

Mid-Term Project – Food Inspection for Cities Chicago & Dallas

Group 20 →

Chinmay Kulkarni

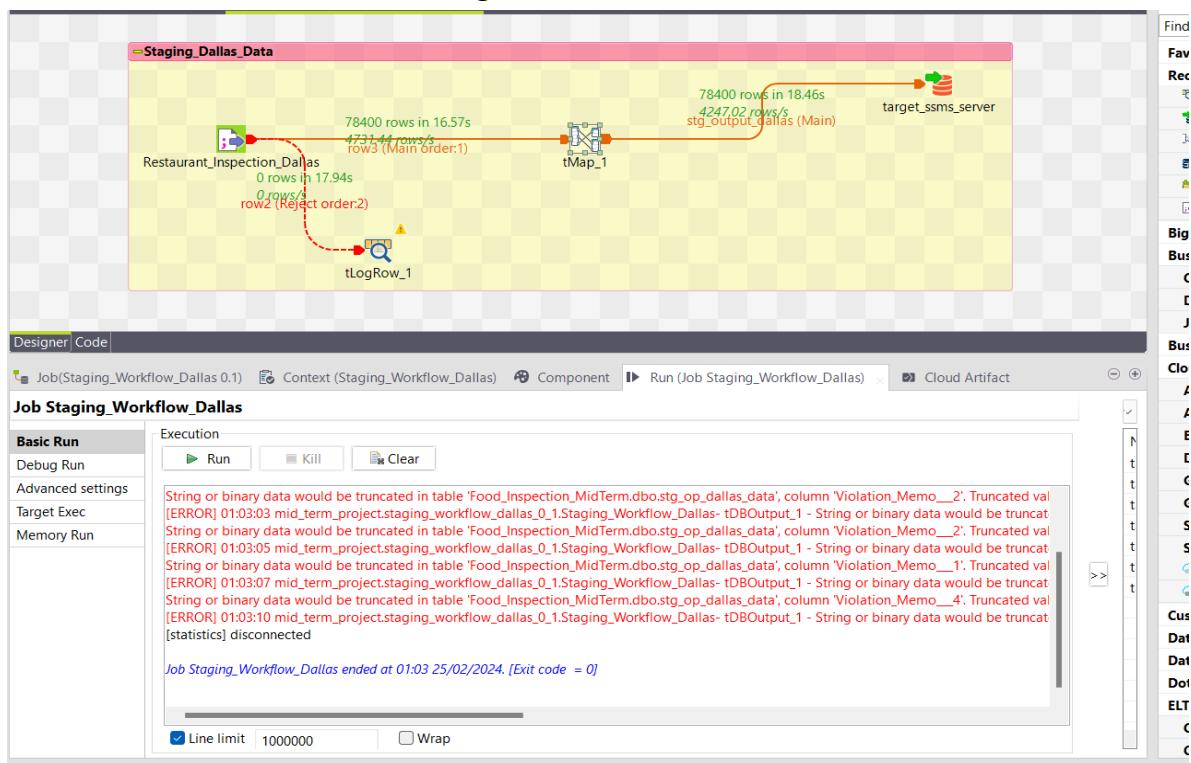
Pranitee Majukar

Rishikesh Chedrodath

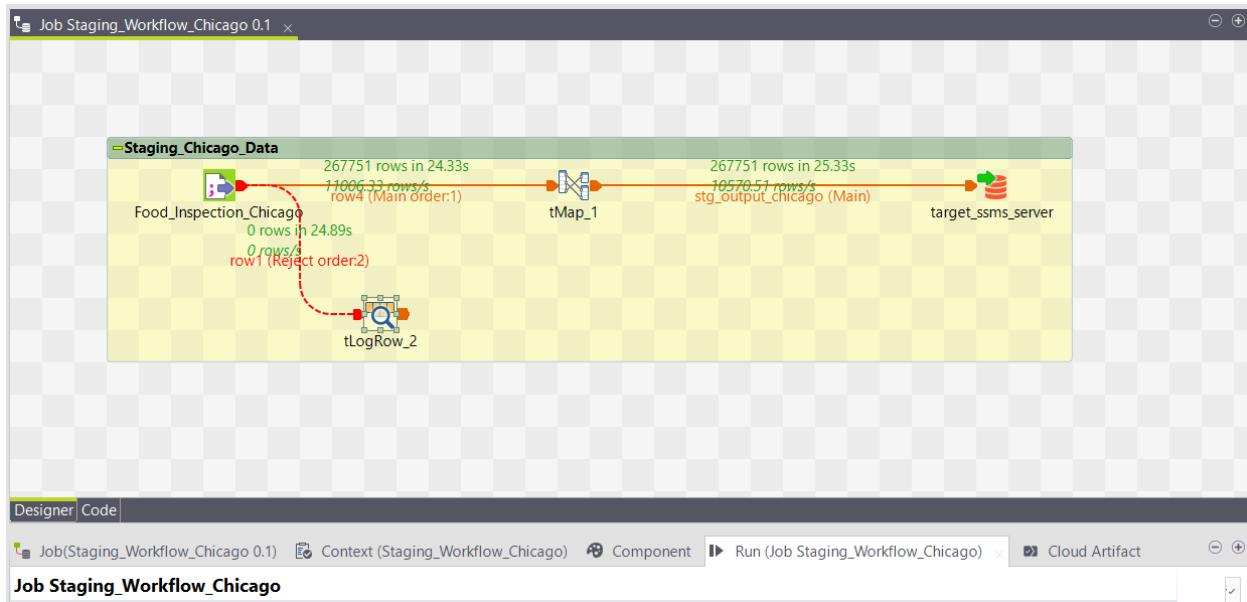
Soumya Nayak

Staging →

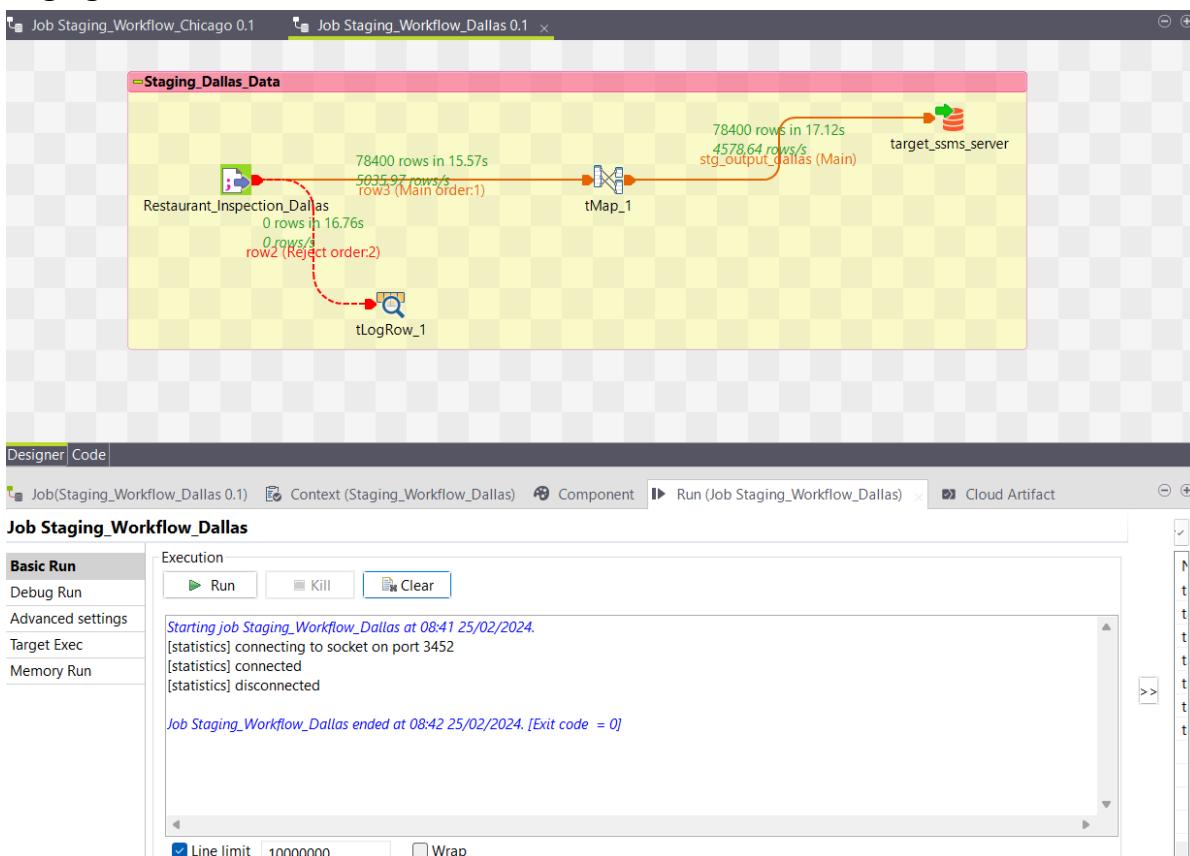
→ Reason Behind Increasing the Lengths of Violations, Memos & other attributes due to truncation error in Dallas and Chicago.



Staging Chicago Dataset in SSMS screenshot →



Staging Dallas Dataset in SSMS screenshot →



