Project Documentation: Data Transformation and Visualization

# 1. Introduction

This project focuses on providing insights to a bank regarding transaction trends, customer spending behavior, and risk assessments. The data is processed through various layers in the lakehouse, transformed, and visualized using Power BI to aid in decision-making.

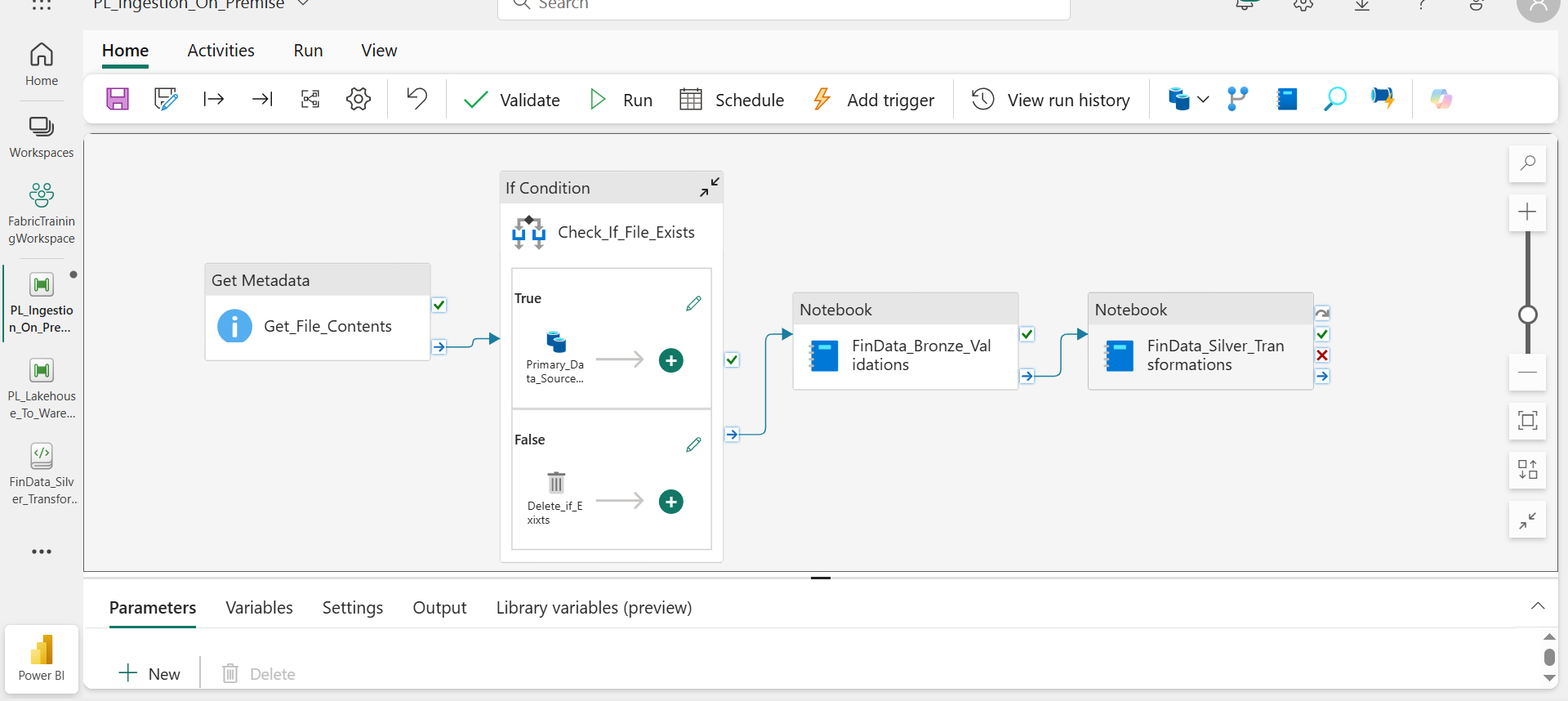
The goal of the project is to automate the financial data flow, including ingestion, processing, transformation, and ultimately visualizing the data in Power BI dashboards, replacing the previous manual process

