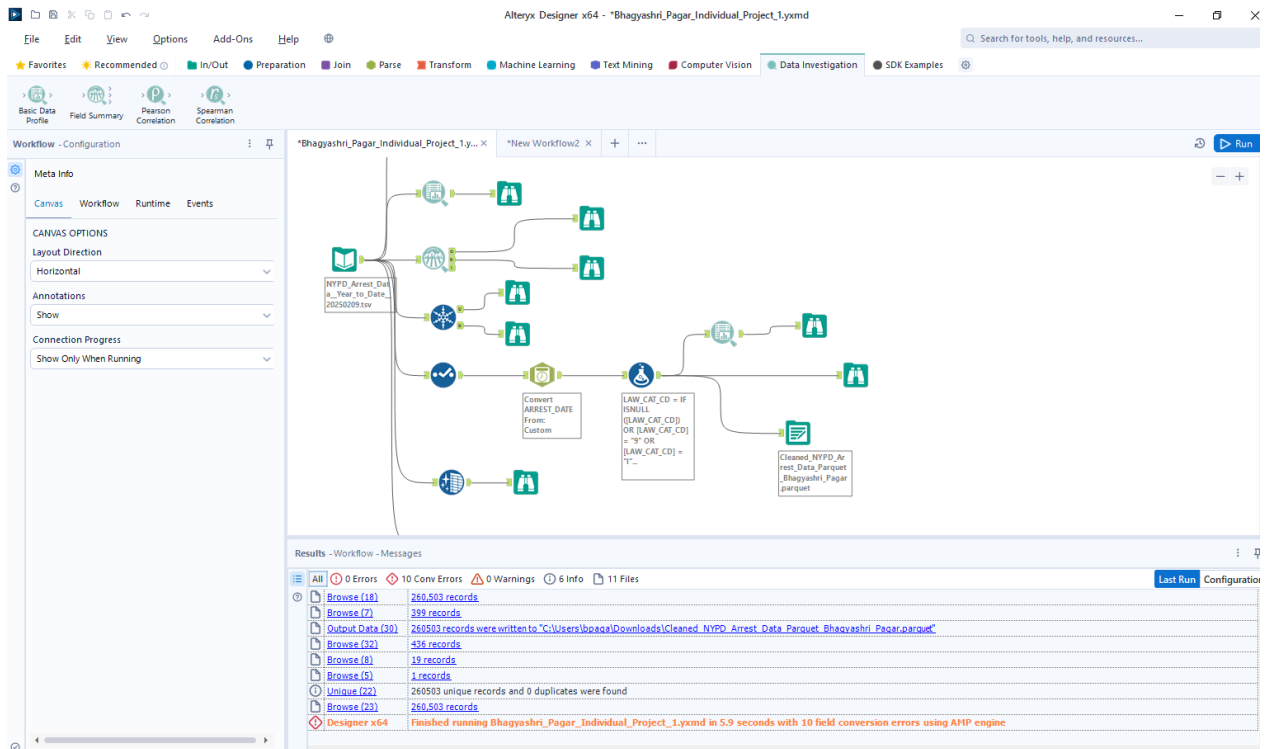


Alteryx, ADF Pipeline Run and Snowflake

Bhagyashri Avinash Pagar

NUID:002310690

Alteryx Workflow which is used for data cleaning and data profiling



Alteryx Designer v64 - *Bhagashri_Pagar_Individual_Project_1.yxmd

File Edit View Options Add-Ons Help

Search for tools, help, and resources...

