Rohan Patel (010745904)

Prof. Kaikai Liu

Homework - 1
Big Query - Austin Crime Dataset

Colab link:

https://colab.research.google.com/drive/1Njc6TLVgHvAADqmD7ZTlHOE7McDCyNUi?usp=sharing

Query 1: Austin Crime Dataset

```
query1 = """ SELECT

 *

 FROM

    `zeta-time-368105.austin_crime_data.crime`

 LIMIT 1000

 """

dfquery = client.query(query1)
```

				1 to 5 of 5 entries	Filter								
index	unique_key	address	census_tract	clearance_date	clearance_status	council_district_code	description	district	latitude	longitude	location	loca
0	20165002898	Austin, TX	NaN	2016-03-14 12:00:00+00:00	Not cleared	<NA>	RAPE	UK	NaN	NaN	null	nan
1	20165004843	Austin, TX	NaN	2016-03-31 12:00:00+00:00	Not cleared	<NA>	RAPE	UK	NaN	NaN	null	nan
2	2016250988	Austin, TX	NaN	2016-09-16 12:00:00+00:00	Not cleared	<NA>	RAPE	UK	NaN	NaN	null	nan
3	20165003839	Austin, TX	NaN	2016-02-09 12:00:00+00:00	Not cleared	<NA>	AGG RAPE OF A CHILD	UK	NaN	NaN	null	nan
4	2016321746	Austin, TX	NaN	2016-03-22 12:00:00+00:00	Not cleared	<NA>	RAPE OF A CHILD	UK	NaN	NaN	null	nan

Query 2: Austin crime trend over the years

```
query2 = """ SELECT

unique_key, description, year, clearance_status, clearance_date

FROM

 `zeta-time-368105.austin_crime_data.crime`



WHERE

 year >= 2012

 AND year <= 2022



ORDER BY

 year
```

```
"""

dfquery = client.query(query2)
```

index	unique_key	description	year	clearance_status	clearance_date
				1 to 5 of 5 entries	Filter
0	20145022722	AGG ROBBERY/DEADLY WEAPON	2014	Not cleared	2014-07-03 12:00:00+00:00
1	2014390620	AGG ASLT STRANGLE/SUFFOCATE	2014	Cleared by Exception	2014-02-19 12:00:00+00:00
2	20145027962	BURGLARY OF RESIDENCE	2014		NaT
3	20142711475	ROBBERY BY ASSAULT	2014	Cleared by Exception	2014-10-01 12:00:00+00:00
4	20142910412	AGG ASLT STRANGLE/SUFFOCATE	2014	Cleared by Arrest	2014-10-29 12:00:00+00:00

Show 25 ∨ per page

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 18 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   unique_key            1000 non-null   Int64
 1   address               1000 non-null   object
 2   census_tract          477 non-null    float64
 3   clearance_date        840 non-null    datetime64[ns, UTC]
 4   clearance_status      840 non-null    object
 5   council_district_code 477 non-null    Int64
 6   description           1000 non-null   object
 7   district              1000 non-null   object
 8   latitude              269 non-null    float64
 9   longitude             269 non-null    float64
 10  location              269 non-null    object
 11  location_description  1000 non-null   object
 12  primary_type          1000 non-null   object
 13  timestamp             1000 non-null   datetime64[ns, UTC]
 14  x_coordinate          547 non-null    Int64
 15  y_coordinate          547 non-null    Int64
 16  year                  1000 non-null   Int64
 17  zipcode               477 non-null    object
dtypes: Int64(5), datetime64[ns, UTC](2), float64(3), object(8)
memory usage: 145.6+ KB
```
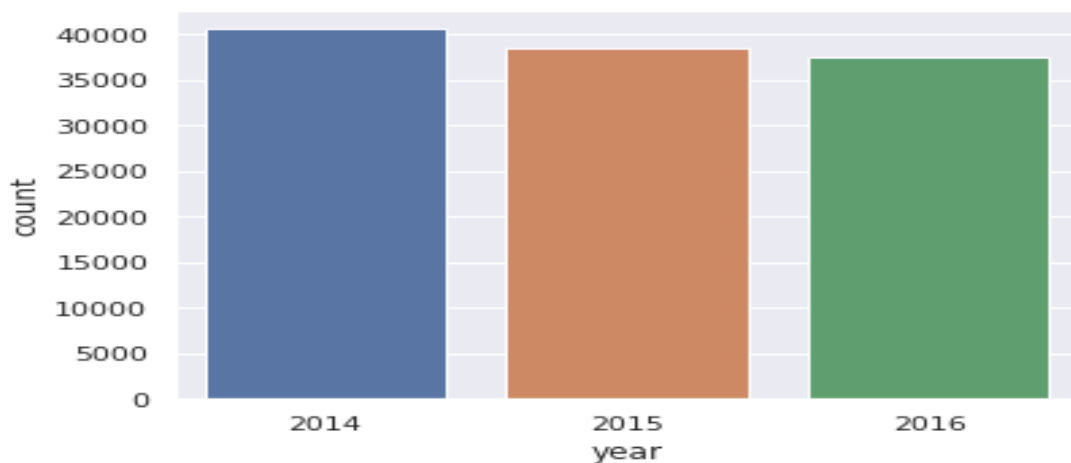
Query 3: 15 highest crime risk locations

