

LAPD Dataset Documentation

Github link: https://github.com/prathushp/End-to-End-LA-Crime-Analytics_Group4

LAPD dataset Profiling:

Overall Information

Basic Information:

Results - Browse (2) - Input												
Record	DR_NO	Date Rptd	DATE OCC	TIME OCC	AREA	AREA NAME	Rpt Dist No	Part 1-2	Crm Cd	Crm Cd Desc	Actions	000
1	211507896	2021 Apr 11 12:00:00 AM	2020 Nov 07 12:00:00 AM	0845	15	N Hollywood	1502	2	354	THEFT OF IDENTITY		
2	201516622	2020 Oct 21 12:00:00 AM	2020 Oct 18 12:00:00 AM	1845	15	N Hollywood	1521	1	230	ASSAULT WITH DEADLY WEAPON,		
3	240913563	2024 Dec 10 12:00:00 AM	2020 Oct 30 12:00:00 AM	1240	09	Van Nuys	0933	2	354	THEFT OF IDENTITY		
4	210704711	2020 Dec 24 12:00:00 AM	2020 Dec 24 12:00:00 AM	1310	07	Wilshire	0782	1	331	THEFT FROM MOTOR VEHICLE - G		
5	201418201	2020 Oct 03 12:00:00 AM	2020 Sep 29 12:00:00 AM	1830	14	Pacific	1454	1	420	THEFT FROM MOTOR VEHICLE - PI		
6	240412063	2024 Dec 11 12:00:00 AM	2020 Nov 11 12:00:00 AM	1210	04	Hollenbeck	0429	2	354	THEFT OF IDENTITY		
7	240317069	2024 Dec 16 12:00:00 AM	2020 Apr 16 12:00:00 AM	1350	03	Southwest	0396	2	354	THEFT OF IDENTITY		
8	201115217	2020 Oct 29 12:00:00 AM	2020 Jul 07 12:00:00 AM	1400	11	Northeast	1133	2	812	CRM AGNST CHLD (13 OR UNDER)		
9	241708596	2024 Apr 20 12:00:00 AM	2020 Mar 02 12:00:00 AM	1200	17	Devonshire	1729	2	354	THEFT OF IDENTITY		
10	242113813	2024 Dec 18 12:00:00 AM	2020 Sep 01 12:00:00 AM	0900	21	Topanga	2196	2	354	THEFT OF IDENTITY		
11	240605846	2024 Feb 06 12:00:00 AM	2020 Jun 20 12:00:00 AM	0001	06	Hollywood	0657	2	812	CRM AGNST CHLD (13 OR UNDER)		
12	242014110	2024 Dec 18 12:00:00 AM	2020 Nov 17 12:00:00 AM	1320	20	Olympic	2023	2	354	THEFT OF IDENTITY		
13	202113531	2020 Sep 06 12:00:00 AM	2020 Sep 05 12:00:00 AM	1500	21	Topanga	2149	1	510	VEHICLE - STOLEN		
14	201710725	2020 Jul 03 12:00:00 AM	2020 Jul 02 12:00:00 AM	0500	17	Devonshire	1762	1	310	BURGLARY		
15	201406733	2020 Feb 16 12:00:00 AM	2020 Feb 13 12:00:00 AM	2300	14	Pacific	1406	1	330	BURGLARY FROM VEHICLE		

Total Number of Records: 1004991

Number of Columns :28

Duplicate records:0

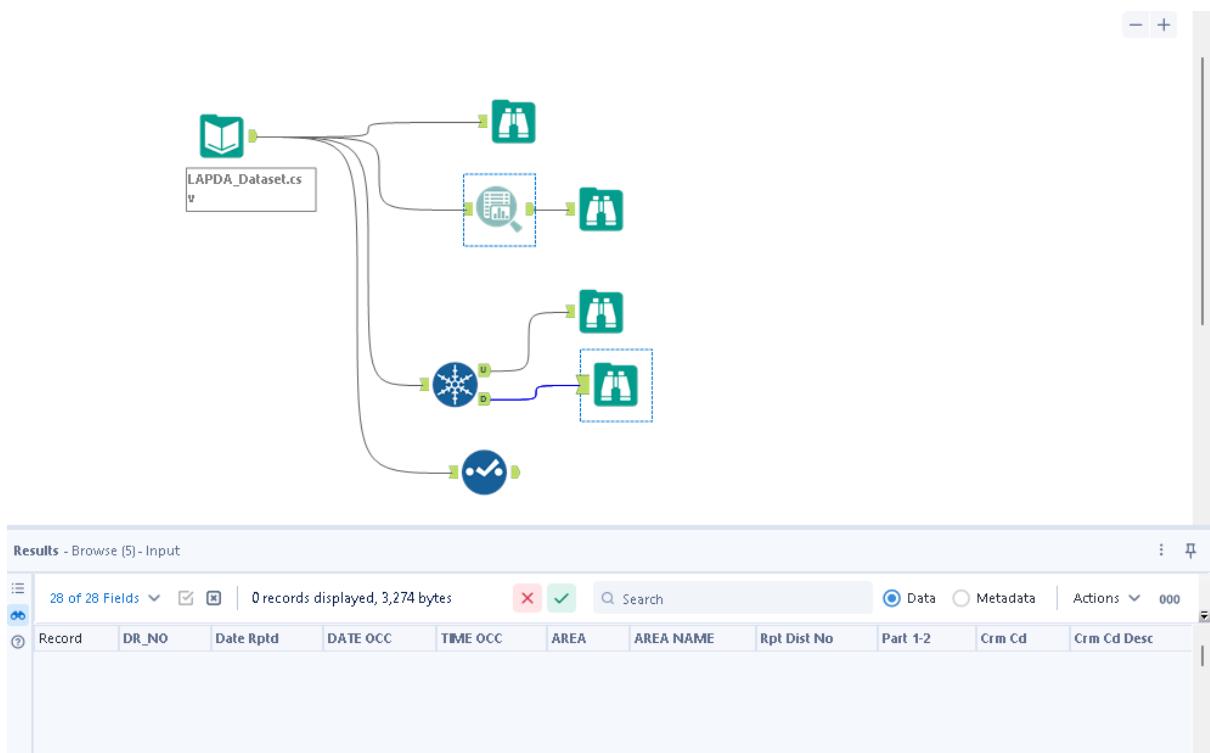
Source to Target DataType

Column_Name	Source Data Type	Target Data type
DR_NO	V_String	V_String
Date Rptd	V_String	Date
DATE OCC	V_String	Date
TIME OCC	V_String	Time
AREA	V_String	Number
AREA NAME	V_String	V_String
Rpt Dist No	V_String	Number
Part 1-2	V_String	Number
Crm Cd	V_String	Number
Crm Cd Desc	V_String	V_String
Mocodes	V_String	To be dropped
Vict Age	V_String	Number
Vict Sex	V_String	V_String
Vict Descent	V_String	V_String
Premis Cd	V_String	Number
Premis Desc	V_String	V_String
Weapon Used Cd	V_String	Number
Weapon Desc	V_String	V_String
Status	V_String	V_String
Status Desc	V_String	V_String
Crm Cd 1	V_String	Number

Crm Cd 2	V_String	Number
Crm Cd 3	V_String	Number
Crm Cd 4	V_String	Number
LOCATION	V_String	V_String
Cross Street	V_String	V_String
LAT	V_String	Double
LON	V_String	Double

Data Quality Assessment

1. Duplicates



Detail : There are no duplicate record while selecting all the columns from the dataset.

2. Missing/Null Values

Verified using basic data profiling tool

Affected Columns: Mocodes ,Vict Sex, Vict Descent ,Premis Cd,Premis Desc, Weapon Used Cd, Weapon Desc ,Status,Crm Cd 1 Crm Cd 2, Crm Cd 3, Crm Cd 4, Cross Street

Results - Browse (7) - Input

Record	FieldName	Name	Value
1	DR_NO	Nulls	0
2	Date Rptd	Nulls	0
3	DATE OCC	Nulls	0
4	TIME OCC	Nulls	0
5	AREA	Nulls	0
6	AREA NAME	Nulls	0
7	Rpt Dist No	Nulls	0
8	Part 1-2	Nulls	0
9	Crm Cd	Nulls	0
10	Crm Cd Desc	Nulls	0
11	Mocodes	Nulls	151619
12	Vict Age	Nulls	0
13	Vict Sex	Nulls	144644
14	Vict Descent	Nulls	144656
15	Premis Cd	Nulls	16
16	Premis Desc	Nulls	588
17	Weapon Used Cd	Nulls	677744
18	Weapon Desc	Nulls	677744
19	Status	Nulls	1
20	Status Desc	Nulls	0

Record	FieldName	Name	Value
10	Crm Cd Desc	Nulls	0
11	Mocodes	Nulls	151619
12	Vict Age	Nulls	0
13	Vict Sex	Nulls	144644
14	Vict Descent	Nulls	144656
15	Premis Cd	Nulls	16
16	Premis Desc	Nulls	588
17	Weapon Used Cd	Nulls	677744
18	Weapon Desc	Nulls	677744
19	Status	Nulls	1
20	Status Desc	Nulls	0
21	Crm Cd 1	Nulls	11
22	Crm Cd 2	Nulls	995831
23	Crm Cd 3	Nulls	1002677
24	Crm Cd 4	Nulls	1004927
25	LOCATION	Nulls	0
26	Cross Street	Nulls	850755
27	LAT	Nulls	0
28	LON	Nulls	0

