

MSA Data Dictionary

Capstone Project by
Group A

Team Members

1. Kavit Mehta (Product Manager)
2. Rohith Gowda Ranganatha (Data Engineer)
3. Yue Zhang (Operations)
4. Ziyan Liu (UI/UX)
5. Zisheng Lu (Data Analyst)

Agenda

01

Introduction & Problem Statement

02

Our Challenges & Decision-Making Journey

03

Technology Roadmap

04

Process Flow Documentation

05

Future Work



MSA Professional Services

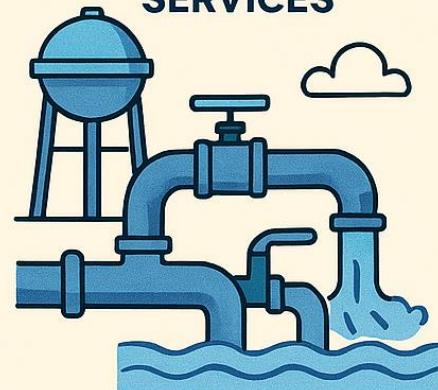
AGRICULTURAL SERVICE PLANNING



CONSTRUCTION MANAGEMENT



WATER DISTRIBUTION SERVICES



PARKS & REC



Who are Mid-State Associates? They are data-driven solution experts leading sustainable development.

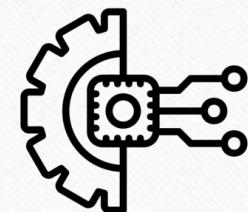
Pain Points – A real need for business intelligence



Scattered ERP Data



Missing Column/Table definitions



Unreaped business value from
AI and automation



Missing Metadata Documentation



Overworked Analysts due to data ambiguity

Problem Statements:

1. How can we extract metadata out of existing data, and automate future documentation?

2. How can we enrich the metadata with business definitions, costing minimum effort to knowledge owners?

Challenges For Us

- No access to client system and technology
- Synthesized data led to reduced semantic awareness about business definitions
- Understanding the intersect of data engineering and automation tools
- Learning from scratch tools like Fabric, Copilot Studio, Power Automate
- Constraints due to “trial” versions and limited access to complete tool features
- Ambiguity in dictionary end-point made metadata visualization harder

What shaped our decisions

1. Knowing the starting point of raw ERP data – a Fabric Lakehouse.
2. Partner wanted pure SQL data types for columns – which are readily extracted within a Fabric Warehouse.
3. Knowing the end point of data dictionary – FreshService ticketing platform.
4. We needed a universal distribution point to visualize our metadata outside of Fabric – a Fabric SQL Database.
5. Our biggest challenge – get knowledge owners to fill in missing business definitions.

Questions we asked ourselves

1. We can build an ETL (Extract-Transform-Load data) pipeline in Fabric that can build a semi-complete data dictionary, but even the partner can do that easily. Is that the real problem we are solving here?
2. As a busy knowledge owner, why would I spend my time looking at lengthy forms and spreadsheets just to contribute to a documentation process I know little about?

