

Case

Mining Process Quality Prediction

Apresentação da Análise Exploratória de
Dados (EDA) e Modelagem

Introdução

O processo de flotação, na indústria de mineração, é utilizado para concentrar o minério de ferro

A porcentagem de Ferro na entrada do processo é providenciado pelo planejamento de Lavra

Além do minério desejando, também, são encontrados impurezas, como Sílica

Assim, o objetivo do processo de flotação é aumentar a concentração de Ferro e remover as impurezas

OBJETIVO DO PROJETO

Prever a % Silica Concentrate (sílica no concentrado).

- O resultado da qualidade do concentrado é realizado no Laboratório.
- O que significa que os engenheiros de processo levam pelo menos uma hora para obter esse valor.
- A predição da impureza permitirá que a tomada de decisão seja realizada antes que o resultado oficial seja disponibilizado.

Hipóteses Iniciais (Case Kaggle)

É possível prever a % de concentrado de sílica a cada minuto?

Com quantas horas de antecedência podemos prever a % de sílica no concentrado?

Isso ajudaria os engenheiros a agir de forma preditiva e otimizada, mitigando a % de ferro que poderia ter ido para os rejeitos?

É possível prever a % de sílica no concentrado sem usar a coluna % de concentrado de ferro

já que elas são altamente correlacionadas

“Sabe quais decisões mudaria se tivesse os dados certos para o ajudar?” (Jeff Winter)

Hipóteses Definidas

Em que condição o processo está operando?

Qual a condição operacional desejada? Quais parâmetros de processo devem ser alterados?

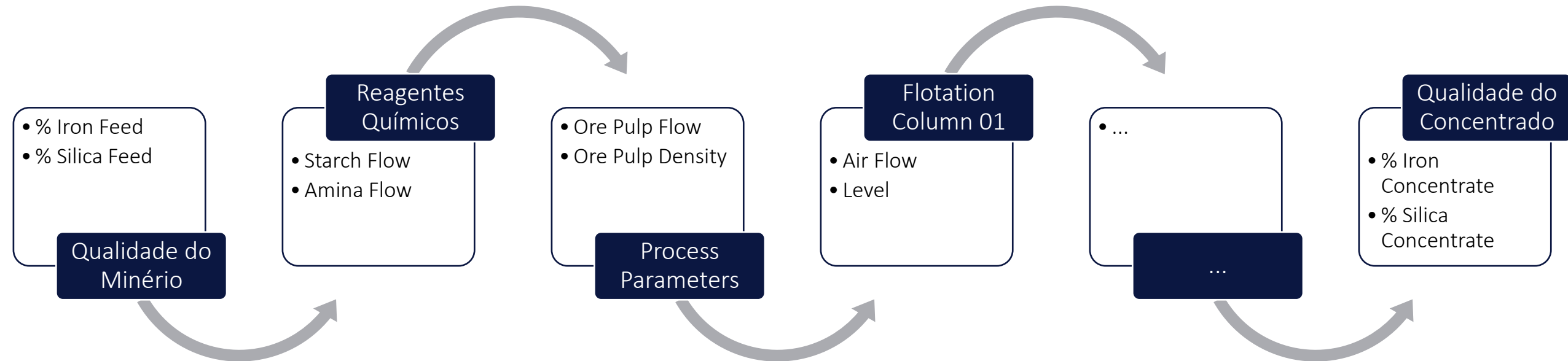
Como um modelo preditivo pode auxiliar nesta tomada de decisão?

É possível prever a % de sílica no concentrado sem usar a coluna % de concentrado de ferro

já que elas são altamente correlacionadas

“Sabe quais decisões mudaria se tivesse os dados certos para o ajudar?” (Jeff Winter)

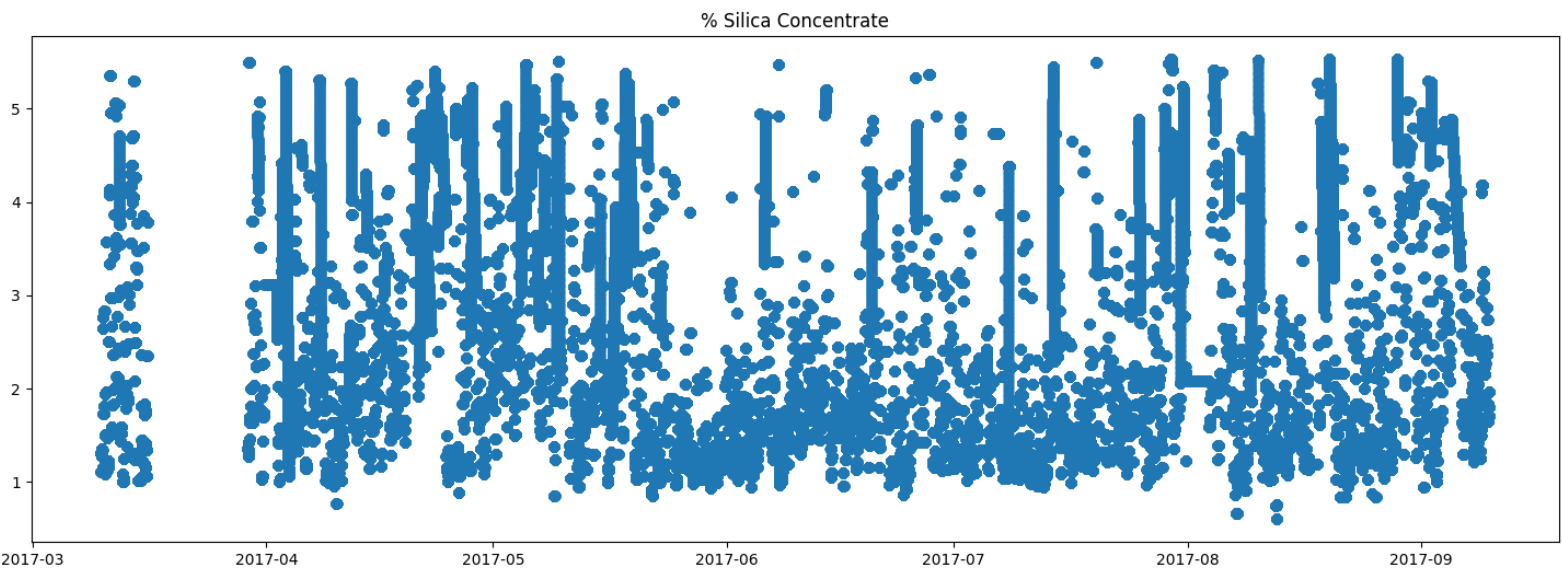
Variáveis (Features)



Necessidade de normalização dos dados na linha do tempo

- Alguns dados foram amostrados a cada 20 segundos
- Outros foram amostrados de hora em hora

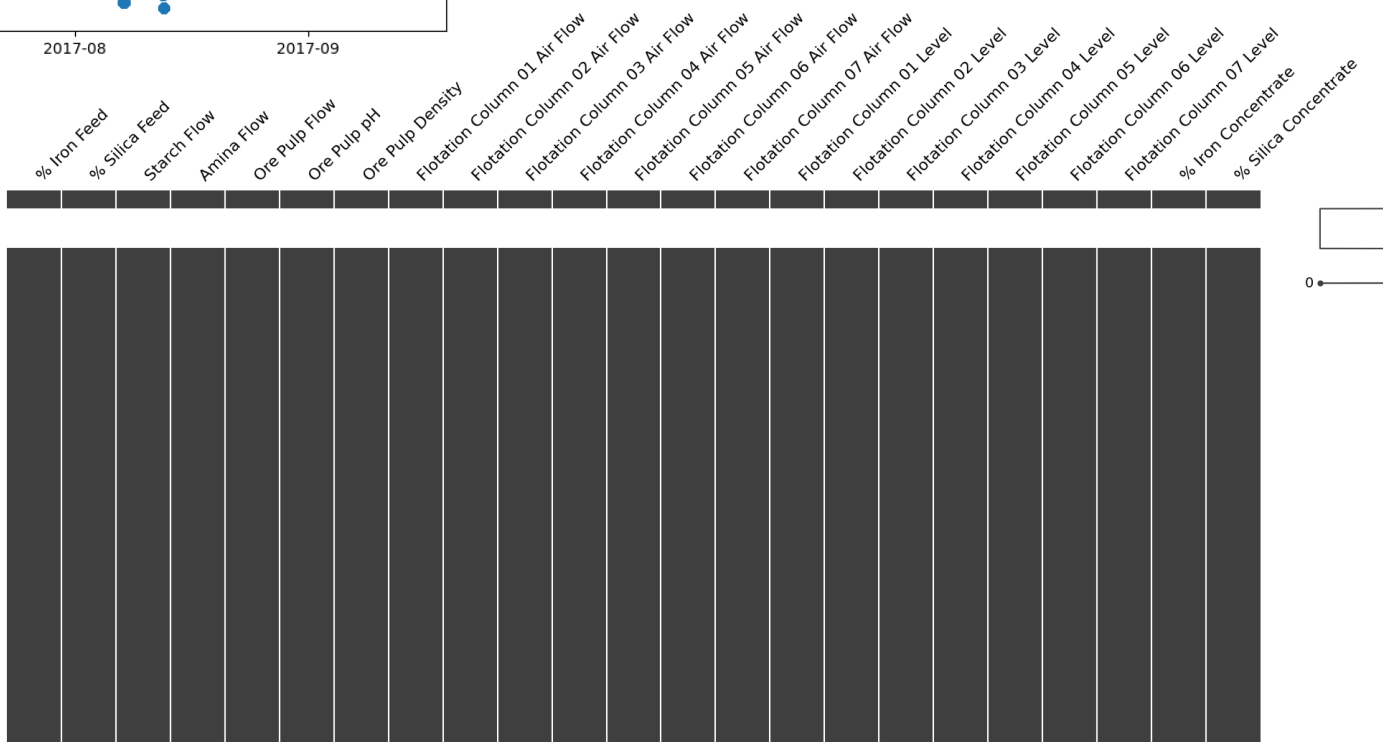
EDA (Análise Exploratória de Dados)



Target data
Timeline

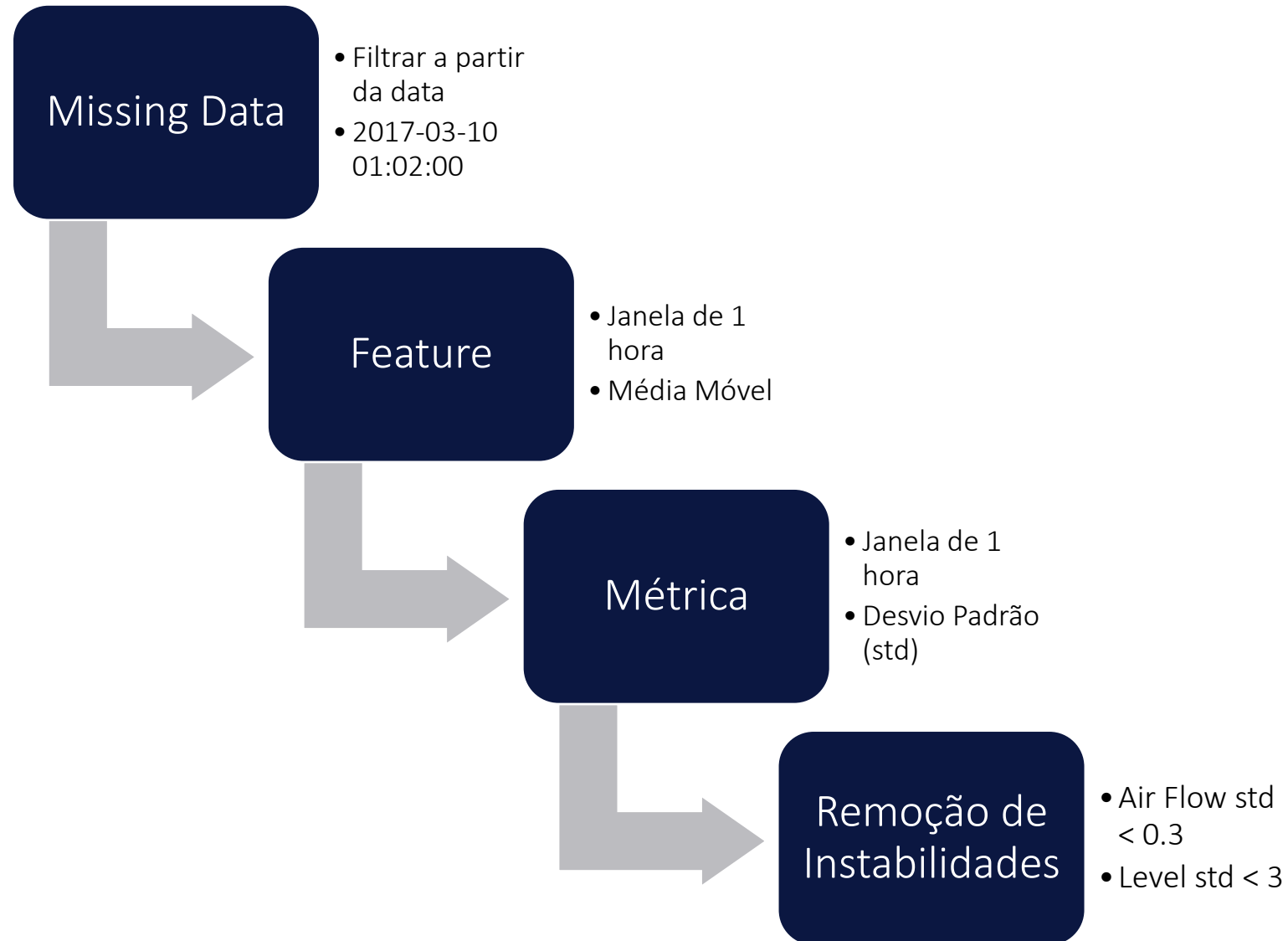
Missing Data

Filtrar a partir da data
2017-03-10
01:02:00



EDA (Análise Exploratória de Dados)

Tratamento dos dados

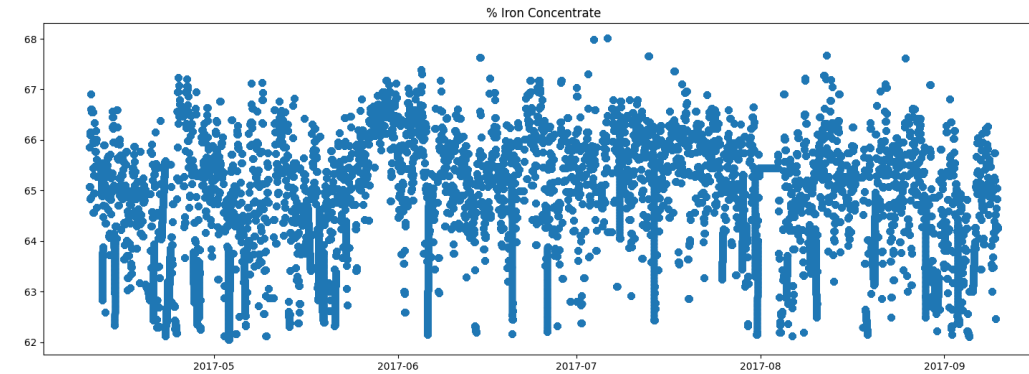
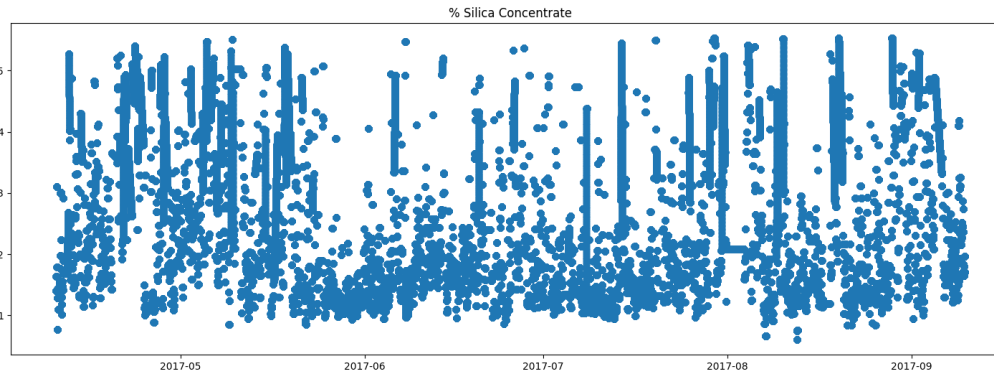


Objetivo:

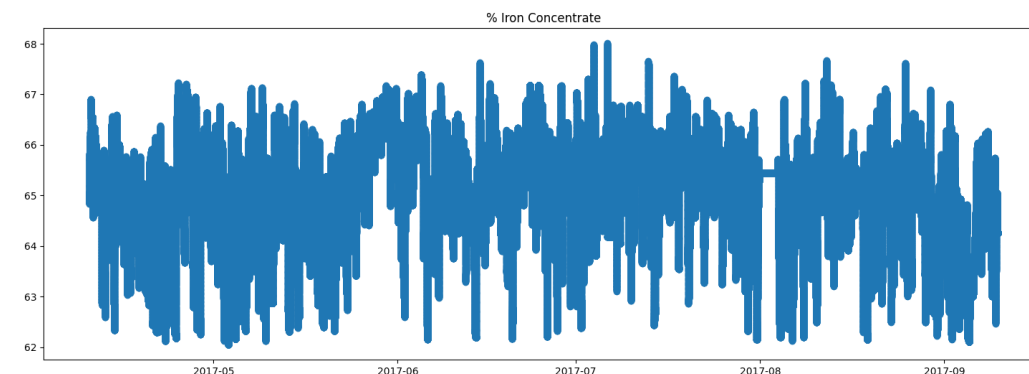
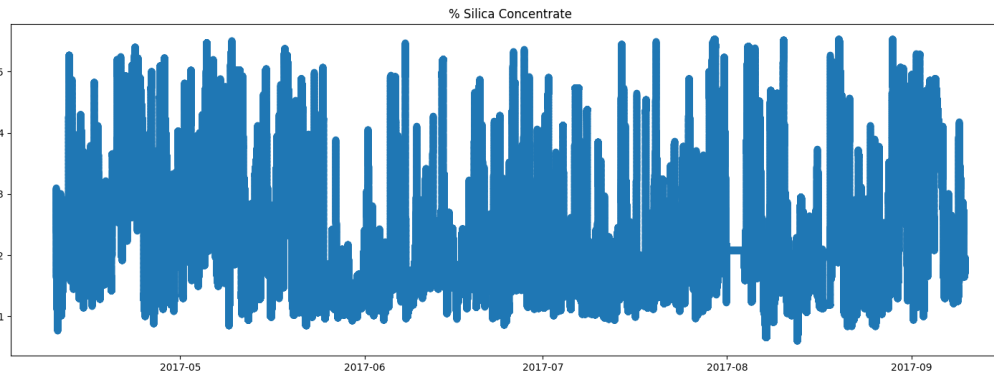
Obter dados que representem condições de operação estável

