

**Macro-Economic Analysis Specification Documentation**

**Contents**

[**1. INTRODUCTION** 3](#_Toc118116245)

[**2. OVERALL DESCRIPTION** 4](#_Toc118116246)

[**2.1 DATABASE DESIGN** 4](#_Toc118116247)

[**2.2 OPERATING ENVIRONMENT** 7](#_Toc118116248)

[**2.2 DESIGN AND IMPLEMENTATION CONSTRAINTS** 7](#_Toc118116249)

[**3. System Features** 8](#_Toc118116250)

[**3.1 SQL Media Pack** 8](#_Toc118116251)

[**3.2 Job Design** 8](#_Toc118116252)

[**3.2 Job Implementation** 8](#_Toc118116253)

[**3.2 Report Design** 13](#_Toc118116254)

[**4. System Testing** 14](#_Toc118116255)

[**4.1 Performance Requirements** 14](#_Toc118116256)

[**5. References** 14](#_Toc118116257)

**1. INTRODUCTION**

**1.1 PURPOSE**

The objective of this project is to generate a report using different data sources namely the Corruption Perception Index Data set and World Development Indicators. The data sets are updated annually with the last year data.

**1.2 PROJECT SCOPE**

The project scope has been limited to only relevant information with respect to the different data sources in compliance with the best practices defined below.

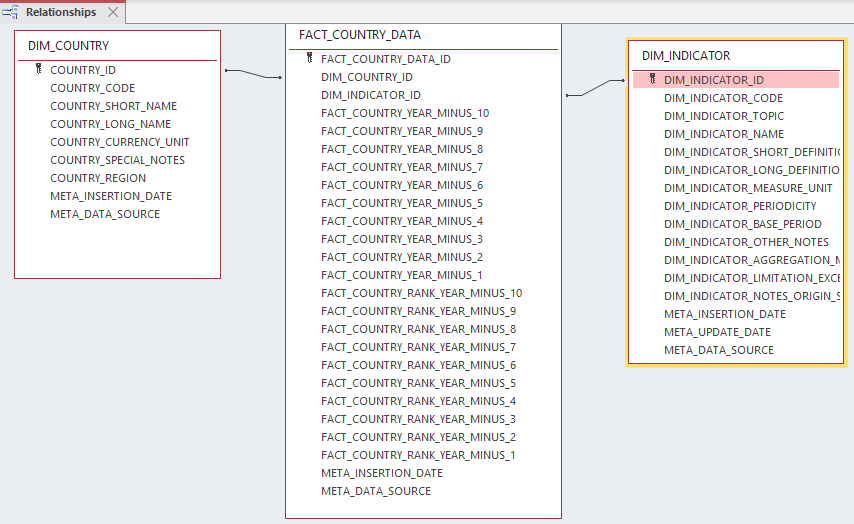
According to the best practices described in the oracle [documentation](https://docs.oracle.com/en/cloud/paas/identity-cloud/uaids/use-best-practices-bulk-loading-data.html#GUID-9E96FDF1-0FCD-4F39-9D84-E9BE7A68D221) the bulk file loading must meet the following specifications:

1. Use a comma as the delimiter between the values
2. Save the file in a CSV format (\*.csv)
3. Limit file size to 52 MB

**2. OVERALL DESCRIPTION**

**2.1 DATABASE DESIGN**

The database comprises of one Fact table (FACT\_COUNTRY\_DATA) with two dimensional tables (DIM\_COUNTRY) and (DIM\_INDICATOR). With increasing file size annually, the database has been designed to cater for the last 10 years data for each country for each indicator.

**

**DIM\_COUNTRY**

|  |  |  |  |
| --- | --- | --- | --- |
| **COLUMN** | **DATA TYPE** | **NULLABLE** | **COMMENTS** |
| COUNTRY\_ID | NUMBER (15,0) | NO (PK) | TECHNICAL UNIQUE IDENTIFIER OF COUNTRY |
| COUNTRY\_CODE | VARCHAR2(20 BYTE) | NO | FUNCTIONAL UNIQUE IDENTIFIER OF COUNTRY |
| COUNTRY\_SHORT\_NAME | VARCHAR2(100 BYTE) | NO | SHORT NAME OF COUNTRY |
| COUNTRY\_LONG\_NAME | VARCHAR2(500 BYTE) | NO | LONG NAME OF COUNTRY |
| COUNTRY\_CURRENCY\_UNIT | VARCHAR2(200 BYTE) | YES | COUNTRY\_CURRENCY\_UNIT |
| COUNTRY\_SPECIAL\_NOTES | VARCHAR2(2000 BYTE) | YES | SPECIAL NOTES |
| COUNTRY\_REGION | VARCHAR2(100 BYTE) | YES | COUNTRY REGION |
| META\_INSERTION\_DATE | DATE | YES | META\_INSERTION\_DATE |
| META\_DATA\_SOURCE | VARCHAR2(20 BYTE) | YES | META\_DATA\_SOURCE |

