

**CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS
CAMPUS TIMÓTEO**

Nathésia de Moura Guerra

**AVALIAÇÃO DA QUALIDADE DE DADOS GEOGRÁFICOS NA
INFORMAÇÃO CORPORATIVA: UM ESTUDO EXPLORATÓRIO**

Timóteo

2017

Nathésia de Moura Guerra

**AVALIAÇÃO DA QUALIDADE DE DADOS GEOGRÁFICOS NA
INFORMAÇÃO CORPORATIVA: UM ESTUDO EXPLORATÓRIO**

Monografia apresentada à Coordenação de Engenharia de Computação do Campus Timóteo do Centro Federal de Educação Tecnológica de Minas Gerais para obtenção do grau de Bacharel em Engenharia de Computação.

Orientador: Leonardo Lacerda Alves

Timóteo

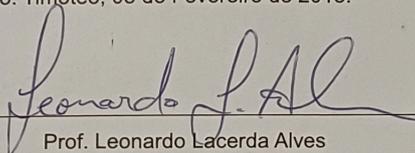
2017

Nathésia de Moura Guerra

**AVALIAÇÃO DA QUALIDADE DE DADOS GEOGRÁFICOS NA
INFORMAÇÃO CORPORATIVA: UM ESTUDO EXPLORATÓRIO**

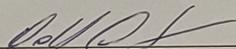
Monografia apresentada ao Curso de
Engenharia de Computação do Centro
Federal de Educação Tecnológica de Minas
Gerais para a obtenção do título de
Engenheiro de Computação.

Trabalho aprovado. Timóteo, 08 de Fevereiro de 2018:



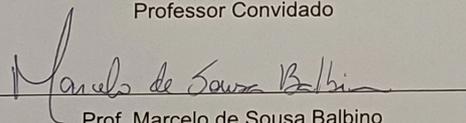
Prof. Leonardo Lacerda Alves

Orientador



Prof. Odilon Correa da Silva

Professor Convidado



Prof. Marcelo de Sousa Balbino

Professor Convidado

Timóteo

2018

Agradecimentos

Agradeço primeiramente à Deus que me permitiu chegar até aqui.

Agradeço à minha família, em especial meus pais Neuza e Joel por me darem suporte. Agradeço ao Jean por me encorajar em cada momento difícil dessa caminhada.

Ao meu orientador Leonardo Lacerda Alves pelo conhecimento compartilhado e paciência para me orientar.

Agradeço aos meus amigos que me deram suporte em todo o desenvolvimento.

Agradeço ao CEFET, aos professores, equipe administrativa e aos demais funcionários que contribuíram fortemente para a minha formação acadêmica.

Resumo

A administração urbana é complexa pois exige conhecimento dos fatores que influenciam no território. Esses fatores sofrem transformações constantes da sociedade e avaliá-los pode ser caro. Para uma gestão eficiente, é necessário identificar as informações dos fenômenos que interferem na administração urbana e agir em tempo hábil, antes que ocorra mudanças significativas no cenário urbano. Uma alternativa para identificar os fenômenos que interferem na gestão urbana é avaliar informações armazenadas não apenas em bancos de dados de entidades da gestão pública. Bancos de dados de associações comerciais, industriais e de prestação de serviços possuem informações úteis para a administração urbana. Os sistemas de informação geográfica (SIG) viabilizam as análises dos dados referenciados geograficamente, permitindo realizar o tratamento automatizado dessas informações. Essa pesquisa avalia informações de um bancos de dados de uma associação. Para essa análise foram utilizados mapas de calor para representar os dados tabulares do banco de dados graficamente. Foi realizado uma investigação com especialistas nas informações armazenadas no banco de dados para avaliar o método e as informações do banco de dados.

Palavras chaves: geocodificação, mapa de calor, atividade econômica.

Abstract

The urban administration is complex because it requires knowledge of the factors that influence the territory. These factors undergo constant transformations in society, and evaluating them can be costly. For efficient management, it is necessary to identify the information about the phenomena that affect urban administration and act in a timely manner before significant changes occur in the urban scenario. An alternative to identifying the phenomena that interfere with urban management is to assess information stored not only in public management databases. Databases from commercial, industrial, and service associations contain useful information for urban administration. Geographic information systems (GIS) enable the analysis of geographically referenced data, allowing for the automated processing of this information. This research evaluates information from an association's database. For this analysis, heat maps were used to graphically represent the tabular data from the database. An investigation was conducted with experts on the information stored in the database to evaluate the method and the information from the database.

Keywords: geocoding, heat map, economic activity.

Lista de ilustrações

Figura 1 – Exemplo da estrutura das informações armazenadas nos arquivos JSON de endereços	20
Figura 2 – Exemplo da estrutura das informações armazenadas nos arquivos JSON de coordenadas	21
Figura 3 – Mapa de calor do comércio em Timóteo	23
Figura 4 – Mapa de calor de serviços em Timóteo	23
Figura 5 – Mapa de calor da indústria metal mecânica em Timóteo	24
Figura 6 – Mapa de calor da indústria alimentícia em Timóteo	25
Figura 7 – Mapa de calor da indústria química em Timóteo	25

Lista de tabelas

Tabela 1 – Erros na geocodificação dos endereços	21
Tabela 2 – Respostas mapa comércio	26
Tabela 3 – Respostas mapa serviço	27
Tabela 4 – Respostas mapa indústria metal-mecânica	27
Tabela 5 – Respostas mapa indústria alimentícia	27
Tabela 6 – Respostas mapa indústria química	28

Lista de abreviaturas e siglas

ACIATI	Associação Comercial Industrial, Agropecuária e de Prestação de Serviços de Timóteo
API	Aplicação
CEP	Código de Endereçamento Postal
JSON	JavaScript Object Notation
SIG	Sistema de Informação Geográfica
UF	Unidade Federativa

Sumário

1	INTRODUÇÃO	10
1.1	Problema	11
1.2	Objetivos	11
2	PROCEDIMENTOS METODOLÓGICOS	13
2.1	Coleta e organização	13
2.2	Geocodificação das empresas	13
2.3	Produção dos mapas temáticos	14
2.4	Avaliação dos mapas	14
3	FUNDAMENTOS HISTÓRICOS, TEÓRICOS E METODOLÓGICOS	15
3.1	Cartografia temática	15
3.2	Sistema de informação geográfica	16
4	RESULTADOS	18
4.1	Coleta e organização	18
4.2	Geocodificação	19
4.3	Produção dos mapas temáticos	22
4.4	Avaliação dos mapas	24
5	CONCLUSÃO	29
5.1	Contribuições	29
5.2	Limitações	30
5.3	Trabalhos futuros	31
	REFERÊNCIAS	33
	ANEXOS	34

1 Introdução

A administração urbana é uma tarefa complexa que exige abrangente conhecimento da realidade de todo o território, assim como sobre os principais fatores que influenciam o ambiente. Isso se dá porque o ambiente urbano contempla um conjunto grande de variáveis e de atores (REZENDE; FREY; BETINI, 2003). Adicionalmente, a administração urbana tem sido desafiada com mudanças sociais, econômicas, políticas e tecnológicas contínuas.

Porém, apenas a percepção mais precisa da realidade socioespacial de cada região permitiria atender melhor aos anseios da população (KURKDJIAN; PEREIRA,). Uma vez que os recursos são limitados, é necessário gerenciá-los de maneira que os mesmos sejam utilizados de forma eficiente na organização urbana, na busca por melhores soluções.

Além disso, é essencial que as informações estejam atualizadas para refletir adequadamente a disponibilidade de recursos e que as soluções sejam encontradas rapidamente. Como a cidade está em constante mudança, é fundamental monitorar o ambiente, reavaliar as informações, identificar as necessidades mais atuais e tomar decisões antes que a realidade se altere.

Diante de um cenário de mudanças, é necessário avaliá-las com instrumentos e procedimentos que proporcionam rapidez e agilidade na monitoração. Alguns estudos de áreas urbanas podem ser custosos pelo extenso número de informações de campos diferentes, dificultando a análise em tempo hábil. Uma possível alternativa para esse problema é utilizar informações além das armazenadas em bancos de dados de entidades da gestão pública. Seria possível cruzar as informações que já são conhecidas por outras instituições, criando representações mais complexas da realidade com custo menor.

No entanto, a integração de diferentes informações pode ser difícil para os usuários, com potencial de deixar as informações confusas. Através da representação das informações em um mapa, a monitoração poderia ser realizada de forma eficiente, apesar da dificuldade de se encontrar dados geográficos precisos e de qualidade. Por outro lado, o planejamento e o conhecimento do território se tornam mais simples por meio de um mapa, uma vez que a comunicação visual pode facilitar muito o entendimento de informações antes textuais e com relações geográficas menos óbvias. Portanto, seria possível elaborar mapas que facilitem a compreensão das informações textuais ou tabulares e que possibilitem a visualização de correlações dessas informações pelo leitor (ARCHELA; THÉRY, 2008).

Kurkdjian e Pereira () defendem que a informação compartilhada com toda a população facilita o diálogo entre os diferentes atores urbanos. Tal compartilhamento permite a identificação das necessidades urbanas por agentes do setor público e de toda a sociedade. Isso também facilitaria a justificativa das ações realizadas para suprir as necessidades urbanas. Por meio do conhecimento dessas atividades, seria possível elaborar políticas públicas de acordo com a necessidade da sociedade e a identificação das características do ambiente seriam utilizadas como fundamento para o planejamento de políticas públicas como o tráfego de trânsito, segurança ou saúde, por exemplo.

Outro exemplo de política pública que pode ser beneficiada por um planejamento urbano compartilhado é o desenvolvimento econômico. E assim como acontece com as demais políticas públicas, muitas das informações necessárias para o gerenciamento e mapeamento dependem de muitas fontes que precisam ser atualizadas continuamente e rapidamente.

