

SPRINT 4

BIG D Modelo de Ensemble, Gráficos e Custo do projeto

Agenda



Objetivo
da Sprint



Arquitetura
com Expurgo



Mapear Dados e
Correlações



Método
Ensemble



Gráficos
iniciais



Mapeamento
de custo



Status
Report



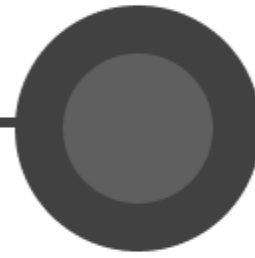
Próximos
Passos

Objetivo da Sprint



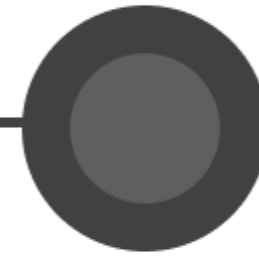
Sprint 1

Entendimento de
negócio, usuário e
arquitetura da solução



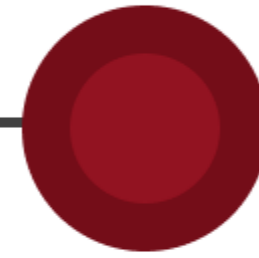
Sprint 2

Estrutura de Ingestão
de dados e Protipação
de baixa fidelidade



Sprint 3

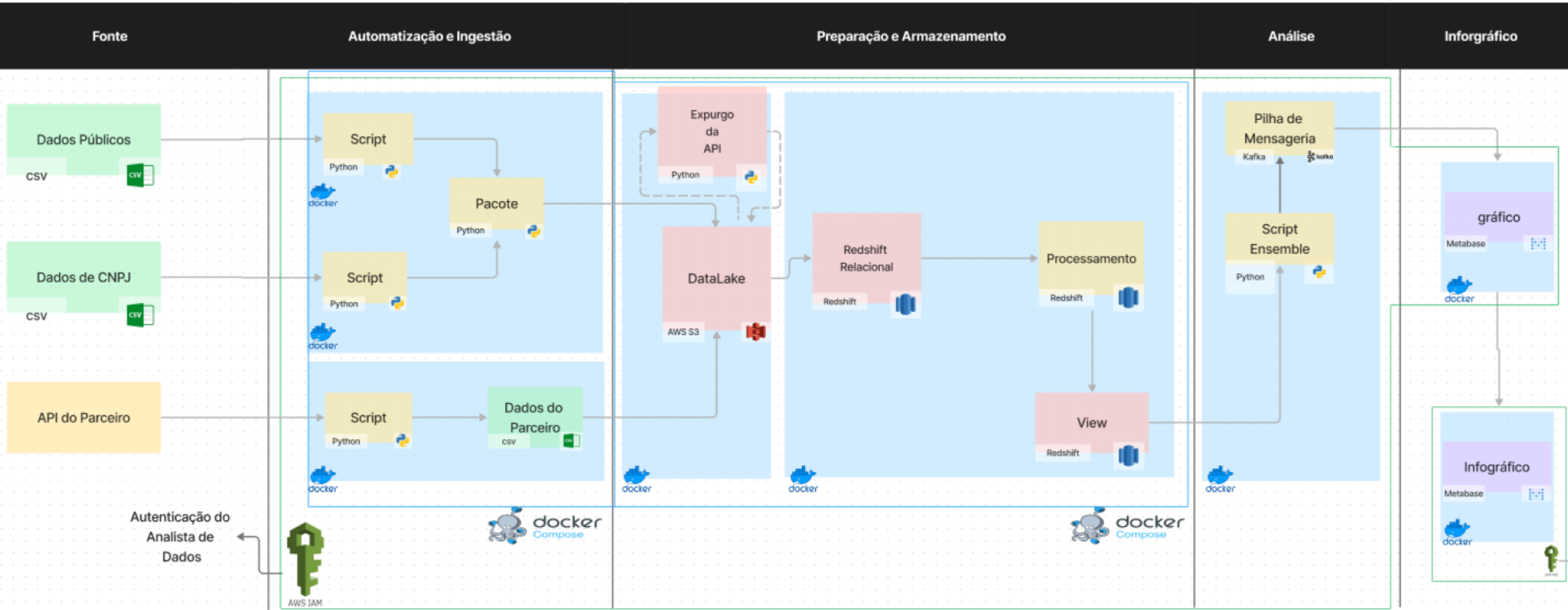
Criação do Data
Warehouse



Sprint 4

Método Ensemble,
Gráficos iniciais e
Custo do projeto

Arquitetura com Expurgo



Matriz de correlação

POF IPM-NM
(Índice de Pobreza Multidimensional -
Não Monetária)