EDA (Análise Exploratória de Dados)

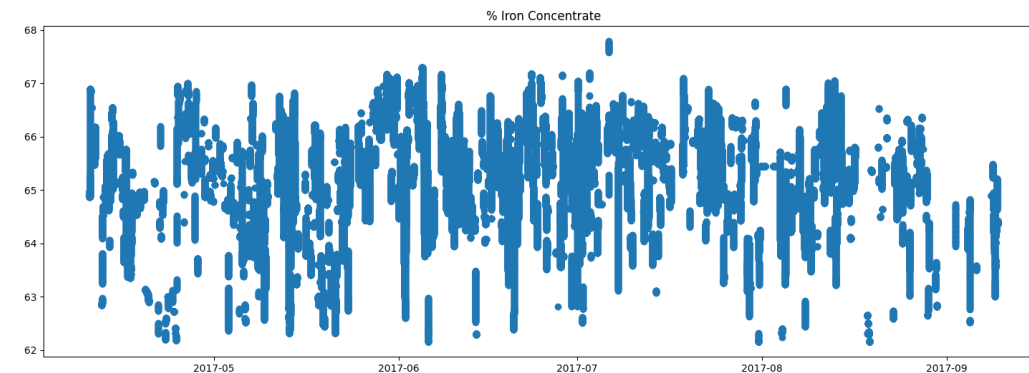
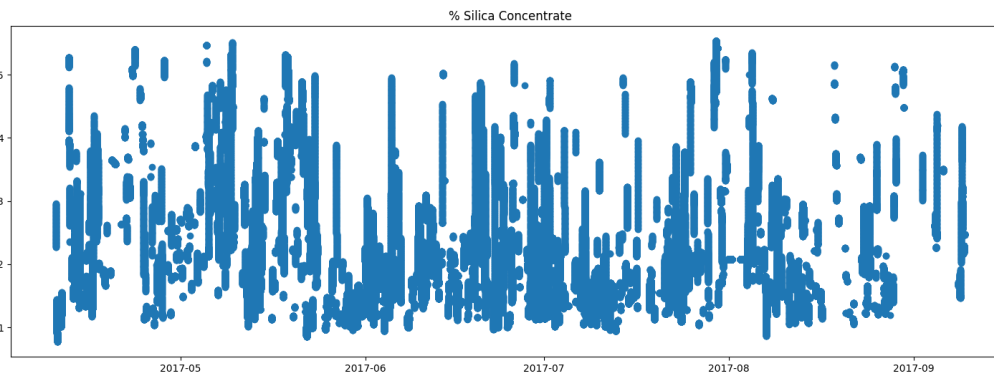
Raw Data



Feature

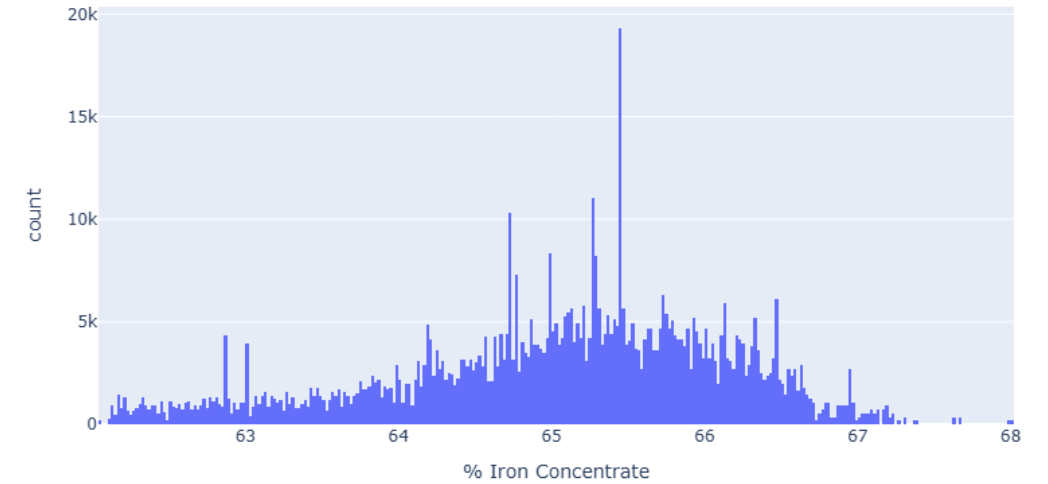
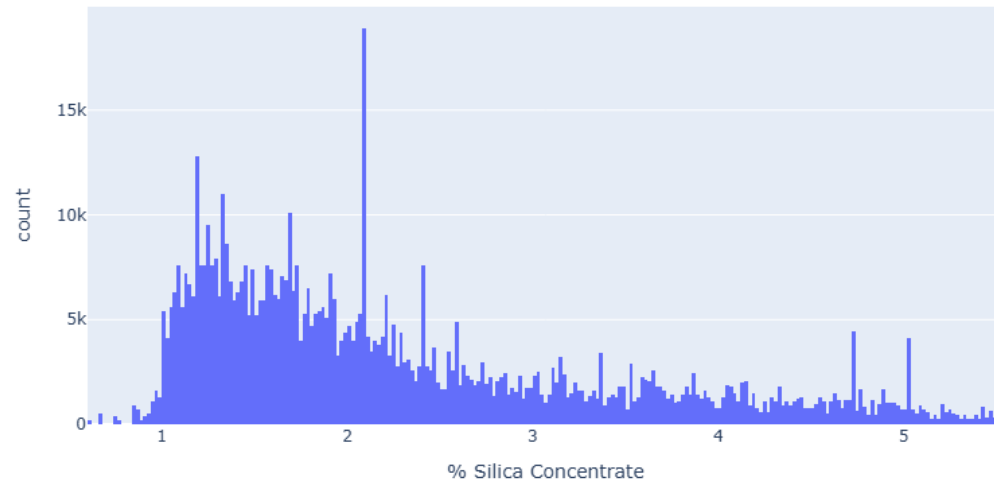


Dado Filtrado

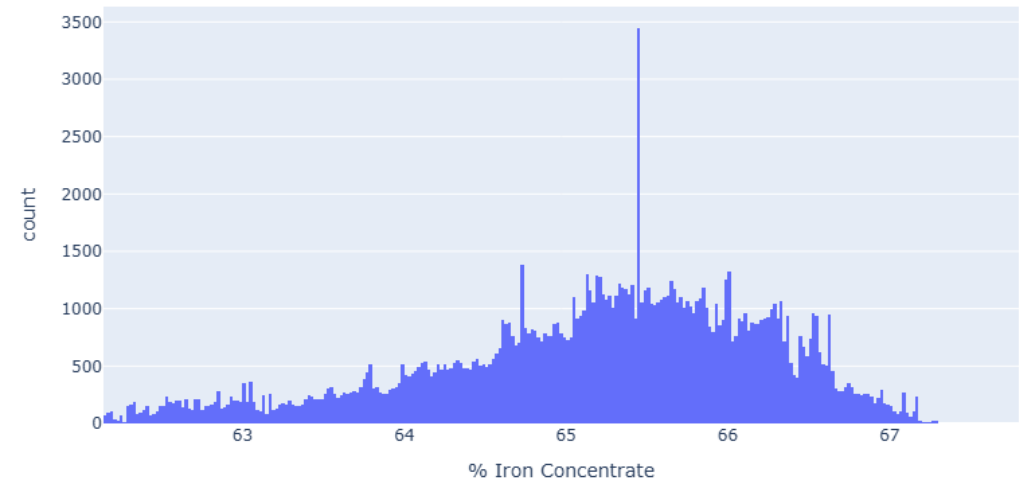
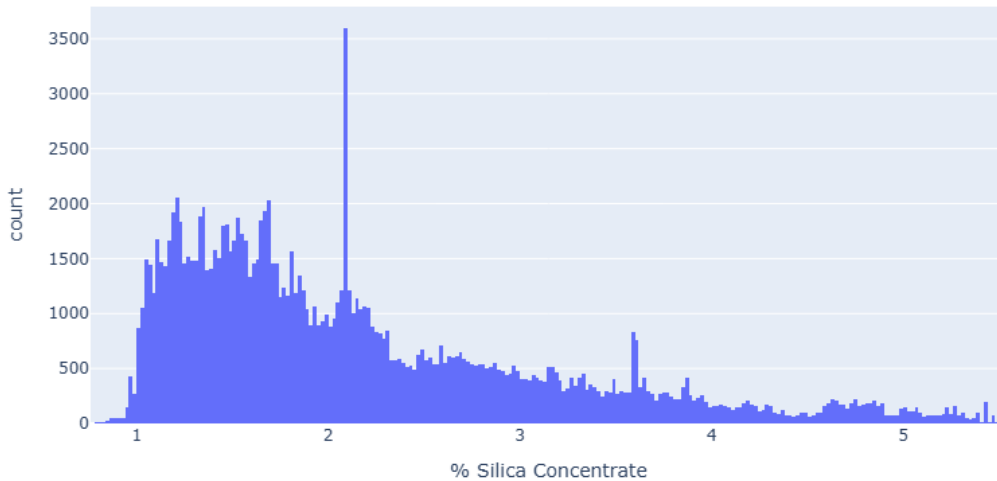


EDA (Análise Exploratória de Dados)

Raw Data

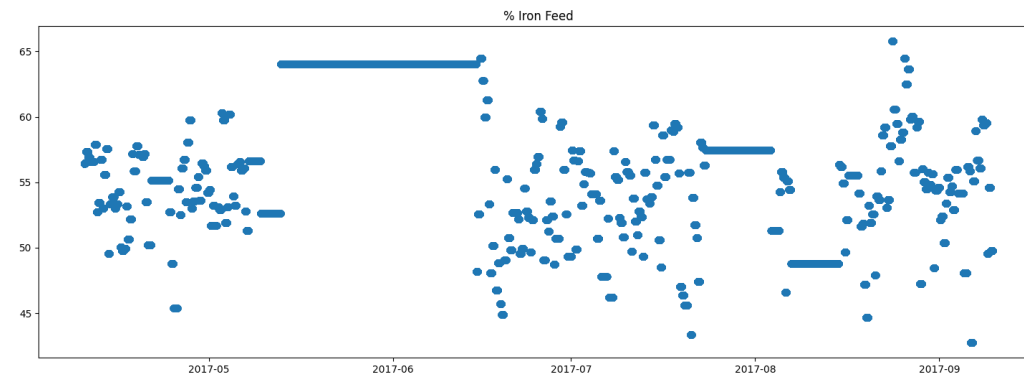
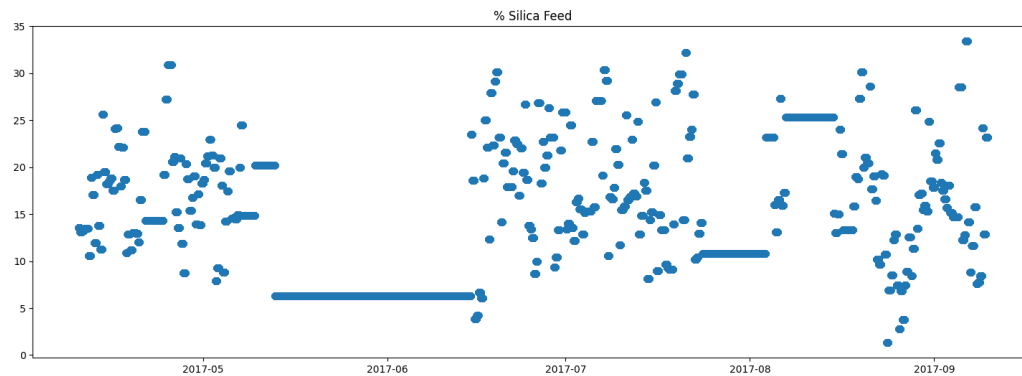


Dado
Filtrado

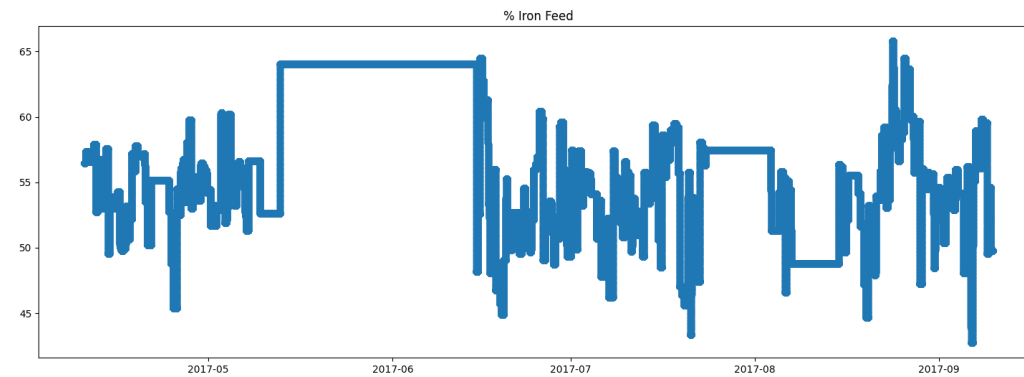
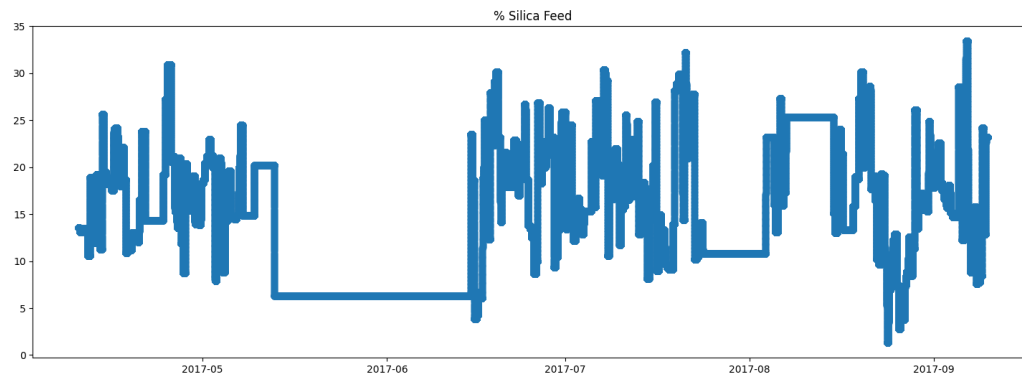


EDA (Análise Exploratória de Dados)

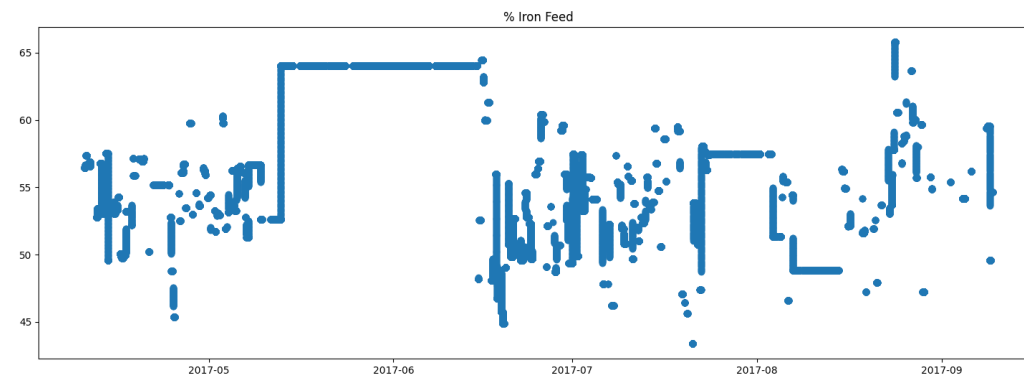
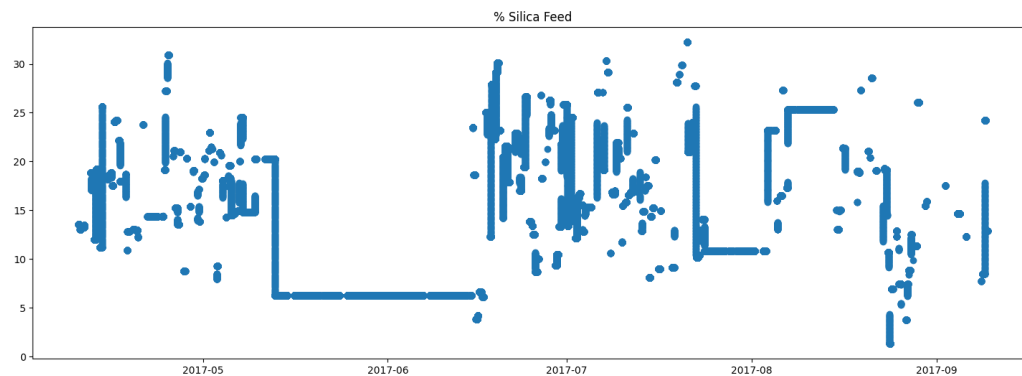
Raw Data



