MEDIUM_SOL

DISHA

2025-03-07

Contents

0.1 Student Performance Analysis

0.1.1 Load Required Libraries

```r
library(data.table)
library(ggplot2)
library(reshape2)
```

```
##
## Attaching package: 'reshape2'

## The following objects are masked from 'package:data.table':
##
##     dcast, melt
```

0.1.2 Create Data Tables

```r
# Student Data
students <- data.table(student_id = c(1, 2, 3, 4, 5, 6),
                       name = c("Alice", "Bob", "Charlie", "David", "Eva", "Frank"),
                       age = c(20, 21, 19, 22, 20, 23))

# Exam Scores Data
exam_scores <- data.table(student_id = c(2, 3, 4, 5, 7),
                          math_score = c(85, 90, 88, 78, 92),
                          science_score = c(80, 85, 89, 75, 95))

# Attendance Data
attendance <- data.table(student_id = c(1, 2, 3, 5, 6),
                         attendance_percentage = c(95, 88, 90, 85, 92))
```

0.1.3 Merge the Data

```
# Merge Students with Exam Scores
merged_data <- merge(students, exam_scores, by = "student_id", all = TRUE)

# Merge the result with Attendance Data
final_data <- merge(merged_data, attendance, by = "student_id", all = TRUE)
```

0.1.4 Handle Missing Values using Mean Imputation

```
for (col in c("math_score", "science_score", "attendance_percentage")) {
  final_data[is.na(get(col)), (col) := mean(final_data[[col]], na.rm = TRUE)]
}
```

0.1.5 Print Final Merged Data Table

```
print("Final Merged Data:")
```

```
## [1] "Final Merged Data:"
```

```
print(final_data)
```

```
## Key: <student_id>
##      student_id    name    age math_score science_score attendance_percentage
##          <num>  <char> <num>      <num>         <num>                  <num>
## 1:           1   Alice    20       86.6          84.8                     95
## 2:           2     Bob    21       85.0          80.0                     88
## 3:           3 Charlie    19       90.0          85.0                     90
## 4:           4   David    22       88.0          89.0                     90
## 5:           5     Eva    20       78.0          75.0                     85
## 6:           6   Frank    23       86.6          84.8                     92
## 7:           7    <NA>    NA       92.0          95.0                     90
```

0.1.6 Reshape Data for Visualization

```
long_data <- melt(final_data, id.vars = c("student_id", "name"),
                  measure.vars = c("math_score", "science_score", "attendance_percentage"),
                  variable.name = "Category", value.name = "Value")
```

0.1.7 Create an Advanced Bar Chart with Trends

```r
ggplot(long_data, aes(x = factor(student_id), y = Value, fill = Category)) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.7)) +
  geom_text(aes(label = round(Value, 1)), vjust = -0.5, position = position_dodge(width = 0.7)) +
  labs(title = "Student Performance Overview", x = "Student ID", y = "Value") +
  scale_fill_manual(values = c("math_score" = "blue", "science_score" = "red", "attendance_percentage" =
                    labels = c("Math Score", "Science Score", "Attendance %")) +
  theme_minimal()
```