Mapear Dados e Correlações

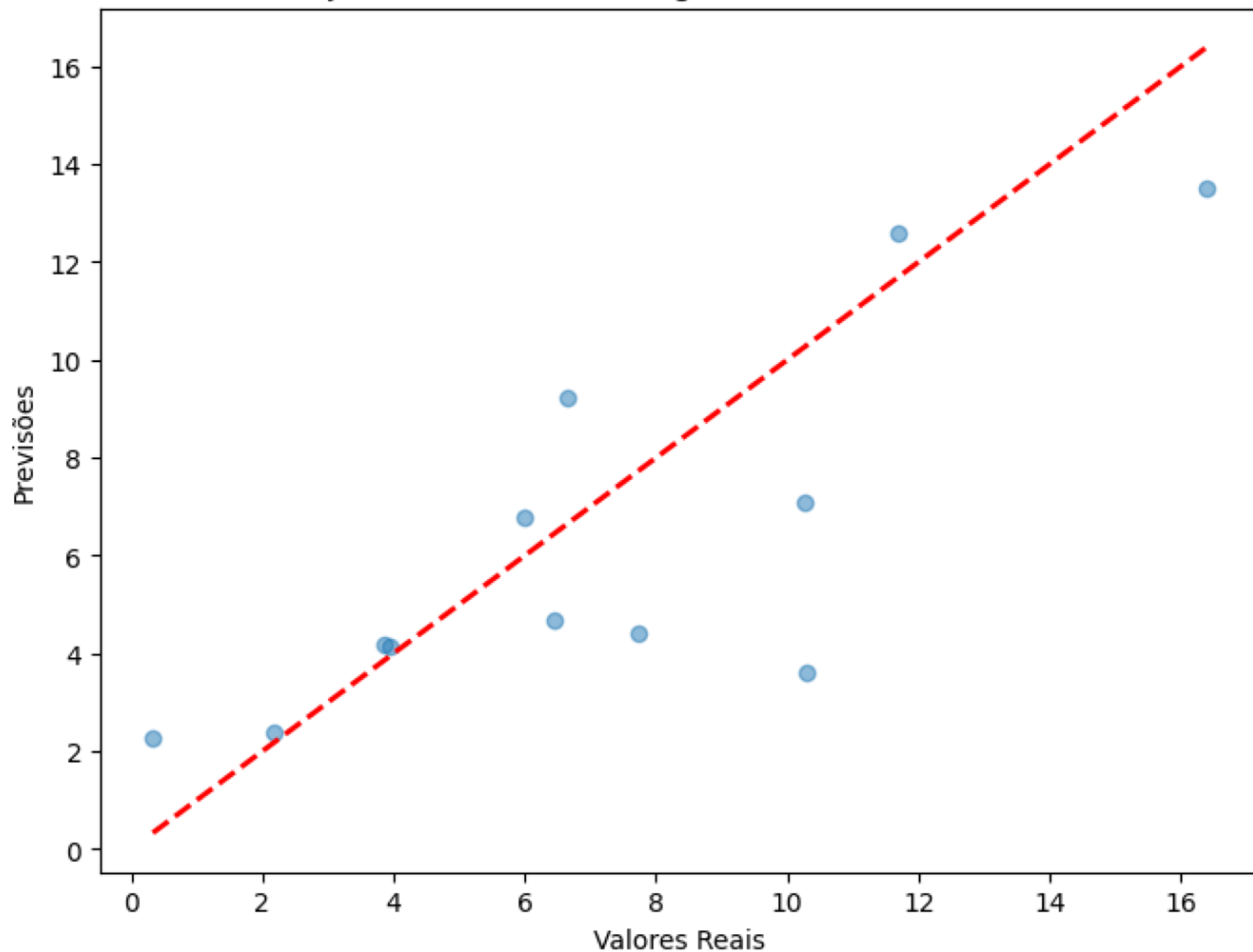
Método Ensemble

Modelagem: Uso do **algoritmo Random Forest**, com avaliação por MSE, e R^2 , etc.

Quanto mais próximo da linha vermelha, melhor a previsão do modelo.

