

# Relatório de Análise de Rotatividade e Desempenho

23 de junho de 2025

## Contents

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Análise Exploratória</b>	<b>1</b>
2.1	Distribuição da Idade . . . . .	1
2.2	Distribuição do Tempo de Serviço . . . . .	2
2.3	Relação entre Tempo de Serviço e Rotatividade . . . . .	3
2.4	Relação entre Avaliação de Desempenho e Rotatividade . . . . .	4
2.5	Boxplot da Distribuição Salarial por Departamento . . . . .	5
2.6	Taxa de Rotatividade por Departamento . . . . .	6
2.7	Relação entre Anos na Empresa e Salário (com Rotatividade) . . . . .	7
<b>3</b>	<b>Correlação das variáveis com a Rotatividade (Attrition) - Gráfico Melhorado</b>	<b>8</b>
3.0.1	Análise da Correlação . . . . .	10
<b>4</b>	<b>Conclusão</b>	<b>11</b>

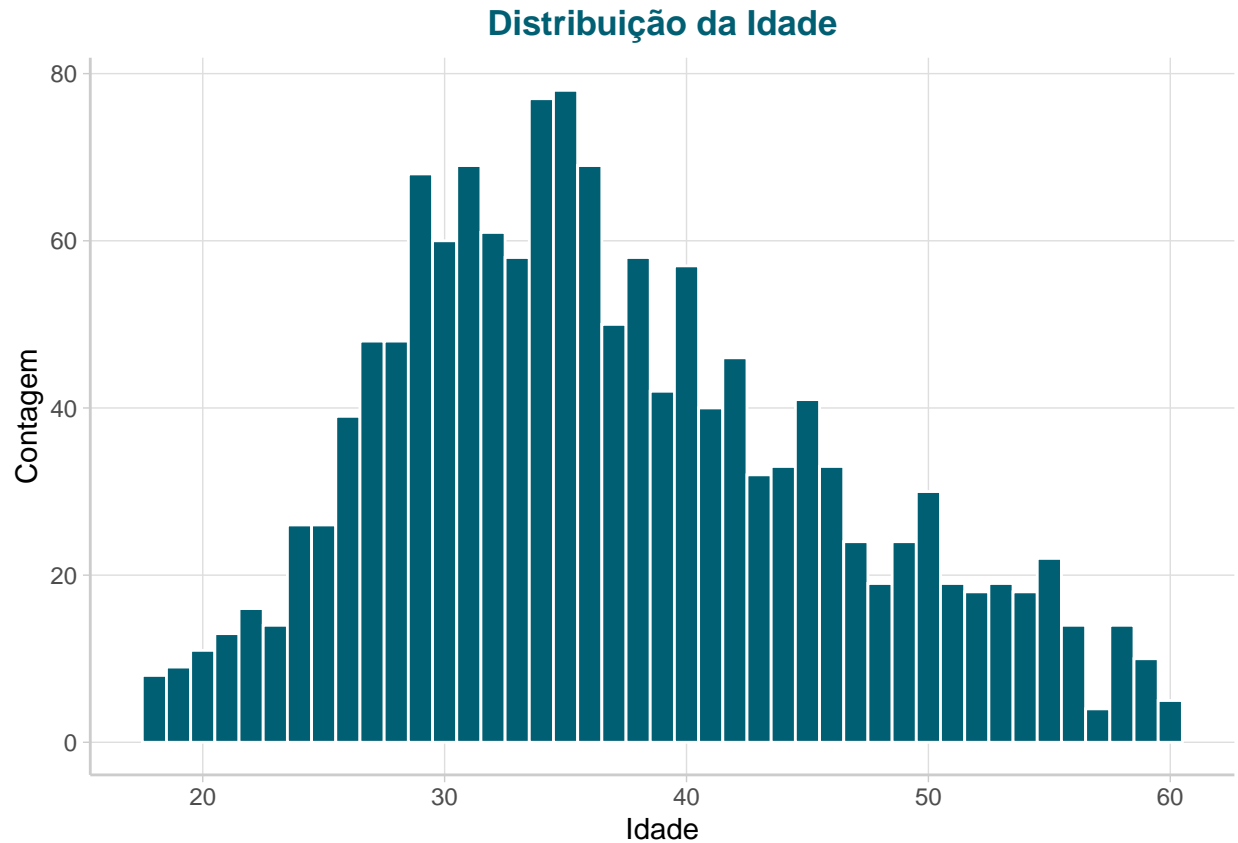
## 1 Introdução

Este relatório tem como objetivo analisar os dados do dataset “ibm\_hr\_data.csv” e identificar padrões e tendências que possam ajudar a entender a rotatividade e o desempenho dos funcionários da IBM.

