Assignment 2

LA TROBE
UNIVERSITY

**Subject:** DATA WAREHOUSING
**Assignment Title:** Data Warehousing and Big Data
**Student Name:** Parshad Nirbhavane
**Student ID:** 2147050

## Contents

## Task 1:

### Firstly, improving the current dimensional model/ warehouse to capture all relevant information that applies to customers, such as demographics.

The goal is to better understand customers by adding more relevant information to the DimCustomer table and potentially creating new dimension tables. The current DimCustomer table has some basic demographic information (e.g., Age, Gender, Education), but we need to capture a fuller picture of the customer to make more informed decisions about how to market and sell to them

### 1) What's Missing in the Current Dimensional Model?

The current customer data lacks key demographic details like income, household size, and occupation.

We also need better geographic information to understand where customers are coming from and how location impacts their shopping behavior.

Proposed Enhancements:

a) Income Level

Classifying customers into income brackets (Low, Medium, High) helps businesses understand their purchasing power. People with higher incomes are often more willing to spend on premium products, while those with lower incomes might look for discounts or budget-friendly options. By recognizing these differences, marketing campaigns can be tailored to match each group's needs, resulting in more effective promotions and ultimately, increased sales.

b) Marital Status

Knowing a customer's marital status—whether they are single, married, or divorced—can significantly impact buying behavior. Families, for example, tend to purchase in bulk, while singles might focus more on convenience and smaller quantities. By tailoring marketing campaigns to match these lifestyle differences, businesses can better meet the needs of families versus individuals.

c) Occupation

A customer's profession or industry plays a big role in their spending habits. Corporate professionals often have more disposable income, while students or entry-level workers may prioritize more affordable options. Understanding this allows businesses to target certain products or services to specific job groups, such as offering high-end products to professionals or budget options to students.

d) Household Size

The number of people in a household can affect shopping behavior. Larger households tend to buy in bulk, while smaller ones may make fewer, smaller purchases. By understanding the size of a customer's household, businesses can tailor promotions, like offering bulk-buy discounts to larger families and single-serve options to smaller households or individuals.

e) Geographic Information (New Dimension: DimGeography)

Customer location—whether country, state, city, or postal code—provides insight into regional buying trends and preferences. Geographic data helps businesses understand differences in customer behavior across regions. By tailoring campaigns and product availability to local preferences, businesses can ensure that the right products are promoted in the right places, enhancing customer satisfaction and boosting sales.

f) Loyalty Program Details

Understanding a customer's loyalty program tier—whether they're in Gold, Silver, or Bronze—and tracking their accumulated points helps businesses engage customers more effectively. Customers in higher tiers can be rewarded with exclusive offers, while lower-tier customers can be incentivized to increase their spending to move up. This creates a sense of reward and motivates customer loyalty, leading to repeat purchases.

## Using current data in the warehouse to better understand the customer base by identifying

Now that we've enhanced the data model with more customer information, the next step is using that data to understand customer behavior and spot sales and marketing opportunities. Here's how:

## What Can Be Done with the Improved Data?

a) Customer Segmentation for Personalized Marketing:

With more detailed demographic information (like income, marital status, and household size), we can break customers into targeted segments. For instance, customers with higher incomes might get promotions for premium products, while larger households could be offered bulk-buy discounts. By tailoring the marketing to each group's specific needs, we can boost engagement, increase sales, and build stronger customer loyalty.

b) Understanding Purchase Behavior by Region:

Using geographic data, we can figure out how customers in different areas shop. Some products might sell better in cities, while others could be more popular in rural areas. For example, a particular brand might do great in metropolitan regions but not perform as well in the suburbs. By recognizing these patterns, the company can optimize inventory and marketing efforts, reducing waste and maximizing sales.

c) Loyalty Program Optimization:

By keeping track of loyalty program participation, the business can spot high-value customers and offer them exclusive rewards. For instance, customers in the "Gold" tier might get early access to sales, while "Silver" tier customers could be incentivized to spend more and move up to "Gold." This approach strengthens customer retention, as people are more likely to stick around and make repeat purchases when they feel valued.

d) Tracking Household Buying Patterns:

Household size data can reveal trends in buying behavior, like whether a family is more likely to buy in bulk or not. Larger families, for example, might buy in bulk more frequently, so offering discounts on bulk items could be an effective strategy. Not only does this boost sales, but it also helps with inventory management by predicting demand for bulk products in specific customer segments.

e) Sentiment Analysis with Customer Feedback:

Analyzing customer feedback and linking it to demographic data (age, income, etc.) can help the company understand how different groups feel about certain products. For instance, if higher-income customers consistently complain about product value, it could indicate a need for higher-quality offerings. This insight can improve both product development and marketing strategies, ensuring they align better with customer expectations.

## Dimensional Model Re-design

### Goal:

The main goal is to optimize the current dimensional model to capture more detailed customer information and improve the business insights. This redesign focuses on adding new attributes, dimensions, measures, and design techniques that give us a better understanding of customer behavior and potential marketing and sales opportunities.

### 1. Enhancing Existing Dimensions

### I. DimCustomer:

The **DimCustomer** table is the heart for storing all the demographic and customer profile data. To really meet the business needs, the following changes are proposed:

**New Attributes:**

- a) **Income Level**: We can categorize customers into income brackets like Low, Medium, and High. This will help segment the market better and tailor marketing strategies for different income groups.
- b) **Marital Status**: Whether a customer is single, married, or divorced is important to know because it can impact their buying habits. We can use this to target customers based on their life stages.
- c) **Occupation**: A customer's job or industry can give us a lot of insight into their spending habits. Corporate workers might spend more, while students might look for cheaper options.
- d) **Household Size**: Adding a field for how many people are in a customer's household can help us tailor bulk purchase offers to families vs individuals.

