

Beatriz Cardoso Silva SP3099849
Guilherme Henrique Santana da Silva SP3109208

PROJETO DE ESTATÍSTICA E PROBABILIDADE

SÃO PAULO

2025

PROJETO DE ESTATÍSTICA E PROBABILIDADE

Trabalho apresentado à disciplina de Estatística e Probabilidade do Instituto Federal de Educação, Ciência e Tecnologia Campus São Paulo como nota parcial para aprovação na disciplina do curso em Estatística e Probabilidade.

SÃO PAULO - SP

2025

SUMÁRIO

INTRODUÇÃO.....4

BASES DE DADOS..... 4

ANÁLISES.....4

INTRODUÇÃO

A partir do conteúdo apresentado na disciplina de Estatística e Probabilidade este relatório contém explicações e descrições acerca dos cálculos feitos a partir de base de dados disponibilizada desenvolvimento do projeto. A partir dos dados, foi possível levantar questionamentos em referência a eles e obter resultados precisos.

BASE DE DADOS

As bases de dados selecionadas para o presente relatório foram extraídas a partir de repositório do GitHub contendo dados referentes ao COVID-19. Dentre as opções, as bases selecionadas foram:

1. Vacinação por faixa etária (<https://github.com/owid/covid-19-data/blob/master/public/data/vaccinations/vaccinations-by-age-group.csv>);
2. Hospitalizações por COVID-19 (<https://github.com/owid/covid-19-data/blob/master/public/data/hospitalizations/covid-hospitalizations.csv>).

ESTATÍSTICA DESCRITIVA

O primeiro levantamento proposto questiona se ocupação hospitalar e a taxa de internação em UTI por COVID-19 apresentaram variação significativa ao longo de períodos entre 2020 e 2023. Para responder a esta questão, recorreremos à estatística descritiva e calculamos a média amostral e a mediana das taxas de ocupação hospitalar e de UTI para cada semestre do período de análise. A média de ocupação por milhão de habitantes provê uma informação instantânea desta ocupação hospitalar, a medida que a mediana da ocupação por milhão de habitantes ajuda a reduzir o

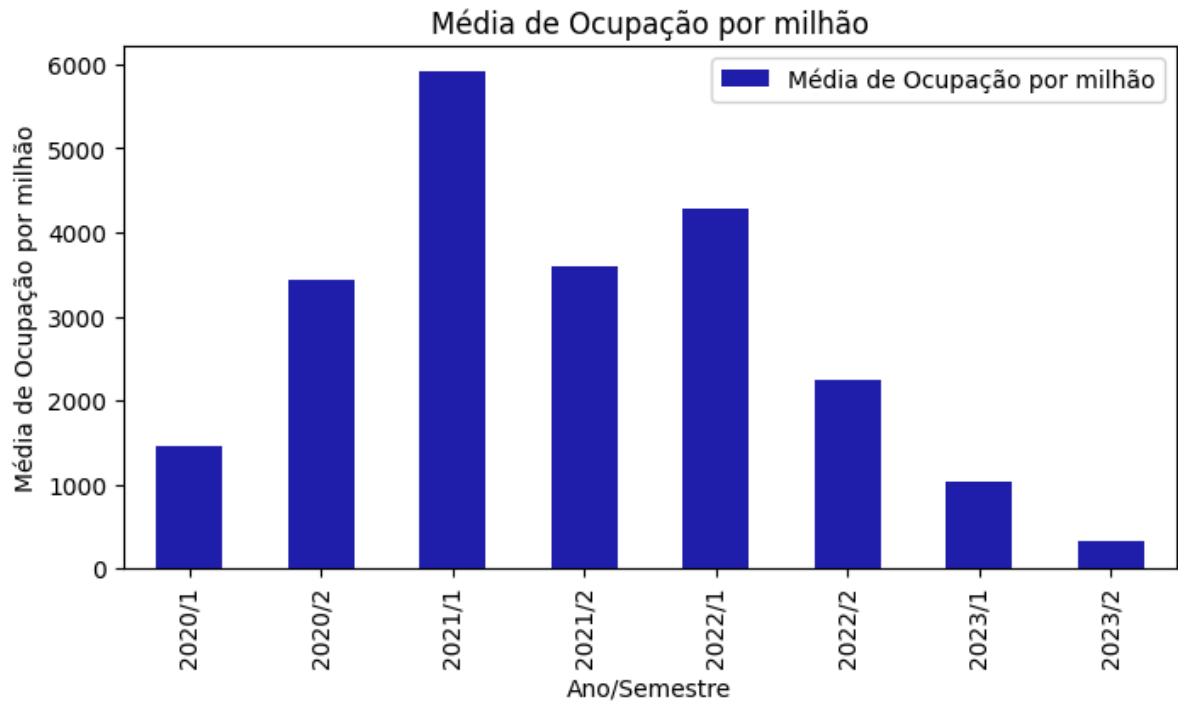
impacto de valores, representa o ponto central da distribuição. Este foi o fluxo de trabalho:

1. Filtrar dados de ocupação hospitalar e de UTI em decorrência da COVID-19 dentro de cada semestre;
2. Criação de tabela pivotada;
3. Cálculo da soma diária da ocupação hospitalar e de UTI;
4. Informações estatísticas descritivas de cada período.

