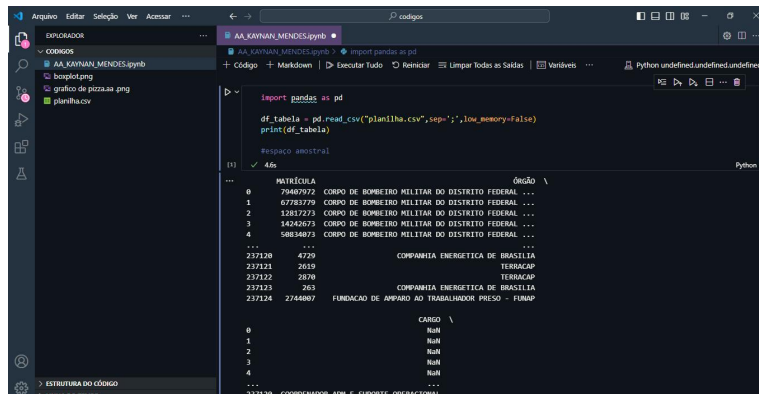


ATIVIDADE AVALIATIVA

NOTA: Eu realizei a atividade através do vscode, baixei a base de dados e alterei o nome para “planilha” e chamei de “df_tabela”, então para que os códigos funcionem perfeitamente é necessário realizar essa alteração.



```
import pandas as pd

df_tabela = pd.read_csv("planilha.csv", sep=";", low_memory=False)
print(df_tabela)

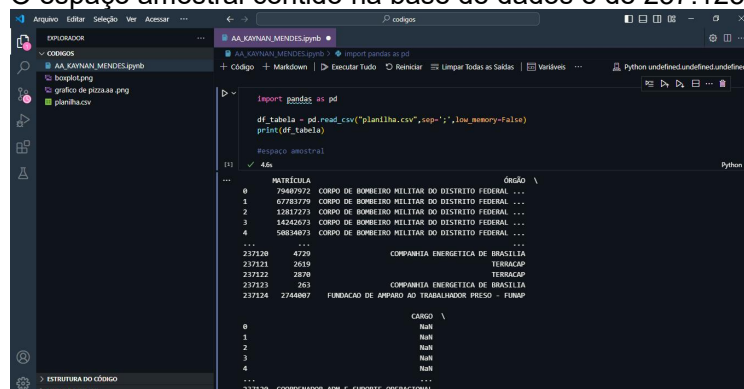
#espaço amostral

...
0 79467972 CORPO DE BOMBEIRO MILITAR DO DISTRITO FEDERAL ...
1 67783779 CORPO DE BOMBEIRO MILITAR DO DISTRITO FEDERAL ...
2 12817273 CORPO DE BOMBEIRO MILITAR DO DISTRITO FEDERAL ...
3 14242673 CORPO DE BOMBEIRO MILITAR DO DISTRITO FEDERAL ...
4 58834873 CORPO DE BOMBEIRO MILITAR DO DISTRITO FEDERAL ...
...
237120 4720 COMPANHIA ENERGETICA DE BRASLIA
237121 2619 TERRACAP
237122 2870 TERRACAP
237123 263 COMPANHIA ENERGETICA DE BRASLIA
237124 2744887 FUNDACAO DE AMPARO AO TRABALHADOR PRESO - FUNAP

...
0 NaN
1 NaN
2 NaN
3 NaN
4 NaN
...
237120 COORDENADOR ADM E SUPORTE OPERACIONAL
```

Qual o espaço amostral compreendido nesta base de dados?

