



DATA DREAM

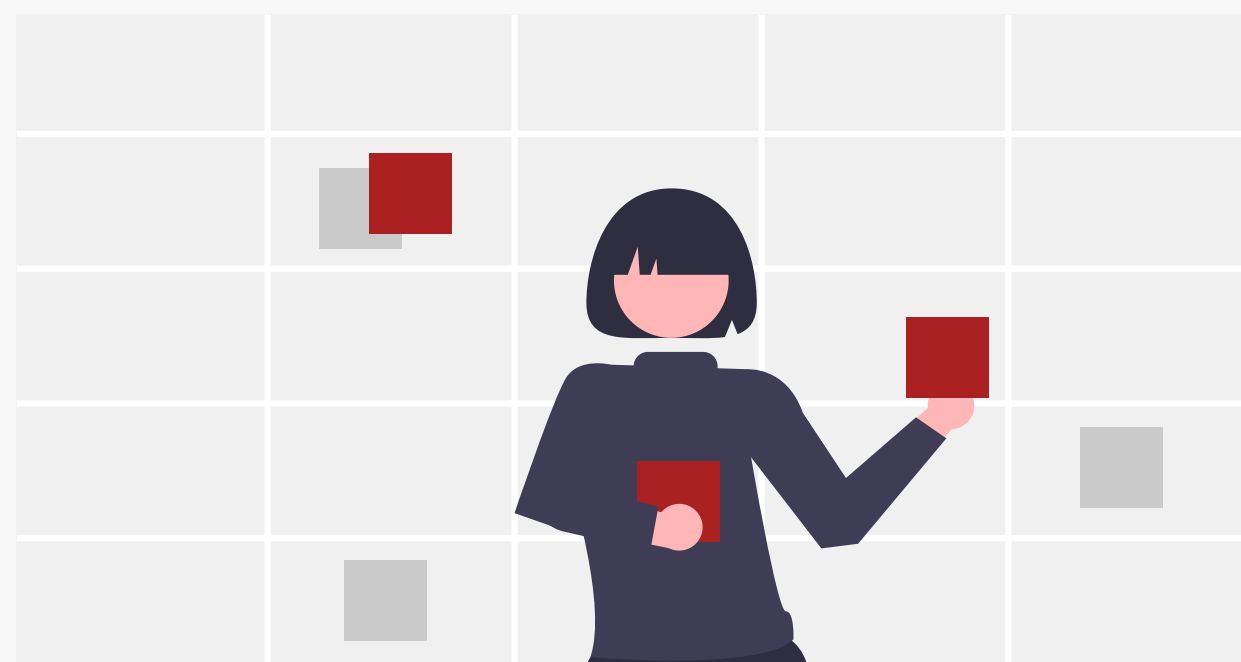
SPRINT REVIEW

INTEGRAÇÃO, GERENCIAMENTO E ANÁLISE DE BIG DATA



DATA DREAM

AGENDA



AJUSTES SPRINT 2

Ajustes com base no feedback

OBJETIVO SPRINT 3

Descrição do objetivo

NOVA ARQUITETURA

Apresentação da arquitetura reformulada

ESTRUTURA DO REDSHIFT

Explicação da utilização

SQL VIEW REDSHIFT

Demonstração das view's criadas

PROTOTIPAÇÃO

Exibição dos protótipos em alta fidelidade

PRÓXIMAS ENTREGAS

Artefatos que serão desenvolvidos na Sprint 4

STATUS REPORT

Cenário de desenvolvimento das tarefas

AJUSTES - SPRINT 2



DATA DREAM

Apresentado pelo parceiro:

Quando realizar o upload de dados CSV no AWS S3, os arquivos CSV localmente devem ser removidos automaticamente.

Realizado:

Foi implementada uma função que, após o carregamento dos dados no bucket, realiza a exclusão dos arquivos CSV localmente.

```
import os
import boto3

aws_access_key_id = os.getenv("aws_access_key_id")
aws_secret_access_key = os.getenv("aws_secret_access_key")
aws_session_token=""
region_name = 'us-east-1'
bucket_name = ''
csv_directory = './csv_tratados'

s3 = boto3.client('s3', aws_access_key_id=aws_access_key_id, aws_secret_access_key=aws_secret_access_key,
region_name=region_name, aws_session_token=aws_session_token)

csv_files = [file for file in os.listdir(csv_directory) if file.endswith('.csv')]

for csv_file in csv_files:
    local_csv_file = os.path.join(csv_directory, csv_file)
    s3_file_key = csv_file

    s3.upload_file(local_csv_file, bucket_name, s3_file_key)

    print(f'0 arquivo {csv_file} foi enviado para o Amazon S3 com a chave {s3_file_key}.')

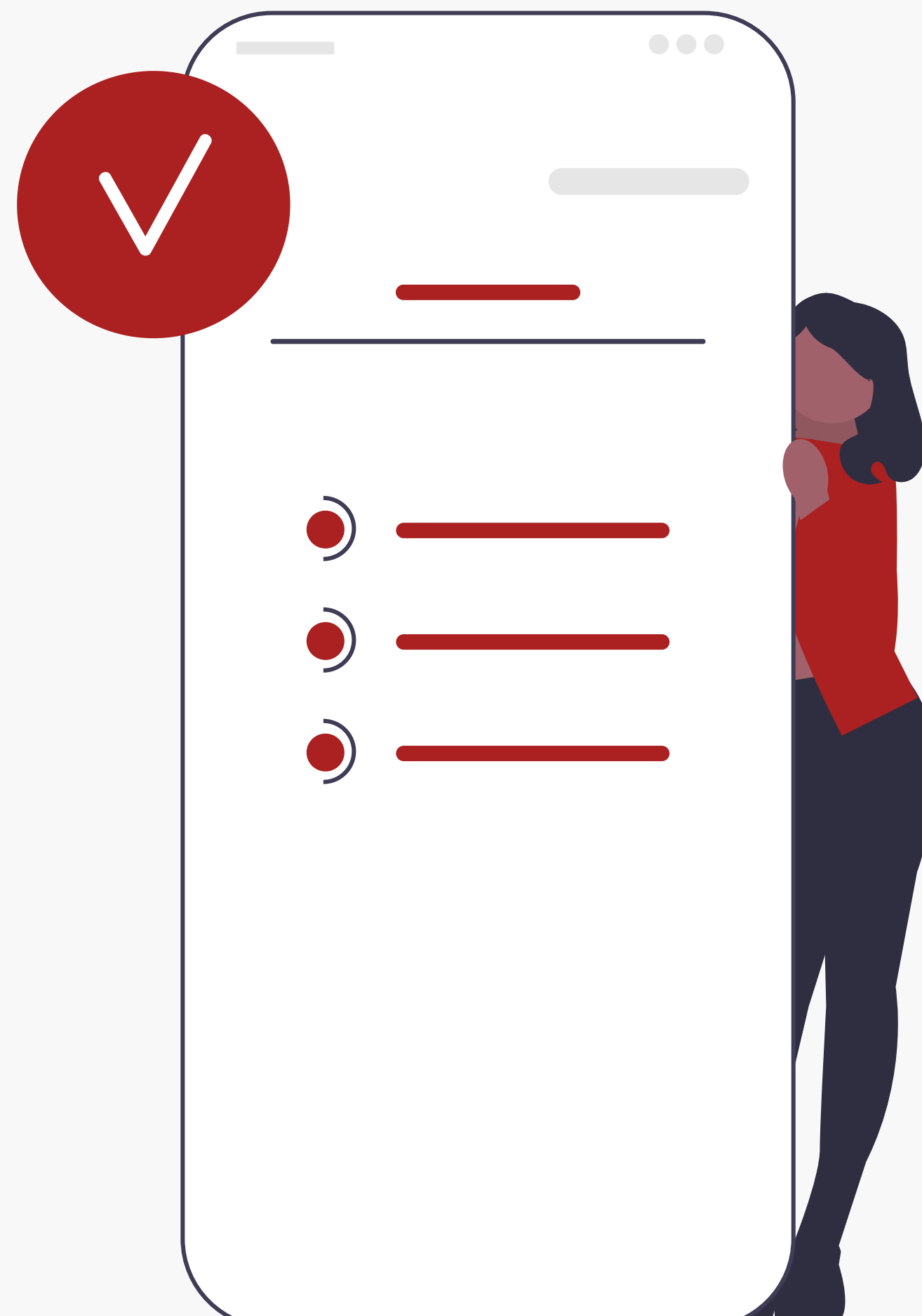
    os.remove(local_csv_file)
    print(f'0 arquivo local {csv_file} foi excluído.')

print('Todos os arquivos CSV foram processados, enviados para o Amazon S3 e os arquivos locais foram excluídos.')
```

