

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.

Alteryx Designer x64 - Crime_Data_Profiling.ymd

File Edit View Options Add-Ins Help

Input Data (1) - Configuration

Connect a File or Database
C:\Users\Bhuvana\Downloads\Crime_Data_from_2020_to_Present_20251123.csv
Set Up a Connection

Options

Name	Value
1 Record Limit	10
2 File Format	Comma Separated Value (*.csv)
3 Search SubDirs	No
4 Output File Name as Field	No
5 Delimiters	,

Preview (first 100 records)

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	Wilshire	0762
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	240214110	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	201405870	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

Crime_Data_Profiling.ymd

Results - Input Data (1) - Output

Record	DR_NO	Date Rptd	DATE OCC	TIME OCC	AREA	AREA NAME
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 Oct 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	240214110	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 Jul 02 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

Alteryx Designer x64 - New Workflow

File Edit View Options Add-Ins Help

Input Data (1) - Configuration

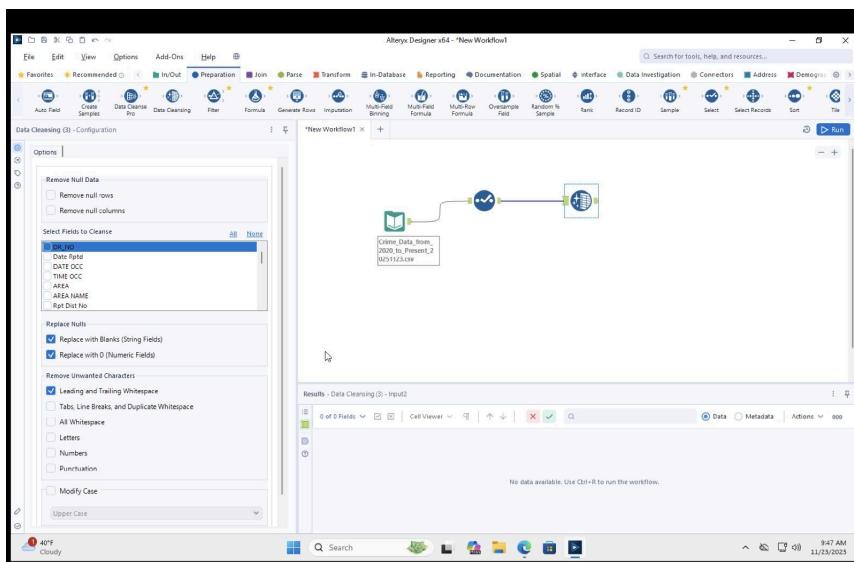
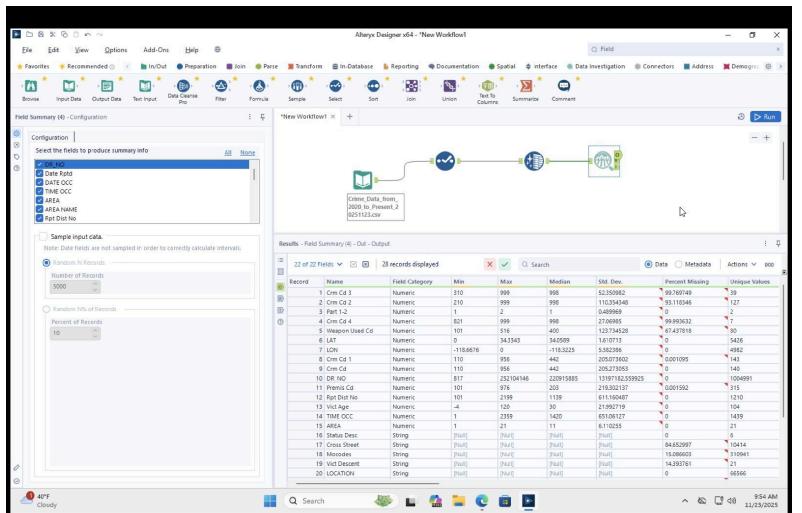
Options

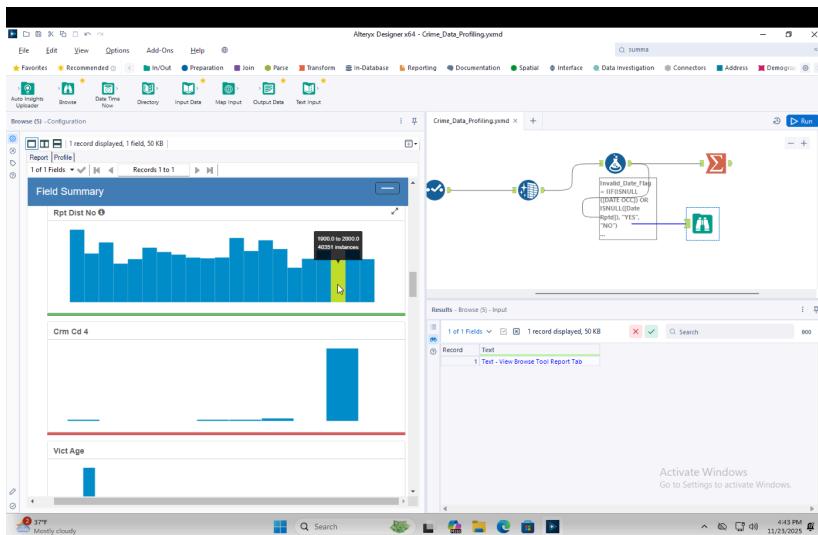
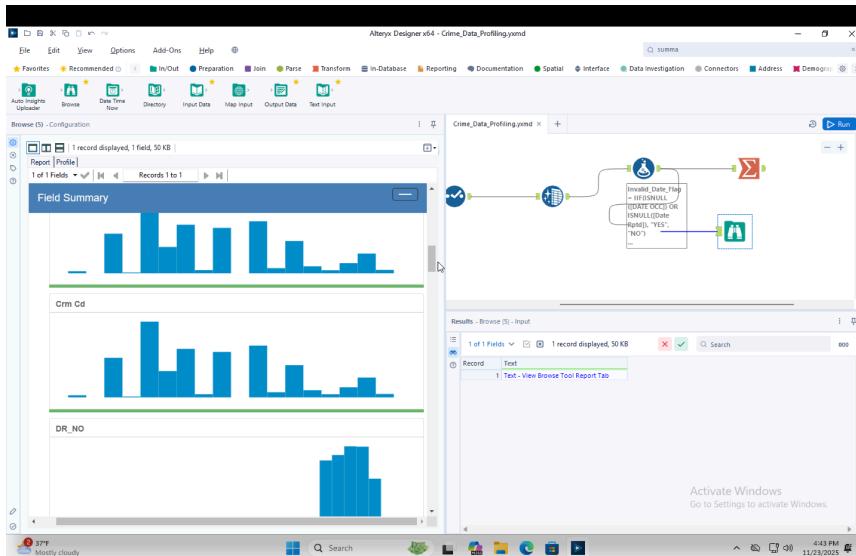
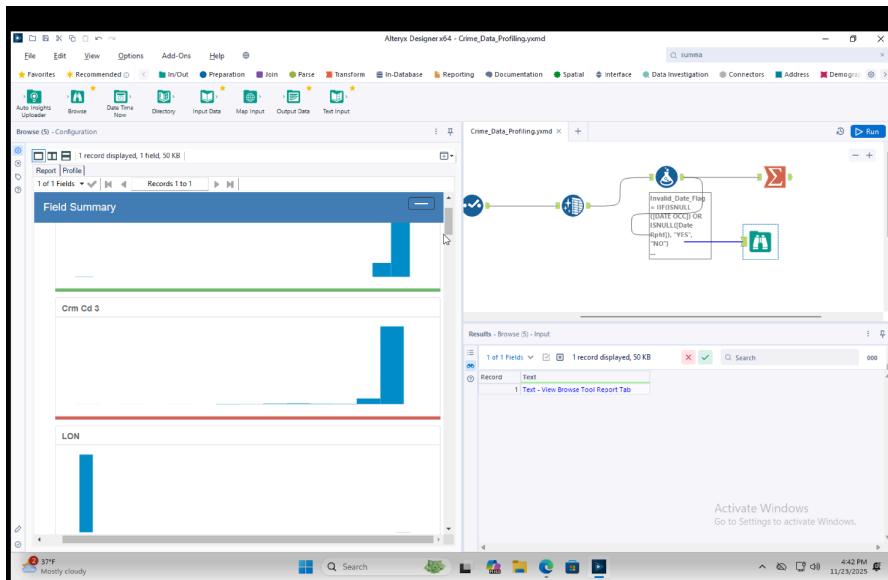
Column	Type	Size	Rename	Description
DR_NO	Int64	8		
Date Rptd	Date	10		
DATE OCC	Date	10		
TIME OCC	Date	10		
AREA	Int64	8		
AREA NAME	V_String	254		
Rpt Dist No	Int64	8		
Part 1-2	Int64	8		
Crm Cd	Int64	8		
Crm Cd Desc	V_String	254		
Motodes	V_String	254		
Vict Age	Int64	8		
Vict Sex	V_String	254		
Vict Descent	V_String	254		
Previous Dw	V_String	254		
Previous Dw Desc	V_String	254		
Weapon Used Cd	Int64	8		
Weapon Des.	V_String	254		
Status	V_String	254		
Status Desc	V_String	254		
Crm Cd 1	Int64	8		
Crm Cd 2	Int64	8		
Crm Cd 3	Int64	8		

