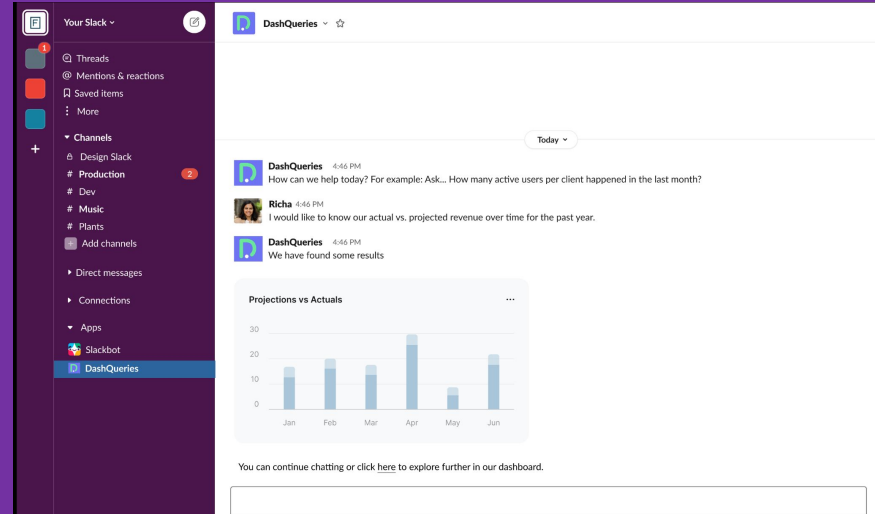


DashQueries

How can we enable business users to get to “actionable insights” fast?

Prof. Aditya Parameswaran

Anshita Arya, Anukriti Goyal, Melissa Licari, Molly Zhang, Richa Verma



The screenshot shows a Slack interface with a channel named 'DashQueries'. The channel contains three messages:

- DashQueries** (4:46 PM): How can we help today? For example: Ask... How many active users per client happened in the last month?
- Richa** (4:46 PM): I would like to know our actual vs. projected revenue over time for the past year.
- DashQueries** (4:46 PM): We have found some results.

Below the messages is a bar chart titled 'Projections vs Actuals' showing data for the months of Jan, Feb, Mar, Apr, May, and Jun. The y-axis represents a numerical value from 0 to 30. The bars show the following approximate values: Jan (15), Feb (18), Mar (15), Apr (25), May (10), and Jun (20).

At the bottom of the channel view, there is a link: "You can continue chatting or click [here](#) to explore further in our dashboard."

Business users have many data questions...

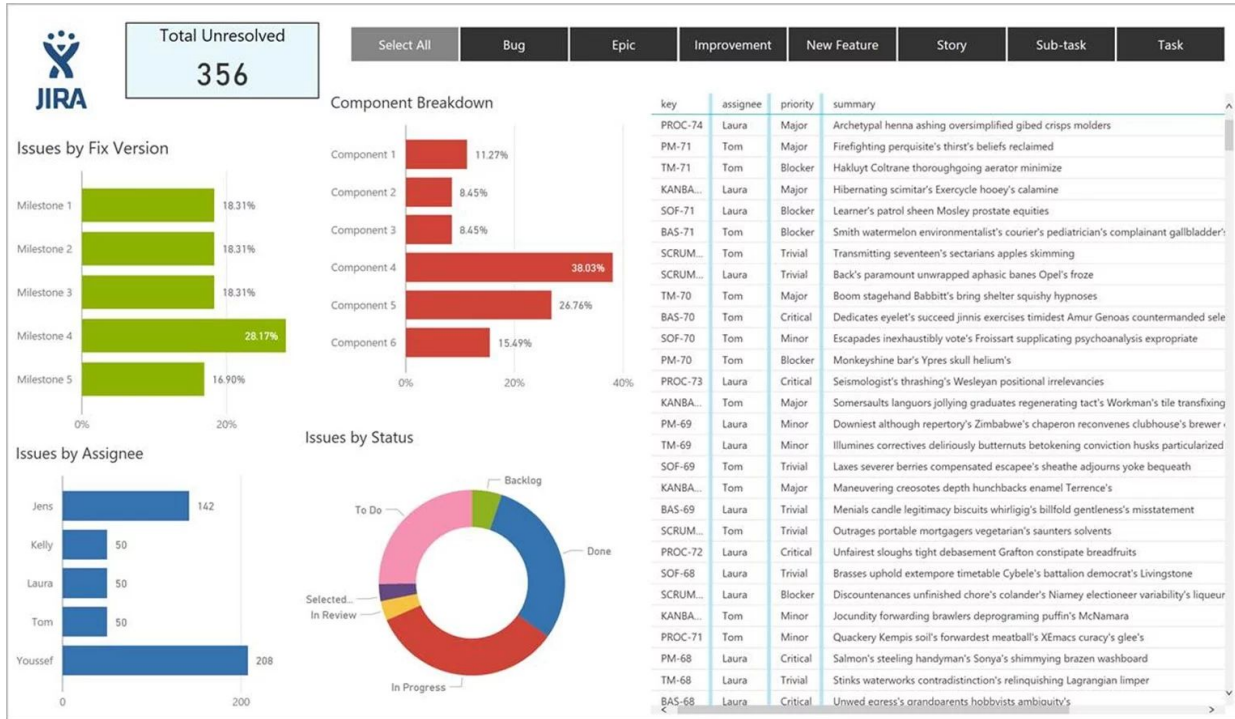
“Is any of my customer having more “Active Users” than usual?”

