SUBJEC – DATA MANAGEMENT, WAREHOUSING, ANALYTICS
ASSIGNMENT – 1

SUBMITTED BY,
Dhruv Doshi
B00883311
dh722257@dal.ca

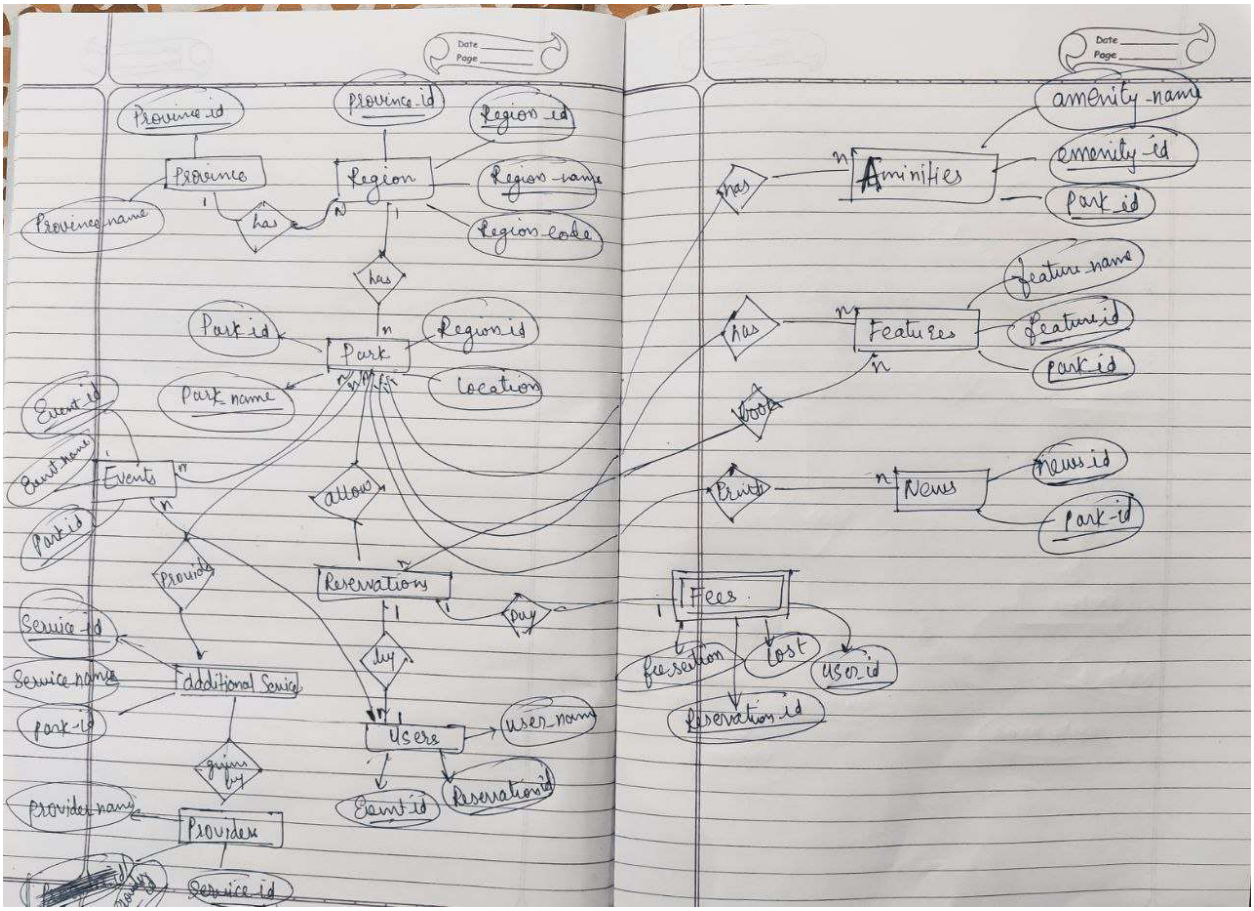
SUBMITETD TO: Prof. SAURABH DEY

Building a Data Model for Nova Scotia on its Provincial Parks.

After going through https://parks.novascotia.ca/ I have founded these 12 entities and listed their reason for the selection in the given table.

ENTITY	REASON FOR SELECTION
Park	Parks are the center point for the whole system
Region	Region is directly connected with park and have distinct value
Province	Province could help further rectify the rules to be considered
Events	All events happening in the park.
Additional Service	Third party services
Providers	Third party services providers
Reservations	User could make reservations from here.
Users	All the data regarding guests
Features	All the additional paid features given by park
Amenities	All the free amenities given by park
Fees	Fees to be paid during registration
News	News published by the park
Staff	All the maintain staff for the park

After completion of finalizing the entities, I move forward to design the initial ER diagram for the parks. The pen and paper ERD is attached as a figure below.

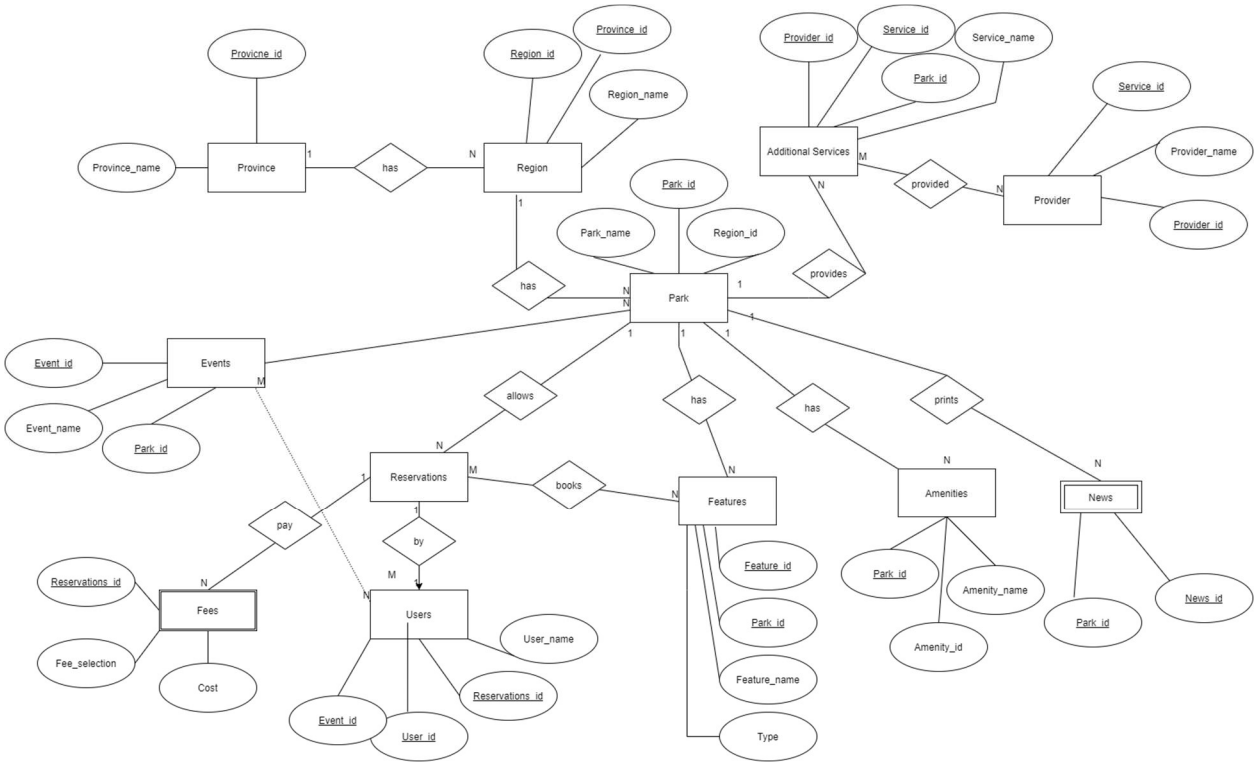


Finding the design issues, or attributes that were not considered, or entities which were found after going through the design. I have tried to cover all the aspects in the initial diagram itself still if any of the part is left then it will be corrected in this segment.