O resultado obtido é observado conforme tabela abaixo:

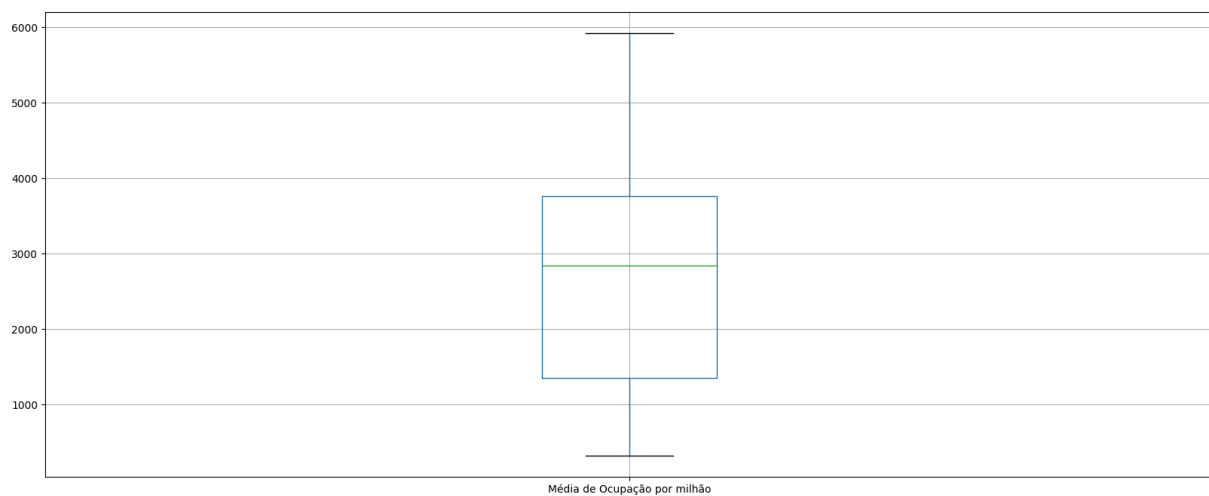
	Ano	Semestre	Média de Ocupação por milhão	Mediana de Ocupação por milhão
0	2020	1	1453.453221	1114.0400
1	2020	2	3433.987701	1503.2705
2	2021	1	5923.007514	6876.2680
3	2021	2	3587.488478	3247.7110
4	2022	1	4275.984851	4121.5310
5	2022	2	2236.248125	2244.3490
6	2023	1	1040.522547	1137.2850
7	2023	2	320.089060	297.8390

Para melhor visualização, utilizamos gráfico de barra e caixas para destacar os principais períodos da pandemia e o provável impacto da campanha de vacinação.

