

Análise do ITBI para Descoberta de Dinâmicas Imobiliárias Intra Urbanas: Um Estado de Caso em Belo Horizonte

Samuel Henrique Miranda Alves¹

¹Departamento de Ciência da Computação (DCC)

Instituto de Ciências Exatas (ICEX)

Universidade Federal de Minas Gerais (UFMG)

Julho 2025

samuelhma91@gmail.com

Resumo. *Identificar tipos específicos de imóveis que apresentem padrões excepcionais pode revelar características interessantes a respeito da dinâmica imobiliária de uma cidade. Este estudo se propõe a investigar grupos especiais de imóveis dentro da cidade de Belo Horizonte que tenham apresentado uma valorização inusitada nos últimos 16 anos, através da técnica de descoberta de subgrupos aplicada sobre a base de dados do ITBI, publicamente disponibilizada pela prefeitura. Os resultados fornecem informações valiosas sobre a presença de tipos específicos de imóveis que apresentam clara distinção valorativa comparadas com a média global da cidade, tanto em termos de valor do m² como também na distribuição dos padrões de acabamento. Por fim, o trabalho também oferece a possibilidade adaptativa da abordagem metodológica, empregada nos experimentos para identificação dessas dinâmicas, em qualquer outra cidade que também contenha dados públicos do ITBI.*

1. Introdução

As cidades sofrem diversas transformações imobiliárias com o passar dos anos. Em específico, grandes centros urbanos, como metrópoles e capitais estaduais, são alvos constantes de empreendimentos imobiliários que podem modificar a economia urbana local. Um exemplo claro dessas transformações é o processo conhecido como *gentrificação* [Smith 2007], em que a renovação urbana, com a chegada de moradores de maior poder aquisitivo, promove a expulsão da população original, devido, principalmente, ao aumento dos custos de moradia. Esse processo pode ser evidenciado, por exemplo, com o aparecimento de imóveis de maior padrão luxuoso em áreas onde existia uma concentração menor desses tipos de construções. Identificar essas modificações torna-se um elemento imperativo para o correto entendimento da economia urbana [Nabuco 2021].

Este trabalho tem o objetivo principal de encontrar grupos excepcionais de imóveis dentro da cidade de Belo Horizonte que tenham se destacado por uma valorização acima da média global da cidade. Essa busca será feita analisando-se os dados da série histórica do ITBI (Imposto sobre a Transmissão de Bens Imóveis) dos últimos 16 anos (2009-2024), disponibilizada publicamente pela prefeitura de Belo Horizonte, por se tratar de uma base de dados que contém informações sobre o preço de imóveis da cidade. Com base nessa informação, calcula-se o valor do m² (preço / área), pois este é um indicador que permite a comparação do valor imobiliário [Soares et al. 2025]. Além disso, a prefeitura de Belo Horizonte também disponibiliza na sua base de dados do ITBI o atributo

denominado *padrão de acabamento*. Este atributo pode assumir 5 valores categóricos (P1, P2, P3, P4, P5) e é calculado com base em uma soma dos tipos de elementos construtivos utilizados para edificar o imóvel. Dessa forma, a classe P1 sugere baixo padrão construtivo, enquanto a classe P5 constitui o valor mais alto para o padrão luxuoso de um imóvel, indicando uma diferenciação intrínseca de valor entre eles [Rangel and Gomes 2007].

Portanto, este projeto busca identificar os tipos de imóveis que produzem variações anormais, primeiramente, no valor do m² e, depois, na concentração de padrões de acabamento. Os imóveis são analisados com base no bairro de localização, tipo construtivo (casa ou apartamento), idade e ano de avaliação (ano de quitação da transação do ITBI). Para detecção dos grupos de imóveis, utilizou-se a estratégia de Descoberta de Subgrupos (*Subgroup Discovery*), através do algoritmo *Beam Search* [Lavrač et al. 2004]. Os resultados encontrados indicam a presença de tipos específicos de imóveis na cidade de Belo Horizonte que apresentam clara distinção valorativa comparadas com a média global da cidade. Todo o código desenvolvido para os experimentos está disponível online para consulta e reprodução pública¹.

O restante do artigo está organizado da seguinte forma: a Seção 2 detalha a respeito da coleta e tratamento dos dados utilizados; a Seção 3 discute os métodos empregados para análise dos dados; a Seção 4 apresenta os resultados encontrados pelo algoritmo, bem como uma interpretação sobre cada um deles; por fim, a Seção 5 encerra o trabalho apontando as principais descobertas, assim como uma discussão sobre limitações da análise e propostas de trabalhos futuros.

2. Base de dados e Tratamento

Este estudo utilizou os dados públicos do ITBI da cidade de Belo Horizonte, disponibilizadas pela própria prefeitura em seu portal de dados aberto². A base do ITBI refere-se às transações onerosas de imóveis constituídas principalmente por vendas. Ela abrange 14 tipos diferentes de atributos, sendo os quais foram utilizados, neste trabalho, metade deles: Bairro, Ano de Construção, Área, Tipo Construtivo Preponderante, Data de Quitação da Transação, Valor Base Cálculo e Padrão de Acabamento. A partir destes, foram construídas mais 3 variáveis, listadas a seguir:

- **Idade** = Ano de Quitação da Transação - Ano de Construção
- **Valor m²** = Valor Base Cálculo / Área
- **Valor m² corrigido** = Aplicação do IGP-M nos valores dos imóveis

Para este trabalho, assumiu-se que o preço do imóvel constituía-se do Valor Base de Cálculo presente nos dados. Esta variável é o maior valor do imóvel resultante entre a comparação da estimativa do valor do m² feita pela própria Prefeitura de Belo Horizonte, tendo como base a Planta Genérica de Valores do Município, e o valor declarado pelo próprio contribuinte. O cálculo do imposto do ITBI é incidido em cima deste valor [Bernardes 2018].