Points performed:

- Region_code is not required as it does not hold any substantial information for this use case.
- In Park we need to have Park_id as a primary key.
- Additional Service attribute should consist of the provider_id as an attribute because that will cast the connection between two tables.
- In Additional Services Park_id needed to be a primary key.
- Users entities need to have user_id attribute and that too need to be part of composite primary key.
- News will be the weak entity as it cannot be identified independently without accessing the parks along with the only way to have a primary key in it is by having the park_id and news_id being a composite key.
- There is multiple 1 to many relationships in the ER Diagram, but these all entity are individual and do not collide hence there will not be any case of fan trap here.
- There are not any many to one relationship hence no occurrence of chasm trap is identified in this ERD.

After performing the steps this will be the final ER Diagram.



Format Ocean Tracking Data and Report

The data given includes the information regarding some research work on the animals as there is vital information about the animals like their sex, age and length. Alongside that we also have information about the lab. The information about the datacenter, offload tag, detection, project and receivers is also given through the dataset. After the initial skimming I could say that the database are in bad shape and they need to be refactored which I will be doing in the later half of the assignment. In this Assignment we initially have what the changes I have performed on the datasets and why I needed to perform those steps. Continuing that it will have initial ER diagram and after that we will discuss the normalization steps and also have an ER diagram with normalization.

Datasets used for this task:

- Otnunit_att_animals
- Otnunit_aat_datacenter
- otnunit_aat_detection
- otnunit_aat_manmade
- otnunit_aat_project
- Otnunit_aat_receivers
- Otunit_aat_recover_offload_details
- Otnunit_aat_tag_releases

After cleaning and filtering the data

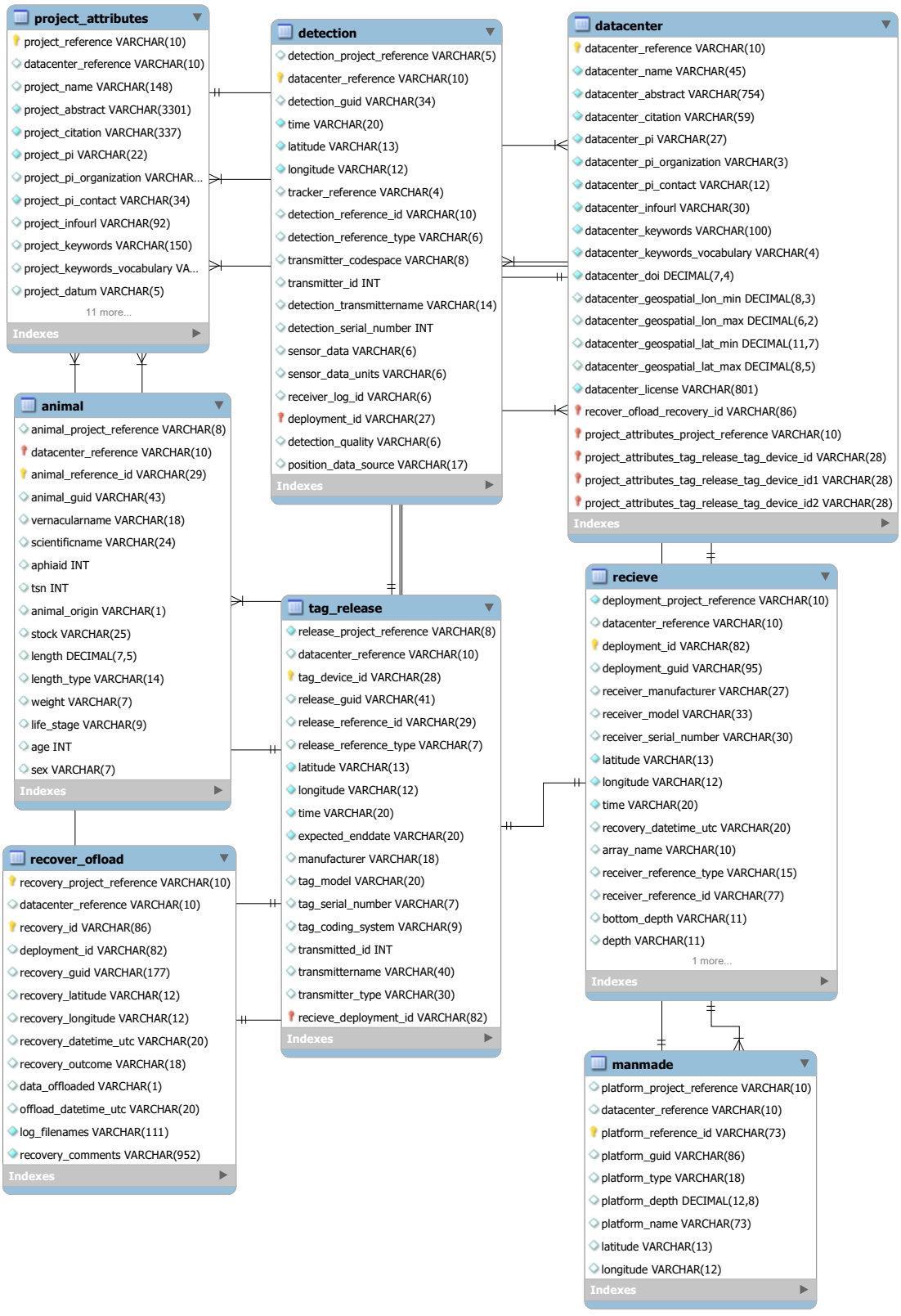
- Otnunit_att_animals
  - Taxonrank – Empty column – Removed completely
  - Animal_Origin – For every null value and aphiaid = 217628 and scientific name = Notorynchus cepedianus we will change the blank with W as per the previous records.
  - Stock – There are three data inputs UNK, Unknown and Blanks which will be replaced by No Data.
  - Length – The records where there is no data for stock are majorly NAN in length hence, we need to make sure there isn't any issue with any average hence we will change them by the average length of similar aphiaid. Where the information couldn't been found it had put the null values.
  - Length_types – Blank rows with No Data.
  - Life_stage – blank rows with No Data.
  - Age – its int data type hence removing nan with NULL.
  - Sex – Sex cannot be predicted hence we put UKNOWN at blank positions.
- Otnunit_aat_datacenter
  - Datacenter distribution statement – Empty hence removed
  - Datacenter date modified – Empty hence removed
  - time_coverage_start – Empty hence removed
  - time_coverage_end – Empty hence removed
  - datacenter_geospatial_lon_min, datacenter_geospatial_lon_max, datacenter_geospatial_lat_min, datacenter_geospatial_lat_max – wherever there is na values changes with NULL
- otnunit_aat_detection
  - Sensor_data, Sensor_data_units – All blank is replaced with unknown.
  - receiver_log_id, detection_quantity – All blank is replaced with unknown.
  - Depth – the whole column does not hold any substantial values, hence removing the whole column.
  - Uncertainty_in_latitude, Uncertainty_in_longitude – the whole column is nan hence removing the column.
  - Depth_data_source, uncertainty_in_depth, other_position_data, dataset quantity – blank column, hence removing them.
- otnunit_aat_manmade
  - platform_depth, latitude_degrees_north, longitude_degrees_north – wherever the data is nan change it with NULL.
- otnunit_aat_project
  - time_coverage_end, time_coverage_start – whole column is empty hence remove the column.
  - geospatial_vertical_positive, project_linestring – whole column is empty hence remove the column.
  - Project_date_modified, project_distribution statement, project_doi, project_references – removing the whole column.
  - Project_infourl – Blank and na cells are <NULL>
  - Project_pi_contact, project_pi, project_citation, project_abstract – Blanks are converted to No Data.

