IS 6480 - Data Warehousing

Wide World Importers

Data Warehouse

Krishna Swathi Peetha (u1475679)

24 July 2024

Table of Contents

1. Executive Summary

This report provides a detailed overview of the data warehouse (DW) implementation for the organization WideWorldImporters.

WideWorldImporters offers a diverse range of items, including novelty items, computing novelties, clothing, packaging materials, mugs, T-shirts, furry footwear, toys, and USB novelties. These items cater to different customer categories such as novelty shops, corporates, supermarkets, gift stores, and computer stores. The primary objective is to increase sales. By understanding customer purchasing patterns, stock group performance, and transaction trends, the organization can strategically allocate resources, develop targeted marketing strategies, and improve the product offerings.

The data warehouse implementation addresses key analytical requirements, including transaction analysis, customer analysis, order analysis, and stock group analysis. The transaction analysis focuses on the total transaction amounts over periods, by customer category, and by store group. Customer analysis includes the number of customers per category and the average credit limit for each category. Order analysis examines the number of orders and the quantity of orders per stock group. Stock group analysis looks at the average unit price per stock group.

The dimensional model for this data warehouse contains one fact table and three dimension tables. The FactTransaction table stores transactional data and links to the DimCustomer, DimOrder, and DimDate tables via warehouse keys. These tables provide detailed information on customers, orders, and dates, supporting comprehensive analysis. The data warehouse architecture involves sourcing data from Azure Blob Storage, loading it into the Lakehouse (Staging Area) through various data pipelines, and then transforming and cleaning the data. Clean tables are then loaded into the Warehouse (Presentation area), where dimensional and fact tables are created. Finally, reports are generated based on the requirements and the data.

The reports generated include insights into total transaction amounts per customer category, total transaction amounts per stock group, the number of orders per stock group, the average unit price per stock group, the number of customers per customer category, and the average credit limit per customer category. These reports provide actionable insights that guide resource allocation and operational improvements.

2. Background

a. Products/Services Provided by the Organization

WideWorldImporters provides a wide range of items classified into categories including novelty items, computing novelties, clothing, packaging materials, mugs, T-shirts, furry footwear, toys, and USB novelties. These items are supplied to various customer categories such as novelty shops, corporates, supermarkets, gift stores, and computer stores. The offerings are diverse and cater to a broad market segment.

b. Objectives for the Organization

The primary objective of WideWorldImporters is to increase sales.

By analyzing which products and customer categories drive the most revenue, targeted marketing and sales efforts can be implemented to boost sales.

By gaining insights into customer purchasing patterns, stock group performance, and transaction trends the objective can be achieved effectively.

c. How a Data Warehouse Would Support the Organization's Objectives

The data warehouse supports WideWorldImporters' objective of increasing sales by providing valuable insights through transaction analysis, order analysis, and stock group analysis.

- Transaction Analysis: By examining total transaction amounts over time and across customer categories, the organization can identify top-performing products and customer segments, enabling targeted marketing and sales efforts to boost revenue.
- Order Analysis: Understanding the number of orders and quantities per stock group helps identify popular products, guiding inventory management and promotional activities to enhance sales.
- Stock Group Analysis: Analyzing the average unit price per stock group offers insights into pricing strategies and product profitability, allowing the organization to refine pricing strategies and optimize revenue and cost management.