A variável “Valor m² corrigido” possui um valor significativo para este projeto. Como será feita a análise comparativa de valores monetários (preço dos imóveis) ao curso de vários anos, faz-se necessário trazer todos estes preços para valor presente, a fim de

¹<https://github.com/SamuelHenr/analise-itbi>

²https://dados.pbh.gov.br/pt_BR/dataset/itbi-relatorios

tornar a investigação mais justa. O índice inflacionário aplicado neste contexto é o IGP-M (Índice Geral de Preços - Mercado) [Nadal 2023], calculado pela Fundação Getúlio Vargas (FGV) [IGP-M 2025]. Os valores foram corrigidos para a data de referência de Maio de 2025 e aplicados em um período mensal. Apesar do algoritmo receber o valor do m² original (neste caso denominado valor nominal) como variável alvo, o valor do m² corrigido pela inflação (denominado valor real) será utilizado posteriormente como uma métrica para identificar o impacto real dos grupos retornados pelo algoritmo. Mais detalhes serão informados nas próximas seções.

Além dessas inclusões, foram aplicados determinados filtros nos dados, na tentativa de eliminar alguns ruídos. Os tratamentos aplicados estão detalhados abaixo:

- Ano de Quitação da Transação > 2008
- Área >= 50 (m²)
- Valor Base Cálculo >= 100.000 (R\$)
- Tipo de Ocupação = Residencial
- Idade >= 0; Idade <= 100

O primeiro filtro (ano de quitação da transação do imóvel) faz-se necessário pois, como este foi o primeiro ano em que os dados começaram a ser publicamente disponibilizados, entende-se que existem ruídos nestas informações, no que diz respeito à presença de transações antigas não tratadas. O segundo e terceiro filtros (área e valor base cálculo) são aplicados visando excluir *outliers* presentes nos dados. O quarto filtro (tipo de ocupação) permite a obtenção de imóveis que são predominantemente utilizados como residências (ao contrário de comércios, por exemplo). Por fim, os limites inferiores e superiores aplicados à idade no último filtro têm o objetivo de eliminar erros de digitação presentes nas variáveis utilizadas para a construção deste atributo.

3. Metodologia

Os experimentos realizados neste trabalho foram divididos em 2 partes: análise comparativa do valor do m² e análise comparativa da distribuição de padrões de acabamento. Além disso, para cada análise, realizou-se 2 tipos de testes: um exame levando em consideração o escopo global da cidade e, a seguir, uma inspeção focada especificamente em cada um dos bairros, tratando-os como um grupo global de referência. Essa decisão foi tomada após analisar-se os resultados retornados pelo algoritmo considerando os dados globais da cidade, que imprimia padrões facilmente identificáveis dentro do contexto urbano do município (por exemplo, as regiões que mais se valorizaram são aquelas localizadas nos bairros mais “nobres”). Isso, por um lado, prova a consistência do algoritmo, por fornecer resultados em acordo com o esperado, porém ofusca o principal propósito deste trabalho de identificar regiões não-características da cidade que revelam tipos de imóveis com uma valorização anormal. Tendo isso em vista, decidiu-se, para o segundo tipo de análise, mudar o escopo dos experimentos, filtrando os dados pelos bairros e aplicando a busca do algoritmo separadamente em cada um deles. Isso permite revelar variações específicas dessas regiões que eram encobertas pela análise global. O entendimento deste tipo de abordagem se tornará mais evidente na apresentação dos resultados da próxima seção.

A criação dos experimentos foram conduzidos utilizando-se os recursos disponíveis da ferramenta *pysubgroup*³ [Lemmerich and Becker 2018]. No que diz respeito ao pri-

³<https://github.com/flemmerich/pysubgroup>

meio tipo de análise (focada no valor do m^2 como variável alvo), por se tratar de um atributo numérico, utilizou-se a métrica *StandardQFNumeric* da biblioteca. Trata-se de uma função de qualidade com uma abordagem mais estatística e que leva em consideração, no seu resultado, o tamanho do subgrupo encontrado, um parâmetro que pode ser ponderado pelo argumento α da função. Ademais, incorporamos nessa análise uma métrica adicional calculada através do valor do m^2 corrigido pela inflação (denominado de valor real), que será usada como apoio interpretativo dos resultados. Note que a variável alvo analisada pelo algoritmo continua sendo o valor original do m^2 (denominado de valor nominal), pois as variações temporais do preço dos imóveis constitui-se uma característica importante a ser capturada pelo algoritmo.

Tratando-se do segundo tipo de análise (focada na distribuição dos padrões de acabamento dos imóveis), empregou-se uma estratégia comparativa diferenciada. Primeiramente, por ser uma variável de natureza categórica e ordinal (*i.e.* existe uma ordem intrínseca e clara associada às categorias: $P1 < P2 < P3 < P4 < P5$), torna-se possível extrair um índice médio dos padrões de acabamento de uma região, calculado por meio da média ponderada de cada tipo de padrão. Este índice procura capturar a tendência média de sofisticação de um bairro ou segmento.

Todavia, existe uma falha estatística da abordagem pela média de uma distribuição. Por ser uma medida de tendência central, ela tende a esconder a polarização de um subgrupo. Sendo assim, a análise baseada no índice de padrão médio não conseguiria sozinha, por exemplo, diferenciar um subgrupo de imóveis predominantemente P3 de um subgrupo polarizado entre P1 e P5 que, por acaso, resulte na mesma média. Para contornar isso, ao invés de construir uma métrica que se propõe a capturar todas as variâncias juntas, mudou-se a abordagem estratégica deste experimento para uma *análise triangular*.