O espaço amostral contido na base de dados é de 237.125.



```
import pandas as pd

df_tabela = pd.read_csv("planilha.csv", sep=";", low_memory=False)
print(df_tabela)

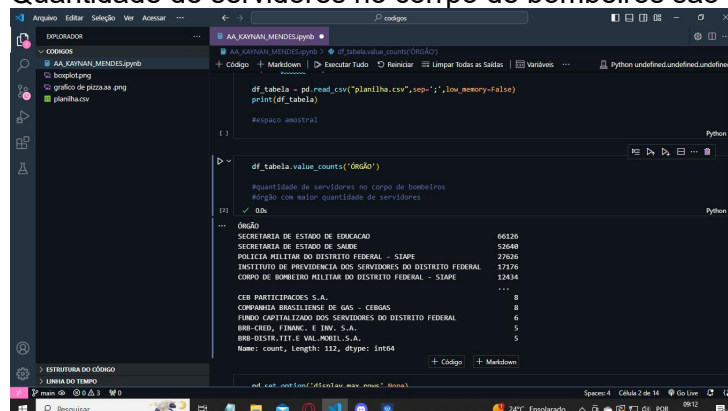
#espaço amostral

...
0 79467972 CORPO DE BOMBEIRO MILITAR DO DISTRITO FEDERAL ...
1 67783779 CORPO DE BOMBEIRO MILITAR DO DISTRITO FEDERAL ...
2 12817273 CORPO DE BOMBEIRO MILITAR DO DISTRITO FEDERAL ...
3 14242673 CORPO DE BOMBEIRO MILITAR DO DISTRITO FEDERAL ...
4 58834873 CORPO DE BOMBEIRO MILITAR DO DISTRITO FEDERAL ...
...
237120 4720 COMPANHIA ENERGETICA DE BRASLIA
237121 2619 TERRACAP
237122 2870 TERRACAP
237123 263 COMPANHIA ENERGETICA DE BRASLIA
237124 2744887 FUNDACAO DE AMPARO AO TRABALHADOR PRESO - FUNAP

...
0 NaN
1 NaN
2 NaN
3 NaN
4 NaN
...
237120 COORDENADOR ADM E SUPORTE OPERACIONAL
```

Quanto servidores estão lotados no corpo de bombeiros?

Quantidade de servidores no corpo de bombeiros são de 12.434.



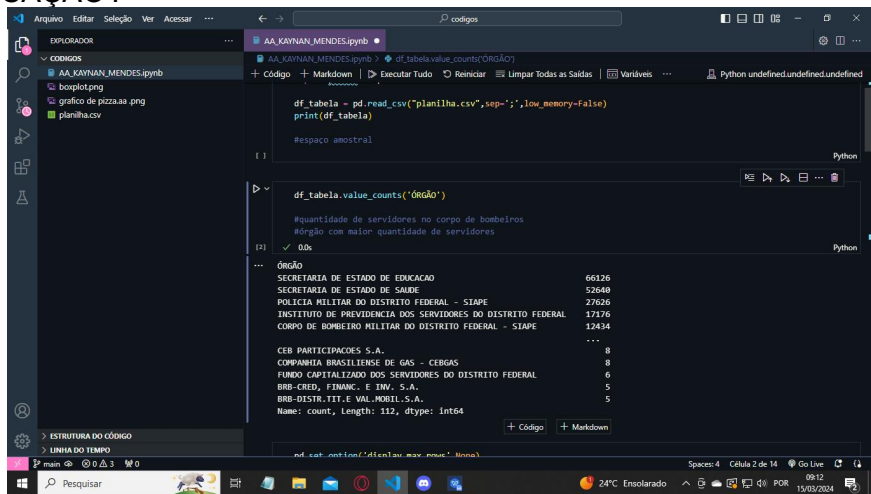
```
df_tabela.value_counts(ORGÃO)

Quantidade de servidores no corpo de bombeiros
Órgão com maior quantidade de servidores

...
0 66126
1 23648
2 27626
3 12719
4 12434
...
0 8
1 8
2 6
3 5
4 5
Name: count, Length: 122, dtype: int64
```

Qual o órgão público com maior número de funcionários?