Entretanto, a distribuição de atividades econômicas pode ser ser mais facilmente monitorada com informações de fontes privadas, além dos dados de instituições de gestão pública. A utilização de informações de instituições particulares não vinculadas à gestão pública tem o potencial ajudar na monitoração, tendo em vista que os bancos de dados dessas instituições armazenam informações úteis e atualizadas. Ainda assim, a viabilidade da criação de representações espaciais a partir da adoção de quaisquer bancos de dados que não tenham compromisso com atividades de administração pública é um problema em aberto.

1.1 Problema

Este trabalho visa responder a seguinte pergunta: As informações disponíveis em bancos de dados privados são suficientes para indicar a real distribuição geográfica de atividades econômicas em uma região?

Com essa pergunta, o presente estudo pretende se concentrar exclusivamente em uma política pública e em um escopo muito reduzido de uso da informação corporativa. Se for verdadeira as hipóteses de viabilidade e precisão de um ou alguns fenômenos socioeconômicos em bancos de dados privados, pesquisas ou produtos baseados em informações privadas de fontes heterogêneas se mostrariam recursos úteis e baratos para o planejamento e monitoração de administração pública.

1.2 Objetivos

Para responder o problema de pesquisa, este trabalho objetiva principalmente avaliar se as informações em um banco de dados de associação comercial são suficientes para indicar a real distribuição geográfica de atividades econômicas em uma região.

Para atender ao objetivo proposto, são indicados os seguintes objetivos específicos:

1. coletar e organizar as informações do banco de dados da Associação Comercial, Industrial, Agropecuária e de Prestação de Serviços de Timóteo (ACIATI);
2. geocodificar os endereços encontrados nos bancos de dados usados;
3. desenvolver mapas temáticos da região da amostra estudada representando as atividades econômicas
4. avaliar os mapas temáticos pela qualidade e precisão das informações disponíveis.

Este trabalho, está estruturado em cinco capítulos. O capítulo 2 trata dos procedimentos metodológicos seguidos para a execução da pesquisa. Em seguida, o capítulo 3 apresenta os principais trabalhos relacionados ao tema deste trabalho. O próximo é o capítulo 4, onde são apresentados

os resultados encontrados com o trabalho. E finalmente o capítulo 5, onde são apresentadas as conclusões encontradas a partir da análise dos resultados, limitações e trabalhos futuros .

2 Procedimentos metodológicos

Os procedimentos para a busca da resposta do problema apresentado no capítulo 1 dividem-se em quatro passos:

1. coletar e organizar as informações do banco de dados da Associação Comercial, Industrial, Agropecuária e de Prestação de Serviços de Timóteo (ACIATI);
2. geocodificar os endereços encontrados no banco de dados utilizado;
3. produzir mapas temáticos da região da amostra estudada representando as atividades econômicas;
4. avaliar os mapas temáticos pela qualidade e precisão das informações disponíveis.

2.1 Coleta e organização

Para coletar as informações do banco de dados da associação, é necessário solicitar o acesso aos dados à associação. Os dados de interesse neste trabalho são os endereços das empresas e informações que identifiquem a área de atuação econômica.

Foi adotada a Associação Comercial, Industrial, Agropecuária e de Prestação de Serviços de Timóteo (ACIATI) como estudo de caso deste trabalho. Os dados foram usados de modo a preservar a privacidade da associação comercial e de seus associados.

Após a coleta dos dados, foi necessário verificar quais informações disponíveis eram importantes para responder o problema proposto. Foi realizado um filtro para verificar as informações que permitem identificar o endereço e atividade econômica.

Todas as informações coletadas foram organizadas para permitir a identificação rápida das informações necessárias para o trabalho. Foi necessário criar um padrão para apresentar as informações de endereços, similar ao formato <Tipo de logradouro> <logradouro>, <número>, <bairro>, <cidade>, <estado>.

Os dados disponíveis foram então separados em categorias de acordo com a atividade econômica desenvolvida por cada empresa. Os dados foram armazenados em diferentes arquivos, de acordo com a atividade econômica, onde cada arquivo armazenava informações de uma categoria econômica.

2.2 Geocodificação das empresas

Para geocodificar as empresas do banco de dados da associação, foi utilizado a Interface de Programação de Aplicativo (API) Google Maps, que converteu os endereços da forma em que estavam armazenados no banco de dados em coordenadas de latitude e longitude.

Em seguida, foram atribuídas as coordenadas aos respectivos endereços do banco de dados. A conversão foi realizada em todos os endereços disponíveis no banco de dados utilizado na pesquisa.

2.3 Produção dos mapas temáticos

Em seguida, foram produzidos os mapas temáticos por meio do renderizador de mapas do Google Maps, fornecendo como entrada os pares de coordenadas de empresas de atividade econômica compatível, e obtendo mapas de calor.

Os mapas de calor deviam representar a área de localização das empresas cadastradas no banco de dados. O mapa devia ser capaz de refletir com diferenças de cores os endereços cadastrados, representando aglomerações e distribuição no território geográfico do município de Timóteo.

2.4 Avaliação dos mapas

Os mapas de calor seriam utilizados para avaliar se as informações disponibilizadas no banco de dados seriam suficientes para indicar a real distribuição geográfica de atividades econômicas na região do estudo de caso. Essa avaliação deveria ser capaz de mensurar a qualidade das informações do banco de dados e a capacidade de interpretação dos usuários-finais.

Adicionalmente, este passo deveria proporcionar uma avaliação da própria visualização do banco de dados, onde seria possível apontar maneiras de melhorar a própria informação coletada e armazenada pela proprietária do banco de dados.

O principal mecanismo de avaliação do mapa seria por meio questionário. Um questionário de interpretação do mapa produzido foi aplicado para especialistas sobre a informação corporativa utilizada. Eles responderam questões sobre a representação das informações nos mapas desenvolvidos e, a partir das respostas, foi realizada a avaliação da qualidade das informações ao se partir do pressuposto que a realidade espacial era reconhecida pelo especialista.

3 Fundamentos históricos, teóricos e metodológicos

As ferramentas de geoprocessamento surgiram para viabilizar as análises dos dados referenciados geograficamente, transformando informações textuais em mapas temáticos. Sistemas de informação geográfica (SIG) é um conjunto poderoso de ferramentas para coletar, armazenar, recuperar, transformar e visualizar dados sobre o mundo real (BURROUGH, 1986). Esses sistemas são responsáveis desde a coleta dos dados até a saída do resultado com a análise, por meio de mapas ou relatórios. Um mapa SIG deve representar fenômenos naturais ou construídos pelo homem de forma legível e intuitiva para o leitor do mapa.

3.1 Cartografia temática

A cartografia temática é uma área onde os mapas representam propriedades não apenas vistas, mas também as conhecidas. Ela surgiu no final do século XVIII e início do século XIX, com a necessidade de mapas mais completos. Seu objetivo é representar fenômenos localizáveis de um determinado tema sobre uma base com um mapa topográfico.

Cada mapa possui um tema específico, identificando um problema proposto. Os mapas temáticos possuem um foco, de acordo com o seu tema. As informações nele apresentadas, são selecionadas com o objetivo de responder algumas perguntas sobre o problema, orientando o leitor do mapa a uma visão ampla sobre a questão abordada. Essa categoria de mapas, exibe as informações necessárias para análise do tema dentro de uma região limitada geograficamente. Cada mapa é construído com técnicas com a intenção de otimizar a visualização e comunicação considerando as aptidões do perfil do leitor.

Um mapa temático proporciona análise crítica de um tema em uma região. Fenômenos geográficos, demográficos, econômicos, sociais, naturais, e vários outros podem ser representados nesse tipo de mapa por cores e símbolos.

Os elementos básicos necessários para desenhar um mapa temático, são dados estatísticos, utilizados para análise e ilustração de um determinado fenômeno de uma área e um mapa base utilizado como a base de referência do local onde o específico fenômeno está ocorrendo (TSORLINI et al., 2015).

Para um mapa coerente com a realidade são necessárias informações atualizadas das variáveis do problema trabalhado. Esses dados podem ser provenientes de fontes de diversas naturezas, bancos de dados, informação verbal, impressas ou digitais podem servir de referências. O autor do mapa seleciona as informações de acordo com a problematização da pergunta. Os dados são limitados pelo tema definido, o autor seleciona as informações relevantes para responder o problema. As informações são referenciadas geograficamente, com localização de onde o fato ocorreu quando foi coletado. Os dados - registros das situações percebidas concretamente de forma sistemática - podem

ser de natureza qualitativa ou quantitativa, ambas, ordenadas ou não (MARTINELLI, 2003).

A definição do tema e dos dados selecionados deve determinar a região territorial representada pelo mapa temático. Utiliza-se uma base com referência geográfica, normalmente um mapa topográfico como fundo do mapa temático.