“What’s the last time this customer paid an invoice?”

“Is my metric really going up?”

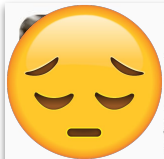


But... rather than getting them answered.. They get “triaged”



And, data teams hate answering these one-off questions..

Which means that these queries get unanswered..



Richa, Product Manager 4:22 PM

Hey, can you tell me the number of opportunities broken down by city?



Molly, Data Analyst 4:24 PM

Sorry, unfortunately, we have a backlog for 2-weeks, can you look at this dashboard: [bad-dashbaord-that-doesnt-work.tableau.com/c/random-product-insights-no-one-want](https://work.tableau.com/c/random-product-insights-no-one-want). It might help!

82%

Data Projects Fail

(Gartner, 2019, 2020, 2021)



Hence, we created “DashQueries”

A SlackBot that helps answer business questions in seconds, rather than days.

Richa Verma 11:47 AM
@DashQueries tell me about average sales by city?

DashQueries APP 11:47 AM
Runnin' the numbers, gimme a sec

Finding the right data for your question: tell me about average sales by city?

In the meantime, go back to your work, I'll ping you in this channel once the answer is ready.

Here's the answer I got to your question: tell me about average sales by city?

city	avg_sales_velocity
Bengaluru	43.8899
Chennai	44.2124
Delhi	45.9489
Hyderabad	44.1106
Kolkata	42.9662
Mumbai	45.1368
Pune	39.364

And the SQL I used:

```
SELECT city, AVG(sales_velocity) AS avg_sales_velocity
FROM sales_data
GROUP BY city
ORDER BY city NULLS LAST;
```

Which basically means this in plain English:

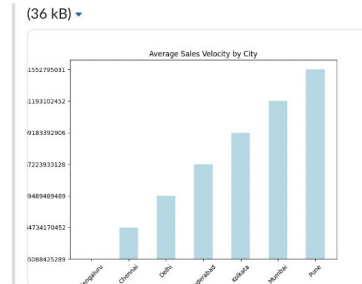
```
--This query will calculate the average sales velocity by city
```

Are you satisfied with this answer?

[Looks good](#) [Data seems wrong](#) [Query seems wrong](#)

Tip: T

Here's the chart I got for you: <https://dashqueries.s3.us-west-2.amazonaws.com/chart-1683139745.png>



And, we used “Praxis” method to understand...

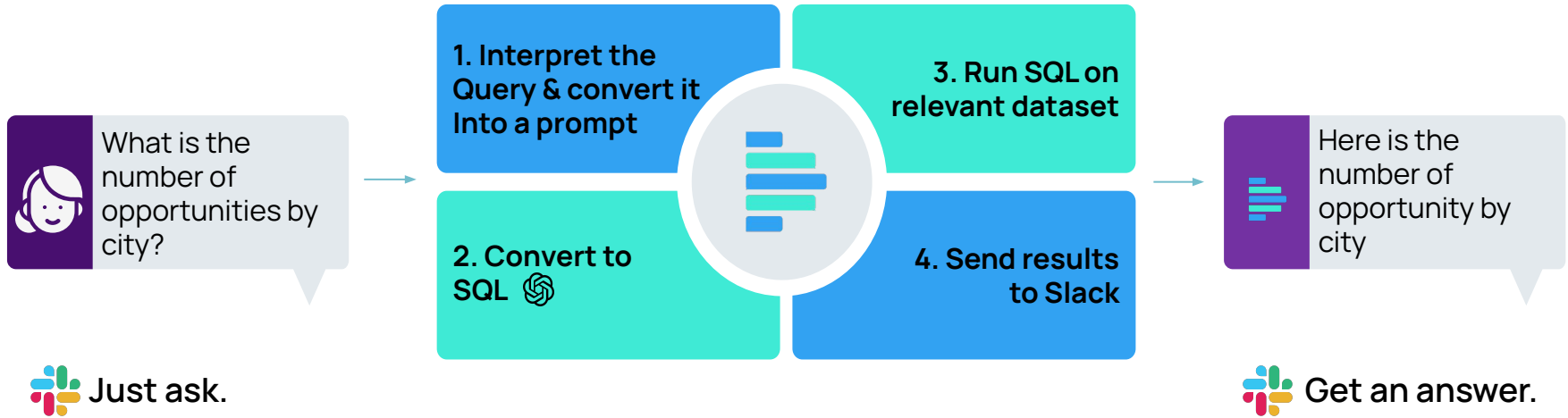
1. Will business users’ actually trust a natural language query based system to make decisions?
2. What are the querying / exploration /scalability limitations of such a tool?
 - 300+ queries ran.
 - 15 user tests in two phases.
 - 40+ informal demos in conferences, etc.



