

A. FEASIBILITY ANALYSIS:

1. The key performance indicators for business, lifestyle, education and safety are those parameters which will judge the improvement in all these sectors. It means, these KPIs will be responsible for enhancing these sectors. So, according to my findings and research, the datasets which we have been provided may be not be sufficient to connect all four sectors, however we can find relations between 2 sectors, like between business and lifestyle on one hand and education and safety on another hand. This has been shown in detail in this report, along-with the KPIs for them.
2. The datasets which I found useful for finding the KPIs for the sectors are given below, along-with their URLs and reason for their selection. The reason will be very brief as asked in the question to provide the response in one sentence.
 - [Crime statistics data](#) : This data is chosen for viewing the crime severity index in a particular region, for a particular year, which helps to determine the crime rate in a particular area, so that safety can be measured in the area.
 - [Graduation rate Data](#): This data set has the names of a board in a particular region along with the graduation rate of students in that region for students graduating in the year 2013 and 2014, this is helpful to compare the graduation rates of different boards in different years.
 - [Number of tourist visitors in Nova scotia](#): It gives information about number of visitors to Nova Scotia, their country of origin, and the year in which they visited, which is very helpful to compare the number of visitors in each year, to know the trend in the number of visitors over the years, to see the business of tourism industry.
 - [Hotel occupancy rates data](#): This dataset has the occupancy rates of the hotels all across Nova Scotia, which is helpful to see the trend of people staying in the hotels from 2006 to 2019, which is a part of city's lifestyle.
3. The tool which was used for collecting the datasets was Microsoft excel. The dataset was downloaded in CSV format from the URLs provided above. In excel, pivot-table was created to calculate the average values of occupancy rates of hotel in a particular year as shown below. This average is taken from the monthly data of each and every year.

Row Labels	Average of Occupancy_rate
2007	0.40125
2008	0.400416667
2009	0.381666667
2010	0.385208333
2011	0.3890625
2012	0.386458333
2013	0.3784375
2014	0.398125
2015	0.415625
2016	0.433229167
2017	0.461979167
2018	0.462291667
2019	0.45525
(blank)	
Grand Total	0.410892857

FIGURE 1

Row Labels	Average of NUM_OF_VISITORS
2006	5509.114583
2007	5570.3125
2008	5414.0625
2009	5445.572917
2010	5089.84375
2011	5085.9375
2012	5184.114583
2013	4925.260417
2014	5101.302083
2015	5387.5
2016	5826.822917
2017	6384.776903
2018	7527.1875
2019	7572.058824
Grand Total	5653.299981

FIGURE 2

Similarly, the average values of total number of visitors to the city was also calculated using pivot table. The formatting of the columns and rows was done using excel. Moreover, un-necessary rows/columns has been removed from the datasets and only those columns are kept which are utilized for analysis of the KPIs. Snipping tool was used for placing the above figures after snipping from the dataset excel sheet.

4. Identifying the Entities and their initial attributes from the selected datasets:

a)- The entity for the number of tourist visitors in nova scotia data is “tourist visitors” and the entity for hotel occupancy data is “tourist accommodations”. The **initial attributes** for the number of tourist visitors data(tourist visitors entity) in nova scotia were mode of entry, month/year, visitor origin, country and number of visitors. Similarly the **initial attributes** for hotel accommodation data(tourist accommodation entity) were Region, Month/year, Occupancy rate, and room nights sold.

b)- The entity for crime statistics data is “Geographical crime statistics” and the entity for graduation rate data is “School board”. The **initial attributes** for Geographical crime statistics were Geography, year, total crime severity index, violent crime severity index, non-violent crime severity index, youth crime severity index, and youth violent crime severity index. The **initial attributes** for School board entity were board name, grade 9 enrolment year, number of graduating students, grade 9 enrolment, and graduation rate.

Filtering the Attributes of the Entities:

The initial attributes which were selected for all the four entities needed to be filtered, since in the analysis, they were not required. So the attributes left for the entities after removing the unwanted ones are given below entity-wise:

Tourist visitors- year, country and number of visitors

Tourist accommodations- year, region, Occupancy rate

Geographical Crime Statistics- year, region, total crime severity index

School board- school board name, graduation year, number of graduating students, grade 9 enrolment and graduation rate.

5. The entities which are present in this analysis consists of weak as well as strong entities. For example tourist visitors is not dependent on any other entity in the system, that tourists/visitors will continue to visit Nova Scotia and they do not depend upon any entity which are present in my system. Tourist accommodations entity is for the visitors/tourists who will come to visit Nova Scotia, and it is partially dependent on tourist visitors, since the tourists who come will need some place to stay. Geographical crime statistics is a weak intangible entity, that is partially dependent on School boards, since school boards are located in a particular region which will have its own crime severity index, that will constitute to a critical safety factor of the region.

School board is an independent entity, that is school board does not depend upon whether the crime is higher or lower in that area, the school boards existence depends upon many factors.

B. DATA MODELLING:

Initial design:

- 1.)- The entities tourist visitors and tourist accommodations are related to one another. They are related in many to many relationship, that is one tourist can go to many different accommodations on different days and one accommodation will have more than one tourist.
- 2.)- The entities geographical crime statistics and school board are related in one to one relationship, since crime rate of that area in which the school is present will be unique. Similarly, the graduation rate of a particular area is higher in a period of time, where the crime severity index is lower for those particular years. This is displayed in data, which will be shown later in this report.
- 3.)- The first rough paper ER diagram is shown below, according to the data sets available. This ER diagram consists of **filtered attributes** of the entities. After removing those initial attributes, the first rough ER diagrams for 2 different relationships is shown below.

