

BCG X e Amigos do Bem

Introdução

Durante a participação no evento BCG X, o grupo Barton Consulting Group realizou uma análise de dados para a ONG "Amigos do Bem". O objetivo principal desta análise foi identificar a próxima localidade que deveria receber a atuação beneficente da organização. Além da identificação do município, a análise também se propôs a discernir qual atividade econômica seria mais viável e impactante para ser implementada. A introdução desta atividade econômica tem o intuito de promover oportunidades de emprego para os habitantes locais, fomentando empregos, assim como, desenvolvimento econômico da região. Os lucros gerados a partir desta iniciativa serão revertidos para a ONG "Amigos do Bem", assegurando a continuidade e expansão de suas atividades filantrópicas e impacto social.

Configuração

Para a realização das análises utilizamos a ferramenta VS Code para a interpretação da linguagem python, onde usamos bibliotecas,são elas:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import lightgbm as lgb
import geopandas as gpd
from shapely.geometry import Point
import random
```

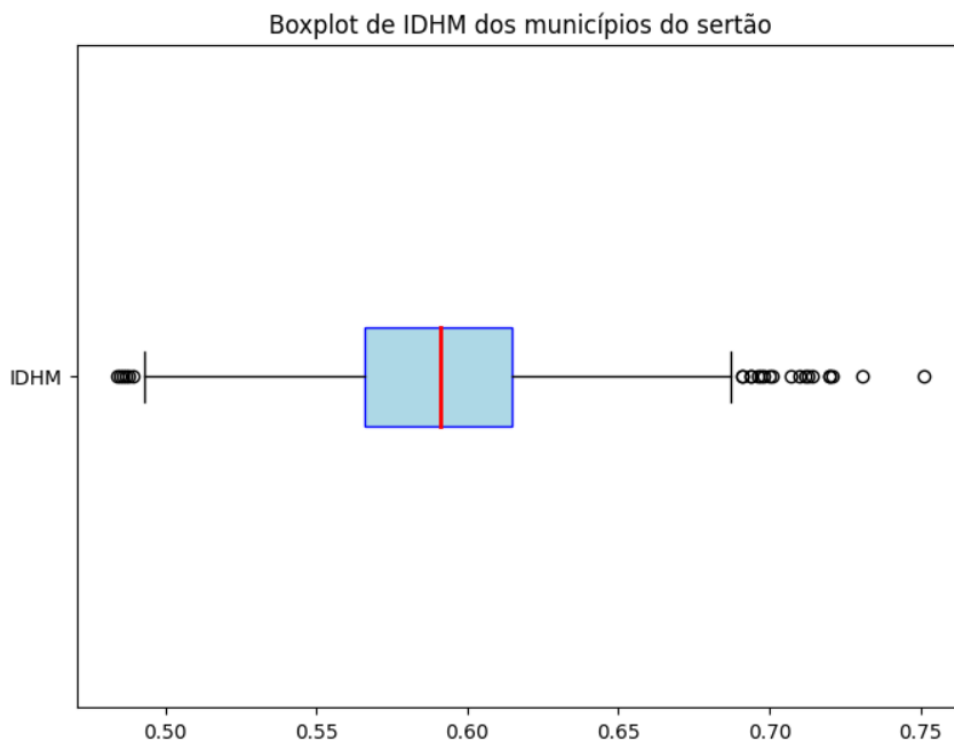
Entendimento dos dados

Inicialmente, recebemos dados do censo, registro administrativo e localidades de todos os municípios do Brasil. Optamos por limitar as bases de dados aos municípios da região semiárida, pois isso nos permitiria focar nas áreas de interesse da ONG "Amigos do Bem".

Após uma primeira filtragem dos dados, ainda enfrentamos problemas relacionados a dados ilegíveis, irrelevantes e nulos. Para extrair informações de maneira mais eficiente, combinamos informações sobre o IDH dos municípios com dados do censo e registros administrativos, junto a informações como latitude, longitude, área e densidade demográfica, pegadas do site oficial do IBGE.

Realizamos análises estatísticas e qualitativas para identificar padrões e tendências nos dados. Definimos parâmetros com base nas análises iniciais, como:

- Como um dos principais tópicos abordados por eles, tanto na apresentação feita na abertura do desafio, assim como em fontes da internet a qual encontramos por meio de pesquisas. O IDH é um item de alta importância para a priorização de municípios, dessa forma fizemos uma análise por meio do gráfico do tipo Boxplot, com ele conseguimos ver a distribuição dos dados como as medidas centrais e medidas que se encontram nos extremos, dessa forma podemos encontrar qual a linha limite que demonstra a partir de qual valor podemos considerar um IDH baixo o suficiente para darmos a prioridade. Dessa forma desenvolvemos o seguinte gráfico:



E por meio deles conseguimos tirar a conclusão que IDHs inferiores a 0,50 são mais propícios de receber uma prioridade maior, visto que são os outliers de menor IDHM do sertão.

