

Data Warehousing and Business Intelligence Project

on

Tourism sector's impact on Ireland Economy

Sindhujan Dhayalan
X17170265

MSc/PGDip Data Analytics – 2018/9

Submitted to: Dr. Horacio González-Vélez

National College of Ireland
Project Submission Sheet – 2017/2018
School of Computing



Student Name:	Sindhujan Dhayalan
Student ID:	X17170265
Programme:	MSc Data Analytics
Year:	2018/9
Module:	Data Warehousing and Business Intelligence
Lecturer:	Dr. Horacio González-Vélez
Submission Due Date:	26/11/2018
Project Title:	Tourism sector's impact on Ireland Economy

I hereby certify that the information contained in this (my submission) is information pertaining to my own individual work that I conducted for this project. All information other than my own contribution is fully and appropriately referenced and listed in the relevant bibliography section. I assert that I have not referred to any work(s) other than those listed. I also include my TurnItIn report with this submission.

ALL materials used must be referenced in the bibliography section. Students are encouraged to use the Harvard Referencing Standard supplied by the Library. To use other author's written or electronic work is an act of plagiarism and may result in disciplinary action. Students may be required to undergo a viva (oral examination) if there is suspicion about the validity of their submitted work.

Signature:	
Date:	November 26, 2018

PLEASE READ THE FOLLOWING INSTRUCTIONS:

1. Please attach a completed copy of this sheet to each project (including multiple copies).
2. **You must ensure that you retain a HARD COPY of ALL projects**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. Please do not bind projects or place in covers unless specifically requested.
3. Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Table 1: Mark sheet – do not edit

Criteria	Mark Awarded	Comment(s)
Objectives	of 5	
Related Work	of 10	
Data	of 25	
ETL	of 20	
Application	of 30	
Video	of 10	
Presentation	of 10	
Total	of 100	

Project Check List

This section capture the core requirements that the project entails represented as a check list for convenience.

- ☒ Used L^AT_EX template
- ☒ Three Business Requirements listed in introduction
- ☒ At least one structured data source
- ☒ At least one unstructured data source
- ☒ At least three sources of data
- ☒ Described all sources of data
- ☒ All sources of data are less than one year old, i.e. released after 17/09/2017
- ☒ Inserted and discussed star schema
- ☒ Completed logical data map
- ☒ Discussed the high level ETL strategy
- ☒ Provided 3 BI queries
- ☒ Detailed the sources of data used in each query
- ☒ Discussed the implications of results in each query
- ☒ Reviewed at least 5-10 appropriate papers on topic of your DWBI project

Tourism sector's impact on Ireland Economy

Sindhujan Dhayalan
X17170265

November 26, 2018

Abstract

Tourism is one of the major economic industrial sector in Ireland. It is contributing to Ireland financial growth enormously and looks promising to grow furthermore in the upcoming years. Statistics show that the growth of tourism sector is directly and indirectly contributing in large proportions to the growth of GDP and job opportunities in Ireland. It is forecast to increase to larger proportion in forthcoming years contributing to the entire countrys economy. This project is focused on analyzing the major factors contributing to the growth of economy in Irelands tourism sector and its impact on living standards of people in Ireland. The following business requirements is analyzed to gain insights sufficient to answer the objective of the project: 1) The type of outbound tourists contributing to the revenue of Ireland tourism sector. 2) The type of tourists spending the most and the places generating most of the expenditure. 3) The Impact of growth in tourism to the people living in Ireland.

1 Introduction

Statistics show that tourism sector in 2017 accounted for 5.9% GDP of Ireland and is expected to grow by 4.2% to 17.9billion(Euros) in 2018(6.0% of GDP). It is expected to rise by 3.8% PA to 26.0billion(EUROS) by 2028 (7.1% of GDP). Travel Tourism generated 39,500 jobs directly in 2017 (1.9% of total employment) and this is expected to rise by 4.9% in 2018 to 41,500(2.0% of total employment). This includes employment by accommodation industry, food industry, leisure industry, travel agents, airlines and other passenger transportation services (excluding commuter services) directly servicing the tourist wttc.org (2018). The core factors contributing to the growth must be analyzed to further withhold and increase the popularity and financial contribution to the country. This project is focused on gaining knowledge on the below three business requirements:

- (Req-1) **Requirement 1:** Type of outbound tourists contributing to the revenue of Irelands tourism sector. Analysis on this requirement will result in knowledge on the growth of revenue by contribution of tourists from different regions of the world. Further this will give insights on the rise and decline in the popularity of Ireland over different regions of the world.

- (Req-2) **Requirement 2:** The type of tourists spending the most, the sites contributing to the expenditure of the tourists. Analysis on this requirement will result in knowledge on destinations contributing highly towards tourists expenditure on categorizing the tourists by gender, purpose and region. The findings from this requirement will aid in enhancing expenditure of different type of tourists.
- (Req-3) **Requirement 3:** The Impact of growth in tourism to the people living in Ireland. This requirement is analyzed to get knowledge on improvement in life standards of people in Ireland resulting from the growth of tourism.

2 Data Sources

Seven Datasets from four different sources are used for analysis. Detailed description of the dataset and its sources are mentioned in this section.

Source	Type	Brief Summary
CSO.ie	Structured	The dataset contains the overseas trips to Ireland by area of residence for years, 2010 - 2017
Eurostat	Structured	Two datasets having data about the Employment by Ireland tourism sector and Expenditure by outbound tourists in Ireland have been sourced from this site
Independent.ie failteireland.ie	Unstructured	Data from Articles about the Top 5 paid attractions in Independent.ie and failteireland.ie over period of 2012 to 2017 is web scrapped and merged using R to create the dataset. Main source of data used in these articles are sourced from failteireland.ie
Statista	Structured	Datasets consisting data related to the economy of Ireland such as revenue from tourism, GDP and GDP per capita have been used in this project

Table 2: Summary of sources of data used in the project

2.1 Source 1: Central Statistics Office

Downloaded the CSO: Overseas Trips to Ireland dataset from : <https://www.cso.ie/en/releasesandpublications/er/ot/overseastraveljanuary2018/>. The dataset provides data on the number of trips taken by tourists from different countries in a period of 8 years from 2010-2017, For this project requirement, only the recent 6 years of data is used and compared with other sources of data to analyze and gain knowledge on the business requirement 1 mentioned in section 1. The publish date of dataset accessed is illustrated in Figure 1.

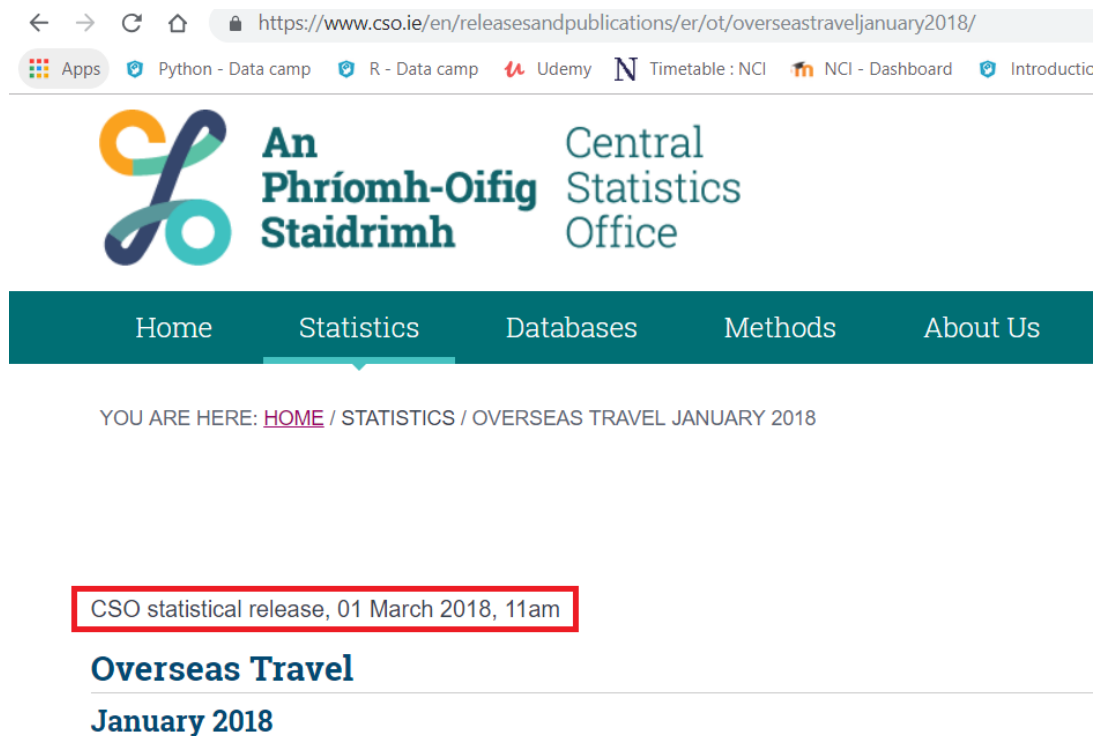


Figure 1: Publish date of dataset - Overseas Trips from source CSO

2.2 Source 2: Eurostat

Downloaded the Eurostat: Employment dataset from: http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=tour_lfs6r2&lang=en. Dataset provides structured data on the Employment in tourism sector based on gender, sectors employed over a period of 10 years from 2008 to 2017. This data is used to gain knowledge on the 3rd business requirement mentioned in section 1. The dataset allows to drill down on gender and sector basis to analyze and derive insights on business requirement 3. The publish date of dataset accessed is illustrated in Figure 2

Downloaded the Eurostat: Expenditure of tourists dataset from: http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=tour_dem_exsex&lang=en. Dataset provides structured data on the expenditure of the tourists in Ireland based on gender, spending over a period of 10 years from 2008 to 2017. This data is used to gain knowledge on the 2nd business requirement mentioned in section 1. The dataset allows to drill down on gender, purpose and region basis to analyze and derive insights on business requirement 2. The publish date of dataset accessed is illustrated in Figure 3