Live Demo



How DashQueries work



Phase 1

Get the bot out of door

R **Richa Verma** 1:38 PM
@DashQueries tell me how many opportunities were won in Delhi that had an opportunity size of more than 60K

DashQueries APP 1:38 PM
Your question is: tell me how many opportunities were won in Delhi that had an opportunity size of more than 60K

1:39 Here's what I found: [{"sql": "\n\nSELECT COUNT(*) \nFROM sales_data \nWHERE city \nLIKE 'Delhi' AND opportunity_size > '60K' AND opportunity_status = 'Won' \nGROUP BY \nopportunity_size \nORDER BY opportunity_size DESC NULLS LAST;"}, {"result": [{"count": 139}], "error": null}

Anshita Arya 10:04 PM
@DashQueries how long were the past sales cycles

DashQueries APP 10:05 PM
Lmc, hold tight

DashQueries APP 10:05 PM
Your question is: how long were the past sales cycles

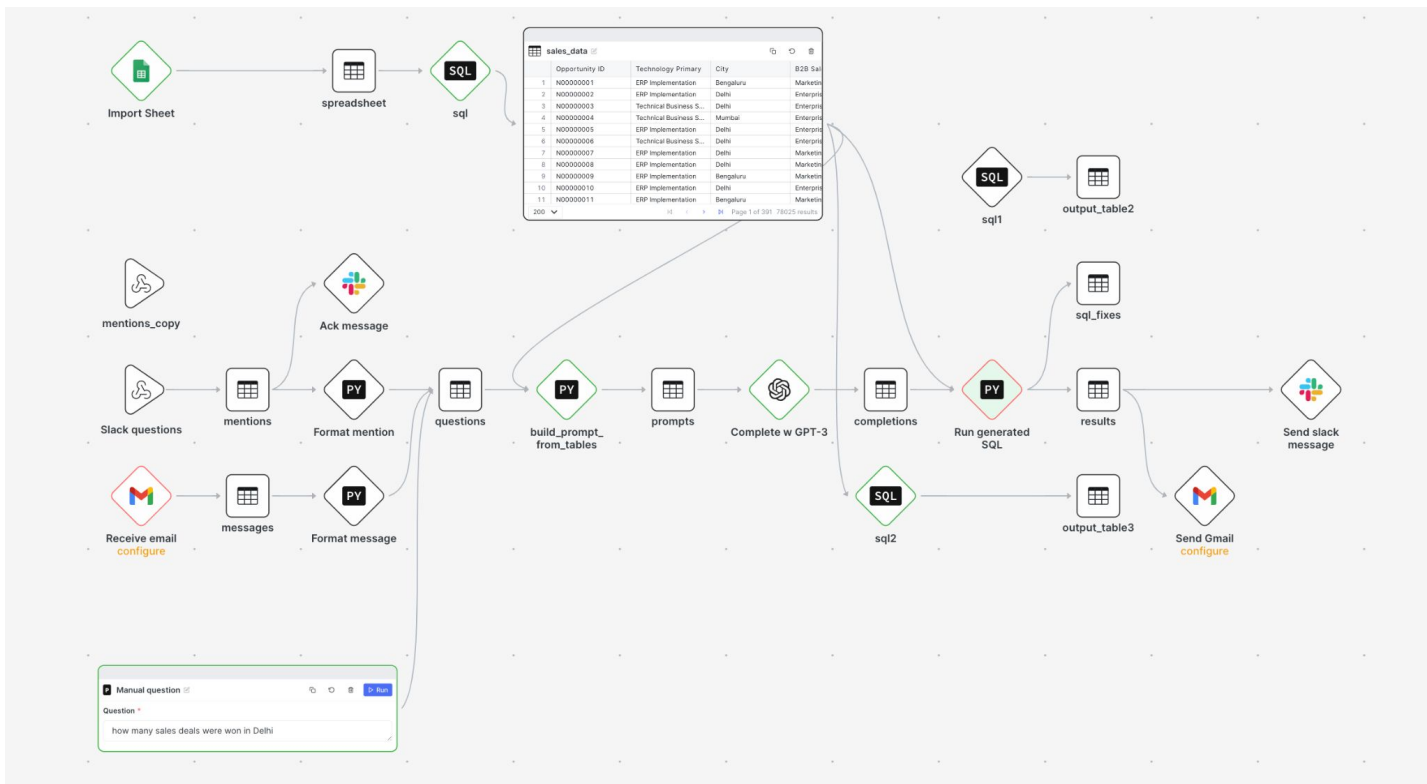
DashQueries APP 10:05 PM
Here's the answer I got:

```
avg_time_days
-----
44685.2
```

