Advanced Sales & Marketing Data Warehousing with Azure

Author : Rohan Chawla

Executive Summary

This report presents a data-driven analysis of customer purchasing behaviour, product sales trends, seasonal fluctuations, and the impact of discounting to refine sales and marketing strategies. The insights will help decision-makers optimize revenue, profitability, and customer engagement. The study investigates which customer segment (store vs individual) contributes most to sales and profit, identifies top-selling products across different regions, examines seasonal demand variations, evaluates customer purchase frequency, and measures the influence of discounting on overall profitability. These findings will support strategic pricing, promotions, and inventory management adjustments.

Historical sales, customer, and product data were extracted, transformed, and loaded (ETL) into a star-schema data warehouse to facilitate structured analysis. Data cleansing, integration, and transformation steps were applied to ensure consistency and accuracy. Through SQL queries and analytical computations, we identified purchasing patterns, regional sales trends, discount effectiveness, and seasonality effects. Key metrics such as total revenue, gross profit, purchase frequency, and seasonal revenue fluctuations were evaluated to provide actionable business insights.

Key Findings:

- Stores generate the highest revenue but operate at a loss, while individual customers contribute lower sales but significantly higher profits. This highlights the need to reassess bulk pricing strategies to ensure store-based sales remain profitable.

- Bikes are the most profitable product category despite having one of the lowest discount rates. This suggests that high-value products do not require aggressive discounting to drive sales, reinforcing the need for a more selective and strategic approach to promotions.

- Seasonality significantly impacts sales, with distinct revenue peaks in certain quarters. Products like Mountain Bikes and Road Bikes experience high seasonal fluctuations, emphasizing the need for timely marketing campaigns and inventory adjustments to maximize sales during peak seasons.

To act on these insights, the company should adjust discounting policies, customize marketing campaigns, optimize inventory planning, and revise bulk pricing models to ensure profitability. Implementing these strategies will enhance sales performance, boost profit margins, and improve customer engagement, driving long-term business growth in a competitive market.

2

Motivation

As a team of Business Analytics students, we selected Sales and Marketing Analytics due to its critical role in business decision making and revenue optimization. Businesses increasingly rely on data driven strategies to enhance customer engagement, optimize marketing efforts, and maximize profitability. This project enables us to apply our analytical skills to real world business challenges, providing valuable experience in data warehousing, ETL processes, and SQL based insights.

1. **Relevance to Industry**

Sales and marketing drive revenue, customer retention, and market positioning. In today's data-driven economy, businesses must analyse customer behaviours, sales trends, and product demand to stay competitive. This project helps refine sales strategies, price optimization , and promotional campaigns for sustainable growth.

2. **Strategic Decision Making**

Data-driven insights are key to customer segmentation, regional sales trends, and discount impact. This project helps businesses identify high-value customers, optimize product distribution, and refine pricing strategies, leading to enhanced marketing, higher conversions, and maximized profitability.

3. **Data Driven Growth**

Analysing historical sales data uncovers seasonal trends, demand patterns, and customer preferences. These insights enable accurate sales forecasting, pricing optimization, and promotion assessment, ensuring proactive decision making and long-term competitive advantage.

4. **Real World Application**

This project provides hands-on experience in data warehousing, ETL, and SQL analytics, essential for retail, e-commerce, manufacturing, and finance. Businesses rely on structured data to drive sales optimization, marketing strategies, and operational efficiency, making this project highly applicable to real-world business environments.

By working on this project, we develop practical expertise in data analytics and decision making, gaining insights that will help businesses optimize sales performance, enhance customer engagement, and drive strategic growth.

3

Insightful Questions

Our analysis is driven by six key business questions, each designed to uncover actionable insights into customer behaviours, sales trends, pricing strategies, and profitability. These questions align with business objectives, real-world scenarios, and a clear sense of purpose, ensuring strategic decision-making and fostering sustainable growth. By addressing these questions, we aim to gain a deeper understanding of how different factors influence business outcomes and identify opportunities for optimization. The questions are as follows:

1. **Customer Segment Contribution to Sales and Profit:**
   Which customer segment (individual vs. store) contributes the most to total sales revenue and profit?

   Understanding the revenue and profit contribution of different customer segments enables businesses to optimize pricing, customer engagement, and retention strategies. While stores may place larger orders, individual customers might yield higher profit margins. Identifying the most valuable segment helps refine sales and marketing efforts for maximum impact.

2. **Regional Sales Trends and Product Preferences:**
   What are the top selling products in different regions, and how do regional customer preferences influence sales trends?

   Consumer demand varies by geography, culture, and economic factors. Analysing regional sales trends allows businesses to optimize inventory, tailor marketing efforts, and expand product offerings in high-demand areas. This enhances stock management efficiency and customer satisfaction.

3. **Impact of Seasonality on Product Sales:**
   How does seasonality impact product sales, and which products experience the highest seasonal fluctuations?

   Seasonal trends affect demand, pricing, and promotional strategies. Identifying peak and off-peak sales periods helps businesses forecast demand, plan marketing campaigns, and manage inventory efficiently. Understanding which products fluctuate seasonally ensures better stock control and revenue maximization.

4

4. **<u>Identifying High-Value Customers:</u>**
How can we identify the most engaged and high-value customers based on their purchase frequency, total spend, and order value?

Customer engagement is a key driver of repeat sales and long-term business sustainability. By analysing purchase frequency and spending patterns, businesses can identify their most loyal and valuable customers. These insights allow for personalized marketing, loyalty programs, and customer retention strategies, maximizing lifetime value.

5. **<u>Effect of Discounts and Offers on Sales and Profitability:</u>**
How do discounts and special offers influence sales volume, revenue, and overall profitability?

Discounts and promotions can boost sales volume but may erode profit margins if not applied strategically. Understanding which types of discounts drive profitable sales helps businesses refine pricing strategies and promotional campaigns to increase revenue without sacrificing profitability.

6. **<u>Profitability of Product Categories and Discount Impact:</u>**
Which product categories yield the highest profitability, and how do different discount strategies impact their profit margins?

Not all high-revenue products generate high profit margins. Some may sell in large volumes but contribute minimal net profit, while others yield high margins despite lower sales volume. Analysing category profitability ensures that businesses prioritize the most valuable products and adjust discounting strategies accordingly.

Strategic Impact

These questions provide a data-driven foundation for optimizing sales, pricing, and marketing strategies. By addressing customer segmentation, regional demand, seasonality, and discount effectiveness, businesses can enhance profitability, improve inventory management, and drive sustainable growth. The insights gained will support long-term strategic planning, helping businesses stay competitive in an evolving market.

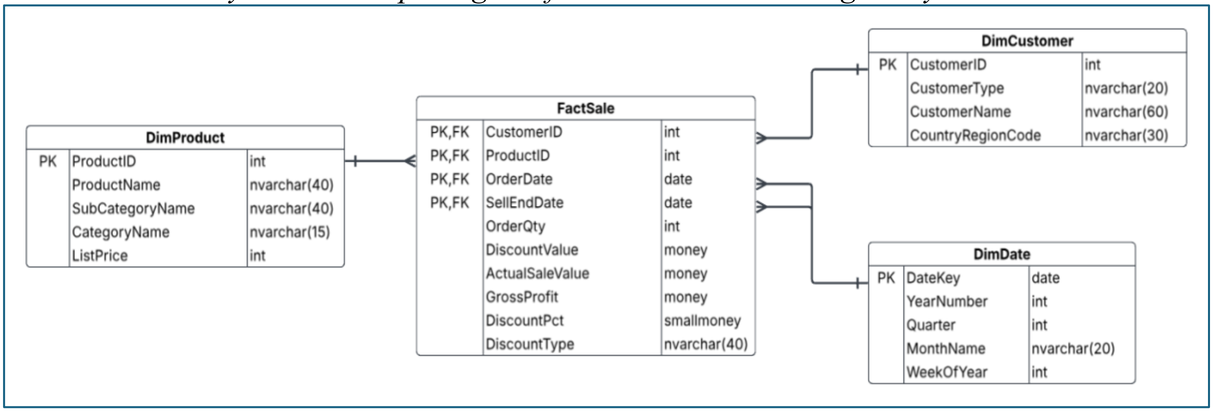
5

Entity-Relationship Diagram (ERD)

The Entity-Relationship Diagram (ERD) represents a star schema model designed to optimize sales and marketing analytics. At its core, the FactSale table captures key transactional data, including order quantity, discount values, actual sales revenue, and gross profit, providing a structured view of sales performance. This fact table is linked to three essential dimension tables: DimCustomer, DimProduct, and DimDate, through well-defined foreign key relationships, ensuring data integrity and consistency.

Figure 1

Star Schema Entity-Relationship Diagram for Sales and Marketing Analytics



*Note: The ERD represents a star schema, where FactSale connects with DimProduct, DimCustomer, and DimDate, facilitating efficient sales and customer analysis.

DimCustomer categorizes customers by type, name, and region, facilitating segmentation and market analysis. DimProduct provides hierarchical product details, enabling businesses to analyse product sales trends and profitability across categories. DimDate allows for time-based trend analysis, supporting insights into seasonality, peak sales periods, and forecasting.

The star schema structure improves query performance, simplifies data retrieval, and supports advanced business intelligence applications. This model empowers businesses to make data-driven strategic decisions, optimize marketing efforts, product distribution, and pricing strategies, and enhance customer engagement through targeted promotions and personalized sales approaches.

Azure Data Factory ETL (Extract Transform Load) Process

Customer Dimension Table

The transformation of the Customer Dimension table in Azure Data Factory follows a structured ETL process to integrate, clean, and refine customer data from multiple sources. The workflow begins with extracting data from external databases, specifically store-based and individual customer data, which are processed separately. Derived column transformations are applied to classify customers by type, ensuring consistency in the dataset. SQL-based source transformations refine the individual customer dataset by constructing full customer names and associating them with regional sales data. A union transformation merges store and individual customer data into a unified structure, followed by column mapping and renaming to standardize the schema. The final transformation stage loads the cleaned and structured data into the DimCustomer table, making it available for analytics and reporting.

