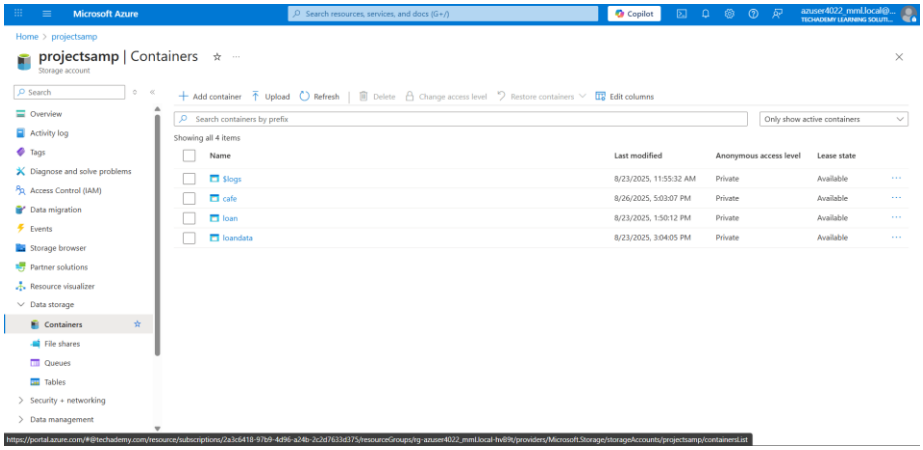
Azure DevOps Coding Challenge

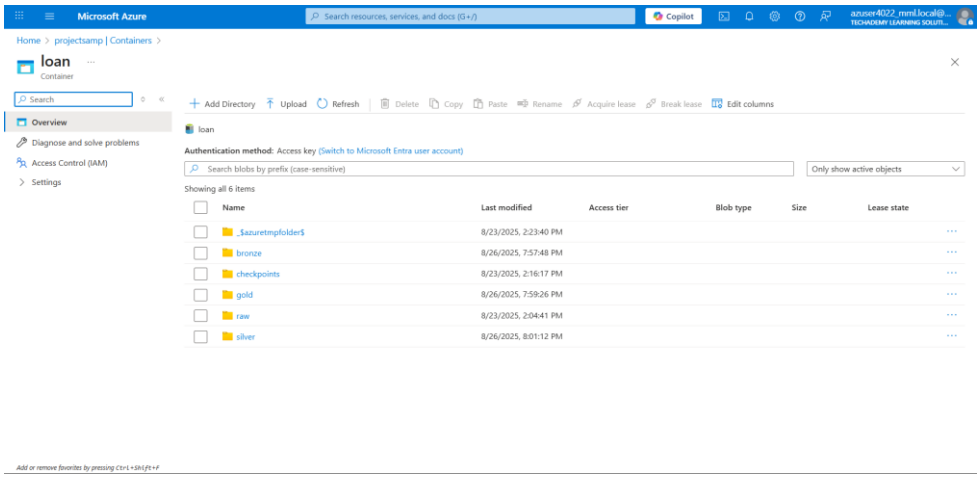
Implementing CI/CD pipline activiry in Azure devops,

Step 1: Create Azure Storage Account

- Provision a Storage Account in Azure.
- Create containers:
    - source → for raw CSV files
    - bronze/silver/gold → for processed data