- Otnunit_aat_receivers
    - Receiver_manufacturer, receiver_seriel_number, receiver_reference_id, deployment_comments – blank row converted to No Data
    - Frequencies_monitored, receiver_coding_scheme, deployed_by – blank column hence removed.
    - Recovery datetime utc – some blank cells are there replaced by NULL
    - Bottom_depth, depth – nan to NULL
    - Expected_reciever_life – removed because it doesn't hold any value

- Otunit_aat_recover_offload_details
    - Log_files, recovery_comments – blank replaced with No Data
    - Recovery_datetime_utc, offload_daytime_utc – nan or blank to NULL
    - Clock synchronized, recovered_by – blank so removed

- Otnunit_aat_tag_releases
    - Tag_frequency, transmitter_type, tag_programming_id – blank hence removed

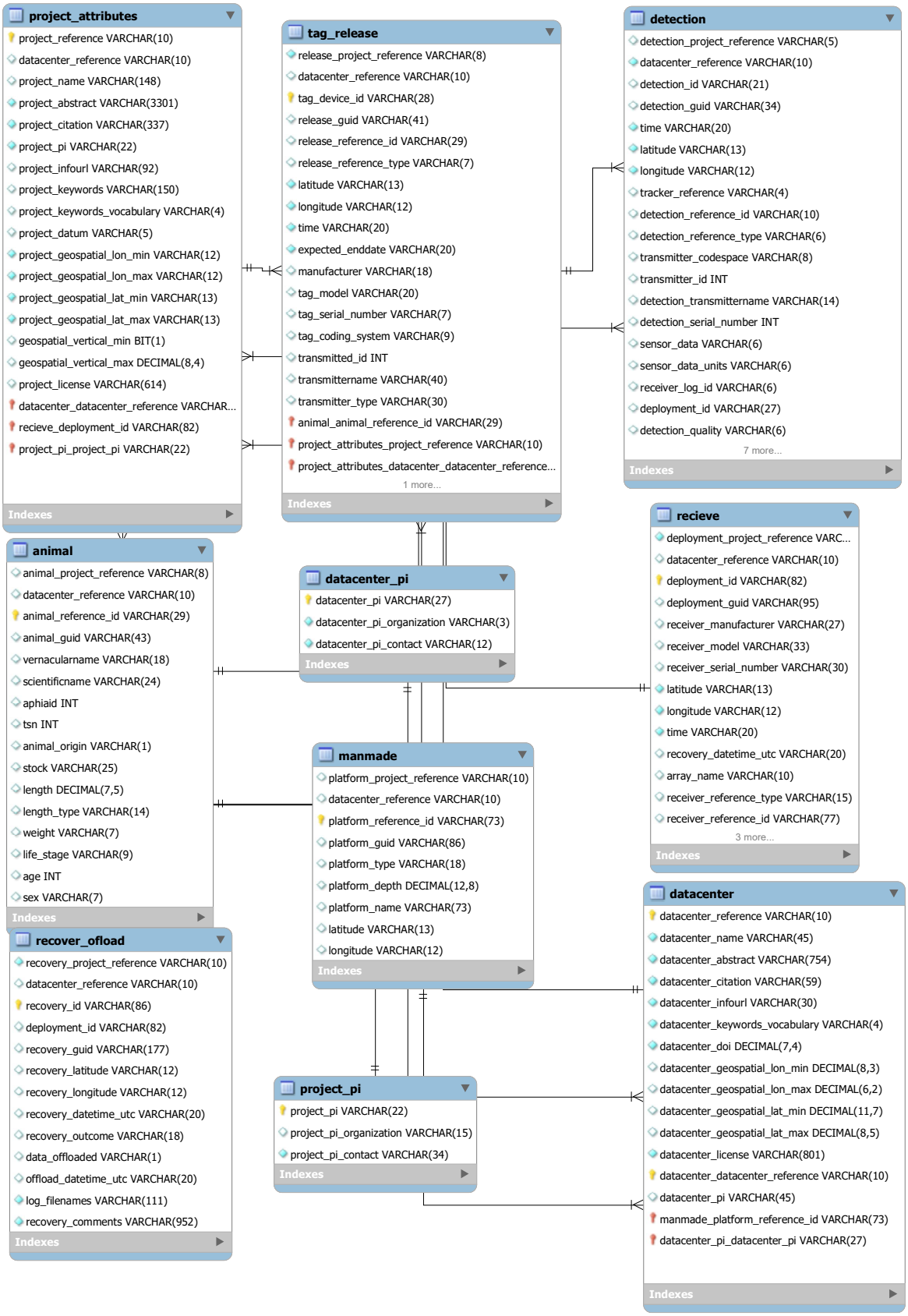
On the following page we could have a pre normalization reverse engineered ERD diagram. That consist of all the relationships and cardinalities.


