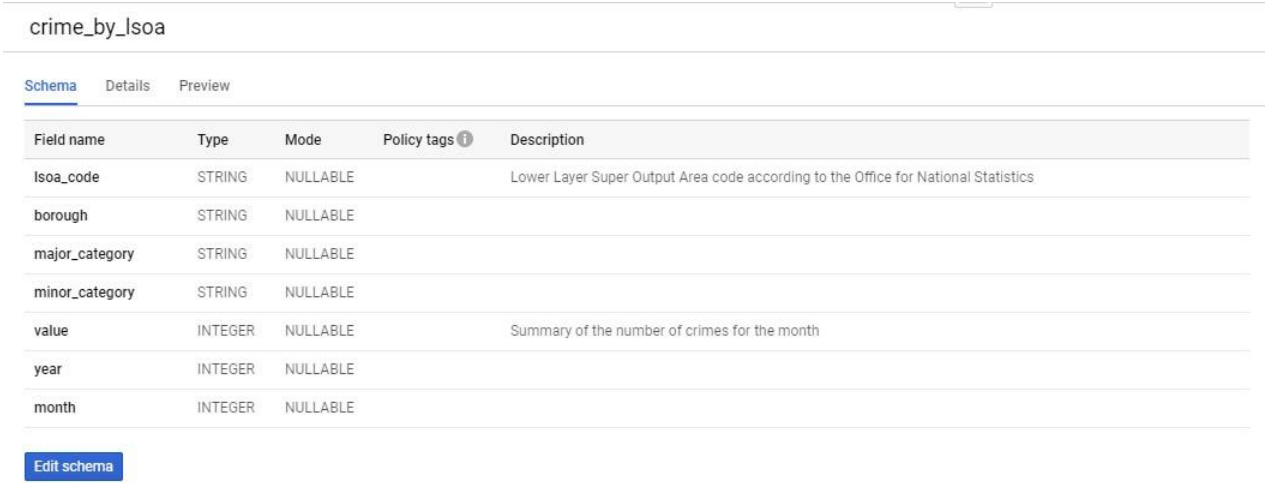
IS 6850-008 Advanced SQL for Analytics

Final Project Report

Name: **Huzefa Saifee**
UNID: **u1274086**

- **Dataset:**
  I am have worked on the public dataset named **London Crime Data** hosted on BigQuery public datasets to analyze different crimes occurred over a period of time.

The analysis performed on the dataset can help the govt. and the forces in tackling crime effectively. For example, by analyzing which type of crimes occur more in which parts of the city to deploy specialized personals/troops in those areas to handle the crime in a much better way possible. Also, it can help in determining whether the steps taken over the years have been effective in reducing the crime or not.



crime_by_lsoa

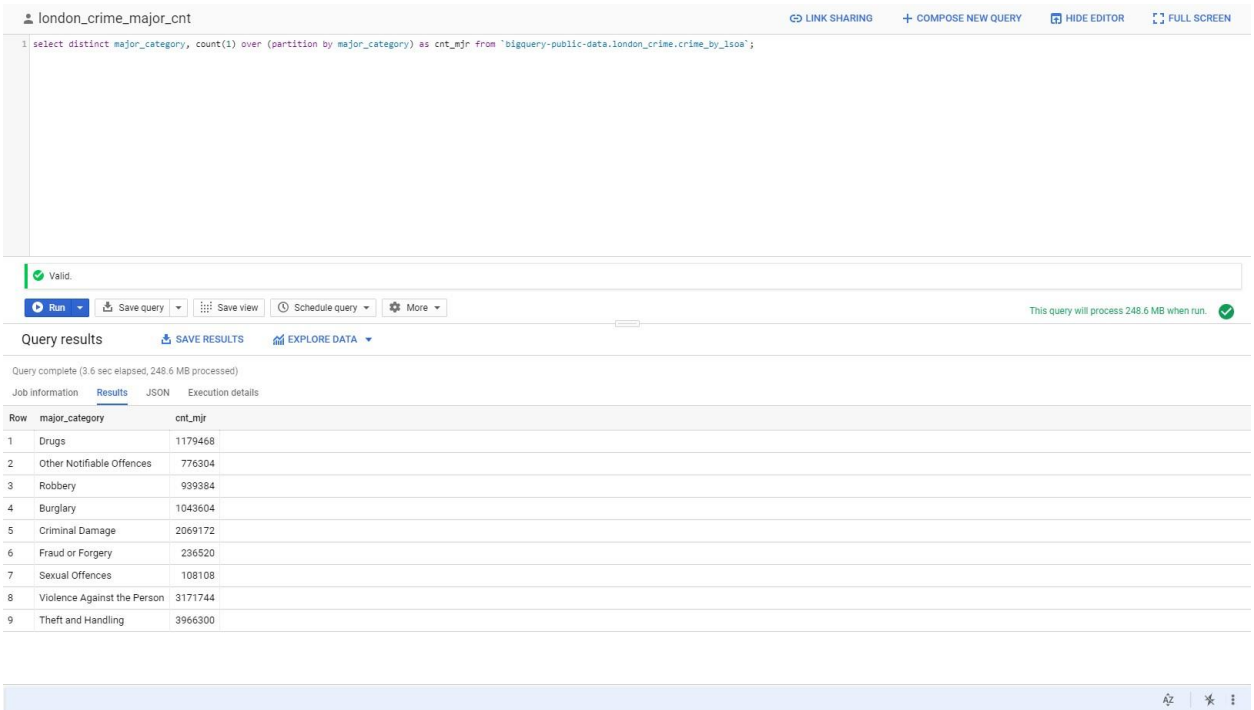
Schema Details Preview

Field name	Type	Mode	Policy tags	Description
lsoa_code	STRING	NULLABLE		Lower Layer Super Output Area code according to the Office for National Statistics
borough	STRING	NULLABLE		
major_category	STRING	NULLABLE		
minor_category	STRING	NULLABLE		
value	INTEGER	NULLABLE		Summary of the number of crimes for the month
year	INTEGER	NULLABLE		
month	INTEGER	NULLABLE		

