**UNIVERSITY OF ECONOMICS AND LAW**
**FACULTY OF INFORMATION SYSTEM**

-----

# FINAL PROJECT REPORT
## SUBJECT: DATA WAREHOUSE AND INTEGRATION

# TOPIC: ANALYZE ADIDAS SALES DATA IN THE USA USING DATA WAREHOUSE.

**Group 5.0 – Class K21416C**

## LECTURER: Mr. Le Ba Thien

**Teaching Assistant: Mr. Nguyen Anh Nhat**

*Ho Chi Minh, December 12th, 2023*

# MEMBER OF GROUP 5.0

| No | Name | Id Student | Contribution |
|---|---|---|---|
| **1** | Hoang Thi Nhat Quyen | K214162151 | 25% |
| **2** | Tran Thi Y Nhu | K214162149 | 25% |
| **3** | Ly Vu Thien An | K214161372 | 25% |
| **4** | Le Tuan Nguyen | K214160993 | 25% |

# COMMITMENT

The project "***Analyze Adidas sales data in the USA using Data Warehouse***" is our research work, performed under guidance and passionate mentoring and support of Msc. Le Ba Thien.

The data collected on the Kaggle platform, simulation results and solutions presented in this project are the results of our analysis, based on the knowledge gained from the course.

All reference sources used in this thesis are clearly published in reputable scientific journals, cited and properly acknowledged.

*Group 5.0*

# ACKNOWLEDGMENTS

First and foremost, Group 5.0 would like to express my sincere gratitude to Msc. Le Ba Thien, our supervisor, for his constant guidance and support over the past 2 months of learning. This project would not have been possible without his dedication and wisdom throughout this journey.

I sincerely thank Msc. Thien for imparting invaluable knowledge and insights that allowed us to improve and successfully complete this project to the best of our abilities. His teachings enabled us to gain a profound understanding of the subject.

*Ho Chi Minh, December 12$^{th}$, 2023*

**Group 5.0**

# TABLE OF CONTENT

# LIST OF ACRONYMS

| Acronym | Description |
|---|---|
| ETL | Extract, transform, and load |
| SCD | Slowly changing dimensions |
| MDX | Multidimensional Expressions |
| KPI | Key Performance Indicator |
| OLAP | Online Analytical Processing |
| DAX | Data Analysis Expressions |
| USA | The United States of America |
| DW | Data Warehouse |
| HR | Human resources |
| SSIS | SQL Server Integration Services |
| EDW | Enterprise Data Warehouse |
| ODS | Operational Data Store |
| DM | Data Mart |

# LIST OF TABLES

# LIST OF FIGURES, CHARTS

# ABSTRACT

Currently, the significant changes in the business landscape after the Covid-19 pandemic have created numerous opportunities and challenges for businesses. Marking the era of technology adoption in business, companies are implementing data analysis processes alongside building their business strategies. Business Intelligence is a widely used and effective approach for businesses to utilize data for their operations. An important component in this process is the Data Warehouse, which serves as the foundation for the subsequent data analysis. This study presents an experimental model built for educational purposes and aims to provide a clear understanding of the role of Data Warehouse in the business model. Based on a dataset of nearly 10,000 transaction records from Adidas and their retailers, the study constructs a Data Warehouse and performs Data Modeling for data analysis purposes. Additionally, it includes an evaluation of the analysis results based on the research findings. Furthermore, the study proposes enhancements to the application of the subject matter it has explored.

*Keywords: Data Warehouse, sales data analysis, OLAP*

# CHAPTER 1. OVERVIEW OF THE PROJECT

Chapter 1 presents the reasons for choosing the topic, the research object and scope, along with the research methods combined with the research process of the topic. Finally, it clarifies the research objectives of the project.

## *1.1. Reason for choosing the topic*

Adidas has been one of the biggest names in the sportswear industry over the years, thanks to the relentless journey of founder Adolf Dassler. [1] This is also considered the best-selling fashion brand of all time, with a record high sales figure in 2019 of 23,640 million euros. However, in the period 2020 - 2021, this number will seriously decrease by more than 5 million euros and is the lowest in the last 5 years [2]. So a close look at their sales data in the United States - Adidas' largest market during this period, helps to gain a clearer view of their business in this million-dollar market and identify the Factors affecting Adidas sales.

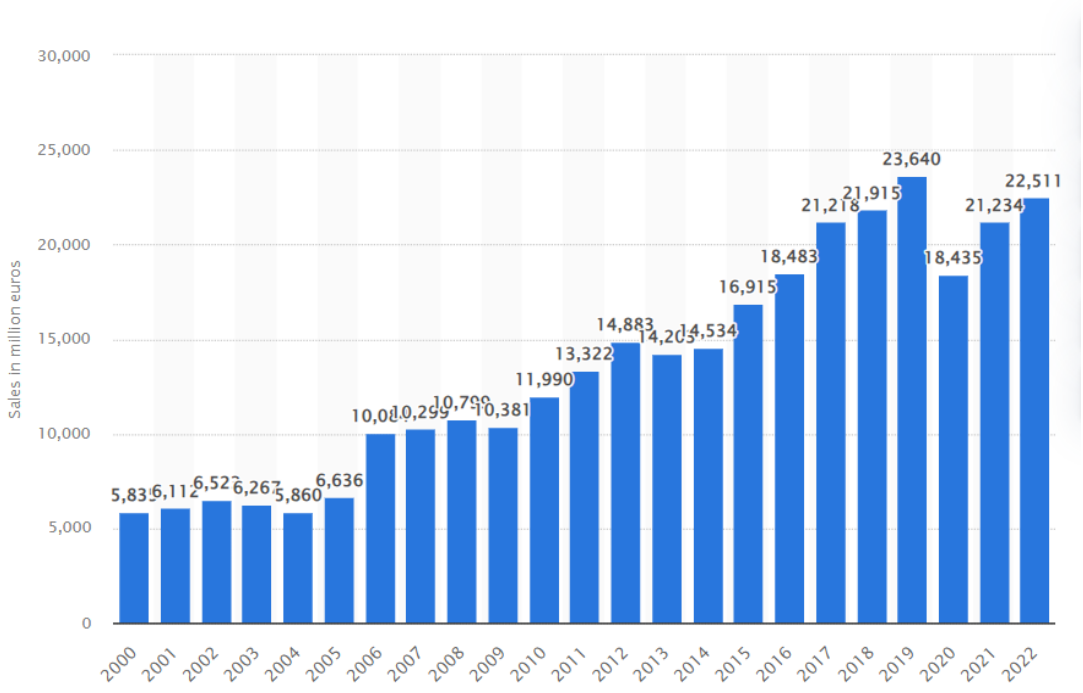



*Figure 1. The adidas Group's net sales worldwide from 2000 to 2022 [3]*

2020 is a special year in world history when the COVID-19 pandemic spread and broke out, the whole world was blocked, the economy stagnated,... In the December edition, Time called 2020 a year “the worst year in history” [4]. Adidas was also not left out of the biggest economic downturn in more than a century when business disruptions due to the COVID-19 pandemic caused profits to drop by 78% in 2020. Not only COVID-19, but also COVID-19. -19 also changed consumers' buying habits from direct to online, as well as the growing demand for outlet goods,... Examining these changes through sales data of Adidas can provide insights into the impact of the pandemic on the sportswear business as well as assist the company in coming up with better growth strategies going forward.

In parallel with the actual situation of the world market, the project also applies the data warehouse to analyze the business situation. Using a Data Warehouse to conduct sales data analysis is an important aspect of the modern data revolution [5]. This approach delivers accurate and reliable results from real-world data, thereby allowing businesses to make more informed decisions. This helps data analysts or Adidas departments have a more comprehensive view of the US business market.

Based on those objective motivations, this thesis proposes a data warehouse application to analyze and understand Adidas' business data in the US in the period 2020 - 2021. Besides, this solution promises will enhance our understanding of Adidas' business model, the consequences of the COVID-19 pandemic, and how the brand is overcoming challenges in the US market. This could benefit not only Adidas but also the wider sports and e-commerce industries.

### *1.2. Objectives of the study*

From the above practices, the project analyzed Adidas's sales data in the US in the period 2020-2021 to have a more comprehensive view to improve the brand's business strategy. Specific goals include:

- **Objective 1**: Apply data warehouse and OLAP queries to understand Adidas' business situation in the US during this volatile period

- **Objective 2**: Evaluate Adidas' sales performance in the US during the period 2020 - 2021.
- **Objective 3**: Research and analyze the sales market situation of Adidas in the US by geographical region.
- **Objective 4**: Propose an effective business strategy for Adidas in the US in the coming time.

### *1.3. Objectives and Scope of work*

- **Objectives**: The research was conducted on a 9,468 rows data set about business activities of sports products of the Adidas brand in the US market
- **Scope**:

- **Time scope**: The study recorded data during the period from January 1st, 2020 to December 31st , 2021
- **Space scope**: The study focuses on sales data in major cities across all fifty States in five different regions of the United States.

### *1.4. Research Methods*

The study analyzed a data set of Adidas company's sports product sales revenue collected secondary from reputable sources on the Internet. Measure the revenue of Adidas sports products business by region in 2020 and 2021.

- ***Method 1***: Research and collect data on Adidas sports product sales revenue in 2020 and 2021, determined based on geographical location characteristics (Region, State, City) . Combine learning about business performance, factors affecting customers' purchases of sports products, and make recommendations for data analysis.

Qualitative research method used in this study aims to collect insights about customers and analyze factors that cannot be measured numerically related to customer needs.

- ***Method 2***: After collecting the primary and secondary data necessary for analysis and research, build a data warehouse of Adidas sports product sales revenue, explore the data using OLAP (Online Analytical Processing) model and perform calculations, query information and data from the built data warehouse, thereby making accurate judgments about the business situation.

## *1.5. Implementation process.*

Research explores and analyzes data based on a total of 6 steps, the steps are as follows:

***Step 1: Identify and research the topic***

In this step, the research team focuses on understanding information, financial situation effects, influencing factors,... related to the topic of Adidas sports product revenue. In addition, collect related articles and studies or have similar goals to find limitations of previous studies, as a basis for developing applicable research directions for businesses. higher career.\

***Step 2: Collect data***

After determining the research topic and the desire for a quality output, collect data on Adidas sports product business revenue within the two years 2020 to 2021, the data includes the following information: related to business scope, profits, product information, order information,...

***Step 3: Preprocess data***

The data set just collected from many different sources will undergo raw processing, bringing it to the most consistent, reasonable and balanced form for later analysis steps.

***Step 4: Build a data warehouse and perform information retrieval to serve the business analysis process.***

With the data sets processed, ready to be imported to build a data warehouse containing all the information that needs to be analyzed. Build an OLAP Cube model from a data warehouse and perform queries and information retrieval using basic

operations on it, in order to provide analysis and evaluation of the accuracy and effectiveness of the model.

***Step 5: Present and visualize analysis results***

From analysis, calculation and measurement, all research results on sales revenue by region and type will be visualized into a Dashboard to easily grasp analytical information.

***Step 6: Propose business solutions for businesses***

And finally, after analyzing and exploring data, combined with qualitative research results, the team proposes suitable business solutions for businesses.

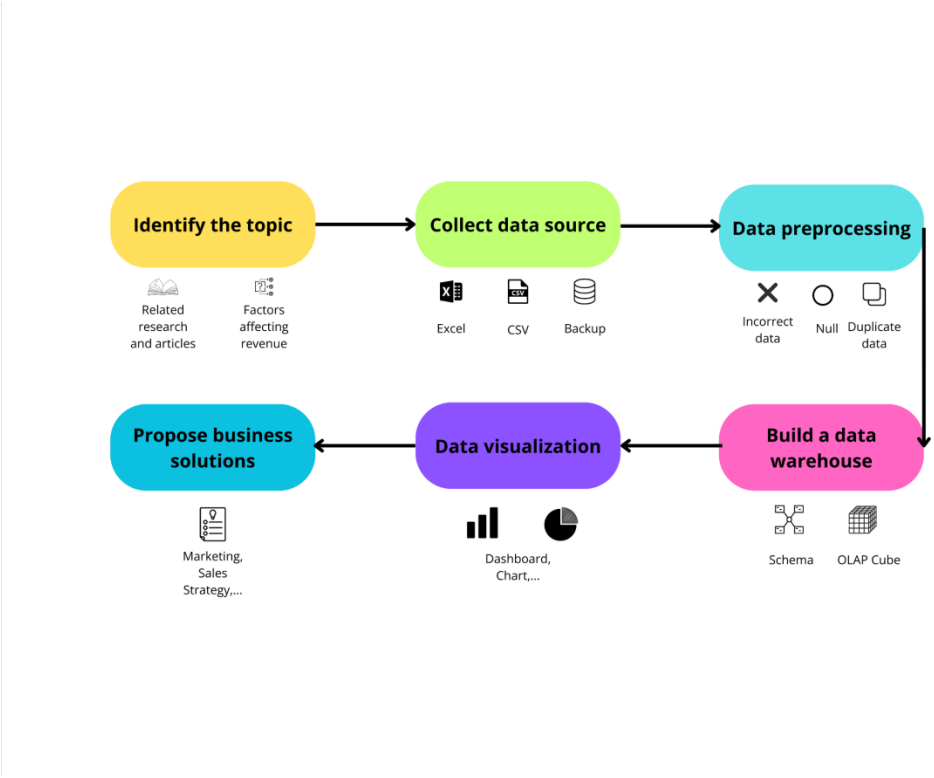



*Figure 2. Research process.*

## *1.6. Structure of the report.*

The report structure consists of 6 chapters, with the following content:

**Chapter 1: Introduction**

The first chapter provides an overview of the research topic, including the rationale for choosing the topic, objectives, research scope, and the target audience. It also

provides information about the research methodology, including data collection methods and the execution process.

**Chapter 2: Literature Review**

This chapter summarizes and evaluates 12 previous research papers and the current state of research on the topics of sales data analysis, data warehousing, OLAP, and sales in the fashion industry. Based on this, it assesses the direction of the study and the methods, theories, and techniques that will be applied.

**Chapter 3: Theoretical Background**

This chapter introduces the theoretical foundation with concepts such as Data Warehouse, ETL (Extract, Transform, Load), OLAP Cube (including the MDX query language). Additionally, the report addresses trend analysis and prediction in fashion sales, supported by relevant models.

**Chapter 4: Proposed Model And Experiment**

This chapter presents a proposed model to address the specific problem, while identifying the requirements and key performance indicators (KPIs). It provides details on the data sources, Extract, Transform, Load (ETL) process, and the structure of the data warehouse built in this research. The chapter also evaluates the level of correlation between the problem requirements and the experimental model.

**Chapter 5: Experimental Results**

This chapter focuses on presenting detailed results from the experimentation process, using query languages like MDX, DAX, and other tools to query the OLAP Cube. Through the specific requirements from Chapter 4, this chapter delves into decoding information from the OLAP Cube, diagnoses trends and characteristics of customers, as well as relationships with other business factors, thus providing important insights and understanding from the analyzed data.

**Chapter 6: Conclusion and Recommendations**

Chapter 6 is the conclusion of the research, providing key findings and suggesting future development directions.

# CHAPTER 2. LITERATURE REVIEW

In this chapter, the study explores and analyzes the empirical results of articles and previous related studies using keywords. From there, we can evaluate previous achievements on this topic as well as recognize the limitations of researching this topic.

## *2.1. Sales data analysis.*

Current economic development poses great challenges for businesses in analyzing sales data to understand customer files and develop business. *Data mining in sales data grouping* research uses qualitative methods using document review and observation with various types of secondary data to analyze sales data.

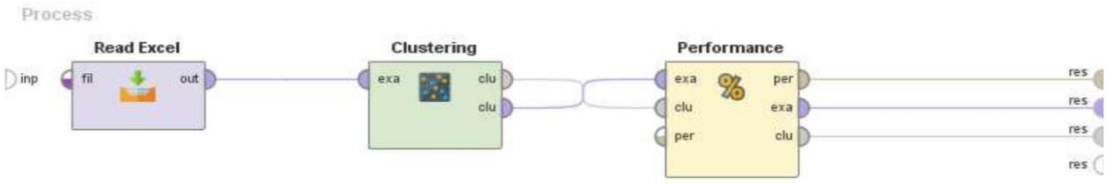



*Figure 3. Research implementation process [6]*

Retail analytics is a term that encompasses various factors that support decision-making in the retail business. Typically, this includes data collection and storage (data warehousing), data analysis involving some statistical or predictive modeling, and decision-making. Traditionally, data analytics has been limited to tracking and visualizing a few key performance indicators (KPIs) [7]. Factors are considered and evaluated such as consumers, advertising, products, personnel… Retail analytics is used by retailers to make better commercial use of that data and increase the value of their business [8].

Traditional retailers face significant commercial competition from online stores, especially at this time of pandemic restrictions. Both shopping malls and stores are now striving to provide more personalized and customer-centric in-store shopping. They are trying to better understand customer behavior to improve and personalize the shopping

experience. To learn more about their customers, retailers are using store-based technologies that essentially mimic e-commerce behavioral tracking and in some cases exceed it [9].

### *2.2. Data Warehouse*

Data warehouses are now widely used to provide strong support for analysis and decision support.

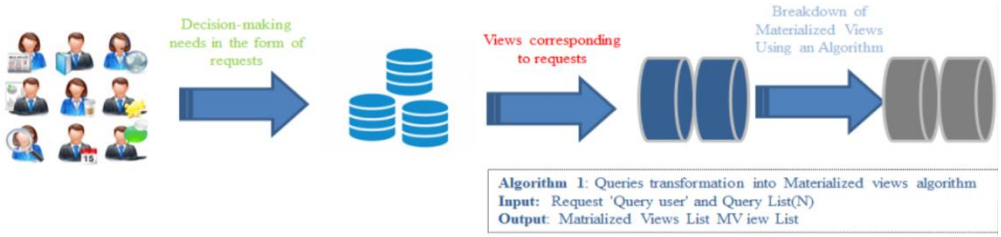



*Figure 4. Generating Materialized Views Attached to Contexts.*

*The Linking context to data warehouse* [10] design study conducts data warehouse design for non-specialists to suit their context. We have proposed a context representation and contextual data collection model as well as a method to create a DW model. Resolving conflicts related to the heterogeneity of both context attributes and data extracted from the data source was achieved using ontologies. Using contextual data makes it possible to automatically create custom blocks that fit the decision maker's context.

Research on Data warehouse for analyzing music sales on a digital media store builds a data warehouse using the Kimball method. The process of extracting, transforming, and loading data (ETL) of child tables needed for data analysis.

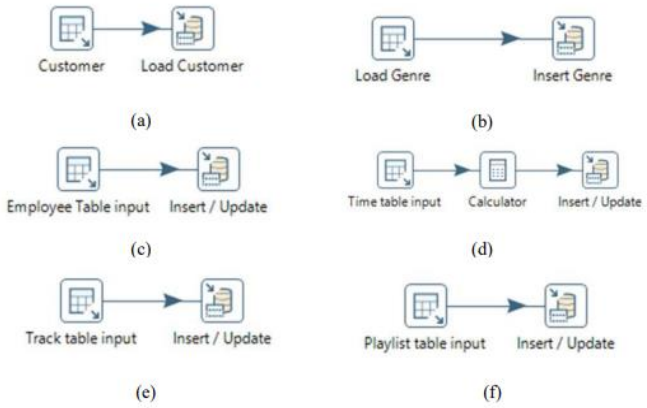



*Figure 5. ETL process for dimension table (a) CustomerDim, (b) GenreDim, (c) EmployeeDim, (d) TimeDim, (e) TrackDim, (f) PlaylistDim*

The author commented that creating reports on the data warehouse is much faster and reports can be generated based on many different parameters. The data warehouse can quickly display all data and sales history, customer shopping preferences, and which products are the best sellers. These reports can help management departments and company executives make quick and accurate data-driven decisions.

The study *Data Warehouse and Data Quality – An Overview* [11] explored data quality in data warehouses (DWs) in depth. Through research. The author emphasized that low-quality data is not only inaccurate data but also includes data that is accurate but has inappropriate semantics. This study provides an overview of data quality issues and opens up new research directions in determining and maintaining data quality in DW. This research is mainly theoretical and emphasizes the quality of input data of the Data Warehouse.

### *2.3. OLAP*

Many members of senior management are in dire need of high-quality information over a variety of time periods. Most organizations are facing some problems related to the data information gap. Therefore, the study Sales System for Golf Course products using Data warehouse and OLAP Technology was born to carry out the mission of conducting a data warehouse using OLAP queries to analyze golf course product data. In

the multidimensional model, golf course product data is organized into multiple dimensions, and each dimension contains multiple levels of abstraction defined by a conceptual hierarchy. In this system, all OLAP operations are used. By using this system, users can investigate sales information quickly. Furthermore, effective sellers will get more customers and have more business because users can decide when and where to get products for their relevant sales problem using this system.

The study *Improve HR decision-making based on data mart and OLAP* [12] has opened a new approach to using human resources data to support decisions in the field of human resource management. Using tools such as SQL Server and Microsoft Excel Pivot Table, the study implemented and analyzed HR "data marts", thereby providing insight and decision support to managers. The ability to combine historical data and advanced techniques such as AI and data mining, opens up the possibility of more accurate analysis and prediction and suggests future directions such as building an HR enterprise data warehouse and comprehensive application of data mining techniques.

The study *Implementing Sales Decision Support System Using Data Mart Based on OLAP, KPI, and Data Mining Approaches* [13] introduces a sales decision support system, combining data warehouse with OLAP and KPI. The special feature is the integration of three data mining methods: decision trees, clustering, and neural networks, to analyze and predict sales data from history. The system supports management in making accurate decisions based on long-term and short-term information, with a web interface that allows easy access to reports and analysis results. This research opens up new directions in using information technology to support effective business decisions.

### *2.4. Sales in the fashion industry*

Demand for fashion products is difficult to predict due to their short life cycle and high volatility due to ever-changing fashion trends. The *AI-Based Fashion Sales Forecasting Methods in Big Data Era* [14] study uses artificial intelligence (AI) methods and AI-based hybrid methods that have proven to be effective in performing sales forecasting fashion in previous studies. With the development and application of big data,

information analysis will certainly bring benefits to forecasting fashion sales, managing operations, and even coordinating the entire fashion supply chain. This article aims to provide an updated review of the more commonly used and effective AI-based fashion sales forecasting methods and to further consider the applicability of these methods in big data.

Analyzing sales data in particular and business data in general not only allows them to gain a deeper understanding of their customer behavior and business development trends, but also allows them to make key decisions. more accurately to improve sales, marketing, customer retention and every other aspect of business. The study *Overview of Big data analytics in E-Commerce* [15] states that: “With the growth of data in the e-commerce field at about 30% to 60% annually, this is a clear opportunity for businesses to grasp and apply big data analysis technology to its business." Shows that business data analysis helps businesses maintain and attract more potential customers in addition to improving business quality and enhancing brand image.

# CHAPTER 3. THEORETICAL BACKGROUND

In this chapter, the research team explores the definitions and theories of concepts and methods used to support the research and analysis process in the study.

## *3.1. Data Warehouse*

### 3.1.1 Definition

Data warehouse is a centralized system for storing and managing data, integrating data from various sources within an organization, then cleaning, standardizing, and storing this data to serve analysis purposes. Data warehouses enable quick access to a large amount of historical –data to identify trends and valuable relationships, thereby supporting effective business decision-making.

### 3.1.2. Main features

Data warehouses have several key features that differentiate them from standard databases.

- Firstly, data warehouses are **"Subject-Oriented"** aimed at providing information for specific topics or domains such as sales, marketing, human resources, etc., instead of focusing on the ongoing operations of the entire organization. This subject orientation helps data warehouses provide in-depth insights into specific issues or business processes.
- Secondly, the data in data warehouses is **"Integrated"** from various sources and synchronized in storage format. The uniformity in naming and encoding makes data retrieval and analysis easier.
- The third feature is **"Non-volatile"** nature - data does not change once it is entered into the data warehouse. They are only readable, with no updates or deletions allowed. This ensures the integrity of data history for trend analysis.
- Lastly, data warehouses support **"Time-Variant"** labeling for each data entry. These labels accurately indicate the historical timing of the data, facilitating timeline-based analysis.

### 3.1.3. Structure

Data warehouses have various architectures depending on the specific needs of each business. The simplest architecture includes a central data warehouse, data sources for input, and users for output to analyze and report. This architecture is easy to deploy but has limitations in scalability and customization.

The **Hub and Spoke** architecture is more common when businesses need to store and analyze separately for each business domain. The central data warehouse is linked to more specialized data marts.

Finally, the **Sandbox** model allows users free access and experimentation without concern for rules, useful for exploring new data.

### 3.1.4. Types of Data Warehouse

Data warehouse is a consolidated, centralized data repository to support analysis and decision making for businesses. Currently, there are 3 main types of data warehouses that have been widely applied in enterprises:

- **Enterprise Data Warehouse (EDW)**: As the name implies, this is an enterprise-level data warehouse, encompassing all business operations of the organization. Typically, global companies use EDW to aggregate and store data from different systems. The functional departments will rely on the EDW to have an overview of the business, supporting risk assessment and long-term decision making.
- **Operational Data Store (ODS)**: Different from EDW, ODS stores transactional data and is updated frequently, daily. After a period, the ODS will stabilize and the data is synchronized to the EDW for long term storage.
- **Data mart**: This is a smaller data warehouse, serving in-depth analyses of a specific department or business unit. For example, customer data mart, store data mart, supplier data mart, etc. allows close storage of data as well as providing detailed and accurate reports on those operations.

### *3.2. Integration Process ETL*

The ETL process comprises three main stages, and it occurs during idle times in the source system [16].

ETL (Extract Transform Load) A process is used to enable companies to move data from multiple sources, reformat and cleanse it, and then load the data into another area for analysis or operational system for support of the organizations business process.

The ETL process consists of three main stages, and ETL processes typically run during off-peak system times.

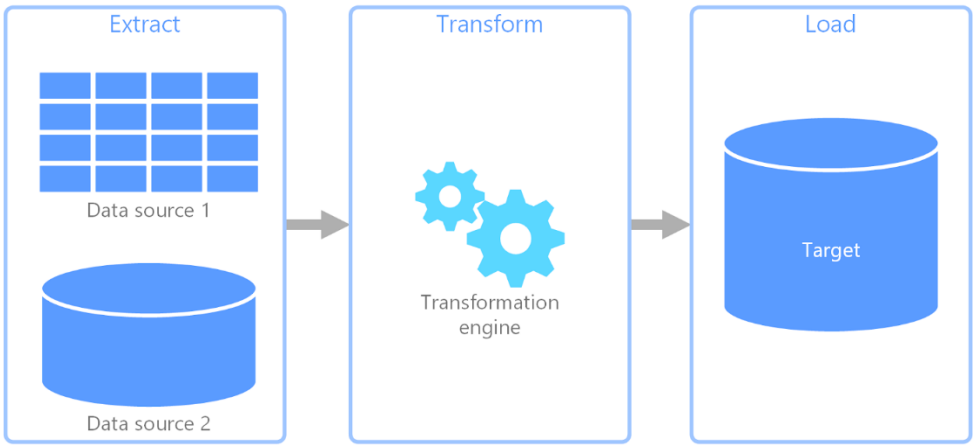



*Figure 6. ETL process*

3.2.1. Extract

This involves extracting non-uniform data from different source systems and transforming it into a polarized data warehouse format, ready for the transformation process. A crucial part of the extraction process includes identifying the exact subset of source data that needs to be sent to the ETL workflow, validating data to confirm whether it is extracted from sources consistently valuable in a specific domain or not.

In practice, this task faces significant technical challenges due to two limitations:

Source data must endure pressure during extraction as other activities take place simultaneously.

Security concerns: administrators may be reluctant to significant interventions in their system configurations, necessitating specific interventions in the source software configuration.

### 3.2.2. Transform

In this step, a set of protocols and methods are applied to data extracted from the source to prepare it for loading into the destination data store room. An important function of data transformation is data purification, aiming to transmit only compatible data to the target. When different systems interact, based on how these systems store data, there can be challenges in their interaction. Certain characters present in one system may not exist in another, leading to quality issues. Therefore, at this step, data must be transformed from the extracted format to the precise, clear, and consistent format compatible with the previously designed data warehouse. In some cases, data may not require transformation.

