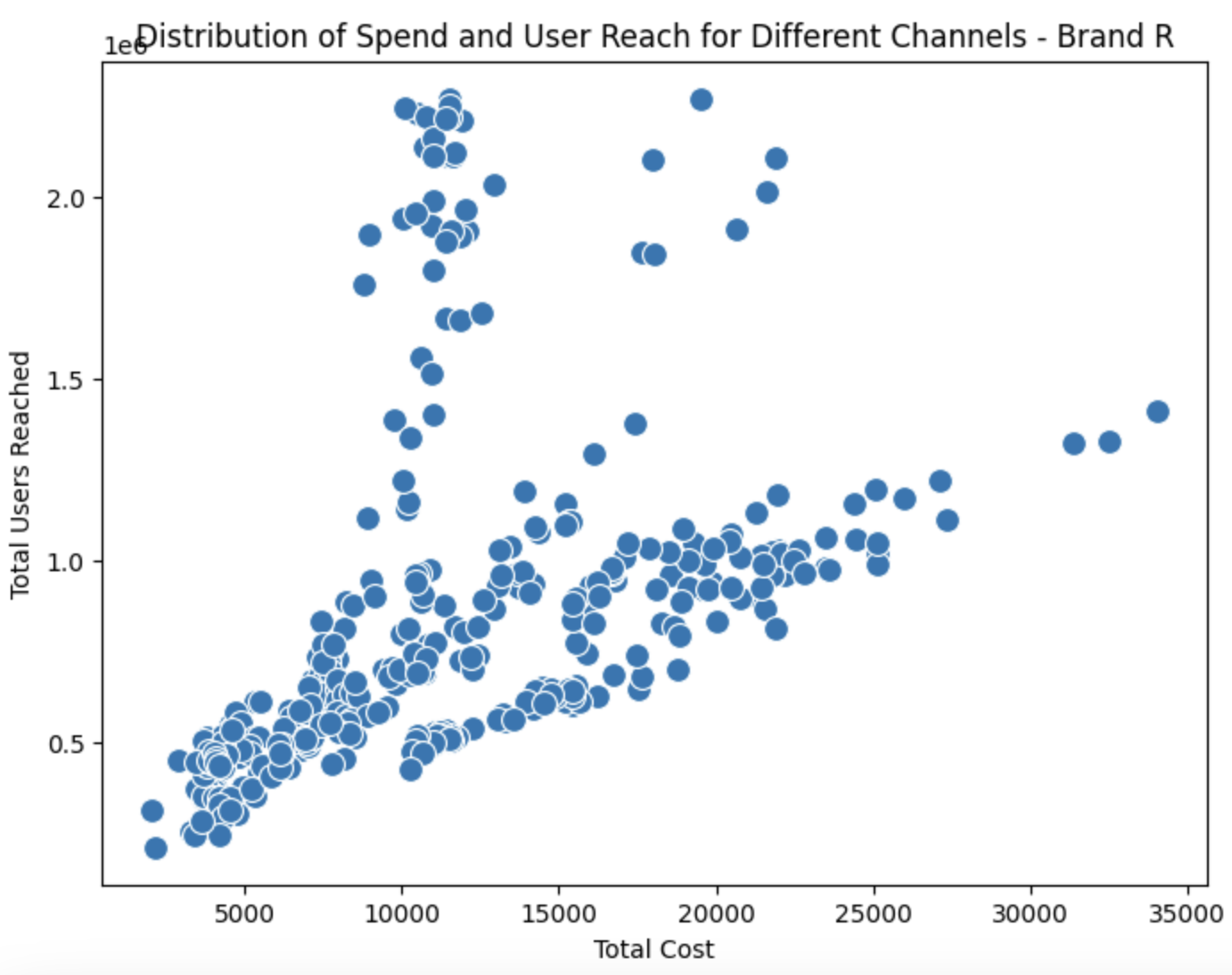
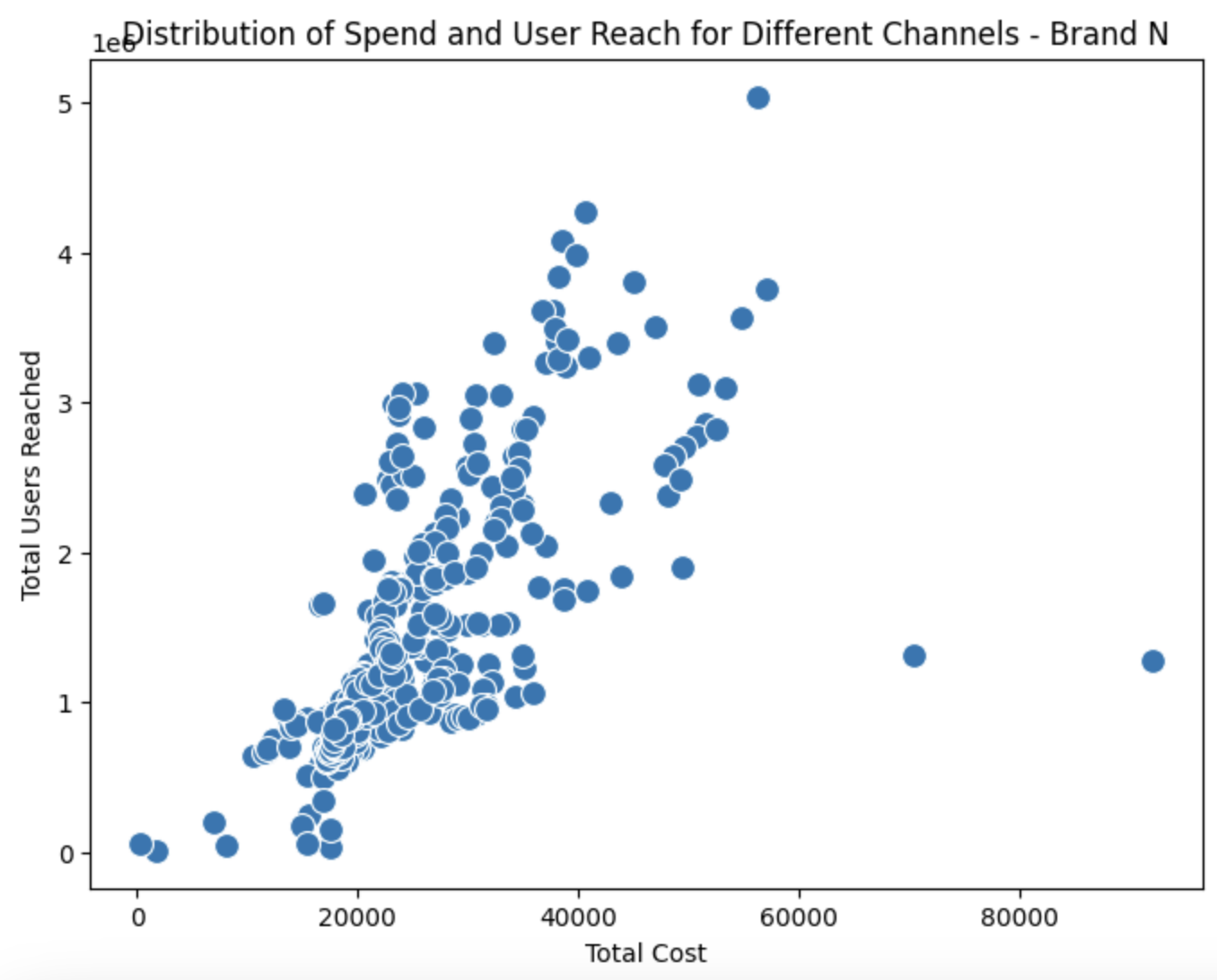
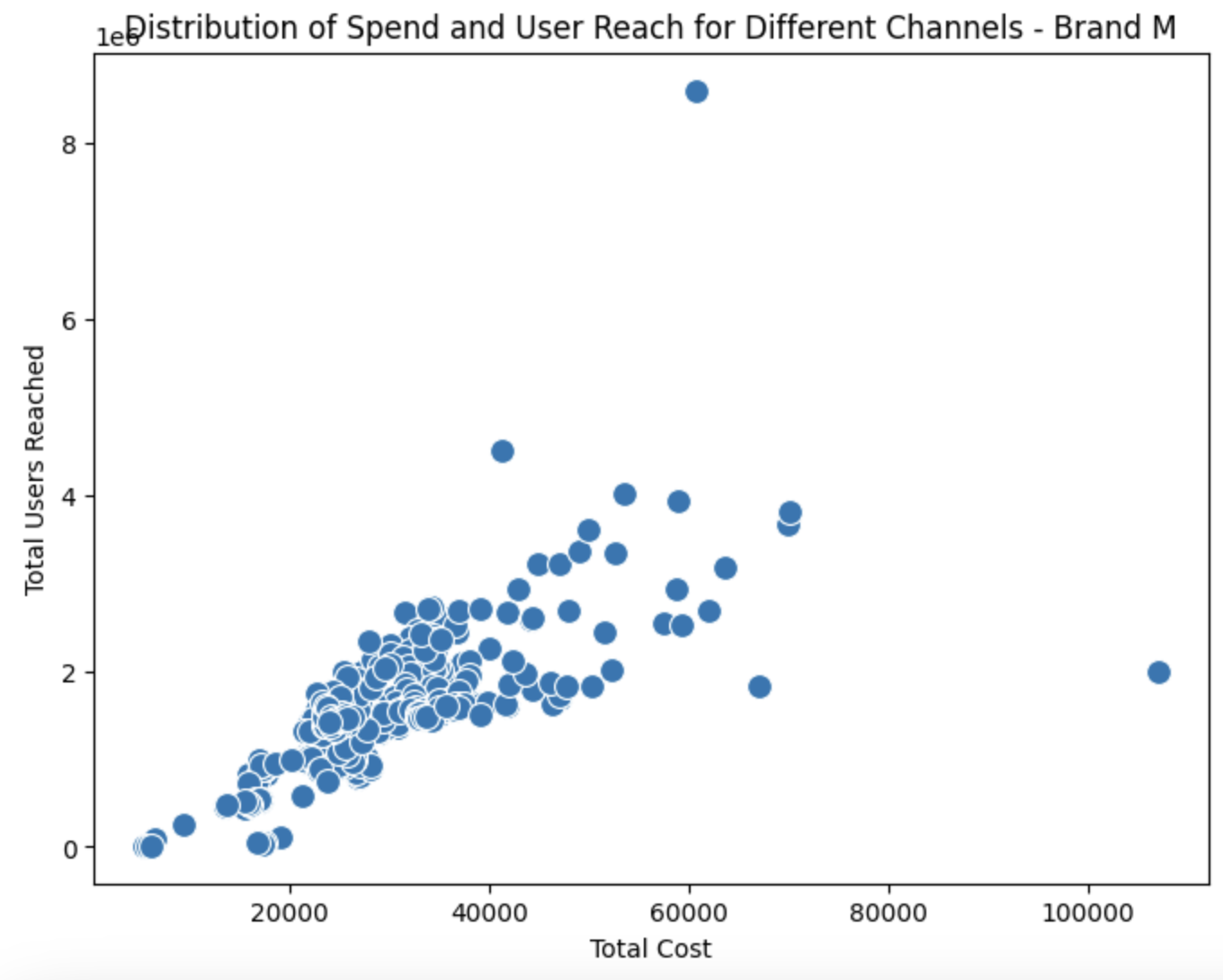
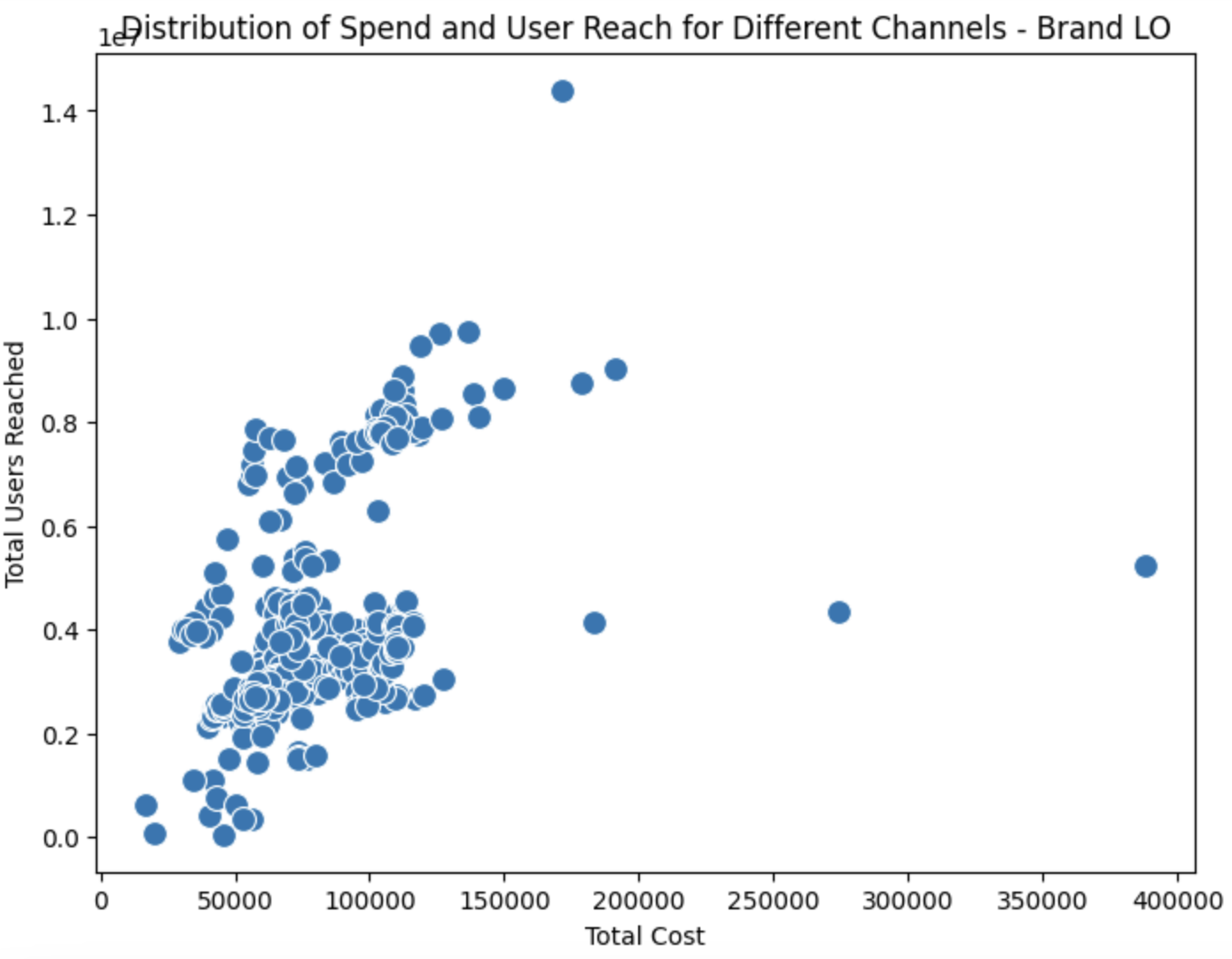
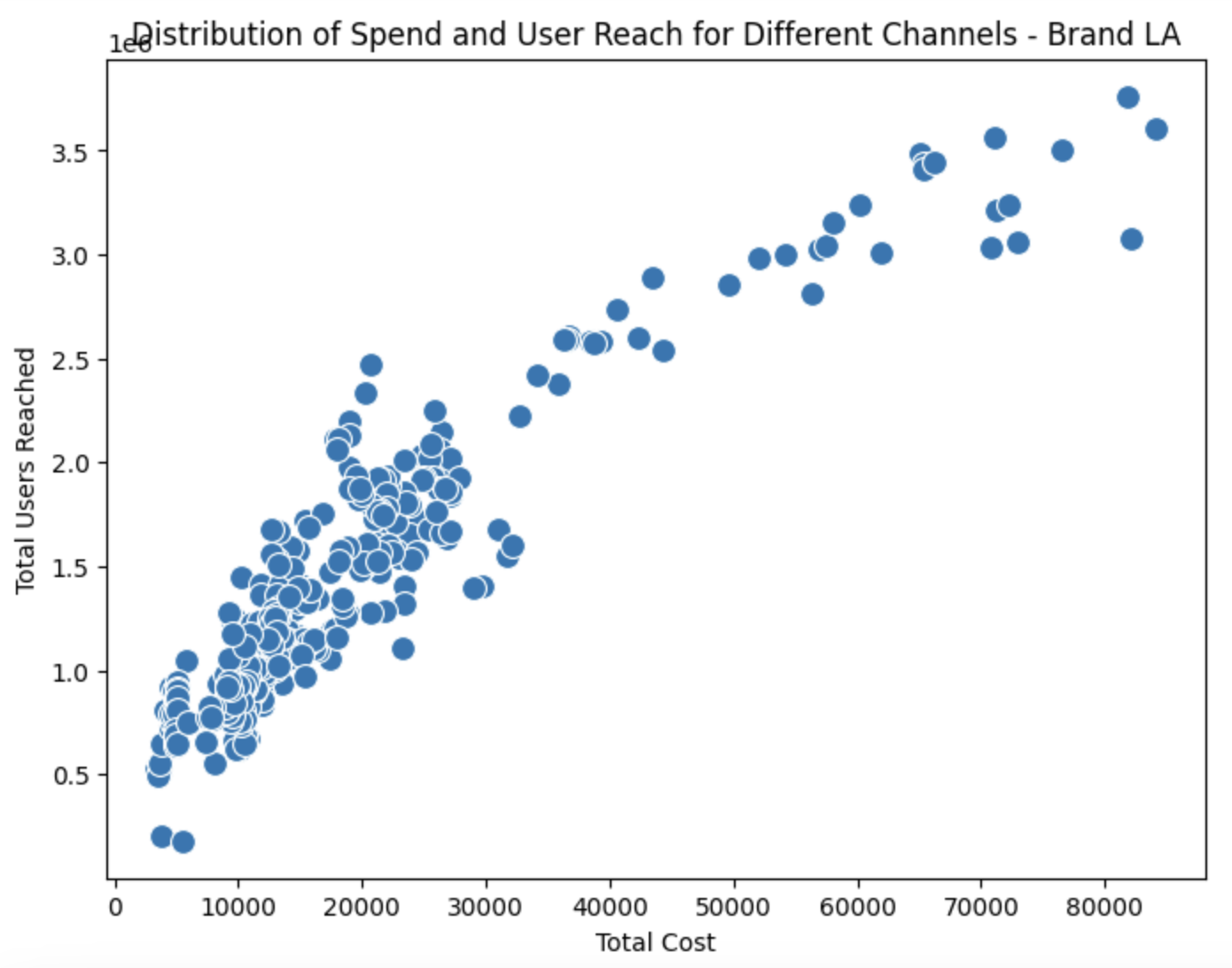
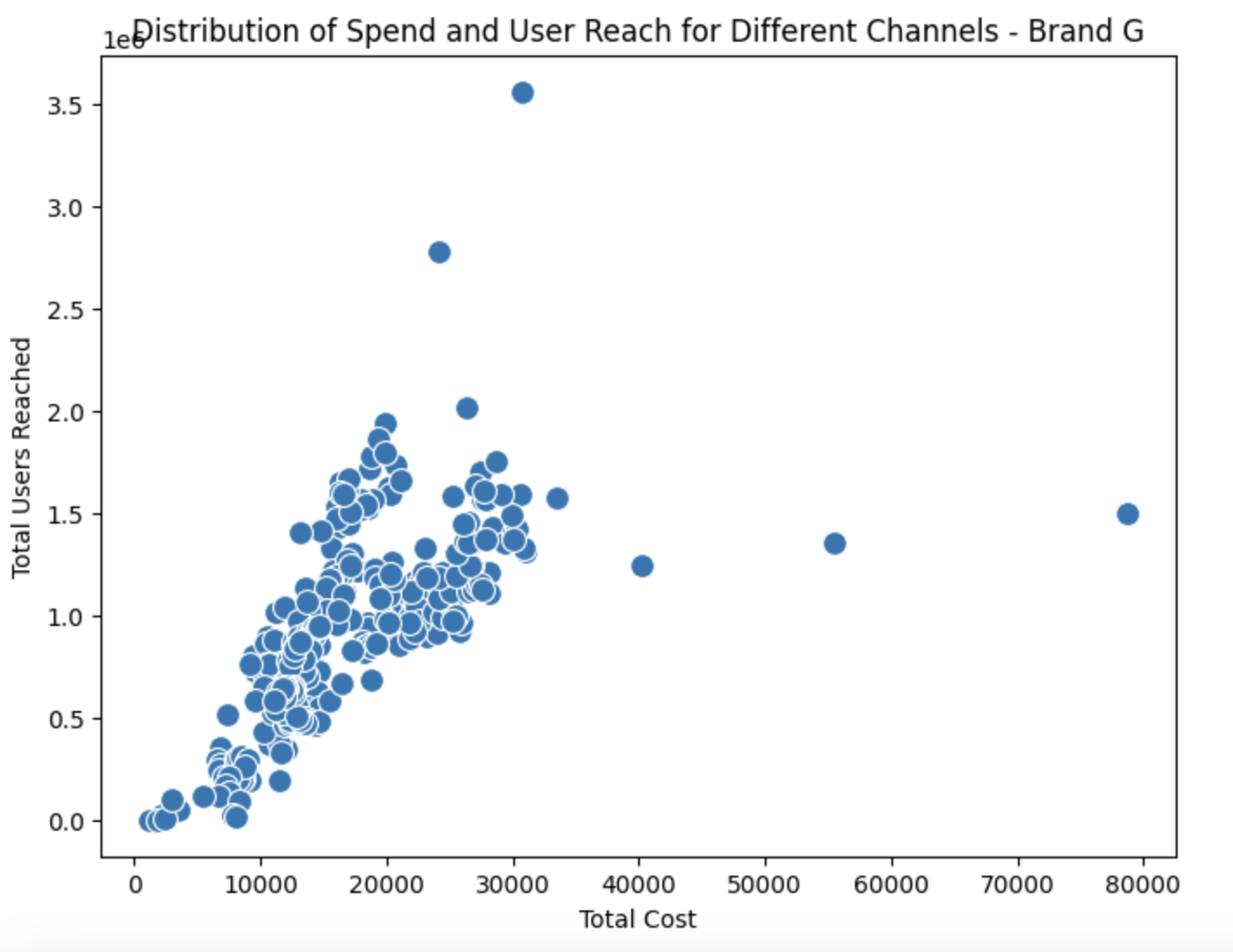
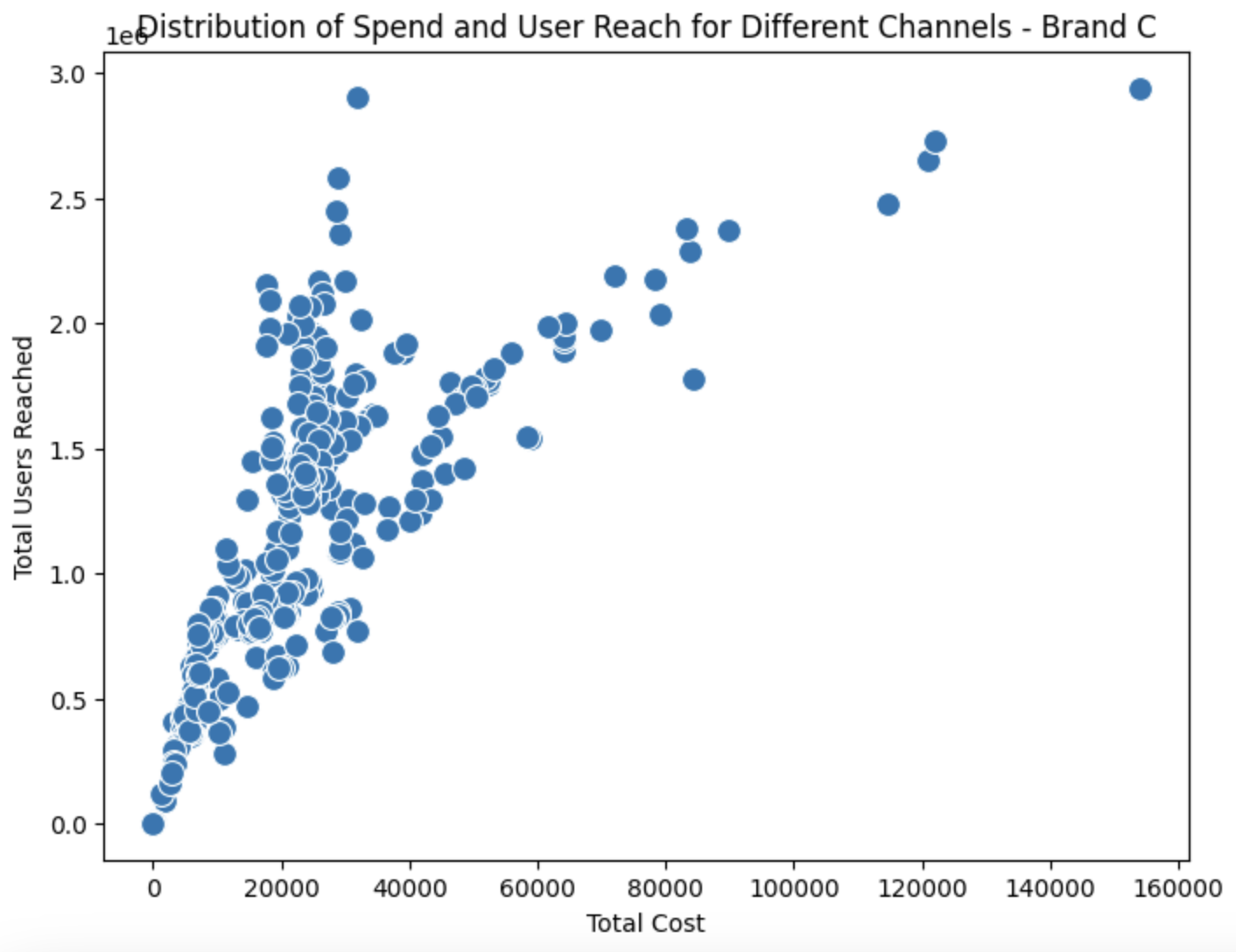
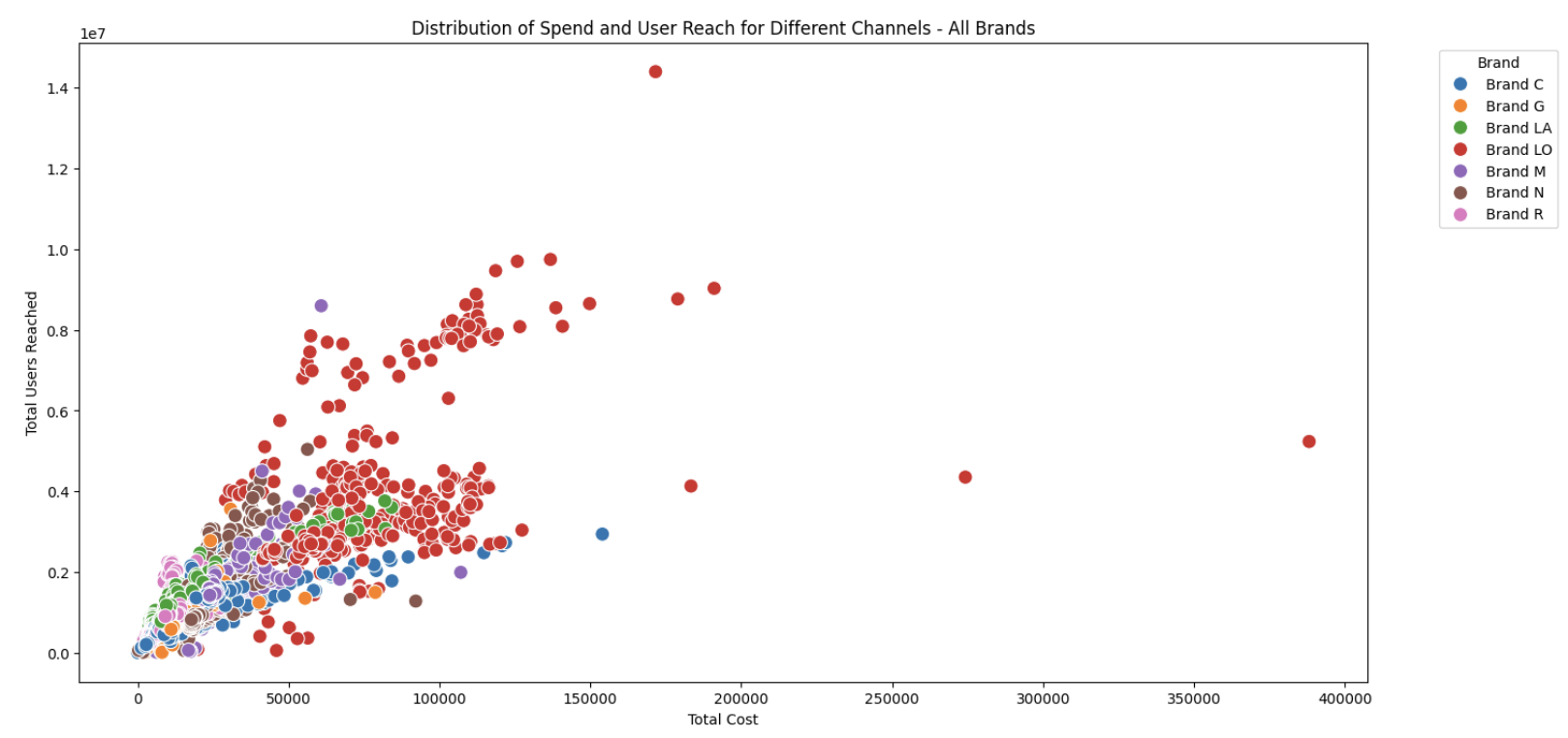
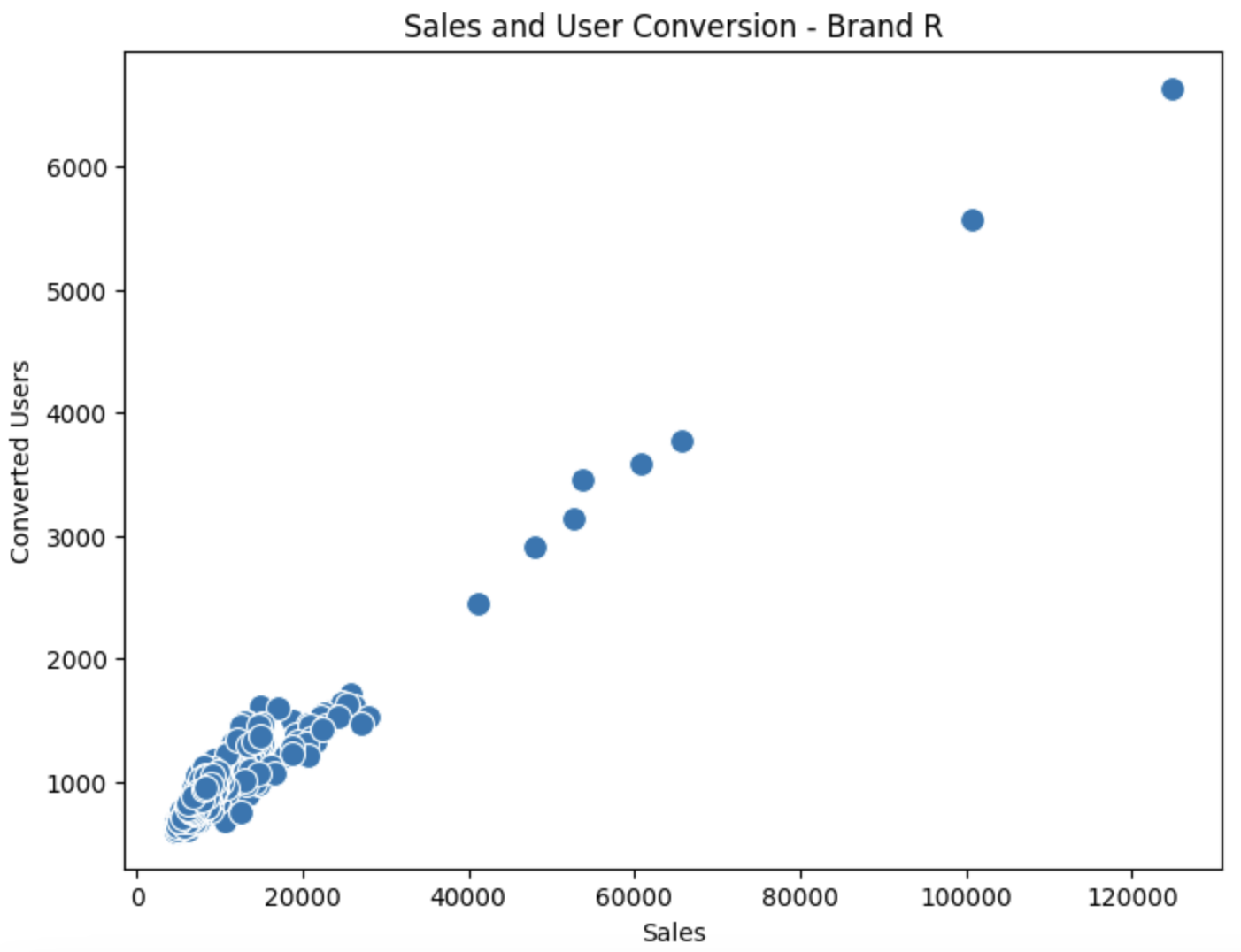
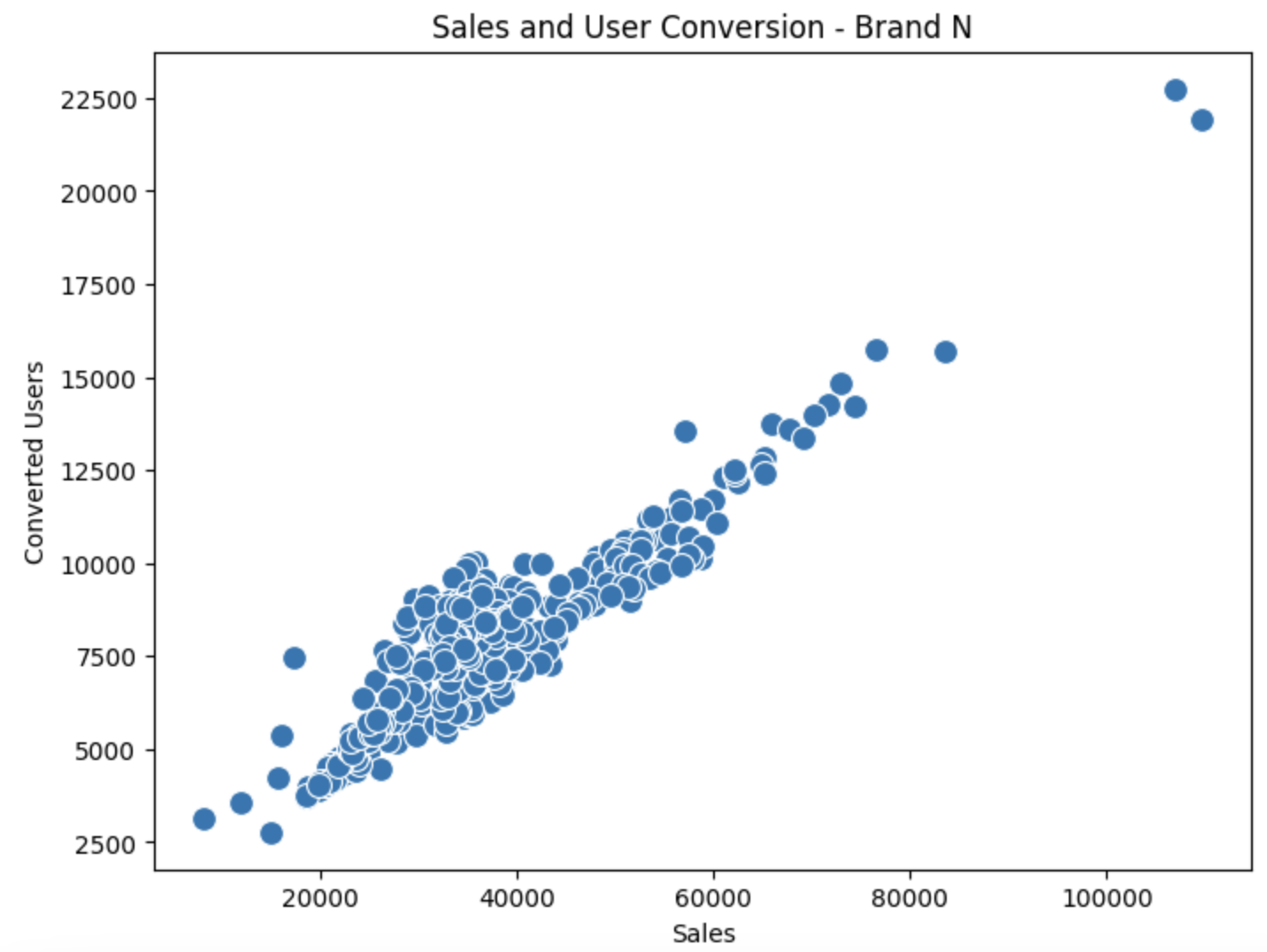
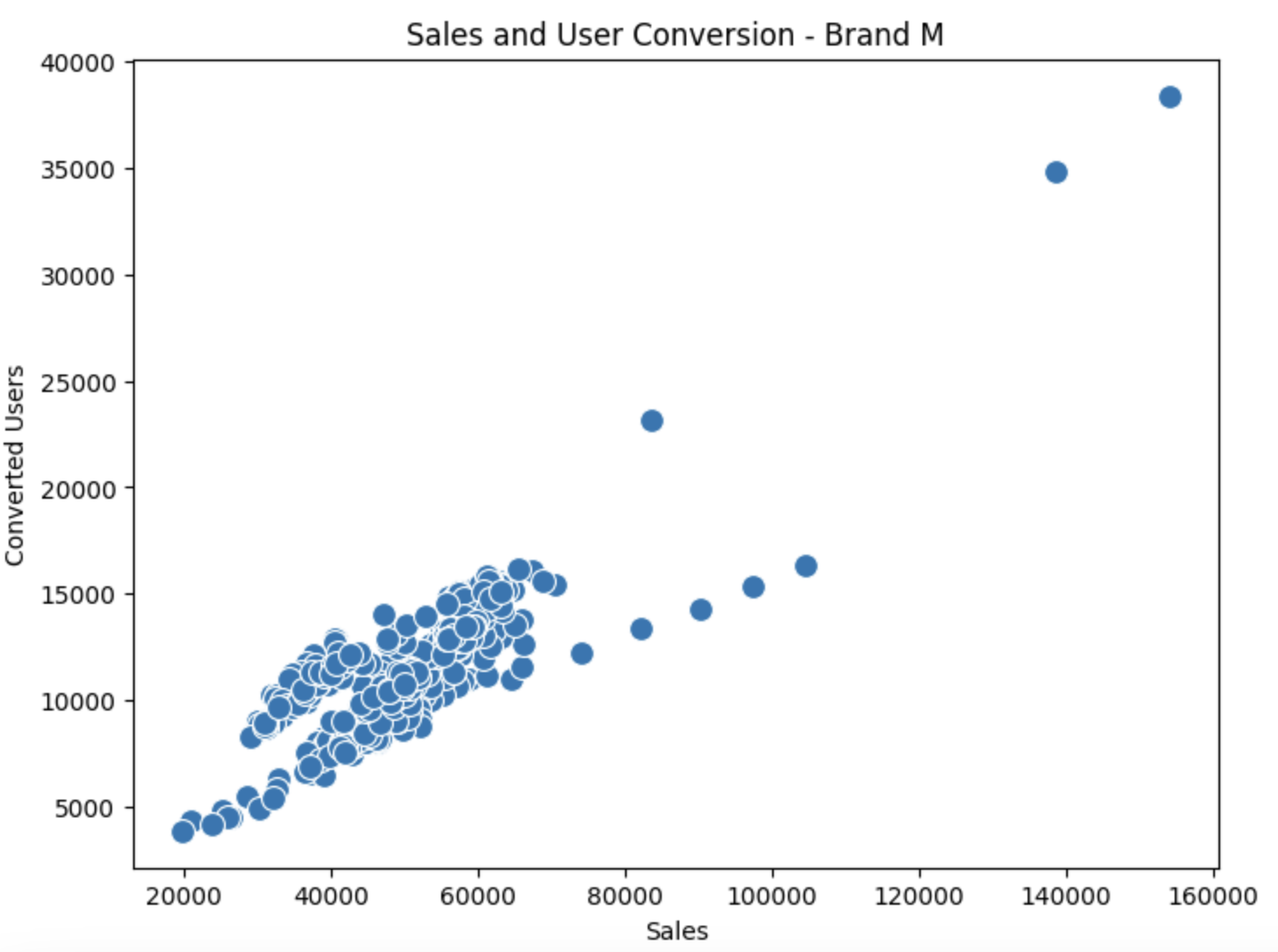
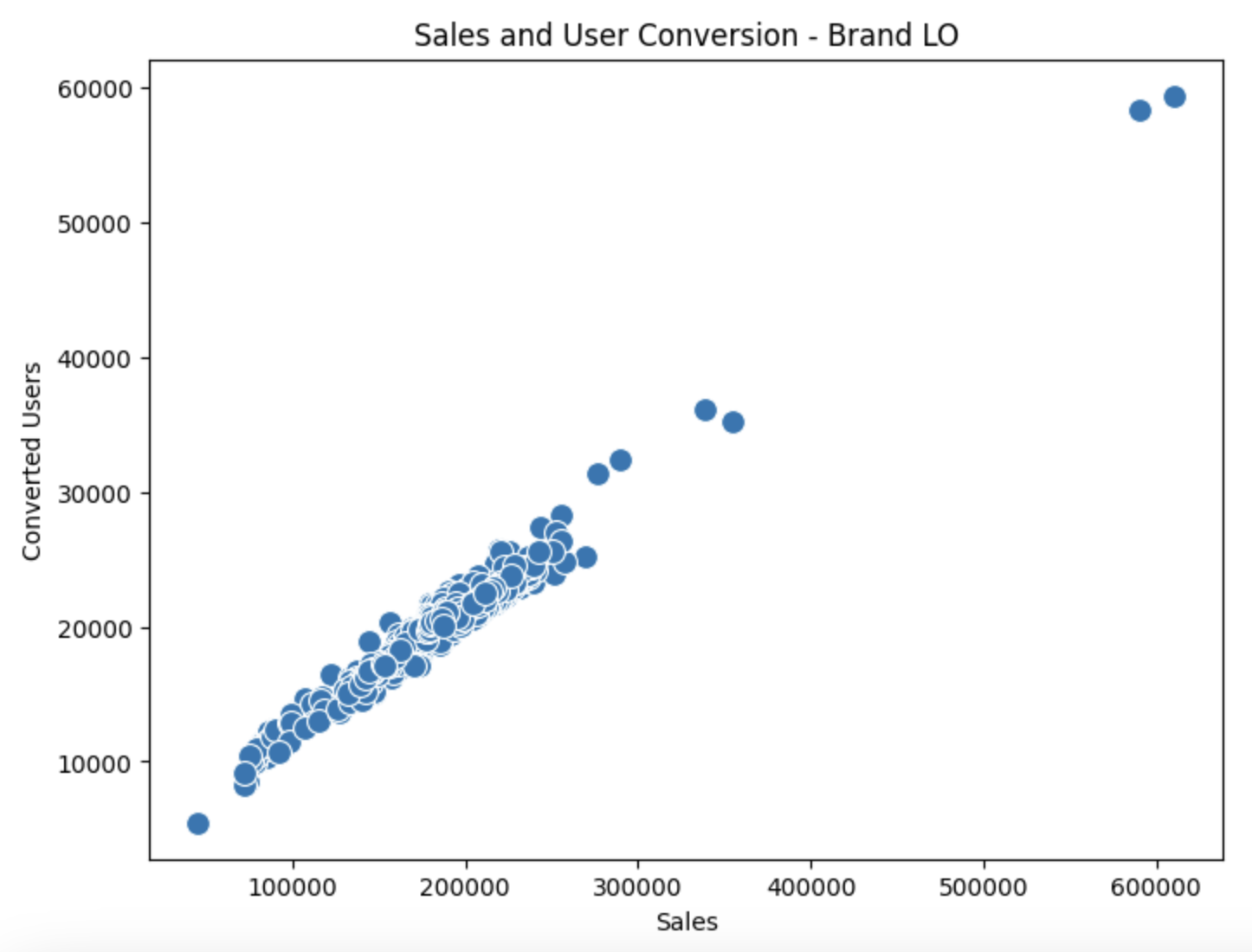
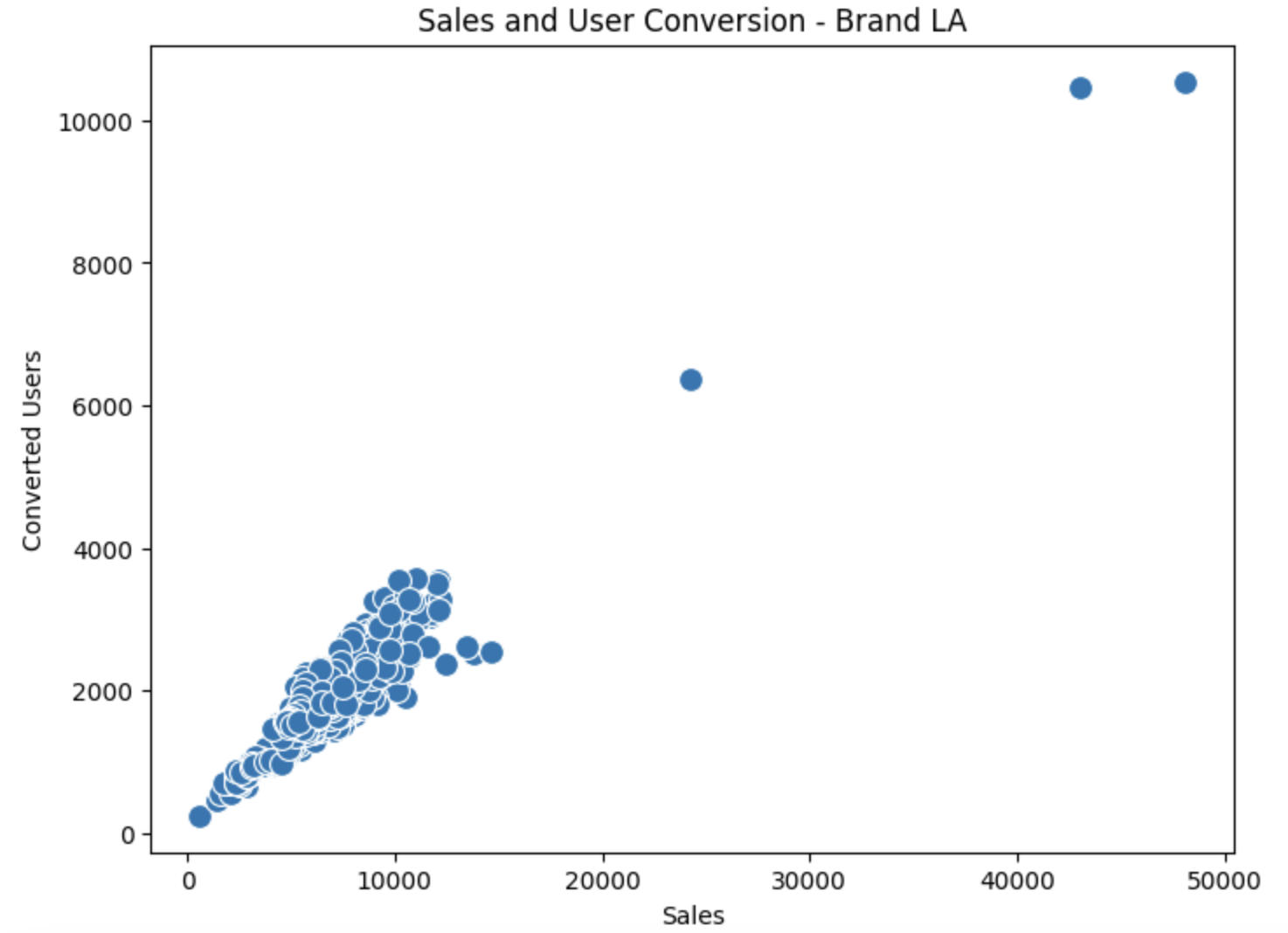
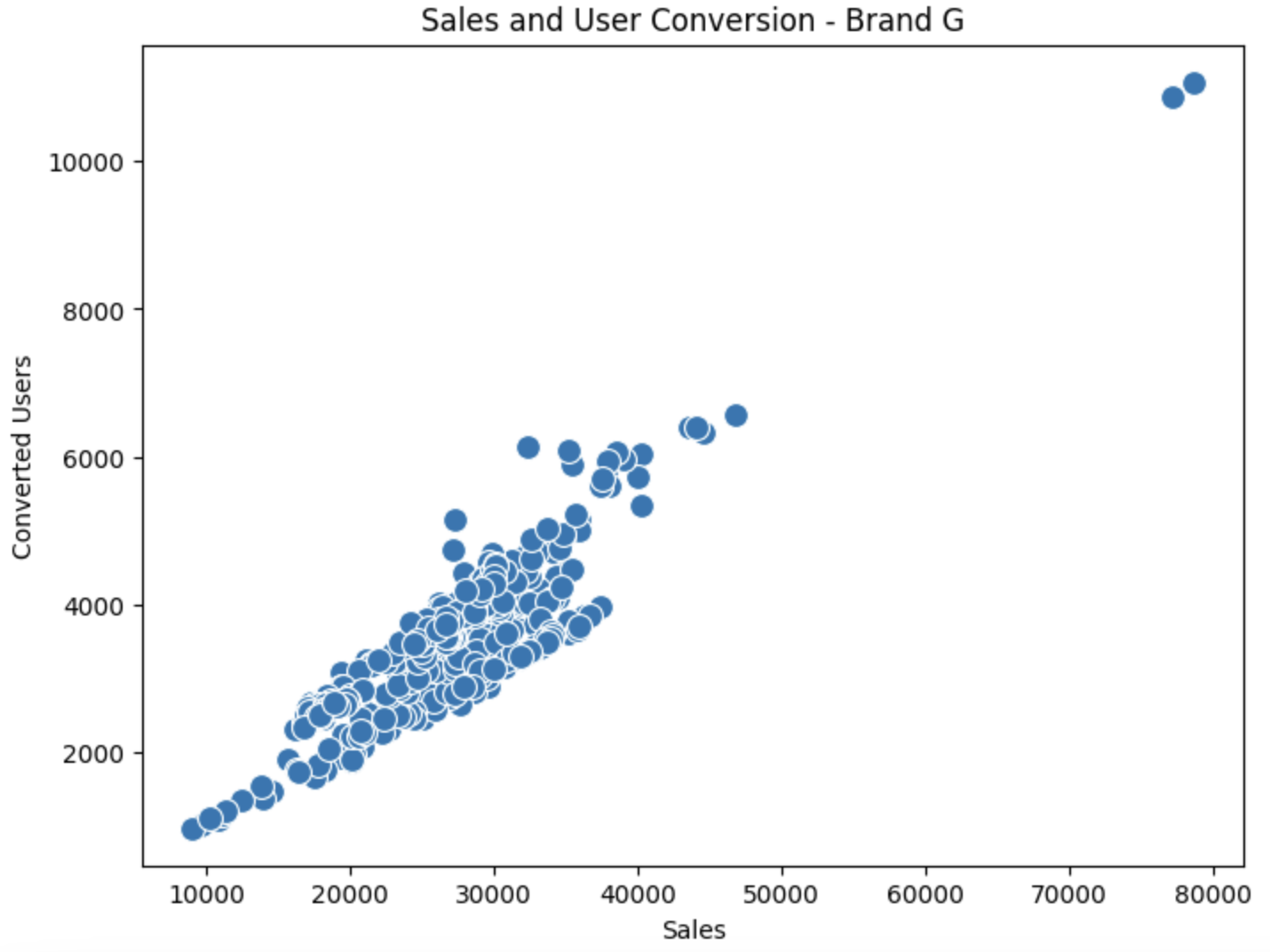
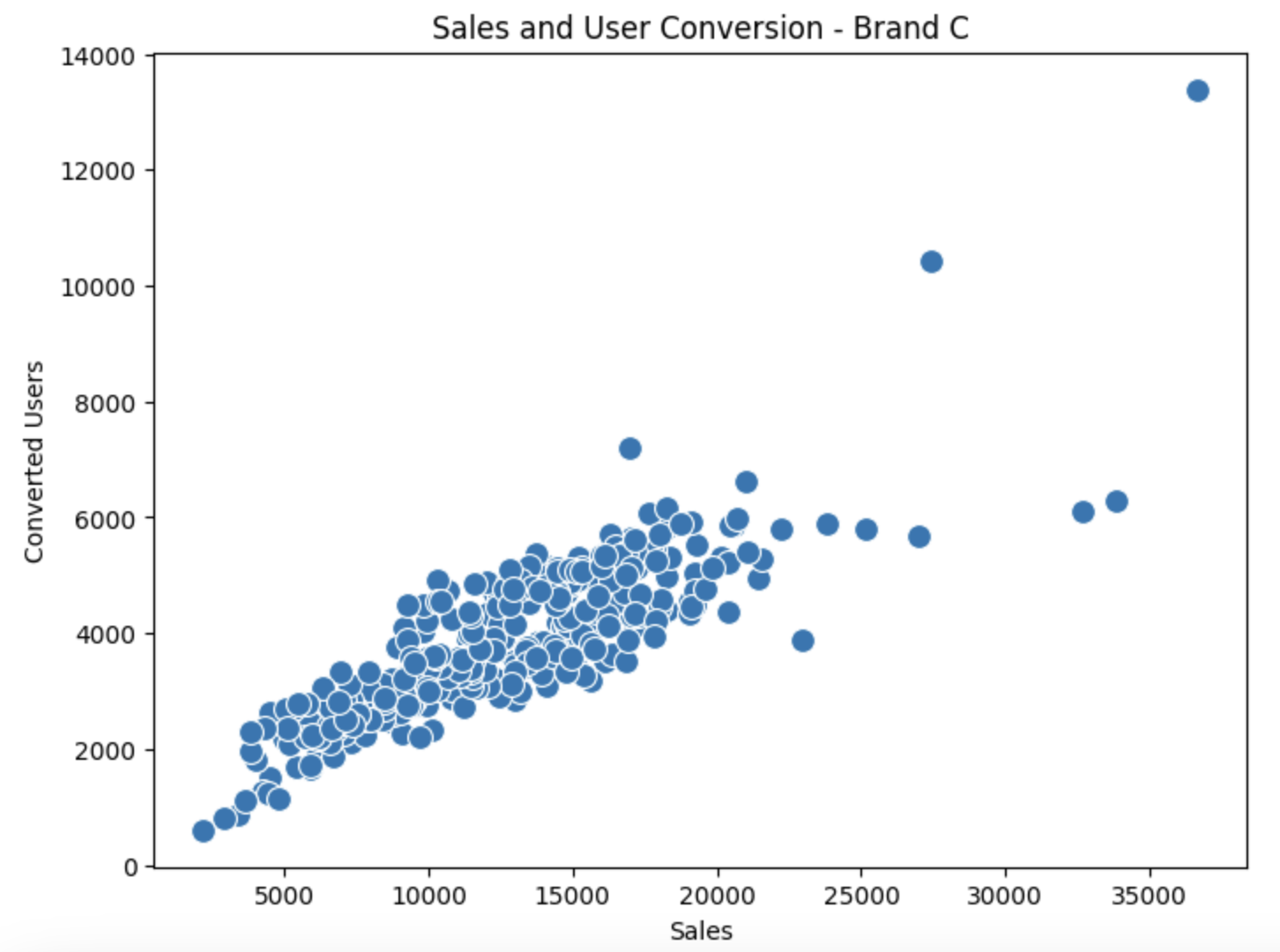
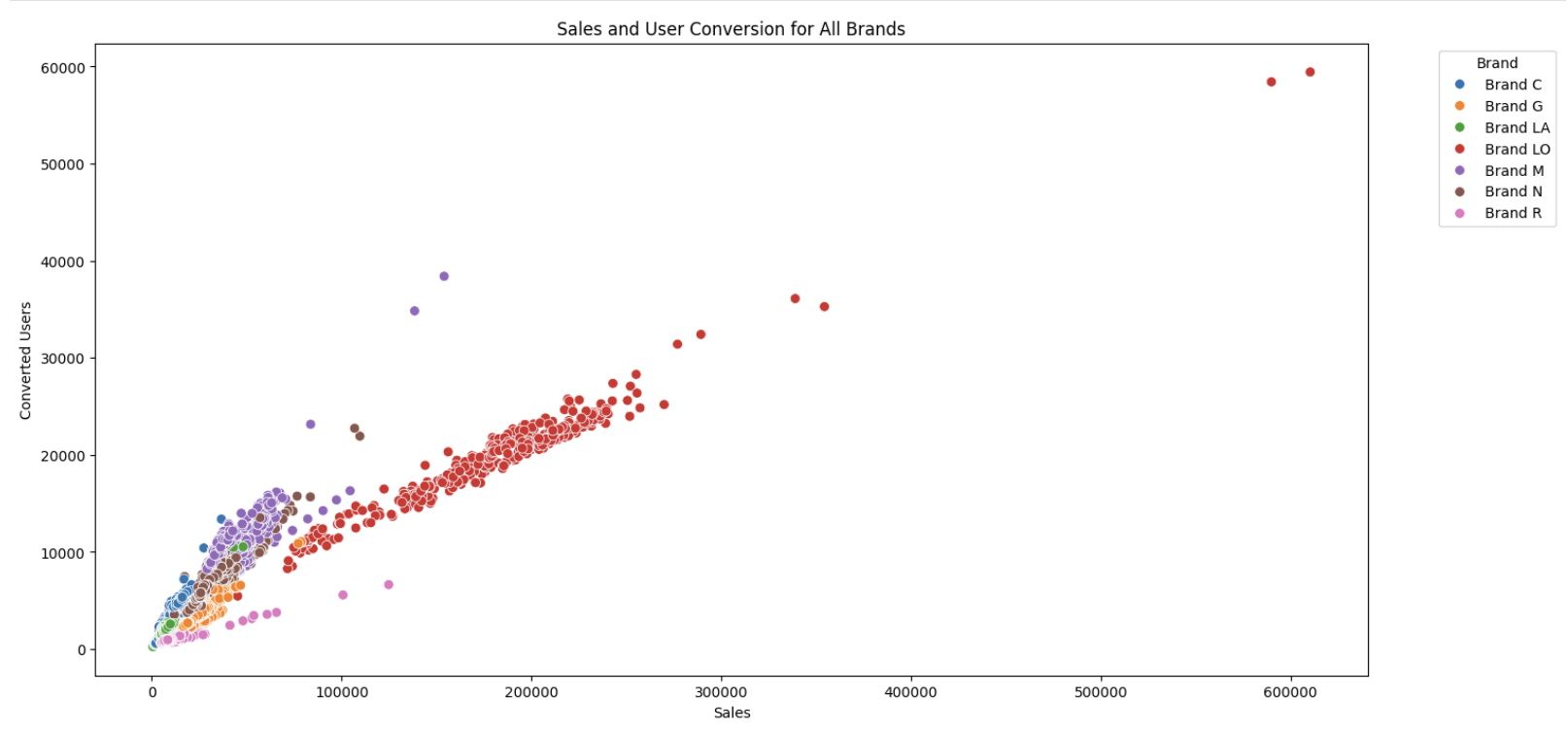
**Descriptive Statistics**

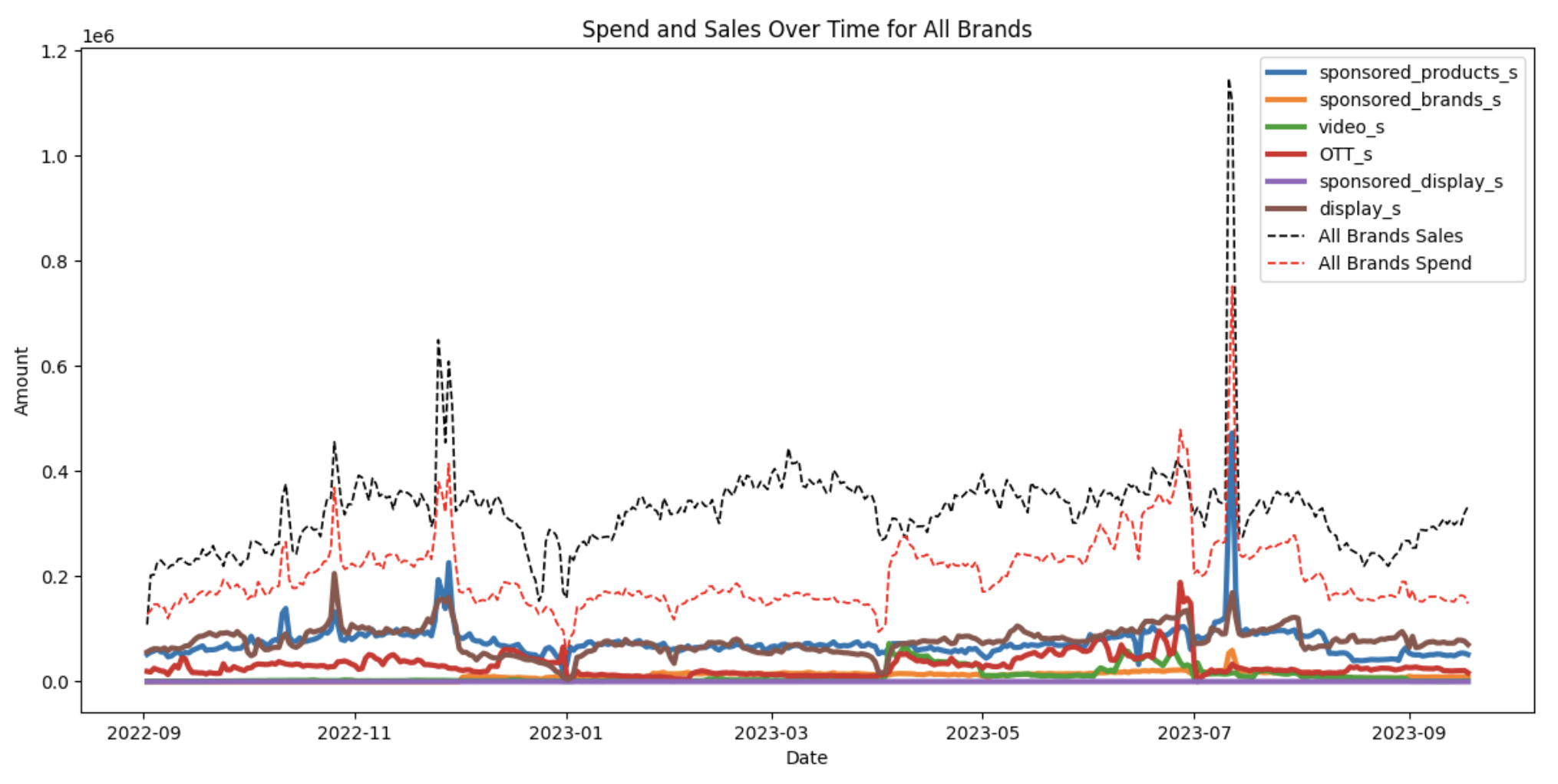
* Graph showing the distribution of spend and user reach of different channels for all brands and individual brands separately

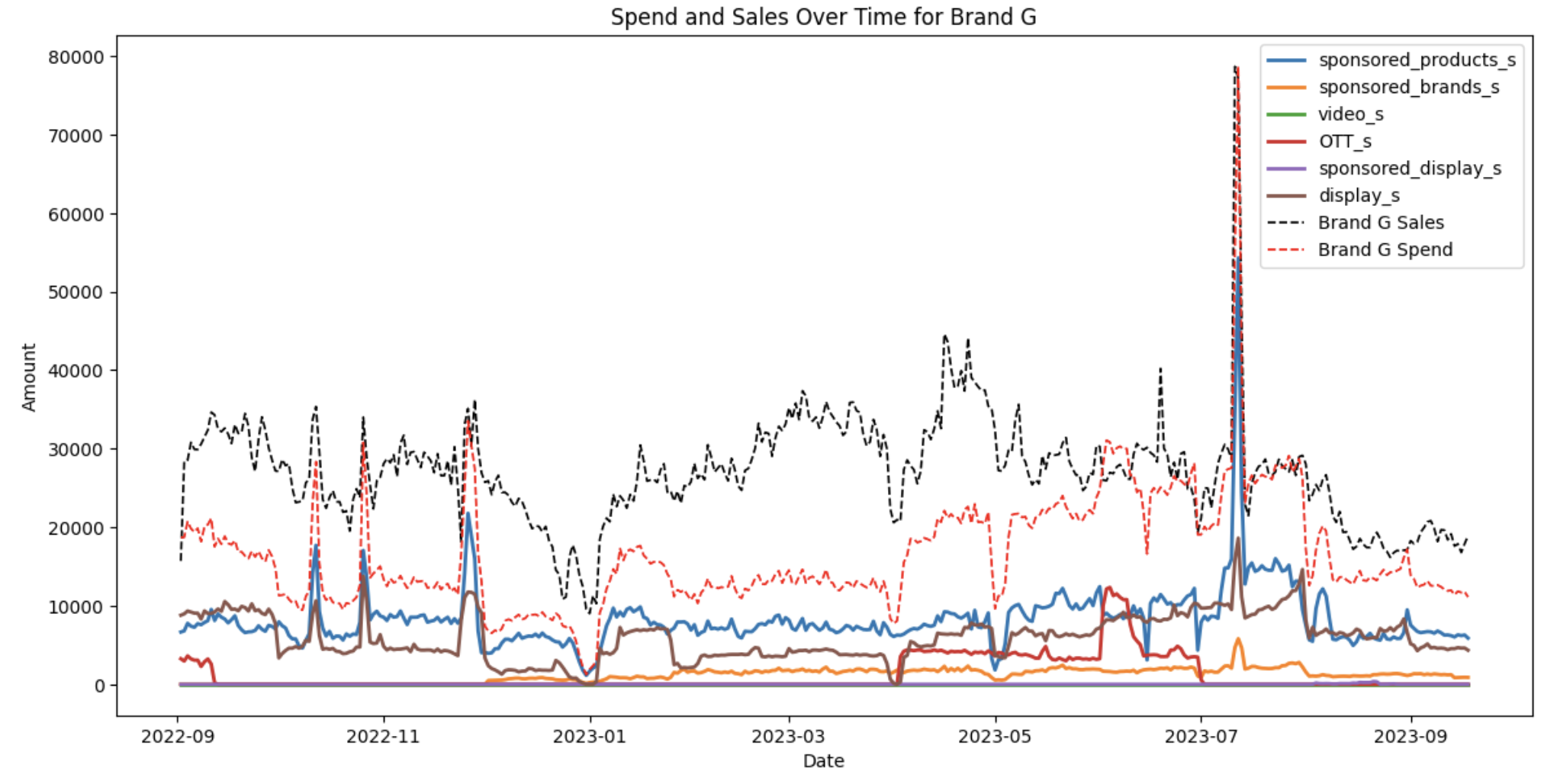
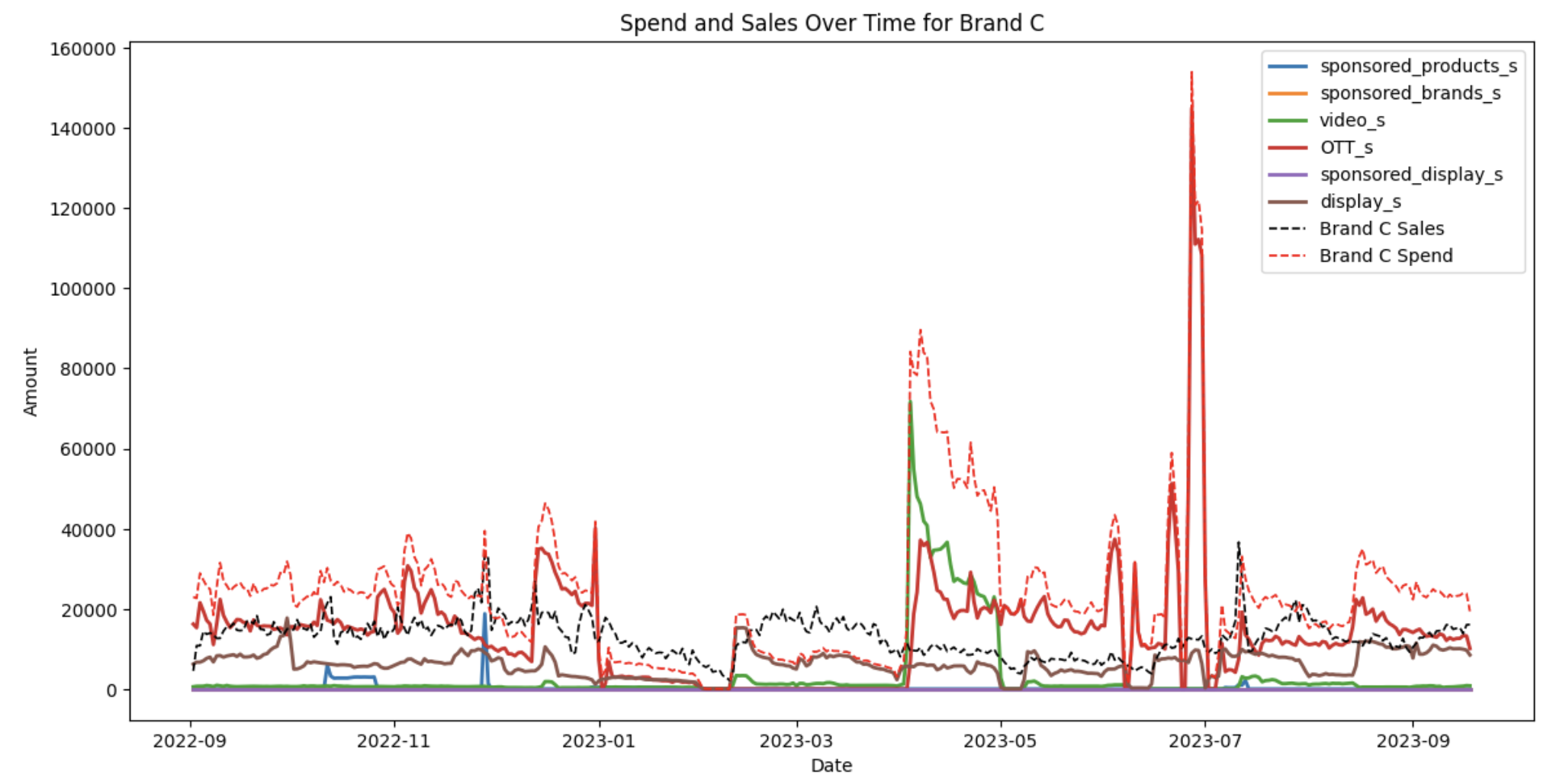


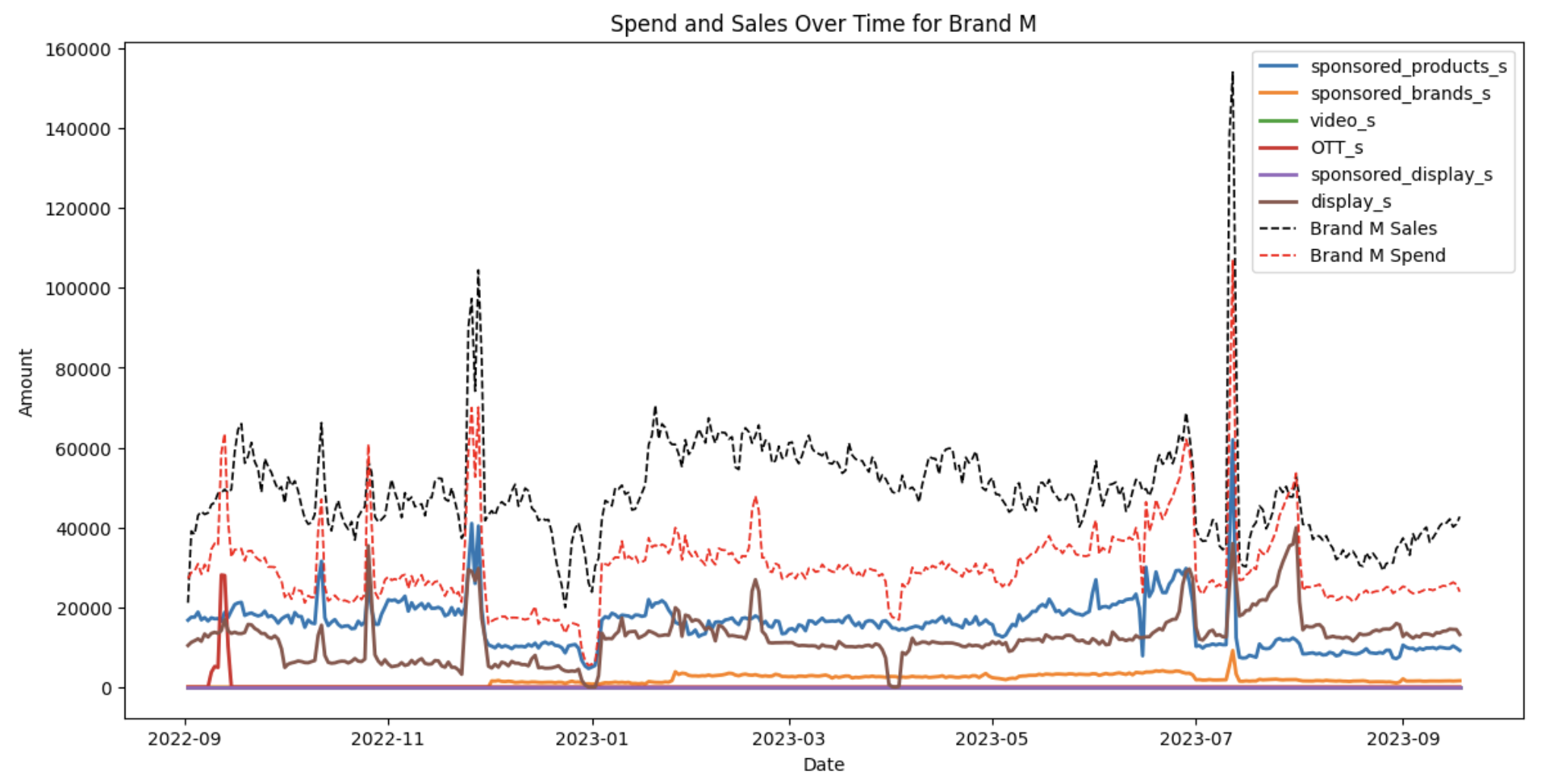
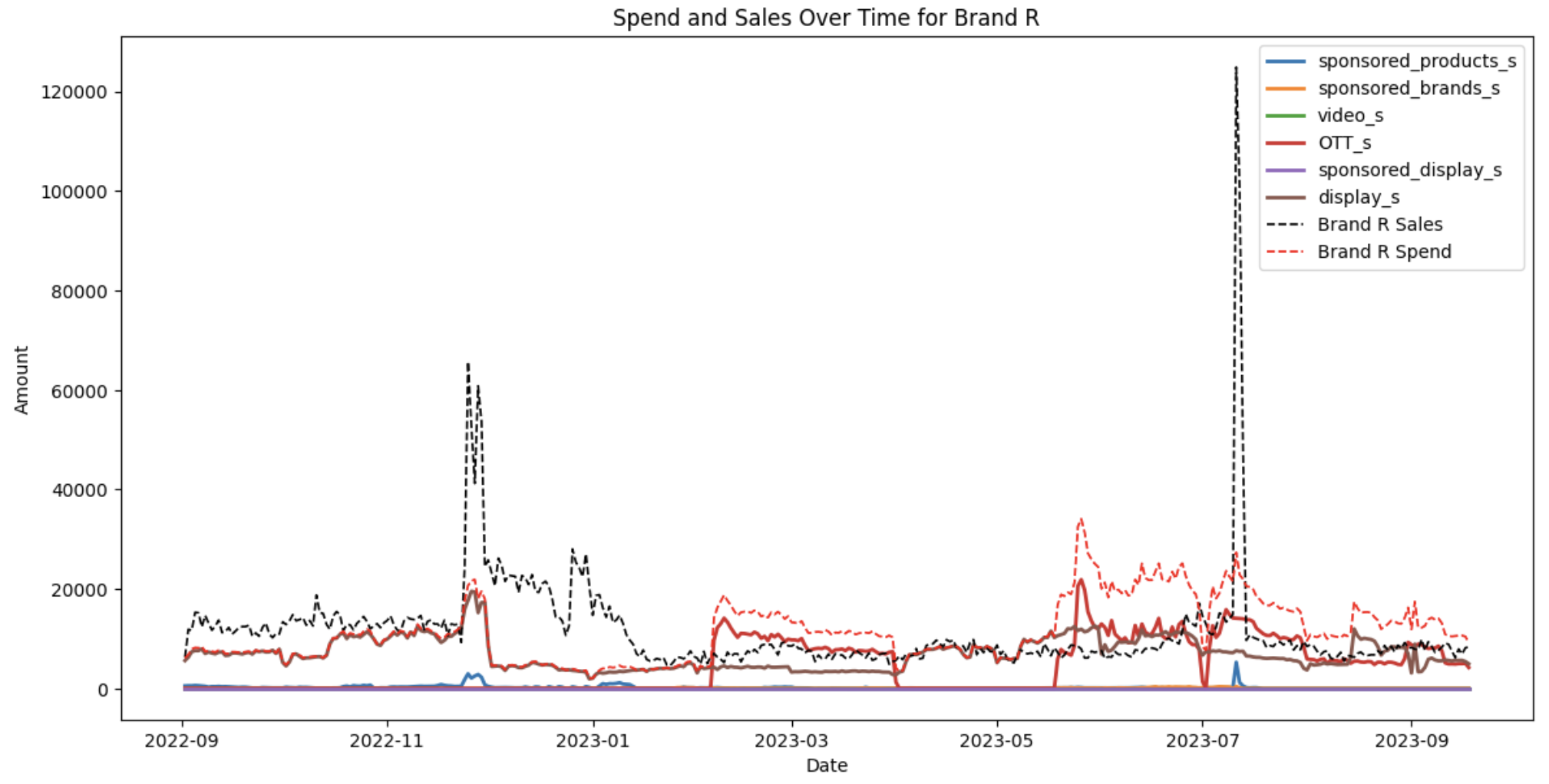
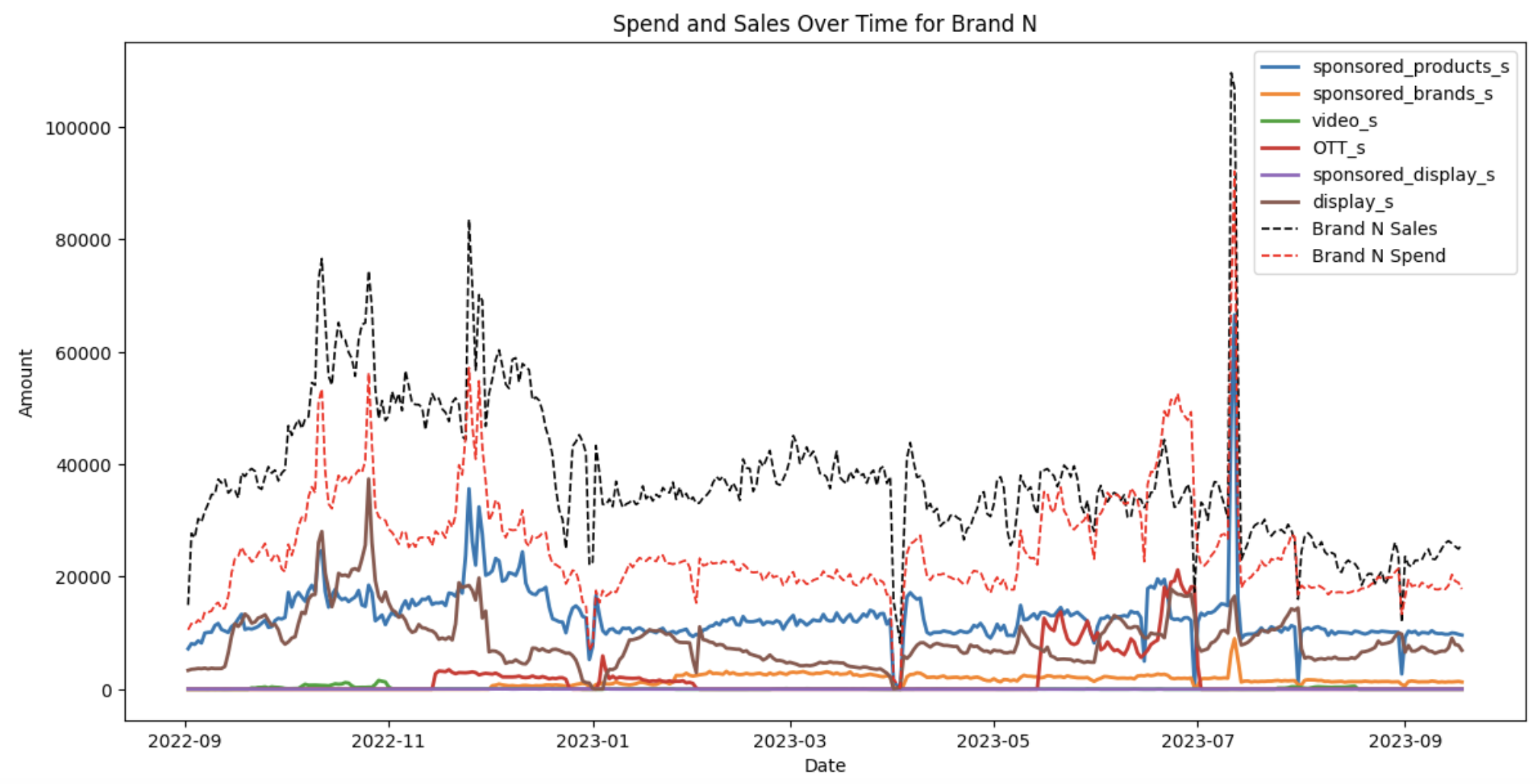
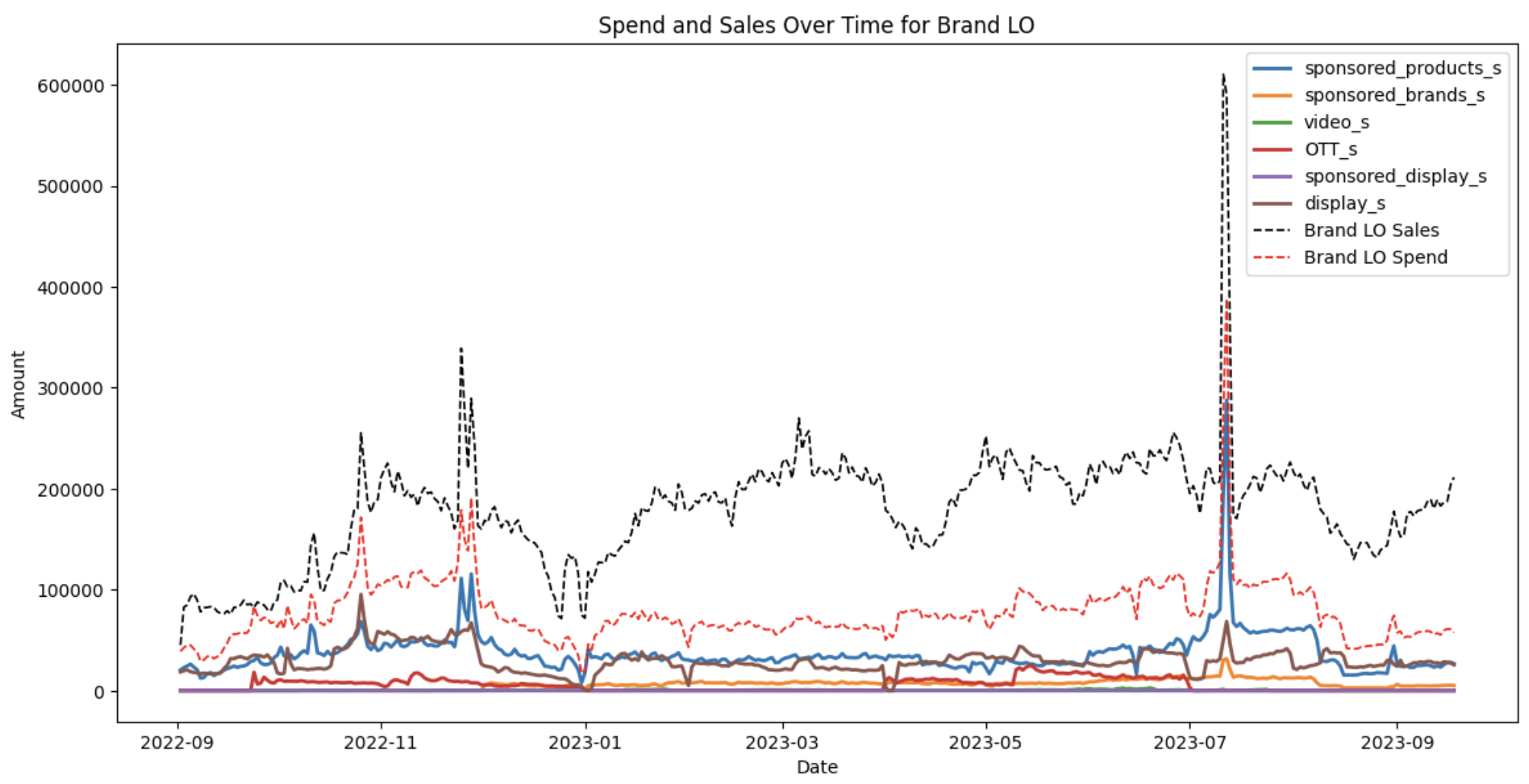
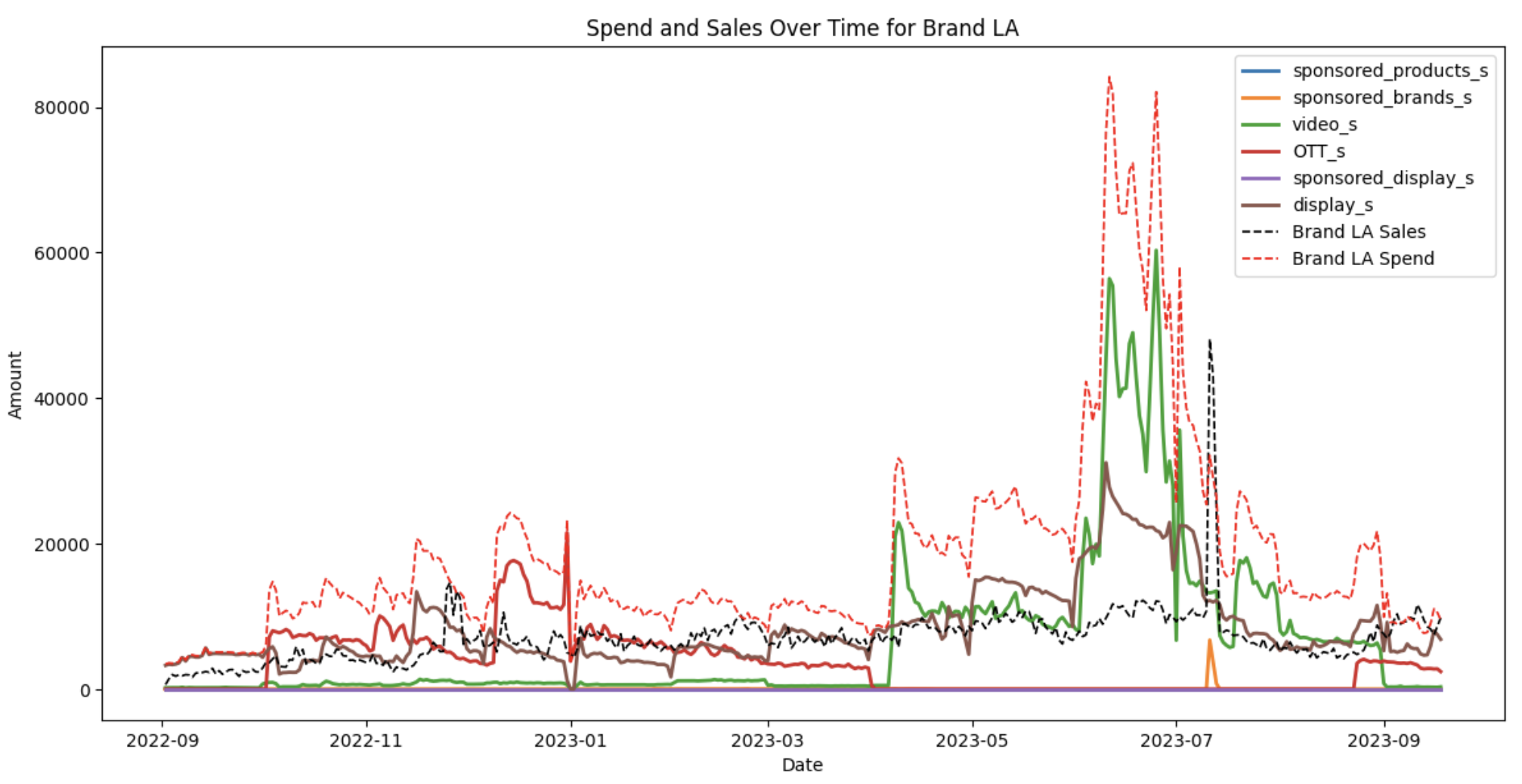
* Graph showing sales and user conversion for all brands and individual brands separately



* Graph showing the spend on each channel and sales over time for all brands combined and individual brands (Sales are much smaller than spending for each channel?)
  + Both the black and red dotted lines are now labeled “All Brands Sales” (please amend) - looking at your codes I think black is sales and red is spend? In that case, sales is still greater than spend? For Brands LA and Brand C, spend is greater than sales.
  + Also, it is confusing to have two y axes. Can we only have one so that spend and sales are shown on the same scale? If total spend if smaller than total sales, then the spending on each channel should not be bigger than total sales. In fact, if I am not mistaken, I think the labels for the right and left y axes are flipped, but anyway, lets just keep one.Have this fixed! ❤️
  + I think we can also remove all the xxxx\_u (which represents the user reach of each channel, just for a cleaner graph Done
  + Same comments apply to individual brands Done







Summary fraction of spend

|  | **All** | **C** | **G** | **LA** | **LO** | **M** | **N** | **R** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Sponsored products** | 36.84% | 0.91% | 50.17% | 0.00% | 47.53% | 52.85% | 25.01% | 1.86% |
| **Sponsored brands** | 5.1% | 0.00% | 6.87% | 0.19% | 7.68% | 5.76% | 5.59% | 0.31% |
| **video** | 4.93% | 12.92% | 0.00% | 35.44% | 0.51% | 0.00% | 0.44% | 0.00% |
| **OTT** | 14.85% | 60.16% | 6.96% | 18.55% | 6.66% | 0.70% | 7.03% | 37.29% |
| **Sponsored display** | 0.01% | 0.00% | 0.05% | 0.00% | 0.00% | 0.00% | 0.04% | 0.00% |
| **display** | 38.28% | 26.01% | 35.95% | 45.82% | 37.62% | 40.69% | 34.90% | 60.55% |
| **ROI** | 60.67% | -45.19% | 63.99% | -61.43% | 130.44% | 64.28% | 51.36% | 1.22% |

* Calculate simple ROI over the entire period (using total spend) for all brands and individual brands separately

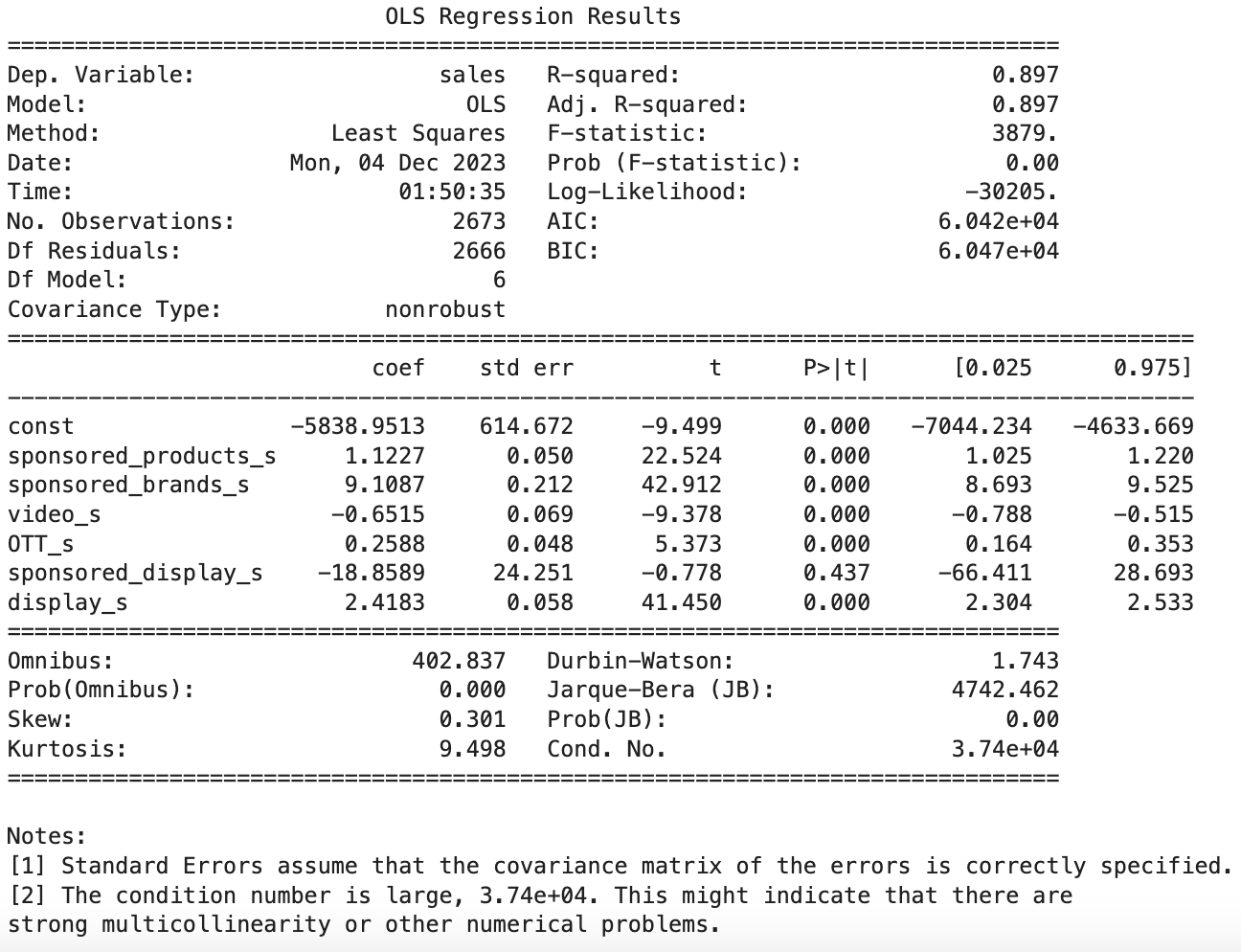
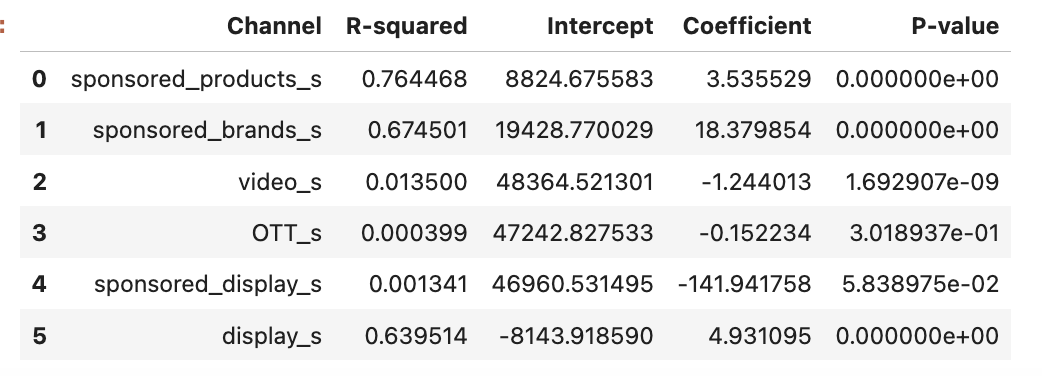
| Overall ROI | | 60.67% |
| --- | --- | --- |
| Individual Brand ROIs | Brand C | -45.19% |
| Brand G | 63.99% |
| Brand LA | -61.43% |
| Brand LO | 130.44% |
| Brand M | 64.28% |
| Brand N | 51.36% |
| Brand R | 1.22% |

* Overall ROI is 60.67%, which suggests that, on average, the marketing campaigns across all brands have been profitable.
* Brand LO shows a high positive ROI of 130.44%, indicating a strong return on investment, while brand G, brand M, and brand N also demonstrate a positive sign for marketing investment.
* Brand LA has a negative ROI of -61.43% and Brand C has a negative ROI of -45.19%, indicating a significant loss. A reassessment of the marketing approach for brand LA and brand C may be needed.
* Identify any common trends
  + whether there is any channel outperforming others consistently

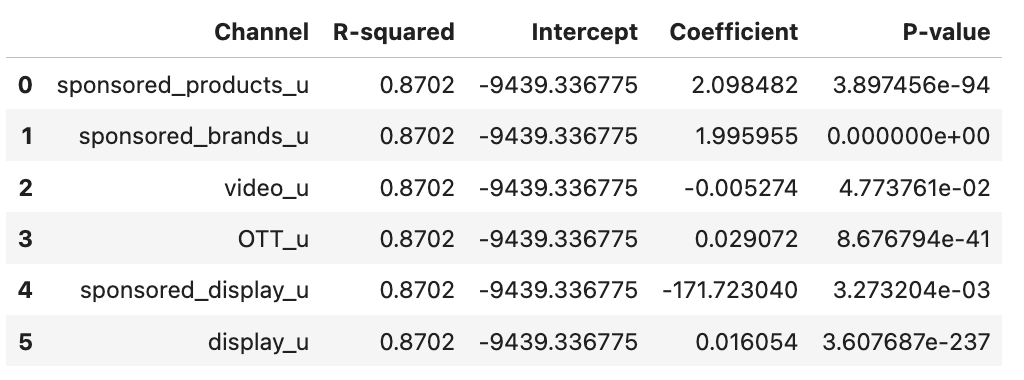
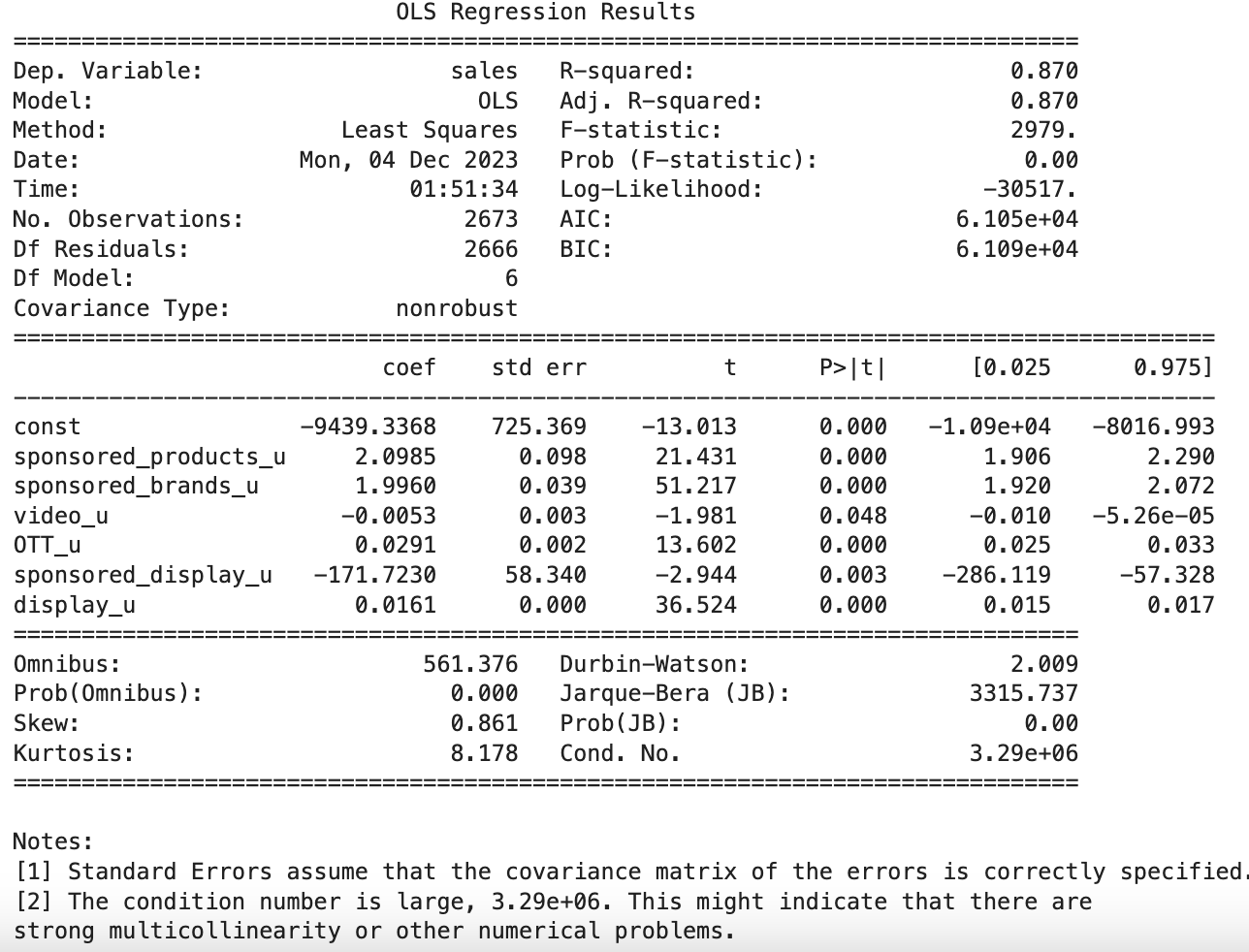
**Simple Regression Models**

* Regress sales on spend of different channels
* Let’s remove all the xxx\_u
* Sorry, I am not familiar with this kind of regression output. Why do we have an R-squared for each channel instead of an overall R-squared, F statistic, etc. for the entire model? And why do we have separate intercepts for each channel? What do they mean?Sorry I misunderstood! Have the model and attributes updated below. Not really sure if I did it in the correct way. I will double check in the afternoon.

One regression:

Multiple regressions: 

* (Regress sales on user reach of different channels) (only do if have time) (Question: do we have data for user reach of different channels? I thought we only have data for spend of different channels) (the xxxx\_u are user reach of different channels)



* Regress sales on user reach for all brands Individual brands separately

| Brand | R-squared | User\_Reach\_Coefficient |
| --- | --- | --- |
| All | 0.443433 | 0.029833 |
| Brand C | 0.098426 | 0.002576 |
| Brand G | 0.110013 | 0.005264 |
| Brand LA | 0.157731 | 0.002381 |
| Brand LO | 0.000552 | -0.000686 |
| Brand M | 0.124745 | 0.006169 |
| Brand N | 0.317301 | 0.008419 |
| Brand R | 0.052882 | 0.005056 |

* For all brands combined, an increase in total users reached is associated with a statistically significant increase in sales.
* Brand N has the highest R-squared value (31.73%), indicating that the model explains a significant portion of the variance in sales for this brand.
* Only Brand LO’s user reached has a negative relationship with sales. On average, a one-unit increase in total users reached is associated with a -0.000686 unit decrease in sales for Brand LO.
* Identify any common trends
  + whether there is any channel outperforming others consistently

How to better Google Lightweight MMM Model:

Client changing their dataset:

* Use spend instead of user reached of each channel as media\_data
* Total cost per channel: how to compute it for best results (use total spend of each channel as costs → prices = jnp.ones(mmm.n\_media\_channels) under optimization)

\*\*Optimized budget allocation with the model above is highly similar as the one with user reached as media\_data and average spend as costs

* What channels to include (collinearity, spend too high on one channel…) Interesting (maybe) results for spend too high on one channel

|  | Remove Sponsored Product (37% of total spend) | | | | Remove Display (38% of total spend) | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| In sample model fit | | Out sample prediction | | In sample model fit | | Out sample prediction | |
| (for all variables) | R square | MAPE | R square | MAPE | R square | MAPE | R square | MAPE |
| adstock | 0.762 | 9.4% | 0.86 | 4.3% | 0.845 | 7.7% | 0.737 | 9.2% |
| hill\_adstock | 0.838 | 6.9% | 0.17 | 12.9% | 0.921 | 4.888% | 0.889 | 6.3% |

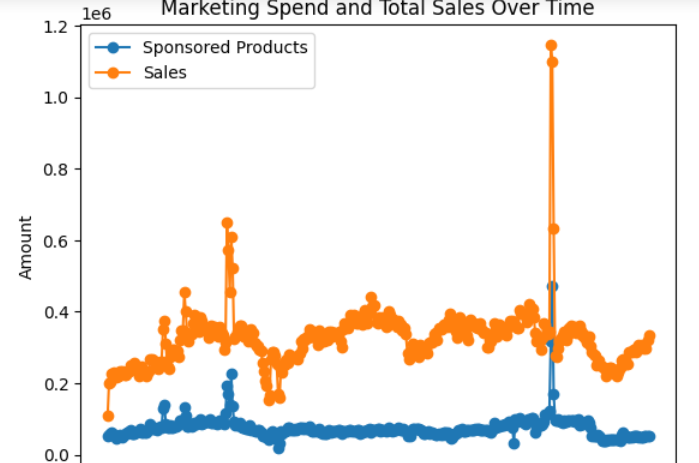
I tried to find a way to explain the huge drop in out of sample prediction accuracy for the hill adstock when removing Sponsored Products compared to the huge increase in prediction accuracy when removing Display.

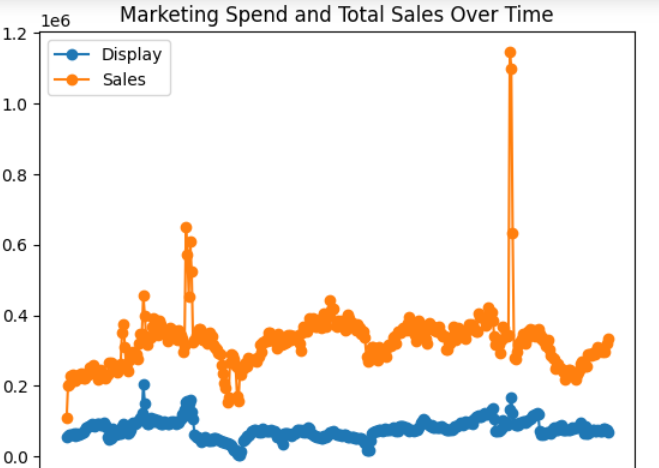
As we can see on the graphs below, Sponsored Products follow quite drastically to the T the distribution of Sales, when one increases the other does. Since the peaks in spend go quite high, they’re might be a need to apply a hill function to the distribution of Sponsored Products. Sponsored Products accounting for more than a third of the total spend, once removed, the hill\_adstock may have predicted quite badly because applying the hill transformation to the dataset without Sponsored Products may not make any sense anymore.

Additionally, Display on the other hand presents much lower peaks in spend and boasts more of the qualities regarding the adstock function. When it increases in spend, there is a small delay in the sales increasing after (compared to sponsored products where it is instantaneous). This is understandable as a Display presents an image and an advertising asset/image stays in the mind of a customer whereas a Sponsored products presents mostly a feature, a product page and the impact on purchase should be more instantaneous, the image of the product page will not stay in your mind long after seeing it. The fact that Display represents almost 40% of total spend and shows purchase delay in the data may also explain why when removing it from the model, the out of sample prediction of the adstock transformed model performed more poorly. The main subject of the adstock transformation is now gone.

These examples show the importance of applying different transformations to different channels. Robyn and Google Lightweight will for the most part apply the same transformation to all the data, but presenting a display or a sponsored product is quite different and one may not be subject to another’s transformation. If it is not possible to apply different transformations to different channels, the channels following a similar pattern that have the strongest weight in the data should be preferred, otherwise applying an inconsistent transformation may create a poor model.

\*\*kind of removing display by removing display





* Try to remove channels that are insignificant (drop sponsored display, which has a very high variance, accounts for a negligibly small fraction of total spend, and correlates negatively with sales and all other channels.)
* Look into seasonality of daily data and maybe compared daily and weekly data

Client changing Google lightweight input

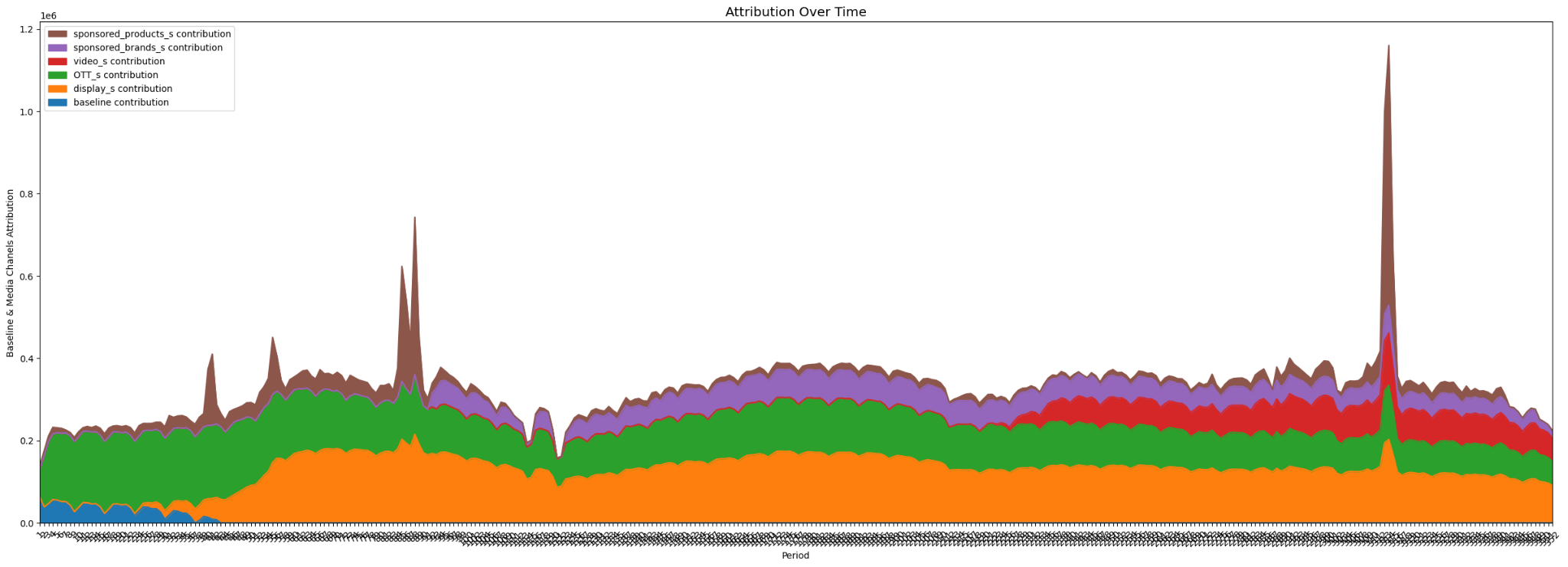
* Custom priors
* No lag effect in the data, understandable because amazon (but video for ex seems to have a small adstock)
* Try out “carryover” vs “hill\_adstock” vs “adstock” LOOK INTO WHICH ONE TO CHOOSE

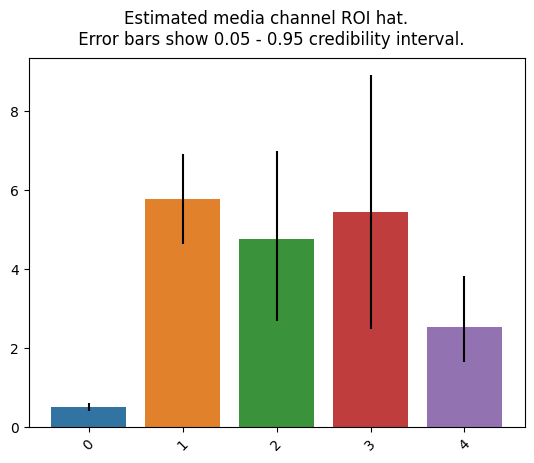
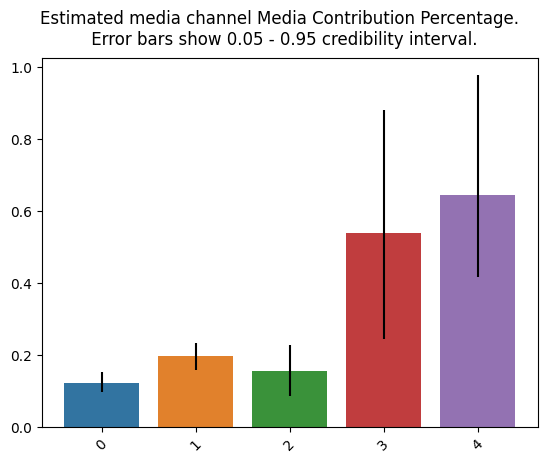
|  | In sample model fit | | Out sample prediction | |
| --- | --- | --- | --- | --- |
| (for all variables) | R square | MAPE | R square | MAPE |
| carryover | 0.853 | 7.374% | 0.570 | 10.382% |
| adstock | 0.872 | 6.556% | 0.827 | 6.373% |
| hill\_adstock | 0.934 | 4.348% | 0.657 | 4.460% |

(Similar results when “sponsored\_display” is dropped)

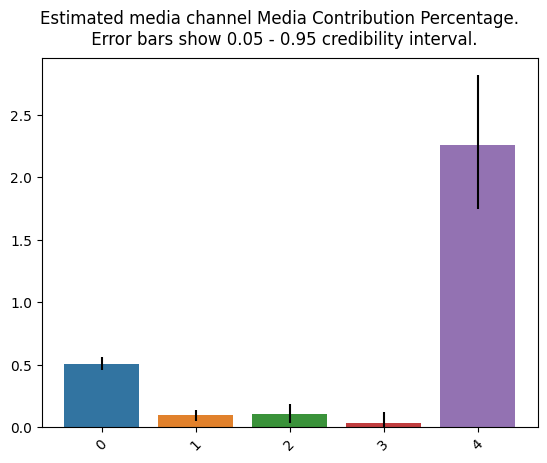
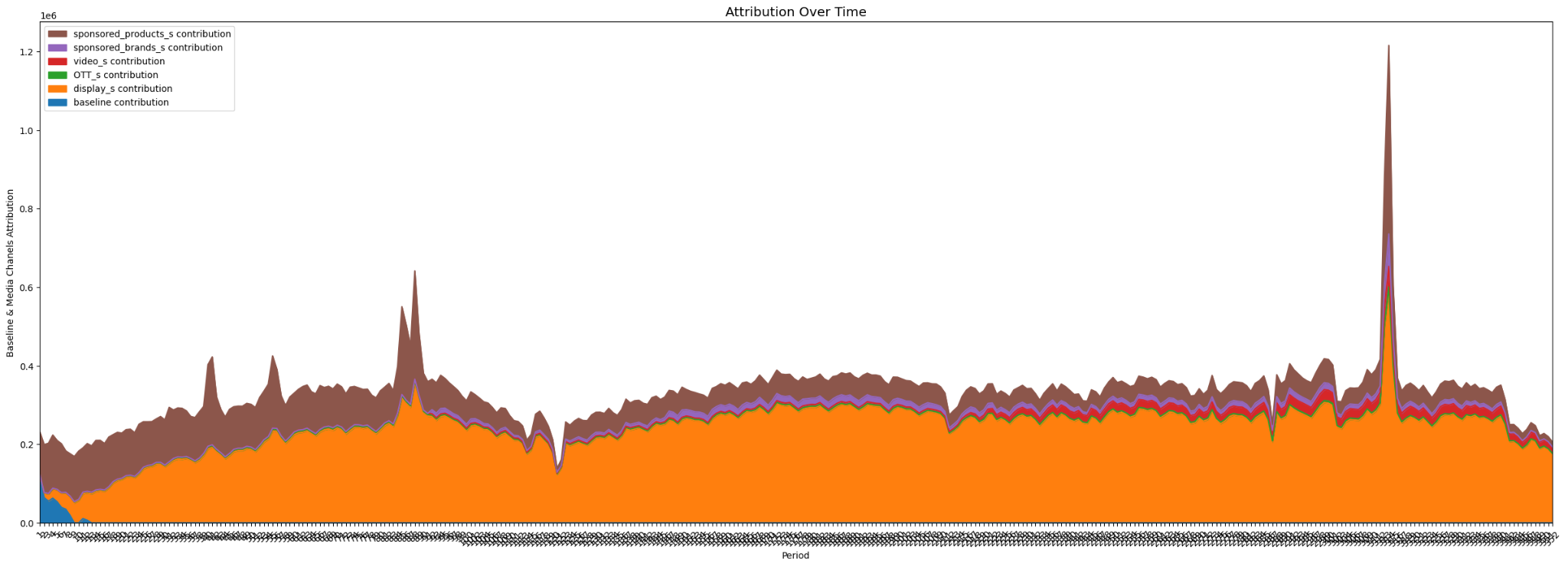
(choose “adstock” or “hill\_adstock”?? “Hill\_adstock” indicates relatively more evenly distributed contribution from the different channels -

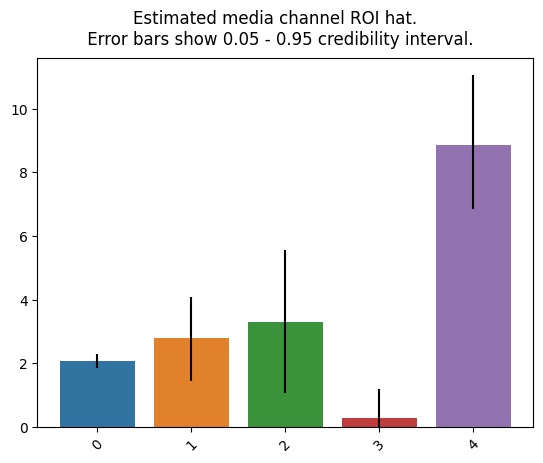
*Hill\_adstock*





*Adstock*



)

* Look more into changing: 1) number\_warmup=1000; 2) number\_samples=1000, 3) number\_chains=2 Default parameters are large enough, increasing warmups and samples to 2k or 1.5k and the number of chains to 3 or 4 gives the same in sample prediction and gives a lower explained out of sample predication because of overfitting
* Check if we can change “custom priors” DOESNOT CHANGE A THING

Optimization Model

* Try out different upper and lower bounds for the channels
* Optimization model for all brands and individuals brands

**Assumptions:**

* Date range: 90 days (2023-6-21 to 2023-9-18)
* Scenario: maximize response

*Specific outputs uploaded*

**Optimization 1: +-40%**

* Lower bound = 0.6; Upper bound = 1.4

|  | Total Response Increased | R square (Train) | R square (Test) |
| --- | --- | --- | --- |
| All Brands Combined | 52.1% | 0.6829 | 0.7856 |
| Brand C | 93.9% | 0.7699 | 0.6423 |
| Brand LA | 14.9% | 0.7702 | 0.3622 |
| Brand M | 3.39% | 0.7668 | 0.8542 |
| Brand N | 53.6% | 0.8761 | 0.3492 |

**Optimization 2: +-100%**

* Lower bound = 0; Upper bound = 2

|  | Total Response Increased | R square (Train) | R square (Test) |
| --- | --- | --- | --- |
| All Brands Combined | 146% | 0.6829 | 0.7856 |
| Brand C | 235% | 0.7699 | 0.6423 |
| Brand LA | 18.4% | 0.7702 | 0.3622 |
| Brand M | 18.7% | 0.7668 | 0.8542 |
| Brand N | 192% | 0.8761 | 0.3492 |

Other

* Do individual brands (least priority) (C, M, N, LA) with Hill Adstock model

|  | With all channels possible and Display Removed | | | | With all channels possible | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| In sample model fit | | Out sample prediction | | In sample model fit | | Out sample prediction | |
| (for all variables) | R square | MAPE | R square | MAPE | R square | MAPE | R square | MAPE |
| Brand C | 0.619 | 18.48% | 0.296 | 50.86% | 0.695 | 13.79% | 0.211 | 21.63% |
| Brand M |  |  |  |  |  |  |  |  |
| Brand N |  |  |  |  |  |  |  |  |
| Brand LA |  |  |  |  |  |  |  |  |

Comparison with Robyn

* Approaches and Methodologies
  + Performance metrics to compare their predictive power through accuracy, R-squared, and any industry-standard benchmarks
  + Compare the cost structure for both models and analyze the potential ROI for each channel implementing each model
* Identify any common conclusions
  + whether there is any channel outperforming others consistently
* Common parameters for optimization
  + Same initial budget
  + Period: 90 days
  + Upper and lower bounds
    - 100% change
    - 40% change

\*Budget optimization comparison for all brands: Robyn results seem to be comparable with GoogleLightweight using “carryover” / “hill\_adstock” (increase spending mostly on sponsored\_products), but not with using “adstock” (increase spending mostly on display)

Do we get the same optimization?

How to interpret why the results are different? (between robyn and google lightweight)

Differentiate between hill and exponent? If you assume otherwise, then the best solution is y

Final Report:

1. Brand Sales optimization (finding strongest model for best predictions)
2. Google Lighweight vs. Robyn comparisons (advantages and disadvantages)
3. How to best use MMMs and limitations