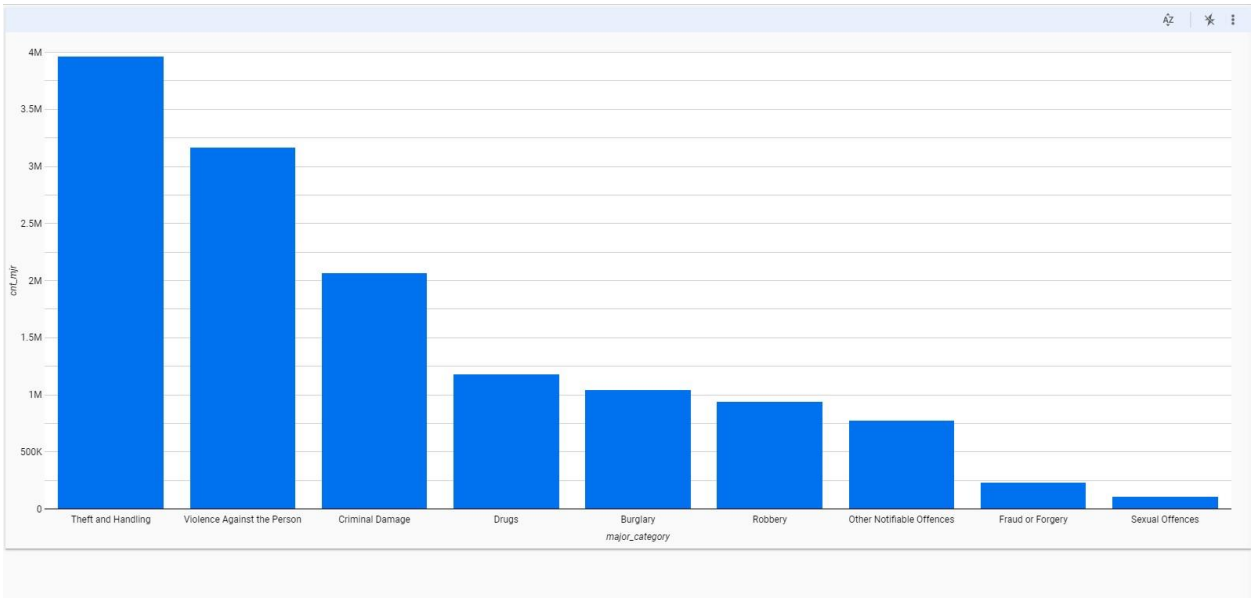
Edit schema

Question 1: Which major categories have how many crimes recorded over the years?

```
SELECT DISTINCT major_category,
  COUNT(1) OVER (PARTITION BY major_category) AS cnt_mjr
FROM `bigquery-public-data.london_crime.crime_by_lsoa`;
```

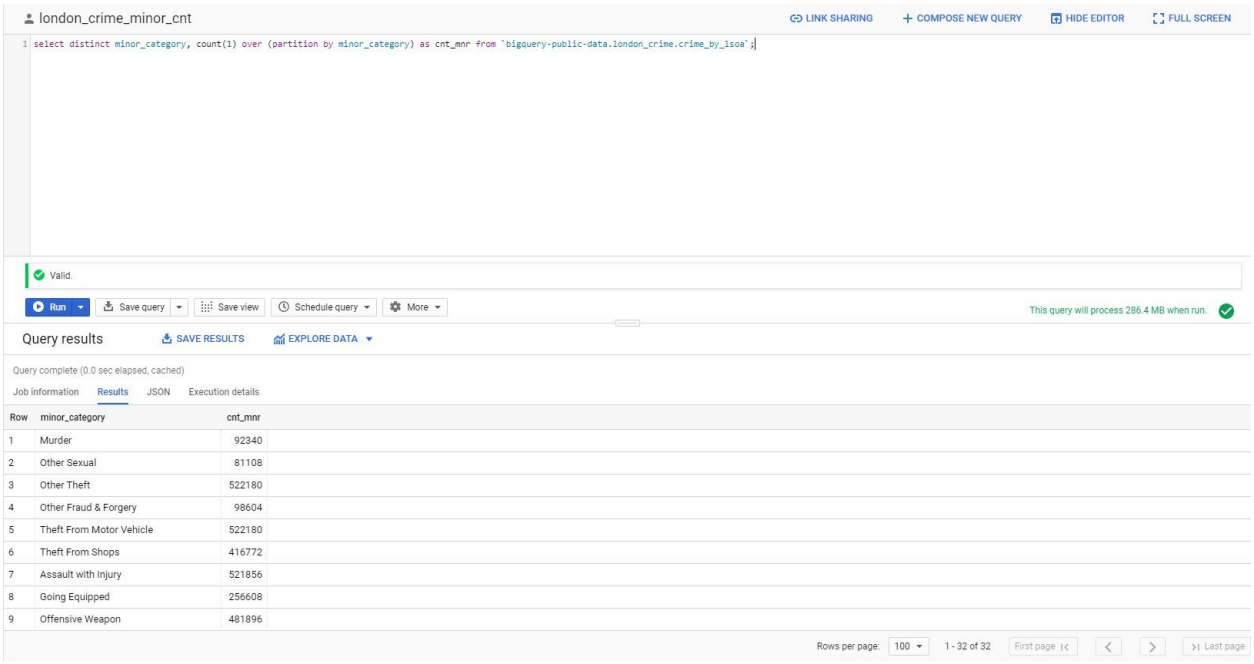


Row	major_category	cnt_mjr
1	Drugs	1179468
2	Other Notifiable Offences	776304
3	Robbery	939384
4	Burglary	1043604
5	Criminal Damage	2069172
6	Fraud or Forgery	236520
7	Sexual Offences	108108
8	Violence Against the Person	3171744
9	Theft and Handling	3966300

