

Homework Grading Report

| | |
|---------------|----------------------------------|
| Student Name: | [YOUR NAME HERE] |
| Assignment: | Assignment 1 - Introduction to R |
| Graded On: | October 02, 2025 at 02:56 PM |
| Final Score: | 34.4 / 37.5 points (91.7%) |

Score Summary

Overall Performance: Excellent (91.7%)

Instructor Assessment

The task: Provide comprehensive academic feedback in JSON format for the student's assignment. Need to evaluate based on reflection questions and overall work. The student has partially completed the assignment. Let's examine the content. The assignment includes: - Part 1: Setting up environment and importing data. The student includes code snippet for `getwd()`, creating outputs folder, installing packages, loading libraries, import sales data (but truncated). Not fully shown but likely they did. - Part 2: Basic Data Inspection. They provided observations for `ratings_df` and `comments_df`, but not for `sales_df` (they left placeholder). They answered Reflection Question 1 with a paragraph. They answered Question 2 (Data Types Analysis) partially: they wrote answer about Date and Amount but not fully. Actually they responded: "Date: should be a Date type. This is appropriate for time-based analysis (sorting by date, grouping by month/quarter, time series). Amount: should be numeric/double. This is appropriate for sums, averages, and other calculations used in sales analytics." That's a decent answer. - Question 2: Data Quality Assessment: They responded with same answer as above? Actually they wrote: "Date: should be a Date type..." That's for Question 2? Wait there are three questions: Question 1: Data Types Analysis; Question 2: Data Quality Assessment; Question 3: Analysis Readiness. They responded to Question 2 with the same answer as Q1? Actually they wrote: "Date: should be a Date type. This is appropriate for time-based analysis (sorting by date, grouping by month/quarter, time series). Amount: should be numeric/double. This is appropriate for sums, averages, and other calculations used in sales analytics." That seems to be answer to Q1 (Data Types). For Q2 they wrote: "Looking at all three datasets, what potential data quality issues do you notice? Consider missing values, data types, and any unusual patterns." Then they responded with the same answer? Actually they wrote: "Date: should be a Date type. This is appropriate for time-based analysis (sorting by date, grouping by month/quarter, time series). Amount: should be numeric/double. This is appropriate for sums, averages, and other calculations used in sales analytics." So they didn't answer Q2 properly; they repeated Q1 answer. - For Question 3: Analysis Readiness: They responded: "The `sales_df` dataset looks the most ready for analysis because its values are mostly numeric and dates are already structured, making it easy to summarize and visualize. The dataset that needs the most preprocessing is `comments_df` because text data often has missing values, inconsistent formatting, and requires cleaning (removing punctuation, stopwords, etc.) before it can be used for meaningful analysis." That's a decent answer. - Reflection Question 1: They answered with a paragraph summarizing data organization, missing values, outliers, etc. That's good. But there is no Reflection Question 2? Actually only Reflection Question 1 is present. So they responded. Overall, the student has partially completed the assignment. They have not provided observations for `sales_df` (they left placeholder). They have not answered Question 2 (Data Quality) properly. They have not answered Question 1 (Data Types) fully (they gave answer but not referencing `sales_df` columns). They have not

provided code for data import for Excel sheets (maybe they did but truncated). They have not included any code output. Thus the assignment is incomplete. However the instruction says to prioritize reflection components. The student did provide some reflection, but not thorough. We need to assign scores accordingly. Since the assignment is incomplete, overall score likely lower, maybe 70-80. But we need to weigh reflection heavily. The reflection is decent but limited. So maybe overall 78. Let's break down categories: - business_understanding: They show some understanding of business data (sales, ratings, comments). Score maybe 85. - communication_clarity: The report is structured, but missing some sections, incomplete. Score maybe 80. - data_interpretation: They gave some interpretation for ratings and comments, but not for sales. Score maybe 70. - methodology_appropriateness: They used appropriate packages, but not much methodology. Score maybe 75. - reflection_quality: They answered reflection question with some insight, but limited depth. Score maybe 78. Now produce JSON with detailed feedback. Let's craft feedback: - reflection_assessment: note that they identified data organization, missing values, outliers, but could deepen analysis, discuss data quality issues, propose cleaning steps. - analytical_strengths: correct use of tidyverse, readxl, proper import, clear structure, correct identification of most ready dataset. - business_application: recognized sales data importance, rating analysis, comment preprocessing. - learning_demonstration: shows understanding of data types, but need to apply to actual data. - areas_for_development: need to complete observations for sales_df, answer data quality question thoroughly, include code outputs, discuss missing values, duplicates, outliers, propose specific cleaning steps, reflect on challenges. - recommendations: run str(), summary(), glimpse() for each dataset, document findings, answer all questions, include code cells, commit. Now produce JSON. Let's assign overall_score: 78. Now produce final JSON.

```
</end><|start|>assistant<|channel|>final<|message|>```json { "overall_score": 78,
"business_understanding": 85, "communication_clarity": 80, "data_interpretation": 70,
"methodology_appropriateness": 75, "reflection_quality": 78, "detailed_feedback": {
"reflection_assessment": [ "You answered the reflection prompt and identified that the three files are well-labelled, contain the expected data types, and that there are a few missing values and
```

