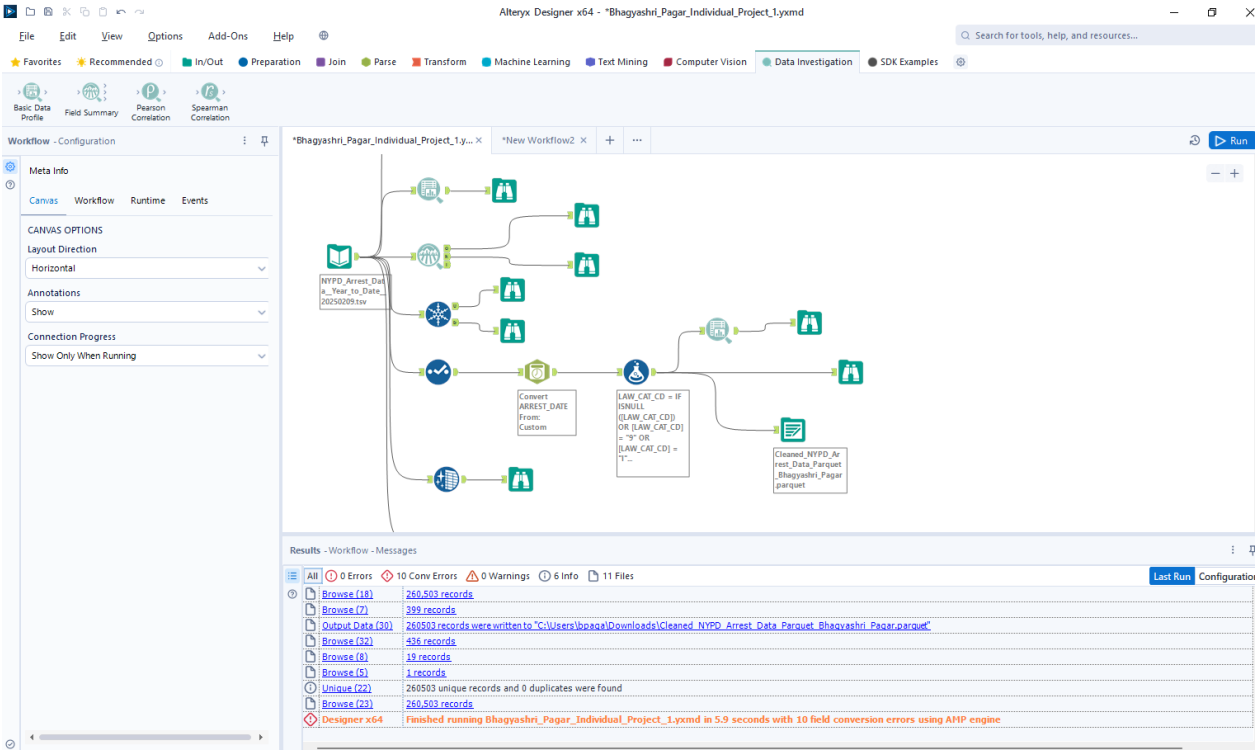
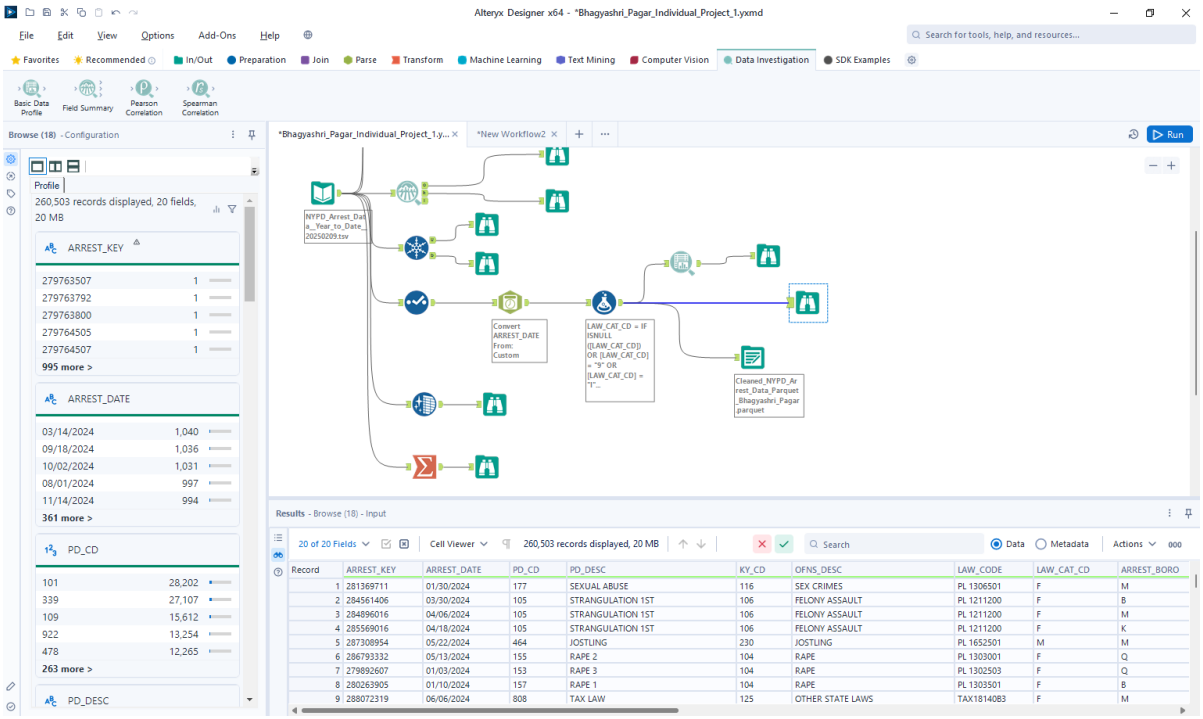
Alteryx, ADF Pipeline Run and Snowflake

