**Section-A**

1. I have seen many datasets that provide information about Nova Scotia and Halifax. All these datasets will be helpful in identifying the key performance indicators to improve business, education, lifestyle and safety of the Halifax region.
2. Name: halifax\_employment\_traded\_2017[4]

URL: <https://www.competeprosper.ca/clusters/data/by-region/halifax>

Reason: This dataset provides information about employment provided by each cluster in the region of Halifax for the year 2017. This can help in analyzing which business cluster provides more employment opportunities.

Name: halifax\_employment\_traded\_2016

URL: <https://www.competeprosper.ca/clusters/data/by-region/halifax>

Reason: This dataset provides information about employment provided by each cluster in the region of Halifax for the year 2016. This can help in analyzing which business cluster provides more employment opportunities.

Name: halifax\_employment\_traded\_2015

URL: <https://www.competeprosper.ca/clusters/data/by-region/halifax>

Reason: This dataset provides information about employment provided by each cluster in the region of Halifax for the year 2015. This can help in analyzing which business cluster provides more employment opportunities.

Name: halifax\_employment\_traded\_2014

URL: <https://www.competeprosper.ca/clusters/data/by-region/halifax>

Reason: This dataset provides information about employment provided by each cluster in the region of Halifax for the year 2014. This can help in analyzing which business cluster provides more employment opportunities.

Name: halifax\_employment\_traded\_2013

URL: <https://www.competeprosper.ca/clusters/data/by-region/halifax>

Reason: This dataset provides information about employment provided by each cluster in the region of Halifax for the year 2013. This can help in analyzing which business cluster provides more employment opportunities.

Name: halifax\_employment\_traded\_2012

URL: <https://www.competeprosper.ca/clusters/data/by-region/halifax>

Reason: This dataset provides information about employment provided by each cluster in the region of Halifax for the year 2012. This can help in analyzing which business cluster provides more employment opportunities.

Name: halifax\_employment\_traded\_2011

URL: <https://www.competeprosper.ca/clusters/data/by-region/halifax>

Reason: This dataset provides information about employment provided by each cluster in the region of Halifax for the year 2011. This can help in analyzing which business cluster provides more employment opportunities.

Name: halifax\_employment\_traded\_2010

URL: <https://www.competeprosper.ca/clusters/data/by-region/halifax>

Reason: This dataset provides information about employment provided by each cluster in the region of Halifax for the year 2010. This can help in analyzing which business cluster provides more employment opportunities.

Name: Education\_Survey

URL: <https://data.novascotia.ca/Education-Post-Secondary-and-Skills-Training/-ARCHIVED-NHS-Education-by-Location-of-Education/ut6n-ijt8>

Reason: This dataset provides information about educated Female and Male count in each Location. It also provides participation percent of females and males.

Name: job\_type\_and\_education\_needed

URL: <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1410032801&pickMembers%5B0%5D=1.4&pickMembers%5B1%5D=2.1>

Reason: This dataset provides information about Job vacancies, the proportion of job vacancies and average offered hourly wage by selected characteristics, quarterly, unadjusted for seasonality for the nova scotia province.

Name: ns\_jobs[5]

URL: <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1410035601&pickMembers%5B0%5D=1.4>

Reason: This dataset provides information about Job vacancies and average offered hourly wage by occupation (broad occupational category), quarterly, unadjusted for seasonality for the nova scotia province.

Name: health\_male\_female\_halifax

URL: <https://www12.statcan.gc.ca/health-sante/82-228/details/page.cfm?Lang=E&Tab=1&Geo1=CMA&Code1=205&Geo2=PR&Code2=01&Data=Rate&SearchText=Halifax&SearchType=Contains&SearchPR=01&B1=All&Custom=&B2=All&B3=All>

Reason: This dataset provides information about health profile, health conditions and well beings of females and males of Halifax and Canada in December 2013.

