



BITS Pilani

Pilani Campus

Data Warehouse for Amul Sales.

Submitted by -

Vaibhav Vyas

Aarnav JP (2024H1030075P)

Amir S Faizal (2024H1030073P)

Submitted to -

Dr. L. Rajya Laxmi

Data Warehouse - SS G515

Table of Contents -



❖	Problem statement
❖	Analytical Requirements
❖	Subject Area Identification
❖	Information Package Diagrams (Frontend and Backend Subjects)
❖	Summarized Dimension Model
❖	Star Schemas
❖	Queries and Results

Problem statement

Operational Overview



❖ **Scale & Scope:**

- One of India's largest dairy cooperatives
- Extensive operations: procurement, production, distribution, retail

❖ **Data Insights Needs:**

- Supplier performance evaluation
- Sales trends analysis
- Inventory management and monitoring
- Customer behavior insights
- Effectiveness of promotional activities

Problem statement(2/2)

Need for Data Warehouse



❖ **Performance Tracking:**

- Identifying top-performing products, regions, and retail outlets

❖ **Operational Efficiency:**

- Understanding supplier reliability and addressing production bottlenecks
- Monitoring warehouse capacities and predicting inventory shortfalls

❖ **Customer & Market Insights:**

- Tracking customer engagement and evolving preferences
- Evaluating product returns, reasons for failures, and demand signals

Analytical Requirements



❖ Order Analysis

- Track sales performance by product, store, and region.
- Identify high-value customers and purchasing trends.
- Analyze seasonal demand fluctuations and promotional impact.

❖ Customer Behavior Analysis

- Monitor customer-wise sales volume and spending patterns.
- Identify top cities by average revenue per customer.
- Examine product enquiries to gauge customer interest.

❖ Supplier and Production Analysis

- Analyze quantity and cost of supplies across suppliers and factories.
- Compare production volumes and costs by product and plant.
- Determine supplier efficiency using supply-to-cost ratios.

❖ Inventory and Distribution Analysis

- Track stock levels across warehouses and identify over/under-utilization.
- Rank warehouses by total inventory weight and dominant category.
- Evaluate distributor-to-wholesaler product flows.

❖ Returns and Failure Analysis

- Identify most common return reasons and failure categories.
- Analyze quantity and frequency of returns by region or store.
- Quantify refunds and losses due to returns.

❖ Product Demand Analysis

- Identify top-enquired products across zones.
- Compare enquiry volume to actual inventory levels.
- Highlight high-demand products understocked in warehouses.

Subject Area Identification



- ❖ In order to organize the analytical scope of the Amul Milk Marketing Company, we identified several subject areas based on the business functions.
- ❖ Each subject area corresponds to one or more fact tables that capture event-driven metrics, supported by relevant dimension tables to enable multi-dimensional analysis.
- ❖ Each of these subject areas has been modeled around a core fact table with supporting normalized and conformed dimensions such as DateDim, ProductDim, CustomerDim, and StoreParlourDim, enabling unified and flexible analytics across domains.

Back-end Operations based subjects :

- ❖ Production and Supplier
 - Supports evaluation of production volumes, costs, and supplier performance to optimize factory and vendor operations
- ❖ Inventory
 - Provides a comprehensive view of product stock across warehouses, enabling monitoring of inventory levels and category-wise aggregation.
- ❖ Distribution
 - Enables tracking of product distribution flows across the supply chain, specifically from distributors to wholesalers.

Front-end Operations based subjects :

- ❖ Sales and Transactions
 - Facilitates detailed and aggregate-level analysis of product sales across parlours, customers, and distributors.
- ❖ Returns and Failure Analysis
 - Captures information related to product returns, including refund amounts and failure reasons, for loss assessment and quality insights.
- ❖ Customer Interest and Enquiries
 - Captures customer interest data through product enquiries at parlours, aiding in demand forecasting and engagement analysis.

Dimensional Modelling(1/4)



- ❖ **Dimensional modeling** is a data design technique used in data warehousing to organize data into fact and dimension tables for efficient analytical querying. It structures business data around key subject areas, using star or snowflake schemas, guided by Information Package Diagrams to capture analysis requirements.

Information Package Diagrams for Front-end subjects :

1) Sales Subject :

DateDim	CustomerDim	ParlourDim	ProductDim	PaymentDim
Date	CustomerID	ParlourName	ProductID	PaymentMethod
Month	FirstName	City	ProductName	PaymentProvider
Quarter	LastName	State	Brand	
Year	Email	Zone	Category	
	Phone		Price	
Measured Facts: Quantity Sold, Sales Amount, Bill Number				

Dimensional Modelling(2/4)



2) Customer interest Analysis Subject :

DateDim	CustomerDim	ParlourDim	ProductDim
Date	CustomerID	ParlourName	ProductName
Month	FirstName	City	Brand
Quarter	LastName	State	Category
Year	Email	Zone	
	Phone		
Measured Facts: None (Factless – logs product interest)			

3) Returns Subject :

DateDim	CustomerDim	ParlourDim	ProductDim	ReasonDim	JunkDim
Date	CustomerID	ParlourName	ProductName	ReasonCode	IsPromotional
Month	FirstName	City	Brand	ReasonDesc	IsOnline
Quarter	LastName	State	Category	ReasonCategory	IsReturnable
Year	Email	Zone	Price		IsClearance
	Phone				
Measured Facts: Quantity Returned, Refund Amount					



