

MidTerm - Group 19

Program:MS Data Architecture and Management

Course: DAMG7370 - Designing Advanced Data Architectures for Business Intelligence

AuthorDetails:

Name:ShubhamAgarwal

NUID:002866699

EmailAddress:agarwal.shubh@northeastern.edu

Name: Pratik Karunakar Poojari

NUID:002645173

EmailAddress:poojari.pr@northeastern.edu

Name: Vaidehi Tawde

NUID:002643847

EmailAddress:v@northeastern.edu

Name:Anuja Shinde

NUID:002643847

EmailAddress:v@northeastern.edu

PROFILING IN ALTERYX

Dallas

Alteryx Designer x64 - Dallas_Data_Profiling.ymd

File Edit View Options Help

Favorites Recommended In/Out Preparation Join Parse Transform In-Database Reporting Documentation Spatial Machine Learning Text Mining Computer Vision Interface Data Investigation

Browse (10) - Configuration

Dallas_Data_Profiling.ymd Chicago_DataProfiling1.ymd + ... Run

Profile 78,400 records displayed, 114 fields, 44 MB

Restaurant Name

MCDONALDS	531
SUBWAY	310
BURGER KING	290
SONIC DRIVE IN	264
WINGSTOP	261
995 more >	

Inspection Type

Routine	77,436
Follow-up	934
Complaint	30

Inspection Date

03/21/2018	117
09/21/2017	110
10/12/2017	109
11/28/2018	103
05/31/2018	102
995 more >	

Results - Browse (10) - Input

114 of 114 Fields Cell Viewer 78,400 records displayed, 44 MB

Record	Restaurant Name	Inspection Type	Inspection Date	Inspection Score	Street Number	Street Name	Street Direction
1	MICKLE CHICKEN	Routine	10/30/2019	100	3203	CAMP WISDOM	W
2	TOM THUMB - JUICE BAR	Routine	04/08/2020	100	2380	FIELD	N
3	BROOKDALE WHITE ROCK	Routine	07/29/2020	97	9271	WHITE ROCK	[null]
4	CHURCH'S CHICKEN #201	Routine	09/14/2020	100	10295	FERGUSON	[null]
5	PEAK PREPARATORY-PRIMARY SCHOOL	Routine	04/25/2017	98	1474	ANNEX	[null]
6	SAMS CLUB #6482 - DEMO ROOM	Routine	04/05/2021	100	5555	BUCKNER	S
7	AMC THEATRES NORTHPARK 15 (BAR)	Routine	04/09/2021	100	8687	CENTRAL	N

```

graph LR
    A[Restaurant_and_Food_Establishments_Inspections_October_2016_to_Present_20240223(t).tsv] --> B(( ))
    B --> C(( ))
    C --> D[dcmsHUBH Table=stg_Dallas_Alteryx]
    
```

Alteryx Designer x64 - Dallas_Data_Profiling.ymd - Browse (10)

Table Profile

78,400 records displayed, 114 fields, 44 MB

Restaurant Name

MCDONALDS	531
SUBWAY	310
BURGER KING	290
SONIC DRIVE IN	264
WINGSTOP	261

Inspection Type

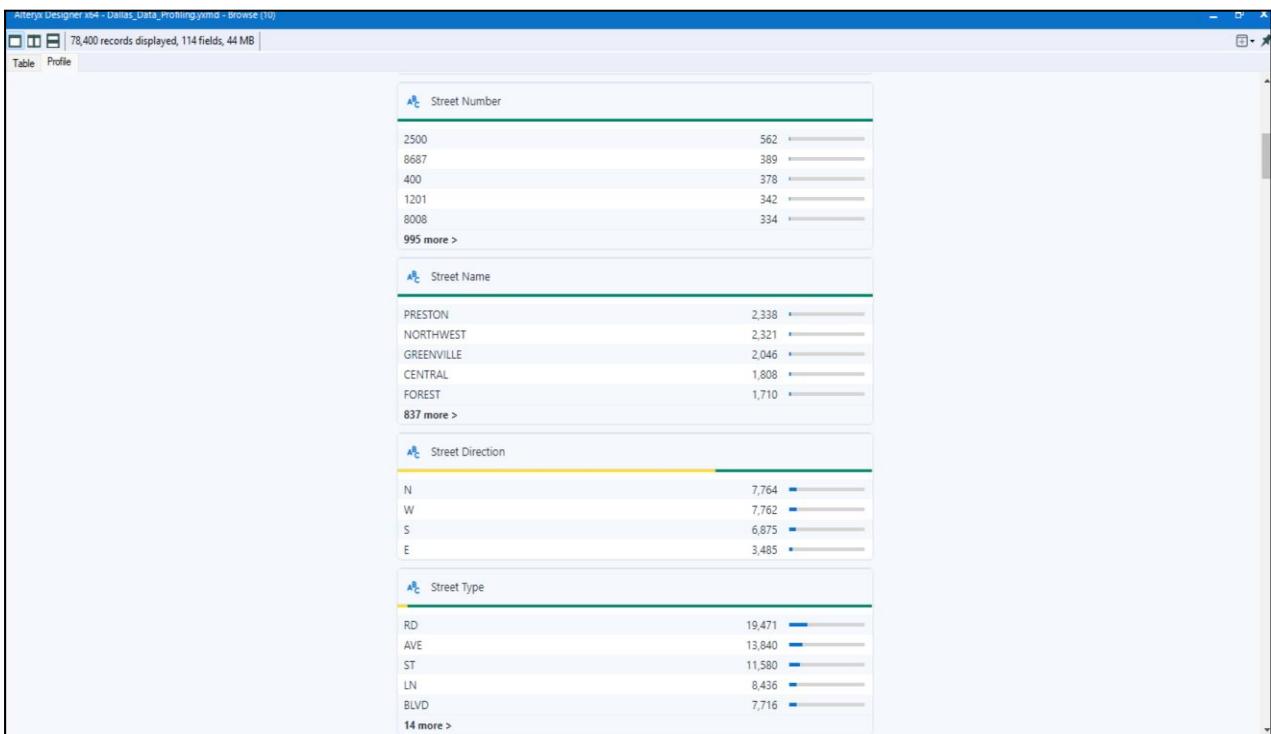
Routine	77,436
Follow-up	934
Complaint	30

Inspection Date

03/21/2018	117
09/21/2017	110
10/12/2017	109
11/28/2018	103
05/31/2018	102

Inspection Score

100	6.525
97	5.219
95	4.926
94	4.696
96	4.645



Chicago

Alteryx Designer x64 - *Chicago_DataProfiling1.ymd

File Edit View Options Help

Browse (37) - Configuration

Profile

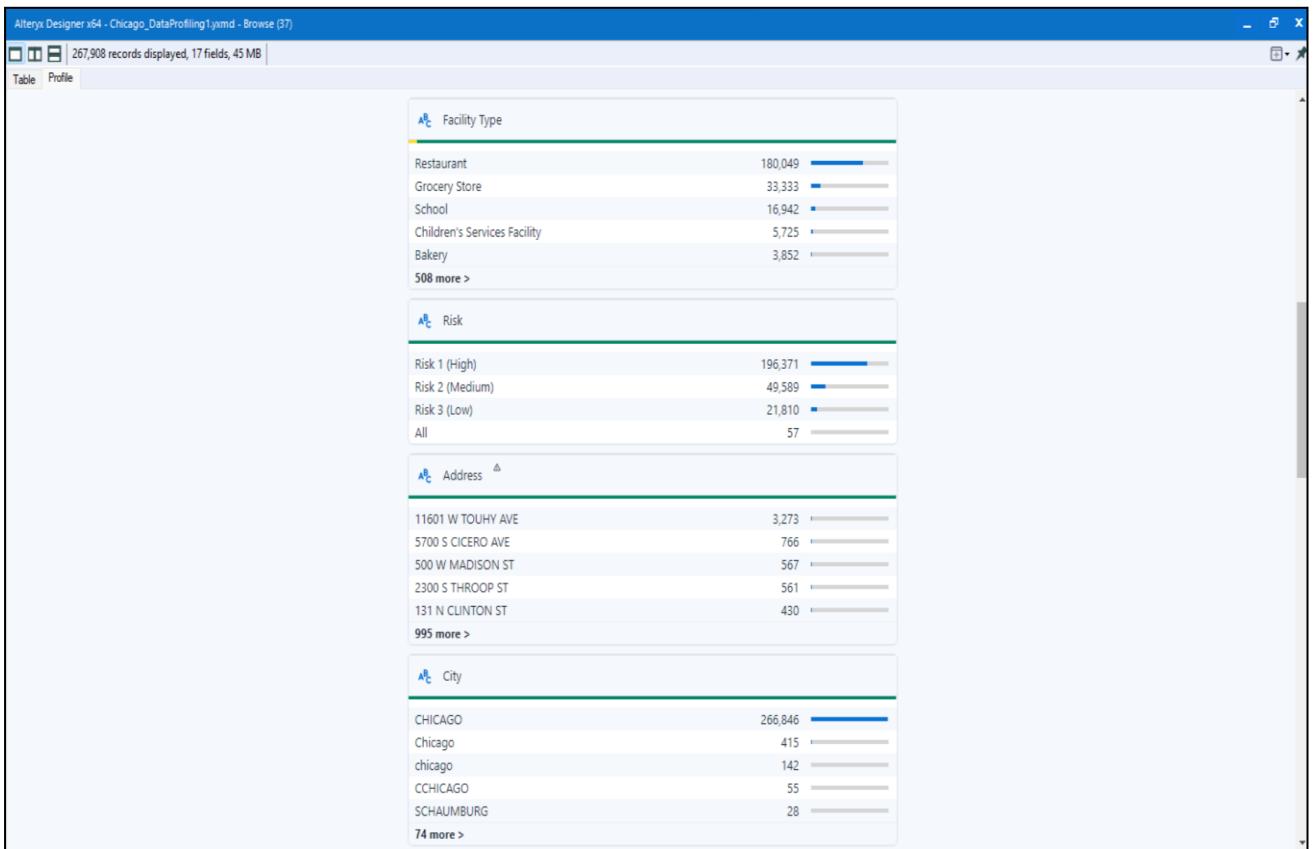
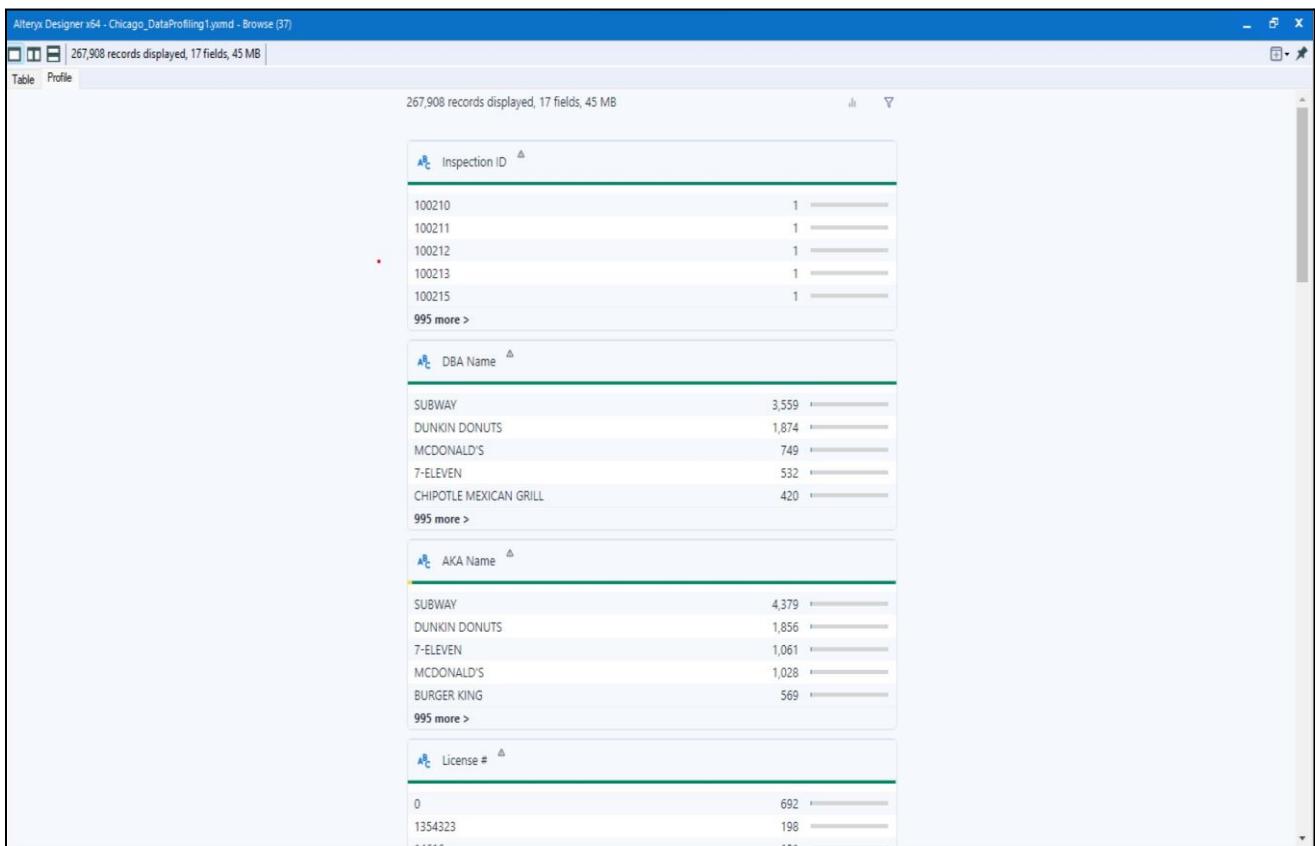
267,908 records displayed, 17 fields, 45 MB

