Ontario Schools Analysis

Preface

Tools Used: Power Bi, MS Excel, PostgreSQl

Database Source:
https://data.ontario.ca/dataset/private-school-contact-information/resource/0f209d98-8dfb-45f1-bcc5-5dc936e740ea

Problem Statement:

The Ontario government is currently facing challenges in analyzing school data, with around 9,000 rows of data to manage in order to track changes over the past years. Furthermore, they require detailed insights into schools' information such as education credits, geographical location, association membership, leadership, and school websites. Therefore, it is necessary to extract key information and develop a dashboard to analyze trends, supporting the government in making accurate, data-driven decisions for optimizing the policy-making process for educational institutions.

Solution:

1. **Dataset Collection**
2. **Understanding the data**
3. **Data cleaning and Extracting Key Insights**
4. **Data Visualization**
5. **Quality Assurance: Data Validation and Functional Validation.**

Data Cleaning

- **Opening Dataset in Excel and Make a Copy of Dataset for security purpose.**
- **Removing Duplicates.**
- **Formatting of columns wherever necessary.**
- **Spelling Check.**
- **Changing Case - Lower/Upper/Proper.**
- **Trimming unwanted spaces.**
- **Removing null values wherever necessary.**
- **Finding & Replacing values.**
- **Uploading the cleaned dataset to PostgreSQL.**

Extracting Key Insights

a) Table Creation Statements:

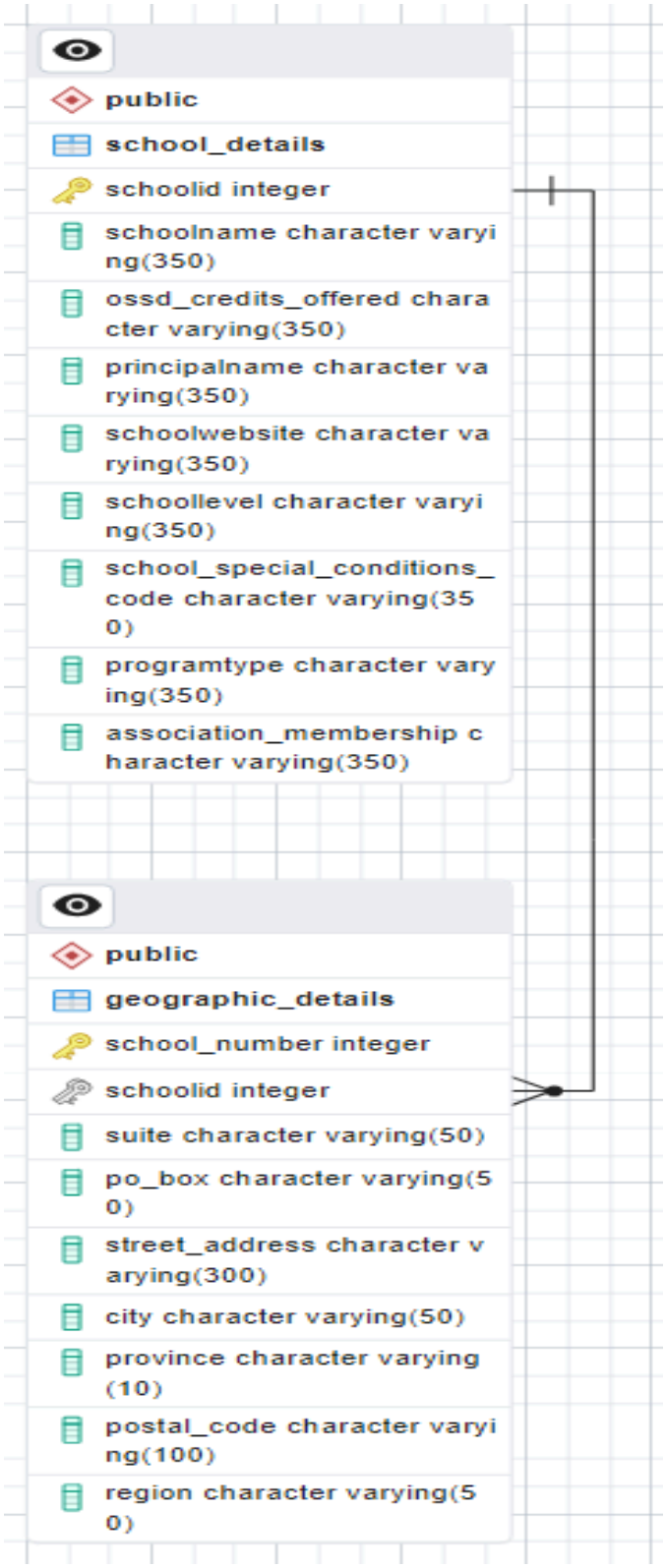
i) School Details table:

```
Create Table School_Details
(SchoolID int primary key,
SchoolName varchar(100),
OSSD_Credits_Offered varchar(100),
PrincipalName varchar(100),
SchoolWebsite varchar(200),
SchoolLevel varchar(200),
School_Special_Conditions_Code varchar(100),
ProgramType varchar(100),
Association_Membership varchar(100)
);

## Creating clustered index on tables
Create index school_details_index on school_details(schoolid);
Cluster school_details using school_details_index;
```

