DADABI

Midterm Project

Food Inspections Chicago & Dallas

Presented by

Apoorv Dhaygude
Sangram Shinde
Kunal Tibe
Shraddha Bhandarkar

Table Of Content

Project Overview

The dataset from Chicago includes detailed records of inspections conducted at restaurants and other food service establishments, spanning from January 1, 2010, to the present day. These inspections are undertaken by personnel from the Food Protection Program of the Chicago Department of Public Health, who adhere to a standardized inspection protocol. The outcomes of these inspections are meticulously documented.

In contrast, the dataset from Dallas aims to convey essential information regarding food service establishments. It covers the name and physical location of each establishment, the date when the inspection took place, the overall score achieved during the inspection, and detailed deductions for each violation identified.

Data Profiling and Understanding

Chicago:

Overview:

Overview | Alerts 8 | Reproduction

Dataset statistics

Number of variables	17
Number of observations	109020
Missing cells	44386
Missing cells (%)	2.4%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	14.1 MiB
Average record size in memory	136.0 B

Variable types

Numeric	5
Text	8
Categorical	3
DateTime	1

Observation:

The total number of records are 109020 with missing cells 44386 (2.4%).
There are 17 variables namely Inspection ID, DBA Name, AKA Name, License #, Facility Type, Risk, Address, City, State, Zip, Inspection Date, Inspection Type, Results, Violations, Latitude, Longitude, Location with variable types being 5 Numeric, 8 Text, 3 Categorical and 1 DateTime.

Variables:

1. Inspection ID is a variable with numeric data type which itself is Unique contains 109020 distinct values (100%) with zero missing values.

2. DBA Name is a variable with text data type which contains 23936 distinct values (22.0%) with zero missing values

3. AKA Name is a variable with numeric data type which contains 22732 distinct values (21.0%) with 662 missing values (0.6%)

Data Profiling and Understanding

Variables:

4. License # is a variable with numeric data type which contains 331377 distinct values (28.8%) with 7 missing values (<0.1%).

5. Facility Type is a variable with text data type which contains 385 distinct values (0.4%) with 1667 missing values (1.5%).

6. Risk is a variable with categorical data type which contains 4 distinct values (<0.1%) with 39 missing values (<0.1%).
7. Address is a variable with text data type which contains 18062 distinct values (16.6%) with 1 missing values (<0.1%).

8. City is a variable with text data type which contains 62 distinct values (0.1%) with 84 missing values (0.1%).

9. State is a variable with categorical data type which contains 5 distinct values (<0.1%) with 43 missing values (<0.1%).

10. Zip is a variable with numerical data type which contains 108 distinct values (0.1%) with 18 missing values (<0.1%).

11. Inspection Date is a variable with date data type which contains 3477 distinct values (3.2%) with 1 missing values (<0.1%).

12. Inspection Type is a variable with text data type which contains 53 distinct values (<0.1%) with 2 missing values (<0.1%).

Data Profiling and Understanding

Variables:

13. Results is a variable with categorical data type which contains 7 distinct values (<0.1%) with 1 missing values (<0.1%).

14.Violations is a variable with text data type which contains 67922 distinct values (99.2%) with 40526 missing values (37.2%).

15.Latitude is a variable with numeric data type which contains 15843 distinct values (14.6 %) with 445 missing values (0.4 %).

16. Longitude is a variable with numeric data type which contains 15843 distinct values (14.6%) with 445 missing values (0.4%).

17. Location is a variable with text data type which contains 15843 distinct values (14.6%) with 445 missing values (0.4%).

Data Profiling and Understanding

Dallas:

Overview:

Overview | Alerts 151 | Reproduction

Dataset statistics

Number of variables	114
Number of observations	6708
Missing cells	571706
Missing cells (%)	74.8%
Duplicate rows	1
Duplicate rows (%)	< 0.1%
Total size in memory	5.8 MiB
Average record size in memory	912.0 B

Variable types

Text	81
Categorical	29
DateTime	2
Numeric	2

Observation:

The total number of records are 6708 with missing cells 571706 (74.8%).
There are 114 variables namely Restaurant Name, Inspection Type , Inspection Date, Inspection Score, Street Number , Street Name, Street Direction , Street Type, Street Unit, Street Address, Zip Code, Violation Description ,Violation Points ,Violation Details , Violation Memo, Inspection Month, Inspection Year and Lat Long Location.

Variables:

1.Restaurant Name is a variable with text data type which contains 4022 distinct values (60.0%) with zero missing values.

2.Inspection Type is a variable with text data type which contains 3 distinct values (<0.1%) with 0 missing values.

3.Inspection Date is a variable with date data type which contains 915 distinct values (13.6%) with 0 missing values .

Data Profiling and Understanding

Variables:

4.Inspection Score is a variable with numeric data type which contains 43 distinct values (0.6%) with 0 missing values.

5.Street Number is a variable with numeric data type which contains 2336 distinct values (34.8%) with 0 missing values.

6.Street Name is a variable with text data type which contains 607 distinct values (9.0%) with 0 missing values.

7.Street Direction is a variable with categorical data type which contains 4 distinct values (0.2%) with 4521 missing values (67.8%).

8.Street Type is a variable with numeric data type which contains 18 distinct values (0.3 %) with 140 missing values (2.1%).

9.Street Unit is a variable with Numeric data type which contains 512 distinct values (22.3%) with 4417 missing values (65.8%).

10.Street Address is a variable with text data type which contains 3903 distinct values (58.2%) with 0 missing values.

11.Zip Code is a variable with text data type which contains 89 distinct values (1.3%) with 0 missing values.

12.Violation Description is a variable with text data type which contains detailed violations.

Data Profiling and Understanding

Variables:

13.Violation Point is a variable with categorical data type which contains violation points given per violation.

14.Inspection Month is a variable with date data type which contains 85 distinct values (1.4%) with 1 missing value (<0.1%).

15.Inspection Year is a variable with categorical data type which contains 8 distinct values (0.1%) with 1 missing value (<0.1%).

16.Lat Long Location is a variable with text data type which contains 4108 distinct values (61.2%) with 1 missing value (<0.1%).

Business Requirements

Examine food inspection results by:

- Inspection Type
- Inspection Result
- Risk Category
- Facility Type
- Violations (Code, Descriptions)
- Business Impacted (DBA (Doing Business As), AKA (Also Know As),
- License

Inspection results:

- All of above with inspection #, license #, violations & inspector comments

Data Preparation (Part 1)

In the process of refining the Dallas dataset for enhanced analysis and integration, the following steps were undertaken to ensure data quality and consistency:

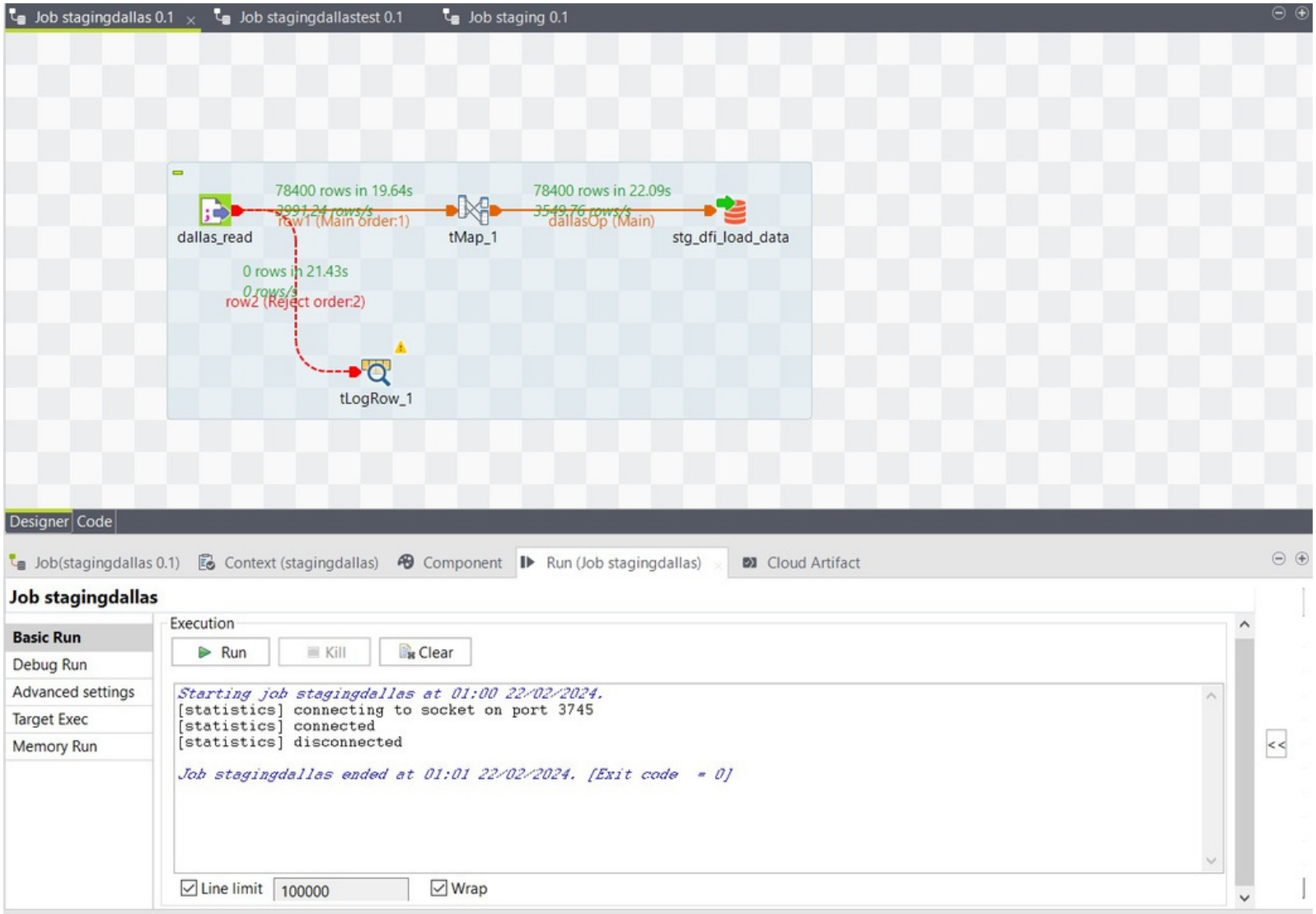
- **Normalization of Geographical Coordinates:**
** - Encountered a column labeled "lat_long_location" with coordinates in an inconsistent format.**
** - Utilized regex to reformat these coordinates, standardizing the data across the dataset.**

- **Separation of Latitude and Longitude:**
** - Used Talend to split the latitude and longitude data into separate columns, enhancing the dataset's granularity for precise geographical analysis during visualizations.**

- **Harmonization of Datasets:**
** - Addressed disparities between the Dallas and Chicago datasets by introducing equivalent columns in the Dallas stage table, populated with default values like "Dallas" for the city and "TX" for the state, ensuring geographic specificity and uniform data structure.**

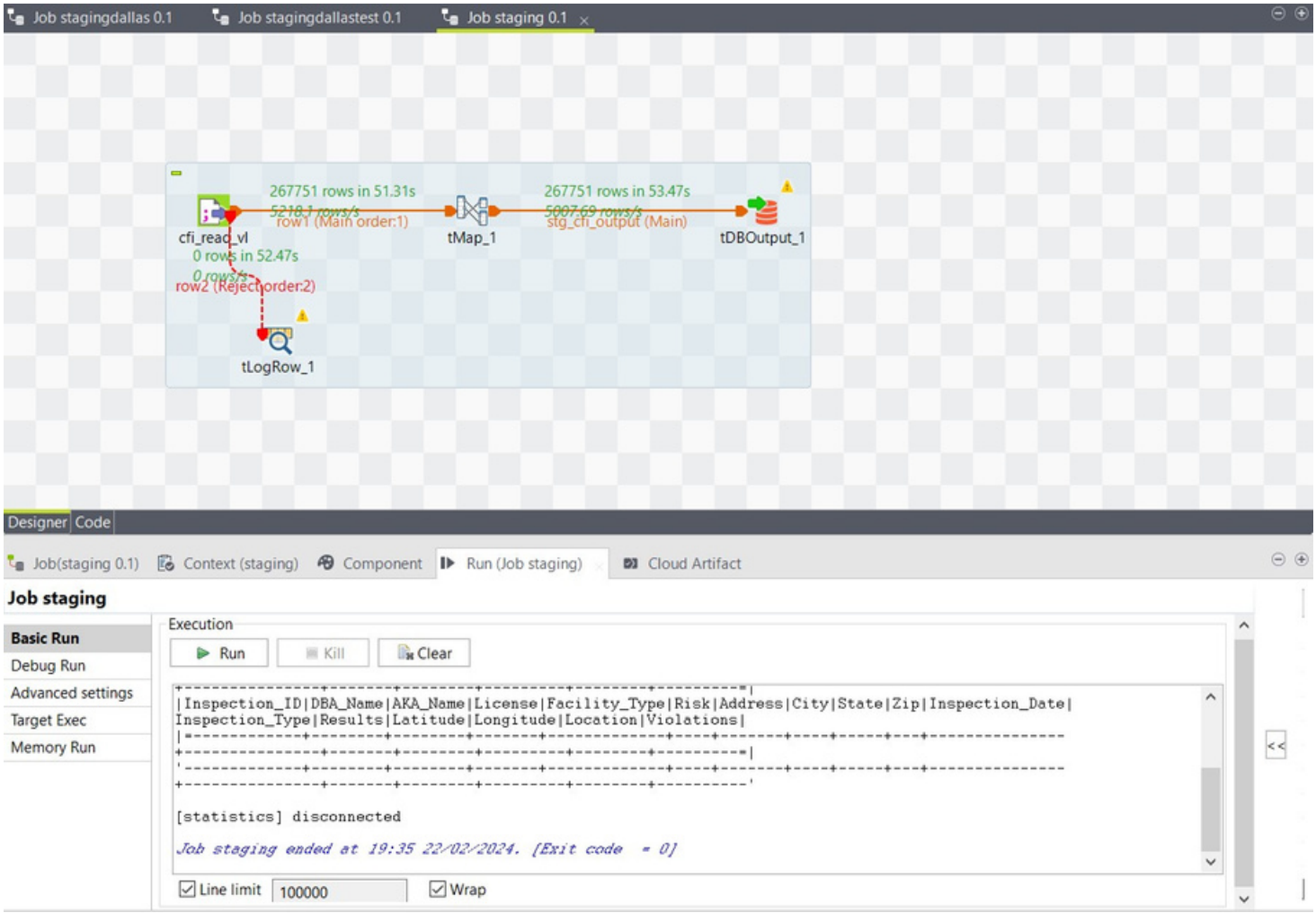
Data Preparation (Part 1)