The step-by-step process and data flow for transforming the Customer Dimension table is detailed below in Figures 2.1 to 2.7.

Figure 2.1

Azure Data Factory ETL for Customer Dimension

7

Figure 2.2

Derived Column Transformation for Customer Type in Azure Data

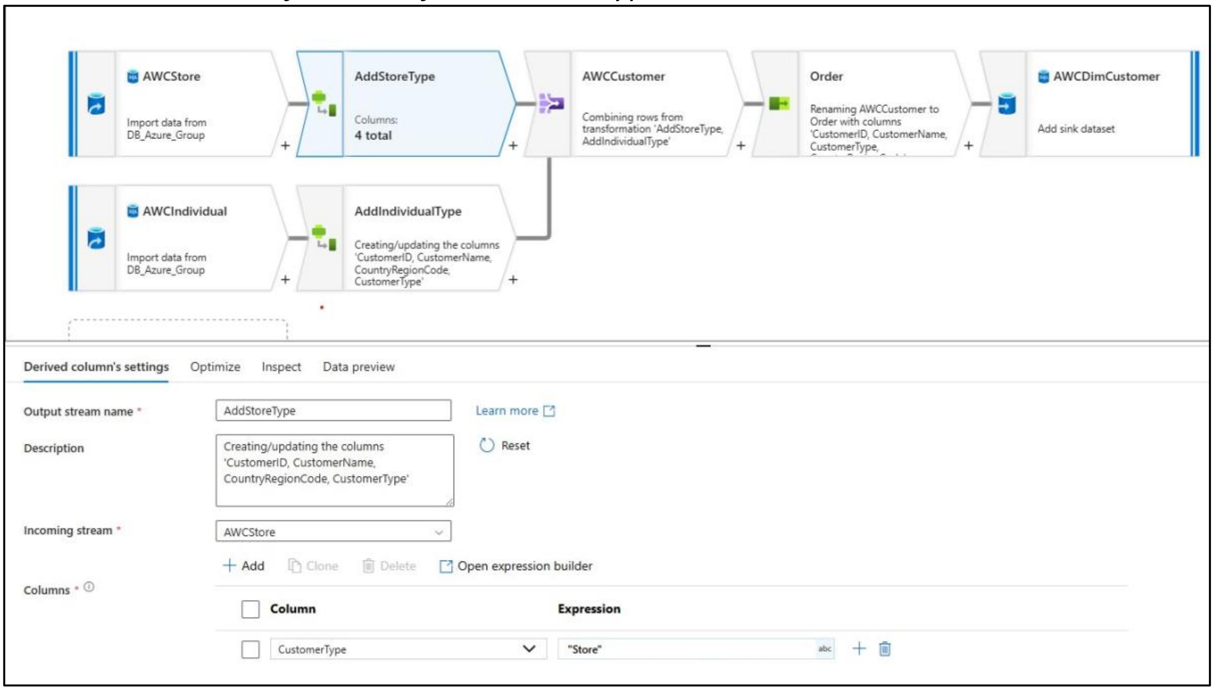


Figure 2.3

Source Query Transformation for Individual Customers in Azure Data

8

Figure 2.4

Derived Column Transformation for Individual Customers

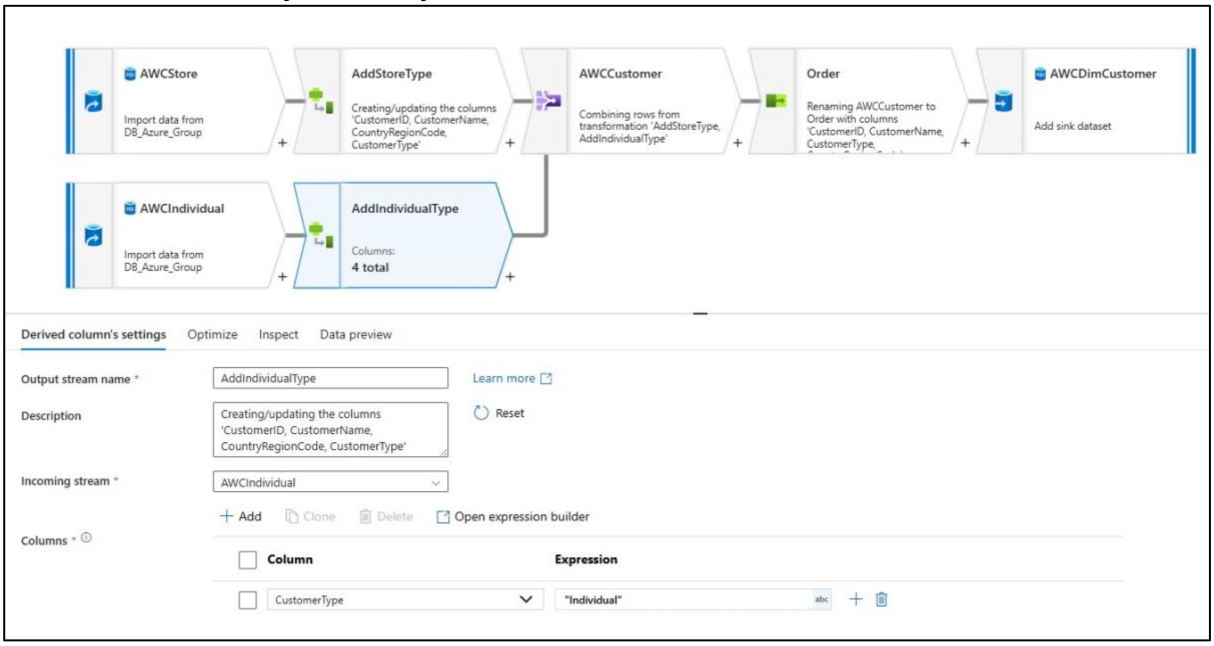


Figure 2.5

Union Transformation for Customer Data

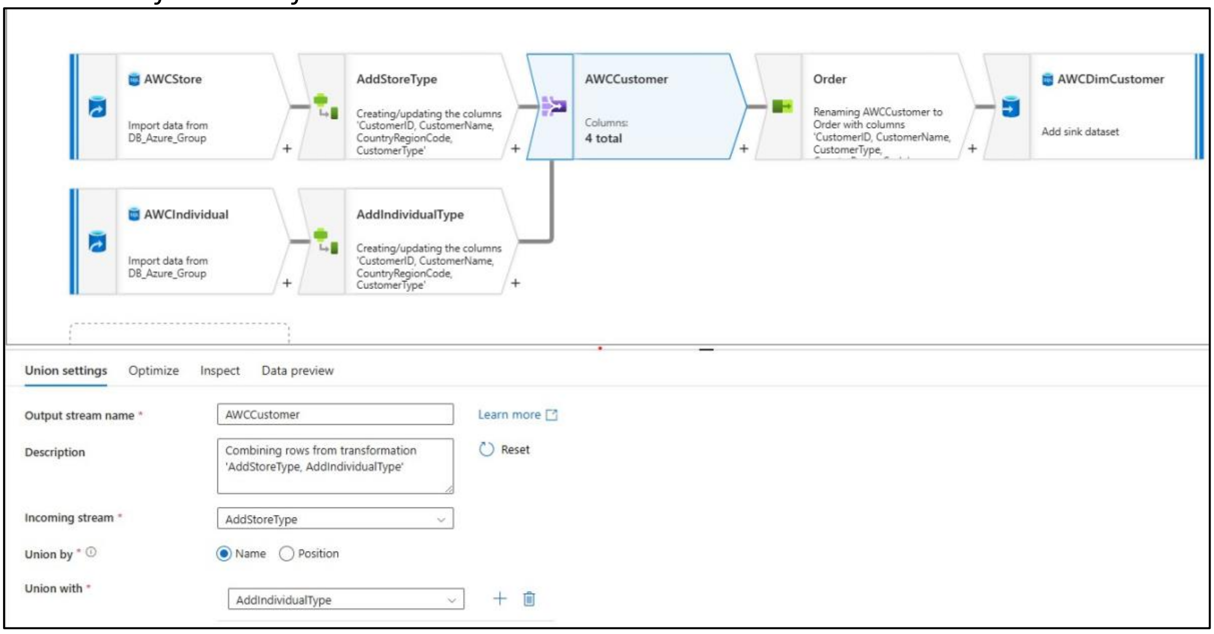


Figure 2.6

Column Mapping and Renaming for Customer Data

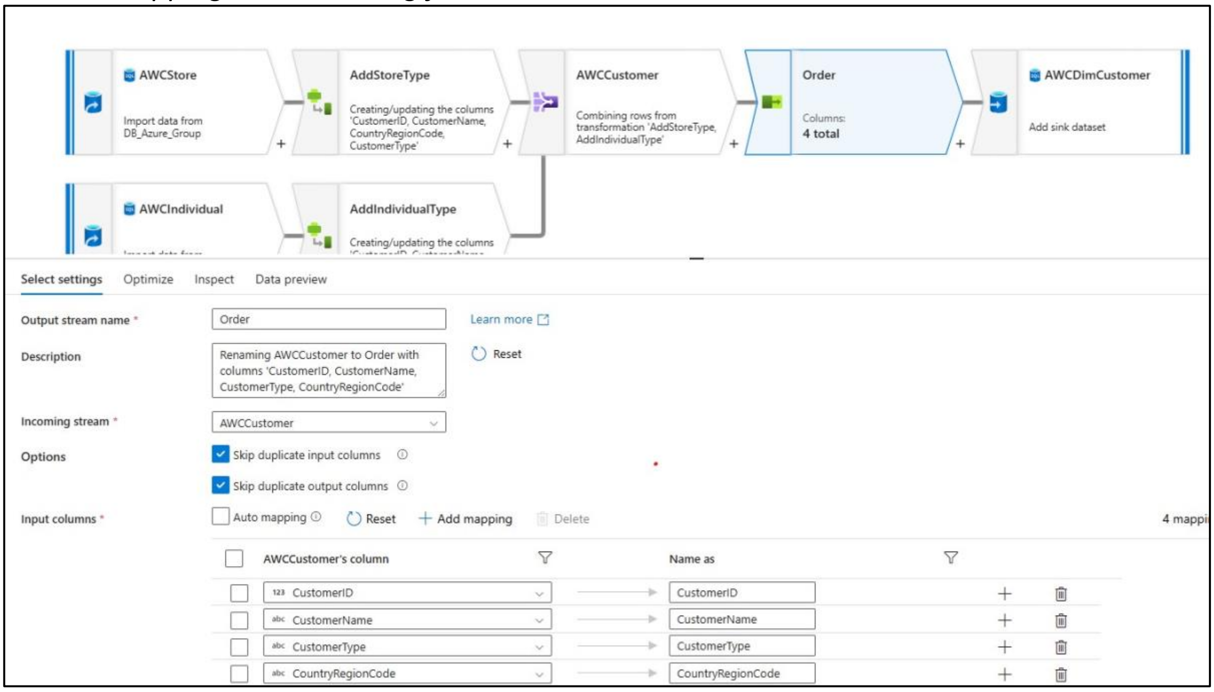
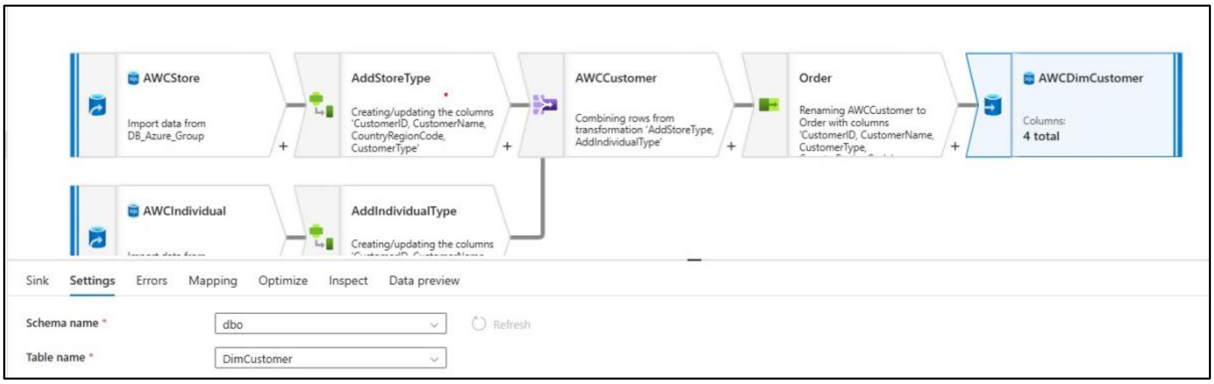


Figure 2.7

Final Data Load into DimCustomer Table



10

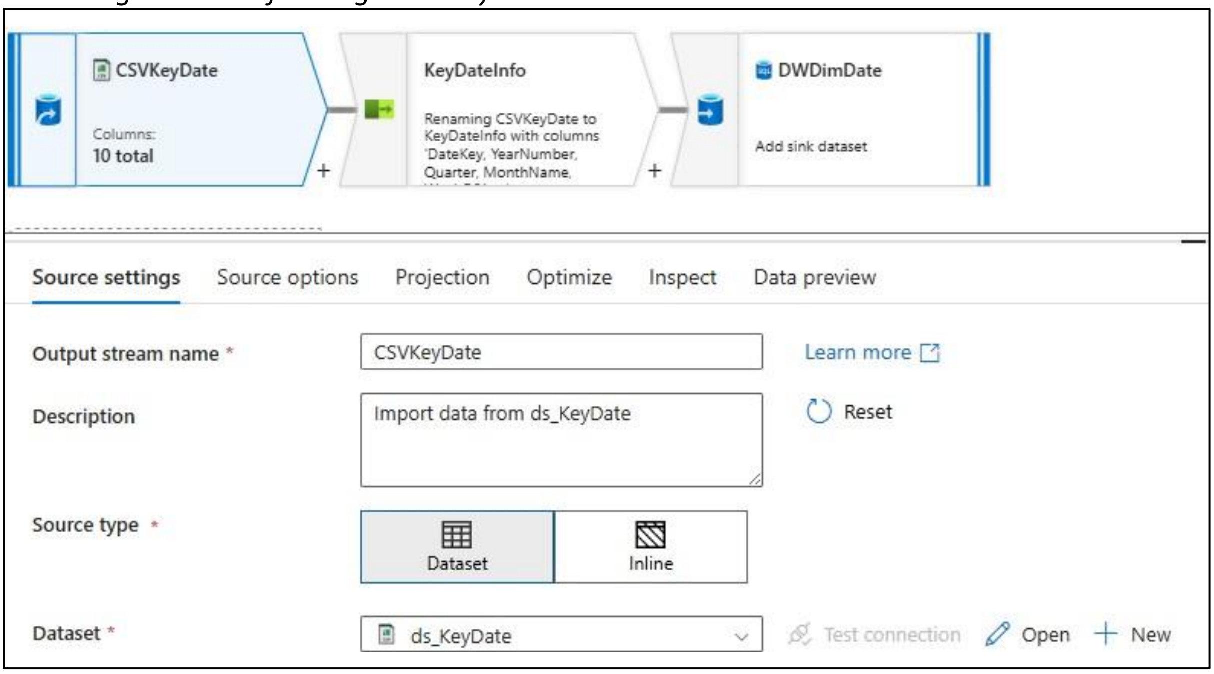
Date Dimension Table

The ETL process for the Date Dimension Table ensures the seamless extraction, transformation, and loading of date-related attributes into the data warehouse, facilitating accurate time-based analysis. The process begins with the extraction phase, where