**Step 2 Projeto – Sirius Cidades**

1. **Objetivo de Negócio**

Criar algoritmo com o objetivo de sugerir municípios onde a Sirius pode prospectar mais alunos para os cursos de MBA da Instituição, com base em fatores estruturais destes municípios e a produtividade dos trabalhadores.

Espera-se acurácia de pelo menos 70%.

1. **Obter Dados**

Dataset Cidades com diversas variáveis explicativas das cidades brasileiras que explicam suas respectivas produtividades.

Para complementar o estudo será necessário que a Sirius forneça um dataset dos alunos, matriculados ou não, contendo as informações abaixo por aluno:

* Cidades de Nascimento
* Cidade Atual
* Instituição de Ensino
* Nível de Escolaridade
* Status – Matriculado ou não
* Forma de Entrada
* Status – Empregado ou não
* Ramo de Emprego
* Renda

1. **Separar Dados**

Por se tratar de uma base de dados extensa, sugere-se a separação de 80% da base de dados para teste, desde que se garanta, de forma proporcional, amostras de todos os clusters obtidos na Step 1 do Projeto (*Clusterização por produtividade das cidades brasileiras*).

Sugere-se, 20% da base de dados para teste, seguindo as mesmas observações citadas na divisão da base de treino.

*Obs.: No aguardo da base de dados dos alunos para, dependendo do tamanho do dataset, tomar a decisão da divisão e estratégia a ser seguida.*

1. **Explorar Dados**

Será inserido a base de dados de teste no BI obtido como saída da Step 1 do Projeto para obtenção de mais insights.