

Trabalho de Aprendizado Descritivo:

Mineração de Dados Urbanos

Curso: Aprendizado Descritivo

Aluno(a): Gruilherme Namen Pimenta

Orientador(a): Prof. Dr. Renato Vimieiro

Belo Horizonte, 30 de Julho de 2024



U F *m* G

Outline

1 Introdução

- Questionamentos
- Objetivos

2 Processamentos dos Dados

- Coleta
- As Regionais de Belo Horizonte
- Características

3 Metodologia

- Arquitetura de Dutos e Filtros

4 Resultados

- Itemsets Frequentes
- Itemsets estatisticamente relevantes
- Utilidade em m^2 construídos
- Utilidade negativa m^2
- Descoberta de Subgrupo - Alvo Padrão de Acabamento P5
- Descoberta de Subgrupo - Alvo Tipo Construtivo Apartamento
- Descoberta de Subgrupo - Gentrificação
- Gentrificação

5 Conclusão e Trabalhos Futuros

Introdução

O processo uso e ocupação do solo dos grandes centros urbanos é alvos de diversos trabalhos da área da sociologia (ANDRADE; MENDONÇA, 2020; VELOSO; ANDRADE, 2019; SOLLA, 2019).

Com o auxílio da Política de Dados Abertos, os municípios têm divulgados muitas informações, possibilitando grandes avanços no uso de técnicas de aprendizado de máquina para auxiliar as políticas sociais e econômicas.

Questionamentos

Questionamentos

- P1. Quais os algoritmos de mineração de dados são úteis para descrever uma base de dados urbanos?
- P2. Quais informações eles podem gerar?
- P3. Qual é a melhor sequência de algoritmos para descrever os dados e formar a arquitetura de dutos?

Objetivos

Objetivo Geral

Adaptar a arquitetura de dutos e filtros de descoberta de conhecimento em banco de dados (NWAGU; OMANKWU; INYIAMA, 2017), utilizando um conjunto mínimo de algoritmos de mineração de dados.

Gerar conhecimento aplicado ao uso e ocupação do solo de áreas urbanas.

Objetivos Específicos

Coleta, limpeza e armazenamento de dados;

Pré processamento de dados;

Definir metodologia para escolha dos algoritmos;

Executar os algoritmos;

Apresentar o conhecimento gerado;

Responder as perguntas de pesquisa;

Coleta

Coleta dos Dados

Os dados forma coletados no Portal de Dados Abertos da Prefeitura de Belo Horizonte,
<https://ckan.pbh.gov.br/>.

Os dados são referentes às guias de IPTU da data de 03/06/2024.

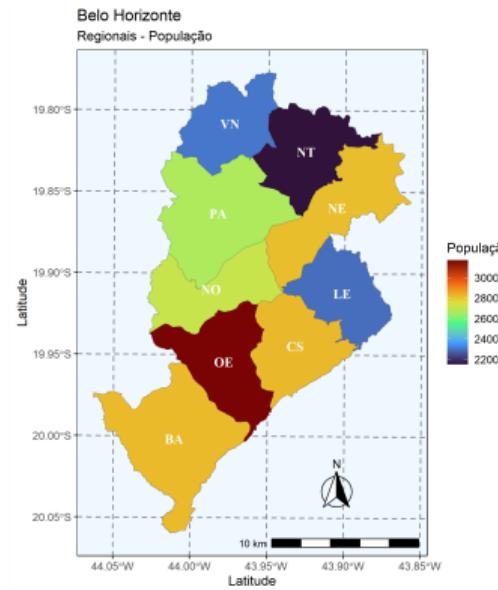
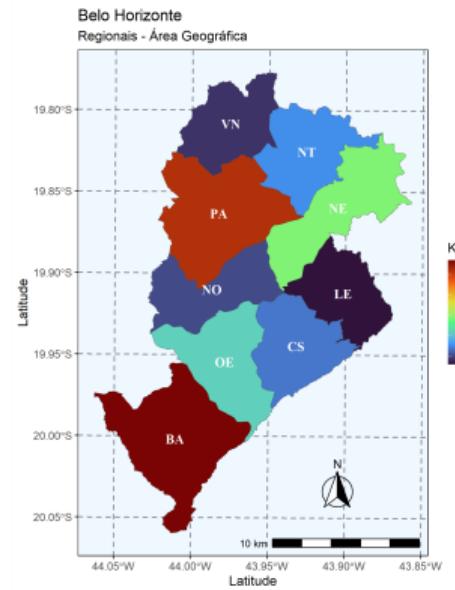
As informações são da Secretaria Municipal de Fazenda.

