University of Ottawa

School of Electrical Engineering and Computer Science

CSI4142 Fundamentals of Data Science

Project Phase 3: OLAP Queries and BI Dashboard

Instructions:

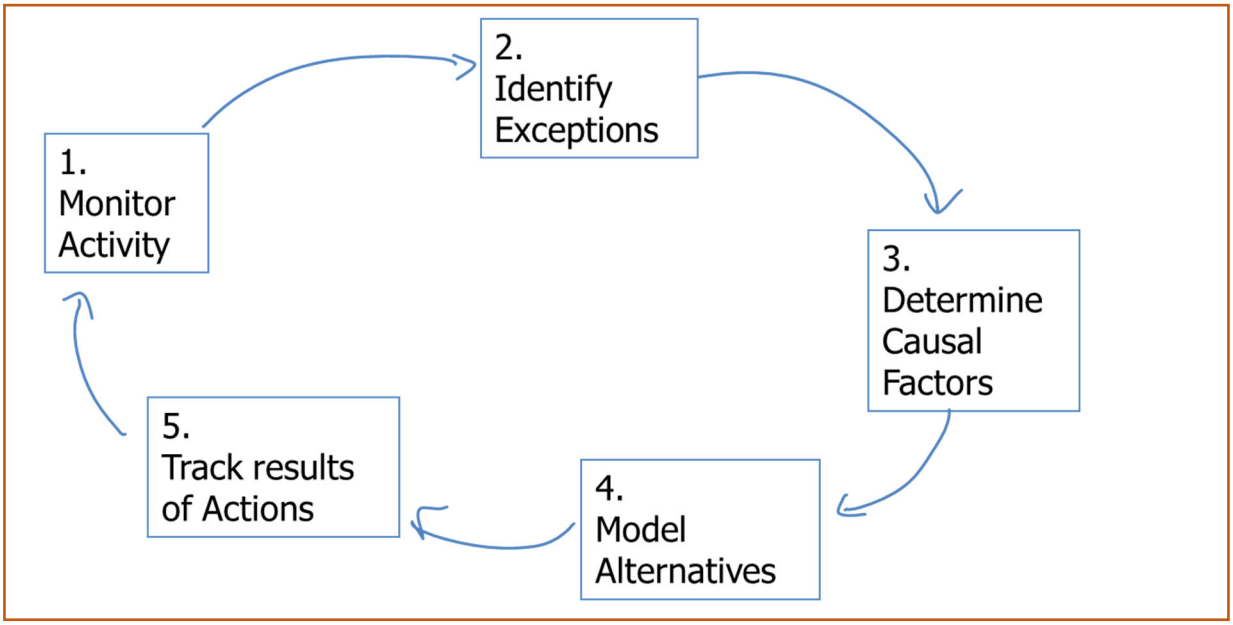
A. This is a team assignment.
B. Submit your documentation via BrightSpace using your team locker.
C. For your source code, you may either submit a zipped file or provide a link to a GitHub repository. You are asked to submit the following details:
   i) the scripts to execute the SQL queries
   ii) screen shots of your Business Intelligence (BI) Dashboard that show the functionality
D. Demonstrate your work during a Zoom meeting with the TA, in the timeslot allocated to you. Note that all team members are required to attend this demonstration and you will be asked to turn your cameras on.

Project Description - World Bank Data Mart

Data Science and Artificial Intelligence (AI) have been very successful to discover important trends in data over time. Increasingly, organizations such as the World Bank provide access to open-source repositories for data analytics and data mining, to enable data scientists to use these resources in their individual projects. Specifically, the World Bank Health Nutrition and Population Statistics (WB-HNP) 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].
During Phase 2 of the project, you created the data mart and staged the data of nine (9) countries of your choice. In this deliverable, you are required to explore your data mart using online analytics processing (OLAP) queries and to create a BI Dashboard.

Part A. OLAP queries (60 marks – 12 queries)

Write OLAP queries exploring the data to answer questions posed during the typical analytical lifecycle as covered in class and as shown below.



You should include a total of 12 queries, in the categories as shown below. (The examples are shown to illustrate the concepts. Teams are free to use their own examples.)

Part 1. Standard OLAP operations – 9 queries in total

a. **Drill down and roll up. – 2 queries** by using concept hierarchies in your data mart, such as (name, region, continent) and (month, quarter, year, decade).
b. **Slice,** where only one dimension is selected. **– 1 query**
For instance, contrast (i) the prevalence of health conditions, (ii) the literacy rates, or (iii) the life expectancies in your nine (9) countries.
c. **Dice,** where one creates a sub-cube. **– 2 queries**
For instance, contrast (i) the prevalence of health conditions, (ii) the literacy rates, or (iii) the life expectancies in Canada versus Mexico.

d. **Combining OLAP operations.** In these queries, we combine the above-mentioned operations. **– 4 queries.** For instance, we may explore the data characteristics i) during different time periods, ii) when certain events were taking place, iii) for different countries and regions, iv) while comparing age groups, or v) contrasting unemployment rates.

Part 2. Explorative operation – 3 queries

Identify general trends using advanced SQL operations. Give *one query* from each one of these categories.

a. **Iceberg queries.** For instance, i) find the five years with the highest population growths, ii) find the five countries with the highest decreases in term of specific health conditions (e.g., tuberculosis) in subpopulations {children, male, female, total} when considering a particular decade.
b. **Windowing queries.** For instance, display the ranking of the countries in terms of the literacy rates, as reported per gender, over the last five years.
c. **Using the Window clause.** For instance, compare the number of hospital beds in Canada in 2019 to that of the previous and next years.

Note: Refer to the Module 4 Data Analytics lecture slides. The PostgreSQL syntax is available at:
https://www.postgresql.org/docs/current/queries-table-expressions.html
https://www.postgresql.org/docs/current/tutorial-window.html
https://www.postgresql.org/docs/current/sql-expressions.html#SYNTAX-WINDOW-FUNCTIONS

Part 3. BI dashboard and Information Visualisation (40 marks)

Create a dashboard that allows the users to explore the data and to visualise trends. Specifically, users should be able to traverse concept hierarchies, including the ability to roll up and drill down, slice and dice, as well as execute Top N or Bottom N queries. Your interface should include graphs and charts. You are encouraged to use Tableau, or any other dashboard tool of your choice.

Reference:

[1]https://datacatalog.worldbank.org/search/dataset/0037652/Health-Nutrition-and-Population-Statistics