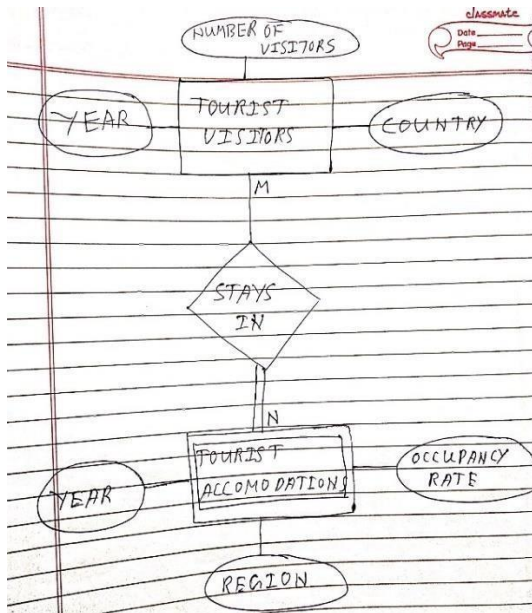


FIGURE 3

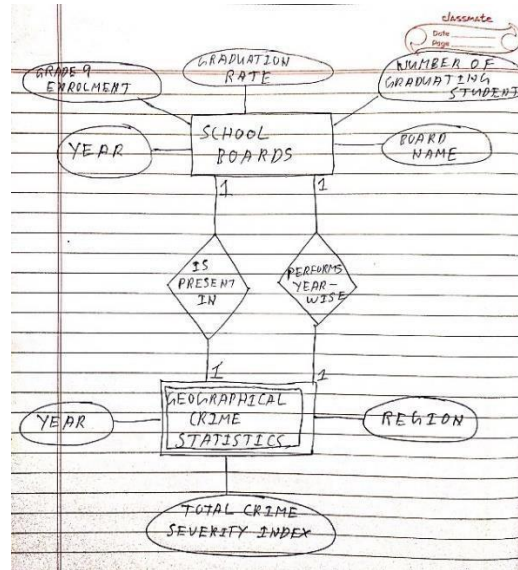


FIGURE 4

The ER diagram in figure 3 shows relationship between tourist visitors and tourist accommodation entities.

CARDINALITY: Since one tourist can go to multiple hotels and similarly one hotel can have multiple tourists, so we have M:N relationship.

PARTICIPATION: There is total participation from tourist accommodations entity, since each accommodation will have some tourists, no accommodation will be completely vacant, so total participation.

The ER diagram in figure 3 shows relationship between School board and logical intangible entity Geographical crime statistics.

CARDINALITY: Since each school board will be located in a particular geography and that school board will not be present in some other region. For example Halifax metropolitan school board will be in Halifax and cape Breton school board will be in Cape Breton, so the crime statistics will be unique for both of these areas. Similarly, crime statistics will be of particular region in particular year, for a particular school board. School board has graduate rate, which is unique for each region.

PARTICIPATION: Each school board does not has the crime severity index, since data is not available for all the school boards, so this is partial participation.

FINAL DESIGN: This is the final ER diagram for Tourist and accommodation entity. The explanation is given below in detail.

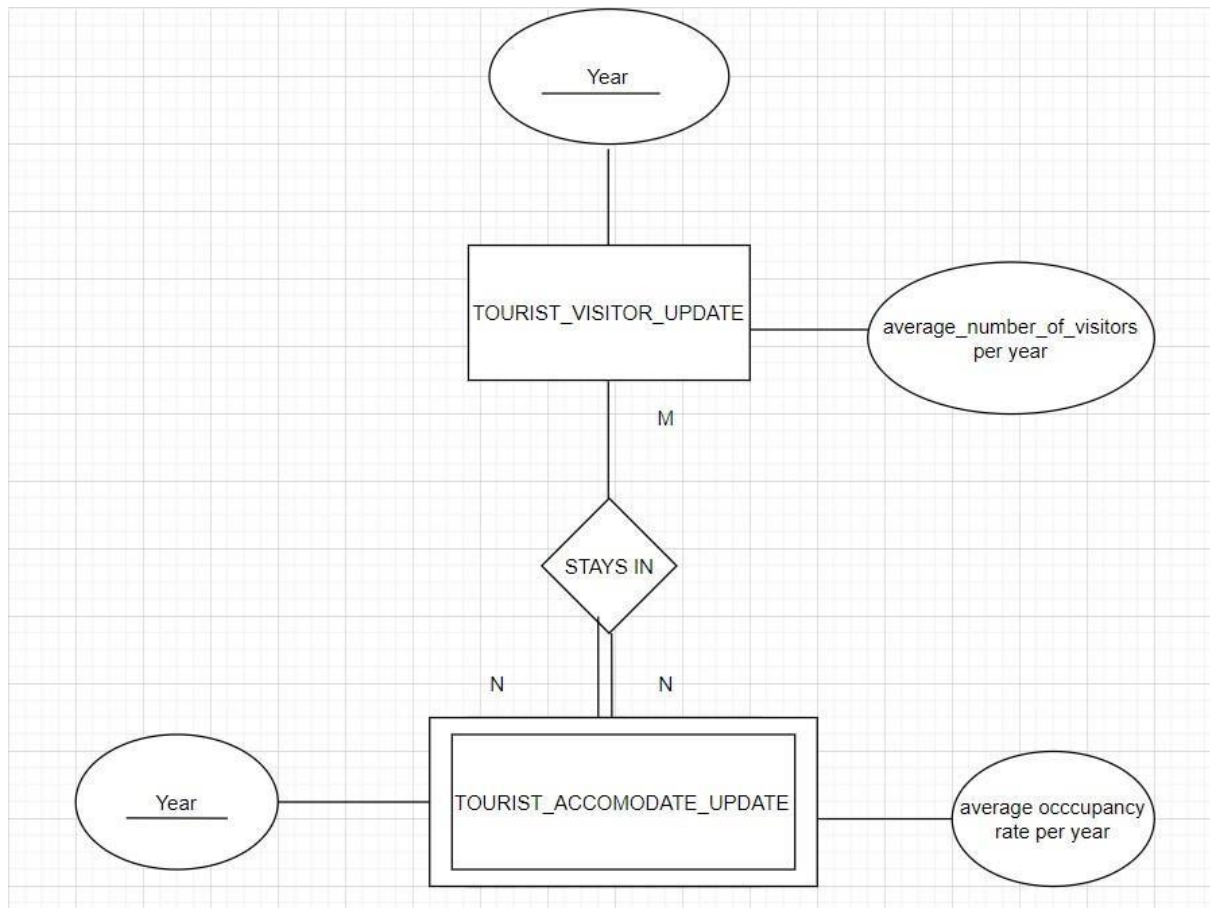


FIGURE 5

Initially the dataset had many attributes, but in the final design it can be observed that only 2 attributes are left in the design. The reason for that is , for this relationship, only these attributes were required. After filtering the data set of number of tourist visitors in Nova Scotia and hotel occupancy rates data, it was observed that in a particular year, when the average number of tourist visitors increased , at the same time the average occupancy rates of the hotels also increased for that year. Similarly, after a certain period of time, the average number of tourist visitors decreased for some years, which was again impacting the occupancy rates of the hotels, since they also decreased for those years. It means that there is a relationship between tourist visitors and accommodations, as displayed by data. Have a look at this trend, which was cut from the excel sheet using snipping tool:

Row Labels	Average of NUM_OF_VISITORS
2006	5509.114583
2007	5570.3125
2008	5414.0625
2009	5445.572917
2010	5089.84375
2011	5085.9375
2012	5184.114583
2013	4925.260417
2014	5101.302083
2015	5387.5
2016	5826.822917
2017	6384.776903
2018	7527.1875
2019	7572.058824
Grand Total	5653.299981

FIGURE 6

Row Labels	Average of Occupancy_rate
2007	0.40125
2008	0.400416667
2009	0.381666667
2010	0.385208333
2011	0.3890625
2012	0.386458333
2013	0.3784375
2014	0.398125
2015	0.415625
2016	0.433229167
2017	0.461979167
2018	0.462291667
2019	0.45525
(blank)	
Grand Total	0.410892857

FIGURE 7

Here in the above two pivot tables created from excel sheet, the average of number of visitors in all the months, gave the average number of visitors in that particular year. Similarly, from the dataset, the average of occupancy rates of all the months was taken to create the average occupancy rate of that particular year. In this way, from 2006 to 2019, the average number of visitors per year is seen and the average occupancy rates per year in the accommodations from 2007 to 2019 is seen.

TREND, OBSERVATIONS and INFERENCE: We can observe from the above data that from 2006 to 2013, the average number of visitors per year was decreasing or was approximately constant, but from 2014 the average number of visitors per year increased till 2019. Now the visitors who come to the country, have to find an accommodation to live. So the average occupancy rate should also follow the similar trend. When the data was observed, it can be seen that the average occupancy rate has also increased from 2014 till 2019 approximately. So by the data it can be deduced, that the visitors stayed in the accommodations. So the KPI for this relation is average number of visitor for the tourism industry, that is more the number of tourists that visit the country, more better chances the tourism industry will get to invest. This KPI will help in improving the tourism Business sector. Similarly, the KPI Occupancy rate is responsible for knowing the profit made by the hotel, that is better the occupancy rate, more will be the profit, and better facilities will be provided by the hotels to the tourists, so the lifestyle will be impacted. This KPI can be helpful for improving the lifestyle sector.

The final ER diagram for School board and Geographical crime statistics entity is shown below.



In this diagram, the entities have only those attributes which are required for further analysis. The dataset which was used initially had total crime severity index of many regions, but the data set of school boards had only 3 regions which matched with the crime index of that area, so only 3 regions for school boards and crime index both are considered. Furthermore, the crime severity index was available for many years, that is from 2010 to 2019, but the graduation rate for particular school boards was available only for 2013 and 2014 pass outs, so only these 2 years crime data is observed, so that there is no inconsistency in the data.

TREND, OBSERVATIONS, and INFERENCE:

School Board	Grade 9 graduation year	Num of Graduating Students	Grade 9 Enrolment	Graduation Rate (%)
Cape Breton-Victoria Regional School Board	2014	1051	1122	93.7
Halifax Regional School Board	2014	3568	3751	95.1
Annapolis Valley Regional School Board	2014	1050	1089	96.4
Cape Breton-Victoria Regional School Board	2013	987	1103	89.5
Halifax Regional School Board	2013	3591	3830	93.8
Annapolis Valley Regional School Board	2013	1013	1093	92.7

FIGURE 9

region	year	TOTAL CRIME SEVERITY INDEX
Cape Breton Region, Nova Scotia, municipal	2014	61
Halifax Metropolitan Area, Nova Scotia, municipal	2017	64
Halifax Metropolitan Area, Nova Scotia, municipal	2010	98
Cape Breton Region, Nova Scotia, municipal	2015	61
Cape Breton Region, Nova Scotia, municipal	2011	65
Cape Breton Region, Nova Scotia, municipal	2013	59
Annapolis County, Nova Scotia, Royal Canadian Mounted Police, rural	2010	82
Halifax Metropolitan Area, Nova Scotia, municipal	2012	76
Annapolis County, Nova Scotia, Royal Canadian Mounted Police, rural	2015	63
Cape Breton Region, Nova Scotia, municipal	2012	73
Annapolis County, Nova Scotia, Royal Canadian Mounted Police, rural	2013	74
Annapolis County, Nova Scotia, Royal Canadian Mounted Police, rural	2014	61
Cape Breton Region, Nova Scotia, municipal	2010	67
Cape Breton Region, Nova Scotia, municipal	2017	59
Halifax Metropolitan Area, Nova Scotia, municipal	2013	68
Halifax Metropolitan Area, Nova Scotia, municipal	2015	64
Halifax Metropolitan Area, Nova Scotia, municipal	2014	66
Halifax Metropolitan Area, Nova Scotia, municipal	2016	62
Annapolis County, Nova Scotia, Royal Canadian Mounted Police, rural	2017	58
Cape Breton Region, Nova Scotia, municipal	2016	48
Annapolis County, Nova Scotia, Royal Canadian Mounted Police, rural	2012	92
Halifax Metropolitan Area, Nova Scotia, municipal	2011	89
Annapolis County, Nova Scotia, Royal Canadian Mounted Police, rural	2016	69
Annapolis County, Nova Scotia, Royal Canadian Mounted Police, rural	2011	77

FIGURE 10

If figure 9 is observed carefully, it is inferred that the graduation rate has increased from the year 2013 to 2014, for the three regions which are Halifax, Cape Breton and Annapolis. Now after careful examination of figure 10, it can be seen that for these three regions the crime severity index is reduced for the years 2013 to 2014. For example, the CSI(Crime severity index) for Halifax region was 68 in 2013, which got reduced to 66 in 2014, similarly for Annapolis it was 74 in 2013 and got reduced to 61 in 2014. So to conclude it means that the KPI here is the graduation rate and the crime severity index. The graduation rate if increased in a particular area for a particular year, then the crime severity index of that area in that particular year is reduced, and this is proved by the data which is available. So the Education sector can be improved by increasing the Graduation rate of the boards, which is a very important KPI for education. Similarly, the Crime severity index is reduced, which is an important KPI

for increasing the safety of the people.

