PROMPT2MAP: MAP GENERATION FROM NATURAL LANGUAGE QUERIES USING LARGE LANGUAGE MODELS (LLM)

José Miguel Cordero Carvacho

Dissertation presented as partial requirement for obtaining the degree of Master in Geographical Information Systems and Science

NOVA Information Management School

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade NOVA de Lisboa

PROMPT2MAP: MAP GENERATION FROM NATURAL LANGUAGE QUERIES USING LARGE LANGUAGE MODELS (LLM)

by

José Miguel Cordero Carvacho

Dissertation presented as partial requirement for obtaining the degree of Master in Geographical Information Systems and Science

Adviser: Professor Pedro da Costa Brito Cabral

November, 2024

Resumo

Prompt2Map: Geração de Mapas a partir de Consultas em Linguagem Natural Utilizando Grandes Modelos de Linguagem (LLM)

Os Sistemas de Informação Geográfica (SIG) e as tecnologias web tornaram a criação de mapas mais acessível do que nunca. No entanto, ainda é necessário um conhecimento técnico das ferramentas de SIG para produzir resultados cartográficos. Esta tese apresenta o Prompt2Map, um sistema que converte consultas em linguagem natural em mapas web utilizando Grandes Modelos de Linguagem (LLMs). A nossa abordagem baseia-se em *Retrieval-Augmented Generation* (RAG), envolvendo uma etapa inicial de extração de dados a partir de fontes geoespaciais, seguida de uma etapa de mapeamento para visualizar os dados. Testes de desempenho realizados com consultas sintéticas demonstram a eficácia do sistema. Discutimos também as implicações éticas desta abordagem. Este trabalho contribui para reduzir a lacuna entre consumidores e produtores de mapas, oferecendo uma interface em linguagem natural para dados geoespaciais autoritativos, aproximando assim a informação espacial do público geral.

Palavras-chave

Grandes Modelos de Linguagem

Sistemas de Informação Geográfica

Cartografia

IA Generativa

RAG

Abstract

Prompt2Map: Map Generation from Natural Language Queries Using Large Language Models (LLM)

Geographic Information Systems (GIS) and web technologies have made map creation more accessible than ever before. However, a technical understanding of GIS tools is still required to produce cartographic outputs. This thesis introduces Prompt2Map, a system that converts natural language queries into web maps using Large Language Models (LLMs). Our approach is based on Retrieval-Augmented Generation (RAG), involving an initial step to extract data from geospatial sources, followed by a mapping step to visualize the data. Performance tests conducted over synthetic prompts demonstrate the system's effectiveness. We also discuss ethical implications of this approach. This work contributes to bridging the gap between map consumers and producers, offering a natural language interface to authoritative geospatial data, thereby bringing spatial information closer to the general public.

Keywords

Large Language Models

Geographic Information Systems

Cartography

Generative AI

Retrieval-augmented generation

Contents

List of Tables

List of Figures

1

Introduction

1.1 Motivation

Geographic Information Systems (GIS) have become indispensable tools in various disciplines, including urban planning, environmental management, public health, and transportation (Longley et al., 2015). These systems enable professionals to visualize, analyze, and interpret spatial data, facilitating informed decision-making based on geographic relationships and patterns. Despite the proliferation of GIS technologies and the increasing availability of geospatial data, the process of creating meaningful maps and conducting spatial analyses often remains confined to experts with specialized training. This expertise barrier limits the accessibility and utility of GIS for a broader audience, including policymakers, educators, and the general public (Haklay, 2013).

The democratization of GIS has been an ongoing pursuit, with efforts focused on making spatial data and mapping tools more accessible to non-specialist users. Advancements in technology, such as web-based GIS platforms and open-source software, have lowered some barriers, allowing users to access spatial data and perform basic analyses without requiring extensive technical knowledge (Malakar & Roy, 2024). However, these platforms often still necessitate familiarity with GIS concepts, terminologies, and interfaces, which can be intimidating for those without specialized training. This gap underscores the need for more intuitive and user-friendly interfaces that can cater to a diverse range of users with varying levels of expertise.

Simultaneously, the field of Artificial Intelligence (AI), particularly Natural Language Processing (NLP), has witnessed significant advancements. Large Language Models (LLMs), such as OpenAI's GPT series, have demonstrated remarkable capabilities in understanding and generating human-like text (Sindhu et al., 2024). These models can interpret nuanced queries, generate coherent responses, and perform tasks like code generation and translation. The integration of LLMs with GIS presents an opportunity to bridge the gap between expert GIS practitioners and non-specialist users. By enabling users to interact with GIS platforms through natural language queries, we can lower the barriers to entry, allowing a wider audience to leverage spatial data for various purposes.

Such integration could revolutionize how users access, analyze, and visualize geospatial information, making GIS more inclusive and user-friendly. For instance, policymakers could generate complex spatial analyses without needing to write queries

in technical language or manipulate GIS software interfaces manually. Educators could create interactive maps for teaching purposes with simple verbal instructions, enhancing the learning experience for students. Additionally, the general public could engage with spatial data to better understand community planning initiatives, environmental changes, and public health trends.

However, this integration is not without challenges. Accurately interpreting natural language queries in a geospatial context requires the LLM to comprehend spatial concepts, relationships, and terminologies. Ensuring that the system retrieves the correct data and generates accurate maps is critical to maintaining trust and utility (Zhang et al., 2023). Ethical considerations, including data privacy, trustworthiness, and potential biases, must also be addressed to ensure responsible deployment. Misinterpretations of queries could lead to incorrect data retrieval and misleading visualizations, which could have serious implications in fields like public health or urban planning.

This thesis proposes the development of Prompt2Map, a system that leverages LLMs to convert natural language queries into interactive web maps. The system aims to facilitate geospatial data retrieval and map generation, making GIS more accessible to non-expert users. Through this work, we seek to contribute to ongoing efforts to democratize GIS and explore the intersection of AI and geospatial technologies. By addressing both technical and ethical challenges, Prompt2Map aspires to provide a reliable and user-friendly tool that empowers a diverse range of users to engage with spatial data effectively.

1.2 Objectives

The objectives of this thesis are threefold. First, to develop the Prompt2Map system, which is capable of converting natural language queries into interactive web maps by leveraging LLMs in conjunction with geospatial data retrieval techniques. The system should be able to interpret a wide range of queries, retrieve relevant data from various sources, and generate accurate and informative maps. This involves integrating advanced NLP capabilities with robust data processing and visualization tools to create a seamless user experience.

Second, to evaluate the effectiveness of Prompt2Map through a series of validation tests using synthetic prompts tailored to specific datasets. The evaluation will analyze metrics such as accuracy in data retrieval and map generation. This comprehensive evaluation aims to ensure that the system meets its performance goals and can handle diverse user needs and data complexities.

Third, to analyze the ethical implications associated with using LLMs for map generation. This includes issues of data privacy, potential biases in data and algorithms, alignment with user intent, risks of misinformation, and the potential for misuse. Strategies for mitigating these risks will be proposed and integrated into the system design.

Ensuring ethical integrity is paramount to maintaining user trust and preventing harmful outcomes resulting from AI-generated maps.

By achieving these objectives, the thesis aims to contribute to making GIS tools more accessible and user-friendly, fostering greater engagement with spatial data among non-specialists, and promoting responsible use of AI in geospatial applications.

1.3 Thesis Structure

The thesis is organized into six chapters. The introduction provides the motivation for the study, outlines the objectives, and summarizes the thesis structure. The background chapter reviews the evolution of GIS towards greater accessibility, discusses the role of neogeography and self-service GIS platforms, introduces LLMs in the context of GIS, and addresses ethical considerations related to AI-generated maps.

The methods chapter details the system architecture of Prompt2Map, describes the datasets used for evaluation, and outlines the methodologies for system testing and ethical assessment. It delves into the technical aspects of how natural language queries are processed, data is retrieved, and maps are generated, providing a comprehensive overview of the system's components and workflows.

The results chapter presents the outcomes of validation tests, showcasing the system's ability to generate maps from natural language queries. It analyzes system performance in terms of accuracy, response time, and scalability, supported by various figures and tables that illustrate the findings. This chapter provides empirical evidence of the system's effectiveness and identifies areas where it excels or requires improvement.

The discussion interprets the findings, discusses the system's effectiveness in meeting its objectives, explores ethical implications, and identifies limitations and areas for improvement. It contextualizes the results within the broader landscape of GIS and AI, offering insights into how Prompt2Map contributes to the field and what future developments could enhance its capabilities.

Finally, the conclusions summarize the key findings, highlight the contributions to the fields of GIS and AI, discuss potential future work, and reflect on the broader impact of natural language interfaces in GIS. This chapter synthesizes the research, drawing connections between the objectives, methods, and results, and proposing directions for future exploration and innovation.

1.4 Background

1.4.1 Democratization of GIS

GIS have evolved significantly since their inception, transitioning from specialized tools used primarily by experts to more widely accessible platforms. Initially, GIS required substantial technical expertise, including knowledge of programming, database management, and spatial analysis techniques. The steep learning curve and high costs associated with GIS software limited its use to specific professional domains.

Advancements in technology and the proliferation of geospatial data have catalyzed the democratization of GIS. The development of user-friendly interfaces, open-source GIS software (e.g., QGIS), and web-based platforms (e.g., ArcGIS Online) has lowered barriers to entry. These tools have enabled a broader audience to engage with spatial data, perform basic analyses, and create maps without requiring extensive technical training. The increasing availability of high-quality geospatial data, coupled with more intuitive tools, has empowered non-specialist users to harness the power of GIS for various applications, from community planning to environmental monitoring.

The concept of neogeography emerged as individuals outside traditional geographic professions began using mapping tools for personal and community projects. Goodchild describes neogeography as the use of geographical techniques by non-expert users for non-professional purposes (Goodchild, 2009). This movement has been facilitated by platforms like OpenStreetMap, which allow users to contribute to and utilize geospatial information collaboratively. Neogeography has expanded the scope of GIS, enabling everyday users to participate in mapping activities, share spatial data, and contribute to collective knowledge bases. This participatory approach has democratized mapmaking, making it a communal and accessible activity.

Despite these advancements, a gap persists between the capabilities of expert GIS practitioners and those accessible to non-specialists. Complex spatial analyses, custom data queries, and advanced cartographic designs often remain beyond the reach of general users. This gap underscores the need for innovative approaches to further democratize GIS, making advanced functionalities more accessible and intuitive. Enhancing user interfaces, integrating intelligent assistants, and leveraging AI-driven tools are potential strategies to bridge this gap (Frez & Baloian, 2023), enabling a wider audience to perform sophisticated GIS tasks without specialized training.

1.4.2 Self-Service GIS

Self-service GIS platforms represent an important step toward making GIS more user-friendly. These platforms provide tools that simplify spatial data analysis and map creation through intuitive interfaces, often featuring drag-and-drop functionalities, guided workflows, and pre-configured templates (Rowland et al., 2020). Users can perform tasks such as data visualization, basic spatial analysis, and thematic mapping

without deep technical knowledge. By abstracting the complexity of GIS operations, self-service platforms enable users to focus on their specific analytical needs and visualization preferences.

However, self-service GIS platforms have limitations. They may not fully accommodate the diverse and complex needs of users, especially when dealing with sophisticated analyses or custom data queries. Users still need to understand GIS concepts, data structures, and the specific functionalities of the platform. The interfaces may not be flexible enough to handle unique or nuanced user requests. For instance, advanced spatial analyses like network analysis, spatial interpolation, or multi-criteria decision analysis often require specialized tools and a deeper understanding of GIS methodologies, which are not typically supported by self-service platforms.

To overcome these limitations, integrating natural language interfaces into GIS platforms has been proposed. Natural language interfaces allow users to interact with systems using everyday language, potentially reducing the learning curve and making GIS functionalities more accessible. By expressing queries and commands in natural language, users can bypass complex menus and technical commands. This approach leverages the advancements in NLP and AI to create more intuitive and responsive GIS tools that can cater to a wider range of user needs and expertise levels. Natural language interfaces can facilitate more dynamic and conversational interactions with GIS systems, enabling users to ask complex questions and receive immediate, contextually relevant responses.

1.4.3 Large Language Models in GIS

Large Language Models have revolutionized the field of NLP, offering new ways to interpret and generate human-like text. Models like GPT-4 have been trained on extensive datasets encompassing a wide range of topics, enabling them to understand context, disambiguate meanings, and generate coherent responses. These models excel at tasks such as language translation, text summarization, and question-answering, demonstrating a high level of proficiency in understanding and generating natural language.

In the context of GIS, LLMs can serve as intermediaries between the user and the system, interpreting natural language queries and translating them into actions that the GIS platform can execute. For example, a user could input a query like "Show me the areas in Lisbon with the highest unemployment rates,"and the LLM would process this request, retrieve the relevant data, and generate the appropriate map.

Furthermore, the concept of Retrieval-Augmented Generation (RAG) enhances LLM capabilities by combining retrieval mechanisms with generation. In a RAG system, the LLM retrieves relevant information from external data sources to inform its responses (Fan et al., 2024). In GIS applications, RAG can enable the system to access geospatial databases dynamically, allowing for up-to-date and context-specific map generation.

This approach ensures that the system can provide accurate and relevant visualizations based on the most current and pertinent data, enhancing the overall utility and reliability of the GIS platform.

Literature on the application of LLMs in GIS showcases various use cases where AI-driven natural language interfaces have significantly improved user interaction and data analysis capabilities. For instance, studies have explored how LLMs can assist in automated map generation and interactive spatial querying (S. Wang et al., 2024). These applications demonstrate the potential of LLMs to transform traditional GIS workflows, making spatial data analysis more efficient and accessible. Additionally, research has highlighted the benefits of integrating LLMs with GIS for educational purposes, allowing students to engage with spatial data through conversational interfaces and enhancing their learning experiences (Mooney et al., 2023; Redican et al., 2024).