### 3.2.3. Load

All transformed data is gradually or entirely added to the destination database, which can be a data warehouse or data mart. This process is broadly categorized based on the requirements of different departments. Some data warehouses may replicate existing information groups through summary information; updating segmented data is often done daily, hourly, or weekly.

Activities related to ETL in the data warehouse consume the most time, and developers focus on maximizing integration and transformation capabilities. A well-designed ETL system extracts data from source systems, enforces standards for consistency and data quality, conforms data so that separate sources can be used together, and ultimately provides data in a ready format for application developers to build applications, and end-users to make decisions.

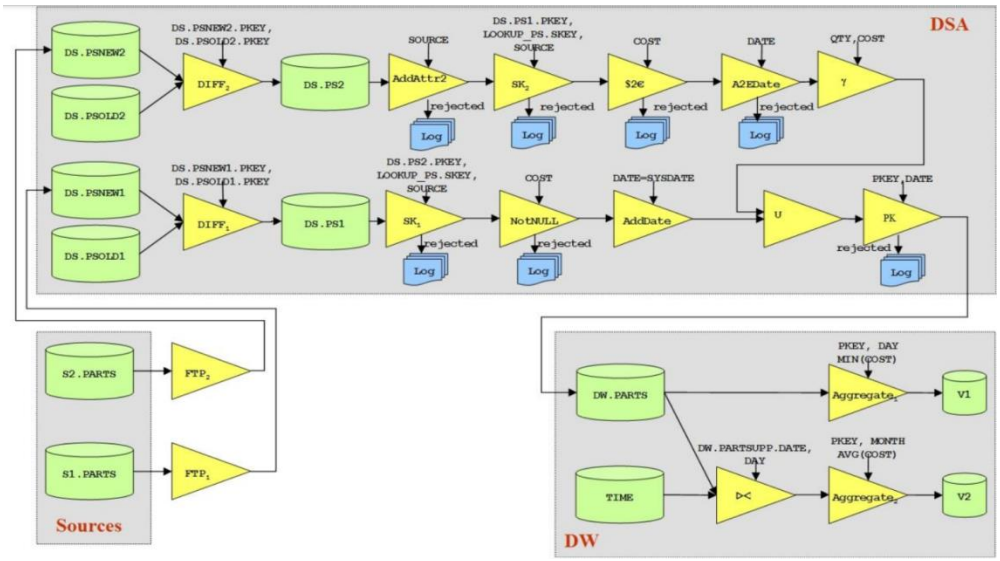



*Figure 7. An example of the ETL workflow.*

## *3.3. OLAP Cube*

3.3.1. Definition

An OLAP cube is a data structure that organizes data into multiple dimensions and measures. A dimension is a category of data that can be used to slice and dice the data, such as time, product, location, customer, etc. A measure is a numeric value that can be aggregated and calculated, such as sales, profit, quantity, etc. An OLAP cube can have multiple dimensions and measures, and each dimension can have multiple levels of hierarchy, such as year, quarter, month, day, etc.

3.3.2. Cube Structure

An OLAP cube is a special data structure that allows the aggregation and analysis of data across multiple dimensions. Specifically, the OLAP cube will aggregate the metrics (called facts, stored in the fact table) based on different dimensions. Common dimensions include time (year, quarter, month, etc.), geography (continent, region, country, etc.), product category, etc.

An OLAP cube typically has 3 dimensions but can have more (called MOLAP - multidimensional OLAP). Each dimension is stored in the corresponding dimension tables. To build an OLAP cube, the original data will first be modeled into fact and

dimension tables according to a star schema or snowflake schema structure. It is then loaded into the OLAP cube for aggregation and analysis.

### 3.3.3. Operations

#### 3.3.3.1. Roll up

In a roll up, the online analytical processing (OLAP) system summarizes data along selected attributes. In other words, this operation displays less detailed data. For example: you can view data for Vancouver, Toronto, New York and Chicago. A roll up operation will provide sales data based on country instead of city, such as Canada, USA.

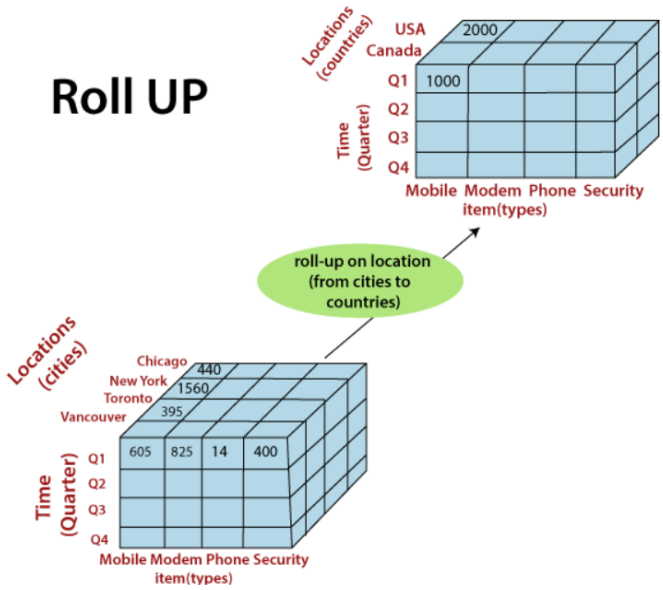



*Figure 8. Roll Up*

#### 3.3.3.2. Drill down

Drill down is the opposite of the roll up operation. Business analysts move down the hierarchy in gradual steps and extract the details they need. For example: they can switch from viewing sales data by quarter to visualizing data by month.

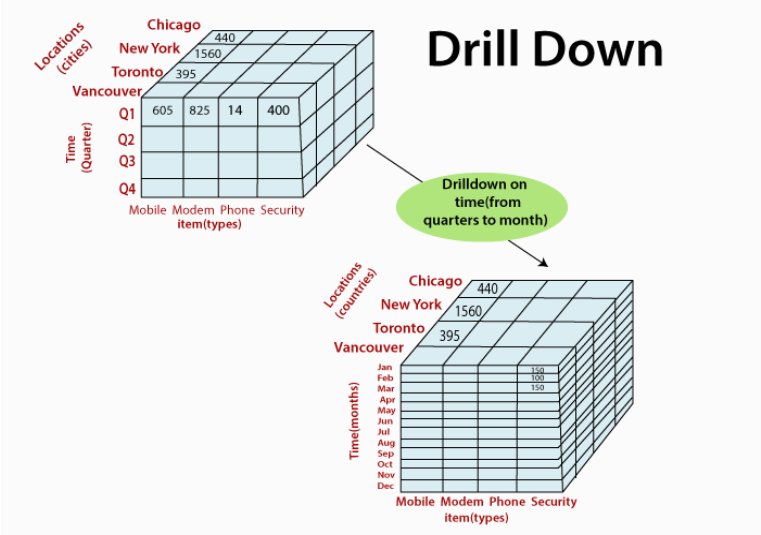



*Figure 9. Drill down*

3.3.3.3. Slice

Data engineers use the slice operation to create a two-dimensional view from the OLAP cube. For example: a MOLAP cube arranges data by city and time unit is quarter. By slicing the cube, data engineers can create a spreadsheet-like table containing data by city for a specific quarter such as “Q1”.

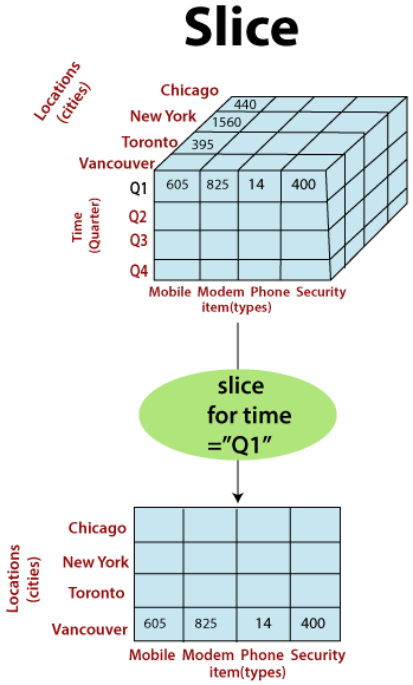



*Figure 10. Slice*

3.3.3.4. Dice

Data engineers use the dice operation to create a smaller sub-cube from an OLAP cube. They identify the required dimensions and construct a smaller cube from the original mega cube.

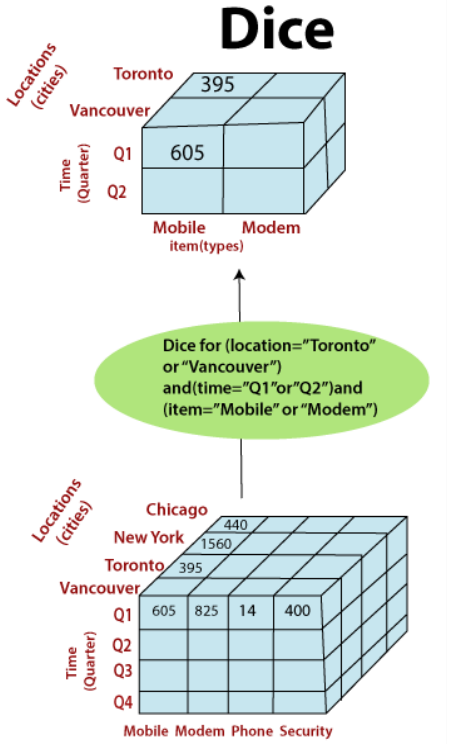



*Figure 11. Dice*

From a city-based mega cube with quarterly timeline, dicing gives us a smaller cube with specific cities like Toronto or Vancouver in Q1 or Q2 for each product category like mobile or modem.

3.3.3.5. Pivot

The pivot operation refers to rotating the OLAP cube along one of its axes to gain a different perspective on the multidimensional data model.

Pivoting the original cube data as above figure changes the axis orientation as follows:

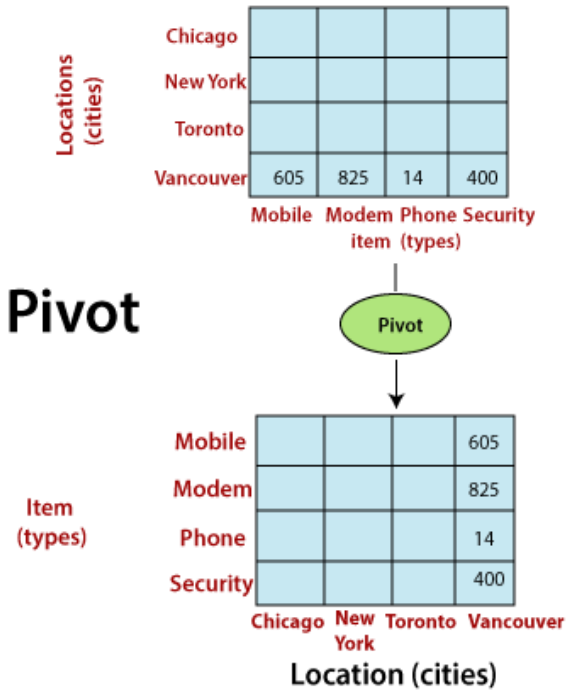



*Figure 12. Pivot*

3.3.4. Query Language - MDX

Multidimensional Expressions (MDX) is a query and data manipulation language for multidimensional data in OLAP databases. It helps define and retrieve multidimensional objects and data from OLAP cubes in an easier and more intuitive way through commands that enable multidimensional querying.

MDX is a specialized query language for OLAP databases, similar to T-SQL for relational databases. Although different in purpose, MDX syntax extends from T-SQL, allowing access to multidimensional data stored in SQL Server Analysis Services OLAP cubes. Hence it is relatively easy to grasp for those already proficient in SQL. MDX uses the concept of tuples to identify and extract data from individual cells or groups of cells in an OLAP cube.

## *3.4. KPIs*

### 3.4.1. Definition of Sales KPIs

Sales KPIs (Key Performance Indicators) are specific sales metrics used to evaluate team and product performance. They are directly connected to company-wide objectives and priorities. While KPIs can be invaluable in monitoring sales success and progress toward longer-term business goals, problems arise when organizations don’t know which KPIs to use.

A good KPI is SMART: Specific, Measurable, Achievable, Realistic, and Timely [17].

### 3.4.2. Advantages and disadvantages of KPIs

***Advantages of KPIs***

- Close Learning Gaps: KPIs help in recognizing and addressing learning gaps among employees. For instance, if a sales goal isn't met, it might indicate a need for additional training in sales techniques. Implementing KPIs allows for the evaluation of employee performance and the effectiveness of training programs.
- Empower Employees to Take Action: Clear and specific KPIs drive employees towards set objectives, guiding their actions. Examples include targets like sending a certain number of sales emails daily or following up with leads within a specified timeframe. These measurable goals help in directing and motivating employee efforts.
- Measure Outcomes and Results: KPIs are essential for measuring the success or failure of specific goals and strategies. They provide a clear metric to assess whether certain actions, like employee training, are paying off. Analyzing KPIs helps in understanding the reasons behind reaching or failing to reach certain objectives.

***Disadvantages KPIs***

- Time Consumption: KPIs often require time to reflect tangible results. Immediate impacts are rarely seen, especially in cases like new employee training programs. It’s important not to rush judgments based on KPIs and allow sufficient time for them to manifest results.
- High Learning Curve: Implementing KPIs can be overwhelming, particularly for those new to this approach. Starting with too many KPIs can lead to confusion and mismanagement, much like trying to do an overly ambitious workout after a long period of inactivity. It’s recommended to start with a manageable number of KPIs and gradually increase as comfort with the process grows.

### 3.4.3. Categories of KPIs (Sale)

In the increasingly competitive context of the retail market, analyzing sales data becomes increasingly important for businesses to understand customers, optimize business strategies, and ensure sustainability. Identify important Key Performance Indicators (KPIs) for sales data analysis.

- **KPIs By Sales Volume:**

The first is KPIs related to sales, one of the most basic indicators of a business strategy.

Total sales is an important indicator to measure a company's financial strength during the study period. At the same time, averaging the sales per transaction provides a view of the average value of each transaction.

- **KPI by profit and performance:**

The project is not only concerned with revenue but also with profits and operational efficiency.

Total operating profit is an important KPI, demonstrating a business's ability to create real value. Operating profit margin and average profit per transaction will help evaluate a business's performance in converting revenue into profit.

- **KPIs By Region and Time:**

Time and area management are important to ensure efficiency and rapid response to market demand. Sales by region, city, and state provide a detailed look at the geography of the market. The average time between the sale and delivery dates is an important business KPI, especially in the increasingly popular online shopping environment.

- **KPIs By Sales Method:**

With the rise of e-commerce and the diversity of sales channels, analytics by sales method is key to better understanding customer preferences.

Sales and sales ratio by mode will help the company evaluate the effectiveness of both retail and online channels.

Overall, carefully combining these KPIs will create a comprehensive view of a company's sales performance, helping them adjust their business strategy and increase competitiveness in a challenging market environment. awake.

## *3.5. Slowly Changing Dimensions (SCDs)*

### 3.5.1. Definition of SCDs

When data is stored in a Data Warehouse, there are attributes in the data dimensions that may change over time. For example, customer information such as address, phone number, or marital status may change after a certain period.

Slowly Changing Dimensions (SCDs) define rules and methods to handle these changes during the data update process. This important concept is used to represent the historical aspect of data in an analytical system. In a data warehouse used for historical data analysis, it is necessary to store different data states. The data is initiated in the operational database, and it will be extracted-transformed-loaded (ETL) into the data warehouse to fit the analytical environment. Dimension attributes are modified over time, and in the data warehouse, it is essential to maintain history. In Facts, attributes that have been modified may be overwritten, so there may not be a need to implement SCD for Fact tables [17].

3.5.2. Types of SCDs

SCD Type 1 overwrites data sequentially. There may be situations where data in the Dimension is missing or needs to be changed. Therefore, when ETL is executed, missing records or changed values are created in the data warehouse. When this new data is ETLed into the analytical database, it needs to be updated in the data warehouse, and the new data will override the old data. The limitation is that historical overwritten information cannot be queried.

SCD Type 2 is the most commonly used type in data warehousing. Since the data warehouse is used for data analysis, historical aspects of the data need to be preserved. When the value of the current record changes, the current record is marked as inactive (inactive – 0), and a new record is inserted. As a result, there will be two records associated with Nguyen Van A in the updated table, but only the latest version is marked as active (active – 1) [18]. Therefore, the result after SCD will still show the previous historical record and can query old results, but it takes up space in the database memory.

Original Record – Dữ liệu gốc

| Cust_ID | Name | City | IsActive |
|---|---|---|---|
| 1001 | Nguyễn Văn A | HCM | 1 |
| 1002 | Nguyễn Văn B | Nam Định | 1 |

Updated Record – Dữ liệu được thay đổi

| Cust_ID | Name | City | IsActive |
|---|---|---|---|
| 1001 | Nguyễn Văn A | HCM | 0 |
| 1001 | Nguyễn Văn A | Hà Nội | 1 |
| 1002 | Nguyễn Văn B | Nam Định | 1 |

*Figure 13. Example of SCD Type 2*

In SCD Type 3, an additional column is added to store the previously appeared value, and the history is maintained in the supplementary column. The limitation of this method is that it only captures the current and previous values, rather than the entire history [18]. SCD Type 3 is commonly used for attributes that undergo infrequent changes, such as name changes or changes with minimal variability.

Original Record – Dữ liệu gốc

| Cust_ID | Name | City |
|---|---|---|
| 1001 | Nguyễn Văn A | HCM |
| 1002 | Nguyễn Văn B | Nam Định |

Updated Record – Dữ liệu được thay đổi

| Cust_ID | Name | Current City | Previous City |
|---|---|---|---|
| 1001 | Nguyễn Văn A | Hà Nội | HCM |
| 1002 | Nguyễn Văn B | Nam Định | |

*Figure 14. Example of SCD Type 3*

By utilizing the history table, the initial main table only contains the latest data, while previous versions are stored in the history table. This ensures data consistency and preserves the historical records, enabling tracking and analysis of changes over time.

Customer Table

| Cust_ID | Name | City |
|---|---|---|
| 1001 | Nguyễn Văn A | Hà Nội |
| 1002 | Nguyễn Văn B | Nam Định |

Customer History Table

| Cust_ID | Name | City | Last_updated_date |
|---|---|---|---|
| 1001 | Nguyễn Văn A | HCM | 11-03-2023 |
| 1001 | Nguyễn Văn A | Hà Nam | 11-05-2023 |
| 1001 | Nguyễn Văn A | Bắc Ninh | 11-06-2023 |

*Figure 15. Example of SCD Type 4*

This method brings several advantages. Firstly, querying the latest data from the main table is fast and efficient, satisfying the requirements for retrieving the most up-to-date information. Secondly, data management becomes easier because previous versions are stored clearly and organized in the history table. This also facilitates coding and data processing, especially when dealing with highly volatile attributes or frequent usage in a large-scale environment.

However, it is important to note certain limitations of this approach. Sometimes, aggregating or combining data between the main table and the history table can be time-consuming and complex. This demands careful design and handling of complex data to ensure consistency and performance in the system.

### *3.6. Trend Analysis and Prediction*

Data analysis is a crucial pillar in the process of making effective decisions and developing strong business strategies. When integrated with OLAP (Online Analytical Processing) models and MDX (Multidimensional Expressions) Query Language, this becomes more flexible and powerful than ever. Data from various sources such as sales, inventory, and production costs can be integrated into an OLAP model to create a multidimensional storage space. Online Analytical Processing (OLAP) is a set of tools to extract useful information from raw data [20]. Charts, graphs, and tables generated from this data help understand trends and relationships.

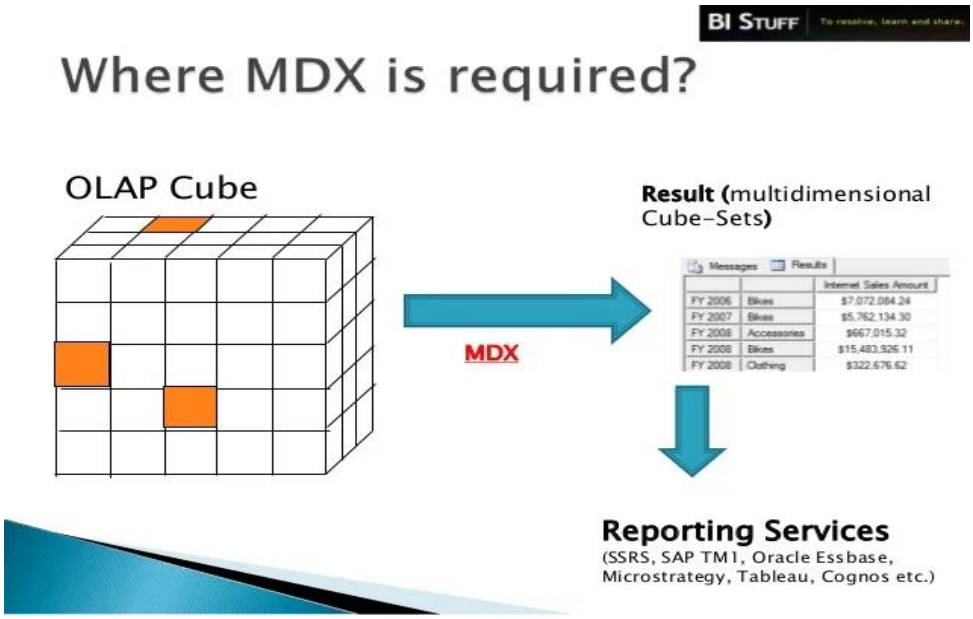



*Figure 16. OLAP and MDX in Analysis*

By combining data from different sources, data analysis can help identify customer characteristics. This includes shopping preferences, purchase frequency, and even factors such as geography that may influence purchasing decisions.

Analysis also helps understand the relationship between sales trends and other business factors such as pricing strategy, advertising campaigns, and brand perception. This is crucial for shaping a comprehensive and cohesive strategy within a business.

Calculation formulas such as totals, averages, or growth rates can be directly applied to data within the OLAP model. This provides flexibility for performing complex operations and detailed analysis. The study is put into practice in the form of MDX query language on the OLAP model. OLAP users need a query language to retrieve values and make the calculations using several measures from fact tables. OLAP tools have their own query languages. One such language is MDX, (Multidimensional Expressions) [20].

MDX allows combining data from multiple dimensions within an OLAP Cube. This helps create a comprehensive view of the data and relationships between factors. For example, to calculate the growth rate of sales over time, MDX can be used to perform this calculation directly on the Cube.

# CHAPTER 4. PROPOSED MODEL AND EXPERIMENT

In this chapter, the study proposes a work model, including from collection stages to data set processing to building a data warehouse to a roadmap for analyzing measurement indicators.

## *4.1. Business problem and proposed model*

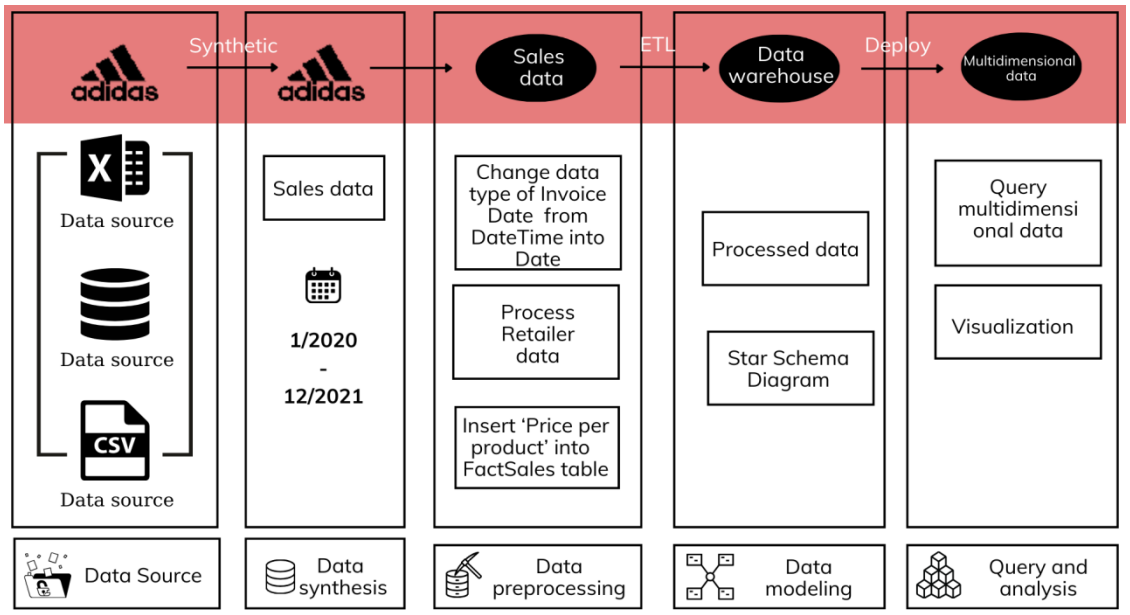


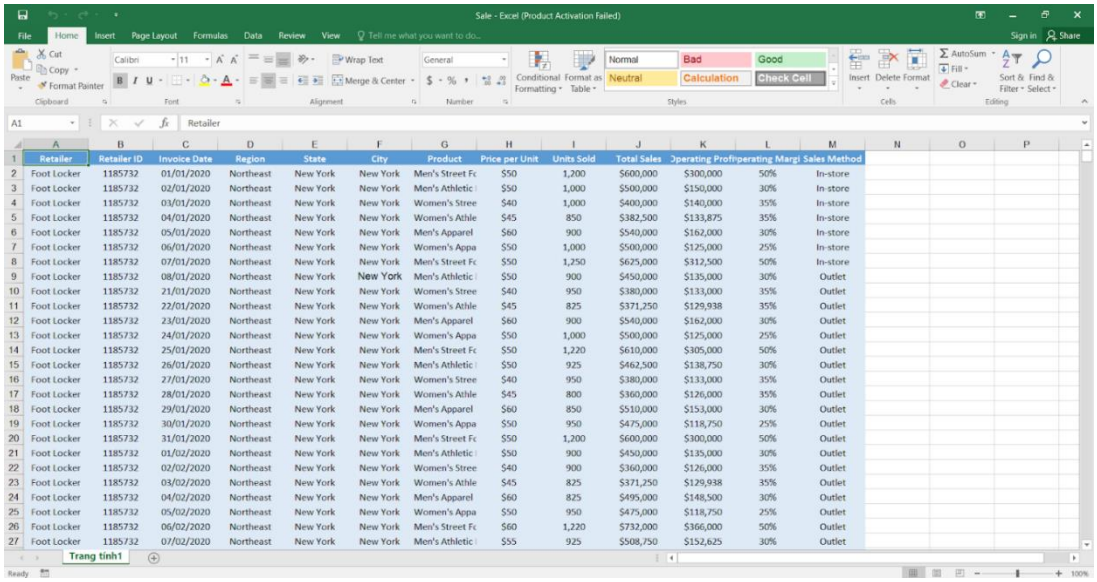

*Figure 17. Prososed model of the research*

Data on Adidas' sales process in the US was collected from the Kaggle platform. Then, the data was integrated into a single source and pre-processed. When the input data was standardized and cleaned, we continued to build a Data Warehouse based on the Star Schema model. Next, we began the process of building Cubes for the Data Warehouse and delved deeper into the data, while presenting information about the data in the form of charts.