Relação entre as variáveis independentes

Ajuste do Modelo de Regressão com Linha Ideal



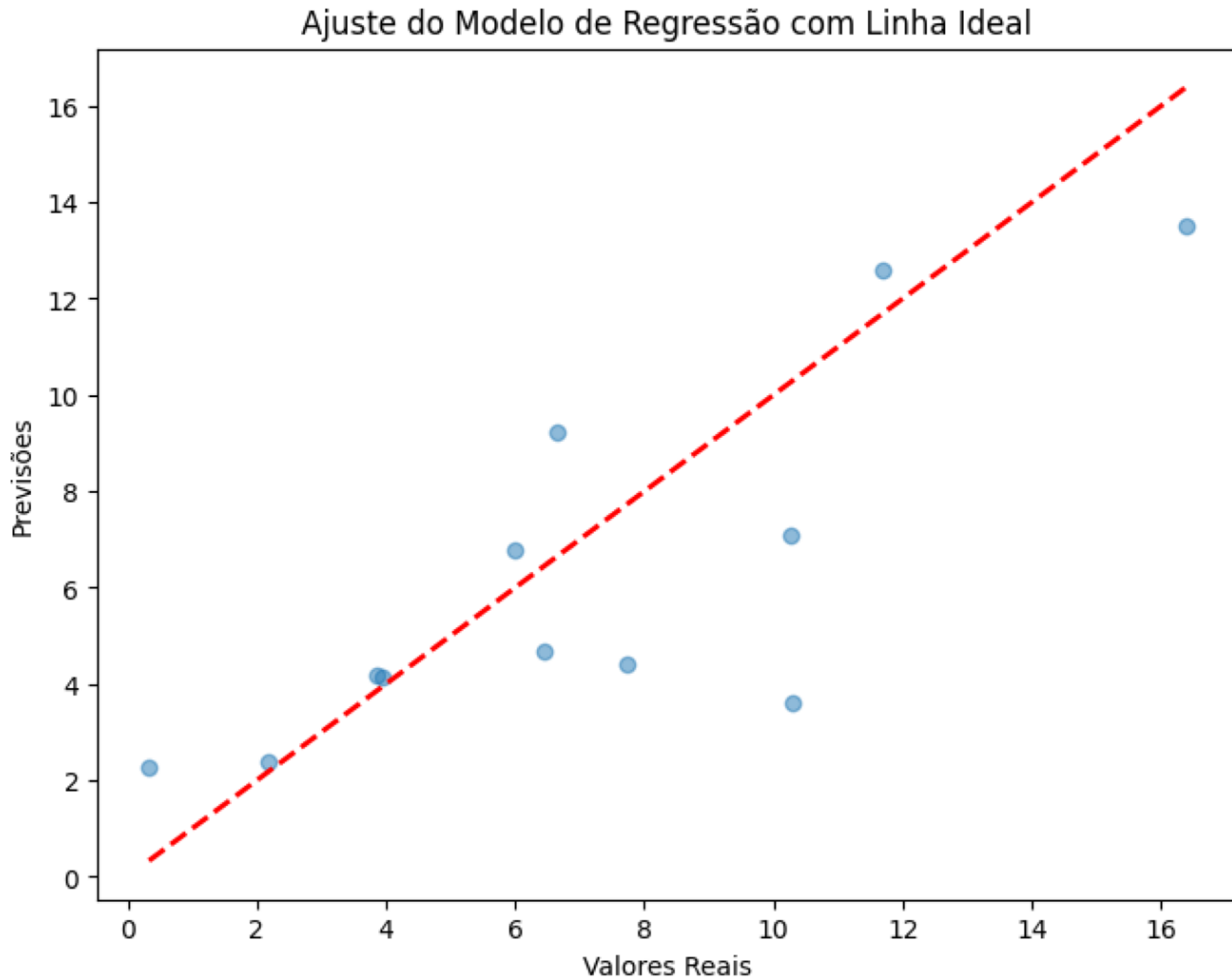
Método Ensemble

Mean Squared Error (MSE): 20.44

média dos quadrados dos erros entre as previsões do modelo e os valores reais

R-squared (R^2): 0.55

Proporção da variabilidade na variável de resposta que é explicada pelo modelo.