Feature

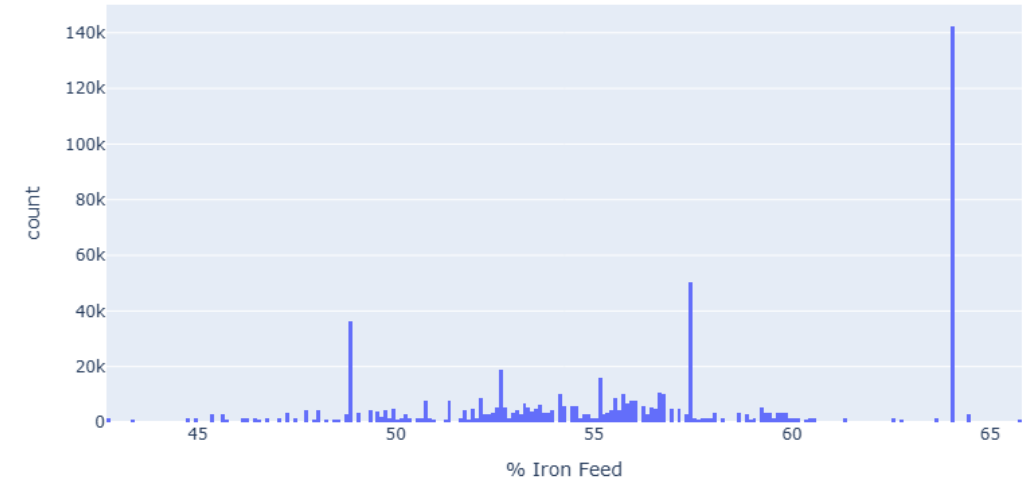
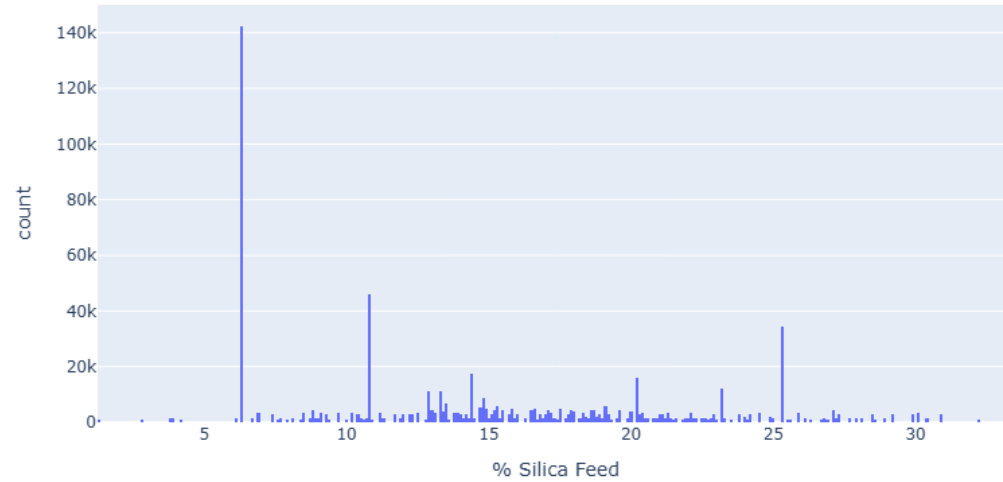


Dado Filtrado

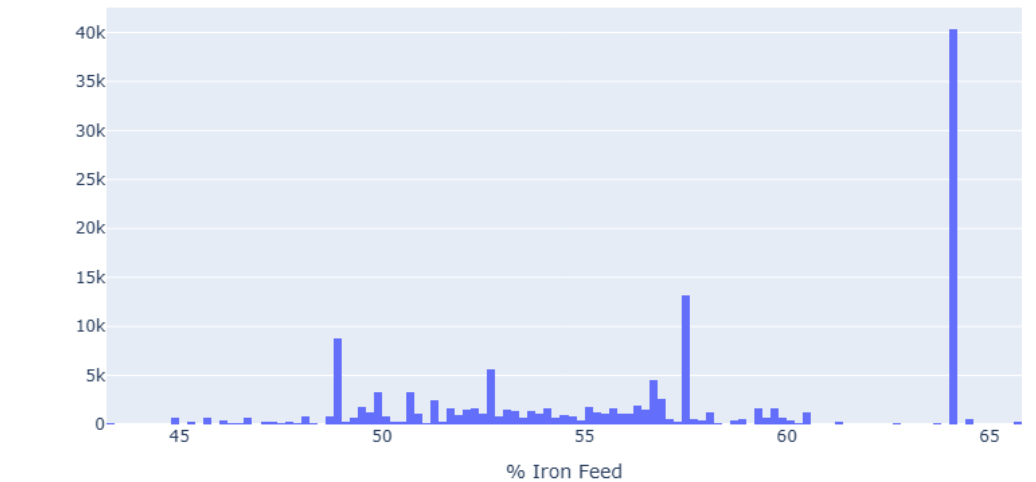
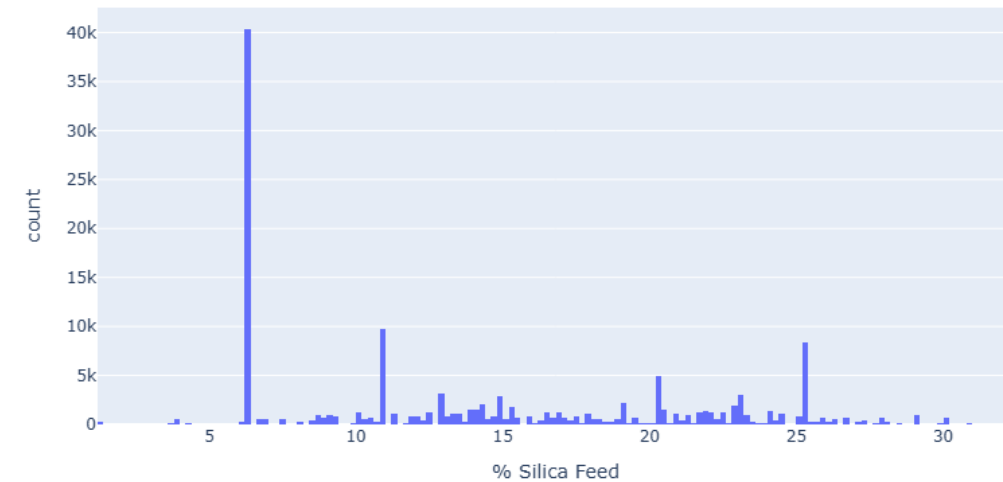


EDA (Análise Exploratória de Dados)

Raw Data

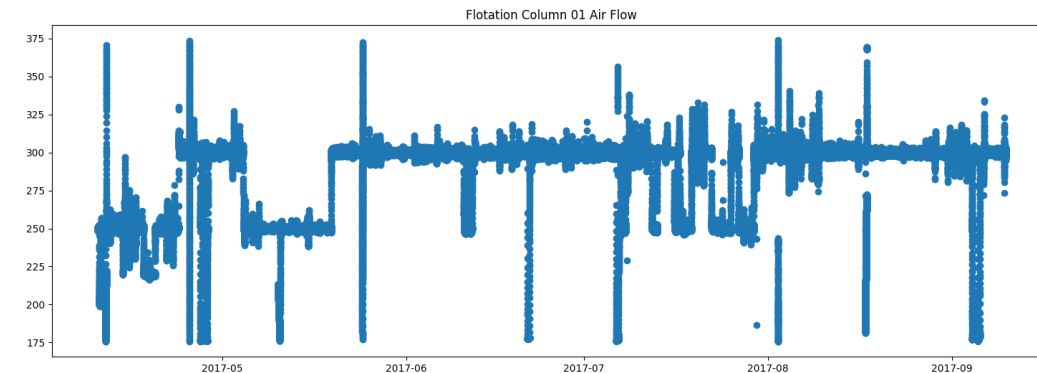
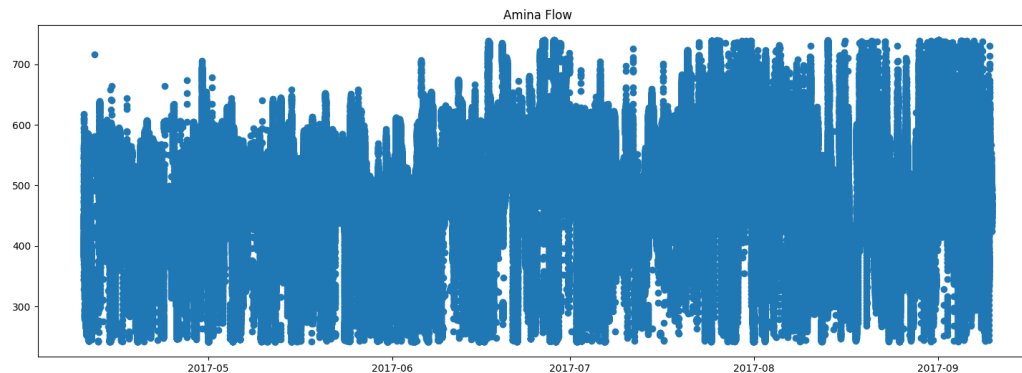


Dado
Filtrado

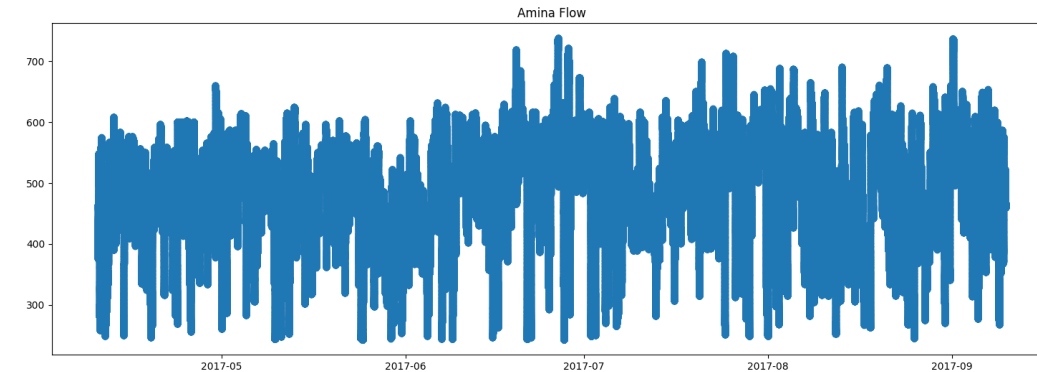
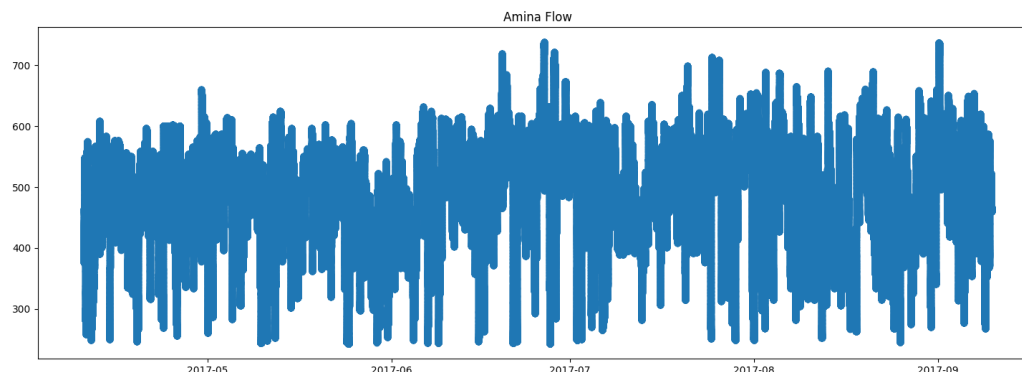


EDA (Análise Exploratória de Dados)

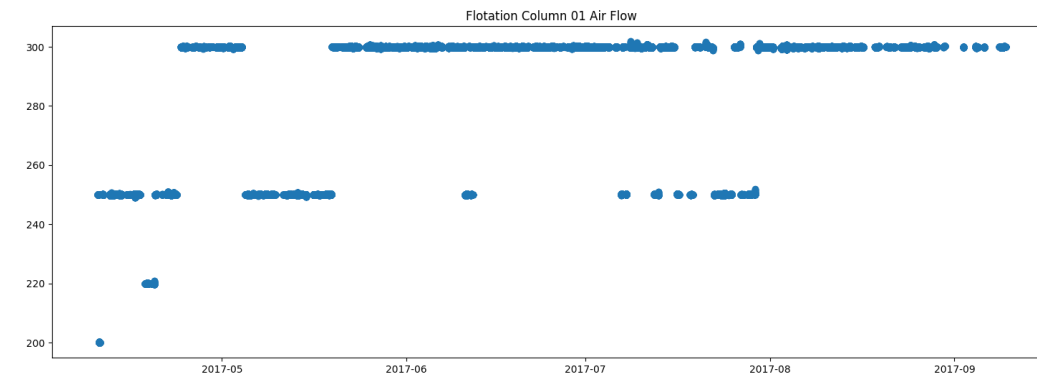
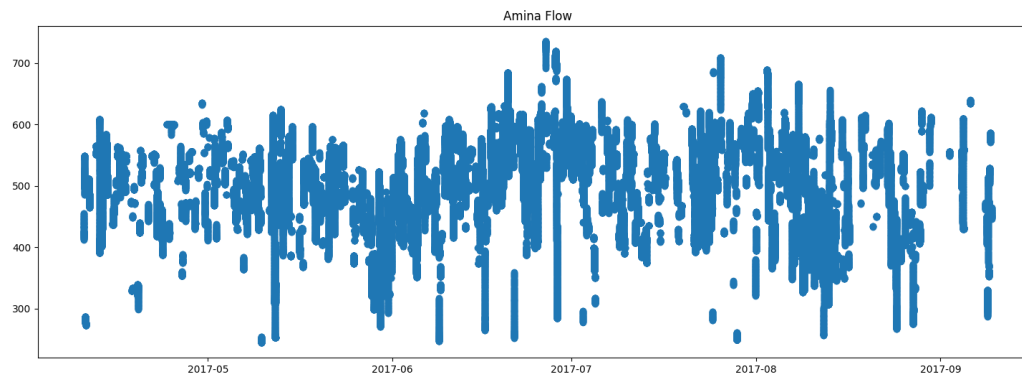
Raw Data



Feature

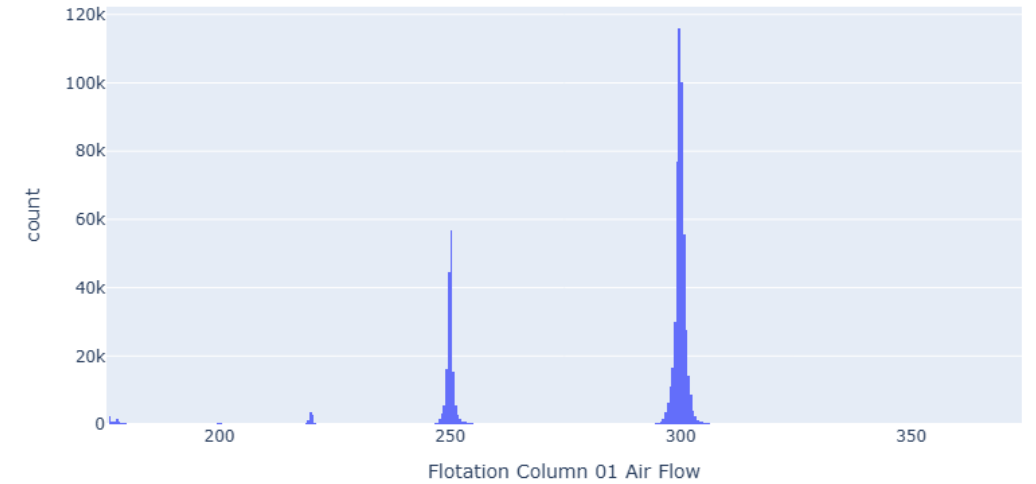
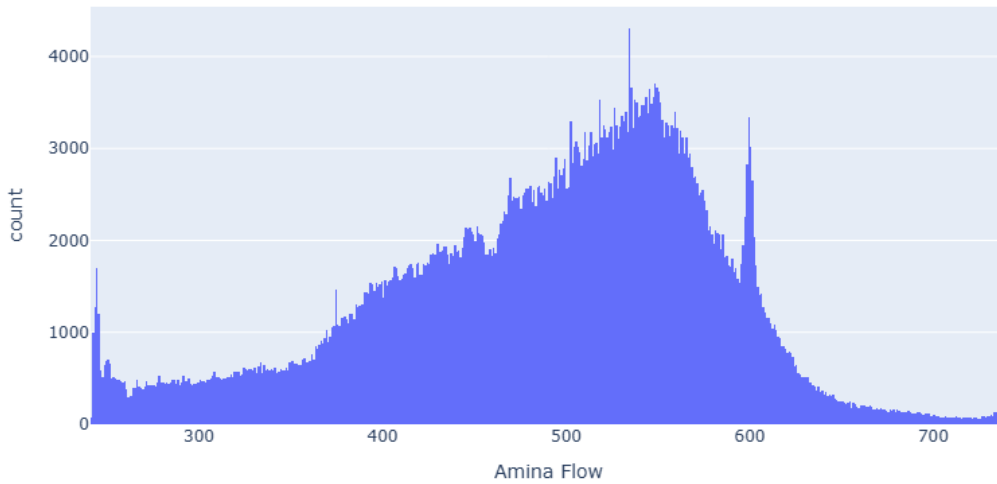


