Created by Professor Juveria Baig for academic use. Use this template for all class deliverables where specified. Save assignments as pdf and upload to e-learning. Do not submit any other formats or your deliverable will receive a 0

Project Title - Data Report

Author: Alaya Sirigiri

Date: September 17th, 2023

Class: ITSS 4V95

For formatting Jupyter notebook, refer to https://www.ibm.com/docs/en/watson-studio-local/1.2.3?topic=notebooks-markdown-jupyter-cheatsheet

Table of Contents:

Business Problem Overview

Problem Statement

Looking the data for the crimes in Los Angeles, California to conduct an analysis of the demographics and how that affects the likelihood of an individual being a victim of a crime.

Hypothesis: If a crime is to be chosen at random, then it is more likely that the individual is of a minority group and in the age range of 18-22, because they are the most at risk and vulnerable.

Data Overview and Pre-processing

Data Overview *Provide information related to data source and update frequency. Use knowledge learned to classify and define data and give insights into the overall quality of data.*

Example:

● The data contains information for about 798,242 crimes as of Spetember 15, 2023.

● There are 27 columns in total. Of these, six columns have a significant number of missing (Null or NaN) values. Some of these columns are not applicable to every crime that is reported, and therefore, many rows will have empty values.

● The characteristics (columns) include DR_NO, Date Rptd, Date OCC, DATE OCC, TIME OCC, AREA, AREA NAME, Rpt Dist No, Part 1-2, Crm Cd, Crm Cd Desc, Mocodes, Vict Age, Vict Sex, Vict Descent, Premis Cd, Premis Desc, Weapon Used Cd, Weapon Desc, Status, Status Desc, Crm Cd 1, Crm Cd 2, Crm Cd 3, Crm Cd 4, LOCATION, Cross Street, LAT, and LON.

Data Dictionary

Column Name	Description	Data Type	Required
DR_NO	Report Number (PK)	Int	Yes
DATE RPTD	Date Crime Reported	Date	Yes
DATE OCC	Date Crime Occurred	Date	Yes
TIME OCC	Time Crime Occurred	Date	Yes
AREA	Geographical Area/ Police Dept. Crime Occurred	String	Yes
AREA NAME	Geographical Area/ Police Dept. Crime Occurred	String	Yes
RRPT DIST NO	Report District Number	Int	Yes
PART 1-2	Crime Classification	Int	Yes
CRM CD	Crime Code	Int	Yes
CRM CD DESC	Description of Crime Code	String	Yes
MOCODES	Modus Operandi Code	Int	No
VICT AGE	Victim Age	Int	No
VICT SEX	Victim Sex	String	No
VICT DESCENT	Victim Descent	String	No
PREMIS CODE	Premis Code	Int	No
PREMIS DESC	Premis Description	String	No
WEAPON USED	Code for Weapon Used	Int	No
WEAPON CD	Code for Weapon Used	Int	No
WEAPON DESC	Weapon Used Description	String	No
STATUS	Status of the Crime	String	No
STATUS DESC	Desctption of the Status of the Crime	String	No
CRM CD 1-4	Additional Crime Codes	Int	No
LOCATION	The Location Crime Occurred	String	No
CROSS STREET	The Name of the Intersecting Street (if applicable)	String	No
LAT	The Latititude where Crime Occurred	Float	No
LON	The Longitude where Crime Occurred	Float	No

Column Name	Description	Data Type	Required
Deg of Crm	Degree of Crime	Int	No

In [1]:
```
from IPython.display import display, HTML
display(HTML("<style>.container { width:100% !important; )</style>"))
```

In [2]:
```
import pandas as pd
import numpy as np
from openpyxl import load_workbook
import matplotlib.pyplot as plt
from sklearn.cluster import DBSCAN
from sklearn.preprocessing import StandardScaler
import seaborn as sns
from datetime import datetime
from scipy.stats import linregress
import os
import random
%matplotlib inline
```

In [3]:
```
path = "/Users/alaya/Desktop/data sci project"
source_data = '/Users/alaya/Desktop/Crime_Data_from_2020_to_Present.csv'
df = pd.read_csv(source_data)
```

Data Pre-Processing

Briefly explain data cleansing, blending and joining activities. Examples: dropping/renaming of important features, null values, creation of new features/attributes,defining data types etc

Graphics/Visuals: It's also a good idea to give an illustration of sample data (5 rows or so) and data workflow (if using Alteryx or Orange)

In [4]:
```
data_sample = pd.read_csv('/Users/alaya/Desktop/data sci project/Crime_Data_fro
print(data_sample)
```

```
       DR_NO               Date Rptd     DATE OCC  TIME OCC  AREA    AREA NAME
\
0   10304468  01/08/2020 12:00:00 AM  2020-01-08      2230     3    Southwest
1  190101086  01/02/2020 12:00:00 AM  2020-01-01       330     1      Central
2  200110444  04/14/2020 12:00:00 AM  2020-02-13      1200     1      Central
3  191501505  01/01/2020 12:00:00 AM  2020-01-01      1730    15  N Hollywood
4  191921269  01/01/2020 12:00:00 AM  2020-01-01       415    19      Mission

   Rpt Dist No  Part 1-2  Crm Cd  \
0          377         2     624
1          163         2     624
2          155         2     845
3         1543         2     745
4         1998         2     740

                                      Crm Cd Desc  ... Crm Cd 1  Crm Cd 2
\
0                          BATTERY - SIMPLE ASSAULT  ...    624.0       NaN
1                          BATTERY - SIMPLE ASSAULT  ...    624.0       NaN
2          SEX OFFENDER REGISTRANT OUT OF COMPLIANCE  ...    845.0       NaN
3           VANDALISM - MISDEAMEANOR ($399 OR UNDER)  ...    745.0     998.0
4  VANDALISM - FELONY ($400 & OVER, ALL CHURCH VA...  ...    740.0       NaN

  Crm Cd 3 Crm Cd 4                                LOCATION Cross Street  \
0      NaN      NaN   1100 W  39TH                      PL          NaN
1      NaN      NaN    700 S  HILL                      ST          NaN
2      NaN      NaN    200 E  6TH                       ST          NaN
3      NaN      NaN   5400    CORTEEN                   PL          NaN
4      NaN      NaN  14400    TITUS                     ST          NaN

       LAT       LON Degree of Crime Deg of Crm
0  34.0141 -118.2978               3          3
1  34.0459 -118.2545               3          3
2  34.0448 -118.2474               0          0
3  34.1685 -118.4019               0          0
4  34.2198 -118.4468               1          1

[5 rows x 30 columns]
```

In [5]:
```
df_crm_cds = new_df = df[['Crm Cd', 'Crm Cd Desc']]


csv_file = '/Users/alaya/Desktop/data sci project/Crime_Data_from_2020_to_Prese

new_df.to_csv(csv_file, index=False, mode='a')
```

In [6]:
```
list(df['Vict Descent'].unique())

# changing the descent values to be more insightful and add value to chart

df['Vict Descent'] = df['Vict Descent'].map({'B':'Black',
 'H':'Hispanic',
 'X': 'Unknown',
 'W': 'White',
 'A': 'Asian',
 'O': 'Other',
 'C' : 'Native American',
 'F': 'Filipino',
 'K': 'Korean',
 'I': 'Pacific Islander' ,
```

```
 'V': 'Vietnamese',
 'Z': 'Asian Indian',
 'J': 'Japanese',
 'P': 'Cambodian',
 'G': 'Guamanian',
 'U': 'Hawaiian',
 'D': 'Laotian',
 'S': 'Samoan',
 'L': 'Fijian',
 '-' : '-'})
```

Exploratory Data Analysis

Exploratory Data Analysis (EDA) is a critical step in the data analysis process that involves examining and visualizing data to gain insights, discover patterns, and identify potential issues. Various techniques can be used during EDA such as below, apply the ones that satisfy goals of the analysis.*

Classification of Data	Quality of Data	Accuracy	Reliability
Transformed Data	Appropriate and descriptive column names	Secondary source	N/A - No record of outages in the database
Qualitative and Quantitative Data	Unique PK for each record	Data collected by LAPD OpenData (City of Los Angeles)	Metadata and dataset creation date November 10, 2020
Row Count: 798,242	No empty rows in the dataset	"Data is as accurate as the data in the database" - Data.gov	Regularly updated (updated daily)
Date: November 10, 2020 - September 15, 2023	Data types: String, Float, Int, Date, Geodata		License: https://shorturl.at/bEH39
	No anomalies in the data		

In [7]:
```
df = pd.read_csv(source_data)
```

In [8]:
```
#information about crime dataset

df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 802956 entries, 0 to 802955
Data columns (total 28 columns):
 #   Column          Non-Null Count   Dtype  
---  ------          --------------   -----  
 0   DR_NO           802956 non-null  int64  
 1   Date Rptd       802956 non-null  object 
 2   DATE OCC        802956 non-null  object 
 3   TIME OCC        802956 non-null  int64  
 4   AREA            802956 non-null  int64  
 5   AREA NAME       802956 non-null  object 
 6   Rpt Dist No     802956 non-null  int64  
 7   Part 1-2        802956 non-null  int64  
 8   Crm Cd          802956 non-null  int64  
 9   Crm Cd Desc     802956 non-null  object 
 10  Mocodes         692304 non-null  object 
 11  Vict Age        802956 non-null  int64  
 12  Vict Sex        697683 non-null  object 
 13  Vict Descent    697675 non-null  object 
 14  Premis Cd       802947 non-null  float64
 15  Premis Desc     802481 non-null  object 
 16  Weapon Used Cd  279525 non-null  float64
 17  Weapon Desc     279525 non-null  object 
 18  Status          802956 non-null  object 
 19  Status Desc     802956 non-null  object 
 20  Crm Cd 1        802946 non-null  float64
 21  Crm Cd 2        59147 non-null   float64
 22  Crm Cd 3        1970 non-null    float64
 23  Crm Cd 4        57 non-null      float64
 24  LOCATION        802956 non-null  object 
 25  Cross Street    128522 non-null  object 
 26  LAT             802956 non-null  float64
 27  LON             802956 non-null  float64
dtypes: float64(8), int64(7), object(13)
memory usage: 171.5+ MB
```