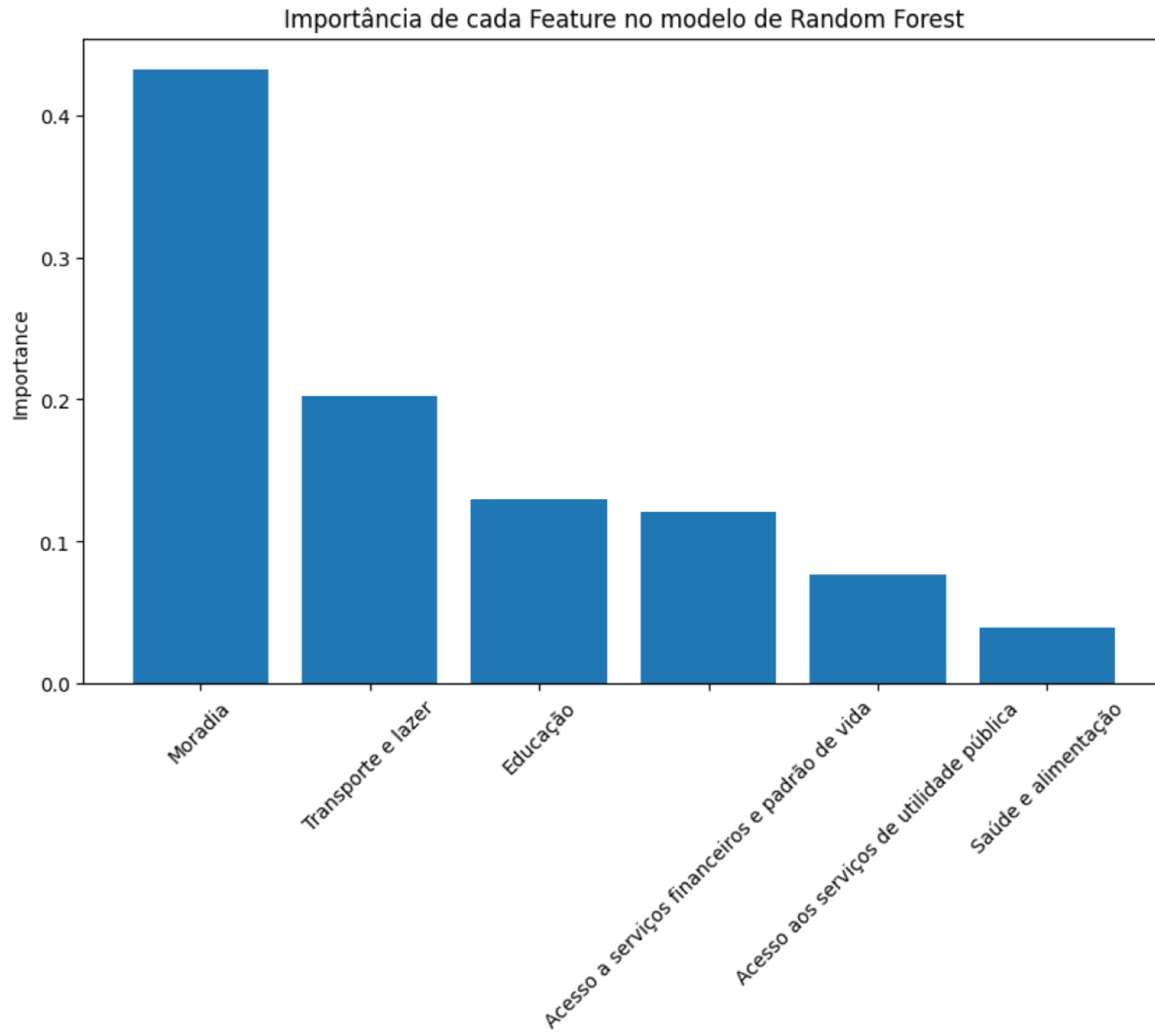


Método Ensemble

Modelagem: Uso do **algoritmo Random Forest**, com avaliação por MSE, MAE e R^2 .