Dado Filtrado

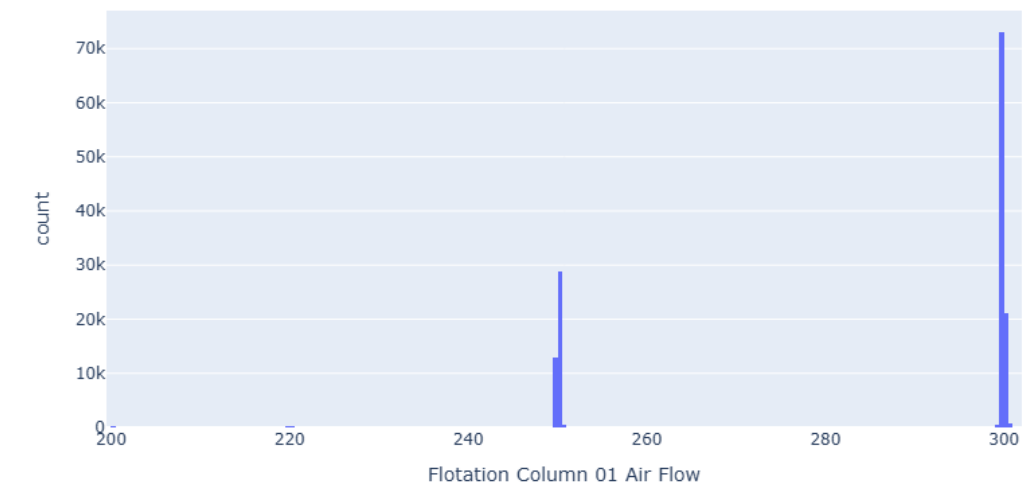
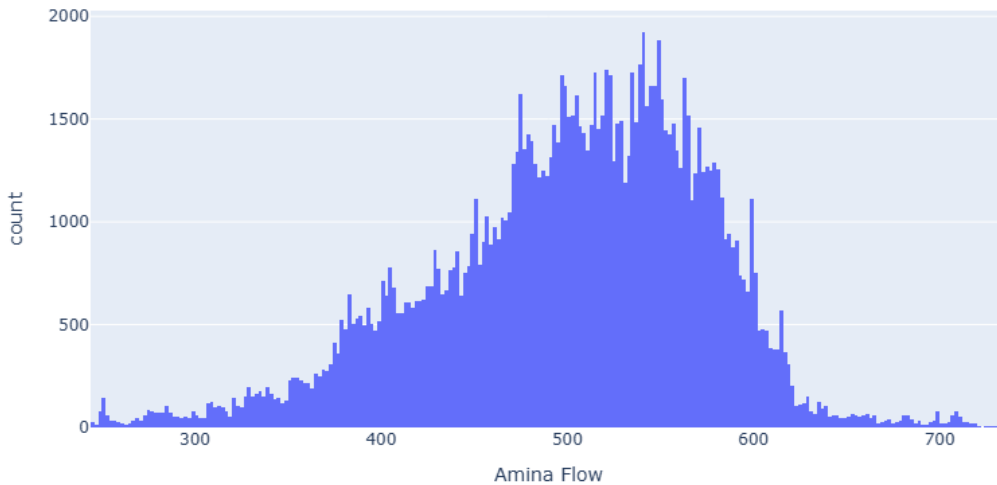


EDA (Análise Exploratória de Dados)

Raw Data

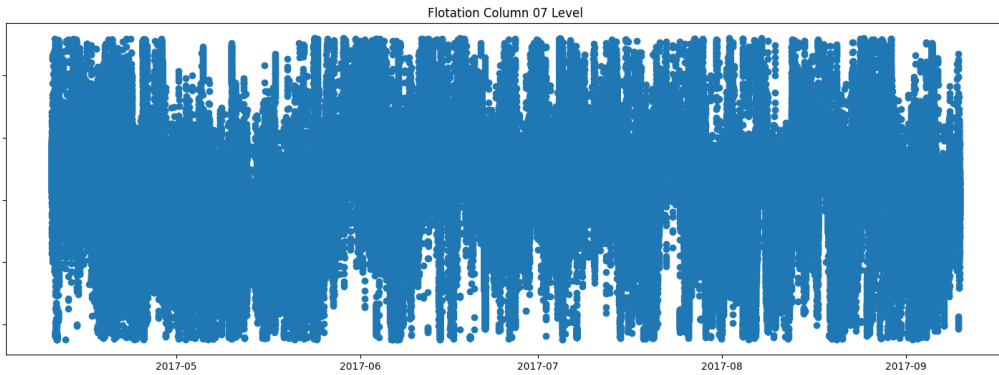
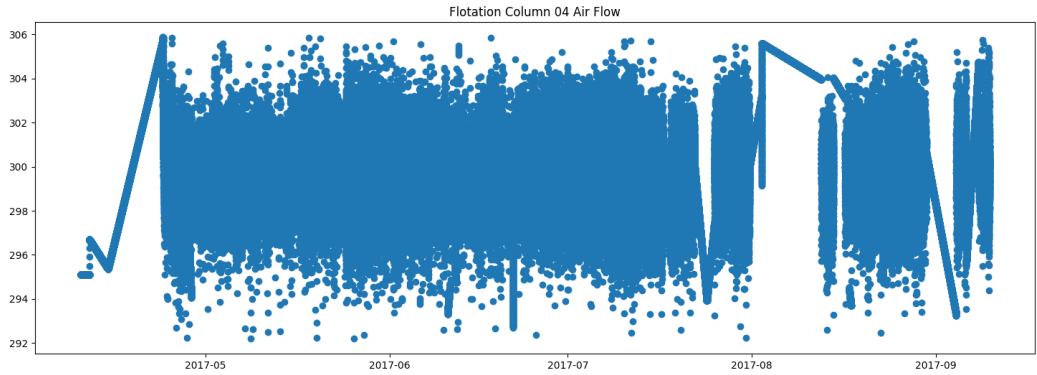


Dado
Filtrado

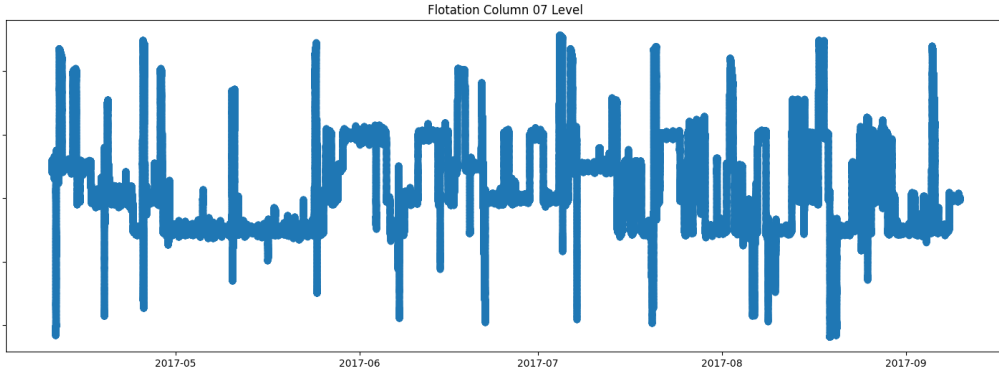
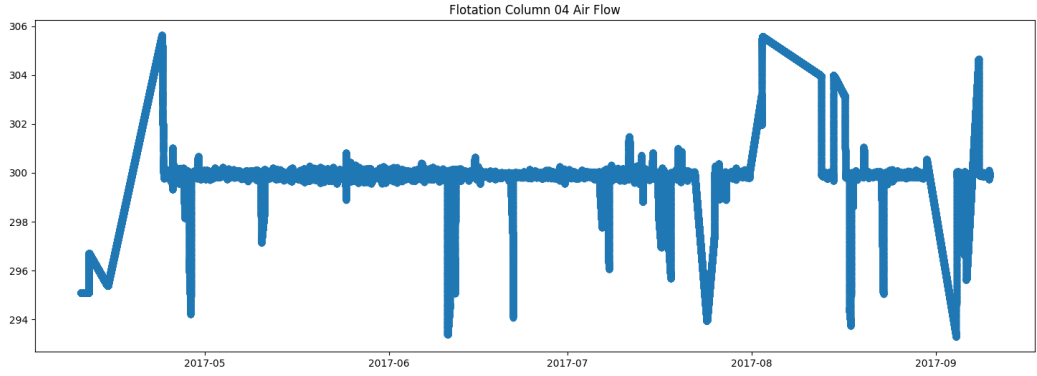


EDA (Análise Exploratória de Dados)

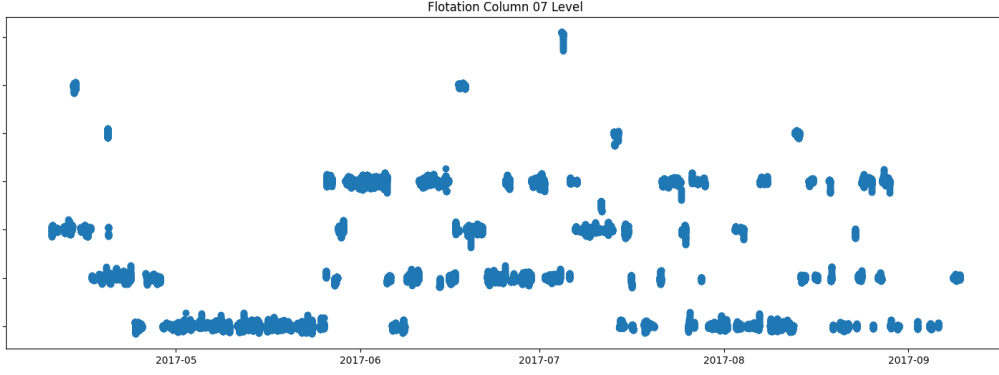
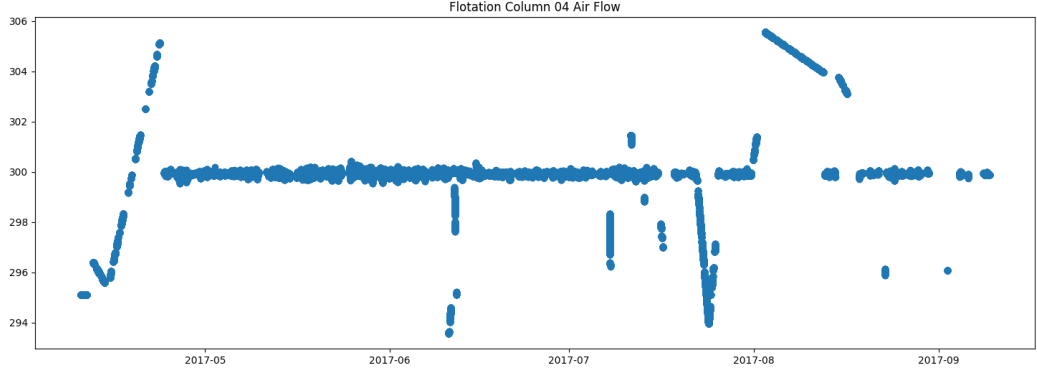
Raw Data



Feature

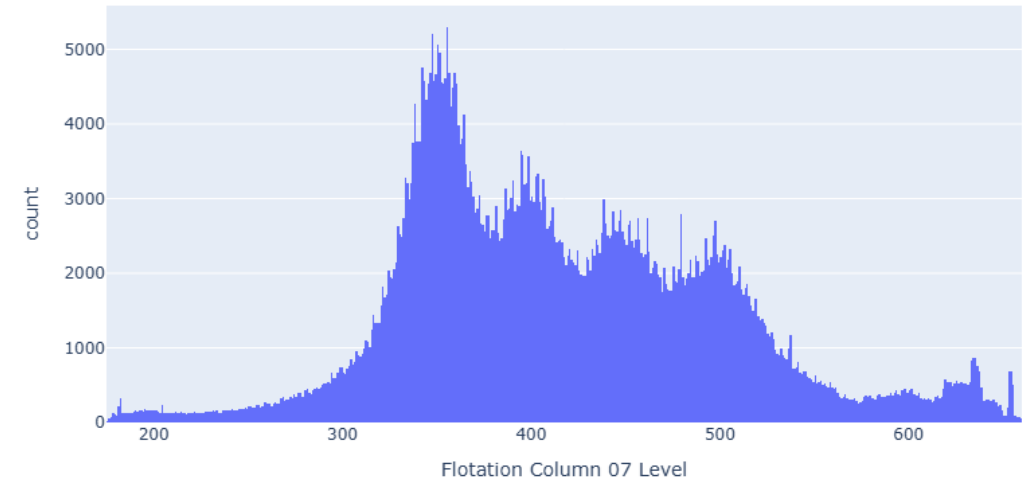
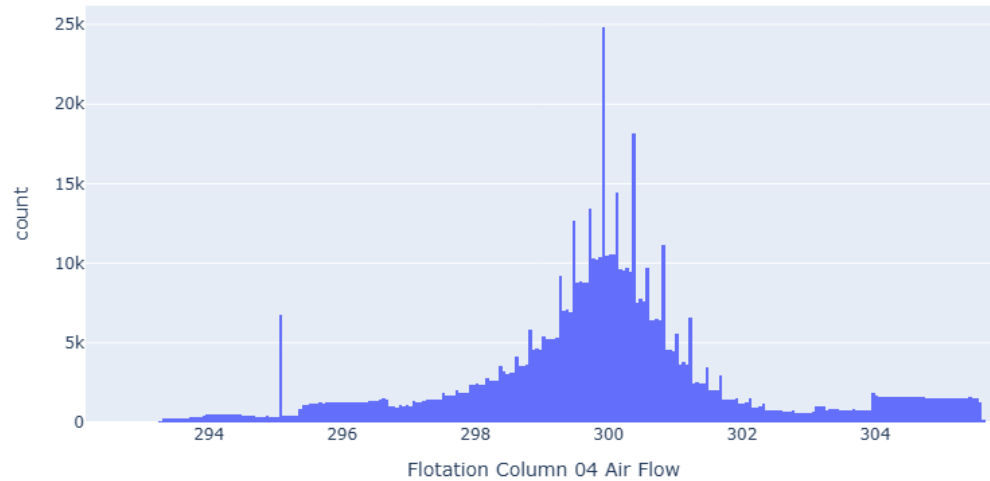


Dado Filtrado

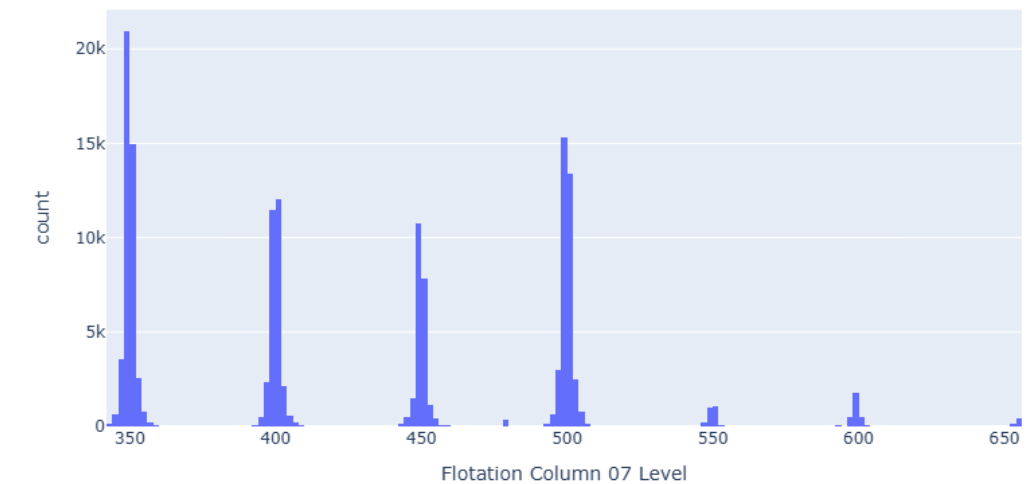
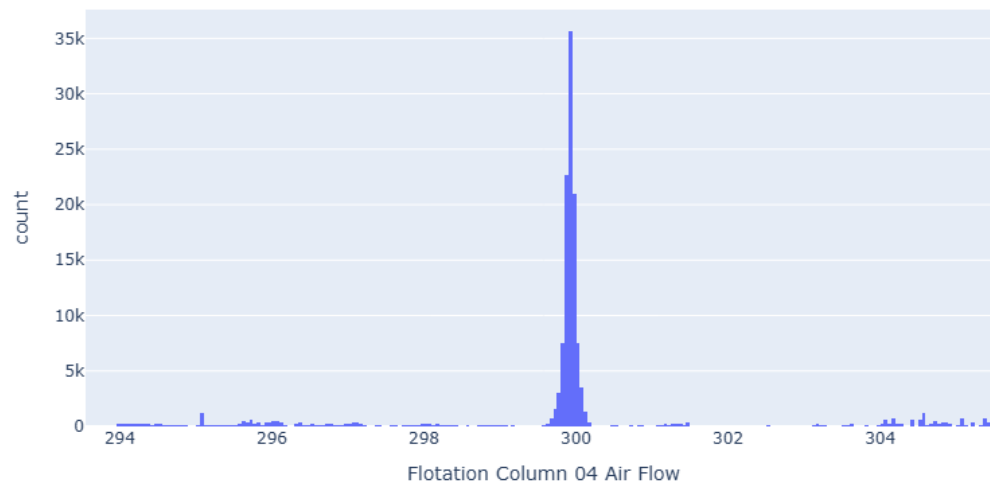


EDA (Análise Exploratória de Dados)