# 2. Pipeline Design

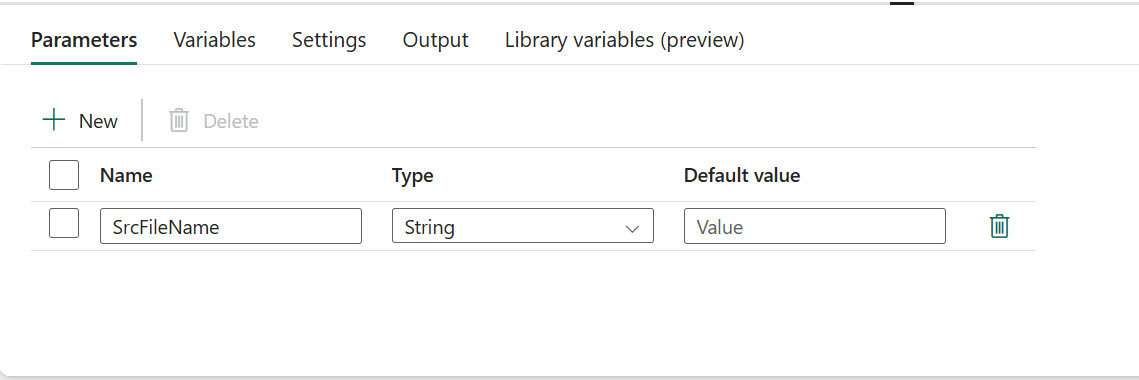
The pipeline is designed to transfer data from the Bronze layer to the Silver layer and then to the Gold layer. The data starts in the Bronze layer, where raw data is validated, then moves to the Silver layer where it is transformed, and finally to the Gold layer for structured and clean data that is ready for reporting and visualization.

The pipeline from Lakehouse to Warehouse is used to ensure structured data is ready for Power BI visualization. Data is transferred from the Lakehouse (fin\_data\_lakehouse) to the Warehouse where structured data is stored.

