ai\_assistant\_usage\_student\_life

**Table of Contents**

[**1.** **Data Overview** 2](#_Toc204170660)

[**2.** **Logical & Physical Model** 3](#_Toc204170661)

[**2.1 Logical Model** 3](#_Toc204170662)

[**2.2 Physical Model** 4](#_Toc204170663)

[**3.** **Lake House Overview** 6](#_Toc204170664)

[**4.** **Ware House Overview** 11](#_Toc204170665)

[**5.** **Power BI Overview** 14](#_Toc204170666)

# **Data Overview**

This dataset simulates 10,000sessions of students interacting with an AI assistant (like Chat-GPT or similar tools) for various academic tasks.

Each row represents a single session, capturing the student’s level, discipline, type of task, session length, AI effectiveness, satisfaction rating, and whether they reused the AI tool later.

**Attributes:**

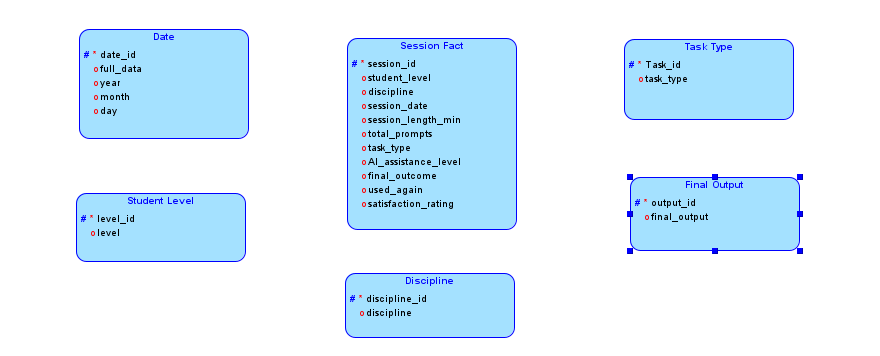
* **Session ID:** Unique session identifier.
* **Student Level:** Academic level: High School, Undergraduate, Graduate.
* **Discipline:** Student’s field of study (e.g., CS, Psychology, etc.).
* **Session Date:** Date of the session.
* **Session Length Min:** Length of AI interaction in minutes.
* **Total Prompts:** Number of prompts/messages used.
* **Task Type:** Nature of the task (e.g., Coding, Writing, Research).
* **AI Assistance Level:** 1–5 scale on how helpful the AI was perceived to be.
* **Final Outcome:** What the student achieved: Assignment Completed, Idea Drafted, etc.
* **Used Again:** Whether the student returned to use the assistant again.
* **Satisfaction Rating:** 1–5 rating of overall satisfaction with the session.

# **Logical & Physical Model**

A Logical Model and a Relational Model were designed. These models define the entities, relationships, and structure required for storing

