

Data_Frame.R

vishnu

2023-11-10

#Create a sample dataset in a CSV file named "employee_data.csv." The dataset should contain the following columns:

#Employee ID: A unique identifier for each employee.

#Name: The name of the employee.

#Department: The department in which the employee works.

#Salary: The monthly salary of the employee (in USD).

#Joining Date: The date on which the employee joined the company (in YYYY-MM-DD format).

```
employee_data <- data.frame(  
  Employee_ID = 101 : 110,  
  Name = c("Vishnu","Aaron","Alwin","Amar","Akhil","Ashwin","Abin","Sidharth",  
           "Mohit","Joe"),  
  Department = c("R&D","IT","Finance","Marketing","HR","Finance","HR","IT",  
                "R&D","Finance"),  
  Salary = c(90000, 68000, 33500, 33500, 90900, 48200, 21300, 66100, 85900, 16400),  
  Joining_Date = c("2020-04-16", "2018-10-07", "2021-05-30", "2018-12-16", "2016-11-20", "201  
9-01-03", "2020-06-01", "2017-06-17", "2020-02-25", "2019-05-14")  
)
```

Save the data frame to a CSV file

```
write.csv(employee_data, "employee_data.csv", row.names = FALSE)
```

#Load the dataset from "employee_data.csv" into a data frame in R.

```
employee_data <- read.csv("employee_data.csv")
```

```
employee_data
```

| ## | Employee_ID | Name | Department | Salary | Joining_Date |
|-------|-------------|----------|------------|--------|--------------|
| ## 1 | 101 | Vishnu | R&D | 90000 | 2020-04-16 |
| ## 2 | 102 | Aaron | IT | 68000 | 2018-10-07 |
| ## 3 | 103 | Alwin | Finance | 33500 | 2021-05-30 |
| ## 4 | 104 | Amar | Marketing | 33500 | 2018-12-16 |
| ## 5 | 105 | Akhil | HR | 90900 | 2016-11-20 |
| ## 6 | 106 | Ashwin | Finance | 48200 | 2019-01-03 |
| ## 7 | 107 | Abin | HR | 21300 | 2020-06-01 |
| ## 8 | 108 | Sidharth | IT | 66100 | 2017-06-17 |
| ## 9 | 109 | Mohit | R&D | 85900 | 2020-02-25 |
| ## 10 | 110 | Joe | Finance | 16400 | 2019-05-14 |

#Display the structure of the data frame, including column names and data types.

```
str(employee_data)
```

```
## 'data.frame': 10 obs. of 5 variables:
## $ Employee_ID : int 101 102 103 104 105 106 107 108 109 110
## $ Name : chr "Vishnu" "Aaron" "Alwin" "Amar" ...
## $ Department : chr "R&D" "IT" "Finance" "Marketing" ...
## $ Salary : int 90000 68000 33500 33500 90900 48200 21300 66100 85900 16400
## $ Joining_Date: chr "2020-04-16" "2018-10-07" "2021-05-30" "2018-12-16" ...
```

#Calculate and add a new column named "Years of Service" to the data frame, representing the number of years each employee has worked in the company.

```
employee_data$Joining_Date <- as.Date(employee_data$Joining_Date,format = "%Y-%m-%d")
employee_data$Current_Date <- Sys.Date() #system current date
employee_data$Years_of_Service <- as.numeric(difftime(employee_data$Current_Date, employee_data$Joining_Date, units = "days") / 365.25)
employee_data$Current_Date <- NULL
employee_data
```

| | Employee_ID | Name | Department | Salary | Joining_Date | Years_of_Service |
|-------|-------------|----------|------------|--------|--------------|------------------|
| ## 1 | 101 | Vishnu | R&D | 90000 | 2020-04-16 | 3.567420 |
| ## 2 | 102 | Aaron | IT | 68000 | 2018-10-07 | 5.092402 |
| ## 3 | 103 | Alwin | Finance | 33500 | 2021-05-30 | 2.447639 |
| ## 4 | 104 | Amar | Marketing | 33500 | 2018-12-16 | 4.900753 |
| ## 5 | 105 | Akhil | HR | 90900 | 2016-11-20 | 6.970568 |
| ## 6 | 106 | Ashwin | Finance | 48200 | 2019-01-03 | 4.851472 |
| ## 7 | 107 | Abin | HR | 21300 | 2020-06-01 | 3.441478 |
| ## 8 | 108 | Sidharth | IT | 66100 | 2017-06-17 | 6.398357 |
| ## 9 | 109 | Mohit | R&D | 85900 | 2020-02-25 | 3.707050 |
| ## 10 | 110 | Joe | Finance | 16400 | 2019-05-14 | 4.492813 |

