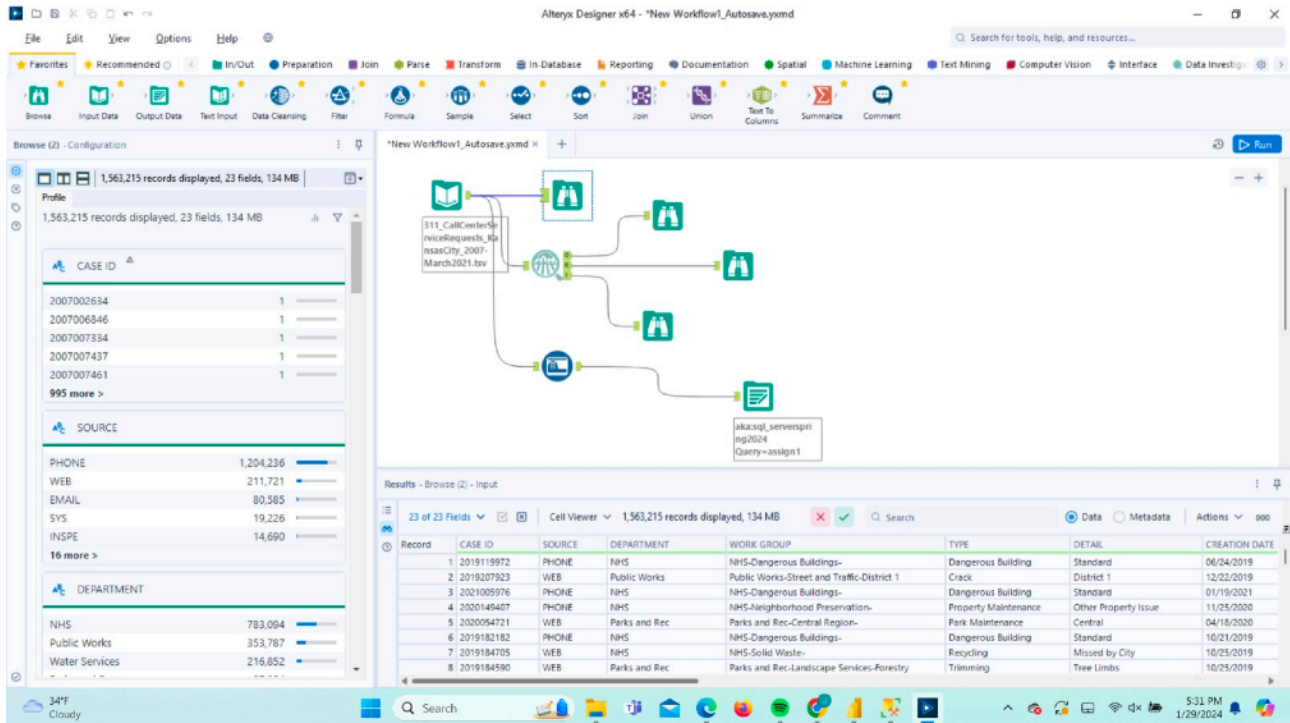


## DAMG 7370 Designing Advanced Data Architectures for Business Intelligence

## ALTERYX Data Profiling -



## Observations Made - Columns Containing Null Values

Column Name	Null Values
Closed Date	12,702
Street Address	24
Zip Code	826
Neighbourhood	46,106
County	66,956
Police District	32,265
Category 2	10,01,657
Category 3	14,09,943

## Field with wrongly assigned values -

Field List:

Field	Type	Size	Renamed	Description
CASE ID	Int32	4		
SOURCE	String	5		
DEPARTMENT	V_String	35		
WORK GROUP	V_String	50		
TYPE	V_String	48		
DETAIL	V_String	48		
CREATION DATE	String	10		
CREATION TIME	String	8		
STATUS	String	5		
EXCEEDED EST TIMEFRAME	String	1		
CLOSED DATE	String	10		
DAYS TO CLOSE	Double	8		
STREET ADDRESS	V_String	39		
ZIP CODE	Int32	4		
NEIGHBORHOOD	V_String	42		
COUNTY	V_String	13		
POLICE DISTRICT	V_String	11		
PARCEL ID NO	Int32	4		
LATITUDE	Double	8		
LONGITUDE	Double	8		
CATEGORY1	V_String	33		
CATEGORY2	V_String	9		
CATEGORY3	V_String	12		

Workflow Tools:

- 311\_CalCenterServiceRequests\_KansasCity\_2007-March2021.csv
- Convert CREATION DATE From: MM/dd/yyyy
- Convert CLOSED DATE From: MM/dd/yyyy
- DSN-task1 Query=dmg737

Results - Select (14) - Messages: 0 Errors, 0 Warnings, 0 Info, 0 Files.

