Employee Attrition Analysis

Nabeel Ghalib

18-05-2024

Project overview

This Data Analysis project aims to provide insights into Attrition of employees from XYZ company. By Analyzing various aspects of the data we can identify trends, make Data-driven recommendation to improve the company.

Problem Statement

XYZ company which was established a few years back is facing around a 15% attrition rate for a couple of years. And it's majorly affecting the company in many aspects. In order to understand why employees are leaving the company and reduce the attrition rate XYZ company has approached an HR analytics consultancy for analyzing the data they have. You are playing the HR analyst role in this project and building a dashboard which can help the organization in making data-driven decisions.

ASK

The key business task is to identify the reason employees are leaving the company,

1. Finding out total employees
2. Calculating the attrition rate
3. Finding out the reason for attrition

Data Preparation

The dataset used is provided by Unified Mentor Private Limited which was provided for my Data Analytics internship program.

Note - The XYZ is a fictional company.

Tools Used

RStudio - Data cleaning, Analyzing, and Visualization
Tableau - Data Visualization

Installing required packages

```r
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.3.3

## Warning: package 'ggplot2' was built under R version 4.3.2

## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.3     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.4     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(tidyr)
library(dplyr)
library(ggplot2)
library(janitor)
```

```
##
## Attaching package: 'janitor'
##
## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

```r
library(forcats) # to reorder by values, variables etc..
library(scales) # to use percent()
```

```
##
## Attaching package: 'scales'
##
## The following object is masked from 'package:purrr':
##
##     discard
##
## The following object is masked from 'package:readr':
##
##     col_factor
```

Importing the dataset

Importing the dataset and storing it in a data frame

```r
employee_attrition_data = read.csv("F:/Rprojects/Rprojects/Projects to work/Employee Attrition data.csv
```

DATA CLEANING

Finding null values and na values

```r
print(paste0("There are ",nrow(employee_attrition_data)," rows" ))
```

```
## [1] "There are 4410 rows"
```

```r
print(paste0("There are ",ncol(employee_attrition_data)," columns"))
```

```
## [1] "There are 29 columns"
```

```r
print(paste0("There are ",n_distinct(employee_attrition_data)," distinct rows"))
```

```
## [1] "There are 4410 distinct rows"
```

```r
print(paste0("There are ",sum(is.null(employee_attrition_data))," null values"))
```

```
## [1] "There are 0 null values"
```

```r
print(paste0("There are ",sum(is.na(employee_attrition_data))," na values"))
```

```
## [1] "There are 111 na values"
```

```r
print(paste0("There are ",sum(is.na(employee_attrition_data$EmployeeID))," na values in EmployeeID"))
```

```
## [1] "There are 0 na values in EmployeeID"
```

Removing na values

```r
employee_attrition_data = employee_attrition_data %>%
  drop_na()
```

Checking Number of rows, columns and distinct values after removing na values

```r
print(paste0("There are ",nrow(employee_attrition_data)," rows"))
```

```
## [1] "There are 4300 rows"
```

```r
print(paste0("There are ",ncol(employee_attrition_data)," columns"))
```

```
## [1] "There are 29 columns"
```

```r
print(paste0("There are ",n_distinct(employee_attrition_data)," distinct rows"))
```

```
## [1] "There are 4300 distinct rows"
```

```r
n_distinct(employee_attrition_data$BusinessTravel)
```

```
## [1] 3
```

```r
n_distinct(employee_attrition_data$Attrition)
```

```
## [1] 2
```

```r
n_distinct(employee_attrition_data$JobRole)
```

```
## [1] 9
```

```r
n_distinct(employee_attrition_data$Gender)
```

```
## [1] 2
```

```r
n_distinct(employee_attrition_data$JobLevel)
```

```
## [1] 5
```

the data is cleaned and ready for analysis.

DATA ANALYSIS

Total Employees

```r
total_employees = employee_attrition_data %>%
  select(EmployeeCount) %>%
  summarise(total_employees = sum(EmployeeCount))

print(paste0("There are ",total_employees," employees"))
```

```
## [1] "There are 4300 employees"
```

Employee Attrition Count and Attrition rate

```r
emp_att_count2 = employee_attrition_data %>%
  select(Attrition) %>%
  count(Attrition, name = 'total_employees') %>%
  summarise(Attrition, total_employees, attrition_rate = round(total_employees/sum(total_employees)* 10

emp_att_count2
```

```
##   Attrition total_employees attrition_rate
## 1        No            3605          83.84
## 2       Yes             695          16.16
```

The attrition count is 695 and the attrition rate is 16.16%

Active Employee

```r
 active_employee = emp_att_count2 %>%
  select(Attrition, total_employees) %>%
  filter(Attrition == "No")

print(paste0('There are ',active_employee$total_employees ,' active employees'))
```

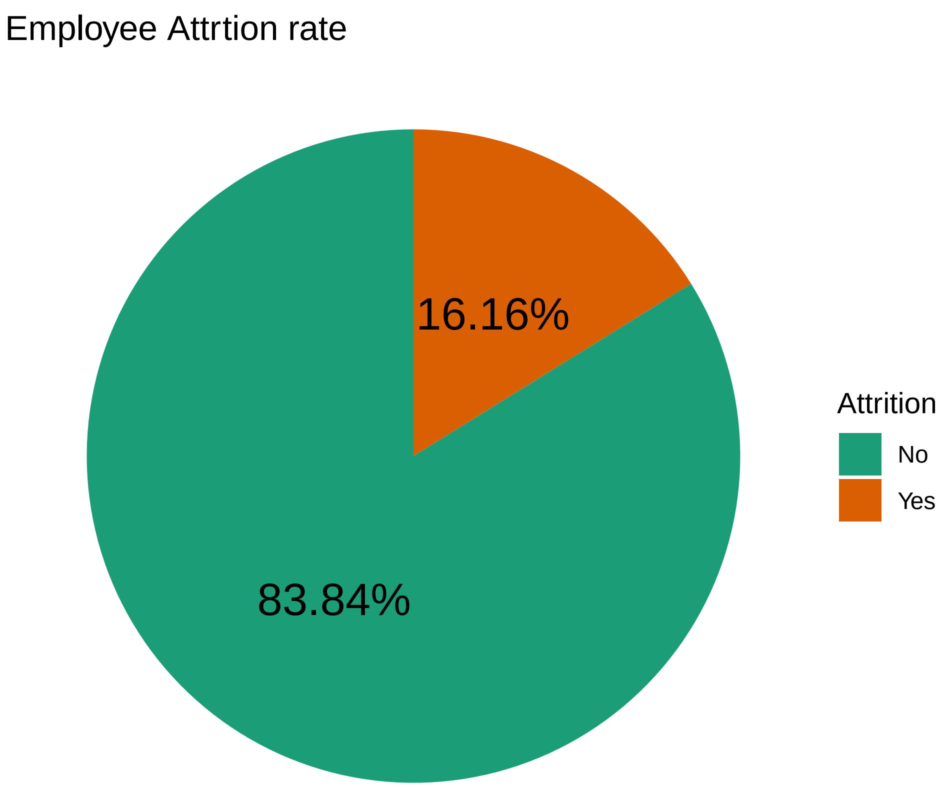
```
## [1] "There are 3605 active employees"
```

Attrition rate pie chart

```r
# pie chart attrition rate
# calculation to label the values in their respective positions

empatt_count_pie = emp_att_count2

empatt_count_pie = empatt_count_pie %>%
  arrange(desc(Attrition)) %>%
  mutate(prop = (total_employees / sum(empatt_count_pie$total_employees)))%>%
  mutate(ypos = cumsum(prop)- 0.5 * prop)

empatt_count_pie
```

```
##   Attrition total_employees attrition_rate      prop       ypos
## 1       Yes             695          16.16 0.1616279 0.08081395
## 2        No            3605          83.84 0.8383721 0.58081395
```