Dallas: Talend Workflow



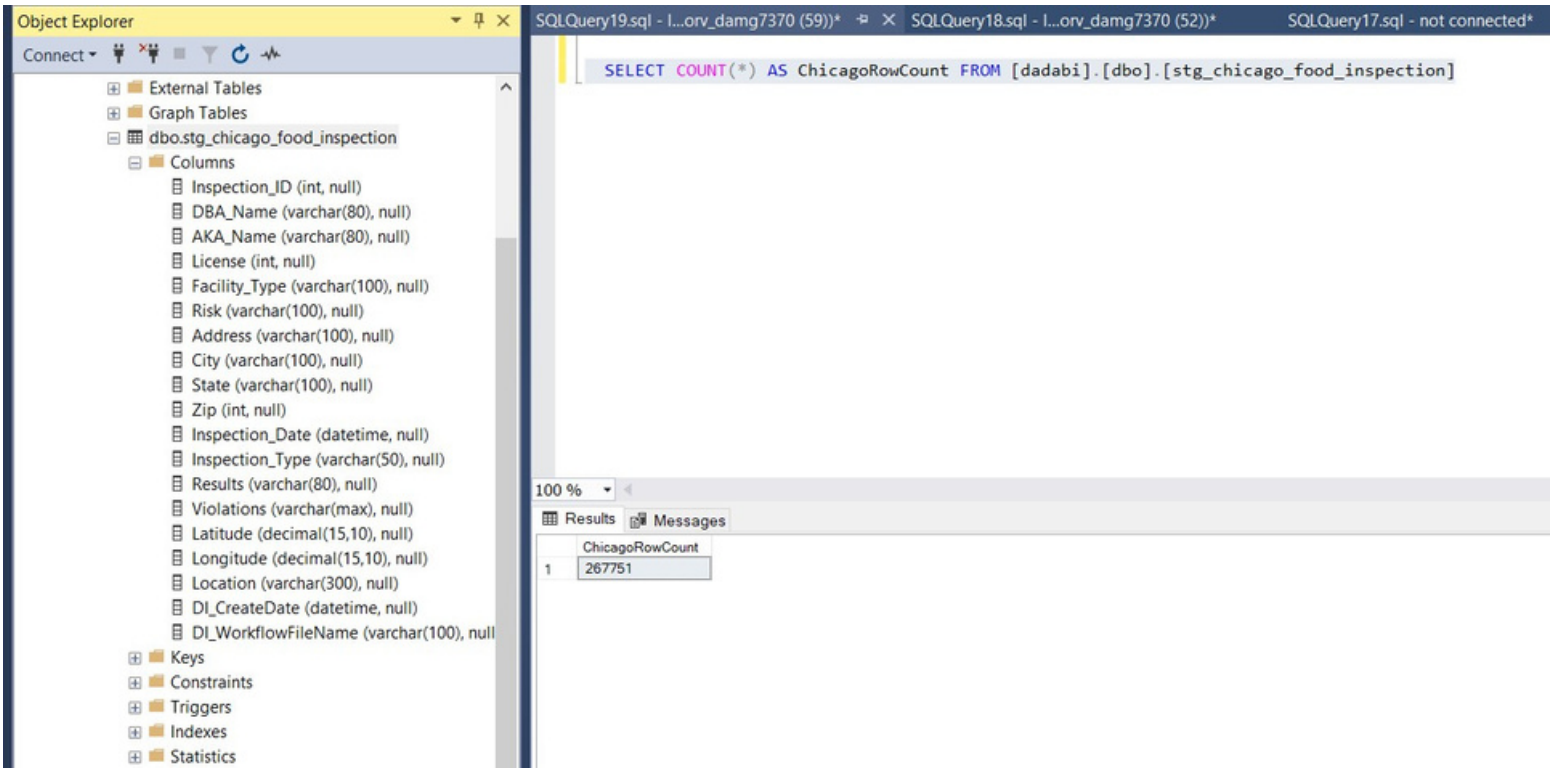
Data Preparation (Part 1)

Chicago : Worflow in Talend

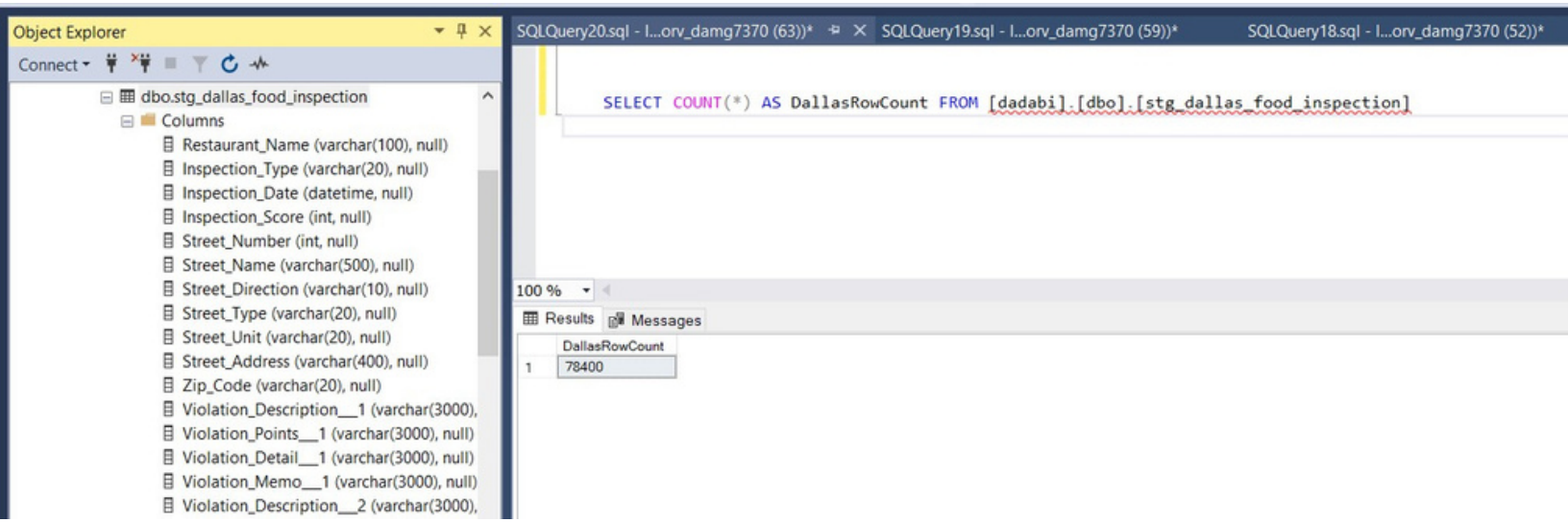


Data Preparation (Part 1)

Chicago: Row Count



Dallas: Row Count



Data Preparation (Part 1)

Dallas: Azure SQL Output

```
      ,[Violation_Description___23]
      ,[Violation_Points___23]
      ,[Violation_Detail___23]
      ,[Violation_Memo___23]
      ,[Violation_Description___24]
      ,[Violation_Points___24]
      ,[Violation_Detail___24]
      ,[Violation_Memo___24]
      ,[Violation_Description___25]
      ,[Violation_Points___25]
      ,[Violation_Detail___25]
      ,[Violation_Memo___25]
      ,[Inspection_Month]
      ,[Inspection_Year]
      ,[Lat_Long_Location]
      ,[DI_CreateDate]
      ,[DI_WorkflowFileName]
  FROM [dadabi].[dbo].[stg_dallas_food_inspection]
```