Information Package Diagrams for Back-end subjects :

1) Distribution Subject :

DateDim	SupplierDim	FactoryDim
Date	SupplierID	FactoryID
Month	SupplierName	FactoryName
Quarter	City	City
Year	State	State
	Zone	ManagerName
Measured Facts: Quantity Supplied, Supply Cost		

Supplier to factory

DateDim	DistributorDim	WholesalerDim	ProductDim
Date	DistributorID	WholesalerID	ProductName
Month	DistributorName	WholesalerName	Brand
Quarter	City	City	Category
Year	State	State	Price
	Zone	Zone	UnitSize
Measured Facts: Quantity Distributed			

Distribution to Wholesaler

Dimensional Modelling(4/4)



2) Production Subject :

DateDim	FactoryDim	ProductDim
Date	FactoryID	ProductName
Month	FactoryName	Brand
Quarter	City	Category
Year	State	Price
	Zone	UnitWeight
Measured Facts: Quantity Produced, Product Cost		

3) Inventory Subject :

DateDim	WarehouseDim	ProductDim
Date	WarehouseID	ProductName
Month	WarehouseName	Brand
Quarter	City	Category
Year	State	Price
	Zone	UnitWeight
Measured Facts: Quantity On Hand		

Summarized Dimension model (1/2)



- ❖ 10 Fact tables have been implemented with 7 being Base and 3 Aggregated Tables

Table Name	Sub-Type	Notes
SalesFact	Normal	Core transactional sales data (BillNumber is a degenerate measure)
ReturnsFact	Normal	Captures product return events
InventoryFact	Normal	Tracks warehouse-level stock snapshots
ProductionFact	Normal	Tracks quantity produced per factory
SupplierSupplyFact	Normal	Captures supply quantity and cost from suppliers
DistributionFact	Normal	Product movement: Distributor → Wholesaler
ProductEnquiryFact	Factless	Logs customer interest in a product
Sales_Aggregate_ByDistributor	Aggregate	Sales grouped by Distributor, Customer, Parlour
SupplierSupply_Aggregate_ByMonth	Aggregate	Monthly aggregation of SupplierSupplyFact
Inventory_Aggregate_ByCategory	Aggregate	Category-wise quantity on hand by warehouse

Quantity sold and sales amount

Quantity Returned

Quantity in hand

Quantity produced and product cost

Production and Supplier facts table belongs to same subjects Production and supply but have been implemented separately.

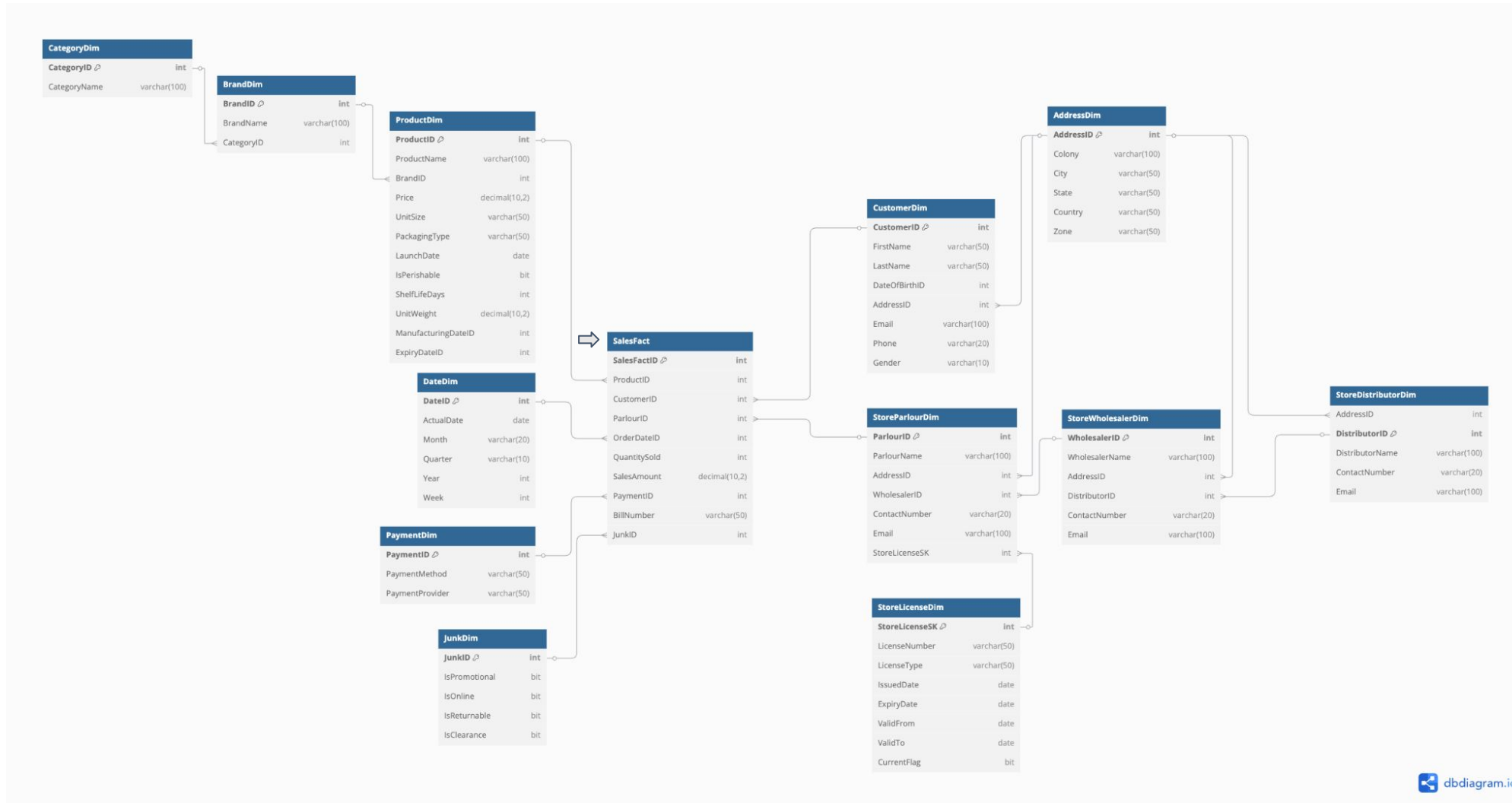
Summarized Dimension model(2/2)