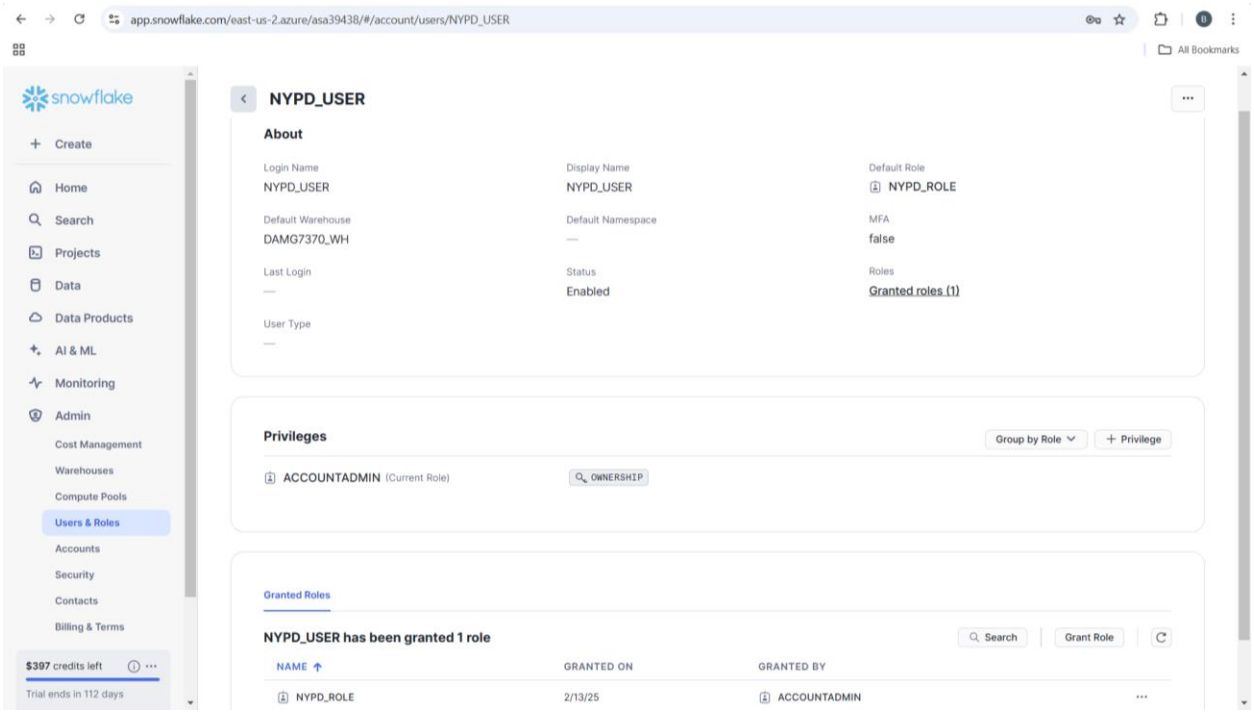
Bhagyashri Avinash Pagar **NUID:002310690**

