

Supplemental Material: Human-Authored and Automatic Dashboard Summaries

JANE HOFFSWELL, Adobe Research, USA

VICTOR S. BURSZTYN, Adobe Research, USA

SHUNAN GUO, Adobe Research, USA

EUNYEE KOH, Adobe Research, USA

Introduction

This document is supplemental material for the paper “*A Case Study of Human-Authored versus Automatic Dashboard Summaries*” (<https://doi.org/10.1145/3706599.3720155>). In this work, we conducted a case study evaluation with five industry professionals to understand the insight selection and summary authoring process for analytic dashboards. We compare these results to a set of automatic, LLM-based dashboard summaries. As part of this case study, we collected fifteen human-authored summaries and nine LLM-generated summaries based on nine sample dashboards of varying sizes (small, medium, and large). This document includes both the dashboard images and the corresponding summaries.

CONTENTS

1	Dashboard: Call Center Overview	2
1.1	Small: Call Center Overview	2
1.2	Medium: Call Center Overview	3
1.3	Example Prompt: Medium Call Center Overview	4
1.4	Large: Call Center Overview	4
2	Dashboard: Human Resources Executive Overview	6
2.1	Small: Human Resources	6
2.2	Medium: Human Resources	7
2.3	Large: Human Resources	8
3	Dashboard: 2023 US Regional Sales Summary	10
3.1	Small: Sales Summary	10
3.2	Medium: Sales Summary	11
3.3	Large: Sales Summary	12

Authors' Contact Information: Jane Hoffswell, jhoffswell@adobe.com, Adobe Research, Seattle, Washington, USA; Victor S. Bursztyn, soaresbu@adobe.com, Adobe Research, San Jose, California, USA; Shunan Guo, sguo@adobe.com, Adobe Research, San Jose, California, USA; Eunyee Koh, eunyee@adobe.com, Adobe Research, San Jose, California, USA.

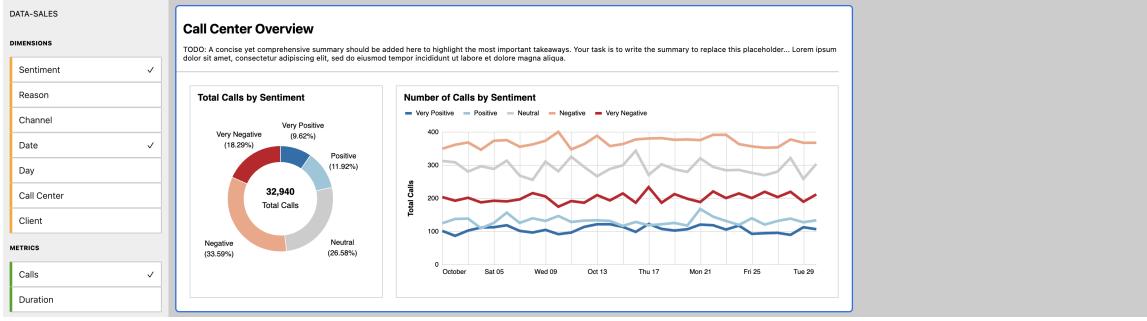


Fig. 1. The small “Call Center Overview” dashboard includes two visualizations that focus on the number of Calls broken down by Sentiment and Date. We generate 28 possible insights for this dashboard.

1 Dashboard: Call Center Overview

For the natural language summaries (both the human-authored and LLM-based), our participants/approach selected an average of 10.6 insights for the “Call Center Overview” dashboard (9 for small, 10.3 for medium, and 13.5* for large). The average summary length was 124.9 words (83.3 words for the small dashboard, 108.3 for medium, and 212* for large); *only two samples (one human, one automatic) are available for the large dashboard.

1.1 Small: Call Center Overview

(1) Human-authored Summary (chs): “*Most calls were Negative throughout the entire period. Its share is about 34% of the total. It is 26% more than the second highest item. The number of different sentiments remained the same throughout the entire period, with some fluctuations. There is no visible correlation between the series. The series have some anomalies. Positive has the most significant anomaly between October 20-21. During this period, it significantly grew and then dropped. A significant anomaly can also be observed for Neutral from October 13 to 17, with a 29% increase followed by a 21% decline.*”

(2) Human-authored Summary (csh): “*Negative calls seemed to dominate the graph at around 50% but that can also be a little misleading, consumers who are unhappy are more likely to report that vs those who were pleased with the service. Only having roughly 20% of the calls being reported in some fashion as positive. Neutral calls accounted for about 27% of the population. A lot of the call sentiments peaked at different times and there is nothing pointing to any correlation between calls or why they may fluctuate.*”

(3) Automatic Summary (cxx): “*The max item ‘Negative’ has 26% more calls than ‘Neutral’. The number of calls in the ‘Neutral’ category decreased by 21% between Oct. 16th and 17th, dropping from 343 to 270. The highest number of calls in the ‘Negative’ category was 400 on Oct. 10th, surpassing its average by 8%. The highest number of calls in the ‘Neutral’ category was 343 on Oct. 16th, which was 18% higher than the average of 291.*”

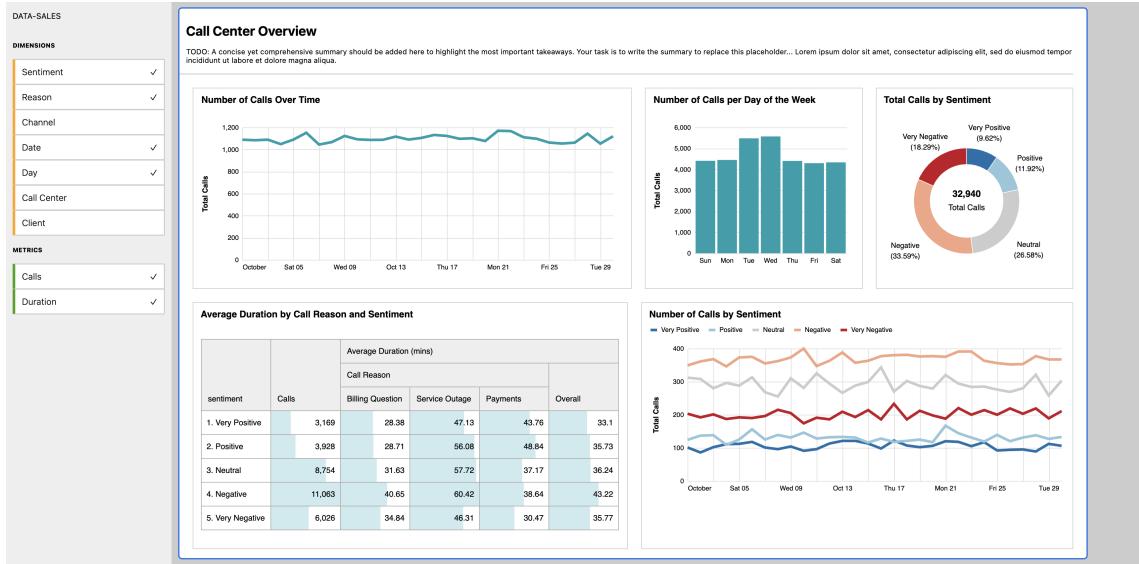


Fig. 2. The medium “Call Center Overview” dashboard includes five visualizations showing the number of Calls and average call Duration broken down by combinations of Sentiment, Reason, Date, and Day. We generate 49 possible insights for this dashboard.

1.2 Medium: Call Center Overview

(1) Human-authored Summary (sch): “The number of calls peaked at 1,170 on Oct 21st. This was 7% more than the daily average of 1,098. There was a notable increase from Oct 4 to Oct 6 (up 10 %) and a notable drop from Oct 21 to Oct 26 (down 10%). There is a considerable increase in the number of calls during Tuesdays and Wednesdays. Combined, these two days account for 34% of total calls and are 17% higher than the average daily calls. Negative is the most common call sentiment (34% of total calls). Very positive is the least common with 3,169 (only 9.62% of total calls). From analyzing call duration, negative sentiment is more common for Billing Questions, and positive sentiment is more common for Payments. There is no strong correlation between the call sentiments across the calendar.”

(2) Human-authored Summary (hcs): “In the month of October, only 22 percent of customers have shown positive sentiment in customer care calls. A high percentage of negative sentiment calls indicates unhappy and sad customers. Calls for billing-related questions demonstrate the most negative sentiment, whereas calls related to payments indicate the highest positive sentiment. Client calls that happen on Tuesday and Wednesday contribute 34% of overall calls compared to the remaining weekdays. The average duration of service-outage related calls is higher compared to other call reasons. Number of calls increased notably by 10 percent at the start of the month (4th to 6th October) and dropped by 10 percent towards the end of the October month.”

(3) Automatic Summary (xcx): “On October 21st, the highest number of calls reached 1,170, which was 7% higher than the average of 1,098. The days with the highest call volume were Wednesday and Tuesday, accounting for 34% of the total calls. The values of calls were highly skewed towards these two days. Wednesday had the highest number of calls, which was 2% higher than Tuesday. The highest value in the negative category for calls was 11,063, accounting for 34% of the total calls.”

1.3 Example Prompt: Medium Call Center Overview

To generate the medium summary, we used OpenAI's GPT-3.5 (*gpt-35-turbo-v0613*), with a decoding temperature of 0.5.

Write a summary of the data report below using one third of the sentences.

Report:

About: "Calls" topped at 1,170 on Oct. 21st. It was 7% more than the average of 1,098.

About: "Wednesday" and "Tuesday" contain the greatest values, with around 5,525.5 in "Calls" (34% in total). The values of "Calls" are highly skewed towards "Tuesday" and "Wednesday" (34% in total). The max item, "Wednesday", is 2% more than the second highest one, "Tuesday", in "Calls".

About: "Negative" had the greatest value, with 11,063 in "Calls" (34% in total).

About: "Negative" had the greatest value, with 11,063 in "Calls" (34% in total).

Summary:

1.4 Large: Call Center Overview

As discussed in the full paper, one participant was unable to complete the study (shc), and as such, we only have one sample of a human-authored summary for the large "Call Center Overview" dashboard.

(1) Human-authored Summary (shc): N/A

(2) Human-authored Summary (hsc): "The billing policy needs immediate attention since 71% of the calls are related to questions regarding billing, followed by Payments and Service Outage, both contributing 14% respectively to total number of calls. Majority of customers contacting the support are frustrated, as almost 52% of the total calls have a negative sentiment (33.59% negative, 18.29% very negative). That is more than half of our customer base, which is an alarming figure. Only 22% of the calls had a positive sentiment. Los Angeles/CA and Baltimore/MD report the highest number of calls reported, which could be a huge indication of regional-level service issues in those areas. Call-Centers handle the most customer support traffic, followed by Chatbot, Email, and Web. Although, the traffic distribution among these 4 channels is fairly balanced, showing that there's no overloading on a single channel. In a week, most calls are recorded on Wednesday and Thursday, up 16.50% and 18.38% respectively compared to the daily average. Support calls through Email go the longest duration, almost 49 minutes, so optimization on this channel is a possibility. Calls with a negative sentiment due to issues related with billing had the highest average duration, about 40 minutes. Company-wise, O.E. Ltd, 'Exact-Rx, Corp', 'Select' and 'Llorens Ltd' make up about 25% of the total calls volume, so a dedicated team to handle these largest clients could be a feasible option. Finally, calls from PEDIFIX, Corp, Eminence Corp, and Pacific Ltd go for the longest duration, taking about 50 minutes to resolve on average."

(3) Automatic Summary (xxc): "The report shows that the highest number of calls were related to billing questions, accounting for 71% of the total calls. Additionally, negative sentiment had the highest number of calls at 34% of the total. The peak number of calls occurred on October 21st with 1,170 calls, which was 7% higher than the average. The call center in Denver had the lowest number of calls at 2,776, accounting for 8% of the total. Negative sentiment had the highest number of calls on October 10th with 400 calls, surpassing its average by 8%. The call channel with the highest number of calls was the call center, accounting for 32% of the total. Wednesday and Tuesday had the highest number of calls, each accounting for 34% of the total. The client "O.E. Ltd" had the highest number of calls, 5% more than the second highest client. Email had the longest average call duration at 49.087 minutes, accounting for 32% of the total. Lastly, negative sentiment had the highest average call duration at 34% of the total."

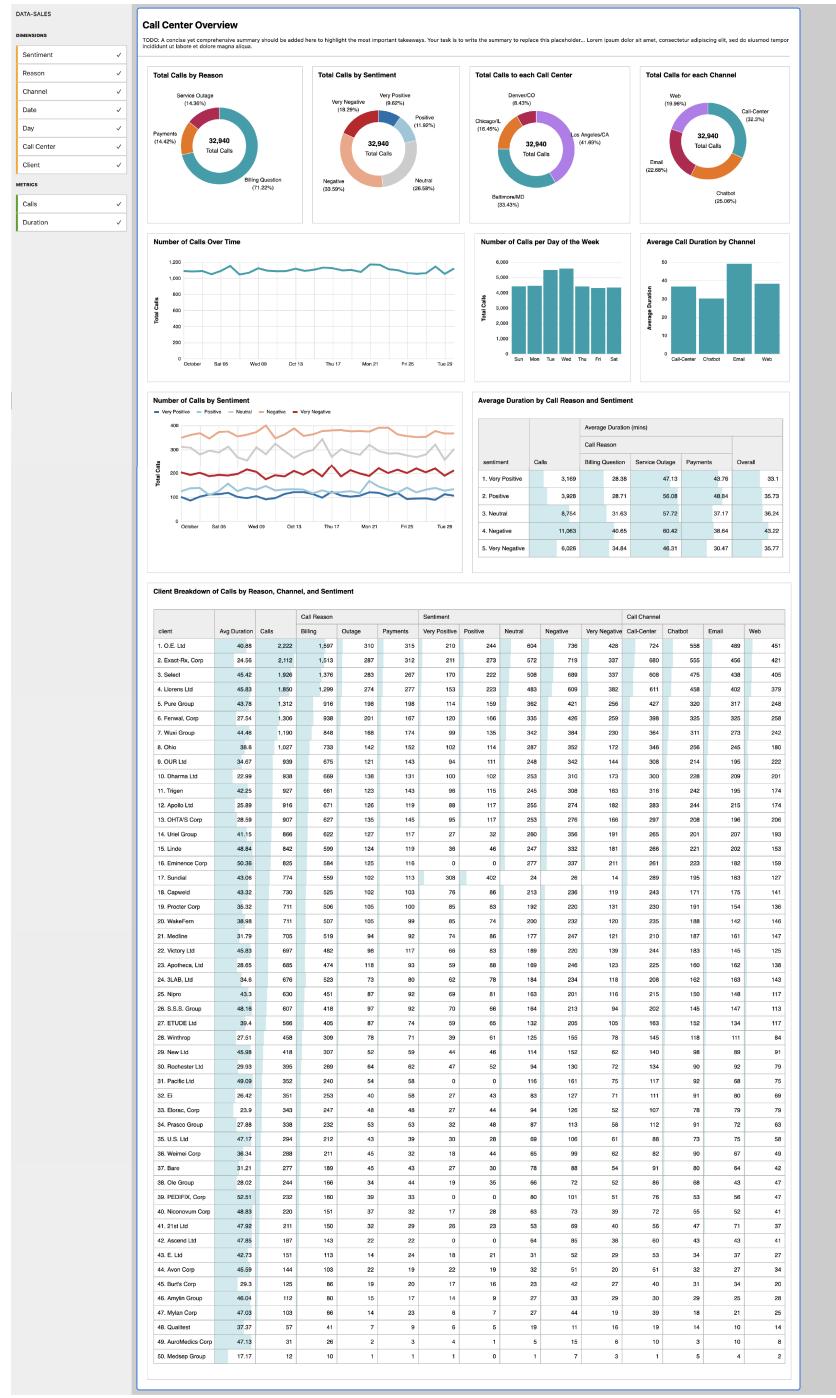


Fig. 3. The large “Call Center Overview” dashboard includes ten visualizations that utilize all of the available attributes: Calls, Sentiment, Reason, Duration, Channel, Date, Day, Client, Call Center. We generate 127 possible insights for this dashboard.

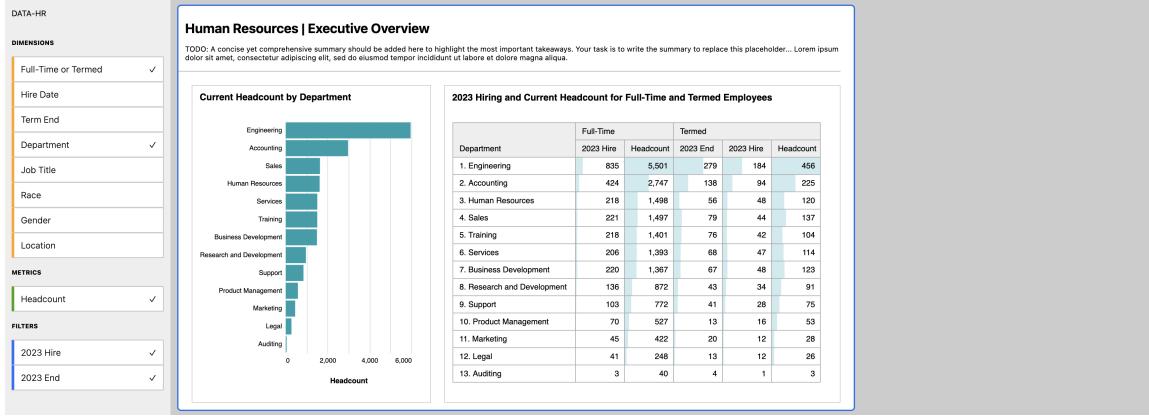


Fig. 4. The small “Human Resources | Executive Overview” dashboard includes two visualizations that focus on the Headcount for Full-Time and Termed employees by Department. We generate 30 possible insights for this dashboard.

2 Dashboard: Human Resources | Executive Overview

Our participants/approach selected an average of 8.6 insights for the “Human Resources” dashboard (7.7 for small, 7.5 for medium*, and 10.3 for large). The average summary length was 108.4 words (85 words for the small dashboard, 98 for medium*, and 138.7 for large); *only two samples (one human, one automatic) are available for the medium dashboard.

2.1 Small: Human Resources

(1) Human-authored Summary (hcs): “*Dashboard indicates different types of employee counts, including current headcount by department, full-time hired employees count 2023, terminated employees count 2023, hired [termed] employees in 2023, and termed employees headcount. All these items form a tail distribution due to 11 departments representing only 54–55% in total. All activities related to the engineering department, like hiring, termination and headcount are highly skewed and almost 100% higher than what constitutes the second-highest activity in accounting. The auditing region contains a very low headcount—about 0.2% in total. Therefore, its hired and termed employee count in 2023 is also low (0.1–0.5%).*”

(2) Human-authored Summary (hsc): “*It’s clear that the “Engineering” department dominates the workforce headcount by a huge margin (Almost 1/3) and has double the amount of workers of our second largest department “Accounting”. It’s showing the same dominance in 2023 full time hiring as well. This department combined with “Accounting” makes up about 50% of the entire workforce while the remaining 11 departments make up the other half. In 2023, for Engineering department, more termed workers were let go than the number of new termed workers who were hired. And lastly, Auditing remains the smallest department, both in terms of overall headcount and new hirings.*”

(3) Automatic Summary (hxx): “*The department with the highest headcount is Engineering, accounting for 30% of the total. The distribution of headcount is highly skewed towards Engineering. In terms of 2023 hiring for full-time employees, Engineering also has the highest value at 30% of the total. The distribution of 2023 hiring for full-time employees is highly skewed towards Engineering as well.*”

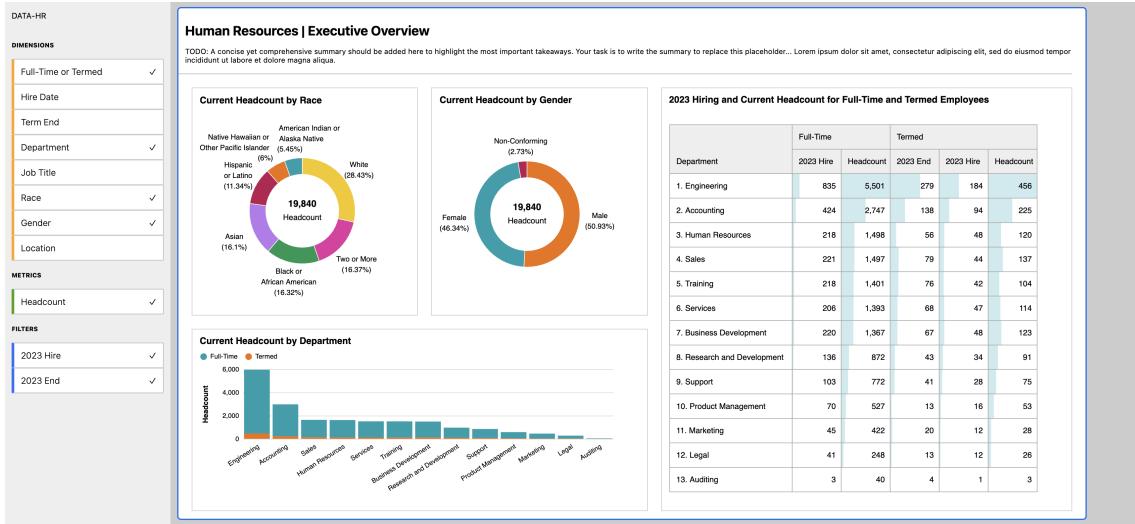


Fig. 5. The medium “Human Resources | Executive Overview” dashboard includes four visualizations, with the two new charts exploring breakdowns of Headcount by Race and Gender. We generate 41 possible insights for this dashboard.

2.2 Medium: Human Resources

As discussed in the full paper, one participant was unable to complete the study (shc), and as such, we only have one sample of a human-authored summary for the medium “Human Resources” dashboard.

(1) Human-authored Summary (chs): “The “White” category leads in headcount, it is 74% more than the second highest category, “Two or More Races”. With a headcount of 5,640, Whites are 28% of the total. The lowest headcounts are seen in the “American Indian or Alaska Native” and “Native Hawaiian or Other Pacific Islander” categories, with around 1,135.5 in “Headcount” (11% in total). There are 10% more Males than Females. Engineering dominates in all series, it is 97% to 102% more than the second highest category, Accounting. The rest of categories show a tail distribution, with 11 items. Auditing has the lowest numbers in all series”

(2) Human-authored Summary (shc): N/A

(3) Automatic Summary (xhx): “The highest headcount by race is for “White” with 5,640 people, accounting for 28% of the total. The smallest headcount by gender is for “Non-Conforming” with 542 people, accounting for 3% of the total. In terms of hiring for 2023, “Engineering” has the highest value with 835 full-time hires, accounting for 30% of the total. The department with the highest headcount is also “Engineering” with 5,501 full-time employees, accounting for 30% of the total. Both the 2023 full-time hires and the full-time headcount are highly skewed towards “Engineering” with 30% of the total in both cases.”

2.3 Large: Human Resources

(1) Human-authored Summary (csh): “*This summary covers everything under the HR scope we have a headcount of gender, ethnicity, employee type. It breakdown how our company is structured with where the most jobs are within each department. It shows there is opportunity outside of working a full-time job, as well as highlighting that the company is always hiring showing how often they receive new hires and within what part of the country.”*

(2) Human-authored Summary (sch): “*New hires for full-time positions were highest on June 30. There were 17 new hires, which was 126% more than the average of 7. For termed positions, hire count was highest last Oct 17 which was 199% more than the average of 1. Most full-time and termed positions come from Research Assistant II, Business Analyst, Human Resources Analyst II, Research Assistant I, and Account Executive. New hires are mostly from Ohio. Highest was on Apr 29 when Ohio had 64 new hires which was 26% higher than the average of 50. From the total headcount, most are full-time positions at 92%. Most common races are White (28%), Two or More (16%), Black or African American (16%), and Asian (16%). The highest headcount comes from Engineering (30% of total). Most new hires are from Engineering as well. End dates for termed employees are highly concentrated on the years 2032 and 2033.”*

(3) Automatic Summary (xxh): “*In 2023, the maximum number of full-time new hires occurred on June 20th, with a headcount of 17, which was 126% higher than the average of 7. The minimum number of full-time hires was 1, which occurred on February 21st, 87% lower than the average. An anomaly in headcount was observed on June 9th, where the value deviated by 239% from the expected value. In terms of new hires by week, Wisconsin had the most significant anomaly on February 5th, with a deviation of 378% from the expected value. Michigan had the lowest number of hires, with 0 on February 5th, which was 100% lower than the average of 1. The maximum number of full-time employees is 365% higher than the second highest category, termed employees. In terms of current headcount by department, Engineering had the highest value with 5,501 full-time employees, accounting for 30% of the total. The distribution of full-time employees is highly skewed towards Engineering. White employees had the highest headcount with 5,640, accounting for 28% of the total. In terms of future end dates for termed employees, the year 2023 had the highest number of terminations with 897, accounting for 36% of the total.”*

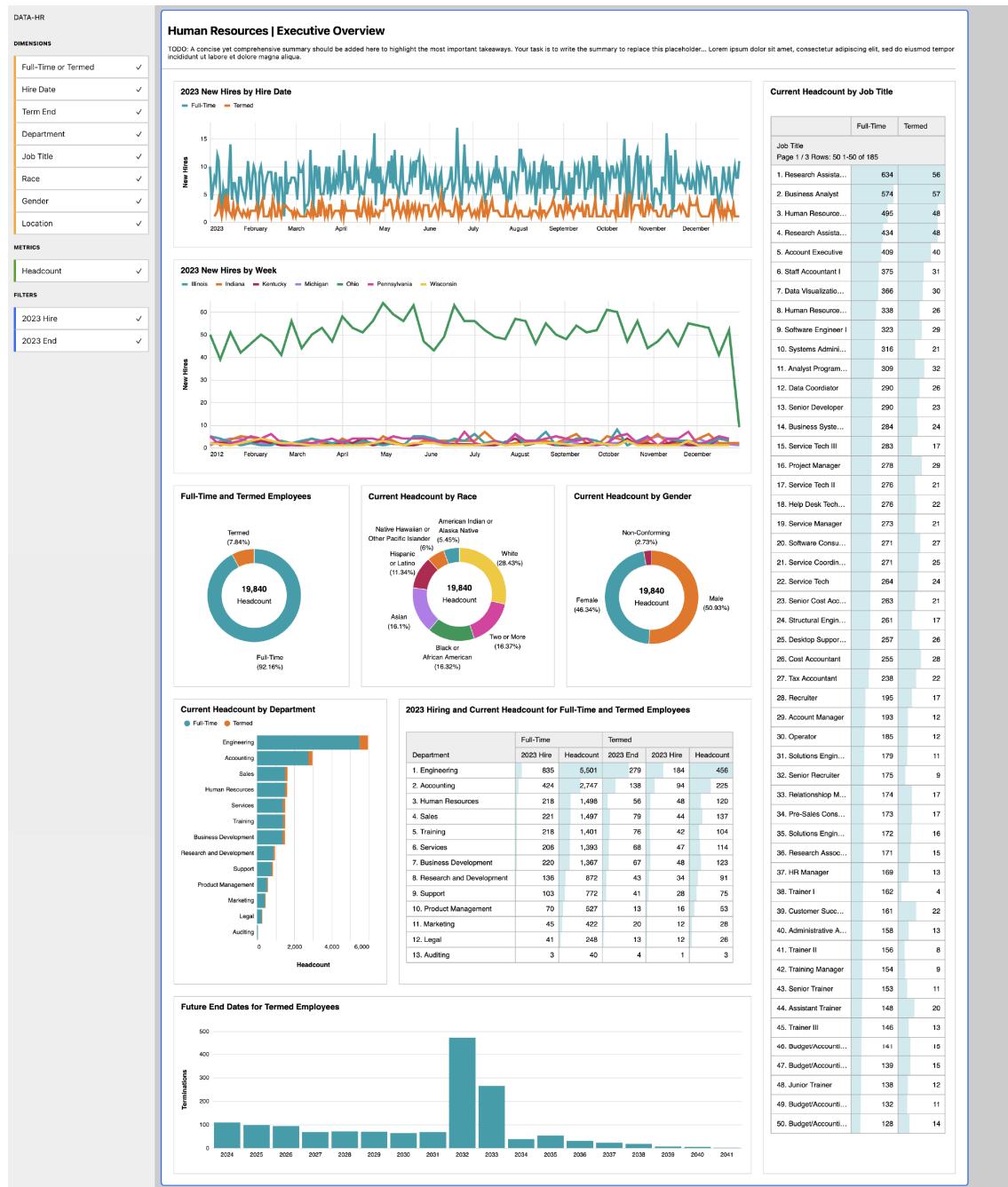


Fig. 6. The large “Human Resources | Executive Overview” dashboard includes nine visualizations based on the attributes: Headcount, Full-Time or Termed, Department, Hire Date, Gender, Race, Location, Job Title, and Term End; this dashboard also uses two data filters (2023 Hire and 2023 End). We generate 107 possible insights for this dashboard.

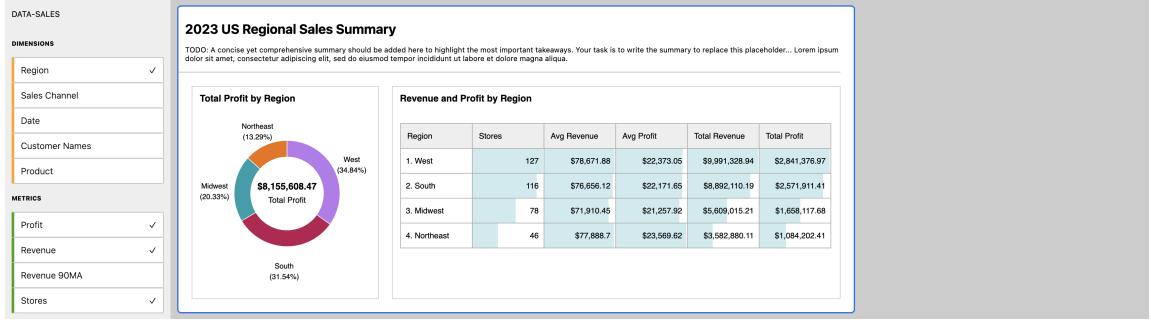


Fig. 7. The small “2023 US Regional Sales Summary” dashboard includes two visualizations that explore the Profit, Revenue, and number of Stores by Region. We generate 16 possible insights for this dashboard.

3 Dashboard: 2023 US Regional Sales Summary

For the natural language summaries (both the human-authored and LLM-based), our participants/approach selected an average of 8.9 insights for the “2023 US Regional Sales Summary” dashboard (5.5* for small, 7.3 for medium, and 12.7 for large). The average summary length was 125.9 words (97.5* words for the small dashboard, 119 for medium, and 151.7 for large); *only two samples (one human, one automatic) are available for the small dashboard.

3.1 Small: Sales Summary

As discussed in the full paper, one participant was unable to complete the study (shc), and as such, we only have one sample of a human-authored summary for the small “2023 US Regional Sales Summary” dashboard.

(1) Human-authored Summary (sch): “*West has the highest profit at \$2.84M, closely followed by South with \$2.57M. Combined, West and South account for \$5.41M which is 66% of total profit. Northeast has the lowest profit at \$1.08M. West and South also have the highest number of stores with 127 and 116, respectively. Combined, West and South account for 243 number of stores which is 66% of all total stores. Northeast has the lowest number of stores at 46. Although with the lowest total profit and number of stores. Northeast has the best Average Profit by Store at \$23,569. Average Profit by Store for all Regions is \$22,222. West has the highest revenue at \$9.99M, followed by South with \$8.89M. Combined, West and South account for \$18.88M which is 67% of total revenue.*”

(2) Human-authored Summary (shc): N/A

(3) Automatic Summary (sxx): “*The highest bars in terms of profit are in the West and South regions, with the West region having 10% more profit than the South region. The West region also has the highest revenue and profit, with 9% more stores and 12% more total revenue compared to the South region. Overall, the West and South regions contribute to the majority of the total profit and total revenue.*”

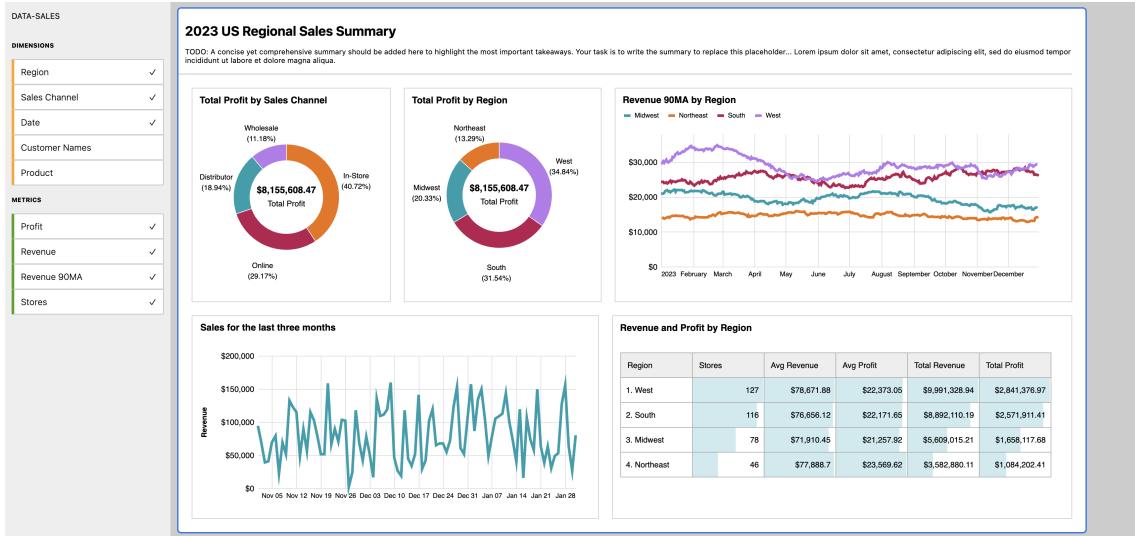


Fig. 8. The medium “2023 US Regional Sales Summary” dashboard includes five visualizations, with the new charts introducing breakdowns by Date, Sales Channel, and Revenue90MA. We generate 44 possible insights for this dashboard.

3.2 Medium: Sales Summary

(1) Human-authored Summary (hsc): “Overall, In-Store and Online sales contribute the most towards total profit, making up more than 2/3 of the total profit. In-store dominates online in terms of profits, contributing 10% more to total profit whereas Wholesale is performing the lowest, making just 11% of the total profit. Regional-wise, West and South make 2/3 of the total profit while Northeast contributes the lowest, just 13% of the total profit. The same pattern exists in revenue generation by region. West and South make 2/3 of the total revenue, whereas Northeast holds 13% of the total revenue. The low amount of In-Store sales in the Northeast region could be a reason for its low profits and revenue, since In-Store sales make up the most of our total profits. In the last 3 months, the highest revenue time period was Jan-24 to Jan-28, which saw an impressive increase of almost 415% compared to average revenue. The biggest peak came on Dec 9th, which was 98% more than the average. Nov-27 marked the lowest amount of Revenue recorded on a single day, 97% less than the average.”

(2) Human-authored Summary (csn): “The West contains the highest profit totals of the Regions with about 35%, as well as the most store in the data set with 127 stores. The South is a close second to the West with total profits accounting for about 31%, and the second most stores with 116. The Midwest accounts for only 20% of the total profits with 78 stores throughout. The Northeast from the outside looking in appears to be the weak link only accounting for 13% of the profits, with only 46 stores.”

(3) Automatic Summary (xsx): “The smallest total profit by sales channel is in "Wholesale" with \$911,677.77 (11% of the total). The highest total profit by region is in "West" and "South" with around \$2,706,644.595 (66% of the total). "West" has 10% more profit than "South". The revenue in the "South" region shows a cyclic pattern every 7 days, with the peak on day 3 and valley on day 4. The maximum sales revenue in the last three months was \$159,642.09 on Dec. 9th, 2023, which is 98% higher than the average of \$80,478.097.”

3.3 Large: Sales Summary

(1) Human-authored Summary (hcs): “The west and south regions contribute 66% of total profit, with 65% of stores located in these regions. The Northeast region generates the lowest profit (13.29%), with the fewest stores located in this region (12.253%). However, the Northeast region generates the highest average profit among all regions. The south region shows a cyclic pattern in revenue generation, with a peak on day 3, followed by a valley on day 4. In-store and online sales channels contribute 70% of total profit. Distributors and wholesalers bring only 30% of the total profit. A strong upward movement trend is observed in online sales profit from November 16th, 2023, to December 24th, 2023, indicating increased demand during the Christmas festive season. Products in the categories of collectibles, serverware, wall frames, accessories, furniture and furniture cushions generate a high amount of profit and revenue. Products in the categories of pillows and ornaments generate low revenue and profit across all channels. Customer channel Medline generates the highest revenue and profit.”

(2) Human-authored Summary (chs): “The Revenue chart displays a pattern of ups and downs, mostly between \$50,000 and \$100,000. A noticeable drop occurs from late November to mid-December, with a few sharp falls. The West region leads in revenue, particularly strong in the first quarter of 2023, and has the most stores. The Northeast, with the fewest stores, has the lowest revenue and profit, indicating a link between number of stores and financial performance. In-store sales lead in profit, having 41% of total profit, which is 40% higher than the next leading channel, “Online.” The Profit by Channel graph is marked by frequent fluctuations without clear long-term patterns, except for notable spikes and drops in profit, ranging from a 938% decrease to a 707% increase. When analyzing customer revenue by channel, “Medline” leads both in revenue and profit, which is more than the second-highest item by 49% in revenue and 33% in profit.”

(3) Automatic Summary (xss): “The revenue for December 9th, 2023, reached a peak of \$159,642.09, which was 98% higher than the average revenue of \$80,478.097. The regions of “West” and “South” had the highest total profits, with “West” being 10% higher than “South”. The sales channels of “In-Store” and “Online” had the highest total profits, with “In-Store” being 40% higher than “Online”. The region of “West” had the most number of stores, with 127 in total. The revenue in the “South” region exhibited a cyclic pattern occurring every 7 days, with peak values on day 3 and valleys on day 4. The online sales channel had the highest maximum profit of \$34,721.41 on January 11th, which was 404% higher than its average profit of \$6,885.426. The region of “West” had the greatest value in terms of revenue and profit, with 127 stores and 35% of the total.”

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009

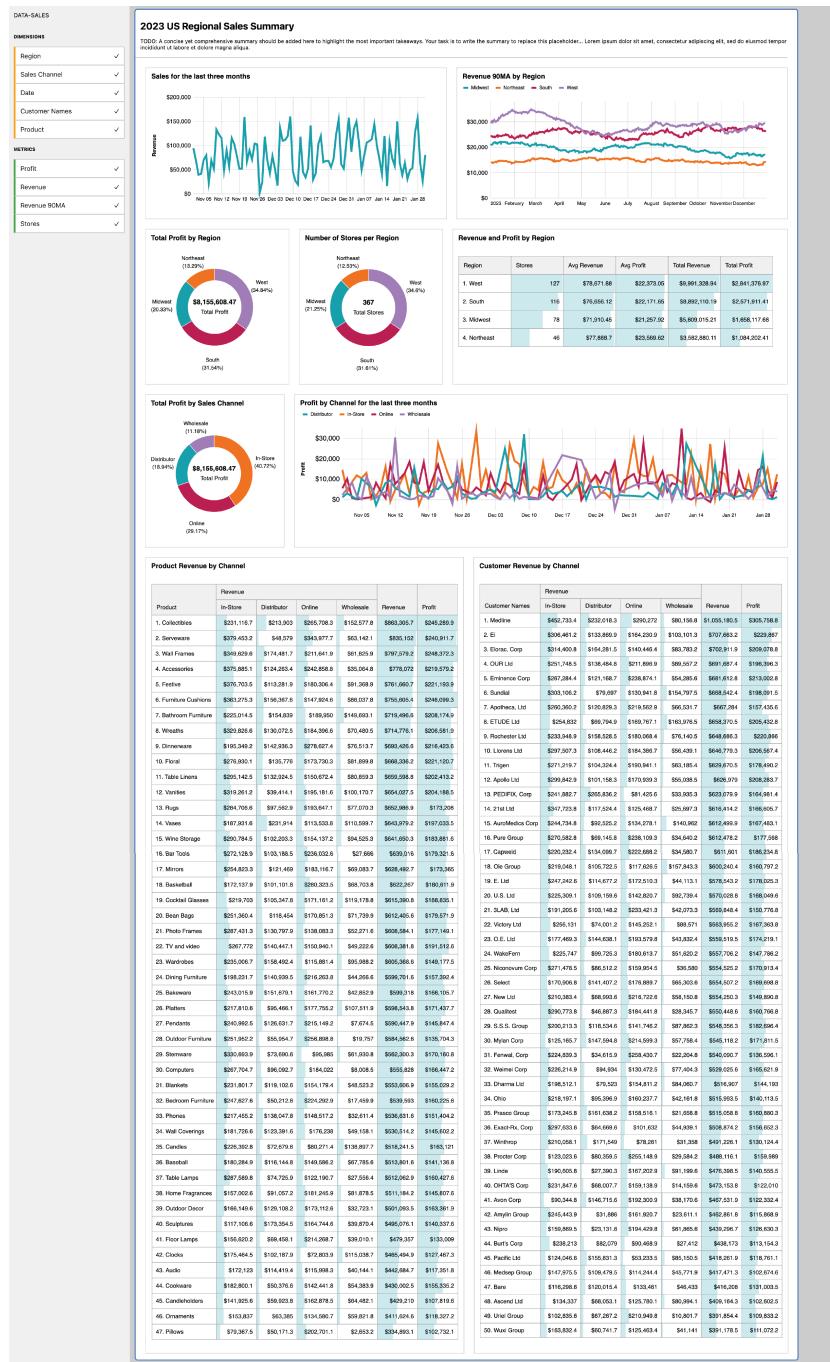


Fig. 9. The large “2023 US Regional Sales Summary” dashboard includes nine visualizations based on the attributes: Profit, Revenue, Region, Sales Channel, Date, Stores, Revenue 90MA, Product, and Customer. We generate 112 possible insights for this dashboard.