Food_Inspections_20240223.tsv

dcm:SHUBH
Table=stg_Chicago_Alteryx

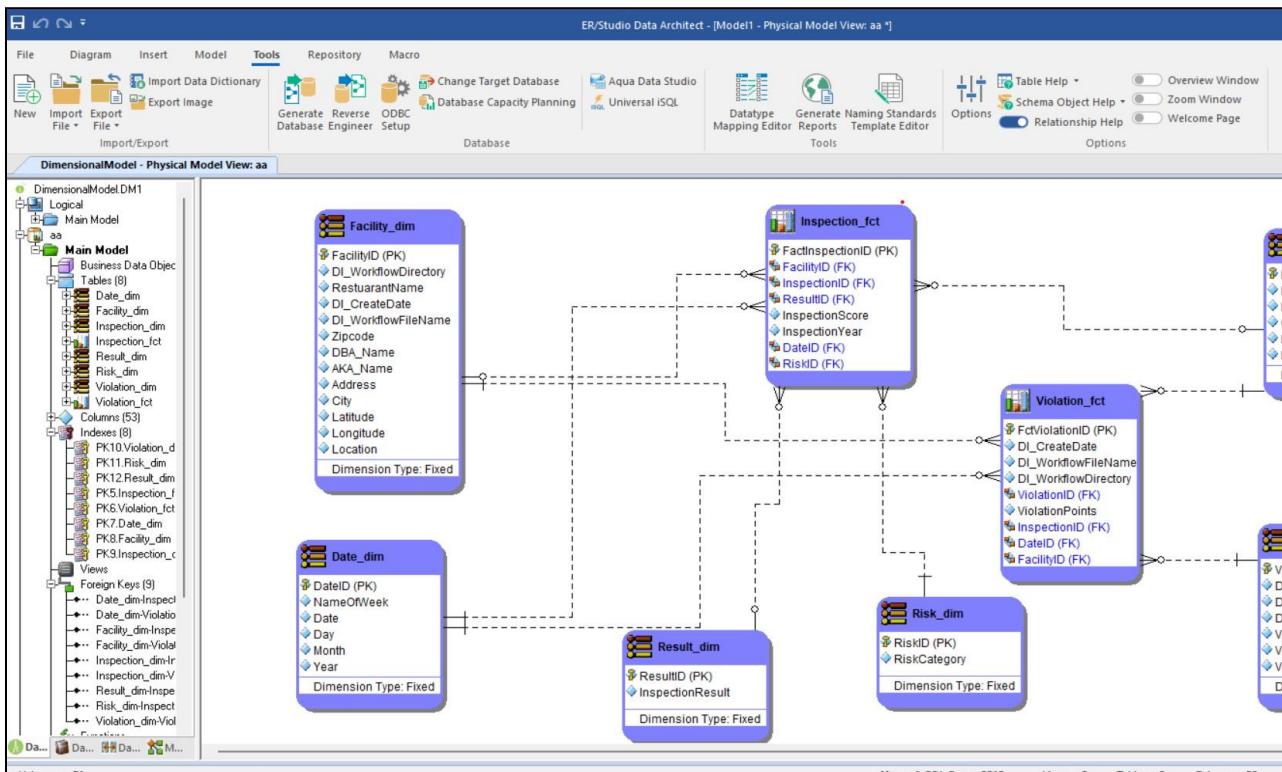
Results - Browse (37) - Input

Record	Inspection ID	DBA Name	AKA Name	License #	Facility Type
1	2589808	LADYBUG & FRIEND'S DAYCARE	LADYBUG & FRIEND'S DAYCARE AND PRESCHOOL	2694470	Children's Services Fac
2	2589815	TACOMBI	TACOMBI	2951873	Restaurant
3	2589772	Thomas, Velma ECC	Thomas, Velma ECC	26891	School
4	2589771	LA SERRE/BAR LA RUE	LA SERRE/BAR LA RUE	2895520	Restaurant
5	2589779	NANDO'S PERI-PERI	NANDO'S PERI-PERI	2379215	Restaurant
6	2589762	MANOLOS TAMALES 4	MANOLOS TAMALES	2745155	Restaurant
7	2589732	THE SIT DOWN	THE SIT DOWN	1904432	Restaurant

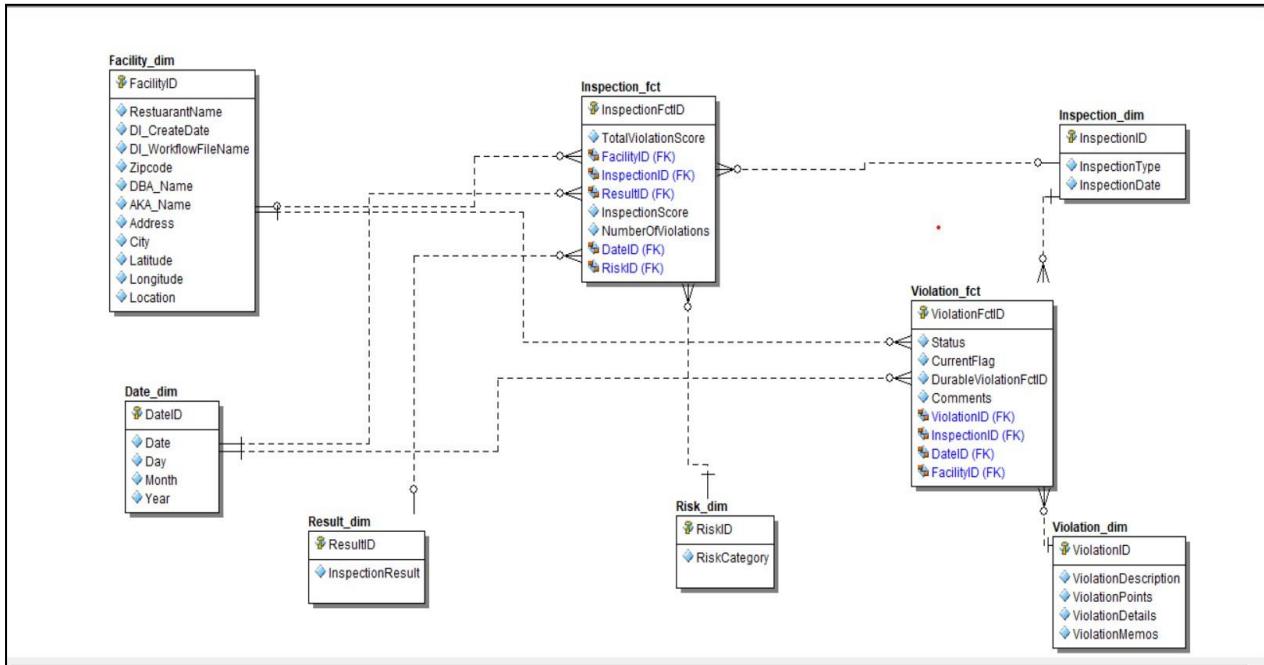


BUILDING DIMENSIONAL MODEL AND DDL SCRIPTS

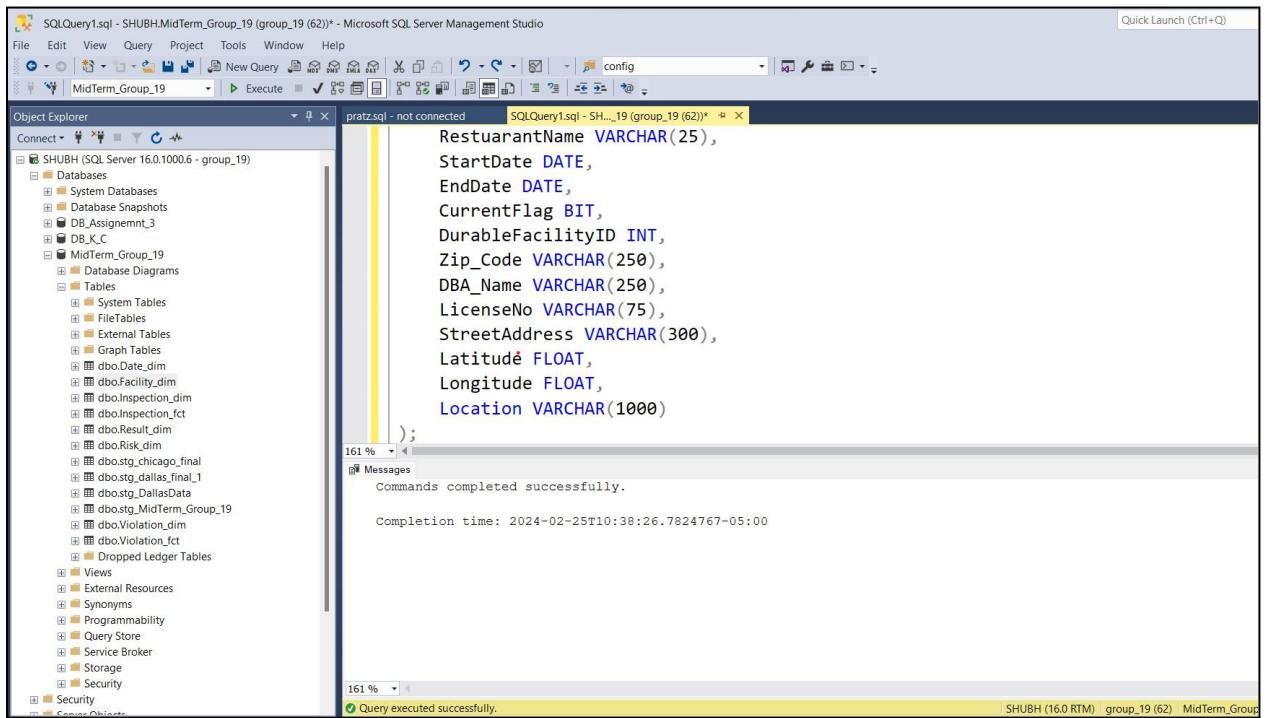
Physical Model



Logical Model



DDL Scripts



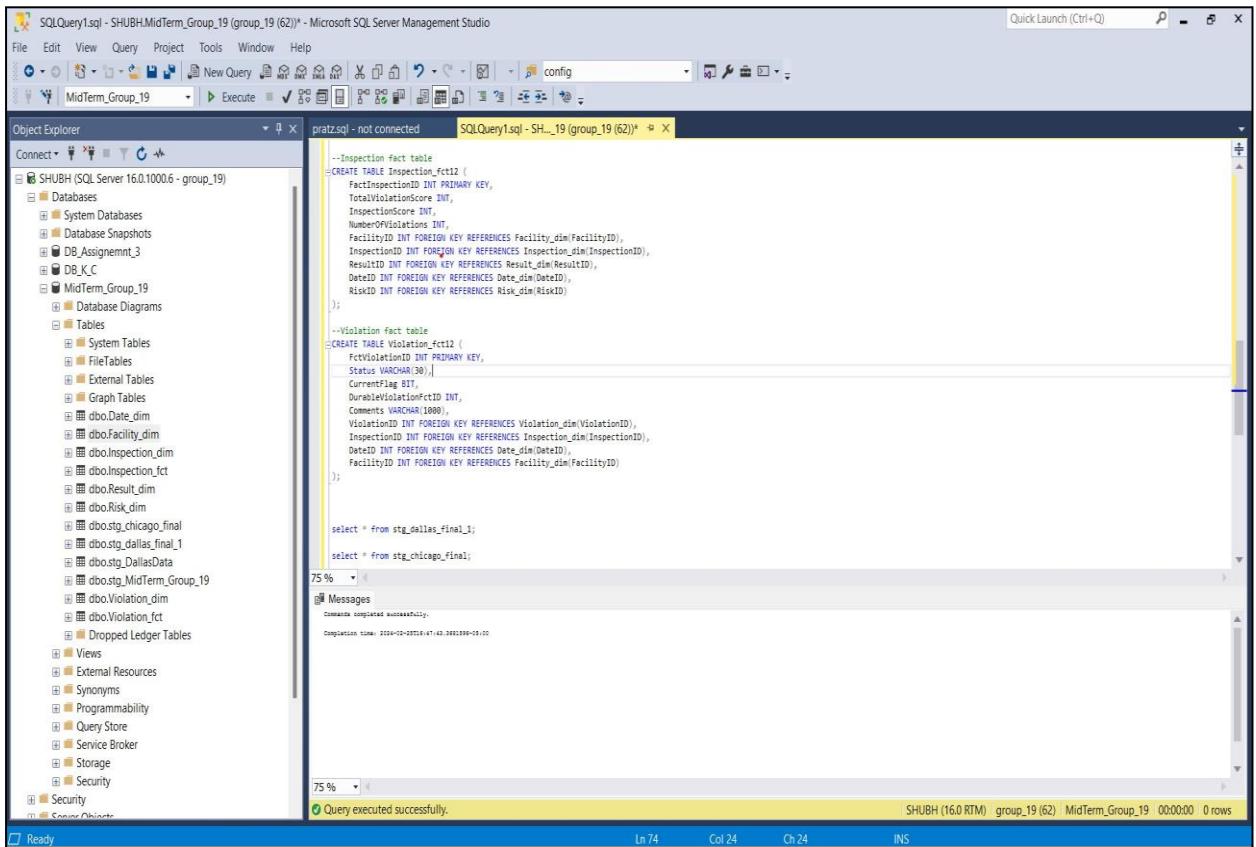
SQLQuery1.sql - SHUBH.MidTerm_Group_19 (group_19 (62)) - Microsoft SQL Server Management Studio

```
CREATE TABLE Restaurant (
    RestaurantName VARCHAR(25),
    StartDate DATE,
    EndDate DATE,
    CurrentFlag BIT,
    DurableFacilityID INT,
    Zip_Code VARCHAR(250),
    DBA_Name VARCHAR(250),
    LicenseNo VARCHAR(75),
    StreetAddress VARCHAR(300),
    Latitude FLOAT,
    Longitude FLOAT,
    Location VARCHAR(1000)
);
```

Commands completed successfully.