As informações coletadas são apresentadas por itens visuais com diferentes formas e cores no mapa. Para representar os diversos temas é preciso recorrer a uma simbologia específica que, aplicada aos modos de implantação - pontual, linear ou zonal, aumentam a eficácia no fornecimento da informação (ARCHELA; THÉRY, 2008). Essa reprodução dos diversos elementos no mapa necessita seguir o padrão de apresentação dos métodos da cartografia temática de acordo com as características dos fenômenos trabalhados. Existem quatro categorias fundamentais para a representação dos dados:

- Qualitativa: Representa os fenômenos, sua localização e quantidade. Normalmente responde a questão "o que há em determinada região?". Pode ser representado por pontos, linhas, texturas ou áreas em um mapa. Nesse tipo de representação, o leitor consegue verificar a diversidade entre os fenômenos, se um é predominante ao outro e quais se manifestam em maior quantidade.
- Quantitativa: Reproduz a relação de proporção entre os atores do mapa. Responde a questão "quanto?". Muito comum a representação por ícones de tamanhos diferentes ou diferença na quantidade de itens sem alterar o tamanho, evidenciando a correlação de acontecimentos de um fenômeno.
- Ordenada: Expõe os fenômenos de forma ordenada, respondendo a questão "em que ordem os fenômenos acontecem?". Os fenômenos são separados por categorias ordenadas e em seguida são dispostos ordenados visualmente no mapa.
- Dinâmica: Usado quando o mapa representa os acontecimentos em tempo real. É a representação da manipulação interativa dos dados geográficos em tempo de execução.

3.2 Sistema de informação geográfica

Os dados geográficos são informações que descrevem objetos ou fatos em relação a sua posição no espaço dentro de um sistema. Essas informações possibilitam análise do espaço geográfico em que estão situadas, porém, associar grandes volumes de dados geográficos pode demandar elevado grau de compreensão, devido a extensas informações de diversas naturezas.

Geoprocessamento é uma área do conhecimento que utiliza métodos e técnicas para analisar dados geográficos transformando-os em informações relevantes em um determinado contexto. O geoprocessamento utiliza um modelo conceitual com informações usadas para representar o ambiente real que será analisado. Com a leitura dos dados geográficos identificados e estruturados, a técnica de geoprocessamento processa e elabora análise das informações trabalhadas para o leitor.

Os sistemas de informação geográfica (SIG), são ferramentas computacionais de geoprocessamento que realizam o tratamento automatizado de dados georreferenciados, trabalhando com

coleta, armazenamento, recuperação transformação e visualização de dados sobre o mundo real (BURROUGH, 1986).

SIG interpreta uma base de dados que representa características de fatos e processos do mundo de maneira ágil. A componente base de dados de um SIG é o elemento chave do sistema, pois é o que reflete a representação da realidade (PEREIRA; SILVA, 2001). As informações geográficas analisadas por um SIG são originadas de diversas fontes e por diferentes formas de coleta, a função desse sistema é ampliar o entendimento do leitor das variáveis analisadas.

Esses sistemas retornam para o usuário análises através do cruzamento dos dados geográficos interligando informações descentralizadas, organizando-as em saídas por meio de mapas temáticos ou relatórios. Os resultados de um SIG são limitados a um tema específico, para direcionar a análise dos dados, eliminando assim variáveis que não influenciam no problema analisado, simplificando a saída das informações de forma legível e intuitiva para o usuário do sistema de informação geográfica.

Para o planejamento de recursos e tomadas de decisões que necessitam análise espacial, os SIG hoje são utilizados como apoio, auxiliando a visualização dos diversos recursos disponíveis. Informações de diferentes áreas como a geologia, construção civil, hidrografia, vegetação, entre outras, podem ser mapeadas por sistemas de informação geográficas, facilitando a gestão de recursos.

4 Resultados

Esse capítulo apresenta os resultados do trabalho.

4.1 Coleta e organização

Para este trabalho, os dados foram disponibilizados pelo diretor da ACIATI, para fins exclusivamente acadêmicos, após solicitação pelos autores. As informações disponibilizadas representam uma fração dos dados cadastrais dos associados da instituição.

As informações estavam disponíveis em planilhas. Ao todo, foram disponibilizadas informações de 473 empresas cadastradas na associação. Foram disponibilizadas as seguintes informações de cada empresa:

- sócios,
- código,
- nome,
- razão social,
- endereço,
- número,
- bairro,
- município,
- UF,
- CEP,
- ramo de atividade.

Antes de as informações serem disponibilizadas, o diretor da associação comercial foi informado sobre o problema que a pesquisa pretendia responder. Por essa razão, foram utilizadas informações que permitiam identificar a localização geográfica e a atividade econômica de cada empresa, conforme apresentado abaixo:

- Localização geográfica: endereço, número, bairro, município, UF e CEP.
- Atividade econômica: nome, razão social e ramo de atividade.

As informações das empresas foram disponibilizadas em quatro planilhas, separadas de acordo com a atividade econômica: comércio, serviço, indústria e indústria metal-mecânica.

Na tentativa de reduzir as limitações da classificação, a categoria de indústria foi dividida em duas outras categorias: indústria alimentícia e indústria química. Dessa forma, a pesquisa se baseou em cinco tipos de atividades econômicas:

1. comércio,
2. serviço,
3. indústria metal-mecânica
4. indústria alimentícia,
5. indústria química.

Para dividir a categoria de indústria, as informações que permitiram identificar as atividades econômicas foram analisadas em conjunto. Uma vez que algumas empresas eram conhecidas pelos autores, foi possível identificar a atividade econômica por meio da razão social e do nome.

Para as outras empresas, foi realizada pesquisa na Internet utilizando o buscador Google, fazendo uso do nome da empresa ou sua localização. Foram realizadas pesquisas nos veículos de comunicação das empresas. Os veículos utilizados na pesquisa foram principalmente as redes sociais das empresas, em sua maioria o Facebook, e sites das próprias empresas.

A separação das empresas em categorias foi baseada no cruzamento das informações disponíveis na pesquisa e banco de dados da associação.

As informações de endereços das categorias de atividades econômicas foram armazenadas em cinco diferentes arquivos de texto no formato JSON. Logo, cada arquivo armazenou as informações de endereço de uma categoria econômica. Os arquivos armazenam um conjunto de *strings*. As strings são compostas pela concatenação das informações da localização geográfica de uma empresa, onde cada *string* representava o endereço de uma empresa.

A figura 1 retrata um exemplo da estrutura do arquivo JSON com as informações de endereço. As informações da figura 1 são compostas por dados que não foram utilizados na pesquisa. Os dados da figura 1 foram utilizados apenas para ilustrar a organização do arquivo JSON. As informações de cada *string* seguiram a seguinte ordem: endereço, número, bairro, município, UF e CEP. Cada informação foi separada por vírgula.

4.2 Geocodificação

A geocodificação dos endereços em coordenadas geográficas foi realizada por uma ferramenta elaborada para o desenvolvimento do trabalho. A ferramenta foi desenvolvida com HTML juntamente com a Google Maps JavaScript Geocoding API.

Figura 1 – Exemplo da estrutura das informações armazenadas nos arquivos JSON de endereços

```
{
  "enderecos": [
    {"e": "R. Dezenove de Novembro, 121, Centro, Timóteo, MG, 35180-008 "},
    {"e": "Av. Amazonas, 5.253, Nova Suiça, Belo Horizonte, MG, Brasil, 30.421-169"}
  ]
}
```

Fonte: elaborada pelo autor

Essa ferramenta utiliza o conceito de geocodificação, convertendo os endereços estáticos adquiridos das planilhas em coordenadas geográficas de latitude e longitude. A ferramenta consiste em uma página HTML simples com código JavaScript.

Essa ferramenta realiza a leitura de um arquivo no formato JSON que contém uma lista de *strings* de endereços. As *strings* foram armazenadas em um *script* na página HTML. Nessa ferramenta existia uma função responsável por realizar requisições ao servidor da Google Maps, onde um endereço é informado para cada requisição. Assim, cada requisição ao servidor retornava dois atributos:

1. Status: informando se a geocodificação foi executada com sucesso ou qual erro ocorreu.
2. Resultado: os pares de coordenadas geográficas de latitude e longitude, caso ocorresse com sucesso a conversão.

As requisições foram realizadas na ordem que os endereços são lidos pelo *script*. Os resultados obtidos por requisições ao servidor são armazenados em um arquivo JSON. Cada resultado com as coordenadas repassada do servidor é associado ao respectivo endereço, pelo número da requisição, seguindo a ordem do *script*. Dessa forma, é possível identificar o par de coordenadas pertence ao respectivo endereço.

De forma análoga a sessão 4.1, foram elaborados cinco arquivos JSON. Cada arquivo foi resultado da iteração da ferramenta com um arquivo JSON com informações de endereços de uma atividade econômica. Dessa forma, cada arquivo JSON obtido pela ferramenta representa os pares de coordenadas de latitude e longitude de uma atividade econômica.

Foi realizado um tratamento para erros na conversão dos endereços para coordenadas. O servidor da API somente geocodifica o endereço que ele consegue identificar o endereço. Logo, endereços sem identificação ficaram sem coordenadas e o item foi descartado.

A cada solicitação realizada ao servidor da API, a ferramenta verificou o status da requisição. Quando ocorria erro, as coordenadas não identificadas foram representadas por um código de erro no arquivo JSON de coordenadas. E quando a ferramenta não tinham terminado a execução do *script*, é realizada a próxima requisição ao servidor, repassando o próximo endereço, até finalizar todos os endereços.

A tabela 1 apresenta as informações dos endereços que não foram geocodificados. Os endereços que não foram geocodificados foram identificados com código de erro no arquivo JSON de coordenadas de latitude e longitude.

Tabela 1 – Erros na geocodificação dos endereços

Setor	Numero de registros no banco de dados	Número de erros
Comércio	263	41
Serviço	162	15
Indústria metal mecânica	16	0
Indústria alimentícia	12	0
Indústria química	7	0
Total	460	56

Fonte: Elaborada pelo autor.

A figura 2 retrata o exemplo da estrutura do arquivo JSON obtido pela ferramenta de geocodificação. No arquivo JSON de coordenadas são armazenados os pares de coordenadas de latitude e longitude e os códigos de erro, caso existam.