A nova estratégia versa rodar três análises separadas e, depois, cruzar os resultados. A primeira análise se trata da medida de tendência central (índice médio dos padrões), que revela bairros consistentemente sofisticados. A segunda análise foca apenas na concentração de imóveis do padrão P5, medindo, assim, a concentração de luxo de uma região, revelando quais são os polos de elite existentes nos dados. Por fim, a terceira análise tem a finalidade inversa de focar apenas na concentração de imóveis de padrão P1, revelando regiões com maior acúmulo de construções mais populares ou mais depreciados. O objetivo consiste em rodar análises mais simples com diferentes focos (média, extremo superior, extremo inferior) e sintetizar os resultados. Essa triangulação permite, assim, pintar um retrato fiel e profundo das personalidades socioeconômicas de cada bairro. Vale ressaltar que considerou-se utilizar, para esta parte do trabalho, a divergência de *Kullback-Leibler*, que diferencia a variação entre duas distribuições probabilísticas. Estatisticamente, este seria o caminho mais completo a seguir. Porém, visando priorizar a fácil interpretabilidade dos resultados e a menor complexidade dos experimentos, optou-se pela abordagem da análise triangular. Além do mais, nas análises 2 e 3 da abordagem triangular, que medem a concentração de uma variável binária categórica, usou-se a métrica *WRAcc* como função de qualidade para o algoritmo [Atzmueller 2015].

4. Resultados e Discussões

Todos os experimentos compreendidos neste estudo foram realizados em uma máquina Intel(R) Core(TM) i5-1334U, CPU com 4.60 GHz de frequência de processamento, 16GB

de memória RAM e sistema operacional Ubuntu 22.04 LTS 64 bits. Como já descrito anteriormente, os experimentos estão divididos em 2 blocos: análise comparativa do valor do m^2 e análise triangular da distribuição dos padrões de acabamento, sendo que para todos eles realizou-se uma análise global da cidade e uma análise focada nos bairros, resultando em um total de 4 experimentos. Os resultados encontrados em cada um deles estão apresentados a seguir, assim como uma interpretação extraída dos padrões retornados. Vale ressaltar que todos os descritores presentes nos resultados foram gerados automaticamente pela ferramenta, ou seja, não houve ajuste manual em nenhum deles.

4.1. Experimento 1: Análise Comparativa do Valor do M^2 da Cidade Inteira

Neste experimento, utilizou-se as variáveis Bairro, Ano de Avaliação, Idade e Tipo Construtivo do imóvel como descritores e a variável Valor do M^2 como o atributo alvo a ser analisado. Além disso, após retornar os resultados, calculou-se a média do Valor do M^2 do subgrupo corrigido pela inflação, a fim de identificar os grupos que realmente tiveram uma valorização acima da inflação. Os resultados são apresentados na tabela 1, sendo que o impacto nominal e real (últimas duas colunas) correspondem à variação da média do valor do m^2 do subgrupo com relação às médias globais da cidade (destacadas abaixo).

Média Global Nominal: R\$ 3.275,19

Média Global Real: R\$ 5.688,29

Tabela 1. Análise dos Subgrupos de Maior Impacto Nominal

Posição	Bairro	Ano Avaliação	Idade	Tipo Construtivo	Tamanho (Nº imóveis)	Impacto Nominal	Impacto Real
1	-	2024	-	Apartamento	18.845	+41,98%	-13,89%
2	-	2024	-	-	21.522	+38,70%	-15,89%
3	Savassi	-	-	-	5.992	+57,45%	+53,56%
4	Savassi	-	-	Apartamento	5.939	+55,74%	+51,12%
5	-	2024	[1:6]	-	3.829	+67,68%	+1,99%
6	-	2024	[1:6]	Apartamento	3.730	+68,15%	+2,27%
7	-	2023	-	Apartamento	17.371	+30,48%	-19,56%
8	-	2023	-	-	19.858	+28,08%	-21,04%
9	Savassi	-	[1:6]	-	1.786	+85,78%	+74,31%
10	Savassi	-	[1:6]	Apartamento	1.786	+85,78%	+74,31%

Para melhor visualização dos resultados, a métrica de qualidade dos subgrupos da tabela 1 foi removida. Contudo, os resultados da tabela estão ordenados em ordem decrescente de acordo com esta métrica, ou seja, o primeiro resultado contém o maior valor, enquanto os demais apresentam valores sequencialmente menores. O resultado original e completo do algoritmo pode ser consultado em arquivo separado⁴, contendo os valores da métrica de qualidade para cada subgrupo. O algoritmo foi executado com os parâmetros *depth* (quantidade de descritores máxima para cada subgrupo) igual a 3 e *k* (quantidade dos melhores subgrupos encontrados) igual a 10.

Percebe-se 3 padrões predominantes nos resultados. Primeiramente, o único bairro presente no top 10 do algoritmo é o *Savassi*. O bairro Savassi se destaca como a área de

⁴https://drive.google.com/file/d/1vMZqBqpKP48MkUdfjwwP7_CRRfJE3XIp/view?usp=sharing

maior e mais consistente valorização, superando a média global da cidade tanto em valores nominais quanto reais. O subgrupo de imóveis na Savassi (posição 3), com 5.992 propriedades, apresenta um valor nominal 57,45% acima da média nominal do valor do m² global da cidade. Mais importante, seu valor real também é 53,56% superior à média do valor real do m² global da cidade, indicando que a valorização do bairro consistentemente supera a inflação. Esta valorização acima da inflação se mantém ao focar apenas em apartamentos (posição 4), que representam a quase totalidade das amostras do bairro (5.939 imóveis), apresentando um impacto nominal de +55,74% e um real de +51,12%. Imóveis novos na Savassi (idade entre 1 e 5 anos) são ainda mais valorizados (posições 9 e 10), com um impacto nominal de +85,78% e um impacto real de +74,31% sobre as respectivas médias globais da cidade.