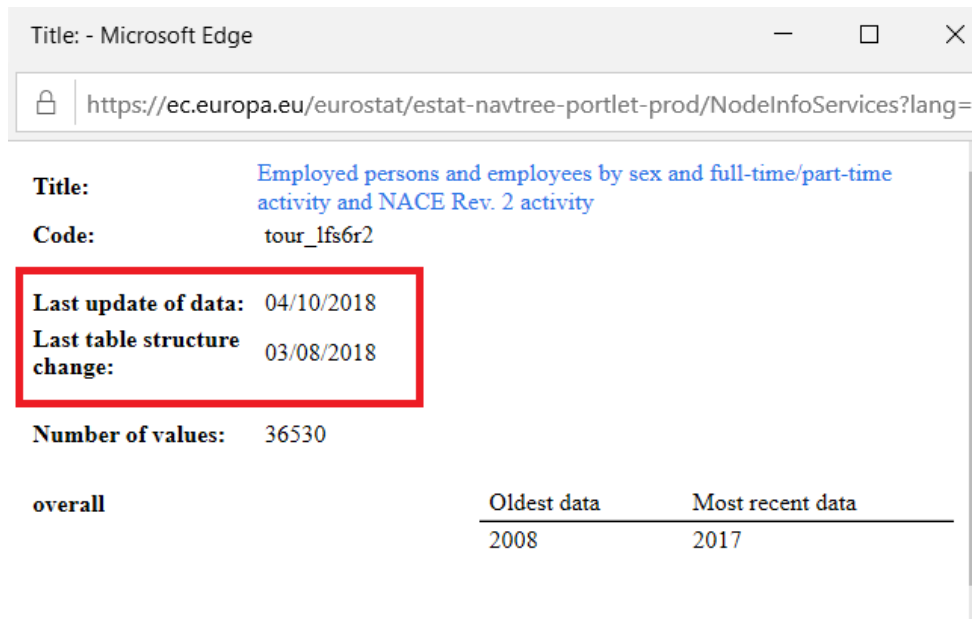


Figure 2: Publish date of dataset - Employment

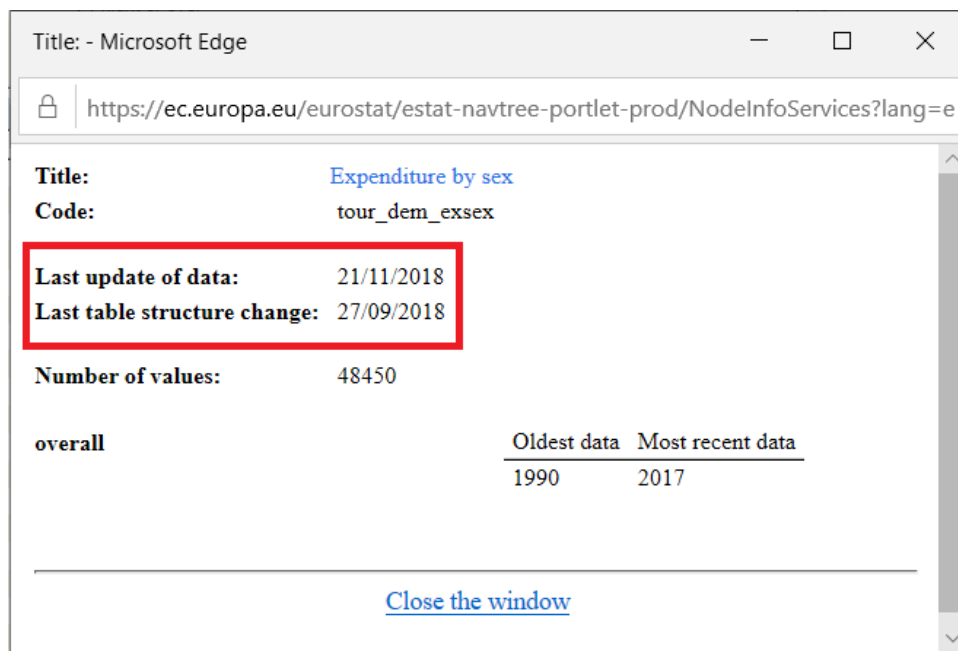


Figure 3: Publish date of dataset - Expenditure

2.3 Source 3: Failteireland.ie and Independent.ie

Web scrapped information on top 5 paid attraction in Ireland from 2012 to 2017. The dataset provides data to analyze on the tourist's preferences, industries attracting tourists, the changes in the popularity of the attractions over a period. The unstructured form of data is taken from the following sources: Independent.ie is a reliable source of news article and failteireland.ie is officially one of Ireland tourism websites. Screen-shot of approved email confirmation by editor of independent.ie for web scraping the data available in articles is illustrated in Figure 4 . The data scrapped from the independent.ie article

is sourced back to failteireland.ie website. Below are the URLs for the articles being web scraped.

1. 2017:<https://www.independent.ie/life/travel/travel-news/revealed-irelands-top-20-tourist-attractions-paid-and-free-36986059.html>.
2. 2016:<http://www.failteireland.ie/Utility/News-Library/Irelands-Most-Popular-Visitor-Attractions-Revealed.aspx>.
3. 2015:<https://www.independent.ie/life/travel/ireland/irelands-top-visitor-attractions-revealed-and-the-numbers-are-booming-34947131.html>.
4. 2014:<https://www.independent.ie/life/travel/ireland/irelands-top-tourist-attractions-revealed-31370553.html>.
5. 2013:<https://www.independent.ie/life/travel/ireland/irelands-top-10-tourist-attractions-revealed--30397521.html>.
6. 2012:<http://www.failteireland.ie/Footer/Media-Centre/Ireland-s-most-popular-tourist-attractions-for-201.aspx>.

The Source of data is from failteireland.com article, but article on few years are not available in failteireland, as an alternative independent.ie articles were used for web scrapping using Rstudio. The data from this source is used as a part of analyzing the business requirement 2 mentioned in section 1.

Re: Request for access to data about top tourism site articles for masters project » Inbox

Pól Ó Conghaile poconghaile@independent.ie via student.ncirl.ie
to Sindhujan ▾

Hi Sindhujan,

This data is from Fáilte Ireland and the CSO.

You are welcome to quote from the article, but you may also wish to go to those two primary sources for the latest tourism data.

Best wishes for your project.

Pól

Pól Ó Conghaile

Travel Editor - Independent.ie & Irish Independent

M +353 087 419-9505

W www.inm.ie

Figure 4: Email approval for quoting independent.ie article data

2.4 Source 4: Statista

3 Datasets consisting of Ireland's economic factors relating to tourism sector is downloaded from Statista. These data sets are key to the analysis of the project's business requirements as GDP per capita income depicts the standard of living of people in the country, GDP depicts the overall economy of the country and Revenue dataset contains the revenue generated from tourism sector. For analysis with other sources of data, a specific period from 2012 to 2017 is used from these data sets. The values from the datasets constitutes as measures in the fact table of the data warehouse to be implemented. The data from this source is used as a part of analyzing the business requirement 1 and 3 mentioned in section 1. The publish date of the datasets sourced from statista are illustrated in Figure 5

1. GDP per capita income: <https://www.statista.com/statistics/377002/gross-domestic-product-gdp-per-capita-in-ireland/>.
2. GDP: <https://www.statista.com/statistics/375217/gross-domestic-product-gdp-in-ireland/>.
3. Revenue: <https://www.statista.com/statistics/660115/tourism-revenue-in-ireland/>.

Revenue publish date			GDP publish date			GDP/capita income date		
DESCRIPTION	SOURCE	MORE INFORMATION	DESCRIPTION	SOURCE	MORE INFORMATION	DESCRIPTION	SOURCE	MORE INFORMATION
Sources Fáilte Ireland; Central Statistics Office Ireland; NISRA; Central Bank of Ireland			Source IMF			Source IMF		
Survey by Central Statistics Office Ireland; NISRA; Fáilte Ireland; Central Bank of Ireland			Survey by IMF			Survey by IMF		
Survey name Fáilte Ireland: Tourism facts 2017			Survey name World Economic Outlook Database October 2018			Survey name World Economic Outlook Database October 2018		
Published by Fáilte Ireland			Published by IMF			Published by IMF		
Source link Fáilte Ireland: Tourism facts 2017, page 4			Source link imf.org			Source link imf.org		
Release date July 2018			Release date October 2018			Release date April 2018		

Figure 5: Publish date of Revenue, GDP and GDP per capita income datasets from statista

3 Related Work

Tourism sector in many countries are contributing as one of the major sector in country's economical growth. Comparing the Tourism receipts and Gross domestic product over a range of 1963-2006 period for turkey analyzed by using state-space time-varying coefficients and rolling window methods resulted that tourism receipts have positive impact on the economic development in Turkey Arslanturk et al. (2011). Another literature paper results show a one-way Granger causality from domestic tourism expenditure against

economic growth in Cyprus, Latvia and Slovakia, implying that tourism growth is of great importance to economic growth in these countries. Same way, in another research relationship between economic development and tourism expenditure indicated that economic development can rise the demand for tourism and lead to the development of tourism sectors in Czech Republic and Poland. Statistics show that tourism development and economic growth mutually influence each other for Estonia and Hungary. Where as other empirical results support the neutrality hypothesis for Bulgaria, Romania and Slovenia, indicating that neither tourism development nor economic growth is sensitive to each other in these 3 countries Chou (2013). Above findings show that tourism impacts differently to the economic growth of different countries, This project is focused on growth of tourism in Ireland in its economic development. The travel and tourism sector of Ireland supported about 122000 jobs in 2017. This is expected to rise by 4.1% in 2018 and rise by 3.6% pa to 180000 jobs in 2028(7.5% of total employment of Ireland) wttrc.org (2018). Similar datasets used in this products such as GDP, GDP per capita income and revenue sourced from eurostat and most visited attractions from failteireland.ie are used for above paper. But this project is focused on the impact on the life of people by the economic growth. this aspects of relationship with growth of tourism sector is addition to the already existing research papers with impact of economy due to tourism in Ireland. the tourism data warehousing model can produce business intelligent data report used to analyze with various levels of granularity, which facilitate understanding the information, save time, store data in one place, and generate efficient reports that can help in requirements of the decision makers in the tourism sector Abdulaziz et al. (2015).