- Notou-se que a maioria das cidades atendidas possuíam uma população com mais de 15 mil habitantes antes da chegada dos Amigos do Bem, sendo a menor delas com 18000.. Neste sentido, foi definido que as cidades com população menor que 15 mil habitantes não seriam priorizadas, considerando também o potencial de impacto, haja vista que uma cidade com maior população, possui mais chances de impactar mais famílias.

Com as informações extraídas, adaptamos a nossa base de dados para que ela siga as informações ditas anteriormente, dessa forma, organizamos de forma com que as restrições ditas anteriormente estejam sendo respeitadas. Notamos por fim, que estávamos lidando com a informação de 4 tipos de IDH, sendo eles: **IDHM**, **IDHM-E** (relacionado à educação), **IDHM-L** (relacionado à longevidade) e **IDHM-R** (relacionado à renda).

Identificamos a necessidade de atuação nos estados de **Alagoas e Piauí**, que apresentavam uma maior predominância dentro das restrições citadas.

Aprimoramento dos dados

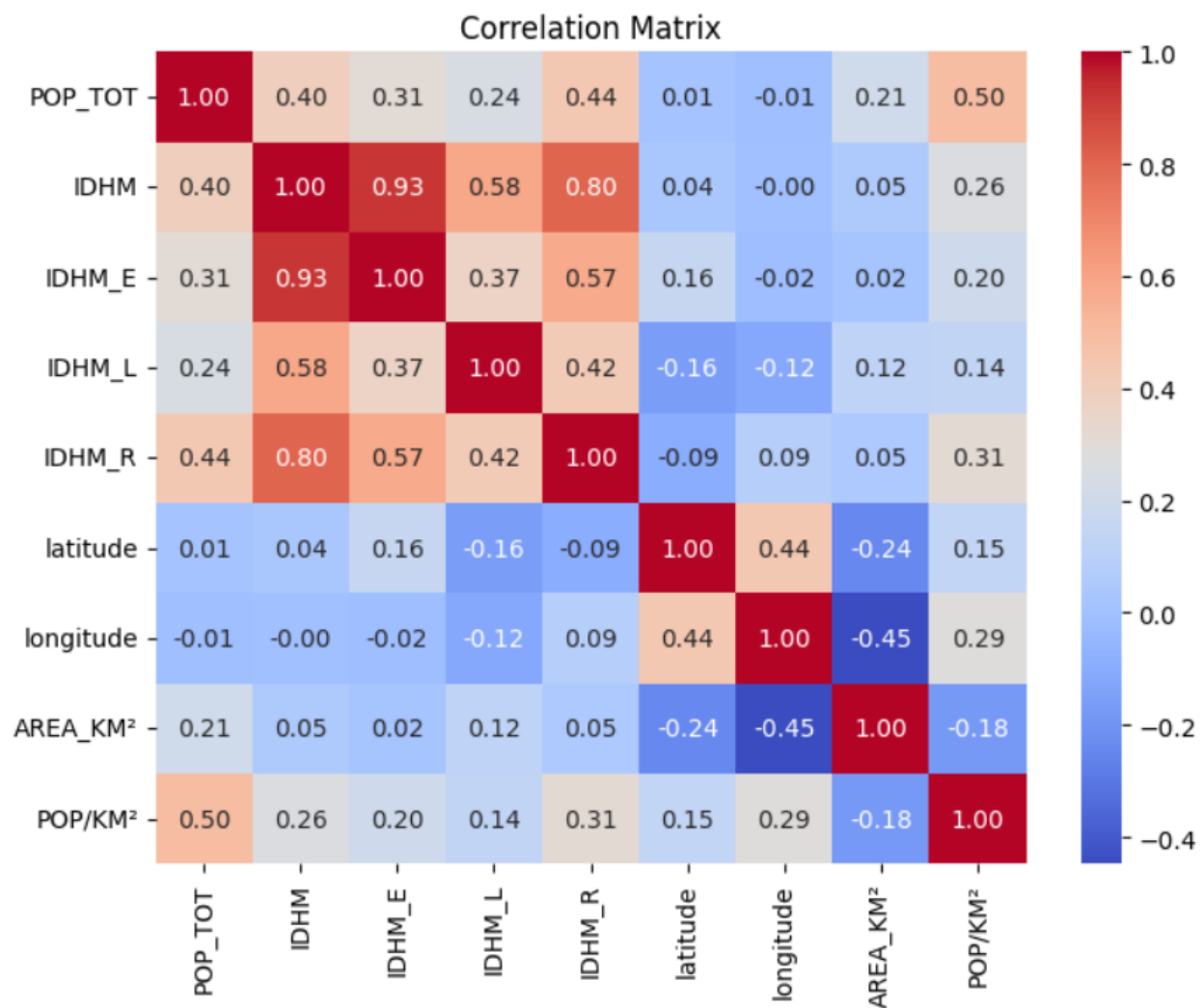
Com a informação de quais features devem ser priorizadas e quais estados estão em uma situação mais alarmante, vamos então dar alguma modificada nos dados coletados visando alcançar novos resultados. Vimos que com a restrição de população sendo maior ou igual à 15 mil habitantes, poderíamos estar excluindo cidades com situações tão alarmantes quanto, porém só por ter menos habitantes elas ficariam de fora, dessa forma resolvemos relaxar a restrição para acima de 6 mil, mas isso não significa que estamos diminuindo a restrição no sentido de quem deve receber a prioridade de auxílio, a prioridade ainda se mantém para municípios com mais de 15 mil habitantes, só fizemos essa alteração para uma visualização melhor dos dados.