Figura 2 – Exemplo da estrutura das informações armazenadas nos arquivos JSON de coordenadas

```
callback(
  {
    "LatLong": [
      {"lat": "-19.538975","lon": "-42.65000470000001"},
      {"lat": "-19.9301439","lon": "-43.97769570000003"}
    ]
  }
)
```

Fonte: elaborada pelo autor

Para converter os endereços armazenados no arquivo JSON em coordenadas com a Google Maps Geocoding API de forma gratuita, existe limitação de requisições ao servidor. Todas as solicitações de geocodificação são contabilizadas, no limite diário e no limite por segundos. Por dia era possível realizar 2.500 solicitações e por segundo era possível realizar até 50 requisições.

Por esse motivo, uma das soluções encontradas para evitar restrições na aplicação desenvolvida, por limites das requisições, foi implantar uma função de intervalo de tempo na aplicação, para controlar o tempo entre as requisições de geocodificação. Dessa forma, uma requisição só era requerida respeitando o limite mínimo de tempo entre a requisição anterior e a próxima requisição.

Outra medida tomada no projeto, na tentativa de não ultrapassar o limite de requisições, foi salvar em um arquivo JSON todas as respostas requisitadas ao servidor da Google Maps Geocoding API. Dessa forma, as conversões de endereço para coordenadas de latitude e longitude puderam ser

realizadas apenas uma vez, já que ambas são informações estáticas. Essa tática evitou requisições desnecessárias ao servidor.

4.3 Produção dos mapas temáticos

Para desenvolver os mapas temáticos, foi utilizado de forma semelhante à conversão de dados, a Google Maps JavaScript API juntamente com HTML. Foi elaborada uma ferramenta que executava a leitura de um arquivo JSON que continha os pares de coordenadas de latitude e longitude. Em seguida a ferramenta realizava requisições ao servidor da Google Maps JavaScript API, repassando as coordenadas. O servidor da API foi responsável por ler as coordenadas de latitude e longitude e apresentá-las em um mapa de calor.

Para cada atividade econômica foi elaborado um mapa de calor, fazendo uso das coordenadas do arquivo JSON da respectiva atividade econômica. Todas as coordenadas geográficas dos arquivos JSON foram utilizadas na elaboração dos mapas de calor. Dessa forma, cada mapa representava os endereços de uma atividade econômica. Ao todo, foram elaborados cinco mapas de calor.

Os mapas temáticos foram elaborados com uma imagem do Google Maps como base, onde foi possível identificar as principais ruas e avenidas da região dos endereços analisados. Para completa-lo, a segunda camada foi composta por pares de coordenadas geográficas representadas por pontos na tonalidade de azul e vermelho. A sobreposição das camadas compuseram o mapa de calor de cada atividade econômica.

Cada ponto era a representação de um conjunto de coordenadas de latitude e longitude de um endereço, onde foram representados na aplicação pela cor azul. A medida que a densidade de pontos próximos crescia, a cor dos pontos era modificada, saindo do azul claro até o vermelho, revelando áreas de maior intensidade de endereços próximos.

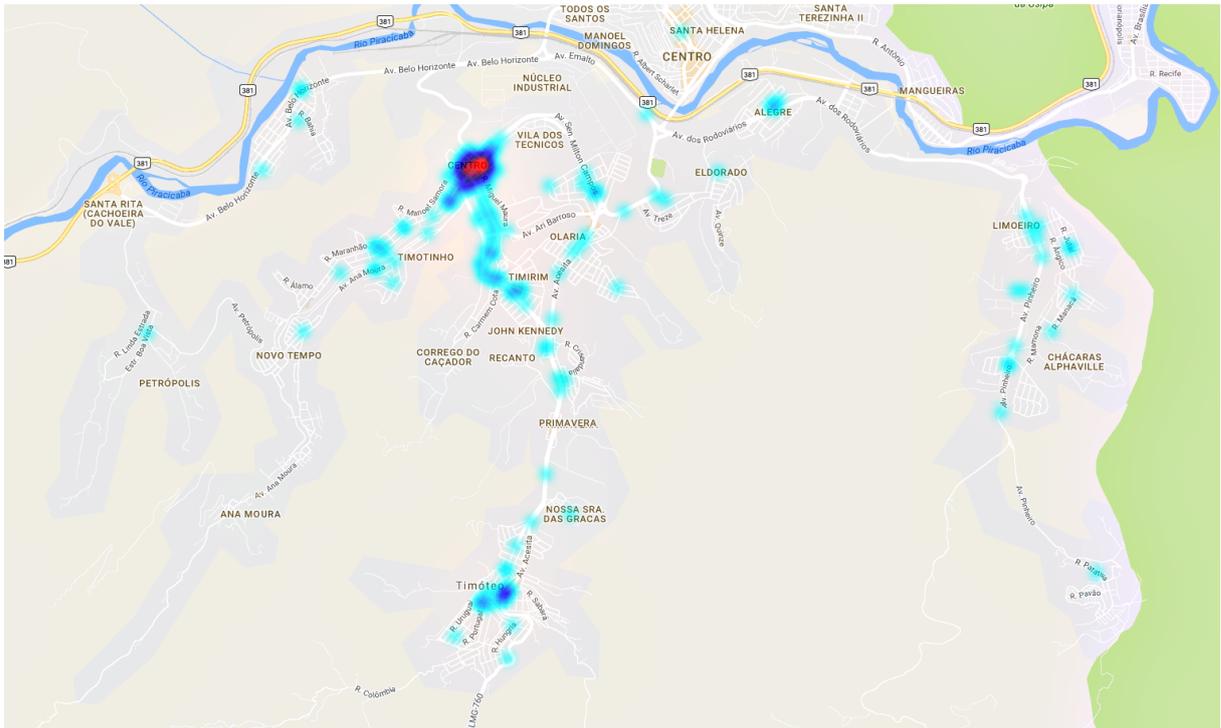
Os pares de coordenadas de latitude e longitude eram pontos aproximados que o servidor da API encontrava para cada endereço. A API identificava o local da via, o logradouro, e em seguida o número do endereço. No processo para identificar o local do endereço, a geocodificação foi realizada inteiramente pela API baseada na distância do início da rua até o número do endereço informado. Por esse motivo, endereços muito próximos podiam ser representados graficamente por pontos sobrepostos no mapa. Além disso, endereços apresentados em vias muito compridas podiam ser representados com erros de quilômetros do local real do endereço e do local onde o ponto foi representado.

A maioria dos endereços analisados neste trabalho foram da cidade de Timóteo, porém eles não se restringiram a esse território. Nas informações analisadas existiam empresas de outras cidades e outros estados. Porém, ao produzir e avaliar os mapas, os endereços foram restritos a um retângulo envolvente que incluía apenas a cidade de Timóteo.

A figura 3 apresenta o mapa de calor do comércio do amostra. Essa categoria de atividade econômica possui o maior número de empresas, com 263 cadastros no banco de dados.

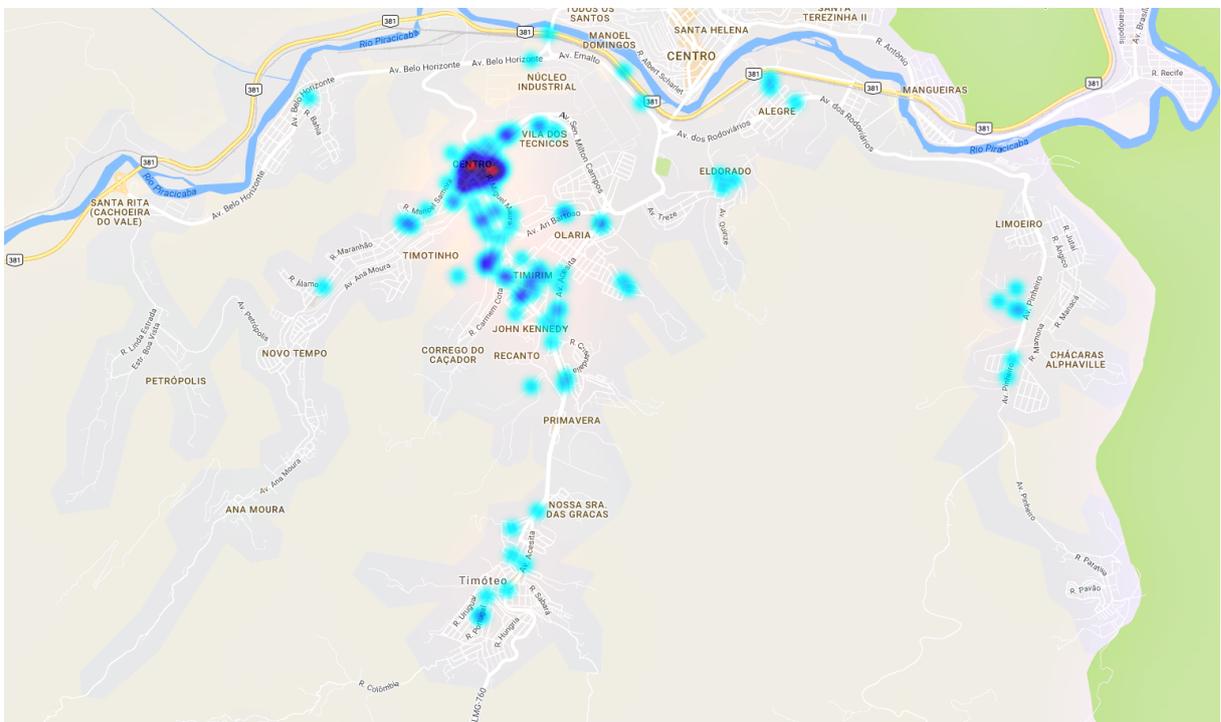
O segundo mapa é apresentado na figura 4, representando o mapa de serviço. O número de empresas nessa categoria é menor que o comércio, com 162 ocorrências no banco de dados.