```r
ggplot(empatt_count_pie, aes(x="", y = prop , fill= Attrition)) +
  geom_bar(stat="identity", width=1) +
  coord_polar("y", start=0) +
  labs(title = 'Employee Attrtion rate') +
  theme_void() + # remove background, grid, numeric labels
  geom_text(aes(y = ypos, label = percent(prop,accuracy = 0.01)), color = 'black',size = 6)+
  scale_fill_brewer(palette="Dark2")
```



- The attrition rate is 16,16%

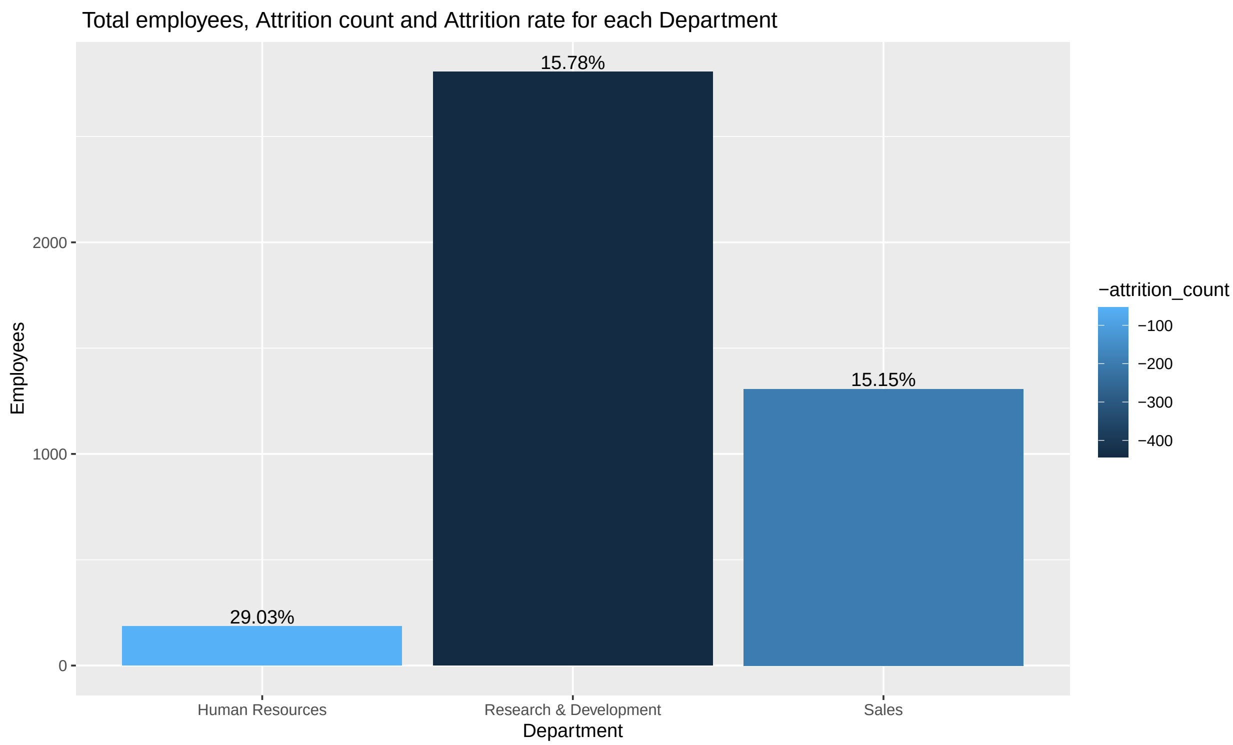
Total employees and Attrition count from each department

```
# merging emp_dep , dep_att by department

dept_att = merge(emp_dep,dep_att, by = c("Department","Department"))

dept_att = dept_att %>%
  arrange(-attrition_count)

dept_att = dept_att %>%
  select(Department,total_employees,attrition_count) %>%
  mutate(attrition_rate = (attrition_count/total_employees)) %>%
  mutate(proportion_of_attrition = (attrition_count/sum(attrition_count)))

dept_att
```

```
##               Department total_employees attrition_count attrition_rate
## 1 Research & Development            2807             443      0.1578197
## 2                  Sales            1307             198      0.1514920
## 3        Human Resources             186              54      0.2903226
##   proportion_of_attrition
```

```
## 1              0.63741007
## 2              0.28489209
## 3              0.07769784
```

```
# Bar graph


ggplot(data = dept_att, aes(x=Department, y = total_employees, attrition_count, fill = - attrition_coun
  geom_col(position = "dodge") + labs(title = " Total employees, Attrition count and Attrition rate for
  geom_text(aes(label = percent(attrition_rate)), vjust = -0.2)
```

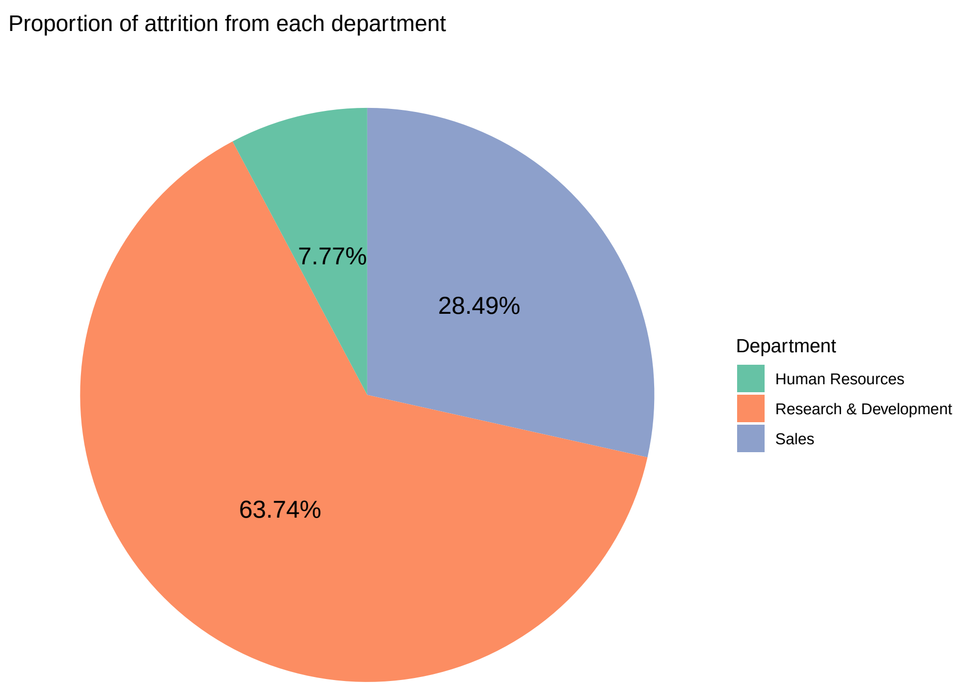



```
# pie chart



dept_att_pie = dept_att

dept_att_pie = dept_att_pie %>%
  arrange(desc(Department)) %>%
  mutate(prop = attrition_count / sum(attrition_count)) %>%
  mutate(ypos = cumsum(prop)- 0.5 * prop)

dept_att_pie
```

```
##               Department total_employees attrition_count attrition_rate
## 1                  Sales            1307             198      0.1514920
## 2 Research & Development            2807             443      0.1578197
## 3        Human Resources             186              54      0.2903226
##   proportion_of_attrition       prop       ypos
```

```
## 1              0.28489209 0.28489209 0.1424460
## 2              0.63741007 0.63741007 0.6035971
## 3              0.07769784 0.07769784 0.9611511
```