O órgão com maior quantidade de servidores é 'SECRETARIA DE ESTADO E EDUCAÇÃO'.



```
df_tabela = pd.read_csv("planilha.csv", sep=';', low_memory=False)
print(df_tabela)

#espaço amostral

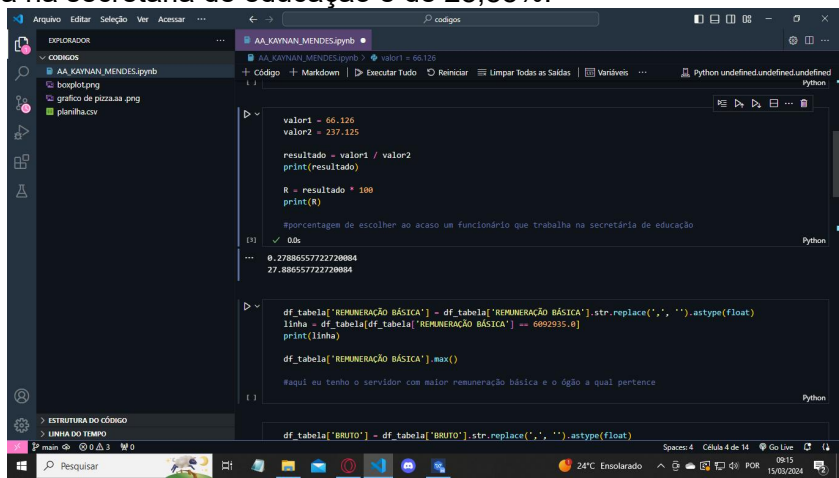
df_tabela.value_counts('ÓRGÃO')

#quantidade de servidores no corpo de bombeiros
#órgão com maior quantidade de servidores

ÓRGÃO
SECRETARIA DE ESTADO DE EDUCACAO      66126
SECRETARIA DE ESTADO DE SAUDE          58649
POLICIA MILITAR DO DISTRITO FEDERAL - SIAPE  27626
INSTITUTO DE PREVIDENCIA DOS SERVIDORES DO DISTRITO FEDERAL  17176
CORPO DE BOMBEIRO MILITAR DO DISTRITO FEDERAL - SIAPE  12434
...
CEB PARTICIPACOES S.A.                  8
COMPANHIA BRASILENSE DE GAS - CEBGAS    8
FUNDO CAPITALIZADO DOS SERVIDORES DO DISTRITO FEDERAL    6
BRB-CRED, FIMMAG, E IMV, S.A.           5
BRB-DISTRIT, E VAL-MODUL, S.A.           5
Name: count, dtype: int64
```

Qual a probabilidade de ao escolher uma linha ao acaso, ser de um funcionário que trabalha na secretaria de educação?

A probabilidade de escolher uma linha ao acaso e ser um funcionário que trabalha na secretaria de educação é de 28,88%.



```
valor1 = 66.126
valor2 = 237.125

resultado = valor1 / valor2
print(resultado)

R = resultado * 100
print(R)

#porcentagem de escolher ao acaso um funcionario que trabalha na secretaria de educação

0.288857722728004
27.8857722728004

df_tabela['REMUNERACAO BASICA'] = df_tabela['REMUNERACAO BASICA'].str.replace(',', '').astype(float)
linha = df_tabela[df_tabela['REMUNERACAO BASICA'] == 6092935.0]
print(linha)

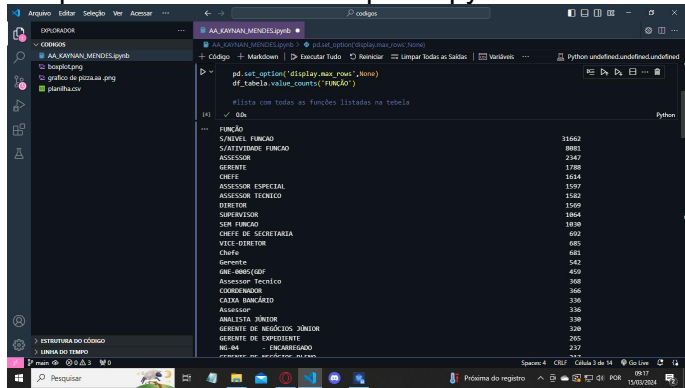
df_tabela['REMUNERACAO BASICA'].max()

#aqui eu tenho o servidor com maior remuneração básica e o órgão a qual pertence

df_tabela['BRUTO'] = df_tabela['BRUTO'].str.replace(',', '').astype(float)
```

Liste todas as funções contidas na base

Resposta na célula 3 do arquivo .ipynb.