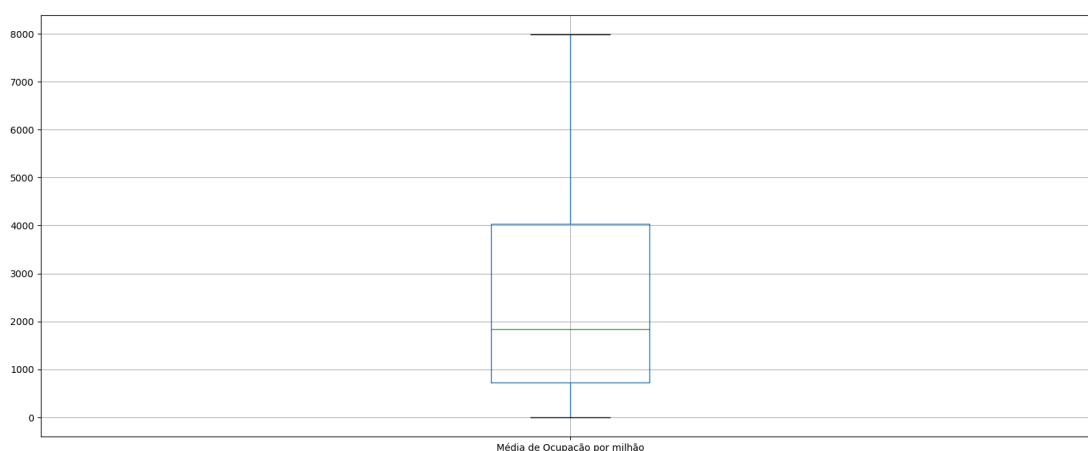
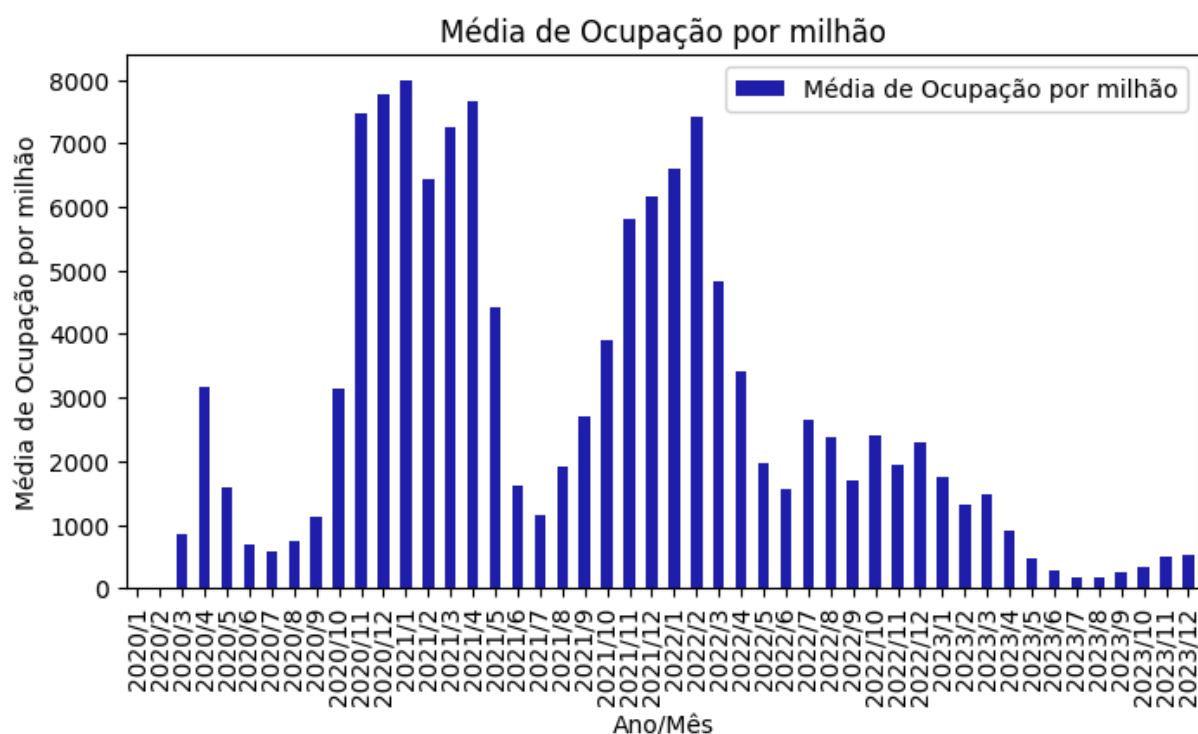


É possível observar com mais clareza a partir do gráfico como a ocupação variou durante o período estudado. No primeiro semestre de 2020, a ocupação se iniciou e houve um aumento considerável no segundo semestre do ano. O primeiro semestre de 2021 apresenta a maior alta de internações pelo vírus, com quase 6 milhões de internados. Em 2022 há uma queda hierárquica.

O Box Plot demonstra a distribuição de forma resumida.



Também utilizamos dados da base para analisar a média de ocupação por meses entre janeiro de 2021 e dezembro de 2023.

