WIE3008
Assignment 2
30 Marks

This assignment assesses your proficiency in three critical aspects of data analysis: Data Integration, Online Analytical Processing (OLAP), and Data Modeling. As a team of five, you will integrate two distinct datasets, provide recommendations for OLAP analytics, and develop a model using SAS (in at least one of the steps). Your findings and strategies will be presented in a **PowerPoint slide format**. Due date: 11/1/2024

Part 1: Data Integration (10 Marks)

Task 1: Dataset Selection and Description (2 Marks)

- Select two datasets from varied sources. Choose a sizable dataset with a range of data types from two data sources, e.g. COVID-19 vaccination and the COVID-19 deaths. Example dataset link: https://www.kaggle.com/code/franklinposso/sql-project-exploratory-analysis-covid-19
- Describe each dataset, including its source, structure, and the type of data it contains.

Task 2: Data Cleaning and Preprocessing (3 Marks)

- Execute necessary data cleaning and preprocessing steps for both datasets.
- Document these steps, explaining their importance and relevance.

Task 3: Data Integration (5 Marks)

- Integrate the two datasets into a single comprehensive dataset.
- Explain the integration technique used (e.g., merging, joining, concatenation), and justify your selection.
- Discuss any challenges encountered during the integration process and the solutions applied.

Part 2: OLAP Analytics (10 Marks)

Task 4: OLAP Analysis Plan (5 Marks)

- Develop a detailed plan for how OLAP analytics could be applied to the integrated dataset.
- Include recommendations for cube design, dimensions, and measures.

Task 5: OLAP Operations and Benefits (5 Marks)

- Recommend specific OLAP operations (e.g., slicing, dicing, drilling down/up, pivoting) that would be beneficial for data analysis.
- Explain the rationale behind each recommendation and how these operations could provide insights or solve specific problems.

Part 3: Modeling (10 Marks)

Task 6: Data Modeling (5 Marks)

- Develop a data model suitable for your integrated dataset.
- Choose between predictive, descriptive, or prescriptive modelling, and justify your choice.

Task 7: Model Implementation and Evaluation (5 Marks)

- Implement the chosen model on the integrated dataset.
- Evaluate the model's performance and effectiveness in providing actionable insights or predictions.
- Discuss the potential applications of your model in real-world scenarios.