Descrição dos Dados

Os dados contém várias informações dos imóveis, tais como endereço, área do terreno, área construída, geometria do terreno e outras. A descrição completa encontra-se no Portal de dados abertos

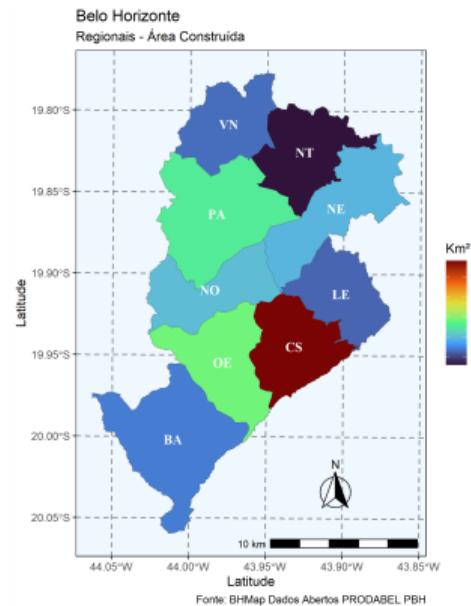
As Regionais de Belo Horizonte

Mapas das Regionais



As Regionais de Belo Horizonte

Mapas das Regionais



Características

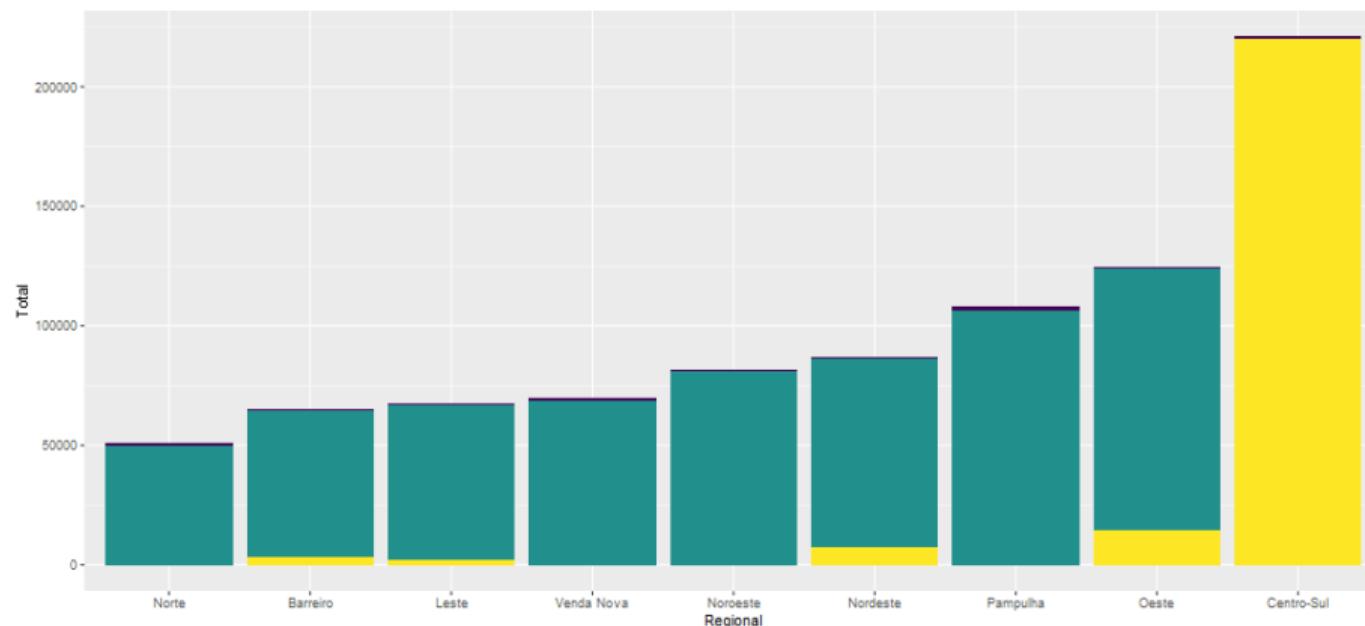
Coleta de Lixo

Figura 1

Coleta de Lixo

Total por Regional

SEM COLETA COLETA ALTERNADA COLETA DIARIA



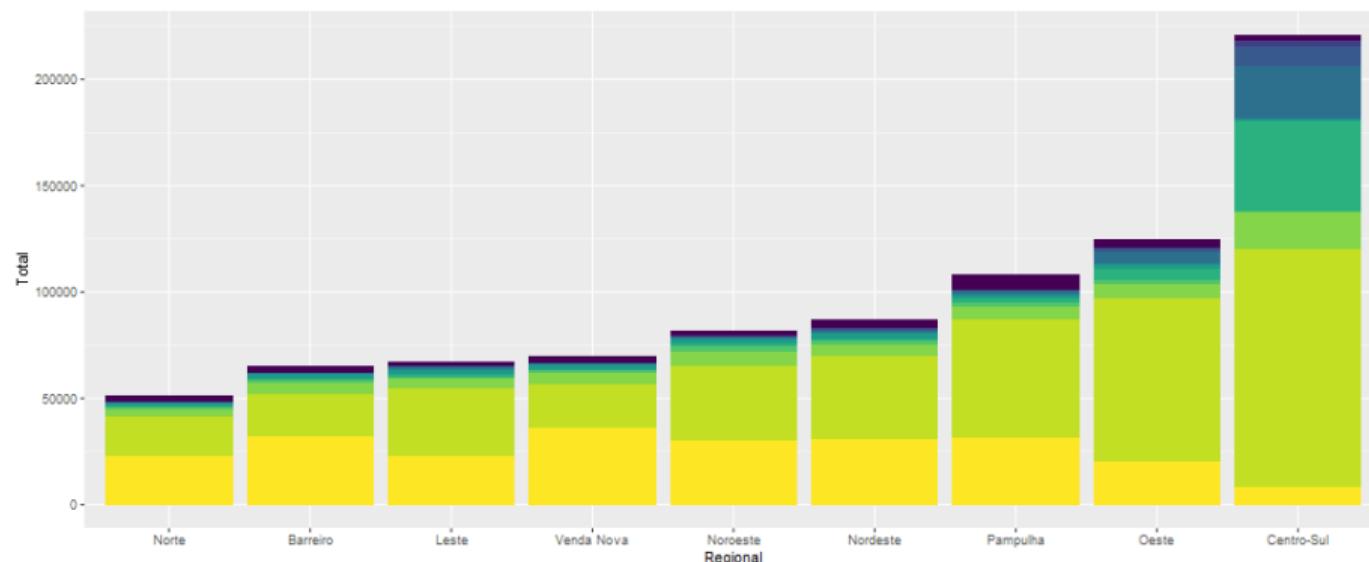
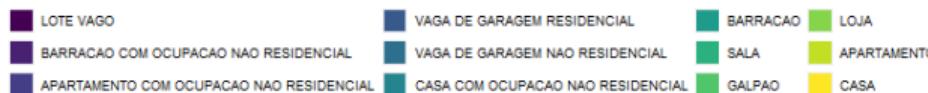
Características