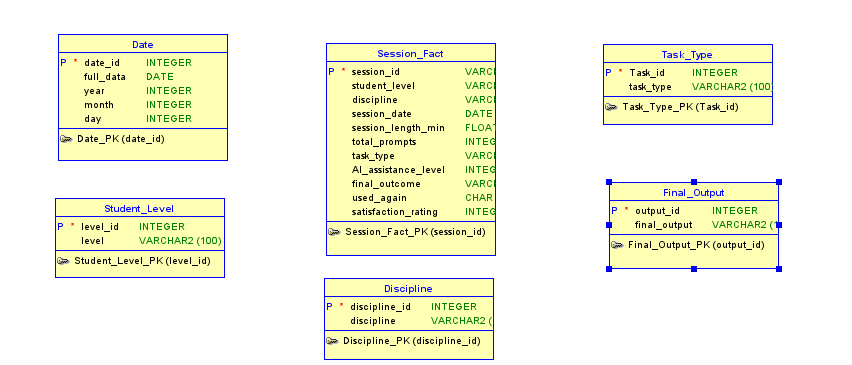
# **2.1 Logical Model**

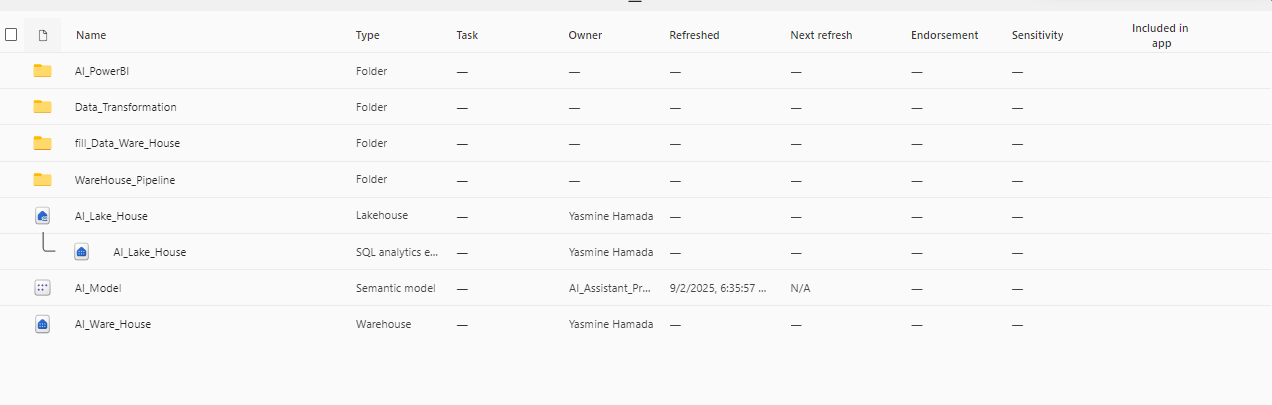
The logical model represents the high-level structure and relationships between the entities in the restaurant dataset without detailing storage or data types.



# **2.2 Physical Model**

The relational model describes the schema in terms of tables, columns, data types, and relationships between keys. It also captures physical characteristics such as the size and data types of each attribute, suitable for implementation in a relational database system.

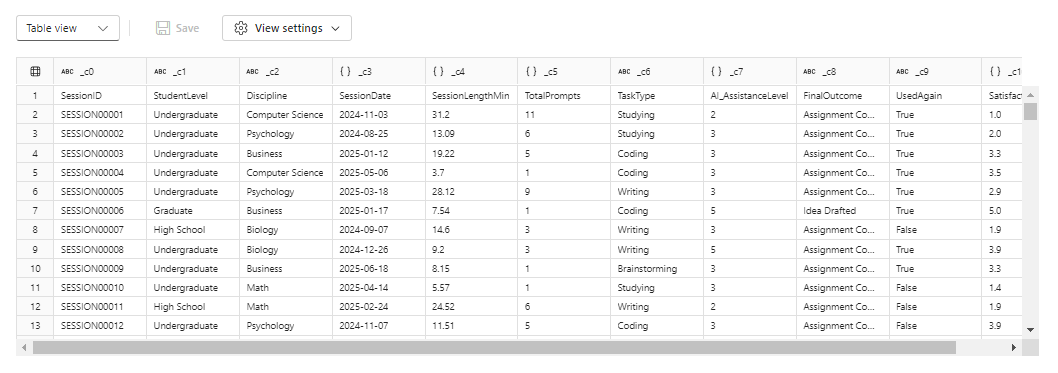
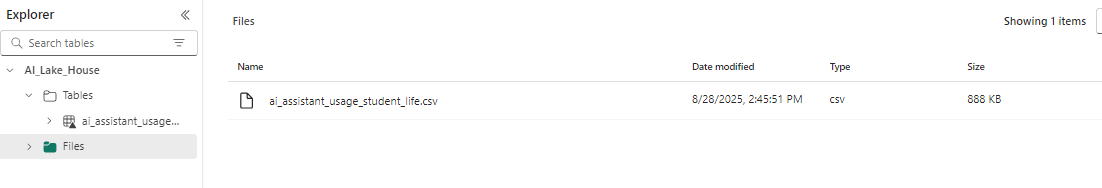