In [195…

```
#the unique crimes and their desc that have been reported in LA

unique_crm_cds = df['Crm Cd'].unique()

#keeping a running total of the number of unique crimes reported
total_unique_crms = 0

#unique crimes + their description as the outpu

for cd in unique_crm_cds:
    desc = df[df['Crm Cd'] == cd]['Crm Cd Desc'].iloc[0]
    total_unique_crms = total_unique_crms + 1
    print(f"Crime Code: {cd}, Description: {desc}")

#print the number of unique crimes in LA
print('\nThe # of unique crimes reported in LA, California is:', total_unique_c
```

```
Crime Code: 624, Description: BATTERY - SIMPLE ASSAULT
Crime Code: 845, Description: SEX OFFENDER REGISTRANT OUT OF COMPLIANCE
Crime Code: 745, Description: VANDALISM - MISDEAMEANOR ($399 OR UNDER)
Crime Code: 740, Description: VANDALISM - FELONY ($400 & OVER, ALL CHURCH VAND
ALISMS)
Crime Code: 121, Description: RAPE, FORCIBLE
Crime Code: 442, Description: SHOPLIFTING - PETTY THEFT ($950 & UNDER)
Crime Code: 946, Description: OTHER MISCELLANEOUS CRIME
Crime Code: 341, Description: THEFT-GRAND ($950.01 & OVER)EXCPT,GUNS,FOWL,LIVE
STK,PROD
Crime Code: 330, Description: BURGLARY FROM VEHICLE
Crime Code: 930, Description: CRIMINAL THREATS - NO WEAPON DISPLAYED
Crime Code: 648, Description: ARSON
Crime Code: 626, Description: INTIMATE PARTNER - SIMPLE ASSAULT
Crime Code: 440, Description: THEFT PLAIN - PETTY ($950 & UNDER)
Crime Code: 354, Description: THEFT OF IDENTITY
Crime Code: 210, Description: ROBBERY
Crime Code: 230, Description: ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT
Crime Code: 310, Description: BURGLARY
Crime Code: 510, Description: VEHICLE - STOLEN
Crime Code: 420, Description: THEFT FROM MOTOR VEHICLE - PETTY ($950 & UNDER)
Crime Code: 761, Description: BRANDISH WEAPON
Crime Code: 236, Description: INTIMATE PARTNER - AGGRAVATED ASSAULT
Crime Code: 662, Description: BUNCO, GRAND THEFT
Crime Code: 350, Description: THEFT, PERSON
Crime Code: 860, Description: BATTERY WITH SEXUAL CONTACT
Crime Code: 480, Description: BIKE - STOLEN
Crime Code: 623, Description: BATTERY POLICE (SIMPLE)
Crime Code: 956, Description: LETTERS, LEWD  -  TELEPHONE CALLS, LEWD
Crime Code: 900, Description: VIOLATION OF COURT ORDER
Crime Code: 888, Description: TRESPASSING
Crime Code: 331, Description: THEFT FROM MOTOR VEHICLE - GRAND ($950.01 AND OV
ER)
Crime Code: 901, Description: VIOLATION OF RESTRAINING ORDER
Crime Code: 886, Description: DISTURBING THE PEACE
Crime Code: 421, Description: THEFT FROM MOTOR VEHICLE - ATTEMPT
Crime Code: 647, Description: THROWING OBJECT AT MOVING VEHICLE
Crime Code: 940, Description: EXTORTION
Crime Code: 810, Description: SEX,UNLAWFUL(INC MUTUAL CONSENT, PENETRATION W/
FRGN OBJ
Crime Code: 922, Description: CHILD STEALING
Crime Code: 812, Description: CRM AGNST CHLD (13 OR UNDER) (14-15 & SUSP 10 YR
S OLDER)
Crime Code: 220, Description: ATTEMPTED ROBBERY
Crime Code: 625, Description: OTHER ASSAULT
Crime Code: 755, Description: BOMB SCARE
Crime Code: 649, Description: DOCUMENT FORGERY / STOLEN FELONY
Crime Code: 815, Description: SEXUAL PENETRATION W/FOREIGN OBJECT
Crime Code: 251, Description: SHOTS FIRED AT INHABITED DWELLING
Crime Code: 320, Description: BURGLARY, ATTEMPTED
Crime Code: 890, Description: FAILURE TO YIELD
Crime Code: 850, Description: INDECENT EXPOSURE
Crime Code: 820, Description: ORAL COPULATION
Crime Code: 668, Description: EMBEZZLEMENT, GRAND THEFT ($950.01 & OVER)
Crime Code: 902, Description: VIOLATION OF TEMPORARY RESTRAINING ORDER
Crime Code: 664, Description: BUNCO, PETTY THEFT
Crime Code: 920, Description: KIDNAPPING - GRAND ATTEMPT
Crime Code: 343, Description: SHOPLIFTING-GRAND THEFT ($950.01 & OVER)
Crime Code: 437, Description: RESISTING ARREST
Crime Code: 753, Description: DISCHARGE FIREARMS/SHOTS FIRED
```

```
Crime Code: 928, Description: THREATENING PHONE CALLS/LETTERS
Crime Code: 910, Description: KIDNAPPING
Crime Code: 760, Description: LEWD/LASCIVIOUS ACTS WITH CHILD
Crime Code: 762, Description: LEWD CONDUCT
Crime Code: 661, Description: UNAUTHORIZED COMPUTER ACCESS
Crime Code: 351, Description: PURSE SNATCHING
Crime Code: 821, Description: SODOMY/SEXUAL CONTACT B/W PENIS OF ONE PERS TO A
NUS OTH
Crime Code: 237, Description: CHILD NEGLECT (SEE 300 W.I.C.)
Crime Code: 903, Description: CONTEMPT OF COURT
Crime Code: 813, Description: CHILD ANNOYING (17YRS & UNDER)
Crime Code: 666, Description: BUNCO, ATTEMPT
Crime Code: 627, Description: CHILD ABUSE (PHYSICAL) - SIMPLE ASSAULT
Crime Code: 805, Description: PIMPING
Crime Code: 763, Description: STALKING
Crime Code: 441, Description: THEFT PLAIN - ATTEMPT
Crime Code: 122, Description: RAPE, ATTEMPTED
Crime Code: 443, Description: SHOPLIFTING - ATTEMPT
Crime Code: 450, Description: THEFT FROM PERSON - ATTEMPT
Crime Code: 520, Description: VEHICLE - ATTEMPT STOLEN
Crime Code: 434, Description: FALSE IMPRISONMENT
Crime Code: 410, Description: BURGLARY FROM VEHICLE, ATTEMPTED
Crime Code: 352, Description: PICKPOCKET
Crime Code: 670, Description: EMBEZZLEMENT, PETTY THEFT ($950 & UNDER)
Crime Code: 951, Description: DEFRAUDING INNKEEPER/THEFT OF SERVICES, $950 & U
NDER
Crime Code: 660, Description: COUNTERFEIT
Crime Code: 654, Description: CREDIT CARDS, FRAUD USE ($950 & UNDER
Crime Code: 250, Description: SHOTS FIRED AT MOVING VEHICLE, TRAIN OR AIRCRAFT
Crime Code: 110, Description: CRIMINAL HOMICIDE
Crime Code: 652, Description: DOCUMENT WORTHLESS ($200 & UNDER)
Crime Code: 933, Description: PROWLER
Crime Code: 950, Description: DEFRAUDING INNKEEPER/THEFT OF SERVICES, OVER $95
0.01
Crime Code: 231, Description: ASSAULT WITH DEADLY WEAPON ON POLICE OFFICER
Crime Code: 345, Description: DISHONEST EMPLOYEE - GRAND THEFT
Crime Code: 822, Description: HUMAN TRAFFICKING - COMMERCIAL SEX ACTS
Crime Code: 814, Description: CHILD PORNOGRAPHY
Crime Code: 932, Description: PEEPING TOM
Crime Code: 622, Description: BATTERY ON A FIREFIGHTER
Crime Code: 471, Description: TILL TAP - PETTY ($950 & UNDER)
Crime Code: 235, Description: CHILD ABUSE (PHYSICAL) - AGGRAVATED ASSAULT
Crime Code: 470, Description: TILL TAP - GRAND THEFT ($950.01 & OVER)
Crime Code: 921, Description: HUMAN TRAFFICKING - INVOLUNTARY SERVITUDE
Crime Code: 906, Description: FIREARMS RESTRAINING ORDER (FIREARMS RO)
Crime Code: 433, Description: DRIVING WITHOUT OWNER CONSENT (DWOC)
Crime Code: 651, Description: DOCUMENT WORTHLESS ($200.01 & OVER)
Crime Code: 806, Description: PANDERING
Crime Code: 943, Description: CRUELTY TO ANIMALS
Crime Code: 653, Description: CREDIT CARDS, FRAUD USE ($950.01 & OVER)
Crime Code: 436, Description: LYNCHING - ATTEMPTED
Crime Code: 949, Description: ILLEGAL DUMPING
Crime Code: 446, Description: PETTY THEFT - AUTO REPAIR
Crime Code: 113, Description: MANSLAUGHTER, NEGLIGENT
Crime Code: 487, Description: BOAT - STOLEN
Crime Code: 438, Description: RECKLESS DRIVING
Crime Code: 451, Description: PURSE SNATCHING - ATTEMPT
Crime Code: 439, Description: FALSE POLICE REPORT
Crime Code: 485, Description: BIKE - ATTEMPTED STOLEN
Crime Code: 944, Description: CONSPIRACY
```

```
Crime Code: 954, Description: CONTRIBUTING
Crime Code: 756, Description: WEAPONS POSSESSION/BOMBING
Crime Code: 942, Description: BRIBERY
Crime Code: 473, Description: THEFT, COIN MACHINE - GRAND ($950.01 & OVER)
Crime Code: 347, Description: GRAND THEFT / INSURANCE FRAUD
Crime Code: 435, Description: LYNCHING
Crime Code: 880, Description: DISRUPT SCHOOL
Crime Code: 444, Description: DISHONEST EMPLOYEE - PETTY THEFT
Crime Code: 475, Description: THEFT, COIN MACHINE - ATTEMPT
Crime Code: 474, Description: THEFT, COIN MACHINE - PETTY ($950 & UNDER)
Crime Code: 931, Description: REPLICA FIREARMS(SALE,DISPLAY,MANUFACTURE OR DIS
TRIBUTE)
Crime Code: 865, Description: DRUGS, TO A MINOR
Crime Code: 349, Description: GRAND THEFT / AUTO REPAIR
Crime Code: 353, Description: DRUNK ROLL
Crime Code: 452, Description: PICKPOCKET, ATTEMPT
Crime Code: 870, Description: CHILD ABANDONMENT
Crime Code: 522, Description: VEHICLE, STOLEN - OTHER (MOTORIZED SCOOTERS, BIK
ES, ETC)
Crime Code: 924, Description: TELEPHONE PROPERTY - DAMAGE
Crime Code: 840, Description: BEASTIALITY, CRIME AGAINST NATURE SEXUAL ASSLT W
ITH ANIM
Crime Code: 948, Description: BIGAMY
Crime Code: 884, Description: FAILURE TO DISPERSE
Crime Code: 904, Description: FIREARMS EMERGENCY PROTECTIVE ORDER (FIREARMS EP
O)
Crime Code: 830, Description: INCEST (SEXUAL ACTS BETWEEN BLOOD RELATIVES)
Crime Code: 432, Description: BLOCKING DOOR INDUCTION CENTER
Crime Code: 882, Description: INCITING A RIOT
Crime Code: 445, Description: DISHONEST EMPLOYEE ATTEMPTED THEFT

The # of unique crimes reported in LA, California is: 138
```