project_attributes
project_reference VARCHAR(10)
datacenter_reference VARCHAR(10)
project_name VARCHAR(148)
project_abstract VARCHAR(3301)
project_citation VARCHAR(337)
project_pi VARCHAR(22)
project_pi_organization VARCHAR...
project_pi_contact VARCHAR(34)
project_infourl VARCHAR(92)
project_keywords VARCHAR(150)
project_keywords_vocabulary VA...
project_datum VARCHAR(5)
11 more...
Indexes
detection
detection_project_reference VARCHAR(5)
datacenter_reference VARCHAR(10)
detection_guid VARCHAR(34)
time VARCHAR(20)
latitude VARCHAR(13)
longitude VARCHAR(12)
tracker_reference VARCHAR(4)
detection_reference_id VARCHAR(10)
detection_reference_type VARCHAR(6)
transmitter_codespace VARCHAR(8)
transmitter_id INT
detection_transmittername VARCHAR(14)
detection_serial_number INT
sensor_data VARCHAR(6)
sensor_data_units VARCHAR(6)
receiver_log_id VARCHAR(6)
deployment_id VARCHAR(27)
detection_quality VARCHAR(6)
position_data_source VARCHAR(17)
Indexes
datacenter
datacenter_reference VARCHAR(10)
datacenter_name VARCHAR(45)
datacenter_abstract VARCHAR(754)
datacenter_citation VARCHAR(59)
datacenter_pi VARCHAR(27)
datacenter_pi_organization VARCHAR(3)
datacenter_pi_contact VARCHAR(12)
datacenter_infourl VARCHAR(30)
datacenter_keywords VARCHAR(100)
datacenter_keywords_vocabulary VARCHAR(4)
datacenter_doi DECIMAL(7,4)
datacenter_geospatial_lon_min DECIMAL(8,3)
datacenter_geospatial_lon_max DECIMAL(6,2)
datacenter_geospatial_lat_min DECIMAL(11,7)
datacenter_geospatial_lat_max DECIMAL(8,5)
datacenter_license VARCHAR(801)
recover_ofload_recovery_id VARCHAR(86)
project_attributes_project_reference VARCHAR(10)
project_attributes_tag_release_tag_device_id VARCHAR(28)
project_attributes_tag_release_tag_device_id1 VARCHAR(28)
project_attributes_tag_release_tag_device_id2 VARCHAR(28)
Indexes
animal
animal_project_reference VARCHAR(8)
datacenter_reference VARCHAR(10)
animal_reference_id VARCHAR(29)
animal_guid VARCHAR(43)
vernacularname VARCHAR(18)
scientificname VARCHAR(24)
aphiaid INT
tsn INT
animal_origin VARCHAR(1)
stock VARCHAR(25)
length DECIMAL(7,5)
length_type VARCHAR(14)
weight VARCHAR(7)
life_stage VARCHAR(9)
age INT
sex VARCHAR(7)
Indexes
tag_release
release_project_reference VARCHAR(8)
datacenter_reference VARCHAR(10)
tag_device_id VARCHAR(28)
release_guid VARCHAR(41)
release_reference_id VARCHAR(29)
release_reference_type VARCHAR(7)
latitude VARCHAR(13)
longitude VARCHAR(12)
time VARCHAR(20)
expected_enddate VARCHAR(20)
manufacturer VARCHAR(18)
tag_model VARCHAR(20)
tag_serial_number VARCHAR(7)
tag_coding_system VARCHAR(9)
transmitted_id INT
transmittername VARCHAR(40)
transmitter_type VARCHAR(30)
recieve_deployment_id VARCHAR(82)
Indexes
recieve
deployment_project_reference VARCHAR(10)
datacenter_reference VARCHAR(10)
deployment_id VARCHAR(82)
deployment_guid VARCHAR(95)
receiver_manufacturer VARCHAR(27)
receiver_model VARCHAR(33)
receiver_serial_number VARCHAR(30)
latitude VARCHAR(13)
longitude VARCHAR(12)
time VARCHAR(20)
recovery_datetime_utc VARCHAR(20)
array_name VARCHAR(10)
receiver_reference_type VARCHAR(15)
receiver_reference_id VARCHAR(77)
bottom_depth VARCHAR(11)
depth VARCHAR(11)
1 more...
Indexes
recover_ofload
recovery_project_reference VARCHAR(10)
datacenter_reference VARCHAR(10)
recovery_id VARCHAR(86)
deployment_id VARCHAR(82)
recovery_guid VARCHAR(177)
recovery_latitude VARCHAR(12)
recovery_longitude VARCHAR(12)
recovery_datetime_utc VARCHAR(20)
recovery_outcome VARCHAR(18)
data_offloaded VARCHAR(1)
offload_datetime_utc VARCHAR(20)
log_filenames VARCHAR(111)
recovery_comments VARCHAR(952)
Indexes
manmade
platform_project_reference VARCHAR(10)
datacenter_reference VARCHAR(10)
platform_reference_id VARCHAR(73)
platform_guid VARCHAR(86)
platform_type VARCHAR(18)
platform_depth DECIMAL(12,8)
platform_name VARCHAR(73)
latitude VARCHAR(13)
longitude VARCHAR(12)
Indexes

Normalization:

- In datacenter table, it includes multivalued attributes keyword datacenter column, so this violated 1NF norms. To solve the issue, create a new table named keywords with two fields named references_datacenter and keywords_datacenter. After doing this step delete the multivalued attribute keyword datacenter from the table.
- Datacenter table has non – prime attributes (datacenter_pi, datacenter_pi_contact, datacenter_pi_organisation) which could uniquely identified from datacenter_pi by putting it as the primary key of the newer generated table.
    - Table 1: Datacenter_pi_details(datacenter_pi (PK), datacenter_pi_contact, datacenter_pi_organisation)
    - In the original table the Datacenter_pi will become primary key and datacenter_pi_contact, datacenter_pi_organisation would be removed from that table.
- Project_Attribute table is not also in 2NF, hence we will continue with the same method we proceed in the previous step.
    - Here the issue lies with project_pi, project_pi_contact, project_pi_organisation where the tables could be designed in the form like where one table is main holding all the information with project_pi as foreign key and then the second table with project_pi as the primary key and project_pi_contact, project_pi_organisation as an attribute.