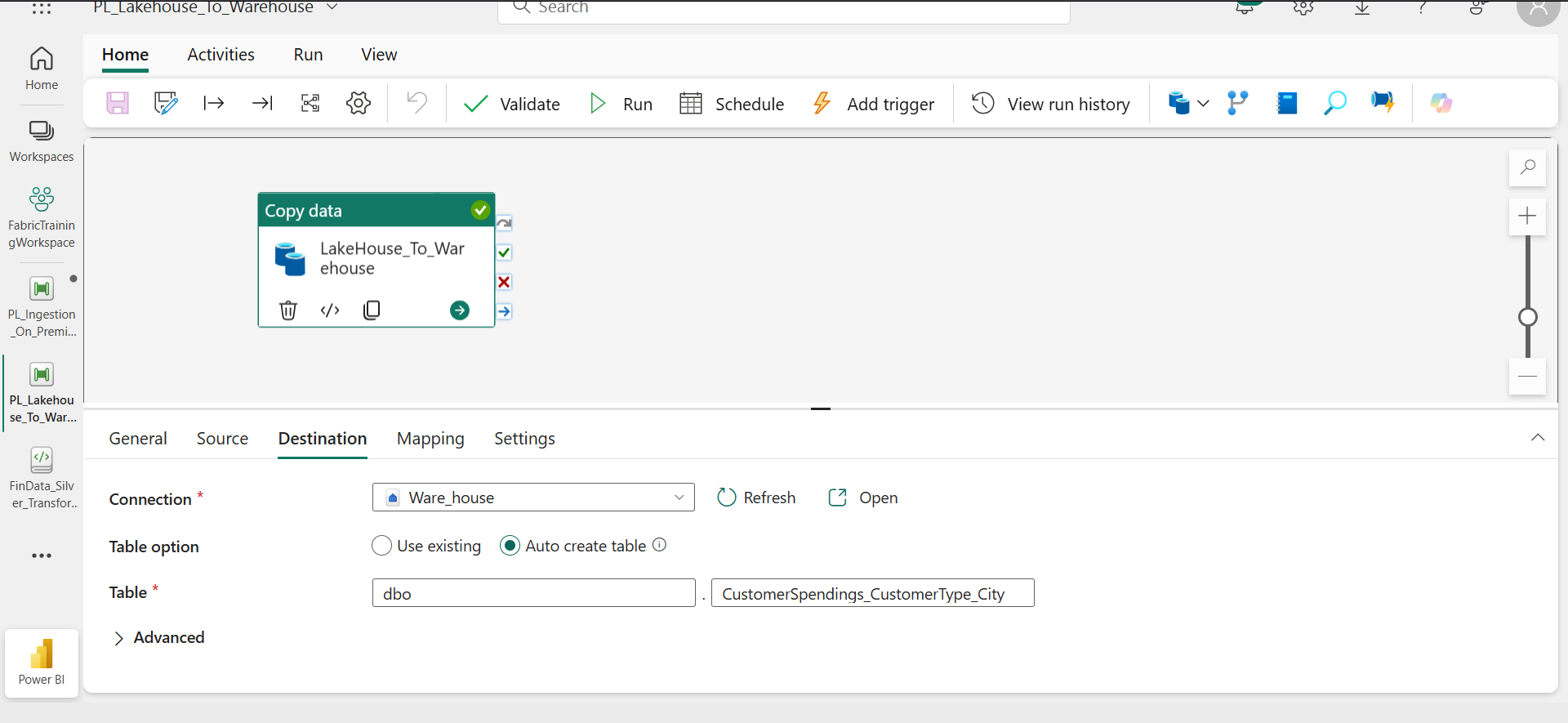
**PL\_Ingestion\_On\_Premise :**



**Pipeline parameters for passing File name dynamically**



**Pipeline To transfer Data From Lakehouse to Warehouse**



# 3. Data Transformation Process

**Here's the summary of the Data Ingestion**:

The primary source of data is **SQL Server**.

Data is transferred from SQL Server to **Microsoft Fabric** (cloud platform).

**Data Gateway** was created to link SQL Server with Fabric.

Data ingestion is done via **pipelines** in Fabric, specifically using **copy activity**

**Bronze Layer**

In the Bronze layer, raw data from SQL is validated and cleaned. The transformations include: dropping duplicates, filling null values, changing data types, and ordering the data for consistency. For example, 'Firm Revenue' is converted to an integer, and null values are filled based on specific business rules.