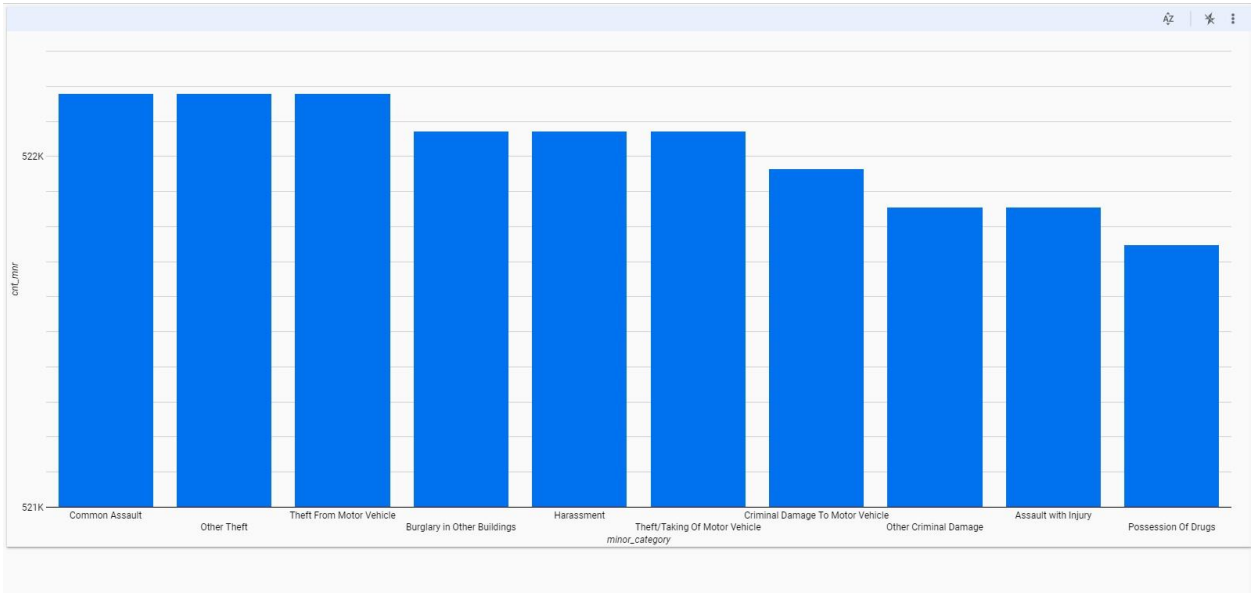


Question 2: Which minor categories have how many crimes recorded over the years?

```
SELECT DISTINCT minor_category,
  COUNT(1) OVER (PARTITION BY minor_category) AS cnt_mnr
FROM `bigquery-public-data.london_crime.crime_by_lsoa`;
```

