

Clique duas vezes (ou pressione "Enter") para editar

```
1 import pandas as pd
2
3 df = pd.read_csv('https://docs.google.com/spreadsheets/d/e/2PACX-1vQXJcD06iN8cJTdXtvX7
4 display(df)
```



	Idade	Demissão	Viagem de negócios	Diárias recebidas	Departamento	Distância de Casa	Escolaridade
0	41	Sim	Viaja raramente	1102	Vendas	1	
1	49	Não	Viaja frequentemente	279	Pesquisa e Desenvolvimento	8	
2	37	Sim	Viaja raramente	1373	Pesquisa e Desenvolvimento	2	
3	33	Não	Viaja frequentemente	1392	Pesquisa e Desenvolvimento	3	
4	27	Não	Viaja raramente	591	Pesquisa e Desenvolvimento	2	
...	
1465	36	Não	Viaja frequentemente	884	Pesquisa e Desenvolvimento	23	
1466	39	Não	Viaja raramente	613	Pesquisa e Desenvolvimento	6	
1467	27	Não	Viaja raramente	155	Pesquisa e Desenvolvimento	4	
1468	49	Não	Viaja frequentemente	1023	Vendas	2	
1469	34	Não	Viaja raramente	628	Pesquisa e Desenvolvimento	8	

1470 rows × 35 columns

```
1 # Convertendo a coluna "Demissão" para valores numéricos (0 para não e 1 para sim)
2 df['Demissão'] = df['Demissão'].map({'Sim': 1, 'Não': 0})
3
4 display(df)
```



	Idade	Demissão	Viagem de negócios	Diárias recebidas	Departamento	Distância de Casa	Escolaridade
0	41	1	Viaja raramente	1102	Vendas	1	
1	49	0	Viaja frequentemente	279	Pesquisa e Desenvolvimento	8	
2	37	1	Viaja raramente	1373	Pesquisa e Desenvolvimento	2	
3	33	0	Viaja frequentemente	1392	Pesquisa e Desenvolvimento	3	
4	27	0	Viaja raramente	591	Pesquisa e Desenvolvimento	2	
...	
1465	36	0	Viaja frequentemente	884	Pesquisa e Desenvolvimento	23	
1466	39	0	Viaja raramente	613	Pesquisa e Desenvolvimento	6	
1467	27	0	Viaja raramente	155	Pesquisa e Desenvolvimento	4	
1468	49	0	Viaja frequentemente	1023	Vendas	2	
1469	34	0	Viaja raramente	628	Pesquisa e Desenvolvimento	8	

1470 rows × 35 columns

```
1 # Selecionando apenas as colunas numéricas
2 dados_numericos = df.select_dtypes(include='int64')
3
4 # Calculando a correlação entre "demissões" e todas as outras colunas numéricas
5 correlacoes = dados_numericos.corrwith(df['Demissão'])
6
7 # Exibindo as correlações
8 print(correlacoes)
```



Idade	-0.159205
Demissão	1.000000
Diárias recebidas	-0.056652
Distância de Casa	0.077924
Escolaridade	-0.031373
Número de contratos de trabalho na empresa	NaN
ID Funcionário	-0.010577
Nível de Satisfação com o ambiente de trabalho	-0.103369

Salário por hora	-0.003561
Nível de envolvimento com o trabalho	-0.130016
Nível hierárquico	-0.169105
Satisfação com o trabalho	-0.103481
Renda mensal	-0.159840
Salário Mensal	0.015170
qtde de empresas que já trabalhou	0.043494
percentual de aumento de salário	-0.013478
score de performance	0.002889
Satisfação nas relações Não trabalho	-0.045872
Jornada padrão de trabalho	NaN
opção de remaunração variável em ações	-0.137145
Total de anos trabalhados	-0.171063
Qtde de treinamentos realizados no último ano	-0.059478
Equilíbrio vida trabalho	-0.063939
Total de anos trabalhados na empresa	-0.134392
Anos trabalhados na função atual	-0.160545
Anos desde a última promoção	-0.033019
Anos trabalhando com o atual gestor	-0.156199

dtype: float64

```

1 # calculando a porcentagem de relação
2 porcentagem_correlacao = round((correlacoes ** 2) * 100,2)
3 print(porcentagem_correlacao.sort_values(ascending=False))

