|  |  |  |
| --- | --- | --- |
| **Student ID** | **Student Name** | **Contribution %** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**SCHOOL OF COMPUTING (SOC)**

**This is the submission template for CA2 task a) to d) by group. Fill group members’ name and student ID below, and your answers to respective tasks in subsequent pages. Put group leader in the first row and only leader makes the submission on Brightspace.**

**Class:**

**Group Number:**

**CA2 Group Template**

**Diploma in Applied AI & Analytics**

**ST1501**

**Data Engineering**

**2025/2026 Semester 1**

1. Fill below tables for data quality issues found and how it is fixed (the changes made instead of tools/procedures). Extend the table if necessary. One sample error with description is shown here and you should **delete** it from your answers. Reminder that you can assume no timetable clash in the appointments and hence do not need to check it. Focus on issues described in topic E including inconsistencies, invalid values, etc.

|  |  |  |  |
| --- | --- | --- | --- |
| File | Row | What’s the problem? | How do you fix it? |
| Customer.dat | 123 | The date column is of invalid range: 2025/12/32. | Delete the row. |
| … | … | … | … |

Submit your SQL INSERT statements for data insertion to the OLTP tables in OLTP\_insert.sql. If you apply bulk loading, submit each of the cleaned data files used during insertion as well.

Paste your database diagram of your OLTP tables with the use of “database/reverse engineer” function in MySQL workbench.

1. Design your DW using star or snowflake schema. Paste your ERD that illustrates your design in the box below, with a short explanation of key design focus and considerations.

Briefly explain your design, such as choice of dimensions/measurements, assumptions and considerations, etc.

1. Submit your CREATE table statements for DW table creations in DW\_create.sql. Use appropriate SQL statements to extract and load data from OLTP tables to the DW and submit all your queries in DW\_insert.sql.

Paste your database diagram for your DW in the box below. Explain with data to show how you verify that your insertion is successful.

1. Describe and discuss your analysis done using the DW. Briefly explain clearly your objective, results and insights for each analysis. Your group should produce 6 different analysis if group size is 3, or 8 different analysis if group size is 4. Separately submit all your queries in DW\_query.sql. (So do not include your statements in this template, only results and discussion)

1.

2.

3.

4.

5.

6.

7.

8.