



# PREVISÃO DO CONSUMO ENERGÉTICO RESIDENCIAL

AQUARELA ANALYTICS

A close-up, artistic photograph of a printed circuit board (PCB). The image shows intricate copper traces and numerous solder points, some of which are highlighted with a greenish-blue glow. The background is dark, making the metallic components stand out.

## OBJETIVOS

- Desenvolver uma solução preditiva robusta aplicando boas práticas de engenharia e ciência de dados para apoiar decisões energéticas segmentadas.

# ESTRATÉGIA APLICADA

- Análise exploratória e tratamento dos dados
- Clusterização de clientes
- Modelagem preditiva por cluster e região
- Construção de pipeline com boas práticas de MLOps
- Visualização e entrega de insights

# ENTENDIMENTO DOS DADOS

## Dados disponíveis:

- clientes.csv: dados cadastrais dos clientes
- consumo.csv: histórico de consumo por cliente/data
- clima.csv: dados climáticos por região/data

## Desafios Enfrentados:

- Regiões incorretas em dados cadastrais
- Dados climáticos ausentes
- Variação no padrão de consumo entre clientes

# ANÁLISE EXPLORATÓRIA

## Clientes

### Dados:

- client\_id: chave identificadora do cliente
- region: Região do cliente

### Problemas:

- Algumas regiões estão cadastradas como “Desconhecida”, representando 5% dos clientes.

### Alternativas para solução de problemas:

- Criação de um modelo classificador de regiões.
- Exclusão dos dados.



# ANÁLISE EXPLORATÓRIA

## Clientes

O modelo de classificação de regiões não atingiu métricas aceitáveis:

- Accuracy: 0.2533
- Precision: 0.2654
- Recall: 0.2533
- F1-Score: 0.2554
- ROC-AUC: 0.5491

As métricas são baixas para e um escolha aleatória de regiões é considerada melhor.

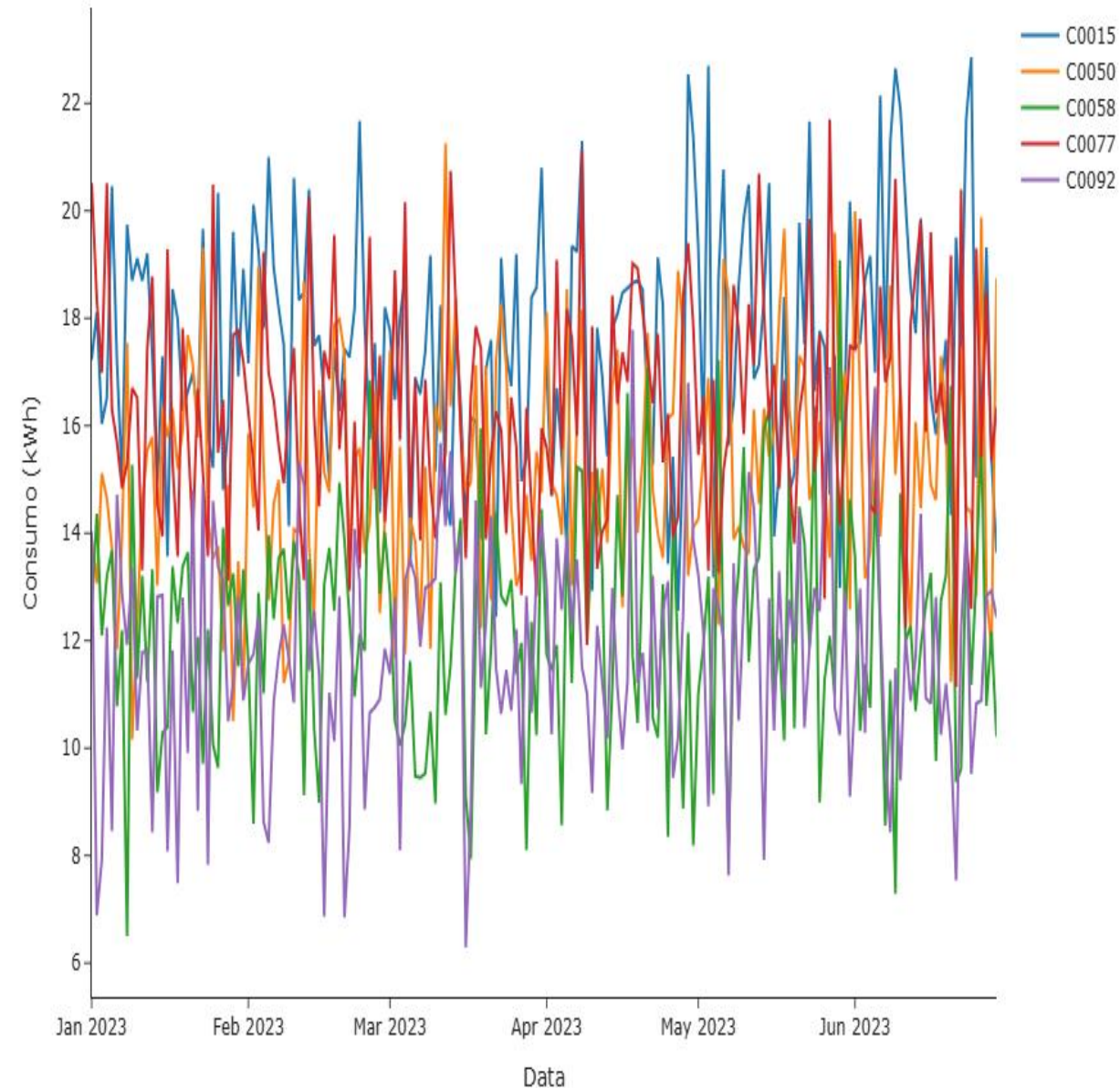
# ANÁLISE EXPLORATÓRIA

## Clientes

A exclusão dos dados foi a alternativa adotada para os dados errados.

O número de 5% apenas é baixa e não causará falta no treinamento dos modelos pois os clientes com região “Desconhecida” não possuem padrão.