**Project Overview**

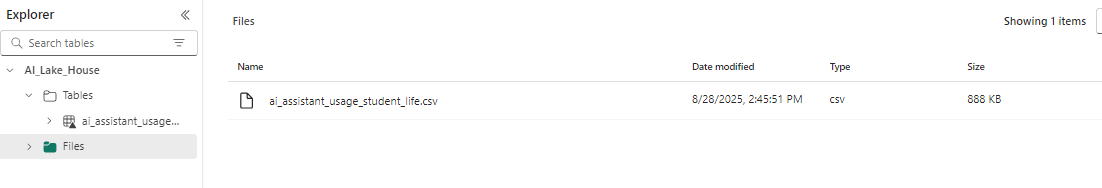
# **Lake House Overview**

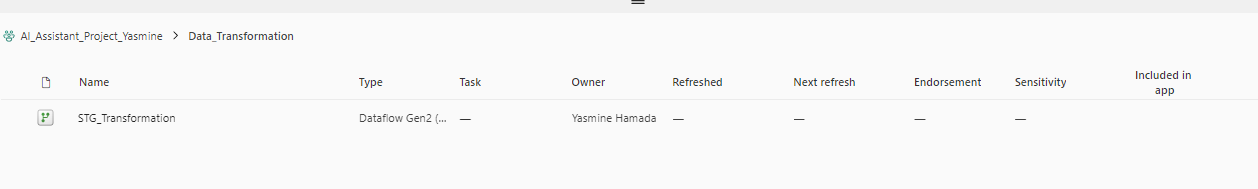
**Raw Layer (Bronze):**

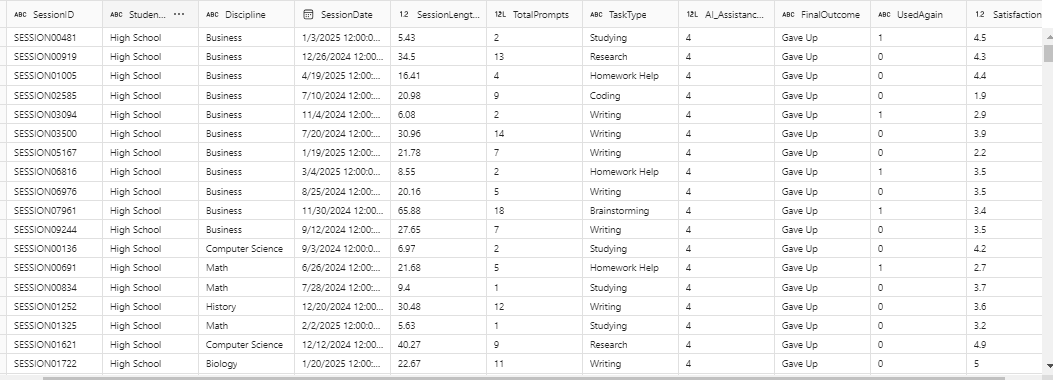
This is where data is first ingested in its original format. It's the landing zone for raw data from various sources.

**ODS/Staging Layer (Silver):**

This is the ODS. In this layer, the raw data is cleaned, transformed, and validated. This may include:

****

****

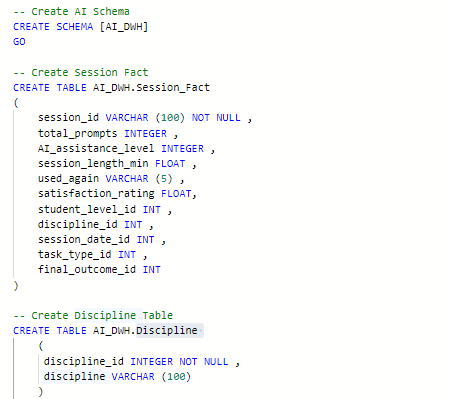


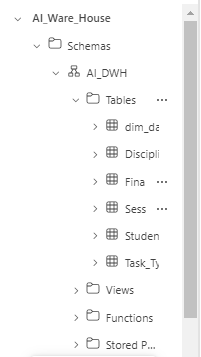
**Gold Layer:**

1. **Ware House:**

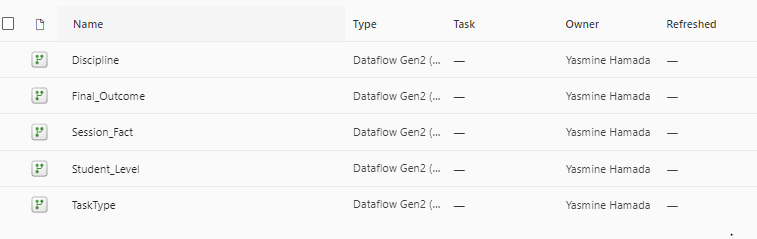
* **Create Schema and Dimensions Table:**
  1. **AI\_DWH Schema**
  2. **Dimensions:**
* Dim Discipline
* Dim Student Level
* Dim Task Type
* Dim Final Outcome
* Dim Date