Results - Select (2) - Output

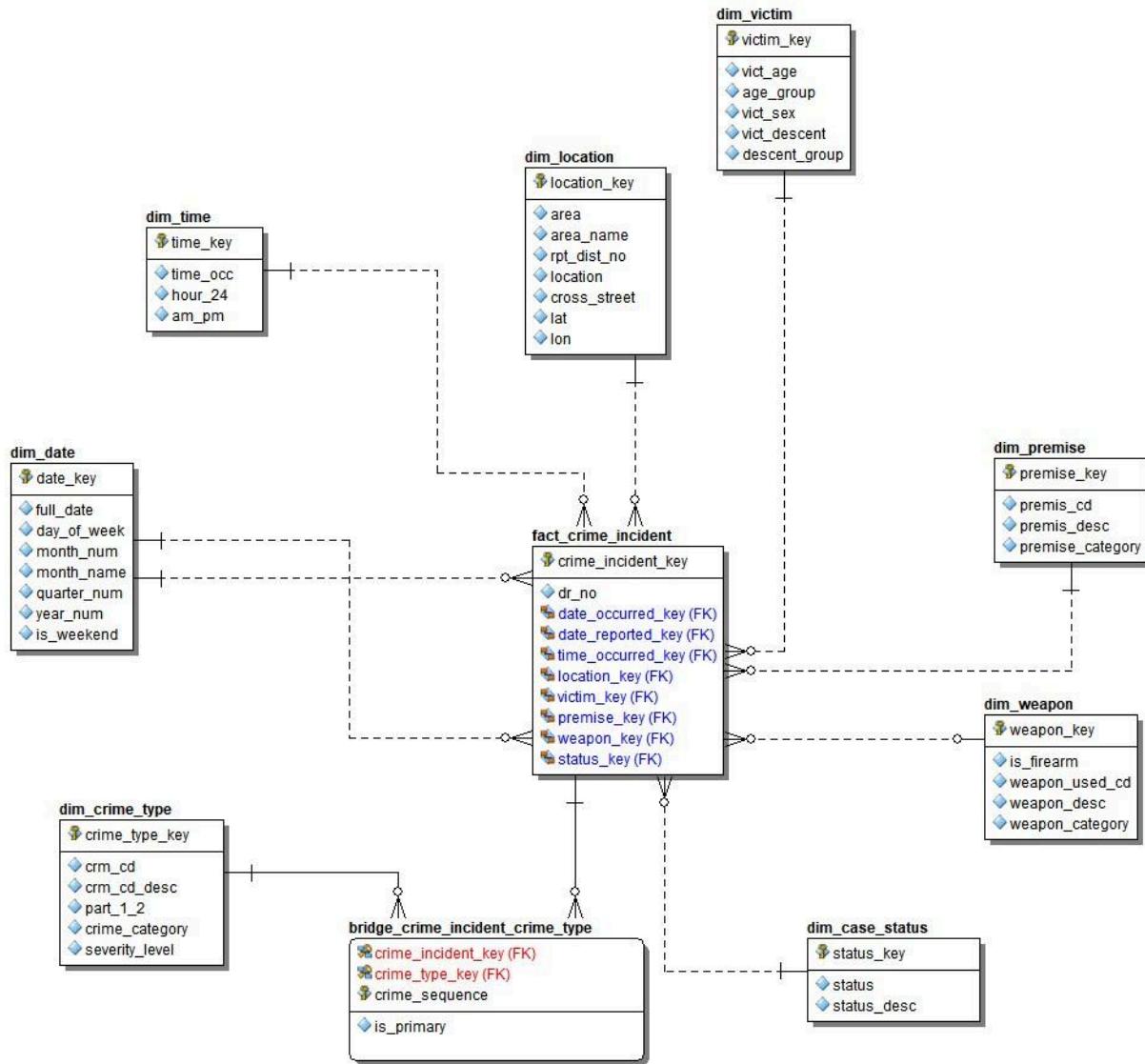
0 of 0 Fields	Get Viewer	Run	Data	Metadata	Addons
---------------	------------	-----	------	----------	--------

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





Data Modelling



Professor Feedback & Model Revisions

1. Consolidate Unnecessary Dimensions

Professor Feedback: "The reporting district dimension is unnecessary complexity. It should be consolidated into the location dimension since reporting districts are subsets of geographic areas."

Change Made:

- **Removed:** Separate DIM_REPORTING_DISTRICT table
- **Consolidated into:** DIM_LOCATION which now includes area, area_name, rpt_dist_no, location, cross_street, lat, lon

Rationale: Reporting districts are always part of a larger geographic area. Having a separate dimension adds unnecessary joins without providing analytical value. Users naturally think of crime locations hierarchically (Area → Reporting District → Specific Location).

2. Simplify Date Hierarchy

Professor Feedback: "You don't need separate dimension tables for date components. A single date dimension with attributes is the standard approach and sufficient for all temporal analysis."

Change Made:

- **Removed:** Multiple date hierarchy tables
- **Created:** Single DIM_DATE with all temporal attributes: full_date, day_of_week, month_num, month_name, quarter_num, year_num, is_weekend

Rationale: A single date dimension is industry standard (Kimball methodology). All date-related queries can use this one dimension. Separate tables for month, quarter, year create unnecessary complexity and slower queries.

3. Add Bridge Table for Multiple Crimes

Professor Feedback: "Your fact table has multiple crime code columns (CRM_CD_1 through CRM_CD_4) which violates normalization principles. Use a bridge table to properly model the many-to-many relationship between incidents and crime types."