Séries Temporais de Clientes com Região 'Desconhecida'



# ANÁLISE EXPLORATÓRIA

## Clima

### Dados:

- region: Região do clima
- date: dia relacionado aos dados de clima
- Temperature: temperatura °C
- Humidity: umidade %

### Problemas:

- Alguns dias com dados “temperature” faltantes.

### Alternativa para solução de problemas:

- Interpolação estes os dias para a região.



# ANÁLISE EXPLORATÓRIA

## Clima

### Solução de problemas:

- Os valores de “temperature” podem receber o valor interpolado entre o dia anterior e o dia seguinte. Esta técnica mantém o padrão e a fluidez da série temporal e não afeta o modelo no treino. Ela segue a tendência e não permite buracos em uma série temporal.

# ANÁLISE EXPLORATÓRIA

## Consumo

### Dados:

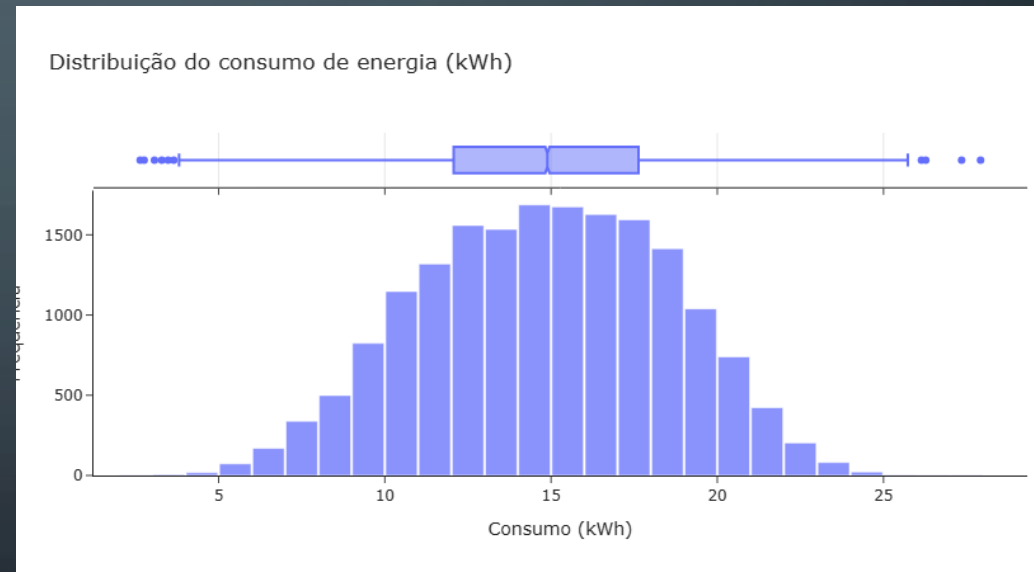
- client\_id: indentificador do cliente
- date: dia relacionado aos dados de clima
- consumption\_kwh: temperatura °C

Não há problemas nestes dados.

# ANÁLISE EXPLORATÓRIA

## CONSUMO

- Foi possível observar que o consumo dos clientes padrões.
- Os clientes não costumam sair de sua zona de atuação.



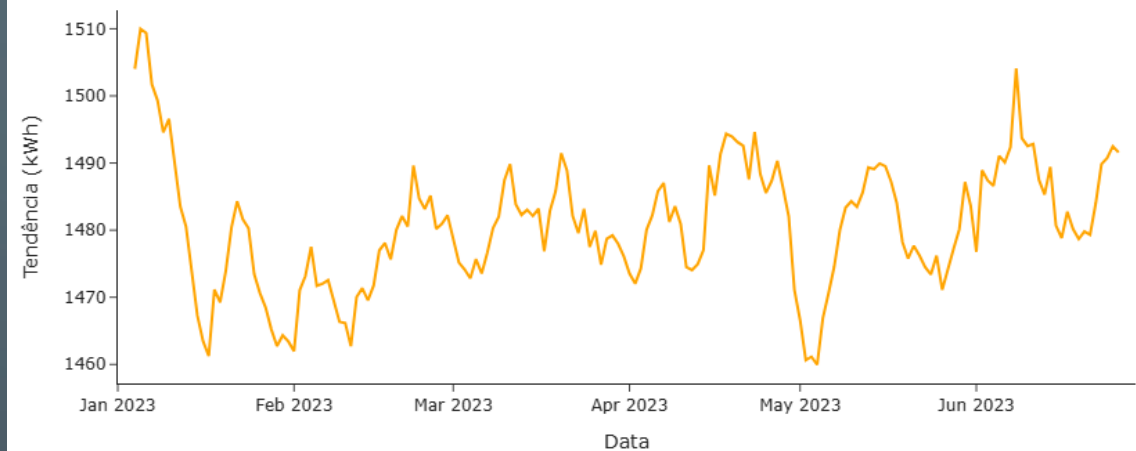
# ANÁLISE EXPLORATÓRIA

Consumo

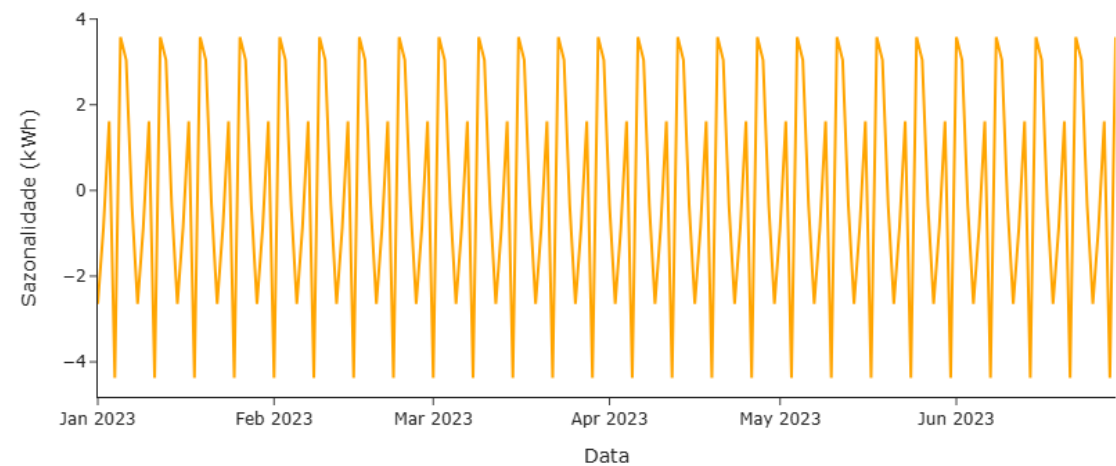
O consumo geral não tem uma tendência definida.