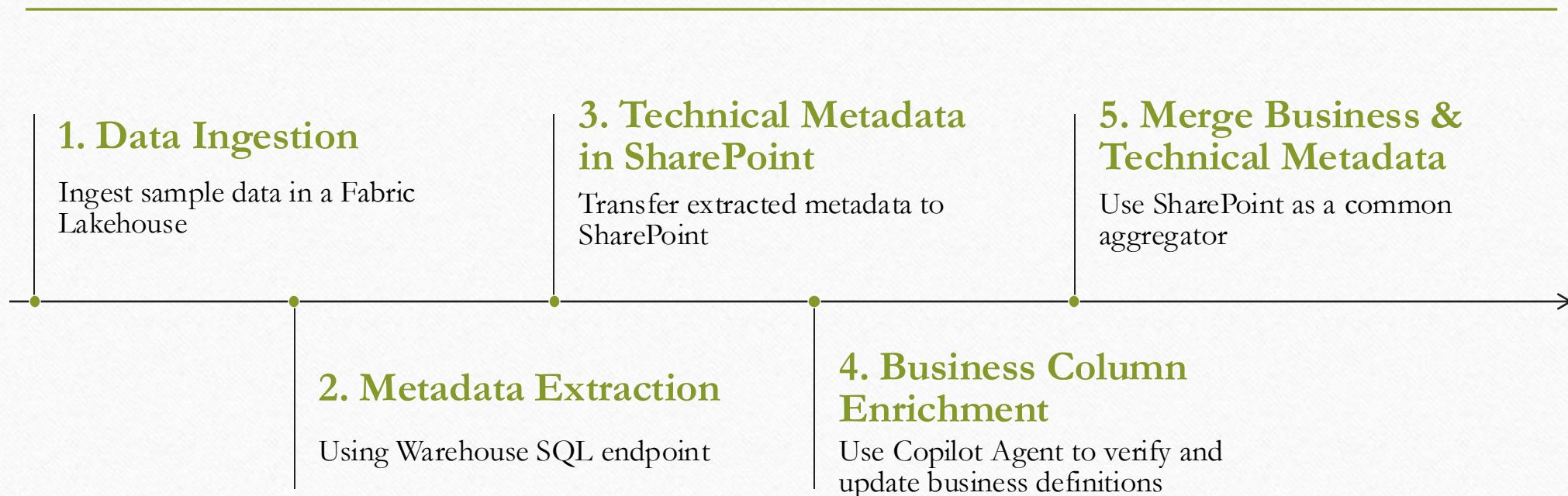
Questions we asked ourselves

3. We can build a simple Microsoft Forms using Power Apps to visualize the metadata and gather business definitions, but where is the innovation in that?
4. Do we focus on building an app that somewhat appeals to knowledge owners, or we build them a knowledge assistant - an AI agent - that will smartly do the heavy lifting for them?
5. If we want to set ourselves apart from bringing a generic solution, should we stop because of a few technological barriers?
6. What are we ultimately learning from this capstone?

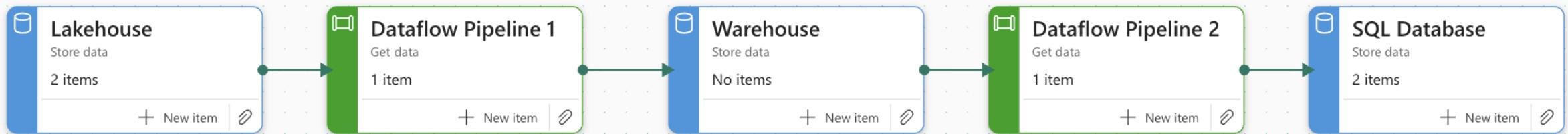
Technology Roadmap

- **Microsoft Fabric** – to extract, transform, and load technical (SQL) metadata
- **Microsoft Copilot Studio** – to build a semantic agent for business definition enrichment
- **Power Automate** – to orchestrate flow between Fabric and our AI agent
- **SharePoint** – to combine technical and business metadata forming a complete data dictionary

Process Flow Documentation



1. Microsoft Fabric



The complete data flow within Fabric.

Lakehouse (Raw Data) ➔ Warehouse (Clean Metadata) ➔ SQL Database (Distribution point)

Fabric  Warehouse  SQL Database  Lakehouse  Dataflow 1  Trials activated: 10 days left 

Power Query

Home Transform Add column View Help

Search (Alt + Q)

Save & validation runs  Check recent runs  Recent  Get data  Recent sources  Enter data  Manage connections  Default data destination  Options  Manage parameters  Query  Manage columns  Reduce rows  Sort  Transform  Combine  Copilot  Export template  Insights  Share 

Queries [44] 

