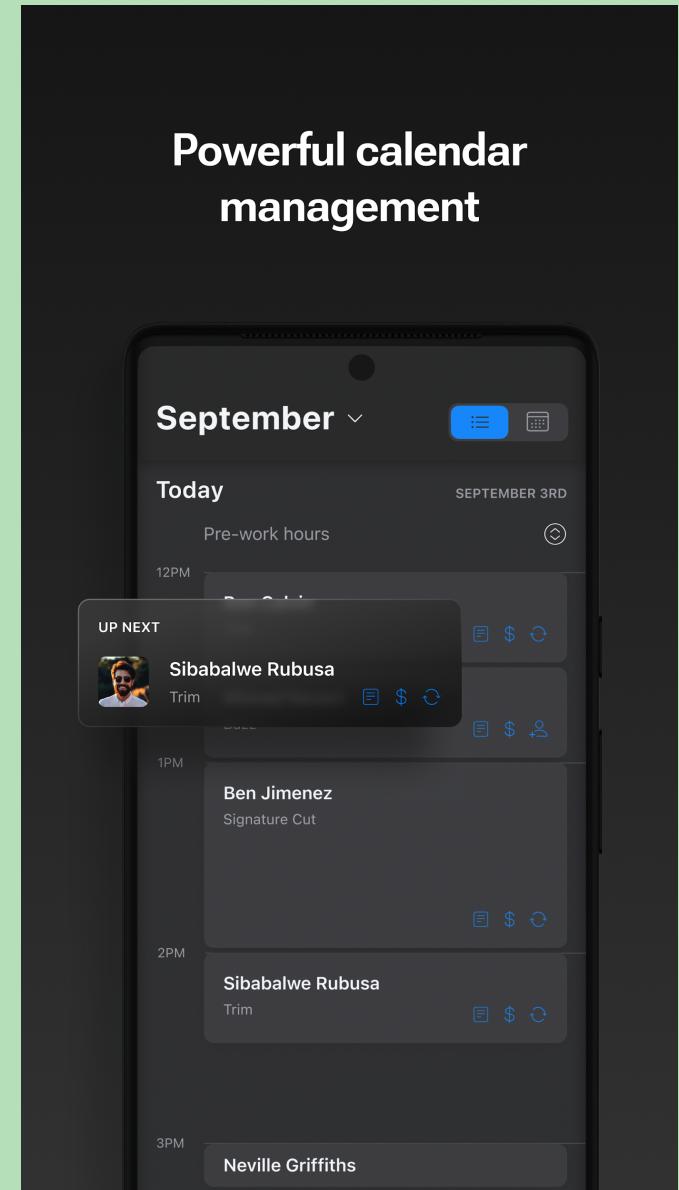
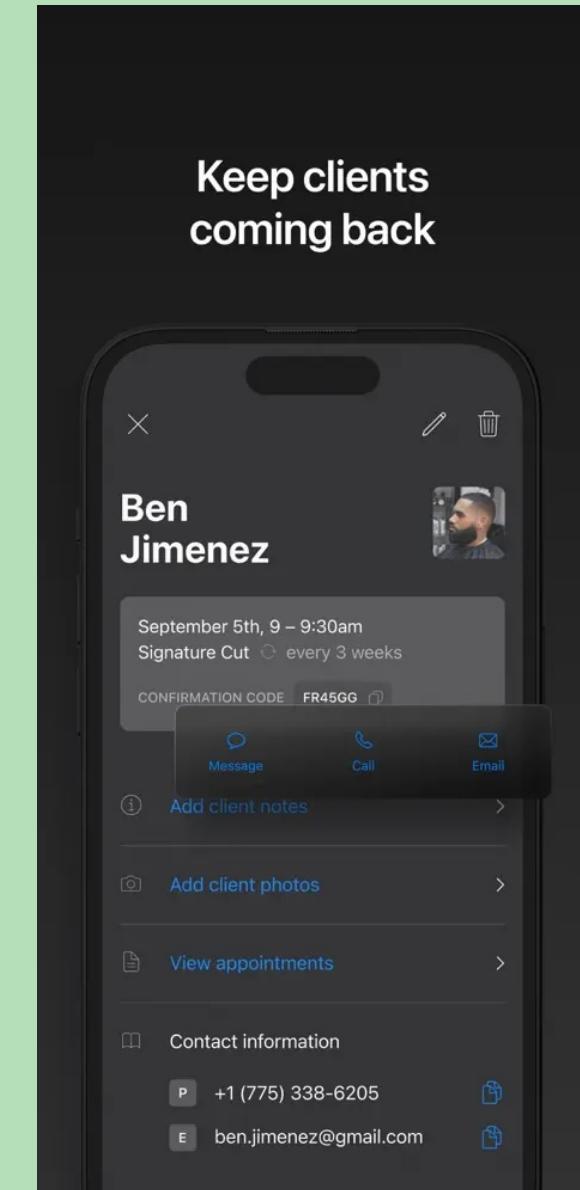
