

**UNIVERSITY OF
WESTMINSTER** 



**INFORMATICS
INSTITUTE OF
TECHNOLOGY**

INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER

NL2SQL DATA VISUALIZATION AGENT

A dissertation by

Mr. Pramuditha Sahas

20201214 | W1810601

Supervised by

Mr. John Sriskandarajah

Submitted in partial fulfilment of the requirements for the BEng (Hons) Software
Engineering degree at the University of Westminster.

April 2025

ABSTRACT

Problem: NL2SQL systems allow non-technical users to query databases without having knowledge of SQL language and the database table structure. However, the users will struggle to understand the output due to lack of knowledge on the table structure and the data is in a RAW form. This project aims to address this gap in NL2SQL system by automating the system to transform the RAW data into visualizations that are more human in readable report structure.

Methodology: The proposed approach is about developing an NL2SQL Data Visualization Agent that uses machine learning algorithms to process SQL outputs. The system consists of a chart selection model that dynamically chooses the most suitable chart types based on data structure and user input. Real-time data processing techniques are used to ensure the system can handle large datasets efficiently. The methodology follows an agile development process, integrating user feedback and testing at each stage to refine the system's functionality.

Initial Results: A pre-trained Hugging Face transformer model is used for NL2SQL translation, and the prototype system was evaluated using it. Several natural language queries were fed to the model which were converted to a corresponding SQL statement and executed on an SQLite database. Based on the data structure of each SQL query result, the model was able to recommend the correct chart types (line, bar, pie, and more) for that result. These initial results show that this Python based implementation is feasible to produce accurate data visualizations consistent with SQL query output, and that integrating NLP with chart prediction is effective.

Subject descriptors:

- Information systems -> Information retrieval -> Query languages -> Natural language interfaces
- Computing methodologies -> Machine learning -> Machine learning applications -> Neural networks
- Human-centered computing -> Visualization -> Visualization application domains -> Data visualization

Keywords: NL2SQL, data visualization, natural language querying, machine learning, real-time processing