LinkedIn Analytics Practicum

Mohamed Berro

Project Manager: Hamin Oh

Introduction/Problem Statement:

The purpose of this project is to assist the Financial Planning and Analysis (FP&A) team by building an AI agent that will provide insights into the LinkedIn Marketing Solution (LMS) business line. The agent will analyze the Monetization DED (MDED) report that is generated daily and output a narrative detailing any key insights that may be pertinent to the CFO and all of his leads. The Monetization report contains an array of metrics that are aggregated across various periods -- ie weekly, monthly, and quarterly-to-date. These metrics detail how the business is performing across various dimensions such as region, vertical, and segment. Additionally, the report includes a one-year revenue projection of three core metrics, Net Bookings (C$), Budgets, and Budget Utilization. This projection, known as Plan, estimates the fiscal year-end values for those three metrics and is broken up into quarterly estimates known as Outlook. Business analysts analyze the MDED to determine if LMS is either at risk or on track of reaching its quarterly Outlook/Plan goals. If a particular line of business is at risk, the root cause must be determined. This task is quite tedious as BA's need to spend considerable amounts of time querying the MDED to determine why a business line is at risk. Ultimately, the goal of this project is to simplify the BA's work by implementing an AI agent to do the tedious work of querying the MDED and presenting a narrative that provides key insights into business line performance. In terms of development milestones, development of the agent will proceed with the completion of each of the following milestones:

Milestones

1. The agent will be capable of identifying any anomalies with the three-core metrics that put the business line at risk of meeting quarterly/yearly performance benchmarks. At the behest of the project's lead, the anomaly detection algorithm should not be anything advanced as LinkedIn already has an anomaly detection process in place. Instead focus should be on implementing a rules-based procedure that will identify the core metrics that are at risk of meeting performance goals.
2. Second, the agent will look to see if there's any relationship between the anomalies. For instance, if there were anomalies detected in both net bookings and budget then it's likely a result of advertisers not putting their money into the system, hence we're not able to utilize that budget and turn it into revenue.
3. Next, the agent will analyze all the underlying dimensions of any identified anomalies to determine the root cause for poor performance.
4. Finally, the agent will generate a narrative that captures key insights from its analysis. This will entail training an LLM on the data gathered from Milestones 1, 2, and 3. Additionally, prompt engineering will be needed to generate a prompt that the model can use to create a

comprehensive narrative detailing the insights discovered from the MDED. This will be a critical milestone that the project will be evaluated on.

My Approach:

This is a multi-faceted and open-ended project with many different paths. I considered multiple solutions and decided to proceed with a two-phase approach. The first phase will involve feature engineering the data by deriving multiple metrics the provide further insights into the data. The second phase will involve building and training a machine learning algorithm such as Isolation Forest and training the model with the features I engineered. Of course, I could go with the deep learning path and build a neural net that automatically learns the data representation without needing to do any feature engineering. However, these types of models require immense amounts of data and can potentially require more compute resources. Given the initial data and compute constraints, I elected with the feature engineering approach – though I later discovered I can produce an immense amount of data from the data on hand using these metrics. Regardless, the metric-driven approach is still a worthwhile endeavor I am hopeful will produce solid results.

Work Completed:

Milestone 1 (complete). For Milestone 1 I have been able to build out a rules-based anomaly detection procedure that detects when the LMS business line is at risk of meeting it's quarterly projections. I have built the procedure for the Net Bookings, Budget, and Budget Utilization metrics. To illustrate, please consider the below scenario using a current date of 09/22/2024, and a quarter end of 09/30/2024:

- **Derive Net Bookings for Last 7 Days (L7D)**

segment	current_num	p/p	y/y	daily_run_rate
ESG	3.9481e+07	6.7503	-0.59059	5.6402e+06

- Since the quarterly Outlook is broken out across the Segment dimensions ESG, OSO, SA, I have filtered the data by Net Bookings and L7D and then aggregated the filtered data by Segment. The current_num represents the revenue for the last 7 days. The p/p represents period-over-period percent change and y/y represents year-over-year percent change for last 7 days. The daily_run_rate is computed by dividing the current_num by 7. Here, daily_run_rate informs us the current trajectory of a particular segment and will be used later to predict quarter-end revenue. We can see from p/p and y/y that segment ESG appear to be experiencing a healthy growth rates in terms of recent performance. However, its yearly growth is concerning and warrants investigation.
- **Derive Net Bookings for Weekly and Monthly Periods**

segment	period_label	p/p_weekly	y/y_weekly	p/p_l7d	y/y_l7d
ESG	2024-08-25	3.0229	4.0312	6.7503	-0.59059
ESG	2024-09-01	-0.65802	0.16694	6.7503	-0.59059
ESG	2024-09-08	11.572	1.1361	6.7503	-0.59059
ESG	2024-09-15	6.7503	-0.59059	6.7503	-0.59059

segment	period_label	p/p	y/y
ESG	2024-6	2.0898	5.4571
	2024-7	-12.35	6.7752
	2024-8	-2.1243	1.4122
	2024-9	19.044	0.21594

- I've also created weekly and monthly p/p and y/y dataframes for further investigation.
- Here, we look at the quarterly-to-date performance period for Net Bookings of the three segments and compare this period to our quarterly outlook projections. The current_num_qtd is our quarter-to-date revenue and current_num_outlook is our outlook projection for that quarter. The otd is outlook-to-date and is derived by multiplying the Outlook by the percent of quarter completed, which is 91%. This field is important in performing a direct comparison between qtd and outlook. Using the aforementioned fields, we can derive qtd_to_otd_percent as the percent of outlook-to-date that is captured by quarter-to-date. qtd_to_otd_variance is the actual amount qtd is off from otd. The projected_qr is projected quarterly revenue and is derived by multiplying the number of days remaining in the quarter by the daily_run_rate and adding the result to current_num_qtd. This field is used to directly compare qtd to outlook – it's what we're expecting the quarter-end revenue to be given recent performance. For projected_qr_to_outlook_variance we determine how much projected_qr is short or over outlook. If projected_qr_to_outlook_variance is negative then that particular segment is at risk of missing its quarterly projections. According to the results provided below ESG is currently showing at risk of missing its quarterly projections and further investigation of this segment's dimensions are needed to understand the root problem(s).