- String → DATE format conversion in staging.
→ `TalendDate.parseDate("MM/dd/yyyy", row4.Inspection_Date)`

Row count of stage table in SSMS →

```

USE Food_Inspection_MidTerm;
SELECT COUNT(*) FROM dbo.stg_op_dallas_data;
SELECT COUNT(*) FROM dbo.stg_op_chicago_data;

```

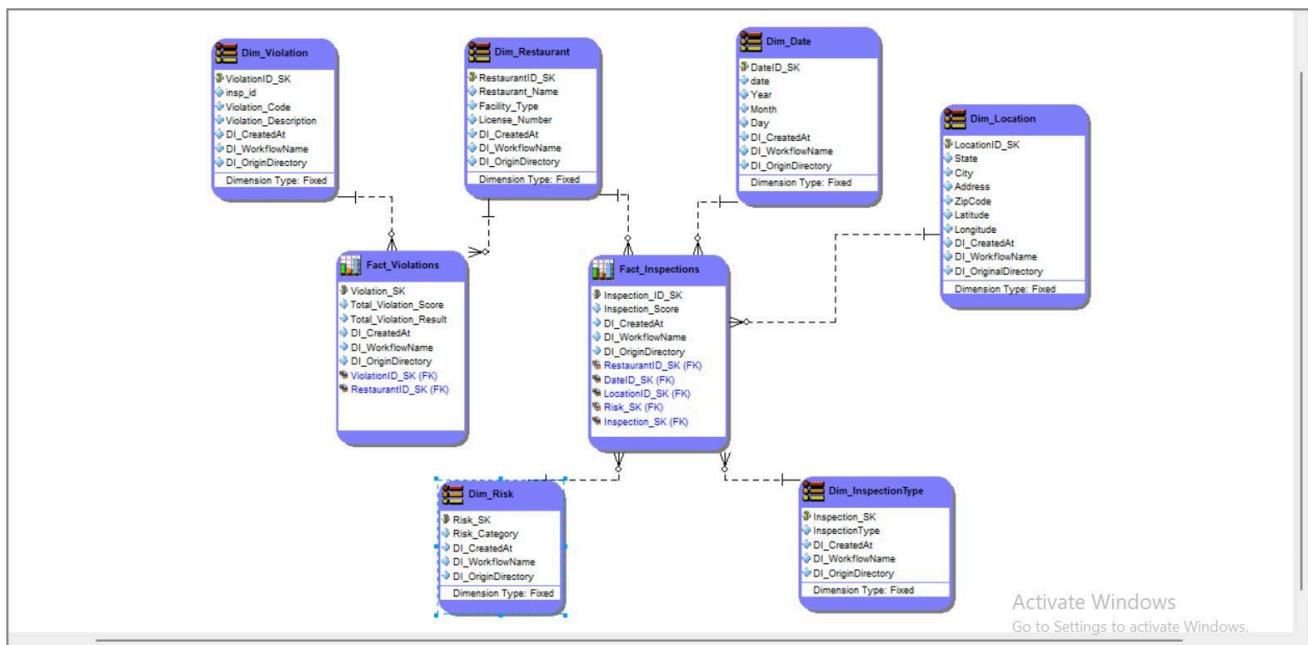
Results

(No column name)
1 78400

(No column name)
1 267034

After running the Talend job and verifying in SSMS, the row count is the same. That is for the Chicago dataset, the total row count is 267034 and for Dallas, it is 78400.

