

Salary

Yang Yuetong

2025-09-15

READ ME

This piece of analysis was collaborated with other group members as part of a school assignment.

```
hr <- read.csv("hr2.csv", stringsAsFactors = TRUE)
```

Which department records the lowest average salary, and what is the most common reason for leaving the current position?

Productions department has the lowest average salary at \$59 953.55, and the most common reason for leaving the position is having found another position.

```
Q3Salary <- hr %>% group_by(department) %>% summarize("Average Salary" = mean(salary))

#Group by department and Termination, with the counts of the respective reason
dpt_TR_group <- hr %>% group_by(department, terminationreason) %>% summarize(Count = n())

#Exclude employees still employed
exclude_employed <- dpt_TR_group %>% filter(terminationreason!= "N/A-StillEmployed")

#Extract top termination reason
Q3MaxCount <- exclude_employed %>% group_by(department) %>% filter(Count==max(Count))

#Combine with Mean salary
Q3Table <- inner_join(Q3Salary, Q3MaxCount, by="department")
colnames(Q3Table) <- c("Department", "Average Salary", "Top Termination Reason", "Count")
kbl(Q3Table, booktabs = T) %>%
  kable_styling(latex_options = c("striped", "scale_down", "hold_position"))
```

Department	Average Salary	Top Termination Reason	Count
Admin Offices	71791.89	no-call, no-show	2
IT/IS	97064.64	career change	3
Production	59953.55	Another position	17
Sales	69061.26	relocation out of area	2
Software Engineering	94989.45	medical issues	2

We hypothesise that employees who are paid lower are more likely to voluntarily terminate their job.

Employment Status and Salary

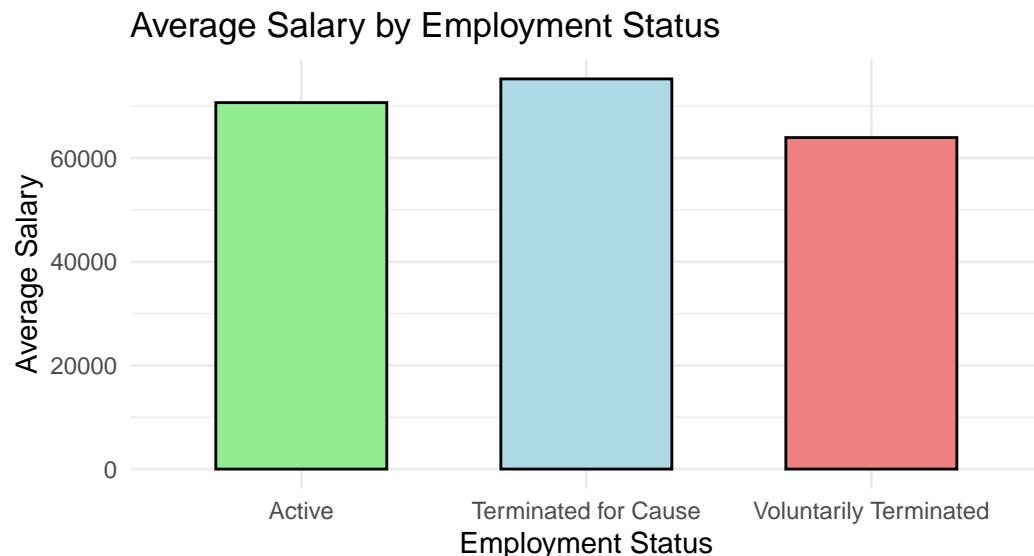
Firstly, active employees earn on average **\$70,694**, while voluntarily terminated employees earn significantly less at an average of **\$63,949**.

Additionally, it seems that those who terminate voluntarily have slightly higher number of absences on average compared to active employees. This may suggest disengagement before resigning, and hence absenteeism can be a potential indicator of an employee's likelihood to terminate.

```
Q4status <- hr %>% group_by(employmentstatus) %>% summarize("Count" = n(),  
                                                             "Average Salary" = mean(salary),  
                                                             "Average Absences" = mean(absences))  
  
kbl(Q4status, booktabs = T) %>%  
  kable_styling(latex_options = c("striped", "scale_down", "hold_position"))
```

employmentstatus	Count	Average Salary	Average Absences
Active	207	70694.03	9.830918
Terminated for Cause	16	75264.88	11.562500
Voluntarily Terminated	88	63949.20	10.954546

```
ggplot(Q4status, aes(x = employmentstatus, y = `Average Salary`, fill = employmentstatus)) +  
  geom_bar(stat = "identity", width = 0.6, color = "black", linewidth = 0.5) +  
  scale_fill_manual(values = c("lightgreen", "lightblue", "lightcoral")) +  
  labs(title = "Average Salary by Employment Status",  
       x = "Employment Status", y = "Average Salary") +  
  theme_minimal() +  
  theme(legend.position = "none")
```