raw date data is imported from an external dataset (ds_KeyDate), containing fundamental time attributes necessary for chronological analysis. In the transformation phase, the extracted data is processed to derive essential time-based attributes such as DateKey, YearNumber, Quarter, MonthName, and WeekOfYear. These transformations standardize and optimize the data structure, enhancing its usability for analytical queries and time-series reporting. Finally, in the loading phase, the transformed data is stored in the DimDate table within the data warehouse, serving as a key reference point for linking time-based trends across various fact tables. This structured approach ensures efficient reporting and supports business intelligence applications.

The step-by-step dataflow for the Date Dimension Table is illustrated in Figures 3.1 to 3.3, detailing the extraction, transformation, and loading processes.

Figure 3.1

Extracting and Transforming Date Keys



11

Figure 3.2

Date Key Transformation and Mapping

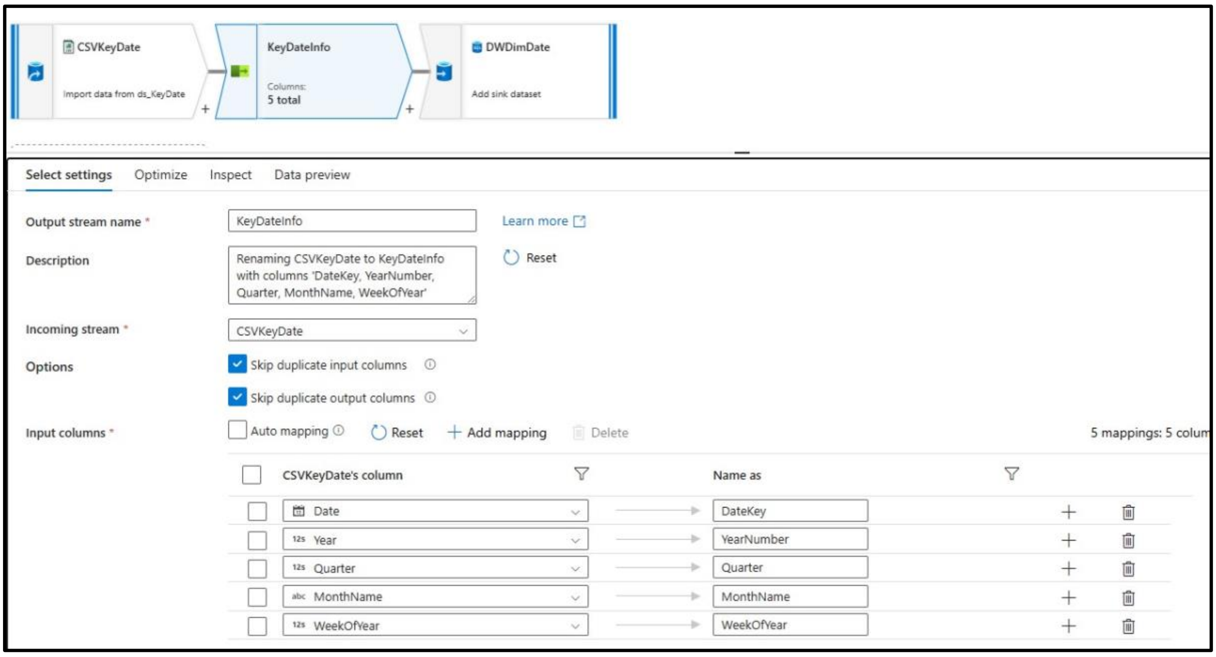
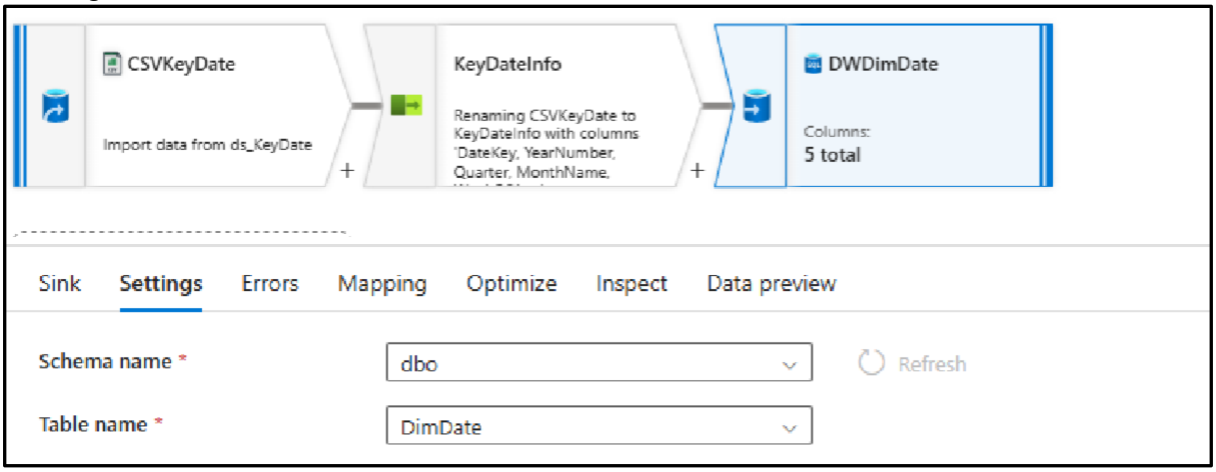


