

Teste de Análise Exploratória

O desafio será fazer uma análise exploratória da base de lojista da Justa de acordo com sua movimentação financeira, para isso temos uma coluna com a primeira categoria dividida em seis grupos:

1. **Novo** (primeira transação = mês de referência), ou seja, é um volume novo em nossa base de dados;
2. **Recorrente** (lojista que transacionou mês passado e transacionou no mês de referência);
3. **Recuperação** (lojista que não transacionou mês passado e transacionou no mês de referência);
4. **Perda** (lojista que transacionou mês passado e não transacionou no mês de referência);
5. **Churn** (lojista que não transaciona a mais de dois meses, baseado no mês de referência);
6. **Nunca Mov** (lojista que está em nossa base e nunca teve uma transação).

Dado essa classificação, aplicar modelos inteligentes para uma análise exploratória da base por critérios que não estão a olho nu, por exemplo: dentre os recorrentes quais são os melhores? Quais outputs o modelo KNN trará com essa base? Quais outros modelos podem ser útil para esta análise?

Exemplo:

Queremos dar um “**zoom**” nessas categorias e **sub classificar todos os heróis**, como por exemplo: o lojista recorrente, gostaríamos de saber **qual deles que traz o maior volume transacional e tem uma constância mensal em sua movimentação financeira** (transaciona todos os meses), podendo sub classificá-los em: recorrente 1 (lojista com maior volume e constância mensal); recorrente 2 (lojista com volume mediano mas com constância mensal); recorrente 3 (lojista com volume mediano e constância mensal mediana); recorrente 4 (lojista com volume baixo e constância mensal baixa) etc... Claro que isso é um exemplo e temos que levar outras variáveis em questão.

Desafio - sub classificar as categorias da base transacional da Justa:

- Utilizar linguagem de programação para clusterização (Python ou R);
- Enviar código formatado e comentado com o passo a passo;
- Quais colunas dessa base não agregaram valor ao estudo;
- Documentar toda análise realizada, até as que não entregaram resultados satisfatórios;
- Enviar base versão final.