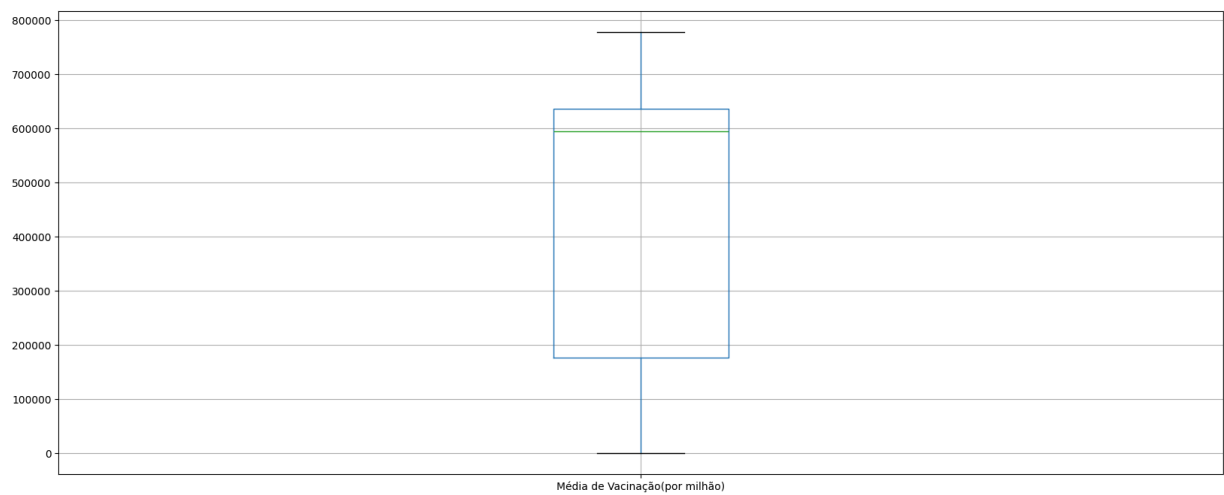
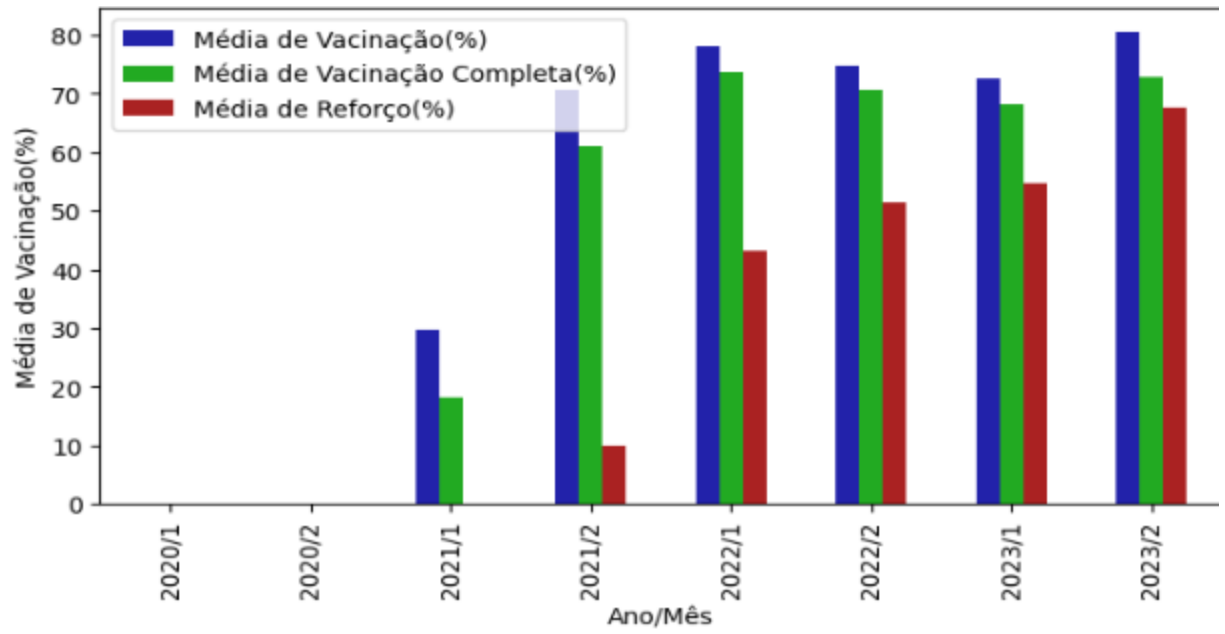