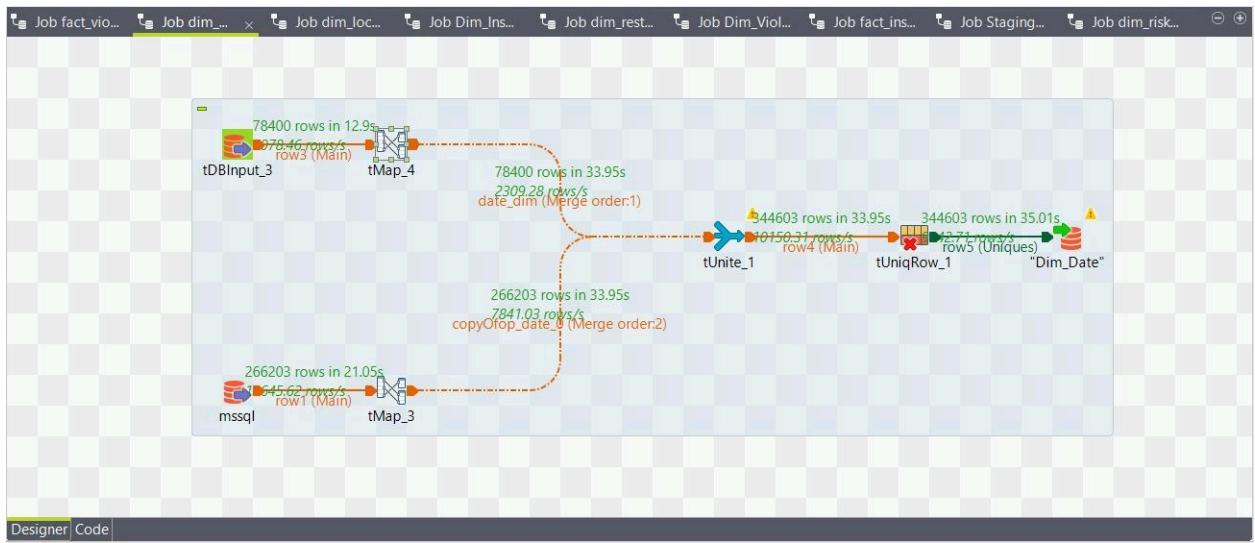
Dimensional Model →



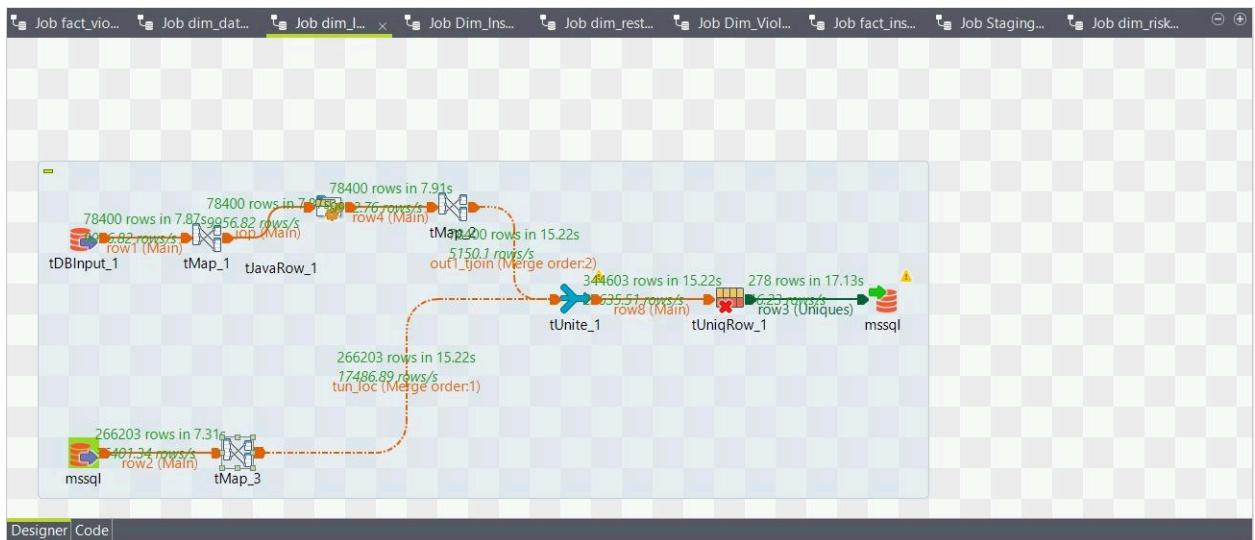
For our dimensional model, we have 2 facts named Fact_Inspections and Fact_Violations. Further, we have 6 dimensions named - Dim_Violation, Dim_Restaurant, Dim_Date, Dim_Location, Dim_InspectionType, and Dim_Risk.

