



RASPAGEM DE DADOS PARA AVALIAÇÃO IMOBILIÁRIA: UM ESTUDO DE CASO PARA MAPEAMENTO DE RISCOS E OPORTUNIDADES

AUTORES

Bruno Henrique Gazzinelli

Engenheiro Civil Especialista em Avaliações e Perícias de Engenharia, Crea-MG 235.299/D, IBAPE 1.113
Belo Horizonte/MG
bruno@bhengenharia.com

PALAVRAS-CHAVE

Raspagem de Dados, Avaliações de Imóveis, Mercado Imobiliário, Engenharia de Avaliações.

Introdução

A raspagem de dados, também conhecida como Web Scraping, é o processo de extração automatizada de informações de websites. A técnica envolve o uso de um software ou script para percorrer páginas da web, identificar e extrair dados relevantes, como texto, imagens, links ou qualquer outro tipo de conteúdo estruturado. Ainda, permite coletar grandes quantidades de dados de maneira eficiente, eliminando a necessidade de extração manual.

Na área de avaliação imobiliária, a raspagem de dados pode ser utilizada para coletar informações sobre propriedades, como endereço, e características físicas, tais como área construída, área total, preço de venda ou aluguel, histórico de transações, entre outros.

Figura 1 - Desenvolvimento da Raspagem e Análise de Dados



Fonte: Autor

Plataforma de Anúncio de Imóveis e Ambientes de Coleta de Dados

Para a concepção deste trabalho, foram escolhidas as plataformas Plataforma 01 e Plataforma 02, de forma a possibilitar a realização de comparativo de dados obtidos de ambas as plataformas e que se diferenciam, além da independência dos anúncios, das declarações de campos textuais, bem como da descrição interna elaborada pelos Anunciantes.

2. Sequência Lógica de Obtenção de Dados

Apresenta-se, a seguir, o algoritmo de sequência lógica desenvolvido para raspagem de dados separado por etapas:

Figura 2 - Sequência Lógica de Obtenção de Dados



Fonte: Autor

Fez-se a escolha da Savassi como região para o estudo de caso por se tratar de um bairro localizado no município de Belo Horizonte, conhecido por sua localização privilegiada, bem como de grande disponibilidade de oferta de imóveis, de diferentes tipologias.

Panorama Geral de Dados

Apresenta-se, a seguir, o Panorama Geral dos Dados, obtido de forma não refinada, através de anúncios para fins de aluguel, no período entre 15 a 22 de Maio de 2023, dividido por Plataforma de Anúncio e Tipologia de Imóvel:

Tabela 1 - Panorama Geral de Imóveis - Total de 1.091 Imóveis

Panorama Geral de Imóveis			
	Plataforma 01	Plataforma 02	Total
Residencial	256,00	56,00	312,00
Comercial	502,00	277,00	779,00

Fonte: Autor

3. Análise técnica

A análise técnica se baseará inicialmente nos dados obtidos após as etapas de estruturação e tratamento de dados, trazendo luz às inconsistências das declarações dos dados de Classe Textuais e Numéricas.

Em sequência, faz-se uma análise com foco na geração de inteligência de mercado, através da análise dos campos de Classe Qualitativa.

3.1 - Inconsistência de Dados

Duplicidade de Anúncios

Tabela 2 - Remoção de Duplicatas de Imóveis

Panorama de Remoção de Duplicatas de Imóveis						
	Plataforma 01			Plataforma 02		
	Iniciais	Duplicatas	Restantes	Iniciais	Duplicatas	Restantes
Residencial	256,00	7,00	249,00	56,00	0,00	56,00
Comercial	502,00	10,00	492,00	277,00	72,00	205,00

Fonte: Autor

Informações Cadastrais Incompletas de Endereço

Verifica-se também, para fins cadastrais, a qualidade da declaração dos endereços de cada um dos anúncios remanescentes:

- Bairro;
- Logradouro;
- Numeração Completa e Válida.