Figure 3.3

Loading Date Dimension Table



12

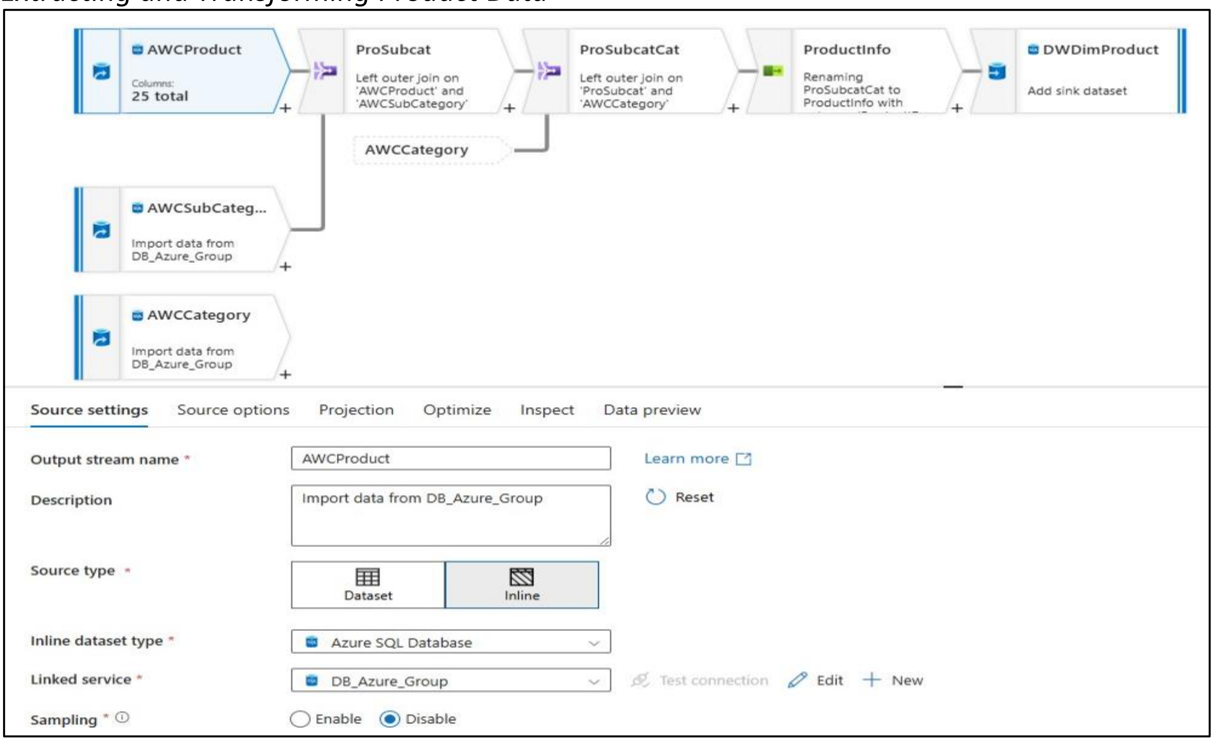
Product Dimension Table

The ETL process for the Product Dimension Table in Azure Data Factory follows a structured workflow to extract, transform, and load product-related data. Initially, Figure 4.1 represents the extraction of product data from the Azure SQL database, where the dataset AWCProduct is ingested. In Figure 4.2, subcategory data is extracted and integrated into the transformation process, ensuring hierarchical structuring. Figure 4.3 details the joining of product and subcategory data using a left outer join, linking ProductSubcategoryID between AWCProduct and AWCSubCategory.
Following this, Figure 4.4 highlights the extraction of category data from the Azure SQL database, which is further utilized in subsequent transformations. In Figure 4.5, the product-subcategory relationship is extended by performing another left outer join with AWCCategory, linking the ProductCategoryID. This ensures a fully structured dataset that integrates product, subcategory, and category information.
The transformation phase is consolidated in Figure 4.6, where the dataset is renamed and mapped with key attributes such as ProductID, ProductName, SubCategoryName, CategoryName, and ListPrice. This step ensures proper column alignment and deduplication. Finally, Figure 4.7 represents the loading of the transformed product data into the DimProduct table within the Azure SQL database, completing the ETL process.

The step-by-step data flow is illustrated in Figures 4.1 to 4.7, ensuring a comprehensive transformation pipeline for the Product Dimension Table in the data warehouse.

Figure 4.1

Extracting and Transforming Product Data



13

Figure 4.2

Extracting and Integrating Subcategory Data

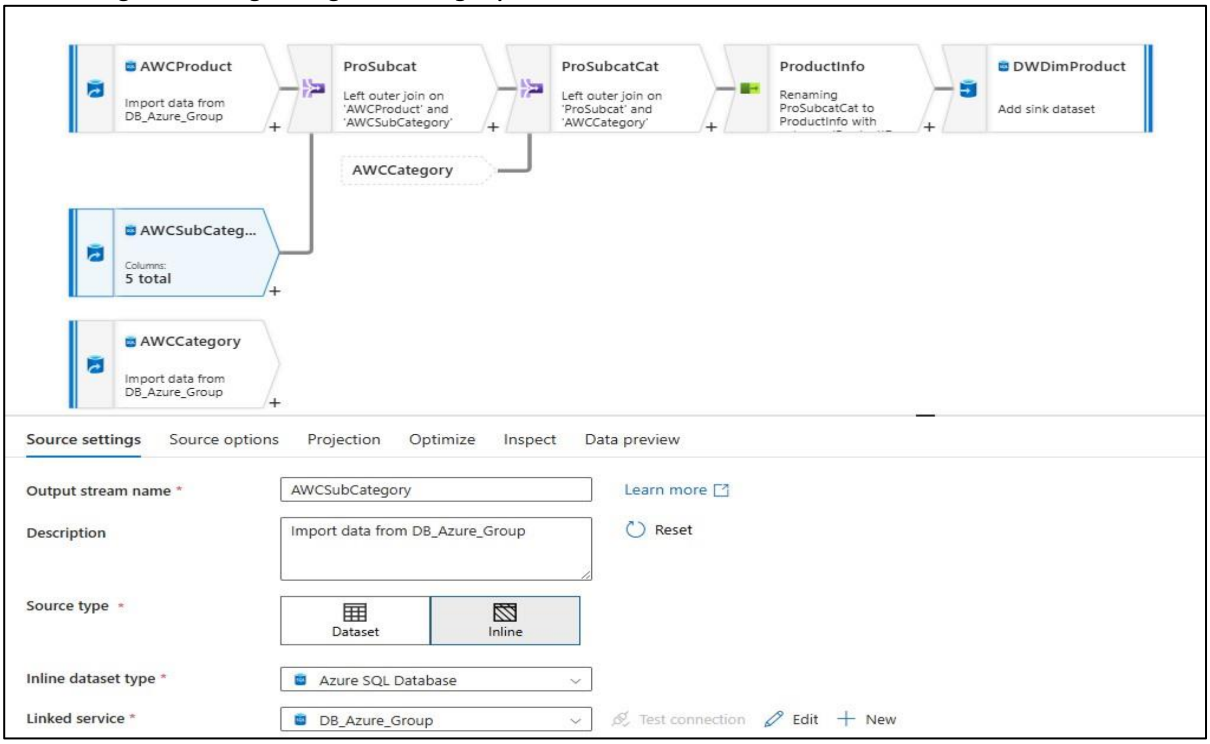
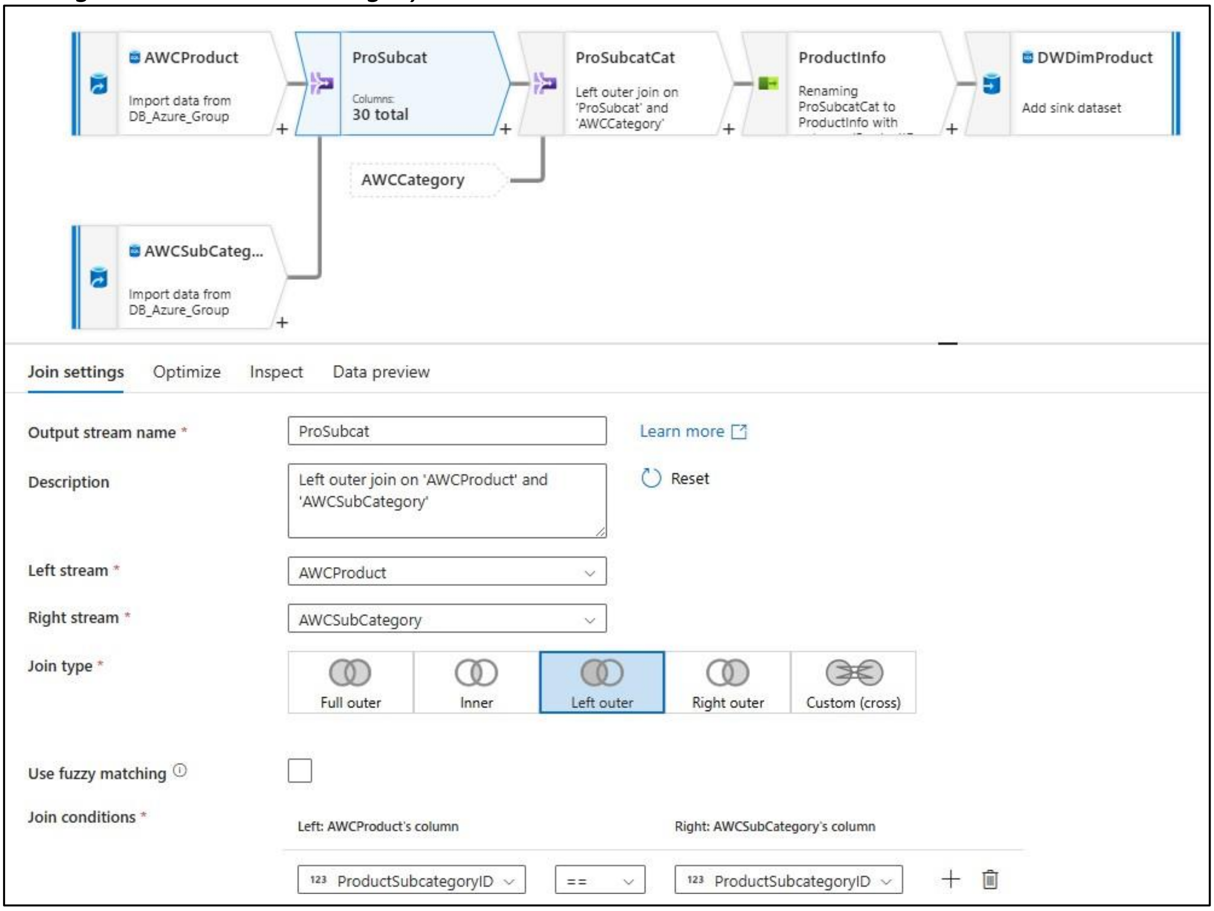