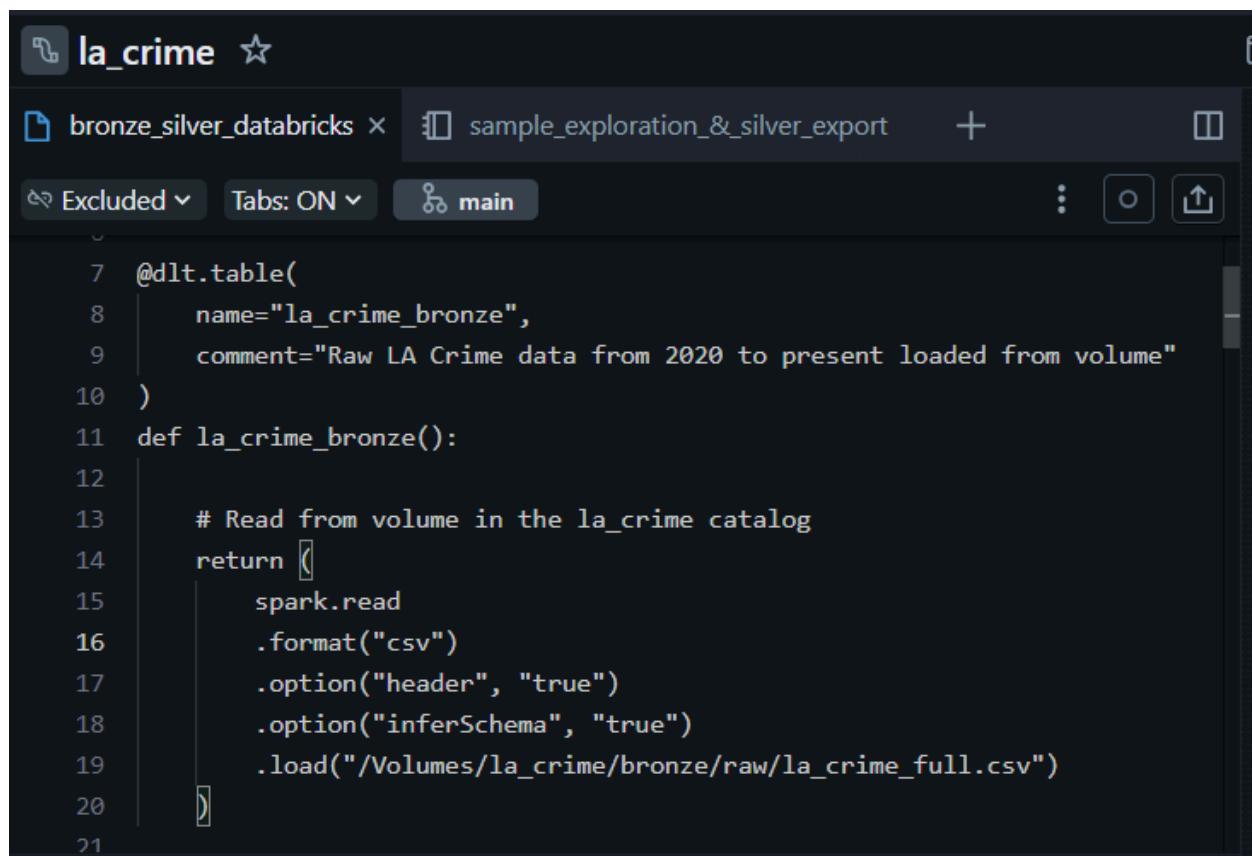
Change Made:

- **Added:** BRIDGE_CRIME INCIDENT_CRIME_TYPE table
- **Attributes:** crime_incident_key (FK), crime_type_key (FK), crime_sequence, is_primary
- **Removed:** Multiple crime code columns from fact table (kept only primary crime reference)

Rationale: Bridge tables are the proper way to handle many-to-many relationships in dimensional models. This allows an incident to have multiple associated crimes while maintaining referential integrity. The crime_sequence indicates order, and is_primary flags the main crime.

Medallion Architecture

Bronze



The screenshot shows a Databricks notebook titled "la_crime". The notebook interface includes a header with tabs for "bronze_silver_databricks" and "sample_exploration_&_silver_export", and a sidebar with "Excluded" and "Tabs: ON" buttons. The main area contains the following Python code:

```
7 @dlt.table(
8     name="la_crime_bronze",
9     comment="Raw LA Crime data from 2020 to present loaded from volume"
10 )
11 def la_crime_bronze():
12
13     # Read from volume in the la_crime catalog
14     return (
15         spark.read
16             .format("csv")
17             .option("header", "true")
18             .option("inferSchema", "true")
19             .load("/Volumes/la_crime/bronze/raw/la_crime_full.csv")
20     )
21 
```

Purpose:

The bronze layer is designed to ingest the **raw LA crime dataset** from 2020 to the present, preserving the source data exactly as it exists. This layer acts as the **foundation for all downstream transformations** and analytics.

Process / ETL Steps:

1. Ingestion:

- The CSV files are loaded from `/Volumes/la_crime/bronze/raw/la_crime_full.csv` using Spark.
- Schema is inferred automatically (`.option("inferSchema", "true")`).
- Headers are preserved to maintain column names.

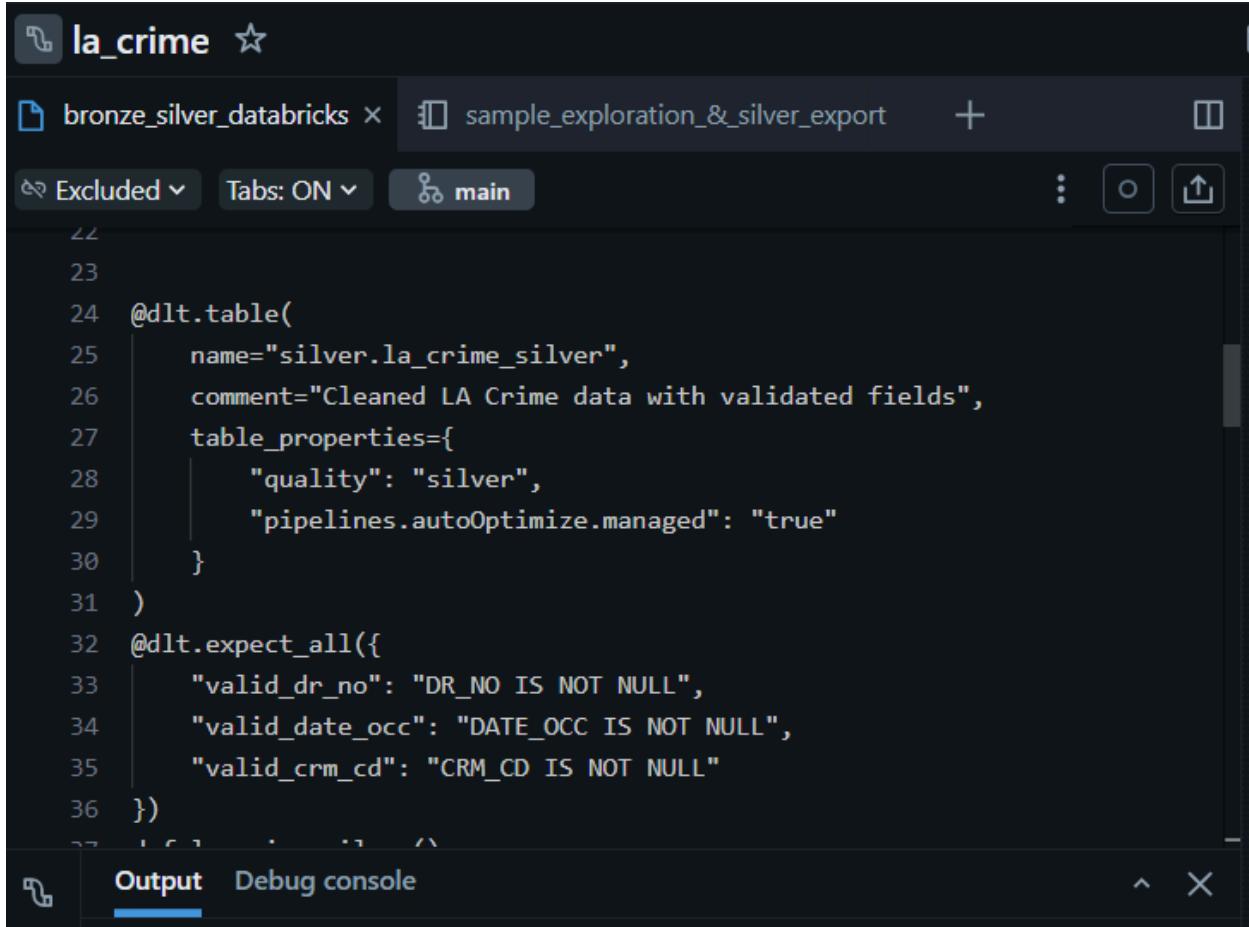
2. No transformations applied:

- All data is retained in its original form, including whitespaces, inconsistent casing, and invalid or missing values.
- This ensures **traceability** and a historical record of all ingested raw data.