Moreover, advancements in function calling and tool integration with LLMs have opened new avenues for enhancing GIS functionalities (Qu et al., 2024). Function calling enables LLMs to execute specific functions or access external tools based on user queries, thereby extending the capabilities of the GIS platform beyond standard data retrieval and visualization tasks. This integration facilitates more sophisticated spatial analyses and dynamic map generation, catering to a broader range of user needs and enhancing the overall user experience.

1.4.4 Retrieval-Augmented Generation

Traditional RAG systems primarily rely on semantic proximity through embeddings to retrieve relevant documents or text snippets from large corpora. These systems encode both queries and documents into high-dimensional vector spaces, enabling the identification of semantically similar content based on their proximity in the embedding space (Gao et al., 2024). This approach is effective for tasks such as question-answering and information retrieval, where the goal is to provide contextually relevant textual information to the user.

In the context of geospatial data, RAG holds significant potential. Geospatial information is inherently dynamic, with frequent updates and changes in data due to various factors such as urban development, environmental changes, and demographic shifts. RAG allows GIS systems to access and incorporate the most recent data, ensuring that the generated maps and analyses reflect the current state of the spatial environment. This is crucial for applications that rely on real-time or near-real-time data, such as disaster management, traffic analysis, and environmental monitoring.

The state-of-the-art in RAG involves sophisticated retrieval algorithms and indexing techniques that enable efficient access to vast amounts of data. Advances in vector-based retrieval, where data is represented in high-dimensional vector spaces, have significantly improved the speed and accuracy of information retrieval in RAG systems (Gao et al., 2024). Additionally, the integration of RAG with transformer-based architectures

has enhanced the model's ability to understand and leverage retrieved information effectively, leading to more coherent and contextually appropriate responses.

1.4.5 Text-to-SQL

Structured Query Language (SQL) is a standardized programming language specifically designed for managing and manipulating relational databases. It provides a robust framework for performing a wide array of operations, including querying data, updating records, inserting new information, and deleting existing entries. SQL's declarative syntax allows users to specify what data they need without detailing the exact procedures to retrieve it, thereby simplifying complex data interactions. This language is fundamental for data analysts, database administrators, and applications that require efficient and precise data management. In the context of Geographic Information Systems, SQL is often extended with spatial functions that enable the querying and analysis of geospatial data, facilitating tasks such as spatial joins, proximity searches, and geographic aggregations.

The task of translating natural language to database SQL queries is well-established in NLP, known as text-to-SQL. This task involves converting a natural language statement into a SQL query that can be executed against a database. Modern text-to-SQL models leverage deep learning architectures, particularly transformer-based models, which excel at capturing the complex relationships between language and database schemas. These models are trained on large datasets that pair natural language questions with their corresponding SQL queries, enabling them to learn the patterns and structures necessary for accurate translation. Techniques such as schema linking, where the model identifies and aligns entities in the natural language query with database schema elements, have been instrumental in improving performance (Z. Li et al., 2024). In addition to handling simple queries, state-of-the-art text-to-SQL models are capable of managing complex queries involving multiple tables, nested subqueries, and advanced SQL functions.

LLMs' ability to perform text-to-SQL translation can be harnessed in GIS to facilitate data retrieval. By integrating LLMs with GIS databases, users can query spatial data using natural language, and the system can generate the corresponding SQL queries to retrieve the data. This integration simplifies the data retrieval process, making it more accessible to users who may not have expertise in SQL or database management. Additionally, it allows for more dynamic and flexible querying capabilities, enabling users to perform complex spatial analyses through simple verbal instructions.

1.4.6 Function Calling

Function calling in LLMs refers to the capability of language models to invoke specific functions or access external tools based on the context of the input they receive. This feature extends the functionality of LLMs beyond text generation, allowing them

to interact with external systems, perform computations, and access specialized tools dynamically. Function calling is a crucial advancement in the development of intelligent assistants, enabling more interactive and context-aware responses.

Function calling works by informing the model about the available external functions, including their names, descriptions, and expected parameters. When the model identifies a need to perform a specific task that aligns with an available function, it generates a function call with the appropriate arguments. This call is then executed by the system, and the results are incorporated into the model's response. The process ensures that the model's outputs adhere to the constraints and requirements of the external tools, enhancing the reliability and accuracy of the responses.

In the context of GIS, function calling enables the integration of specialized mapping and spatial analysis functions. For instance, when a user requests the generation of a choropleth map based on certain criteria, the LLM can invoke a predefined mapping function with the necessary parameters, such as data layers, color schemes, and attribute values. This automation streamlines the map creation process, allowing users to generate complex visualizations with simple natural language commands.

Function calling also facilitates the extension of the system's capabilities without necessitating continuous retraining of the LLM. As new functions are developed or existing ones are updated, they can be integrated into the system's function library, and the LLM can be instructed to utilize them as needed. This modular approach ensures that Prompt2Map remains adaptable and can incorporate advanced GIS functionalities seamlessly, catering to evolving user needs and technological advancements.

Moreover, function calling enhances the system's ability to maintain consistency and standardization in map generation. By relying on predefined functions with well-defined parameters, the system ensures that the generated maps adhere to specific cartographic principles and visualization standards. This consistency is crucial for producing reliable and professional-quality maps that meet user expectations and project requirements. Additionally, the use of predefined functions facilitates replicability, as the same functions can be invoked with identical parameters to reproduce maps under similar conditions. This replicability is essential for verifying results, conducting comparative analyses, and ensuring that map generation processes can be reliably repeated in future applications.

1.4.7 Ethical Considerations

The deployment of LLMs in GIS introduces a range of ethical considerations that are crucial to address for ensuring responsible and trustworthy use. Among these considerations are bias and fairness, data privacy, accuracy and the potential for misuse, as well as trustworthiness and reproducibility. Each of these factors plays a significant role in shaping the ethical landscape of AI-integrated GIS applications.

Bias and fairness stand out as prominent concerns in the integration of LLMs with

GIS. LLMs are trained on extensive corpora of text, which inherently contain biases reflecting societal prejudices. These biases can inadvertently manifest in the model's outputs, leading to unfair or discriminatory results (Gallegos et al., 2024; Y. Wang et al., 2023). In the context of GIS, such biases may result in skewed data retrieval or the misrepresentation of specific areas or populations, thereby perpetuating existing inequalities. To mitigate these risks, it is essential to engage in careful dataset selection, implement robust bias detection and correction mechanisms, and maintain continuous monitoring of system outputs. These measures help ensure that the GIS outputs remain fair and unbiased, fostering equitable use across diverse user groups.

Data privacy is another critical ethical consideration, particularly because geospatial data often encompasses sensitive information, especially when dealing with demographic data at granular levels. Protecting user privacy and adhering to data protection regulations, such as the General Data Protection Regulation (GDPR), is paramount. GIS systems must incorporate measures to anonymize personal data, secure data storage and transmission, and prevent the inadvertent disclosure of sensitive information. Ensuring data privacy not only complies with legal standards but also builds user trust, which is essential for the widespread adoption of AI-driven GIS tools.

Accuracy and the potential for misuse present additional ethical challenges in the deployment of LLM-integrated GIS systems. LLMs have the capability to generate incorrect or misleading information, a phenomenon known as hallucinations, where the model produces plausible yet false outputs (Shuster et al., 2021). In GIS applications, inaccuracies in data retrieval or map generation can lead to significant consequences, including incorrect conclusions or flawed decision-making. Furthermore, there is a risk that the system could be exploited to generate maps that infringe on privacy, compromise security, or spread misinformation. To address these issues, it is vital to implement validation checks, establish transparency measures, and enforce strict usage policies. These strategies help mitigate the risks associated with inaccuracies and misuse, ensuring that the GIS outputs are reliable and ethically sound.

Trustworthiness and reproducibility are essential for users to confidently rely on AI-generated maps. For GIS systems to be trustworthy, they must be transparent about how outputs are produced and allow for the reproducibility of results (Z. Li & Ning, 2023; S. Wang et al., 2024; Zhang et al., 2023). Providing users with access to information about data sources, processing methods, and inherent limitations enables them to understand and verify the results effectively. Reproducibility ensures that the same inputs will consistently yield the same outputs, which is critical for validating the reliability of GIS analyses and fostering user confidence in AI-driven tools.

Addressing these ethical considerations is imperative for the responsible deployment of LLM-integrated GIS systems. By proactively identifying potential risks and implementing comprehensive mitigation strategies, such systems can enhance trust, fairness, and reliability. Ensuring that GIS applications are ethically sound not only safeguards against misuse and biases but also promotes equitable access and utilization,

ultimately contributing to more informed and just decision-making processes across various domains.

2

Methods

2.1 System Architecture

The Prompt2Map system is designed to convert natural language queries into interactive web maps by integrating LLMs with geospatial data retrieval and mapping functionalities. The system architecture comprises two main components: the Geospatial Retriever and the Map Generator. This structure is inspired by the Retrieval-Augmented Generation (RAG) architecture, although it differs in that it retrieves geospatial features instead of documents.

The system workflow involves several key steps. When a user inputs a natural language query expressing their mapping needs, the Geospatial Retriever interprets the query and retrieves relevant geospatial data from the data source. Subsequently, the Map Generator transforms the retrieved data and the user's intent into an interactive web map. Figure 2.1 shows this process in detail.

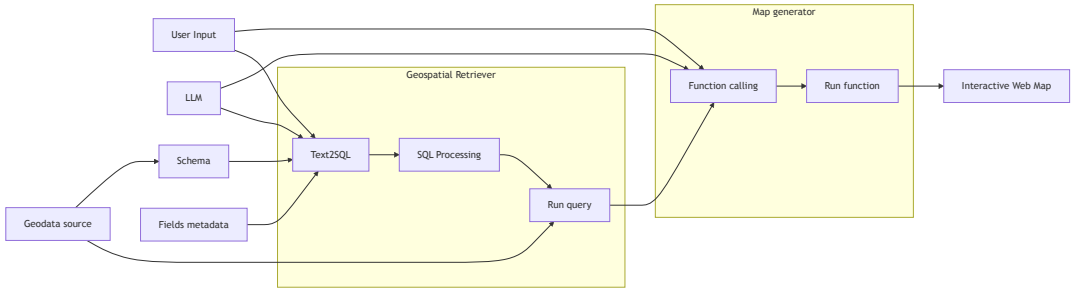


Figure 2.1: System logical architecture

2.1.1 Geospatial Retriever

The Geospatial Retriever is responsible for interpreting the user's natural language query and retrieving the relevant geospatial data from the data source. This process involves several subcomponents, including geospatial data sources, text-to-SQL translation, SQL dialect adaptation, SQL processing and validation, and query execution.

2.1.1.1 Geospatial Data Sources

Geospatial data sources constitute the foundation of any GIS application, encompassing a variety of data formats and storage systems that manage the spatial relationships and attributes of features on the Earth's surface. These data sources can take many forms, such as vector and raster datasets, and are stored in specific formats like Shapefiles, GeoPackage, GeoJSON, GeoParquet, and specialized geospatial databases such as PostGIS and DuckDB. Each data source has unique capabilities, optimized for particular types of spatial data, and supports varying levels of data complexity and query functionality. The type of geospatial data source used in a system like Prompt2Map defines the range and complexity of queries that can be performed.

2.1.1.2 Text-to-SQL Translation

As Prompt2Map adopts a RAG-inspired approach, the first step in addressing user queries is to retrieve data relevant to the request. In this context, the target structured language is SQL, which will be executed against DuckDB to query the geospatial data sources. Traditional text-to-SQL models are trained on generic SQL, focusing on "core SQL"keywords like SELECT, GROUP BY, WHERE, etc. Database-specific syntax, such as spatial extensions, are not inherently recognized by these models.

To address this limitation, instead of fine-tuning the model or extensively modifying the prompts to include spatial extensions, the system translates the user query into regular SQL without explicit mentions of geospatial columns. Subsequently, the query is processed to incorporate geospatial features deterministically. This two-step approach ensures that the text-to-SQL translation remains efficient and cost-effective while enabling the integration of spatial functionalities required for accurate data retrieval.

2.1.1.3 SQL Processing and Validation

The generated SQL query undergoes a processing phase that includes validation and correction. Validation involves checking the query for syntactic correctness, security (e.g., preventing SQL injection), and adherence to read-only constraints to protect data integrity. This ensures that the query is safe to execute and will not compromise the underlying database.

Beyond mere validation, the system actively repairs queries by identifying and rectifying potential issues that could lead to incorrect or empty results. For example, if the WHERE clause contains incorrect literals or references nonexistent fields, the system attempts to correct these errors by aligning them with the actual data schema and available metadata. This dual approach of validation and correction enhances the reliability and accuracy of the data retrieval process, ensuring that the resulting datasets are both relevant and correctly structured for subsequent mapping tasks.

After validation, the SQL query is converted to a geospatial DuckDB SQL query by injecting the necessary spatial features. This deterministic modification ensures that the query leverages DuckDB's spatial functions appropriately, enabling accurate and efficient data retrieval based on the user's natural language input.

2.1.1.4 Query Execution

The validated and corrected SQL query is executed against the geospatial data source. The choice of DuckDB as the database engine allows for efficient handling of both single-table and multi-table environments. DuckDB's support for various geospatial data formats and its in-process architecture facilitate seamless data access and manipulation, ensuring that the system can handle a wide range of geospatial queries with high performance.

2.1.2 Map Generator

The Map Generator component transforms the retrieved geospatial data and the user's intent into an interactive web map. This process involves function calling with LLMs, mapping function selection, parameter specification, and map rendering.