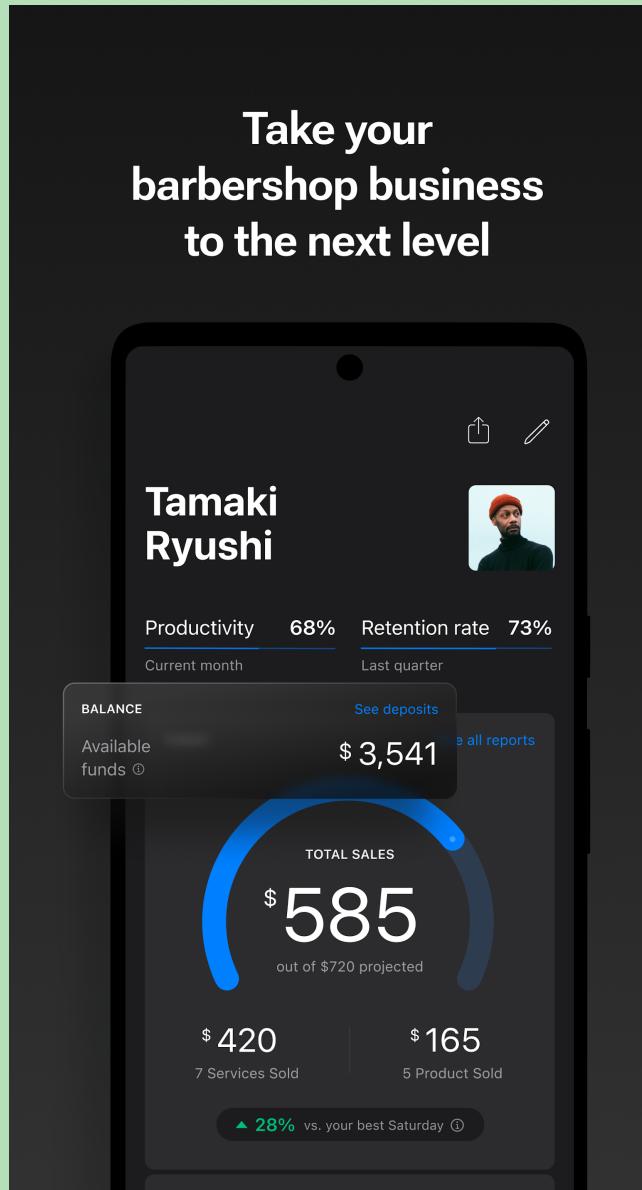
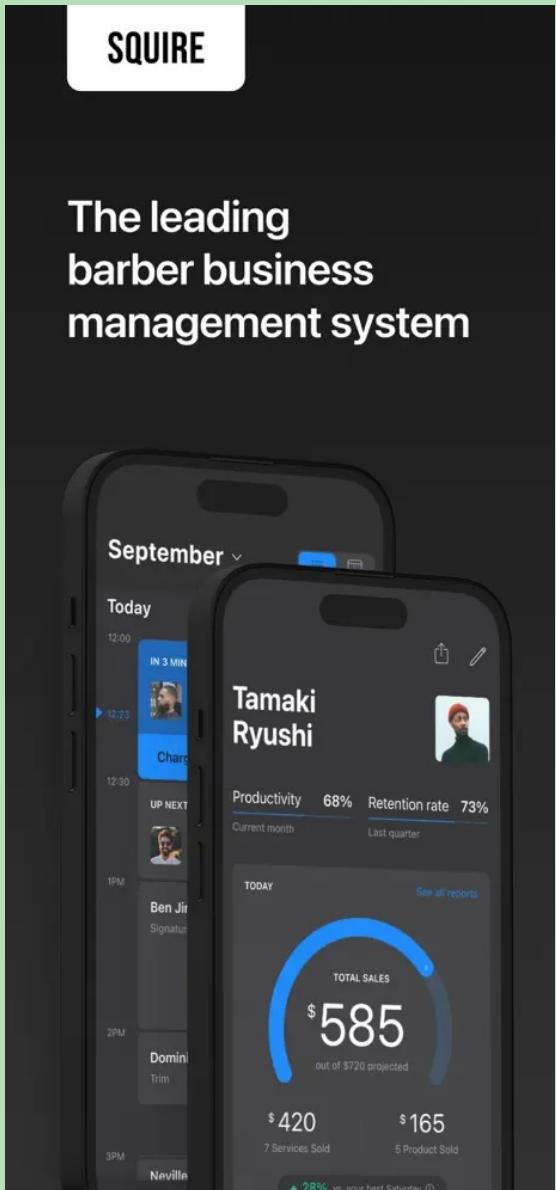