## *4.2. Data Preprocessing*

The structure of the initial data files is in various disparate formats, comprising multiple file types. The team conducted ETL (Extract, Transform, Load) processes to

standardize the data format and store it in a Data Warehouse. The Data Warehouse model consists of a Fact table and six Dimension tables. The Fact table is based on the Sale data file, with columns linking values to the Dimension tables.



| Retailer | Retailer ID | Invoice Date | Region | State | City | Product | Price per Unit | Units Sold | Total Sales | Operating Profit | Operating Margi | Sales Method |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Foot Locker | 1185732 | 01/01/2020 | Northeast | New York | New York | Men's Street Fc | $50 | 1,200 | $600,000 | $300,000 | 50% | In-store |
| Foot Locker | 1185732 | 02/01/2020 | Northeast | New York | New York | Men's Athletic | $50 | 1,000 | $500,000 | $150,000 | 30% | In-store |
| Foot Locker | 1185732 | 03/01/2020 | Northeast | New York | New York | Women's Stree | $40 | 1,000 | $400,000 | $140,000 | 35% | In-store |
| Foot Locker | 1185732 | 04/01/2020 | Northeast | New York | New York | Women's Athle | $45 | 850 | $382,500 | $133,875 | 35% | In-store |
| Foot Locker | 1185732 | 05/01/2020 | Northeast | New York | New York | Men's Apparel | $60 | 900 | $540,000 | $162,000 | 30% | In-store |
| Foot Locker | 1185732 | 06/01/2020 | Northeast | New York | New York | Women's Appa | $50 | 1,000 | $500,000 | $125,000 | 25% | In-store |
| Foot Locker | 1185732 | 07/01/2020 | Northeast | New York | New York | Men's Street Fc | $50 | 1,250 | $625,000 | $312,500 | 50% | In-store |
| Foot Locker | 1185732 | 08/01/2020 | Northeast | New York | New York | Men's Athletic | $50 | 900 | $450,000 | $135,000 | 30% | Outlet |
| Foot Locker | 1185732 | 21/01/2020 | Northeast | New York | New York | Women's Stree | $40 | 950 | $380,000 | $133,000 | 35% | Outlet |
| Foot Locker | 1185732 | 22/01/2020 | Northeast | New York | New York | Women's Athle | $45 | 825 | $371,250 | $129,938 | 35% | Outlet |
| Foot Locker | 1185732 | 23/01/2020 | Northeast | New York | New York | Men's Apparel | $60 | 900 | $540,000 | $162,000 | 30% | Outlet |
| Foot Locker | 1185732 | 24/01/2020 | Northeast | New York | New York | Women's Appa | $50 | 1,000 | $500,000 | $125,000 | 25% | Outlet |
| Foot Locker | 1185732 | 25/01/2020 | Northeast | New York | New York | Men's Street Fc | $50 | 1,220 | $610,000 | $305,000 | 50% | Outlet |
| Foot Locker | 1185732 | 26/01/2020 | Northeast | New York | New York | Men's Athletic | $50 | 925 | $462,500 | $138,750 | 30% | Outlet |
| Foot Locker | 1185732 | 27/01/2020 | Northeast | New York | New York | Women's Stree | $40 | 950 | $380,000 | $133,000 | 35% | Outlet |
| Foot Locker | 1185732 | 28/01/2020 | Northeast | New York | New York | Women's Athle | $45 | 800 | $360,000 | $126,000 | 35% | Outlet |
| Foot Locker | 1185732 | 29/01/2020 | Northeast | New York | New York | Men's Apparel | $60 | 850 | $510,000 | $153,000 | 30% | Outlet |
| Foot Locker | 1185732 | 30/01/2020 | Northeast | New York | New York | Women's Appa | $50 | 950 | $475,000 | $118,750 | 25% | Outlet |
| Foot Locker | 1185732 | 31/01/2020 | Northeast | New York | New York | Men's Street Fc | $50 | 1,200 | $600,000 | $300,000 | 50% | Outlet |
| Foot Locker | 1185732 | 01/02/2020 | Northeast | New York | New York | Men's Athletic | $50 | 900 | $450,000 | $135,000 | 30% | Outlet |
| Foot Locker | 1185732 | 02/02/2020 | Northeast | New York | New York | Women's Stree | $40 | 900 | $360,000 | $126,000 | 35% | Outlet |
| Foot Locker | 1185732 | 03/02/2020 | Northeast | New York | New York | Women's Athle | $45 | 825 | $371,250 | $129,938 | 35% | Outlet |
| Foot Locker | 1185732 | 04/02/2020 | Northeast | New York | New York | Men's Apparel | $60 | 825 | $495,000 | $148,500 | 30% | Outlet |
| Foot Locker | 1185732 | 05/02/2020 | Northeast | New York | New York | Women's Appa | $50 | 950 | $475,000 | $118,750 | 25% | Outlet |
| Foot Locker | 1185732 | 06/02/2020 | Northeast | New York | New York | Men's Street Fc | $60 | 1,220 | $732,000 | $366,000 | 50% | Outlet |
| Foot Locker | 1185732 | 07/02/2020 | Northeast | New York | New York | Men's Athletic | $55 | 925 | $508,750 | $152,625 | 30% | Outlet |

*Figure 18. Table Sale format .xlsx (Excel File)*

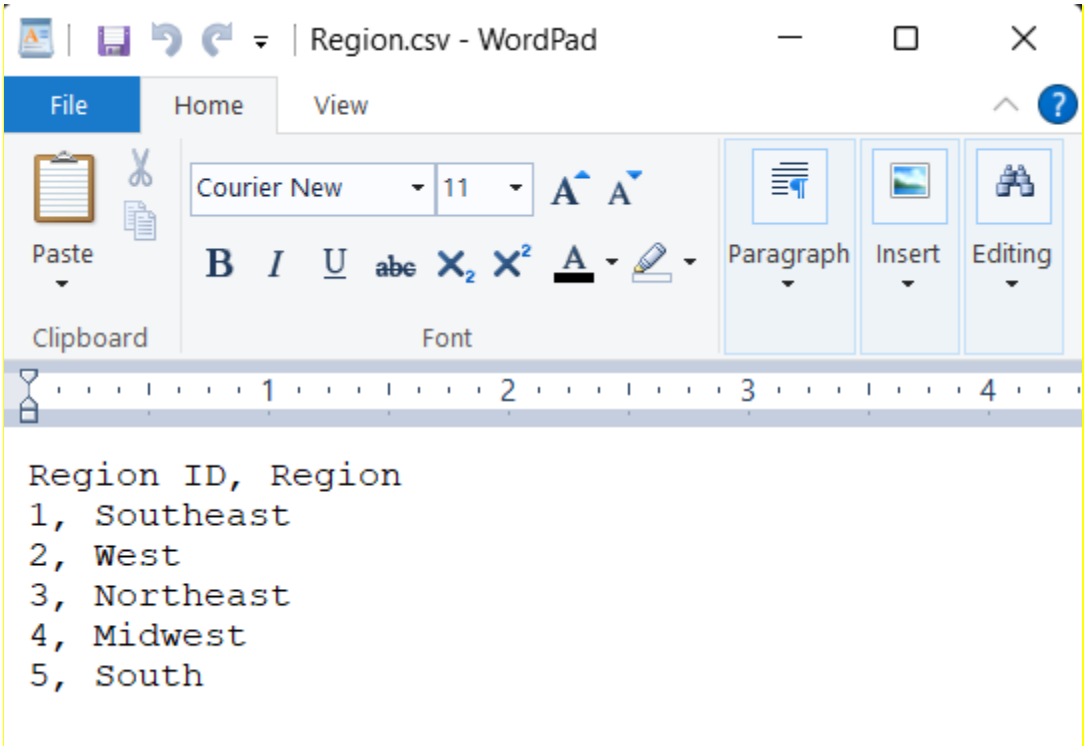


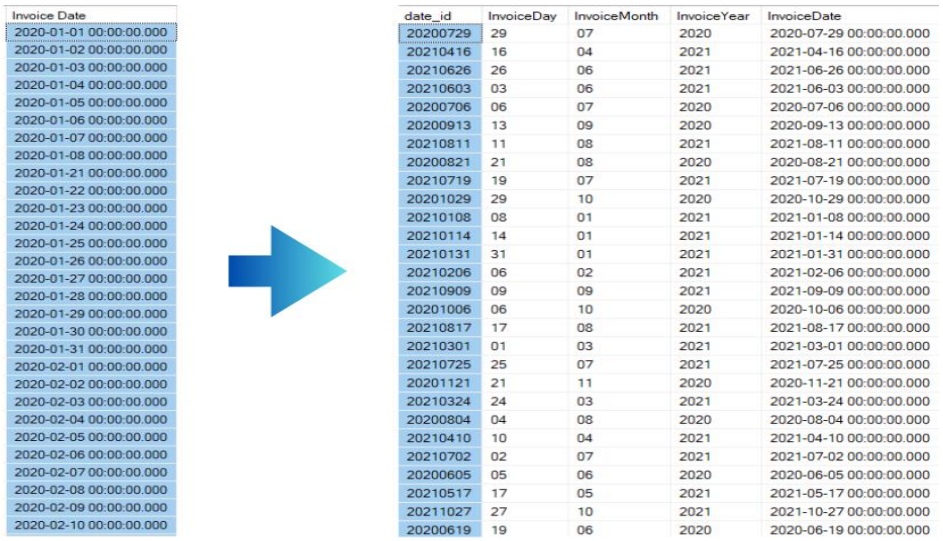

*Figure 19. Table Region format CSV (Flat file)*

| | Retailer | Retailer_ID |
|---|---|---|
| 1 | Amazon | 1185732 |
| 2 | Foot Locker | 1128299 |
| 3 | Foot Locker | 1185732 |
| 4 | Foot Locker | 1197831 |
| 5 | Kohl's | 1128299 |
| 6 | Kohl's | 1185732 |
| 7 | Kohl's | 1189833 |
| 8 | Kohl's | 1197831 |
| 9 | Sports Direct | 1128299 |
| 10 | Sports Direct | 1185732 |
| 11 | Sports Direct | 1197831 |
| 12 | Walmart | 1128299 |
| 13 | Walmart | 1185732 |
| 14 | Walmart | 1197831 |
| 15 | West Gear | 1128299 |
| 16 | West Gear | 1185732 |

*Figure 20. Table Retailer format .bak (OLE DB)*

The data was collected from various sources, including Database (file.bak) and other flat files (csv, excel). Therefore, the team summarized and extracted data from different formats into a common source, the SQL Database, for convenient processing of issues in the dataset. Overall, the data is relatively clean so the data pre-processing process is quite simple and does not take much time.

After consolidating data from different sources, while filtering out duplicate and NULL data rows, we have a dbo.sale dataset with 13 columns and 9648 data rows in the Database. However, during analysis, we realized the dataset still has many shortcomings that need to be standardized.

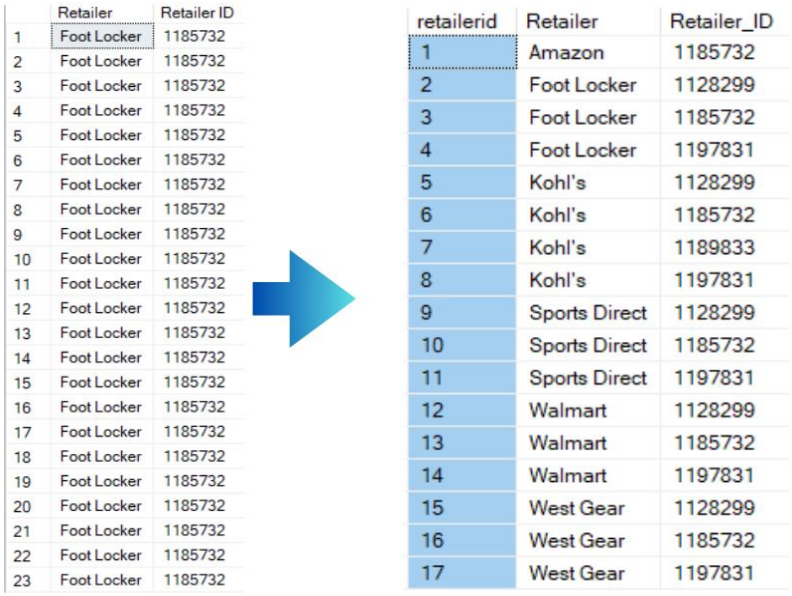


| Invoice Date |
|---|
| 2020-01-01 00:00:00.000 |
| 2020-01-02 00:00:00.000 |
| 2020-01-03 00:00:00.000 |
| 2020-01-04 00:00:00.000 |
| 2020-01-05 00:00:00.000 |
| 2020-01-06 00:00:00.000 |
| 2020-01-07 00:00:00.000 |
| 2020-01-08 00:00:00.000 |
| 2020-01-21 00:00:00.000 |
| 2020-01-22 00:00:00.000 |
| 2020-01-23 00:00:00.000 |
| 2020-01-24 00:00:00.000 |
| 2020-01-25 00:00:00.000 |
| 2020-01-26 00:00:00.000 |
| 2020-01-27 00:00:00.000 |
| 2020-01-28 00:00:00.000 |
| 2020-01-29 00:00:00.000 |
| 2020-01-30 00:00:00.000 |
| 2020-01-31 00:00:00.000 |
| 2020-02-01 00:00:00.000 |
| 2020-02-02 00:00:00.000 |
| 2020-02-03 00:00:00.000 |
| 2020-02-04 00:00:00.000 |
| 2020-02-05 00:00:00.000 |
| 2020-02-06 00:00:00.000 |
| 2020-02-07 00:00:00.000 |
| 2020-02-08 00:00:00.000 |
| 2020-02-09 00:00:00.000 |
| 2020-02-10 00:00:00.000 |

| date_id | InvoiceDay | InvoiceMonth | InvoiceYear | InvoiceDate |
|---|---|---|---|---|
| 20200729 | 29 | 07 | 2020 | 2020-07-29 00:00:00.000 |
| 20210416 | 16 | 04 | 2021 | 2021-04-16 00:00:00.000 |
| 20210626 | 26 | 06 | 2021 | 2021-06-26 00:00:00.000 |
| 20210603 | 03 | 06 | 2021 | 2021-06-03 00:00:00.000 |
| 20200706 | 06 | 07 | 2020 | 2020-07-06 00:00:00.000 |
| 20200913 | 13 | 09 | 2020 | 2020-09-13 00:00:00.000 |
| 20210811 | 11 | 08 | 2021 | 2021-08-11 00:00:00.000 |
| 20200821 | 21 | 08 | 2020 | 2020-08-21 00:00:00.000 |
| 20210719 | 19 | 07 | 2021 | 2021-07-19 00:00:00.000 |
| 20201029 | 29 | 10 | 2020 | 2020-10-29 00:00:00.000 |
| 20210108 | 08 | 01 | 2021 | 2021-01-08 00:00:00.000 |
| 20210114 | 14 | 01 | 2021 | 2021-01-14 00:00:00.000 |
| 20210131 | 31 | 01 | 2021 | 2021-01-31 00:00:00.000 |
| 20210206 | 06 | 02 | 2021 | 2021-02-06 00:00:00.000 |
| 20210909 | 09 | 09 | 2021 | 2021-09-09 00:00:00.000 |
| 20201006 | 06 | 10 | 2020 | 2020-10-06 00:00:00.000 |
| 20210817 | 17 | 08 | 2021 | 2021-08-17 00:00:00.000 |
| 20210301 | 01 | 03 | 2021 | 2021-03-01 00:00:00.000 |
| 20210725 | 25 | 07 | 2021 | 2021-07-25 00:00:00.000 |
| 20201121 | 21 | 11 | 2020 | 2020-11-21 00:00:00.000 |
| 20210324 | 24 | 03 | 2021 | 2021-03-24 00:00:00.000 |
| 20200804 | 04 | 08 | 2020 | 2020-08-04 00:00:00.000 |
| 20210410 | 10 | 04 | 2021 | 2021-04-10 00:00:00.000 |
| 20210702 | 02 | 07 | 2021 | 2021-07-02 00:00:00.000 |
| 20200605 | 05 | 06 | 2020 | 2020-06-05 00:00:00.000 |
| 20210517 | 17 | 05 | 2021 | 2021-05-17 00:00:00.000 |
| 20211027 | 27 | 10 | 2021 | 2021-10-27 00:00:00.000 |
| 20200619 | 19 | 06 | 2020 | 2020-06-19 00:00:00.000 |

*Figure 21. Processing Invoice Date table*

Firstly, for the 'Invoice Date' data column, after being extracted into the Database from the Excel file, it has the 'DateTime' data type. After analysis, the team decided to split the 'Invoice Date' column into columns including date_id; InvoiceDay; InvoiceMonth; InvoiceYear while keeping the 'Invoice Date' column and adding those columns into the Dim.Date table with date_id as the primary key.

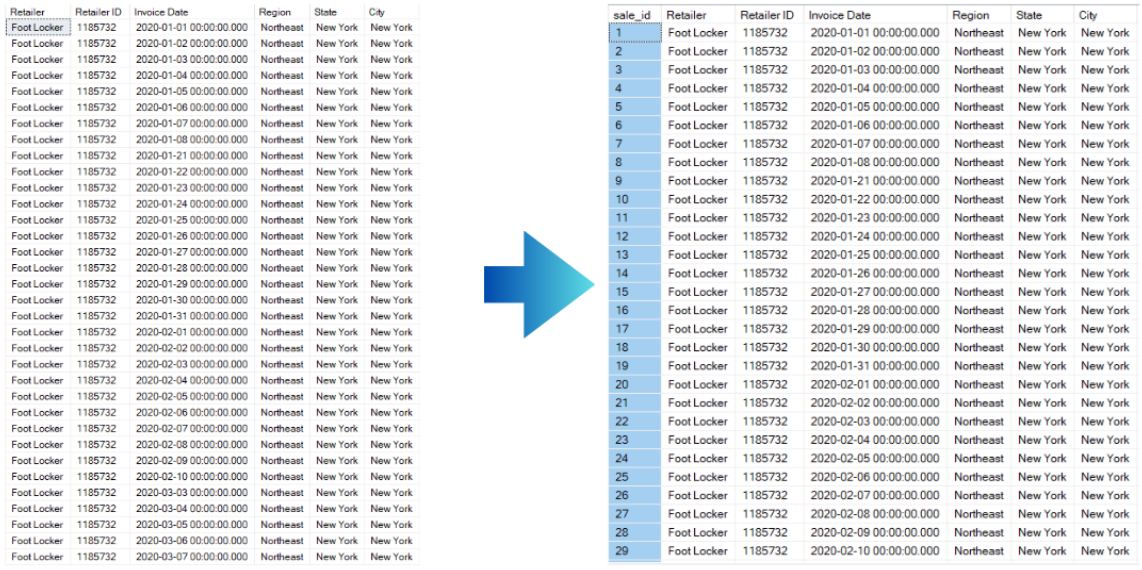


| | Retailer | Retailer ID |
|---|---|---|
| 1 | Foot Locker | 1185732 |
| 2 | Foot Locker | 1185732 |
| 3 | Foot Locker | 1185732 |
| 4 | Foot Locker | 1185732 |
| 5 | Foot Locker | 1185732 |
| 6 | Foot Locker | 1185732 |
| 7 | Foot Locker | 1185732 |
| 8 | Foot Locker | 1185732 |
| 9 | Foot Locker | 1185732 |
| 10 | Foot Locker | 1185732 |
| 11 | Foot Locker | 1185732 |
| 12 | Foot Locker | 1185732 |
| 13 | Foot Locker | 1185732 |
| 14 | Foot Locker | 1185732 |
| 15 | Foot Locker | 1185732 |
| 16 | Foot Locker | 1185732 |
| 17 | Foot Locker | 1185732 |
| 18 | Foot Locker | 1185732 |
| 19 | Foot Locker | 1185732 |
| 20 | Foot Locker | 1185732 |
| 21 | Foot Locker | 1185732 |
| 22 | Foot Locker | 1185732 |
| 23 | Foot Locker | 1185732 |

| retailerid | Retailer | Retailer_ID |
|---|---|---|
| 1 | Amazon | 1185732 |
| 2 | Foot Locker | 1128299 |
| 3 | Foot Locker | 1185732 |
| 4 | Foot Locker | 1197831 |
| 5 | Kohl's | 1128299 |
| 6 | Kohl's | 1185732 |
| 7 | Kohl's | 1189833 |
| 8 | Kohl's | 1197831 |
| 9 | Sports Direct | 1128299 |
| 10 | Sports Direct | 1185732 |
| 11 | Sports Direct | 1197831 |
| 12 | Walmart | 1128299 |
| 13 | Walmart | 1185732 |
| 14 | Walmart | 1197831 |
| 15 | West Gear | 1128299 |
| 16 | West Gear | 1185732 |
| 17 | West Gear | 1197831 |

*Figure 22. Processing Retailer table*

For the data in the 'Retailer' and 'Retailer ID' columns, the team found they were not compatible with each other. This happened because one Retailer has more than one Retailer ID and one Retailer ID has multiple Retailers. To fix this issue, the team decided not to remove the 'Retailer ID' column. Instead, the team would create a retailer id for each data row with the same 'Retailer ID' and 'Retailer', make it the primary key and put it into a new table called Dim.Retailer for data consistency.



| Retailer | Retailer ID | Invoice Date | Region | State | City |
|---|---|---|---|---|---|
| Foot Locker | 1185732 | 2020-01-01 00:00:00.000 | Northeast | New York | New York |
| Foot Locker | 1185732 | 2020-01-02 00:00:00.000 | Northeast | New York | New York |
| Foot Locker | 1185732 | 2020-01-03 00:00:00.000 | Northeast | New York | New York |
| Foot Locker | 1185732 | 2020-01-04 00:00:00.000 | Northeast | New York | New York |
| Foot Locker | 1185732 | 2020-01-05 00:00:00.000 | Northeast | New York | New York |
| Foot Locker | 1185732 | 2020-01-06 00:00:00.000 | Northeast | New York | New York |
| Foot Locker | 1185732 | 2020-01-07 00:00:00.000 | Northeast | New York | New York |
| Foot Locker | 1185732 | 2020-01-08 00:00:00.000 | Northeast | New York | New York |
| Foot Locker | 1185732 | 2020-01-21 00:00:00.000 | Northeast | New York | New York |
| Foot Locker | 1185732 | 2020-01-22 00:00:00.000 | Northeast | New York | New York |
| Foot Locker | 1185732 | 2020-01-23 00:00:00.000 | Northeast | New York | New York |
| Foot Locker | 1185732 | 2020-01-24 00:00:00.000 | Northeast | New York | New York |
| Foot Locker | 1185732 | 2020-01-25 00:00:00.000 | Northeast | New York | New York |
| Foot Locker | 1185732 | 2020-01-26 00:00:00.000 | Northeast | New York | New York |
| Foot Locker | 1185732 | 2020-01-27 00:00:00.000 | Northeast | New York | New York |
| Foot Locker | 1185732 | 2020-01-28 00:00:00.000 | Northeast | New York | New York |
| Foot Locker | 1185732 | 2020-01-29 00:00:00.000 | Northeast | New York | New York |
| Foot Locker | 1185732 | 2020-01-30 00:00:00.000 | Northeast | New York | New York |
| Foot Locker | 1185732 | 2020-01-31 00:00:00.000 | Northeast | New York | New York |
| Foot Locker | 1185732 | 2020-02-01 00:00:00.000 | Northeast | New York | New York |
| Foot Locker | 1185732 | 2020-02-02 00:00:00.000 | Northeast | New York | New York |
| Foot Locker | 1185732 | 2020-02-03 00:00:00.000 | Northeast | New York | New York |
| Foot Locker | 1185732 | 2020-02-04 00:00:00.000 | Northeast | New York | New York |
| Foot Locker | 1185732 | 2020-02-05 00:00:00.000 | Northeast | New York | New York |
| Foot Locker | 1185732 | 2020-02-06 00:00:00.000 | Northeast | New York | New York |
| Foot Locker | 1185732 | 2020-02-07 00:00:00.000 | Northeast | New York | New York |
| Foot Locker | 1185732 | 2020-02-08 00:00:00.000 | Northeast | New York | New York |
| Foot Locker | 1185732 | 2020-02-09 00:00:00.000 | Northeast | New York | New York |
| Foot Locker | 1185732 | 2020-02-10 00:00:00.000 | Northeast | New York | New York |
| Foot Locker | 1185732 | 2020-03-03 00:00:00.000 | Northeast | New York | New York |
| Foot Locker | 1185732 | 2020-03-04 00:00:00.000 | Northeast | New York | New York |
| Foot Locker | 1185732 | 2020-03-05 00:00:00.000 | Northeast | New York | New York |
| Foot Locker | 1185732 | 2020-03-06 00:00:00.000 | Northeast | New York | New York |
| Foot Locker | 1185732 | 2020-03-07 00:00:00.000 | Northeast | New York | New York |

| sale_id | Retailer | Retailer ID | Invoice Date | Region | State | City |
|---|---|---|---|---|---|---|
| 1 | Foot Locker | 1185732 | 2020-01-01 00:00:00.000 | Northeast | New York | New York |
| 2 | Foot Locker | 1185732 | 2020-01-02 00:00:00.000 | Northeast | New York | New York |
| 3 | Foot Locker | 1185732 | 2020-01-03 00:00:00.000 | Northeast | New York | New York |
| 4 | Foot Locker | 1185732 | 2020-01-04 00:00:00.000 | Northeast | New York | New York |
| 5 | Foot Locker | 1185732 | 2020-01-05 00:00:00.000 | Northeast | New York | New York |
| 6 | Foot Locker | 1185732 | 2020-01-06 00:00:00.000 | Northeast | New York | New York |
| 7 | Foot Locker | 1185732 | 2020-01-07 00:00:00.000 | Northeast | New York | New York |
| 8 | Foot Locker | 1185732 | 2020-01-08 00:00:00.000 | Northeast | New York | New York |
| 9 | Foot Locker | 1185732 | 2020-01-21 00:00:00.000 | Northeast | New York | New York |
| 10 | Foot Locker | 1185732 | 2020-01-22 00:00:00.000 | Northeast | New York | New York |
| 11 | Foot Locker | 1185732 | 2020-01-23 00:00:00.000 | Northeast | New York | New York |
| 12 | Foot Locker | 1185732 | 2020-01-24 00:00:00.000 | Northeast | New York | New York |
| 13 | Foot Locker | 1185732 | 2020-01-25 00:00:00.000 | Northeast | New York | New York |
| 14 | Foot Locker | 1185732 | 2020-01-26 00:00:00.000 | Northeast | New York | New York |
| 15 | Foot Locker | 1185732 | 2020-01-27 00:00:00.000 | Northeast | New York | New York |
| 16 | Foot Locker | 1185732 | 2020-01-28 00:00:00.000 | Northeast | New York | New York |
| 17 | Foot Locker | 1185732 | 2020-01-29 00:00:00.000 | Northeast | New York | New York |
| 18 | Foot Locker | 1185732 | 2020-01-30 00:00:00.000 | Northeast | New York | New York |
| 19 | Foot Locker | 1185732 | 2020-01-31 00:00:00.000 | Northeast | New York | New York |
| 20 | Foot Locker | 1185732 | 2020-02-01 00:00:00.000 | Northeast | New York | New York |
| 21 | Foot Locker | 1185732 | 2020-02-02 00:00:00.000 | Northeast | New York | New York |
| 22 | Foot Locker | 1185732 | 2020-02-03 00:00:00.000 | Northeast | New York | New York |
| 23 | Foot Locker | 1185732 | 2020-02-04 00:00:00.000 | Northeast | New York | New York |
| 24 | Foot Locker | 1185732 | 2020-02-05 00:00:00.000 | Northeast | New York | New York |
| 25 | Foot Locker | 1185732 | 2020-02-06 00:00:00.000 | Northeast | New York | New York |
| 26 | Foot Locker | 1185732 | 2020-02-07 00:00:00.000 | Northeast | New York | New York |
| 27 | Foot Locker | 1185732 | 2020-02-08 00:00:00.000 | Northeast | New York | New York |
| 28 | Foot Locker | 1185732 | 2020-02-09 00:00:00.000 | Northeast | New York | New York |
| 29 | Foot Locker | 1185732 | 2020-02-10 00:00:00.000 | Northeast | New York | New York |

*Figure 23. Reference Retailer ID from Retailer table to Sales table*

The last issue the team handled in the above Database was to create a sale_id column for orders in the dataset and make it the primary key for the Fact.Sales table, significantly reducing query time.

After handling the above issues in the dataset, we can move on to building the Database Schema to serve the in-depth analysis process by querying the Cube with the MDX query language.

