

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

## Entrega 1

**Luiz Henrique A. P. Bittencourt - 321111628**

Centro Universitário UNA – Rua dos Aimorés 1451, Lourdes-Belo Horizonte  
Departamento de Tecnologia – Análise e desenvolvimento de sistemas

Belo Horizonte – 29 de Maio de 2023

## **1.PROPOSTA DE PROJETO:**

Para a realização do trabalho proposto para a UC de Análise de dados e Big Data, foi pensado para a confecção uma análise comparativa dos resultados de dados provenientes de diferentes fontes afim de ilustrar e mostrar a confiabilidade dessas plataformas futuramente aqui propostas. A ideia inicial foi consumir dados referentes aos municípios Brasileiros como um todo, buscando dentre eles algumas informações genéricas que pudessem assim tanto trazer a informação, mostrar métodos diferentes de consulta e testar a confiabilidade.

Para realizar essas consultas, foram levantadas três plataformas que serão colocadas frente a frente para testar os resultados obtidos e sua confiabilidade, sendo elas, o software de inteligência artificial da empresa openAI, o chatGPT, um software de geoprocessamento de código livre, chamado QGIS onde por meio dele serão consumidos dados provenientes das base de dados do governo e juntamente com essa base de dados do governo, o consumo de dados no portal do IBGE que tem uma interface mais amigável e fácil de ser utilizada.

### **1.1.FONTES DE DADOS:**

- **chatGPT:**

A controversa plataforma da empresa openAI foi a ideia inicial de se realizar buscas em bancos de dado, o software permite que você faça prompts (comandos) a ele que ditem como agir e o que ele deve fazer. Por meio desses comandos é possível dizer ao software que ele deve funcionar como um banco de dados SQL aceitando apenas queries da linguagem e onde qualquer outro tipo de comando será ignorado ou indicado como erro. A ideia partiu de diversas interações com a plataforma postadas em diferentes locais da internet, mostrando a real capacidade dela. Porém, apesar da facilidade e acessibilidade, o objetivo do presente trabalho não é de apenas usar o chatGPT como sua principal fonte de dados, esse papel ficará com o portal do IBGE onde será confirmado os dados e testada a confiabilidade da IA, o objetivo de incluir o chatGPT no presente estudo é trazer luz sobre as capacidades da mesma e diferente usos ele pode ter, ao mesmo tempo que testaremos a sua confiabilidade de trazer dados reais.

- **Arquivos de shapeFile consultados por meio do software QGIS:**

A segunda fonte de dados que utilizaremos no presente estudo seria a consulta por meio de filtro da tabela de dados do shapefile de municípios disponibilizadas pelo governo brasileiro. Arquivos com extensão shapefile, são arquivos de geoprocessamento que utilizam de referências geográficas como coordenadas e polígonos desenhados nessas referencias para representar feições do mundo real, sejam abstratas ou reais. Por exemplo uma feição abstrata do mundo real seria o limite entre dois estados e uma feição real, seria o curso de um rio. Dentro desses arquivos, existe uma tabela de atributos onde cada polígono tem atribuídos a ele, dados, como por exemplo nome, extensão, numero de

habitantes, etc. Esse tipo de informação é de certo modo de difícil acesso, não se acha facilmente uma plataforma onde se pode filtrar cidades brasileiras pelo tamanho de seu território e ao mesmo tempo ver uma representação gráfica das mesmas. Portanto, considerou-se usar esse meio, como uma segunda e mais importante fonte de dados.

- **Portal do IBGE:**

Por ultimo e mais importante, vem o portal do IBGE que vai servir não só para confirmar ou refutar dados recebidos das duas fontes anteriores como também ser uma terceira fonte de dados, por meio do portal do IBGE se tem acesso gratuito e livre a diferentes dados de todos os estados e municípios brasileiros, onde é possível realizar buscas de dados específicos, porém mais generalistas dos temas citados anteriormente.

## **1.2. FERRAMENTAS:**

Para a realização do trabalho aqui propostos, será necessário a utilização apenas de um navegador, para duas das fontes de dados, portal IBGE e chatGPT. Porém, para a terceira fonte de dados, serão necessárias duas ferramentas/arquivos.

A ferramenta em questão que será obrigatória para a realização dessas consultas, é o software de código livre chamado QGIS, esse software é especializado em geoprocessamento, onde por meio dele é possível realizar a confecção de mapas de consultas simples ou complexas, existem funções e plugins no software que permitem até mesmo de forma automática traçar a vazão de um dado rio, apenas perfil de elevação de um terreno. Porém, o software precisa de um arquivo contendo os dados os quais ele precisa trabalhar, nesse ponto é que se encontra os arquivos de extensão shapefile, que é apenas uma das extensões utilizadas em geoprocessamento. Dentro desse arquivo, quando aberto pelo software de geoprocessamento, é possível verificar a presença de polígonos diferentes, linhas e pontos, cada uma dessas formas representando uma feição do mundo real. Para a realização deste trabalho, será utilizado o shapefile de municípios, disponibilizado de forma gratuita pelo governo brasileiro, por meio do site do IBGE.

## **2.RESULTADOS ESPERADOS**

Durante a realização do trabalho e ao chegar a conclusão, se espera que ao relacionar os dados entre as três fontes utilizadas, alguns comportamentos sejam esperados:

- 1- **Discordância entre dados do IBGE:** Isso se deve ao fato do portal ser atualizado com base num banco de dados e o arquivo utilizado tem datas de criação onde os dados presentes nele, correspondem a data, podendo assim serem desatualizados

- 2- **Erros do chatGPT por conta de estar desatualizado:** O chatGPT apesar de ser uma plataforma de inteligência artificial bem confiável, ele tem um problema quando tratamos dados dessa forma. A base com que a IA foi treinada, data do fim de 2021, portanto qualquer alteração posterior a essa data, não será contabilizada pelo software, deixando-o assim, desatualizado.
- 3- **Dados em acordo, porém podendo haver pequenas discordâncias:** É esperado que os dados do IBGE ambos estejam completamente certos, dadas as limitações da idade do shapefile, porém é esperado que existam pequenas diferenças entre os dois, podendo ser por conta de diferentes fatores, porém ambos estarão bem próximos.
- 4- **Todos os dados em acordo na maioria dos casos:** Também é esperado que todos os dados que futuramente serão buscados por meios das três fontes, sejam alinhados uns com os outros na maioria das vezes, porém tendo um comportamento destoante partindo mais vezes do chatGPT.

### 3.REFERENCIAS BIBLIOGRAFICAS:

Shape file município do Brasil; Disponível em:

[https://www.ibge.gov.br/geociencias/organizacao-do-territorio/malhas\\_territoriais/15774-malhas.html](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/>;