

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 displays the Alteryx Designer x64 interface for a workflow named 'Crime_Data_Profiling.yxmd'. The workflow consists of the following tools:

- Input Data (1) - Configuration:** Set up to connect to a CSV file at 'C:\Users\Bhuvana\Downloads\Crime_Data_from_2020_to_Present_20251123.csv'. Options include Record Limit, File Format (Comma Separated Value (.csv)), Search SubDirs, Output File Name as Field (No), and Delimiters.
- Select Tool:** Used to filter or transform fields.
- Formula Tool:** Configured to create four validation flags based on the following logic:
 - Invalid_Date_Flag = IF(ISNULL([DATE OCC]) OR ISNULL([Date Rptd]), "YES", "NO")
 - Invalid_Time_Flag = IF(ISNULL([TIME OCC]) OR ISNULL([Time Rptd]), "YES", "NO")
 - Invalid_Geo_Flag = IF(ISNULL([AREA]) OR ISNULL([Area Name]), "YES", "NO")
 - Invalid_Age_Flag = IF(ISNULL([Age]), "YES", "NO")
- Summarize Tool:** Used to aggregate the data, likely to count the occurrences of each flag.

The results pane shows a preview of the first 100 records of the crime data. The data includes fields such as DR_NO, Date Rptd, DATE OCC, TIME OCC, AREA, AREA NAME, and Rpt Dist No.

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
2	201516622	2020 Oct 21 12:00:00 AM	2020 Oct 18 12:00:00 AM	1845	15	N Hollywood
3	240913563	2024 Dec 10 12:00:00 AM	2020 Oct 30 12:00:00 AM	1240	09	Van Nuys
4	210704711	2020 Dec 24 12:00:00 AM	2020 Dec 24 12:00:00 AM	1310	07	Wilshire
5	201418201	2020 Oct 03 12:00:00 AM	2020 Sep 29 12:00:00 AM	1830	14	Pacific
6	240412063	2024 Dec 11 12:00:00 AM	2020 Nov 11 12:00:00 AM	1210	04	Hollenbeck
7	240317069	2024 Dec 16 12:00:00 AM	2020 Apr 16 12:00:00 AM	1350	03	Southwest
8	201115217	2020 Apr 29 12:00:00 AM	2020 Jul 07 12:00:00 AM	1400	11	Northeast
9	241708596	2024 Apr 20 12:00:00 AM	2020 Mar 02 12:00:00 AM	1200	17	Devonshire
10	242113813	2024 Dec 18 12:00:00 AM	2020 Sep 01 12:00:00 AM	0900	21	Topanga
11	240605946	2024 Feb 06 12:00:00 AM	2020 Jun 20 12:00:00 AM	0001	06	Hollywood
12	242014110	2024 Dec 18 12:00:00 AM	2020 Nov 17 12:00:00 AM	1320	20	Olympic
13	202113531	2020 Sep 06 12:00:00 AM	2020 Sep 05 12:00:00 AM	1500	21	Topanga
14	201710725	2020 Jul 03 12:00:00 AM	2020 Feb 01 12:00:00 AM	0500	17	Devonshire
15	201406733	2020 Feb 16 12:00:00 AM	2020 Feb 13 12:00:00 AM	2300	14	Pacific
16	201406970	2020 Feb 01 12:00:00 AM	2020 Feb 01 12:00:00 AM	1658	14	Pacific
17	201820230	2020 Nov 08 12:00:00 AM	2020 Nov 08 12:00:00 AM	0730	18	Southeast

Alteryx Designer v6.4 - New Workflow

File Edit View Options Add-Ons Help

Input Data Output Data Test Input Data Connect Pro Filter Formula Join Append Transform In-Database Reporting Documentation Spatial Interface Data Investigation Connectors Address Designer