Creation Date - String data type to date type

Creation Time- String to Time data type

Closed date- String to Date data type

ZIP code-Double to Integer type

Days to Close-to be changed to Int data type

Input Data (1) - Configuration

Connect a File or Database

C:\Users\jainam\Downloads\311\_CalCenterServiceRequests\_KansasCity\_2007-March2021.csv

Options

Name	Value
1 Record Limit	1
2 File Format	Comma Separated Value (*.csv)
3 Search SubDirs	Checked
4 Output File Name as Field	Not checked

Preview first 100 records

CASE ID	SOURCE	DEPARTMENT	WORK GROUP
1	2019119972	PHONE	NHS
2	2019207923	WEB	Public Works
3	2021005976	PHONE	NHS
4	2020149407	PHONE	NHS
5	2020054721	WEB	Parks and Rec
6	2019182182	PHONE	NHS
7	2019184705	WEB	NHS
8	2019184580	WEB	Parks and Rec
9	2020095175	PHONE	NHS
10	2019094486	PHONE	NHS
11	2020021142	PHONE	Water Services
12	2019200956	WEB	Public Works
13	2019208957	WEB	Public Works
14	2019208975	WEB	Public Works
15	2019208976	WEB	Public Works
16	2019208977	WEB	Public Works

Workflow Tools:

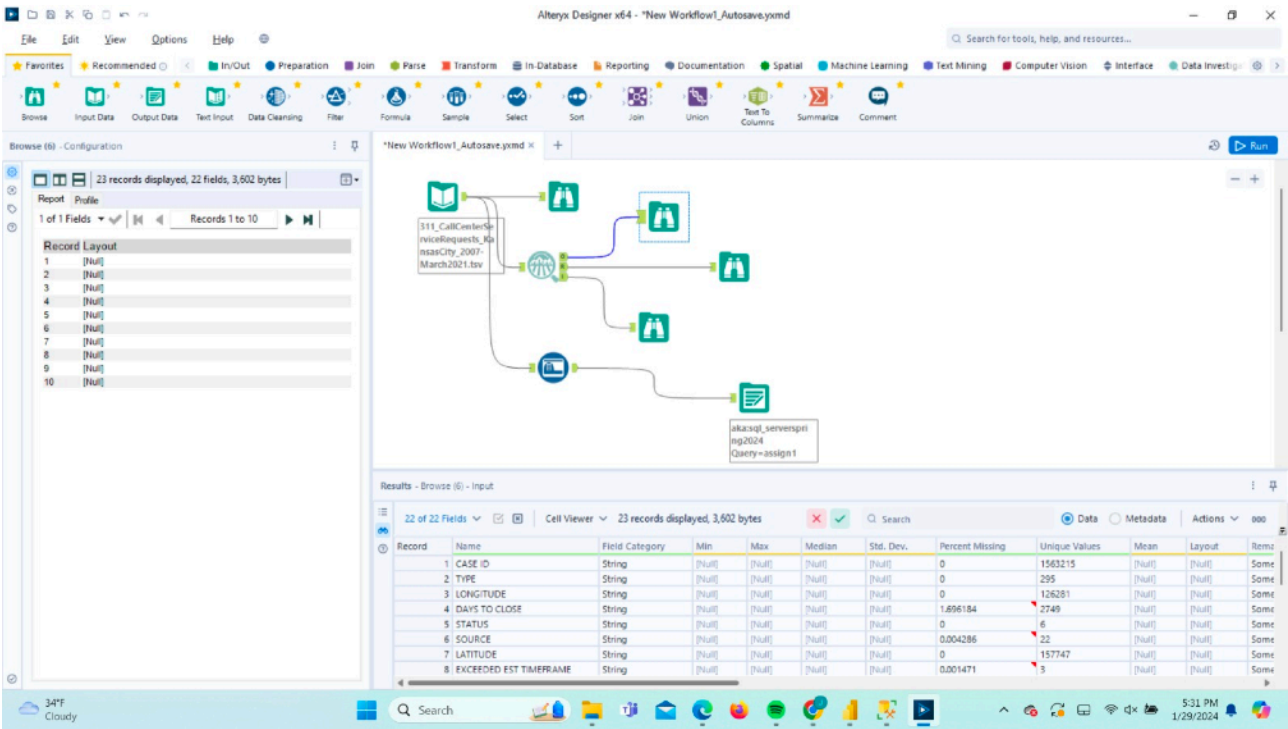
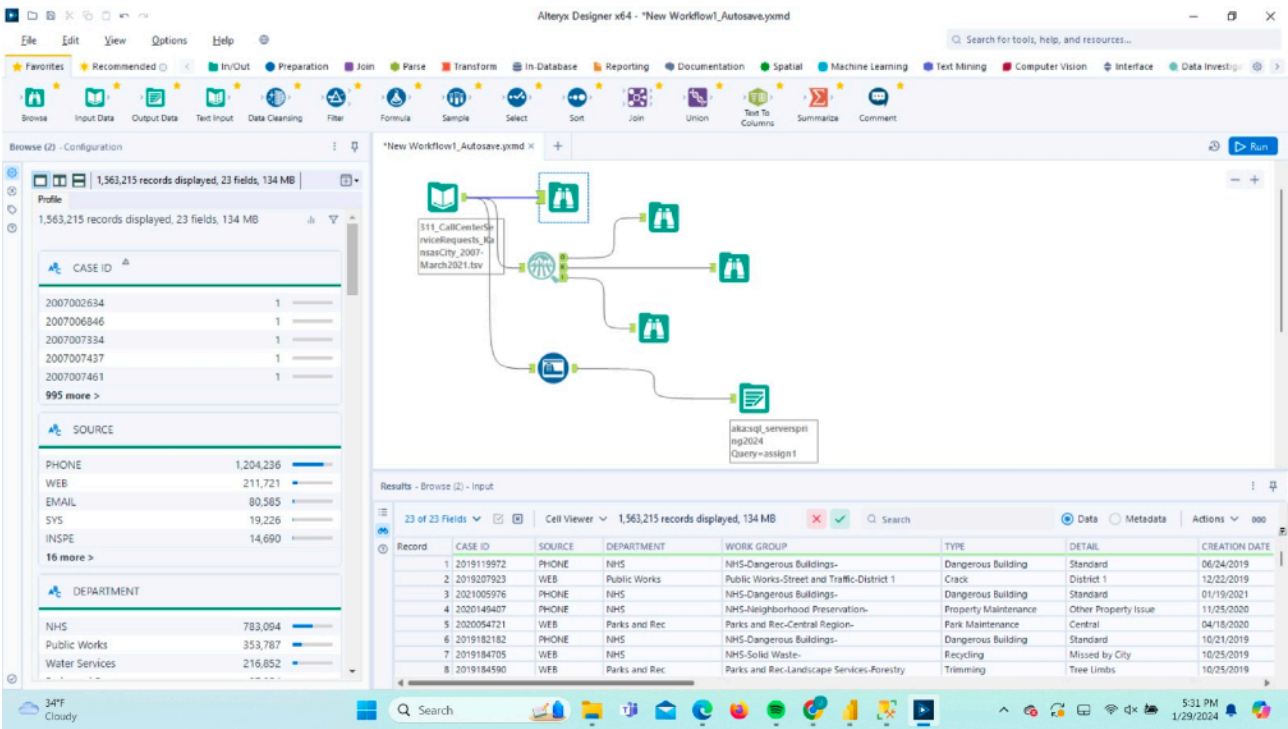
- 311\_CalCenterServiceRequests\_KansasCity\_2007-March2021.csv
- Convert CREATION DATE From: MM/dd/yyyy
- Convert CLOSED DATE From: MM/dd/yyyy
- DSN-task1 Query=dmg737