Favorites Recommended In/Out Preparation Join Parse Transform Machine Learning Text Mining Computer Vision Data Investigation SDK Examples

Browse (18) - Configuration

Profile

260,503 records displayed, 20 fields, 20 MB

ARREST_KEY

279763507 1

279763792 1

279763800 1

279764505 1

279764507 1

995 more >

ARREST_DATE

03/14/2024 1,040

09/18/2024 1,036

10/02/2024 1,031

08/01/2024 997

11/14/2024 994

361 more >

PD_CD

101 28,202

339 27,107

109 15,612

922 13,254

478 12,265

263 more >

PD_DESC

*Bhagashri_Pagar_Individual_Project_1.yxmd

*New Workflow2

Run

Results - Browse (18) - Input

20 of 20 Fields Cell Viewer 260,503 records displayed, 20 MB

Record	ARREST_KEY	ARREST_DATE	PD_CD	PD_DESC	KY_CD	OFNS_DESC	LAW_CODE	LAW_CAT_CD	ARREST_BORO
1	281369711	01/30/2024	177	SEXUAL ABUSE	116	SEX CRIMES	PL 1306501	F	M
2	284561406	03/30/2024	105	STRANGULATION 1ST	106	FELONY ASSAULT	PL 1211200	F	B
3	284998016	04/06/2024	105	STRANGULATION 1ST	106	FELONY ASSAULT	PL 1211200	F	M
4	285569016	04/18/2024	105	STRANGULATION 1ST	106	FELONY ASSAULT	PL 1211200	F	K
5	287308954	05/22/2024	464	JOUSTING	230	JOUSTING	PL 1652501	M	M
6	286793332	05/13/2024	153	RAPE 2	104	RAPE	PL 1303001	F	Q
7	279892607	01/03/2024	153	RAPE 3	104	RAPE	PL 1302503	F	Q
8	280263905	01/10/2024	157	RAPE 1	104	RAPE	PL 1303501	F	B
9	288072319	06/06/2024	808	TAX LAW	125	OTHER STATE LAWS	TAX1814083	F	M

Snowflake Environment Setup

app.snowflake.com/east-us-2.azure/asa39438/#/account/users/NYPD_USER

Home Search Projects Data Data Products AI & ML Monitoring Admin Cost Management Warehouses Compute Pools Users & Roles Accounts Security Contacts Billing & Terms

\$397 credits left Trial ends in 112 days

NYPD_USER

About

Login Name: NYPD_USER

Display Name: NYPD_USER

Default Warehouse: DAMG7370_WH

Default Namespace: —

Last Login: —

Status: Enabled

User Type: —

Default Role: NYPD_ROLE

MFA: false

Roles: Granted roles (1)

