Data Consumption using Azure Purview

Business Overview
Purview is a single data governance solution from Azure that lets you manage and regulate your data across on-premises, multi-cloud, and software-as-a-service (SaaS) environments. With automated data discovery, sensitive data categorization, and end-to-end data lineage, you can create a holistic, up-to-date map of your data environment. It allows data curators to manage and safeguard their data estates and data consumers to locate important reliable data.

By delivering data scanning and categorization as a service for assets throughout your data estate, Azure Purview automates data discovery. Metadata and descriptions of data assets identified are combined into a comprehensive map of your data estate. There are purpose-built apps above this map that enable data discovery, access management, and data landscape analytics environments.

Data Map makes your data more meaningful by visualizing your data assets and their interactions throughout your data estate. The data map is used to find information and control who has access.

Data Catalog helps you identify reliable data sources by exploring and searching your data assets. The data catalog identifies data sources by aligning your assets with business terminology and data classification.

Data Insights provides a snapshot of your data estate, allowing you to see what kind of data you have and where it is stored.

Data Pipeline
A data pipeline is a technique for transferring data from one system to another. The data may or may not be updated, and it may be handled in real-time (or streaming) rather than in batches. The data pipeline encompasses everything from harvesting or acquiring data using various methods to storing raw data, cleaning, validating, and transforming data into a query-worthy format, displaying KPIs, and managing the above process.

Agenda
This project is the second part of a two projects series. The first project focuses on data ingestion and preparation for Azure Purview using Azure Logic Apps, Azure Storage Account, Azure Data Factory, and Azure SQL Databases. This project focuses on data consumption and analysis using Azure Purview.

Dataset Description
This project uses Hospital data with 30 unique data fields, a few of which include:
- Hospital Name
- Address
- Hospital Type
- Mortality national comparison
- Patient experience national comparison footnote
- Hospital overall rating
- Meets criteria for meaningful use of EHRs

Tech Stack

➔Services: Azure Logic Apps, Azure Storage Account, Azure Data Factory, Azure SQL Databases, DBeaver, Azure Purview

Key Takeaways
- Exploring Azure Purview
- Creation o Azure Purview Account
- Creation and Scanning of Purview Collection
- Exploring Purview UI
- Understanding Azure Purview Glossary
- Understanding Azure Purview Assets
- Introduction to Purview Access Control
- Best Practices in Azure Purview
- Understanding limitations of Azure Purview

Project Architecture