Recommend Action

Column Names	Replace Null with	Reason
Mocodes	Unkown	To maintain String formmat
Vict Sex	X (As per site X is for unknown)	To maintain String formmat
Vict Descent	X(As per site X is for unknown)	To maintain String formmat
Premis Cd	-1	To maintain number format
Premis Desc	Unknown	To maintain String formmat
Weapon Used Cd	-1	To maintain number format

Weapon Desc	Unknown	To maintain String format
Status	Unknown	To maintain String format
Crm Cd [1-4]	-1	To maintain number format
Cross Street	Unknown	To maintain String format

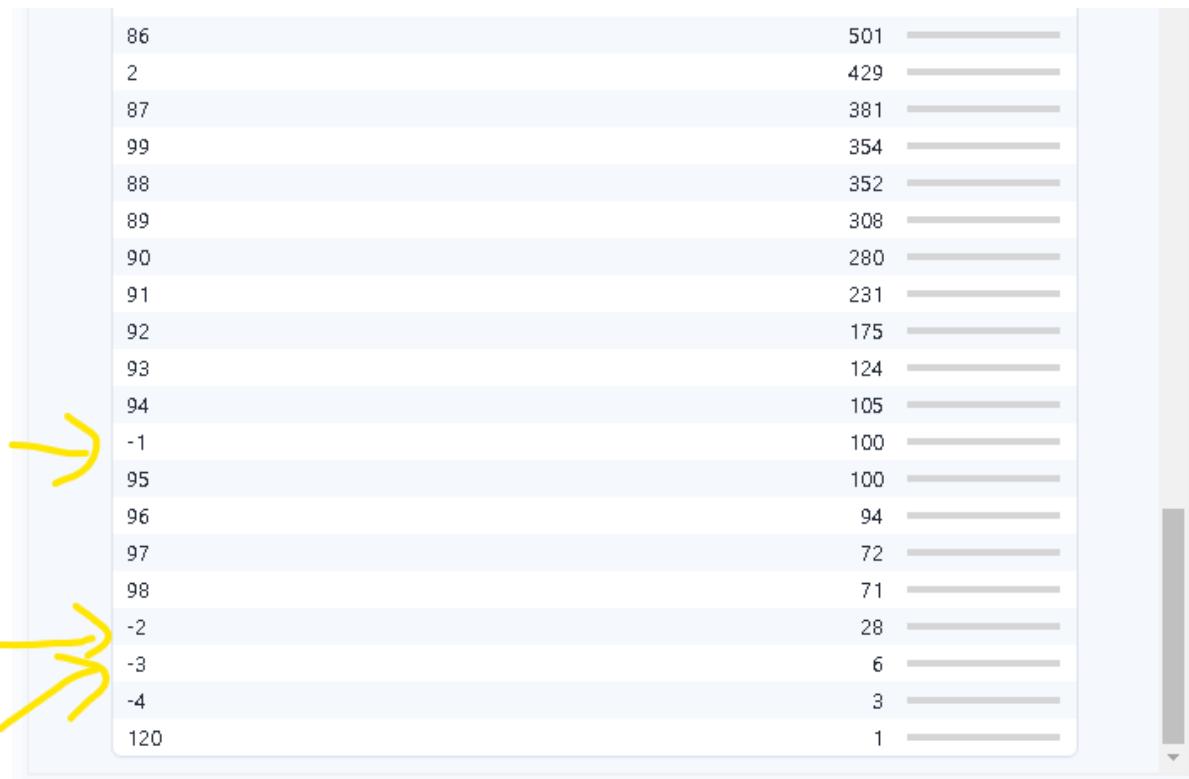
2. Leading and trailing whitespaces

There are no leading and trailing whitespaces in any of the columns in the entire dataset

3. Negative values and '0' present in age

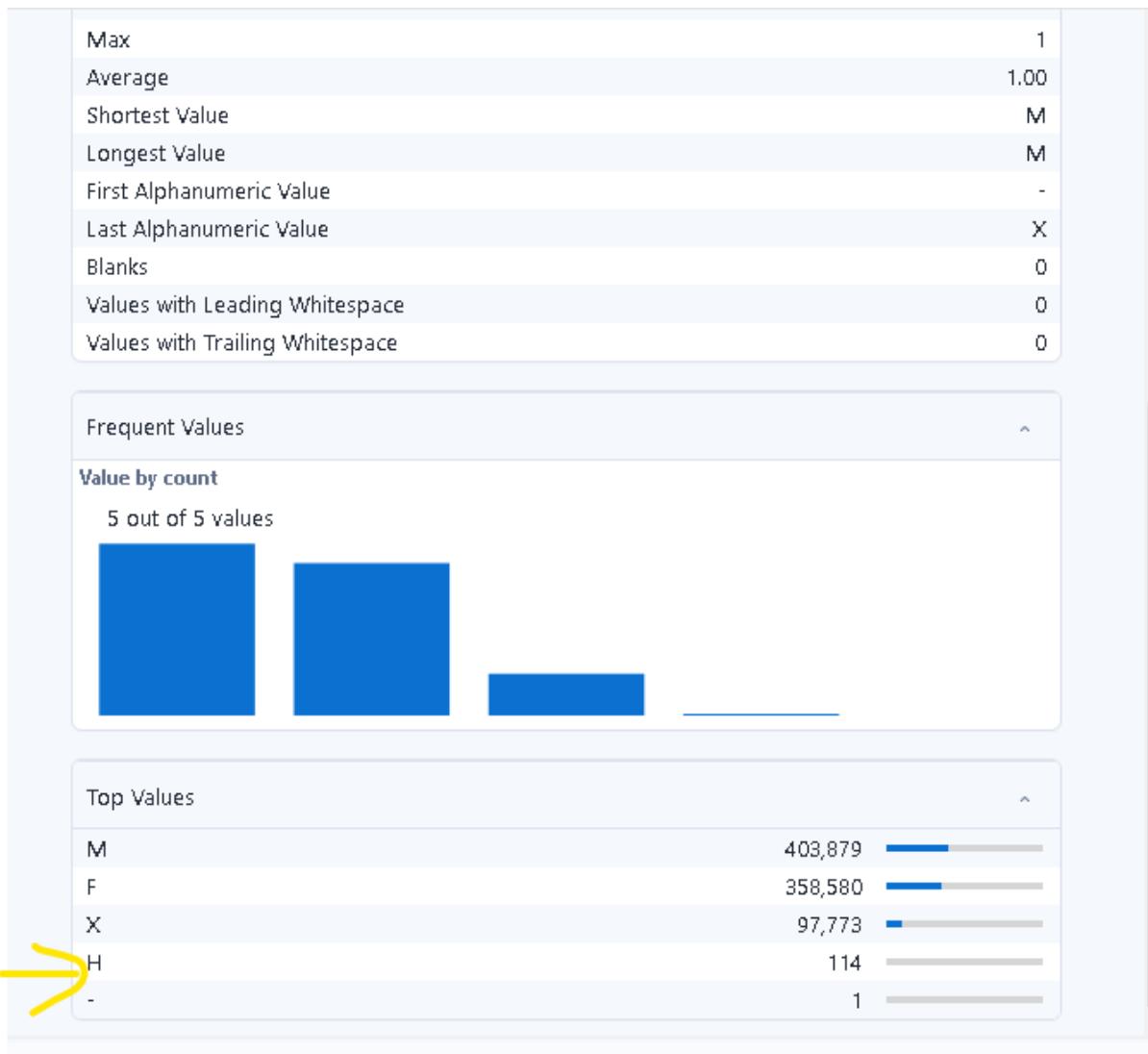
Some values for the age column are having '0' and in negative as well