2.1.2.1 Function Calling with LLMs

The system utilizes the function-calling feature provided by major LLM API providers, including OpenAI, Claude, and Groq. Function calling allows the LLM to invoke predefined functions with specific parameters based on the user's query and the data characteristics. In this context, the functions correspond to mapping operations, such as generating a choropleth map, a heatmap, or a proportional symbol map. Each function has defined parameters, such as data layers, visual variables, styling options, and interactivity settings.

The LLM analyzes the metadata of the retrieved data, including geometry types, attributes, and spatial relationships. Based on this analysis and the user's intent, the LLM selects the most appropriate mapping function. For example, if the data consists of polygons with quantitative attributes and the user requests to visualize rates or densities, the LLM may select a choropleth mapping function. For point data representing events or occurrences, a heatmap function might be appropriate. Once the mapping function is selected, the LLM specifies the parameters required for the function. These include data layers to be visualized, visual variables to represent attributes through color, size, shape, or texture, styling options like color schemes and classification methods, and interactivity settings such as tooltips, pop-ups, and legends.

The mapping function is executed to render the interactive web map. The system employs web mapping libraries such as Leaflet or Plotly, which support interactive features and responsive design. The final map allows users to interact with the data,

providing functionalities like zooming, panning, clicking on features to display attribute information, and toggling layers on and off.

2.2 Implementation

The logical design described in the previous section was implemented as an open-source Python package, available via Python Package Index (PyPI)[1] and Github[2]. [Figure @fig:codeimp] shows a Unified Modeling Language (UML) diagram with the project classes and their dependencies.

The implementation of Prompt2Map follows a modular architecture, organized into three primary modules: interfaces, providers, and application. This structure adheres to the principles of Clean Architecture, ensuring that the core application logic remains decoupled from external dependencies and infrastructure concerns. Each module serves a distinct purpose within the system, facilitating scalability, maintainability, and ease of integration with various components. In this design paradigm, the application layer is decoupled from the infrastructure layer (providers) by relying on abstract interfaces rather than concrete implementations.

This approach ensures that the core business logic remains unaffected by changes in external systems or technologies. Specifically, Prompt2Map defines interfaces such as GeoDatabase, Embedding, and LLM within the interfaces module, which outline the essential functionalities required by the application. The providers module contains concrete implementations like GeoDuckDB and OpenAIProvider, which fulfill these interfaces. This modular structure offers significant advantages, including ease of maintenance, scalability, and flexibility. For instance, integrating support for a different geospatial database, such as PostgreSQL with PostGIS, would simply involve creating a new class that implements the GeoDatabase interface without altering the core application logic. This decoupling facilitates adaptability to evolving technological landscapes and simplifies the extension of system capabilities.

Python serves as the primary programming language for implementing Prompt2Map, chosen for its versatility, extensive library support, and strong presence in both the AI and geospatial communities. Python's rich ecosystem includes powerful libraries like GeoPandas, which extends Pandas to enable spatial operations on geospatial data, making it an ideal tool for managing and analyzing geographic information within the system.

2.2.1 Application

The application component constitutes the core functionality of the prompt2map package, orchestrating the end-to-end process of converting natural language queries into

[1] https://pypi.org/project/prompt2map/
[2] https://github.com/josemcorderoc/prompt2map

interactive web maps. This component is responsible for managing the workflow from query interpretation to data retrieval and finally to map generation. By leveraging the abstract interfaces defined within the interfaces module and utilizing the concrete implementations provided by the providers, the application layer ensures seamless interaction between different system modules. This modular approach not only enhances the system's scalability but also facilitates easier maintenance and future enhancements.

Within the application component, key classes such as 'SQLGeoRetriever', 'LLMPrompt2SQL', 'LLMMapGenerator', and 'Prompt2Map' play pivotal roles. 'SQLGeoRetriever' handles the translation of natural language queries into SQL statements and manages the retrieval of geospatial data from the database. 'LLMPrompt2SQL' utilizes the capabilities of LLMs to accurately convert user prompts into executable SQL queries, ensuring that the data retrieval process aligns with the user's intent. 'LLMMapGenerator' is tasked with transforming the retrieved geospatial data into visually coherent and interactive maps, selecting appropriate visualization techniques based on the nature of the data and the user's requirements. Finally, the 'Prompt2Map' class serves as the primary interface for users, integrating the functionalities of the retriever and generator to deliver a cohesive mapping experience. This structured separation of responsibilities within the application component ensures that each aspect of the system operates efficiently and cohesively.

2.2.2 Interfaces

The interfaces component serves as the foundational layer of the prompt2map package, defining the abstract contracts that dictate how different modules interact with one another. By establishing clear and standardized interfaces, the system ensures that each component adheres to a consistent set of functionalities, promoting interoperability and flexibility. This abstraction is crucial for maintaining a decoupled architecture, allowing individual modules to be developed, tested, and maintained independently without impacting the overall system integrity.

Key interfaces within this component include 'Embedding', 'LLM', 'GeoDatabase', 'GeoRetriever', 'Prompt2SQL', and 'MapGenerator'. The 'Embedding' interface outlines the methods required for generating numerical representations of textual data, which are essential for tasks like similarity matching and query understanding. The 'LLM' interface defines the interaction methods with LLMs, facilitating tasks such as prompt-based query generation and response handling. 'GeoDatabase' encapsulates the functionalities needed to interact with geospatial databases, including schema retrieval and geospatial data querying. 'GeoRetriever' specifies the methods for fetching geospatial data based on user queries, ensuring that the application can retrieve relevant and accurate datasets. 'Prompt2SQL' bridges the gap between natural language inputs and structured SQL queries, enabling the seamless translation of user intents into executable database commands. Lastly, the 'MapGenerator' interface outlines the

methods for creating interactive maps from geospatial data, ensuring that visualizations are generated consistently and effectively. By defining these interfaces, the system promotes a high level of abstraction and modularity, allowing for easy integration of new functionalities and technologies as the system evolves.

2.2.3 Providers

The providers component encapsulates the concrete implementations of the abstract interfaces defined within the interfaces module. This layer is responsible for managing interactions with external services and tools, effectively bridging the gap between the system's core functionalities and the underlying technologies that enable them. By centralizing these implementations within the providers, the system achieves a high degree of flexibility and adaptability, allowing for seamless integration of new services or replacement of existing ones without disrupting the core application logic.

Within the providers module, classes such as 'GeoDuckDB' and 'OpenAIProvider' play critical roles. 'GeoDuckDB' implements the 'GeoDatabase' interface, managing all interactions with DuckDB, the chosen geospatial database engine. This includes handling data storage, executing spatial queries, and managing database connections. DuckDB's in-process architecture and robust support for various geospatial data formats make it an ideal choice for efficient data retrieval and management within Prompt2Map. On the other hand, 'OpenAIProvider' implements both the 'LLM' and 'Embedding' interfaces, facilitating interactions with the OpenAI API. This provider manages the communication with LLMs, enabling the system to leverage advanced natural language processing capabilities for tasks such as text-to-SQL translation and function calling. By abstracting these external dependencies into dedicated provider classes, the system ensures that changes or updates to external services can be accommodated with minimal impact on the overall architecture. Additionally, this separation of concerns enhances the maintainability and scalability of the system, allowing developers to focus on core functionalities without being bogged down by the complexities of external service integrations.

2.2.4 External Dependencies

The prompt2map package leverages several external dependencies to achieve its functionality, each chosen for its robustness, performance, and compatibility with the system's requirements.

2.2.4.1 OpenAI API

The OpenAI API is integrated into Prompt2Map to leverage the advanced natural language processing capabilities of LLMs like GPT-4. This API facilitates the system's

ability to interpret and translate natural language queries into executable SQL statements through the text-to-SQL functionality. By utilizing the OpenAI API, Prompt2Map can effectively handle complex and nuanced user inputs, ensuring accurate data retrieval and meaningful map generation. The API provides access to powerful language models that excel in understanding context, disambiguating meanings, and generating coherent and contextually relevant responses. Additionally, the OpenAI API supports function calling, allowing Prompt2Map to invoke predefined mapping functions with specific parameters based on user queries. This integration enhances the system's ability to maintain consistency and standardization in map generation, as the predefined functions ensure that all maps adhere to established cartographic principles and visualization standards. Moreover, the use of the OpenAI API abstracts the complexities of managing and deploying LLMs, enabling Prompt2Map to focus on delivering a seamless and user-friendly GIS experience without the overhead of maintaining sophisticated NLP infrastructure.

2.2.4.2 DuckDB

DuckDB is employed as the primary geospatial database engine within Prompt2Map, selected for its high performance, in-process architecture (Raasveldt & Mühleisen, 2019), and comprehensive support for geospatial data types and functions. Unlike traditional SQL databases that require separate server infrastructure, DuckDB operates entirely within the application's process, eliminating the need for additional setup and reducing latency in data retrieval. Its geospatial extension provides robust capabilities for handling spatial queries, including spatial joins, proximity searches, and geographic aggregations, which are essential for generating accurate and informative maps. DuckDB's ability to efficiently manage large datasets and perform complex queries makes it an ideal choice for Prompt2Map, ensuring swift data processing and reliable performance. Furthermore, DuckDB"s compatibility with various geospatial data formats, such as GeoJSON, Shapefiles, and GeoParquet, enhances the system's flexibility and ease of data integration, allowing Prompt2Map to seamlessly incorporate diverse geospatial datasets without compromising on performance or functionality.

2.3 Data

The Prompt2Map system is designed to work with various structured spatial data sources. For this thesis, we use the Portuguese 2021 Census Data. This dataset includes demographic, socioeconomic and housing information structured in a single-table format, suitable for testing basic SQL queries and simple map visualizations. The data attributes encompass population counts (total population, age groups), socioeconomic indicators (employment rates, education levels, household income), and geographic units (administrative boundaries at different levels, such as municipalities and regions).

The data is obtained from the Portuguese National Statistics Institute (INE) (Instituto Nacional de Estatística (INE), 2021).

2.4 Evaluation

Evaluating the performance of Prompt2Map poses several challenges due to the system's unique characteristics, including the integration of natural language processing, geospatial data retrieval, and map generation functionalities. Traditional evaluation metrics used in NLP, such as BLEU scores or F1 scores, may not fully capture the system's effectiveness in converting natural language queries into interactive web maps. Therefore, a comprehensive evaluation strategy is required to assess the system's performance across different dimensions, including accuracy, efficiency, usability, and scalability.

Nevertheless, for the scope of this thesis, the evaluation focuses on the text-to-SQL translation component of Prompt2Map, which corresponds to the Geospatial Retriever in the system architecture. This evaluation aims to assess the system's ability to accurately interpret natural language queries, generate SQL statements that retrieve relevant geospatial data, and execute these queries against the geospatial database. The evaluation of the Map Generator component, which involves the creation of interactive web maps based on the retrieved geospatial data, is more subjective and user-dependent, making it challenging to define objective metrics. Therefore, the focus of this evaluation is on the Geospatial Retriever, which forms the foundation of the system's functionality.

2.4.1 Metrics

2.4.1.1 The problem with text-to-SQL metrics

Can a question be represented as a SQL query? This is the main question that text-to-SQL systems aim to answer. However, evaluating the performance of these systems is challenging due to the inherent ambiguity and complexity of natural language queries. Even for human annotators, determining the correct SQL query for a given question can be a non-trivial task, as it requires an understanding of the underlying database schema, the semantics of the question, and the intended query result. Therefore, developing robust evaluation metrics that can accurately assess the performance of text-to-SQL systems is crucial for advancing the field and enabling fair comparisons between different approaches.

There are multiple benchmarks for text-to-SQL evaluation, such as Spider (Yu et al., 2018), BIRD (J. Li et al., 2023) and WikiSQL (Zhong et al., 2017). These benchmarks provide standardized datasets with natural language questions paired with their corresponding SQL queries, allowing researchers to evaluate the performance of their systems on common tasks. The two main evaluation metrics for the text-to-SQL task are Exact Matching (EM) and Execution Accuracy (EX), introduced by Spider. EM measures

the syntax-level equivalence between the generated SQL query and the ground truth query, while EX evaluates the execution results of the generated query.

$$\text{EM} = \frac{\text{Number of exact matches}}{\text{Total number of queries}} \tag{2.1}$$

$$\text{EA} = \frac{\text{Number of correct executions}}{\text{Total number of queries}} \tag{2.2}$$

However, those metrics have limitations that can lead to biased results and inaccurate performance assessments (Kim et al., 2024). While EM evaluates the syntax-level equivalence of queries, it often leads to high false negative rates, as the same logical intent can be represented through various query structures and alises. On the other hand, EX focuses on the execution results of the generated queries, providing a more robust evaluation metric. However, EX is not without its limitations, as it can produce false positives when incorrect queries yield correct results, and false negatives when extra fields are included in the output.

2.4.1.2 Soft F1-Score

The hardness of EM and EX is that both require exact matches between the generated SQL query and the ground truth. This is a very strict requirement, as even small variations in the generated query can lead to a mismatch. To address this issue, the Soft F1-score was introduced as a more lenient evaluation metric that considers the data similarity between the generated and ground truth queries.

The original F1-score is a metric commonly used in classification tasks to balance the model's ability to correctly identify positive samples (precision) and to capture all positive samples (recall).

Precision is the proportion of true positive samples out of the total number of samples predicted as positive. It evaluates the model's accuracy in identifying positive samples (Equation 2.3). Recall is the proportion of true positive samples out of the total number of positive samples in the dataset. It measures the model's ability to capture all positive samples (Equation 2.4). The F1-score is the harmonic mean of precision and recall, providing a balanced evaluation metric that considers both aspects of the model's performance (Equation 2.5).