100 %

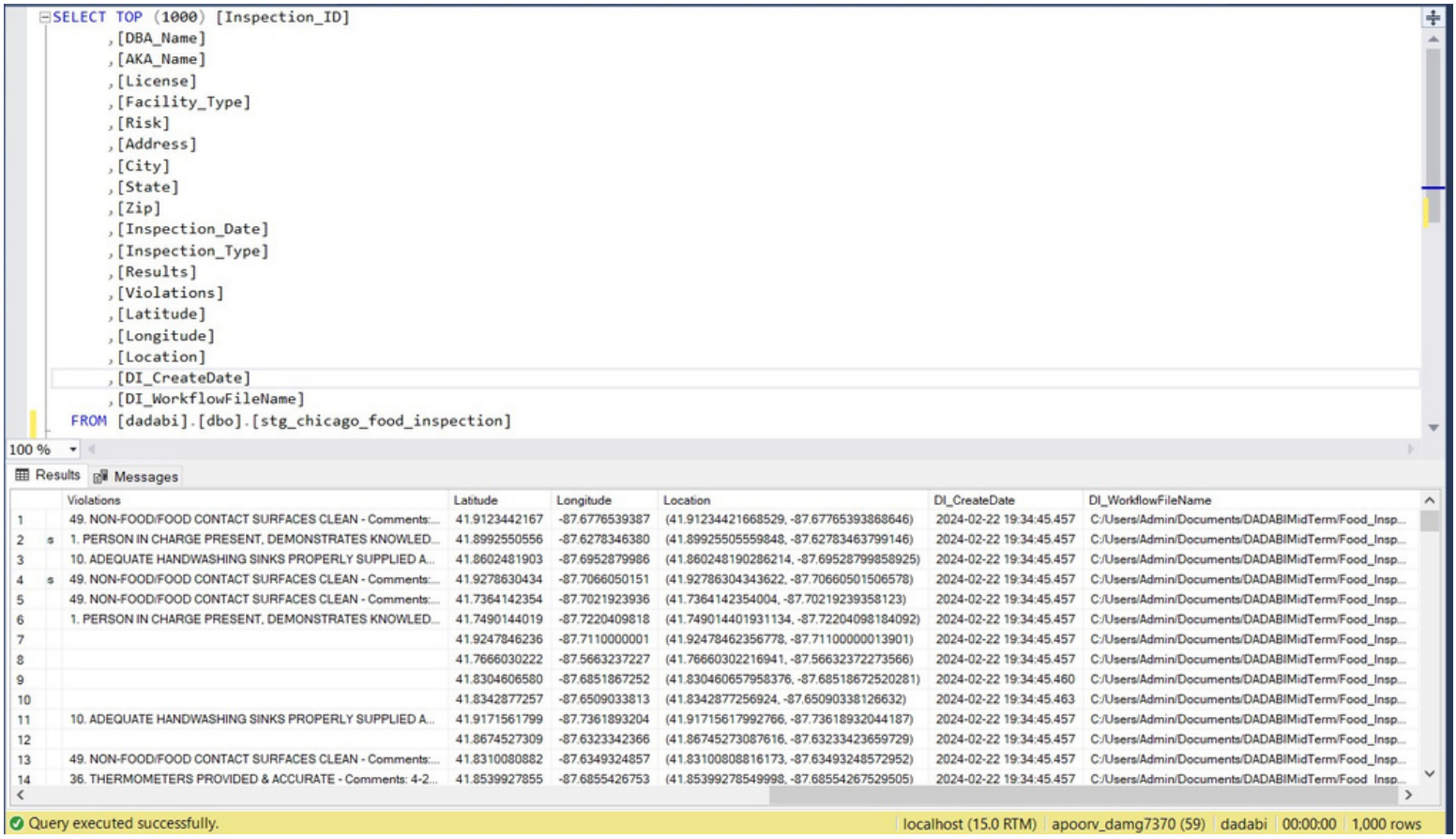
Results | Messages

	scription___25	Violation_Points___25	Violation_Detail___25	Violation_Memo___25	Inspection_Month	Inspection_Year	Lat_Long_Location	DI_CreateDate	DI_WorkflowFileName
1					Dec 2020	FY2021	39.934878, -84.370281	2024-02-22 01:00:42.973	C:\Users\Admin\Documents\DADABIMidTerm\Cleaned_D...
2					Jan 2022	FY2022	32.896003, -96.722504	2024-02-22 01:00:42.973	C:\Users\Admin\Documents\DADABIMidTerm\Cleaned_D...
3					Feb 2022	FY2022	38.922717, -75.79054	2024-02-22 01:00:42.973	C:\Users\Admin\Documents\DADABIMidTerm\Cleaned_D...
4					Jul 2020	FY2020	32.761505, -96.857818	2024-02-22 01:00:42.973	C:\Users\Admin\Documents\DADABIMidTerm\Cleaned_D...
5					Aug 2018	FY2018	32.851172, -96.816318	2024-02-22 01:00:42.977	C:\Users\Admin\Documents\DADABIMidTerm\Cleaned_D...
6					Apr 2021	FY2021	32.656463, -96.750349	2024-02-22 01:00:42.977	C:\Users\Admin\Documents\DADABIMidTerm\Cleaned_D...
7					Apr 2020	FY2020	32.85442, -96.730528	2024-02-22 01:00:42.980	C:\Users\Admin\Documents\DADABIMidTerm\Cleaned_D...
8					Feb 2018	FY2018	32.720331, -96.828492	2024-02-22 01:00:42.977	C:\Users\Admin\Documents\DADABIMidTerm\Cleaned_D...
9					May 2018	FY2018	32.875308, -96.762062	2024-02-22 01:00:42.977	C:\Users\Admin\Documents\DADABIMidTerm\Cleaned_D...
10					Oct 2021	FY2022	32.734166, -96.677554	2024-02-22 01:00:43.060	C:\Users\Admin\Documents\DADABIMidTerm\Cleaned_D...
11					Oct 2019	FY2020	32.919456, -96.751347	2024-02-22 01:00:42.977	C:\Users\Admin\Documents\DADABIMidTerm\Cleaned_D...
12					Mar 2020	FY2020	32.789624, -96.700127	2024-02-22 01:00:43.093	C:\Users\Admin\Documents\DADABIMidTerm\Cleaned_D...
13					Mar 2020	FY2020	32.833226, -96.829073	2024-02-22 01:00:43.127	C:\Users\Admin\Documents\DADABIMidTerm\Cleaned_D...
14					Nov 2020	FY2021	32.732539, -96.648457	2024-02-22 01:00:42.980	C:\Users\Admin\Documents\DADABIMidTerm\Cleaned_D...
15					Mar 2020	FY2020	32.813867, -96.752756	2024-02-22 01:00:43.093	C:\Users\Admin\Documents\DADABIMidTerm\Cleaned_D...
16					Mar 2020	FY2020	32.92416, -96.838315	2024-02-22 01:00:43.127	C:\Users\Admin\Documents\DADABIMidTerm\Cleaned_D...

Query executed successfully. | localhost (15.0 RTM) | apoorv_damg7370 (63) | dadabi | 00:00:00 | 1,000 rows

Data Preparation (Part 1)

Chicago: Azure SQL Output



```
SELECT TOP (1000) [Inspection_ID]
      ,[DBA_Name]
      ,[AKA_Name]
      ,[License]
      ,[Facility_Type]
      ,[Risk]
      ,[Address]
      ,[City]
      ,[State]
      ,[Zip]
      ,[Inspection_Date]
      ,[Inspection_Type]
      ,[Results]
      ,[Violations]
      ,[Latitude]
      ,[Longitude]
      ,[Location]
      ,[DI_CreateDate]
      ,[DI_WorkflowFileName]
  FROM [dadabi].[dbo].[stg_chicago_food_inspection]
```

100 %

Results | Messages

	Violations	Latitude	Longitude	Location	DI_CreateDate	DI_WorkflowFileName
1	49. NON-FOOD/FOOD CONTACT SURFACES CLEAN - Comments:...	41.9123442167	-87.6776539387	(41.91234421668529, -87.67765393868646)	2024-02-22 19:34:45.457	C:/Users/Admin/Documents/DADABIMidTerm/Food_Insp...
2	1. PERSON IN CHARGE PRESENT, DEMONSTRATES KNOWLED...	41.8992550556	-87.6278346380	(41.89925505559848, -87.62783463799146)	2024-02-22 19:34:45.457	C:/Users/Admin/Documents/DADABIMidTerm/Food_Insp...
3	10. ADEQUATE HANDWASHING SINKS PROPERLY SUPPLIED A...	41.8602481903	-87.6952879986	(41.860248190286214, -87.69528799858925)	2024-02-22 19:34:45.457	C:/Users/Admin/Documents/DADABIMidTerm/Food_Insp...
4	49. NON-FOOD/FOOD CONTACT SURFACES CLEAN - Comments:...	41.9278630434	-87.7066050151	(41.92786304343622, -87.70660501506578)	2024-02-22 19:34:45.457	C:/Users/Admin/Documents/DADABIMidTerm/Food_Insp...
5	49. NON-FOOD/FOOD CONTACT SURFACES CLEAN - Comments:...	41.7364142354	-87.7021923936	(41.7364142354004, -87.70219239358123)	2024-02-22 19:34:45.457	C:/Users/Admin/Documents/DADABIMidTerm/Food_Insp...
6	1. PERSON IN CHARGE PRESENT, DEMONSTRATES KNOWLED...	41.7490144019	-87.7220409818	(41.749014401931134, -87.72204098184092)	2024-02-22 19:34:45.457	C:/Users/Admin/Documents/DADABIMidTerm/Food_Insp...
7		41.9247846236	-87.7110000001	(41.92478462356778, -87.71100000013901)	2024-02-22 19:34:45.457	C:/Users/Admin/Documents/DADABIMidTerm/Food_Insp...
8		41.7666030222	-87.5663237227	(41.76660302216941, -87.56632372273566)	2024-02-22 19:34:45.457	C:/Users/Admin/Documents/DADABIMidTerm/Food_Insp...
9		41.8304606580	-87.6851867252	(41.830460657958376, -87.68518672520281)	2024-02-22 19:34:45.460	C:/Users/Admin/Documents/DADABIMidTerm/Food_Insp...
10		41.8342877257	-87.6509033813	(41.8342877256924, -87.65090338126632)	2024-02-22 19:34:45.463	C:/Users/Admin/Documents/DADABIMidTerm/Food_Insp...
11	10. ADEQUATE HANDWASHING SINKS PROPERLY SUPPLIED A...	41.9171561799	-87.7361893204	(41.91715617992766, -87.73618932044187)	2024-02-22 19:34:45.457	C:/Users/Admin/Documents/DADABIMidTerm/Food_Insp...
12		41.8674527309	-87.6323342366	(41.86745273087616, -87.63233423659729)	2024-02-22 19:34:45.457	C:/Users/Admin/Documents/DADABIMidTerm/Food_Insp...
13	49. NON-FOOD/FOOD CONTACT SURFACES CLEAN - Comments:...	41.8310080882	-87.6349324857	(41.83100808816173, -87.63493248572952)	2024-02-22 19:34:45.457	C:/Users/Admin/Documents/DADABIMidTerm/Food_Insp...
14	36. THERMOMETERS PROVIDED & ACCURATE - Comments: 4-2...	41.8539927855	-87.6855426753	(41.85399278549998, -87.68554267529505)	2024-02-22 19:34:45.457	C:/Users/Admin/Documents/DADABIMidTerm/Food_Insp...

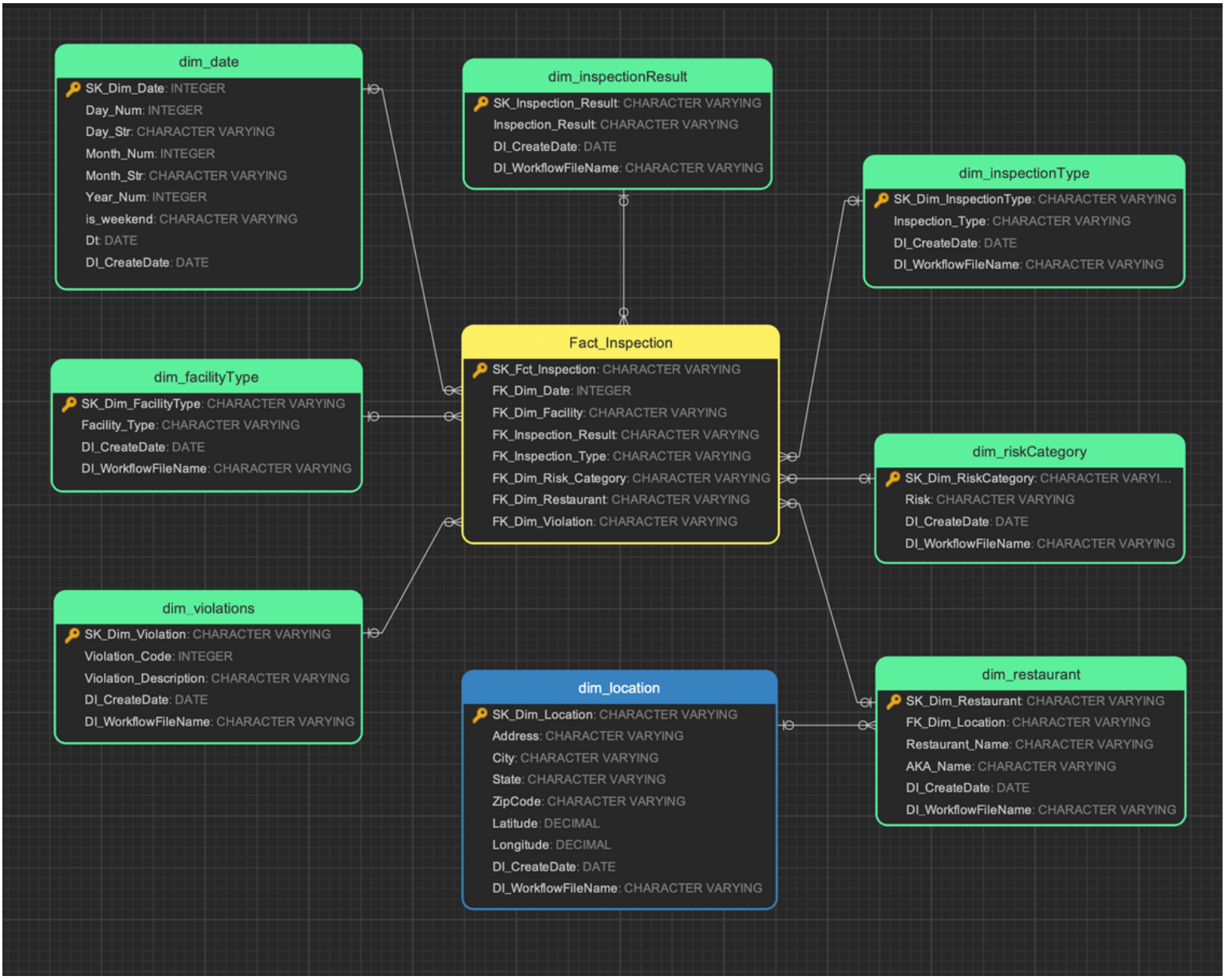
Query executed successfully. | localhost (15.0 RTM) | apoorv_damg7370 (59) | dadabi | 00:00:00 | 1,000 rows

Data Preparation (Part 1)

- **Handling Null Values:**
  - **For categorical variables, null entries were replaced with the term 'missing' to maintain data integrity.**
  - **For numerical variables, null values were filled with -999, allowing for the distinction between genuine data points and imputed values during analyses.**
- **Data Transformation and Stage Table Creation:**
  - **Developed a cleaned stage table that underwent various transformations, aligning with analytical objectives and data standards.**
- **Schema Management with XML Generator:**
  - **Crafted an XML generator to dynamically generate the required XML schema based on the table structure, reducing manual coding efforts and enhancing accuracy.**
- **Date and Workflow Metadata:**
  - **For the DI_CREATE_DATE field, Talend's date expression was used to accurately generate data integration timestamps.**
  - **The DI_Workflow_FileName field was populated through a context variable in Talend, capturing workflow metadata consistently.**
  -

This structured approach to data preprocessing not only facilitated the seamless integration and analysis of the Dallas dataset but also underscored the importance of meticulous data preparation in achieving reliable and insightful analytical outcomes.

Dimensional Modeling (Part 2)



Dimensional Modeling (Part 2)

In the report presented herein, our group has developed a star schema data warehouse model aimed at enhancing the analytical capabilities for monitoring food inspection processes. The centerpiece of this schema is the Fact_Inspection table, which serves as a repository for the key metrics and attributes associated with each inspection event.