# Forget SQL! Create your own AI Data Analyst

Artur Kuzmin

# Squire



# Questions

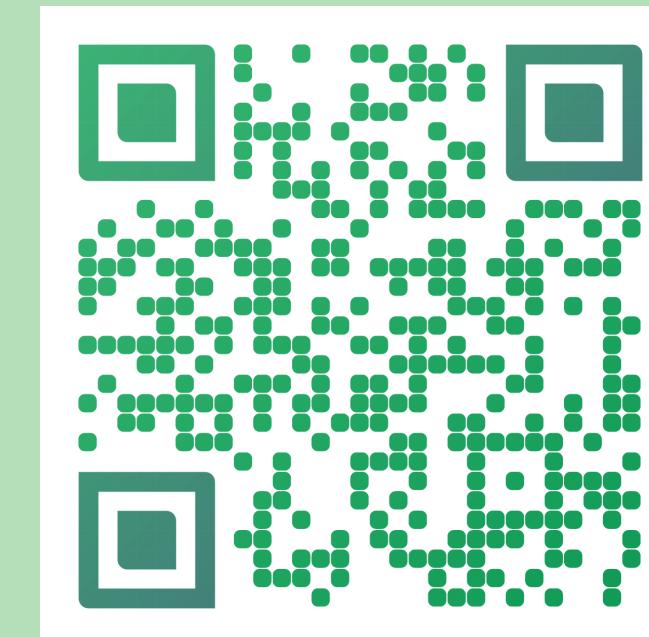
- “How many rental barbers we've got in the last 30 days?”
- “I need a test account: a commission barber in New York with over 30 bookings last week.”
- “What are the most common types of SMS we send?”
- “Find me a barber with email [someemail@gmail.com](mailto:someemail@gmail.com)”

A close-up selfie of Artur Kuzmin, a man with short brown hair, smiling at the camera. He is positioned next to a kangaroo, which is partially visible on the left side of the frame. They are outdoors, with a sandy beach and the ocean visible in the background under a clear blue sky.

# Artur Kuzmin

Director of Engineering at Squire,  
Google Developer Expert

[linkedin.com/in/arturkuzmin/](https://linkedin.com/in/arturkuzmin/)  
[twitter.com/ArtursTwit](https://twitter.com/ArtursTwit)



Solution

AI Data Analyst you can chat with

# The goal of this talk

- Build an AI Data Analyst that:
- knows your db structure
- knows your companies context
- can answer questions about your data

# Tooling highlights:

- Google Ai Studio
- Bigquery
- Kaggle
- Google Colab



zon  
IFT



io



Exasol



ORACLE  
DATA BASE



ake



[ROCKSET]

VERTICA



data.



Timescale



yugabyte

# Writing SQL queries

```
SELECT *
FROM barbers
WHERE created_at > '2024-01-01'
```

# Writing SQL queries

- Requires an interface to run queries
- Another app to juggle
- Requires knowledge of SQL
- Learning curve

# AI Data Analyst

- Works in Slack
- Returns an answer
- Still returns the SQL query

# Part 1: Ask a chat bot to generate SQL

- Write me an sql query to get a rental barber from my db

# Part 1: Ask a chat bot to generate SQL

- Write me an sql query to get a rental barber from my db
- To provide a precise SQL query, we need more information about your database schema...

# Part 1: Ask a chat bot to generate SQL

```
SELECT *
FROM barbers
WHERE is_rental = 1;
```

# Part 2: Provide table structure

Write me an sql query to get a rental barber from my db.

My barbers table has the following columns:  
id, name, shop\_id, created\_at

The shops table has shops.kind column which is 0 for rental and 1 for commission

# Part 2: Provide table structure

```
SELECT *
FROM barbers
  JOIN shops ON barbers.shop_id = shops.id
WHERE shops.kind = 0
LIMIT 1;
```

# 225 tables

225 tables

600,000 tokens

## Run settings

Reset

### Model

Gemini 1.5 Pro 002

### Token Count

628,891 / 2,000,000

### Temperature

0

## Tools

JSON mode



Edit schema

Code execution



## System Instructions

Optional tone and style instructions for the model

User

Generate an SQL query that would return me all rental barbers registered within the last 30 days

Model 54.1s

# Google AI Studio

```
SELECT b.id, b."firstName", b."lastName"  
FROM barbers AS b  
JOIN shops AS s ON b.shopId = s.id  
WHERE s.kind = 0 -- Assuming 'kind' 0 indicates a rental shop  
AND b.createdAt >= NOW() - INTERVAL '30 days';
```

Use code [with caution](#).