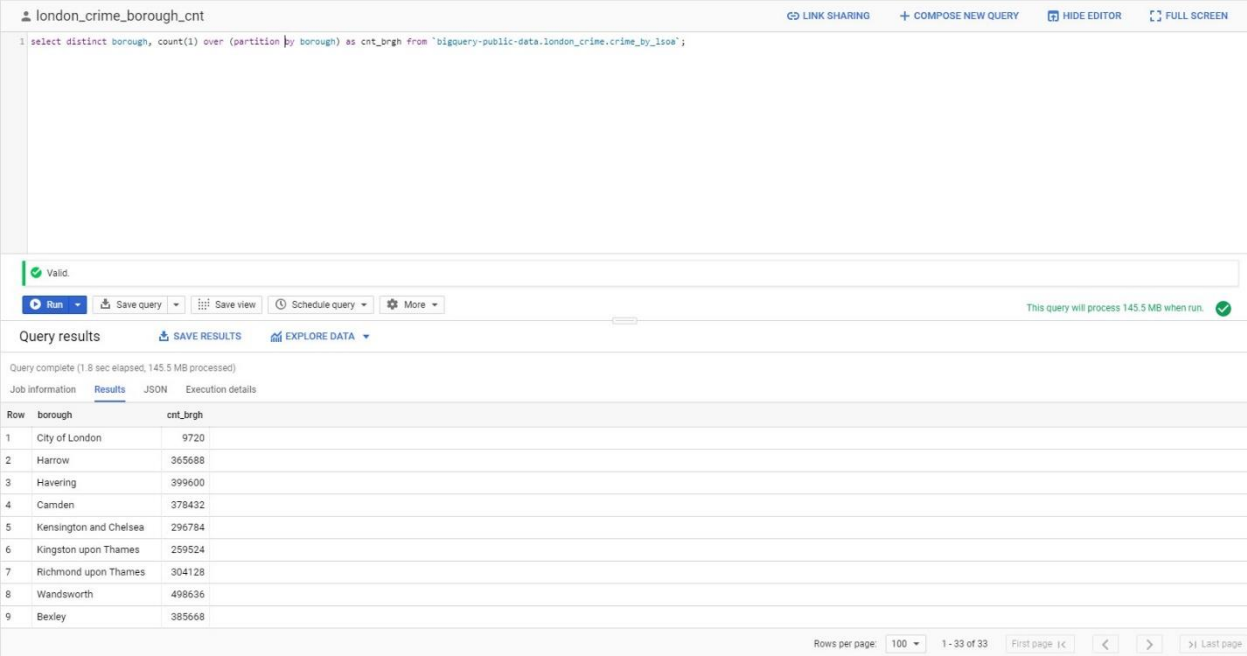


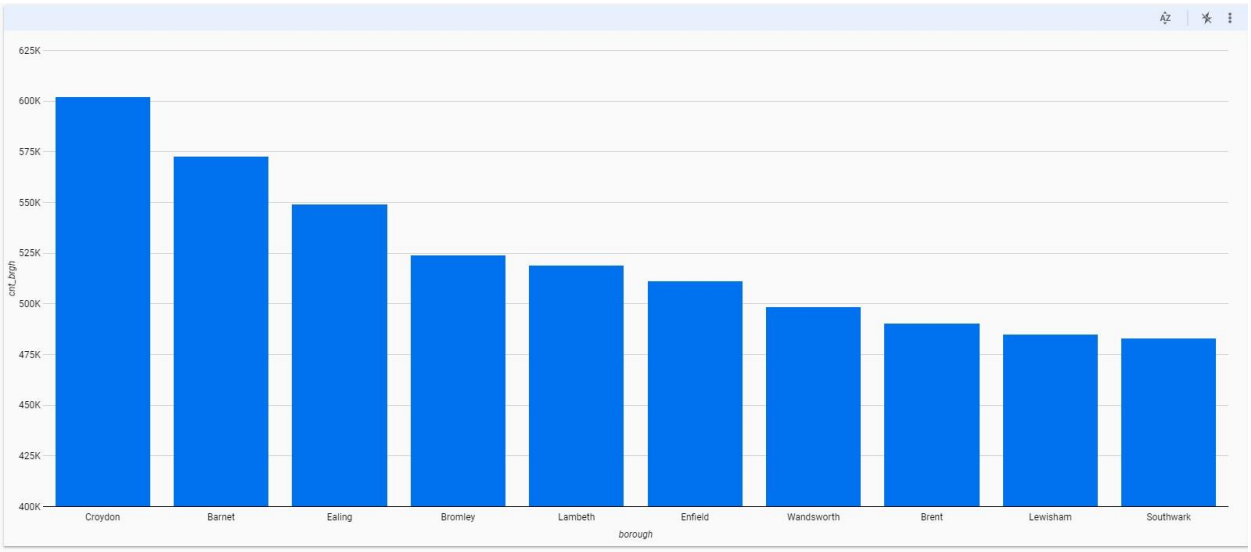
Row	minor_category	cnt_mnr
1	Murder	92340
2	Other Sexual	81108
3	Other Theft	522180
4	Other Fraud & Forgery	98604
5	Theft From Motor Vehicle	522180
6	Theft From Shops	416772
7	Assault with Injury	521856
8	Going Equipped	256608
9	Offensive Weapon	481896



Question 3: Which boroughs have how many crimes recorded over the years?

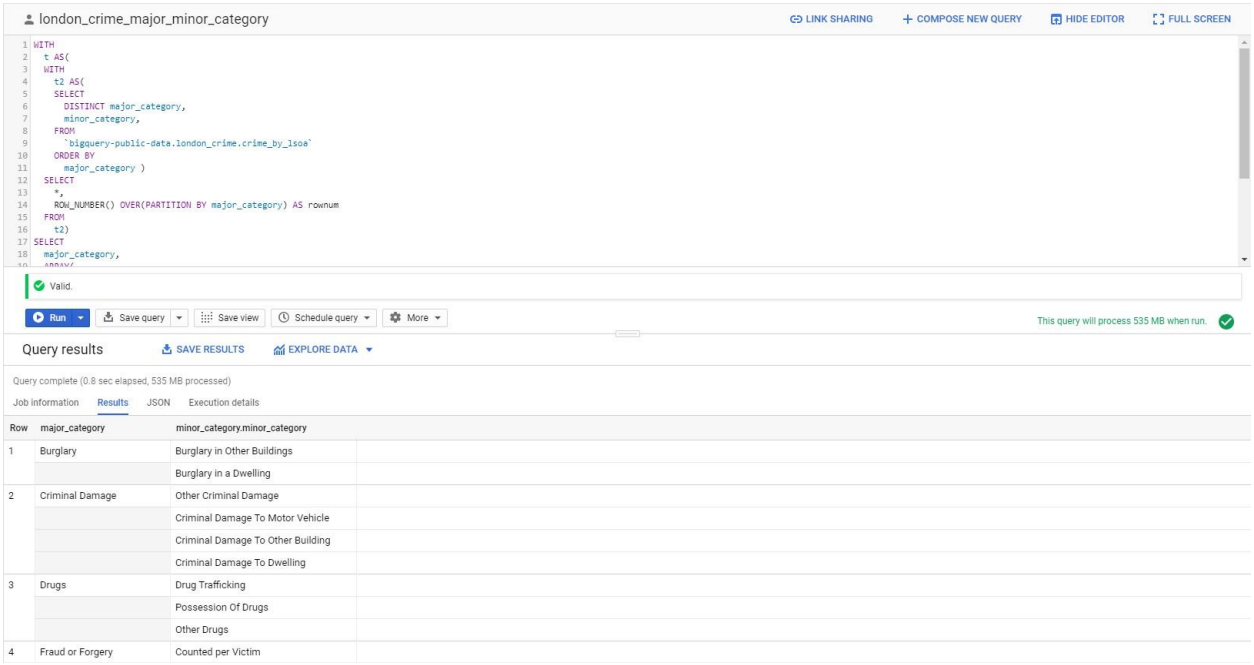
```
SELECT DISTINCT borough,
  COUNT(1) OVER (PARTITION BY borough) AS cnt_brgh
FROM `bigquery-public-data.london_crime.crime_by_lsoa`;
```