C. DATA DEFINITION LANGUAGE(DDL) and DATA MANIPULATION LANGUAGE(DML)

The design is ready, and according to that design the database tables are constructed. This project consists of total 8 tables. Each table is described below, along with the query used for the tables.

1)- CRIME_INDEX: This is the first table, which was designed initially after looking at the dataset. Below is the table along with the data in it. This table basically has 3 columns region, year and total crime severity index. The region column has the name of the region in of which the crime severity index is provided, year is the year in which the CSI is recorded and crime severity index is a parameter to know a crime level of a particular region.

```
SELECT * FROM homefilescenter.crime_index;
```



The screenshot shows a database interface with a 'Result Grid' tab selected. Above the grid, there are controls for 'Filter Rows' (a text box) and 'Export' (a button with a document icon). To the right of the export button is a 'Wrap Cell Content' checkbox, which is currently checked. The table below has three columns: 'region', 'year', and 'total_crime_severity_index'. The data rows are as follows:

region	year	total_crime_severity_index
Cape Breton Region, Nova Scotia, municipal	2011	65
Cape Breton Region, Nova Scotia, municipal	2013	59
Annapolis County, Nova Scotia, Royal Canadian...	2010	82
Halifax Metropolitan Area, Nova Scotia, municipal	2012	76
Annapolis County, Nova Scotia, Royal Canadian...	2015	63
Cape Breton Region, Nova Scotia, municipal	2012	73
Annapolis County, Nova Scotia, Royal Canadian...	2013	74
Annapolis County, Nova Scotia, Royal Canadian...	2014	61
Cape Breton Region, Nova Scotia, municipal	2010	67
Cape Breton Region, Nova Scotia, municipal	2017	59
Halifax Metropolitan Area, Nova Scotia, municipal	2013	68

FIGURE 11

This Crime_index table has the data populated from the dataset excel CSV file. Homefilescenter is the name of the database. This table shows the crime severity index of Cape Breton region, Halifax region and Annapolis county region for various years. This is just a snapshot of the data, not all the rows are visible here. However, the entire table is uploaded to the folders, it can be viewed from there.

2)- GRADUATION_INFO: This is the second table of the database. It consists of 5 columns – school_board, grade_9_graduation_year, num_of_graduating_students, grade_9_enrolment and graduation_rate. School_board gives the name of the board, grade_9_graduation_year gives the year of passing/graduation, num_of_graduating_students gives the total number of students graduated from that board in that year, grade_9_enrolment gives number of students enrolled in that year and graduation rate means the ratio of number of enrolled students to number of students passed in percentage. The table can be viewed from the folders. Here is a select statement to display a snapshot of data in the table.

```
select * from homefilescenter.graduation_info;
```

	school_board	grade_9_graduation_year	num_of_graduating_students	grade_9_enrolment	graduation_rate
▶	Cape Breton-Victoria Regional School Board	2014	1051	1122	93.70000
	Halifax Regional School Board	2014	3568	3751	95.10000
	Annapolis Valley Regional School Board	2014	1050	1089	96.40000
	Cape Breton-Victoria Regional School Board	2013	987	1103	89.50000
	Halifax Regional School Board	2013	3591	3830	93.80000
	Annapolis Valley Regional School Board	2013	1013	1093	92.70000

FIGURE 12

3)- CRIME_INDEX_UPDATE: This table is an updated version of the crime_index table. It is basically for renaming the longer names of the places in the original datasets to a much shorter name. For instance, in crime_index it can be observed that the values in the first column are quite long, so to simplify that another column is added to the table, whose name is region_name and in this column simplified name is written. The query used for this operation is shown below, along with the result of the operation.

```
create table homefilescenter.crime_index_update as(
select region,year,total_crime_severity_index,
case when region="Halifax Metropolitan Area, Nova Scotia, municipal" then "Halifax"
when region="Cape Breton Region, Nova Scotia, municipal" then "Cape Breton"
when region="Annapolis County, Nova Scotia, Royal Canadian Mounted Police, rural" then "Annapolis"
end as region_name
from homefilescenter.crime_index);
```

FIGURE 13

The name of the new column is region_name and the new regions simplified names are “Halifax”, “Cape Breton” and “Annapolis”. The table data snapshot for crime_index_update is shown below:

```
75 • SELECT * FROM homefilescenter.crime_index_update;
```

```
76
```

```
77
```

region	year	total_crime_severity_index	region_name
Annapolis County, Nova Scotia, Royal Canadian...	2010	82	Annapolis
Cape Breton Region, Nova Scotia, municipal	2010	67	Cape Breton
Halifax Metropolitan Area, Nova Scotia, municipal	2010	98	Halifax
Annapolis County, Nova Scotia, Royal Canadian...	2011	77	Annapolis
Cape Breton Region, Nova Scotia, municipal	2011	65	Cape Breton
Halifax Metropolitan Area, Nova Scotia, municipal	2011	89	Halifax
Annapolis County, Nova Scotia, Royal Canadian...	2012	92	Annapolis
Cape Breton Region, Nova Scotia, municipal	2012	73	Cape Breton
Halifax Metropolitan Area, Nova Scotia, municipal	2012	76	Halifax
Annapolis County, Nova Scotia, Royal Canadian...	2013	74	Annapolis
Cape Breton Region, Nova Scotia, municipal	2013	59	Cape Breton

FIGURE 14

Region_name column is added and a new table is created.

