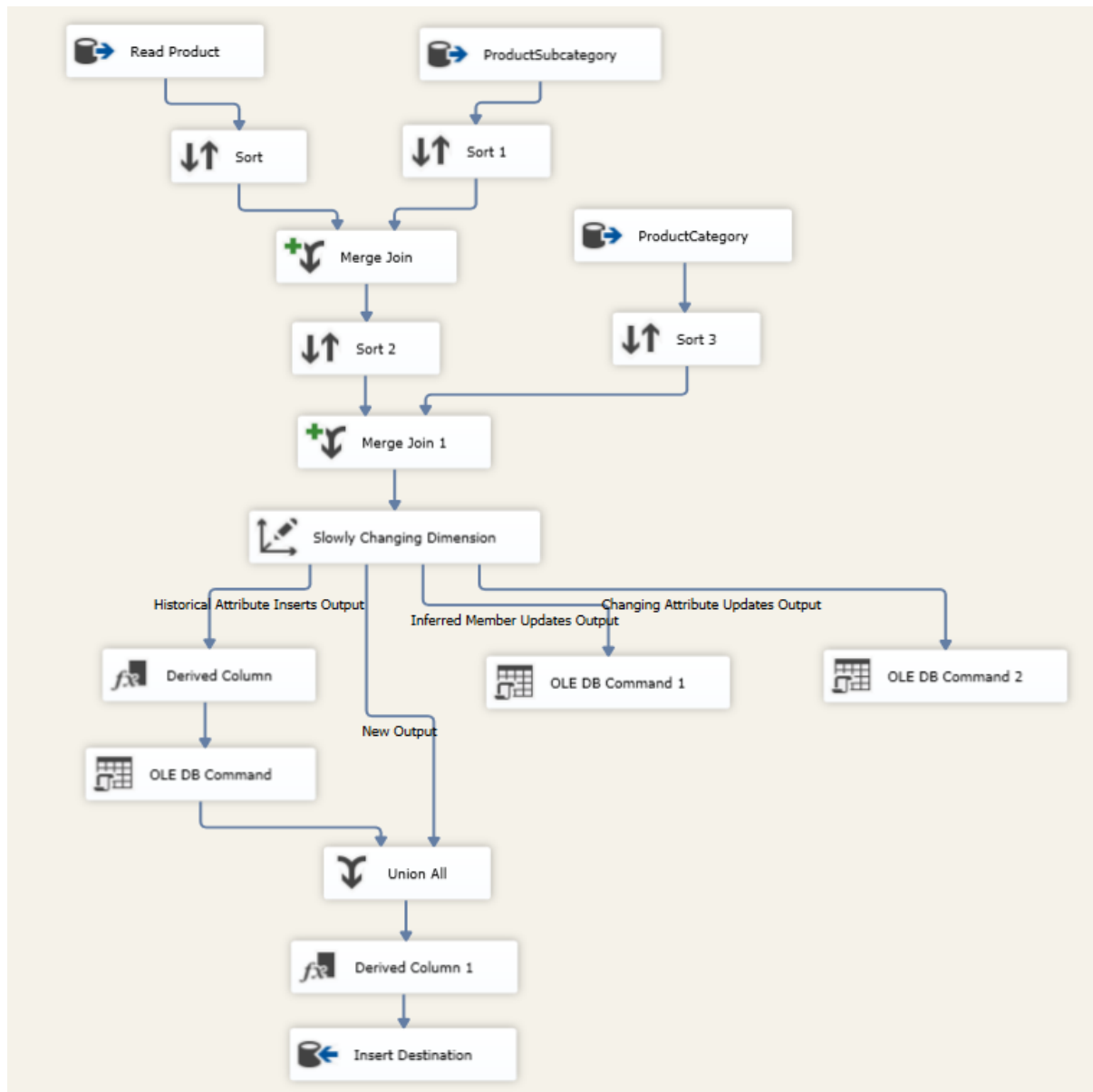


Figure 4-4. Data Flow Task for DimProduct

(source: Authors)

DimProduct is constructed by integrating data from table Production.Product and the Subcategory Name and Category Name from two tables: Production.ProductSubcategory and Production.ProductCategory.

Slowly Changing Dimension is used to record the historical changes of DaysToManufacture, ListPrice, ProductSubcategoryName, ReorderPoint, SafetyStockLevel, Size, SizeUnitMeasureCode, StandardCost by two new columns (Active From, Active To).

#### d) DimCustomer

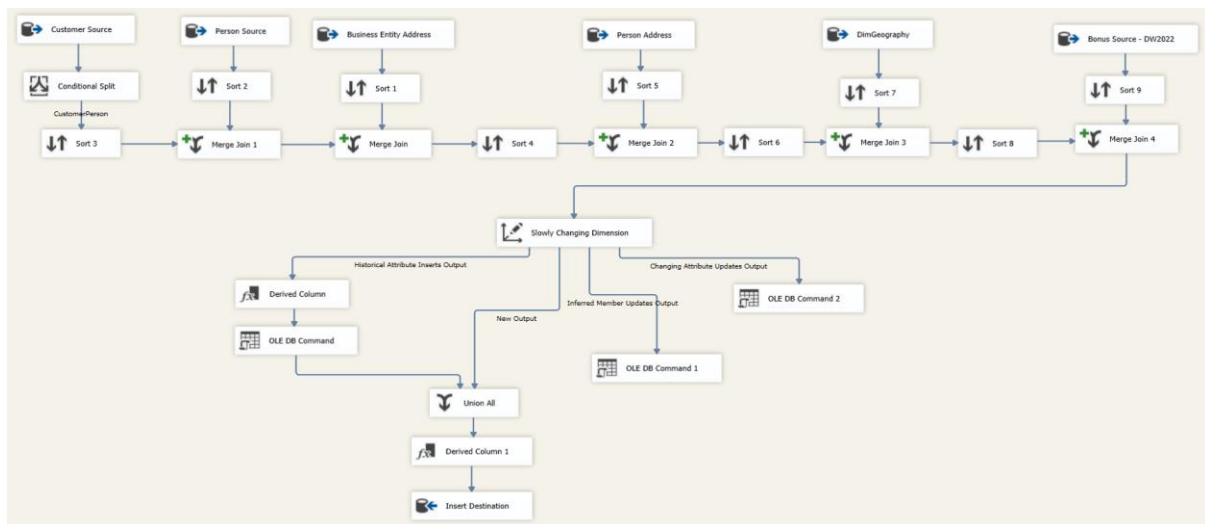


Figure 4-5. Data Flow Task for DimCustomer

(source: Authors)

