**Individual Project**

ArunBalaji Selvamohan

Business Analytics, Conestoga College

INFO8076-23F-Sec3: SQL and Data Analysis

**Jey Kumaresan**

October 22, 2023

**Introduction**

Education plays a vital role in reshaping the socioeconomic conditions in any country. Different educational facilities, like kindergarten, elementary school, high school education, college, and universities, contribute to society by shaping the future of young minds. To understand the importance of new educational facilities in a country needs an everyday database collection of all schools and colleges and their location. The increasing immigrant population, international students, and other reasons are considered to ensure the quality of education and infrastructure in Canada. The topic “*The Open Database of Educational Facilities”* (ODEF) aims to unravel the detailed analysis of institutions, their spatial distribution, and the resources and skills they offer during their study. Also, it’s an opportunity to understand the educational institutes in different demographics' performance and impact on society.

**Data Cleaning**

The Open Database of Educational Facilities**(**ODEF) represents the crucial information of Canadian institutions nationwide. However, we should have a meaningful dataset to make an informed decision, and we need to do data cleaning. This data set required different data cleaning techniques, such as ensuring data accuracy, handling the missing values, consistency in data format, removing duplicate values, and ensuring data quality. Our database has undergone a rigorous data-cleaning process to ensure clean data. The following tables are extracted for data upload,

* Facility – Containing the facility name, authority name, international standard classification of education (ISCED) level, and official language minority status.
* Location - address, unit, street number, street name, municipality name, province, postal code, longitude, latitude, and geocoding source.

**ERD Diagram**

****

**SQL Queries**

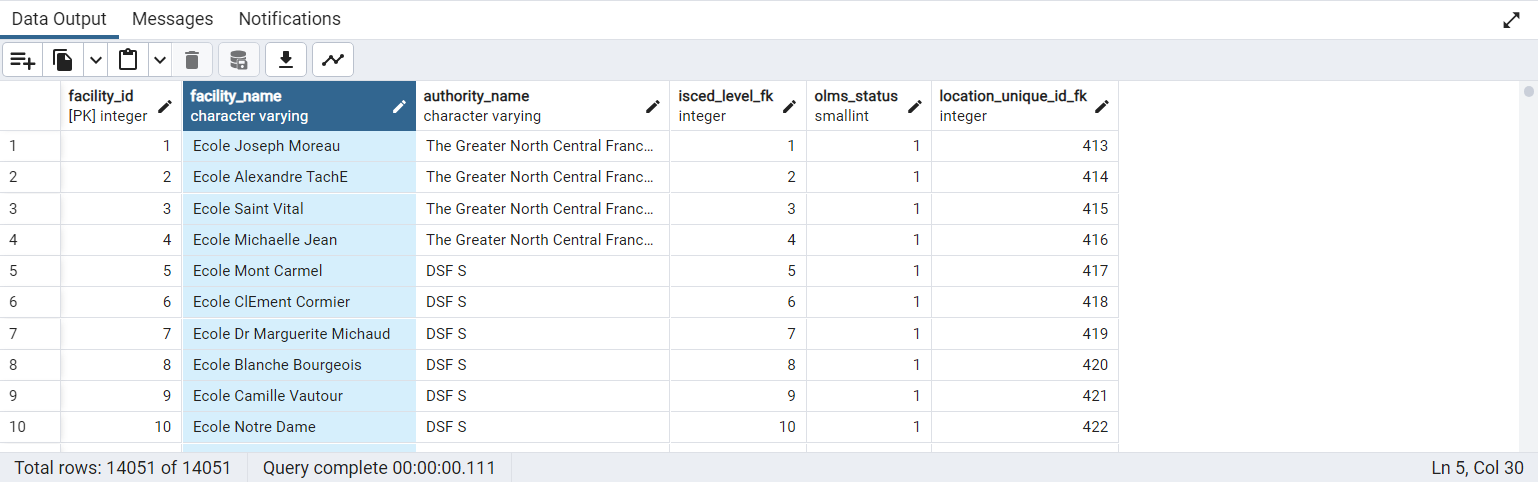
The following SQL queries are created for data analysis purposes and retrieved results:

1. ***Retrieve All Educational Facilities:***

SELECT facility\_id, facility\_name, authority\_name, isced\_level\_fk, olms\_status, location\_unique\_id\_fk

FROM public.facility;

***Output***



***Insights***

This query provides a comprehensive list of all educational facilities across Canada from the database. It provides the foundational idea about our primary analysis.