Talend Workflows →

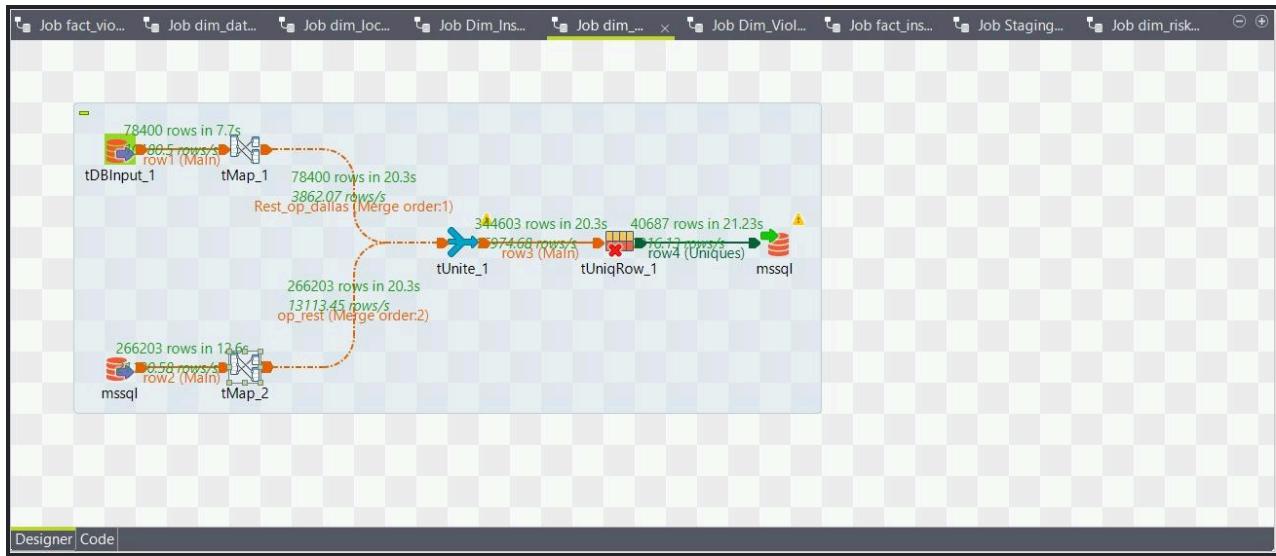
- Dim_Date –



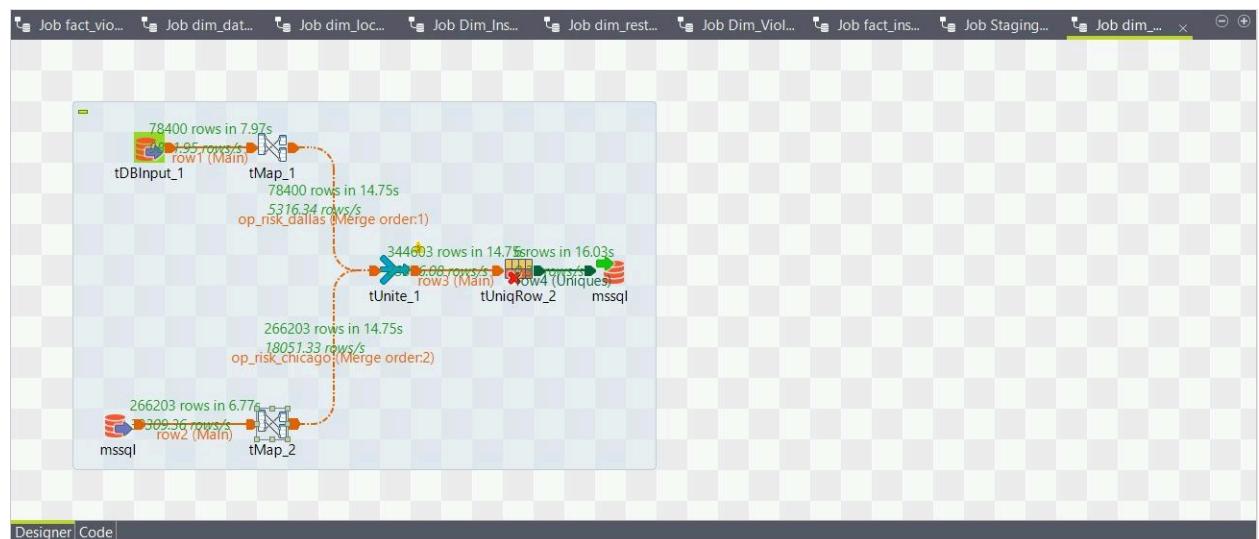
- *Dim_Location –*