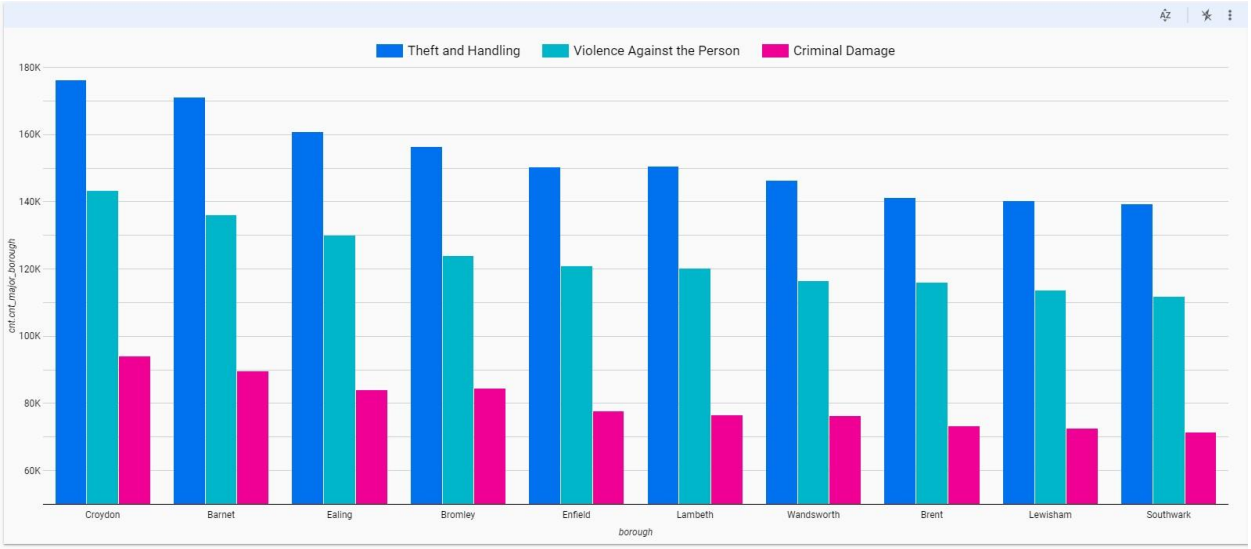


Question 4: Which major category has how many cases recorded in every borough?

```
WITH t2 AS (
  WITH t AS (
    SELECT DISTINCT borough, major_category,
      COUNT(1) OVER (PARTITION BY borough, major_category) AS cnt_major_borough,
    FROM `bigquery-public-data.london_crime.crime_by_lsoa`)
  SELECT *,
    DENSE_RANK() OVER(PARTITION BY borough ORDER BY cnt_major_borough DESC)AS rank
  FROM t)
SELECT borough,
  ARRAY( SELECT AS STRUCT major_category,
    cnt_major_borough
  FROM t2 b
  WHERE a.borough = b.borough
    AND rank <= 3) cnt
FROM t2 a
WHERE rank = 1;
```

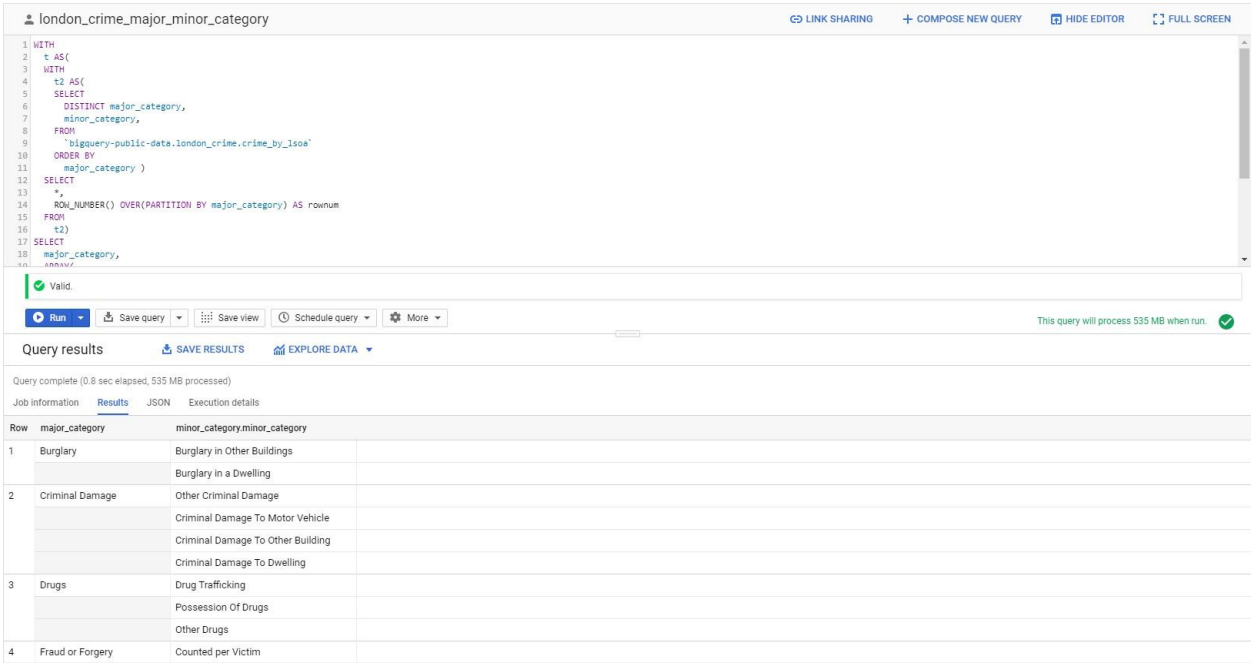


Row	major_category	minor_category.minor_category
1	Burglary	Burglary in Other Buildings
		Burglary in a Dwelling
2	Criminal Damage	Other Criminal Damage
		Criminal Damage To Motor Vehicle
		Criminal Damage To Other Building
		Criminal Damage To Dwelling
3	Drugs	Drug Trafficking
		Possession Of Drugs
		Other Drugs
4	Fraud or Forgery	Counted per Victim


Theft and Handling
Violence Against the Person
Criminal Damage
180K
160K
140K
120K
100K
80K
60K
cnt.cnt_major_borough
Croydon
Barnet
Ealing
Bromley
Enfield
Lambeth
Wandsworth
Brent
Lewisham
Southwark
borough

Question 5: Which Minor Categories falls under which Major Category?

```
WITH t AS(
  WITH t2 AS(
    SELECT DISTINCT major_category, minor_category,
    FROM `bigquery-public-data.london_crime.crime_by_lsoa`
    ORDER BY major_category )
  SELECT *,
    ROW_NUMBER() OVER(PARTITION BY major_category) AS rownum
  FROM t2)
SELECT major_category,
  ARRAY(
  SELECT AS STRUCT minor_category
  FROM t b
  WHERE b.major_category = a.major_category) minor_category
FROM t a
WHERE rownum=1
```