## 2 Análise Exploratória

### 2.1 Distribuição da Idade

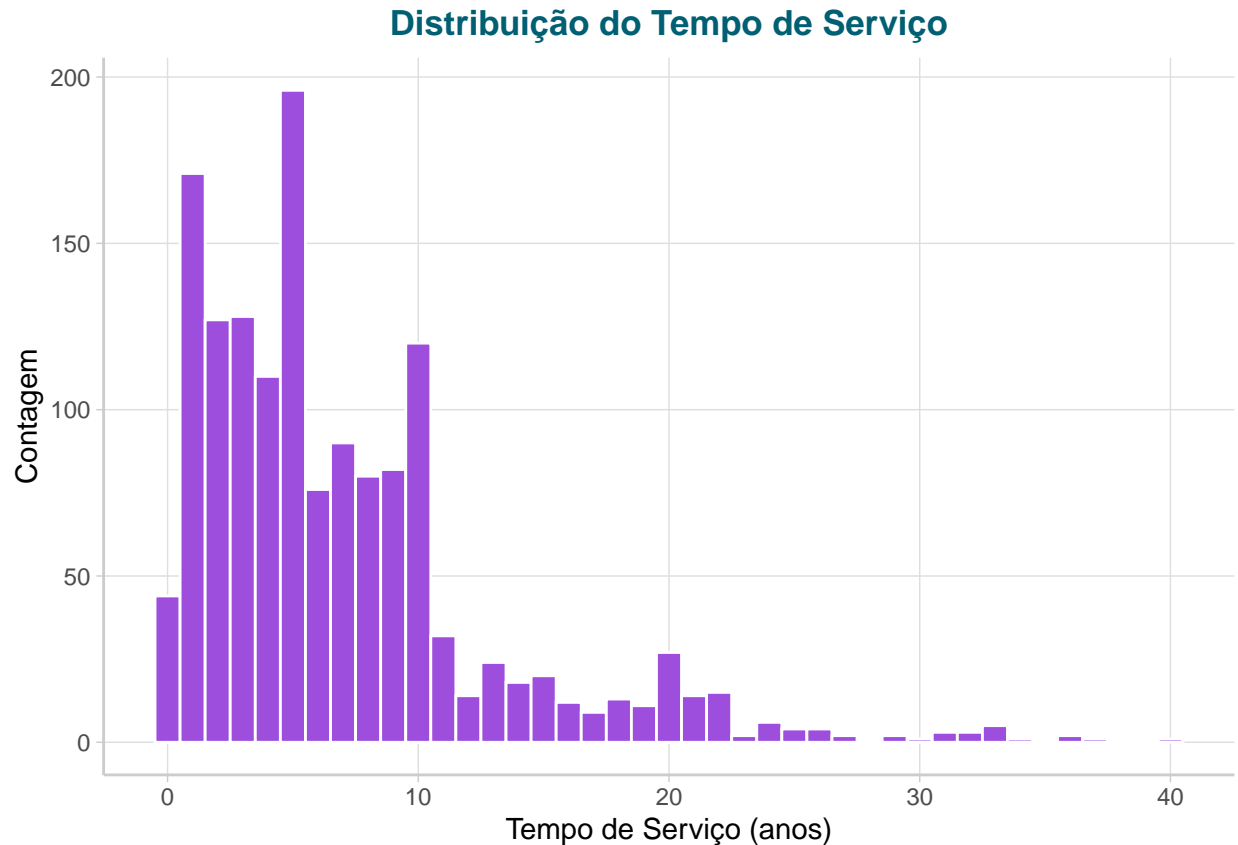
```
# Distribuição da idade
ggplot(data, aes(x = Age)) +
  geom_histogram(binwidth = 1, fill = color_palette$primary, color = "white") +
  labs(title = "Distribuição da Idade", x = "Idade", y = "Contagem")
```



O histograma da idade revela que a força de trabalho da IBM neste dataset é maioritariamente composta por profissionais em meio de carreira, com uma concentração significativa de funcionários entre os 30 e os 40 anos. A representação de trabalhadores mais jovens (abaixo dos 25) e mais velhos (acima dos 55) é consideravelmente menor.

