**1. Qual problemas estamos querendo resolver?**

Como selecionar estrategicamente municípios com alto potencial para atração de alunos para os cursos de MBA da Instituição Sirius, com base em dados concretos e relevantes, maximizando o ROI (retorno sobre investimento) e minimizando riscos e gastos desnecessários?

**2. Quais são os padrões nos dados que levam a soluções?**

Estamos analisando diferentes municípios para identificar onde a Sirius poderia encontrar mais alunos para nossos MBAs à distância. Ao olhar para os dados disponíveis, aqui estão os pontos principais que considero:

Densidade e Urbanização: Queremos áreas com muita gente e boa urbanização. Isso sugere que as pessoas têm acesso à tecnologia e podem estar mais inclinadas a fazer um curso online.

Educação Atual: Se um município tem muitas pessoas com formação média ou alta, ou muitos estudantes universitários, é um bom sinal. Isso mostra que eles valorizam a educação e podem estar interessados em um MBA.

Economia Local: Municípios com um bom PIB e salários elevados são locais onde as pessoas podem ter dinheiro para investir em sua educação. Além disso, áreas com muitos trabalhadores no setor de serviços (terciário) podem ter mais pessoas interessadas em um MBA.

Infraestrutura: Se uma cidade tem bons serviços de saúde e saneamento, é um sinal de que é bem desenvolvida. Isso pode se traduzir em mais pessoas com acesso e interesse em cursos online.

Cultura: Cidades com muitas atividades culturais e de lazer geralmente têm pessoas abertas a novas ideias e aprendizados.

Empregos: Em lugares com baixo desemprego, as pessoas são mais estáveis e podem pensar em melhorar suas habilidades com um MBA.

Então, para identificar as melhores cidades, vamos olhar para estas áreas-chave nos dados e ver quais municípios se destacam!

**3. Qual modelo "parece" resolver o problema?**

Ao analisar nosso objetivo, que é entender e prever os municípios com maior probabilidade de ter potenciais alunos para o programa de MBA à distância da Sirius, percebemos que alguns modelos se destacam como opções promissoras:

Modelos de Classificação: Esses modelos poderiam nos ajudar a categorizar cidades como "prováveis" ou "improváveis" para prospecção com base nas características de cada município e nos dados dos alunos atuais. Por exemplo, poderíamos usar Árvores de Decisão.

Sistemas de Recomendação: Baseando-nos nas cidades que já são fontes de muitos de nossos alunos, poderíamos usar este modelo para identificar outras cidades com perfis semelhantes, aumentando assim nossas chances de encontrar potenciais alunos. Esse enfoque se assemelha à lógica de "clientes que compraram X também se interessaram por Y".

Em minha opinião, começaria com os Modelos de Classificação e depois migrar para os Sistemas de Recomendação. A classificação pode dar uma visão clara das cidades em que devemos ou não focar, enquanto o sistema de recomendação poderia nos dar insights sobre novos municípios que ainda não consideramos, mas que têm um perfil semelhante aos nossos atuais pontos de sucesso.

**4. Por que chegamos nesta definição de modelo? Argumente e mostre os prós e contras do modelo.**

Modelos de Classificação:

Prós:

Simplicidade e interpretabilidade: Modelos como Árvores de Decisão são facilmente interpretáveis, permitindo-nos entender quais características das cidades são mais relevantes para a prospecção.

Eficiência: Tendem a ser rápidos de treinar e podem lidar bem com um conjunto diversificado de características.

Flexibilidade: Podemos ajustar e adaptar o modelo com base no feedback e nos resultados das campanhas de prospecção, tornando nosso enfoque cada vez mais preciso.

Contras:

Generalização excessiva: Modelos muito simplificados, como a Árvore de Decisão, podem não capturar toda a complexidade do dataset, podendo ser muito generalistas ou muito específicos.

Possibilidade de Overfitting: Em particular, modelos como Florestas Aleatórias, se não forem corretamente otimizados, podem se ajustar demais aos dados de treinamento e ter um desempenho inferior nos dados de teste.

Sistemas de Recomendação:

Prós:

Personalização: Ao analisar as características das cidades de onde já temos alunos, podemos identificar padrões que nos permitem sugerir outras cidades com perfis semelhantes.

Escalabilidade: Uma vez que o sistema esteja operacional, pode ser facilmente escalado para considerar mais características ou dados adicionais.

Abordagem Proativa: Em vez de apenas reagir aos dados que temos, estamos ativamente procurando novas oportunidades, expandindo nosso alcance.

Contras:

Complexidade: Dependendo da abordagem escolhida, sistemas de recomendação podem ser mais difíceis de interpretar e ajustar do que os modelos tradicionais.

Possibilidade de Bias: Se baseado muito fortemente nos dados existentes, podemos nos tornar míopes, focando apenas em cidades muito semelhantes às que já conhecemos e potencialmente perdendo outras oportunidades.

Conclusão:

Optamos por começar com a classificação por sua capacidade de fornecer uma visão clara e acionável das cidades em que devemos focar. Isso nos dá uma base sólida. Posteriormente, o sistema de recomendação entra como uma ferramenta estratégica, permitindo-nos expandir nossa busca e identificar oportunidades que poderíamos ter perdido. Juntos, esses modelos oferecem uma combinação robusta de eficácia e proatividade, alinhada com nossa missão de expandir a base de alunos do MBA à distância da Sirius.

**5. Além desse, há algum outro modelo que também resolveria o problema? Qual? Por quê?**

1. Modelos de Regressão:

Descrição: Esses modelos visam estimar uma variável contínua a partir de uma ou mais variáveis independentes. Por exemplo, podemos tentar prever o número potencial de alunos que um município pode trazer para o MBA, com base em suas características.

Por que considerar: Se tivéssemos dados sobre o número exato de alunos oriundos de cada município, um modelo de regressão poderia prever quantos alunos outros municípios com características semelhantes poderiam trazer.

2. Modelos de Clusterização (como K-Means ou DBSCAN):

Descrição: Esses modelos segmentam o dataset em grupos (clusters) de registros semelhantes.

Por que considerar: Eles poderiam ser usados para identificar grupos de cidades com perfis semelhantes, ajudando-nos a direcionar estratégias de marketing ou prospecção para cada segmento.

3. Redes Neurais:

Descrição: São modelos que se inspiram no funcionamento do cérebro humano e são particularmente bons para detectar padrões complexos.

Por que considerar: Para um conjunto de dados grande e complexo, as redes neurais podem capturar interações não-lineares entre as variáveis que outros modelos podem não detectar. Porém, requerem mais dados e são menos interpretáveis.