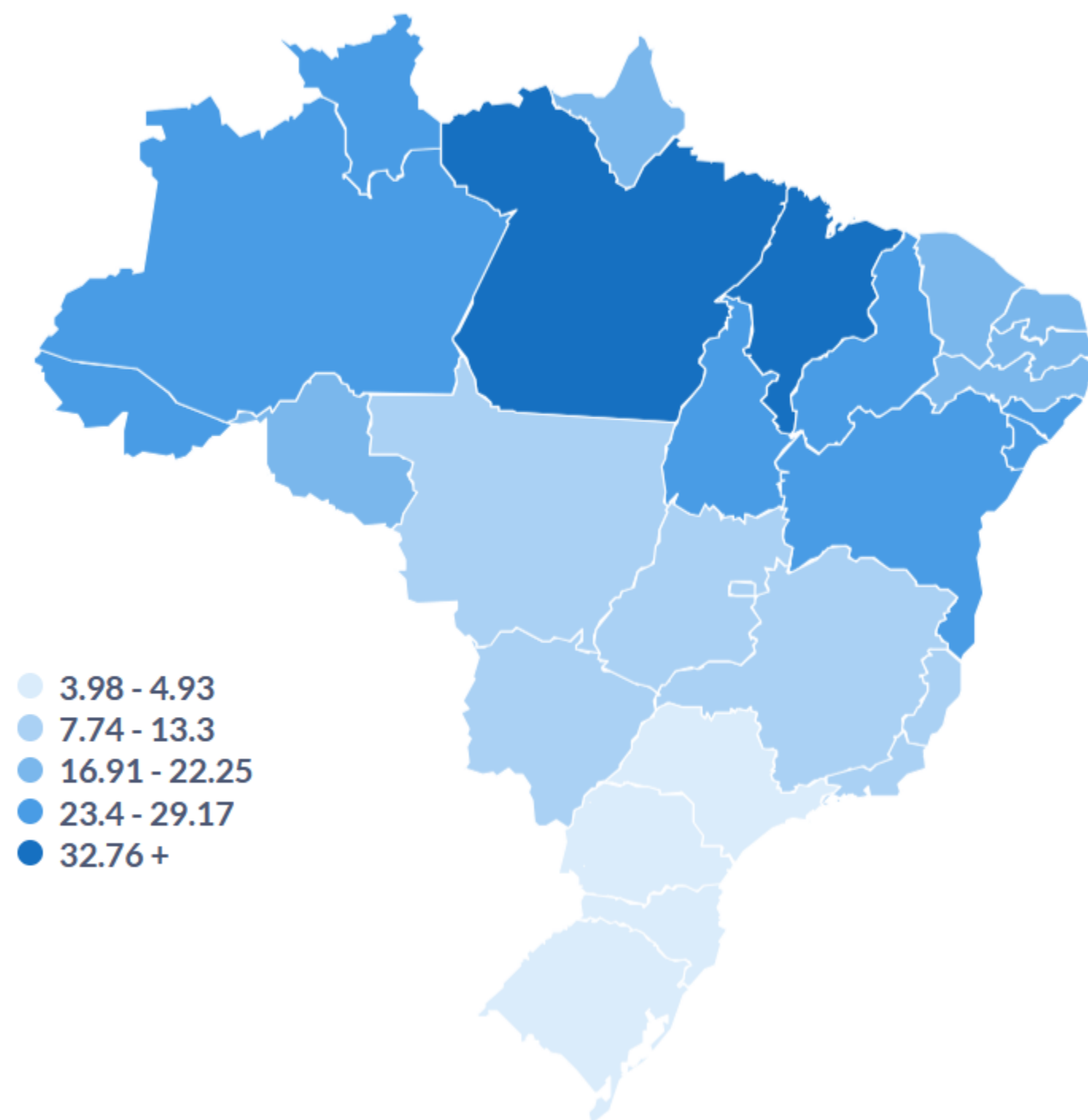
Cruzamento com Matriz de Correlação

Entender a natureza dos dados

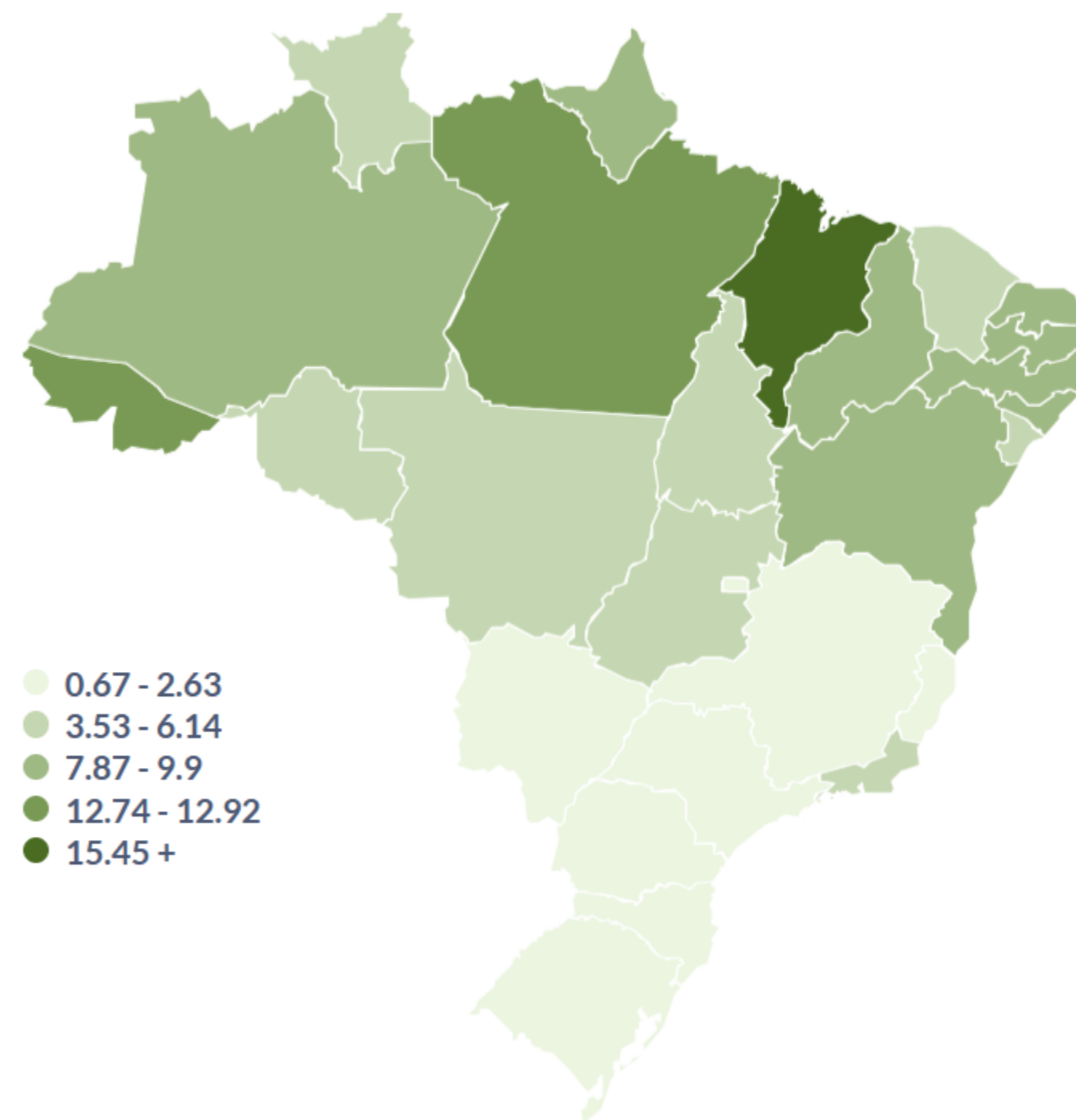


Gráficos iniciais

MAPA (IPM-NM) 2008-2009 POR UF



MAPA (IPM-NM) 2017-2018 POR UF



