

CSCI 5408: Assignment 1

Problem 3

Visited <https://data.novascotia.ca> for creating an ERD for the real-estate client.

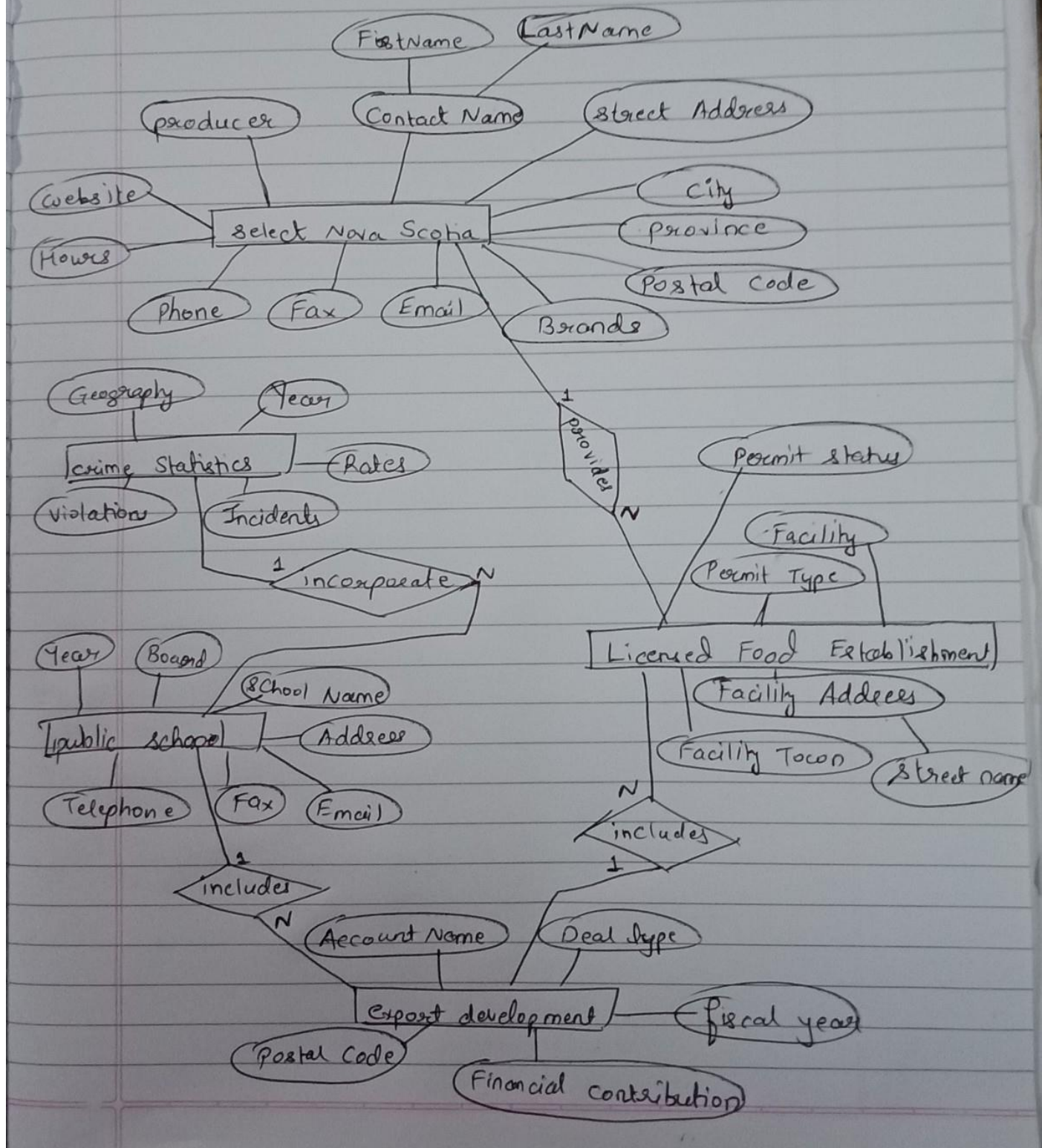
- It is a Nova Scotia Government's Open Data Portal
- It contains all the data relating to the province
- There are in total 643 datasets, but I will be focusing on 5 entities relating to **Education** and **Lifestyle**.