- **Fact_Inspection Table: This pivotal table is equipped with unique identifiers for each recorded inspection and includes references to various dimension tables that provide context and granularity to the inspection data. These references are manifested through foreign keys that link to respective dimensions such as date, facility type, and inspection results.**
- **Dimension Tables Overview:**
  - **The dim_date table is a temporal dimension that allows us to dissect the data chronologically, providing insights into the distribution of inspections over time.**
  - **Through the dim_facilityType table, we can categorize the data by the nature of the establishments inspected.**
  - **The outcomes of the inspections are elucidated by the dim_inspectionResult table.**
  - **Our schema differentiates the inspection methods via the dim_inspectionType table.**
  - **The dim_riskCategory table aids in assessing the severity and urgency of the findings from each inspection.**
  - **A catalogue of infractions is housed within the dim_violations table, each characterized by a unique code and description.**
  - **The dim_location table enriches our dataset with geographical precision, enabling location-based analytics.**
  - **Lastly, the dim_restaurant table captures details specific to each restaurant subject to inspection, providing a clear link between inspections and individual establishments.**

Dimensional Modeling (Part 2)

DDL Script

```
CREATE TABLE [dim_date] (
  [SK_Dim_Date] int NOT NULL,
  [Day_Num] int NOT NULL,
  [Day_Str] varchar(20) NOT NULL,
  [Month_Num] int NOT NULL,
  [Month_Str] varchar(50) NOT NULL,
  [Year_Num] int NOT NULL,
  [is_weekend] varchar(50) NOT NULL,
  [Dt] date NOT NULL,
  [DI_CreateDate] date NOT NULL,
  CONSTRAINT [_copy_3] PRIMARY KEY CLUSTERED ([SK_Dim_Date])
WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON)
)
GO

CREATE TABLE [dim_facilityType] (
  [SK_Dim_FacilityType] varchar(100) NOT NULL,
  [Facility_Type] varchar(100) NOT NULL,
  [DI_CreateDate] date NOT NULL,
  [DI_WorkflowFileName] varchar(100) NOT NULL,
  CONSTRAINT [_copy_4] PRIMARY KEY CLUSTERED ([SK_Dim_FacilityType])
WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON)
)
GO

CREATE TABLE [dim_inspectionResult] (
  [SK_Inspection_Result] varchar(10) NOT NULL,
  [Inspection_Result] varchar(80) NOT NULL,
  [DI_CreateDate] date NOT NULL,
  [DI_WorkflowFileName] varchar(200) NOT NULL,
  CONSTRAINT [_copy_7] PRIMARY KEY CLUSTERED ([SK_Inspection_Result])
WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON)
)
GO
```

Dimensional Modeling (Part 2)

DDL Script

```
CREATE TABLE [dim_inspectionType] (
  [SK_Dim_InspectionType] varchar(10) NOT NULL,
  [Inspection_Type] varchar(50) NOT NULL,
  [DI_CreateDate] date NOT NULL,
  [DI_WorkflowFileName] varchar(200) NOT NULL,
  CONSTRAINT [_copy_6] PRIMARY KEY CLUSTERED ([SK_Dim_InspectionType])
WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON)
)
GO

CREATE TABLE [dim_location] (
  [SK_Dim_Location] varchar(100) NOT NULL,
  [Address] varchar(100) NULL,
  [City] varchar(100) NULL,
  [State] varchar(100) NULL,
  [ZipCode] varchar(100) NULL,
  [Latitude] decimal(15,10) NULL,
  [Longitude] decimal(15,10) NULL,
  [DI_CreateDate] date NULL,
  [DI_WorkflowFileName] varchar(100) NULL,
  CONSTRAINT [_copy_8] PRIMARY KEY CLUSTERED ([SK_Dim_Location])
WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON)
)
GO

CREATE TABLE [dim_restaurant] (
  [SK_Dim_Restaurant] varchar(20) NOT NULL,
  [FK_Dim_Location] varchar(100) NOT NULL,
  [Restaurant_Name] varchar(100) NOT NULL,
  [AKA_Name] varchar(100) NOT NULL,
  [DI_CreateDate] date NOT NULL,
  [DI_WorkflowFileName] varchar(100) NOT NULL,
  CONSTRAINT [_copy_1] PRIMARY KEY CLUSTERED ([SK_Dim_Restaurant])
WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON)
)
GO
```

Dimensional Modeling (Part 2)

DDL Script

```
CREATE TABLE [dim_riskCategory] (
  [SK_Dim_RiskCategory] varchar(100) NOT NULL,
  [Risk] varchar(100) NOT NULL,
  [DI_CreateDate] date NOT NULL,
  [DI_WorkflowFileName] varchar(100) NOT NULL,
  CONSTRAINT [_copy_5] PRIMARY KEY CLUSTERED ([SK_Dim_RiskCategory])
WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON)
)
GO

CREATE TABLE [dim_violations] (
  [SK_Dim_Violation] varchar(100) NOT NULL,
  [Violation_Code] int NOT NULL,
  [Violation_Description] varchar(3000) NOT NULL,
  [DI_CreateDate] date NOT NULL,
  [DI_WorkflowFileName] varchar(100) NOT NULL,
  CONSTRAINT [_copy_9] PRIMARY KEY CLUSTERED ([SK_Dim_Violation])
WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON)
)
GO

CREATE TABLE [Fact_Inspection] (
  [SK_Fct_Inspection] varchar(100) NOT NULL,
  [FK_Dim_Date] int NOT NULL,
  [FK_Dim_Facility] varchar(100) NOT NULL,
  [FK_Inspection_Result] varchar(100) NOT NULL,
  [FK_Inspection_Type] varchar(100) NOT NULL,
  [FK_Dim_Risk_Category] varchar(100) NOT NULL,
  [FK_Dim_Restaurant] varchar(100) NOT NULL,
  [FK_Dim_Violation] varchar(100) NOT NULL,
  CONSTRAINT [_copy_2] PRIMARY KEY CLUSTERED ([SK_Fct_Inspection])
WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON)
)
GO
```

Dimensional Modeling (Part 2)

DDL Script

```
ALTER TABLE [dim_restaurant] ADD CONSTRAINT [fk_dim_restaurant_dim_location] FOREIGN KEY ([FK_Dim_Location]) REFERENCES [dim_location] ([SK_Dim_Location])
GO
ALTER TABLE [Fact_Inspection] ADD CONSTRAINT [fk_Fact_Inspection_dim_date] FOREIGN KEY ([FK_Dim_Date]) REFERENCES [dim_date] ([SK_Dim_Date])
GO
ALTER TABLE [Fact_Inspection] ADD CONSTRAINT [fk_Fact_Inspection_dim_facilityType] FOREIGN KEY ([FK_Dim_Facility]) REFERENCES [dim_facilityType] ([SK_Dim_FacilityType])
GO
ALTER TABLE [Fact_Inspection] ADD CONSTRAINT [fk_Fact_Inspection_dim_inspectionResult] FOREIGN KEY ([FK_Inspection_Result]) REFERENCES [dim_inspectionResult] ([SK_Inspection_Result])
GO
ALTER TABLE [Fact_Inspection] ADD CONSTRAINT [fk_Fact_Inspection_dim_inspectionType] FOREIGN KEY ([FK_Inspection_Type]) REFERENCES [dim_inspectionType] ([SK_Dim_InspectionType])
GO
ALTER TABLE [Fact_Inspection] ADD CONSTRAINT [fk_Fact_Inspection_dim_riskCategory] FOREIGN KEY ([FK_Dim_Risk_Category]) REFERENCES [dim_riskCategory] ([SK_Dim_RiskCategory])
GO
ALTER TABLE [Fact_Inspection] ADD CONSTRAINT [fk_Fact_Inspection_dim_restaurant] FOREIGN KEY ([FK_Dim_Restaurant]) REFERENCES [dim_restaurant] ([SK_Dim_Restaurant])
GO
ALTER TABLE [Fact_Inspection] ADD CONSTRAINT [fk_Fact_Inspection_dim_violations] FOREIGN KEY ([FK_Dim_Violation]) REFERENCES [dim_violations] ([SK_Dim_Violation])
GO
```

Data Loading (Part 3)

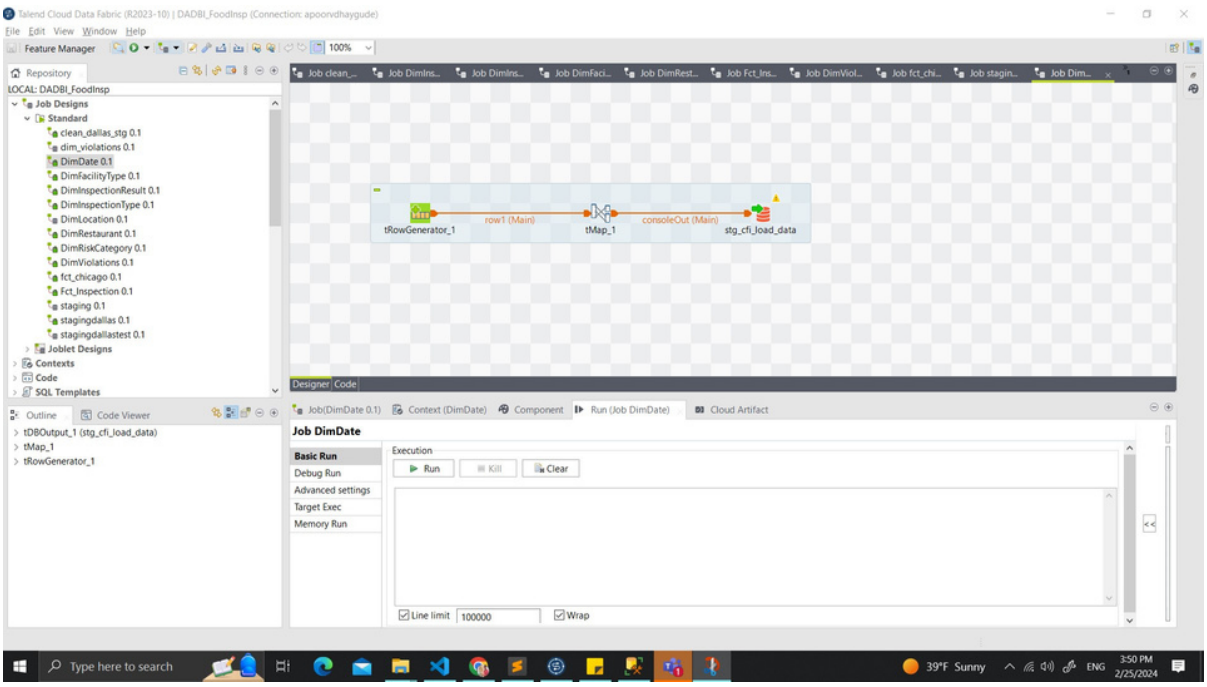
In the course of developing our data warehouse, the initial step involved the establishment of the dim_location table. This was a strategic move, as it forms the basis for the dim_restaurant table, with location data being a precursor to restaurant information.

- Dimension Population Process:
  - Location Dimension: Our first action was to populate the dim_location table, which included critical geographic identifiers such as Address, City, State, ZipCode, Latitude, and Longitude.
  - Restaurant Dimension: Following the completion of the dim_location table, we proceeded to populate the dim_restaurant table. This included a meticulous join operation, where the Address column from the dim_restaurant was linked to the dim_location through the appropriate foreign key relationship, ensuring referential integrity.

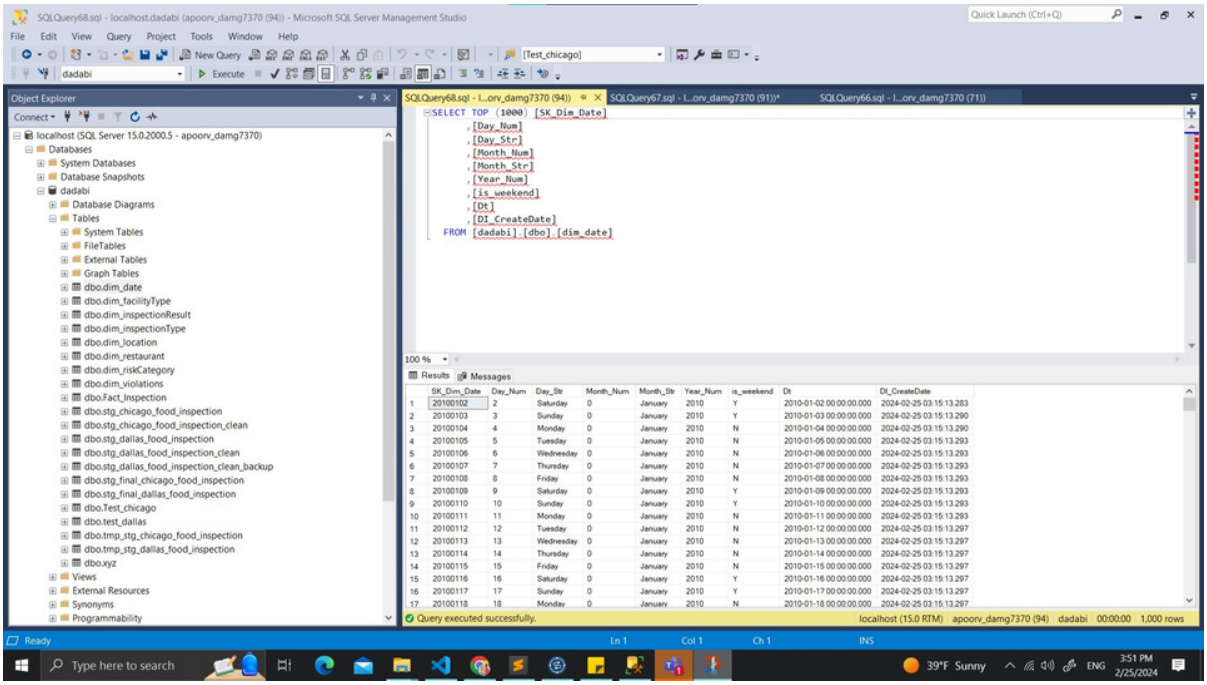
Upon successfully establishing these foundational dimensions, our attention then turned to the additional dimension tables.