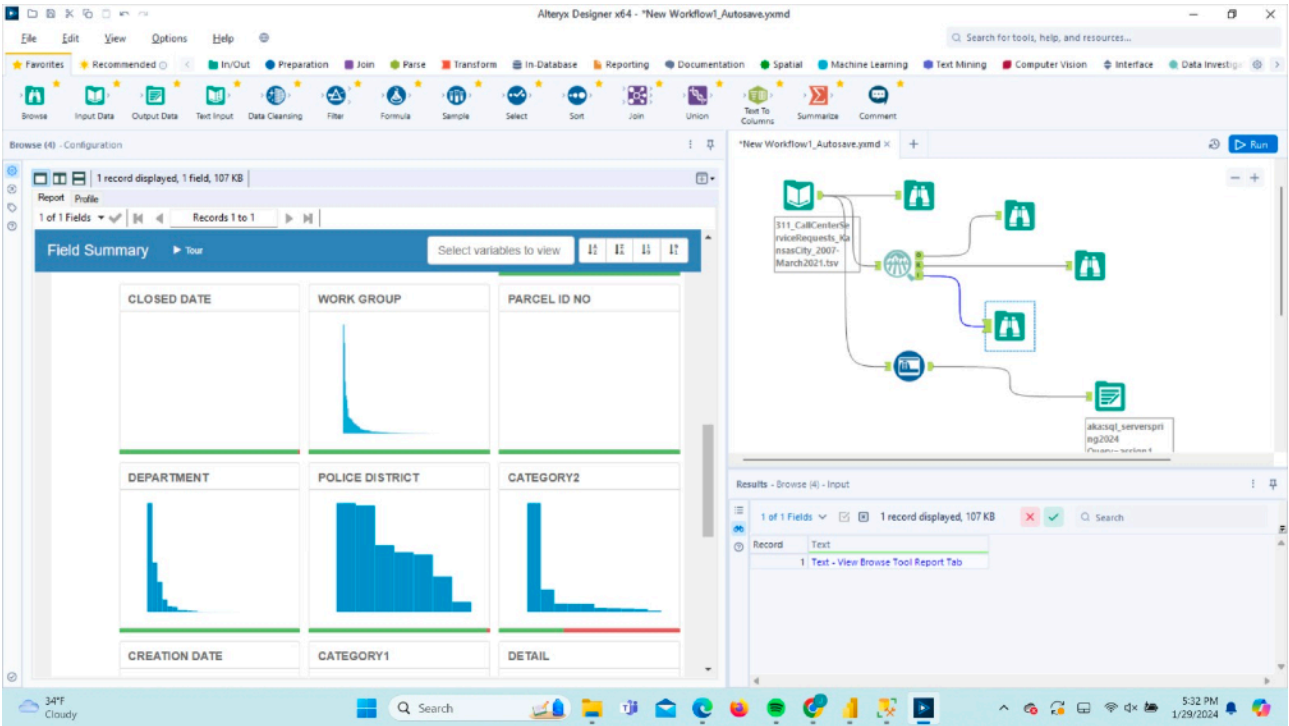
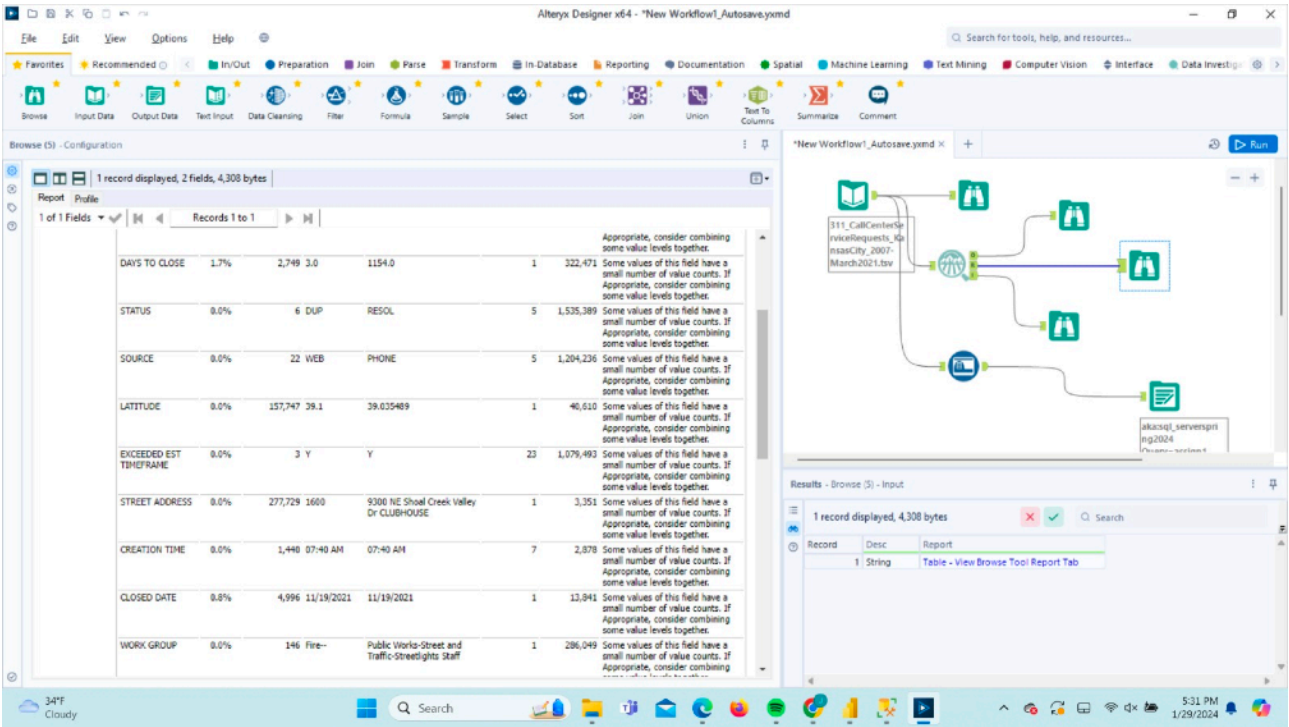
Results - Input Data (1) - Messages: 0 Errors, 0 Warnings, 0 Info, 1 Files.