On the following page you could find the post normalization ERD diagram


project_attributes
project_reference VARCHAR(10)
datacenter_reference VARCHAR(10)
project_name VARCHAR(148)
project_abstract VARCHAR(3301)
project_citation VARCHAR(337)
project_pi VARCHAR(22)
project_infourl VARCHAR(92)
project_keywords VARCHAR(150)
project_keywords_vocabulary VARCHAR(4)
project_datum VARCHAR(5)
project_geospatial_lon_min VARCHAR(12)
project_geospatial_lon_max VARCHAR(12)
project_geospatial_lat_min VARCHAR(13)
project_geospatial_lat_max VARCHAR(13)
geospatial_vertical_min BIT(1)
geospatial_vertical_max DECIMAL(8,4)
project_license VARCHAR(614)
datacenter_datacenter_reference VARCHAR...
recieve_deployment_id VARCHAR(82)
project_pi_project_pi VARCHAR(22)
Indexes
tag_release
release_project_reference VARCHAR(8)
datacenter_reference VARCHAR(10)
tag_device_id VARCHAR(28)
release_guid VARCHAR(41)
release_reference_id VARCHAR(29)
release_reference_type VARCHAR(7)
latitude VARCHAR(13)
longitude VARCHAR(12)
time VARCHAR(20)
expected_enddate VARCHAR(20)
manufacturer VARCHAR(18)
tag_model VARCHAR(20)
tag_serial_number VARCHAR(7)
tag_coding_system VARCHAR(9)
transmitted_id INT
transmittername VARCHAR(40)
transmitter_type VARCHAR(30)
animal_animal_reference_id VARCHAR(29)
project_attributes_project_reference VARCHAR(10)
project_attributes_datacenter_datacenter_reference...
1 more...
Indexes
detection
detection_project_reference VARCHAR(5)
datacenter_reference VARCHAR(10)
detection_id VARCHAR(21)
detection_guid VARCHAR(34)
time VARCHAR(20)
latitude VARCHAR(13)
longitude VARCHAR(12)
tracker_reference VARCHAR(4)
detection_reference_id VARCHAR(10)
detection_reference_type VARCHAR(6)
transmitter_codespace VARCHAR(8)
transmitter_id INT
detection_transmittername VARCHAR(14)
detection_serial_number INT
sensor_data VARCHAR(6)
sensor_data_units VARCHAR(6)
receiver_log_id VARCHAR(6)
deployment_id VARCHAR(27)
detection_quality VARCHAR(6)
7 more...
Indexes
animal
animal_project_reference VARCHAR(8)
datacenter_reference VARCHAR(10)
animal_reference_id VARCHAR(29)
animal_guid VARCHAR(43)
vernacularname VARCHAR(18)
scientificname VARCHAR(24)
aphiaid INT
tsn INT
animal_origin VARCHAR(1)
stock VARCHAR(25)
length DECIMAL(7,5)
length_type VARCHAR(14)
weight VARCHAR(7)
life_stage VARCHAR(9)
age INT
sex VARCHAR(7)
Indexes
datacenter_pi
datacenter_pi VARCHAR(27)
datacenter_pi_organization VARCHAR(3)
datacenter_pi_contact VARCHAR(12)
Indexes
recieve
deployment_project_reference VARC...
datacenter_reference VARCHAR(10)
deployment_id VARCHAR(82)
deployment_guid VARCHAR(95)
receiver_manufacturer VARCHAR(27)
receiver_model VARCHAR(33)
receiver_serial_number VARCHAR(30)
latitude VARCHAR(13)
longitude VARCHAR(12)
time VARCHAR(20)
recovery_datetime_utc VARCHAR(20)
array_name VARCHAR(10)
receiver_reference_type VARCHAR(15)
receiver_reference_id VARCHAR(77)
3 more...
Indexes
manmade
platform_project_reference VARCHAR(10)
datacenter_reference VARCHAR(10)
platform_reference_id VARCHAR(73)
platform_guid VARCHAR(86)
platform_type VARCHAR(18)
platform_depth DECIMAL(12,8)
platform_name VARCHAR(73)
latitude VARCHAR(13)
longitude VARCHAR(12)
Indexes
recover_ofload
recovery_project_reference VARCHAR(10)
datacenter_reference VARCHAR(10)
recovery_id VARCHAR(86)
deployment_id VARCHAR(82)
recovery_guid VARCHAR(177)
recovery_latitude VARCHAR(12)
recovery_longitude VARCHAR(12)
recovery_datetime_utc VARCHAR(20)
recovery_outcome VARCHAR(18)
data_offloaded VARCHAR(1)
offload_datetime_utc VARCHAR(20)
log_filenames VARCHAR(111)
recovery_comments VARCHAR(952)
Indexes
project_pi
project_pi VARCHAR(22)
project_pi_organization VARCHAR(15)
project_pi_contact VARCHAR(34)
Indexes
datacenter
datacenter_reference VARCHAR(10)
datacenter_name VARCHAR(45)
datacenter_abstract VARCHAR(754)
datacenter_citation VARCHAR(59)
datacenter_infourl VARCHAR(30)
datacenter_keywords_vocabulary VARCHAR(4)
datacenter_doi DECIMAL(7,4)
datacenter_geospatial_lon_min DECIMAL(8,3)
datacenter_geospatial_lon_max DECIMAL(6,2)
datacenter_geospatial_lat_min DECIMAL(11,7)
datacenter_geospatial_lat_max DECIMAL(8,5)
datacenter_license VARCHAR(801)
datacenter_datacenter_reference VARCHAR(10)
datacenter_pi VARCHAR(45)
manmade_platform_reference_id VARCHAR(73)
datacenter_pi_datacenter_pi VARCHAR(27)
Indexes

Opportunities in Halifax

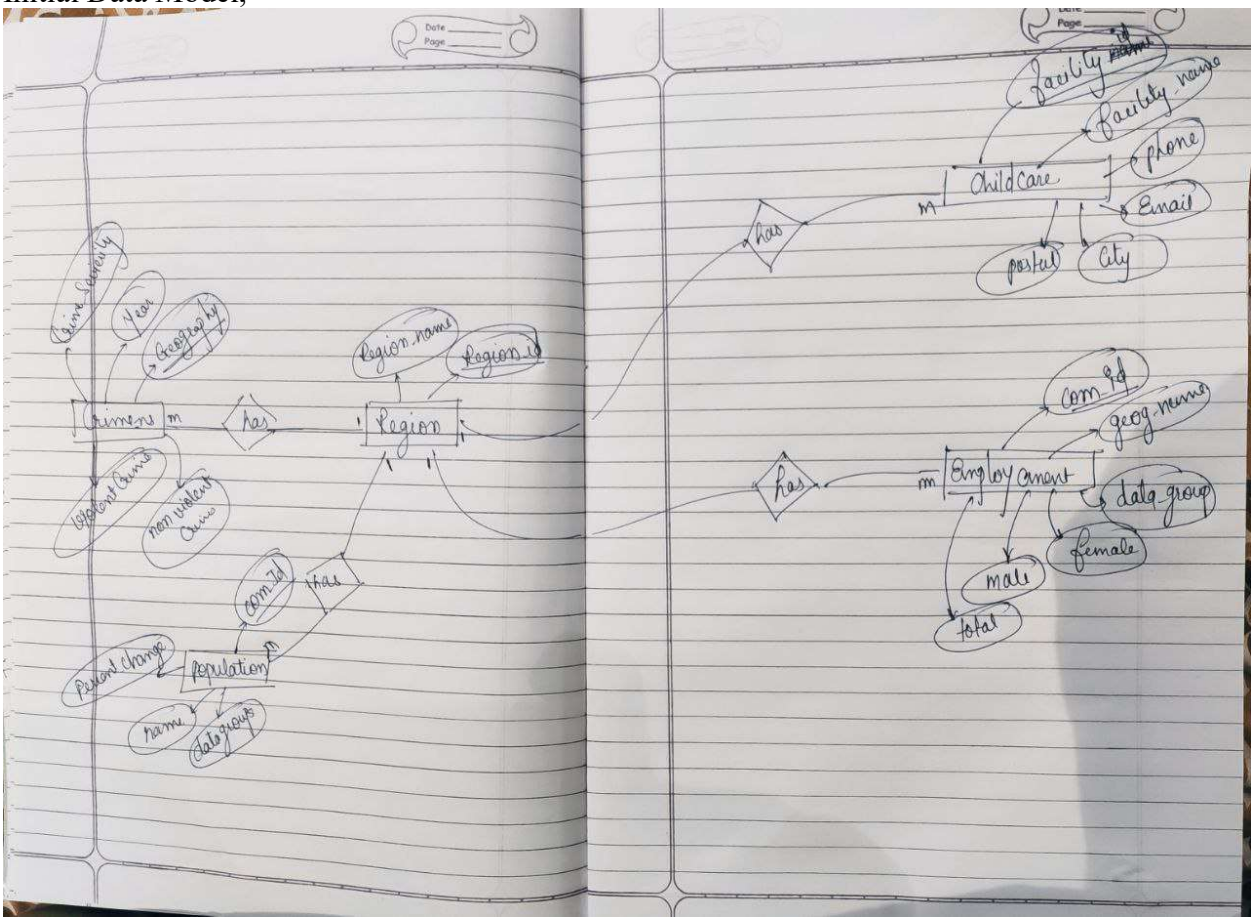
Here in this part of the assignment we need to gather 5 entities for the real-estate client who is looking to buy some property in nova scotia region.