Tabela 3 - Declaração incompleta de Endereço

Panorama de Declaração de Informação Cadastral Incompleta de Endereço						
	Plataforma 01			Plataforma 02		
	Bairro	Logradouro	Nº Completo	Bairro	Logradouro	Nº Completo
Residencial	47,00	133,00	69,00	12,00	23,00	21,00
Comercial	59,00	220,00	213,00	60,00	91,00	54,00

Fonte: Autor

Indisponibilidade de Imagens para Conferência de Características

Verificou-se, ainda, os anúncios que não possuem imagens para verificação das características declaradas pelo Anunciante, apresentadas a seguir:

Tabela 4 - Indisponibilidade de Imagens para Conferência

Panorama de Indisponibilidade de Imagens para Conferência				
	Plataforma 01		Plataforma 02	
	Disponíveis	Indisponíveis	Disponíveis	Indisponíveis
Residencial	249,00	0,00	49,00	7,00
Comercial	469,00	23,00	161,00	44,00

Fonte: Autor

3.2 – Observações e Inteligência de Mercado

Num olhar distinto da busca por inconsistências na declaração dos dados, pode-se verificar também informações valiosas quando analisados dados da Classe Qualitativa. Assim, desenvolvem-se algumas perspectivas interessantes a seguir:

Quantidade de Declarações de Características Qualitativas Distintas

Realizando-se o levantamento quantitativo das características declaradas na Classe Qualitativa, pode-se observar o seguinte universo de distribuição:

Tabela 5 - Características Qualitativas - Levantamento Quantitativo

Declaração de Características Qualitativas - Qtde de Termos Independentes		
	Plataforma 01	Plataforma 02
	Caixa de Características Interna	Caixa de Características Interna
Residencial	128,00	60,00
Comercial	86,00	32,00

Fonte: Autor

Histograma de Frequência de Termos e Nuvem de Palavras das Características Qualitativas

Em seguida, realizou-se, quantitativamente, o levantamento do histograma de frequência dos termos, para cada um dos cenários considerados. Apresenta-se, a seguir, o mais representativo:

Conclusões e recomendações

Dado o Estudo de Caso em tela, pode-se chegar às seguintes constatações, de forma resumida:

Riscos

- Diferentes estruturas de publicidade de dados entre as plataformas geram um nível alto de incompatibilidade para comparações diretas;
- A não realização de conferência visual, técnica que demanda investimentos mais robustos em recursos computacionais e tempo de execução, diminui o nível de confiabilidade da análise dos dados.

Oportunidades

- A geração de um histograma de frequência, bem como uma nuvem de palavras, pode indicar qualitativamente quais são os termos mais declarados e, por consequência, com maior percepção de valor para aquele determinado mercado;
- A realização da Raspagem de Dados, feita periodicamente, pode reduzir as inconsistências e o descarte de dados, bem como fornecer informações valiosas quanto à velocidade de comercialização de um determinado imóvel e das flutuações de preços declarados no mercado.

Considerações finais e observações para trabalhos futuros

O levantamento de informações via Raspagem de Dados, realizado de forma rápida, demonstra-se muito mais adequado à geração de conhecimento de mercado do que à aplicação direta nos modelos de avaliação imobiliária.

Referências bibliográficas

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 14653-1 – Norma brasileira para avaliação de bens – Parte 1: Procedimentos gerais**. São Paulo: ABNT, 2019.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 14653-2 – Avaliação de bens - Parte 2: Imóveis urbanos**. São Paulo: ABNT, 2011.

MACHADO, Bruno Guia Neto. **Avaliação Imobiliária com Redes Neurais Artificiais**. Instituto Superior de Engenharia de Lisboa, Janeiro de 2022. Dissertação de Mestrado.

SANTOS, João Manuel Azevedo. **Real Estate Market Data Scraping and Analysis for Financial Investments**. September 10, 2018. Mestrado Integrado em Engenharia Informática e Computação.