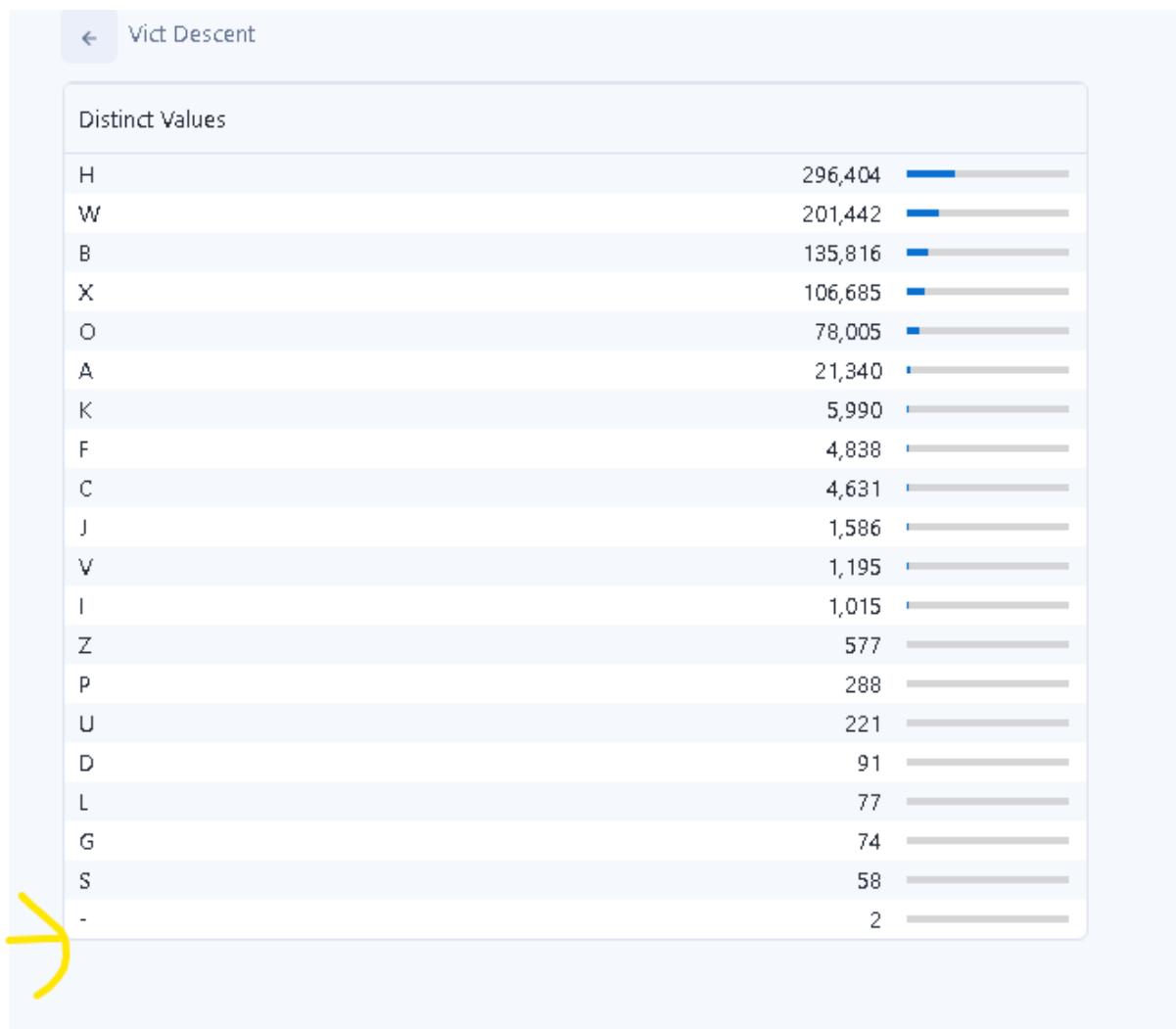
Action: Replace all those values with 99999

4. Some values in **Vict Sex** field are as 'H' and '-' (M=Male ,F- Female, X=Unknown)



Action- Replace 'H' and '-' with X(unknown as per site)

5. **Vict Descent** field has ‘-’ value



Action- Replace ‘-’ with X(Unknown as per site)

6. Some values under **Prem Desc** field have * at the end

The screenshot shows a Microsoft Excel spreadsheet with a list of locations and their corresponding counts. The columns are labeled 'Location' and 'Count'. A vertical scroll bar is visible on the right side of the spreadsheet. Yellow arrows have been drawn on the left side of the image to point to specific rows: one arrow points to 'PATIO*', another to 'SPECIALTY SCHOOL/OTHER', another to 'MEMBERSHIP STORE (COSTCO,SAMS CLUB)*', and another to 'PHARMACY INSIDE STORE OR SUPERMARKET*'. The data is as follows:

MTA - RED LINE - NORTH HOLLYWOOD	1,023
MISSIONS/SHELTERS	1,017
THE GROVE	985
TRANSIENT ENCAMPMENT	840
BEAUTY/BARBER SHOP	787
LIBRARY	777
MOBILE HOME/TRAILERS/CONSTRUCTION TRAILERS/RV'S/M...	777
BAR/SPORTS BAR (OPEN DAY & NIGHT)	771
TRANSITIONAL HOUSING/HALFWAY HOUSE	760
GROUP HOME	758
MTA - RED LINE - 7TH AND METRO CENTER	689
PATIO*	662
MTA PROPERTY OR PARKING LOT	647
WEBSITE	640
MTA - RED LINE - PERSHING SQUARE	568
MTA - RED LINE - UNION STATION	568
AUTO SUPPLY STORE*	560
SPECIALTY SCHOOL/OTHER	552
CAR WASH	540
BEAUTY SUPPLY STORE	519
MEMBERSHIP STORE (COSTCO,SAMS CLUB)*	514
TOBACCO SHOP	488
HARDWARE/BUILDING SUPPLY	456
THE BEVERLY CONNECTION	452
PHARMACY INSIDE STORE OR SUPERMARKET*	451
AUTO DEALERSHIP (CHEVY, FORD, BMW, MERCEDES, ETC.)	440
IFMFI RV STORE	436

Plan of action: Replace("*,")

7. Lat field

In latitude field there is '0' as well

34.1903	2,705
34.1938	2,623
34.0998	2,598
34.0636	2,487
34.1576	2,454
34.0736	2,449
34.2011	2,405
34.0981	2,366
34.0761	2,343
34.0483	2,299
34.2048	2,252
0	2,240
33.9456	2,116
34.0944	2,044
34.1976	1,999
34.1794	1,942
34.0467	1,938

Action- Replace '0' with '99999'

8. Lon field

Longitude has '0' for some records

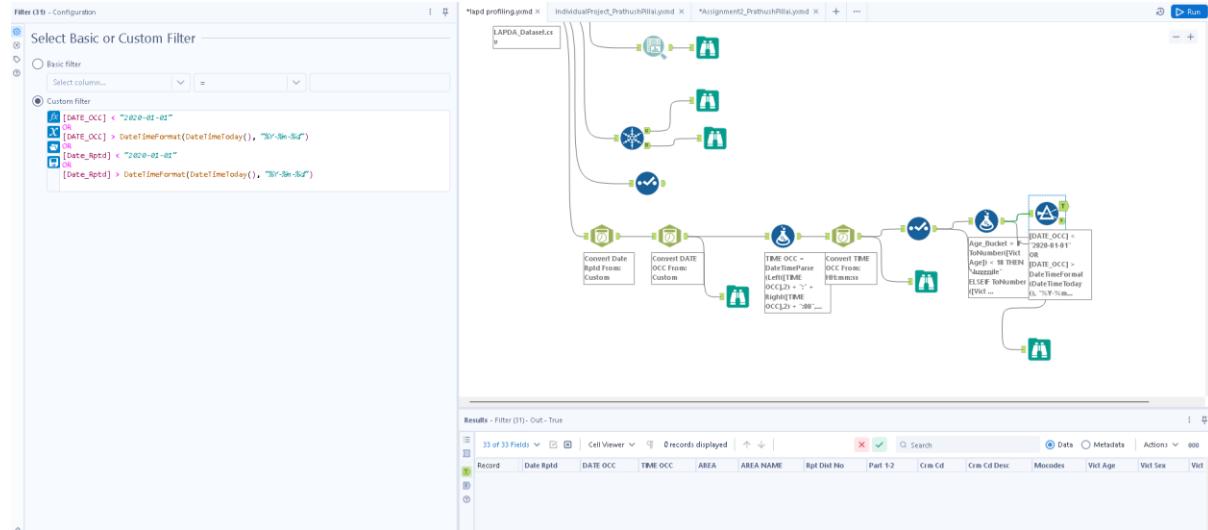
-118.2783	3,759
-118.2916	3,731
-118.309	3,629
-118.4662	2,725
-118.2695	2,655
-118.3002	2,568
-118.2375	2,551
-118.3965	2,517
-118.287	2,515
-118.4487	2,460
-118.6059	2,406
-118.5361	2,312
-118.3703	2,252
0	2,240
-118.3092	2,181
-118.263	2,175
-118.2917	2,161
-118.3563	2,153
-118.2631	2,146
-118.2565	2,128
-118.2586	2,028

Action- Replace '0' with '-99999'

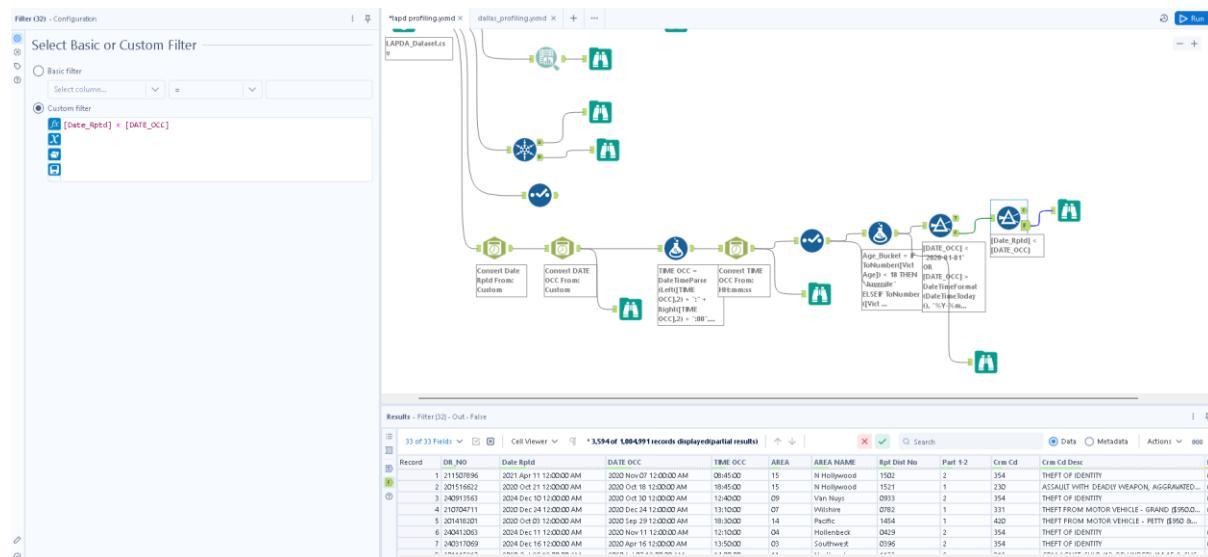
Crm Cd1 and Crm Cd fields are the same

9 . Valid Data range:

Using the filter tool checked the records to make sure nothing is beyond 2020 and no records of future are present which showed 0 records falling in the criteria:



10. Checking if there are any data where Reported Date is before Occurrence Date which is impossible. Used the filter to check and there are no such records



Summary Statistics

1 of 1 Fields | Records 1 to 1 | Report | String/Character Fields

Name	% Missing	Unique Values	Shortest Value	Longest Value	Min Value Count	Max Value Count	Remarks
LAT	0.0%	5,426	0	34.2124	1	5,705	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Crm Cd 3	99.8%	39	998	998	1	1,002,677	This field has over 10% missing values. Consider imputing these values. Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
LON	0.0%	4,982	0	-118.4092	1	7,550	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Crm Cd 2	93.1%	127	860	860	1	935,831	This field has over 10% missing values. Consider imputing these values. Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Status Desc	0.0%	6	UNK	Adult Arrest	7	802,862	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Cross Street	84.7%	10,414	5	S CENTRAL AV	1	850,755	This field has over 10% missing values. Consider imputing these values. Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Crm Cd 1	0.0%	143	354	354	1	115,174	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Crm Cd	0.0%	140	354	354	1	115,190	Some values of this field have a small number of value counts. If Appropriate,

Report | Profile | 1 record displayed, 2 fields, 4,720 bytes | Records 1 to 1

Date Rptd	0.0%	1,896	2021 Apr 11 12:00:00 AM	2021 Apr 11 12:00:00 AM	1	929	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Mocodes	15.1%	310,941	0377	0907 0450 1407 1822 0342 1100 0919 1310 1312 0432	1	151,619	This field has over 10% missing values. Consider imputing these values. Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
DR_NO	0.0%	1,004,991	0817	211507896	1	1	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Premis Cd	0.0%	315	501	501	1	261,284	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
DATE OCC	0.0%	1,879	2020 Nov 07 12:00:00 AM	2020 Nov 07 12:00:00 AM	1	1,164	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Rpt Dist No	0.0%	1,210	1502	1502	1	5,403	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Vict Descent	14.4%	21	H	H	2	296,404	This field has over 10% missing values. Consider imputing these values. Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
LOCATION	0.0%	66,566	D	14600 SYLVAN ST	1	2,598	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Premis Desc	0.1%	307	BANK	VEHICLE STORAGE LOT (CARS, TRUCKS, RV'S, BOATS, TRAILERS, ETC.)	1	261,284	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Vict Sex	14.4%	6	M	M	1	403,879	This field has over 10% missing values.

Report | Profile |

1 of 1 Fields | Records 1 to 1 |

Crm Cd Desc	0.0%	140 ARSON	CRM AGNST CHLD (13 OR UNDER) (14-15 & SUSP 10 YRS OLDER)	1	115,190	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Status	0.0%	7 IC	IC	1	802,862	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
AREA NAME	0.0%	21 Newton	N Hollywood	33,133	69,670	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Crm Cd 4	100.0%	7 998	998	1	1,004,927	This field has over 10% missing values. Consider imputing these values. Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Weapon Desc	67.4%	80 AXE	STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)	1	677,744	This field has over 10% missing values. Consider imputing these values. Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Vict Age	0.0%	104 0	120	1	269,222	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
TIME OCC	0.0%	1,439 0845	0845	19	35,200	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Weapon Used Cd	67.4%	80 200	200	1	677,744	This field has over 10% missing values. Consider imputing these values. Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
AREA	0.0%	21 15	15	33,133	69,670	Some values of this field have a small number of value counts. If Appropriate, consider combining some value levels together.
Part 1-2	0.0%	2 2	2	402,346	602,645	

Data Cleaning:

For cleaning have done the following steps in alteryx before exporting to csv file:

1. Date Rptd and Date Occ field has been converted from string to Date/Time

The screenshot shows the Alteryx Data Cleaning interface. On the left, the "DateTime (13) - Configuration" panel is open, showing the following settings:

- Select the format to convert: String to Date/Time format (radio button selected).
- Select the string field to convert: Date Rptd.
- Specify the new column name: Date_Rptd.
- Specify your DateTime Language: English.
- Select the format that matches the incoming string field: yyyy-MM-dd hh:mm:ss.
- Specify the format of the incoming string field: %Y %b %d %I:%M:%S %p.
- Example: 2000 Jan 02 03:04:05 AM.
- Output: 2000-01-02 03:04:05.

On the right, the workflow diagram titled "lapd profiling.ymd" is visible. It consists of two "Convert DATE" tools, each with a "Custom" configuration, connected by a green line. The first "Convert DATE" tool has an incoming connection from a blue "Source" tool and an outgoing connection to a green "Sink" tool. The second "Convert DATE" tool also has an incoming connection from the first one and an outgoing connection to the green "Sink" tool. The green "Sink" tool has a connection leading to a green "Metrics" tool, which is connected to a green "Audit" tool. A legend on the right side of the interface lists various symbols corresponding to different types of tools and components.

2. For the Time Occ since it was in 4 digits made it to HH:MM:SS format and converted it from string to date/time

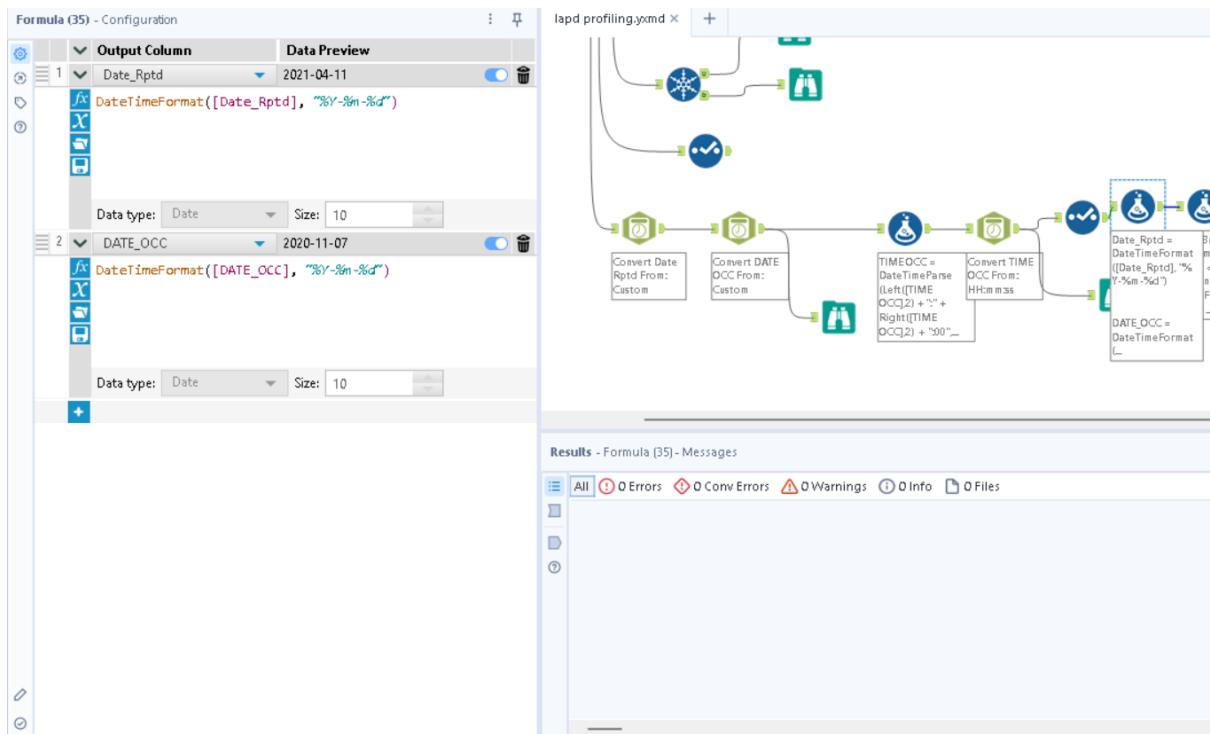
Output Column: TIME OCC
Data Preview: 08:45:00
fx DateTimeParse(Left([TIME OCC],2) + ":" + Right([TIME OCC],2) + ":00", "%H:%M:%S")
Data type: V_String Size: 254

Results - Formula (18) - Messages: All 0 Errors 0 Conv Errors 0 Warnings 0 Info 0 Files

