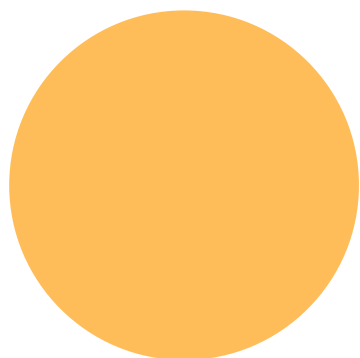




GABRIEL DELUCCA

CONSULTORIA EM PROCESSOS GERENCIAIS



ANÁLISE DE DADOS RH



(31) 99175-9049



gabrieldbarros@gmail.com

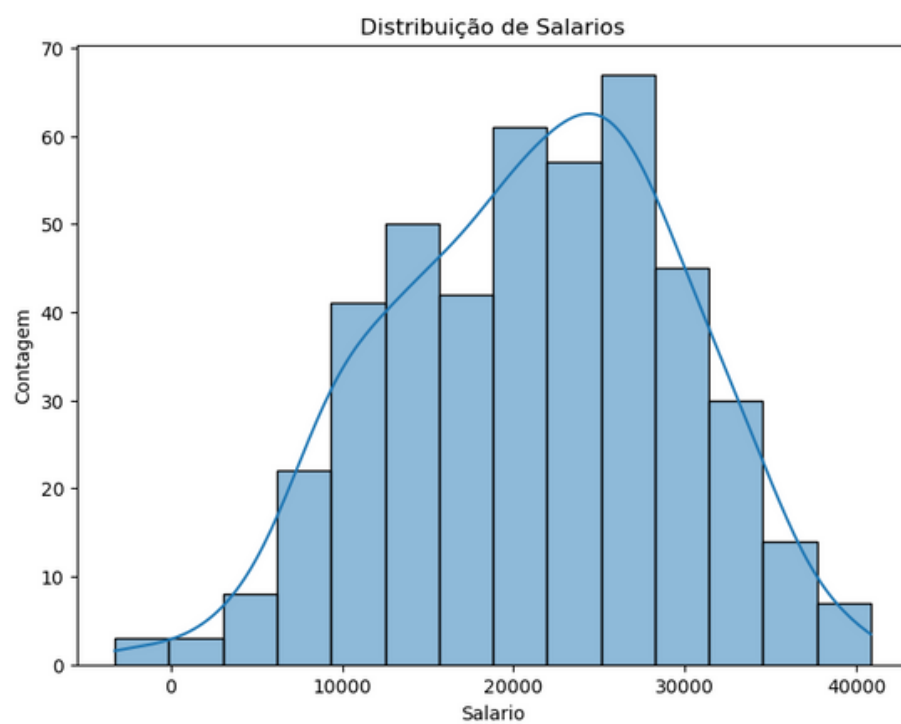
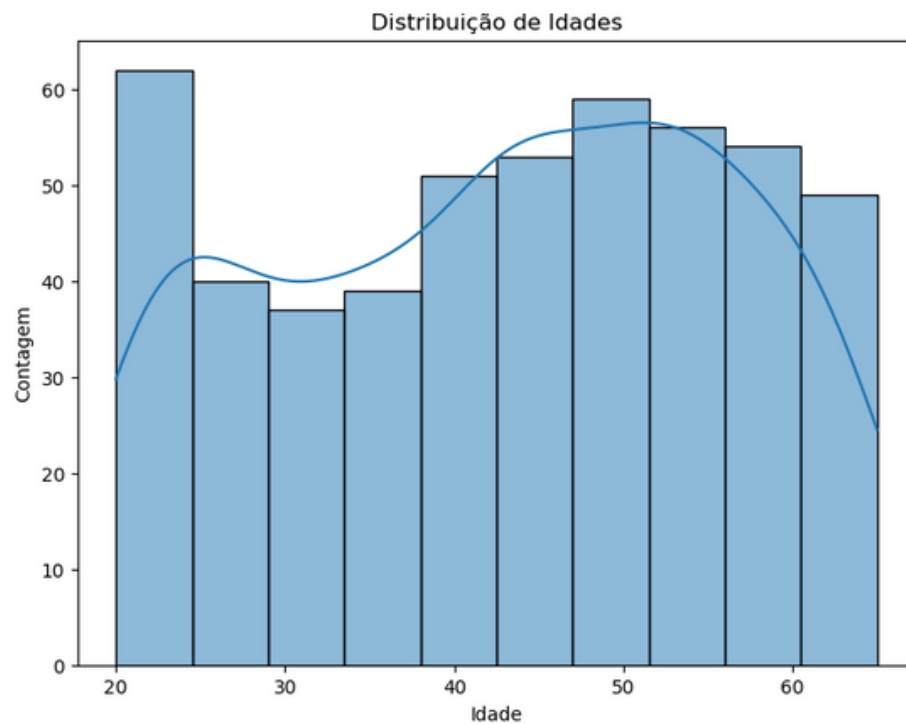


