



팬심M 고객 세분화 프로젝트

AI_14_최유진

AI_14_오동호

▶ 팀 역할 및 구성도

훈련생	역할	담당 업무
최유진	팀장	<ul style="list-style-type: none">▶ 데이터 전처리 및 분석▶ 가설 검정
오동호	팀원	<ul style="list-style-type: none">▶ 데이터 전처리▶ 데이터 시각화(Tableau)

CONTENTS

●	●	●	●
Part_1	Part_2	Part_3	Part_4
프로젝트 개요	데이터 소개	데이터 분석 및 전처리	가설 검증
●	●	●	●
Part_5	Part_6	Part_7	Part_8
RFM 분석 및 적용	데이터 시각화	인사이트 도출	마무리

프로젝트 개요

01. 프로젝트 개요

- 프로젝트 목적

마케팅을 보다 효과적으로 할 순 없을까?



01. 프로젝트 개요

구매 행동을 기반으로 고객을 분류하는

효과적인 맞춤 마케팅

프로모션 효과가 클 것 같은 고객을 타겟팅 하고싶은데..

VIP 고객들을 어떻게 파악하지?

새로운 서비스를 구매력이 좋은 고객에게 먼저?



Customer Segmentation

고객의 행동 유형을 기반으로 고객을 분류하는 것

고객 개인의 특성에 맞춘 세밀한 마케팅으로 효과를 증대시키고 매출을 극대화하는 개인화 마케팅에 활용

RFM Analysis

CRM 마케팅 분야에서 가장 널리 사용되는 고객 세분화 모델 RFM

Recency (최근 구매 여부): 고객이 얼마나 최근에 구매했는가?

Frequency (구매 빈도): 고객이 얼마나 자주 방문했는가?

Monetary (구매 금액): 고객이 돈을 얼마나 썼는가?

Analysis Method

분석 방법

- 가설 검증
- RFM 세분화
- 시계열 분석을 통해 주차 별 서비스 사용량 파악 및 시각화
- 등급 별 필요한 action item or 등급을 상승시키기 위한 action item 도출

데이터 : 

데이터 소개

02.데이터 소개

▶ 셀럽 활동 데이터



channel_url_x

- 메시지가 전송된 채팅방 url 입니다.



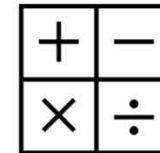
id_type

- 팬과 셀럽의 구분입니다.



message

- 전송된 메시지 내용입니다.



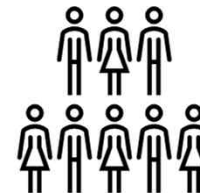
Message_count

- sender_id 기준
총 메시지의 수 입니다..



created_at

- 메시지 전송 시간입니다.



트위치, 아프리카, 유튜브 수

-각 채널 팔로워의 수 입니다.

02.데이터 소개

▶ 셀럽 활동 데이터



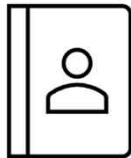
채널주인

- 본인 채팅방 url 입니다.



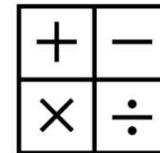
sender_id

- 메시지 전송자의 이메일 입니다.



sender_nickname

- 메시지 전송자의 nickname 입니다.



channel_name

- 현재 채팅방 name 입니다.

데이터 분석 및 전처리

03.데이터 분석 및 전처리

▶ 데이터 프레임

	channel_url_x	type	message	created_at	id type	channel_member_count_x	message_count	channel_created_at	registration date	total_shot	트위치	아프리카 (애청자/팬클럽)	유튜브 수	메인 플랫폼 링크	콘텐츠 카테고리	메인 평균 방송 시간 (9개)/ 업로드 주기	평균 시청 횟수/ 최근 조회수	커뮤니티 링크
0	sendbird_group_channel_38649690_0c2f6fa0c6fd20...	MESG	으어~ 오늘도 재 밧영ㅎ습니다	2022-02-27 01:03:05	celeb	135.0	426.0	1.623506e+09	2021-06-12 13:53:34	4610.0	12000.0	0	4620	https://www.twitch.tv/witching_	게임/ 토크	2H	600회	https://tgd.kr/s/witching_
1	sendbird_group_channel_38649690_0c2f6fa0c6fd20...	MESG	@@	2022-02-27 01:03:06	celeb	135.0	426.0	1.623506e+09	2021-06-12 13:53:34	4610.0	12000.0	0	4620	https://www.twitch.tv/witching_	게임/ 토크	2H	600회	https://tgd.kr/s/witching_
2	sendbird_group_channel_38649690_0c2f6fa0c6fd20...	MESG	요즘 계속 업뎃 중이라 토크이 자꾸 만료되네 요ㅋㅋ	2022-03-01 13:43:28	celeb	135.0	426.0	1.623506e+09	2021-06-12 13:53:34	4610.0	12000.0	0	4620	https://www.twitch.tv/witching_	게임/ 토크	2H	600회	https://tgd.kr/s/witching_
3	sendbird_group_channel_38649690_0c2f6fa0c6fd20...	MESG	ㅋㅋㅋㅋ	2022-03-01 22:03:24	celeb	135.0	426.0	1.623506e+09	2021-06-12 13:53:34	4610.0	12000.0	0	4620	https://www.twitch.tv/witching_	게임/ 토크	2H	600회	https://tgd.kr/s/witching_
4	sendbird_group_channel_38649690_0c2f6fa0c6fd20...	MESG	img 1646136816463	2022-03-01 22:03:32	celeb	135.0	426.0	1.623506e+09	2021-06-12 13:53:34	4610.0	12000.0	0	4620	https://www.twitch.tv/witching_	게임/ 토크	2H	600회	https://tgd.kr/s/witching_

03.데이터 분석 및 전처리

▶ 데이터 프레임

커뮤니티 링크	성별	열 공개 여부	채널주인				read_receipt_secured
			channel_name_secured	sender_nickname_secured	sender_id_secured		
https://tgd.kr/s/witching_	여자	O	sendbird_group_channel_38649690_0c2f6fa0c6fd20...	김위*	위칭	wit****g_@dau*****	[pdk****ol@gma***** hye*****27@gma***** ho...
https://tgd.kr/s/witching_	여자	O	sendbird_group_channel_38649690_0c2f6fa0c6fd20...	김위*	위칭	wit****g_@dau*****	[pdk****ol@gma***** hye*****27@gma***** ho...
https://tgd.kr/s/witching_	여자	O	sendbird_group_channel_38649690_0c2f6fa0c6fd20...	김위*	위칭	wit****g_@dau*****	[pdk****ol@gma***** hye*****27@gma***** ho...
https://tgd.kr/s/witching_	여자	O	sendbird_group_channel_38649690_0c2f6fa0c6fd20...	김위*	위칭	wit****g_@dau*****	[pdk****ol@gma***** hye*****27@gma***** ho...
https://tgd.kr/s/witching_	여자	O	sendbird_group_channel_38649690_0c2f6fa0c6fd20...	김위*	위칭	wit****g_@dau*****	[pdk****ol@gma***** hye*****27@gma***** ho...