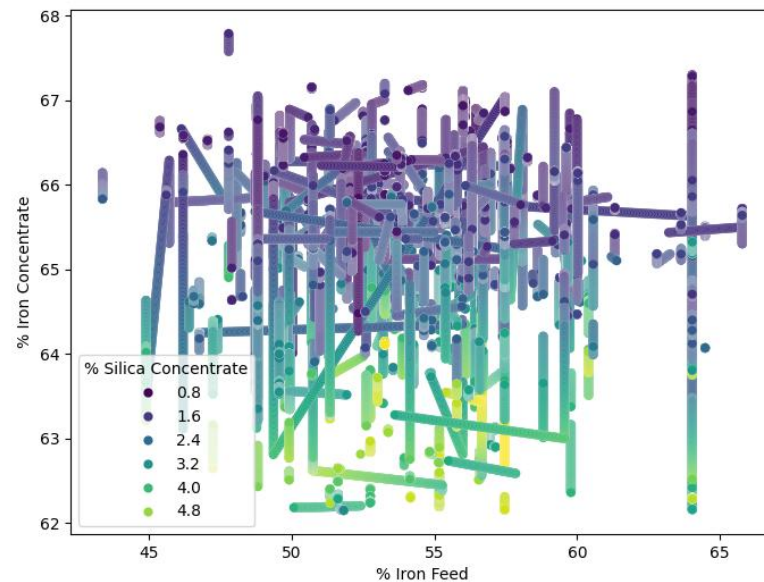
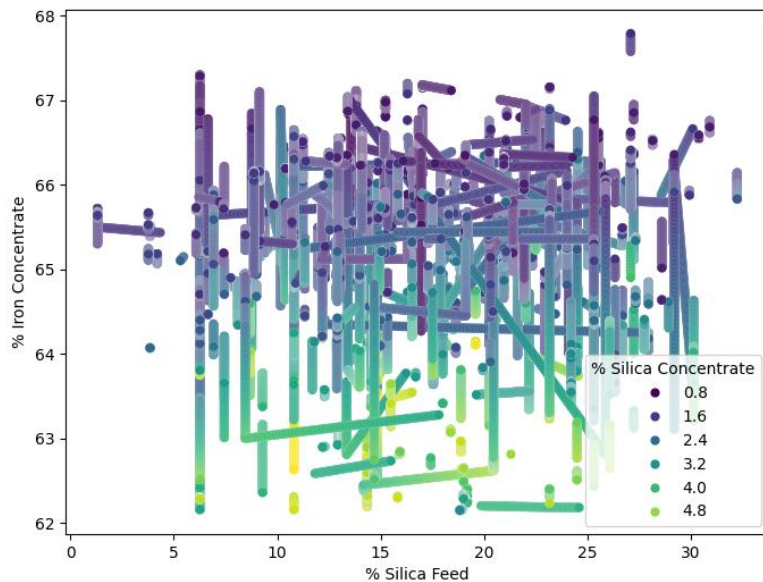
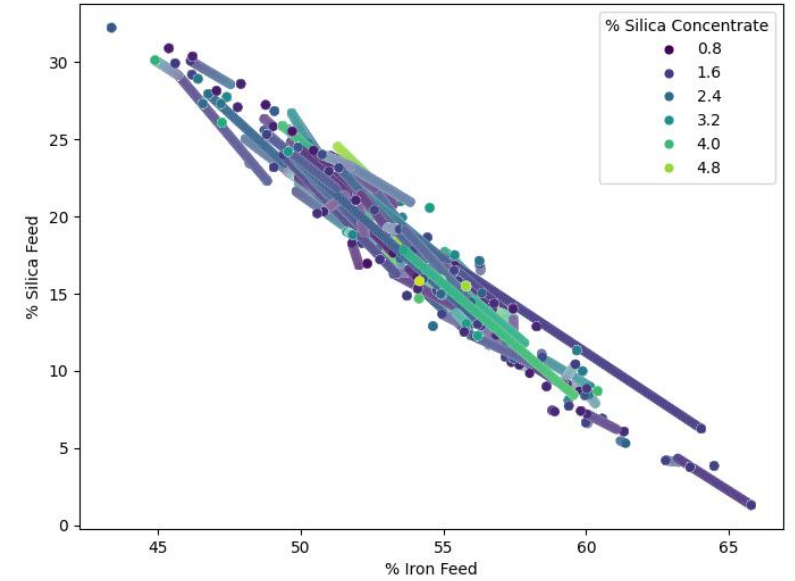
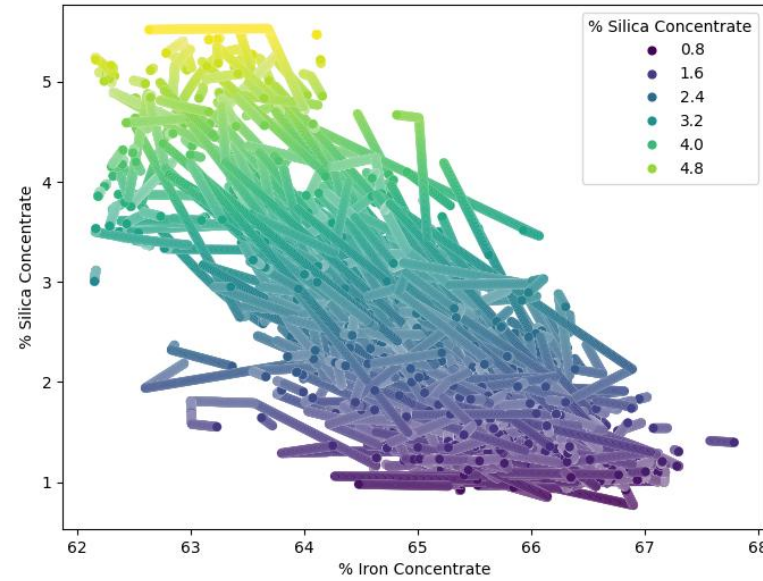
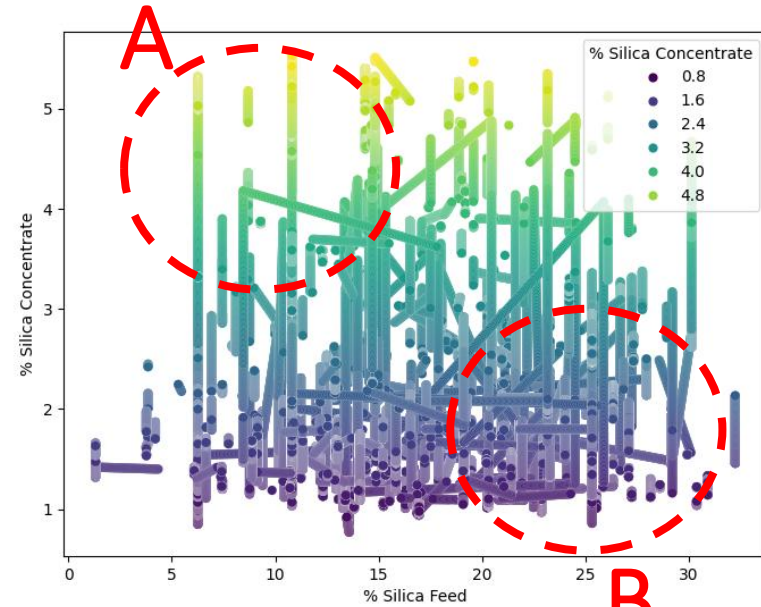
Raw Data



Dado
Filtrado



EDA (Análise Exploratória de Dados)



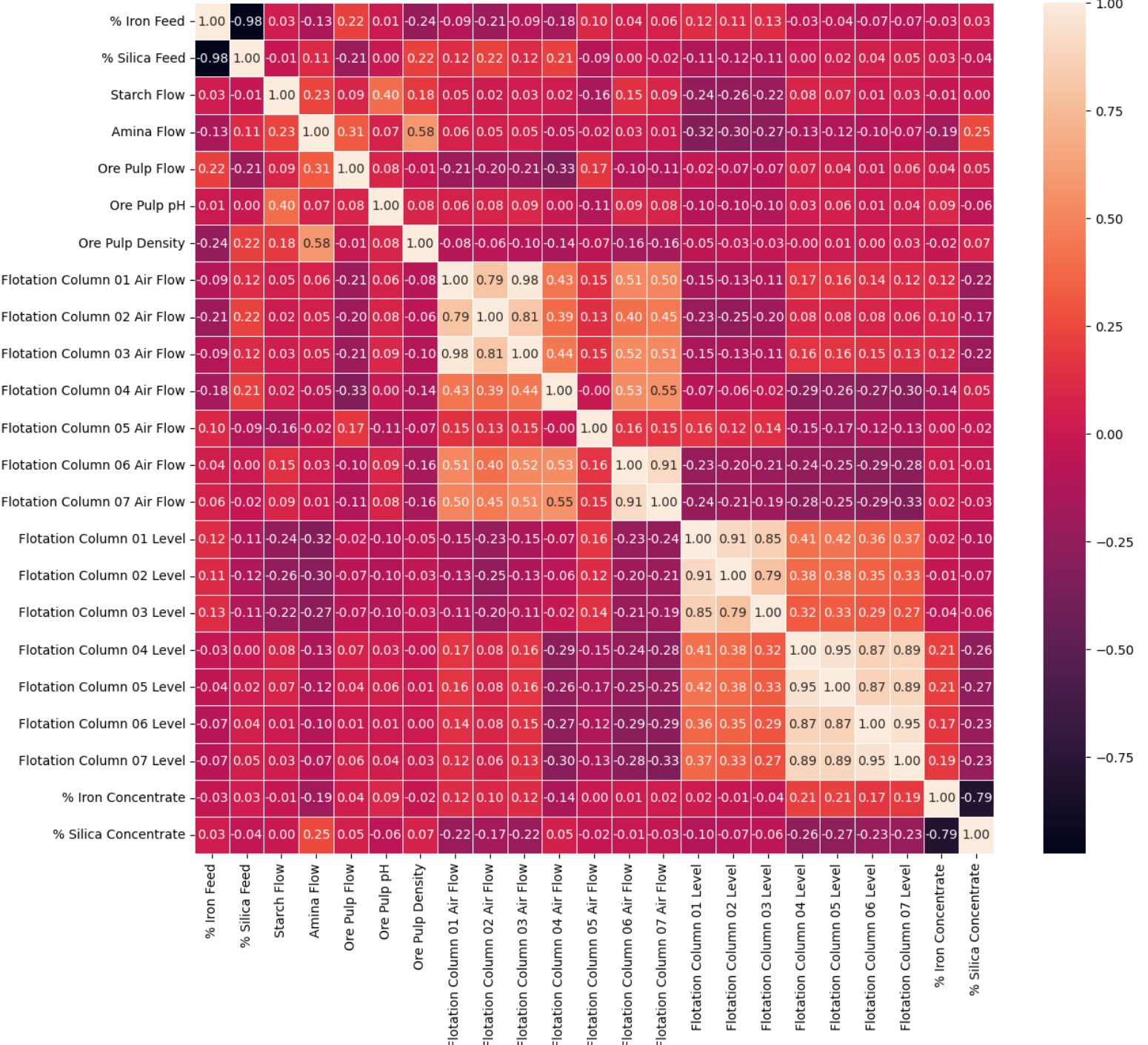
Região
A

- Minério com baixo teor de sílica (baixa impureza), mas o concentrado resultante apresenta teor de impureza elevado.
- Oportunidade de melhor aproveitamento do minério com baixo teor de impureza.

Região
B

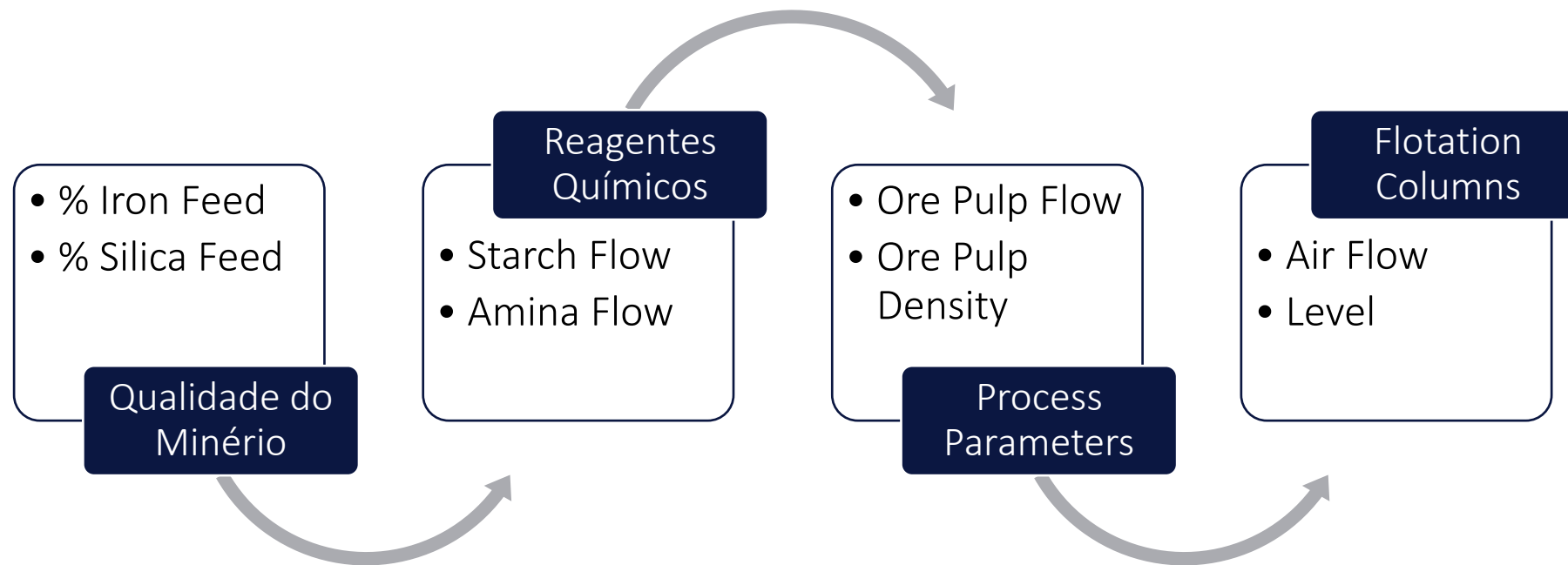
- Baixo teor de sílica no concentrado, mesmo que o teor no minério seja mais elevado.
- Esta é uma condição desejada – o processo executa sua função de remoção da impureza.
- Oportunidade de redução de minério no rejeito e de consumo de reagentes, com redução de variabilidade.

EDA (Análise Exploratória de Dados)



Modelo Predictivo
% Silica Concentrate

Features



No treinamento do modelo preditivo não se utilizou do dado % de concentrado de ferro

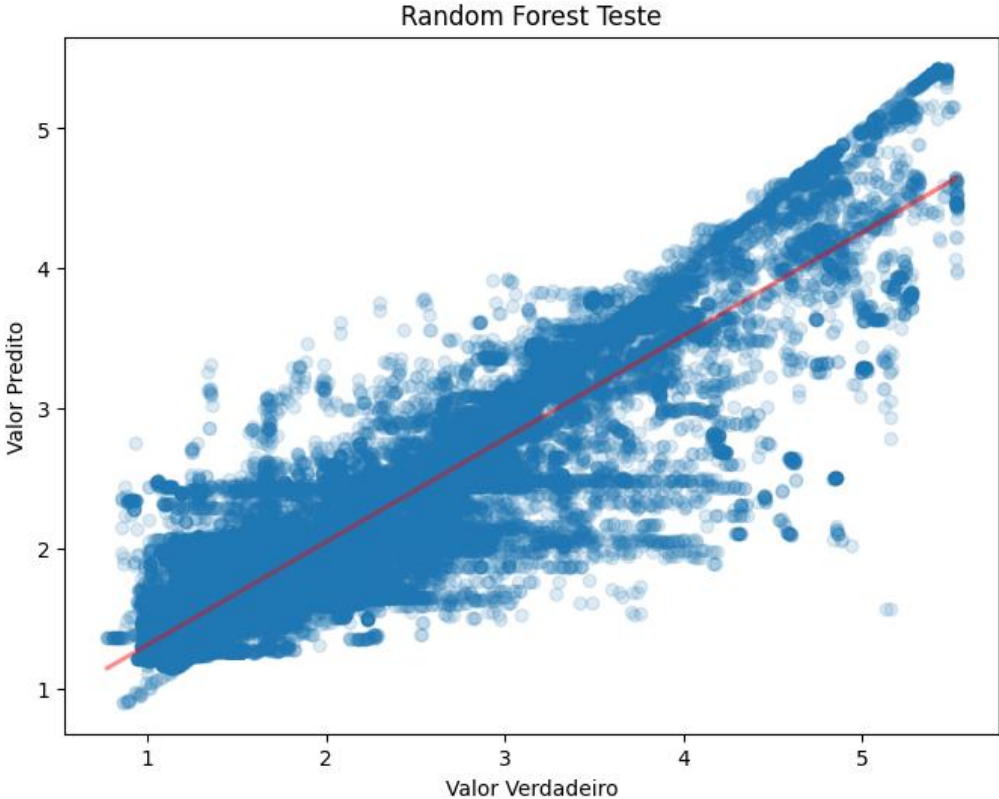
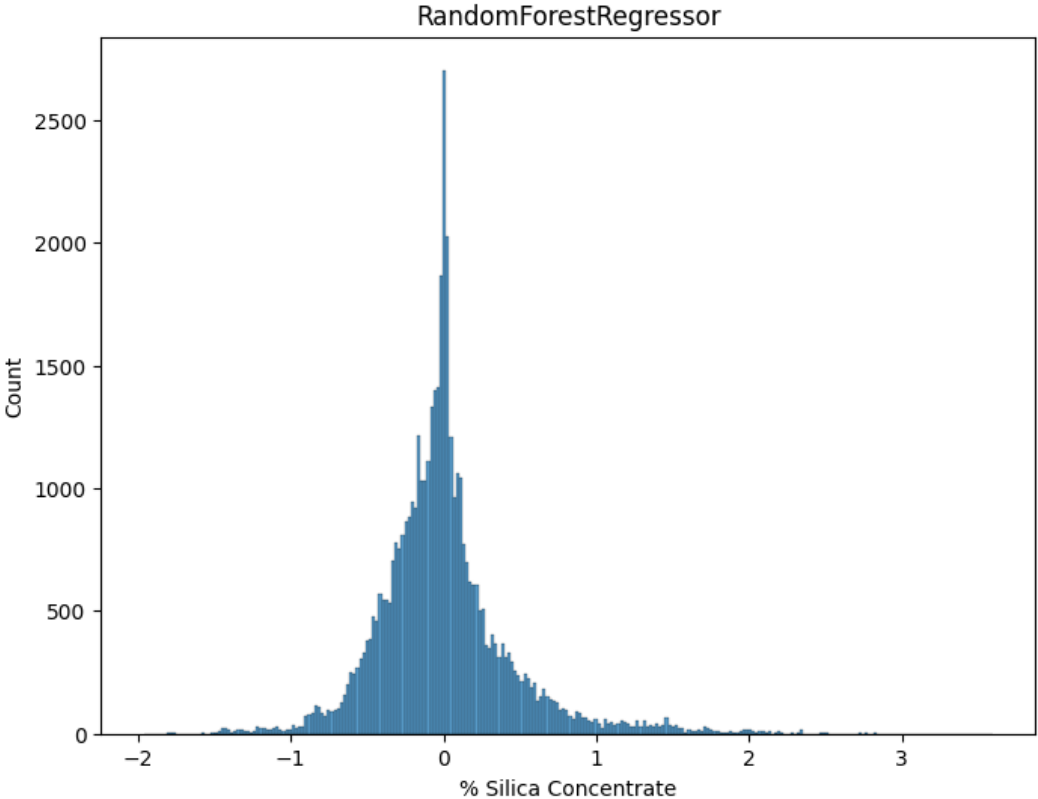
- Este é um dado providenciado pelo laboratório. Dessa forma, não é aplicável considerá-lo disponível para realizar a predição, pois o tempo do resultado será semelhante ao da variável desejada.

