Sri Lanka Institute of Information Technology

Data warehousing and Business Intelligence
(IT3021)

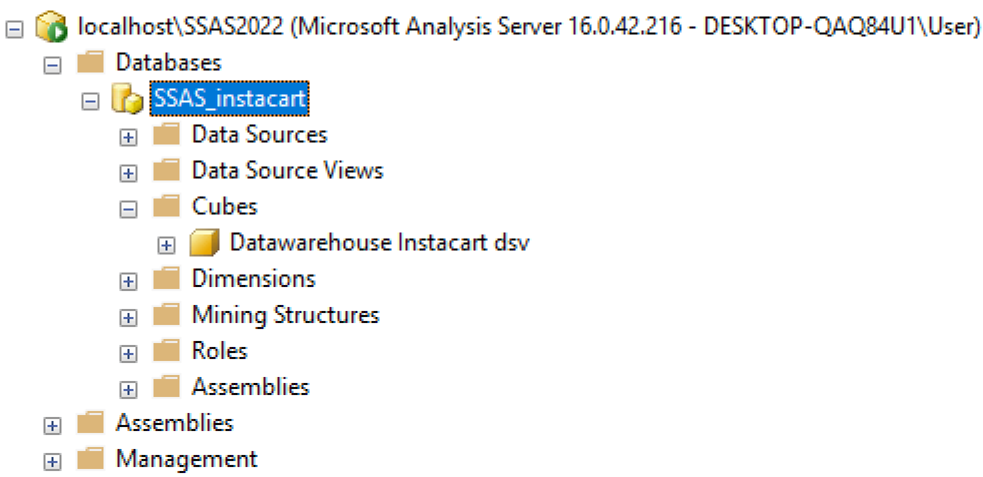
Continuous Assignment – 2025, Semester 1
Assignment 2

Completed By:
Himsara P V D
IT22367112

Table of Contents

1.Data source

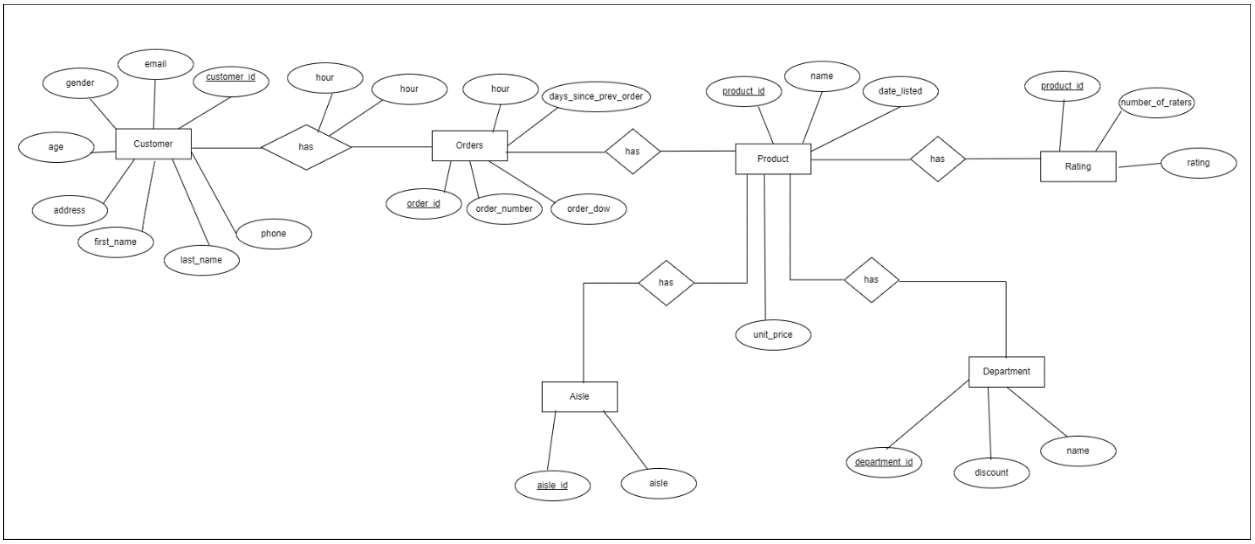
The source data warehouse (DS_Instacart_DW) was populated with transactional Instacart data that was transformed and loaded into a Snowflake schema. SQL Server Analysis Services (SSAS) was employed to build the cube on the basis of this carefully developed warehouse. Shown in the second photo is the following schema:

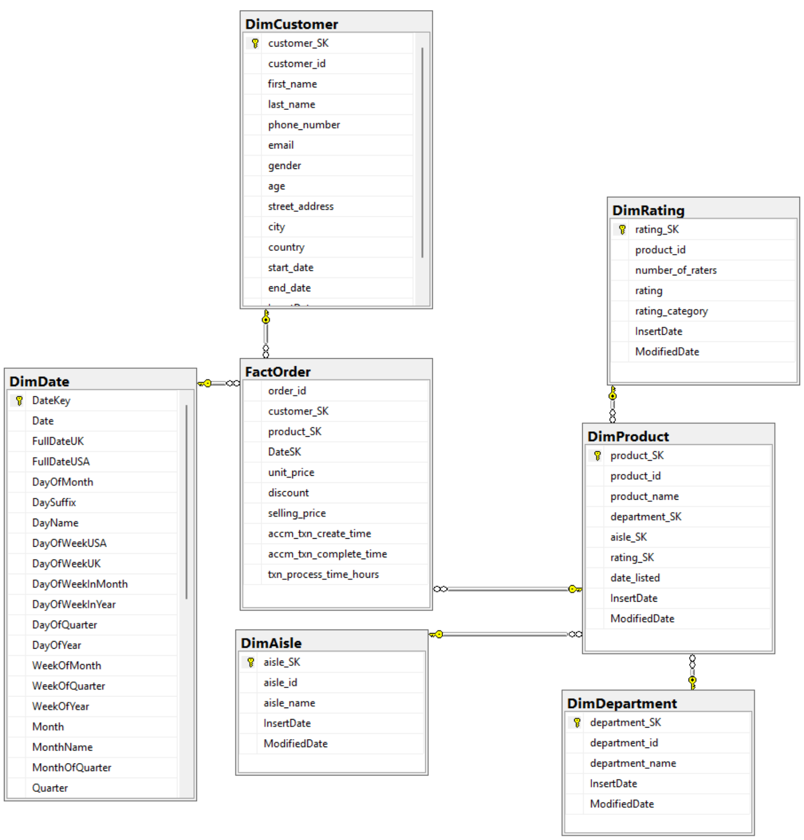


- FactOrders being the fact table containing transactional information like order_id, unit_price, discount, selling_price, and txn_process_time_hours
- Six dimensional tables:
  DimCustomer, DimProduct, DimRating, DimAisle, DimDepartment, and DimDate, all linked to the fact table through foreign key surrogate keys. These dimensions allow in-depth slicing and dicing of data from various viewpoints such as product categories, customer demographics, time, and ratings.
- The third picture demonstrates the Cube_Browser window in SSAS with the Cube_Instacart_DW being currently browsed. It contains
  An organized measure group for Fact Order that includes:
  - Order Id
  - Unit Price
  - Discount
  - Transaction Processing Time Hours
  - Fact Order Count

A KPI group of measures, e.g., KPI Fast Delivery, that is used to examine delivery efficiency per product
An example of a data grid displaying a list of products and their respective measures such as Txn Process Time Hours and KPI measures

This cube structure permits users to conduct OLAP operations such as aggregating measures and drilling down and slicing by dimensions (product or date) to reveal information such as
Which products offer the quickest processing time
How discounts and ratings affect ordering behaviour
Trends in delivery performance by product and department





2. SSAS Cube implementation

A data structure called an OLAP cube, also known as a hypercube or multidimensional cube, allows OLAP databases to do near-instantaneous data analysis.

The most significant parts of a cube are its dimensions and measurements.

- • Dimensions – These are the dimensions that come from the data source.
- • Measure group – This has a similar concept to the fact table of the data warehouse. Here all the measures of the OLAP cube are present.

For the creation of the new project SQL Sever Data Tools was used as below:

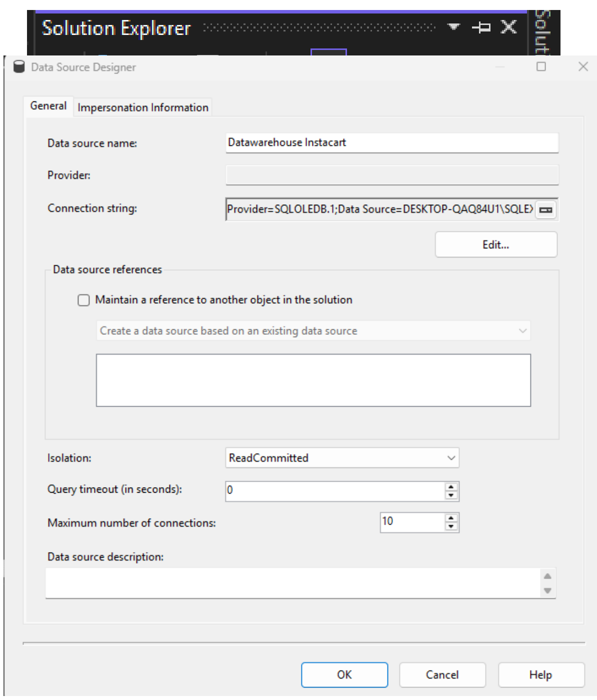
- **• Analysis Services -> Multidimensional -> Analysis Services Multidimensional and Data Mining Project**

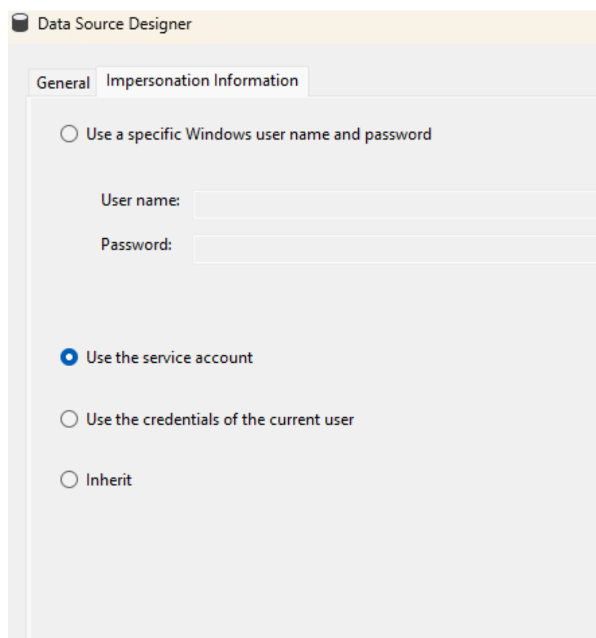
2.1 Cube Implementation

2.1.1 Creating the Data Source

In this scenario, a data warehouse built from the Instacart dataset serves as the data source. The source is defined in the cube project as DS_Instacart_DW, as shown in the Solution Explorer.

- The cube connects to the SQL Server data warehouse via DS_Instacart_DW using a connection string that includes the local SQL Server instance.
- The data source references the warehouse created from transformed Instacart transactional and dimensional data.





