National College of Ireland

Data Warehousing and Business Intelligence Project

on

Quality of Life

Zainul Abedin Khatik
x18110096

MSc/PGDip Data Analytics – 2018/9

Submitted to: Dr. Simon Caton

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

Student Name:	Zainul Abedin Khatik
Student ID:	x18110096
Programme:	MSc Data Analytics
Year:	2018/9
Module:	Data Warehousing and Business Intelligence
Lecturer:	Dr. Simon Caton
Submission Due Date:	26/11/2018
Project Title:	Quality of Life

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 LaTeX 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

Quality of Life

Zainul Abedin Khatik
x18110096

November 26, 2018

Abstract

Quality of life is a multidimensional concept which consists of various categories such as well-being, happiness,economic indices but a few of the components have a broader concept of quality of life, but no concept can record accurately complexity of life. This project aims to compare different parameters to measure quality of life.

1 Introduction

How should we measure the progress of any country? What parameters should be looked into while judging the quality of life. Even though the measuring of quality of life is really difficult, it should have some clarity on what parameters it is heavily dependent on. GDP remains undeniably the important parameter to measure economic results, but it is really essential it should complement it with other indicators such as happiness, Cost of living, Health care index and other parameters. After the recession of 2008 researchers have looked into other indicators to judge the quality of life. Nobel Prize-winning economists Joseph Stiglitz and Amartya Sen and their team have criticized GDP as a prime indicator of progress (Jeroen, V. (2017)) -Jeroen (2017).In business query 2 Section 7 I will be comparing the happiness, quality of life and gdp are correlated from the data I have collected. Also, Banerjee (2016) puts forward that for a nation to strengthen terms of development, the nation should assess its GDP growth, quality of life of its citizen and human happiness. The study shows that there is a high correlation between macroeconomic development parameters such as education, health, quality of life index and GDP growth rate, expressed through the Human Development Index. In business query 2 Section 7 I will be comparing the happiness, quality of life and GDP are correlated from the data I have collected. Lastly, Education is key component of individual well-being. The author aims to presents different ways to analyze the quality of life and to draw special attention the link between the education, quality of life and employment rate in Romania and in general. Studies shows that attaining a high education qualification level increases to get a job and offers an individual to get higher salary. Also, high level of education decreases the risk of falling into poverty, precisely because people who invest in education find that to get a job it is easier and have good stability in the labour market - Aceleanu (2012). The organisation suhc as OECD (Better Life Index), Eurofound (Quality of Life Survey), The Economist Intelligence Unit (Quality of life index),European Bank for Reconstruction and Development (Report of Life Satisfaction), Legatum Institute London, UK (The 2011 Legatum Prosperity Index), EU Commission

(Wellbeing-Aggregate Report) have their own method to calcualte the quality of life, I have taken some of the parameters which are highly dependent on quality of life.

(Req-1) my first requirement - I have compared GDP, Cost of living and Health care index(Both of the parameters of the Quality of Life indicators dataset).

(Req-2) my second requirement - I will be comparing the happiness, quality of life and GDP are correlated from the data I have collected.

(Req-3) my third requirement - I will be comparing the education expenditure, employment rate, GDP, quality of life index and quality of education in top universities

2 Data Sources

Here you should present and formally describe your sources of data used in the project.

2.1 Source 1: Statista

The structured dataset is sourced from Statista : `https://www.statista.com/statistics/268127/employment-rate-in-selected-industrialized-countries/` which provides 2 columns of information about employment rate on 36 countries. This dataset consists of employment rate in OCED (Organization for Economic Co-operation and Development) countries in the year 2017.

These 2 columns are names of the 36 countries and their employment rate in percentage

Both the columns in the dataset are relevant to this project.

This dataset addresses the business requirement listed in Section 1, it will be used to answer the 3rd BI query to analyze that how employment rate, GDP, quality of life index and quality of education.

2.2 Source 2: World Happiness Report

This unstructured dataset is sourced from: `http://worldhappiness.report/download/` which provides 3 columns of information about the happiness rankings of 156 countries. I have taken 4 years of data i.e. 2017, 2016, 2015 and 2014.

1. 2017: `https://s3.amazonaws.com/happiness-report/2017/HR17.pdf` - page 20 (table 1)

2. 2016: `https://s3.amazonaws.com/happiness-report/2016/HR-V1_web.pdf` - page 20 (table 1)

3. 2015: https://s3.amazonaws.com/happiness-report/2015/WHR15_Sep15.pdf - page 26 (table 1)

4. 2013: https://s3.amazonaws.com/happiness-report/2013/WorldHappinessReport2013_online.pdf - Page 22 (table 1)

These 3 columns are the rank of the country, country name, and the happiness score.

All the 3 columns are relevant to my project. I have considered first 50 countries. The 156 countries consist of 3 pdfs, I have considered the first the pdf.

This dataset addresses the business requirement listed in Section 1, it will be used to answer the second query to analyze that how happiness, GDP and quality of index are correlated.

2.3 Source 3: Numbeo

This semi-structured dataset is sourced from: https://www.numbeo.com/quality-of-life/rankings_by_country.jspwhich provides 11 columns of information about the quality of life indicators on 66 countries. I have taken 5 years of data i.e. 2017, 2016, 2015, 2014 and 2013.

Link of 5 years of data:

1. 2017: https://www.numbeo.com/quality-of-life/rankings_by_country.jsp?title=2017
2. 2016: https://www.numbeo.com/quality-of-life/rankings_by_country.jsp?title=2016
3. 2015: https://www.numbeo.com/quality-of-life/rankings_by_country.jsp?title=2015
4. 2014: https://www.numbeo.com/quality-of-life/rankings_by_country.jsp?title=2014
5. 2013: https://www.numbeo.com/quality-of-life/rankings_by_country.jsp?title=2013

These columns are rank of the country, name of the country, quality of life index, purchasing power index, safety index, health care index, cost of living index, property price to income ratio, traffic commute time, index pollution, index climate index.

I have used rank of the country, name of the country, quality of life index and cost of living columns to build business requirement query.

This dataset addresses the business requirement listed in Section 1, it will be used to answer the first BI query, second BI query and third BI query. In the first BI query, it will be used to compare and analyze GDP and Cost of living and Health care index

(Both of the parameters of the Quality of Life indicators dataset).
In the second query, it will be used to analyze that how happiness, GDP and quality of index are correlated.
In the third BI query, it will be used to compare the education expenditure, employment rate, GDP, quality of life index and the quality of education.

2.4 Source 4: 2.4.1 The World bank

The structured dataset is sourced from:
`https://data.worldbank.org/indicator/NY.GDP.PCAP.CD` which provides 62 columns of information about the GDP per capita on 265 countries

These columns represent the year, which starts from the year 1960 to 2017

I have taken 5 years of data i.e. 2017,2016, 2015, 2014, and 2013.

This dataset addresses the business requirement listed in Section 1, it will be used to answer the first BI query and second BI query.

2.5 Source 4: 2.4.2 The World bank

The structured dataset is sourced from: `https://data.worldbank.org/indicator/SE.XPD.TOTL.GD.ZS` which provides on 265 countries

These columns represent the year, which starts from the year 1960 to 2017.

I have taken 5 years of data i.e. 2017,2016, 2015, 2014, and 2013.

This dataset addresses the business requirement listed in Section 1, it will be used to compare the education expenditure, employment rate, GDP, quality of life index and quality of education.

2.6 Source 5: Center for World University Rankings (CWUR)

The semi-structured dataset is sourced from:
`https://cwur.org/` which provides 13 columns of information on 1000 university rankings. I have taken 5 years of data i.e. 2017, 2016, 2015, 2014 and. 2013.

Link of 5 years of data:

1. 2017: `https://cwur.org/2017.php`
2. 2016: `https://cwur.org/2016.php`
3. 2015:`https://cwur.org/2015.php`
4. 2014:`https://cwur.org/2014.php`

5. 2013:https://cwur.org/2013.php

These columns are world rank, institution, location, national rank, quality of education, alumni employment, quality of faculty, publications, influence, citations, broad impact, patents and score.

I have used quality of education to generate a build business requirement query.

This dataset addresses the business requirement listed in 1, it will be used to compare the education expenditure, employment rate, GDP, quality of life index and quality of education

2.7 Source 6: Github

The structured data source is sourced from:
https://github.com/lukes/ISO-3166-Countries-with-Regional-Codes/blob/master/all/all.csv which provides 11 columns of information about countries with region and sub region on 250 countries.

These columns are name of the country, alpha-2, alpha-3, country code, iso 3166-2, region sub-region, intermediate-region, region-code, sub-region-code and intermediate-region-code

I have used the name of country, region and sub-region. This dataset addresses the business requirement listed in 1, will help to build a hierarchy
Region to Sub region to Country.

Source	Type	Brief Summary
Statista	Structured	Employment rate in OCED countries in year 2017
World Happiness Report	Unstructured	Happiness Ranking and Score Country Wise (4 years of data)
Numbeo.com	Semi- Structured	Quality of life indicators (Last 5 years of data)
World Bank	Structured	GDP country wise (Last 5 years of data)
World Bank	Structured	Education Expenditure (Last 5 years of data)
Center for World University Ranking	Semi- Structured	World University Ranking of 1000 institution.
Github	Structured	All the country data, region and sub-region wise.

Summary of sources of data used in the project

3 Related Work

1. For more than a half century, the most widely accepted measure of a countrys economic progress is measured with respect to its Gross Domestic Product (GDP). Lately, concerns have emerged that macroeconomic statistics such as GDP, don't give an adequate point by point image of the living conditions that ordinary people experience. Economic progress, environmental progress and social progress of the country do not go hand in hand with the Gross Domestic Product (GDP) (Walter, R. (2015)) Walter (2015)
   Robert F. Kennedy once said that a countrys gross domestic product (GDP) measures everything except that which makes life worthwhile (Costanza et al. (2014)) - Costanza et al. (2014) After the recession of 2008 researchers have looked into other indicators to judge the quality of life. Nobel Prize-winning economists Joseph Stiglitz and Amartya Sen and their team have criticized GDP as a prime indicator of progress (Jeroen, V. (2017)) - Jeroen (2017)
   I have compared GDP and Cost of living and health care index (Both of the parameters are of the Quality of Life indicators dataset).
   According to the statistics which I found that the cost of living and health care index in developing when compared with GDP is higher than to cost of living and health care index in developed countries. There is a correlation in developed countries between GDP, Cost of living and health care index but not in developing countries.
   It clearly shows the indicator cost of living should be compared with GDP to have a better understanding of the quality of life in developing countries.

2. (Banerjee, D.(2016)) - Banerjee (2016) puts forward that for a nation to strength in terms of development, the nation should assess its GDP growth, quality of life of its citizen and human happiness. The study shows that there is a high correlation between macroeconomic development parameters such as education, health, quality of life index and GDP growth rate, expressed through the Human Development Index. Further, it also has a correlation with quality of life-based on the understanding urban socio-economic classes, which is measured in the study.
   Quality of life is studied by using five parameters such as quality of the house, education, health care, transportation, and entertainment facilities. The human happiness is assessed on the basis of the understanding of respondents towards change in their money-related conditions and consumption expenditure impacting their quality of life.