Entities	Attributes
Licensed Food Establishments	<ul style="list-style-type: none"> • Facility • Permit Type • Permit Status • Facility Address • Facility Town
Select Nova Scotia (Agriculture)	<ul style="list-style-type: none"> • Producer • Contact Name • Street Address • City • Province • Postal Code • Phone • Fax • Email • Website • Hours • Brands
Public School	<ul style="list-style-type: none"> • Year • Board Code • Board Name • School Code • School Name • Mail Address Line 1 • Mail Address Line 2 • Mailing Postal Code • Civic Address Line 1 • Civic Address Line 2 • Civic Postal Code • Telephone • Fax • Email • Georeference
Crime Statistics	<ul style="list-style-type: none"> • Geography • Year

	<ul style="list-style-type: none">• Violations• Incidents• Rates
Export Development and Investment Attraction	<ul style="list-style-type: none">• Account Name• nsbi_sector• Deal Type• NSBI Financial Contribution• Place Name• nsbi_county• PostalCode• fiscal year• Longitude• Latitude• New Georeferenced Column

Here is the initial Chen model that I created on paper for the above entities and attributes.

Problem 3 - Chen Model



Cleaning and Transforming the datasets:

To avoid any data type mismatch, I have made all the columns as text.

1. Crime_Statistics_-_Incidents_and_rates_for_selected_offences.csv

Operations Performed:

- The file contains 7675 records out of which I have considered first 200 records.
- Replacing blank values in **incidents** column with -1.

VIOLATIONS	INCIDENTS
Sexual assault, level 1	
Attempted murder	
Total Drug violations	
Sexual assault, level 3, aggravated	
Total Drug violations	
Total other Controlled Drugs and Substances Act drugs, trafficking, production or distribution	
Total breaking and entering	
Sexual assault, level 2, weapon or bodily harm	
Total theft of motor vehicle	
Total sexual violations against children	
Total cannabis, trafficking, production or distribution	
Total breaking and entering	
Possession, cocaine	
Total theft of motor vehicle	
Total breaking and entering	
Total other Controlled Drugs and Substances Act drugs, trafficking, production or distribution	
Total cocaine, trafficking, production or distribution	
Total theft of motor vehicle	
Assault, level 2, weapon or bodily harm	
Possession, cannabis	
Total theft of motor vehicle	
Total cannabis, trafficking, production or distribution	
Total cannabis, trafficking, production or distribution	
Total breaking and entering	
Assault, level 3, aggravated	
Sexual assault, level 2, weapon or bodily harm	
Total impaired driving	
Total other Controlled Drugs and Substances Act drugs, trafficking, production or distribution	

- Replacing blank values in **rates** column with -1.

VIOLATIONS	INCIDENTS	RATES
Sexual assault, level 1	-1	
Attempted murder	-1	
Total Drug violations	-1	
Sexual assault, level 1	1	
Sexual assault, level 3, aggravated	-1	
Total Drug violations	-1	
Total theft of motor vehicle	3	
Total impaired driving	7	
Total other Controlled Drugs and Substances Act drugs, trafficking, production or distribution	-1	
Total breaking and entering	-1	
Sexual assault, level 2, weapon or bodily harm	-1	
Total theft of motor vehicle	-1	
Total sexual violations against children	-1	
Total cannabis, trafficking, production or distribution	-1	
Total breaking and entering	-1	
Possession, cocaine	-1	
Total theft of motor vehicle	-1	
Assault, level 1	34	
Total impaired driving	14	
Total theft of motor vehicle	2	
Total breaking and entering	-1	
Total other Controlled Drugs and Substances Act drugs, trafficking, production or distribution	-1	
Total cocaine, trafficking, production or distribution	-1	
Total theft of motor vehicle	-1	
Assault, level 2, weapon or bodily harm	-1	
Possession, cannabis	-1	
Total theft of motor vehicle	-1	
Total cannabis, trafficking, production or distribution	-1	

2. List_of_Licensed_Food_Establishments.csv

Operations Performed:

- The file contains 6606 records out of which I have considered first 200 records.
- Replacing blank value in **facility_address** with UNKNOWN.

B	C	D
Permit Type	Permit Status	Facility Address
Eating Establishment	Active	

- Replacing blank value in **facility_town** with UNKNOWN.