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{2.3}$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{2.4}$$

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{2.5}$$

In the context of text-to-SQL evaluation, the definition of what is a true/false positive/negative is what differences the metrics. EM considers a true positive when

the generated query is exactly the same as the ground truth. EX considers a true positive when the generated query returns the exact same results as the ground truth. The Soft F1-score relaxes the requirement of exact matches, considering the similarity between the generated and ground truth queries, as shown in Figure 2.2.

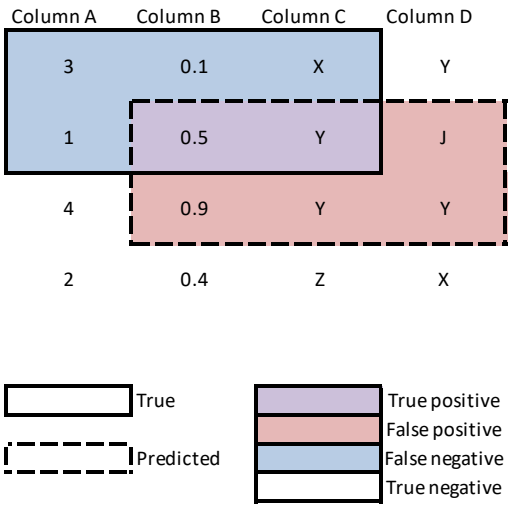


Figure 2.2: True and False Positives and Negatives in SQL evaluation for Soft F1-score. The solid border represents the ground truth, while the dashed border represents the generated query.

2.4.1.3 Consistency

To assess the consistency of the Prompt2Map outputs across multiple runs, we introduced a metric termed Consistency Entropy. The entropy (Equation 2.6) is the amount of bits required to encode the information in a message, and it is a measure of uncertainty. The normalized entropy (Equation 2.7) is the entropy divided by the maximum possible entropy, which is the logarithm of the number of possible outputs. This normalization ensures that the entropy value is between 0 and 1, making it comparable across different scenarios. The Consistency Entropy (Equation 2.8) is defined as 1 − normalized entropy, where a higher score indicates higher consistency.

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i) \tag{2.6}$$

$$H_{rel}(X) = \frac{H(X)}{\log_2 n} \tag{2.7}$$

$$\text{Consistency Entropy} = 1 - H_{rel}(X) \tag{2.8}$$

For example, consider two different output distributions for a given prompt. If the system consistently generates the same output "AAAAA"across all runs, the entropy of this distribution is zero, indicating no uncertainty, and the consistency score is 1,

reflecting perfect consistency. Conversely, if the system produces a uniform distribution of outputs "A", "B", "C", "D", and "E"with equal probability, the entropy reaches its maximum value, indicating high uncertainty, and the consistency score is minimized, reflecting low consistency.

This might lead to cases that are unintuitive. For instance, the normalized entropy of AAAAABBBBB is the same as ABCDEFGHIJ (both are 1), but the second one has more diversity. This is because the normalized entropy is compares the distribution of the outputs, not the outputs themselves. Both ouputs are uniformly distributed, there for the same normalized entropy (the maximum value).

As diversity is also a relevant aspect of the system's performance, we introduce a second metric, the relative mode frequency (Equation 2.9). This metric provides a simple yet effective measure of the most common output generated by the system across multiple runs. A high mode frequency indicates that the system consistently produces a specific output, which can be beneficial for tasks requiring deterministic results. By combining Consistency Entropy and mode frequency, the evaluation framework captures both the diversity and stability of the system's outputs, offering a comprehensive assessment of its performance.

$$\text{Mode Frequency (X)} = \frac{\text{Number of occurrences of the mode}}{\text{Total number of runs}} = \frac{\max_{x \in X} \text{count}(x)}{|X|} \tag{2.9}$$

This metric penalizes having a high diversity of outputs. For example, the relative mode frequency of AAAAABBBBB is $\frac{1}{2}$, while the relative mode frequency of ABCDEFGHIJ is 0.1. This metric complements the Consistency Entropy by providing a measure of the most common output generated by the system, offering insights into the system's deterministic behavior.

2.4.2 Evaluation Questions

The metrics presented in the previous section are used to evaluate the performance of the Prompt2Map system across a series of evaluation questions. These questions are designed to test the system's ability to accurately interpret natural language queries, generate SQL statements that retrieve relevant geospatial data, and execute these queries against the geospatial database. By evaluating the system's performance on a diverse set of queries, we can assess its effectiveness in handling various types of user inputs and producing accurate and informative maps.

Each question is accompanied by the expected SQL query, which serves as the ground truth for evaluating the system's performance. The evaluation questions are designed to test the system's ability to handle different types of queries, including aggregation, filtering, and spatial analysis tasks. By evaluating the system's performance on these questions, we can gain insights into its strengths and limitations, identify areas for improvement, and guide future development efforts.

Table 2.1: Evaluation questions for the Prompt2Map system.

ID	Question	Expected SQL
Q1	Which Freguesias have the highest percentage of buildings constructed before 1945?	`SELECT c.Freguesia, (c.N_EDIFICIOS_CONSTR_ANTES_1919 + c.N_EDIFICIOS_CONSTR_1919A1945)/ (c.N_EDIFICIOS_CLASSICOS)* 100 AS percentage, c.geometry FROM censo2021pt.censo2021_freguesia c ORDER BY percentage DESC;`
Q2	How does the absolute distribution of one- or two-floor buildings vary across Municipios?	`SELECT c.Municipio, SUM(c.N_EDIFICIOS_1OU2_PISOS) AS building_count, ST_Union(c.geometry)AS geometry FROM censo2021pt.censo2021_freguesia c GROUP BY c.Municipio;`
Q3	In which Distritos is the absolute number of single-family or two-family homes highest?	`SELECT c.Distrito, SUM(c.N_EDIFICIOS_CLASSICOS_1OU2_ALOJ )AS single_family_homes, ST_Union(c.geometry)AS geometry FROM censo2021pt.censo2021_freguesia c GROUP BY c.Distrito ORDER BY single_family_homes DESC;`
Q4	What is the ratio of rented housing to owned housing in different Freguesias?	`SELECT c.Freguesia, (c.N_RHABITUAL_ARRENDADOS)/ (c.N_RHABITUAL_PROP_OCUP)AS rent_own_ratio, c.geometry FROM censo2021pt.censo2021_freguesia c ORDER BY rent_own_ratio DESC;`
Q5	What percentage of buildings are constructed with three or more apartments in each Freguesia?	`SELECT c.Freguesia, (c.N_EDIFICIOS_CLASSICOS_3OUMAIS_ALOJ / c.N_EDIFICIOS_CLASSICOS)* 100 AS multi_family_density, c.geometry FROM censo2021pt.censo2021_freguesia c ORDER BY multi_family_density DESC;`
Q6	Which Municipios have the highest proportion of individuals aged 65 or older?	`SELECT c.Municipio, SUM(c.N_INDIVIDUOS_65_OU_MAIS)/ SUM(c.N_INDIVIDUOS)* 100 AS age_65_plus_percentage , ST_Union(c.geometry)AS geometry FROM censo2021pt.censo2021_freguesia c GROUP BY c.Municipio ORDER BY age_65_plus_percentage DESC;`
Q7	Which Freguesias have the highest male-to-female ratio in the population?	`SELECT c.Freguesia, (c.N_INDIVIDUOS_H / NULLIF(c.N_INDIVIDUOS_M, 0))AS male_to_female_ratio, c.geometry FROM censo2021pt.censo2021_freguesia c ORDER BY male_to_female_ratio DESC;`
Q8	Where are parking-accessible residential accommodations more common across Municipios relative to all residences?	`SELECT c.Municipio, SUM(c.N_RHABITUAL_COM_ESTACIONAMENTO )/ SUM(c.N_CLASSICOS_RES_HABITUAL)* 100 AS parking_accessible, ST_Union(c.geometry)AS geometry FROM censo2021pt.censo2021_freguesia c GROUP BY c.Municipio ORDER BY parking_accessible DESC;`
Q9	What is the percentage of individuals with no formal education across Distritos?	`SELECT c.Distrito, SUM(c.N_INDIVIDUO_ENSINCOMP_NENHUM) / SUM(c.N_INDIVIDUOS)* 100 AS no_education_percentage , ST_Union(c.geometry)AS geometry FROM censo2021pt.censo2021_freguesia c GROUP BY c.Distrito ORDER BY no_education_percentage DESC;`
Q10	In which Freguesias is the unemployment rate highest?	`SELECT c.Freguesia, ((c.N_INDIVIDUOS_DESEMPREGADOS_1EMP + c.N_INDIVIDUOS_DESEMPREGADOS_NOVOEMP)/ NULLIF(c.N_INDIVIDUOS_COM_ATIVIDADE_ECONOMICA, 0))* 100 AS unemployment_rate, c.geometry FROM censo2021pt.censo2021_freguesia c ORDER BY unemployment_rate DESC;`

2.4.3 Aggregation

All the questions were tested 10 times using the prompt2map package. For each test, the system generated a SQL query based on the natural language question, which was used to compute the previously described metrics. The 10 results were aggregated as

it follows:

- Completion rate: percentage of successful completions (i.e., the system generated a valid SQL query that executed without errors) out of the total number of runs.
- Macro Precision: the average precision across all runs.
- Macro Recall: the average recall across all runs.
- Macro Soft F1-score: the average Soft F1-score across all runs.
- Frequency of the mode: the frequency of the most common output generated by the system across all runs.
- Consistency Entropy: the normalized entropy of the model's output distribution across all runs.

3

Results

3.1 Introduction

This chapter presents the results of the system's performance in generating SQL queries based on natural language questions. The evaluation focuses on key metrics that provide a comprehensive view of the accuracy, consistency, and reliability in translating user queries into executable SQL statements. The analysis highlights both the strengths and areas needing improvement, offering insights into the system's capabilities and limitations.

3.2 Summary of Results

The system was evaluated on ten different questions, each designed to test various aspects of SQL query generation. Table 3.1 summarizes the performance metrics for each question.

Table 3.1: Performance metrics for each question.

Question ID	Completion Rate	Macro Precision	Macro recall	Macro Soft F1-score	Frequency of the mode	Consistency Entropy
Q1	100%	0.8202	0.9907	0.8858	0.4000	0.1390
Q2	100%	0.9988	0.9988	0.9988	1.0000	1.0000
Q3	100%	0.9667	0.8722	0.8772	0.9000	0.5310
Q4	100%	0.9316	0.9165	0.9187	0.9000	0.5310
Q5	100%	0.9672	0.9081	0.9367	1.0000	1.0000
Q6	100%	0.9588	0.9988	0.9738	0.9000	0.5310
Q7	100%	0.7737	1.0000	0.8516	0.5000	0.0628
Q8	100%	0.9988	0.9988	0.9988	1.0000	1.0000
Q9	100%	0.9630	0.9630	0.9630	1.0000	1.0000
Q10	90%	0.8831	0.7537	0.7680	0.5556	0.1472

As shown in Table 3.1, the system generally performed well, with most questions achieving a 100% completion rate and high macro precision and recall scores. However, certain questions exhibited lower scores in specific metrics, indicating areas where the system faced challenges.

3.3 Scatter Matrix

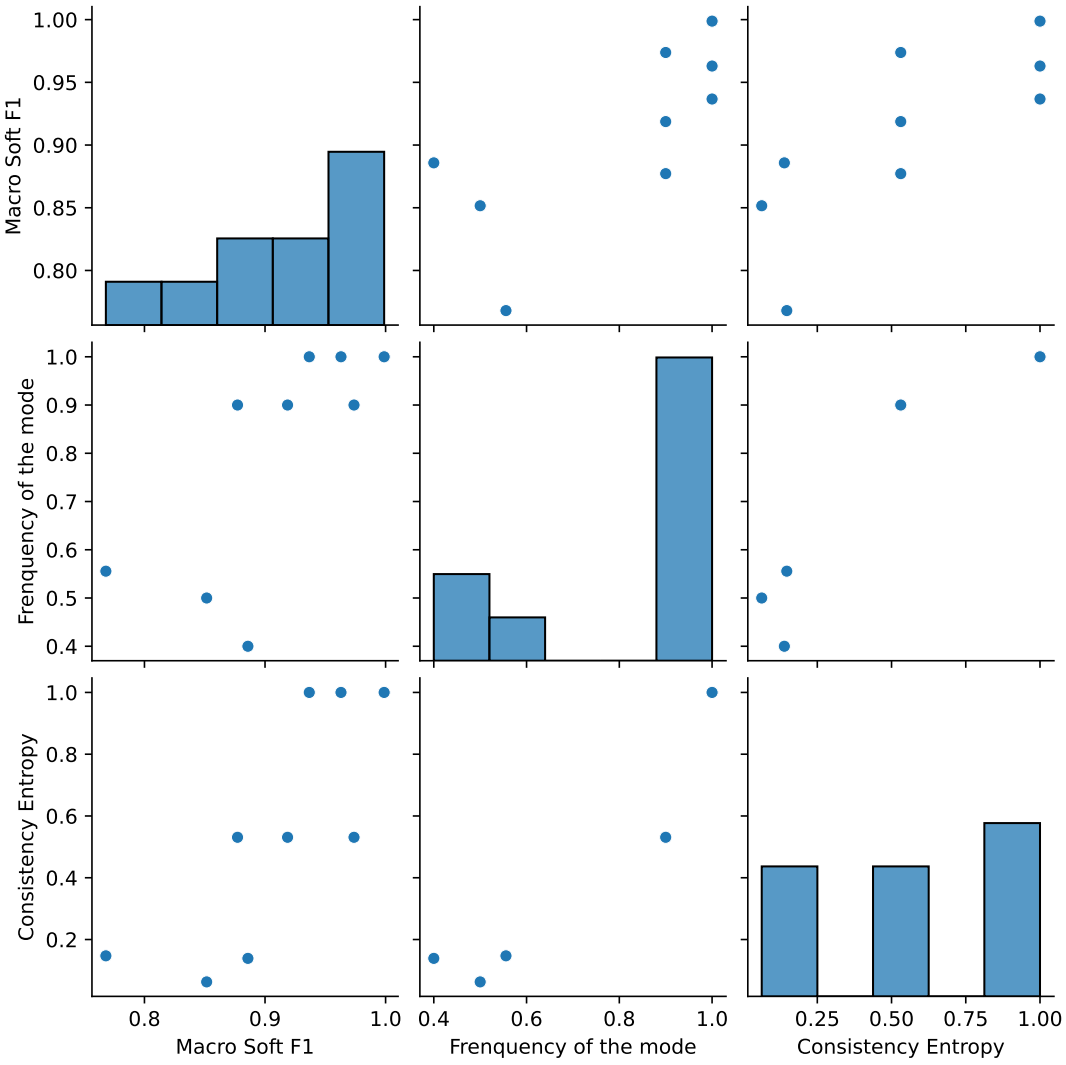


Figure 3.1: Scatter matrix of evaluation metrics. Each point represents a question.