```
ggplot(data = dept_att_pie, aes (x=" ", y = prop, fill = Department))+
  geom_bar(stat= "identity", width = 1) +
  coord_polar("y", start = 0) +
  labs(title = "Proportion of attrition from each department") +
  theme_void() +
  geom_text(aes(y = ypos, label = percent(prop, accuracy = 0.01)), color = "black", size = 5) +
  scale_fill_brewer(palette="Set2")
```



- Highest attrition count is from Research & Development Department, Out of 2807 employees 443 left (63.74%)

- Highest attrition rate (%) is from Human Resources Department, Out of 186 employees 54 left (29.03%)

- Highest proportion of attrition is 64% from Research & Development Department

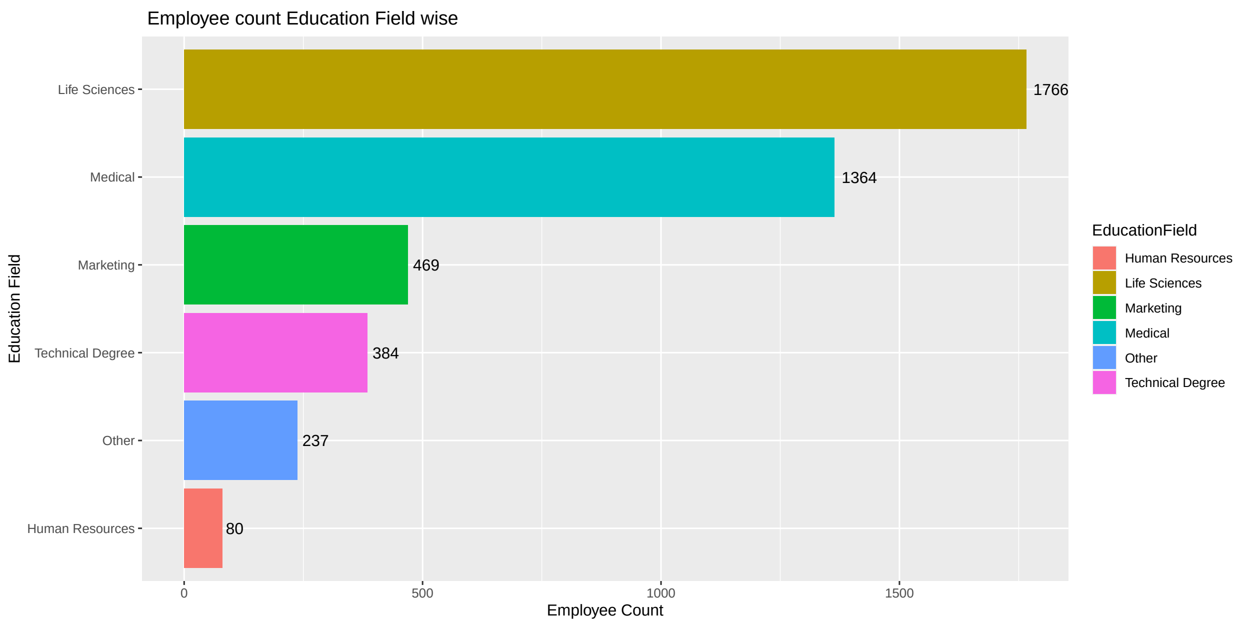
Education field wise total employees and attrition

```
eduf_att_tot = employee_attrition_data %>%
  select(EducationField,Attrition) %>%
  group_by(EducationField) %>%
```

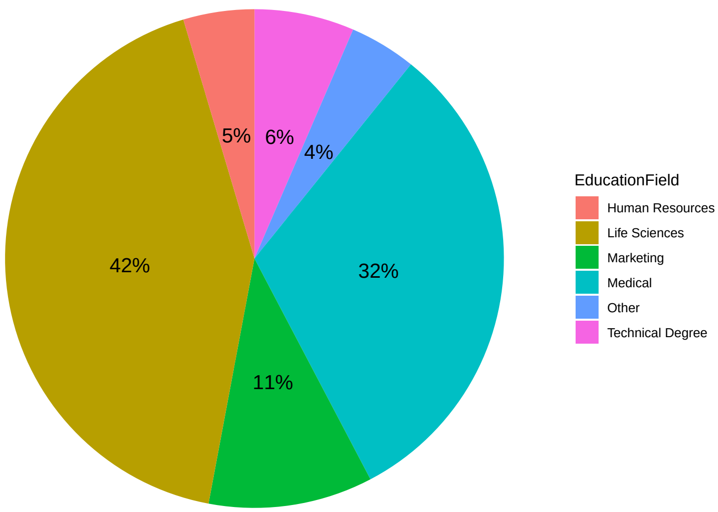
```r
  count(Attrition, name ='attrition_count') %>%
  reframe(EducationField,Attrition, attrition_count, total_employees=sum(attrition_count)) %>%
  arrange(-total_employees,EducationField)

eduf_att_tot = eduf_att_tot %>%
  filter(Attrition == "Yes")

eduf_att_tot = eduf_att_tot %>%
  select(EducationField, attrition_count , total_employees)

eduf_att_tot
```

```
## # A tibble: 6 x 3
##   EducationField   attrition_count total_employees
##   <chr>                      <int>           <int>
## 1 Life Sciences                295            1766
## 2 Medical                      219            1364
## 3 Marketing                     74             469
## 4 Technical Degree              45             384
## 5 Other                         30             237
## 6 Human Resources               32              80
```

```r
# Horizontal bar chart for education field employee count

ggplot(data = eduf_att_tot,aes(x = reorder(EducationField, total_employees), y = total_employees, fill =
  geom_bar(stat = "identity") +
  coord_flip() +
  labs(title = " Employee count Education Field wise", x= 'Education Field', y = 'Employee Count') +
  geom_text(aes(label = total_employees), hjust = -0.2)
```



```
# pie chart

eduf_att_tot_pie = eduf_att_tot %>%
  arrange(desc(EducationField)) %>%
  mutate(prop = attrition_count/sum(attrition_count)) %>%
  mutate(ypos = cumsum(prop) -0.5 * prop)

ggplot(data = eduf_att_tot_pie,aes(x= "", y = prop,fill= EducationField)) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar("y", start = 0)+
  theme_void()+
  geom_text(aes(y = ypos , label = percent(prop, accuracy = 1)),color="black",size = 5)
```



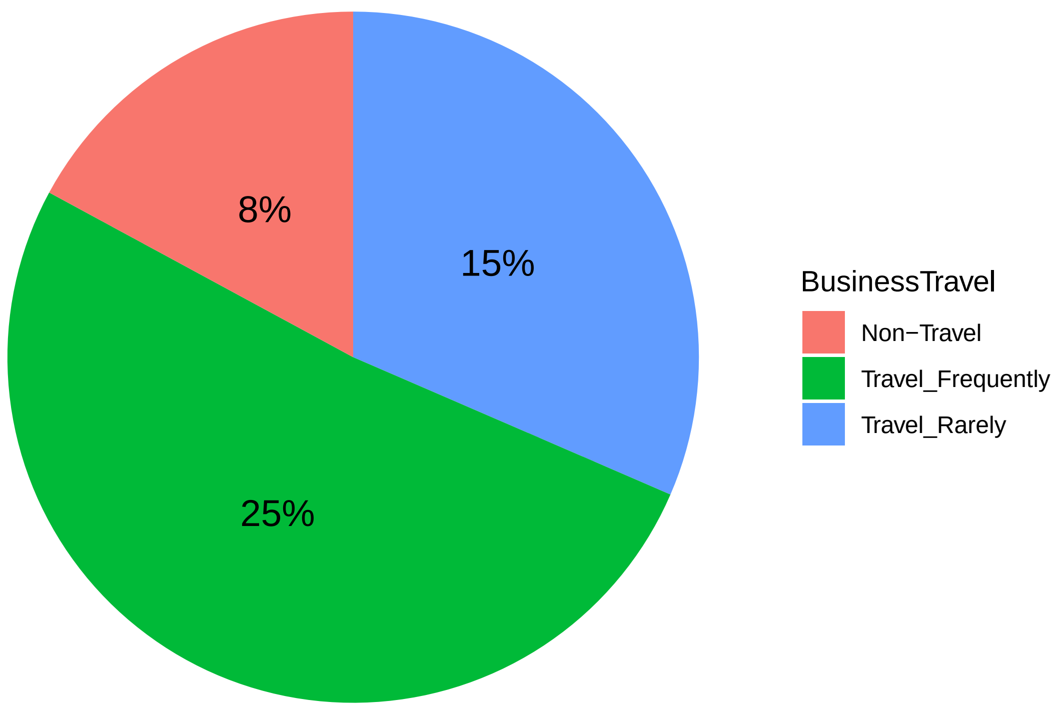
The Highest attrition is from Life Sciences Education Field and then Medical

Total employees and attrition count Business Travel wise

```
bus_trav_att = employee_attrition_data %>%
  select(BusinessTravel, Attrition) %>%
  group_by(BusinessTravel) %>%
  count(Attrition , name = 'attrition_count') %>%
  reframe(BusinessTravel, Attrition,attrition_count , total_employees = sum(attrition_count), attrition_

bus_trav_att = bus_trav_att %>%
  select(BusinessTravel ,Attrition, attrition_count, total_employees, attrition_rate) %>%
  filter(Attrition == "Yes")

bus_trav_att = bus_trav_att %>%
  select(BusinessTravel , attrition_count, total_employees, attrition_rate)

bus_trav_att
```

```
## # A tibble: 3 x 4
##   BusinessTravel    attrition_count total_employees attrition_rate
##   <chr>                       <int>           <int> <chr>
## 1 Non-Travel                     36             440 8%
## 2 Travel_Frequently             199             809 25%
## 3 Travel_Rarely                 460            3051 15%
```