Dimensions and Facts Workflow and DDL Script

Dim Date Table - Workflow

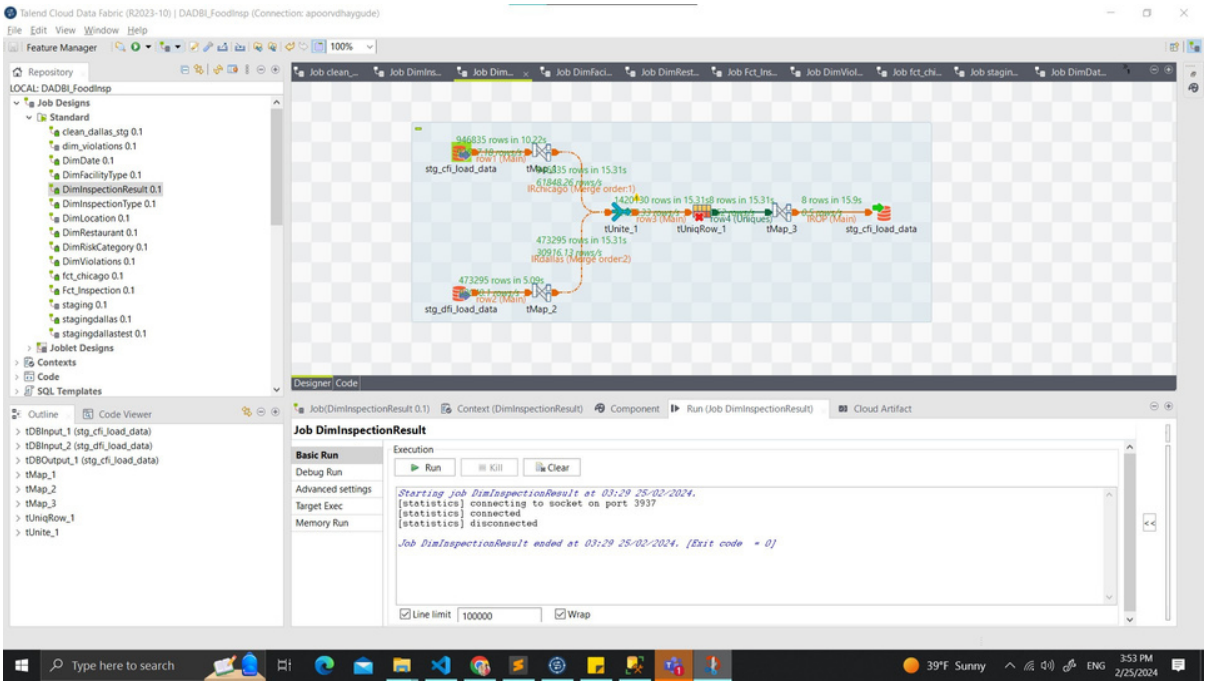


Dim Date Table- DDL Script

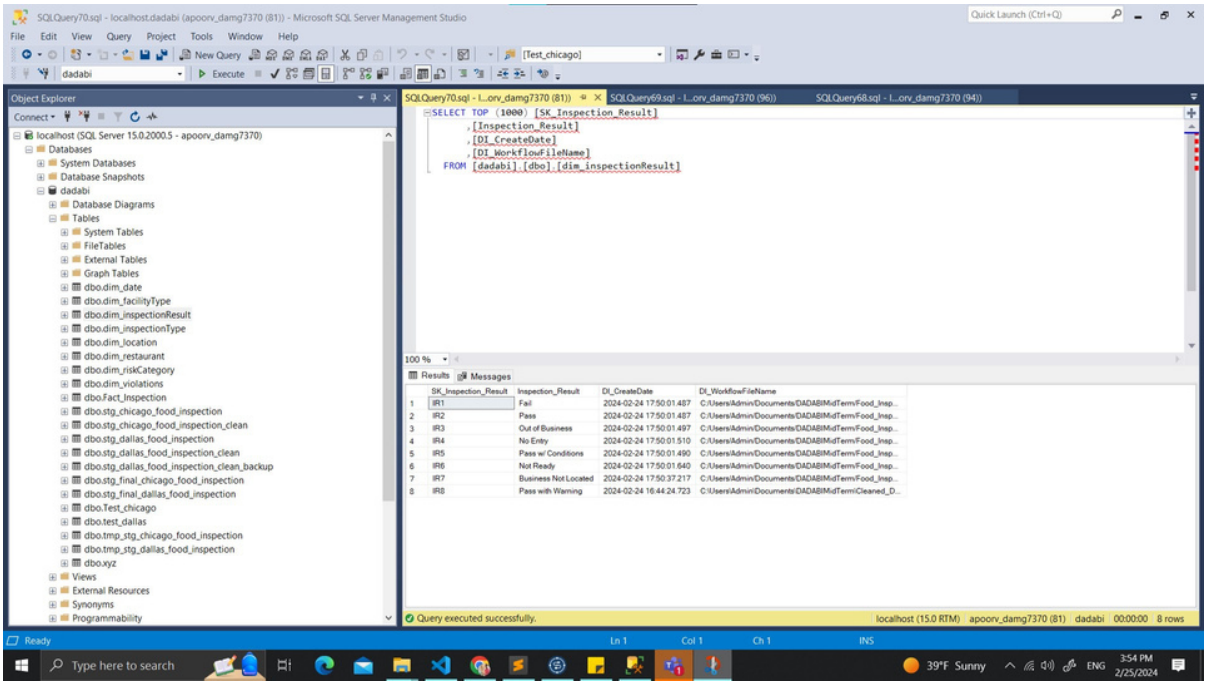


Dimensions and Facts - Workflow

Dim Inspection Result Table - Workflow

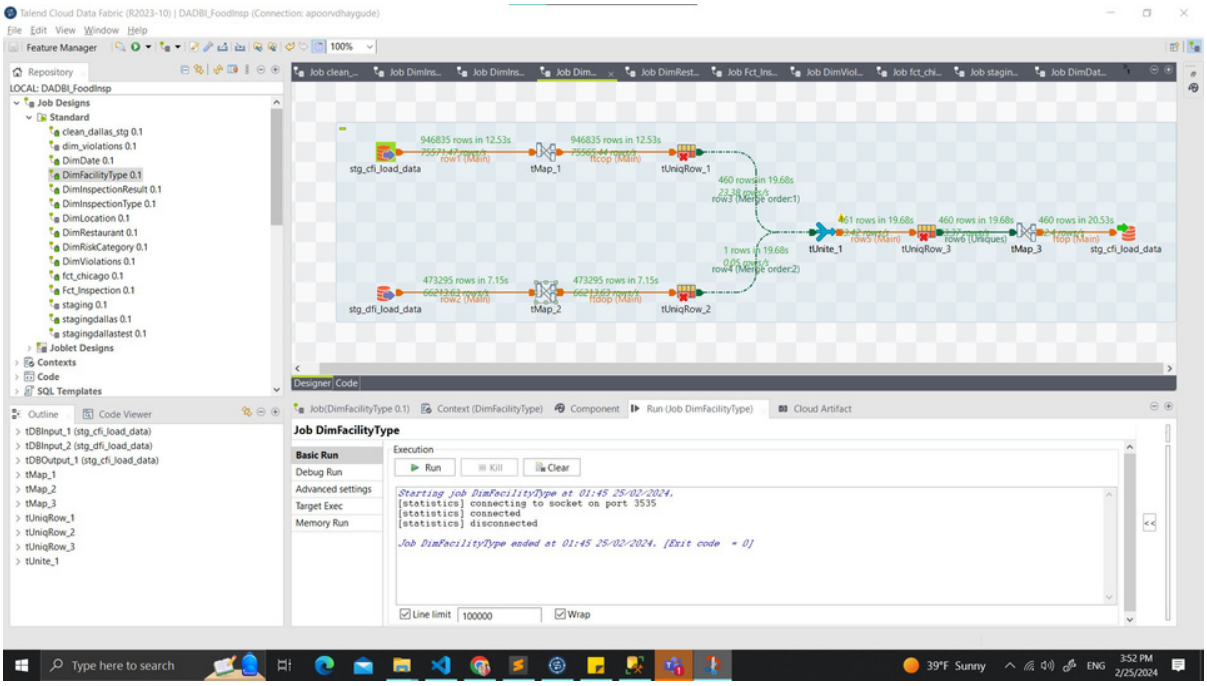


Dim Inspection Result Table - DDL Script

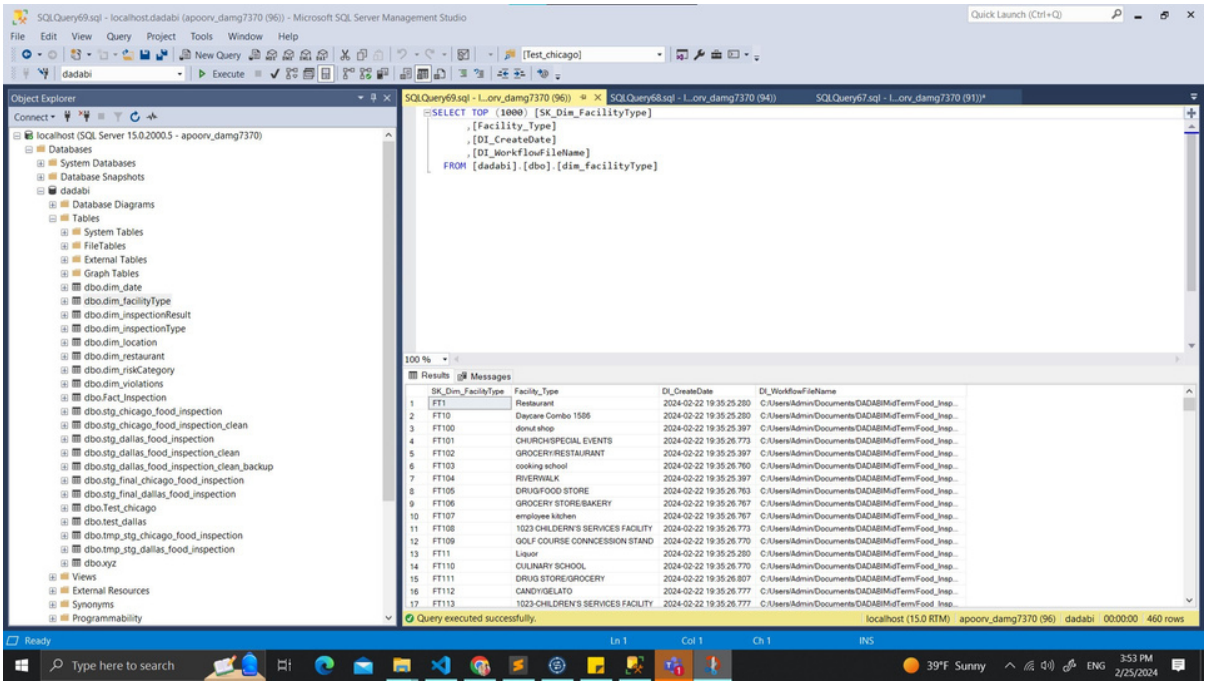


Dimensions and Facts - Workflow

Dim Facility Table - Workflow