The scatter matrix in Figure 3.1 provides a clear visualization of the relationships among the performance metrics, including macro Soft F1-score, frequency of the mode, and consistency entropy. These metrics are positively correlated, indicating that better performance in one area is often associated with better performance in the others. For instance, the macro Soft F1-score correlates strongly with the frequency of the mode, suggesting that when the system achieves a good balance between precision and recall, it also produces more consistent outputs. Similarly, the macro Soft F1-score correlates with consistency entropy, meaning that higher accuracy is associated with more predictable outputs. The frequency of the mode also correlates with consistency entropy, reinforcing the observation that a higher proportion of consistent outputs is linked to greater predictability.

In Figure 3.1, these relationships are reflected in the positioning of data points. Questions with high performance metrics are concentrated in the upper-right corner of the graph, showing strong alignment between accuracy, consistency, and predictability. These include Q2, Q5, Q8, and Q9, which achieved high macro Soft F1-scores and consistency metrics.

In contrast, questions with lower performance metrics are scattered in the lower-left region of the graph, reflecting reduced consistency and accuracy. Examples include Q1, Q7, and Q10, which had lower macro Soft F1-scores and consistency metrics, indicating areas where the system struggled to generate accurate and consistent SQL queries.

3.4 Detailed Examination of Questions

The performance of the system can be further understood by analyzing the characteristics of both high-performing and low-performing questions. These two groups show distinct differences in terms of language clarity, SQL complexity, and data mapping requirements, which significantly influenced the system's ability to generate accurate and consistent outputs.

3.4.1 High-Performing Questions

High-performing questions, such as Q2, Q5, Q8, and Q9, demonstrated strong performance across all evaluated metrics. These questions achieved a 100% completion rate, high macro precision and recall scores, and perfect or near-perfect macro Soft F1-scores. Additionally, they exhibited high frequency of the mode and consistency entropy values, indicating that the system consistently generated similar SQL queries for these prompts.

Several factors contributed to the strong performance of these questions. Firstly, the prompts were phrased clearly and unambiguously, reducing the likelihood of misinterpretation. For instance, Q2 ("How does the absolute distribution of one- or two-floor buildings vary across Municipios?") explicitly specifies the variable of interest and the geographical aggregation level. Secondly, the required SQL queries for these questions involved standard aggregation functions and joins that are commonly encountered, such as `SUM` and `GROUP BY` clauses. This simplicity facilitated the system's ability to map the natural language to SQL syntax accurately. Lastly, the attributes mentioned in the questions corresponded directly to column names in the database, minimizing the need for complex transformations or reasoning. For example, in Q8, terms like "parking-accessible residential accommodations" directly map to database columns such as N_RHABITUAL_COM_ESTACIONAMENTO.

3.4.2 Low-Performing Questions

In contrast, low-performing questions such as Q1, Q7, and Q10 exhibited lower scores in macro Soft F1-score, frequency of the mode, and consistency entropy. Q10, in particular, had the lowest macro Soft F1-score of 0.7680 and a completion rate of 90%, indicating that the system sometimes failed to generate a valid SQL query for this prompt.

Several challenges contributed to the lower performance of these questions. The prompts contained complex phrasing or multiple clauses, increasing the difficulty for the system to parse and interpret them correctly. For example, Q1 ("Which Freguesias have the highest percentage of buildings constructed before 1945?") requires calculating a percentage based on multiple columns, which adds complexity. Additionally, some questions implied certain computations or data manipulations that were not explicitly stated. In Q7 ("Which Freguesias have the highest male-to-female ratio in the population?"), the system needed to recognize the requirement to handle potential division by zero, as some Freguesias might have zero females, necessitating the use of `NULLIF` in the SQL query.

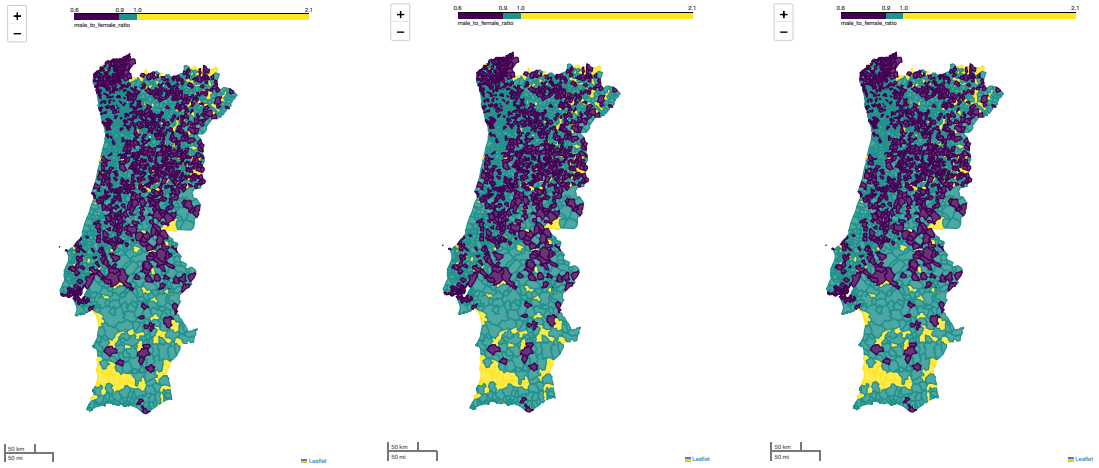
Furthermore, low-performing questions often required the use of advanced SQL functions or error handling mechanisms that the system was less adept at generating. In Q10, the calculation of the unemployment rate involves adding two columns and dividing by another, with a potential division by zero, which complicates the query. The attributes mentioned in the questions did not always have a direct one-to-one mapping with database columns, requiring the system to infer or transform terms. This increased the likelihood of errors in the generated SQL.

3.5 Visual Analysis of Generated Maps

To further evaluate the system's performance, Figure 3.2 presents the generated maps for the two lowest-performing questions based on the macro Soft F1-score: Q7 and Q10. Visual inspection of these maps provides additional insights into the system's capabilities and limitations.

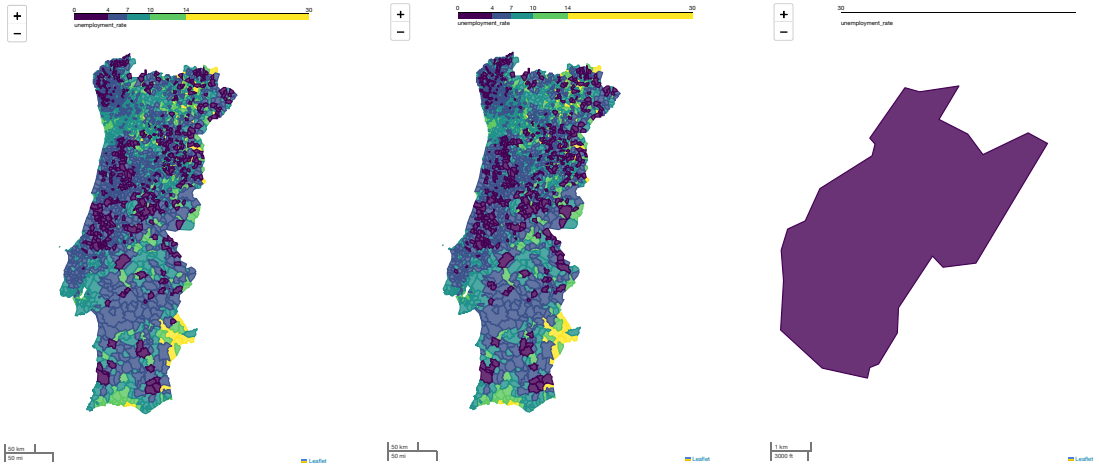
3.5.1 Analysis of Q7 Map

For Q7, both generated maps are identical to the ground truth map, despite the low macro Soft F1-score. In this case, the recall is very high, indicating that the generated query is retrieving more data that it needs at the cost of losing precision. The ground truth is a subset of the generated datasets, but the extra columns are penalized, leading to a lower precision score. This discrepancy highlights a limitation of the macro Soft F1-score metric, as it penalizes the model for false positives (extra columns) even when the generated maps accurately reflect the ground truth. The high recall suggests that the system is capable of capturing all relevant data points, but the metric's sensitivity to precision results in a lower overall score.



(a) Ground Truth (b) Generated Map 1 (c) Generated Map 2

(d) Comparison for Question 7: Which Freguesias have the highest male-to-female ratio in the population?



(e) Ground Truth (f) Generated Map 1 (g) Generated Map 2

(h) Comparison for Question 10: In which Freguesias is the unemployment rate highest?

Figure 3.2: Comparison of ground truth and generated maps by prompt2map for questions Q7 and Q10. Each row corresponds to a single question.

3.5.2 Analysis of Q10 Map

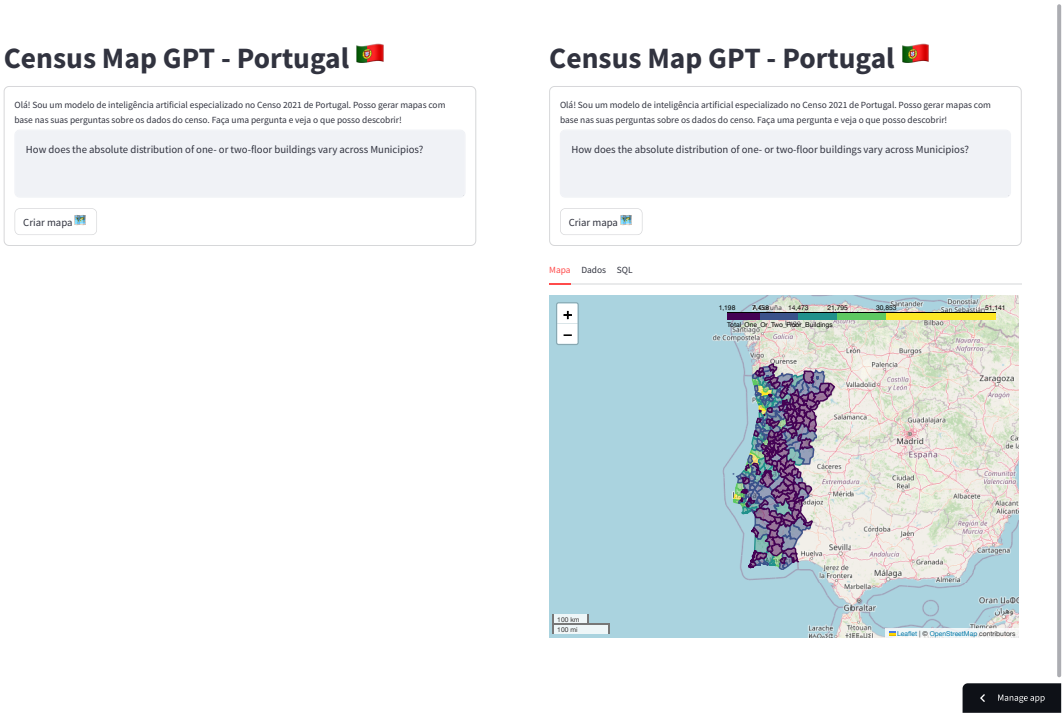
For Q10, Generated Map 1 might seem identical to the ground truth map at first glance, but it actually contains fewer rows. This discrepancy arises because the underlying SQL query is incorrect. The query groups the data by Freguesia name alone, but there are multiple Freguesias with the same name across different Municipios. Therefore, the correct approach would be to group by both Freguesia and Municipio to ensure accurate aggregation. On the other hand, Generated Map 2 displays a single feature, which we could assume represents the Freguesia with the highest unemployment rate. This result could also answer the original question, indicating potential ambiguity in the prompt that leads to inconsistent outputs. The system's interpretation of the question might vary, causing it to generate different results based on its understanding of the query requirements.

3.6 Web Application

The Prompt2Map system includes a user-friendly web application built using Streamlit, named Census Map GPT. This application allows users to generate maps based on census data through natural language prompts. The frontend captures the user's intent and translates it into SQL queries, which are then used to retrieve and visualize the relevant data.