This dimension table gets data sources from multiple tables. Firstly, we read the data from Sales.Customer table and filter out stores. Secondly, we get information about customers from tables Person.Person, Person.Address. We also input the GeographyKey from DimGeography. In addition, because the OLTP data source does not support other information about customers such as Marital Status, Gender, etc, we decided to get this information from an additional source - Adventure Work Datawarehouse 2022.

Finally, when the attributes are complete, Slowly Changing Dimension is used to keep track of changes for AccountNumber, GeographyKey, TerritoryID through two new columns (Active From, Active To).

### e) DimPromotion

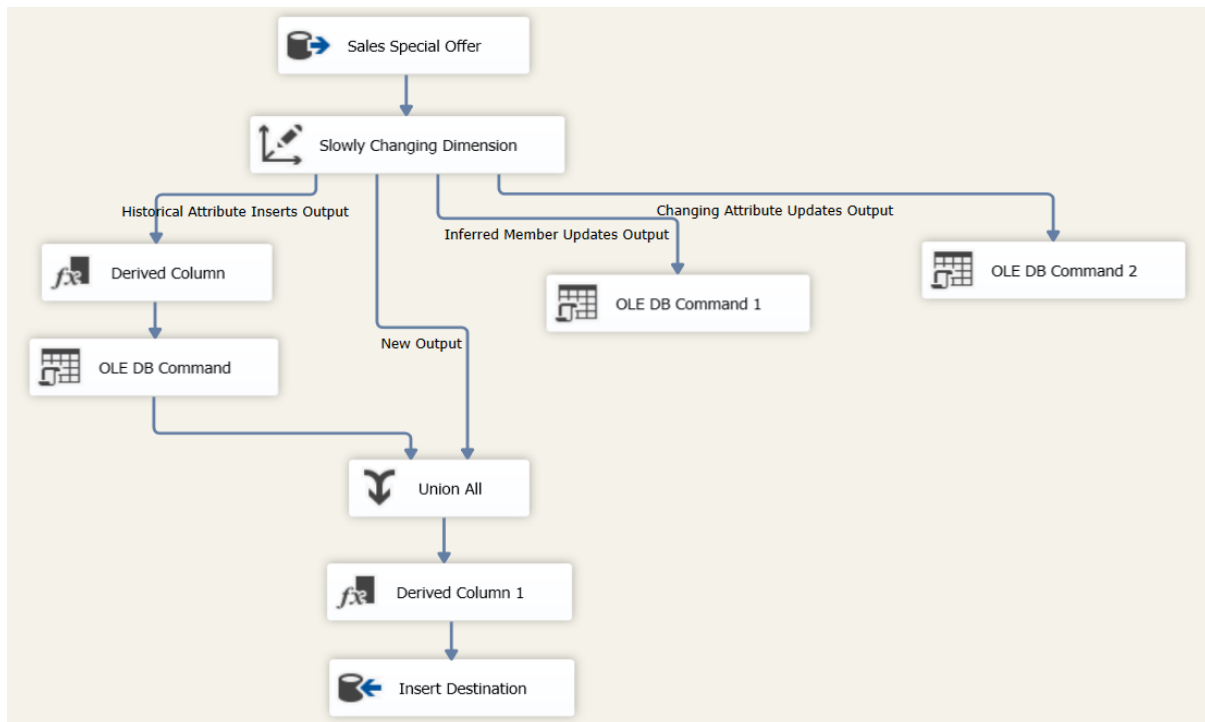


Figure 4-6. Data Flow Task for DimPromotion

(source: Authors)

Data is fetched from Sales.SpecialOffer and loaded into data warehouse via Slowly Changing Dimension. The historical attributes include DiscountPct, PromotionType.

### f) DimSalesReason

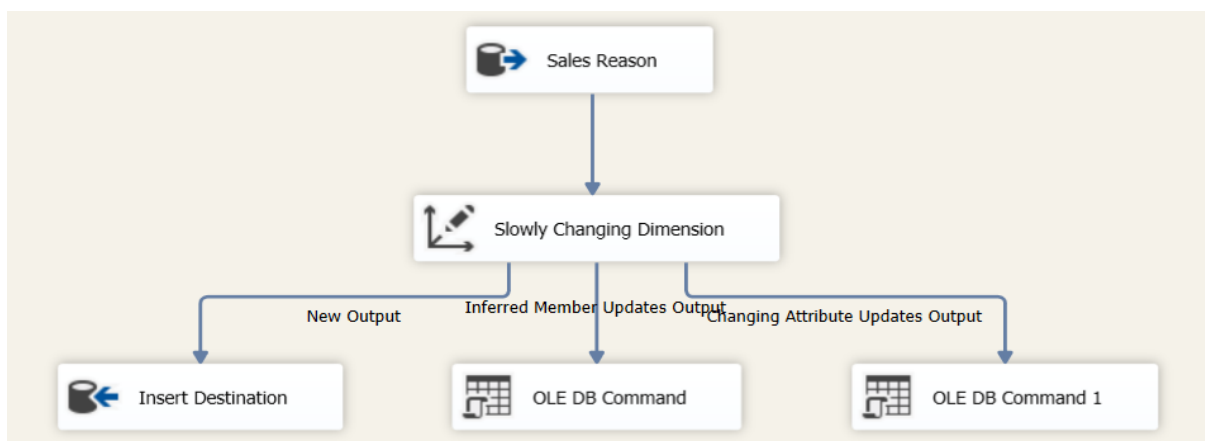


Figure 4-7. Data Flow Task for DimSalesReason

(source: Authors)

All attributes are obtained from SalesReason table. Then these values are overwritten by Slowly Changing Dimension.