   Human Development Index has a relationship between with the economic development of a country. The research shows that when there is a positive change or a negative change in their financial condition over a period of time their consumption also changes on basis of the financial condition. This shows that the economic development, a prospect of improved financial condition and a better quality of life build a happy state of mind of the people.
   To study this research, in 2012-13 a sample of 850 respondents taken through questionnaire in Mumbai and its suburban areas.

   Analysis
   H0 : There is no relationship between economic development and quality of life

Ha1: There is positive correlation between economic development and quality of life.

Result :
Its proved that there is positive and strong association with economic development expressed through quantitative approach of GDP and the quality of life as estimated through HDI. Hence the null hypothesis is rejected and alternative hypothesis accepted(Banarjee, 2016, pp. 76-85) - Banerjee (2016)

H02 : There is no relationship between economic development and human happiness Ha2 : There is positive correlation between economic development and human happiness.
Result :
The results of sample taken shows that 54.976

I have compared quality of life index, happiness score of the country and GDP. After analyzing I found that quality of life index, happiness score of the country and GDP are correlated.

3. Education is key component of individual well-being. The author aims to presents different ways to analyze the quality of life and to draw special attention the link between the education, quality of life and employment rate in Romania and in general. Studies shows that attaining a high education qualification level increases to get a job and offers an individual to get higher salary. Also, high level of education decreases the risk of falling into poverty, precisely because people who invest in education find that to get a job it is easier and have good stability in the labour market. However, if an individual gets better training, there is an increased productivity level and quality, all directly or indirectly affecting the quality of life.
Furthermore the author shows that in Romania, according to ICCV and Eurostat studies in recent years the quality of life has been deteriorated because of the current crisis. Due to reduction in the income, increasing instability and uncertainty in the labour market and the economy the quality of life of Romanians has been affected. In Romania the relationship between quality of life, employment and education have suffered tremendously in recent years due to imbalances disclosed by the weak link between education and labour market needs, poor investment in education, increasing the duration of transition from school to work, difficulty in balancing work-life balance, and low participation in lifelong learning. (Aceleanu, M. (2012)) - Aceleanu (2012)
I will be comparing the education expenditure, employment rate, GDP, quality of life index and quality of education

This will show that the government should invest in education infrastructure because of which the individual will have less chance of being unemployed. Also, the country which has high quality of education will have maximum number of world rankings universities which will eventually contribute more to GDP and will improve the quality of life of an individual.

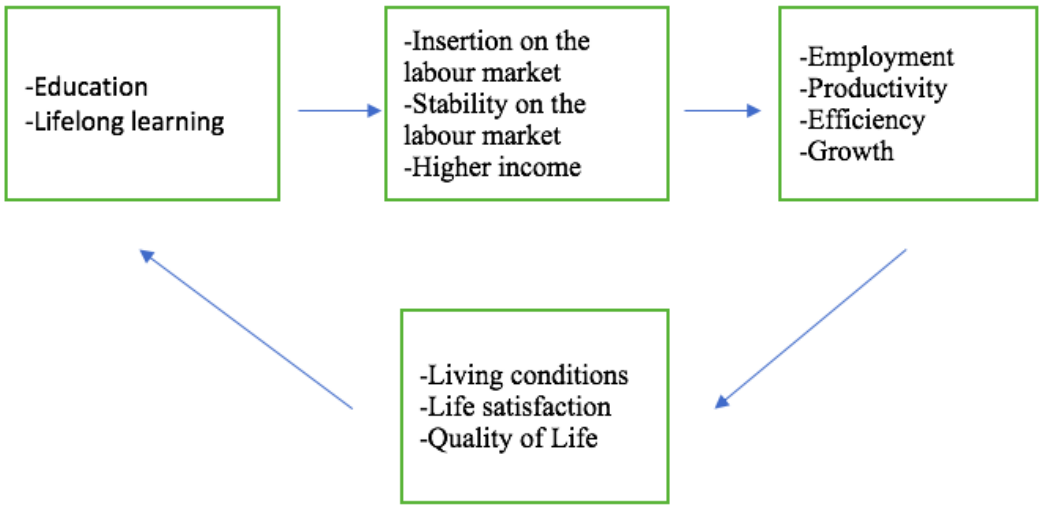


Figure: Relationship between education, employment and quality of life.
Aceleanu, Mirela Ionela. (2012)

Figure 1: Results for BI Query 1

4 Data Model

For implementation of this project Microsoft SQL Server 2017 is being used. For the transformation and loading of the data in data warehouse system services like Database engine, Analysis services and integration services from Microsoft SQL server are being used. This data warehouse system has one fact table and 7 dimension tables. The fact table contains all the measures required to answer the business queries mentioned in Section 1. The 7 dimension table contains information regarding GDP, government education expenditure, quality of life indicators, employment rate, happiness score, world university rankings and countries with region and sub-region. These dimensions are mapped with the fact table by writing SQL query of joining. It is done by The primary key column of the dimension table to the fact table. The above mentioned BI queries in Section 1, can be answered from this design of the data model. Measures from the fact table will serve as an answer to the Business Intelligence queries and can get the details from the dimension table whenever required.

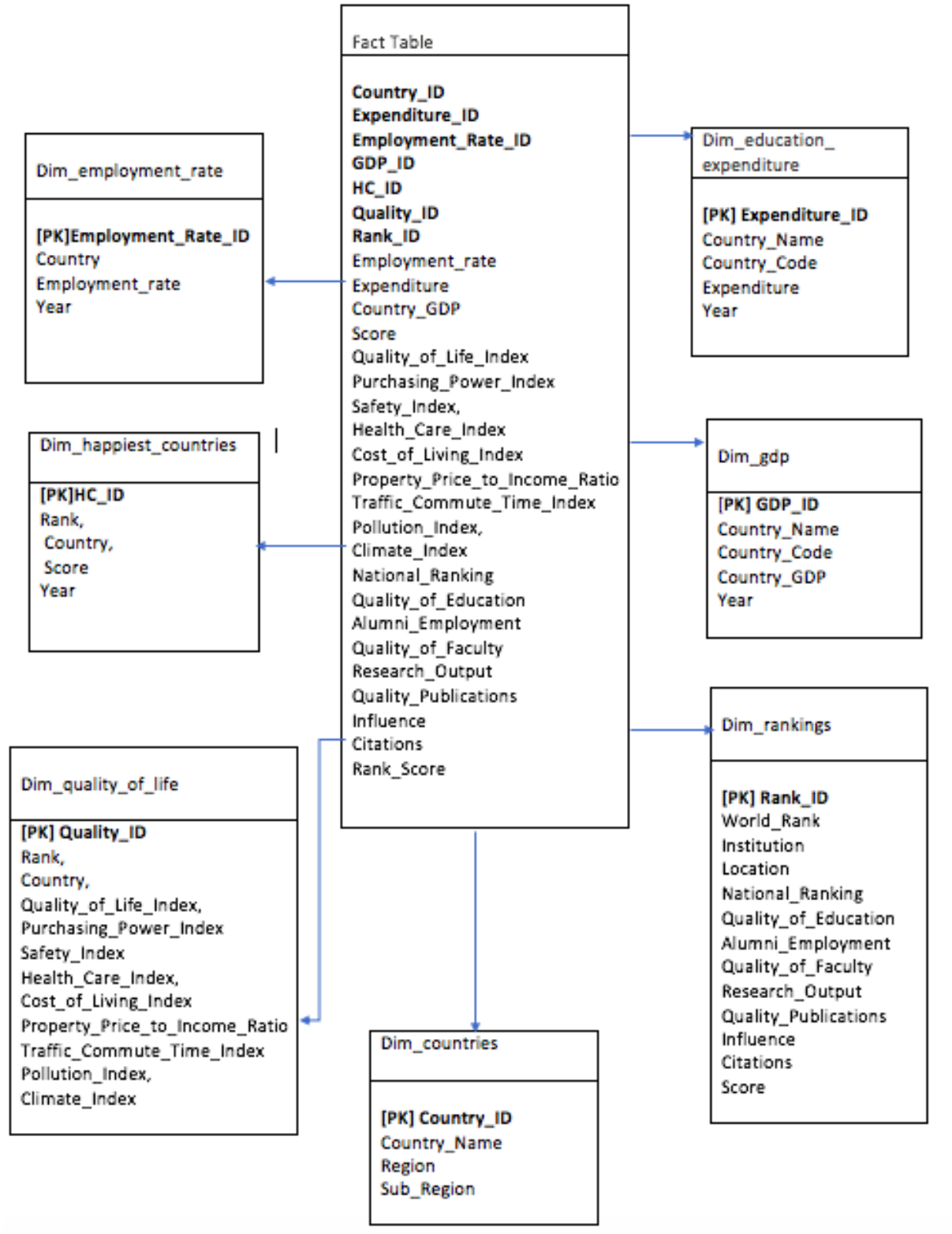


Figure 2: Star Schema
Figure 2.

Dimensions:

1. Dimension table 1 Countries with Region and Sub Region
   Name: Dim_countries
   Primary key: Country_ID
   Hierarchy: $Region \rightarrow Subregion \rightarrow CountryName$

2. Dimension table 2 - Government Education Expenditure
   Name: Dim_education_expenditure
   Fields: Country_Name, Country_Code, Expenditure and Year
   Primary Key: Expenditure_ID

3. Dimension table 3 Employment Rate
   Name: Dim_employment_rate
   Fields: Country, Employment_rate and Year
   Primary Key: Employment_Rate_ID

4. Dimension table 4 GDP
   Name: Dim_gdp
   Fields: Country_Name, Country_Code, Country_GDP and Year
   Primary Key: GDP_ID

5. Dimension table 5 Happiness Score country wise
   Name: Dim_happiest_countries
   Fields: Rank, Country, Score and Year
   Primary Key: HC_ID

6. Dimension table 6 Quality of life indicators
   Name: Dim_quality_of_life
   Fields: Rank, Country, Quality_of_Life_Index, Purchasing_Power_Index, Safety_Index,Health_Care_Index,Cost_of_Living_Index, Property_Price_to_Income_Ratio, Traffic_Commute_Time_Index, Pollution_Index,Climate_Index Primary Key: Quality_ID

7. Dimension table 7 World University Rankings
   Name: Dim_rankings
   Fields: World_Rank, Institution, Location, National_Ranking , Quality_of_Education, Alumni_Employment, Quality_of_Faculty, Research_Output, Quality_Publications , Influence, Citations , Score
   Primary Key: Rank_ID
   Hierarchy: $WorldRank \rightarrow NationalRank$

Fact table: Fact_Table Fields: Country_ID, Expenditure_ID, Employment_Rate_ID, GDP_ID, HC_ID, Quality_ID, Rank_ID, Employment_rate, Expenditure, Country_GDP, Score, Quality_of_Life_Index, Purchasing_Power_Index, Safety_Index,Health_Care_Index,

Cost_of_Living_Index, Property_Price_to_Income_Ratio, Traffic_Commute_Time_Index, Pollution_Index,Climate_Index, National_Ranking , Quality_of_Education, Alumni_Employment, Quality_of_Faculty, Research_Output, Quality_Publications , Influence, Citations , Rank_Score

5 Logical Data Map

In this section, describe your logical data map, i.e. how every row of every data source is handled such that it is a part of your star schema.