```r
# pie

bus_trav_att_pie = bus_trav_att %>%
  arrange(desc(BusinessTravel)) %>%
  mutate(prop = (attrition_count/total_employees)) %>%
  mutate(ypos = cumsum(prop) - 0.5 * prop)



ggplot(bus_trav_att_pie, aes(x ="" , y = prop, fill = BusinessTravel))+
  geom_bar(stat = "identity", width = 1) +
  coord_polar("y", start = 0)+
  theme_void()+
  geom_text(aes(y = ypos , label = percent(prop, accuracy = 1)),color="black",size = 5)
```



Out of 809 employees 199 employees which is 25% have left in Travel frequently

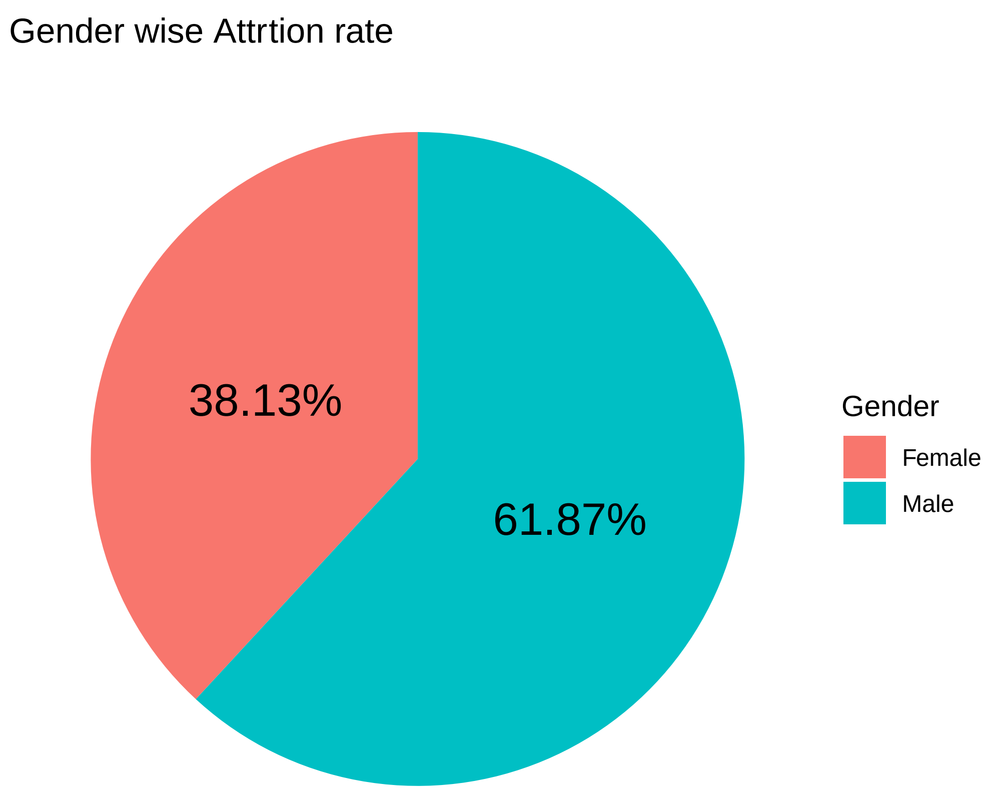
Out of 3051 employees 460 have left in Travel rarely which is 15%

Employee count and attrition count Gender wise

```r
gend_tot_att = merge(gend_tot,gend_att, by = c("Gender","Gender"))

gend_tot_att
```

```
##   Gender total_employees attrition_count attrition_rate
## 1 Female            1729             265          38.13
## 2   Male            2571             430          61.87
```

```r
gend_pie = gend_tot_att

gend_pie = gend_pie %>%
  arrange(-attrition_rate) %>%
  mutate(prop = (attrition_count/sum(attrition_count))) %>%
  mutate(ypos= cumsum(prop) - 0.5 * prop)


ggplot( data = gend_pie , aes(x= "", y = prop, fill = Gender)) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar("y", start = 0) +
  labs(title = 'Gender wise Attrtion rate') +
  theme_void() + # remove background, grid, numeric labels
  geom_text(aes(y = ypos, label = percent(prop, accuracy = 0.01)), color = 'black',size = 6)
```



Most Attrition is from **Male**

Marital status wise employees and attrition rate

```
mar_stat_tot = employee_attrition_data %>%
  select(MaritalStatus) %>%
  count(MaritalStatus, name = "total_employees")

marstatfull = employee_attrition_data %>%
  select(MaritalStatus, Attrition) %>%
  filter(Attrition == "Yes") %>%
  count(MaritalStatus,Attrition, name = "attrition_count") %>%
  summarise(MaritalStatus, attrition_count, attrition_rate = percent(attrition_count/sum(attrition_coun
```

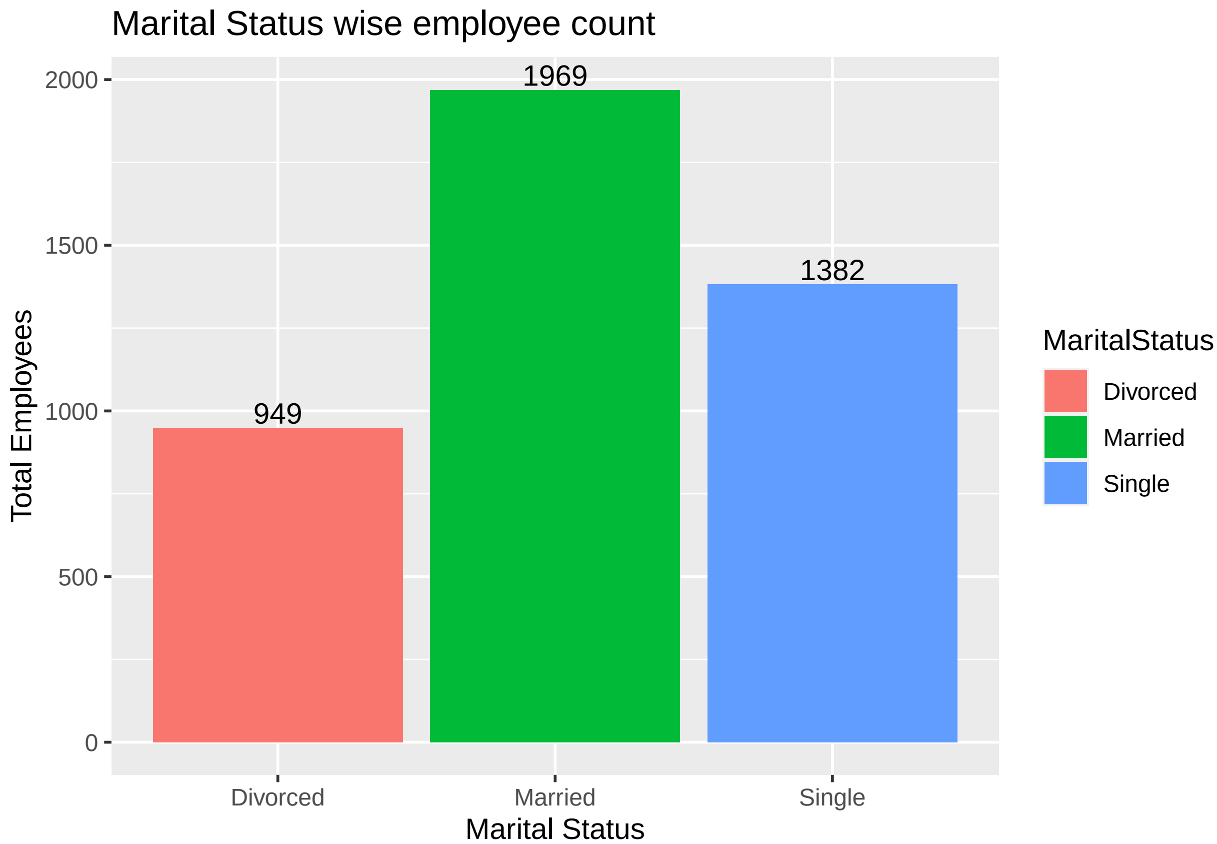
```
## Warning: Returning more (or less) than 1 row per `summarise()` group was deprecated in
## dplyr 1.1.0.
## i Please use `reframe()` instead.
## i When switching from `summarise()` to `reframe()`, remember that `reframe()`
##   always returns an ungrouped data frame and adjust accordingly.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
mar_stat_full = merge(mar_stat_tot,marstatfull, by = "MaritalStatus","MaritalStatus")