Completion time: 2024-02-25T10:38:26.7824767-05:00

Query executed successfully.



SQLQuery1.sql - SHUBH.MidTerm_Group_19 (group_19 (62)) - Microsoft SQL Server Management Studio

```
--Inspection fact table
CREATE TABLE Inspection_fct12 (
    FacilityID INT PRIMARY KEY,
    InspectionScore INT,
    NumberOfViolations INT,
    FacilityID INT FOREIGN KEY REFERENCES Facility_dim(FacilityID),
    InspectionID INT FOREIGN KEY REFERENCES Inspection_dim(InspectionID),
    ResultID INT FOREIGN KEY REFERENCES Result_dim(ResultID),
    DateID INT FOREIGN KEY REFERENCES Date_dim(DateID),
    RiskID INT FOREIGN KEY REFERENCES Risk_dim(RiskID)
);

--Violation fact table
CREATE TABLE Violation_fct12 (
    FctViolationID INT PRIMARY KEY,
    Status VARCHAR(30),
    CurrentFlag BIT,
    DurationViolationFctID INT,
    Comments VARCHAR(1000),
    ViolationID INT FOREIGN KEY REFERENCES Violation_dim(ViolationID),
    InspectionID INT FOREIGN KEY REFERENCES Inspection_dim(InspectionID),
    DateID INT FOREIGN KEY REFERENCES Date_dim(DateID),
    FacilityID INT FOREIGN KEY REFERENCES Facility_dim(FacilityID)
);

select * from stg_dallas_final_1;
select * from stg_chicago_final;
```

Commands completed successfully.

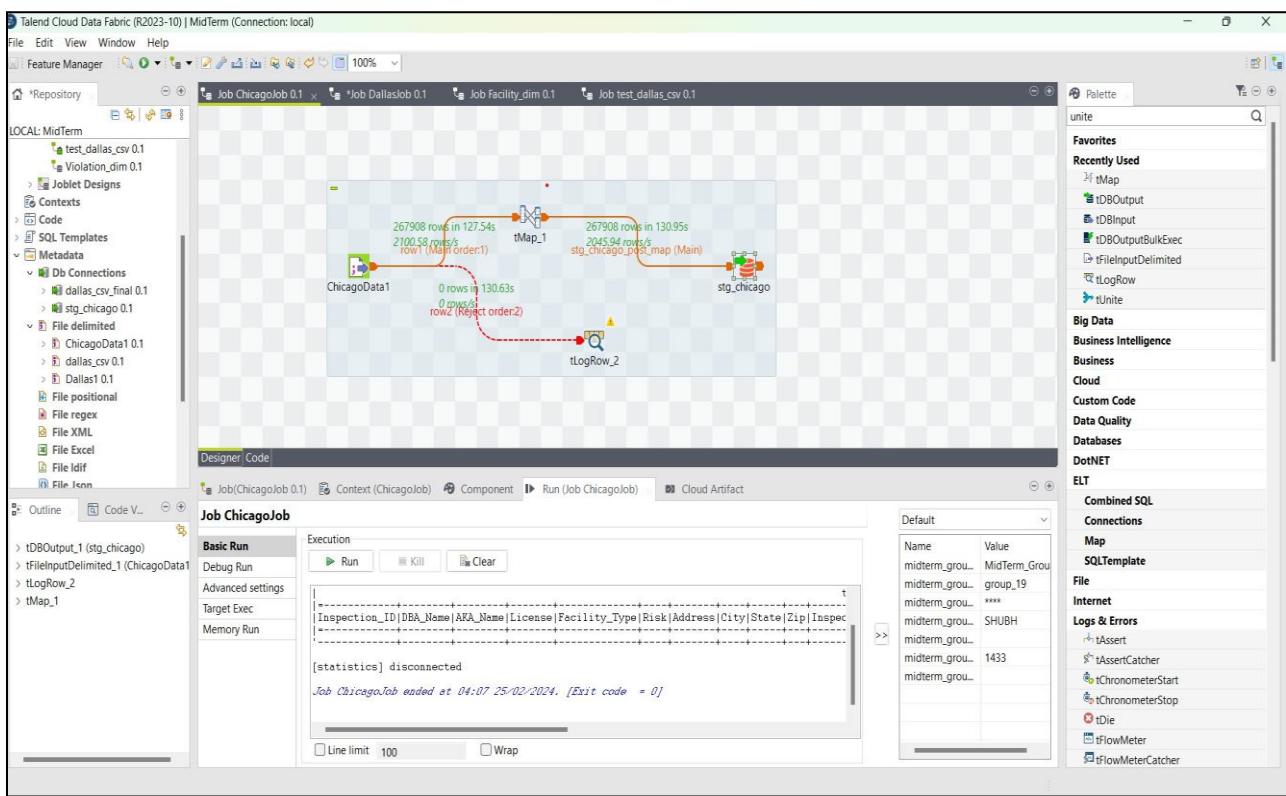
Completion time: 2024-02-25T10:47:43.0000000-05:00

Query executed successfully.

[-]  MidTerm_Group_19
[-]  Database Diagrams
[-]  Tables
[+]  System Tables
[+]  FileTables
[+]  External Tables
[+]  Graph Tables
[+]  dbo.Date_dim
[+]  dbo.Facility_dim
[+]  dbo.Inspection_dim
[+]  dbo.Inspection_fct
[+]  dbo.Result_dim
[+]  dbo.Risk_dim
[+]  dbo.stg_chicago_final
[+]  dbo.stg_dallas_final_1
[+]  dbo.stg_DallasData
[+]  dbo.stg_MidTerm_Group_19
[+]  dbo.Violation_dim
[+]  dbo.Violation_fct
[-]  Dropped Ledger Tables
[+]  Views
[+]  External Resources
[+]  Synonyms
[+]  Programmability
[+]  Query Store
[+]  Service Broker
[+]  Storage
[+]  Security
[+]  Security
[+]  Server Objects
[+]  Replication
[+]  Always On High Availability
[+]  Management
[-]  Information Schema Catalog

STAGING IN TALEND

Chicago Staging in Talend:



ScreenShot in MSSQL server of Chicago Data

The screenshot shows the Microsoft SQL Server Management Studio (SSMS) interface. The Object Explorer on the left shows the database structure, including 'SHUBH' and 'MidTerm_Group_19' databases. The central pane displays a query results window for the 'MidTerm_Group_19' database. The query executed was:

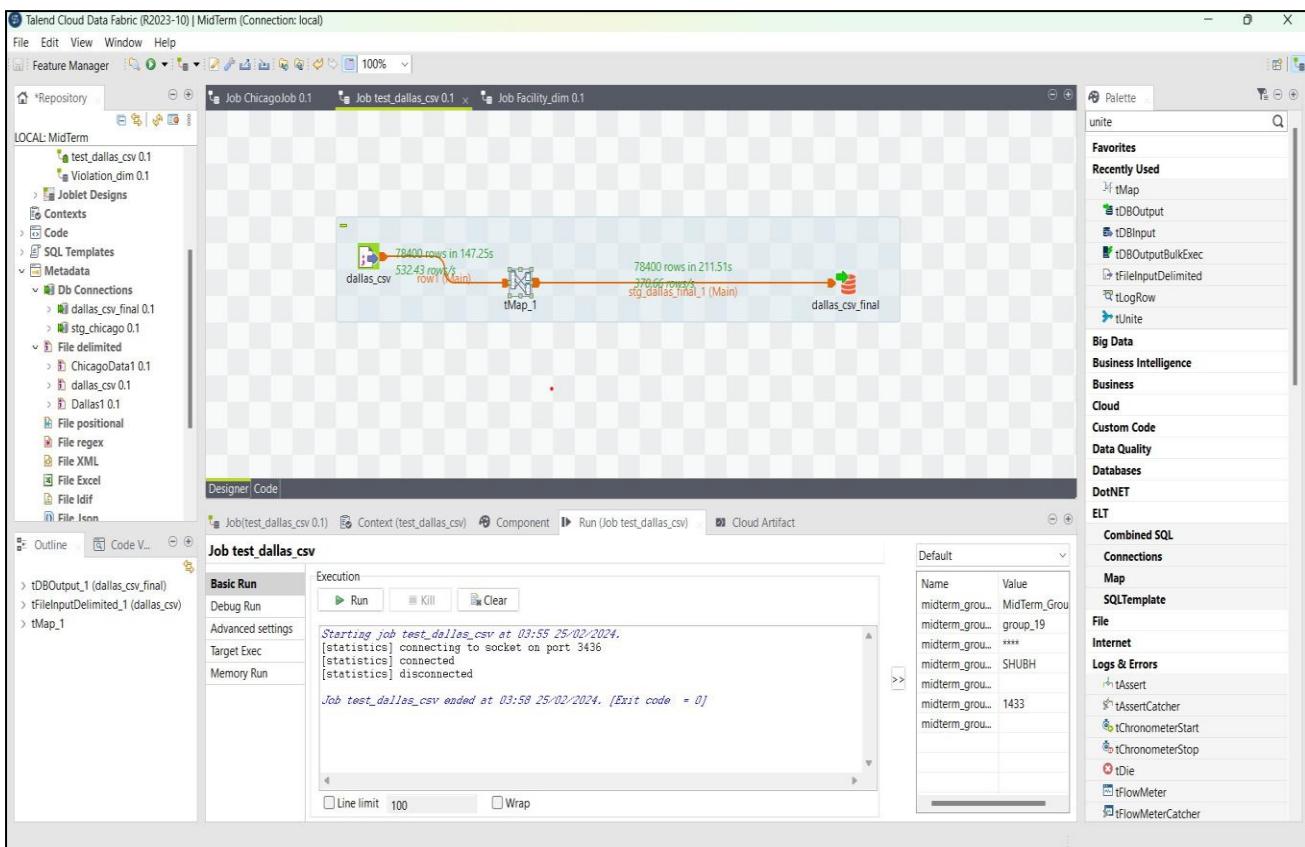
```
use MidTerm_Group_19;

select * from stg_dallas_final_1;
select * from stg_chicago_final;
```

The results show data for various restaurants and schools in Chicago and Dallas. The data is presented in two tables:

Inspection_ID	DBA_Name	AKA_Name	License	Facility_Type	Risk	Address	City	State	Zip	Inspect
1	2569808	LADYBUG & FRIEND'S DAYCARE	2694470	Children's Services Facility	Risk 1 (High)	7648-54 W TOUHY AVE	CHICAGO	IL	60631	2024-0
2	2569815	TACOMBI	2951873	Restaurant	Risk 1 (High)	1442 N MILWAUKEE AVE	CHICAGO	IL	60622	2024-0
3	2569772	Thomas, Velma ECC	26891	School	Risk 1 (High)	3625 S Horne (2100W) (St. Maurice)	CHICAGO	IL	60609	2024-0
4	2569771	LA SERRE/BAR LA RUE	2999520	Restaurant	Risk 3 (Low)	820 W FULTON MARKET BLDG	CHICAGO	IL	60607	2024-0
5	2569779	NANDO'S PERI-PERI	2379215	Restaurant	Risk 1 (High)	670 W DIVERSEY PKWY	CHICAGO	IL	60614	2024-0
6	2569762	MANOLOS TAMALES 4	2745155	Restaurant	Risk 1 (High)	6641 S CICERO AVE	CHICAGO	IL	60638	2024-0
7	2569732	THE SIT DOWN	1904432	Restaurant	Risk 1 (High)	1312 E 53RD ST	CHICAGO	IL	60615	2024-0
8	2569743	LA BESOTTEUR LA RUE	2998111	Restaurant	Risk 1 (High)	501 S WOODLAWN MARKET BLDG	CHICAGO	IL	60615	2024-0
9	2569749	NICKY'S CHINESE FOOD	126747	Restaurant	Risk 1 (High)	501 S WOODLAWN AVE	CHICAGO	IL	60615	2024-0
10	2569760	HACHI SUSHI	2636936	Restaurant	Risk 1 (High)	3448 E 118TH ST	CHICAGO	IL	60617	2024-0
11	2569669	CUBANITO EXPRES	2652053	Restaurant	Risk 1 (High)	4720 S PULASKI RD	CHICAGO	IL	60632	2024-0
12	2569669	HUA'S CHINESE CHOP SUEY EXPRESS	2147319	Restaurant	Risk 1 (High)	301 E 91ST ST	CHICAGO	IL	60617	2024-0
13	2569654	THE PROMONTORY	2269538	Restaurant	Risk 1 (High)	5307-5319 S LAKE PARK AVE	CHICAGO	IL	60615	2024-0
14	2569645	NOORI CHICKEN CHICAGO	2662358	Restaurant	All	3223 N CLARK ST	CHICAGO	IL	60657	2024-0
15	2569646	Ellington Elementary School	23101	School	Risk 1 (High)	243 N Parkside	CHICAGO	IL	60644	2024-0

Dallas Staging in Talend:



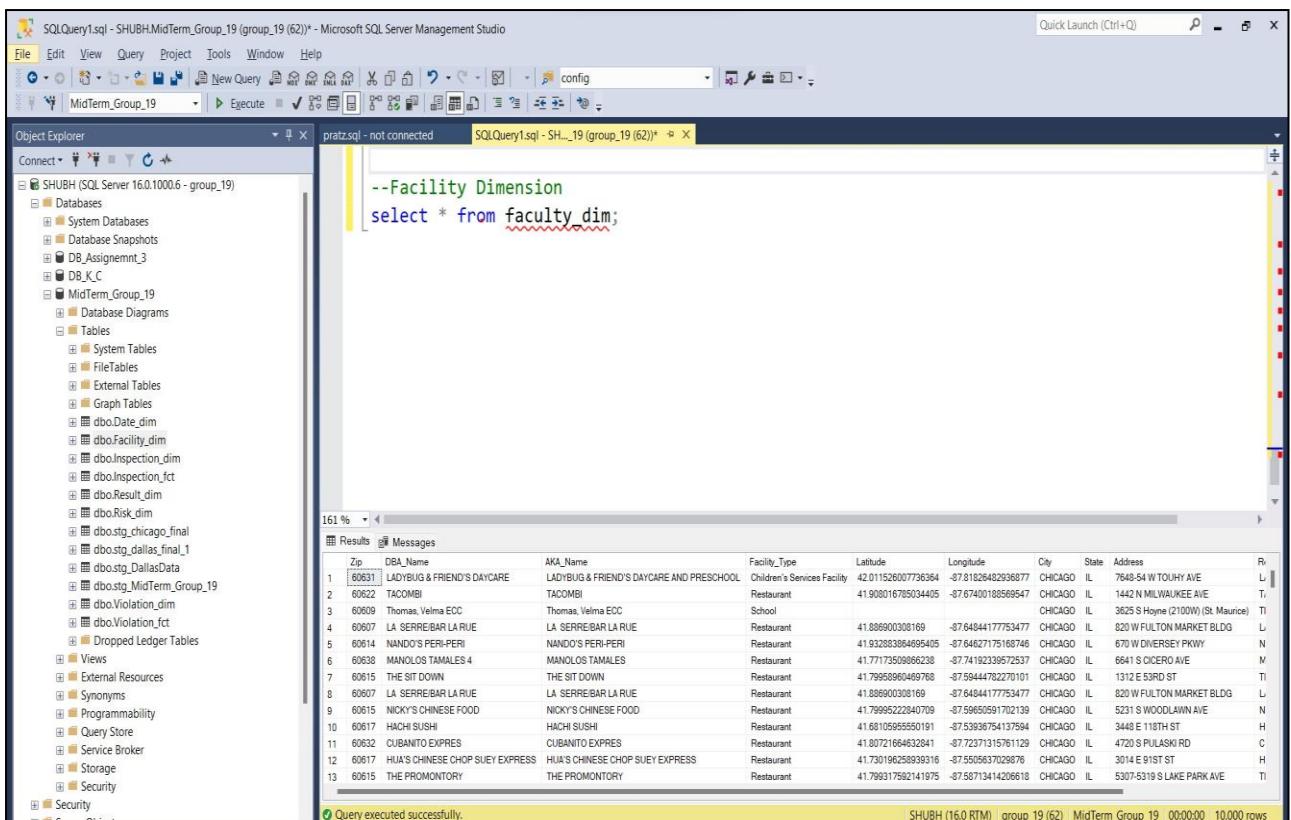
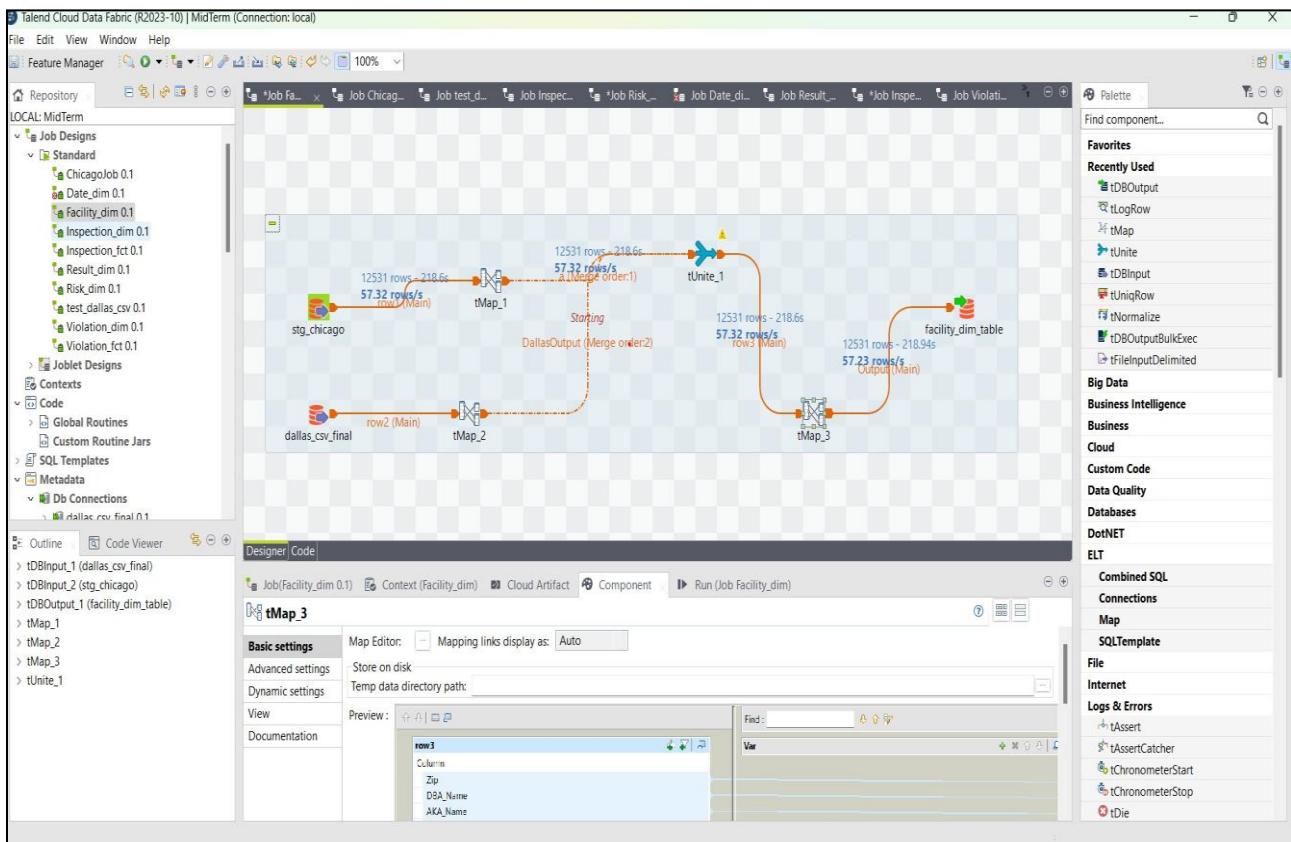
ScreenShot in MSSQL server of Dallas Data

The screenshot shows Microsoft SQL Server Management Studio (SSMS) with a query window titled 'SQLQuery1.sql - SHUBH.MidTerm_Group_19 (group_19 (59)) Executing...'. The query runs a USE statement for 'MidTerm_Group_19' and then selects all columns from two tables: 'stg_dallas_final_1' and 'stg_chicago_final'. The results pane displays the data from both tables. The Object Explorer on the left shows the database structure, including tables like 'Facility_dim', 'Inspection_dim', and 'Violation_dim'.

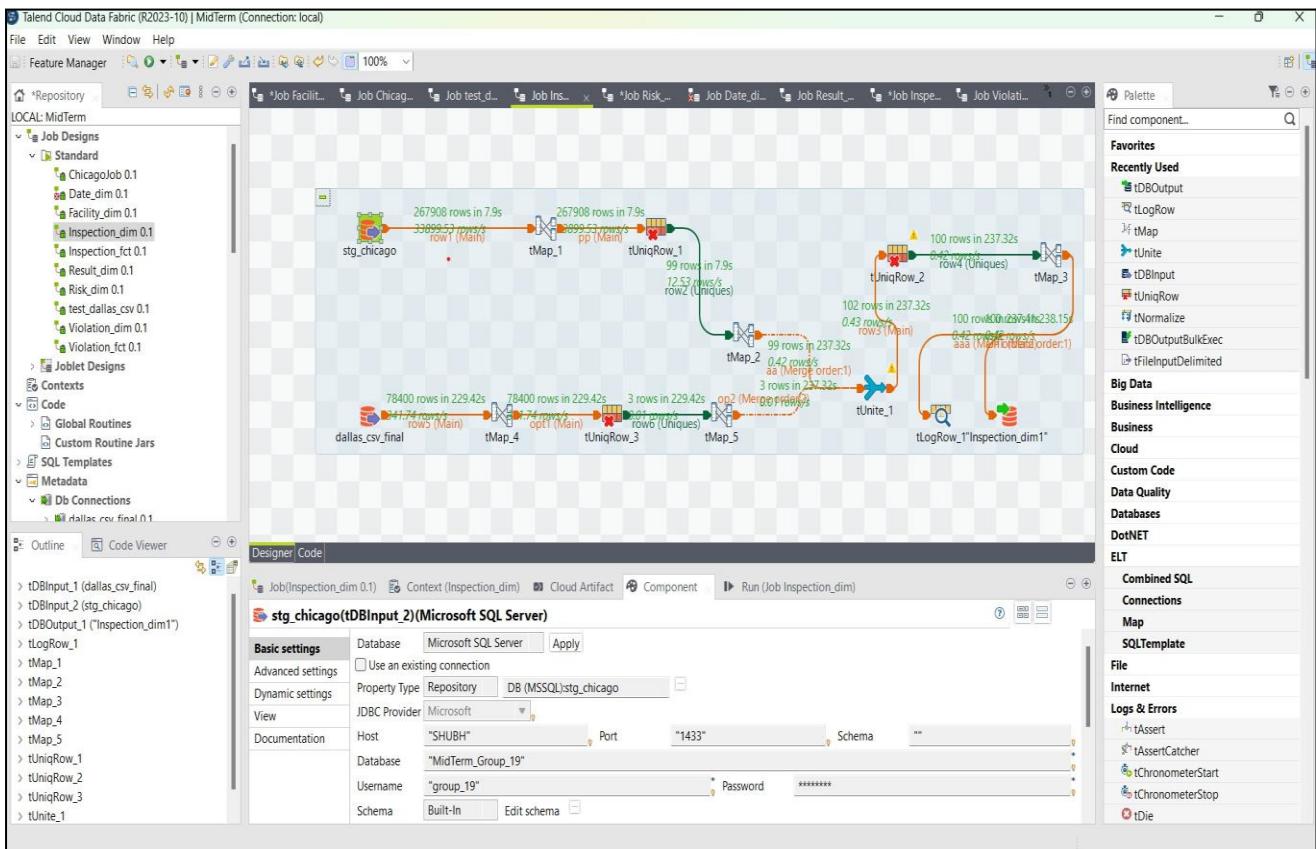
Restaurant_Name	Inspection_Type	Inspection_Date	Inspection_Score	Street_Number	Street_Name	Street_Direction	Street_Type	Street_Unit	Street_Address	Zip_Code	Violation_L
MICKLE CHICKEN	Routine	10/30/2019	100	3203	CAMP WISDOM	W	RD		3203 W CAMP WISDOM RD	75237	
TOM THUMB - JUICE BAR	Routine	04/08/2020	100	2380	FIELD	N	ST		2380 N FIELD ST	75021	
BROOKDALE WHITE ROCK	Routine	07/29/2020	97	9271	WHITE ROCK	TRL			9271 WHITE ROCK TRL	75238	*21 RFSM
CHURCH'S CHICKEN #201	Routine	09/14/2020	100	10295	FERGUSON	RD			10295 FERGUSON RD	75228	
PEAK PREPARATORY-PRIMARY SCHOOL	Routine	04/25/2017	98	1474	ANNEX	AVE			1474 ANNEX AVE	75204	*21 RFSM
SAMS CLUB #6482 - DEMO ROOM	Routine	04/05/2021	100	5555	BUCKNER	S	BLVD		5555 S BUCKNER BLVD	75227	
AMC THEATRES NORTHPARK 15 (BAR)	Routine	04/09/2021	100	8887	CENTRAL	N	EXPW	#3000	8887 N CENTRAL EXPW #3000	75225	
SAM'S CLUB #282 - DEMO ROOM	Routine	11/14/2021	100	2900	WHEATLAND	W	RD		2900 N WHEATLAND RD	75237	
UCHI DALLAS	Routine	03/11/2020	98	2817	MAPLE	AVE	#110		2817 MAPLE AVE #110	75201	*32 Equipr
UNREFINED BAKERY	Routine	03/23/2020	100	2222	MCKINNEY	AVE	#180		2222 MCKINNEY AVE #180	75201	
KIMBALL HIGH SCHOOL	Routine	11/18/2019	100	3806	WESTMORELAND	S	RD		3806 S WESTMORELAND	75233	
SHELL EMPIRE	Routine	07/15/2020	97	8390	STEMMONS	N	FRWY	#B	8390 N STEMMONS FRWY #B	75247	*18 Toxic
NOVA ACADEMY	Routine	10/08/2020	100	6459	SCYENE	RD			6459 SCYENE RD	75227	
UNPREPARED SECONDARY	Routine	12/01/2020	100	2625	ELM	ST			2625 ELM ST	75226	
AFC SUSHI	Routine	04/12/2021	98	1441	BECKLEY	N	AVE		1441 N BECKLEY AVE	75208	*29 Conce
COLLEGATION A&P	Routine	03/11/2020	98	2200	FRUIT	RD	RD		2200 FRUIT RD	75201	