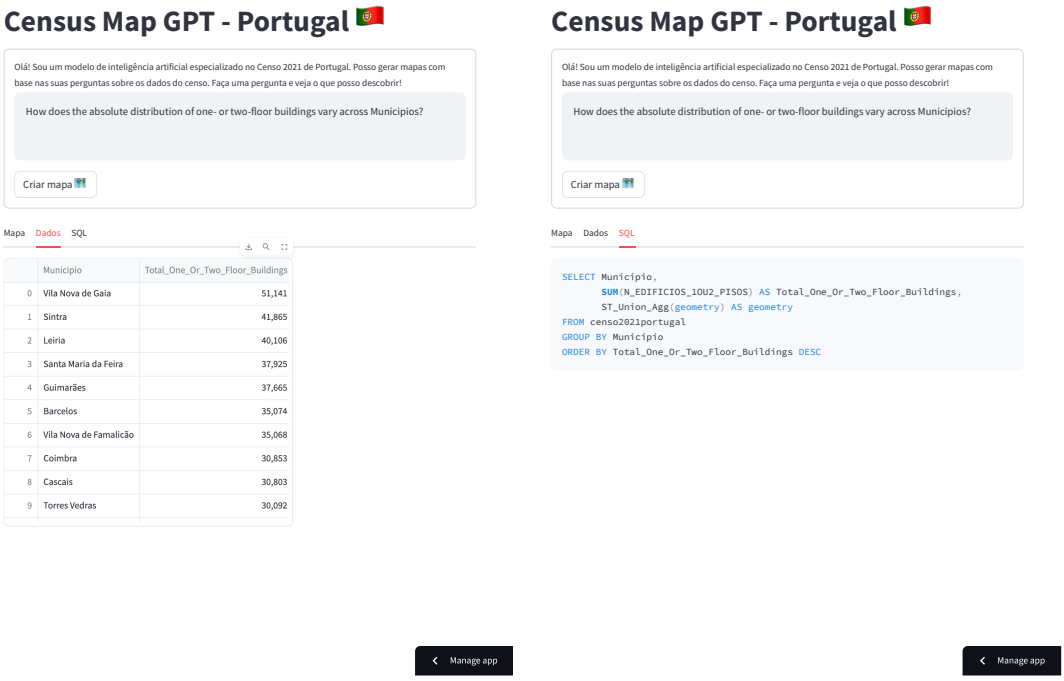
Figure 3.3 illustrates the interface of the Census Map GPT application. Subfigure 3.3(a) shows the prompt input area where users can type their queries. Subfigure 3.3(b) displays the generated map based on the user's prompt. Subfigure 3.3(c) presents the data retrieved from the database, and subfigure 3.3(d) shows the SQL query generated from the user's prompt.

The web application enhances accessibility by allowing users without technical expertise in SQL or data visualization tools to interact with complex datasets effectively. By leveraging natural language processing, the system simplifies the process of data retrieval and visualization, making it more intuitive and efficient.



(a) Prompt to generate a map

(b) Generated map



(c) Generated data

(d) Generated SQL

Figure 3.3: Census Map GPT UI for Prompt2Map

4

Discussion

4.1 Interpretation of Results

The evaluation of Prompt2Map demonstrates its capability to effectively translate natural language prompts into SQL queries for generating accurate and reproducible geospatial maps. The system achieved high completion rates and macro precision across most test cases, reflecting its reliability in handling well-defined prompts. The scatter matrix analysis revealed positive correlations among metrics such as macro F1-score, frequency of the mode, and consistency entropy. These correlations emphasize that consistent and accurate outputs often go hand-in-hand, with better performance in one area reinforcing reliability across others.

High-performing questions exemplified how clear language, direct schema mappings, and standard SQL constructs enable robust query generation. Conversely, low-performing questions exposed the system's limitations in handling ambiguous prompts or advanced SQL requirements, such as division by zero or multi-column computations. These challenges underscore the trade-offs between simplicity and versatility in natural language interfaces for GIS.

4.2 Comparison with Existing Work

4.2.1 Implementation Status of Existing Solutions

Prompt2Map exists within a broader context of natural language GIS tools, each with varying degrees of implementation and readiness. Systems like GeoGPT (Kim et al., 2024), GeoQAMap (Feng et al., 2023), and Aino[1] represent different approaches to integrating LLMs with geospatial systems. GeoGPT is a research-oriented framework designed to handle complex workflows and automate GIS tasks, serving as a valuable proof of concept but lacking commercialization. GeoQAMap operates as an experimental system capable of generating SPARQL queries and visualizing results based on geospatial knowledge bases like Wikidata. In contrast, Aino is a fully implemented and commercially available platform for geospatial data analysis and visualization.

[1] https://aino.world/

4.2.2 Comparison to Prompt2Map

Prompt2Map shares similarities with these systems but also has unique strengths and limitations. Like GeoQAMap and Aino, it uses natural language as the primary interface, offering accessibility to non-expert users. Prompt2Map automates the mapping process in a manner similar to GeoQAMap's SPARQL generation and GeoGPT's GIS workflow automation, but it is more specialized, focusing specifically on generating maps rather than handling a broad spectrum of geospatial operations.

In terms of implementation, Prompt2Map is a functional, open-source Python package available on GitHub, making it more accessible than research prototypes like GeoGPT and GeoQAMap. However, it lacks the commercial polish and collaborative features of Aino, which positions it as a business-ready product. Additionally, Prompt2Map allows tailored map generation with specific parameters, aligning with the extensibility goals of GeoGPT and Aino. Yet, it does not yet match GeoGPT's integration with professional GIS tools and geoprocessing.

Prompt2Map occupies a promising middle ground between research prototypes and commercial solutions. With further development, it has the potential to bridge this gap by integrating external datasets, such as OpenStreetMap, or connecting to web-based GIS platforms. This would enhance its analytical capabilities and broaden its appeal to a wider audience.

4.2.3 Addressing Ethical Challenges

Prompt2Map effectively addresses several ethical concerns identified in the literature. Misleading information, a key risk in AI-generated maps (Zhang et al., 2023), can occur if SQL queries are incorrect or poorly generated. By incorporating rigorous testing and evaluation metrics, such as the macro F1-score, Prompt2Map reduces the likelihood of such errors. Unanticipated features, often indicative of low precision, are similarly mitigated through robust evaluation and continuous system improvements.

Reproducibility, a significant challenge for many generative systems, is a core strength of Prompt2Map. Each generated map is traceable to a specific SQL query and dataset, ensuring outputs are transparent and replicable. This aligns with calls for reproducibility in geospatial research (Zhang et al., 2023), further supporting the system's reliability. While Prompt2Map does not process sensitive individual-level data, it remains crucial to maintain robust data privacy measures, particularly if the system integrates more complex datasets in the future.

4.3 Implications for GIS and NLP

The integration of natural language interfaces into GIS systems, as exemplified by Prompt2Map, offers transformative potential. By lowering the barrier to entry for non-expert users, these systems democratize access to spatial analysis and visualization

tools. This aligns with the broader vision of self-service GIS systems (Rowland et al., 2020), and contributes to a growing body of work exploring the intersection of GIS and NLP technologies.

Prompt2Map's approach demonstrates how SQL-based systems can serve as a foundation for ethical and reproducible GIS tools. The potential to integrate Linked Data presents an exciting avenue for future work, enabling richer data interoperability and expanding the system's analytical capabilities. Such enhancements could position Prompt2Map as a key tool for advancing equity and accessibility in geospatial science (S. Wang et al., 2024). Moreover, the system's simplicity and usability make it well-suited for educational applications, urban planning, and public information dissemination.

5

Conclusions

This thesis set out to address the challenges of making GIS more accessible to non-expert users by leveraging advancements in AI, specifically LLMs. The primary objective was to develop Prompt2Map, a system that allows users to interact with geospatial data and generate web maps through natural language queries. By integrating LLMs with RAG, Prompt2Map bridges the gap between traditional GIS workflows and intuitive, user-friendly interfaces. The main contributions of this research include the design and implementation of Prompt2Map, showcasing its capability to interpret natural language, retrieve relevant geospatial data, and generate accurate, reproducible maps. The system's development involved addressing technical challenges, such as natural language understanding, SQL generation, and geospatial visualization.

5.1 Contributions

This research makes significant contributions to both academic and practical domains. Academically, it advances the intersection of AI and GIS, demonstrating how LLMs can be effectively utilized to interpret spatial queries and automate map generation. The integration of RAG techniques further enriches the field, offering a framework for retrieving and utilizing real-time geospatial data in natural language-driven systems.

Practically, Prompt2Map underscores the importance of making geospatial tools accessible to a wider audience. By simplifying interactions with spatial data, the system empowers policymakers, educators, and the general public to engage with geospatial information for decision-making, education, and community initiatives. Performance tests with synthetic prompts validated the system's ability to handle diverse user queries reliably, providing a robust foundation for its applicability in real-world scenarios.

These findings highlight the transformative potential of natural language interfaces to bridge the gap between map producers—GIS experts—and map consumers, such as non-specialist users. By fostering more intuitive interactions with geospatial data, Prompt2Map paves the way for more inclusive and equitable access to spatial insights. In the long term, it has the potential to influence the design of future geospatial technologies, contributing to both research and practice.

5.2 Limitations

While this research demonstrates the viability of natural language-driven GIS tools, certain limitations remain. The use of synthetic prompts for evaluation, while effective for initial testing, may not fully capture the diversity of real-world user queries. The system's reliance on LLMs introduces challenges related to data quality, ambiguity in language interpretation, and scalability when handling complex datasets or queries.

Additionally, Prompt2Map's performance depends on the quality and structure of the underlying geospatial data. Issues such as outdated information or inconsistencies in data formats could impact the accuracy and reliability of the generated maps. Addressing these limitations is essential for enhancing the system's robustness and applicability in diverse contexts.

5.3 Future Directions

To improve Prompt2Map, several enhancements are recommended. First, incorporating real-time data retrieval capabilities would enable the system to generate maps reflecting current conditions, such as traffic patterns or weather events. Second, expanding the system's natural language understanding capabilities to handle more complex queries and ambiguous language constructs would improve user experience. Third, integrating advanced features, such as multi-layered map visualizations and interactive data exploration tools, would cater to a broader range of user needs.

Future research could explore the integration of Prompt2Map with external geospatial platforms like OpenStreetMap or ArcGIS, enabling richer analyses and data interoperability. Conducting user studies would provide insights into the system's usability and inform further refinements. Additionally, addressing ethical challenges, such as ensuring fairness and mitigating biases in LLM-generated outputs, remains a critical area for ongoing investigation.

5.4 Final Thoughts

Prompt2Map represents a significant step toward bridging the gap between map consumers and producers, making geospatial information more accessible and actionable. By enabling users to interact with GIS through natural language, the system lowers barriers to entry and fosters greater engagement with spatial data.

The aspirations for Prompt2Map extend beyond its current capabilities. By continuing to innovate and address challenges, this research hopes to inspire further exploration at the intersection of AI and GIS. Ultimately, Prompt2Map exemplifies the transformative potential of bringing spatial information closer to the public, empowering individuals and communities to make informed decisions based on geographic insights.

Bibliography

Fan, W., Ding, Y., Ning, L., Wang, S., Li, H., Yin, D., Chua, T.-S., & Li, Q. (2024). A Survey on RAG Meeting LLMs: Towards Retrieval-Augmented Large Language Models. *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 6491–6501. https://doi.org/10.1145/3637528.3671470 (cit. on p. 5).

Feng, Y., Ding, L., & Xiao, G. (2023). GeoQAMap - Geographic Question Answering with Maps Leveraging LLM and Open Knowledge Base [ISSN: 1868-8969], *277*. https://doi.org/10.4230/LIPIcs.GIScience.2023.28 (cit. on p. 31).

Frez, J., & Baloian, N. (2023). Bridging the gap: Enhancing geospatial analysis with natural language and scenario generation language [Cited by: 0]. *Lecture Notes in Networks and Systems*, *842 LNNS*, 252–263. https://doi.org/10.1007/978-3-031-48642-5_24 (cit. on p. 4).

Gallegos, I. O., Rossi, R. A., Barrow, J., Tanjim, M. M., Kim, S., Dernoncourt, F., Yu, T., Zhang, R., & Ahmed, N. K. (2024). Bias and Fairness in Large Language Models: A Survey. *Computational Linguistics*, 1–83. https://doi.org/10.1162/coli_a_00524 (cit. on p. 9).

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, M., & Wang, H. (2024, March). Retrieval-Augmented Generation for Large Language Models: A Survey [arXiv:2312.10997 [cs]]. https://doi.org/10.48550/arXiv.2312.10997 (cit. on p. 6).

Goodchild, M. (2009). NeoGeography and the nature of geographic expertise. *Journal of Location Based Services*, *3*(2), 82–96. https://doi.org/10.1080/17489720902950374 (cit. on p. 4).

Haklay, M. ((2013). Neogeography and the Delusion of Democratisation [Publisher: SAGE Publications Ltd]. *Environment and Planning A: Economy and Space*, *45*(1), 55–69. https://doi.org/10.1068/a45184 (cit. on p. 1).

Instituto Nacional de Estatística (INE). (2021). Censos 2021 - Resultados Definitivos [Accessed: 2024-10-07]. https://tabulador.ine.pt/censos2021/ (cit. on p. 18).

Kim, H., Jeon, T., Choi, S., Choi, S., & Cho, H. (2024, October). FLEX: Expert-level False-Less EXecution Metric for Reliable Text-to-SQL Benchmark [arXiv:2409.19014]. https://doi.org/10.48550/arXiv.2409.19014 (cit. on pp. 19, 31).

