# Task4.r

aravi

2023-11-10

```
print(getwd())
```

```
## [1] "C:/Users/aravi/Desktop/StatisticsUsingR/Labworks/Task4"
```

```
setwd("C:\\Users\\aravi\\Desktop\\StatisticsUsingR\\Labworks\\Task4")
print(getwd())
```

```
## [1] "C:/Users/aravi/Desktop/StatisticsUsingR/Labworks/Task4"
```

```
# Read the CSV file into a data frame
employee_data <- read.csv("employee_data.csv")

# Display the structure of the data frame
print(employee_data)
```

```
##    Employee_ID                 Name              Department Salary
## 1       123abc             John Doe   Information Technology   5000
## 2       456def           Jane Smith               Marketing   4000
## 3       789ghi        David Johnson                 Finance   6000
## 4       012jkl       Sarah Williams         Human Resources   4500
## 5       345mno        Michael Brown                   Sales   5500
## 6       678pqr          Emily Davis                 Finance   5200
## 7       901stu       Jessica Wilson        Customer Service   3800
## 8       234vwx Christopher Martinez              Operations   4800
## 9        567yz      Amanda Anderson   Information Technology   6500
## 10       890ab        Daniel Thomas       Quality Assurance   4200
## 11       123cd         Laura Garcia                 Finance   5800
## 12      456efg           Robert Lee     Business Development   5100
## 13      789hij   Jennifer Hernandez     Business Development   4700
## 14      012klm        Matthew Clark                Logistics   4300
## 15      345nop         Lauren Lewis                Logistics   5400
## 16      678qrs       Anthony Walker                Marketing   3900
## 17      901tuv         Rebecca Hall           Administration   4700
## 18      234wxy          Kevin Young   Information Technology   4400
## 19        567z    Stephanie Turner Research and Development   6200
## 20        890a         Thomas Allen        Customer Service   4100
## 21       123bc       Michelle White              Operations   4900
## 22       456de            Ryan Hall   Information Technology   6800
## 23       789fg         Rachel Scott       Quality Assurance   4500
## 24       012hi         Andrew Green                 Finance   5300
## 25       345jk       Samantha Adams     Business Development   5000
## 26       678lm          Jason Baker     Business Development   4600
## 27       901no       Melissa Turner                Logistics   4200
## 28       234pq           Erica Hill                Logistics   5600
## 29       567rs         Patrick Cook                Marketing   4000
##    Joining_Date
## 1    2020-01-15
## 2    2015-07-20
## 3    2020-03-10
## 4    2018-11-05
## 5    2017-09-12
## 6    2020-02-01
## 7    2016-04-18
## 8    2014-12-30
## 9    2020-05-25
## 10   2019-09-08
## 11   2017-03-21
## 12   2017-04-07
## 13   2019-08-14
## 14   2018-02-27
## 15   2020-06-10
## 16   2019-10-23
## 17   2017-05-06
## 18   2020-03-19
## 19   2019-11-01
## 20   2018-06-14
## 21   2017-01-27
## 22   2015-08-09
## 23   2019-12-22
## 24   2018-07-05
## 25   2020-03-21
## 26   2019-09-03
## 27   2018-01-16
## 28   2017-07-29
## 29   2020-05-12
```

```
# Calculate and add a new column for years of service
employee_data$Years_of_Service <- 2023 - as.numeric(substring(employee_data$Joining_Date,first = 1, last = 4))
print(employee_data)
```

```
##    Employee_ID                Name              Department Salary
## 1       123abc             John Doe   Information Technology   5000
## 2       456def           Jane Smith               Marketing   4000
## 3       789ghi        David Johnson                 Finance   6000
## 4       012jkl       Sarah Williams         Human Resources   4500
## 5       345mno        Michael Brown                   Sales   5500
## 6       678pqr          Emily Davis                 Finance   5200
## 7       901stu       Jessica Wilson        Customer Service   3800
## 8       234vwx Christopher Martinez             Operations   4800
## 9        567yz      Amanda Anderson   Information Technology   6500
## 10       890ab        Daniel Thomas        Quality Assurance   4200
## 11       123cd         Laura Garcia                 Finance   5800
## 12      456efg           Robert Lee    Business Development   5100
## 13      789hij   Jennifer Hernandez    Business Development   4700
## 14      012klm        Matthew Clark                Logistics   4300
## 15      345nop         Lauren Lewis                Logistics   5400
## 16      678qrs       Anthony Walker                Marketing   3900
## 17      901tuv         Rebecca Hall           Administration   4700
## 18      234wxy          Kevin Young   Information Technology   4400
## 19        567z     Stephanie Turner Research and Development   6200
## 20        890a         Thomas Allen         Customer Service   4100
## 21       123bc       Michelle White             Operations   4900
## 22       456de           Ryan Hall   Information Technology   6800
## 23       789fg         Rachel Scott        Quality Assurance   4500
## 24       012hi         Andrew Green                 Finance   5300
## 25       345jk       Samantha Adams    Business Development   5000
## 26       678lm          Jason Baker    Business Development   4600
## 27       901no       Melissa Turner                Logistics   4200
## 28       234pq           Erica Hill                Logistics   5600
## 29       567rs         Patrick Cook                Marketing   4000
##    Joining_Date Years_of_Service
## 1    2020-01-15                3
## 2    2015-07-20                8
## 3    2020-03-10                3
## 4    2018-11-05                5
## 5    2017-09-12                6
## 6    2020-02-01                3
## 7    2016-04-18                7
## 8    2014-12-30                9
## 9    2020-05-25                3
## 10   2019-09-08                4
## 11   2017-03-21                6
## 12   2017-04-07                6
## 13   2019-08-14                4
## 14   2018-02-27                5
## 15   2020-06-10                3
## 16   2019-10-23                4
## 17   2017-05-06                6
## 18   2020-03-19                3
## 19   2019-11-01                4
## 20   2018-06-14                5
## 21   2017-01-27                6
## 22   2015-08-09                8
## 23   2019-12-22                4
## 24   2018-07-05                5
## 25   2020-03-21                3
## 26   2019-09-03                4
## 27   2018-01-16                5
## 28   2017-07-29                6
## 29   2020-05-12                3
```

```
# Create a new data frame for senior employees
senior_employees <- employee_data[employee_data$Years_of_Service >= 5, ]
print(senior_employees)
```

```
##    Employee_ID                Name              Department Salary Joining_Date
## 2       456def           Jane Smith               Marketing   4000   2015-07-20
## 4       012jkl       Sarah Williams         Human Resources   4500   2018-11-05
## 5       345mno        Michael Brown                   Sales   5500   2017-09-12
## 7       901stu       Jessica Wilson        Customer Service   3800   2016-04-18
## 8       234vwx Christopher Martinez             Operations   4800   2014-12-30
## 11       123cd         Laura Garcia                 Finance   5800   2017-03-21
## 12      456efg           Robert Lee    Business Development   5100   2017-04-07
## 14      012klm        Matthew Clark                Logistics   4300   2018-02-27
## 17      901tuv         Rebecca Hall           Administration   4700   2017-05-06
## 20        890a         Thomas Allen         Customer Service   4100   2018-06-14
## 21       123bc       Michelle White             Operations   4900   2017-01-27
## 22       456de           Ryan Hall   Information Technology   6800   2015-08-09
## 24       012hi         Andrew Green                 Finance   5300   2018-07-05
## 27       901no       Melissa Turner                Logistics   4200   2018-01-16
## 28       234pq           Erica Hill                Logistics   5600   2017-07-29
##    Years_of_Service
## 2                 8
## 4                 5
## 5                 6
## 7                 7
## 8                 9
## 11                6
## 12                6
## 14                5
## 17                6
## 20                5
## 21                6
## 22                8
## 24                5
## 27                5
## 28                6
```

```
# Print average salary by department
print(tapply(employee_data$Salary, employee_data$Department, mean))
```

```
##           Administration     Business Development         Customer Service
##                 4700.000                 4850.000                 3950.000
##                  Finance         Human Resources   Information Technology
##                 5575.000                 4500.000                 5675.000
##                Logistics                Marketing               Operations
##                 4875.000                 3966.667                 4850.000
##        Quality Assurance Research and Development                    Sales
##                 4350.000                 6200.000                 5500.000
```

```
# Print highest and lowest salaries with employee details
cat("Highest Salary:\n")
```

```
## Highest Salary:
```

```
print(employee_data[employee_data$Salary == max(employee_data$Salary), c("Employee_ID", "Name", "Salary")])
```

```
##    Employee_ID      Name Salary
## 22       456de Ryan Hall   6800
```