Table 2: Logical Data Map describing all transformations, sources and destinations for all components of the data model illustrated in Figure 2

Source	Column	Destination	Column	Type	Transformation
1	Employment rate in OECD countries in 2017	Dim employment rate	Country	Dimension	I have changed the column name for readability
1	Employment rate	Dim employment rate	Employment rate	Fact	
1		Percentage sign			I have deleted this column which had percentage sign
1		Dim employment rate	Employment rate id	Fact	I have created this id because it act as a foreign key in fact table
2	Unstructured	Dim happiest countries	Rank	Dimension	The column is spilted using strsplit and
2	Unstructured	Dim happiest countries	Country	Dimension	The column was cleaned
2	Unstructured	Dim happiest countries	Score	Fact	It represents the happiness score of the country. There were
2	Unstructured	Dim happiest countries	Year	Dimension	This column is added so that it can be differentiated with years
2	Unstructured	Dim happiest countries	HC ID	Fact	I have created this id because it act as a foreign key in fact table

Continued on next page

Table 2 – *Continued from previous page*

Source	**Column**	**Destination**	**Column**	**Type**	**Transformation**
4.1		Dim_gdp	GDP_ID	Fact	I have created this id because it act as a foreign key in fact table

6 ETL Process

ETL flow:

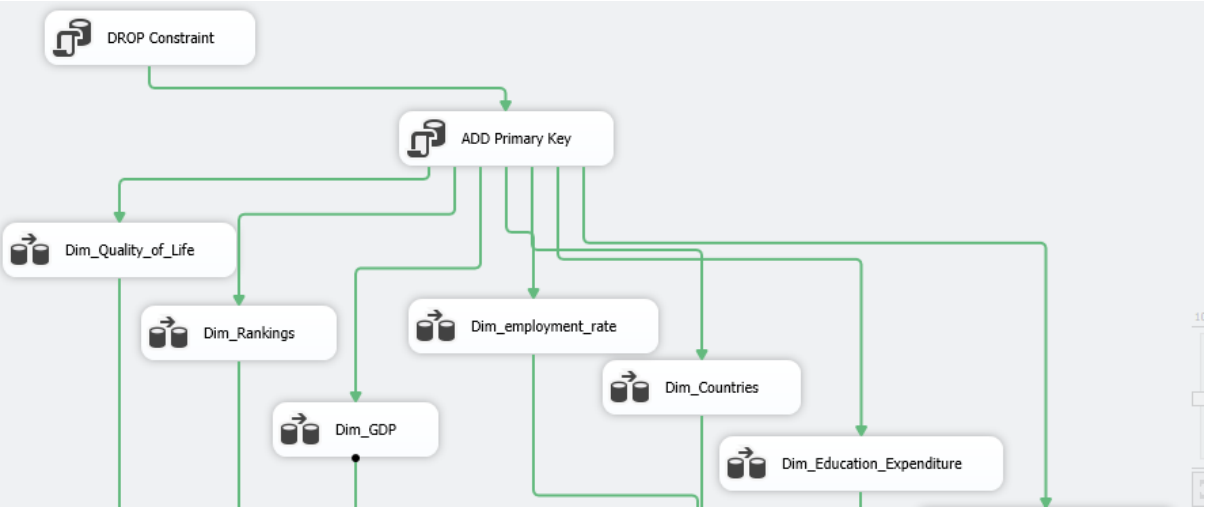


Figure 3: ETL flow 1
Figure 3.

1.) The SSIS SqlTask called Add primary Key will add primary key constraints to all the dimension tables.

2.) Dim_Quality_of_Life will load the dimension table from the staging table .

3.) Dim_Rankings will load the dimension table from the staging area table.

4.) Dim_GDP will load the dimension table dim GDP after necessary transformations.

5.) Dim_employment_rate will load data from the staging tables to the dim dimension.

6.) Dim_countries will load data from into the dimension table from the staging table.

7.) Dim_education_expenditure will load data into the respective dimension table after the necessary transformations.

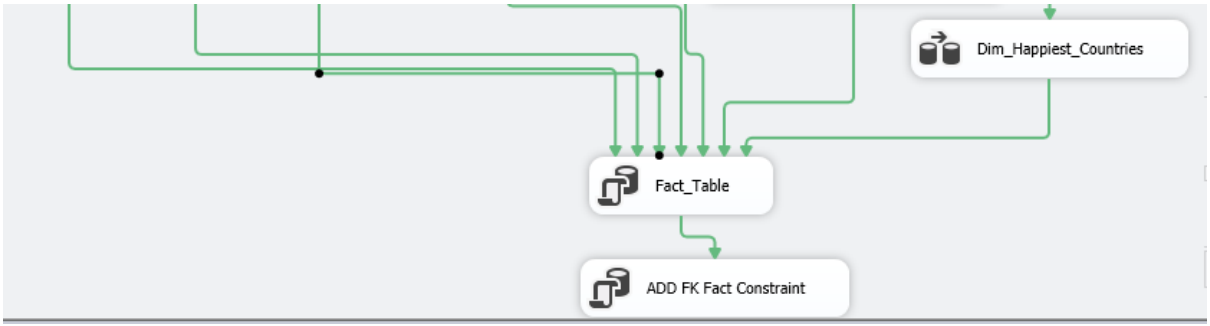


Figure 4: ETL flow 2
Figure 4.

8.) Dim_Happiest_Countries will load the dimension table from the staging area tables

9.) Finally the Sqltask called Fact_Table will load all facts by joining on the common columns in the dimension tables.

10.) After Fact table load is complete, we finally add the Foreign key constraints to fact table as it improves fact_table load times.

Loading:

The data is been loaded from R to SMSS using the following code.

library(RODBC) Setting up dB Connectionn string Exporting Refined Data to SQL SERVER sqlConnString ¡- odbcDriverConnect("Driver=SQL Server;Server=MOLAP; Database=Qualit
$true")try(sqlDrop(sqlConnString, "stage_quality_of_life_combine"), silent = TRUE)sqlSave(sqlConnSt
"stage_quality_of_life_combine", rownames = F)

Simlarly, I have loaded other files using the same code.

7 Application

Comparing Quality of life indicators, GDP, Employment Rate, Happiness score of different countries,Government Education Expenditure and Top World Ranking Universities.

7.1 BI Query 1: Comparison GDP, Cost of living and Health care index(Both of the parameters of the Quality of Life indicators)

For this query, the contributing sources of data are: Source 3 i.e Quality of life indicators dataset and source 4.1 GDP per capita.

The general findings are that developing countries such as Brazil and India have cost of living and health care index on a higher side as we compare to GDP of their countries However, if we look at the developed countries such as France, Germany, Switzerland the cost of living, health care index and GDP are in co relation as shown in figure below.

Figure 5: Results for BI Query 1

7.2 BI Query 2: Comparison of happiest countries, quality of life and Country GDP

For this query, the contributing sources of data are Source 2 i.e Happiest Countries, Source 3 i.e Quality of life indicators and Source 4.1 Country GDP.

The general findings are that Country GDP, Quality of life index and Happiness score(higher the rank, greater the score) are correlated with each other.I have considered 3 years of data i.e. 2015, 2016 and 2017. findings is illustrated in Figure 6.

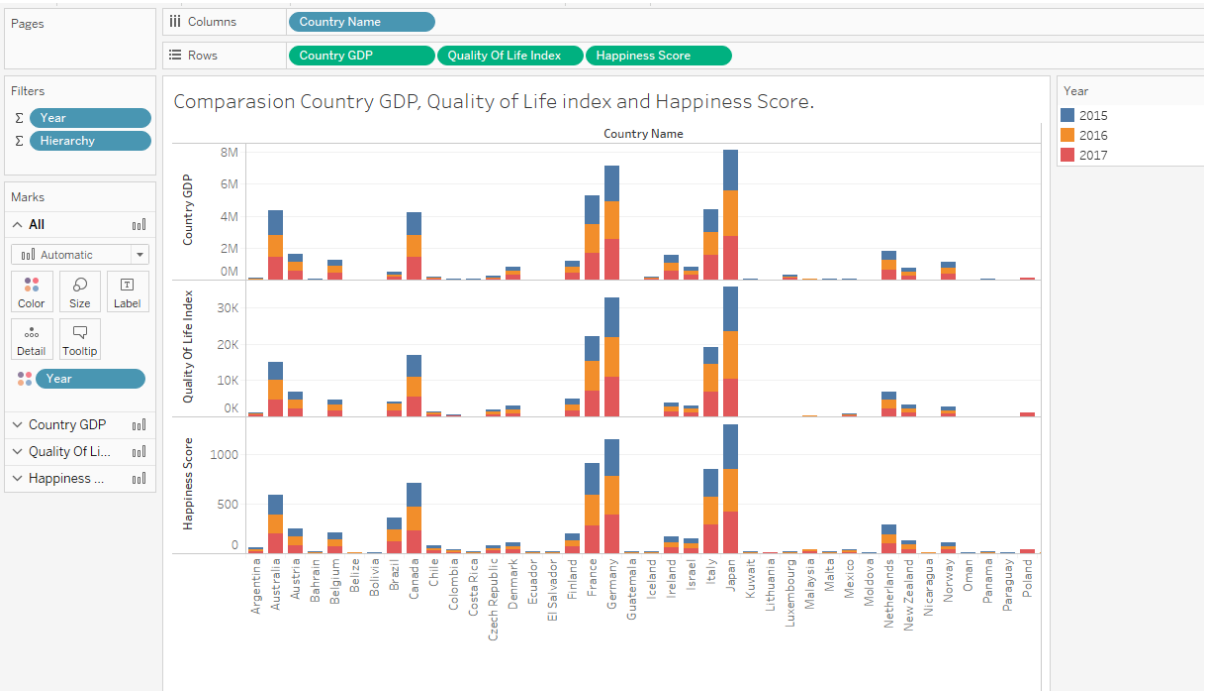
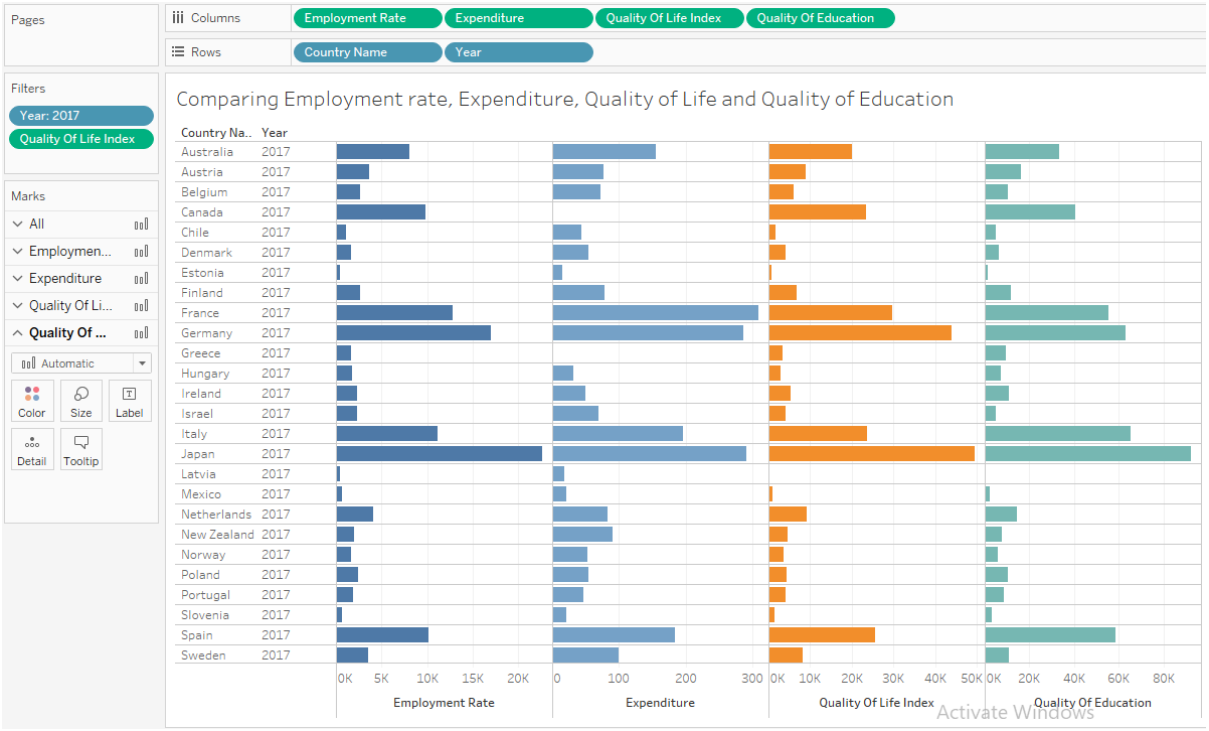


Figure 6: Results for BI Query 2

7.3 BI Query 3: Comparison the education expenditure, employment rate, GDP, quality of life index and quality of education in top universities

For this query, the contributing sources of data are: Source 1 i.e. statista dataset, Source 4.2 i.e. Education Expenditure, Source 3 i.e. Quality of life indicator and Source 5 i.e University rankings dataset.

