Employee Growth Per Department based on Training Scores and Promotions

Neil Bhutada

2/19/2022

Introduction

Analyzing Employee growth and performance across various business units is useful for any company because they can understand how various units in the company are growing. If there is inadequate growth in the business units then companies can take initiatives to increase growth. Analyzing Employee growth and performance across various business units is useful for an employee because they can understand how other business units are doing in comparison to theirs, and know what departments will have better opportunities. As I would be venturing into the corporate world, all of the questions mentioned above are interesting to me. Hence, I chose to analyse/visualize a data set about Employee Growth and Performance.

```
url<-"https://raw.githubusercontent.com/neilbhutada/STAT-479/main/Portfolio-1/employee_promotion.csv"
data<- read_csv(url,show_col_types = FALSE)
```

Plot 1: Proportion of Employees promoted per Department

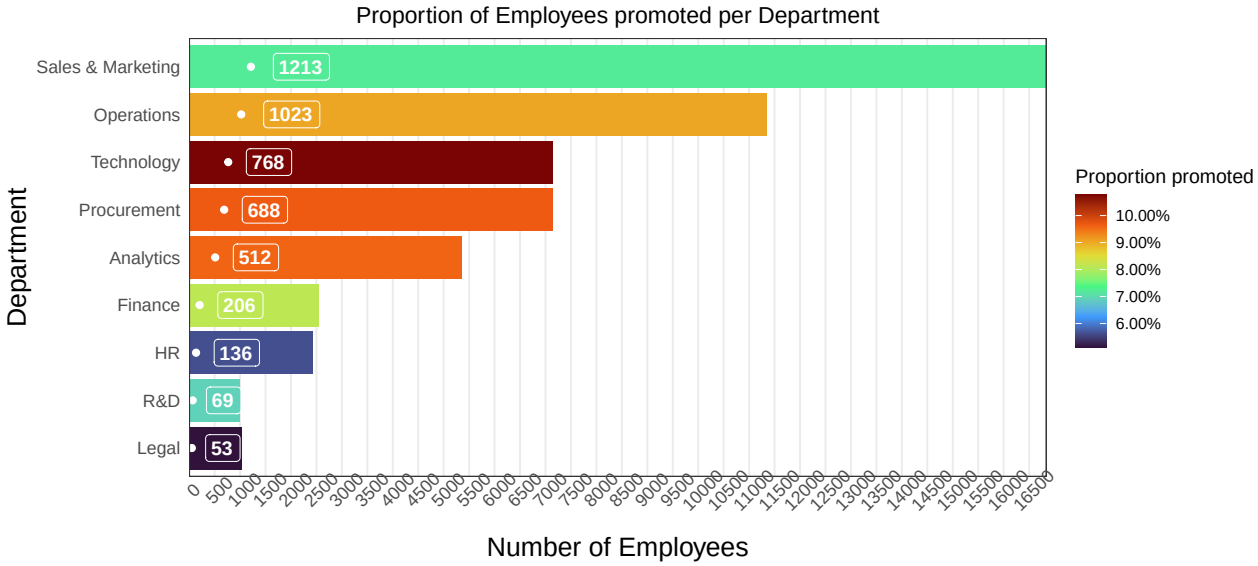
Information about the number and proportion of promotions are very good indicator about increasing opportunities and growth in the department. Hence, I made a plot to relay this information.

For creating this plot the data had to be grouped by the department name. Then for each department the following is plotted/calculated: i) the total number of employees working in the department (using geom_col) ii) proportion of employees promoted (Filling the columns of geom_col) iii) Number of employees promoted (labelled and plotted with white dot)

I used the 'scale fill Viridis turbo' option for coloring the bars because of it's distinct colors, it will be easier to identify the different promotions. I arranged the bars in descending order of their sizes.

```
data %>%
  group_by(department)%>%
  summarize(
    count = n(),
    prop_promoted = mean(is_promoted),
    sum_promoted = sum(is_promoted)
  ) %>%
  ggplot(aes(y = reorder(department, sum_promoted), fill = prop_promoted))+
  geom_col(aes(x= count))+
  geom_point(aes(x = sum_promoted), color = "white")+
  scale_x_continuous(breaks = seq(0, 17000, 500),expand = expansion(0))+
  theme_bw()+
 theme(
   axis.text.x = element_text(angle = 45, size = 10),
   axis.text.y = element_text(size = 10),
   panel.grid.major.y = element_blank(),
    axis.ticks = element_blank(),
   panel.grid.minor.x = element_blank(),
```

```
  axis.title.x = element_text(size = 15),
  axis.title.y = element_text(size = 15),
  plot.title = element_text(hjust = 0.5)
  )+
labs(
  x = "Number of Employees",
  y = "Department",
  title = "Proportion of Employees promoted per Department"
)+
scale_fill_viridis_c("Proportion promoted",option ="turbo", labels = percent)+
geom_label(aes(x = sum_promoted, label = sum_promoted),hjust = -0.4,
           color = "white", fontface = "bold")
```



Looking at the plot, Sales and Marketing has the highest number of promotions; however, Operations and Technology are growing better because they have higher proportions of promotions along with a decent number of promotions. HR and Legal really lack in terms of growth opportunities. Interestingly, Procurement and Analytics also seem to have decent growth opportunities because of the relatively greater proportions of promotions.

