

1 Feasibility Analysis

The startup HalifaxInfo needs large amount of data relevant to identify key performance indicators in business, lifestyle, safety and education in Halifax region. After going through the links, there was some data which is found valid to the scenario along with some unorganized data. I found the entities and tried to organize a structured form of data from those data sets. Here is my analysis about each KPI:

1.1 Education

There are many data sets regarding education in given web links. The basic entities which needs to be considered while developing a KPI for education is schools, education boards, medium of teaching, grad school details, school facilities etc. [1]. After exploring all the sources data sets, the conclusion is that there is ample amount of data regarding education to develop a KPI. However, the data regarding programs offered by various schools and child education is not there in source data sets which is considerable aspect while developing KPI.

1.2 Lifestyle

After exploring all the data sources, the conclusion is that there is ample amount of valid data to develop the lifestyle KPI. There data sets have relevant data about the entities like housing, transit details, food establishments, hotels, shopping places and public park details [2]. However, the data related to tourism places, bars and pubs, movie theatres is not there which could have enhanced the quality of lifestyle KPI.

1.3 Safety

The data sources from the given web links has limited amount of data regarding the safety KPI [3]. The data regarding hospitals, residential care facilities, ambulances and crime area details are there in sources. The information about police stations, Accident prone areas and disease-based hospital details would have been more useful to develop safety KPI.

1.4 Business

After exploring all the data sources through given links, the data sets regarding business are very limited and these data sets are not enough to develop a good business KPI [1]. However, the details like business establishments, export details, business statistics and business registration office details are available in given data sets which will help a lot in developing business KPI.

2 List of Data Sets

2.1 Education

2.1.1 Nova Scotia Public School Contact Information

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

Reason: This data set gives the details of all public schools and board details in nova scotia region. This will be the key source of information to develop the education KPI.

2.1.2 Schools Offering the Pre-primary Program

URL: <https://data.novascotia.ca/Education-Early-Childhood/Schools-Offering-the-Pre-primary-Program/2k3n-wed5/data>

Reason: This data set will help us with the details of schools which offers pre-primary programs in region of Nova Scotia.

2.1.3 Schools with French Programs

URL: <https://data.novascotia.ca/Education-Primary-to-Grade-12/Schools-with-French-Programs/kdgs-3rvd>

Reason: This data set gives us information about all the schools which offer French program in nova scotia region.

2.1.4 Graduation Information by Board

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

Reason: This data set helps with the information of various graduation boards across nova scotia.

2.1.5 Gaelic Courses Taught by Regional Centre for Education, School and Grade

URL: <https://data.novascotia.ca/Education-Primary-to-Grade-12/Gaelic-Courses-Taught-by-Regional-Centre-for-Educa/s2fr-akki>

Reason: This data set helps with the data about schools which offer Gaelic course across nova scotia.

2.1.6 Early Childhood Education Training Institutions

URL: <https://data.novascotia.ca/Education-Early-Childhood/Early-Childhood-Education-Training-Institutions/5zyd-q83j>

Reason: This data set helps with details of all the institutions which offer early childhood training institutions in Halifax region.

2.2 Safety

2.2.1 Hospitals

URL: <https://data.novascotia.ca/Health-and-Wellness/Hospitals/tmfr-3h8a>

Reason: This data set helps us with information regarding all the hospitals and its details like contact numbers, address and location coordinates.

2.2.2 Long Term Care and Residential Care Facilities

URL: <https://data.novascotia.ca/Health-and-Wellness/Long-Term-Care-and-Residential-Care-Facilities/x76a-axw2>

Reason: This data set helps the safety KPI with the details of long-term care and residential facilities offered across nova scotia.

2.2.3 Justice Centers/Courts

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

Reason: This data set helps with the data of justice centers and courts across nova scotia region.

2.2.4 EHS Ambulance Fees in Nova Scotia

URL: <https://data.novascotia.ca/Health-and-Wellness/EHS-Ambulance-Fees-in-Nova-Scotia/dc9f-gb4a>

Reason: This data set helps safety KPI with the information regarding the ambulance fee and Transportation category.

2.2.5 Crime Statistics - Incidents and rates for selected offences

URL: <https://data.novascotia.ca/Crime-and-Justice/Crime-Statistics-Incidents-and-rates-for-selected-/m862-kmjj>

Reason: This data set helps with information of crime incidents and violations happened in nova scotia region every year.

2.2.6 Crime

URL: http://catalogue-hrm.opendata.arcgis.com/datasets/f6921c5b12e64d17b5cd173cafb23677_0/data

Reason: This data set helps with details of all crimes happened in every street of Halifax area.

2.3 Lifestyle

2.3.1 List of Licensed Food Establishments

URL: <https://data.novascotia.ca/Permits-and-Licensing/List-of-Licensed-Food-Establishments/2shh-dv8p>

Reason: This data set help lifestyle KPI with all the details of licensed food establishments across nova scotia.

2.3.2 Select Nova Scotia

URL: <https://data.novascotia.ca/Agriculture-and-Agri-business/Select-Nova-Scotia/t8ij-8svj>

Reason: This data set is developed by the Province of Nova Scotia to encourage Nova Scotians to shop and buy locally. This data set also contains data about Nova Scotia agri-food and seafood products, and other locally made goods.