## 2.2 Distribuição do Tempo de Serviço

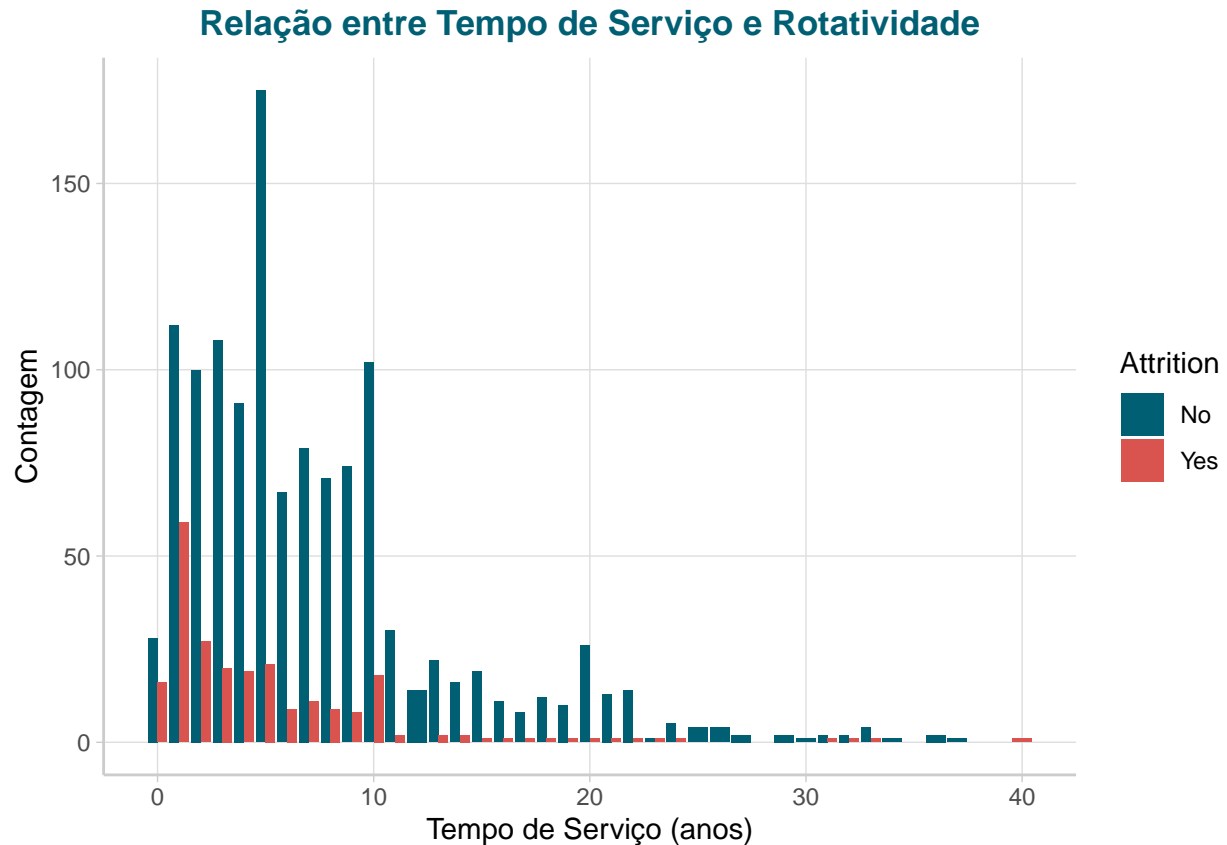
```
# Distribuição do tempo de serviço
ggplot(data, aes(x = YearsAtCompany)) +
  geom_histogram(binwidth = 1, fill = color_palette$accent, color = "white") +
  labs(title = "Distribuição do Tempo de Serviço", x = "Tempo de Serviço (anos)", y = "Contagem")
```



A análise do tempo de serviço na empresa mostra que uma grande parte dos funcionários tem até 10 anos de casa, com a maior concentração a ocorrer nos primeiros 5 anos. Isto sugere um fluxo constante de novos talentos, mas também pode indicar que os primeiros anos são críticos para a retenção a longo prazo. A cauda longa do gráfico aponta para a existência de um núcleo de funcionários com uma longa carreira na empresa.

## 2.3 Relação entre Tempo de Serviço e Rotatividade

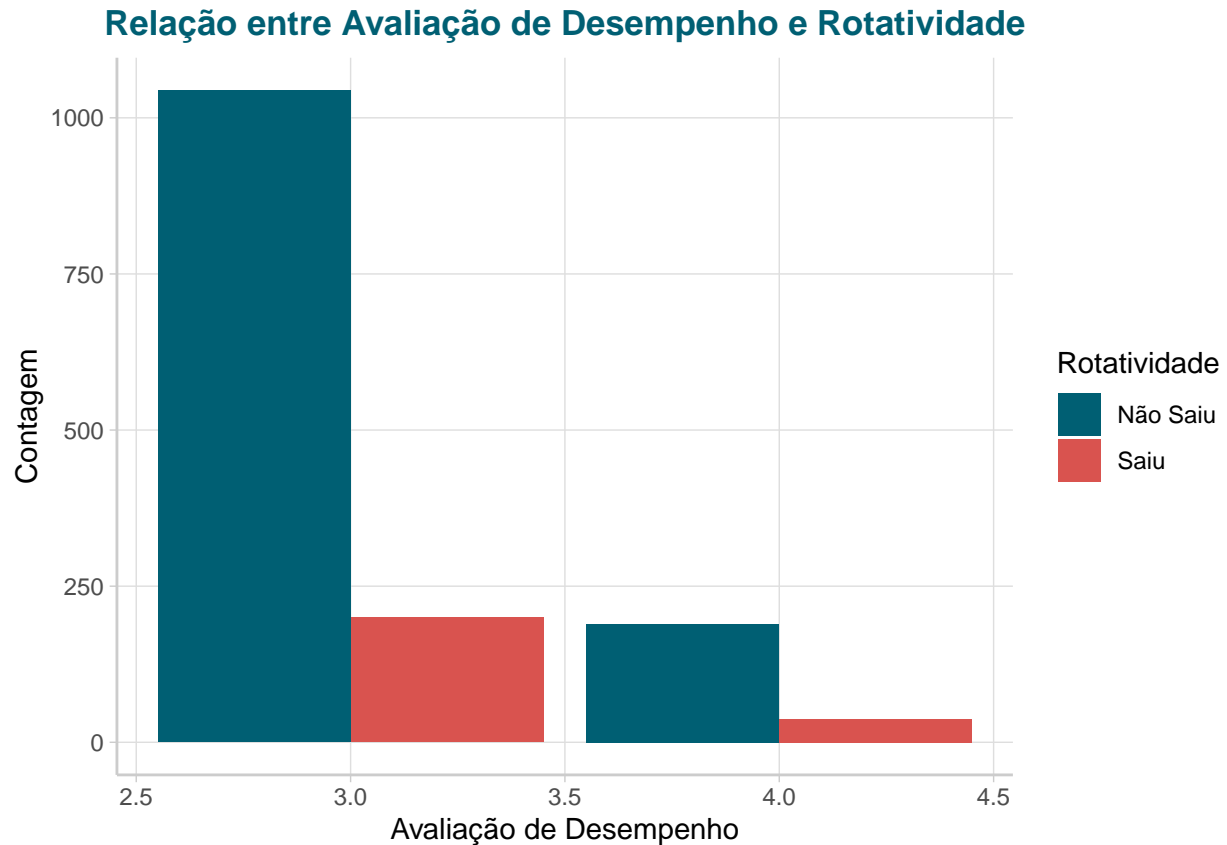
```
# Relação entre tempo de serviço e rotatividade
ggplot(data, aes(x = YearsAtCompany, fill = Attrition)) +
  geom_bar(position = "dodge") +
  labs(title = "Relação entre Tempo de Serviço e Rotatividade", x = "Tempo de Serviço (anos)", y = "Con")
  scale_fill_manual(values = c("No" = color_palette$primary, "Yes" = color_palette$secondary))
```