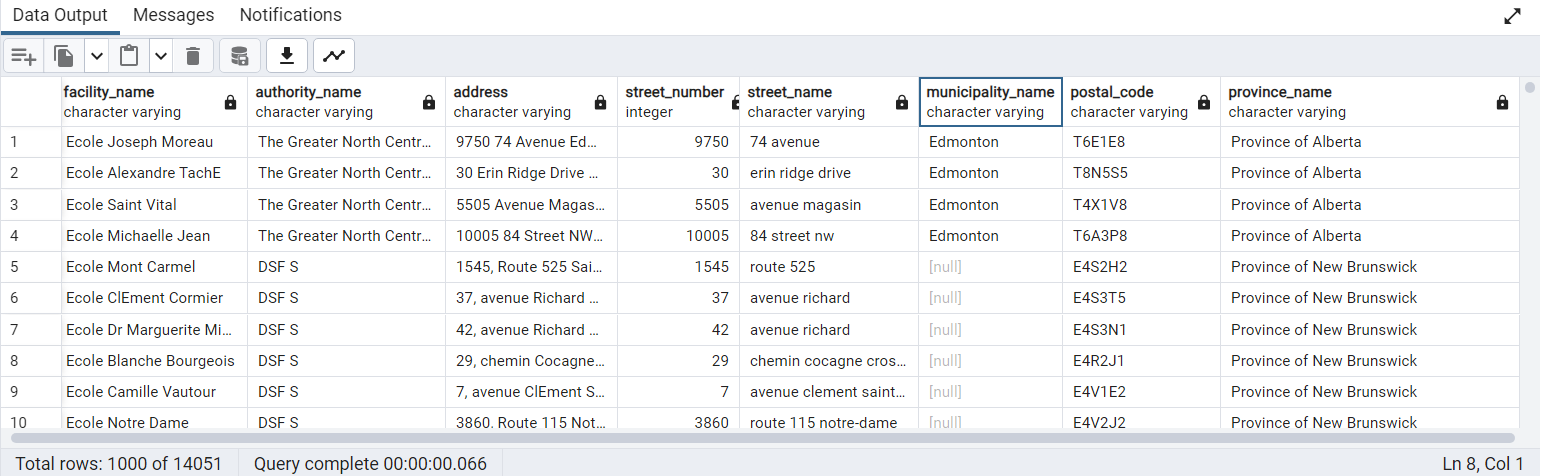
1. ***Retrieve All Facilities With Location Details:***

SELECT f.facility\_name, f.authority\_name, l.address, l.street\_number, l.street\_name, l.municipality\_name, l.postal\_code, pr.province\_name

FROM public.facility f

INNER JOIN public.location\_detail l ON f.location\_unique\_id\_fk = l.location\_unique\_id\_pk

INNER JOIN public.province\_detail pr ON l.province\_unique\_id\_fk = pr.province\_unique\_id\_pk;

***Output*** 

***Insights***

This query combines the facility and its locations to give more spatial analysis about understanding the distribution of different facilities across Canada with their geographical areas like address, unit, street name, municipal name, postal codes, and provinces.

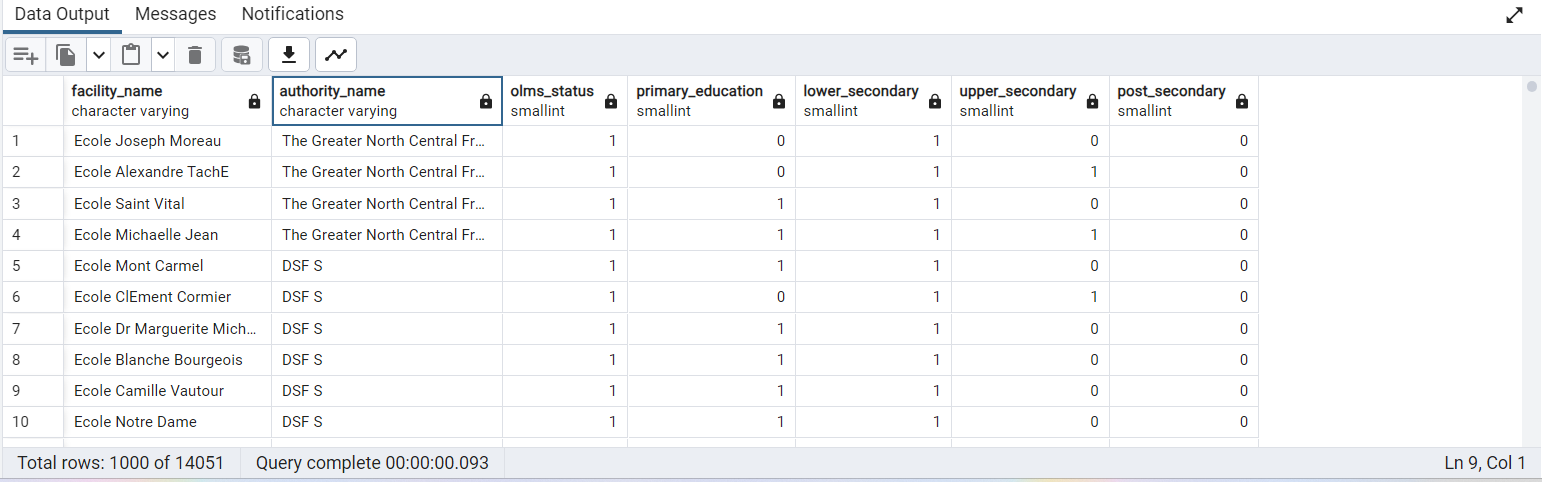
1. ***Retrieve Educational Facilities With ISCED Details:***

