

# Mobile 고객 이탈 분석

박유진, 서승우, 송정우

Table of Contents

01	02	03	04	05
				느낀점
	이변량 분석	추가 분석	관계 정리	
가설수립				





# 01 가설수립

## 01 가설 수립

1. COLLEGE에 따라 CHURN에 차이가 있다.
2. REPORTED\_SATISFACTION에 따라 CHURN에 차이가 있다.
3. CONSIDERING\_CHANGE\_OF\_PLAN에 따라 CHURN에 차이가 있다.
4. INCOME에 따라 CHURN에 차이가 있다.
5. OVERAGE에 따라 CHURN에 차이가 있다.
6. LEFTOVER에 따라 CHURN에 차이가 있다.
7. HANDSET\_PRICE에 따라 CHURN에 차이가 있다.
8. AVERAGE\_CALL\_DURATION에 따라 CHURN에 차이가 있다.





## 02 이변량 분석

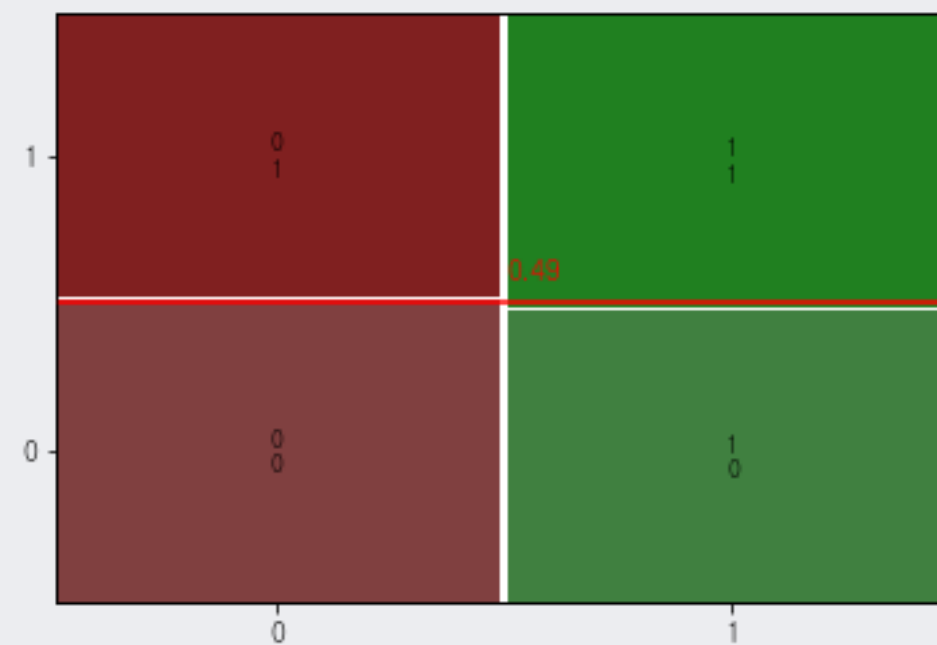
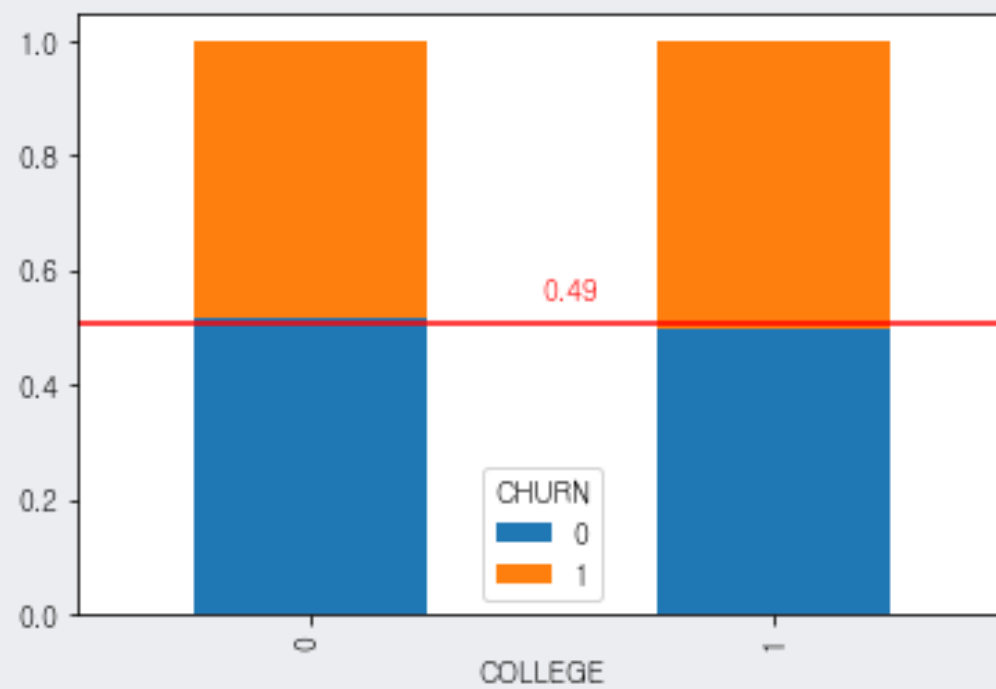
## 02 이변량 분석

