

Can LLMs replace a business analyst?

AND, THE SUCCESS RATE IS...

We found that LLM SQL query generation for real business use cases such as querying sales data was 75% successful. The query did particularly well on continuous variables, and with aggregate and filter functions. However, we had to re-engineer the prompt to handle "categorical variable" better.

Further, most users showed comfort in asking questions, and trusting the results once the Slackbot got 2-3 right answers and built initial trust.

INTRODUCTION

BI tools don't support open-ended queries, but with the advancement of Large Language Models (LLMs), natural language based interfaces can help business users run data queries from their communication tools such as Slack itself. This can significantly improve data based decision-making in organizations. However, we are not sure whether such tools will work accurately, and if the business users are going to trust them.

Authors

Anshita Arya, Anukriti Goyal, Melissa Licari, Molly Zhang, Richa Verma, Master of Information Management System,'23 School of Information,

Advisor:

Dr. Aditya Parameswaran, School of Information



OBJECTIVE

This research tries to answer the following questions:

- (a) will business users' actually trust a natural language query based system to make decisions, and
- (b) what are the querying/exploration limitations of such a tool?

METHODOLOGY

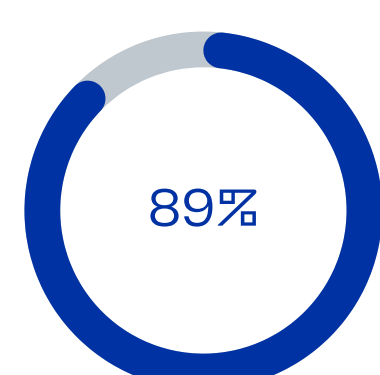
To understand this, we are using a Praxis approach where we:

1. Created a tool that users can ask data queries from Slack.
2. Analyzed queries run, to understand accuracy and potential limitations.
3. Informal interviews with users on how their experience was, and what they expect such a tool to do.

RESULTS/FINDING

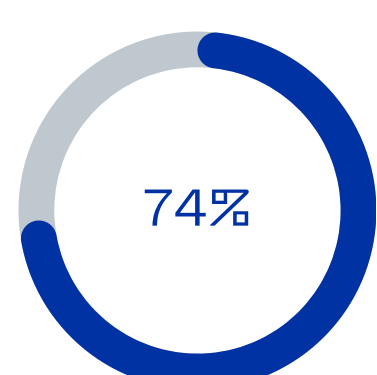
These results are based on the queries run by the research team and the interviewees who were business users. However, there can be some bias in data because of limited user set. We are expanding our interview pool further.

