



Promotion Analysis

for the Google Merchandise Store



Data Analyst Candidate

April Chiu

aprilchiu0503@gmail.com

Dataset

- An obfuscated [Google Analytics 4-property dataset](#) for BigQuery
- Contains Business/eCommerce data from [Google Merchandise Store](#)

The screenshot shows the Google Cloud Platform BigQuery interface. The top navigation bar includes 'Google Cloud Platform', 'Select a project', and a search bar. Below the navigation is a toolbar with various icons. The main area displays a table titled 'events_20210131' with a timestamp of '2021-01-31'. The table has three tabs: 'SCHEMA', 'DETAILS', and 'PREVIEW'. The 'PREVIEW' tab is selected, showing 15 rows of data. The columns are: Row, event_date, event_timestamp, event_name, event_params, event_previous_timestamp, event_value_in_usd, event_bundle_sequence_id, event_server_timestamp_offset, user_id, user_pseudo_id, privacy_info, user_properties, and us. The data rows show various events like 'page_view', 'scroll', and 'user_engagement' with corresponding timestamps and user IDs.

Row	event_date	event_timestamp	event_name	event_params	event_previous_timestamp	event_value_in_usd	event_bundle_sequence_id	event_server_timestamp_offset	user_id	user_pseudo_id	privacy_info	user_properties	us
1	20210131	1612069510766593	page_view	▼ (10+ rows)	-	6595101026	-	-	1026454.42271112504	▼ (...)	▼ (0 rows)	1	
2	20210131	1612069529243877	scroll	▼ (10+ rows)	-	9011338476	-	-	1026454.42271112504	▼ (...)	▼ (0 rows)	1	
3	20210131	1612069515781635	page_view	▼ (8 rows)	-	-6830522854	-	-	1026454.42271112504	▼ (...)	▼ (0 rows)	1	
4	20210131	1612069530073506	user_engagement	▼ (10 rows)	-	-8264942910	-	-	1026454.42271112504	▼ (...)	▼ (0 rows)	1	
5	20210131	1612069510766593	session_start	▼ (5 rows)	-	6595101026	-	-	1026454.42271112504	▼ (...)	▼ (0 rows)	1	
6	20210131	1612069510766593	first_visit	▼ (6 rows)	-	6595101026	-	-	1026454.42271112504	▼ (...)	▼ (0 rows)	1	
7	20210131	1612111667188347	session_start	▼ (5 rows)	-	-2377566162	-	-	1029692.9551304564	▼ (...)	▼ (0 rows)	1	
8	20210131	1612111667188347	first_visit	▼ (6 rows)	-	-2377566162	-	-	1029692.9551304564	▼ (...)	▼ (0 rows)	1	
9	20210131	1612111672202517	page_view	▼ (10 rows)	-	2958840217	-	-	1029692.9551304564	▼ (...)	▼ (0 rows)	1	
10	20210131	1612111667188347	page_view	▼ (10+ rows)	-	-2377566162	-	-	1029692.9551304564	▼ (...)	▼ (0 rows)	1	
11	20210131	1612111672464586	user_engagement	▼ (10 rows)	-	9262234012	-	-	1029692.9551304564	▼ (...)	▼ (0 rows)	1	
12	20210131	1612066896782877	user_engagement	▼ (10 rows)	-	-8116520843	-	-	1031480.8260955562	▼ (...)	▼ (0 rows)	1	
13	20210131	1612066887624268	page_view	▼ (10 rows)	-	6304919096	-	-	1031480.8260955562	▼ (...)	▼ (0 rows)	1	
14	20210131	1612066882613427	first_visit	▼ (6 rows)	-	1630445792	-	-	1031480.8260955562	▼ (...)	▼ (0 rows)	1	
15	20210131	1612066882613427	page_view	▼ (10+ rows)	-	1630445792	-	-	1031480.8260955562	▼ (...)	▼ (0 rows)	1	

- **Traffic source data:** information about where website visitors originate. This includes data about organic traffic, paid search traffic, display traffic, etc.
- **Content data:** information about the behavior of users on the site. This includes the URLs of pages that visitors look at, how they interact with content, etc.
- **Transactional data:** information about the transactions that occur on the Google Merchandise Store website.
- **Time period:** 2020-11-01 to 2021-01-31

Objective

Increase purchase revenue for our eCommerce store!

The screenshot shows the Google Merchandise Store homepage. At the top, there's a navigation bar with links for New, Apparel, Lifestyle, Stationery, Collections, Shop by Brand, Sale (highlighted in red), and Campus Collection. Below the navigation is a search bar and a shopping cart icon. On the left, there's a sidebar with links for Login, Sign Up, and Help. The main content area features a large banner with the text "For Everyone" and "A new classic for your closet!" Below the banner is a blue "Shop now" button. To the right of the banner, there are four black t-shirts arranged in a 2x2 grid. Each t-shirt has "Google" printed on the back and "Create Design Code Build for everyone" printed on the front. The store's logo and name are visible at the top left of the page.



- Did promotions ([homepage carousels](#)) help the business grow?
- If yes, how can we [apply promotions](#) to improve our business in the future?

Analysis Process



#0 Get the data, dive deeper into the data, and [structure our analysis](#)



#1 Answer the first question: “[Did promotions help our business grow?](#)”



#2 Discover [successful promotions](#) and [popular item categories](#)



#3 Identify [product preference](#) among customers using cluster analysis



#4 Create [promotional strategies](#) to improve our business



#5 Design an [A/B test plan](#) for launching new promotion





Phase #0

Get the data, dive deeper into the data,
and **structure our analysis**

Get the data

- Create and manage service accounts and service account keys
- Create credentials with scopes
- Download query results for the dataset to Pandas DataFrame
- Events in the data:
 - **`view_promotion, select_promotion`**
 - **`view_item_list, select_item, view_item`**
 - **`add_to_cart`**
 - **`begin_checkout, purchase`**

```
In [2]: os.environ['GOOGLE_APPLICATION_CREDENTIALS'] = \
'//Users/aprilchiu2000/Desktop/LINE/service_account_key/graphical-tape-338802-82aed3845959.json'

scopes = ('https://www.googleapis.com/auth/bigquery',
          'https://www.googleapis.com/auth/cloud-platform',
          'https://www.googleapis.com/auth/drive')

credentials = Credentials.from_service_account_file('/Users/aprilchiu2000/Desktop/LINE/service_account_key/graphical-ta
credentials = credentials.with_scopes(scopes)
client = bigquery.Client(credentials=credentials)
```

```
In [3]: query = \
"""
SELECT DISTINCT
    event_date,
    FORMAT_TIMESTAMP('%Y-%m-%d %H:%M:%S', TIMESTAMP_MICROS(event_timestamp)) AS event_time,
```

```
In [4]: client = bigquery.Client()
query_job = client.query(query)
df = query_job.to_dataframe()
print("query success:", df.shape[0], "rows")
```

query success: 3980420 rows

```
In [5]: df.head()
```

	event_date	event_time	user_pseudo_id	continent	sub_continent	country	region	city	device_category	mobile_brand_name	item_revenue_in
0	2020-11-03	2020-11-03 11:04:19	8825068.2274309139	Europe	Southern Europe	(not set)	(not set)	(not set)	desktop	Apple	...
1	2020-11-03	2020-11-03 07:43:31	10673312.8932527724	Americas	Northern America	United States	Georgia	(not set)	desktop	Google	...
2	2020-11-03	2020-11-03 13:50:36	15626554.5768081465	Americas	Northern America	United States	New York	New York	mobile	Apple	...
3	2020-11-03	2020-11-03 22:34:53	16462431.2283594309	Americas	South America	Colombia	(not set)	(not set)	mobile	Apple	...
4	2020-11-03	2020-11-03 21:31:24	16765047.7521791154	Americas	South America	Ecuador	Pichincha	Quito	desktop	Google	...

5 rows × 51 columns

Structure the Analysis

Customer Journey Funnel



Events in This Funnel

view_promotion, select_promotion
- only happen when customers engage with promotions

view_item_list, select_item, view_item

add_to_cart

begin_checkout, purchase

- Enable us to:**
- Identify whether customers **respond to promotions** in each session
 - Uncover how many customers **bounce/churn** in each stage

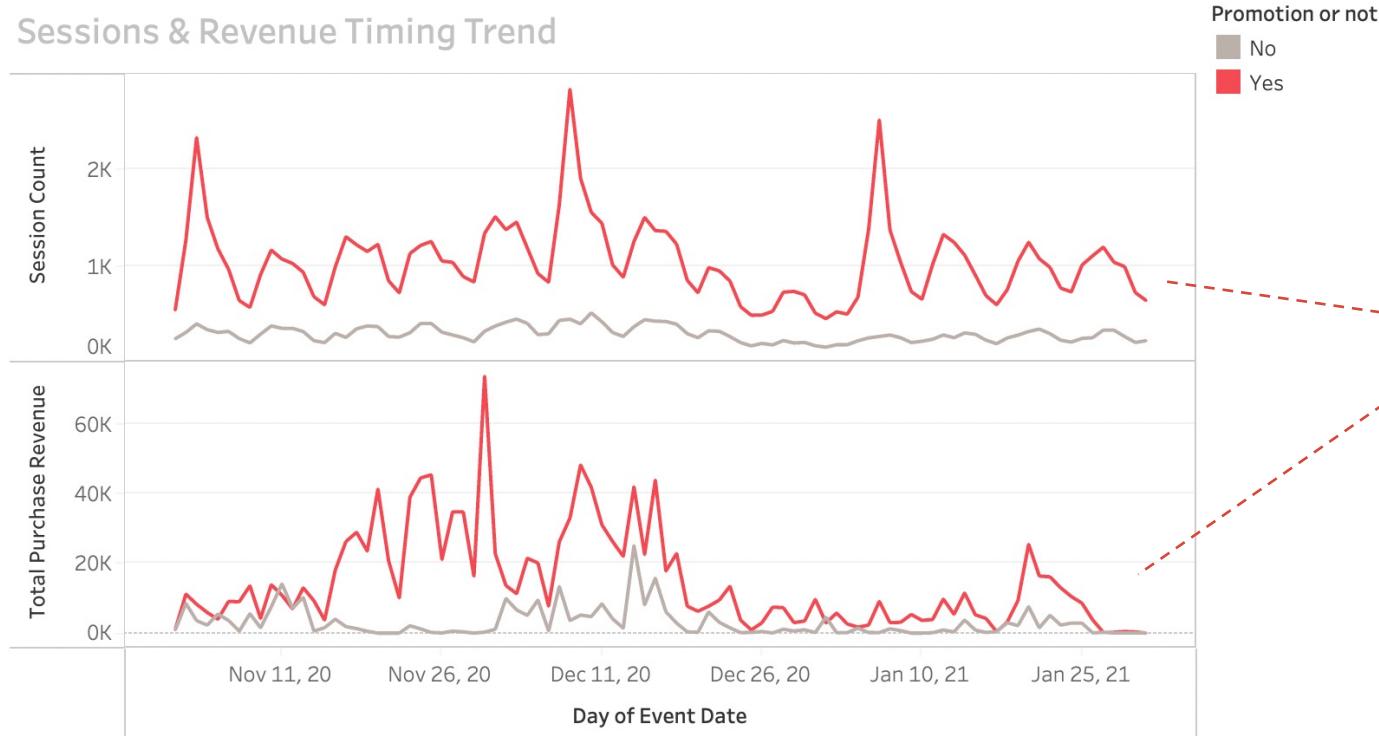
Phase #1



Answer the first question:
“Did promotions help our business grow?”

Did promotions help our business grow?

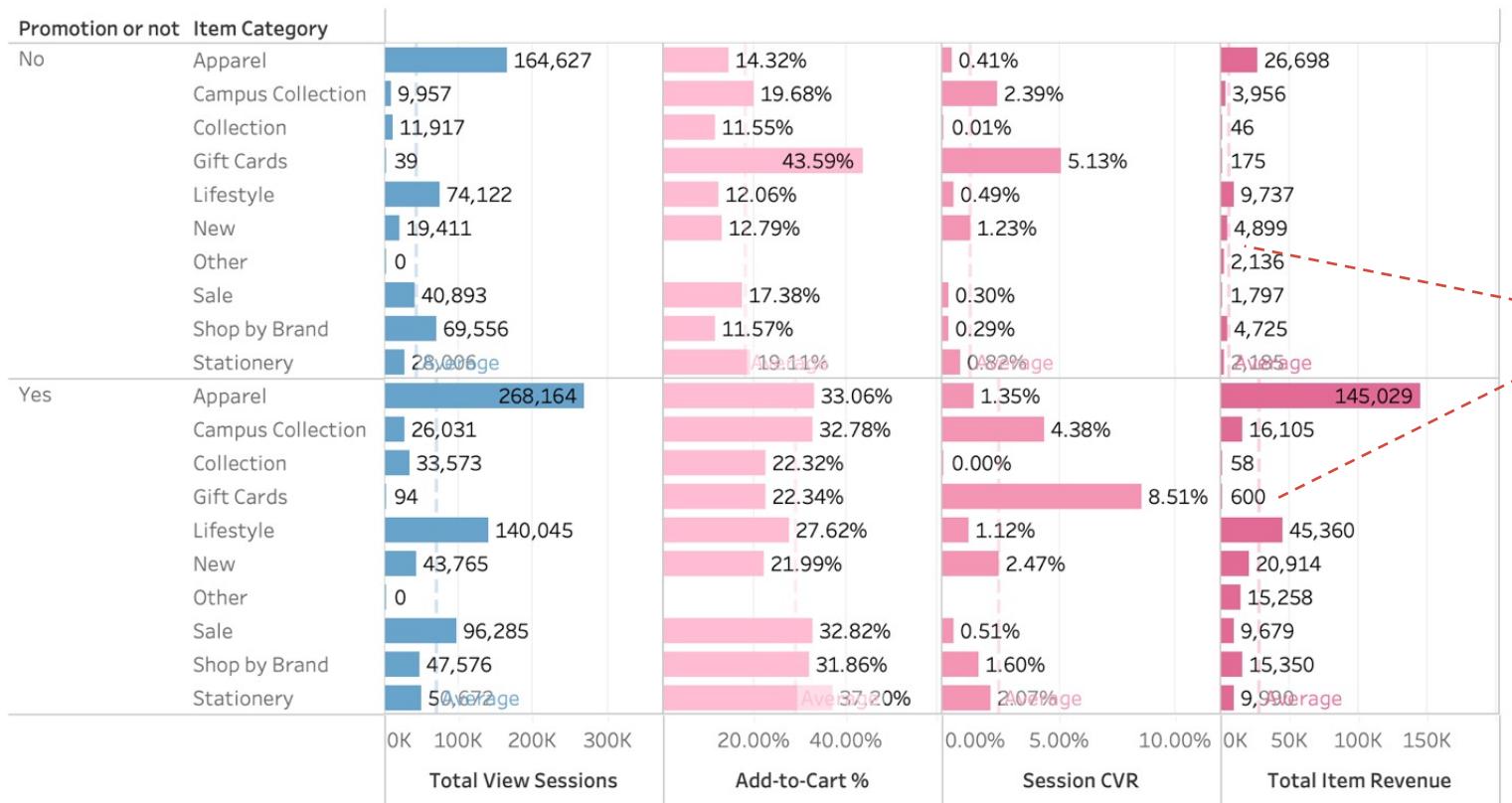
From sessions and total purchase revenue timing trend by promotion or not:



Daily customer count and total purchase revenue of sessions with promotions are mostly higher than sessions without promotions.

Did promotions help our business grow?

According to **bounce and abandonment of each item category by promotion or not:**



The average of all metrics of sessions with promotions are higher than that of sessions without promotions

$$\text{* Add-to-Cart Rate} = \frac{\text{Number of sessions added products to cart}}{\text{Number of sessions}} \times 100\%$$

$$\text{* Session CVR} = \frac{\text{Number of sessions purchased products}}{\text{Number of sessions}} \times 100\%$$

Did promotions help our business grow?



Yes! Homepage carousels on the website did help our business grow.



However, how to improve business using promotions?

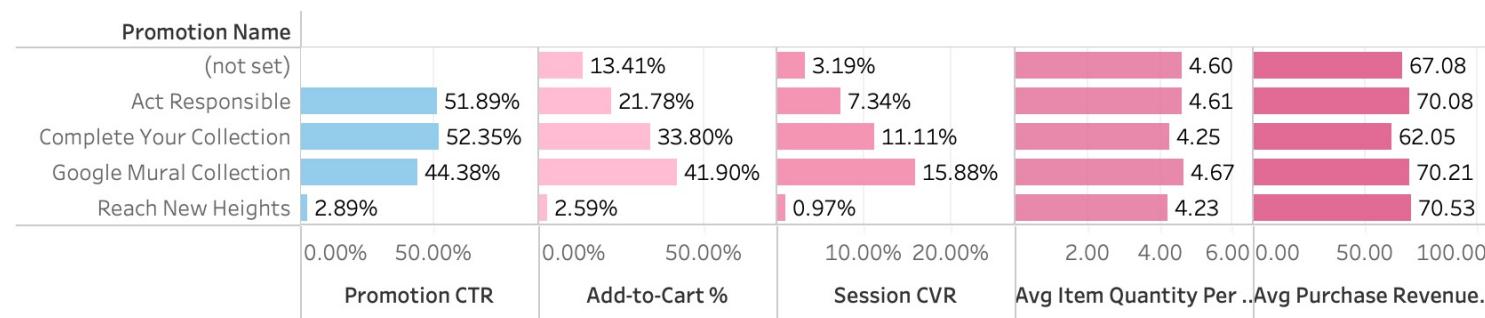
Phase #2



Discover **successful promotions**
and **popular item categories**

Discover Successful Promotions I

To measure promotion effectiveness, we calculate **metrics for each promotion**:



Metrics including:

$$\text{Promotion CTR} = \frac{\text{Number of sessions responded to this promotion}}{\text{Number of sessions viewed this promotion}} \times 100\%$$

$$\text{Item Quantity Per Transaction} = \frac{\text{Total quantity of purchased products in this promotion}}{\text{Number of sessions purchased item(s)}} \times 100\%$$

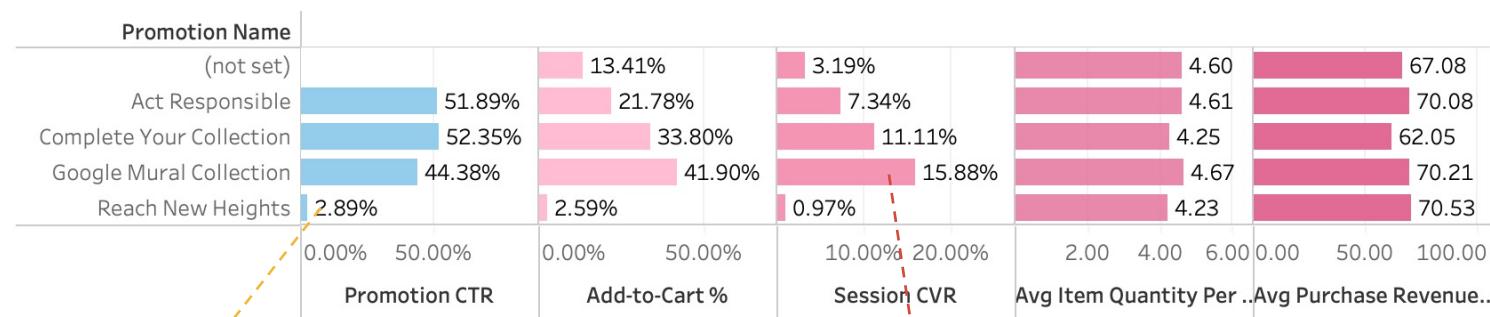
$$\text{Add - to - Cart Rate} = \frac{\text{Number of sessions added item(s) to cart}}{\text{Number of sessions viewed this promotion}} \times 100\%$$

$$\text{Purchase Revenue Per Transaction} = \frac{\text{Total purchase revenue in this promotion}}{\text{Number of sessions purchased item(s)}} \times 100\%$$

$$\text{Session CVR} = \frac{\text{Number of sessions purchased item(s)}}{\text{Number of sessions viewed this promotion}} \times 100\%$$

Discover Successful Promotions II

To measure promotion effectiveness, we calculate **metrics for each promotion**:



Among all promotions, “Reach New Heights” has the lowest performance.

“Act Responsible” and “Complete Your Collection” have relatively higher promotion CTRs, however, “Google Mural Collection” has the highest add-to-cart rate (41.9%) and CVR (15.9%).

Profits including item quantity and purchase revenue per transaction among promotions are similar.

Discover Popular Item Categories I

Also, we get **the percentage of unique items by item category among all unique items viewed by customers** to identify the most popular categories (mostly viewed by customers) for each promotion:

	(not set)	Promotion Name					Grand Total
		Act Responsible	Complete Your Collection	Google Mural Collection	Reach New Heights		
Accessories	4.78%	0.49%	0.35%	0.00%	0.13%	1.15%	
Apparel	11.25%	15.13%	16.70%	17.88%	13.15%	14.82%	
Campus Collection	7.53%	7.40%	8.44%	9.87%	8.72%	8.39%	
Collection	3.90%	4.11%	3.51%	5.03%	5.34%	4.38%	
Gift Cards	0.44%	0.00%	0.00%	0.00%	0.00%	0.09%	
Lifestyle	12.93%	16.12%	18.28%	19.55%	16.93%	16.76%	
New	4.78%	5.92%	4.57%	4.10%	4.56%	4.79%	
Sale	15.59%	20.72%	20.74%	22.16%	20.31%	19.90%	
Shop by Brand	33.22%	21.88%	18.28%	11.73%	23.57%	21.73%	
Stationery	5.58%	8.22%	9.14%	9.68%	7.29%	7.98%	

For instance, there's a total of 608 unique items viewed in the "Act Responsible" promotion, and among those unique items, 3 unique items are in the "Accessories" category, which accounted for 0.5% of the total unique items viewed (3/608) in this promotion.

Discover Popular Item Categories II

Also, we get **the percentage of unique items by item category among all unique items viewed by customers** to identify the most popular categories (mostly viewed by customers) for each promotion:

		Promotion Name					
	(not set)	Act Responsible	Complete Your Collection	Google Mural Collection	Reach New Heights	Grand Total	
Accessories	4.78%	0.49%	0.35%	0.00%	0.13%	1.15%	
Apparel	11.25%	15.13%	16.70%	17.88%	13.15%	14.82%	
Campus Collection	7.53%	7.40%	8.44%	9.87%	8.72%	8.39%	
Collection	3.90%	4.11%	3.51%	5.03%	5.34%	4.38%	
Gift Cards	0.44%	0.00%	0.00%	0.00%	0.00%	0.09%	
Lifestyle	12.93%	16.12%	18.28%	19.55%	16.93%	16.76%	
New	4.78%	5.92%	4.57%	4.10%	4.56%	4.79%	
Sale	15.59%	20.72%	20.74%	22.16%	20.31%	19.90%	
Shop by Brand	33.22%	21.88%	18.28%	11.73%	23.57%	21.73%	
Stationery	5.58%	8.22%	9.14%	9.68%	7.29%	7.98%	

“Apparel” and “Lifestyle” are popular categories among all customers.

“Sale” is popular among customers purchasing with promotions.
“Shop by Brand” is relatively more popular among customers purchasing without promotions.

Customers who don't respond to/engage with promotions tend to **browse around the website**.

Insights I – Promotions

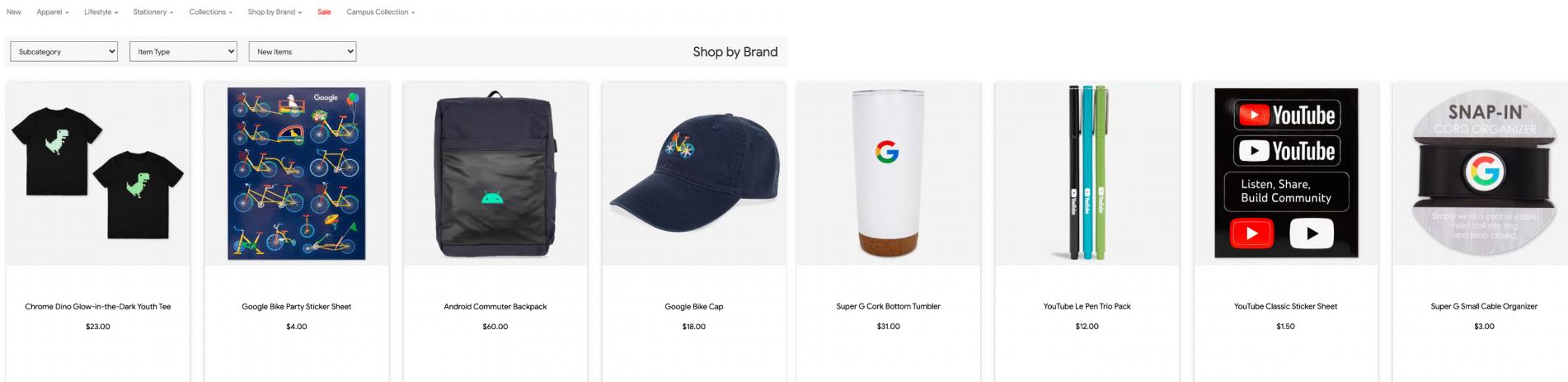
- **The causes of the poor performance of “Reach New Heights”**
 - Dig into the events and find this promotion may be **the first slide of our homepage carousel**
 - The content on this promotion may not be attractive to our customers, so they choose to skip it
- **Dive deeper into the the best performance promotion, “Google Mural Collection”**
 - This collection was conceived by a team of Googlers and a San Francisco based artist Monica Garwood
 - Great to invite talented artists or designers of special collections to **create artworks for advertising products**



Sources: [Mural Painting At GooglePlex](#)

Insights II – Item Categories

- Promote more items in “Apparel” and “Lifestyle”
 - Those items are popular among sessions with promotions and without promotions
 - They are most of our customers’ fav!
- Create promotional strategy specifically for “Shop by Brand”
 - Those items are popular among customers who don’t respond to promotions
 - Attract customers to our promotion and guide them to this item category efficiently



Sources: [Shop by Brand](#)

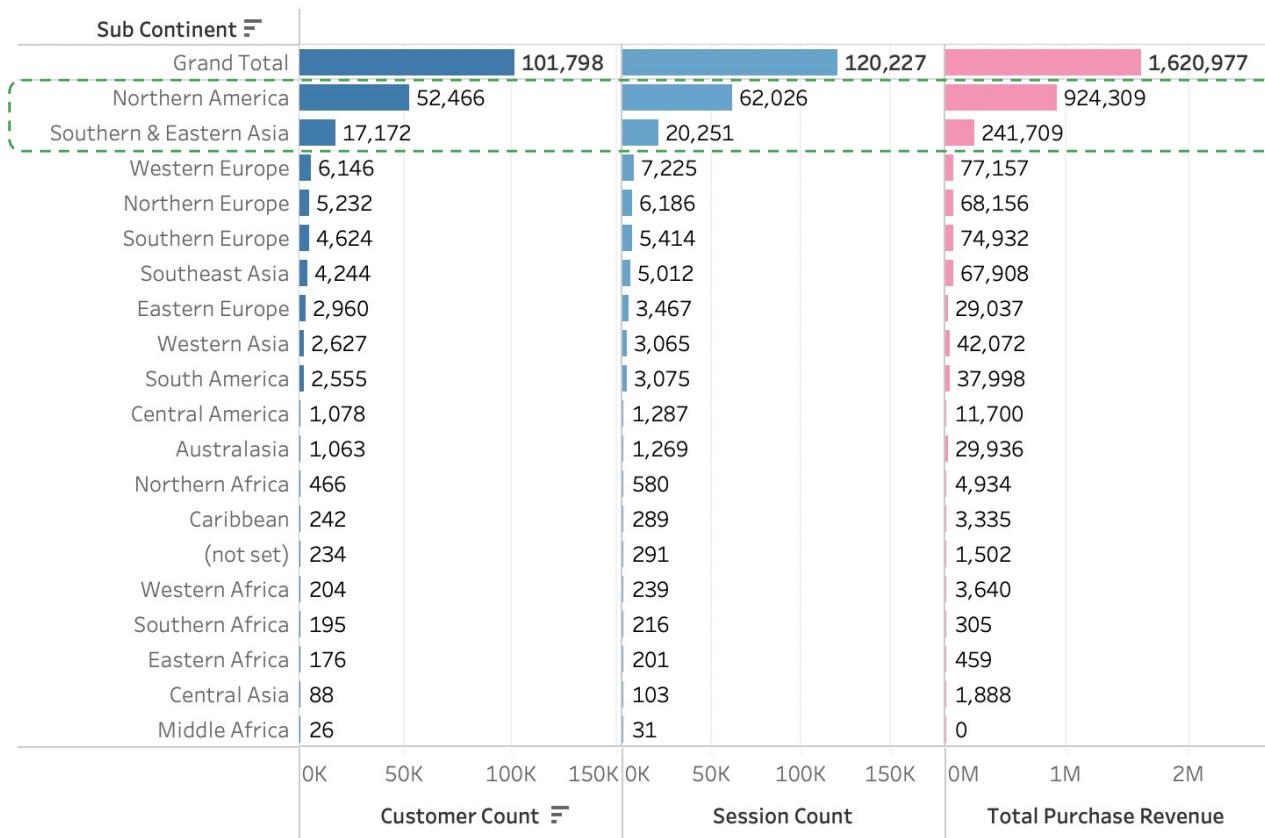


Phase #3

Identify **product preference**
among customers using cluster analysis

Our Main Markets

Customers, Sessions & Revenue by Country



Focusing on our two main markets:

- **Northern America**
 - Countries including the United States and Canada
 - Customer count: 52,466 (51.5% of total customers)
 - Session count: 62,026 (51.6% of total sessions)
- **Southern & Eastern Asia**
 - Countries including India, China, Taiwan, Japan, etc.
 - Customer count: 17,172 (16.9% of total sessions)
 - Session count: 20,251 (16.8% of total sessions)

Create Attributes for Cluster Analysis

First, we calculate **metrics for each item viewed/added to a cart/purchased by customers** in these two regions separately:



1. Number of unique sessions viewed this item



2. Number of unique sessions added this item to cart
3. Number of unique sessions purchased this item



4. Total purchased quantity of this item
5. Total purchase revenue of this item



6. Average price of this item among all transactions



Popularity of each item among customers

--> **Revenue each item bring to us**

--> **Important attribute (price range)**

Focus on Top Items

Second, we **narrow down the scope of products** in the region by only including products which viewed/added to cart/purchased sessions are greater than **at most 50% of all data points** in the region.



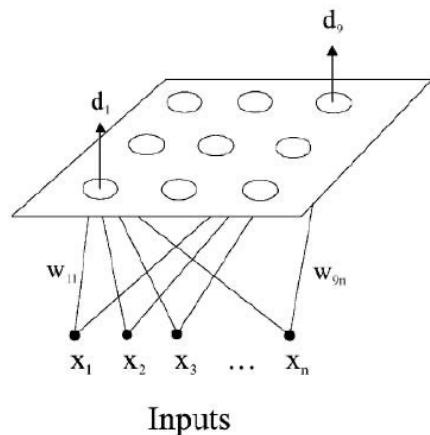
Original: 430 items -----> Top Products: 257 items

Only include top items



Original: 428 items -----> Top Products: 316 items

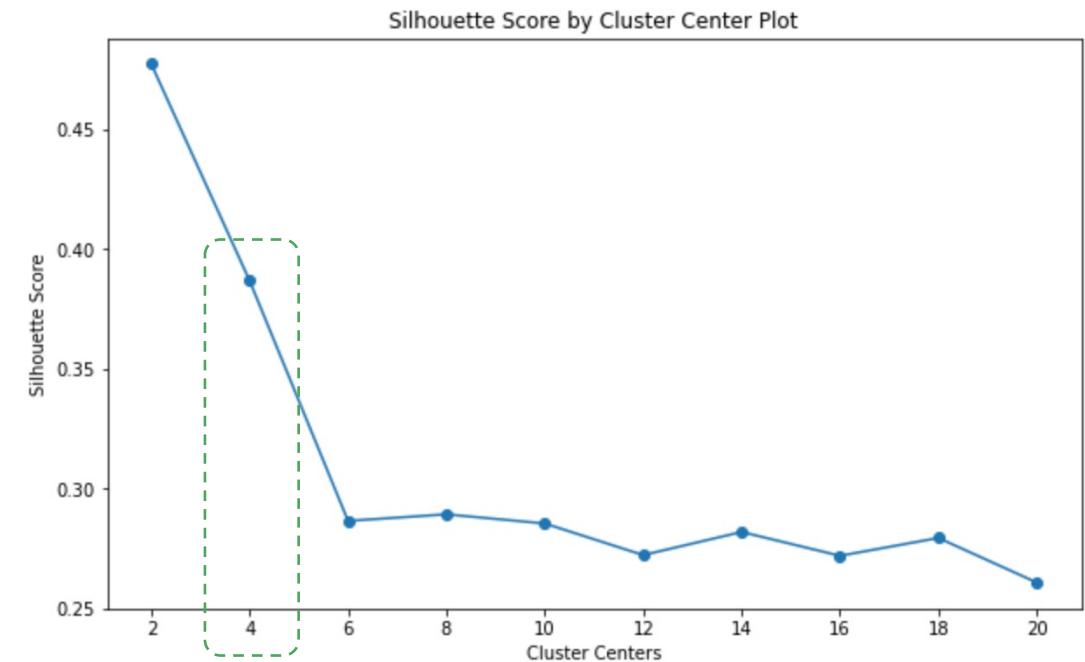
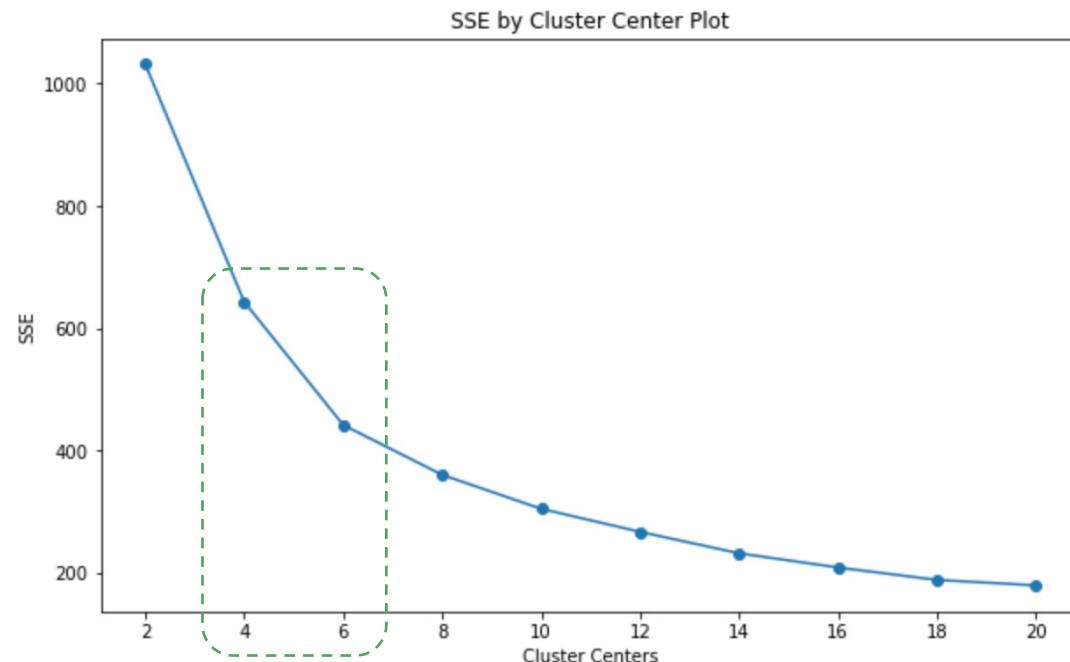
We then conduct **Kohonen Clustering Network** on items based on item popularity, profits and price.



Find the Optimal value of K

Take the results of Northern America for instance:

Step 1: Choose the optimal value of K for clustering



Cluster Items to Identify Product Preference



Take the results of Northern America for instance:

Step 2: Group those items into 4 clusters and find insights for each cluster

Cluster #1

Highly viewed and added to cart by customers, however, CVRs are relatively lower.

e.g., Google Crewneck Sweatshirt Navy, Google Badge Heavyweight Pullover

Cluster #2

Like items in Cluster #1, but sessions of view/add-to-cart and revenue from those items are lower.

e.g., Google Crewneck Sweatshirt Green, Google Canteen Bottle Black

Cluster #3

Slightly viewed by customers compared to items in other groups but add-to-cart rates of those items are relatively high.

e.g., Google Land & Sea Cotton Cap, Google KeepCup

Cluster #4

Has high add-to-cart rates and CVRs, which indicates that items in this group are popular among customers.

e.g., Google Men's Puff Jacket Black, Google Heather Green Speckled Tee

Insights I – Items

- **Top items we can focus on promoting in Northern America**

- **Cluster #1:** Items with **high add-to-cart rates** but relatively low CVRs. If customers purchase items in this group, it will bring us more revenue due to the high item prices.
 - **Cluster #4:** Items with **high add-to-cart rates and CVRs**, which are all popular among customers.



- **Top items we can focus on promoting in Southern/Eastern Asia**

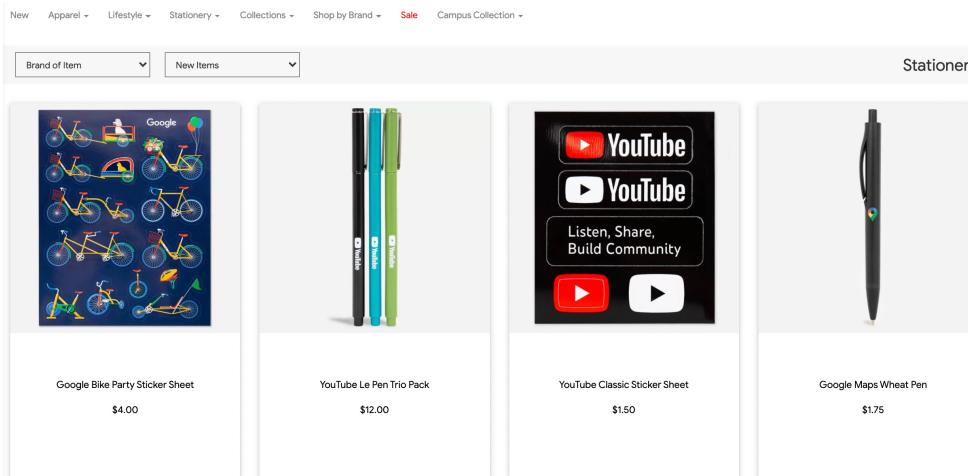
- **Cluster #3:** Items with **high average price, add-to-cart rates and CVRs**. Those items are popular and bring us great revenue because of their high prices.
 - **Cluster #4:** Items with **high average price, add-to-cart rates**, but CVRs of those items are lower than that of items in Cluster #3.



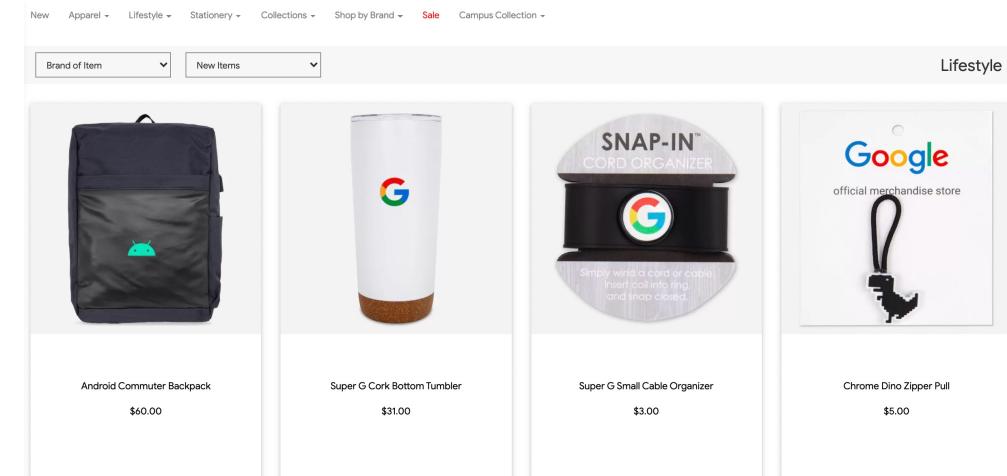
Insights II – Clearance Sales

- Some interesting insights from Cluster #3 in Northern America

- Most products in the "Stationery" and some in the "Lifestyle" are relatively inexpensive, are slightly viewed but are often added to a cart after being viewed by customers.
- In the future, when we focus on having clearance sales but not increasing profits, it will be suitable to promote items in those two categories in the sales.



Sources: [Stationery](#)



Sources: [Lifestyle](#)

Phase #4



Create **promotional strategies**
to improve our business

Promotional Strategies



Promote popular items with high value on homepage

→ Convert potential customers and increase purchase revenue!



Create a new promotion for “Shop by Brand”

→ Improve engagement in promotions and customer experience!



Invite artists or designers to create practical artworks

→ Promote item collections and emphasize brand image!

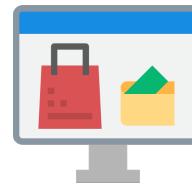
Additional Ideas



Launch seasonal clearance sales for

“Stationery” and “Lifestyle”

→ Improve clearance sales of our store
instead of only focusing on profits



Optimize the content on the first slide of

our homepage carousel

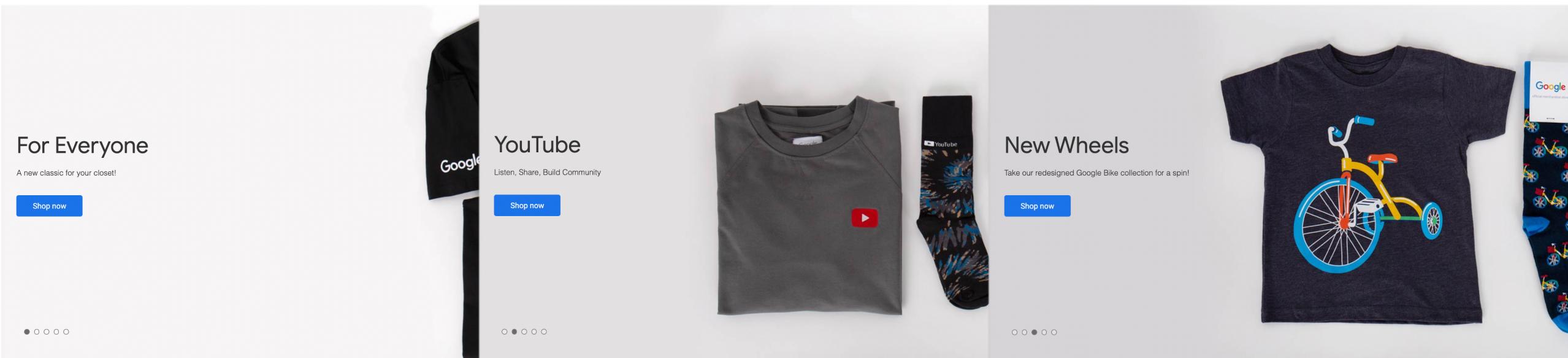
→ Make visitor’s very first experience of our
site greater!

Phase #5



Design an **A/B test plan**
for launching new promotion

Set Up the Experiment I



Sources: [Google Merchandise Store](#)

- According to the historical data, homepage carousels on our website have been **merely promoting specific item categories/collections**.
- We plan to improve our business in the Northern American region by **displaying popular and high-value products among customers** on a slide of our homepage carousel.

Set Up the Experiment II



Promote popular items with high value

on homepage

→ Convert potential customers and
increase purchase revenue!

- In our experiment, we are creating a slide showing items with high add-to-cart rates and CVRs in Northern America based on our cluster analysis, and testing two different contents:
 - The original slide which shows specific item category
 - The new slide promoting popular products
- We would like to evaluate the impact (if any) on revenue. Our hypothesis is: "**Showing a homepage carousel with popular and high-value items will increase revenue.**"
- **Objective:** Increase purchase revenue

Define Goal/Success Metrics



- Primary metric

Revenue Per Session: Does our business grow because of the new promotion?

$$\text{Revenue Per Session} = \frac{\text{Total purchase revenue}}{\text{Number of sessions clicked this promotion}} \times 100\%$$

- We include all potentially affected users, but no unaffected users (users who never select the promotion) who may distort our results.

- Secondary metrics

Promotion CTR: Does the content of the new promotion attract customers?

$$\text{Promotion CTR} = \frac{\text{Number of sessions clicked this promotion}}{\text{Number of sessions}} \times 100\%$$

CVR: Are customers actually buying more products?

$$\text{CVR} = \frac{\text{Number of sessions purchased item(s)}}{\text{Number of sessions clicked this promotion}} \times 100\%$$

Evaluate the Sample Size

We calculate the ideal sample size by the following formula:

$$N = \frac{2 \sigma^2 (Z_{1-\frac{\alpha}{2}} + Z_{1-\beta})^2}{\delta^2}$$

- α : set as 5% $Z_{1-\frac{\alpha}{2}} \approx 1.96$
- β : set as 20% $Z_{1-\beta} \approx 0.84$
- **Baseline metric**: The average revenue per session from 2021/1/1 to 2021/1/31 in Northern America, which is **around \$5.92**.
- σ : The standard deviation of revenue per session is **around \$5.3**.
- δ : We first set our desired growth on revenue per session as 10%, so the expected revenue per session after launching our new promotion will be $5.92 * (1+10\%) = 6.512$, and **the absolute MDE will be $6.512 - 5.92 = 0.592$** .

Sample size for each group:

$$N = \frac{2 \times 5.3^2 \times (1.96 + 0.84)^2}{0.592^2} \approx 1256$$

That is, we should collect a total of **$1,256 * 2 = 2,512$ subjects** for the experiment.

Test Duration and Experiment Summary

- Because groups of active users may be different every day (referring to day-of-the-week effects), it's necessary to run the experiment for at least a full week or run for a usage cycle such as multiples of 7 days.



In conclusion, our experiment design is as follows:

- For our experiment, we are creating a slide showing items with high add-to-cart rates and CVRs on our website for customers in Northern America based on our cluster analysis.
- The hypothesis of our experiment is “Showing a homepage carousel with popular and high-value items will increase revenue per session for customers who click this promotion.”
- The primary metric is revenue per session, and secondary metrics are promotion CTR and CVR.
- We will run our experiment for 7 to 14 days, and potentially prolong it depending on whether the subjects reach 2,448.

Summing Up



First, we downloaded query results for the dataset to DataFrame and structured the funnel.



From comparison of matrices, we found promotions did help our business grow.



We discovered successful promotions and popular item categories among our customers.



Then we identified product preference using cluster analysis for Northern America and S/E Asia.



Based on our analysis, we created several promotional strategies based on our analysis.



Finally, we designed an A/B test plan for launching new promotion.