3. Usage:

- Serves as a source for the **silver layer**, where data cleaning and validation occurs.

SILVER



The screenshot shows a Databricks notebook titled "la_crime". The code in the notebook is as follows:

```
22
23
24 @dlt.table(
25     name="silver.la_crime_silver",
26     comment="Cleaned LA Crime data with validated fields",
27     table_properties={
28         "quality": "silver",
29         "pipelines.autoOptimize.managed": "true"
30     }
31 )
32 @dlt.expect_all({
33     "valid_dr_no": "DR_NO IS NOT NULL",
34     "valid_date_occ": "DATE_OCC IS NOT NULL",
35     "valid_crm_cd": "CRM_CD IS NOT NULL"
36 })
```

The notebook has tabs for "bronze_silver_databricks" and "sample_exploration_&_silver_export". The current tab is "main". The bottom navigation bar shows "Output" (which is selected) and "Debug console".

Purpose:

The silver layer transforms raw data into a **clean, validated, and standardized dataset** that can be reliably used for reporting and analytics. It ensures **data quality** while preserving the integrity of the original data.

Process / ETL Steps:

1. **Read Bronze Data:**
 - The silver table reads directly from `la_crime_bronze`.
2. **Field Cleaning and Standardization:**
 - Trim leading/trailing spaces from string fields (`DR_NO`, `CRM_CD_DESC`, `MOCODES`, `PREMIS_DESC`, `WEAPON_DESC`, `LOCATION`, `CROSS_STREET`).

- **Uppercase** key categorical fields (`AREA_NAME`, `VICT_SEX`, `VICT_DESCENT`, `STATUS`).
- **Normalize spacing** in the `LOCATION` field using regex to replace multiple spaces with a single space.

3. Validation of Critical Fields:

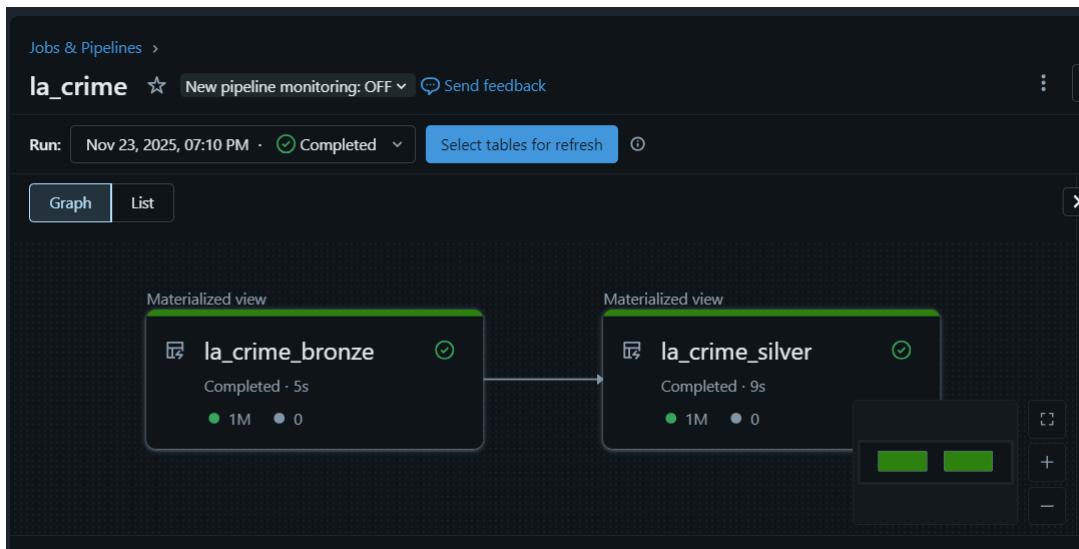
- Ensure `DR_NO`, `DATE_OCC`, and `CRM_CD` are not null (@`dlt.expect_all`).
- Remove records missing these essential identifiers.

4. Data Type and Value Corrections:

- `VICT_AGE` is set to NULL if the value is non-positive.
- Latitude and longitude are set to NULL if both are zero, preventing invalid coordinates.

5. Additional Quality Controls:

- Automatically managed table optimization (`pipelines.autoOptimize.managed=true`) ensures **performance improvements** for queries.
- Silver layer acts as a **trusted dataset**, ready for consumption by gold-layer aggregations and analytics.



GOLD

```
1 import snowflake.snowpark as snowpark
2 from snowflake.snowpark.functions import col, concat, lpad, year, month, dayofweek, when, lit, row_number, monotonically_increasing_id
3 from snowflake.snowpark.window import Window
4
5 def main(session: snowpark.Session):
6     """Create simplified Gold dimensional model with essential columns only"""
7
8     print("Creating Simplified Gold Layer Dimensional Model...")
9     silver_df = session.table("SILVER.LA_CRIME_SILVER")
10    print(f"Silver records: {silver_df.count()}")
11    print(f"Silver columns: {silver_df.columns}")
12
13    # ===== DIM_DATE =====
14    print("\nCreating DIM_DATE (simplified)...")
15    dates_distinct = silver_df.select(col("DATE_OCC")).distinct().filter(col("DATE_OCC").isNotNull())
16
17    dim_date = dates_distinct.select(
18        to_char(col("DATE_OCC"), "YYYYMMDD").cast("integer").alias("DATE_KEY"),
19        col("DATE_OCC").alias("FULL_DATE"),
20        dayofweek(col("DATE_OCC")).alias("DAY_OF_WEEK"),
21        month(col("DATE_OCC")).alias("MONTH_NUM"),
22        when(month(col("DATE_OCC")).in_([1,2,3]), lit("Q1")).when(month(col("DATE_OCC")).in_([4,5,6]), lit("Q2")),
23        .when(month(col("DATE_OCC")).in_([7,8,9]), lit("Q3")).otherwise(lit("Q4")).alias("QUARTER_NUM"),
24        year(col("DATE_OCC")).alias("YEAR_NUM")
25    )
26    dim_date.write.mode("overwrite").save_as_table("GOLD.DIM_DATE")
27    print(f"    ✓ Created: {dim_date.count()} records")
28
29    # Create DIM_DATE for DATE_RPTD as well
30    print("\nCreating DIM_DATE entries for DATE_RPTD...")
31    dates_rpt_distinct = silver_df.select(col("DATE_RPTD")).distinct().filter(col("DATE_RPTD").isNotNull())
```

The CSV exported from Databricks Silver is uploaded to a Snowflake table stage using the [PUT](#) command (`PUT file:///path/to/la_crime_silver.csv @%LA_CRIME_SILVER;`) before loading it into the `SILVER.LA_CRIME_SILVER` table for Gold-layer processing.

The **Gold layer** represents the **curated, analytics-ready dimensional model** of LA crime data, derived from the Silver layer. It follows a **star schema** design with fact and dimension tables, suitable for **BI reporting, dashboards, and analytics**.

This Python Snowflake notebook builds **simplified dimension and fact tables** to support streamlined analysis while maintaining traceability to the Silver data.

Process / ETL Steps

1. Read Silver Layer

- Data is read from `SILVER.LA_CRIME_SILVER` using Snowpark.
- Key validation checks and initial record counts are performed.