**3-** **Session Fact**

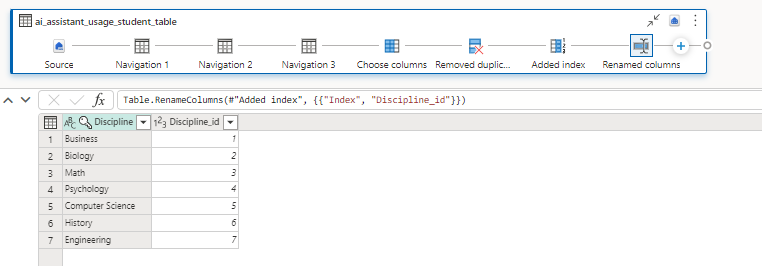
****



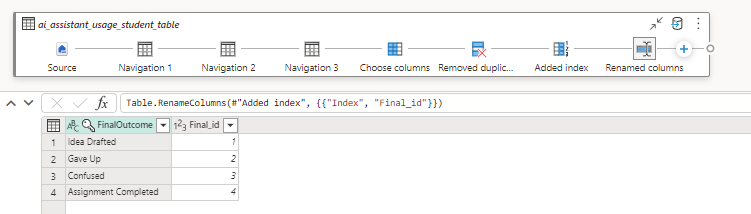
**Fill Data in Dimensions:**



**Fill Dim Discipline:**



**Fill Dim Final Outcome:**



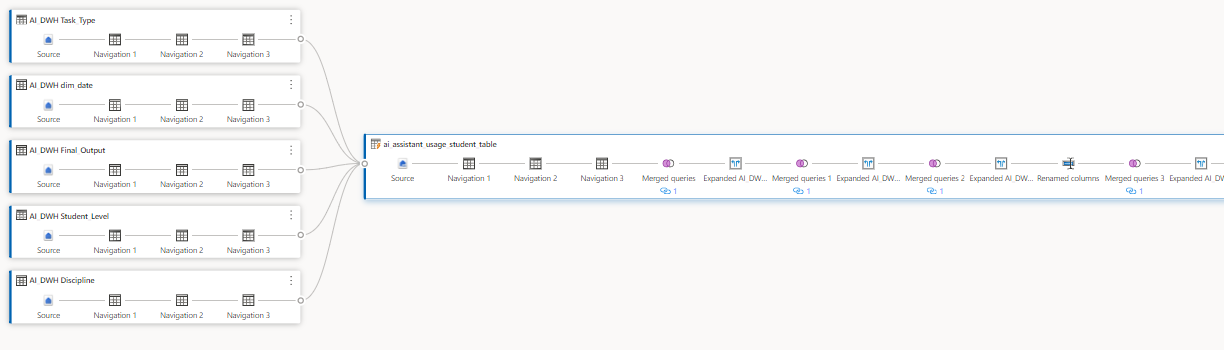
**Fill Dim Date:**

The Dim Date dimension table was created and fully populated using a custom SQL query. This query was designed to dynamically generate all necessary dates and their attributes to support analytical reporting, ensuring a complete and continuous timeline for the data**.**

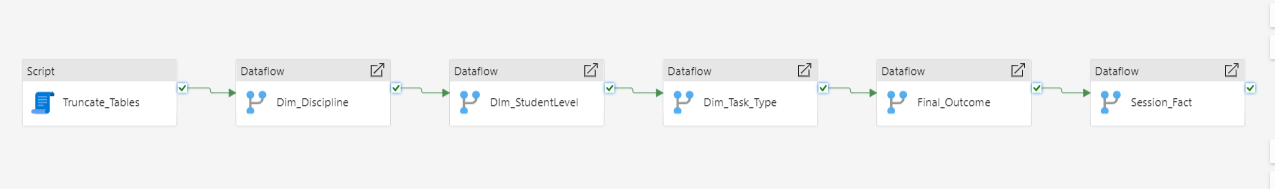
# **SSAS Overview**

In this project, SQL Server Analysis Services (SSAS) was used to deploy the multidimensional schema and enhance analytical capabilities. The model integrates fact and dimension tables from the Data Warehouse and exposes key business measures to support advanced reporting and analysis.

