Data Modeling and Visualization (DMV) Course

Module 5: Data Modeling from Different Data Sources for Visualization Modular Assignment 5

MODULE 5: Data Modeling from Different Data Sources for Visualization

Duration: 4 hours

Problem Statement

Integrate data from multiple sources (CSV, JSON, database) to create unified visualizations analyzing social media engagement across platforms.

Dataset

Multiple data sources provided: - "Social_Media_Posts.csv" (structured data) - "User_Engagement.json" (semi-structured data) - "Platform_Metrics.xml" (semi-structured data) - Sample streaming data simulation

What is Expected

1. Connect to and import different data sources
2. Clean and transform data from each source
3. Merge datasets appropriately
4. Create unified dashboard
5. Demonstrate real-time data handling concepts

Rubrics

1. **Multi-Source Data Connection** – 25 points
   - Successfully import CSV data (8 points)
   - Parse and import JSON data (8 points)
   - Handle XML data extraction (9 points)
2. **Data Transformation & Cleaning** – 25 points
   - Standardize data formats across sources
   - Handle missing values and inconsistencies
   - Create common key fields for joining
3. **Data Integration** – 25 points
   - Merge datasets using appropriate join types
   - Validate data integrity after merging
   - Create unified data model
4. **Unified Visualization Dashboard** – 20 points
   - Create dashboard showing integrated metrics

- Include filters for different platforms
- Display engagement trends and comparisons

5. **Streaming Data Simulation** – 5 points
   - Demonstrate understanding of real-time data concepts
   - Show how streaming data would be incorporated

General Assignment Guidelines

Submission Requirements

- All code files and processed datasets
- Documentation of data integration steps
- Final unified dashboard or visualization
- Reflection document explaining approach and findings

Evaluation Criteria

- Technical accuracy and implementation
- Understanding of multi-source data integration concepts
- Quality of data transformation and cleaning
- Professional presentation of unified results
- Adherence to best practices

Time Allocation Recommendations

- **Hour 1**: Data source exploration and connection setup
- **Hour 2**: Data cleaning and transformation for each source
- **Hour 3**: Data integration and unified model creation
- **Hour 4**: Dashboard creation and documentation

This assignment is designed to provide comprehensive coverage of multi-source data integration fundamentals while maintaining beginner-friendly complexity and practical applicability.