ii) Geographical Detail Table:

```
Create Table Geographic_details
(School_Number int primary key,
 SchoolID int references School_Details(SchoolID),
 Suite int,
 PO_Box int,
 Street_Address varchar(300),
 City varchar(50),
 Province varchar(10),
 Postal_Code varchar(100),
 Region varchar(50),
 School_Website varchar(300)

## Creating clustered index on tables
create index geographic_details_index on geographic_details(school_number);
cluster geographic_details using geographic_details_index;
```

b) Database Schematics:



c) Database Queries:

Query-1 : Find the school names with no condition code of in eastern region.

```
select sd.schoolname as School_Name, sd.school_special_conditions_code as
Condition_Code,GD.region as Region
from school_details sd
INNER JOIN
geographic_details GD
on sd.schoolid = GD.schoolid
where GD.region = 'East Region' and sd.school_special_conditions_code = 'Not applicable'
order by sd.schoolname;
```

Insights: There are over 205 schools in the eastern region with no special condition code applicable.

	school_name character varying (350)	condition_code character varying (350)	region character varying (50)
1	A+ Academy of Advancement	Not applicable	East Region
2	Abraar School - Elementary	Not applicable	East Region
3	Academia Stella Maris	Not applicable	East Region
4	AcadEmie Providence Soeurs Antonines	Not applicable	East Region
5	Academy of New Echo Education	Not applicable	East Region
6	Academy of Scholars Inc.	Not applicable	East Region
7	Afzal Islamic Montessori & Academy	Not applicable	East Region
8	Agincourt Montessori School Inc	Not applicable	East Region
9	Ahlul Bayt Islamic School	Not applicable	East Region
10	Al-Qasim Academy	Not applicable	East Region
11	Al Azhar Islamic School	Not applicable	East Region
12	Al Furqan School	Not applicable	East Region
13	Al Haadi (9828770 Canada Institute)	Not applicable	East Region
14	Alathena International Elementary School	Not applicable	East Region

Total rows: 205 of 205 | Query complete 00:00:00.114

Query-2 : Calculate the count of OSSD credits for schools in each city and ranking them accordingly.

```
Select GD.city as City, count(sd.ossd_credits_offered) as Credits_Count
from
geographic_details GD
INNER JOIN
school_details sd
on sd.schoolid = GD.schoolid
group by GD.city
order by Credits_Count desc;
```

Insights: Toronto has got the maximum number of schools with OSSD credits.

	city character varying (50)	credits_count bigint
1	Toronto	142
2	Mississauga	53
3	Markham	50
4	North York	45
5	Richmond Hill	42
6	Scarborough	28
7	Ottawa	26
8	Brampton	24
9	Oakville	18
10	Thornhill	13
11	Etobicoke	11
12	Hamilton	10
13	TORONTO	9

Total rows: 383 of 383 Query complete 00:00:00.132

Query-3: Counting the number of elementary, secondary and Elem/Sec. schools within each city in Ontario.

```
select GD.city as City,
sum(case when sd.schoollevel = 'Secondary' then 1 else 0 end) as Secondary_Schools,
sum(case when sd.schoollevel = 'Elementary' then 1 else 0 end) as Elementary_Schools,
sum(case when sd.schoollevel = 'Elem/Sec' then 1 else 0 end) as Elem_Secs
from
school_details sd
inner join
geographic_details GD
on sd.schoolid = GD.schoolid
group by GD.city
order by GD.city;
```

Insights: Toronto has got the maximum number of secondary schools, elementary and elem/sec schools.

	city character varying (50)	secondary_schools bigint	elementary_schools bigint	elem_secs bigint
1	Addison	0	1	0
2	Ailsa Craig	0	1	0
3	Ajax	0	4	1
4	AJAX	0	1	0
5	Alexandria	0	0	1
6	Algonquin Highlands	1	1	0
7	Algonquin Park	1	0	0
8	Alliston	0	1	1
9	Alma	1	1	0
10	Amaranth	0	2	0
11	Ancaster	2	2	0
12	Apsley	0	0	1
13	Ariss	0	1	0

Total rows: 383 of 383 Query complete 00:00:00.119

Query-4: The name of principals and the number of schools they headed.

```
Select S.principalname as Principal_Name, count( distinct SD.schoolid) as SchoolCount
from
school_details S
inner Join
school_details SD
on S.principalname = SD.principalname
group by Principal_Name
order by SchoolCount desc;
```

Insights:Hassan Mirzai has headed maximum number of schools in Ontario as a principal.

	principal_name character varying (350)	schoolcount bigint
1	Hassan Mirzai	5
2	Anchuan Jiang	4
3	Simon Huynh	4
4	Yehudis Cagen	4
5	Christian Bayly	4
6	Bilal Rashid	4
7	Alex Fan	4
8	Ana Morales	3
9	Salman Qureshi	3
10	Becky Yin	3
11	Sheileen Krone	3
12	Samantha Schmidt	3