A sazonalidade do consumo é de 7 dias (1 semana).

Tendência do Consumo Diário



Sazonalidade do Consumo Diário

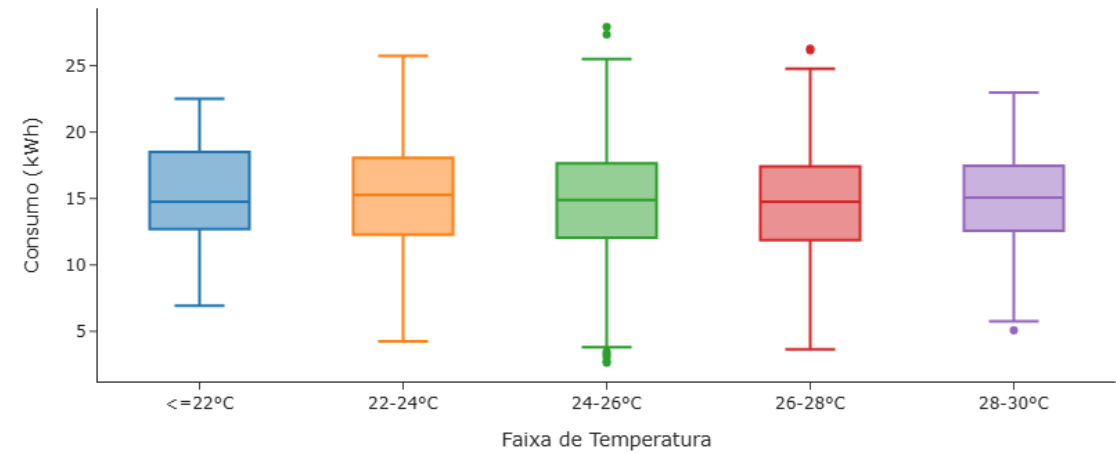


# ANÁLISE EXPLORATÓRIA

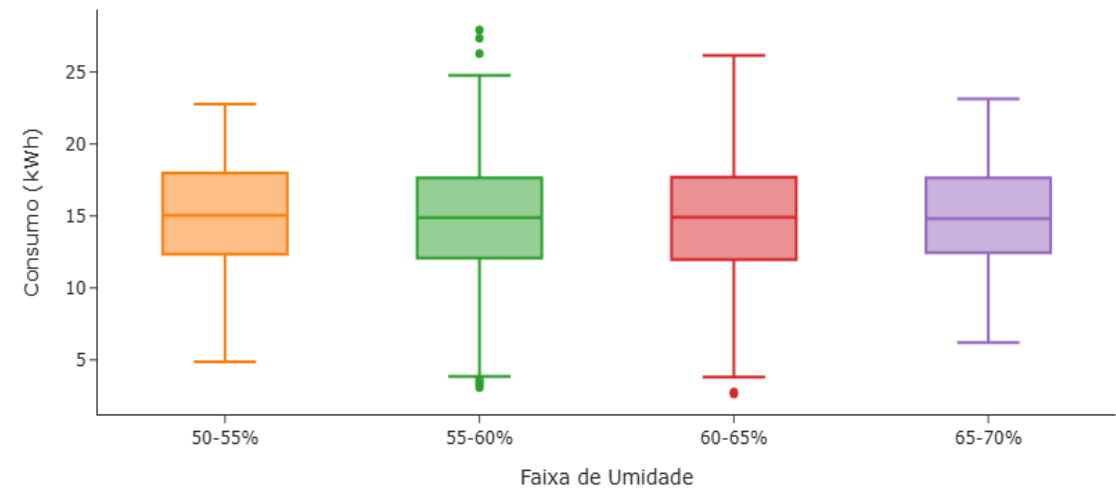
Consumo

As faixas de temperatura e umidade não tem diferenças entre a quantidade consumida de energia.

