

Inblund call Data EDA

yonggi.yeom

2024-03-26

수익률 증가 프로젝트 1-1.

결론

카이제곱검정으로 신뢰수준 95%에서 숙련사용자와 비숙련사용자의 취소방어율 차이가 존재합니다.

숙련사용자의 노하우를 비숙련 사용자에게 전수 또는 공유하므로서 전체 취소 방어율 상승에 기여할 수 있을것으로 예상합니다.

Inbound call 기록 확인

앞서 진행한 취소 후원자 EDA를 통해 취소의 40% 이상이 선 인입되는 전화로 발생하고 있다.

후원자 관리센터 데이터 불러오기

필요 패키지 불러오기

개인정보 제외한 데이터부터 공개

```
glimpse(call_cols)

## Rows: 2,377
## Columns: 9
## $ serialno      <chr> "CI000009", "IM000214", "SR0000000000000000098", "SR...
## $ `source code` <chr> "CONV", "ETC", "ETC", "ETC", "ETC", "ETC", "ETC", "ETC", "ETC"
## $ user          <chr> "jihye.kim@sgsupport.com", "sangok.oh@sgsupport.com", "sangok.oh@sgsupport.com"
## $ `task group`  <chr> "2M DNH", "Inbound", "OUTBOUND", "5, 6주년 리워드 ...
## $ `task closed date` <dtm> 2023-07-11 14:26:18, 2023-12-22 12:03:33, 2023-07-...
## $ `task type`   <chr> "Fix", "Inbound", "Inbound", "General", "General", "General", "General", "General", "General"
## $ outcome       <chr> "Successful", "Upgraded", "Maintained", "Successful", "Successful", "Successful", "Successful", "Successful", "Successful"
## $ `pledge status` <chr> "Unrealised", "Active", "Active", "Active", "Active", "Active", "Active", "Active", "Active"
## $ `donation amount` <dbl> 10000, 30000, 20000, 10000, 30000, 30000, 50000, 30000, 30000
```

```
skimr::skim(call_cols)
```

Data summary


| | |
|-------------------|-----------|
| Name | call_cols |
| Number of rows | 2377 |
| Number of columns | 9 |

| | |
|------------------------|---|
| Column type frequency: | |
| character | 7 |
| numeric | 1 |
| POSIXct | 1 |
| Group variables | |
| None | |

Variable type: character

| skim_variable | n_missing | complete_rate | min | max | empty | n_unique | whitespace |
|---------------|-----------|---------------|-----|-----|-------|----------|------------|
| serialno | 0 | 1 | 4 | 20 | 0 | 2019 | 0 |
| source code | 0 | 1 | 2 | 8 | 0 | 8 | 0 |
| user | 0 | 1 | 22 | 27 | 0 | 12 | 0 |
| task group | 0 | 1 | 5 | 31 | 0 | 14 | 0 |
| task type | 0 | 1 | 3 | 12 | 0 | 6 | 0 |
| outcome | 0 | 1 | 3 | 11 | 0 | 8 | 0 |
| pledge status | 0 | 1 | 5 | 10 | 0 | 6 | 0 |

