

Análise de dados exploratória – Consumo de dados dos municípios do Brasil usando diferentes métodos

Entrega 2

Luiz Henrique A. P. Bittencourt - 321111628

Centro Universitário UNA – Rua dos Aimorés 1451, Lourdes-Belo Horizonte

Departamento de Tecnologia – Análise e desenvolvimento de sistemas

Resumo: O presente trabalho descreve e detalha o consumo de dados referente aos municípios do Brasil, por meio de três diferentes meios, dois menos conhecidos e um utilizado diariamente por muitos brasileiros, porém não da forma apresentada no presente estudo.

1.INTRODUÇÃO:

No Brasil, uma das principais fontes de dados sobre os municípios é o IBGE, onde ele por meio de pesquisas “porta em porta”, coleta dados sobre os municípios e seus habitantes. Por meio desses dados o IBGE é capaz de trazer diferentes indicadores que trazem um relatório completo sobre todos os estados e regiões do país, dados esses que serão consumidos no presente trabalho de três formas diferentes: Consultas ao site do IBGE, Arquivos de extensão ShapeFile e o controverso chatGPT.

2.OBJETIVOS:

O presente trabalho tem como objetivo testar e colocar frente a frente dados de diferentes fontes, desse modo testando ao mesmo tempo a confiabilidade do chatGPT colocando-o frente a frente com os dados do IBGE, enquanto os mesmos são comparados entre si.

Os dados utilizados para essa comparação são os dos municípios brasileiros, sendo filtrados por algumas queries/filtros simples que traziam os dados também de forma simples.

3.METODOLOGIA:

Como dito anteriormente o presente trabalho colocou frente a frente três fontes de dados diferentes, duas da mesma fonte e uma controversa, sendo elas respectivamente o site do IBGE, arquivos de extensão ShapeFile do próprio IBGE e consultas feitas na IA chatGPT, da OpenAI.

Para os dados do IBGE, foram utilizados dados do censo, o censo é realizado há mais de 150 anos, tendo seu 150º aniversário sendo comemorado em 2022 segundo o Senado Brasileiro, censo esse que trouxe um dado abstrato, porém assustador para nós brasileiros atuais, onde 1,5 milhões de escravos foram contabilizados

Os dados do censo do IBGE além de serem completos e trazerem um retrato da população, tem a vantagem que por serem dados do governo, serem públicos e terem seu acesso liberado a todos os interessados. Dentre esses dados existem não só os presentes no próprio site de consultas do IBGE, mas também arquivos de extensão ShapeFile, que são arquivos utilizados em trabalhos e consultas de geoprocessamento. Para consumir esses dados dos arquivos ShapeFile, foi utilizado o software livre chamado QGIS, onde é possível abrir o arquivo e fazer consultas tanto por meio de filtros como simplesmente clicando nos polígonos presentes na tela. Esses arquivos, contam não só com os polígonos que representam os estados e os municípios mas também contam com diferentes dados sobre os mesmos.

Até então, foram introduzidas duas das três plataformas utilizadas na confecção desse trabalho, sendo eles os arquivos de extensão ShapeFile e o site para consultas do IBGE, porém esse trabalho vai ter uma terceira fonte de dados, o controverso chatGPT. Apesar de ser mal visto e muitas vezes errar, o chatGPT tal qual inteligências artificiais variadas vem ganhando cada vez mais espaço no mundo moderno. Porém, apesar de

todos os pontos contrários a utilização do mesmo, ele foi utilizado de uma forma diferente no presente trabalho, ao invés de simplesmente perguntar dados que teriam a resposta entregue normalmente e em formato de texto comprido, foi utilizado um prompt para que a IA funcionasse como um banco de dados dos municípios do Brasil, e que só responderia por meio de queries de SQL e retornava os dados pedidos.

Para a realização do trabalho, as três fontes de dados foram filtradas pelas seguintes questões:

1. Município com a maior extensão de área.
2. Quantos municípios tem acima de 500.000 habitantes.
3. Quantos municípios tem acima de 1 milhão de habitantes.
4. Quantos municípios no estado de Minas Gerais.
5. Quantos municípios na região Sudeste do Brasil.
6. E número de municípios totais.

Ao iniciar as consultas foi necessário fazer alguns preparativos, sendo eles o download do arquivo dos municípios do Brasil disponibilizado no site do IBGE e dizer ao chatGPT a forma com que ele teria de se comportar daquele ponto em diante.

Para ditar ao chatGPT como ele deveria se comportar foi produzido o seguinte prompt, onde por meio desse prompt foi informado quais os dados que gostaríamos que ele trouxesse, sendo esses dados: Nome do Município, número de habitantes, área em km² e data de fundação; a data de fundação por fim não foi utilizada para realizar as consultas por conta das outras fontes de dados não terem acesso facilitado a essa informação e ser uma informação volátil. Abaixo, a imagem com o prompt utilizado e a resposta do chatGPT:

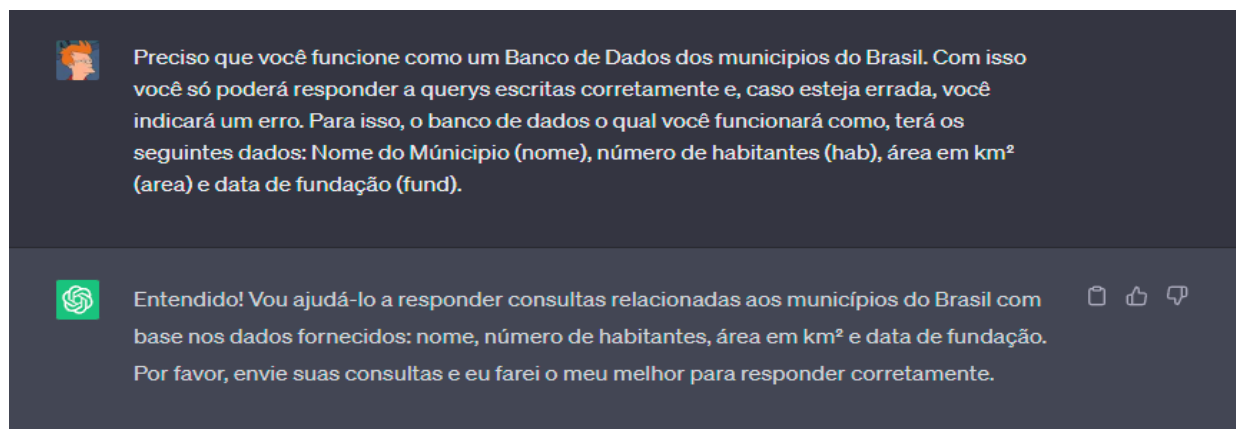


Figura 1 - Prompt chatGPT

As consultas por meio do software QGIS, apesar de uma interface user friendly, é de certa forma complexo, porém com a prática é um software fácil de ser utilizado. Abaixo uma captura de tela da interface do software com o arquivo dos municípios brasileiros já carregada.