Boxplot do Consumo por Faixa de Temperatura



Boxplot do Consumo por Faixa de Umidade



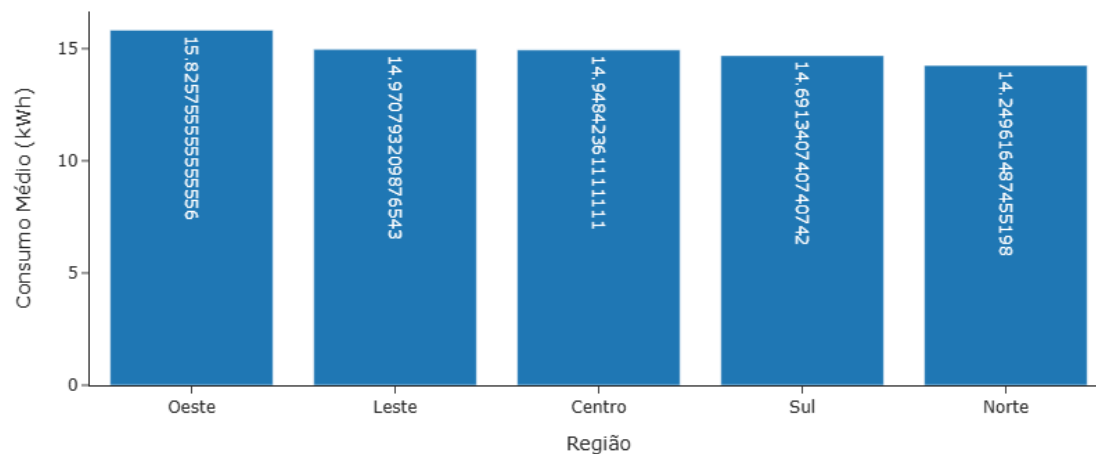


# ANÁLISE EXPLORATÓRIA

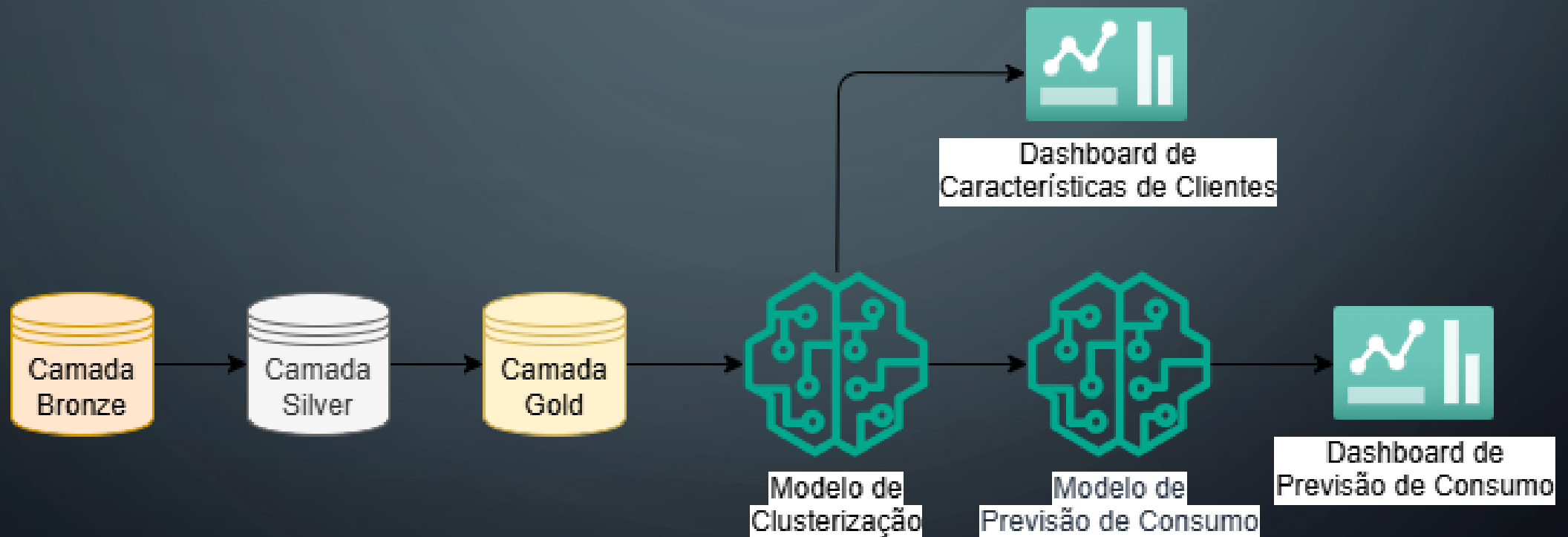
## Consumo

As regiões não possuem uma diferença significativa de consumo de energia, portanto não é um caracterizador forte para a série temporal.

Consumo Médio de Energia por Região



# PIPELINE DE PRODUÇÃO



# BANCO DE DADOS

O banco de dados utilizado foi o duckDB.

A arquitetura medallion foi escolhida para organização dos dados pois cada camada representa uma etapa de processamento e refinamento dos dados, garantindo qualidade, consistência e escalabilidade.

Os dados inicialmente estão em um CSV e após a criação, processamento e refinamento ficam disponíveis na camada gold para serem utilizados em análises e treinamento de modelos

# MLOPS

As boas práticas de MLOps foram seguidas para versionamento e escalabilidade de modelos.

Cada modelo teve a separação clara de cada etapa em módulo.

A imagem representa a pipeline e organização dos módulos de um modelo:



# MLOPS

A abordagem “Champion” do MLFlow foi utilizada no modelo de segmentação de cliente. A cada novo modelo ele é comparado com o modelo “Champion atual” e caso sua métrica seja superior ele é escolhido como novo “Champion” e passa a ser utilizado na inferência.

Já para o modelo de previsão nenhuma abordagem de comparação entre versões foi escolhido pois a cada novo dia o modelo deve ser retreinado com novos dados, isso não se adequa a abordagem Champion.

Validação temporal apropriada para séries temporais: utilizamos uma divisão cronológica dos dados, garantindo que o modelo fosse treinado apenas com informações do passado e validado com dados futuros. Isso evita vazamento de informação e simula com fidelidade o cenário de produção, onde previsões são feitas com base em dados históricos.



# SEGMENTAÇÃO DE CLIENTES

O objetivo da segmentação de clientes é gerar cluster para clientes com padrões semelhantes.

## Benefícios:

- Aumento da performance preditiva
- Personalização de ofertas
- Planejamento regional
- Segmentação climática

# SEGMENTAÇÃO DE CLIENTES

Os dados utilizados foram médias de consumo, temperatura e umidade para cada cliente.

A métrica a ser utilizada foi Silhouette Score pois ela mede o quão bem cada ponto foi atribuído ao seu cluster. Quanto maior melhor.

Para a escolha dos modelos foram testados Kmeans e DBScan. O campeão foi Kmeans com métrica de 0.53 e 5 clusters.