Este gráfico evidencia uma tendência clara: a rotatividade é mais acentuada entre os funcionários com menos tempo de serviço. O risco de um funcionário sair é maior nos primeiros anos na empresa, diminuindo progressivamente à medida que o tempo de serviço aumenta. Este padrão sugere que a integração inicial e o desenvolvimento de carreira nos primeiros anos são cruciais para a retenção.

## 2.4 Relação entre Avaliação de Desempenho e Rotatividade

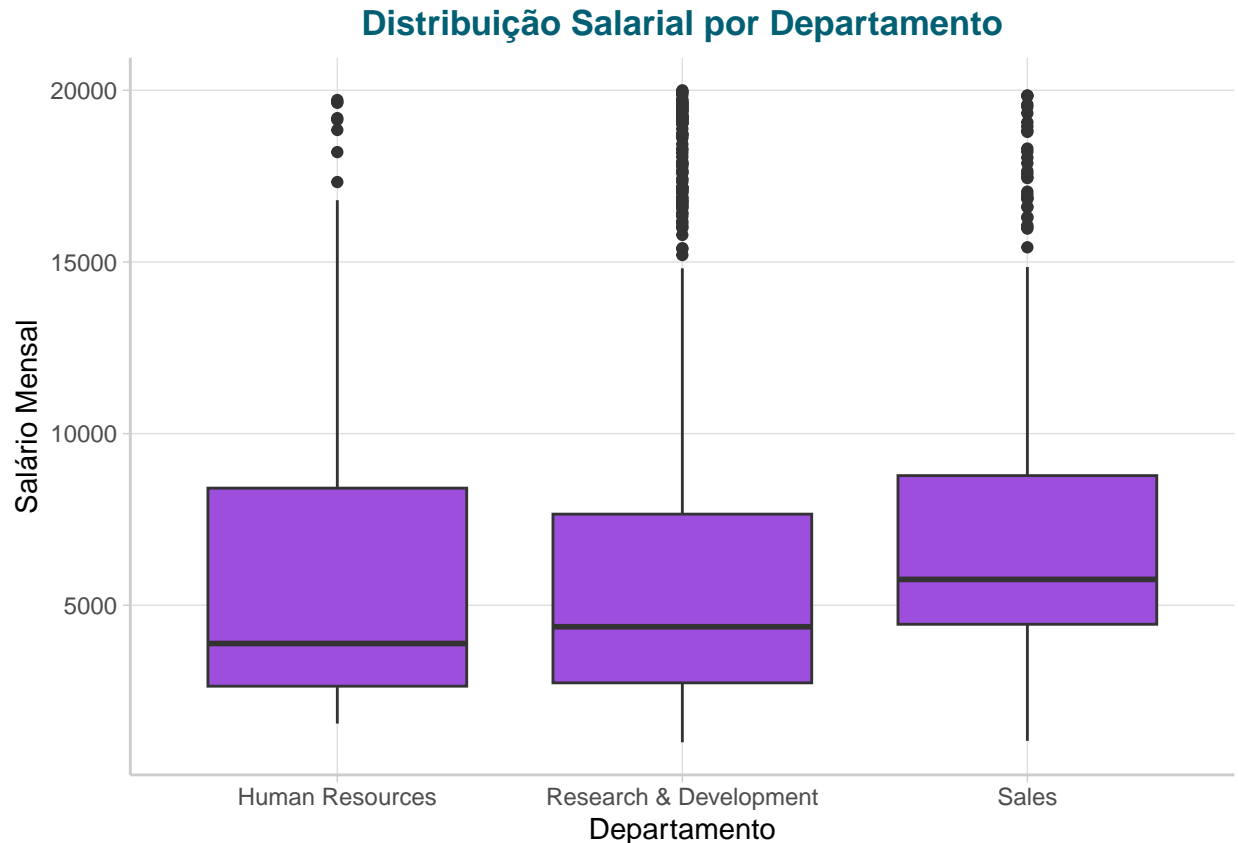
```
ggplot(data, aes(x = PerformanceRating, fill = Attrition)) +
  geom_bar(position = "dodge") +
  labs(title = "Relação entre Avaliação de Desempenho e Rotatividade",
       x = "Avaliação de Desempenho",
       y = "Contagem",
       fill = "Rotatividade") +
  scale_fill_manual(values = c("No" = color_palette$primary, "Yes" = color_palette$secondary),
                   labels = c("Não Saiu", "Saiu"))
```