Plot 2: Ridge Density Plot for Average Training Scores of Employees

Information about how employees are improving their soft skills, wiz. relayed by the average scores they get during their soft-skill training sessions, are crucial to understand for improving team work and total output. Soft-skill training also speaks volumes about the work-culture. Hence, I made a ridge density plot depicting the distribution of average training scores per department.

Data preparation:

1) Calculate number of employees per department
2) Inner join this table with the parent table that has information about the training scores

Plot Creation:

1) Created geom_density_ridges with scale = 1 so that the peaks can be clearly seen without any overlap
2) The plot was filled according to number of employees for quick understanding and comparisons.

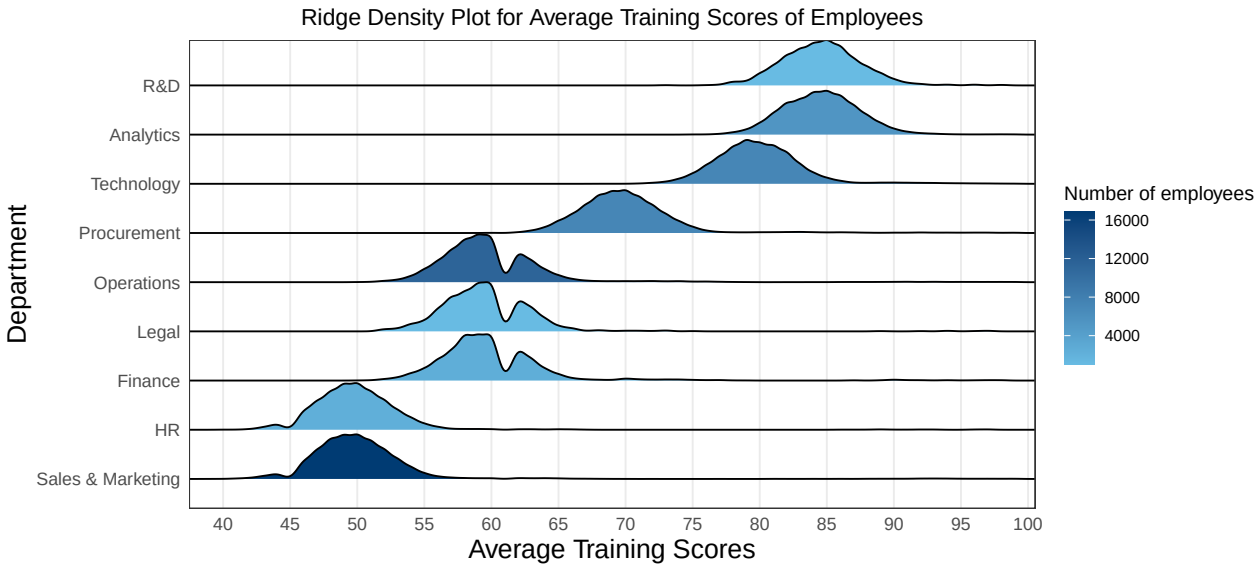
```
employees_per_department <- data %>% group_by(department) %>%
  summarise(count = n())
```

```
ridge_plot <- data %>%
  inner_join(employees_per_department, by = "department")

ridge_plot$department <- factor(ridge_plot$department, levels
                                = rev(c("R&D", "Analytics", "Technology",
                                        "Procurement","Operations",
                                        "Legal", "Finance",
                                        "HR","Sales & Marketing")),
                                        ordered = TRUE)

ridge_plot%>%
  ggplot(aes(y = department, x = avg_training_score, fill = count))+
  geom_density_ridges(scale = 1)+
  scale_fill_gradient("Number of employees", low = "#68BBE3",
                      high = "#003B73",
                      guide = "colorbar")+
  theme_bw()+
  scale_x_continuous(breaks = seq(0, 100, 5), expand = expansion(0))+
  labs(
    x = "Average Training Scores",
    y = "Department",
    title = "Ridge Density Plot for Average Training Scores of Employees"
  )+theme(
   axis.text.x = element_text(size = 10),
   axis.text.y = element_text(size = 10),
   panel.grid.major.y = element_blank(),
    axis.ticks = element_blank(),
   panel.grid.minor.x = element_blank(),
   axis.title.x = element_text(size = 15),
   axis.title.y = element_text(size = 15),
   plot.title = element_text(hjust = 0.5)
   )
```

```
## Picking joint bandwidth of 0.511
```



R&D and Analytics have done the better in soft-skill trainings. Most departments have reasonably one

peak. Legal, Finance, and Operations has two peaks at around 60 and 65, and their shapes align. Similarly, HR and Sales&Marketing have density shapes that are aligned and similar in ranges. The most concerning department is HR because of low scores, lower number of employees, and high requirement of soft-skills in the job