3. Requirements

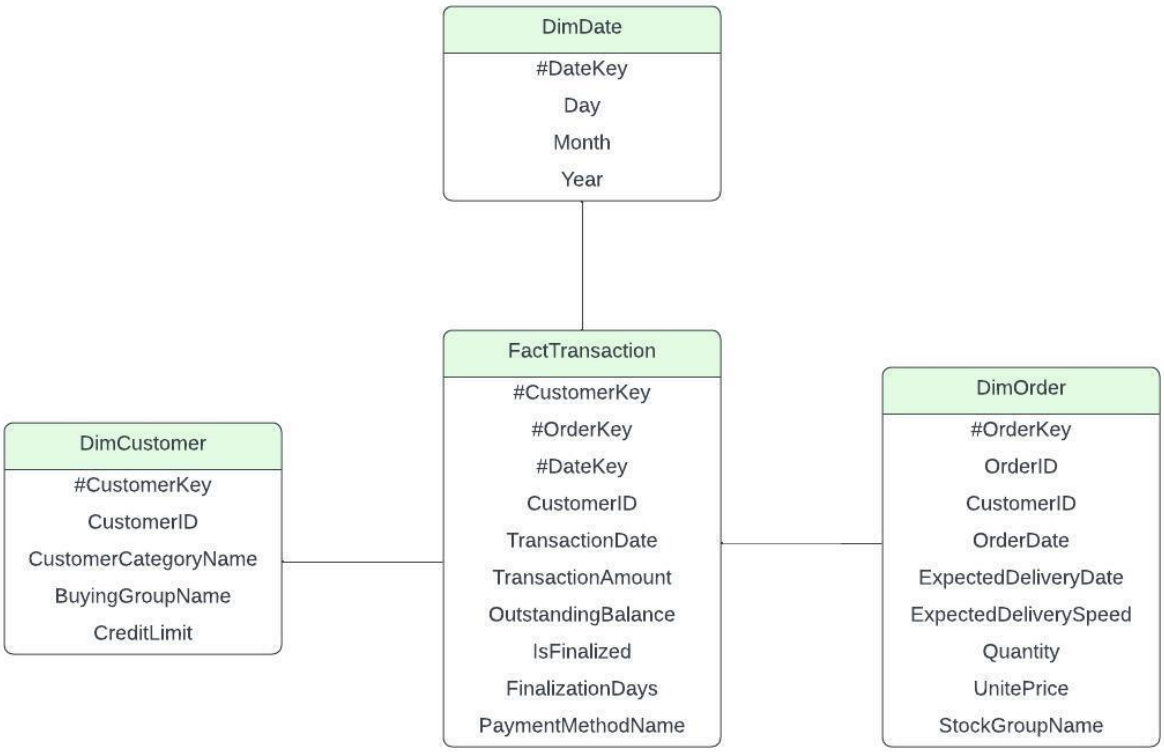
a. Summary of primary requirements

Requirement	Short Description	Status
Transaction Analysis		
Transactions over period	Acquire the trend of total transaction amounts over the years.	Planned for future
Transactions per Customer Category	Acquire total transactions per customer category. Acquire which customer category has the highest and least transactions per customer category.	Completed
Transactions per Store Group	Acquire total transactions per store group. Acquire which customer category has the highest and least transactions per store group.	Completed
Customer Analysis		
Customers per customer category	Acquire the number of customers under each customer category.	Completed
Credit Limit for customer catgory	Acquire the average credit limit offered to each customer category.	Completed
Order Analysis		
Orders per stock group	Acquire the orders that came for each stock group	Completed
Quantity of orders	Acquire the number of orders that came for each stock group.	Completed
Stock Group Analysis		
Unit price per stock group	Acquire the average unit price per stock group	Completed

b. Completed bus matrix

	Dimensions								
Business processes	**Buying Groups**	**Customer**	**CustomerCategories**	**Transactions**	**Orders**	**OrderLines**	**StockGroups**	**PaymentMethods**	**Date**
Transaction Analysis									
Transactions over period				x					x
Transactions per Customer Category		x	x	x					
Transactions per Store Group				x			x		
Customer Analysis									
Customers per customer category		x	x						
Credit Limit for customer catgories		x	x						
Order Analysis									
Orders per stock group					x	x	x		
Quantity of orders					x	x			
Stock Group Analysis									
Unit price per stock group						x	x		

4. Dimensional Model



This dimensional model contains one fact table and three dimension tables. Below is a breakdown of the schema components:

Fact Table:

FactTransaction: This table stores transactional data and includes the following columns:
#CustomerKey: Foreign key linking to the DimCustomer table.
#OrderKey: Foreign key linking to the DimOrder table.
#DateKey: Foreign key linking to the DimDate table.
CustomerID: The ID of the customer involved in the transaction.
TransactionDate: The date the transaction occurred.
TransactionAmount: The amount of money involved in the transaction.
OutstandingBalance: The remaining balance after the transaction.
IsFinalized: Indicates whether the transaction is finalized.

FinalizationDays: Number of days taken to finalize the transaction.
PaymentMethodName: The name of the payment method used.

Dimension Tables:

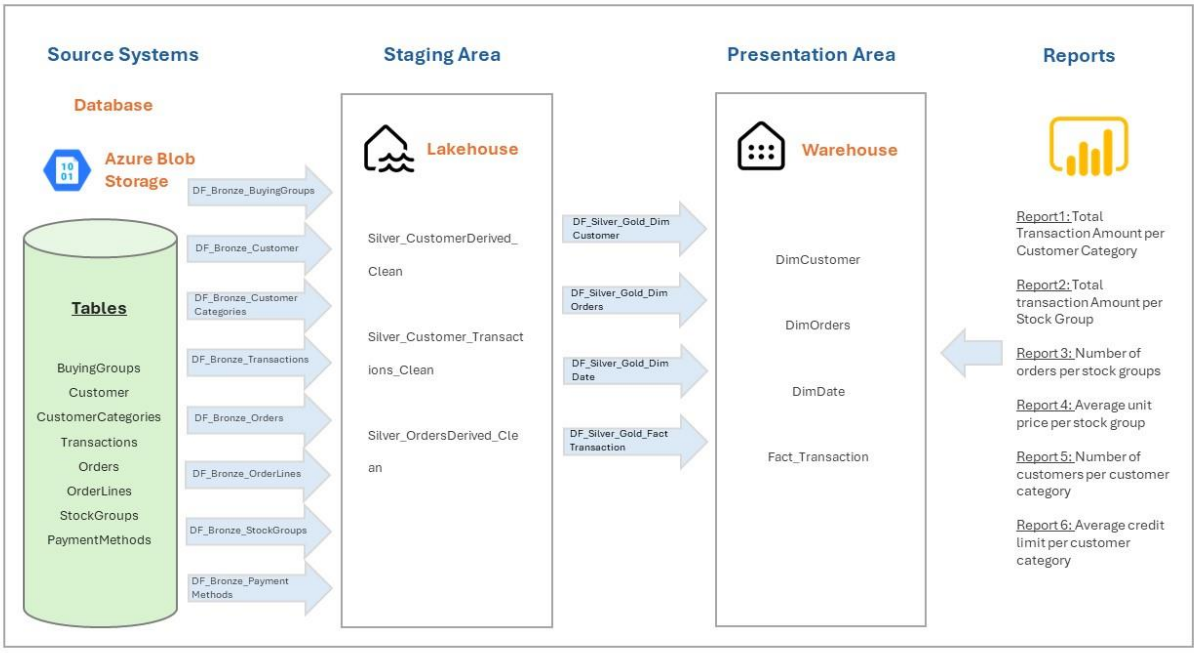
DimDate: This table contains date information and includes:
#DateKey: Primary key for the date dimension.
Day: The day of the month.
Month: The month.
Year: The year.