SQL



20/54



# DB Helper



## System Instructions

Optional tone and style instructions for the model



```
"transactionId" varchar(255)  
)  
with (parallel_workers = 8);
```

```
alter table public.account_entries
```



Compare

Get code

:

## Run settings

Reset

### Model

Gemini 1.5 Pro 002

### Token Count

628,813 / 2,000,000

### Temperature



0



### Tools

#### JSON mode



Edit schema

#### Code execution



## ^ System Instructions

Optional tone and style instructions for the model

User

Generate an SQL query that would return me all rental barbers registered within the last 30 days

Model 54.1s

```
SELECT b.id, b."firstName", b."lastName"  
FROM barbers AS b  
JOIN shops AS s ON b."shopId" = s.id  
WHERE s.kind = 0 -- Assuming 'kind' 0 indicates a rental shop  
AND b."createdAt" >= NOW() - INTERVAL '30 days';
```

Use code [with caution](#).

SQL



22/54

## Run settings

Reset

✖ Model

Gemini 1.5 Pro 002 ▾

⌚ Token Count

628,891 / 2,000,000

🌡 Temperature

0

## ^ Tools

JSON mode



Edit schema

Code execution





## ^ System Instructions

Optional tone and style instructions for the model

## User

Generate an SQL query that would return me all rental barbers registered within the last 30 days

## Model 54.1s

```
SELECT b.id, b."firstName", b."lastName"  
FROM barbers AS b  
JOIN shops AS s ON b.shopId = s.id  
WHERE s.kind = 0 -- Assuming 'kind' 0 indicates a rental shop  
AND b."createdAt" >= NOW() - INTERVAL '30 days';
```

Use code [with caution](#).

SQL



Type something



Run



## Run settings

Reset

## Model

Gemini 1.5 Pro 002

## Token Count

628,891 / 2,000,000

## Temperature

0

## ^ Tools

## JSON mode



Edit schema

## Code execution



## User

Generate an SQL query that would return me all commission barbers registered within the last 30 days who had more than 10 bookings

## Model 53.1s

```
SELECT b.id, b."firstName", b."lastName"
FROM barbers AS b
JOIN shops AS s ON b."shopId" = s.id
WHERE s.kind = 1 -- Assuming 'kind' 1 represents commission-based shops
      AND b."createdAt" >= NOW() - INTERVAL '30 days'
      AND (SELECT COUNT(*)
            FROM appointments AS a
            WHERE a."barberId" = b.id) > 10;
```



Use code [with caution](#).

SQL

## User

Generate an SQL query that would return me all commission barbers registered within the last 30 days who had more than 10 bookings

## Model 53.1s

```
SELECT b.id, b."firstName", b."lastName"
FROM barbers AS b
JOIN shops AS s ON b."shopId" = s.id
WHERE s.kind = 1 -- Assuming 'kind' 1 represents commission-based shops
      AND b."createdAt" >= NOW() - INTERVAL '30 days'
      AND (SELECT COUNT(*)
            FROM appointments AS a
            WHERE a."barberId" = b.id) > 10;
```



Use code [with caution](#).

SQL

## Optional tone and style instructions for the model

User

Generate an SQL query that would return me all rental barbers registered within the last 30 days

Model 54.1s

```
SELECT b.id, b."firstName", b."lastName"  
FROM barbers AS b  
JOIN shops AS s ON b."shopId" = s.id  
WHERE s.kind = 0 -- Assuming 'kind' 0 indicates a rental shop  
AND b."createdAt" >= NOW() - INTERVAL '30 days';
```

# Gemini Free tier

<b>Limit Type</b>	<b>Gemini 1.5 Pro</b>	<b>Gemini 1.5 Flash</b>
Requests per minute (RPM)	2	15
Tokens per minute (TPM)	32,000	1,000,000
Requests per day (RPD)	50	1,500

# Gemini Pay-as-you-go

Pricing Type	Gemini 1.5 Pro	Gemini 1.5 Flash
Input* (per 1M tokens)	\$1.25	\$0.075
Output* (per 1M tokens)	\$5.00	\$0.30
Context Caching (per 1M tokens)	\$0.3125	\$0.01875

# AI Studio Free tier

AI Studio is  
Free

# AI Studio

- Experiments
- Prompt Gallery
- Model tuning

# BigQuery Copilot

# BigQuery

BigQuery is a fully managed, AI-ready data analytics platform that helps you maximize value from your data.



## Explorer

+ ADD



Untitled query



com\_get... OID



Search BigQuery resources



Viewing resources.

SHOW STARRED ONLY



squire-commander



Queries



(Classic) Queries (5)



Notebooks



Data canvases



Data preparations



Workflows



External connections



analytics\_173126600



firebase\_crashlytics



com\_getsquire\_Squire\_Co...



com\_getsquire\_commande...



com\_getsquire\_rosemont\_...



firebase.messaging



firebase.predictions



firebase.predictions\_v2



play\_store\_stats



smsmay



com\_getsquir...

