

Data Modeling and Visualization (DMV) Course

Module 8: Multi-Source Analytics with R Integration Modular Assignment 8

MODULE 8: Multi-Source Analytics with R Integration

Combining: Data Visualization using R + Multi-Source Data Integration

Duration: 4 hours

Problem Statement

Integrate data from multiple sources (CSV, JSON, API, database simulation) to analyze digital marketing performance across channels. Use R for data processing, integration, and advanced visualization while creating a comprehensive analytical framework for multi-channel attribution.

Dataset

Multiple data sources provided: - “Digital_Campaigns.csv” (structured campaign data) - “Social_Engagement.json” (semi-structured social media data) - “Web_Analytics.xml” (semi-structured website data) - “Customer_Journey.txt” (unstructured customer interaction logs) - Real-time data simulation for live metrics

What is Expected

1. Connect to and process data from all 4 different source types using R
2. Implement ETL pipeline to clean, transform, and integrate data
3. Perform advanced analytics including attribution modeling
4. Create comprehensive visualization suite using ggplot2 and plotly
5. Build automated reporting pipeline with R Markdown
6. Demonstrate real-time data processing capabilities

Rubrics

1. **Multi-Source Data Integration** – 30 points
 - Process CSV data effectively (7 points)
 - Parse and integrate JSON data (8 points)
 - Extract and transform XML data (8 points)
 - Handle unstructured text data (7 points)
2. **ETL Pipeline Development** – 25 points
 - Implement data cleaning and validation
 - Create standardized data model
 - Handle data quality issues appropriately
3. **Advanced Analytics Implementation** – 20 points

- Perform multi-channel attribution analysis
 - Create customer journey mapping
 - Implement statistical modeling for insights
4. **Comprehensive Visualization Suite** – 20 points
- Create 5+ different advanced visualizations
 - Include interactive elements using plotly
 - Implement dashboard-style layouts
5. **Automated Reporting and Real-time Processing** – 5 points
- Build automated R Markdown reporting
 - Demonstrate real-time data handling concepts
 - Create reproducible analytical pipeline

General Assignment Guidelines

Skill Integration Expectations

Students should demonstrate:

- **Technical Proficiency:** Competent use of R with multiple data sources and advanced analytics
- **Analytical Thinking:** Structured approach to multi-source data integration
- **Communication Skills:** Clear presentation of complex analytical results
- **Business Acumen:** Understanding of digital marketing analytics applications

Submission Requirements

- All R code files and processed datasets from multiple sources
- Comprehensive R Markdown report with integrated analysis
- Documentation of ETL pipeline and data integration approach
- Reflection document explaining findings and business recommendations

Evaluation Criteria

- **Integration over Isolation:** Focus on combining multiple data sources effectively
- **Process over Product:** Emphasize robust ETL methodology and data quality
- **Application over Theory:** Prioritize practical multi-channel analytics
- **Communication over Complexity:** Value clear explanation of integrated insights

Time Allocation Recommendations

- **Hour 1:** Multi-source data exploration and ETL pipeline setup
- **Hour 2:** Data integration, cleaning, and unified model creation
- **Hour 3:** Advanced analytics and comprehensive visualization development
- **Hour 4:** Automated reporting, real-time concepts, and documentation

This Assignment is designed to provide comprehensive coverage of advanced multi-source analytics integration while maintaining beginner-friendly complexity and practical applicability.