2.1.2 Creating the Data Source View

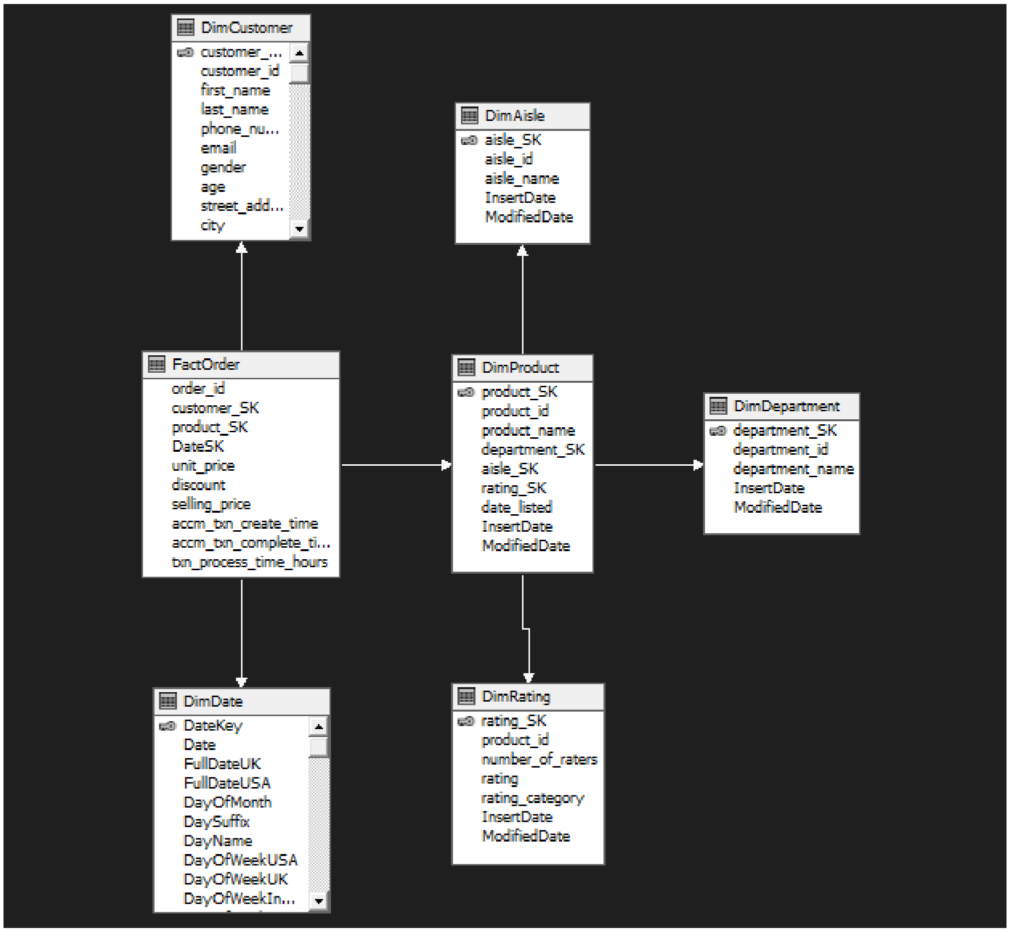
In SQL Server Analysis Services (SSAS), the Data Source View (DSV) acts as a logical layer that defines which tables from the data warehouse are visible to the cube. The analysis service can access only those tables that are included in the DSV. Hence, creating a DSV is a critical step before cube design.

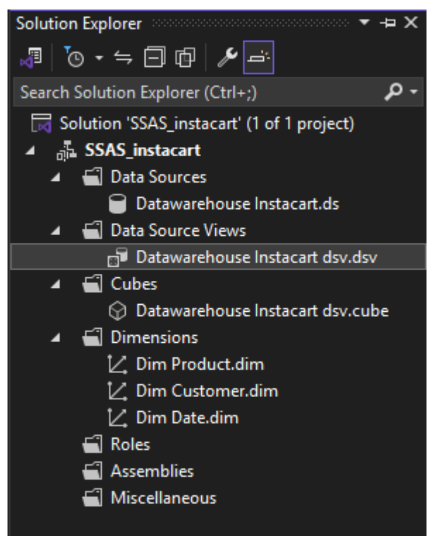
For the Instacart project, the DSV was created using the previously defined data source, DS_Instacart_DW. After selecting the required fact and dimension tables, the relationships were established based on foreign key references. A proper name was then assigned to the DSV: DSV_Instacart_DW.

The DSV includes the following key components:

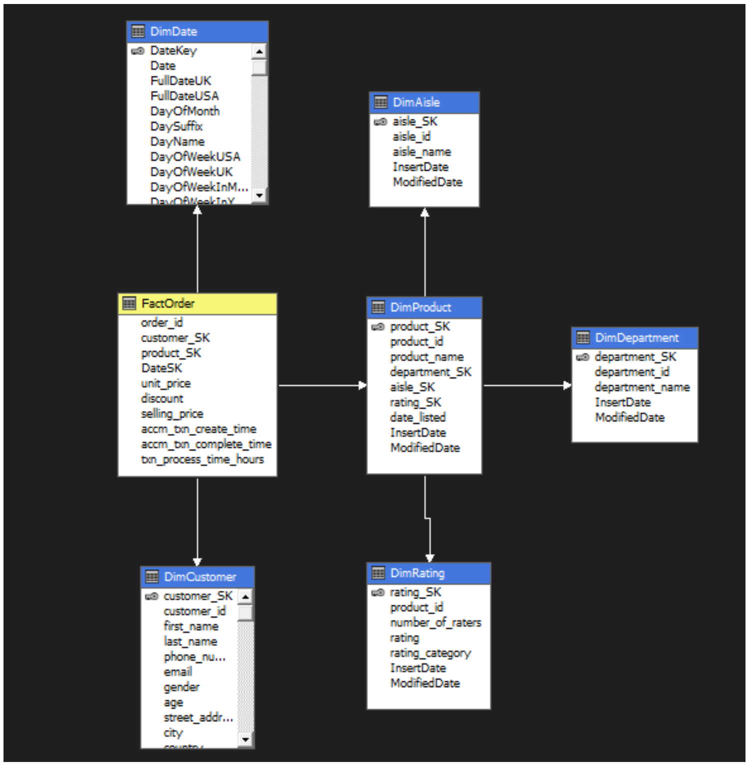
- Fact Table:
  FactOrder – Contains transactional details such as unit_price, selling_price, discount, and txn_process_time_hours.
- Dimension Tables:
  DimCustomer – Customer demographics and contact details
  DimProduct – Product metadata linked to DimAisle and DimDepartment
  DimRating – Product ratings and rating categories
  DimDate – Date-related attributes used for time-based slicing and filtering
  DimAisle and DimDepartment – Hierarchical classification of products

These tables are interconnected to form a Snowflake schema, with appropriate foreign key relationships clearly defined within the DSV.