And the SQL I used:

```
SELECT AVG(last_updated - sales_velocity) AS avg_time_days
FROM sales_data
WHERE opportunity_status ILIKE 'Won'
ORDER BY avg_time_days DESC NULLS LAST;
```

Created a simple LLM pipeline



A working slackbot with basic UX

- Tested with 5 users
- And, internal testing ~100 questions

Questions asked:

- Can you actually tell me the number of opportunities that have sales velocity more than 70 and client revenue more than 1M?
- Tell me how many opportunities were won in Delhi that had an opportunity size of more than 60K
- Why are deals won in Delhi more than Mumbai?
- Where are the largest opportunity sizes
- How many clients have done repeated sales with

Richa Verma 9:34 AM
@DashQueries tell me which city has the most opportunities won?

DashQueries APP 9:34 AM
Runnin' the numbers, gimme a sec

I am trying my best to answer your question: tell me which city has the most opportunities won?

How I work? 🤖

It's simple, I take your question and interpret (1) table you are curious about, and (2) write an SQL query for it.
Once the results are ready, I get them back to you.

Kind of questions I can answer:

1. Tell me whether sales have increased in a <region>
2. Which opportunities were won/lost?
3. Tell me the average sales for the last quarter/month?
4. Tell me which product has been selling the most?

Questions I can't answer:

1. Why sales changed?
2. Predict the sales in future..
3. What's the purpose of life?

In the meantime, go back to your work, I'll ping you in this channel once the answer is ready.

Here's the answer I got:

city	opportunitieswon
Mumbai	5333

And the SQL I used:

```
SELECT city, COUNT(*) AS OpportunitiesWon
FROM sales_data
WHERE opportunity_status ILIKE 'won'
GROUP BY city
ORDER BY OpportunitiesWon DESC NULLS LAST
LIMIT 1;
```

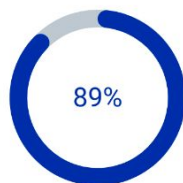
Which basically means this in plain English:

```
-- This query returns the city with the most Opportunities Won.
```

Are you satisfied with this answer? (Yes or No)

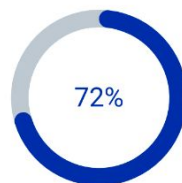
What did we learn?

Questions answered



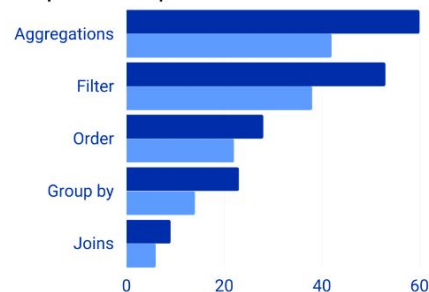
n = 69

Satisfactory answers



n = 62

How complex were the queries?

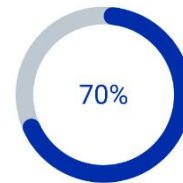


What kind of errors did the bot produce?



n = 29/69

Making assumptions



n = 10

Will the users trust the bot?

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



n = 12

This was us...



60% of the time it works every time

What did we learn?

Trust Increased With Usage

- First impressions were key to building trust
- Data users were more skeptical than non-technical users
- *"I like that it shows me the SQL code so that I can run it myself to check it"*