The general findings are that the Employment rate, Education Expenditure, Quality of life index and quality of education are correlated as illustrated in Figure 7.3.



Results for BI Query 3

7.4 Discussion

1) Jeroen (2017) mentions that the quality of life should be dependent on one indicator to measure quality of life. In BI query 1 I have compared GDP, Cost of living and health care of developing countries and developed countries. According to my findings countries such as Puerto Rico, Mongolia, Brazil and India are higher cost of living and health care index.
But, Countries such as Switzerland, Sweden, Austria, Canada have Cost of living and health care index when we compare it with their GDP.
This shows GDP should not be the sole factor to calculate the measure of quality of life.
2)Banerjee (2016) mentions that the GDP, happiness and quality of life are highly correlated. According to statistics I found out by analyzing 3 years of data it shows that all the three parameters go hand in hand.
3)Aceleanu (2012) mentions that in Romania the quality of life is deteriorated due to imbalances disclosed to by the weak link between quality of life, education and employment. According to my findings it shows that there is a strong link in between.If employment rate of country and government spends on education expenditure this improve quality of education will which eventually make quality o life index better.

8 Conclusion and Future Work

To conclude, GDP undoubtedly is of the main parameters to judge countries health but it should not be the one sole parameter to measure quality of life. Also, In contrast to BI query my findings show that the happiness of country, GDP and quality of life index are correlated. Lastly,the government should invest in education infrastructure because of which the individual will have less chance of being unemployed. Also, the country which has high quality of education will have maximum number of world rankings universities

which will eventually contribute more to GDP and will improve the quality of life of an individual.

After the recession researchers are now calculating what should be parameters that will be successor to GDP to judge quality of life. The new set parameters should have current knowledge, sociology collectively, economics, psychology and measures to calculate sustainable wellbeing,. The new metrics must garner broad support from stakeholders in the coming conclaves Costanza et al. (2014)

References

Aceleanu, M. (2012), 'mpact of economic development on quality of life and human happiness: a study on urban socio economic classes of suburban mumbai', *International Journal of Business Economic Development* **4**(3), pp. 717–730.

Banerjee, D. (2016), 'mpact of economic development on quality of life and human happiness: a study on urban socio economic classes of suburban mumbai', *International Journal of Business Economic Development* **4**(3), pp. 76–85.

Costanza, R., Kubiszewski, I. & Giovannini, E. (2014), 'Development: Time to leave gdp behind', *Nature* **505**, pp. 283285.

Jeroen, V. (2017), 'Dont worry, be happy', *Alternatives Journal (AJ) - Canada's Environmental Voice* **43**(1), p22–24. 3p.

Walter, R. (2015), 'Recent and future developments related to 'gdp and beyond", *Review of Income Wealth* **61**(1), p18. 7p.