Britain, the United States, Germany, and France are the four major sources of market for the Irelands tourism sector OLeary & Deegan (2005). they are still topping as the major source of Irish tourism market and similar datasets used in the above paper is used for this project sourced from CSO.ie. but in this project, the dataset is used to analyze data from different angle. where in, we analyze on how Irish tourism survived the effects of Britain's decline in tourist with the revenue in business query 1. Complete analysis emerges when main purpose for visiting Ireland is queried. Main purpose is drilled down into three categories: holiday, business and visiting friends and relatives. holiday-makers are the most valuable contributors to Irish tourism promotional initiatives Deegan & Di-neen (2003). One of the dataset from eurostat used in this project is also drilled down into similar hierarchy in this paper, but the analysis angle is different. this paper handles the dataset to compare it with the most visited attractions to get insights on business requirement 2 mentioned in section 1.

4 Data Model

The Data model of the project is designed by using dimensional model(star schema) following the Kimball's bottom up approach for data warehousing. Star schema is used to map the denormalized descriptive dimensions to the fact table and then further process the OLAP cube. Design of the schema is illustrated in Figure 6. Using the schema will enable in simpler SQL query, usage of lesser joins and faster performance and quick data warehouse to build Kimball et al. (2013). The Schema consists of 6 dimensions, detailed description of the sources and uses of dimensions and fact table for analysis of business requirements are explained in this section.

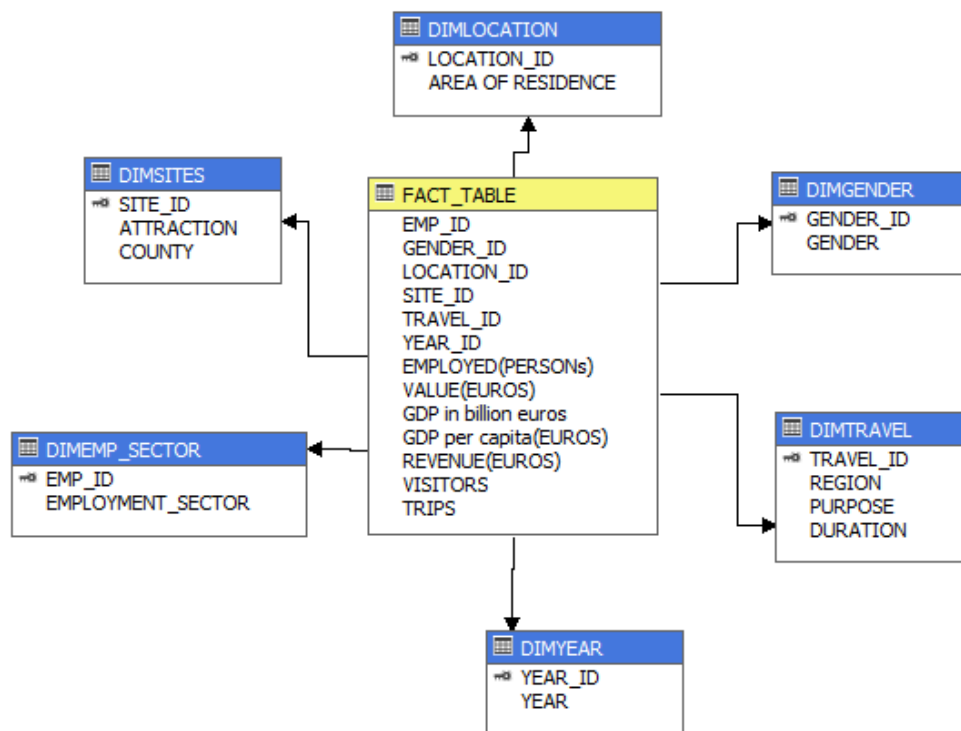


Figure 6: Star schema

Dimension 1: DIMYEAR, Source: Can be derived from all the 4 sources:

DIMYEAR consists of YEAR ID(primary key) and YEAR.it consist of years ranging from 2008 to 2017 derived from the employment dataset from the Eurostat source.DIMYEAR can be used to link and compare all the measures in the fact table by filtering out the years. For this project,data from recent 6 years(2012-2017) is used to correlate with other datasets from different sources.YEAR column in DIMYEAR links to all the measures in the FACT_TABLE and is major part of analysis in all the 3 business query.

Dimension 2: DIMEMP_SECTOR, Source: Eurostat:

EMP_ID	EMPLOYMENT_SECTOR
1	Accommodation
2	Accommodation and food service activities
3	Air transport
4	All NACE Activities
5	Travel agency, tour operator reservation service and related activities

Figure 7: DIMEMP SECTOR Attributes

DIMEMP_SECTOR consists of EMP_ID(primary key) and EMPLOYMENT_SECTOR data from the Employment dataset sourced from Eurostat. it links to the EMPLOYED(people) measure in FACT_TABLE and can be filtered with various categories available illustrated in Figure 7. The data from this dimension is a part in analysis of Business intelligence query 3.

Dimension 3: DIMGENDER, source: Eurostat:

DIMGENDER consists of GENDER ID and GENDER data from the Employment dataset sourced from Eurostat. DIMGENDER also is linked with Expenditure dataset sourced from Eurostat. it can be used with the measures REVENUE(EUROS) and EMPLOYED(people) in fact table and can be drilled down between male and female attributes. This dimension is used as part of analysis in business intelligence query 2 and 3.

Dimension 4: DIMSITES, source: Failteireland.ie:

SITE_ID	ATTRACTION	COUNTY
1	Book of Kells Trinity College	Dublin
2	Cliffs of Moher	Co. Clare
3	Dublin Zoo	Dublin
4	Guinness Storehouse	Dublin
5	National Aquatic Centre	Dublin

Figure 8: DIMSITES Attributes

DIMSITES consists of SITE ID, ATTRACTION and COUNTY data from the unstructured data sourced from articles published in Independent.ie and failteireland.ie. it is mapped with the measure VISITORS in fact table and can be drilled down by ATTRACTION and COUNTY attributes illustrated in Figure 8. it is used as part of analysis in business intelligence query 2.

Dimension 5: DIMTRAVEL, source: Eurostat:

TRAVEL_ID	REGION	PURPOSE	DURATION
1	Domestic	For leisure or with family and friends	1 night or over
2	Domestic	Professional, business	1 night or over
3	Outbound	For leisure or with family and friends	1 night or over
4	Outbound	Professional, business	1 night or over

Figure 9: DIMTRAVEL Attributes

DIMTRAVEL consists of TRAVEL ID, REGION, PURPOSE and DURATION data in the expenditure dataset sourced from Eurostat. it is mapped with measure VALUE(EUROS) in fact table and can be drilled down by Region and purpose attributes illustrated in Figure 9. Duration attribute is just used as description and is not a category. the attributes from this dimension is used as part of analysis in BI query 2.

DIMENSION 6: DIMLOCATION, source: CSO:

DIMLOCATION consists of LOCATION ID and AREA OF RESIDENCE attributes in the TRIPS dataset sourced from CSO. it is mapped with TRIPS measure in fact table and can be filtered by the area of residence attribute. it is used as a part of analysis in business intelligence query 1.

FACT TABLE, source: contains measures from all the 4 sources:

The Fact table consists of primary key of all the dimensions, they are EMP ID, GENDER ID, LOCATION ID, SITE ID, TRAVEL ID, YEAR ID and are the foreign keys in the fact table. It also consists of measures with granularity of year and they are related to the dimension tables such as, EMPLOYED(PEOPLE) linked with DIMYEAR, DIMGENDER and DIMEMP SECTOR dimensions. VALUE(EUROS) linked with DIMYEAR, DIMGENDER and DIMTRAVEL dimensions. GDP in billion euros linked with DIMYEAR dimension. GDP per capita(EUROS) linked with DIMYEAR dimension. REVENUE(EUROS) linked with DIMYEAR dimension. VISITORS linked with DIMYEAR and DIMLOCATION. TRIPS linked with DIMYEAR and DIMSITES.

5 Logical Data Map

The logical data map is described in the below table, All the transformations made were done by using R programming.

Table 3: Logical Data Map describing all transformations, sources and destinations for all components of the data model illustrated in Figure 6

Source	Column	Destination	Column	Type	Transformation
1	AREA OF RESIDENCE	DIM_LOCATION	AREA OF RESIDENCE	Dimension	Special characters in few of the countries name have been removed.
1	TRIPS	FactTable	TRIPS	Fact	Value of trips in years not needed removed.
2	YEAR	DIMYEAR	YEAR	Dimension	Years with unwanted data removed.
2	EMPLOYMENT_SECTOR	DIMEMP_SECTOR	EMPLOYMENT_SECTOR	Dimension	Changed into factor from character data type and proper naming of the factor levels was done.
2	GENDER	DIMGENDER	GENDER	Dimension	GENDER - 'TOTAL' fields removed as it was not needed.
2	EMPLOYED (PEOPLE)	FactTable	EMPLOYED (PEOPLE)	Fact	Multiplied by 1000 as per the units mentioned in separate column in raw-table
2	PURPOSE	DIMTRAVEL	PURPOSE	Dimension	PURPOSE - 'TOTAL' fields removed as it was not needed
2	DURATION	DIMTRAVEL	DURATION	Dimension	Removed few unwanted fields.
2	REGION	DIMTRAVEL	REGION	Dimension	Removed 'All countries of the world' field as it was not needed.
2	VALUE (EUROS)	FactTable	VALUE (EUROS)	Fact	Merged the Units Column with Value(Euros) column by multiplying with 1000 units.
3	Attraction	DIMSITES	ATTRACTION	Dimension	No transformation made.
3	County	DIMSITES	COUNTY	Dimension	No transformation made.

