<https://github.com/bhavapriyapj>

BHAVAPRIYA.PJ

**R PROGRAMMING MOST IMPORTANT QUESTION AND ANSWER**

**DATAFRAMES**

1. **Create a dataframe with the following details**

**Name = c("Alice", "Bob", "Charlie", "David")**

**Age = c(25, 30, 22, 28)**

**Salary = c(50000, 60000, 45000, 70000)**

* **Change the column name Salary to Income.**
* **Add a new column Department = c("HR", "IT", "Finance", "Marketing").**
* **Remove the column Age.**

data<-data.frame(

Name = c("Alice", "Bob", "Charlie", "David"),

Age = c(25, 30, 22, 28) ,

Salary = c(50000, 60000, 45000, 70000)

)

colnames(data)[3]<-"income"

print(data)

data$department<- c("HR", "IT", "Finance", "Marketing")

print(data)

data$Age<-NULL

print(data)

**2.Given the following DataFrame:**

**data <- data.frame(**

**Product = c("Laptop", "Tablet", "Phone", "Monitor"),**

**Price = c(70000, 30000, 25000, 15000),**

**Stock = c(50, 100, 200, 80)**

**)**

* **Extract only the Price column.**
* **Extract the row where Product is "Phone".**
* **Sort the dataframe based on Price in descending order.**

data <- data.frame(

Product = c("Laptop", "Tablet", "Phone", "Monitor"),

Price = c(70000, 30000, 25000, 15000),

Stock = c(50, 100, 200, 80))

extractedcol<-data[,3]

print(extractedcol)

product<-data[data$Product=="Phone",]

sorteddata<-data[order (-data$Price),]

print(product)

print(sorteddata)

3.Given the dataframe

students <- data.frame(

Name = c("Rahul", "Priya", "Amit", "Neha", "Vikram"),

Marks = c(85, 76, 90, 65, 80),

Attendance = c(90, 80, 85, 75, 60)

)

Perform the following operations:

1. Find students who scored more than 80 marks.
2. Find students with both marks > 75 and attendance > 80.
3. Replace marks of Neha with 70.

students <- data.frame(

Name = c("Rahul", "Priya", "Amit", "Neha", "Vikram"),

Marks = c(85, 76, 90, 65, 80),

Attendance = c(90, 80, 85, 75, 60))

scored<-students[students$Marks>75 & students$Attendance>80,]

print(scored)

students$Marks[students$Name=="Neha"]<-70

print(students)

1. Create two dataframes:

df1 <- data.frame(ID = c(1, 2, 3), Name = c("A", "B", "C"))

df2 <- data.frame(ID = c(3, 4, 5), Age = c(25, 30, 22))

Perform the following:

1. Merge these dataframes using the ID column.
2. Append df2 to df1 using rbind().

df1 <- data.frame(ID = c(1, 2, 3), Name = c("A", "B", "C"))

df2 <- data.frame(ID = c(3, 4, 5), Age = c(25, 30, 22))

print(df1)

print(df2)

merge(df1,df2,by ="ID")

rbind(df2,df1)

1. Given the dataframe:

Region = c("North", "South", "East", "West", "North")

Sales = c(1000, 1500, 1200, 1800, 1600)

Perform the following:

1. Find the **total sales** using sum().
2. Find the **average sales per region** using aggregate().

sale <- data.frame(

Region = c("North", "South", "East", "West", "North"),

Sales = c(1000, 1500, 1200, 1800, 1600)

)

print(sales)

total<-sum(sale$Sales)

print(total)

avg\_sales <- aggregate(Sales ~ Region, data = sale, FUN = mean)

print("Average Sales per Region:")

print(avg\_sales)

1. **Given the dataframe:**

**data <- data.frame(**

**ID = c(1, 2, 3, 4, 5),**

**Score = c(78, NA, 85, 90, NA)**

**Perform the following:**

1. **Find the rows with missing values.**
2. **Replace NA values with the mean of the Score column.**
3. **Remove all rows with NA values.**

data <- data.frame(

ID = c(1, 2, 3, 4, 5),

Score = c(78, NA, 85, 90, NA)

)

print(data)

missing<-data[is.na(data$Score),]

print(missing)

meandata<-mean(data$Score,na.rm=TRUE)

print(meandata)

remove\_row<-na.omit(data)

print(remove\_row)

1. Given the dataframe:

ID = c(101, 102, 103, 104, 105),

Score = c(78, 56, 89, 45, 90)

Perform the following:

1. Find the **mean, median, and standard deviation** of the Score column using mean(), median(), and sd().
2. Use apply() to find the **maximum value** in each column.
3. Use lapply() to return the **square of each Score**.

data <- data.frame(

ID = c(101, 102, 103, 104, 105),

Score = c(78, 56, 89, 45, 90)

)

print(data)

mean\_data<-mean(data$Score)

print(mean\_data)

median\_data<-median(data$Score)

print(median\_data)

std<-sd(data$Score)

print(std)

max\_value<-apply(data,2,max)

print(max\_value)

square\_score<-lapply(data["Score"], function(x) x^2)

print(square\_score)

1. Given the dataframe:

Name = c("John", "Mary", "Alex", "Sarah"),

Salary = c(50000, 60000, 45000, 70000)

Perform the following:

1. If Salary is **greater than 55000**, add a new column Bonus = "Yes", otherwise Bonus = "No".
2. Increase Salary by **10% for all employees** and update the dataframe.

employees <- data.frame(

Name = c("John", "Mary", "Alex", "Sarah"),

Salary = c(50000, 60000, 45000, 70000)

)

employees$Bonus<-ifelse(employees$Salary>50000,"yes","No")

print(employees)

employees$Salary<-employees$Salary\*1.10

print(employees)

1. Given the dataframe:

Student = c("Alice", "Bob", "Charlie"),

Subject = c("Math", "Science", "English"),

Marks = c(85, 90, 88)

Perform the following:

1. Convert the **long format to wide format** using dcast().
2. Convert the **wide format back to long format** using melt().

library(reshape2)

data <- data.frame(

Student = c("Alice", "Bob", "Charlie"),

Subject = c("Math", "Science", "English"),

Marks = c(85, 90, 88)

)

wide\_data <- dcast(data, Student ~ Subject, value.var = "Marks")

print(wide\_data)

long\_data <- melt(wide\_data, id.vars = "Student", variable.name = "Subject", value.name = "Marks")

print(long\_data)

10.Given the dataframe

Name = c("Alice", "Bob", "Alice", "David"),

Age = c(25, 30, 25, 28),

Score = c(90, 85, 90, 80)

Perform the following:

1. Identify duplicate rows using duplicated().
2. Remove duplicate rows and print the updated dataframe.

data <- data.frame(

Name = c("Alice", "Bob", "Alice", "David"),

Age = c(25, 30, 25, 28),

Score = c(90, 85, 90, 80))

duplicates <- duplicated(data)

print("Duplicate Rows:")

print(data[duplicates, ])

unique\_data <- data[!duplicated(data), ]

print("Updated Dataframe without Duplicates:")

print(unique\_data)