TRANSFORMATION AND INTEGRATION INTO FACT AND DIM TABLES IN TALEND

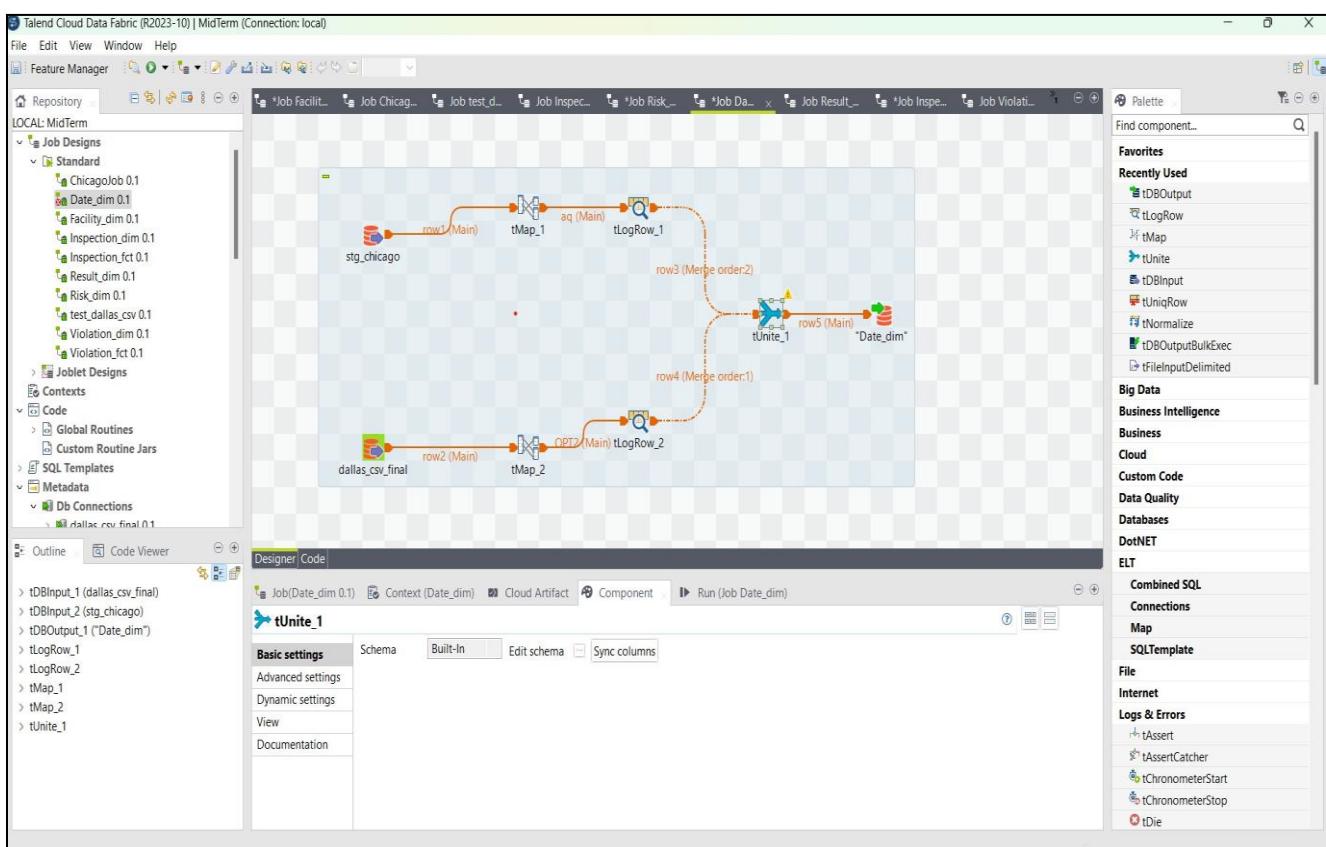
Facility Dimension



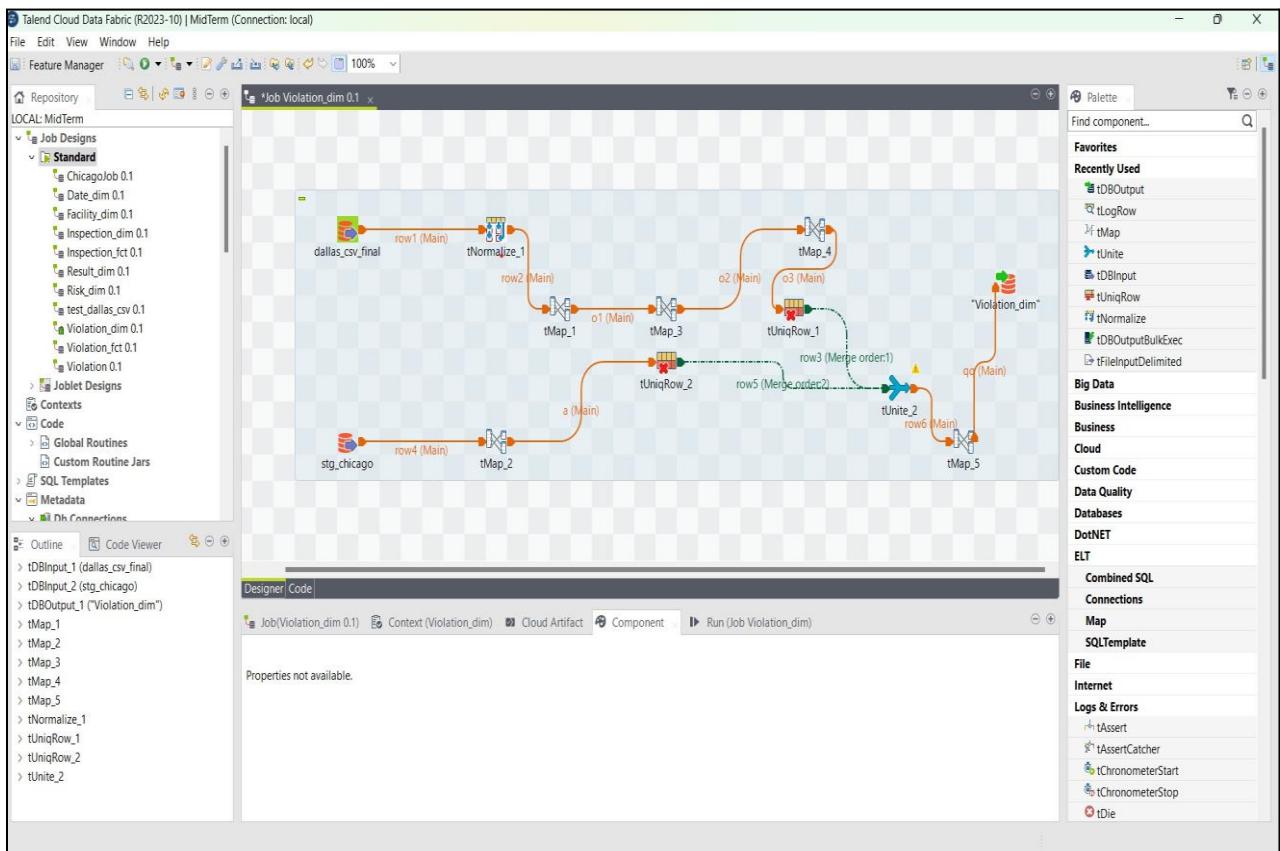
Inspection Dimension



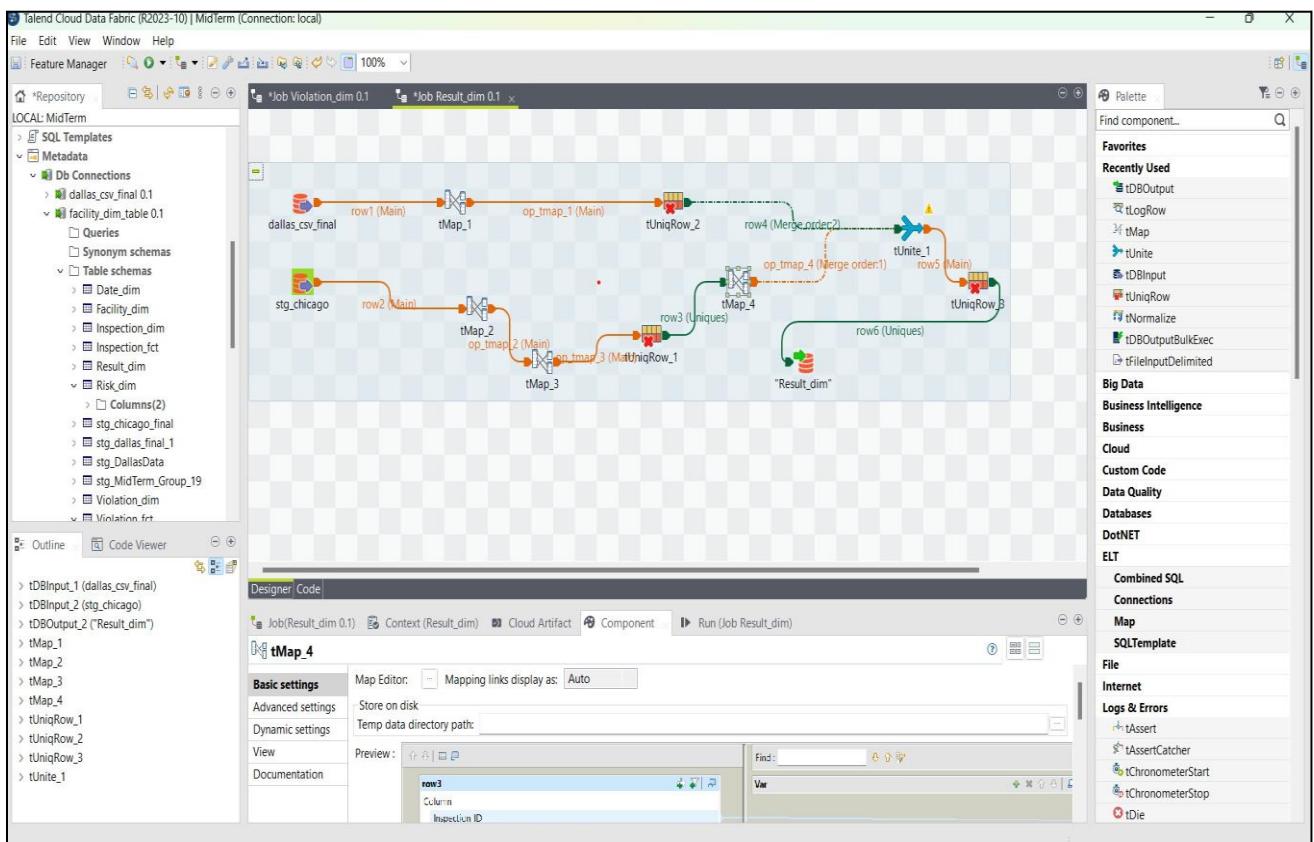
Date Dimension



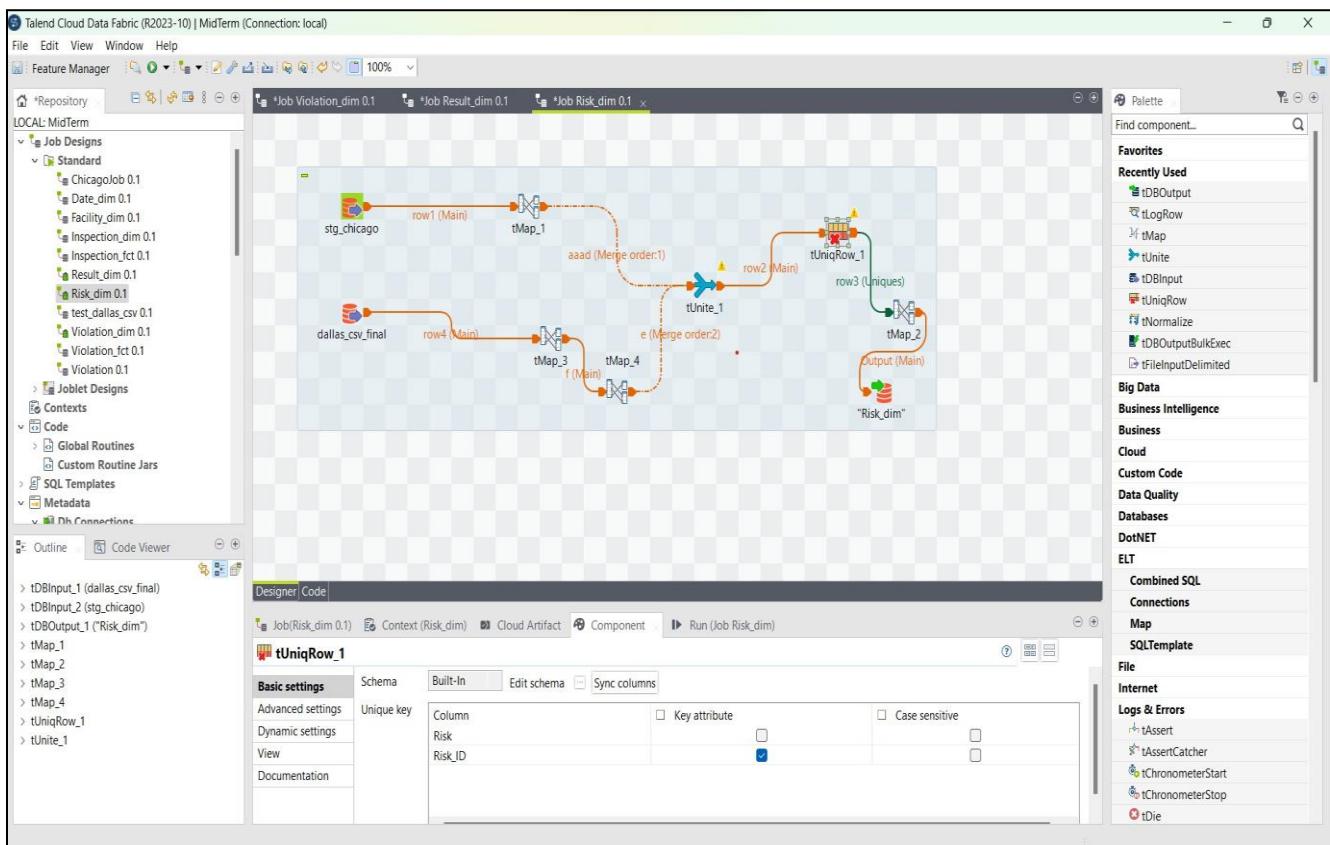
Violation Dimension



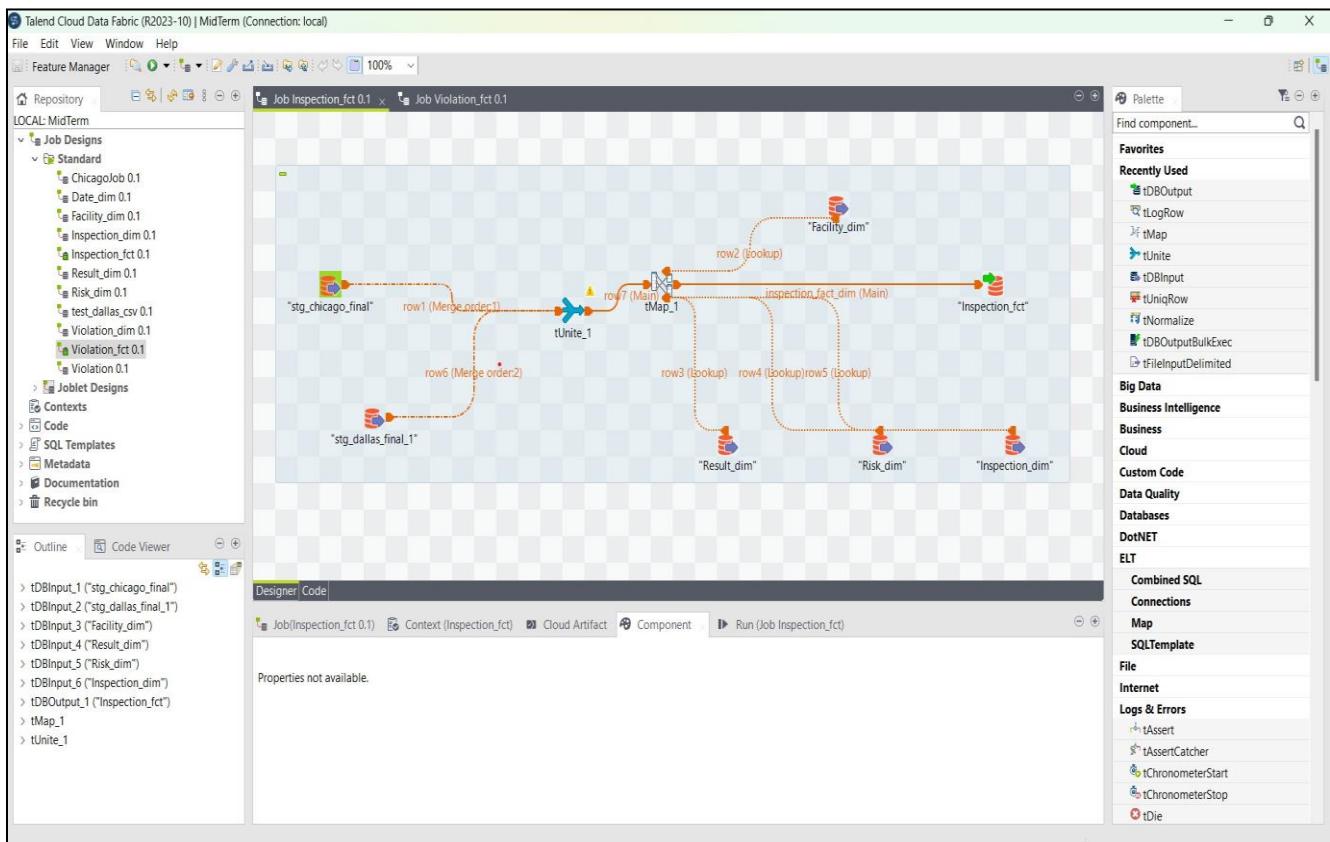
Result Dimension



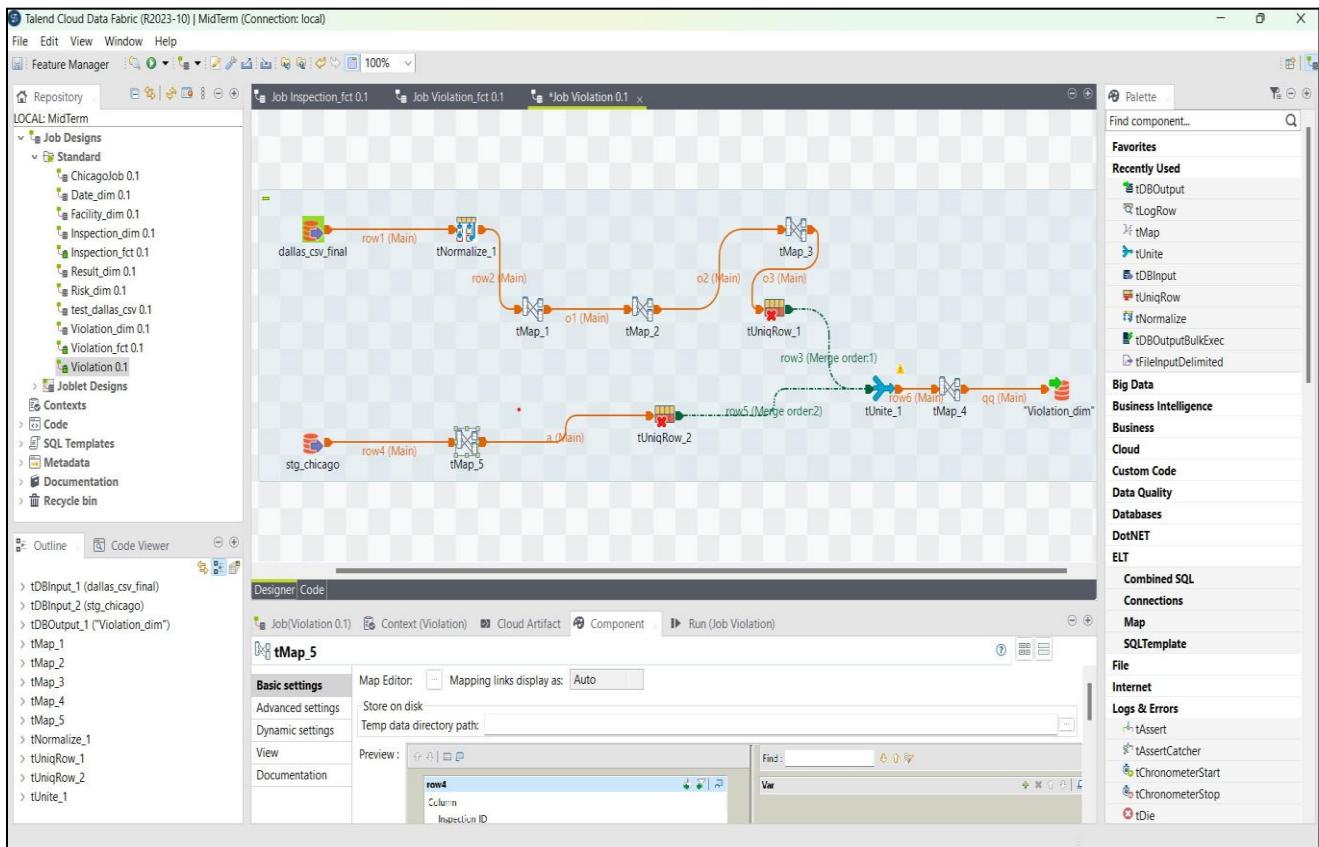
Risk Dimension



Inspection Fact