## *4.3. Experimental Model*

### 4.3.1. Building the Data Warehouse

#### 4.3.1.1. Data Warehouse Bus Architecture Matrix

| Business process | Dimensions: Date | City | Region | Product | Method | Retailer |
|---|---|---|---|---|---|---|
| Revenue | x | x | x | x | x | x |
| Quantity of unit sold | x | x | x | x | x | x |
| Operating margin | x | x | x | x | x | x |
| Return on sale | x | | x | | | |

*Figure 24. Bus Matrix*

#### 4.3.1.2. ETL process (Extract, Transform, Load)

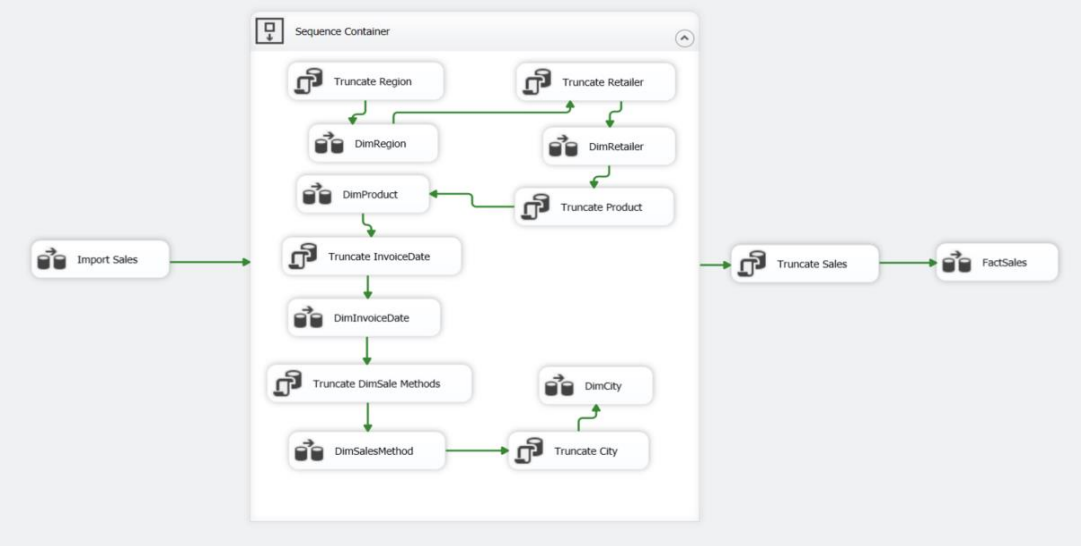



*Figure 25. Overview of the ETL Process using SSIS*

The research identifies the DimRegion, DimRetailer, DimProduct, DimInvoiceDate, and DimSalesMethod tables as Dimension tables. These tables contain descriptive data about objects in the system, such as regions, retailers, products, invoice dates, and sales methods. The FactSales table is referred to as the fact table, containing actual data on sales volume.

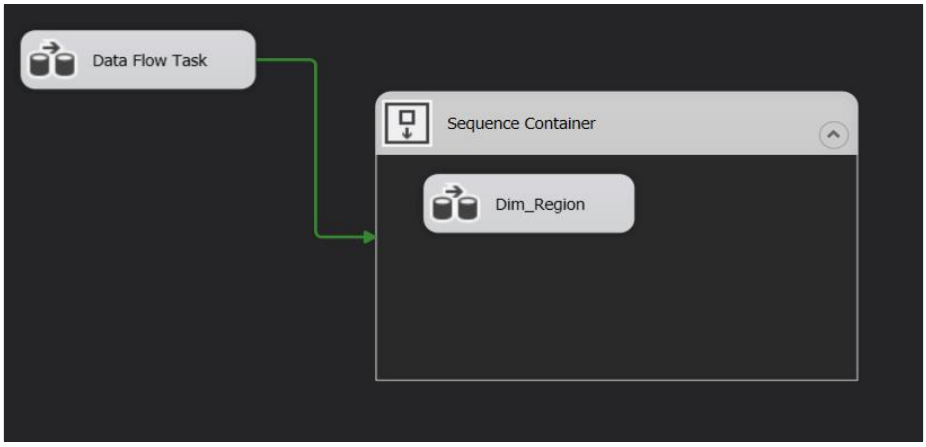



*Figure 26. Create Sequence containers for Dimension tables*

The first step is to import new sales data into the FactSales table. This involves bringing in new sales data from an external source. The table's data includes information about products, retailers, invoice dates, and sales methods, which can establish relationships with Dimension tables through represented values. This is Extract step in ETL process.

Truncate the DimRegion, DimRetailer, DimProduct, DimInvoiceDate, DimSalesMethod, and DimCity tables. This step clears all data in these tables, necessary to initiate a new data cycle.

The final step is to merge the Dim and Fact tables, with detailed steps as follows.

- **ETL Dimension Tables**

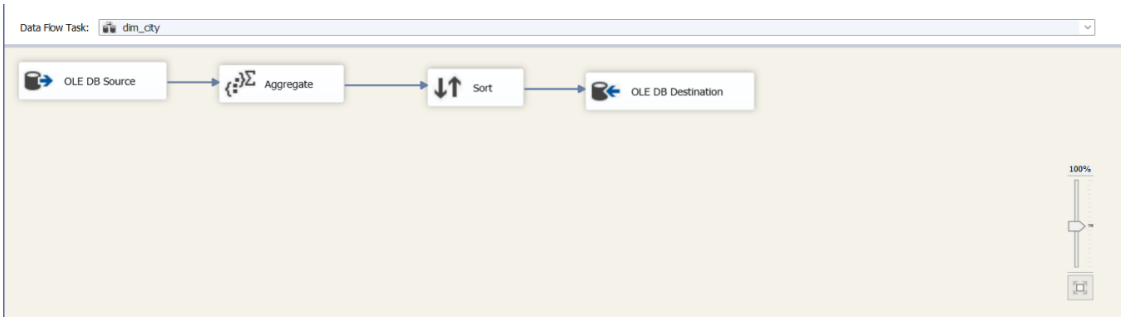



*Figure 27. ETL Dim Table*

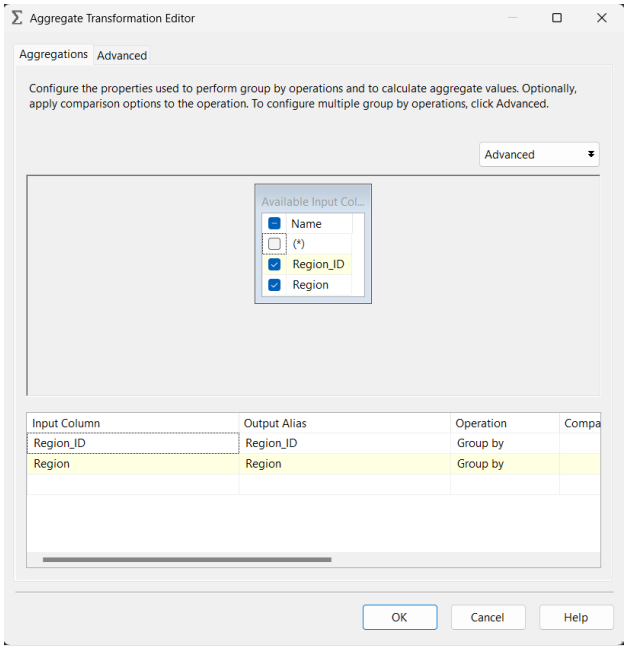



*Figure 28. Aggregate columns in Dimension table*

Calculating data columns using Aggregate Transformation: Selecting the columns to be calculated for transformation, various operations such as sum, group by, etc., can be performed for each column. The returned result will be standardized tables; this is a crucial step in organizing the data, eliminating duplicates, and reducing data size.

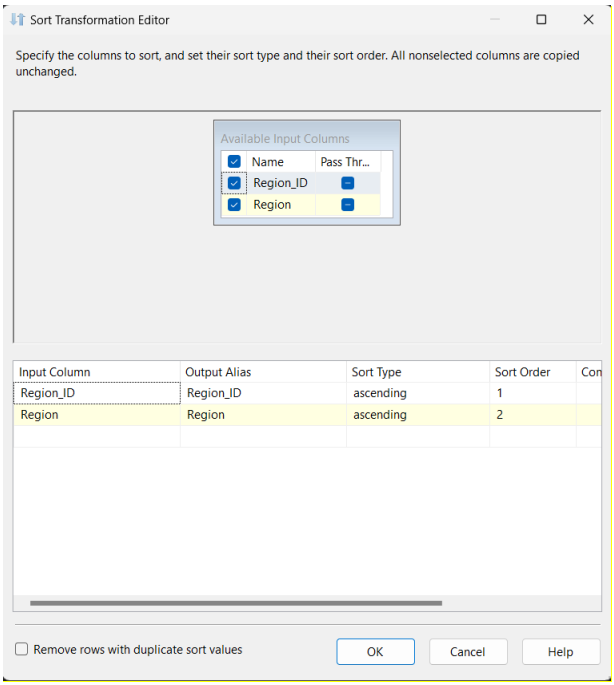



*Figure 29. Sort columns in Dimension tables*

Sort data: by ascending or descending

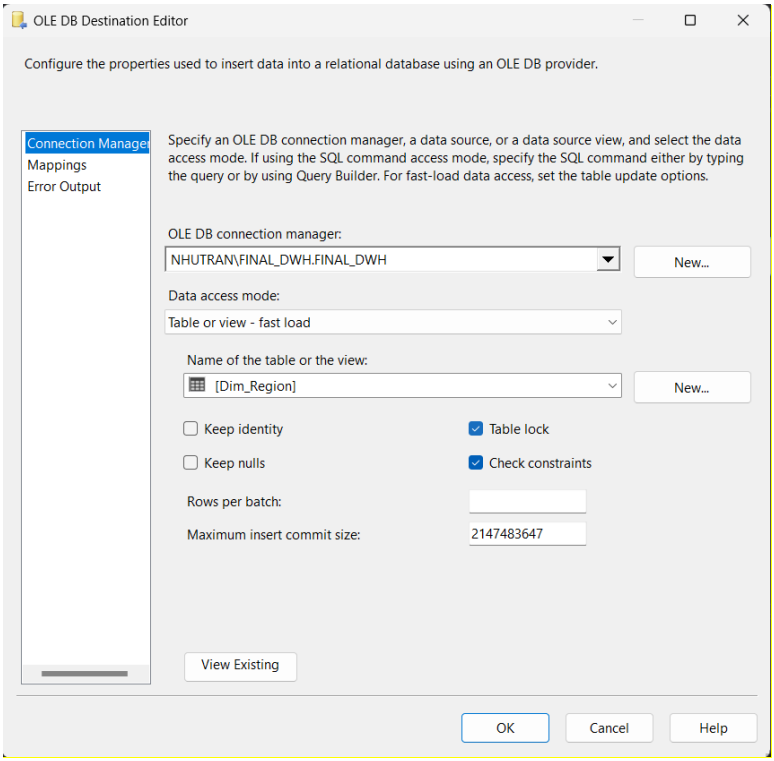



*Figure 30. Create Dimension table*

Final step is Load to OLE DB

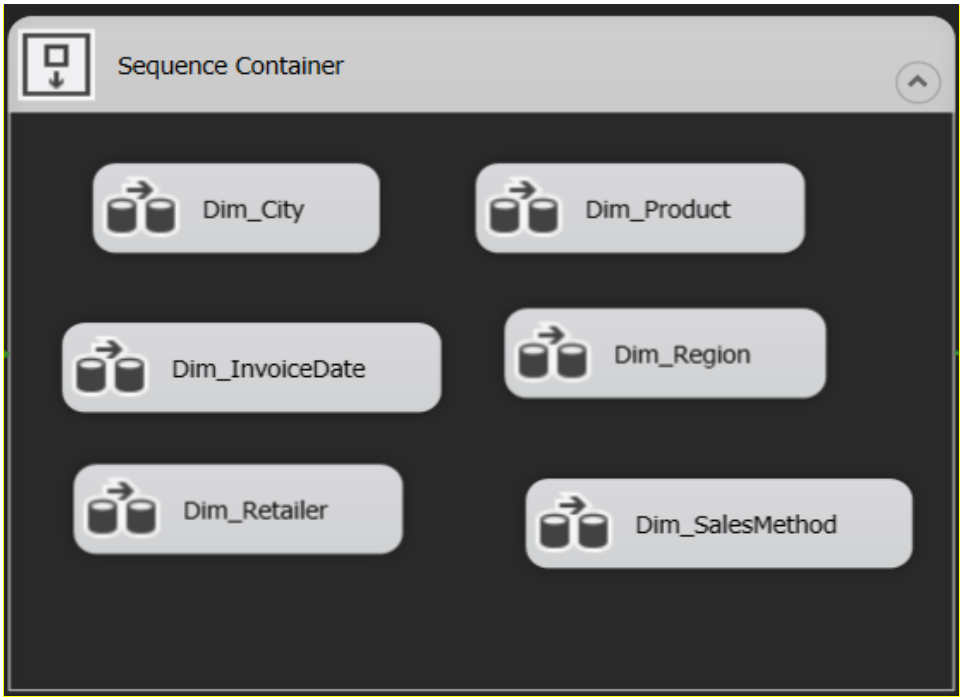



*Figure 31. Dimension tables*

The ETL process is executed with various types of table data. The data source is loaded and aggregated, with group by and sum calculations applied. As the data is often

diverse in type, it needs to be appropriately formatted before loading, necessitating a data column formatting process. Subsequently, keys are generated, and the data loading process into SQL is carried out.

- **ETL Fact Tables**

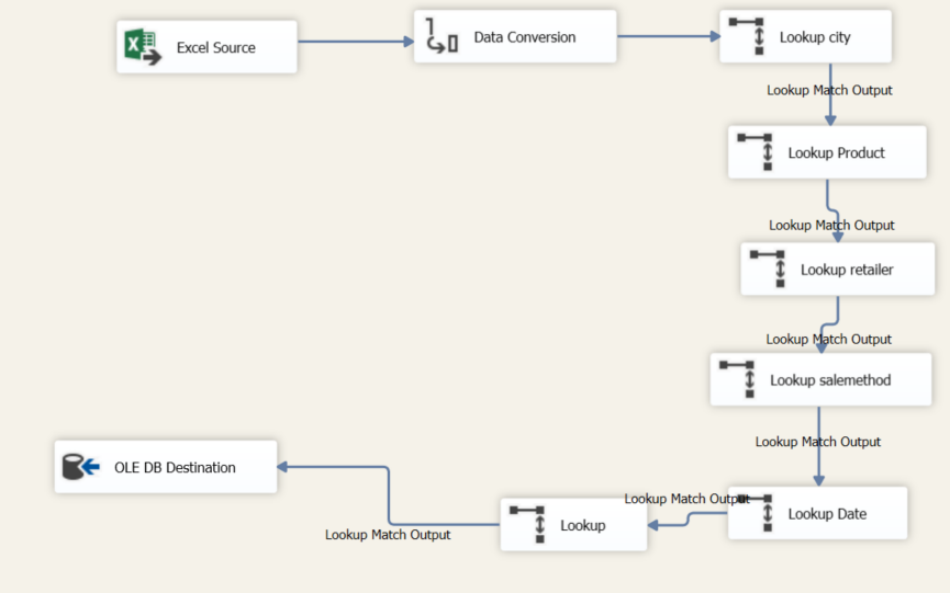



*Figure 32. Create Fact Table*

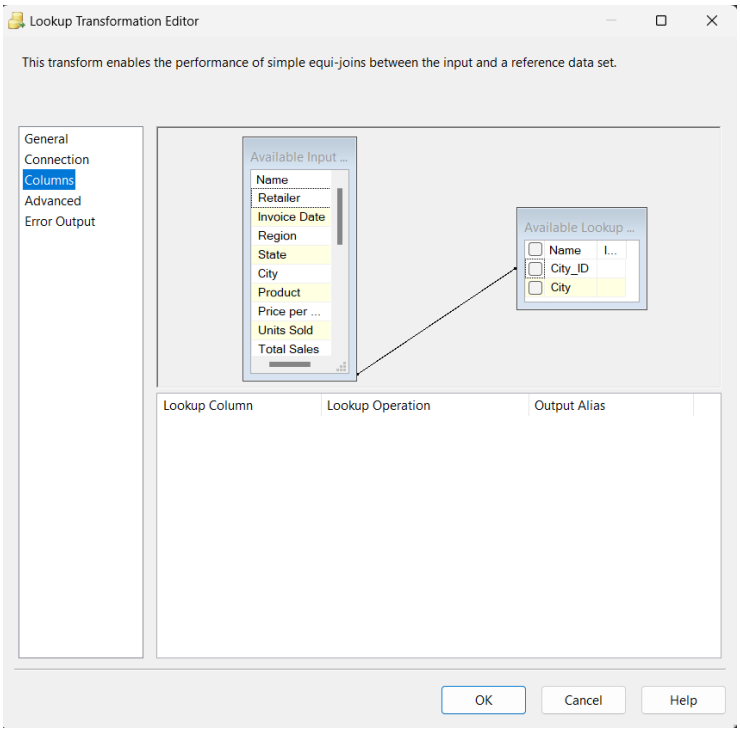



*Figure 33. Look up from Fact_sales to Dimension tables*

The Fact table is initialized from the initial Sales data file and is connected to look up values in Dimension tables to retrieve key values into the Fact table. Through the Star Schema in the Figure, ID values are looked up in the Fact table, followed by the process of establishing relationships.

Final step is Load to OLE DB

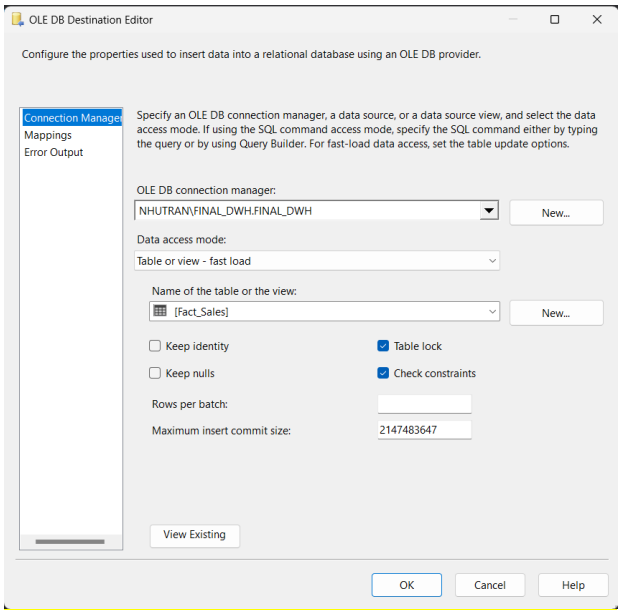



*Figure 34. Dimension tables editor*

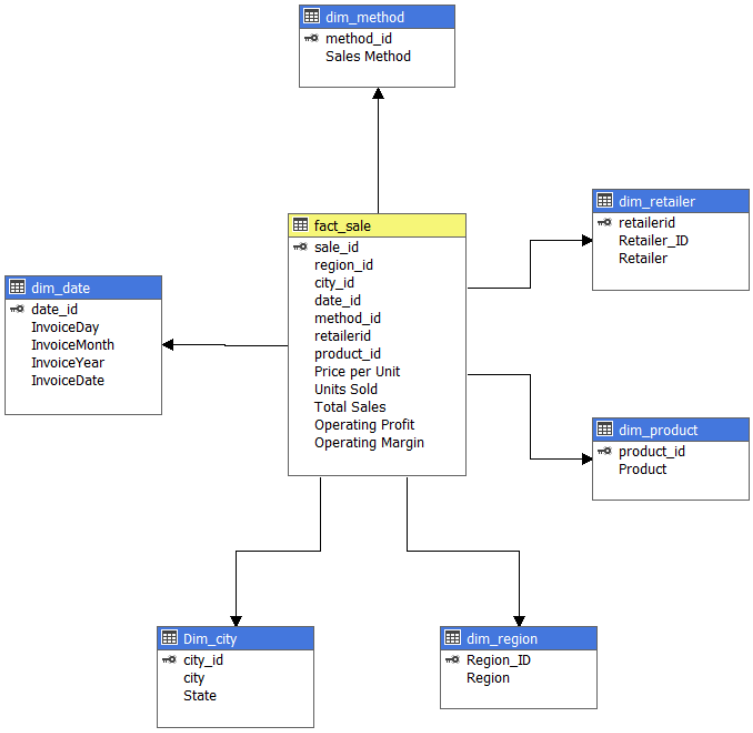



*Figure 35. Star Schema*

To facilitate the analysis process, a Sales Data Mart is designed for analysis purposes. These queries are used to filter, group, and summarize the model data. Then, a well-designed model provides tables for filtering, grouping, and summarizing model data. This design aligns well with the principles of a star schema. [21]:

- Dimension Tables supporting filtering and grouping
- DimRetailer: Describes information about retailers collaborating with Adidas.
- DimInvoiceDate: Describes information about invoice dates.
- DimProduct: Describes Adidas' product lines.
- DimSalesMethod: Describes information about sales methods.
- DimCity: Describes the names of cities in the United States where retailers are located.
- DimRegion: Describes geographical regions in the United States.
- FactSales: Describes information about sales orders.

4.3.2. Implement OLAP

After successfully building the Data Warehouse in the above section, continue to build an OLAP analytical database to serve as a premise for the query process later.

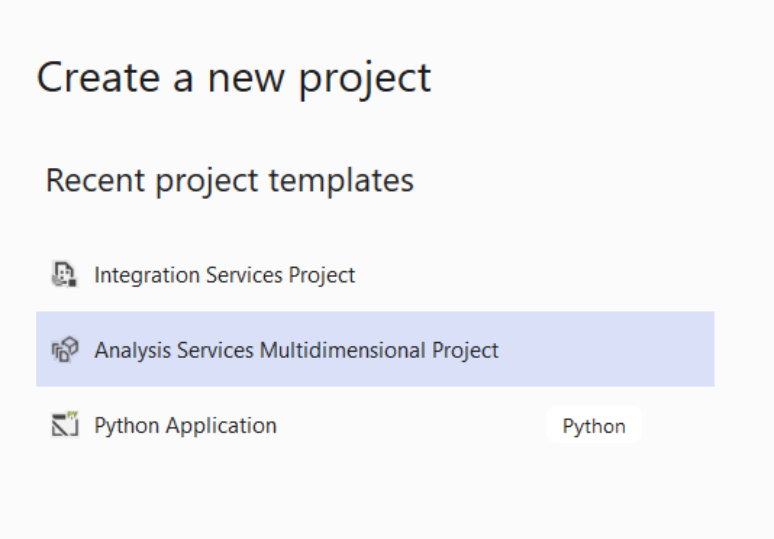



*Figure 36. Create a new project*

The first step in the process is to start Visual Studio, select "Create a new project" and select the application package that supports the cube creation process, "Analysis Services Multidimensional Project" to create a new project.

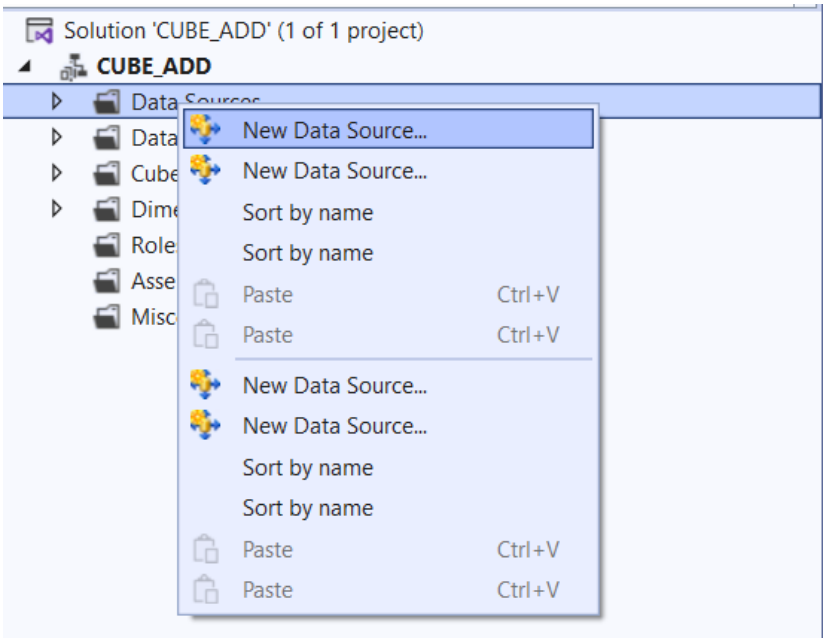



*Figure 37. Create a data connector*

In the Solution Explorer window of Project ***CUBE_ADD***, right-click the Data Source to create a connection to the data for analysis.

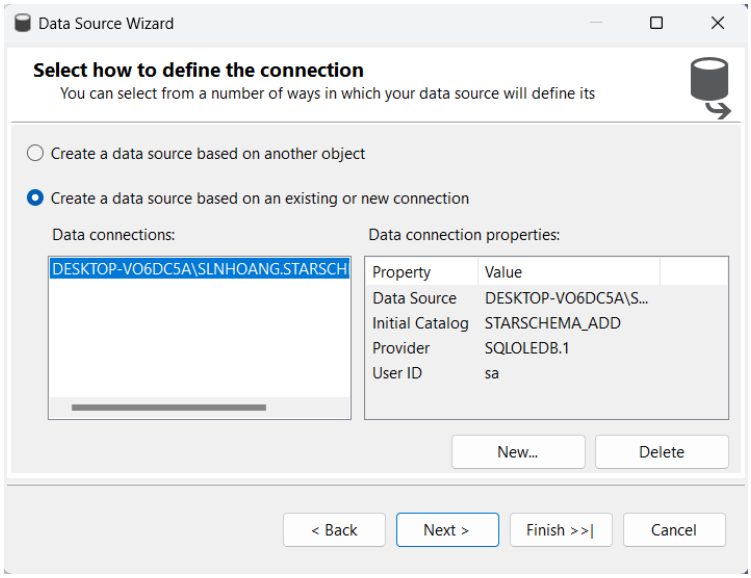



*Figure 38. Data connection*

In this section, implement the data warehouse that wants to create a cube. Then name the Data Source and click Finish to complete the connection to the database.

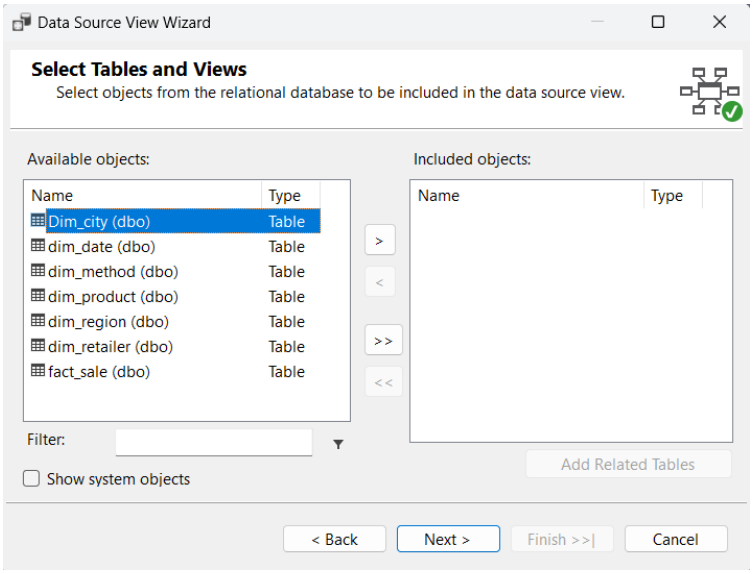



*Figure 39. Identify input data sources*

Create a Data Source View to get the necessary data tables needed for analysis. Right-click Data Source View in the *Solution Explorer* window and select *New Data Source View*. Next, select data tables that suit your analysis requirements in the *Select Tables and Views* section.