Variable type: numeric

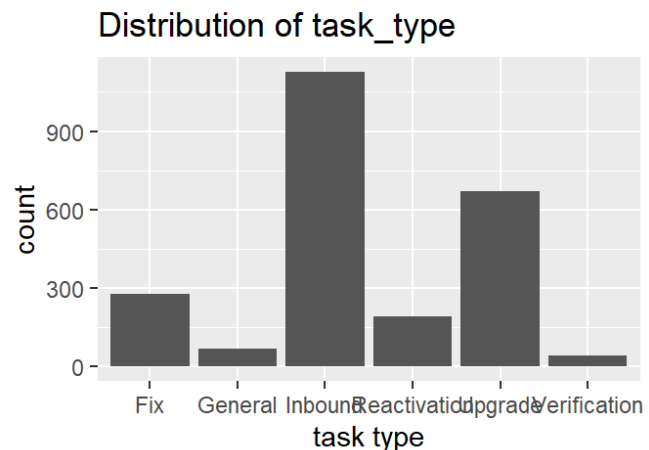
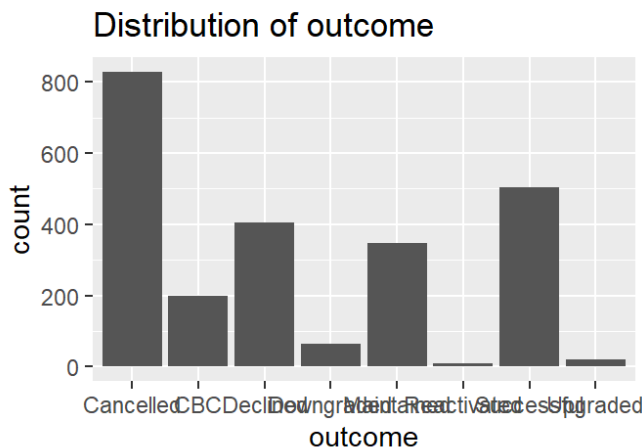
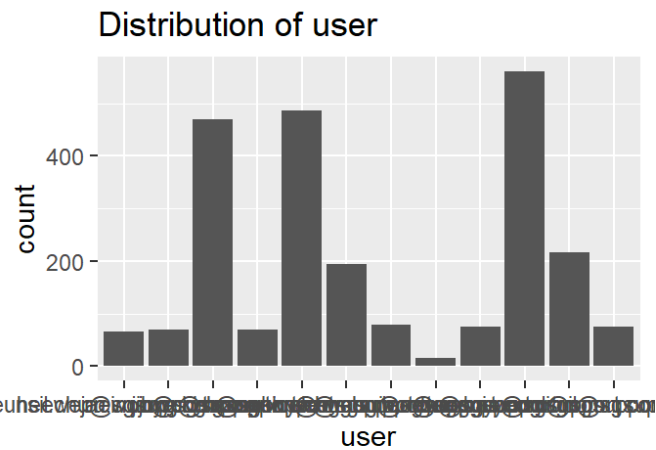
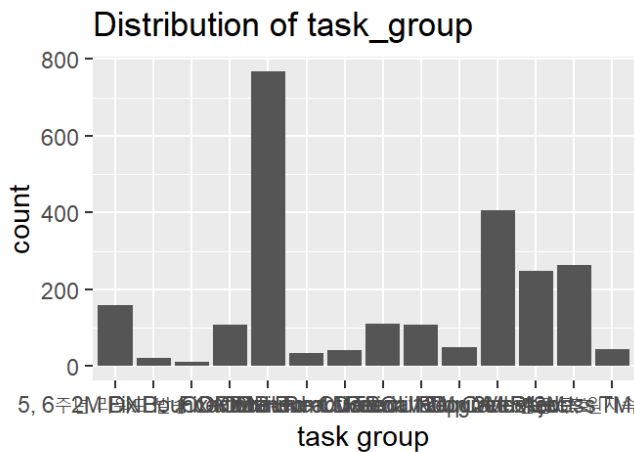
| skim_variable | n_missing | complete_rate | mean | sd | p0 | p25 | p50 | p75 | p100 | hist |
|-----------------|-----------|---------------|----------|----------|------|-------|-------|-------|-------|---|
| donation amount | 0 | 1 | 23250.53 | 159874.2 | 4000 | 10000 | 10000 | 20000 | 5e+06 |  |

Variable type: POSIXct

| skim_variable | n_missing | complete_rate | min | max | median | n_unique |
|------------------|-----------|---------------|---------------------|---------------------|---------------------|----------|
| task closed date | 0 | 1 | 2023-07-03 13:09:44 | 2024-01-08 14:08:18 | 2023-10-13 16:45:56 | 2318 |

해당 파일은 결측치가 없지만, chr 특성이 컬럼이 대다수를 차지하기 때문에 분석 방법을 숫자 컬럼과 다르게 사용해야 한다.