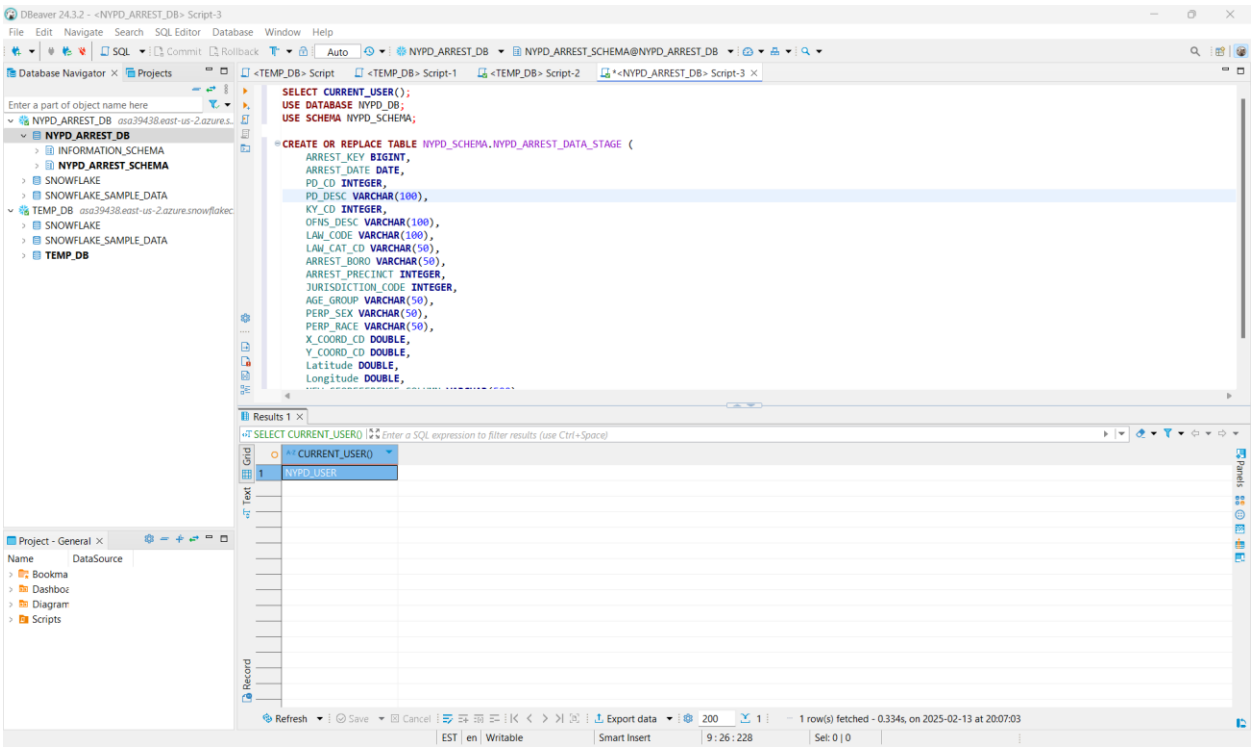
Alteryx Workflow which is used for data cleaning and data profiling



