Preface:

For this Final Project, I will be looking at the dataset regarding crime reports in the city of Los Angeles. This dataset contains every single report in the city of Los Angeles in the past 2 years along with the attributes related to it, such as the victim’s age and the kind of crime that has been committed. [1]

Assumptions:

* For this Final Project we will only be looking at the data up to April 5, 2022 (the day this dataset was collected).
* We will also be looking at the first crime code as the type of crime committed. As there are instances where a person is arrested on multiple charges.

Design Process:

* 6 tables will be created.
  + The ‘Report’ table is a time dimensional table that contains records of every single crime committed including time of occurence.
  + The ‘Area’ table contains the name of the area associated with the code.
  + The ‘Victim\_Desc’ table contains the description of every single victim
  + The ‘Status’ table is a slowly changing dimension as the status of the crime can change.
  + The ‘Premise’ table contains the description of the premise that is associated with its code.
  + The ‘Weapon\_Use’ table is a hierarchical table where it will break down whether or not the weapon used is a firearm.

ETL Process:

* The data will be extracted from the CSV
* The data will then be transformed and loaded into each table; we make sure that each table contains an ID column that we can trace back to the ‘Report’ table.
* This process was chosen as it is the most simplest yet effective method to undergo this task.

Professional Document:

Between the Year 2020 and 2022, the city of Los Angeles saw 445,167 crimes reported. The city is divided into 21 zones, most crime took place in Zone 12 (77th Street) where 28,939 reports were filled (6.50% of the total crime in 2 years), Zone 16 (Foothill) saw the least amount of reports filled at 15,657 reports (only 3.52%).

Victims were more likely to be self-reported as male than female (191,300 vs. 165,386). Both sexes exhibit a similar trend regarding the ethnic background of each victim. People of a Hispanic background are the most likely to be a victim, followed by Caucasians and African Americans.

As of the time the dataset was collected. Out of all of the 445,167 reports that were filled, only 41,471 ends up in arrests. Most reports end up with an investigation that needs to be continued (362,430).

The most rampant crime that was experienced correlates with the most occurring premise of the incident. 510 (Vehicle- stolen) is the most reported crime, along with the premise of incident being on the streets.

The city also saw 20,752 incident that involves a firearm a rate of 4.67%, compared to the national average of 13.55%. Which means that while Los Angeles is experiencing some firearm problem, it is nowhere close to the national average.[2]

References:

1. https://data.lacity.org/Public-Safety/Crime-Data-from-2020-to-Present/2nrs-mtv8
2. https://worldpopulationreview.com/state-rankings/gun-violence-by-state