**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**](https://console.cloud.google.com/bigquery?project=bigquery-public-data&p=bigquery-public-data&d=london_crime&t=crime_by_lsoa&page=table) 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.

A screenshot of a social media post

Description automatically generated

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`;

A screenshot of a social media post

Description automatically generated

A screenshot of a cell phone

Description automatically generated

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`;

A screenshot of a social media post

Description automatically generated

A picture containing drawing

Description automatically generated

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`;

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

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;

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

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

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

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

A screenshot of a social media post

Description automatically generated

A screenshot of a social media post

Description automatically generated

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

A screenshot of a social media post

Description automatically generated

A screenshot of a cell phone

Description automatically generated

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;

A screenshot of a social media post

Description automatically generated

A picture containing drawing

Description automatically generated

**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>