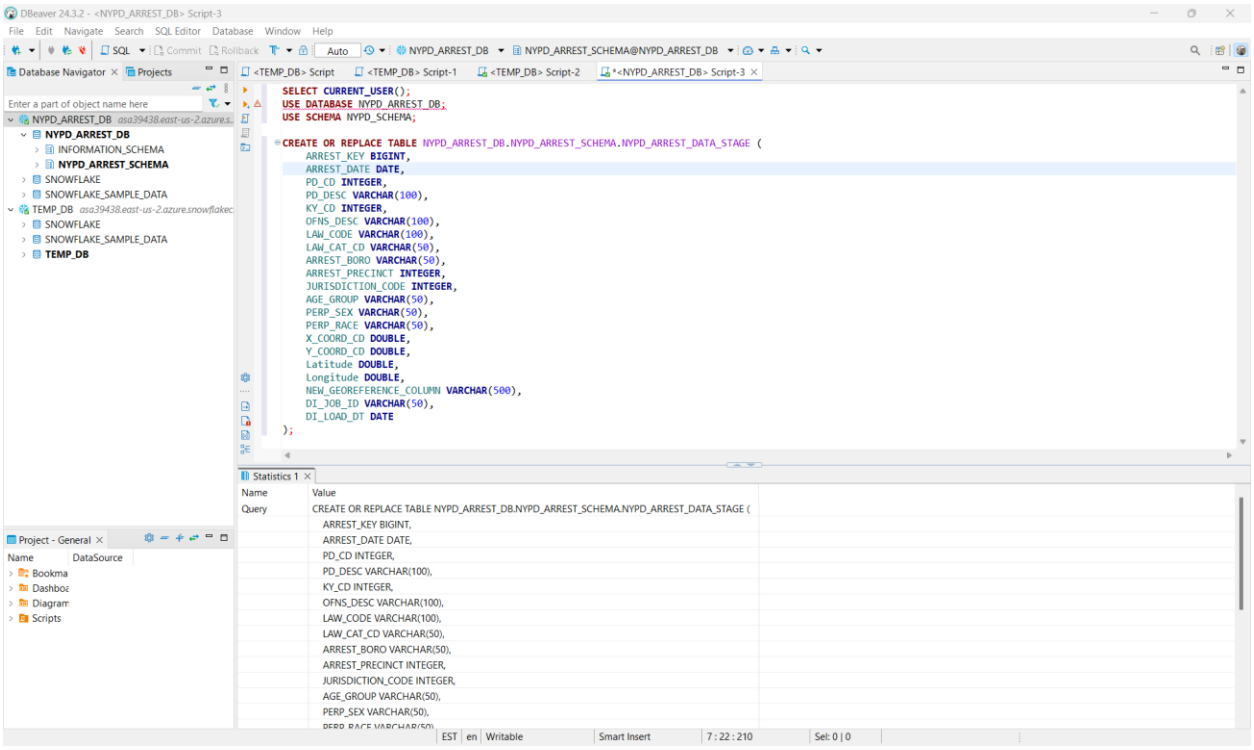


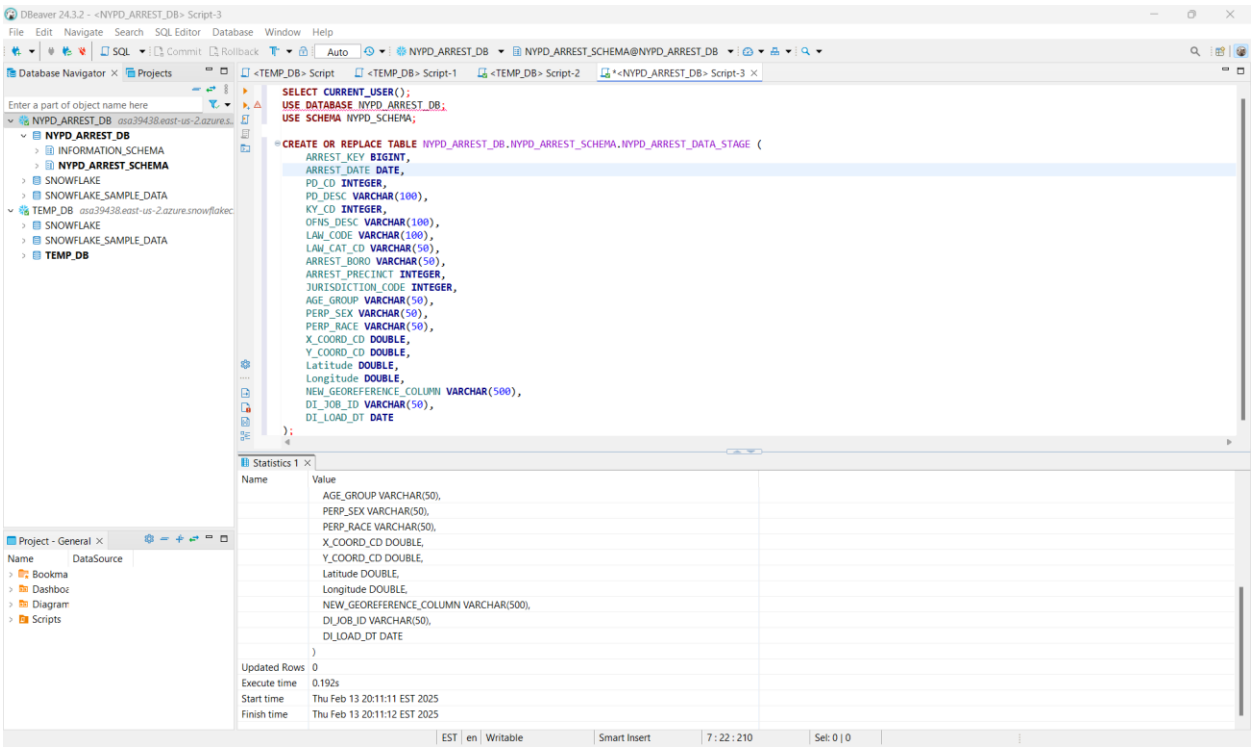
Snowflake Environment Setup