Continued on next page

Table 3 – *Continued from previous page*

Source	Column	Destination	Column	Type	Transformation
3	Visitors	FactTable	VISITORS	Fact	Removed special characters and changed into numeric data type.
4	GDP per capita (EUROS)	FactTable	GDP per capita (EUROS)	Fact	Removed unwanted columns and rows of descriptions, Converted from dollars to Euros(1 \$ = 0.87 Euros).
4	GDP in billion euros	FactTable	GDP in billion euros	Fact	Removed unwanted columns and rows of descriptions, Converted from dollars to Euros(1 \$ = .87 Euros).
4	REVENUE (EUROS)	FactTable	REVENUE (EUROS)	Fact	Removed unwanted columns and rows of descriptions, Converted Revenue column in millions as per the units given in source by multiplying with 1000000.

6 ETL Process

6.1 Extraction:

Extraction is a process of gathering raw data from different sources and storing it in the ETL environment before any restructuring of data is carried out. Kimball et al. (2013) The same approach from Kimball is used in this project. The required raw structured and unstructured data from different sources are gathered. The structured data from different sources such as Statista, Eurostat and CSO were extracted in the formats of .CSV and .XLSX. The unstructured data is extracted from 6 online articles published by independent.ie and failteireland.ie by using the execute process task loaded with R code in SQL Server Integration Services.

6.2 Transform:

The Data gathered is then restructured to improve the quality of data. This is done by using execute process task loaded with R code in SSIS. The detailed description of the cleaning process done by using R code for each dataset from different sources is below:

Source 1: CSO: The data extracted from CSO in .XLS format was in a complex arrangement. The Column names were labelled appropriately, the Year column with unwanted data have been removed. Area of residence column has been changed into factor and the level names were labelled appropriately. The final data frame was written into a .CSV file.

Source 2: EUROSTAT: The Expenditure dataset in .CSV format was read in R; the column headers were labelled appropriately for better understanding and unwanted columns were removed. The data types of columns YEAR, GENDER, PURPOSE, REGION was changed into factors. The unwanted levels in DURATION, GENDER, PURPOSE and REGION were removed. VALUE(EUROS) column was merged with UNITS column by multiplying VALUE(EUROS) with 1000. The final data frame is the written into .CSV file loading. The Employment dataset in .CSV format was read in R; the Column names were changed into appropriate names. The columns EMPLOYMENT_SECTOR, YEAR, GENDER was changed into factor. The unwanted levels of factors were removed. The column EMPLOYED(PEOPLE) were merged with UNITS column by multiplying the column with 1000. The total values in GENDER column was removed. The levels of factor EMPLOYMENT_SECTOR was modified as required. The final data frame was written into .CSV file.

SOURCE 3: FAILTEIRELAND.IE & INDEPENDENT.IE aticle.: The unstructured data from the 6 article sources were web scraped using R. the data from the 6 articles were merged into columns ATTRACTION, VISITORS and YEAR. the special characters present in the values of all the columns were removed. The values in ATTRACTION was split into two columns as ATTRACTION & COUNTY to be used as drilldown for Business query. The columns ATTRACTION, COUNTY and YEAR were changed as factor. The final data frame was written into a .CSV file.

SOURCE 4: STATISTA: The GDP dataset from the source in .XLSX format was read in R and the unwanted descriptions and texts were removed. The column names were labelled appropriately. The column GDP in billion euros was changed to numeric data type and converted from dollars to euros by multiplying with 0.87. Column YEAR is changed into factor datatype. The GDP per capita dataset from the source in .XLSX format was read in R, unwanted columns are removed. Column GDP per capita(EUROS)

was changed to numeric and converted from dollars to euros. Column YEAR was changed to factor datatype. Both the columns of both the dataset were merged together and written as a single .CSV file read for loading. Revenue dataset from the source in .XLSX is read into R, the unwanted descriptions were removed, and the table was indexed to the required data of rows and columns. The column REVENUE(EUROS) were changed as numeric and multiplied by units 1000000. The column YEAR was changed into factor. The resulting data frame is written into .CSV file format.

6.3 LOAD

The cleaned and transformed data in the form of flat files are imported and stored into staging area in SSIS using the raw tables created in SQL server database. Each dataset are imported by using data flow task in SSIS, which is designed as illustrated in Figure 10, to avoid any duplication and update of existing records during multiple ETL runs.

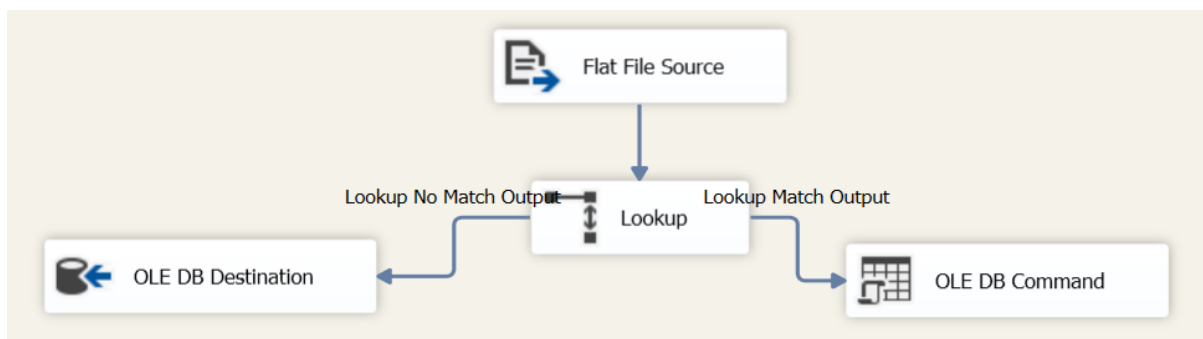


Figure 10: Data flow task structure

The Imported flat file source is processed to the OLE DB Destination (Raw table created in SQL query) using a lookup transformation, the lookup transformation uses the OLE DB destination as reference and sends the no match output to OLE DB destination. The data with rows which has match to the reference data is updated by using the SQL query with OLE DB command Sheldon (2012). The data loaded in the raw tables are then imported to dimension table using Execute SQL Task with the help of SQL query. The SQL query is used to create dimension tables and insert the appropriate data from raw tables into the specific dimension table created. the below sample DROP table query is used before the query for creating each dimensional table to avoid any duplication while multiple ETL runs.

```

IF OBJECT_ID(TABLENAME,'U') IS NOT NULL
DROP TABLE TABLENAME;
  
```

After loading the dimension table with the descriptive data, another execute SQL task is used to create Fact table and load the measures in to the fact table. The sql query used in the execute SQL task has an DROP table query to avoid any duplication of records.

```

IF OBJECT_ID('FACT_TABLE', 'U') IS NOT NULL
DROP TABLE FACT_TABLE;
  
```

The query then creates the fact table, inserting the measures from the raw table using inner join and adding the primary keys of all the dimensional table. The primary keys acts as foreign key referencing the dimensional tables in the fact table. The complete SQL queries are mentioned in the appendix. Cube deployment: The OLAP cube is created using the SQL server analysis service. The measures from the fact table and dimensions

from the dimension table are added to build the cube. During the process of building the cube, hierarchy of the dimensions can be specified. once the cube is build. it is processed successfully as illustrated in Figure 11

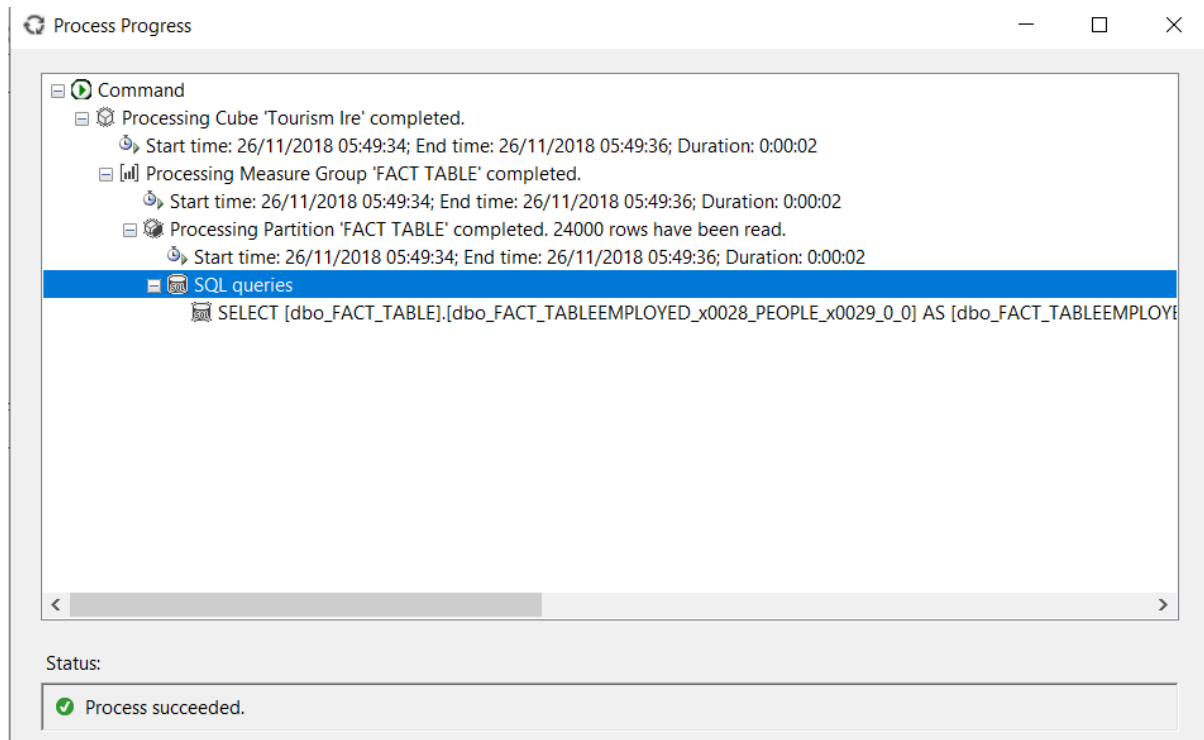


Figure 11: OLAP CUBE

After cube is deployed, the same is being automated in SSIS using Analysis services processing task. this done inside a sequence container where the deployment of cube is divided into tow analysis services processing task. one loaded with dimension tables and other handling the fact table.