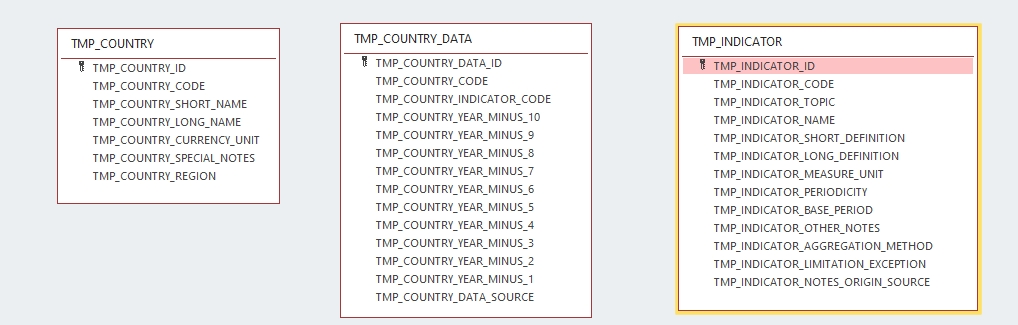
**DIM\_INDICATOR**

|  |  |  |  |
| --- | --- | --- | --- |
| **COLUMN** | **DATA TYPE** | **NULLABLE** | **COMMENTS** |
| COUNTRY\_DIM\_INDICATOR\_ID | NUMBER(15,0) | NO (PK) | TECHNICAL UNIQUE IDENTIFIER OF SERIES |
| DIM\_INDICATOR\_CODE | VARCHAR2(50 BYTE) | YES | FUNCTIONAL UNIQUE IDENTIFIER OF SERIES |
| DIM\_INDICATOR\_TOPIC | VARCHAR2(400 BYTE) | YES | TOPIC OF SERIES |
| DIM\_INDICATOR\_NAME | VARCHAR2(400 BYTE) | YES | NAME OF SERIES |
| DIM\_INDICATOR\_SHORT\_DEFINITION | VARCHAR2(1000 BYTE) | YES | SHORT DEFINITION OF SERIES |
| DIM\_INDICATOR\_LONG\_DEFINITION | VARCHAR2(4000 BYTE) | YES | LONG DEFINITION OF SERIES |
| DIM\_INDICATOR\_MEASURE\_UNIT | VARCHAR2(20 BYTE) | YES | MEASURE UNIT |
| DIM\_INDICATOR\_PERIODICITY | VARCHAR2(40 BYTE) | YES | PERIODICITY |
| DIM\_INDICATOR\_BASE\_PERIOD | VARCHAR2(50 BYTE) | YES | BASE PERIOD |
| DIM\_INDICATOR\_OTHER\_NOTES | VARCHAR2(2000 BYTE) | YES | OTHER NOTES |
| DIM\_INDICATOR\_AGGREGATION\_METHOD | VARCHAR2(50 BYTE) | YES | AGGREGATION METHOD |
| DIM\_INDICATOR\_LIMITATION\_EXCEPTION | VARCHAR2(4000 BYTE) | YES | LIMITATIONS AND EXCEPTION |
| DIM\_INDICATOR\_NOTES\_ORIGIN\_SOURCE | VARCHAR2(4000 BYTE) | YES | NOTES ORIGIN SOURCE |
| META\_INSERTION\_DATE | DATE | YES | DATE UPLODED |
| META\_UPDATE\_DATE | DATE | YES | UPDATED DATE |
| META\_DATA\_SOURCE | VARCHAR2(20 BYTE) | YES | DATA SOURCE |