Name: ns\_health\_male\_female

URL: <https://www12.statcan.gc.ca/health-sante/82-228/details/page.cfm?Lang=E&Tab=1&Geo1=PR&Code1=12&Geo2=PR&Code2=01&Data=Rate&SearchText=Nova%20Scotia&SearchType=Contains&SearchPR=01&B1=All&Custom=&B2=All&B3=All&GeoLevel=PR&GeoCode=12>

Reason: This dataset provides information about health profile, health conditions and well beings of females and males of nova scotia and Canada in December 2013.

Name: crime\_safety\_home\_alone[2]

URL: <https://data.novascotia.ca/Crime-and-Justice/Perceived-safety-from-crime-when-home-alone-at-nig/nw2i-ymps>

Reason: This dataset provides data on the level of perceived safety from crime when home at night reported by Nova Scotians and is based on a question included in the Atlantic Quarterly survey administered by Corporate Research Associates (CRA).

Name: courts\_list

URL: <https://data.novascotia.ca/Crime-and-Justice/Justice-Centres-Courts/72n8-bk3g>

Reason: This dataset lists the location of Justice Centres/Courts throughout the province

Name: crime\_count\_adults\_years

URL: <https://data.novascotia.ca/Crime-and-Justice/Average-Daily-Counts-at-Adult-Correctional-Facilit/m8rv-4ygy>

Reason: This dataset provides information about average daily counts at adult correctional facilities by year.

Name: crime\_stastics\_year

URL: <https://data.novascotia.ca/Crime-and-Justice/Crime-Statistics-Crime-Severity-Index/w64p-5ue3>

Reason: The dataset includes crime statistics at the provincial and police service levels. Statistics included are youth crime severity index, youth violent crime severity index and youth non-violent crime severity index for provincial-level data.

Name: trip\_purpose\_exit\_survey

URL: <https://data.novascotia.ca/Business-and-Industry/Nova-Scotia-Visitor-Exit-Survey-Trip-Purpose-by-Ye/kb9w-de7x>

Reason: This dataset provides the purpose for the visit over the years. This helps us in identifying which performance indicator is main for Halifax.

Name: Invest\_Nova\_Scotia\_Fund\_Applications

URL: <https://data.novascotia.ca/Business-and-Industry/Invest-Nova-Scotia-Fund-Applications/bcxb-v7fu>

Reason: This dataset provides information about the number of Invest Nova Scotia applications received, declined, and approved. This information tells us about the new business opportunities in the Nova scotia.

Name:Nova\_Scotia\_Visitor\_Exit\_Survey\_\_Tourism\_Regions\_Visited\_\_2010\_\_2015\_and\_2017

URL: <https://data.novascotia.ca/Business-and-Industry/Nova-Scotia-Visitor-Exit-Survey-Tourism-Regions-Vi/e4rq-n5mz/data>

Reason: This dataset provides the purpose for the visit over the years. This helps us in identifying which performance indicator is main for Halifax.

Name: Register\_of\_Private\_Career\_Colleges

URL: <https://data.novascotia.ca/Education-Post-Secondary-and-Skills-Training/Register-of-Private-Career-Colleges/k3yz-94y3>

Reason: This dataset provides information about registered private career colleges and programs that can provide information on education opportunities.

Name: Discovering\_Opportunities\_\_DO\_\_Schools

URL: <https://data.novascotia.ca/Education-Primary-to-Grade-12/Discovering-Opportunities-DO-Schools/jr2d-k7va>

Reason: This dataset provides information about a list of public schools offering the Discovering Opportunities program. Includes school board, school, school address and implementation year.

Name: Grade\_9\_Enrolment\_\_2012-13\_\_and\_Graduating\_Students\_\_2015-16\_\_by\_Gender

URL: <https://data.novascotia.ca/Education-Primary-to-Grade-12/Grade-9-Enrolment-2012-13-and-Graduating-Students-/naes-ujpk>

Reason: This dataset provides information about the students who enrolled for grade 9 over a period by gender. This information helps in providing educational information.