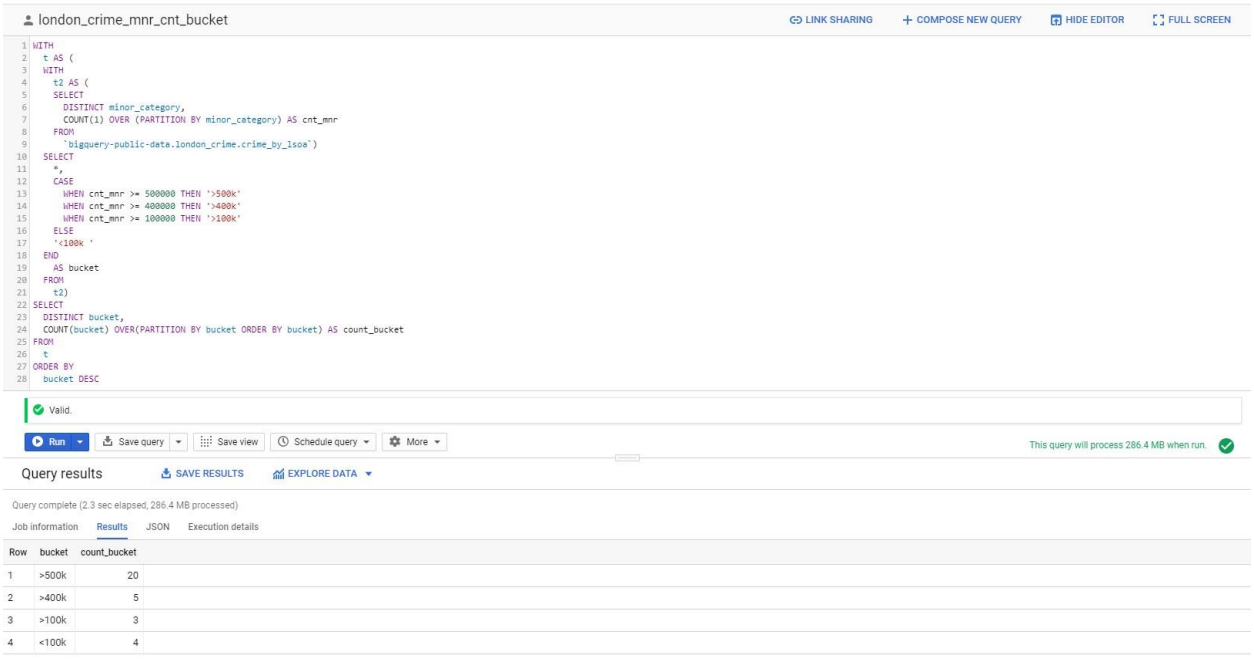


Row	major_category	minor_category.minor_category
1	Burglary	Burglary in Other Buildings
		Burglary in a Dwelling
2	Criminal Damage	Other Criminal Damage
		Criminal Damage To Motor Vehicle
		Criminal Damage To Other Building
		Criminal Damage To Dwelling
3	Drugs	Drug Trafficking
		Possession Of Drugs
		Other Drugs
4	Fraud or Forgery	Counted per Victim

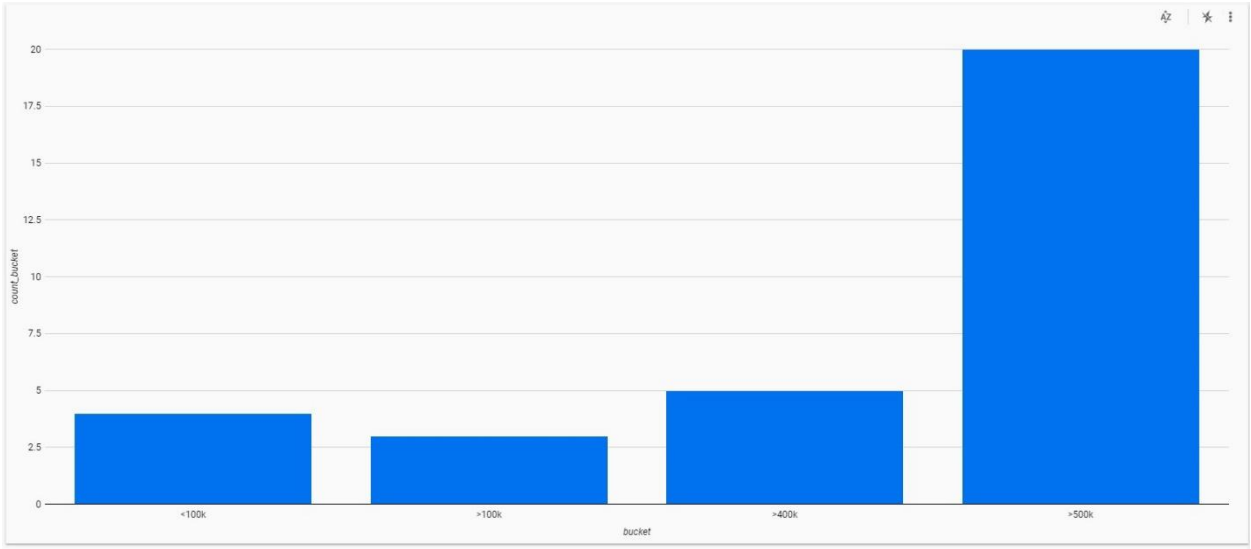
	major_category	minor_category.minor_category
1.	Burglary	Burglary in Other Buildings
2.	Burglary	Burglary in a Dwelling
3.	Criminal Damage	Criminal Damage To Other Building
4.	Criminal Damage	Other Criminal Damage
5.	Criminal Damage	Criminal Damage To Dwelling
6.	Criminal Damage	Criminal Damage To Motor Vehicle
7.	Drugs	Drug Trafficking
8.	Drugs	Possession Of Drugs
9.	Drugs	Other Drugs
10.	Fraud or Forgery	Counted per Victim
11.	Fraud or Forgery	Other Fraud & Forgery
12.	Other Notifiable Offences	Going Equipped
13.	Other Notifiable Offences	Other Notifiable
14.	Robbery	Personal Property
15.	Robbery	Business Property
16.	Sexual Offences	Other Sexual
17.	Sexual Offences	Rape
18.	Theft and Handling	Motor Vehicle Interference & Tampering
19.	Theft and Handling	Theft/Taking Of Motor Vehicle
20.	Theft and Handling	Other Theft
21.	Theft and Handling	Theft/Taking of Pedal Cycle
22.	Theft and Handling	Theft From Motor Vehicle
23.	Theft and Handling	Handling Stolen Goods
24.	Theft and Handling	Other Theft Person
25.	Theft and Handling	Theft From Shops
26.	Violence Against the Person	Offensive Weapon
27.	Violence Against the Person	Assault with Injury
28.	Violence Against the Person	Other violence
29.	Violence Against the Person	Common Assault
30.	Violence Against the Person	Wounding/GBH
31.	Violence Against the Person	Murder
32.	Violence Against the Person	Harassment

