

**SECP3133 HIGH PERFORMANCE DATA PROCESSING**

**SEMESTER 2 2024/2025**

**ASSIGNMENT 2: MASTERING BIG DATA HANDLING LOGBOOK**

|  |  |
| --- | --- |
| **GROUP** | DataStorm |
| **GROUP MEMBERS** | NUR FARAH ADIBAH BINTI IDRIS (A22EC0245)  NUR ALEYSHA QURRATU’AINI BINTI MAT SALLEH (A22EC0241) |
| **SECTION** | 02 |
| **LECTURER** | DR. ARYATI BINTI BAKRI |

|  |  |  |  |
| --- | --- | --- | --- |
| **Date** | **Group Member** | **Task Description** | **Result** |
| 24/5/2025 | Both | Selected datasets that are bigger than 700MB | Dataset chosen: Transaction |
| 24/5/2025 | Farah Adibah | Create GitHub file submission | Created a folder for submission |
| 25/5/2025 | Nur Aleysha | Pandas coding | Completed Task 2 using Pandas |
| 25/5/2025 | Nur Aleysha | Polars coding | Completed Task 3 using Polars |
| 26/5/2025 | Farah Adibah | Dask coding | Completed Task 3 using Dask |
| 26/5/2025 | Farah Adibah | Graph Performance coding | Completed Task 4 |
| 30/5/2025 | Both | Finalized all tasks in Google Colab | All tasks are finalized |
| 30/5/2025 | Nur Aleysha | Markdown report: Introduction and Objectives, Task 1, Task 2, Task 3 | Completed the sections in markdown report |
| 31/5/2025 | Farah Adibah | Markdown report: Task 3, Task 4, Task 5 | Completed the sections in markdown report |
| 1/6/2025 | Both | Submit all files for submission in GitHub | Completed Assignment 2 |