

World Bank Data Dictionary – for use with Module 4 “World Bank Mini-Analysis”

Overview

This document provides the necessary details documenting the source data files that are a subset of columns extracted from original World Bank Indicator datasets. The World Bank provides free access and use of a wide variety of datasets spanning many areas of interest including health, demography, business and socio-economics, and agriculture. Our mini-analysis uses only two years of data extracted from the original source data (2000 and 2019) in order to limit complexity.

Links to the original source data from the World Bank datasets are available below.

- [Mortality caused by road traffic injury \(per 100,000 population\) | Data \(worldbank.org\)](#)
- [Mobile cellular subscriptions \(per 100 people\) | Data \(worldbank.org\)](#)
- The geocodes dataset is a reduced extraction of the geocodes provided in the indicator datasets.

You can visit [World Bank Open Data | Data](#) for more details on other dataset availability.

Data Details

The details below describe the contents of each .csv file used as source data and the requirements for constructing the database and table schema and table relationships.

Data (.csv) details:

| | |
|----------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|
| Country/Region Codes – filename: WBGeoCodes.csv | |
| Data set contains the ISO-3 character geographic codes and descriptions for 218 individual countries and 48 world regions. | |
| Data Source | World Development Indicators |
| <u>ColumnName</u> | <u>Column and value description</u> |
| CountryRegion3Code | ISO-3-Character Country/Region Code |
| CountryRegionName | The descriptive name of the country or region |
| IsCountry | Boolean (binary) value indicating 0=Not a specific country // 1=Is a specific country |

| | |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------|
| Mobile cellular subscriptions (per 100 people) – filename: WBCellPhoneUse.csv | |
| Data set contains Mobile cellular telephone subscriptions for 218 individual countries and 48 world regions (years 2000 and 2019) that are subscriptions to a public mobile telephone service that provide access to the PSTN using cellular technology. The indicator includes (and is split into) the number of postpaid subscriptions, and the number of active prepaid accounts (i.e. that have been used during the last three months). The indicator applies to all mobile cellular subscriptions that offer voice communications. It excludes subscriptions via data cards or USB modems, subscriptions to public mobile data services, private trunked mobile radio, telepoint, radio paging and telemetry services. | |
| Data Source | World Development Indicators |
| <u>ColumnName</u> | <u>Column and value description</u> |
| CountryRegion3Code | ISO-3-Character Country/Region Code |
| IndicatorName | The descriptive name of the indicator |
| IndicatorCode | World Development Indicator Code |
| Y2000 | Cell Phone subscriptions per 100 people for the year 2000 (0 or "blank" can be treated as missing) |
| Y2019 | Cell Phone subscriptions per 100 people for the year 2019 (0 or "blank" can be treated as missing) |

| | |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------|
| Mortality caused by road traffic injury (per 100,000 population) – filename: WBTrafficFatalities.csv | |
| Data set contains mortality caused by road traffic injury (per 100,000 population) for 218 individual countries and 48 world regions for years 2000 and 2019. | |
| Data Source | World Health Organization, Global Health Observatory Data Repository (http://apps.who.int/ghodata/). |
| <u>ColumnName</u> | <u>Column and value description</u> |
| CountryRegion3Code | ISO-3-Character Country/Region Code |
| IndicatorName | The descriptive name of the indicator |
| IndicatorCode | World Development Indicator Code |
| Y2000 | Mortality caused by road traffic injury (per 100,000 population) in year 2000 (0 or "blank" can be treated as missing) |
| Y2019 | Mortality caused by road traffic injury (per 100,000 population) in year 2019 (0 or "blank" can be treated as missing) |

Database Details:

| | | |
|---------------------------|----------------------|-----------------|
| <u>Database structure</u> | | |
| Database Name | WBIndicatorsDataBase | |
| | | |
| TableName | WBGeoCodes | |
| <u>ColumnType</u> | <u>ColumnName</u> | <u>DataType</u> |
| primary key | CountryRegion3Code | text |
| value | CountryRegionName | text |
| value | IsCountry | integer |
| | | |
| TableName | WBIndicatorValues | |
| <u>ColumnType</u> | <u>ColumnName</u> | <u>DataType</u> |
| primary key/foreign key | CountryRegion3Code | text |
| primary key | IndicatorCode | text |
| primary key | IndicatorName | text |
| primary key | Year | integer |
| value | Value | real |

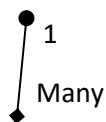
A primary key comprised of multiple fields is required for this schema.

Table Relationship Requirements:

In order to effectively store the data, the WBGeoCodes and WBIndicatorValues tables must be designed to enforce referential integrity at the schema level. This permits storing just one instance of the geocode details and relate them to the shared “key” in the indicator data. The diagram below describes the requirements for correctly defining the foreign keys in each table and enabling this cross-table reference so a 1:many relationship is established and enforced.

Relational Structure Diagram

WBGeoCodes (CountryRegion3Code, CountryRegionName, IsCountry)



WBIndicatorValues (CountryRegion3Code, IndicatorCode, IndicatorName, Year, Value)

Note: Italic, double underline denotes contributing to foreign key relationship. Primary is denoted by bold.