Name: Grade\_9\_Enrolment\_\_2012-13\_\_and\_Graduating\_Students\_\_2015-16\_\_by\_Board

URL: <https://data.novascotia.ca/Education-Primary-to-Grade-12/Grade-9-Enrolment-2012-13-and-Graduating-Students-/e2ba-nibn>

Reason: This dataset provides information about Graduation Information arranged by the School Board; the percentage of the number of students receiving a High School Graduation Diploma compared to the number of students in grade 9 three years earlier.

Name: Graduation\_Information\_by\_Board

URL: <https://data.novascotia.ca/Education-Primary-to-Grade-12/Graduation-Information-by-Board/fmxh-aef7>

Reason: This dataset provides information about the students who graduated by the board. This information helps in providing educational information.

Name: Graduation\_Information\_by\_Gender

URL: <https://data.novascotia.ca/Education-Primary-to-Grade-12/Graduation-Information-by-Gender/jr6h-hrx6>

Reason: This dataset provides information about the students who graduated by gender. This information helps in providing educational information.

Name: Graduation\_Rates

URL: <https://data.novascotia.ca/Education-Primary-to-Grade-12/Graduation-Rates/fqau-nfyv>

Reason: This dataset provides information about the student's graduation rates over the year. This information helps in providing educational information.

Name:Nova\_Scotia\_Business\_Inc.\_Export\_Development\_and\_Investment\_Attraction\_Activity

URL: <https://data.novascotia.ca/Business-and-Industry/Nova-Scotia-Business-Inc-Export-Development-and-In/9rd7-7dm6>

Reason: This dataset provides information about new exporters and attracting innovative, globally competitive companies that establish a business location in Nova Scotia. This information helps in providing information on business opportunities.

Name: \_ARCHIVED\_\_Business\_Establishments\_2010-2011

URL: <https://data.novascotia.ca/Business-and-Industry/-ARCHIVED-Business-Establishments-2010-2011/wa8g-ji9a>

Reason: This dataset provides information about industries established in the period 2010-2011, which can provide information on business opportunities.

Name: Innovacorp\_Incubation\_Resident\_Companies\_2018-2019

URL: <https://data.novascotia.ca/Business-and-Industry/Innovacorp-Incubation-Resident-Companies-2018-2019/qhzg-5qkx/data>

Reason: This dataset provides information about the companies established by innovacorp Enterprise centre and their description. This information will help with business opportunities.

Name: NS\_schools

URL: <https://data.novascotia.ca/Education-Primary-to-Grade-12/Nova-Scotia-Map-Directory-of-Public-Schools/hcdu-5rtg>

Reason: This dataset provides information about all public schools in Nova Scotia, including School Board office contact information.

Name: Nova\_Scotia\_Communities\_Engaged\_in\_Age-Friendly\_Community\_Planning

URL: <https://data.novascotia.ca/Population-and-Demographics/Nova-Scotia-Communities-Engaged-in-Age-Friendly-Co/aasp-xptj>

Reason: This dataset provides information about the list of communities across Nova Scotia that are engaged in Age-Friendly Community planning.

This information will help in analyzing lifestyle.

Name: Gaelic\_language\_and\_cultural\_and\_supporting\_organizations\_in\_Nova\_Scotia

URL: <https://data.novascotia.ca/Arts-Culture-and-History/Gaelic-language-and-cultural-and-supporting-organi/dnv2-xrwf>

Reason: This dataset provides information about the organizations that listed are mandated or include some Gaelic language and cultural programming and/or support the same. This information will help in analyzing lifestyle

1. I didn’t use any programming languages and the tools used are Microsoft Word, Excel, notepad
2. 1) I have identified < halifax\_employment\_traded\_2017> , <halifax\_employment\_traded\_2016>, < halifax\_employment\_traded\_2015>, <halifax\_employment\_traded\_2014>, < halifax\_employment\_traded\_2013>, <halifax\_employment\_traded\_2012>, <halifax\_employment\_traded\_2011>,<halifax\_employment\_traded\_2010> useful.But all these data sets have common elements and, I merged all these 7 data sets and finally extracted a single dataset <halifax\_employment\_traded\_2010\_2017> and made a entity set “Business job reports”

2) I have identified <job\_type\_and\_education\_needed> useful. After exploring the data I found some fields are important and created two entity sets “experience\_jobs” and “education\_jobs” from the dataset.