DimCustomer: This table contains customer-related information and includes:
#CustomerKey: Primary key for the customer dimension.
CustomerID: The ID of the customer.
CustomerCategoryName: The category or type of the customer.
BuyingGroupName: The buying group to which the customer belongs.
CreditLimit: The credit limit assigned to the customer.

DimOrder: This table contains order-related information and includes:
#OrderKey: Primary key for the order dimension.
OrderID: The ID of the order.
CustomerID: The ID of the customer who placed the order.
OrderDate: The date the order was placed.
ExpectedDeliveryDate: The expected date of delivery.
ExpectedDeliverySpeed: The expected speed of delivery.
Quantity: The quantity of items ordered.
UnitPrice: The price per unit of the items ordered.
StockGroupName: The name of the stock group the items belong to.

Relationships:
The FactTransaction table is connected to the DimDate table via #DateKey, to the DimCustomer table via #CustomerKey, and to the DimOrder table via #OrderKey.

5. Data Warehouse Architecture



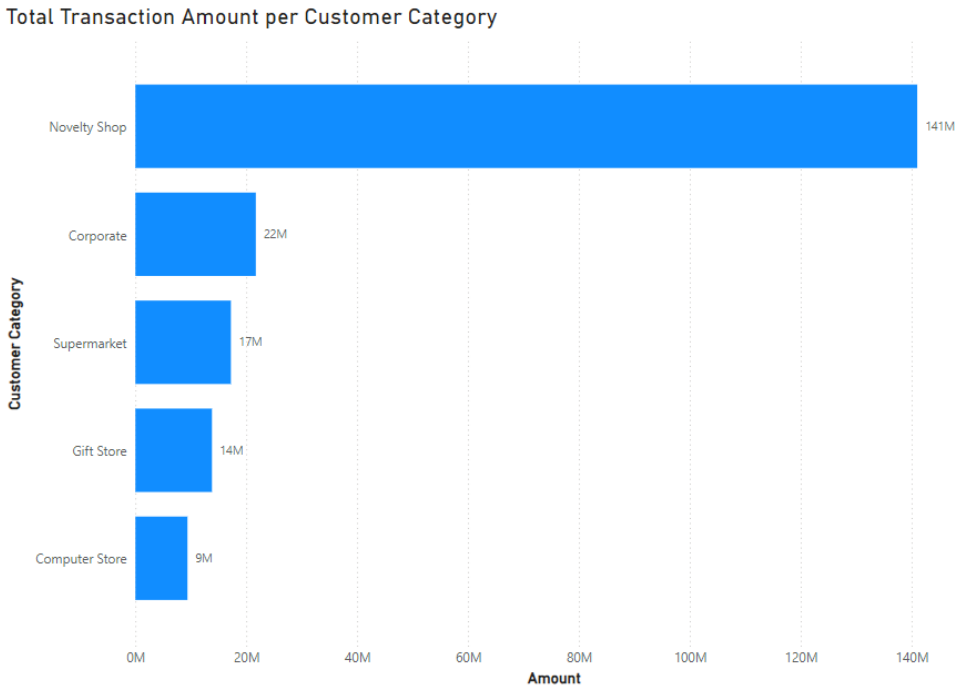
The data of WideWorldImporters organization was sourced from Azure Blob Storage and loaded into the Lakehouse (Staging Area) through different data pipelines like DF_Bronze_Orders, DF_Bronze_Customers, etc.

Derived tables were created by combining relevant tables: the CustomerDerived table combined data from Customer, CustomerCategories, and BuyingGroups tables, while the OrdersDerived table combined data from Orders, OrderLines, and StockGroups tables. After transforming, the data was then cleaned, resulting in the tables: CustomerDerived_Clean, CustomerTransactions_Clean, and OrdersDerived_Clean in the Lakehouse.

Next, the clean tables were loaded from the Lakehouse, transformed into dimension and fact tables and are loaded into the Warehouse (Presentation area). Dimensional and fact tables were in the Warehouse. Finally, reports were generated based on the requirements and these dimensional and fact tables.