GABRIEL DELUCCA

CONSULTORIA EM PROCESSOS GERENCIAIS

(31) 99175-9049 gabrieldbarros@gmail.com

Visualizando a Distribuição de Variáveis Quantitativas





GABRIEL DELUCCA

CONSULTORIA EM PROCESSOS GERENCIAIS

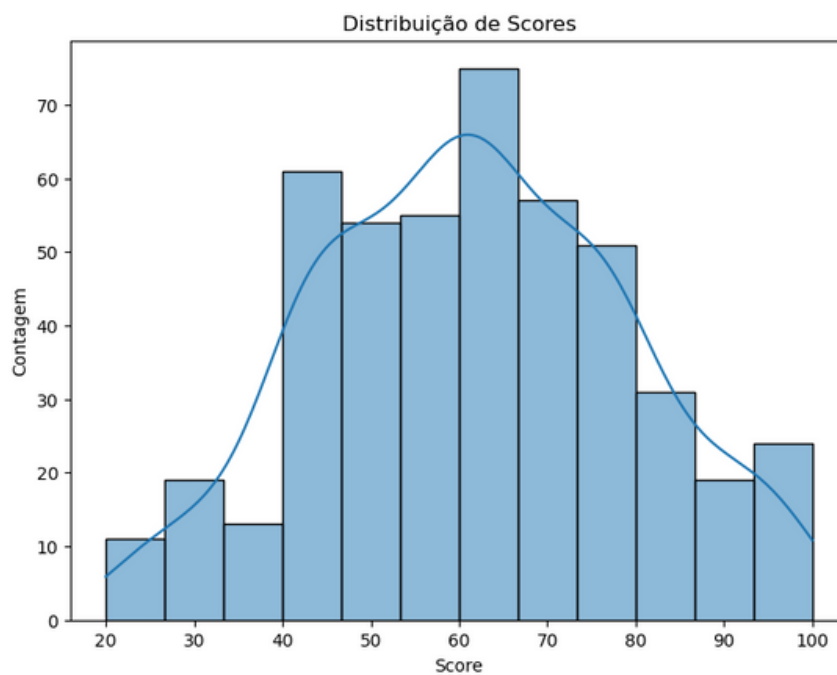


(31) 99175-9049

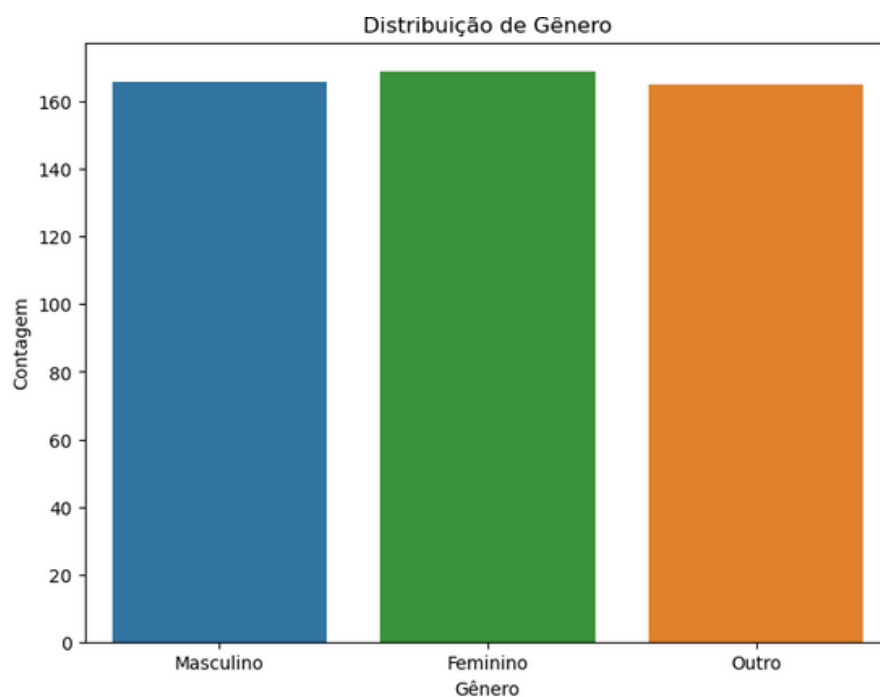


gabrielbarros@gmail.com

Visualizando a Distribuição de Variáveis Quantitativas



Visualizando a Distribuição de Variáveis Qualitativas/ Categóricas





GABRIEL DELUCCA

CONSULTORIA EM PROCESSOS GERENCIAIS

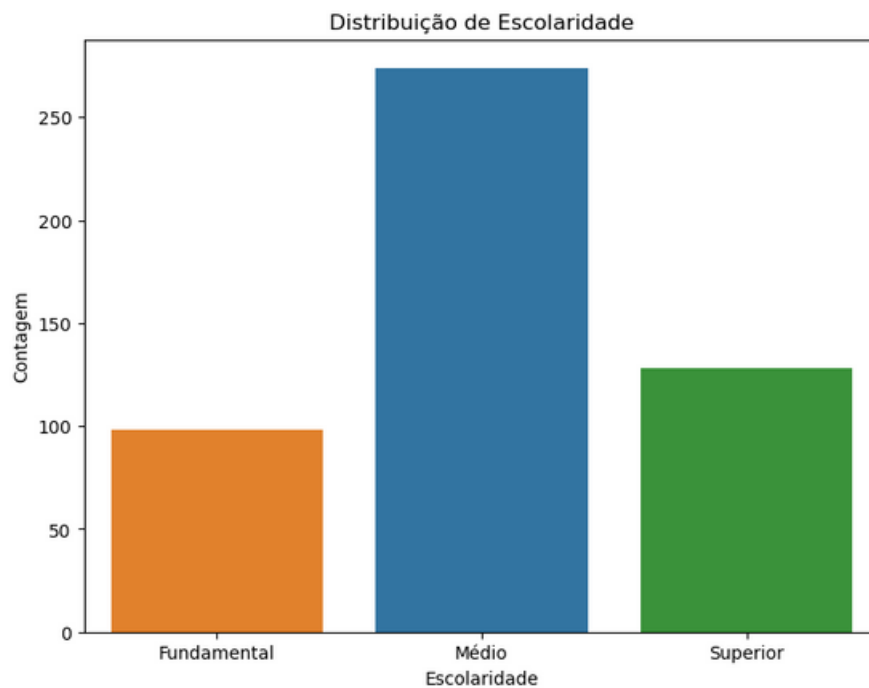


(31) 99175-9049



gabrieldbarros@gmail.com

Visualizando a Distribuição de Variáveis Qualitativas/ Categóricas





GABRIEL DELUCCA

CONSULTORIA EM PROCESSOS GERENCIAIS

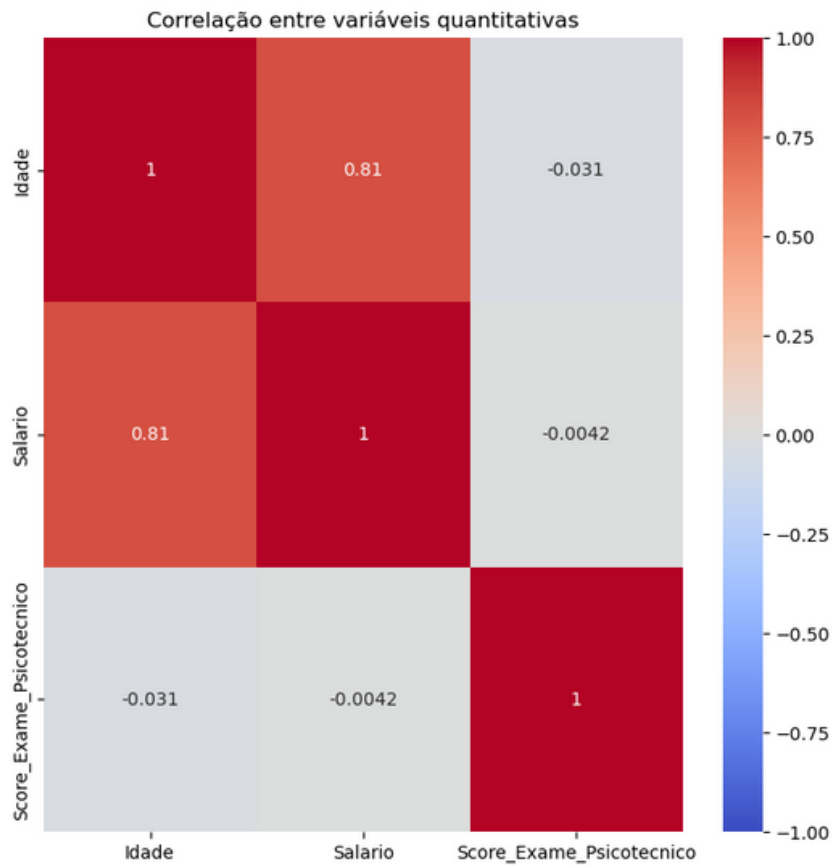


(31) 99175-9049



gabrielbarros@gmail.com

Correlação Entre Variáveis Quantitativas