Perfil

Figura 2

Perfil dos Imóveis

Total por Regional



Fonte: Portal de Dados Abertos da PBH

Características

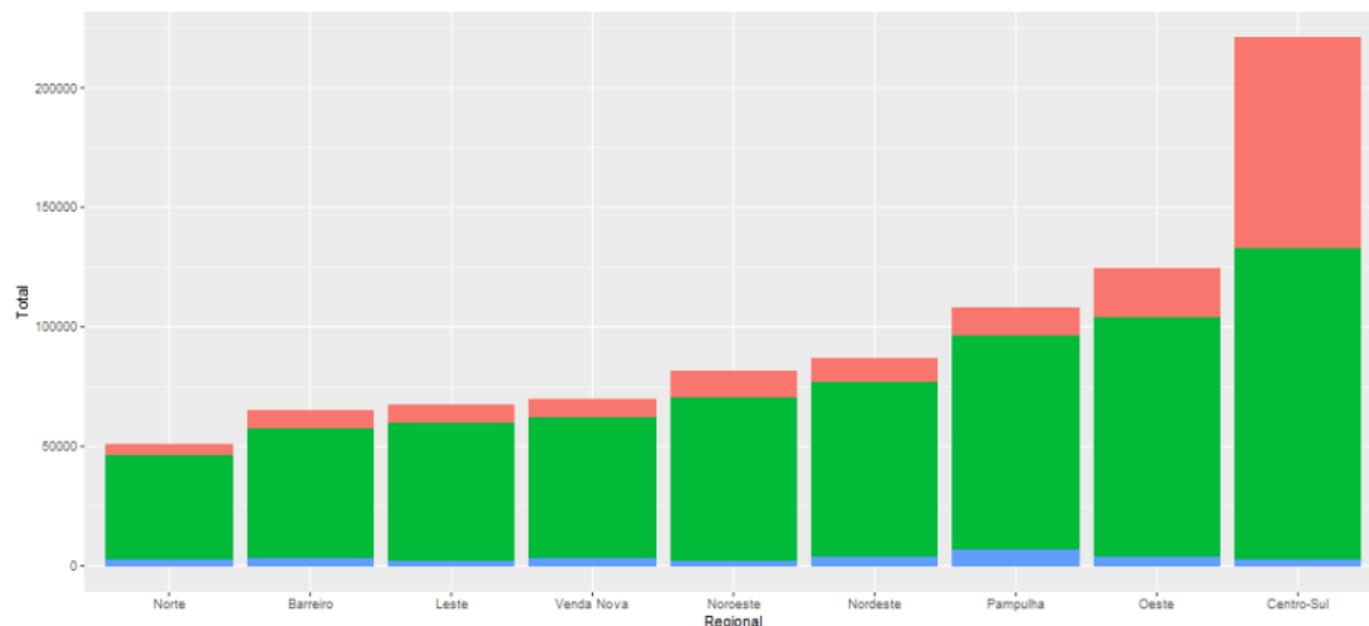
Ocupação

Figura 3

Perfil da Ocupação

Total por Regional

■ NAO RESIDENCIAL ■ RESIDENCIAL ■ TERRITORIAL



Características

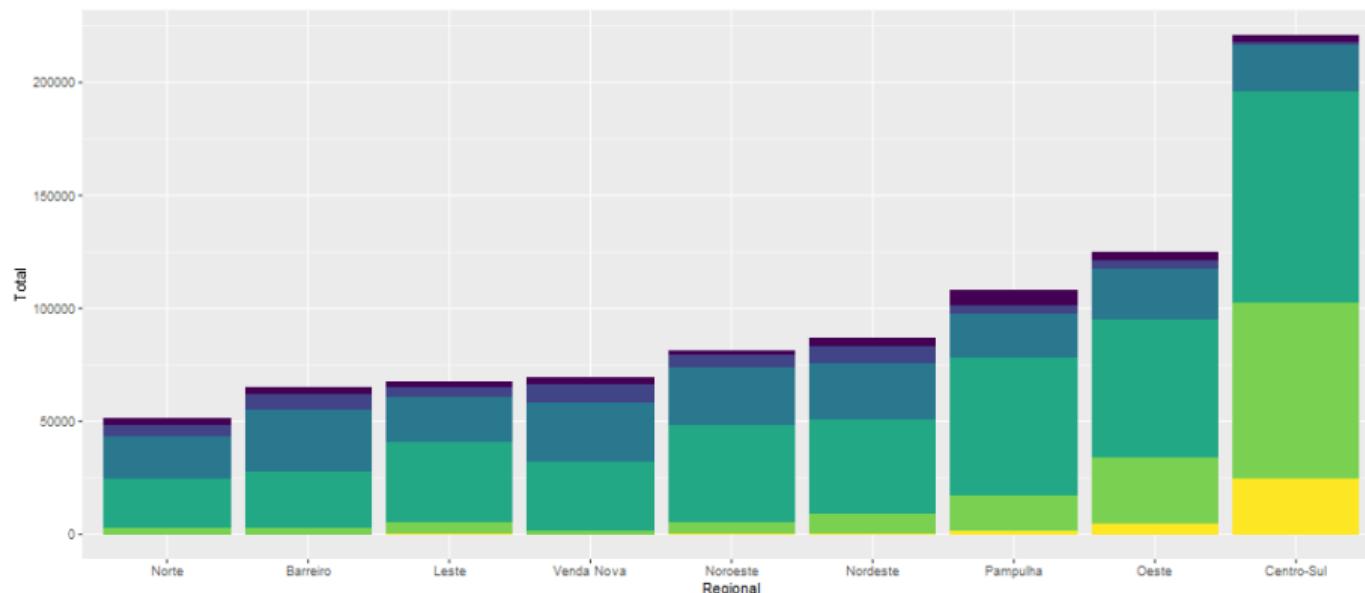
Qualidade do Acabamento

Figura 4

Qualidade do Imóvel

Total por Regional

TE P2 P4
P1 P3 P5



Fonte: Portal de Dados Abertos da PBH

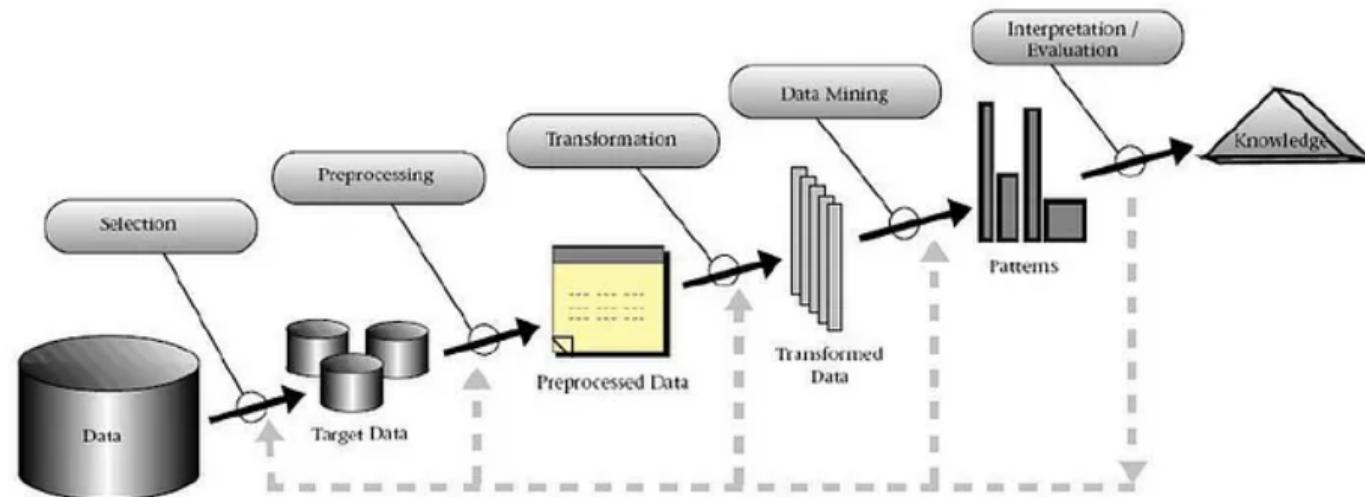
Metodologia

Para a escolha dos algoritmos foi utilizado o Princípio da Descrição de Comprimento Mínimo (DCM) da mesma forma que foi utilizado na tese de Proença (PROENÇA, 2021).

Para guiar o processo de criação da arquitetura foi utilizado a metodologia *Goal Question Metric* (GQM) (CALDIERA; ROMBACH, 1994), metodologia muito utilizada no campo da Engenharia de Software (SOMMERVILLE, 2011).

Estrutura da Arquitetura de Dutos e Filtros

Figura: Exemplo de um pipeline



Definição das Fases

Pré-Processamento

Uso da ferramenta **GritBot** da RuleQuest Research para detectar anomalias (*outliers*).

Transformação

Outlier: Maiores áreas construídas;

Item: Tipo Construtivo X Padrão de Acabamento X Outlier;

Transações: agrupamento por CEP;

Utilidade: área construída;

Utilidade Negativa: área do terreno vago;

Mineração de Dados

Algoritmos

FPClose - Algoritmo eficiente para obter os itemsets fechados mais freqüentes;

OpusMiner - Algorítmico para obter os itemsets estatisticamente relevantes;

FHM Freq - Algoritmo para obter os itemsets de maior utilidade e suporte.