SELECT f.facility\_name,f.authority\_name,f.olms\_status, i.primary\_education, i.lower\_secondary, i.upper\_secondary, i.post\_secondary

FROM public.facility f

LEFT JOIN public.isced\_detail i ON f.isced\_level\_fk = i.isced\_level\_pk;

***Output***

******

***Insights***

This query helps to identify the ISCED level, which means the international standard classification of educational level. This query shows how different educational institutions are categorized and compares them based on primary, secondary, and post-secondary educational stages.

1. ***Count Educational Facilities In Each Province:***

SELECT p.province\_name, COUNT(f.facility\_id) AS facility\_count

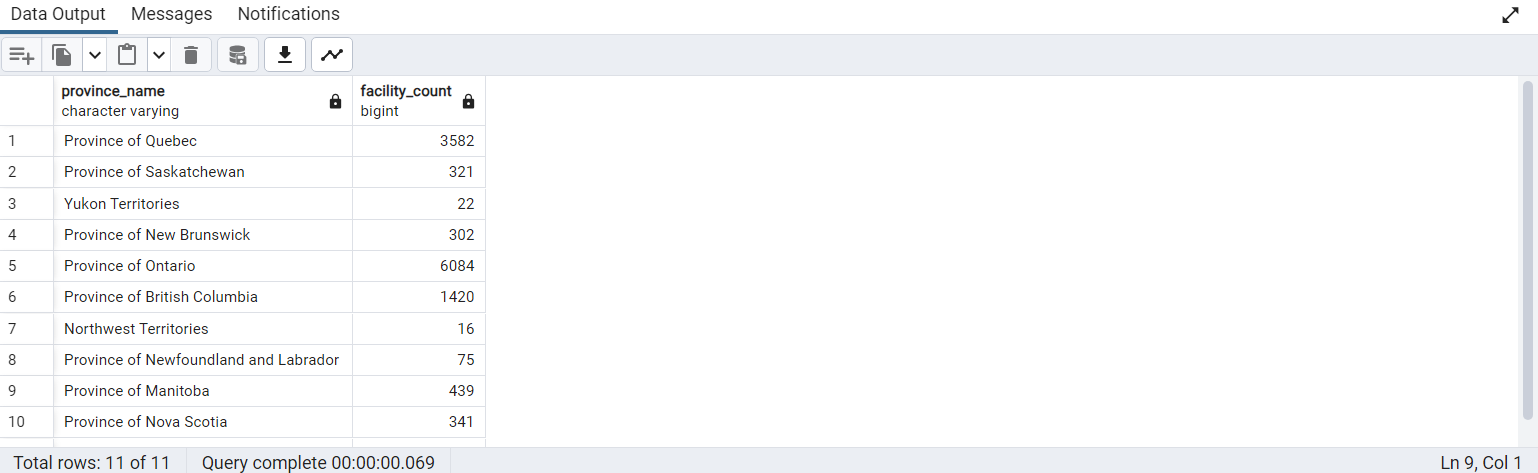
FROM public.facility f

JOIN public.location\_detail l ON f.location\_unique\_id\_fk = l.location\_unique\_id\_pk

JOIN public.province\_detail p ON l.province\_unique\_id\_fk = p.province\_unique\_id\_pk

GROUP BY p.province\_name;

***Output***

******

***Insights***

This query provides information about count-wise educational facilities per province and distribution across different regions.

