Fórum de Debate do Módulo 2

No Brasil, a cultura orientada à dados ainda é incipiente, pouco divulgada, e poucas são as empresas que realmente investem na implantação da cultura orientada à dados. Pensar em uma organização orientada à dados é ter em mente que a tomada de decisões desta organização deve ser fundamentada por dados e pelas análises de dados.

Tendo em vista a grande disponibilidade de dados que existe hoje, um dos desafios das organizações é saber identificar quais dados, sejam eles internos ou externos, são relevantes e necessários para compor uma solução de análise, de forma a garantir que suas decisões sejam mais assertivas.

Dada a variedade de tipos de dados e de fontes de dados disponíveis, é imprescindível que a organização selecione suas fontes e determine os dados necessários. Conhecemos alguns portais de dados abertos governamentais, como por exemplo:

* Portal Brasileiro de Dados Abertos: <http://www.dados.gov.br/>.
* Portal da Transparência: <http://www.portaldatransparencia.gov.br/>.
* Portal de Dados Abertos Dataprev: <http://dadosabertos.dataprev.gov.br/>.
* Portal de Dados e APIs do IBGE: [https://servicodados.ibge.gov.br/api/docs](https://servicodados.ibge.gov.br/api/docs )
* Entre outros, inclusive internacionais.

Pesquise dois conjuntos de dados que possam ser combinados e analisados. Comente sobre as necessidades para coletar esses dados e como você imagina que deva ser uma solução, indicando ainda que relação ou ligação você identifica entre esses conjuntos de dados. Lembre-se de descrever a fonte em que eles se localizam e o formato em que se encontram.

Boa noite, pessoal!

Após a leitura de algumas respostas de meus colegas, percebi que muitos estão destacando a questão da saúde. Por isso, resolvi abordar outro tema não menos importante: a educação.

Em primeiro lugar, acessando o conjunto de dados de “**Microdados do Exame Nacional do Ensino Médio – Enem**” (<http://dados.gov.br/dataset/microdados-do-exame-nacional-do-ensino-medio-enem>) no **Portal Brasileiro de Dados Abertos**, percebi que há uma desatualização das informações disponíveis se comparadas ao que consta no site oficial do **INEP** (<http://inep.gov.br/web/guest/microdados>), pois este último disponibiliza informações das provas até o ano de 2019, enquanto que a primeira fonte disponibiliza apenas até o ano de 2014. Acredito que muito provavelmente não há um sincronismo e atualização efetiva para deixar ambos os portais pareados. Seria interessante que houvesse essa sintonia entre as partes, a fim de manter as informações sempre atualizadas, na medida do possível.

Em segundo lugar, aproveitei para baixar o conjunto de dados em formato CSV das provas de ENEM aplicadas em 2019 (MICRODADOS\_ENEM\_2019.csv) e analisar os campos disponíveis. Percebi que existem diversas informações sobre os candidatos, como número de inscrição, estado civil, sexo, local de nascimento, notas nas provas em cada disciplina e inclusive informações referentes à localização de residência dos candidatos e de suas escolas, incluindo município e UF. Além disso, aproveitei para baixar os dados referentes ao **Orçamento da Despesa** no **Portal da Transparência**, referentes ao ano de 2019 (<http://www.portaltransparencia.gov.br/download-de-dados/orcamento-despesa>). Trata-se de um conjunto de dados também em formato CSV (2019\_OrcamentoDespesa.zip.csv), com diversas informações de valores gastos por diversos setores e em diversos programas orçamentários, seja comércio exterior, aviação civil, agropecuária, além das próprias iniciativas relacionadas à educação. Então, seria possível relacionar os dados de provas do ENEM com os dados de despesas em prol da educação brasileira e avaliar possíveis deficiências existentes nos alunos em sua absorção e entendimento das disciplinas, com base nas notas médias dos alunos por disciplina e por região específica, seja residencial ou escolar, e investir mais em carga horária dos professores nestas matérias no ensino médio e, consequentemente, injetando mais dinheiro nesse sentido. Outra aplicação interessante para esta conclusão seria investir mais em aulas EAD transmitidas por canais abertos e por região, servindo de reforço aos alunos que tenham dificuldade em certas disciplinas, mais do que em outras. Desta forma, seria possível obter um conhecimento maior e mais profundo do rendimento escolar periódico, analisar possíveis mudanças de cenário e apurar melhor o planejamento financeiro de cada estado e/ou município para fins de educação presencial e à distância.

Por fim, em termos de soluções para a coleta de dados, acredito que estes dados deveriam ser coletados e relacionados anualmente, agregados com notas médias em cada disciplina por município/UF de escola e residência, juntamente com as despesas em educação pública presencial e à distância. Após estes serem unidos e relacionados, seria possível serem disponibilizados via API para serem consumidos por outras aplicações de iniciativas educacionais.