평균이나 표준편차 등으로 데이터 확인하기 어렵게 때문에 분포부터 확인 한다



사용자 마다 전화 수신량의 차이가 있고 인바운드의 기록이 전체 기록 중 대부분을 차지 하고 있음을 알수 있다.
또한 전화의 마무리의 대다수가 후원취소로 확인된다

```
call_cols %>%
  filter(outcome == "Cancelled") %>%
  group_by(`task group`) %>%
  summarise(n = n()) %>%
  data.frame %>%
  mutate(perc = (n / sum(n))*100) %>%
  arrange(desc(n))
```

| ## | task.group | n | perc |
|-------|---------------------------------|-----|-----------|
| ## 1 | Inbound | 407 | 49.095296 |
| ## 2 | WebAccess | 248 | 29.915561 |
| ## 3 | OUTBOUND | 45 | 5.428227 |
| ## 4 | 2M DNH | 31 | 3.739445 |
| ## 5 | Upgrade 12M | 29 | 3.498191 |
| ## 6 | Fix NDNH | 18 | 2.171291 |
| ## 7 | 긴급 구호 TM | 15 | 1.809409 |
| ## 8 | Reactivation 2M Reject | 14 | 1.688782 |
| ## 9 | ODP Burnt Donor | 12 | 1.447527 |
| ## 10 | ODP Pre-enrol Material Required | 10 | 1.206273 |

인바운드 외에도 다른 경로로 취소가 발생하고 있지만 인바운드와 후원자 관리센터를 거치지 않는 취소가 전체의 80%를 차지하고있다

앞선 데이터 분석에서도 확인한 내용 처럼, 후원자 관리 센터를 거치지 않고 후원취소가 이루어지는 후원자들에 대한 관리가 필요하다는 사실을 두가지 데이터로 확인 하였다

조금더 자세한 분석을 위해 센터를 거치지 않는 task 제외한 데이터 확인

```
call_cols %>%
  filter(outcome == "Cancelled") %>%
  filter(`task group` != "WebAccess") %>%
  group_by(`task group`) %>%
  summarise(n = n()) %>%
  data.frame %>%
  mutate(perc = (n / sum(n))*100) %>%
  arrange(desc(n))
```

```
##           task.group    n      perc
## 1           Inbound 407 70.051635
## 2           OUTBOUND  45  7.745267
## 3             2M DNH  31  5.335628
## 4         Upgrade 12M  29  4.991394
## 5             Fix NDNH  18  3.098107
## 6         긴급 구호 TM   15  2.581756
## 7   Reactivation 2M Reject  14  2.409639
## 8             ODP Burnt Donor  12  2.065404
## 9 ODP Pre-enrol Material Required  10  1.721170
```

인바운드에 포함된 예약전화(outbound)까지 포함하면 77% 이상이 인바운드를 통해 취소가 이루어 지고 있다
가장 많은 데이터를 차지하는 인바운드 데이터와 그 전화의 결과만 살펴본다

```
call_cols %>%
  filter(`task group` == "Inbound") %>%
  group_by(outcome) %>%
  summarise(n = n()) %>%
  data.frame %>%
  mutate(perc = n / sum(n),
         per = scales::percent(perc)) %>%
  select(-perc) %>%
  arrange(desc(n))
```

```
##      outcome    n    per
## 1  Cancelled 407 52.99%
## 2  Maintained 292 38.02%
## 3  Downgraded  46  5.99%
## 4   Upgraded  14  1.82%
## 5  Reactivated   9  1.17%
```

예상대로 전화의 53% 정도는 취소로 종료되지만 해지방어로 적지 않은 비율로 발상하고 있음을 확인 할수 있다.

가설설정 및 검증

사용자별 전화 수신량을 차이가 인바운드 전화의 결과를 다르게 가져가지는 않는지 즉 사용자별 취소 방어율과 취소율이 통계적으로 유의미한 수준으로 다른지 확인