Input Data (1) 2593215 records were read from "C:\Users\jainam\Downloads\311\_CalCenterServiceRequests\_KansasCity\_2007-March2021.csv".

Your device needs to restart to install updates. Select a time to restart.

Reading and Profiling data -







## Populating Data on SQL Server

The screenshot shows the Microsoft SQL Server Enterprise Manager interface. The left pane displays the 'Object Explorer' with the 'master' database selected. The right pane shows a query window with the following SQL query:

```
SELECT TOP (1000) [RecordID]
, [CASE_ID]
, [SOURCE]
, [DEPARTMENT]
, [WORK_GROUP]
, [TYPE]
, [DETAIL]
, [CREATION_TIME]
, [STATUS]
, [EXCEEDED_EST_TIMEFRAME]
, [DAYS_TO_CLOSE]
, [STREET_ADDRESS]
, [ZIP_CODE]
, [NEIGHBORHOOD]
, [CITY]
, [POLICE_DISTRICT]
, [PARCEL_ID_NO]
, [LATITUDE]
```

The query results are displayed in a table with the following columns: RecordID, CASE\_ID, SOURCE, DEPARTMENT, WORK\_GROUP, TYPE, DETAIL, CREATION\_TIME, STATUS, EXCEEDED\_EST\_TIMEFRAME, DAYS\_TO\_CLOSE, and ST. The table contains 15 rows of data, showing various service requests such as 'NHS Dangerous Buildings', 'NHS Dangerous Buildings', 'NHS Dangerous Buildings', 'NHS Dangerous Buildings', 'NHS Dangerous Buildings', 'NHS Dangerous Buildings', 'NHS Dangerous Buildings', 'NHS Dangerous Buildings', 'NHS Dangerous Buildings', 'NHS Dangerous Buildings', 'NHS Dangerous Buildings', 'NHS Dangerous Buildings', 'NHS Dangerous Buildings', 'NHS Dangerous Buildings', and 'NHS Dangerous Buildings'.