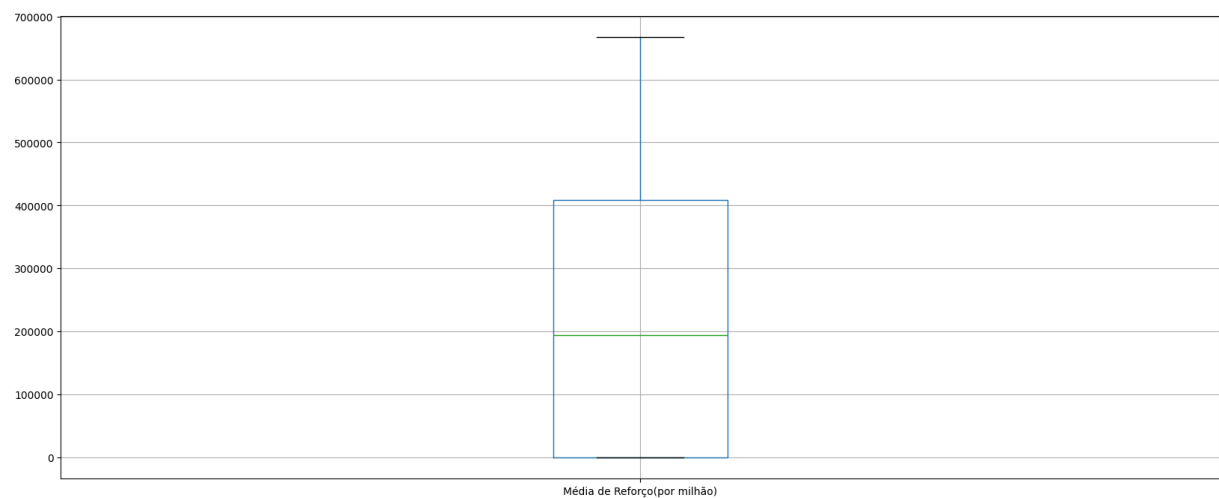
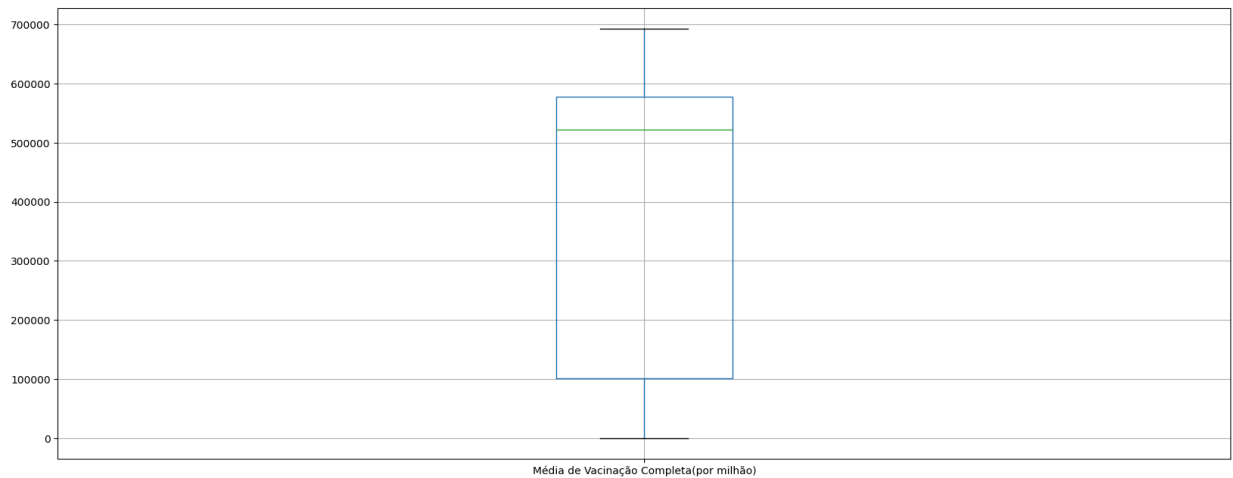
O próximo passo foi referente a segunda fonte de dados, vacinados contra a COVID-19 por faixa etária. Foi analisada a evolução da vacinação contra a COVID-19 em diferentes países ao longo dos semestres entre os anos de 2020 e 2023. Os dados incluem percentual da população vacinada com pelo menos uma dose, percentual da população totalmente vacinada e percentual da população que recebeu a dose de reforço. Os passos para coleta foram:

1. Filtrar os dados por ano e mês;
2. Agrupar as informações;
3. Calcular as médias de cada categoria de vacinação;
4. Criar uma tabela consolidada com os valores.

	Ano	Semestre	Média de Vacinação(%)	Média de Vacinação Completa(%)	Média de Reforço(%)
0	2020	1	0.000000	0.000000	0.000000
1	2020	2	0.001466	0.000000	0.000000
2	2021	1	29.649594	18.267481	0.019917
3	2021	2	70.680426	61.117873	10.003291
4	2022	1	78.087965	73.624454	43.083532
5	2022	2	74.742719	70.724362	51.500763
6	2023	1	72.712957	68.281949	54.695259
7	2023	2	80.636751	72.896930	67.720496