Figure 4.3

Joining Product and Subcategory Data



14

Figure 4.4

Extracting Category Data from Azure SQL Database

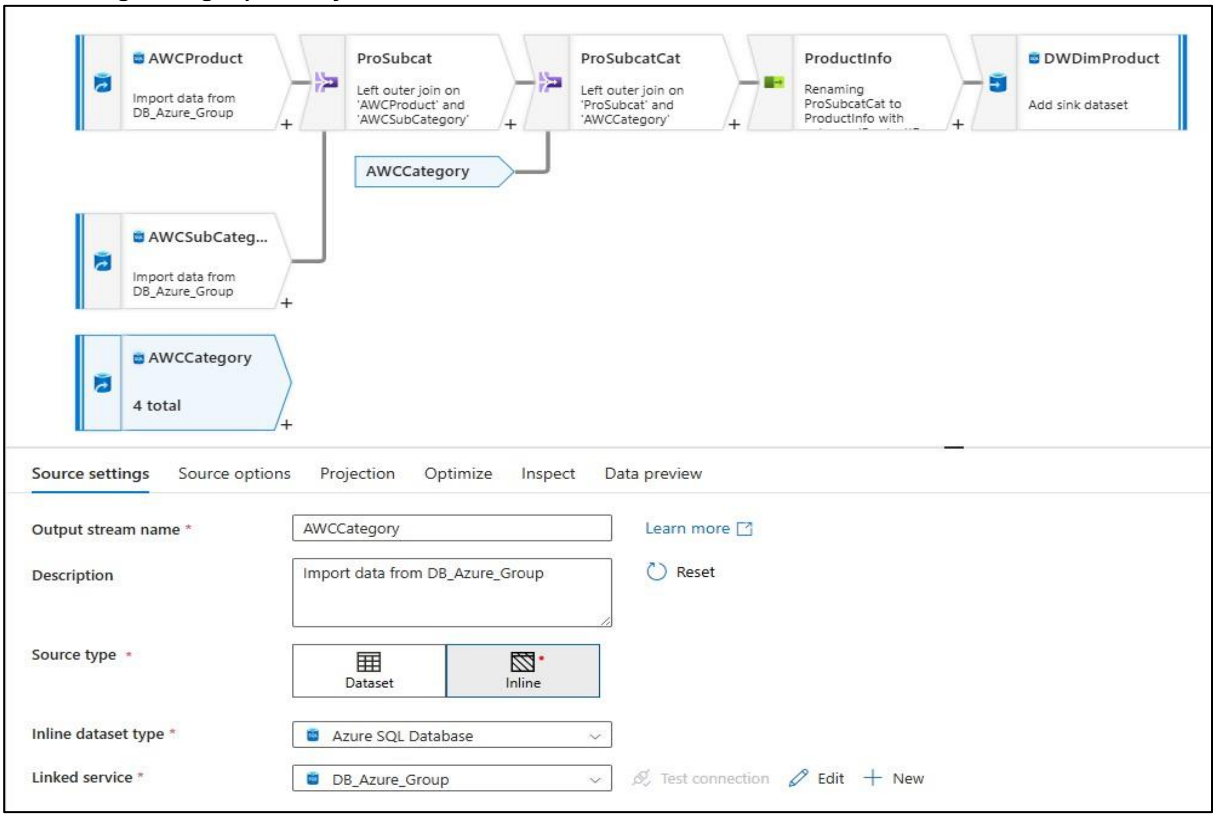
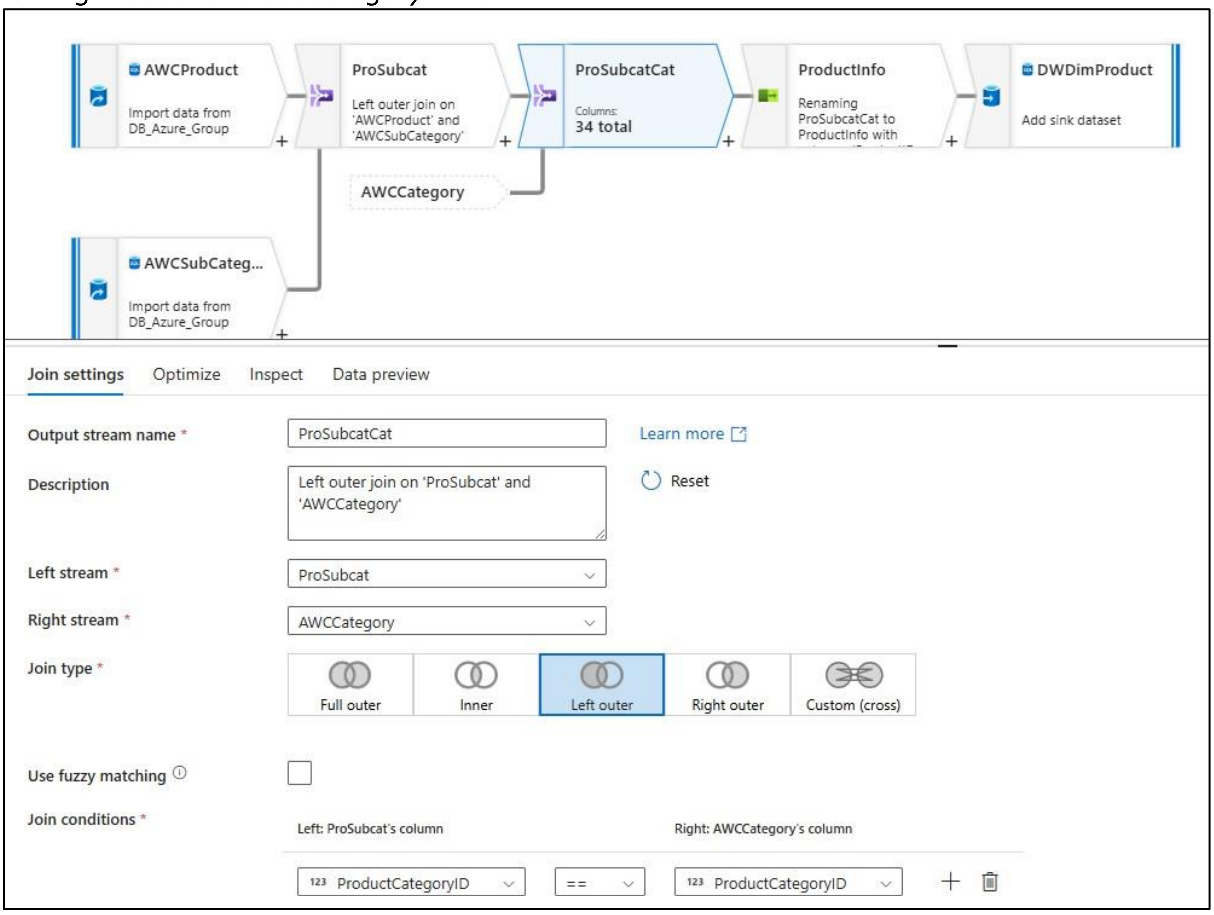