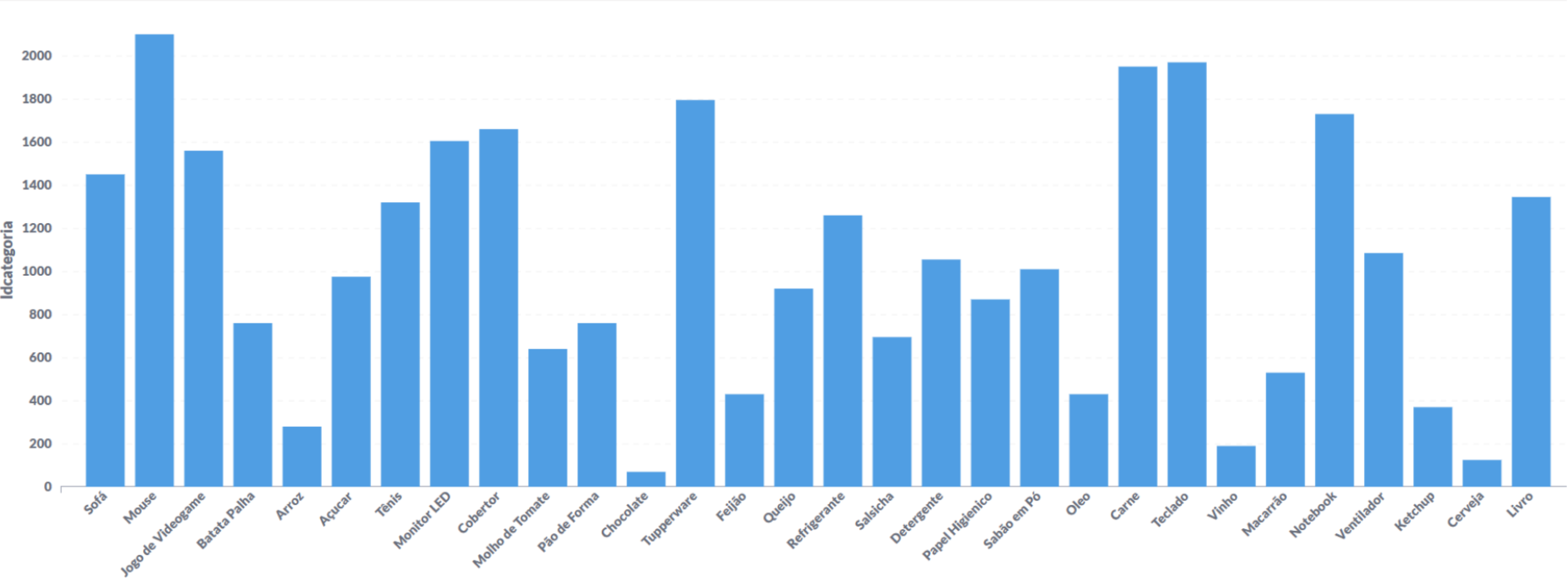
Gráficos iniciais



Categoria Mais Lucrativa Por Cnpj1

Filtrar

Sumarizar



Mapeamento de custo

- Custo Inicial: \$0
 - Custo Mensal: **\$646.09**
 - Custo Total em **12 Meses: \$7,753.08**
 - Moeda: USD
-

