

End-to-End LA Crime Analytics (Medallion Architecture)

Alteryx Data Profiling –

1. Input & Select Tool

- Loaded the LA Crime dataset (2020–Present).
- Standardized data types: Date fields → Date, numeric fields → Int/Double, descriptive fields → String.

2. Data Cleansing

- Replaced nulls in numeric fields with 0 and in string fields with blanks.
- Removed leading/trailing whitespace.

3. Field Summary & Browse Tools

- Generated automated profiling for:
 - Numeric fields (min, max, unique counts)
 - Date fields (missingness, earliest/latest)
 - String fields (unique values, length)

4. Formula Tool – Data Quality Flags Added

Created four validation columns:

- Invalid_Date_Flag
- Invalid_Time_Flag
- Invalid_Geo_Flag
- Invalid_Age_Flag

These flags help identify missing or out of range values.

5. Summarize Tool

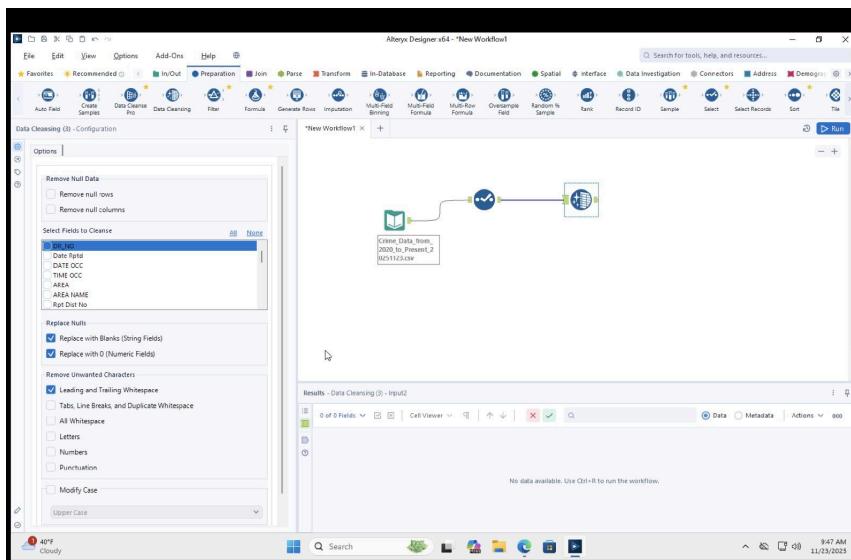
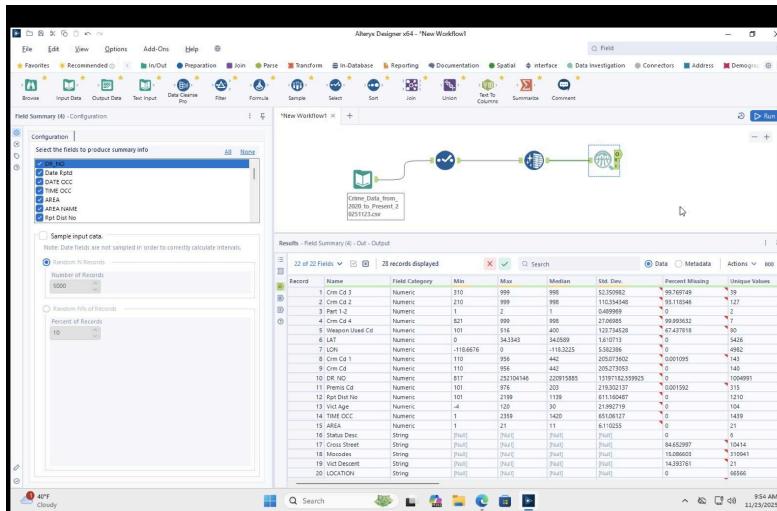
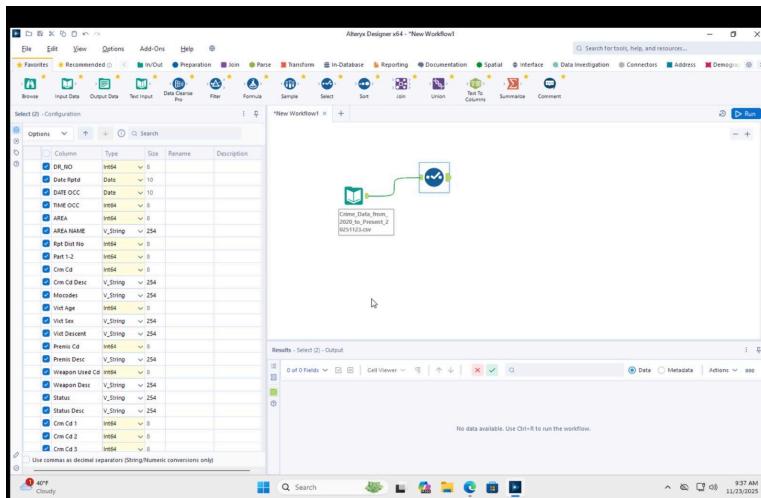
- Counted invalid occurrences for each flag
- All four flags returned counts equal to total records, indicating significant data quality issues.