Figure 4.5

Joining Product and Subcategory Data



15

Figure 4.6

Transforming Product Data for DimProduct Table

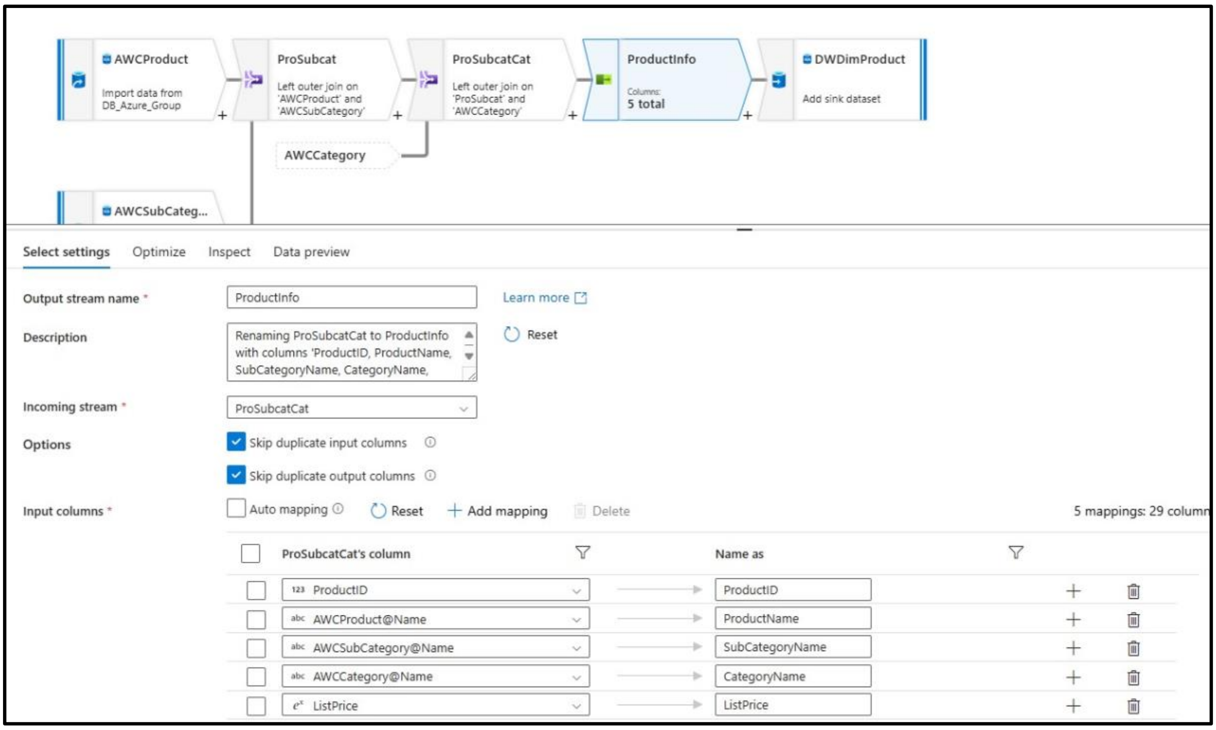
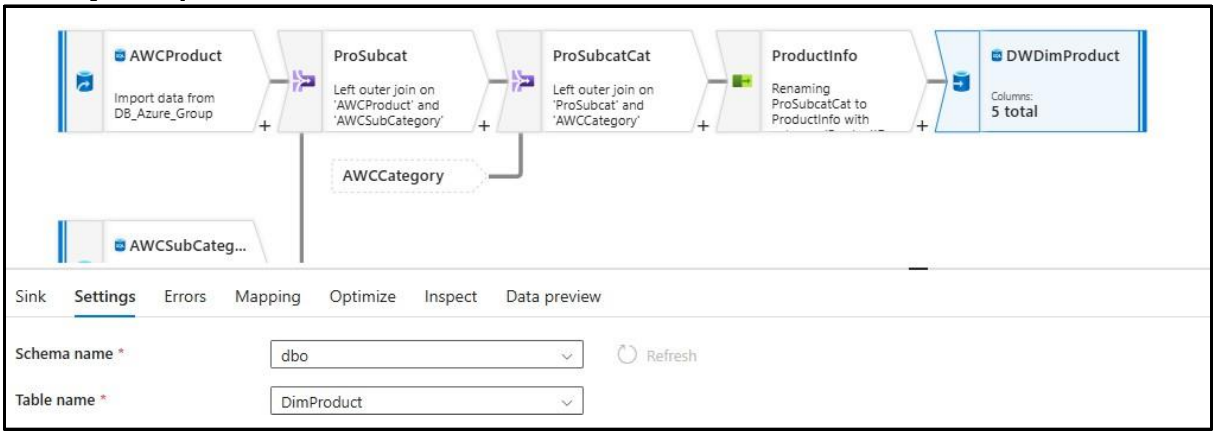


Figure 4.7

Loading Transformed Product Data into DimProduct Table



16

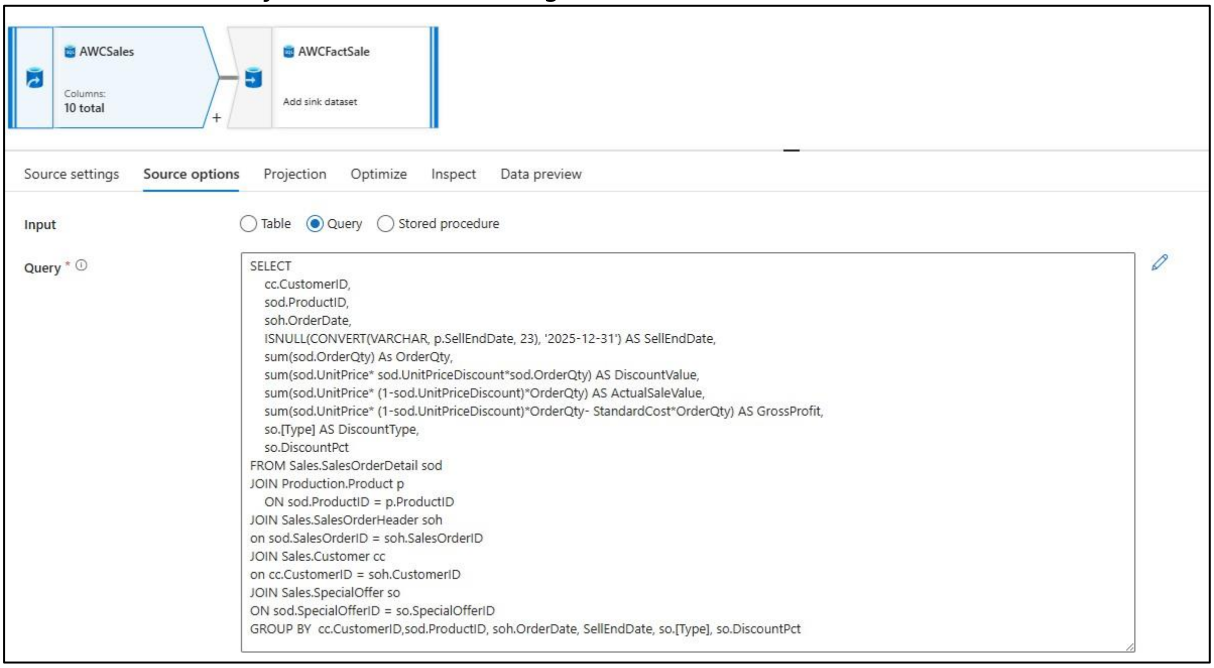
Centralised Fact Table

The centralized fact table transformation and loading process involves structuring sales transaction data to facilitate analytical insights. In Figure 5.1, the Fact Sale Data Transformation and Loading step involves extracting data from the AWCSales source table and applying SQL transformations. The transformation aggregates sales order details, computing total order quantities, discount values, actual sales values, and gross profit. This transformation also ensures data completeness by incorporating necessary joins with related tables, such as Product, SalesOrderHeader, Customer, and SpecialOffer, to enrich the fact table with essential attributes.

Subsequently, in Figure 5.2, the Fact Sale Data Loading Process step maps the transformed data to the AWCFactSale table in the dbo schema. This ensures the structured integration of sales transactions into the data warehouse for downstream analysis. The complete data flow from extraction to transformation and final loading into the FactSale table is detailed in Figures 5.1 and 5.2.

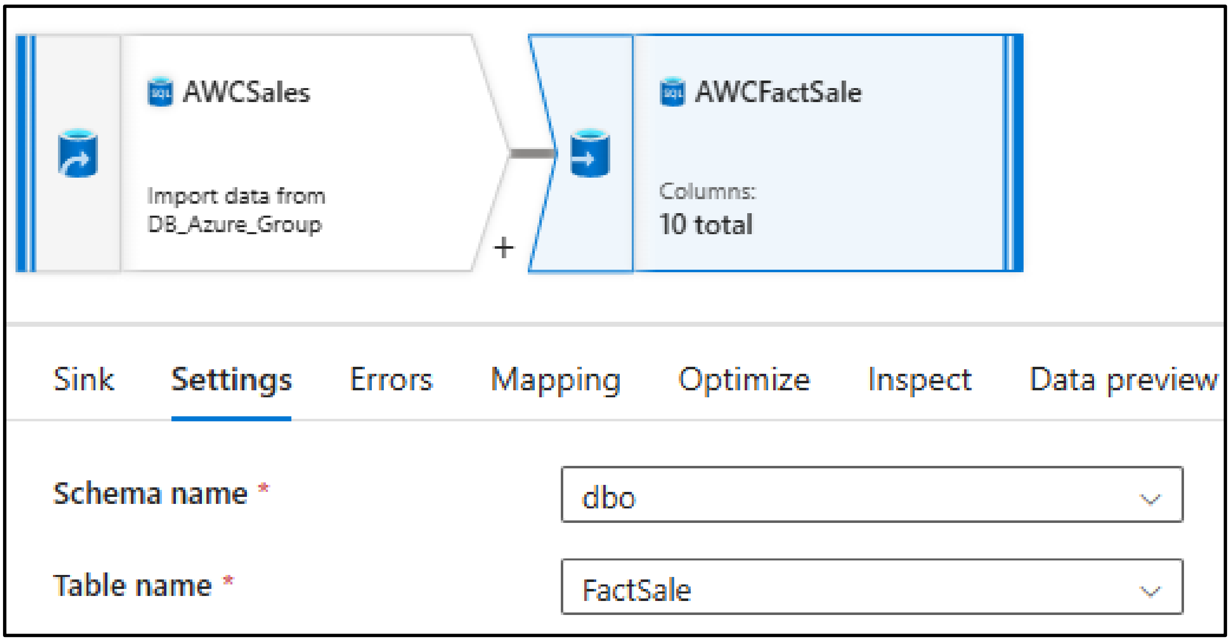
Figure 5.1

Fact Sale Data Transformation and Loading



17

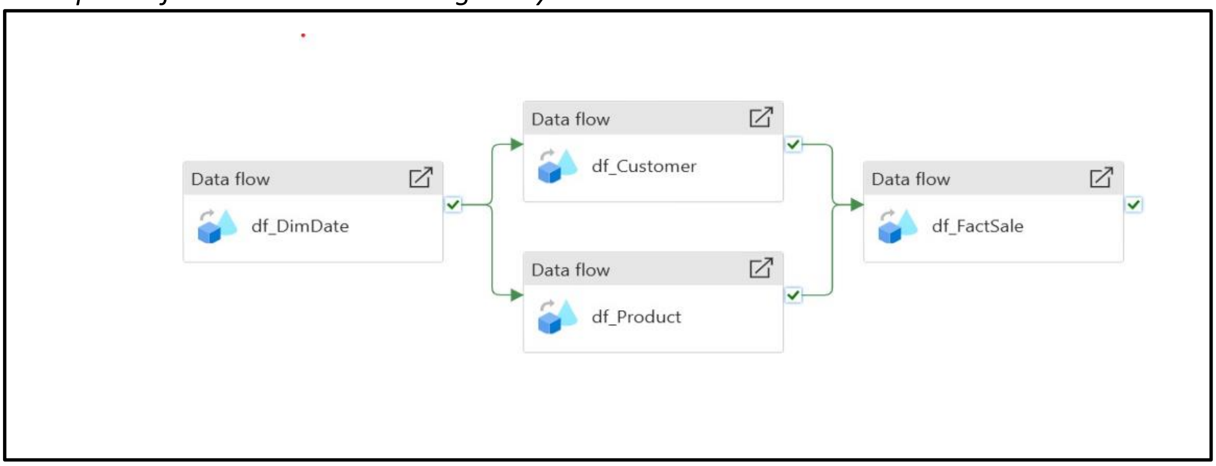
Figure 5.2
Fact Sale Data Loading Process



ETL Pipeline Ingestion

The ETL Pipeline for Sales and Marketing Analytics in Figure 6 represents the structured data integration process designed to enable business insights. This pipeline consists of multiple interconnected data flows, beginning with df_DimDate, which processes and standardizes date-related attributes for temporal analysis. The df_Customer and df_Product flows are responsible for extracting, transforming, and enriching customer and product data, ensuring they contain relevant attributes for downstream analytics. These dimensional data flows are then linked to df_FactSale, which consolidates transactional sales data, including order details, discount values, and gross profit calculations. The pipeline ensures data consistency, transformation accuracy, and seamless integration, facilitating advanced sales and marketing analysis.

Figure 6
ETL Pipeline for Sales and Marketing Analytics



SQL Statement

1. Analysing Sales Revenue and Profit Contribution by Customer Segment: Individual vs Store

```
SELECT

    c.CustomerType,
    SUM(f.ActualSaleValue) AS TotalSalesRevenue,
    SUM(f.GrossProfit) AS TotalProfit
FROM FactSale f
JOIN DimCustomer c ON f.CustomerID = c.CustomerID
GROUP BY c.CustomerType
ORDER BY TotalSalesRevenue DESC;
```

Figure 2

Customer Segment Sales Revenue and Profit Comparison

	CustomerType	TotalSalesRevenue	TotalProfit
1	Store	80487704.3222	-2316039.1101
2	Individual	29358677.2207	11687942.8795

The analysis highlights a clear contrast between revenue generation and profitability across customer segments. Stores contribute the highest total sales revenue ($80.48M), but they operate at a loss with a total profit of -$2.31M, whereas individual customers generate lower revenue ($29.35M) but significantly higher profit ($11.68M). This indicates that while store-based sales drive volume, they may be impacted by high operational costs, aggressive discounting, or lower profit margins per transaction.

In contrast, individual customers, despite generating less revenue, yield a positive profit, suggesting better margins per sale. This insight underscores the need to evaluate the pricing, discounting, and cost structures associated with store-based sales while recognizing the profitability potential of individual customers.

2. Analysing Top-Selling Products Across Regions and the Impact of Regional Customer Preferences on Sales Trends

```
WITH ProductSales AS (
    SELECT
        c.CountryRegionCode,
        p.ProductName,
        p.SubCategoryName,
        SUM(f.OrderQty) AS TotalQuantitySold,
```

16

```
        SUM(f.ActualSaleValue) AS TotalRevenue

    FROM FactSale f
    JOIN DimProduct p ON f.ProductID = p.ProductID
    JOIN DimCustomer c ON f.CustomerID = c.CustomerID
    GROUP BY c.CountryRegionCode, p.ProductName,
p.SubCategoryName
), RankedProducts AS (
    SELECT
        CountryRegionCode,
        ProductName,
        SubCategoryName,
        TotalQuantitySold,
        TotalRevenue,
        RANK() OVER (PARTITION BY CountryRegionCode ORDER BY
TotalRevenue DESC) AS Rank
    FROM ProductSales
)
SELECT CountryRegionCode, ProductName, SubCategoryName,
TotalQuantitySold, TotalRevenue
FROM RankedProducts
WHERE Rank = 1
ORDER BY TotalRevenue DESC;


WITH CountryQuarterRevenue AS (
    SELECT
        c.CountryRegionCode,
        d.Quarter,
        SUM(f.ActualSaleValue) AS TotalRevenue
    FROM FactSale f
    JOIN DimCustomer c ON f.CustomerID = c.CustomerID
    JOIN DimDate d ON f.OrderDate = d.DateKey
    GROUP BY c.CountryRegionCode, d.Quarter
), RankedRevenue AS (
    SELECT
        CountryRegionCode,
        Quarter,
        TotalRevenue,
        RANK() OVER (PARTITION BY CountryRegionCode ORDER BY
TotalRevenue DESC) AS Rank
    FROM CountryQuarterRevenue
)
SELECT CountryRegionCode, Quarter, TotalRevenue
FROM RankedRevenue
WHERE Rank = 1
ORDER BY TotalRevenue DESC;
```

Figure 3

Top-Selling Products and Regional Sales Trends

	CountryRegionCode	ProductName	SubCategoryName	TotalQuantitySold	TotalRevenue
1	US	Mountain-200 Black, 38	Mountain Bikes	1699	2386094.4902
2	CA	Mountain-200 Black, 38	Mountain Bikes	536	735839.9495
3	AU	Road-150 Red, 62	Road Bikes	111	397187.97
4	GB	Mountain-200 Black, 42	Mountain Bikes	240	390017.1553
5	FR	Mountain-200 Black, 38	Mountain Bikes	252	380029.6215
6	DE	Touring-1000 Yellow, 60	Touring Bikes	143	197812.0102

	CountryRegionCode	Quarter	TotalRevenue
1	US	1	16898250.6531
2	CA	1	4383315.1482
3	AU	1	2882815.0232
4	FR	2	2091032.0563
5	GB	2	2064151.7809
6	DE	2	1404926.338

The analysis of top-selling products by region reveals clear variations in consumer preferences across different countries. The Mountain-200 Black bike emerges as the best-selling product in multiple regions, particularly in the United States, Canada, France, and Great Britain, suggesting strong demand for this product category in these markets. Meanwhile, Touring and Road Bikes dominate in countries like Germany and Australia, indicating regional differences in product preference.
This insight suggests that consumer purchasing behavior is influenced by factors such as infrastructure, climate, and market positioning.

Additionally, an analysis of quarterly sales trends across different regions highlights seasonal variations in revenue generation. The United States consistently leads in sales, particularly in Quarter 1, followed by Canada and Australia, indicating strong demand early in the year. Conversely, European countries such as France, Germany, and the UK experience peak sales in Quarter 2, reflecting possible differences in consumer spending habits or economic cycles. Understanding these fluctuations allows businesses to optimize inventory planning and targeted promotions based on seasonal demand, ensuring maximum profitability and efficient supply chain management.

3. Analysing the Impact of Seasonality on Product Sales and Identifying Products with the Highest Seasonal Fluctuations

```
WITH QuarterlySales AS (
    SELECT
        d.Quarter,
        p.ProductName,
        p.SubCategoryName,
        SUM(f.ActualSaleValue) AS TotalRevenue
    FROM FactSale f
    JOIN DimProduct p ON f.ProductID = p.ProductID
```

```
    JOIN DimDate d ON f.OrderDate = d.DateKey
    GROUP BY d.Quarter, p.ProductName, p.SubCategoryName
),
FluctuationAnalysis AS (
    SELECT
        ProductName,
        SubCategoryName,
        MAX(TotalRevenue) AS MaxRevenue,
        MIN(TotalRevenue) AS MinRevenue,
        MAX(TotalRevenue) - MIN(TotalRevenue) AS RevenueFluctuation
    FROM QuarterlySales
    GROUP BY ProductName, SubCategoryName
),
PeakQuarter AS (
    SELECT
        QS.ProductName,
        QS.SubCategoryName,
        QS.Quarter AS PeakQuarter
    FROM QuarterlySales QS
    JOIN (
        SELECT ProductName, MAX(TotalRevenue) AS MaxRevenue
        FROM QuarterlySales
        GROUP BY ProductName
    ) AS MaxRevenueData
    ON QS.ProductName = MaxRevenueData.ProductName
    AND QS.TotalRevenue = MaxRevenueData.MaxRevenue
),
RankedFluctuation AS (
    SELECT
        F.ProductName,
        F.SubCategoryName,
        P.PeakQuarter,
        F.MaxRevenue,
        F.MinRevenue,
        F.RevenueFluctuation,
        RANK() OVER (ORDER BY F.RevenueFluctuation DESC) AS Rank
    FROM FluctuationAnalysis F
    JOIN PeakQuarter P ON F.ProductName = P.ProductName
)
SELECT
    ProductName,
    SubCategoryName,
    PeakQuarter,
    MaxRevenue,
    MinRevenue,
    RevenueFluctuation
FROM RankedFluctuation
WHERE Rank <= 10  -- Retrieve the top 10 products with the highest
seasonal fluctuations
ORDER BY RevenueFluctuation DESC;

WITH QuarterlySales AS (
    SELECT
```

```
            d.Quarter,
            p.ProductName,
            p.SubCategoryName,
            SUM(f.ActualSaleValue) AS TotalRevenue
        FROM FactSale f
        JOIN DimProduct p ON f.ProductID = p.ProductID
        JOIN DimDate d ON f.OrderDate = d.DateKey
        GROUP BY d.Quarter, p.ProductName, p.SubCategoryName
),
FluctuationAnalysis AS (
        SELECT
            Quarter,
            MAX(TotalRevenue) AS MaxRevenue,
            MIN(TotalRevenue) AS MinRevenue,
            MAX(TotalRevenue) - MIN(TotalRevenue) AS RevenueFluctuation
        FROM QuarterlySales
        GROUP BY Quarter
)
SELECT
        Quarter,
        MAX(RevenueFluctuation) AS HighestFluctuation
FROM FluctuationAnalysis
GROUP BY Quarter
ORDER BY Quarter;
```

