

Universidade Federal de Minas Gerais  
Departamento de Ciência da Computação

DCC057 Mineração de Dados  
**Trabalho Prático 2 — Proposta**

Alexander Thomas Mol Holmquist  
10 de fevereiro de 2021

# 1 Introdução

É amplamente conhecido a dificuldade que é, no processo de procura de um novo lar, de tomar uma decisão definitiva. Tantas vezes, não se sabe como analisar a situação. Este projeto visa auxiliar o corretor no processo de acompanhamento do cliente para tomada de decisão.

Em qualquer área do conhecimento, é muito útil prover uma categorização do que se tem pela frente. No nosso caso, é claramente benéfico para o cliente saber quais tipos de imóveis são disponibilizados pela corretora, e isso de acordo com um critério específico do cliente, que deve ser perseguido heurísticamente pelo corretor em questão. Precisamente com este objetivo, procuramos integrar a mineração de dados ao ambiente de correção de imóveis, através de técnicas simples de agrupamento.

## 2 Exemplos

Pode ser que se encontre um agrupamento trivial, que indique, de forma elegante, que o preço é um atributo sumarizante para o conjunto de dados, como: "casa muito barata, casa barata, casa de preço médio, casa de preço alto, casa de preço muito alto".

É possível, ainda, que o modelo forneça um resultado mais interessante, como: "casa muito barata e com dois banheiros, casa barata e nova, casa barata e velha, casa nova e cara, casa nova e velha".

Em qualquer caso, essa categorização será realizada de forma heurística, com base em características destacantes dos grupos porventura encontrados.

## 3 Recursos disponíveis

O plano inicial é utilizar a plataforma *Lemonade* para a realização das tarefas. Caso não seja possível, será construído um *Jupyter Notebook* escrito em Python 3, com o passo-a-passo da mineração de dados.

Quanto aos dados, o foco é em uma única base de dados de imóveis de Boston, amplamente utilizada para fins didáticos. Está disponível, mas não se restringe à fonte[1]. Geralmente ela é aproveitada para demonstrar conceitos relacionados a regressão. Porém, como foi dito, serão aplicadas técnicas de agrupamento.

## 4 Critérios de sucesso

O critério de sucesso do negócio é aumentar a facilidade da busca dos clientes por um novo imóvel. Pode ser medido de acordo com o tempo que levou para ele tomar a decisão.

O objetivo principal de mineração de dados é agrupar os dados em cinco categorias. Um valor de fronteira para razão de corte, como 5%, será utilizado, mais ainda não está determinado. Segundo, um objetivo secundário é a apresentação de uma metodologia para tomada de decisão que pode ser generalizada para outras fontes de dados irrelacionadas. Esta hipótese está sob análise e pode ser que seja descartada. Neste caso, a aplicabilidade da metodologia gerada pode ser utilizada como medida de sucesso.

## 5 Plano inicial

Pretende-se seguir à risca a metodologia CRISP-DM. Assim, os passos antecipados são:

1. Inserir os dados na plataforma *Lemonade*
2. Familiarizar-se com os dados, e anotar conclusões e hipóteses, assim como pontos fracos
3. Realizar o tratamento dos dados, se necessário
4. Selecionar o modelo de agrupamento apropriado
5. Construir um ambiente de testes para o modelo, e determinar estatísticas que possam ser usadas para analisar a sua qualidade
6. Verificar se os resultados obtidos condizem com a proposta. Se não, decidir o que fazer (pode ser que desistir com justificativa seja a opção correta, por conta do contexto didático).

Observação: durante todo o processo acima, um relatório deve ser gradativamente construído. No fim, ele deve ser reorganizado e será considerado como a "aplicação" (fase 6 do CRISP-DM) do presente projeto de mineração de dados.

## 6 Bibliografia

[1] [https://raw.githubusercontent.com/rashida048/Datasets/master/home\\_data.csv](https://raw.githubusercontent.com/rashida048/Datasets/master/home_data.csv)