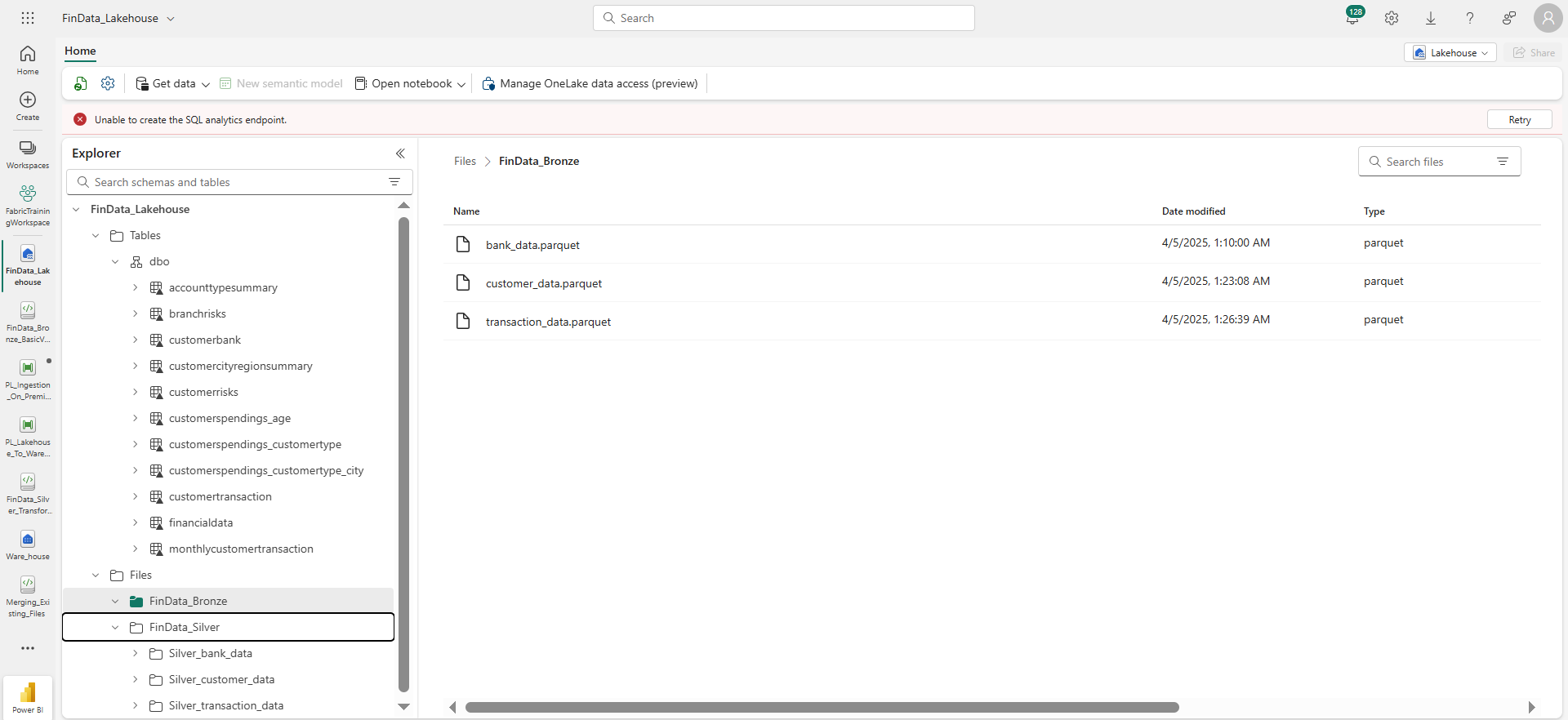
**Silver Layer**

The Silver layer focuses on more complex transformations, including aggregating data for transaction trends, customer spending behavior, and risk assessment. Data is categorized based on age groups, customer types, and city regions.

**Transaction per Month**: Transformed the data to show transactions by month.

**Transaction Based on Account Type**: Calculated the sum of transactions based on different account types (e.g., fixed, savings).

**Transaction Based on Region and City**: Analyzed transaction trends based on geographical regions and cities.



**Gold Layer:**

Connected Power BI Desktop to the warehouse using your Fabric account, imported tables like Branch Risks, Customer, City Region, Customer Risk, Customer Spending, and Customer Spending by Customer Type and City. Then, you created four interactive tabs based on this data.