1. ***Retrieve Educational Facilities Based On Specific Province:***

SELECT f.facility\_name, l.address, l.municipality\_name, p.province\_name

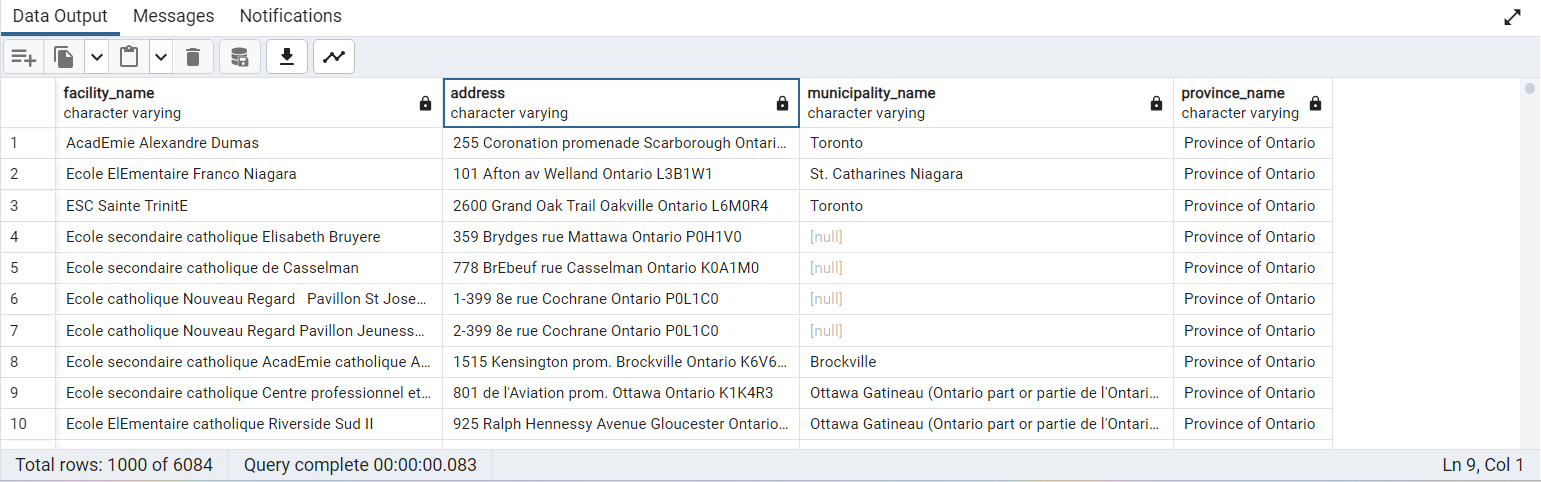
FROM public.facility f

JOIN public.location\_detail l ON f.location\_unique\_id\_fk = l.location\_unique\_id\_pk

JOIN public.province\_detail p ON l.province\_unique\_id\_fk = p.province\_unique\_id\_pk

WHERE p.province\_name = 'Province of Ontario';

***Output***

******

***Insights***

This query gives the privilege of selecting a facilities list with province-based searching and gives an idea of studying how the regional disparity occurs.

1. ***Retrieve Educational Facilities with Census Subdivision Information:***

SELECT f.facility\_name, l.address, l.municipality\_name, c.census\_subdivision\_name

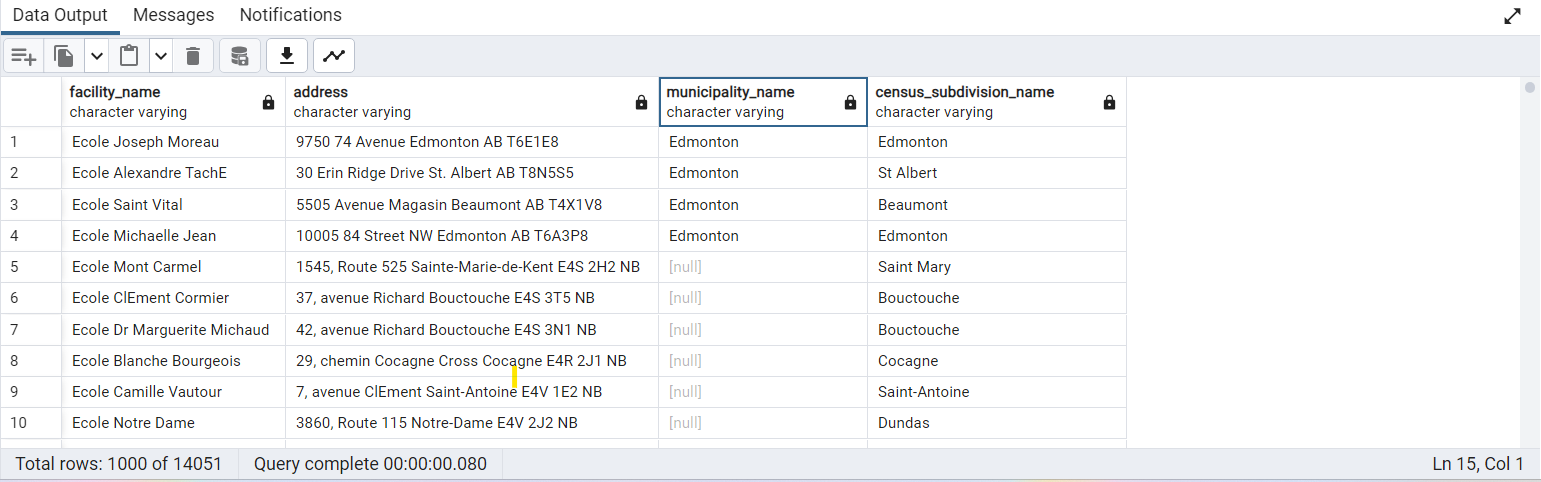
FROM public.facility f

JOIN public.location\_detail l ON f.location\_unique\_id\_fk = l.location\_unique\_id\_pk

JOIN public.census\_subdivision c ON l.census\_subdivision\_unique\_id\_fk = c.census\_subdivision\_unique\_id\_pk;

O

***Output***



***Insights***

This query gives the facility for retrieving the educational facilities based on demographic analysis, helping to understand how many institutions are there for local populations.