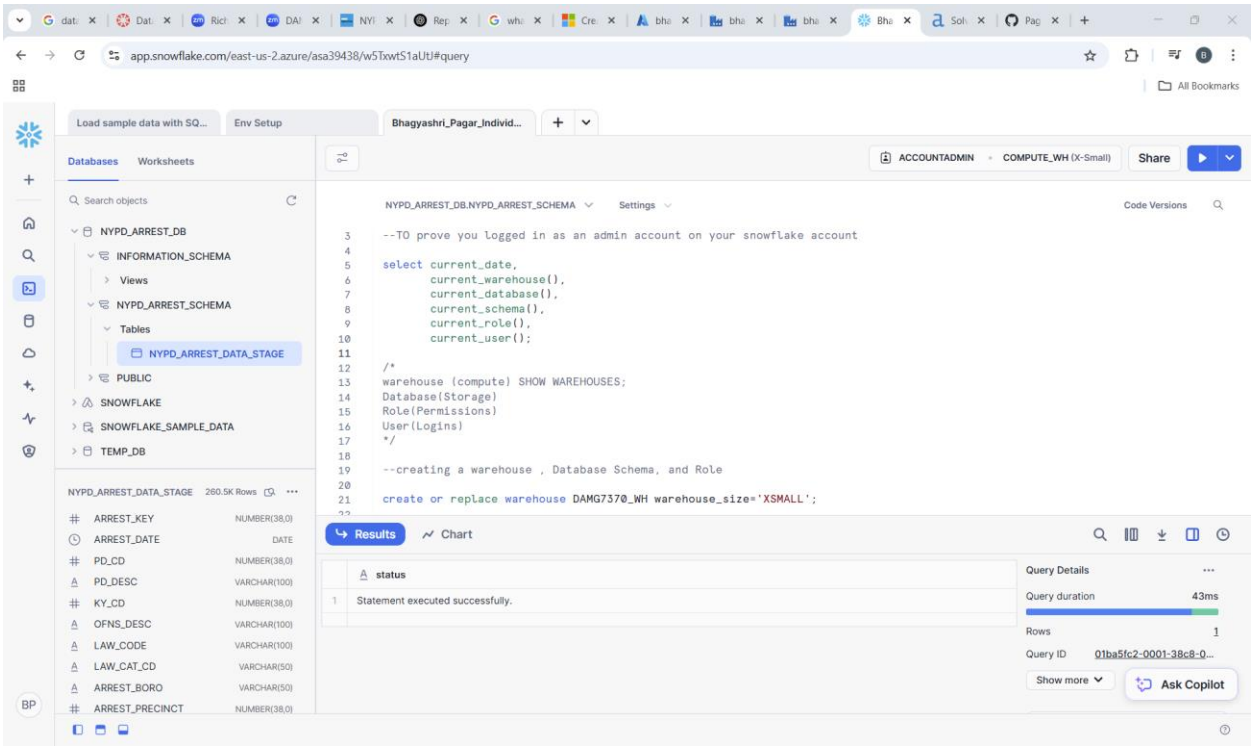


Snowflake NYPD_ARREST_SCHEMA is created in snowflake



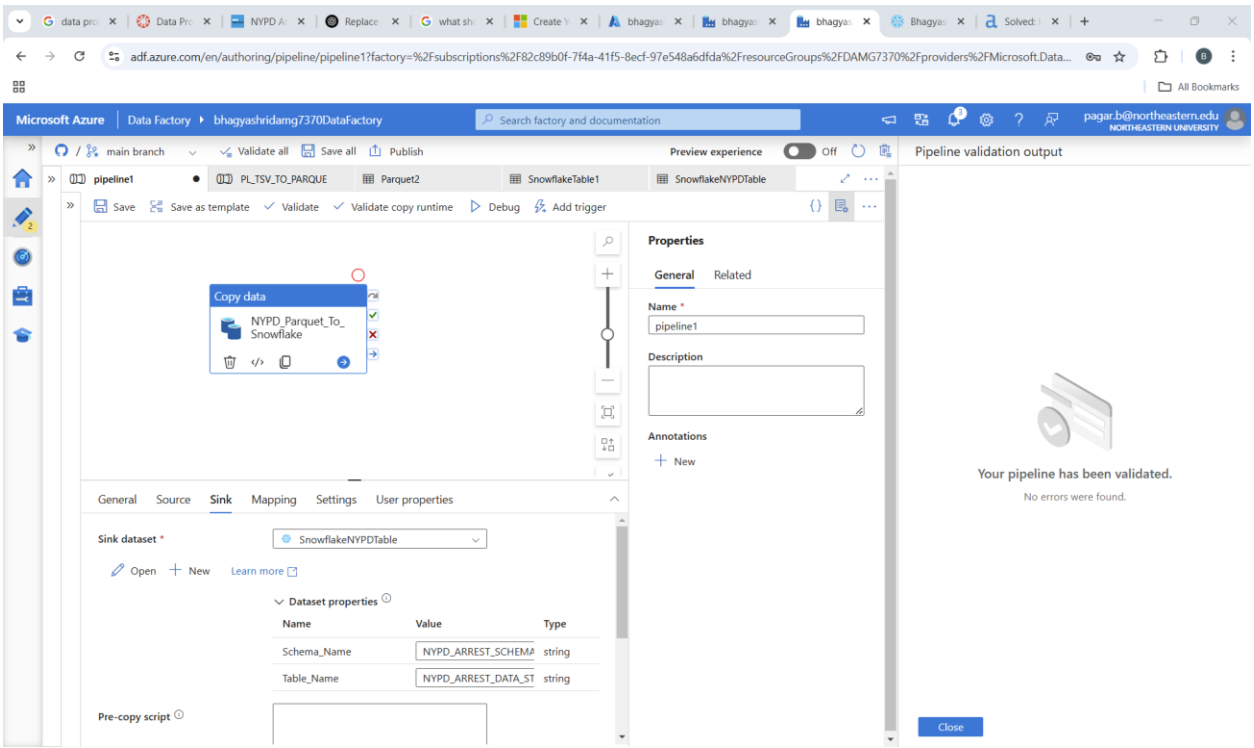


