

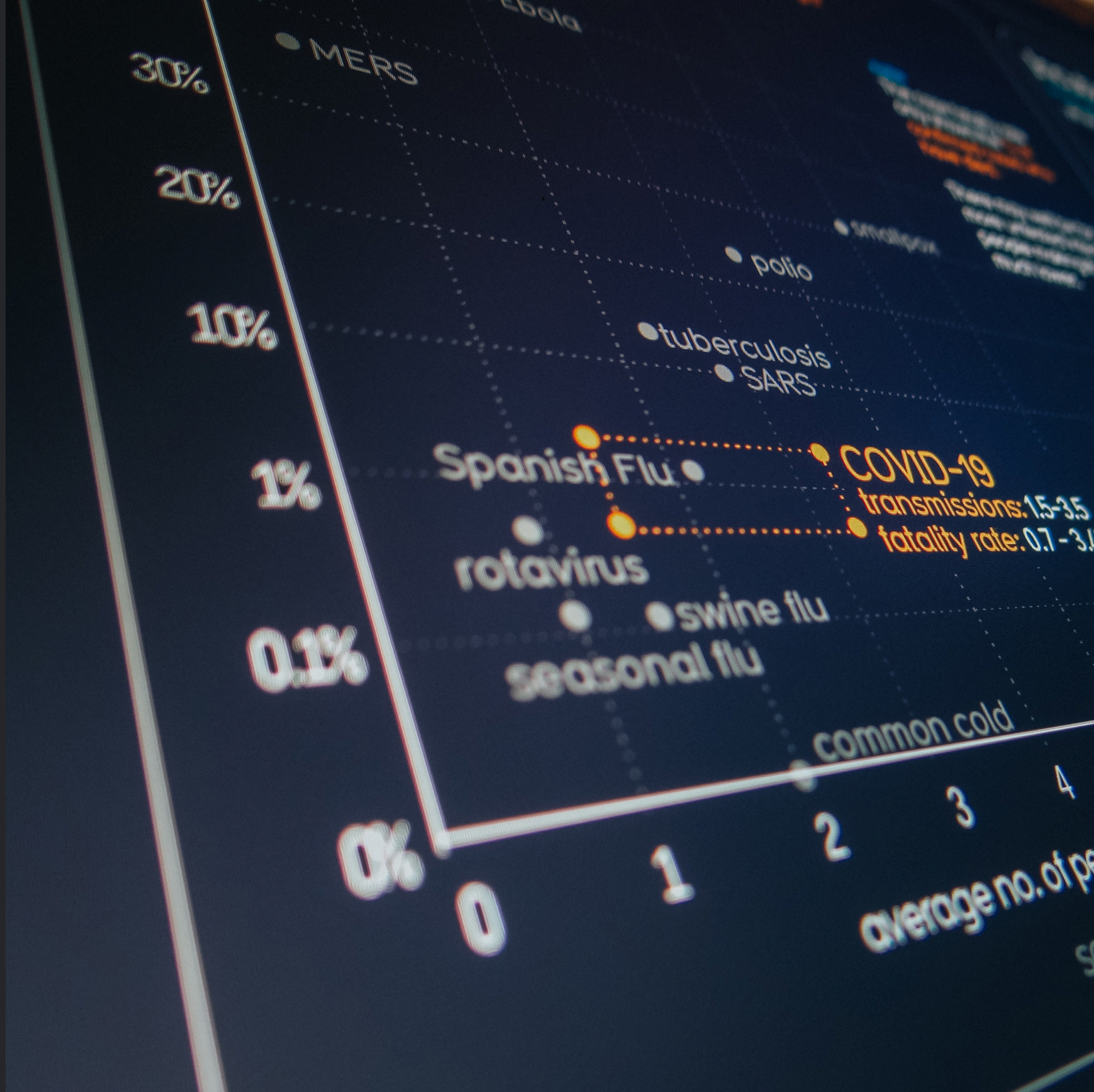
ELC1098 - MINERAÇÃO DE DADOS

Análise de Dados de Pacientes em UTIs: Aplicação do Clustering e Árvore de Decisão

DIEGO RIBEIRO CHAVES
LUCAS XAVIER PAIRÉ
PROF. DR. JOAQUIM V.C. ASSUNÇÃO

Motivação

Analisar dados de pacientes internados em UTIs para identificar padrões relacionados ao risco de óbito e melhorar a alocação de recursos.