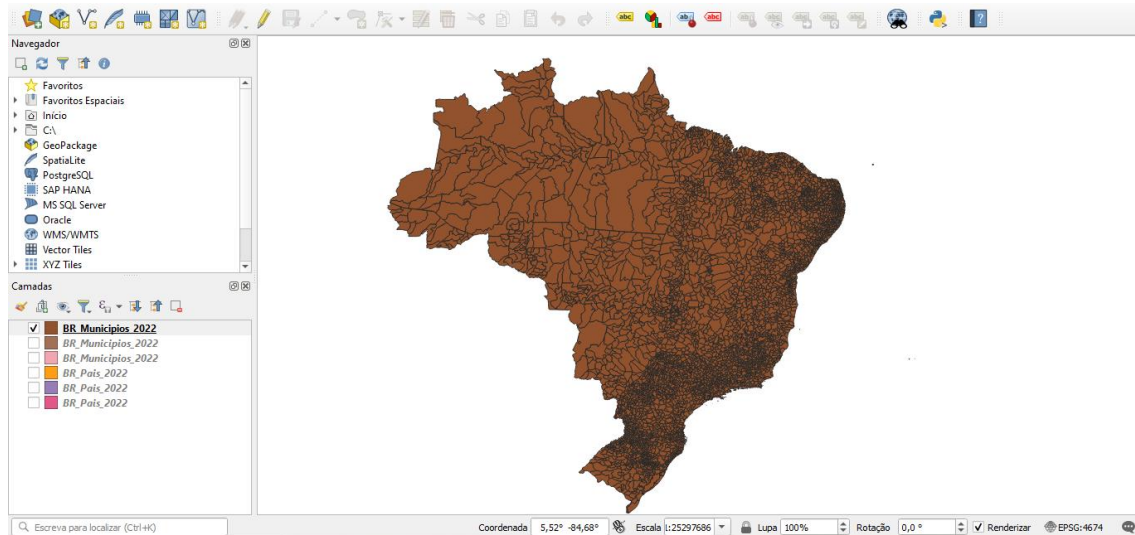


Figura 2 - shapeFile Municípios do Brasil

Dentro desse arquivo, se percebe vários polígonos, esses polígonos representam cada um dos municípios brasileiros, por isso percebe-se grande densidade de linhas nas regiões sudeste, sul e nordeste. A partir daí, é possível abrir uma tabela de dados onde se apresenta todos os dados do arquivo carregado, como mostra a imagem abaixo:

QGIS: BR_Municipios_2022 — Total de feições: 5572, Filtrado: 5572, Selecionado: 0

	CD_MUN	NM_MUN	SIGLA_UF	AREA_KM2
1	1100015	Alta Floresta D'...	RO	7067,127
2	1100023	Ariquemes	RO	4426,571
3	1100031	Cabixi	RO	1314,352
4	1100049	Cacoal	RO	3793,000
5	1100056	Cerejeiras	RO	2783,300
6	1100064	Colorado do O...	RO	1451,060
7	1100072	Corumbiara	RO	3060,321
8	1100080	Costa Marques	RO	4987,177
9	1100098	Espigão D'Oeste	RO	4518,038
10	1100106	Guajará-Mirim	RO	24856,877
11	1100114	Jaru	RO	2944,128
12	1100122	Ji-Paraná	RO	6896,649
13	1100130	Machadinho D'...	RO	8509,270
14	1100148	Nova Brasilândia	RO	1702,000

Mostrar todos os feições

Figura 3 - Interface de dados

Porém, além dessa consulta geral, é possível selecionar apenas um polígono e consumir as informações dele, como o exemplo abaixo da cidade de Belo Horizonte:

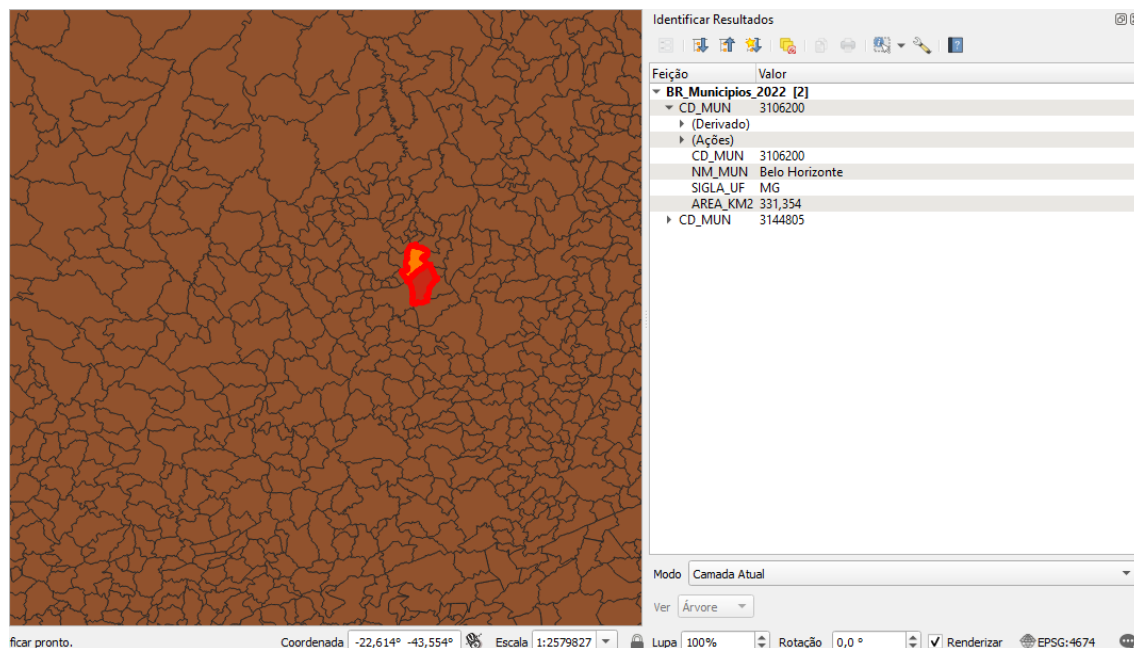


Figura 4 - Dados específicos

Agora, para a terceira fonte de dados, o site do próprio IBGE é bem fácil de ser utilizado e as consultas são simplificadas, não só de dados fáceis como os utilizados na presente análise, mas também dados mais nichado e específicos, como de recursos hídricos e até cobertura vacinal de doenças quais a mídia não presta tanta atenção.