DimCustomer
customer_...
customer_id
first_name
last_name
phone_nu...
email
gender
age
street_add...
city
DimAisle
aisle_SK
aisle_id
aisle_name
InsertDate
ModifiedDate
FactOrder
order_id
customer_SK
product_SK
DateSK
unit_price
discount
selling_price
accm_txn_create_time
accm_txn_complete_ti...
txn_process_time_hours
DimProduct
product_SK
product_id
product_name
department_SK
aisle_SK
rating_SK
date_listed
InsertDate
ModifiedDate
DimDepartment
department_SK
department_id
department_name
InsertDate
ModifiedDate
DimDate
DateKey
Date
FullDateUK
FullDateUSA
DayOfMonth
DaySuffix
DayName
DayOfWeekUSA
DayOfWeekUK
DayOfWeekIn...
DimRating
rating_SK
product_id
number_of_raters
rating
rating_category
InsertDate
ModifiedDate


Solution Explorer
Search Solution Explorer (Ctrl+;)
Solution 'SSAS_instacart' (1 of 1 project)
SSAS_instacart
Data Sources
Datawarehouse Instacart.ds
Data Source Views
Datawarehouse Instacart dsv.dsv
Cubes
Datawarehouse Instacart dsv.cube
Dimensions
Dim Product.dim
Dim Customer.dim
Dim Date.dim
Roles
Assemblies
Miscellaneous

2.1.3 Creating the Cube



Using the previously developed Data Source View, the cube was deployed via the Cube Wizard in SQL Server Data Tools (SSDT). The wizard walks you through the process of selecting the DSV, designating the fact table to be the measure group table and subsequently selecting the proper measures and dimensions.
The cube in this case was named Cube_Instacart_DW, and the fact table we chose was FactOrder with the most important business measures of unit_price, discount, selling_price, and txn_process_time_hours.

They added the following dimensions:
DimCustomer
DimProduct
DimDate
DimRating
DimAisle

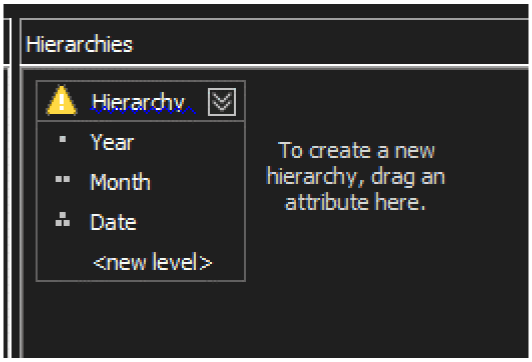
2.1.4 Creating Hierarchies and Dimension Structures

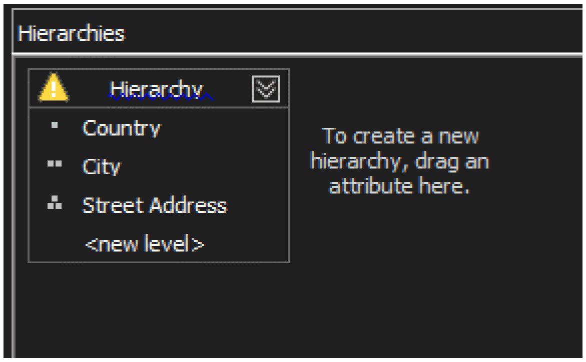
Once a cube structure had been developed, separate dimensions were seen in the Dimensions folder of the solution. Attributes in the Data Source View were moved into the Attributes pane of each of the dimensions.

Next, meaningful hierarchies were built by bringing multiple related attributes into the Hierarchy pane to enable advanced navigation and drill-down in the cube.

Examples from the implementation
• DimDate:
A hierarchy was established with levels
Year to Quarter to Month to Date
It facilitates time-series analysis of transaction metrics and performance across various time spans.
(See the "Dim Date" tab in the cube structure design window)

• DimCustomer
A geographic hierarchy was built with
Country to city to street address
Facilitates slicing customer data by region to measure purchase patterns per region.





2.1.5 Creating KPIs

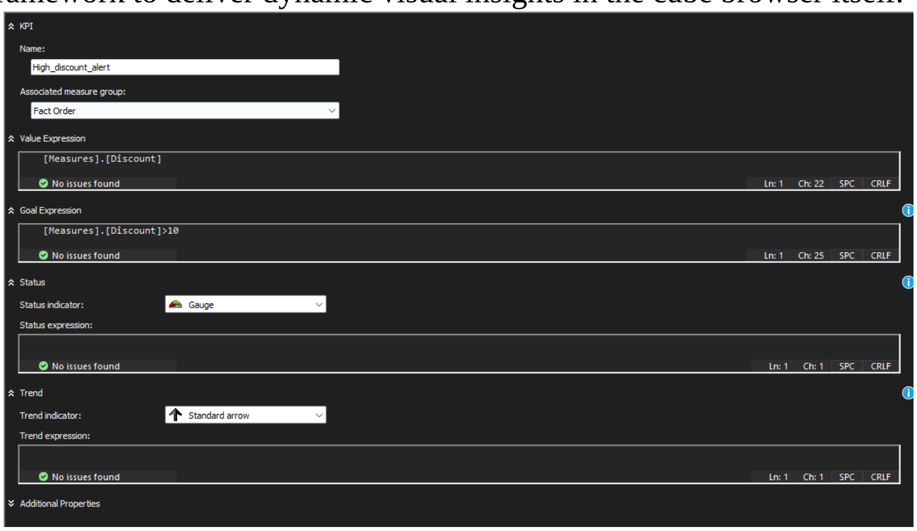
Key Performance Indicators (KPIs) are measurable values that demonstrate how effectively certain objectives are being achieved. KPIs in a cube environment offer a quick overview of key business metrics and allow decision-makers to track and monitor performance trends in real-time.

In this project, KPIs have been defined based on measures available in the Cube_Instacart_DW. Each KPI uses a Value Expression to compute the actual measure and a Goal Expression to set a threshold or target. A gauge-style status indicator is used to visually represent performance.