Questions answered



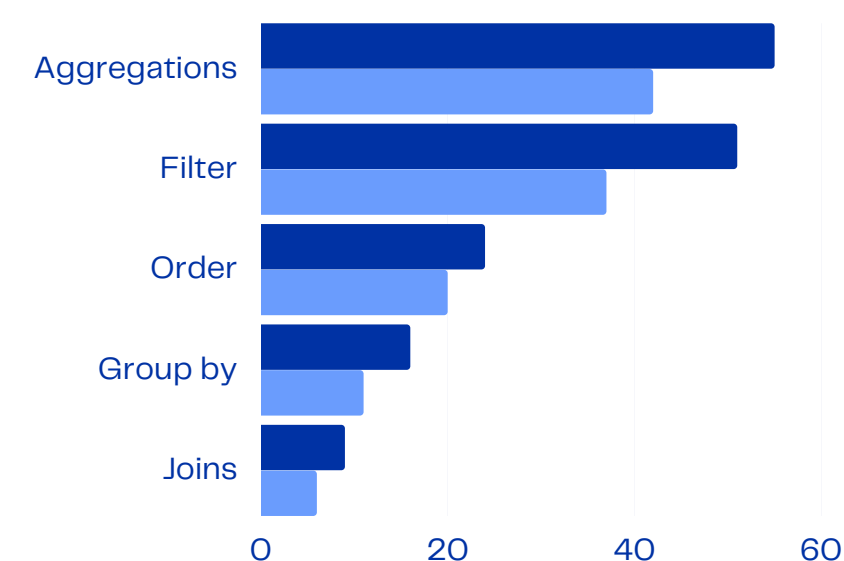
n = 61

Satisfactory answers

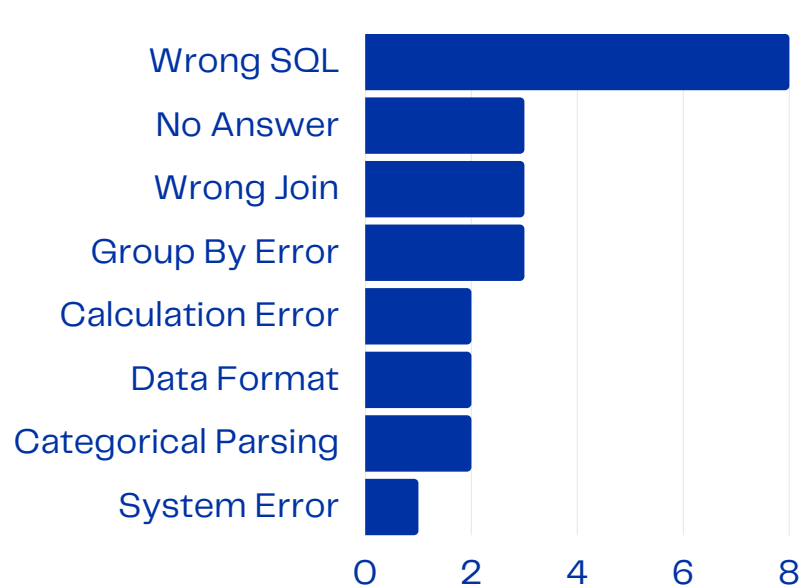


n = 54

How complex were the queries?

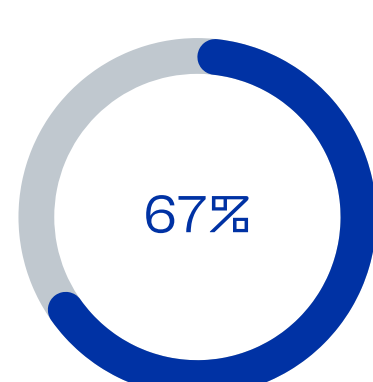


What kind of errors did the bot produce?