Os resultados de maior qualidade (posições 1 e 2) revelam um padrão econômico muito interessante sobre os imóveis avaliados nos últimos anos. Os imóveis avaliados em 2024 (21.522 unidades) possuem um preço nominal 38,70% acima da média global do valor nominal do m² da cidade. No entanto, quando a inflação é descontada, o valor real desses imóveis é, na verdade, **15,89% inferior** à média do valor real do m² global da cidade. Essa diferença de valorização com o desconto da inflação é ainda mais acentuada para os imóveis de 2023. Eles apresentam um valor nominal 28,08% acima da média global do valor nominal do m² da cidade, mas um valor real **21,04% abaixo** da média do valor real do m² global da cidade (posição 8). Isso significa que, embora os preços nominais nos últimos anos estejam altos, a valorização desses imóveis não foi suficiente para acompanhar a inflação acumulada, fazendo com que, em termos de poder de compra real, eles valessem menos que a média real histórica da cidade. Encontrar subgrupos com essa característica é exatamente o motivo pelo qual a abordagem de usar as duas métricas (Nominal e Real) é tão poderosa.

Em particular, a análise da idade dos imóveis revela uma exceção importante a essa diferença de valorização com a correção da inflação nos últimos anos. Imóveis muito novos (idade entre 1 e 5 anos) avaliados em 2024 tiveram um valor nominal massivo, 67,68% acima da média global do valor nominal do m² da cidade. Diferente da tendência geral de 2024, esse subgrupo de imóveis novos conseguiu ter um impacto real positivo de +1,99% (posição 5). O mesmo ocorre ao focar apenas em apartamentos novos (idade entre 1 e 5 anos), com um impacto real de +2,27% (posição 6). Isso indica que o mercado em 2024 pagou um prêmio tão alto por imóveis novos que, mesmo em um cenário de valorização real negativa para o mercado geral, a novidade foi um fator suficiente para, por pouco, superar a inflação.

4.2. Experimento 2: Análise Comparativa do Valor do M² de Cada Bairro

Apesar dos resultados do experimento anterior revelarem características interessantes sobre os dados imobiliários da cidade, o tipo de abordagem utilizado (comparação com os dados da cidade inteira) favorecem o aparecimento de subgrupos convencionalmente usuais [Herrera et al. 2010]. Especificamente, apenas 1 bairro apareceu nos top 10 resultados. Aumentar a quantidade de subgrupos retornados também não ajudaria, pois o algoritmo imprimiria apenas bairros de regiões “nobres”, como Belvedere, Lourdes, Santo Agostinho, etc. Como o objetivo deste trabalho é analisar a dinâmica de diversas regiões da cidade, foi necessário alterar a estratégia comparativa. Para esta análise, filtrou-se os dados para cada bairro e rodou-se o algoritmo individualmente para cada grupo. Os

parâmetros do algoritmo foram alterados para *depth* igual a 2 e *k* igual a 5. Abaixo, estão apresentados os principais achados deste experimento, analisados juntamente com o Modelo *Gemini Pro*⁵. O resultado original e completo do algoritmo pode ser consultado em arquivo separado⁶.

4.2.1. A Dominância da “Miragem Inflacionária” em 2024

O padrão mais forte e generalizado em praticamente todos os bairros, de diferentes perfis e localizações, é o de uma forte valorização nominal em 2024, que não se converte em valorização real quando comparada à média do valor real do m² do próprio bairro. Este fenômeno foi observado de forma consistente desde bairros como Buritis (Nominal: +49.74%, Real: -13.87%) e Castelo (Nominal: +53.35%, Real: -9.74%) até bairros como Planalto (Nominal: +29.74%, Real: -18.82%) , Padre Eustáquio (Nominal: +37.61%, Real: -18.35%) e Barreiro (Nominal: +30.63%, Real: -18.21%). Isso sugere que a alta nos preços de 2024 é um reflexo de um movimento macroeconômico de inflação, e não necessariamente de uma valorização intrínseca generalizada dos imóveis na cidade. A maioria dos imóveis vendidos em 2024, na prática, representou um poder de compra menor que a média histórica de seus respectivos bairros.

4.2.2. Imóveis Novos como principal Vetor de Valorização Real

A análise em larga escala confirma que o antídoto mais comum para a superação da valorização do m² com a correção da inflação é a novidade. Segmentos de imóveis recém-construídos ou com poucos anos de idade são frequentemente os únicos que conseguem apresentar um impacto real positivo. Abaixo estão listados alguns exemplos notáveis de ganho real:

- **Santo Antônio:** Imóveis com menos de 2 anos superaram a média real do bairro em **+30,02%**.
- **Anchieta:** Imóveis novos (idade < 2) avaliados em 2024 tiveram um impacto real de **+28,34%**.
- **Estoril:** O mesmo perfil de imóveis (idade < 2, ano 2024) gerou um ganho real de **+21,25%**.
- **Palmares:** Imóveis de 1 a 5 anos de idade em 2024 tiveram um ganho real de **+19,14%**.