In [10]:
```
#count of null rows and empty rows

print('number of rows with null values:', df.isna().any(axis=1).sum())
print('number of empty rows:',df.isna().all(axis=1).sum())
```

```
number of rows with null values: 802949
number of empty rows: 0
```

Descriptive Statistics:

Summary statistics: Calculate measures such as mean, median, mode, standard deviation, and range to describe the central tendency and spread of the data. Frequency distributions: Create histograms, bar charts, or frequency tables to visualize the distribution of categorical or numerical data.

Summary Statistics

In [11]:
```
#describing the crime dataset

print(df.describe())
```

```
              DR_NO       TIME OCC           AREA    Rpt Dist No  \
count  8.029560e+05  802956.000000  802956.000000  802956.000000
mean   2.158016e+08    1335.403294      10.715737    1117.999182
std    1.071699e+07     654.164612       6.092093     609.194026
min    8.170000e+02       1.000000       1.000000     101.000000
25%    2.101176e+08     900.000000       6.000000     622.000000
50%    2.201057e+08    1415.000000      11.000000    1142.000000
75%    2.218191e+08    1900.000000      16.000000    1617.000000
max    2.399165e+08    2359.000000      21.000000    2199.000000

            Part 1-2         Crm Cd       Vict Age      Premis Cd  \
count  802956.000000  802956.000000  802956.000000  802947.000000
mean        1.414290     500.743798      29.847791     305.763698
std         0.492599     207.829477      21.762934     216.577524
min         1.000000     110.000000      -3.000000     101.000000
25%         1.000000     331.000000       9.000000     101.000000
50%         1.000000     442.000000      31.000000     203.000000
75%         2.000000     626.000000      45.000000     501.000000
max         2.000000     956.000000     120.000000     976.000000

       Weapon Used Cd       Crm Cd 1       Crm Cd 2      Crm Cd 3    Crm Cd 4  \
count   279525.000000  802946.000000  59147.000000  1970.000000   57.000000
mean       362.840966     500.481701    957.431045   983.892893  990.368421
std        123.726239     207.618224    111.585024    50.704245   28.594225
min        101.000000     110.000000    210.000000   434.000000  821.000000
25%        310.000000     331.000000    998.000000   998.000000  998.000000
50%        400.000000     442.000000    998.000000   998.000000  998.000000
75%        400.000000     626.000000    998.000000   998.000000  998.000000
max        516.000000     956.000000    999.000000   999.000000  999.000000

                LAT            LON
count  802956.000000  802956.000000
mean       33.977505    -118.020515
std         1.809752       6.275190
min         0.000000    -118.667600
25%        34.013600    -118.429600
50%        34.058400    -118.321500
75%        34.163100    -118.273900
max        34.334300       0.000000
```

In [12]:
```
#describes the numerical and categorical columns
print('NUMERICAL/QUANTITATIVE')
print(df.describe())
print('\nxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
print('CATEGORICAL/QUALITATIVE')
print(df.describe(include='object'))
```

```
NUMERICAL/QUANTITATIVE
              DR_NO       TIME OCC           AREA    Rpt Dist No  \
count  8.029560e+05  802956.000000  802956.000000  802956.000000
mean   2.158016e+08    1335.403294      10.715737    1117.999182
std    1.071699e+07     654.164612       6.092093     609.194026
min    8.170000e+02       1.000000       1.000000     101.000000
25%    2.101176e+08     900.000000       6.000000     622.000000
50%    2.201057e+08    1415.000000      11.000000    1142.000000
75%    2.218191e+08    1900.000000      16.000000    1617.000000
max    2.399165e+08    2359.000000      21.000000    2199.000000

            Part 1-2         Crm Cd       Vict Age      Premis Cd  \
count  802956.000000  802956.000000  802956.000000  802947.000000
mean        1.414290     500.743798      29.847791     305.763698
std         0.492599     207.829477      21.762934     216.577524
min         1.000000     110.000000      -3.000000     101.000000
25%         1.000000     331.000000       9.000000     101.000000
50%         1.000000     442.000000      31.000000     203.000000
75%         2.000000     626.000000      45.000000     501.000000
max         2.000000     956.000000     120.000000     976.000000

       Weapon Used Cd       Crm Cd 1       Crm Cd 2     Crm Cd 3    Crm Cd 4  \
count   279525.000000  802946.000000  59147.000000  1970.000000   57.000000
mean       362.840966     500.481701    957.431045   983.892893  990.368421
std        123.726239     207.618224    111.585024    50.704245   28.594225
min        101.000000     110.000000    210.000000   434.000000  821.000000
25%        310.000000     331.000000    998.000000   998.000000  998.000000
50%        400.000000     442.000000    998.000000   998.000000  998.000000
75%        400.000000     626.000000    998.000000   998.000000  998.000000
max        516.000000     956.000000    999.000000   999.000000  999.000000

                LAT            LON
count  802956.000000  802956.000000
mean       33.977505    -118.020515
std         1.809752       6.275190
min         0.000000    -118.667600
25%        34.013600    -118.429600
50%        34.058400    -118.321500
75%        34.163100    -118.273900
max        34.334300       0.000000

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

CATEGORICAL/QUALITATIVE
                    Date Rptd                 DATE OCC AREA NAME  \
count                  802956                   802956    802956
unique                   1357                     1357        21
top     02/03/2023 12:00:00 AM  12/02/2022 12:00:00 AM   Central
freq                      924                     1130     53901

             Crm Cd Desc Mocodes Vict Sex Vict Descent Premis Desc  \
count             802956  692304   697683       697675      802481
unique               138  268499        5           20         306
top     VEHICLE - STOLEN    0344        M            H      STREET
freq               85807   32929   331958       246453      202762

                                        Weapon Desc  Status  Status Desc  \
count                                        279525  802956       802956
unique                                           79       6            6
top     STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)      IC  Invest Cont
```

```
freq                                          149916  642828       642828

                                        LOCATION Cross Street
count                                     802956       128522
unique                                     63423         9612
top      800 N  ALAMEDA                       ST     BROADWAY
freq                                        1465         2143
```