Permutation Testing

To test if lower salary is indeed related to increased voluntary termination, permutation testing can be conducted.

H0 (null hypothesis): Salary is unrelated to voluntary termination

H1 (alternative hypothesis): Employees with lower salaries are more likely to voluntarily terminate their employment.

```
set.seed(123)
status <- hr$employmentstatus
salary <- hr$salary
meanquit <- hr[hr$employmentstatus == "Voluntarily Terminated", "salary"] %>% mean()
meanactive <- hr[hr$employmentstatus == "Active", "salary"] %>% mean()
obs <- meanactive - meanquit

#Permutation
N <- 10^4
result <- numeric(N)
for (i in 1:N){
  #Take a permutation sample w/o replacement
  shufflestatus <- sample(status, size=length(status), replace=FALSE)
  #For each permutation, calculate test statistic
  result[i] <- mean(salary[shufflestatus=="Active"])-
    mean(salary[shufflestatus=="Voluntarily Terminated"])
}
#One sided test
pvalue <- (sum(result>=obs) + 1) / (N+1)
glue('p-value is {round(pvalue,4)}.')
```

p-value is 0.0117.

Since $p\text{-value} = 0.0117 < 0.05$, we can reject Null Hypothesis and lower salary is likely associated with higher voluntary termination.

Termination Reason and Salary

To further investigate reasons for termination, we create a table to visualise what are frequent reasons cited for those who voluntarily terminate their job.

As seen below, “more money” constitutes the 3rd most frequently cited reason for voluntary termination. Correspondingly, those who cite “more money” as a termination reason also has **lower salary** at an average of \$57 148. It should be noted that other reasons including finding another position, or unhappiness also constitute the top 2 reasons for voluntary termination.

```
Q4reason <- hr %>% filter(employmentstatus == "Voluntarily Terminated") %>%
  group_by(terminationreason) %>%
  summarize("Average Salary" = mean(salary), "Count" = n())

kbl(arrange(Q4reason, desc(Count)), booktabs = T) %>%
  kable_styling(latex_options = c("striped", "scale_down", "hold_position"))
```

Voluntary Termination Across Departments

Categorised by departments, **productions** department has the **lowest average salary** (\$59,954) and the **highest termination rate** (35.9%).

terminationreason	Average Salary	Count
Another position	65684.84	19
unhappy	61348.93	14
more money	57148.45	11
career change	74915.12	8
hours	65264.86	7
relocation out of area	63408.83	6
return to school	55671.00	5
medical issues	76264.50	4
military	59981.75	4
retiring	60527.75	4
maternity leave - did not return	63800.00	3
attendance	50428.00	1
no-call, no-show	55000.00	1
performance	93093.00	1