# 4. Power BI Visualizations

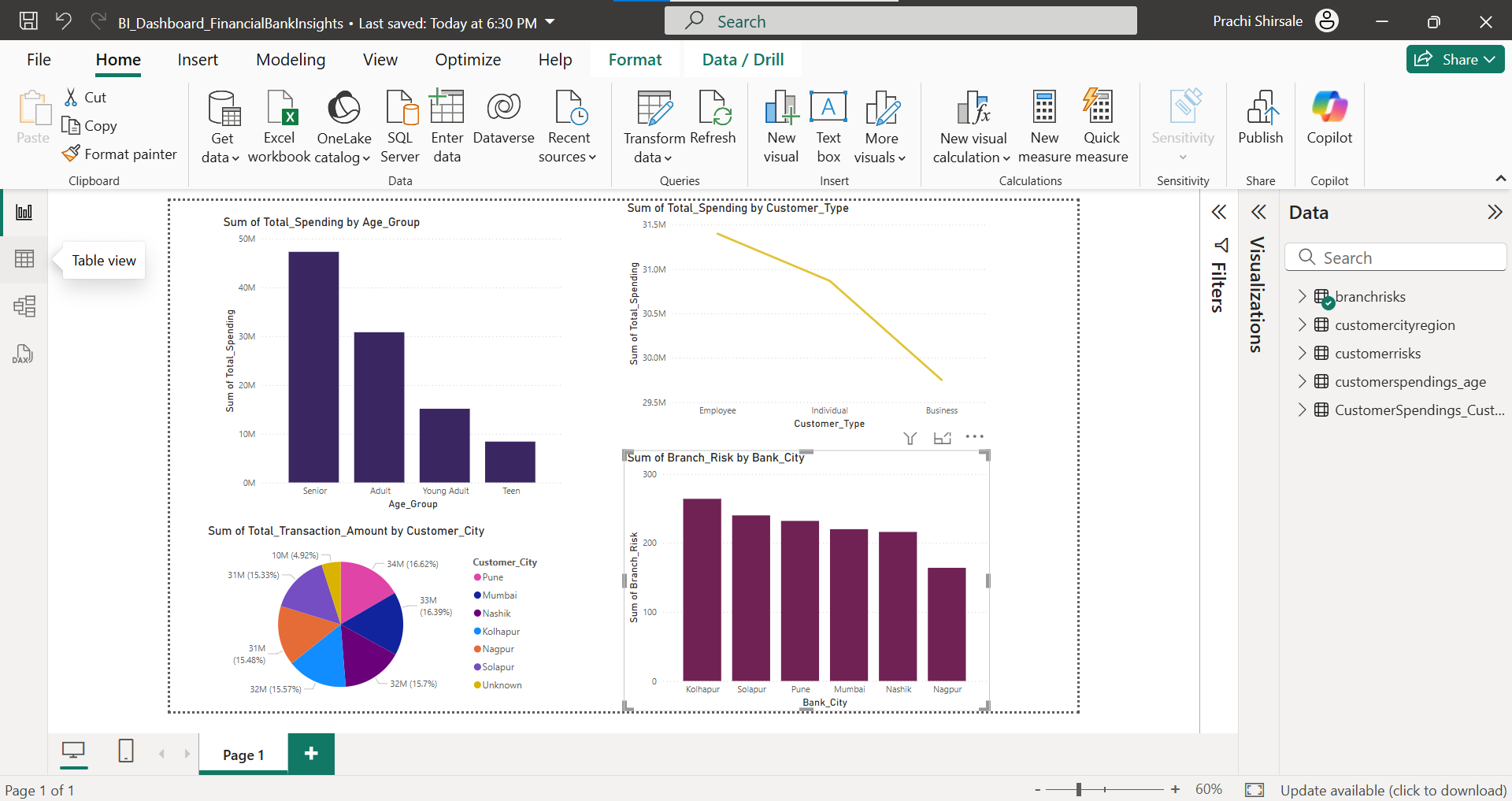
Power BI Desktop is used to visualize the transformed data. The following visualizations are created:

1. Sum of Total Spending by Age Groups (Bar Graph) - Analysis of total spending by different age groups: Senior, Adult, Young, and Teen.

2. Sum of Total Spending by Customer Type (Line Graph) - Insights into spending behavior based on customer types: Employee, Individual, and Business Person.

3. Sum of Total Transaction Amount by Customer City (Pie Chart) - Geographical insights on transaction amounts by city.

4. Sum of Branch Risk by Branch City (Bar Graph) - Analysis of branch risks based on different cities.



# 5. Conclusion

The project successfully demonstrates how data transformation and visualization can assist the bank in gaining insights into transaction trends, customer spending behavior, and risk assessments. The transformations applied provide a clean, structured data set, while the visualizations allow for better decision-making regarding resources, marketing, and customer management.

**Sum of Total Spending by Age Groups (Bar Graph):** This will help the bank understand which age group is spending the most. This analysis can help tailor marketing strategies and promotional offers to the most active age groups.

**Sum of Total Spending by Customer Type (Line Graph):** By visualizing spending patterns of different customer types (employee, individual, business person), the bank can identify trends for each customer category and create customized financial products or incentives for each type.

**Sum of Total Transaction Amount by Customer City (Pie Chart):** This pie chart will allow the bank to identify which cities are driving the most transactions. This insight can help the bank allocate resources and marketing efforts to cities with the highest transaction volumes.

**Prachi (INT - 31)**