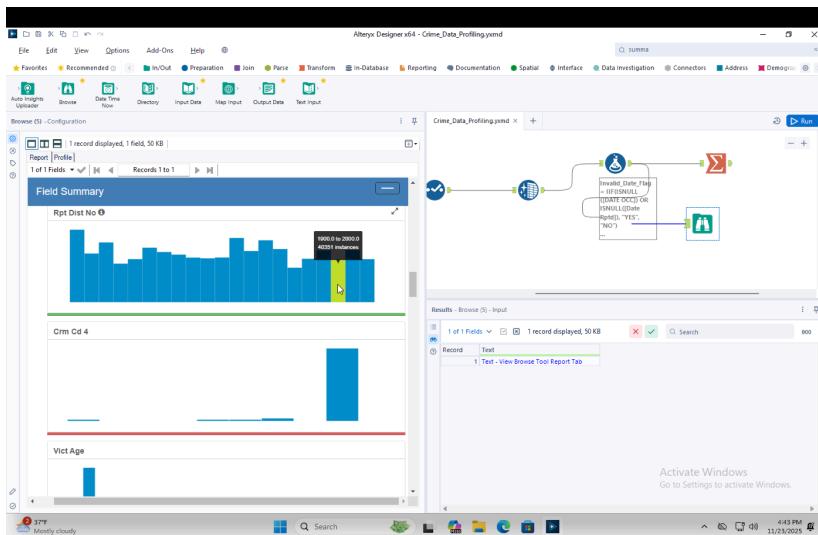
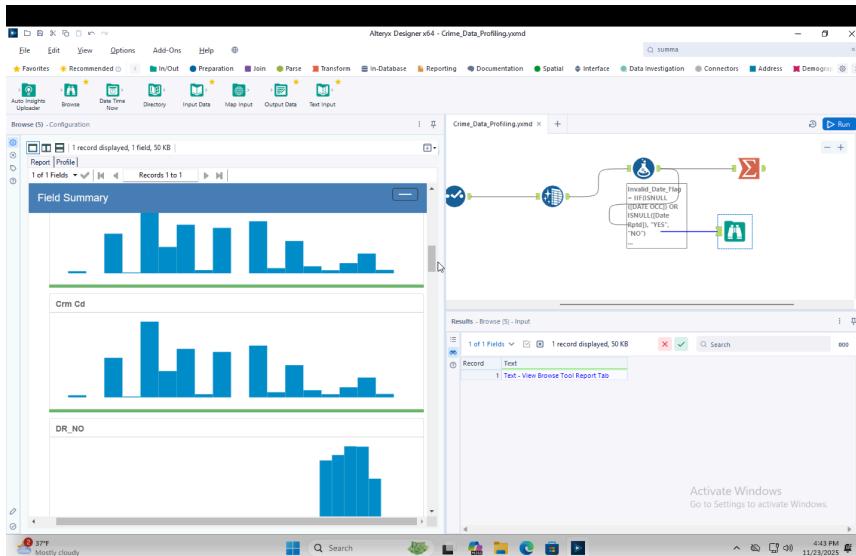
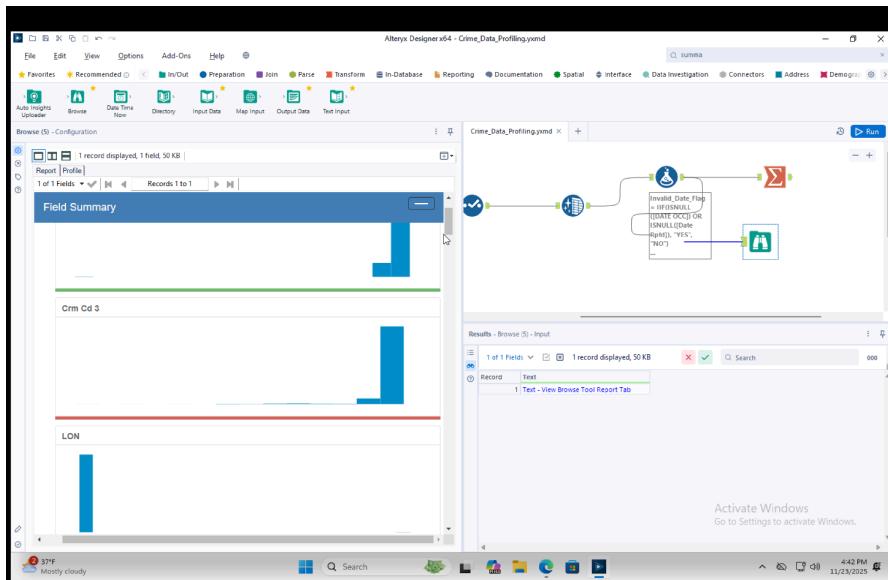
The screenshot shows the Alteryx Designer interface with the following components:

- Input Data (1) - Configuration:** A configuration panel for connecting to a file or database. It shows the connection to "C:\Users\Bhuvana\Downloads\Crime_Data_from_2020_to_Present_20251123.csv".
 - Options section with settings for Record Limit (1), File Format (Comma Separated Value (*.csv)), and Delimiters (,).
 - Preview (first 100 records) showing the crime data structure.
- Crime_Data_Profiling.ymd:** The main workflow diagram.
 - An input step from the configuration panel feeds into a "Summa" tool.
 - A formula tool contains the logic for generating the Invalid_Date_Flag:

```
Invalid_Date_Flag = IF(IIFNULL([DATE OCC]) OR IIFNULL([Date Rptd]), "YES", "NO")
```
 - The formula tool connects to a "Count" tool, which then connects to a summation symbol (Σ).
 - The workflow ends at a "Browse" tool.
- Results - Input Data (1) - Output:** A table showing the first 28 of 28 fields with 4,027 of 1,004,991 records displayed (partial results). The table includes columns: Record, DR_NO, Date Rptd, DATE OCC, TIME OCC, AREA, AREA NAME, and Rpt Dist No.

Record	DR_NO	Date Rptd	DATE OCC	TIME OCC	AREA	AREA NAME	Rpt Dist No
1	211507896	2021 Apr 11 12:00:00 AM	2020 Nov 07 12:00:00 AM	0845	15	N Hollywood	1502
2	201516622	2020 Oct 21 12:00:00 AM	2020 Oct 18 12:00:00 AM	1845	15	N Hollywood	1521
3	240913563	2024 Dec 10 12:00:00 AM	2020 Oct 30 12:00:00 AM	1240	09	Van Nuys	0933
4	210704711	2020 Dec 24 12:00:00 AM	2020 Dec 24 12:00:00 AM	1310	07	Wilsshire	0782
5	201418201	2020 Oct 03 12:00:00 AM	2020 Sep 29 12:00:00 AM	1830	14	Pacific	1454
6	240412063	2024 Dec 11 12:00:00 AM	2020 Nov 11 12:00:00 AM	1210	04	Hollenbeck	0429
7	240317069	2024 Dec 16 12:00:00 AM	2020 Apr 16 12:00:00 AM	1350	03	Southwest	0396
8	201115217	2020 Oct 29 12:00:00 AM	2020 Jul 07 12:00:00 AM	1400	11	Northeast	1133
9	241708596	2024 Apr 20 12:00:00 AM	2020 Mar 02 12:00:00 AM	1200	17	Devonshire	1729
10	242113813	2024 Dec 18 12:00:00 AM	2020 Sep 01 12:00:00 AM	0900	21	Topanga	2196
11	240605946	2024 Feb 06 12:00:00 AM	2020 Jun 20 12:00:00 AM	0001	06	Hollywood	0657
12	242014110	2024 Dec 18 12:00:00 AM	2020 Nov 17 12:00:00 AM	1320	20	Olympic	2023
13	202113531	2020 Sep 06 12:00:00 AM	2020 Sep 05 12:00:00 AM	1500	21	Topanga	2149
14	201710725	2020 Jul 03 12:00:00 AM	2020 Jul 02 12:00:00 AM	0500	17	Devonshire	1762
15	201406733	2020 Feb 16 12:00:00 AM	2020 Feb 13 12:00:00 AM	2300	14	Pacific	1406
16	201405970	2020 Feb 01 12:00:00 AM	2020 Feb 01 12:00:00 AM	1658	14	Pacific	1494
17	201820230	2020 Nov 08 12:00:00 AM	2020 Nov 08 12:00:00 AM	0730	18	Southeast	1844





Alteryx Designer x64 - Crime_Data_Profiling.youd

File Edit View Options Add-Ins Help

Favorites Recommended In/Out Preparation Joins Parse Transform In-Database Reporting Documentation Spatial Interface Data Investigation Connectors Address Demographic

Auto Insights Uploader Browse Save New Directory Input Data Map Input Output Data Test Input

Formula Bar - Configuration Output Column Data Preview Crime_Data_Profiling.youd + Run

1 $\text{IF}(\text{ISNULL}([\text{TIME OCC}]) \text{ OR } \text{ISNULL}([\text{Date Rptd}]), \text{"YES"}, \text{"NO"})$