A relação entre a avaliação de desempenho e a rotatividade mostra que a grande maioria dos funcionários, quer os que saem quer os que ficam, possui uma avaliação de desempenho de nível 3. Embora não haja uma diferença drástica na *proporção* de rotatividade entre os diferentes níveis de avaliação, o volume absoluto de saídas vem do grupo de desempenho médio. Isto pode indicar que o sistema de avaliação pode não estar a diferenciar eficazmente o desempenho ou que outros fatores, para além do desempenho medido, estão a impulsionar a rotatividade.

## 2.5 Boxplot da Distribuição Salarial por Departamento

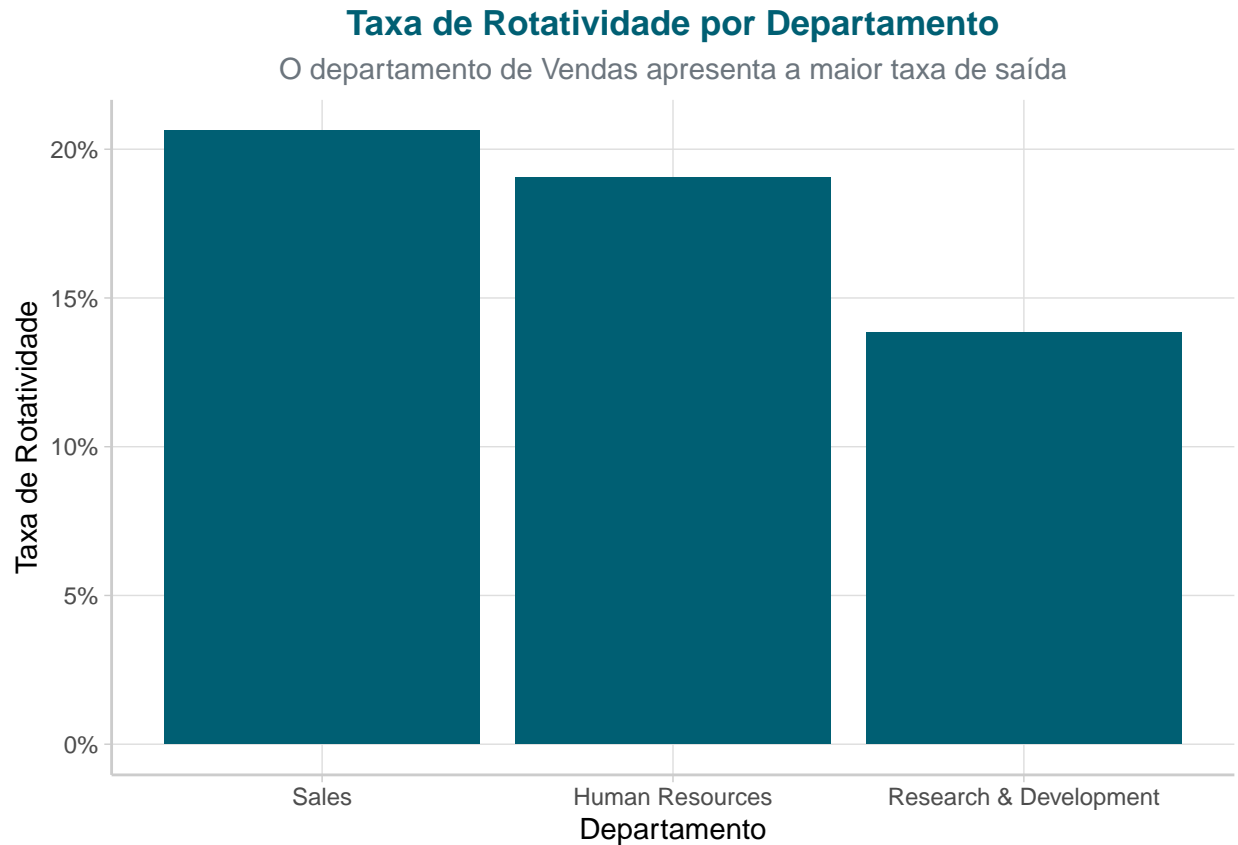
```
# Boxplot da distribuição salarial por departamento
ggplot(data, aes(x = Department, y = MonthlyIncome)) +
  geom_boxplot(fill = color_palette$accent) +
  labs(title = "Distribuição Salarial por Departamento", x = "Departamento", y = "Salário Mensal")
```