C	D	E
Permit Status	Facility Address	Facility Town
Active	UNKNOWN	

3. Nova_Scotia_Business_Inc._Export_Development_and_Investment_Attraction_Activity.csv

Operations Performed:

- The file contains 1468 records out of which I have considered first 200 records.
- Moved **fiscal_year** beside **NSBI Financial Contribution (Commitment/Announced)** for better understanding and organization of data.

D	E	
fiscal year	NSBI Financial Contribution (Commitment/Announced)	P
2018 - 2019	10,850	D
2018 - 2019	3,750	H
2018 - 2019	8,100	H
2018 - 2019	15,000	B
2018 - 2019	15,000	T
2018 - 2019	12,500	D
2018 - 2019	2,520	H
2018 - 2019	25,000	H
2018 - 2019	2,450	H
2018 - 2019	15,000	N
2018 - 2019	4,775	D
2018 - 2019	8,400	H
2018 - 2019	14,950	H
2019 - 2020	10,000	H

4. Nova_Scotia_Public_School_Contact_Information.csv

Operations Performed:

- The file contains 1111 records out of which I have considered first 200 records.
- Formatted **Civic Address Line 1** column.

G	H	I
Mail Address Line 2	Mailing Postal Code	Civic Address Line 1
Shag Harbour	B0W 3B0	#5961 Hwy #3

5. Select_Nova_Scotia.csv

Operations Performed:

- Replaced blank values in **contact_name** column with UNKNOWN.

Producer	Contact Name
Acadian Seaplants	
Amos Pewter	
Amos Pewter Halifax	
Amos Pewter Peggy's Cove	
AquaPrime Mussel Ranch Ltd	
Art with MoJo	
Authentic Seacoast Distilling Ltd	
Baddeck Farmers Market	
Bandha Bar	
Big Cove Foods	
Blomidon Farms	
Blomindon Estate Winery	
Blue Barn Farms	

- Formatted **street_address** data and replaced blank values with UNKNOWN.

Contact Name	Street Address
Carol Pounder	
Shauna Austin	
Peter LeBlanc	
Gabrielle Breault	
Brent Larson	#101-647 Bedford Hwy
Sally & Michael Knight	
Robert Crow	

- Replacing blank values in **city** column with UNKNOWN.

Street Address	City
UNKNOWN	
UNKNOWN	
2062 Hwy 14 Centre Rawdon	
UNKNOWN	

- Formatting values in **province** column.

City	Province
Annapolis Royal	
Annapolis Royal	
Westchester Station	
Lake Echo	
Halifax	
Centerville	
Lakeville	
Cape Breton	
Cleveland	
Canning	
Cape North	

- Replacing blank values in **postal_code**, **phone**, **fax**, **email**, **website**, **hours** and **brands** with UNKNOWN.

Postal Code	Phone	Fax	Email	Website
	1 (902) 867-7479		manager@antigonishfarmersmarket.org	http://www.antigonishfarmersmarket.ca
			barringtonflorist@gmail.com	http://www.barringtonflorist.ca
	1 (902) 383-2198		claudia.gahlinger@gmail.com	
			carolpounderart@gmail.com	http://www.carolpounderart.com
				https://www.facebook.com/ChesterSeasideFarmerArtisanMarket
			sheila.leblanc@usainteanne.ca	http://farmersmarketsnovascotia.ca/find-a-market/cheticamp/
			petermleblanc@gmail.com	http://www.frontporchfarm.ca/
	(902) 679-2514		marketmanager@kentville.ca	http://www.kentvillefarmersmarket.com
	(902) 527-7721		macdougallmeadows@gmail.com	https://macdougallmeadows.com
			merigomishcommunitycentre@gmail.com	http://pcnagr.ca/index.php/farms/detail/farm_merigomish_cafe_market
			info@petitepatriechocolate.com	http://www.petitepatriechocolate.com
	1 (902) 225-1031		renerobichaud@eastlink.ca	http://www.pureproductagent.com
	(902) 674-2517			
	(902) 526-4440		strattonscsa@gmail.com	http://www.strattonsfarm.ca
	(902) 883-2698	(902) 632-2422	sarah.withrow@eastlink.ca	http://www.withrowsfarmmarket.ca

The ERD for the above cleaned datasets:

