

AskData

Product Requirements Document · GenAI Analytics Assistant · MVP

Field	Detail
Product	AskData: natural-language questions answered from governed metrics
Company	StreamCo (fictional subscription business, synthetic data)
Author / PM	Jaykumar Raichura
Builds on	FinSight governed fact table and metric glossary
Status	MVP prototype complete; evaluation v2
Stakeholders	Head of Data, FP&A, Product, Regional Ops, Security & Privacy, Data Engineering

1. Summary

Business users at StreamCo wait two to three days for analysts to answer simple data questions, and analysts spend much of their week on repeat requests. Generic AI chatbots are not an option: they produce confident numbers that do not match finance. AskData lets users ask questions in plain English and returns an answer computed from FinSight's governed metric layer, with the SQL, assumptions and a confidence label shown on every answer.

2. Problem

- Speed: ad-hoc data questions take 2–3 days through the analyst queue.
- Analyst capacity: repeat lookups crowd out deeper analysis.
- Trust: free-form AI answers can invent numbers or use different definitions from finance.

Problem statement: How might we let business users get trusted data answers in seconds without creating a new source of wrong numbers?

3. Goals and non-goals

Goals

- Answer supported questions in seconds, with numbers that match FinSight exactly.
- $\geq 90\%$ accuracy on a golden evaluation set before launch; $\geq 95\%$ at GA.
- 100% safe handling of guardrail cases (PII, forecasts, unsupported metrics, prompt injection).
- Reduce repeat analyst requests by 40% within 90 days.

Non-goals (MVP)

- Forecasting or predictions (handled by the FinSight financial model).
- Customer-level or personal data.
- Free-form SQL on raw tables.
- Write actions of any kind.

4. Users

Persona	Job to be done	Pain today
Regional Ops Lead	Get a number for a meeting in minutes	Waits days; doesn't know SQL
Product Manager	Explore churn, tiers and adoption ad hoc	Analyst queue slows decisions
Data/BI Analyst	Hand off repeat lookups and verify AI output	Spends hours on the same questions
FP&A / Governance	Ensure every number matches finance	Distrusts AI-generated figures

5. Discovery insights

- Users don't trust an answer they can't check, so showing the SQL and assumptions is a requirement, not a nice-to-have.
- Most questions are combinations of a few patterns: metric × filter × period × breakdown.
- Users want the assistant to say "I don't know" rather than guess.

6. Solution approach

The LLM never calculates numbers. It maps the question to a structured request (metric, filters, breakdown, period, ranking). SQL is then compiled from the governed semantic layer and executed against the FinSight fact table. This removes the two biggest risks of text-to-SQL: invented numbers and inconsistent definitions.

Approach	Accuracy	Trust risk	Decision
Generic chatbot, no data access	Low	Invents numbers	Rejected
LLM writes SQL on raw tables	Medium	Wrong joins and definitions	Rejected
LLM → semantic layer → compiled SQL	High	Limited to defined metrics	Selected (MVP)
Fine-tuned small model	High once trained	Needs labeled data	Revisit at V2

7. User stories

ID	User story	Priority
US-1	As an ops lead, I want to ask a metric question in plain English and get a number in seconds.	Must
US-2	As any user, I want to see the SQL, assumptions and metric definition behind each answer so I can trust it.	Must
US-3	As any user, I want breakdowns by tier, region and month with a chart.	Must
US-4	As governance, I want the assistant to refuse PII, forecasts and unsupported metrics.	Must
US-5	As a user, I want the assistant to ask me to clarify vague questions.	Should

ID	User story	Priority
US-6	As a PM, I want a descriptive "why" breakdown of variance and churn.	Should
US-7	As an analyst, I want thumbs up/down feedback to feed the eval set.	Should
US-8	As a user, I want proactive anomaly alerts.	Won't (V2)

8. Functional requirements

ID	Requirement	Acceptance criteria
FR-1	Parse question into metric, filters, breakdown, period and rank (JSON schema)	Schema-valid output on 100% of requests
FR-2	Compile SQL only from semantic-layer definitions	No free-text SQL executed
FR-3	Show answer, chart, SQL, assumptions, definition and confidence	All six present on every numeric answer
FR-4	Guardrails: PII, out-of-range dates, forecasts, unknown metrics, injection	100% of guardrail golden cases handled
FR-5	Clarifying question when no metric is identifiable	Offers 3–4 specific options
FR-6	Defaults stated explicitly (e.g., full year, latest month)	Assumption shown whenever a default is used
FR-7	Feedback capture on each answer	Logged with question and parsed request

9. AI requirements

- Output: strict JSON matching the request schema; temperature 0.
- Context: metric names, synonyms, dimensions and valid values from the semantic layer only; no raw data in the prompt.
- Unsupported: model must return type "unsupported" rather than approximate.
- Latency target: p95 under 3 seconds end to end.
- Cost target: under \$0.05 per question.

10. Evaluation plan

- Golden set: 45 questions across 8 categories (lookups, filters, breakdowns, trends, variance/why, ambiguous, guardrails, paraphrases).
- Held-out set: 15 new questions written after tuning, used only for reporting.
- Correct = metric, filters, breakdown, period and rank all match; numbers verified against SQL.
- Launch gate: $\geq 90\%$ held-out accuracy and 100% guardrail pass rate.
- Feedback loop: every thumbs-down is reviewed and added to the golden set.

11. Guardrails and responsible AI

Risk	Control
Invented numbers	LLM never computes; all numbers come from executed SQL
Personal data exposure	Aggregated metrics only; PII requests refused
Overtrust in "why" answers	Labeled descriptive, not causal
Prompt injection	Read-only access; SQL compiled from layer, not user text
Silent assumptions	Every default shown to the user

12. Success metrics

Metric	Baseline	Target
Time to answer	2–3 days	< 10 seconds
Held-out accuracy	n/a	≥ 90% (prototype: 93%)
Guardrail pass rate	n/a	100%
Repeat analyst requests	100%	–40% in 90 days
Weekly active users	0	60% of licensed users
Thumbs-up rate	n/a	≥ 80%

13. Release plan

Phase	Scope	Exit criteria
MVP	Q&A on 11 governed metrics, breakdowns, trends, guardrails, SQL transparency	Held-out ≥ 90%, guardrails 100%
V1	"Why" driver breakdowns, follow-up suggestions, feedback loop	Thumbs-up ≥ 80%
V2	Anomaly alerts, more metrics (CAC, NPS) once governed, model cost optimization	Precision target met

14. Open questions

- Which metrics most requested but unsupported (e.g., CAC, NPS) should be added to the semantic layer first?
- Should low-confidence answers route to an analyst automatically?