QUERY

SHARE

COPY



&lt;

SCHEMA

DETAILS

PREVIEW

TABLE EXPLORER

PREVIEW

INSIGHTS

LINEAR

Row

platform

1

ANDROID

bundle\_identifier

com.getsquire.commander

event\_id

1993848666587889693

is\_fatal

null

ei

F

Results per page: 50 ▾ 1 – 50 of 135519

&lt;

&lt;

&gt;

&gt;

&gt;|



▶ Data preparations

▶ Workflows

▶ External connections

▶ analytics\_173126600



▶ firebase\_crashlytics



▶ com\_getsquire\_Squire\_Co...



▶ com\_getsquire\_commande...



▶ com\_getsquire\_rosemont...



▶ firebase.messaging



▶ firebase.predictions



▶ firebase.predictions\_v2



▶ play\_store\_stats



▶ smsmay



Res

## Generate SQL with Gemini

Generate me an sql query to see the referral source distribution of my users in my Google Analytics data

[EDIT](#)

```
SELECT
    trafficSource.referralPath,
    COUNT(*) AS referral_count
FROM
    `bigquery-public-data.google_analytics_sample.ga_sessions_20170801`
GROUP BY
    trafficSource.referralPath;
```

This query will process 8.32 KB when run.

< 1/1 >

Refine



Edit Table Sources

Query Summary

[Use query with caution](#)

Rate this suggestion

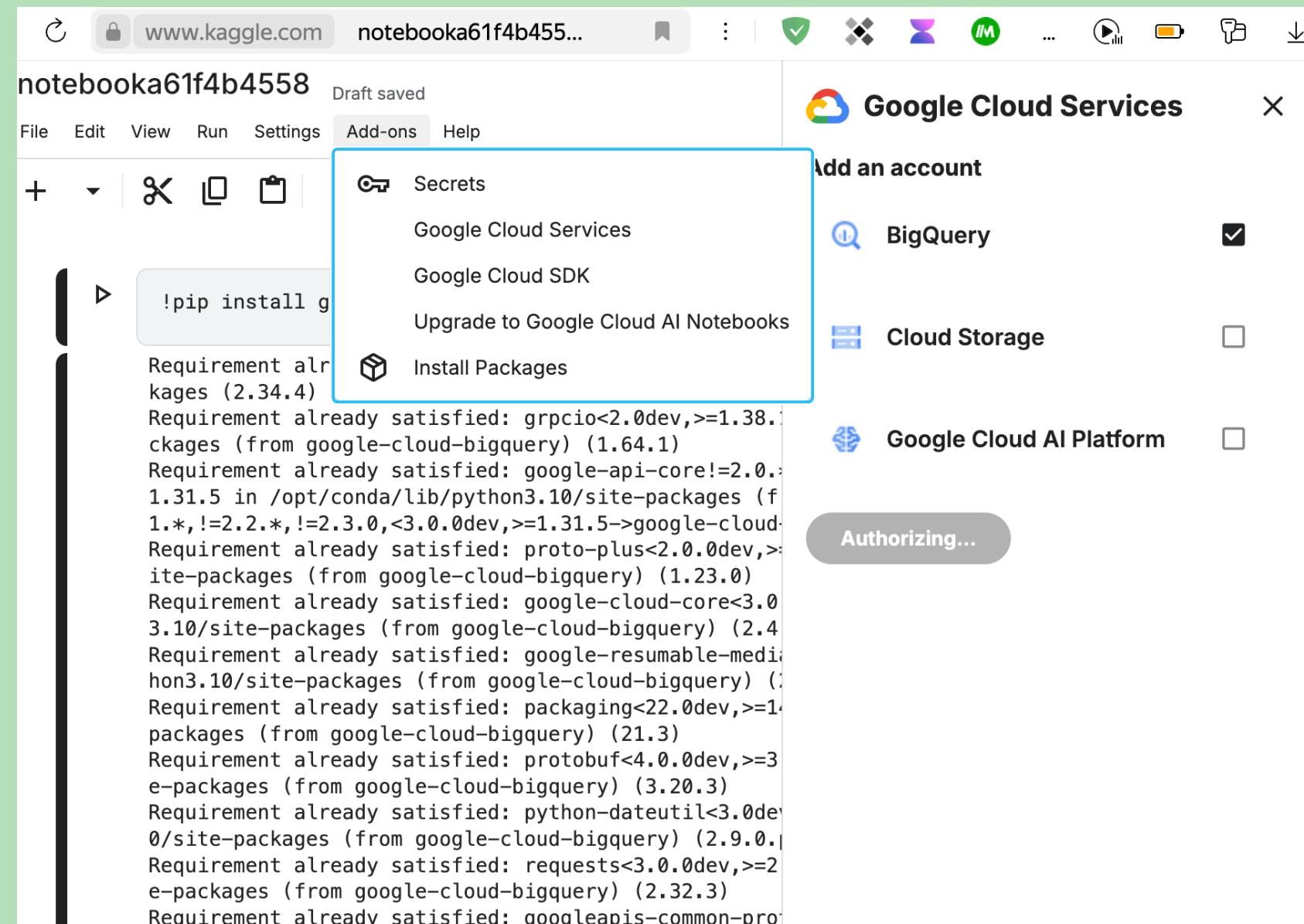


[INSERT](#)