The following KPIs were implemented:

high_discount_alert

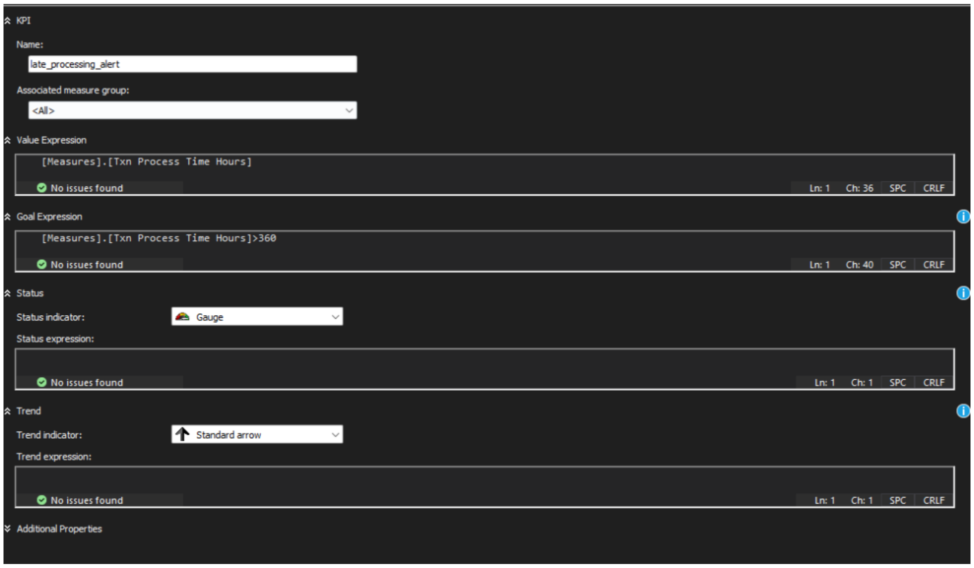
- This KPI measures the [Discount] measure.
- It alerts if the discount percentage exceeds 10%.
- Similar to the earlier KPI, it employs a gauge to depict the state and has a trend indicator to track the direction of discounting activity.
- It is helpful to track profitability risk associated with over-discounting and can be used to guide pricing or promotional decisions. Both of these KPIs are important in ensuring that service level expectations and revenue margins are safeguarded. They utilize SSAS's KPI framework to deliver dynamic visual insights in the cube browser itself.



late_processing_alert

- This KPI tracks the [Txn Process Time Hours] metric.
- It alerts upon processing time over 360 hours, showing operational delays.
- Performance is represented visually through a gauge indicator.

- This assists the stakeholders in recognizing transactions that are inordinately delayed to be addressed proactively.



3.Demonstration of OLAP Operations

OLAP stands for Online Analytical Processing. This enables easy understanding of data and easy handling of data in making important business decisions. OLAP is an integral part of business intelligence (BI) where it helps greatly in trend analysis and other data analysis functions from various perspectives .

There are 5 main OLAP operations:

1. Drill Down – Drilling down converts less detailed information into more detailed information. It's possible to accomplish so by working your way down the concept hierarchy.
2. Roll Up – This is the opposite of drilling down. This performs aggregations on the data cube. This can be performed by climbing up the concept hierarchy.
3. Slice– It takes a single dimension from the OLAP cube and turns it into a new sub-cube.
4. Dice – Here a sub cube is selected from the OLAP cube by selecting two or more dimensions
5. Pivot – This acts as a rotation operation, where the current view is rotated to get a new view.

Once all cube components—including dimensions, measure groups, and KPIs—had been finalized, the cube was deployed through SQL Server Data Tools (SSDT). The deployment process involved connecting to the designated Analysis Services server and publishing the cube structure for browsing and analysis.

According to the deployment summary, the following actions took place:

• The Instacart_SSAS database was processed

• The Cube_Instacart_DW was deployed and processed

• The Fact Order measure group was also processed

Roll up

Row Labels	Fact Order Count
\"Constant Comment\" Decaffeinated Black Tea Blend	
tea	
beverages	1
\"Darn Good\" Chili Mix	
soup broth bouillon	
canned goods	1
0 Calorie Strawberry Dragonfruit Water Beverage	
energy sports drinks	
beverages	2
0% Fat Black Cherry Greek Yogurt y	
yogurt	
dairy eggs	1
0% Fat Organic Greek Vanilla Yogurt	
yogurt	
dairy eggs	2
0% Fat Vanilla Greek Yogurt	
yogurt	
dairy eggs	1
0% Greek Strained Yogurt	
yogurt	
dairy eggs	36
0% Greek Yogurt Black Cherry on the Bottom	
yogurt	
dairy eggs	2
1 % Lowfat Milk	
milk	
dairy eggs	4
1 Apple + 1 Pear Fruit Bar	
energy granola bars	
snacks	1
1 Liter	
water seltzer sparkling water	
beverages	20
1 Ply Paper Towels	
paper goods	

Roll up | Drill down | Slice | Dice | Pivot

The **Roll Up** operation displayed here aggregates the **Fact Order Count** based on the **product hierarchy**, summarizing data across broader product categories like sub-aisles and aisles rather than analyzing individual products.

From the pivot table:
• **0% Greek Strained Yogurt** has the highest number of orders (36), followed by **1 Liter water seltzer sparkling water** (20), and **1% Lowfat Milk** (4).
• Other items such as **"Darn Good!" Chili Mix**, **"Constant Comment" Decaffeinated Black Tea Blend**, and **1 Apple + 1 Pear Fruit Bar** each have only **1** order, indicating relatively low demand or isolated purchases.
• The data is organized hierarchically—from product to subcategory (e.g., yogurt, milk, snacks) to department (e.g., dairy eggs, beverages, canned goods)—allowing users to see both fine and coarse levels of transactional summaries.

Drill down

The **Drill Down** operation shown here breaks down aggregated values into finer levels of detail, allowing deeper exploration of the data hierarchy. Instead of viewing summarized totals, users can analyze metrics at a more granular level.

In this case, the **Selling Price** has been drilled down by **Year** → **Month** → **Date**. Each row represents a specific time unit (likely a month within a year), revealing the total selling price for that period:

• For example, period **12** shows the highest selling price at **105,378.19**, followed closely by **4** with **102,655.16**, and **5** at **102,199.81**.

• Other periods such as **11 (99,039.57)** and **6 (94,213.38)** display lower selling activity in comparison.