7 Application

The business requirements mentioned in section1, are analyzed using business queries in this section. Tableau is used for visualization of the data.

7.1 BI Query 1: Countries contributing the most for Ireland tourism sector in the recent years.

The query is analyzed by comparing 4 countries with the most tourists in the world to Ireland and the revenue generated from tourism industry in the past 4 years. The data for the comparison has been extracted from CSO and Statista. The Figure 12 has been visualized using tableau and the number of Trips dataset sourced from CSO has been filtered down to the 4 countries with most visitors and is filtered for the recent 4 years of data with revenue dataset sourced from Statista visualized for the same year.

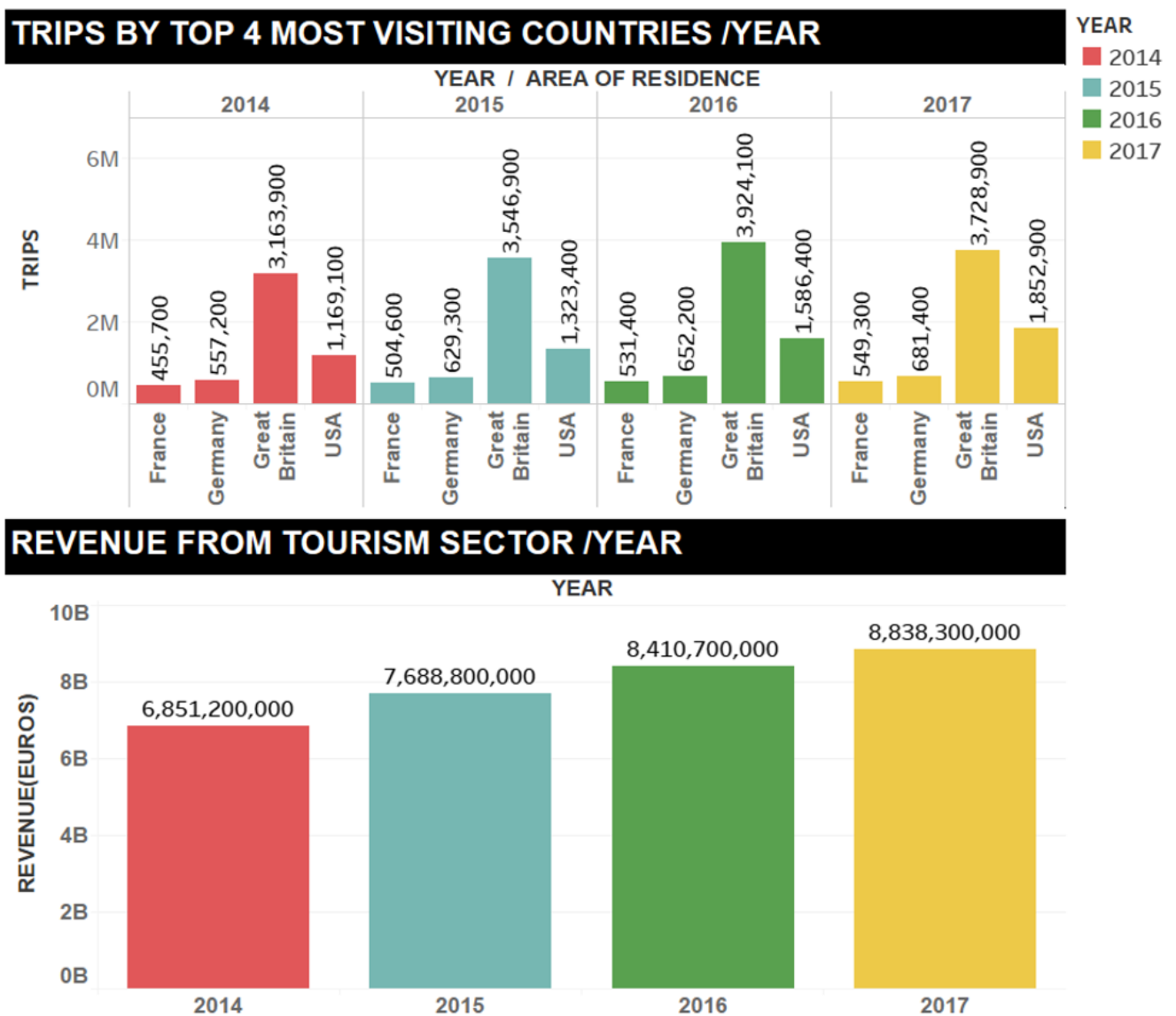


Figure 12: Visualization of Data for BI Query 1

By comparing the data illustrated in Figure 12, we see that tourists from Great Britain

have been the main source of revenue for tourism industries over the last 4 years but trips from Great Britain have declined in year 2017 compared to the continuous growth in the previous years. But this did not affect revenue of tourism in Ireland as the popularity of Ireland has increased among the tourists from other countries mainly USA, France and Germany. Specifically, USA travelers (16% increase) have compensated to Great Britains tourist decline to Ireland in 2017. The Ireland tourism sector should take the measures to attract and hold their popularity among Great Britain as it has been the major source of revenue in the previous years and also try to attract trips from far countries such as USA as it will directly contribute to Air tickets, longer accommodation stays and worldwide reach.hence further enhancing the opportunities in tourism sector.

7.2 BI Query 2: What type of tourists are contributing the most to Ireland's top visited attractions.

This Query is intended to find out the type of tourists contributing to the most visited sites in Ireland over the years 2013,2015 & 2017. To analyze and get insights about the query, the top 4 most visited paid attractions in Ireland sourced from Independent.ie fail-teIreland articles is filtered in tableau and compared with tourist expenditure in Ireland sourced from Eurostat for the same years. the expenditure is drilled down from Domestic and outbound to purpose and then further drilled down to gender classification.

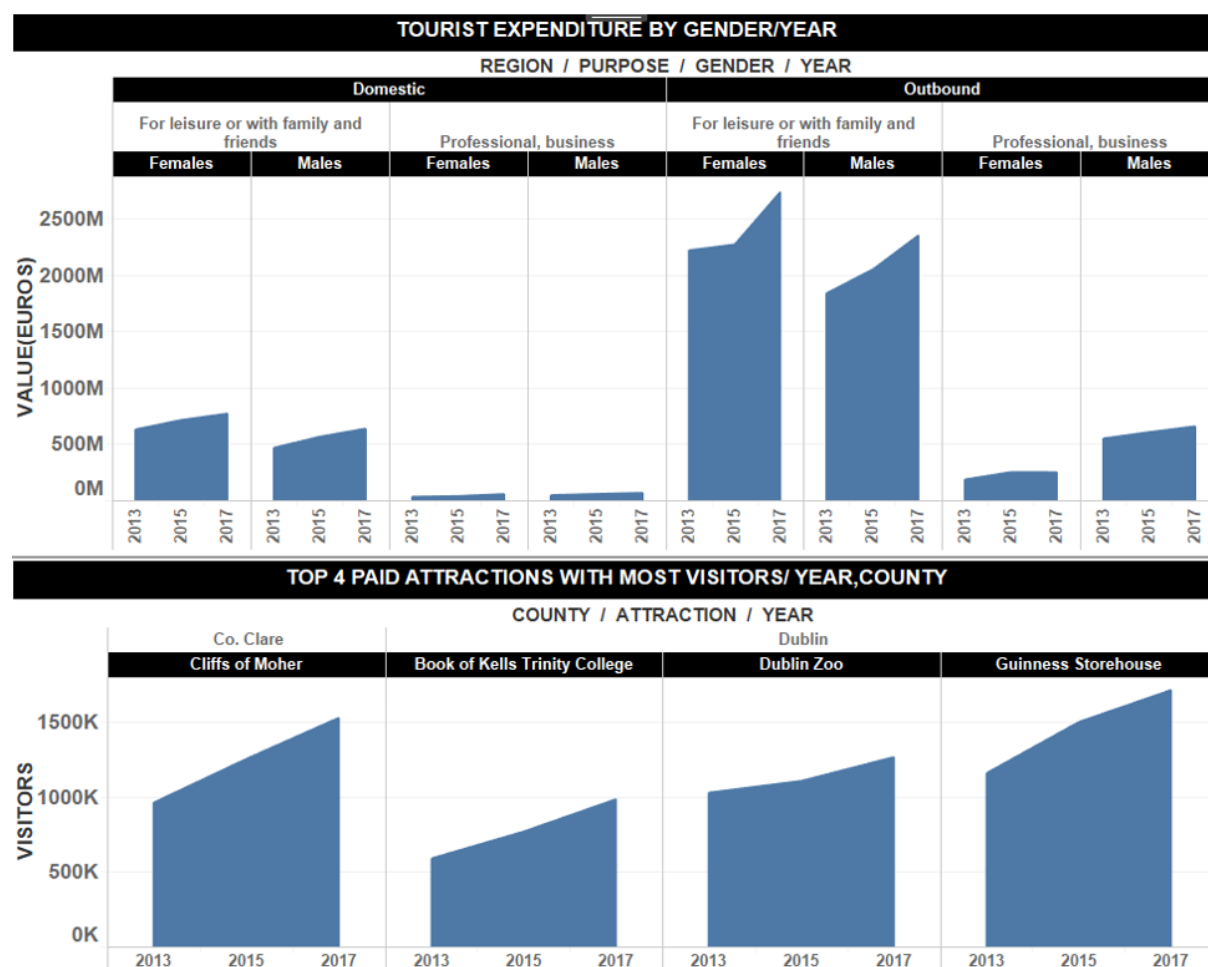


Figure 13: Visualization of Data for BI Query 2

From the graphical representation illustrated in Figure 13 designed using tableau, we can infer that outbound female tourist with purpose of leisure or with family and friends are spending the most in Ireland and must be spending in Dublin. As Dublin seems to be popular county the tourists spend as the county has 3 top paid attractions in Ireland of the top 4 attractions. It is also inferred that the people travelling to Ireland for professional purposes are very less compared to leisure purposes and are mostly males. Both the purpose travellers are preferring sites in Dublin, Only place out of Dublin from the top 4 for the last 5 years is Cliffs of Moher, of whichs popularity has over taken Dublin zoo and topped the 2nd place in recent years.