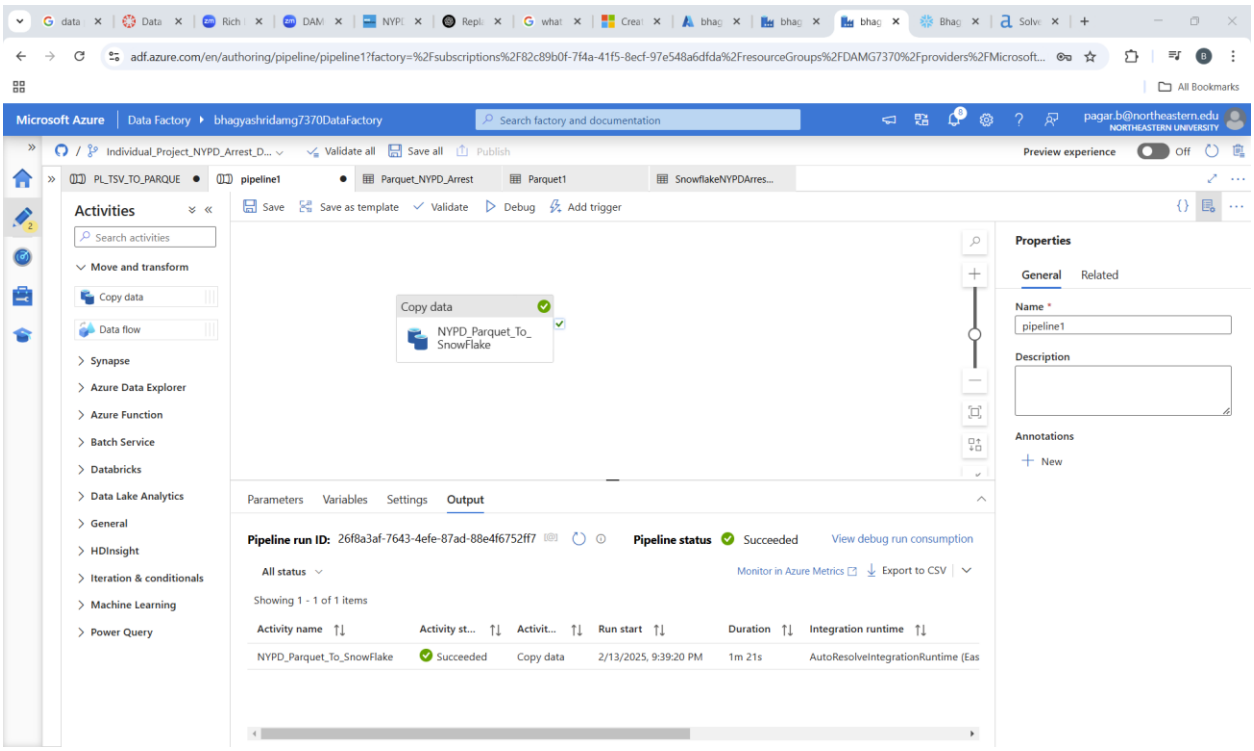
NYPD_ARREST_DATA_STAGE Table is Created in the snowflake