### II. DimProductCategory:

To improve product-related data, we can add the following enhancements to the **DimProductCategory** table:

- a) **Product Type**: Classifying products into more specific types like essential, premium, luxury can give us better analytics.
- b) **Brand Loyalty**: Tracking how loyal a customer is to a specific brand within the product category can help us measure and improve brand performance.

### III. DimTransactionDate (Expanded to DimTime):

Instead of just tracking the transaction date, we should expand **DimTransactionDate** into a full **DimTime** table to enable more complex time-based analysis:

- a) **Time of Day**: Knowing the hour of a transaction helps us understand peak shopping times.
- b) **Day of the Week**: We need to capture which days are busier to target promotions for specific times.
- c) **Holiday Indicator**: Knowing if a transaction happened during a holiday lets us see how holidays impact sales and adjust promotions accordingly.

## 2. Introducing New Dimensions

### a) DimGeography:

This dimension will store geographic information, which is crucial to understanding regional customer behavior and buying preferences.

**Attributes:**

- a) Country
- b) State/Region
- c) City
- d) Postal Code
- e) Store Location (for in-store purchases)

### b) DimLoyaltyProgram:

To better understand customer loyalty, we should add a new dimension, **DimLoyaltyProgram**, to track:

- a) **Loyalty Program Tier**: Whether a customer is in the Gold, Silver, or Bronze tier of our loyalty program.
- b) **Points Accumulated**: How many loyalty points a customer has earned over time.
- c) **Enrollment Date**: The date when the customer joined the loyalty program.
- d) **Customer Status**: Whether the customer is active or inactive in the program.

## 3. New Measures in the Fact Table (FactTransactionSummary)

In addition to the usual measures like **TotalQuantity** and **TotalSales**, we should add new metrics to the **FactTransactionSummary** table for better analysis:

- a) **Discount Amount**: The total discount applied to each transaction.
- b) **Profit Margin**: The profit made on each transaction after subtracting costs.
- c) **Customer Lifetime Value (CLV)**: The total value that a customer has contributed over their lifetime.
- d) **Average Transaction Value**: How much customers are spending on average in each transaction, which can give us insight into buying habits.

## 4. Dimension Design Techniques

### a) Slowly Changing Dimensions (SCD Type 2) for DimCustomer:

We should implement SCD Type 2 for the **DimCustomer** dimension to track historical changes in customer data like marital status, income, or occupation. This way, we can analyze how a customer's profile changes over time and how that affects their behavior.

### b) Role-Playing Dimensions for Time-Related Data:

**DimTransactionDate** should be split into multiple role-playing dimensions for different types of date analysis, such as:

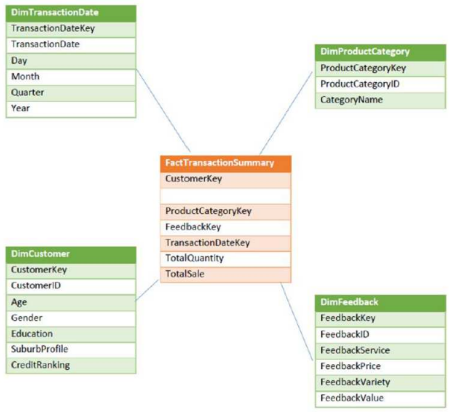
- a) Transaction Date
- b) Order Date
- c) Delivery Date

## 5. Sentiment Analysis and Enhanced Feedback (DimFeedback Enhancements):

We should enhance the **DimFeedback** table to improve our analysis of customer feedback. The following new fields are suggested:

- a) **Feedback Sentiment**: A score that represents the customer's feedback as Positive, Neutral, or Negative.
- b) **Net Promoter Score (NPS)**: A metric to see how likely customers are to recommend our products.
- c) **Customer Suggestions**: A field to capture any specific suggestions provided by customers in their feedback.

Before:

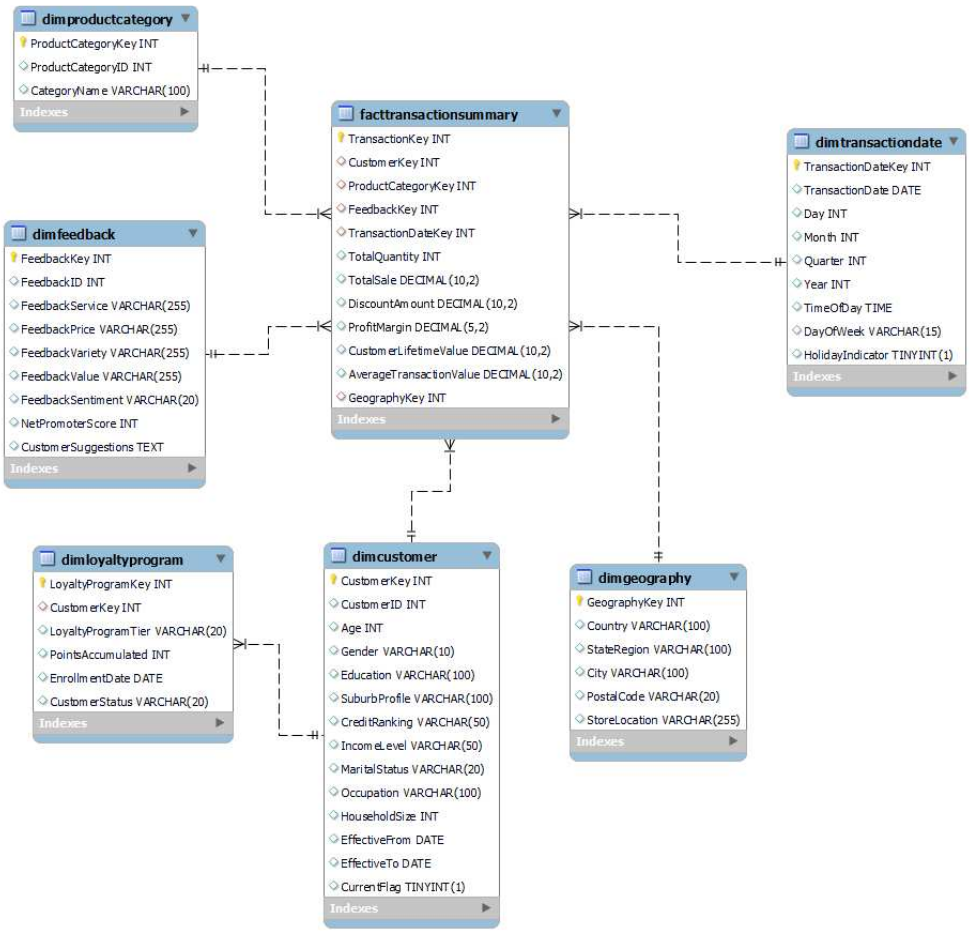



## 5. Sentiment Analysis and Enhanced Feedback (DimFeedback Enhancements):

We should enhance the **DimFeedback** table to improve our analysis of customer feedback. The following new fields are suggested:

- a) **Feedback Sentiment**: A score that represents the customer's feedback as Positive, Neutral, or Negative.

- b) **Net Promoter Score (NPS)**: A metric to see how likely customers are to recommend our products.
- c) **Customer Suggestions**: A field to capture any specific suggestions provided by customers in their feedback.




To really get a better grasp on our customer base and figure out their behaviors, we can use the current data in our warehouse to identify potential sales and marketing opportunities. Here's a few ways this data model could be used to find some valuable insights:

1. **Segmenting Customers**:

   The **DimCustomer** table gives us a lot of info like **Age**, **Gender**, **Education**, and **SuburbProfile**. By grouping customers based on these characteristics, we can start to see patterns. For example, maybe certain age groups prefer certain products. If we can figure that out, we can create marketing campaigns that target those specific groups, rather than a blanket approach. Like, if customers with a higher **CreditRanking** seem to make bigger purchases, we could push more expensive products towards them.

2. **Product Preferences**:

   **Feedback and Satisfaction**:
   The **DimFeedback** table is full of customer feedback, covering areas like **Price**, **Variety**, and **Service**. By connecting this feedback to customer data, we can see if certain groups are more or less satisfied with what we're offering. If, for example, younger customers are always

unhappy about the variety, we'd know to expand our options for that particular demographic. This gives us a good chance to make improvements that directly affect customer happiness.

3. **Regional Insights**:

   Though we don't have geography details here, combining this with external data could be a game-changer. We could look at buying patterns by region and tweak marketing or even adjust inventory to match regional needs. If customers in one area buy in bulk more often, we might want to make sure that region is well-stocked with bulk items.

4. **Time-Based Purchase Patterns**:

   The **DimTransactionDate** table lets us analyze when purchases are happening most often, whether it's daily, monthly, or seasonally. We could find that certain customer groups, like younger buyers, tend to shop more during holiday periods. Knowing this, we could time promotions or special deals to match their buying habits, rather than missing out on key times.

## Task2: Data Analytics Platform for Geographical Expansion Using Microsoft Technology Stack

The objective was to create a data warehouse and perform analytical operations by constructing a multidimensional cube using SQL Server Data Tools (SSDT), which is integrated into Visual Studio. The cube creation is based on the Wide World Importers Data Warehouse (DW), which was accessed through the BUS5WB server.

### 1. Connecting to the Data Source

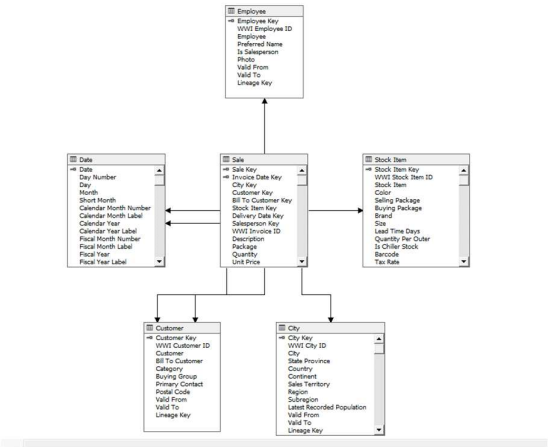
The first step involved establishing a connection to the data warehouse stored on the BUS5WB server. Using SSDT within Visual Studio, the connection was configured by creating a data source to link the Wide World Importers Data Warehouse to the project. This data source is essential because it forms the foundation for building cubes and dimensions in SSDT.

### 2. Creating Data Source Views

The next step was to create a Data Source View (DSV), which acts as a virtual representation of the underlying tables in the data warehouse. This step allows for the selection of relevant tables and columns that will be used in the cube.

Steps to Create the Data Source View:

- A new Data Source View was created by selecting all relevant tables (fact and dimension tables) from the data warehouse.
- The Fact and Dimension tables were selected, including Sale, Customer, City, Stock Item, Employee, and Date.




## 3. Cube Creation and Configuration

After the DSV was created, the process of building the cube began. The cube is the primary structure used for analyzing multidimensional data, and it allows for slicing and dicing the data along different dimensions such as time, location, and product.