% Silica Concentrate

(Modelagem Preditiva)

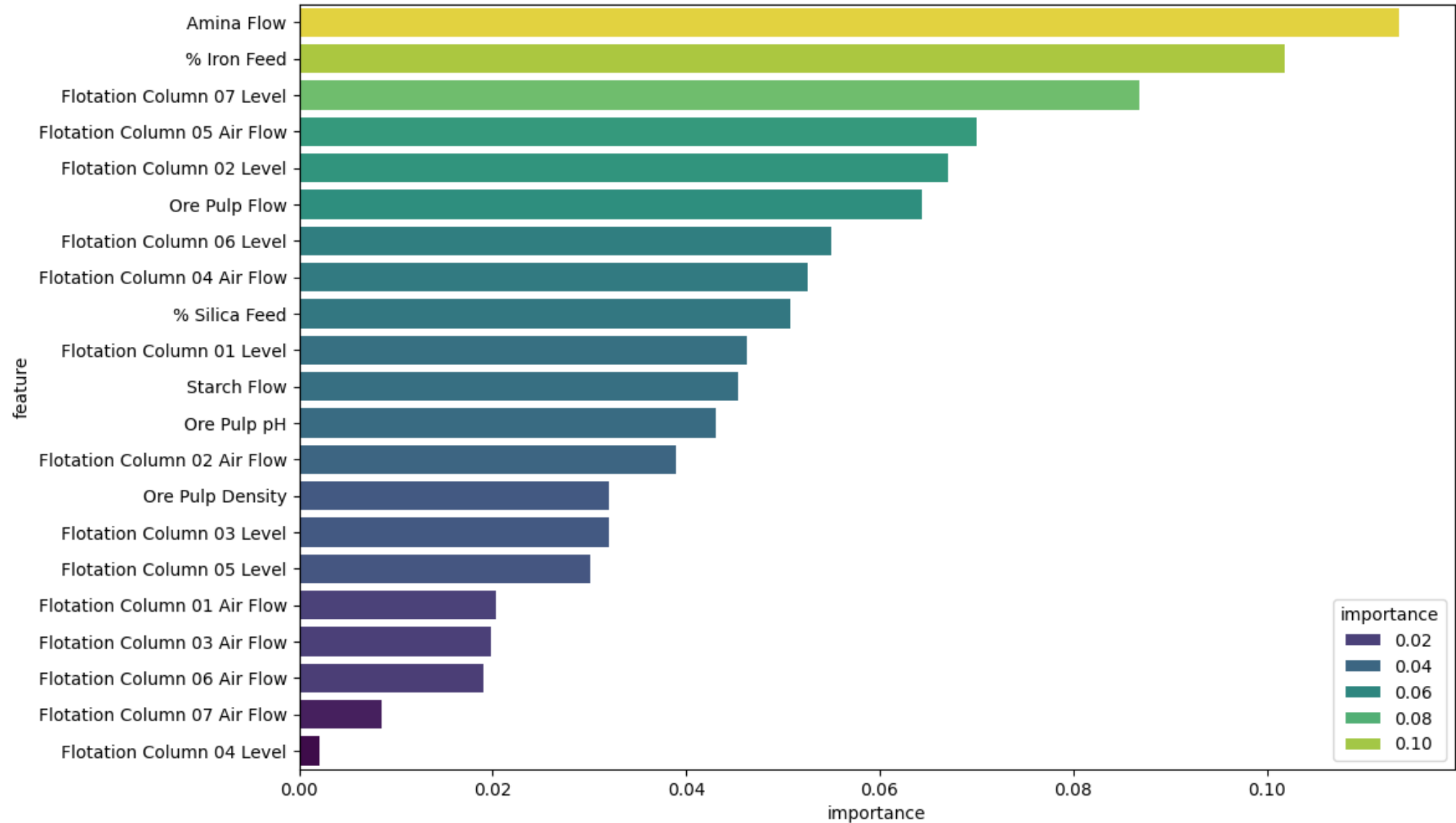
Modelo escolhido:
RandomForest

| Label | Mean Absolute Error (MAE) | Mean Squared Error (MSE) | Root Mean Squared Error (RMSE) | R-squared (R2) Score |
|-------|---------------------------|--------------------------|--------------------------------|----------------------|
| Train | 0,29 | 0,18 | 0,43 | 0,81 |
| Test | 0,29 | 0,19 | 0,43 | 0,81 |



Variáveis mais Importantes

(RandomForest)



Para que serve o Modelo Preditivo?

Em que condição o processo está operando?

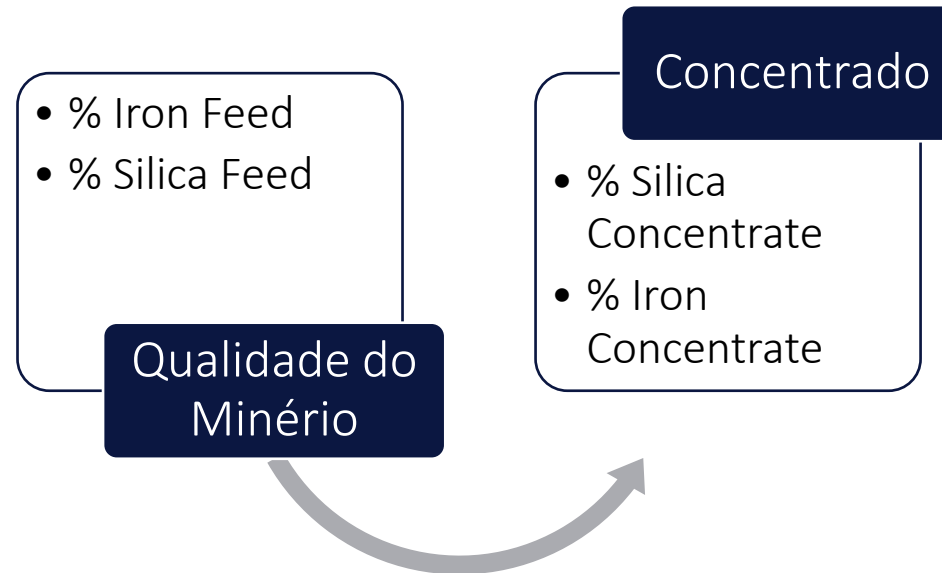
Qual a condição operacional desejada?

Quais parâmetros de processo devem ser alterados?

Análise de Clusters

(Padrões Operacionais)

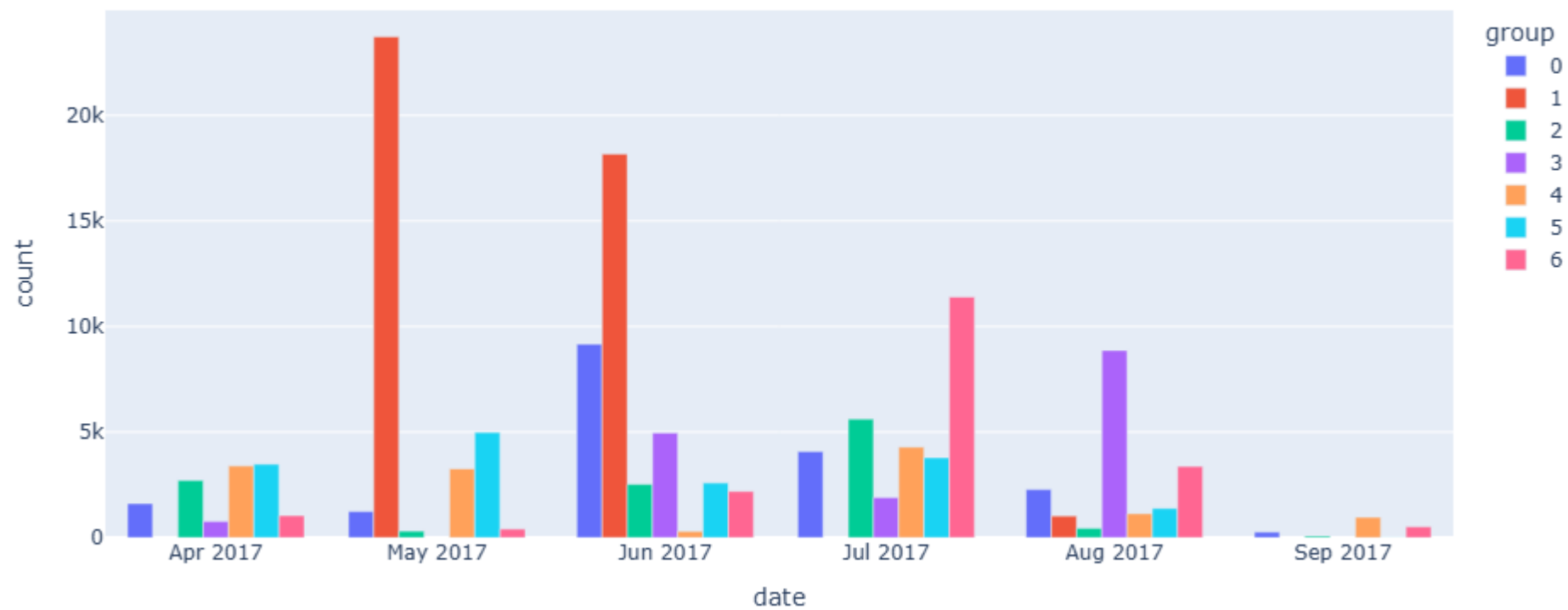
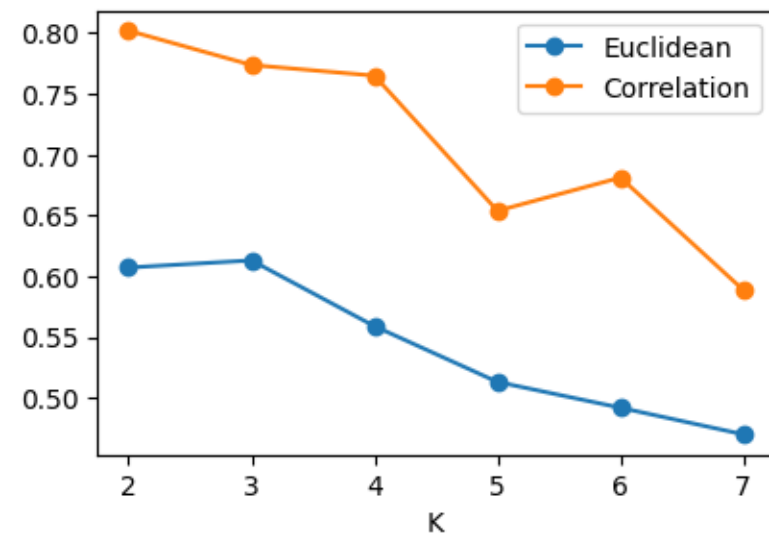
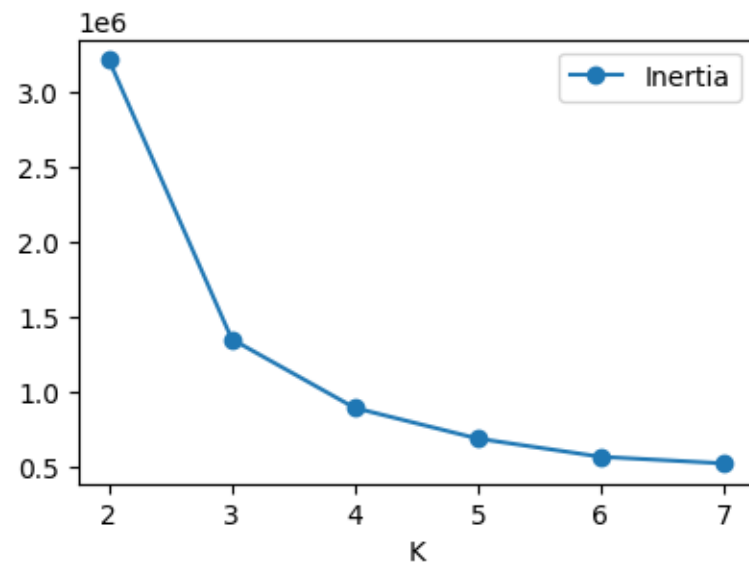
Cenário 1 - Análise Clusters



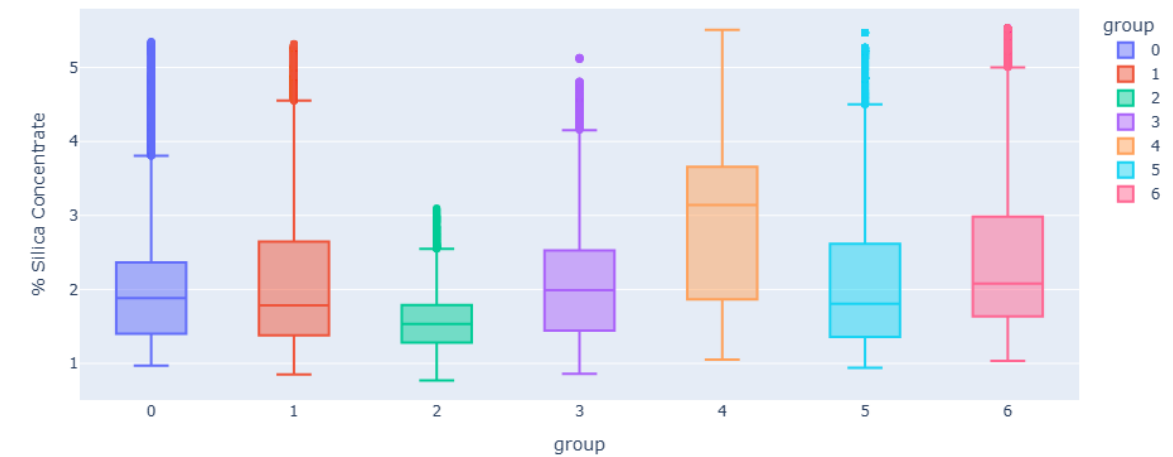
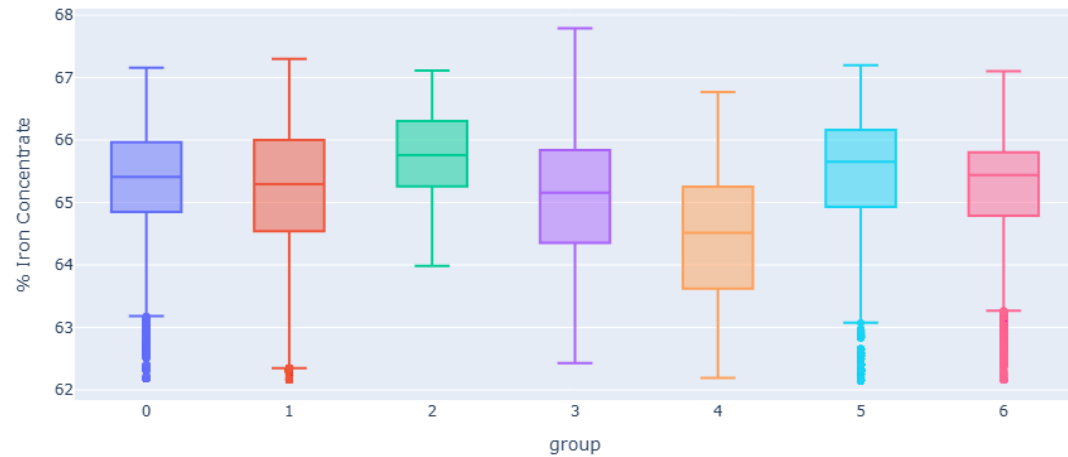
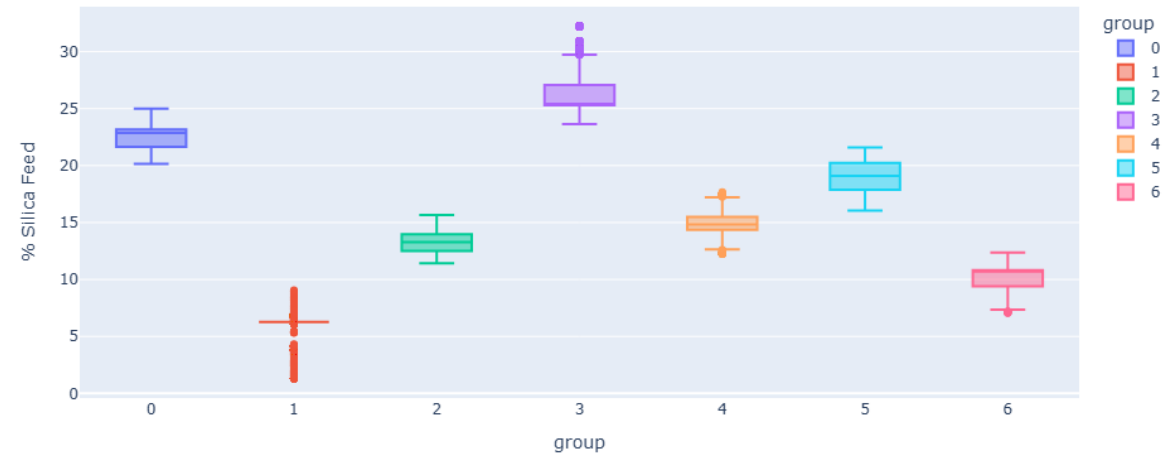
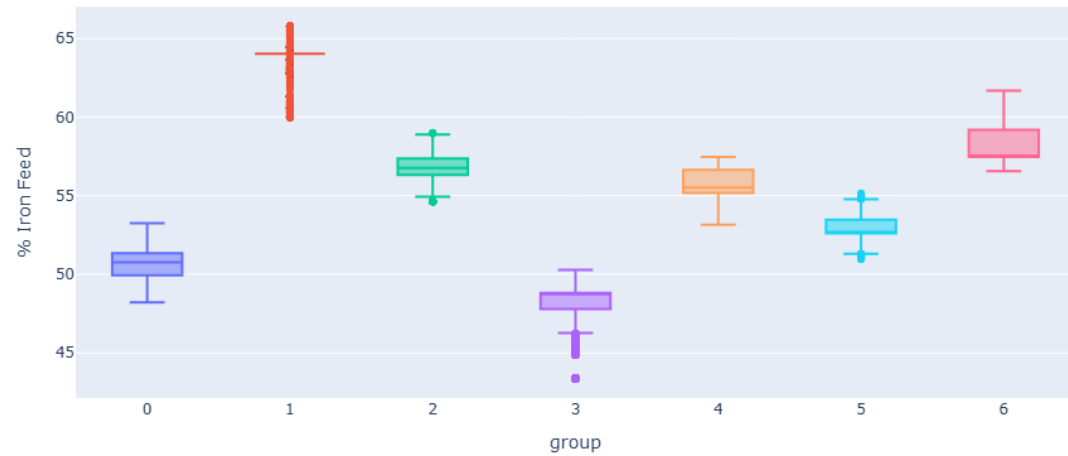
No cenário 1 da análise de clusters, foram empregados dados de qualidade do minério e do concentrado