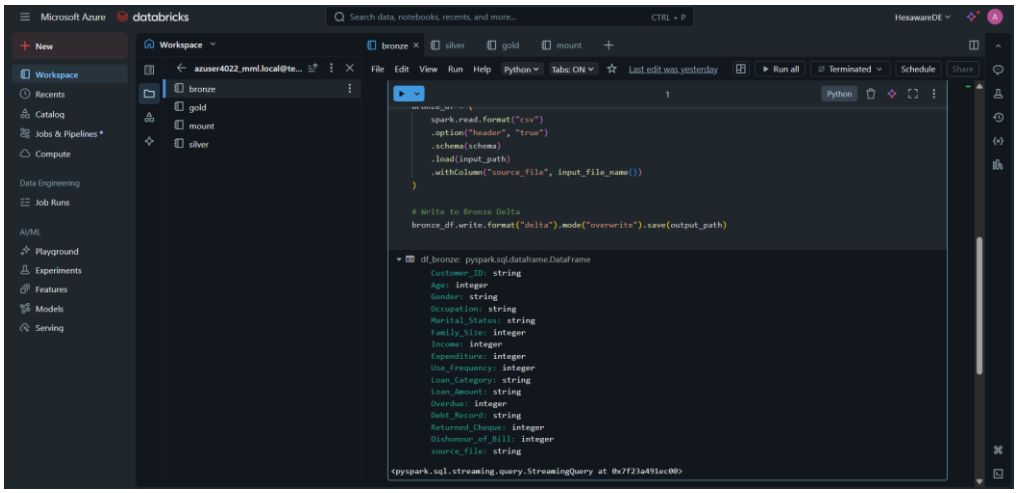
- Upload the raw loan CSV files into the source container.



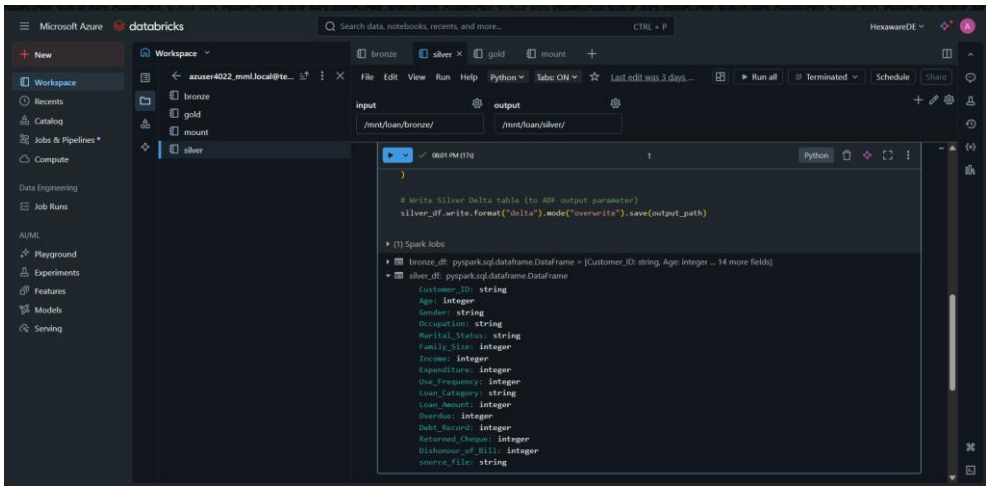
Step 2: Set up Azure Databricks

- Create a Databricks workspace and a Spark cluster.
- Mount the ADLS Gen2 containers into Databricks.

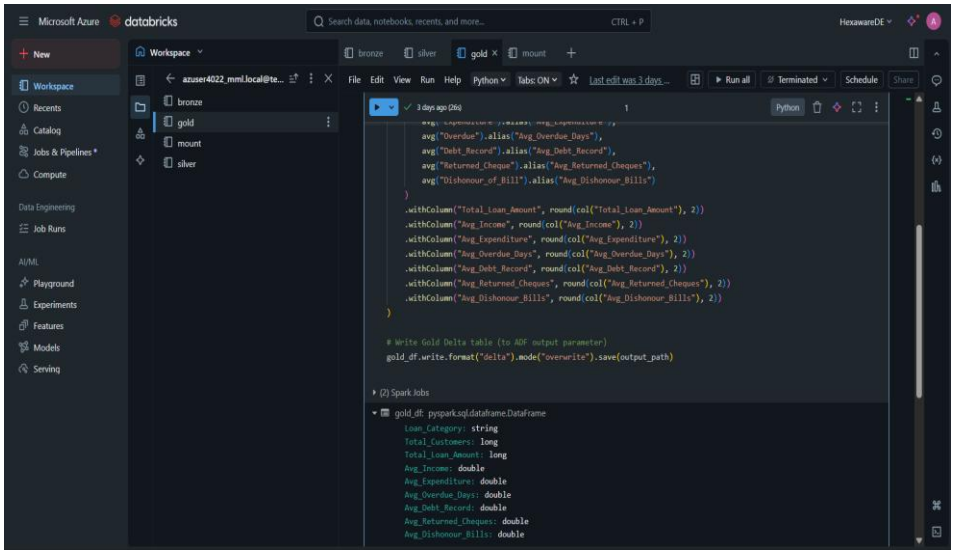
- Create and run notebooks:
    - **Bronze notebook** → ingest raw CSV into Bronze Delta tables.