O boxplot revela diferenças significativas na estrutura salarial entre os departamentos. O departamento de Vendas (Sales) apresenta a maior dispersão salarial, com alguns funcionários a atingirem rendimentos muito elevados (outliers), o que é típico de funções com componentes variáveis baseadas em comissões. O departamento de Investigação e Desenvolvimento (Research & Development) mostra uma mediana salarial competitiva. Em contraste, o departamento de Recursos Humanos (Human Resources) apresenta a faixa salarial mais baixa e mais concentrada.

## 2.6 Taxa de Rotatividade por Departamento

```
# Taxa de rotatividade por departamento
data %>%
  group_by(Department) %>%
  summarize(turnover_rate = mean(Attrition == "Yes", na.rm = TRUE), .groups = 'drop') %>%
  ggplot(aes(x = reorder(Department, -turnover_rate), y = turnover_rate)) +
  geom_bar(stat = "identity", fill = color_palette$primary) +
  scale_y_continuous(labels = scales::percent) +
  labs(
    title = "Taxa de Rotatividade por Departamento",
    subtitle = "O departamento de Vendas apresenta a maior taxa de saída",
    x = "Departamento",
    y = "Taxa de Rotatividade"
  )
```



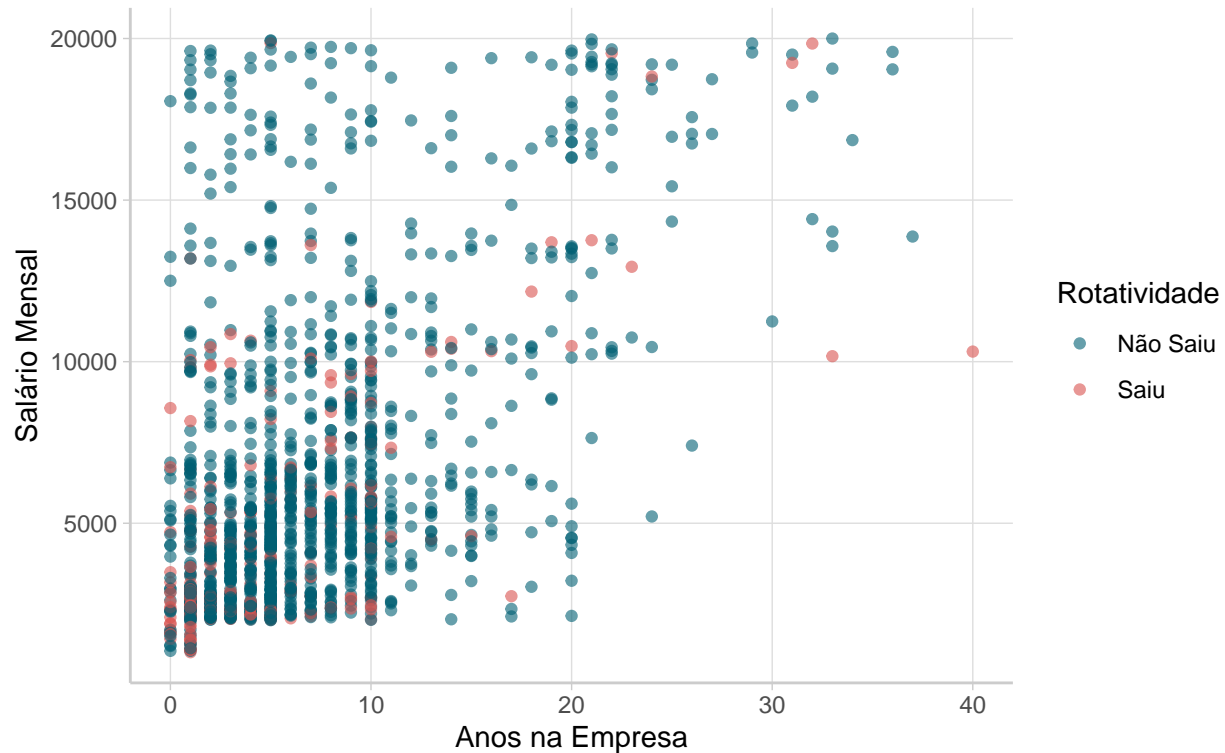
Ao analisar a taxa de rotatividade por departamento, observa-se que o departamento de Vendas (**Sales**) sofre a taxa mais elevada, seguido pelos Recursos Humanos (**Human Resources**). O departamento de Investição e Desenvolvimento (**Research & Development**) apresenta a menor taxa de rotatividade. Estes dados, combinados com a análise salarial, sugerem que a natureza de alta pressão e a estrutura de remuneração variável das Vendas podem contribuir para uma maior saída de funcionários.