Serviços Incluídos:

- Amazon Redshift: \$527.04/mês
 - S3 Standard: \$0.57/mês
 - AWS Lambda: \$0/mês
 - Amazon EC2: \$118.26/mês
-

Total nos Primeiros 12 Meses:

- Amazon Redshift: \$6,324.48
 - S3 Standard: \$6.84
 - AWS Lambda: \$0
 - Amazon EC2: \$1,419.12
-

Principais Configurações dos Serviços AWS:

- Amazon Redshift: 16 RPU's, 3h de execução/dia
 - S3 Standard: 22 GB de armazenamento/mês
 - AWS Lambda: 50 solicitações/dia
 - Amazon EC2: Instância t3a.medium, Plano de 3 anos
-


Projeto | Cliente


Arquitetura Cloud | Integration


Data


08/12/2023

Indicadores

 Concluído

















 Em andamento

 Atrasado

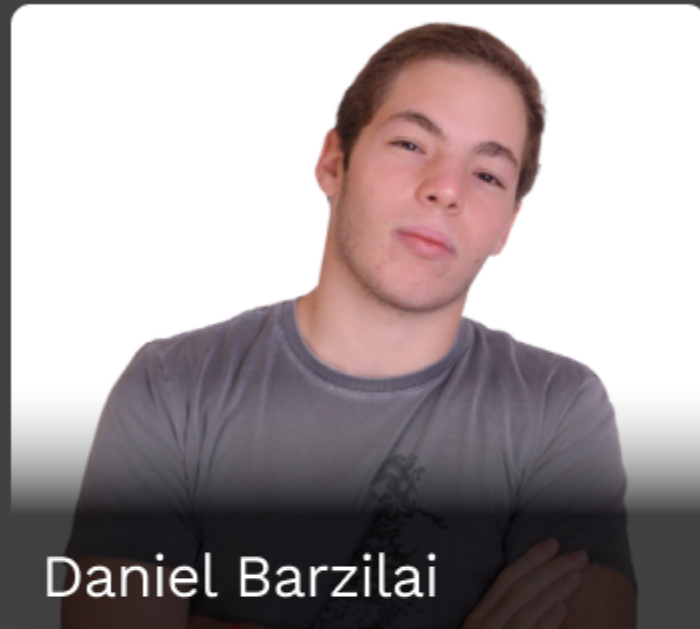
 Pausado

Objetivo - Sprint 4

Modelo de Ensemble, Gráficos e Custo do projeto

Processo	Status	Responsável	Observação	Indicador
Curadoria dos dados	Concluído	Grupo Big D.	Listar as fontes e categorias de dados	
Adicionar dados nos buckets	Concluído	Grupo Big D.	Configurar e provisionar buckets	
Análise de custo do projeto	Concluído	Grupo Big D.	Criar documento com custos das ferramentas mensalmente	
Criar Redshift e IAM Role	Concluído	Grupo Big D.	Implementar ferramentas na AWS	
Criar dicionário para tabelas POF	Concluído	Grupo Big D.	Criação de código para automatizar ajustes em tabelas	
Criar plano de comunicação	Concluído	Grupo Big D.	Documentar planos e estratégias do projeto	
Conexão entre Redshift e Endpoint	Concluído	Grupo Big D.	Implementar ferramentas na AWS	
Refatorar pacotes Python para Ingestão	Concluído	Grupo Big D.	Revisar códigos do projeto	
Script de tratamento dos dados	Concluído	Grupo Big D.	Criação de código para automatizar ajustes em tabelas	
Criar Metabase e conectar Redshift	Concluído	Grupo Big D.	Iniciar a criação de gráficos do projeto	
Criar Notebook Ensemble	Concluído	Grupo Big D.	Código para implementar modelo de regressão	
Script de expurgo	Concluído	Grupo Big D.	Código para remover dados da API do cliente do banco de dados e ferramentas	
Atualizar Arquitetura	Concluído	Grupo Big D.	Arquitetura com expurgo	
Matriz de correlação	Concluído	Grupo Big D.	Artefato para identificar correlação entre features da tabela	
Criação Infográficos	Pausado	Grupo Big D.	Visualização de dados para criar um infográfico	
Relatório de análise de eficácia	Pausado	Grupo Big D.	Documentação para métrica da geração de infográficos	

