

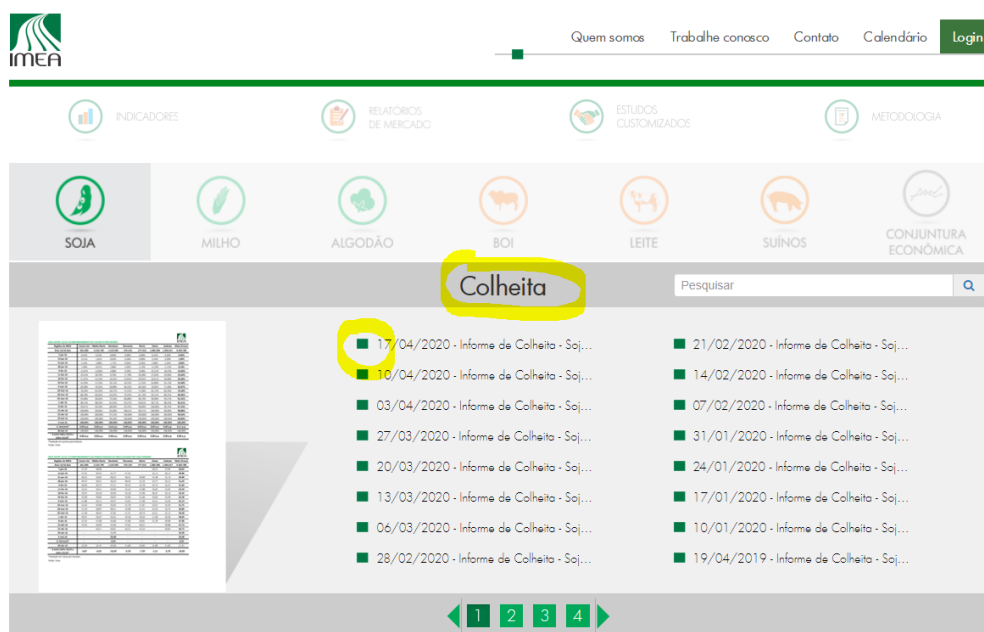
Desafio – Engenheiro de Dados Sênior / Especialista

Contexto:

Precisamos obter a evolução percentual diária da colheita de soja por “Regiões do IMEA” para criarmos alguns gráficos no Tableau. No entanto, os dados estão em uma tabela dentro de um PDF e são separados em relatórios semanais.

Seu desafio consiste em obter os dados das **últimas duas safras** e garantir que os dados da **nova safra** sejam obtidos automaticamente após as novas publicações.

Link do site: <http://www.imea.com.br/imea-site/relatorios-mercado-detalle?c=4&s=8>



17 de abril de 2020

SOJA: ACOMPANHAMENTO DE COLHEITA POR REGIÕES SAFRA 19/20

Regiões do IMEA	Centro-Sul	Médio-Norte	Nordeste	Noroeste	Norte	Oeste	Sudeste	Mato Grosso
3-jan-20	0,00%	0,39%	0,00%	0,09%	0,00%	0,21%	0,35%	0,23%
10-jan-20	0,94%	3,08%	0,07%	1,80%	1,63%	2,01%	1,34%	1,79%
17-jan-20	3,88%	8,04%	1,44%	5,01%	6,23%	6,49%	6,65%	5,86%
24-jan-20	11,23%	20,29%	5,36%	11,27%	16,92%	17,60%	12,39%	14,42%
31-jan-20	19,53%	35,63%	10,81%	24,86%	25,36%	39,15%	22,15%	26,67%
7-fev-20	30,78%	60,87%	21,11%	41,34%	40,25%	60,79%	35,57%	44,51%
14-fev-20	41,23%	71,33%	36,28%	56,23%	59,03%	76,33%	52,27%	58,22%
21-fev-20	58,31%	87,90%	47,50%	75,41%	76,81%	88,75%	66,31%	73,18%
28-fev-20	77,45%	95,26%	64,29%	89,06%	84,44%	93,63%	78,11%	84,10%
6-mar-20	91,40%	99,21%	74,70%	98,29%	89,48%	96,40%	88,42%	91,47%
13-mar-20	97,24%	100,00%	88,06%	99,52%	98,36%	99,35%	96,48%	96,86%
20-mar-20	99,42%	100,00%	91,83%	100,00%	99,77%	100,00%	99,17%	98,38%
27-mar-20	100,00%	100,00%	95,93%	100,00%	100,00%	100,00%	100,00%	99,30%
3-abr-20	100,00%	100,00%	99,02%	100,00%	100,00%	100,00%	100,00%	99,83%
10-abr-20	100,00%	100,00%	99,67%	100,00%	100,00%	100,00%	100,00%	99,94%
17-abr-20	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%
Δ Semanal*	0,00 p.p.	0,00 p.p.	0,33 p.p.	0,00 p.p.	0,00 p.p.	0,00 p.p.	0,00 p.p.	0,06 p.p.
19-abr-19	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%
Δ entre Safra 18/19 e Safra 19/20*	0,00 p.p.	0,00 p.p.	0,00 p.p.	0,00 p.p.	0,00 p.p.	0,00 p.p.	0,00 p.p.	0,00 p.p.

*Variação em pontos percentuais
Fonte: Imea

Resultado esperado:

1. Criação do pipeline de dados com recorrência semanal;
2. Uma tabela no banco de dados para armazenar os registros (respeitando as boas práticas);
3. Disponibilizar acesso ao banco de dados (público) ou “dump” para possível teste no Tableau;
4. Criação de repositório no GitHub com todo o código.

Tecnologias obrigatórias:

1. Linguagem Python;
2. Banco de dados PostgreSQL ou similar;
3. Versionamento de código no GitHub.

Opcionais (não obrigatório):

1. Utilização de serviços na AWS (EC2, ECR, RDS e EKS);
2. Docker;
3. SQLAlchemy (Python);
4. GitHub Actions para CI/CD.

Prazo de entrega esperado: *15 dias

**se precisar de mais tempo é só avisar*

Boa sorte!

Att.

**Time de Business Intelligence
Sementes Jotabasso**