6. Current Dashboards/Reports

a. Total Transaction Amount per Customer Category

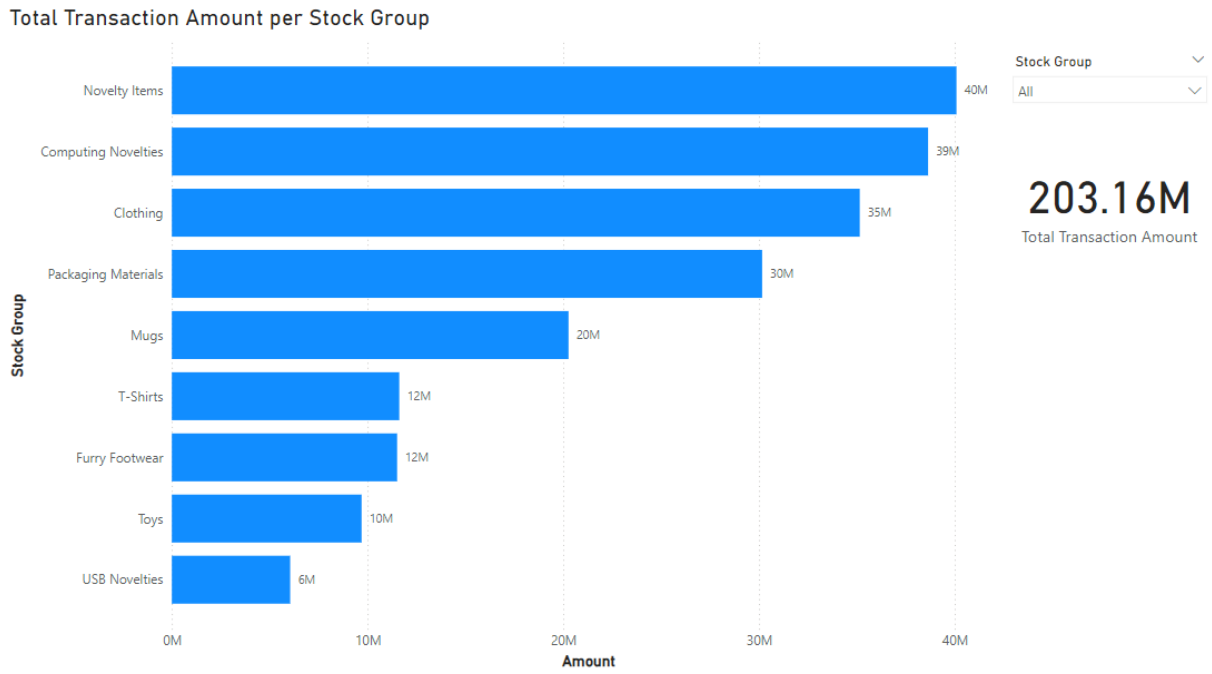


This visualization shows the total transaction amount per customer category, highlighting the distribution of revenue across different customer categories.

High-performing categories include Novelty Shop and Corporate, which generate the highest transaction amounts. This suggests that these customer categories are significant revenue generators and there should be the focus on continuous marketing and sales efforts.

By understanding the revenue contributions of each customer category, the organization can allocate resources more effectively, develop strategies to boost sales in underperforming segments, and keep focusing on top.

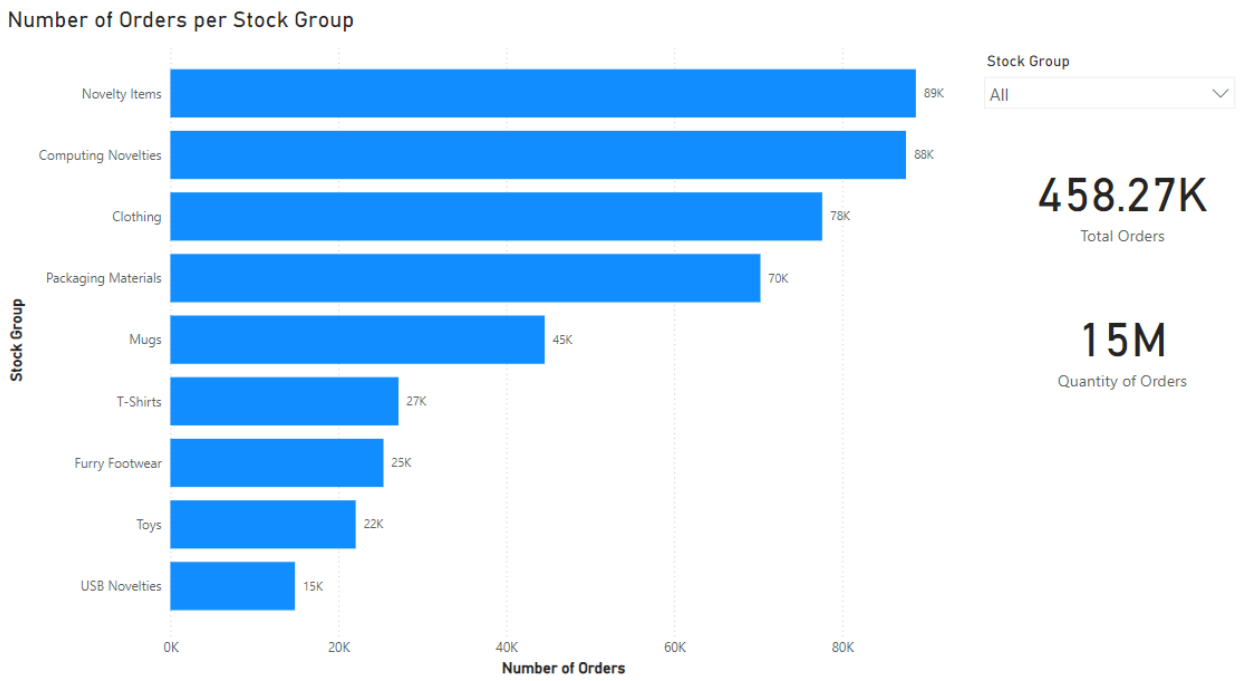
b. Total Transaction Amount per Stock Group



This chart displays the total transaction amount per stock group, providing insights into the financial performance of different product categories. The slicer on the right can be used to select single or multiple sales groups, and the overall transaction amount for the selected groups will be displayed in a total transaction amount card.

Novelty Items and Computing Novelties are the top-performing groups, indicating strong market demand and significant revenue generation. These categories should continue to receive adequate inventory and marketing support to maintain and grow their market share. On the other hand, lower-performing groups like Toys and USB Novelties highlight areas that may need strategic intervention, such as promotional activities.