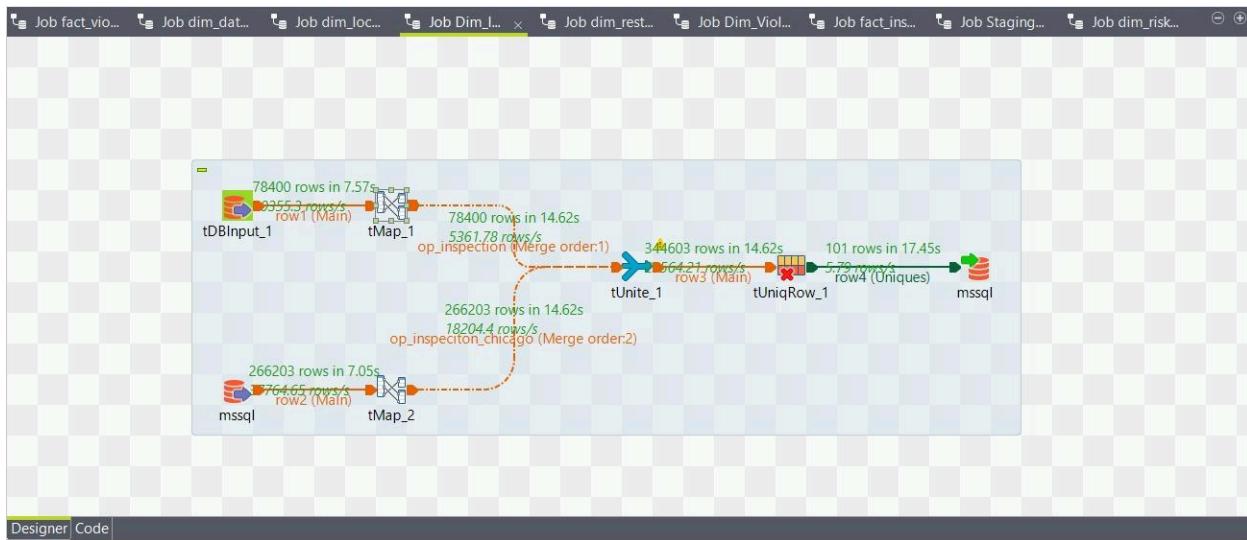
- *Dim_Restaurant –*



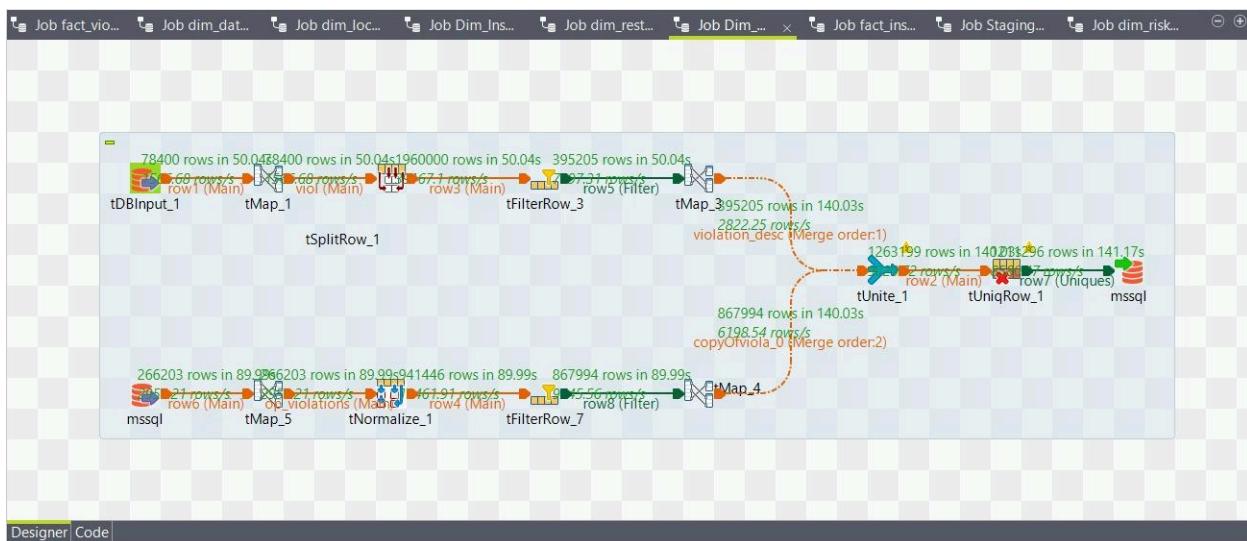
- *Dim_Risk –*



- *Dim_InspectionType –*