# SEGMENTAÇÃO DE CLIENTES

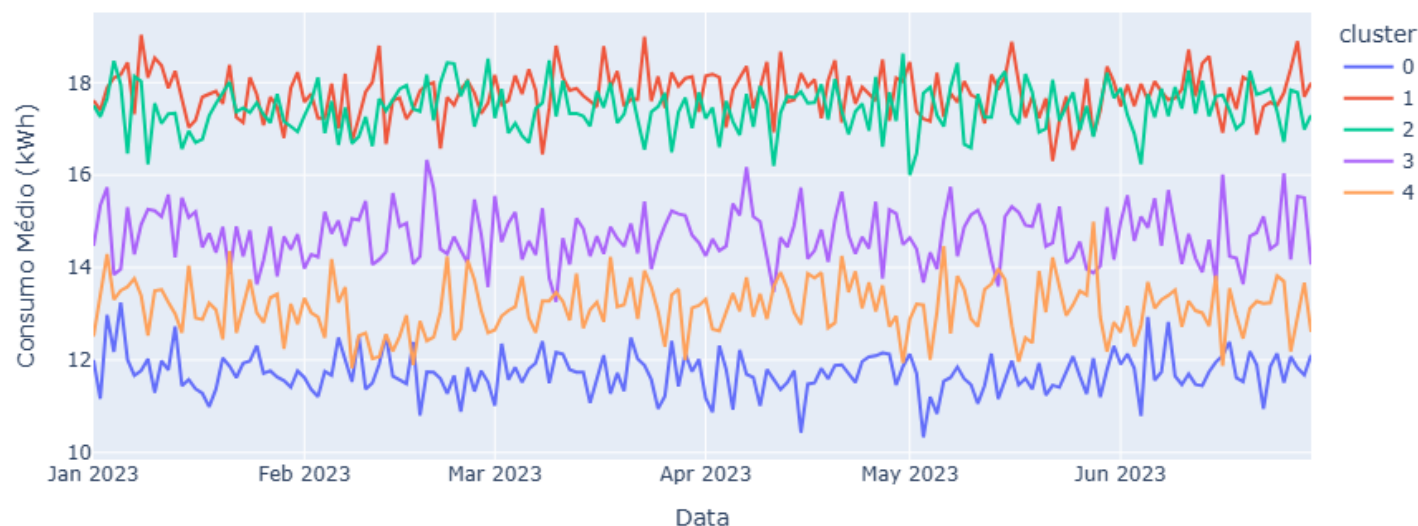
## Padrões de Consumo por Cluster:

Clusters 1 e 2 apresentam os maiores consumos médios diários, indicando grupos com clientes de maior demanda energética.

Clusters 0 e 4 têm os menores consumos, sugerindo clientes com comportamento mais econômico ou perfil residencial menor.

Cluster 3 está em um nível intermediário de consumo, podendo representar um perfil de cliente de classe média ou padrão misto.

Consumo Médio Diário por Cluster



# SEGMENTAÇÃO DE CLIENTES

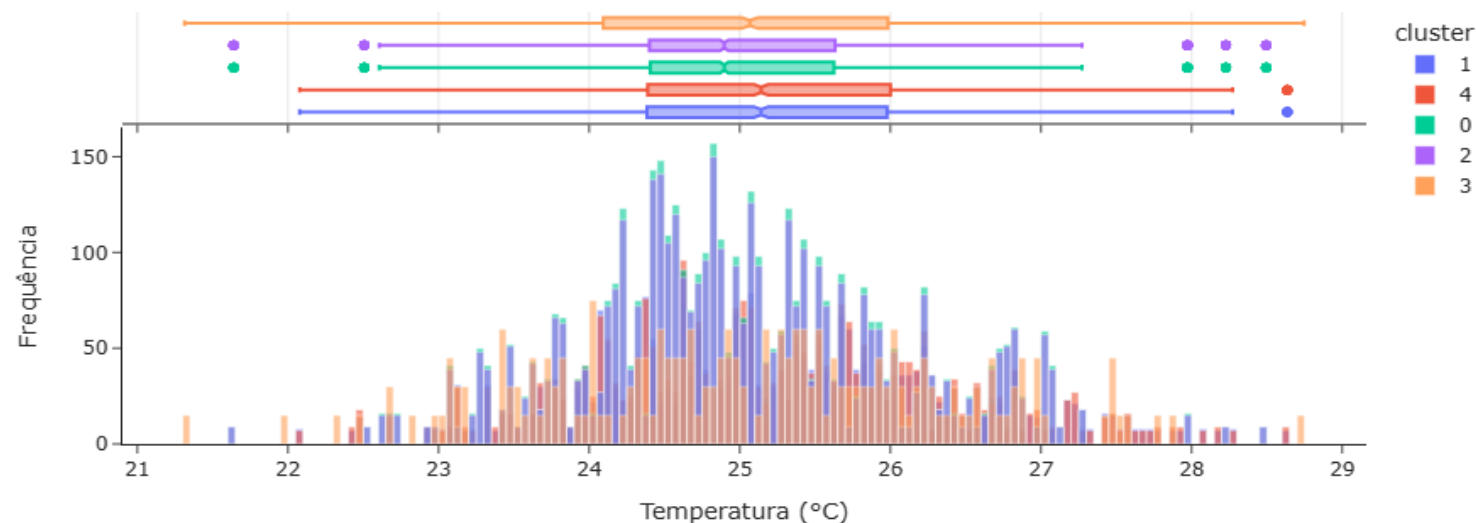
## Perfil Climático por Cluster:

Cluster 3 não possui faixa climática (temperatura e umidade) bem definida porém se concentra na região Sul.