-  vw_actualesales
-  vw_actualdata
-  vw_actualesales (2)
-  vw_annualdirectwages
-  vw_annualrevenue
-  vw_budget
-  vw_budgetsales
-  vw_cashreceipts
-  vw_ceashboarddata
-  vw_clientdata
-  vw_clientliason
-  vw_contractagreement
-  vw_customer
-  vw_customertransactions
-  vw_customertransactionsettlement
-  vw_dataarea
-  vw_datefinancial
-  vw_employelabor
-  vw_financialdimension

1 let
2 Source = Lakehouse.Contents(null),
3 Navigation = Source{[workspaceId = "1faab650-f75a-43fb-b795-55aa4ad03b07"]}[Data],
4 #"Navigation 1" = Navigation{[lakehouseId = "ccf36406-ec1c-419c-ad32-3754206e8921"]}[Data],
5 #"Navigation 2" = #"Navigation 1"{[Id = "vw_actualesales", ItemKind = "Table"]}[Data]

A _C PKFactTActualSales	A _C FKActivity	A _C FKDataArea	A _C FKProject	A _C FKDateFinancial	A _C FKProjectFundid
1 acy 0421927239	jxs 3	rrv	nxx D18509599.09.114...	otr 84918710	wvz 5036423991
2 xbw 2735836992	jxs 3	rrv	oqr X99855799.94.00...	otr 84918710	wvz 5036423991
3 wkv 0053831149	jxs 3	rrv	cmb C32855548.84.3...	otr 84918710	wvz 5036423991
4 dyu 2832005408	jxs 3	rrv	dnu O55018069.29.06...	otr 84918710	wvz 5036423991
5 dpn 2597241143	jxs 3	rrv	tge X36456904.24.15...	otr 84918710	wvz 5036423991

Fabric lets you create dataflow channels with smart transformation tools to move data between a lakehouse, a warehouse, and a SQL database. You just have to ensure there is a source, and a destination for your data.

Query settings 

Properties 

Name: vw_actualesales

Applied steps 

Source  Navigation  Navigation 1  X Navigation 2 

Data destination 

X Warehouse 

The screenshot shows the Fabric Data Platform interface. The top navigation bar includes links for Fabric, Warehouse (selected), SQL Database, Lakehouse, and Dataflow 1, along with a search bar and user information. The main menu has options for Home, Management, and Help, with a Share button on the right.

In the left sidebar, the **Explorer** pane is open, showing a tree structure of the warehouse schema:

- + Warehouses
- Warehouse
 - Schemas
 - DataDictionary
 - Tables
 - ColumnMetadata
 - Stg_ColumnMetadataSrc
 - Stg_TableMetadataSrc
 - TableMetadata
 - Views
 - Functions
 - Stored Procedures
 - RefreshColumnMetadata
 - RefreshTableMetadata

The central pane is the **Query Editor**, titled "Refresh Metadata". It contains the following SQL code:

```
1 EXEC DataDictionary.RefreshTableMetadata;
2 EXEC DataDictionary.RefreshColumnMetadata;
```

A callout box highlights the stored procedures:

We created “stored procedures” within our Fabric Warehouse, that let you refresh your entire metadata at a single click – Just run the above query to ingest any new metadata while keeping the previously documented data intact.

Refresh Metadata ColumnMetadata X TableMetadata

Data preview - ColumnMetadata Showing 1000 rows Search

	schema_name	table_name	column_name	column_defi...	data_type	is_nullable	column_defa...	notes
1	dbo	vw_generalledger	FKFinancialDimen...	NULL	bigint	YES	NULL	NULL
2	dbo	vw_salesmanager	PKSalesManager	NULL	bigint	YES	NULL	NULL
3	dbo	vw_generalledger	FKMainAccount	NULL	bigint	YES	NULL	NULL
4	dbo	vw_generalledger	MainAccount	NULL	bigint	YES	NULL	NULL
5	dbo	vw_fnbihcmempl...	Recid_	NULL	bigint	YES	NULL	NULL

A semi-complete data dictionary is ready, with just the column definitions and table definitions remaining to be filled.

