From SQL to Insights

Using LLMs For Automated Business Intelligence Report Generation

LLM Course – Spring 2024: Final Project

Table of contents

01

Introduction and Problem Statement

Overview of Project Goals, Related Work, and Proposed Solutions

Introduction

- Develop a Business Intelligence (BI) Assistant to convert natural language questions into SQL queries.
- Automatically execute queries, retrieve data, and generate user-friendly reports with essential charts
- Address the complexities and time-consuming nature of traditional data analysis methods

Previous Works

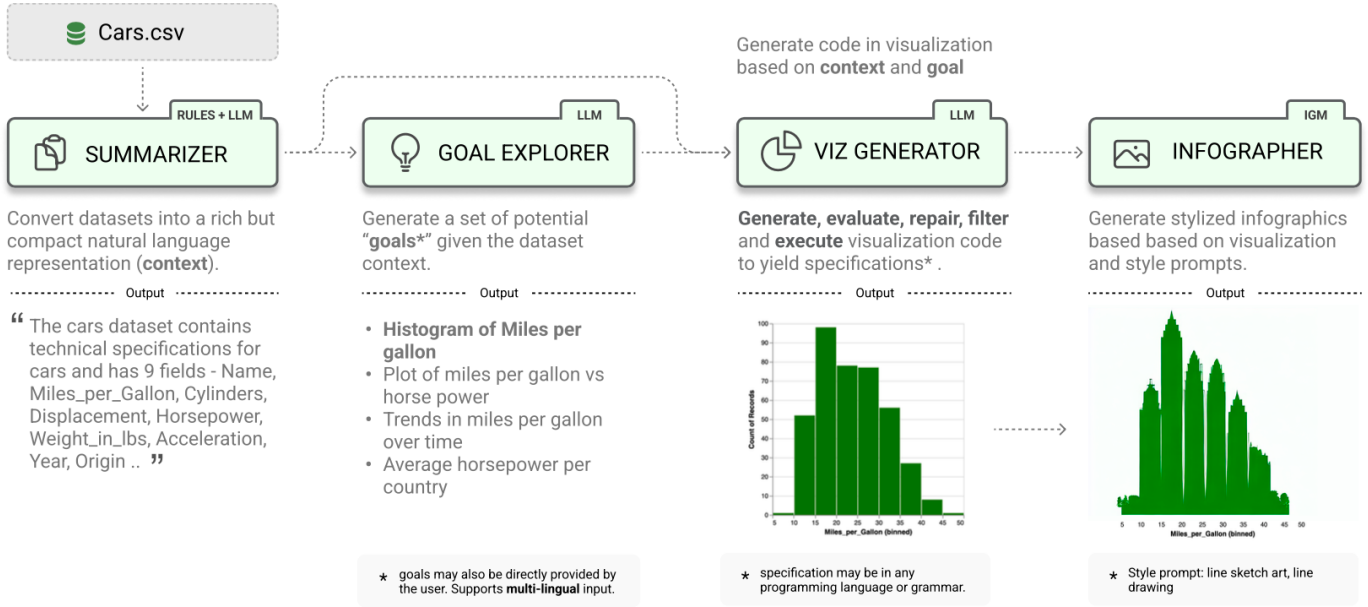


Figure 1: LIDA generates visualizations and infographics across 4 modules - data summarization, goal exploration, visualization generation and infographics generations. Example output from each module is shown.

