**PROJECT DOCUMENTATION**

**👨‍💻 Developed by: Rajesh Kumar Rout**

**Platform: Microsoft Fabric**

**Tools & Technologies: PySpark, Delta Lake, Lakehouse, Data Warehouse, Power BI**

**Project Overview**

This project focuses on **analyzing financial data** for a banking system to generate actionable insights regarding:

* Transaction trends
* Customer spending behavior
* Risk assessment

The data flows from **raw ingestion** to a **gold layer star schema**, and finally into a **Power BI dashboard**, via **Microsoft Fabric Lakehouse and Warehouse**.

**Architecture Overview**

**Importing data into MSSQL-**

* Imported data from three Excel files into a Microsoft SQL Server database in tabular format.

**Bronze layer-**

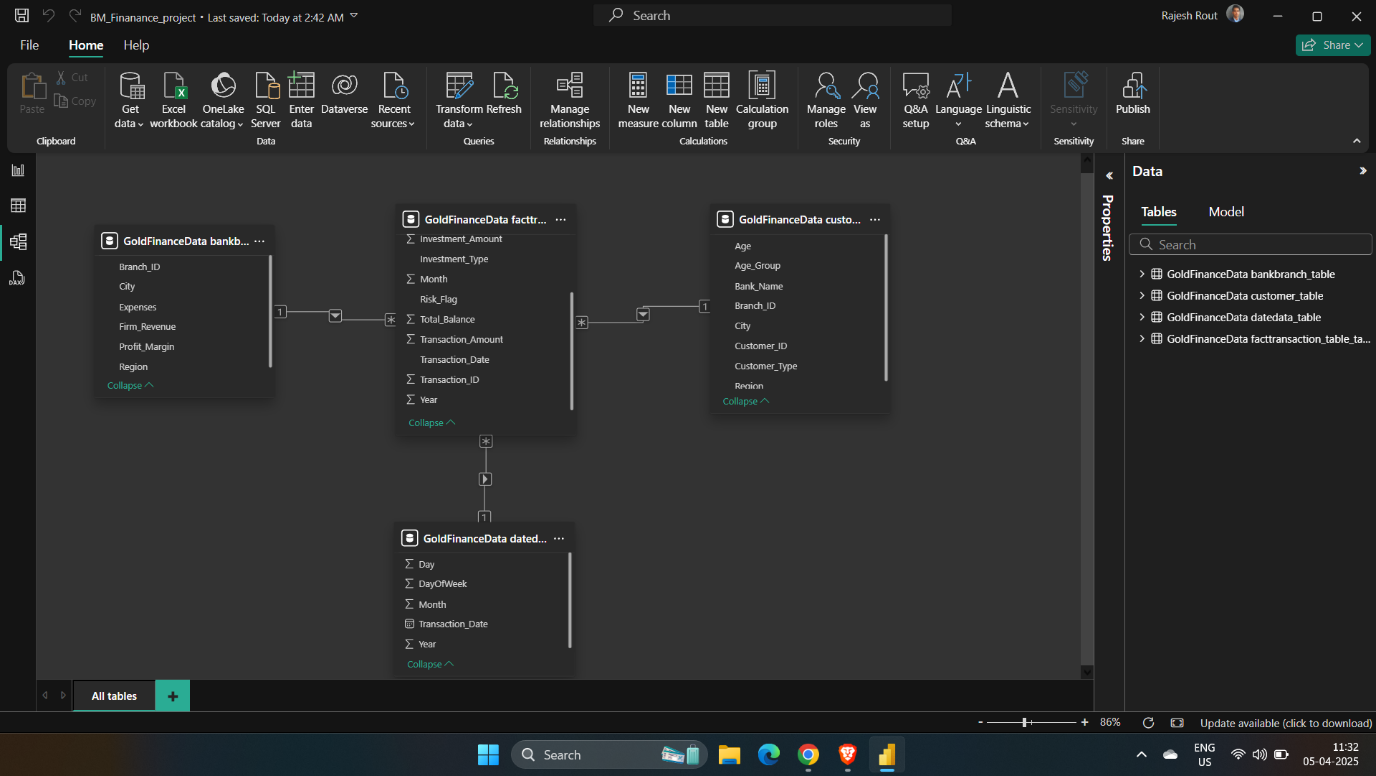
* Connected on-premises SQL Server to Microsoft Fabric using the On-premises Data Gateway, enabling secure and reliable data transfer without exposing the source directly.
* Configured the data pipeline to ingest data into the Bronze layer of a Fabric Lakehouse, ensuring the raw data was stored efficiently for further transformation.
* Utilized OneLake storage to manage the ingested data centrally, setting the stage for building a structured multi-layered data architecture.