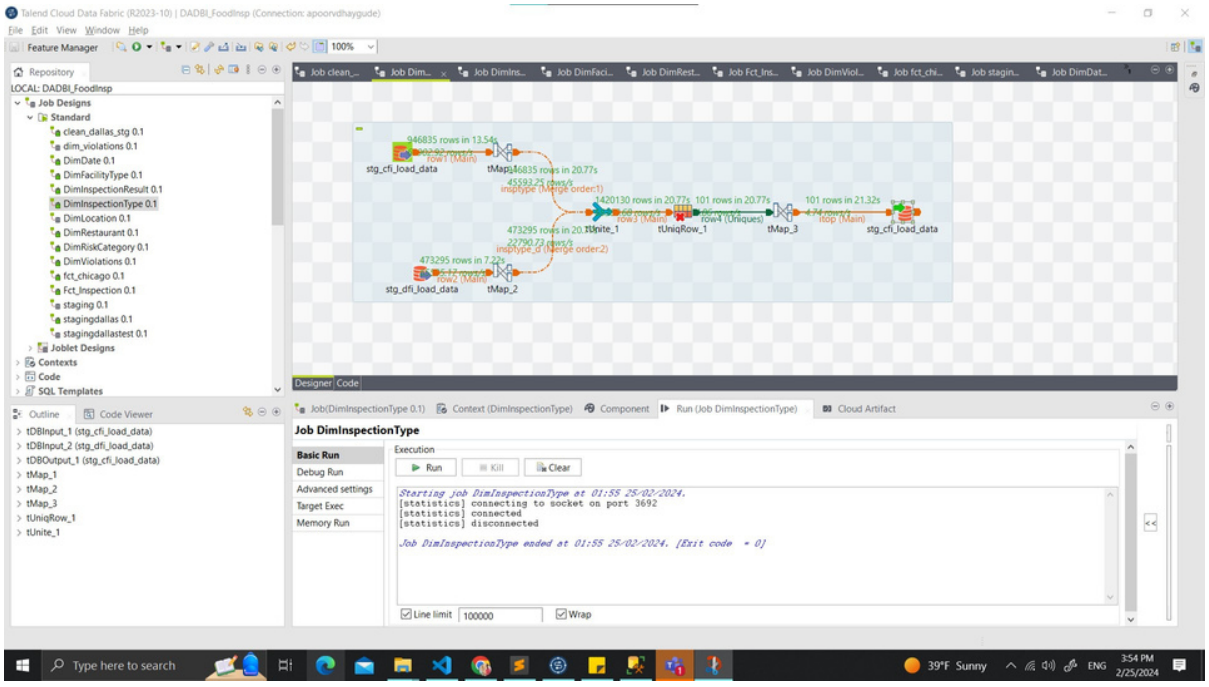


Dim Facility Table - DDL Script

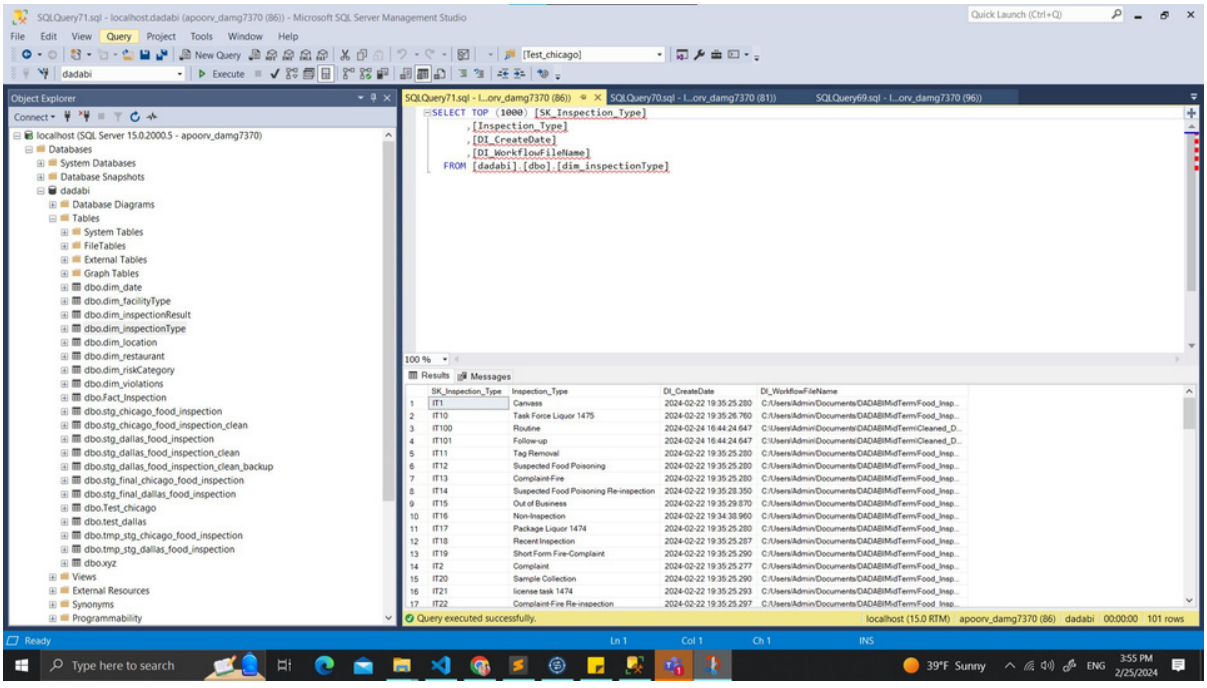


Dimensions and Facts - Workflow

Dim Inspection Type Table - Workflow

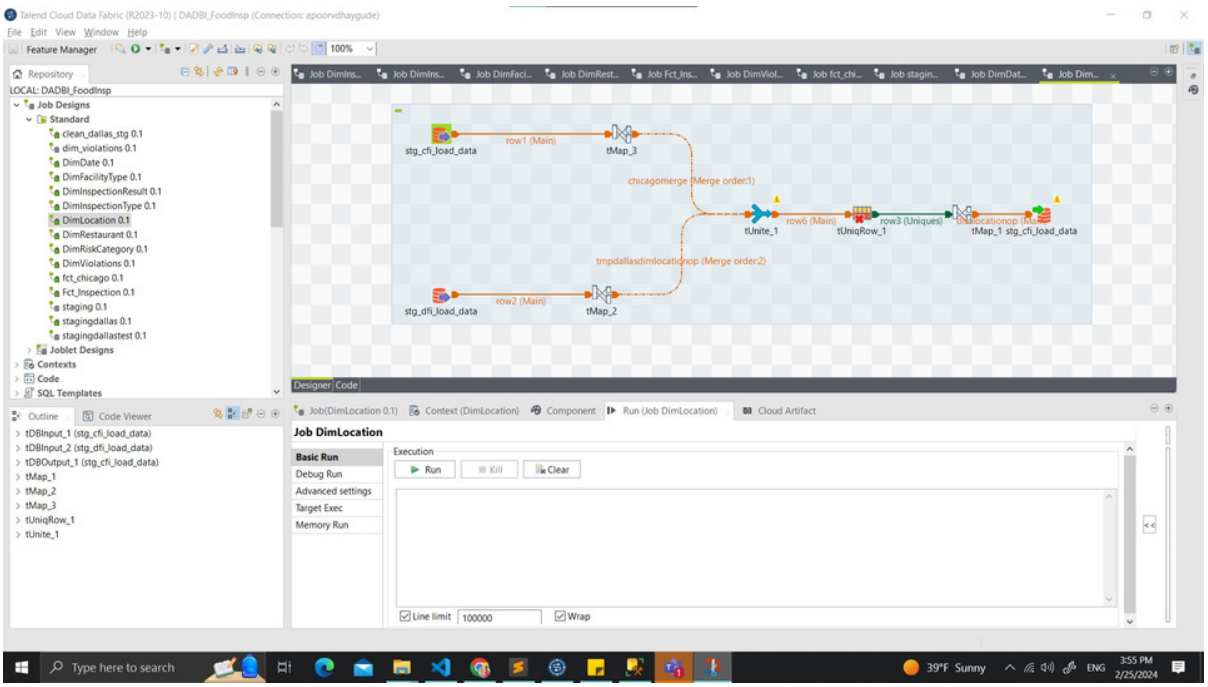


Dim Inspection Type Table - DDL Script

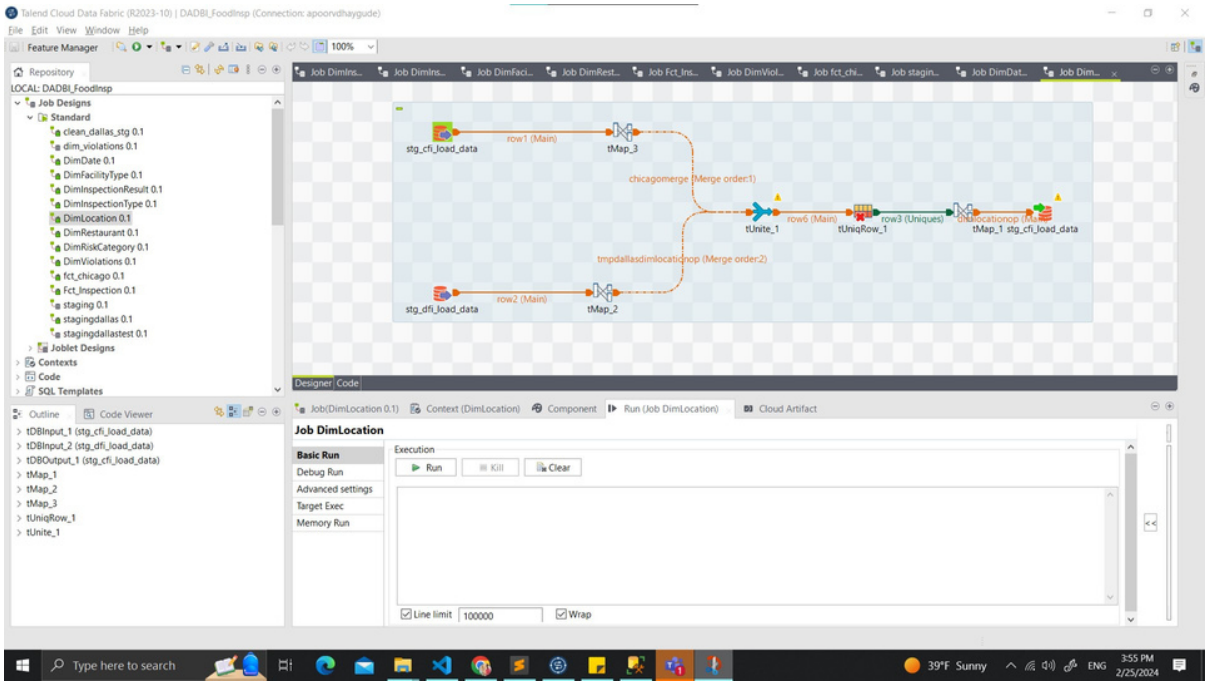


Dimensions and Facts - Workflow

Dim Location Table - Workflow

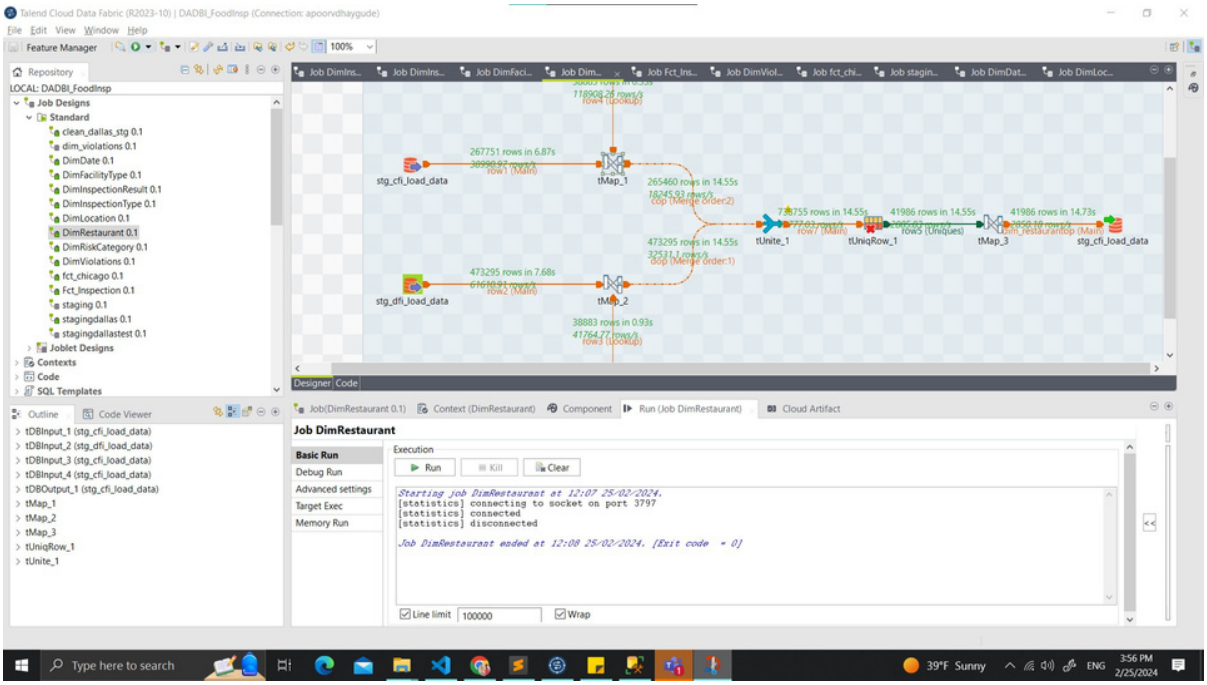


