

Desafio Técnico Cientista de Dados

Aqui na Geofusion nós prezamos bastante pela cultura lean e nossa equipe técnica participa ativamente no processo de inovação. O desafio técnico se baseia no case fictício descrito a seguir, mas representa algo bastante corriqueiro no nosso dia a dia. Do ponto de vista técnico, observaremos como você organiza seu código e suas ideias, seu conhecimento nas tecnologias utilizadas, manipulação de dados, etc.

Case

Uma **empresa alimentícia** situada no **RJ** deseja abrir filiais na cidade de **São Paulo**, tendo como público alvo adultos de **25 a 50 anos**, das **classes A** (rendas A1 e A2) e **B** (rendas B1 e B2).

Essa empresa já atua no Rio de Janeiro e antes de fazer o investimento **quer estimar quanto poderia faturar em cada um dos bairros** de São Paulo também.

Como o mercado de São Paulo não é bem conhecido para essa empresa, ela deseja **classificar os bairros de São Paulo para saber se o potencial é alto, médio ou baixo**. Utilizando como base o potencial que ela já conhece dos bairros do Rio de Janeiro.

Os analistas dessa empresa também gostariam de **segmentar os bairros de São Paulo** de acordo com o perfil de **renda e idade** para direcionar ações de marketing diferenciadas.

O contato dessa empresa também sinalizou que outras áreas da empresa possuem problemas muito parecidos com esse e que também devem procurar a Geofusion. Internamente também temos recebido feedback da equipe comercial e eles dizem que outros clientes também estão procurando soluções como essa e que poderíamos utilizar as mesmas técnicas para criar um produto que resolva esse tipo de problema.

Desafio Cientista de Dados

O Engenheiro de dados ajudou a extrair os dados de faturamento e potencial dos bairros do Rio de Janeiro provenientes do nosso cliente, juntamente com os dados de sociodemografia dos bairros do Rio de Janeiro e São Paulo.

O CSV anexado possui esses dados para que você possa realizar as suas análises.

Agora, monte uma análise dos bairros de São Paulo para:

- 1 - Estimar o faturamento que uma loja teria em cada um dos bairros
- 2 - Classificar o potencial de cada bairro como Alto, Médio ou Baixo
- 3 - Segmentar os bairros de São Paulo de acordo com a renda e a idade, e indicar aqueles com maior aderência ao público alvo.

Como resultado da análise, é necessário que seja gerado um documento (doc, pdf, ppt, etc) para pessoas não técnicas, apresentando um racional de como os dados foram analisados. Utilize gráficos, tabelas e descreva para apresentar as suas conclusões.

Dado que as técnicas aplicadas podem ser utilizadas para outros clientes com problemas semelhantes, é necessário gerar também o código que te levou a essa análise. Isso se faz necessário pois a operacionalização será feita pelos nossos Engenheiros de Machine Learning juntamente com o time de desenvolvedores. Dessa forma, é necessário também o código que te levou às conclusões apresentadas.

Outra coisa que gostaríamos de ouvir de você é: dado a natureza do problema apresentado, você conhece algum outro dado externo, de fontes públicas ou de fontes privadas, que poderiam ser utilizados para agregar mais valor para o resultado? Se sim, explique o por quê.

Dicionário dos dados:

| | |
|--------------|--------------------------------------|
| codigo | Código do bairro |
| nome | Nome do bairro |
| cidade | Cidade |
| estado | Estado |
| população | População total |
| popAte9 | População - até 9 anos |
| popDe10a14 | População - de 10 a 14 anos |
| popDe15a19 | População - de 15 a 19 anos |
| popDe20a24 | População - de 20 a 24 anos |
| popDe25a34 | População - de 25 a 34 anos |
| popDe35a49 | População - de 35 a 49 anos |
| popDe50a59 | População - de 50 a 59 anos |
| popMaisDe60 | População - 60 anos ou mais |
| domiciliosA1 | Quantidade de Domicílios de Renda A1 |
| domiciliosA2 | Quantidade de Domicílios de Renda A2 |
| domiciliosB1 | Quantidade de Domicílios de Renda B1 |
| domiciliosB2 | Quantidade de Domicílios de Renda B2 |
| domiciliosC1 | Quantidade de Domicílios de Renda C1 |
| domiciliosC2 | Quantidade de Domicílios de Renda C2 |
| domiciliosD | Quantidade de Domicílios de Renda D |
| domiciliosE | Quantidade de Domicílios de Renda E |
| rendaMedia | Renda Média por Domicílio |
| faturamento | Faturamento Total no Bairro |
| potencial | Potencial do Bairro |