Metodologia

Origem dos dados

Dados do BRATECA, advindos de 10 hospitais brasileiros.

Origem dos dados

Dados do BRATECA, advindos de 10 hospitais brasileiros.

Pré-Processamento

Inicialmente, os dados estavam em 3 *datasets*, contendo informações acerca de internação, exames e prescrições.

Origem dos dados

Dados do BRATECA, advindos de 10 hospitais brasileiros.

Pré-Processamento

Inicialmente, os dados estavam em 3 *datasets*, contendo informações acerca de internação, exames e prescrições.

Os dados continham informações duplicadas, pois um mesmo paciente fazia o mesmo exame várias vezes.

Origem dos dados

Dados do BRATECA, advindos de 10 hospitais brasileiros.

Pré-Processamento

Inicialmente, os dados estavam em 3 *datasets*, contendo informações acerca de internação, exames e prescrições.

Os dados continham informações duplicadas, pois um mesmo paciente fazia o mesmo exame várias vezes.

A agregação foi feita por paciente e hospital, utilizando funções lógicas, aritméticas e estatísticas.

Origem dos dados

Dados do BRATECA, advindos de 10 hospitais brasileiros.

Pré-Processamento

Inicialmente, os dados estavam em 3 *datasets*, contendo informações acerca de internação, exames e prescrições.

Os dados continham informações duplicadas, pois um mesmo paciente fazia o mesmo exame várias vezes.

A agregação foi feita por paciente e hospital, utilizando funções lógicas, aritméticas e estatísticas.

Os motivos de alta foram remapeados.

Os exames foram pivotados.

Os valores faltantes foram calculados.