## 2.7 Relação entre Anos na Empresa e Salário (com Rotatividade)

```
# Relação entre anos na empresa e salário (com rotatividade)
ggplot(data, aes(x = YearsAtCompany, y = MonthlyIncome, color = Attrition)) +
  geom_point(alpha = 0.6) +
  scale_color_manual(values = c("No" = color_palette$primary, "Yes" = color_palette$secondary),
    labels = c("Não Saiu", "Saiu")) +
  labs(
    title = "Relação entre Anos na Empresa e Salário",
    subtitle = "A rotatividade concentra-se em salários mais baixos e menor tempo de casa",
    x = "Anos na Empresa",
    y = "Salário Mensal",
    color = "Rotatividade"
  )
```

## Relação entre Anos na Empresa e Salário

A rotatividade concentra-se em salários mais baixos e menor tempo de casa



Este gráfico de dispersão confirma a intuição de que o salário mensal tende a aumentar com o tempo de serviço na empresa. É particularmente revelador que os pontos que representam os funcionários que saíram (*Attrition* = “Yes”, em vermelho) se concentram predominantemente na zona de salários mais baixos e com menos anos de empresa. Isto reforça a conclusão de que a compensação e a falta de progressão na carreira nos primeiros anos são fatores críticos que impulsionam a rotatividade.

### 3 Correlação das variáveis com a Rotatividade (*Attrition*) - Gráfico Melhorado

```
# Converter variáveis categóricas para numéricas para correlação
data_corr <- data %>%
  mutate(
    AttritionNum = ifelse(Attrition == "Yes", 1, 0),
    OverTimeNum = ifelse(OverTime == "Yes", 1, 0)
  )

# Selecionar variáveis numéricas, incluindo a nova OverTimeNum
num_vars <- data_corr %>%
  select_if(is.numeric) %>%
  select(-EmployeeCount, -StandardHours, -EmployeeNumber, -AttritionNum) # Remover colunas irrelevantes

# Calcular a correlação
cor_results <- cor(num_vars, data_corr$AttritionNum, use = "pairwise.complete.obs")
```



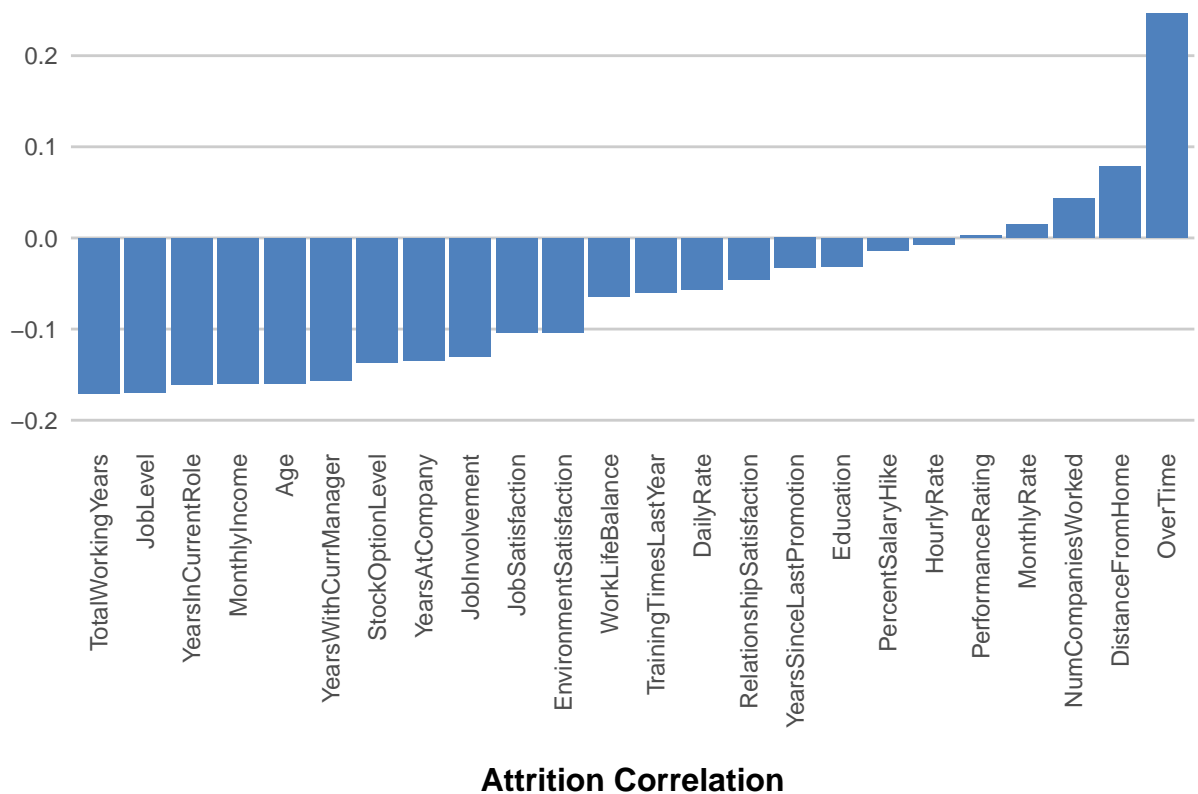
```

# Criar e limpar o dataframe
cor_df <- data.frame(
  Variable = rownames(cor_results),
  Correlation = cor_results[, 1]
) %>%
  # Renomear 'OverTimeNum' para 'OverTime' para melhor visualização
  mutate(Variable = ifelse(Variable == "OverTimeNum", "OverTime", Variable)) %>%
  filter(Variable != "AttritionNum") %>%
  arrange(Correlation)

# Garantir que a variável é um fator com a ordem correta para o gráfico
cor_df$Variable <- factor(cor_df$Variable, levels = cor_df$Variable)

# Gerar o gráfico de barras idêntico à imagem
ggplot(cor_df, aes(x = Variable, y = Correlation)) +
  geom_bar(stat = "identity", fill = "#4F81BD") +
  scale_y_continuous(breaks = seq(-0.2, 0.3, by = 0.1), limits = c(-0.2, 0.25)) +
  labs(x = "Attrition Correlation", y = "") +
  theme_bw() +
  theme(
    axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1, size = 9),
    axis.title.x = element_text(size = 12, face = "bold", margin = margin(t = 15, b = 10)),
    panel.grid.major.y = element_line(color = "grey80"),
    panel.grid.major.x = element_blank(),
    panel.grid.minor.y = element_blank(),
    panel.border = element_blank(),
    axis.ticks = element_blank()
  )

