Shivam Gupta
B00810723
Assignment - 1

DATA WAREHOUSE MANAGEMENT

ASSIGNMENT 1

A.Feasibility Analysis

1. Yes. I have enough datasets to identify KPIs to improve business, safety, education, lifestyle.
2. I have found these datasets:

Business Domain:

- Census_long_form_labour_by_industry
  - https://data.novascotia.ca/Employment-and-Labour/Census-Long-Form-Labour-by-Industry-1991-and-1996/25vn-wikd
- Census_long_form_labour_by_occupation
  - https://data.novascotia.ca/Employment-and-Labour/Census-Long-Form-Labour-Occupation-1996-2006/sgix-yfew
- Nova Scotia Co-Operatives
  - https://data.novascotia.ca/browse?q=Nova%20Scotia%20Co-operatives&sortBy=relevance
- Capital Projects
  - http://catalogue-hrm.opendata.arcgis.com/datasets/capital-projects
- Business_Establishments_2010_2011
  - https://data.novascotia.ca/Business-and-Industry/Business-Establishments-2010-2011/wa8g-ji9a
- Aquaculture_Production_2006-2011
- https://data.novascotia.ca/browse?q=Aquaculture%20Production%202006-2011&sortBy=relevance
- Agriculture_Funding_Program
  - https://data.novascotia.ca/Agriculture-and-Agri-business/Agriculture-Funding-Programs-Details/jv92-pedy

Education Domain:

- Nova-Scotia-Public-School-Contact-Information
  - https://data.novascotia.ca/Education-Primary-to-Grade-12/Nova-Scotia-Public-School-Contact-Information/iyap-ttn5

Lifestyle:

- HRM_PARKS
  - http://catalogue-hrm.opendata.arcgis.com/datasets/hrm-parks

Safety:

- Crime
  - http://catalogue-hrm.opendata.arcgis.com/datasets/crime

Shivam Gupta
B00810723
Assignment - 1

3. I have used MS Excel to gather the data and to select the fields.

4. I have found many entities from the datasets I gathered:

DATASETS	TABLES OBTAINED
Census_long_form_labour_by_industry	Labour_by_Occupation, AllCanadaLocation
Census_long_form_labour_by_occupation	Labour_by_Idustry, AllCanadaLocation
Nova Scotia Co-Operatives	Nova_Scotia_Co-operatives
Capital Projects	Capital_Projects, Halifax_location, ProjectsInfo
Business_Establishments_2010_2011	Business_Establishment, Canada_Location
Aquaculture_Production_2006-2011	Aquaculture, Canada_Location
Agriculture_Funding_Program	Agriculture_Funding_Program, Canada_Location
Nova-Scotia-Public-School-Contact-Information	School, Board_School, Area
HRM_PARKS	Parks
Crime	Crime, CrimeTypes

NOTE: Cleaning of data is done by omitting non-relevant fields.