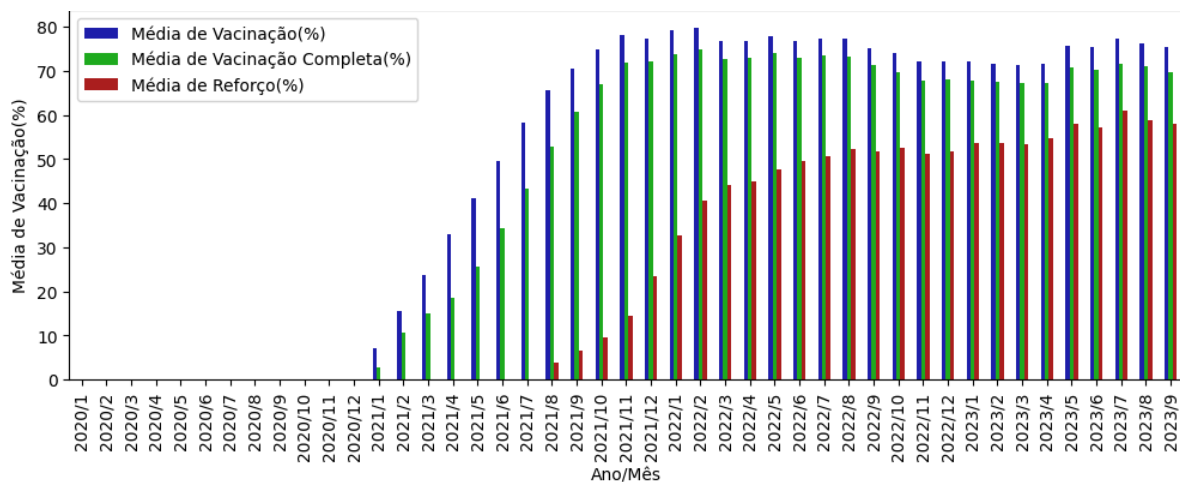
Um gráfico e box-plots para melhor visualização e uma compreensão intuitiva das tendências ao longo do tempo destas informações também foi feito.





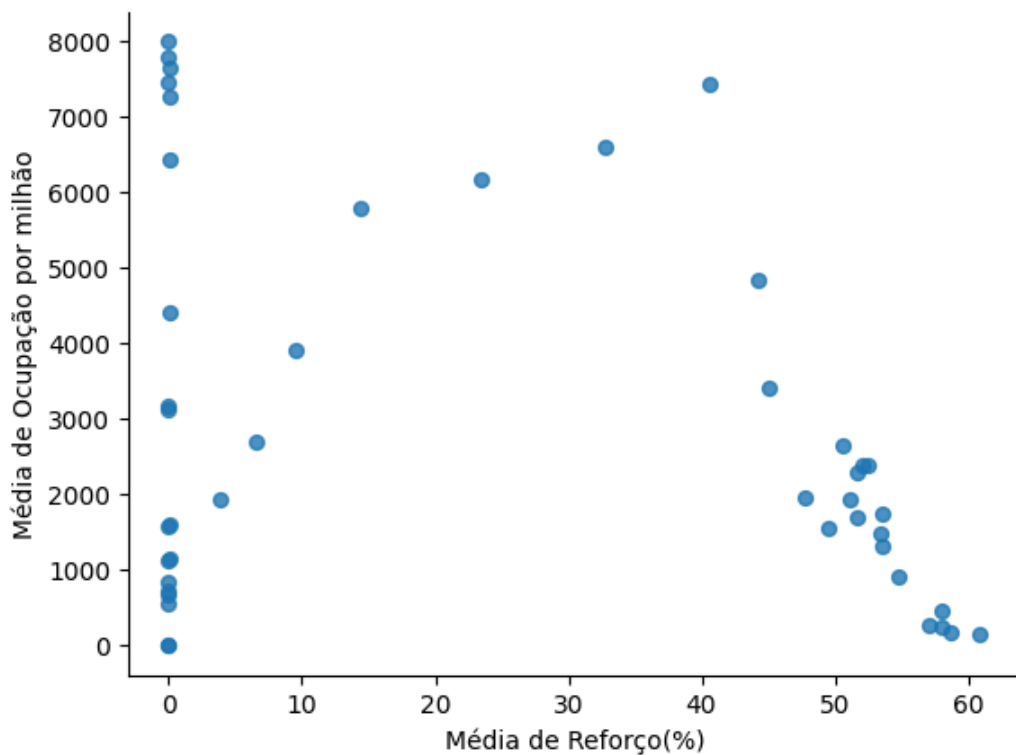
A taxa de vacinação apresentou um crescimento acelerado a partir do segundo semestre de 2021, a medida em que a distribuição das vacinas se tornou mais acessível e amplamente distribuída.

Para observação mais precisa, contemplamos também informações entre nos meses de 2020 a 2023.



A próxima análise visou explorar a relação entre a ocupação média dos hospitais e as taxas de vacinação de reforço usando um gráfico de dispersão.

A regressão linear quantifica a relação entre a taxa média mensal de vacinação de reforço e a ocupação média mensal de hospitais e UTIs.