Em muitos bairros, o segmento de imóveis novos é o que transforma um resultado de perda real em ganho. No bairro **Serra**, por exemplo, enquanto 2024 como um todo teve um impacto real de -17,16% , os imóveis novos (idade < 2) no mesmo ano tiveram um impacto real positivo de +12,59%. O mesmo ocorre no **Gutierrez** (real de +10,88% para imóveis com menos de 2 anos de idade em 2024) e no **Prado** (real de +9,82% para imóveis entre 1 e 4 anos de idade em 2023). Isso sugere que a renovação do estoque imobiliário (verticalização e novas construções) é o motor mais consistente de criação de riqueza real no mercado de Belo Horizonte, conseguindo superar a inflação de forma robusta em diversas regiões.

⁵gemini.google.com

⁶https://drive.google.com/file/d/1XsPSGGB1_nGGafgS-DZshFU0225VDt6o/view?usp=sharing

4.2.3. A Descoberta de Nichos e Dinâmicas Particulares

A análise em múltiplos bairros revelou dinâmicas que não eram visíveis em uma escala menor, mostrando como diferentes bairros possuem personalidades de mercado distintas.

- **A Valorização Geográfica em Bairros “Nobres”:**
 - **Lourdes e Savassi:** Confirmam ser mercados à parte, onde o maior valor está em **casas antigas**, com ganhos reais de **+139,33%** e **+181,86%**, respectivamente.
 - **Funcionários e Carmo:** Exibem este mesmo padrão, com casas antigas gerando ganhos reais massivos de **+148,30%** e **+68,89%**, respectivamente. Esta é uma característica de bairros centrais, consolidados e de alto padrão, onde o valor do terreno e o potencial construtivo são imensos.
- **Bairros de Valorização Antiga:**
 - **Luxemburgo:** Diferente da maioria, seus padrões de maior valorização real não são recentes. O ano de **2017** se destaca, com imóveis novos (idade < 2) daquele ano gerando **+63,75%** de ganho real.
 - **Calafate:** Também mostra força em anos passados, com imóveis de **2015** gerando ganhos reais de até **+48,45%**.
- **Bairros com Dinâmica de Renovação Intensa:**
 - **Santo Antônio, Serra, Sion, Anchieta, Estoril, Palmares, Barro Preto e outros:** Nestes bairros, os padrões de maior destaque e com ganhos reais positivos estão quase sempre ligados à idade < X. Isso indica que são áreas passando por um forte processo de modernização e verticalização, onde o mercado valoriza o novo em detrimento do estoque mais antigo.
- **Bairros com Desempenho Real Consistentemente Positivo:**
 - **Jardim Vitória e Céu Azul:** Diferente da maioria, estes bairros mostram vários subgrupos com ganho de valor real. No **Jardim Vitória**, apartamentos em geral tiveram um ganho real de **+7,29%** e os avaliados em 2021, de **+2,82%**. No **Céu Azul**, imóveis de 1 a 4 anos tiveram ganho real de **+23,61%**.

4.3. Experimento 3: Análise Triangular dos Padrões da Cidade Inteira

Para este experimento, mudou-se a perspectiva de análise do *preço* para o *perfil* dos imóveis. Agora, avaliou-se a cidade de acordo com a distribuição dos padrões de acabamento dos imóveis, conforme desenvolvido na Seção 3. Assim, como anteriormente, o experimento inicial focará em uma análise global da cidade e, depois, a análise será feita por bairros. Os parâmetros do algoritmo foram iguais ao Experimento 1: *depth* igual a 3 e *k* igual a 10. O resultado original e completo do algoritmo pode ser consultado em arquivo separado⁷. Abaixo estão resumidos os principais achados desta análise, novamente com a ajuda do *Gemini Pro*.

4.3.1. Relatório da Análise 1: Padrão de Acabamento MÉDIO

O objetivo desta primeira análise é identificar segmentos com um padrão de qualidade médio geral significativamente superior à média da cidade. O “Índice de Padrão Médio” da

⁷<https://drive.google.com/file/d/1X-ouhiqZxoNCCfLbCpsLn77ZzTzPljgr/view?usp=sharing>

cidade é de **3.22** (em uma escala de 1 a 5). Esta análise é excelente para identificar segmentos de grande volume com qualidade superior (imóveis novos) e bairros que são, em sua totalidade, estruturalmente mais bem acabados que a média da cidade.

- **Principal Fator de Qualidade: Idade do Imóvel**

O fator mais forte que eleva o padrão médio de um imóvel na cidade é a sua idade. O segmento de apartamentos com 1 a 5 anos de idade, composto por mais de 82.000 imóveis, possui um índice médio de 3.54, quase 10% acima da média geral. Isso confirma que novas construções em toda a cidade tendem a ter um padrão médio superior ao estoque mais antigo.

- **Bairros de Estrutura Superior:** Geograficamente, um grupo seletivo de bairros se destaca por ter um padrão médio estruturalmente mais alto:

- **Belvedere:** É o líder absoluto, com um índice médio de **4.55** (+41,46% acima da média).

- **Santo Agostinho:** Segue com um índice de **4.09** (+26,91%).

- **Savassi e Lourdes:** Também apresentam um perfil de alta qualidade, com índices de **3.95** e **3.86**, respectivamente.

- **O Caso do Buritis:** É notável a presença do **Buritis** entre os principais resultados, com um índice médio de **3.61** (+12,16% acima da média). Isso indica que, embora não seja um polo de extremo luxo, o bairro possui uma qualidade construtiva média geral sólida e superior à da cidade.