- Objetivo: observar padrões evidenciados pela qualidade, que indiquem comportamentos do processo

Cenário 1

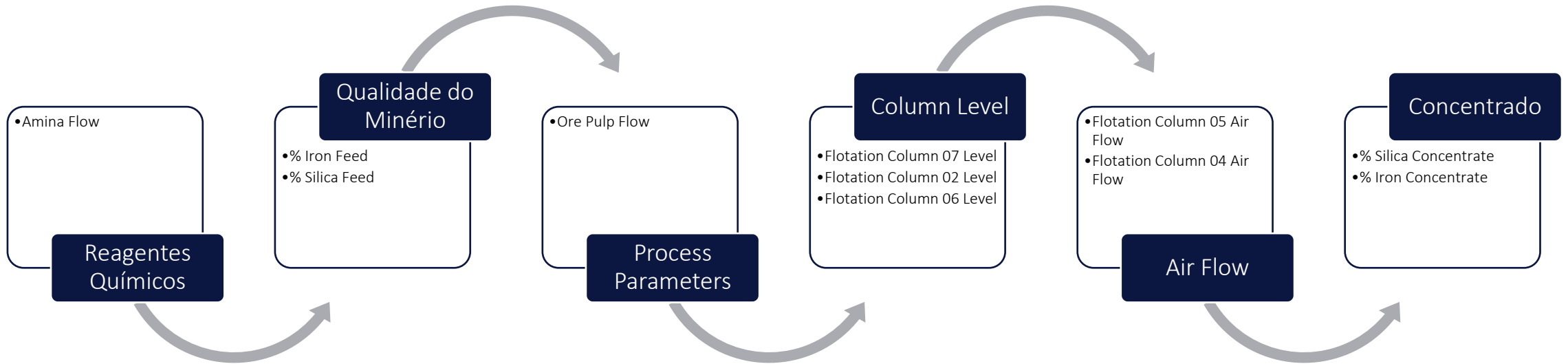


Cenário 1



- Grupo 1, se assemelha a REGIÃO A de oportunidade, com baixos teores de sílica e elevado teor de Ferro no minério. No entanto, o efeito do processo é minimizado na transformação do concentrado.
- Grupo 3, se assemelha da REGIÃO B de oportunidade, com teores de sílica elevado no minério, e o processo consegue reduzir a impureza no concentrado. Observa-se que apresenta baixo teor de ferro no minério e bom efeito no processo de concentração.
- Grupo 6, apresenta uma oportunidade a ser investigada. Pergunta: o que é necessário para reduzir o teor de sílica e aumentar teor de Fe no concentrado (assim como observa-se no grupo 2)? Quais variáveis do processo podem gerar este efeito?

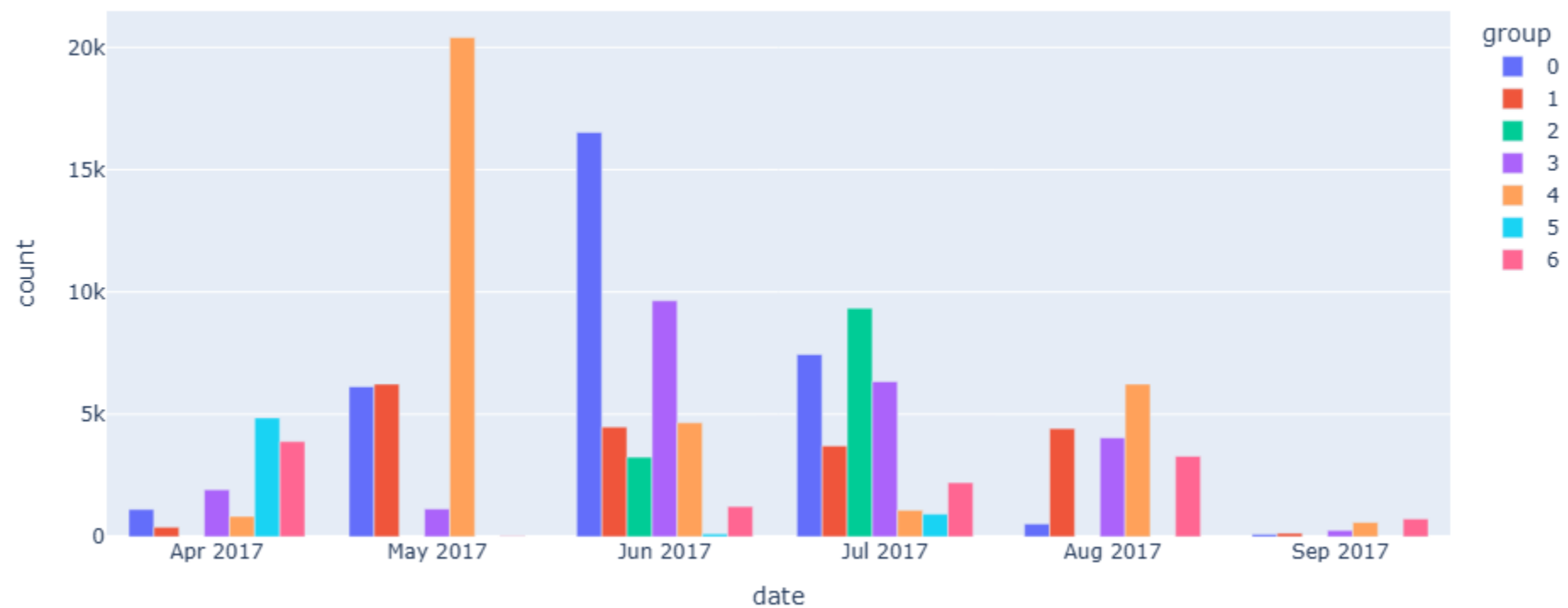
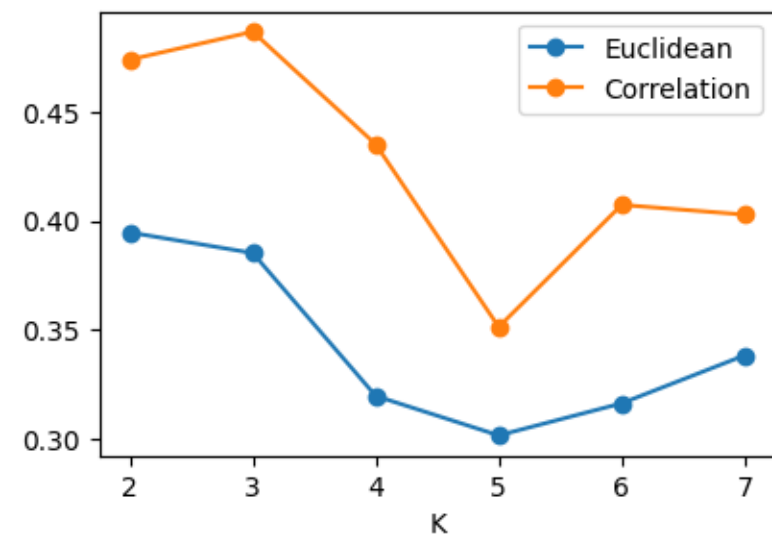
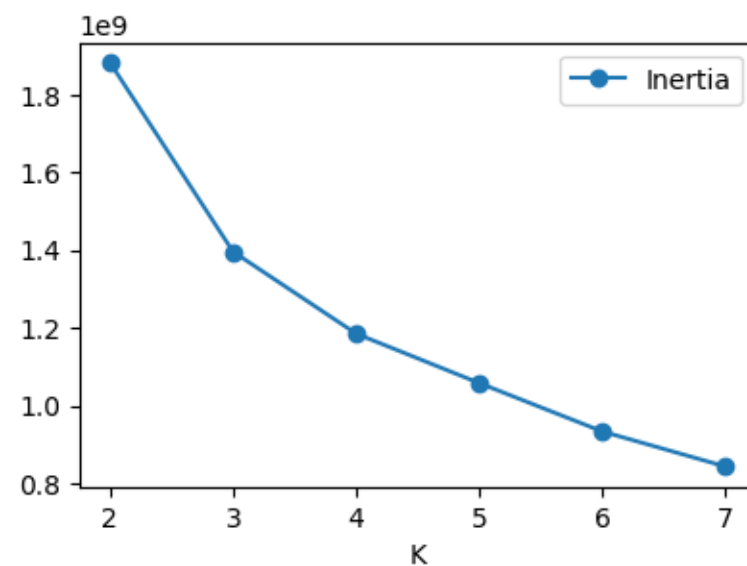
Cenário 2 - Análise Clusters



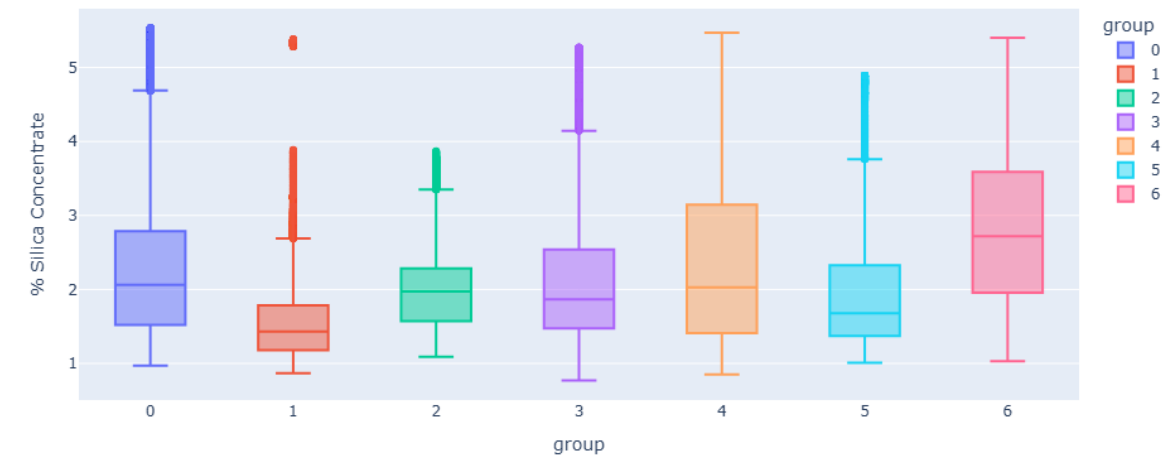
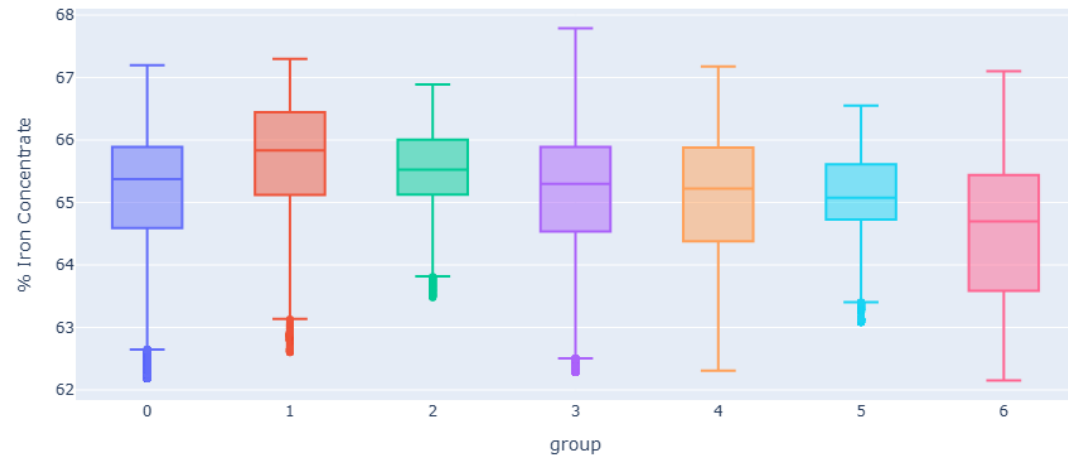
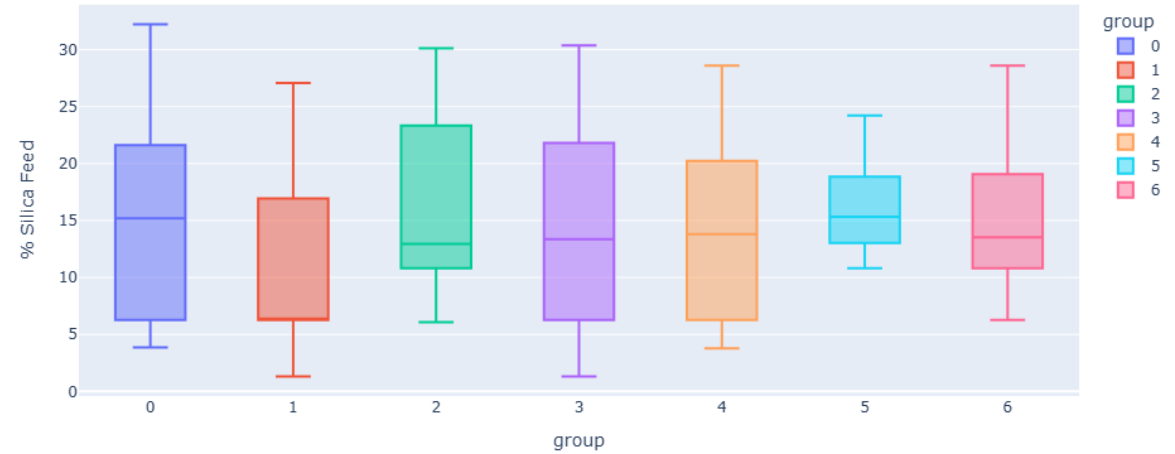
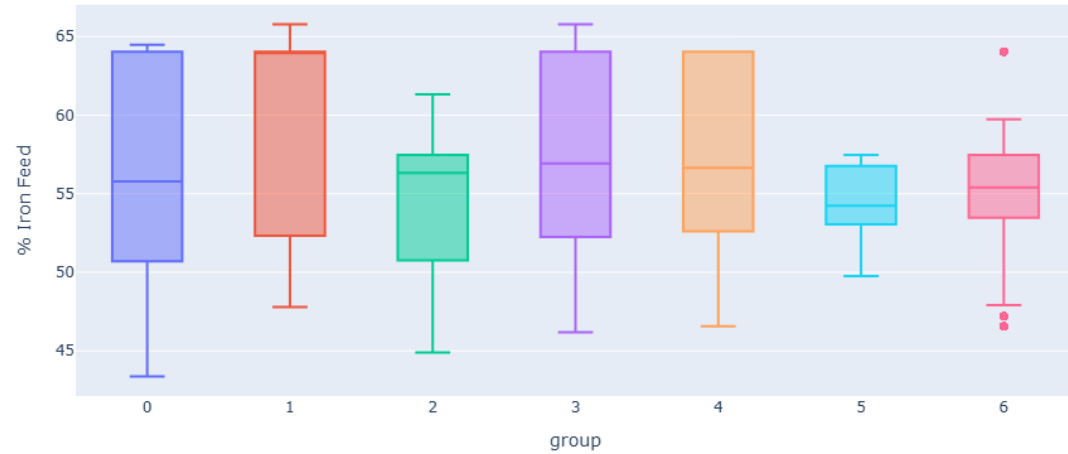
No cenário 2 da análise de clusters, foram empregados os dados com maior nível de importância, obtidos pelo modelo RandomForest

- Adicionalmente, incluiu-se os dados de % Iron Concentrate e % Silica Concentrate