7.3 BI Query 3: Impact of growth in tourism to the people living in Ireland.

This query is to get insights on the impact of growth in tourism sector and what it means to the economy of the country. For this query, The employment in tourism sector dataset is sourced from Eurostat and GDP GDP per capita is sourced from Statista. The employment dataset is drilled down to Accommodation and food service activities and travel agency and 'All NACE Activities' activities of the total 5 activities, its further can be classified by gender into males ad females. The Data from both the sources are compared between years 2013,2015 & 2017.

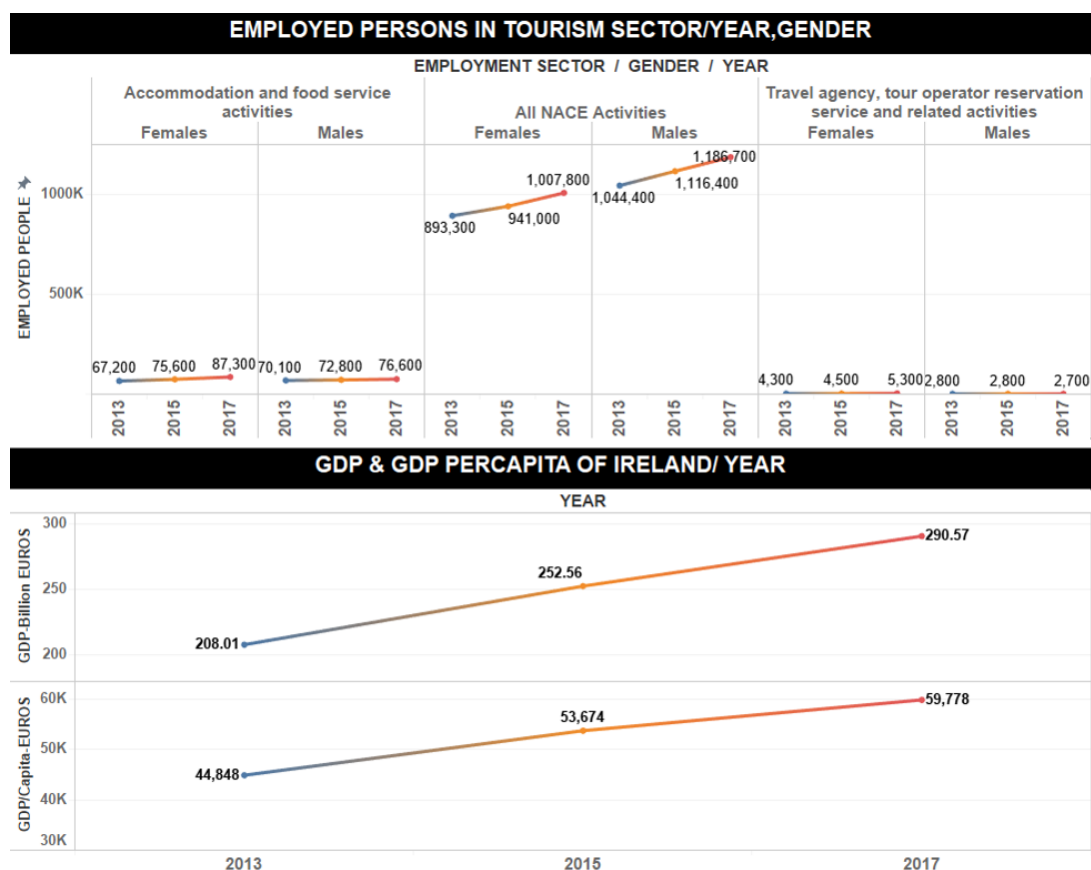


Figure 14: Visualization of Data for BI Query 3

From the visualization illustrated in Figure 14 designed in tableau, We can understand that females employed are rising in Accommodation and food service activities and also

in travel agencies related activities compared to male. But male employment is still at 10% high in All NACE activities compared to women in NACE activities. But women employment in All NACE activities is displaying a steeper increase in year 2017. comparing these positively increasing line in each year to parallel increasing GDP and GDP per capita from the figure. we can infer that impact of growth in tourism sector is greatly benefiting Ireland's GDP and GDP per capita resulting in better life standard of people and women employment is having a steeper rise in the recent year and looks to be growing over the years resulting additional source of income to households meaning increase in life standards of people.

7.4 Discussion

Analyzing the business query 1 derived from business requirement 1 in section 1, we see that decline in tourists from Great Britain did not impact on the revenue generated from tourism even though Great Britain have been the top tourist contributor to Ireland tourism. this is because of increase in tourists from USA and other European countries. It can be inferred that Britain, the United States, Germany, and France are the four major sources of market for tourism for the last 15 years comparing the data from BI query 1 and the literature reviewed published in 2003 stating the same top 4 visitor countries. O'Leary & Deegan (2005). But with the issue of Brexit and Great Britain's decline in tourists in Ireland. Irish tourism sector has to attract tourists from other regions and make use of the increase of tourists from USA to enhance world reach of Irish tourism.

Business query 2 derived from business requirement 2, clearly depicts that outbound female tourists visiting Ireland for leisure or with family and friends spend the most and due to the high concentration of attraction in Dublin topping the most paid attractions. they must be making the most expenses in Dublin. the other business insights to be taken from this query are Ireland's tourism sector must try expand the concentration of tourist to other counties, As this will create opportunities of more tourism revenue and a wider spread of tourism expenditure helping in development of attractions and life standards of people outside Dublin. the concentration of most expenditure by the purpose of leisure visitors is similar to that of literature Deegan & Dineen (2003) published 10 years ago but the difference in this query is finding the gender who are contributing to that expenditure and the sites which are generating those expenditure.

The visualization of Query 3 derived from business requirement 3 provides us insights on the increment in women employment in tourism sector in the recent years with steady increase in GDP and GDP per capita income. which shows that more people in households are getting a source of income which is increasing each year as per GDP per capita, resulting better life standards of the people. other insights to be taken are the increase in gender equality in employment in tourism sector.

8 Conclusion and Future Work

The business queries used in this project produced insights required by the business requirement. But there is improvement in the depth of the analysis that can be carried out. that is future work can be considered on the context of expenditure the outbound female tourists make in Dublin, prospects of improving the professional visitors to Ireland, the gender equality in employment of tourism sector to other industrial sectors in Ireland. Combining the business requirements, future work can be considered on USA tourists

expenditure in Ireland by gender, region and attractions. these future queries can help in enhancing tourism business.

References

- Abdulaziz, T. A., Moawad, I. F. & Abu-Alam, W. M. (2015), Building data warehouse system for the tourism sector, *in* ‘2015 IEEE Seventh International Conference on Intelligent Computing and Information Systems (ICICIS)’, pp. 410–417.
- Arslanturk, Y., Balcilar, M. & Ozdemir, Z. A. (2011), ‘Time-varying linkages between tourism receipts and economic growth in a small open economy’, *Economic Modelling* **28**(1), 664 – 671.
URL: <http://www.sciencedirect.com/science/article/pii/S0264999310001008>
- Chou, M. (2013), ‘Does tourism development promote economic growth in transition countries? a panel data analysis’, *Economic Modelling* **33**, 226–232. cited By 48.
- Deegan, J. & Dineen, D. J. (2003), ‘The changing contribution of tourism in a dynamic economy: The case of ireland’, *Tourism Economics* **9**(2), 147–164.
URL: <https://doi.org/10.5367/000000003101298321>
- Kimball, R., Ross, M. & Kimball, R. (2013), *The data warehouse toolkit : the definitive guide to dimensional modeling : [electronic book].*, Indianapolis, Ind. : John Wiley, 2013.
URL: <http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,cookie,shibdb=cat05743aAN=nci.31479site=eds-livescope=sitcustid=ncirlib>
- OLeary, S. & Deegan, J. (2005), ‘Irelands image as a tourism destination in france: Attribute importance and performance’, *Journal of Travel Research* **43**(3), 247–256.
URL: <https://doi.org/10.1177/0047287504272025>
- Sheldon, R. (2012), ‘Implementing Lookup Logic in SQL Server Integration Services’. [Online; April 26,2012].
URL: <https://www.red-gate.com/simple-talk/sql/ssis/implementing-lookup-logic-in-sql-server-integration-services/>
- wttc.org (2018), ‘Travel and tourism economic impact 2018 ireland’.
URL: <https://www.wttc.org/-/media/files/reports/economic-impact-research/countries-2018/ireland2018.pdf>