Select (3) - Configuration

Options

Column Type Size Rename Description

Column	Type	Size	Rename	Description
DR_NO	Int64	8		
Date Field	Date	10		
DATE OCC	Date	10		
TIME OCC	Int64	8		
AREA	Int64	8		
AREA NAME	V_String	254		
Rpt Dist No	Int64	8		
Part 1-2	Int64	8		
Com Cat	Int64	8		
Com Cat Desc	V_String	254		
Miscodes	V_String	254		
Viol Age	Int64	8		
Viol Sex	V_String	254		
Viol Descrit	V_String	254		
Premis Cat	Int64	8		
Premis Desc	V_String	254		
Weapon Used Cat	Int64	8		
Weapon Desc	V_String	254		
Status	V_String	254		
Status Desc	V_String	254		
Com Cat 1	Int64	8		
Com Cat 2	Int64	8		
Com Cat 3	Int64	8		

Use commas as decimal separators (String/numeric conversions only)

Results - Select (3) - Output

0 of 0 Fields

No data available. Use Ctrl+R to run the workflow.

Alteryx Designer v6.4 - New Workflow

File Edit View Options Add-Ons Help

Input Data Output Data Test Input Data Connect Pro Filter Formula Join Append Transform In-Database Reporting Documentation Spatial Interface Data Investigation Connectors Address Designer

Field Summary (6) - Configuration

Configuration

Search the fields to produce summary info

Number of Records

Percent of Records

Results - Field Summary (6) - Output

22 of 22 Fields

Field	Name	Field Category	Min	Max	Median	Std. Dev.	Percent Missing	Unique Values
1	Com Cat 3	Numeric	210	999	999	52.209862	99.107168	2
2	Com Cat 2	Numeric	210	999	999	110.254540	99.110346	127
3	Part 1-2	Numeric	1	2	1	0.409889	0	2
4	Com Cat 4	Numeric	821	999	999	27.003855	99.993032	7
5	Weapon Used Cat	Numeric	101	516	400	123.734229	67.437918	80
6	LAT	Numeric	0	34.2345	34.0209	1.610713	0	5405
7	LOCN	Numeric	-116.6676	0	-116.3225	5.062386	0	4962
8	Com Cat 1	Numeric	110	999	442	205.073602	0.0071095	743
9	Com Cat	Numeric	110	999	442	205.273503	0	140
10	DR NO	Numeric	817	252104148	220915885	11997162.559923	0	1008891
11	Premis Cat	Numeric	101	516	202	219.202127	0.0011902	210
12	Rpt Dist No	Numeric	101	2108	1108	611.60407	0	1210
13	Viol Age	Numeric	-4	120	30	21.982719	0	104
14	TIME OCC	Numeric	1	2108	1405	633.68127	0	1489
15	AREA	Numeric	1	21	11	6.110335	0	21
16	Status Desc	String	[Null]	[Null]	[Null]	[Null]	0	9
17	Cross Street	String	[Null]	[Null]	[Null]	[Null]	84.652997	10414
18	Miscodes	String	[Null]	[Null]	[Null]	[Null]	13.886025	110941
19	Viol Descrit	String	[Null]	[Null]	[Null]	[Null]	14.303781	51
20	LOCATION	String	[Null]	[Null]	[Null]	[Null]	0	68566

Alteryx Designer v6.4 - New Workflow

File Edit View Options Add-Ons Help

Input Data Output Data Test Input Data Connect Pro Filter Formula Join Append Transform In-Database Reporting Documentation Spatial Interface Data Investigation Connectors Address Designer

Data Cleansing (3) - Configuration

Options

Remove Null Data

Remove null rows

Remove null columns

Select Fields to Cleanse

Replace Nulls

Replace with Blanks (String Fields)

Replace with 0 (Numeric Fields)