segment	current_num_qtd	current_num_outlook	otd	qtd_to_otd_percent	qtd_to_otd_variance	projected_qr	projected_qr_to_outlook_variance	at_risk
ESG	3.9165e+08	4.5956e+08	4.1916e+08	93.438	-2.7504e+07	4.3678e+08	-2.2783e+07	1

Milestone 2 (complete). For milestone 2, I have determined the relationships for the core metrics depending on the performance of each metric. Below are those relationships:

1. Revenue (Bookings) Down, Budget Down
   - Explanation: Advertisers are not putting enough money into the system.
   - Reason: Indicates a **demand-side issue**, such as reduced budgets due to economic constraints or shifting ad spend.

2. Revenue (Bookings) Down, Budget Flat
- Explanation: Inefficiencies in utilizing available budgets.
- Reason: Declining budget utilization due to supply-side constraints or performance issues.

3. Revenue (Bookings) Down, Budget Up

- Explanation: Growing budgets are not effectively converted into bookings.

4. Revenue (Bookings) Flat, Budget Down
- Explanation: Improved utilization offsets reduced budgets.
- Reason: Efficiency improvements in budget utilization compensating for lower budgets.

5. Revenue (Bookings) Flat, Budget Up
- Explanation: Utilization issues prevent increased budgets from driving bookings growth.
- Reason: Stagnant performance despite increased budgets, likely due to limited inventory or audience mismatches.

6. Revenue (Bookings) Up, Budget Down
- Explanation: Higher utilization compensates for lower budgets.
- Reason: **Optimized utilization and performance leading to higher bookings despite reduced budgets.

7. Revenue (Bookings) Up, Budget Flat
- Explanation: Improved utilization drives bookings growth.
- Reason: Healthy performance with better conversion of budgets into bookings.

8. Revenue (Bookings) Up, Budget Up
- Explanation: Balanced growth in both advertiser demand and utilization.
- Reason: Strong overall performance with effective utilization of increased budgets.

Milestone 3 (complete). For this section I want to cover some of the details of my approach, particularly, provide some explanations of the metrics I've used and their importance. The hope is for the reader to gain an understanding of my reasoning for the significance of my approach. The first phase of my approach involves deriving the metrics, identifying how to apply the metrics, and effectively leveraging these metrics to produce key insights. Given that the data is broken out across multiple dimensions and time periods, there are an immense number of permutations that can be applied to the metrics. I initially decided to go with a one-layer dimension breakdown, e.g. segment-to-product, and analyze how each dimension impacts segment.

Since the metrics I integrated into my solution are targeted at identifying the various attributes influencing the performance of segment, this approach should not only help the model better learn the data representations, but tailor it to more effectively spot faults in segment. Because ESG's performance is determined by the cumulative impact of its various dimensions, I believe it is important to identify the components that have the most impact. Metrics are important for identifying impact, however each metric only tells part of the story and focusing on only one or two metrics can be misleading.

				y/y_change
chart_label	period_label	segment	product	
Quarterly-to-date	FY25Q1	ESG	FEED	-1.869
			INMAILS	40.743
			LAN	5.4126
			OTHER	7.3118

For example, if the product INMAILS had a year-over-year growth of 40.7% while FEED declined by only 1.9%, it could possibly lead one to believe that INMAILS had more of a significant impact on segment's performance. However, this observation can be misleading as year-over-year change captures overall growth dynamics but does not reflect alignment, proportionality, or sensitivity within the segment. It does not account for how much of segment's revenue each product represents or how sensitive segment's performance is to changes in product's performance.

				y/y_change	cont_percent
chart_label	period_label	segment	product		
Quarterly-to-date	FY25Q1	ESG	FEED	-1.869	-56.611
			INMAILS	40.743	111.55
			LAN	5.4126	38.892
			OTHER	7.3118	6.1712

Now, if we look at the contribution percentage, which is the year-to-year variance of product divided by the year-to-year variance of segment – basically, this metric reveals the extent to which a product is driving the segment's performance as a percentage of the segment's total movement. Here we see that though FEED had the lowest year-over-year percent change in terms of magnitude, we see that FEED actually has the second highest contribution percentage, in terms of magnitude at -56.6%. So even though FEED had a miniscule year over year decline, it contributed significantly to how much revenue segment gained or lost over the year. However, Contribution Percentage focuses on raw contribution without adjusting for product size or sensitivity.

				product_y/y_change	cont_percent	proportion	weighted_cont
chart_label	period_label	segment	product				
Quarterly-to-date	FY25Q1	ESG	FEED	-1.869	-56.611	0.73775	-41.764
			INMAILS	40.743	111.55	0.066684	7.4384
			LAN	5.4126	38.892	0.17501	6.8065
			OTHER	7.3118	6.1712	0.020557	0.12686

Taking this analysis further, we can observe the weighted contribution value, which adjusts for the product's size relative to the segment, providing a nuanced view of proportional impact. Note that this value is not a percentage.

				product_y/y_change	cont_percent	proportion	weighted_cont	rev_elasticity
chart_label	period_label	segment	product					
Quarterly-to-date	FY25Q1	ESG	FEED	-1.869	-56.611	0.73775	-41.764	-1.3032
			INMAILS	40.743	111.55	0.066684	7.4384	0.059781
			LAN	5.4126	38.892	0.17501	6.8065	0.44999
			OTHER	7.3118	6.1712	0.020557	0.12686	0.33311

Here, the Revenue Elasticity captures the segment's sensitivity to changes in the product's revenue, identifying products whose revenue changes disproportionately affect segment performance. It shows where small changes in product revenue may have outsized or muted effects on segment performance. Revenue elasticity values greater than 1 or less than -1 is considered significant, suggesting that increases or decreases in product's revenue generate more-than-proportional growth or decline in the segment's revenue. Values greater than .3 or less than -.3 are considered moderately significant. A moderate positive revenue elasticity suggests that increases in product's revenue result in steady, predictable growth for segment. This alignment supports the segment's performance without creating excessive volatility

				product_y/y_change	cont_percent	proportion	weighted_cont	rev_elasticity	CP_RE
chart_label	period_label	segment	product						
Quarterly-to-date	FY25Q1	ESG	FEED	-1.869	-56.611	0.73775	-41.764	-1.3032	73.775
			INMAILS	40.743	111.55	0.066684	7.4384	0.059781	6.6684
			LAN	5.4126	38.892	0.17501	6.8065	0.44999	17.501
			OTHER	7.3118	6.1712	0.020557	0.12686	0.33311	2.0557

The last metric I will discuss in this example, though I have incorporated other metrics into my solution, is the result from applying revenue elasticity to contribution percentage. One of the more important metrics, CP_RE has a dual perspective in uniquely capturing both the scale and responsiveness of a product's impact on the segment. It is good for identifying products that amplify or dampen segment performance. Values greater than 50 suggest that the product is not only a key driver in the segment, but also that the segment's performance is highly responsive to changes in the product's revenue.

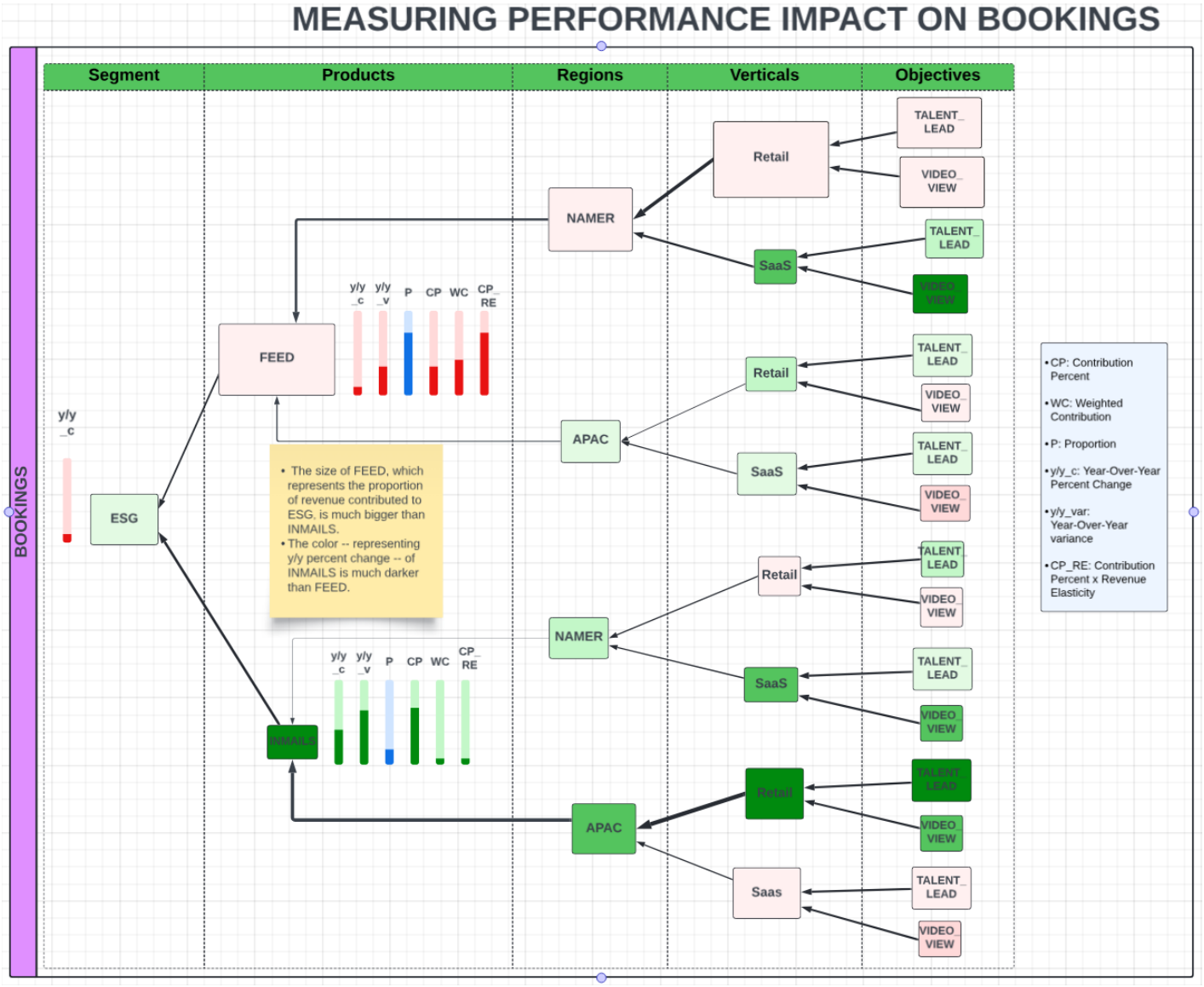
chart_label	period_label	segment	product	product_y/y_change	cont_percent	proportion	weighted_cont	rev_elasticity	CP_RE	segment_y/y_change
Quarterly-to-date	FY25Q1	ESG	FEED	-1.869	-56.611	0.73775	-41.764	-1.3032	73.775	2.4356
			INMAILS	40.743	111.55	0.066684	7.4384	0.059781	6.6684	2.4356
			LAN	5.4126	38.892	0.17501	6.8065	0.44999	17.501	2.4356
			OTHER	7.3118	6.1712	0.020557	0.12686	0.33311	2.0557	2.4356

Finally, we can see the segment year-over-year change, resulting from the combined impacts of the four products. What's striking is notice the large gap between INMAILS percent change and segment's. that there is a 38% difference between INMAILS year over year change and segment's. Even though its percent change is 40% and its contribution percentage is 111.5%, its stellar performance had a negligible impact on segment's percent change. This low level of segment responsiveness to INMAILS is made evident by INMAILS miniscule revenue elasticity and low CP_RE values, both of which excels in measuring how responsive segment is to product shifts. This alone shows why it's important to understand the strengths and limitations of metrics. Leveraging one or two metrics can be misleading.

Another observation is that the average percent change of the four products is close to 13%, yet segment's percent change isn't close to this value either – at little over 2%. That's because FEED accounts for such a significant proportion of segment's revenue. As shown by its hefty weighted contribution, revenue elasticity, and CP_RE, FEED is truly the main driver of segment's performance. Even though segment is highly responsive to FEED's drastic negative performance, segment still has a positive growth. Shifting our focus from FEED to LAN, we can see from LAN's solid revenue elasticity and CP_RE values, that its performance has some influence over segment, though not to the level of FEED. Finally, one last interesting observation I want to highlight is the revenue_elasticity of INMAILS compared to OTHER. Notice that its OTHER's elasticity value is significantly higher than INMAILS though the two products have similar revenue proportions. Clearly, if anything, INMAILS should have the higher revenue elasticity. But not only is it smaller – it's much smaller. This occurrence is an anomaly that happened purely from happenstance and not because OTHER exhibits performant attributes. Revenue elasticity is calculated using segment percent change and product percent change. Since segment change is largely determined by the more dominant products such as FEED and LAN, and OTHER's percent change is coincidentally closer in value to segment change than INMAILS is, the revenue elasticity is higher for OTHER than for INMAILS. With that said revenue elasticity is still a useful metric, that becomes more powerful when combined with other metrics – that's why CP_RE is so important.

The combination of these metrics should hopefully give the agent a deeper understanding of how the performance of each product uniquely impacts the outcome of segment. The agent should compare the metrics with other metrics both within the product and across the products to understand the relationship dynamics of the products. The goal here is to have the agent identify not only the products that are driving segment's performance, but also the product's that are showing great promise.

Regarding my implementation, I have been able to successfully implement these metrics, along with others, for all of the dimensions. Please see the below figure for a visual aid of the metrics and the general relationships of the data.



Milestone 4 (complete). For this milestone, I have implemented a multi-agent architecture using LangChain that leverages a few-shot approach to instruct my agents on how to process the analysis from my interpretation algorithm. My architecture is as follows:

- Of the 16 total agents, there are
    - A set of agents that handles Bookings dimension analysis. Each dimension is as follows:
        - **Product**: Here, there are two agents. One agent handles processing the interpretations of the metrics that measures product's impact on segment.

Another agent will then enhance the output of the interpretation agent with the analysis performed on the metrics such as CTR and Impressions
        - **Region:** Since there is no region specific metrics, I have 1 agent to handle the metric interpretations for region's impact on segment.
        - **Vertical:** As with region, there is only 1 agent to handle vertical's impact on segment.
        - **Objective:** Like with product, there are two agents. One agent handles processing the interpretations of the metrics that measures product's impact on segment. Another agent will then enhance the output of the interpretation agent
    - A set of agents that handles Budgets dimension analysis. Each dimension is as follows:
        - **Product**: Here, there are two agents. One agent handles processing the interpretations of the metrics that measures product's impact on segment. Another agent will then enhance the output of the interpretation agent with the analysis performed on the metrics such as CTR and Impressions
        - **Region:** Since there is no region specific metrics I have 1 agent to handle the metric interpretations for region's impact on segment.
        - **Vertical:** As with region, there is only 1 agent to handle vertical's impact on segment.
        - **Objective:** Like with product, there are two agents. One agent handles processing the interpretations of the metrics that measures product's impact on segment. Another agent will then enhance the output of the interpretation agent
    - An agent to analyze the 3 core metrics and determine their relationships
    - Two agents to analyze the outputs from the Booking's and Budget's agents and extract the split from each dimension that has the most significant performance impact on segment.
    - One agent to analyze the findings from the two prior agents, identify the most significant contributors and integrate all its discoveries with the core metric relationships to formulate a final narrative.

- Below is the output of my objective agents:

Objective to Segment Analysis (Enhanced)

Segment: ESG
Objective: LEAD_GENERATION

- **Noteworthy Metrics and Values**:
  - **Influence**: -4.9566 (High negative influence)
  - **Contribution Percentage**: 276.47% (Significant contribution to segment decline)
  - **Weighted Contribution**: 99.70% (Size-adjusted negative impact)
  - **Contribution-Adjusted Growth**: -13.70330 (Underperformance relative to expectations)
  - **YoY Percent Change**: -13.75% (Significant year-over-year decline)
  - **Impressions**: -12.016 (Decline in impressions)
  - **Clicks**: 84.336 (Increase in clicks)
  - **CTR**: 109.51% (Significant increase in click-through rate)
- **Explanation**: LEAD_GENERATION is a substantial drag on the ESG segment, with a **high negative Influence (-4.9566)** and **Contribution Percentage (276.47%)** indicating its critical role in exacerbating the segment's downturn. The **Weighted Contribution (99.70%)** highlights its disproportionate impact, suggesting urgent attention is needed to address its underperformance. Despite a **-12.016 decline in impressions**, the **84.336 increase in clicks** and a **CTR of 109.51%** suggest that while fewer users are seeing the content, those who do are more engaged. This indicates a potential issue with reach rather than content quality, necessitating a strategy to improve visibility.

Segment: ESG
Objective: BRAND_AWARENESS

- **Noteworthy Metrics and Values**:
  - **Influence**: 3.6202 (High positive influence)
  - **Contribution Percentage**: -201.93% (Major mitigation of the segment's negative trend)
  - **Weighted Contribution**: -21.58% (Significant stabilizing effect despite the segment's decline)
  - **YoY Percent Change**: 33.87% (Strong year-over-year growth)
  - **Impressions**: 48.15 (Increase in impressions)
  - **Clicks**: 28.806 (Increase in clicks)
  - **CTR**: -13.057% (Decrease in click-through rate)
- **Explanation**: BRAND_AWARENESS acts as a key stabilizing force for the segment, as indicated by its **high Influence (3.6202)** and **Contribution Percentage (-201.93%)**, which demonstrate its capacity to offset the segment's decline. Its **Weighted Contribution (-21.58%)** shows its size-adjusted impact in mitigating losses, while its **strong YoY growth (33.87%)** highlights its potential as a growth driver. The **increase in impressions (48.15)** and **clicks (28.806)** indicates effective outreach, although the **CTR decrease (-13.057%)** suggests that while more users are seeing the content, engagement per impression may need to be improved.

Segment: ESG
Objective: VIDEO_VIEW

- **Noteworthy Metrics and Values**:
  - **Influence**: -0.7085 (High negative influence)
  - **Contribution Percentage**: 39.52% (Significant contribution to segment decline)
  - **Weighted Contribution**: 2.38% (Neutral size-adjusted impact)
  - **YoY Percent Change**: -11.75% (Moderate year-over-year decline)
  - **Impressions**: -12.46 (Decline in impressions)
  - **Clicks**: -25.535 (Decline in clicks)
  - **CTR**: -14.936% (Decrease in click-through rate)
- **Explanation**: VIDEO_VIEW's **high negative Influence (-0.7085)** and **Contribution Percentage (39.52%)** indicate that it is a notable factor in the segment's decline. The **Weighted Contribution (2.38%)** suggests a neutral impact when adjusted for size, while the **YoY decline of -11.75%** signals a need for strategic intervention to prevent further deterioration. The **-12.46 decline in impressions** and **-25.535 decline in clicks** further emphasize the need for a renewed focus on content strategy, as the **CTR decrease (-14.936%)** indicates that the current video content may not be resonating with the audience.

- Below is the output of my agent who handles core metric relationships:

Segment: ESG

- **Period Type: Quarterly-to-date**
  - **Scenario Identified**: Revenue (Bookings) Down, Budget Down
  - **Explanation**:
    - **Bookings**: The current bookings show a year-over-year percentage change of +2.4356%, indicating a slight increase compared to the previous year. However, the percentage change from the previous period is -7.1949%, suggesting a decline in bookings compared to the last quarter.
    - **Budgets**: The current budgets show a year-over-year percentage change of -1.7928%, indicating a decrease in budget allocation compared to the previous year. The percentage change from the previous period is -7.4757%, indicating a significant reduction in budgets compared to the last quarter.
    - **Budget Utilization**: The budget utilization shows a year-over-year percentage change of +1.3441%, indicating improved efficiency in utilizing the available budget. However, the percentage change from the previous period is -0.17347%, suggesting a slight decline in utilization efficiency compared to the last quarter.

  Given that both bookings and budgets are down, this scenario indicates that advertisers in the ESG segment are not putting enough money into the system, leading to a demand-side issue. The decline in bookings despite a slight increase in year-over-year bookings suggests that the current economic conditions or strategic shifts may be affecting advertiser spending.

Overall Insights

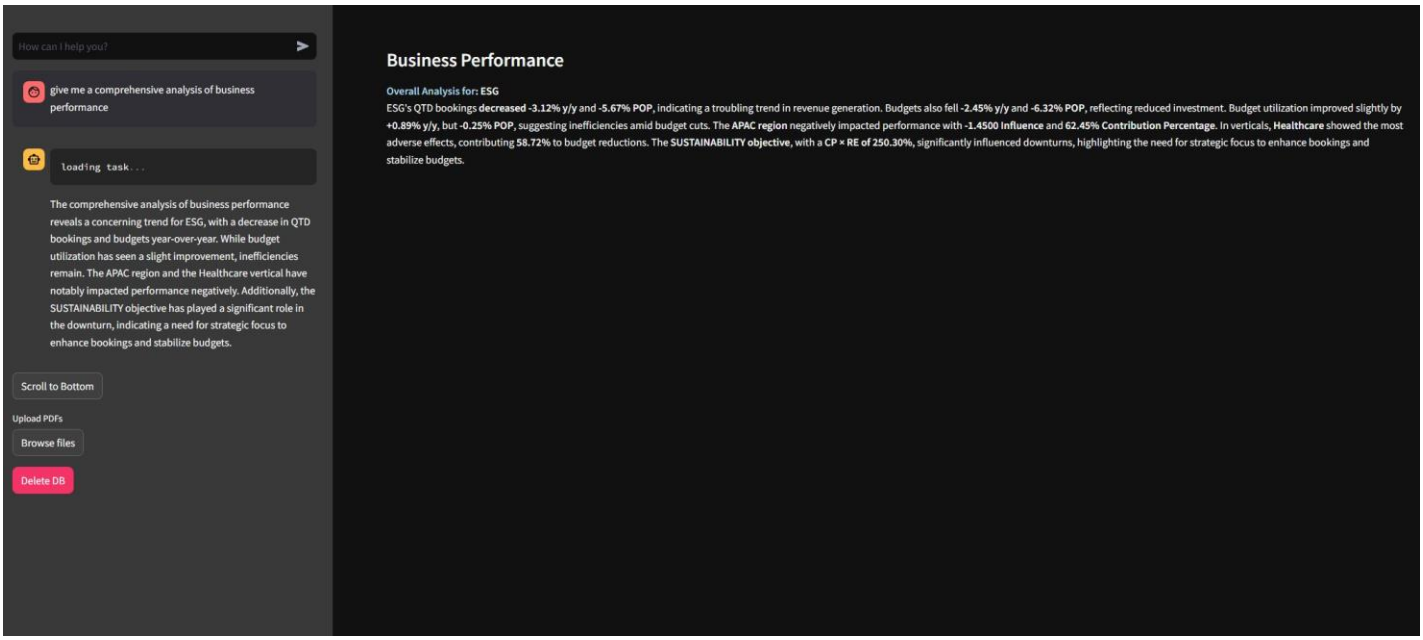
- The analysis of the ESG segment reveals a concerning trend where both bookings and budgets are declining in the Quarterly-to-date period. This indicates a potential demand-side issue, where advertisers are reducing their spending, possibly due to economic constraints or shifts in advertising strategies.
- The slight improvement in budget utilization year-over-year suggests that there may be some efficiency in how budgets are being spent, but the overall decline in both bookings and budgets indicates that this efficiency is not enough to offset the reduced investment in advertising.
- The relationship between **Bookings**, **Budgets**, and **Budget Utilization** in this case highlights the importance of maintaining adequate budget levels to drive bookings growth. Without sufficient investment, even efficient utilization cannot compensate for the lack of demand.

In summary, the ESG segment is facing challenges that require attention to both budget allocation and strategies to enhance bookings, particularly in the context of the current economic landscape.

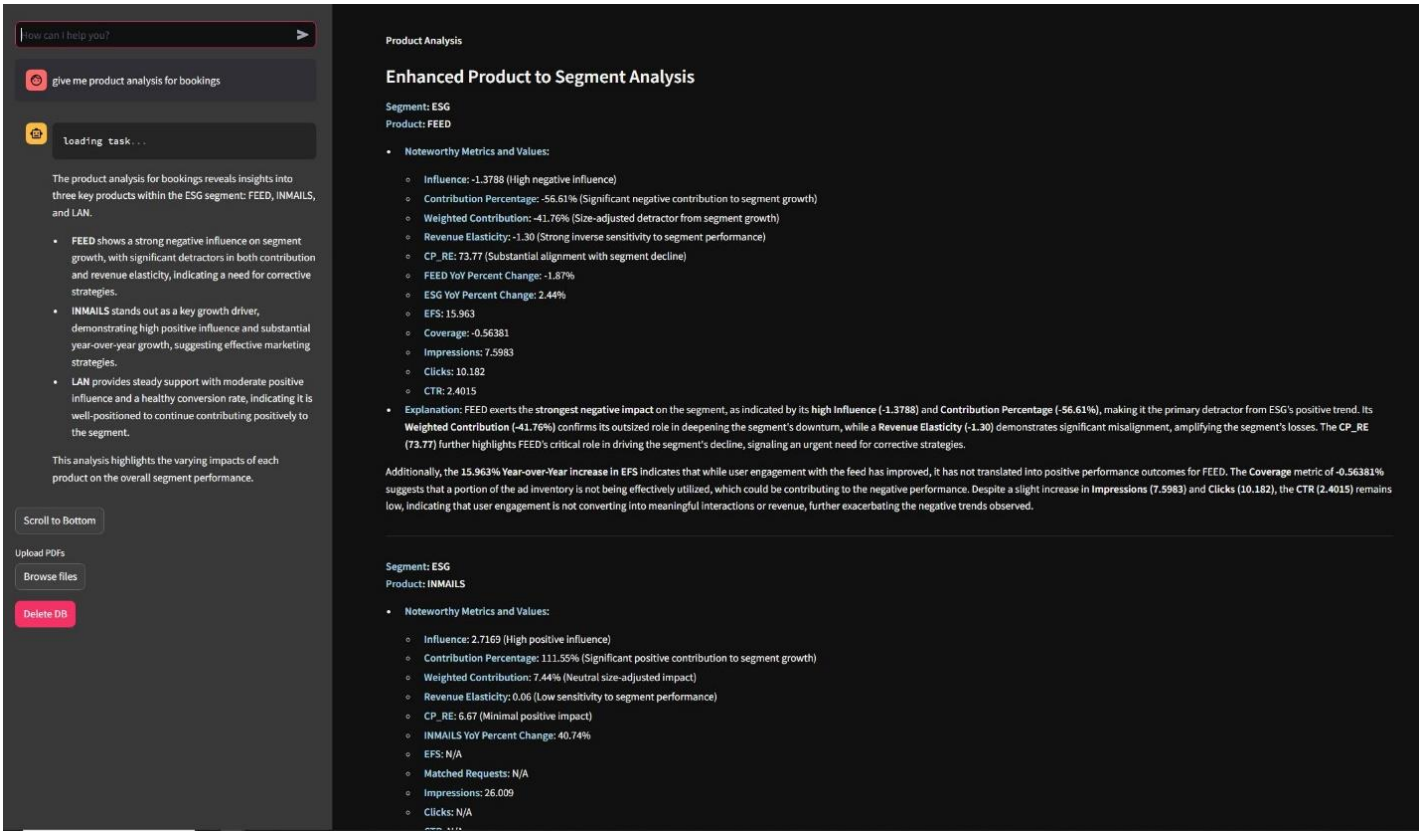
- Below is the final narrative produced from the agent's analysis for Q1:
  - "ESG's QTD bookings rose +2.44% y/y but fell -7.19% POP, indicating a troubling short-term decline. Budgets decreased -1.79% y/y and -7.48% POP, reflecting reduced advertiser investment. Budget utilization improved slightly by +1.34% y/y but declined -0.17% POP, suggesting efficiency gains are insufficient. EMEAL negatively impacts the segment with an Influence of -1.1990 and a Contribution Percentage of 66.88%. Professional Services also detracts significantly, contributing 63.13% to budget declines. Conversely, NAMER shows potential with a Revenue Elasticity of 1.36, while BRAND_AWARENESS stabilizes the segment with a 33.87% YoY growth, highlighting areas for strategic focus."