2. Dimension Tables Creation

- **DIM_DATE**: Contains unique dates for DATE_OCC and DATE_RPTD. Derived columns include:
 - DATE_KEY (YYYYMMDD as integer)
 - FULL_DATE
 - DAY_OF_WEEK, MONTH_NUM, QUARTER_NUM, YEAR_NUM
 - Duplicate dates from reporting and occurrence dates are merged.
- **DIM_TIME**: Simplifies TIME_OCC into:
 - TIME_KEY (integer)
 - HOUR_24, MINUTE, AM_PM
- **DIM_LOCATION**: Aggregates by AREA and RPT_DIST_NO:
 - Location name, cross street, latitude, longitude
 - Surrogate LOCATION_KEY assigned with row_number()
- **DIM_CRIME_TYPE**: Maps CRM_CD and description, and categorizes crimes:
 - CRIME_CATEGORY (e.g., Theft, Burglary)
 - SEVERITY_LEVEL (High, Medium, Low)
 - Surrogate CRIME_TYPE_KEY
- **DIM_VICTIM**: Standardizes victim attributes:
 - AGE_GROUP buckets
 - DESCENT_GROUP mapping (e.g., Asian, Hispanic)
 - Surrogate VICTIM_KEY

- **DIM_PREMISE**: Standardizes premises descriptions and categories:
 - Surrogate **PREMISE_KEY**
 - **DIM_WEAPON**: Categorizes weapons and flags firearms:
 - Surrogate **WEAPON_KEY**
 - **DIM_CASE_STATUS**: Unique case statuses with **STATUS_KEY**.
-

3. Fact Table Creation

- **FACT_CRIME INCIDENT**: Core fact table linking:
 - **DR_NO, DATE_OCC, DATE_RPTD, TIME_OCC, AREA, VICTIM, PREMISE, WEAPON, STATUS**
 - Joins with dimension tables to replace raw attributes with **surrogate keys**.
 - Fact table uses `row_number()` to generate **CRIME INCIDENT KEY**.
-

4. Bridge Table

- **BRIDGE_CRIME INCIDENT CRIME TYPE** handles **multiple crimes per incident**:
 - Maps **DR_NO** to multiple **CRM_CD** fields (**CRM_CD, CRM_CD_1 ...4**)
 - Assigns sequence and primary/secondary flags
 - Surrogate keys from **DIM_CRIME_TYPE**

```
39  
40 | show tables;
```

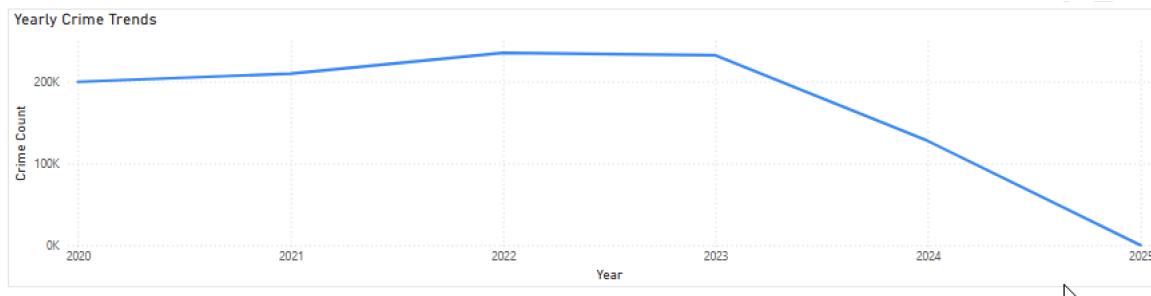
↳ Results ⚡ Chart

	created_on	name	database_name	schema_name	# rows	# bytes
1	2025-11-23 21:16:36.190 -0800	BRIDGE_CRIME INCIDENT_CRIME_TYPE	LA_CRIME	GOLD	2023753	7951872
2	2025-11-23 21:16:24.997 -0800	DIM_CASE_STATUS	LA_CRIME	GOLD	6	9216
3	2025-11-23 21:16:21.814 -0800	DIM_CRIME_TYPE	LA_CRIME	GOLD	140	24064
4	2025-11-23 21:16:19.187 -0800	DIM_DATE	LA_CRIME	GOLD	1901	29184
5	2025-11-23 21:16:20.958 -0800	DIM_LOCATION	LA_CRIME	GOLD	1210	53248
6	2025-11-23 21:16:23.474 -0800	DIM_PREMISE	LA_CRIME	GOLD	314	21504
7	2025-11-23 21:16:20.164 -0800	DIM_TIME	LA_CRIME	GOLD	1439	24576
8	2025-11-23 21:16:22.631 -0800	DIM_VICTIM	LA_CRIME	GOLD	2937	33280
9	2025-11-23 21:16:24.225 -0800	DIM_WEAPON	LA_CRIME	GOLD	80	12432

Interpretation of LA_CRIME Dashboards

Crime Rates Over Time

Yearly Crime Trends



- Crime rates showed a steady increase from 2020 to 2022, peaking at approximately 220,000 incidents in 2022
- A **sharp decline began in late 2023**, dropping to around 50,000 incidents by 2025, representing a **77% decrease** from the peak
- This significant downward trend suggests effective crime prevention measures or changes in reporting practices in recent years

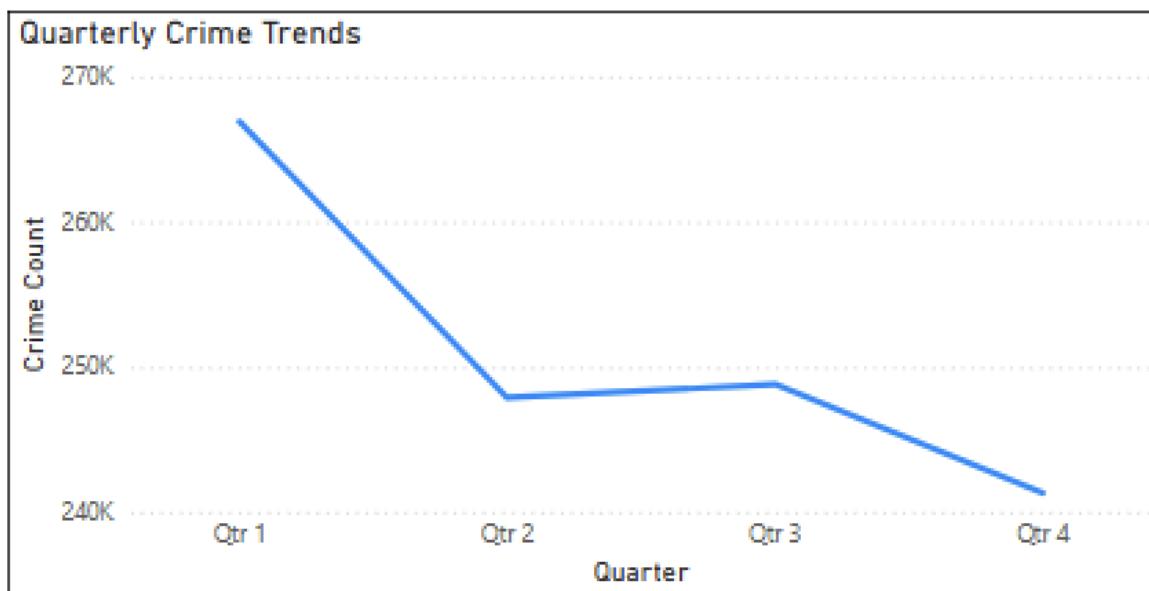
Monthly Crime Trends



- **January experiences the highest crime rate** (~93,000 incidents), while **December has the lowest** (~78,000 incidents), showing a **16% decrease** throughout the year

- Crime rates demonstrate a **consistent downward seasonal pattern**, with a notable drop from January to February, followed by fluctuations between 81,000-88,000 incidents.
- The data suggests **winter months (particularly early January) are peak crime periods**, with crimes declining as the year progresses toward the holidays