DateTime (21) - Configuration:
Select the format to convert:
String to Date/Time format
TIME OCC
Specify the new column name: TIME_OCC
Specify your DateTime Language: English
Select the format that matches the incoming string field:
MM/dd/yy
MM/dd/yyyy
Month dd, yyyy
Month, yyyy
yyyy-MM-dd
yyyyMMdd
yyyy/MM/dd
HH:MM:SS
Custom
Specify the format of the incoming string field:
Example Output

Results - DateTime (21) - Messages: All 0 Errors 0 Conv Errors 0 Warnings 0 Info 0 Files

3. Then for the Date_Rptd and Date_Occ since only the date part is need and time is just a placeholder/relevant extracted only the date part



4 .Used the select tool to rename the columns as per Dimensional model and also dropped fields which were not required and changed data types of some fields:

	<input type="checkbox"/>	Column	Type	Size	Rename	Description
>	<input checked="" type="checkbox"/>	DR_NO	V_String	254	File_Number	
	<input checked="" type="checkbox"/>	Date_Rptd	Date	10	Date_Reported	
	<input checked="" type="checkbox"/>	DATE_OCC	Date	10	Date_Occured	
	<input checked="" type="checkbox"/>	TIME_OCC	Time	8	Time_Occured	
	<input checked="" type="checkbox"/>	Age_Bucket	V_String	214...	Victim_Age_Category	
	<input checked="" type="checkbox"/>	Time_Bucket	V_WString	107...	Period_of_Day	
	<input checked="" type="checkbox"/>	AREA	Int64	8	Area	
	<input checked="" type="checkbox"/>	AREA NAME	V_String	254	Area_Name	
	<input checked="" type="checkbox"/>	Rpt Dist No	Int64	8	Reported_District_No	
	<input checked="" type="checkbox"/>	Part 1-2	Int64	8	Part_1_2	
	<input checked="" type="checkbox"/>	Crm Cd	Int64	8	Crime_Code	
	<input checked="" type="checkbox"/>	Crm Cd Desc	V_String	254	Crime_Code_Desc	
	<input checked="" type="checkbox"/>	Vict Age	Int64	8	Victim_Age	
	<input checked="" type="checkbox"/>	Vict Sex	V_String	254	Victim_Sex	
	<input checked="" type="checkbox"/>	Vict Descent	V_String	254	Victim_Descent	
	<input checked="" type="checkbox"/>	Premis Cd	Int64	8	Premis_Code	
	<input checked="" type="checkbox"/>	Premis Desc	V_String	254	Premis_Description	
	<input checked="" type="checkbox"/>	Weapon Used Cd	Int64	8	Weapon_Used_Code	
	<input checked="" type="checkbox"/>	Weapon Desc	V_String	254	Weapon_Desc	
	<input checked="" type="checkbox"/>	Status	V_String	254	Arrest_Status	
	<input checked="" type="checkbox"/>	Status Desc	V_String	254	Arrest_Status_Desc	

Options Search

<input type="checkbox"/>	Column	Type	Size	Rename	Description
<input checked="" type="checkbox"/>	Vict Sex	V_String	▼ 254	Victim_Sex	
<input checked="" type="checkbox"/>	Vict Descent	V_String	▼ 254	Victim_Descent	
<input checked="" type="checkbox"/>	Premis Cd	Int64	▼ 8	Premis_Code	
<input checked="" type="checkbox"/>	Premis Desc	V_String	▼ 254	Premis_Description	
<input checked="" type="checkbox"/>	Weapon Used Cd	Int64	▼ 8	Weapon_Used_Code	
<input checked="" type="checkbox"/>	Weapon Desc	V_String	▼ 254	Weapon_Desc	
<input checked="" type="checkbox"/>	Status	V_String	▼ 254	Arrest_Status	
<input checked="" type="checkbox"/>	Status Desc	V_String	▼ 254	Arrest_Status_Desc	
<input checked="" type="checkbox"/>	LOCATION	V_String	▼ 254	Location	
<input checked="" type="checkbox"/>	Cross Street	V_String	▼ 254	Cross_Street	
<input checked="" type="checkbox"/>	LAT	Double	▼ 8	Latitude	
<input checked="" type="checkbox"/>	LON	Double	▼ 8	Longitude	
<input type="checkbox"/>	*Unknown	Unknown	▼ 0		Dynamic or Unknown Columns
<input type="checkbox"/>	Date Rptd	V_String	▼ 254		
<input type="checkbox"/>	DATE OCC	V_String	▼ 254		
<input type="checkbox"/>	TIME OCC	V_String	▼ 254		
<input type="checkbox"/>	Mocodes	V_String	▼ 254		
<input type="checkbox"/>	Crm Cd 2	V_String	▼ 254		
<input type="checkbox"/>	Crm Cd 3	V_String	▼ 254		
<input type="checkbox"/>	Crm Cd 4	V_String	▼ 254		
<input type="checkbox"/>	Crm Cd 1	V_String	▼ 254		

Then In Databricks from the bronze layer(created via csv in volume) ,did the following cleaning in the silver layer and also changed the datatypes which was to be done:

For Victim_Sex field replace all nulls, 'H' , '-' with X

For Victim_Descent field replace all nulls , '-' with X

For Victim_Age replace '0' and all negative values with '99999'

For Premis_Code field replace all nulls with -1

For Premis_Description field replace nulls with 'Unknown' ,also remove the "*" at the end of the values

For Weapon_Used_Code field replace nulls with '-1'

For Weapon_Desc field replace nulls with 'Unknown'

For Arrest_Status field replace nulls with 'Unknown'

For Cross_Street field replace nulls with 'Unknown'

For Latitude replace '0' with '99999'

For longitude replace '0' with '-99999'

```
# =====#
# Victim_Sex + replace nulls, 'H', '-' > "X"
df = df.withColumn(
    "victim_Sex",
    when(col("Victim_Sex").isin("", "H", "-") | col("Victim_Sex").isNull(), "X")
    .otherwise(col("Victim_Sex"))
)

# Victim_Descent + replace nulls, '-' > X
df = df.withColumn(
    "victim_Descent",
    when(col("Victim_Descent").isin("", "-") | col("Victim_Descent").isNull(), "X")
    .otherwise(col("Victim_Descent"))
)

# Victim_Age + replace 0 and negative with 99999
df = df.withColumn(
    "victim_Age",
    when((col("Victim_Age") <= 0) | col("Victim_Age").isNull(), lit(99999))
    .otherwise(col("Victim_Age"))
)

# Premis_Code + replace null with -1
df = df.withColumn(
    "premis_Code",
    when(col("Premis_Code").isNull() | (col("Premis_Code") == ""), lit(-1))
    .otherwise(col("Premis_Code"))
)

# Premis_Description + replace nulls + remove "*"
df = df.withColumn(
    "premis_Description",
    regexp_replace(
        when(col("Premis_Description").isNull() | (col("Premis_Description") == ""), "Unknown")
        .otherwise(col("Premis_Description"))
),
```

```

# Weapon_Used_Code + nulls → -1
df = df.withColumn(
    "Weapon_Used_Code",
    when(col("Weapon_Used_Code").isNull() | (col("Weapon_Used_Code") == ""), lit(-1))
    .otherwise(col("Weapon_Used_Code"))
)

# Weapon_Desc → nulls → Unknown
df = df.withColumn(
    "Weapon_Desc",
    when(col("Weapon_Desc").isNull() | (col("Weapon_Desc") == ""), "Unknown")
    .otherwise(col("Weapon_Desc"))
)

# Arrest_Status → nulls → Unknown
df = df.withColumn(
    "Arrest_Status",
    when(col("Arrest_Status").isNull() | (col("Arrest_Status") == ""), "Unknown")
    .otherwise(col("Arrest_Status"))
)

# Cross_Street + nulls + Unknown
df = df.withColumn(
    "Cross_Street",
    when(col("Cross_Street").isNull() | (col("Cross_Street") == ""), "Unknown")
    .otherwise(col("Cross_Street"))
)

# Latitude → 0 → 99999
df = df.withColumn(
    "Latitude",
    when(col("Latitude") == 0, lit(99999)).otherwise(col("Latitude"))
)