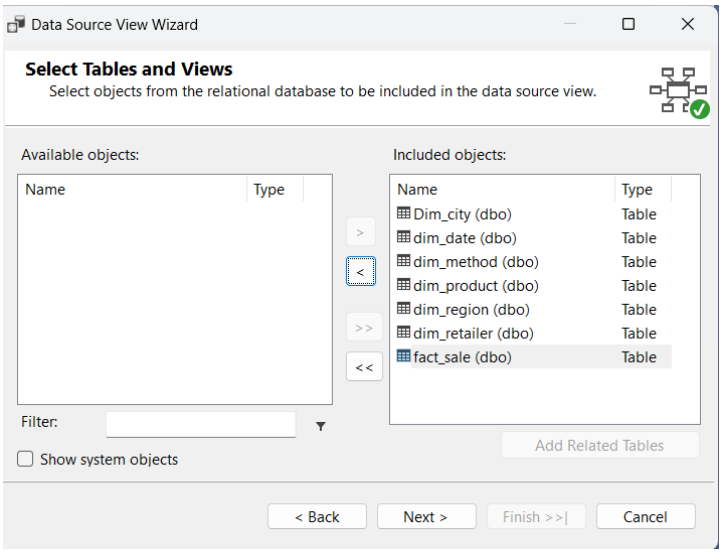



*Figure 40. Choose the table to parsing*

After completing the above step, the fact and dim tables will be displayed visually using the schema.

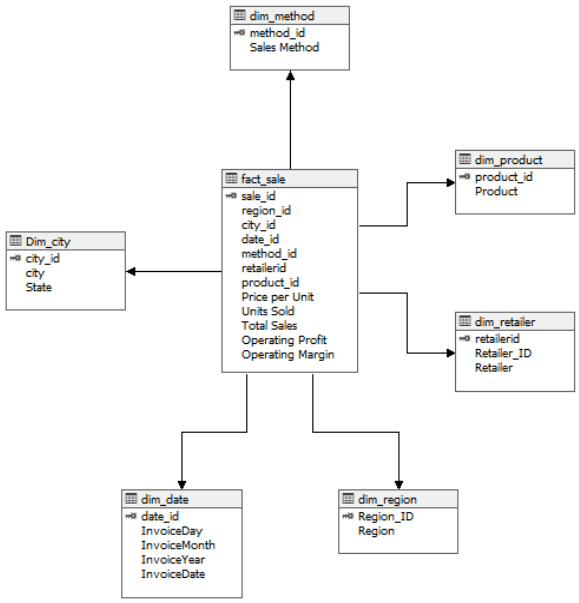



*Figure 41. Star schema diagram.*

After creating the Data Source and Data Source View, we create cube data for analysis by right-clicking on the Cube in Solution Explorer and selecting New Cube.

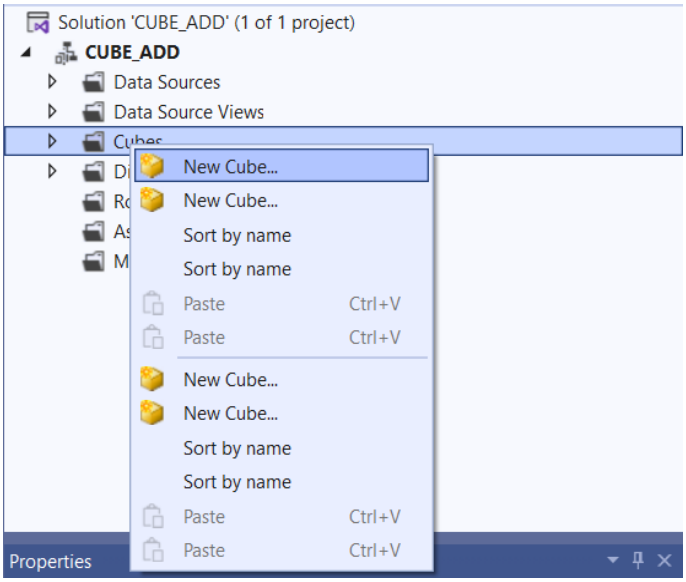



*Figure 42. Create new cube*

During the cube creation process, we will select the Fact table in the data warehouse and select appropriate measures to support the query process afterward.

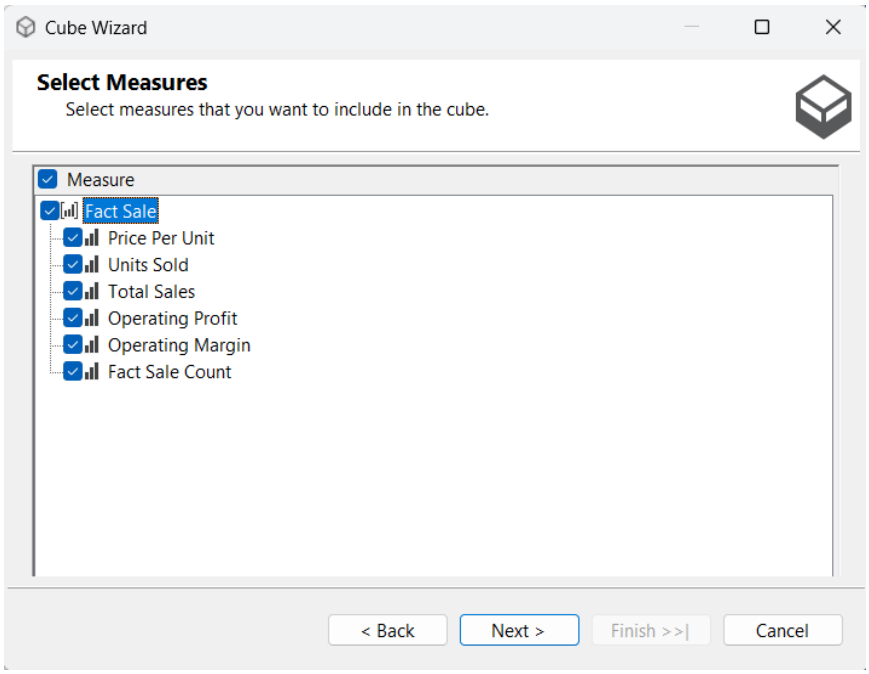



*Figure 43. Select Measures*

Here, select the Measures with the following uses: First, *Price Per Unit* helps measure the average price of each product and is often used to track price trends and evaluate sales performance. Next, *Units Sold* helps track the number of products sold. This is an important indicator to evaluate sales performance and predict future demand. Third is *Total Sales*, this measure helps calculate the total value of all products sold and is also a main measure to measure revenue. Besides, *Operating Profit* helps measure profits from business activities after deducting all related costs and fees. Next is *Operating Margin*, which helps calculate a percentage, measures profit from operations compared to sales, and is used to evaluate the financial performance of a business. Finally and also the most widely used measure, *Fact Sale Count*. It helps count the number of sales transactions and is used to measure specific values of sales data.

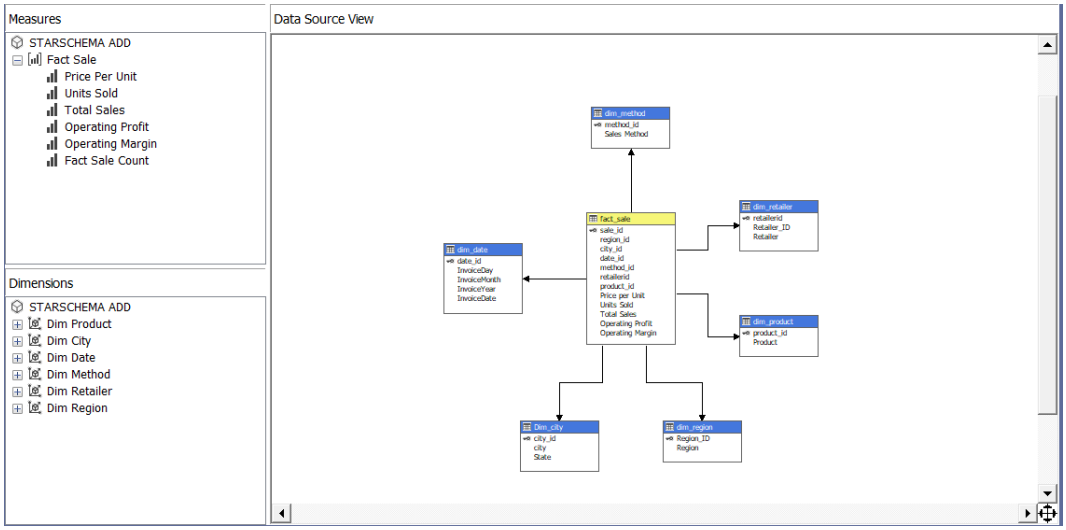



*Figure 44. Complete the cube*

After performing all the above steps, the cube will be complete and ready to serve the next query process.

### 4.3.3. MDX query

MDX is a powerful language for multidimensional data querying, helping us discover information from complex data systems. For the Adidas data set, we can use

MDX to query important parameters to serve the analysis process, visualize data, and meet the set KPIs.

- ***KPIs By Sales Volume***

First, KPIs By Sales Volume, this is an important indicator to measure the financial strength of a company during the research period. Here execute the queries:

```
SELECT NON EMPTY { [Measures].[Total Sales] } ON COLUMNS
FROM [STARSCHEMA ADD]
```

100 %

Messages | Results

| Total Sales |
|---|
| 899902125 |

*Figure 45. Total revenue in the period 2020 - 2021*

```
SELECT
  NON EMPTY { [Measures].[Total Sales] } ON COLUMNS,
  NON EMPTY { [Dim Date].[Invoice Year].[Invoice Year].ALLMEMBERS } ON ROWS
FROM [STARSCHEMA ADD]
```

100 %

Messages | Results

| | Total Sales |
|---|---|
| 2020 | 182080675 |
| 2021 | 717821450 |

*Figure 46. Total revenue of Adidas by year*

```
SELECT NON EMPTY { [Measures].[Units Sold] } ON COLUMNS,
NON EMPTY { ([Dim Product].[Product].[Product].ALLMEMBERS ) } ON ROWS
FROM [STARSCHEMA ADD]
```

100 %

Messages | Results

| | Units Sold |
|---|---|
| Men's Apparel | 306683 |
| Men's Athletic Footwear | 435526 |
| Men's Street Footwear | 593320 |
| Women's Apparel | 433827 |
| Women's Athletic Footwear | 317236 |
| Women's Street Footwear | 392269 |

*Figure 47. Total products sold in the period 2020 – 2021*

- ***KPIs by profit and performance***

Next, KPIs by profit and performance, are important indicators to measure the profit and performance of a business. Here execute the queries one after another:

```
SELECT NON EMPTY { [Measures].[Operating Profit], [Measures].[Total Sales] } ON COLUMNS,
NON EMPTY { ([Dim Date].[Invoice Year].[Invoice Year].ALLMEMBERS ) } ON ROWS
FROM [STARSCHEMA ADD]
```

100 %

Messages | Results

| | Operating Profit | Total Sales |
|---|---|---|
| 2020 | 63375662.5799999 | 182080675 |
| 2021 | 268759098.87 | 717821450 |

*Figure 48. Total revenue and profit by year*

```
-- Operating margin Region
SELECT NON EMPTY { [Measures].[Operating Margin] } ON COLUMNS,
NON EMPTY { ([Dim Region].[Region].[Region].ALLMEMBERS ) } ON ROWS
FROM [STARSCHEMA ADD]
```

100 %

Messages | Results

| | Operating Margin |
|---|---|
| Midwest | 814.830000000002 |
| Northeast | 975.230000000002 |
| South | 806.800000000002 |
| Southeast | 513.06 |
| West | 971.099999999998 |

*Figure 49. Operating margin by region*

- ***KPIs By Sales Method:***

Third is KPIs By Sales Method, analysis by sales method is the key to better understanding customer needs for each market. Here execute the queries one after another:

```
SELECT NON EMPTY { [Measures].[Total Sales] } ON COLUMNS,
NON EMPTY { ([Dim Method].[Sales Method].[Sales Method].ALLMEMBERS ) } ON ROWS
FROM [STARSCHEMA ADD]
```

100 %

Messages | Results

| | Total Sales |
|---|---|
| In-store | 356643750 |
| Online | 247672882 |
| Outlet | 295585493 |

*Figure 50. Query revenue by method*

```
SELECT NON EMPTY { [Measures].[Units Sold] } ON COLUMNS,
NON EMPTY { ([Dim Method].[Sales Method].[Sales Method].ALLMEMBERS ) } ON ROWS
FROM [STARSCHEMA ADD]
```

100 %

Messages | Results

| | Units Sold |
|---|---|
| In-store | 689990 |
| Online | 939093 |
| Outlet | 849778 |

*Figure 51. Query the total quantity of goods sold by method*

```
SELECT NON EMPTY { [Measures].[Total Sales] } ON COLUMNS,
NON EMPTY { ([Dim Retailer].[Retailer].[Retailer].ALLMEMBERS * [Dim Method].[Sales Method].[Sales Method].ALLMEMBERS ) }
FROM [STARSCHEMA ADD]
```

100 %

Messages | Results

| | | Total Sales |
|---|---|---|
| Amazon | In-store | 22366250 |
| Amazon | Online | 28909731 |
| Amazon | Outlet | 26422931 |
| Foot Locker | In-store | 76525000 |
| Foot Locker | Online | 72943290 |
| Foot Locker | Outlet | 70626430 |
| Kohl's | In-store | 29566250 |
| Kohl's | Online | 30992229 |
| Kohl's | Outlet | 41556274 |
| Sports Direct | In-store | 55048500 |
| Sports Direct | Online | 59225997 |
| Sports Direct | Outlet | 68196500 |
| Walmart | In-store | 16411250 |
| Walmart | Online | 15069494 |
| Walmart | Outlet | 43077666 |
| West Gear | In-store | 156726500 |
| West Gear | Online | 40532141 |
| West Gear | Outlet | 45705692 |

*Figure 52. Sales of retailers by method*

- ***KPIs By Region and Time:***

Finally, KPIs By Region and Time, Sales by Region, City, and State also provide an overall and detailed view of the local market. Here execute the queries:

```
SELECT NON EMPTY { [Measures].[Total Sales] } ON COLUMNS,
NON EMPTY { ([Dim Date].[Invoice Year].[Invoice Year].ALL
FROM [STARSCHEMA ADD]
```

100 %

Messages | Results

| | | Total Sales |
|---|---|---|
| 2020 | 01 | 16253746 |
| 2020 | 02 | 14997988 |
| 2020 | 03 | 17660577 |
| 2020 | 04 | 24607006 |
| 2020 | 05 | 16918014 |
| 2020 | 06 | 8829819 |
| 2020 | 07 | 17146013 |
| 2020 | 08 | 19877980 |
| 2020 | 09 | 18304436 |
| 2020 | 10 | 10836269 |
| 2020 | 11 | 8622300 |
| 2020 | 12 | 8026527 |
| 2021 | 01 | 55225396 |
| 2021 | 02 | 46102165 |
| 2021 | 03 | 39148532 |
| 2021 | 04 | 47732964 |
| 2021 | 05 | 63589681 |
| 2021 | 06 | 65917553 |
| 2021 | 07 | 78334681 |
| 2021 | 08 | 72288221 |
| 2021 | 09 | 59357023 |
| 2021 | 10 | 53074764 |
| 2021 | 11 | 59235040 |
| 2021 | 12 | 77815430 |

*Figure 53. Revenue fluctuations during the months of the year*

```
SELECT
  NON EMPTY { [Measures].[Total Sales] } ON COLUMNS,
  NON EMPTY {
    TOPCOUNT(
      [Dim City].[City].[City].MEMBERS,
      10,
      [Measures].[Total Sales]
    )
  } ON ROWS
FROM [STARSCHEMA ADD]
```

100 %

Messages | Results

| | Total Sales |
|---|---|
| Charleston | 39974797 |
| New York | 39801235 |
| San Francisco | 34539220 |
| Miami | 31600863 |
| Portland | 30545652 |
| Orlando | 27682851 |
| Seattle | 26330718 |
| Los Angeles | 25634913 |
| Houston | 25456882 |
| Albany | 24427804 |

*Figure 54. Top 10 cities with the largest revenue*

```
-- top 10 bang có doanh thu cao nhất
SELECT
  NON EMPTY { [Measures].[Total Sales] } ON COLUMNS,
  NON EMPTY {
    TOPCOUNT(
      [Dim city].[State].[State].MEMBERS,
      10,
      [Measures].[Total Sales]
    )
  } ON ROWS
FROM [STARSCHEMA ADD]
```

100 %

Messages | Results

| | Total Sales |
|---|---|
| New York | 64229039 |
| California | 60174133 |
| Florida | 59283714 |
| Texas | 46359746 |
| South Carolina | 39974797 |
| Maine | 30545652 |
| Washington | 26330718 |
| North Carolina | 23956531 |
| Louisiana | 23750781 |
| Hawaii | 22282457 |

*Figure 55. Top 10 states with the highest revenue*

```
SELECT NON EMPTY { [Measures].[Total Sales], [Measures].[Units Sold] } ON COLUMNS,
NON EMPTY { ([Dim Region].[Region].[Region].ALLMEMBERS ) } ON ROWS
FROM [STARSCHEMA ADD]
```

100 %

Messages | Results

| | Total Sales | Units Sold |
|---|---|---|
| Midwest | 135800459 | 391337 |
| Northeast | 186324067 | 501279 |
| South | 144663181 | 492260 |
| Southeast | 163171236 | 407000 |
| West | 269943182 | 686985 |

*Figure 56. Revenue and number of sales by region*

Above are queries that help solve the proposed KPIs. This is also the prerequisite step to visualize data into charts in the next step.

### 4.3.4. DAX

The research continues to use the DAX calculation language, determine measurement values and perform analysis to make comments on Adidas' business situation in 2020 and 2021.

First, create a new report to build a dashboard from the measures set by the group.

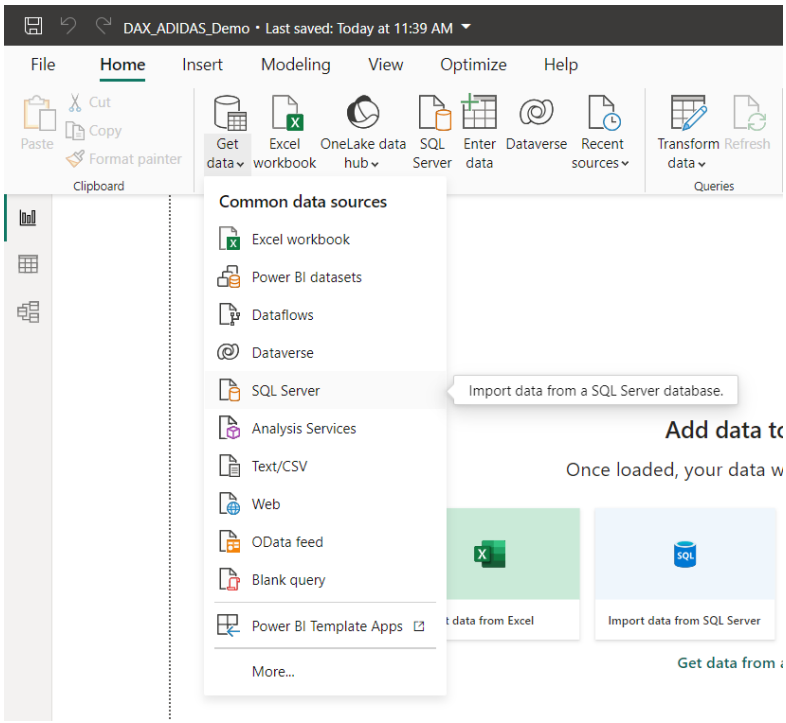



*Figure 57. Get data from a data warehouse in SQL Server.*

Then, in the box that appears, enter the server name and database name. We will choose Data Connectivity mode as Import so that the system automatically imports all data in the database, or choose DirectQuery to perform import by query. Here the research team chooses the method of automatically importing all tables in the database, then selects OK.

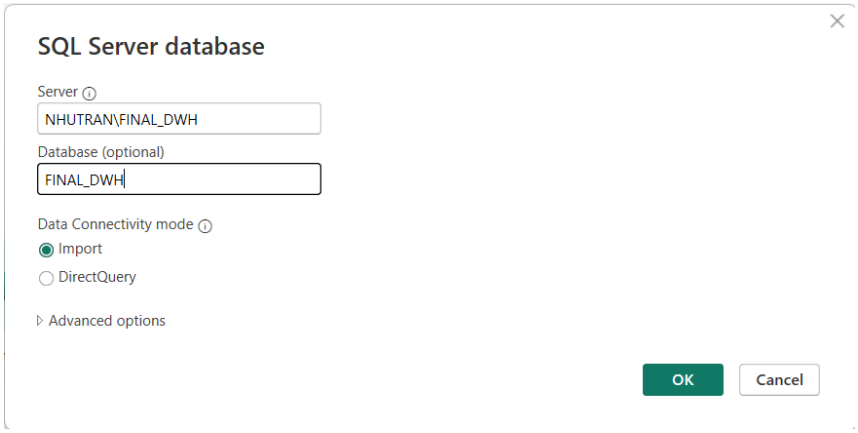



*Figure 58. Fill in server name and database*

Then select all dim tables and created tables to perform calculations and visualize measurement indicators, finally click Load.

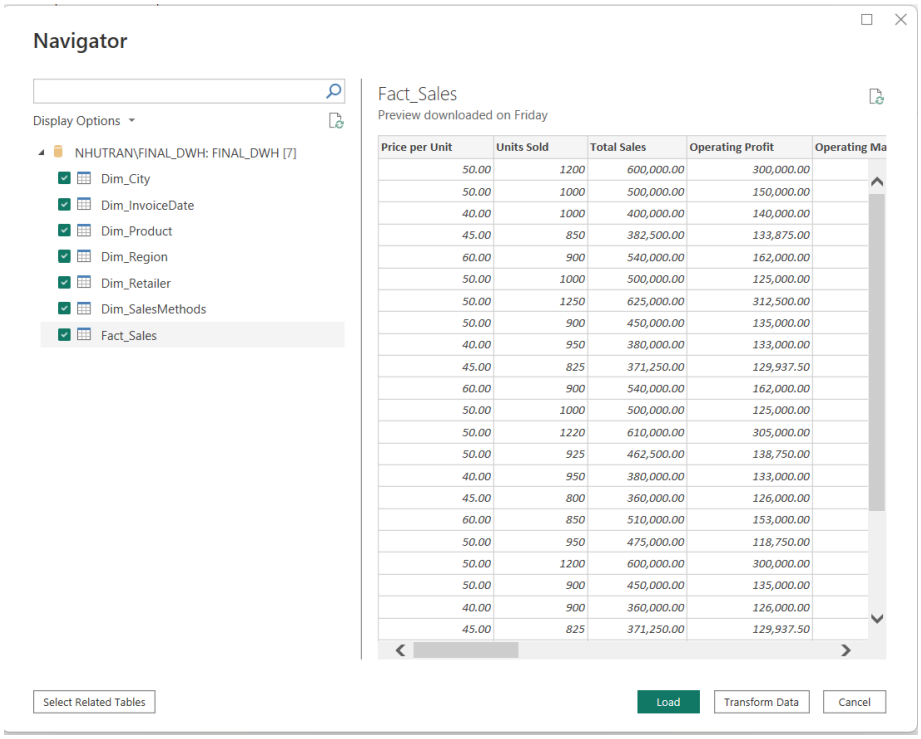



*Figure 59. Select dimension tables and fact table.*

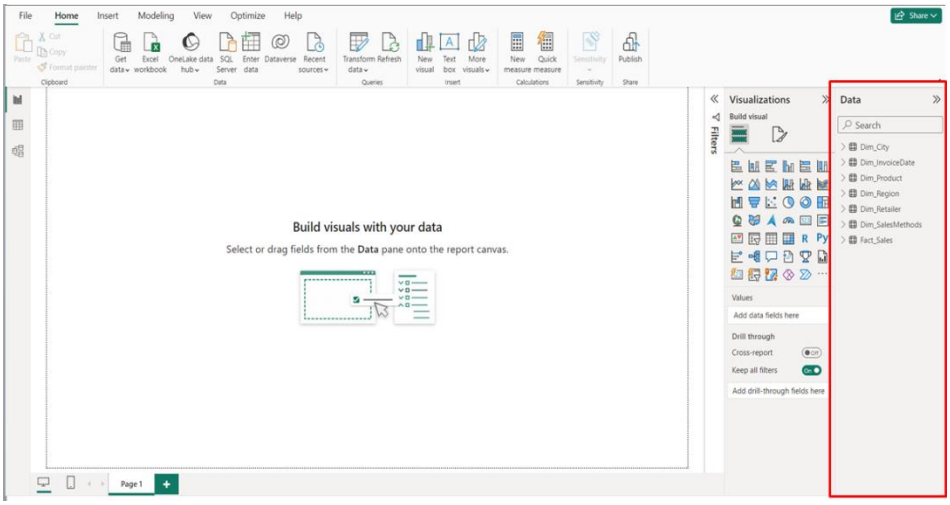



*Figure 60. Data loaded from SQL Server to Power BI.*

After loading the input data, we will determine the measurement indicators based on the group problem we have set.

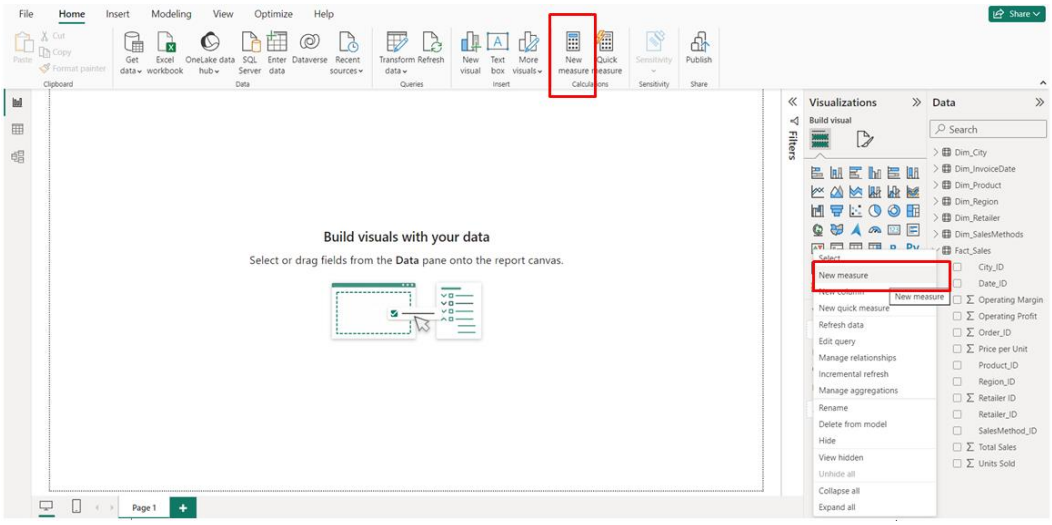



*Figure 61. Identify measures.*

You can add a new measure by right-clicking on the table to add a measure or selecting the create New measure command as shown in the picture. And then enter the calculation formula for that measure.

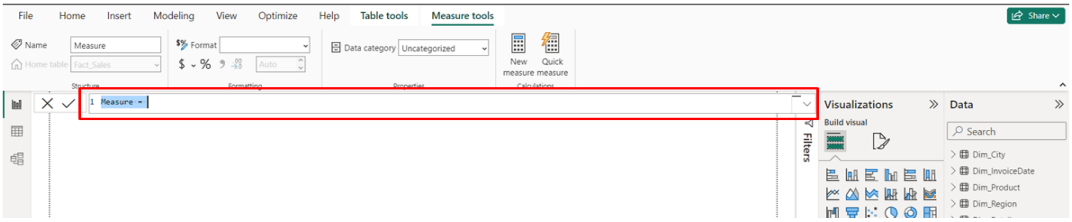



*Figure 62. Create measure's formula.*

Here, the research team analyzed a number of measurement indicators as follows:

- ***Total sales***: this is the total revenue. With this index, the team aims to perform analysis on different aspects such as by region, by time,...
- ***Operating profit***: this is the total profit before interest and taxes or income before interest and taxes is an indicator used to evaluate the company's ability to earn profits, equal to income minus expenses, but not yet minus interest and income tax. The study will use operators for this measure such as sum, average, min, max,... to analyze the business situation of the enterprise.