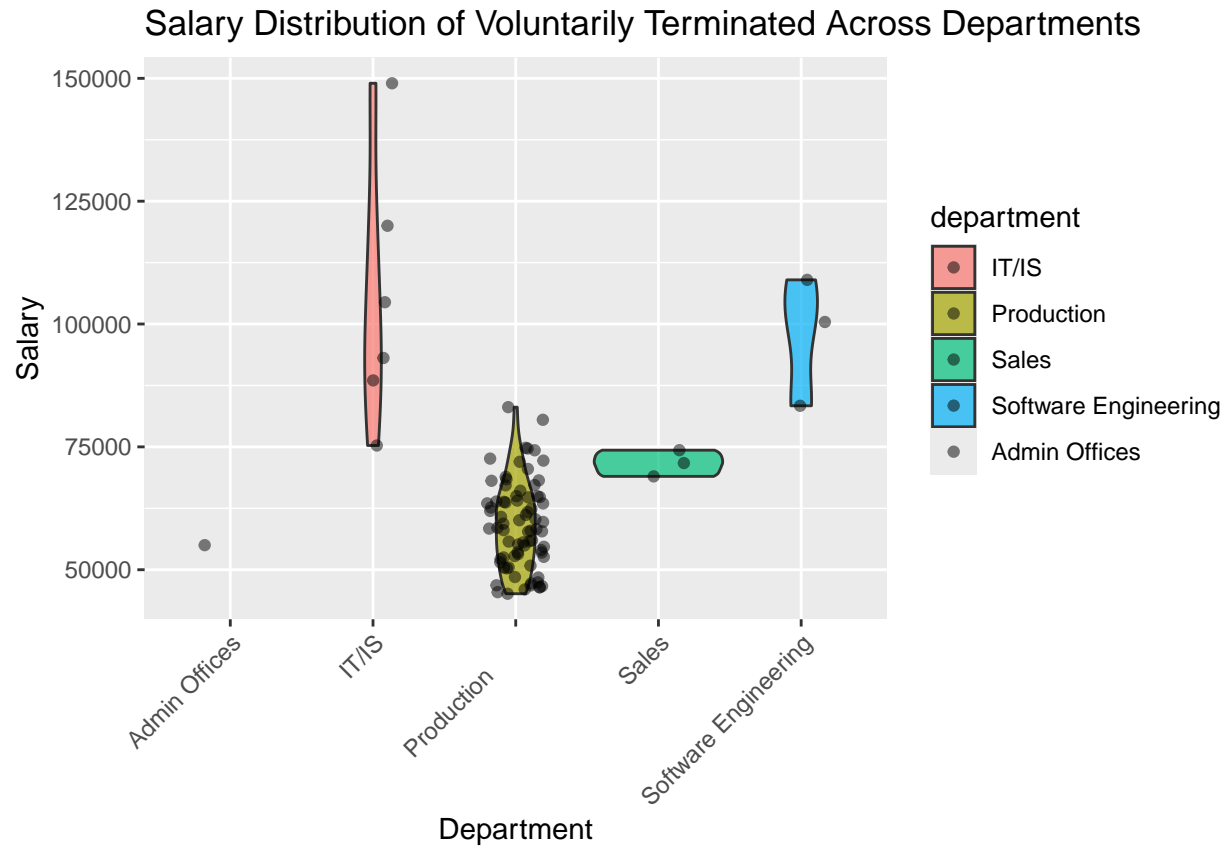
```
Q4department <- hr %>% group_by(department) %>%
  summarize("Proportion terminated (%)" =
    round(sum(employmentstatus=="Voluntarily Terminated")*100/n(),4),
    "Average Salary" = mean(salary))
kbl(Q4department, booktabs = T) %>%
  kable_styling(latex_options = c("striped", "scale_down", "hold_position"))
```

department	Proportion terminated (%)	Average Salary
Admin Offices	11.1111	71791.89
Executive Office	0.0000	250000.00
IT/IS	12.0000	97064.64
Production	35.8852	59953.55
Sales	9.6774	69061.26
Software Engineering	27.2727	94989.45