```

```

# Latitude → 0 → 99999
df = df.withColumn(
    "Latitude",
    when(col("Latitude") == 0, lit(99999)).otherwise(col("Latitude"))
)

# Longitude → 0 → -99999
df = df.withColumn(
    "Longitude",
    when(col("Longitude") == 0, lit(-99999)).otherwise(col("Longitude"))
)

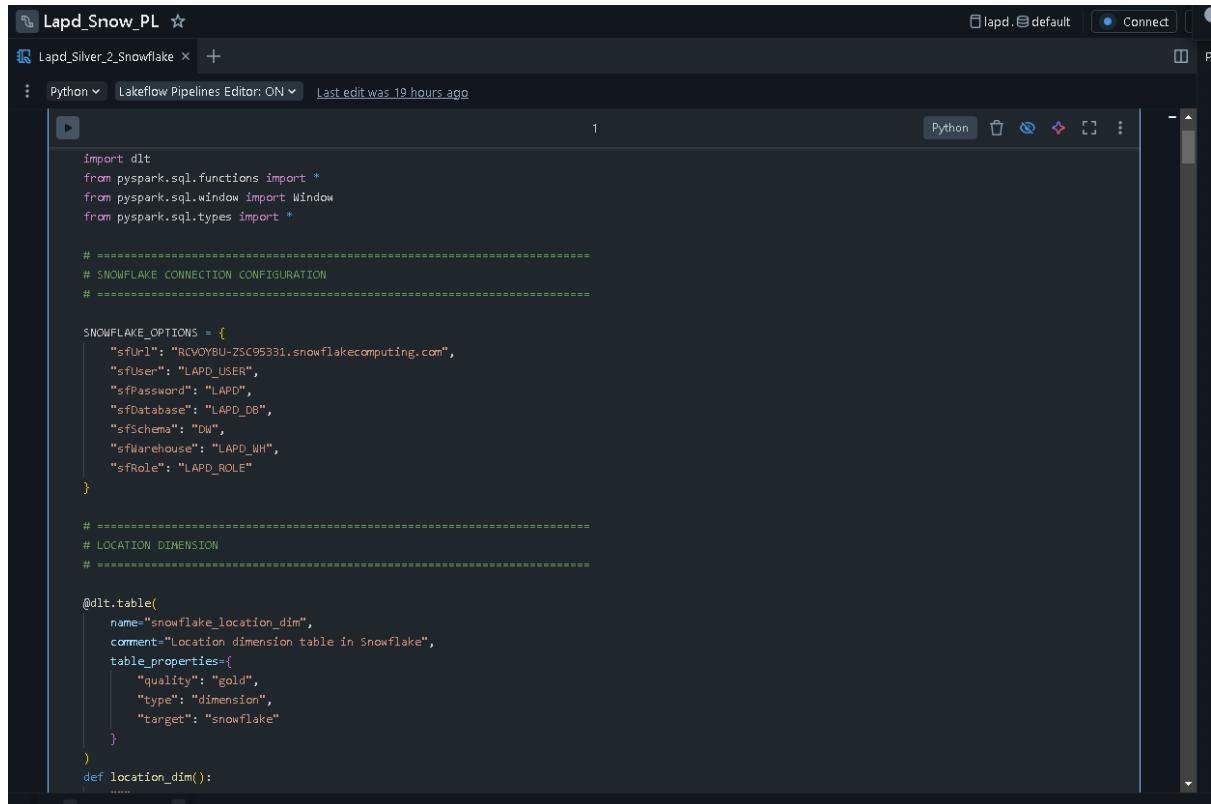
# =====#
# DATATYPE CASTING (STRICTLY AS YOU REQUESTED)
# =====#

df = (
    df
    .withColumn("Date_Reported", to_date("Date_Reported"))
    .withColumn("Date_Occured", to_date("Date_Occured"))
    .withColumn("Time_Occured", date_format(to_timestamp("Time_Occured", "HH:mm:ss"), "HH:mm:ss"))
    .withColumn("Area", col("Area").cast("INT"))
    .withColumn("Reported_District_No", col("Reported_District_No").cast("INT"))
    .withColumn("Part_1_2", col("Part_1_2").cast("INT"))
    .withColumn("Crime_Code", col("Crime_Code").cast("INT"))
    .withColumn("Victim_Age", col("Victim_Age").cast("INT"))
    .withColumn("Premis_Code", col("Premis_Code").cast("INT"))
    .withColumn("Weapon_Used_Code", col("Weapon_Used_Code").cast("INT"))
    .withColumn("Latitude", col("Latitude").cast("DOUBLE"))
    .withColumn("Longitude", col("Longitude").cast("DOUBLE"))
)

```

```
return df
```

Below is the snapshot of the code used to load the facts and dimensions onto snowflake ie silver to gold



```
import dlt
from pyspark.sql.functions import *
from pyspark.sql.window import Window
from pyspark.sql.types import *

# =====
# SNOWFLAKE CONNECTION CONFIGURATION
# =====

SNOWFLAKE_OPTIONS = {
    "sfurl": "RCVOYBU-ZSC95331.snowflakecomputing.com",
    "sfuser": "LAPD_USER",
    "sfpassword": "LAPD",
    "sfdatabase": "LAPD_DB",
    "sfschema": "DLM",
    "sfwarehouse": "LAPD_WH",
    "sfRole": "LAPD_ROLE"
}

# =====
# LOCATION DIMENSION
# =====

@dlt.table(
    name="snowflake_location_dim",
    comment="Location dimension table in Snowflake",
    table_properties={
        "quality": "gold",
        "type": "dimension",
        "target": "snowflake"
    }
)
def location_dim():
    ....
```



```
) def location_dim():
    """
    Create Location dimension from silver data and write to Snowflake
    """
    silver_df = dlt.read("lapd.silver.lapd_silver")

    # Business key columns
    bk_cols = ["Area", "Area_Name", "Reported_District_No", "Location", "Cross_Street"]

    location_df = (
        silver_df
        .select(
            "Area",
            "Area_Name",
            "Reported_District_No",
            "Location",
            "Cross_Street",
            "Latitude",
            "Longitude"
        )
        .filter(col("Area").isNotNull())
        .dropDuplicates(bk_cols)
    )

    w = Window.orderBy(*bk_cols)
    location_df = location_df.withColumn("Location_ID", row_number().over(w))

    final_df = location_df.select(
        col("Location_ID").cast("int"),
        col("Area").cast("int"),
        col("Area_Name"),
        ....
```

```

fact_dt = fact_dt.join(
    descent_df,
    fact_df.Victim_Descent == descent_df.Victim_Descent,
    "left"
).drop(descent_df.Victim_Descent)

# Join with Premise dimension
fact_df = fact_df.join(
    premise_df,
    fact_df.Premis_Code == premise_df.Premis_Code,
    "left"
).drop(premise_df.Premis_Code)

# Join with Weapon dimension
fact_df = fact_df.join(
    weapon_df,
    fact_df.Weapon_Used_Code == weapon_df.Weapon_Used_Code,
    "left"
).drop(weapon_df.Weapon_Used_Code)

# Join with Arrest dimension
fact_df = fact_df.join(
    arrest_df,
    (fact_df.Arrest_Status == arrest_df.Arrest_status) &
    (fact_df.Arrest_Status_Desc == arrest_df.Arrest_status_desc),
    "left"
).drop(arrest_df.Arrest_status, arrest_df.Arrest_status_desc)

```

Tables 8 Performance 8

Ran the pipeline:

Lapd_Snow_PL

Python | Lakeflow Pipelines Editor: ON | Last edit was 20 hours ago

Tables 8 Performance 8

LAPD DLT Pipeline -- Silver to Snowflake Gold Layer
Creates Dimension and Fact tables directly in Snowflake

import dtl
from awswrangler.dql.functions import *

Name	Catalog	Schema	Type	Duration	Output rows	Expectations	Dropped	Warnings	Failed	Incrementalization
snowflake_arrest_dim	lapd	default	Materialized view	11s	7	Not defined	-	-	-	Full recompute
snowflake_crimeType_dim	lapd	default	Materialized view	11s	140	Not defined	-	-	-	Full recompute
snowflake_descent_dim	lapd	default	Materialized view	13s	19	Not defined	-	-	-	Full recompute
snowflake_fact_crime	lapd	default	Materialized view	15s	1M	Not defined	-	-	-	Full recompute
snowflake_location_dim	lapd	default	Materialized view	7s	151K	Not defined	-	-	-	Full recompute
snowflake_premis_dim	lapd	default	Materialized view	8s	315	Not defined	-	-	-	Full recompute
snowflake_victim_dim	lapd	default	Materialized view	10s	296	Not defined	-	-	-	Full recompute
snowflake_weapon_dim	lapd	default	Materialized view	11s	80	Not defined	-	-	-	Full recompute

Output of facts and dimensions on snowflake:

```

12
13 | select * from arrest_dim
14
15
16
Results (just now)
Table | Chart

```

#	ARREST_SK	ARREST_STATUS	ARREST_STATUS_DESC	INT_DLLOAD_DT	DL_SOURCE_ID
1	1	AA	Adult Arrest	2025-11-23 03:40:36.032	DATABRICKS_DLT
2	2	AO	Adult Other	2025-11-23 03:40:36.032	DATABRICKS_DLT
3	3	CC	UNK	2025-11-23 03:40:36.032	DATABRICKS_DLT
4	4	IC	Invest Cont	2025-11-23 03:40:36.032	DATABRICKS_DLT
5	5	JA	Juv Arrest	2025-11-23 03:40:36.032	DATABRICKS_DLT
6	6	JO	Juv Other	2025-11-23 03:40:36.032	DATABRICKS_DLT
7	7	Unknown	UNK	2025-11-23 03:40:36.032	DATABRICKS_DLT

My Workspace > Untitled 1.sql

1 | select * from dw.date_dim

Results (just now)