mar_stat_full
```

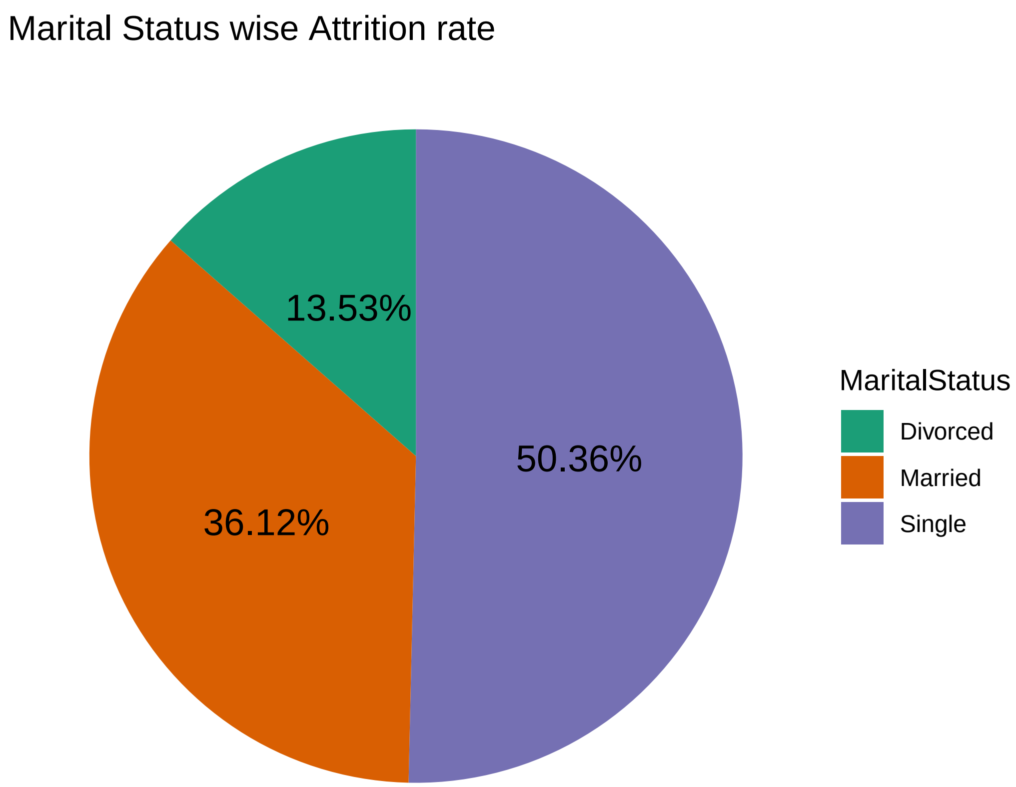
```
##   MaritalStatus total_employees attrition_count attrition_rate
## 1      Divorced             949              94         13.53%
## 2       Married            1969             251         36.12%
## 3        Single            1382             350         50.36%
```

```
ggplot(data = mar_stat_full,aes(x=MaritalStatus , y = total_employees, fill = MaritalStatus)) +
  geom_col(position = "dodge",stat = "identity")+
  labs(title = "Marital Status wise employee count", x = "Marital Status", y = "Total Employees") +
  geom_text(aes(label = total_employees, vjust = -0.2))
```

```
## Warning in geom_col(position = "dodge", stat = "identity"): Ignoring unknown
## parameters: `stat`
```




```
mar_stat_full_pie = mar_stat_full

mar_stat_full_pie = mar_stat_full_pie %>%
  arrange(-attrition_count) %>%
  mutate(prop = (attrition_count/sum(attrition_count))) %>%
  mutate(ypos = (cumsum(prop) - 0.5 * prop))


ggplot(data = mar_stat_full_pie, aes(x="",y = prop , fill = MaritalStatus))+
  geom_bar(stat= "identity", width = 1)+
  coord_polar("y", start = 0)+
  labs(title = "Marital Status wise Attrition rate")+
  theme_void()+
  geom_text(aes(y = ypos  , label = percent(prop,accuracy = 0.01)), color = "Black",size = 5) +
  scale_fill_brewer(palette = "Dark2")
```



Highest attrition are from those who are single

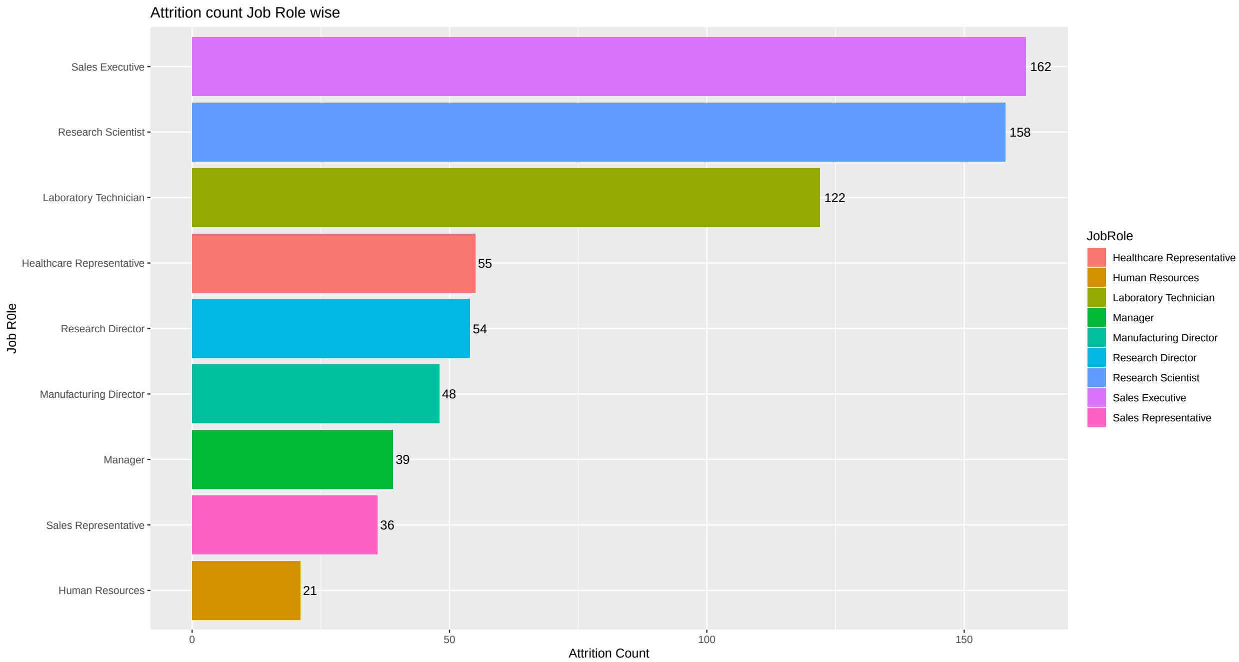
Attrition Job Role wise

```
# total employees

jr_emp = employee_attrition_data %>%
  select(JobRole) %>%
  count(JobRole, name = 'total_employees')

# attrition count

jr_att = employee_attrition_data %>%
  select(JobRole, Attrition) %>%
  filter(Attrition == 'Yes') %>%
  count(JobRole, name = 'attrition_count')

# merged

jr_emp_att = merge(jr_emp, jr_att, by = "JobRole")

jr_emp_att = jr_emp_att %>%
  select(JobRole,total_employees,attrition_count) %>%
```

```r
  mutate(attrition_rate = percent(attrition_count / total_employees)) %>%
  mutate(prop_of_att= percent(attrition_count / sum(attrition_count))) %>%
  arrange(- attrition_count)

jr_emp_att
```

```
##                    JobRole total_employees attrition_count attrition_rate
## 1          Sales Executive             956             162         16.95%
## 2       Research Scientist             859             158         18.39%
## 3    Laboratory Technician             757             122         16.12%
## 4 Healthcare Representative            377              55         14.59%
## 5        Research Director             235              54         22.98%
## 6   Manufacturing Director             422              48         11.37%
## 7                  Manager             299              39         13.04%
## 8     Sales Representative             241              36         14.94%
## 9          Human Resources             154              21         13.64%
##   prop_of_att
## 1      23.31%
## 2      22.73%
## 3      17.55%
## 4       7.91%
## 5       7.77%
## 6       6.91%
## 7       5.61%
## 8       5.18%
## 9       3.02%
```