1. ***Count Facilities Based On Each Metropolitan Area:***

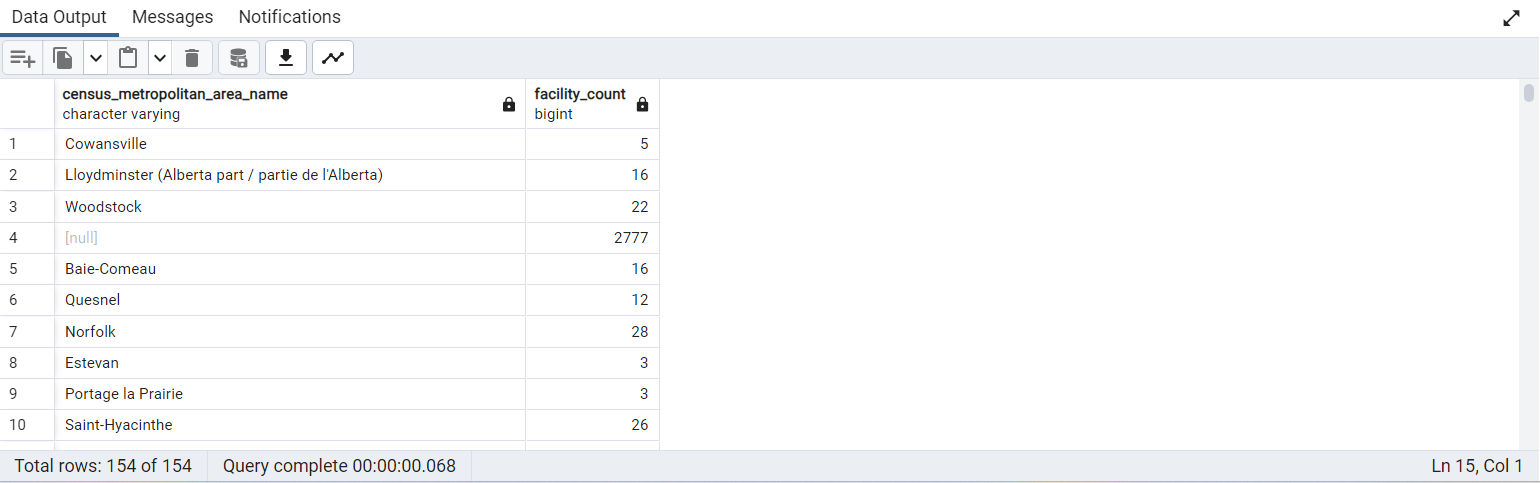
SELECT cm.census\_metropolitan\_area\_name, COUNT(f.facility\_id) AS facility\_count FROM public.facility f

JOIN public.location\_detail l ON f.location\_unique\_id\_fk = l.location\_unique\_id\_pk

JOIN public.census\_metropolitan cm ON l.census\_metropolitan\_area\_unique\_id\_fk = cm.census\_metropolitan\_area\_unique\_id\_pk

GROUP BY cm.census\_metropolitan\_area\_name;

***Output***

******

***Insights***

This query allows the facility to find count-wise the facilities distributed across the census metropolitan cities and provides insights into comparing how many institutes are there in urban cities.

1. ***Retrieve Facilities in a Specific Census Metropolitan Area:***

SELECT f.facility\_name, f.authority\_name, l.address, l.street\_name, l.municipality\_name, cm.census\_metropolitan\_area\_name