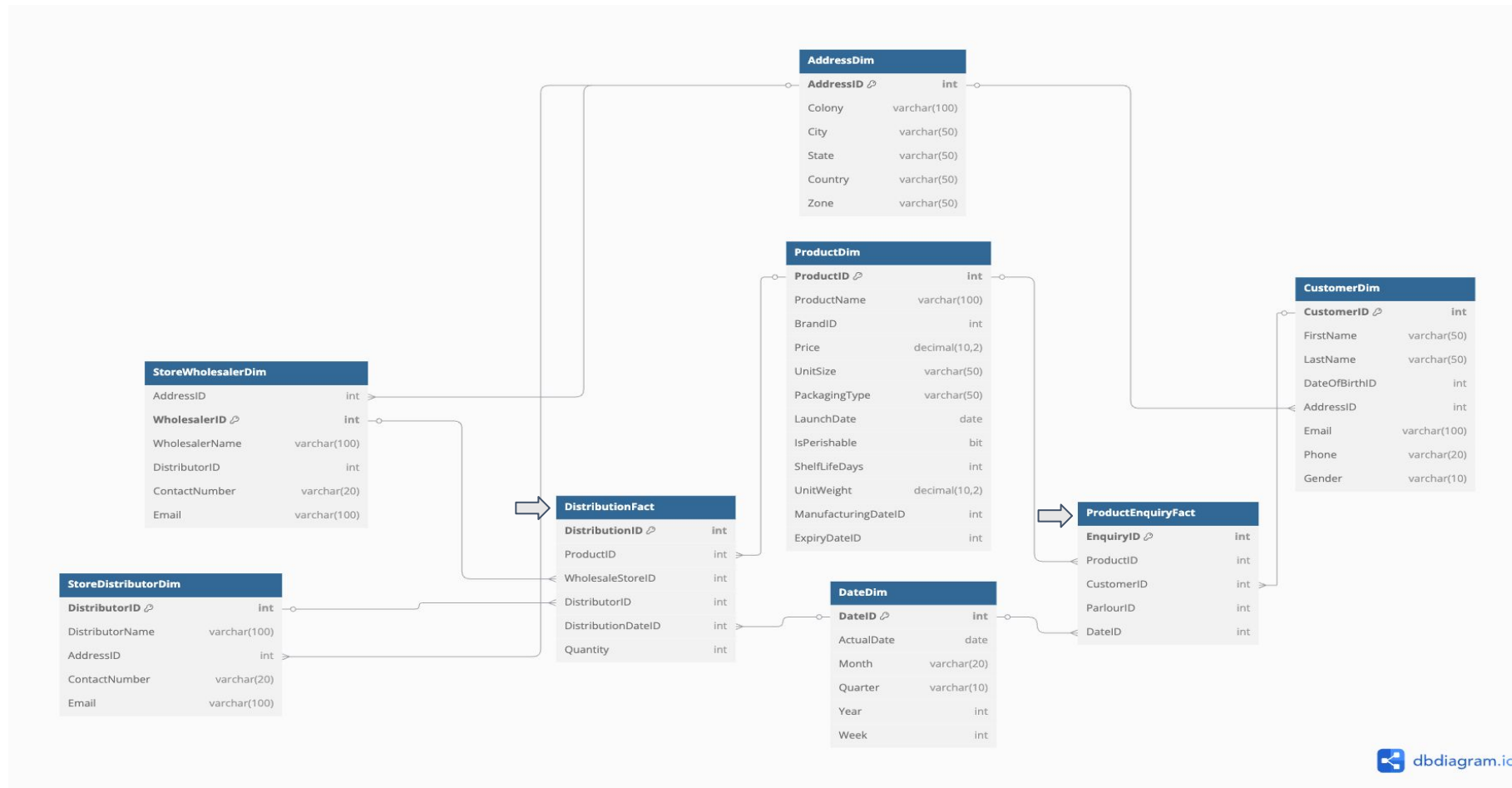
- ❖ Following are the conformed dimension tables, that have been implemented along, which analysis of the measures/facts have been done. We have implemented 16 dimension.