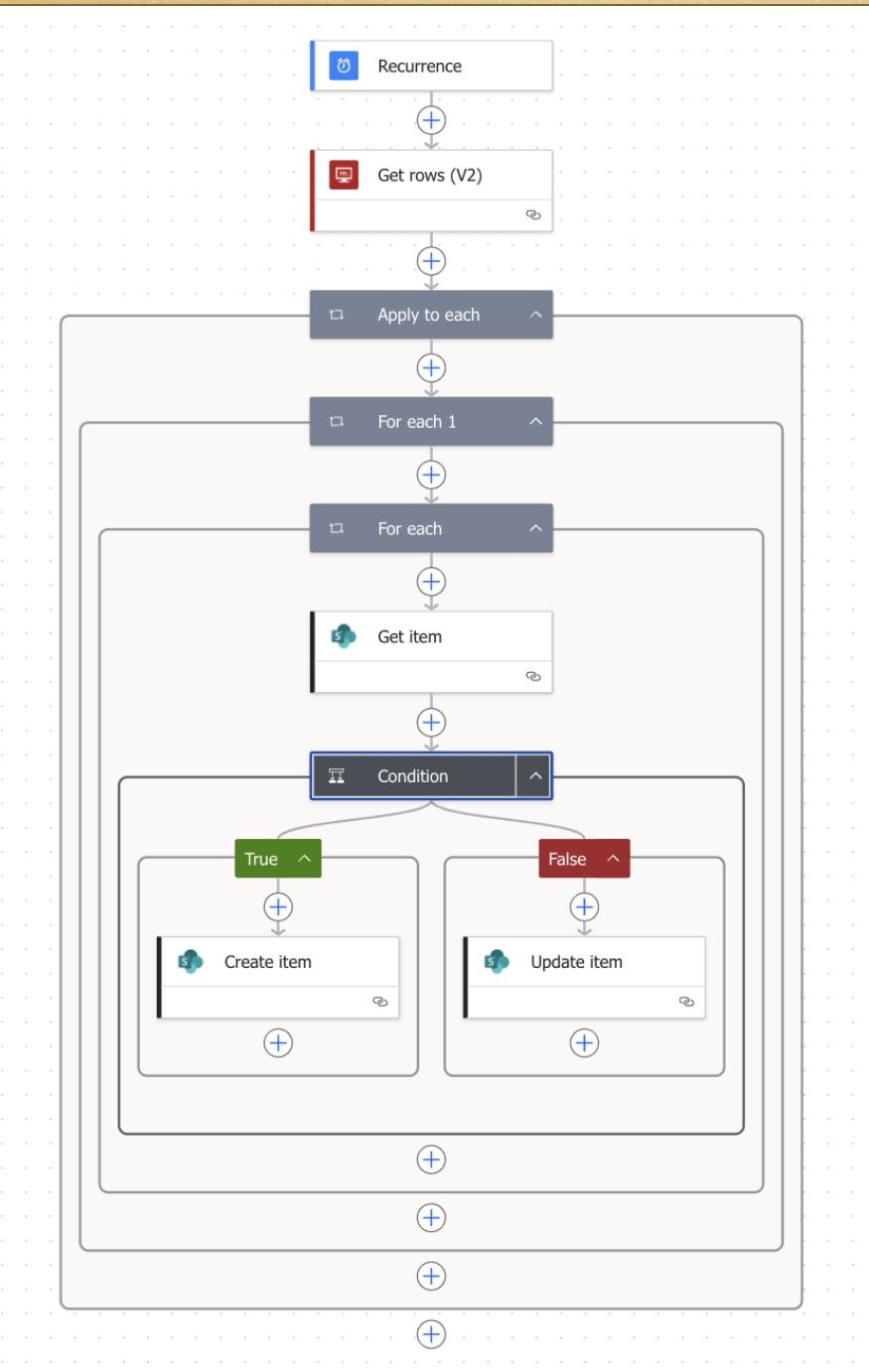
Refresh Metadata ColumnMetadata X TableMetadata