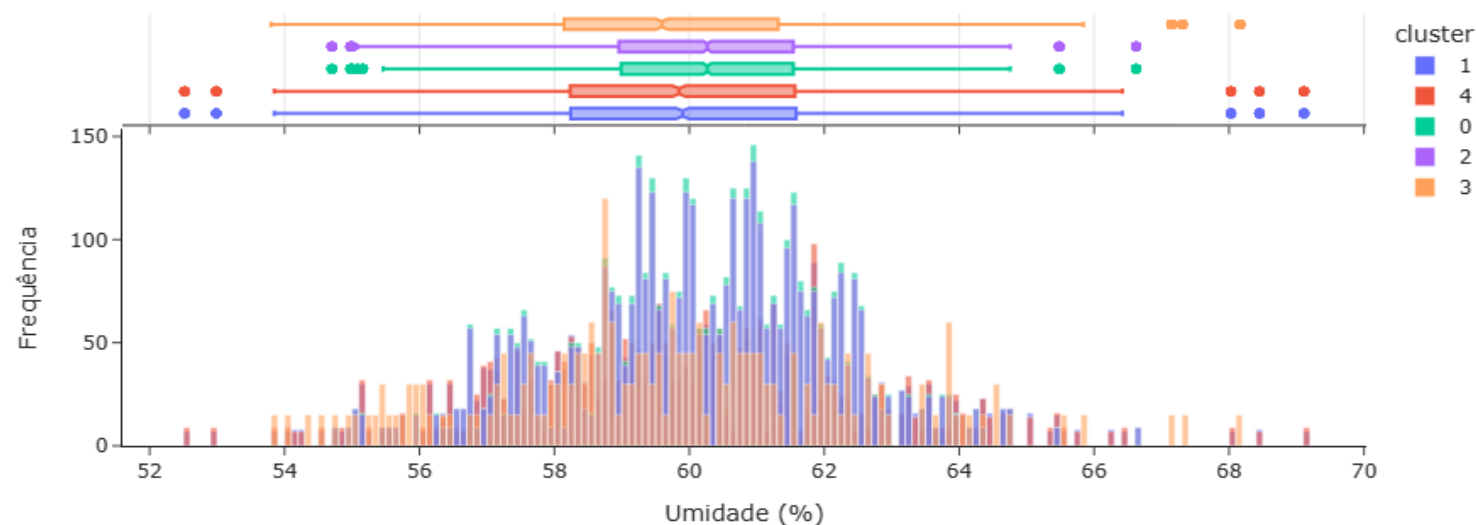
Clusters 1 e 4 compartilham faixas semelhantes de temperatura e umidade, o que pode indicar atuação em regiões com clima semelhante sendo Centro e Oeste.

O mesmo ocorre entre os clusters 0 e 2, que também compartilham perfis climáticos similares sendo Norte e Leste.

Distribuição da Temperatura por Cluster



Distribuição da Umidade por Cluster



# PREVISÃO DE CONSUMO

O objetivo da previsão é estimar o consumo energético de clientes residenciais para o próximo dia.

Benefícios:

- Precificação dinâmica
- Campanhas direcionadas
- Orçamento e controle financeiro
- Planejamento estratégico
- Alocação de recursos
- Antecipação de gargalos



# PREVISÃO DE CONSUMO

O objetivo da previsão é estimar o consumo energético de clientes residenciais para o próximo dia.

Benefícios:

- Precificação dinâmica
- Campanhas direcionadas
- Orçamento e controle financeiro
- Planejamento estratégico
- Alocação de recursos
- Antecipação de gargalos

# PREVISÃO DE CONSUMO

Os dados utilizados foram dados estatístico de clima e classificação como temperature ou umidade alta,média e baixa, variabilidade e tendência de consumo por cliente e features de tempo para consumo, temperatura e umidade.

As métricas a serem utilizadas são: erro média absoluto (MAE), raiz do erro quadrático médio (RMSE), normalização da raiz do erro quadrático médio (NRMSE) e coeficiente de determinação ( $R^2$ ).

Para a escolha dos modelos foram testados XGBoost, RandomForest e LightGBM. O campeão foi XGBoost com métrica NRMSE de 0.15.

Foi gerado um modelo para cada combinação cluster e região.

# PREVISÃO DE CONSUMO

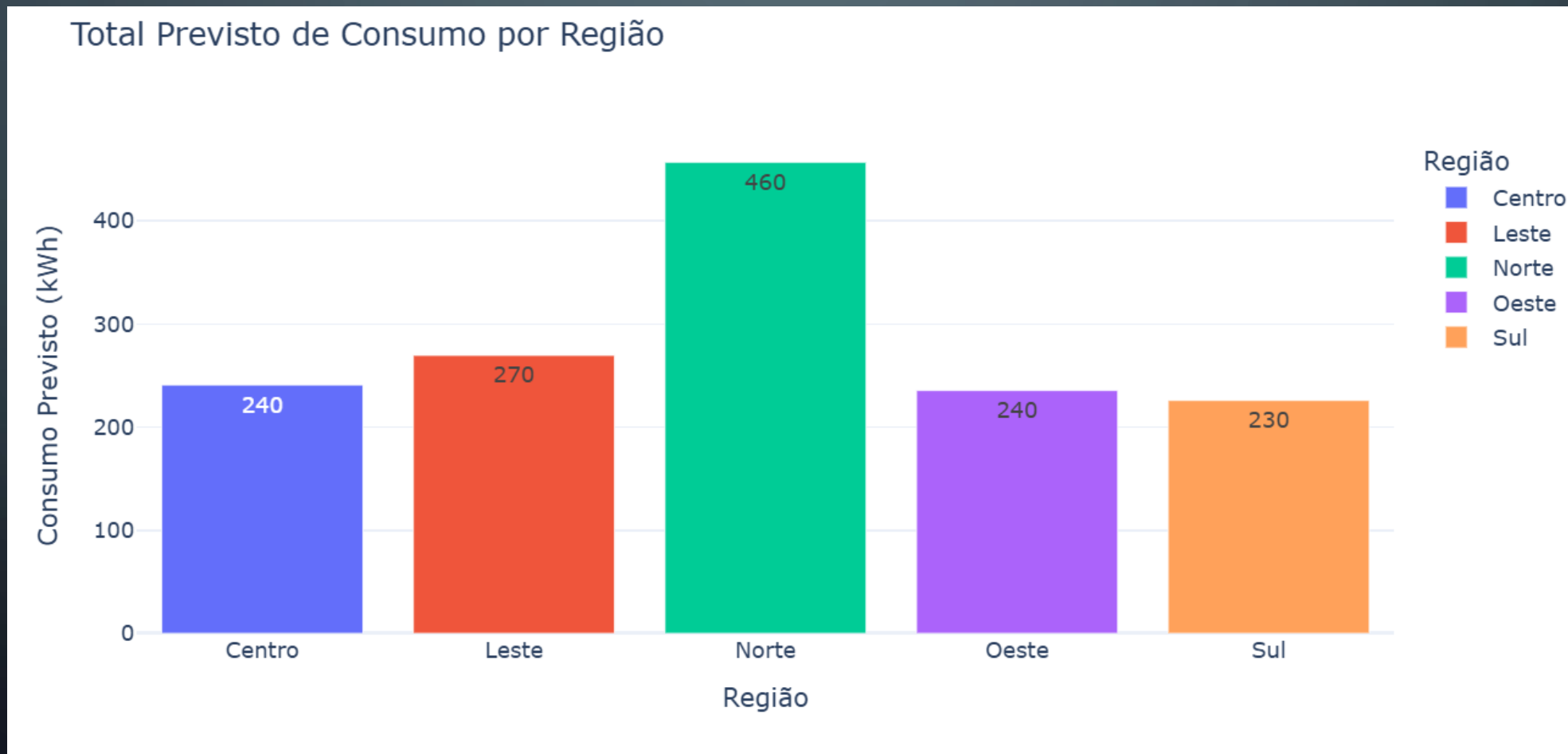
Os dados utilizados foram dados estatístico de clima e classificação como temperature ou umidade alta,média e baixa, variabilidade e tendência de consumo por cliente e features de tempo para consumo, temperatura e umidade.