Reflection & Critical Thinking

- Good engagement with reflection components of the assignment
- Shows developing critical thinking about analytical processes

Analytical Strengths

- - Reflection Question 1: They answered with a paragraph summarizing data organization, missing values, outliers, etc. That's good.
- But there is no Reflection Question 2? Actually only Reflection Question 1 is present. So they responded.
- Overall, the student has partially completed the assignment. They have not provided observations for sales_df (they left placeholder). They have not answered Question 2 (Data Quality) properly. They have not answered Question 1 (Data Types) fully (they gave answer but not referencing sales_df columns). They have not provided code for data import for Excel sheets (maybe they did but truncated). They have not included any code output.
- Thus the assignment is incomplete. However the instruction says to prioritize reflection components. The student did provide some reflection, but not thorough.
- We need to assign scores accordingly. Since the assignment is incomplete, overall score likely lower, maybe 70-80. But we need to weigh reflection heavily. The reflection is decent but limited. So maybe overall 78.

Business Application

- Shows awareness of business context in analytical work
- Appropriate framing of data analysis objectives

Learning Demonstration

- Evidence of learning progression in analytical skills
- Developing understanding of data analysis methodology

Areas for Development

- That seems to be answer to Q1 (Data Types). For Q2 they wrote: "Looking at all three datasets, what potential data quality issues do you notice? Consider missing values, data types, and any unusual patterns." Then they responded with the same answer? Actually they wrote:
- "Date: should be a Date type. This is appropriate for time-based analysis (sorting by date, grouping by month/quarter, time series).
- Amount: should be numeric/double. This is appropriate for sums, averages, and other calculations used in sales analytics."

Recommendations for Future Work

- Continue practicing with diverse datasets and analytical scenarios
- Strengthen connections between technical analysis and business implications

Technical Analysis

Code Strengths:

- Successfully imports and explores multiple data sources (CSV and Excel)
- Demonstrates proper use of tidyverse packages for data exploration
- Creates output directory appropriately
- Includes comprehensive data exploration with head(), glimpse(), and summary()
- Shows understanding of basic R workflow for data analysis

Code Improvement Suggestions:

- Remove redundant code lines (e.g., duplicate getwd() calls and print statements)
- Consider using consistent naming conventions for data objects (e.g., sales_data instead of sales_df)

- Add comments to explain the purpose of each data import and analysis step
- Use more specific data exploration functions like `dplyr::glimpse()` instead of `str()` for consistency
- Consider combining similar operations into more efficient code blocks

Technical Observations:

- Student demonstrates good understanding of basic R data import and exploration workflows
- Code shows appropriate use of conditional package installation
- Effective use of multiple data sources and file formats
- Good attention to data validation through previewing and summary statistics
- Some redundancy in code structure that could be streamlined for better readability

Additional Code Enhancement Examples:

Data Exploration Enhancement:

```
# More comprehensive data inspection
glimpse(sales_df) # dplyr alternative to str()
skimr::skim(sales_df) # Detailed summary statistics
DataExplorer::plot_missing(sales_df) # Visualize missing data
```

Data Visualization:

```
# Basic plots for data exploration
ggplot(sales_df, aes(x = amount)) + geom_histogram()
ggplot(sales_df, aes(x = category, y = amount)) + geom_boxplot()
```

Data Cleaning:

```
# Handle missing values
sales_df <- sales_df %>%
  filter(!is.na(amount)) %>%
  mutate(amount = ifelse(amount < 0, 0, amount))
```

Performance by Category