

Cancellation Data EDA

yonggi.yeom

2024-03-26

수익률 증가 프로젝트 1

결론

아래 EDA에서 확인한 결과부터 서술한다

- 인바운드를 통해 후원 취소 처리되는 경우는 40%에 달한다
- 후원자 관리 센터를 거치지 않는 후원취소도 24.5%에 달하는 만큼 해당 후원자들과 컨택이 필요하다
- 기간지정 후원자들이 그렇지 않은 후원자보다 총후원금이 높다
 - 기간지정 후원자의 기간을 늘릴 수 있을까?
 - 취소시 총 후원금이 낮은 후원자들에게는 기간지정을 유도하는 방법은?
- 취소가 가장 많이 발생하는 2시 30분경에는 인바운드 피하기

자세한 내용은 아래 본문 참고

수익률 인수분해

현재 개선이 가능한 부분을 찾기 위해 데이터가 아닌 사업모델을 확인하였고

내부 회의를 통해서 제어가 가능한 부분인

기존 후원자의 이탈에 대해서 확인 하기로 결정

자세한 결정 과정에 대해서는 공개 불가

이탈 후원자의 현황을 보여주는 데이터부터 확인

취소 데이터 불러오기

필요 패키지 불러오기

개인정보 제외한 데이터부터 공개

```
glimpse(can_sel)
```

```
## Rows: 147
## Columns: 12
## $ sourcecode      <chr> "LLF", "LLF", "FIN", "FIN", "ODP", "ODP", "ODP...
## $ canceldate      <dtm> 2023-12-04, 2023-12-29, 2023-12-28, 2023-12-2...
## $ cancelcode      <chr> "C7", "C1", "C1", "C1", "C24", "C7", "C24", "C...
## $ cancelreasondescription <chr> "Financial constraints", "Not using credit car...
## $ serialno        <chr> "DR002291", "DR002383", "FF005626", "FF006553"...
## $ bank            <chr> "KFTC", "KFTC", "KFTC", "KFTC", "KFTC", "KFTC"...
## $ frequency       <chr> "1", "1", "1", "1", "1", "1", "1", "1", "1", "1"...
## $ donationamount  <dbl> 10000, 20000, 10000, 10000, 15000, 15000, 1500...
## $ dob             <dtm> 1988-02-27, 1961-11-14, 2003-12-26, 1941-02-1...
## $ gender          <chr> "Female", "Male", "Male", "Female", "Female", ...
## $ firstdeduction  <dtm> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ totalcollected <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2...
```

취소 후원자에 대한 기본 정보를 담고 있는 데이터이며

취소 데이터와 관련이 있는 후원자 관리 센터의 기록을 더하여 확인 예정이다

모든 내부 데이터에는 약정 별 키 값인 **시리얼번호**가 존재한다 해당 키값을 사용해 데이터 병합을 위해서는 데이터 테이블 마다 컬럼명을 동일하게 맞춰 준다

개인정보 제외한 데이터부터 공개

```
colnames(call)[2] <- "serialno" #키 컬럼 변경

call_sel <-
  call %>%
  select(serialno, user, `task group`, `task closed date`)

call_sel <-
  call_sel %>%
  arrange(desc(`task closed date`)) %>%
  distinct(serialno, .keep_all = TRUE) #데이터 전처리

#시리얼번호별 최신 데이터만 남기기
```

후원자 관리센터 데이터에서는 전화 및 인입이력을 시간순으로 저장하고 있어 키값으로 사용하는 **시리얼번호** 별로

여러개의 데이터가 존재할 수 있기 때문에 가장 최신의 데이터 남겨 취소 데이터와 병합할 예정이다 (현재는 취소 데이터를 중점으로 보기위한 목적)

Task 종료 시점을 나타내는 컬럼에서 날짜, 시간 정보를 추출

```
call_sel <-
  call_sel %>%
  mutate(date = as.Date(`task closed date`),
         time = format(strptime(`task closed date`, format="%Y-%m-%d %H:%M:%S"), "%H:%M:%S"))
  %>%
  select(- `task closed date`)

call_sel$time <- hms(call_sel$time)
```

```
glimpse(call_sel)
```

```
## Rows: 2,019
## Columns: 5
## $ serialno      <chr> "DR002378", "FF001712", "DR001102", "FF002298", "FF005250..."
## $ user          <chr> "yonggi.yeom@sgsupport.com", "sangok.oh@sgsupport.com", "...
## $ `task group`  <chr> "Inbound", "Inbound", "Inbound", "Upgrade 12M", "Inbound"..."
## $ date          <date> 2024-01-08, 2024-01-08, 2024-01-02, 2024-01-02, 2023-12-..."
## $ time          <Period> 14H 8M 18S, 12H 49M 34S, 9H 54M 27S, 9H 45M 47S, 17H 1..."
```