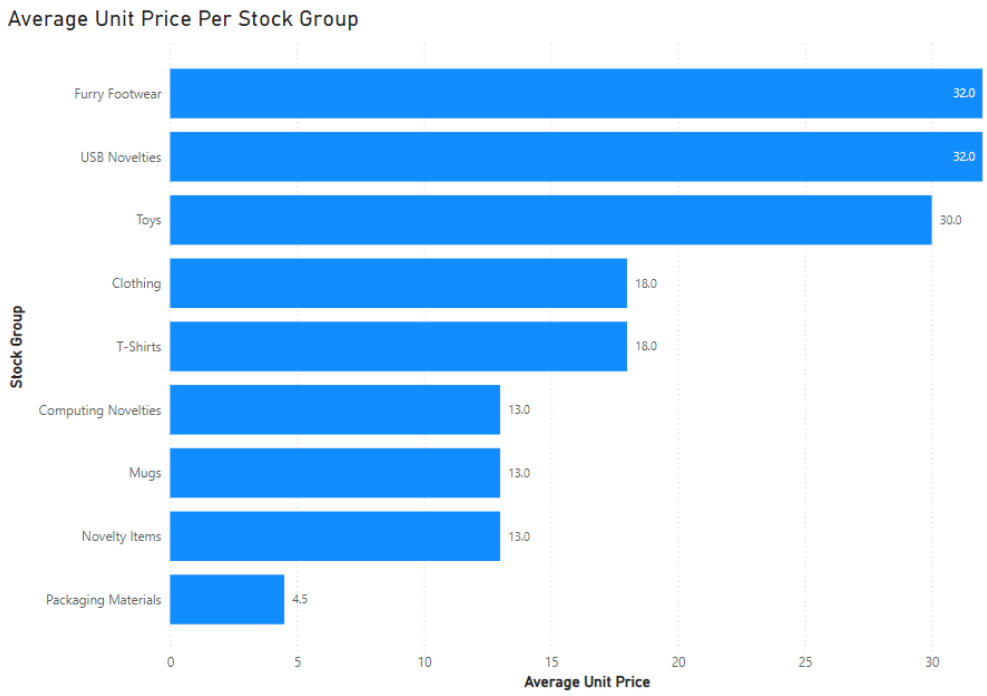
c. Number of Orders per Stock Group



This chart illustrates the number of orders per stock group, offering insights into customer purchasing patterns across different product categories. The slicer on the right can be used to select single or multiple sales groups, and the total orders and quantity of orders for the selected groups will be displayed in the respective cards.

Novelty Items and Computing Novelties have the highest number of orders, indicating strong customer demand. This suggests that these categories are popular and should be prioritized in terms of inventory management and marketing efforts. Mid-level categories like Clothing and Packaging Materials also show significant order numbers, showing opportunities for further growth. Lower-order categories such as Toys and USB Novelties may need strategic interventions to boost their sales.

d. Average Unit Price per Stock Group

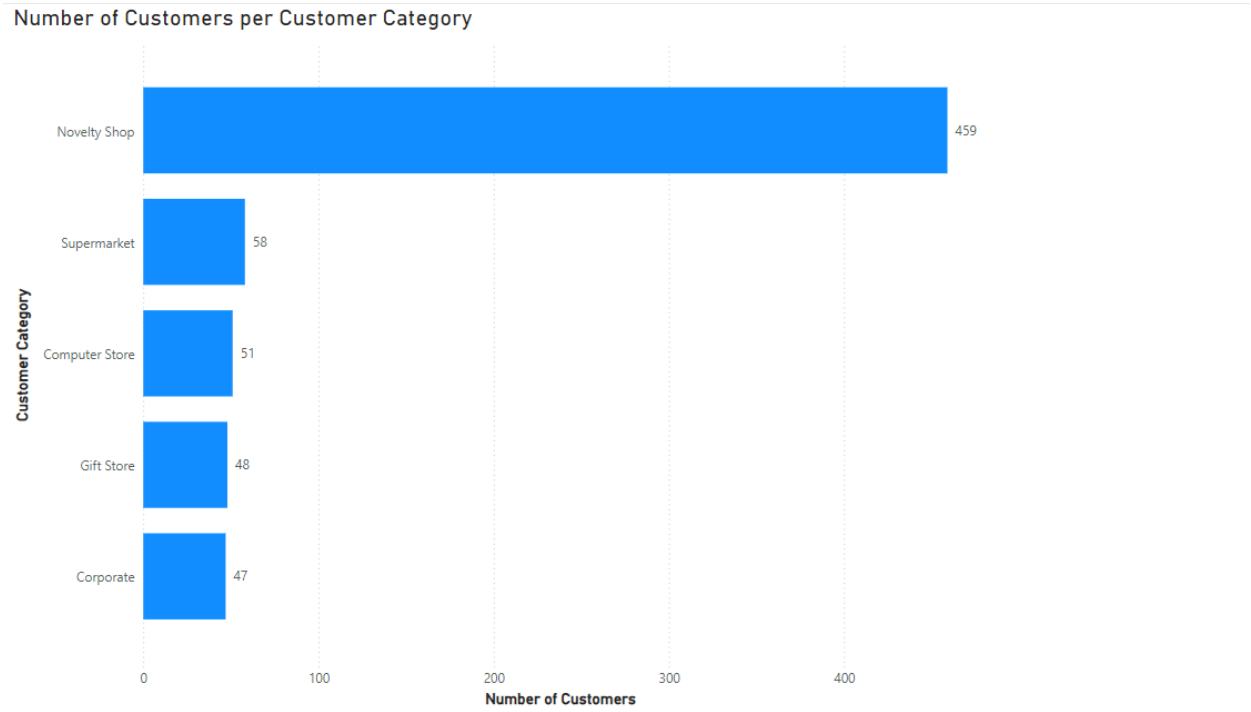


This chart displays the average unit price per stock group, providing insights into the pricing strategies of different product categories.

Furry Footwear and USB Novelties have the highest average prices, indicating a premium pricing strategy. Mid-range pricing is observed for Clothing and T-Shirts. Lower-priced categories like Packaging Materials may increase volume sales but require careful cost management to ensure profitability.