Dim Location Table - DDL Script

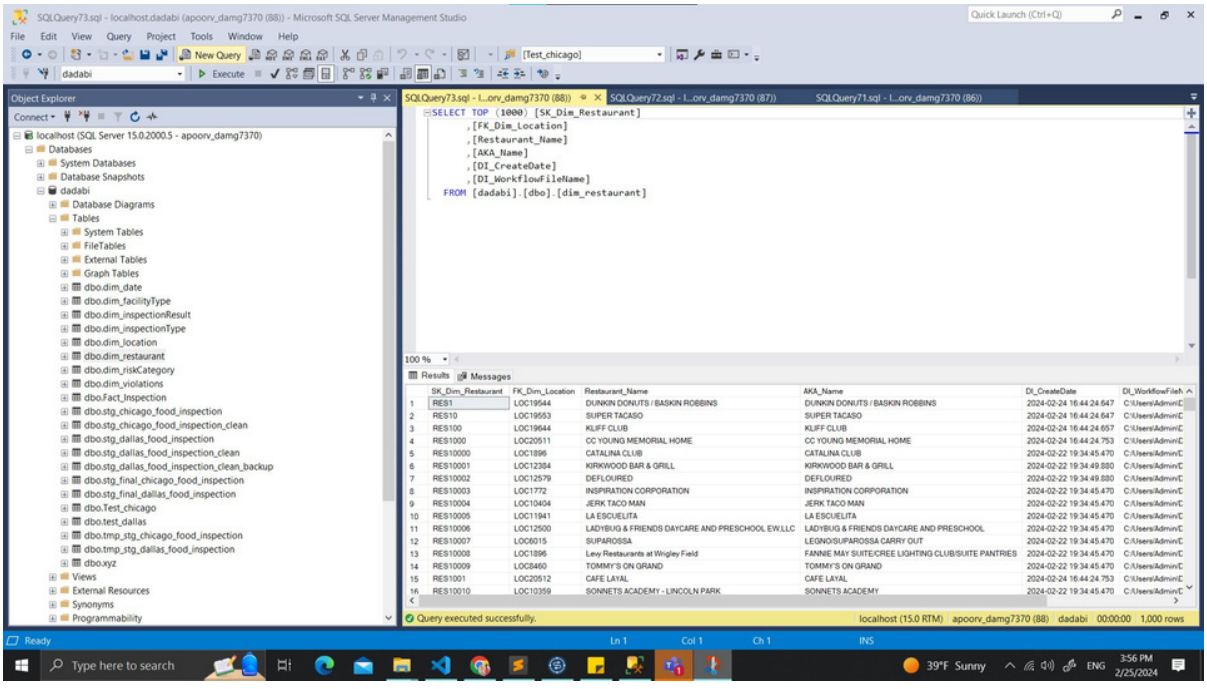


Dimensions and Facts - Workflow

Dim Restaurant Table - Workflow

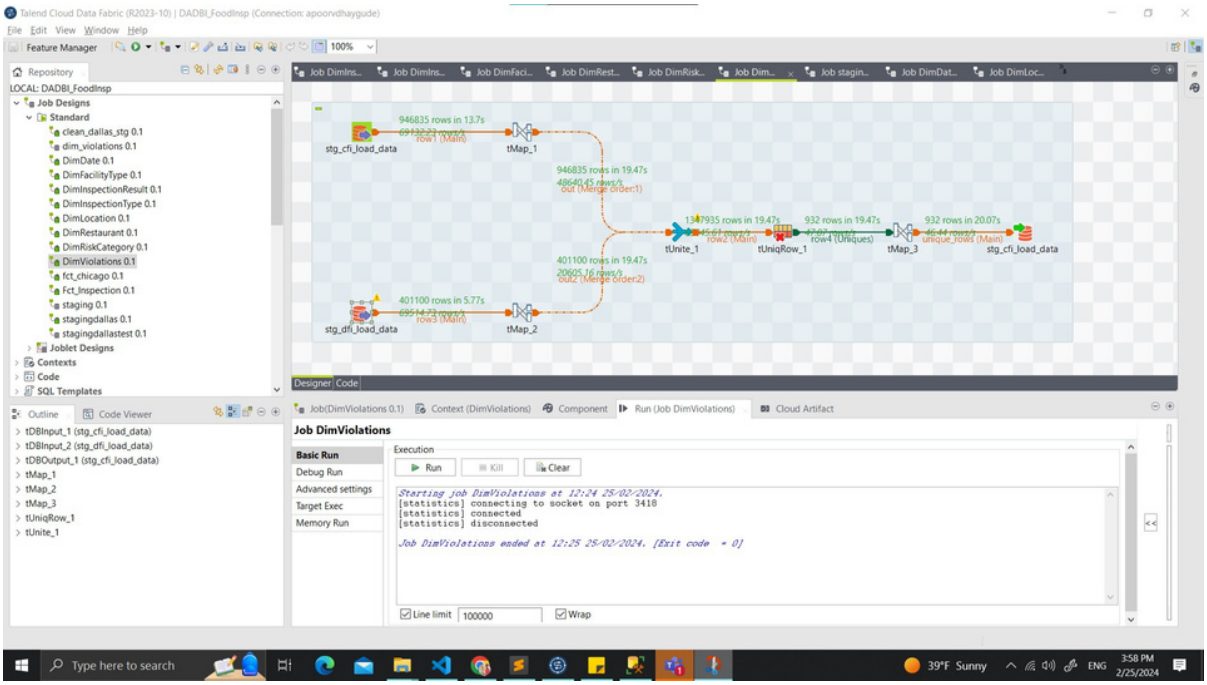


Dim Restaurant Table - DDL Script

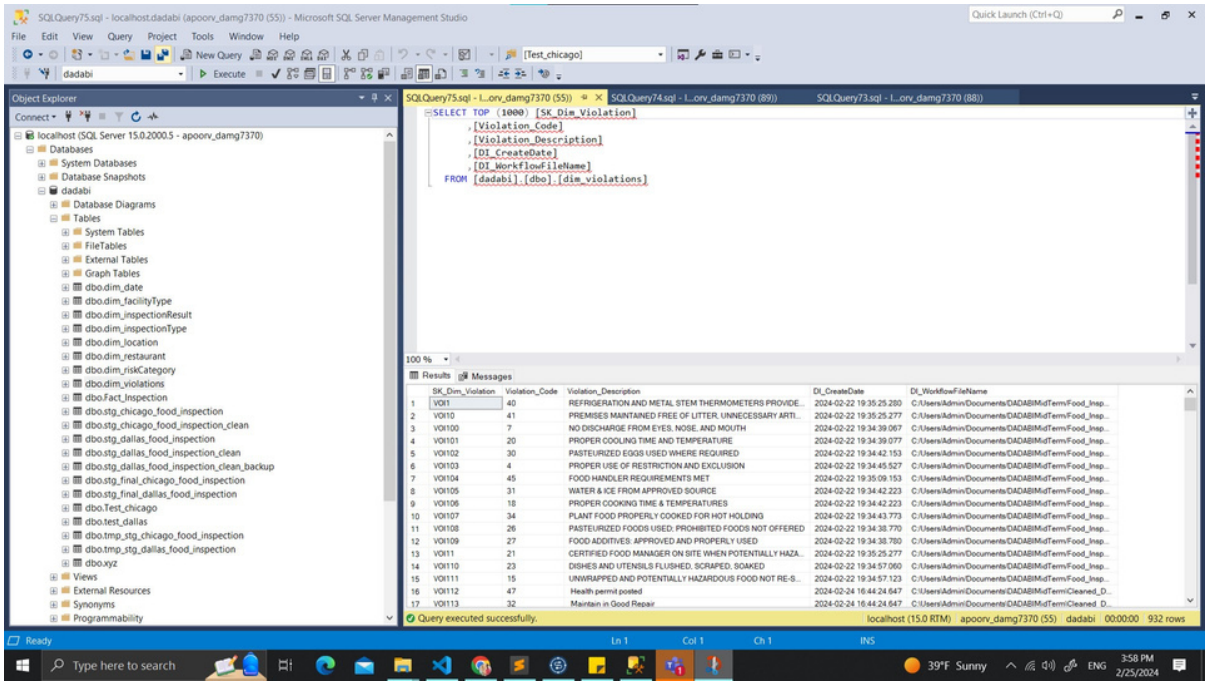


Dimensions and Facts - DDL Scripts

Dim Violations Table - Workflow

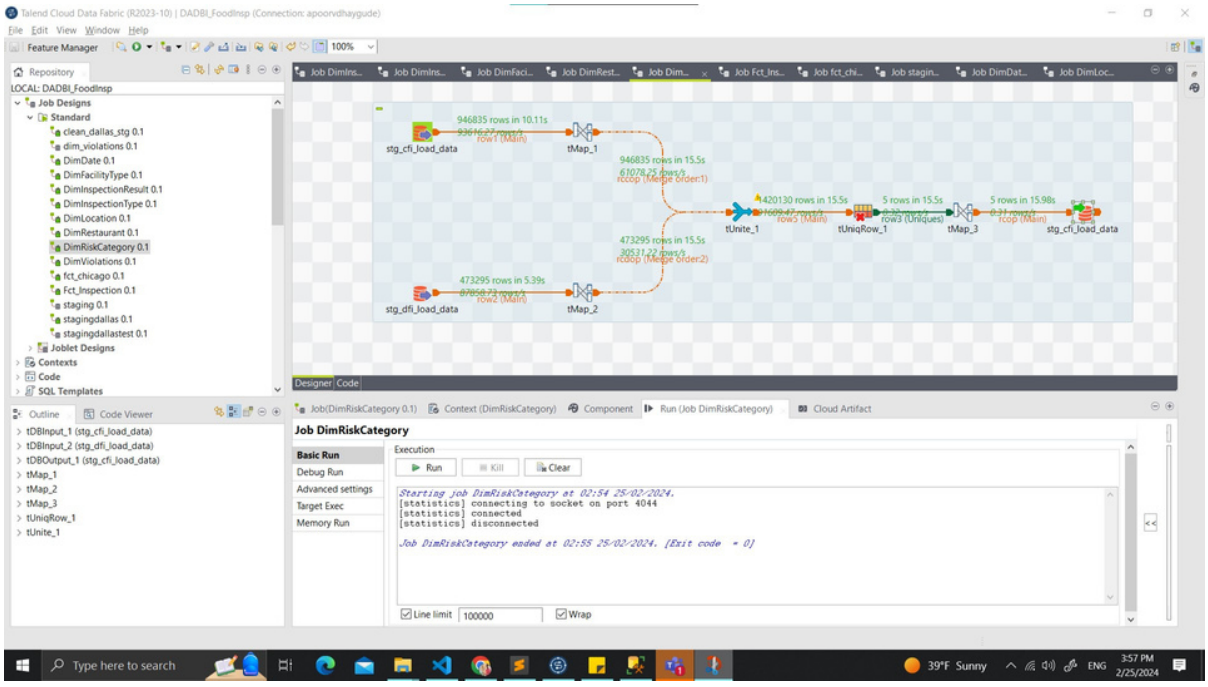


Dim Violations Table - DDL Script

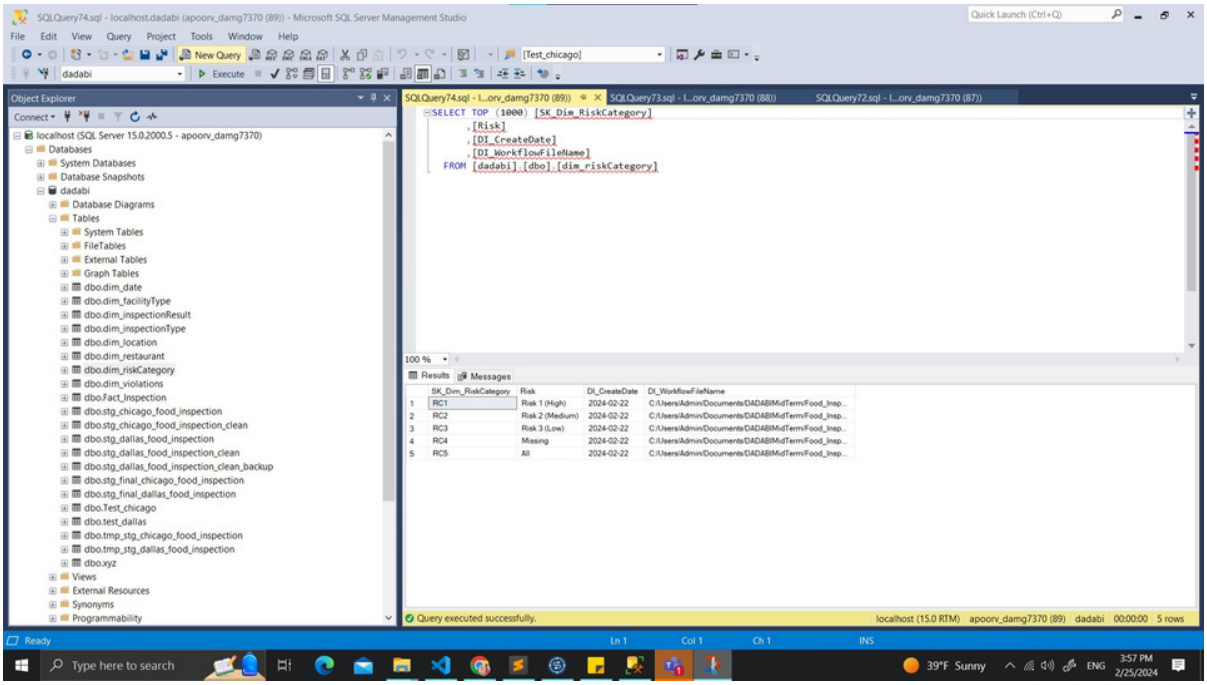