## g) DimDate

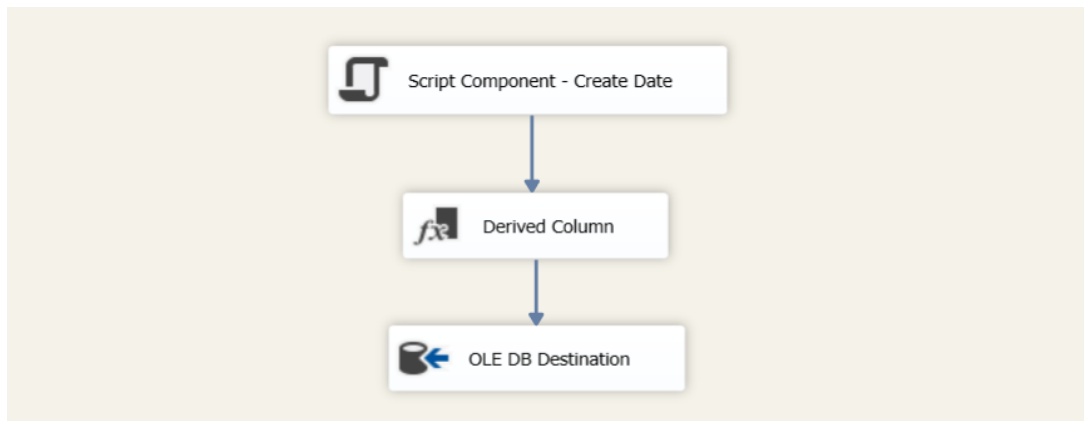


Figure 4-8. Data Flow Task for DimDate

(source: Authors)

This dimension table is specially created from a script component as a source. In this component, we create a new column named “Date”, which is the date from 01/01/2005 to 12/31/2014.

```

public override void CreateNewOutputRows()
{
    DateTime startDate = new DateTime(2005, 1, 1);
    DateTime endDate = new DateTime(2014, 12, 31);

    for (DateTime date = startDate; date <= endDate; date = date.AddDays(1))
    {
        OutputBuffer.AddRow();
        OutputBuffer.Date = date;
    }
}
  
```

Figure 4-9. Script to create “Date” column

(source: Authors)

Next, we added columns to specify the date by submitting the following syntax to Derived Column component:

Derived Column Name	Derived Column	Expression	Data Type
Year	<add as new column>	YEAR(Date)	four-byte signed integer [DT_I4]
Month	<add as new column>	MONTH(Date)	four-byte signed integer [DT_I4]
Day	<add as new column>	DAY(Date)	four-byte signed integer [DT_I4]
Quarters	<add as new column>	DATEPART("quarter",Date)	four-byte signed integer [DT_I4]
WeekOfYear	<add as new column>	DATEPART("week",Date)	four-byte signed integer [DT_I4]
DayOfYear	<add as new column>	DATEPART("dayofyear",Date)	four-byte signed integer [DT_I4]
DayOfMonth	<add as new column>	DATEPART("day",Date)	four-byte signed integer [DT_I4]
DayOfWeek	<add as new column>	DATEPART("weekday",Date)	four-byte signed integer [DT_I4]
MonthName	<add as new column>	MONTH(Date) == 1 ? "January" : (MONTH(Dat...	Unicode string [DT_WSTR]

Figure 4-10. Derive new columns for DimDate

(source: Authors)

We kindly note that these data should be loaded at once in the beginning of a new year.

## h) FactSales

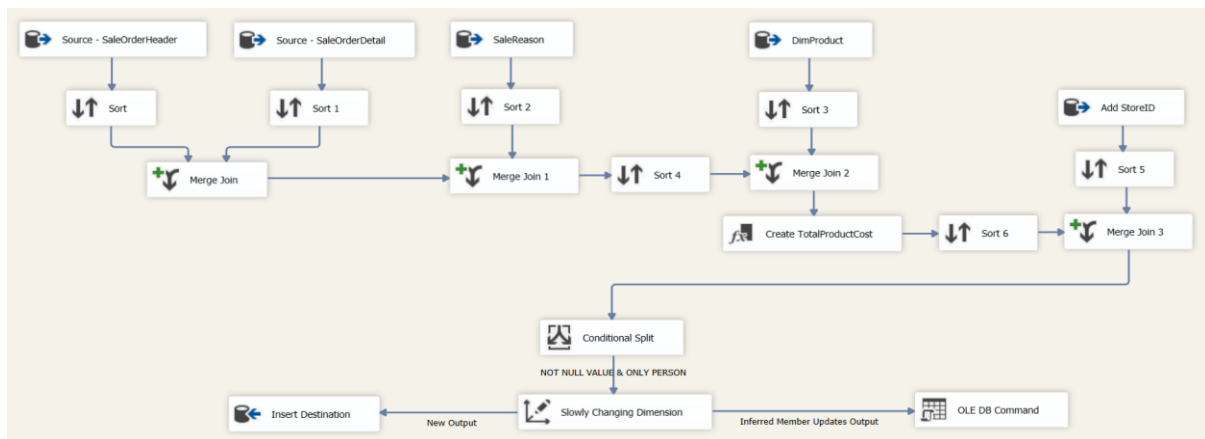


Figure 4-11. Data Flow Task for FactSales

(source: Authors)

From the merge between SalesOrderHeader and SalesOrderDetail, we get the SalesReasonID from SalesReason table, standard cost to calculate the total product cost and StoreID to exclude the orders that belong to stores. At the end, we filter out rows recording null values in any attributes and overwritten values onto the FactSales table.