By understanding the average unit prices, the organization can refine its pricing strategies to maximize revenue and market competitiveness.

e. Number of Customers per Customer Category



This visualization shows the number of customers per customer category, highlighting the distribution of the customer base across different categories.

Novelty Shop has the highest number of customers, indicating a broad customer base and strong market scope. This category should be a focal point for customer retention and engagement strategies. Other categories like Supermarket and Computer Store also have a substantial number of customers.

Understanding the customer distribution helps the organization tailor its marketing and sales efforts to different segments effectively.

f. Average Credit Limit per Customer Category

Average Credit Limit per Customer Category

CustomerCategoryName	Average of CreditLimit
Supermarket	2,703.28
Computer Store	2,638.42
Corporate	2,620.31
Novelty Shop	2,577.93
Gift Store	2,521.46
Total	**2,614.85**

This table shows the average credit limit per customer category, providing insights into the financial capabilities of different customer segments.

Supermarket and Computer Store customers have the highest average credit limits. On the other hand, categories like Gift Store and Novelty Shop have lower average credit limits.

By understanding the average credit limits, the organization can tailor its product offerings and payment terms to match the financial profiles of different customer categories, enhancing overall sales and customer satisfaction.

7. Future Development

a. Dashboards/Reports/Analyses Supported by the Current Data Warehouse

Trend Analysis of Total Transaction Amounts Over Time

This dashboard will visualize the trend of total transaction amounts over different periods (e.g., monthly, quarterly, yearly). It will help in identifying seasonal trends, growth patterns, and potential inconsistencies in sales performance. By understanding these trends, the organization can make informed decisions about marketing strategies and sales forecasting.

The current data warehouse already contains the necessary transactional data and date dimensions. This analysis can be easily supported by leveraging existing data. The DimDate table and FactTransaction table will be utilized to visualize transaction amounts over different time periods.

b. Dashboards/Reports/Analyses Requiring Updates to the Dimensional Model

Product Sales Performance by Region

The current data warehouse doesn't support regional sales performance analysis. Gaining deeper insights, such as locally demanded products, impact of regional marketing campaigns, can be very beneficial to the organization.

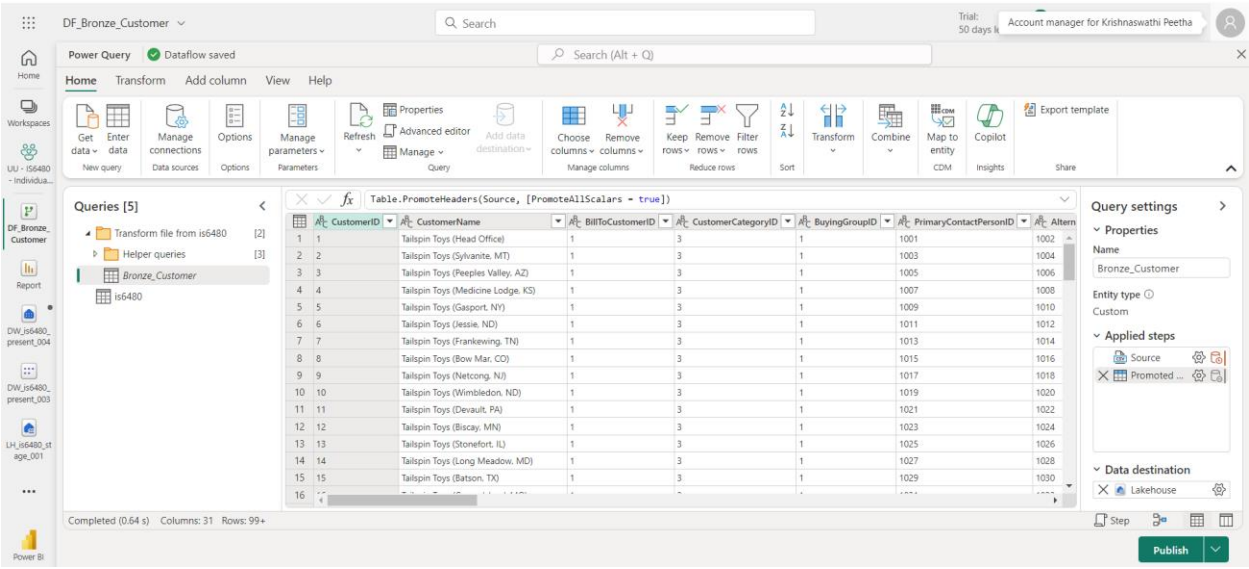
To enhance regional sales performance analysis, the dimensional model would need updates to include more detailed geographic attributes in the DimCity table (e.g., city name, state name).

By developing these additional dashboards and reports, WideWorldImporters can further leverage its data warehouse to gain deeper insights and support strategic decision-making.

