

Data Exploration-Segment 2

The data source we chose is the Austin Crime Data from 2017-2022. The dataset for this timeframe is comprised of 1 crime table containing 200,000+ rows of data and 27 columns. The data was loaded in from the CSV file using Pandas. It was also cleaned and transformed using Pandas. The data was then loaded into the database and queried using Postgres. Once loaded into the database, it was then hosted locally on Postgres.

For segment 2, Using Pandas, several columns were dropped to remove data that we didn't feel was applicable to our analysis. The dataset, as seen below, was then split into two tables in preparation of fulfilling the join requirement.

1 SELECT * FROM public."crime reports"

2 LIMIT 100

3

Scratch Pad

Data Output

Explain

Messages

Notifications

	index bigint	Incident Number bigint	Highest Offense Code bigint	Occurred Date text	Occurred Time bigint	Report Date text	Report Time bigint	Location Type text	Zip Code double precision	Council District double precision	APD Sector text	APC text
1	0	2017471291	600	02/16/2017	1800	02/16/2017	1822	PARKING /DROP LOT/ GARAGE	78721		1 CH	4
2	1	20172171183	607	08/05/2017	1800	08/05/2017	1800	SHOPPING MALL	78746		8 DA	5
3	2	20205047456	607	11/20/2020	1330	11/20/2020	1522	DEPARTMENT / DISCOUNT STORE	78748		5 FR	2
4	3	20191441181	700	05/23/2019	2000	05/24/2019	1343	PARKING /DROP LOT/ GARAGE	78705		9 BA	2
5	4	20175050403	601	12/19/2017	2200	12/20/2017	626	PARKING /DROP LOT/ GARAGE	78741		3 HE	2
6	5	201940773	402	01/04/2019	730	01/04/2019	1304	HWY / ROAD / ALLEY/ STREET/ SIDEWALK	78701		3 CH	1
7	6	20208001307	601	03/20/2020	1955	03/20/2020	2248	GOVERNMENT / PUBLIC BUILDING	78731		10 BA	7
8	7	20188004000	601	10/12/2018	750	10/12/2018	1003	PARKING /DROP LOT/ GARAGE	78741		3 HE	4
9	8	20172330290	601	08/21/2017	830	08/21/2017	830	RESIDENCE / HOME	78741		3 HE	2
10	9	20195049971	600	10/06/2019	1200	12/06/2019	812	RESIDENCE / HOME	78729		6 AD	4
11	10	20193140226	600	11/10/2019	225	11/10/2019	225	HWY / ROAD / ALLEY/ STREET/ SIDEWALK	78701		9 GE	2
12	11	20205050707	600	12/09/2020	1100	12/11/2020	1630	BANK / SAVINGS & LOAN	78702		3 CH	2
13	12	20188000144	601	01/03/2018	1605	01/10/2018	1625	PARKING /DROP LOT/ GARAGE	78757		7 BA	6
14	13	20188002761	600	07/07/2018	6	07/13/2018	1105	BAR / NIGHTCLUB	78701		9 GE	2
15	14	20208005193	620	10/19/2020	2200	10/29/2020	1254	PARKING /DROP LOT/ GARAGE	78745		2 FR	3
16	15	20185051374	601	12/20/2018	1820	12/20/2018	2101	HWY / ROAD / ALLEY/ STREET/ SIDEWALK	78704		9 DA	1
17	16	20202460149	600	09/02/2020	347	09/02/2020	421	PARKING /DROP LOT/ GARAGE	78703		9 BA	2
18	17	20205046661	600	11/12/2020	1430	11/16/2020	1329	COMMERCIAL / OFFICE BUILDING	78741		2 FR	6
19	18	20185047132	601	11/23/2018	1830	11/23/2018	2110	PARKING /DROP LOT/ GARAGE	78745		2 FR	4
20	19	20215008058	601	03/01/2021	0	03/03/2021	807	PARKING /DROP LOT/ GARAGE	78759		7 AD	3
21	20	201990714	500	04/09/2019	1206	04/09/2019	1206	RESIDENCE / HOME	78744		2 FR	5
22	21	2018760682	601	03/16/2018	2300	03/17/2018	919	PARKING /DROP LOT/ GARAGE	78753		4 ED	1
23	22	2020540751	500	02/22/2020	2245	02/23/2020	1003	RESIDENCE / HOME	78758		4 ED	2
24	23	20195050418	601	12/08/2019	1630	12/09/2019	1343	PARKING /DROP LOT/ GARAGE	78753		7 ED	4
25	24	20213450457	700	12/11/2021	634	12/11/2021	634	PARKING /DROP LOT/ GARAGE	78758		4 ED	4
26	25	20185010306	607	03/18/2018	1345	03/19/2018	1419	DEPARTMENT / DISCOUNT STORE	78746		8 DA	5

Query Editor

Query History

1 SELECT * FROM public."family violence"

2 LIMIT 100

3

Scratch Pad

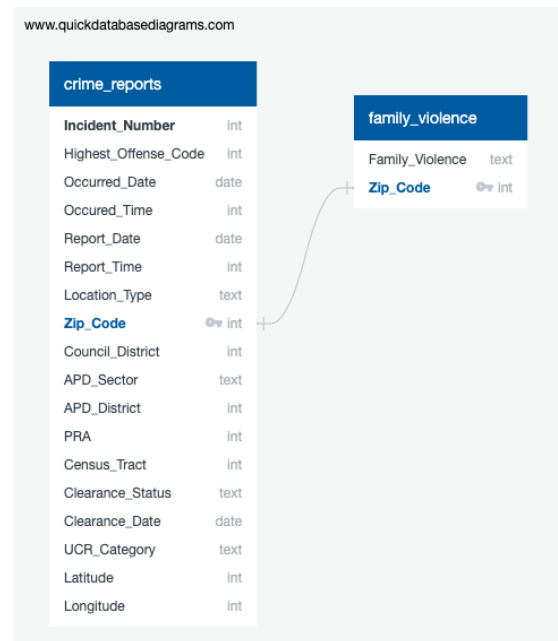
Data Output

Explain

Messages

	index bigint	Family Violence text	Zip Code double precision
1	0	N	78721
2	1	N	78746
3	2	N	78748
4	3	N	78705
5	4	N	78741
6	5	N	78701
7	6	N	78731
8	7	N	78741
9	8	N	78741
10	9	N	78729
11	10	N	78701
12	11	N	78702
13	12	N	78757
14	13	N	78701
15	14	N	78745
16	15	N	78704
17	16	N	78703
18	17	N	78741
19	18	N	78745
20	19	N	78759
21	20	N	78744
22	21	N	78753
23	22	N	78758
24	23	N	78753
25	24	N	78758

The ERD below was created in quickdatabasediagrams.com.



Finally, in order to fulfil the requirement for joining two tables, the dataset that was previously split into two tables was joined back together using SQL Query.

```
Query Editor  Query History

1  SELECT *
2  FROM public."family violence"
3  JOIN public."crime reports" ON
4  public."family violence"."Zip Code" = public."crime reports"."Zip Code"
5  WHERE public."crime reports"."Zip Code" = 78732;
```

Scratch Pad Data Output Explain Messages Notifications

Successfully run. Total query runtime: 125 msec.
441 rows affected.