n = 24/61

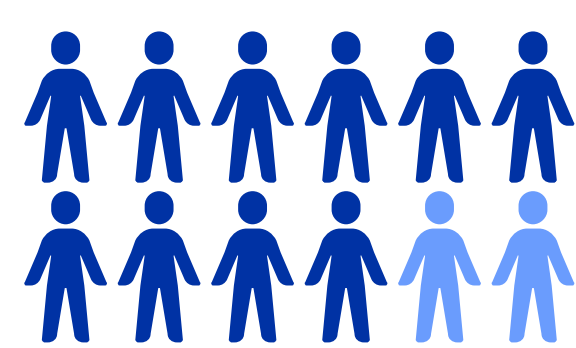
Making assumptions



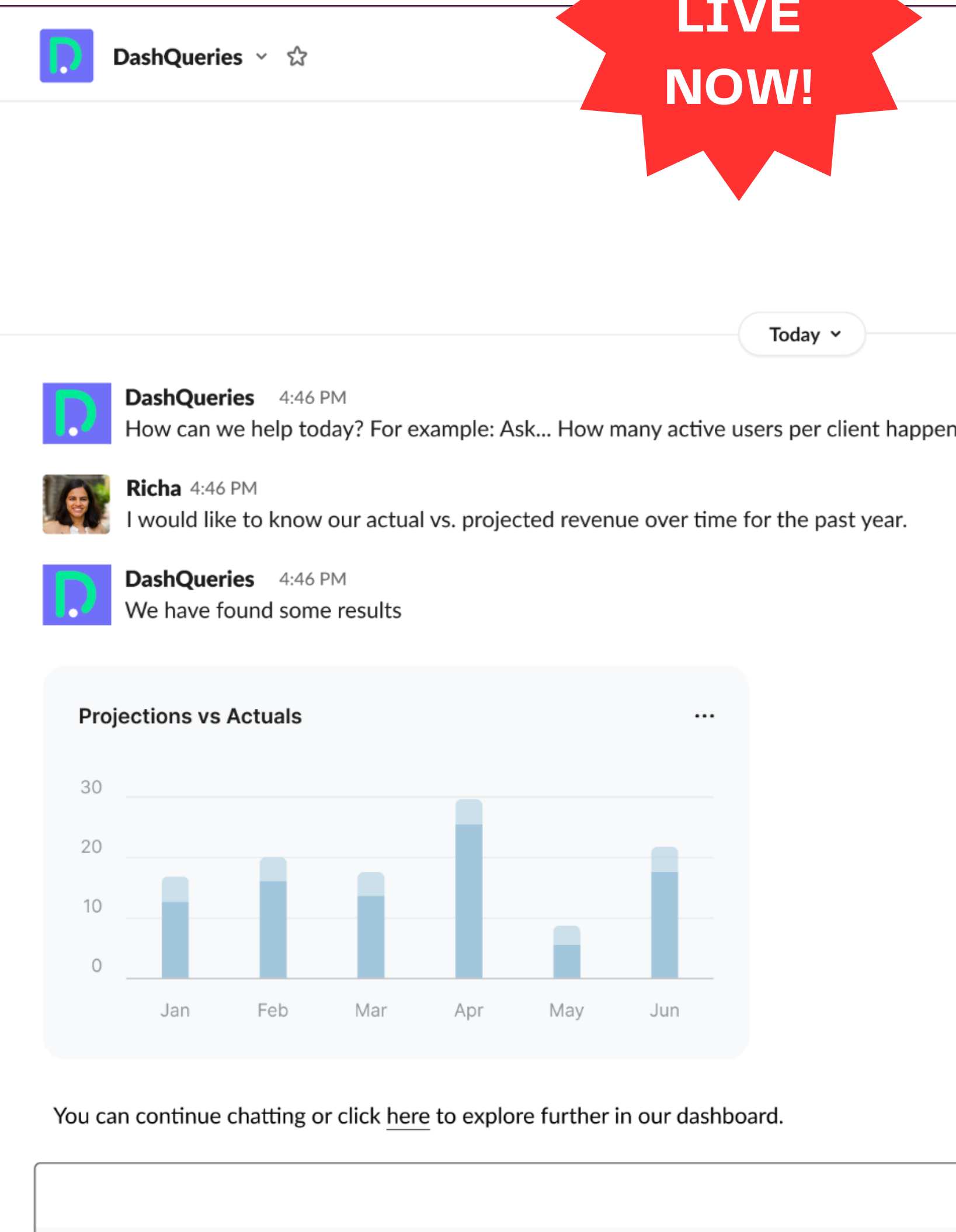
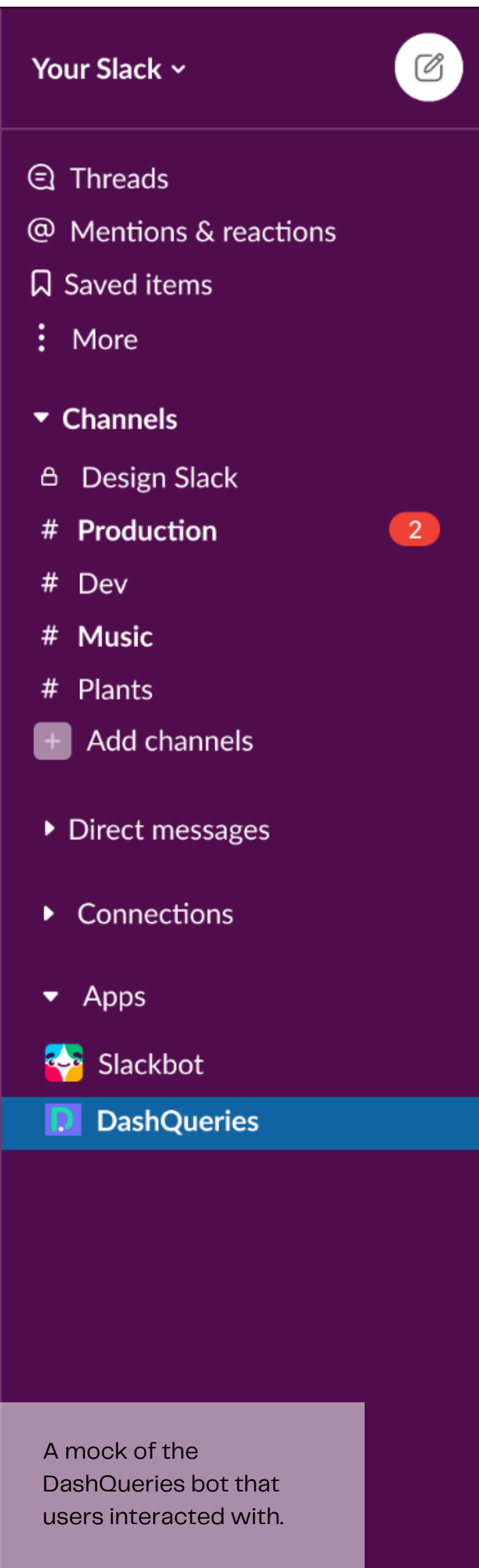
n = 9

Will the users trust the bot?

Depends on how accurately it answers the questions and produce results.



n = 12



OTHER FINDINGS

Engineering the prompt correctly matters, fixing the prompt to understand categorical variables and id variables variables better helped improve accuracy.

The LLM was able to generate some decent assumptions for representing vague queries (67% satisfaction) and handling dates as a unique case.

Getting feedback on whether the bot produced the correct result built trust with the user. Need to see whether that can help improve accuracy of the query generation.

CONCLUSION

LLMs are able to do a decent job of generating SQL query with prompt based on schema and a few rows of data. However, the chances that it will get a query right is roughly 75%. We still need to test if that's good enough and what changes do users' expect in such an interaction.

Further, we need to test how the bot performs on some production databases. We are doing POCs on some real company data.

Top comments from interviewees - positive & constructive.

"Love it! Super simple, yet super powerful! Especially helpful for business users who want to focus on their work"

"It will be great if I can get to visualize this data or play around in an Excel like interface."

"This can be a first line of defense for any data analyst team. Can kill 80% of their backlog."

"How can the app tell me that I should ask a specific question?"

"This is really going to eliminate my job."

"Can the app tell me that my answer cannot be answered then answering it incorrectly?"

- A data analyst