#Create a new data frame named "Senior Employees" containing records of employees who have worked for the company for 5 or more years.

```
Senior_Employees <- subset(employee_data, Years_of_Service >= 5)
Senior_Employees
```

| | Employee_ID | Name | Department | Salary | Joining_Date | Years_of_Service |
|------|-------------|----------|------------|--------|--------------|------------------|
| ## 2 | 102 | Aaron | IT | 68000 | 2018-10-07 | 5.092402 |
| ## 5 | 105 | Akhil | HR | 90900 | 2016-11-20 | 6.970568 |
| ## 8 | 108 | Sidharth | IT | 66100 | 2017-06-17 | 6.398357 |

#Calculate and print the average salary of employees in each department.

```
avg_salary_by_dept <- aggregate(Salary ~ Department, data = employee_data, FUN = mean)
print(avg_salary_by_dept)
```

| | Department | Salary |
|------|------------|--------|
| ## 1 | Finance | 32700 |
| ## 2 | HR | 56100 |
| ## 3 | IT | 67050 |
| ## 4 | Marketing | 33500 |
| ## 5 | R&D | 87950 |

#Determine the highest and lowest salaries in the entire dataset and identify the employees with these salaries.

```
highest_salary <- max(employee_data$Salary)
lowest_salary <- min(employee_data$Salary)
```

```
employees_with_highest_salary <- subset(employee_data, Salary == highest_salary)
employees_with_lowest_salary <- subset(employee_data, Salary == lowest_salary)
```

```
cat("Highest Salary: $", highest_salary, "\n")
```

```
## Highest Salary: $ 90900
```

```
cat("Employees with Highest Salary:\n")
```

```
## Employees with Highest Salary:
```

```
print(employees_with_highest_salary)
```

```
##   Employee_ID Name Department Salary Joining_Date Years_of_Service
## 5           105 Akhil          HR   90900   2016-11-20         6.970568
```

```
cat("Lowest Salary: $", lowest_salary, "\n")
```

```
## Lowest Salary: $ 16400
```

```
cat("Employees with Lowest Salary:\n")
```

```
## Employees with Lowest Salary:
```

```
print(employees_with_lowest_salary)
```

```
##   Employee_ID Name Department Salary Joining_Date Years_of_Service
## 10           110 Joe      Finance  16400   2019-05-14         4.492813
```

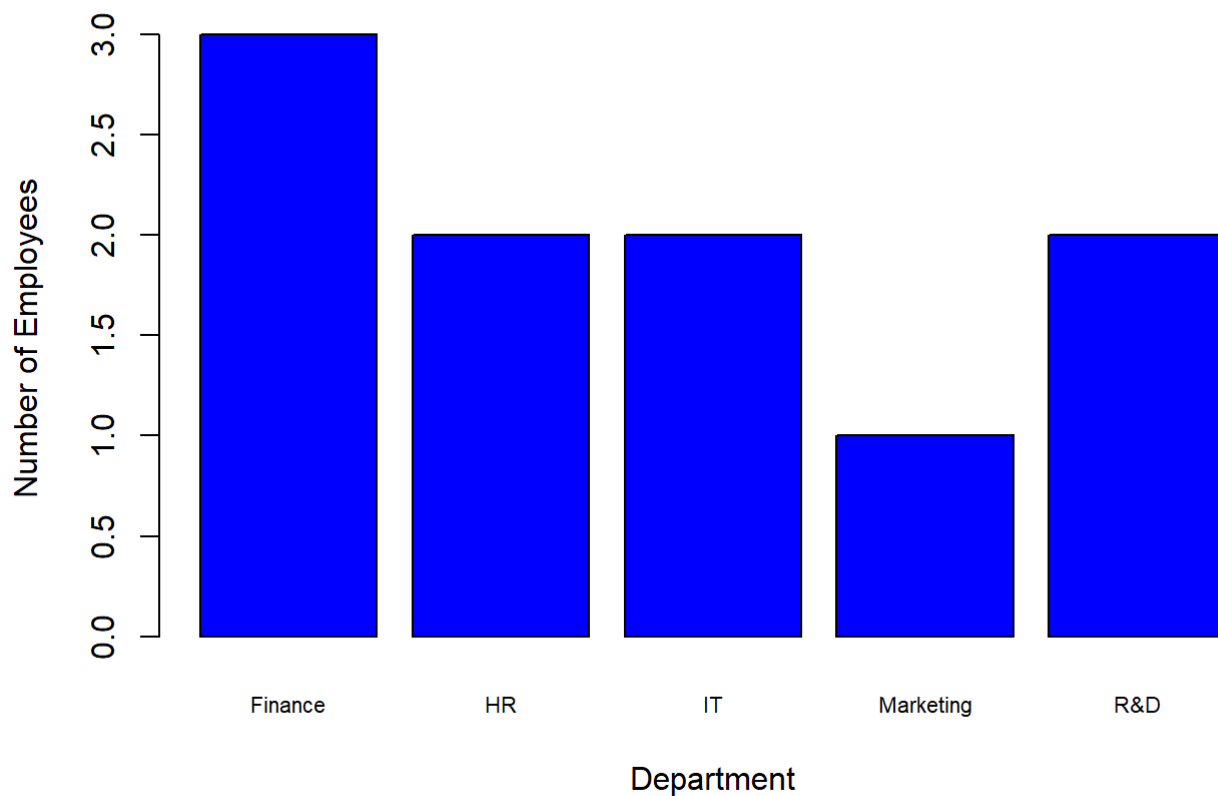
```
#Create a bar plot to visualize the number of employees in each department.
```

```
# Create a bar plot with blue bars
```

```
department_counts <- table(employee_data$Department)
```

```
barplot(department_counts,  
        main = "Number of Employees in Each Department",  
        xlab = "Department",  
        ylab = "Number of Employees",  
        col = "blue",  
        border = "black",  
        cex.names = 0.7  
)
```