4)- GRADUATION_INFO_UPDATE: Similarly, a new table is created for graduation_info, for getting a common region name to join both graduation_info_update with crime_index_update. In this table also, region_name is added, which will be a common column for both the tables. The query for adding the column and creating a new table is shown below, along with the query to view the newly created table.

```
create table homefilescenter.graduation_info_update as(
select school_board,grade_9_graduation_year,num_of_graduating_students,grade_9_enrolment,graduation_rate,
      case when school_board="Cape Breton-Victoria Regional School Board" then "Cape Breton"
            when school_board="Halifax Regional School Board" then "Halifax"
            when school_board="Annapolis Valley Regional School Board" then "Annapolis"
            end as region_name
from homefilescenter.graduation_info);
```

FIGURE 15

```
71 • select * from homefilescenter.graduation_info_update;
```

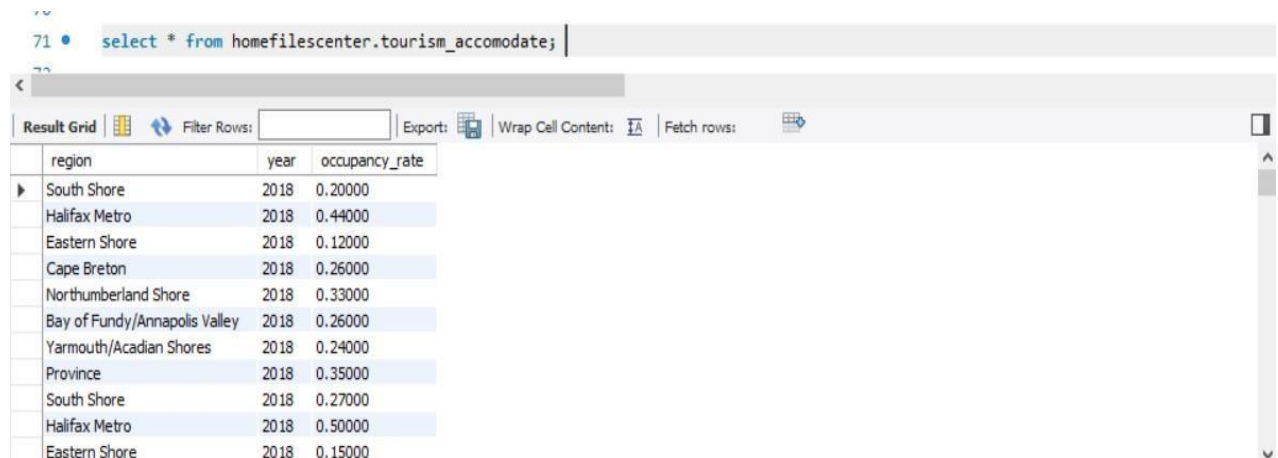
```
72
```

school_board	year	num_of_graduating_students	grade_9_enrolment	graduation_rate	region_name
Annapolis Valley Regional School Board	2013	1013	1093	92.70000	Annapolis
Cape Breton-Victoria Regional School Board	2013	987	1103	89.50000	Cape Breton
Halifax Regional School Board	2013	3591	3830	93.80000	Halifax
Annapolis Valley Regional School Board	2014	1050	1089	96.40000	Annapolis
Cape Breton-Victoria Regional School Board	2014	1051	1122	93.70000	Cape Breton
Halifax Regional School Board	2014	3568	3751	95.10000	Halifax

FIGURE 16

5)-TOURISM ACCOMMODATE: This table is basically for tracking the occupancy rate of the tourists who visit Nova Scotia. It consists of 3 columns, region, year and occupancy rate. Region tells in which region the accomodation is present, the year(primary key) tells in which year the tourist came and the

occupancy rate tells how much time the accommodation was occupied. The table can be viewed after populating the data from the CSV file as shown below:



The screenshot shows a database interface with a query bar containing the SQL statement: `select * from homefilescenter.tourism_accomodate;`. Below the query bar is a 'Result Grid' with a table of data. The table has three columns: 'region', 'year', and 'occupancy_rate'. The data is as follows:

region	year	occupancy_rate
South Shore	2018	0.20000
Halifax Metro	2018	0.44000
Eastern Shore	2018	0.12000
Cape Breton	2018	0.26000
Northumberland Shore	2018	0.33000
Bay of Fundy/Annapolis Valley	2018	0.26000
Yarmouth/Acadian Shores	2018	0.24000
Province	2018	0.35000
South Shore	2018	0.27000
Halifax Metro	2018	0.50000
Eastern Shore	2018	0.15000

FIGURE 17

The data comes month wise, that is the 2018 written is month wise, so the average is calculated for every year and a new table is created which shows the average occupancy rate of a particular year. Since all the regions are in Nova Scotia, so region column is omitted from the new table. The average for each year is taken so that it can be compared with the average number of visitors for a particular year from tourism_visitor table. So the new table is Tourism_accomodate_update.

6)- TOURISM ACCOMODATE UPDATE: This table consists of 2 columns average occupancy rate per year of accommodations and the year(primary key). It is created using the query as shown below. Avg function is used to calculate the average of the rows.

```
create table homefilescenter.tourism_accomodate_update
as(select year,avg(occupancy_rate) as average_occupancy_rate FROM homefilescenter.tourism_accomodate
group by year
order by year);
```

FIGURE 18

After this query the table is created, which can be viewed from the select statements as shown below:

76 • `SELECT * FROM homefilescenter.tourism_accomodate_update;`

77

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content: `IA`

year	average_occupancy_rate
2007	0.401250000
2008	0.400416666
2009	0.381666666
2010	0.385208333
2011	0.389062500
2012	0.386458333
2013	0.378437500
2014	0.398125000
2015	0.415625000
2016	0.433229166
2017	0.461979166

FIGURE 19

7)- TOURISM VISITOR: This table is for the details of the tourists visiting to Nova Scotia. It has year, num_of_visitors and country as columns. Year is the primary key column and consists of the year in which the tourist has visited the country, num_of_visitors shows the total number of visitors and country shows the place where the visitor belongs to. In this table also, average needs to be calculated for each year to make the relationship, since month wise data is not available in the data sets, so taking the average of all the months gives the number of visitors for that particular year, which is related with average occupancy rate of that particular year. The initial table is shown below :

72 • `select * from homefilescenter.tourism_visitor;`

73

Result Grid | Filter Rows: | Export: | Wrap Cell Content: `IA` | Fetch rows: `1A`

year	country	num_of_visitors
2006	Canada	5400
2006	Canada	3400
2006	Canada	16600
2006	Canada	7000
2006	United States	800
2006	United States	1200
2006	United States	300
2006	United States	300
2006	United States	1400
2006	United States	200
2006	United States	500