Appendix

R code example

```
#Quality of life indicators

Sys.setenv(JAVA_HOME="C://Program␣Files//Java//jre1.8.0_151")
.libPaths("C:\\Users\\MOLAP\\Desktop\\R")
#set directory
setwd("C:/Users/MOLAP/Desktop/Quality_of_life")
install.packages(rvest)
##Import dataset from Numbeo - YEAR 2013
##https://www.numbeo.com/quality-of-life/rankings_by_country.jsp?
   title=2013
install.packages("XML")
install.packages("RCurl")
install.packages("readxl")
library(XML)
library(RCurl)
library(readxl)
url_file<-readHTMLTable("https://www.numbeo.com/quality-of-life/
   rankings_by_country.jsp?title=2013-Q1")
```

```
table_to_extracted<-getURL(url_file)
quality_of_life_2013<-readHTMLTable(table_to_extracted,which = 3)
colnames(quality_of_life_2013) <- gsub("␣","_",colnames(quality_
   of_life_2013))
quality_of_life_2013$Rank<-as.character(quality_of_life_2013$Rank
   )
quality_of_life_2013$Country<-as.character(quality_of_life_2013$
   Country)
quality_of_life_2013$Quality_of_Life_Index<-as.character(quality_
   of_life_2013$Quality_of_Life_Index)
quality_of_life_2013$Purchasing_Power_Index<-as.character(quality
   _of_life_2013$Purchasing_Power_Index)
quality_of_life_2013$Safety_Index<-as.character(quality_of_life_2
   013$Safety_Index)
quality_of_life_2013$Health_Care_Index<-as.character(quality_of_
   life_2013$Health_Care_Index)
quality_of_life_2013$Cost_of_Living_Index<-as.character(quality_
   of_life_2013$Cost_of_Living_Index)
quality_of_life_2013$Property_Price_to_Income_Ratio<-as.character
   (quality_of_life_2013$Property_Price_to_Income_Ratio)
quality_of_life_2013$Traffic_Commute_Time_Index<-as.character(
   quality_of_life_2013$Traffic_Commute_Time_Index)
quality_of_life_2013$Pollution_Index<-as.character(quality_of_
   life_2013$Pollution_Index)
quality_of_life_2013$Climate_Index<-as.character(quality_of_life_
   2013$Climate_Index)
for(i in 0:length(quality_of_life_2013$Rank)-1) {
  quality_of_life_2013$Rank[i+1]<-i+1}
quality_of_life_2013$Year <- "2013"
#########################################################2014
   ##########################################################################

url_file<-"https://www.numbeo.com/quality-of-life/rankings_by_
   country.jsp?title=2014"
table_to_extracted<-getURL(url_file)
quality_of_life_2014<-readHTMLTable(table_to_extracted,which = 3)
colnames(quality_of_life_2014) <- gsub("␣","_",colnames(quality_
   of_life_2014))
quality_of_life_2014$Rank<-as.character(quality_of_life_2014$Rank
   )
quality_of_life_2014$Country<-as.character(quality_of_life_2014$
   Country)
quality_of_life_2014$Quality_of_Life_Index<-as.character(quality_
   of_life_2014$Quality_of_Life_Index)
quality_of_life_2014$Purchasing_Power_Index<-as.character(quality
   _of_life_2014$Purchasing_Power_Index)
quality_of_life_2014$Safety_Index<-as.character(quality_of_life_2
   014$Safety_Index)
quality_of_life_2014$Health_Care_Index<-as.character(quality_of_
   life_2014$Health_Care_Index)
quality_of_life_2014$Cost_of_Living_Index<-as.character(quality_
```

```
   of_life_2014$Cost_of_Living_Index)
quality_of_life_2014$Property_Price_to_Income_Ratio<-as.character
   (quality_of_life_2014$Property_Price_to_Income_Ratio)
quality_of_life_2014$Traffic_Commute_Time_Index<-as.character(
   quality_of_life_2014$Traffic_Commute_Time_Index)
quality_of_life_2014$Pollution_Index<-as.character(quality_of_
   life_2014$Pollution_Index)
quality_of_life_2014$Climate_Index<-as.character(quality_of_life_
   2014$Climate_Index)
quality_of_life_2014$Climate_Index<-0
for(i in 0:length(quality_of_life_2014$Rank)-1) {
  quality_of_life_2014$Rank[i+1]<-i+1
}
quality_of_life_2014$Year <- "2014"
sapply(quality_of_life_2014, class)
#########################################################2015
   ##########################################################################

##Import dataset from Numbeo
##https://www.numbeo.com/quality-of-life/rankings_by_country.jsp?
   title=2015
url_file<-"https://www.numbeo.com/quality-of-life/rankings_by_
   country.jsp?title=2015"
table_to_extracted<-getURL(url_file)
quality_of_life_2015<-readHTMLTable(table_to_extracted,which = 3)
colnames(quality_of_life_2015) <- gsub("␣","_",colnames(quality_
   of_life_2015))
quality_of_life_2015$Rank<-as.character(quality_of_life_2015$Rank
   )
quality_of_life_2015$Country<-as.character(quality_of_life_2015$
   Country)
quality_of_life_2015$Quality_of_Life_Index<-as.character(quality_
   of_life_2015$Quality_of_Life_Index)
quality_of_life_2015$Purchasing_Power_Index<-as.character(quality
   _of_life_2015$Purchasing_Power_Index)
quality_of_life_2015$Safety_Index<-as.character(quality_of_life_2
   015$Safety_Index)
quality_of_life_2015$Health_Care_Index<-as.character(quality_of_
   life_2015$Health_Care_Index)
quality_of_life_2015$Cost_of_Living_Index<-as.character(quality_
   of_life_2015$Cost_of_Living_Index)
quality_of_life_2015$Property_Price_to_Income_Ratio<-as.character
   (quality_of_life_2015$Property_Price_to_Income_Ratio)
quality_of_life_2015$Traffic_Commute_Time_Index<-as.character(
   quality_of_life_2015$Traffic_Commute_Time_Index)
quality_of_life_2015$Pollution_Index<-as.character(quality_of_
   life_2015$Pollution_Index)
quality_of_life_2015$Climate_Index<-as.character(quality_of_life_
   2015$Climate_Index)
quality_of_life_2015$Climate_Index<-0
for(i in 0:length(quality_of_life_2015$Rank)-1) {
```

```
  quality_of_life_2015$Rank[i+1]<-i+1
}
quality_of_life_2015$Year <- "2015"
##################################################2016
   ##########################################################################

url_file<-"https://www.numbeo.com/quality-of-life/rankings_by_
   country.jsp?title=2016"
table_to_extracted<-getURL(url_file)
quality_of_life_2016<-readHTMLTable(table_to_extracted,which = 3)
colnames(quality_of_life_2016) <- gsub("␣","_",colnames(quality_
   of_life_2016))
quality_of_life_2016$Rank<-as.character(quality_of_life_2016$Rank
   )
quality_of_life_2016$Country<-as.character(quality_of_life_2016$
   Country)
quality_of_life_2016$Quality_of_Life_Index<-as.character(quality_
   of_life_2016$Quality_of_Life_Index)
quality_of_life_2016$Purchasing_Power_Index<-as.character(quality
   _of_life_2016$Purchasing_Power_Index)
quality_of_life_2016$Safety_Index<-as.character(quality_of_life_2
   016$Safety_Index)
quality_of_life_2016$Health_Care_Index<-as.character(quality_of_
   life_2016$Health_Care_Index)
quality_of_life_2016$Cost_of_Living_Index<-as.character(quality_
   of_life_2016$Cost_of_Living_Index)
quality_of_life_2016$Property_Price_to_Income_Ratio<-as.character
   (quality_of_life_2016$Property_Price_to_Income_Ratio)
quality_of_life_2016$Traffic_Commute_Time_Index<-as.character(
   quality_of_life_2016$Traffic_Commute_Time_Index)
quality_of_life_2016$Pollution_Index<-as.character(quality_of_
   life_2016$Pollution_Index)
quality_of_life_2016$Climate_Index<-as.character(quality_of_life_
   2016$Climate_Index)
for(i in 0:length(quality_of_life_2016$Rank)-1) {
  quality_of_life_2016$Rank[i+1]<-i+1
}
quality_of_life_2016$Year <- "2016"
##################################################2017
   ##########################################################################

url_file<-"https://www.numbeo.com/quality-of-life/rankings_by_
   country.jsp?title=2017"
table_to_extracted<-getURL(url_file)
quality_of_life_2017<-readHTMLTable(table_to_extracted,which = 3)
colnames(quality_of_life_2017) <- gsub("␣","_",colnames(quality_
   of_life_2017))
quality_of_life_2017$Rank<-as.character(quality_of_life_2017$Rank
   )
quality_of_life_2017$Country<-as.character(quality_of_life_2017$
   Country)
```

```
quality_of_life_2017$Quality_of_Life_Index<-as.character(quality_
   of_life_2017$Quality_of_Life_Index)
quality_of_life_2017$Purchasing_Power_Index<-as.character(quality
   _of_life_2017$Purchasing_Power_Index)
quality_of_life_2017$Safety_Index<-as.character(quality_of_life_2
   017$Safety_Index)
quality_of_life_2017$Health_Care_Index<-as.character(quality_of_
   life_2017$Health_Care_Index)
quality_of_life_2017$Cost_of_Living_Index<-as.character(quality_
   of_life_2017$Cost_of_Living_Index)
quality_of_life_2017$Property_Price_to_Income_Ratio<-as.character
   (quality_of_life_2017$Property_Price_to_Income_Ratio)
quality_of_life_2017$Traffic_Commute_Time_Index<-as.character(
   quality_of_life_2017$Traffic_Commute_Time_Index)
quality_of_life_2017$Pollution_Index<-as.character(quality_of_
   life_2017$Pollution_Index)
quality_of_life_2017$Climate_Index<-as.character(quality_of_life_
   2017$Climate_Index)
for(i in 0:length(quality_of_life_2017$Rank)-1) {
  quality_of_life_2017$Rank[i+1]<-i+1}
quality_of_life_2017$Year <- "2017"
sapply(quality_of_life_2017, class)
#------------------------------R BIND ALL THE QUALITY OF LIFE
   TABLES---------
quality_of_life_combine <- rbind(quality_of_life_2013,quality_of_
   life_2014,quality_of_life_2015,quality_of_life_2016,quality_of
   _life_2017)
quality_of_life_combine$Climate_Index <- as.numeric(quality_of_
   life_combine$Climate_Index)
#dnstcts
quality_of_life_combine$Climate_Index[quality_of_life_combine$
   Year==2013]<-quality_of_life_combine$Climate_Index[quality_of_
   life_combine$Year==2017]
#sapply(quality_of_life_2017, FUN=function(x) {sum(is.na(x))})
quality_of_life_combine$Rank <- ass(quality_of_life_combine$Rank)
quality_of_life_combine$Quality_of_Life_Index <- as.numeric(
   quality_of_life_combine$Quality_of_Life_Index)
quality_of_life_combine$Purchasing_Power_Index <- as.numeric(
   quality_of_life_combine$Purchasing_Power_Index)
quality_of_life_combine$Safety_Index <- as.numeric(quality_of_
   life_combine$Safety_Index)
quality_of_life_combine$Health_Care_Index <- as.numeric(quality_
   of_life_combine$Health_Care_Index)
quality_of_life_combine$Cost_of_Living_Index <- as.numeric(
   quality_of_life_combine$Cost_of_Living_Index)
quality_of_life_combine$Property_Price_to_Income_Ratio <- as.
   numeric(quality_of_life_combine$Property_Price_to_Income_Ratio
   )
quality_of_life_combine$Traffic_Commute_Time_Index <- as.numeric(
   quality_of_life_combine$Traffic_Commute_Time_Index)
quality_of_life_combine$Pollution_Index <- as.numeric(quality_of_
```

```
   life_combine$Pollution_Index)
quality_of_life_combine$Climate_Index <- as.numeric(quality_of_
   life_combine$Climate_Index)
quality_of_life_combine$Year <- as.integer(quality_of_life_
   combine$Year)
#sapply(quality_of_life_combine, class)
quality_of_life_combine <- data.frame(quality_of_life_combine)
#-----------------------------END R BIND ALL THE QUALITY OF LIFE
   TABLES---------
library(RODBC)
#########Setting up dB Connectionn string ###############
############## Exporting Refined Data to SQL SERVER######
sqlConnString <- odbcDriverConnect("Driver=SQL␣Server;Server=
   MOLAP;␣Database=Quality_Of_Life;trusted_connection=true")
try(sqlDrop(sqlConnString,"stage_quality_of_life_combine"),silent
   =TRUE)
sqlSave(sqlConnString,quality_of_life_combine,tablename = "stage_
   quality_of_life_combine",rownames = F )

#GDP

Sys.setenv(JAVA_HOME="C://Program␣Files//Java//jre1.8.0_151")
#set directory
setwd("C:/Users/MOLAP/Desktop/Quality_of_life")
library(readr)
########## reading dataset