```r
# bar chart attrition count

ggplot(data = jr_emp_att,aes(x= reorder(JobRole,attrition_count), y = attrition_count, fill = JobRole))
  geom_col(position = "dodge") +
  coord_flip()+
  labs(title = "Attrition count Job Role wise", x = "Job ROle" , y = "Attrition Count")+
  geom_text(aes(label = attrition_count, hjust= -0.2))
```



The most attrition is from Sales Executive, Research Scientist, and Laboratory Technician

Least attrition is from Human Resources

Attrition Job Level wise

```
jl  = employee_attrition_data %>%
  select(Attrition,JobLevel) %>%
  group_by(JobLevel) %>%
  count( Attrition, name = "attrition_count") %>%
  reframe(Attrition, JobLevel,attrition_count ,total_employees = sum(attrition_count))


 jl_emp_att = jl %>%
   select(JobLevel, attrition_count, total_employees,Attrition) %>%
   filter(Attrition == "Yes")

 jl_emp_att = jl_emp_att %>%
  select(JobLevel,attrition_count,total_employees) %>%
  mutate(attrition_rate = (attrition_count/ total_employees)*100) %>%
  mutate(proportion_of_attrition = (attrition_count/ sum(attrition_count))*100)

jl_emp_att
```

```
## # A tibble: 5 x 5
##   JobLevel attrition_count total_employees attrition_rate proportion_of_attrit~1
##      <int>           <int>           <int>          <dbl>                  <dbl>
## 1        1             249            1582           15.7                   35.8
```

```
## 2         2               275             1563            17.6                      39.6
## 3         3                96              641            15.0                      13.8
## 4         4                51              313            16.3                       7.34
## 5         5                24              201            11.9                       3.45
## # i abbreviated name: 1: proportion_of_attrition
```

Highest attrition count is from Job level 2 and 1 , 275 and 249 employees have left

Highest Attrition rate is 17.59% and Highest proportion of attrition is 39.56% from Job level 2

Attrition monthly income wise

```
employee_attrition_data %>%
  select(MonthlyIncome) %>%
  summary(MonthlyIncome)
```

```
##  MonthlyIncome
##  Min.   : 10090
##  1st Qu.: 29260
##  Median : 49360
##  Mean   : 65060
##  3rd Qu.: 83803
##  Max.   :199990
```

```
employee_attrition_data %>%
  select(MonthlyIncome,Attrition) %>%
  filter(Attrition == 'Yes') %>%
  summarise(average_income = mean(MonthlyIncome))
```

```
##   average_income
## 1       61564.22
```

```
employee_attrition_data %>%
  select(MonthlyIncome, Attrition) %>%
  filter(Attrition == "Yes" & MonthlyIncome >= 61564) %>%
  count(name = "attrition_count") %>%
  summarise(monthly_income = ">= 61564", attrition_count)
```

```
##   monthly_income attrition_count
## 1       >= 61564             238
```

```
employee_attrition_data %>%
  select(MonthlyIncome, Attrition) %>%
  filter(Attrition == "Yes" & MonthlyIncome < 61564) %>%
  count(name = "attrition_count") %>%
  summarise(monthly_income = "< 61564", attrition_count)
```

```
##   monthly_income attrition_count
## 1        < 61564             457
```

Minimum salary is 10090

Maximum salary is 199990

Out of 695 employees 457 left who has salary is below 61564

Monthly income bin Attrition count

```
monthly_income_att = full_join(abcdefgh, ijklmn)
```

```
## Joining with ‘by = join_by(monthly_income, attrition_count)‘
```

```
monthly_income_att
```

```
##     monthly_income attrition_count
## 1    10000 - 20000              24
## 2    20001 - 30000             178
## 3    30001 - 40000              62
## 4    40001 - 50000             104
## 5    50001 - 60000              86
## 6    60001 - 70000              63
## 7    70001 - 80000              26
## 8    80001 - 90000              24
## 9   90001 - 100000              15
## 10 100001 - 120000              34
## 11 120001 - 140000              18
## 12 140001 - 160000              12
## 13 160001 - 180000              25
## 14 180001 - 200000              24
```

Most attrition comes from those who got salary from 20k to 30k and then 40k to 50k after that as salary increases attrition decreases

Age wise attrition

```
employee_attrition_data %>%
  select(Age,Attrition) %>%
  summary(Age)
```

```
##       Age         Attrition
##  Min.   :18.00   Length:4300
##  1st Qu.:30.00   Class :character
##  Median :36.00   Mode  :character
##  Mean   :36.93
##  3rd Qu.:43.00
##  Max.   :60.00
```

```r
employee_attrition_data %>%
  select(Age,Attrition) %>%
  filter(Attrition == "Yes") %>%
  summary(Age)
```

```
##       Age         Attrition
##  Min.   :18.00   Length:695
##  1st Qu.:28.00   Class :character
##  Median :32.00   Mode  :character
##  Mean   :33.69
##  3rd Qu.:39.00
##  Max.   :58.00
```

```r
employee_attrition_data %>%
  select(Age , Attrition , EmployeeCount) %>%
  filter(Attrition == "Yes" & Age >= 33) %>%
  summarise(Age = ">=33", sum(EmployeeCount))
```

```
##    Age sum(EmployeeCount)
## 1 >=33                318
```

```r
employee_attrition_data %>%
  select(Age , Attrition , EmployeeCount) %>%
  filter(Attrition == "Yes" & Age < 33) %>%
  summarise(Age = "<33", sum(EmployeeCount))
```

```
##   Age sum(EmployeeCount)
## 1 <33                377
```

Aveage age of employee is 36

```r


ag12 =full_join(ag1,ag2)
```

Average age for the employees that leave is 33

```
## Joining with `by = join_by(Age, attrition_count)`
```

```r
ag34 =full_join(ag3,ag4)
```

```
## Joining with `by = join_by(Age, attrition_count)`
```

```r
ag56 =full_join(ag5,ag6)
```

```
## Joining with `by = join_by(Age, attrition_count)`
```

```r
ag1234 = full_join(ag12, ag34)
```

```
## Joining with `by = join_by(Age, attrition_count)`
```

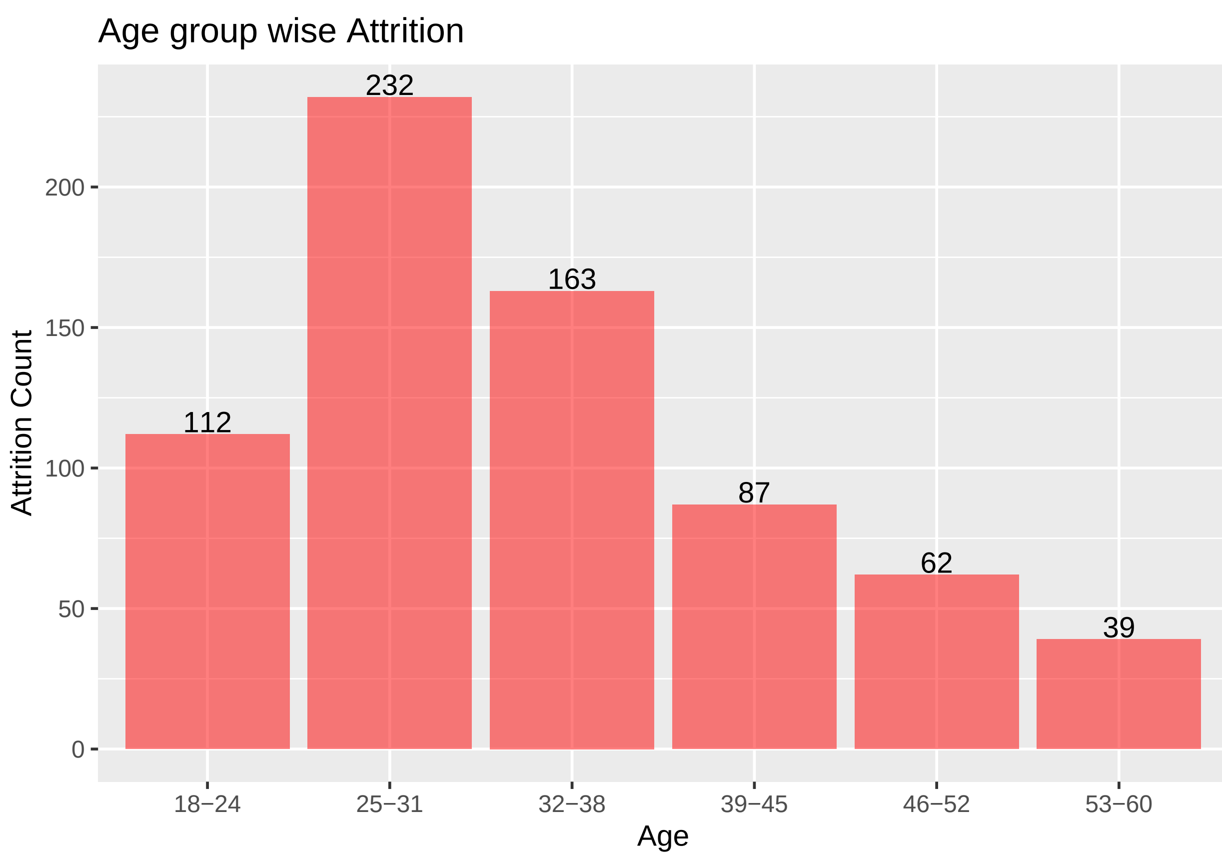
```r
age_att = full_join(ag1234, ag56)
```