Remove Unwanted Characters

Leading and Trailing Whitespace

Tab, Line Breaks, and Duplicate Whitespace

All Whitespace

Letters

Numbers

Punctuation

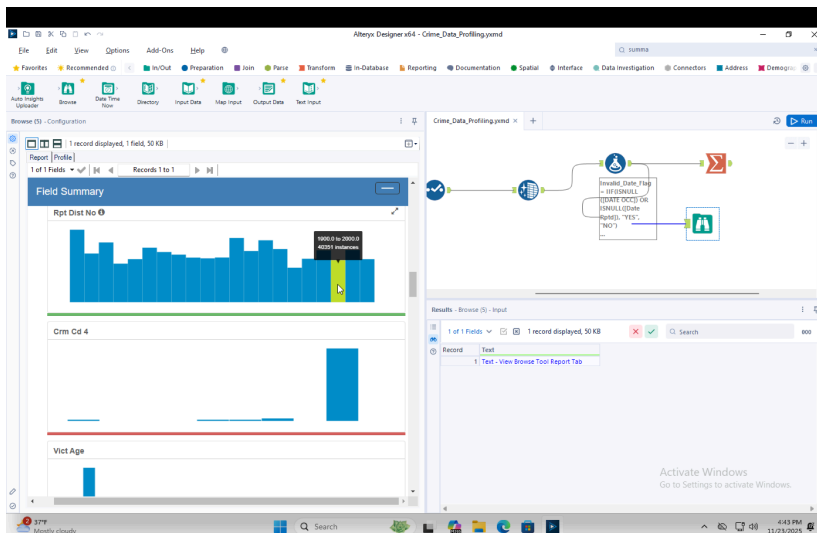
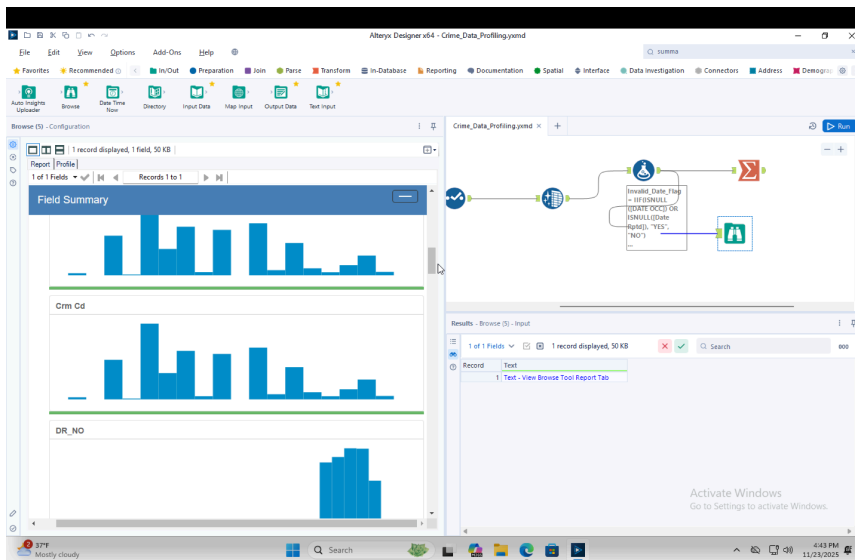
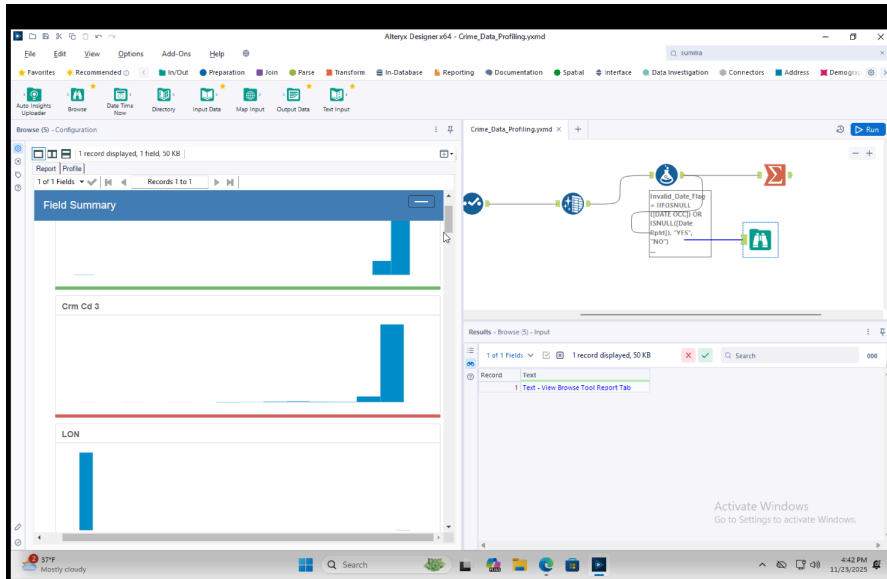
Modify Case