csv_data<-read.csv("gdp_per_capita_world_bank_dollars.csv",skip=4
   )
csv_data$Indicator.Name<-NULL
csv_data$Indicator.Code<-NULL
x <- csv_data[,c(55:60)]
y<-csv_data[,c(1:2)]
gdp_year_by_year <- data.frame(y,x)
# gdp_year_by_year$X2012[is.na(gdp_year_by_year$X2012)]<-0
# gdp_year_by_year$X2013[is.na(gdp_year_by_year$X2013)]<-0
# gdp_year_by_year$X2014[is.na(gdp_year_by_year$X2014)]<-0
# gdp_year_by_year$X2015[is.na(gdp_year_by_year$X2015)]<-0
# gdp_year_by_year$X2016[is.na(gdp_year_by_year$X2016)]<-0
# gdp_year_by_year$X2017[is.na(gdp_year_by_year$X2017)]<-0

z<-rbind(gdp_year_by_year$X2012,gdp_year_by_year$X2013,gdp_year_
   by_year$X2014,gdp_year_by_year$X2015,gdp_year_by_year$X2016,
   gdp_year_by_year$X2017)
z<-data.frame(z)
gdp<-c(z$X1, z$X2, z$X3, z$X4, z$X5, z$X6, z$X7, z$X8, z$X9, z$X1
   0, z$X11, z$X12, z$X13, z$X14, z$X15, z$X16, z$X17, z$X18, z$X
   19, z$X20, z$X21, z$X22, z$X23, z$X24, z$X25, z$X26, z$X27, z$
   X28, z$X29, z$X30, z$X31, z$X32, z$X33, z$X34, z$X35, z$X36, z
   $X37, z$X38, z$X39, z$X40, z$X41, z$X42, z$X43, z$X44, z$X45,
   z$X46, z$X47, z$X48, z$X49, z$X50, z$X51, z$X52, z$X53, z$X54,
    z$X55, z$X56, z$X57, z$X58, z$X59, z$X60, z$X61, z$X62, z$X63
   , z$X64, z$X65, z$X66, z$X67, z$X68, z$X69, z$X70, z$X71, z$X7
```

```
   2, z$X73, z$X74, z$X75, z$X76, z$X77, z$X78, z$X79, z$X80, z$X
   81, z$X82, z$X83, z$X84, z$X85, z$X86, z$X87, z$X88, z$X89, z$
   X90, z$X91, z$X92, z$X93, z$X94, z$X95, z$X96, z$X97, z$X98, z
   $X99, z$X100, z$X101, z$X102, z$X103, z$X104, z$X105, z$X106,
   z$X107, z$X108, z$X109, z$X110, z$X111, z$X112, z$X113, z$X114
   , z$X115, z$X116, z$X117, z$X118, z$X119, z$X120, z$X121, z$X1
   22, z$X123, z$X124, z$X125, z$X126, z$X127, z$X128, z$X129, z$
   X130, z$X131, z$X132, z$X133, z$X134, z$X135, z$X136, z$X137,
   z$X138, z$X139, z$X140, z$X141, z$X142, z$X143, z$X144, z$X145
   , z$X146, z$X147, z$X148, z$X149, z$X150, z$X151, z$X152, z$X1
   53, z$X154, z$X155, z$X156, z$X157, z$X158, z$X159, z$X160, z$
   X161, z$X162, z$X163, z$X164, z$X165, z$X166, z$X167, z$X168,
   z$X169, z$X170, z$X171, z$X172, z$X173, z$X174, z$X175, z$X176
   , z$X177, z$X178, z$X179, z$X180, z$X181, z$X182, z$X183, z$X1
   84, z$X185, z$X186, z$X187, z$X188, z$X189, z$X190, z$X191, z$
   X192, z$X193, z$X194, z$X195, z$X196, z$X197, z$X198, z$X199,
   z$X200, z$X201, z$X202, z$X203, z$X204, z$X205, z$X206, z$X207
   , z$X208, z$X209, z$X210, z$X211, z$X212, z$X213, z$X214, z$X2
   15, z$X216, z$X217, z$X218, z$X219, z$X220, z$X221, z$X222, z$
   X223, z$X224, z$X225, z$X226, z$X227, z$X228, z$X229, z$X230,
   z$X231, z$X232, z$X233, z$X234, z$X235, z$X236, z$X237, z$X238
   , z$X239, z$X240, z$X241, z$X242, z$X243, z$X244, z$X245, z$X2
   46, z$X247, z$X248, z$X249, z$X250, z$X251, z$X252, z$X253, z$
   X254, z$X255, z$X256, z$X257, z$X258, z$X259, z$X260, z$X261,
   z$X262, z$X263, z$X264)
gdp<-data.frame(gdp)
gdp_year_by_year<-gdp_year_by_year[rep(seq_len(nrow(gdp_year_by_
   year)), each=6),]
gdp_year_by_year<-cbind(gdp_year_by_year$Country.Name,gdp_year_by
   _year$Country.Code,gdp)

# f<-c(2012,2013,2014,2015,2016,2017)
# f<-data.frame(f)
# f<-f[rep(seq_len(nrow(f)), each=264),]
#https://stackoverflow.com/questions/11121385/repeat-rows-of-a-
   data-frame
# f<-data.frame(f)
u <- rep(c(2012,2013,2014,2015,2016,2017), times = 264)
gdp<-cbind(gdp_year_by_year,u)
colnames(gdp)<-c("Country_Name","Country_Code","Country_GDP","
   Year")
gdp
#Dealing with missing vaules
#sapply(gdp, FUN=function(x) {sum(is.na(x))})
#sapply(gdp, class)
gdp$Country_GDP[is.na(gdp$Country_GDP)] <- 0
gdp$Country_Name <- as.character(gdp$Country_Name)
gdp$Country_Code <- as.character(gdp$Country_Code)
gdp$Country_GDP <- as.numeric(gdp$Country_GDP)
gdp$Year <- as.integer(gdp$Year)
# which(is.na(gdp$Expenditure))
```

```
# x<-which(is.na(gdp$Expenditure))
# x
# gdp$Country_Name[x]
# unique(gdp$Country_Name[x])
#End

#########Setting up dB Connectionn string ############
############## Exporting Refined Data to SQL SERVER######
sqlConnString <- odbcDriverConnect("Driver=SQL␣Server;Server=
   MOLAP;␣Database=Quality_Of_Life;trusted_connection=true")
try(sqlDrop(sqlConnString,"stage_gdp"),silent=TRUE)
sqlSave(sqlConnString,gdp,tablename = "stage_gdp",rownames = F )

#Happiness score of the countries

#setwd("~/Desktop/Quality_of_life")
##Import dataset from https://cwur.org/2013.php
setwd("C:/Users/MOLAP/Desktop/Quality_of_life")
library(readr)
library(XML)
library(RCurl)
url_file<-read_html("https://cwur.org/2013.php")
world_universities_rankings_2013 <- getURL(url_file)
#rankings <-readHTMLTable(getURL(url_file),which = 1)
rankings_2013 <-readHTMLTable(world_universities_rankings_2013,
   which = 1)
colnames(rankings_2013)<-c("World_Rank","Institution","Location",
   "National_Ranking","Quality_of_Education","Alumni_Employment",
   "Quality_of_Faculty","Quality_Publications","Influence","
   Citations","Patents","Score")
rankings_2013$Patents <- NULL
rankings_2013$Year<-"2013"

#-------------------------------------------------
#Converting int to char to remove + sign
rankings_2013$Quality_of_Education <- as.character(rankings_2013$
   Quality_of_Education)
rankings_2013$Quality_of_Education <- gsub("\\+","",rankings_2013
   $Quality_of_Education)
#converting back to int to load
rankings_2013$Quality_of_Education
#Converting int to char to remove + sign
rankings_2013$Alumni_Employment <- as.character(rankings_2013$
   Alumni_Employment)
rankings_2013$Alumni_Employment <- gsub("\\+","",rankings_2013$
   Alumni_Employment)
#converting back to int to load
rankings_2013$Alumni_Employment
#Converting int to char to remove + sign
rankings_2013$Quality_of_Faculty <- as.character(rankings_2013$
   Quality_of_Faculty)
rankings_2013$Quality_of_Faculty <- gsub("\\+","",rankings_2013$
```

```
   Quality_of_Faculty)
#converting back to int to load
rankings_2013$Quality_of_Faculty
#Converting int to char to remove + sign
rankings_2013$Quality_Publications <- as.character(rankings_2013$
   Quality_Publications)
rankings_2013$Quality_Publications <- gsub("\\+","",rankings_2013
   $Quality_Publications)
#converting back to int to load
rankings_2013$Quality_Publications
#Converting int to char to remove + sign
rankings_2013$Influence <- as.character(rankings_2013$Influence)
rankings_2013$Influence <- gsub("\\+","",rankings_2013$Influence)
#converting back to int to load
rankings_2013$Influence
#Converting int to char to remove + sign
rankings_2013$Citations <- as.character(rankings_2013$Citations)
rankings_2013$Citations <- gsub("\\+","",rankings_2013$Citations)
#converting back to int to load
rankings_2013$Citations
#Converting int to char to remove + sign
rankings_2013$Patents <- as.character(rankings_2013$Patents)
rankings_2013$Patents <- gsub("\\+","",rankings_2013$Patents)
#converting back to int to load
rankings_2013$Patents

##########################################################2014
   ##########################################################################

url_file<-"https://cwur.org/2014.php"
world_universities_rankings_2014 <- getURL(url_file)
#rankings <-readHTMLTable(getURL(url_file),which = 1)
rankings_2014 <-readHTMLTable(getURL(url_file),which = 1)
colnames(rankings_2014)<-c("World_Rank","Institution","Location",
   "National_Ranking","Quality_of_Education","Alumni_Employment",
   "Quality_of_Faculty","Research_Output","Quality_Publications",
   "Influence","Citations","Patents","Score")
rankings_2014$Research_Output <- NULL
rankings_2014$Patents <- NULL
rankings_2014$Year<-"2014"

#check data type
sapply(rankings_2014, class)
#Converting int to char to remove + sign
rankings_2014$Quality_of_Education <- as.character(rankings_2014$
   Quality_of_Education)
rankings_2014$Quality_of_Education <- gsub("\\+","",rankings_2014
   $Quality_of_Education)

#Converting int to char to remove + sign
rankings_2014$Alumni_Employment <- as.character(rankings_2014$
```

```
   Alumni_Employment)
rankings_2014$Alumni_Employment <- gsub("\\+","",rankings_2014$
   Alumni_Employment)

#Converting int to char to remove + sign
rankings_2014$Quality_of_Faculty <- as.character(rankings_2014$
   Quality_of_Faculty)
rankings_2014$Quality_of_Faculty <- gsub("\\+","",rankings_2014$
   Quality_of_Faculty)

#Converting int to char to remove + sign
rankings_2014$Quality_Publications <- as.character(rankings_2014$
   Quality_Publications)
rankings_2014$Quality_Publications <- gsub("\\+","",rankings_2014
   $Quality_Publications)

#Converting int to char to remove + sign
rankings_2014$Influence <- as.character(rankings_2014$Influence)
rankings_2014$Influence <- gsub("\\+","",rankings_2014$Influence)

#Converting int to char to remove + sign
rankings_2014$Citations <- as.character(rankings_2014$Citations)
rankings_2014$Citations <- gsub("\\+","",rankings_2014$Citations)

#Converting int to char to remove + sign
rankings_2014$Patents <- as.character(rankings_2014$Patents)
rankings_2014$Patents <- gsub("\\+","",rankings_2014$Patents)

#check data type
sapply(rankings_2014, class)
##########################################################2015
   ##########################################################################

url_file<-"https://cwur.org/2015.php"
world_universities_rankings_2015 <- getURL(url_file)

rankings_2015 <-readHTMLTable(getURL(url_file),which = 1)
colnames(rankings_2015)<-c("World_Rank","Institution","Location",
   "National_Ranking","Quality_of_Education","Alumni_Employment",
   "Quality_of_Faculty","Research_Output","Quality_Publications",
   "Influence","Citations","Patents","Score")
rankings_2015$Year<-"2015"
rankings_2015$Research_Output <- NULL
rankings_2015$Patents <- NULL
#check data type
sapply(rankings_2015, class)
#Converting int to char to remove + sign
rankings_2015$Quality_of_Education <- as.character(rankings_2015$
   Quality_of_Education)
rankings_2015$Quality_of_Education <- gsub("\\+","",rankings_2015
   $Quality_of_Education)
```

```
#Converting int to char to remove + sign
rankings_2015$Alumni_Employment <- as.character(rankings_2015$
   Alumni_Employment)
rankings_2015$Alumni_Employment <- gsub("\\+","",rankings_2015$
   Alumni_Employment)

#Converting int to char to remove + sign
rankings_2015$Quality_of_Faculty <- as.character(rankings_2015$
   Quality_of_Faculty)
rankings_2015$Quality_of_Faculty <- gsub("\\+","",rankings_2015$
   Quality_of_Faculty)

#Converting int to char to remove + sign
rankings_2015$Quality_Publications <- as.character(rankings_2015$
   Quality_Publications)
rankings_2015$Quality_Publications <- gsub("\\+","",rankings_2015
   $Quality_Publications)

#Converting int to char to remove + sign
rankings_2015$Influence <- as.character(rankings_2015$Influence)
rankings_2015$Influence <- gsub("\\+","",rankings_2015$Influence)

#Converting int to char to remove + sign
rankings_2015$Citations <- as.character(rankings_2015$Citations)
rankings_2015$Citations <- gsub("\\+","",rankings_2015$Citations)

#Converting int to char to remove + sign
rankings_2015$Patents <- as.character(rankings_2015$Patents)
rankings_2015$Patents <- gsub("\\+","",rankings_2015$Patents)

#check data type
sapply(rankings_2015, class)