Users Want a Feedback Loop

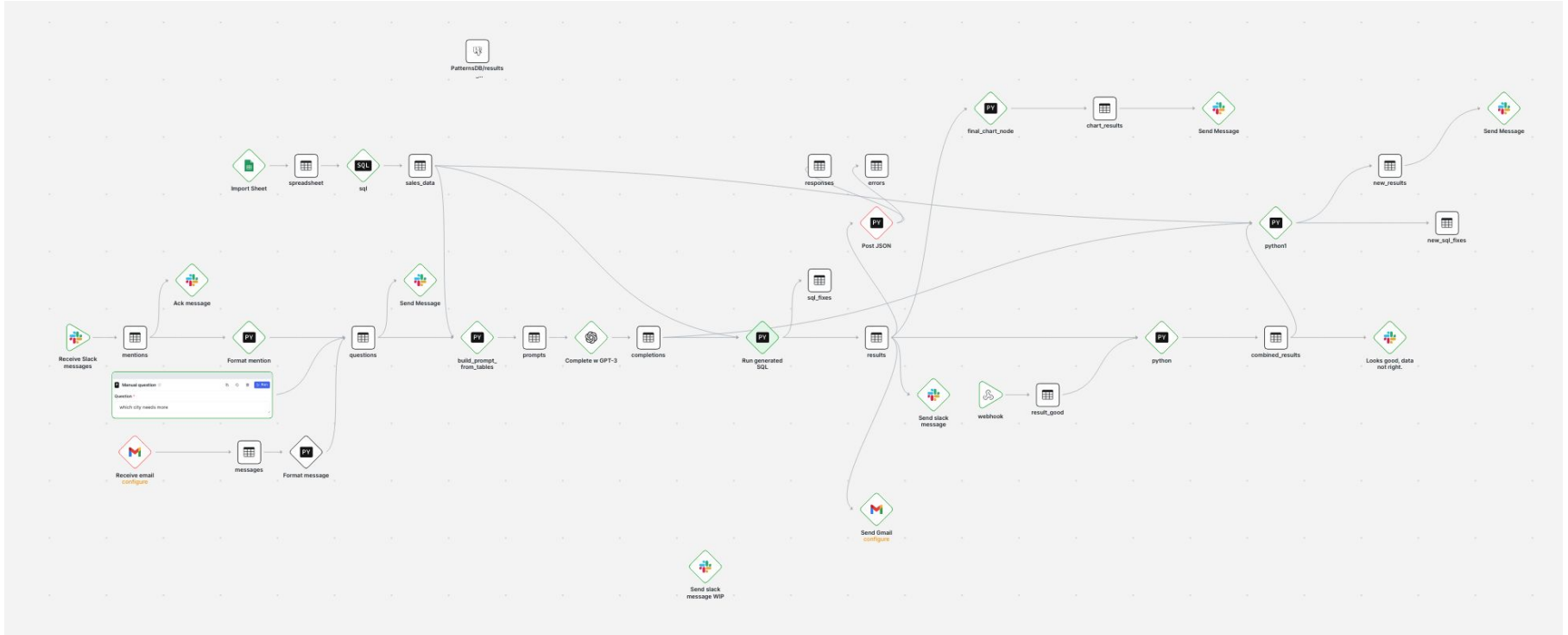
- Some responses were not optimal/what the user was looking for and the user would have to rephrase their question for a better response.
- Users wanted a way to tell us whether the response was satisfactory or not.

Phase 2

Get the bot out of door



Made our bot more robust



Improvements

1. UX Improvements


- Number and frequency of slack messages bot sent
- Feedback loop (Looks Good, Wrong Data, Wrong Query)
- Chart response
- Length of the results table

2. Result Quality improvements

- Categorical parsing
- Recognition of previous vs. current query
- Error fixes

```
SELECT opportunity_id, client_revenue_sizing, opportunity_size_usd
FROM sales_data
WHERE client_revenue_sizing >= '500K'
GROUP BY opportunity_id, client_revenue_sizing, opportunity_size_usd
ORDER BY COUNT(opportunity_id) DESC NULLS LAST;
```

 **Richa Verma** 7:51 PM
@DashQueries how many opportunities have a size greater than 500k?

 **DashQueries APP** 7:52 PM
whipping it up for you

Finding the right data for your question: how many opportunities have a size greater than 500k?

In the meantime, go back to your work, I'll ping you in this channel once the answer is ready.

7:52 Here's the answer I got to your question: how many opportunities have a size greater than 500k?

```
count
-----
14680
```

And the SQL I used:


```
SELECT COUNT(*)
FROM sales_data
WHERE opportunity_size_usd > 500000
OR (client_revenue_sizing ILIKE 'More than 1M'
OR client_revenue_sizing ILIKE '500K to 1M'
OR client_revenue_sizing ILIKE '250K to 500K')
OR (opportunity_size ILIKE 'More than 100K'
OR opportunity_size ILIKE '50K to 100K'
OR opportunity_size ILIKE '30K to 50K');
```

A responsive bot that handles data

- Tested with 10 users
- And, internal testing ~100 questions

Questions asked:

- Please give me a line graph of the client revenue sizing compared to the client employee sizing
- What are my biggest revenue streams?
- Tell me about year to year growth of opportunities?
- How much business was done last year in all cities?
- In which technology do we have the highest potential for revenue increase

 **Molly** 12:58 PM
@DashQueries what are my biggest revenue streams?

 **DashQueries APP** 12:58 PM
Runnin' the numbers, gimme a sec

Finding the right data for your question: what are my biggest revenue streams?