3) I have identified <ns\_jobs> useful. After exploring the data I found all fields are important and created one entity set “ns\_jobs” from the dataset.

4) I found < Grade\_9\_Enrolment\_\_2012-13\_\_and\_Graduating\_Students\_\_2015-16\_\_by\_Board> and < Graduation\_Information\_by\_Board> shows same information but difference in years. So I added both and made one dataset from which “graduation information by board” entity set is made considering all fields.

5) I found < Grade\_9\_Enrolment\_\_2012-13\_\_and\_Graduating\_Students\_\_2015-16\_\_by\_Gender > and < Graduation\_Information\_by\_Gender> shows same information but difference in years. So I added both and made one dataset from which “graduation information by gender” entity set is made considering all fields.

6) I identified < Register\_of\_Private\_Career\_Colleges> useful and made “private colleges” entity set considering four attributes.

7) I identified <ns\_schools> useful and made “schools” entity set considering four attributes.

8) I found <graduation\_rates> useful. So considering all the attributes I made “graduation rates” entity set.

9) I identified < Education\_Survey> useful and made “Education survey” entity set considering four attributes.

10) I identified < DO\_organization> useful and made “Discovering opportunities program” entity set considering five attributes.

11)Identified “court lists” entity set from <court\_list> dataset considering all attributes

12) Identified “crimes” entity set from < crime\_stastics\_year > dataset considering all attributes

13) Identified “Crime count” entity set from < crime\_count\_adults\_years > dataset considering all attributes

14) ) I found <ns\_health\_male\_female > useful. So considering all the attributes I made “hospitals” entity set.

15) ) I found <crime\_safety\_home\_alone> useful. So considering all the attributes I made “crime surveys” entity set.

16)I found “community” entity set from <Nova\_Scotia\_Communities\_Engaged\_in\_Age-Friendly\_Community\_Planning> dataset considering four important attributes.

17) “Gaelic Organization” is an important entity set found from <Gaelic\_language\_and\_cultural\_and\_supporting\_organizations\_in\_Nova\_Scotia> dataset with 3 attributes

18) ) Identified “native language” entity set from < Native\_Language\_of\_Nominees\_-\_Total\_Number\_by\_Language> dataset considering all attributes

19) ) I found <Labour\_Market\_Classification\_of\_Nominees > useful. So considering all the attributes I made “labour market” entity set.

20) “visitor survey” is an important entity set found from <Nova\_Scotia\_Visitor\_Exit\_Survey\_\_Tourism\_Regions\_Visited\_\_2010\_\_2015\_and\_2017> dataset with 3 attributes

21) “Trip survey” is an important entity set found from <trip\_purpose\_exit\_survey> dataset with 3 attributes

22) some disjoint datasets are derived from the < job\_type\_and\_education\_needed > data set considering some useful attributes

|  |  |
| --- | --- |
| **Strong entities** | **Weak entities** |
| Educational Institutions | DO schools |
| Education survey | Business job reports |
| Graduation rate | Crime count |
| Private colleges | Trip purpose |
| Schools |  |
| Graduation information by board |  |
| Graduation information by gender |  |
| Business Development Program |  |
| Business Establishments |  |
| Jobs |  |
| Recruiting strategies |  |
| Work type |  |
| Position type |  |
| Education type |  |
| experience |  |
| Duration |  |
| Court lists |  |
| Crimes |  |
| Crime surveys |  |
| Health |  |
| communities |  |
| Art program |  |
| Gaelic organizations |  |
| Native languages |  |
| Labour market |  |
| Visitors survey |  |

**Section-B**

1. I have identified the relationship between the entity sets with respect to the real-world scenario and finding KPI in the Halifax region. I found the cardinality between the entity sets. I left some entity sets for which I cant find the relationship with my data. I have taken the nova scotia data so that Halifax can be compared to nova scotia.
2. The pictures below are the initial sketch of the business, safety, education

, and lifestyle respectively.

