The Los Angeles Police Department's transition to a NIBRS-compliant crime and arrest reporting system

Prathamesh Shankar Agare
MS in Data Science
FAU Erlangen-Nürnberg Erlangen, Germany
prathamesh.agare@fau.de

Project Description

The Los Angeles Police Department (LAPD) is working toward an upgrade of their present crime and arrest records' system into one that meets the new requirements of the Federal Bureau of Investigation, known as the National Incident-Based Reporting System (NIBRS). These new standards cover a nationwide scope concerning making data on crimes more consistent, accurate, and specific. Not only will this include input from the community on trends of crime affecting Los Angeles, but also in the future, it will include individuals beyond their immediate surroundings.

At the moment, while LAPD builds this new system, the public will have sight of information only from the old one. To make the records the more accurate during this transition period, the department has switched from updating crime reports each week to updates every other week.

Project Question

What is the relationship between the number of reported crimes and arrests made within different age groups in Los Angeles Police Department (LAPD)? Are younger or older age groups more likely to be arrested relative to crime occurrences?

Data sources

Data Source 1: Crime Data from 2020 to Present (2024)

- Metadata Context: https://project-open-data.cio.gov/v1.1/schema/catalog.jsonld
- Metadata Catalog ID: https://data.lacity.org/data.json
- License: http://creativecommons.org/publicdomain/zero/1.0/legalcode
- URL: https://data.lacity.org/api/views/2nrs-mtv8/rows.csv?accessType=DOWNLOAD
- Data Type: CSV

Description:

This dataset shows crime incidents in Los Angeles from 2020 to now. It includes information like the age and gender of people involved, the location of the crime, and a description of what happened.

Data Source 2: Arrest Data from 2020 to Present (2024)

- Metadata Context: https://project-open-data.cio.gov/v1.1/schema/catalog.jsonld
- Metadata Catalog ID: https://data.lacity.org/data.json
- License: http://creativecommons.org/publicdomain/zero/1.0/legalcode
- URL: https://data.lacity.org/api/views/amvf-fr72/rows.csv?accessType=DOWNLOAD
- Data Type: CSV

Description:

This dataset arrest records in Los Angeles from 2020 to now, including details such as arrest date, time, location (latitude/longitude), demographics (age, sex, descent), charge descriptions, and booking information. It provides insights into criminal activity, geographic patterns, and offender profiles.

Data Pipeline

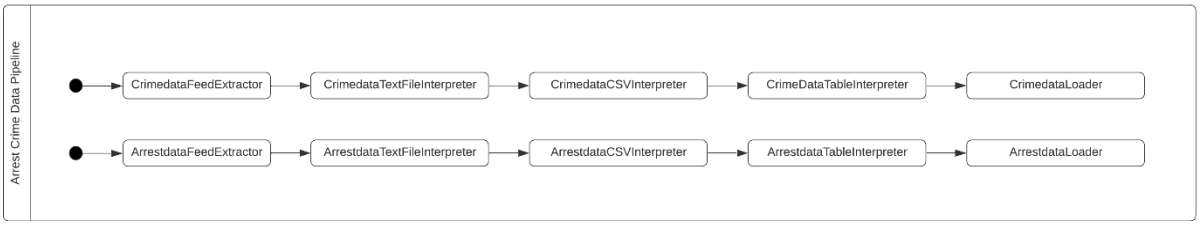
ETL pipelines are processes that extract data from different sources, transform it into a usable format, and load it into a target system, e.g. a database or data warehouse. ETL thus automates data integration processes, making it consistent and ready for later analysis/reporting.

The ETL pipeline is build with Jayvee. Jayvee is the Domain-specific language (DSL) to model data pipelines. Basically, it extracts public datasets including Crime Data from 2020 to Present (2024) and Arrest Data from 2020 to Present (2024).

The process of data cleaning only contains records with positive age value and will filter out all invalid entries with negative age records. Gender will also be validated such that only "male" and "female" entries will be accepted for the purposes of data consistency and accuracy.

The data contains column values with embedded double quotes (") that could disrupt parsing or processing. To resolve this, an enclosing mechanism using double quotes (enclosing: "") was implemented to ensure proper handling and avoid errors during data manipulation in jayvee.

The loaded crime and arrest data will be examined at an analytical level with respect to trends, contiguities, and attributions in between patterns of crime phenomena, different demographic factors, and geographical influences on arrest rates.



Result and Limitations

Output Data and Discussion:

To analyze the relationship between the crime reported and the arrests made by the LAPD for different age groups, a study on the arrest-to-crime ratio for each age group is considered. It indicates the chances or possibilities that would be qualified as arrested against a number of crimes reported for certain selected age groups. The younger groups are usually within the range of 18–24 and often show a higher arrest-to-crime ratio compared to an elder group. This could probably point to a sharper focus by law enforcement on their behavior or the behavioral patterns they show. These older-aged groups have fewer ratios because it may not be as easy to identify such incidents or because they have other forms of crime in that age group. These trends can be defined through the actual comparison of occurrences, arrests, and age, all taken along with other factors such as crime types and various socioeconomic aspects.

Output Data Format:

The output from the data pipeline is an SQLite file. The choice of SQLite here is that it's easy and it's cross-platform across various systems. Sometimes it connects directly to pandas to do the work of actual analysis.

Potential Issues:

Extraction Delay - Jayvee pipeline faced issue while extracting SQLite file due to large amount of data, hence pipeline taking too much time to execute.