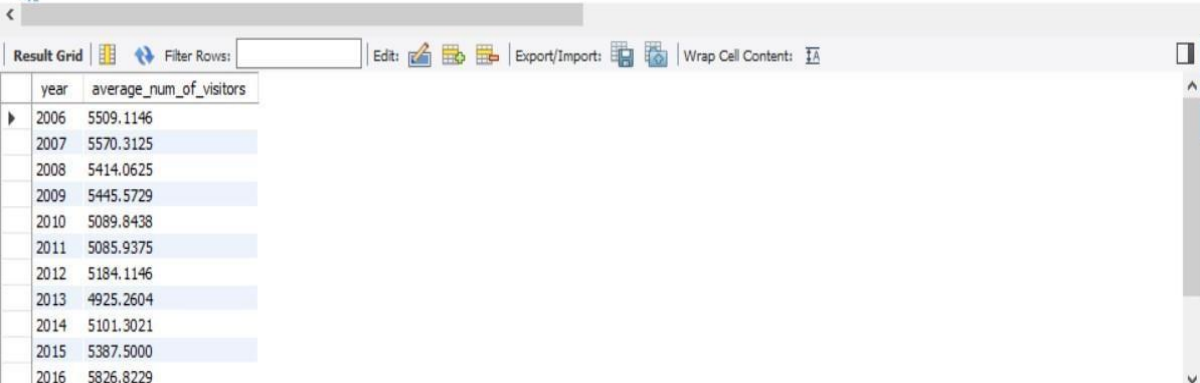
8)- TOURISM_VISITOR_UPDATE: This table is created after taking average of number of visitors for each month, to give the average number of visitors for each year. This country column will be removed from tourism_visitor table, since that is not required for analysis. The year column will act as primary key in this table, since it will be unique. The table creation query along with the result is shown below.


```
create table homefilescenter.tourism_visitor_update
as(select year,avg(NUM_OF_VISITORS) as average_num_of_visitors FROM homefilescenter.TOURISM_VISITOR
group by year
order by year);
```

FIGURE 21

Now to view this table, the select statement is used as shown below.

73 • select * from homefilescenter.tourism_visitor_update;



year	average_num_of_visitors
2006	5509.1146
2007	5570.3125
2008	5414.0625
2009	5445.5729
2010	5089.8438
2011	5085.9375
2012	5184.1146
2013	4925.2604
2014	5101.3021
2015	5387.5000
2016	5826.8229

FIGURE 22

BUILDING THE RELATION IN THE TABLES AND CALCULATING THE RESULTS(KPIs)

1)- After looking at the tables above, it can be clearly observed that the Graduation rate of the school boards are higher for 2014 as compared to 2013. The crime severity index of those regions, where the graduation rate is higher, is decreasing from 2013 to 2014. So it can be deduced that, the region where the graduation rate is higher, it has lesser crime and should be considered to be a safer place. So graduation rate and crime severity index are the two critical KPIs for education and safety sectors, since higher graduation rate means more are getting educated, that is good for the society. Similarly lower crime severity index means, lower number of severe crimes, that will be vital for the safety of the region. Now to show this deduction in tabular format, join operation is required between the linking tables.

2)- The second pair of tables is of tourism and accommodations. It can be observed from the tables that the average number of visitors from the year 2006 to 2013 has not shown any definite trend(mostly decreased number of visitors), but from 2014 to 2019, it is clearly visible that the number of visitors has increased. Now for the accommodations, the occupancy rates have shown very similar trends. The occupancy rates from 2007 to 2013 are mixed but from 2014 to 2018, they have increased.

So it can be deduced, that the number of visitors increasing from 2014 has an impact on the accommodation, and thus the occupancy rates also have increased. So the KPI here for the business is average number of visitors, which is getting increased and due to that it means more profit for the tourism industry. Similarly the KPI for lifestyle sector is occupancy rates of the hotels, higher occupancy rate is a measure of quality of the accommodations.

JOIN OPERATIONS FOR DISPLAYING THE FINAL RESULT:

The tables will be joined for showing the trends and deducing the result.

a)- GRADUATION_INFO_UPDATE will be joined by CRIME_INDEX_UPDATE

b)- TOURISM_VISITOR_UPDATE will be joined by TOURISM_ACCOMMODATE_UPDATE

a)- The query for the join operation is shown below, along with the results.

```

7  select a.*,b.total_crime_severity_index from
8  homefilescenter.graduation_info_update a, homefilescenter.crime_index_update b
9  where a.region_name=b.region_name
10 and a.year=b.year;

```

school_board	year	num_of_graduating_students	grade_9_enrolment	graduation_rate	region_name	total_crime_severity_index
Annapolis Valley Regional School Board	2013	1013	1093	92.70000	Annapolis	74
Cape Breton-Victoria Regional School Board	2013	987	1103	89.50000	Cape Breton	59
Halifax Regional School Board	2013	3591	3830	93.80000	Halifax	68
Annapolis Valley Regional School Board	2014	1050	1089	96.40000	Annapolis	61
Cape Breton-Victoria Regional School Board	2014	1051	1122	93.70000	Cape Breton	61
Halifax Regional School Board	2014	3568	3751	95.10000	Halifax	66

Here the tables GRADUATION_INFO_UPDATE and CRIME_INDEX_UPDATE are joined using the year and region_name columns. The result can now be observed in this single table. The graduation rate has increased from 2013 to 2014 for all the boards(Annapolis, Halifax, Cape Breton) in these regions and the crime severity index has been decreased from 2013 to 2014 for these regions. NOTE: The graduation rate was only available for the years 2013, 2014, so only these are compared. So the KPIs are graduation rate and total crime severity index.

b) The query for the second join operation is shown below, along with the results.

```

67 • select a.*,b.average_occupancy_rate from
68 homefilescenter.tourism_visitor_update a,homefilescenter.tourism_accomodate_update b
69 where a.year=b.year;

```

year	average_num_of_visitors	average_occupancy_rate
2007	5570.3125	0.40125000
2008	5414.0625	0.40041666
2009	5445.5729	0.38166666
2010	5089.8438	0.38520833
2011	5085.9375	0.38906250
2012	5184.1146	0.38645833
2013	4925.2604	0.37843750
2014	5101.3021	0.39812500
2015	5387.5000	0.41562500
2016	5826.8229	0.43322916
2017	6384.7769	0.46197916

FIGURE 23

Here the tables TOURISM_VISITOR_UPDATE and TOURISM_ACCOMODATE_UPDATE are joined using the year column. Now it is clearly visible that from 2007 to 2013 the average number of visitors has shown a mixed response, but from 2014 to 2019 they have increased. The average occupancy rate from 2007 to 2013 has shown very similar trend and from 2014 to 2019, it has also increased. So the visitors impact is shown in the occupancy rate. More the visitors, higher the occupancy rate, this is displayed by the data.