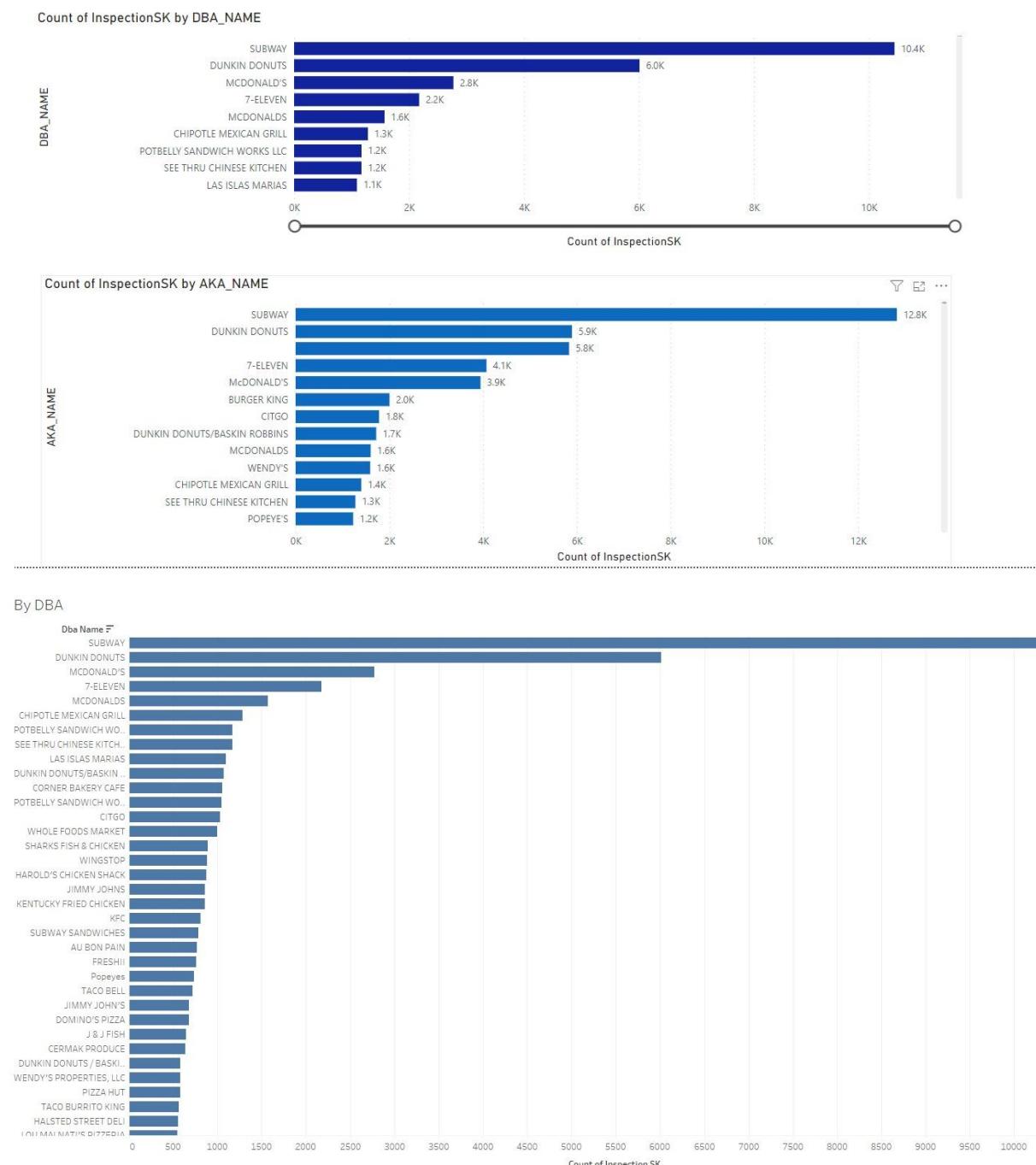
Violation Fact



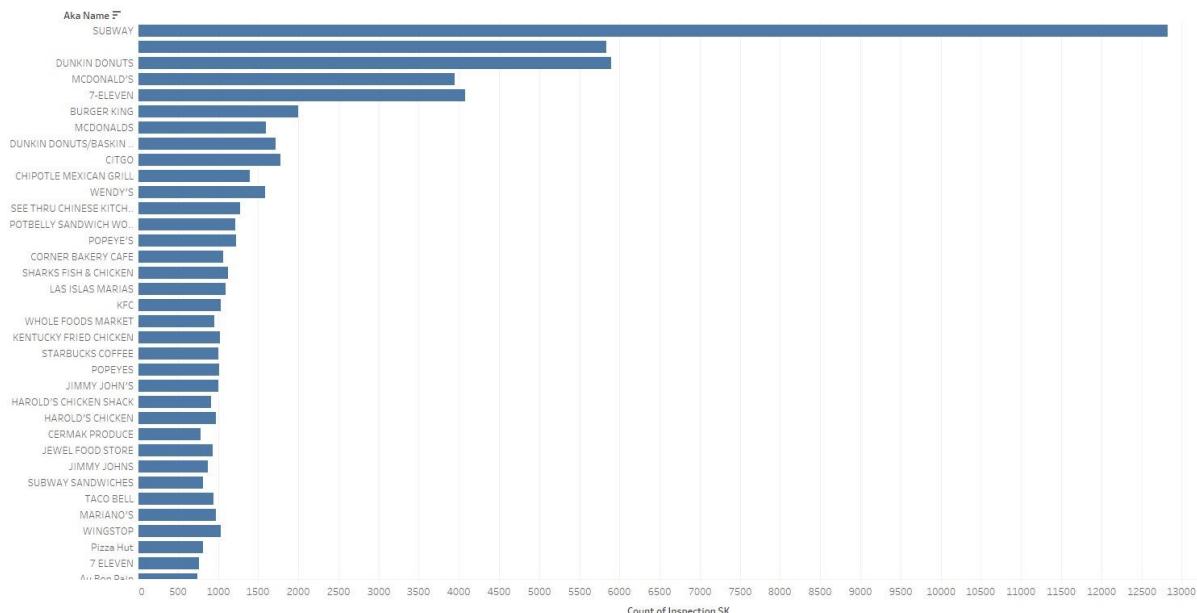
Part 04

Visualisations in Power BI and Tableau

Count of InspectionSK by DBA name and AKA name



By AKA



```

SELECT
    fp.AKA_NAME,
    COUNT(fi.InspectionSK) AS InspectionCount
FROM
    DIM_FoodPlaces AS fp
JOIN
    FACT_FoodInspection AS fi
    ON fp.FoodPlaceSK = fi.FoodPlaceSK
GROUP BY
    fp.AKA_NAME
ORDER BY
    InspectionCount DESC;