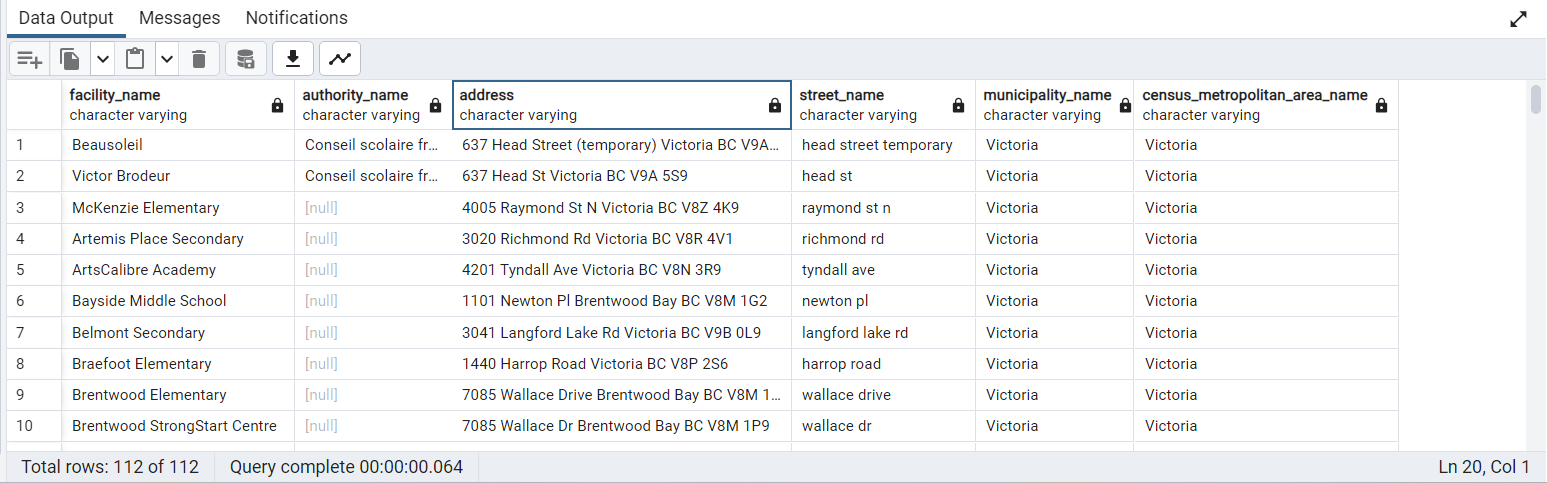
FROM public.facility f

JOIN public.location\_detail l ON f.location\_unique\_id\_fk = l.location\_unique\_id\_pk

JOIN public.census\_metropolitan cm ON l.census\_metropolitan\_area\_unique\_id\_fk = cm.census\_metropolitan\_area\_unique\_id\_pk

WHERE cm.census\_metropolitan\_area\_name = 'Victoria';

***Output***

******

***Insights***

This query helps to retrieve the data for specific metropolitan area facilities lists so that we can plan the infrastructure and resource allocation for that facility's development.

1. ***Retrieve Facilities with a Specific Authority Name:***

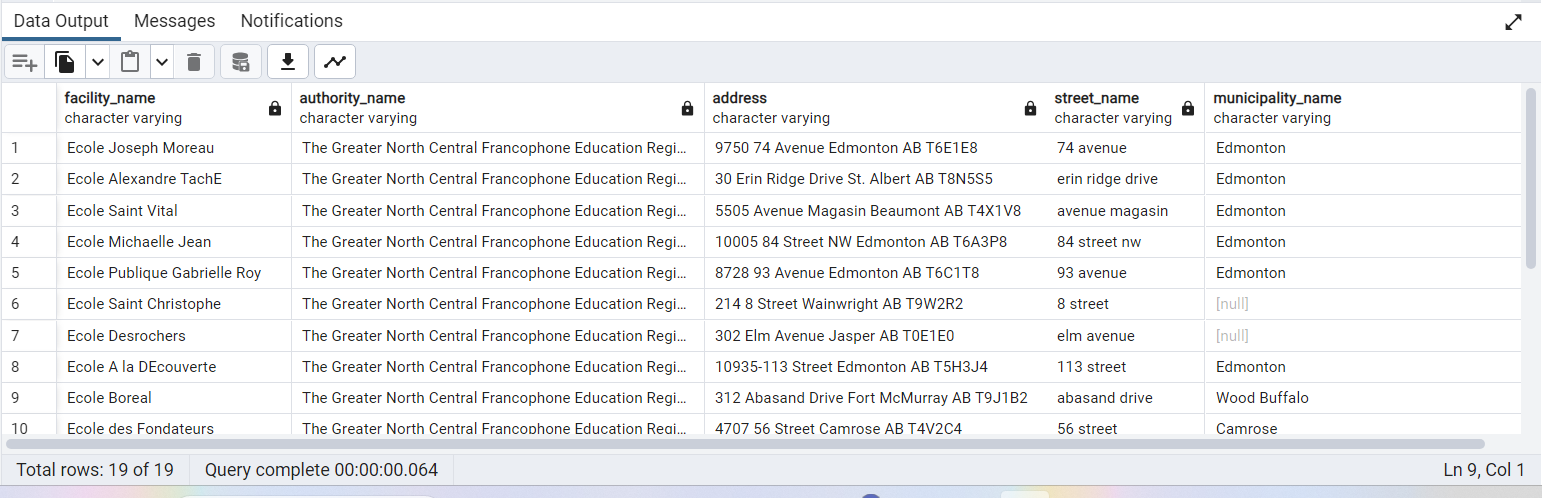
SELECT f.facility\_name, f.authority\_name, l.address, l.street\_name, l.municipality\_name

