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 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 - Covid-19 Tracking and Lifestyle Trends Data Mart

Data science and artificial intelligence (AI) have been successfully used to study trends in our behaviours over time. However, our daily routines changed abruptly with the onset of the COVID-19 pandemic. In Canada, and many other countries, lockdown procedures were implemented, thus leaving citizens with little choice to adapt their lifestyles accordingly. For instance, people increasingly turned to online shopping, while participation in outdoor activities increased. Many non-essential businesses, notably in the hospitality sector, also adapted by offering only delivery or curbside pickup, leading to changes in consumer behaviour and traffic patterns.

During Phase 2 of the project, you created the database and staged the data.

In this deliverable, you are required to explore the data using OLAP queries and to create a Business Intelligence Dashboard. Refer to the lecture slides, as well as the practical lecture as presented by the TA.

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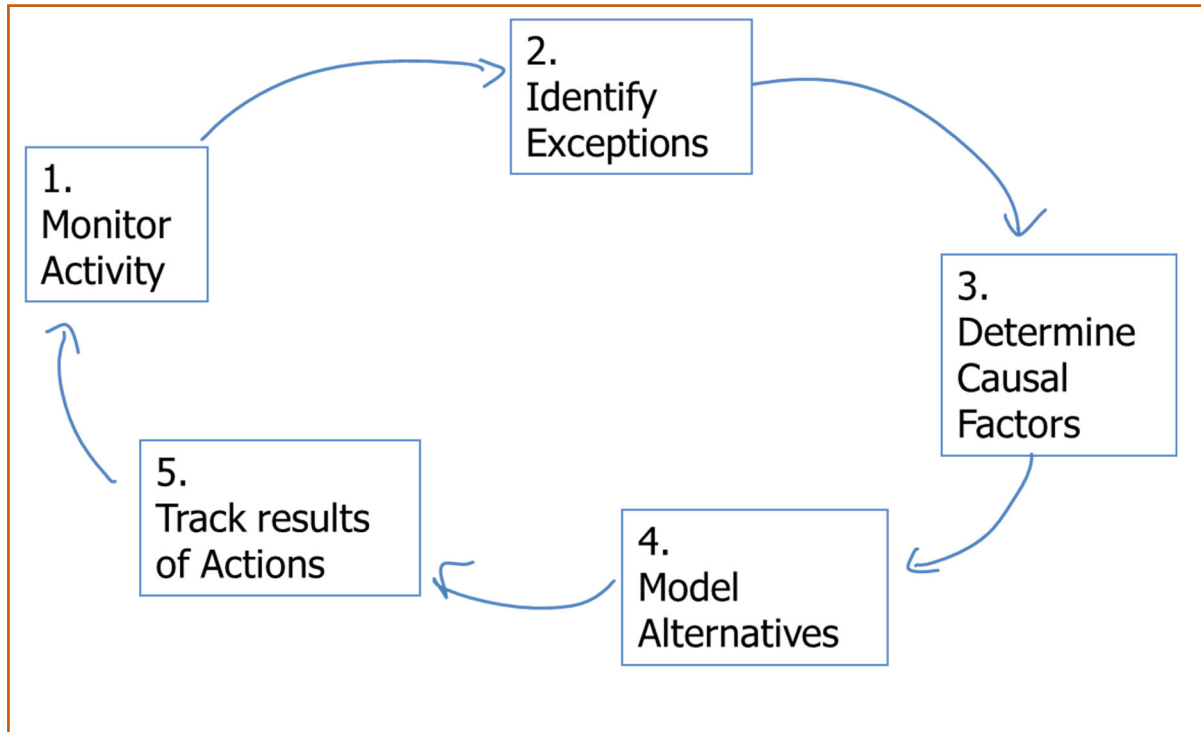
You should include a total of 12 queries, in the categories as shown below. (The examples are shown to illustrate the concepts. Teams are free to use their own examples.)

Part 1. Standard OLAP operations – 9 queries in total

   a. **Drill down and roll up. – 2 queries**
      i   For instance, explore the total number of positive cases in your data mart; drill down to a month (April 2020), and drill down to a specific day.
      ii For instance, explore the total number of resolved cases in your data mart; drill down to a week (first week of April 2020), and drill down to a specific day.
      iii For instance, consider all the unresolved cases in Toronto City, roll up to GTA, and roll up to all data in your data mart.
   b. **Slice,** where only one dimension is selected. **– 2 queries**
      For instance, provide the (i) the number of cases in a specific PHU (resolved, unresolved and fatal), (ii) the number cases across the PHUs when a specific special measure was in place, (iii) the mobility levels in Ottawa, etc.

c. **Dice,** where one creates a sub-cube. **– 2 queries**
   For instance, (i) provide the number of fatal cases during a period of two months, e.g., February and March, in Peel and Ottawa, (ii) provide the number of unresolved cases when contrasting two mobility locations, e.g., parks and transit, in Peel and Ottawa.
d. **Combining OLAP operations.** In these queries, we combine the above-mentioned operations. **– 3 queries**
   For instance, we may aim to explore the number of cases i) during different periods of the year, ii) when certain types of measures are in place, iii) for different types of outcomes and weather conditions iv) contrasting mobility levels in Ottawa and Peel, v) comparing sunny versus rainy days, etc.

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, find the five days with the highest numbers of resolved outcomes, find the location with the highest mobility in terms of visits to parks, to grocery stores and pharmacies, etc.
b. **Windowing queries.** For instance, show the ranking of the PHUs in terms of the number of cases per week, per outcome, per month, etc.
c. **Using the Window clause.** For instance, compare the number of resolved cases in Ottawa to that of the previous and next months, etc.

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. Your interface should include graphs and charts. You are encouraged to use Tableau, or any other dashboard tool of your choice.

Note: Refer to the practical lecture as presented on 12 March 2021.