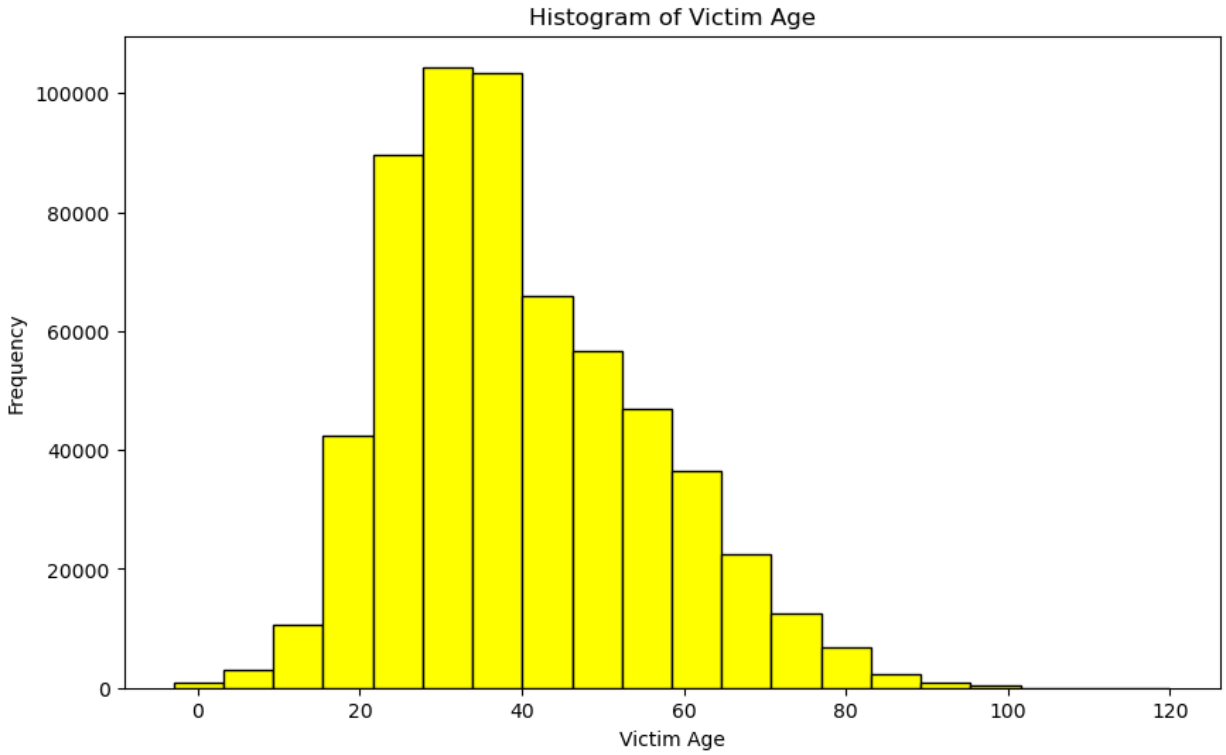
Frequency Statistics

Age Demographics of Victims

Understanding the age demographics of the victims of crimes in the Los Angeles area will allow for increased awareness and better allocation of resources. If individuals of an age group are disproportionately victims of crimes, then better policing and watchdog efforts can be made in neighborhoods that are heavily populated by said demographic.

If the victim's age is unknown, the LAPD has declared their age as '0' in the database. For analysis purposes, we will not consider the ages of the victims, who's age is unknown. Since each crime is unique in nature, imputation will not be used as it will create a disproportionate view of the targeted victims' age group. Outliers in the dataset have also been removed.

In [13]:
```
# filtered df with no 0 for vict age

# creating a new data from without the data where the victim is older than 0
mid_df = df[df['Vict Age'] != 0]

# create the new the column we will use
new_df = mid_df['Vict Age']

# dims and color
plt.figure(figsize=(10, 6))
plt.hist(new_df, bins=20, color='yellow', edgecolor='black')

# labels
plt.xlabel('Victim Age')
plt.ylabel('Frequency')
plt.title('Histogram of Victim Age')

#output
plt.show()
```



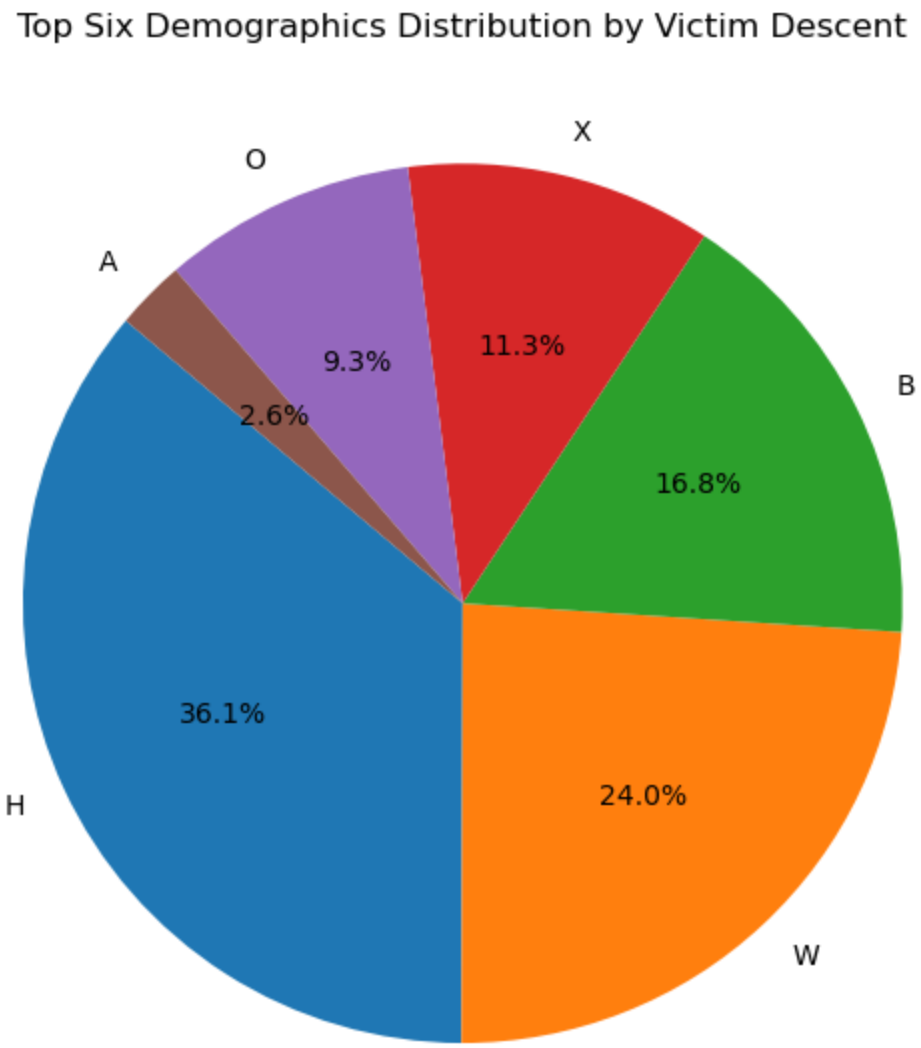
Descent Demographics of Victims

Understanding the descent demographics of the victims of crimes in the Los Angeles area will allow for increased awareness and better allocation of resources. If individuals of a certain descent are disproportionately victims of crimes, then better policing and watchdog efforts can be made in neighborhoods that are heavily populated by said demographic.

For simplicity and conciseness, this report will only consider from which race the most number of victims yielded. Additionally, in some crimes, the victim's descent may never have been identified due to the nature of the crime or other factors; such victims will be categorized as 'Unknown.'

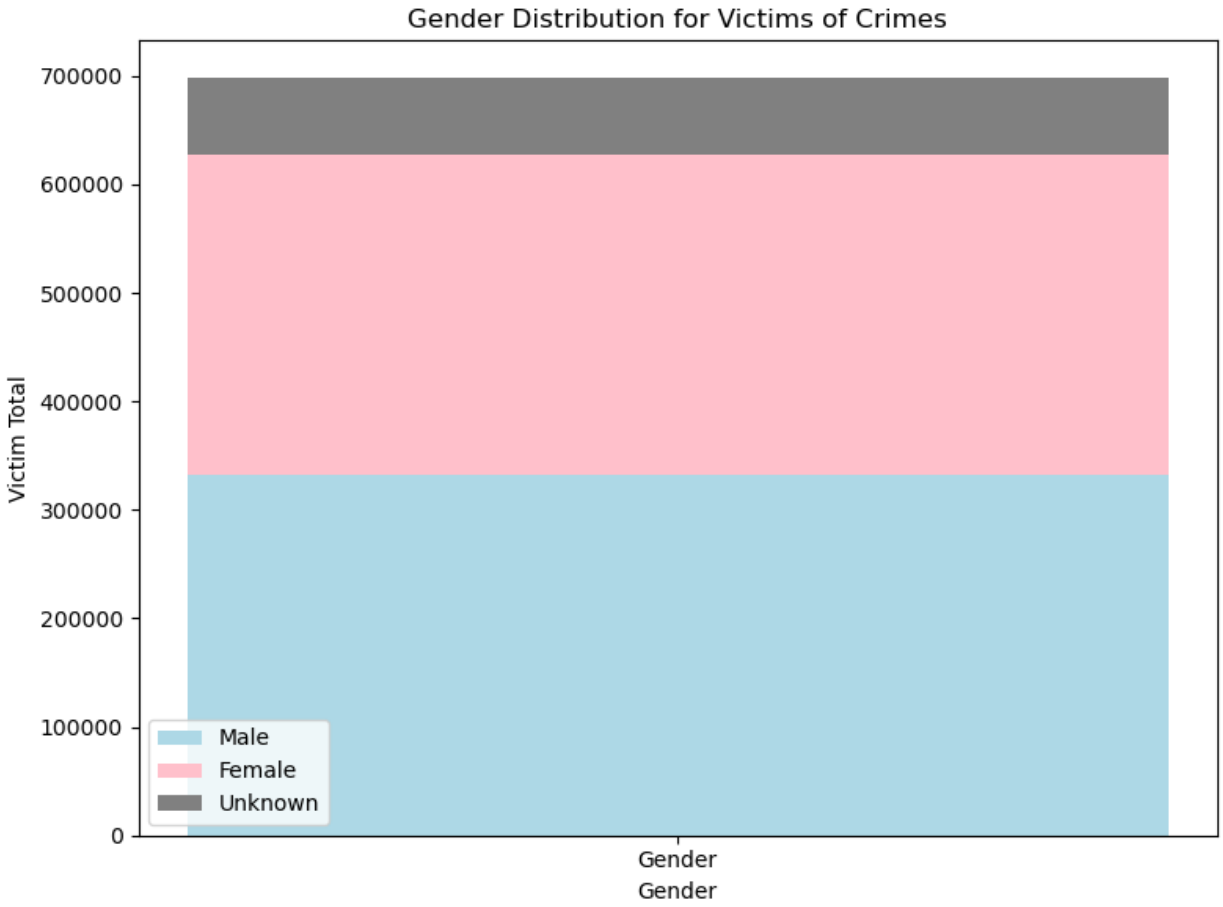
In [14]:
```
#we will be using only the top six decents reported for conciseness

#iterates through the rows and instance of each descent/race
v_desc_count = df['Vict Descent'].value_counts()
#limit to top 6 demographics
v_desc_count = v_desc_count.head(6)
#category names
race_labels = v_desc_count.index
#get total counts for each race
v_desc_total = v_desc_count.values


#figure size of the pie chart
plt.figure(figsize=(8, 6))
#create the pie chart
plt.pie(v_desc_total, labels=race_labels, autopct='%1.1f%%', startangle=140)
#create the descriptive chart title
plt.title('Top Six Demographics Distribution by Victim Descent')

#adding data labels
```

```
plt.gca().set_aspect('equal')
plt.tight_layout()

#display the output as a pie chart
plt.show()
```