## 4.3. Data Analytics

### 4.3.1 Report on Sales Performance

The Sales Performance dashboard, as shown in Figure 4-12, is divided into two sections: a section with general information and a section with charts that make the data easier to compare. The dashboard also has slicers that allow users to filter by country and time. The company is operating until 2014 and this report aims to provide ideas for enhancing the company's productivity in the near future by summarizing the sales performance in 2013.

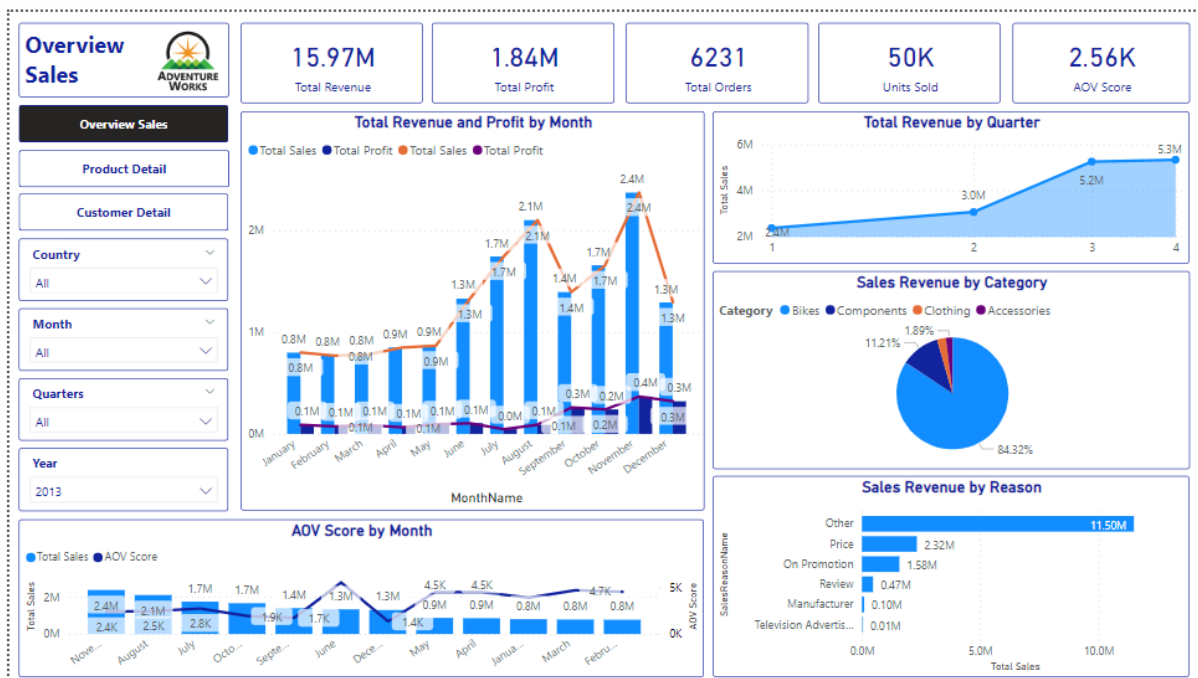


Figure 4-12. Overview of the Sales Performance Dashboard

(Source: Authors)

## Sales Performance Overview

In 2013, total revenue is 15.97 million, profit is 1.84 million. The enterprise has 6,231 orders with the number of products sold being 50 thousand. The average amount of money customers used to spend per order is 2.56 thousand.

## Revenue and Profit Growth Analysis

### Revenue:

The chart shows that revenue increased intermittently in 2013, with the highest being November (~\$2,400,000) and the lowest being March (~\$780,000). From January to August, profits always increased, especially during the period from May to August, more than doubling from nearly ~\$1,000,000 to more than \$2,000,000.

However, in September there was a sharp decrease in revenue, then a sharp increase in October and November, and a sharp decrease in December (down nearly 46% from \$2,400,000 to \$1,300,000).

### *Profit:*

Profits in 2013 are unstable. From January to June, profits remained at a level of ~\$62,000 to ~\$90,000, with the lowest being in April and the highest in May. After that, profits dropped sharply from ~\$84,000 (June ) to ~8,500 (July), a decrease of about 10 times. Profits trended up rapidly through the end of the year, peaking at ~\$364,000 in November, then falling about 12% in December, leaving ~\$320,000.

In general, it can be seen that the growth rates of revenue and profit are not the same. Although May's revenue was \$1,750,000, profit was only ~\$84,000, % profit was 4.7%. Meanwhile, in December, revenue was less, ~\$1,290,000, but profit was up to ~\$320,000, % profit was 24.8%. It can be concluded that businesses have good strategic plans to optimize profits in the last months. In general, not only did profits increase, but also revenue in the last months of the year was higher than in the first months.

The revenue comparison among quarters shows that the revenue had consistent growth from the first quarter to the Fourth quarter. But there was a significant increase from the second quarter (\$3.000.000) to the third quarter (\$5.200.000), increasing 1.74 times. The data suggests that businesses should focus more on Q1, and Q2 to maximize revenue by planning some campaigns to increase purchasing power.

### **Sales Revenue by Sales Reason**

From the chart, it can be seen that customers are affected by many different reasons. However, Price, On Promotion, and Review account for significant revenue, 14.5%, 9.9%, and 2.9% respectively. Businesses should apply product price reduction strategies to increase the number of products sold, and at the same time have many promotion programs to attract customers. In addition, businesses should also run additional marketing and seeding campaigns to increase brand identity.