숙련 사용자는 비숙련 사용자의 전화결과에 차이를 보일것이다.

```
call_in <-
  call_in %>%
    filter(user != "yonggi") %>%
    filter(user != "heeweon") # 실제 전화를 받지 않는 사용자 제외

call_in <-
  call_in %>%
    mutate_if(is.character, as.factor) #분석을 위해 특성 변경

call_in$serialno <- as.character(call_in$serialno) #분석에 필요없는 ID값

glimpse(call_in)
```

```
## Rows: 714
## Columns: 9
## $ serialno      <chr> "IM000214", "RI000059", "3680", "3757", "SR00000000...
## $ `source code` <fct> ETC, REACT, Homepage, Homepage, ETC, Homepage, Home...
## $ `task group`   <fct> Inbound, Inbound, Inbound, Inbound, Inbound, Inbound...
## $ `task closed date` <dtm> 2023-12-22 12:03:33, 2023-12-12 11:38:11, 2023-12-...
## $ `task type`    <fct> Inbound, Inbound, Inbound, Inbound, Inbound, Inbound...
## $ outcome        <fct> defense, cancelled, defense, cancelled, defense, de...
## $ `pledge status` <fct> Active, Cancelled, Cancelled, Cancelled, Active, Ac...
## $ `donation amount` <dbl> 30000, 10000, 50000, 10000, 20000, 30000, 50000, 25...
## $ user           <fct> sangok, wonhee, sangok, sangok, myunghoon, wonhee, ...
```

```
skimr::skim(call_in)
```

Data summary

| Name | call_in |
|------------------------|---------|
| Number of rows | 714 |
| Number of columns | 9 |
| <hr/> | |
| Column type frequency: | |
| character | 1 |
| factor | 6 |
| numeric | 1 |
| POSIXct | 1 |
| <hr/> | |
| Group variables | None |

Variable type: character

| skim_variable | n_missing | complete_rate | min | max | empty | n_unique | whitespace |
|---------------|-----------|---------------|-----|-----|-------|----------|------------|
| serialno | 0 | 1 | 4 | 20 | 0 | 622 | 0 |

Variable type: factor

| skim_variable | n_missing | complete_rate | ordered | n_unique | top_counts |
|---------------|-----------|---------------|---------|----------|---------------------------------------|
| source code | 0 | 1 | FALSE | 7 | FIN: 277, LLF: 264, ODP: 132, SG: 16 |
| task group | 0 | 1 | FALSE | 1 | Inb: 714 |
| task type | 0 | 1 | FALSE | 1 | Inb: 714 |
| outcome | 0 | 1 | FALSE | 2 | can: 363, def: 351 |
| pledge status | 0 | 1 | FALSE | 6 | Can: 422, Act: 246, Unr: 25, Hol: 11 |
| user | 0 | 1 | FALSE | 9 | won: 203, myu: 166, jae: 163, san: 46 |

Variable type: numeric

| skim_variable | n_missing | complete_rate | mean | sd | p0 | p25 | p50 | p75 | p100 | hist |
|-----------------|-----------|---------------|----------|--------|------|-------|-------|-------|-------|------|
| donation amount | 0 | 1 | 39589.63 | 290536 | 4000 | 10000 | 15000 | 20000 | 5e+06 | |

Variable type: POSIXct

| skim_variable | n_missing | complete_rate | min | max | median | n_unique |
|------------------|-----------|---------------|---------------------|---------------------|---------------------|----------|
| task closed date | 0 | 1 | 2023-07-03 14:31:24 | 2024-01-08 12:49:34 | 2023-10-18 15:12:01 | 714 |