불러온 두개의 데이터를 병합하여 EDA 실시

Data EDA

```
glimpse(can_call)
```

```
## Rows: 147
## Columns: 16
## $ sourcecode      <chr> "LLF", "LLF", "FIN", "FIN", "ODP", "ODP", "ODP..."
## $ canceldate      <dtm> 2023-12-04, 2023-12-29, 2023-12-28, 2023-12-2..."
## $ cancelcode      <chr> "C7", "C1", "C1", "C1", "C24", "C7", "C24", "C..."
## $ cancelreasondescription <chr> "Financial constraints", "Not using credit car..."
## $ serialno        <chr> "DR002291", "DR002383", "FF005626", "FF006553"..."
## $ bank            <chr> "KFTC", "KFTC", "KFTC", "KFTC", "KFTC", "KFTC"..."
## $ frequency       <chr> "1", "1", "1", "1", "1", "1", "1", "1", "1", "1"..."
## $ donationamount  <dbl> 10000, 20000, 10000, 10000, 15000, 15000, 1500..."
## $ dob            <dtm> 1988-02-27, 1961-11-14, 2003-12-26, 1941-02-1..."
## $ gender          <chr> "Female", "Male", "Male", "Female", "Female", "...
## $ firstdeduction  <dtm> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N..."
## $ totalcollected <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2..."
## $ user            <chr> "jaeim.jung@sgsupport.com", "yonggi.yeom@sgsup..."
## $ `task group`    <chr> "Inbound", "WebAccess", "WebAccess", "WebAcces..."
## $ date            <date> 2023-12-04, 2023-12-29, 2023-12-28, 2023-12-2..."
## $ time            <Period> 11H 54M 6S, 13H 34M 2S, 15H 19M 29S, 12H 50..."
```

Task별로 입력된 취소 후원자 숫자와 평균 시간대를 확인

```
can_call %>%
  mutate(hour = hour(time),
         minu = minute(time)) %>%
  group_by(`task group`) %>%
  summarise(count = n(),
            avg_h = median(hour),
            avg_m = median(minu)) %>%
  data.frame %>%
  mutate(sum = sum(count),
         perc = round((count / sum)*100,2)) %>%
  select(task.group, count, perc, avg_h, avg_m) %>%
  arrange(desc(count))
```

##	task.group	count	perc	avg_h	avg_m
## 1	Inbound	59	40.14	12.0	29.0
## 2	WebAccess	36	24.49	13.5	25.5
## 3	OUTBOUND	11	7.48	16.0	34.0
## 4	Upgrade 12M	9	6.12	11.0	27.0
## 5	<NA>	9	6.12	NA	NA
## 6	후원 지속콜	5	3.40	17.0	28.0
## 7	2M DNH	4	2.72	13.5	34.5
## 8	긴급 구호 TM	4	2.72	9.5	24.0
## 9	Fix NDNH	3	2.04	14.0	13.0
## 10	ODP Burnt Donor	3	2.04	17.0	30.0
## 11	ODP Pre-enrol Material Required	2	1.36	15.0	13.0
## 12	Reactivation 2M Reject	2	1.36	15.0	33.5

인바운드로 들어와 처리되는 후원취소건수는 전체의 40%에 달하며 후원자 관리센터를 거치지 않고 처리되는 "WebAccess"도 24.5%를 차지하고 있어 해당 취소를 후원자 관리 센터에서 다룰 필요가 있을 것이다.

후원 권유의 성격을 지니고 있는 전화에서도 취소가 발생하고 있어 실제 발생했던 전화 기록과 스크립트를 대조하여 어떤 문구에서

취소가 발생되었는지도 확인할 필요가 있을 것이다.

인바운드로 인입되는 취소후원자 데이터를 통해 개발사 별로 총 후원금 확인

```
can_call %>%
  filter(`task group` == "Inbound") %>%
  group_by(sourcecode) %>%
  summarise(median = median(totalcollected))
```

```
## # A tibble: 5 × 2
##   sourcecode median
##   <chr>      <dbl>
## 1 FIN        85000
## 2 Homepage   270000
## 3 LLF        200000
## 4 ODP        37500
## 5 REACT      50000
```

평균 총 후원금이 높은 두개의 개발사를 중점으로 확인

```
can_call %>%
  filter(sourcecode == "Homepage" | sourcecode == "LLF") %>%
  group_by(cancelreasondescription) %>%
  summarise(n = n()) %>%
  arrange(desc(n))
```

```
## # A tibble: 9 × 2
##   cancelreasondescription      n
##   <chr>                    <int>
## 1 Not using credit card / Auto debit account anymore    16
## 2 Financial constraints                                15
## 3 Stop at 12 months / signed up for 1 year only         9
## 4 Donor refuse to provide reason                       8
## 5 Donation to other charity                           2
## 6 Signed up for a few months only                     2
## 7 Unemployed                                           2
## 8 Donated more than 1 year, 2 years, etc                1
## 9 Left the country                                     1
```

두 개발사의 취소 사유에서 보여주는 것처럼, 개발 당시 부터 기간을 지정하여 인입되어올 가능성이 높다

하지만 다른 개발사 보다 취소시 총후원금이 높기 때문에

취소 요청시 기간 지정을 유도하여 총 후원금을 높이는 전략도 생각해 볼수 있을 것이다.

```
can_call %>%
  filter(cancelreasondescription == "Financial constraints" |
         cancelreasondescription == "Donor refuse to provide reason") %>%
  mutate(hour = hour(time),
         minu = minute(time)) %>%
  summarise(mean = mean(donationamount),
            avg_h = median(hour),
            avg_m = median(minu))
```

```
## # A tibble: 1 × 3
##   mean avg_h avg_m
##   <dbl> <dbl> <dbl>
## 1 15384.    14    27
```

또한 취소 사유중 가장 많은 비중을 차지하는 사유거절과 경제적 사유로 후원취소를 요청하는 후원자는

2시30분경에 15,000원 정도를 후원하는 후원자라는 단순 통계를 이용하여

해당 시간대에 인바운드 대신 취소가 많이 발생하지 않는 정보수정 전화를 진행하는 것도 하나의 방법이 될수 있을 것이다.