    - **Silver notebook** → clean, standardize, and transform data.

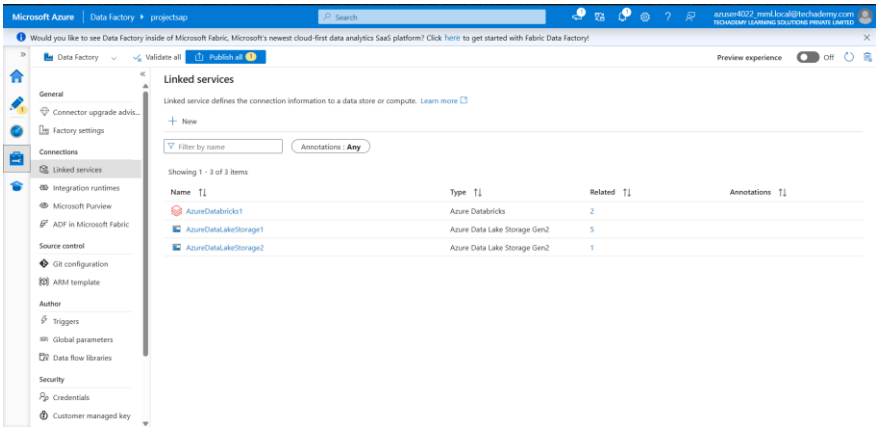


    - **Gold notebook** → aggregate metrics (loan amounts, overdue trends, repayment stats).



Step 3: Create Azure Data Factory (ADF)

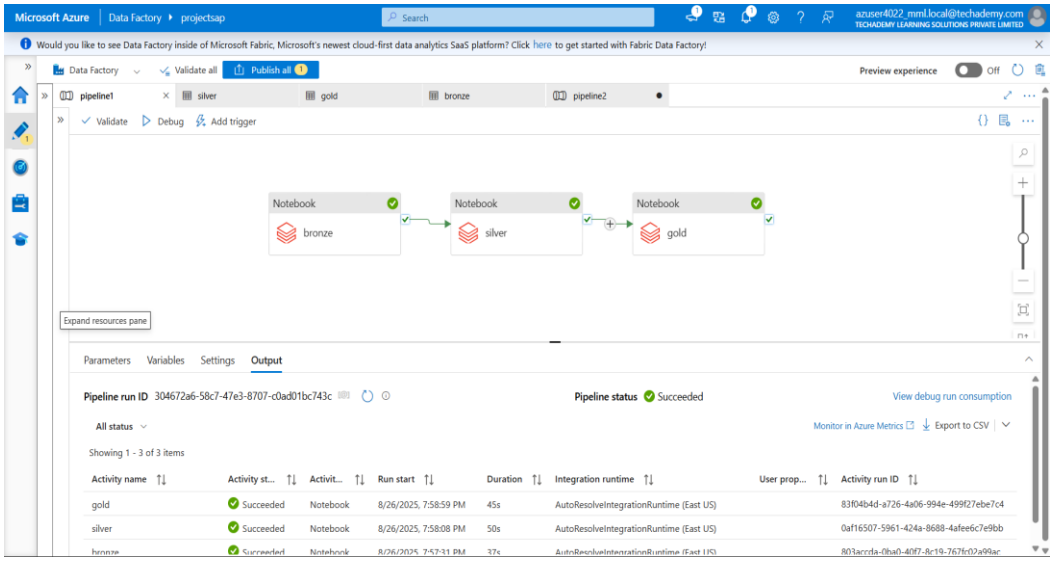
- Launch ADF Studio in the portal.
- Build a pipeline with activities:
  - Databricks Notebook Activity → call Bronze, Silver, Gold notebooks.
- Configure linked services:
  - to Azure Storage (ADLS)
  - to Databricks workspace