검증을 위해 사용자를 숙련과 비숙련 그룹으로 나눔

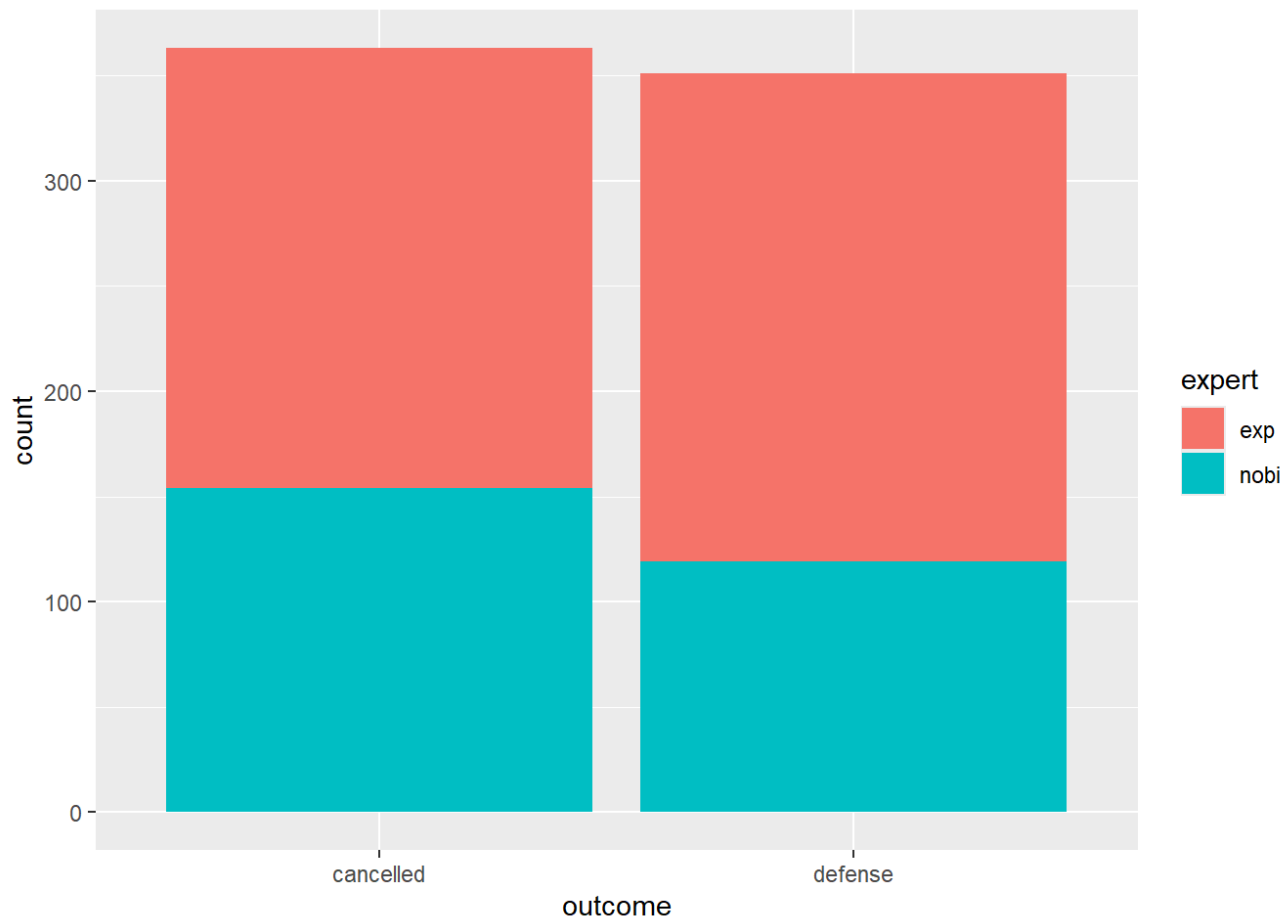
```
call_in<-
  call_in %>%
  mutate(expert = ifelse(user == "myunghoon" | user == "sangok"
    |user == "jihye" | user == "wonhee", "exp", "nobi"))

call_table <- xtabs(~expert + outcome, data = call_in)

prop.table(call_table, margin = 1) *100

##      outcome
## expert cancelled defense
##  exp  47.39229 52.60771
##  nobi  56.41026 43.58974

ggplot(call_in , aes(x = outcome, fill = expert)) + geom_bar()
```



인바운드 처리결과를 숙련, 비숙련으로 나눈 결과로 숙련은 취소방어가 비숙련에 비해 9% 정도 높은 것으로 확인되었다

해당 표가 통계적으로 의미가 있는지 카이제곱 검정으로 진행한다

```
call_table %>%
  chisq.test() %>%
  broom::tidy()
```

```
## # A tibble: 1 × 4
##   statistic p.value parameter method
##   <dbl>    <dbl>     <int> <chr>
## 1      5.13  0.0235         1 Pearson's Chi-squared test with Yates' continuity...
```

가설 설정 H_0 : 두 그룹의 차이는 없다. H_1 : 두 그룹의 차이는 존재 한다.

P값이 0.02로 유의수준 0.05보다 낮기 때문에 귀무가설 H_0 를 기각하고 대립 가설 H_1 을 채택 할수 있다

즉 숙련사용자들의 취소방어가 높기 때문에 해당 숙련 사용자들의 취소 방어 전화 중 교육 자료로 생성하여 비숙련 사용자에게 공유하는 방식으로하여 취소 방어률을 높일수 있을 것입니다.