As métricas a serem utilizadas são: erro média absoluto (MAE), raiz do erro quadrático médio (RMSE), normalização da raiz do erro quadrático médio (NRMSE) e coeficiente de determinação ( $R^2$ ).

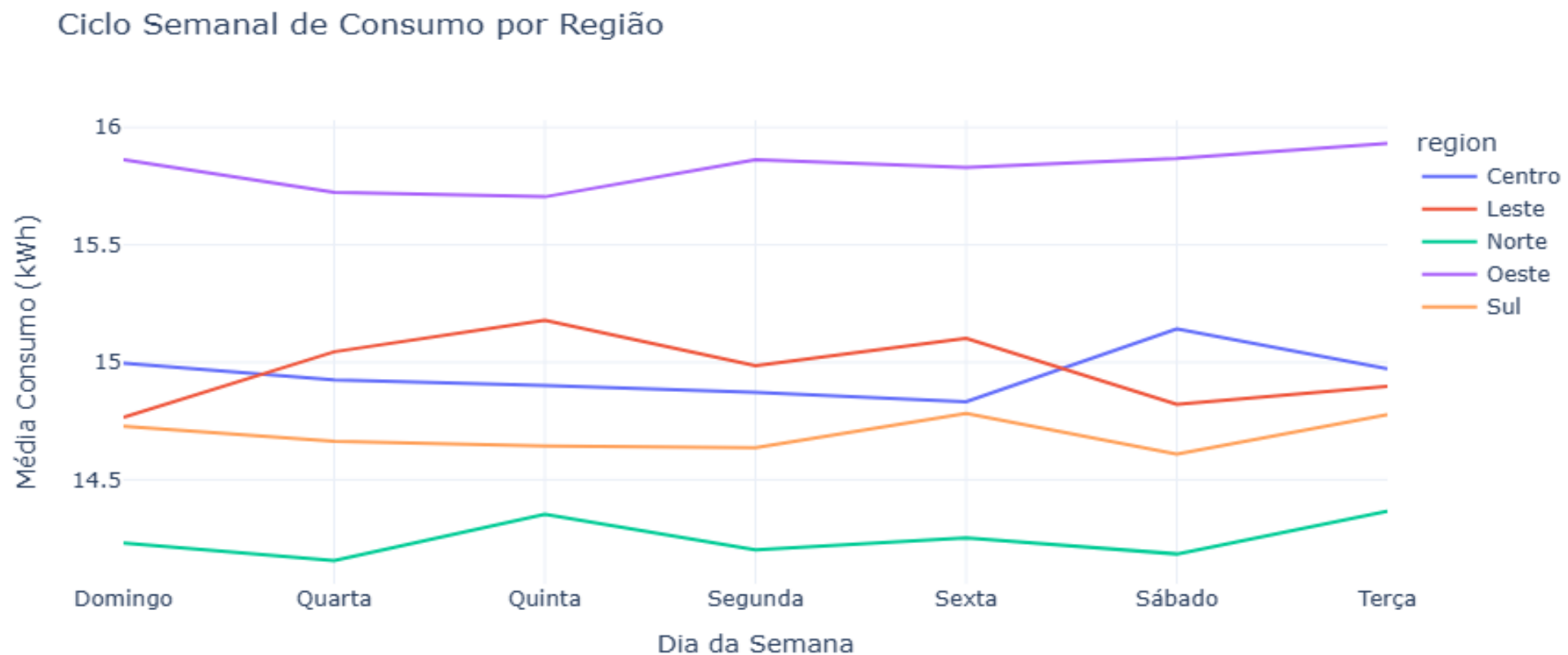
Para a escolha dos modelos foram testados XGBoost, RandomForest e LightGBM. O campeão foi XGBoost com métrica NRMSE de 0,15%.

Foi gerado um modelo para cada combinação cluster e região.