In the meantime, go back to your work, I'll ping you in this channel once the answer is ready.

Here's the answer I got to your question: what are my biggest revenue streams?

	client_revenue_sizing	sum
100K or less	1763516989	
More than 1M	197752656	
500K to 1M	154573140	
250K to 500K	127992201	
100K to 250K	103599025	

And the SQL I used:

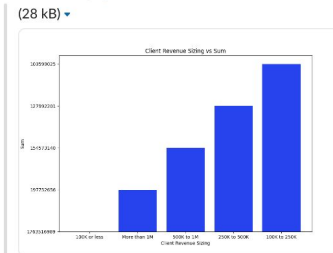
```
SELECT client_revenue_sizing, SUM(opportunity_size_usd)
FROM sales_data
WHERE opportunity_size_usd IS NOT NULL
GROUP BY client_revenue_sizing
ORDER BY SUM(opportunity_size_usd) DESC NULLS LAST;
```

Which basically means this in plain English:

--This query uses the opportunity_size_usd column from the sales_data table to determine the biggest revenue streams by summing up the total revenue for each client_revenue_sizing.

Are you satisfied with this answer?

Here's the chart I got for you: <https://dashqueries.s3.us-west-2.amazonaws.com/chart-1683143968.png>

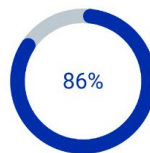


What did we learn?

Query Analysis

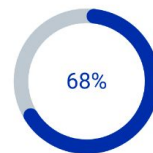
- Specific queries yielded better results
- Most errors were wrong variable selection
- Performed well for complex queries
- Slow response time, but faster than asking a person

Questions answered



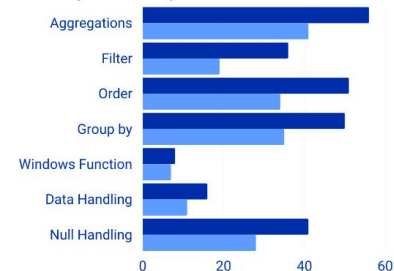
n = 77

Satisfactory answers

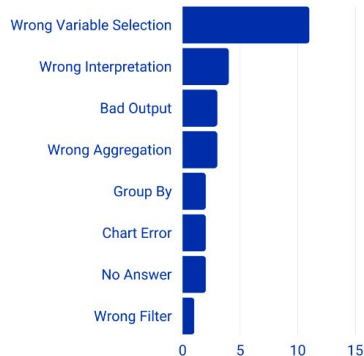


n = 66

How complex were the queries?

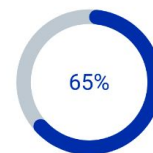


What kind of errors did the bot produce?



n = 33/77

Making assumptions



n = 31/48

Will you ask the bot the questions instead of a person?

8/10 users agreed to use it for quick answers and not ask a person instead.