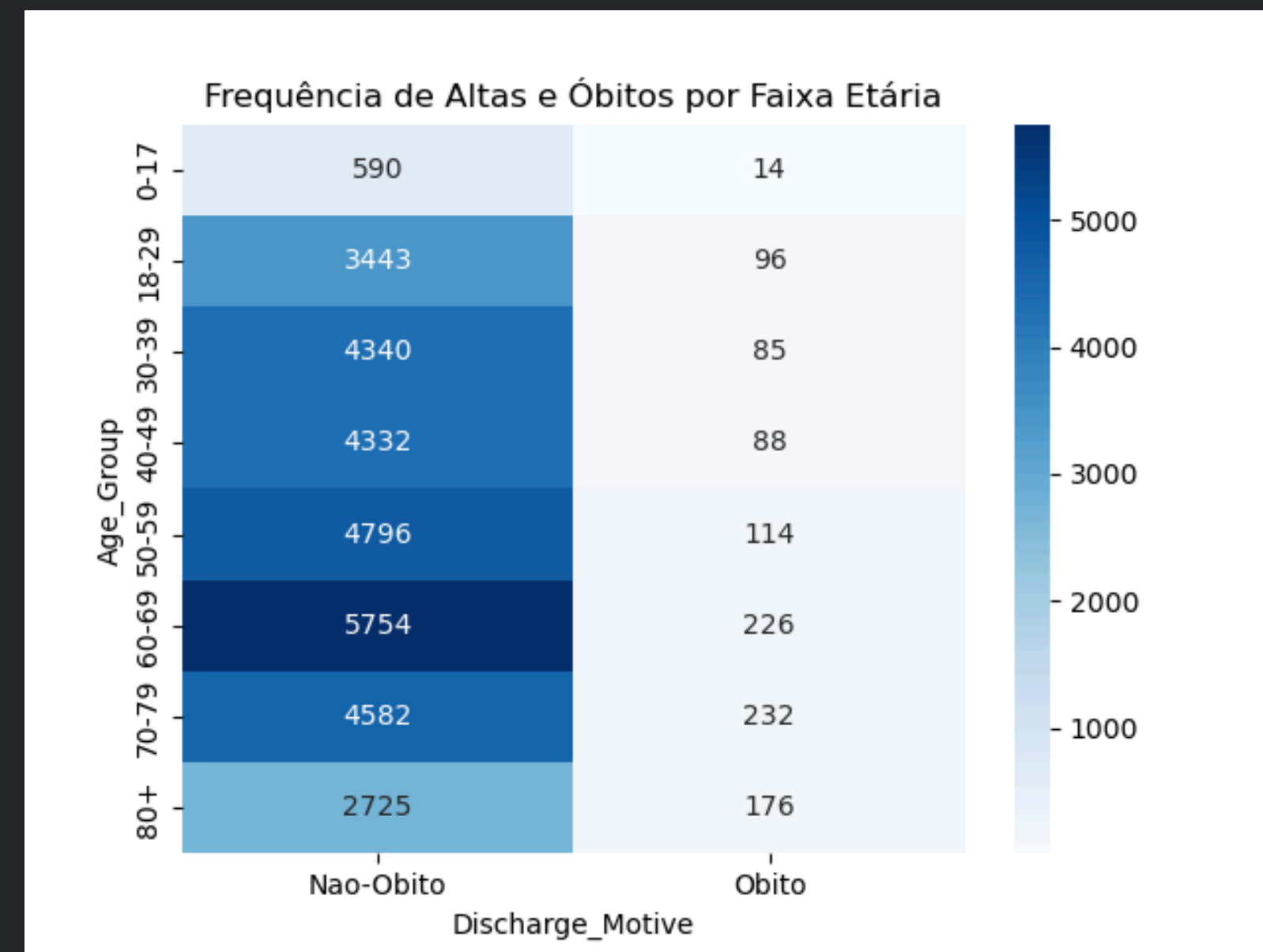