For that work these are the five entities which had been finalized

- Region,
- Crime_ns,
- Population,
- Childcare and
- Employment

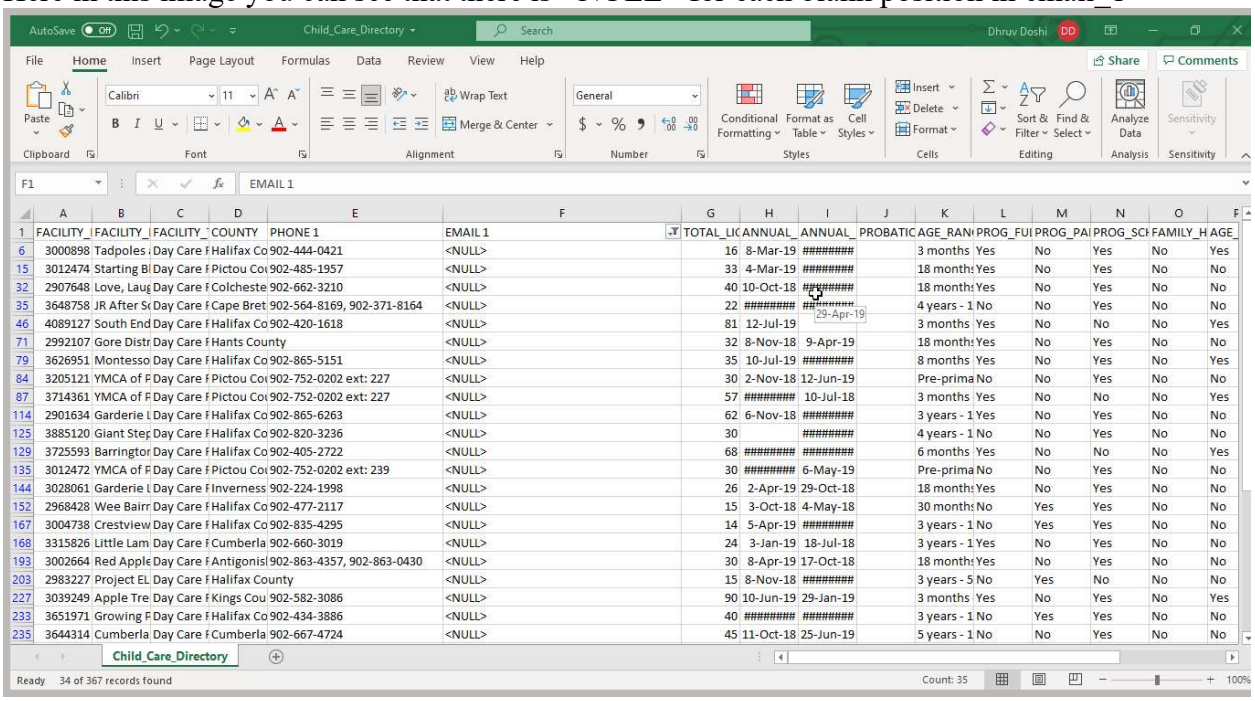
To gather the data sets https://data.novascotia.ca/ is used as per stated in the assignment document.

Initial Data Model,



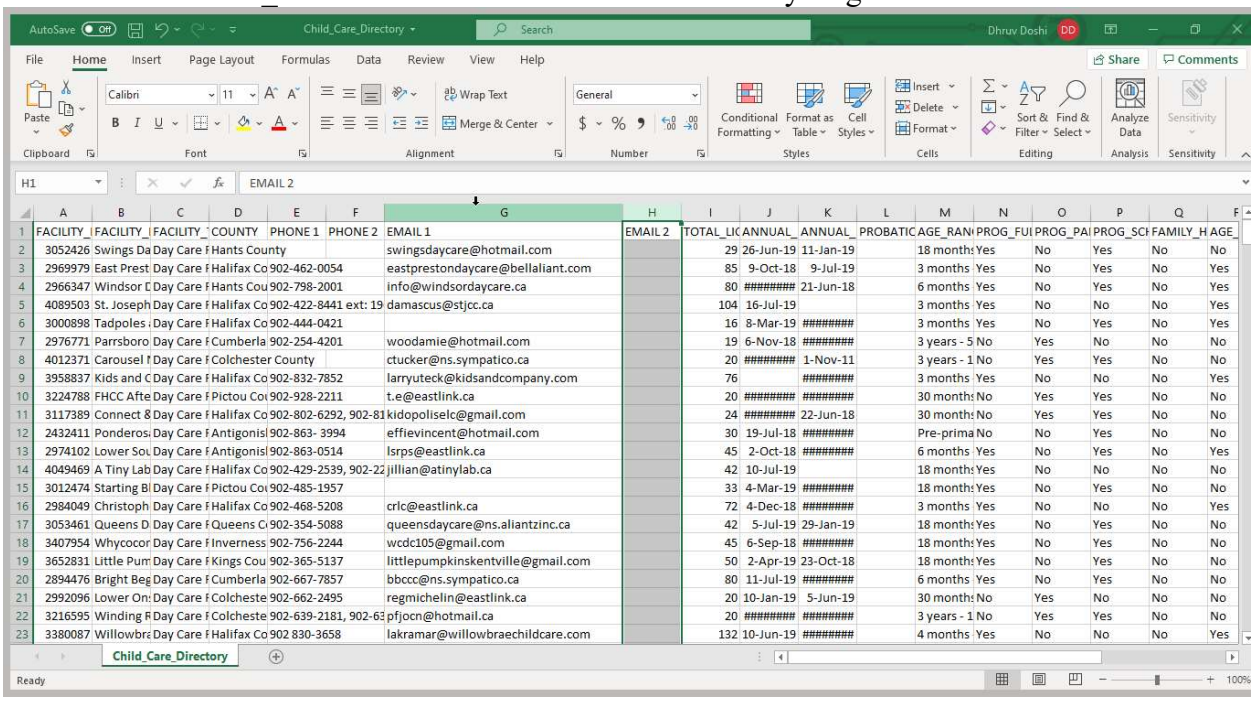
After finding the data sets I have used several Excel techniques to narrow down the data by cleaning the given data. Broadly wherever the data field in blank, it is filled by NULL values for numerical and No Data for the Varchar data types.