A picture containing text, whiteboard

Description automatically generated

**A picture containing text, whiteboard

Description automatically generated**

A picture containing text

Description automatically generated

A picture containing text, whiteboard

Description automatically generated

1. I have identified certain disjoint/overlap subtypes in the above ERD diagrams. Considering all these subtypes I have drawn an extended ERD using draw.io tool[1]
2. Below are the extended ERD diagrams for the business, safety, education, lifestyle respectively.

A picture containing text, map

Description automatically generated

A close up of a map

Description automatically generated

A picture containing text, map

Description automatically generatedA picture containing text, map

Description automatically generated

1. I have validated my Extended ERD against the fan trap and chasm trap problem but I didn’t find any anti-patterns(1 to n and nto1) that result in fan trap problem. There are no issues between the entity and its attributes. I had a ternary relationship in the business ERD but it was eliminated in the Extended ERD by using disjoint subtypes. As far as my knowledge I don’t find any issues in the Extended ERD.

**Section- C**

I have populated tables[3] using the data from the selected datasets. I used MySQL Workbench to create my database and executed the following questions.

1. Which business cluster provided more jobs in the year 2017?

Query: select Cluster from business\_development\_report order by year\_2017 desc limit 1;

Result:

A screenshot of a cell phone

Description automatically generated

2. Which are in Halifax has more schools?

Query: select CPOSTCODE,CIVIC1,CIVIC2,the\_geom,count(SCHLNAME) from schools group by CPOSTCODE order by count(SCHLNAME) desc limit 1;

Result:

A screenshot of a social media post

Description automatically generated

3. Which street in Halifax has more number of reported crimes?

Query: select GEOGRAPHY from crime\_stastics\_year order by TOTAL\_CRIME\_SEVERITY\_INDEX desc limit 1;

Result:

A screenshot of a cell phone

Description automatically generated

All these queries are saved in an SQL file named”**script.sql**” and provided it for the reference.

**Section-D**

I have identified functional dependencies between the attributes in the entity set and found out there is a scope for normalization in the data. While importing the data sets to my SQL there is a need to clear data. I have cleaned many entity sets eliminating 1nf, 2nf, and 3nf as far as possible. I have created new entity sets with their names ending with new for easy identification. For example “communities” is the name of the entity set before normalization. “communities\_new” is the name of the same entity set after normalization.

I have created a dump file containing all the entity sets before normalization and named it “Dumpfile\_beforenormalization”. I have created a dump file containing all the entity sets after normalization and named it ”Dumpfile\_afternormalization”.

**References:**

[1]draw.io - free flowchart maker and diagrams online, “Flowchart Maker & Online Diagram Software,” Draw.io, 2020. [Online]. Available: [https://www.draw.io/.](%20https:/www.draw.io/) [Accessed: 10-Feb-2020].

[2]“Nova Scotia Government - Open Data Portal | Open Data | Nova Scotia,” Socrata, 2018. [Online]. Available: <https://data.novascotia.ca/>. [Accessed: 8-Feb-2020]

[3]W3schools.com-THE WORLD'S LARGEST WEB DEVELOPER SITE,” SQL FOREIGN KEY Constraint”W3schools.com, 2020. [Online]. Available: <https://www.w3schools.com/sql/>. [Accessed: 10-Feb-2020].

[4]LibGuides: Data and Statistics: Nova Scotia Data,” Libguides.com, 2013. [Online]. Available: <https://dal.ca.libguides.com/data/novascotia>. [Accessed: 8-Feb-2020].

[5]“Halifax Regional Municipality,” Arcgis.com, 2017. [Online]. Available: <http://catalogue-hrm.opendata.arcgis.com/>. [Accessed: 8-Feb-2020]