### **Sales Revenue by Category**

It can be seen that 'Bikes' is the Category that brings the most revenue to Adventure Works, accounting for 84.32%, followed by Components accounting for 11.21%.



Clothing and Accessories only account for a small proportion, 2.58% and 1.89% respectively. To promote sales of Clothing and Accessories, businesses should apply combo sales campaigns at preferential prices. The combos combine Bikes and Clothing or Accessories.

### **AOV Score by Month**

The AOV index has an intermittent downward trend in 2013, including 3 slight increases in February, June and November. In addition, the AOV index decreased sharply, doubling in July, from \$5600 to \$2800. Although the AOV index decreased, revenue still increased, this shows that the number of orders increased but the average amount of money customers used to spend per order decreased.

It can be predicted that the business was implementing an effective marketing campaign to find customers, but the overall business strategies are still not good, the business needs to adjust business policies accordingly more with the market.

### **4.3.2 Report on Sales Products Detail**

The Sales Products Detail dashboard contains overall data cells and visual charts illustrating detailed information about Adventure Works company products. Additionally, there are filters for time, region, and product type to refine and display detailed reports.

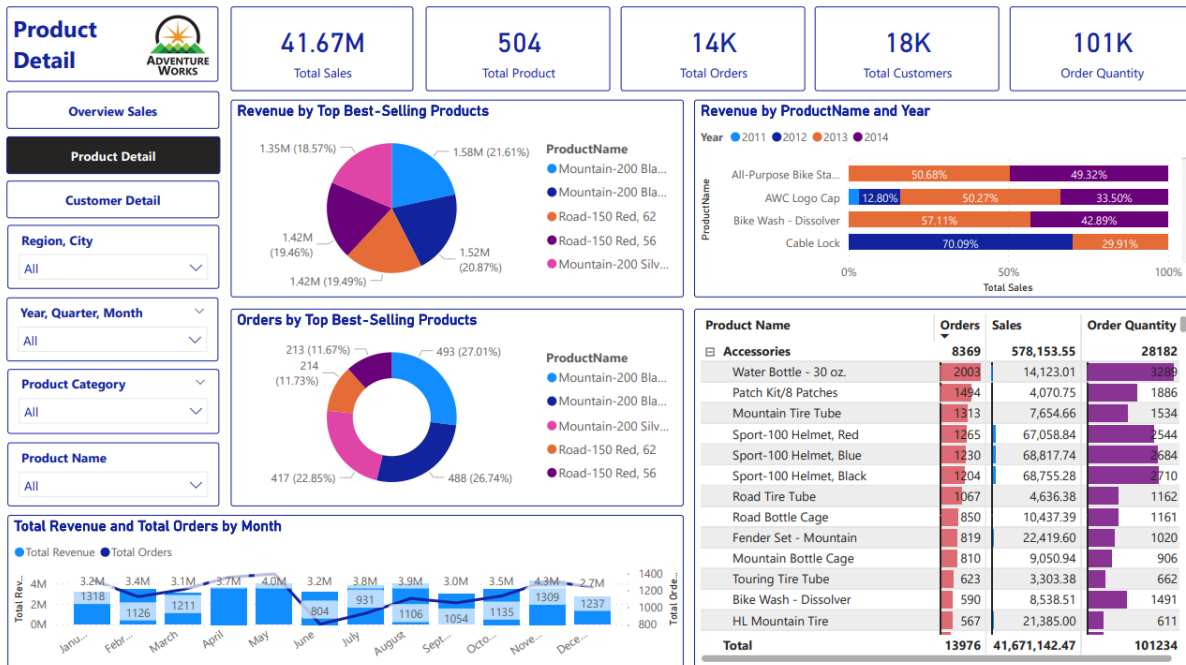


Figure 4-13. Overview of the Product Detail Dashboard

(Source: Authors)

## Product Detail Overview

The dashboard indicates that the sales amount of products is 41,67 millions and the total number of company products is 504. Up to 2014, the business had 14 thousand orders, with 101 thousand products sold. Bikes are truly the number 1 product of Adventure Works as the revenue of it is 36,70 millions (the total revenue is 41,67 millions).

## Revenue by Top Best-Selling Products

The pie chart "Revenue by Top Best-Selling Products" illustrates the top 5 best-selling products based on company revenue, ranked from high to low: Mountain-200 Black, 38 (21.61%), Mountain-200 Black, 42 (20.87%), Road-150 Red, 62 (19.49%), Road-150 Red, 56 (19.46%), Mountain-300 Silver, 38 (18.57%). This indicates that these are the products that significantly impact Adventure Works' revenue the most. Therefore, the business should prioritize focusing on marketing and advertising efforts or expanding related products associated with these items.

## Revenue by ProductName and Year

The dashboard presents product sales figures across the years from 2011 to 2014 (the most recent years). It is evident that the product with the highest sales is Mountain-200 Black, 38. In 2013, the sales of this product constituted 51.12%, possibly indicating that the company undertook sales promotion campaigns or that it was a year when cycling events became more popular. Based on this chart, businesses can assess the sales performance of different product types in recent years to devise effective sales strategies.