**FACT\_COUNTRY\_DATA**

|  |  |  |  |
| --- | --- | --- | --- |
| **COLUMN** | **DATA TYPE** | **NULLABLE** | **COMMENTS** |
| FACT\_COUNTRY\_DATA\_ID | NUMBER(15,0) | No | TECHNICAL UNIQUE IDENTIFIER OF COUNTRY DATA |
| DIM\_COUNTRY\_ID | NUMBER(15,0) | No | FOREIGN KEY OF COUNTRY ID |
| DIM\_INDICATOR\_ID | NUMBER(15,0) | No | FOREIGN KEY OF INDICATOR ID |
| FACT\_COUNTRY\_YEAR\_MINUS\_10 | NUMBER(38,5) | Yes | INDICATOR VALUE FOR CURRENT YEAR MINUS 10 |
| FACT\_COUNTRY\_YEAR\_MINUS\_9 | NUMBER(38,5) | Yes | INDICATOR VALUE FOR CURRENT YEAR\_MINUS\_9 |
| FACT\_COUNTRY\_YEAR\_MINUS\_8 | NUMBER(38,5) | Yes | INDICATOR VALUE FOR CURRENT YEAR\_MINUS\_8 |
| FACT\_COUNTRY\_YEAR\_MINUS\_7 | NUMBER(38,5) | Yes | INDICATOR VALUE FOR CURRENT YEAR\_MINUS\_7 |
| FACT\_COUNTRY\_YEAR\_MINUS\_6 | NUMBER(38,5) | Yes | INDICATOR VALUE FOR CURRENT YEAR\_MINUS\_6 |
| FACT\_COUNTRY\_YEAR\_MINUS\_5 | NUMBER(38,5) | Yes | INDICATOR VALUE FOR CURRENT YEAR\_MINUS\_5 |
| FACT\_COUNTRY\_YEAR\_MINUS\_4 | NUMBER(38,5) | Yes | INDICATOR VALUE FOR CURRENT YEAR\_MINUS\_4 |
| FACT\_COUNTRY\_YEAR\_MINUS\_3 | NUMBER(38,5) | Yes | INDICATOR VALUE FOR CURRENT YEAR\_MINUS\_3 |
| FACT\_COUNTRY\_YEAR\_MINUS\_2 | NUMBER(38,5) | Yes | INDICATOR VALUE FOR CURRENT YEAR\_MINUS\_2 |
| FACT\_COUNTRY\_YEAR\_MINUS\_1 | NUMBER(38,5) | Yes | INDICATOR VALUE FOR CURRENT YEAR\_MINUS\_1 |
| FACT\_COUNTRY\_RANK\_YEAR\_MINUS\_10 | NUMBER(15,0) | Yes | INDICATOR RANK FOR CURRENT YEAR MINUS 10 |
| FACT\_COUNTRY\_RANK\_YEAR\_MINUS\_9 | NUMBER(15,0) | Yes | INDICATOR RANK FOR CURRENT YEAR MINUS 9 |
| FACT\_COUNTRY\_RANK\_YEAR\_MINUS\_8 | NUMBER(15,0) | Yes | INDICATOR RANK FOR CURRENT YEAR MINUS 8 |
| FACT\_COUNTRY\_RANK\_YEAR\_MINUS\_7 | NUMBER(15,0) | Yes | INDICATOR RANK FOR CURRENT YEAR MINUS 7 |
| FACT\_COUNTRY\_RANK\_YEAR\_MINUS\_6 | NUMBER(15,0) | Yes | INDICATOR RANK FOR CURRENT YEAR MINUS 6 |
| FACT\_COUNTRY\_RANK\_YEAR\_MINUS\_5 | NUMBER(15,0) | Yes | INDICATOR RANK FOR CURRENT YEAR MINUS 5 |
| FACT\_COUNTRY\_RANK\_YEAR\_MINUS\_4 | NUMBER(15,0) | Yes | INDICATOR RANK FOR CURRENT YEAR MINUS 4 |
| FACT\_COUNTRY\_RANK\_YEAR\_MINUS\_3 | NUMBER(15,0) | Yes | INDICATOR RANK FOR CURRENT YEAR MINUS 3 |
| FACT\_COUNTRY\_RANK\_YEAR\_MINUS\_2 | NUMBER(15,0) | Yes | INDICATOR RANK FOR CURRENT YEAR MINUS 1 |
| FACT\_COUNTRY\_RANK\_YEAR\_MINUS\_1 | NUMBER(15,0) | Yes |  |
| META\_INSERTION\_DATE | DATE | Yes | META\_INSERTION\_DATE |
| META\_DATA\_SOURCE | VARCHAR2(20 BYTE) | Yes | META\_DATA\_SOURCE |

There are also 3 temporary tables that are truncated before each loading annually.