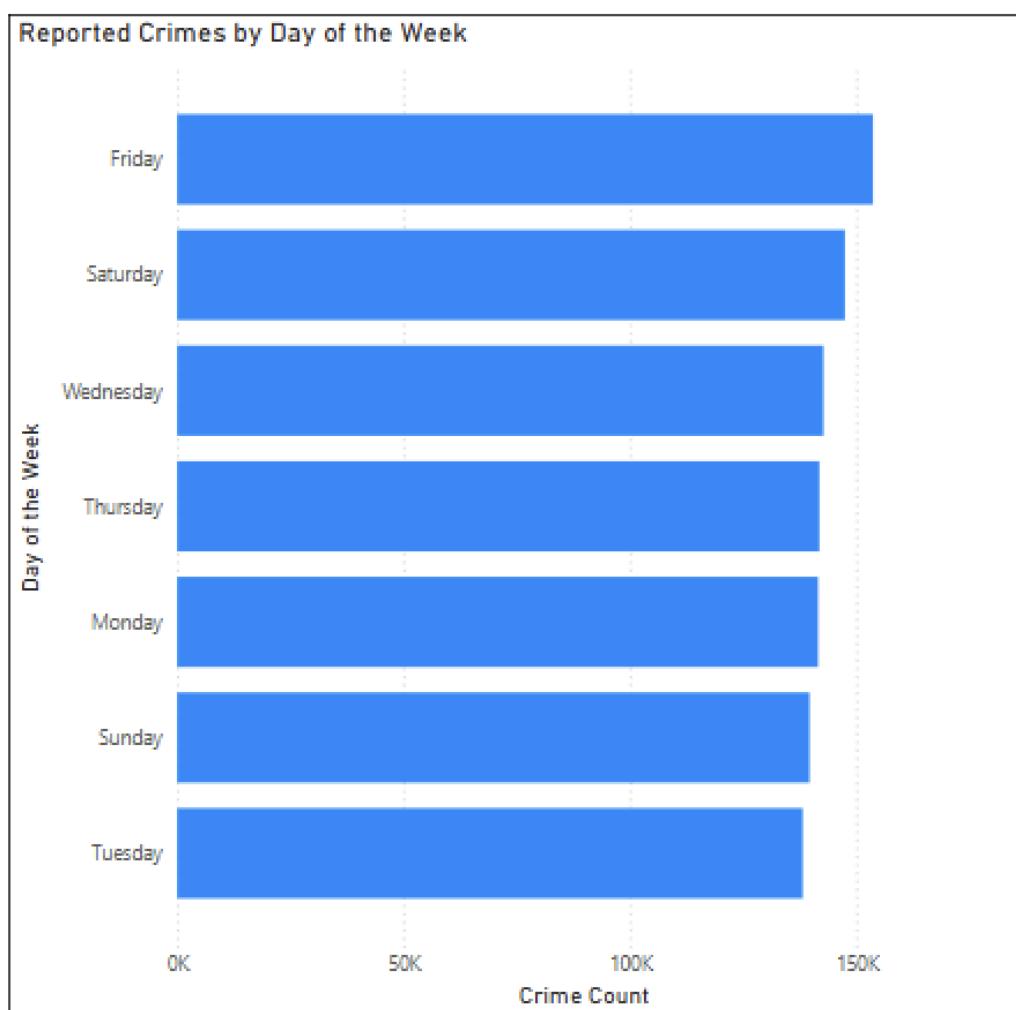
Quarterly Time Trends



- **Q1 shows the highest crime activity** at 270,000 incidents, followed by a **sharp 8% drop** to Q2 (248,000 incidents).
- **Q2 and Q3 remain relatively stable** (248,000-249,000 incidents), indicating consistent mid-year crime levels.
- **Q4 experiences a decline** to 242,000 incidents, confirming that crime rates decrease toward the end of the calendar year, with the most significant decrease occurring between Q1 and Q2.

Day and Time Factors

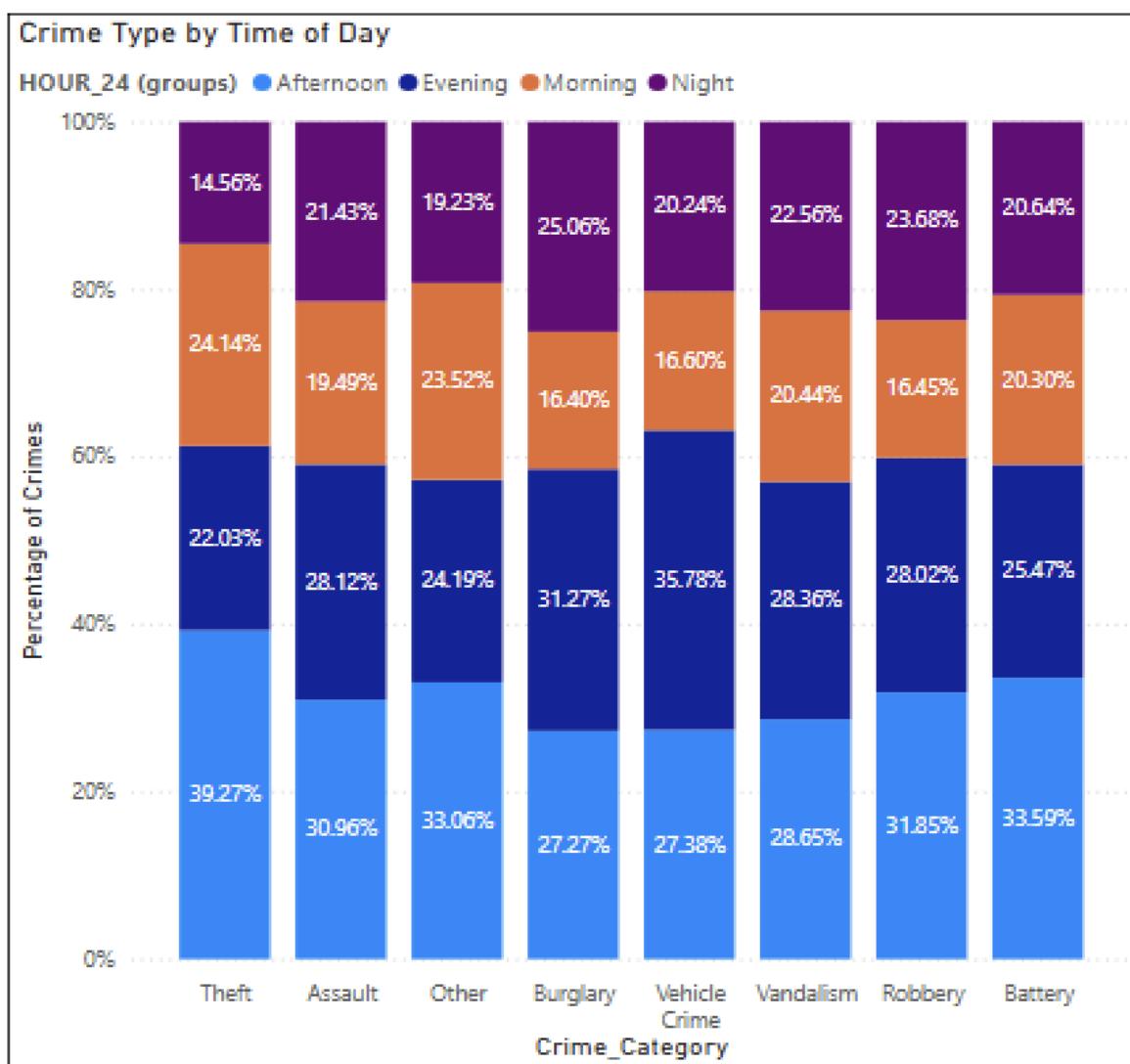
Reported Crimes by Day of the Week



- There is a clear correlation between day of week and crime frequency. **Friday experiences the highest crime rate** at approximately 155,000 incidents, while **Tuesday has the lowest** at around 135,000 incidents, representing a **15% difference**.
- Weekend and end-of-week patterns emerge, with Friday and Saturday showing elevated crime levels, suggesting increased criminal activity as the work week ends and weekend activities begin.