Question 6: How many minor categories falls under different buckets?

```
WITH t AS (
  WITH t2 AS (
    SELECT DISTINCT minor_category,  COUNT(1) OVER (PARTITION BY minor_category) AS cnt_mnr
    FROM `bigquery-public-data.london_crime.crime_by_lsoa`)
  SELECT *,
    CASE
      WHEN cnt_mnr >= 500000 THEN '>500k'
      WHEN cnt_mnr >= 400000 THEN '>400k'
      WHEN cnt_mnr >= 100000 THEN '>100k'
    ELSE '<100k '
  END AS bucket
  FROM t2)
SELECT DISTINCT bucket,
  COUNT(bucket) OVER(PARTITION BY bucket ORDER BY bucket) AS count_bucket
FROM t
ORDER BY bucket DESC
```



Row	bucket	count_bucket
1	>500k	20
2	>400k	5
3	>100k	3
4	<100k	4


20
17.5
15
12.5
10
7.5
5
2.5
0
count_bucket
<100k
>100k
>400k
>500k
bucket

Question 7: In each bucket found above, which top 3 minor categories had highest number of crimes recorded?

```
WITH t AS(
  WITH t2 AS (
    WITH t3 AS (
      SELECT DISTINCT minor_category,
        COUNT(1) OVER (PARTITION BY minor_category) AS cnt_mnr
      FROM `bigquery-public-data.london_crime.crime_by_lsoa`)
    SELECT *,
      CASE
        WHEN cnt_mnr >= 500000 THEN '>500k'
        WHEN cnt_mnr >= 400000 THEN '>400k'
        WHEN cnt_mnr >= 100000 THEN '>100k'
      ELSE '<100k '
    END AS bucket
    FROM t3)
  SELECT *,
    DENSE_RANK() OVER(PARTITION BY bucket ORDER BY cnt_mnr) AS rank
  FROM t2)
SELECT bucket,
  ARRAY(
  SELECT AS STRUCT ARRAY_AGG(minor_category) as minor_categpry,
    ARRAY_AGG(cnt_mnr) as count
  FROM t b
  WHERE b.bucket = a.bucket
    AND rank <= 3) bckt
FROM t a
WHERE rank = 1
ORDER BY bucket DESC
```

london_crime_bucket_minor_cnt

LINK SHARING | + COMPOSE NEW QUERY | HIDE EDITOR | FULL SCREEN

```
WITH
  t AS(
  WITH
    t2 AS (
    WITH
      t3 AS (
      SELECT
        DISTINCT minor_category,
        COUNT(1) OVER (PARTITION BY minor_category) AS cnt_mnr
      FROM
        `bigquery-public-data.london_crime.crime_by_lsoa`)
    SELECT
      *,
      CASE
        WHEN cnt_mnr >= 500000 THEN '>500k'
        WHEN cnt_mnr >= 400000 THEN '>400k'
        WHEN cnt_mnr >= 100000 THEN '>100k'
      ELSE
```

Valid.

Run | Save query | Save view | Schedule query | More

This query will process 286.4 MB when run.

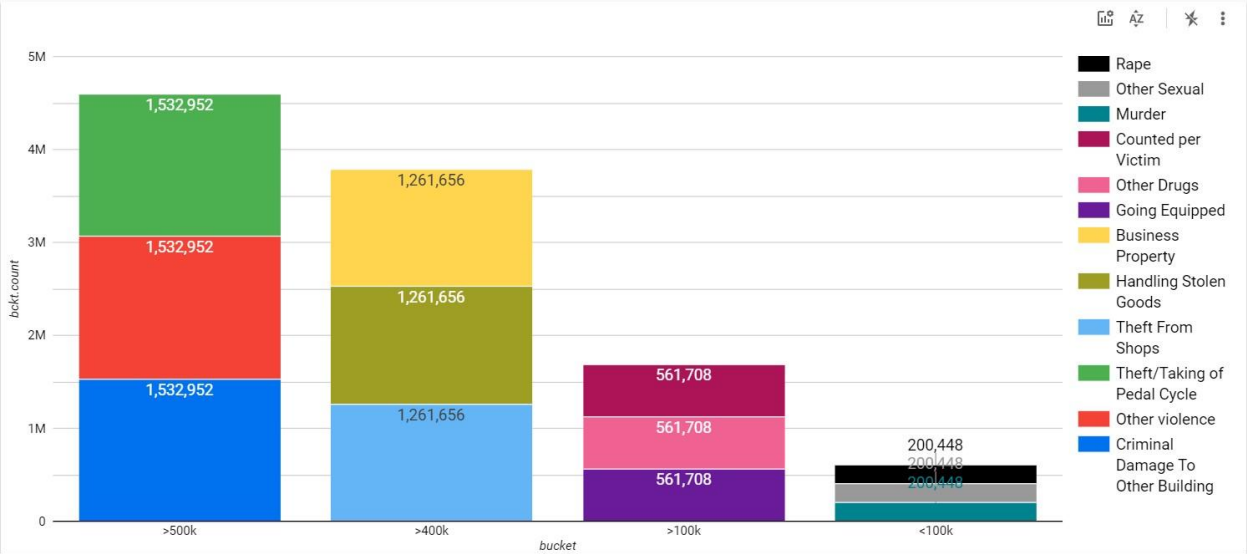
Query results

SAVE RESULTS | EXPLORE DATA

Query complete (0.0 sec elapsed, cached)