Steps to Create the Cube:

- A new cube was created by selecting Use Existing Tables.
- The FactSales table was used as the Measure Group for the cube. The measure group consists of numerical data (measures) that we want to analyze. The main measures selected include Total Sales, Profit, and Quantity.
- All dimensions, including Date, City, Stock Item, Customer, and Employee, were selected for the cube.

## 1. Data Source and Cube Design

The cube is designed based on data extracted from the Wide World Importers Data Warehouse (DW). Using SSDT, a connection to the BUS5WB server was established, and the data warehouse was integrated into the cube structure.

### Data Source View (DSV)

The Data Source View (DSV) consists of several fact and dimension tables that form the backbone of the cube. These tables include:

- Fact Table: Sale
- Dimension Tables: Employee, Customer, City, Stock Item, and Date

This structure forms a star schema, with the Sale fact table at the center, connected to various dimensions for multidimensional analysis.

## 2. Key Components of the Cube

### Measure Group: Fact Table

The central fact table is Sale, which contains key measures used for analysis. These measures allow the company to assess key performance indicators (KPIs) related to sales, profitability, and regional success.

Key Measures include:

- Total Sales: Total value of sales transactions.
- Profit: The profit associated with each sale (derived from unit price and cost).
- Quantity Sold: The total number of products sold in each transaction.
- Unit Price: The selling price of individual stock items.

### Dimension Tables

1. Employee Dimension: Represents the sales employees, including whether they are salespeople, their ID, and other personal details.
2. Customer Dimension: Represents customers, including their category, buying group, and postal code, providing insights into customer behavior across regions.
3. City Dimension: Represents geographical data, including city, country, region, and population. This dimension is critical for analyzing the company's geographical expansion.
4. Stock Item Dimension: Contains product information such as stock item name, brand, size, and tax rate. This dimension enables product performance analysis.
5. Date Dimension: Time-based analysis is possible using the date dimension, which allows for drilling down from Year > Quarter > Month > Day.

## 3. Multidimensional Analysis for Geographical Expansion

The cube was built to enable multidimensional analysis across different aspects of the business, with a focus on regional performance. By leveraging the relationship between fact and dimension tables, the company can extract meaningful insights across several dimensions.

### 3.1 Sales Performance by Region

The cube allows the CEO to analyze Total Sales by City, Country, and Region. This is essential for evaluating how well different regions are performing and identifying high-growth areas. By combining the City and Sales dimensions, the company can pinpoint regions that are generating the most revenue.

Example Query:

- Total Sales by Region over a specific time period (using the Date dimension to view yearly, quarterly, and monthly sales trends).
- Profitability by Region: Analyze which regions generate the most profit relative to sales.

### 3.2 Customer Behavior by Region

Using the Customer and City dimensions, the cube provides insights into customer purchasing patterns across different geographical areas. By analyzing Quantity Sold and Total Sales by customer category and postal code, the company can identify key customer segments in specific regions.

Example Query:

- Top Customers by Region: Identify high-value customers in each region.
- Customer Buying Patterns: Analyze customer behavior by region and customer category, helping tailor marketing strategies to each region's demographics.

### 3.3 Product Performance Analysis

The Stock Item dimension allows the company to track product sales across different regions. By analyzing Quantity Sold and Total Sales by Stock Item and City, the company can determine which products perform well in specific geographical areas. This analysis helps in making data-driven decisions about which products to promote or discontinue in certain regions.

Example Query:

- Product Demand by Region: Identify which products have high demand in specific regions.
- Top-Selling Products: Determine which products are driving the most revenue in each region.

### 3.4 Time-Based Analysis

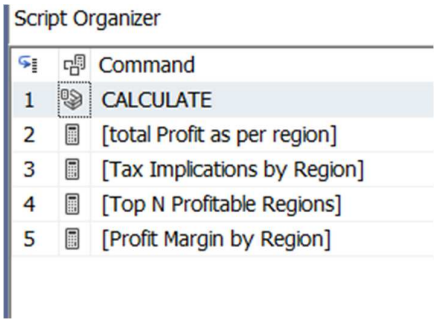
Using the **Date** dimension, the cube enables time-based analysis to track sales trends over time. The CEO can evaluate **year-over-year** growth, assess **seasonal trends**, and monitor the impact of geographical expansion on sales.

**Example Query**:

- **Yearly Sales Growth**: Compare sales performance across different years and regions.
- **Seasonal Product Trends**: Analyze which products are most popular during certain times of the year.

### 4. Calculation for region wise sales and profit:

- Total Profit as per Region: This measure calculates the total profit for each region by subtracting the total costs from sales. It provides insights into which regions are driving profitability, helping the company focus its expansion efforts on areas with the greatest potential for financial growth.
- Tax Implications by Region: This measure calculates the tax burden for each region, based on local tax rates and sales volumes. It allows the company to account for the impact of regional taxation on overall profitability, ensuring that tax considerations are factored into strategic expansion decisions.




- Top N Profitable Regions: This measure identifies the top N most profitable regions based on total profit. It helps the company prioritize its expansion by focusing on regions that are

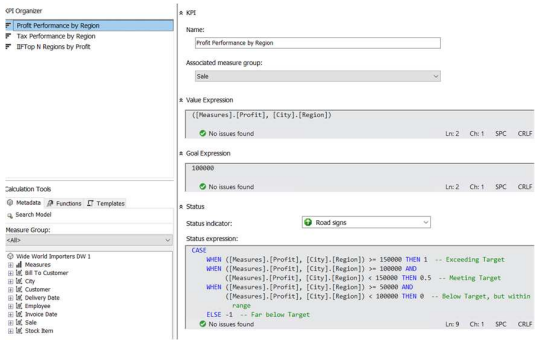
already generating strong financial returns, ensuring that resources are directed to the highest-performing areas.