2 $\text{IF}(\text{ISNULL}([\text{TIME OCC}]) \text{ OR } \text{ISNULL}([\text{Date Rptd}]), \text{"YES"}, \text{"NO"})$

3 $\text{IF}([\text{LAT}] < \text{LOW} \text{ OR } [\text{LAT}] > \text{HIGH}, \text{"YES"}, \text{"NO"})$

4 $\text{IF}([\text{LONG}] < \text{LOW} \text{ OR } [\text{LONG}] > \text{HIGH}, \text{"YES"}, \text{"NO"})$

5 $\text{IF}([\text{Vict Age}] < 0 \text{ OR } [\text{Vict Age}] > 120, \text{"YES"}, \text{"NO"})$

6 $\text{IF}([\text{Vict Age}] < 0 \text{ OR } [\text{Vict Age}] > 120, \text{"YES"}, \text{"NO"})$

7 $\text{IF}([\text{Vict Age}] < 0 \text{ OR } [\text{Vict Age}] > 120, \text{"YES"}, \text{"NO"})$

8 $\text{IF}([\text{Vict Age}] < 0 \text{ OR } [\text{Vict Age}] > 120, \text{"YES"}, \text{"NO"})$

Crime Data Team, Import to Present [0051123.csv]

Results - Formula [6] - Output

32 of 32 Fields \rightarrow 4,941 of 1,004,991 records displayed(partial results)

Record	DR_NO	Date Rptd	TIME OCC	AREA	NAME	Rpt DR No	Part 1-2	Crm Cd	Category	
1	21107096	[Null]	[Null]	845	N Hollywood	1502	2	354	THEFT OF	
2	20151622	[Null]	[Null]	1645	N Hollywood	1521	1	230	CRIM ALLEGED	
3	20151623	[Null]	[Null]	1345	9	Wiltshire	253	254	THEFT OF	
4	21070411	[Null]	[Null]	1316	7	Wiltshire	782	1	331	THEFT OF
5	20141624	[Null]	[Null]	1820	14	Pacific	1454	1	420	CRIM ALLEGED
6	20141625	[Null]	[Null]	1250	12	9	254	254	THEFT OF	
7	20131599	[Null]	[Null]	1350	3	Southwest	396	2	334	THEFT OF
8	20111527	[Null]	[Null]	1300	11	Northwest	1133	2	812	CRIM ALLEGED
9	20111528	[Null]	[Null]	1729	2	2	354	354	THEFT OF	
10	24115113	[Null]	[Null]	500	21	Los Angeles	2156	2	254	THEFT OF
11	24060526	[Null]	[Null]	1	6	Hollywood	657	2	812	CRIM ALLEGED
12	24204110	[Null]	[Null]	1320	20	Olympic	2033	2	354	THEFT OF
13	24204111	[Null]	[Null]	1650	21	Los Angeles	2149	2	254	THEFT OF
14	20171025	[Null]	[Null]	500	17	Downtown	1762	1	354	CRIM ALLEGED
15	20164711	[Null]	[Null]	7100	14	Parcde	1406	1	110	BURGLAR

Mostly cloudy 11/23/2025 4:37 PM

The screenshot shows the Alteryx Designer interface with the title "Alteryx Designer x64 - Crime_Data_Profiling.ymd". The top menu bar includes File, Edit, View, Options, Add-Ins, Help, and various tool icons. The ribbon tabs are Preparation, Join, Parse, Transform, In-Database, Reporting, Documentation, Spatial, Interface, Data Investigation, Connectors, Address, and Demographic.

The main workspace displays a workflow titled "Crime_Data_Profiling.ymd". The workflow starts with a "Text Input" tool reading from a CSV file named "Crime_Data_from_2020_to_Present_2 (251123.csv)". This is followed by a "Date Time Now" tool, which then connects to a "Join" tool. The "Join" tool has two inputs: one from a "Text Input" tool containing the field "DR_NO" and another from a "Map Input" tool containing the field "Rpt Dist No". The output of the "Join" tool is connected to a "Transform" tool. Inside the "Transform" tool, there is a script block with the following code:

```
Invalid_Date_Flag = IF(LEN(DR_NO) < 10 OR ISNULL(DR_NO) OR ISNULL(Date_Rptd), "YES", "NO")
```

The output of the "Transform" tool is then connected to a "Summarize" tool, which is set up to sum the "Count" field for four new fields: "Invalid_Date_Flag", "Invalid_Time_Flag", "Invalid_Geo_Flag", and "Invalid_Age_Flag". The "Actions" panel on the left shows the configuration for these summarized fields. The "Results" panel at the bottom shows the output of the "Summarize" tool, which consists of a single record with four columns: Count, Count2, Count3, and Count4, all having a value of 1004991.

At the bottom right, there is a watermark: "Activate Windows Go to Settings to activate Windows." The system tray shows the date as 11/23/2025 and the time as 4:38 PM.