Figura 5 - Interface do Portal do IBGE

4. ANÁLISE E CONSOLIDAÇÃO:

4.1. APRESENTAÇÃO DOS RESULTADOS:

No presente tópico, serão apresentados os resultados coletados nas três fontes utilizadas e serão colocadas lado a lado para conferir não só os resultados mas também a veracidade e confiabilidade dos mesmos. Os resultados serão organizados seguindo as perguntas apresentadas na metodologia:

1. Município com a maior extensão de área:

A primeira pergunta realizada às fontes de dados foi o município de maior extensão territorial, a resposta esperada dessa pergunta, seria Altamira no estado do Pará, sendo essas as respostas obtidas:

- **chatGPT:** Para obter a resposta por meio dessa fonte, foi feita uma query e obtivemos a seguinte resposta:

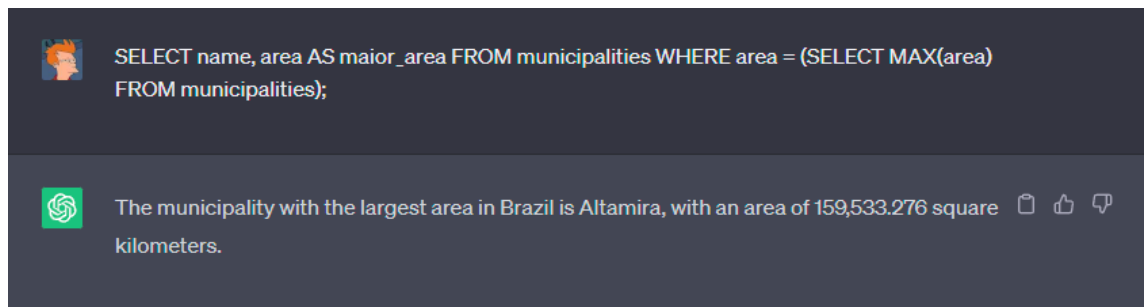


Figura 6 - Resposta chatGPT

- **ShapeFile:**

Para realizar consultas e filtrações no arquivo do ShapeFile, utilizamos o filtro da tabela de atributos do mesmo, a partir dela é possível filtrar as feições (polígonos dos municípios) que condizem com a filtração realizada, para filtrar o município de maior área em KM², foi utilizada a seguinte função: `AREA_KM2 = maximum(AREA_KM2)`. Essa função filtrará dentre todas as entradas de dados do arquivo e retornará a que tem o maior valor no campo informado.

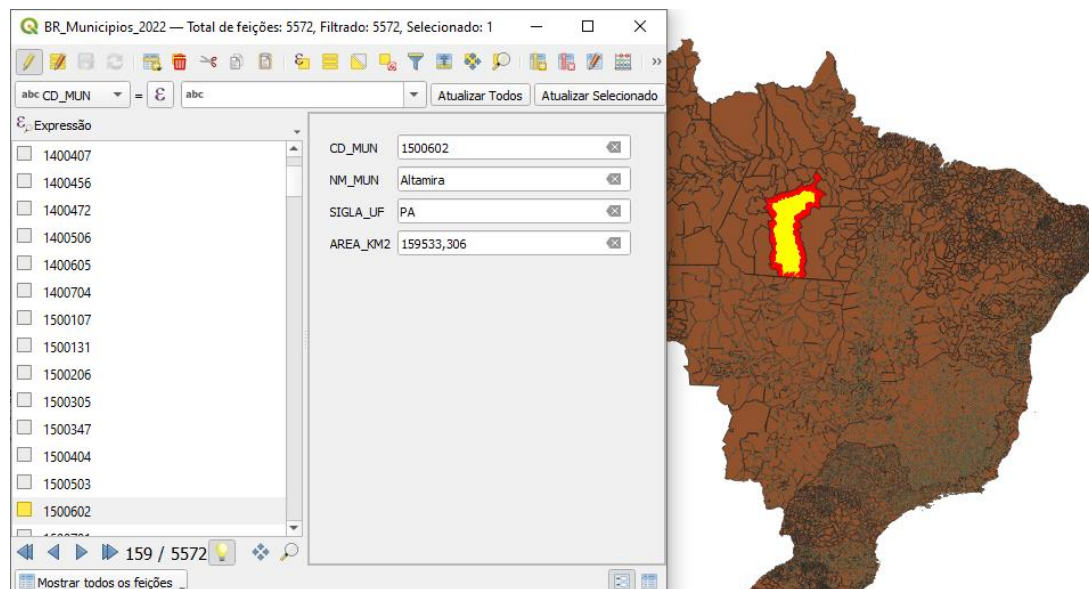


Figura 7 - Resposta QGIS

Um ponto interessante sobre o QGIS é que por ser um software de geoprocessamento ele além de informar o município referente, também o seleciona dentre as demais feições presentes no shapefile. E quanto ao resultado, foi informado o município correto e com a área bem próxima a informada pelo chatGPT.

- **Portal IBGE:**

Indo agora para a terceira fonte de dados, temos o portal do IBGE, onde nele realizamos a pesquisa diretamente pelo município de Altamira e obtivemos o seguinte resultado.

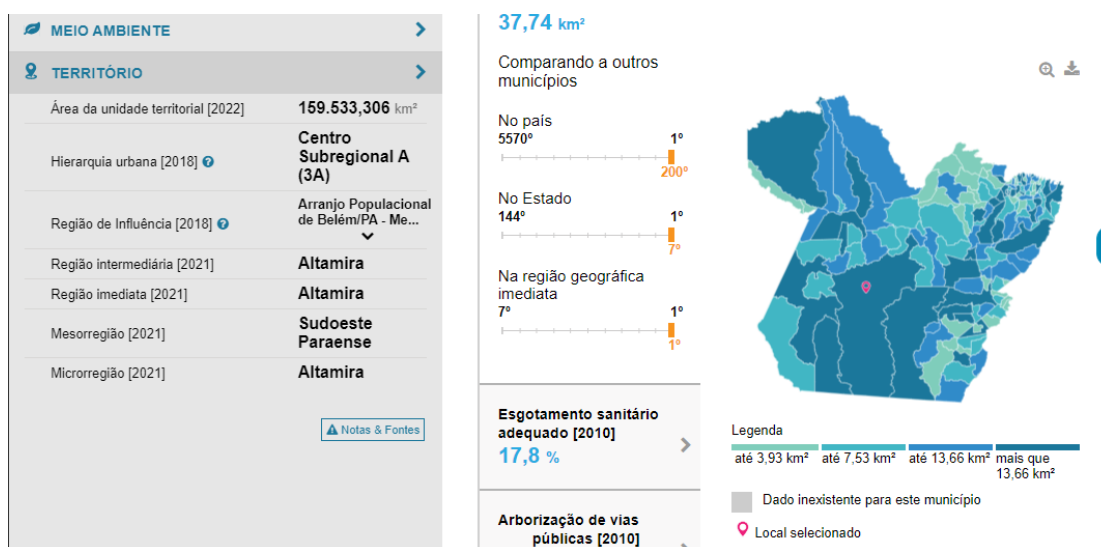


Figura 8 - Informação no portal IBGE

Também de acordo com os dados apresentados nas outras fontes utilizadas.