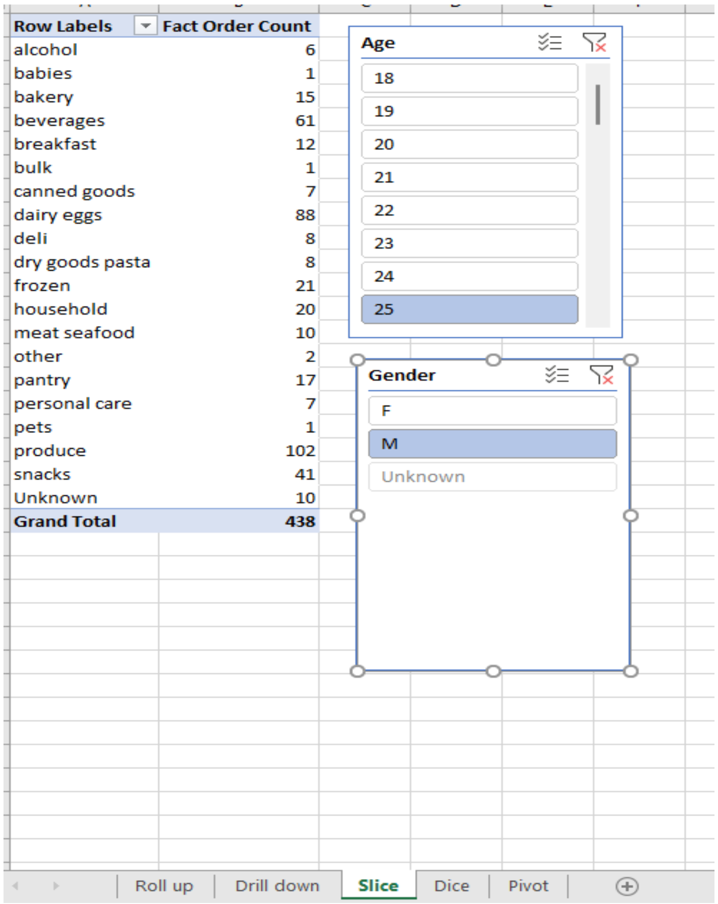
• The **Grand Total** of selling prices across all time periods is **1,197,572.82**.

This breakdown helps stakeholders analyze revenue trends over time, enabling better forecasting, promotional planning, and understanding of seasonal sales patterns.

Row Labels	Selling Price
1	99910.16
10	101732.4
11	99039.57
12	105378.19
2	95993.62
3	98607.44
4	102655.16
5	102199.81
6	94213.38
7	100369.89
8	100199.35
9	97273.85
Grand Total	**1197572.82**

Slice

The **Slice** operation helps isolate a single layer of data from the multidimensional cube, offering a focused view based on a specific dimension. In this case, slicing has been applied across different dimensions for targeted insights.



Row Labels	Fact Order Count
alcohol	6
babies	1
bakery	15
beverages	61
breakfast	12
bulk	1
canned goods	7
dairy eggs	88
deli	8
dry goods pasta	8
frozen	21
household	20
meat seafood	10
other	2
pantry	17
personal care	7
pets	1
produce	102
snacks	41
Unknown	10
Grand Total	**438**

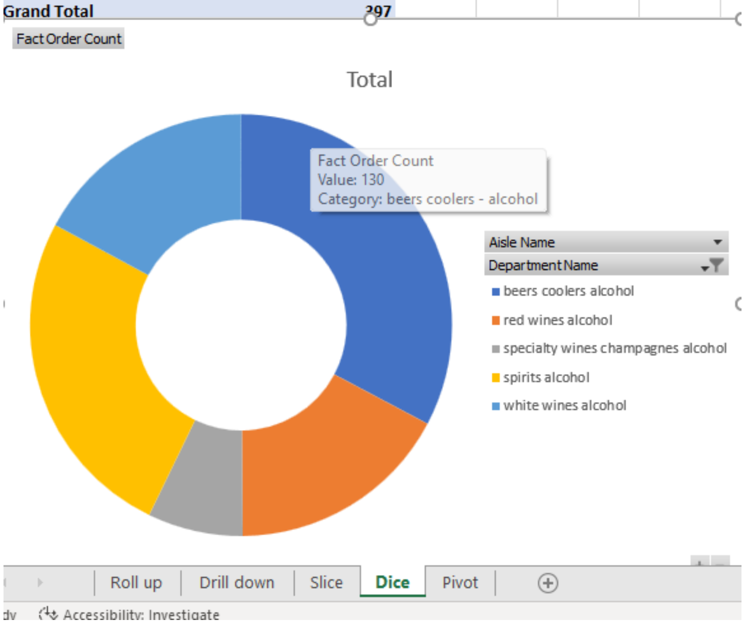
In the provided Excel pivot table:
• The data is sliced by **Age = 25** and **Gender = M**, using the slicers on the right.
• Once these slicers were applied, the **Fact Order Count** was displayed specifically for male customers aged 25 across all departments.
• For instance, departments like **produce (102 orders)**, **dairy eggs (88 orders)**, and **beverages (61 orders)** show strong purchase activity under the selected slice.
• Other departments such as **personal care (7 orders)** and **canned goods (7 orders)** show relatively lower activity, contributing to a **grand total of 438 orders**.

This slicing technique enables granular analysis of customer behavior within specific demographic segments. It is particularly valuable for targeted marketing, demand forecasting, and optimizing product assortments based on user profile segments.

Dice

The **Dice** operation is used to create a sub-cube by selecting specific values across multiple dimensions, allowing for multidimensional analysis of targeted data segments.

Row Labels	Fact Order Count
beers coolers	
alcohol	130
red wines	
alcohol	68
specialty wines champagnes	
alcohol	29
spirits	
alcohol	102
white wines	
alcohol	68
Grand Total	**297**



In this scenario:
• Dicing was applied across the **Department Name** and **Aisle Name** dimensions, focusing specifically on the **alcohol** department and its various aisles.
• The **Fact Order Count** measure was aggregated for each department-aisle combination, producing detailed insights such as:
o **beers coolers** (130 orders)
o **red wines** (68 orders)
o **spirits** (102 orders)
o **specialty wines champagnes** (29 orders)
o **white wines** (68 orders)
• A **donut chart** visualization was used to display the distribution of order counts across these combinations, offering an intuitive snapshot of segment performance.

