LEVEL: 5

MODULE TITLE: Big Data Analytics

MODULE CODE: MOD008909
Assessment Element: 010

Word Count: 6000 words (or equivalent)

Introduction

This assessment is based on the production of a data warehouse using data from the Office for National Statistics. You are to demonstrate knowledge of the creation process for a data warehouse that will link together data on Gross Disposable Household Income (GDHI) per head to Life Expectancy. The assignment is split into 2 elements which evaluate the module learning outcomes as follows:

1. Knowledge and Understanding. Understand and critically appraise the underlying theories, statistical and mathematical concepts associated with big data analytics.
2. Knowledge and Understanding Demonstrate knowledge of different types of tools for data collection, data cleaning, data integration and data visualisation.
3. Intellectual, practical, affective and transferrable skills. Analyse large datasets and extract statistics and features.
4. Intellectual, practical, affective and transferrable skills. Critically discuss the challenges with big data e.g., privacy, security, storage, and scaling.
5. Intellectual, practical, affective and transferrable skills. Train and validate a range of models for machine learning.

E010-2 Task 2 (70%)

Presentation

You should create a presentation of at least 10 minutes but no more than 15 minutes about the design of the data warehouse, the decisions taken using appropriate graphics and the results produced by the data warehouse.

Learning Outcomes 1, 2, 3, and 4:

You **MUST** design and present an **OLAP data warehouse** to assess **life expectancy** in relation to **Gross Disposable Household Income (GDHI)** using a NoSQL DBMS such as **MongoDB, Hadoop, or similar**. The data warehouse **MUST** support the following capabilities:

For Marks Between 0-49% (MUST Criteria):

1. **Basic Data Warehouse Design:**
   - You **MUST** design the data warehouse to provide:
     - **Life expectancy at birth** for a given area.
     - **Life expectancy at age 60** for a given area.
     - **Life expectancy at birth** for a specific income band.
     - **Life expectancy at age 60** for a specific income band.
     - **Expected income** for a given area.
2. **Basic Data Manipulation:**
   - You **MUST** demonstrate how the original data can be manipulated to fit the design.
3. **Basic Structure and Design Choices:**
   - You **MUST** provide a basic structure for the warehouse, including clear decisions on design choices.
4. **Presentation:**
   - You **MUST** create a **10-15 minute presentation** explaining the design and implementation of the data warehouse.
   - The presentation **MUST** include **basic graphics** to support the explanation.

Marks in this range are awarded based on the correct and complete implementation of the **MUST** requirements. Failure to meet all **MUST** criteria would result in marks below 40%.

For Marks Between 50-69% (SHOULD Criteria):

In addition to fulfilling the **MUST** requirements:

1. **Design Choices and Justification:**
   - You **SHOULD** provide clear **design choices** and **justify** these decisions.
2. **Enhanced Data Manipulation:**
   - You **SHOULD** demonstrate how the data can be manipulated effectively and provide **evidence** of a well-thought-out approach to integrating the data into the warehouse.
3. **Warehouse Structure:**
   - The structure of the data warehouse **SHOULD** allow for **additional data mining** capabilities beyond the basic requirements.
4. **Presentation:**
   - You **SHOULD** present a **well-structured, logical** explanation of your data warehouse design and the processes involved.
   - The presentation **SHOULD** use **visual aids** effectively to support the findings and the design decisions.

Marks in this range are awarded for meeting all **MUST** criteria and the correct implementation of **SHOULD** criteria, demonstrating a deeper understanding of the

warehouse design, better decision-making, and an intermediate level of analysis and presentation.

For Marks of 70% or Above (MAY Criteria):

In addition to fulfilling the **MUST** and **SHOULD** requirements:

1. **Advanced Design Choices:**
   - You **MAY** present **innovative design choices**, demonstrating original thinking in the design of the data warehouse.
2. **Advanced Data Manipulation and Integration:**
   - You **MAY** demonstrate advanced techniques for **data manipulation** and the integration of **complex datasets** into the warehouse.
3. **Structure for Additional Data Mining:**
   - The structure of the warehouse **MAY** be designed to support **advanced data mining** capabilities and other functionalities beyond the scope of the task.
4. **Presentation:**
   - You **MAY** deliver a **fluent, engaging presentation** that demonstrates a high level of technical competence and deep insight into the design and structure of the data warehouse.
   - The presentation **MAY** include **sophisticated visual aids** and multimedia elements that enhance clarity and understanding.

Marks of 70% and above are awarded for implementing **MAY** criteria, where students show creativity, in-depth analysis, and superior understanding of data warehouse design and NoSQL DBMS integration.

Marking Criteria

- **0-49%:** Awarded for implementing **MUST** criteria, including basic data warehouse design, data manipulation, and presentation structure.
- **50-69%:** Awarded for implementing **SHOULD** criteria, which include justification of design choices, effective data manipulation, and a well-structured presentation with supporting visuals.
- **70% and above:** Awarded for implementing **MAY** criteria, which demonstrate originality in design, advanced data manipulation, additional mining capabilities, and sophisticated presentation skills.