Em uma primeira análise podemos observar que existe uma correlação positiva entre idades e salários.



GABRIEL DELUCCA

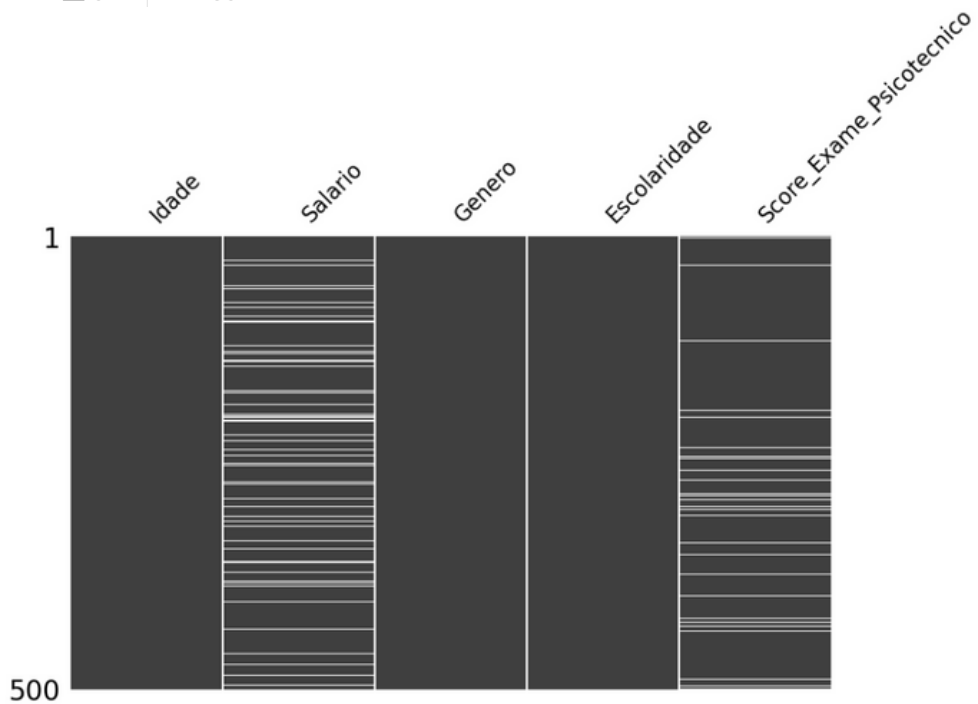
CONSULTORIA EM PROCESSOS GERENCIAIS



(31) 99175-9049



gabrielbarros@gmail.com



Colunas salário e score exame psicotecnico com valores ausentes. A coluna mais relevante é salário.

Teste de Normalidade Para Decidir a Estratégia de Tratamento de Valores Ausentes

```
print(valores_ausentes)
```

```
Idade          0
Salario        53
Genero         0
Escolaridade   0
Score_Exame_Psicotecnico  30
dtype: int64
```



GABRIEL DELUCCA

CONSULTORIA EM PROCESSOS GERENCIAIS



(31) 99175-9049



gabrieldbarros@gmail.com

Teste de Normalidade Para Decidir a Estratégia de Tratamento de Valores Ausentes

```
print(porcentagem_ausentes)
```

```
Idade          0.0
Salario        10.6
Genero         0.0
Escolaridade   0.0
Score_Exame_Psicotecnico  6.0
dtype: float64
```

```
[54]: from scipy import stats

# Extraia a coluna "salario" em uma série
salario = df_dados_originais['Salario']

# Aplique o teste de Shapiro-Wilk
stat, p_valor = stats.shapiro(salario)

# Imprima o resultado do teste
print(f"Estatística de teste: {stat}")
print(f"Valor-p: {p_valor}")

# Verifique a hipótese nula com base no p-valor
alfa = 0.05 # Nível de significância
if p_valor > alfa:
    print("Não há evidências para rejeitar a hipótese nula (os dados parecem seguir uma distribuição normal).")
else:
    print("A hipótese nula é rejeitada (os dados não seguem uma distribuição normal).")

Estatística de teste: nan
Valor-p: 1.0
Não há evidências para rejeitar a hipótese nula (os dados parecem seguir uma distribuição normal).
```

Dessa forma, podemos considerar utilizar média para preenchimento dos valores ausentes em salário.