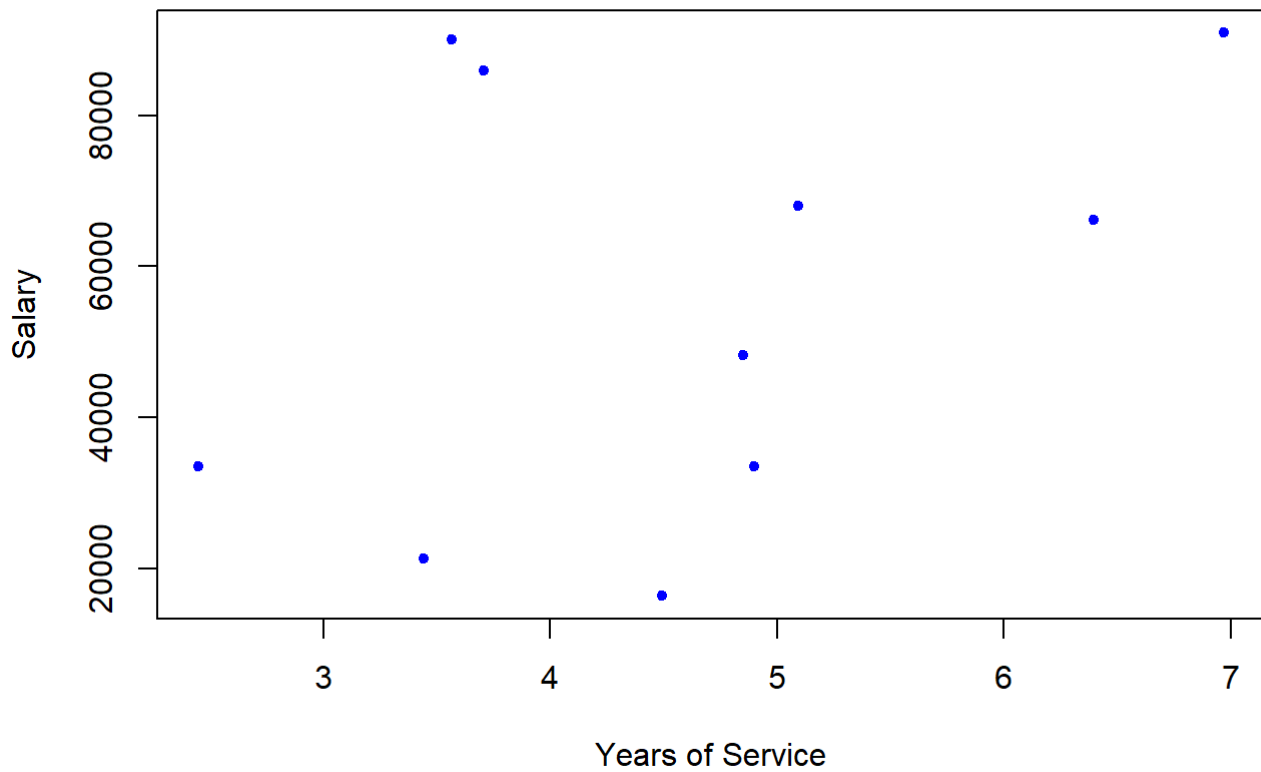
Number of Employees in Each Department



```
#Generate a scatter plot to explore the relationship between years of service and salary.
```

```
plot(employee_data$Years_of_Service, employee_data$Salary,  
      main = "Scatter Plot of Years of Service vs. Salary",  
      xlab = "Years of Service",  
      ylab = "Salary",  
      pch = 20,  
      col = "blue"  
)
```

Scatter Plot of Years of Service vs. Salary



```
#Save the "Senior Employees" data frame as a CSV file named "senior_employees.csv."  
write.csv(Senior_Employees, "Senior_Employees.csv", row.names = FALSE)
```