Subject Code	**WQD 7009**
Subject Name	**Big Data Applications and Analytics**
Subject Leader	Dr. Riyaz Ahamed
Weightage	15%
Assignment	Individual Assignment
Handout date	7th November 2024
Submission date	1st December 2024
Instructions	1. This is an individual assignment that carries 15% of the total marks available for the module. 2. The output of this assignment should be a report between **1500 and 1800** words. 3. You are required to submit your report in soft copy (**through Spectrum UM**). Kindly ensure your name and matric ID are written on the cover sheet. 4. A mark of 0 and barring from sitting the final examination may be implemented for those who do not submit any assignments. **GENTLE REMINDER** 5. Plagiarism is a serious offence, and plagiarized work will result in an F grade. 6. Failure to submit the report by the deadline will be penalized with zero marks.

Tasks

1) Choose any dataset related to climate change and sustainability from openly available internet resources. The chosen dataset must be approved by the module leader and should have been created or developed within the **last two years**. The module leader may advise on additional tasks within the framework of this project to fulfill the module's learning outcomes.

2) Climate change and sustainability are critical, interconnected global issues impacting ecosystems, economies, and communities. Key components include carbon emissions, renewable energy adoption, climate policies, global temperature trends, conservation efforts, the environmental impacts of industry, adaptation strategies, and sustainable practices for resource management. Data analytics plays a crucial role in this area, providing insights that help governments, organizations, and individuals understand patterns, forecast future scenarios, and develop strategies for mitigation and adaptation.

3) You have the option to select your own dataset in addition to the ones on this list. Refer to the links below to find datasets related to climate change or sustainability. Update your dataset in the Excel file, specifically in the individual assignment dataset sheet. **Please check whether your classmates have already selected a dataset to avoid duplicating analyses.**
   https://www.kaggle.com/search?q=climate+change+in%3Adatasets

   https://www.kaggle.com/search?q=sustainability+in%3Adatasets

4) Propose your selected dataset to the module leader for approval by November 10th, 2024.

5) Perform various data manipulation and querying operations on the selected dataset using the HBase NoSQL platform.

6) Utilize HBase to analyze the selected dataset and execute up to 20 different queries, including DDL and DML commands, to derive meaningful results from the dataset's parameters. Additionally, execute 5 HBase shell commands to check the status of HBase operations.

Deliverables

The output should be in terms of

1. Provide a brief introduction to the chosen dataset. Draw a table and explain the most important parameters of the chosen dataset. There is no need to include all columns or entities for the project if they are not relevant.
2. Present the query results and provide explanations for the HBase queries, including meaningful analysis. Include a minimum of 5 DDL and 10 DML queries, along with 5 shell commands. Attach descriptions and screenshots as evidence for all the queries.
3. Screenshots – include only the input queries and final results (no need to attach the processing steps in the screenshot; input and final output only).
4. For DML queries, ensure that meaningful analysis is included, such as analyzing trends in carbon emissions over time, renewable energy adoption rates, or global temperature changes. Provide relevant insights and explanations that demonstrate how the data can inform policies and strategies for climate change mitigation and sustainability practices.
5. Create a Word file to include your screenshots of the given queries, an overview of your results, and explanations for each result.

Assessment Marking Criteria - Total: 15 Marks

1. Introduction about the dataset - 5 Marks
2. HBase queries, result and explanation – 10 Marks.

Rubric

Introduction about the dataset - 5 Marks			
Low **(1 Mark)**	Fair **(2 Marks)**	Above average **(3-4 Marks)**	Excellent **(5 Marks)**
No explanation on the selected dataset. Parameters are not identified.	Made attempt to explain the dataset used for the assignment. Minimum evidence for the parameters table.	Moderately explained the dataset used for the assignment. Clear evidence and table for parameter description.	Good, explanation of the dataset All the parameters are clearly explained in the table.

HBase queries and result – 10 Marks			
Low **(1-3 Marks)**	Fair **(4-6 Marks)**	Above average **(7-8 Marks)**	Excellent **(9-10 Marks)**
Minimum queries are executed and incomplete descriptions for the queries. No evidence is attached for the command execution.	Queries are complete but do not meet specifications partly or fully. Minimum pieces of evidence are attached for the queries execution and description.	All the necessary queries are executed and shown in the screenshot. Clearly evidence and description for the queries.	Good, all the 10 queries are executed successfully, Clearly, pieces of evidence are attached to show the description and relationship.