# BigQuery free tier

- 10 GB storage
- 1 TB of queries per month
- Gemini is Free w/o billing account

# Kaggle Notebooks + BigQuery



# AI Analyst

1. Receive a question
2. Generate SQL
3. Execute SQL query
4. Display the results

# How to give the tables structure?

# RAG

## Retrieval Augmented Generation

# RAG - Langchain

```
self.prompt = PromptTemplate.from_template("""  
Given the following database schema:  
{context}"""
```

```
Generate an SQL query to answer this question{question}  
Return only the SQL query without any explanation.  
""")
```

```
sql_query = self.llm_chain.invoke({  
    "context": context,  
    "question": question  
})
```

# Vanna AI

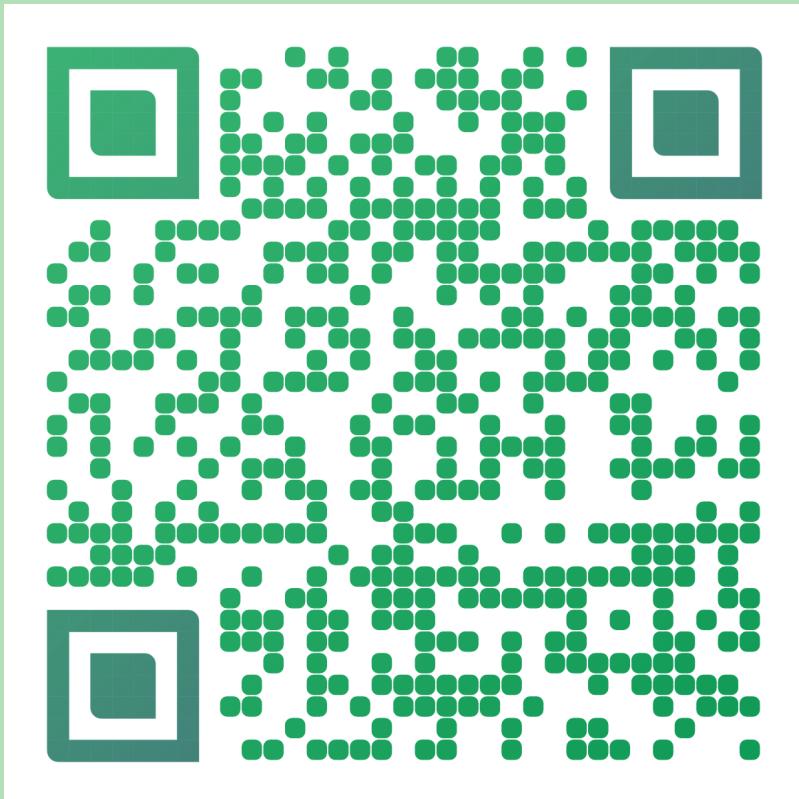
- Can be connected to any\* db
- Can be connected to any LLM api
- Has a few pre-created UIs

# Vanna AI

- vanna.create()
- vanna.connect()
- vanna.train()
- vanna.ask()

# Kaggle notebook

Start using it now!  
[pxl.to/ai-analyst](https://pxl.to/ai-analyst)





**Artur Badretdinov**  Sep 25th at 9:09 AM

How many bookings we had from platforms starting with "ROCK on..."?



4 replies



**Chat with Database** APP Sep 25th at 9:09 AM

SQL Query:

**Run Query**

```
SELECT COUNT(*) AS booking_count
FROM app_db_poly.appointments
WHERE platform LIKE 'ROCK on%'
LIMIT 30;
```

Here are the results (up to 30 rows and 6 columns):

BOOKING_COUNT
9134



**Artur Badretdinov**  Oct 1st at 12:13 PM

Give me a SELECT query returning barbers of brands, where those brands' subscriptions.status = 'past\_due'.

