

GLM DATA CHAT

Date: September 15, 2025

Project: Data Chat App with GLM AI

What I Did Today

I made a data chat app that lets people ask questions about their data in simple English. The app uses the GLM-4 model from Ollama to understand questions and give answers.

The Model I Used

GLM-4 Model

- Works really well for data questions
- Good at understanding what people ask
- Can handle math and make charts

Why Not GLM-4.5?

- GLM-4.5 is 200GB in size - too big for my computer!
- GLM-4 works fine and is much smaller
- GLM-4.5 has 355 billion parts, GLM-4 has much less

What My App Can Do

The app lets users:

- Upload CSV files with data
- Ask questions in plain English
- Get answers with numbers and charts
- See SQL code that was made automatically

Questions That Worked Well

From my screenshots, these worked great:

Customer Info:

- "Which education level is most common?" → Secondary (2396 people)
- "Which month had most contacts?" → May (1398 contacts)

Health Data:

- "What's the highest number of pregnancies?" → 17
- "What's the average age?" → 33.24 years
- "What's the median blood pressure?" → 72

Money Stuff:

- "Average balance of married customers?" → €1463.20
- "How many have housing loans?" → 691 people

Problems I Found ⚠️

Easy questions: Work perfectly

Hard questions: Some work, some don't

Very complex questions: Need more work tomorrow

Screenshots:

The screenshot shows the 'Data Analysis Chat' interface. On the left, there's a sidebar with 'Model Settings' (GLM-4 9B) and 'Upload Dataset'. The main area displays a chat conversation. The first message is a system prompt: 'The minimum account balance in the dataset is the lowest amount recorded among all the accounts. This figure can help you understand the starting point of financial health for the customers in your dataset, and it may indicate the threshold below which some accounts might be considered at risk or in need of assistance.' The second message is a user query: 'Which job category has the highest number of customers?'. The third message is the AI response, which includes a SQL query:

```
SELECT job, COUNT(*) AS count
FROM data
GROUP BY job
ORDER BY count DESC
LIMIT 1
```

 and a table result showing 'management' with a count of 909. The fourth message is another system prompt: 'The job category with the most customers is at the top of the list. It's clear that this particular job category is quite popular and has a significant customer base compared to others!'. The fifth message is a user query: 'Which marital status is most common among customers?'. The sixth message is the AI response, which includes a SQL query:

```
SELECT marital, COUNT(*) AS count
FROM data
GROUP BY marital
ORDER BY count DESC
LIMIT 1
```

 and a table result showing 'married' with a count of 2787. At the bottom, there's a text input field with the placeholder 'Ask a question about your data...' and a send button.

The screenshot shows the 'Data Analysis Chat' interface. The first message is a user query: 'Which education level is most frequent among customers?'. The second message is the AI response, which includes a SQL query:

```
SELECT education, COUNT(*) AS frequency
FROM data
GROUP BY education
ORDER BY frequency DESC
LIMIT 1
```

 and a table result showing 'secondary' with a frequency of 2206. The third message is a system prompt: 'The most common education level among customers is the one with the highest count in our dataset. This suggests that a significant portion of our customer base has this particular educational background, which could be an interesting demographic to target for marketing or product development purposes.' The fourth message is a user query: 'Which month has the highest number of contacts?'. The fifth message is the AI response, which includes a SQL query:

```
SELECT month, COUNT(*) AS contact_count
FROM data
GROUP BY month
ORDER BY contact_count DESC
LIMIT 1
```

 and a table result showing 'may' with a contact count of 1388. The sixth message is another system prompt: 'The month with the most contacts is the clear winner! This month really stands out from the rest with a significantly higher number than any other. It's like it's the busiest time for reaching out, so it might be worth focusing on what makes this period so special or planning around it to maximize engagement.' The seventh message is a user query: 'What is the average balance of married customers?'. At the bottom, there's a text input field with the placeholder 'Ask a question about your data...' and a send button.

Model Settings
Select Model
GLM-4 9B - More accur...
Upload Dataset

Basic Analytics

What is the maximum number of pregnancies recorded?

```
SELECT MAX(Pregnancies) FROM data;
```

Max(Pregnancies)

17

The highest number of pregnancies recorded in the dataset is quite impressive! It's a significant figure that highlights either an exceptional case or a pattern worth investigating. This could indicate a unique situation for the individual involved or possibly a trend within the group being studied.

What is the median blood pressure?

```
SELECT APPROX_QUANTILE(BloodPressure, 0.5) AS MedianBloodPressure FROM data;
```

MedianBloodPressure

72

The median blood pressure from the dataset is right in the middle of all recorded values. It's a good indicator that half of the people have higher blood pressures and half have lower ones, providing a central tendency for this health metric.

How many patients have glucose level greater than 150?

```
SELECT COUNT(*) FROM data WHERE Glucose > 150;
```

Count(*)

140

Ask a question about your data...

Model Settings
Select Model
GLM-4 9B - More accur...
Upload Dataset

```
SELECT UNNEST(CORR(*)) AS correlation FROM data ORDER BY correlation DESC LIMIT 1;
```

What is the average glucose level for diabetic vs non-diabetic patients?

```
WITH diabetic AS (
  SELECT AVG(Glucose) as avg_diabetic FROM data WHERE Outcome = 1
), non_diabetic AS (
  SELECT AVG(Glucose) as avg_non_diabetic FROM data WHERE Outcome = 0
)
SELECT avg_diabetic, avg_non_diabetic FROM diabetic, non_diabetic;
```

avg_diabetic	avg_non_diabetic
0	141.2575 109.98

What is the average BMI for diabetic vs non-diabetic patients?

```
WITH diabetic AS (
  SELECT AVG(BMI) as avg_bmi_diabetic FROM data WHERE Outcome = 1
), non_diabetic AS (
  SELECT AVG(BMI) as avg_bmi_non_diabetic FROM data WHERE Outcome = 0
)
SELECT avg_bmi_diabetic, avg_bmi_non_diabetic FROM diabetic, non_diabetic;
```

avg_bmi_diabetic	avg_bmi_non_diabetic
0	30.1425 30.3042

Ask a question about your data...

Tomorrow's Plan

1. Fix hard questions that don't work yet
2. Make app faster for big data files
3. Add more chart types for better visuals
4. Test more complex data analysis

What I Learned

- GLM-4 is great for data questions
- Simple questions work better than complex ones
- People can now ask data questions without knowing SQL
- The app saves lots of time for data analysis