OBJETIVO SPRINT REVIEW 3

Criação do Cubo de Dados

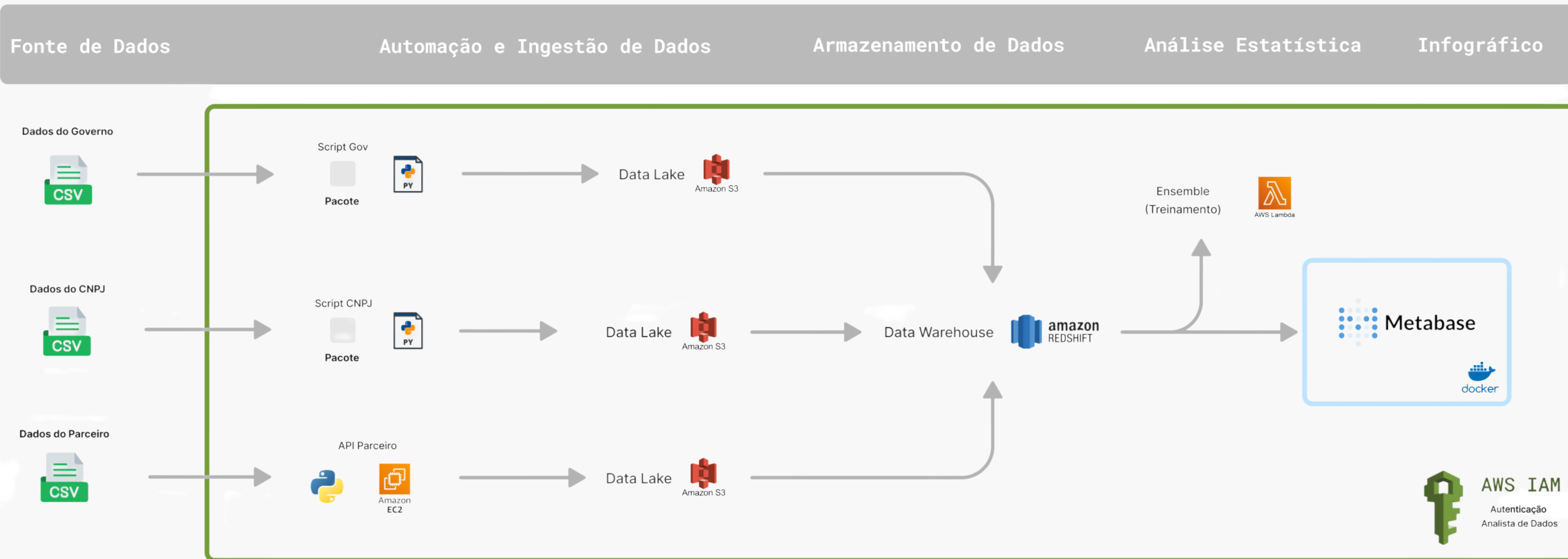
Data Lake/Data Warehouse Alimentado pelo ETL