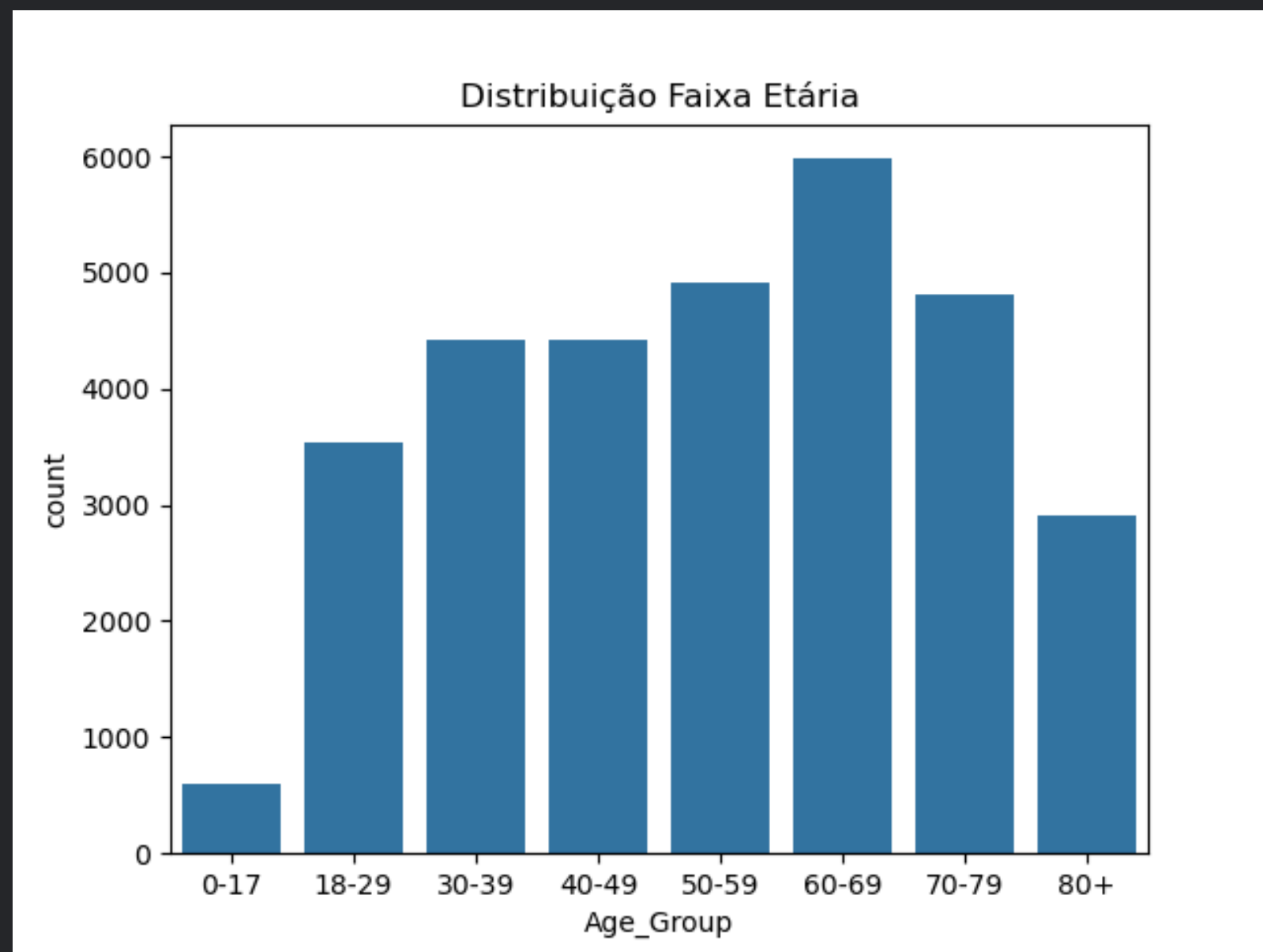
Ao término do pré-processamento

