



Enterprise Data Platform

Automating Data Question-Answering Task

2023.10.30

01. Introduction

02. Company / Product Overview

03. Question-Answering Automation

Smart Pattern Discovery

B-to-the-A

Text-to-SQL

Dialogue

Signal

I. 2015~

AUGMENTED
ANALYTICS

- Smart Pattern Discovery
- Raw Data Visualization



II. 2020~

B-T-O-T-H-E-A

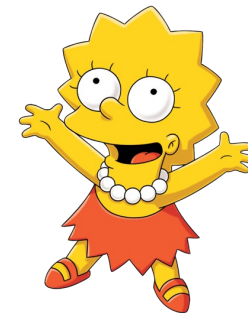
- BI to the Analytics
 - Integrated Workflow
 - Ad-hoc Question on Metrics
 - Data Access (SQL Editor)

III. 2023~

DATA
EXPERIENCE

- Data Q-A Automation
 - Text-to-SQL
 - Dialogue
 - Signal

Data Experience
in answering
data questions !!!!

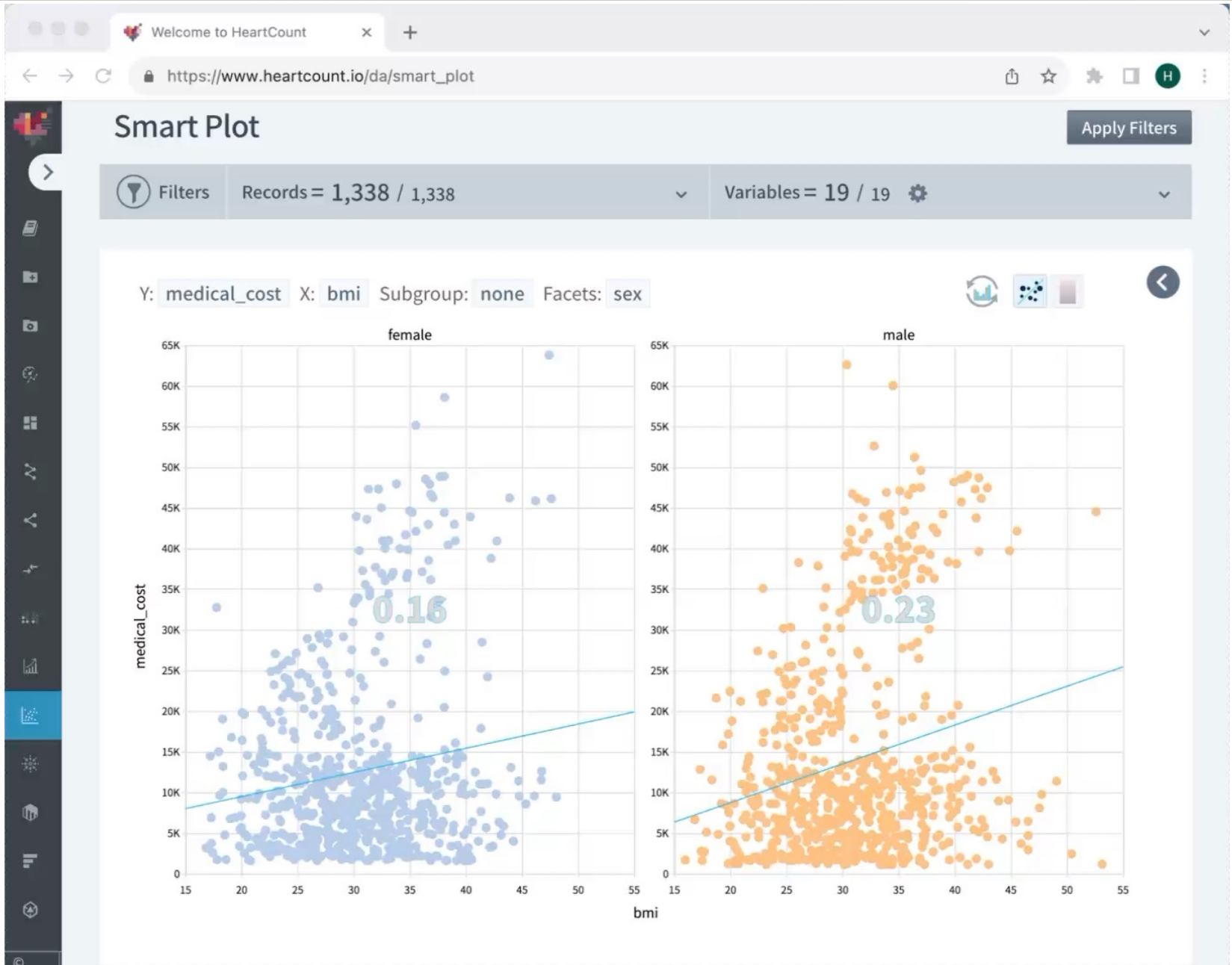


HEARTCOUNT

an enterprise data consumption platform which unifies end-to-end analytical workflow from question to answer, ensuring an effortless data consumption experience.

Enterprise Customers in KR/JP





B-to-the-A

End-to-End Data Consumption Tool: BI and Analytics Integrated



See Forest

check metrics and access analytics-ready datasets for further questions



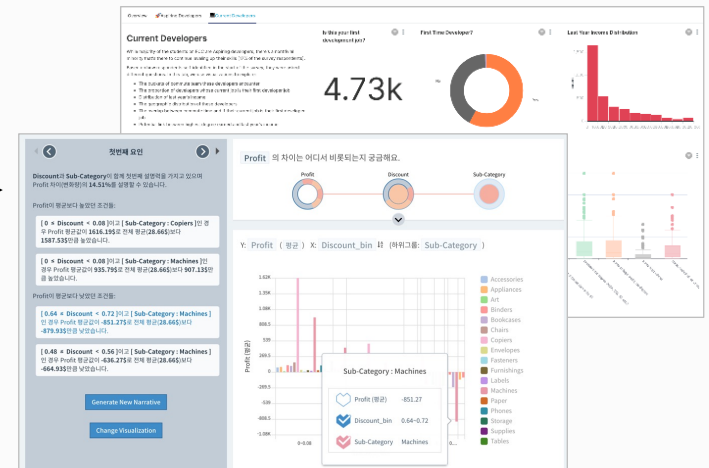
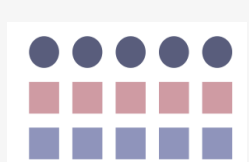
See Trees

EDA and augmented analytics with the datasets to answer questions

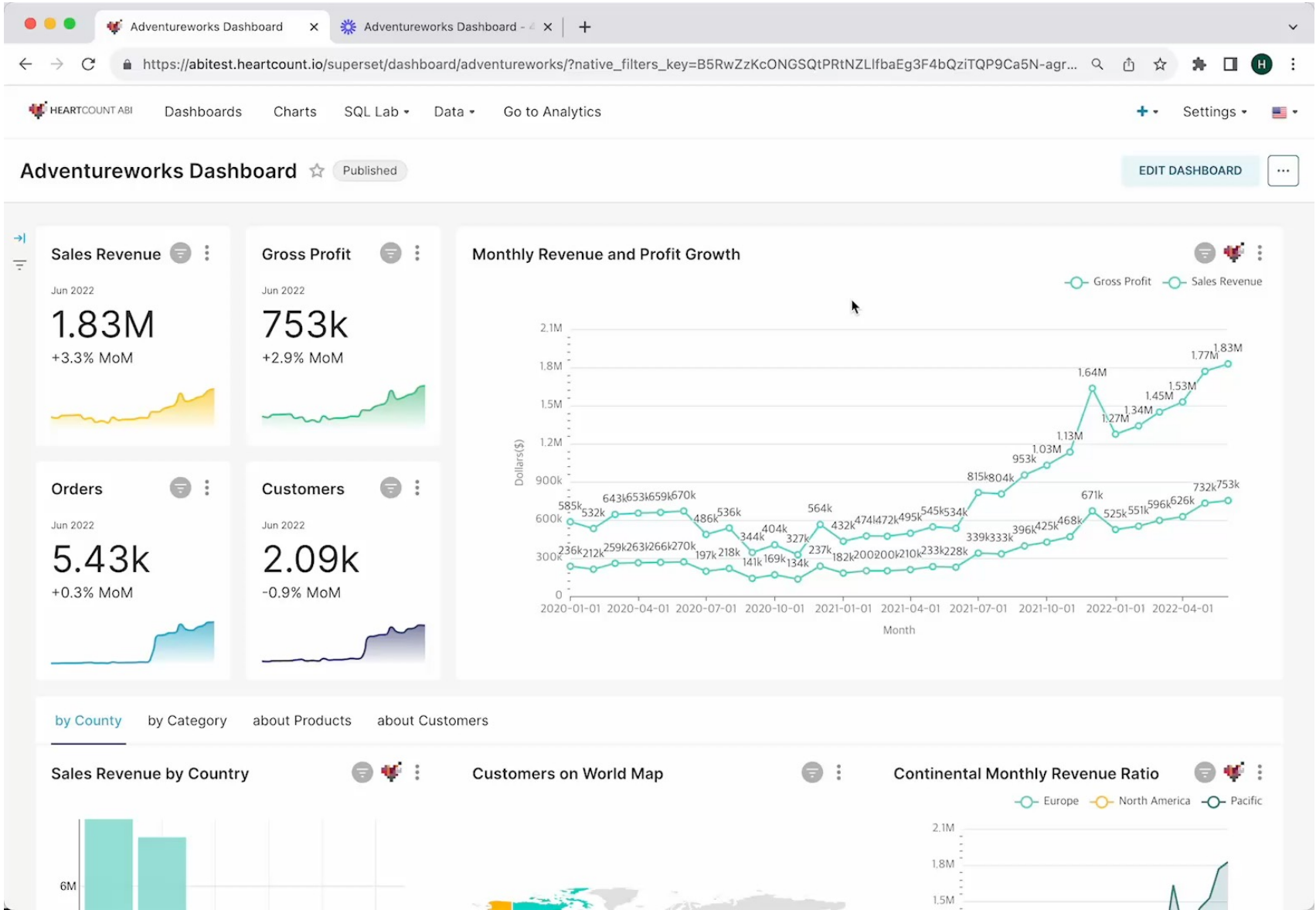
Data Source



Analytics-Ready Dataset



B-to-the-A: on-the-fly EDA in response to ad-hoc questions





- I've studied SQL, but I still find data extraction challenging.
- Even with a well-organized dataset, I'm not sure what questions to ask.
- If only someone could preemptively inform me about the major changes in Metrics

Automating Data Question-Answering

ad-hoc
Question
Answering

How many? ▶ **Text-to-SQL**

Why/How? ▶ **Dialogue**


What's New? ▶ **Signal**

Life-Changing
Data Experience

Time-to-Insight

▶ **Automate Everything But Thought**

How Many? ► Text-to-SQL

Question	Counting	Why, How		
	Question-Answer	Questions	Q-Chart	C-Narrative
Work	Answering Counting Question	Asking Questions	Chart Visualization for Question	Key Narrative Generation from the Chart
Effort	Data(set) Extraction	Analytical Reasoning	Tool Learning	Quantitative Interpretation
	*LLM Text-to-SQL	Proprietary Algorithm		

10월 5일 10:20



Please extract
Last Month's Sales Total and
of Customer by Category?

```
select
  product_category,
  state,

  sum(sales) as total_sales_amount,
  count(distinct customer_id) as number_of_customers

from sales
where date_trunc('year', sold_at) = 2021
group by 1, 2
```

product_category	state	year	total_sales_amount	number_of_customers
brewing supplies	NY	2021	612	27
brewing supplies	VT	2021	172	9
coffee beans	NY	2021	575	27
coffee beans	VT	2021	256	12
merchandise	NY	2021	415	16
merchandise	VT	2021	325	14

10월 5일 11:20



Oops... By Category and
Region, Please?

10월 7일 12:20



Can I Ask the same on a daily
basis?

Why don't you Learn SQL?

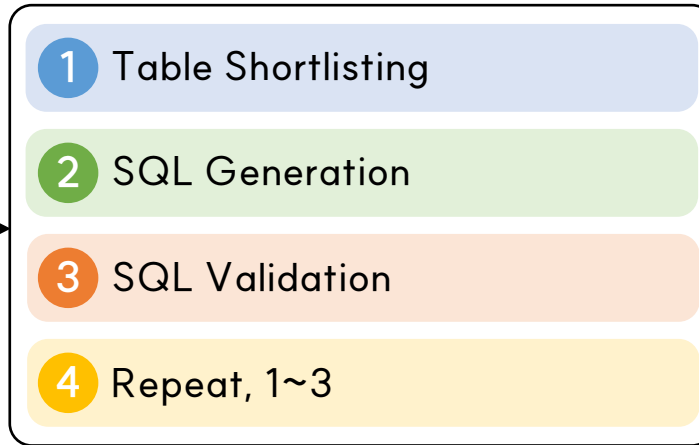


Question: Extract total Sales Amount and Unique Customer Count by Category on a daily basis for the month of May 2021

The screenshot shows the Superset SQL Lab interface. On the left, the 'DATABASE' is set to 'postgresql PostgreSQL' and the 'SCHEMA' is 'adventureworks'. Under 'SEE TABLE SCHEMA', three tables are listed: 'adventureworks_customers', 'adventureworks_product_categories', and 'adventureworks_sales'. The 'adventureworks_sales' table is expanded, showing columns: 'orderdate' (DATE), 'stockdate' (DATE), 'ordernumber' (VARCHAR), 'productkey' (INTEGER), 'customerkey' (INTEGER), 'territorykey' (INTEGER), 'orderlineitem' (INTEGER), and 'orderquantity' (INTEGER). The main editor area contains a text input field with the placeholder 'Enter a question and we'll answer it with a SQL.' and a 'GENERATE QUERY' button. Below the input field, a SQL editor shows a single line: '1 SELECT ...'. At the bottom of the editor, there is a 'RUN' button, a 'LIMIT: 1 000' dropdown menu, a 'SAVE AS' button, and a 'COPY LINK' button. The bottom of the interface has tabs for 'RESULTS' and 'QUERY HISTORY'.

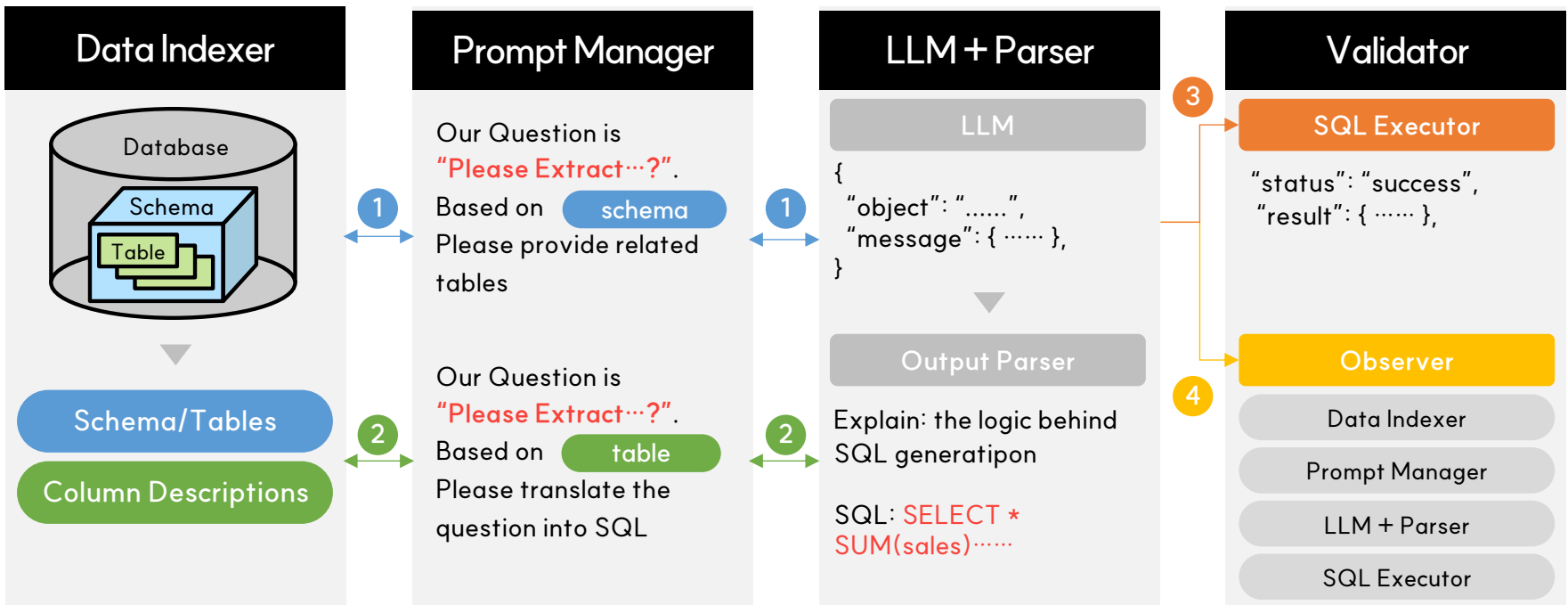


Total Sales and Unique Customer # by Category on a daily basis?




```
SELECT sum(sales) as Total_sales
      , sum(profit) as Total_profit
FROM sales_table
WHERE year =
YEAR(CURDATE())
GROUP BY month;
```

Total_sales	Total_profit
100,000	50,000



Why Difference? How to Increase? ▶ Dialogue

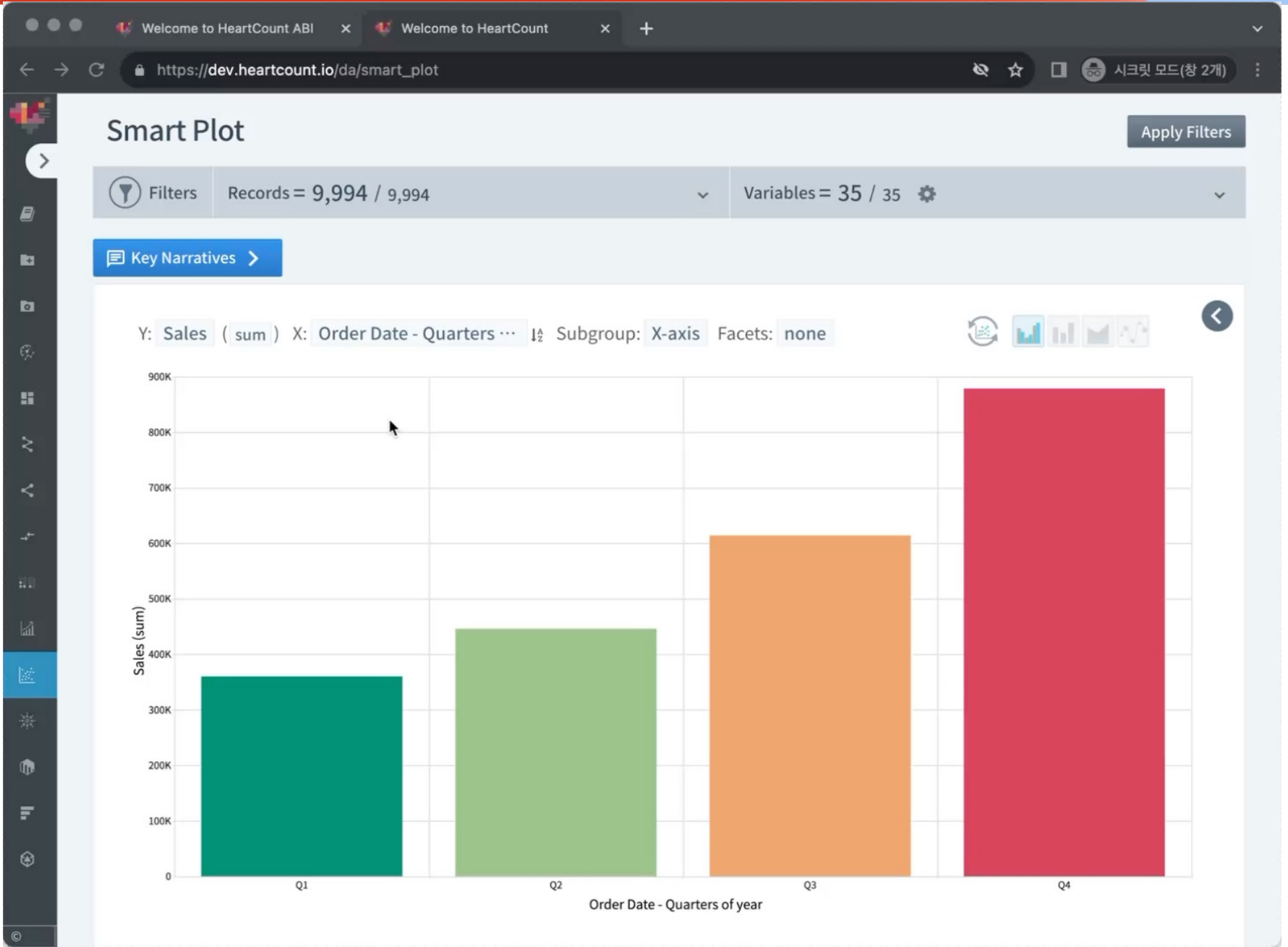
Question	Counting	Why, How		
	Question-Answer	Questions	Q-Chart	C-Narrative
Work	Answering Counting Question	Asking Questions	Chart Visualization for Question	Key Narrative Generation from the Chart
Effort	Data(set) Extraction	Analytical Reasoning	Tool Learning	Quantitative Interpretation
	*LLM Text-to-SQL	Proprietary Algorithm		

SOCRATES



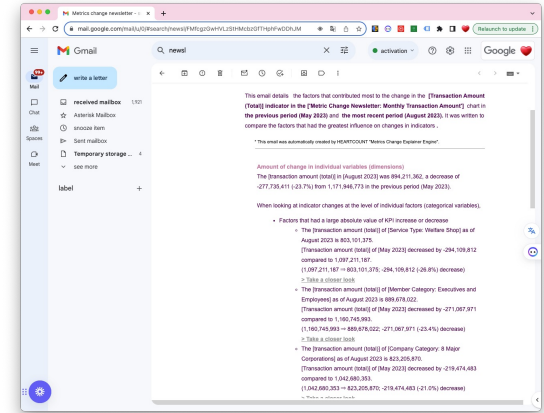
Socratic Dialogue

- For those not familiar with the analytical reasoning process,
- HC Dialogue offers accurate facts and new perspectives to help construct informed opinions,
- posing the insightful questions you might not have thought to ask.



What's New on Metrics? ▶ Signal

Data Newsletter Explaining Changes on Key Metrics



Outlier Detection on All the Metrics



You received a **Signal**
Jan 30 at 6:58pm

[Category=Clothing] Sales
Total increase 34% on 30 Jan 2020

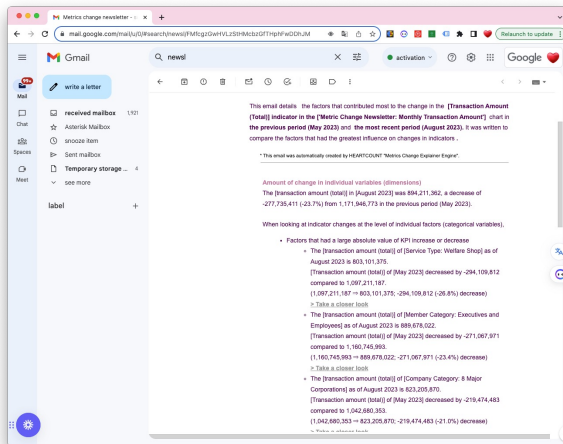
- Sales Average: 9% ⇄
- Sales Record Count: 31%. ↗



Between 01-01-2020 and 01-30-2020

[Explore](#)

Data Newsletter : Newsletter that regularly seeks out the causes of changes in specific Metrics.