ARQUITETURA



DATA DREAM



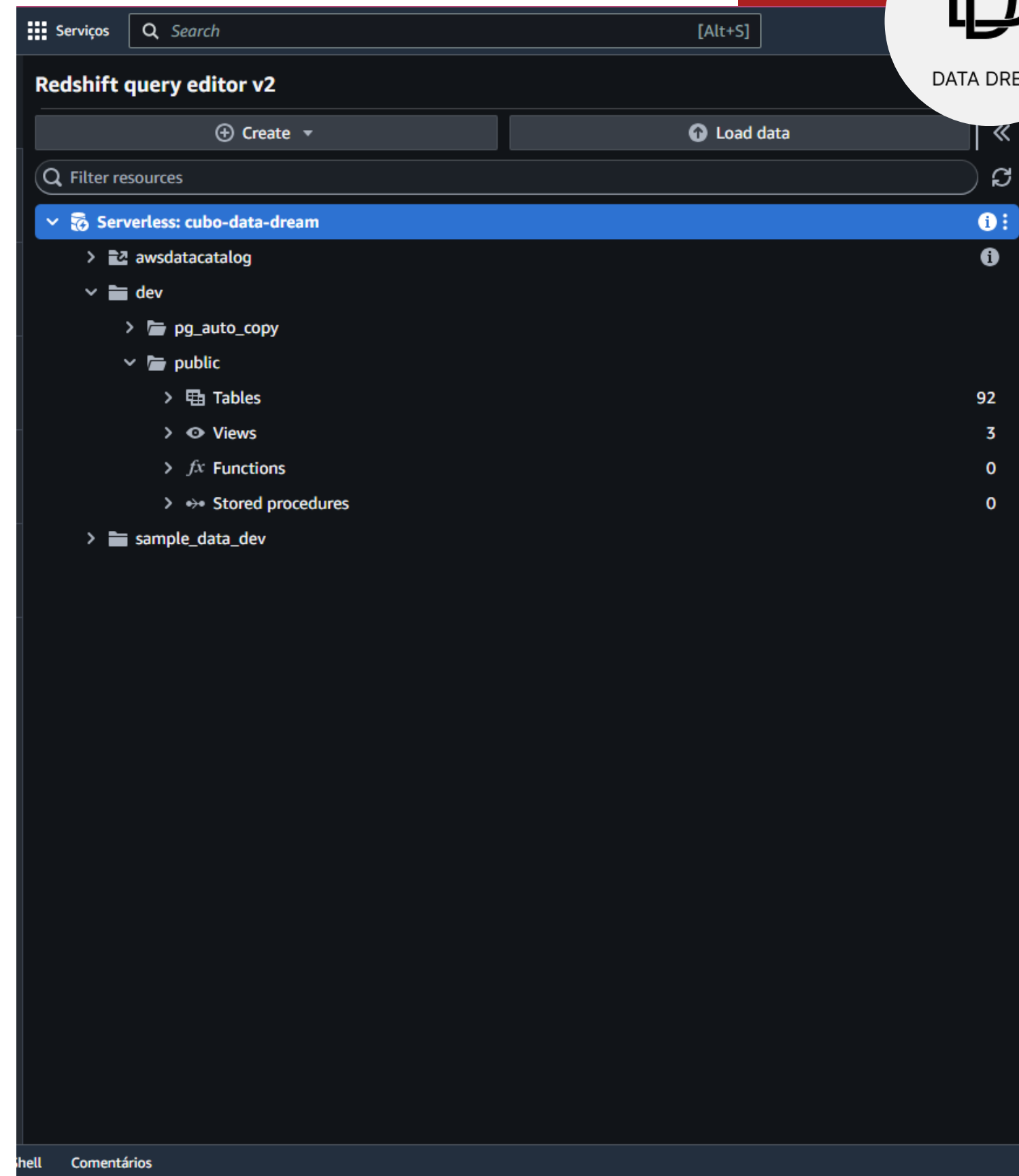
REDSHIFT

Cubo de dados



Serviço de data warehouse que consolida e armazena um grande volume de dados de diferentes fontes em um único local centralizado.

Para o contexto do atual projeto, o Redshift será utilizado como o ambiente OLAP, armazenando todas as views.





VIEW'S REDSHIFT

Demonstração

MELHORIAS CONTÍNUAS

Erro encontrado:

ERROR: Load into table 'inep-ana2016' failed. Check 'sys_load_error_detail' system table for details.

[ErrorId: 1-655e62dc-52331880e6f518133bbb401]

Solução:

Revisar o pipeline e aperfeiçoar a estrutura de ingestão





DESIGN

Prototipação Inicial

EXPLORE E
EXTRAIA
INSIGHTS



DATA DREAM



Faça Login com sua conta

Email

integration@gmail.com

Senha

[Esqueceu ?](#)