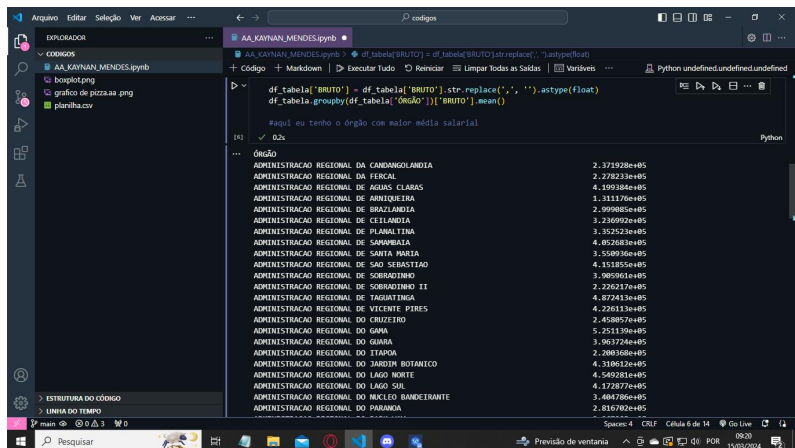
```
pd.set_option('display.max_rows', None)
df_tabela.value_counts('FUNCAO')

#lista com todas as funções listadas na tabela

FUNCAO
FUNCAO      31662
S/FUNVEL FUNCAO      8881
S/ATIVIDADE FUNCAO      2447
ASSESSOR      1788
GERENTE      1524
CHIEFE      1597
ASSESSOR ESPECIAL      1582
DIRETOR      1548
SUPERVISOR      1404
SRV FUNCAO      1328
CHIEFE DE SECRETARIA      692
VICE-DIRETOR      685
CHIEFE      681
Gerente      542
DIRE-ASSESSOR      499
Assessor Tecnico      368
COORDENADOR      366
CAIXA REMATE      336
Assessor      336
ANALISTA JUNIOR      336
GERENTE DE NEGOCIOS JUNIOR      328
GERENTE DE EXPEDIENTE      265
ME-04 - EXPEDIENTE      217
```

Qual órgão publico possui a maior média salarial?

O órgão público com maior média salarial é 'BRB-CRED, FINANC. E INV. S.A.'.



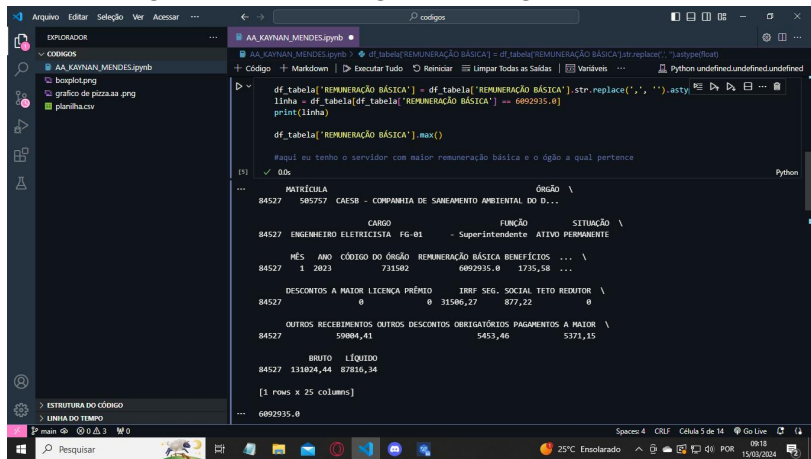
```
df_tabela["BRUTO"] = df_tabela["BRUTO"].str.replace(",", "").astype(float)
df_tabela.groupby(df_tabela["Órgão"])[["BRUTO"]].mean()

# aqui eu tenho o órgão com maior média salarial
```