03.데이터 분석 및 전처리

▶ 데이터 프레임 정보

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 152285 entries, 0 to 541749
Data columns (total 25 columns):
#   Column                Non-Null Count  Dtype
---  -
0   channel_url_x          152285 non-null object
1   type                   152285 non-null object
2   message                152285 non-null object
3   created_at             152285 non-null object
4   id type                152285 non-null object
5   channel_member_count_x 152285 non-null float64
6   message_count          152285 non-null float64
7   channel_created_at     152285 non-null float64
8   registration date      152285 non-null object
9   total_shot             152285 non-null float64
10  트위치                 144615 non-null float64
11  아프리카(애청자/팬클럽) 144615 non-null object
12  유튜브 수              144615 non-null object
13  메인 플랫폼 링크       144615 non-null object
```

```
14  콘텐츠 카테고리        144615 non-null object
15  메인 평균 방송시간(9개)/업로드주기 144615 non-null object
16  평균 시청 횟수/최근 조회수 144615 non-null object
17  커뮤니티 링크          144615 non-null object
18  성별                   144615 non-null object
19  얼굴 공개 여부         144615 non-null object
20  채널주인               151319 non-null object
21  channel_name_secured    152285 non-null object
22  sender_nickname_secured 152285 non-null object
23  sender_id_secured       152285 non-null object
24  read_receipt_secured    152285 non-null object
dtypes: float64(5), object(20)
memory usage: 30.2+ MB
```

- 수가 다른 데이터가 존재(붉은 박스)
- 결측치가 있을 것으로 보여 추후 확인 후 제거
- 데이터 자료형을 바꿔주어야 하는 컬럼 확인(푸른 박스)

03.데이터 분석 및 전처리

▶ 데이터 프레임 정보

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 152285 entries, 0 to 541749
Data columns (total 25 columns):
#   Column                Non-Null Count  Dtype
---  -
0   channel_url_x          152285 non-null object
1   type                   152285 non-null object
2   message                152285 non-null object
3   created_at             152285 non-null object
4   id type                152285 non-null object
5   channel_member_count_x 152285 non-null float64
6   message_count           152285 non-null float64
7   channel_created_at      152285 non-null float64
8   registration date       152285 non-null object
9   total_shot             152285 non-null float64
10  트위치                 144615 non-null float64
11  아프리카(애청자/팬클럽) 144615 non-null object
12  유튜브 수              144615 non-null object
13  메인 플랫폼 링크        144615 non-null object
```

```
14  콘텐츠 카테고리          144615 non-null object
15  메인 평균 방송시간(9개)/업로드주기 144615 non-null object
16  평균 시청 횟수/최근 조회수 144615 non-null object
17  커뮤니티 링크            144615 non-null object
18  성별                     144615 non-null object
19  얼굴 공개 여부           144615 non-null object
20  채널주인                 151319 non-null object
21  channel_name_secured      152285 non-null object
22  sender_nickname_secured   152285 non-null object
23  sender_id_secured         152285 non-null object
24  read_receipt_secured      152285 non-null object
dtypes: float64(5), object(20)
memory usage: 30.2+ MB
```

- 수가 다른 데이터가 존재(붉은 박스)
- 결측치가 있을 것으로 보여 추후 확인 후 제거
- 데이터 자료형을 바꿔주어야 하는 컬럼 확인(푸른 박스)

03.데이터 분석 및 전처리

▶ 기준(index) Feature 설정

sender_nickname_secured	sender_id_secured
유메*	ezu***ll@nav*****
정다**	jda***ol@gma*****
반달*	ban*****15@gma*****
정다**	jda***ol@gma*****
후_*	np242@kor*****

마스킹 처리가 되어도 중복이 발생하지 않는
sender_id_secured를 기준 id로 활용

03.데이터 분석 및 전처리

▶ 컬럼 추가(own_channel)

```
celeb_df.insert(21, 'own_channel', celeb_df['channel_url_x'] == celeb_df['채널주인'])
```

own_channel	channel_name_secured	sender_nickname_secured	sender_id_secured	read_receipt_secured
True	김선*	김선*	wng****33@nav*****	[pdk****ol@gma*****, hye*****27@gma*****, ho...
True	정다**	정다**	jda***ol@gma*****	[pdk****ol@gma*****, hye*****27@gma*****, ho...
True	김재	김재	kim***e0@gma*****	[pdk****ol@gma*****, hye*****27@gma*****, ho...
True	짱스	짱스	yoo****36@nav*****	[pdk****ol@gma*****, hye*****27@gma*****, ho...
False	소누*****	유완*	yas*****to@nav*****	[pdk****ol@gma*****, hye*****27@gma*****, ho...

- 'channel_url_x' 컬럼과 '채널주인' 값이 일치할 경우 본인의 채팅방에서 대화하고 있는 셀럽 해당 정보를 'own_channel'(자신의 채널) 컬럼 추가를 통해 구분 가능하도록 하였음.
- True(자신의 채팅방) - False(다른 셀럽의 채팅방)

03.데이터 분석 및 전처리

▶ 셀럽 간 대화 데이터 제거

```
len(celeb_df.loc[celeb_df['own_channel'] == True])
```

147101 ← 자신의 채팅방에서 대화

```
len(celeb_df.loc[celeb_df['own_channel'] == False])
```

5184 ← 타 셀럽의 채팅방에서 대화

```
print(float(5184/152285))
```

0.034041435466395244

타 셀럽의 채팅방에서 행한 채팅은 전체의
약 **3.4%**

- 팬심은 팬과 소통하는 것이 핵심인 플랫폼이기에 셀럽끼리 대화한 데이터는 제거

03.데이터 분석 및 전처리

▶ 제거 후 셀럽 수 확인

```
celeb_df['sender_id_secured'].nunique()
```

421



```
celeb_df['sender_id_secured'].nunique()
```

225

셀럽 간 대화 데이터 제거 후 데이터 내 셀럽의 수가
421명에서 **225명**으로 감소

03.데이터 분석 및 전처리

▶ 결측치 확인

```
celeb_df['아프리카(애청자/팬클럽)'].isna().sum()
```

7070

```
celeb_df['유튜브 수'].isna().sum()
```

7070

```
celeb_df['트위치'].isna().sum()
```

7070

t	트위치	아프리카 (애청자/ 팬클럽)	유튜브 수	메인 플랫폼 링크	콘텐츠 카테고리	메인 평균 방송 시간 (9 개)/ 업로 드주 기	평균 시청 횟수/ 최근 조회 수	커뮤 니티 링크	성별	얼굴 공개 여부
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

아프리카, 트위치, 유튜브 수 컬럼 결측치 확인

03.데이터 분석 및 전처리

▶ 결측치 제거

```
celeb_df = celeb_df.dropna(subset=['트위치'], how='any', axis=0)
```

```
celeb_df.isnull().sum()
```

```
channel_url_x      0
type               0
message            0
created_at         0
id type            0
channel_member_count_x  0
message_count      0
channel_created_at  0
registration date  0
total_shot         0
트위치             0
아프리카(애청자/팬클럽)  0
유튜브 수         0
메인 플랫폼 링크  0
콘텐츠 카테고리   0
메인 평균 방송시간(9개)/업로드주기  0
평균 시청 횟수/최근 조회수  0
커뮤니티 링크    0
성별              0
얼굴 공개 여부    0
채널주인          0
own_channel       0
channel_name_secured  0
sender_nickname_secured  0
sender_id_secured  0
read_receipt_secured  0
dtype: int64
```

```
celeb_df['sender_id_secured'].nunique()
```

421



```
celeb_df['sender_id_secured'].nunique()
```

212

- 모든 결측치 제거 확인
- 셀럽 수 421명에서 212명으로 감소

03.데이터 분석 및 전처리

▶ 컬럼 추가(total_follower)

```
df_celeb_sum = celeb_df[['트위치', '아프리카(애청자/팬클럽)', '유튜브 수']].apply(lambda series:series.sum(), axis=1)
```

```
celeb_df.insert(13, 'total_follower', df_celeb_sum)
```

total_shot	트위치	아프리카(애청자/팬클럽)	유튜브 수	total_follower
0.0	753.0	0.0	0.0	753.0
0.0	9525.0	0.0	0.0	9525.0
0.0	1504.0	0.0	0.0	1504.0
0.0	0.0	0.0	0.0	0.0
0.0	788.0	0.0	0.0	788.0