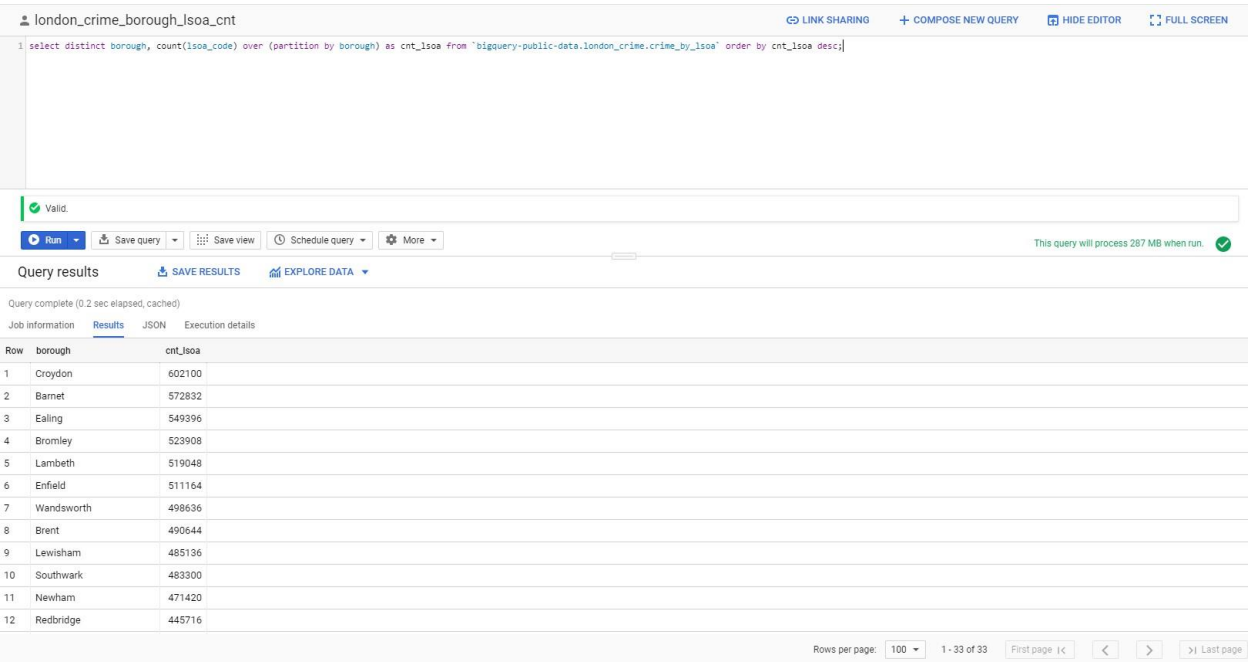
Job information | Results | JSON | Execution details

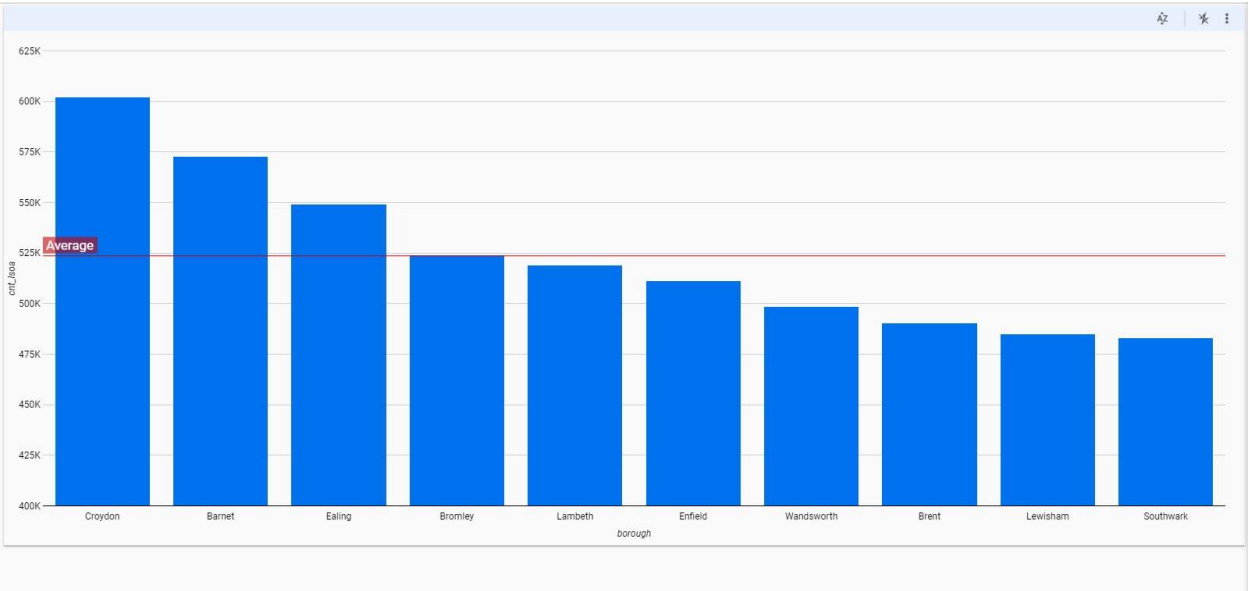
Row	bucket	bckt.minor_categpry	bckt.count
1	>500k	Criminal Damage To Other Building	503928
		Other violence	512028
		Theft/Taking of Pedal Cycle	516996
2	>400k	Theft From Shops	416772
		Business Property	418716
		Handling Stolen Goods	426168
3	>100k	Counted per Victim	137916
		Other Drugs	167184
		Going Equipped	256608
4	<100k	Rape	27000



Question 8: Which borough had most LSOA_Codes?

```
SELECT DISTINCT borough,
  COUNT(lsoa_code) OVER (PARTITION BY borough) AS cnt_lsoa
FROM `bigquery-public-data.london_crime.crime_by_lsoa`
ORDER BY cnt_lsoa DESC;
```





Link to all visualizations:

https://datastudio.google.com/s/pq0eF9DNEto

Conclusions:

The most major category of crimes that happen in the London are *"Theft and Handling"* and most minor crimes are *"Common Assault"*.

The Borough that has highest number of cases is *"Croydon"* and top 3 major categories are same over every borough and they are: *"Theft and Handling", "Violence Against the Person",* and *"Criminal Damage"*.

There are only three boroughs that have more than average number of LSOA Codes, which are: *"Croydon", "Barnet", and "Ealing"*.

Methodologies:

Analyze the Data: I used ***Windows functions***, ***Array Functions***, ***Array_Agg***, ***Struct***, ***Case***, and ***Rank*** methods that I've learned in the class.

Queries and Code: Provided above.

Visualization: Using Google Data Studio.

GitHub Link:

https://github.com/HzSaifee/Final_Report_Advanced_SQL