Órgão	BRUTO
ADMINISTRACAO REGIONAL DA CANDANGOLANDIA	2.371928e+05
ADMINISTRACAO REGIONAL DA FERCAL	2.228213e+05
ADMINISTRACAO REGIONAL DE AGUAS CLARAS	4.199384e+05
ADMINISTRACAO REGIONAL DE ABRILIO LIRA	1.311176e+05
ADMINISTRACAO REGIONAL DE BRASILEIRIA	2.099865e+05
ADMINISTRACAO REGIONAL DE CEILANDIA	3.236992e+05
ADMINISTRACAO REGIONAL DE PLAMALINA	3.352523e+05
ADMINISTRACAO REGIONAL DE SAMPALVA	4.852683e+05
ADMINISTRACAO REGIONAL DE SANTA MARIA	3.504930e+05
ADMINISTRACAO REGIONAL DE SAO SEBASTIAO	4.151855e+05
ADMINISTRACAO REGIONAL DE SOBRADIMHO	3.905961e+05
ADMINISTRACAO REGIONAL DE SOBRADIMHO II	2.280217e+05
ADMINISTRACAO REGIONAL DE TAGUATINGA	4.872413e+05
ADMINISTRACAO REGIONAL DE VICENTE PIRES	4.226113e+05
ADMINISTRACAO REGIONAL DO CRUZEIRO	2.458897e+05
ADMINISTRACAO REGIONAL DO GAMA	5.251139e+05
ADMINISTRACAO REGIONAL DO GUARA	3.963724e+05
ADMINISTRACAO REGIONAL DO ITACARA	2.280740e+05
ADMINISTRACAO REGIONAL DO JARDIM BOTANICO	4.318612e+05
ADMINISTRACAO REGIONAL DO LAGO NORTE	4.549281e+05
ADMINISTRACAO REGIONAL DO LAGO SUL	4.127287e+05
ADMINISTRACAO REGIONAL DO NUCLEO BANDEIRANTE	3.484786e+05
ADMINISTRACAO REGIONAL DO PARANAO	2.816702e+05

O servidor com maior remuneração básica pertence a qual órgão?

O servidor com maior remuneração básica pertence a 'CAESB-COMPANHIA DE SANEAMENTO AMBIENTAL DO DISTRITO FEDERAL'.



```
df_tabela["REMUNERACAO BASICA"] = df_tabela["REMUNERACAO BASICA"].str.replace(",", "").astype(float)
linha = df_tabela[df_tabela["REMUNERACAO BASICA"] == 6092935.0]
print(linha)

df_tabela["REMUNERACAO BASICA"].max()

# aqui eu tenho o servidor com maior remuneração básica e o órgão a qual pertence
```

MATRÍCULA	Órgão	CARGO	FUNÇÃO	SITUAÇÃO
84527 565757	CAESB - COMPANHIA DE SANEAMENTO AMBIENTAL DO D...			
84527	ENGENHEIRO ELETRICISTA	FG-01	Superintendente	ATIVO PERMANENTE

MÊS	ANO	CÓDIGO DO ÓRGÃO	REMUNERAÇÃO BÁSICA	BENEFÍCIOS
84527	1	2023	731502	6092935.0 1735,58

DESCONTOS A MAIOR LICENÇA	PRÊMIO	IRRF SEG. SOCIAL	TETO	RENTIST
84527	0	31586,27	877,22	0

OUTROS RECEBIMENTOS	OUTROS DESCONTOS	OBRIGATORIOS	PAGAMENTOS A MAIOR
84527	59004,41	5453,46	5371,15

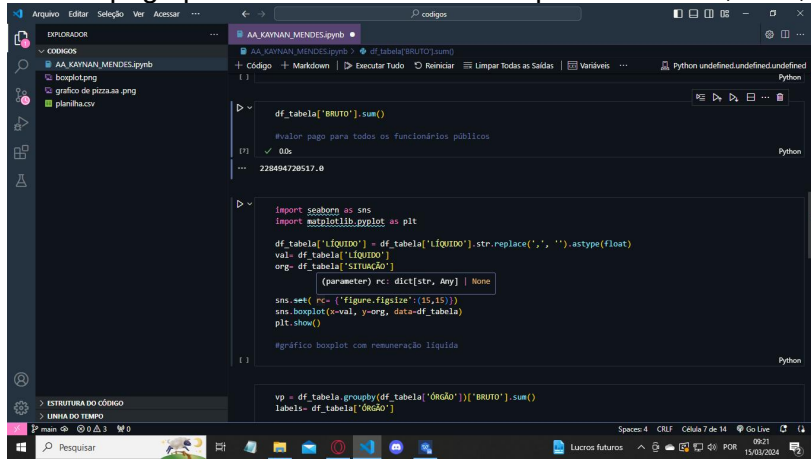
BRUTO	LÍQUIDO
84527 1318024,44	87816,34

[1 rows x 25 columns]

6092935.0

Qual o valor pago para todos os funcionários públicos?