Entre com sua senha



Logar

MENU

Dashboard

Análises

FILTROS

PRODUTO

Selecione

REGIÃO

Selecione

DATA

Selecione

CANAL

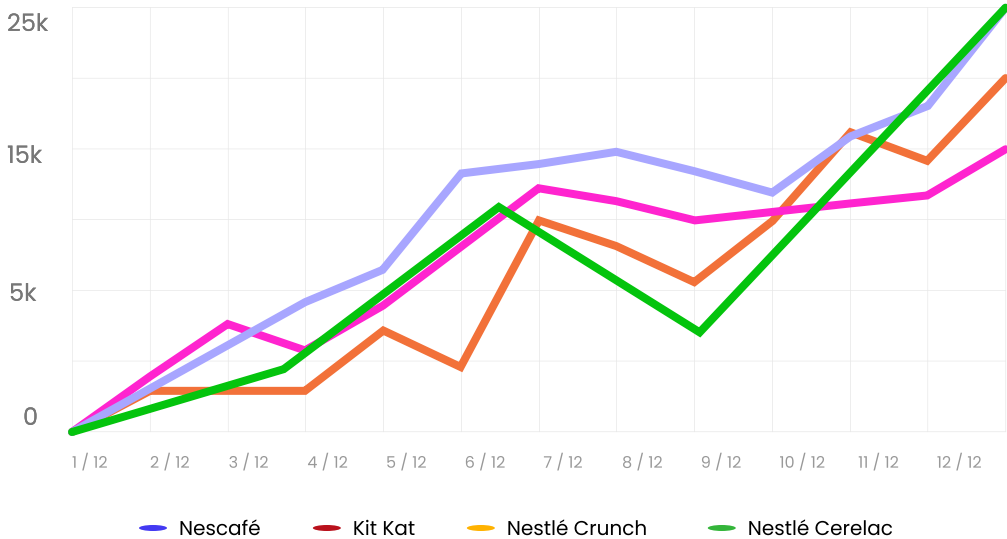
Selecione

Logout

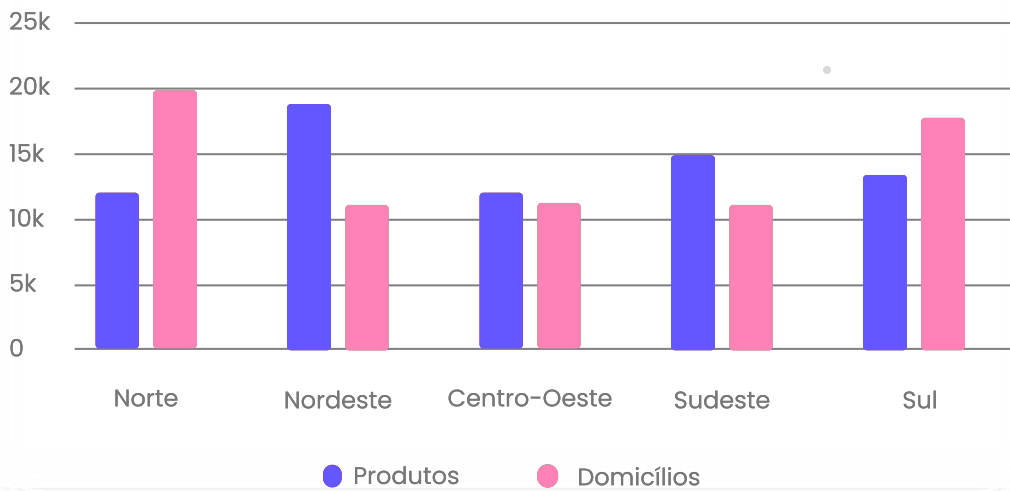
Dashboard de Tendências de Consumo

Entenda as Preferências dos Consumidores e Avalie o Desempenho dos Produtos no Mercado

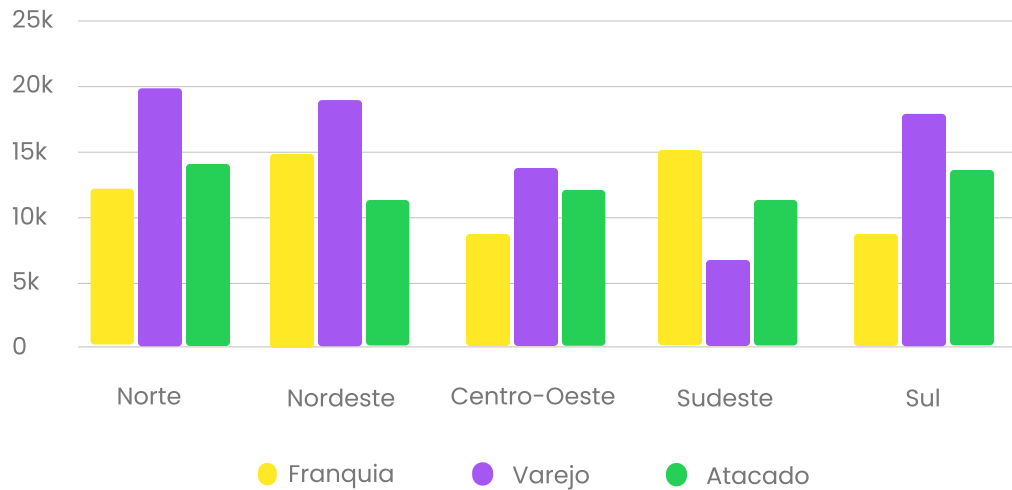
Vendas x Tempo



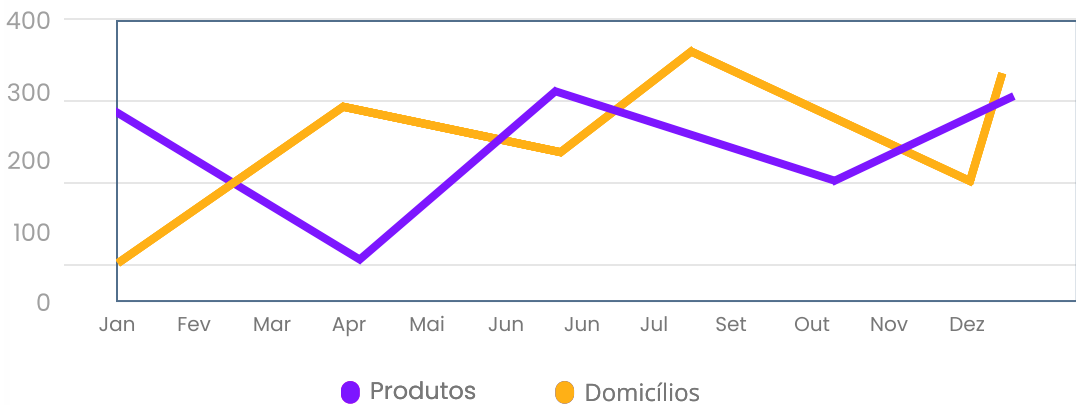
Quantidade de consumo de produtos X Região