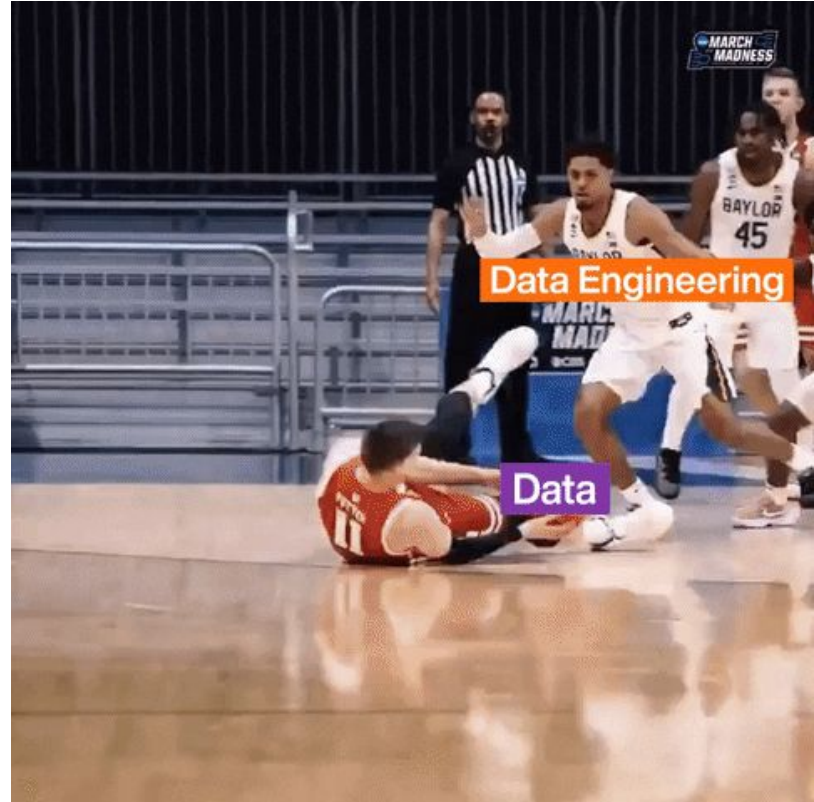


n = 10

Learning & Conclusion



Diverse User
Personas



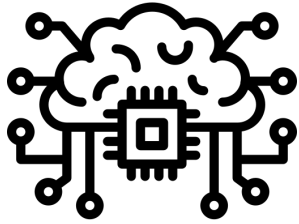
Learning & Conclusion



Building
Trust



Learning & Conclusion



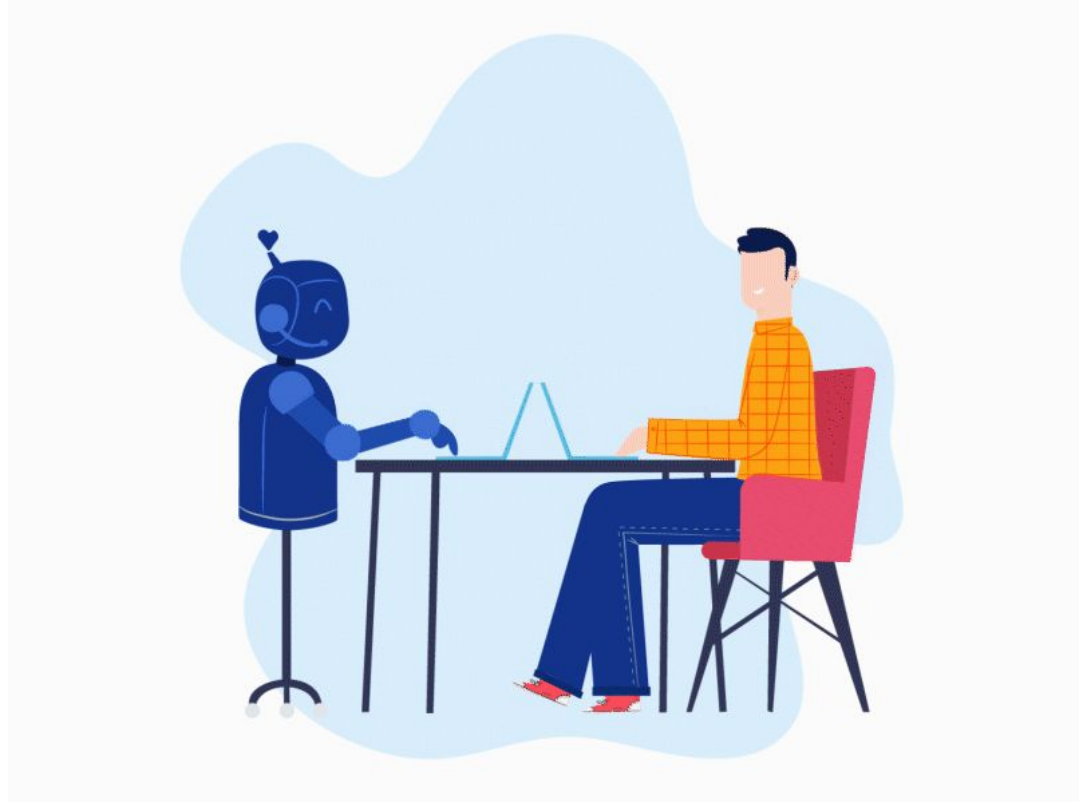
Capabilities
of LLM



Future Work



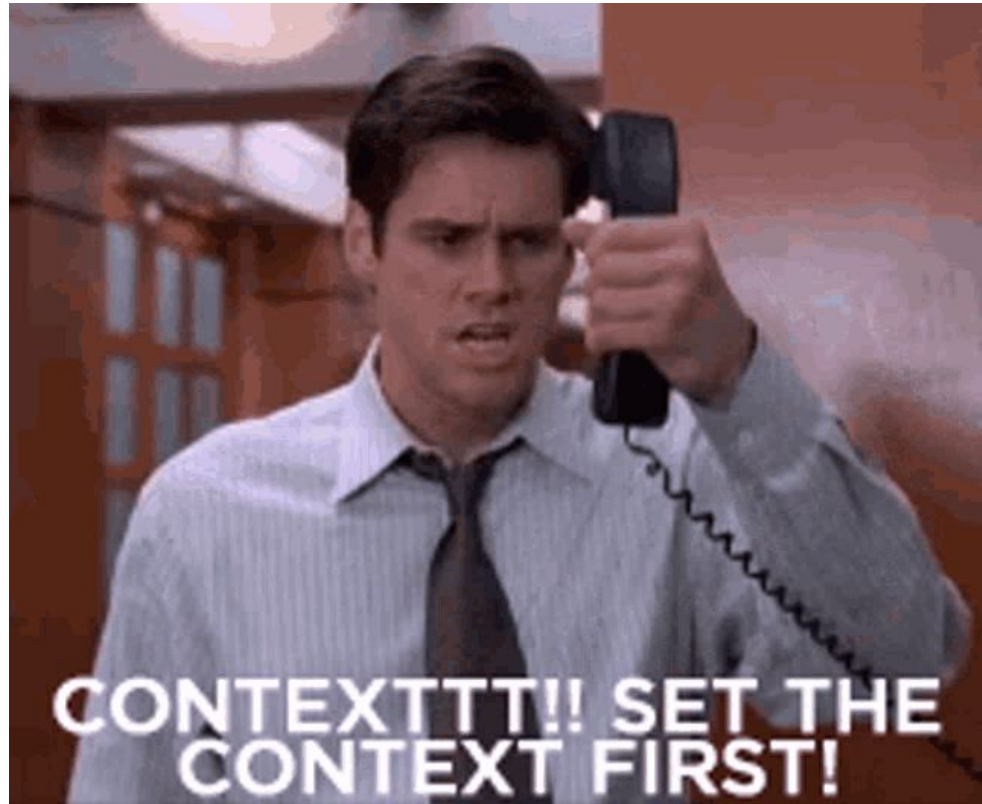
Interactive
Feedback



Future Work



Semantic
Context



Future Work



Extension to BI,
Productivity Tools
(Zoom, Meet,
Email)

Q1 Performance - Saved

Search

File Home Insert Formulas Data Review View Automate Help

Comments Catch Up Editing Share

Aptos (Body) 11 B General

A1 fx Country

Country	Customer	Product	Discount Band	Units Sold	Manufacturing Price	Sale Price	Gross Sales	Discounts	Sales	COGS
France	VanArsdel	Biodegradable Cleaning Products	Low	4345	\$10	\$352	\$1,529,440	\$15,294	\$1,514,146	\$
USA	VanArsdel	Biodegradable Cleaning Products	Low	4245	\$10	\$352	\$1,494,240	\$14,942	\$1,479,298	\$
Australia	VanArsdel	Biodegradable Cleaning Products	None	4124	\$10	\$352	\$1,451,648	\$0	\$1,451,648	\$
Mexico	VanArsdel	Biodegradable Cleaning Products	Medium	3954	\$10	\$352	\$1,391,808	\$83,508	\$1,308,300	\$
France	Contoso Pharmaceuticals	Biodegradable Cleaning Products	Low	4952	\$10	\$127	\$628,904	\$6,289	\$622,615	\$