Privileges

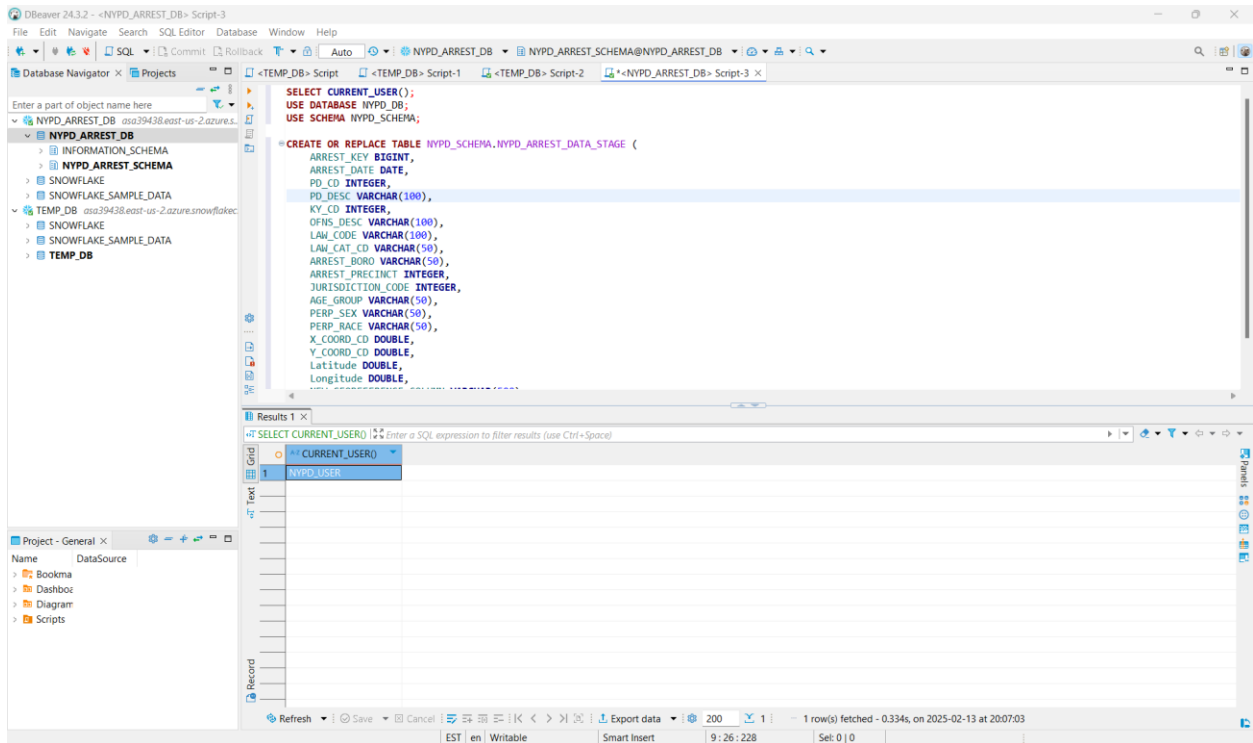
Group by Role + Privilege

ACCOUNTADMIN (Current Role) OWNERSHIP

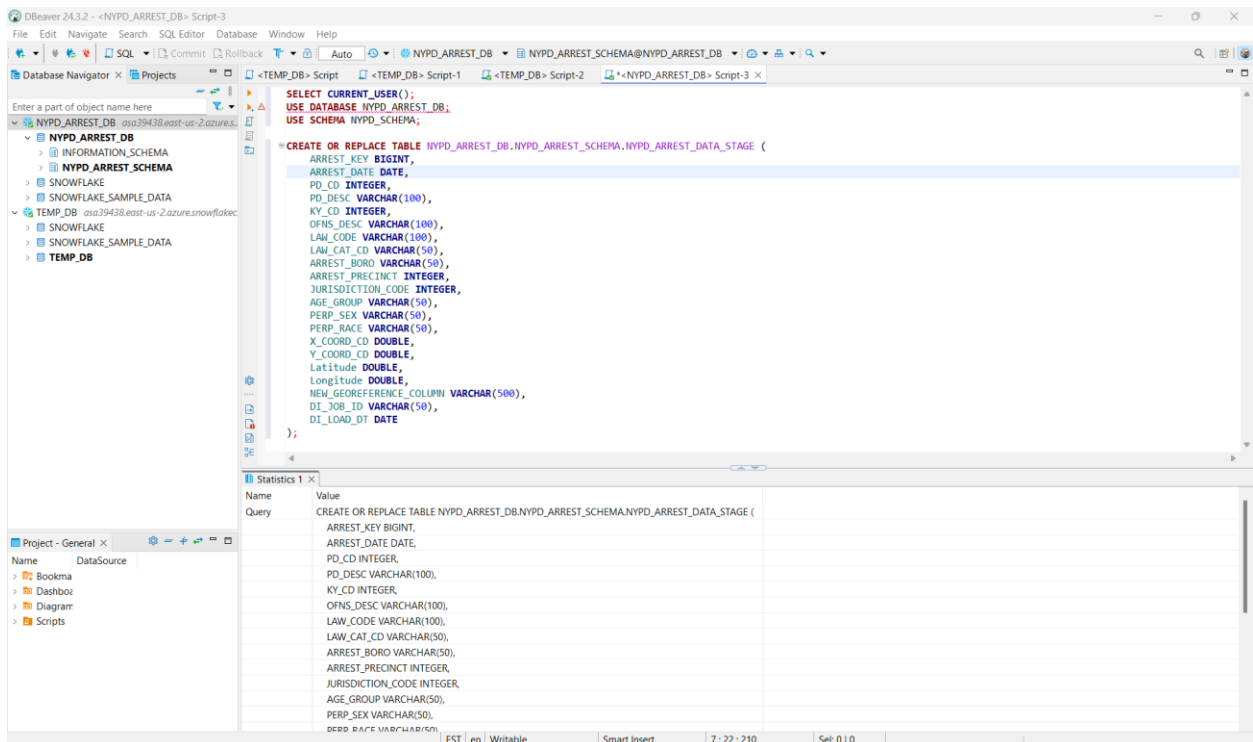
Granted Roles

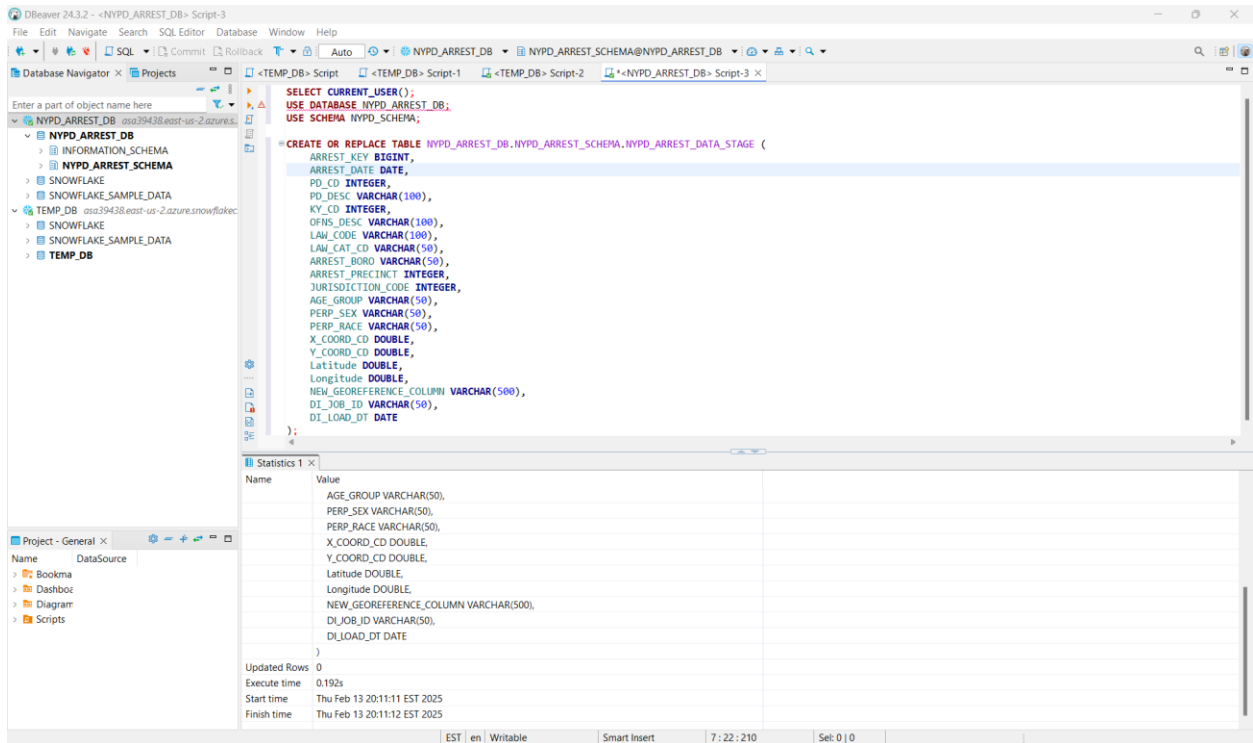
NYPD_USER has been granted 1 role

NAME	GRANTED ON	GRANTED BY
NYPD_ROLE	2/13/25	ACCOUNTADMIN

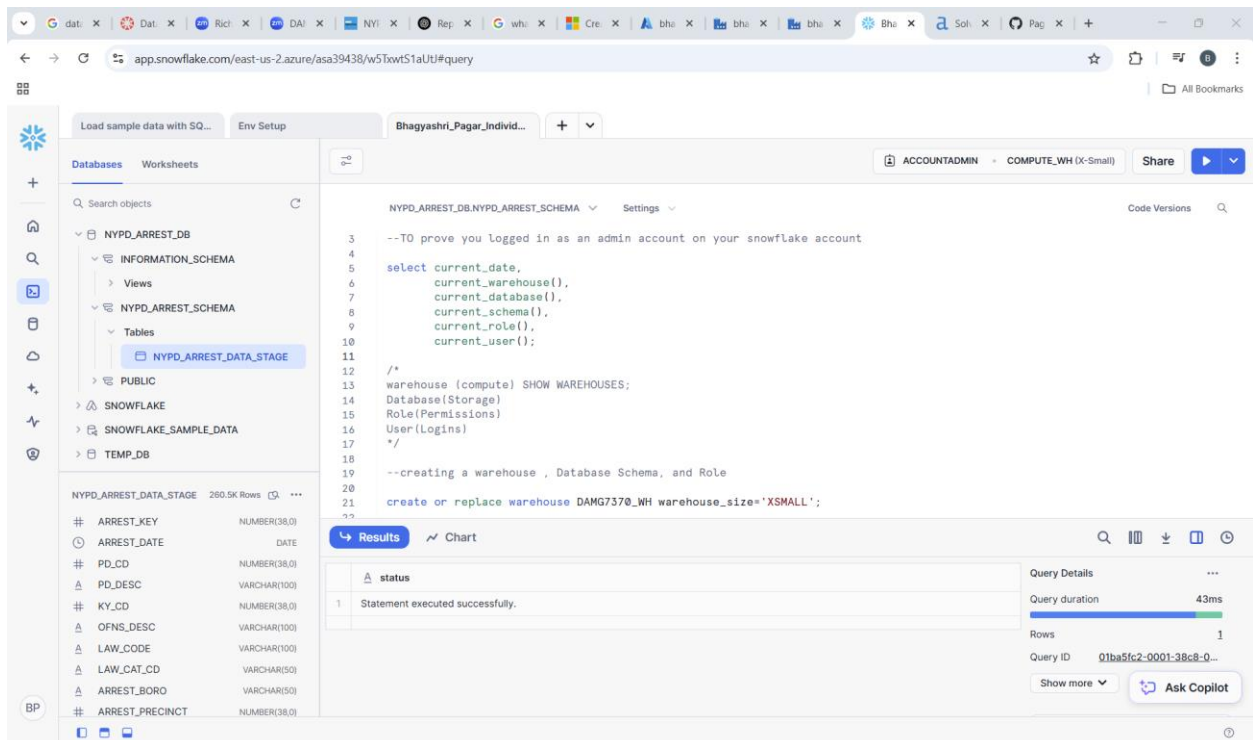


Snowflake NYPD_ARREST_SCHEMA is created in snowflake



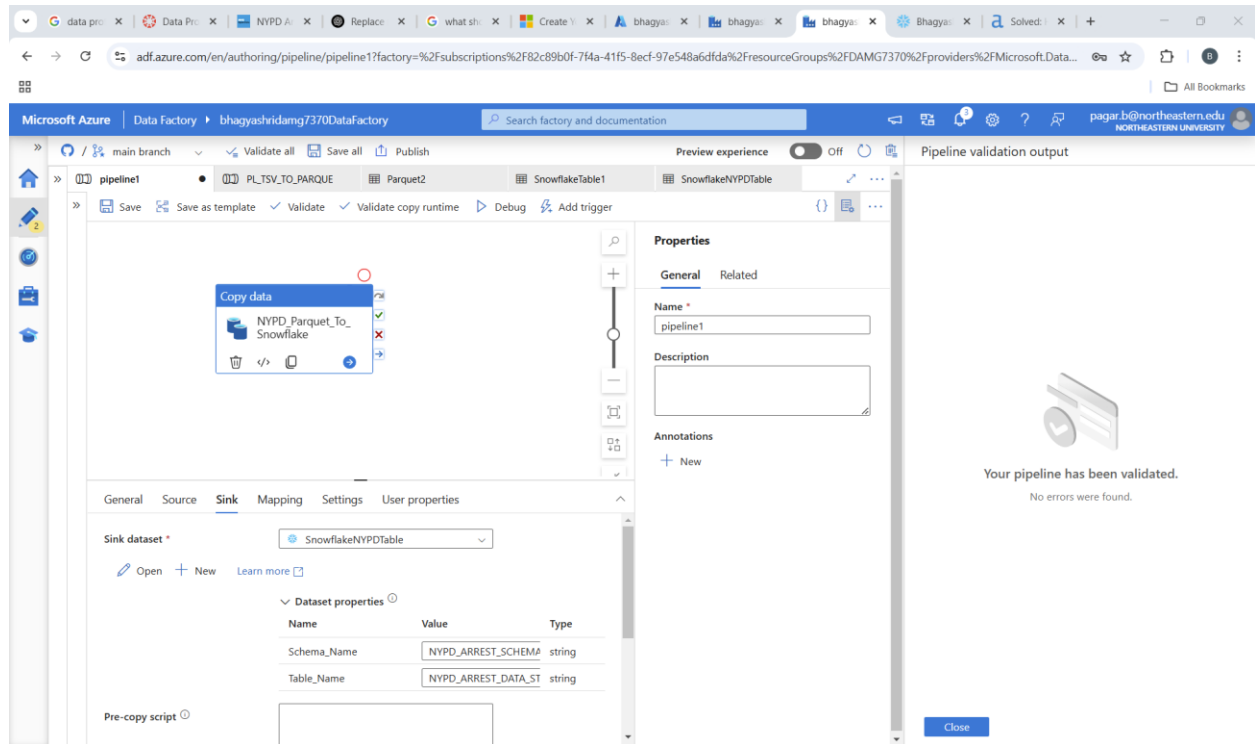


NYPD_ARREST_DATA_STAGE Table is Created in the snowflake