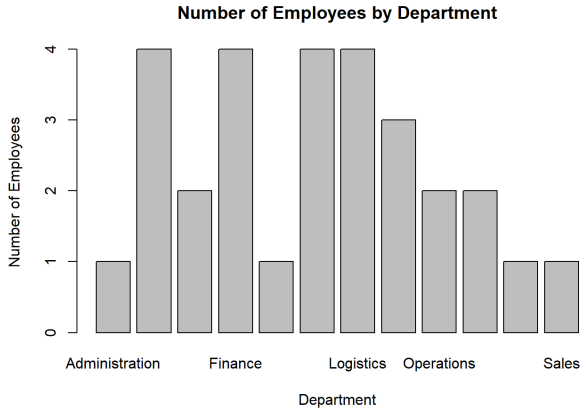
```
cat("\nLowest Salary:\n")
```

```
##
## Lowest Salary:
```

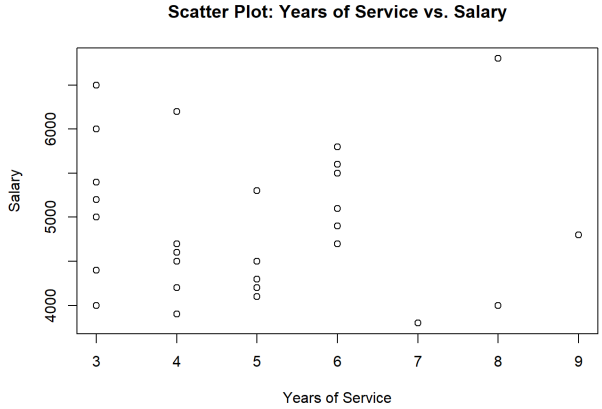
```
print(employee_data[employee_data$Salary == min(employee_data$Salary), c("Employee_ID", "Name", "Salary")])
```

```
##   Employee_ID           Name Salary
## 7      901stu Jessica Wilson   3800
```

```
# Bar plot for the number of employees in each department
barplot(table(employee_data$Department), main="Number of Employees by Department", xlab="Department", ylab="Number of Employ
ees")
```




```
# Scatter plot for years of service vs. salary
plot(employee_data$Years_of_Service, employee_data$Salary, main="Scatter Plot: Years of Service vs. Salary", xlab="Years of
Service", ylab="Salary")
```




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