Canal X Região X Venda



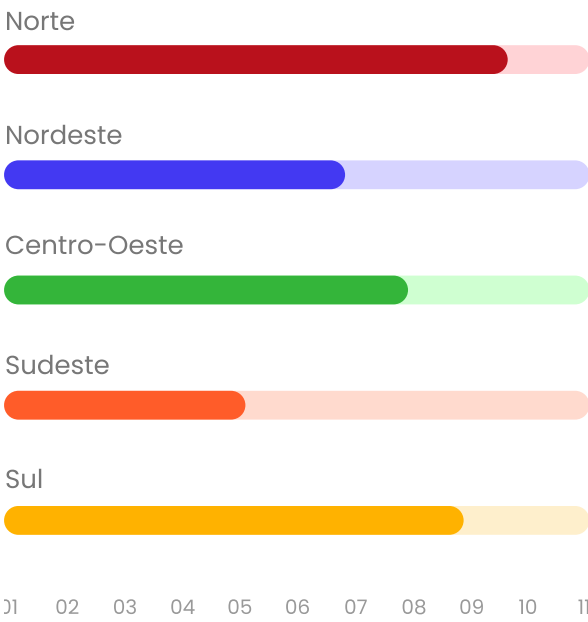
Quantidade de consumo de produtos X Tempo



Principal Categoria X Região

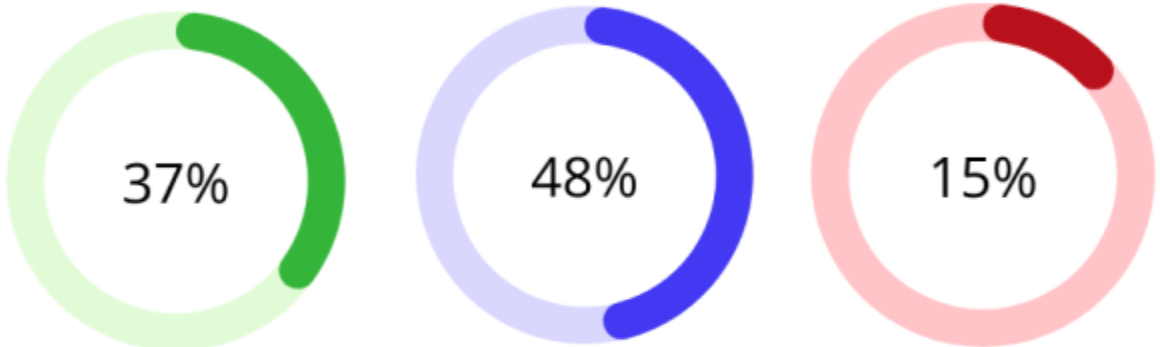


Quantidade de domicílios X Região



Quantidade de dados por fonte

Governo CNPJ Parceiro











Últimos dados atualizados

March, 01, 2020 #MS-415646	100kb	csv
February, 10, 2021 #RV-126749	250kb	csv
April, 05, 2020 #FB-212562	300kb	csv
June, 25, 2019 #QW-103578	105kb	csv

1-4 of 8

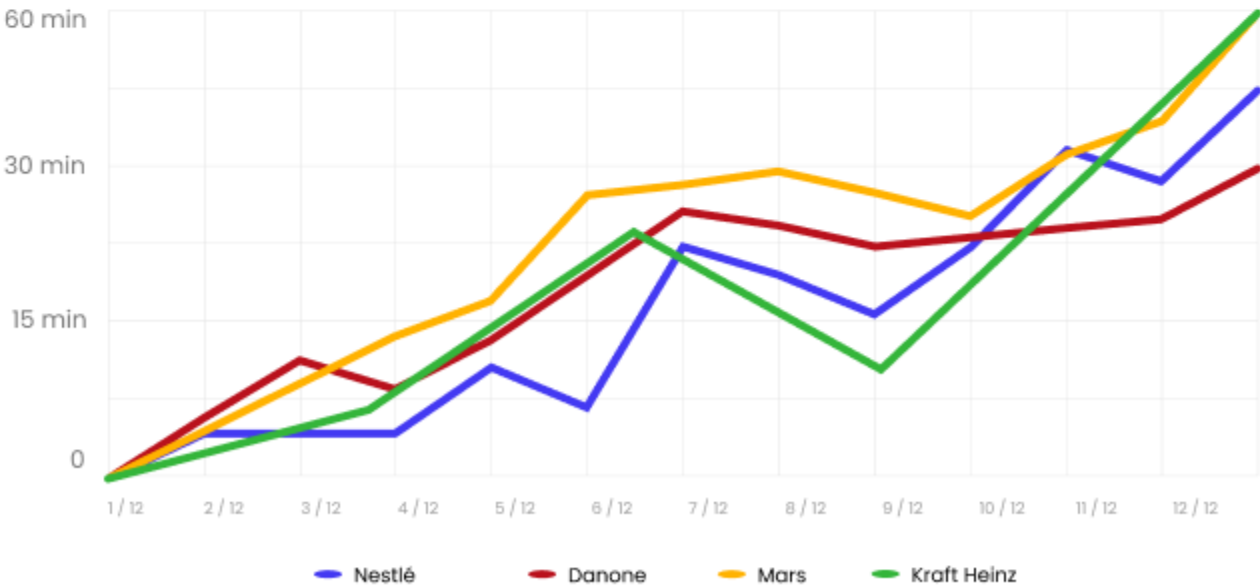
Empresas Analisadas

+ Nova Empresa

 Nestlé	Alimentício	 Kraft Heinz Company	Alimentício
 Mondelez Internation	Alimentício	 Unilever	Alimentício
 Danone	Alimentício	 Kellogg's	Alimentício
 Mars, Incorporated	Alimentício	 General Mills	Alimentício