- ***Unit sold***: this is the number of products sold. With this measure, research can analyze the level of consumption by product, by region, by year,...; top 5, top 10 best sellers;... using some related operators such as sum, count,...
- ***Operating margin***: This is the profit ratio on revenue, a financial ratio used to monitor the profitability of a joint stock company. It reflects the relationship between net profit available to shareholders and company revenue. With this measure, researchers can perform related operators such as average, min, max,...
- ***Sale ratio by region***: this is the ratio of revenue of one region to the total revenue of all regions.

```
Sales Ratio by region =
DIVIDE (
    SUMX (
        VALUES ( Dim_Region[Region_ID] ),
        CALCULATE ( SUM ( Fact_Sales[Total Sales]) )
    ),
    CALCULATE ( SUM ( Fact_Sales[Total Sales] ), ALL ( Dim_Region ) )
)
```

*Figure 63. Formula of Sale ratio by region*

- ***Profit ratio by region****: is the profit ratio of one region to the total profit of all regions.*

```
Profit Ratio by region =
DIVIDE (
    SUMX (
        VALUES ( Dim_Region[Region_ID] ),
        CALCULATE ( SUM ( Fact_Sales[Operating Profit] ) )
    ),
    CALCULATE ( SUM ( Fact_Sales[Operating Profit] ), ALL ( Dim_Region ) )
)
```

*Figure 64. Formula of Profit ratio by region*

- ***Return on Sales***: Profit ratio on revenue is a financial ratio used to monitor the profitability of a joint stock company. It reflects the relationship between net profit available to shareholders and company revenue.

```
ROS = DIVIDE(SUM('Fact_Sales'[Operating Profit]), SUM('Fact_Sales'[Total Sales]), 0)
```

*Figure 65. Formula of ROS.*

4.3.5. Slowly changing dimension (SCD Type 2)

In the process of managing and storing data, tracking changes in information over time is an important challenge for businesses. For data warehousing systems, understanding fluctuations in data is important to ensure the accuracy and reliability of

information. One of the popular techniques used to handle this problem is Slowly Changing Dimension (SCD), of which SCD type 2 is one of the most important methods.

SCD type 2 is an effective method for tracking changes in data over time while also retaining a history of previous values. This becomes important when we need to track fluctuations in the attributes of an entity or object in data, such as product information, and price fluctuations,...

In this essay, we will focus on modeling and implementing the SCD type 2 process for the Adidas sales data set. In this way, we can not only maintain data consistency but also track the development and changes of information over time, facilitating analysis and decision-making based on historical data.

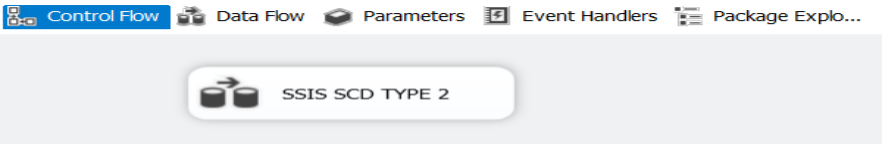



*Figure 66. Create new data flow task*

The first step of the process is to create a new project on SSIS. After that, we will create a new data flow task. Then we will design a new data flow, we will drag and drop the OLE DB source to add an input data source. Here, create *OLE DB Connection Manager* and use the *dbo.sale* table (Adidas's original sales table) as the source table.

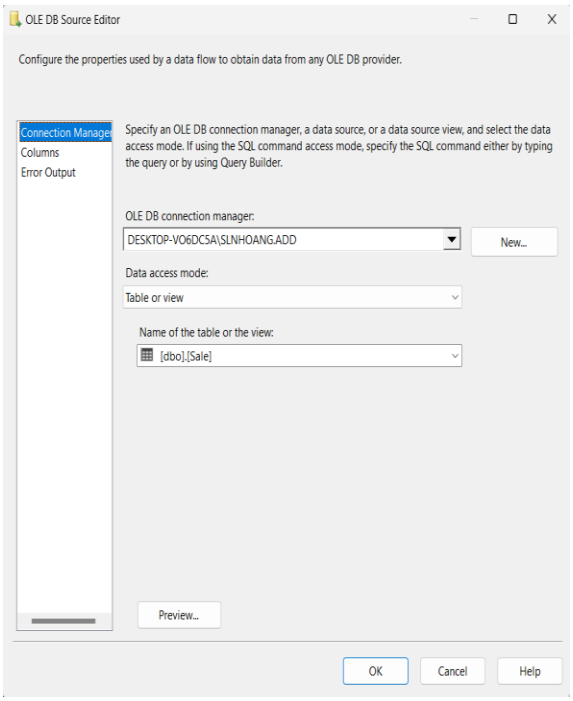



*Figure 67. Create source table*

The next step in the process is to initialize *SSIS Slowly Changing Dimension* to initiate SCD type 2. Here, we proceed to *Select a Dimension Table and Keys* that suit our requirements. We choose the table to perform SCD type 2 as *[dbo].[insert_Sale]* and select at least one business key as *Sale_ID*.

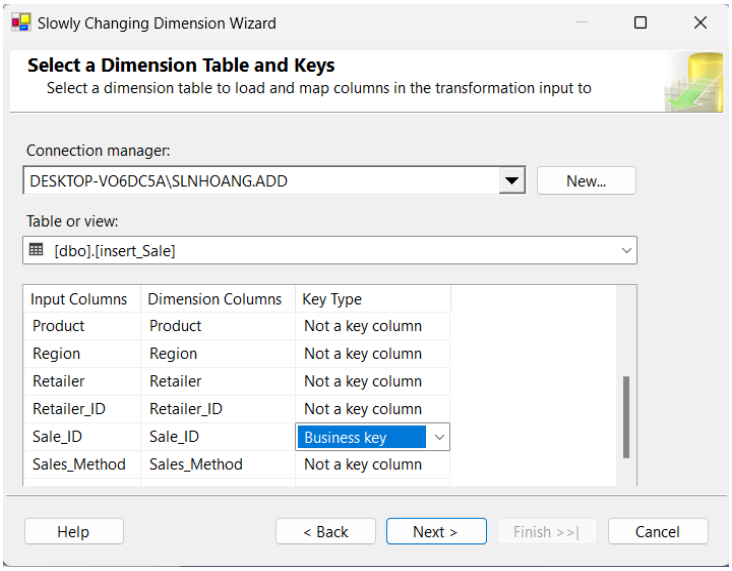



*Figure 68. Select a Dimension Table and Business Keys*

The most important step of the process is that we will Select dimension columns and change the type to suit the analysis requirements. There are 3 types of attributes to choose from:

- First, Fixed Attributes are attributes that do not change over time. In this case, the value [product] will be chosen as the selected fixed attribute because they will not change the product name no matter how many times it is sold.
- The second is Changing Attributes. These are attributes that may change over time. Columns like [Region], [Retailer], [Sale_Method] are chosen because they can change after each order.
- Finally, Historical Attributes: These are the attributes whose change history you want to track. Columns like [Units_Sold], [Total_Sales], and [Operating_Profit] can be considered historical attributes because they provide information about sales history and help analysts track fluctuations in revenue, price increases,...

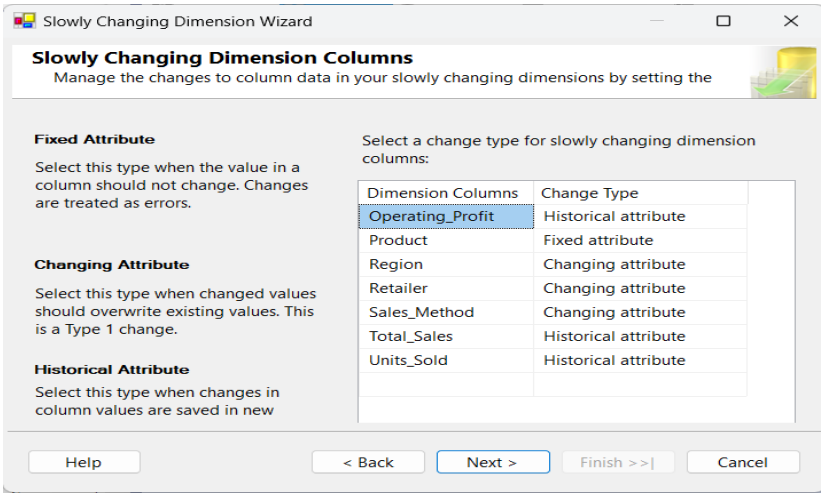



*Figure 69. Select dimension columns and change type*

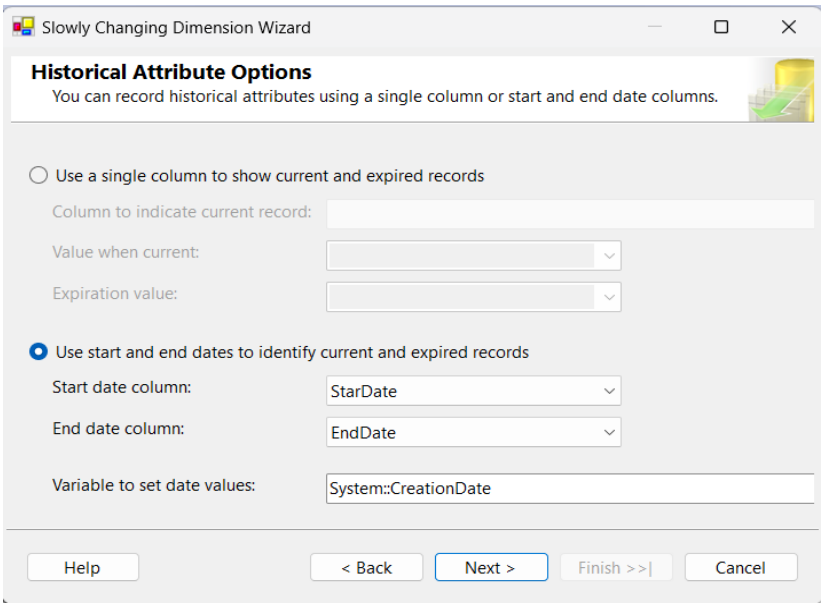



*Figure 70. Select historical attribute options*

SCD type 2 needs to use star_date and end_date to confirm whether the current record has expired and when the update time is. For this table star date is the date the new record was inserted. Then we click and finish, the data flow will automatically change. And finally, we choose to run this thread to complete the SCD type 2 process.

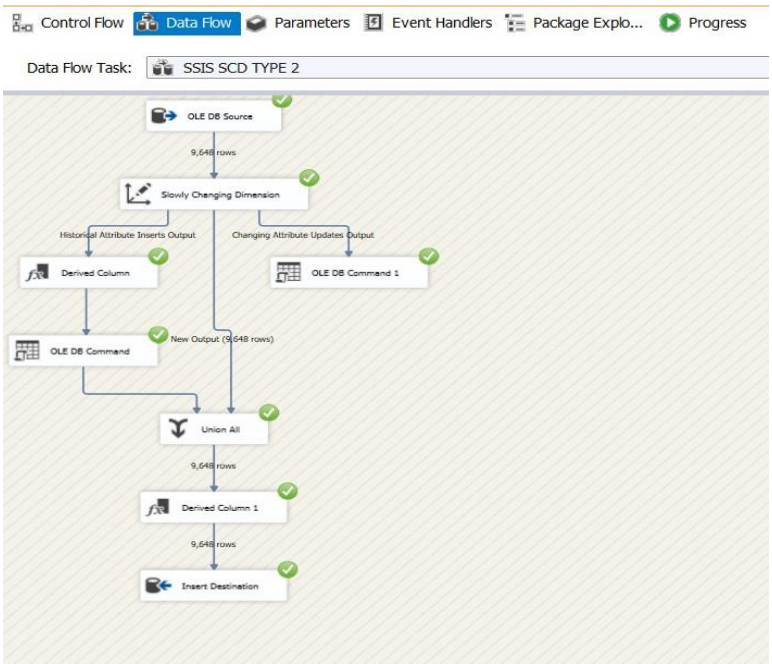



*Figure 71. Successfully initialized data flow for SCD type 2*

After successfully initializing the SCD type 2 data stream, we conducted testing on the Adidas sales data set in 3 typical cases:

- In the first case, with the Region column as the Changing Attributes value, we convert the region from Northeast to Souther in the sale_ID = 1 column. After the change is complete, we will see the results, in the insert_Sale record table, we will insert an additional row. where the Region value has been changed and the old row of data remains unchanged.

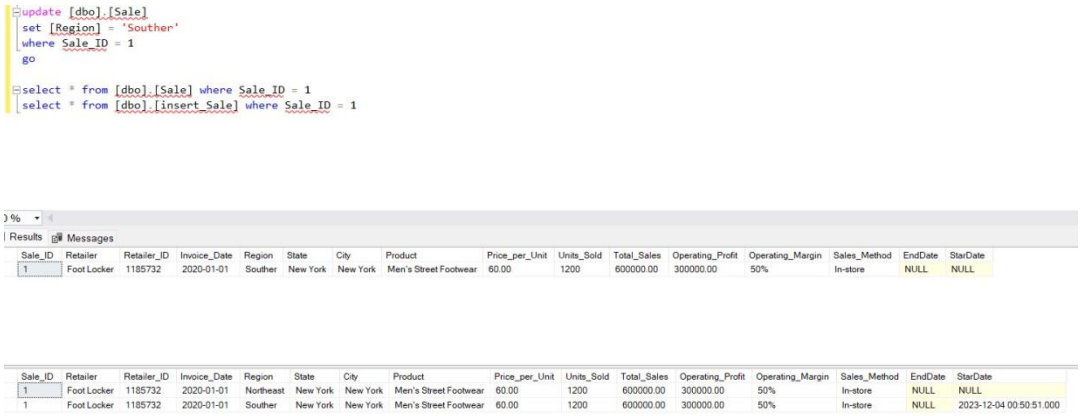



*Figure 72. Change the value for the Region column*

- In the second case, we change the value of the Unit_Sold - Historical Attributes column. We change the data of the Unit_Sold column from 1400 to 1000, and the result will be that the data is inserted with a new row of changed data. changed, at the same time the old data was also changed to 1000.

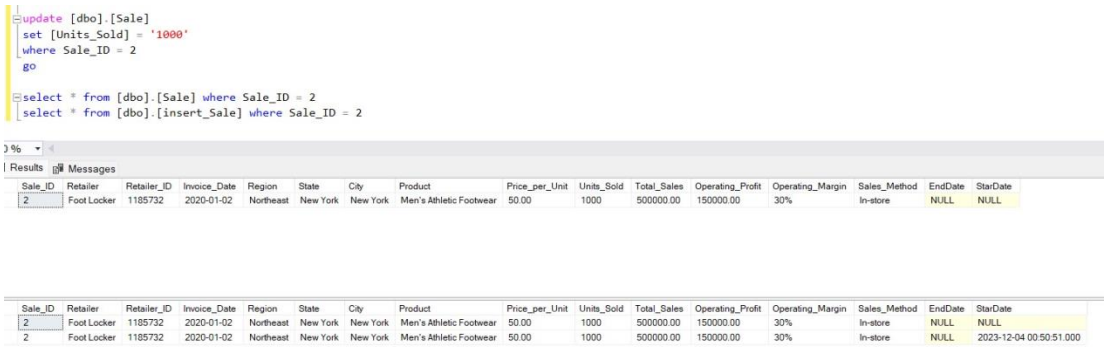



*Figure 73. Change the Unit_Sold value*

- In the third case, when we change the value of Product - Fixed Attributes, the data is similar to adding a new value row in the Insert_Sale table, but the values in both the old row and the newly created row do not change.

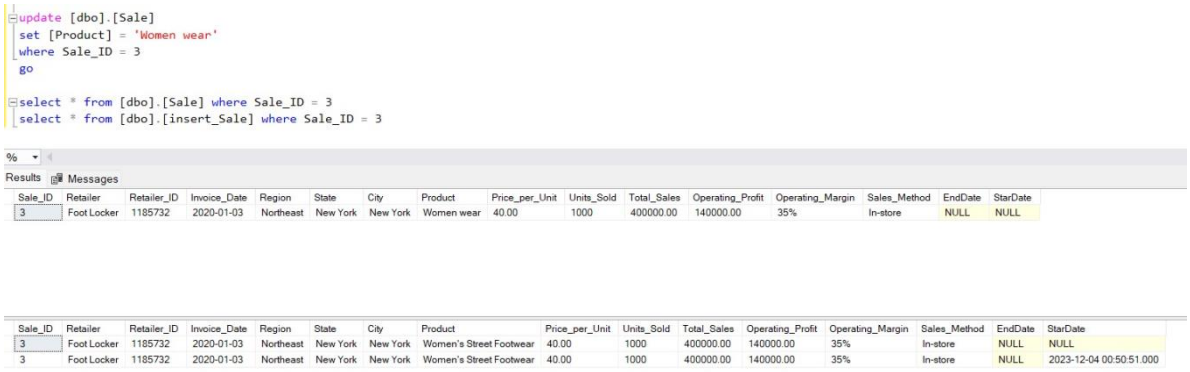



*Figure 74. Change value Product*

# CHAPTER 5: EXPERIMENTAL RESULTS

In this chapter, the study will present all query results and acceptance results from the analysis process based on data from the data warehouse with measurement indicators selected from business requirements.

## *5.1. Overview of Adidas Business Situation in the United States*

### 5.1.1. Revenue Over Time

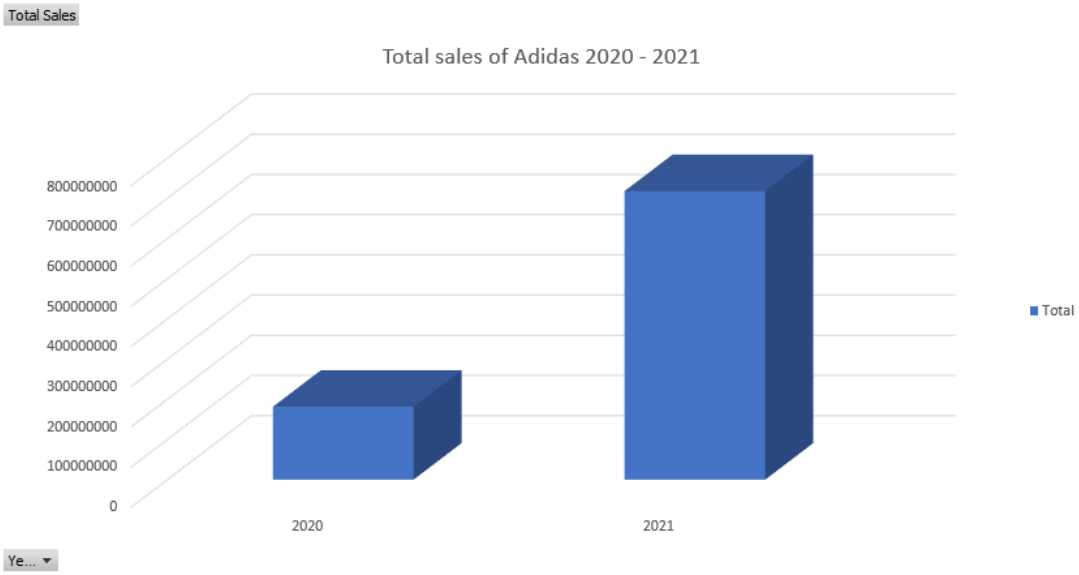



*Figure 75. Total revenue of Adidas by year*

Comparing the total revenue in 2020 and 2021 reveals a significant difference. For a detailed perspective, the team conducted an evaluation through the following analyses.

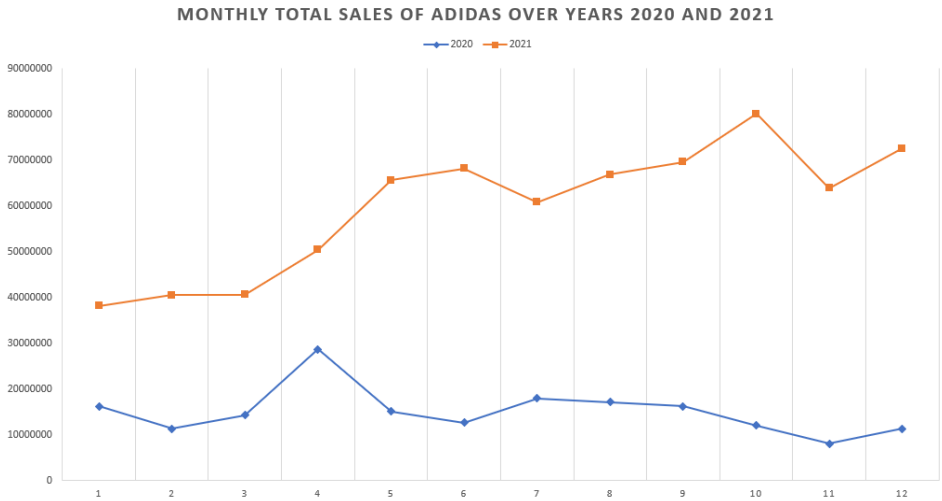



*Figure 76. Revenue variations by month in 2021 and 2022*

The chart results show that 2020 had a peak, mainly in April, and overall throughout the year, there was little fluctuation in revenue compared to 2021. In 2021, the improvement is evident in the second half of the year after a positive shift in the pandemic situation in the United States. Given the period of the Covid-19 pandemic, this abnormality in revenue is understandable.

### 5.1.2. Revenue by method

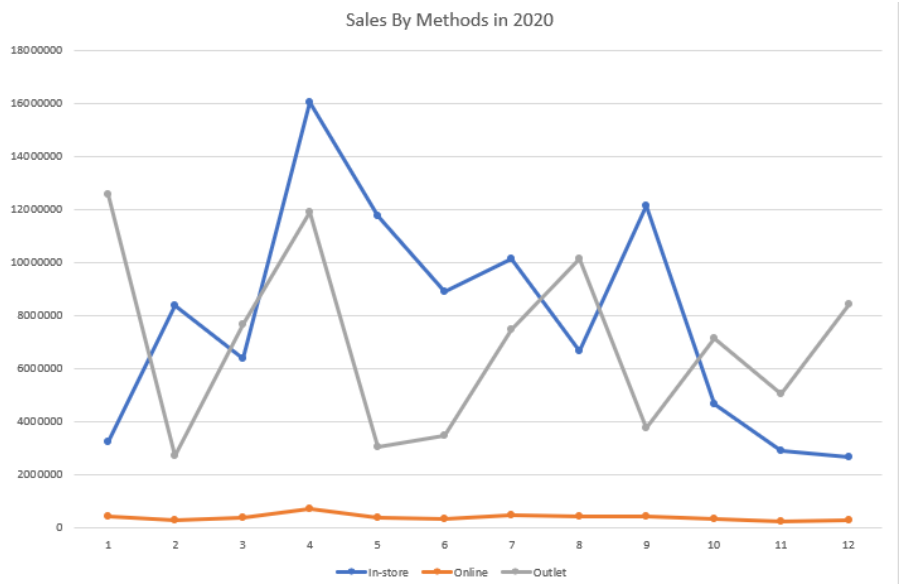



*Figure 77. Revenue by Method in 2020*

Given the pandemic situation, there has been evident disruption and significant constraints in the supply chain, with the Online method showing a certain disadvantage in this regard. This prompted Adidas to strongly emphasize the Outlet and In-store methods in 2020. In section 5.1.1., the changes and fluctuations in April show a strong growth in the In-store and Outlet forms, without the need for any intermediary delivery methods.

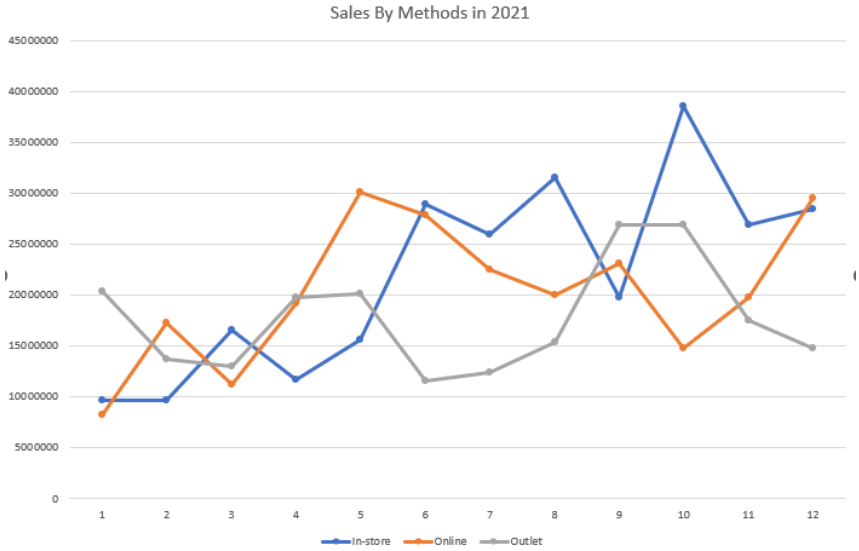



*Figure 78. Revenue by Method in 2021*

It can be seen that after 2020, Adidas and retailers have become more stable in terms of business conditions, and the development of the online method is clearly evident. The revenue difference among the three methods also shows that Adidas has evenly invested in all three forms, yielding quite positive results.

### 5.1.3. Revenue by Geography

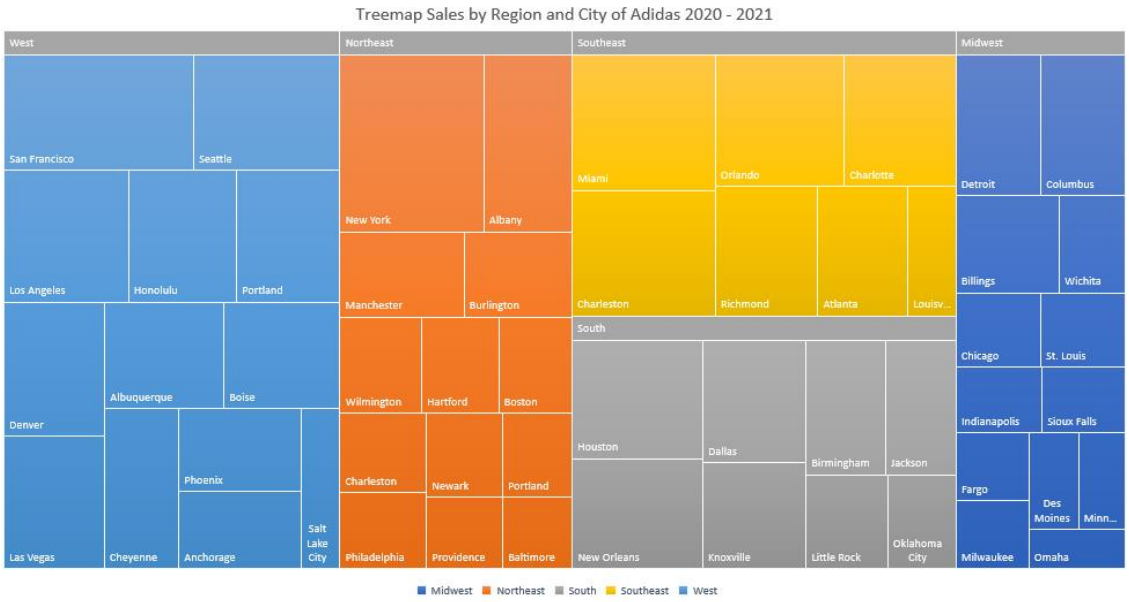



*Figure 79. Treemap Sales by Region and City of Adidas 2020 - 2021*

Orders are concentrated on the West Coast of the United States, reflecting a strong consumer cultural trend in this area. Detailed analysis of geography will be addressed in the following sections.