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

Validate all Save all Publish Preview experience Off Pipeline validation output

pipeline1 PL_TSV_TO_PARQUE Parquet2 SnowflakeTable1 SnowflakeNYPDTable

Copy data

NYPD_Parquet_To_Snowflake

Sink dataset * SnowflakeNYPDTable

Open + New Learn more

Dataset properties

Name	Value	Type
Schema_Name	NYPD_ARREST_SCHEME	string
Table_Name	NYPD_ARREST_DATA_ST	string

Pre-copy script

Properties

General Related

Name * pipeline1

Description

Annotations + New

Your pipeline has been validated. No errors were found.

Close

Microsoft Azure | Data Factory | bhagyashridamg/7370DataFactory

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

Copy data

NYPD_Parquet_To_Snowflake

Parameters Variables Settings **Output**

Pipeline run ID: 26f8a3af-7643-4efe-87ad-88e4f6752ff7 Pipeline status: Succeeded

All status

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

Properties

General Related

Name: pipeline1

Description

Annotations

+ New

Microsoft Azure | Data Factory | bhagyashridamg/7370DataFactory

Details Refresh

Activity run id: c5db0ae3-8a9c-48d1-ab72-1e28d38d6413

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

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 DILUs: 4
Used parallel copies: 1
Duration: 00:00:22

Details

	Working duration	Total duration
Queue	00:00:00	00:00:06
Transfer	00:00:14	00:00:14

Azure Blob Storage → Snowflake

Data written: 80,393 MB
Rows written: 260,503
Rows skipped: 0

How satisfied or dissatisfied are you with the performance of this copy activity?