Data preview - TableMetadata Showing 1000 rows Search

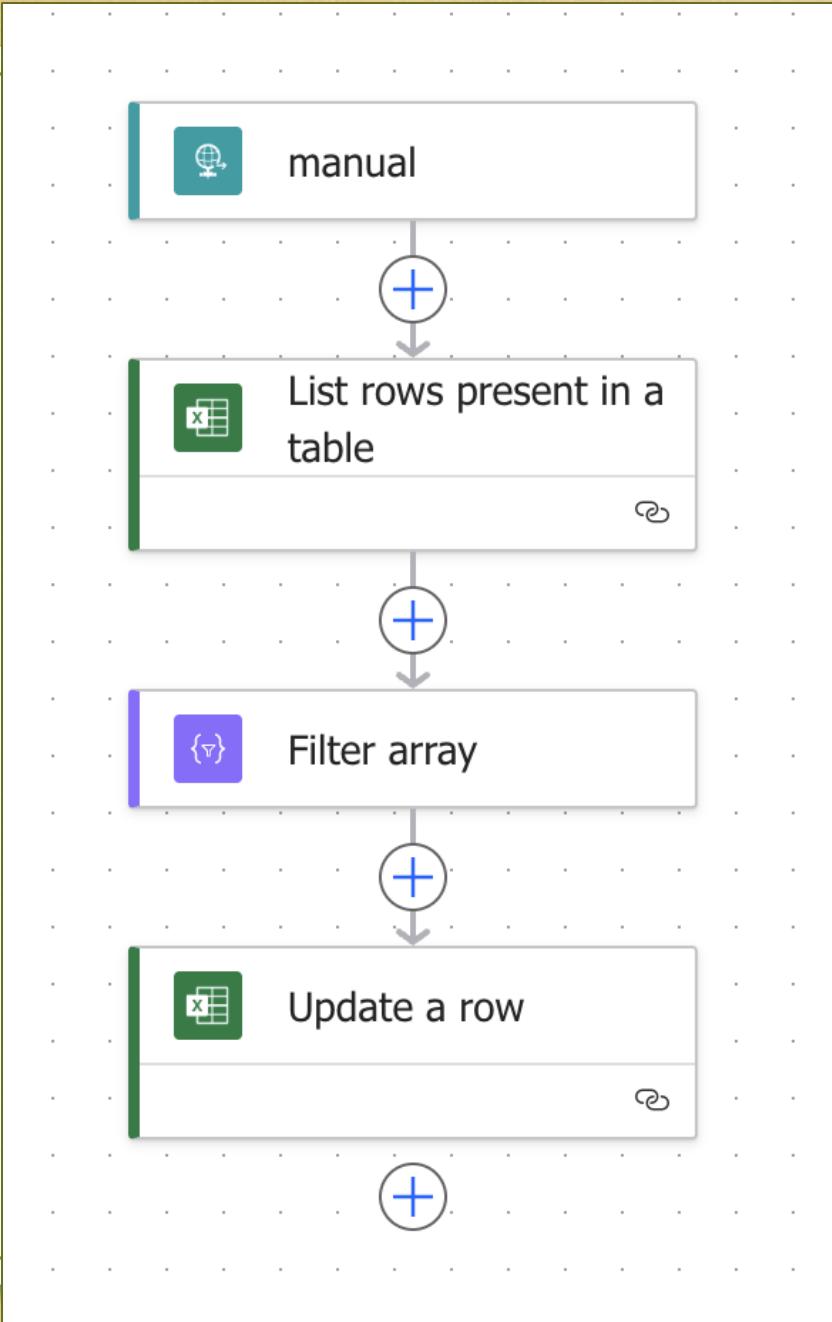
	schema_name	table_name	table_definition	notes
1	dbo	vw_employeelabor	NULL	NULL
2	dbo	vw_fnbihcmemployment	NULL	NULL
3	dbo	vw_unapprovedtimesheetlines	NULL	NULL
4	dbo	vw_projectmanager	NULL	NULL
5	dbo	vw_project	NULL	NULL