• Factors that had a large absolute value of KPI increase/decrease rate

- The [transaction amount (total)] of [Category Name: Culture and Leisure] in [August 2023] is 5,160,200.
- [Transaction amount (total)] of [May 2023] increased by 4,941,900 compared to 218,300.
- (218,300 ⇒ 5,160,200; 4,941,900 (2,263.8%) increase)

> [Take a closer look](#)

- [Payment amount range: 5 million won range]
- [Transaction amount (total)] in [August 2023] is 0
- [Transaction amount (total)] of [May 2023] decreased by -5,158,000 compared to 5,158,000.
- (5,158,000 ⇒ 0; -5,158,000 (-100.0%) decrease)

> [Take a closer look](#)

- The [transaction amount (total)] of [orderer age group: 80s] in [August 2023] is 0.
- [Transaction amount (total)] of [May 2023] decreased by -109,700 compared to 109,700.
- (109,700 ⇒ 0; -109,700 (-100.0%) decrease)

> [Take a closer look](#)

> [Explore more factors](#)

Before Your Boss Asks, Understand Major Changes on Metrics

[Demonstration] Dashboard for analysis of automobile non-life insurance increase/decrease factors ★ Draft

[EDIT DASHBOARD](#)

Monthly increase/decrease amount, increase/decrease rate, number of increase/decrease cases by variable (1) compared to the previous month

Base year month_str	division	summary	Increase/decrease compared to previous month	Rank of increase or decrease	Month-on-month change rate	Ranking of increase/decrease rate	Number of cases increase/decrease compared to the previous month	Ranking of number of increases and decreases
2022-12-01	Total insurance amount	In case of [Degree of Damage: Total Loss], the fluctuation range is +7.48 billion won (1105.2 → 118.00 billion won , +6.8%) .	7,479,681,149	One	6.8%	16	4104	One
2022-12-01	Total insurance amount	When [Collateral: Property], the fluctuation range is +5.96 billion won (87.14 → 93.10 billion won , +6.8%) .	5,955,693,623	2	6.8%	15	3533	2
2022-12-01	Total insurance amount	When [Vehicle_Major Category: Domestic], the fluctuation range is +4.46 billion won (677.0 → 72.17 billion won , +6.6%) .	4,463,393,942	3	6.6%	20	3203	3
2022-12-01	Total insurance amount	When [Vehicle_Major Category: Foreign Made], the fluctuation range is +3.17 billion won (44.43 → 47.60 billion won , +7.1%) .	3,168,413,760	4	7.1%	13	983	6
2022-12-01	Total	In the case of [Vehicle_Medium Classification: Passenger Vehicle Medium], the change is +2.81 billion won (11.12 → 13.93 billion won , +6.3%) .	2,812,145,204	5	7.6%	10	1339	4

Variables (2) Increase/decrease amount, increase/decrease rate, number of increase/decrease cases compared to the previous month by combination

Base year month_str	division	summary	Increase/decrease compared to previous month	Rank of increase or decrease	Month-on-month change rate	Ranking of increase/decrease rate	Number of cases increase/decrease compared to the previous month	Ranking of number of increases and decreases
2022-12-01	Total insurance amount	In case of [Degree of Damage: Total Loss], the fluctuation range is +7.48 billion won (1105.2 → 118.00 billion won , +6.8%) .	7,479,681,149	One	6.8%	16	4104	One
2022-12-01	Total insurance amount	When [Collateral: Property , Vehicle_Middle Category: Passenger Vehicle], the fluctuation range is +1.42 billion won (22.36 → 23.78 billion won , +6.4%) .	1,422,859,497	14	6.4%	142	655	13
2022-12-01	Total insurance amount	When [Region: Foreign , Vehicle Category: Foreign], the change is +1.39 billion won (22.05 → 23.44 billion won , +6.3%) .	1,393,640,949	15	6.3%	144	384	26
2022-12-01	Total insurance amount	When [Region: Foreign , Damage Level: Total Loss], the fluctuation range is +1.36 billion won (21.79 → 23.15 billion won , +6.2%) .	1,355,678,078	16	6.2%	148	370	28
2022-12-01	Total insurance	In the case of [Vehicle_Major Category: Domestic , Vehicle_Middle Category: Passenger Midsize], the fluctuation range is +1.25 billion won (17.06 → 18.31 billion won , +7.3%) .	1,249,279,865	17	7.3%	119	868	10

Thank You