### Total Revenue and Total Orders by Month

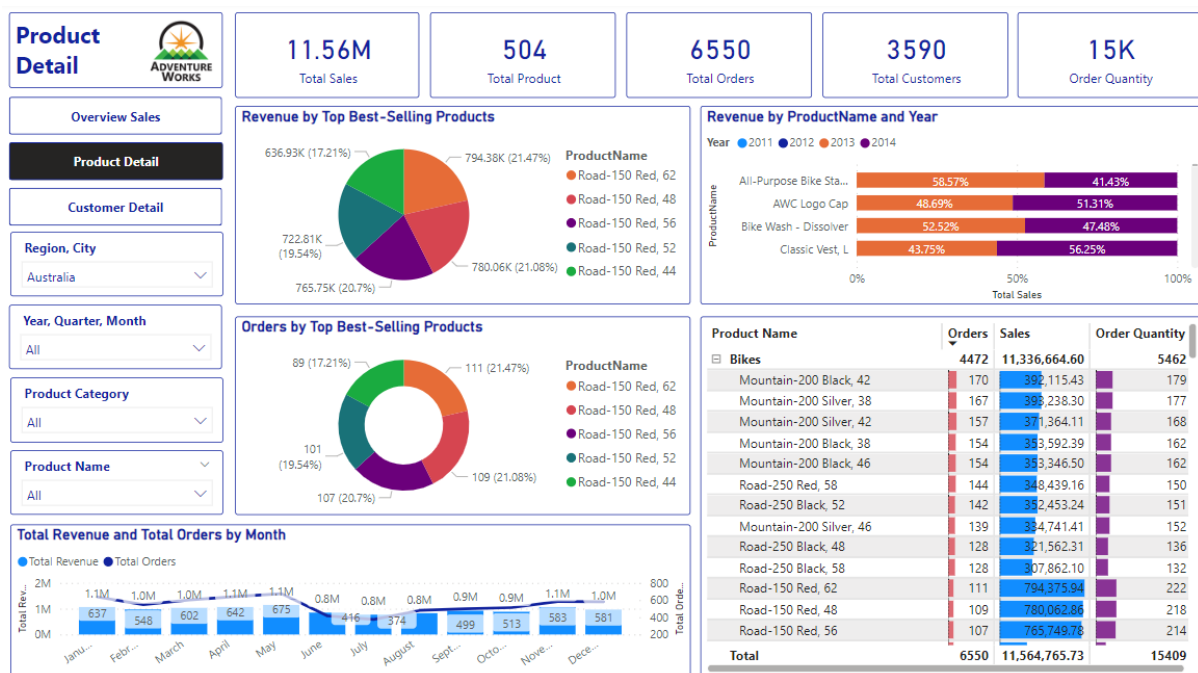


Figure 4-14. Total Revenue and Total Orders by Month in 2013

(Source: Authors)

The revenue and total order chart by month helps the business identify revenue trends for each product category. As bicycles are the highest-selling product category, let's examine the revenue trend for bicycles by month. In the year 2013, the peak sales for bicycles occurred in November, generating revenue of 2.4 million, while the lowest sales were recorded from February to April, with revenue at 0.8 – 0.9 million. It is noticeable that in the early months of the year (from January to May), customers tend to purchase fewer bicycles. From May onwards, bicycle sales gradually increase, indicating a growing demand from consumers. The higher sales figures from June to

December may be attributed to the summer season and year-end festivities, as users typically engage in mountain biking activities during these months. Therefore, the business should intensify its marketing and advertising strategies from May to December to boost sales more effectively.

### The top best-selling products by Region

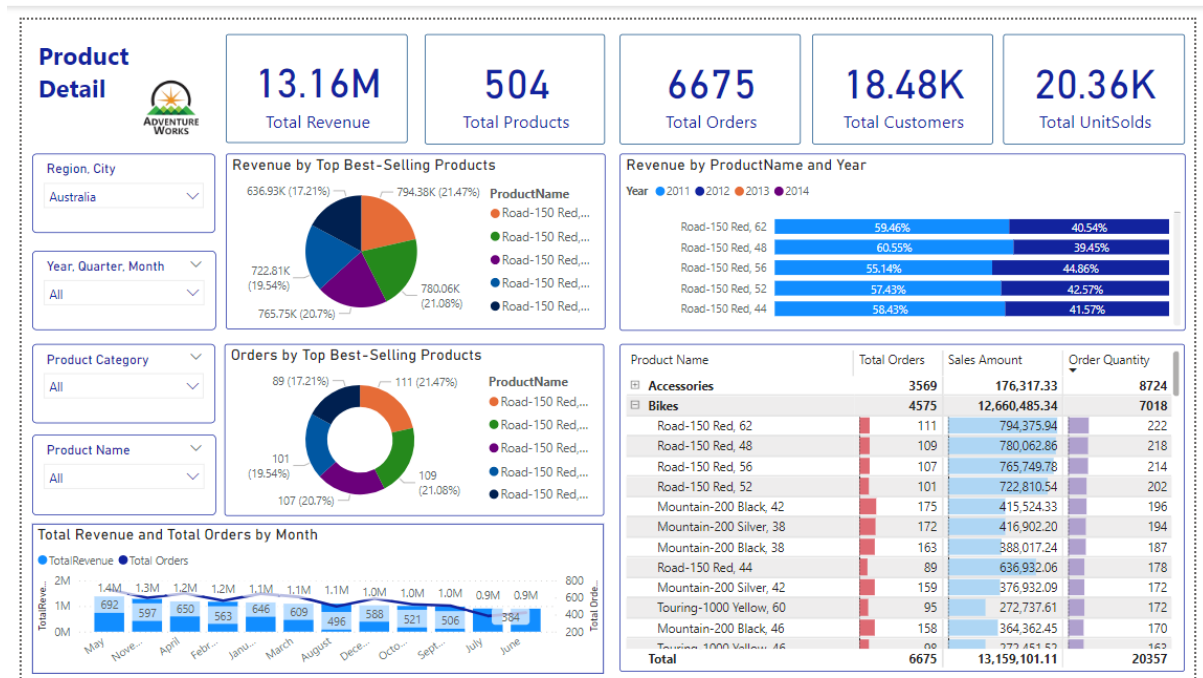


Figure 4-14. The top best-selling products by Region

(Source: Authors)

The chart depicting the top-selling products by region helps businesses discern market trends in specific locations. For instance, in Australia, the best-selling product is the Road-150 Red, generating revenue of 794.38 thousand with 222 units sold. Therefore, Adventure Works should consider exploring opportunities to expand sales by either increasing the availability of this particular product or introducing related items.