Microsoft LIDA – https://arxiv.org/abs/2303.02927

Previous Works

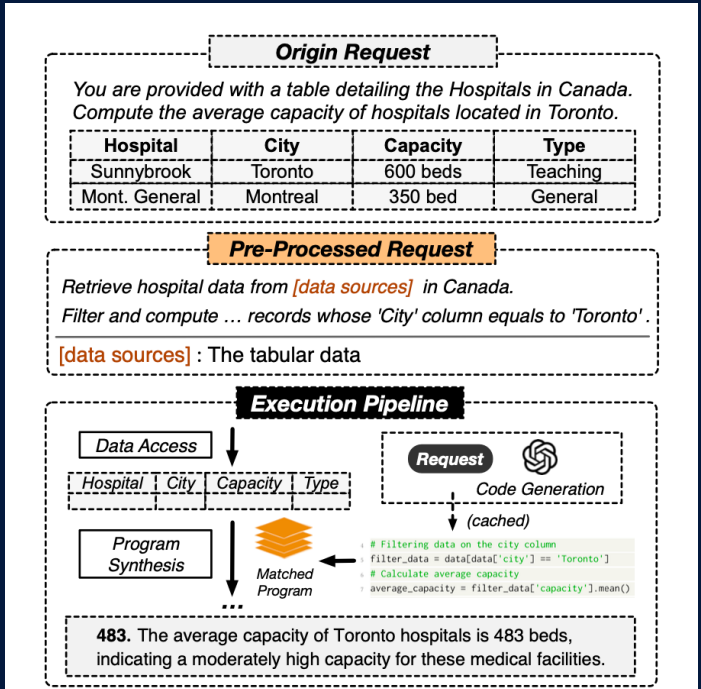


Figure 4: LLMDB for NL-based Data Analytics

LLM-Enhanced Data Management- https://arxiv.org/abs/2402.02643

Previous Works

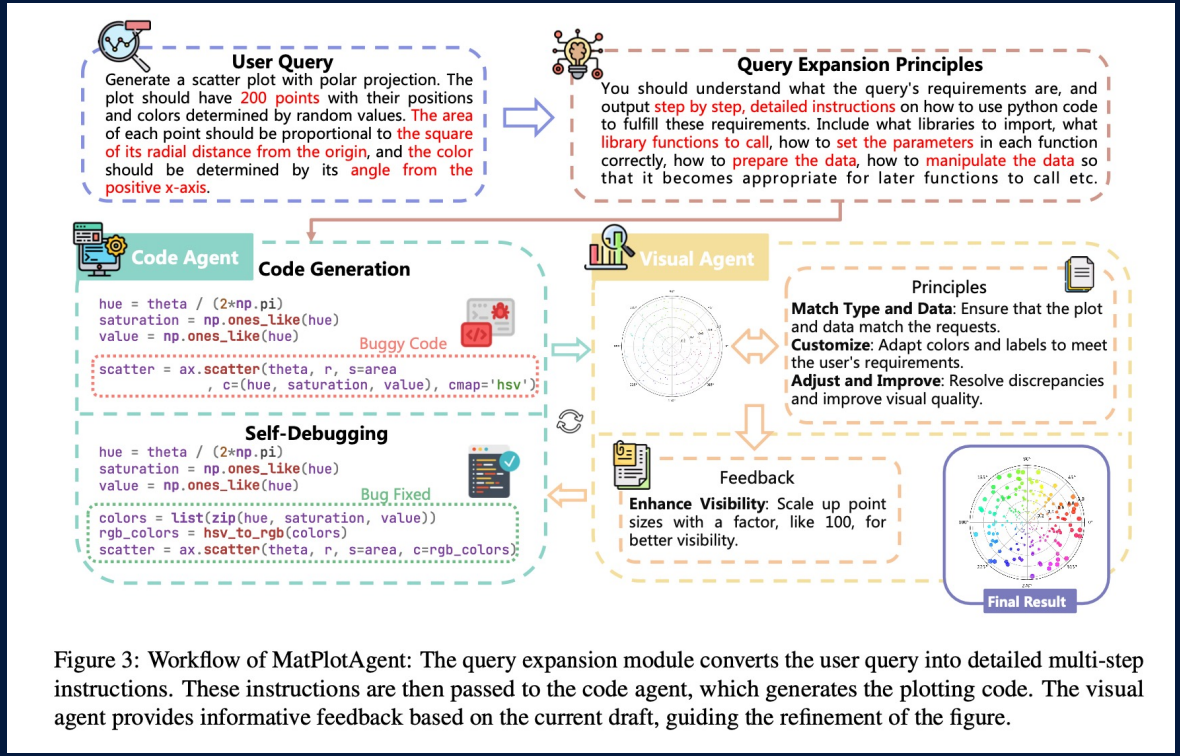


Figure 3: Workflow of MatPlotAgent: The query expansion module converts the user query into detailed multi-step instructions. These instructions are then passed to the code agent, which generates the plotting code. The visual agent provides informative feedback based on the current draft, guiding the refinement of the figure.