O gráfico de dispersão sugere uma tendência em que maiores taxas de vacinação com dose de reforço estão associadas a menores taxas de ocupação hospitalar. No entanto, outros fatores podem influenciar essa relação, conforme as variações encontradas.

Implementamos o Teorema de Bayes para calcular a probabilidade condicional de hospitalização dado um determinado *status* de vacinação. A partir do resultado é possível analisar que a probabilidade de uma hospitalização dado que um indivíduo tomou a segunda dose é de aproximadamente 0,0189%. O valor baixo sugere uma associação entre vacinação e taxas reduzidas de hospitalização.

Por fim, a fim de verificar a diferença entre as taxas médias de vacinação de faixas etárias de um país no segundo semestre de 2021.

As etapas foram:

1. Filtrar dados de dois grupos etários específicos (10-14 anos e 60-69 anos) entre 01-07-2021 e 31-12-2021 na Espanha.
2. Média das taxas de vacinação;
3. Boxplots para cada grupo;
4. Gráficos QQ.

Vacinados entre 10 e 14 anos:

```
count    27.000000
mean     25.629259
std      19.253010
min       0.010000
25%       0.505000
50%      38.480000
75%      41.910000
max      43.830000
```

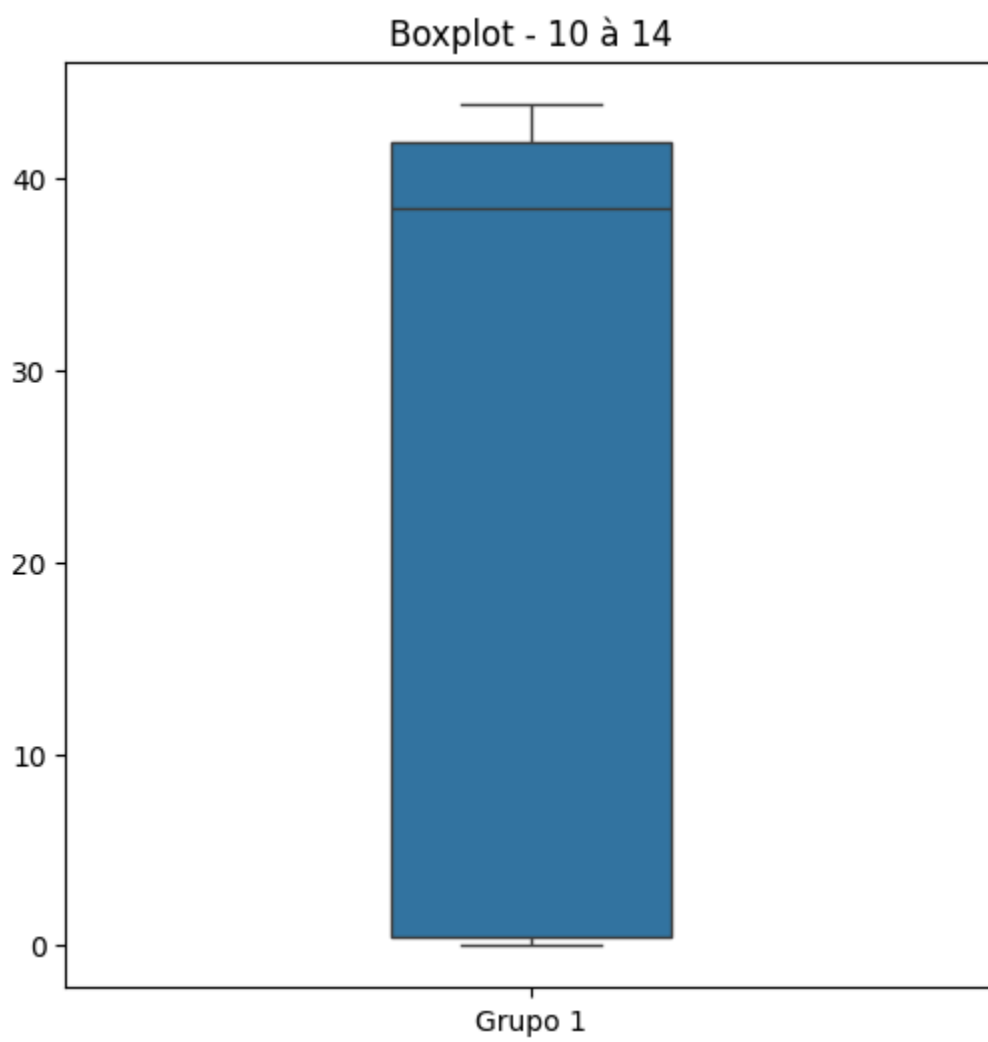
Name: people_fully_vaccinated_per_hundred, dtype: float64