Table Chart

id	#DATE_SK	FULL_DATE	#DAY_OF_WEEK	DAY_NAME	#MONTH_NUM	MONTH_NAME	#QUARTER	QUARTER_NAME	#YEAR	MONTH_START_DATE	MONTH_END_DATE	IS_WEEKEND
1	20100101	2010-01-01	5	Fri	1	Jan	1	Q1	2010	2010-01-01	2010-01-31	N
2	20100102	2010-01-02	6	Sat	1	Jan	1	Q1	2010	2010-01-01	2010-01-31	N
3	20100103	2010-01-03	0	Sun	1	Jan	1	Q1	2010	2010-01-01	2010-01-31	N
4	20100104	2010-01-04	1	Mon	1	Jan	1	Q1	2010	2010-01-01	2010-01-31	Y
5	20100105	2010-01-05	2	Tue	1	Jan	1	Q1	2010	2010-01-01	2010-01-31	N
6	20100106	2010-01-06	3	Wed	1	Jan	1	Q1	2010	2010-01-01	2010-01-31	N
7	20100107	2010-01-07	4	Thu	1	Jan	1	Q1	2010	2010-01-01	2010-01-31	N
8	20100108	2010-01-08	5	Fri	1	Jan	1	Q1	2010	2010-01-01	2010-01-31	N
9	20100109	2010-01-09	6	Sat	1	Jan	1	Q1	2010	2010-01-01	2010-01-31	N
10	20100110	2010-01-10	0	Sun	1	Jan	1	Q1	2010	2010-01-01	2010-01-31	N
11	20100111	2010-01-11	1	Mon	1	Jan	1	Q1	2010	2010-01-01	2010-01-31	Y
12	20100112	2010-01-12	2	Tue	1	Jan	1	Q1	2010	2010-01-01	2010-01-31	N
13	20100113	2010-01-13	3	Wed	1	Jan	1	Q1	2010	2010-01-01	2010-01-31	N
14	20100114	2010-01-14	4	Thu	1	Jan	1	Q1	2010	2010-01-01	2010-01-31	N
15	20100115	2010-01-15	5	Fri	1	Jan	1	Q1	2010	2010-01-01	2010-01-31	N
16	20100116	2010-01-16	6	Sat	1	Jan	1	Q1	2010	2010-01-01	2010-01-31	N
17	20100117	2010-01-17	0	Sun	1	Jan	1	Q1	2010	2010-01-01	2010-01-31	N
18	20100118	2010-01-18	1	Mon	1	Jan	1	Q1	2010	2010-01-01	2010-01-31	Y
19	20100119	2010-01-19	2	Tue	1	Jan	1	Q1	2010	2010-01-01	2010-01-31	N
20	20100120	2010-01-20	3	Wed	1	Jan	1	Q1	2010	2010-01-01	2010-01-31	N
21	20100121	2010-01-21	4	Thu	1	Jan	1	Q1	2010	2010-01-01	2010-01-31	N
22	20100122	2010-01-22	5	Fri	1	Jan	1	Q1	2010	2010-01-01	2010-01-31	N
23	20100123	2010-01-23	6	Sat	1	Jan	1	Q1	2010	2010-01-01	2010-01-31	N

5
6 | select * from dw.TIME_DIM
7
8

Results (just now)

Table Chart

ID	TIME_KEY	TIME_24_HR	HOUR	MINUTE	PERIOD_OF_DAY	DT_SOURCE_ID	DLLOAD_DT
	0000	0.1%	00:00	0.1%	Night	100.0%	11/21/2025 11/21/2025
	0001	0.1%	00:01	0.1%	Morning		
	+98 more	+98 more	0	59	+2 more	LAPD_DW	2025-11-22
1	0000	00:00	0	0	Night	LAPD_DW	2025-11-22
2	0001	00:01	0	1	Night	LAPD_DW	2025-11-22
3	0002	00:02	0	2	Night	LAPD_DW	2025-11-22
4	0003	00:03	0	3	Night	LAPD_DW	2025-11-22
5	0004	00:04	0	4	Night	LAPD_DW	2025-11-22
6	0005	00:05	0	5	Night	LAPD_DW	2025-11-22
7	0006	00:06	0	6	Night	LAPD_DW	2025-11-22
8	0007	00:07	0	7	Night	LAPD_DW	2025-11-22
9	0008	00:08	0	8	Night	LAPD_DW	2025-11-22
10	0009	00:09	0	9	Night	LAPD_DW	2025-11-22
11	0010	00:10	0	10	Night	LAPD_DW	2025-11-22
12	0011	00:11	0	11	Night	LAPD_DW	2025-11-22
13	0012	00:12	0	12	Night	LAPD_DW	2025-11-22
14	0013	00:13	0	13	Night	LAPD_DW	2025-11-22
15	0014	00:14	0	14	Night	LAPD_DW	2025-11-22
16	0015	00:15	0	15	Night	LAPD_DW	2025-11-22
17	0016	00:16	0	16	Night	LAPD_DW	2025-11-22
18	0017	00:17	0	17	Night	LAPD_DW	2025-11-22
19	0018	00:18	0	18	Night	LAPD_DW	2025-11-22

15
16
17 Ctrl+L to generate
18 select * from crimetype_dim
19
20

Results (just now)

Table | Chart

	# CRIME_SK	# CRIME_CODE	A CRIME_CODE_DESC	# PART_1_2	INF DLLOAD_DT	A DL_SOURCE_ID
1	140	956	ARSON 0.7% ASSAULT WITH DEADLY WEAP. 0.7% +98 more	1 2	11/22/2025 11/22/2025	DATABRICKS_DLT 100.0%
1	1	110	CRIMINAL HOMICIDE	1	2025-11-23 03:40:27.530	DATABRICKS_DLT
2	2	113	MANSLAUGHTER, NEGLECT	1	2025-11-23 03:40:27.530	DATABRICKS_DLT
3	3	121	RAPE, FORCIBLE	1	2025-11-23 03:40:27.530	DATABRICKS_DLT
4	4	122	RAPE, ATTEMPTED	1	2025-11-23 03:40:27.530	DATABRICKS_DLT
5	5	210	ROBBERY	1	2025-11-23 03:40:27.530	DATABRICKS_DLT
6	6	220	ATTEMPTED ROBBERY	1	2025-11-23 03:40:27.530	DATABRICKS_DLT
7	7	230	ASSAULT WITH DEADLY WEAPON, AGGRAVATED	1	2025-11-23 03:40:27.530	DATABRICKS_DLT
8	8	231	ASSAULT WITH DEADLY WEAPON ON POLICE OFFICER	1	2025-11-23 03:40:27.530	DATABRICKS_DLT
9	9	235	CHILD ABUSE (PHYSICAL) - AGGRAVATED	1	2025-11-23 03:40:27.530	DATABRICKS_DLT
10	10	236	INTIMATE PARTNER - AGGRAVATED ASSAULT	1	2025-11-23 03:40:27.530	DATABRICKS_DLT
11	11	237	CHILD NEGLECT (SEE 300 W.I.C.)	2	2025-11-23 03:40:27.530	DATABRICKS_DLT
12	12	250	SHOTS FIRED AT MOVING VEHICLE, TRAIN	1	2025-11-23 03:40:27.530	DATABRICKS_DLT
13	13	251	SHOTS FIRED AT INHABITED DWELLING	1	2025-11-23 03:40:27.530	DATABRICKS_DLT
14	14	310	BURGLARY	1	2025-11-23 03:40:27.530	DATABRICKS_DLT
15	15	320	BURGLARY, ATTEMPTED	1	2025-11-23 03:40:27.530	DATABRICKS_DLT
16	16	330	BURGLARY FROM VEHICLE	1	2025-11-23 03:40:27.530	DATABRICKS_DLT

27
28 select * from descent_dim
29
30
31
32
33
34

Results (just now)

Table | Chart

	# DESCENT_SK	A VICTIM_DESCENT	A DESCENT_DESCRIPTION	INF DLLOAD_DT	A DL_SOURCE_ID
1	19	A B +17 more	American Indian Asian +17 more	11/22/2025 11/22/2025	DATABRICKS_DLT 100.0%
1	1	A	Asian	2025-11-23 03:41:02.707	DATABRICKS_DLT
2	2	B	Black	2025-11-23 03:41:02.707	DATABRICKS_DLT
3	3	C	Chinese	2025-11-23 03:41:02.707	DATABRICKS_DLT
4	4	D	Cambodian	2025-11-23 03:41:02.707	DATABRICKS_DLT
5	5	F	Filipino	2025-11-23 03:41:02.707	DATABRICKS_DLT
6	6	G	Guamanian	2025-11-23 03:41:02.707	DATABRICKS_DLT
7	7	H	Hispanic/Latino/Mexican	2025-11-23 03:41:02.707	DATABRICKS_DLT
8	8	I	American Indian	2025-11-23 03:41:02.707	DATABRICKS_DLT
9	9	J	Japanese	2025-11-23 03:41:02.707	DATABRICKS_DLT
10	10	K	Korean	2025-11-23 03:41:02.707	DATABRICKS_DLT
11	11	L	Laotian	2025-11-23 03:41:02.707	DATABRICKS_DLT
12	12	O	Other	2025-11-23 03:41:02.707	DATABRICKS_DLT
13	13	P	Pacific Islander	2025-11-23 03:41:02.707	DATABRICKS_DLT

54
55 select * from location_dim
56
57
58
59
60
61
62

Results (just now)