Table Name	Sub-Type	What It Stores
CustomerDim	Normal	Information about individual customers including name, contact, gender, birth-date, and address
AddressDim	Outrigger	Geographical details such as colony, city, state, country, and zone shared across multiple entities
DateDim	Roleplaying (RP: MFG/EXP)	Daily date records with role-playing views for Manufacturing and Expiry tracking
CategoryDim	Normal	Product categories such as Dairy, Beverage used for product classification
BrandDim	Derived	Brand details for each category; acts as child of CategoryDim
ProductDim	Derived	Detailed product data including brand, pricing, packaging, perishable info, and MFG/EXP dates
PaymentDim	Normal	Accepted payment methods and associated providers (e.g., UPI-PayTM)
StoreDistributorDim	Normal	Distributor store information including contact and address (1 per zone)
StoreWholesalerDim	Normal	Wholesaler store details linked to a distributor and serving specific cities
StoreParlourDim	Normal	Individual Amul parlours linked to wholesalers, including licensing and contact details
StoreLicenseDim	SCD Type 1	Latest license details for parlours including issue/expiry dates and validity flag
FactoryDim	Normal	Information about Amul manufacturing factories and their managers
WarehouseDim	Normal	Storage site details for inventory with capacity, refrigeration, and location
SupplierDim	Normal	Registered third-party suppliers including contact and registration details
ReasonDim	Derived	Defined reasons for returns categorized by type, including code and description
JunkDim	Derived	Binary flag combinations (e.g., promotional, online, returnable, clearance) used across facts

Star Schemas (1 / 5)



Star Schemas (2 / 5)

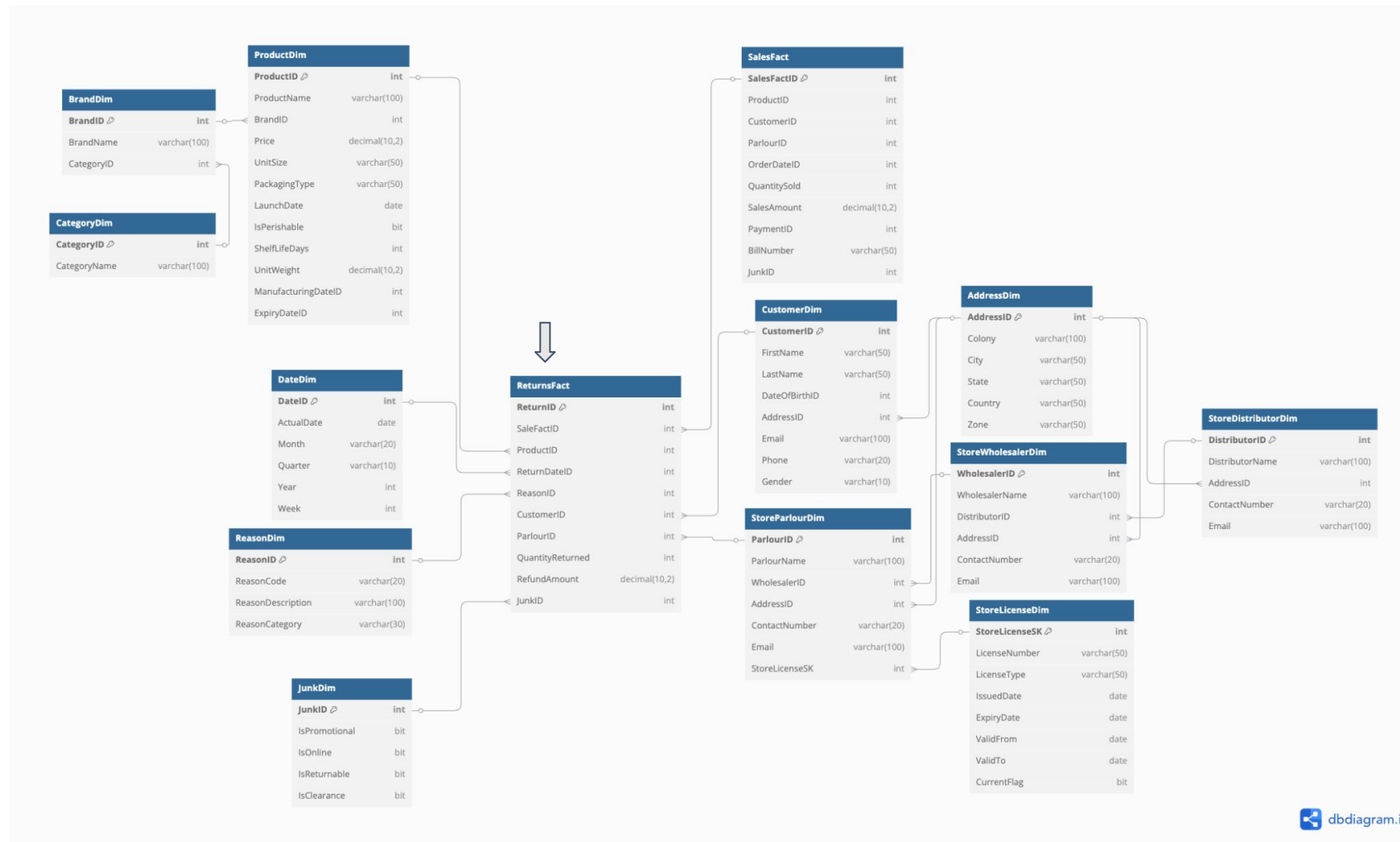


Star Schemas (3 / 5)