O valor pago para todos os funcionários públicos é 228,4947205170.



```
df_tabela["BRUTO"].sum()

#valor pago para todos os funcionarios publicos
```

228494720517.0

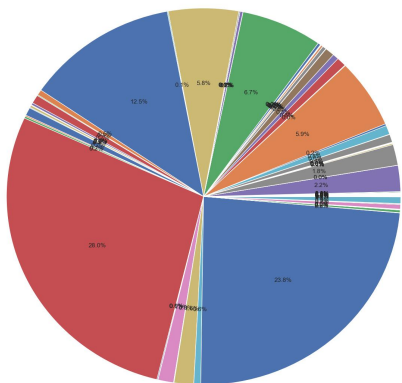
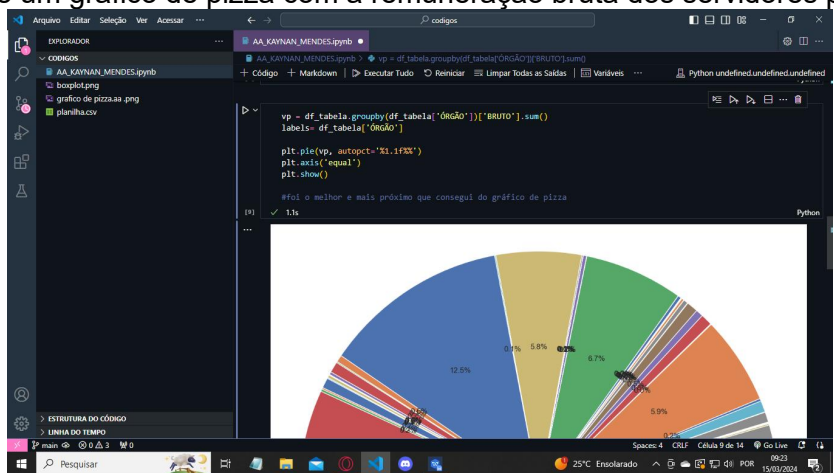
```
df_tabela["LÍQUIDO"] = df_tabela["LÍQUIDO"].str.replace(",", "").astype(float)
val = df_tabela["LÍQUIDO"]
org = df_tabela["SITUAÇÃO"]

sns.set(rc={"figure.figsize":(15,15)})
sns.boxplot(x=val, y=org, data=df_tabela)
plt.show()

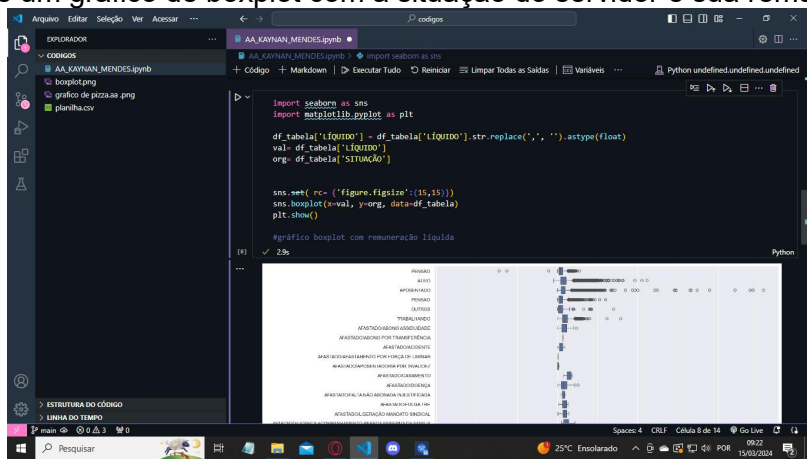
#gráfico boxplot com remuneração líquida
```