2. Power Automate

- Power Automate lets you create data flows to internally form a chain of command between different Microsoft 365 tools.
- We needed two Power Automate flows:
 1. Fabric to SharePoint - Runs on a schedule, pulls from Fabric, and **only updates technical fields** for matching rows or creates them if they don't exist yet.
 2. Copilot Agent to SharePoint - Triggered by “When an item is created/modified” in the list, writing **business** columns (definitions, notes) based on Copilot input.



- Flow #1: Fabric to SharePoint
- Lets you set an automatic recurrence based on your desired frequency and business needs
- Allows connecting SQL Database to SharePoint list
- Allows conditioning to add/modify new or existing rows



- Flow #2: Copilot Agent to SharePoint
- Lets the agent trigger the flow on command (here we have set a manual trigger, because we still couldn't get permission to integrate Copilot agent with other Microsoft apps while using the .wisc domain)

3. Copilot Studio Agent

- We built a fully functioning Copilot Studio Agent that lets users:
 - Ask for table/column definitions within the chat
 - Update said definitions from the same chat
- There is no need to scroll through a form.
- This agent can be integrated in any of the partner's Microsoft channels. Users can access it at a single click whether they are in MS Teams, or Outlook, or any other platform.

m365.cloud.microsoft/chat/?auth=2&home=1

Star 🔍 R New Chrome available :

M365 Copilot

Copilot Studio My agents > Metadata Documentation Agent

Last updated December 5, 2025 at 3:22:13 AM

Update Share ...

Search

Chat

Agents

Metadata Documentation Agent

Snowflake Cortex Agents

All agents

Create agent

Conversations

Library

Create

Apps

Rohith Gowda Rangan...

Copilot Chat

Describe Configure

Name: Metadata Documentation Agent

Description: Helps employees look up and update column definitions.

Instructions

Metadata Documentation Assistant — Smart & Trusted Behavior Instructions

Primary Purpose: Help employees find and update metadata stored in the MetadataDictionary SharePoint List, with focus on definition updates.

Main Workflow

When the user enters a column name (even misspelled or unclear):
Automatically interpret the intended column
Perform fuzzy matching only within MetadataDictionary values
Match to real existing column names — never invented ones

Agent preview

New chat

Metadata Documentation Agent

Helps employees look up and update column definitions.

- One simply needs to assign the agent instructions in natural language – Prompt Engineering
- Allows you to link the source i.e., the SharePoint list containing business definitions for read/write operations via Power Automate

AI-generated content may be incorrect | Copilot Studio capabilities in M365 are processed by the Copilot Studio service. By using this experience, you agree to these [Terms](#) | [Privacy and cookies](#)



Metadata Documentation Agent > Hi



Hi



Metadata Documentation Agent

Hello! 🌟 How can I assist you today? Are you looking up a column definition or planning to update one?



look up column definition for actual sales amount



Metadata Documentation Agent

The closest match in the DataDictionary for “actual sales amount” is:

Column Name: ActualSalesAmount

Current Definition: *“Represents the total sales amount recorded for a given transaction or reporting period.”*

Table Name: SalesSummary

Would you like to **update this information?**

yes, update the definition

show more details for ActualSalesAmount



Message Copilot



AI-generated content may be incorrect

Future Work

1. Create and Integrate Copilot Agent into Microsoft 365 Ecosystem
2. Configure Power Automate triggers with agent
3. Encourage stakeholders to use Data Dictionary chatbot by showcasing ease of use
4. Bring aggregated SharePoint file to FreshService as a ground-truth source
5. Provide feedback/difficulties in actual implementation

Thank You

It was a pleasure working with you.