```

Demissão	100.00
Total de anos trabalhados	2.93
Nível hierárquico	2.86
Anos trabalhados na função atual	2.58
Renda mensal	2.55
Idade	2.53
Anos trabalhando com o atual gestor	2.44
opção de remaunração variável em ações	1.88
Total de anos trabalhados na empresa	1.81
Nível de envolvimento com o trabalho	1.69
Nível de Satisfação com o ambiente de trabalho	1.07
Satisfação com o trabalho	1.07
Distância de Casa	0.61
Equilíbrio vida trabalho	0.41
Qtde de treinamentos realizados no último ano	0.35
Diárias recebidas	0.32
Satisfação nas relações Não trabalho	0.21
qtde de empresas que já trabalhou	0.19
Anos desde a última promoção	0.11
Escolaridade	0.10
Salário Mensal	0.02
percentual de aumento de salário	0.02
ID Funcionário	0.01
score de performance	0.00
Salário por hora	0.00
Número de contratos de trabalho na empresa	NaN
Jornada padrão de trabalho	NaN

dtype: float64

```
1 dados_numericos.groupby('Demissão')[['Total de anos trabalhados']].mean()
```



Total de anos trabalhados

Demissão		
0		11.862936
	0.75	
1		8.244726
	0.75	

```
1 dados_numericos.groupby('Demissão')[['Nível hierárquico']].quantile([0.5, 0.75])
```



Nível hierárquico

Demissão		
0	0.50	2.0
	0.75	3.0
1	0.50	1.0
	0.75	2.0

```
1 dados_numericos.groupby('Demissão')[['Anos trabalhados na função atual']].quantile([0.
```



Anos trabalhados na função atual

Demissão		
0	0.75	7.0
1	0.75	4.0

```
1 dados_numericos.groupby('Demissão')[['Renda mensal']].quantile([0.75])
```



Renda mensal

Demissão		
0	0.75	8834.0
1	0.75	5916.0

```
1 dados_numericos.groupby('Demissão')[['Satisfação com o trabalho']].quantile([0.25, 0.5
```



Satisfação com o trabalho

Demissão		
0	0.25	2.0
	0.50	3.0
	0.75	4.0
1	0.25	1.0
	0.50	3.0
	0.75	3.0

```
1 dados_numericos.groupby('Demissão')[['Nível de Satisfação com o ambiente de trabalho']]
```



Nível de Satisfação com o ambiente de trabalho

Demissão	
0	2.771290
1	2.464135

```
1 dados_numericos.groupby('Demissão')[['Nível de Satisfação com o ambiente de trabalho']]
```



Nível de Satisfação com o ambiente de trabalho

Demissão		
0	0.50	3.0
	0.75	4.0
1	0.50	3.0
	0.75	4.0

- Nível hierárquico <= 2
- Anos trabalhados na função atual <= 4
- Renda mensal <= 5.900
- Satisfação com o trabalho <= 3

```
1 df = pd.read_csv('https://docs.google.com/spreadsheets/d/e/2PACX-1vQXJcD06iN8cJTdXtvX7
2
3 colunas_object = df.select_dtypes(include=['object'])
4 colunas_object.info()
```




```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1470 entries, 0 to 1469
Data columns (total 9 columns):
#    Column                Non-Null Count  Dtype

```

```
---  -----  -----  ----
0  Demissão      1470 non-null  object
1  Viagem de negócios  1470 non-null  object
2  Departamento  1470 non-null  object
3  Área de Formação  1470 non-null  object
4  Gênero        1470 non-null  object
5  Cargo         1470 non-null  object
6  Estado civil  1470 non-null  object
7  Maior de idade  1470 non-null  object
8  Faz hora extra  1470 non-null  object
dtypes: object(9)
memory usage: 103.5+ KB
```


✓ Filtrado por pessoas que pediram demissão

```
1 demissaoSim = colunas_object.loc[df['Demissão'] == 'Sim']
2 percentagemGroupby = demissaoSim.groupby(['Viagem de negócios'])[['Viagem de negócios']]
3 round(percentagemGroupby, 2)
```

 Viagem de negócios

Viagem de negócios	
Não Viaja	5.06
Viaja frequentemente	29.11
Viaja raramente	65.82

```
1 demissaoSim = colunas_object.loc[df['Demissão'] == 'Sim']
2 percentagemGroupby = demissaoSim.groupby(['Departamento'])[['Departamento']].count()/der
3 round(percentagemGroupby, 2)
```

 Departamento

Departamento	
Pesquisa e Desenvolvimento	56.12
Recursos HumaNãos	5.06
Vendas	38.82

```
1 Comece a programar ou gere código com IA.
```