Figura 3 – Mapa de calor do comércio em Timóteo



Fonte: elaborada pelo autor

Figura 4 – Mapa de calor de serviços em Timóteo

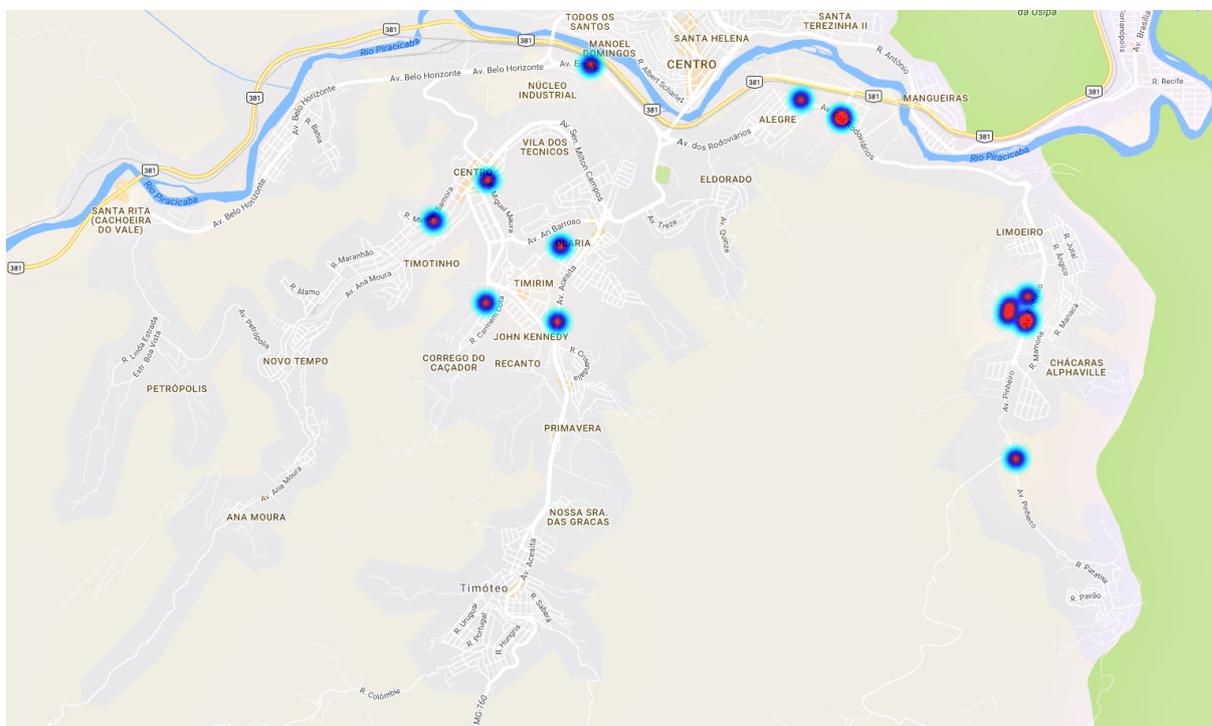


Fonte: elaborada pelo autor

A terceira imagem apresenta o mapa de calor da categoria de indústria metal mecânica. O

mapa é representado na figura 5. Essa categoria econômica possui 16 registros de empresas no banco de dados.

Figura 5 – Mapa de calor da indústria metal mecânica em Timóteo



Fonte: elaborada pelo autor

O mapa de calor da categoria indústria alimentícia é apresentado na figura 6. O banco de dados contém 12 empresas no ramo de indústria alimentícia registradas no banco de dados.

Por último, é apresentado o mapa de calor da categoria de indústria química na figura 7. Essa categoria contém o menor número de empresas no banco de dados, apenas sete.

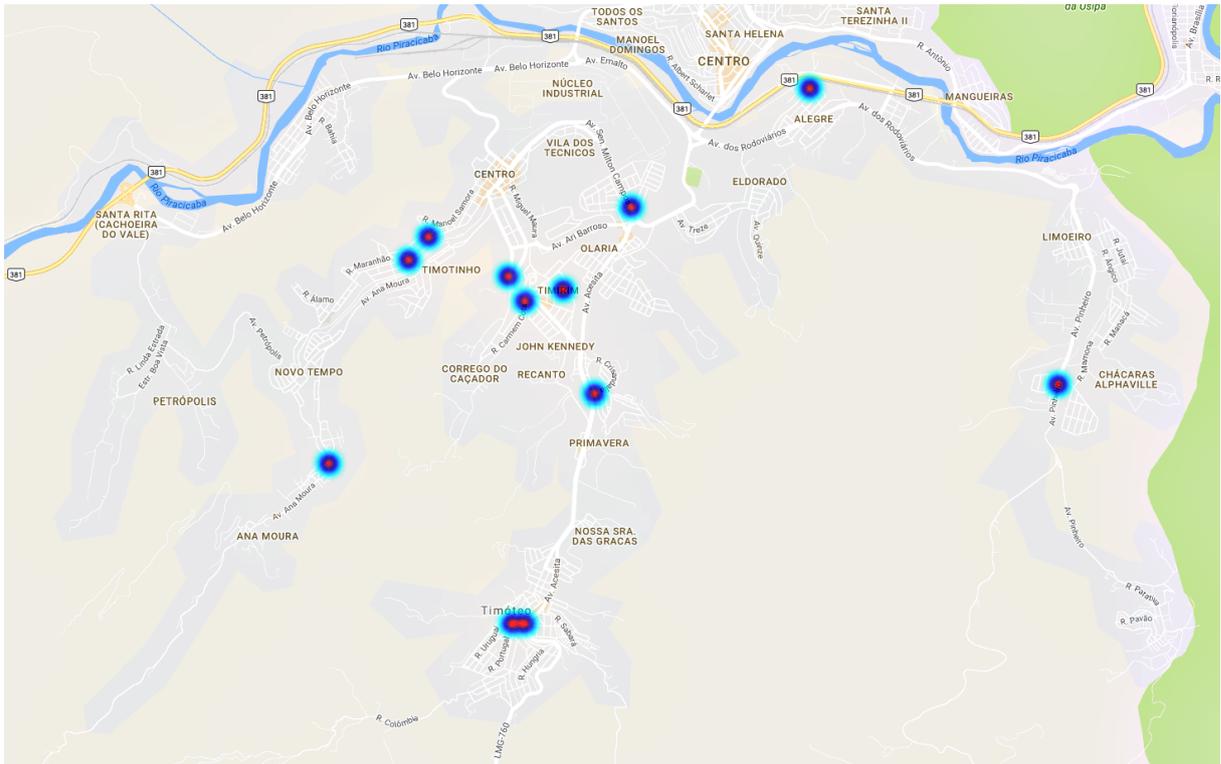
4.4 Avaliação dos mapas

Após o desenvolvimento dos mapas, as informações do banco de dados foram analisadas. Essa análise foi baseada nas informações apresentadas nos mapas de calor e em um questionário respondido por especialistas nas empresas cadastradas na ACIATI.

Para analisar as informações, os mapas foram apresentados a oito colaboradores da associação. Essas pessoas foram selecionadas para analisarem os mapas pois utilizam os dados diariamente em suas atividades profissionais, além de conhecerem as atividades econômicas.

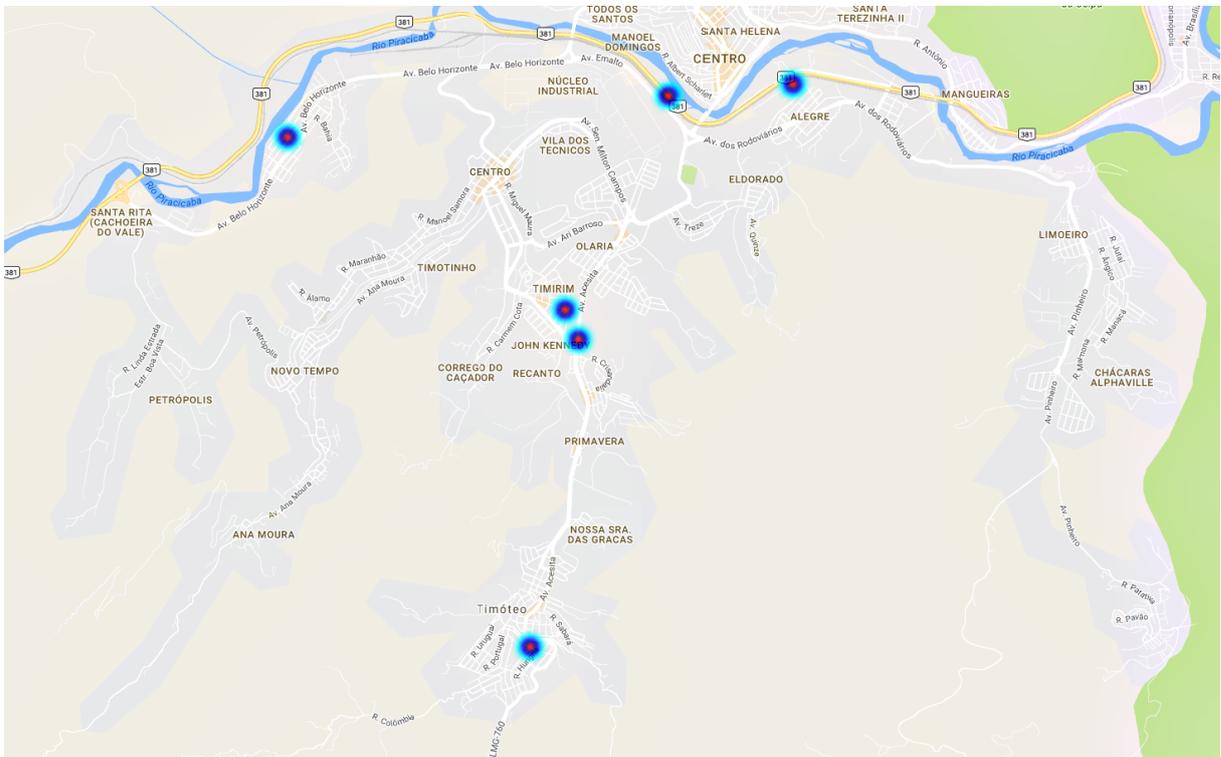
Foi elaborado um questionário com cinco perguntas sobre o tipo de atividade econômica que cada mapa de calor representa. Cada questão apresentava um mapa de calor, sem a legenda da categoria da atividade econômica, e ao respondente foi pedido que selecionasse a resposta que continha a opção que ele acreditava apresentar a atividade econômica do mapa correspondente. Em cada questão, foi apresentado um mapa de calor com as seguintes opções:

Figura 6 – Mapa de calor da indústria alimentícia em Timóteo



Fonte: elaborada pelo autor

Figura 7 – Mapa de calor da indústria química em Timóteo



Fonte: elaborada pelo autor

1. comércio,
2. serviço,
3. indústria metal mecânica
4. indústria de alimentos,
5. indústria química.