Here in this image you can see that there is <NULL> for each blank position in email_1



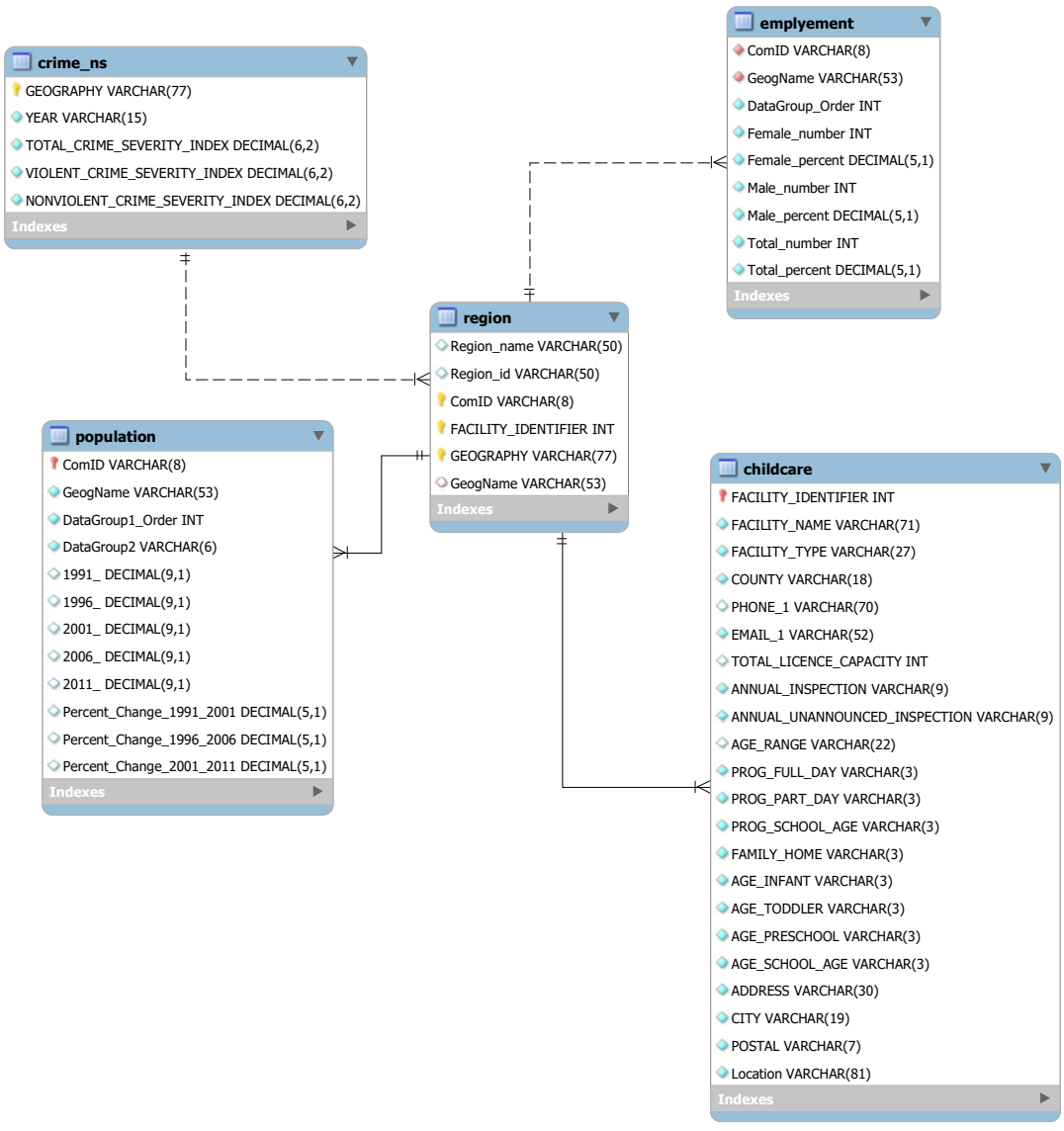
FACILITY_	FACILITY_	FACILITY_	COUNTY	PHONE 1	EMAIL 1	TOTAL_LIC	ANNUAL_	ANNUAL_	PROBATIC	AGE_RAN	PROG_FU	PROG_PA	PROG_SC	FAMILY_H	AGE_
3000898	Tadpoles	Day Care	Halifax Co	902-444-0421	<NULL>	16	8-Mar-19	########		3 months	Yes	No	Yes	No	Yes
3012474	Starting B	Day Care	Pictou Co	902-485-1957	<NULL>	33	4-Mar-19	########		18 months	Yes	No	Yes	No	No
2907648	Love, Laug	Day Care	Colcheste	902-662-3210	<NULL>	40	10-Oct-18	########		18 months	Yes	No	Yes	No	No
3648758	JR After S	Day Care	Cape Bret	902-564-8169, 902-371-8164	<NULL>	22	########	########		4 years - 1	No	No	Yes	No	No
4089127	South End	Day Care	Halifax Co	902-420-1618	<NULL>	81	12-Jul-19			3 months	Yes	No	No	No	Yes
2992107	Gore Distr	Day Care	Hants County		<NULL>	32	8-Nov-18	9-Apr-19		18 months	Yes	No	Yes	No	No
3626951	Montesso	Day Care	Halifax Co	902-865-5151	<NULL>	35	10-Jul-19	########		8 months	Yes	No	Yes	No	Yes
3205121	YMCA of F	Day Care	Pictou Co	902-752-0202 ext: 227	<NULL>	30	2-Nov-18	12-Jun-19		Pre-prima	No	No	Yes	No	No
3714361	YMCA of F	Day Care	Pictou Co	902-752-0202 ext: 227	<NULL>	57	########	10-Jul-18		3 months	Yes	No	No	No	Yes
2901634	Garderie L	Day Care	Halifax Co	902-865-6263	<NULL>	62	6-Nov-18	########		3 years - 1	Yes	No	Yes	No	No
3885120	Giant Step	Day Care	Halifax Co	902-820-3236	<NULL>	30		########		4 years - 1	No	No	Yes	No	No
3725593	Barrington	Day Care	Halifax Co	902-405-2722	<NULL>	68	########	########		6 months	Yes	No	No	No	Yes
3012472	YMCA of F	Day Care	Pictou Co	902-752-0202 ext: 239	<NULL>	30	########	6-May-19		Pre-prima	No	No	Yes	No	No
3028061	Garderie L	Day Care	Inverness	902-224-1998	<NULL>	26	2-Apr-19	29-Oct-18		18 months	Yes	No	Yes	No	No
2968428	Wee Bairn	Day Care	Halifax Co	902-477-2117	<NULL>	15	3-Oct-18	4-May-18		30 months	No	Yes	Yes	No	No
3004738	Crestview	Day Care	Halifax Co	902-835-4295	<NULL>	14	5-Apr-19	########		3 years - 1	No	Yes	Yes	No	No
3315826	Little Lam	Day Care	Cumberla	902-660-3019	<NULL>	24	3-Jan-19	18-Jul-18		3 years - 1	Yes	No	Yes	No	No
3002664	Red Apple	Day Care	Antigonis	902-863-4357, 902-863-0430	<NULL>	30	8-Apr-19	17-Oct-18		18 months	Yes	No	Yes	No	No
2983227	Project EL	Day Care	Halifax County		<NULL>	15	8-Nov-18	########		3 years - 5	No	Yes	No	No	No
3039249	Apple Tre	Day Care	Kings Cou	902-582-3086	<NULL>	90	10-Jun-19	29-Jan-19		3 months	Yes	No	Yes	No	Yes
3651971	Growing F	Day Care	Halifax Co	902-434-3886	<NULL>	40	########	########		3 years - 1	No	Yes	Yes	No	No
3644314	Cumberla	Day Care	Cumberla	902-667-4724	<NULL>	45	11-Oct-18	25-Jun-19		5 years - 1	No	No	Yes	No	No

Whereas here Email_2 had been deleted as it does not had any single value in it.