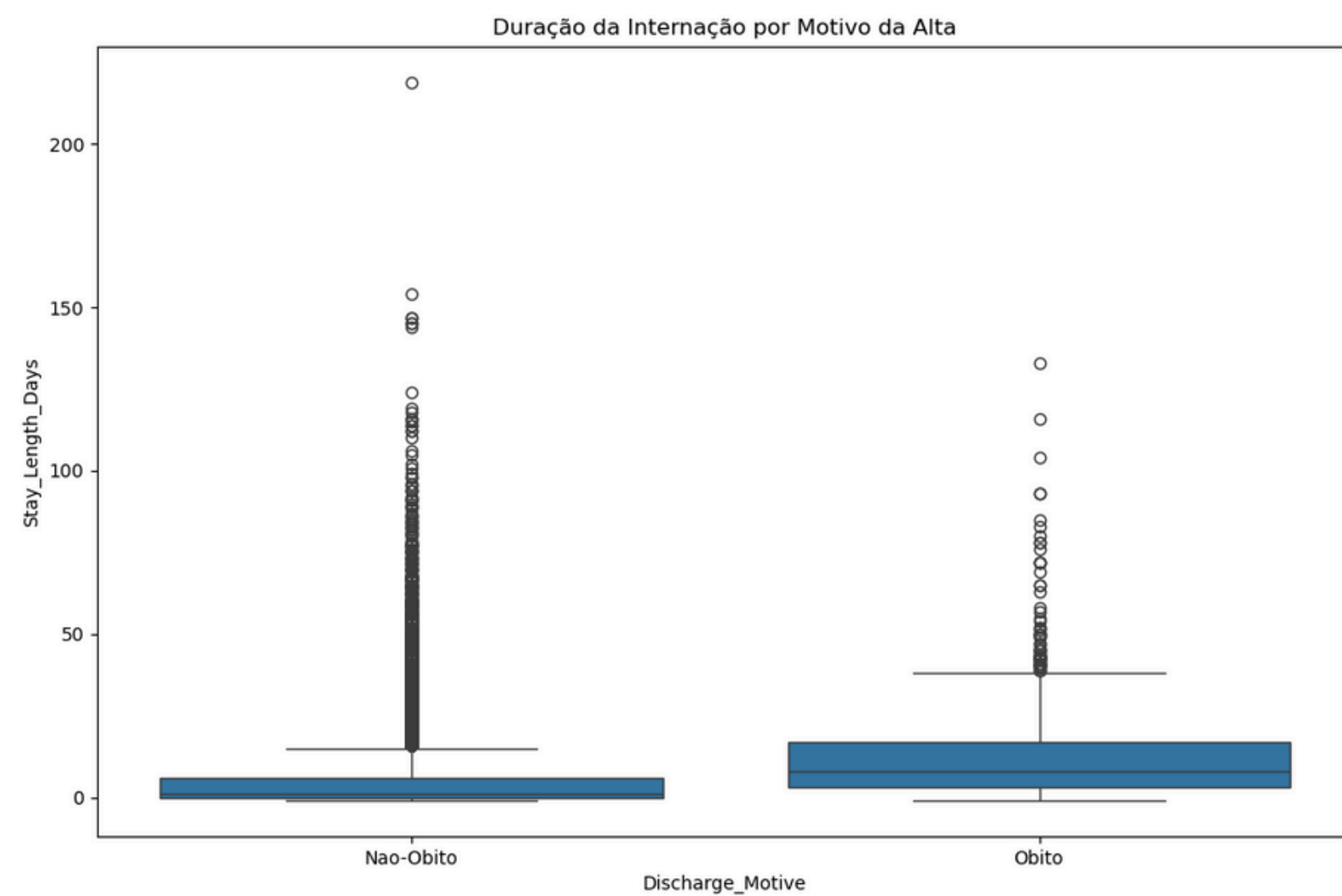
Cada linha representava um paciente, com todos os valores de exames e demais informações preenchidas. O motivo de alta estava remapeado, e tinha valor binário: óbito ou não-óbito.