# PREVISÃO DE CONSUMO



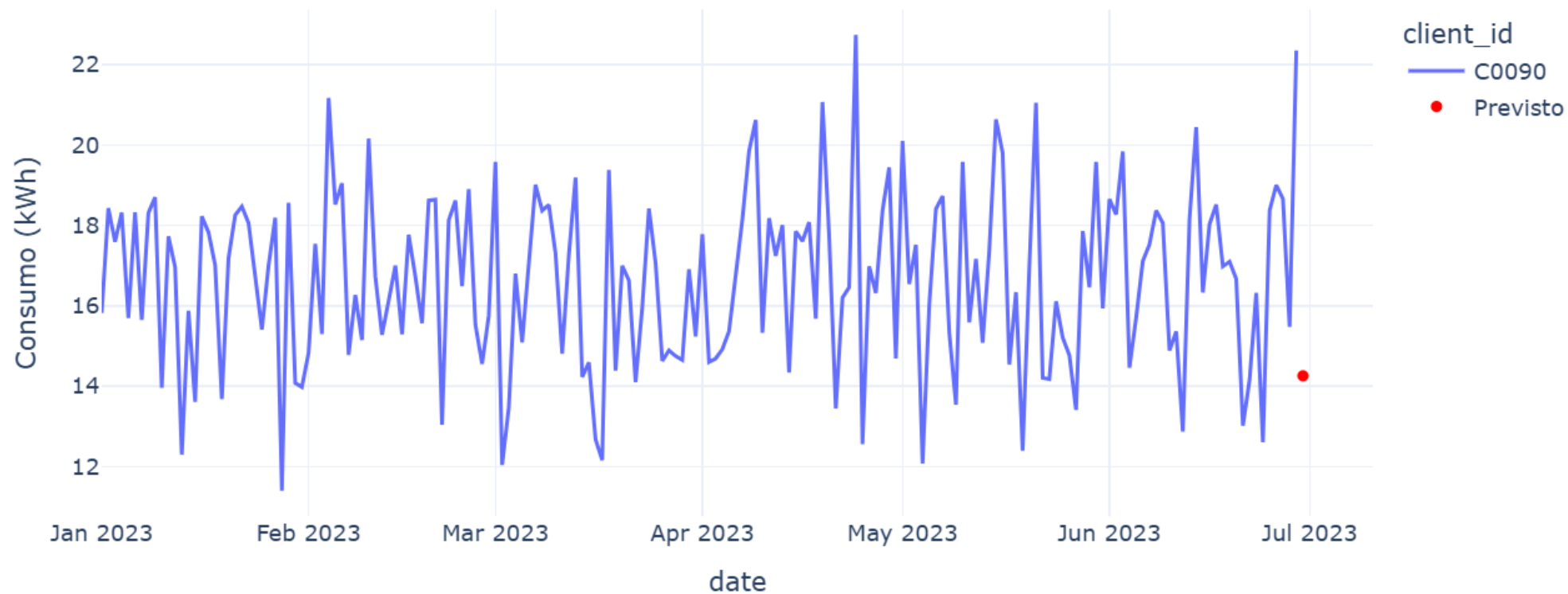
# PREVISÃO DE CONSUMO





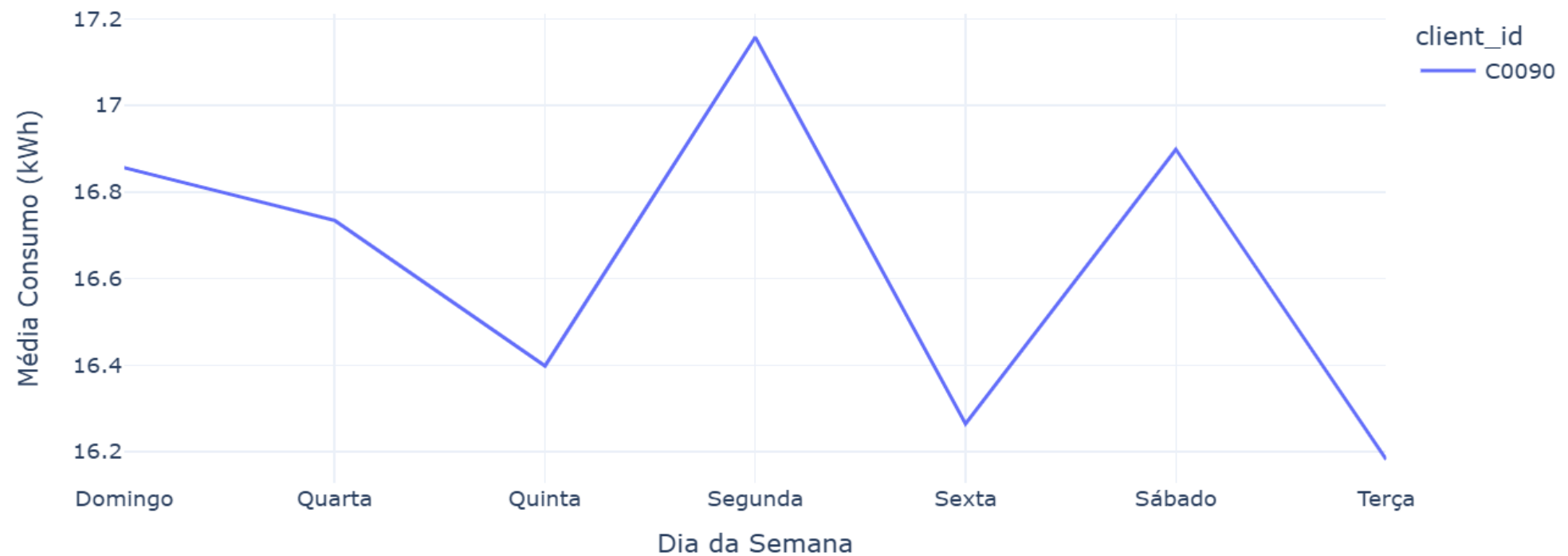
# PREVISÃO DE CONSUMO

Consumo Real e Previsto por Cliente



# PREVISÃO DE CONSUMO

Ciclo Semanal de Consumo por Cliente



# PRÓXIMOS PASSOS

**Uso de variáveis socioeconômicas:** incorporar dados como renda média, número de moradores ou tipo de imóvel pode aumentar o poder preditivo do modelo e permitir análises mais ricas.

**Avaliação de impacto por política energética:** uso do modelo como ferramenta para simular efeitos de mudanças tarifárias ou campanhas de economia de energia por perfil de cliente.

**Agente de IA para suporte e tomada de decisão:** desenvolver um agente conversacional baseado em IA que auxilie analistas ou consumidores finais na interpretação das previsões, simulações de consumo futuro, alertas sobre consumo atípico ou recomendações personalizadas de economia de energia. Esse agente poderia ser integrado ao dashboard, oferecendo respostas em linguagem natural e decisões baseadas nos dados em tempo real.



## DETALHES DE CONTATO

[joao060220@gmail.com](mailto:joao060220@gmail.com)