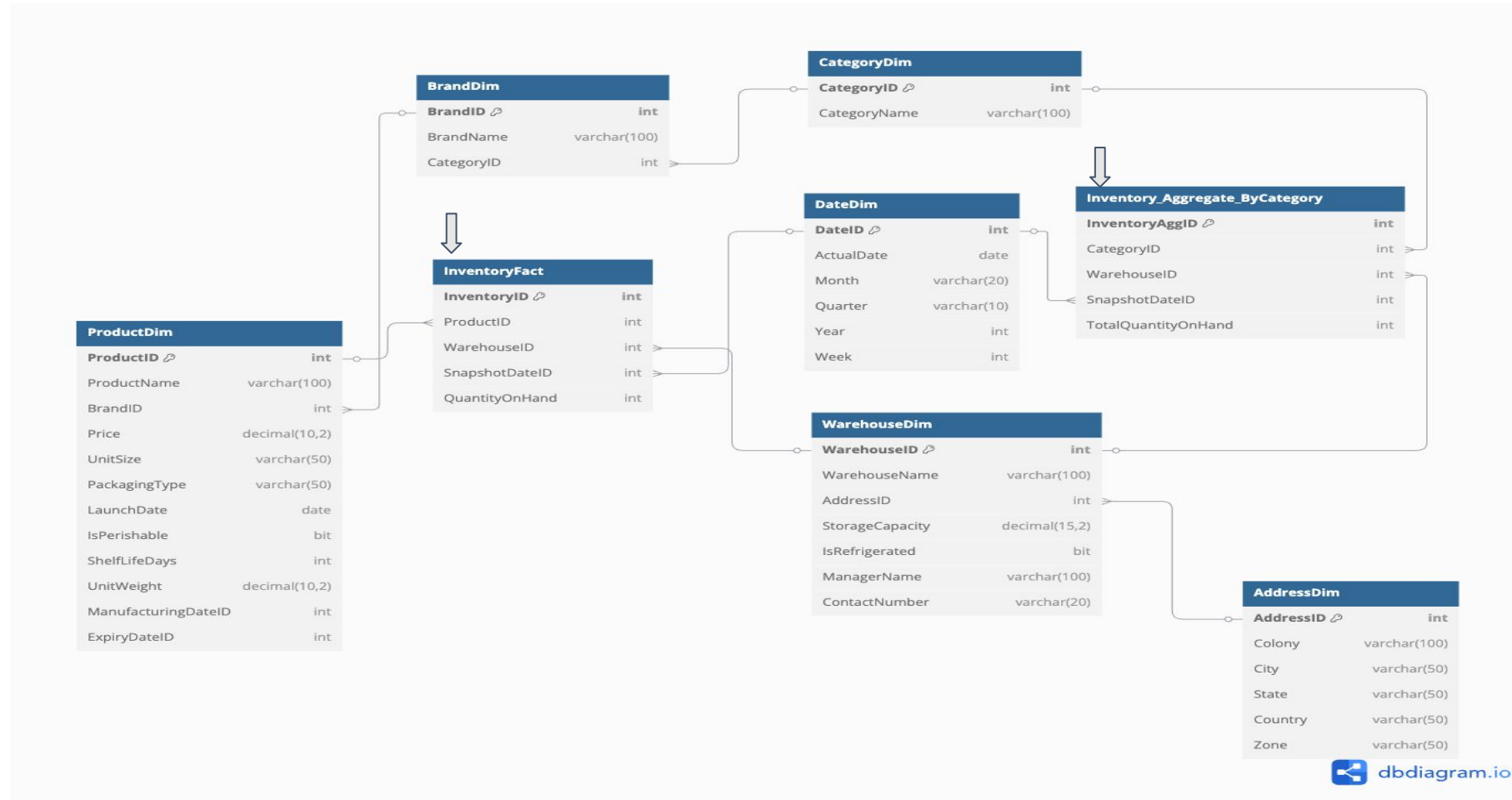
innovate

achieve

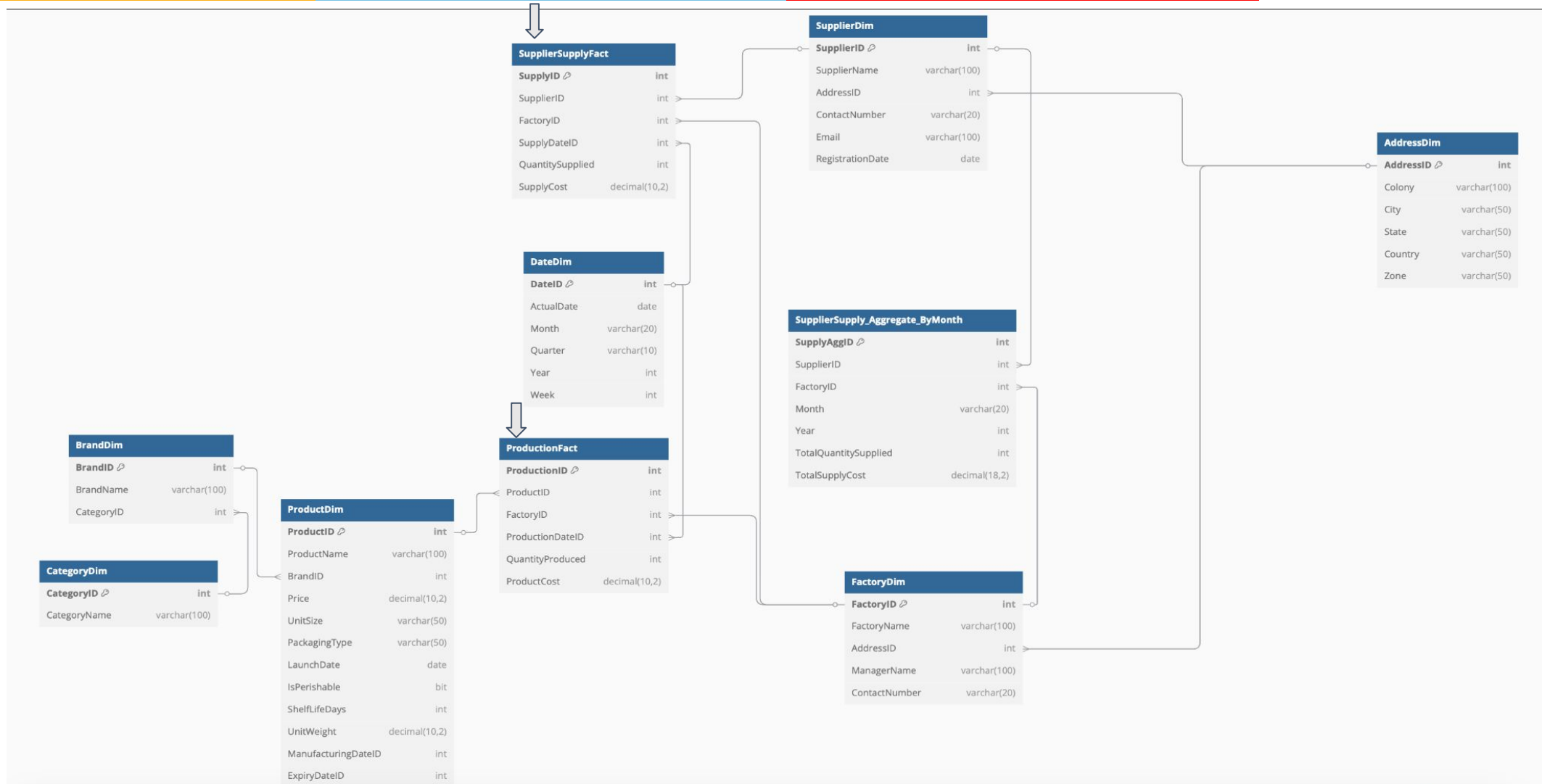
lead



Star Schemas (4 / 5)



Star Schemas (5 / 5)



Analysis Queries



- ❖ Total 20 queries have been implemented. In the following section Few of the queries.

Analysis Queries : Sales



Query 1 : Top 10 Amul Parlours wrt Total Sales

Objective : Identify the 10 parlours generating the highest sales.

```
SELECT TOP 10
    sp.ParlourName,
    a.City,
    SUM(sf.SalesAmount) AS TotalSales
FROM SalesFact sf
JOIN StoreParlourDim sp ON sf.ParlourID = sp.ParlourID
JOIN AddressDim a ON sp.AddressID = a.AddressID
GROUP BY sp.ParlourName, a.City
ORDER BY TotalSales DESC;
GO
```

	ParlourName	City	TotalSales
1	Amul Parlour 2	Bangalore	124592.36
2	Amul Parlour 146	Chennai	124418.62
3	Amul Parlour 25	Butwal	124061.94
4	Amul Parlour 156	Butwal	122482.92
5	Amul Parlour 128	Rangpur City	121488.20
6	Amul Parlour 158	Rangpur City	120870.15
7	Amul Parlour 31	Bhubaneswar	120821.84
8	Amul Parlour 94	Pokhara	120400.19
9	Amul Parlour 187	Coimbatore	119710.11
10	Amul Parlour 170	Kathmandu	119585.60

Analysis Queries : Sales



Query 2 : Top 5 Best-Selling Products

Objective : Find the products with the highest total sales amount.

```
SELECT TOP 5
    p.ProductName,
    SUM(sf.SalesAmount) AS TotalSales
FROM SalesFact sf
JOIN ProductDim p ON sf.ProductID = p.ProductID
GROUP BY p.ProductName
ORDER BY TotalSales DESC;
GO
```

	ProductName	TotalSales
1	Amul Fresh Dahi - Family Pack #2	43778.60
2	Amul Premium Paneer - 1kg #1	41706.60
3	Amul Kool Badam - Mini #2	41362.41
4	Amul Kool Caf+ - Pouch #3	39037.91
5	Amul Taaza - Cup #4	38996.04