Análise exploratória dos dados

Focou-se em buscar relações entre as variáveis e o desfecho do paciente.







Ao término da análise exploratória dos dados

Não foi possível encontrar relações simples, indicando uma maior complexidade de interação entre as variáveis.

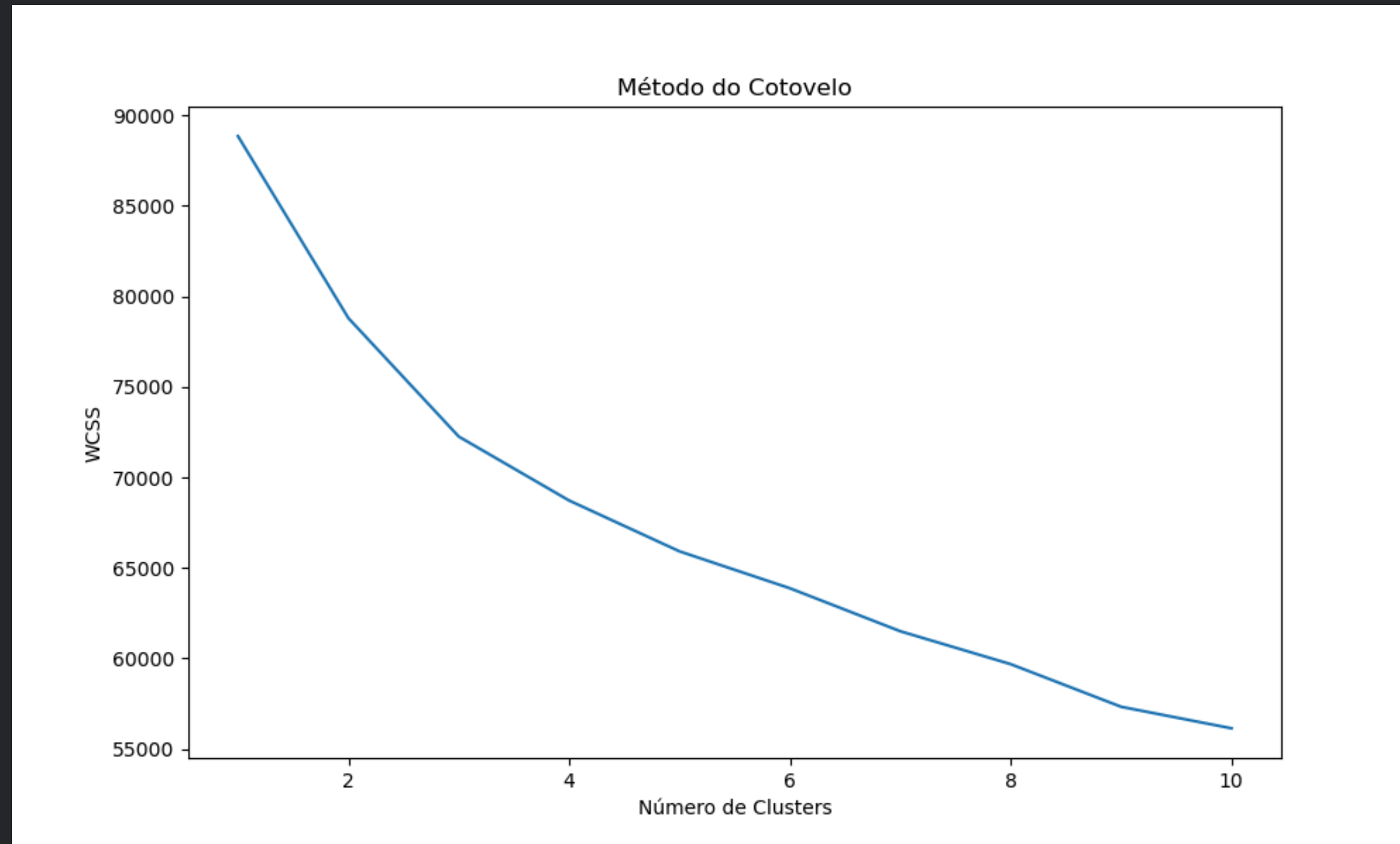
Ao término da análise exploratória dos dados

Não foi possível encontrar relações simples, indicando uma maior complexidade de interação entre as variáveis.

Por essa razão, decidiu-se empregar técnicas mais sofisticadas, nesse caso, *clustering* e árvore de decisão.