FROM public.facility f

JOIN public.location\_detail l ON f.location\_unique\_id\_fk = l.location\_unique\_id\_pk

WHERE f.authority\_name = 'The Greater North Central Francophone Education Region';

***Output***

******

***Insights***

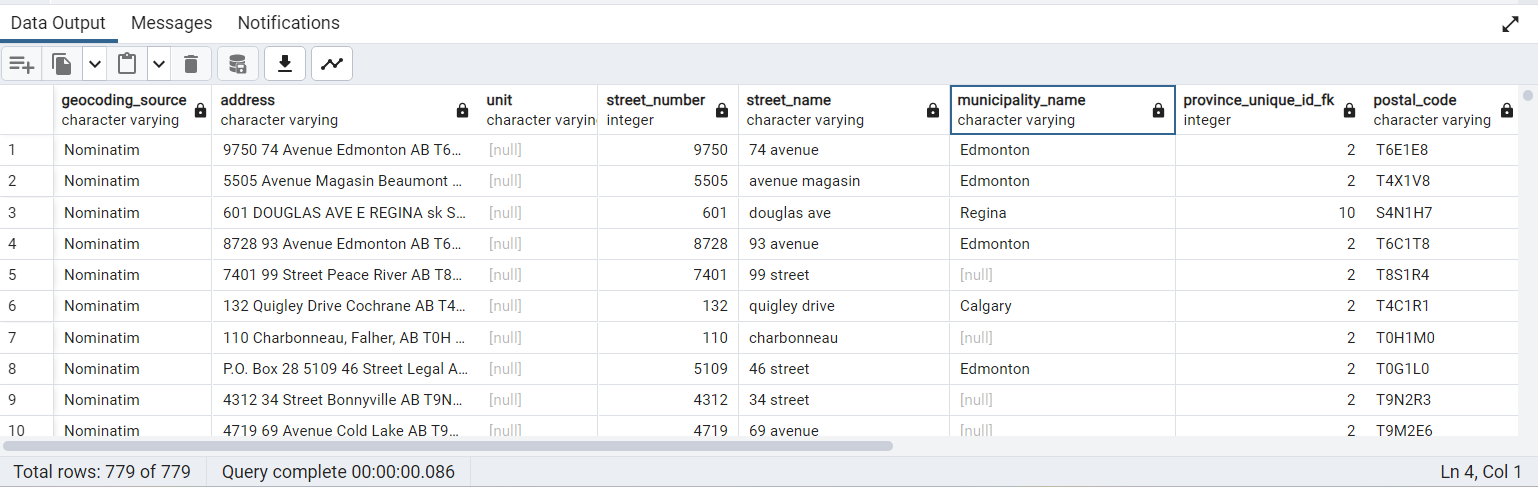
It helps examine the facilities associated with specific authority names and provides the distribution of facilities under the specific authority’s name.

1. ***Retrieve Facilities with Specific Geocoding Source:***

SELECT geocoding\_source, address, unit, street\_number, street\_name, municipality\_name, province\_unique\_id\_fk, postal\_code, longitude, latitude, source\_id, census\_subdivision\_unique\_id\_fk, census\_metropolitan\_area\_unique\_id\_fk

FROM public.location\_detail WHERE geocoding\_source = 'Nominatim';

***Output***

******

***Insights***

It gives the detail of specific geo coding sources accurate and reliable data associated with the educational facilities.

**Views**

1. **Overview of Educational facilities**

CREATE VIEW educational\_facilities\_overview AS

SELECT f.facility\_id, f.facility\_name, f.authority\_name, f.olms\_status, l.address, l.unit, l.street\_number, l.street\_name,

l.municipality\_name, l.province\_unique\_id\_fk, l.postal\_code, l.longitude, l.latitude, l.geocoding\_source, i.primary\_education, i.lower\_secondary,