This dicing approach helps stakeholders analyze specific product groupings within a department, aiding in category-level strategy, promotional planning, and inventory prioritization.

Pivot

The **Pivot** operation allows the sub-cube data to be rotated across different dimensional axes to provide alternative views and perspectives of the same measure. This is useful for comparing values by switching row and column fields in the pivot table.

In this scenario:

• The Pivot was applied on the **Fact Order Count** measure across the **Department Name** dimension.

• Rather than displaying each department as a vertical list, the departments—such as **alcohol, bakery, beverages, dairy eggs, frozen**, and more—have been pivoted into **columns**, with the Fact Order Count shown in a single horizontal row.

• This structure makes it easy to compare departmental performance side by side. For example, **produce** has the highest number of orders (**10,909**), followed by **dairy eggs (7,557)** and **beverages (5,354)**.

• In contrast, departments like **bulk (43)**, **meat seafood (22)**, and **missing (185)** show much lower order volumes.

• The **Grand Total** across all departments is **40,096 orders**, offering a comprehensive view of customer purchase activity.

This pivot technique enhances visual clarity, allowing stakeholders to quickly identify trends, outliers, and focus areas across product categories—similar to rotating perspectives in business intelligence dashboards.

	Column Labels										
	alcohol	babies	bakery	beverages	breakfast	bulk	canned goods	dairy eggs	deli	dry goods pasta	frozen
Fact Order Count	397	239	1362	5354	677	43	731	7557	1048	626	2285

	household	international	meat seafood	missing	other	pantry	personal care	pets	produce	snacks	Unknown	Grand Total
Fact Order Count	1188	212	904	185	58	1728	731	174	10909	2679	1009	40096

Roll up | Drill down | Slice | Dice | Pivot

4.PowerBI Reports

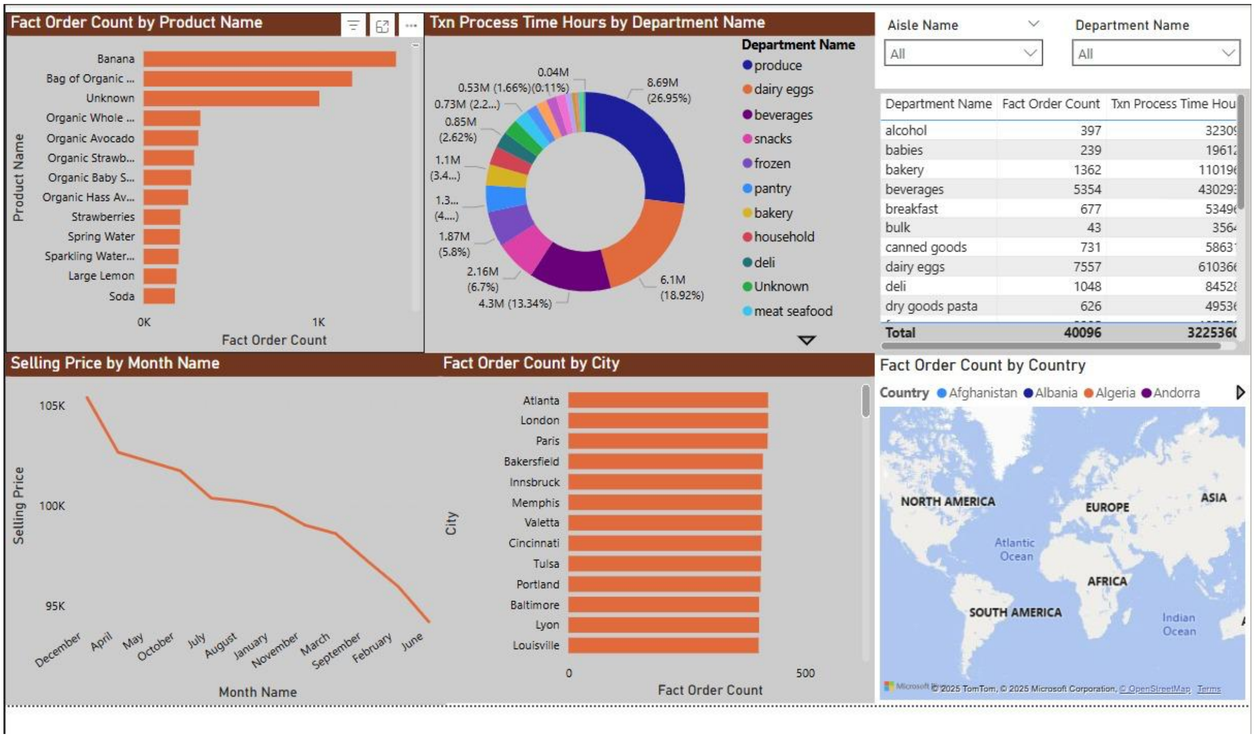
4.1 Report with Matrix Visualization

Objective:
To create a tabular view that presents detailed user information with categorical groupings.

Implementation:
• **Visual Used:** Matrix visual.
• **Data Fields:** The matrix displayed user data including First Name, Last Name, Email, Phone Number, and Country.
• **Grouping:** Rows were grouped by user names alphabetically, while columns were categorized by personal information fields (e.g., Email, Phone Number, Country).
• **Data Preparation:** The dataset was cleaned and organized to ensure uniform formatting of phone numbers and email addresses. Duplicate entries were removed, and missing values were addressed.
• **Modeling:** Data types were validated for consistency (e.g., phone numbers as text, email format validation), and the table was structured to support easy sorting and filtering by field..