O questionário foi implementado por meio do programa Web Google Forms. Após a criação do formulário, o caminho para respondê-lo foi enviado por e-mail aos colaboradores que participaram da pesquisa e que responderam as perguntas. Ao responder todas as perguntas do formulário, o colaborador conseguia visualizar as respostas com as legendas dos mapas e a correta categoria da atividade econômica. As respostas foram anônimas, sem a identificação do respondente e cada respondente foi orientado a responder uma única vez.

Foram realizados alguns filtros no questionário. Um dos filtros implantado foi para limitação de respostas, onde cada questão somente permitia assinalar uma alternativa como resposta. Outro filtro implantado foi para garantir que cada atividade econômica somente podia ser assinalada uma única vez. Um terceiro filtro foi a obrigatoriedade de responder todas as questões.

As tabelas de 2 à 6 apresentam as respostas dos colaboradores da pesquisa. Cada tabela apresenta as respostas de um mapa apresentado no formulário.

As respostas do mapa de comércio são apresentadas na tabela 2. É possível identificar que 50% das respostas foram de acordo com a atividade econômica representada no mapa.

Tabela 2 – Respostas mapa comércio

Setor	Número de respostas	Porcentagem
Comércio	4	50
Serviço	2	25
Indústria metal mecânica	1	12,5
Indústria alimentícia	1	12,5
Indústria química	0	0
Total	8	100

Fonte: Elaborada pelo autor.

A próxima é a tabela 2, onde são apresentadas as respostas do mapa de serviço. No mapa de serviço, 37,5% dos respondentes assinalaram que o mapa correspondia ao setor de serviço.

A terceira é a tabela 4, representando as respostas do mapa de indústria metal-mecânica. É possível identificar que 37,5% dos colaboradores identificaram que o mapa representava o setor de indústria metal-mecânica.

A tabela 5 apresenta todas as respostas do mapa do setor de indústria alimentícia, onde 37,5% das respostas identificaram que o mapa pertencia a essa categoria.

Tabela 3 – Respostas mapa serviço

Setor	Número de respostas	Porcentagem
Comércio	5	62,5
Serviço	3	37,5
Indústria metal mecânica	0	0
Indústria alimentícia	0	0
Indústria química	0	0
Total	8	100

Fonte: Elaborada pelo autor.

Tabela 4 – Respostas mapa indústria metal-mecânica

Setor	Número de respostas	Porcentagem
Comércio	4	50
Serviço	0	0
Indústria metal mecânica	3	37,5
Indústria alimentícia	0	0
Indústria química	1	12,5
Total	8	100

Fonte: Elaborada pelo autor.

Tabela 5 – Respostas mapa indústria alimentícia

Setor	Número de respostas	Porcentagem
Comércio	2	25
Serviço	1	12,5
Indústria metal mecânica	2	25
Indústria alimentícia	3	37,5
Indústria química	0	0
Total	8	100

Fonte: Elaborada pelo autor.

Por último, é apresentado a tabela 6, representado as respostas do mapa de indústria química. Nesse mapa, 50 % das respostas do questionário apresentaram que o mapa correspondia à categoria de indústria química.

Tabela 6 – Respostas mapa indústria química

Setor	Número de respostas	Porcentagem
Comércio	2	25
Serviço	1	12,5
Indústria metal mecânica	1	12,5
Indústria alimentícia	0	0
Indústria química	4	50
Total	8	100

Fonte: Elaborada pelo autor.

5 Conclusão

Este trabalho objetivou avaliar a qualidade e a precisão dos dados geográficos presentes em um banco de dados corporativo. Para isso, foi realizada a coleta das informações do banco de dados com os endereços de empresas foram elaborados mapas de calor que representavam a distribuição geográfica das empresas e suas atividades econômicas. Os mapas de calor foram analisados por especialistas e foi possível identificar as contribuições e limitações do método utilizado.

5.1 Contribuições

Neste trabalho, foi utilizada uma metodologia de geocodificação de endereços a partir de um banco de dados convencional. Porém, ao invés de uma visualização tabular de endereços, foi experimentada uma visualização geográfica das atividades econômicas, identificando os pontos de maior e menor aglomeração de atividades econômicas. A utilização de poucos endereços para geocodificação permitiu analisar os endereços geocodificados em tempo hábil para a pesquisa.

Adicionalmente, este trabalho fez uso de dados de uma associação comercial real, o que permitiu construir um estudo de caso válido. Também, isso permitiu representar todos os endereços da cidade de Timóteo, onde a associação se encontra, que foram geocodificados em um mapa. A análise baseada em uma única associação e em uma área limitada também possibilitou mapas com números de informações reduzido. Isso facilitou a identificação das informações pelo leitor do mapa, um usuário real de informação corporativa da associação, e facilitou a compreensão dos resultados obtidos.

Enquanto parte dos estudos presentes na literatura acontecem em grandes centros urbanos, este estudo aconteceu em município de pequeno porte, com população inferior a 100 mil habitantes. Isso produz impactos nos dados geográficos e na qualidade da visualização geográfica. Nesta pesquisa, por exemplo, os endereços se concentraram, na maior parte das vezes, em ruas e avenidas pequenas, comparadas ao tamanho de vias públicas de metrópoles. Isso proporcionou a redução do erro da geocodificação dos endereços analisados para alguns metros em sua maioria. Isso parece favorecer uma representação geográfica mais próxima da realidade.

Finalmente, este trabalho permitiu utilizar as informações de um banco de dados na geocodificação de empresas, sem repetição. Garantidamente, cada empresa foi representada uma única vez no mapa. Isso foi possível mesmo sem um identificador único para as empresas, como o número do Cadastro Nacional de Pessoa Jurídica (CNPJ), visto que cada empresa está representada apenas uma única vez no banco de dados onde a pesquisa foi baseada. Isso parece viável para regiões pequenas ou com elevada dispersão geográfica de empresas. Por outro lado, possivelmente esse fenômeno não acontece em regiões densamente povoadas ou em centros de serviços ou centros de compras.

5.2 Limitações

A principal limitação presente neste trabalho foi o desenvolvimento e análise baseados em um número muito reduzido de empresas. A maior parte das empresas estão localizadas na mesma cidade, o que restringe os tipos de serviços oferecidos. Logo, a representação desse grupo de empresas está longe de representar as empresas do mundo, do país ou mesmo do estado.

Uma segunda e importante limitação foi a utilização de informações de uma única associação. A associação não possui o cadastro de todas as empresas da região de interesse. Por outro lado, bancos de dados de outras entidades, tais como prefeitura ou corpo de bombeiros, poderiam representar melhor a maior parte das empresas da região e contribuiriam para maior precisão geográfica e econômica.

Outra limitação ocorreu com a representação gráfica dos endereços. Os mapas ficaram limitados a um retângulo envolvente sobre a cidade de Timóteo. No entanto, havia empresas com o registro de endereços em cidades vizinhas e até de outros estados. Porém, não foi viável apresentar todos os endereços do banco de dados, visto que o campo territorial a ser analisado seria muito extenso. Esse fenômeno dificultaria muito a interpretação dos mapas por usuários, especialistas ou não. Por esse motivo, os mapas não conseguiriam apresentar adequadamente as informações sobre onde ocorre maior concentração de empresas.

A quarta limitação identificada correspondeu à própria extensão dos logradouros do estudo de caso. A cidade onde os endereços estão concentrados tem entre 80 mil e 100 mil habitantes, a cidade não possui ruas e avenidas muito grandes, e isso garantiu que os erros da geocodificação se limitassem a um intervalo de metros. No entanto, alguns endereços analisados estavam localizados em rodovias e foi possível observar que a margem de erro na geocodificação foi muito maior. A precisão da geocodificação de empresas e atividades econômicas, portanto, parece representar um gargalo importante para mapas temáticos como esse.

Outra limitação do trabalho foi a divisão das empresas em tão poucas categorias de atividades econômicas. Apenas cinco categorias foram utilizadas na pesquisa. Essas categorias não parecem refletir todas as empresas do banco de dados e por consequência também não parecem representar adequadamente as várias atividades econômicas que podem acontecer, em áreas urbanas ou rurais. Embora tenha servido ao propósito exploratório deste trabalho, outras tentativas de classificação são necessárias.