Step 4: Configure Pipeline Activities

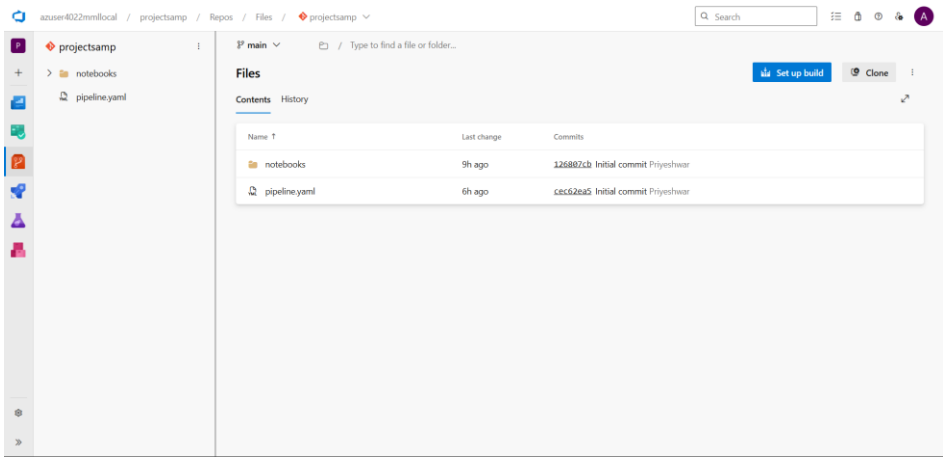
- Parameterize input/output paths.
- Chain activities so Silver runs after Bronze, Gold runs after Silver.
- Validate and debug the pipeline.

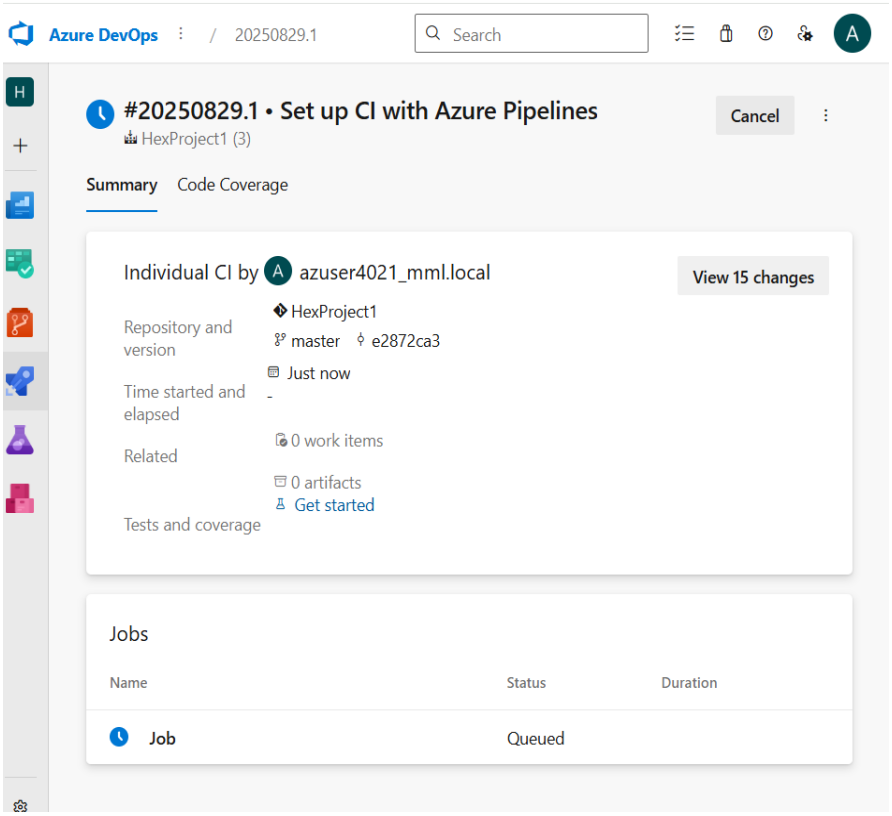
- Trigger execution (manual or schedule).