Total rows: 1000 of 1467 | Query complete 00:00:00.131

Query-5: Extracting the count of duplicate school websites.

```
Select schoolwebsite, count(schoolwebsite)
from
school_details
group by schoolwebsite
having count(schoolwebsite) > 1
order by count(schoolwebsite) DESC;
```

Insights: "www.atlastforestschools.com" has been duplicated multiple times.

	schoolwebsite character varying (350)	count bigint
1	www.atlastforestschools.com	7
2	www.caasda.com	4
3	www.torahhigh.org	4
4	www.mua.ca	3
5	www.careerquestcanada.com	3
6	blytheducation.com	3
7	www.blytheducation.com	3
8	www.rotherglen.com	3
9	www.canadaonlineeducation.com	3
10	www.westfieldeducation.ca	3
11	www.tmsschool.ca	2
12	http://www.academiestececile.ca	2
13	thestudyacademy.ca	2

Total rows: 41 of 41 | Query complete 00:00:00.136

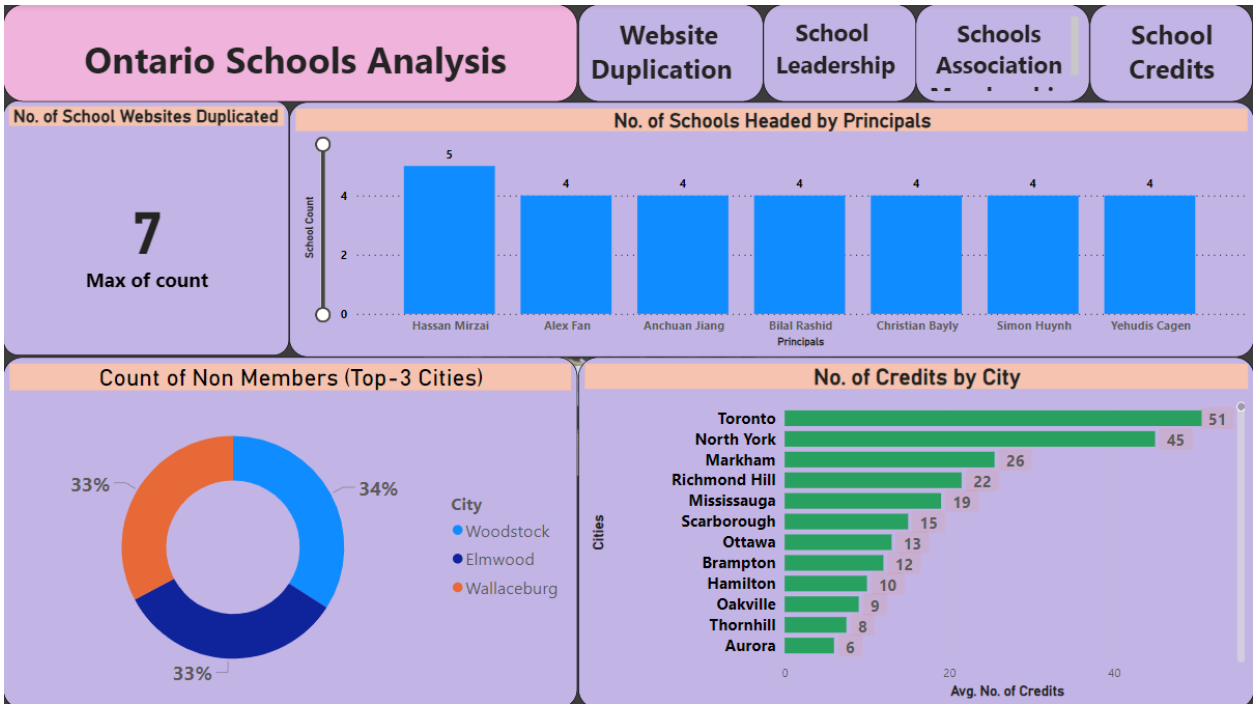
Query-6: Top-3 cities with average number of schools without association membership.

```
select gd.city as City, avg(gd.schoolid) as Average from
geographic_details gd inner join school_details S
on gd.schoolid = S.schoolid
where
S.association_membership != 'No'
group by gd.city
order by
Average desc
limit 3;
```

Insights: Woodstock has got the maximum average number of schools without association membership.

	city character varying (50)	average numeric
1	Woodstock	1564.0000000(
2	Elmwood	1517.0000000(
3	Wallaceburg	1503.0000000(

Data Visualization



KPI:-

- Website Duplication (Query-5)
- School Leadership Management (Query-4)
- Schools Without Association Membership (Query-6)
- Ontario School Credits (Query-2)

Quality Testing

1. **Functional Validation:-**

   Testing each feature to ensure it functions as required and verifying that all filters and action filters on the report operate correctly according to the specifications.

2. **Data Validation: -**

   Ensuring the accuracy and quality of the data by matching the values in Power BI reports with the SQL query results.