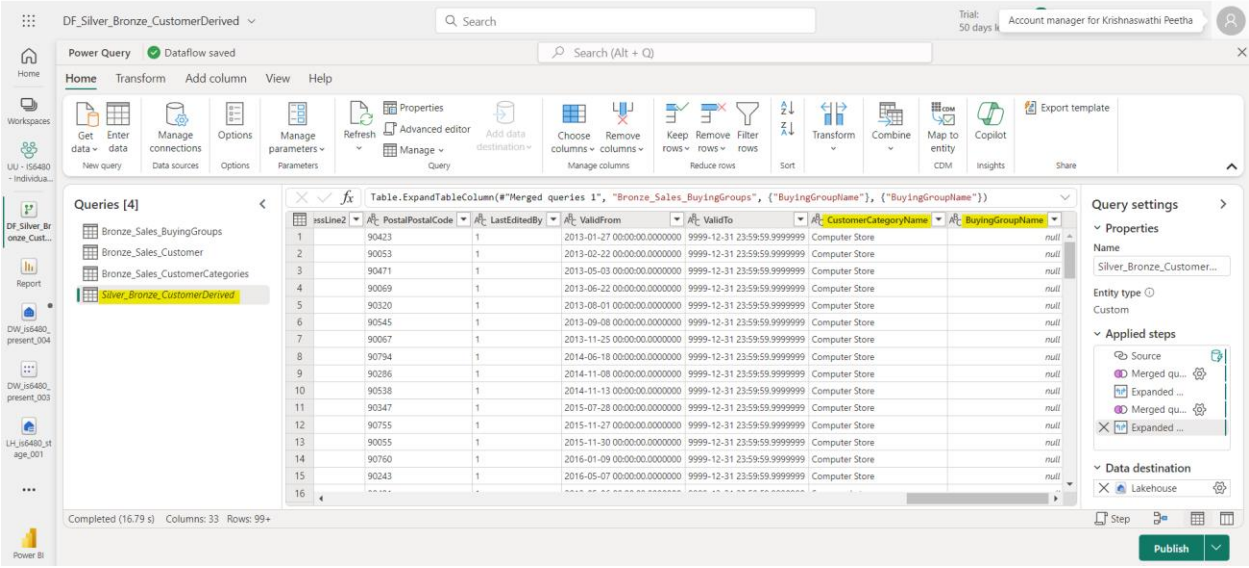
8. Appendix

a. Screenshots of Pipelines, Lakehouse and Warehouse

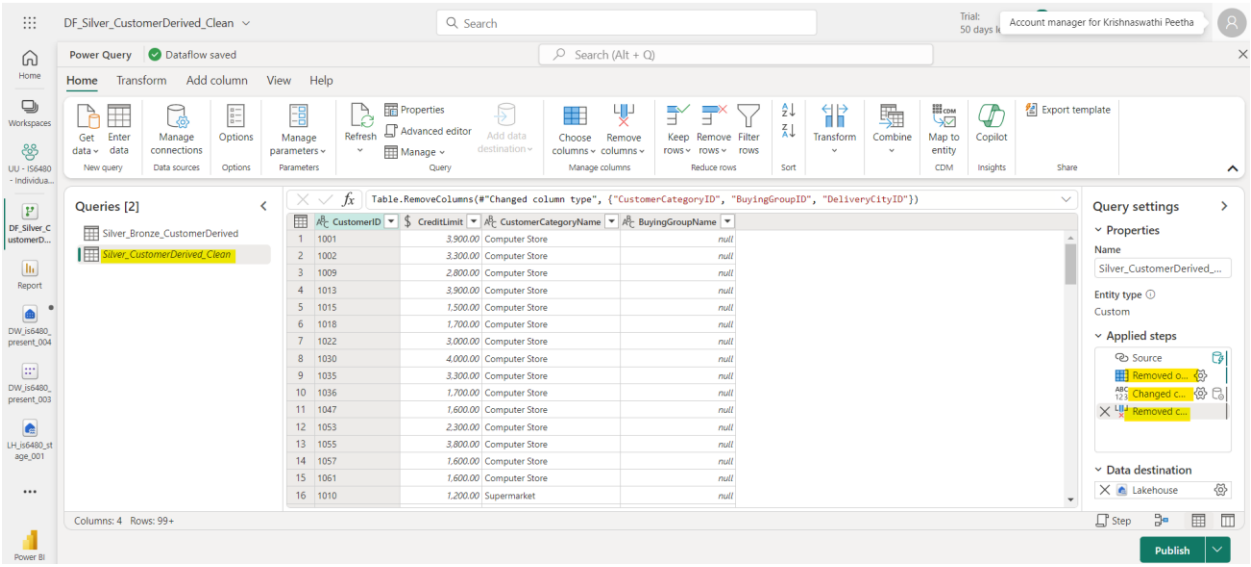
Screenshot 1: Raw data Bronze_Customer is loaded to Lakehouse from source



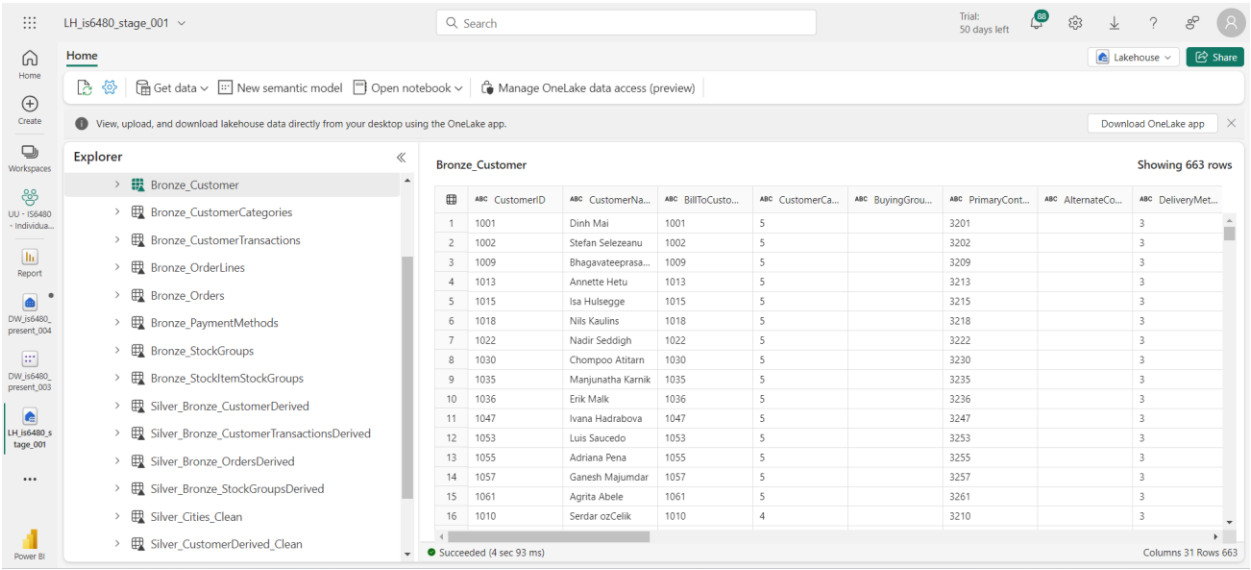
Screenshot 2: Fetched Bronze_Customer, Bronze_SalesCategories, Bronze_BuyingGroups from Lakehouse and transformed into single Silver_Bronze_CustomerDerived, loaded this new table to Lakehouse