Appendix

R Code

```
library(readr)
library(rvest)
library(stringr)
```



```

library(dplyr)
library(tidyr)
library(readxl)

#Cleaning Expenditure.CSV file
Exp <- read_csv("C:/Masters/#DWBI/DW&BI_Project/Tourism_in_ireland/
Eurostat-Employment&/Expenditure_for_ireland/Expenditure_by_sex(S)/
tour_dem_exsex_1_Data.csv",
               col_types = cols('Flag and Footnotes' = col_skip(),
                                GEO = col_skip(), Value = col_number()))
ABC <- c("TIME", "SEX", "PURPOSE", "DURATION", "PARTNER")
Exp[ABC] <- lapply(Exp[ABC], factor)
cols <- c("YEAR", "GENDER", "UNIT", 'PURPOSE', 'DURATION', 'REGION',
'VALUE_IN_EUROS')
colnames(Exp) <- cols
#Merging the Units Column with Value in Euros column by multiplying
with 1000 units
Exp$'VALUE IN EUROS' <- Exp$'VALUE IN EUROS' * 1000
Exp$UNIT <- NULL
Exp <- Exp[-grep("4_nights_or_over", Exp$DURATION, perl=T),]
Exp <- Exp[-grep("From_1_to_3_nights", Exp$DURATION, perl=T),]
Exp <- Exp[-grep("Total", Exp$GENDER, perl=T),]
Exp <- Exp[-grep("Total", Exp$PURPOSE, perl=T),]
Exp <- Exp[-grep("All_countries_of_the_world", Exp$REGION, perl=T),]
names(Exp)[6] <- "VALUE(EUROS)"
write_csv(Exp, file='C:/Masters/#DWBIPROJECT/Expenditure.csv', row.names
= FALSE)

#Cleaning Employment.csv file
Emp <- read_csv("C:/Masters/#DWBI/DW&BI_Project/Tourism_in_ireland/
Eurostat-Employment&/Employment_for_ireland/Employed_persons_and_employees
by_sex_and_full-timepart-time_activity_and_Nace_Rev._2_activity/
Employment.csv",
               col_types = cols('Flag and Footnotes' = col_skip(),
                                GEO = col_skip(), Value = col_number()))
FAC <- c("TIME", 'WORKTIME', 'SEX', 'NACE_R2', 'WSTATUS')
Emp[FAC] <- lapply(Emp[FAC], factor)
Emp <- Emp[-grep("Full-time", Emp$WORKTIME, perl=T),]
Emp <- Emp[-grep("No_response", Emp$WORKTIME, perl=T),]
Emp <- Emp[-grep("Part-time", Emp$WORKTIME, perl=T),]
Emp <- Emp[-grep("Employees", Emp$WSTATUS, perl=T),]
Emp$UNIT <- NULL
Emp$WORKTIME <- NULL
Emp$WSTATUS <- NULL
Emp$Value <- Emp$Value * 1000
colnames(Emp) <- c("EMPLOYMENT_SECTOR", "YEAR", "GENDER", "EMPLOYED(PEOPLE)")
Emp <- Emp[-grep("Total", Emp$GENDER, perl=T),]
levels(Emp$EMPLOYMENT_SECTOR)[4] <- 'All_NACE_Activities'
write_csv(Emp, file = 'C:/Masters/#DWBIPROJECT/Rawemp.csv', row.names
= FALSE)

```

```

#Cleaning Revenue.csv file
#install.packages("readxl")
Rev <- read_excel("C:/Masters/#DWBI/DW&BI_Project/Tourism_in_ireland/
Statista_-Revenue/Revenue.xlsx",sheet = 2)
#Removing unwanted columns and rows
Rev <- Rev[3:8,1:2]
R <- c("YEAR","REVENUE(EUROS)")
colnames(Rev) <- R
Rev$'REVENUE(EUROS)'<- as.numeric(Rev$'REVENUE(EUROS)')
Rev$YEAR <- as.factor(Rev$YEAR)
# Converting Revenue column in millions as per the units given in source
by * with 1000000
Rev$'REVENUE(EUROS)' <- Rev$'REVENUE(EUROS)' * 1000000
write.csv(Rev, file = 'C:/Masters/#DWBIPROJECT/Revenue.csv', row.names =
FALSE)

#TOP5sites - webscrapping texts from News aritcle Independent.ie & failte
ireland using R

#Top 5 visited sites in ireland - 2017 - Independent.ie
url <-read_html("https://www.independent.ie/life/travel/travel-news/
revealed-irelands-top-20-tourist-attractions-paid-and-free-36986059.html")
Top <- url %>% html_nodes("ol")
Top <- Top[1]
Top <- html_text(Top)
Places <- c(substr(Top,1, 27),substr(Top,40,65),substr(Top,78,87),
substr(Top,100,130),substr(Top,143,181))
Visitors <- c(substr(Top,31,39),substr(Top,69,77),substr(Top,91,99),
substr(Top,134,142),substr(Top,185,191))
Year <- c('2017')
Visitors <- gsub("[^0-9]", "", Visitors)
Visitors <- as.numeric(Visitors)
Top2017 <- data.frame(Places,Visitors,Year)

#Top 5 visited sites in ireland - 2016 - Failteireland
url <-read_html("http://www.failteireland.ie/Utility/News-Library/
Ireland%E2%80%99s-Most-Popular-Visitor-Attractions-Reveale.aspx")
Top2016 <- url %>% html_nodes("ul")
Top2016 <- html_text(Top2016)
Top2016 <- Top2016[5]
Visitors <- c(substr(Top2016,27,35),substr(Top2016,87,95),
substr(Top2016,123,131),substr(Top2016,171,179),substr(Top2016,211,217))
Visitors <- gsub("[^0-9]", "", Visitors)
Visitors <- as.numeric(Visitors)
Year <- c('2016')
Top2016 <- data.frame(Places,Visitors,Year)

#Top 5 visited sites in ireland - 2015 - Independent.ie
url <-read_html("https://www.independent.ie/life/travel/ireland/
irelands-top-visitor-attractions-revealed-and-the-numbers-are-booming-
34947131.html")

```

```

Top <- url %>% html_nodes("ol")
Top <- Top[1]
Top <- html_text(Top)
Visitors <- c(substr(Top,31,39),substr(Top,77,85),substr(Top,99,107),
substr(Top,142,148),substr(Top,173,179))
Year <- c('2015')
Visitors <- gsub("[^0-9]", "", Visitors)
Visitors <- as.numeric(Visitors)
Top2015 <- data.frame(Places,Visitors,Year)

#Top 5 visited sites in ireland - 2013 - Independent.ie
url <-read_html("https://www.independent.ie/life/travel/ireland/irelands-
top-10-tourist-attractions-revealed--30397521.html")
Top <- url %>% html_nodes(".p402_hide")
Top <- Top[2]
Top <- html_text(Top)
Top <- str_replace_all(Top, "[^[:alnum:]]", "_")
Visitors <- c(substr(Top,448,456),substr(Top,533,539),substr(Top,473,481),
substr(Top,577,583),substr(Top,611,618))
Visitors <- gsub("[^0-9]", "", Visitors)
Visitors <- as.numeric(Visitors)
Year <- c('2013')
Top2013 <- data.frame(Places,Visitors,Year)

#Top 5 visited sites in ireland - 2014 - Independent.ie
url <-read_html("https://www.independent.ie/life/travel/ireland/irelands-
top-tourist-attractions-revealed-31370553.html")
Top <- url %>% html_nodes(".p402_hide")
Top <- Top[2]
Top <- html_text(Top)
Top <- str_replace_all(Top, "[^[:alnum:]]", "_")
Visitors <- c(substr(Top,570,578),substr(Top,635,643),substr(Top,669,677),
substr(Top,724,730),substr(Top,766,772))
Visitors <- gsub("[^0-9]", "", Visitors)
Visitors <- as.numeric(Visitors)
Year <- c('2014')
Top2014 <- data.frame(Places,Visitors,Year)

#Top 5 visited sites in ireland - 2012 - Failteireland
url <-read_html("http://www.failteireland.ie/Footer/Media-Centre/Ireland
-s-most-popular-tourist-attractions-for-201.aspx")
Top2012 <- url %>% html_nodes("p")
Top2012 <- html_text(Top2012)
Top2012 <- Top2012[15]
Top2012 <- str_replace_all(Top2012, "[^[:alnum:]]", "_")
Visitors <- c(substr(Top2012,59,67),substr(Top2012,199,205),substr
(Top2012,132,140),substr(Top2012,267,273),substr(Top2012,340,346))
Visitors <- gsub("[^0-9]", "", Visitors)
Visitors <- as.numeric(Visitors)
Year <- c('2012')
Top2012 <- data.frame(Places,Visitors,Year)

```

```

Top <- rbind(Top2016,Top2017,Top2013,Top2014,Top2012,Top2015)
Top$Places <- str_replace_all(Top$Places, "[()]", "")
Top <- Top %>% separate(Places, c("Attraction", "County"), ",")
Top$County <- as.character(Top$County)
Top$County <- ifelse(is.na(Top$County),
                    'Dublin', Top$County)
Top$County <- as.factor(Top$County)
Top$Attraction <- as.factor(Top$Attraction)
write.csv(Top, file = 'C:/Masters/#DWBIPROJECT/Top5sites.csv')

#GDP & GDP per capita income
GDPCI <- read_excel("C:/Masters/#DWBI/DW&BI_Project/Tourism_in_ireland/
Statista_Revenue/GDP_per_capita_income.xlsx",sheet = 2)
GDP <- read_excel("C:/Masters/#DWBI/DW&BI_Project/Tourism_in_ireland/
Statista_Revenue/GDP.xlsx",sheet = 2)
#Removing unwanted columns and rows for GDPCI
GDPCI <- GDPCI[3:8,1:2]
colnames(GDPCI) <- c("YEAR","GDP_per_capita(EUROS)")
GDPCI$'GDP per capita(EUROS)'<- as.numeric(GDPCI$'GDP per capita(EUROS)')
GDPCI$YEAR <- as.factor(GDPCI$YEAR)
# converting $ to Euros( 1$ = 0.87 Euros as of 08-11-2018)
GDPCI$'GDP per capita(EUROS)' <- GDPCI$'GDP per capita(EUROS)' * 0.87
#Removing unwanted columns and rows for GDP
GDP <- GDP[3:8,1:2]
colnames(GDP) <- c("YEAR","GDP_in_billion_euros")
GDP$'GDP in billion euros' <- as.numeric(GDP$'GDP in billion euros')
GDP$YEAR <- as.factor(GDP$YEAR)
# converting $ to Euros( 1$ = 0.87 Euros as of 08-11-2018)
GDP$'GDP in billion euros' <- GDP$'GDP in billion euros' * 0.87
TGDP <- cbind.data.frame(GDPCI,GDP$'GDP in billion euros')
colnames(TGDP) <- c("YEAR","GDP_per_capita(EUROS)","GDP_in_billion_euros")
write.csv(TGDP, file = 'C:/Masters/#DWBIPROJECT/TotalGDP.csv')

#TRIPS to Ireland - CSO
TRIP <- read_excel("C:/Masters/#DWBI/DW&BI_Project/Tourism_in_ireland/
CSO_Trips_to_ireland/OT2018M01TBL7.xls", sheet = 1)
colnames(TRIP) <- c("AREA_OF_RESIDENCE","2010","2011","2012","2013","2014",
"2015","2016","2017")
TRIP$'2010' <- NULL
TRIP$'2011' <- NULL
TRIP <- TRIP[-c(1, 22, 23, 24, 25, 26, 27), ]
AREA <- TRIP[,1]
YEAR <- c('2012','2013','2014','2015','2016','2017')
value2012 <- TRIP[,2]
value2013 <- TRIP[,3]
value2014 <- TRIP[,4]
value2015 <- TRIP[,5]
value2016 <- TRIP[,6]
value2017 <- TRIP[,7]
TT12 <- data.frame(AREA,'2012',value2012)
colnames(TT12) <- c("AREA_OF_RESIDENCE","YEAR","TRIPS")