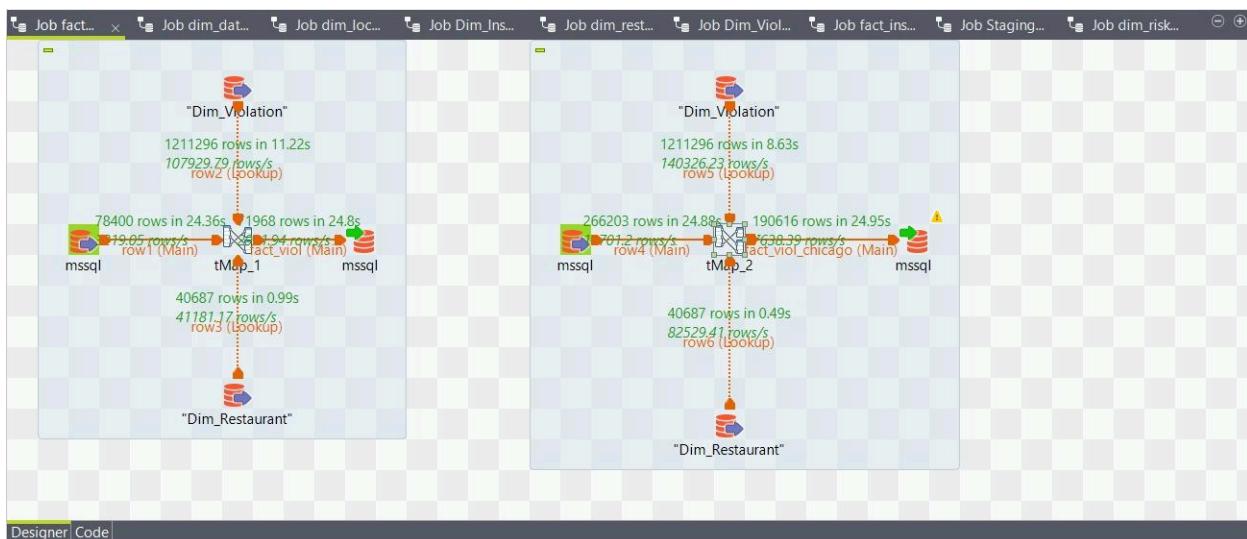
- *Dim_Violation –*



- *Fact_Inspections –*

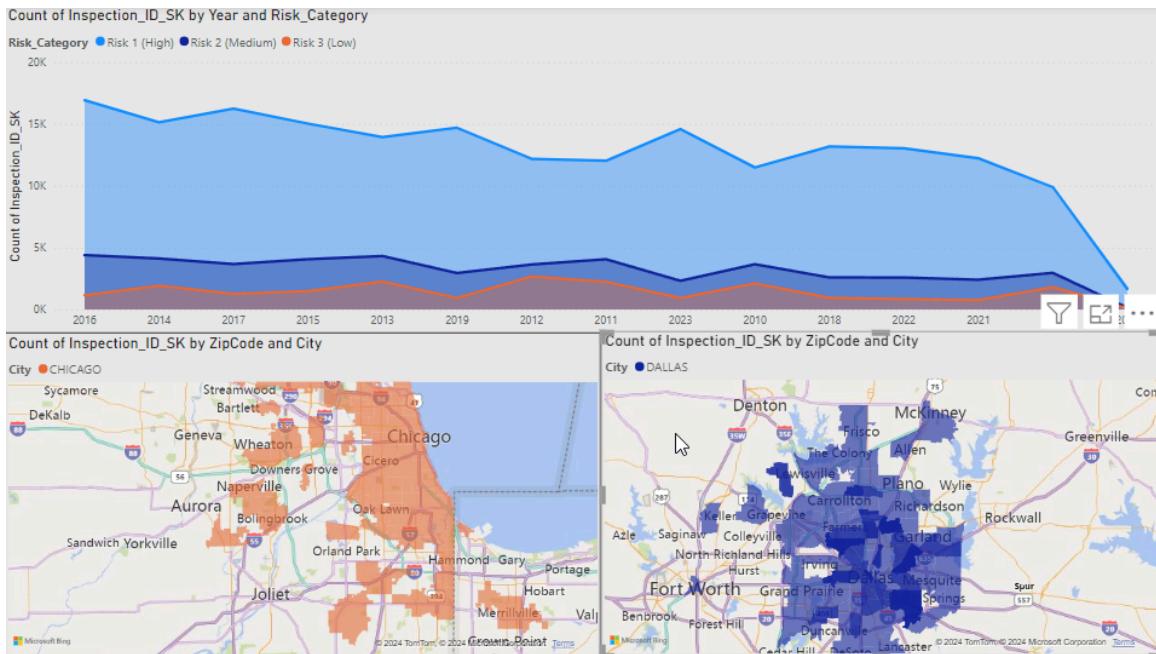
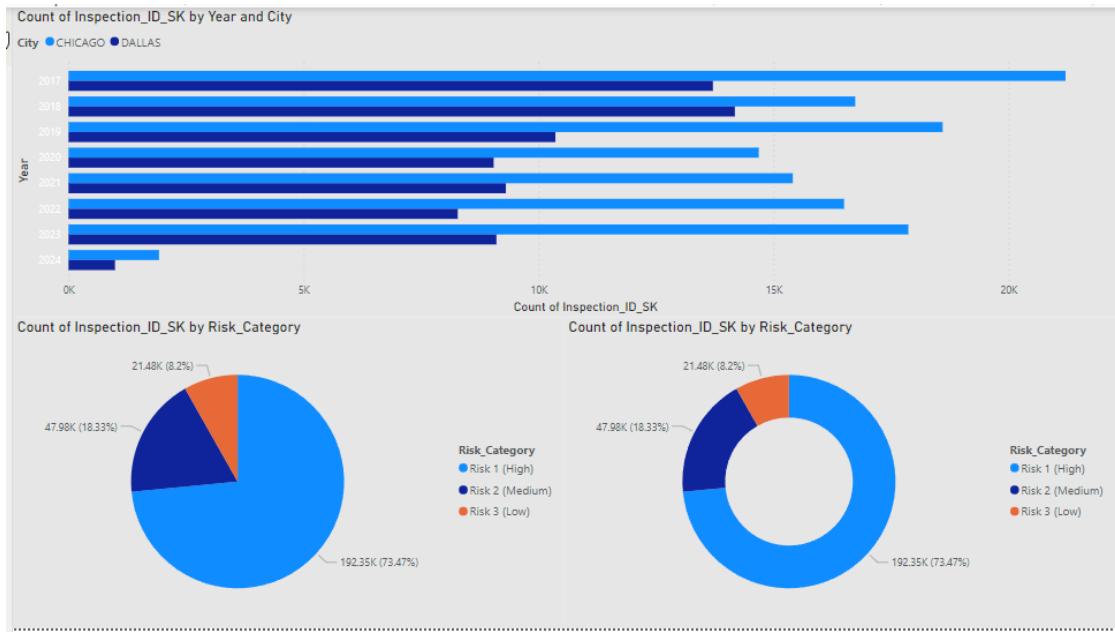


