

Group 1 - Entity Relationship Diagram (ERD) Description

Master's in Business Analytics – Culminating Project

Week of: 2

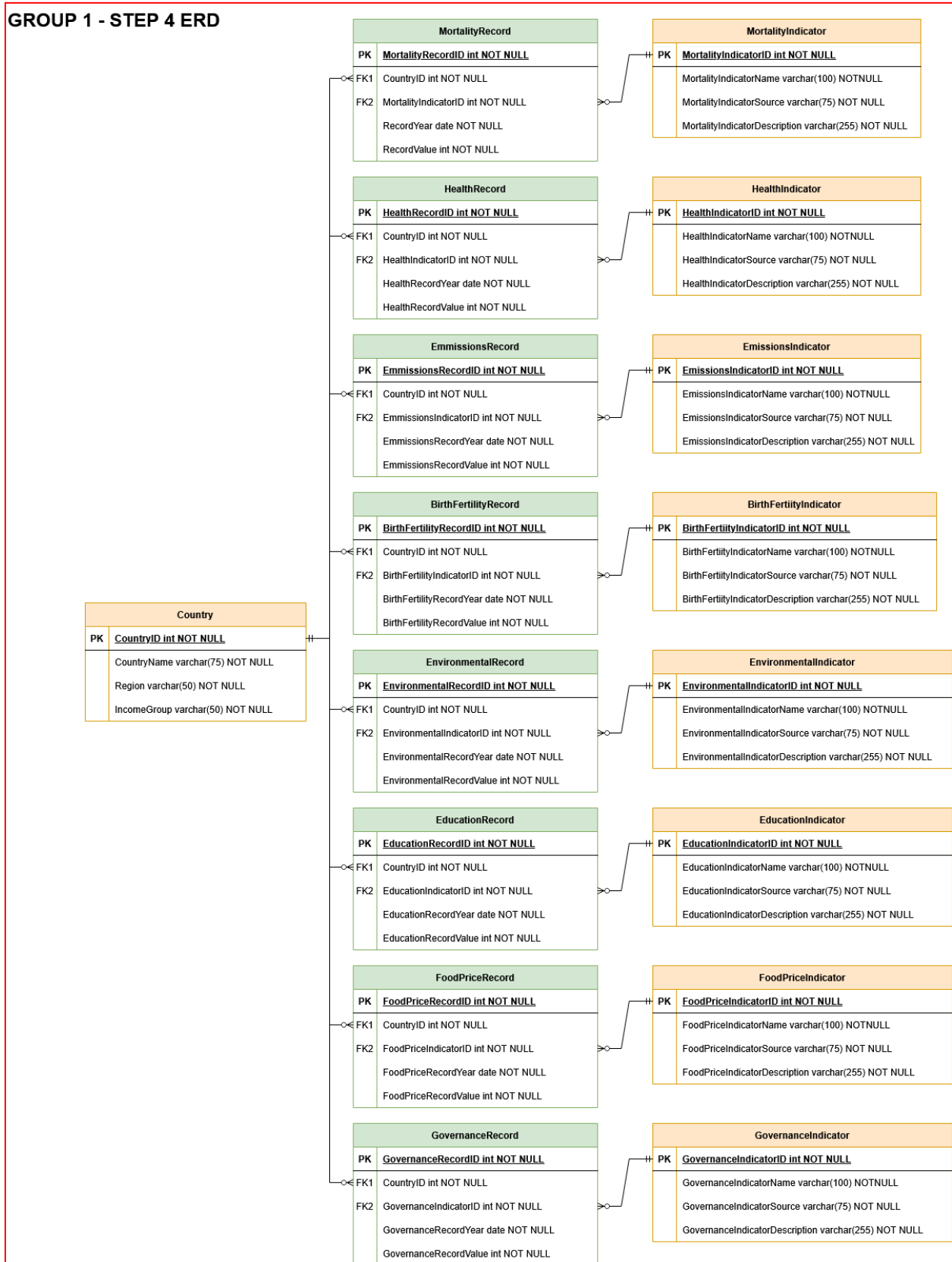
Joshua Cabal, Namrata Patil, Christopher Yeung, Celine Abrahamian

Step 2 General ERD Description:

This ERD represents three databases: Food Prices for Nutrition, World Development Indicators, and World Governance Indicators. We chose these databases because they contain indicators for our areas of interest: education, food, governance, and mortality. We found indicators related to education and mortality in the World Development Indicators database, food cost and food insecurity indicators in the Food Prices for Nutrition database, and government corruption, stability, efficacy, and law indicators in the World Governance Indicator database.

For this ERD, rather than list every single indicator from each database in their perspective database entity boxes as attributes (which would require several pages to cover the almost 1,500 indicators in the World Development Indicators database), we instead opted to depict general umbrella topics of only the indicators that are related to our four areas of interest and treated these subcategories as attributes in their associated database entity. For example, indicators related to gender disparity in education are represented in the World Development Indicator entity as the attribute "GenderDisparities," and indicators related to political instability are represented in the World Governance Indicator entity as the attribute "PoliticalInstability."

Figure: Refined Entity Relationship Diagram



Step 4 Refined ERD Description:

The **Country** table will store all countries available from the databases. Below is the data description table for this entity. These columns were chosen because they are metadata which exist for all countries in the World Bank website.

Column Name	Description
CountryID	PK
CountryName	Full name of the country
IncomeGroup	Can attain one of the following values: Low, Lower middle, Upper middle, High, Not classified
Region	Can attain one of the following values: South Asia, Europe & Central Code, Middle East & North Africa, East Asia & Pacific, Sub-Saharan Africa, Latin America & Caribbean, North America

The **Indicator** tables will store all the available indicators for that specific series group. As in, each table is the selected list of indicators from the actual exported database / data report. There will be 30+ records within each indicator table.

The **Record** tables will store the values for each unique combination of indicator, country, and year. The cardinality for these tables is as follows: each record must have one and only one country and indicator related to it. However, a given country may not have any records for a given indicator.