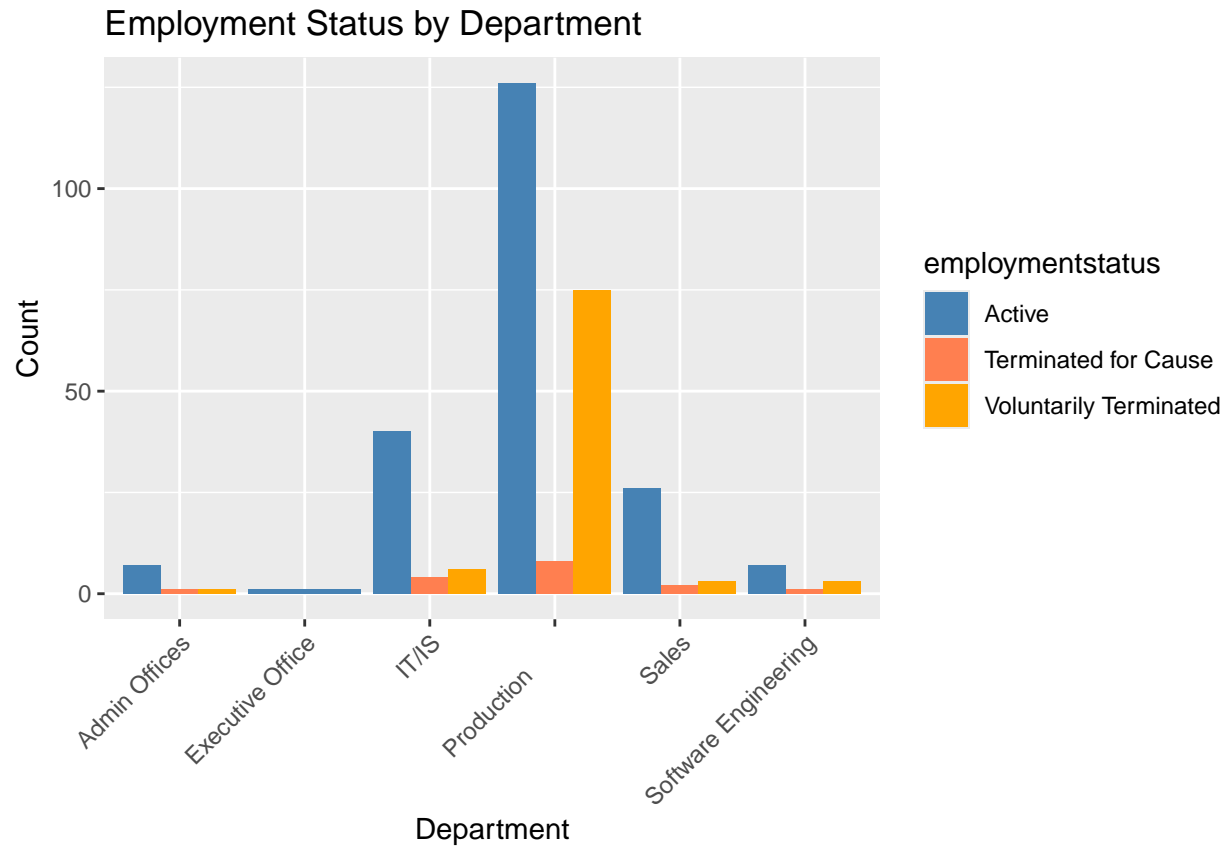
Based on the plot below, among those who voluntarily terminated their employment, those in **production** have the **lowest salary ranges**, tightly clustered between \$50k–70k, corroborating with the above findings. More terminations are observed in the productions department compared to others, suggesting that these employees may feel underpaid compared to those in other departments, causing them to leave.

```
all_terminated <- hr %>% filter(employmentstatus == "Voluntarily Terminated") %>%
  select("department", "salary", "performancescore", "jobsatisfactionscore", "terminationreason")

ggplot(all_terminated, aes(x = department, y = salary, fill = department)) +
  geom_violin(alpha = 0.7) +
  geom_jitter(width = 0.2, alpha = 0.5, color = "black") +
  labs(title = "Salary Distribution of Voluntarily Terminated Across Departments",
    x = "Department", y = "Salary") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



```
ggplot(hr, aes(x = department, fill = employmentstatus)) +
  geom_bar(position = "dodge") +
  labs(title = "Employment Status by Department",
       x = "Department", y = "Count") +
  scale_fill_manual(
    values = c(
      "Active" = "steelblue",
      "Voluntarily Terminated" = "orange",
      "Terminated for Cause" = "coral"
    )
  ) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