MatPlotAgent- https://arxiv.org/abs/2402.11453

Previous Works

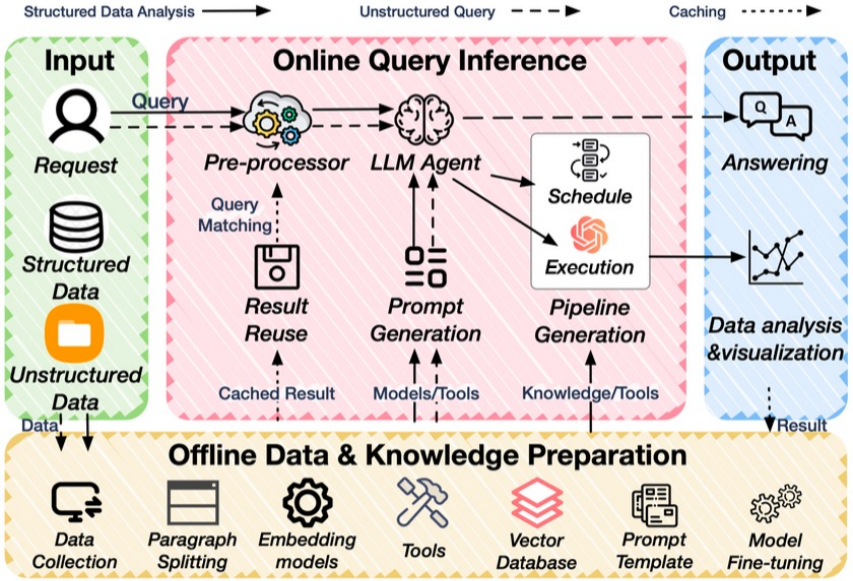


Figure 1: Chat2Data Workflow Overview.

Chat2Data- https://dbgroup.cs.tsinghua.edu.cn/ligl/papers/chat2data_demo_vldb2024.pdf

Previous Works

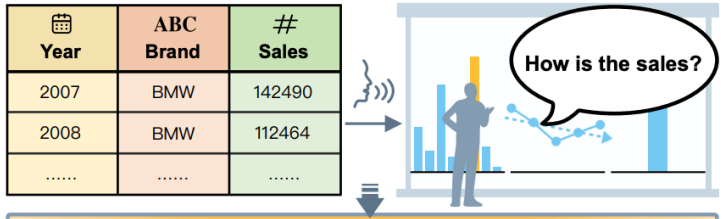


Fig. 1. The user asks the system a question about car sales. The system decomposes this complex question into three simple questions and answers with a set of annotated charts.

Talk2Data- https://arxiv.org/abs/2107.14420

Problem?

They have been made for data analyst!

Our Innovation

Non-Technical User Friendly

Our BI Assistant simplifies data analysis for non-technical users through natural language queries

Data Storytelling

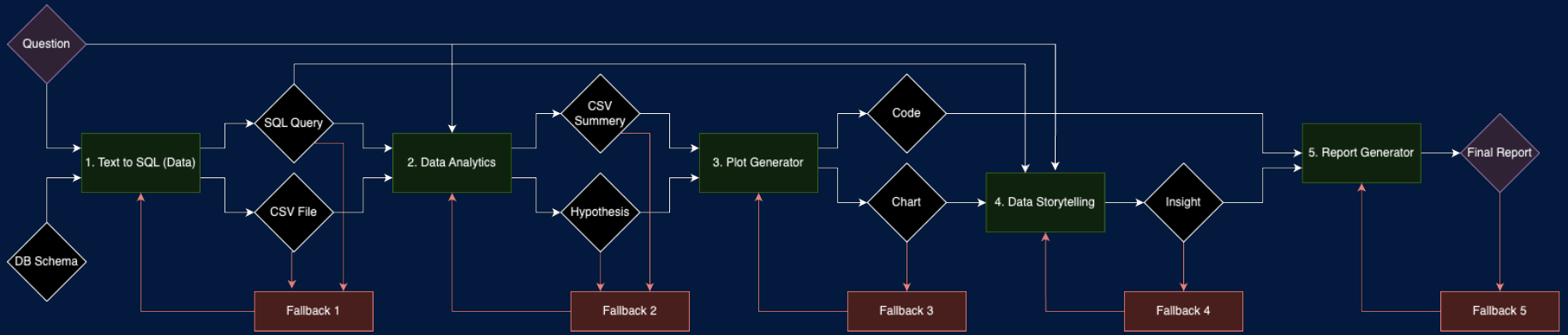
Transforms complex data into compelling narratives and visualizations, making insights easily understandable and actionable

02

Technical Architecture and Implementation

System Design, Tools Utilized, and Implementation Details

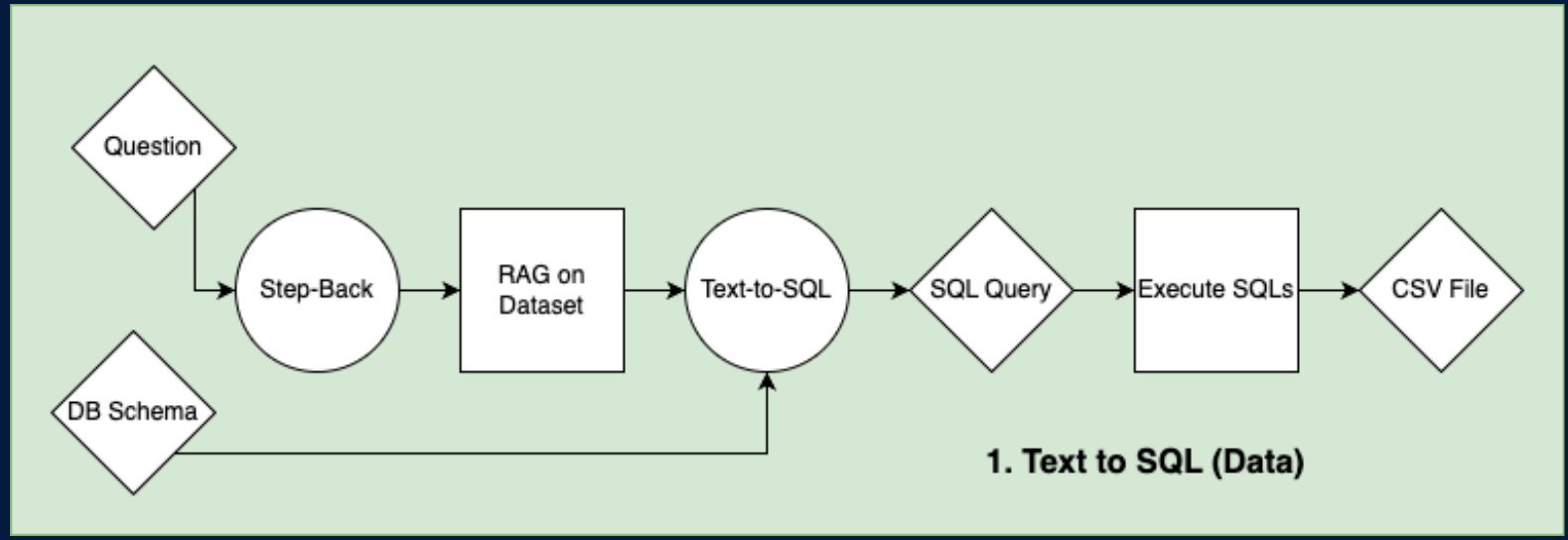
Architecture



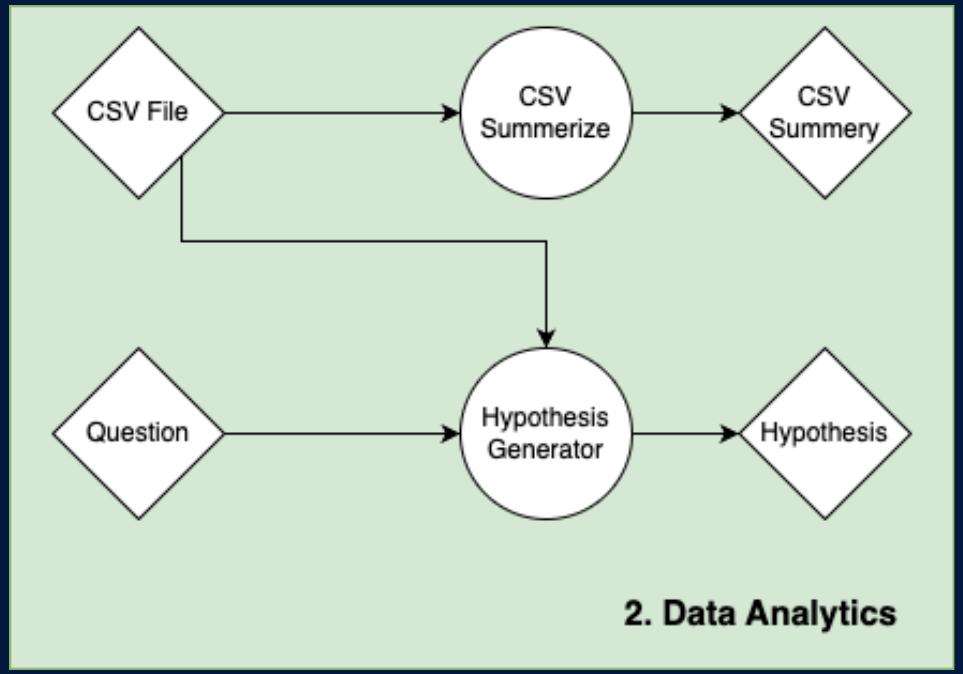
Models, Tools and Framworks