**

|  |  |  |
| --- | --- | --- |
| **Temporary table** | **Data source** | **Comments** |
| TMP\_COUNTRY | WDICountry.csv | Only 6 columns from the csv have been retained for the purpose of this project |
| TMP\_COUNTRY\_DATA | WDIDATA.csv  CPI.csv | Only the last 10 years of data have been retained. |
| TMP\_INDICATOR | WDISeries.csv | Only 12 columns from the csv have been retained for the purpose of this project |

*Note : All these temporary tables are truncated before each loading. The files are saved in UTF8-BOM encoding format.*

**2.2 OPERATING ENVIRONMENT**

Operating environment for the airline management system is as listed below.

* Oracle database
* Talend Open Studio
* Desktop Power BI

**2.2 DESIGN AND IMPLEMENTATION CONSTRAINTS**

The CSV files from the different sources are rendered according to a predefined format and placed in a predefined repository to be read by the Talend Job.

An indicator CPI.INDEX have been created for the purpose of the Indicator Corruption Perception Index Data. A manual entry have been done for this indicator prior to Talend job execution.

**3. System Features**

**3.1 SQL Media Pack**

The database have been designed in the normalized format to redundancies as far as possible. As mentioned earlier, data for only the last 10 years have been retained due to file size constraints. Extendable tablespace and indexes on the different foreign keys have been created on the respective tables for performance tuning.

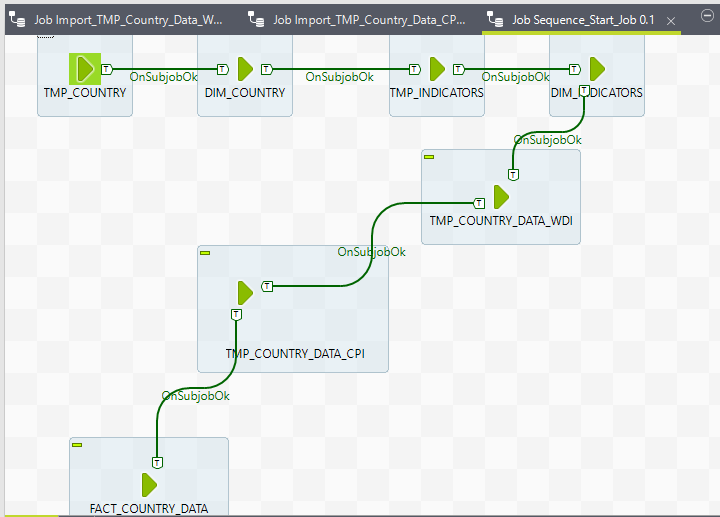
**3.2 Job Design**

The jobs have been designed to comply to the best practices.

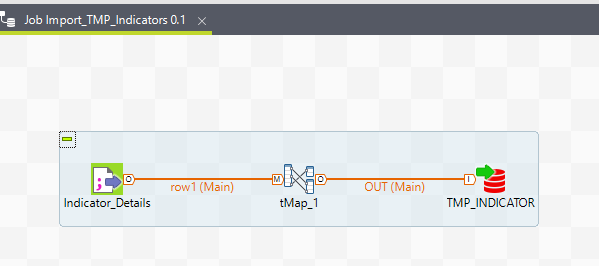
* Using DB connection exported as context file
* Data inconsistences have been identified and data type conversion have been applied wherever applicable
* Errors have been exported in output files.
* Creation of sequence job which will trigger the loading of data in predefined order.

**3.2 Job Implementation**

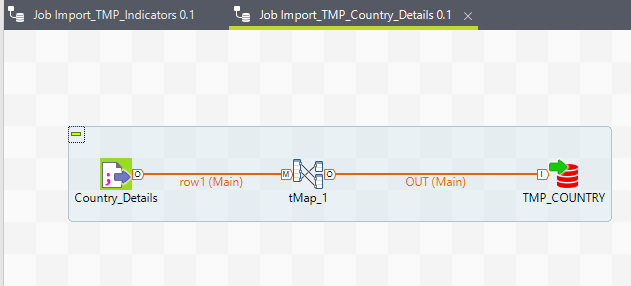
Starting Job



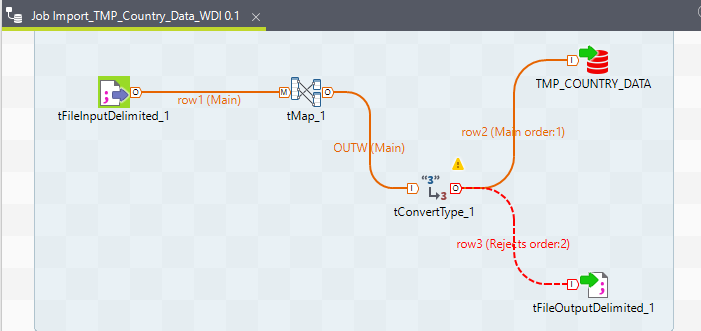
**Tmp Indicators JOB** will load the indicator details in the temporary indicator table. Prior to loading the temporary table has been truncated.