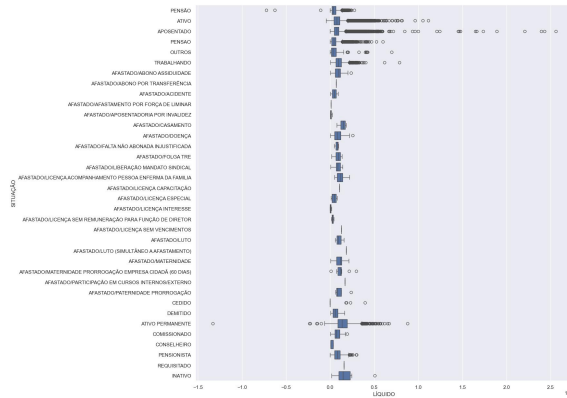
```
vp = df_tabela.groupby(df_tabela["Órgão"])[["BRUTO"]].sum()
tabela = df_tabela
```

Elabore um gráfico de pizza com a remuneração bruta dos servidores por órgão



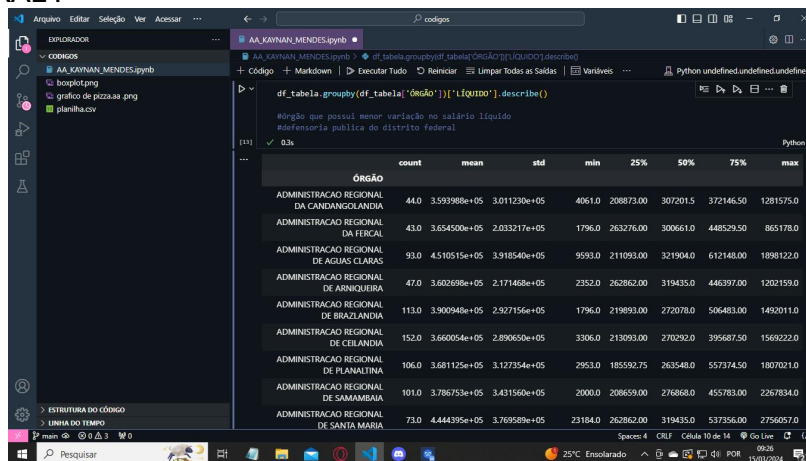
Elabore um gráfico de boxplot com a situação do servidor e sua remuneração líquida





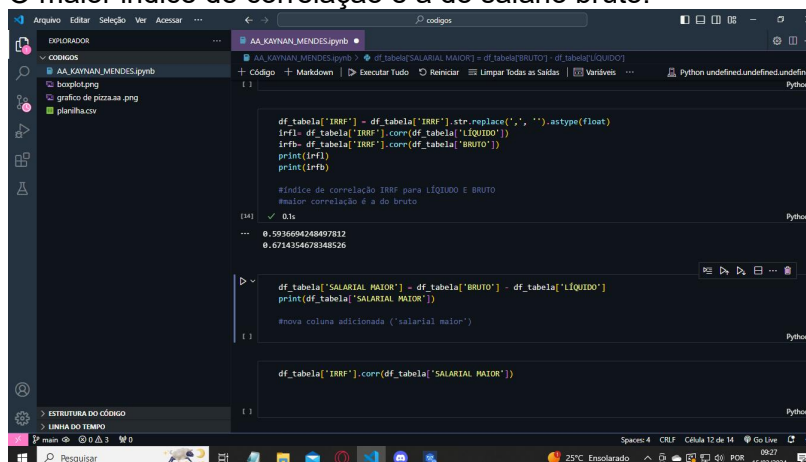
Identifique qual órgão possui salário líquido com menos variações, utilizando para isto medidas de dispersão.

O órgão com menor variação é 'DEFENSORIA PUBLICA DO DISTRITO FEDERAL'.



Calcule o índice de correlação entre o IRRF e os salários líquido e bruto. Em qual dos casos o índice de correlação foi maior?

O maior índice de correlação é a do salário bruto.



Adicione uma nova coluna que irá conter a diferença entre o salario bruto e liquido, e responda as seguintes questões;

