Practical 2

Data condensation using RStudio

Summarize the Data:-

summary(data)

Variable	Min	1st Qu.	Median	Mean	3rd Qu.	Max
gender	0.000	0.000	0.000	0.482	1.000	1.000
race_ethnicity	-	-	-	-	-	-
parental_level_of_education	-	-	-	-	-	-
lunch	0.000	0.000	1.000	0.645	1.000	1.000
test_preparation_course	0.000	0.000	0.000	0.358	1.000	1.000
math_score	0.00	57.00	66.00	66.09	77.00	100.00
reading_score	17.00	59.00	70.00	69.17	79.00	100.00
writing_score	10.00	57.75	69.00	68.05	79.00	100.00
total_score	27.0	175.0	205.0	203.3	233.0	300.0
average_score	9.00	58.33	68.33	67.77	77.67	100.00

Explore Relationships Between Variables

1 Distribution of Scores

Overall statistics

> overall_stats <- data.frame(Mean = colMeans(studentdata[, c("math_score", "reading_score", "writing_score")]),Variance = apply(studentdata[, c("math_score", "reading_score", "writing_score")], 2, var))

> print("Overall Statistics:")

> print(overall_stats)

Variable	Mean	Variance
math_score	66.089	229.9190
reading_score	69.169	213.1656
writing_score	68.054	230.9080

```
# Step 1: Print Gender Distribution
cat("Gender Distribution:\n")
gender_distribution <- table(studentdata$gender)
print(gender_distribution)

# Step 2: Print Test Preparation Course Distribution
cat("\nTest Preparation Course Distribution:\n")
test_preparation_course_distribution <- table(studentdata$test_preparation_course)
print(test_preparation_course_distribution)

# Step 3: Print Parental Education Level Distribution
cat("\nParental Education Level Distribution:\n")
parental_education_distribution <- table(studentdata$parental_level_of_education)
print(parental_education_distribution)

# Step 4: Print Lunch Distribution
cat("\nLunch Distribution:\n")
lunch_distribution <- table(studentdata$lunch)
print(lunch_distribution)
```

ouput:-

Gender Distribution

Gender	Count
Male	518
Female	482

Test Preparation Course Distribution

Course Completion Status	Count
Not Completed	642
Completed	358

Parental Education Level Distribution

Education Level	Count
Associate's Degree	222
Bachelor's Degree	118
High School	196
Master's Degree	59
Some College	226
Some High School	179

Lunch Distribution

Lunch Type	Count
Standard Lunch	355
Free/Reduced Lunch	645

Condense the data:-

```
unique_gender <- length(unique(studentdata$gender))
> unique_test_preparation <- length(unique(studentdata$test_preparation_course))
> unique_parental_education <- length(unique(studentdata$parental_level_of_education))
> unique_lunch <- length(unique(studentdata$lunch))
>
> # Quantitative variables: Calculate summary statistics
> math_summary <- summary(studentdata$math_score)
> reading_summary <- summary(studentdata$reading_score)
> writing_summary <- summary(studentdata$writing_score)
cat("Qualitative Variables:\n")
cat("Unique Genders:", unique_gender, "\n")
cat("Unique Test Preparation Courses:", unique_test_preparation, "\n")
cat("Unique Parental Education Levels:", unique_parental_education, "\n")
```

```
cat("Unique Lunch Types:", unique_lunch, "\n\n")
cat("Quantitative Variables:\n")
 cat("Math Score Summary:\n", math_summary, "\n")
cat("Reading Score Summary:\n", reading_summary, "\n")
cat("Writing Score Summary:\n", writing_summary, "\n")
```

output:-

Here is the **output in table format** as requested:

Qualitative Variables Table

Variable	Unique Values
Gender	2
Test Preparation Courses	2
Parental Education Levels	6
Lunch Types	2

Quantitative Variables Table

Statistic	Math Score	Reading Score	Writing Score
Min	0	17	10
1st Quartile	57	59	57.75
Median	66	70	69
Mean	66.089	69.169	68.054
3rd Quartile	77	79	79
Max	100	100	100

This provides a clear and structured view of your summarized data.

Use of Data Condensation

1. **Simplifies Data**: Condensing data reduces complexity by summarizing large datasets into key metrics, making it easier to interpret.
2. **Highlights Trends**: It helps identify patterns, such as averages, ranges, or distributions, at a glance.
3. **Improves Decision-Making**: Summarized data enables quicker and more informed decisions, especially in areas like performance analysis or resource allocation.
4. **Efficient Comparisons**: Condensation facilitates comparisons across variables or groups without overwhelming details.
5. **Data Visualization**: Summarized information is ideal for creating charts or tables to present insights effectively.

In essence, data condensation transforms raw data into actionable insights, aiding both analysis and communication.