```
query3 = """ SELECT
    location_description  AS loc,
    COUNT(unique_key) AS loc_count
FROM
 `zeta-time-368105.austin_crime_data.crime`
WHERE
 year < 2022
GROUP BY location_description
ORDER BY
 location_description


LIMIT 1000
    """

dfquery = client.query(query3)
```

1 to 5 of 5 entries Filter
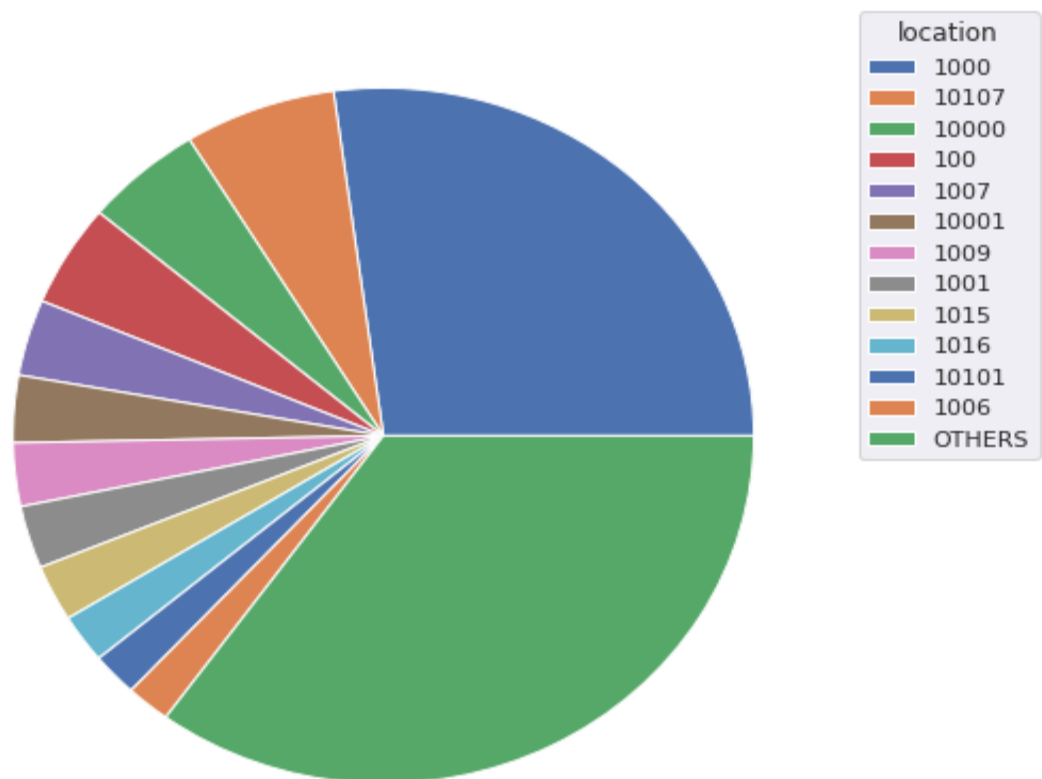
index	loc	loc_count
0	1 OLMOS DR	1
1	10 BLOCK N IH 35 NB	1
2	10 N PLEASANT VALLEY RD	1
3	10-4710 S LAMAR BLVD SVRD SB	1
4	100 ACADEMY DR	1

Show 25 ∨ per page

1 to 25 of 1000 entries Filter

index	loc	loc_count	location
0	1 OLMOS DR	1	1
1	10 BLOCK N IH 35 NB	1	10
2	10 N PLEASANT VALLEY RD	1	10
3	10-4710 S LAMAR BLVD SVRD SB	1	10-4710
4	100 ACADEMY DR	1	100
5	100 BASTROP HWY NB	1	100
6	100 BLK CHALMERS STREET	1	100
7	100 BLOCK BRUSHY ST	1	100
8	100 BLOCK CHALMERS AVE	1	100
9	100 BLOCK CHICON ST	1	100
10	100 BLOCK COLEMAN ST	1	100
11	100 BLOCK E 4TH ST	1	100
12	100 BLOCK E 5TH ST	4	100
13	100 BLOCK E 6TH ST	2	100
14	100 BLOCK E 7TH ST	4	100
15	100 BLOCK E CESAR CHAVEZ ST	1	100
16	100 BLOCK E MILTON ST	1	100
17	100 BLOCK E OLTORF ST	1	100
18	100 BLOCK E RUNDBERG LN	1	100

location	count
1000	914
10107	223
10000	173
100	160
1007	119
10001	105
1009	99
1001	98
1015	88
1016	76
10101	68
1006	65
1011	64
1003	62
101	60
1005	55
1012	48
1008	40
1004	40
10100	40
1010	38
10201	36
10111	35
102	33
10205	33

Show 25 per page

15 highest crime risk location in 2022

location
- 1000
- 10107
- 10000
- 100
- 1007
- 10001
- 1009
- 1001
- 1015
- 1016
- 10101
- 1006
- OTHERS

Query 4: Crime type - 2022