Application Preview

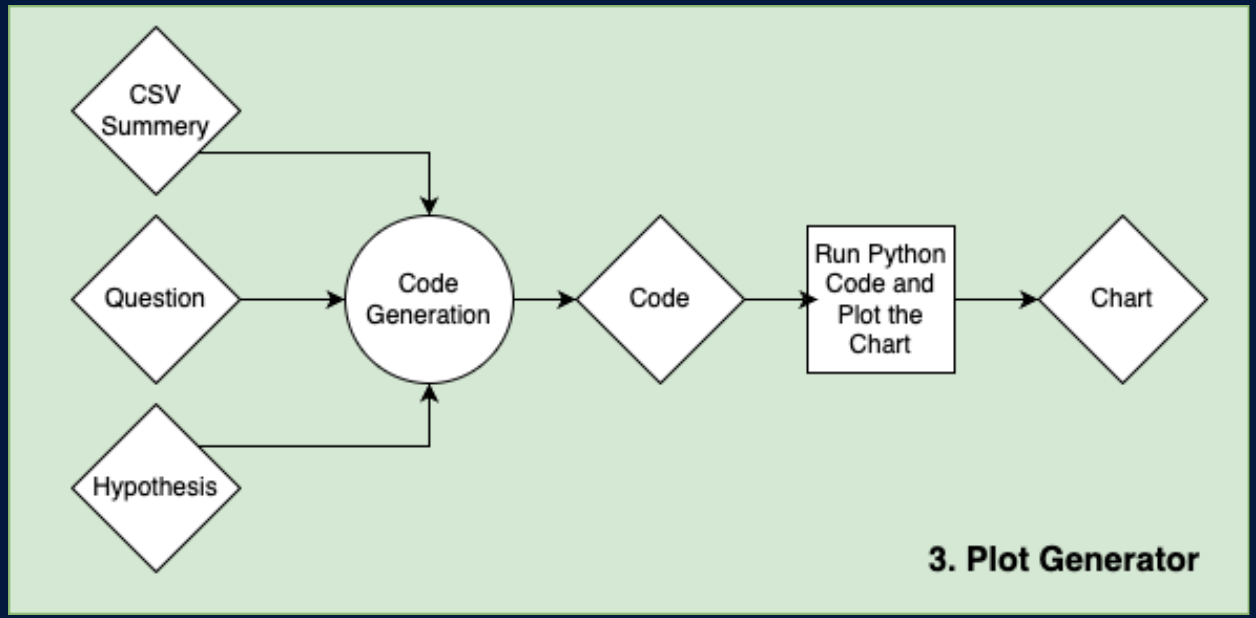
Node 1 : Text2SQL



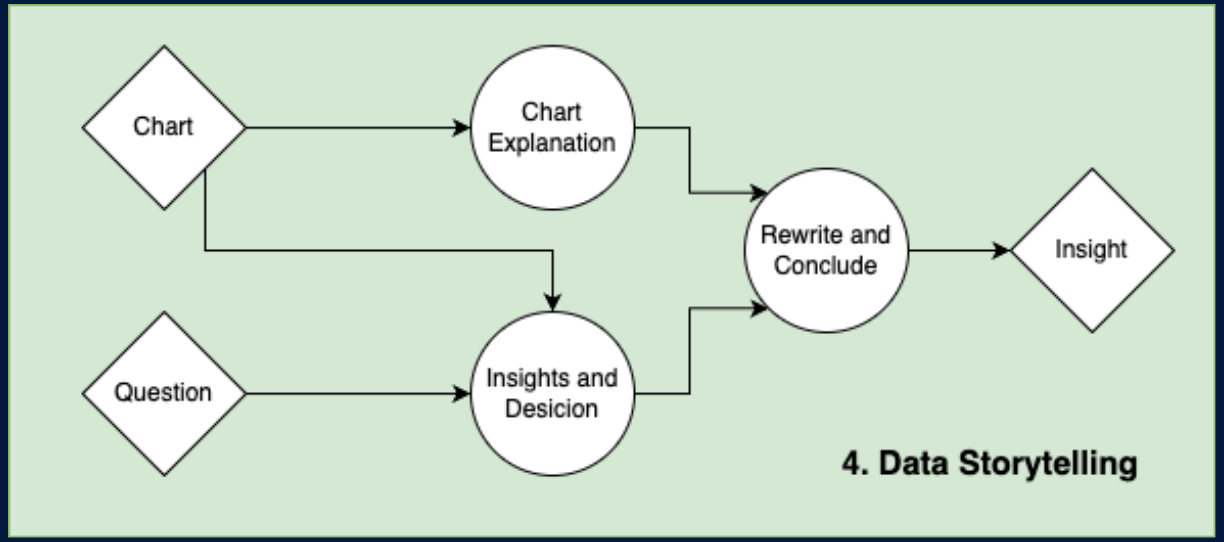
Node 2 : Data Analytics



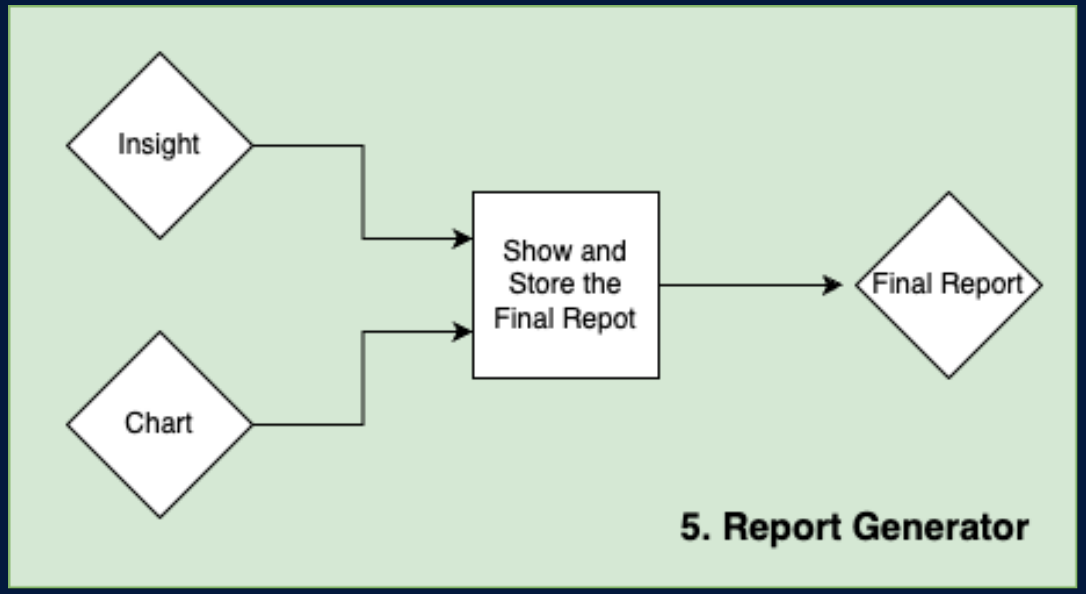
Node 3 : Plot Generator



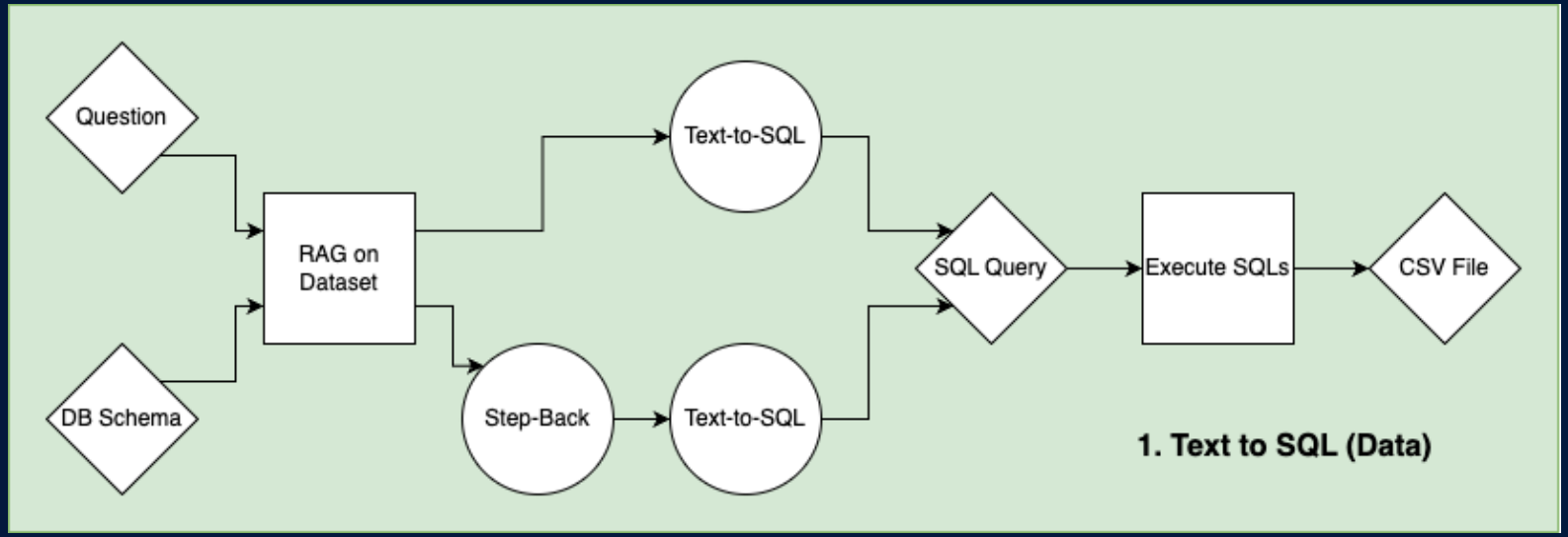