Data Populated in the snowflake

The screenshot shows the DBeaver 24.3.2 interface with the following SQL scripts:

```
--  
-- TABLE: FACT_ARREST  
  
--  
-- ALTER TABLE FACT_ARREST ADD CONSTRAINT RefDIM_DATE12  
-- FOREIGN KEY (DATE_SK)  
-- REFERENCES DIM_DATE (DATE_SK)  
--  
-- ALTER TABLE FACT_ARREST ADD CONSTRAINT RefDIM_LOCATION13  
-- FOREIGN KEY (LOCATION_SK)  
-- REFERENCES DIM_LOCATION (LOCATION_SK)  
--  
-- ALTER TABLE FACT_ARREST ADD CONSTRAINT RefDIM_PERP14  
-- FOREIGN KEY (PERP_SK)  
-- REFERENCES DIM_PERP (PERP_SK)  
--  
-- ALTER TABLE FACT_ARREST ADD CONSTRAINT RefDIM_CRIME15  
-- FOREIGN KEY (CRIME_SK)  
-- REFERENCES DIM_CRIME (CRIME_SK)  
  
SELECT Count(*) FROM NYPD_ARREST_SCHEMA.NYPD_ARREST_DATA_STAGE;  
SELECT * FROM NYPD_ARREST_SCHEMA.NYPD_ARREST_DATA_STAGE;
```

The query results show the following data:

ARREST_KEY	ARREST_DATE	PD_CD	PD_DESC	KY_CD	OFFNS_DESC	LAW_CODE	LAW_CAT_CD	ARREST_BORO	ARREST_PRECINCT
281,369,731	2024-01-30	177	SEXUAL ABUSE	116	SEX CRIMES	PL 1306501	F	M	25
284,561,406	2024-03-30	105	STRANGULATION 1ST	106	FELONY ASSAULT	PL 1211200	F	B	44
284,896,016	2024-04-06	105	STRANGULATION 1ST	106	FELONY ASSAULT	PL 1211200	F	M	19
285,569,016	2024-04-18	105	STRANGULATION 1ST	106	FELONY ASSAULT	PL 1211200	F	K	69
287,308,954	2024-05-22	464	JOSTLING	230	JOSTLING	PL 1652501	M	M	18
286,793,332	2024-05-13	155	RAPE 2	104	RAPE	PL 1303001	F	Q	112
279,892,607	2024-01-03	153	RAPE 3	104	RAPE	PL 1302503	F	Q	113
280,263,905	2024-01-10	157	RAPE 1	104	RAPE	PL 1303501	F	B	42
288,072,319	2024-06-06	808	TAX LAW	125	OTHER STATE LAWS	TAX18140B3	F	M	13
288,408,753	2024-06-12	105	STRANGULATION 1ST	106	FELONY ASSAULT	PL 1211200	F	B	52
280,676,776	2024-01-17	263	ARSON 2,3,4	114	ARSON	PL 150150H	F	M	5
280,733,020	2024-01-18	153	RAPE 3	104	RAPE	PL 1302503	F	R	41

Row counts using SQL query:

The screenshot shows the DBeaver 24.3.2 interface with the following SQL query:

```
--  
-- TABLE: FACT_ARREST  
  
--  
-- ALTER TABLE FACT_ARREST ADD CONSTRAINT RefDIM_DATE12  
-- FOREIGN KEY (DATE_SK)  
-- REFERENCES DIM_DATE (DATE_SK)  
--  
-- ALTER TABLE FACT_ARREST ADD CONSTRAINT RefDIM_LOCATION13  
-- FOREIGN KEY (LOCATION_SK)  
-- REFERENCES DIM_LOCATION (LOCATION_SK)  
--  
-- ALTER TABLE FACT_ARREST ADD CONSTRAINT RefDIM_PERP14  
-- FOREIGN KEY (PERP_SK)  
-- REFERENCES DIM_PERP (PERP_SK)  
--  
-- ALTER TABLE FACT_ARREST ADD CONSTRAINT RefDIM_CRIME15  
-- FOREIGN KEY (CRIME_SK)  
-- REFERENCES DIM_CRIME (CRIME_SK)  
  
SELECT Count(*) FROM NYPD_ARREST_SCHEMA.NYPD_ARREST_DATA_STAGE;
```

The query results show the following data:

COUNT(*)
521,006