Table | Chart

	# LOCATIONID	# AREA	A AREA_NAME	# REPORTING_DISTRICT	A LOCATION	A CROSS_STREET	# LATITUDE	# LONGITUDE	A DL_SOURCE_ID	INF DLLOAD_DT
1	1	1	Central	101	00900 FIGUEROA	TR	34.0663	-118.2442	DATABRICKS_DLT	2025-11-23 03:41:08.919
2	2	1	Central	101	1000 ALPINE	ST	34.0649	-118.2465	DATABRICKS_DLT	2025-11-23 03:41:08.919
3	3	1	Central	101	1000 BARTLETT	ST	34.0646	-118.2474	DATABRICKS_DLT	2025-11-23 03:41:08.919
4	4	1	Central	101	1000 EVERETT	PL	34.0693	-118.2488	DATABRICKS_DLT	2025-11-23 03:41:08.919
5	5	1	Central	101	1000 EVERETT	ST	34.0701	-118.2484	DATABRICKS_DLT	2025-11-23 03:41:08.919
6	6	1	Central	101	1000 FIGUEROA	TR	34.0683	-118.245	DATABRICKS_DLT	2025-11-23 03:41:08.919
7	7	1	Central	101	1000 FIGUEROA TER		34.0684	-118.2449	DATABRICKS_DLT	2025-11-23 03:41:08.919
8	8	1	Central	101	1000 JOELS	PL	34.0696	-118.2462	DATABRICKS_DLT	2025-11-23 03:41:08.919
9	9	1	Central	101	1000 MARVIEW	AV	34.0698	-118.2476	DATABRICKS_DLT	2025-11-23 03:41:08.919
0	10	1	Central	101	1000 WHITE KNOLL	DR	34.0695	-118.2469	DATABRICKS_DLT	2025-11-23 03:41:08.919
1	11	1	Central	101	1000 W COLLEGE	ST	34.0667	-118.2452	DATABRICKS_DLT	2025-11-23 03:41:08.919
2	12	1	Central	101	1000 W SUNSET	BL	34.0685	-118.246	DATABRICKS_DLT	2025-11-23 03:41:08.919
3	13	1	Central	101	1100 ALPINE	ST	34.0662	-118.2476	DATABRICKS_DLT	2025-11-23 03:41:08.919
4	14	1	Central	101	1100 FIGUEROA	TR	34.0695	-118.2469	DATABRICKS_DLT	2025-11-23 03:41:08.919

```

45 | select * from premis_dim
46
47
48
49
50
51
52
53

```

Results (just now) 315 rows 197ms

Table **Chart**

#	PREMIS_SK	#	PREMIS_CODE	A PREMIS_DESC	INT_DLLOAD_DT	A DLSOURCE_ID
1	1	315	-1	Unknown RETIRED (DUPLICATE) DO NOT USE... +98 more	2025-11-22 03:41:18.329	DATABRICKS_DLT 100.0%
2	2		101	STREET	2025-11-23 03:41:18.329	DATABRICKS_DLT
3	3		102	SIDEWALK	2025-11-23 03:41:18.329	DATABRICKS_DLT
4	4		103	ALLEY	2025-11-23 03:41:18.329	DATABRICKS_DLT
5	5		104	DRIVEWAY	2025-11-23 03:41:18.329	DATABRICKS_DLT
6	6		105	PEDESTRIAN OVERCROSSING	2025-11-23 03:41:18.329	DATABRICKS_DLT
7	7		106	TUNNEL	2025-11-23 03:41:18.329	DATABRICKS_DLT
8	8		107	VACANT LOT	2025-11-23 03:41:18.329	DATABRICKS_DLT
9	9		108	PARKING LOT	2025-11-23 03:41:18.329	DATABRICKS_DLT
10	10		109	PARK/PLAYGROUND	2025-11-23 03:41:18.329	DATABRICKS_DLT
11	11		110	FREEWAY	2025-11-23 03:41:18.329	DATABRICKS_DLT
12	12		111	BUS-CHAPTER/PRIVATE	2025-11-23 03:41:18.329	DATABRICKS_DLT

```

6 | select * from victim_dim
7
8
9
0

```

Results (just now) 296 rows 53ms

Table **Chart**

#	VICTIM_SK	#	VICTIM_AGE	A VICTIM_AGE_CATEGORY	A VICTIM_SEX	INT_DLLOAD_DT	A DLSOURCE_ID
1	1	296	2	60+ Juvenile +4 more	F M +1 more	2025-11-22 03:41:15.290	DATABRICKS_DLT 100.0%
2	2		2	Juvenile	F	2025-11-23 03:41:15.290	DATABRICKS_DLT
3	3		2	Juvenile	M	2025-11-23 03:41:15.290	DATABRICKS_DLT
4	4		3	Juvenile	X	2025-11-23 03:41:15.290	DATABRICKS_DLT
5	5		3	Juvenile	F	2025-11-23 03:41:15.290	DATABRICKS_DLT
6	6		3	Juvenile	M	2025-11-23 03:41:15.290	DATABRICKS_DLT
7	7		4	Juvenile	X	2025-11-23 03:41:15.290	DATABRICKS_DLT
8	8		4	Juvenile	F	2025-11-23 03:41:15.290	DATABRICKS_DLT
9	9		4	Juvenile	M	2025-11-23 03:41:15.290	DATABRICKS_DLT
10	10		5	Juvenile	X	2025-11-23 03:41:15.290	DATABRICKS_DLT
			5	Juvenile	F	2025-11-23 03:41:15.290	DATABRICKS_DLT

```

60 | select * from weapon_dim
61
62
63
64
65
66
67
68

```

Results (just now) 80 rows 139ms

Table **Chart**

#	WEAPON_SK	#	WEAPON_USED_CODE	A WEAPON_DESC	INT_DLLOAD_DT	A DLSOURCE_ID
1	1	80	-1	AIR PISTOL/REVOLVER/RIFLE/BB... ANTIQUE FIREARM +78 more	2025-11-22 03:41:05.782	DATABRICKS_DLT 100.0%
2	2		101	Unknown	2025-11-23 03:41:05.782	DATABRICKS_DLT
3	3		102	REVOLVER	2025-11-23 03:41:05.782	DATABRICKS_DLT
4	4		103	HAND GUN	2025-11-23 03:41:05.782	DATABRICKS_DLT
5	5		104	RIFLE	2025-11-23 03:41:05.782	DATABRICKS_DLT
6	6		105	SHOTGUN	2025-11-23 03:41:05.782	DATABRICKS_DLT
7	7		106	SAWED OFF RIFLE/SHOTGUN	2025-11-23 03:41:05.782	DATABRICKS_DLT
8	8		107	UNKNOWN FIREARM	2025-11-23 03:41:05.782	DATABRICKS_DLT
9	9		108	OTHER FIREARM	2025-11-23 03:41:05.782	DATABRICKS_DLT
10	10		109	AUTOMATIC WEAPON/SUB-MACHINE GUN	2025-11-23 03:41:05.782	DATABRICKS_DLT
11	11		110	SEMI-AUTOMATIC PISTOL	2025-11-23 03:41:05.782	DATABRICKS_DLT
12	12		111	SEMI-AUTOMATIC RIFLE	2025-11-23 03:41:05.782	DATABRICKS_DLT
13	13		112	STARTER PISTOL/REVOLVER	2025-11-23 03:41:05.782	DATABRICKS_DLT
14	14		113	TOY GUN	2025-11-23 03:41:05.782	DATABRICKS_DLT
			114	SIMULATED GUN	2025-11-23 03:41:05.782	DATABRICKS_DLT

Results (just now)

Table | Chart

Q D 1,004,991 rows 69ms ⚡ ↻

#	# CRIME_KEY	# DAYS_TO_REPORT	FILE_NUMBER	# PREMIS_SK	# WEAPON_SK	# CRIME_SK	TIME_KEY	# ARREST_SK	# LOCATION_ID	# VICTIM_SK	# DATE_SK	# DESCENT_SK
1	1	0	010304468	132	63	2230		2	18922	103	20200108	2
2	2	1	0817	2	1	58	1700	4	121389	296	20200919	18
3	3	1	190101086	3	64	63	0330	4	4328	71	20200101	7
4	4	1	190101087	133	63	65	0510	1	3335	164	20200101	2
5	5	0	190326475	2	1	58	2130	1	48136	296	20200301	12
6	6	0	191501505	133	1	82	1730	4	106019	223	20200101	17
7	7	0	191921269	122	1	81	0415	4	137229	90	20200101	18
8	8	1	200100001	9	1	58	2000	1	275	296	20200225	18
9	9	1	200100002	2	1	58	1740	2	4759	296	20200814	18
10	10	1	200100003	2	1	58	2200	4	4186	296	20200814	18
11	11	1	200100005	2	1	58	2045	2	6237	296	20201205	18
12	12	1	200100501	193	64	3	0030	4	4327	70	20200101	7
13	13	0	200100502	117	1	41	1315	4	4036	65	20200102	7
14	14	0	200100504	184	1	134	0040	4	3253	296	20200104	18
15	15	0	200100507	133	1	18	0200	4	36	65	20200104	2
16	16	0	200100508	3	63	65	0900	2	679	179	20200104	17
17	17	0	200100509	2	56	16	2200	4	5693	83	20200104	1
18	18	0	200100510	9	75	126	0955	4	333	101	20200105	12