### 5.1.4. Revenue by Retailer

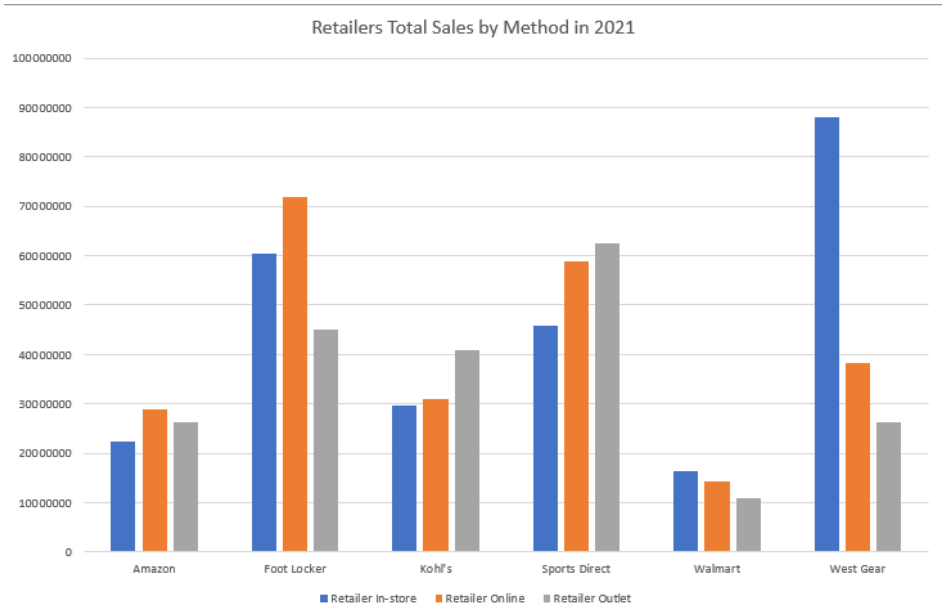



*Figure 80. Retailers Total Sales by Method in 2021*

Assessing based on the year 2021 with stable growth, each retailer employs different business methods. Foot Locker focuses on Online sales, while West Gear performs well in-store. This variation may be attributed to the number of stores and warehouses that retailers have in different geographical locations.

## *5.2. Adidas Sales Performance in the United States during the Period 2020-2021*

- ***Analysis of the quantity of products sold by the business in 2020 and 2021***

Based on the collected business results, a total of 7 product lines recorded revenue, excluding promotional items. The next research will conduct an exploratory data analysis on the quantity of products sold for each product line that Adidas provides in the United States. This helps managers identify popular items, any shortcomings in product lines that may deter users from consuming them. Additionally, businesses can formulate pricing and cost strategies tailored to the output standards for each product type.

With a total of 2.48 million units sold over the two years, 2020 witnessed the sale of 462.35 thousand products, accounting for approximately 18.7%, while 2021 saw the sale of 2.02 million products, constituting around 81.3%. It is evident that there was a significant growth in Adidas' business from 2020 to 2021.

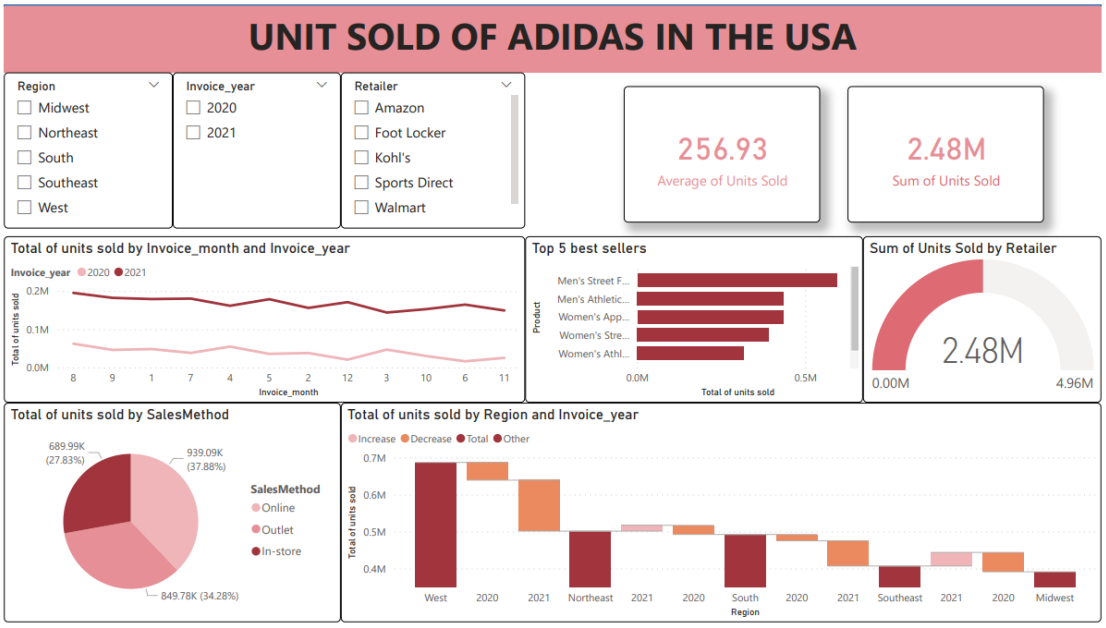



*Figure 81. Unit sold of Adidas in the USA.*

As shown on the dashboard, the quantity of products sold over the months in both years appears relatively consistent, with modest variations between months. There is a positive trend in business towards the end of each year. Based on the product sales data collected from 2020 and 2021, we identified the top 5 best-selling products of Adidas in the overall market. These products include:

- Men's Street Footwear
- Men's Athletic Footwear
- Women's Apparel
- Women's Street Footwear
- Women's Athletic Footwear

Among the top 5 product categories, Men's Street Footwear is favored by male consumers, while Women's Apparel is a popular choice among female consumers. Regarding the shopping methods, the majority of customers prefer online purchases, which is easily understandable given the prevalent use of e-commerce in the current era, offering significant advantages in the shopping experience. The consumption of outlet products is also substantial, accounting for 34.28% of the total products. This can be explained by the nature of outlet items, often discounted or promoted by sellers as they

are typically excess stock. Lastly, purchasing products in-store represents 27.83%, a figure that is not negligible in the overall landscape. However, consumers tend to prioritize faster and more incentive-driven shopping methods.

- ***Analyzing the operational profitability of Adidas is crucial for evaluating business efficiency and profit-generating capabilities.***

The average value of the profit margin on revenue, calculated from the profit margin through each transaction update, is 0.42. This figure will be used to assess Adidas' investment capability compared to industry peers. A positive aspect is that Adidas' operating margin in 2020 and 2021 ranges from 0.1 to 0.8 (i.e., non-negative), indicating that the business operates at a profit. In 2020, Adidas' business operations were not yet stable, as evident from a significant decline in the operating margin during the second quarter. However, by 2021, the path of the operating margin has become more stable, with results across months showing moderate variations within the range of 0.4 to 0.5.

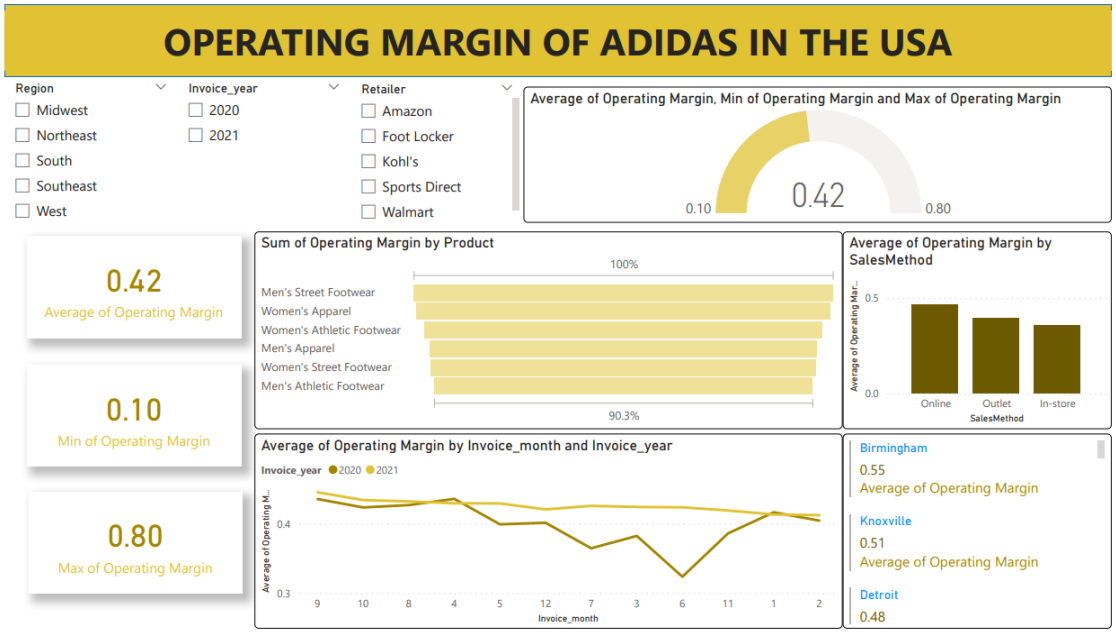



*Figure 82. Operating margin of Adidas in the USA.*

Reality remains accurate as the preference for online shopping is prioritized by many consumers. As seen on the dashboard, the average operating margin for the online shopping method is the highest, followed by the outlet, and ultimately in-store.

According to the dashboard monitoring, the operating margin by region is arranged in descending order as follows:

*Table 1. Operating margin by region.*

| Region | Operating margin | City with the highest operating margin |
|---|---|---|
| South | 0,47 | 0,55 |
| Midwest | 0,41 | 0,48 |
| Southeast | 0,42 | 0,46 |
| Northeast | 0,41 | 0,46 |
| West | 0,40 | 0,46 |

According to the analysis, the northern region of the USA appears to be the area where Adidas's business efficiency is evaluated the highest. With an average operating margin of 0.47 over two years, this could be seen as a potential focus for Adidas to invest in production and promote products to retain existing customers while attracting new customers based on the potential indicated by the operating margin figure.

### *5.3. Adidas Market Situation by Geographic Region in the USA*

- ***Comparing Performance by Region: Analyzing and comparing sales between different geographical regions in the USA.***

Looking at the regional slice, the data provides profound insights into the differences in business performance as well as preferences and consumption trends in

each region. Based on the specific percentage of sales of each product compared to total sales in each region, we can see the diversity in consumer demand.

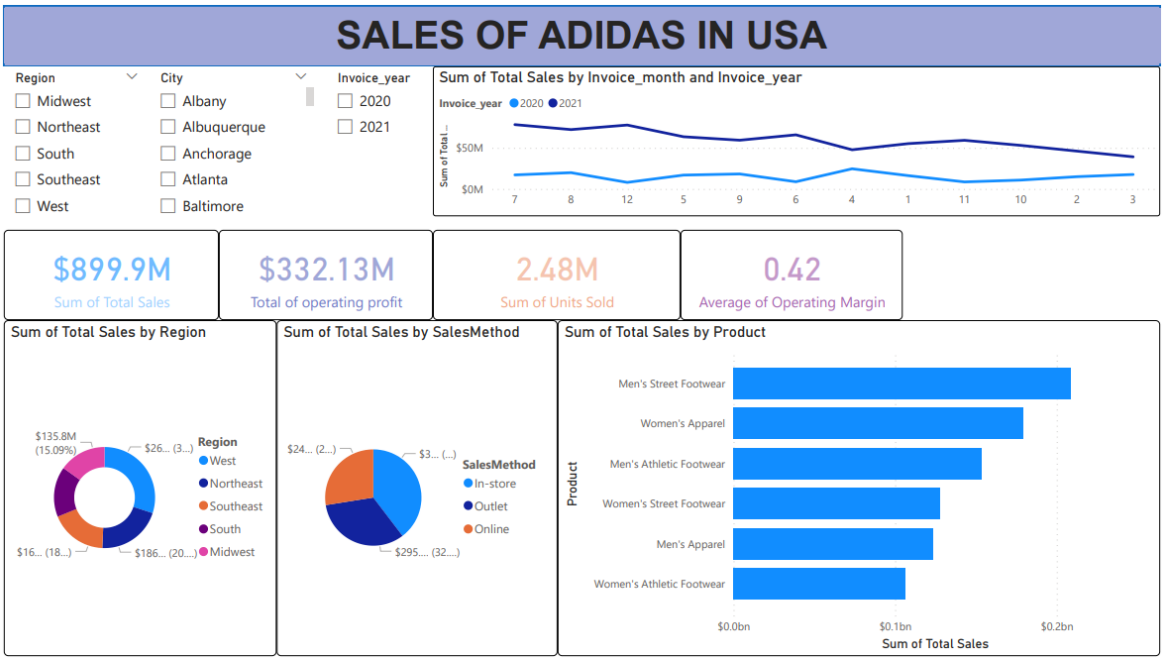



*Figure 83. Sales of Adidas in the USA*

For the West - which accounts for 30% of nationwide sales with a total turnover of $269.94 million but the lowest "Operating margin" among the 5 regions of the US with an average of 0.4 for the whole region in the 2 years of analysis, 0.02 lower than the average (0.42). Here, the strong appeal comes from street fashion products, especially Men's Street Footwear accounting for 23.21% of sales in this region, along with Women's Apparel with 19.9%, and Men's Athletic Footwear with 17.08% as key items, reflecting modern urban fashion trends and attention to personal style. With the same production costs but different Original product price causes products here to have lower profits. However, with the largest sales volume, profits in the West region are superior compared to other regions at $269.94 million. So this is an appropriate situation that requires no changes.

Behind the West is the Northeast, with 20.7% of total national sales, along with the Southeast accounting for 18.13% and the Midwest having the lowest share at 15.09%, all demonstrating a strong preference for products in the same order as the West region. The similarities between these regions reflect a common lifestyle and fashion taste trend. In

terms of profitability, these 3 regions show profitability approaching or on par with the average, indicating stable development in this area during 2020 and 2021. To maintain and further develop in subsequent years, the most effective and optimal method is to develop new products and create more product designs to maintain the competitive edge for the business.

Last but not least, although the South only has the 4th highest turnover out of 5 regions at $144.66 million, it demonstrates tremendous profitability at 0.47, exceeding the average by 0.05 and completely surpassing other regions. This not only shows that this is a viable and high potential business area that needs to be exploited and expanded in scale so that both revenue and profit can reach high levels, but also demonstrates sales effectiveness. This is a region with not too dense population, so the issue for businesses is retaining potential existing customers, thereby maintaining stable revenue and profits.

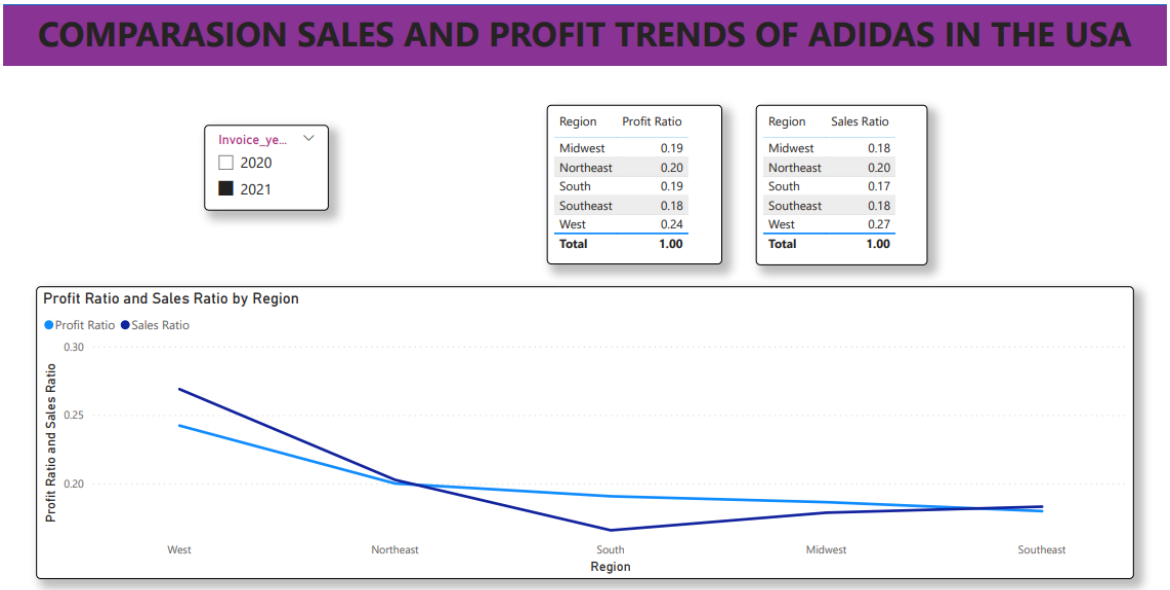



*Figure 84. Line chart of sales ratio and profit ration of Adidas in the USA in 2020*

- With regards to the differences between Profit ratio and Sales ratio by regions, we can extract some important information from the chart:
- With the area has Sales Ratio > Profit Ratio (West), Adidas needs to review the selling price structure and costs of the product/service. Although the “Price per

Unit" in this area is stable and the selling amount is outstanding, the profit is still lower than others just only because the original price of the products is too high, causing "Operating Profits" to be smaller than others

- With the area has Sales Ratio ≤ Profit Ratio (others area), the enterprises have the ability to well control operating expenses and sales expenses and products/services have high profit margins thanks to applying appropriate pricing strategies in these area. Moreover, the business operations are effective, profitable and it needs to be maintained.

- ***Target Markets and Regional Opportunities: Evaluating key target regions, development opportunities, and challenges for Adidas in these regions.***

Based on the statistics presented in the chart, the analysis should focus on 3 main issues related to the Adidas brand's operations in each distinct geographical region, including: Overall assessment of Adidas' current situation and position, development opportunities and potential challenges that may affect the business by region.

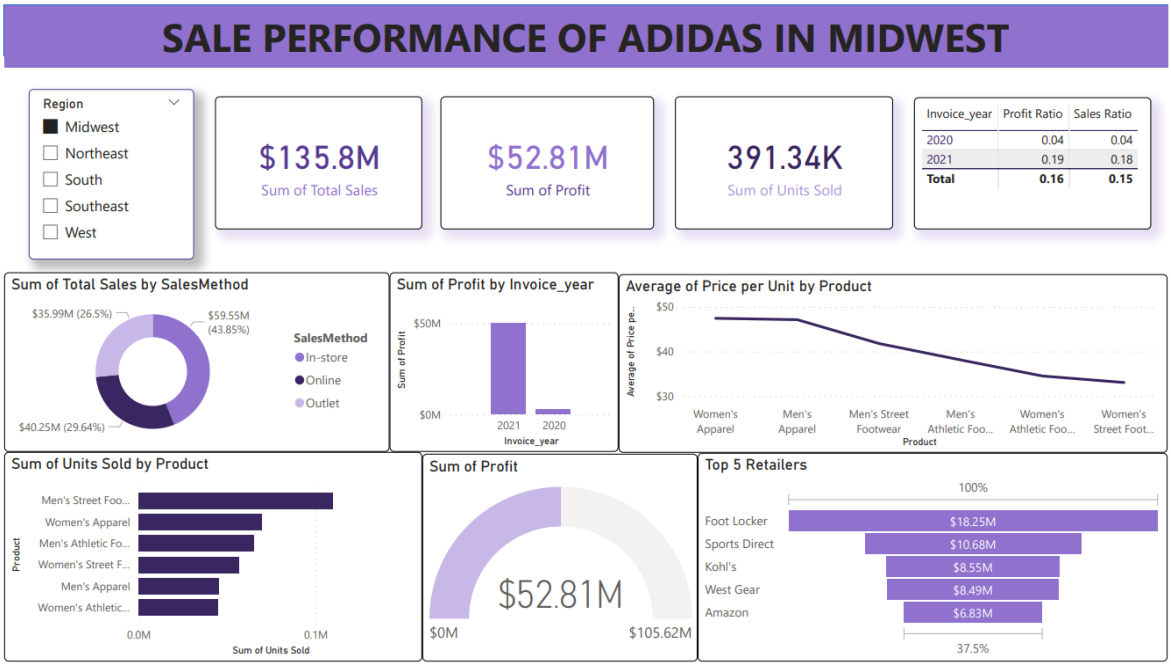



*Figure 85. Sale performance of Adidas in Midwest.*

- **Dashboard Overview**

- **Total Sales**: $135.8M.
- **Total Profit**: $52.81M.
- **Units Sold**: 391.34K.

- **Target Markets and Regional Opportunities**

- **In-store vs. Online Sales**: In the Midwest, in-store sales is the dominant channel, comprising 43.85% of total sales, which suggests that the physical retail presence is still crucial in this region. This could point towards an opportunity to strengthen in-store experiences or enhance omnichannel strategies.
- **Product preferences**: There is an uneven distribution between the types of footwear sold, with 'Men's Street Shoes' showing superiority in terms of sales compared to other products. This shows that men's products are more popular than women's products but there is still room to grow the women's sneaker segment.
- **Price Points**: Women's Apparel has the highest average price per unit sold, which may indicate a higher willingness to pay for these items, suggesting this could be a target for increased marketing or introduction of premium product lines.

- **Challenges**

- **Online Sales and Outlet**: Online sales and Outlet account for a much smaller portion than In-store in sales channels, which can be both a challenge and an opportunity for businesses. Enhancing your online shopping experience, offering exclusive online deals, or improving your digital marketing strategy can help increase this sales channel. For Outlet, businesses can come up with a strategy to strongly target this channel to increase reach.
- **Profit growth**: Although profits increased significantly from 2020 to 2021, ensuring this trend continues could be a challenge. Strategies that improve profit margins, such as cost reduction or higher-margin products, will be beneficial.
- **Retailer Dependence**: The data shows that Foot Locker is a significant retailer for Adidas in the Midwest. While this presents a strong partnership opportunity, there is also a risk associated with reliance on a single retailer. Diversifying sales channels could mitigate this risk.

**Conclusion:** For Adidas in the Midwest, the key target regions could be where in-store purchases are most prevalent, and the product focus might be on men's footwear, which seems popular. Opportunities lie in boosting women's product lines and expanding online sales. Challenges include increasing the online sales proportion and managing dependence on major retailers. Overall, a balanced approach to enhance both in-store, also online presence and outlet, coupled with targeted marketing and product development, would likely serve Adidas well in the Midwest.

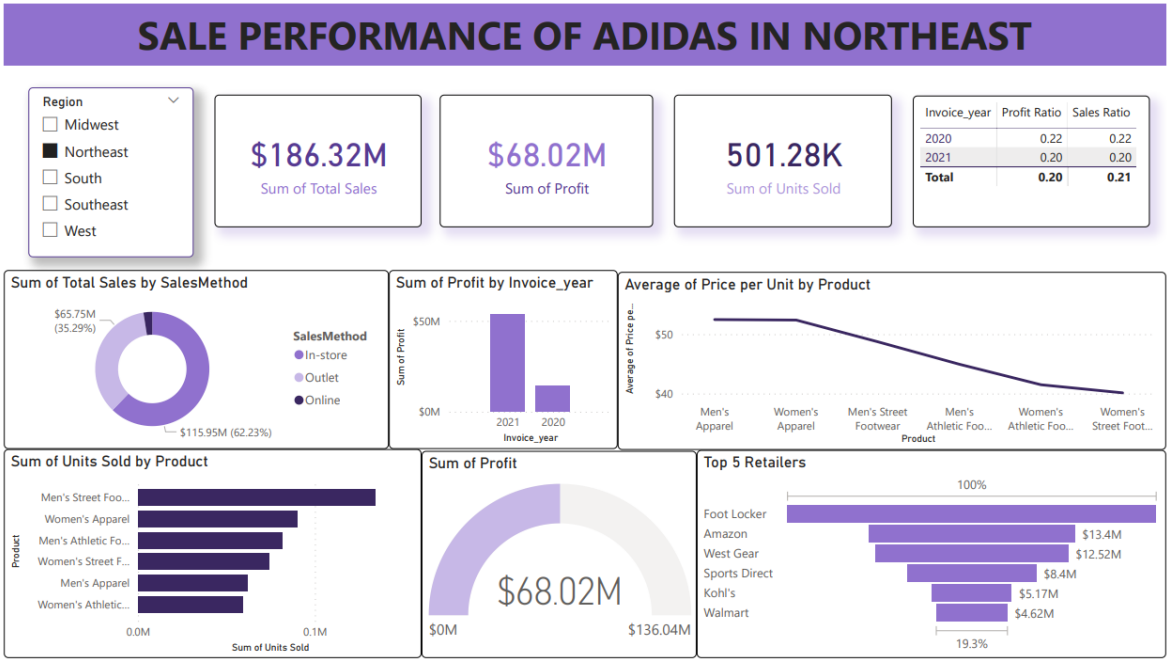



*Figure 86. Sales performance of Adidas in the Northeast*

- **Dashboard Overview**

- **Total Sales**: $186.32M
- **Total Profit**: $68.02M
- **Units Sold**: 501.28K

- **Target Markets and Regional Opportunities**

- **In-store vs. Outlet vs. Online Sales**: The Northeast shows a strong preference for in-store and outlet shopping, with in-store purchases being the dominant sales method. This suggests that physical retail experiences are highly valued and that there might be an opportunity to further enhance these experiences or to expand the number of outlets.
- **Product Preferences**: The popularity of 'Men's Street Footwear' and 'Women's Apparel' suggests these are strong categories in the Northeast. Adidas could focus on expanding these lines or providing region-specific products that cater to these preferences.
- **Pricing Strategy**: With 'Men's Apparel' and 'Women's Apparel' sharing the highest price per unit and also a significant number of units sold, this

category could be positioned as a premium offering. For women's products, the data suggests a potential to explore higher price points or premium product lines, especially in athletic footwear.

- **Challenges**

- **Online Sales**: Given the relatively small share of online sales, there is an opportunity to grow this channel. This may involve digital marketing campaigns, improved e-commerce user experience, or exclusive online offerings.
- **Year-over-Year Profit Growth**: While the profit has increased from 2020 to 2021, maintaining and furthering this growth can be a challenge, especially in a competitive market. Strategies for cost optimization and market expansion can be pivotal.
- **Retailer Performance**: Like the Midwest problem, the reliance on Foot Locker as a top retailer is evident, suggesting that maintaining strong relationships with key retailers is crucial. However, this also indicates a potential risk of over-reliance on a single retailer, emphasizing the need for a diversified retailer strategy.

**Conclusion**: For Adidas in the Northeast, the key target markets are those where in-store and outlet sales dominate. There is a clear opportunity to enhance retail experiences and possibly expand the number of outlets. Online sales represent a growth opportunity, and product development should focus on men's street footwear and women's apparel, which have shown strong sales. The challenge lies in growing online sales and ensuring profitability sustains its upward trend. Diversification of retailer partnerships and continued investment in physical retail experiences appear to be the strategic priorities for the region.