2.3.3 Bus Routes

URL: http://catalogue-hrm.opendata.arcgis.com/datasets/e3b2bfdd61154176822c00602504c950_0/data?selectedAttribute=ROUTE_NUM

Reason: This data set helps us with bus routes in Halifax region which is one of key entity in developing lifestyle KPI.

2.3.4 Bus Stops

URL: http://catalogue-hrm.opendata.arcgis.com/datasets/29de9d04a3454e11a1e0a1f78a27bc07_0/data

Reason: This data set provides the information about bus stops in Halifax region which help bus route entity in lifestyle KPI.

2.3.5 Public Housing Units - Nova Scotia Families

URL: <https://data.novascotia.ca/Community-Services/Public-Housing-Units-Nova-Scotia-Families/nxzm-xxps>

Reason: This data set helps with the details of all public housing details across nova scotia which is helps to improve housing entity.

2.3.6 Tourism Nova Scotia Listed Operators

URL: <https://data.novascotia.ca/Arts-Culture-and-History/Tourism-Nova-Scotia-Listed-Operators/2h2s-6bg4>

Reason: This data set helps with the information regarding tourism nova scotia listed operators which helps tourism entity in lifestyle KPI.

2.3.7 Ferry Routes

URL: http://catalogue-hrm.opendata.arcgis.com/datasets/1bac39377d2643dfa5c68e694151a312_0/data?geometry=-63.626%2C44.645%2C-63.496%2C44.666

Reason: This data set helps with the information regarding ferry routes in Halifax region.

2.3.8 HRM Parks

URL: http://catalogue-hrm.opendata.arcgis.com/datasets/3df29a3d088a42d890f11d027ea1c0be_0/data

Reason: This dataset gives us the information about all the park details in Halifax.

2.3.9 Civic Addresses

URL: http://catalogue-hrm.opendata.arcgis.com/datasets/2bc8323870fe44eab50630404713be6a_0

Reason: This data set gives us the information about all the geographical points around Halifax area.

2.3.10 Street Name Routes

URL: http://catalogue-hrm.opendata.arcgis.com/datasets/6d93ecf9e2a348b891e1d5f21c436fe1_0/data

Reason: This data set give us the street details across Halifax regional municipality.

2.4 Business

2.4.1 Nova Scotia Co-operatives

URL: <https://data.novascotia.ca/Business-and-Industry/Nova-Scotia-Co-operatives/k29k-n2db/data>

Reason: This data set gives the information regarding all the Co-operatives in Halifax region.

2.4.2 Business Establishments 2010-2011

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

Reason: This data set gives the information about all the business establishments in Halifax in the years of 2010 and 2011.

2.4.3 Business Establishments 2007

URL: <https://data.novascotia.ca/Population-and-Demographics/-ARCHIVED-Business-Establishments-2007/9mdt-jt2i/data>

Reason: This data set gives the information about all the business establishments happened in Halifax in the year of 2007.

2.4.4 Business Registration Unit Office Locations and Hours

URL: <https://data.novascotia.ca/Business-and-Industry/Business-Registration-Unit-Office-Locations-and-Ho/n48k-avit/data>

Reason: This data set gives the information about all the business registration offices in the region of Halifax.

2.4.5 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 data set gives the information about all the investments and export details across Halifax regional board.

3 programming languages and tools used

3.1 Microsoft Excel

All the data is extracted from given web links in CSV format which is supported by excel. The table modification and columns deletions are performed in excel. Data refactoring is also done in excel by deleting the empty rows, removing redundant data.

4 Finding entity sets from data sets

4.1 Education KPI

- The data set 'Nova Scotia Public School Contact Information' consists of 14 columns where only 10 columns has been considered to get the entity set of all school details in Halifax.
- A new entity set 'Board lists' is also created from the data set 'Nova Scotia Public School Contact Information' with board numbers and board details.
- The data set 'Schools Offering the Pre-primary Program' is used to create a weak entity set 'Schools with pre-primary program' by deleting few common rows from public school contact data set.
- The data set 'Schools with French Programs' is used to create a weak entity set 'French program schools' by deleting few common rows from public school contact data set.
- The data set 'Graduation Information by Board' is used to create a weak entity set of 'Graduate board details' by deleting few common columns from public school contact data set.
- The data set 'Gaelic Courses Taught by Regional Centre for Education, School and Grade' is used to create a weak entity set 'Gaelic Course Schools' by deleting few common from public school contact data set.
- The data set 'Early Childhood Education Training Institutions' is created to get a new entity set which contains all details of childhood education institutions.