Li, J., Hui, B., Qu, G., Yang, J., Li, B., Li, B., Wang, B., Qin, B., Geng, R., Huo, N., Zhou, X., Chenhao, M., Li, G., Chang, K., Huang, F., Cheng, R., & Li, Y. (2023). Can LLM

Already Serve as A Database Interface? A BIg Bench for Large-Scale Database Grounded Text-to-SQLs. *Advances in Neural Information Processing Systems*, *36*, 42330–42357. Retrieved 2024-11-22, from https://proceedings.neurips.cc/paper_files/paper/2023/hash/83fc8fab1710363050bbd1d4b8cc0021-Abstract-Datasets_and_Benchmarks.html (cit. on p. 18).

Li, Z., & Ning, H. (2023). Autonomous gis: The next-generation ai-powered gis [Cited by: 7; All Open Access, Gold Open Access, Green Open Access]. *International Journal of Digital Earth*, *16*(2), 4668–4686. https://doi.org/10.1080/17538947.2023.2278895 (cit. on p. 9).

Li, Z., Wang, X., Zhao, J., Yang, S., Du, G., Hu, X., Zhang, B., Ye, Y., Li, Z., Zhao, R., et al. (2024). Pet-sql: A prompt-enhanced two-stage text-to-sql framework with cross-consistency. *arXiv preprint arXiv:2403.09732* (cit. on p. 7).

Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2015). *Geographic information science and systems*. John Wiley & Sons. (Cit. on p. 1).

Malakar, K. D., & Roy, S. (2024). GIS for All: Challenges and Future Directions. In K. D. Malakar & S. Roy (Eds.), *Mapping Geospatial Citizenship: The Power of Participatory GIS* (pp. 73–85). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-63107-8_6 (cit. on p. 1).

Mooney, P., Cui, W., Guan, B., & Juhász, L. Towards understanding the geospatial skills of chatgpt: Taking a geographic information systems (gis) exam [Cited by: 7; All Open Access, Hybrid Gold Open Access]. In: Cited by: 7; All Open Access, Hybrid Gold Open Access. 2023, 85–94. https://doi.org/10.1145/3615886.3627745 (cit. on p. 6).

Qu, C., Dai, S., Wei, X., Cai, H., Wang, S., Yin, D., Xu, J., & Wen, J.-R. (2024). Tool learning with large language models: A survey. *arXiv preprint arXiv:2405.17935* (cit. on p. 6).

Raasveldt, M., & Mühleisen, H. (2019). DuckDB: An Embeddable Analytical Database. *Proceedings of the 2019 International Conference on Management of Data*, 1981–1984. https://doi.org/10.1145/3299869.3320212 (cit. on p. 17).

Redican, K., Gonzalez, M., & Zizzamia, B. (2024). Assessing chatgpt for gis education and assignment creation [Cited by: 0]. *Journal of Geography in Higher Education*. https://doi.org/10.1080/03098265.2024.2397332 (cit. on p. 6).

Rowland, A., Folmer, E., & Beek, W. (2020). Towards Self-Service GIS—Combining the Best of the Semantic Web and Web GIS [Number: 12 Publisher: Multidisciplinary Digital Publishing Institute]. *ISPRS International Journal of Geo-Information*, *9*(12), 753. https://doi.org/10.3390/ijgi9120753 (cit. on pp. 4, 33).

Shuster, K., Poff, S., Chen, M., Kiela, D., & Weston, J. (2021, November). Retrieval Augmentation Reduces Hallucination in Conversation. In M.-F. Moens, X. Huang, L. Specia, & S. W.-t. Yih (Eds.), *Findings of the Association for Computational Linguistics: EMNLP 2021* (pp. 3784–3803). Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.findings-emnlp.320 (cit. on p. 9).

Sindhu, P, P. R., B, S. M., & S, K. (2024). The Evolution of Large Language Model: Models, Applications and Challenges. *2024 International Conference on Current Trends in Advanced Computing (ICCTAC)*, 1–8. https://doi.org/10.1109/ICCTAC61556.2024.10581180 (cit. on p. 1).

Wang, S., Hu, T., Xiao, H., Li, Y., Zhang, C., Ning, H., Zhu, R., Li, Z., & Ye, X. (2024). GPT, large language models (LLMs) and generative artificial intelligence (GAI) models in geospatial science: A systematic review. *International Journal of Digital Earth*, *17*(1), 2353122. https://doi.org/10.1080/17538947.2024.2353122 (cit. on pp. 6, 9, 33).

Wang, Y., Zhong, W., Li, L., Mi, F., Zeng, X., Huang, W., Shang, L., Jiang, X., & Liu, Q. (2023, July). Aligning Large Language Models with Human: A Survey [arXiv:2307.12966 [cs]]. https://doi.org/10.48550/arXiv.2307.12966 (cit. on p. 9).

Yu, T., Zhang, R., Yang, K., Yasunaga, M., Wang, D., Li, Z., Ma, J., Li, I., Yao, Q., Roman, S., Zhang, Z., & Radev, D. (2018, October). Spider: A Large-Scale Human-Labeled Dataset for Complex and Cross-Domain Semantic Parsing and Text-to-SQL Task. In E. Riloff, D. Chiang, J. Hockenmaier, & J. Tsujii (Eds.), *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (pp. 3911–3921). Association for Computational Linguistics. https://doi.org/10.18653/v1/D18-1425 (cit. on p. 18).

Zhang, Q., Kang, Y., & Roth, R. (2023). The Ethics of AI-Generated Maps: DALL·E 2 and AI's Implications for Cartography (Short Paper). https://doi.org/10.4230/LIPIcs.GIScience.2023.93 (cit. on pp. 2, 9, 32).

Zhong, V., Xiong, C., & Socher, R. (2017). Seq2sql: Generating structured queries from natural language using reinforcement learning. *CoRR*, *abs/1709.00103* (cit. on p. 18).

A

Prompt2Map architecture

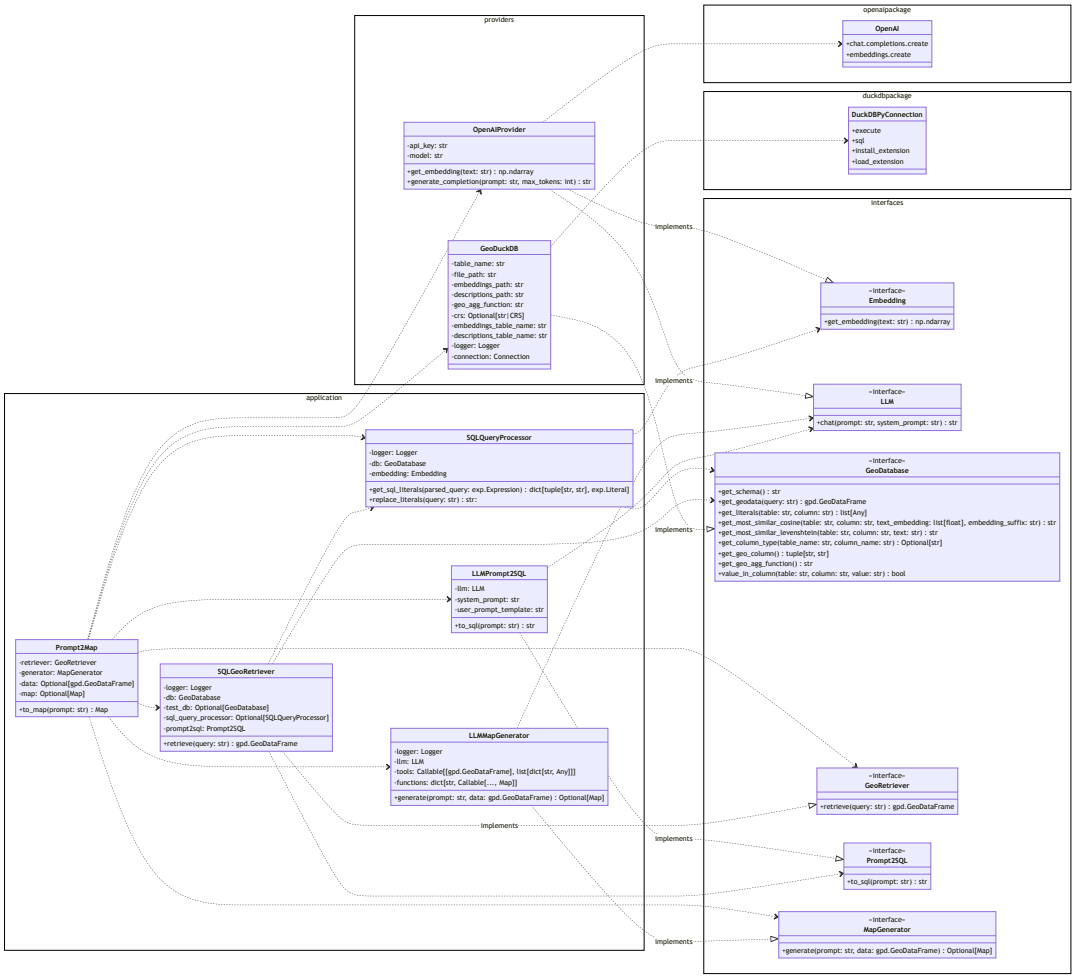


Figure A.1: UML class diagram of Prompt2Map implementation in Python

B

Portuguese 2021 Census Data

Table B.1: Description of table fields, including building and housing details.

Variable	Description
N_EDIFICIOS_CLASSICOS	Number of classic buildings
N_EDIFICIOS_CLASSICOS_1OU2_ALOJ	Number of classic buildings, built to have 1 or 2 housing units
N_EDIFICIOS_CLASSICOS_3OUMAIS_ALOJ	Number of classic buildings, built to have 3 or more housing units
N_EDIFICIOS_CLASSICOS_OUTRO	Number of classic buildings of another type
N_EDIFICIOS_1OU2_PISOS	Number of buildings with 1 or 2 floors
N_EDIFICIOS_3OU4_PISOS	Number of buildings with 3 or 4 floors
N_EDIFICIOS_5OU_MAIS_PISOS	Number of buildings with 5 or more floors
N_EDIFICIOS_SEM_ELEVADOR	Number of buildings without an elevator
N_EDIFICIOS_COM_ELEVADOR	Number of buildings with an elevator
N_EDIFICIOS_EXCLUS_RESIDENCIAL	Number of buildings exclusively residential
N_EDIFICIOS_PINCIPAL_RESIDENCIAL	Number of buildings primarily residential
N_EDIFICIOS_PINCIPAL_NAO_RESIDENCIAL	Number of buildings primarily non-residential
N_EDIFICIOS_CONSTR_ANTES_1919	Number of buildings built before 1919
N_EDIFICIOS_CONSTR_1919A1945	Number of buildings built between 1919 and 1945
N_EDIFICIOS_CONSTR_1946A1960	Number of buildings built between 1946 and 1960
N_EDIFICIOS_CONSTR_1961A1980	Number of buildings built between 1961 and 1980
N_EDIFICIOS_CONSTR_1981A1990	Number of buildings built between 1981 and 1990
N_EDIFICIOS_CONSTR_1991A2000	Number of buildings built between 1991 and 2000
N_EDIFICIOS_CONSTR_2001A2005	Number of buildings built between 2001 and 2005
N_EDIFICIOS_CONSTR_2006A2010	Number of buildings built between 2006 and 2010
N_EDIFICIOS_CONSTR_2011A2015	Number of buildings built between 2011 and 2015
N_EDIFICIOS_CONSTR_2016A2021	Number of buildings built between 2016 and 2021
N_EDIFICIOS_COM_NEC_REPARACAO	Number of buildings needing repair
N_EDIFICIOS_COM_NEC_REPARACAO_LIGEIRAS	Number of buildings needing minor repairs
N_EDIFICIOS_COM_NEC_REPARACAO_MEDIAS	Number of buildings needing moderate repairs
N_EDIFICIOS_COM_NEC_REPARACAO_PROFUNDAS	Number of buildings needing major repairs
N_EDIFICIOS_SEM_NEC_REPARACAO	Number of buildings not needing repairs
N_ALOJAMENTOS_TOTAL	Total number of housing units
N_ALOJAMENTOS_FAMILIARES	Number of family housing units
N_ALOJAMENTOS_FAM_CLASSICOS	Number of classic family housing units
N_ALOJAMENTOS_FAM_N_CLASSICOS	Number of non-classic family housing units
N_ALOJAMENTOS_COLECTIVOS	Number of collective housing units
N_CLASSICOS_RES_HABITUAL	Number of classic family housing units for habitual residence
N_ALOJAMENTOS_FAM_CLASS_RES_SECUNDARIA	Number of classic family housing units for secondary residence
N_ALOJAMENTOS_VAGOS_TOTAL	Number of vacant classic family housing units
N_ALOJAMENTOS_FAM_CLASS_VAGOS_VEND_ARRE	Number of vacant classic family housing units for sale or rent
N_ALOJAMENTOS_FAM_CLASS_VAGOS_OUTR_MOTIVOS	Number of vacant classic family housing units for other reasons