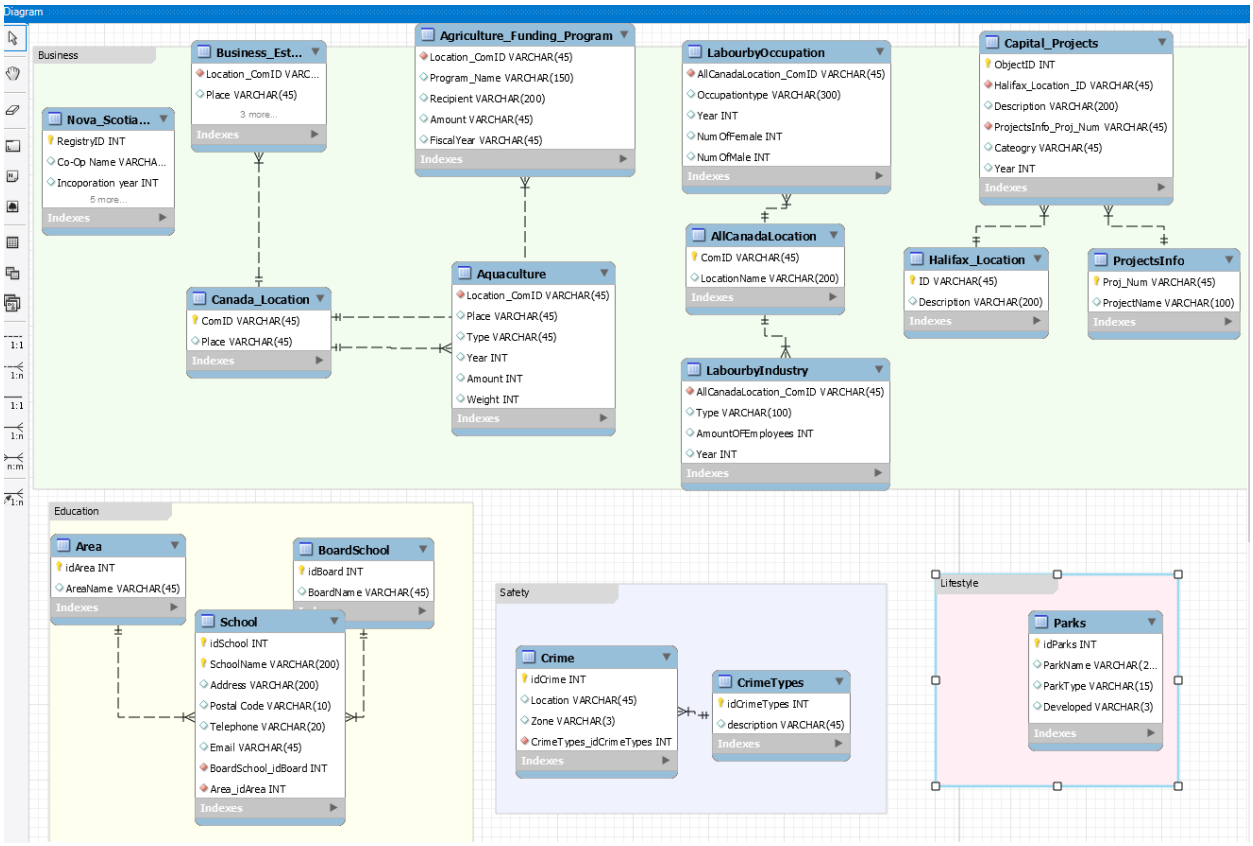
5. Strong and Weak Entity Sets:

Strong Entity Sets	Weak Entity Sets
Labour_by_Occupation, AllCanadaLocation	
Labour_by_Idustry, AllCanadaLocation	
Nova_Scotia_Co-operatives	
Capital_Projects, Halifax_location	ProjectsInfo
Business_Establishment, Canada_Location	
Aquaculture, Canada_Location	
Agriculture_Funding_Program, Canada_Location	
School, Area	School_board
Parks	
Crime, CrimeTypes	

Shivam Gupta
B00810723
Assignment - 1

B. Data Modelling:

Initial Design and final design: Free from Design issues. Please refer the attached ERD .mwb file for referring to schema.



C. DDL and DML

Created ERD and used forward engineering to create database and import wizard for getting data into tables.

Notes:

- *Please refer the assignment1.sql for the script of the schema created.*
- *Please refer all the scripts and tables to populate data.*

• Which business organization or type of business organization has highest employees?

```
SELECT TypeOFBusiness, sum (TotalEmployees) employees FROM mydb.business_establishment
where Location_comid='cnt1209' and TypeOfBusiness <> 'All Industries'
group by typeofbusiness order by employees desc limit 1;
```

- Which area in Halifax region has more schools?

```
select AreaName from (SELECT  AreaName, count(*) numberschool FROM mydb.school, mydb.area where area.idArea=school.Area_idArea and AreaName in ('Halifax NS','Dartmouth NS','Bedford NS', 'barrington NS') group by AreaName) as table1 where numberschool = (select max(numberschool) from (SELECT  AreaName, count(*) numberschool FROM mydb.school, mydb.area where area.idArea=school.Area_idArea and AreaName in ('Halifax NS','Dartmouth NS','Bedford NS', 'barrington NS') group by AreaName) as table2);
```

- Which street in Halifax region has a greater number of reported crimes?

```
SELECT location from (SELECT location, count (*) numcrime FROM mydb.crime group by location) as table1 where numcrime = (select max(numcrime) from (SELECT location, count (*) numcrime FROM mydb.crime group by location) as table2);
```

D. Normalization

All the tables are in 2NF form and are normalised.