Cenário 2



Cenário 2

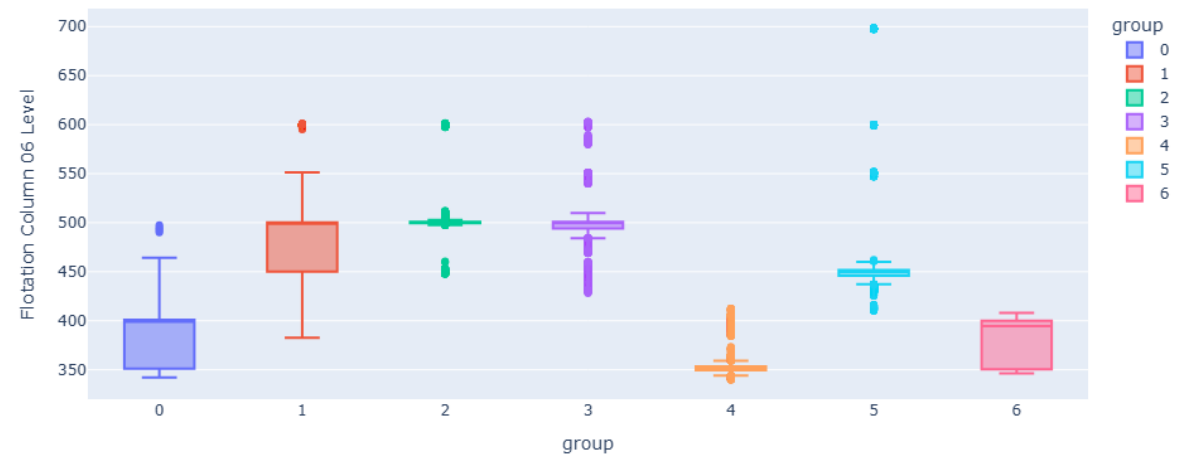
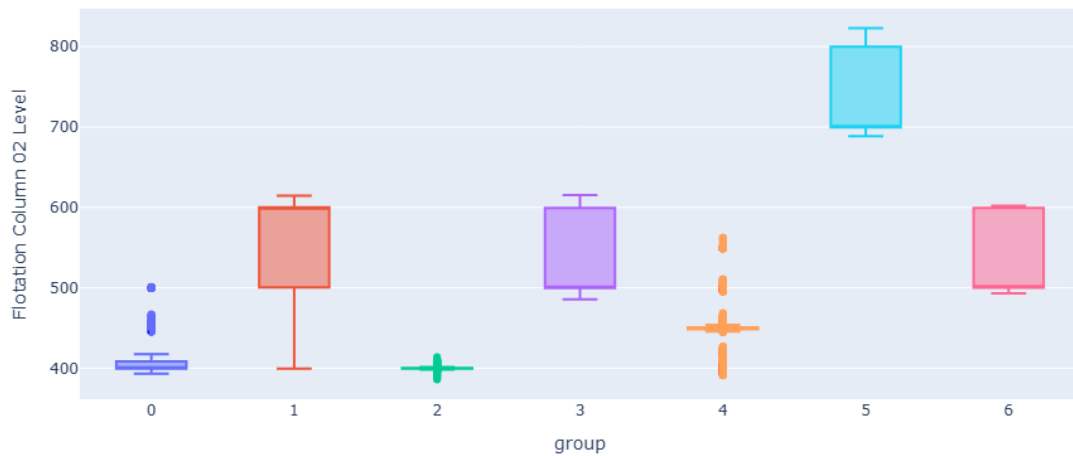
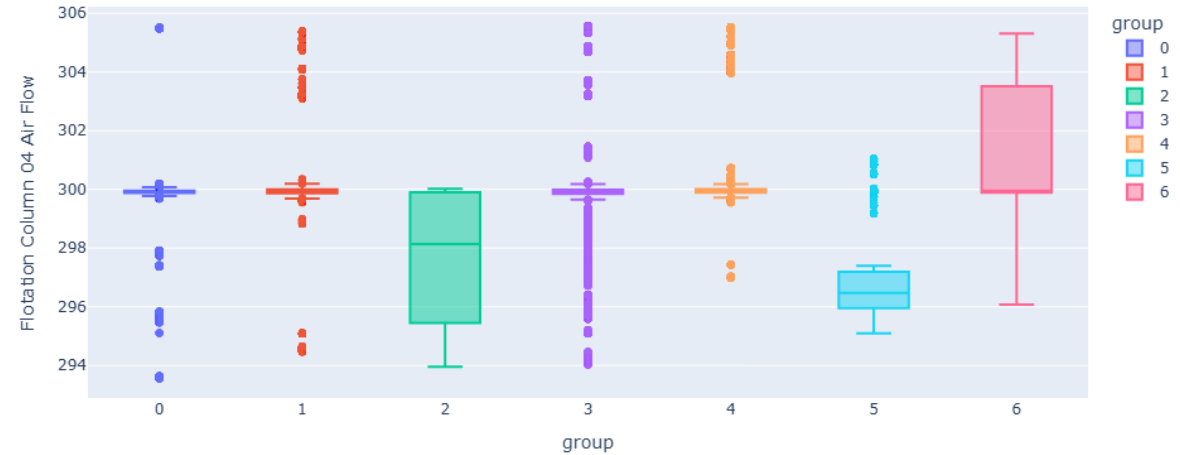
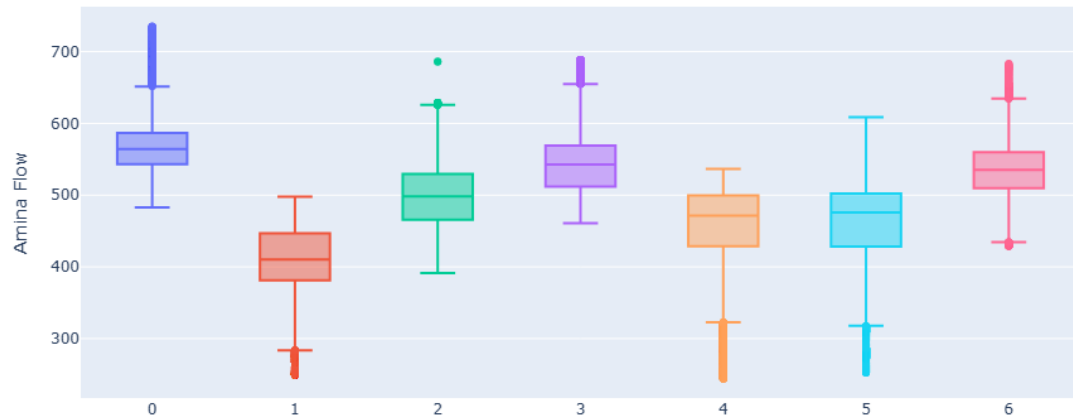


- Grupo 2, é uma referência de operação desejada.

De modo semelhante a análise do Cenário 1

- Grupo 6, apresenta uma oportunidade a ser investigada. Pergunta: o que é necessário para reduzir o teor de sílica e aumentar teor de Fe no concentrado (assim como observa-se no grupo 2)? Quais variáveis do processo podem gerar este efeito?

Cenário 2



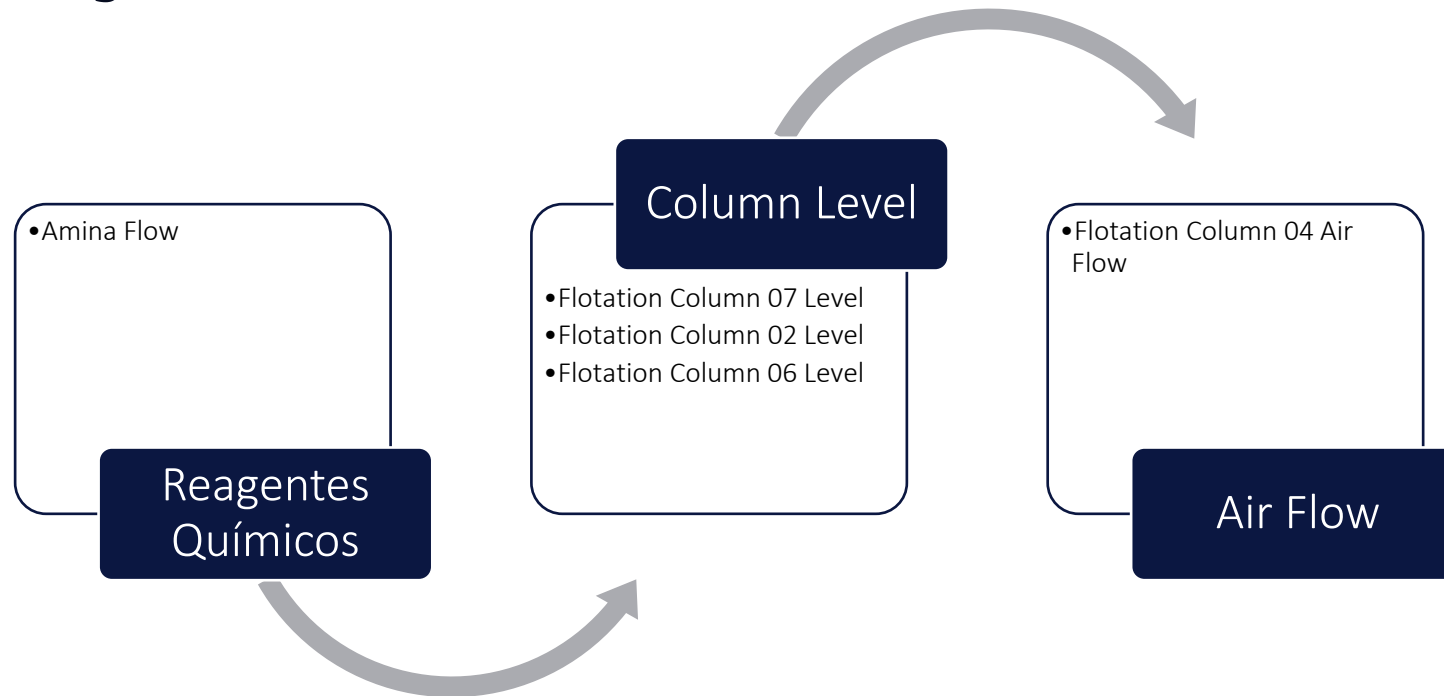
- Observa-se as diferenças de distribuição dos parâmetros operacionais entre Grupo 6 e Grupo 2

A alteração destes parâmetros, pode gerar um efeito desejado de redução do teor de sílica no concentrado?
(sendo mantidos os demais parâmetros)

Análise de Monte Carlo

(Utilização do modelo Preditivo
para Simulações Operacionais)

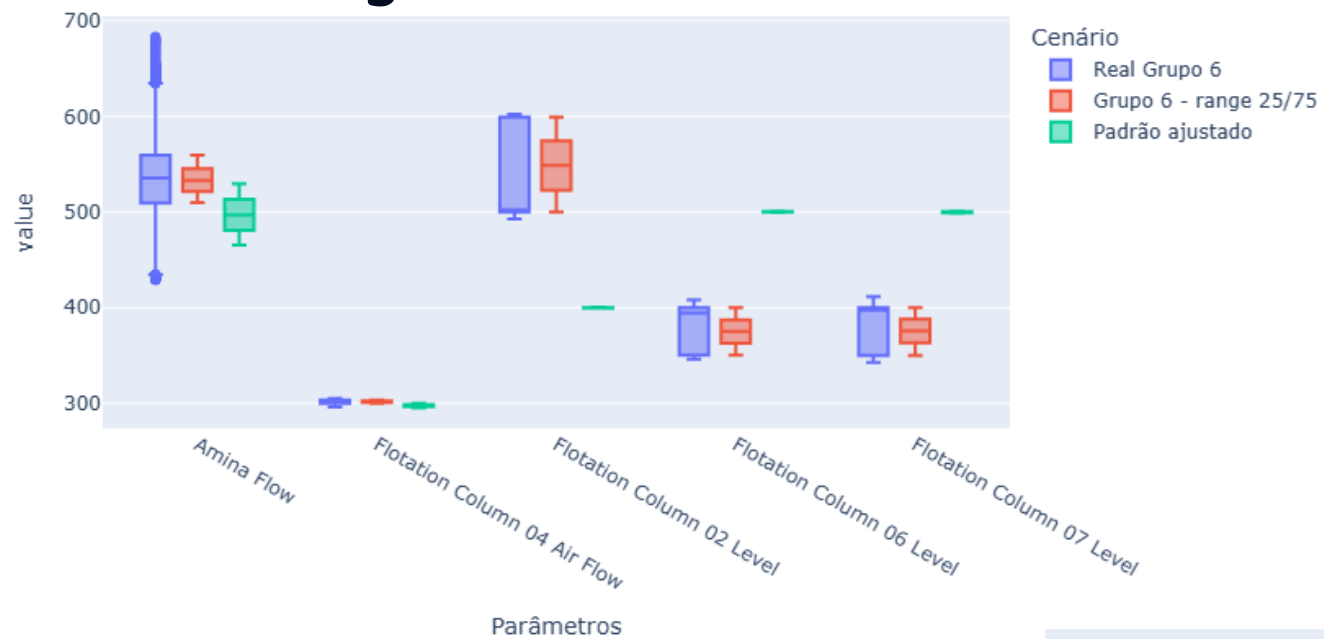
Simulação - Variáveis



Qualidade do concentrado não é interpretada como uma variável manipulável, por isso não foram consideradas na simulação de alteração de padrão operacional

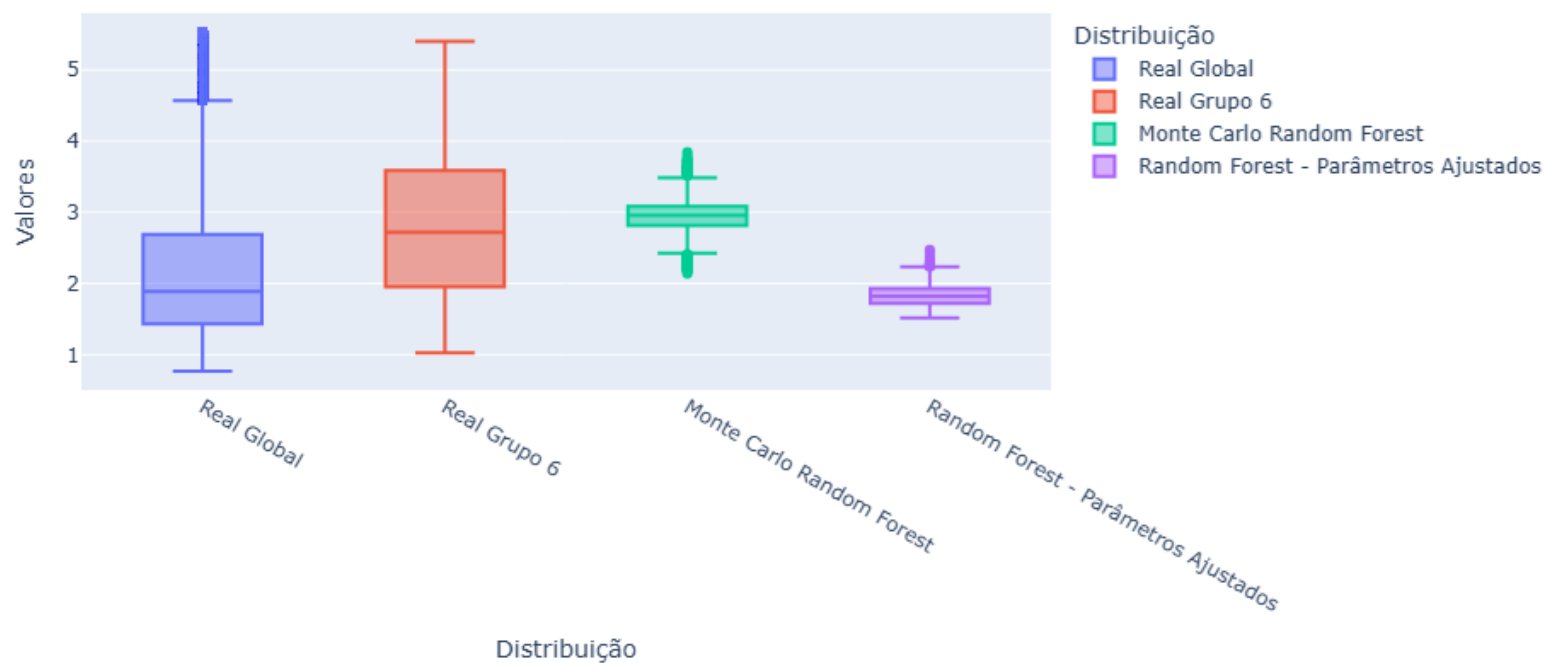
- Na Simulação, foi utilizado os limites de 25% e 75% no range da distribuição das variáveis
- Simulando a alteração do range do Grupo 6 para o range equivalente do Grupo 2.

Simulação Monte Carlo



Variáveis manipuladas para alteração de padrão operacional

% de Sílica no Concentrado



Oportunidades

Oportunidades

Avaliar efeito do período da janela das métricas (Features)

Exemplo: usar 2 horas dos ciclos para construir as métricas Média, Máximo, Mínimo e Desvio Padrão.

Avaliar métricas estatísticas (Features)

Exemplo: métricas de Máximo, Mínimo e Desvio Padrão.

Utilizar a predição de % Iron Concentrate na Predição

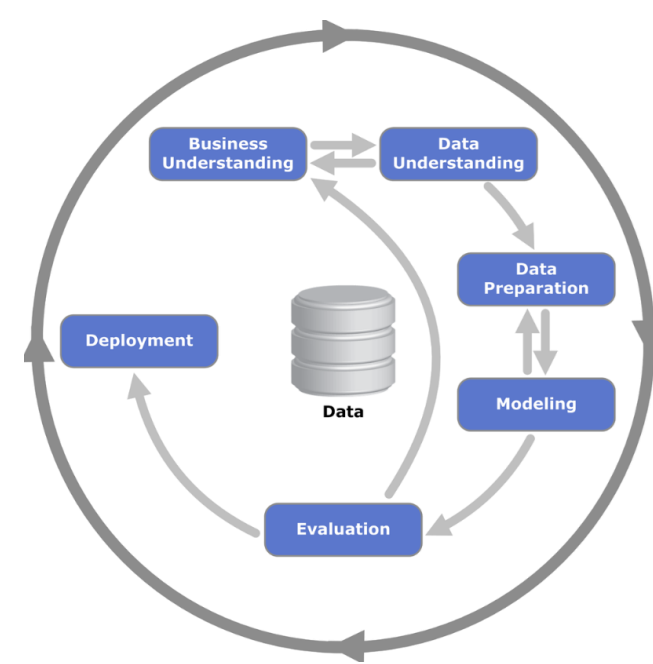
Usar os modelos em cadeia.

Oportunidades

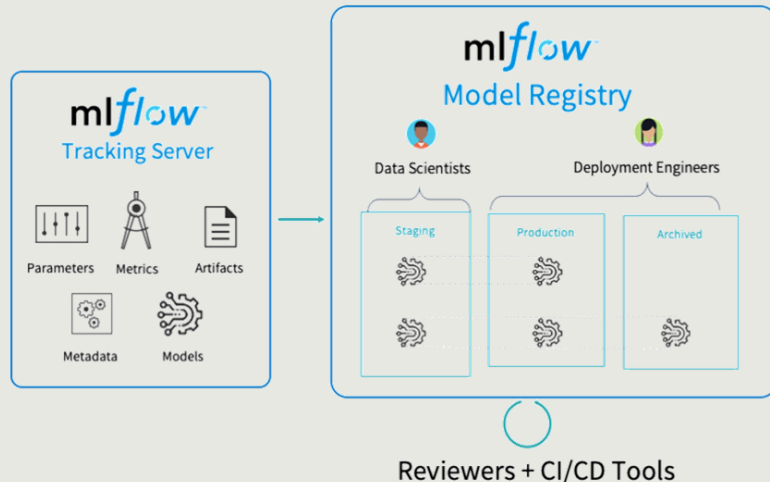
Como utilizar os Modelos

Cientista de
Dados

Business Understanding Data Understanding Data Preparation Modeling Evaluation



Engenheiro de
Machine Learning



Servir os Models
Deploy

- Criar serviço agendado e recorrente
- Criar API para consumo por outras aplicações

Usuário

Aplicações para o Usuário

- Avaliação Preditiva da Qualidade
- Simulação de Parâmetros operacionais



O Projeto desenvolvido para o Case pode ser encontrado no Github.

- Foi construído uma estrutura para servir os modelos criados via FastAPI
- Contempla ativação de Github Actions para atualização da imagem no DockerHub (<https://hub.docker.com/repository/docker/lucasrigobello/mining-process-quality-prediction/general>)
- Imagem Docker e Estrutura para trabalhar com Kubernetes



<https://github.com/lucasrigobello/mining-process-quality-prediction>

Muito obrigado!

Lucas da Silva Rigobello

lucasrigobello@gmail.com