```

Results		Messages
AKA_NAME	InspectionCount	
SUBWAY	12827	
DUNKIN DONUTS	5896	
	5835	
7-ELEVEN	4073	
MCDONALD'S	3945	
BURGER KING	2003	
CITGO	1780	
DUNKIN DONUTS/BASKIN ROBBINS	1720	
McDONALDS	1602	
WENDY'S	1591	
CHIPOTLE MEXICAN GRILL	1401	
SEE THRU CHINESE KITCHEN	1276	

```

    select count(*) from FACT_FoodInspection

SELECT
    fp.DBA_NAME,
    COUNT(fi.InspectionSK) AS InspectionCount
FROM
    DIM_FoodPlaces AS fp
JOIN
    FACT_FoodInspection AS fi
    ON fp.FoodPlaceSK = fi.FoodPlaceSK
GROUP BY
    fp.DBA_NAME
ORDER BY
    InspectionCount DESC;

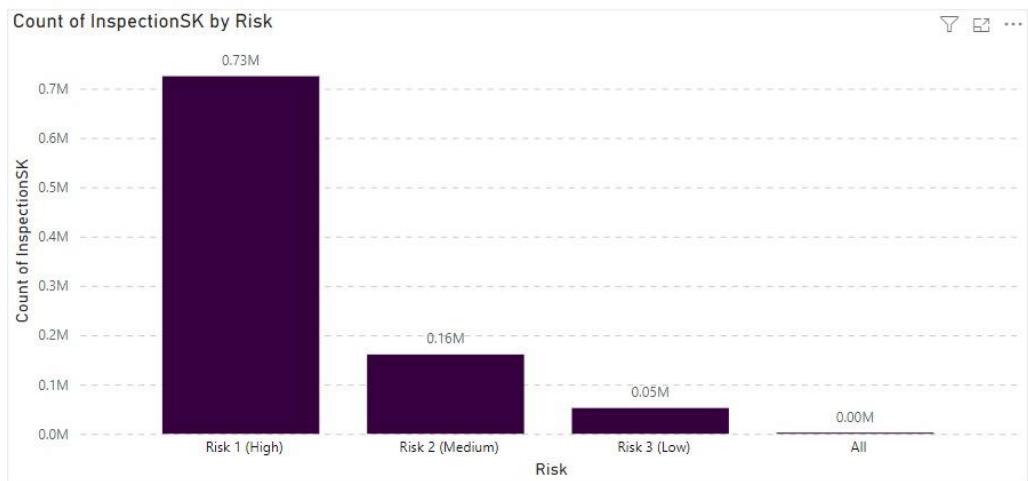
```

00 % ▾

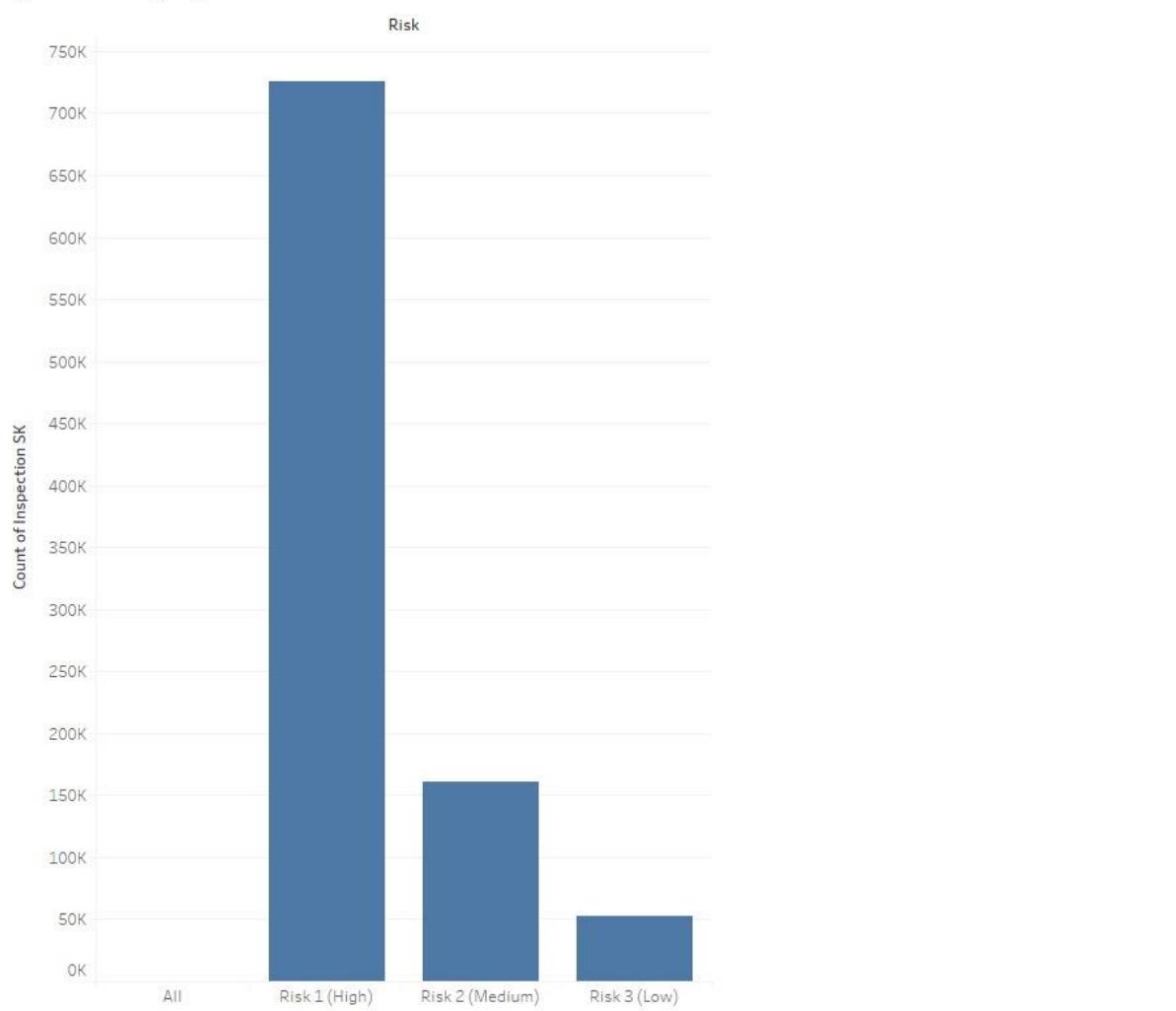
Results Messages

	DBA_NAME	InspectionCount
1	Subway	10447
2	DUNKIN DONUTS	6010
3	McDonald's	2772
4	7-Eleven	2176
5	McDONALDS	1574
6	CHIPOTLE MEXICAN GRILL	1284
7	POTBELLY SANDWICH WORKS LLC	1173
8	See Thru Chinese Kitchen	1172
9	LAS ISLAS MARIAS	1093
10	DUNKIN DONUTS/BASKIN ROBBINS	1070
11	CORNER BAKERY CAFE	1055
12	POTBELLY SANDWICH WORKS	1046
13	CITGO	1031
14	WHOLE FOODS MARKET	999
15	SHARKS FISH & CHICKEN	890
16	WINGSTOP	881
17	HAROLD'S CHICKEN SHACK	872
18	Jimmy Johns	857
19	KENTUCKY FRIED CHICKEN	856
20	KFC	800

Count Of Inspection SK by Risk



By Risk Category



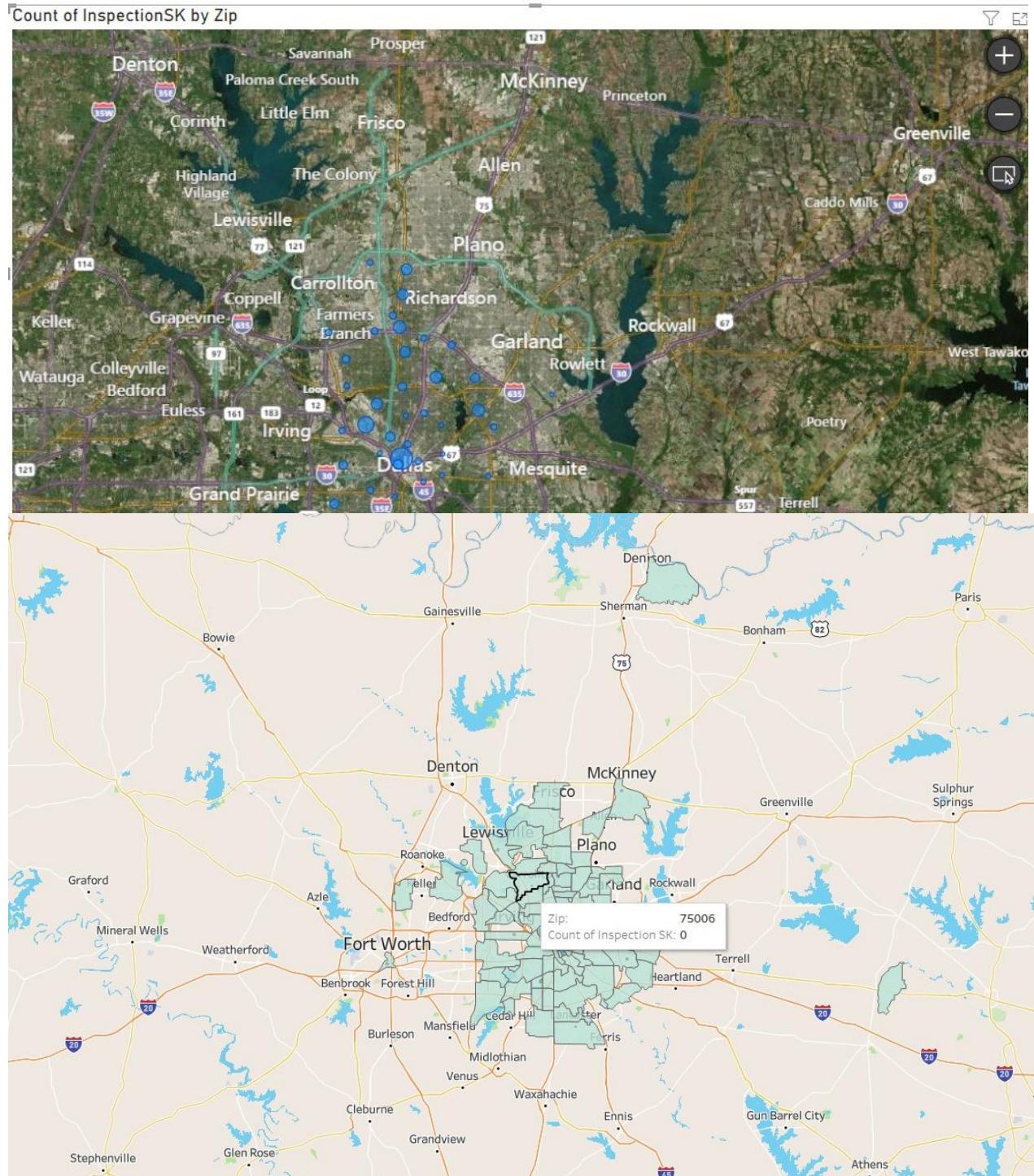
```
SELECT
    r.Risk,
    COUNT(fi.InspectionSK) AS InspectionCount
FROM
    DIM_Risk AS r
JOIN
    FACT_FoodInspection AS fi
    ON r.RiskSK = fi.RiskSK
GROUP BY
    r.Risk
ORDER BY
    InspectionCount DESC;
```

3 %

Results Messages

Risk	InspectionCount
Risk 1 (High)	725209
Risk 2 (Medium)	160840
Risk 3 (Low)	52253
All	57