Em uma conversa com o Amigos do Bem, descobrimos que uma feature muito relevante que deve ser levada em consideração era a proximidade com as cidades que eles já ajudam, isso se dá pelo fato de que apesar da grande força de vontade de ajudar os municípios, a logística de deslocamento nessas áreas é muito complicada, portanto quanto mais próximo de centros que eles já ajudam favorece ainda mais a escolha de uma cidade. Portanto, pegamos as 10 primeiras cidades com os piores IDHM e calculamos para cada uma delas a distância de cada uma até as 4 cidades que eles já ajudam, **Mauriti, Inajá, Catimbau e Torrões**.

Modelagem dos dados

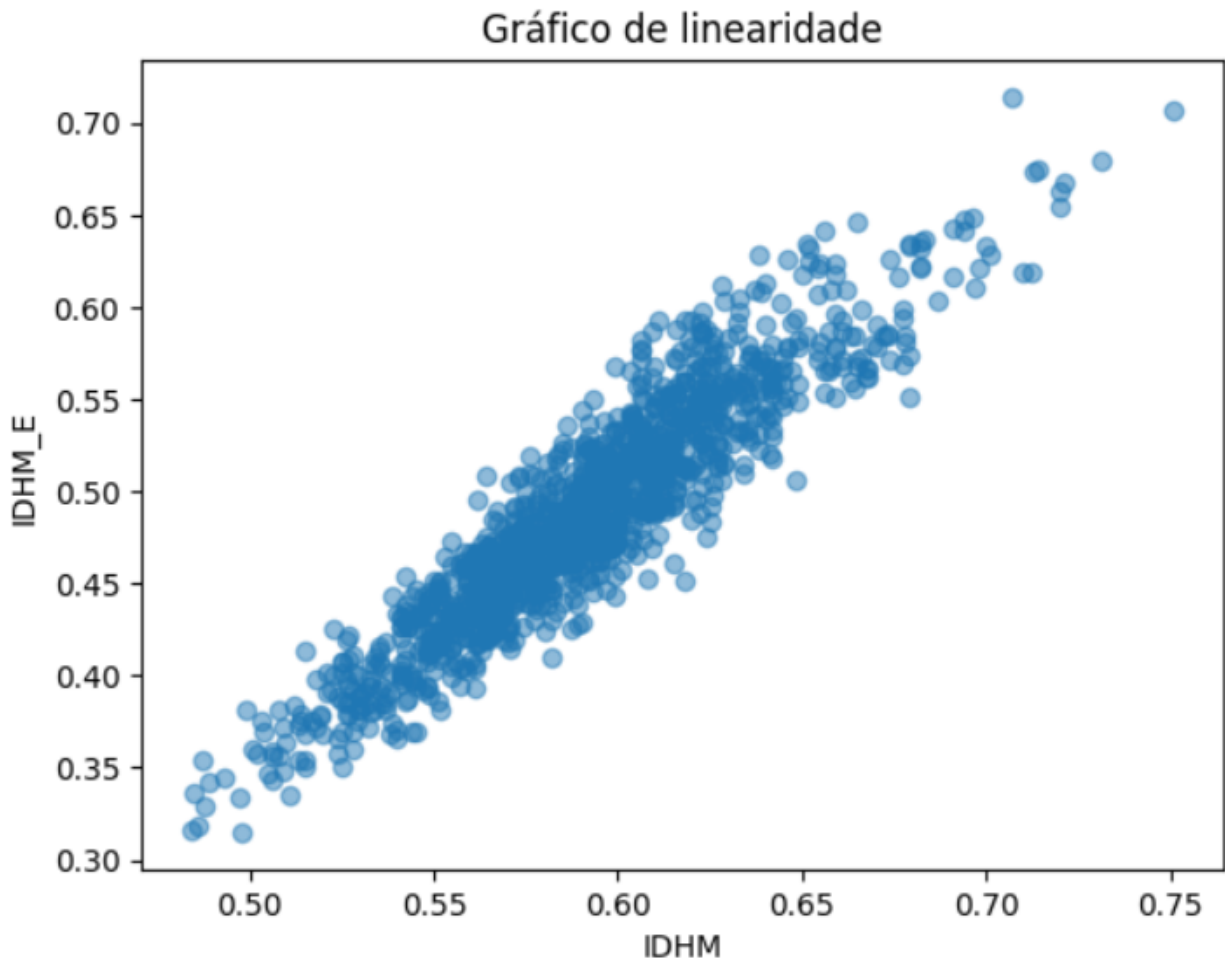
Pensando em uma análise mais sofisticada dos fatores de impacto e dados selecionados, foi decidido com suporte de profissionais do BCG X o uso de machine learning para identificação e comprovação da relevância de alguns índices para o objetivo final: aumentar o IDHM das cidades afetadas, através da melhora da qualidade de vida, educação e renda.