Criação de uma nova coluna faixa etária para cálculo da média salarial por faixa etária.

```
Faixa_Etaria
Menos de 25      13108.546536
25-34            14945.150178
35-44            20980.622184
45-54            24736.492690
55 ou mais      29364.334600
Name: Salario, dtype: float64
```

Cálculo da mediana de salário para faixa etária

```
Faixa_Etaria
Menos de 25      12432.948937
25-34            15387.784406
35-44            21410.761236
45-54            24283.117285
55 ou mais      29345.327349
Name: Salario, dtype: float64
```




GABRIEL DELUCCA

CONSULTORIA EM PROCESSOS GERENCIAIS

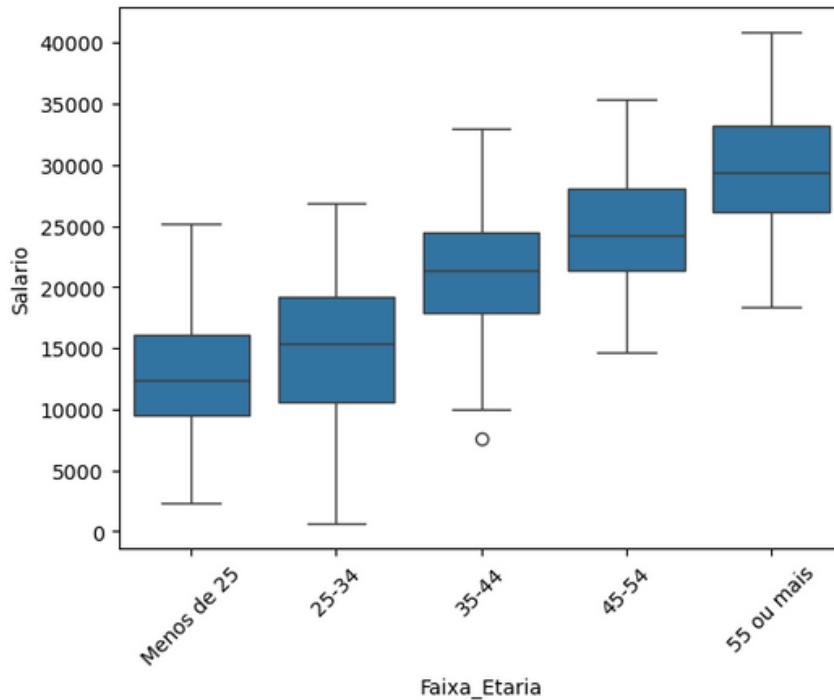


(31) 99175-9049



gabrieldbarros@gmail.com

Boxplots



Execução de Teste ANOVA

valor referência < 0,05

```
# Teste ANOVA

# Import
import scipy.stats as stats

# Realize um teste de ANOVA para verificar as diferenças nas médias
resultado_anova = stats.f_oneway(*[grupo['Salario'] for nome, grupo in
                                   df_dados_originais.groupby('Faixa_Etaria', observed=False)])

# Verifique o resultado do teste
if resultado_anova.pvalue < 0.05:
    print("Há evidências de diferenças significativas nas médias de salário entre as faixas etárias.")
else:
    print("Não há evidências de diferenças significativas nas médias de salário entre as faixas etárias.")
```

Há evidências de diferenças significativas nas médias de salário entre as faixas etárias.



Relatório Final

Com base em nossa análise observamos que:

- A variável Salario apresentava valores negativos. Como não temos mais detalhes sobre isso, decidimos converter de valor negativo para valor ausente.
- As variáveis Salario e Score_Exame_Psicotecnico apresentavam problemas de valores ausentes.
- Tratamos os valores ausentes com a imputação da média pois as duas variáveis seguem uma distribuição normal.
- Identificamos que não há relação entre as variáveis gênero e escolaridade, ou seja, elas são independentes.
- Verificamos que há relação entre idade e salário, independente do valor individual da idade ou de faixa etária.
- Nossas conclusões e decisões foram baseadas em testes estatísticos.

À disposição para maiores esclarecimentos.

Att.

Gabriel Delucca Barros
gabrielbarrosconsultoria@gmail.com
(31) 99175-9049