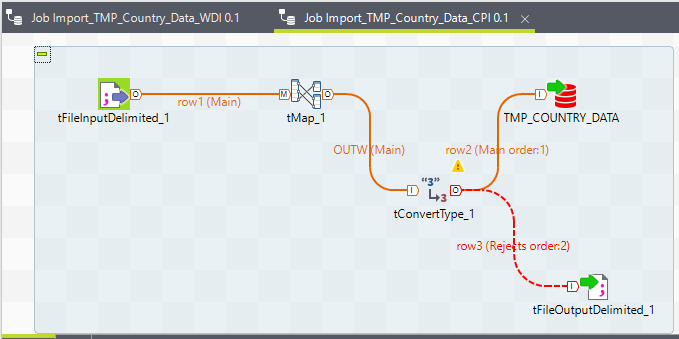
Tmp Country JOB will load the country details in the temporary country table. Prior to loading the temporary table has been truncated.



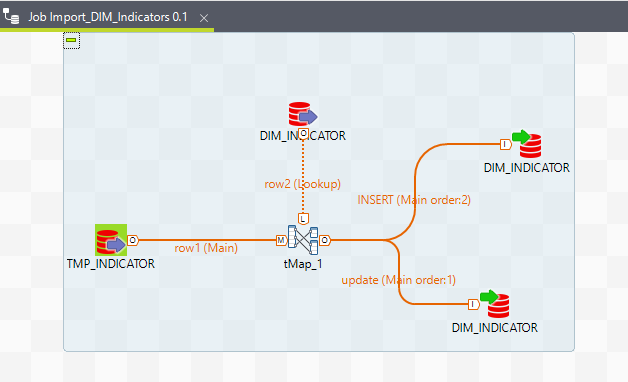
Tmp Country Data job will load the country details in the temporary country data table. Prior to loading the temporary table has been truncated. For this job the data source is the World development Indicators. All incorrect data will be reported in the rejects file.



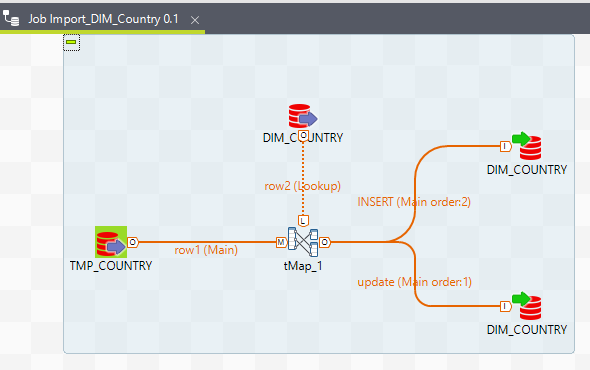
**Tmp Country Data job** will load the country details in the temporary country data table. Prior to loading the temporary table has been truncated. For this job the data source is the Corruption Perception Index Data Set. All incorrect data will be reported in the rejects file.



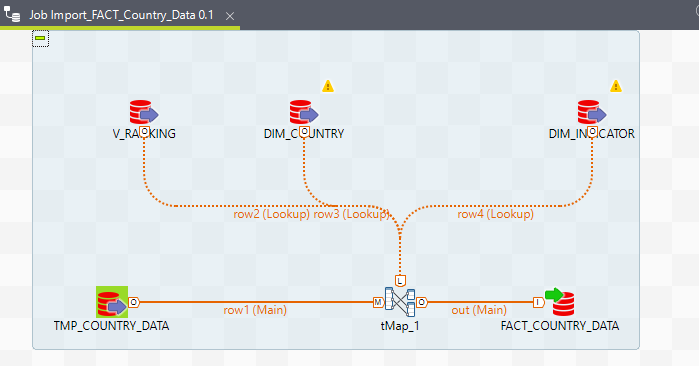
**Dim Indicator Job** will load the indicator details from the temporary table to dimension indicator table. Prior to insertion, data will be validated. In case the data already exists, same will be updated with the last update date else new data will be inserted in the table.