Para analisarmos como podemos atuar no aumento do IDHM optamos por uma Matriz de Correlação com Heatmap, ela é útil, pois ela nos dá uma visualização gráfica e analítica da porcentagem de correlação entre as features de um dataset, dessa forma conseguimos ver qual a feature que mais afeta o IDHM de forma proporcional, para que com isso conseguimos atuar nessa feature com o intuito de conquistar o nosso objetivo, dessa forma, desenvolvemos a seguinte matriz:



Na matriz, quanto maior a porcentagem ou quanto mais próximo de vermelho a cor for, significa uma maior correlação entre os dados. Com isso, podemos notar que o índice **IDHM_E** possui uma correlação de 93% com a coluna de **IDHM**.

Analisando mais profundamente a relação entre essas duas colunas, podemos notar o seguinte comportamento:



O crescimento das duas variáveis, se dá de forma linear e diretamente proporcional, portanto conseguimos influenciar no aumento do **IDHM** através de políticas de melhorias para o **IDHM_E** como por exemplo construção de escolas, qualificação na formação dos professores, melhorias na infraestrutura das escolas já existentes em condições precárias, entre outras.

A análise revela também que os dados de **população e área geográfica** dos municípios têm correlações relativamente baixas com o **IDH**. No entanto, destaca-se que a **densidade populacional** tem uma influência ligeiramente maior no **IDH**, em comparação com outras variáveis relacionadas à **população e área**.

Resultados obtidos

Com essa primeira análise, concluiu-se que as opções de cidades que deveriam ser priorizadas para a expansão da ONG seriam:

- **Inhapi, AL:** Segundo IBGE, Alagoas é o estado com o segundo menor IDH, além da segunda menor Receita realizada em 2017. Contabilizando as médias entre as cidades do semiárido, Alagoas entra como o **primeiro menor IDH**. Está em quarto menor despesas empenhadas. Além disso, tem o **segundo menor rendimento e está com a pior nos índices de educação e renda**, de acordo com a média de cidades do semiárido. E ainda com a **segunda pior média de longevidade**. Vale destacar, que Inhapi é a cidade com o menor IDHM, ressaltando a necessidade de atuação mais próxima da ONG e um alto potencial de impacto, visto a **população de mais de 18.000 habitantes** (valor tal muito próximo a média de população das cidades atendidas pela ONG, na época em que essa começou a desenvolver atividades nelas). Ademais, vale destacar localização, sendo **muito próxima de Inajá**, cidade que já é atendida pela ONG, Inhapi ainda conta com proximidade com outros municípios que precisam de apoio, como **Manari, Mata Grande e Olivença**, que apesar de populações menores, possuem IDHMs igualmente baixos e poderiam ser afetados indiretamente com a atuação em Inhapi.
- **Cocal, PI:** Segundo o IBGE, o Piauí ocupa o **terceiro lugar de IDH mais baixo**, além disso, também tem a **terceira menor receita e segunda menor despesa empenhada**. Ademais, se forem considerados por IDHMs pela média por cidade, têm o **segundo menor IDHM-E e IDHM-R**. Além disso, tem o terceiro pior em longevidade. Soma-se isso a estatística de que das 10 cidades resgatadas com os menores valores de IDHM e com população acima de 6000 habitantes, de 10, 4 são do Piauí. Dessa forma, a cidade de Cocal foi escolhida, considerando sua **população de mais de 27.000 habitantes**, em adição com o baixo IDHM e sua localização próximo a outras duas cidades também em subdesenvolvimento (**Cocal dos Alves e Caxingó**) que poderiam ser afetadas indiretamente. Destaca-se que esse não tem proximidade com nenhuma cidade já atendida pela ONG, contudo a instalação em uma nova localidade seria propício para o crescimento descentralizado e a oportunidade de atuação em novas regiões.