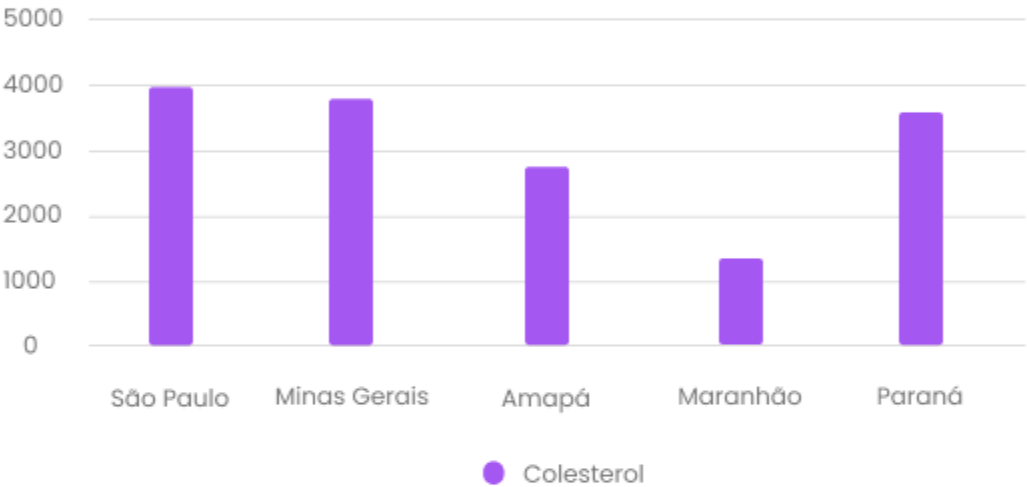
1-8 of 24

Análises X Tempo de Execução

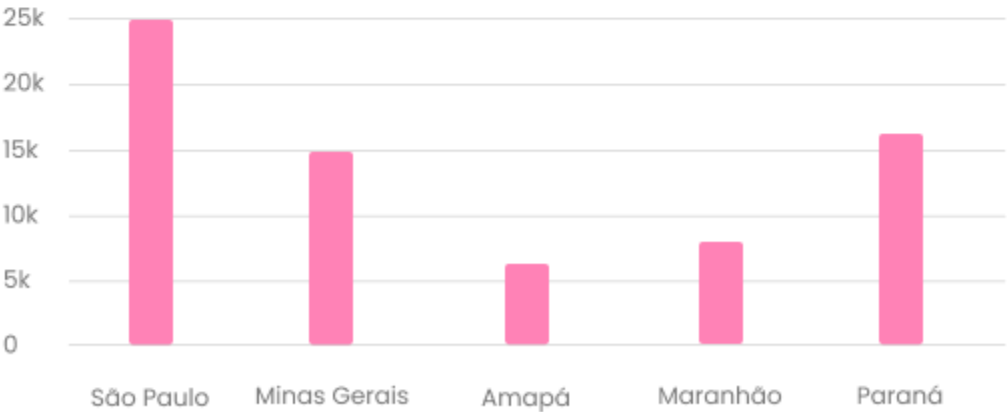


Como vender chocolate?