##########################################################2015
   ##########################################################################


url_file<-"https://cwur.org/2016.php"
world_universities_rankings_2016 <- getURL(url_file)
#rankings <-readHTMLTable(getURL(url_file),which = 1)
rankings_2016 <-readHTMLTable(getURL(url_file),which = 1)
colnames(rankings_2016)<-c("World_Rank","Institution","Location",
   "National_Ranking","Quality_of_Education","Alumni_Employment",
   "Quality_of_Faculty","Research_Output","Quality_Publications",
   "Influence","Citations","Patents","Score")
rankings_2016$Year<-"2016"
rankings_2016$Research_Output <- NULL
rankings_2016$Patents <- NULL
#check data type
sapply(rankings_2016, class)
```

```
#Converting int to char to remove + sign
rankings_2016$Quality_of_Education <- as.character(rankings_2016$
   Quality_of_Education)
rankings_2016$Quality_of_Education <- gsub("\\+","",rankings_2016
   $Quality_of_Education)

#Converting int to char to remove + sign
rankings_2016$Alumni_Employment <- as.character(rankings_2016$
   Alumni_Employment)
rankings_2016$Alumni_Employment <- gsub("\\+","",rankings_2016$
   Alumni_Employment)

#Converting int to char to remove + sign
rankings_2016$Quality_of_Faculty <- as.character(rankings_2016$
   Quality_of_Faculty)
rankings_2016$Quality_of_Faculty <- gsub("\\+","",rankings_2016$
   Quality_of_Faculty)

#Converting int to char to remove + sign
rankings_2016$Quality_Publications <- as.character(rankings_2016$
   Quality_Publications)
rankings_2016$Quality_Publications <- gsub("\\+","",rankings_2016
   $Quality_Publications)

#Converting int to char to remove + sign
rankings_2016$Influence <- as.character(rankings_2016$Influence)
rankings_2016$Influence <- gsub("\\+","",rankings_2016$Influence)

#Converting int to char to remove + sign
rankings_2016$Citations <- as.character(rankings_2016$Citations)
rankings_2016$Citations <- gsub("\\+","",rankings_2016$Citations)

#Converting int to char to remove + sign
rankings_2016$Patents <- as.character(rankings_2016$Patents)
rankings_2016$Patents <- gsub("\\+","",rankings_2016$Patents)

#check data type
sapply(rankings_2016, class)
#############################################################2017
   ##########################################################################

url_file<-"https://cwur.org/2017.php"
world_universities_rankings_2017 <- getURL(url_file)
#rankings <-readHTMLTable(getURL(url_file),which = 1)
rankings_2017 <-readHTMLTable(getURL(url_file),which = 1)
colnames(rankings_2017)<-c("World_Rank","Institution","Location",
   "National_Ranking","Quality_of_Education","Alumni_Employment",
   "Quality_of_Faculty","Research_Output","Quality_Publications",
   "Influence","Citations","Patents","Score")
rankings_2017$World_Rank <- gsub("T(.*)\\%$","",rankings_2017$
   World_Rank)
```

```
rankings_2017$Year<-"2017"
rankings_2017$Research_Output <- NULL
rankings_2017$Patents <- NULL
#check data type
sapply(rankings_2017, class)
#Converting int to char to remove + sign
rankings_2017$Quality_of_Education <- as.character(rankings_2017$
   Quality_of_Education)
rankings_2017$Quality_of_Education <- gsub("\\+","",rankings_2017
   $Quality_of_Education)

#Converting int to char to remove + sign
rankings_2017$Alumni_Employment <- as.character(rankings_2017$
   Alumni_Employment)
rankings_2017$Alumni_Employment <- gsub("\\+","",rankings_2017$
   Alumni_Employment)

#Converting int to char to remove + sign
rankings_2017$Quality_of_Faculty <- as.character(rankings_2017$
   Quality_of_Faculty)
rankings_2017$Quality_of_Faculty <- gsub("\\+","",rankings_2017$
   Quality_of_Faculty)

#Converting int to char to remove + sign
rankings_2017$Quality_Publications <- as.character(rankings_2017$
   Quality_Publications)
rankings_2017$Quality_Publications <- gsub("\\+","",rankings_2017
   $Quality_Publications)

#Converting int to char to remove + sign
rankings_2017$Influence <- as.character(rankings_2017$Influence)
rankings_2017$Influence <- gsub("\\+","",rankings_2017$Influence)

#Converting int to char to remove + sign
rankings_2017$Citations <- as.character(rankings_2017$Citations)
rankings_2017$Citations <- gsub("\\+","",rankings_2017$Citations)

#Converting int to char to remove + sign
rankings_2017$Patents <- as.character(rankings_2017$Patents)
rankings_2017$Patents <- gsub("\\+","",rankings_2017$Patents)

#check data type
sapply(rankings_2017, class)

#-----------------------------------------------END----------
#BINDING ALL THE YEARS
rankings <- rbind(rankings_2013,rankings_2014,rankings_2015,
   rankings_2016,rankings_2017)
library(RODBC)
#########Setting up dB Connectionn string ###############
############## Exporting Refined Data to SQL SERVER######
```

```
sqlConnString <- odbcDriverConnect("Driver=SQL␣Server;Server=
   MOLAP;␣Database=Quality_Of_Life;trusted_connection=true")
sqlSave(sqlConnString,rankings,tablename = "stage_rankings",
   rownames = F )

#Rankings

##Import dataset from https://cwur.org/2013.php
setwd("C:/Users/MOLAP/Desktop/Quality_of_life")
library(readr)
library(XML)
library(RCurl)
url_file<-read_html("https://cwur.org/2013.php")
world_universities_rankings_2013 <- getURL(url_file)
#rankings <-readHTMLTable(getURL(url_file),which = 1)
rankings_2013 <-readHTMLTable(world_universities_rankings_2013,
   which = 1)
colnames(rankings_2013)<-c("World_Rank","Institution","Location",
   "National_Ranking","Quality_of_Education","Alumni_Employment",
   "Quality_of_Faculty","Quality_Publications","Influence","
   Citations","Patents","Score")
rankings_2013$Patents <- NULL
rankings_2013$Year<-"2013"

#-------------------------------------------------
#Converting int to char to remove + sign
rankings_2013$Quality_of_Education <- as.character(rankings_2013$
   Quality_of_Education)
rankings_2013$Quality_of_Education <- gsub("\\+","",rankings_2013
   $Quality_of_Education)
#converting back to int to load
rankings_2013$Quality_of_Education
#Converting int to char to remove + sign
rankings_2013$Alumni_Employment <- as.character(rankings_2013$
   Alumni_Employment)
rankings_2013$Alumni_Employment <- gsub("\\+","",rankings_2013$
   Alumni_Employment)
#converting back to int to load
rankings_2013$Alumni_Employment
#Converting int to char to remove + sign
rankings_2013$Quality_of_Faculty <- as.character(rankings_2013$
   Quality_of_Faculty)
rankings_2013$Quality_of_Faculty <- gsub("\\+","",rankings_2013$
   Quality_of_Faculty)
#converting back to int to load
rankings_2013$Quality_of_Faculty
#Converting int to char to remove + sign
rankings_2013$Quality_Publications <- as.character(rankings_2013$
   Quality_Publications)
rankings_2013$Quality_Publications <- gsub("\\+","",rankings_2013
   $Quality_Publications)
#converting back to int to load
```

```
rankings_2013$Quality_Publications
#Converting int to char to remove + sign
rankings_2013$Influence <- as.character(rankings_2013$Influence)
rankings_2013$Influence <- gsub("\\+","",rankings_2013$Influence)
#converting back to int to load
rankings_2013$Influence
#Converting int to char to remove + sign
rankings_2013$Citations <- as.character(rankings_2013$Citations)
rankings_2013$Citations <- gsub("\\+","",rankings_2013$Citations)
#converting back to int to load
rankings_2013$Citations
#Converting int to char to remove + sign
rankings_2013$Patents <- as.character(rankings_2013$Patents)
rankings_2013$Patents <- gsub("\\+","",rankings_2013$Patents)
#converting back to int to load
rankings_2013$Patents

##########################################################2014
   ########################################################################

url_file<-"https://cwur.org/2014.php"
world_universities_rankings_2014 <- getURL(url_file)
#rankings <-readHTMLTable(getURL(url_file),which = 1)
rankings_2014 <-readHTMLTable(getURL(url_file),which = 1)
colnames(rankings_2014)<-c("World_Rank","Institution","Location",
   "National_Ranking","Quality_of_Education","Alumni_Employment",
   "Quality_of_Faculty","Research_Output","Quality_Publications",
   "Influence","Citations","Patents","Score")
rankings_2014$Research_Output <- NULL
rankings_2014$Patents <- NULL
rankings_2014$Year<-"2014"

#check data type
sapply(rankings_2014, class)
#Converting int to char to remove + sign
rankings_2014$Quality_of_Education <- as.character(rankings_2014$
   Quality_of_Education)
rankings_2014$Quality_of_Education <- gsub("\\+","",rankings_2014
   $Quality_of_Education)

#Converting int to char to remove + sign
rankings_2014$Alumni_Employment <- as.character(rankings_2014$
   Alumni_Employment)
rankings_2014$Alumni_Employment <- gsub("\\+","",rankings_2014$
   Alumni_Employment)

#Converting int to char to remove + sign
rankings_2014$Quality_of_Faculty <- as.character(rankings_2014$
   Quality_of_Faculty)
rankings_2014$Quality_of_Faculty <- gsub("\\+","",rankings_2014$
   Quality_of_Faculty)
```

```
#Converting int to char to remove + sign
rankings_2014$Quality_Publications <- as.character(rankings_2014$
   Quality_Publications)
rankings_2014$Quality_Publications <- gsub("\\+","",rankings_2014
   $Quality_Publications)

#Converting int to char to remove + sign
rankings_2014$Influence <- as.character(rankings_2014$Influence)
rankings_2014$Influence <- gsub("\\+","",rankings_2014$Influence)

#Converting int to char to remove + sign
rankings_2014$Citations <- as.character(rankings_2014$Citations)
rankings_2014$Citations <- gsub("\\+","",rankings_2014$Citations)

#Converting int to char to remove + sign
rankings_2014$Patents <- as.character(rankings_2014$Patents)
rankings_2014$Patents <- gsub("\\+","",rankings_2014$Patents)

#check data type
sapply(rankings_2014, class)
#########################################################2015
   ##########################################################################

url_file<-"https://cwur.org/2015.php"
world_universities_rankings_2015 <- getURL(url_file)

rankings_2015 <-readHTMLTable(getURL(url_file),which = 1)
colnames(rankings_2015)<-c("World_Rank","Institution","Location",
   "National_Ranking","Quality_of_Education","Alumni_Employment",
   "Quality_of_Faculty","Research_Output","Quality_Publications",
   "Influence","Citations","Patents","Score")
rankings_2015$Year<-"2015"
rankings_2015$Research_Output <- NULL
rankings_2015$Patents <- NULL
#check data type
sapply(rankings_2015, class)
#Converting int to char to remove + sign
rankings_2015$Quality_of_Education <- as.character(rankings_2015$
   Quality_of_Education)
rankings_2015$Quality_of_Education <- gsub("\\+","",rankings_2015
   $Quality_of_Education)

#Converting int to char to remove + sign
rankings_2015$Alumni_Employment <- as.character(rankings_2015$
   Alumni_Employment)
rankings_2015$Alumni_Employment <- gsub("\\+","",rankings_2015$
   Alumni_Employment)

#Converting int to char to remove + sign
rankings_2015$Quality_of_Faculty <- as.character(rankings_2015$
```

```
   Quality_of_Faculty)
rankings_2015$Quality_of_Faculty <- gsub("\\+","",rankings_2015$
   Quality_of_Faculty)

#Converting int to char to remove + sign
rankings_2015$Quality_Publications <- as.character(rankings_2015$
   Quality_Publications)