Analysis Queries : Returns



Query 3 : Return Quantity Breakdown by Reason

Objective : Summarize the quantity of returns for each reason, categorized and described.

```
SELECT
    r.ReasonCategory,
    r.ReasonDescription,
    SUM(rf.QuantityReturned) AS TotalReturned
FROM ReturnsFact rf
JOIN ReasonDim r ON rf.ReasonID = r.ReasonID
GROUP BY r.ReasonCategory, r.ReasonDescription
ORDER BY TotalReturned DESC;
GO
```

	ReasonCategory ▾	ReasonDescription ▾	TotalReturned ▾
1	Return	Packaging damaged during delivery	1028
2	Inventory	Suspected contamination	976
3	Inventory	Product found defective on shelf	963
4	Return	Product had too short shelf life	948
5	Inventory	Product SKU no longer valid	917
6	Supply	Supplier delay in dispatch	902
7	Return	Incorrect product delivered	900
8	Supply	Order sent to wrong store	896
9	Return	Received an expired product	895
10	Supply	Delivery not received	894
11	Return	Seal broken at delivery	877
12	Return	Customer disliked the taste	859
13	Supply	Excess supply beyond requested qty	856
14	Inventory	Cold chain not maintained	843
15	Inventory	Product label incorrect	839

Analysis Queries : Product Enquiry



Query 4 : Top 5 Most Enquired Products

Objective : Identify which products received the most enquiries at parlours. Presuming that an enquiry in the parlours would be lodged only for out of stock or not available products and hence this analysis will Supports demand forecasting and promotional planning. Helps identify which products customers are most interested in

```
SELECT TOP 5
    p.ProductName,
    COUNT(*) AS EnquiryCount
FROM ProductEnquiryFact e
JOIN ProductDim p ON e.ProductID = p.ProductID
GROUP BY p.ProductName
ORDER BY EnquiryCount DESC;
GO
```

	ProductName	EnquiryCount
1	Amul Dairy Whitener - Tetra Pack #7	13
2	Amul Tiramisu Delight - 1L #8	12
3	Amul Kool Caf+ - Stick Pack #7	12
4	Amul Homogenized - Stick Pack #1	12
5	Amul Kitchen Ghee - Box #9	12

Analysis Queries : Supplier



Query 5 : Top 10 Suppliers with Least Supply-to-Cost Ratio

Objective : Identify the most efficient suppliers by analyzing how much they supply per unit cost. This ratio-based diagnostic report is useful for supplier benchmarking and renegotiating contracts with inefficient suppliers.

```
SELECT TOP 10
    s.SupplierName,
    fd.FactoryName,
    SUM(f.QuantitySupplied) AS TotalSupplied,
    SUM(f.SupplyCost) AS TotalCost,
    CAST(SUM(f.QuantitySupplied) * 1.0 / NULLIF(SUM(f.SupplyCost), 0) AS DECIMAL(10,4)) AS SupplyToCostRatio
FROM SupplierSupplyFact f
JOIN SupplierDim s ON f.SupplierID = s.SupplierID
JOIN FactoryDim fd ON f.FactoryID = fd.FactoryID
GROUP BY s.SupplierName, fd.FactoryName
ORDER BY SupplyToCostRatio ASC;
GO
```

	SupplierName	FactoryName	TotalSupplied	TotalCost	SupplyToCostRatio
1	Mother Milk Point	Amul Factory - Ahmedabad	125426	296718.19	0.4227
2	Hari Farms	Amul Factory - Gurgaon	119803	283030.01	0.4233
3	Gopal Agro	Amul Factory - Trivandrum	138872	327328.78	0.4243
4	Mother Farms	Amul Factory - Khulna City	138106	324645.21	0.4254
5	Madhav Dairy	Amul Factory - Bangalore	119963	281701.71	0.4259
6	Desi Milk Suppliers	Amul Factory - Kolkata	142880	335447.41	0.4259
7	Gopal Milk Point	Amul Factory - New Delhi	143724	337101.64	0.4264
8	Anand Milk Point	Amul Factory - Siliguri	142697	334121.39	0.4271
9	Kisan Farms	Amul Factory - Dhangadhi	129688	303199.45	0.4277
10	Shiv Farms	Amul Factory - Bhubaneswar	147542	344280.95	0.4286

Analysis Queries : Production



Query 6 : Most expensive products to produce for each category

Objective : Identify products with highest average production cost, per category. Assists in pricing strategy and product profitability analysis.

Can indicate if costlier SKUs need value engineering or revision.

```
SELECT
    p.ProductName,
    AVG(f.ProductCost) AS AverageProductionCost
FROM ProductionFact f
JOIN ProductDim p ON f.ProductID = p.ProductID
GROUP BY p.ProductName;
GO
```

	CategoryName	ProductName	AvgProductionCost
1	Butter	Amul Farmhouse Butter - Mini #10	1051.875238
2	Cheese	Amul Emmental - Tetra Pack #8	1122.864285
3	Cream	Amul Honey Yogurt - Glass Bottle #2	1045.831666
4	Ghee	Amul Daily Ghee - 1L #1	979.241000
5	Ice Cream	Amul Organic Paneer - Mini #5	1055.555294
6	Lassi	Amul Dairy Whitener - 250g #5	972.765333
7	Milk	Amul Homogenized - Jar #9	1018.506666
8	Milk Powder	Amul Stirred Dahi - Pouch #7	1040.208947
9	Paneer	Amul Rajbhog - Pouch #1	1077.399375
10	Yogurt	Amul Kool Chocolate - Cup #1	1070.443333