In [15]:
```
#understanding the distribution of male, female, and other victims (vict gender

#count the instances for M, F, and X by row
gender_total = df['Vict Sex'].value_counts()

m = gender_total.get('M', 0)
f = gender_total.get('F', 0)
x = gender_total.get('X', 0)

#figure size of the stacked column chart
plt.figure(figsize=(8, 6))
#creating the stacked column chart with colors and labels
plt.bar('Gender', m, label='Male', color='lightblue')
plt.bar('Gender', f, bottom=m, label='Female', color='pink')
plt.bar('Gender', x, bottom=m + f, label='Unknown', color='grey')

#axis and chart titles
```

```
plt.xlabel('Gender')
plt.ylabel('Victim Total')
plt.title('Gender Distribution for Victims of Crimes')

#create legend
plt.legend()

#show output
plt.tight_layout()
plt.show()
```



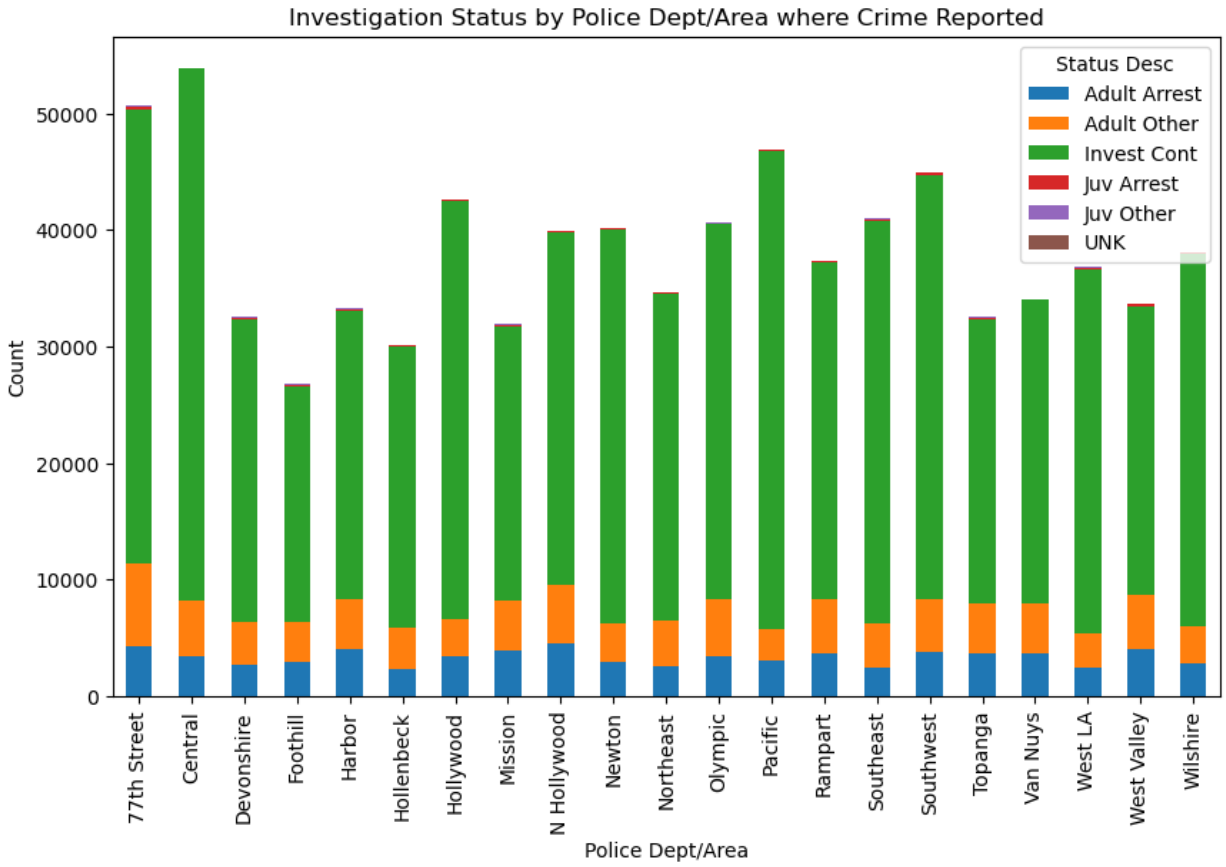
Week Two

Unsolved Cases by Police Department and Area

Looking at how many cases are unsolved by Police Department and Area can help better allocate resources. If there is an area which has disproportionate number of unsolved cases, better training, resources, and practices should be applied at that divison.

In [16]:
```
# stacked column chart - solved, under investigation, and the other. stack by 1

#understanding the how many investigations are open by the area

#creating a new data frame to pivot the data
pivot_df = df.pivot_table(index='AREA NAME', columns='Status Desc', aggfunc='si

pivot_df.plot(kind='bar', stacked=True, figsize=(10, 6))

#axis labels and title
```

```
plt.xlabel('Police Dept/Area')
plt.ylabel('Count')
plt.title('Investigation Status by Police Dept/Area where Crime Reported')

#output
plt.show()
```



Classfication of Crimes by Penal Degree and Offense

The Process of Classification:

To classify, these crimes, there must be a set of rules applied for each crime to be assigned to a specific category. For a criminal offense to be brought against a perpetrator, the legal system (prosecutor/denfense/judge) must be involved. Since each jurisdiction operates differently and crimes are tried differently given the circumstances, for simplicity, objective rules will be used in this case.

The rules for felonies and misdemeanors for this purpose are as follows:

Felonies

- more severe criminal act and can be heavily penalized by the courts
- typically more violent and crude
- significantly longer prison times and harsher legal consequences
- for analysis purposes, the degree of crime assigned is 1

Misdemeanors

- less severe criminal act
- not as violent or cruel
- less severe penalties; little to no jail time and a fine
- for analysis purposes, the degree of crime assigned is 2

Infraction

- a criminal act
- not severe enough to warrant an arrest
- least severe criminal act
- for analysis purposes, the degree of crime assigned is 3

In [17]:
```
#using a sample list of crime codes to assign values of criminal offense to
#used chatgpt to identify the crimes + offense

crime_code_to_degree = {
    740: '1', 121: '1', 230: '1', 210: '1', 310: '1', 220: '1', 110: '1',
    902: '2', 440: '2', 443: '2', 471: '2', 950: '2',
    624: '3', 330: '3', 626: '3', 351: '3', 850: '3', 820: '3'
}

#adding a new column to the dataset
df['Deg of Crm'] = df['Crm Cd'].map(crime_code_to_degree)
df['Deg of Crm'] = df['Deg of Crm'].fillna('0')



#saving to csv
df.to_csv('/Users/alaya/Desktop/data sci project/Crime_Data_from_2020_to_Preser
```

In [24]:
```
data_sample = pd.read_csv('/Users/alaya/Desktop/data sci project/Crime_Data_fro
print(data_sample)
```

```
       DR_NO               Date Rptd                DATE OCC  TIME OCC  AREA
\
0   10304468  01/08/2020 12:00:00 AM  01/08/2020 12:00:00 AM      2230     3
1  190101086  01/02/2020 12:00:00 AM  01/01/2020 12:00:00 AM       330     1
2  200110444  04/14/2020 12:00:00 AM  02/13/2020 12:00:00 AM      1200     1
3  191501505  01/01/2020 12:00:00 AM  01/01/2020 12:00:00 AM      1730    15
4  191921269  01/01/2020 12:00:00 AM  01/01/2020 12:00:00 AM       415    19

     AREA NAME  Rpt Dist No  Part 1-2  Crm Cd  \
0    Southwest          377         2     624
1      Central          163         2     624
2      Central          155         2     845
3  N Hollywood         1543         2     745
4      Mission         1998         2     740

                                      Crm Cd Desc  ...   Status Desc  \
0                          BATTERY - SIMPLE ASSAULT  ...   Adult Other
1                          BATTERY - SIMPLE ASSAULT  ...   Invest Cont
2          SEX OFFENDER REGISTRANT OUT OF COMPLIANCE  ...  Adult Arrest
3           VANDALISM - MISDEAMEANOR ($399 OR UNDER)  ...   Invest Cont
4  VANDALISM - FELONY ($400 & OVER, ALL CHURCH VA...  ...   Invest Cont

   Crm Cd 1 Crm Cd 2 Crm Cd 3  Crm Cd 4  \
0     624.0      NaN      NaN       NaN
1     624.0      NaN      NaN       NaN
2     845.0      NaN      NaN       NaN
3     745.0    998.0      NaN       NaN
4     740.0      NaN      NaN       NaN

                                LOCATION  Cross Street      LAT       LON
\
0   1100 W  39TH                     PL           NaN  34.0141 -118.2978
1    700 S  HILL                     ST           NaN  34.0459 -118.2545
2    200 E  6TH                      ST           NaN  34.0448 -118.2474
3   5400    CORTEEN                  PL           NaN  34.1685 -118.4019
4  14400    TITUS                    ST           NaN  34.2198 -118.4468

  Deg of Crm
0          3
1          3
2          0
3          0
4          1

[5 rows x 29 columns]
```

In [25]:
```
#distribution of felonies, misdemeanors, etc.
```

Data Transformation:

Data normalization: Scale numerical features to have a similar range or distribution. Data encoding: Convert categorical variables into numerical format using techniques like one-hot encoding or label encoding. Data imputation: Fill missing values using methods such as mean, median, or regression imputation.