아프리카, 트위치, 유튜브 팔로워 수 합산

03.데이터 분석 및 전처리

▶ 컬럼 추가(total_follower)

	created_at	message_count	total_follower	sender_id_secured
0	2022-02-27 01:03:05	2792.0	16620.0	wit****g_@dau****
1	2022-02-27 01:03:06	2792.0	16620.0	wit****g_@dau****
2	2022-03-01 13:43:28	2792.0	16620.0	wit****g_@dau****
3	2022-03-01 22:03:24	2792.0	16620.0	wit****g_@dau****
4	2022-03-01 22:03:32	2792.0	16620.0	wit****g_@dau****
5	2022-03-01 22:03:37	2792.0	16620.0	wit****g_@dau****
6	2022-03-02 23:04:45	2792.0	16620.0	wit****g_@dau****
7	2022-03-02 23:04:50	2792.0	16620.0	wit****g_@dau****
8	2022-03-02 23:04:55	2792.0	16620.0	wit****g_@dau****
9	2022-03-02 23:05:00	2792.0	16620.0	wit****g_@dau****

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 140031 entries, 0 to 541749
Data columns (total 4 columns):
#   Column                Non-Null Count  Dtype
---  -
0   created_at            140031 non-null  datetime64[ns]
1   message_count         140031 non-null  float64
2   total_follower        140031 non-null  float64
3   sender_id_secured     140031 non-null  object
dtypes: datetime64[ns](1), float64(2), object(1)
memory usage: 5.3+ MB
```

RFM 분석을 위하여 필요한 특성만 추출

- Created_at(메시지 전송 시간)
- Message_count(총 메시지 수)
- Total_follower(총 팔로워 수)
- Sender_id_secured(셀럽 id)

가설 검정

04.가설 검증

▶ 가설 설정

	가설
R (메시지 전송 날짜)	셀럽의 가장 최근 메시지 전송 날짜의 분포가 가장 오래된 메시지부터 가장 최근 메시지까지 상위 25% 단위로 잘라 점수를 산정해도 될 만큼 골고루 분포해 있을 것이다.
F (총 메시지 수)	셀럽의 메시지 총량을 상위 25% 단위로 잘라 점수를 산정해도 될 만큼 골고루 분포해 있을 것이다.
M (총 팔로워 수)	셀럽의 총 팔로워 수를 상위 25% 단위로 잘라 점수를 산정해도 될 만큼 골고루 분포해 있을 것이다.

- 셀럽 데이터에 RFM모델을 적용하여 상위 25% 단위로 나누어 R,F,M 각각 1점부터 4점까지 점수를 부여해 고객 세분화
- 위의 가설 검증을 통해 기준과 데이터가 적절한지 확인하고 근거를 제시

04.가설 검정

▶ 데이터 분석 및 가설 검정

1. R(메시지 전송 날짜)

```
df3 = df.groupby('sender_id_secured')['created_at'].max().reset_index()
```

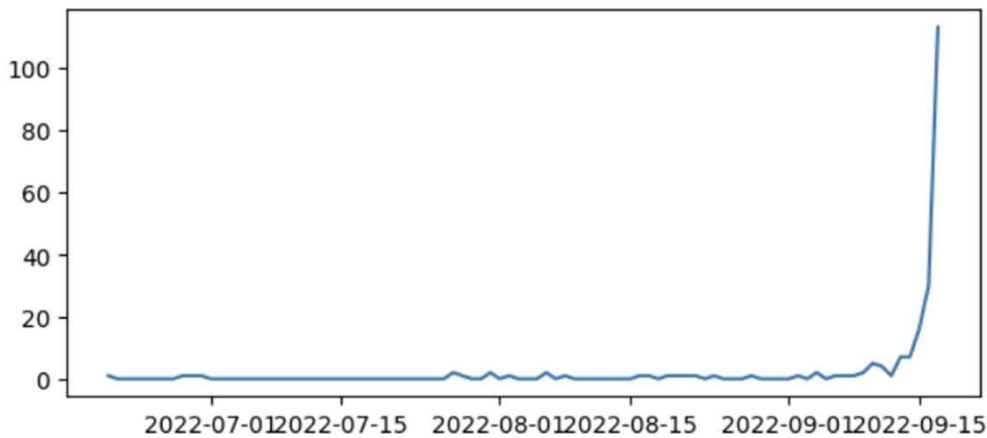
	sender_id_secured	created_at
0	0ga**a0@nav*****	2022-09-14 18:53:12
1	2sa**am@gma*****	2022-09-17 22:39:28
2	382**73@nav*****	2022-09-11 23:56:54
3	595**15@nav*****	2022-09-02 20:55:09
4	982***71@nav*****	2022-09-17 19:29:33
5	MAM***44@gma*****	2022-09-10 21:06:37
6	Npi****dy@gma*****	2022-09-14 22:07:48
7	a03**14@nav*****	2022-09-14 00:59:31
8	abx***17@gma*****	2022-09-17 16:33:16
9	ado*****te@gma*****	2022-09-16 20:20:29
10	alz****31@nav*****	2022-09-17 23:11:05

셀럽 별 가장 최근 메시지를 기준으로
데이터 프레임 생성

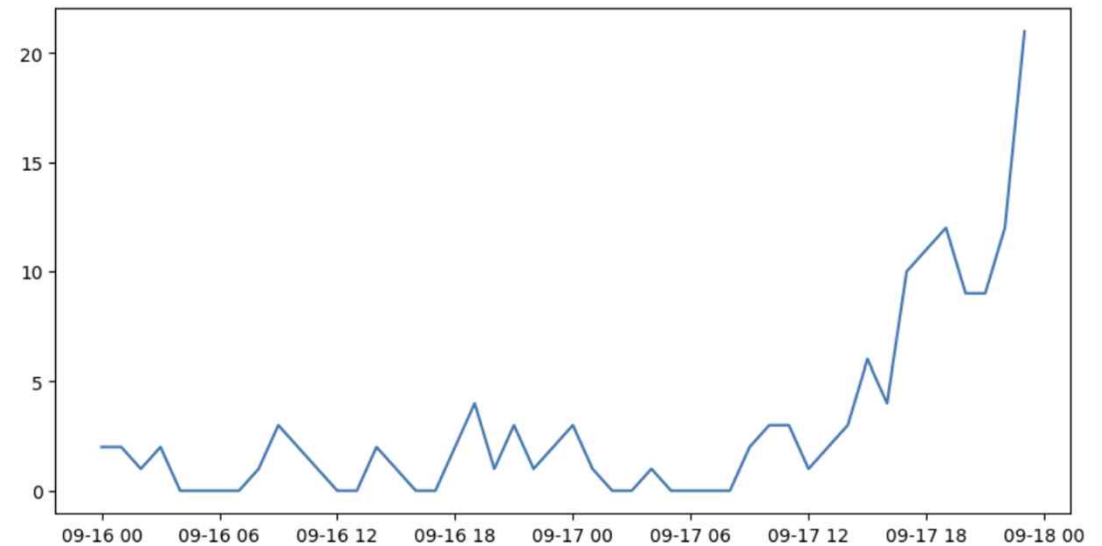
04.가설 검정

▶ 데이터 분석 및 가설 검정

1. R(메시지 전송 날짜)



• 전체 기간 분포도



• 9/16 ~ 9/17분포도

가설 기각

- 총 212명 중 9월 이후로도 메시지를 보낸 셀럽이 191명 그렇지 않은 셀럽이 21명으로 파악됐다
- 가장 빠른 일자인 9월 17일에 마지막으로 보낸 셀럽은 212명중 113명으로 전체 셀럽의 약 50%에 해당한다

➡ 따라서, R(메시지 전송 날짜)에 대하여 설계한 가설은 거짓 새로운 기준 설계 필요

04.가설 검정

▶ 새로운 가설 설정 및 컬럼 추가

	sender_id_secured	created_at	month_cnt
0	0ga**a0@nav*****	2022-09-14 18:53:12	12
1	2sa**am@gma*****	2022-09-17 22:39:28	256
2	382**73@nav*****	2022-09-11 23:56:54	151
3	595**15@nav*****	2022-09-02 20:55:09	8
4	982***71@nav*****	2022-09-17 19:29:33	77
5	MAM***44@gma*****	2022-09-10 21:06:37	4
6	Npi****dy@gma*****	2022-09-14 22:07:48	1
7	a03**14@nav*****	2022-09-14 00:59:31	1
8	abx***17@gma*****	2022-09-17 16:33:16	32
9	ado*****te@gma*****	2022-09-16 20:20:29	61
10	alz*****31@nav*****	2022-09-17 23:11:05	688

← 가장 최근 한 달간 총 메시지 수

새로운 가설(R) : 데이터 수집일 기준 가장 최근 한 달간 셀럽의 총 메시지 수를 기준으로 상위 25% 단위로 잘라 점수를 산정해도 될 만큼 골고루 분포해 있을 것이다.

04.가설 검정

▶ 가설 검정을 위한 데이터 통합

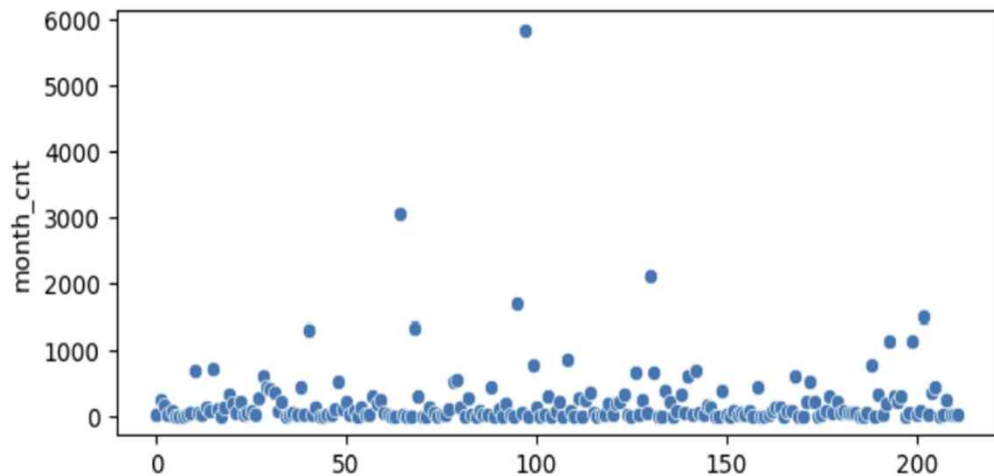
	sender_id_secured	created_at	month_cnt	message_count	total_follower
0	0ga**a0@nav*****	2022-09-14 18:53:12	12	198.0	2855.0
1	2sa**am@gma*****	2022-09-17 22:39:28	256	3011.0	5537.0
2	382**73@nav*****	2022-09-11 23:56:54	151	832.0	0.0
3	595**15@nav*****	2022-09-02 20:55:09	8	8.0	0.0
4	982***71@nav*****	2022-09-17 19:29:33	77	114.0	0.0
5	MAM***44@gma*****	2022-09-10 21:06:37	4	5.0	5495.0
6	Npi****dy@gma*****	2022-09-14 22:07:48	1	1.0	195000.0
7	a03**14@nav*****	2022-09-14 00:59:31	1	1.0	0.0
8	abx***17@gma*****	2022-09-17 16:33:16	32	32.0	1095.0
9	ado*****te@gma*****	2022-09-16 20:20:29	61	248.0	0.0
10	alz****31@nav*****	2022-09-17 23:11:05	688	7481.0	8758.0

새로운 가설 검정 및 R,F,M 관련 특성을 한 번에 다루기 위해
통합 데이터프레임 생성

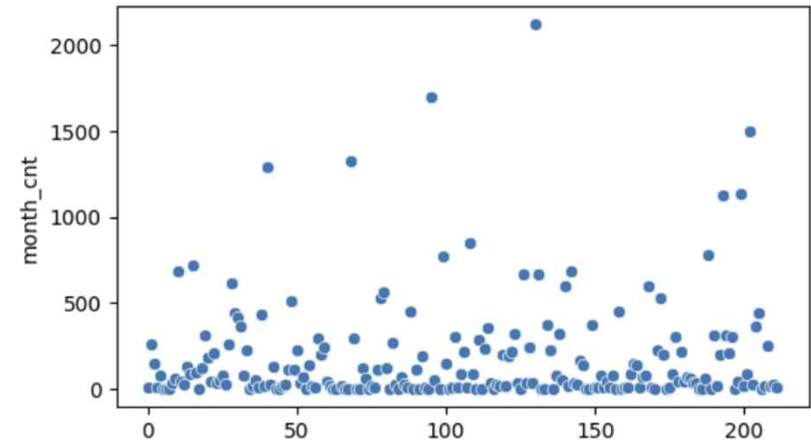
04.가설 검정

▶ 데이터 분석 및 가설 검정

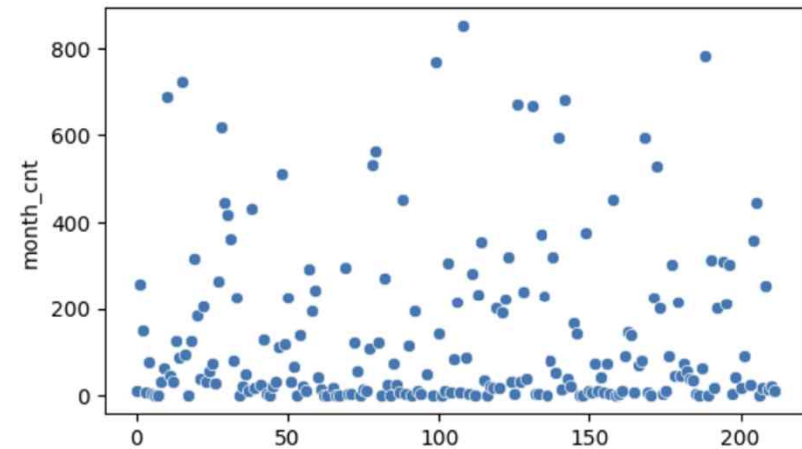
1. R(한 달간 메시지 총량)



• 전체 데이터 분포



• 총량 3000이하 분포



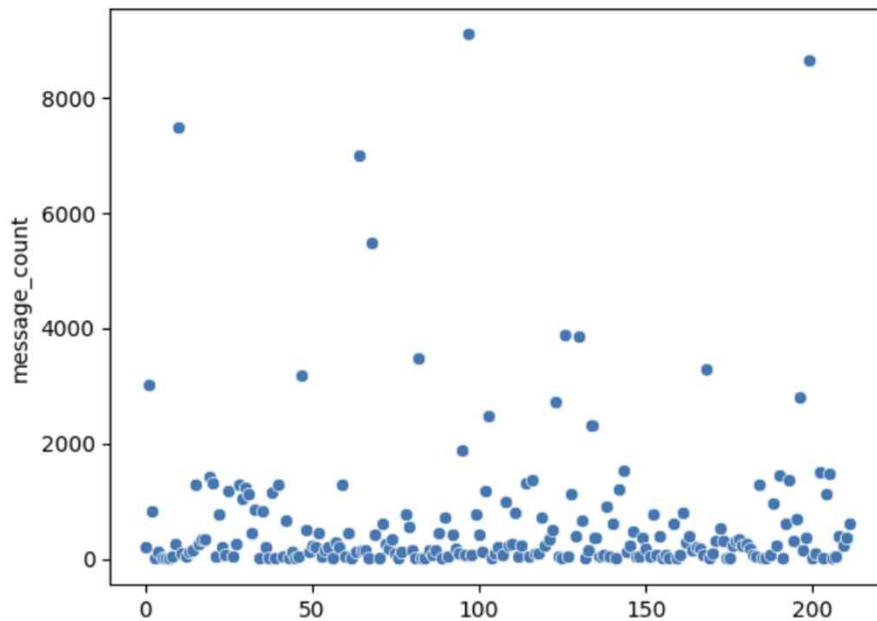
• 총량 1000이하 분포

➤ 분포가 적절하여 가설 **True**

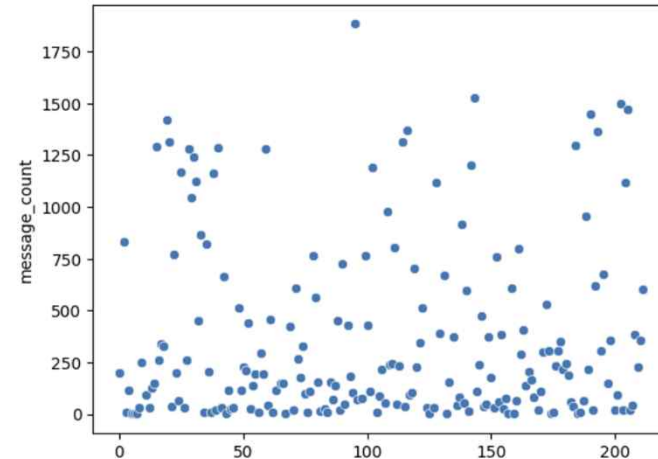
04.가설 검정

▶ 데이터 분석 및 가설 검정

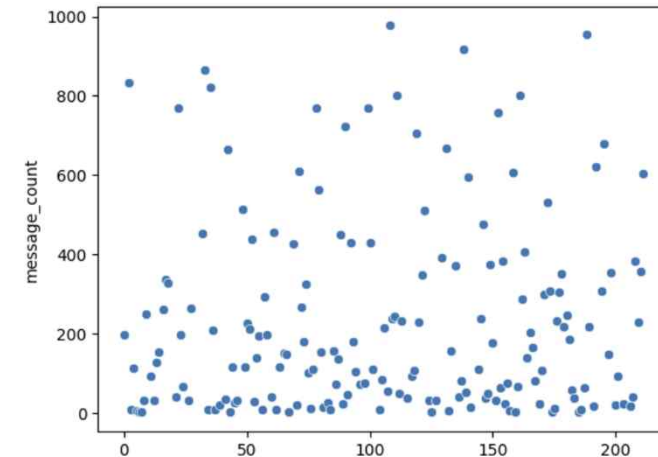
2. F(메시지 총량)



• 전체 데이터 분포



• 총량 2000이하 분포



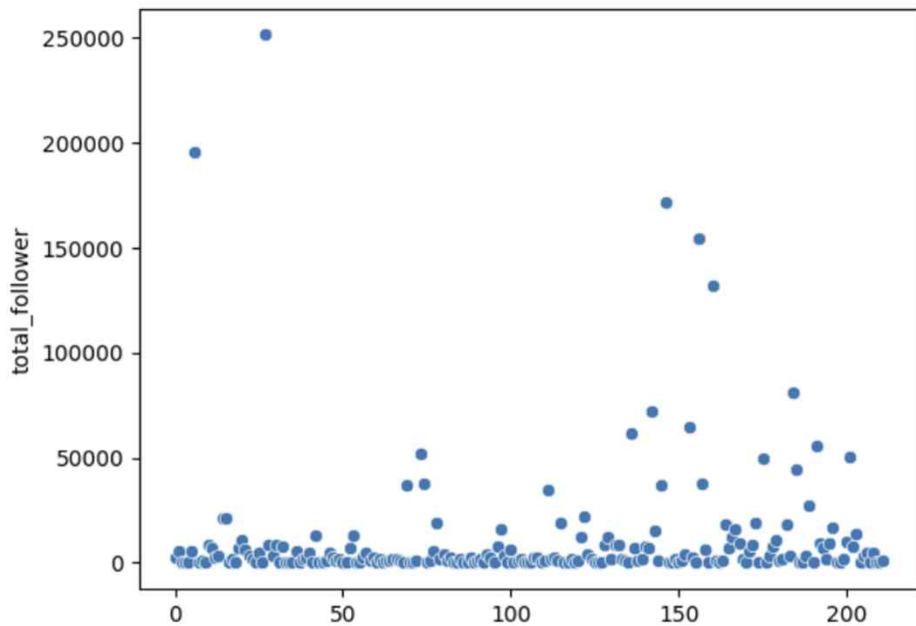
• 총량 1000이하 분포

➤ 분포가 적절하여 가설 **True**

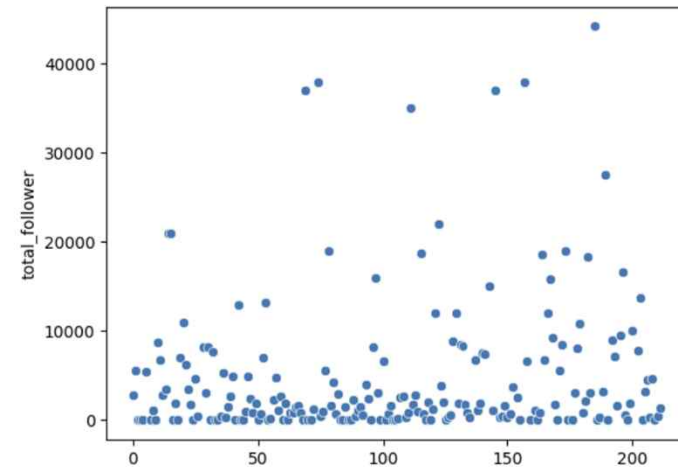
04.가설 검정

▶ 데이터 분석 및 가설 검정

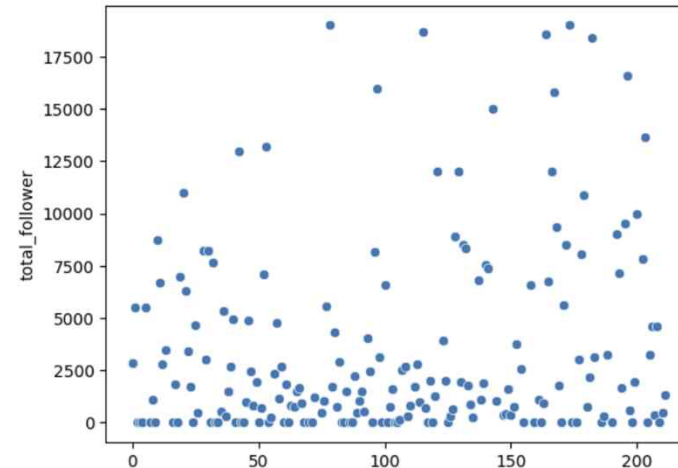
3. M(총 팔로워 수)



• 전체 데이터 분포



• 총량 50000이하 분포



• 총량 20000이하 분포

➤ 분포가 적절하여 가설 **True**

RFM 기반 분석

with Tableau

05. RFM 기반 분석

RFM 분석 기법

각각의 기준에 따라 고객을 4분위로 분류하고 점수 부여

Score	Recency	Frequency	Monetary
특징	작을수록 거래일이 최근	클수록 거래 빈도 많음	클수록 구매 금액 큼
4	0%-25%	75%-100%	75%-100%
3	20%-50%	50%-75%	50%-75%
2	50%-75%	25%-50%	25%-50%
1	75%-100%	0%-25%	0%-25%

05. RFM 기반 분석

기준 별 점수를 합산하여 동일한 점수의 고객들을 하나의 그룹으로 분류함

Segment	특징	Recency Score	Frequeuncy Score	Monetary Score
VIP 고객	최근에 구매했고, 자주 그리고 큰 금액을 소비하는 고객	4	4	4
충성 고객	많은 금액을 소비하고, 프로모션에 잘 반응하는 고객	2 - 4	3 - 4	4
잠재 충성 고객	최근 고객이고, 많은 금액을 소비하며, 1번 이상 구매한 고객	3 - 4	1 - 3	1 - 3
신규 고객	최근에 구매했으나 자주 오지 않은 고객	4	< 2	< 2
잠재 고객	최근에 구매했으나 많이 소비하지 않는 고객	3 - 4	< 2	< 2
관심 필요 고객	최근성, 빈도, 구매 금액이 모두 평균 이상인 고객	3 - 4	3 - 4	3 - 4
잠드려는 고객	최근성, 빈도, 구매 금액이 모두 평균 이하인 고객	2 - 3	< 3	< 3
이탈 우려 고객	많은 금액을 소비하고, 자주 구매했으나 오래 방문하지 않은 고객	< 3	2 - 4	2 - 4
놓치면 안될 고객	매우 큰 금액을 소비하고, 매우 자주 구매했으나 오래 방문하지 않은 고객	< 2	4 - 4	4 - 4
겨울잠 고객	적은 금액을 소비하고, 방문 빈도가 낮으며, 오래 전에 방문한 고객	2 - 3	2 - 3	2 - 3
이탈 고객	최근성, 빈도, 구매 금액이 모두 가장 낮은 고객	< 2	< 2	< 2

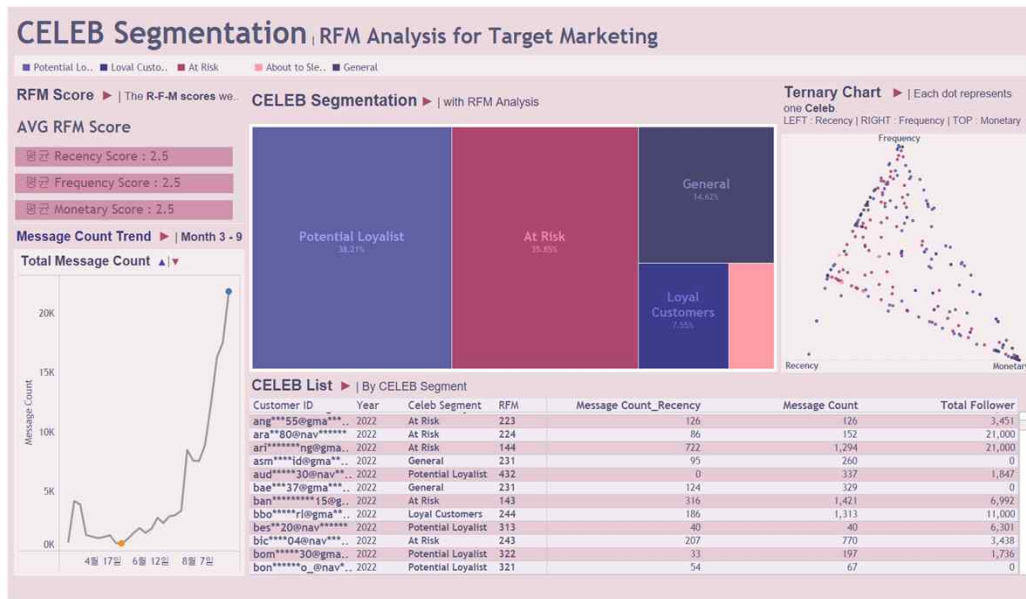
데이터 시각화

(Tableau)

06. 데이터 시각화 (Tableau)

마케팅 시나리오

RFM Celeb Segmentation



1 유저

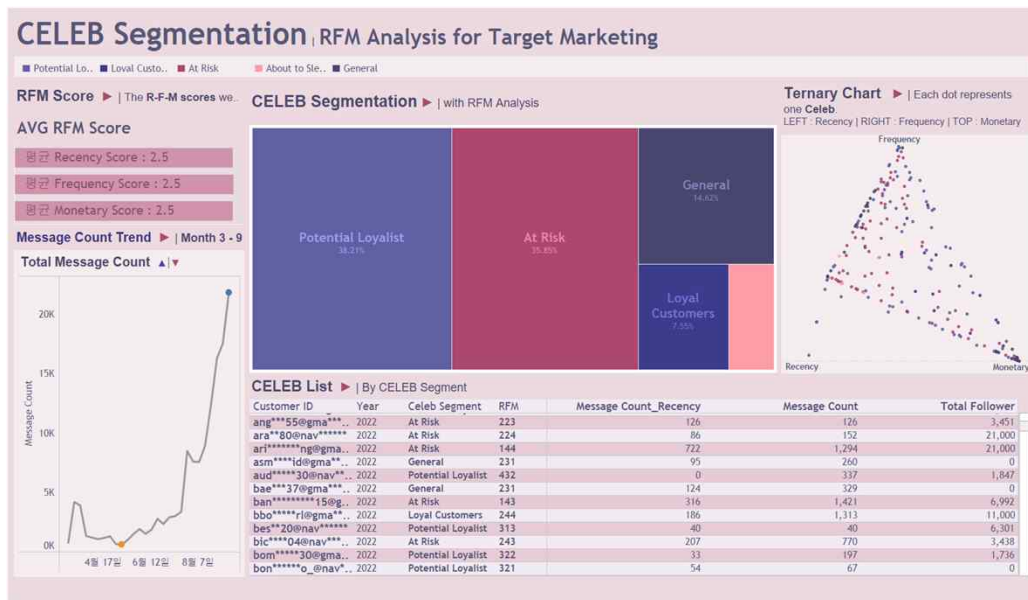
마케터가 직접 분석할 수 있도록 인터랙티브 기능 활용

분석형 대시보드 > 세그먼트 선택 후 해당 고객 데이터 보도록

06. 데이터 시각화 (Tableau)

마케팅 시나리오

RFM Celeb Segmentation



2 기능

비즈니스 목표

효율적인 개인화 마케팅을 통한 매출 증대

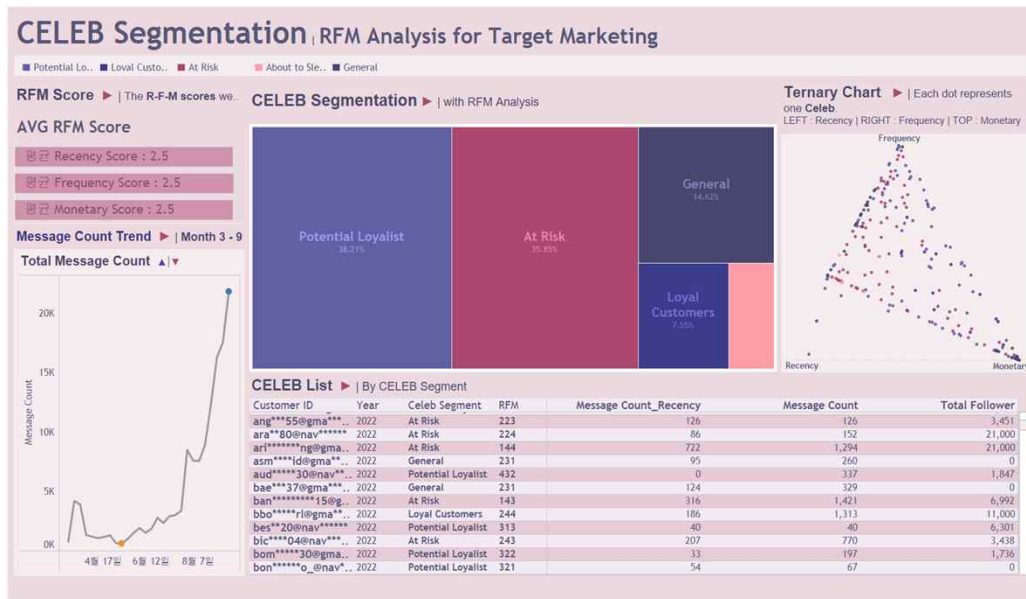
기능목표

1. 고객 세그먼트 별 비중을 한눈에 볼 수 있도록
2. 해당하는 고객 리스트와 정보를 보여줄 것
3. 서비스 사용량 트렌드 시계열 그래프 표시

06. 데이터 시각화 (Tableau)

마케팅 시나리오

RFM Celeb Segmentation



View & Data & Calc

RFM Score

$STR([Recency\ Score]) + STR([Frequency\ Score]) + STR([Monetary\ Score])$

Recency (최근성)

$\{FIXED\ [Sender\ Id\ Secured],\ YEAR([Created\ At]) : MAX([Month\ Cnt])\}$

Frequency (구매 빈도)

$\{FIXED\ [Sender\ Id\ Secured],\ YEAR([Created\ At]) : MAX([Message\ Count])\}$

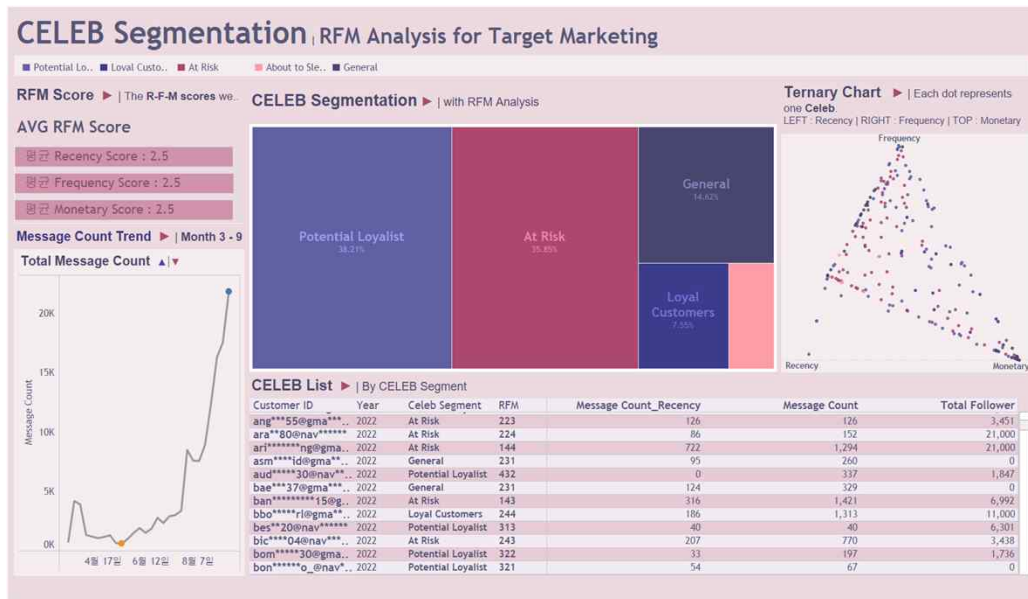
Monetary (구매 금액)

$\{FIXED\ [Sender\ Id\ Secured],\ YEAR([Created\ At]) : MAX([Total\ Follower])\}$

06. 데이터 시각화 (Tableau)

마케팅 시나리오

RFM Celeb Segmentation



3 디자인

컬러

파스텔 핑크 계열 Pink, Red, Purple

글꼴

Trebuchet MS (웹 호환)

레이아웃 & 디테일

트렌드 지표 : 최저, 최고, 강조 + 전월 대비 비교 간트 차트

Ternary Plot : 세그먼트 클릭 시 해당 고객 하이라이트

CELEB Segmentation | RFM Analysis for Target Marketing

■ Potential Lo.. ■ Loyal Custo.. ■ At Risk ■ About to Sle.. ■ General

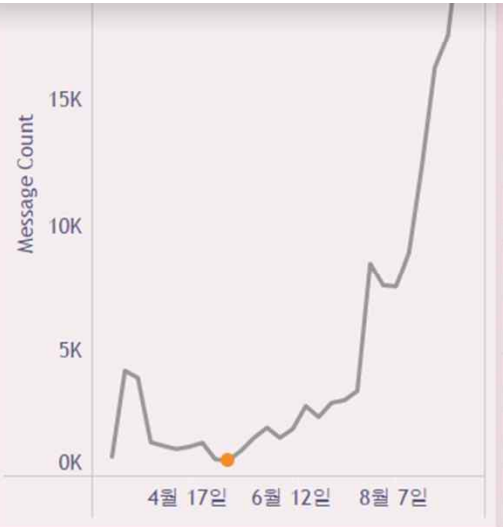
RFM Score ▶ | The R-F-M scores we..

AVG RFM Score

- 평균 Recency Score : 2.5
- 평균 Frequency Score : 2.5
- 평균 Monetary Score : 2.5

1

RFM 스코어



CELEB Segmentation ▶ | with RFM Analysis



CELEB List ▶ | By CELEB Segment

Customer ID	Year	Celeb Segment	RFM
ang***55@gma***	2022	At Risk	223
ara**80@nav*****	2022	At Risk	224
ari*****ng@gma..	2022	At Risk	144
asm*****id@gma**	2022	General	231
aud*****30@nav**	2022	Potential Loyalist	432
bae***37@gma***	2022	General	231
ban*****15@g..	2022	At Risk	143
bbo*****rl@gma**	2022	Loyal Customers	244
bes**20@nav*****	2022	Potential Loyalist	313
bic****04@nav***	2022	At Risk	243
bom*****30@gma..	2022	Potential Loyalist	322
bon*****o_@nav*..	2022	Potential Loyalist	321

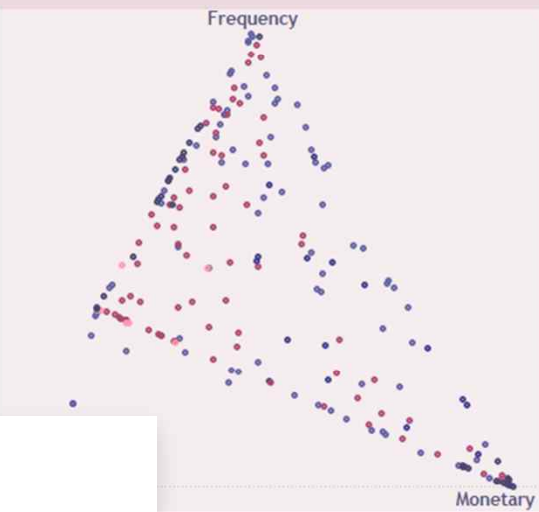
2

RFM 세그멘테이션

3

디자인 / 대시보드 액션

Ternary Chart ▶ | Each dot represents one Celeb.
LEFT : Recency | RIGHT : Frequency | TOP : Monetary



Customer ID	Year	Celeb Segment	RFM	Total Follower
ang***55@gma***	2022	At Risk	223	3,451
ara**80@nav*****	2022	At Risk	224	21,000
ari*****ng@gma..	2022	At Risk	144	21,000
asm*****id@gma**	2022	General	231	0
aud*****30@nav**	2022	Potential Loyalist	432	1,847
bae***37@gma***	2022	General	231	0
ban*****15@g..	2022	At Risk	143	6,992
bbo*****rl@gma**	2022	Loyal Customers	244	11,000
bes**20@nav*****	2022	Potential Loyalist	313	6,301
bic****04@nav***	2022	At Risk	243	3,438
bom*****30@gma..	2022	Potential Loyalist	322	1,736
bon*****o_@nav*..	2022	Potential Loyalist	321	0

06. 데이터 시각화 (Tableau)

[Tableau Link](#)

CELEB Segmentation | RFM Analysis for Target Marketing

■ Potential Lo.. ■ Loyal Cust.. ■ At Risk ■ About to Sle.. ■ General

RFM Score ▶ | The R-F-M scores we..

AVG RFM Score

평균 Recency Score : 2.5

평균 Frequency Score : 2.5

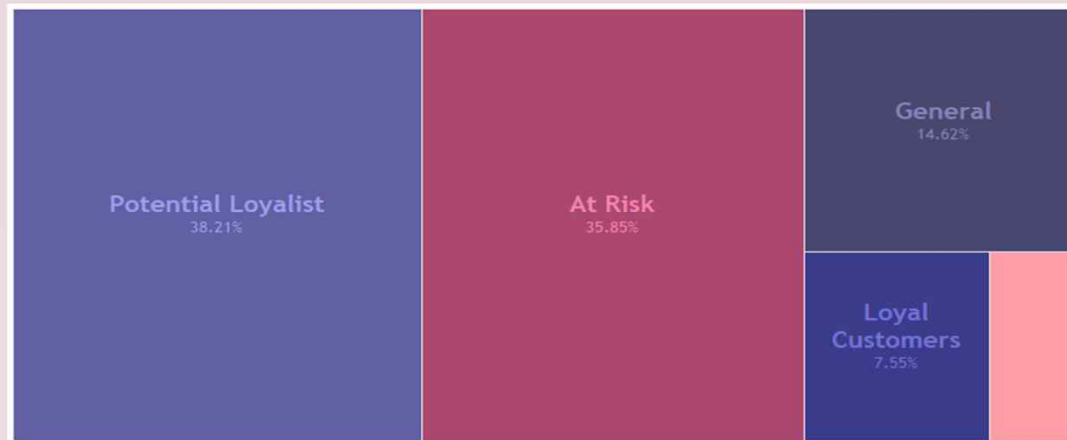
평균 Monetary Score : 2.5

Message Count Trend ▶ | Month 3 - 9

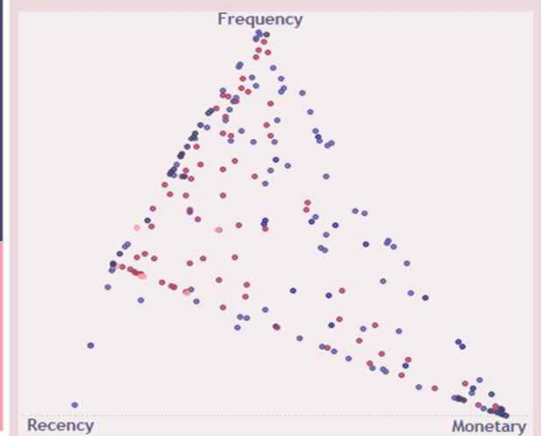
Total Message Count ▲ | ▼



CELEB Segmentation ▶ | with RFM Analysis



Ternary Chart ▶ | Each dot represents one Celeb.
LEFT : Recency | RIGHT : Frequency | TOP : Monetary



CELEB List ▶ | By CELEB Segment

Customer ID	Year	Celeb Segment	RFM	Message Count_Recency	Message Count	Total Follower
ang***55@gma***	2022	At Risk	223	126	126	3,451
ara***80@nav*****	2022	At Risk	224	86	152	21,000
ari*****ng@gma..	2022	At Risk	144	722	1,294	21,000
asm****id@gma**..	2022	General	231	95	260	0
aud****30@nav**..	2022	Potential Loyalist	432	0	337	1,847
bae***37@gma***..	2022	General	231	124	329	0
ban*****15@g..	2022	At Risk	143	316	1,421	6,992
bbo*****r@gma**..	2022	Loyal Customers	244	186	1,313	11,000
bes**20@nav*****	2022	Potential Loyalist	313	40	40	6,301
bic****04@nav***..	2022	At Risk	243	207	770	3,438
bom*****30@gma..	2022	Potential Loyalist	322	33	197	1,736
bon*****o_@nav**..	2022	Potential Loyalist	321	54	67	0

Action Item 도출

07. Action Item 도출

- Action Item

그렇다면 등급을 높이려면 어떻게 해야 할까?



07. Action Item 도출

- Action Item

```
rr_df['month_cnt'].quantile(q=(0.25, 0.5, 0.75), interpolation='nearest')
```

```
0.25    10
0.50    57
0.75   228
Name: month_cnt, dtype: int32
```

```
rr_df['message_count'].quantile(q=(0.25, 0.5, 0.75), interpolation='nearest')
```

```
0.25    52
0.50   217
0.75   678
Name: message_count, dtype: int32
```

```
rr_df['total_follower'].quantile(q=(0.25, 0.5, 0.75), interpolation='nearest')
```

```
0.25    391
0.50   1967
0.75   7856
Name: total_follower, dtype: int32
```

Recenecy (최근 한달 서비스 이용량)

1점 : 10 ↓ 2점 : 10 ↑ 3점 : 57 ↑ 4점 : 228 ↑

Frequency (전체 기간 서비스 이용량)

1점 : 52 ↓ 2점 : 52 ↑ 3점 : 217 ↑ 4점 : 678 ↑

Monetary (총 팔로워 수)

1점 : 391 ↓ 2점 : 391 ↑ 3점 : 1967 ↑ 4점 : 7856 ↑

마치며..

마치며..

프로젝트를 진행하며 느낀 점

- 실무에서 원하는 데이터를 100% 수집 및 활용하기는 많이 어렵겠구나 생각 들었습니다.
- 그래서 프로젝트 기획단계에서 최대한 원하는 데이터 수집 · 적재할 수 있도록 기획을 잘해야 된다고 생각 하지만, 이 또한 어려운 일인 것 같습니다.
- 시간이 더 있었다면 미처 캐치 못한 Feature 간 상관관계 분석 및 더 다양하고 유의미한 가설검정, 보다 디테일한 RFM 가중치 설정, RFM Score 최고점이 나오지 않은 이유에 대한 고찰, 다양한 Action Item 발굴, 그에 따른 기업 방향성 제시 등을 진행할 것 같습니다.
- SQL과 Tableau 활용 측면에서 역량이 부족하다 느꼈습니다. 추가 학습을 통해 해당 역량을 강화 할 예정입니다.

Thanks !

[Tableau Link](#)

코드 : <https://github.com/kjso3887/-CP2>