- Profit Margin by Region: This measure calculates the profit margin for each region by dividing total profit by total sales. It highlights the operational efficiency of each region, helping the company assess whether it should focus on increasing sales or improving cost management in specific areas.

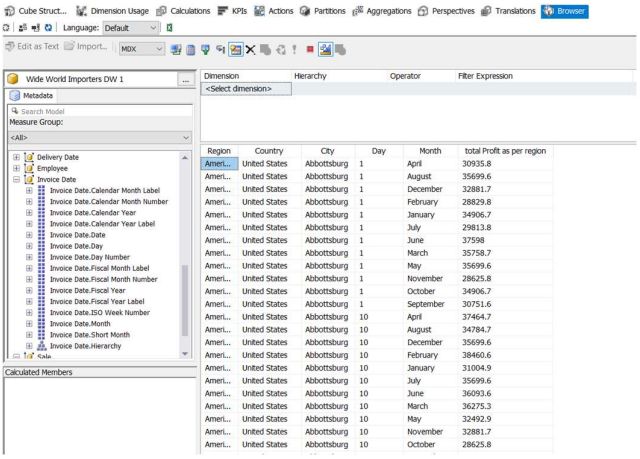
## 5. Reporting and Key Performance Indicators (KPIs)

Once the cube was deployed, KPIs were defined to help the company track the success of its geographical expansion efforts. Some of the key KPIs include:

- Profit Performance by Region: This measure calculates the overall profit performance for each region by analyzing total profit generated. It helps identify regions that contribute significantly to the company's bottom line and provides key insights into which areas should be prioritized for growth and investment.
- Tax Performance by Region: This measure evaluates the tax performance for each region, accounting for regional tax rates and sales data. It helps assess how much tax impacts profitability across different regions, ensuring that tax obligations are considered when planning expansion or investment strategies.

1. 


- Top N Regions by Profit: This measure ranks and displays the top N regions by profit. It allows the company to focus on regions that generate the most revenue, ensuring that expansion efforts are targeted at areas that are already financially successful.



Regional Profitability Analysis: The report shows the total profit by region, allowing the CEO to assess how profitable specific cities, like Abbottsburg, are. This is crucial for deciding whether further investment or expansion into this region is warranted.

Time-Based Insights: The data is broken down by month and day, providing an understanding of seasonal trends and monthly performance. This time-based analysis helps the CEO understand which periods yield higher profits, guiding decisions on resource allocation and market timing for expansion.

Decision-Making for Geographical Expansion: By evaluating the profit performance in each region, the CEO can identify high-performing regions that are ripe for expansion, as well as underperforming areas that might require strategic adjustments. This report offers the data necessary for making data-driven decisions about which regions to target in the expansion plan.

In the context of solving the CEO's problem—focused on geographical expansion and sales performance—MDX provides a robust, flexible way to extract, calculate, and manipulate data from a cube. Here's why MDX was an ideal choice for your solution:

- Handling Multidimensional Data:

    1. The company's data is organized in a cube with multiple dimensions such as City, Region, Time (Year/Month), Product Category, and Sales Metrics. MDX is specifically designed for working with these multidimensional structures, allowing you to query data across different dimensions seamlessly.
    2. For example, MDX allows you to slice and dice data by region, year, and product category in one query. This is essential for analyzing the company's performance in specific regions and identifying high-potential areas for geographical expansion.

- Aggregating and Calculating Measures:

    1. In the cube, you've created calculated measures like Total Profit as per Region, Tax Implications by Region, and Profit Margin by Region. MDX is used to calculate and aggregate these measures dynamically across various dimensions.
    2. For example, you can use MDX to aggregate Total Sales by Year and Category, helping the CEO assess which regions and products are driving the most revenue. This type of analysis would be difficult to achieve efficiently without MDX.

- Interactive Data Exploration:

    1. MDX allows users to drill down into summary data to explore detailed information. By querying hierarchical data, such as Year > Quarter > Month or Region > City, MDX provides flexibility to view both high-level summaries and granular details.

    2. This helps the CEO move from a broad view of geographical performance (e.g., overall sales by region) to more specific insights (e.g., monthly sales trends in a particular city).

- Optimized for Performance:
    1. MDX is highly optimized for OLAP cubes, making it faster and more efficient than SQL for querying large, complex datasets in multidimensional structures. When dealing with massive datasets across multiple dimensions, performance becomes a

key consideration. MDX ensures that your queries run efficiently, even when working with large volumes of data.

## POWER PIVOT:

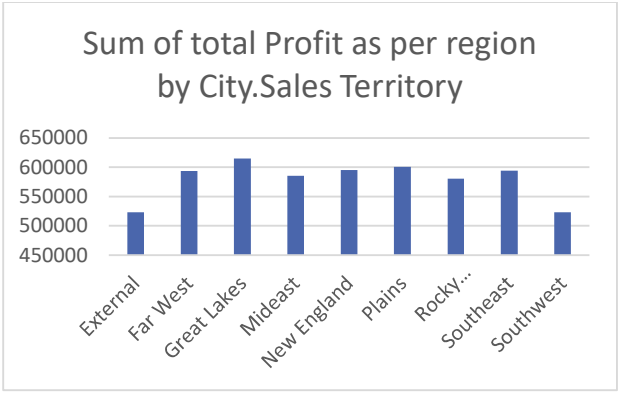
This is how the data looks like when you import from the server you have calculated in your cubes:

The graph provides valuable insights that directly address the CEO's goal of making data-driven decisions to support the company's geographical expansion. Here's how it can help:

| City.City | City.Region | Profit Margin by Region | City.Sales Territory | City.State Province | Invoice Date.Month | Tax Implications by Region | total Profit a |
|---|---|---|---|---|---|---|---|
| Abbottsburg | Americas | 0.495967323 | Southeast | North Carolina | April | 295882.75 | |
| Abbottsburg | Americas | 0.498298382 | Southeast | North Carolina | August | 312461.67 | |
| Abbottsburg | Americas | 0.497273211 | Southeast | North Carolina | December | 305533.43 | |
| Abbottsburg | Americas | 0.498741676 | Southeast | North Carolina | February | 296190.29 | |
| Abbottsburg | Americas | 0.497119628 | Southeast | North Carolina | January | 318952.99 | |
| Abbottsburg | Americas | 0.496990967 | Southeast | North Carolina | July | 307853.8 | |
| Abbottsburg | Americas | 0.499316052 | Southeast | North Carolina | June | 305426.94 | |
| Abbottsburg | Americas | 0.496278477 | Southeast | North Carolina | March | 305156.69 | |
| Abbottsburg | Americas | 0.495862771 | Southeast | North Carolina | May | 314339.48 | |
| Abbottsburg | Americas | 0.496650282 | Southeast | North Carolina | November | 303615.3 | |
| Abbottsburg | Americas | 0.498560184 | Southeast | North Carolina | October | 305621.72 | |
| Abbottsburg | Americas | 0.497572972 | Southeast | North Carolina | September | 298293.05 | |
| Absecon | Americas | 0.4471371 | Mideast | New Jersey | April | 202991.43 | |
| Absecon | Americas | 0.445020784 | Mideast | New Jersey | August | 208335.98 | |
| Absecon | Americas | 0.446431414 | Mideast | New Jersey | December | 207292.76 | |
| Absecon | Americas | 0.444804673 | Mideast | New Jersey | February | 191870.82 | |
| Absecon | Americas | 0.44369987 | Mideast | New Jersey | January | 210966.85 | |
| Absecon | Americas | 0.448295103 | Mideast | New Jersey | July | 216661 | |
| Absecon | Americas | 0.447325071 | Mideast | New Jersey | June | 201827.61 | |
| Absecon | Americas | 0.449780432 | Mideast | New Jersey | March | 207924.08 | |
| Absecon | Americas | 0.448957052 | Mideast | New Jersey | May | 211300.75 | |
| Absecon | Americas | 0.447313855 | Mideast | New Jersey | November | 206592.18 | |
| Absecon | Americas | 0.450739989 | Mideast | New Jersey | October | 213602.77 | |
| Absecon | Americas | 0.447756495 | Mideast | New Jersey | September | 205016.49 | |

## Sales Performance:

- Great Lakes comes out on top with the highest profit, which means sales are doing really well there. This region probably has a strong customer base and demand, making it a key area for further investment or growth.
- Far West, Plains, and Southeast also show good profits, though not as high as Great Lakes. These regions are performing quite well and could be looked at for expansion to take advantage of their success.




## Sales Performance:

- Great Lakes comes out on top with the highest profit, which means sales are doing really well there. This region probably has a strong customer base and demand, making it a key area for further investment or growth.

- Far West, Plains, and Southeast also show good profits, though not as high as Great Lakes. These regions are performing quite well and could be looked at for expansion to take advantage of their success.
- External and Southwest, though, have the lowest profits. Sales in these regions seem to be struggling compared to others. The company might need to focus more attention here, maybe with more marketing or operational improvements.

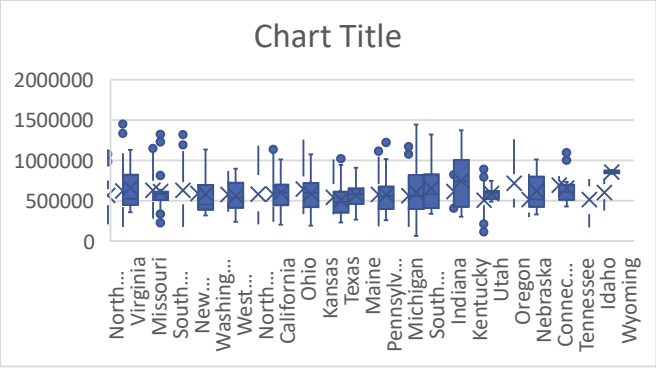
### Customer Behavior:

- In regions like Great Lakes and Far West, it seems like there's a high concentration of loyal or big-spending customers. These places keep delivering strong profits, showing that customers here might have stronger brand loyalty or more spending power.
- New England, Mideast, and Rocky Mountain regions are more in the middle. They seem to have steady but not super high customer demand. These might be places where customers are buying regularly but with room for growth if the company puts in a little more effort.
- In the Southwest and External regions, it looks like customers are either less engaged or maybe more price-sensitive, leading to lower profits. The company may not have as strong a presence here or just needs to work on getting customers more interested.

### Sales Performance:

The box plot gives us a clear picture of how sales are performing across different states by showing how profits are distributed. Here's what it says about sales performance:

- Top-Performing States: States like Wyoming, Idaho, and Nebraska have outliers with really high profits, meaning some cities or regions within these states are doing exceptionally well. These areas might have loyal customers or great sales strategies that's driving those high numbers.
- Mixed Performance: States like Texas, California, and North Carolina show a lot of variation in their profits. Some parts of these states are doing well, while others might be struggling. This tells us that sales performance can be pretty hit or miss depending on where you look, and might need some deeper investigation.