i.upper\_secondary, i.post\_secondary

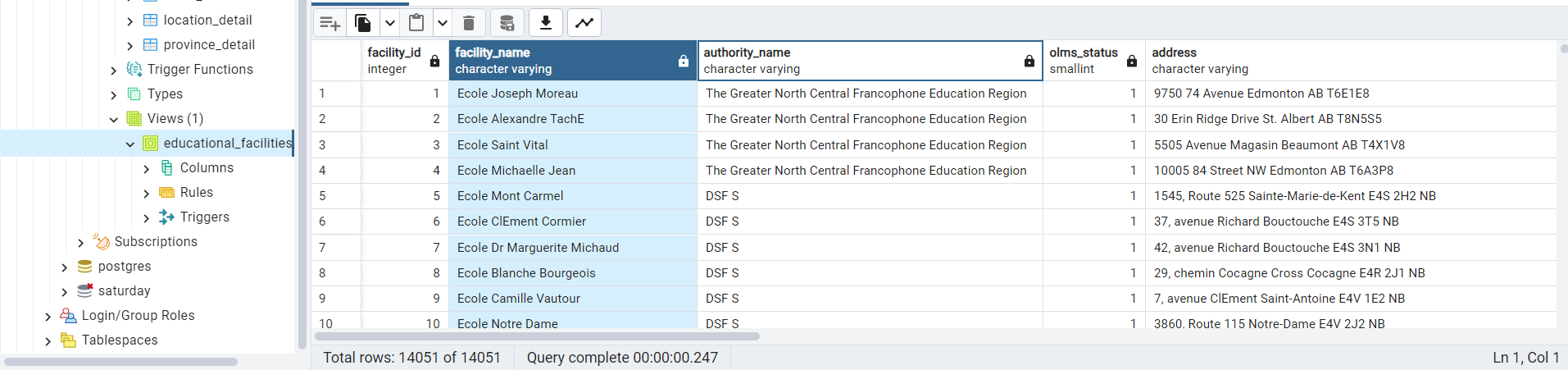
FROM public.facility f INNER JOIN

public.location\_detail l ON f.location\_unique\_id\_fk = l.location\_unique\_id\_pk

LEFT JOIN

public.isced\_detail i ON f.isced\_level\_fk = i.isced\_level\_pk;

***Output***

******

***Insights***

This provides a comprehensive overview of educational facilities, including the details about their location, ISCED, and OLEM status.

1. ***Educational Facilities by Province:***

CREATE VIEW facilities\_by\_province AS

SELECT

p.province\_name,

COUNT(f.facility\_id) AS facility\_count

FROM

public.facility f

JOIN

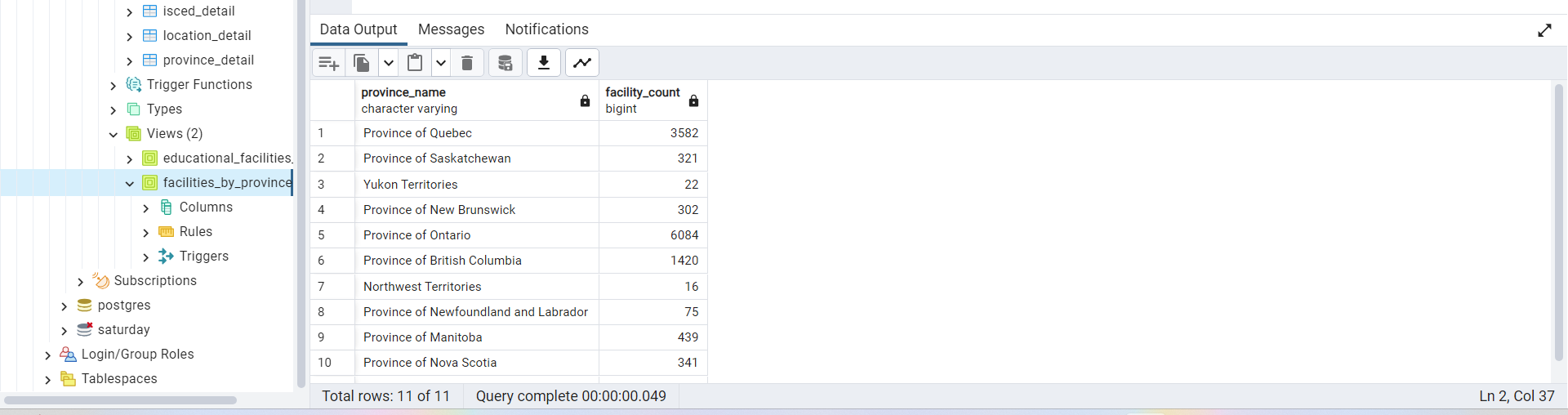
public.location\_detail l ON f.location\_unique\_id\_fk = l.location\_unique\_id\_pk

JOIN

public.province\_detail p ON l.province\_unique\_id\_fk = p.province\_unique\_id\_pk

GROUP BY p.province\_name;

***Output***

******

***Insights***

It summarizes different provinces and their facilities, including a count of each province's facilities.

**Conclusion**

In conclusion, the ODEF Provides comprehensive analysis and decision-making in educational facilities development, resource allocation, and educational policy formulation.

This database provides various dimensions of educational institutes like address and location. Province, census division, and ISCED level provide more insights to stakeholders to understand where changes and infrastructure are needed.

Creating 2 views can provide a handy data overview of “Educational Facilities Overview" and "Educational Facilities by Province".