FHN - Algoritmo para processar itemsets de utilidade negativa;

Cortana - *Software* para processar subgrupos de interesse.

Resultados

Regionais Distantes do Centro

Caracterizadas pela maioria de casas de padrões de acabamento entre um e três e lotes vagos;

A Pampulha se destaca por ter casas de padrão 4 e não apresentar casas com acabamento padrão 1.

Regionais Vizinhas ao Centro

Regionais Oeste e Nordeste predominam casas com padrão igual às regionais mais distantes;

Leste e Noroeste possuem um padrão misto de casas, barracões, comércio local e lotes vagos, com padrões variando do 1 ao 3.

Regional Centro Sul

Todos os itemset apresentaram apenas um item, com padrões variando do 3 ao 4;
Indicação de segmentação ordenada do espaço.

Itemsets estatisticamente relevantes

Resultados

Indicou a forte presença conjugada de salas comerciais de alto padrão de acabamento com a existência de vagas de garagem;

Indicou adaptação de imóveis de moradia para a finalidade comercial conjugados com imóveis comerciais;

Indicou a presença conjugadas de imóveis de maior área construída próximos a imóveis padrão de mesmo acabamento ou superiores;

A regional Centro-Sul possui uma grande quantidade de salas de alto padrão conjugadas com vagas de garagem.

Utilidade em m² construídos

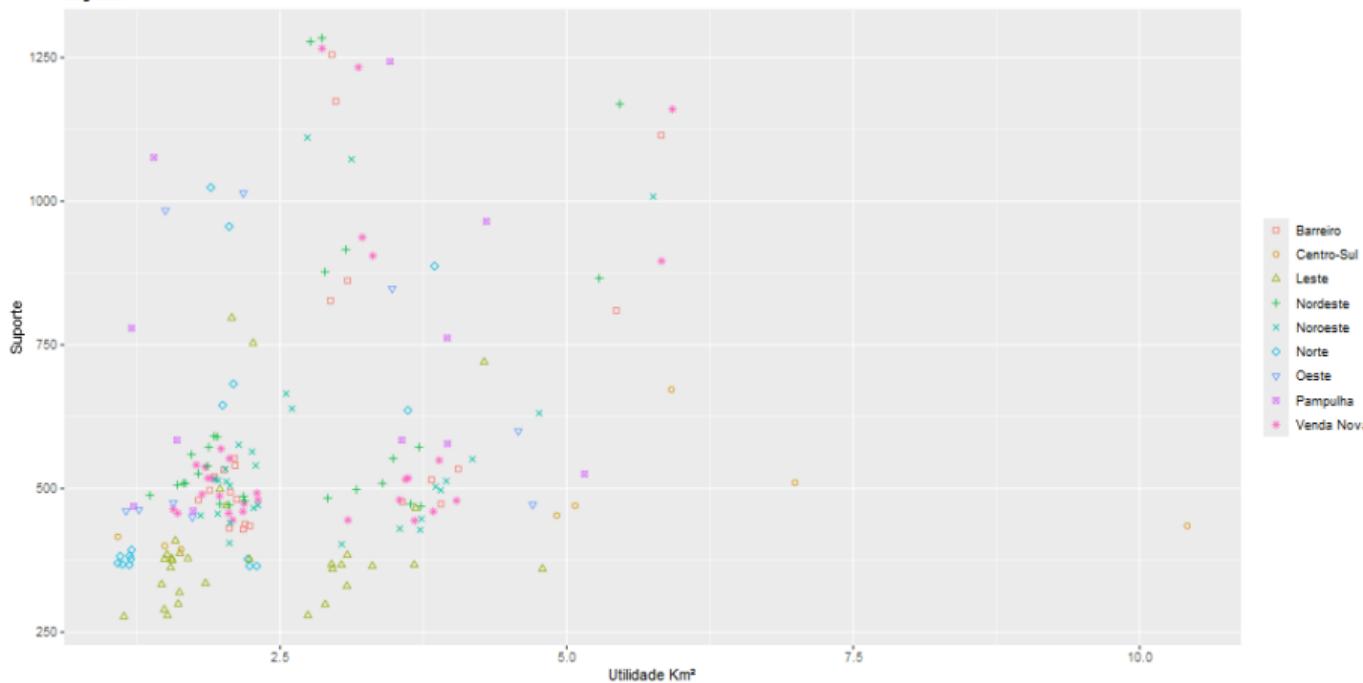
Presença de comércio local nas regionais Barreiro, Nordeste, Norte e Venda Nova;
Regional Oeste apresenta uma grande quantidade de apartamentos com acabamento 3, totalizando mais de 4 km² construídos;
A regional Noroeste possui uma grande área construída destinada a galpões;
A regional Centro-Sul possui muitos centros comerciais com muita área construída.