```python
query4 = """
SELECT primary_type AS crime_type, count(unique_key) AS count

FROM `zeta-time-368105.austin_crime_data.crime`

WHERE

 year < 2022

GROUP BY

 primary_type



LIMIT 1000

    """



dfquery = client.query(query4)

df = dfquery.to_dataframe()

df
```

index	crime_type	count
0	Rape	1801
1	Theft	54514
2	Robbery	2859
3	Aggravated Assault	1888
4	Agg Assault	4092
5	Auto Theft	6231
6	Burglary / Breaking & Entering	5670
7	Burglary	10098
8	Theft: All Other Larceny	13539
9	Theft: BOV	10545
10	Theft: Shoplifting	4284
11	Theft: from Building	221
12	Theft: Pocket Picking	448
13	Murder	61
14	Theft: Auto Parts	231
15	Theft: Coin Op Machine	153
16	Theft: Purse Snatching	7
17	Homicide: Murder & Nonnegligent Manslaughter	32
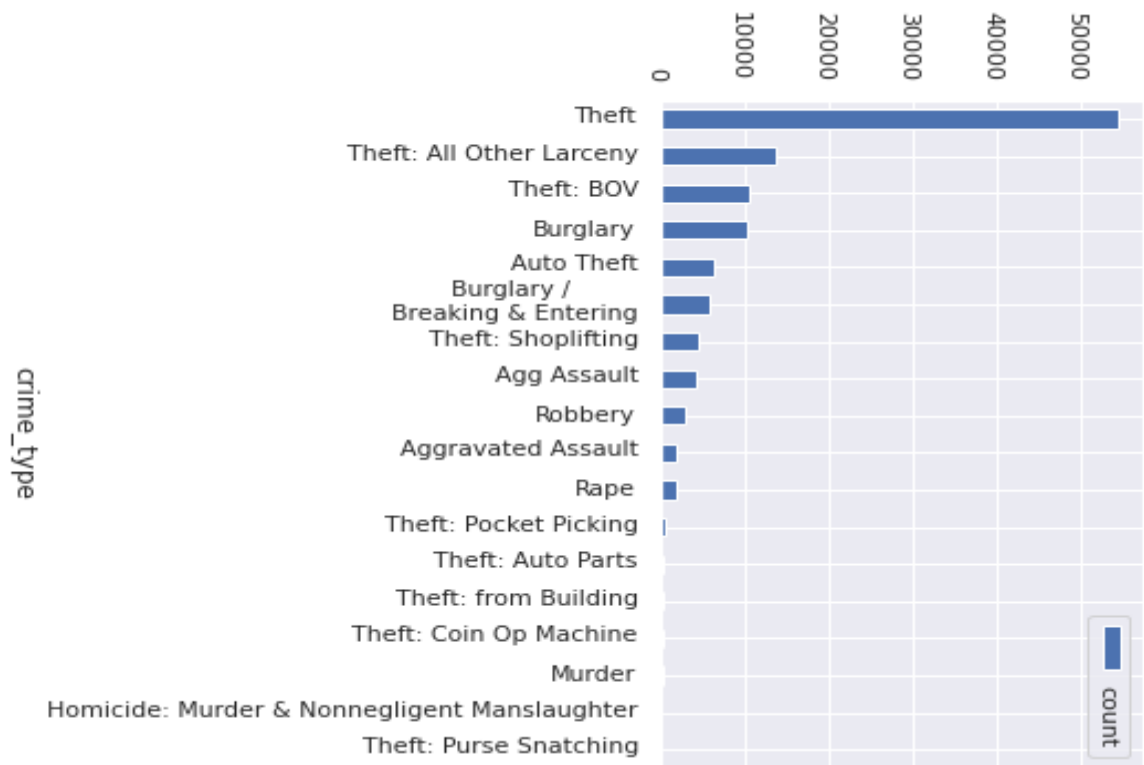
Interactive Pie Chart:

High risk locations in Austin

Others
1000

27.2%

72.8%
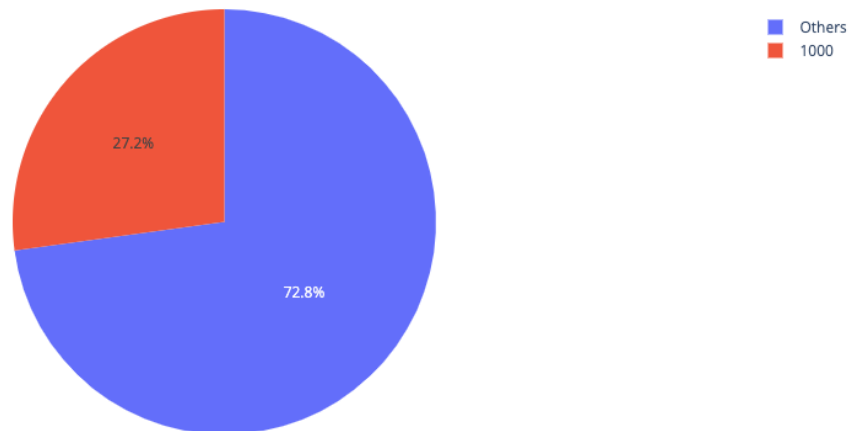
Query 5: Crime heat map for Austin until the year 2022

```
query5 = """ SELECT
 count(unique_key) AS count,
 latitude, longitude
FROM
 `zeta-time-368105.austin_crime_data.crime`

WHERE
 year < 2022

GROUP BY
 latitude, longitude

LIMIT
 1000 """

dfquery = client.query(query5)
```

<div align="right">1 to 5 of 5 entries   Filter</div>

index	count	latitude	longitude
0	82174	NaN	NaN
1	28	30.264979	-97.746598
2	1	30.242457	-97.78394
3	1	30.236083	-97.844218
4	1	30.281456	-97.741

Show 25 ∨ per page