4.3.2. Relatório da Análise 2: Concentração de ALTO Padrão (P5)

O objetivo desta segunda análise é identificar os bolsões de extremo luxo, ou seja, segmentos com uma concentração de imóveis P5 muito acima da média. A fins de comparação, a concentração global de imóveis P5 na cidade é de **4,79%**. Esta análise confirma a existência de uma elite imobiliária muito bem definida geograficamente. Enquanto a cidade tem menos de 5% de imóveis P5, bairros como Belvedere e Santo Agostinho concentram a grande maioria desses ativos, especialmente em seus lançamentos mais recentes.

- **Elite Geográfica:** A concentração de luxo é massivamente dominada por poucos bairros.

- **Belvedere:** Novamente em primeiro lugar, com impressionantes **57,82%** de seus imóveis sendo P5. Seus apartamentos chegam a 61,28% de P5.

- **Santo Agostinho:** Um forte segundo lugar, com **37,95%** de concentração P5.

- **Savassi e Lourdes:** Formam um segundo escalão de elite, com cerca de **22%** e **20%** de concentração P5, respectivamente.

- **A combinação explosiva:** O padrão mais extremo de luxo ocorre ao combinar localização de elite com imóveis novos.

- Em **Santo Agostinho**, imóveis com 1 a 5 anos de idade possuem uma concentração de P5 de **60,98%**. Isso significa que 3 em cada 5 imóveis novos neste bairro são do mais alto padrão.

4.3.3. Relatório da Análise 3: Concentração de Padrão BÁSICO (P1)

O objetivo desta terceira análise é identificar os segmentos com a maior concentração de imóveis de padrão mais simples. A fins de comparação, a concentração global de

imóveis P1 na cidade é muito baixa, de apenas **0,72%**. Ao contrário do luxo (P5) que é geograficamente concentrado, o padrão básico (P1) está mais associado ao **tipo e idade do imóvel** (casas mais antigas) em toda a cidade.

- **O Perfil do Imóvel Básico:** O fator mais determinante para um padrão P1 é o tipo construtivo.
 - **Casas:** Este segmento tem uma concentração de P1 de **3,88%**, mais de 5 vezes a média da cidade.
 - **Casas Antigas:** A tendência se acentua com a idade. Casas com mais de 31 anos têm **4,55%** de concentração P1.
- **Localização do Padrão Básico:** Apenas um bairro apareceu com destaque nesta análise.
 - **Santa Mônica:** Apresentou uma concentração de P1 de **2,94%**, mais de 4 vezes a média da cidade, indicando um perfil de imóveis mais simples nesta região.

4.3.4. Relatório Geral e Cruzamento de Dados: As Personalidades dos Segmentos

Nenhum subgrupo específico domina as três análises. No entanto, ao cruzar as informações, podemos identificar as personalidades de diferentes segmentos e bairros, confirmando a validade da abordagem triangular.

- **Personalidade 1: O Polo de Luxo Consolidado (Ex: Bairro Belvedere)**
 - **Análise 1 (Média):** Possui o maior índice de padrão médio da cidade (4.55).
 - **Análise 2 (Luxo):** Possui a maior concentração de P5 (57,82%).
 - **Análise 3 (Básico):** Não aparece nos resultados, indicando uma baixíssima concentração de P1.
 - **Diagnóstico:** Perfil claro de um bairro de elite, com alta qualidade média e uma concentração massiva do padrão mais alto.
- **Personalidade 2: O Sólido Padrão Médio-Alto (Ex: Bairro Buritis)**
 - **Análise 1 (Média):** Apresenta um índice de padrão médio alto (3.61), superior à média da cidade.
 - **Análise 2 (Luxo):** Não aparece nos top resultados, indicando que sua concentração de P5 não é excepcional.
 - **Análise 3 (Básico):** Não aparece nos top resultados, indicando uma baixa concentração de P1.
 - **Diagnóstico:** Este cruzamento revela a verdadeira identidade do Buritis. Não é um bairro de extremo luxo (poucos P5), mas também não é básico. É uma região **homogênea e de padrão consistentemente bom**, provavelmente com uma vasta maioria de imóveis P3 e P4, o que eleva sua média geral.
- **Personalidade 3: O Segmento Polarizado (Ex: “Casas”)**

Este é o achado mais interessante, que confirma exatamente a hipótese sobre a limitação da média.

 - **Análise 1 (Média):** O segmento de “Casas” não se destacou por uma média excepcionalmente alta ou baixa.
 - **Análise 2 (Luxo):** Também não se destacou como um polo de P5.

- **Análise 3 (Básico): Dominou completamente os resultados**, mostrando que as casas têm uma concentração de P1 muito acima da média.
- **Diagnóstico:** A média do padrão das casas é “normal”, mas isso esconde uma **polarização interna**. O segmento de casas em BH contém tanto as mansões de luxo que vimos na análise por bairro no Experimento 2 (em Lourdes, Funcionários, Savassi, etc.) quanto uma quantidade muito significativa de imóveis de padrão P1, que puxam a média para baixo e a equilibram. A análise triangular foi a única forma de revelar essa dualidade.

4.4. Experimento 4: Análise Triangular dos Padrões de Cada Bairro

Finalmente, este último experimento repete a abordagem da comparação triangular utilizada anteriormente, mas agora isolou-se os dados dos bairros para identificarmos as variações específicas de cada uma dessas regiões. Os parâmetros do algoritmo foram iguais ao Experimento 2: *depth* igual a 2 e *k* igual a 5. O resultado original e completo do algoritmo pode ser consultado em arquivo separado⁸. Abaixo estão resumidos os principais achados desta análise, novamente com a ajuda do *Gemini Pro*.

4.4.1. Impressões Gerais: As Três “Tribos” de Bairros em BH

Ao cruzar os resultados das três análises para múltiplos bairros, é possível agrupá-los em três perfis principais, que definem a dinâmica de seu mercado imobiliário.