Milestone 5 (optional; complete). I have completed the bonus milestone of incorporating a question-and-answer feature to my LLM and overlayed a graphical user interface on my implementation using Streamlit. Please see the screenshots of the conversation with the LLM via the user interface below.

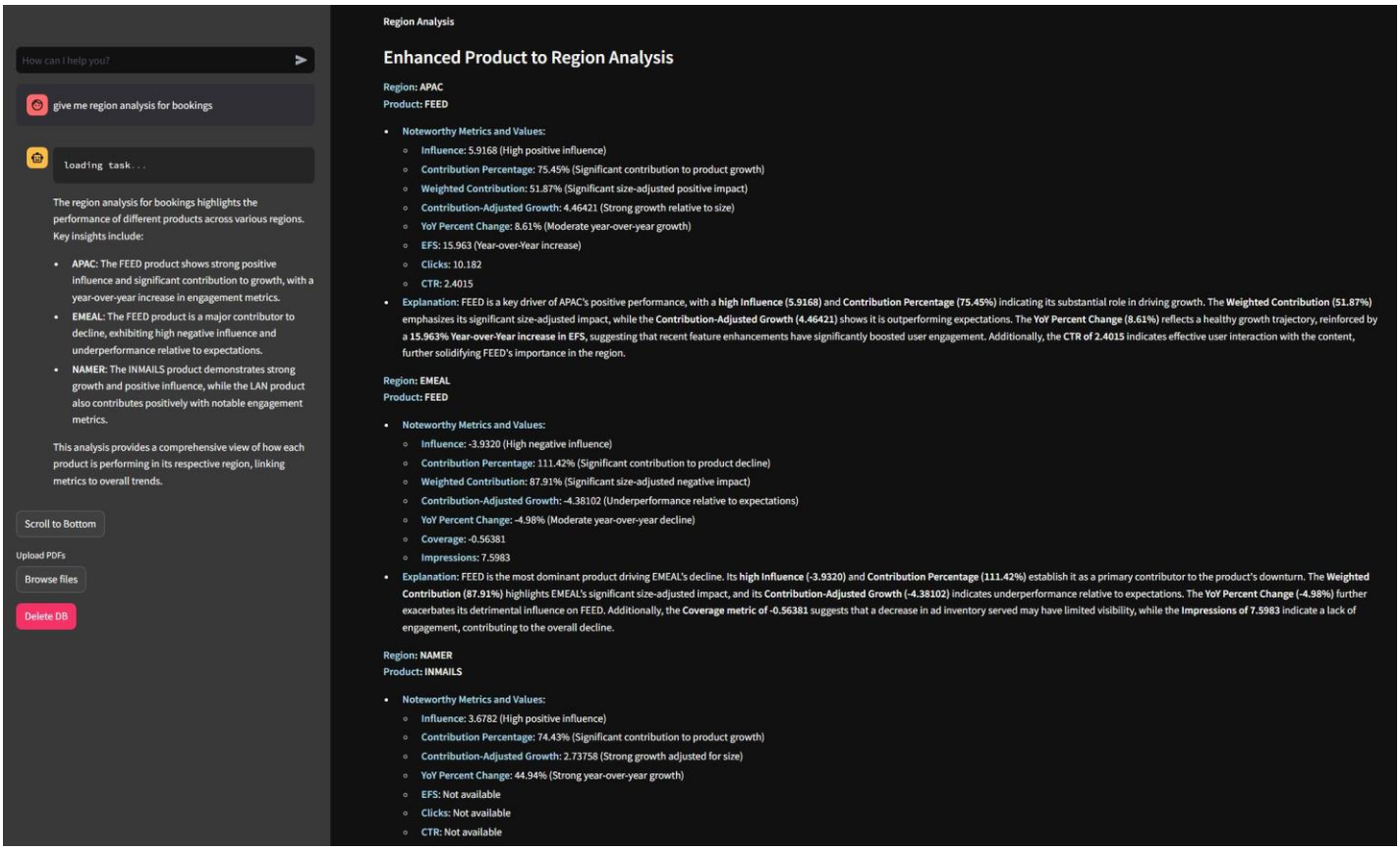
- Asking the agent to provide a comprehensive analysis:



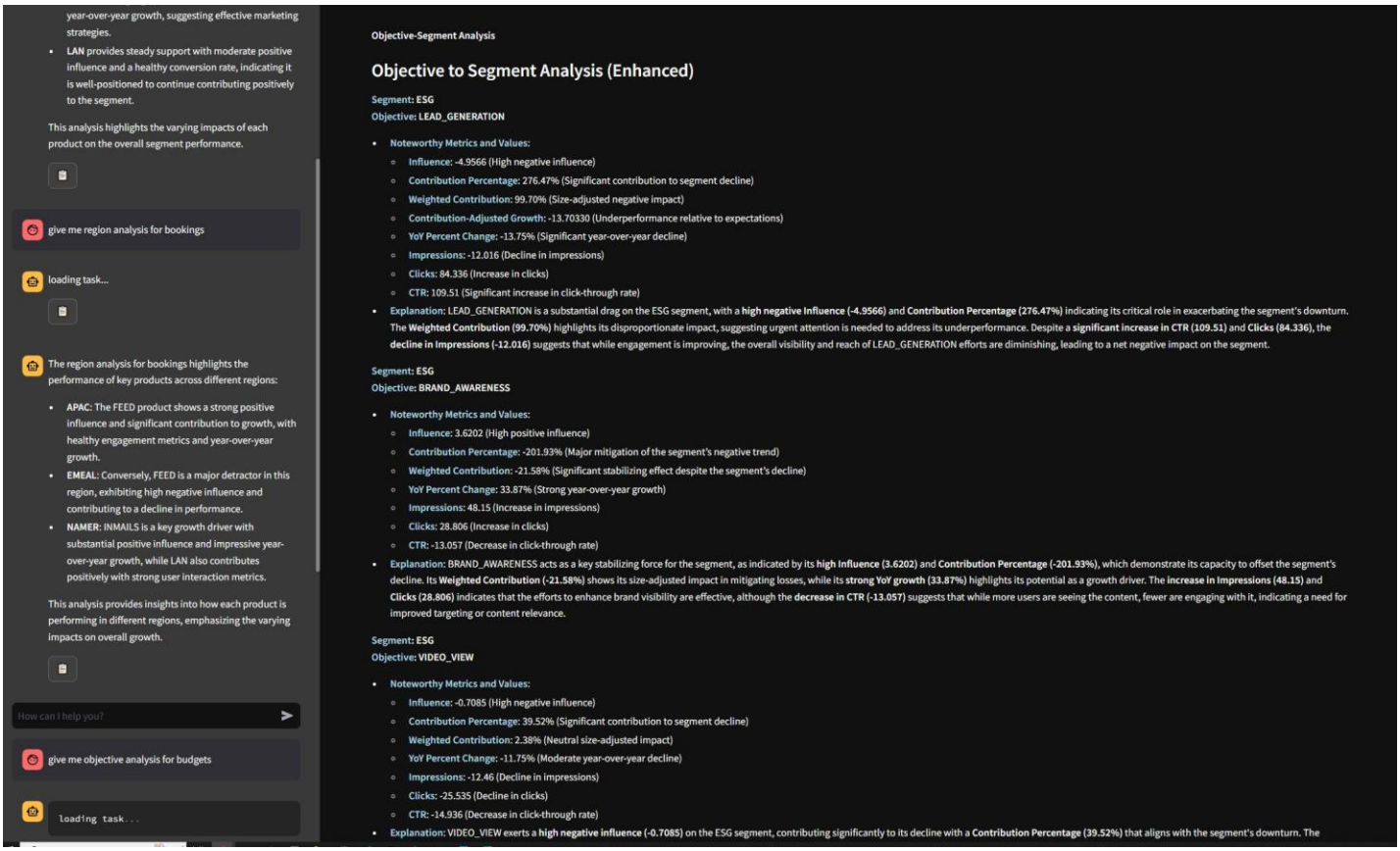
- Asking the agent to provide product insights:



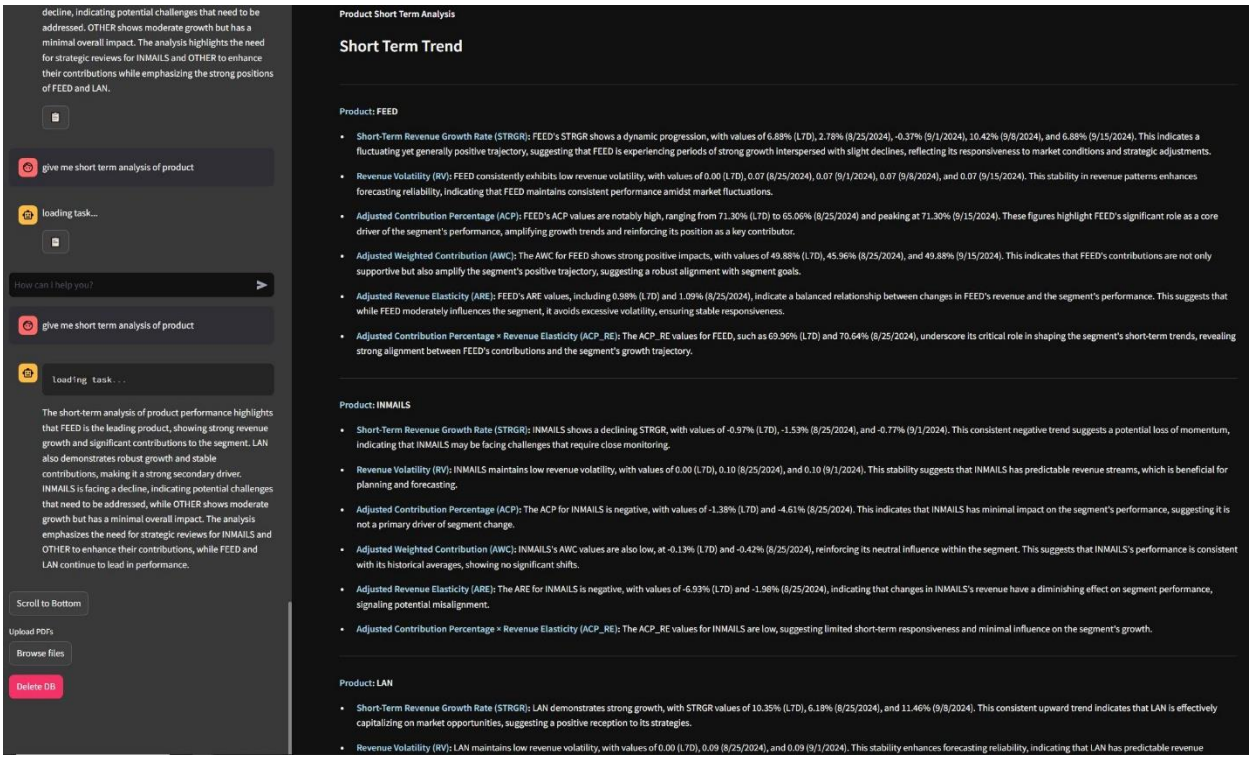
- Asking the agent for analysis of region:



- Asking the agent for Objective insights:



- Asking the agent for short-term trends



- Asking the app for product analysis only for region EMEAL.