rankings_2015$Quality_Publications <- gsub("\\+","",rankings_2015
   $Quality_Publications)

#Converting int to char to remove + sign
rankings_2015$Influence <- as.character(rankings_2015$Influence)
rankings_2015$Influence <- gsub("\\+","",rankings_2015$Influence)

#Converting int to char to remove + sign
rankings_2015$Citations <- as.character(rankings_2015$Citations)
rankings_2015$Citations <- gsub("\\+","",rankings_2015$Citations)

#Converting int to char to remove + sign
rankings_2015$Patents <- as.character(rankings_2015$Patents)
rankings_2015$Patents <- gsub("\\+","",rankings_2015$Patents)

#check data type
sapply(rankings_2015, class)

##############################################################2015
   ##########################################################################


url_file<-"https://cwur.org/2016.php"
world_universities_rankings_2016 <- getURL(url_file)
#rankings <-readHTMLTable(getURL(url_file),which = 1)
rankings_2016 <-readHTMLTable(getURL(url_file),which = 1)
colnames(rankings_2016)<-c("World_Rank","Institution","Location",
   "National_Ranking","Quality_of_Education","Alumni_Employment",
   "Quality_of_Faculty","Research_Output","Quality_Publications",
   "Influence","Citations","Patents","Score")
rankings_2016$Year<-"2016"
rankings_2016$Research_Output <- NULL
rankings_2016$Patents <- NULL
#check data type
sapply(rankings_2016, class)
#Converting int to char to remove + sign
rankings_2016$Quality_of_Education <- as.character(rankings_2016$
   Quality_of_Education)
rankings_2016$Quality_of_Education <- gsub("\\+","",rankings_2016
   $Quality_of_Education)

#Converting int to char to remove + sign
rankings_2016$Alumni_Employment <- as.character(rankings_2016$
   Alumni_Employment)
```

```
rankings_2016$Alumni_Employment <- gsub("\\+","",rankings_2016$
   Alumni_Employment)

#Converting int to char to remove + sign
rankings_2016$Quality_of_Faculty <- as.character(rankings_2016$
   Quality_of_Faculty)
rankings_2016$Quality_of_Faculty <- gsub("\\+","",rankings_2016$
   Quality_of_Faculty)

#Converting int to char to remove + sign
rankings_2016$Quality_Publications <- as.character(rankings_2016$
   Quality_Publications)
rankings_2016$Quality_Publications <- gsub("\\+","",rankings_2016
   $Quality_Publications)

#Converting int to char to remove + sign
rankings_2016$Influence <- as.character(rankings_2016$Influence)
rankings_2016$Influence <- gsub("\\+","",rankings_2016$Influence)

#Converting int to char to remove + sign
rankings_2016$Citations <- as.character(rankings_2016$Citations)
rankings_2016$Citations <- gsub("\\+","",rankings_2016$Citations)

#Converting int to char to remove + sign
rankings_2016$Patents <- as.character(rankings_2016$Patents)
rankings_2016$Patents <- gsub("\\+","",rankings_2016$Patents)

#check data type
sapply(rankings_2016, class)
#############################################################2017
   ##########################################################################

url_file<-"https://cwur.org/2017.php"
world_universities_rankings_2017 <- getURL(url_file)
#rankings <-readHTMLTable(getURL(url_file),which = 1)
rankings_2017 <-readHTMLTable(getURL(url_file),which = 1)
colnames(rankings_2017)<-c("World_Rank","Institution","Location",
   "National_Ranking","Quality_of_Education","Alumni_Employment",
   "Quality_of_Faculty","Research_Output","Quality_Publications",
   "Influence","Citations","Patents","Score")
rankings_2017$World_Rank <- gsub("T(.*)\\%$","",rankings_2017$
   World_Rank)
rankings_2017$Year<-"2017"
rankings_2017$Research_Output <- NULL
rankings_2017$Patents <- NULL
#check data type
sapply(rankings_2017, class)
#Converting int to char to remove + sign
rankings_2017$Quality_of_Education <- as.character(rankings_2017$
   Quality_of_Education)
rankings_2017$Quality_of_Education <- gsub("\\+","",rankings_2017
```

```
   $Quality_of_Education)

#Converting int to char to remove + sign
rankings_2017$Alumni_Employment <- as.character(rankings_2017$
   Alumni_Employment)
rankings_2017$Alumni_Employment <- gsub("\\+","",rankings_2017$
   Alumni_Employment)

#Converting int to char to remove + sign
rankings_2017$Quality_of_Faculty <- as.character(rankings_2017$
   Quality_of_Faculty)
rankings_2017$Quality_of_Faculty <- gsub("\\+","",rankings_2017$
   Quality_of_Faculty)

#Converting int to char to remove + sign
rankings_2017$Quality_Publications <- as.character(rankings_2017$
   Quality_Publications)
rankings_2017$Quality_Publications <- gsub("\\+","",rankings_2017
   $Quality_Publications)

#Converting int to char to remove + sign
rankings_2017$Influence <- as.character(rankings_2017$Influence)
rankings_2017$Influence <- gsub("\\+","",rankings_2017$Influence)

#Converting int to char to remove + sign
rankings_2017$Citations <- as.character(rankings_2017$Citations)
rankings_2017$Citations <- gsub("\\+","",rankings_2017$Citations)

#Converting int to char to remove + sign
rankings_2017$Patents <- as.character(rankings_2017$Patents)
rankings_2017$Patents <- gsub("\\+","",rankings_2017$Patents)

#check data type
sapply(rankings_2017, class)

#-----------------------------------------------END----------
#BINDING ALL THE YEARS
rankings <- rbind(rankings_2013,rankings_2014,rankings_2015,
   rankings_2016,rankings_2017)
library(RODBC)
#########Setting up dB Connectionn string ###############
############## Exporting Refined Data to SQL SERVER######
sqlConnString <- odbcDriverConnect("Driver=SQL␣Server;Server=
   MOLAP;␣Database=Quality_Of_Life;trusted_connection=true")
sqlSave(sqlConnString,rankings,tablename = "stage_rankings",
   rownames = F )

#Employment rate

#set directory
setwd("C:/Users/MOLAP/Desktop/Quality_of_life")
```

```
library(readxl)
csv_data<-read_xlsx("statista_employment_rate_in_oecd_countries_2
   017.xlsx",skip=4 ,sheet = 2)
#employment rate in %
colnames(csv_data)<-c("Country","Employment_rate")
employment_rate<-data.frame(csv_data)
employment_rate$NA. <- NULL
employment_rate$Year <- "2017"
#sapply(employment_rate, class)
library(RODBC)
sqlConnString <- odbcDriverConnect("Driver=SQL␣Server;Server=
   MOLAP;␣Database=Quality_Of_Life;trusted_connection=true")
try(sqlDrop(ConnString,"stage_employment_rate"),silent=TRUE)
sqlSave(sqlConnString,employment_rate,tablename = "stage_
   employment_rate",rownames = F )

#Countries

library(RODBC)
setwd("C:/Users/MOLAP/Desktop/Quality_of_life")
countries<- read.csv("stage_countries.csv",header = T,na.strings
   = c(""))
countries<-countries[,c(1,4,6,7)]
countries<-countries[-c(which(is.na(countries$region))),]
colnames(countries)<-c("Country_Name","Country_Code","Region","
   Sub_Region")

ConnString <- odbcDriverConnect("Driver=SQL␣Server;Server=MOLAP;␣
   Database=Quality_Of_Life;trusted_connection=true")
try(sqlDrop(sqlConnString,"stage_countries"),silent=TRUE)
sqlSave(sqlConnString,countries,tablename = "stage_countries",
   rownames = F )

#Education Expenditure


#setwd("~/Desktop/Quality_of_life")
#set directory
colsumsetwd("C:/Users/MOLAP/Desktop/Quality_of_life")
library(readr)
########## reading dataset
Readdata<-read.csv("API_SE.XPD.TOTL.GD.ZS_DS2_en_csv_v2_10181208.
   csv",skip=4)
Readdata$Indicator.Name<-NULL
Readdata$Indicator.Code<-NULL
x<-Readdata[c(1:264),c(55:60)]
y<-Readdata[,c(1:2)]
expenditure<-data.frame(y,x)
expenditure
z<-rbind(expenditure$X2012,expenditure$X2013,expenditure$X2014,
   expenditure$X2015,expenditure$X2016,expenditure$X2017)
z<-data.frame(z)
```

```
education_expenditure<-c(z$X1, z$X2, z$X3, z$X4, z$X5, z$X6, z$X7
   , z$X8, z$X9, z$X10, z$X11, z$X12, z$X13, z$X14, z$X15, z$X16,
    z$X17, z$X18, z$X19, z$X20, z$X21, z$X22, z$X23, z$X24, z$X25
   , z$X26, z$X27, z$X28, z$X29, z$X30, z$X31, z$X32, z$X33, z$X3
   4, z$X35, z$X36, z$X37, z$X38, z$X39, z$X40, z$X41, z$X42, z$X
   43, z$X44, z$X45, z$X46, z$X47, z$X48, z$X49, z$X50, z$X51, z$
   X52, z$X53, z$X54, z$X55, z$X56, z$X57, z$X58, z$X59, z$X60, z
   $X61, z$X62, z$X63, z$X64, z$X65, z$X66, z$X67, z$X68, z$X69,
   z$X70, z$X71, z$X72, z$X73, z$X74, z$X75, z$X76, z$X77, z$X78,
    z$X79, z$X80, z$X81, z$X82, z$X83, z$X84, z$X85, z$X86, z$X87
   , z$X88, z$X89, z$X90, z$X91, z$X92, z$X93, z$X94, z$X95, z$X9
   6, z$X97, z$X98, z$X99, z$X100, z$X101, z$X102, z$X103, z$X104
   , z$X105, z$X106, z$X107, z$X108, z$X109, z$X110, z$X111, z$X1
   12, z$X113, z$X114, z$X115, z$X116, z$X117, z$X118, z$X119, z$
   X120, z$X121, z$X122, z$X123, z$X124, z$X125, z$X126, z$X127,
   z$X128, z$X129, z$X130, z$X131, z$X132, z$X133, z$X134, z$X135
   , z$X136, z$X137, z$X138, z$X139, z$X140, z$X141, z$X142, z$X1
   43, z$X144, z$X145, z$X146, z$X147, z$X148, z$X149, z$X150, z$
   X151, z$X152, z$X153, z$X154, z$X155, z$X156, z$X157, z$X158,
   z$X159, z$X160, z$X161, z$X162, z$X163, z$X164, z$X165, z$X166
   , z$X167, z$X168, z$X169, z$X170, z$X171, z$X172, z$X173, z$X1
   74, z$X175, z$X176, z$X177, z$X178, z$X179, z$X180, z$X181, z$
   X182, z$X183, z$X184, z$X185, z$X186, z$X187, z$X188, z$X189,
   z$X190, z$X191, z$X192, z$X193, z$X194, z$X195, z$X196, z$X197
   , z$X198, z$X199, z$X200, z$X201, z$X202, z$X203, z$X204, z$X2
   05, z$X206, z$X207, z$X208, z$X209, z$X210, z$X211, z$X212, z$
   X213, z$X214, z$X215, z$X216, z$X217, z$X218, z$X219, z$X220,
   z$X221, z$X222, z$X223, z$X224, z$X225, z$X226, z$X227, z$X228
   , z$X229, z$X230, z$X231, z$X232, z$X233, z$X234, z$X235, z$X2
   36, z$X237, z$X238, z$X239, z$X240, z$X241, z$X242, z$X243, z$
   X244, z$X245, z$X246, z$X247, z$X248, z$X249, z$X250, z$X251,
   z$X252, z$X253, z$X254, z$X255, z$X256, z$X257, z$X258, z$X259
   , z$X260, z$X261, z$X262, z$X263, z$X264)
education_expenditure<-data.frame(education_expenditure)
expenditure<-expenditure[rep(seq_len(nrow(expenditure)), each=6)
   ,]
expenditure<-cbind(expenditure$Country.Name,expenditure$Country.
   Code,education_expenditure)
# f<-c(2012,2013,2014,2015,2016,2017)
# f<-data.frame(f)
# f<-f[rep(seq_len(nrow(f)), each=264),]
#https://stackoverflow.com/questions/11121385/repeat-rows-of-a-
   data-frame
# f<-data.frame(f)
u <- rep(c(2012,2013,2014,2015,2016,2017), times = 264)
education_expenditure<-cbind(expenditure,u)
colnames(education_expenditure)<-c("Country_Name","Country_Code",
   "Expenditure","Year")
#------------end
# sapply(education_expenditure, FUN=function(x) {sum(is.na(x))})
# sapply(education_expenditure, class)
```

```
education_expenditure$Expenditure[is.na(education_expenditure$
   Expenditure)] <- 0
education_expenditure$Country_Name <- as.character(education_
   expenditure$Country_Name)
education_expenditure$Country_Code <- as.character(education_
   expenditure$Country_Code)
education_expenditure$Expenditure <- as.numeric(education_
   expenditure$Expenditure)
education_expenditure$Year <- as.integer(education_expenditure$
   Year)
#########Setting up dB Connectionn string ############
############## Exporting Refined Data to SQL SERVER######
library(RODBC)
sqlConnString <- odbcDriverConnect("Driver=SQL␣Server;Server=
   MOLAP;␣Database=Quality_Of_Life;trusted_connection=true")
try(sqlDrop(sqlConnString,"stage_education_expenditure"),silent=
   TRUE)
sqlSave(sqlConnString,education_expenditure,tablename = "stage_
   education_expenditure",rownames = F )
```