4.2 Safety KPI

- The data set 'Hospitals' is used to create a strong entity set 'Hospital List' by deleting 3 columns.
- The data set 'Long Term Care and Residential Care Facilities' is used to create a string entity set 'Residential care facilities' by deleting 3 columns.
- The data set 'Justice Centers/Courts' is used to create a strong entity set 'Justice centers' by deleting 2 columns.
- The data set 'EHS Ambulance Fees in Nova Scotia' is used to create a strong data set 'Ambulance Services' without deleting any columns.
- The data set 'Crime Statistics - Incidents and rates for selected offences' is used to create a weak entity set 'Crime Statistics' and another table 'crime type details' with crime type description details.
- The data set 'Crime' is used to create a strong entity set 'Crime Details' which has details like crime id, crime type id, occurrence place etc. by deleting 3 columns.

4.3 Lifestyle KPI

- The data set 'List of Licensed Food Establishments' is used to create a string entity set 'food establishments' by adding a surrogate key in it.
- The data set 'Select Nova Scotia' is used to create a new 'store details' by adding surrogate key in.
- The data set 'Bus Routes' is used to create a strong entity set 'Bus Route Details' by deleting 4 columns.
- The data set 'Bus Stops' is used to create a strong entity set 'Bus Stop Details' by deleting 6 columns.

- The data set 'Public Housing Units - Nova Scotia Families' is used to create a strong entity set 'Public House Units' by deleting 3 columns.
- The data set 'Tourism Nova Scotia Listed Operators' is used to create a strong entity set 'Tourist operators' by deleting 2 columns and adding a surrogate key.
- The data set 'Ferry Routes' is used to create a strong entity set 'Ferry Details' without deleting any columns.
- The data set 'HRM Parks' is used to create a strong entity set 'Park Details' by deleting 9 columns and linking to 'civic details' entity set.
- The data set 'Civic Addresses' is used to create a strong entity set 'civic details' by deleting 6 columns and linking to entity set 'Street Details'.
- The data set 'Street Name Routes' is used to create a strong entity set 'Street Details' by deleting 7 columns and having that as foreign key link to civic details entity set.

4.4 Business KPI

- The data set 'Nova Scotia Co-operatives' is used to create strong entity set 'Halifax Co-operatives' by deleting 4 columns.
- The data set 'Business Establishments 2010-2011' and 'Business Establishments 2007' is merged to create new Strong entity set 'Halifax business establishments'.
- A new entity set 'Region details' is created from the data in business establishments entity set.
- A new entity set 'business sector' is created from entity sets of 'Halifax co-operatives' and 'business establishments' which contains all the sector details.
- The data set 'Business Registration Unit Office Locations and Hours' is used to create strong entity set 'business registration office details' by adding a surrogate key.
- The data set 'Nova Scotia Business Inc. Export Development and Investment Attraction Activity' is used to create a strong entity set 'Halifax business export activities' by adding a surrogate key.

5 List of strong and weak entities

5.1 Education KPI

Strong Entity Sets: Schools, Boards.

Weak Entity Sets: French Schools, Pre-Primary Schools, graduation boards and Gaelic course schools.

5.2 Safety KPI

Strong Entity Sets: Hospitals and justice centers, crime description details, crime activity details, Region details, residential care details and ambulance facilities.

Weak Entity Sets: crime statistics.

5.3 Lifestyle KPI

Strong Entity Sets: Housing, buses stops, bus routes, hotels, food establishments, store details, street details, civic details and ferry details.

Weak Entity Sets: There are no weak entity set found in the design.

5.4 Business KPI

Strong Entity Sets: Halifax business export activities, business sector detail, business region details, Halifax Cooperatives.

Weak Entity Sets: Halifax business establishments and business registration office details.

2 Data Modelling

- Paper rough design for all KPI's has been attached in submission folder.
- Extended ERD (2nd design) has been drawn in a tool called 'Visual Paradigm' and all those files has been attached in submission folder.
- I don't find any design issues from the 2nd design, hence there is no 3rd ERD.

3 DDL and DML

- As per question, I have created separate databases for both un-normalized and normalized version.
- All the DDL for those databases has been uploaded in submission folder.

3.1 DML for given questions

1) Which business organization or type of business organization has highest employees?

- `select b.Industry_Type,sum(a.Total) as Employee_Count from
halifaxinfo_normalised.halifax_business_establishments as a inner join
halifaxinfo_normalised.business_sector_details as b
on a.Industry_ID=b.Industry_ID group by b.Industry_Type order by
Employee_Count DESC limit 1;`

2) Which area in Halifax region has more schools?

- `select Postal_Code,count(School_Code) as Number_of_Schools from
halifaxinfo_normalised.halifax_school_info
group by Postal_Code limit 1;`

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

- `select b.Region_Name,sum(a.INCIDENTS) as incident_count from
halifaxinfo_normalised.crime_statistics as a inner join
halifaxinfo_normalised.region_details as b on
a.Region_Id=b.Region_Id group by b.Region_Name order by
incident_count DESC limit 1 ;`

4 Normalization

As per question, all the tables have been normalized and separate databases has been created. The scripts of those two databases has also been attached in submission folder.

5 References

[1] Online Source for some of the data which is in the tables <http://catalogue-hrm.opendata.arcgis.com/>

[2] Online Source for some of the data which is in the table <https://dal.ca.libguides.com/data/novascotia>

[3] Online Source for some of the data which is in the tables <https://data.novascotia.ca/>