



PROCESSO SELETIVO: DATA RESEARCH & ECONOMICS

SETEMBRO DE 2022

Case do Mercado de Combustíveis

A empresa distribuidora de combustíveis chamada Otto S.A., requisitou à 4intelligence um projeto de inteligência de dados com o objetivo de identificar com maior precisão os mercados consumidores de combustíveis. A empresa precisa de informação mais granular do que a existe atualmente para decidir como alocará os recursos reservados para ampliação da estrutura de distribuição.

A Agência Nacional do Petróleo (ANP) divulga dados de venda de combustíveis por estado e tipo de combustível - etanol, gasolina e diesel - na frequência mensal. Esses dados de 2003 a 2020, foram disponibilizados para você via e-mail na planilha 'venda_distribuidoras_anp'. A unidade de medida é metros cúbicos.

O desafio proposto pelo cliente é desagregar esses dados estaduais até o nível do município. Por exemplo, por meio dos dados da ANP, identificamos que em 2020, foram vendidos 7,0 milhões de metros cúbicos de diesel no estado de Minas Gerais, mas queremos saber quanto de diesel foi vendido no município de Santa Rita do Sapucaí, sede da 4intelligence, no mesmo período. Para atender a demanda da Otto S.A. será necessário entender quais são os principais *drivers* do consumo de combustíveis e como esses fatores determinantes estão distribuídos dentro de cada estado para poder estimar o consumo em cada município.

1. O primeiro passo é compreender os dados já disponibilizados pela ANP. Com a base de dados enviada, responda às seguintes perguntas através de visualizações: Qual é a sazonalidade da venda de combustíveis? Essa sazonalidade é igual em todos os estados? Como o consumo de cada tipo de combustível está distribuído entre os estados brasileiros? Qual é o tamanho total do mercado e como o mercado evoluiu durante os últimos 5 anos?
2. Agora que já temos um entendimento inicial dos dados por UF, precisamos pensar sobre os determinantes locais do consumo de combustível. Quais são as características econômicas de um município que estão por trás de um consumo elevado de combustível e quais são os aspectos municipais que explicam uma necessidade menor de abastecimento?
3. Com base nos fatores identificados no item anterior, quais bases de dados nos ajudariam a quantificar as dinâmicas de cada município? Quais informações você julga importante para estimar o consumo de combustível? Recomendamos conferir: (i) Produto Interno Bruto dos Municípios e as suas aberturas setoriais (IBGE), (ii) população (IBGE), (iii) Frota de Veículos (SENATRAN), (iv) Relação Anual de Informações Sociais (Ministério da Economia), (v) Pesquisa Agrícola Municipal (IBGE), (vi) Pesquisa Industrial Anual (IBGE)
4. Finalmente, de posse de todas as informações levantadas no item anterior, qual seria a sua estratégia para estimar o consumo de combustíveis de um município? Explique detalhadamente as fórmulas/cálculos a serem realizados e os conceitos econômicos por trás da sua metodologia. Não é necessário aplicar a metodologia usando os dados enviados e as bases de dados citadas por você no item 2, apenas des-

crever o processo.

Diretrizes

Para resolver esse case, considere as recomendações abaixo:

- Não são questões de certo ou errado! Queremos entender a sua linha de raciocínio e como você incorpora o arcabouço econômico à resolução do desafio proposto.
- O não cumprimento de algum item não será desclassificatório.
- O teste deve ser feito individualmente, sem auxílio de terceiros.
- **Para responder ao item 1**, o candidato deve nos enviar de volta um código limpo e reproduzível por meio de um link do github na linguagem que quiser. Prefira notebooks: Rmarkdown para R e Jupyter para Python. Devemos ser capazes de reproduzir qualquer resultado que você queira nos mostrar.
- Os arquivos de input enviados por nós não podem conter nenhuma alteração. Todas as modificações têm que ser feitas em código. Assim, não serão aceitos scripts que operem sobre uma base diferente da versão enviada ao candidato.
- **Para responder aos itens 2 a 4**, o candidato deverá nos enviar um pdf com as respostas de até 4 páginas sem contar figuras. O tempo de dedicação à correção das provas é escasso. Por conta disso, esperamos respostas completas, mas ao mesmo tempo sucintas. Você pode (e deve!) utilizar dados a fim de embasar seus argumentos.