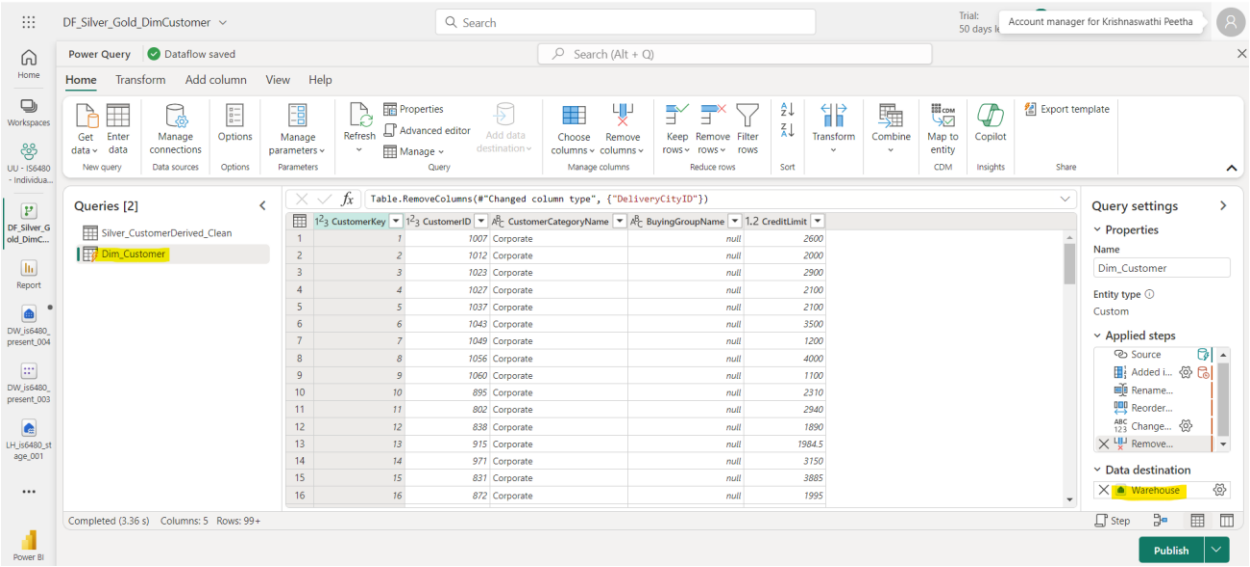
Screenshot 3: Derived Silver_Bronze_CustomerDerived from Lakehouse, cleaned the table (removed columns, changed data type) and loaded back to Lakehouse



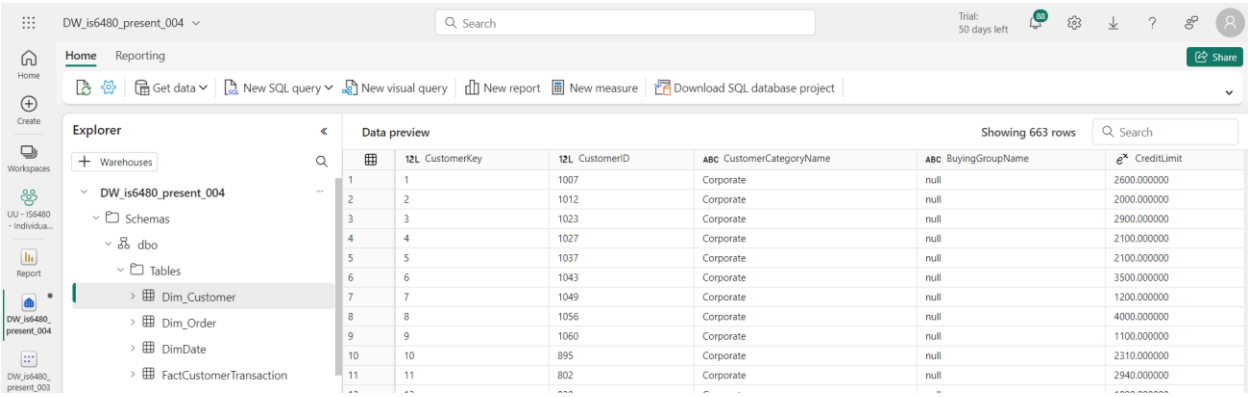
Screenshot 4: Lakehouse



Screenshot 5: Loaded Silver_Bronze_CustomerDerived from Lakehouse. Took reference, added warehouse key and created DimCustomer. Loaded this to Warehouse



Screenshot 6: Warehouse



App Link-

https://app.powerbi.com/Redirect?action=OpenApp&appId=1992a2bf-f893-455d-a478-27af77a3d9e8&ctid=5217e0e7-539d-4563-b1bf-7c6dcf074f91