**Silver Layer(using pyspark notebook):**

* Cleaned data stored in Delta format tables.
* Perform operations like Drop duplicates, fill the null values
* Tables: customer\_data, transaction\_data, bank\_data.

**⭐ Gold Layer (Star Schema):**

Built using PySpark with the following structure:

Apply Bussiness logic to transform silver layer data to gold layer (BI consumable data)

**Dimension Tables:**

* **bankbranch\_table**: Customer details (Customer\_ID, Age, Region, etc.)
* **customer\_table**: Bank branch financials (Branch\_ID, Revenue, Profit, etc.)
* **datedata\_table**: Date-based hierarchy (Year, Month, Day)

**Fact Table:**

**facttransaction\_table**: Transactional and investment facts (amounts, types, balances)

**🔄 Gold Layer to Fabric Data Warehouse**

To support Power BI connectivity and faster querying, gold layer tables were **moved from Lakehouse to Warehouse** using Pipeline:

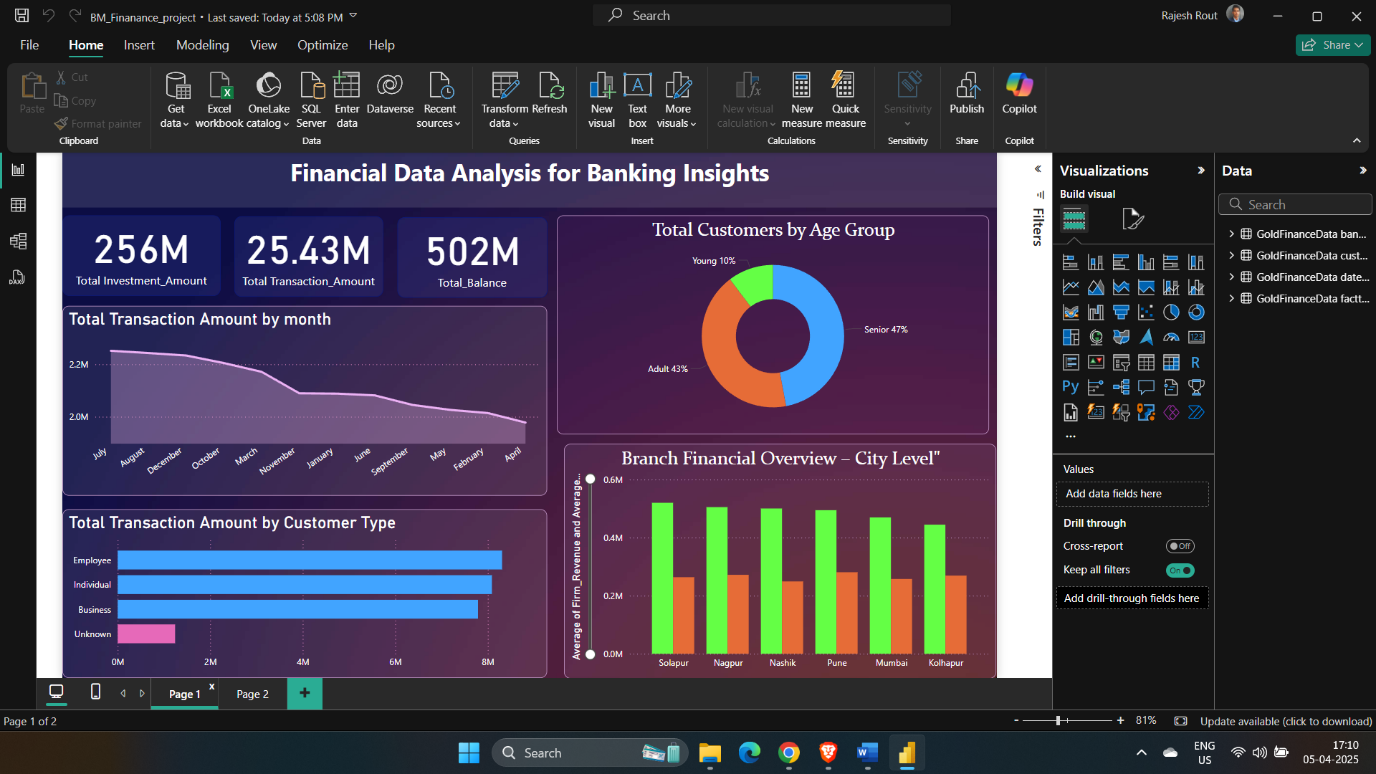
**📊 Power BI Reports**

Using Fabric's Direct Lake connection, the following insights were visualized:

**1. Transaction Trends**

* Monthly volume of transactions
* Total transaction by customer

**2. Customer Spending Behavior**

* Average spending by age group
* Customer distribution by account type

**3. Risk Assessment**

* Branches with low profit margins
* Customers with high withdrawals vs balance
* City and region wise analysis of branch performance