### 4.3.2 Report on Customer Detail

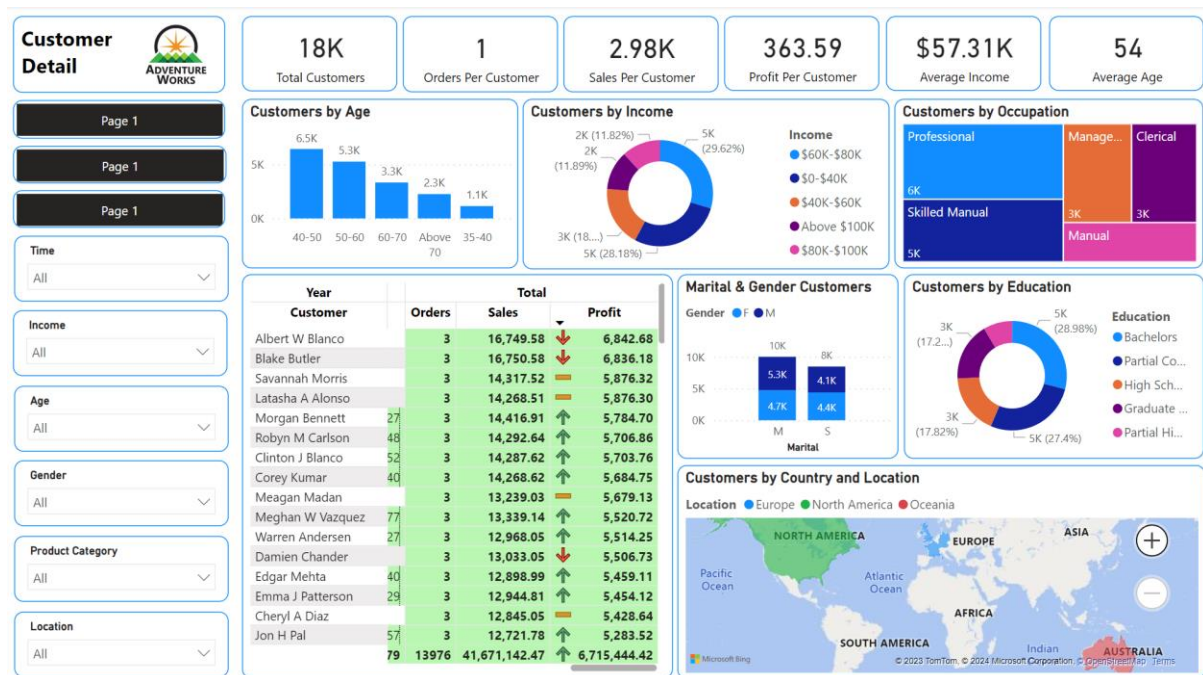


Figure 4-15. Customer Detail Dashboard

(Source: Authors)

The Dashboard provides comprehensive information about AdventureWorks' customers. Users can track information by region, time, or customer segment based on age, income, and gender.

#### Customer Overview

Firstly, the Dashboard offers an overview of customer information. It indicates that there are a total of 18 thousand customers, with an average of 1 order per customer. The average revenue is \$2.98 thousand, average profit is \$363.31, average age is 54, and the average customer income is \$57.31 thousand.

#### Customer by Country

The distribution of customers across regions is displayed on a map, showing that AdventureWorks' customers are mainly concentrated in three main regions: North America, Europe, and Oceania. The statistics and map show that North America has the highest concentration of customers (9,384 people, with 7,814 in the United States and 1,570 in Canada), followed by Europe (5,503 people, with 1,780 in Germany, 1,810 in France, and 1,913 in the United Kingdom), and finally Oceania with 3,590 customers

in Australia. It is evident that Germany has the lowest number of customers, while the United States has the highest. There might be opportunities to expand the customer base in Asia, Africa and South America.

### **Customer by Marital Status, Gender, Age, Income, Education, Occupation**

Furthermore, the Dashboard provides demographic information about customers such as gender, marital status, age, income, education, and occupation. The customer age chart indicates that the majority of customers fall within the 40-50 age range, with the lowest being in the 35-40 age range. This 40-50 age range tends to be especially responsive to Facebook Ads but is most likely not to use TikTok or Instagram. Therefore, you can focus on running advertising campaigns and marketing content through the Facebook platform to reach customers. The income chart shows that the most common income range among customers is \$60 thousand - \$80 thousand, accounting for 29.62%, which is higher than the average customer income. The least common income range is \$80 thousand - \$100 thousand, accounting for 11.82%. In terms of gender and marital status, it is evident that there are more male customers than female customers, with married male customers being the largest group and single female customers having an advantage. They might make purchases not only for themselves, but for their family members as well. Families may be interested in deals that include travel bike racks and group purchases. We can develop product groups for families with preferential prices or promotions for products and accessories included when purchased in groups to promote sales.

### **Customer by sales, order, profit and year**

In terms of orders, revenue, and profit, the total revenue of customers tends to increase with an increase in the number of orders, but the profit is not always guaranteed to increase. It can be observed that occasional customers have an increase in orders and revenue but a decrease in profit. While most customers have an increase in orders and revenue, their profit also increases. The customer with the highest revenue over 4 years is Blake Butler with 3 orders, \$16,750.58 revenue, and \$6,836.18 profit. However, the customer with the highest profit over 4 years is Albert W Blanco also with 3 orders,

\$16,749.58 revenue, and \$6,842.68 profit. The lowest revenue and profit received from a customer over the past 4 years are \$2.29 and \$1.43 respectively with 1 order. From this information, we can determine which customers bring in the most revenue and profits for us, so we can design appropriate care programs, incentives and gratitude to retain customers and drive more purchases. In addition, for customers with low orders, revenue and profits, we will also have activities to remind customers of the brand and products, encouraging customers to come back to buy.