3 replies



**Chat with Database** APP Oct 1st at 12:13 PM

SQL Query:

**Run Query**

```
SELECT b.id,
       b.first_name,
       b.last_name,
       b.shop_id,
       brand.id AS brand_id,
       brand.name AS brand_name
  FROM app_db_poly.barbers b
    JOIN app_db_poly.shops s ON b.shop_id = s.id
    JOIN app_db_poly.brands brand ON s.brand_id = brand.id
    JOIN app_db_poly.subscriptions sub ON brand.id = sub.owner_id
 WHERE sub.status = 'past_due'
  LIMIT 30
```

[See less](#)

47/54

Here are the results (up to 30 rows and 6 columns):



**Artur Badretdinov**  Sep 24th at 4:25 PM

Give me count of migration requests split by it's value. return only value of dataMigrationRequest. Also, don't show Excluded By SQUIRE

2 replies



**Chat with Database** APP Sep 24th at 4:25 PM

SQL Query:

**Run Query**

```
SELECT value:dataMigrationRequest AS migration_request_value,  
       COUNT(*) AS request_count  
  FROM app_db_poly.configuration  
 WHERE TRY_PARSE_JSON(value):dataMigrationRequest IS NOT NULL  
   AND value:dataMigrationRequest != 'Excluded By SQUIRE'
```

[See more](#)

Here are the results (up to 30 rows and 6 columns):

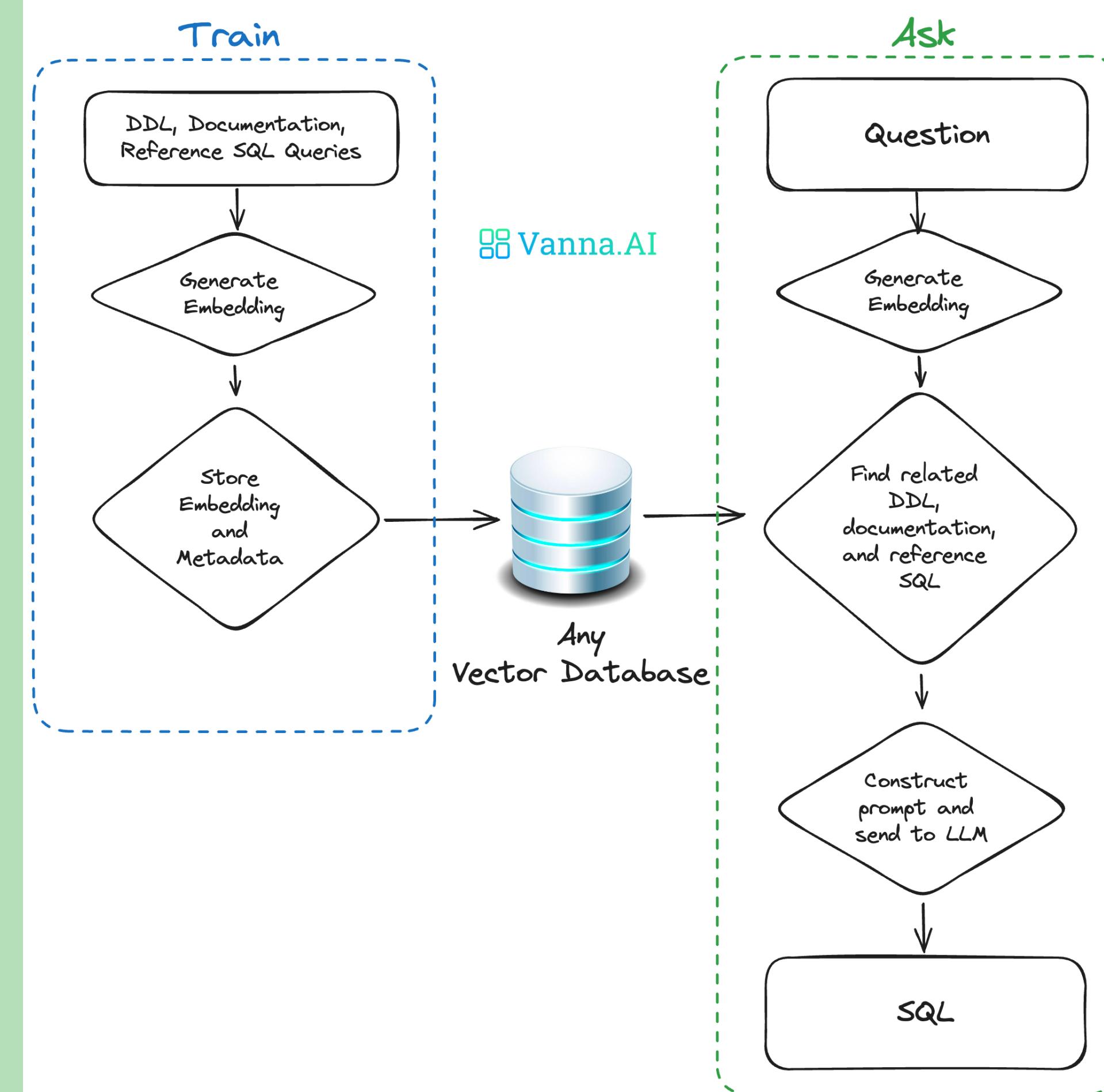
MIGRATION_REQUEST_VALUE	REQUEST_COUNT
"thecut"	14
"none"	267
"schedulicity"	1