Step 5: Set up Azure DevOps (CI/CD)

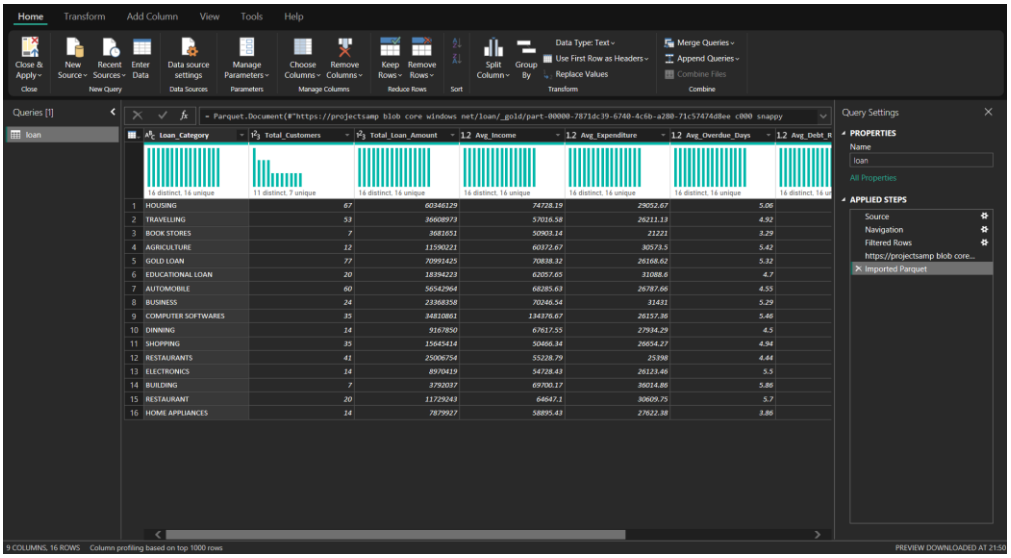
- Store notebooks and ADF JSON definitions in an Azure Repos Git repo.
- Create an Azure Pipelines YAML (azure-pipelines.yml) for CI/CD.
- Add steps to:
  - Deploy Databricks notebooks via databricks-cli.
  - Deploy ADF pipelines via ARM template deployment.
- Run pipeline → validate successful deployment.