In [26]:
```
#number of 0 in the vict age columns is 196113

print(df['Vict Age'].value_counts()[0])
```

```
197418
```

In [27]:
```
#data imputation using median for victim age (greater than 0)

vict_median_age = df[df['Vict Age'] > 0]['Vict Age'].median()

#df.loc[df['Vict Age'] == 0, 'Vict Age'] = vict_median_age
#print(vict_median_age)
#df.loc[df['Vict Age'] == 'Unknown', 'Vict Age'] = 0
```

Outlier Detection

Visual inspection: Use box plots, scatter plots, or histograms to identify outliers. Statistical methods: Apply techniques like the Z-score or IQR (Interquartile Range) to detect outliers.

In [28]:
```
#we can also see that there are outliers in the vict age column (the min is -3
#we can take the IQR of the vict age

#calc the IQR
q1 = df['Vict Age'].quantile(0.25)
q3 = df['Vict Age'].quantile(0.75)
iqr = q3 - q1

#outliers using bounds
lower = q1 - 1.5 * iqr
upper = q3 + 1.5 * iqr

outliers = (df['Vict Age']<lower) | (df['Vict Age'] > upper)

#replacing outliers with the median

df.loc[outliers, 'Vict Age'] = vict_median_age
```

Clustering, Classification, Time-series, Text analysis

Clustering

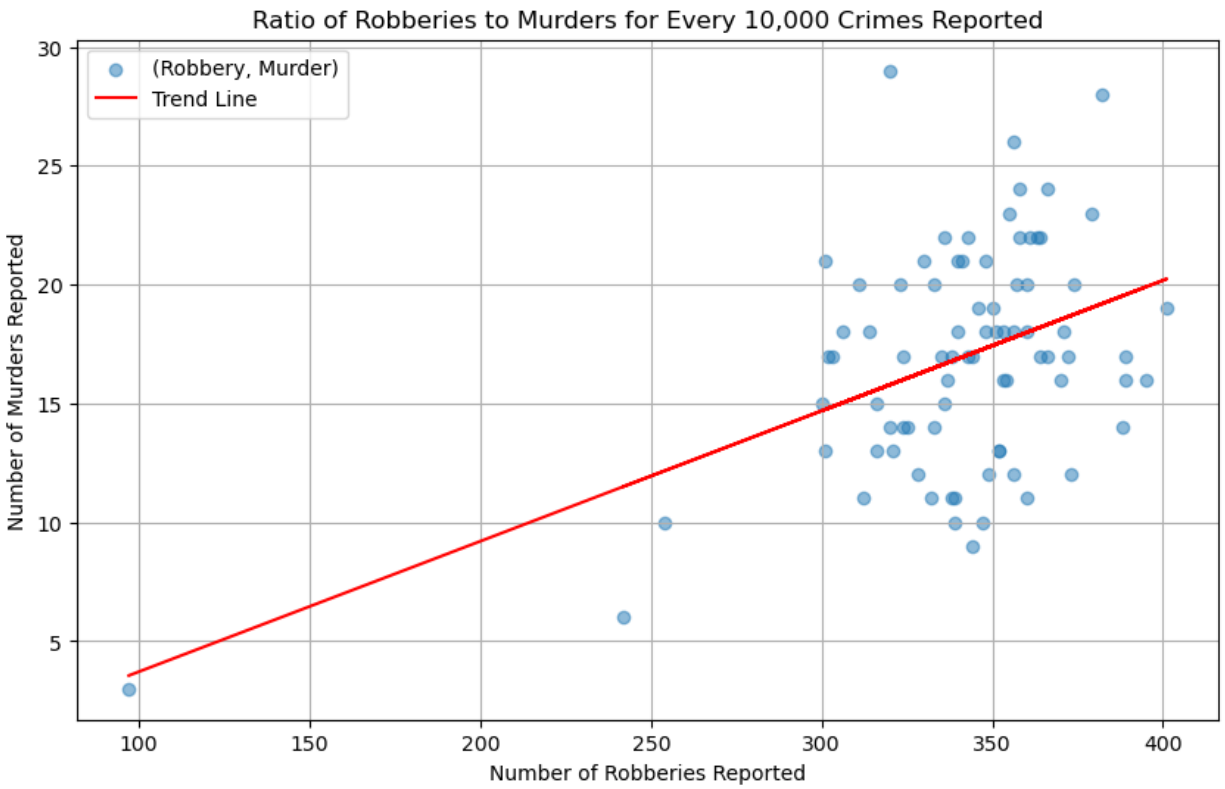
Robberies to Murders Distribution

The Ratio of Robberies to Murders

Understanding the correlation between robberies and murder can help the PD understand the relationship between the two crimes. If there is an increase in the number of robberies happening in a neighborhood, then it can be presumed that there will be an increase in the number of violent crimes, like homicide. It can be assumed that the robberies were done with malicious intent. On the flip side, if there is not a high number of homicides associated with many robberies, other factors may be at play. For example, a high number of robberies might occur in lower socioeconomic neighborhoods, meaning more government funding should be allocated to such areas to allow for development.

In [29]:
```
#lists to store the coordinates
rob = []
murder = []
```

```
#iterating by 10,000
rows = 10000

#for every 10,000 crimes reported, how many are robbery to how many are murder
for i in range(0, len(df), rows):
    r = df[i:i + rows]
    rob.append(r['Crm Cd'].eq(210).sum())
    murder.append(r['Crm Cd'].eq(110).sum())

#scatter plot
plt.figure(figsize=(10, 6))
plt.scatter(rob, murder, marker='o', alpha=0.5, label=f'(Robbery, Murder)')
slope, intercept, r_value, p_value, std_err = linregress(rob, murder)
line = slope * np.array(rob) + intercept
plt.plot(rob, line, color='red', label='Trend Line')

# axis labels and title
plt.xlabel('Number of Robberies Reported')
plt.ylabel('Number of Murders Reported')
plt.title('Ratio of Robberies to Murders for Every 10,000 Crimes Reported')
plt.grid(True)
plt.legend()

#output
plt.show()
```



Time Series Analysis

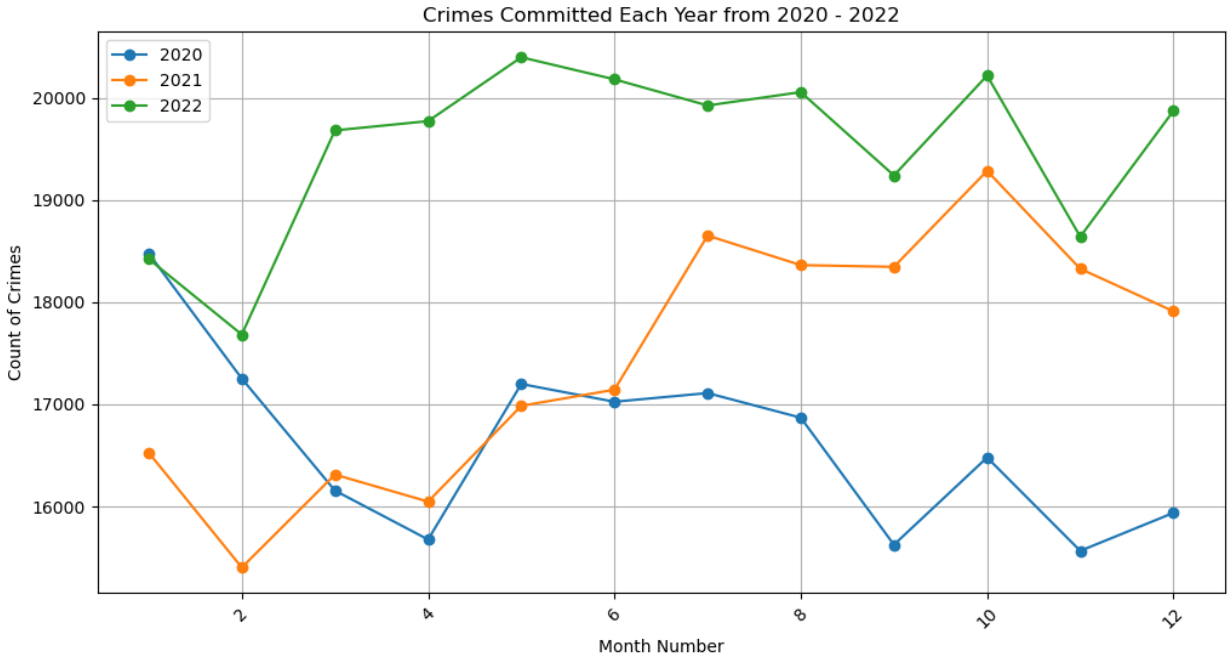
Crime Trends in Los Angeles

The Number of Crimes Reported Each Year

Understanding historical trends can help gauge areas of improvement at a high level. If there is a significant increase in the number of crimes from past years, more patrolling and resources need to be used. If there is a decrease in the number of crimes, then it can be assumed that the current techniques being used are effective in protecting the Los Angeles area.

