TADDR: Towards Agent Driven Data Reports

Kishan Bharadwaj Shridhar*

Walmart Global Tech India

Abstract

With the advent of agentic artificial intelligence (AI), data visualization has evolved into a critical step, analogous to exploratory data analysis (EDA) in traditional machine learning workflows. This work presents **TADDR**, an agent-driven framework designed for the *Agentic VIS Challenge 2025*. The framework accepts a CSV dataset as input and produces a structured, narrative-driven report enriched with visual analytics. TADDR is optimized for runtime and token efficiency, while demonstrating adaptability across diverse domains such as sports, finance, and healthcare. The proposed system highlights how autonomous agents can be orchestrated to replicate and extend the decision-making processes of human data analysts.

Index terms: Data Visualization, Agentic AI, Large Language Models, Autonomous Agents

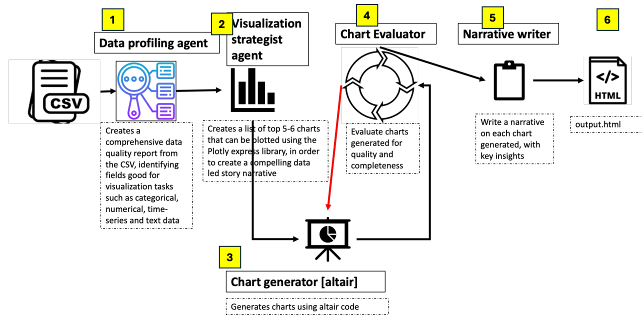
1 Introduction

This work introduces an agentic AI approach developed in response to the Agentic VIS Challenge 2025 [1]. The primary objective of the system is to automatically generate a narrative-led, visualization-rich data story in HTML/PDF format. Unlike traditional static pipelines, TADDR leverages agent autonomy to dynamically adapt its outputs to the characteristics of the incoming dataset. The framework is composed of five interconnected agentic components, each emulating a step in the data analysis workflow. Figure 1 provides a high-level architectural overview of the proposed solution. The design philosophy emphasizes modularity, interpretability, and iterative evaluation of outputs. Subsequent sections detail the technical assumptions, methodological choices, and areas for potential enhancement.

2 Agentic System

The development of TADDR was guided by tracing the methodological workflow typically followed by a human data analyst: data profiling, visualization design, chart creation, quality evaluation, and narrative writing. While the current implementation focuses on the Altair library for visualization, the architecture allows for extensibility to other visualization frameworks.

* e-mail: kishanbs.2016@mitb.smu.edu.sg



https://www.google.com/imgres?q=data%20profiling%20icon&imgurl=https%3A%2F%2Fcdn-icons-png.flaticon.com%2F512%2F2857%2F2857321.png&imgrefurl=https%3A%2F%2Fwww.flaticon.com%2Ffree-icon%2Fdata-classification_2857321&docid=N_bDfvfT_F7-BM&tbnid=QsjjVQYQpqEALM&vet=12ahUKEwiDlojVh7KPAxXb1DgGHS5rEVIQM3oECB0QAA..i&w=512&h=512&hcb=2&ved=2ahUKEwiDlojVh7KPAxXb1DgGHS5rEVIQM3oECB0QAA

Figure 1: Architecture of TADDR

2.1 Data profiling agent

The Data Profiling Agent is responsible for constructing a comprehensive overview of the incoming dataset. This step, analogous to classical EDA, identifies missing values, outliers, field cardinality, and variable types (categorical, numerical, temporal, and textual). Given a dataset ***df***, the agent systematically assesses data quality and produces a structured report. This output serves as the foundation for downstream tasks by the Visualization Strategist Agent.

2.2 Visualisation Strategist

The Visualization Strategist Agent leverages large language model (LLM)-driven reasoning to recommend candidate chart types. Drawing on the profiling report, it selects fields most suitable for visualization, ensuring adherence to best practices in visual design. Recommendations are restricted to basic Python chart types [2], with a maximum of three fields per chart (X-axis, Y-axis, and an optional third dimension for color or filtering). Each recommendation is accompanied by a "thought" rationale, explicitly stating the expected insight. This ensures transparency in the decision-making process and provides a traceable link between dataset characteristics and chart design.

2.3 Chart Generator

The **Chart Generator** module converts the strategist's specifications into executable visualization code. Implemented with reusable Altair templates, this module generates publication-quality figures complete with annotations, legends, and coherent color palettes. The generator can be extended as either a callable tool or a fully autonomous agent.

2.4 Evaluator-Optimiser loop

The Evaluator–Optimizer Loop builds upon the LangGraph evaluator–optimizer paradigm [3]. Charts produced by the generator are evaluated for completeness, readability, and interpretability. If deficiencies are detected—such as missing labels, poor scaling, or incomplete use of fields—the evaluator returns the specification to the generator for refinement. This iterative cycle continues until charts meet predefined quality thresholds. Only validated charts progress to the narrative-writing stage.

2.5 Narrative Generator

The **Narrative Generator** transforms validated charts into a coherent analytical report. Each chart is accompanied by a descriptive narrative of approximately 350–400 words, providing interpretation and contextual insights. The generator also produces:

- **Embedded figures:** Interactive visualizations directly included in the HTML report.
- **Dynamic titles:** Concise, data-driven report titles synthesized after chart validation.
- **Executive summaries:** A 2–3 sentence bullet-point synopsis highlighting key insights.

The combination of visual analytics and narrative enhances interpretability for both expert and non-expert audiences.

2.6 Report Generator

The Report Generator compiles all charts and narratives into a single cohesive document (output.html). The final report balances visual storytelling with explanatory text, offering a holistic perspective of the dataset under analysis.

3 Adaptability to Other Domains

The agentic framework was implemented to two other datasets, on a sports domain and survey data (both completely unseen by the agent while iterating through the workflow). The reports generated on the two domains can be viewed here.

This link has the sample report generated using TADDR for a sports dataset (domain of cricket, which is a famous sport in the UK and the Indian subcontinent and Australia).

This link has the sample report generated on a survey dataset downloaded from Stats New Zealand [4].

The actual output of TADDR generated for the challenge can be accessed from **here**.

4 Conclusion

This work introduces **TADDR**, a novel framework for automated, agent-driven data storytelling. By decomposing the visualization process into modular agents, TADDR emulates the reasoning of human analysts while leveraging LLM-based autonomy for adaptability. Preliminary results demonstrate generalizability across datasets and domains, though future research is required to optimize narrative coherence, integrate multimodal data, and extend visualization libraries.

The generated output of TADDR submitted for the conference is available at https://www.visagent.org/api/output/93fb72e4-9b08-4bc9-bfb4-31d633bcd4e5.

The full agentic implementation code can be accessed from my github repository.

5 References

[1] VIS Agentic Challenge 2025. [Online]. Available: https://www.visagent.org/

[2] Plotly Python Graphing Library. [Online]. Available: https://plotly.com/python/basic-charts/

[3] LangGraph Tutorials: Evaluator-Optimizer Workflows. [Online]. Available: https://langchain-ai.github.io/langgraph/tutorials/workflows/#evaluator-optimizer

[4] Survey data downloaded from https://www.stats.govt.nz/large-datasets/csv-files-for-download/