Count of InspectionSK by zipcode



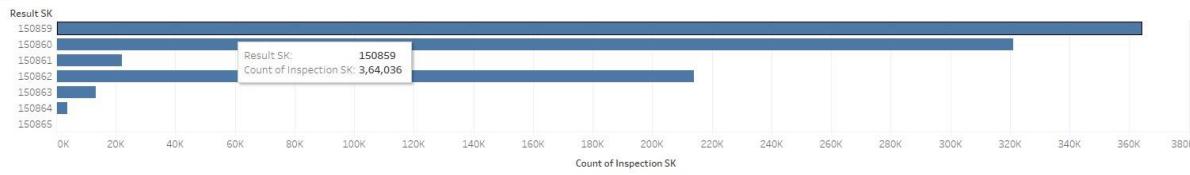
Count of InspectionSK by Result



938.36K

Count of InspectionSK

By Result

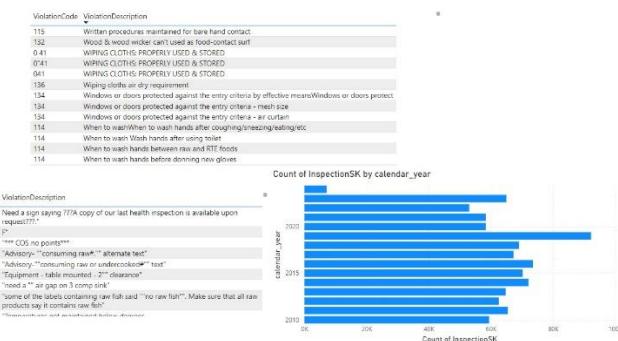


```
--by result
SELECT
    dr.Result,
    COUNT(fi.InspectionSK) AS InspectionCount
FROM
    DIM_Result AS dr
JOIN
    FACT_FoodInspection AS fi
    ON dr.ResultSK = fi.ResultSK
GROUP BY
    dr.Result
```

Results Messages

Result	InspectionCount
Pass	364036
Fail	321156
Pass w/ Conditions	213922
Out of Business	22169
No Entry	13378
Not Ready	3620
Business Not Located	78

Count by InspectionSk by Violation

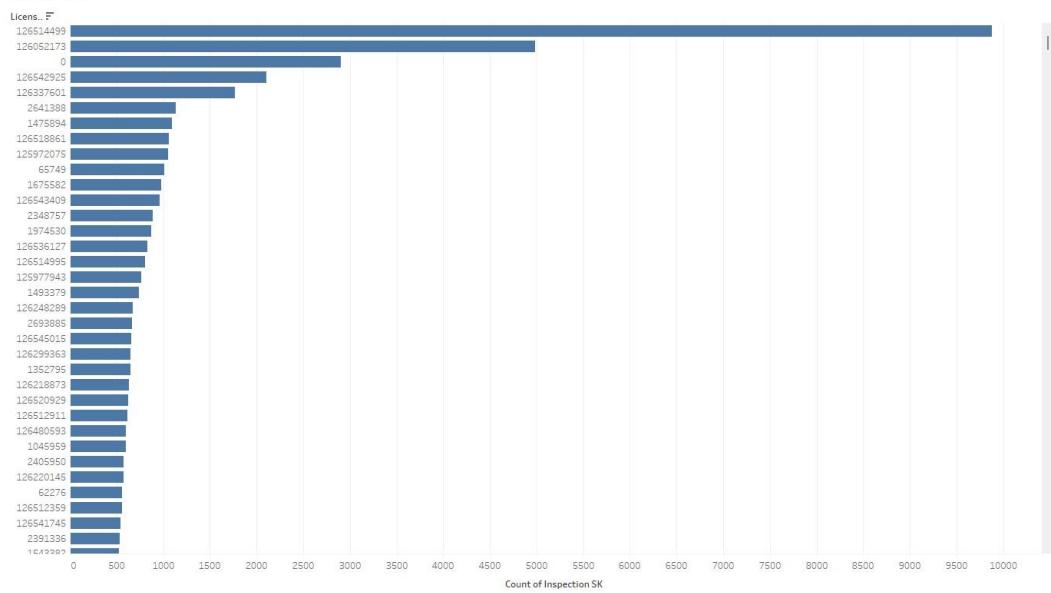


By Violation Code and Description

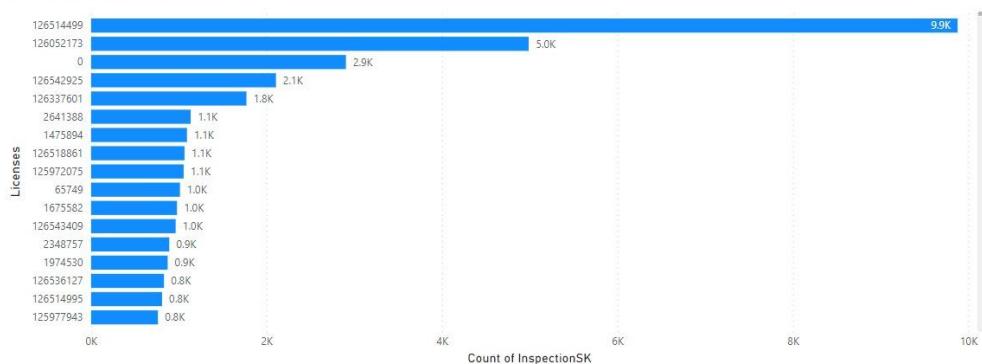
Violation Code	Violation Description
1Sanitize all Equipment	Sanitize all Equipment
1Prevent contamination b...	Prevent contamination by..
1Outdoor bars single serv...	Outdoor bars single servi..
1Outdoor bars dispense ic...	Outdoor bars dispense ice..
1Hair restraints/exer	1Outdoor bars dispense ice thr
1HACCP-Plan submitt...	
1Cleaning floors;spills, dri...	Cleaning floors;spills, drip..
19	Outfitter Operations Reduced Oxygen HACCP P.. Reduces oxygen package .. Reduced Oxygen HACCP Pl..
17901	YOUR BUSSINESS NAME ..
17	DATE ALL FOOD PRODUC.. Eggs and milk products, p.. Private homes/ living or s.. Self-Service Food Market Eggs and milk products, p.. Private homes/ living or sl.. Self-Service Food Market
160	CITATION will be issued i..
16	Discard PHF is exceeds ti.. must have grease trap re.. Self-Service Food Market Utensils must be stored ?.. Discard PHF is exceeds ti.. Self-Service Food Market
150	KEEP SANITIZER IN BUCK.. Restrict community gath.. sanitation level weak bel..
147	Central Preparation Facili.. Central Preparation Facili.. Conditions of Permit-in u.. Constructing / Remodelin.. Food prep/utensil wash/.. Handler-Certificate Not o.. Handwashing signage Health permit posted

Count of InspectionSK by Licenses

By License



Count of InspectionSK by Licenses

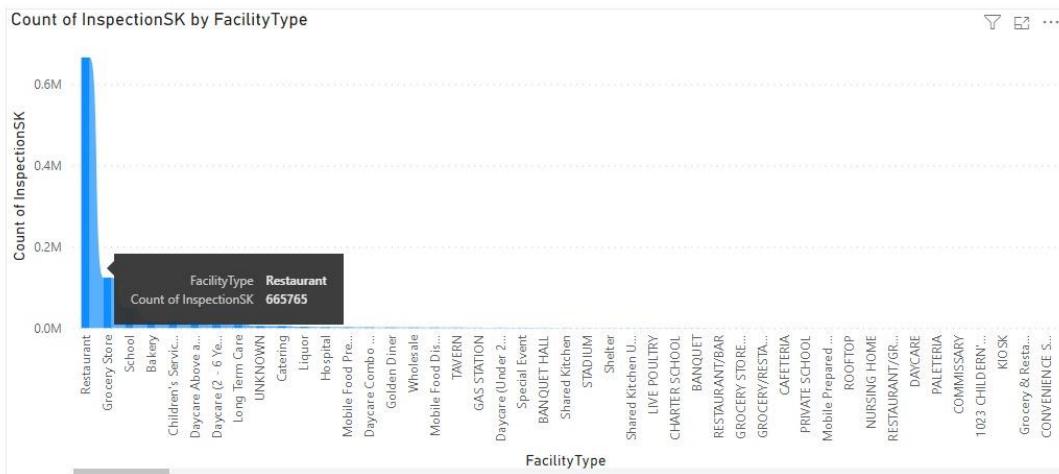


```
--By Facility Type
SELECT
    f.FacilityType,
    COUNT(fi.InspectionSK) AS InspectionCount
FROM
    DIM_Facility AS f
JOIN
    FACT_FoodInspection AS fi
ON f.FacilitySK = fi.FacilitySK
GROUP BY
    f.FacilityType
ORDER BY
    InspectionCount DESC;
```

Results Messages

Licenses	InspectionCount
126514499	9878
126052173	4988
0	2905
126542925	2107
126337601	1771
2641388	1135
1475894	1093
126518861	1065
125972075	1055
65749	1013
1675582	979
126543409	964
2348757	891
1974530	873
126536127	831
126514995	809
125977943	762
1493379	741
126248289	678
2693885	668

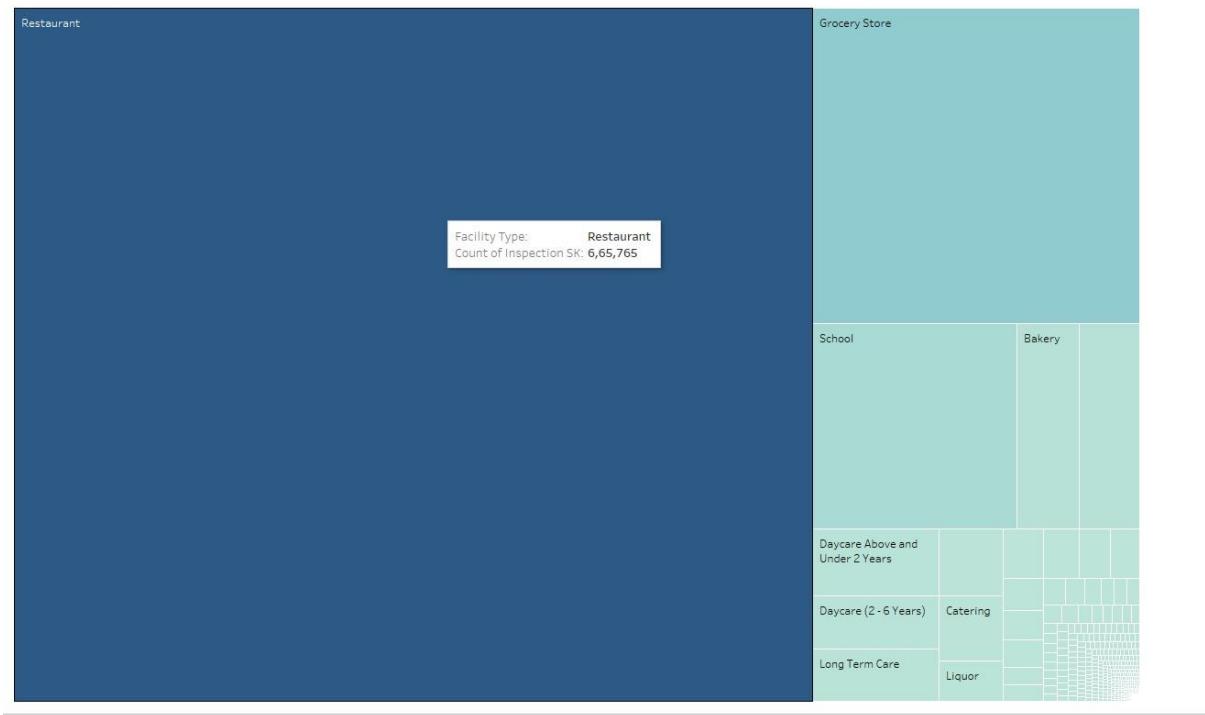
Count of Inspection SK by Facility Type



938.36K

Count of InspectionSK

By Facility Type



```

--by facility type
SELECT
    f.FacilityType,
    COUNT(fi.InspectionSK) AS InspectionCount
FROM
    DIM_Facility AS f
JOIN
    FACT_FoodInspection AS fi
    ON f.FacilitySK = fi.FacilitySK
GROUP BY
    f.FacilityType
ORDER BY
    InspectionCount DESC;

```

Results

FacilityType	InspectionCount
Restaurant	665765
Grocery Store	124173
School	50231
Bakery	15407
Children's Services Facility	14914
Daycare Above and Under 2 Years	10183
Daycare (2 - 6 Years)	8058
Long Term Care	7893
UNKNOWN	5222
Catering	5034
Liquor	3115
Hospital	2384
Mobile Food Preparer	2144
Daycare Combo 1586	1844
Golden Diner	1729
Wholesale	1596
Mobile Food Dispenser	1433
TAVERN	1331
GAS STATION	880
Daycare (Under 2 Years)	802


```

--by inspection type
SELECT
    dit.InspectionType,
    COUNT(fi.InspectionSK) AS InspectionCount
FROM
    DIM_InspectionType AS dit
JOIN
    FACT_FoodInspection AS fi
    ON dit.InspectionTypeSK = fi.InspectionTypeSK
GROUP BY
    dit.InspectionType
ORDER BY
    InspectionCount DESC;

```

Results

InspectionType	InspectionCount
Canvass	512099
Complaint	133040
License	105313
Canvass Re-Inspection	86890
Complaint Re-Inspection	34531
Short Form Complaint	24834
License Re-Inspection	23041
Suspected Food Poisoning	4951
Non-Inspection	3073
Tag Removal	2197
License-Task Force	1952
Consultation	1937
Recent Inspection	1563
Suspected Food Poisoning Re-inspection	697
Complaint-Fire	634
Task Force Liquor 1475	583
Short Form Fire-Complaint	280
Special Events (Festivals)	182
Complaint-Fire Re-inspection	143
Other	774

Thank You