## **CHAPTER 5. CONCLUSION**

### **5.1. Results**

The project successfully created a Business Intelligence system for AdventureWorks' Sales module. This system effectively blends data analytics with visualization to provide detailed insights into sales, meeting both academic standards and practical business needs.

Our team not only grasped the theoretical aspects of BI but also applied them practically. The experience enhanced our capabilities in data collection, warehouse design, and complex data analysis. Learning and discussing with Mr. Thien proved invaluable, offering hands-on insights and bridging the gap between classroom learning and real-world application.

Despite its success, the project highlighted certain limitations, particularly in advanced data analysis techniques and cutting-edge visualization methods. These areas, identified as potential growth points, present an opportunity for further development and exploration in the realm of BI.

In a limited time, we still have many shortages for this project. Firstly, we have just loaded data from the OLTP source to the data warehouse by SSIS, but we have not built a data mart or OLAP cube by SSAS for more complete and consistent data. Secondly, the sales data focuses only on individual customers and ignores the store segment. Thirdly, some attributes are not considered carefully enough and can be left out or redundantly loaded into data warehouse like Person ID of Dim Customer, and dim table for Currency. Finally, our report dashboard has just shown two aspects of sales in detail, including customers and products, while there are many other factors impacting sales performance such as location, and promotion.

### **5.2. Future work**

In the future, our team aims to implement the construction of a Data Mart or OLAP cube using SSAS to load data more comprehensively and consistently. Secondly, we will incorporate the retail store segment for analysis to gain a more holistic view of the business and derive more accurate insights. Thirdly, the team will optimize the data warehouse by carefully considering and selecting data attributes. Lastly, we will analyze



crucial factors such as promotional programs and sales channels to provide deeper insights into sales performance, aiding the business in making more informed decisions for upcoming campaigns.

## REFERENCE

- (1) Geetika, S., & Bharat, A. B. (2014, April). Data Warehouse Designing: Dimensional Modelling and E-R Modelling. *International Journal of Engineering Inventions*, 28-34.
- (2) Mitri, Michel, (2015). Pedagogical Opportunities Of Microsoft's Adventure Works Business Case And Data Model. *SAIS 2015 Proceedings*. 20.
- (3) Korobeyko, T. (2022, July 19). Business intelligence architecture: key components, benefits, and BI team. *Itransition.com*.  
<https://www.itransition.com/business-intelligence/architecture>.
- (4) Stojanovic, F. (2023, November 13). *What is a KPI? Definition, types, examples and best practices*. Databox. <https://databox.com/what-is-a-kpi>.
- (5) *What is a Key Performance Indicator (KPI)? Guide & examples*. (n.d.). Qlik. Retrieved January 4, 2024, from <https://www.qlik.com/us/kpi>.
- (6) *The adventure works cycles: What are the business problems?* (n.d.). Scribd. Retrieved January 4, 2024, from <https://fr.scribd.com/document/469130717/document>.

### PROJECT MEMBERS EVALUATION

No.	Full Name	Tasks	Completion (%)
1	Vũ Thị Phương Anh (Team Leader)	<ul style="list-style-type: none"> <li>- Planning the project outline and managing tasks of the whole team, keep track the deadlines of project.</li> <li>- Defining business requirements, data preparation, project scope and objectives.</li> <li>- Writing report for Business case, ETL Process Overview, Proposing BI Solution, ETL Process.</li> <li>- Collecting and uniting technical files: SSIS Package and SQL query.</li> <li>- In charge of practicing ETL.</li> </ul>	100
2	Đinh Thị Thúy An	<ul style="list-style-type: none"> <li>- Defining business requirements, data preparation, project scope and objectives.</li> <li>- Writing report for Research process, Structure of the report, Schema Types, Business Challenges, Result Structure, Results.</li> <li>- Analyzing business requirements and designing data warehouse structure.</li> <li>- In charge of designing data warehouse model.</li> </ul>	100
3	Nguyễn Thị Thanh Bình	<ul style="list-style-type: none"> <li>- Defining business requirements, data preparation, project scope and objectives.</li> <li>- Writing report for Scope of the project, Overview about BI, Business Requirements, Data Analytics.</li> <li>- Collect report writings and format the project report.</li> <li>- In charge of building Sales Performance Dashboard.</li> </ul>	100
4	Lê Thị Thùy Linh	<ul style="list-style-type: none"> <li>- Defining business requirements, data preparation, project scope and objectives.</li> <li>- Writing report for Objectives, KPIs, Business Requirements, Data Analytics.</li> <li>- Tracking and ensuring the plagiarism of</li> </ul>	100

		project report. - In charge of building Product Detail Dashboard.	
5	Mai Lê Ngọc Trâm	<ul style="list-style-type: none"> <li>- Defining business requirements, data preparation, project scope and objectives.</li> <li>- Writing report for Research objects, Datawarehouse and Datamart, Business Context, Data Analytics.</li> <li>- Collecting and sum up the dashboard.</li> <li>- In charge of building Customer Detail Dashboard.</li> </ul>	100

## CHECK PLAGIARISM

The image shows a plagiarism checker interface. On the left is a preview of a document titled "VIETNAM NATIONAL UNIVERSITY HOCHIMINH CITY UNIVERSITY OF ECONOMICS AND LAW" with the UEL logo and "FINAL REPORT:". On the right is a "Match Overview" sidebar showing a total match percentage of 15% and a list of 8 matches with their respective percentages.

Match Number	Source	Match Percentage
1	Submitted to University... Student Paper	8%
2	www.geeksforgeeks.org Internet Source	<1%
3	www.itransition.com Internet Source	<1%
4	Submitted to Colorado ... Student Paper	<1%
5	Submitted to University... Student Paper	<1%
6	m.moam.info Internet Source	<1%
7	www.scnsoft.com Internet Source	<1%
8	Submitted to American... Student Paper	<1%