1. **A Elite Consolidada:** São bairros que possuem um **Índice de Padrão Médio muito alto**, uma **alta concentração de imóveis P5** e uma **concentração quase nula de imóveis P1**. Eles são consistentemente luxuosos em todas as métricas.
 - **Exemplos Clássicos:** Belvedere, Santo Agostinho, Lourdes, Savassi, Sion, Carmo.
 - **Características Principais:** Nestes bairros, a “sofisticação” é a norma. A análise interna revela que os imóveis novos (idade < 3) são ainda mais luxuosos, mostrando que o mercado continua a se intensificar no topo da pirâmide. O bairro **Lourdes** é um exemplo perfeito, onde não há sequer imóveis P1 para análise.
2. **Os Bairros de Contraste e “Ilhas de Luxo”:** Este é um grupo fascinante. São bairros que, no geral, possuem um **Índice de Padrão Médio baixo ou mediano** e uma **concentração de P5 baixíssima** (muitas vezes abaixo de 1%). No entanto, a análise revela “agulhas no palheiro”: nichos muito específicos onde a concentração de P5 explode.
 - **Exemplos Notáveis:**
 - **União:** Possui apenas 3,13% de P5, mas um segmento de imóveis super novos (idade < 1) de 2022 teve **89,47%** de concentração P5.
 - **Nova Suíça:** Com apenas 1,67% de P5, encontrou um nicho de imóveis novos (idade < 1) em 2023 com **81,36%** de P5.
 - **Ipiranga:** Com 6,57% de P5, revelou um padrão de imóveis novos (idade < 2) em 2017 com **90,74%** de P5.

⁸<https://drive.google.com/file/d/1Z1emYiaJSA07pn4qExqKB7wa2wtJP1DG/view?usp=sharing>

- **Característica Principal:** Estes bairros não são uniformes. Eles abrigam um mercado imobiliário predominantemente padrão, mas são alvo de empreendimentos de altíssimo padrão muito pontuais e recentes, que criam estas “ilhas de luxo” e mudam o perfil de quadras específicas.
3. **Os Bairros de Perfil Popular e Tradicional:** Este grupo é caracterizado por um **Índice de Padrão Médio mais baixo**, uma **concentração de P5 nula ou quase nula**, e uma **concentração de P1 acima da média** da cidade (que é de 0,72%).
- **Exemplos Clássicos:**
 - **Jardim Vitória:** Índice médio de 2.59 e **nenhum imóvel P5**. Em contrapartida, sua concentração de P1 é de 4,75%, e o segmento de **casas** eleva essa taxa para **15,33%**.
 - **Céu Azul:** Sem imóveis P5, possui uma concentração de P1 de 4,87%, com as **casas mais antigas** (idade ≥ 28) sendo os principais tipos de imóveis deste padrão P1.
 - **Santa Mônica:** Sem imóveis P5, tem uma alta concentração de P1 de 2,94%, especialmente em **apartamentos mais antigos** (idade ≥ 29).
 - **Característica Principal:** A dinâmica destes bairros é definida pela predominância de imóveis de padrão mais simples, especialmente no estoque de casas mais antigas.

4.4.2. Análise Detalhada de Casos Excepcionais

Alguns bairros merecem um destaque especial por revelarem dinâmicas únicas. Esta análise final, cruzando as três abordagens, permitiu criar um mapa detalhado da personalidade socioeconômica de cada bairro, revelando padrões que nenhuma análise isolada conseguiria capturar.

1. Belvedere - O Epicentro do Luxo Intensivo

- **Linha de Base:** O Belvedere já parte de um patamar extraordinário, com um **Índice de Padrão Médio de 4.55** e **57,82%** de todos os seus imóveis sendo P5.
- **Padrão Excepcional:** Mesmo neste contexto, o algoritmo encontrou que os **apartamentos com menos de 4 anos** são ainda mais exclusivos, elevando a concentração de P5 para **72,95%**. Em 2011 e 2012, o índice médio dos imóveis novos atingiu o teto de **5.00**.
- **Diagnóstico:** O Belvedere não é apenas um bairro de luxo; é um mercado que continua em um ciclo de **intensificação do luxo**, onde cada nova construção eleva ainda mais o padrão já altíssimo da região. A ausência quase total de imóveis P1 reforça seu status de enclave homogêneo de altíssimo padrão.

2. Bairro União - A Explosão Pontual de 2022

- **Linha de Base:** O União é um bairro de bom padrão, com um índice médio de 3.31 e uma concentração de P5 de 3,13%.
- **Padrão Excepcional:** Em 2022, um grupo de **57 imóveis com menos de 1 ano** de idade registrou uma inacreditável concentração de P5 de **89,47%**.

- **Diagnóstico:** Este achado provavelmente corresponde a um ou dois grandes empreendimentos de altíssimo padrão entregues e comercializados naquele ano. É um exemplo perfeito de uma “ilha de luxo” temporal. Esse evento foi tão forte que elevou o índice médio de padrão daquele nicho para 4.79, muito acima da média de 3.31 do bairro. Isso mostra como lançamentos específicos podem distorcer completamente o perfil de uma área em um curto período.

3. **Buritis vs. Santo Antônio - Duas Realidades de “Bom Padrão”**

Ambos são bairros valorizados, mas a análise triangular revela suas personalidades distintas.