A sexta limitação mais importante identificada neste trabalho foi o uso de endereços exclusivamente urbanos. O banco de dados analisado contém informações de empresas com endereços dentro do limite urbano da cidade, onde as ruas e avenidas são identificadas pela aplicação de geocodificação. Dessa forma, não foi possível observar como a aplicação se comportaria para geocodificar endereços rurais ou em localidades com baixa organização. Nem todos os endereços rurais foram reconhecidos pela aplicação utilizada.

De forma geral, as empresas analisadas não são de grande porte. Nas informações disponibilizadas, havia apenas um único registro de endereço para cada empresa. Isso não permitiu analisar empresas que apresentam mais de um endereço, como é o caso de empresas que possuem endereços diferentes para fábrica e escritório. No banco de dados analisado, possui empresas que possuem

mais de um registro, onde cada registro está cadastrado com um endereço diferente, porém com o mesmo nome fantasia. Seguindo o mesmo princípio, foi identificado que existem empresas onde fisicamente possui mais de um endereço na cidade, porém no banco de dados é identificado apenas uma representação.

Outro fator limitante da pesquisa foi a ausência de CNPJ ou algum identificador único e invariável das empresas. Essa informação é fundamental para utilização de dados de diferentes bancos de dados, com fontes heterogêneas. Esse identificador deve ser utilizado para garantir a unicidade de cada elemento, especialmente quando um grande número de empresas for mobilizado.

Por último, foi utilizada a Google Maps JavaScript API. Essa aplicação é fechada e paga, o que restringe trabalhos do tipo aos serviços de terceiros. Caso a aplicação utilizada passe por mudanças, ou fique indisponível, não seria possível a utilização da aplicação utilizando os mesmos métodos.

5.3 Trabalhos futuros

Como resultado de um primeiro estudo exploratório, algumas ideias de trabalhos futuros parecem estar disponíveis.

Como trabalho futuro, é possível elaborar pesquisas com número maior de empresas. Também, parece adequado utilizar outros bancos de dados de diferentes origens, como, por exemplo, corpo de bombeiros, prefeituras e associações regionais. Essa diferença de bancos exigiria tratamento das informações para geocodificação. É preciso identificar as diferenças de informações dos bancos de dados, prováveis falhas e conseqüentemente pontos para correção. Também é necessário conhecer como cruzar os dados das empresas e verificar se os mesmos são armazenados com as mesmas informações nos diferentes bancos de dados, favorecendo a unicidade e a correção de endereços em caso de desatualização.

Uma segunda sugestão de trabalho futuro é a pesquisa sobre o processamento de banco de dados de empresas que possuam mais de um endereço, identificando melhorias para armazenamento e tratamento das informações. Há ainda uma grande resistência para a adoção em larga escala de bancos de dados geográficos por recém-formados em cursos de computação, o que limita a existência de softwares e bancos geográficos. Estudos sobre a engenharia de software geográfico e sobre a educação em computação para a produção de sistemas de informação geográficos são necessários.

Outra sugestão de trabalho futuro é realizar pesquisas com atividades econômicas dispersas geograficamente. Isso pode ampliar as diferenças de atividades econômicas, visto que as naturezas das atividades estão ligadas diretamente às necessidades e ou facilidades de uma região.

Alterações na amostra e criação de novas amostras são outros caminhos naturais de trabalhos futuros. É preciso realizar novas pesquisas em cidades maiores, com ruas com extensões maiores para verificar a precisão do geocodificador.

Adicionalmente, amostras que incluem novos exemplares de atividades econômicas e classificações de atividades são necessárias.

Para reduzir a dependência de serviços proprietários e fechados, é preciso estudar como

serviços geoespaciais padronizados podem ser adotados para garantir a continuidade de serviços, com maior ou menor precisão, mesmo que um fornecedor deixe de existir ao longo do tempo.

Por último, o usuário e sua capacidade de interpretar e manipular dados geográficos, mapas e sistemas de informação comunitários precisam receber atenção em trabalhos futuros. É possível que parte da responsabilidade pela baixa adoção de sistemas de informação geográficos na administração pública seja pela baixa aptidão dos usuários, especialmente dos tomadores de decisão das entidades que mais dependem deste tipo de informação.

Referências

- ARCHELA, R. S.; THÉRY, H. Orientação metodológica para construção e leitura de mapas temáticos. *Revista Franco-Brasileira de Geografia*, v. 3, 2008. Citado nas páginas 10 e 16.
- BURROUGH, P. A. Principles of geographical information systems for land resources assessment. Taylor & Francis, 1986. Citado nas páginas 15 e 17.
- KURKDJIAN, M. d. L. N. d. O.; PEREIRA, M. N. *Infogeo*. Citado na página 10.
- MARTINELLI, M. *Mapas da geografia e cartografia temática*. [S.l.]: Editora Contexto, 2003. Citado na página 16.
- PEREIRA, G. C.; SILVA, B.-C. N. Geoprocessamento e urbanismo. UNESP; AGETEO, 2001. Citado na página 17.
- REZENDE, D. A.; FREY, K.; BETINI, R. C. Governança e democracia eletrônica na gestão urbana. *Anais do Seminário Internacional em Gestão Urbana*, v. 1, 2003. Citado na página 10.
- TSORLINI, A. et al. Generation of thematic maps with wizards based on cartographic principles: The case study of ocad thematic mapper. In: *Proceedings of the 27 th International Cartographic Conference, Rio de Janeiro, Brazil*. [S.l.: s.n.], 2015. Citado na página 15.

Anexos

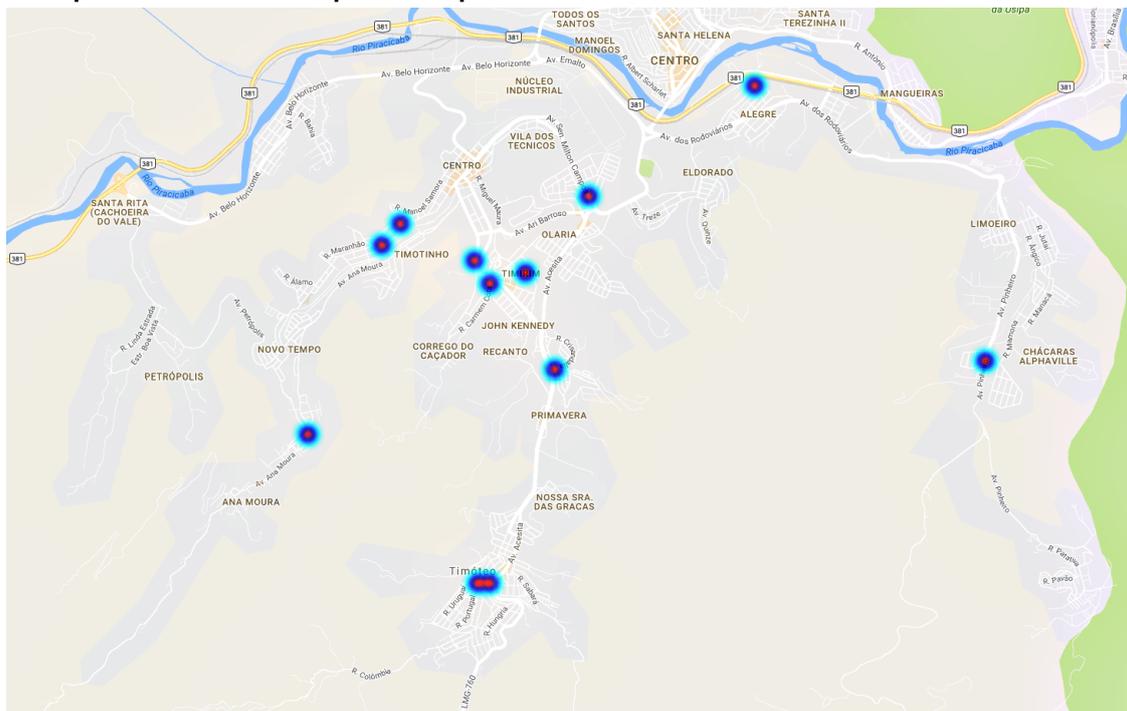
Visualização de informação geográfica

Este formulário tem como objetivo avaliar as informações geográficas das atividades econômicas de Timóteo MG

Este levantamento é para fins acadêmicos.

***Obrigatório**

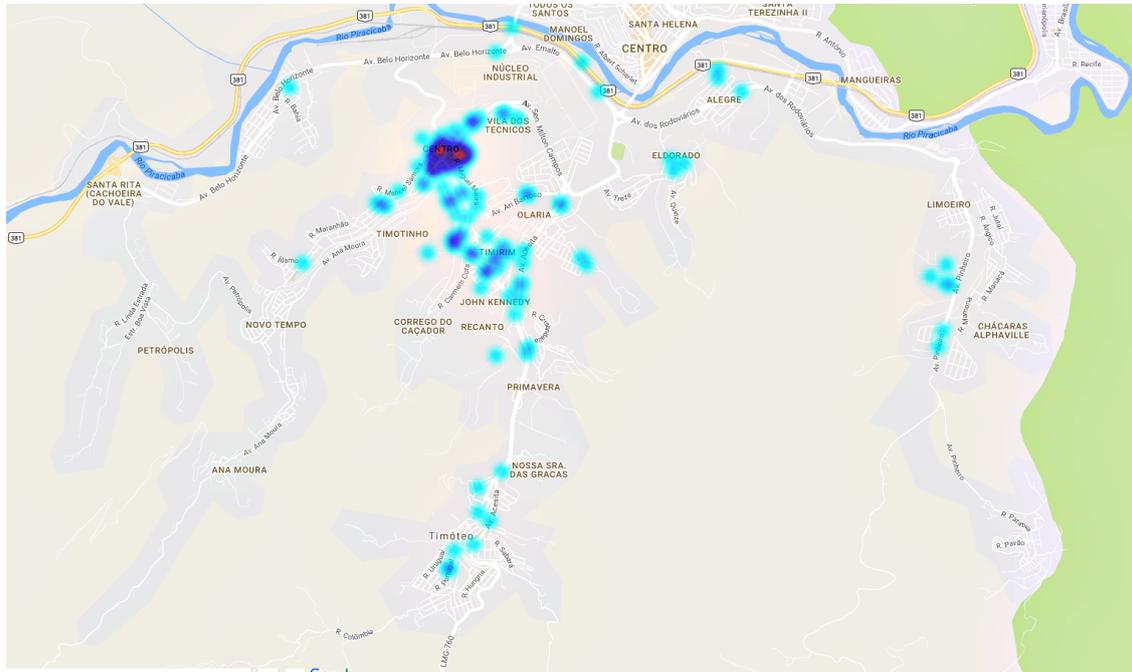
1. O mapa de calor abaixo representa qual atividade econômica na cidade de Timóteo? *



Marcar apenas uma oval.