Figure 4

Seasonal Sales Fluctuations and High-Impact Products

	ProductName	SubCategoryName	PeakQuarter	MaxRevenue	MinRevenue	RevenueFluctuation
1	Mountain-100 Silver, 42	Mountain Bikes	1	484838.574	83682.2552	401156.3188
2	Mountain-100 Black, 42	Mountain Bikes	1	475873.59	95680.968	380192.622
3	Mountain-100 Black, 38	Mountain Bikes	1	511718.6836	133565.231	378153.4526
4	Mountain-100 Black, 44	Mountain Bikes	1	496798.528	135758.9742	361039.5538
5	Mountain-100 Silver, 38	Mountain Bikes	1	456278.658	100512.206	355766.452
6	Mountain-100 Silver, 44	Mountain Bikes	1	461038.644	116832.1576	344206.4864
7	Mountain-100 Black, 48	Mountain Bikes	1	465073.622	141454.2699	323619.3521
8	Mountain-100 Silver, 48	Mountain Bikes	1	407318.802	87379.744	319939.058
9	Road-150 Red, 48	Road Bikes	1	561072.736	271948.52	289124.216
10	Mountain-200 Silver, 38	Mountain Bikes	2	1071663.2711	812823.6102	258839.6609

	Quarter	HighestFluctuation
1	1	1111656.8018
2	2	1241739.8585
3	3	1049698.5573
4	4	997123.736

The analysis of seasonal sales fluctuations reveals significant variations in revenue across different quarters, indicating strong seasonal demand patterns for certain products. The highest seasonal fluctuations are observed in products such as Mountain Bikes and Touring Bikes, which experience drastic revenue shifts between peak and off-peak seasons.

23

This suggests that demand for these products is likely influenced by factors such as weather conditions, outdoor activity trends, and promotional campaigns.

Additionally, by analyzing quarterly revenue fluctuations, we can see that specific quarters exhibit greater volatility, with Quarter 1 and Quarter 3 showing the most significant revenue changes. This implies that product sales are not evenly distributed throughout the year, requiring businesses to adjust inventory levels and marketing strategies accordingly. Understanding these fluctuations enables more effective demand forecasting, ensuring that high-demand products are adequately stocked during peak periods while minimizing excess inventory during slower seasons.

4. Analysing Customer Engagement and High-Value Identification Based on Product Purchase Count and Total Spend

```
SELECT top 5

    c.CustomerID,
    c.CustomerType,
    COUNT(c.CustomerID) AS ProductCount,
    SUM(f.ActualSaleValue) AS TotalSpent

FROM FactSale f
JOIN DimCustomer c ON f.CustomerID = c.CustomerID
GROUP BY c.CustomerID, c.CustomerType
ORDER BY TotalSpent DESC, ProductCount DESC;
```

Figure 5

Top High-Value and Most Engaged Customers

	CustomerID	CustomerType	ProductCount	TotalSpent
1	29818	Store	298	877107.1945
2	29715	Store	366	853849.1793
3	29722	Store	530	841908.7733
4	30117	Store	436	816755.5761
5	29614	Store	451	799277.8969

The analysis identifies the top five most engaged and high-value customers based on their total product purchases and spending. As seen in the output, all the top customers belong to the store customer type, reinforcing that businesses rather than individual consumers contribute the highest transaction volume and spending. The highest spender, CustomerID 29818, has spent $877,107.19 while purchasing 298 products, indicating a strong purchasing pattern.

24

Interestingly, while some customers have a higher product count, their total spending is lower than others with fewer purchases. This suggests that certain stores focus on high-ticket items, while others purchase a higher volume of lower-priced products. The insights from this analysis help in identifying the most valuable customers who should be targeted for retention strategies, personalized promotions, and exclusive incentives to maximize long-term revenue potential.

5. Analysing the Influence of Discounts on Sales Volume, Revenue, and Overall Profitability

```
SELECT
    f.DiscountType,
    AVG(f.DiscountPct) AS AvgDiscount,
    SUM(f.OrderQty) AS TotalUnitsSold,
    SUM(f.ActualSaleValue) AS TotalRevenue,
    SUM(f.GrossProfit) AS TotalProfit,
    SUM(f.GrossProfit) / NULLIF(SUM(f.ActualSaleValue), 0) *
100 AS ProfitMarginPercentage
FROM FactSale f
GROUP BY f.DiscountType
ORDER BY TotalRevenue DESC;
```

Figure 6

Top High-Value and Most Engaged Customers

	CustomerID	CustomerType	ProductCount	TotalSpent
1	29818	Store	298	877107.1945
2	29715	Store	366	853849.1793
3	29722	Store	530	841908.7733
4	30117	Store	436	816755.5761
5	29614	Store	451	799277.8969

The analysis evaluates the impact of discounts of sales volume, total revenue, and overall profitability. By grouping sales data based on discount types, we can observe how different discounting strategies influence business performance. The average discount percentage provides insights into the relative discount applied across transactions, while total units sold highlights how discounts drive purchase volume. Additionally, total revenue and total profit indicate the financial contribution of each discount type.

One key metric, profit margin percentage, is crucial in determining whether discounts are improving or hurting profitability. If higher discounts result in a significant increase in sales volume but a decline in profit margins, it suggests that deep discounting may not be sustainable. Conversely, if discounts maintain healthy profit margins while boosting revenue, they can be strategically leveraged to optimize sales. These findings enable businesses to fine-tune discount strategies focusing on those that maximize revenue without significantly eroding profit margins.

6. Analysing the Profitability of Product Categories and the Impact of Different Discount Strategies on Profit Margins

```
SELECT

    p.CategoryName,
    SUM(f.GrossProfit) AS TotalProfit,
    AVG(f.DiscountPct) AS AvgDiscountPct
FROM FactSale f
JOIN DimProduct p ON f.ProductID = p.ProductID
GROUP BY p.CategoryName
ORDER BY TotalProfit DESC;
```

Figure 7

Most Profitable Product Categories and Discount Influence

	CategoryName	TotalProfit	AvgDiscountPct
1	Bikes	7936394.0554	0.0072
2	Accessories	636394.2755	0.0012
3	Components	490233.1704	0.0001
4	Clothing	308882.2681	0.0022

The analysis of product category profitability and discount strategies reveals significant variations across different categories. Bikes emerge as the most profitable category, generating the highest total profit while maintaining a relatively low average discount percentage (0.0072). This suggests that bikes are high-margin products that do not require deep discounting to drive sales, making them a key revenue contributor.

Accessories and Components also contribute significantly to profitability, but their average discount percentages are slightly higher. This indicates that discounting may play a role in stimulating sales volume for these categories, although the overall impact on profitability remains strong. Clothing, despite generating profit, shows the highest discount percentage among the categories, suggesting that this category relies more on price reductions to drive sales.

These insights highlight the need for category-specific discounting strategies. High-margin products like bikes can sustain profitability with minimal discounts, while lower-margin categories may benefit from strategic promotions to balance revenue growth and profitability.

26

Summary Of Insights

One of the most interesting insights from the analysis is that while store customers generate higher total revenue, individual customers contribute significantly more to overall profitability. This is particularly intriguing because it challenges the common assumption that bulk purchases from stores are always more profitable. The negative profit margin associated with store customers suggests that bulk pricing models or high discount rates may be eroding profit margins. This insight is crucial for adjusting pricing strategies, optimizing discount policies, and ensuring bulk sales remain profitable without compromising revenue generation.

Another key finding is that bikes are the most profitable product category despite having one of the lowest average discount rates. This suggests that high-value products do not necessarily require heavy discounting to drive sales. Instead, demand for such products remains stable, allowing businesses to maintain profitability with minimal promotional incentives. This insight can help companies fine-tune their discounting approach, focusing price reductions on lower-margin categories while protecting high-value products from unnecessary discounts that might erode profit margins.

Lastly, the impact of seasonality on sales fluctuations highlights the importance of timing promotions and inventory planning. Certain quarters show significant revenue swings, with specific products experiencing high seasonal demand peaks. Understanding these patterns allows businesses to align marketing efforts, optimize stock levels, and launch well-timed promotions to maximize revenue during peak seasons while minimizing excess inventory during low-demand periods. This data-driven approach ensures efficient resource allocation and improved sales performance throughout the year.

27

References

Blattberg, R. C., Kim, B. D., & Neslin, S. A. (2008). Database marketing: Analyzing and managing customers. Springer Science & Business Media. https://link.springer.com/book/10.1007/978-0-387-72579-6

Reinartz, W. J., & Kumar, V. (2002). The mismanagement of customer loyalty. Harvard Business Review, 80(7), 86–94. https://hbr.org/2002/07/the-mismanagement-of-customer-loyalty

Ailawadi, K. L., Beauchamp, J. P., Donthu, N., Gauri, D. K., & Shankar, V. (2009). Communication and promotion decisions in retailing: A review and directions for future research. Journal of Retailing, 85(1), 42–55. https://doi.org/10.1016/j.jretai.2008.11.002

Winer, R. S. (2001). A framework for customer relationship management. California Management Review, 43(4), 89–105. https://doi.org/10.2307/41166102

OpenAI. (2024). ChatGPT [Large language model]. https://chatgpt.com/