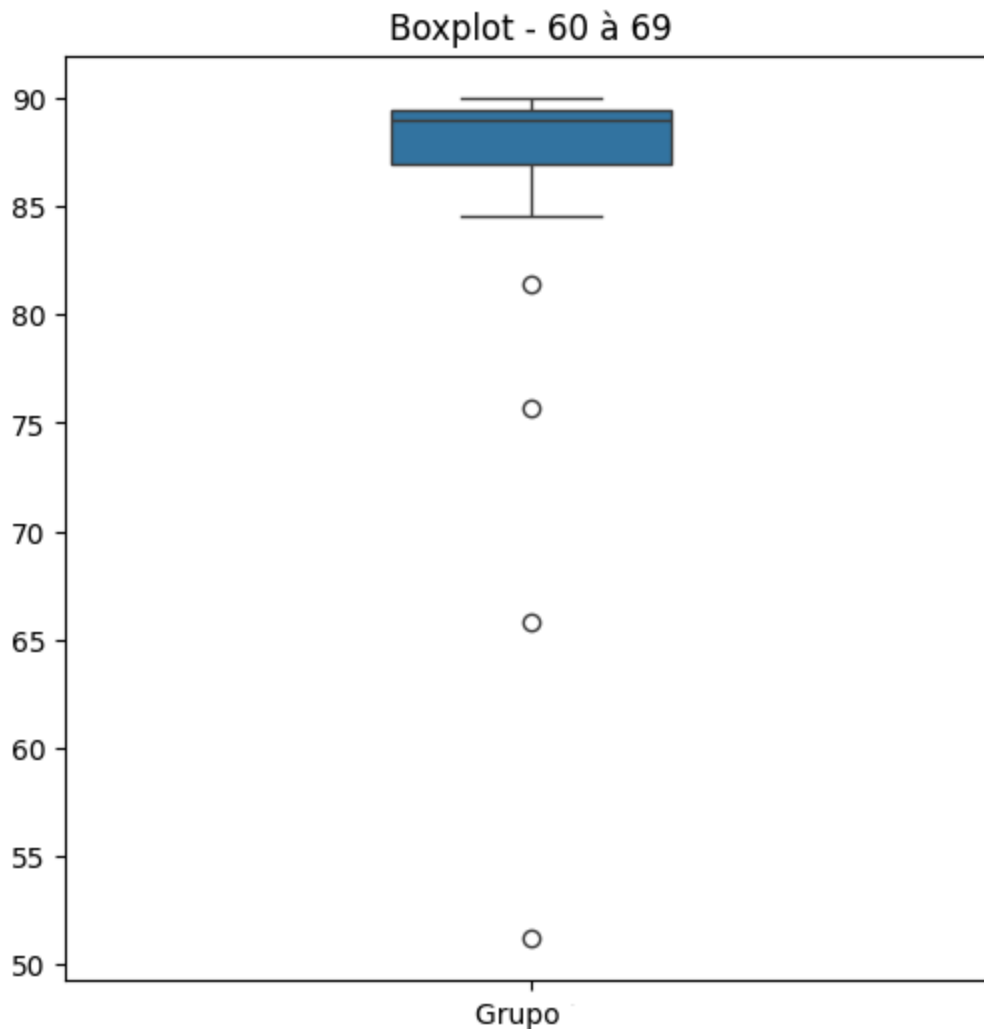
Vacinados entre 60 e 69 anos:

```
count    27.000000
mean     85.643704
std       8.640967
min      51.200000
25%      86.950000
50%      88.980000
75%      89.405000
max      89.970000
```

Name: people_fully_vaccinated_per_hundred, dtype: float64

Os box-plots ilustrando os grupos:





O box-plot do grupo de 10 à 14 anos mostra uma mudança estável da porcentagem de vacinados nessa faixa-etária, com o valor mínimo sendo 0, o máximo sendo 43,0% e no meio do período tendo o valor de 38%. Isso resultou em um box-plot com mediano próximo do topo, junto de bigodes bem próximos dos quartis.

Por outro lado os dados do grupo de 60 à 69 anos, mesmo mostrando também resultados positivos para a vacinação, deixa bem visível alguns outliers abaixo de 85. Tendo em conta que o valor mínimo foi 51,2%, mas na metade do período já estava em 88,98% percebemos um aumento brusco na porcentagem durante esse período, que logo depois ficou bem estável, chegando no valor máximo de quase 98,97%

Finalmente, gráficos QQ gerados para ambos os grupos, o que fornece uma orientação visual da normalidade.