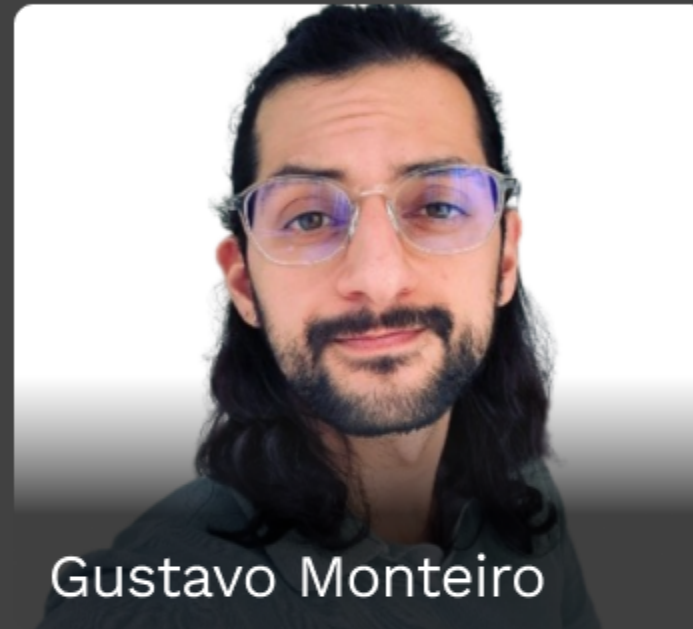
Time 3D



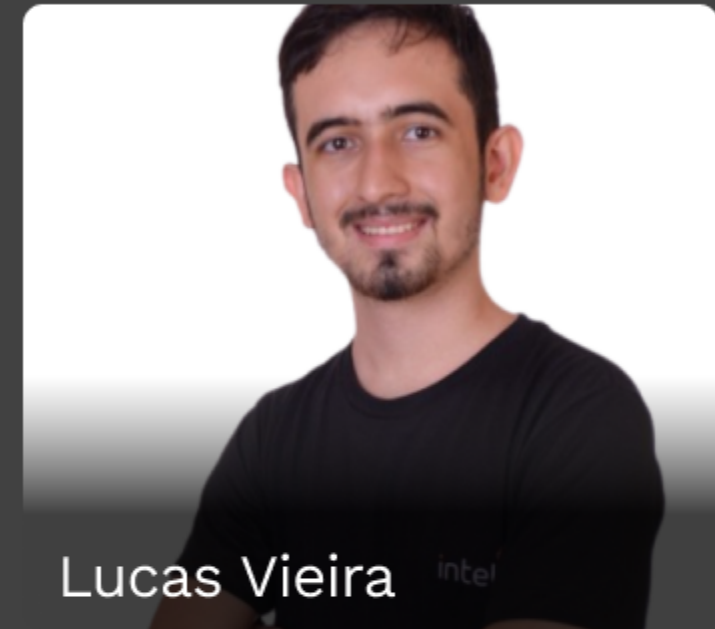
Daniel Barzilai



Dayllan Alho



Gustavo Monteiro



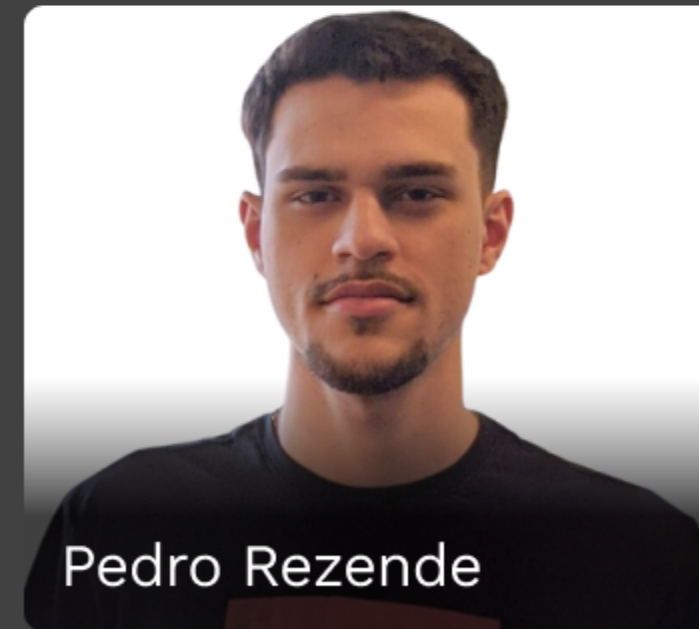
Lucas Vieira



Luiz Augusto



Matheus Fidelis



Pedro Rezende



Muito
Obrigado