# Onboard new team members with AI Analyst

- Get answers about data
- Learn where the data is stored
- Quickly see what changes in the db based on API calls

# SQL Assistant Evolution

- Ask a chat bot to generate SQL
- Also, provide it table structure
- Give it all table structures
- Teach it to actually run queries
- Get it closer to users - Slack

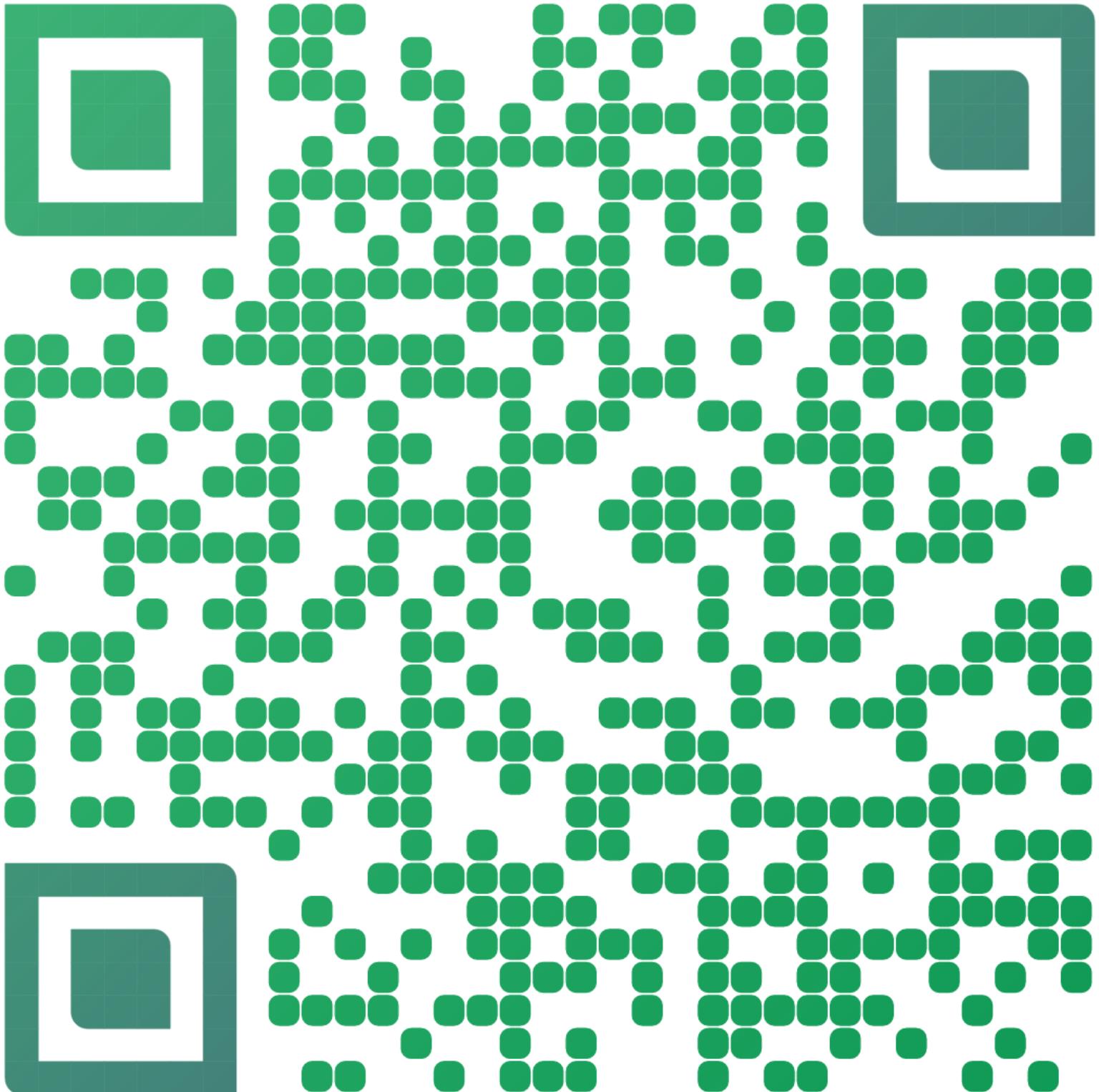


# Considerations

- Everyone needs broader context
- To give or not to give
- Security

# One-line notes

- Interested in the bot? Message me!
- LangChain / other tools
- Function calling by LLMs
- Chat2Db



# Artur Kuzmin

Director of Engineering at Squire,  
Google Developer Expert

[linkedin.com/in/arturkuzmin/](https://linkedin.com/in/arturkuzmin/)  
[twitter.com/ArtursTwit](https://twitter.com/ArtursTwit)