```



### 3.0.1 Análise da Correlação

O gráfico de correlação acima destaca os fatores mais significativamente associados à rotatividade (**Attrition**) dos funcionários. As principais conclusões são:

#### Fatores que Aumentam a Rotatividade (Correlação Positiva):

- **Horas Extra (OverTime):** Este é, de longe, o fator com a maior correlação positiva. Funcionários que trabalham horas extraordinárias têm uma probabilidade muito maior de deixar a empresa.
- **Distância de Casa (DistanceFromHome):** Uma maior distância entre a casa do funcionário e o local de trabalho está associada a uma maior taxa de rotatividade.
- **Número de Empresas Anteriores (NumCompaniesWorked):** Funcionários com um histórico de trabalho em várias empresas têm uma ligeira tendência a sair.

#### Fatores que Diminuem a Rotatividade (Correlação Negativa):

- **Estabilidade e Experiência:** Fatores como **TotalWorkingYears** (total de anos de trabalho), **JobLevel** (nível do cargo), **YearsInCurrentRole** (anos no cargo atual), **MonthlyIncome** (salário) e **Age** (idade) apresentam as correlações negativas mais fortes. Isto indica que funcionários mais experientes, mais velhos, com salários mais altos e em cargos de maior nível são significativamente mais propensos a permanecer na empresa.
- **Satisfação e Envolvimento:** Variáveis como **JobInvolvement** (envolvimento no trabalho), **JobSatisfaction** (satisfação com o trabalho) e **EnvironmentSatisfaction** (satisfação com o ambiente de trabalho) também mostram uma correlação negativa, o que sugere que funcionários mais satisfeitos e envolvidos tendem a ficar.

- **Equilíbrio (WorkLifeBalance):** Um melhor equilíbrio entre a vida profissional e pessoal está associado a uma menor probabilidade de saída.

Em resumo, os esforços para reter talentos devem focar-se em gerir a carga de trabalho para evitar horas extras excessivas, e em promover a satisfação, o desenvolvimento de carreira e a estabilidade para os funcionários.

## 4 Conclusão

A presente análise dos dados de Recursos Humanos da IBM permitiu extrair conclusões detalhadas sobre os fatores que impulsionam a rotatividade dos funcionários. O perfil demográfico da empresa revela uma força de trabalho madura, concentrada na faixa dos 30-40 anos, com um fluxo constante de novos talentos que tendem a ter um tempo de serviço inferior a 10 anos.

A análise de correlação identificou de forma inequívoca que a realização de **horas extraordinárias (OverTime)** é o principal fator associado à decisão de um funcionário sair da empresa. Outros fatores de risco incluem a distância entre a casa e o trabalho e um historial de passagem por múltiplas empresas.

Em contrapartida, os fatores de retenção mais fortes estão ligados à **estabilidade, experiência e compensação**. Funcionários com mais anos de carreira, em cargos de nível superior, com salários mais elevados e há mais tempo na empresa e no mesmo cargo demonstraram uma probabilidade significativamente menor de sair. A satisfação com o trabalho, o ambiente e o equilíbrio entre a vida profissional e pessoal também se revelaram cruciais para a retenção.

Observou-se ainda uma dinâmica departamental clara, com o departamento de **Vendas a registar a taxa de rotatividade mais elevada**, enquanto o de Investigação e Desenvolvimento se mostrou o mais estável.

Em suma, os resultados sugerem que as estratégias de retenção da IBM devem focar-se em três pilares centrais: **1) gerir a carga de trabalho e a cultura de horas extraordinárias, 2) garantir uma progressão de carreira e compensação justas**, especialmente nos primeiros anos de serviço, e **3) promover um ambiente de trabalho positivo e equilibrado**.