ANSWERS TO DML QUESTIONS:

Which region has more schools.

By looking at the grade_9_enrolment column of graduation_info_update table and joined table, it is clearly visible that maximum number of enrollments were in Halifax region school board, so it means the maximum number of schools are present in Halifax. Second number is of Cape Breton and last one is Annapolis for the year 2013 to 2014. This answer is based on the dataset available for the year and the region.

Which region has more number of reported crimes?

By looking at the crime severity indexes, it can be observed that the Halifax metropolitan area has the highest crime rate among the three Cape Breton and Annapolis. However the data has changed year

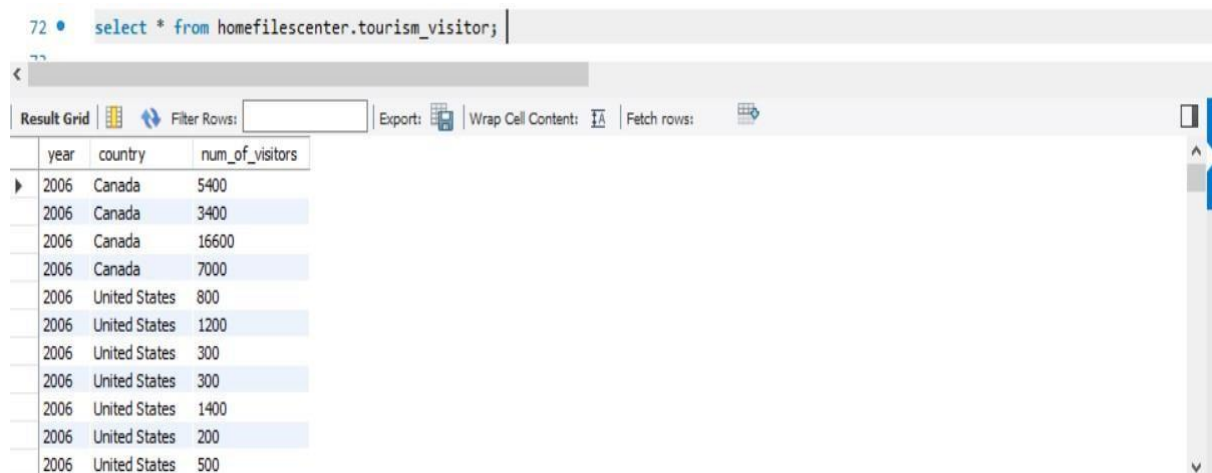
by year, so it cannot be deduced conclusively that Halifax has the highest crime. All the results given are based on the data sets available. For more detail analysis, more data sets are required.

Which business organization has highest number of employees

The data set for answering this question was not found.

D. NORMALIZATION:

It is a process of removing the redundant data from the tables. First all the un-normalized tables are displayed. Here are the un-normalized tables.



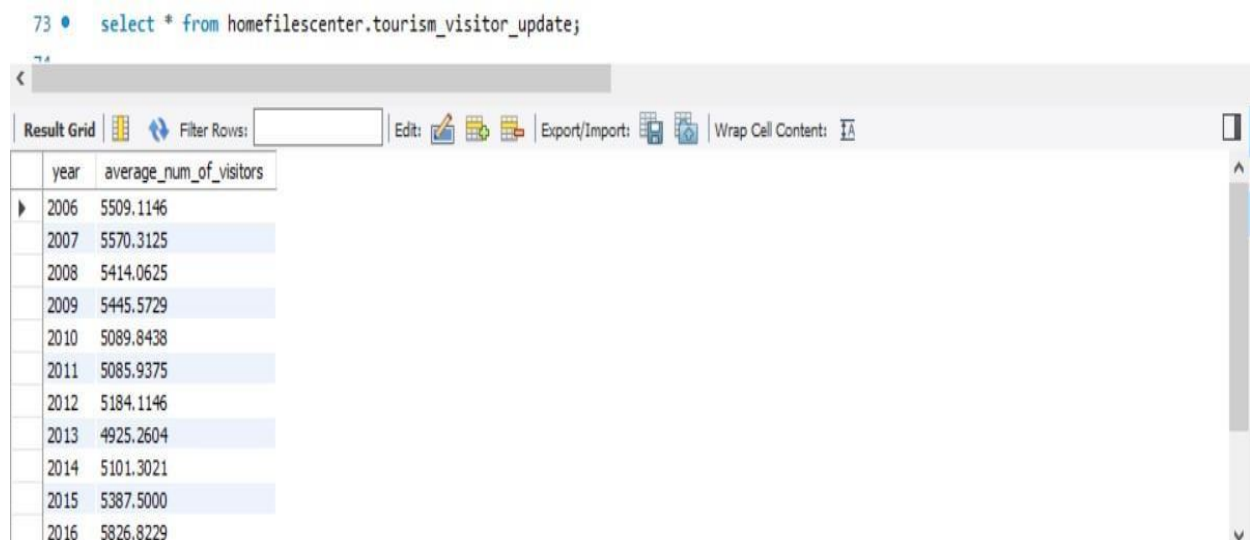
The screenshot shows a SQL query window with the query: `select * from homefilescenter.tourism_visitor;`. The result grid displays a table with three columns: year, country, and num_of_visitors. The data is as follows:

year	country	num_of_visitors
2006	Canada	5400
2006	Canada	3400
2006	Canada	16600
2006	Canada	7000
2006	United States	800
2006	United States	1200
2006	United States	300
2006	United States	300
2006	United States	1400
2006	United States	200
2006	United States	500

FIGURE 24

After taking the average for each year and making year as the primary key it is now satisfying the 2NF.