**Fill Session Fact:**

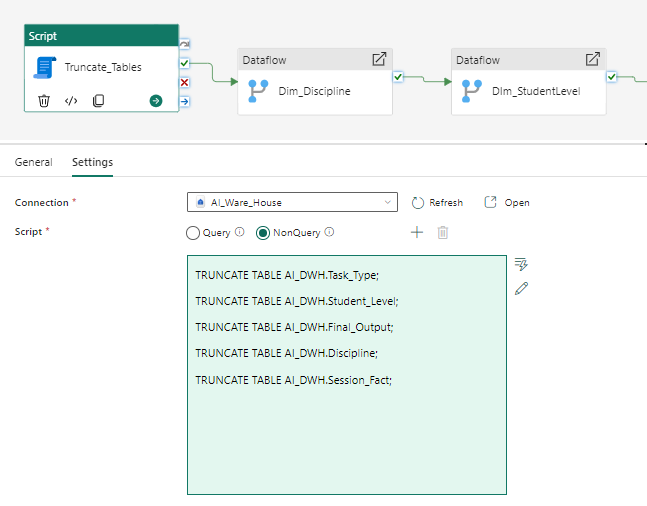


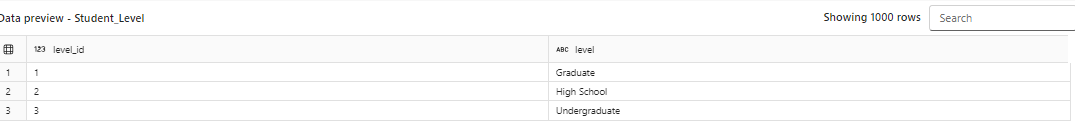
**Ware House Pipeline:**

****

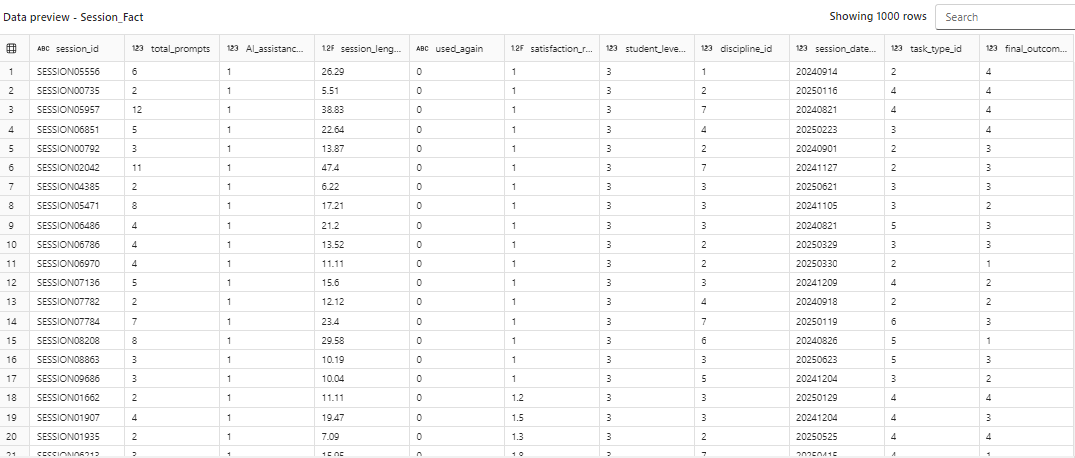
**Truncate Script:**

The data warehouse pipeline includes a dedicated script to perform a TRUNCATE operation on dimension tables. This step ensures that all existing data is removed, preparing the tables for a full reload with the latest information.

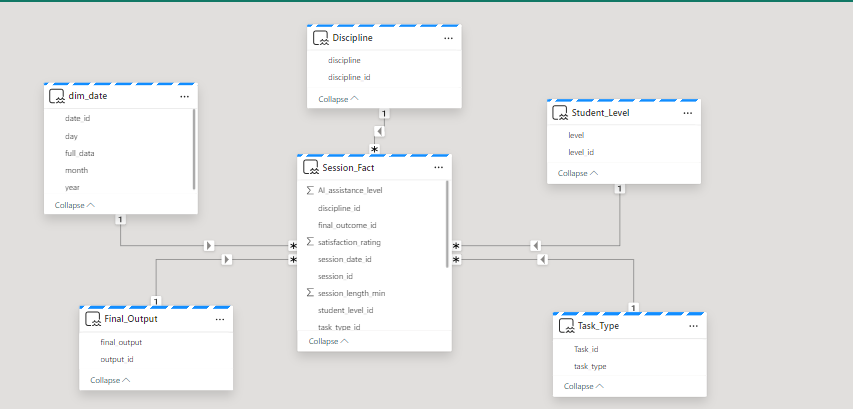


**Student Level Table:**

**Session Fact Table:**

****

**Data Model:**

****

**Implemented Measures:**

* **Total Session**: The total number of sessions.
* **Total Prompt**: The sum of all prompts used across all sessions.
* **Average Rating**: The average satisfaction rating given by users.
* **Average Session Length Min**: The average duration of a session in minutes.



# **Power BI Overview**

Power BI was utilized to create data visualizations and build interactive dashboards and reports based on the refined data models within Azure Synapse Analytics (or Microsoft Fabric) and the underlying Restaurant Data Warehouse. This enabled dynamic exploration of key business metrics such as restaurant distribution, average ratings, total votes, and average cost, providing valuable insights to support decision-making.