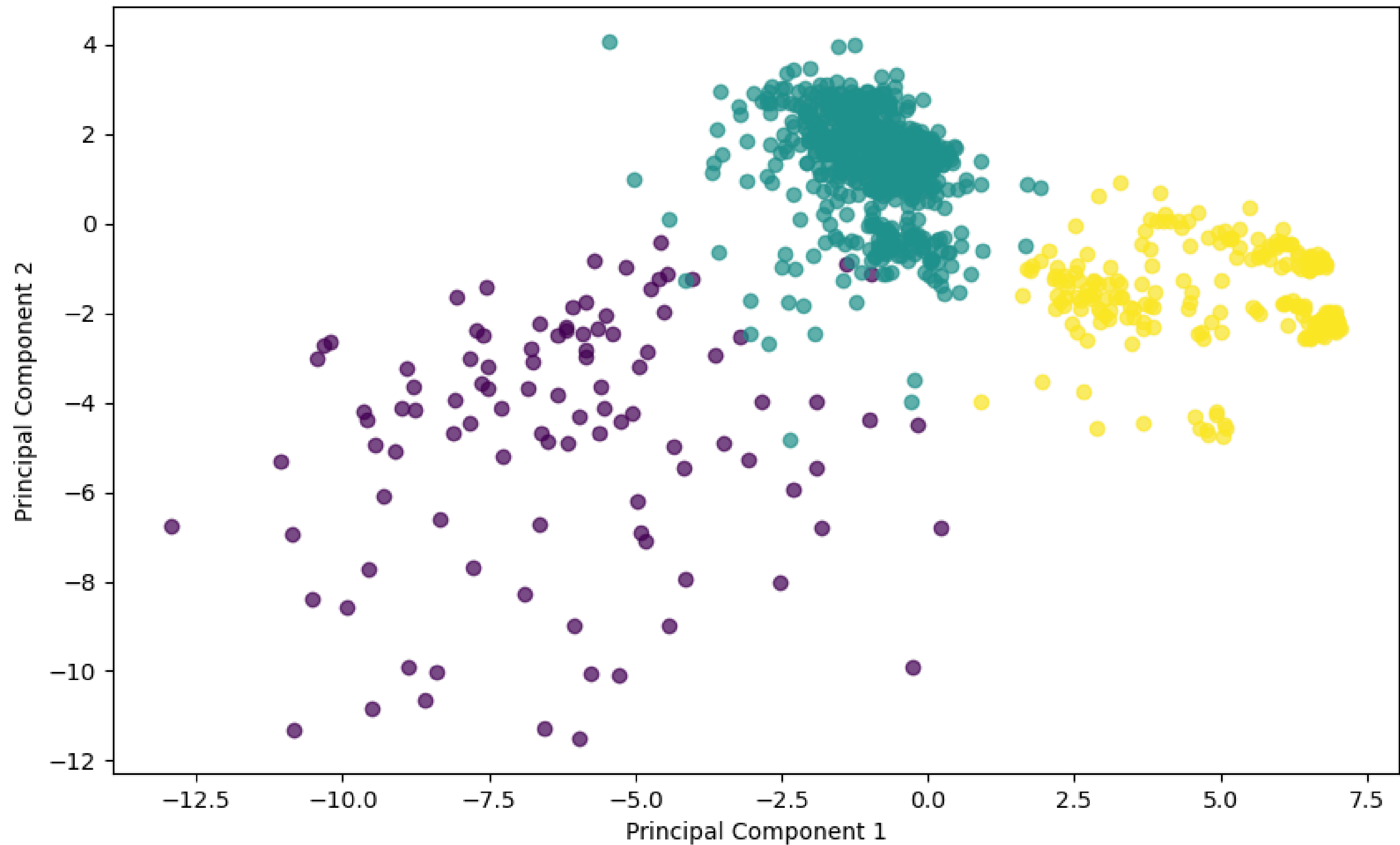
Decidiu-se focar os esforços em pacientes que vieram a óbito, devido a importância desses casos para os hospitais.