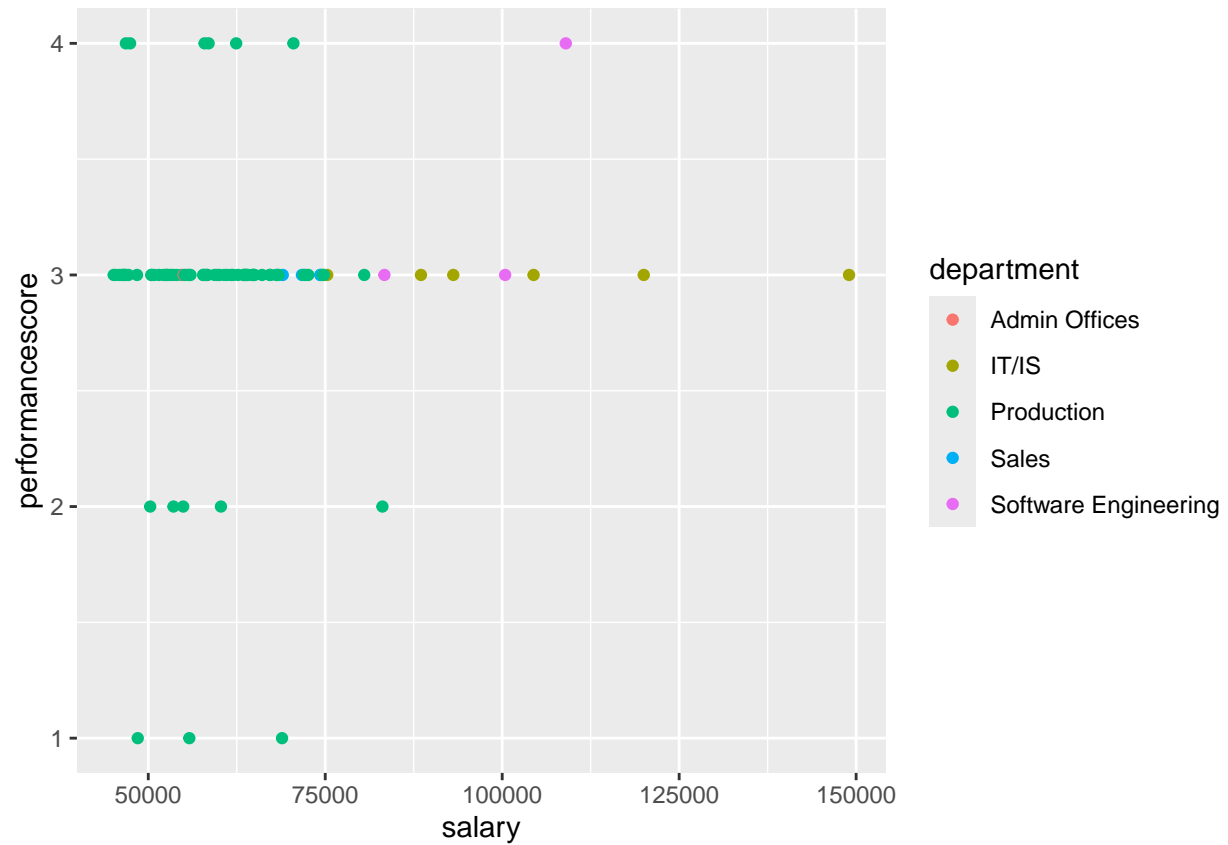


Production has the highest termination rate, mostly voluntarily.