2. Municípios com mais de 500 mil habitantes:

Para apresentar os resultados das presente e consequentes perguntas, vamos apenas apresentar os resultados e métodos que diferem, uma vez que o conteúdo se tornaria redundante.

- **chatGPT:**

Resultado obtido a partir da query no chatGPT:

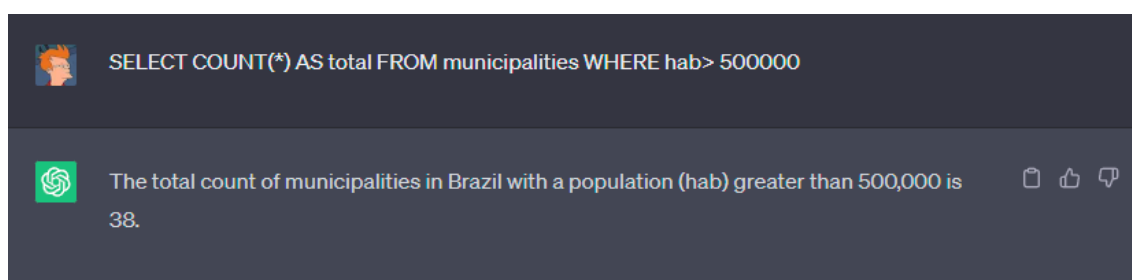


Figura 9 - Resposta chatGPT

- **ShapeFile e Portal do IBGE:**

Infelizmente o arquivo utilizado e disponibilizado não contém o dado. No portal do IBGE por sua vez, não foi possível também verificar pois não possível fazer uma seleção específica como essa. Por conta dessa limitação, apenas uma pergunta extra

sobre número de habitantes, onde será comparado os dados do chatGPT com dados conferidos um a um.

Sendo essa a resposta e a query feita ao chatGPT:

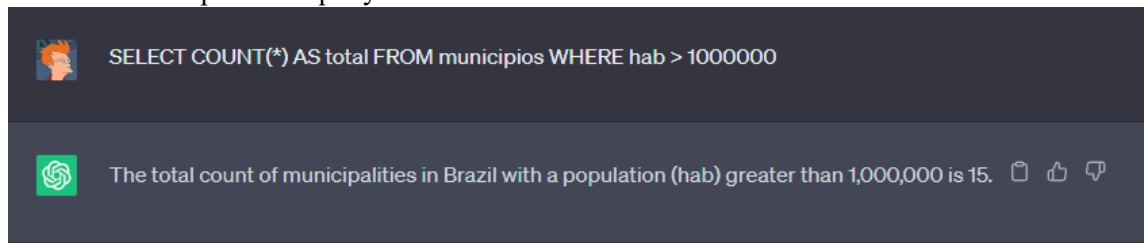


Figura 10 - Resposta chatGPT

A informação foi posteriormente confirmada um a um, mediante a lista encontrada na internet desses municípios.

3. Municípios no estado de Minas Gerais:

- **chatGPT:** Abaixo a query e o resultado obtido.

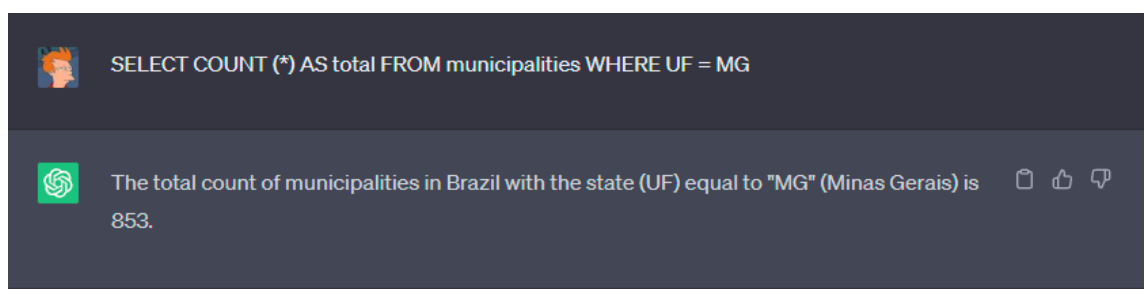


Figura 11 - Resposta chatGPT

- **ShapeFile:**
Utilizando o filtro para o arquivo por meio da função: "SIGLA_UF" LIKE 'MG', obtivemos o retorno de 853 feições, que correspondem ao número de municípios do estado. Nota-se que a camada selecionada no canto esquerdo inferior, continua sendo a BR_Municipios, mostrando que apenas as camadas selecionadas estão sendo exibidas.

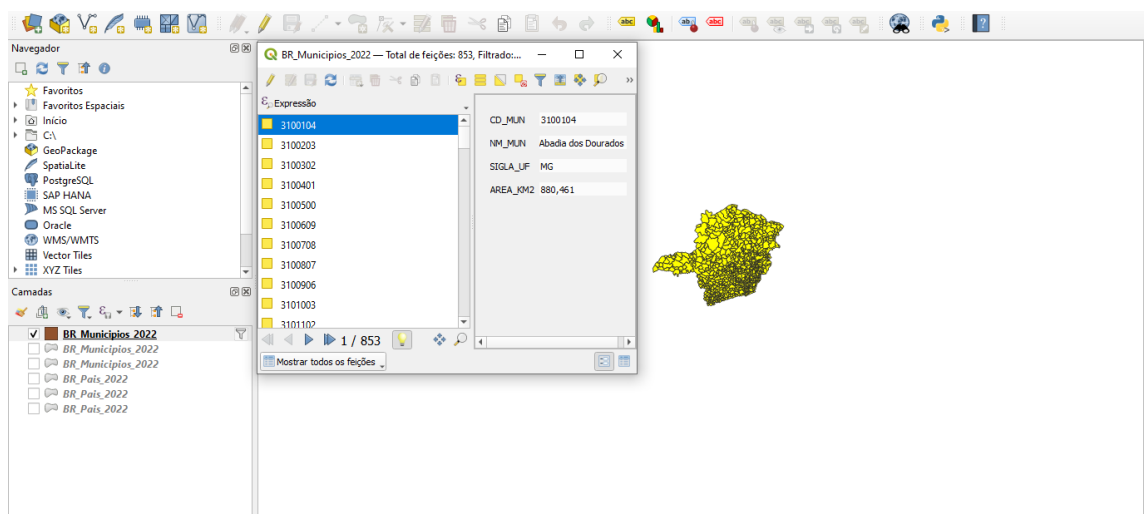


Figura 12 - Resposta QGIS

- **Portal IBGE**

No portal do IBGE a pesquisa foi realizada a partir dos municípios e escolhendo o estado que iria verificar, com isso no canto superior esquerdo da pesquisa aparecem quantos municípios existem, como mostrado abaixo.