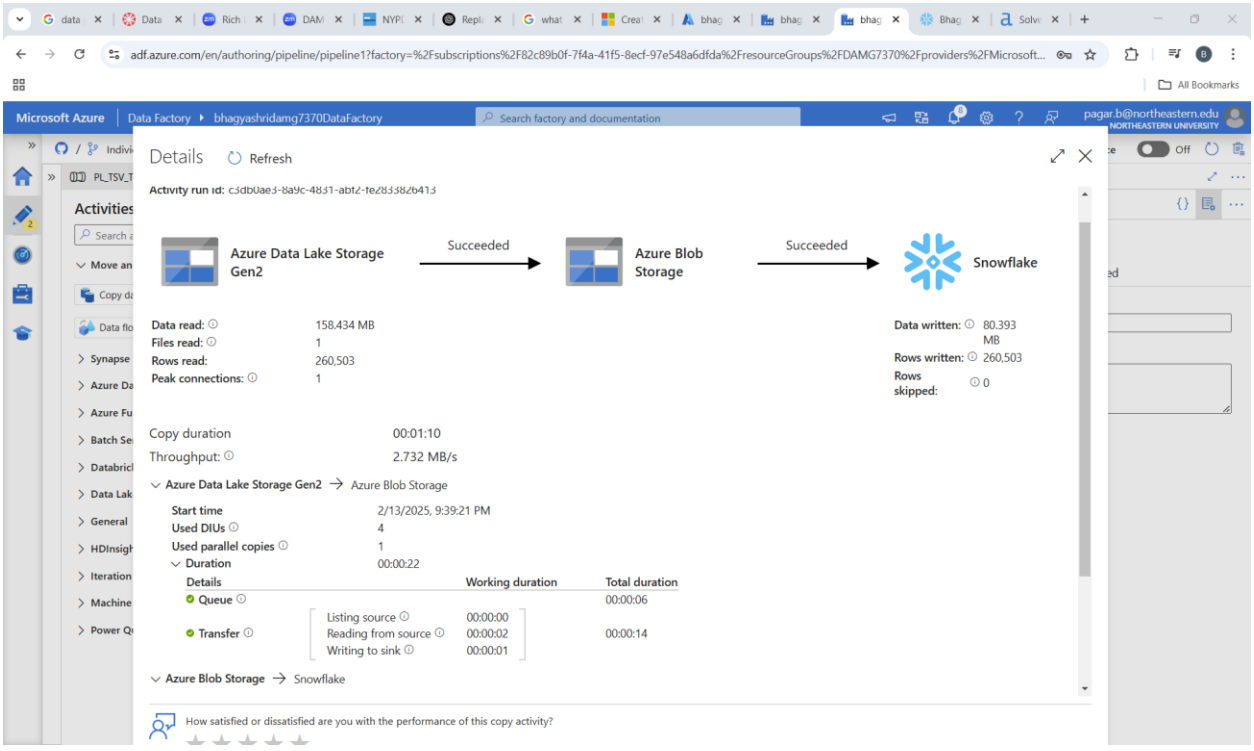


ADF Pipeline Run

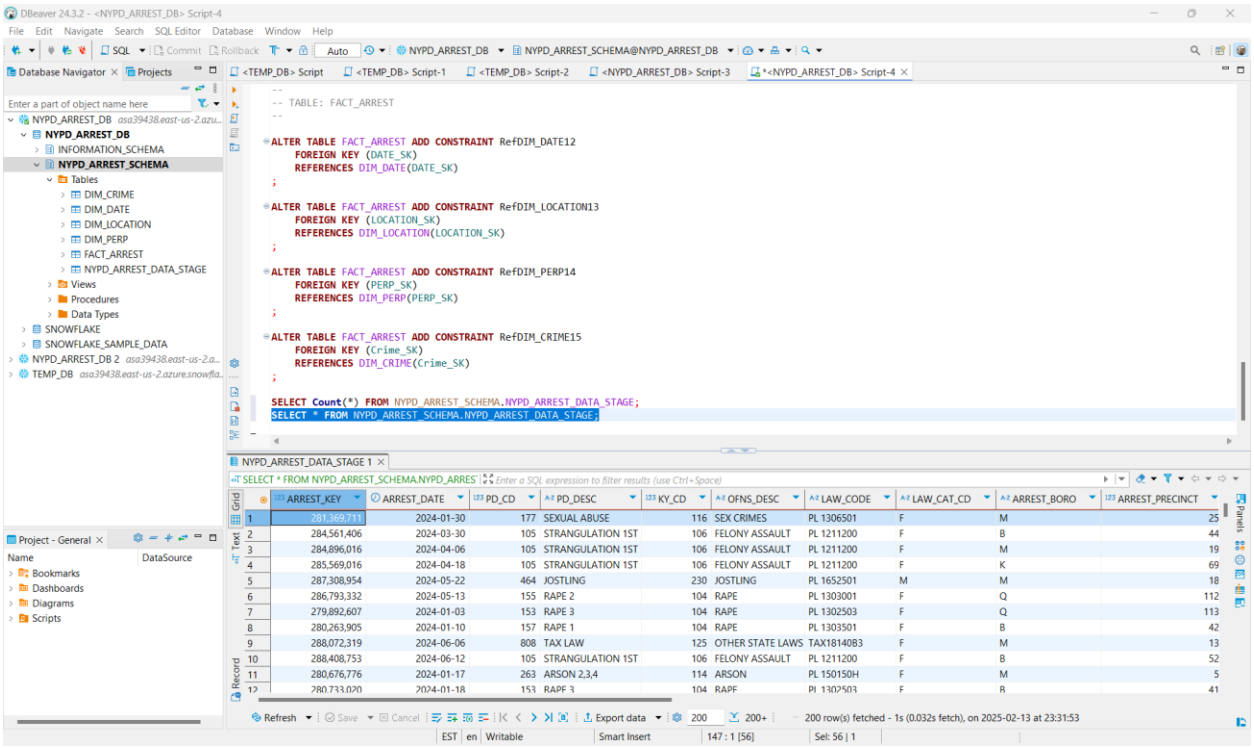
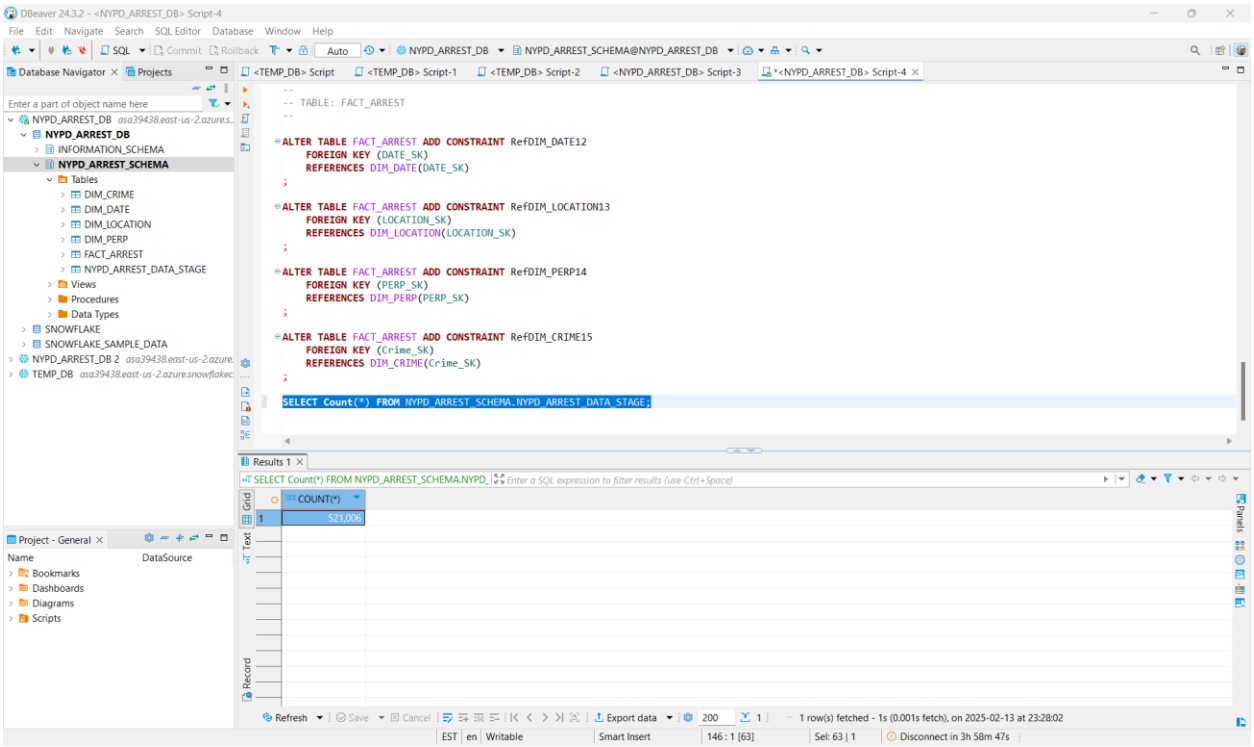
In this I have directly loaded cleaned data from alteryx in parquet file format to snowflake database.