Job satisfaction and Performance scores

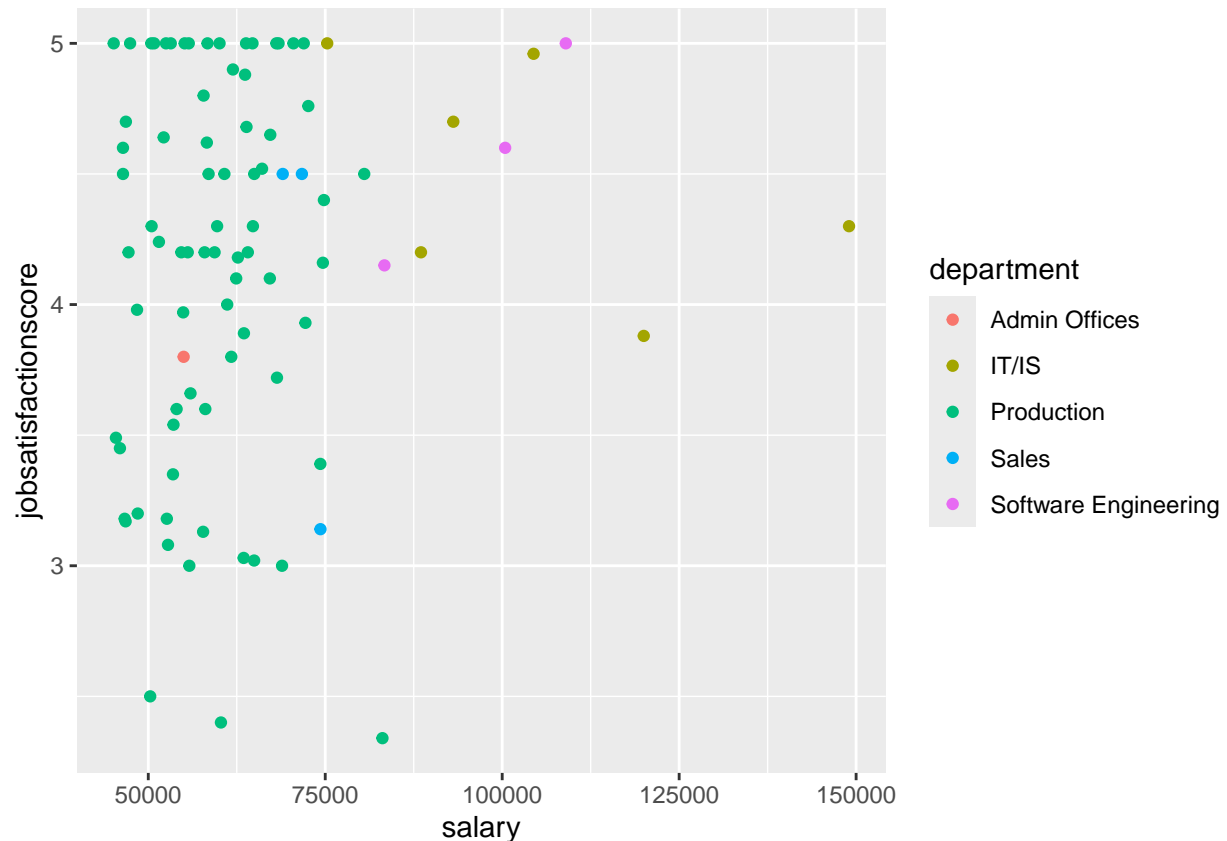
Are these terminated employees performing worse or are less satisfied (if reflected in the scores)?

```
ggplot(data = all_terminated, mapping = aes(x = salary, y = performancescore)) +  
  geom_point(aes(color = department))
```



Most terminated employees have a performance score of 3. We cannot conclude that they perform worse.

```
ggplot(data = all_terminated, mapping = aes(x = salary, y = jobsatisfactionscore)) +  
geom_point(aes(color = department))
```



Job satisfaction score seems quite varied for production employees.

Voluntary Termination Across Roles

Building on the previous results, we segment employees into Director, Manager, Senior, and Regular Employee groups and compare the average pay of voluntary leavers and active employees within each peer group, thereby isolating seniority from department mix.

```
#voluntarily terminated employees across roles
df_tag <- hr %>% filter(employmentstatus == "Voluntarily Terminated")

directors <- df_tag %>% filter(str_detect(position, regex("Director", ignore_case = T)) &
                              !str_detect(position, regex("Manager|Sr|Senior", T)))

managers <- df_tag %>% filter(str_detect(position, regex("Manager", ignore_case = T)) &
                              !str_detect(position, regex("Director|Sr|Senior", T)))

seniors <- df_tag %>% filter(str_detect(position, regex("Sr|Senior", ignore_case = T)) &
                              !str_detect(position, regex("Director|Manager", T)))

indv <- df_tag %>% filter(!str_detect(position,
                                       regex("Director|Manager|Sr|Senior", ignore_case = T)))

volTerm <- bind_rows(
  directors %>% mutate(role_group = "Director"),
  managers %>% mutate(role_group = "Manager"),
  seniors %>% mutate(role_group = "Senior"),
```



```

  indiv %>% mutate(role_group = "Regular Employee")
) %>%
  group_by(role_group) %>%
  summarise(
    no_vol_leavers = n(), # count (sum of people) in each cat
    mean_salary_volLeavers = mean(salary, na.rm = TRUE), # mean salary per cat
    .groups = "drop"
  ) %>%
  arrange(factor(role_group, levels = c("Director", "Manager", "Senior", "Regular Employee")))

#active employees across roles
df_tag <- hr %>% filter(employmentstatus == "Active")

directorsS <- df_tag %>% filter(str_detect(position, regex("Director", ignore_case = T)) &
  !str_detect(position, regex("Manager|Sr|Senior", T)))

managersS <- df_tag %>% filter(str_detect(position, regex("Manager", ignore_case = T)) &
  !str_detect(position, regex("Director|Sr|Senior", T)))

seniorsS <- df_tag %>% filter(str_detect(position, regex("Sr|Senior", ignore_case = T)) &
  !str_detect(position, regex("Director|Manager", T)))

indvS <- df_tag %>% filter(!str_detect(position,
  regex("Director|Manager|Sr|Senior", ignore_case = T)))

Active <- bind_rows(
  directorsS %>% mutate(role_group = "Director"),
  managersS %>% mutate(role_group = "Manager"),
  seniorsS %>% mutate(role_group = "Senior"),
  indvS %>% mutate(role_group = "Regular Employee")
) %>%
  group_by(role_group) %>%
  summarise(
    no_stayers = n(), # count (sum of people) in each cat
    mean_salary_stayers = mean(salary, na.rm = TRUE), # mean salary per cat
    .groups = "drop"
  ) %>%
  arrange(factor(role_group, levels = c("Director", "Manager", "Senior", "Regular Employee")))

comparison <- full_join(volTerm, Active, by = "role_group", c("Voluntarily Terminated, Active"))
Q4roles <- comparison %>%
  mutate(perLeave = round(no_vol_leavers*100/(no_vol_leavers + no_stayers),3),
    #Calculate % of people who leave
    perPayDiff = if_else(!is.na(mean_salary_volLeavers) & !is.na(mean_salary_stayers),
    #Calculate % difference in pay
    round((mean_salary_volLeavers - mean_salary_stayers)*100 / mean_salary_stayers,3), NA_real_))
colnames(Q4roles) <- c("Role Type", "No. of Voluntarily Terminated", "Mean Salary (Voluntarily Terminated)",
  "No. of Active", "Mean Salary (Active)", "Proportion Terminated (%)",
  "Pay Difference (Voluntarily Terminated - Active) (%)")
kbl(Q4roles, booktabs = T, align='c') %>%
  kable_styling(latex_options = c("striped", "scale_down", "hold_position"))