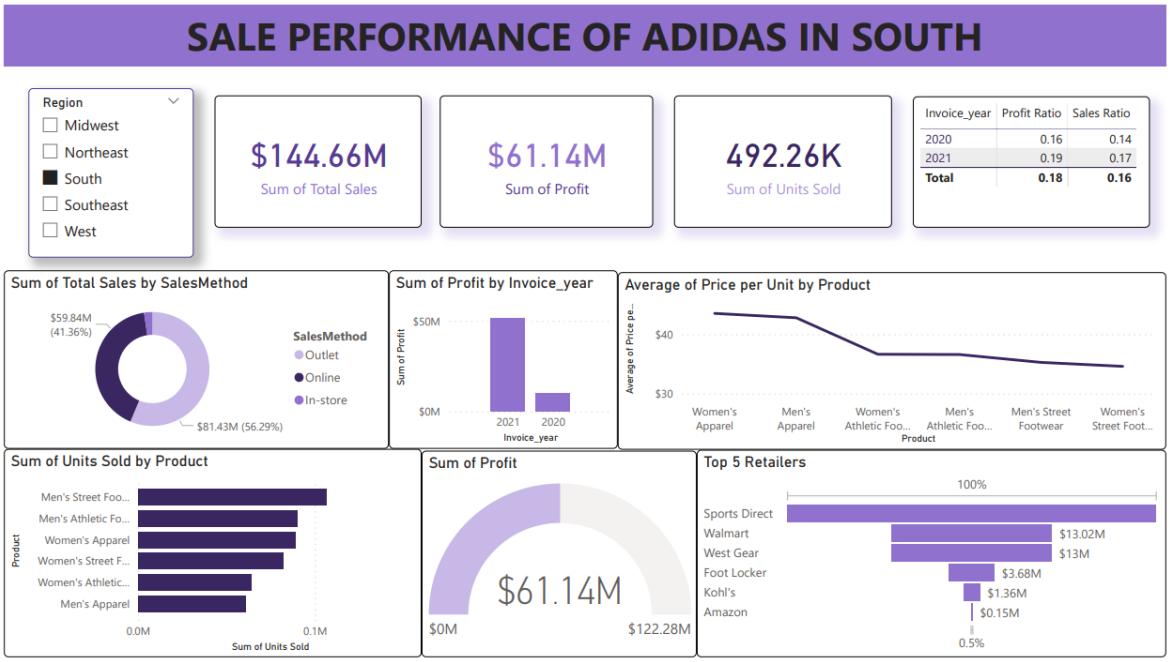



*Figure 87. Sales performance of Adidas in South*

- **Dashboard Overview**

- **Total Sales**: $144.66M
- **Total Profit**: $61.14M
- **Units Sold**: 492.26K

- **Target Markets and Regional Opportunities**

- **Comparations of Sale methods**: The Southern region also shows a preference for Outlet and Online sales while the In-store sales channel is much less effective. This is in complete contrast to Northeast - where In-store excels, showing that customers strongly prefer to shop directly at Outlet. This suggests that Adidas should continue to invest in its Outlets and possibly explore new locations for stores to exploit regional tastes.
- **Product preference**: The difference in sales volume of all products in this area is not much, meaning that the preference that customers have for these products is relatively equal. Therefore, the appropriate product strategy here is Mass marketing for all products because consumers have needs for all.
- **Pricing strategy**: Compared to the above two regions, the price structure shows that the southern region is clearly different from other regions, shown by the fact

that the "Average of Price per Unit" column uses divisions in the range $30 to $40 instead of up to $50 like the 2 areas above. From there, it shows that the prices of goods in this area are quite low, and it is possible to develop more high-end products with higher prices in these popular categories.

- **Challenges**

- **Online Sales**: The almost negligible online sales volume points to a significant area for growth. Adidas could consider strategies to enhance their online presence, improve the digital shopping experience, and offer incentives for customers to shop online.
- **Retailer Dynamics**: Dependency on Sports Direct is significant and while partnerships with Kohl's and Amazon are very low and need to be strengthened. Besides, Adidas also needs to diversify its retail channels to reduce the risk of over-reliance on a few large retailers.

**Conclusion:** For Adidas in the South, in-store and outlet sales are the dominant channels, and men's footwear is the leading product category. There's an opportunity to enhance the physical retail experience and explore new outlet locations. Online sales represent a significant growth opportunity, and there's a challenge in sustaining profit growth. A diversified retail strategy and a strengthened online presence would likely be effective strategies in this region.

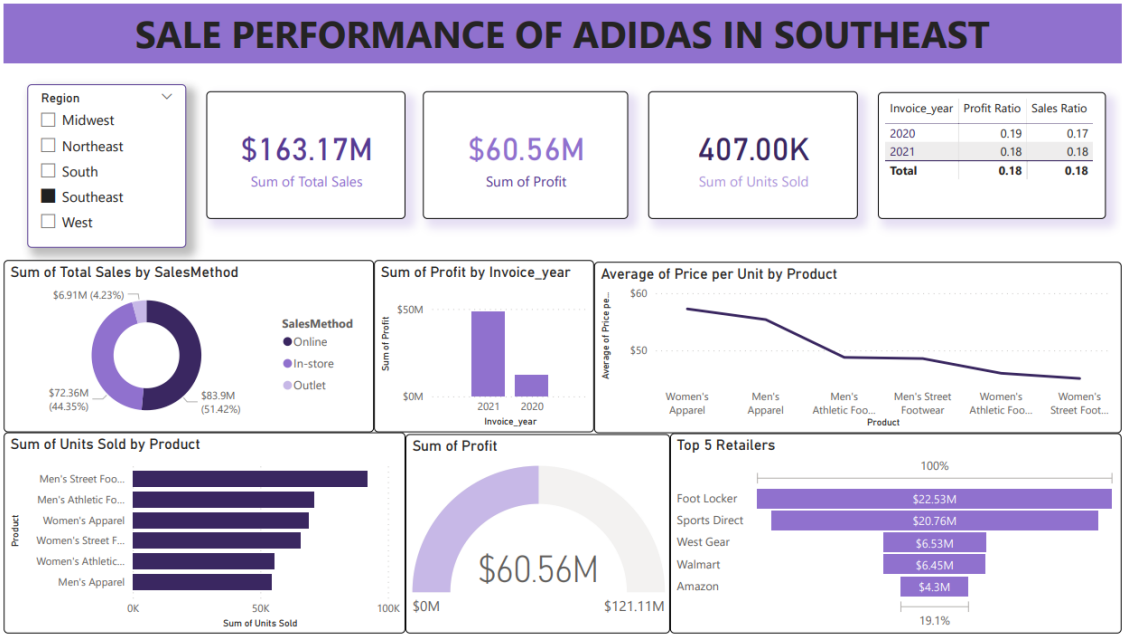



*Figure 88. Sales performance of Adidas in Southeast*

- **Dashboard Overview**

- **Total Sales**: $163.17M
- **Total Profit**: $60.56M
- **Units Sold**: 407.00K

- **Target Markets and Regional Opportunities**

- **Sales channels**: The Southeast region has a relatively different distribution between sales of sales channels, the 2 sales channels that bring the highest sales are Online and In-store, showing that both channels are important. important to Adidas. Revenue from online sales accounts for more than 50% in this region, demonstrating the opportunity to grow this channel with targeted digital strategies.
- **Product Preferences**: Men's footwear, both street and athletic, appears to be the most popular, which could be the focus for targeted marketing and inventory strategies. Women's apparel also has a significant share, which might indicate a market for potential expansion.
- **Pricing strategy**: Average prices are similar to other regions, with Women's Apparel being more expensive than most other products and other regions

representing a premium product. This shows that the current pricing strategy is not consistent across regions and may need adjustment.

- **Challenges**

- **Outlet**: With the proportion of direct sales at Outlets being relatively limited, restructuring the number of stores and increasing the development of other platforms such as e-commerce can be a great opportunity for effective growth. after.
- **High average selling price**: Setting the unit price of the product high can affect the way potential customers perceive the value of the brand, and directly affect revenue and sales.
- **Retailer Reliance**: The significant reliance on a few top retailers like Foot Locker and Sports Direct could be risky because Adidas needs to keep relationships with them to maintain the profits of the business. Therefore, diversifying sales channels could help mitigate this risk.

**Conclusion**: For Adidas in the Southeast, in-store and outlet sales are significant, and men's footwear is a strong category, with women's apparel also being prominent. Online sales are an area ripe for growth. The challenge of bias sale channels, possibly by exploring new market segments or improving operational efficiencies. It would be beneficial for Adidas to invest in enhancing its online presence and diversifying its retailer base to sustain and grow its market share in the Southeast.

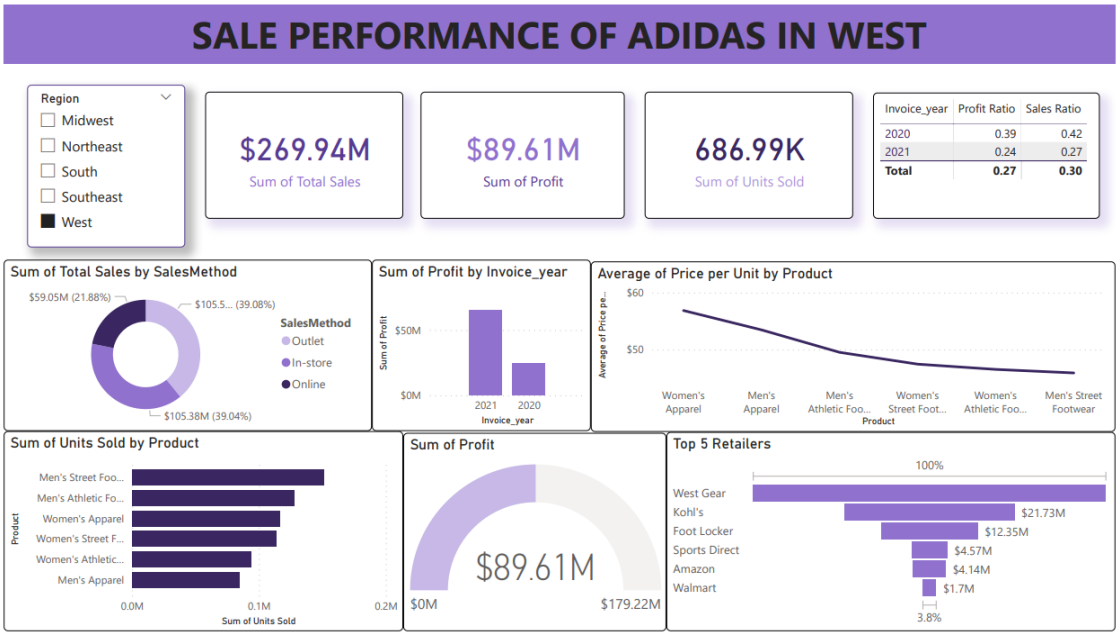



*Figure 89. Sales performance of Adidas in West*

- **Dashboard Overview**

- **Total Sales**: $269.94M
- **Total Profit**: $89.61M
- **Units Sold**: 686.99K

- **Target Markets and Regional Opportunities**

- **Sales channel**: In the West, Outlet and In-store sales methods are equally dominant in revenue (39.208% for Outlet and 39.04% for In-store), showing that direct sales channels are special. effective in this area. This suggests that there may be opportunities to further expand store locations or enhance the in-store shopping experience.
- **Product preferences**: Similar to the South region, products in the West also have a similar number of products sold, without too much difference between products. It would be advantageous to centralize marketing and availability of all products in these areas.
- **Pricing strategy**: Average price score shows that apparel costs more than shoes and the only 2 products have an average price higher than $50 in the West region.

This may indicate that there is still opportunity to introduce higher priced items or premium product lines, especially in the unisex apparel category.

- **Challenges**

- **Online Sales**: Online sales are notably low, which suggests that there is considerable room to grow this sales channel. Developing a stronger online presence and e-commerce platform could be beneficial.
- **Profit trends**: Profits appear to have increased from 2020 to 2021 but this is the region with the lowest growth rate of all regions. However, the West region has a lot of potential for maintaining and continuing to grow this profit because this is the region with the leading revenue and profit, especially when this is a competitive market like the West.
- **Retailer Dynamics**: The reliance on specific retailers such as West Gear and Kohl's is prominent. To reduce risk and increase market reach, Adidas could benefit from diversifying its retailer partnerships and enhancing its direct-to-consumer channels.

**Conclusion**: For Adidas in the West, there is a clear preference for outlet sales, and the product focus could be on men's footwear. Opportunities may lie in expanding the outlet model and enhancing online sales capabilities. The challenges include ensuring sustained growth in profit and reducing reliance on a few top retailers by diversifying sales channels and possibly increasing the emphasis on direct-to-consumer marketing and sales.

- ***Target Markets and Regional Opportunities: Evaluating key target regions, development opportunities, and challenges for Adidas in these regions.***

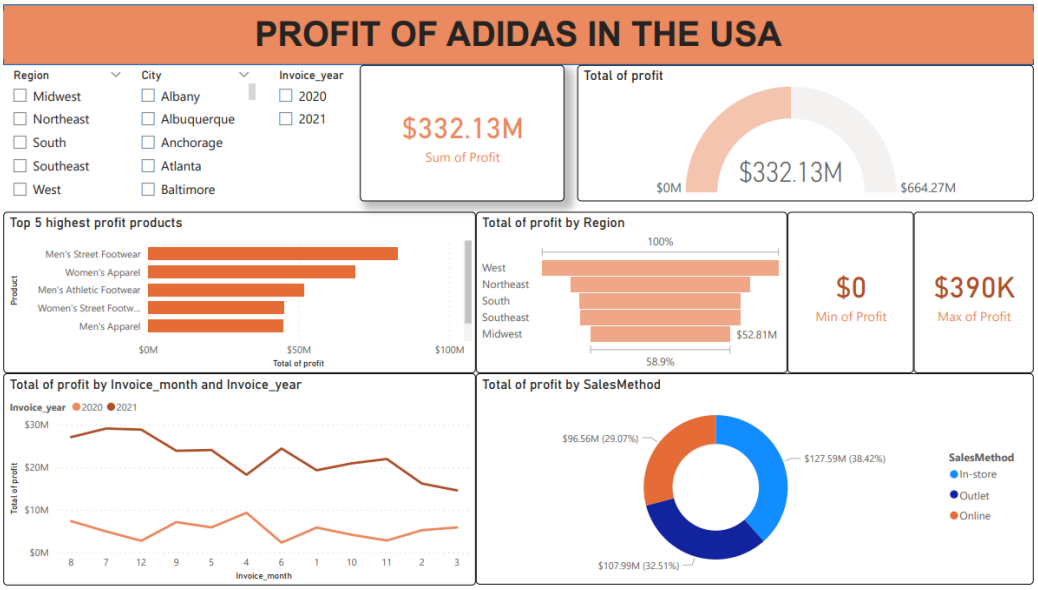



*Figure 90. Profit of Adidas in the USA*

Geographical factors profoundly influence Adidas' business strategy, as clearly demonstrated through the differences in revenue and profits of each region. Therefore, it is necessary to thoroughly assess these factors in order to develop appropriate strategies for the company's growth. Some factors to consider include:

- **Key Target Regions**:
  - West: This region has the highest total profit, making it a key target area for Adidas. Focusing on maintaining and increasing market share here is essential.
  - **Northeast**: Although not the highest, it shows significant profit, indicating it is a substantial market for Adidas. The company could look into ways to optimize or expand their presence here.
  - **South and Southeast**: With lower profit figures, these regions represents a potential growth opportunity. Both south and southeast show superiority in the Online channel (over 40%), this shows a large development potential for e-commerce in this region.
  - **Midwest**: The profit data for these regions isthe lowest one, but sales volume across all three distribution channels in this area is relatively balanced, showing that there is still potential for continued growing.
- **Development Opportunities**:

- **Product Focus**: The top 5 highest profit products are led by Men's Street Footwear and Women's Apparel. These categories could be the focus of new product development or marketing campaigns.
- **Sales Method**: In-store sales represent for a significant portion of profits but Online is the lowest one, showing that traditional sales are still dominant. Future trends are showing that digital channels are essential for growth, so enhancing online customer experience and e-commerce capabilities could be a major growth opportunity.

- **Challenges**:

- **Regional Disparities**: There are stark differences in regional performance, with the West outperforming others by a wide margin. Understanding why other regions are lagging—be it due to less brand recognition, distribution challenges, or competitive pressure—is crucial for addressing these disparities.
- **Seasonal Fluctuations**: The "Total of profit by invoice_month and invoice_year" chart indicates that seasonal sales trends do not exist from year to year, as evidenced by the amount of sales shown symmetrically across the "Total of profit" axis. This can make inventory management difficult but creates opportunities for a year-round marketing strategy
- **Product Saturation**: The data suggests that Adidas might be relying on a few key products for its profit. Diversifying the product line to reduce dependency on these products could be a strategic move to mitigate risk.

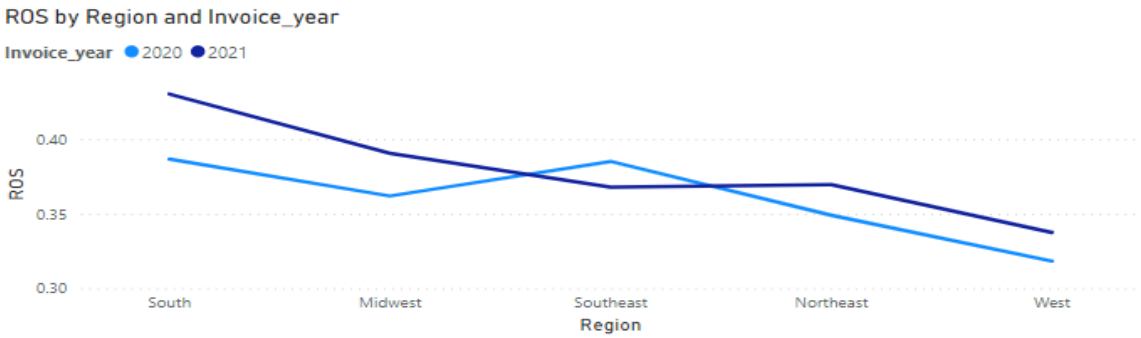



*Figure 91. ROS on Region and Invoice_year*

There are a range of insights that can be drawn through the chart "ROS by Region and Invoice_year":

- **Trend by Year**: There are two lines representing two different years, 2020 and 2021. The line for 2021 starts higher than the line for 2020, indicating that the ROS was generally higher in 2021 across all regions compared to 2020.
- **Regional comparison**: The chart includes five regions: South, Midwest, Southeast, Northeast and West. There is a notable difference in ROS between these regions in both years, which can be divided into 2 groups including:
- ROS in 2020 > ROS in 2021 (Southeast): This might be due to a decrease in revenue or an increase in costs / original price of products, causing profits to decrease. It proves that business operations are ineffective or because there are too many competitors. Businesses need to analyze their strategies carefully and invest more heavily in this area if necessary
- ROS in 2021 > ROS in 2020 (Remaining Areas): This shows that the business performance of the business has improved compared to the previous year, bringing higher profits and the business seems to be operating more efficiently than before. However, to maintain this trend for the following years, businesses need to closely follow customer purchasing trends.

# CHAPTER 6: CONCLUSION AND RECOMMENDATIONS

Finally, the research team will provide general comments and conclusions about the research on building a data warehouse for the purpose of analyzing revenue data. From there there are recommendations on this topic.

## *6.1. Conclusions*

This comprehensive analysis provides a multifaceted view of Adidas' business situation in the United States, encompassing revenue trends, sales methods, geographic variations, and retailer dynamics. By examining performance metrics over 2020 and 2021, critical insights have emerged, guiding strategic considerations for the brand's continued success.

A meticulous evaluation of revenue over time reveals a substantial difference between 2020 and 2021. The impact of the COVID-19 pandemic is evident, with 2021 showcasing an upward trajectory, especially in the latter half of the year following positive shifts in the pandemic situation. The anomaly in revenue during the pandemic period underscores the brand's resilience and adaptability.

Detailed analyses of revenue by method elucidate the adaptive strategies employed by Adidas. In response to supply chain disruptions in 2020, emphasis on In-store and Outlet methods was pronounced. However, in 2021, stability prevailed, and the online method witnessed notable development, reflecting a balanced investment across all three forms.

Geographically, a nuanced examination of sales across regions unravels distinctive consumer preferences and market dynamics. The West emerges as a frontrunner, driven by a fervor for street fashion products, while the South demonstrates tremendous profitability, surpassing other regions. The Northeast and Southeast share common preferences, indicating opportunities for strategic product development.

Retailer-specific analyses reveal diverse approaches among retailers, with Foot Locker prioritizing online sales and West Gear excelling in in-store performance. These variations stem from factors such as store distribution and regional presence, emphasizing the importance of tailored strategies for each partner.

## *6.2. Proposed Business Strategy for Adidas in the Coming Period*

### 6.2.1. Product Innovation and Improvement.

Proposals for product development focus on high-performance product lines or those with growth potential.

Overviewing Adidas' sales data in the USA market, product lines such as “Men’s Street Footwear”, “Women’s Apparel”, and “Men’s Athletic Footwear” are performing well with significant growth potential compared to other products. Therefore, Adidas could introduce more designs and styles to improve these lines to attract more customers. Additionally, the company should continuously research market trends and understand customer preferences to propose innovations in design and technology, such as using sustainable or flexible materials in athletic shoes or integrating smart technology into products.

Adidas also needs to assess customer needs across all regions in the USA to see which products are popular and stable in sales, which have excessive inventory leading to reduced profits, and suggest products that particularly suit the needs of customers in those regions. Moreover, Adidas might consider partnerships with athletes or celebrities to develop unique products, enhancing the appeal and uniqueness of their products. To these days, Adidas has achieved significant success by collaborating with big names like Kanye West, Lionel Messi, etc., but there are still many markets where Adidas is gradually losing its position to its biggest competitor, Nike.

### 6.2.2. Optimizing Marketing Strategy.

Recommendations on marketing and advertising strategies, based on understanding customer behavior and the performance of past campaigns.

For marketing strategies to be most effective, Adidas' data analysts need to thoroughly analyze customer shopping behavior and assess the profitability and efficiency of previous advertising campaigns to identify the most cost-effective approaches. Personalized marketing strategies based on customer data are likely to be very effective. With three main sales channels - In-store, Online, and Outlet - analyzing the revenue from these channels, combined with data on best-selling products, will help solve the problem of which advertisements to run on which channels, instead of having to run the same ads across all three.

Adidas could also leverage social media and digital communication channels to enhance interaction and engagement with customers. The percentage of profit from products sold In-store in 2020-2021 was 38.42% of total sales product. This indicates a less preference for online shopping compared to offline shoping. Therefore, Adidas should builds an interactive marketing campaigns, like virtual try-on systems using new technologies like AI or online contests, to attract online customers and increase interaction and brand recognition. Using big data analytics could enhance accuracy, optimizing the effectiveness of marketing campaigns.

### 6.2.3. Expanding and Optimizing Distribution Channels.

Suggestions for expanding or optimizing current distribution channels, such as enhancing online sales or expanding the retail network.

Adidas needs to assess the effectiveness of its current distribution channels and control the improvements or add new channels to increase brand reach. Strengthening online sales aligns well with the company's development direction. To keep up with the continuously changing online trends, businesses need to optimize the online shopping experience, like improving the website interface, enhancing interactive features, and providing high-quality customer support to increase trust and loyalty. Additionally, opening new stores in potential areas after analyzing profitability and risks is a good

approach to reach new customers. Understanding shopping behavior in different regions is crucial, as customer preferences for distribution methods, like fast shipping or convenient returns, vary by location.

Moreover, Adidas needs to continuously research the market to identify opportunities for expanding distribution channels in emerging or untapped areas, keeping pace with societal needs and changes to avoid obsolescence. If a distribution channel becomes ineffective, the analytics team should reassess and close it if it affects the company's profits too much. Furthermore, developing partnership programs with retailers and distributors to expand product reach is a good way to attract new customers and allow them to experience the products.

### 6.2.4. Focusing on Target Markets and Potential Regions.

Proposing strategies focused on specific geographical areas or target markets with growth opportunities.

In analyzing sales data, Adidas should allocate resources geographically to identify areas with significant growth potential to boost market development, thereby expanding their customer base. Alongside geographical market analysis, the company should also devise specific marketing and product strategies for each area, based on the unique preferences and needs of customers there. The simplest and most effective way is to base these on best-selling products to gauge their needs. Understanding local culture and tastes is also vital to integrating into business strategies, thereby creating a competitive difference. To address this, companies can use local employees to learn about the habits and preferences of the locals, giving an overall view of their lifestyle.

Developing promotional programs and implementing special marketing activities targeted at these areas, based on their needs, habits, and preferences, as mentioned earlier, helps draw their attention to the products, thereby boosting sales. Moreover, integrating cultural and geographical data into advertising will

create a strong connection with customers, making them feel more valued as they see businesses understanding and catering to their desires. If a company lacks local staff to understand the culture, conducting market surveys to understand the needs and wants of customers in targeted areas is a perfect alternative.

#### 6.2.5. Focusing on Sustainability and Social Responsibility

Suggestions for integrating sustainable goals and social responsibility into the business strategy.

To maintain and develop sustainably, Adidas needs to create appropriate sustainable initiatives, like using recycled materials or minimizing environmental impact during production. This contributes significantly to creating a positive image of the business in consumers' eyes. This not only improves relationships with existing customers but also attracts new customers who support environmental issues. Moreover, integrating social responsibility goals into business strategies significantly improves the company's image. Activities include supporting local communities or creating scholarships for students in difficult circumstances, helping those in need in society.

Additionally, integrating sustainability standards throughout the supply chain, from raw materials to production and distribution, makes the business more stable in the future, avoiding lawsuits for negatively impacting the environment. Adidas can also organize events and community activities to raise awareness about the environment and social responsibility, with attendees potentially being loyal customers from the area. Such events are also an excellent opportunity for the company to launch environmentally friendly products and test their reception. Along with benefits for the business, developing training and capacity-building programs for the community also contributes to improving the quality of life and fostering sustainable development.

# REFERENCE


[1] D. Theo Thiên, "Adidas: Hành trình từ nhà kho tiệm giặt là đến một trong những thương hiệu giày thể thao lớn nhất thế giới," *Cafebiz,* 2023.

[2] V. Theo Hòa, "Các "ông lớn" thời trang thể thao thiệt hại trước đại dịch Covid-19," *Tạp chí Tài chính,* 2020.

[3] "adidas revenue 2000-2022 | Statista," 29 4 2023. [Trực tuyến]. Available: https://www.statista.com/statistics/268416/net-sales-of-the-adidas-group-worldwide-since-2000/.

[4] T. Mộc, "Thế giới năm 2020 đầy biến động," *Báo Công an Nhân dân điện tử,* 2021.

[5] H. Thanh, "Các hãng thời trang khốn đốn trong năm đại dịch COVID-19," *Vietnam+ (VietnamPlus),* 2021.

[6] R. Simiao, C. Patrick và C. Tsun-Ming Jason, AI-Based Fashion Sales Forecasting Methods in Big Data era, 2018.

[7] Randhawa, "Retail Analytics," *International Series in Operations Research & Management Science,* 2019.

[8] Soumendra Mohanty, Madhu Jagadeesh và Harsha Srivatsa , "Big Data Imperatives Enterprise Big Data Warehouse," *BI Implementations and Analytics,* 2013.

[9] Gregorczuk, "Smart-Stores Saving Bricks and Mortar Retail or a Privacy Problem?," *Law, Technology and Humans,* 2022.

[10] A. Bouchra, K. Larbi, W. Abderrahim Ait và S. Abderrahim, "Linking context to data warehouse design," *International Journal of Advanced Computer Science and Applications,* tập 1, p. 10, 2019.

[11] Helena Brajković, Danijela Jakšić và Patrizia Poščić, "Data Warehouse and Data Quality – An Overview," *Proceedings of the Central European Conference on Information and Intelligent Systems,* p. 22, 2020.

[12] H. Alaa Khalaf, U. Maysaa Abd, H. Hisham Noori, M. Zahraa Abdulkareem và S. Gadeer Mustafa, "Improve HR Decision-Making based on Data MaRt and OLAP," *Journal of physics,* tập 1, p. 1530, 2020.

[13] JASIM MOHAMMED DAHR, ALAA KHALAF HAMOUD, IHAB AHMED NAJM và MOHAMMED IMAD AHMED, "Implementing sales decision support system using data mart based on olap, kpi, and data mining approaches," *Journal of Engineering Science and Technology,* tập 1, p. 17, 2022.

[14] R. Simiao, C. Patrick và C. Tsun-Ming Jason, AI-Based Fashion Sales Forecasting Methods in Big Data era, 2018.

[15] T. Lê Triệu, “Overview of Big data analytics in Ecommerce,” *TNU Journal of Science and Technology,* 2020.

[16] Panos Vassiliadis và Alkis Simitsis, “Extraction, tranformation and loading”.

[17] RICHMOND, “What Is the Importance of KPIs for Performance Measurement?,” *Unboxed Technology,* 2021.

[18] A. Dinesh, “Implementing slowly changing dimensions (SCDs) in data warehouses,” *SQL Shack,* 2021.

[19] Hanhtm, “SCD (Slowly Changing Dimension) là gì? Các loại SCD và ví dụ cụ thể,” *techover.io,* 2023.

[20] S. Ibrahim H., “THE USE OF ONLINE ANALYTICAL PROCESSING (OLAP) FOR BUSINESS INTELLIGENCE,” *ResearchGate,* 2007.

[21] Peter-Myers, “Understand star schema and the importance for Power BI - Power BI,” *Microsoft Learn,* 2023.

[22] A. Dinesh, “Implementing slowly changing dimensions (SCDs) in data warehouses,” *SQL Shack ,* 2021.