### Sales Performance:

The box plot gives us a clear picture of how sales are performing across different states by showing how profits are distributed. Here's what it says about sales performance:

- Top-Performing States: States like Wyoming, Idaho, and Nebraska have outliers with really high profits, meaning some cities or regions within these states are doing exceptionally well. These areas might have loyal customers or great sales strategies that's driving those high numbers.
- Mixed Performance: States like Texas, California, and North Carolina show a lot of variation in their profits. Some parts of these states are doing well, while others might be struggling. This tells us that sales performance can be pretty hit or miss depending on where you look, and might need some deeper investigation.
- Struggling Regions: States like Missouri and South Dakota show low profits with very little variation, which suggests weak sales performance across the board. These places could benefit from more marketing, better pricing strategies, or other efforts to improve sales.

## Customer Behavior:

The differences in profit distributions also give us a peek into customer behavior:

- Steady Customers: States like Virginia and North Dakota have small box plots with little variation, suggesting that customer buying habits are pretty consistent in these areas. Customers here seem to have steady purchasing behaviors without a lot of big changes.
- Diverse Customers: In states like California, Texas, and North Carolina, where we see more variation, it could mean that different regions or customer groups have very different buying patterns. Some areas might be seeing more sales than others, showing how customer preferences may vary widely within the same state.
- Big Spenders: The outliers in states like Wyoming, Idaho, and Nebraska could point to key customers or big purchases in certain spots. These regions might have a few high-value customers who are making a big difference in the profit numbers.

## Power BI:


Looking at the dashboard, here's a breakdown of what we're seeing:

## Profit Margin by Region and State Province

In the top left, we've got a scatter plot that shows the profit margin for each state. States like Alabama, Alaska, and California have varying profit margins, but the difference isn't too big. This gives us an idea of which regions are keeping their costs down while still generating solid profits.

### Profit and Tax Implications by City

Right next to that, there's a bar chart showing the profit and tax implications for different cities. It looks like cities like North Box, Tempopolis, and West Sylvania are pulling in strong profits while also dealing with higher tax implications. This helps the CEO see where tax is eating into profits and may require some adjustments.

### Total Profit by Region

In the top right, we see a gauge showing the total profit across all regions, sitting at 4.59 billion so far. The goal might be higher, and this gives a quick overview of how much more we need to go to hit the overall target.

### Profit by Category

Below that, there's a pie chart breaking down profit by category. It looks like the Novelty Shop category is responsible for the majority of profits, with about 61.18%, while the N/A category brings in 38.82%. This lets the CEO know where most of the money is coming from in terms of product categories.

### Tax Amount by Sales Territory and Salesperson

On the bottom, there's a line chart showing tax amounts by different sales territories. Southeast seems to have the highest tax amounts, which might indicate that the company is doing a lot of business there, but it's also getting taxed more heavily. We also see that the chart is filtered by whether the salesperson is true, so this data is only showing where salespeople are involved in these transactions.

### Sales Territory Breakdown

Finally, there's a table on the right breaking down sales by state, sales territory, and the total profit. This is a detailed view that shows us how each region is performing at a granular level, which is great for making decisions about where to invest and where to cut back.

### SSRS Report MDX:

### What the Report Shows:

City and Sales Territory:

- This section tells us which cities belong to which sales territories (like Southeast, Mideast, etc.). It's a good starting point for understanding how each region is performing geographically.

Profit Margin by Region:

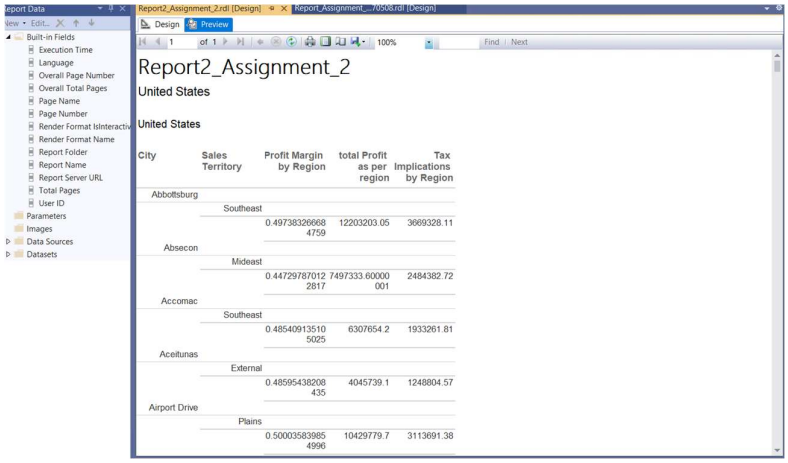
This column shows the profit margin for each city or region. Basically, it tells us how efficiently each region is generating profit. Regions with higher margins are likely more cost-effective, which means they're doing a great job at turning sales into profit without overspending

Total Profit by Region:

- This is a big one—the total profit generated in each region gives a direct insight into which areas are making the most money. For example, Abbottsburg is pulling in over 12 million in profit! Knowing which regions are top earners can help the CEO decide where to invest more resources or expand further.

Tax Implications by Region:

- The report also lists the tax costs for each region. This is super helpful for the CEO, as it shows where the company is paying the most in taxes. High-profit areas might also have high tax implications (like Abbottsburg), so it's something to consider when planning expansion—is it worth it to invest more in a high-tax region?




## How This Helps the CEO's Expansion Plans:

- Targeting High-Performing Regions: The CEO can use this report to figure out which regions are making the most money and are cost-efficient. For instance, areas like Abbottsburg and Airport Drive show strong profits and could be great candidates for further expansion.
- Managing Costs and Taxes: The report also highlights where tax costs are high. The CEO can use this information to be strategic—balancing the benefits of high profits with the costs of operating in a region with high taxes.
- Improving Efficiency: By looking at profit margins, the CEO can see which regions are operating efficiently and which ones might need a bit more help. If an area like Absecon has lower margins, they might look at reducing costs or boosting sales there before expanding.