Dimensions and Facts - DDL Scripts

Dim Risk Category Table - Workflow

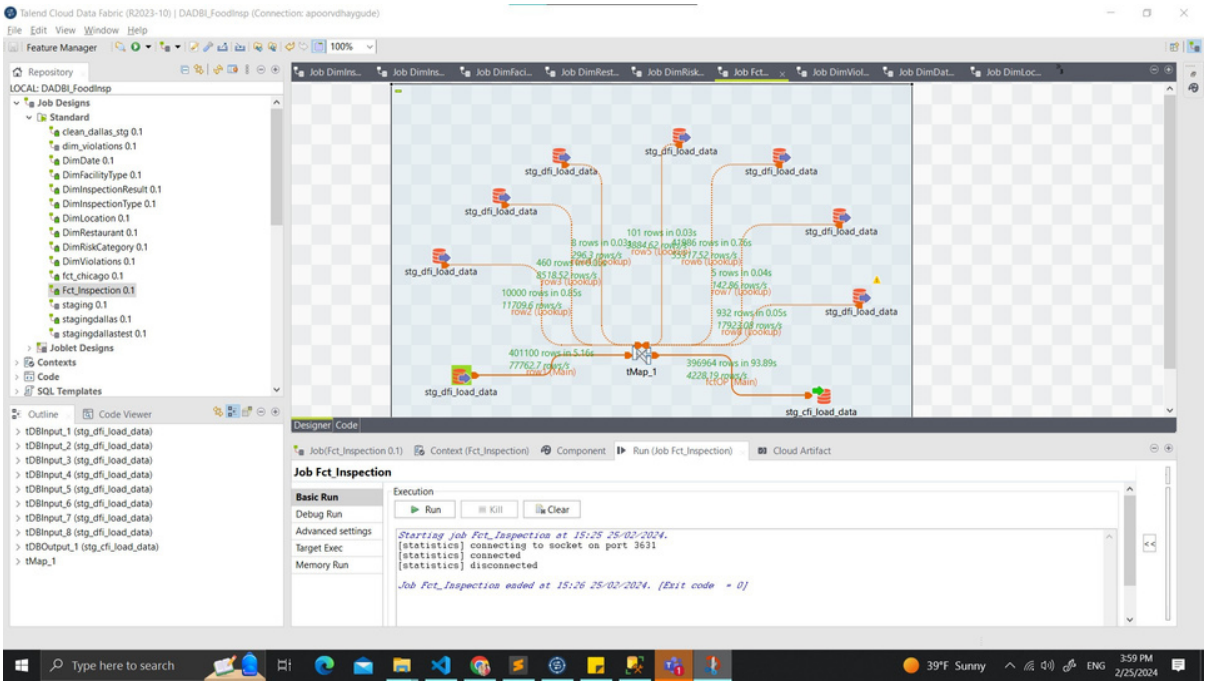


Dim Risk Category Table - DDL Script

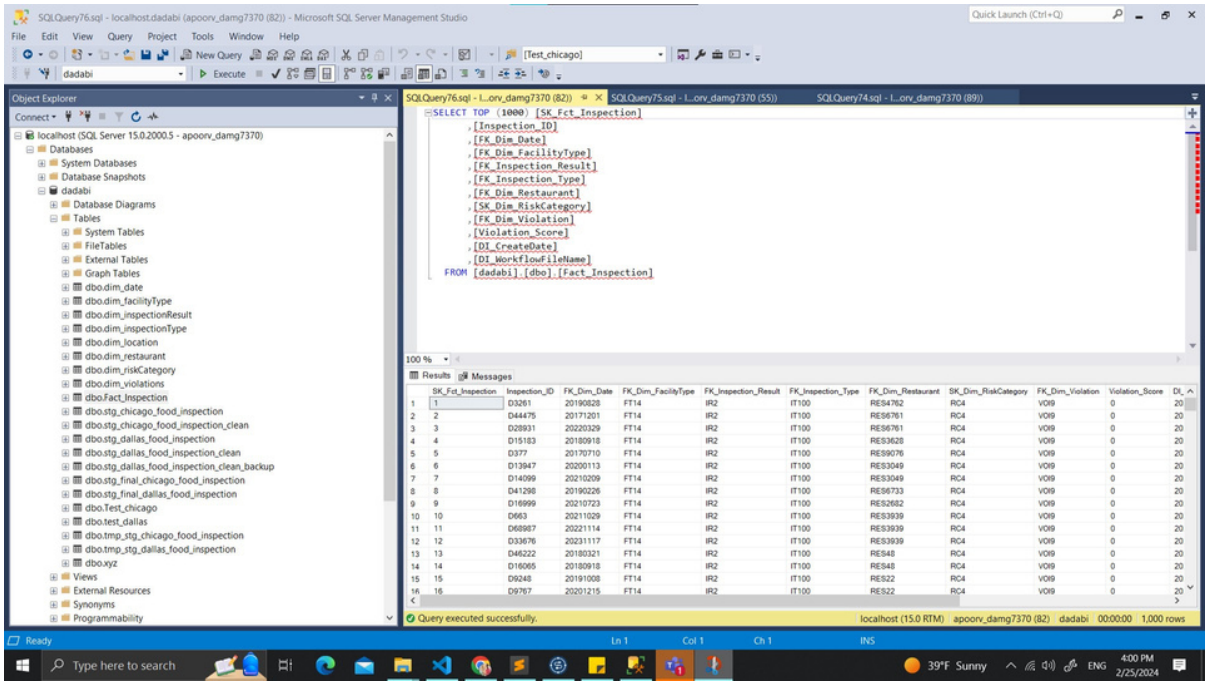


Dimensions and Facts - DDL Scripts

Fact Inspection Table - Workflow

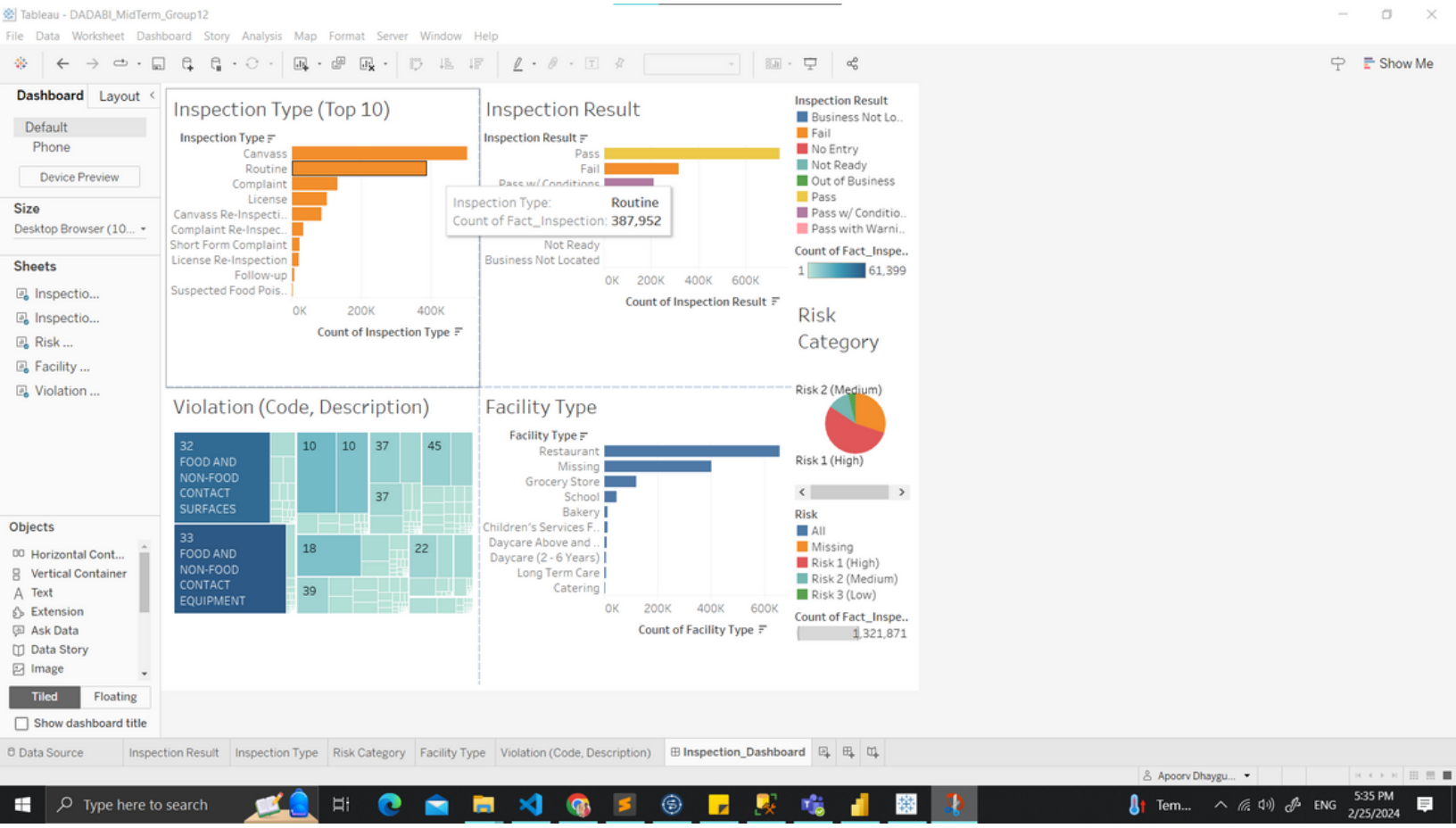


Fact Inspection Table - DDL Script



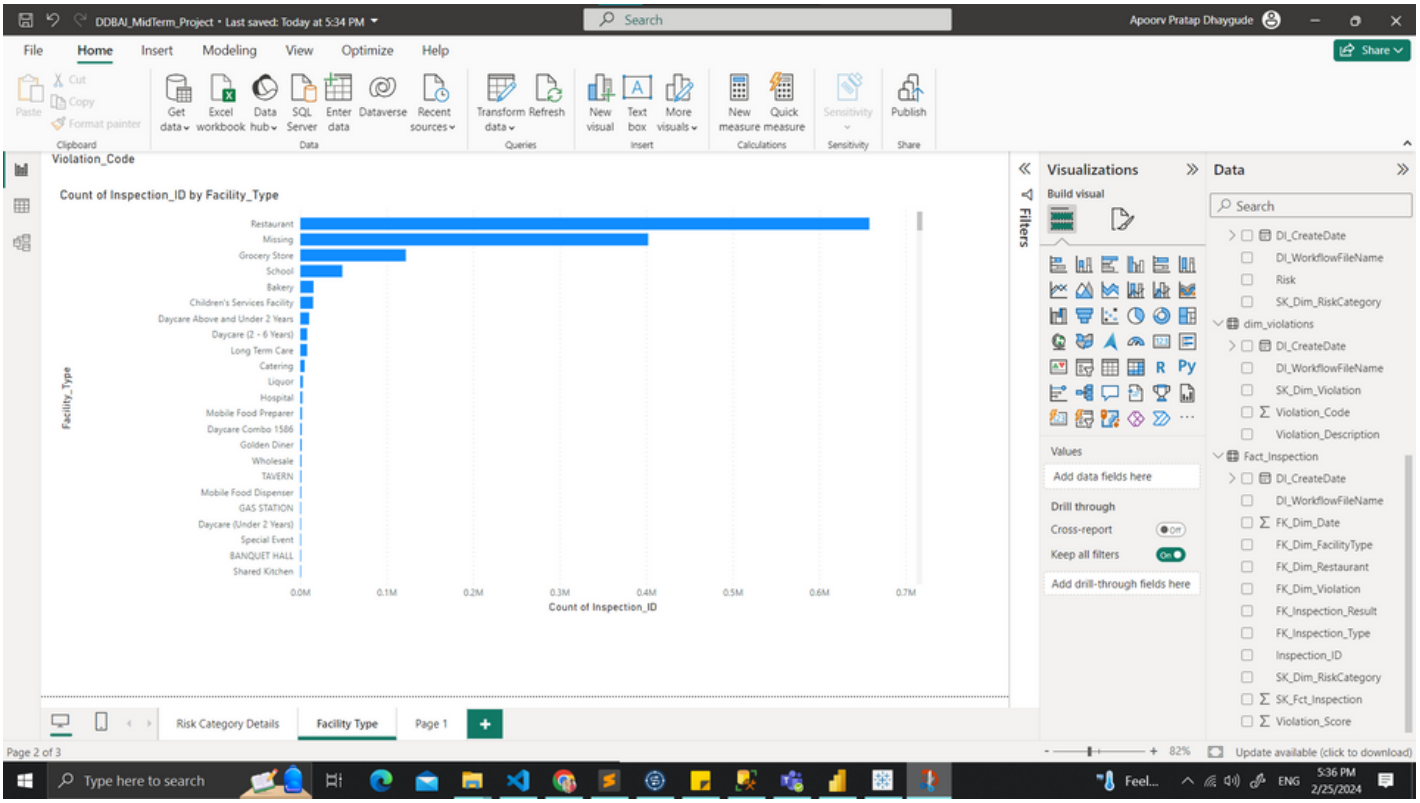
Visualizations

Tableau



Visualizations

PowerBI



Queries