Clustering - determinando o número de *clusters*



WCSS - Within-Cluster Sum of Squares

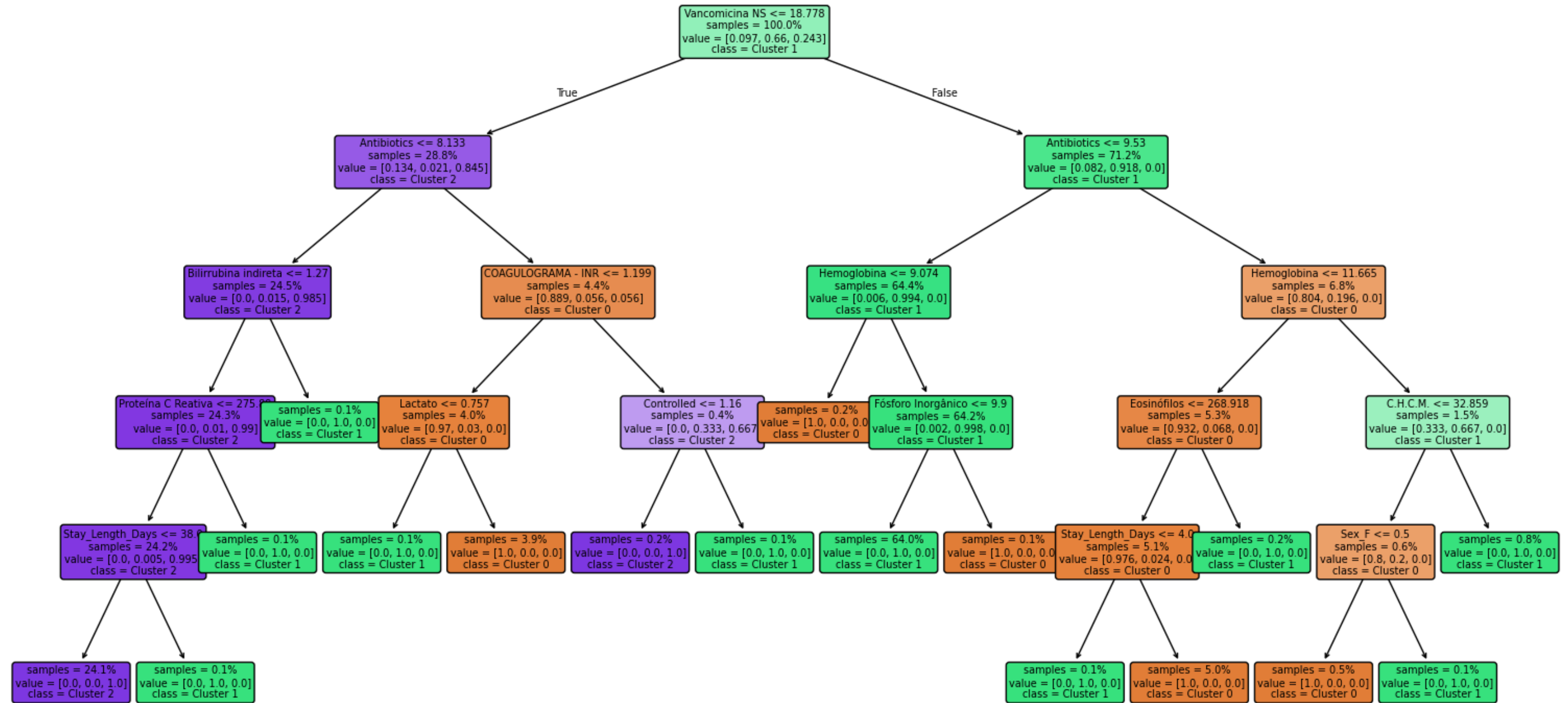
Clustering dos Pacientes que Vieram a Óbito



Avaliação do *clustering*

Foi possível observar 3 classes, sendo duas delas mais bem definidas e uma terceira um pouco menos definida.

A partir disso, decidiu-se investigar quais características tinham maior influência na classificação de um paciente em uma classe.



Avaliação da árvore de decisão

Foi possível constatar que os principais fatores são os resultados dos exames, seguido pelo tempo de permanência.

Isso indica a importância da avaliação dos exames laboratoriais no contexto hospitalar.

Avaliação da árvore de decisão

E como ficaram as métricas do modelo?

Classe	Precisão	<i>Recall</i>	<i>F1-Score</i>	Suporte
0	0.91	0.87	0.89	23
1	0.98	0.99	0.99	133
2	1.00	0.98	0.99	51

Problemas e possíveis soluções

O principal problema foi o desbalanceamento dos dados.

Problemas e possíveis soluções

O principal problema foi o desbalanceamento dos dados.

Para trabalhar com isso, seria interessante utilizar técnicas de balanceamento por *oversampling* ou *undersampling*, como *SMOTE*.

SMOTE possibilita a criação de exemplos sintéticos da classe minoritária.

O que podemos concluir

As técnicas utilizadas, combinadas com a análise visual e estatística dos dados, proporcionaram uma compreensão valiosa sobre os fatores que influenciam os desfechos dos pacientes, especificamente os que foram a óbito.

O bom desempenho do modelo de árvore de decisão, apesar de algumas dificuldades com a classe minoritária, confirma que os dados contêm informações relevantes para prever o risco de óbito.

O que podemos concluir

Esse estudo contribui para o entendimento dos fatores de risco associados ao óbito em pacientes críticos, oferecendo uma base para o desenvolvimento de ferramentas preditivas em ambientes hospitalares, o que pode melhorar o cuidado e a tomada de decisões clínicas.

Para trabalhos futuros, seria interessante explorar técnicas de balanceamento de classes para melhorar a performance do modelo, além de investigar outras abordagens de aprendizado de máquina mais avançadas, como *Random Forests* e *XGBoost*.

Obrigado!