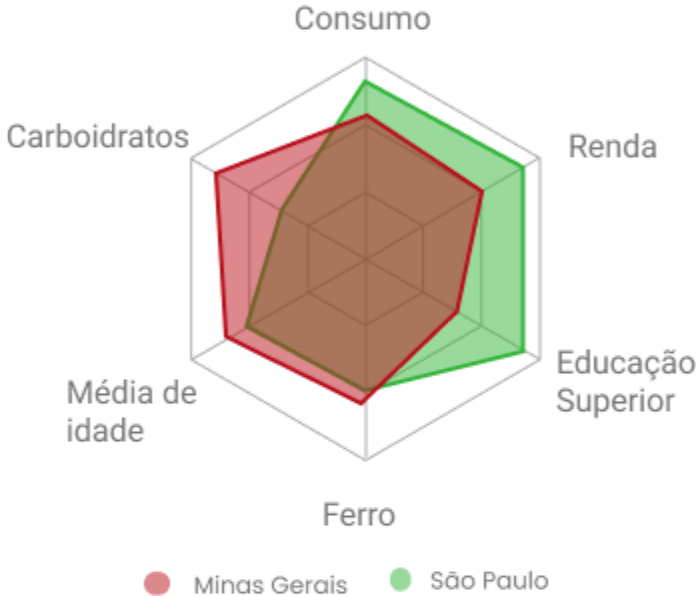
1 Estado X Consumo alimentício



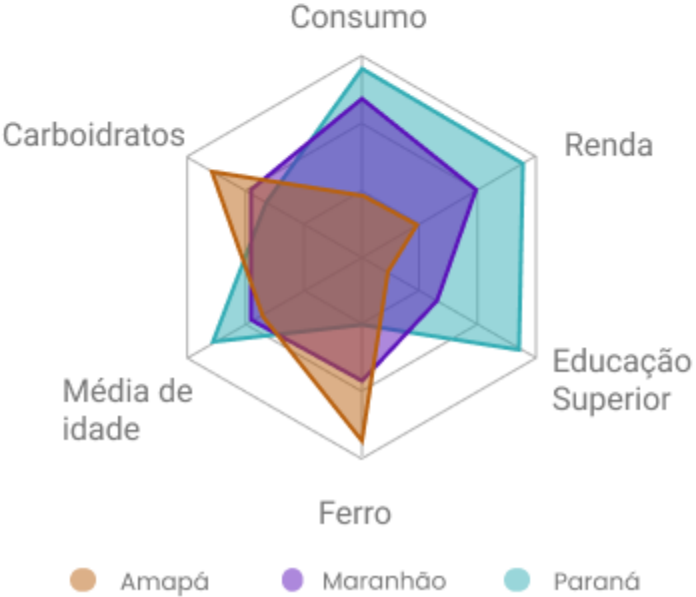
2 Renda total



3 Perfil de consumidor – MG e SP



4 Perfil de consumidor – AP, MA e PR



5 Conclusão

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nulla quam velit, vulputate eu pharetra nec, mattis ac neque. Duis vulputate commodo lectus, ac blandit elit tincidunt id. Sed rhoncus, tortor sed eleifend tristique, tortor mauris molestie elit, et lacinia ipsum quam nec dui. Quisque nec mauris sit amet elit iaculis pretium sit amet quis magna. Aenean velit odio, elementum in tempus ut, vehicula eu diam. Pellentesque rhoncus aliquam mattis. Ut vulputate eros sed felis sodales nec vulputate justo hendrerit. Vivamus varius pretium ligula, a aliquam odio euismod sit amet. Quisque laoreet sem sit amet orci ullamcorper at ultricies metus viverra. Pellentesque arcu mauris, malesuada quis ornare accumsan, blandit sed diam.

SPRINT 4

PRÓXIMOS PASSOS

1

Criação de Infográfico

2

Modelo de Ensemble
Processamento em Big Data

3

Análise de eficácia
Relatório Infográfico

4

Análise financeira e
Plano de comunicação

VISÃO GERAL PROJETO

Arquitetura de Ingestão de Dados.
Entedimento do Negócio;
Entedimento de UX;

Sprint 1

Documentação de análise de impacto ético;
Data Lake/Data Warehouse Alimentado pelo ETL.

Sprint 3

Pipeline de BigData na AWS;
Documentação Final;
Apresentação Final.

Sprint 5

Sprint 2

Prototipação da visualização de dados;
Estrutura de Ingestão de dados com
armazenamento.

Sprint 4

Criação de Infográfico e relatório;
Plano de comunicação.
Modelo de Ensemble;
Análise financeira;

STATUS REPORT

Projeto | Cliente

Integração, Gerenciamento e Análise de Big Data | Integration

Data

24 de Novembro

Indicadores



Concluído



Em andamento



Atrasado



Dependência

Legenda

RW - Review

PR - Programação

UX - User Experience

Status do Projeto

Tema	Processo	Previsão	Status	Plano de ação	Responsável	Indicador
PR	Ingestão dos dados no S3	09/11/2023	Concluído		Inteli	
PR	Atualização automática dos dados	08/12/2023	Atrasado	Continuação na Sprint 4	Inteli	
PR	Códigos postais	20/11/2023	Concluído		Inteli	
PR	Configuração do RedShift	23/11/2023	Concluído		Inteli	
PR	Integração dos buckets com o RedShift	23/11/2023	Concluído		Inteli	
PR	Criação das views	08/12/2023	Em andamento	Continuação na Sprint 4	Inteli	
UX	Prototipação do Infográfico	23/11/2023	Concluído		Inteli	
PR	Ajuste do erro dos pacotes	08/12/2023	Em andamento	Continuação na Sprint 4	Inteli	
PR	Inserção dos dados da API	08/12/2023	Em andamento	Continuação na Sprint 4	Inteli	

Dependências

Objetivo - Sprint 3

Criação do Cubo de Dados



DATA DREAM

OBRIGADO

DATA DREAM - SPRINT 3