University of Ottawa
School of Electrical Engineering and Computer Science

CSI4142 Fundamentals of Data Science

Project Phase 1: Conceptual Design – Dimensional Model

Instructions:

This is a team assignment. Submit your conceptual design using your group locker in Bright Space. Your uploaded document should be in PDF format.

All teams are also required to meet with a TA during a scheduled Zoom meeting, to discuss your design. All students are required to turn their cameras on during these meetings, to enable us to verify your identity. Please have your student card and/or a valid photo id (e.g., a driver's licence) ready for inspection.

Project Description - World Bank Health Nutrition and Population Statistics

Data science and Artificial Intelligence (AI) have been very successful to discover important trends in data over time. Increasingly, to organizations such as the World Bank provide access to open-source repositories for data analytics and data mining, in order to enable data scientists to use these resources in their individual projects.

Specifically, the World Bank Health Nutrition and Population Statistics database "provides key health, nutrition and population statistics gathered from a variety of international and national sources. Themes include global surgery, health financing, HIV/AIDS, immunization, infectious diseases, medical resources and usage, noncommunicable diseases, nutrition, population dynamics, reproductive health, universal health coverage, and water and sanitation" [1].

This repository has been used in numerous studies, including a study on adaptation to sea level rise in Ho Chi Minh City, Vietnam [2], forecasting healthcare demand and spending in Singapore [3], increasing maize nutrition uptake for increased food security [4], and reflecting on the impact of aging populations [5].

The database includes data from across the globe, collected quarterly from 1960 to 2020, involving more than 240 countries. As such it provides us with a rich repository to track and to detect numerous trends. Specifically, the data may be grouped into *health*

indicators such as the number of people living with a disease such as tuberculosis, *population statistics* including births, ages, and genders, *living conditions* such as access to drinking water and sanitation, *levels of employments*, *educational statistics* like school enrollment and literacy rate, as well as important *quality of life* indicators including access to medical care and physicians, amongst many others.

While this repository contains many interesting attributes, a major drawback is that it does not contain any information about environmental or geopolitical events. As such, the data may be enriched by considering major incidents such as rainfall after a drought, or a peaceful transition after armed conflict. Data about such occurrences may be obtained from external sources and would considerably aid data scientist and experts in understanding and interpreting events.

Suppose that your team was hired to design and to implement a data mart to gather not only the data as contained in the original repository, but also to include other "external" sources.

As a first step, your task is to complete an initial conceptual design for an enriched World Bank Health Nutrition and Population data mart.

Specifically, your design should address the following questions:

1. What dimensions and attributes should we include to identify and to track trends, in terms of health, nutrition and population, over time, in individual countries, regions and continents?
   Our aim is to obtain clear profiles of countries that have experienced increases (or decreases) in so-called "key indicators" such as the number of births. Further, we are interested in the details about countries that experienced growth (or not). We also wish to track major occurrences, incidents, experiences and events (e.g., disease outbreaks or sudden increases (or decreases) in levels of unemployment) and to determine the causes and effects.
2. Following 1, what key indicators do we need to store to obtain a clear picture of trends in countries, regions, and continents? For instance, the number of births may be used as a key indicator, but there are additional attributes that should be added. (Note that the key indicators are modelled as "facts/measures" in our data mart, and they will be used in online analytic processing (OLAP) queries.)
3. What were the impacts of specific environmental incidents (e.g., earthquakes or droughts) on key indicators (such as the number of births)?
4. How (and when) did geopolitical events (e.g., peace agreements or social unrest), affect key indicators (such as the number of births)?

5. Are there any additional external sources, other than environmental or geopolitical data, that could potentially be added to the dataset, to further enrich it?

Deliverables:

Create a dimensional model detailing your initial design of the **enriched World Bank Health Nutrition and Population data mart**.

Your PDF document should include the following details.

1. Declare the grain of your data mart.
2. Detail all the dimensions and dimensional attributes. You should list the domains and show sample values. (*e.g., Age: integer, minimum = 0 and maximum = 130, Sample value = 35*).
3. Detail all the measures/facts. You should list the domains and sample values. (*e.g., Age: integer, minimum = 0 and maximum = 130, Sample value = 35*).
4. Remember to detail all your assumptions.
5. Add a list of additional references you used when creating your model.
6. Submit a summary of your team's work plan, including the times and dates you met and how you divided the work.

Links to sources:

Below a list of the sources used in this document. Note that this list is not complete and that you are encouraged to expand your search.

[1] https://datacatalog.worldbank.org/search/dataset/0037652/Health-Nutrition-and-Population-Statistics
[2] https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2017WR021344
[3] https://onlinelibrary.wiley.com/doi/full/10.1111/rode.12528
[4] https://academic.oup.com/jxb/article/67/12/3763/2885016?login=true
[5] https://direct.mit.edu/adev/article/33/2/56/9903/Population-Aging-and-Potential-Growth-in-Asia