```

From this, we can see that contrary to our initial findings, a lower salary does not necessarily mean that the

Role Type	No. of Voluntarily Terminated	Mean Salary (Voluntarily Terminated)	No. of Active	Mean Salary (Active)	Proportion Terminated (%)	Pay Difference (Voluntarily Terminated - Active) (%)
Manager	8	84575.75	39	73914.00	17.021	14.425
Senior	1	104437.00	11	93228.27	8.333	12.023
Regular Employee	79	61347.94	153	65922.08	34.052	-6.939
Director	NA	NA	4	159857.25	NA	NA

employee is more likely to terminate their contract.

For **higher roles like Managers and Seniors**, departures are not explained by lower pay. In these groups, leavers are on average better paid than stayers (~12-14% higher), suggesting pull factors (external offers, role fit, scope) rather than compensation.

For **regular employees**, departures can be explained by lower pay. Leavers on average have a lower salary than their counterparts who stay (~6% lower).

However, we do have to interpret with caution for small samples. Senior has only one leaver and Director has none, limiting statistical confidence for those groups.

Conclusion

In conclusion, our analysis supports the hypothesis that **lower salary is associated with higher voluntary termination, more often for regular employees**.

1. Employees with lower salary are significantly more likely to leave (p-value = 0.0117), with many citing compensation (“more money”) or dissatisfaction (“unhappy”) as reasons to leave. On the other hand, those leaving for career change or health-related reasons earned more, suggesting these exits are less preventable.
2. Looking into departments, the **productions** department with the lowest mean salary has the highest rates of resignation.
3. When considering role types, regular employees who voluntarily terminate their employment have lower salary than those who stay (~6% lower). However, higher ranking roles (Manager, Seniors) who voluntarily terminate their employment typically have higher salary than active ones, suggesting other pull factors besides from low pay among this group.

Proposed Strategy

To retain employees, HR can ensure pay is competitive for roles with higher resignation rates, with greater attention to the **Productions** department.

The company can **offer higher salaries or bonuses** for individuals with lower salaries, especially regular employees, to reduce likelihood of voluntary termination.

HR can also **track employees with increased absenteeism** as this may serve as an indicator of likelihood of resignation, with those who resign having slightly higher absentee rates than active employees.

Finally, for higher ranking roles like managers or seniors, HR can explore career growth opportunities, including trainings or internal mobility offers so that employees do not feel that they must leave to progress.