- Mid-week days (Tuesday through Thursday) demonstrate more consistent and lower crime rates, indicating that crime activity varies significantly based on weekly patterns and social behaviors

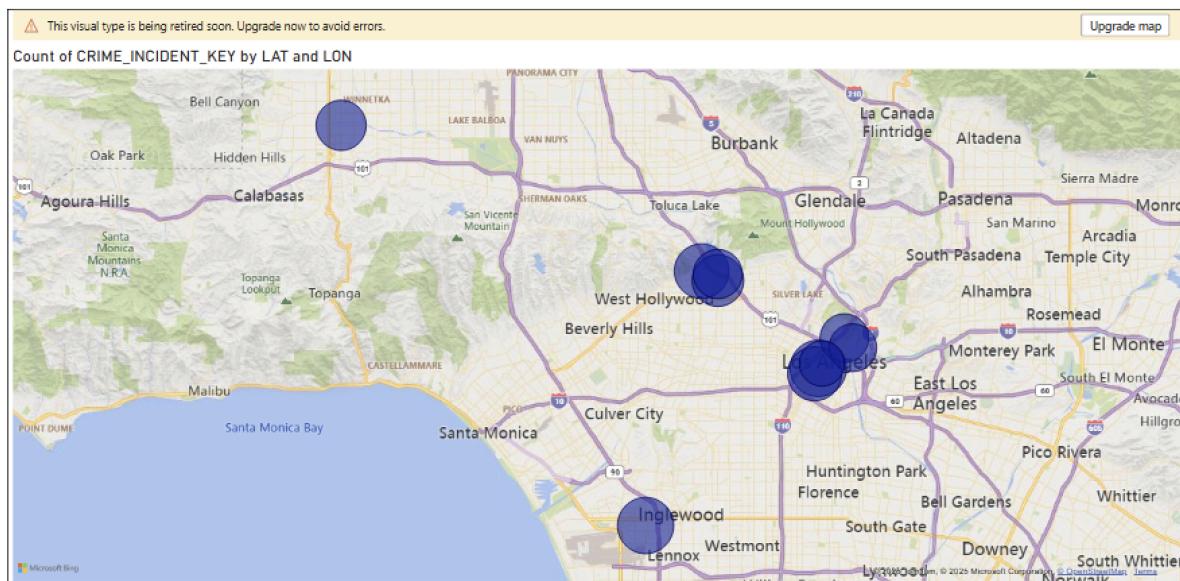
Crime Type by Time of the Day



- Theft crimes peak during afternoon hours (39.27%), making it the predominant daytime crime, while showing relatively balanced distribution across other time periods, suggesting opportunistic property crimes when people are active.

- **Burglary shows a distinct nighttime pattern** with 31.27% occurring at night and only 16.40% in the morning, indicating criminals target properties during dark hours when detection risk is lower
- **Violent crimes (Assault, Battery) are distributed more evenly across all time periods**, with slight increases during evening and night hours (Battery: 33.59% afternoon, 25.47% evening), while **vehicle-related crimes concentrate in afternoon and evening hours** (35.78% evening) when parking lots and streets are busiest.

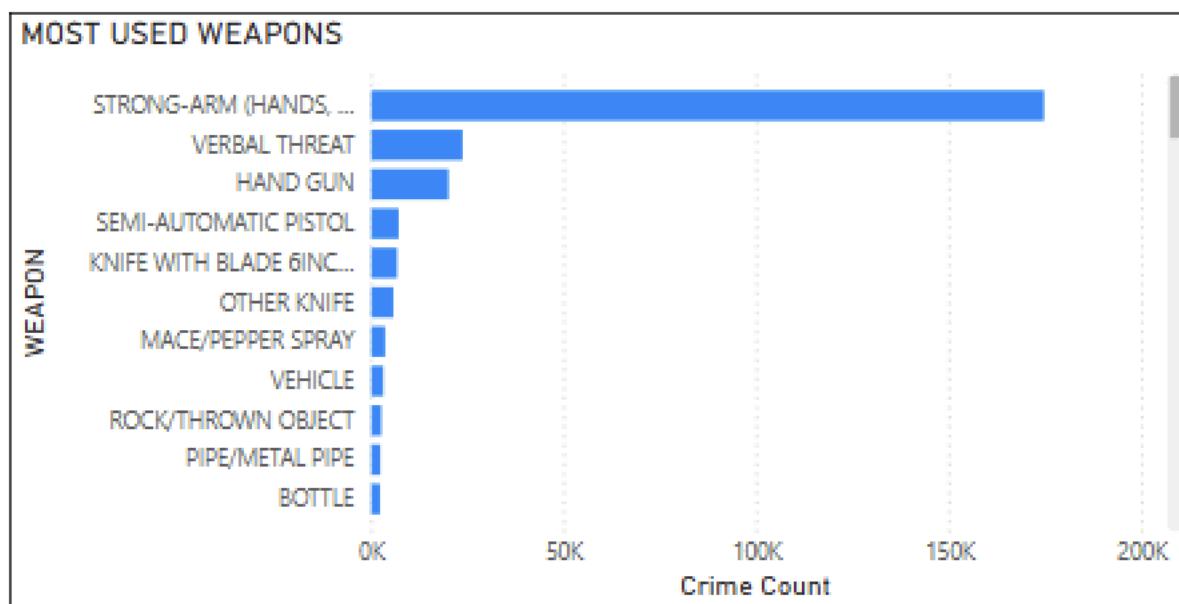
Crime Hotspots by Location



- Central Los Angeles represents the most significant crime hotspot, with the largest concentration of incidents shown by the dominant bubble in the downtown/central area. This urban core experiences substantially higher crime rates compared to all other areas, likely due to higher population density, commercial activity, and socioeconomic factors.
- Secondary hotspots are concentrated in West Hollywood and the Hollywood area, forming a corridor of elevated crime activity through central-west LA. Inglewood (south) also shows notable crime concentration, indicating that crime patterns extend beyond just the downtown core to include entertainment districts and densely populated neighborhoods.
- **The San Fernando Valley (northern area) and coastal regions (Santa Monica, Malibu) demonstrate significantly lower crime rates**, as indicated by smaller or

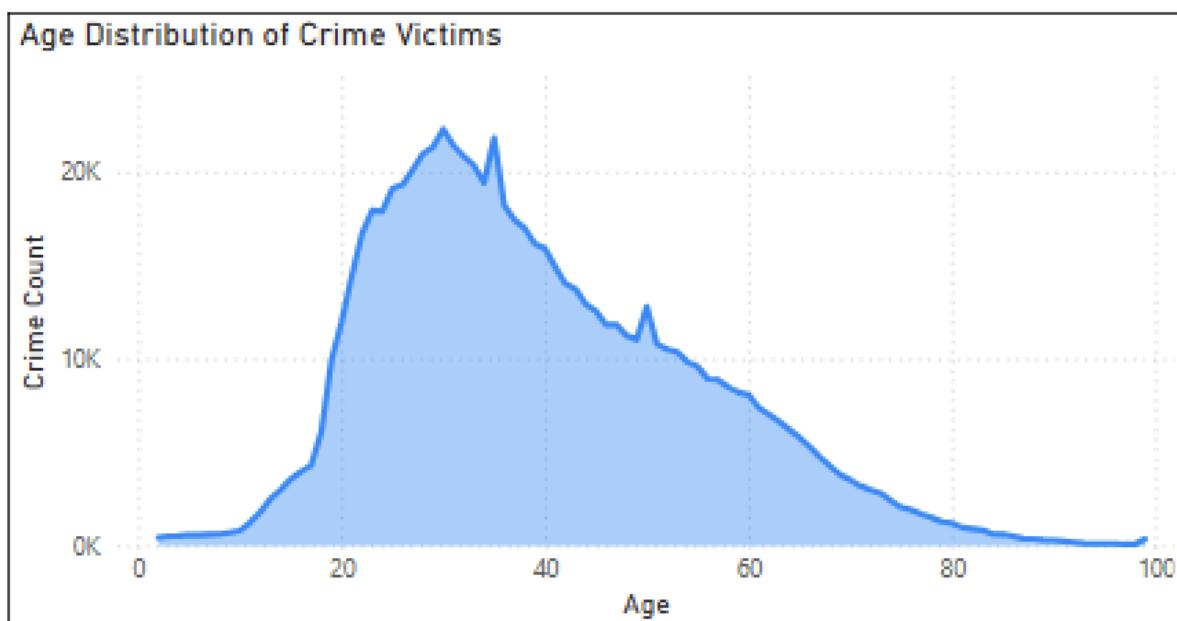
absent bubbles. Similarly, **outlying areas like Pasadena, Glendale, and eastern suburbs show better performance** with minimal crime concentrations, suggesting that distance from the urban core and different community characteristics correlate with lower crime rates.