## Task 3:

### Integration Strategy:

The first step in the integration process is to thoroughly analyze and understand both data warehouses—ValeurDW and Wide World Importers DW. ValeurDW leans more towards customer-focused data, with dimensions like Feedback and Customer Demographics, while Wide World Importers DW is more oriented around products, sales, and transactions. It's essential to identify any overlapping dimensions between the two, such as Customer, Product/Stock Item, and Date. Moreover, we need to ensure that both systems maintain compatible levels of data detail, like sales transactions and customer feedback, to facilitate a smooth integration.

Once we've gained a good understanding of both systems, the next task is to align and map the dimensions from each warehouse. For example, Valeur's Customer Dimension, which holds more detailed demographic information such as age and gender, needs to be mapped to Wide World Importers' Customer Dimension. At this point, it's crucial to clean and deduplicate records to ensure consistency. Similarly, Valeur's Product Category must be aligned with Wide World's Stock Item, and if necessary, we should create a unified product category mapping. The date dimensions in both systems must also be standardized to ensure consistent time-based reporting. One particularly unique aspect of ValeurDW is its Feedback Dimension, which should be integrated into the combined warehouse to enrich the analysis of customer behavior and satisfaction.

With the dimensions mapped, the next step is integrating the fact tables. This involves aligning ValeurDW's FactTransactionSummary with the Sale fact table in Wide World Importers. Merging sales data such as Total Sales and Quantity from both systems is key to consolidating the metrics. However, we also need to make sure that historical data lines up well, avoiding any duplication or misalignment of crucial metrics like profit and revenue. The ETL (Extract, Transform, Load) processes may need to be adjusted to handle the new data sources and dimensions from ValeurDW.

One important part of this integration process is managing historical data. Since ValeurDW likely contains years of historical sales records, we need to clean and standardize this data to match Wide World Importers' formats, ensuring consistency across dates, currencies, and categories. Additionally, we need to be mindful of any data overlaps that may exist between the two systems, ensuring that historical data is consolidated properly.

Ensuring data governance and maintaining quality is another critical component of the integration. Data from both warehouses must adhere to Wide World Importers' governance standards. This includes validating the data to confirm its accuracy, establishing security protocols to protect customer information, and removing duplicates and inconsistencies from the customer, product, and sales data.

Before deploying the integrated system, extensive testing is required. It's crucial to check data integrity to ensure that no data is lost or duplicated during the integration process. Validating both existing and new reports is also necessary, especially for key metrics like sales and customer feedback. Additionally, the ETL pipelines must be stress-tested to confirm that they can handle the increased data load and complexity.

Once the integration is complete, ongoing monitoring and maintenance will be necessary to ensure long-term success. Regular system checks for data quality and integrity will help identify any issues before they become significant problems. The ETL processes should be updated as needed, particularly when new data sources are added. Finally, conducting regular data audits will ensure that governance standards are maintained, keeping the system functioning at a high level of quality.

By following this detailed strategy, ValeurDW and Wide World Importers DW will integrate smoothly, creating a comprehensive data warehouse that supports deeper analysis and more informed decision-making, which will ultimately support the company's expansion goals.

### Dimensional Modeling and processing point of view:

The first step in integrating the Valeur Data Warehouse (DW) with Wide World Importers DW is to take a good hard look at both systems to see how they're set up. ValeurDW focuses more on customer-related stuff, like feedback and customer demographics, while Wide World Importers DW is all about products, sales, and transactions. We really need to figure out where they overlap—like with Customer, Product, or Date dimensions—and make sure the level of detail in each system matches up for smoother integration.

Once we've gotten a good understanding of both systems, we'll need to line up the dimensions properly. So, for example, Valeur's Customer Dimension might have more info, like age and gender, than Wide World Importers' version. We'll have to map those out, clean up duplicates, and make sure it all fits together. Same thing goes for products: Valeur's Product Category needs to be mapped to Wide World's Stock Item. And then, there's the Date Dimension; we'll need to standardize those to make sure reports show the right time-based data. The Feedback Dimension from Valeur can also be brought in to add some richness to customer analysis.

Now, moving onto the fact tables. We have to merge Valeur's FactTransactionSummary with Wide World Importers' Sale fact table. We'll be combining sales metrics like total sales and quantity from both systems. But, we've got to make sure the historical data from Valeur matches up and there's no double-counting. Plus, the ETL processes will probably need some tweaking to make sure they can handle Valeur's data.

Handling historical data is another big deal. ValeurDW has years of old sales data that will need cleaning up and standardizing—things like dates, currencies, and categories—so that it fits into Wide World Importers' format. Plus, we have to be super careful not to overlap or duplicate any data that's already in both systems.

On top of all that, maintaining data quality and sticking to data governance rules is key. We'll need to validate all the data to make sure it's accurate and consistent. Security protocols will also be important, especially when it comes to handling sensitive customer information. And, of course, we'll need to cleanse the data, removing duplicates and fixing inconsistencies.

Before we go live with the integrated system, we've got to test everything thoroughly. We'll check data integrity, making sure nothing's missing or duplicated. We'll also validate the reports, making sure both old and new ones give the right info. The ETL processes will be stress-tested to make sure they can handle the new data without breaking down.

Once the integration is done, it's important to keep an eye on things. Regular system checks and data audits will make sure everything stays in good shape. The ETL processes might need updating as we bring in new data sources, but we've got to make sure the system scales properly and keeps working as we grow.

By addressing all these key points, the ValeurDW and Wide World Importers DW integration will be smooth, creating a unified system that helps the business grow and expand, while supporting better decision-making and deeper customer insights.