- **Buritis:** Tem um **índice médio alto (3.61)**, mas uma **concentração de P5 modesta (3.57%)**. Seu perfil de luxo é difuso e ligado a imóveis novos de forma geral, como os de 2011, que tiveram 26% de P5.
- **Santo Antônio:** Tem um **índice médio ligeiramente inferior (3.47)**, mas uma **concentração de P5 três vezes maior (10.15%)**. Seus picos de luxo são muito mais intensos, como os imóveis novos de 2022 (idade < 2), que alcançaram 65,41% de concentração P5.
- **Diagnóstico Comparativo:** O Buritis é um bairro de **qualidade média consistentemente boa**, com muitos imóveis P3 e P4. O Santo Antônio, por outro lado, parece ser mais **polarizado**, com um estoque geral bom, mas com nichos de extremo luxo muito mais pronunciados, especialmente nos lançamentos recentes.

5. Conclusão e Trabalhos Futuros

Analisar a dinâmica imobiliária de uma cidade pode ser uma tarefa trabalhosa e exaustiva, devido ao grande volume de dados gerados a cada ano. Este trabalho se propôs a analisar os dados históricos do ITBI da cidade de Belo Horizonte, publicamente disponibilizadas pela prefeitura. Apesar de ser uma base de dados bastante limitada (por conter registros apenas de imóveis que foram transmitidos), os experimentos provaram que ainda é possível extrair informações relevantes em um contexto assim, desde que os dados e as técnicas aplicadas sejam trabalhadas da forma correta. Os subgrupos encontrados trouxeram insights valiosos com relação à dinâmica imobiliária urbana, conseguindo identificar nichos de mercados específicos para alguns grupos de imóveis. Os resultados apresentados neste estudo foram confirmados por uma especialista na área⁹.

Um ponto interessante para este trabalho consiste no fato de que a abordagem metodológica aqui empregada pode ser adaptada para ser aplicada nos dados de qualquer outra cidade que também disponibilize publicamente seus dados do ITBI. Isso se deve ao fato de que foi possível, neste estudo, identificar dinâmicas interessantes na cidade mesmo com poucas variáveis. Entretanto, o atributo padrão de acabamento nem sempre estará presente nos dados do ITBI das outras cidades, pois a prefeitura de Belo Horizonte executa o trabalho manual de construir esta variável para ser incluída em suas bases de dados imobiliárias.

Com relação a trabalhos futuros, sugere-se avaliar os experimentos 3 e 4 (análise comparativa da distribuição dos padrões de acabamento) com a métrica da divergência

⁹ Ana Luiza Nabuco: Doutora em Economia pela UFMG e doutora em Estudos Urbanos pela *École des Hautes Études en Sciences Sociales* (EHESS), Paris. Pesquisadora Associada ao Centro de Desenvolvimento e Planejamento Regional (CEDEPLAR)/UFMG

de *Kullback-Leibler*, em substituição à abordagem triangular, como sugerido na Seção 3. Acredita-se que este caminho revelará nuances mais profundas sobre a variação deste padrão ao longo dos subgrupos. Finalmente, um último experimento consiste em juntar as 2 variáveis alvo aqui estudadas para avaliar suas variações em conjunto, uma estratégia conhecida como *Mineração de Modelos Excepcionais* [Duivesteijn et al. 2015]. Isso é perfeitamente possível neste caso, pois os padrões baixos geralmente possuem um metro quadrado também pequeno e os padrões altos apresentam um valor de metro quadrado mais elevado, revelando um modelo de regressão crescente. O algoritmo, assim, buscaria por subgrupos onde a relação Preço \sim Padrão quebra a norma da cidade. Isso nos permitiria entender as diferentes lógicas de precificação que governam os micro-mercados dentro da cidade. Trata-se de um caminho analítico fascinante e muito avançado.

Referências

- Atzmueller, M. (2015). Subgroup discovery. *WIREs Data Mining and Knowledge Discovery*.
- Bernardes, F. C. (2018). A definição do conceito de valor venal como base de cálculo do iptu e do itbi. *Congresso Pernambuco de Direito Tributário*.
- Duivesteijn, W., Feelders, A. J., and Knobbe, A. (2015). Exceptional model mining. *Data Mining and Knowledge Discovery - SPRINGER*, 30:47–98.
- Herrera, F., Carmona, C. J., González, P., and del Jesus, M. J. (2010). An overview on subgroup discovery: foundations and applications. *Knowledge and Information Systems - SPRINGER*, 29:495–525.
- IGP-M (2025). <https://portal.fgv.br/especiais/igp-m-resultados>. [Online; Acessado em Julho de 2025].
- Lavrač, N., Kavšek, B., Flach, P., and Todorovski, L. (2004). Subgroup discovery with cn2-sd. *The Journal of Machine Learning Research*, 5:153–188.
- Lemmerich, F. and Becker, M. (2018). pysubgroup: Easy-to-use subgroup discovery in python. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 658–662.
- Nabuco, A. L. (2021). Terra urbana, propriedade e gentrificação na periferia: dinâmica imobiliária e capitalismo patrimonial no vetor norte de belo horizonte, brasil, 1999 a 2019. *FACE - UFMG*.
- Nadal, V. W. D. (2023). A correção monetária de créditos tributários: uma análise sobre a utilização da taxa selic e do igp-m. *Revista Direito Tributária Atual*, 48.
- Rangel, L. A. D. and Gomes, L. F. A. M. (2007). Determinação do valor de referência do aluguel de imóveis residenciais empregando o método todim. *Revista Brasileira de Pesquisa Operacional*, 27.
- Smith, N. (2007). Gentrificação, a fronteira e a reestruturação do espaço urbano. *Revista GEOUSP Espaço e Tempo*, 11.
- Soares, I. M., Teixeira, E. C., and Tupy, I. S. (2025). Análise da relação entre criminalidade e valor por metro quadrado dos imóveis residenciais no município de belo horizonte. *Revista Desenvolvimento Regional em debate*, 15.