**Dim Country Job** will load the country details from the temporary table to dimension country table. Prior to insertion, data will be validated. In case the data already exists, same will be updated with the last update date else new data will be inserted in the table.



Fact Country Data will load the country details from the temporary table to fact country data table. Prior to insertion, the corresponding foreign keys will be retrieved for country and indicator. The ranking for each indicator will be calculated for each year and inserted in the fact table. Historical data will be kept. Each year the last 10 years will be uploaded in the database.



**3.2 Report Design**

Below is a snapshot of the report dashboard. The report has been developed using Desktop power BI and Access database. Normally the report should connect to the oracle DB directly. Due to missing data access component on my machine, the data from Oracle DB have been exported to Access DB.

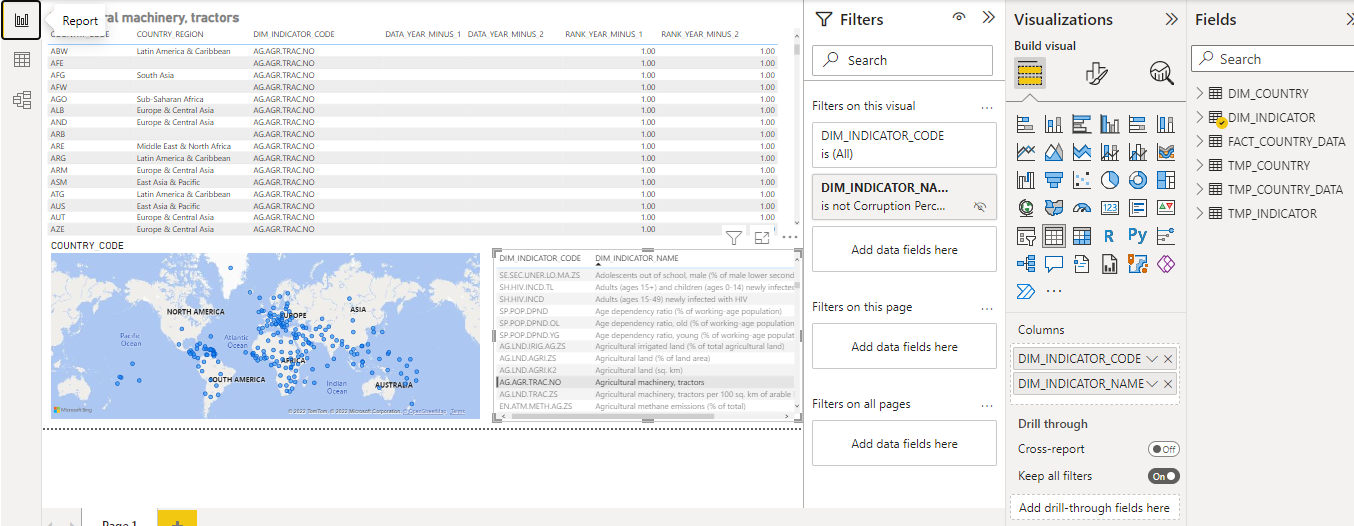
The data access component has not been installed due to lack of admin privilege on my machine and time constraint for support to do the needful.

The report is self-explanatory. The user can view the data for the past 1o years.

To this project, a dashboard has been created with 3 visuals. On the top, a tabular format of the report with country code, region, indicator code, data for current year minus 1 , data for current year minus 2 and the ranking for the respective years for the respective indicators.

Below is a list of indicators which will change the tabular report accordingly upon selection.

At the bottom geographical location of the different countries.



**4. System Testing**

**4.1 Performance Requirements**

* Successful execution of the sql commands in oracle DB
* Successful data loading from CSV files to the oracle database. No loss of data. All the lines in the CSV files have been loaded in the respective tables.
* Ranking for each indicator for each year have been calculated in the job.
* No data loss have been detected.
* Data inconsistencies have been identified.   
  3 missing countries have been detected in the World Development Indicators.   
  The 3 countries have been added manually as remedial actions for foreign keys errors.
* The report has been developed using desktop Power BI to meet and exceed the requirement specifications.

**5. References**

1. <https://docs.oracle.com/en/cloud/paas/identity-cloud/uaids/use-best-practices-bulk-loading-data.html#GUID-9E96FDF1-0FCD-4F39-9D84-E9BE7A68D221>
2. <https://github.com/schinien/MCBConsultantAssessment>