PROGRAMMING FOR DATA ANALYSIS – GROUP ASSIGNMENT

Coursework Title

Analysis of Cyber-Attack Events

Overview

In this assignment, you will work collaboratively to analyze a collection of datasets related to cyber-attack incidents. You are required to merge, clean, and explore these datasets using data analytics techniques in R to uncover meaningful insights and trends. Your analysis should identify data quality issues, detect patterns, and generate actionable insights. The work must demonstrate depth and originality that extends beyond classroom examples. You are expected to apply concepts such as data exploration, transformation, visualization, and statistical analysis, providing clear justification for every analytical choice made.

All tasks must be executed exclusively in R and RStudio, including data preprocessing, analysis, and visualization. No external tools (e.g., Excel, Power BI) are permitted. Finally, you will document your entire process, analysis, and findings in a well-structured analytical report that clearly explains each step and its rationale.

This assignment aims to enable you to:

- Apply R programming and data analytics techniques to clean, integrate, and analyze real-world data.
- Develop the ability to transform raw data into meaningful and interpretable results that support decision-making.
- Enhance collaborative skills by completing a team-based analytical project.

Assignment Type

Group Assignment (Maximum: 4 members per group)

Dataset Description

The dataset comprises information about website defacement incidents. Each record includes:

- Incident Details: Date of defacement, reporting source (Notify), and affected URL.
- Technical Attributes: Server IP address, hosting country, web server version, and character encoding.
- Impact Variables: Ransom amount (in thousands), downtime (in days), and estimated financial losses.

You will receive three datasets (in Moodle) that must be merged, cleaned, and prepared for analysis. Details of the datasets are as shown below:

HackingData_Part3	Text Document	17,894 KB	⊘
HackingData_Part2	Microsoft Excel Worksheet	15,480 KB	⊘
HackingData_Part1	Microsoft Excel Comma Separated Values File	20,046 KB	⊘
HackingData Dataset Description	Text Document	1 KB	⊘

General Requirements

- The analysis must be meaningful and data-driven, supporting decision-making through evidence-based findings.
- All data preprocessing, analysis, and visualization must be performed entirely using R scripts within RStudio.
- External tools (e.g., Excel, Power BI, Tableau) are not permitted.
- Follow good programming practices, including clear commenting, consistent variable naming, and proper indentation.
- You may perform additional background research to enhance your understanding of the dataset and domain.
- Demonstrate a structured analytical approach and ensure your visualizations effectively communicate insights.

Deliverables

You must submit the following via APU Learning Management System (Moodle):

1. R Script

- Submit a zipped .R script file containing your complete code.

- Filename = RScript_Group<group number>.Zip

- The program must execute successfully in RStudio without errors.

2. Analytical Report

- Submit in Microsoft Word or PDF format.

- Maximum length: 100 pages, including charts, tables, and appendices.

- Formatting: Times New Roman, font size 12 pt, 1.5 spacing.

- Filename = AnalysisResport_Group<group number>.Docx or .Pdf

Suggested Report Structure & Chapters

Chapter 1 – Introduction (Group Work)

1.1 Project Description

Introduce the project and its objectives. Provide an overview of the dataset(s) and the context of cyber-attack analysis.

1.2 Data Description

Provide a brief description of the datasets.

1.3 Data Preparation

Describe how the data was imported into RStudio and prepared for analysis.

Document the preprocessing steps including but not limited to:

- Checking and handling duplicates
- Checking and handling missing values
- Identifying and treating outliers
- Checking and fixing for any other issues

Include explanations for each step and justify your approach.

Provide R code snippets with explanations and outputs for each step taken.

1.4 Assumptions

List any assumptions (if any) made in relation to the data preparation and analysis

Chapter 2 – Analysis (Individual)

2.1 Student Name and TP Number:

Each member must complete their own section, beginning on a new page.

2.1.1 Objectives

State your research objectives.

2.1.2 Analysis Questions

Formulate several guiding questions for each objective.

Data Preparation

Describe any additional preparation work you did to perform or enhance your individual analysis.

Include explanations for each step and justify your approach.

Provide R code snippets with explanations and outputs for each step taken.

2.1.4 Data Analysis

Conduct exploratory and inferential analyses relevant to your objectives.

Include summary statistics and visualizations with interpretations.

Clearly indicate which objective each analysis supports.

Include explanations for each step and justify your approach.

Provide R code snippets with explanations and outputs for each step taken.

2.1.5 Hypothesis Formulation and Testing

Formulate hypotheses based on your findings and perform appropriate statistical tests in R.

Interpret and explain your test results comprehensively.

Include explanations for each step and justify your approach.

Provide R code snippets with explanations and outputs for each step taken.

Repeat sections 2.1.1–2.1.5 for each group member. Group member 2 sections will start as 2.2 and so on

Chapter 3 – Group Hypothesis

3.1 Group Hypothesis Formulation

Develop one **composite group hypothesis** that integrates all individual members' findings – at least 1 hypothesis from each member.

3.2 Group Hypothesis Testing

Perform the hypothesis test collaboratively

Interpret and explain your test results comprehensively.

Include explanations for each step and justify your approach.

Provide R code snippets with explanations and outputs for each step taken.

3.3 Overall Conclusion

Summarize key insights and provide overall conclusions from the collective analysis.

Chapter 4 – Summary

4.1 Limitations and Recommendations

Discuss the limitations of your analysis and suggest areas for future improvement or further research.

4.2 Workload Matrix & AI Usage Declaration

Provide a summary each member's contribution and usage of AI in doing the analysis and producing the report.

References

List all references in **APA format**.

Assessment Criteria

Component	Criteria
Analysis (70%)	Appropriateness of techniques, justification, analytical coverage, data handling, visualization, programming practices, and inclusion of advanced or creative analytical elements.
Findings and Discussion (30%)	Quality of hypothesis formulation, relevance of findings, depth of discussion, clarity of conclusions, and originality of interpretation.

Academic Integrity

You are expected to uphold the highest standards of academic integrity. Plagiarism or collusion will be treated as a serious academic offence and dealt with according to the regulations of Asia Pacific University of Technology and Innovation (APU) and De Montfort University.