- *Fact_Violation –*



Visualizations →

PowerBI →



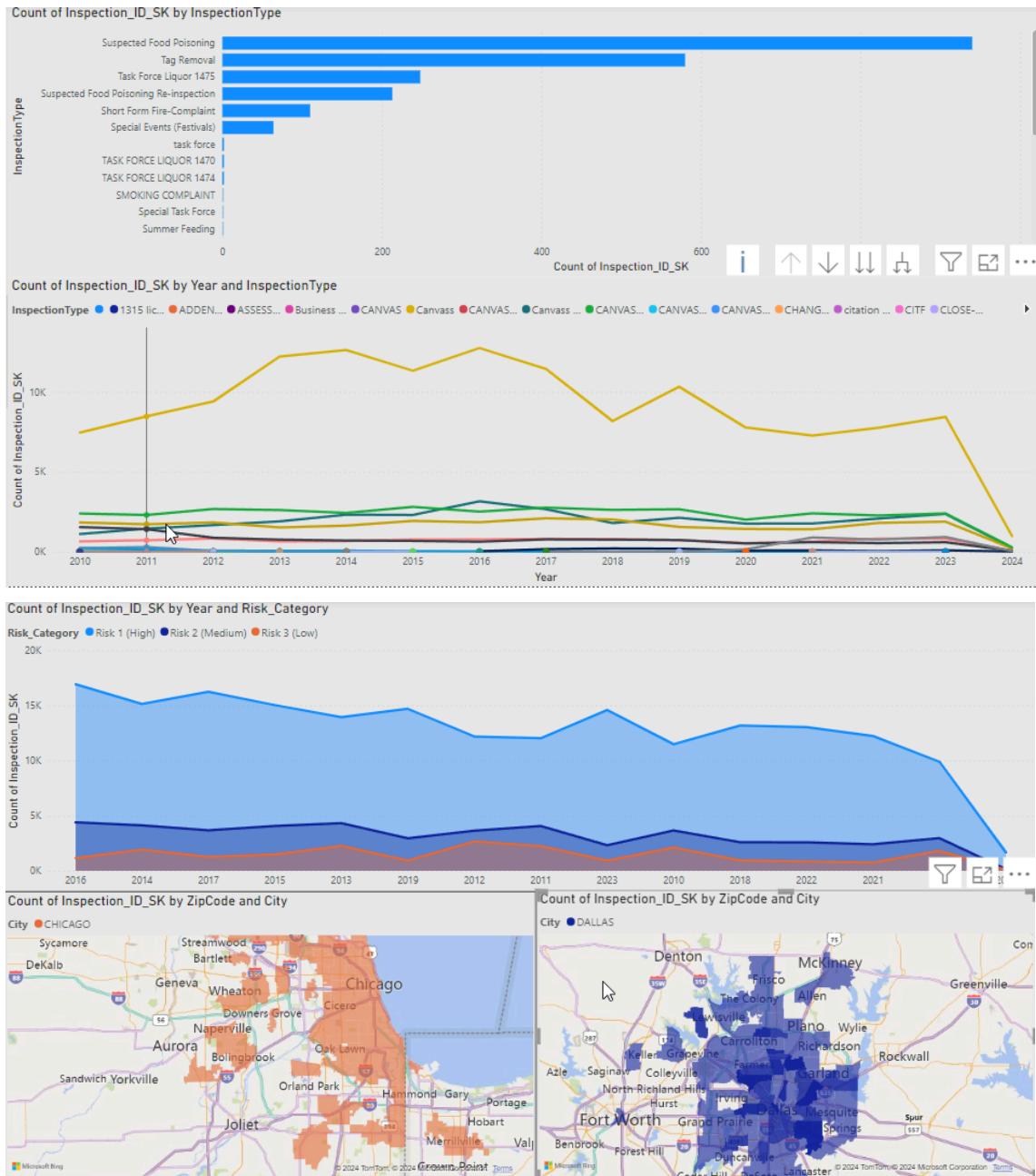
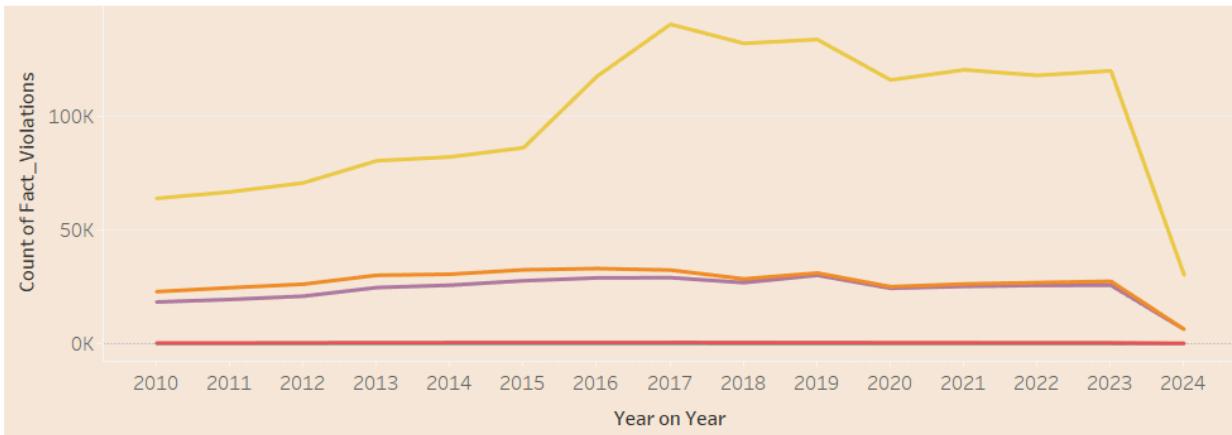
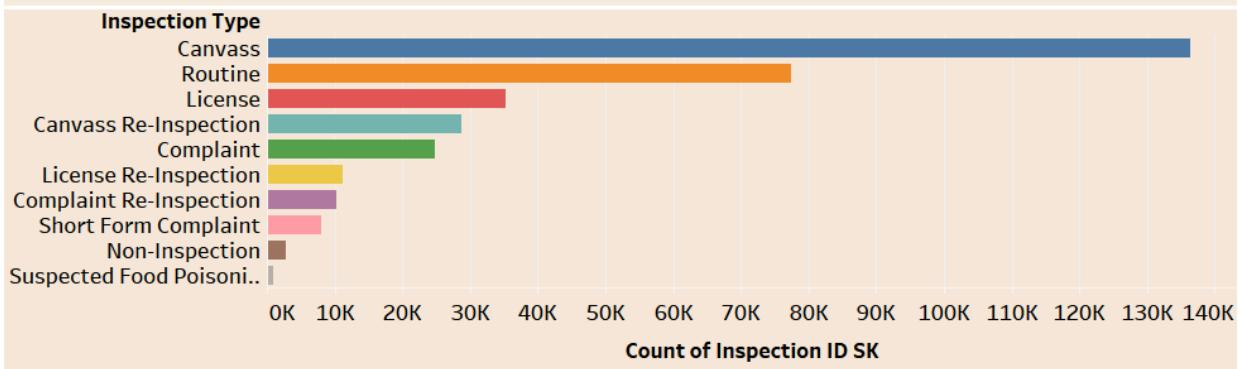