```
## Joining with `by = join_by(Age, attrition_count)`
```

```r
age_att
```

```
##     Age attrition_count
## 1 18-24             112
## 2 25-31             232
## 3 32-38             163
## 4 39-45              87
## 5 46-52              62
## 6 53-60              39
```

```r
ggplot(data = age_att,aes(x= Age, y = attrition_count)) +
  geom_col(fill = alpha('red',0.5))+
  labs(title = "Age group wise Attrition", x = "Age", y = "Attrition Count")+
  geom_text(aes(label = attrition_count, vjust = -0.1))
```



The Highest Attrition is from the age group 25-31

After that the attrition keeps decreasing

Worklife balance wise Attrition

```
employee_attrition_data %>%
  select(WorkLifeBalance,Attrition) %>%
  filter(Attrition == "Yes") %>%
  count(WorkLifeBalance,name = "attrition-count")
```

```
##   WorkLifeBalance attrition-count
## 1               1              73
## 2               2             167
## 3               3             375
## 4               4              80
```

The most attrition is from who have rated 3 and then 2 for worklifebalance

Salary hike wise attrition

```
employee_attrition_data %>%
  select(PercentSalaryHike) %>%
  summary()
```

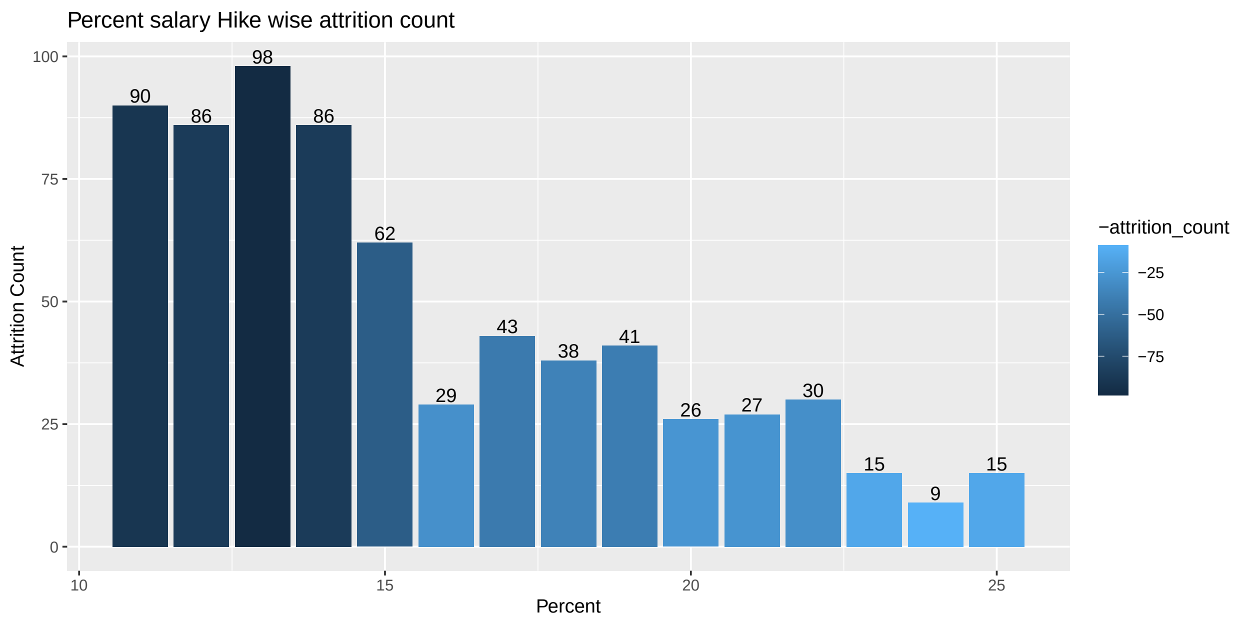
```
##  PercentSalaryHike
##  Min.   :11.00
##  1st Qu.:12.00
##  Median :14.00
##  Mean   :15.21
##  3rd Qu.:18.00
##  Max.   :25.00
```

```
salary_hike =employee_attrition_data %>%
  select(PercentSalaryHike,Attrition) %>%
  filter(Attrition == "Yes") %>%
  group_by(PercentSalaryHike) %>%
  count(name = "attrition_count")

salary_hike
```

```
## # A tibble: 15 x 2
## # Groups:   PercentSalaryHike [15]
##    PercentSalaryHike attrition_count
##                <int>           <int>
##  1                11              90
##  2                12              86
```

```
## 3                13             98
## 4                14             86
## 5                15             62
## 6                16             29
## 7                17             43
## 8                18             38
## 9                19             41
## 10               20             26
## 11               21             27
## 12               22             30
## 13               23             15
## 14               24              9
## 15               25             15
```

```
# bar
salary_hike_pie = salary_hike

ggplot(salary_hike_pie,aes(x= PercentSalaryHike, y = attrition_count, fill = -attrition_count))+
  geom_col()+
  labs(title = "Percent salary Hike wise attrition count", x= "Percent", y ="Attrition Count")+
  geom_text(aes(label = attrition_count, vjust= -0.2))
```

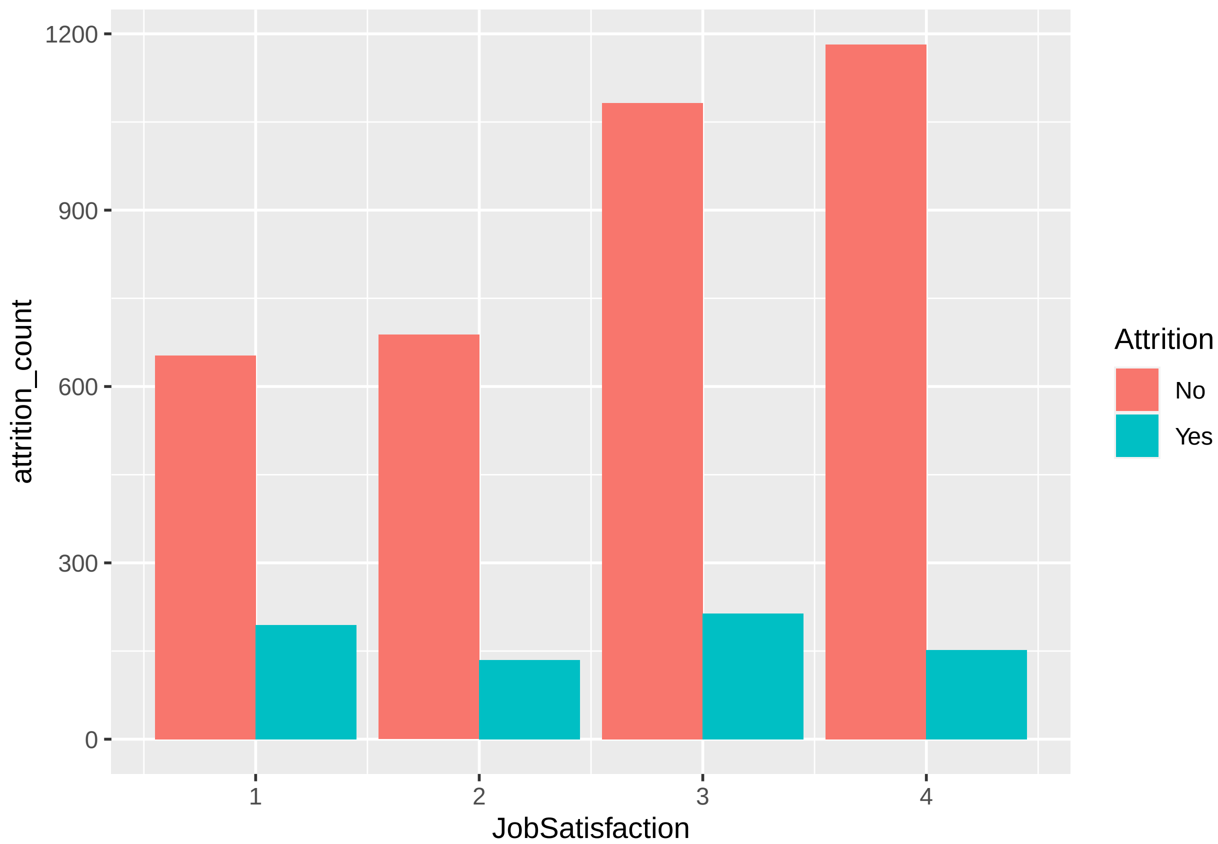


There is high attrition from 10 to 15 % salary hike as percent salary hike increases the attrition decreases

Job satisfaction wise attrition