Most Used Weapons



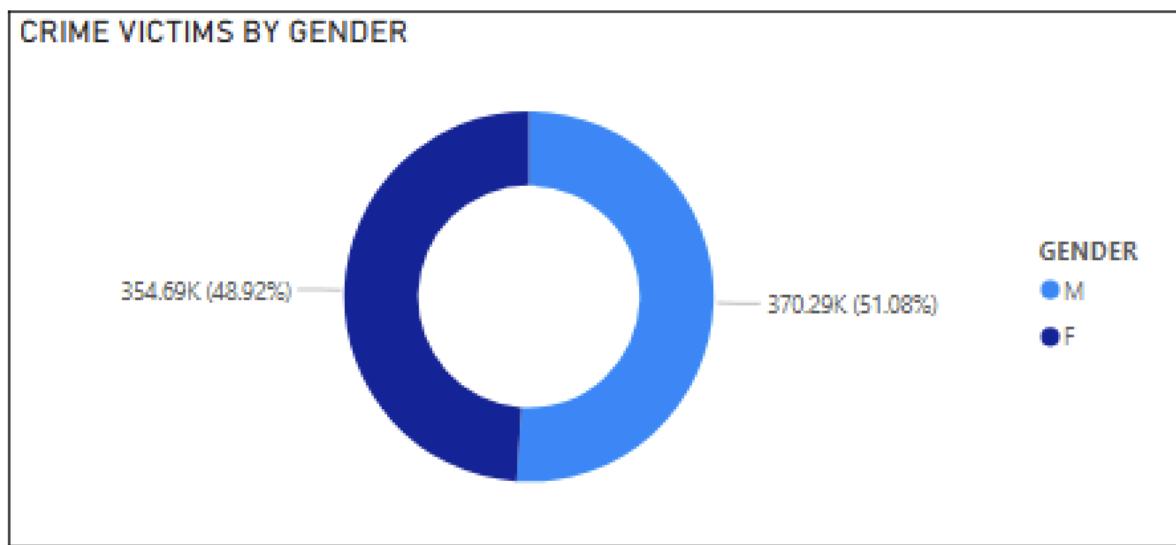
- "**Strong-arm" (hands, fists, feet) overwhelmingly dominates as the most common "weapon"** with approximately 200,000 incidents, representing the majority of crimes. This indicates that **most violent crimes in LA involve physical force rather than actual weapons**, suggesting a prevalence of simple assaults and confrontations
- **Verbal threats rank second** (~50,000 incidents), followed by **handguns** (~40,000 incidents) as the most common actual weapon used in crimes, highlighting that when weapons are involved, firearms are the primary choice
- **A diverse range of weapons follows** including semi-automatic pistols, knives, mace/pepper spray, and improvised weapons (vehicles, rocks, pipes), but each represents a relatively small proportion of total incidents. This distribution indicates that while weapon variety exists, **unarmed physical altercations constitute the primary crime method** in Los Angeles

Age Distribution of Crime Victims



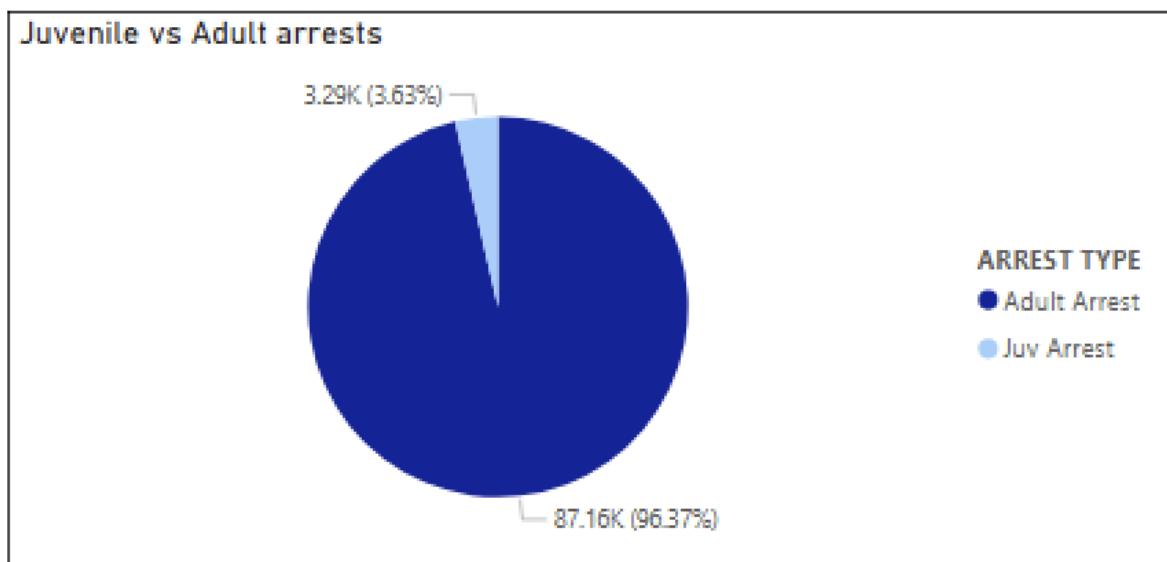
- Crime victimization follows a distinct bell curve pattern, peaking between ages **25-35** with approximately 20,000-22,000 incidents per age year. This demographic represents young to middle-aged adults who are most likely to be active in public spaces, nightlife, and economic activities where crimes commonly occur.
- Victimization is extremely low for minors (ages 0-17) and begins rising sharply in late teens, suggesting that **exposure to crime risk increases significantly as individuals enter adulthood** and engage more independently in society.
- A steady decline occurs after age **40**, with senior citizens (70+) experiencing the lowest victimization rates. The secondary smaller peak around age 50 may reflect continued workplace and community exposure, but overall, **older adults face considerably lower crime risk** than their younger counterparts.

Crime Victims by Gender



- Crime victimization is nearly equally distributed between genders, with males accounting for 51.08% (370,290 victims) and females 48.92% (354,690 victims), representing only a **2.16 percentage point difference**.
- The slight male predominance may reflect higher exposure to certain crime types such as street crimes, robbery, or violence in public spaces where males may be statistically more present.
- This **near-equal distribution challenges common assumptions** about crime victimization being heavily gender-skewed, indicating that both men and women in Los Angeles face similar overall crime risks, though the types of crimes they experience may differ.

Juvenile vs Adult Arrests



- **Adult arrests vastly dominate at 96.37% (87,160 arrests)** compared to juvenile arrests at only 3.63% (3,290 arrests), indicating that the **overwhelming majority of crime-related arrests involve adults aged 18 and older**.
- The **extremely low juvenile arrest rate** suggests either that minors commit significantly fewer crimes, that juvenile cases are handled differently (diversion programs, warnings), or that juvenile crimes have lower reporting/arrest rates.
- This data point reinforces that **crime in Los Angeles is predominantly an adult issue**, with law enforcement resources and criminal justice proceedings primarily focused on adult offenders rather than youth crime prevention.