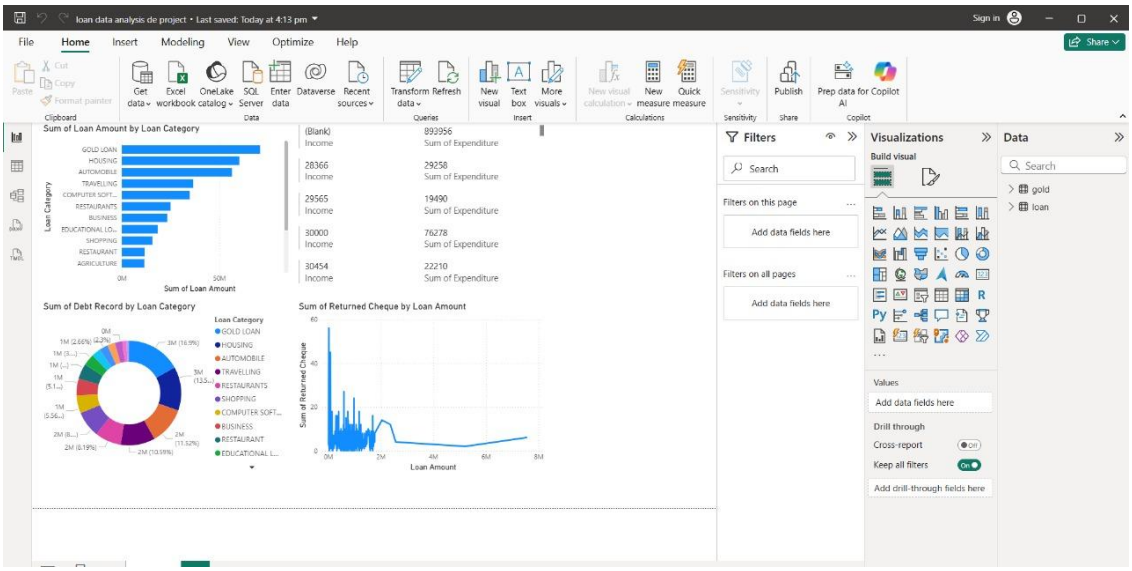




Step 6: Verify & Visualize

- In ADF Monitoring → confirm Bronze → Silver → Gold pipeline run success.
- In Databricks → check Delta tables created for Bronze/Silver/Gold.
- Connect Power BI to Gold tables.
- Build dashboards for loan insights (loan distribution, overdue trends, repayment performance).


Home
Transform
Add Column
View
Tools
Help
Queries [1]
loan
Query Settings
PROPERTIES
Name
loan
All Properties
APPLIED STEPS
Source
Navigation
Filtered Rows
https://projectsamp.blob.core...
Imported Parquet
Loan_Category
Total_Customers
Total_Loan_Amount
Avg_Income
Avg_Expenditure
Avg_Overdue_Days
Avg_Debt_R
1 HOUSING 67 60346129 74728.19 29052.67 5.06
2 TRAVELLING 53 36608973 57016.58 26211.13 4.92
3 BOOK STORES 7 3681651 50903.14 21221 3.29
4 AGRICULTURE 12 11590221 60372.67 30573.5 5.42
5 GOLD LOAN 77 70991425 70838.32 26168.62 5.32
6 EDUCATIONAL LOAN 20 18294223 62057.65 31088.6 4.7
7 AUTOMOBILE 60 56542964 68285.63 26787.66 4.55
8 BUSINESS 24 23368358 70246.54 31431 5.29
9 COMPUTER SOFTWARES 35 34810881 134376.67 28157.38 5.48
10 DINNING 14 9167850 67617.55 27934.29 4.5
11 SHOPPING 35 15645414 50486.34 26854.27 4.94
12 RESTAURANTS 41 25006754 55228.79 25398 4.44
13 ELECTRONICS 14 8970419 54728.43 26123.46 5.5
14 BUILDING 7 3792037 69700.17 36014.86 5.86
15 RESTAURANT 20 11729243 64647.1 30609.75 5.7
16 HOME APPLIANCES 14 7879927 58895.43 27622.38 3.86
9 COLUMNS, 16 ROWS
Column profiling based on top 1000 rows
PREVIEW DOWNLOADED AT 21:50


loan data analysis de project • Last saved: Today at 4:13 pm
Sign in
File
Home
Insert
Modeling
View
Optimize
Help
Share
Sum of Loan Amount by Loan Category
Loan Category
Sum of Loan Amount
(Blank)
Income
893956
Sum of Expenditure
28366
Income
29258
Sum of Expenditure
29565
Income
19490
Sum of Expenditure
30000
Income
76278
Sum of Expenditure
30454
Income
22210
Sum of Expenditure
Sum of Debt Record by Loan Category
Loan Category
Sum of Returned Cheque by Loan Amount
Sum of Returned Cheque
Loan Amount
Filters
Search
Filters on this page
Add data fields here
Filters on all pages
Add data fields here
Visualizations
Build visual
Values
Add data fields here
Drill through
Cross-report
Keep all filters
Add drill-through fields here
Data
Search
gold
loan