Analysis Queries : Distribution



Query 8 : To identify the most efficient parlours in each zone based on their Distribution-to-Sales Efficiency, calculated as the ratio of units sold to units distributed

Objective : Enables zone-wise efficiency comparison, supporting decisions related to promotional planning, stock allocation, and expansion.

```
BestParlours AS (
  SELECT
    r.Zone,
    r.ParlourID,
    r.SellThroughRatio
  FROM RatioByParlour r
  JOIN (
    SELECT Zone, MAX(CAST(TotalSold AS FLOAT) / NULLIF(TotalDistributed, 0)) AS MaxRatio
    FROM RatioByParlour
    GROUP BY Zone
  ) maxr
  ON r.Zone = maxr.Zone AND r.SellThroughRatio = maxr.MaxRatio
)
SELECT
  ad.Zone,
  sp.ParlourName,
  sd.DistributorName,
  ad.City
FROM BestParlours bp
JOIN StoreParlourDim sp ON bp.ParlourID = sp.ParlourID
JOIN StoreWholesalerDim sw ON sp.WholesalerID = sw.WholesalerID
JOIN StoreDistributorDim sd ON sw.DistributorID = sd.DistributorID
JOIN AddressDim ad ON sp.AddressID = ad.AddressID
ORDER BY ad.Zone;
```

```
WITH DistributionToParlour AS (
  SELECT
    Zone,
    ParlourName,
    DistributorName,
    City
  FROM
    SalesByParlour s
    JOIN DistributionToParlour d
    ON s.ParlourID = d.ParlourID
  WHERE
    s.TotalSold > 0
  ORDER BY
    Zone,
    ParlourName,
    DistributorName,
    City
)
SELECT
  Zone,
  ParlourName,
  DistributorName,
  City
FROM DistributionToParlour
```

	Zone	ParlourName	DistributorName	City
1	Central Nepal	Amul Parlour 183	Amul Distributor - Central Nepal	Kathmandu
2	East India	Amul Parlour 60	Amul Distributor - East India	Siliguri
3	Eastern Nepal	Amul Parlour 42	Amul Distributor - Eastern Nepal	Biratnagar
4	North Bangladesh	Amul Parlour 194	Amul Distributor - North Bangladesh	Rajshahi City
5	North India	Amul Parlour 165	Amul Distributor - North India	Amritsar
6	South Bangladesh	Amul Parlour 189	Amul Distributor - South Bangladesh	Barisal City
7	South India	Amul Parlour 41	Amul Distributor - South India	Mysore
8	West India	Amul Parlour 99	Amul Distributor - West India	Ahmedabad
9	Western Nepal	Amul Parlour 161	Amul Distributor - Western Nepal	Dhangadhi

Analysis Queries : Inventory



Query 9 : Warehouse Storage Capacity Analysis

Objective : Identify warehouses operating at or above 90% of their total storage capacity using current product weight stored. This diagnostic query helps detect high-utilization warehouses, enabling distribution rebalancing, overflow prevention, and operational risk reduction. Especially critical for perishable goods management.

```
SELECT
  w.WarehouseID,
  w.WarehouseName,
  w.StorageCapacity,
  SUM(i.QuantityOnHand * p.UnitWeight) AS TotalWeightStored,
  ROUND(100.0 * SUM(i.QuantityOnHand * p.UnitWeight) / w.StorageCapacity, 2) AS UtilizationPercent
FROM InventoryFact i
JOIN ProductDim p ON i.ProductID = p.ProductID
JOIN WarehouseDim w ON i.WarehouseID = w.WarehouseID
GROUP BY w.WarehouseID, w.WarehouseName, w.StorageCapacity
HAVING SUM(i.QuantityOnHand * p.UnitWeight) >= 0.9 * w.StorageCapacity;
GO
```

	WarehouseID	WarehouseName	StorageCapacity	TotalWeightStored	UtilizationPercent
1	29	Kolkata Central Warehouse	379251508.00	360288933.00	95.000000
2	9	Surat Central Warehouse	355732867.00	408673298.00	114.880000
3	3	Jaipur Central Warehouse	333170739.42	346497569.00	104.000000
4	32	Amritsar Central Warehouse	393797132.00	361236274.00	91.730000
5	26	Trivandrum Central Warehouse	381609604.00	371714023.00	97.410000
6	12	Coimbatore Central Warehouse	316514860.19	341836049.00	108.000000
7	6	Khulna City Central Warehouse	360708262.00	334349077.00	92.690000
8	21	Kochi Central Warehouse	342087938.00	376349307.00	110.020000
9	1	Rangpur City Central Warehouse	393434002.00	357602202.00	90.890000
1...	10	Bhubaneswar Central Warehouse	364912251.00	369370949.00	101.220000
1...	13	Barisal City Central Warehouse	387426724.00	373507084.00	96.410000
1...	5	Lalitpur Central Warehouse	349207624.00	347346353.00	99.470000

Thank you!