```

```

TT13 <- data.frame(AREA,'2013',value2013)
colnames(TT13) <- c("AREA_OF_RESIDENCE","YEAR","TRIPS")
TT14 <- data.frame(AREA,'2014',value2014)
colnames(TT14) <- c("AREA_OF_RESIDENCE","YEAR","TRIPS")
TT15 <- data.frame(AREA,'2015',value2015)
colnames(TT15) <- c("AREA_OF_RESIDENCE","YEAR","TRIPS")
TT16 <- data.frame(AREA,'2016',value2016)
colnames(TT16) <- c("AREA_OF_RESIDENCE","YEAR","TRIPS")
TT17 <- data.frame(AREA,'2017',value2017)
colnames(TT17) <- c("AREA_OF_RESIDENCE","YEAR","TRIPS")
TRIP <- rbind(TT12,TT13,TT14,TT15,TT16,TT17)
TRIP$'AREA OF RESIDENCE' <- as.factor(TRIP$'AREA OF RESIDENCE')
K <- c(1,2,3,7,15)
levels(TRIP$'AREA OF RESIDENCE')[K] <- c("Africa","Asia","Australia,
New_Zealand_and_Other_Oceania","Central,South_and_other_America",
"Other_Europe")
write.csv(TRIP, file = 'C:/Masters/#DWBIPROJECT/TRIPS.csv' )

#####SQL QUERY#####
#for dimensional table creation
IF OBJECT_ID('FACT_TABLE', 'U') IS NOT NULL
    DROP TABLE FACT_TABLE;

IF OBJECT_ID('DIMYEAR', 'U') IS NOT NULL
    DROP TABLE DIMYEAR;

CREATE TABLE [dbo].[DIMYEAR](
[YEAR_ID] int identity(1,1) NOT NULL PRIMARY KEY,
[YEAR] [smallint] NULL )

ON [PRIMARY]
GO

INSERT INTO [dbo].[DIMYEAR](
    [YEAR])
Select distinct YEAR from RAWEMP;

IF OBJECT_ID('DIMLOCATION', 'U') IS NOT NULL
    DROP TABLE DIMLOCATION;

CREATE TABLE [dbo].[DIMLOCATION](
[LOCATION_ID] int identity(1,1) NOT NULL PRIMARY KEY,
[AREA OF RESIDENCE] [varchar](40) NULL )

ON [PRIMARY]
GO

INSERT INTO [dbo].[DIMLOCATION](
    [AREA OF RESIDENCE])
Select distinct [AREA OF RESIDENCE] from RAWTRIPS;

```

```
IF OBJECT_ID('DIMSITES', 'U') IS NOT NULL
    DROP TABLE DIMSITES;
```

```
CREATE TABLE [dbo].[DIMSITES](
    [SITE_ID] int identity(1,1) NOT NULL PRIMARY KEY,
    [ATTRACTION] [varchar](29) NULL,
    [COUNTY] [varchar](10) NULL)
ON [PRIMARY]
GO
```

```
INSERT INTO [dbo].[DIMSITES](
    [ATTRACTION] ,[COUNTY])
Select DISTINCT [ATTRACTION] ,[COUNTY] from RAWSITES;
```

```
IF OBJECT_ID('DIMGENDER', 'U') IS NOT NULL
    DROP TABLE DIMGENDER;
```

```
CREATE TABLE [dbo].[DIMGENDER](
    [GENDER_ID] int identity(1,1) NOT NULL PRIMARY KEY,
    [GENDER] [varchar](50) NULL )
ON [PRIMARY]
GO
```

```
INSERT INTO [dbo].[DIMGENDER](
    [GENDER])
Select distinct GENDER from RAWEMP;
```

```
IF OBJECT_ID('DIMTRAVEL', 'U') IS NOT NULL
    DROP TABLE DIMTRAVEL;
```

```
CREATE TABLE [dbo].[DIMTRAVEL](
    [TRAVEL_ID] int identity(1,1) NOT NULL PRIMARY KEY,
    [REGION] [varchar](50) NULL,
    [PURPOSE] [varchar](50) NULL,
    [DURATION] [varchar](50) NULL)
ON [PRIMARY]
GO
```

```
INSERT INTO [dbo].[DIMTRAVEL](
    [REGION],[PURPOSE],[DURATION])
Select distinct [REGION],[PURPOSE],[DURATION] from RAWEXP;
```

```
IF OBJECT_ID('DIMEMP_SECTOR', 'U') IS NOT NULL
    DROP TABLE DIMEMP_SECTOR;
```

```

CREATE TABLE [dbo].[DIMEMP_SECTOR](
    [EMP_ID] [int] identity (1,1) NOT NULL PRIMARY KEY,
    [EMPLOYMENT_SECTOR] [varchar](71) NULL)

ON [PRIMARY]
GO

INSERT INTO [dbo].[DIMEMP_SECTOR](
    [EMPLOYMENT_SECTOR])
Select distinct [EMPLOYMENT_SECTOR] from RAWEMP;

#for fact table creation
USE [TourismIre]
GO

IF OBJECT_ID('FACT_TABLE', 'U') IS NOT NULL
    DROP TABLE FACT_TABLE;

/***** Object:  Table [dbo].[DIMEMP_SECTOR]      Script Date:
13/11/2018 18:31:58 *****/
SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

CREATE TABLE [dbo].[FACT_TABLE](
    [EMP_ID] [int] NOT NULL FOREIGN KEY REFERENCES
    [dbo].[DIMEMP_SECTOR](EMP_ID),
    [GENDER_ID] [int] NOT NULL FOREIGN KEY REFERENCES
    DIMGENDER(GENDER_ID),
    [LOCATION_ID] [int] NOT NULL FOREIGN KEY REFERENCES
    DIMLOCATION(LOCATION_ID),
    [SITE_ID] [int] NOT NULL FOREIGN KEY REFERENCES
    DIMSITES(SITE_ID),
    [TRAVEL_ID] [int] NOT NULL FOREIGN KEY REFERENCES
    DIMTRAVEL(TRAVEL_ID),
    [YEAR_ID] [int] NOT NULL FOREIGN KEY REFERENCES
    DIMYEAR(YEAR_ID),
    [EMPLOYED(PEOPLE)] [int] NULL,
    [VALUE(EUROS)] [numeric](18, 2) NULL,
    [GDP in billion euros] [real] NULL,
    [GDP per capita(EUROS)] [real] NULL,
    [REVENUE(EUROS)] [numeric](18, 2) NULL,
    [VISITORS] [int] NULL,
    [TRIPS] [int] NULL
) ON [PRIMARY]
GO

```

```

INSERT INTO [dbo].[FACT_TABLE]
    ([EMP_ID]
    ,[GENDER_ID]
        ,[LOCATION_ID]
        ,[SITE_ID]
    ,[TRAVEL_ID]
    ,[YEAR_ID]
    ,[EMPLOYED(PEOPLE)]
    ,[VALUE(EUROS)]
    ,[GDP in billion euros]
    ,[GDP per capita(EUROS)]
    ,[REVENUE(EUROS)]
    ,[VISITORS]
        ,[TRIPS])

    Select distinct j.[EMP_ID]
    ,[GENDER_ID]
        ,[LOCATION_ID]
        ,l.[SITE_ID]
    ,[TRAVEL_ID]
    ,[YEAR_ID]
    ,[EMPLOYED(PEOPLE)]
    ,[VALUE(EUROS)]
    ,[GDP in billion euros]
    ,[GDP per capita(EUROS)]
    ,[REVENUE(EUROS)]
    ,[VISITORS]
        ,[TRIPS]

    From [dbo].[RAWEXP] a
        join [dbo].[RAWEMP] b
            on a.YEAR = b.YEAR and a.GENDER = b.GENDER
        join [dbo].[RAWGDP] c
            on a.YEAR = c.YEAR
        join [dbo].[RAWREV] d
            on a.YEAR = d.YEAR
        Join [dbo].[RAWSITES] e
            on a.YEAR = e.YEAR
        Join [dbo].[RAWTRIPS] f
            on a.YEAR = f.YEAR
        join [dbo].[DIMGENDER] g
            on a.GENDER = g.GENDER
        join [dbo].[DIMTRAVEL] h
            on a.REGION = h.REGION and h.PURPOSE = a.PURPOSE
        join [dbo].[DIMYEAR] i
            on a.YEAR = i.YEAR
        join [dbo].[DIMEMP_SECTOR] j
            on b.EMPLOYMENT_SECTOR = j.EMPLOYMENT_SECTOR
        join [dbo].[DIMLOCATION] k
            on f.[AREA OF RESIDENCE] = k.[AREA OF RESIDENCE]
        join [dbo].[DIMSITES] l

```



```
on l.ATTRACTION = e.ATTRACTION
join[dbo].[DIMSITES] m
on m.COUNTY = e.COUNTY
GO
```