Node 4 : Data Storytelling



Node 5 : Report Generator



Node 1 : Text2SQL

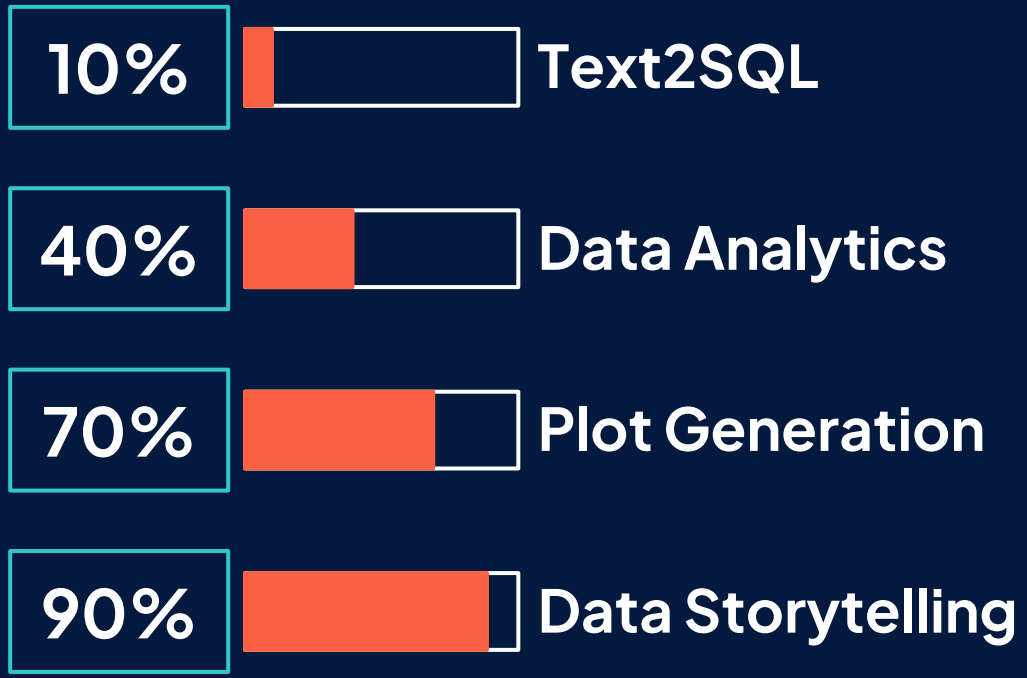


03

Evaluation and Results

Testing Methodology, Performance Metrics, and Analysis

Error Rates



Mercury is the closest planet to the Sun

Venus is the second planet from the Sun

Earth is the third planet from the Sun

Neptune is very far from the Sun

Final Evaluation

Metrics

What are the criteria to assess the final output?