## SQL Queries-

### (1) Service Requests Over Time:

- What is the overall trend in Service Requests over the years 2018-2021?
- How have Service Requests changed on a monthly basis?

```
SELECT YEAR([CREATION TIME]) AS Year, COUNT(*) AS
TotalRequests
FROM dbo.damg7370
WHERE YEAR([CREATION TIME]) BETWEEN 2018 AND 2021
GROUP BY YEAR([CREATION TIME])
ORDER BY Year;
```

```
SELECT YEAR([CREATION TIME]) AS Year, MONTH([CREATION
TIME]) AS Month, COUNT(*) AS MonthlyRequests
FROM dbo.damg7370
WHERE YEAR([CREATION TIME]) BETWEEN 2018 AND 2021
GROUP BY YEAR([CREATION TIME]), MONTH([CREATION TIME])
ORDER BY Year, Month;
```

**(2) Volume of service requests received from different sources:**

- What is the overall trend in Service Requests over Sources?

```
SELECT [SOURCE], COUNT(*) AS TotalRequests
FROM dbo.damg7370
GROUP BY [SOURCE]
ORDER BY TotalRequests DESC;
```

**(3) Volume of service requests received by Department:**

- What is the overall trend in Service Requests received by Departments?

```
SELECT [DEPARTMENT], COUNT(*) AS TotalRequests
FROM dbo.damg7370
GROUP BY [DEPARTMENT]
ORDER BY TotalRequests DESC;
```

**(4) Top 10 Performance Metrics (Response Time) per CATEGORY and Type of Request:**

- What are the top 10 cases whose response time was fastest? Categorize it with Category1 and Type of Request.

```
SELECT TOP 10 [CATEGORY1], [TYPE], DATEDIFF(day,
[CREATION TIME], [closed_date]) AS ResponseTime
FROM dbo.damg7370
ORDER BY ResponseTime;
```

**(5) Geographical Visualization:**

- What are the Top 10 areas where most number of request were raised?

```
SELECT TOP 10 [NEIGHBORHOOD], COUNT(*) AS TotalRequests
```

```
FROM dbo.damg7370
```

```
GROUP BY [NEIGHBORHOOD]
```

```
ORDER BY TotalRequests DESC;
```

### **(6) Departmental Workload Comparison:**

- How does the workload vary among different departments and work groups? Create a visual representation to highlight the distribution.

```
SELECT [DEPARTMENT], [WORK GROUP], COUNT(*) AS  
TotalRequests
```

```
FROM dbo.damg7370
```

```
GROUP BY [DEPARTMENT], [WORK GROUP];
```

### **(7) Response Time Analysis:**

- Visualize the distribution of response times for each department. Are there any outliers or patterns in response times?

```
SELECT [DEPARTMENT], DATEDIFF(day, [CREATION TIME],  
[closed_date]) AS ResponseTime
```

```
FROM dbo.damg7370;
```

### **(8) Service Request Status Composition:**

- Create a visualization to show the composition of service request statuses (open, closed, in progress). How has this composition changed over the years 2018-2021?

```
SELECT [STATUS], YEAR([CREATION TIME]) AS Year, COUNT(*) AS  
TotalRequests
```

```
FROM dbo.damg7370
```

```
WHERE YEAR([CREATION TIME]) BETWEEN 2018 AND 2021
```

```
GROUP BY [STATUS], YEAR([CREATION TIME])
```

```
ORDER BY Year, [STATUS];
```

**(9) Time to Closure Analysis:**

- Visualize the average days to close service requests for each category1. Are there categories with consistently longer closure times?

```
SELECT [CATEGORY1], AVG(DATEDIFF(day, [CREATION TIME],  
[closed_date])) AS AvgDaysToClose
```

```
FROM dbo.damg7370
```

```
GROUP BY [CATEGORY1];
```

- Show top 10 (If you need help on how to restrict top 10 contact us and we can guide / help you)

```
SELECT TOP 10 [CATEGORY1], AVG(DATEDIFF(day, [CREATION  
TIME], [closed_date])) AS AvgDaysToClose
```

```
FROM dbo.damg7370
```

```
GROUP BY [CATEGORY1]
```

```
ORDER BY AvgDaysToClose DESC;
```

**(10) Workload Efficiency:**

- Create a visualization to show the relationship between workload (number of service requests) and efficiency (days to close) for each department?

```
SELECT [DEPARTMENT], COUNT(*) AS TotalRequests,  
AVG(DATEDIFF(day, [CREATION TIME], [closed_date])) AS  
AvgDaysToClose
```

```
FROM dbo.damg7370
```

```
GROUP BY [DEPARTMENT];
```