N_RHABITUAL_ACESSIVEL_CADEIRAS_-RODAS	Number of classic family housing units for habitual residence accessible to wheelchairs
N_RHABITUAL_NAO_ACESSIVEL_-CADEIRAS_RODAS	Number of classic family housing units for habitual residence not accessible to wheelchairs
N_RHABITUAL_COM_ESTACIONAMENTO	Number of classic family housing units for habitual residence with parking
N_RHABITUAL_SEM_ESTACIONAMENTO	Number of classic family housing units for habitual residence without parking
N_RHABITUAL_PROP_OCUP	Number of classic family housing units for habitual residence owned by occupants
N_RHABITUAL_ARRENDADOS	Number of classic family housing units for habitual residence rented
N_RHABITUAL_OCUPADOS_OUTR_SITUACAO	Number of classic family housing units for habitual residence, other situation
N_RHABITUAL_AREA_49	Number of classic family housing units for habitual residence, with usable area up to 49 m²
N_RHABITUAL_AREA_50_99	Number of classic family housing units for habitual residence, with usable area between 50 and 99 m²
N_RHABITUAL_AREA_100_149	Number of classic family housing units for habitual residence, with usable area between 100 and 149 m²
N_RHABITUAL_AREA_150_199	Number of classic family housing units for habitual residence, with usable area between 150 and 199 m²
N_RHABITUAL_AREA_200MAIS	Number of classic family housing units for habitual residence, with usable area 200 m² or more
N_RHABITUAL_1_2_DIV	Number of classic family housing units for habitual residence, with 1 or 2 rooms
N_RHABITUAL_3_4_DIV	Number of classic family housing units for habitual residence, with 3 or 4 rooms
N_RHABITUAL_5_MAIS_DIV	Number of classic family housing units for habitual residence, with 5 or more rooms
N_AGREGADOS_DOMESTICOS_PRIVADOS	Number of private households
N_AGREGADOS_INSTITUCIONAIS	Number of institutional households
N_ADP_1OU2_PESSOAS	Number of private households with 1 or 2 people
N_ADP_3OU4_PESSOAS	Number of private households with 3 or 4 people
N_ADP_5EMAIS_PESSOAS	Number of private households with 5 or more people
N_INDIVIDUOS	Total number of individuals
N_INDIVIDUOS_H	Number of male individuals
N_INDIVIDUOS_M	Number of female individuals
N_INDIVIDUOS_0A14	Number of individuals aged 0 to 14 years
N_INDIVIDUOS_15A24	Number of individuals aged 15 to 24 years
N_INDIVIDUOS_25A64	Number of individuals aged 25 to 64 years
N_INDIVIDUOS_65_OU_MAIS	Number of individuals aged 65 or more years
N_INDIVIDUOS_0A14_H	Number of individuals aged 0 to 14 years - Men
N_INDIVIDUOS_15A24_H	Number of individuals aged 15 to 24 years - Men
N_INDIVIDUOS_25A64_H	Number of individuals aged 25 to 64 years - Men
N_INDIVIDUOS_65_OU_MAIS_H	Number of individuals aged 65 or more years - Men
N_INDIVIDUOS_0A14_M	Number of individuals aged 0 to 14 years - Women
N_INDIVIDUOS_15A24_M	Number of individuals aged 15 to 24 years - Women
N_INDIVIDUOS_25A64_M	Number of individuals aged 25 to 64 years - Women
N_INDIVIDUOS_65_OU_MAIS_M	Number of individuals aged 65 or more years - Women
N_INDIVIDUOS_0A4	Number of individuals aged 0 to 4 years
N_INDIVIDUOS_5A9	Number of individuals aged 5 to 9 years
N_INDIVIDUOS_10A14	Number of individuals aged 10 to 14 years
N_INDIVIDUOS_15A19	Number of individuals aged 15 to 19 years
N_INDIVIDUOS_20A24	Number of individuals aged 20 to 24 years
N_INDIVIDUOS_25A29	Number of individuals aged 25 to 29 years
N_INDIVIDUOS_30A34	Number of individuals aged 30 to 34 years
N_INDIVIDUOS_35A39	Number of individuals aged 35 to 39 years
N_INDIVIDUOS_40A44	Number of individuals aged 40 to 44 years

N_INDIVIDUOS_45A49	Number of individuals aged 45 to 49 years
N_INDIVIDUOS_50A54	Number of individuals aged 50 to 54 years
N_INDIVIDUOS_55A59	Number of individuals aged 55 to 59 years
N_INDIVIDUOS_60A64	Number of individuals aged 60 to 64 years
N_INDIVIDUOS_65A69	Number of individuals aged 65 to 69 years
N_INDIVIDUOS_70A74	Number of individuals aged 70 to 74 years
N_INDIVIDUOS_75_OU_MAIS	Number of individuals aged 75 or more years
N_INDIVIDUOS_0A4_H	Number of individuals aged 0 to 4 years - Men
N_INDIVIDUOS_5A9_H	Number of individuals aged 5 to 9 years - Men
N_INDIVIDUOS_10A14_H	Number of individuals aged 10 to 14 years - Men
N_INDIVIDUOS_15A19_H	Number of individuals aged 15 to 19 years - Men
N_INDIVIDUOS_20A24_H	Number of individuals aged 20 to 24 years - Men
N_INDIVIDUOS_25A29_H	Number of individuals aged 25 to 29 years - Men
N_INDIVIDUOS_30A34_H	Number of individuals aged 30 to 34 years - Men
N_INDIVIDUOS_35A39_H	Number of individuals aged 35 to 39 years - Men
N_INDIVIDUOS_40A44_H	Number of individuals aged 40 to 44 years - Men
N_INDIVIDUOS_45A49_H	Number of individuals aged 45 to 49 years - Men
N_INDIVIDUOS_50A54_H	Number of individuals aged 50 to 54 years - Men
N_INDIVIDUOS_55A59_H	Number of individuals aged 55 to 59 years - Men
N_INDIVIDUOS_60A64_H	Number of individuals aged 60 to 64 years - Men
N_INDIVIDUOS_65A69_H	Number of individuals aged 65 to 69 years - Men
N_INDIVIDUOS_70A74_H	Number of individuals aged 70 to 74 years - Men
N_INDIVIDUOS_75_OU_MAIS_H	Number of individuals aged 75 or more years - Men
N_INDIVIDUOS_0A4_M	Number of individuals aged 0 to 4 years - Women
N_INDIVIDUOS_5A9_M	Number of individuals aged 5 to 9 years - Women
N_INDIVIDUOS_10A14_M	Number of individuals aged 10 to 14 years - Women
N_INDIVIDUOS_15A19_M	Number of individuals aged 15 to 19 years - Women
N_INDIVIDUOS_20A24_M	Number of individuals aged 20 to 24 years - Women
N_INDIVIDUOS_25A29_M	Number of individuals aged 25 to 29 years - Women
N_INDIVIDUOS_30A34_M	Number of individuals aged 30 to 34 years - Women
N_INDIVIDUOS_35A39_M	Number of individuals aged 35 to 39 years - Women
N_INDIVIDUOS_40A44_M	Number of individuals aged 40 to 44 years - Women
N_INDIVIDUOS_45A49_M	Number of individuals aged 45 to 49 years - Women
N_INDIVIDUOS_50A54_M	Number of individuals aged 50 to 54 years - Women
N_INDIVIDUOS_55A59_M	Number of individuals aged 55 to 59 years - Women
N_INDIVIDUOS_60A64_M	Number of individuals aged 60 to 64 years - Women
N_INDIVIDUOS_65A69_M	Number of individuals aged 65 to 69 years - Women
N_INDIVIDUOS_70A74_M	Number of individuals aged 70 to 74 years - Women
N_INDIVIDUOS_75_OU_MAIS_M	Number of individuals aged 75 or more years - Women
N_INDIVIDUO_ENSINCOMP_NENHUM	Number of individuals with no completed level of education
N_INDIVIDUO_ENSINCOMP_1BAS	Number of individuals with one completed level of education - 1st cycle
N_INDIVIDUO_ENSINCOMP_2BAS	Number of individuals with one completed level of education - 2nd cycle
N_INDIVIDUO_ENSINCOMP_3BAS	Number of individuals with one completed level of education - 3rd cycle
N_INDIVIDUO_ENSINCOMP_SEC_E_POS-SEC	Number of individuals with one completed level of education - Secondary or post-secondary
N_INDIVIDUO_ENSINCOMP_SUP	Number of individuals with one completed level of education - Higher education
N_INDIVIDUO_ENSINCOMP_NENHUM_H	Number of individuals with no completed level of education - Men
N_INDIVIDUO_ENSINCOMP_1BAS_H	Number of individuals with one completed level of education - 1st cycle - Men
N_INDIVIDUO_ENSINCOMP_2BAS_H	Number of individuals with one completed level of education - 2nd cycle - Men
N_INDIVIDUO_ENSINCOMP_3BAS_H	Number of individuals with one completed level of education - 3rd cycle - Men
N_INDIVIDUO_ENSINCOMP_SEC_E_POS-SEC_H	Number of individuals with one completed level of education - Secondary or post-secondary - Men

N_INDIVIDUO_ENSINCOMP_SUP_H	Number of individuals with one completed level of education - Higher education - Men
N_INDIVIDUO_ENSINCOMP_NENHUM_M	Number of individuals with no completed level of education - Women
N_INDIVIDUO_ENSINCOMP_1BAS_M	Number of individuals with one completed level of education - 1st cycle - Women
N_INDIVIDUO_ENSINCOMP_2BAS_M	Number of individuals with one completed level of education - 2nd cycle - Women
N_INDIVIDUO_ENSINCOMP_3BAS_M	Number of individuals with one completed level of education - 3rd cycle - Women
N_INDIVIDUO_ENSINCOMP_SEC_E_POS-SEC_M	Number of individuals with one completed level of education - Secondary or post-secondary - Women
N_INDIVIDUO_ENSINCOMP_SUP_M	Number of individuals with one completed level of education - Higher education - Women
N_INDIVIDUOS_COM_ATIVIDADE_ECO-NOMICA	Number of individuals with economic activity
N_INDIVIDUOS_SEM_ATIVIDADE_ECO-NOMICA	Number of individuals without economic activity
N_INDIVIDUOS_EMPREGADOS	Number of individuals with economic activity - Employed
N_INDIVIDUOS_DESEMPREGADOS_1EMP	Number of individuals with economic activity - Unemployed, seeking first job
N_INDIVIDUOS_DESEMPREGADOS_-NOVOEMP	Number of individuals with economic activity - Unemployed, seeking new job
N_INDIVIDUOS_ESTUDANTES	Number of individuals without activity - Students
N_INDIVIDUOS_DOMESTICOS	Number of individuals without economic activity - Domestic workers
N_INDIVIDUOS_REFORMADOS	Number of individuals without economic activity - Retired
N_INDIVIDUOS_EMPREG_SECT_PRIM	Number of individuals employed in the primary sector
N_INDIVIDUOS_EMPREG_SECT_SEC	Number of individuals employed in the secondary sector
N_INDIVIDUOS_EMPREG_SECT_TERC	Number of individuals employed in the tertiary sector
N_INDIVIDUOS_NAC_ESTRANGEIRA	Number of individuals with foreign nationality
N_INDIVIDUOS_RESID_FORA_PAIS	Number of individuals who have lived outside Portugal
N_INDIVIDUOS_COM_ATIVIDADE_ECO-NOMICA_H	Number of individuals with economic activity - Men
N_INDIVIDUOS_SEM_ATIVIDADE_ECO-NOMICA_H	Number of individuals without economic activity - Men
N_INDIVIDUOS_EMPREGADOS_H	Number of individuals with economic activity - Employed - Men
N_INDIVIDUOS_DESEMPREGADOS_-1EMP_H	Number of individuals with economic activity - Unemployed, seeking first job - Men
N_INDIVIDUOS_DESEMPREGADOS_-NOVOEMP_H	Number of individuals with economic activity - Unemployed, seeking new job - Men
N_INDIVIDUOS_ESTUDANTES_H	Number of individuals without activity - Students - Men
N_INDIVIDUOS_DOMESTICOS_H	Number of individuals without economic activity - Domestic workers - Men
N_INDIVIDUOS_REFORMADOS_H	Number of individuals without economic activity - Retired - Men
N_INDIVIDUOS_EMPREG_SECT_PRIM_H	Number of individuals employed in the primary sector - Men
N_INDIVIDUOS_EMPREG_SECT_SEC_H	Number of individuals employed in the secondary sector - Men
N_INDIVIDUOS_EMPREG_SECT_TERC_H	Number of individuals employed in the tertiary sector - Men
N_INDIVIDUOS_NAC_ESTRANGEIRA_H	Number of individuals with foreign nationality - Men
N_INDIVIDUOS_RESID_FORA_PAIS_H	Number of individuals who have lived outside Portugal - Men
N_INDIVIDUOS_COM_ATIVIDADE_ECO-NOMICA_M	Number of individuals with economic activity - Women
N_INDIVIDUOS_SEM_ATIVIDADE_ECO-NOMICA_M	Number of individuals without economic activity - Women
N_INDIVIDUOS_EMPREGADOS_M	Number of individuals with economic activity - Employed - Women
N_INDIVIDUOS_DESEMPREGADOS_-1EMP_M	Number of individuals with economic activity - Unemployed, seeking first job - Women
N_INDIVIDUOS_DESEMPREGADOS_-NOVOEMP_M	Number of individuals with economic activity - Unemployed, seeking new job - Women
N_INDIVIDUOS_ESTUDANTES_M	Number of individuals without activity - Students - Women

N_INDIVIDUOS_DOMESTICOS_M	Number of individuals without economic activity - Domestic workers - Women
N_INDIVIDUOS_REFORMADOS_M	Number of individuals without economic activity - Retired - Women
N_INDIVIDUOS_EMPREG_SECT_PRIM_M	Number of individuals employed in the primary sector - Women
N_INDIVIDUOS_EMPREG_SECT_SEC_M	Number of individuals employed in the secondary sector - Women
N_INDIVIDUOS_EMPREG_SECT_TERC_M	Number of individuals employed in the tertiary sector - Women
N_INDIVIDUOS_NAC_ESTRANGEIRA_M	Number of individuals with foreign nationality - Women
N_INDIVIDUOS_RESID_FORA_PAIS_M	Number of individuals who have lived outside Portugal - Women
shape_area_m2	Polygon area in square meters
shape_length_m	Polygon perimeter in meters
Freguesia	Parish name
Municipio	Municipality name
Distrito	District name

C&SIG

UNIGIS PT

NOVA IMS Information Management School