France	Contoso Pharmaceuticals	Recycled Can	Low	4903	\$5	\$127	\$622,681	\$6,227	\$616,454	\$
Canada	Proseware	Natural Personal Care Products	Low	1954	\$270	\$302	\$590,108	\$5,901	\$584,207	\$58
Canada	Proseware	Biodegradable Cleaning Products	High	1990	\$220	\$302	\$600,980	\$84,137	\$516,843	\$48
USA	VanArsdel	Compostable Food Packaging	Low	1254	\$3	\$380	\$476,520	\$4,765	\$471,755	\$
USA	Proseware	Natural Personal Care Products	Medium	1465	\$224	\$302	\$442,430	\$26,546	\$415,884	\$38
USA	Proseware	Sustainable Snack Packaging	Medium	1410	\$120	\$302	\$425,820	\$25,549	\$400,271	\$18
Australia	Proseware	Natural Personal Care Products	Medium	1369	\$224	\$302	\$413,438	\$24,806	\$388,632	\$31
USA	VanArsdel	Compostable Food Packaging	Medium	1052	\$3	\$380	\$399,760	\$23,986	\$375,774	\$
Canada	VanArsdel	Compostable Food Packaging	Low	980	\$312	\$380	\$372,400	\$3,724	\$368,676	\$58
Australia	Proseware	Sustainable Snack Packaging	Medium	1298	\$120	\$302	\$391,996	\$23,520	\$368,476	\$17
Australia	VanArsdel	Compostable Food Packaging	None	954	\$3	\$380	\$362,520	\$0	\$362,520	\$
Canada	Contoso Pharmaceuticals	Biodegradable Cleaning Products	Low	2785	\$110	\$127	\$353,695	\$3,537	\$350,158	\$23

Calculation Mode: Automatic General Workbook Statistics Give Feedback to Microsoft 100%

Questions?