Now since this resultant table is in 2NF and does not has any transitive dependencies, it is also in 3NF.

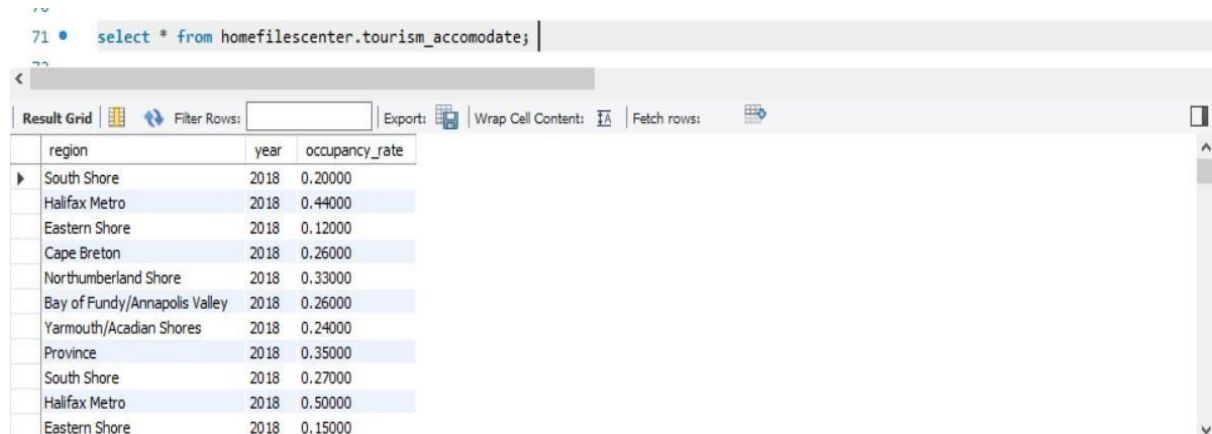


The screenshot shows a SQL query window with the query: `select * from homefilescenter.tourism_visitor_update;`. The result grid displays a table with two columns: year and average_num_of_visitors. The data is as follows:

year	average_num_of_visitors
2006	5509.1146
2007	5570.3125
2008	5414.0625
2009	5445.5729
2010	5089.8438
2011	5085.9375
2012	5184.1146
2013	4925.2604
2014	5101.3021
2015	5387.5000
2016	5826.8229

FIGURE 25

Similarly the table shown below is in un-normalized state.

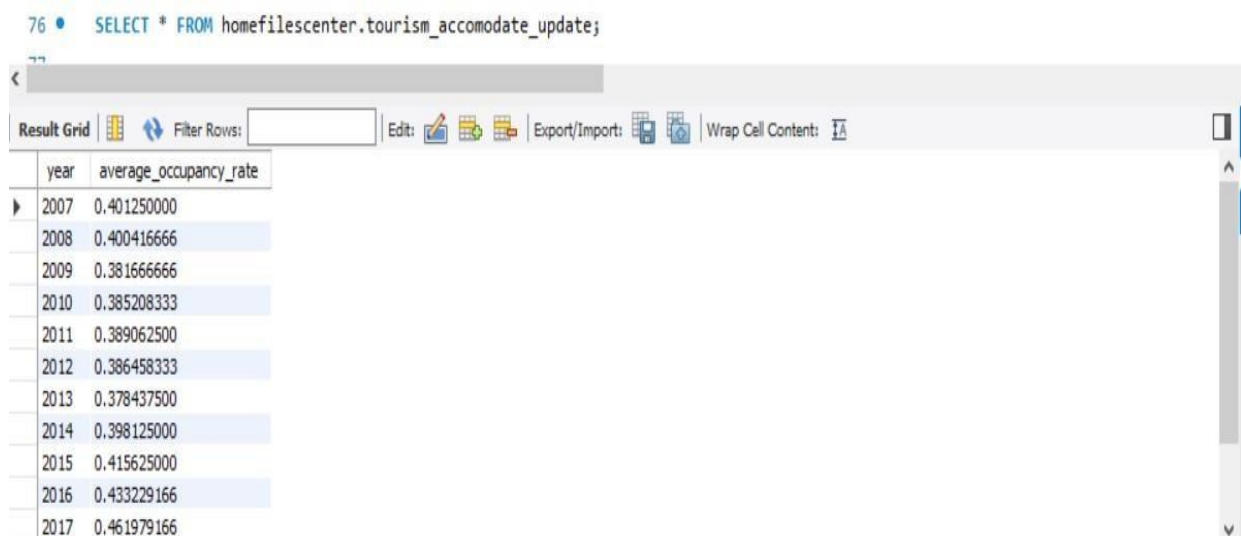


```
71 • select * from homefilescenter.tourism_accomodate;
```

region	year	occupancy_rate
South Shore	2018	0.20000
Halifax Metro	2018	0.44000
Eastern Shore	2018	0.12000
Cape Breton	2018	0.26000
Northumberland Shore	2018	0.33000
Bay of Fundy/Annapolis Valley	2018	0.26000
Yarmouth/Acadian Shores	2018	0.24000
Province	2018	0.35000
South Shore	2018	0.27000
Halifax Metro	2018	0.50000
Eastern Shore	2018	0.15000

FIGURE 26

The primary key of this table is year. After removing the region column, there is single primary key which is year, so the table is in 2NF. Moreover, there is no transitive dependency, so it becomes in 3NF.



```
76 • SELECT * FROM homefilescenter.tourism_accomodate_update;
```

year	average_occupancy_rate
2007	0.401250000
2008	0.400416666
2009	0.381666666
2010	0.385208333
2011	0.389062500
2012	0.386458333
2013	0.378437500
2014	0.398125000
2015	0.415625000
2016	0.433229166
2017	0.461979166

FIGURE 27

So all the tables are normalized and have been uploaded in the folder.

CONCLUSION:

The Key performance indicators which have been found with this data analysis are Graduation rate, Crime severity index, Number of visitors/tourists and the Occupancy rates of the accommodations. Each KPI has its own impact in business, education, lifestyle and safety. The impact analysis has been explained above in detail, along with the relation between them. All the files have been added to the project folder including the datasets and the tables.