Upper Case

Results - Data Cleansing (3) - Input2

0 of 0 Fields

No data available. Use Ctrl+R to run the workflow.



Alteryx Designer x64 - Crime_Data_Profiling.ymd

Formula (6) - Configuration

Output Column: Data Preview

Invalid_Date_Flag: YES
IF([ISNULL([DATE OCC]) OR ISNULL([Date Rptd])], "YES", "NO")

Invalid_Time_Flag: NO
IF([TIME OCC] < 0 OR [TIME OCC] > 2359, "YES", "NO")

Invalid_Geo_Flag: NO
IF([LAT] = 0 OR [LON] = 0 OR [LAT] > 39 OR [LAT] < -35 OR [LON] > 127 OR [LON] < -129, "YES", "NO")

Invalid_Age_Flag: NO
IF([Vict Age] < 0 OR [Vict Age] > 120, "YES", "NO")

Crime_Data_Profiling.ymd

Results - Formula (6) - Output

Record	CR_NO	Date Rptd	DATE OCC	TIME OCC	AREA	AREA NAME	Rpt Dist No	Part 1-2	Crn Cd	Crn Cd Desc
1	211507896	[Null]	[Null]	845	15	N Hollywood	1502	2	354	THEFT OF
2	201916632	[Null]	[Null]	1945	15	N Hollywood	1521	1	230	ASSAULT
3	240913983	[Null]	[Null]	1340	9	Van Nuys	933	2	354	THEFT OF
4	210704711	[Null]	[Null]	1310	7	Whittier	782	1	331	THEFT FR
5	201416201	[Null]	[Null]	1830	14	Pacific	1454	1	433	THEFT OF
6	240412083	[Null]	[Null]	1210	4	Hollywood	429	2	354	THEFT OF
7	240317089	[Null]	[Null]	1330	3	Southwest	386	2	354	THEFT OF
8	201116217	[Null]	[Null]	1400	11	Northwest	1113	2	812	CRN AGA
9	241708536	[Null]	[Null]	1200	17	Downshire	1729	2	354	THEFT OF
10	242113813	[Null]	[Null]	900	21	Topanga	2156	2	354	THEFT OF
11	240605846	[Null]	[Null]	1	6	Hollywood	657	2	812	CRN AGA
12	242014110	[Null]	[Null]	1320	20	Olympic	2029	Activate Windows	354	THEFT OF
13	202113531	[Null]	[Null]	1500	21	Topanga	2149	Go to Settings to activate Windows	354	THEFT OF
14	201702023	[Null]	[Null]	500	17	Downshire	1782	1	230	BURGLES
15	301406711	[Null]	[Null]	7300	14	Pasadena	1416	1	110	RUBGLAR

Alteryx Designer x64 - Crime_Data_Profiling.ymd

Fields:

Field	Type
CR_NO	Int64
Date Rptd	Date
DATE OCC	Date
TIME OCC	Int64
AREA	Int64
AREA NAME	V_String
Rpt Dist No	Int64
Part 1-2	Int64
Crn Cd	Int64
Crn Cd Desc	V_String
Mercedes	V_String

Actions:

Field	Action	Output Field Name
Invalid_Date_Flag	Count	Count1
Invalid_Time_Flag	Count	Count2
Invalid_Geo_Flag	Count	Count3
Invalid_Age_Flag	Count	Count4

Crime_Data_Profiling.ymd

Results - Summarize (7) - Output

Record	Count1	Count2	Count3	Count4
1	1004991	1004991	1004991	1004991