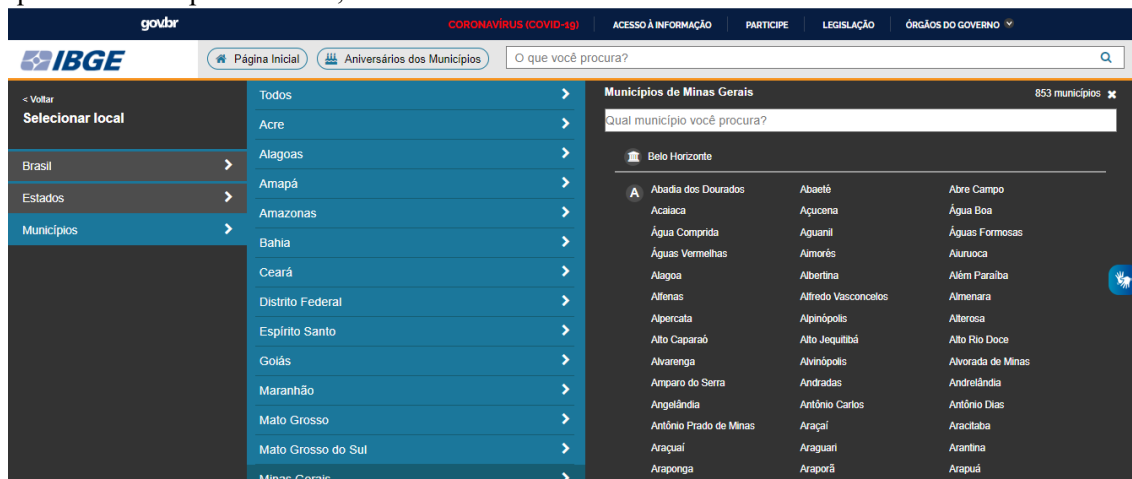


Figura 13 - Informação portal IBGE

4. NÚMERO TOTAL DE MUNICIPIOS DO SUDESTE:

- **chatGPT:** A query fornecida ao chatGPT e seu resultado, foram os seguintes:

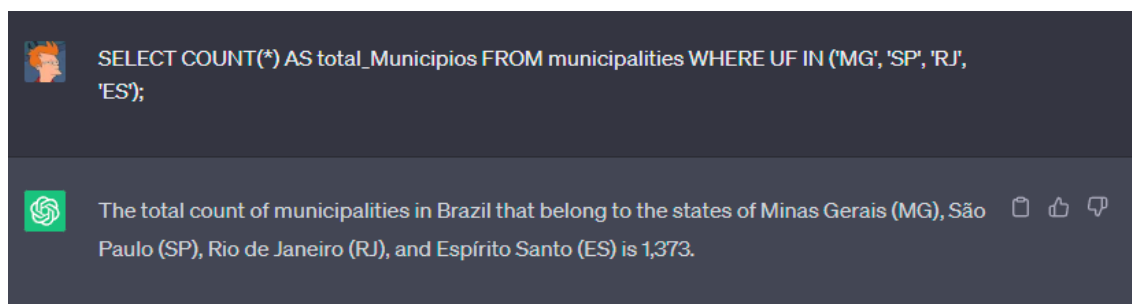
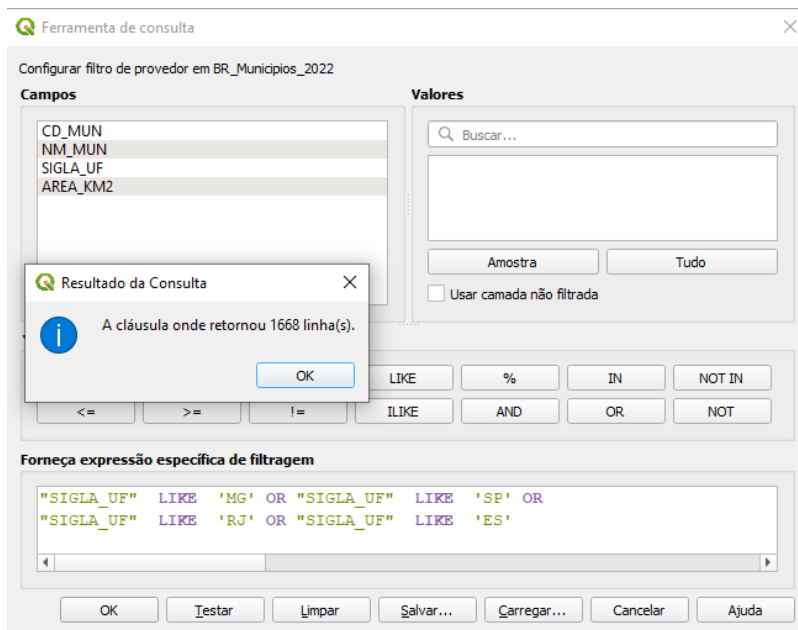


Figura 14 - Resposta chatGPT

- **ShapeFile:** Para o arquivo consultado no QGIS a função de filtro foi um pouco mais complexa, uma vez que os dados não contêm a região que o município se encontra, dessa forma foi realizada a mesma função feita para filtrar Minas Gerais e foram adicionados o operador OR para que fossem filtrados e coletados também municípios com a SIGLA_UF referente aos outros estados da região, retornando 1668, dado esse que difere e muito do fornecido pelo chatGPT:



Agora, abaixo o arquivo renderizado mostrando apenas municípios do Sudeste brasileiro:

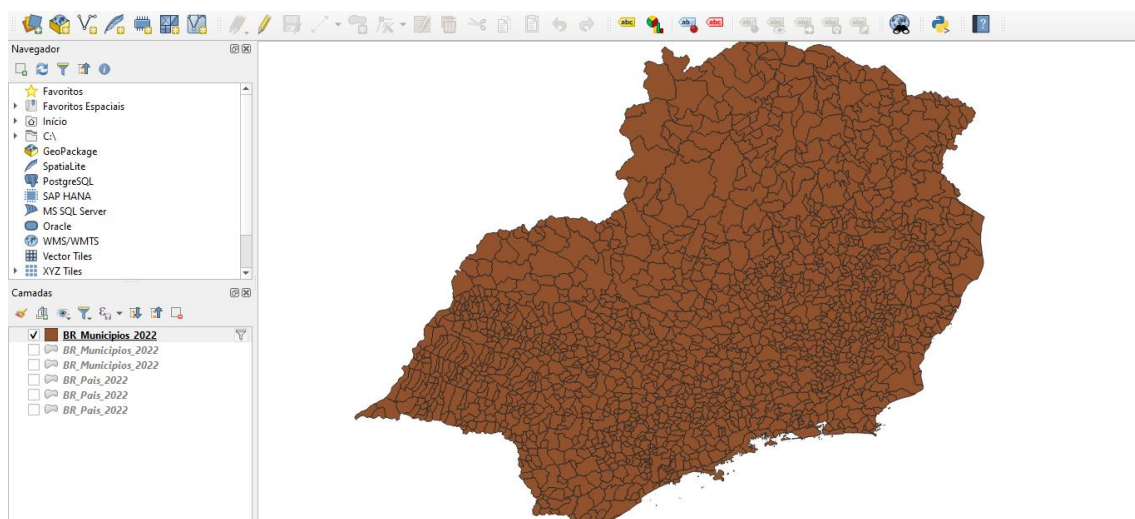


Figura 15 - Resposta QGIS

- **Portal IBGE:**

No portal do IBGE a pesquisa seguiu o mesmo processo também da do estado de Minas Gerais, onde foi realizada a busca estado por estado que deram os seguintes resultados:

RJ:

Municípios de Rio de Janeiro

92 municípios ✕

ES:

Municípios de Espírito Santo

78 municípios ✕

SP:

Municípios de São Paulo

645 municípios ✕

Com esses números dos respectivos estados, temos o total de toda a região sudeste do Brasil, que seria MG: 853 + RJ: 92 + ES: 78 + SP: 645, que daria um total de: 1668, valor esse que está de acordo com o obtido através do filtro no shapeFile e diminuindo a credibilidade do chatGPT.

5. NÚMERO TOTAL DE MUNICÍPIOS NO BRASIL:

- **chatGPT:** O número total obtido no chatGPT e sua query, foram as seguintes:

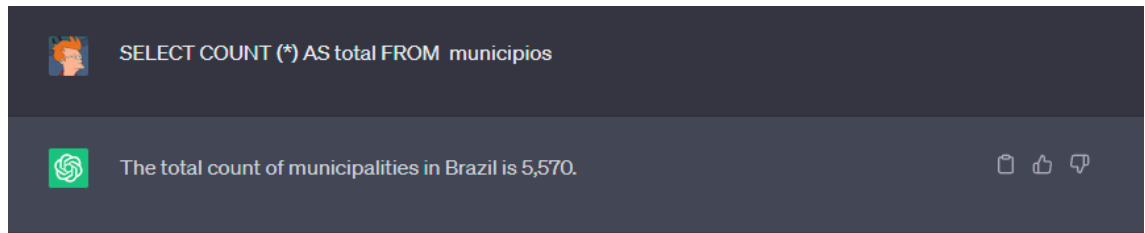


Figura 16 - Resposta chatGPT

- **ShapeFile:** O arquivo consultado por meio do QGIS, não precisou de ter um filtro aplicado a ele para obter o resultado de todos os municípios, uma vez que por se tratar apenas de um arquivo de municípios, foi necessário somente pegar o número total de entrada de dados, sendo o total de: 5572

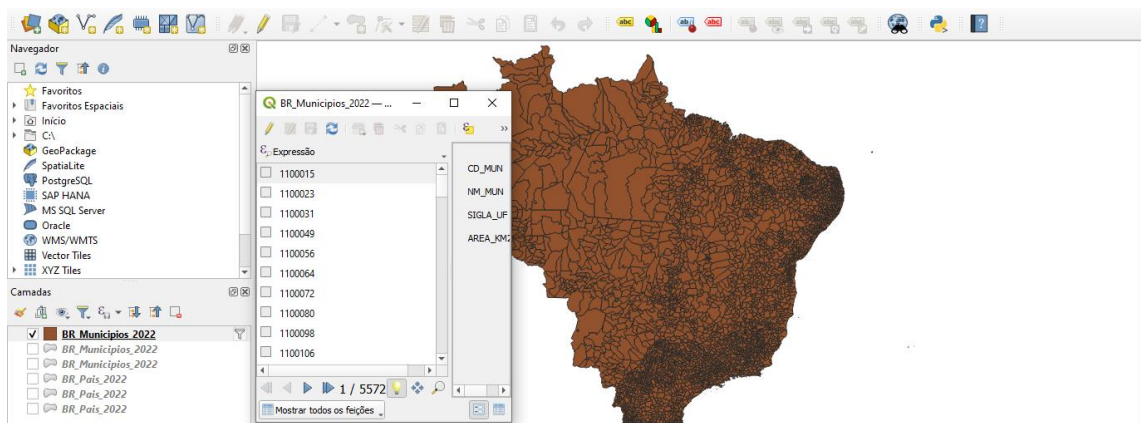


Figura 17 - Resposta QGIS

- **Portal IBGE:** No portal do IBGE foi necessário apenas fazer uma consulta de um panorama geral do país e com isso obter o número total de municípios, 5570. Dessa vez confirmando o dado do chatGPT e contradizendo o arquivo do próprio IBGE. Porém essa diferença pode ser devido a idade do arquivo de shapeFile, sendo desatualizado por ser uma mídia mais trabalhosa de ser trabalhada e atualizada, enquanto o próprio site consome um banco de dados atualizado diariamente.