Microsoft Azure
Data Factory
bhagyashridamg7370DataFactory
Search factory and documentation
pagar.b@northeastern.edu
NORTHEASTERN UNIVERSITY
Individual_Project_NYPD_Arrest_D...
Validate all
Save all
Publish
Preview experience
Off
PL_TSV_TO_PARQUE
pipeline1
Parquet_NYPD_Arrest
Parquet1
SnowflakeNYPDArres...
Activities
Search activities
Move and transform
Copy data
Data flow
Synapse
Azure Data Explorer
Azure Function
Batch Service
Databricks
Data Lake Analytics
General
HDInsight
Iteration & conditionals
Machine Learning
Power Query
Save
Save as template
Validate
Debug
Add trigger
Copy data
NYPD_Parquet_To_SnowFlake
Properties
General
Related
Name *
pipeline1
Description
Annotations
New
Parameters
Variables
Settings
Output
Pipeline run ID: 26f8a3af-7643-4efe-87ad-88e4f6752ff7
Pipeline status
Succeeded
View debug run consumption
All status
Monitor in Azure Metrics
Export to CSV
Showing 1 - 1 of 1 items
Activity name
Activity st...
Activit...
Run start
Duration
Integration runtime
NYPD_Parquet_To_SnowFlake
Succeeded
Copy data
2/13/2025, 9:39:20 PM
1m 21s
AutoResolveIntegrationRuntime (Eas


Microsoft Azure
Data Factory
bhagyashridamg7370DataFactory
Search factory and documentation
pagar.b@northeastern.edu
NORTHEASTERN UNIVERSITY
Details
Refresh
Activity run id:
Azure Data Lake Storage Gen2
Succeeded
Azure Blob Storage
Succeeded
Snowflake
Data read:
158.434 MB
Files read:
1
Rows read:
260,503
Peak connections:
1
Data written:
80.393 MB
Rows written:
260,503
Rows skipped:
0
Copy duration
00:01:10
Throughput:
2.732 MB/s
Azure Data Lake Storage Gen2
Azure Blob Storage
Start time
2/13/2025, 9:39:21 PM
Used DIUs
4
Used parallel copies
1
Duration
00:00:22
Details
Working duration
Total duration
Queue
00:00:06
Transfer
Listing source
00:00:00
Reading from source
00:00:02
Writing to sink
00:00:01
00:00:14
Azure Blob Storage
Snowflake
How satisfied or dissatisfied are you with the performance of this copy activity?

Data Populated in the snowflake



Row counts using SQL query: