MARKETING ANALYTICS ON E-COMMERCE MULTICHANNEL Direct Messaging

Group 6 – BlackPink
120020362 Catherine Febriani
121090709 Yang Yanzhi
121020353 Fu Xiaohan
120090794 Hu Yuwen

CONTENTS

01

Introduction to The Marketing Challenge

Project Background
Marketing Problem Significance

Project Background

Digital Market Evolution:

- Shift towards **online direct marketing** in the 2000s
- Emphasis on platforms like Facebook and Twitter

Challenges in Modern Marketing:

- Increased complexity allowing for personalized marketing
- Variety of digital marketing options—choosing the best approach
- Necessity for targeted advertising on search engines

WHAT WE NEED TO DO...

Retargeting Campaigns

Optimize Marketing Message Delivery

Predict Future Performance

Data Sources:
Data from Kechinov, M. (2023a): "Direct messaging campaigns dataset overview"
Data from Kechinov, M. (2023b): "E-commerce multichannel direct messaging 2021-2023"

MARKETING PROBLEM

How to optimize the effect of campaign on consumers by adjusting the characteristics of the campaign message

WE SEEK TO TEST…

- Relationship between consumer’s different behaviors
- The effects of different campaign message’s characteristics on consumer’s behaviors
- How to adjust the characteristics to optimize the effect brought by the campaign message
- A way to predict the effect of future campaign message on consumer’s behaviors

02

Data and Analytic Approach

Dataset Introduction
Data Preparation and Variables

Dataset Introduction

About the Dataset

- Retail company's multichannel messaging campaigns for over two years, extracting from REES46 (a multinational marketing company)
- Focused on a medium-sized retail company from Russia
- Combined with 2 datasets

19 COLUMNS

"CAMPAIGNS"

All messages that are related to the campaigns.

- Describes the characteristics of each campaign
- bulk / ~~trigger~~ / ~~transactional~~
- Attains most of the campaign's characteristics (IV)

32 COLUMNS

"MESSAGES-DEMO"

All messages received by each user, describing the delivery methods and consumer's actions

- Consumer's actions as our DVs, identifying user behaviors towards each message
- Delivery methods (e.g. delivery channel) as IVs (filtered to include only bulk campaigns)

Data Sources
Data from Kechinov, M. (2023a): "Direct messaging campaigns dataset overview"
Data from Kechinov, M. (2023b): "E-commerce multichannel direct messaging 2021-2023"

Analytic Methods Used

Descriptive analytics helps us understand the current state by exploring relationships and patterns, while **predictive analytics** forecasts future trends and outcomes through regression modeling.

DESCRIPTIVE ANALYTICS

Correlation Analysis

Explore whether variables are related and how strongly

Clustering Analysis

Segment the types of message sent

PREDICTIVE ANALYTICS

Regression Analysis

Find out the influencing factors on different consumer's behaviors

Data Preparation And Variables

Table 1: Data description for “campaign” dataset

Column Name	Description	Variables Type
id	Unique campaign ID only for the specific campaign type	
campaign_type	Campaign type (bulk, trigger, transactional)	
channel	Channel (email, mobile_push, web_push, sms)	decision variables
topic	Meaning of a campaign (sale out, happy birthday, etc.)	decision variables
started_at	Bulk campaign start datetime	decision variables
finished_at	Bulk campaign finish datetime	decision variables
total_count	Total recipients in bulk campaign	decision variables
subject_length	Email subject length	decision variables
subject_with_personalization	Subject contains recipient's name	decision variables
subject_with_deadline	Subject has deadline meaning	decision variables

Data Preparation And Variables

Column Name	Description	Variables Type
subject_with_emoji	Subject has emoji symbols	decision variables
subject_with_bonuses	Subject mentions bonuses for actions	decision variables
subject_with_discount	Subject mentions a discount	decision variables
subject_with_saleout	Subject mentions a sale out	decision variables

Data Preparation And Variables

Table 2: Data description for "message-demo" dataset

Column Name	Description	Variables Type
id	Message sequence ID *will not be used	
message_id	Message unique ID	
campaign_id	Campaign ID from campaigns.csv)	
message_type	Campaign type (bulk, trigger, transactional)	decision variables
client_id	Client ID	
channel	Message channel (email, web_push, mobile_push, sms)	decision variables
category	Category *will not be used	
platform	Device type used to open a message	
email_provider	Public email provider (for email messages)	
stream	Additional identifier of data source (desktop, ios and android)	
date	date in YYYY-MM-DD when a message was sent	

Data Preparation And Variables

Column Name	Description	Variables Type
sent_at	Datetime when a message was sent	
is_opened	Boolean flag if a message was opened by a recipient	outcome variables
opened_first_time_at	First time when a message was opened	
opened_last_time_at	Last time when a message was opened (can be equal to opened_first_time_at, if the message was opened only once)	
is_clicked	Boolean flag if a message was clicked by a recipient	outcome variables
clicked_first_time_at	First time when a message was clicked	
clicked_last_time_at	Last time when a message was clicked (can be equal to clicked_first_time_at, if the message was clicked only once)	
is_unsubscribed	Boolean flag if a recipient clicked unsubscribe link in a message	outcome variables
unsubscribed_at	Datetime when a recipient clicked unsubscribe link in a message	
is_hard_bounced	Whether the message was hard bounced	

Data Preparation And Variables

Column Name	Description	Variables Type
is_soft_bounced	Whether the message was soft bounced	
soft_bounced_at	Datetime when a message was "soft bounced"	
is_complained	Boolean flag if a recipient clicked SPAM button in email client	outcome variables
complained_at	Datetime when the message has been complained	
is_blocked	Boolean flag if a delivery attempt was temporarily blocked by email provider	outcome variables
blocked_at	Datetime when a delivery attempt was temporarily blocked by email provider	
is_purchased	Boolean flag if a recipient clicked any link in a message, opened a website or mobile app and made a purchase	outcome variables
purchased_at	Datetime when a recipient made a purchase after click on email or other message	
created_at	Datetime when the message is created *will not be used	
updated_at	Datetime when the message is updated *will not be used	

Data Cleaning

Table: Data description for ***omitted variables*** *in campaign dataset*

Column Name	Description	Action for values	Action for column
campaign_type	Campaign type (bulk, trigger, transactional)	Only retain values for bulk-type message	deleted
ab_test	Bulk campaign with A/B test mode *will not be used	Deleted messages used for ab_test	deleted
warmup_mode	Bulk campaign with warmup mode	Deleted messages used for warm_up mode	deleted
hour_limit	Hour limit for a bulk campaign with warmup mode		deleted
is_test	Whether it's a test campaign (bulk campaigns only)	Deleted messages used for test	deleted
subject_with_emoji; subject_with_bonus; subject_with_saleout	Whether the sent messages contains emoji/ bonus info/ sale-out info	/	deleted

```
print(df['subject_with_emoji'].value_counts()
print(df['subject_with_bonuses'].value_counts
subject_with_emoji
True     1220393
Name: count, dtype: int64
subject_with_bonuses
False     1220393
Name: count, dtype: int64
```

Data Cleaning

Main process	Coding part	Reason
1 Random sampling	`message = df2[df2.message_type=='bulk'].sample(frac=0.2, random_state=123)` `message`	The original dataset was too large to process

2 Merge datasets	Foreign key
	`df = pd.merge(message, df1, on='campaign_id', how='left')`
	`df`
		To combine IVs and DVs together

3 Delete the omitted columns	`df = df.drop(['client_id','started_at','finished_at','campaign_type','topic'],axis=1)`
	`## Only one type of value in these columns so drop these columns (cannot give us any insight in the analysis)`
	`df = df.drop(['subject_with_emoji','subject_with_bonuses','subject_with_saleout','message_type','stream'],axis=1)`
		To simplify the result table

4 Dummy coding	`## Transform "channel" into dummy variable "is_email"`
	`df.channel = df.channel.replace({'mobile_push':0,'email':1})`
	`df = df.rename(columns={'channel':'is_email'})`
		To take the effect of categorical variables into account

03

Descriptive Analysis

Correlation Analysis
Clustering Analysis

Relationship between IVs and different DVs

For positive actions, the differences are mainly from three IVs

In [24]:
```
var = ['is_email','total_count', 'subject_length', 'subject_with_personalization','subject_with_deadline', 'subject_with_discount']

df.groupby("is_opened")[var].mean()
```

Out[24]:

is_opened	is_email	total_count	subject_length	subject_with_personalization	subject_with_deadline	subject_with_discount
0.0	0.343609	712449.352848	107.488693	0.000827	0.007246	0.000827
1.0	0.437165	690666.824807	113.332783	0.002505	0.008472	0.002505

In [25]:
```
df.groupby("is_clicked")[var].mean()
```

Out[25]:

is_clicked	is_email	total_count	subject_length	subject_with_personalization	subject_with_deadline	subject_with_discount
0.0	0.349711	709791.005215	108.075805	0.001012	0.007505	0.001012
1.0	0.957085	675335.098081	128.218327	0.005382	0.000000	0.005382

In [26]:
```
df.groupby("is_purchased")[var].mean()
```

Out[26]:

is_purchased	is_email	total_count	subject_length	subject_with_personalization	subject_with_deadline	subject_with_discount
0.0	0.356559	709407.896373	108.303385	0.001061	0.00742	0.001061
1.0	1.000000	653827.441520	127.883041	0.005848	0.00000	0.005848

Relationship between IVs and different DVs

For negative actions, there are some initial insights…

In [27]:
```
df.groupby("is_unsubscribed")[var].mean()
```
Out[27]:

is_unsubscribed	is_email	total_count	subject_length	subject_with_personalization	subject_with_deadline	subject_with_discount
0.0	0.372774	700049.163317	108.752872	0.001114	0.007712	0.001114
1.0	0.026409	901874.492460	99.161834	0.000018	0.001363	0.000018

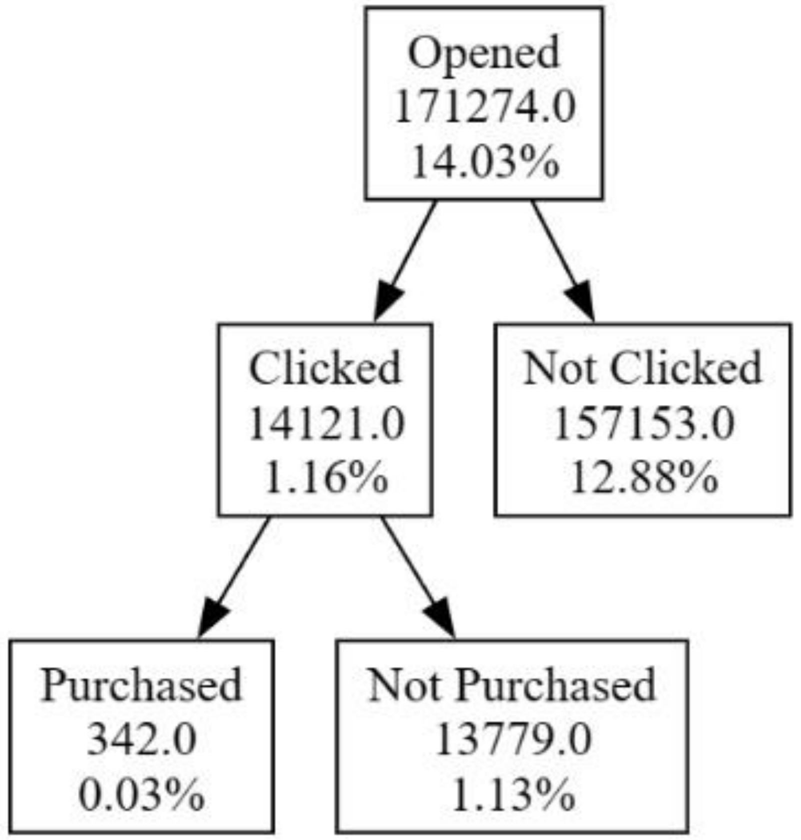
In [28]:
```
df.groupby("is_complained")[var].mean()
```
Out[28]:

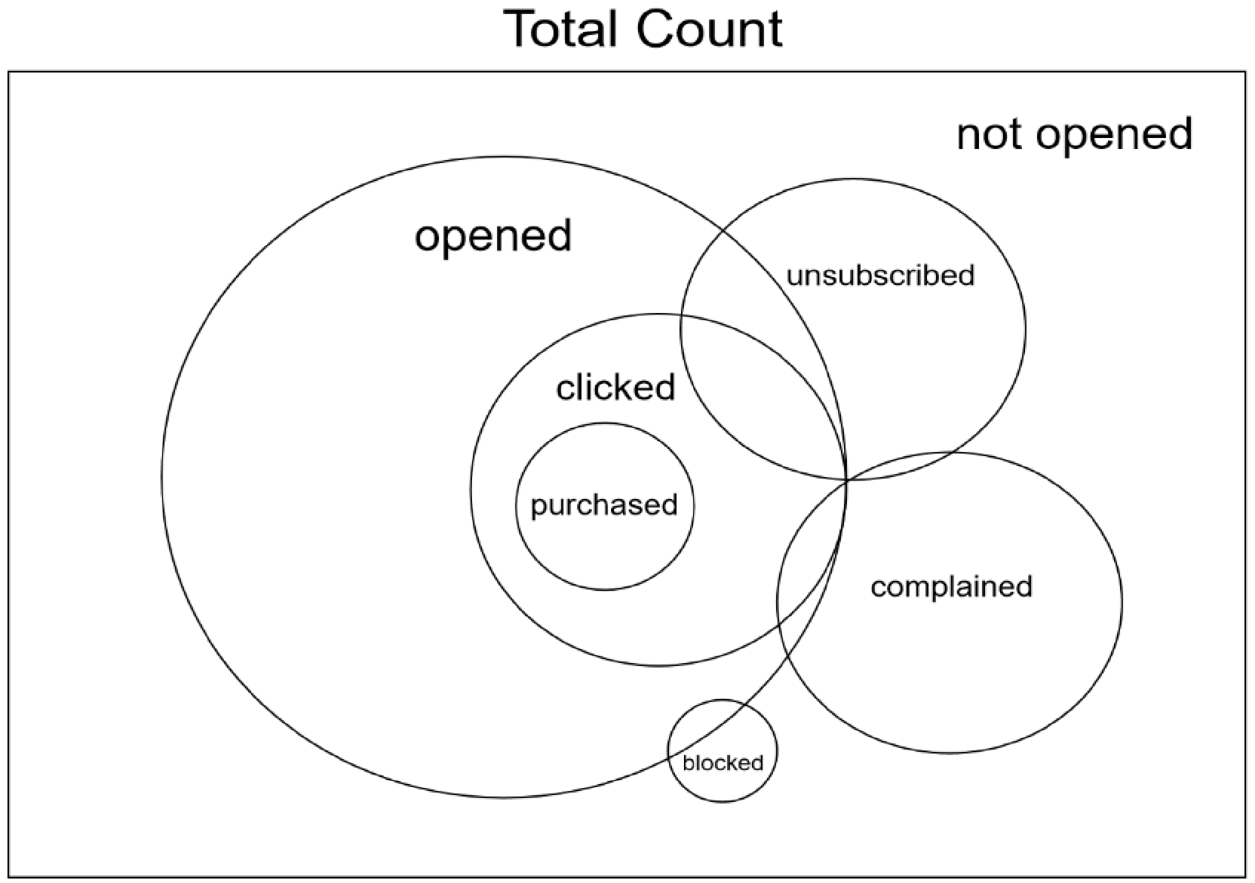
is_complained	is_email	total_count	subject_length	subject_with_personalization	subject_with_deadline	subject_with_discount
0.0	0.356548	709424.959067	108.304796	0.001063	0.00742	0.001063
1.0	1.000000	599392.618785	122.044199	0.000000	0.00000	0.000000

In [29]:
```
df.groupby("is_blocked")[var].mean()
```
Out[29]:

is_blocked	is_email	total_count	subject_length	subject_with_personalization	subject_with_deadline	subject_with_discount
0.0	0.356737	709393.679753	108.308847	0.001063	0.007418	0.001063
1.0	1.000000	377663.200000	114.400000	0.000000	0.000000	0.000000

Relationship between DVs

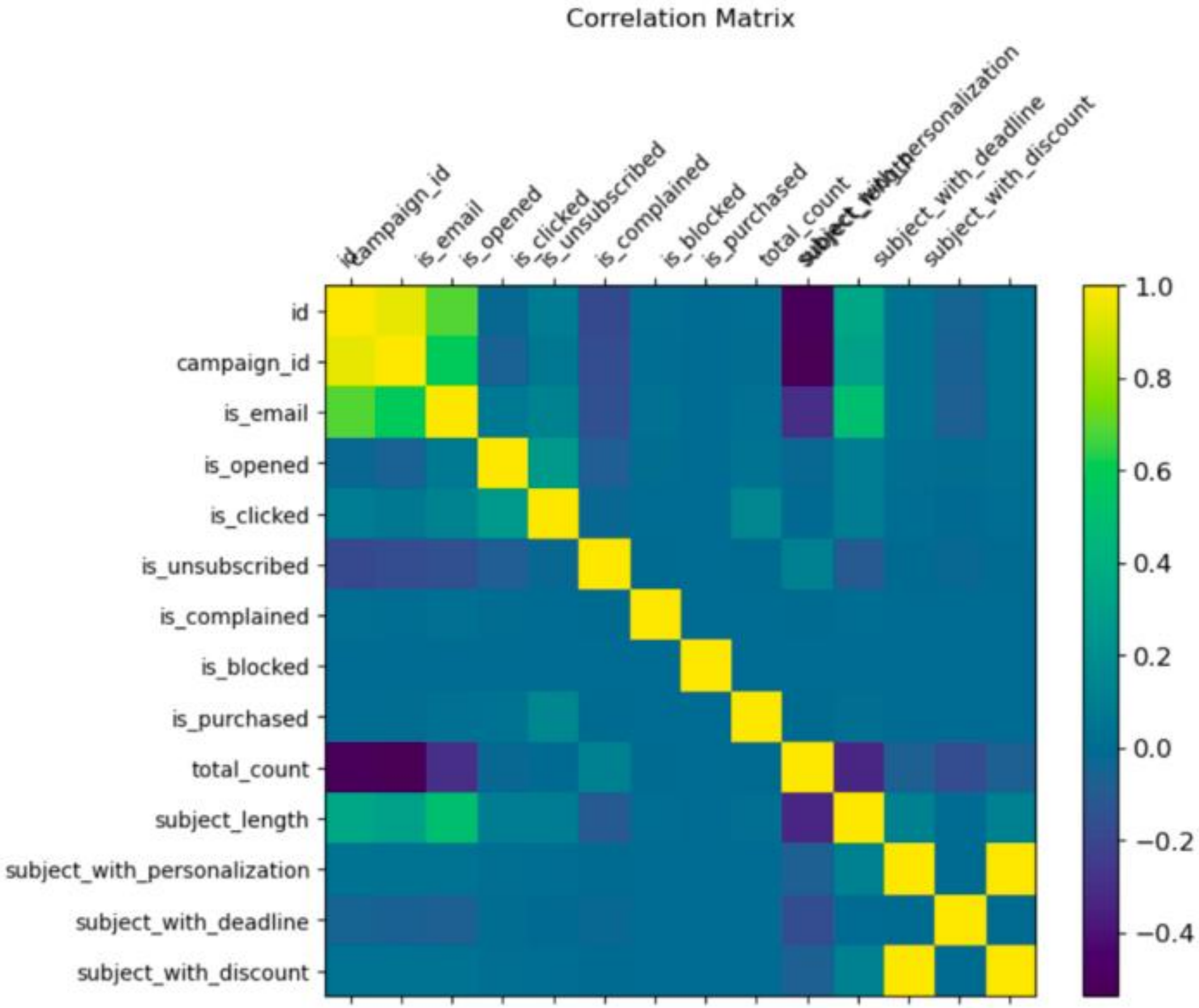




- Conversion chain:
  Open → Click → Purchase
- There is no clear relationship between negative actions of consumers
- Most of the people who send negative feedback do not open the message

- We will mainly focus on the effect of IVs on Open, Click, and Purchase behavior, especially the **Purchase action** which has the lowest conversion rate.

Correlation Matrix



- **“subject_with_personalization” and “subject_with_discount” are highly correlated,** so we deleted one of them
- **"total_count"** means the number of receivers of the message, this is **meaningless in the real world** since receiver has no idea about the total count. So we decided to deduct this IV

04

In-Depth Analysis and Insights

Regression Analysis

Regression Analysis
-- includes 3 stages

- There is a clear conversion chain among the relationship between: Open, Click, and Purchase
- It is more reasonable to test the Click action among all the users who have opened the message (is_open=1); and test the Purchase action among all clicked users (is_click=1)

- All the **DVs are dummy variables**

Logistic Regression Model

Conversion chain: **Open → Click → Purchase**

Stage	IV	Sample Set
Stage 1	is_opened	The whole sample set
Stage 2	is_clicked	The opened dataset (is_open = 1)
Stage 3	is_purchased	The click dataset (is_click = 1)

Logistic Regression

Stage 1: Open

```
Optimization terminated successfully.
         Current function value: 0.400384
         Iterations 6
                           Logit Regression Results
==============================================================================
Dep. Variable:              is_opened   No. Observations:              1220393
Model:                          Logit   Df Residuals:                  1220388
Method:                           MLE   Df Model:                            4
Date:                Tue, 16 Apr 2024   Pseudo R-squ.:                 0.01283
Time:                        19:06:47   Log-Likelihood:            -4.8863e+05
converged:                       True   LL-Null:                   -4.9497e+05
Covariance Type:            nonrobust   LLR p-value:                     0.000
================================================================================================
                                coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------------------------
Intercept                    -3.2245      0.015   -209.049      0.000      -3.255      -3.194
subject_length                0.0122      0.000     85.105      0.000       0.012       0.013
is_email                      0.1560      0.006     26.367      0.000       0.144       0.168
subject_with_personalization  0.1030      0.060      1.721      0.085      -0.014       0.220
subject_with_deadline         0.2363      0.029      8.187      0.000       0.180       0.293
================================================================================================
```

- **DV:** "is_opened"
- **IV:** "subject_length", "is_email" and "subject_with_deadline"
- **p<0.05 and positive coef:**
  - All three DVs have **significant positive relationship** with the likelihood of recipients opening email
- **Interpretation:**
  - In the form of an email, increasing the length of the message and adding an deadline can encourage more users to open the message.

Logistic Regression

Stage 2: Click

Logit Regression Results

Dep. Variable:	is_clicked	No. Observations:	171274
Model:	Logit	Df Residuals:	171269
Method:	MLE	Df Model:	4
Date:	Tue, 16 Apr 2024	Pseudo R-squ.:	0.2172
Time:	19:06:50	Log-Likelihood:	-38169.
converged:	False	LL-Null:	-48763.
Covariance Type:	nonrobust	LLR p-value:	0.000

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-9.0473	0.115	-78.714	0.000	-9.273	-8.822
subject_length	0.0358	0.001	40.632	0.000	0.034	0.038
is_email	3.0050	0.042	71.020	0.000	2.922	3.088
subject_with_personalization	-2.1123	0.136	-15.476	0.000	-2.380	-1.845
subject_with_deadline	-11.5545	111.164	-0.104	0.917	-229.431	206.322

- **DV:** "is_opened"
- **IV:** "subject_length", "is_email" , "subject_with_personalization"
- **$p<0.05$ and positive coef:** "subject_length" and "is_email" have **significant positive relationship** with the likelihood of recipients opening email
- **$p>0.05$ and positive coef:** "**subject_with_personalization**" has a **significant negative** correlation with "is_clicked".
- **Interpretation:**
  - Using **email** and increasing the **length** of the message can **encourage** more users to click on the message links.
  - However, messages with **personalization** will **reduce** users' clicks on the message links.

Logistic Regression

Stage 3: Purchase

Due to multicollinearity we removed "subject_with_deadline"

Logit Regression Results

Dep. Variable:	is_purchased	No. Observations:	14121
Model:	Logit	Df Residuals:	14117
Method:	MLE	Df Model:	3
Date:	Tue, 16 Apr 2024	Pseudo R-squ.:	0.01237
Time:	19:06:51	Log-Likelihood:	-1590.4
converged:	False	LL-Null:	-1610.3
Covariance Type:	nonrobust	LLR p-value:	1.159e-08

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-19.7527	1761.039	-0.011	0.991	-3471.325	3431.819
subject_length	$\boxed{-0.0187}$	0.006	-3.179	$\boxed{0.001}$	-0.030	-0.007
is_email	18.4983	1761.039	0.011	0.992	-3433.074	3470.070
subject_with_personalization	1.0994	0.794	1.385	0.166	-0.457	2.655

- **DV:** "is_purchased"
- **IV:** "subject_length"
- **p>0.05 and negative coef:**
  - The result shows that "**subject_length**" has a **significant negative** correlation with "is_purchased".
- **Interpretation:** In the case where the user has opened and clicked the link, too long messages reduce user purchase.

Logistic Regression

Brief Summary and Interpretation

	Stage 1	Stage 2	Stage 3
"is_email"	open (+)	click (+)	
"subject_length"	open (+)	click (+)	purchase (−)
"subject_with_deadline"	open (+)		
"subject_with_personalization"		click (−)	

Email is an <u>effective way</u> to increase opens and clicks

Messages with **deadline** are more <u>attractive</u> to receivers to <u>open</u>.

- **Long message** is more <u>attractive</u> to receivers to <u>open and click</u>.
- But too long message will also <u>reduce</u> users' desire to <u>buy</u>.
- Balancing the length of the message is an issue that requires careful consideration.

Messages with **personalization** will <u>reduce users' clicks</u> on the message links after they open it.

Why ?

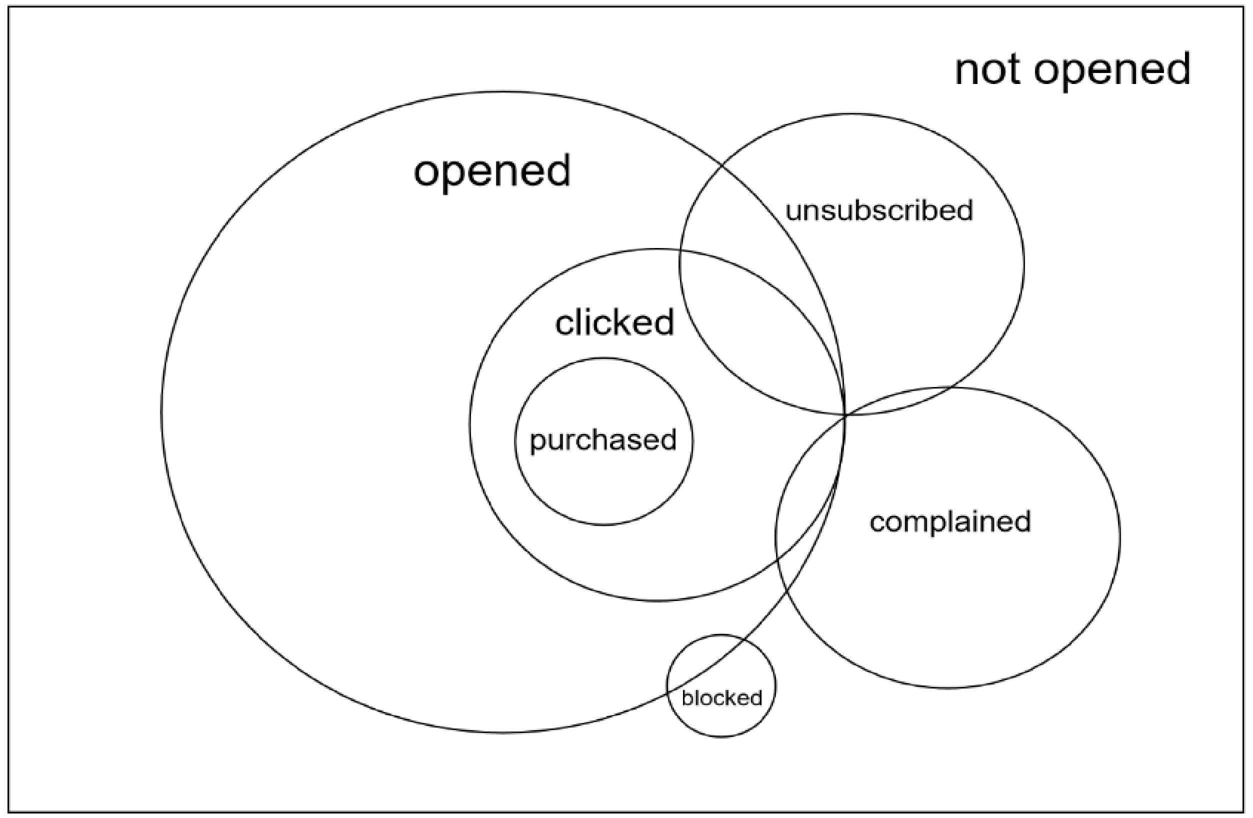
Worry about information security

DON'T miss it!

Other DVs (negative feedback)

DVs	The Whole Dataset	The Open Dataset	The Click Dataset
Unsubscribe	**Significant negative correlation**	**Significant negative correlation**	No significant relationships
Complained	No significant relationships	No significant relationships	No significant relationships
Block	No significant relationships	No significant relationships	No overlap

Total Count



Logit Regression on Other DVs (negative feedback)

Unsubscribe action

(In the entire sample set)

```
                           Logit Regression Results
==============================================================================
Dep. Variable:        is_unsubscribed   No. Observations:            1220393
Model:                          Logit   Df Residuals:                1220388
Method:                           MLE   Df Model:                          4
Date:                Tue, 16 Apr 2024   Pseudo R-squ.:               0.08868
Time:                        19:07:02   Log-Likelihood:          -2.0848e+05
converged:                       True   LL-Null:                 -2.2877e+05
Covariance Type:            nonrobust   LLR p-value:                   0.000
==================================================================================================
                                 coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------------------------------
Intercept                     -2.0329      0.022    -91.324      0.000      -2.077      -1.989
subject_length                -0.0054      0.000    -24.600      0.000      -0.006      -0.005
is_email                      -2.9767      0.027   -111.130      0.000      -3.029      -2.924
subject_with_personalization  -1.1516      1.001     -1.151      0.250      -3.113       0.810
subject_with_deadline         -2.1330      0.115    -18.619      0.000      -2.357      -1.908
==================================================================================================
```

- The result shows that "**subject_length**", "**is_email**" and "**subject_with_deadline**" have a significant negative correlation with "is_unsubscribed".
- In the form of an email, increasing the length of the message and adding an deadline can reduce the probability of being unsubscribed.

(In the opened sample set)

```
                           Logit Regression Results
==============================================================================
Dep. Variable:        is_unsubscribed   No. Observations:             171274
Model:                          Logit   Df Residuals:                 171269
Method:                           MLE   Df Model:                          4
Date:                Tue, 16 Apr 2024   Pseudo R-squ.:                0.1438
Time:                        19:07:04   Log-Likelihood:              -5340.7
converged:                      False   LL-Null:                     -6237.8
Covariance Type:            nonrobust   LLR p-value:                   0.000
==================================================================================================
                                 coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------------------------------
Intercept                    -25.8350   5633.374     -0.005      0.996   -1.11e+04    1.1e+04
subject_length                -0.0291      0.003     -9.772      0.000      -0.035      -0.023
is_email                      25.1664   5633.374      0.004      0.996    -1.1e+04   1.11e+04
subject_with_personalization  -0.0029      1.019     -0.003      0.998      -2.000       1.994
subject_with_deadline          0.3295   4.47e+04   7.37e-06      1.000   -8.76e+04   8.76e+04
==================================================================================================
```

- The result shows that "**subject_length**" has a significant negative correlation with "is_unsubscribed".
- In the case where the user has already opened the message, <u>the longer the length</u> of the message, <u>the less likely it will be unsubscribed</u>.

Why? A higher number of words may correspond to more detailed content.

DV: is_unsubscribed

Sample set	Result	Interpretation
Entire Sample Set	The result shows that "subject_length", "is_email" and "subject_with_deadline" have a significant negative correlation with "is_unsubscribed".	In the form of an email, increasing the length of the message and adding an deadline can reduce the probability of being unsubscribed. This may be because messages with rich content are more difficult to view as spam and therefore unsubscribed.
Opened Sample Set	The result shows that "subject_length" has a significant negative correlation with "is_unsubscribed".	In the case where the user has already opened the message, the longer the length of the message, the less likely it will be unsubscribed. The reasons may be similar to the above. A higher number of words may correspond to more detailed content. Users find it valuable after opening and reading, so they are less likely to unsubscribe.
Clicked Sample Set	The result indicates that no significant relationship exists.	-

DV: is_complained

DV	Result	Interpretation
Entire Sample Set	The result indicates that no significant relationship exists.	-
Opened Sample Set	The result indicates that no significant relationship exists.	-
Clicked Sample Set	The result indicates that no significant relationship exists.	-

DV: is_blocked

DV	Result	Interpretation
Entire Sample Set	The result indicates that no significant relationship exists.	-
Opened Sample Set	The result indicates that no significant relationship exists.	-
Clicked Sample Set	Because there is no overlap between click and block, we did not analyze it in the clicked sample set.	-

DV: “is_unsubscribed”, “is_complained”, “is_blocked”
IV: “subject_length”, “subject_with_personalization”, “subject_with_deadline”

- Analysis of negative feedback has yielded limited results.
- Except for unsubscription, other conditions (complaints and blocking) have no significant relationship with the characteristics we studied.
- Form of an email: increasing the length of the message and adding an deadline can reduce the probability of being unsubscribed.
- User has already opened the message: the longer the length of the message, the less likely it will be unsubscribed.
    - This reflects that message length also remains an important factor in negative feedback.

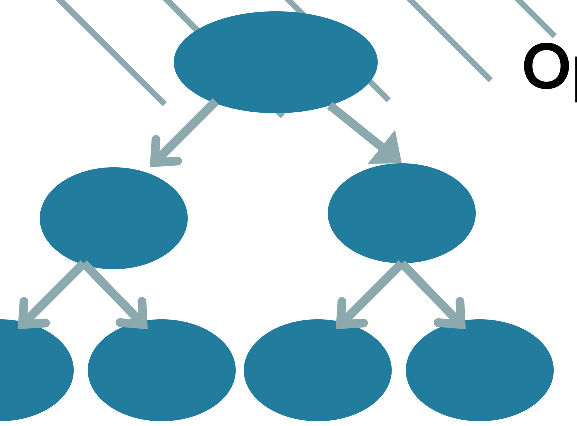
05

Business Recommendations

Key insights and strategic actions
Future directions for research / improvement

Key Insights and Strategic Actions

Research Findings



Open → Click → Purchase

- No clear relationship between negative actions of consumers
- Most of the people who send negative feedback do not open the message

Open: Increase message length and add a deadline
Click (msg links): Increase message length, but avoid personalization
Purchase: for users who have opened and clicked the link --> avoid long messages

(new model)

Marketing Problem

Relationship between consumer's different behaviors

The effects of different campaign message's characteristics on consumer's behaviors

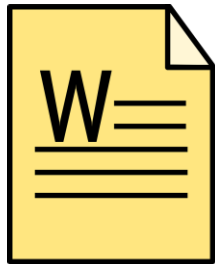
How to adjust the characteristics to optimize the effect brought by the campaign message

A way to predict the effect of future campaign message on consumer's behaviors

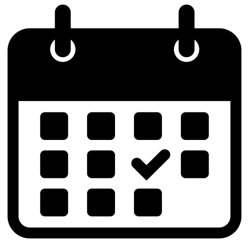
Key Insights and Strategic Actions

Business Recommendation

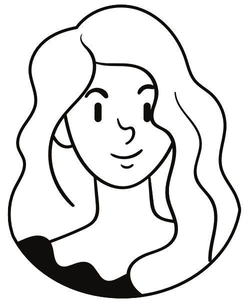
Email → Use email instead of SMS



Length → Be detailed but not too long



Deadline → Appropriate display event time



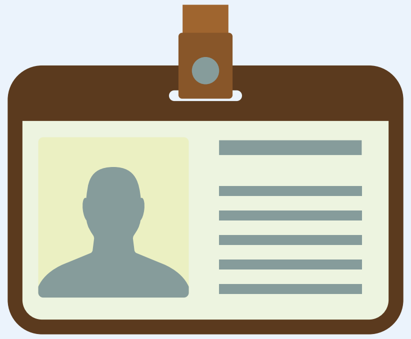
Personalize → Reduce personalization that may raise information security concerns

Internal personalization ✅

Push based on region and consumption habits

External personalization ❎

Personalization that the recipient can perceive, such as real name



Tuesday, April 23, 2024 at 17:21:08 China Standard Time

Subject: Looking for a Chinese tattoo? Why get branded like a Chinese outcast? Get a necklace instead! - I would like to get a tattoo of my late husband's l...
Date: Sunday, 21 April 2024 at 16:13:58 China Standard Time
From: Chinese Characters 漢字
To: xxxxxxxx@gmail.com

Chinese Characters 漢字 · 27.6K followers
A space to share stuff about Chinese characters and Chinese languages

Looking for a Chinese tattoo? Why get branded like a Chinese outcast? Get a necklace instead!

Robert Matthews (馬學進), Univ. retired multilingual Eng/Fre/Chi as For. Lang teacher (40+ yrs experience)
Posted Apr 6

Discreet and respectful, without negative connotations. A great suggestion by Lynne LiKeywords for search engines: Chinese {character, name, surname} + {necklace, pendant}

I would like to get a tattoo of my late husband's last name, in Hanzi. Would that seem disrespectful? I am not Chinese. What would be the best character(s) for Wong?

Thanks for asking.

Well, in the Chinese culture, people rarely get a tattoo featuring a surname. In my humble opinion, you could choose to wear a pendant featuring the Chines...

Read more »
27 upvotes 2 comments

7 upvotes

Example

- ✅ Use email instead of SMS
- ✅ Detailed but not too long
- ✅ Appropriate display of event time
- ✅ Reduce personalization

How difficult is ... characters?

Fluorite, kn...
Answered Tu...

We mostly use Arab... statistics work.And we use a simplified system of characters in some formal reports,but that's not...

Read more »
2 upvotes 3 comments

How does the written Chinese language deal with novel concepts? Can new characters be invented, or do they combine existing characters to express new ideas?

平岡 行雄, former I used to be a high school language teacher
Answered April 10

Chinese characters are used in Japan as well as in China, but there are slight differences in their daily usage.

As an ordinary Japanese, these small differences are interes...

Interested to get one of your own necklaces? Click the link below to purchase for half price
before 30 April!

Read more in Chinese Characters 漢字

Future Directions for Research / Improvement

Research Limitations

1. Low level of fit

Pseudo R-squ. : 0.2172

- Can only explain a small part of the variation
- Reason: small number of IVs/Unsuitable model

2. Further research for "Personalization"

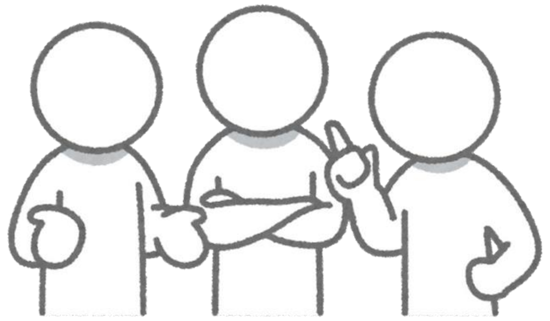
The reason for such phenomenon

- Variables not considered/Interactive effect

Future actions to take

- Add variables or introduce interactions to improve our model

3. Limited findings about negative feedbacks

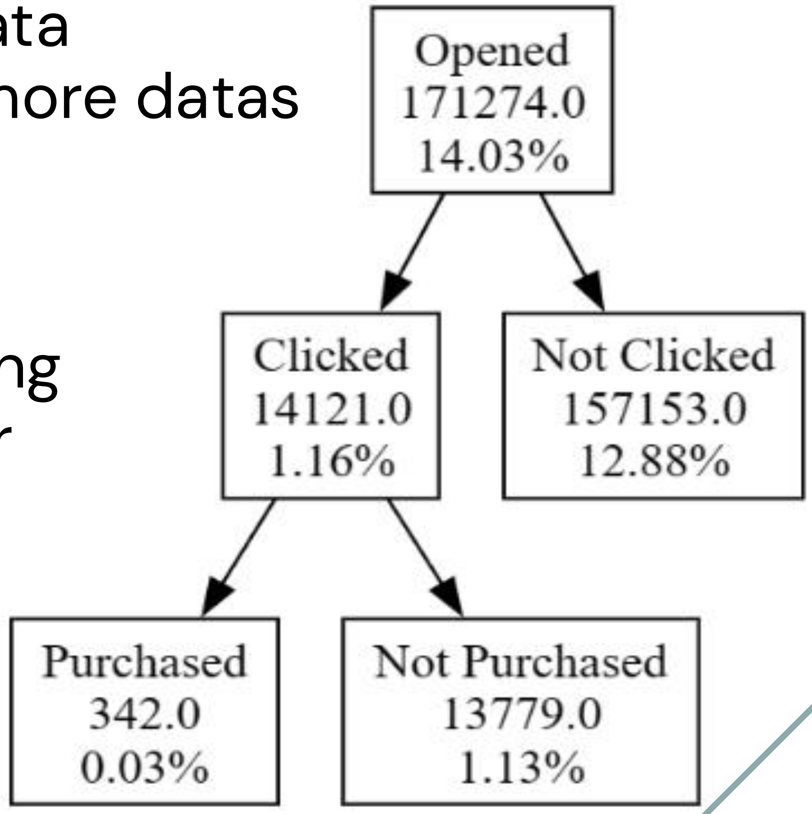


No significant relationships

- lack of more relevant data
- Solution: try to collect more datas

4. Further research to optimize buying process

- Since our decision tree showed 1.16% of consumers clicking and 0.03% of them purchasing, we suggest urgent further research in optimizing methods to improve these performances



THANK YOU

NOW, QNA TIME!