Department Name	April	August	December	February	January	July	June	March	May	November	October	September	**Total**
alcohol	39	31	32	28	43	38	22	32	33	35	27	37	**397**
babies	23	22	27	19	15	16	28	25	13	13	23	15	**239**
bakery	129	95	117	113	98	130	108	100	129	107	121	115	**1362**
beverages	444	451	512	418	450	437	432	442	454	414	450	450	**5354**
breakfast	59	65	69	58	55	55	62	55	43	53	57	46	**677**
bulk	5	3	5	1	1	2	4	4	5	3	5	5	**43**
canned goods	61	63	60	57	49	64	45	64	72	68	64	64	**731**
dairy eggs	646	614	674	642	617	646	587	633	620	642	655	581	**7557**
deli	94	75	108	99	83	101	76	81	80	79	93	79	**1048**
dry goods pasta	47	63	54	55	60	58	42	51	41	42	49	64	**626**
frozen	169	194	188	153	191	188	183	180	237	200	205	197	**2285**
household	104	98	104	76	90	117	103	88	104	112	97	95	**1188**
international	17	20	16	15	20	25	7	16	21	23	21	11	**212**
meat seafood	64	82	87	61	84	75	66	65	80	77	83	80	**904**
missing	18	14	20	15	8	12	11	13	20	20	19	15	**185**
other	8	4	5	7	3		5	4	5	4	10	3	**58**
pantry	161	127	158	131	136	140	155	152	155	125	141	147	**1728**
personal care	76	66	64	60	58	57	49	68	57	58	51	67	**731**
pets	16	16	18	14	17	16	18	6	13	13	17	10	**174**
produce	938	934	889	876	942	930	855	944	904	900	919	878	**10909**
snacks	259	226	219	196	250	221	209	219	239	228	202	211	**2679**
Unknown	76	78	103	85	81	76	86	79	80	87	99	79	**1009**
Total	**3453**	**3341**	**3529**	**3179**	**3351**	**3404**	**3153**	**3321**	**3405**	**3303**	**3408**	**3249**	**40096**

4.2 Report with Slicers



Objective:
To deliver a comprehensive and interactive dashboard for analyzing sales performance, operational metrics, and product-level insights through visual storytelling and dynamic filtering.
Implementation:
• **Visuals Used:**
o *Bar charts* – Fact Order Count by Product Name, City, and Department.
o *Donut chart* – Transaction Process Time Hours by Department.
o *Line chart* – Selling Price Trend by Month Name.
o *Map visual* – Fact Order Count by Country.
o *Table visual* – Combined Fact Order Count and Transaction Hours by Department.
• **Slicers Added:**
o Aisle Name and Department Name slicers enabled user-driven filtering.
o Implemented interactivity allowing cascading filters — selecting a Department filters the Aisle options accordingly.
• **Data Modeling:**
o Relationships were established between Fact and Dimension tables to ensure accurate cross-filtering among visuals.
o Data formatting and categorization were applied to support meaningful aggregations and insights.

4.3 Report with Drill Down

(Country>City>Street Address)

Objective:
To enable users to explore sales trends over time by drilling down from *country to city to street address*.

Implementation:

• **Visual Used**: Bar Chart

• **Hierarchy Setup**:

o Created a country hierarchy in the *dimCustomer* table: Country→ City→ Street Address.

o Enabled drill-down and drill-up options in the chart.

• **Insights Gained**:

o Identified declining order trends over time.

• **Data Modeling**: Ensured date column was marked as a date table to support time intelligence functions.

4.4 Report with drill through

Objective:
To let users right-click on a high-level visual (e.g., by department or city) and navigate to a detailed view.
Implementation:

• Drill-Through Page:

o Created a new page to show detailed product-level data: *First Name,Last Name, Phone Number*

• Setup:

o Added drill-through filters on *Country vs Fact Order Count*

o Users can right-click on any visual in the main dashboard (e.g., map or bar chart) to drill into this detail.

• Visuals Used: Table view

First Name	Last Name	Email	Phone Number	Country
Abbey	Owens	Abbey_Owens619@gmail.com	2-204-272-0317	Mali
Abbey	Shields	Abbey_Shields8804@naiker.biz	2-520-337-6062	Cambodia
Abdul	Avery	Abdul_Avery5872@dionrab.com	1-111-868-0106	Sudan, Sout
Abdul	Bristow	Abdul_Bristow4970@bulaffy.com	4-768-031-8766	Dominica
Abdul	Haines	Abdul_Haines5860@grannar.com	6-165-410-8680	New Zealan
Abdul	Lindop	Abdul_Lindop1798@famism.biz	0-823-688-2500	Mauritania
Abdul	Simpson	Abdul_Simpson4577@nickia.com	3-305-500-2844	Tajikistan
Ada	Bishop	Ada_Bishop4963@liret.org	4-182-124-1146	Indonesia
Adalie	Waterson	Adalie_Waterson9711@bretoux.com	6-727-616-1238	El Salvador
Adela	Andrews	Adela_Andrews9248@womeona.net	0-712-254-6835	Micronesia
Adela	Boden	Adela_Boden6177@extex.org	0-623-356-1516	Malaysia
Aeris	Welsch	Aeris_Welsch7102@zorer.org	5-347-067-0534	San Marino
Aiden	Mitchell	Aiden_Mitchell6024@ubusive.com	4-317-304-1817	New Zealan
Aiden	Taylor	Aiden_Taylor7556@iatim.tech	0-045-287-5628	Croatia
Aileen	Campbell	Aileen_Campbell5931@corti.com	6-343-136-3215	Colombia
Alan	Bradshaw	Alan_Bradshaw7513@acrit.org	7-731-340-5608	Uganda
Alan	Bright	Alan_Bright8339@cispeto.com	5-655-845-5288	Nicaragua
Alan	Bullock	Alan_Bullock2532@twace.org	0-887-321-0500	Norway
Alan	May	Alan_May4455@elnee.tech	3-882-366-0684	Vietnam
Alan	Power	Alan_Power2723@tonsy.org	2-352-411-1787	Tonga
Alba	Harvey	Alba_Harvey8703@joiniaa.com	4-783-604-5080	Georgia
Aleksandra	Doherty	Aleksandra_Doherty6353@womeona.net	3-831-378-6081	Dominican F
Aleksandra	Norman	Aleksandra_Norman6568@nickia.com	2-680-142-3707	Egypt
Aleksandra	Owen	Aleksandra_Owen3228@fuliss.net	4-444-835-2338	Ukraine
Alessandra	Larsen	Alessandra_Larsen2585@mafthy.com	8-667-270-4005	Cameroon
Alessia	Craig	Alessia_Craig3229@supunk.biz	0-206-116-3252	Barbados
Alessia	John	Alessia_John3038@ubusive.com	6-488-877-1135	Congo, Rept
Alessia	Tindall	Alessia_Tindall8561@gmail.com	3-415-157-1713	Mexico