In [104…
```
import geopandas as gpd
from shapely.geometry import Point, Polygon

%matplotlib inline
```

In [30]:
```
df['DATE OCC'] = pd.to_datetime(df['DATE OCC'])

#dates from 2020 - today
start_date = '2020-01-01'
end_date_22 = datetime.now().strftime('2022-12-31')

line_data = df[(df['DATE OCC'] >= start_date) & (df['DATE OCC'] <= end_date_22)

#crime counts grouped by the months
crime_cnt_month = line_data.groupby([line_data['DATE OCC'].dt.year, line_data['
crime_cnt_month = crime_cnt_month.unstack(level=0, fill_value=0)

#each year is a line on the graph
plt.figure(figsize=(12, 6))
for year in crime_cnt_month.columns:
    plt.plot(crime_cnt_month.index, crime_cnt_month[year], label=str(year), mar

#axis label and title
plt.xlabel('Month Number')
plt.ylabel('Count of Crimes')
plt.title('Crimes Committed Each Year from 2020 - 2022')

plt.xticks(rotation=45)

#legend
plt.legend()

#output
plt.grid(True)
#plt.tight_layout()
plt.show()
```



Plotting Data on a Map

Homocide Frequency Trends in Los Angeles by Area

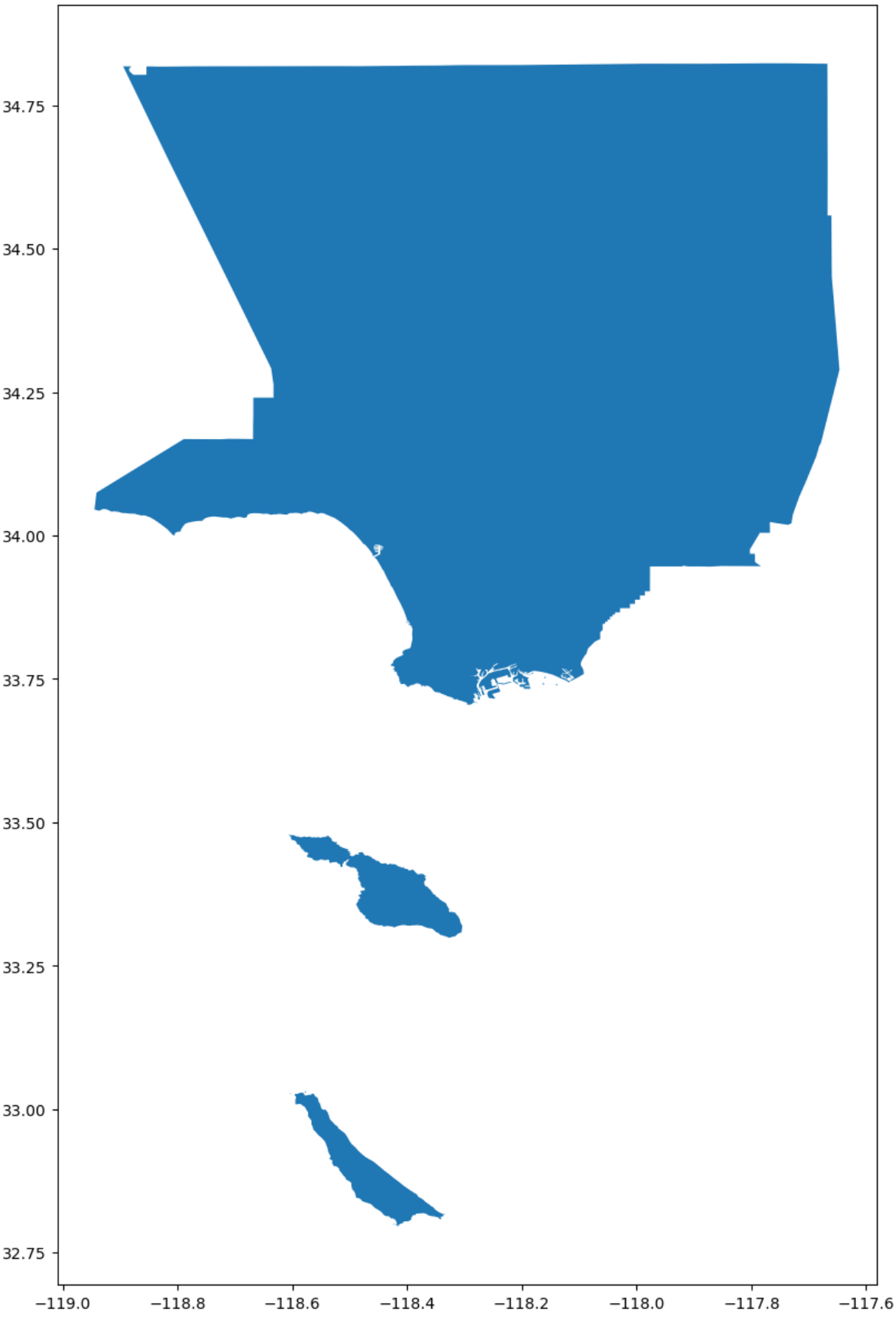
tried following this tutorial for creating a map with data points plotted

https://towardsdatascience.com/geopandas-101-plot-any-data-with-a-latitude-and-longitude-on-a-map-98e01944b972

In [173…
```
la_map = gpd.read_file('/Users/alaya/Desktop/data sci project/County_Boundary.s
```

In [174…
```
fig, ax = plt.subplots(figsize=(15,15))
la_map.plot(ax=ax)
```

Out[174]: <AxesSubplot:>



In [175…

```
crs = 'epsg:4326'
df.head()
```

Out[175]:

	DR_NO	Date Rptd	DATE OCC	TIME OCC	AREA	AREA NAME	Rpt Dist No	Part 1-2	Crm Cd	Crm Cd Desc	.
0	10304468	01/08/2020 12:00:00 AM	2020-01-08	2230	3	Southwest	377	2	624	BATTERY - SIMPLE ASSAULT	.
1	190101086	01/02/2020 12:00:00 AM	2020-01-01	330	1	Central	163	2	624	BATTERY - SIMPLE ASSAULT	.
2	200110444	04/14/2020 12:00:00 AM	2020-02-13	1200	1	Central	155	2	845	SEX OFFENDER REGISTRANT OUT OF COMPLIANCE	.
3	191501505	01/01/2020 12:00:00 AM	2020-01-01	1730	15	N Hollywood	1543	2	745	VANDALISM - MISDEAMEANOR ($399 OR UNDER)	.
4	191921269	01/01/2020 12:00:00 AM	2020-01-01	415	19	Mission	1998	2	740	VANDALISM - FELONY ($400 & OVER, ALL CHURCH VA...	.

5 rows × 29 columns

In [176…
```
df = df.drop(df[df['LAT'] <= 20].index)
```

In [177…
```
geometry = [Point(xy) for xy in zip(df['LON'], df['LAT'])]
```

In [178…
```
geo_df = gpd.GeoDataFrame(df, #specify our data
                          crs = crs, #specify our coordinate reference system
                          geometry = geometry) #specify the geometry list we cr
geo_df.head()
```

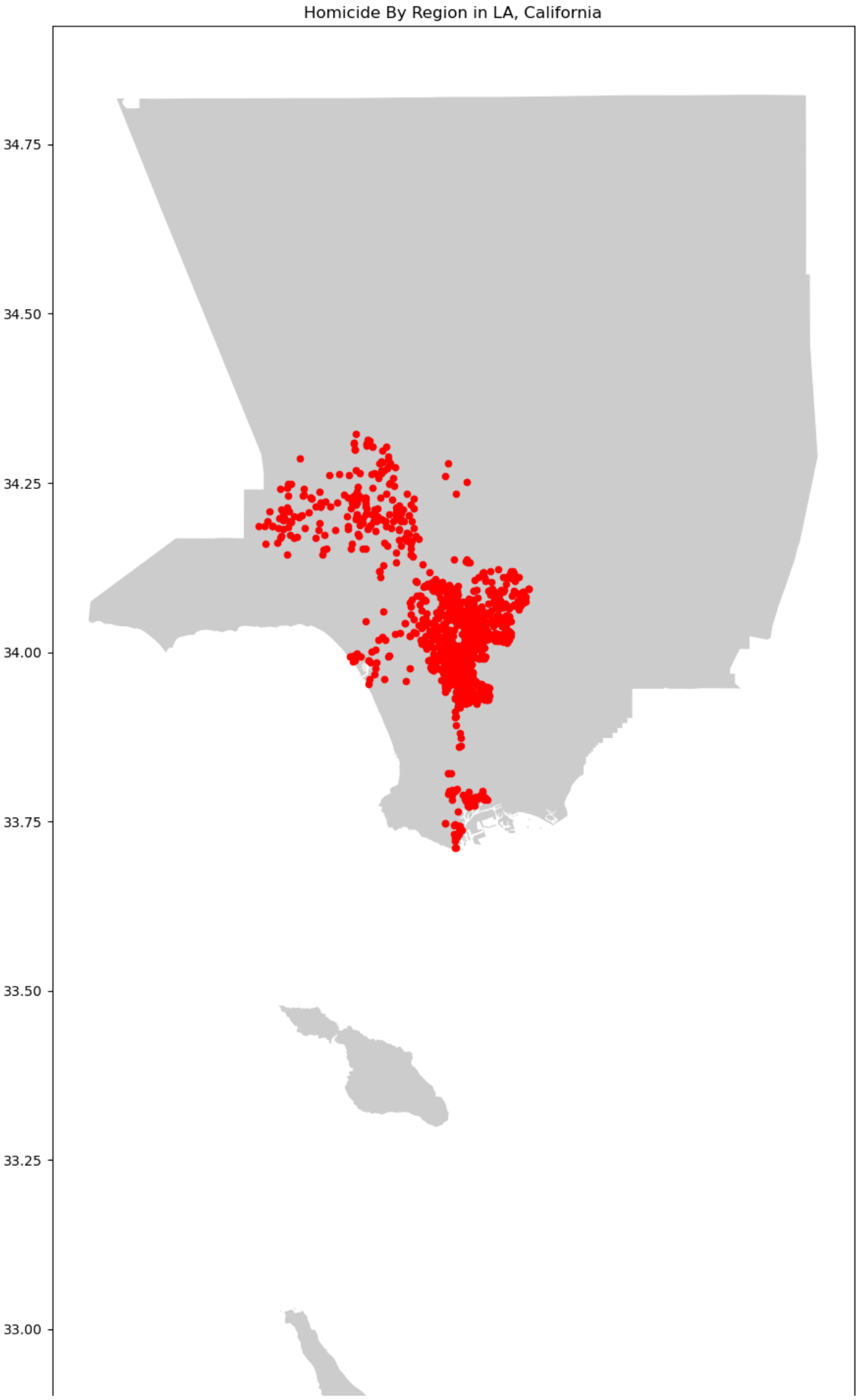
Out[178]:

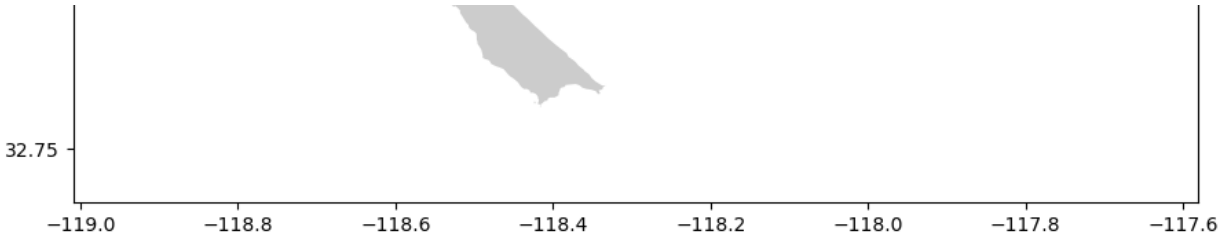
	DR_NO	Date Rptd	DATE OCC	TIME OCC	AREA	AREA NAME	Rpt Dist No	Part 1-2	Crm Cd	Crm Cd Desc	.
0	10304468	01/08/2020 12:00:00 AM	2020-01-08	2230	3	Southwest	377	2	624	BATTERY - SIMPLE ASSAULT	.
1	190101086	01/02/2020 12:00:00 AM	2020-01-01	330	1	Central	163	2	624	BATTERY - SIMPLE ASSAULT	.
2	200110444	04/14/2020 12:00:00 AM	2020-02-13	1200	1	Central	155	2	845	SEX OFFENDER REGISTRANT OUT OF COMPLIANCE	.
3	191501505	01/01/2020 12:00:00 AM	2020-01-01	1730	15	N Hollywood	1543	2	745	VANDALISM - MISDEAMEANOR ($399 OR UNDER)	.
4	191921269	01/01/2020 12:00:00 AM	2020-01-01	415	19	Mission	1998	2	740	VANDALISM - FELONY ($400 & OVER, ALL CHURCH VA...	.

5 rows × 30 columns

In [210…

```
fig, ax = plt.subplots(figsize=(15, 20))
la_map.plot(ax=ax, alpha=0.4, color='grey')

#geo_df[geo_df['Crm Cd'] == 236].plot(ax=ax,
                                     #markersize=20,
                                     #color='blue',
                                     #marker='^',
                                     #label='INTIMATE PARTNER - AGGRAVATED ASSAULT
geo_df[geo_df['Crm Cd'] == 110].plot(ax=ax,
                                     markersize=20,
                                     color='red',
                                     marker='o',
                                     label='CRIMINAL HOMICIDE')

plt.title('Homicide By Region in LA, California')

plt.show()
```


Homicide By Region in LA, California
34.75
34.50
34.25
34.00
33.75
33.50
33.25
33.00



In [211…
```
#I think that the lat and long values are too close together for this map to be
```

Looking at the map above, we can see that most of the homicides that have been reported have occurred in Central LA. Knowing this, the LAPD can allocate more resources to neighborhoods with higher murder rates. For further analysis, the murder rate can be compared with the poverty rates of areas in LA to see if there are patterns between socioeconomic status and violent crimes. With this information, there can be a better allocation of resources and support in lower-income areas to reduce the poverty rate and, subsequently, the violent crime rates.

Predicitive Analysis

Predicting the Target Variable - Crime Code Description

Using the dataset to create a model that predicts the Crime Code Description (crime description) using a crime's other features (attributes).

In the model built in Orange, I will create two datasets - one for training and one for testing the model. When setting the parameters for the model, I also skipped the Deg of Crm since that was an experimental column with mostly null values. I also skipped the following values: Crm Cd and Crm Cd 1-4 (crime code and additional crime codes). Crm Cd and Crm Cd Desc are synonymous, so if the model knows the Crm Cd, it can predict Crm Cd Desc with 100% accuracy.

With the rest of the attributes found in the dataset, the prediction model was able to predict the nature of the crimes with 79.8% accuracy. I split the data into two parts - one part for training and one for testing. I used 20% of the dataset to train the model and 20% of the data to test the model.

I used a classification tree to train the model since I was trying to predict categorical data values using primarily categorical data as inputs.

Note: Orange can only take 16 unique values for each categorical attribute so I filtered the data to include only the top 16 values from each column.

In [88]:
```
#orange can only take 16 unique values for each categorical attribute
#filtering the data to include only the top 16 values from each column w more t

top_area_names = df['AREA NAME'].value_counts().nlargest(16).index
top_descent_names = df['Vict Descent'].value_counts().nlargest(16).index
```

```
top_weapon_names = df['Weapon Desc'].value_counts().nlargest(16).index
top_crm_names = df['Crm Cd Desc'].value_counts().nlargest(16).index

#creating the filtered data frame here
filtered_df = df[df['AREA NAME'].isin(top_area_names) & df['Vict Descent'].isir

#filtered_df = df[df['AREA NAME'].isin(top_area_names)

#saving the df to a tester csv (for me to review)
filtered_df.to_csv("tester.csv", index = False)
```

In [91]:
```
#data splitting ---> use this in orange for predictive analysis

from sklearn.model_selection import train_test_split

#setting the training data set to 25% and test data to 20% of all the data
train, test = train_test_split(filtered_df, train_size = 0.25, test_size = 0.2,

#saving to csv files to use in orange
train.to_csv("train_crime_data.csv", index = False)
test.to_csv("test_crime_data.csv", index = False)
```

Insights

From the histogram, it can be seen that 25 to 30-year-olds are disproportionately affected by crimes in Los Angeles. More resources should be allocated to neighborhoods with young couples and students.

From the bar chart and pie chart, it can be seen that most of the victims were of Hispanic, White, and African descent (respectively). Hispanic individuals have disproportionately been victims of crimes in the Los Angeles area, with 36.1% of crimes having Hispanic victims.

Comparatively, the smallest percentage of crimes have Asian victims, with 2.6% having Asian victims.

From these visualizations, it can be assumed that a Hispanic individual is most likely to be a victim of a crime reported in the Los Angeles area. In contrast, an individual of Asian descent is least likely to be a victim.

From the crimes reported in the Los Angeles area, there is a fairly even distribution between male and female victims of crimes. It can be inferred that though the crimes may vary in nature, both men and women are equally at risk of being a victim of a crime.

To gain further insights into this dataset, we can create visualizations to see the distribution of gender and age to see if individuals of specific demographics are more likely to be a target of a crime. Similarly, we can also create visualizations to see how gender and type of crime might be related to each other.

Crime tends to spike, historically speaking, towards the middle of the year - summer to fall time. More resources and patrolling efforts can be made during this time to reduce the number of crimes committed.

There is a positive correlation between the number of murders reported, and the number of robberies reported. If a neighborhood is reporting a larger number of robberies, then it can be assumed that the number of violent crimes will also increase proportionally. Though the relationship between the two variables is only moderately strong, there should still be an increase in patrolling efforts since the risk of "failure" to do so is high (potentially life-threatening).

The majority of crimes reported in Los Angeles are still under investigation. Central has the highest registered number of crimes, while Foothill has the lowest number. Most crimes with charges associated with them include an adult being detained or arrested (or other). Juvenile arrests take up a very insignificant portion of investigations.

Central LA also has a higher frequency of murders that have been reported, as seen in the map above. More patrolling and better allocation of resources can help reduce this number. The poverty rate in Central LA is 12.2%, which is more than double the rate in California and the United States. Noticing such patterns can help better understand where support and resources should be allocated.

The predictive model can help investigators better understand crime trends and MOs (modus operandi). By understanding which groups of individuals and areas are more likely to be targets of what crimes, police departments and operatives can be better equipped in their investigations while enabling them to take precautions and prevent similar crimes.

For example, suppose young, minority men in an area are predicted to be violent crime victims. In that case, resource allocation can be better in densely populated areas with such demographics. By understanding the profiles of victims of various crimes, there can be better policing initiatives taken and training to assess these issues.

The predictive model can help investigators better understand crime trends and MOs (modus operandi). By understanding which groups of individuals and areas are more likely to be targets of what crimes, police departments and operatives can be better equipped in their investigations while enabling them to take precautions and prevent similar crimes. For example, suppose young, minority men in an area are predicted to be violent crime victims. In that case, resource allocation can be better in densely populated areas with such demographics. By understanding the profiles of victims of various crimes, there can be better policing initiatives taken and training to assess these issues. Such crime forecasting can also be used to understand risks, improve public awareness, and help with predictive policing as a whole.