FACILITY_	FACILITY_	FACILITY_	COUNTY	PHONE 1	PHONE 2	EMAIL 1	EMAIL 2	TOTAL_LIC	ANNUAL_	ANNUAL_	PROBATIC	AGE_RAN	PROG_FU	PROG_PA	PROG_SC	FAMILY_H	AGE_
3052426	Swings Da	Day Care	Hants County			swingsdaycare@hotmail.com		29	26-Jun-19	11-Jan-19		18 months	Yes	No	Yes	No	Yes
2969979	East Prest	Day Care	Halifax Co	902-462-0054		eastprestondaycare@bellaliant.com		85	9-Oct-18	9-Jul-19		3 months	Yes	No	Yes	No	Yes
2966347	Windsor D	Day Care	Hants Cou	902-798-2001		info@windsordaycare.ca		80	########	21-Jun-18		6 months	Yes	No	Yes	No	Yes
4089503	St. Joseph	Day Care	Halifax Co	902-422-8441 ext: 19		damascus@stjcc.ca		104	16-Jul-19			3 months	Yes	No	No	No	Yes
3000898	Tadpoles	Day Care	Halifax Co	902-444-0421				16	8-Mar-19	########		3 months	Yes	No	Yes	No	Yes
2976771	Parrsboro	Day Care	Cumberla	902-254-4201		woodamie@hotmail.com		19	6-Nov-18	########		3 years - 5	No	Yes	Yes	No	No
4012371	Carousel I	Day Care	Colchester County			ctucker@ns.sympatico.ca		20	########	1-Nov-11		3 years - 1	No	Yes	Yes	No	No
3958837	Kids and C	Day Care	Halifax Co	902-832-7852		larryuteck@kidsandcompany.com		76		########		3 months	Yes	No	No	No	Yes
3224788	FHCC Afte	Day Care	Pictou Co	902-928-2211		t.e@eastlink.ca		20	########	########		30 months	No	Yes	Yes	No	No
3117389	Connect &	Day Care	Halifax Co	902-802-6292, 902-81		kidopoliselc@gmail.com		24	########	22-Jun-18		30 months	No	Yes	Yes	No	No
2432411	Ponderos	Day Care	Antigonis	902-863- 3994		effievincent@hotmail.com		30	19-Jul-18	########		Pre-prima	No	No	Yes	No	No
2974102	Lower Sou	Day Care	Antigonis	902-863-0514		lsrps@eastlink.ca		45	2-Oct-18	########		6 months	Yes	No	Yes	No	Yes
4049469	A Tiny Lab	Day Care	Halifax Co	902-429-2539, 902-22		jillian@atinylab.ca		42	10-Jul-19			18 months	Yes	No	No	No	No
3012474	Starting B	Day Care	Pictou Co	902-485-1957				33	4-Mar-19	########		18 months	Yes	No	Yes	No	No
2984049	Christoph	Day Care	Halifax Co	902-468-5208		crlc@eastlink.ca		72	4-Dec-18	########		3 months	Yes	No	No	No	Yes
3053461	Queens D	Day Care	Queens C	902-354-5088		queensdaycare@ns.aliantzinc.ca		42	5-Jul-19	29-Jan-19		18 months	Yes	No	Yes	No	No
3407954	Whycocor	Day Care	Inverness	902-756-2244		wcdc105@gmail.com		45	6-Sep-18	########		18 months	Yes	No	Yes	No	No
3652831	Little Pum	Day Care	Kings Cou	902-365-5137		littlepumpkinskentville@gmail.com		50	2-Apr-19	23-Oct-18		18 months	Yes	No	Yes	No	No
2894476	Bright Beg	Day Care	Cumberla	902-667-7857		bbccc@ns.sympatico.ca		80	11-Jul-19	########		6 months	Yes	No	Yes	No	No
2992096	Lower On	Day Care	Colcheste	902-662-2495		regmichelin@eastlink.ca		20	10-Jan-19	5-Jun-19		30 months	No	Yes	No	No	No
3216595	Winding R	Day Care	Colcheste	902-639-2181, 902-63		pfjocn@hotmail.ca		20	########	########		3 years - 1	No	Yes	Yes	No	No
3380087	Willowbra	Day Care	Halifax Co	902 830-3658		lakramar@willowbraechildcare.com		132	10-Jun-19	########		4 months	Yes	No	No	No	Yes

The following page consists the ERD obtained by MySql workbench.


crime_ns
GEOGRAPHY VARCHAR(77)
YEAR VARCHAR(15)
TOTAL_CRIME_SEVERITY_INDEX DECIMAL(6,2)
VIOLENT_CRIME_SEVERITY_INDEX DECIMAL(6,2)
NONVIOLENT_CRIME_SEVERITY_INDEX DECIMAL(6,2)
Indexes
emplyement
ComID VARCHAR(8)
GeogName VARCHAR(53)
DataGroup_Order INT
Female_number INT
Female_percent DECIMAL(5,1)
Male_number INT
Male_percent DECIMAL(5,1)
Total_number INT
Total_percent DECIMAL(5,1)
Indexes
region
Region_name VARCHAR(50)
Region_id VARCHAR(50)
ComID VARCHAR(8)
FACILITY_IDENTIFIER INT
GEOGRAPHY VARCHAR(77)
GeogName VARCHAR(53)
Indexes
population
ComID VARCHAR(8)
GeogName VARCHAR(53)
DataGroup1_Order INT
DataGroup2 VARCHAR(6)
1991_ DECIMAL(9,1)
1996_ DECIMAL(9,1)
2001_ DECIMAL(9,1)
2006_ DECIMAL(9,1)
2011_ DECIMAL(9,1)
Percent_Change_1991_2001 DECIMAL(5,1)
Percent_Change_1996_2006 DECIMAL(5,1)
Percent_Change_2001_2011 DECIMAL(5,1)
Indexes
childcare
FACILITY_IDENTIFIER INT
FACILITY_NAME VARCHAR(71)
FACILITY_TYPE VARCHAR(27)
COUNTY VARCHAR(18)
PHONE_1 VARCHAR(70)
EMAIL_1 VARCHAR(52)
TOTAL_LICENCE_CAPACITY INT
ANNUAL_INSPECTION VARCHAR(9)
ANNUAL_UNANNOUNCED_INSPECTION VARCHAR(9)
AGE_RANGE VARCHAR(22)
PROG_FULL_DAY VARCHAR(3)
PROG_PART_DAY VARCHAR(3)
PROG_SCHOOL_AGE VARCHAR(3)
FAMILY_HOME VARCHAR(3)
AGE_INFANT VARCHAR(3)
AGE_TODDLER VARCHAR(3)
AGE_PRESCHOOL VARCHAR(3)
AGE_SCHOOL_AGE VARCHAR(3)
ADDRESS VARCHAR(30)
CITY VARCHAR(19)
POSTAL VARCHAR(7)
Location VARCHAR(81)
Indexes

References

[1]"Flowchart Maker & Online Diagram Software", *App.diagrams.net*, 2021. [Online]. Available: https://app.diagrams.net/. [Accessed: 01- Jun- 2021]

[2]*Data.novascotia.ca*, 2021. [Online]. Available: https://data.novascotia.ca/. [Accessed: 01- Jun- 2021]

[3]"How to import a CSV file into a MySQL database?", *Medium*, 2021. [Online]. Available: https://medium.com/@AviGoom/how-to-import-a-csv-file-into-a-mysql-database-ef8860878a68. [Accessed: 01- Jun- 2021]