Relevancy

how accurately addresses the question

Insightfulness

depth and usefulness of the insights

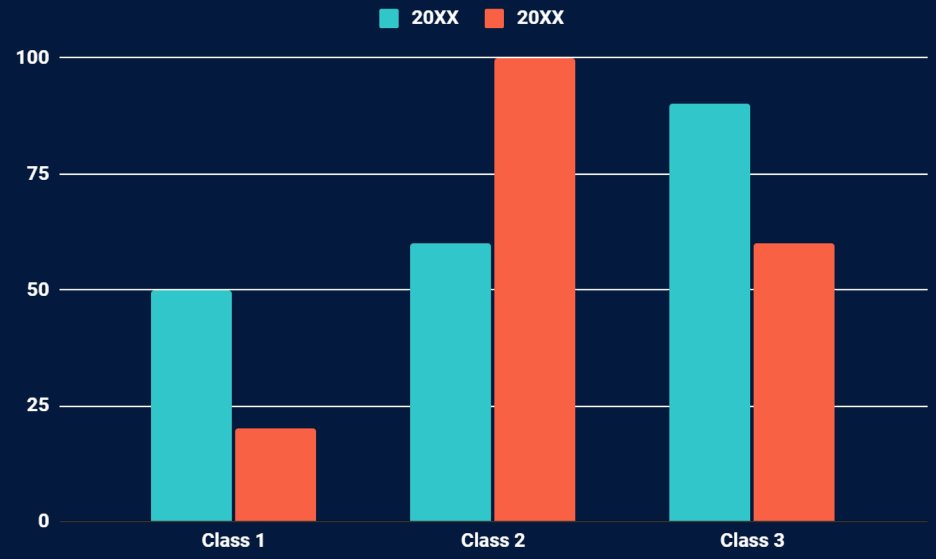
Visualization Quality

correctness, clarity, and readability

Data Storytelling

narrative to a coherent and engaging story

Statistics



20XX

Earth is the third planet from the Sun

20XX

Venus is the second planet from the Sun

Follow the link in the graph to modify its data and then paste the new one here. **For more info, click here**

04

Conclusion and Future Directions

Challenges, Insights, and Potential Enhancements

What are the main challenges?

Cost

The use of LLMs, especially with multistep reasoning and prompting, incurs notable costs

Latency

The multistep nature of the system results in longer response times compared to most digital applications

Accuracy

LLMs can produce highly confident but incorrect or irrelevant answers, posing a challenge of hallucination

Cost Analysis

Introduction

Despite being red, Mars is actually a cold place. It's full of iron oxide dust

- **2XXX** — Mercury is very small
- **2XXX** — Jupiter is very big
- **2XXX** — Venus is a hot planet
- **2XXX** — Neptune is an ice giant

Advantages

Features

Web Application

Coming Soon at:

talkwithyourdata.xyz

Future Road

Admission

Saturn is a gas giant and has several rings

Academic journey

Mercury is the closest planet to the Sun

Support services

Despite being red, Mars is a cold planet

Graduation

Venus is the second planet from the Sun

Our Team

AmirHossein Safdarian

Master of Software Engineering
NLP Lab – University of Tehran

Mahdi Salmani

Master of Artificial Intelligence
NLP Lab – University of Tehran

Milad Mohammadi

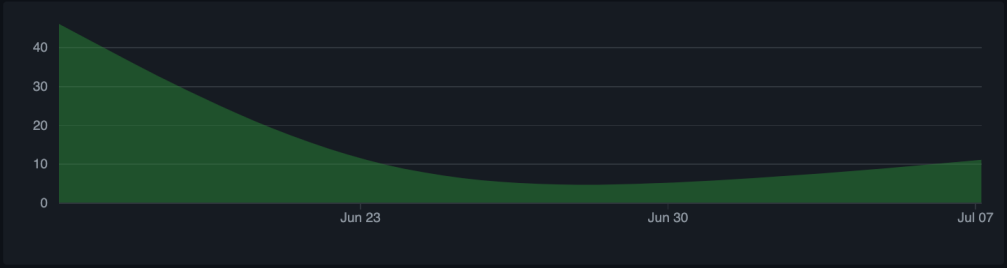
Master of Information Technology
NLP Lab – University of Tehran

Code Report

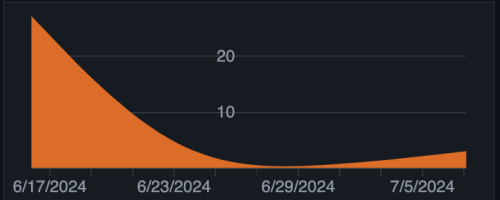
Jun 16, 2024 – Jul 7, 2024

Contributions: Commits

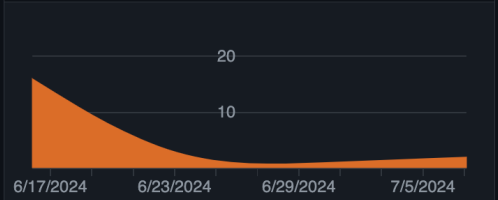
Contributions to master, excluding merge commits



safdarian #1
30 commits 953 ++ 261 --



mohammadi-milad-mim #2
19 commits 226 ++ 92 --



m-salmani78 #3
16 commits 786 ++ 427 --

“Not only can they complete the task much quicker than we can, but in theory, we wouldn’t even need to know how to use these tools in the first place.”

—Harrison Chase, Co-Founder and CEO of LangChain

Thanks!