Tableau →

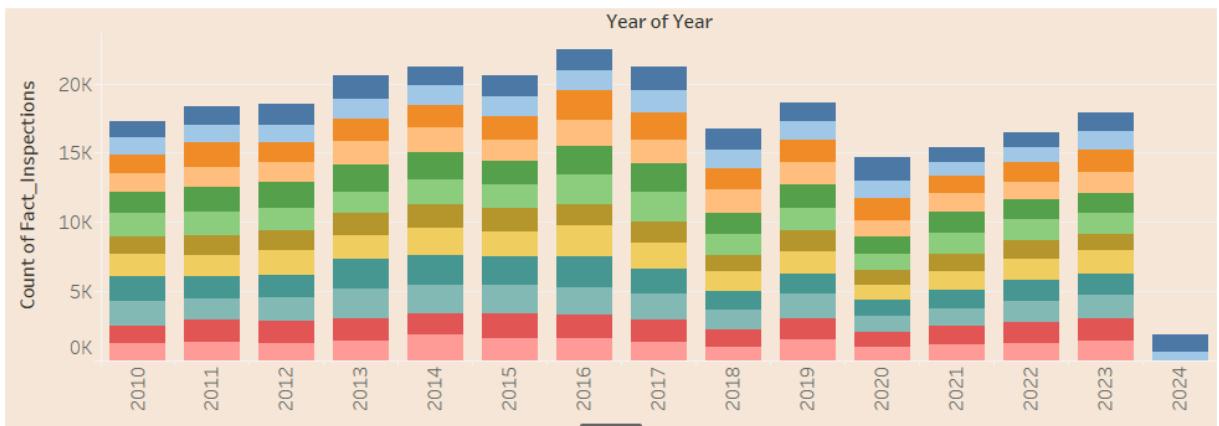
Yearly Violation Count and Results



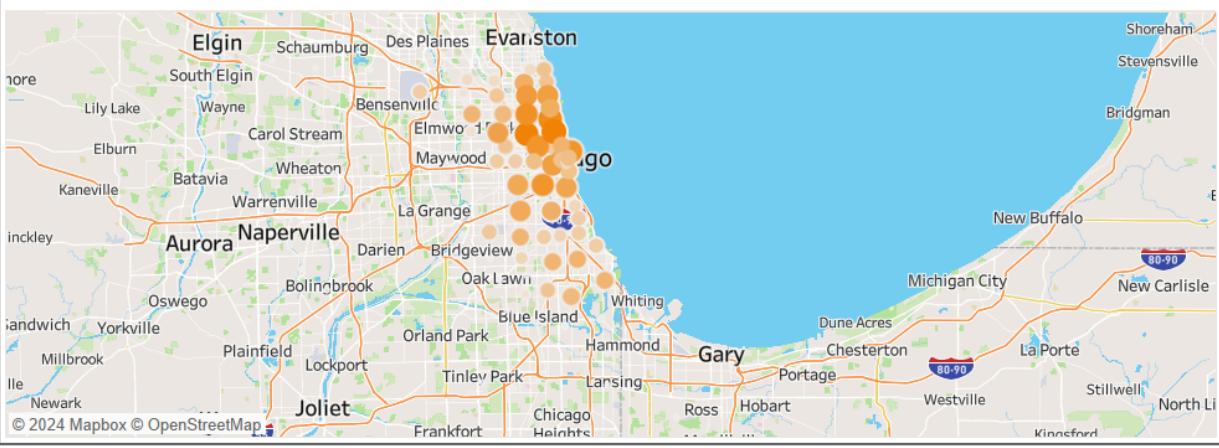
Distribution of Top 10 Inspection Types



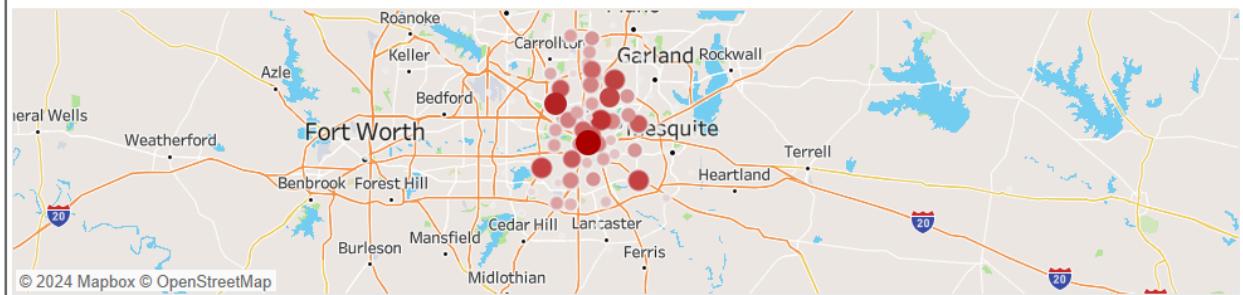
Inspection Count Done Over Time (Month) - Chicago



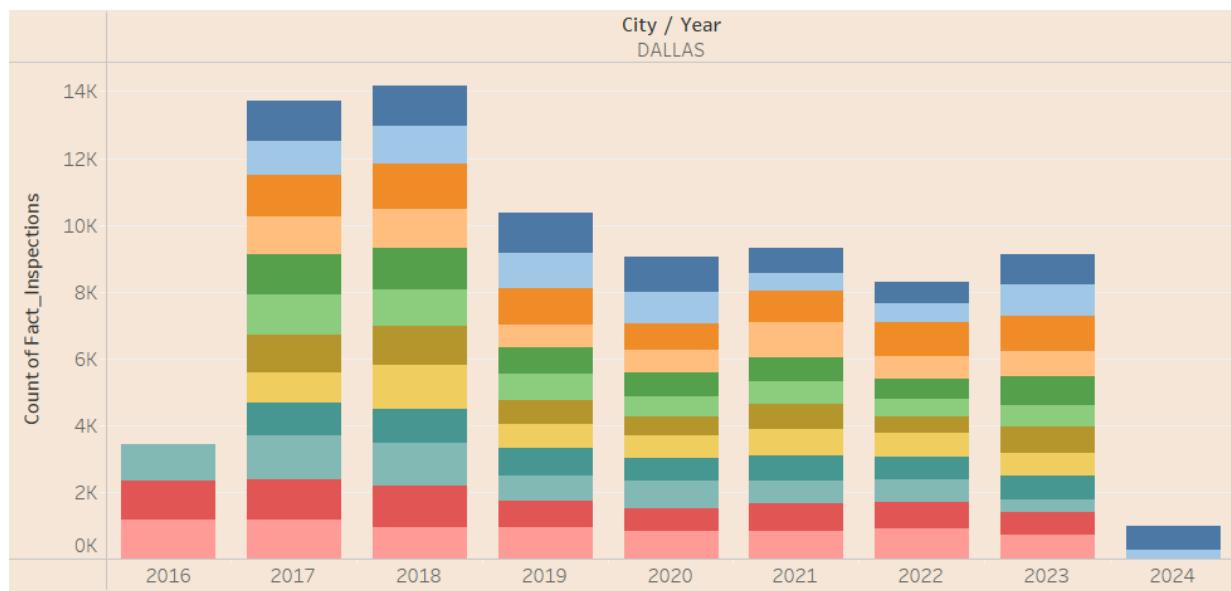
Count of Inspections for Chicago City



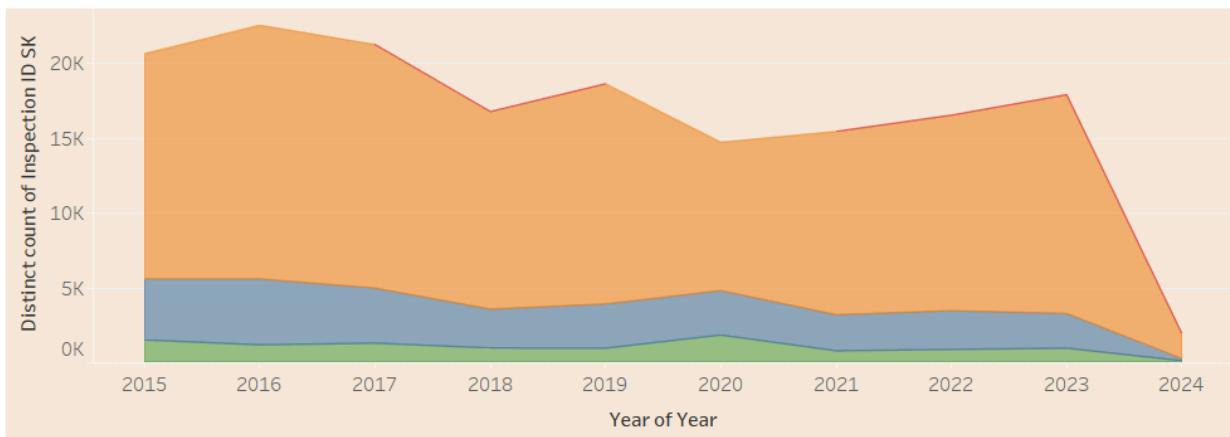
Count of Inspections for Dallas City



Inspection Count Done Over Time (Month) - Dallas



Risk Category for 10 Year based on Inspections Count



Top 10 Location with their Result