```
js_att = employee_attrition_data %>%
  select(JobSatisfaction,Attrition) %>%
  group_by(JobSatisfaction) %>%
```

```
  count(Attrition, name = 'attrition_count') %>%
  reframe(JobSatisfaction, Attrition , attrition_count, total=sum(attrition_count), attrition_rate = pe

# grouped bar chart

ggplot(js_att, aes(fill=Attrition, y=attrition_count, x=JobSatisfaction)) +
    geom_bar(position="dodge", stat="identity")
```




```
js_att = js_att %>%
  filter(Attrition == 'Yes') %>%
  mutate(prop_att = percent(attrition_count/sum(attrition_count)))

js_att
```

```
## # A tibble: 4 x 6
##   JobSatisfaction Attrition attrition_count total attrition_rate prop_att
##             <int> <chr>               <int> <int> <chr>          <chr>
## 1               1 Yes                   194   847 23%            27.9%
## 2               2 Yes                   135   823 16%            19.4%
## 3               3 Yes                   214  1296 17%            30.8%
## 4               4 Yes                   152  1334 11%            21.9%
```

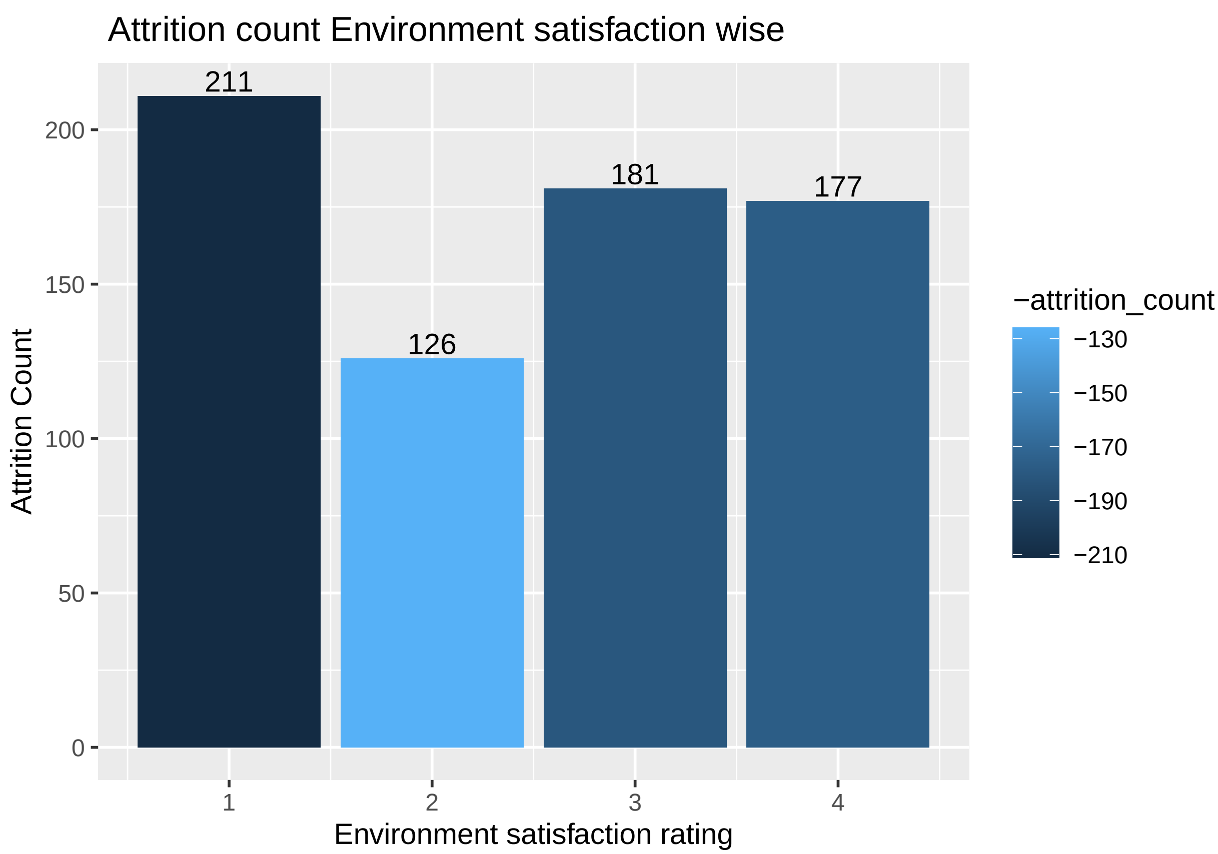
The Most attrition comes from those who rated 3 and then 1

Environment Satisfaction wise Attrition

```
env_sat_att = employee_attrition_data %>%
  select(EnvironmentSatisfaction, Attrition) %>%
  filter(Attrition == "Yes") %>%
  count(EnvironmentSatisfaction, name = 'attrition_count')


env_sat_att
```

```
##   EnvironmentSatisfaction attrition_count
## 1                       1             211
## 2                       2             126
## 3                       3             181
## 4                       4             177
```

```
# bar

env_sat_bar = env_sat_att

ggplot(env_sat_bar, aes(fill= -attrition_count, y=attrition_count, x=EnvironmentSatisfaction)) +
    geom_bar(stat= "identity")+
  labs(title = " Attrition count Environment satisfaction wise", x= 'Environment satisfaction rating', y
  geom_text(aes(label = attrition_count, vjust = -0.2))
```



Summary of Analysis

Overall attrition - The attrition rate is 16.16%

Department - Most attrition is from Research & Development Department and then followed by Sales

Education Field - The Highest attrition is from Life Sciences Education Field and then Medical Education Field

Business Travel - Out of 809 employees 199 employees which is 25% have left in Travel frequently, Out of 3051 employees 460 have left in Travel rarely which is 15%

Gender - Male have most attrition

Marital Status - Single have most attrition

Job Role - The most attrition is from Sales Executive, Research Scientist, and then Laboratory Technician

Job Level - Highest attrition count is from Job level 2 and then 1

Monthly Income - The lower the salary the higher the attrition, as salary increases attrition decreases

Age - Most attrition comes from the age group 25-31 as age increases attrition decreases

Work Life Balance - The most attrition is from who have rated 3 and then 2 for worklifebalance

Percent Salary Hike - From 10 - 15% salary hike has most attritions as salary hike increases above 15% attrition decreases

Job Satisfaction - Most attrition comes from those who rated 3 and then 1

Environment Satisfaction - Highest attrition is from those who rated 1 for Environment Satisfaction

Conclusion

- **Department** Employees from **Research & Development** department are more likely to quit than other departments
- **Business Travel** Employees that **Travel Frequently** are more likely to quit
- **Education Field** Employees who have Life science and then Medical education field have high attrition

- **Monthly Income** Employees who have low income tend to leave for other companies that pay better salary

- **Age** Young employees tend to leave the jobs for better opportunities

- **Environment Satisfaction** Those who are not satisfied with their work environment leave the company

- **Salary Hike** Those who get less salary hike tend to leave the company for more salary