Utilidade em m² construídos

Figura 5

Relação Suporte x Utilidade

Regional

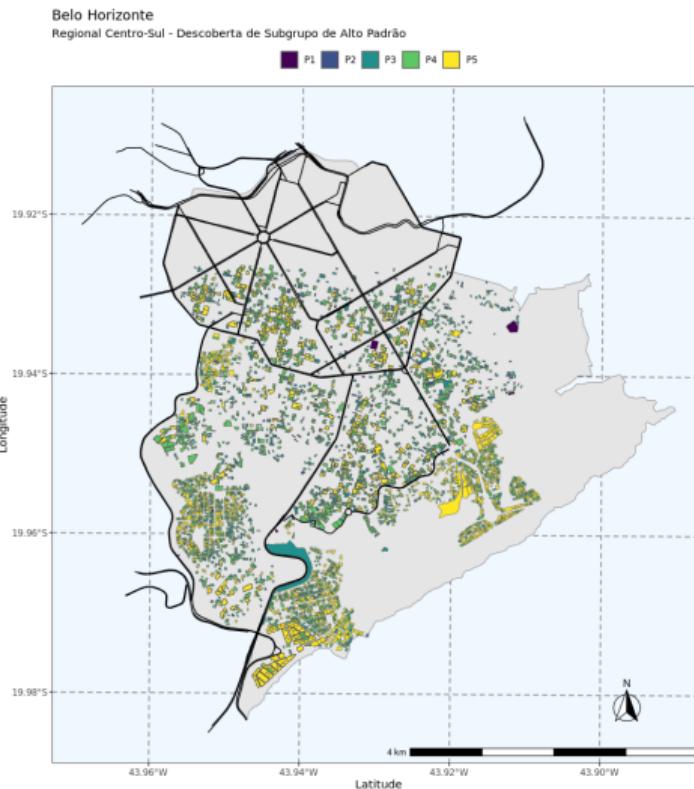


Fonte: Portal de Dados Abertos da PBH

Utilidade negativa m²

As regionais Pampulha e Norte tiveram um impacto negativo em lotes vagos, indicando áreas vagas próximas a terrenos com construção.

Descoberta de Subgrupo - Alvo Padrão de Acabamento P5



Descoberta de Subgrupo - Alvo Tipo Construtivo Apartamento



Descoberta de Subgrupo - Gentrificação

Zona Homogenia	Núcleos Familiáres
CS407	10
CS313	12
CS210, CS212	13
CS411	23



Respostas

- P1. Os algoritmos de mineração de dados mais úteis para descrever uma base de dados foram: FPMax, OpusMiner, FHM Freq, FHN e Descoberta de Subgrupo. Esta sequência de algoritmos descrevem os dados de forma mais genérica à mais específica.
- P2. Eles conseguem, em sequência, gerarem um conjunto de informações mais genéricas até as mais específicas.
- P3. A melhor sequência de execução dos algoritmos foi a apresentada no trabalho, a grande descoberta do trabalho foi que para utilizar o algoritmo de descoberta de subgrupos de forma mais eficiente, a melhor forma, é compreender a base de dados para segmentar muito bem os registros que serão processados.

Conclusão

O trabalho demonstrou a eficácia de uma arquitetura de dutos e filtros com algoritmos de mineração de dados para descrever o espaço urbano de Belo Horizonte de forma compacta, revelando padrões e tendências no uso e ocupação do solo a partir de dados imobiliários.

Trabalhos Futuros

Destaca-se a incorporação de dados espaço-temporais para análise de séries históricas.

A integração com outras fontes de dados, como redes sociais e imagens de satélite.

E uso de outras técnicas de mineração de dados como a mineração de dados de mesma localidade, para evitar a segmentação por regionais.

OBRIGADO!