- Comércio
- Serviço
- Indústria Metal Mecânica
- Indústria de Alimentos
- Indústria Química

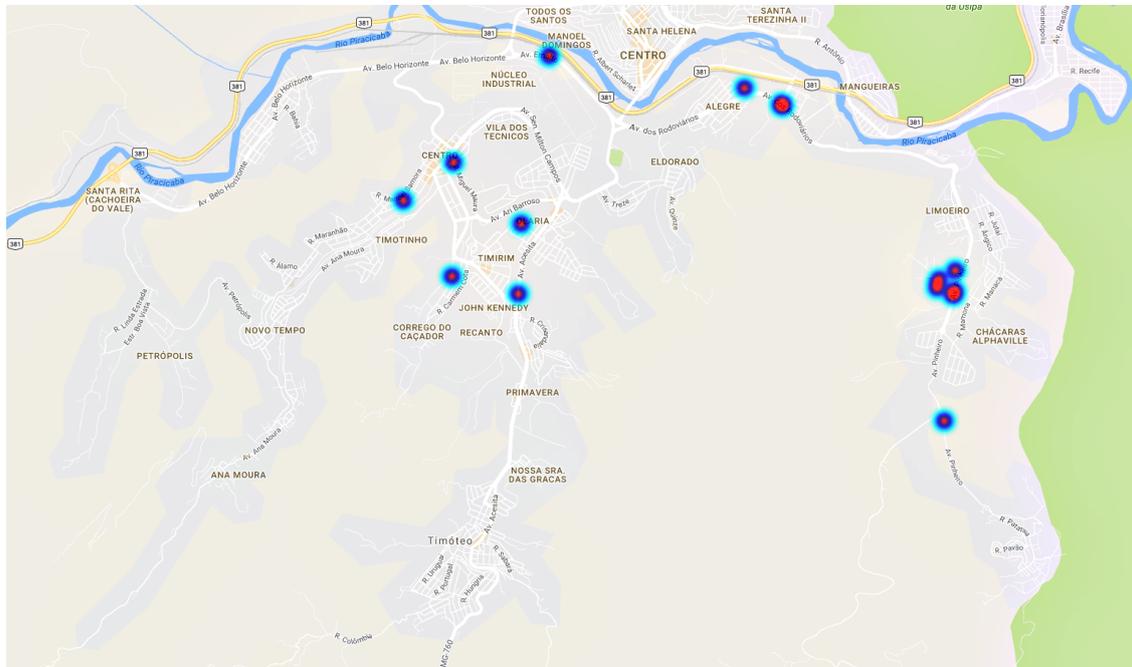
2. O mapa de calor abaixo representa qual atividade econômica na cidade de Timóteo? *



Marcar apenas uma oval.

- Comércio
 Serviço
 Indústria Metal Mecânica
 Indústria de Alimentos
 Indústria Química

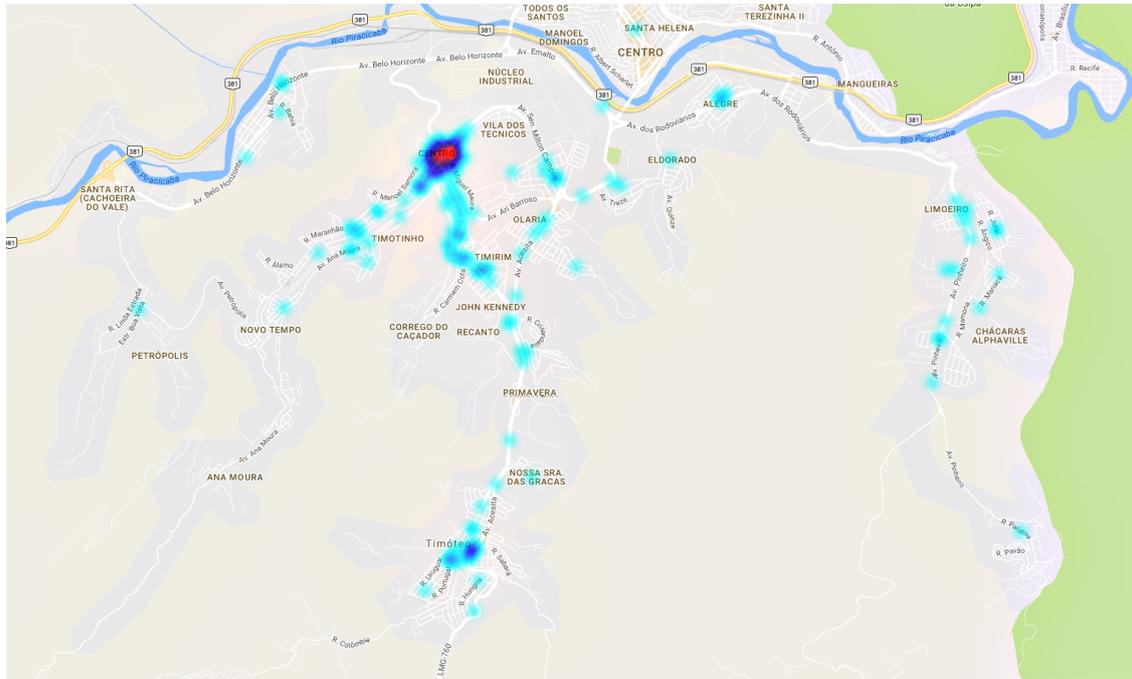
3. O mapa de calor abaixo representa qual atividade econômica na cidade de Timóteo? *



Marcar apenas uma oval.

- Comércio
 Serviço
 Indústria Metal Mecânica
 Indústria de Alimentos
 Indústria Química

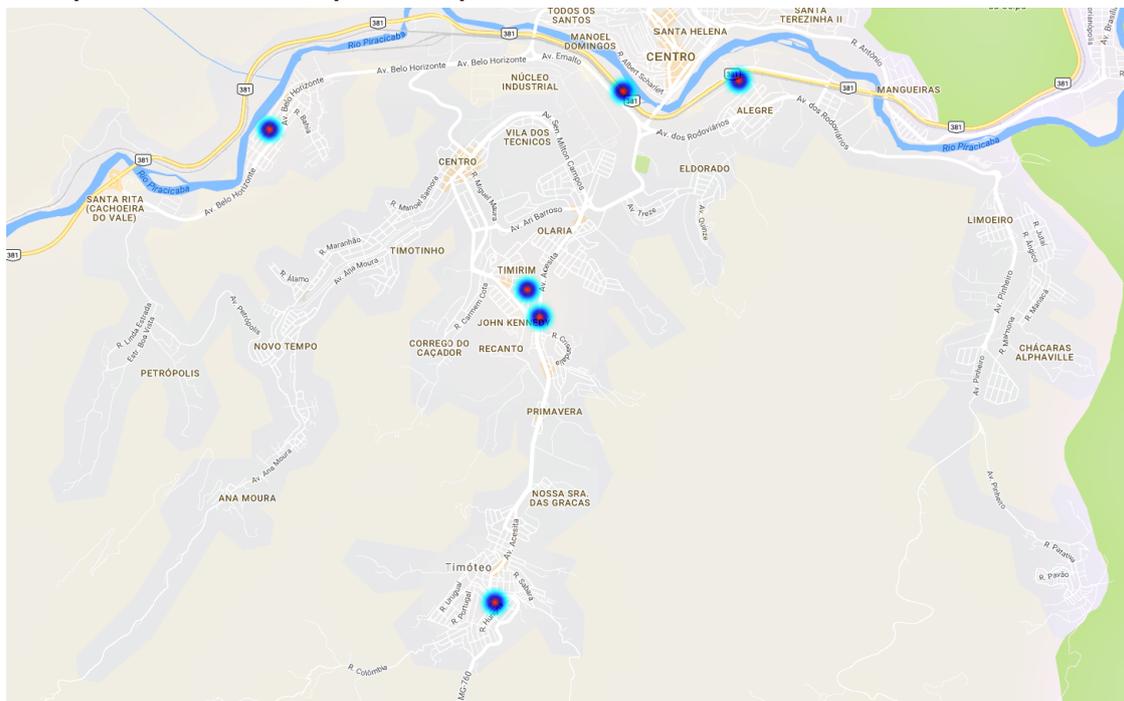
4. O mapa de calor abaixo representa qual atividade econômica na cidade de Timóteo? *



Marcar apenas uma oval.

- Comércio
- Serviço
- Indústria Metal Mecânica
- Indústria de Alimentos
- Indústria Química

5. O mapa de calor abaixo representa qual atividade econômica na cidade de Timóteo? *



Marcar apenas uma oval.

- Comércio
- Serviço
- Indústria Metal Mecânica
- Indústria de Alimentos
- Indústria Química