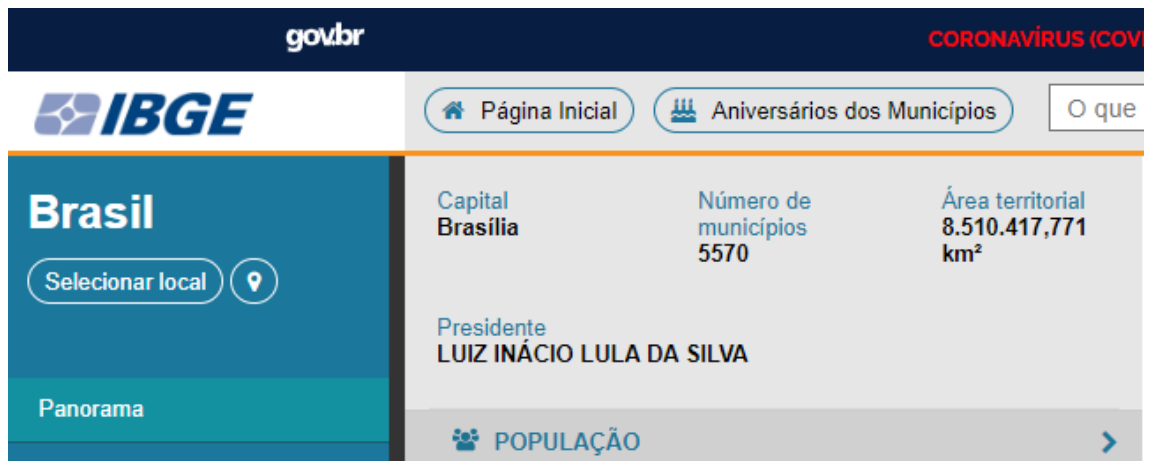


Figura 18 - Informação portal IBGE

Antes de partir para conclusões e interpretações, uma última pergunta foi feita, uma vez que o Brasil é o maior país católico do mundo, foi realizado um prompt ao chatGPT para buscar o número de municípios com nome de Santo, buscando por “São”, “Santo” ou “Santa”. E esse foi o resultado obtido:

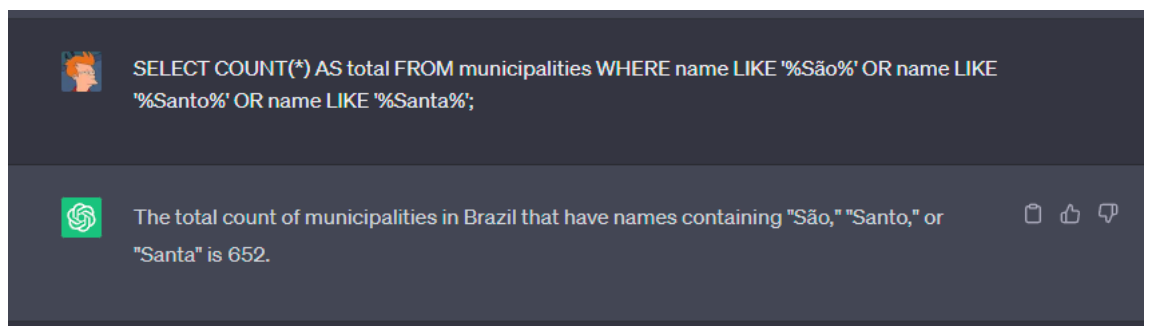


Figura 19 - Resposta chatGPT

Enquanto no ShapeFile foi retornado e seguinte valor com a seguinte função:

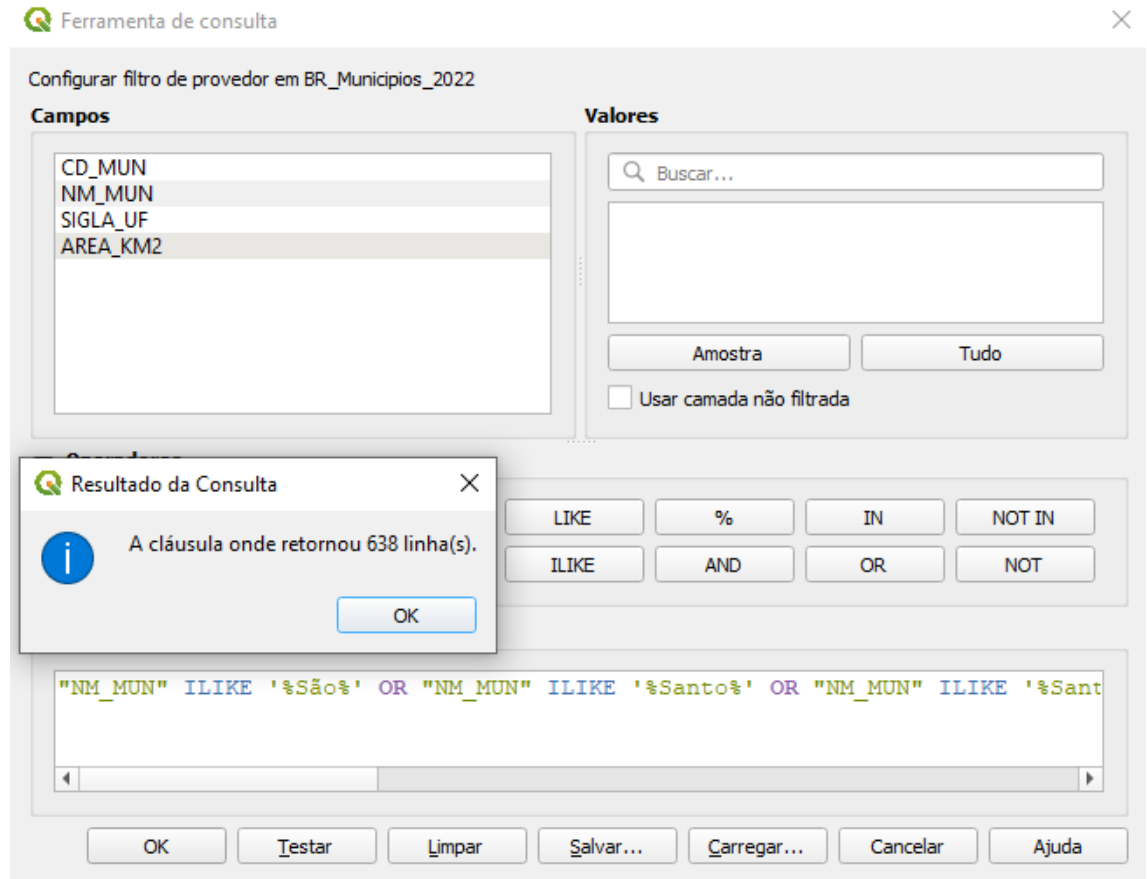


Figura 20 - Resposta QGIs

Para ilustrar, a seguir a imagem dos locais onde estão estes municípios:



Figura 21 - Distribuição espacial

4.2. ANÁLISE DOS RESULTADOS

Tendo em mãos os resultados, é possível levantar alguns pontos sobre os bancos consumidos e a confiabilidade dos dados fornecidos. Partindo da primeira fonte consumida e a mais controversa, temos os resultados fornecidos pelo chatGPT, que apesar de ter entregado alguns resultados e terem sido certos, não indico usá-lo dessa forma. O motivo de não indicar foi a dificuldade apresentada pelo software de entender os prompts e de se comportar da maneira que foi solicitado, a razão de ter o incluído no presente trabalho se deve ao fato já citado anteriormente do mesmo ser considerado polêmico e não confiável, tendo em mente que os dados que treinaram a IA são do início de 2022, pode ser que alguns dados estejam errados, o próprio dado do número total de municípios pode ter sido corretamente indicado de forma errada. Por esses fatos, apesar da facilidade e da comodidade de utilizar de tal software para realizar pesquisas, ela não se mostra muito confiável e também apresenta dificuldades em entender prompts enviados. Partindo agora para a análise dos resultados apresentados pelas duas outras plataformas, temos resultados confiáveis, uma vez que ambos são disponibilizados pelo próprio IBGE e que trabalha incessantemente para atualizar esses mesmos dados. Voltando nossa atenção ao arquivo shapeFile e ao software QGIS temos um cenário interessante, onde os arquivos e o software não são voltas especificamente para a utilidade aqui apresentada, essa funcionalidade utilizada no presente trabalho é na verdade a metade do processo de criação do produto final proposto pelo software, que seria a análise espacial e confecção de mapas por meio do mesmo software, como forma de exemplificar o fato aqui dito, foi realizado também utilizando o software um layout de um mapa com os municípios de MG, aplicando as boas práticas de confecção de mapas:

Estado de Minas Gerais e sua capital

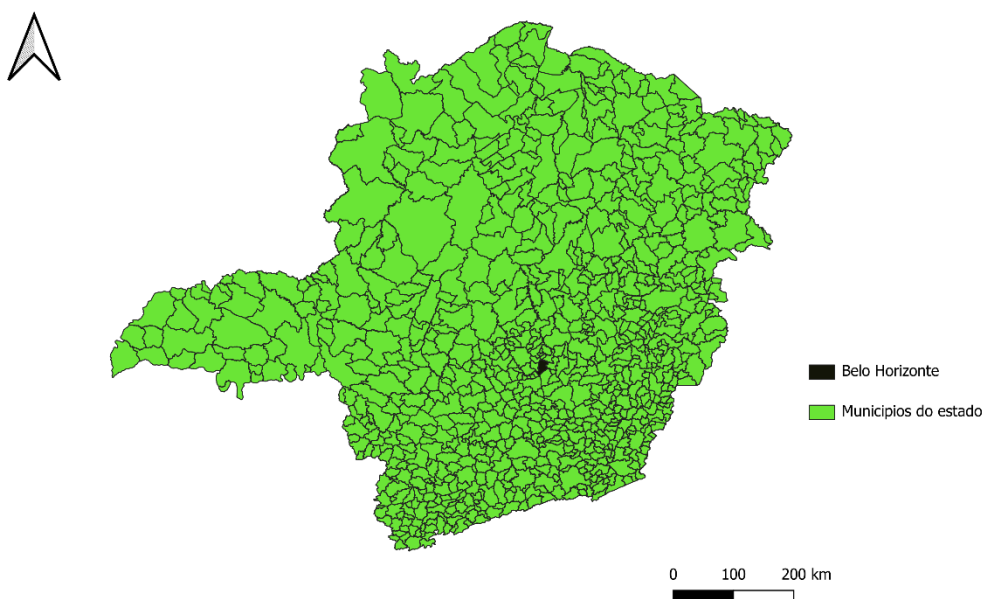


Figura 22 - Layout de Mapa - Exemplo

Com isso, apesar não ser o objetivo final do software, a consulta e consumo de dados filtrados é um uso interessante, porém não é de fácil acesso. Já o portal do IBGE contém muitos dados específicos, porém o portal se mostra um tanto quanto “engessado” para consultas que fujam do padrão de informações que eles disponibilizam, não é possível fazer uma busca específica de dados ou conjunto de dados. Tendo isso em mente, apesar das limitações de entrada, para consumo de dados mais específicos o QGIS se mostra mais interessante, foi dado apenas o exemplo utilizado nesse trabalho, porém existem shapeFiles de inúmeras fontes, como por exemplo a rede hidrográfica completa brasileira, onde a partir dela é possível fazer pesquisas super específicas. Outra vantagem que não foi muito abordada no presente trabalho é a possibilidade de shapeFiles terem dados mais específicos e variados, por exemplo geolocalização de hospitais na capital mineira, onde dentro da tabela de atributos poderiam conter dados como, possui pronto-socorro, possui heliporto, informações mais específicas nesse sentido. A área do geoprocessamento por ser muito abrangente, consegue chegar e coincidir com muitas áreas e isso permite essa gama de informações e conteúdos variados.

Porém, caso sua pesquisa seja mais generalista e menos abrangente, seria mais indicado o próprio site do IBGE, pesquisas rápidas com temas generalistas são indicadas para tal plataforma.

5.CONCLUSÕES:

Consumo de dados é uma prática cotidiana da maioria das pessoas, porém passa despercebido por não ser diretamente nomeado dessa maneira, portanto acredito que levantar métodos de consumir dados diferente da forma com que estamos acostumados é uma ótima maneira de abrir a mente da população e assim diminuir um pouco do estigma que existe em cima do termo de consumo de banco de dados ou termos próximos. Demonstrando que uma simples consulta no portal do IBGE ou mesmo uma pergunta ao chatGPT pode ser uma forma de consumir dados e fazer pesquisas, essa barreira e bloqueio mental pode ser desmontada. Apesar de alguns exemplos aqui demonstrados não terem sido fiéis a realidade em alguns momentos, isso se dá ao fato da data dos dados consumidos por essas plataformas serem mais antigos, porém na maioria foram resultados positivos. Contudo, o olhar e o senso crítico para toda e qualquer tipo de informação consumida e fornecida, uma vez que da mesma forma que é extremamente fácil consumir dados na internet, também é fácil disponibilizar, por esse motivo foram utilizados nesse trabalho, dados de um instituto governamental de extrema confiança.

REFERENCIAS BIBLIOGRAFICAS:

Shape file município do Brasil; Disponível em:

<https://www.ibge.gov.br/geociencias/organizacao-do-territorio/malhas-territoriais/15774-malhas.html>;

Portal IBGE; Disponível em: <https://www.ibge.gov.br/>;

chatGPT; openAI; Disponível em: <https://openai.com/blog/chatgpt>;

QGIS PROJECT; Disponível em: <https://qgis.org/en/site/>;

WESTIN, Ricardo; Disponível em:

<https://www12.senado.leg.br/noticias/especiais/arquivo-s/1o-censo-do-brasil-feito-ha-150-anos-contou-1-5-milhao-de-escravizados>;