<RMP data analysis task>

RMP(Retail Media product) : a sponsored ad solution for retail businesses to monetize their current asset, inventory, and commerce data, which feeds Moloco's recommendation engine to provide the most-buyable selection to end customers upon their request.(GSshop, Brandi 등의 쇼팽앱에서 '이런 상품 어때요?'라는 상품 광고/추천 배너에 MOLOCO recommendation engine을 이용해 적절한 상품을 제공하는 것)

Basic terms

- Impression : the item is shown to the user
- Click : the user click the impression item
- Purchase : the user purchase the clicked item
- D# Revenue : revenue from purchasing within #days of clicking on the item
- Ad spending : spending of advertisers. In RMP, spending is the form of CPC(cost per click)
- CTR : Click-through rate, click count per impression count
- CR : Conversion rate, purchase count per click count
- D# ROAS : Day# Revenue On Ad Spending, revenue from Ad / Ad spending

Role of ODS(Operational Data Scientist) team in RMP product :

- Data analysis requested from Biz team, ML team using bigquery
- Make internal/external dashboard using Looker(Sponsored Ad performance status dashboard)

Weekly RMP meeting : Sharing the analysis status

- Weekly ODS meeting every Thursday
- Weekly RMP – ODS meeting every Tuesday
- Weekly RMP Dev meeting every Tuesday
- Weekly RMP – Biz meeting every Wednesday

Task1. RMP performance analysis based on requested categories and respond categories

<Background>

In the case of PDP (Product Detail Page, inventory_id: prd-nbst), the GS SHOP requests selected items for Sponsored AD for each context item.

- context item: the item that made a request (request from)
- selected item: respond item (response to)

Each item has categories 1, 2, 3 (large, medium, small), and the categories of the respond item and the requested item are (a) the same or (b) different. If we know there're differences in performance between (a) and (b), then it would be helpful to set product policy and ML model training.

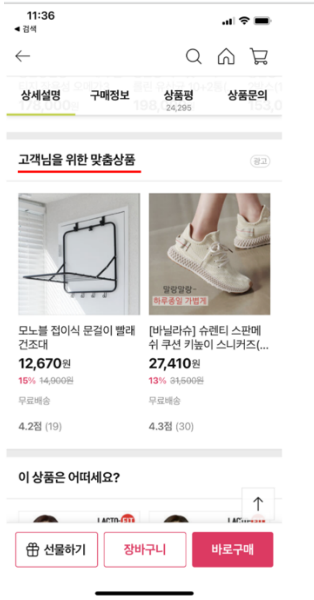


Fig 1. Product Detail Page inventory in GSshop

<Purpose>

We want to figure out if the difference between the requested categories and the response categories has an impact on advertising performance.

<Method>

Set two groups,

- A : (1) Respond to the request, (2) impression occurred, and (3) the requested item's category and the respond items' category were **the same**.
- B : (1) Respond to the request, (2) impression occurred, and (3) the requested item's category and the respond items' category were **different**.

And compare the following performance indexes between two groups,

- CTR, CR, D7 ROAS, etc

<Result>

	category type	imp	ratio(%)	CTR(%)	CR(%)	Ad_spending	revenue	D7_ROAS(%)
0	different category	4424126	66.247481	3.814878	1.230632	10155390.0	55437741.0	545.894751
1	same first category	826415	12.374854	5.496633	1.239406	2772340.0	19569365.0	705.878969
2	same second category	266224	3.986475	6.086979	1.493366	964210.0	6924164.0	718.117837
3	same third/all category	728819	10.913437	7.789177	1.868978	3635670.0	40992989.0	1127.522272
4	different category with empty	408179	6.112129	3.324032	1.709906	827700.0	13262607.0	1602.344690
5	same category with empty	24417	0.365624	2.891428	2.407932	58830.0	1357235.0	2307.045725

Fig 2. Performance comparison based on category

1. If the requested item's category and the response item's category were the same, the performance was better in all of CTR, CR, D7 ROAS than the different category case.
2. Also, the more sub-categories are the equal, the better performance in all of CTR, CR, D7 ROAS is.

Task2. Discovery of Evergreen Seller clusters in RMP products

<Background>

There are a variety of sellers in the RMP product. If we know the typical characteristics of sellers with **good performance or evergreen sellers (not churn-out)**, it will be helpful to track the performance of the RMP product and onboard new clients. However, because products within the e-commerce domain have a high churn-in/out rate of sellers, it's difficult to find out the typical characteristics of these sellers.

<Purpose>

First, We need to define who is the evergreen sellers in the Moloco RMP product (only for 'GSSHOP').
After that, we divide evergreen sellers into some seller clusters and try to analyze the differences between clusters.

<Method>

1. Define the evergreen sellers based on the seller's campaign operation data
2. Cluster the evergreen sellers
2.1. (Data Exploration) Check which data we can use
2.2. (Research Methodologies) Research which clustering algorithm is the best suitable for this task
2.3. (Feature engineering) define features to be used in clustering
2.4. (Find Clusters) cluster sellers into several seller groups and compare between groups
2.5. (Implications) Find the implications (significant differences and its reason) between groups

<Result>

Definition of Evergreen Seller : Seller who operates campaigns until now and whose valid date(daily campaign spending over ₩1,000) ratio is over 0.7

Row	seller_title	seller_id	seller_end	seller_duration	cnt	ratio
1	(주)컨트롤케이앤제이	1031583	2022-01-11	252	238	0.9444444444444444
2	(주)인씨엘	1028688	2022-01-11	255	246	0.9647058823529412
3	엑스앤오	1040941	2022-01-11	255	253	0.9921568627450981
4	주식회사 다솜인터내셔널	1003350	2022-01-11	179	177	0.9888268156424581
5	금성덴탈	1044013	2022-01-11	255	238	0.9333333333333333
6	(주)커커	1034555	2022-01-11	250	232	0.928
7	주식회사네이쳐리빙	1011216	2022-01-11	253	253	1.0
8	주식회사 한스갤러리	1031744	2022-01-11	93	87	0.9354838709677419
9	빼다 BBEDA	1045023	2022-01-11	252	245	0.9722222222222222
10	피피컴퍼니	1045385	2022-01-11	199	197	0.9899497487437185
11	(주)서울코리아	1038168	2022-01-11	245	240	0.9795918367346939
12	피카소	1042769	2022-01-11	246	228	0.926829268292683
13	(주)디바인바이오	1040333	2022-01-11	253	235	0.9288537549407114
14	D102 메리움점	1043443	2022-01-11	255	252	0.9882352941176471
15	가온누리	1030686	2022-01-11	230	225	0.9782608695652174
16	(주)가담에스앤에이치	1028599	2022-01-11	227	209	0.920704845814978

Fig 3. Total 72 Evergreen Sellers

First analysis based on seller features

- Seller features

- Valid date ratio
- Avg. # of campaign per month
- Avg. # of items per campaign
- Avg. daily budget
- Avg. item unit price
- # of unique categories

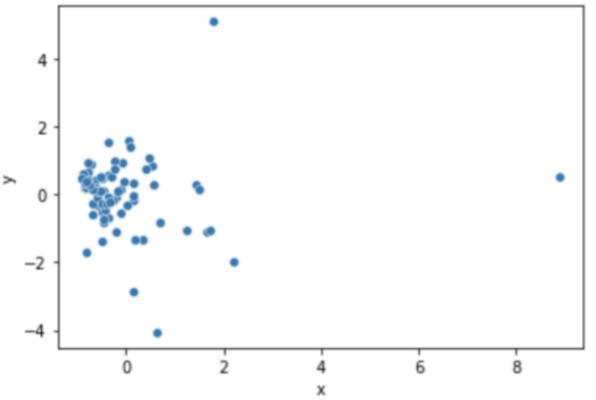


Fig 4. Distribution of seller with PCA algorithm

- The distribution of seller is concentrated except for a few
- Even after seller clustering, there is no evergreen seller group with distinct characteristic
- New approach : group the sellers by their performance(CTR, CR, ROAS)

Second analysis based on seller performance

- Seller Performance
  - CTR(%)
  - CR(%)
  - D7 ROAS(%)
- Normalize & K-means algorithm

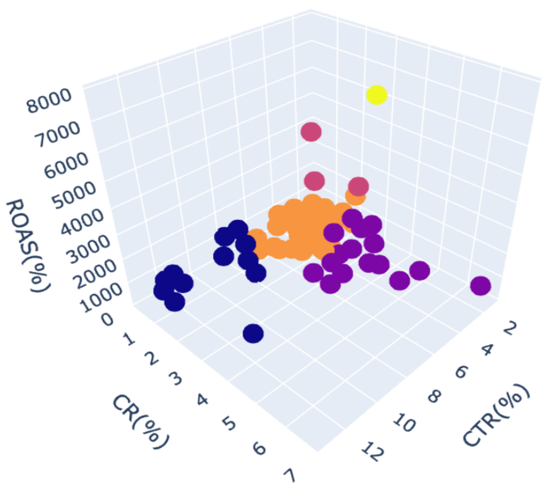


Fig 5. Distribution of Evergreen sellers after k-means clustering

- 3 effective evergreen seller groups
  - Group1 : High CR, Low CTR
  - Group2 : Low CR, High CTR
  - Group3 : Highest ROAS

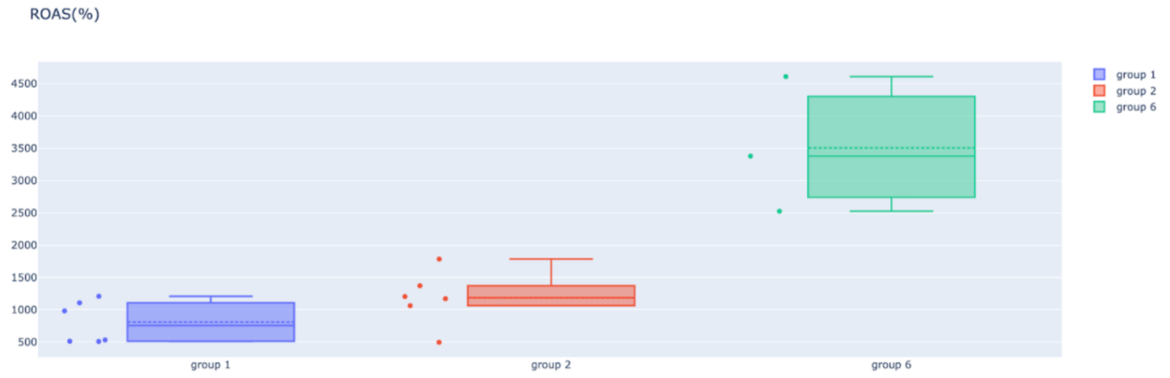


Fig 6. ROAS distribution

Fig 7. CTR distribution

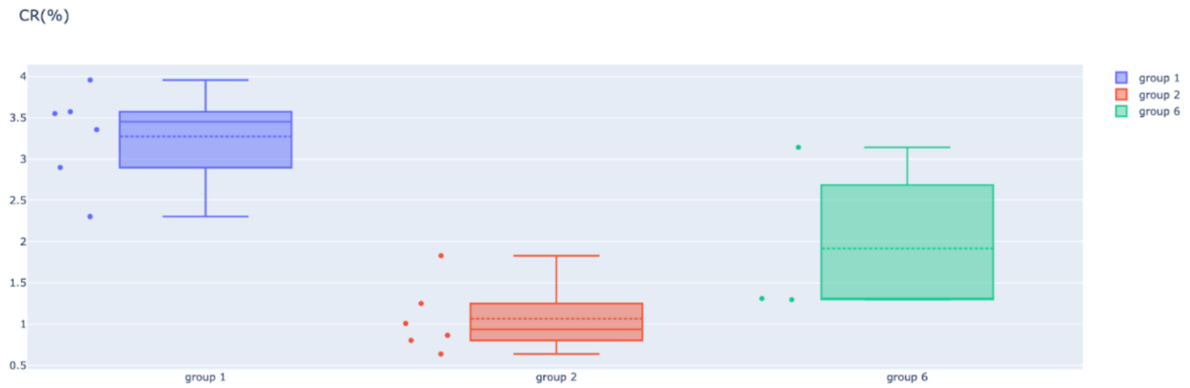


Fig 8. CR distribution

- Characteristics differences among evergreen seller groups

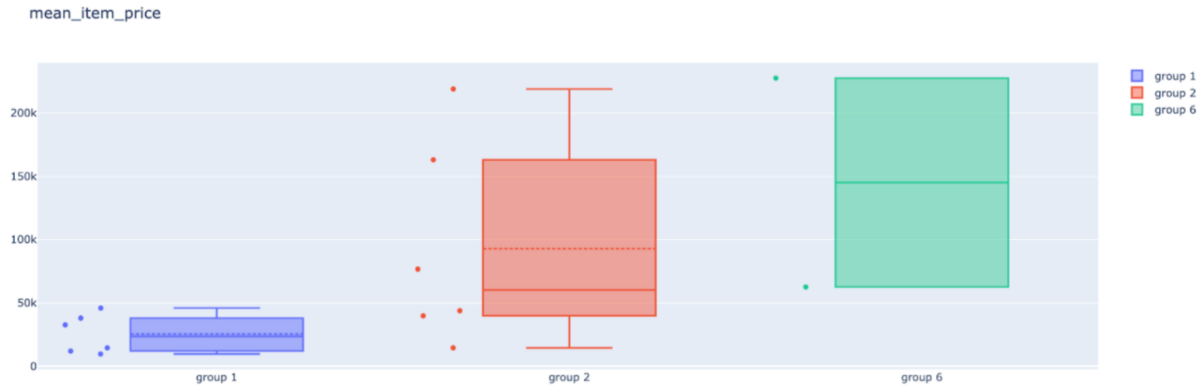


Fig 9. mean item price for each evergreen seller groups

Fig 10. mean target cpc for each evergreen seller groups

- Seller who operates campaign with high price of items show low CR
- Seller who operates campaign with high target cpc show low CTR
- Seem to be no clear characteristic determining the high performing sellers

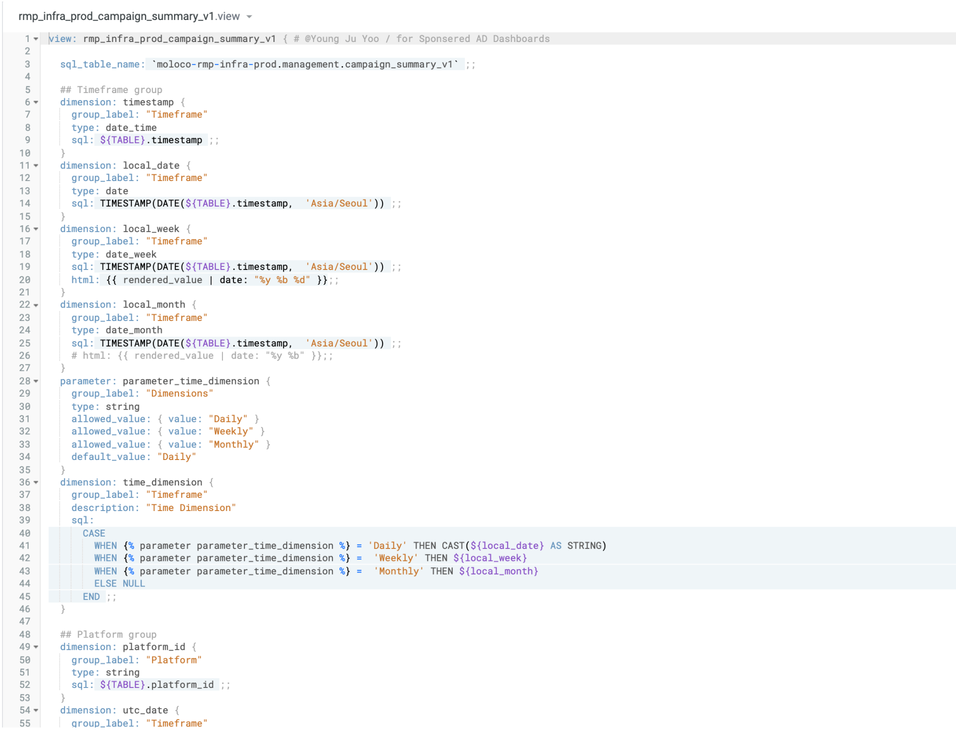
Task3. Sponsored AD[Bucketplace] Looker Dashboard

<Background & Purpose>

Make looker dashboard for Bucketplace sponserd ad.

<Method>

Make campaign_summary, campaign_digest, decision_request, decision_response, user_signals view file and create dashboard



rmp_infra_prod_campaign_summary_v1.view

```
view: rmp_infra_prod_campaign_summary_v1 { # @Young Ju Yoo / for Sponsered AD Dashboards

  sql_table_name: `moloco-rmp-infra-prod.management.campaign_summary_v1` ;;

  ## Timeframe group
  dimension: timestamp {
    group_label: "Timeframe"
    type: date_time
    sql: ${TABLE}.timestamp ;;
  }
  dimension: local_date {
    group_label: "Timeframe"
    type: date
    sql: TIMESTAMP(DATE(${TABLE}.timestamp,  'Asia/Seoul')) ;;
  }
  dimension: local_week {
    group_label: "Timeframe"
    type: date_week
    sql: TIMESTAMP(DATE(${TABLE}.timestamp,  'Asia/Seoul')) ;;
    html: {{ rendered_value | date: "%y %b %d" }};;
  }
  dimension: local_month {
    group_label: "Timeframe"
    type: date_month
    sql: TIMESTAMP(DATE(${TABLE}.timestamp,  'Asia/Seoul')) ;;
    # html: {{ rendered_value | date: "%y %b" }};;
  }
  parameter: parameter_time_dimension {
    group_label: "Dimensions"
    type: string
    allowed_value: { value: "Daily" }
    allowed_value: { value: "Weekly" }
    allowed_value: { value: "Monthly" }
    default_value: "Daily"
  }
  dimension: time_dimension {
    group_label: "Timeframe"
    description: "Time Dimension"
    sql:
      CASE
        WHEN {% parameter parameter_time_dimension %} = 'Daily' THEN CAST(${local_date} AS STRING)
        WHEN {% parameter parameter_time_dimension %} =  'Weekly' THEN ${local_week}
        WHEN {% parameter parameter_time_dimension %} =  'Monthly' THEN ${local_month}
        ELSE NULL
      END ;;
  }

  ## Platform group
  dimension: platform_id {
    group_label: "Platform"
    type: string
    sql: ${TABLE}.platform_id ;;
  }
  dimension: utc_date {
    group_label: "Timeframe"
```

Fig 11. View file(Campaign summary)

<Result>

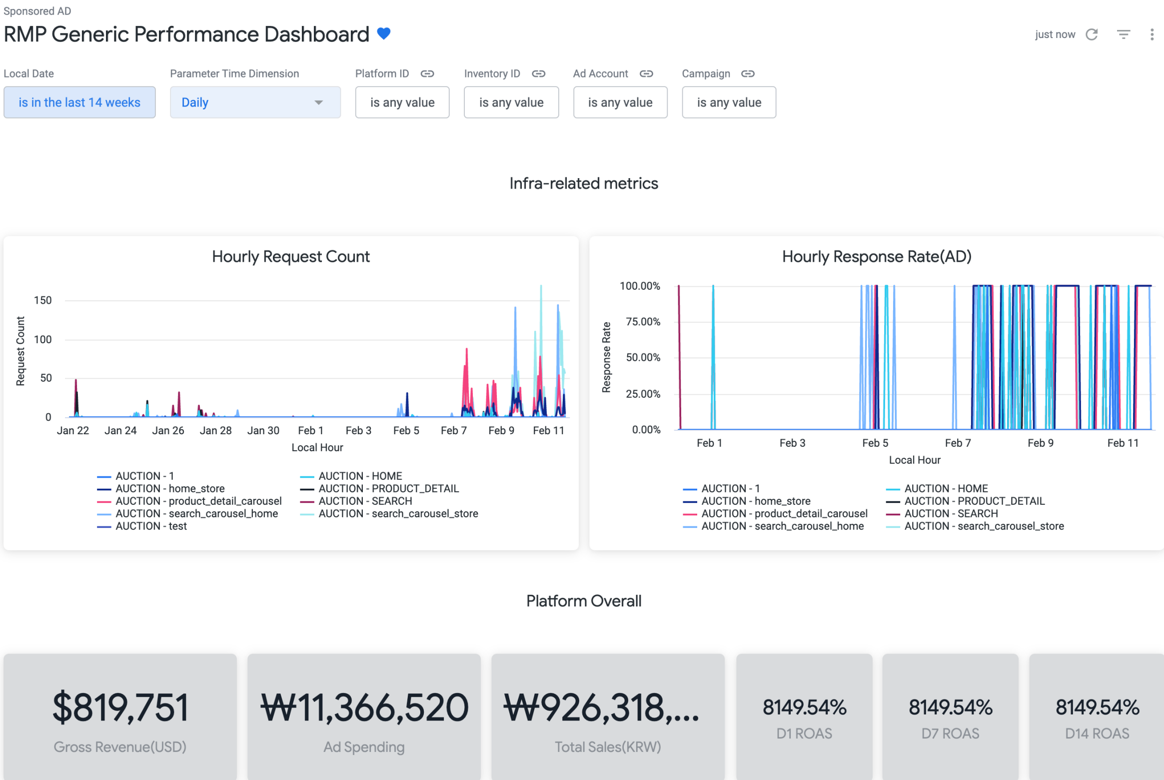


Fig 12. Bucketplace Internal Dashboard(1)

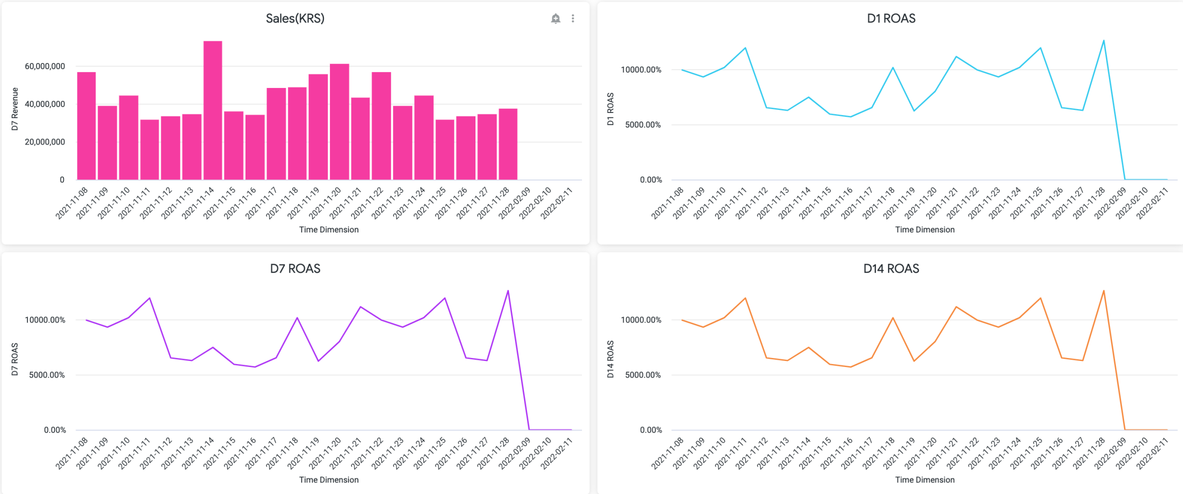


Fig 13. Bucketplace Internal Dashboard(2)

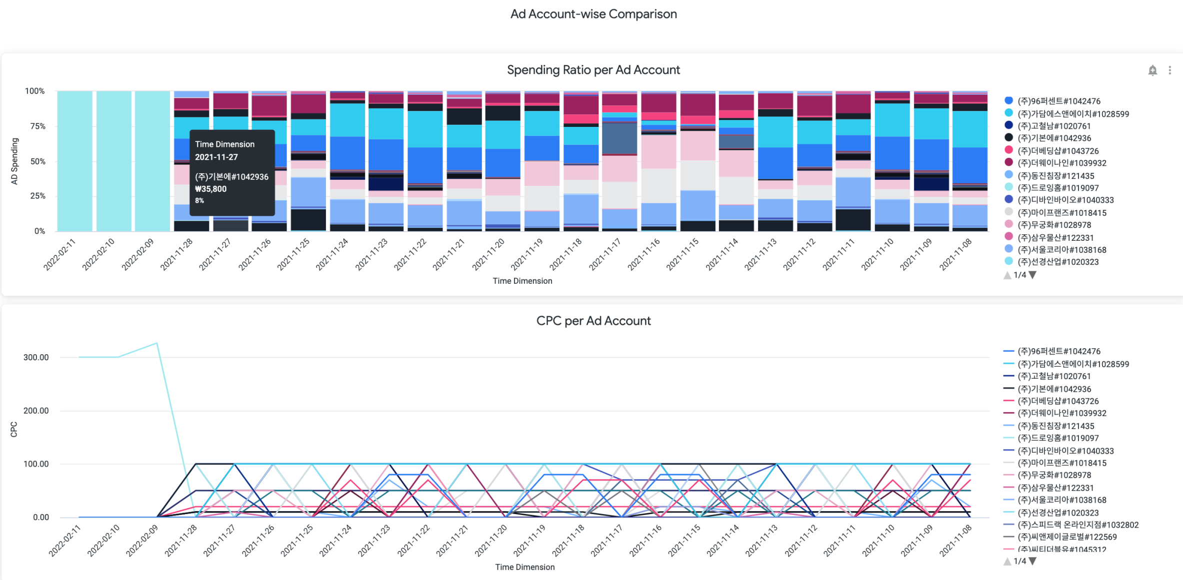


Fig 14. Bucketplace Internal Dashboard(3)

Task4. RMP Value pricing Investigation

<Background>

In the RMP internal auction step, each candidate's value price is calculated by CPC x relevancy_score.

So, the top 6 candidates who have the highest CPC x relevancy_score are selected. Because relevancy_score represents the expected CTR (Click-Through Rate, impression-to-click), this value-pricing formula (Cost per Click x expected Click Rate) makes us expect to maximize sellers' Ad spending compared to impression opportunities.

However, until now, we only check the CTR, CR, and ROAS as the performance indexes that are not related to the value-price directly. So, it is necessary to check whether our value-price formula reflects the real situation well.

platform_id	mtid	mtid_per_item	rank	cpc	score
GSSHOP	ChA---MqJGhCjb5bSRcmwDTuEMbvkZAGGhQIARoQvtCk47t5SOiS1vwKGNdlICACKgAyAA	ChA6NQaIN3RDh5wXCf6B0FYNEMbvkZAGGhQIARoQvtCk47t5SOiS1vwKGNdlICACKgAyAA	0	100.0	0.07033801823854446
GSSHOP	ChA---MqJGhCjb5bSRcmwDTuEMbvkZAGGhQIARoQvtCk47t5SOiS1vwKGNdlICACKgAyAA	ChCN2FmFKTFNf49d8ISxwjF0EMbvkZAGGhQIARoQvtCk47t5SOiS1vwKGNdlICACKgAyAhgB	1	100.0	0.066044382750988
GSSHOP	ChA---MqJGhCjb5bSRcmwDTuEMbvkZAGGhQIARoQvtCk47t5SOiS1vwKGNdlICACKgAyAA	ChAvGXuEkEBEllbhJJDdnLXhEMbvkZAGGhQIARoQvtCk47t5SOiS1vwKGNdlICACKgAyAhgC	2	100.0	0.05926716700196266
GSSHOP	ChA---MqJGhCjb5bSRcmwDTuEMbvkZAGGhQIARoQvtCk47t5SOiS1vwKGNdlICACKgAyAA	ChDSCCYDAGlDHYwinPEnuKgcEMbvkZAGGhQIARoQvtCk47t5SOiS1vwKGNdlICACKgAyAhgD	3	100.0	0.05584614723920822
GSSHOP	ChA---MqJGhCjb5bSRcmwDTuEMbvkZAGGhQIARoQvtCk47t5SOiS1vwKGNdlICACKgAyAA	ChBt3tQ65YJOy70s99AjaLCaEMbvkZAGGhQIARoQvtCk47t5SOiS1vwKGNdlICACKgAyAhgE	4	100.0	0.04614824801683426
GSSHOP	ChA---MqJGhCjb5bSRcmwDTuEMbvkZAGGhQIARoQvtCk47t5SOiS1vwKGNdlICACKgAyAA	ChAmy-3E2ONFXYZxlqXlo4F4EMbvkZAGGhQIARoQvtCk47t5SOiS1vwKGNdlICACKgAyAhgF	5	100.0	0.037515074014663696
GSSHOP	ChA---MqJGhCjb5bSRcmwDTuEMbvkZAGGhQIARoQvtCk47t5SOiS1vwKGNdlICACKgAyAA	ChCj3AYRgoxIR6H77DHWHdXJEMbvkZAGGhQIARoQvtCk47t5SOiS1vwKGNdlICACKgAyAhgG	6	70.0	0.040931664407253265
GSSHOP	ChA---MqJGhCjb5bSRcmwDTuEMbvkZAGGhQIARoQvtCk47t5SOiS1vwKGNdlICACKgAyAA	ChARvp4EEP9NplOyly459jpfEMbvkZAGGhQIARoQvtCk47t5SOiS1vwKGNdlICACKgAyAhgH	7	70.0	0.034544359892606735
GSSHOP	ChA---MqJGhCjb5bSRcmwDTuEMbvkZAGGhQIARoQvtCk47t5SOiS1vwKGNdlICACKgAyAA	ChBFVZXAtVlJl7GsRGmU2WDPEMbvkZAGGhQIARoQvtCk47t5SOiS1vwKGNdlICACKgAyAhgI	8	100.0	0.019597411155700684
GSSHOP	ChA---MqJGhCjb5bSRcmwDTuEMbvkZAGGhQIARoQvtCk47t5SOiS1vwKGNdlICACKgAyAA	ChD9YbSN9mxEELWnaYhibUoZEMbvkZAGGhQIARoQvtCk47t5SOiS1vwKGNdlICACKgAyAhgJ	9	100.0	0.016598261892795563
GSSHOP	ChA---MqJGhCjb5bSRcmwDTuEMbvkZAGGhQIARoQvtCk47t5SOiS1vwKGNdlICACKgAyAA	ChCZ5JuJJYJOB7L-AuYjitfUEMbvkZAGGhQIARoQvtCk47t5SOiS1vwKGNdlICACKgAyAhgK	10	100.0	0.014208987355232239
GSSHOP	ChA---MqJGhCjb5bSRcmwDTuEMbvkZAGGhQIARoQvtCk47t5SOiS1vwKGNdlICACKgAyAA	ChBI1ogjr5JEsb0odFPSkY0hEMbvkZAGGhQIARoQvtCk47t5SOiS1vwKGNdlICACKgAyAhgL	11	100.0	0.013976757414638996
GSSHOP	ChA---MqJGhCjb5bSRcmwDTuEMbvkZAGGhQIARoQvtCk47t5SOiS1vwKGNdlICACKgAyAA	*null*	12	20.0	0.02286849357187748

Fig 15. RMP Internal Auction Table

<Purpose>

The value-price formula maximizes the expected sellers' Ad spending compared to the impression opportunities. So we want to check that candidates with higher value-price scores showed higher (Ad spending)/(impression count).

Unfortunately, relevancy_scores are not normalized and have different scales for each internal auction. So we cannot compare value-price (or relevancy_score) with (Ad spending)/(impression count) directly. Instead of using the value-price score, the rank in the internal auction can be used.

Max 6 candidates are responded to each request and all 0-5 ranked candidates get an opportunity for an impression. Therefore, it will be verified whether the high-ranked candidates showed higher (Ad spending)/(impression count).

Moreover, if there's no relationship between rank and (Ad spending)/(impression count), we can suggest a new value-pricing formula that makes us expect to maximize (Ad spending)/(impression count) well.

<Method>

Hypothesis: The responded items with higher rank show the higher (Ad spending)/(impression count).

Steps

1. decide conditions to gather data: time period, platform (GS SHOP), inventory id, etc.
2. gather data (by joining the internal auction table with history_imp/click tables)
3. check the relationship between rank and (Ad spending)/(impression count)
4. If there's no relationship, propose a new value-pricing formula instead of CPC x relevancy_score.

<Result>

Row	rank	imp_cnt	click_cnt	purchase_cnt	Ad_spending	revenue	CTR	CR	ROAS	spending_per_imp	purchase_per_imp
1	0	7073434	262102	5081	2.048932E7	2.08519046E8	3.705442080890272	1.9385582712073925	1017.6962729851454	289.6658115421732	0.07183215394389768
2	1	2089646	163743	2345	1.258153E7	8.6880979E7	7.835920533908613	1.4321222891970955	690.5438289301858	602.0890619750905	0.11221996452987731
3	2	1590244	88785	1172	6333280.0	3.9335975E7	5.58310548569905	1.3200428000225262	621.0995724174519	398.25838047494597	0.073699381981633
4	3	1490116	77521	975	5375230.0	3.3707258E7	5.202346662944361	1.2577237135743862	627.0849433419593	360.72560793924765	0.06543114764219699
5	4	1438810	72159	855	4889430.0	3.1522012E7	5.01518616078565	1.1848833825302458	644.6970710287293	339.82457725481476	0.05942410742210577
6	5	1424445	71006	766	4685320.0	2.7094558E7	4.984818648666674	1.0787820747542463	578.2861789589612	328.9224926199327	0.05377533004082292

Fig 16. Result table based on item rank

- The higher rank, the higher CTR and ad spending per impression except rank0
- In rank0, the lowest CTR and ad spending per impression but the highest ROAS was shown.
- We need to find the cause for extremely low CTR and ad spending per impression in rank0
- Also, since it shows great relation between rand and CTR, ad spending per impression with CPC x relevancy score formula, we don't have to propose a new value-pricing formula

<Additional analysis about extremely low performance of rank0>

- We supposed that the lowest CTR of rank0 is due to high impression caused by high repeated impression count.
- Repeated impression : Hourly repeated impressions of the same item to the same user
- But duplicate impression occurs evenly at all rank → duplicate impression is not cause of low performance of rank0.

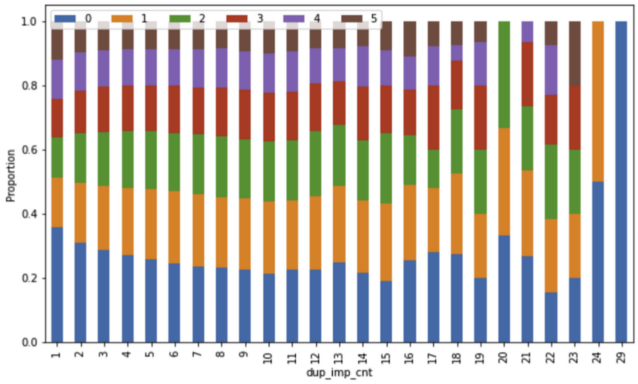


Fig 17. Proportion of rank for each repeated impression count

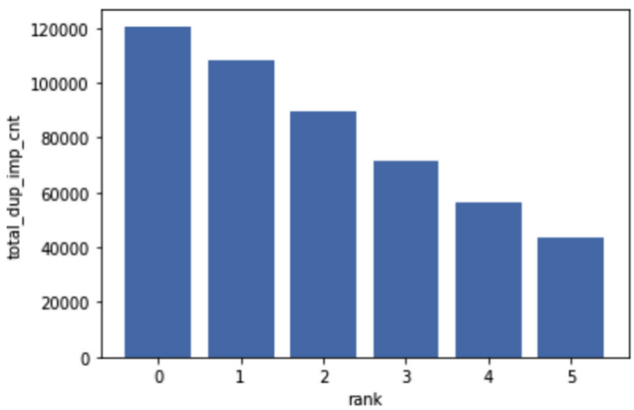


Fig 18. Total repeated impression count for each rank

Task5. Investigating the different characteristics of Impression model & Price multiplier

<Background>

The ML team has run two experiments for GSSHOP HOME sponsored ads.

- Baseline model = recommends items based on user signals data
- CTR optimization model = recommends items based on maximized CTR

- Sales amount optimization model = recommends items based on maximized ROAS

Have to check the characteristics of those different ranking scores.

Investigate the metrics'(CTR, CR, ROAS, etc.) trends by the rank in internal auctions.

<Purpose>

check the characteristics of those different ranking scores - such as which items are usually recommended by each and which advertisers benefit more.

<Method>

Same with Task4

<Result>

The metric trends by the rank

1. Baseline model

rank	imp_cnt	click_cnt	purchase_cnt	Ad_spending	revenue	spending_per_imp	CTR	CR	ROAS	purchase_per_imp
0	69541	2096	38	169930.0	3066500.0	244.359	3.014	1.813	1804.567	0.055
1	29745	1265	13	100620.0	339647.0	338.275	4.253	1.028	337.554	0.044
2	25110	664	5	48160.0	59560.0	191.796	2.644	0.753	123.671	0.02
3	24594	577	4	41570.0	65340.0	169.025	2.346	0.693	157.181	0.016
4	24431	535	7	39070.0	170028.0	159.92	2.19	1.308	435.188	0.029
5	24206	524	7	34550.0	132573.0	142.733	2.165	1.336	383.713	0.029

Fig 19. Result table of the baseline model

The higher rank, the higher CTR except rank0

2. CTR optimized

rank	imp_cnt	click_cnt	purchase_cnt	Ad_spending	revenue	spending_per_imp	CTR	CR	ROAS	purchase_per_imp
0	69339	2242	16	184460.0	451019.0	266.026	3.233	0.714	244.508	0.023
1	30096	1405	12	112600.0	480823.0	374.136	4.668	0.854	427.019	0.04
2	25472	662	8	49230.0	148447.0	193.271	2.599	1.208	301.538	0.031
3	24560	541	5	39090.0	263250.0	159.161	2.203	0.924	673.446	0.02
4	24169	540	7	38580.0	189420.0	159.626	2.234	1.296	490.98	0.029
5	24067	533	4	35210.0	85420.0	146.3	2.215	0.75	242.602	0.017

Fig 20. Result table of the CTR optimized model

The higher rank, the higher CTR as model optimized CTR except rank 0

3. Sale amount optimized

rank	imp_cnt	click_cnt	purchase_cnt	Ad_spending	revenue	spending_per_imp	CTR	CR	ROAS	purchase_per_imp
0	55677	1407	16	112830.0	779231.0	202.651	2.527	1.137	690.624	0.029
1	23892	1025	9	80170.0	335392.0	335.552	4.29	0.878	418.351	0.038
2	19969	545	5	40270.0	89449.0	201.663	2.729	0.917	222.123	0.025
3	19536	459	2	34340.0	62720.0	175.778	2.35	0.436	182.644	0.01
4	19286	446	4	32590.0	134403.0	168.983	2.313	0.897	412.406	0.021
5	19408	456	2	31090.0	22710.0	160.192	2.35	0.439	73.046	0.01

Fig 21. Result table of the sales amount optimized model

Expected result is the higher rank, the higher ROAS but result is not. It should be test again after gather more data