

**University of Ottawa**  
**School of Electrical Engineering and Computer Science**  
**CSI4142 Fundamentals of Data Science**  
**Project Description**

**Instructions:**

The project is an extensive endeavour involving the design and implementation of a data mart, complemented by exploration through various data science techniques and analytical tools. The project is segmented into several phases, each with specific deliverables:

1. **Project Planning and Group Registration** (Deliverable A):
  - Teams, consisting of three students each, are formed.
  - Teams must register their groups and outline their project approach.
2. **Conceptual Design** (Deliverable B):
  - Teams are tasked with designing an initial conceptual model for an enriched data mart.
  - This phase includes defining dimensions, attributes, and key indicators essential for tracking and analyzing trends over time. An example includes tracking financial trends by incorporating dimensions like time, geography, financial products, and customer demographics.
  - Teams are required to comprehensively detail all dimensions, measures/facts, and assumptions. They must also create a checklist to ensure avoidance of common design mistakes.
3. **Physical Design & Data Staging** (Deliverable C):
  - This phase involves translating the conceptual design into a physical database structure. It includes the selection of suitable database systems and technologies, creation of tables, and establishment of relationships between different data entities.
  - Teams engage in data staging, encompassing Extract, Transform, Load (ETL) processes. This involves extracting data from various sources, such as internal databases, external APIs, or open-source datasets. The transformation step ensures data is cleaned, normalized, and formatted correctly, and the load phase involves populating the data mart with this processed data.
4. **Data Analytics, OLAP Queries & BI Dashboard** (Deliverable D):
  - Teams utilize Online Analytical Processing (OLAP) for multi-dimensional data analysis. This includes crafting OLAP queries to perform operations such as drill-down, roll-up, slice, dice, and pivot to analyze data from

different perspectives. For example, a team might analyze sales data to compare performance across various regions or time periods.

- Additionally, teams are responsible for developing a Business Intelligence (BI) dashboard for data visualization and trend identification. Tools like Power BI, Tableau, or Apache Superset may be used to create interactive dashboards that allow users to explore data through charts, graphs, and maps. A well-constructed dashboard can highlight key performance indicators, uncover historical trends, and aid in decision-making processes.

5. **Data Mining** (Deliverable E):

- The final phase involves applying data mining techniques to discover patterns, classify data, and predict trends. Teams might use algorithms for classification, clustering, or anomaly detection to glean meaningful insights from the data. For instance, a machine learning model could be trained to predict customer churn or to segment customers based on purchasing behavior.
- This phase also includes exploring and summarizing data, preprocessing data, and selecting features for further analysis.

6. **Project Quiz** (Deliverable F):

- A quiz or presentation is scheduled to demonstrate the work and knowledge accumulated throughout the project.
- This is planned for April 10, 2024, with confirmation pending.