### 1. COLLEGE(대학졸업 여부) --> CHURN

1: 졸업, 0: 미졸업

COLLEGE	0	1
CHURN		
0	0.515697	0.49895
1	0.484303	0.50105

카이제곱통계량 4.982102934693503  
p-value 0.025610833478220536  
자유도 1



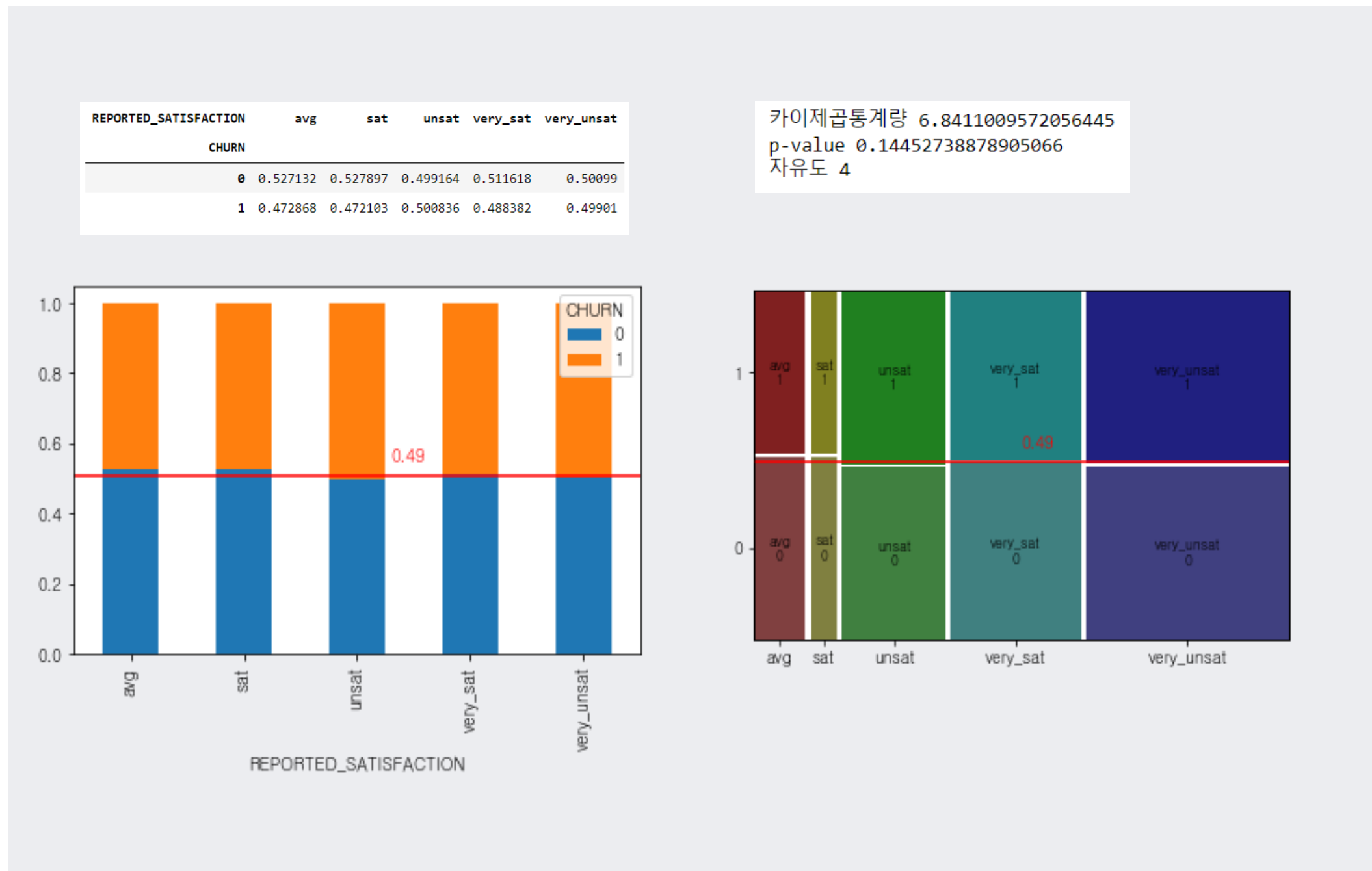
#### 분석 결과

- 전체평균 비율과, 대학졸업 평균 비율이 거의 동일하다는 점에서 COLLEGE와 CHURN은 차이가 없다.
  - 즉, 이 둘은 관계가 없다고 볼 수 있다.

#### 추가 분석

- 대학졸업여부 보다는 다른 변수를 분석해서 유의미한 결과를 확인하는 게 더 효율적으로 보인다.

### 2. REPORTED\_SATISFACTION --> CHURN



#### 분석 결과

- 단변량 분석에서 예상과 달리, 만족도의 항목별 이탈 비율과 이탈 전체 평균이 큰 차이를 보이지 않는다.
- 수치화를 통해 보더라도, 자유도가 4이지만 카이제곱 통계량이 약 6.8로 2~3배가 나지도 않는다.
  - 또한, p-value값도 약 14%의 수치를 보인다.
- 결론적으로 이변량 분석에서 시각화와 가설검정을 확인해보면, 만족도는 이탈여부에 차이가 따르지 않는다.

#### 추가 분석

- 하지만 과연 그럴까?
  - 지금 당장은 이탈을 하지 않았더라도 앞으로 이탈을 할 확률이 커보인다.
  - 아마도 이탈하지 않은 상황에서 설문에 참여한 것으로 예상된다.
- 변경 계획 설문 등과 함께 분석해볼 필요가 있다.
- 만족도를 서비스에 따라서 분석해볼 필요가 있다.

### 3. CONSIDERING\_CHANGE\_OF\_PLAN --> CHURN

CONSIDERING_CHANGE_OF_PLAN	actively_looking_into_it	considering	never_thought	no	perhaps
CHURN					
0	0.511565	0.510853	0.513304	0.500274	0.475992
1	0.488435	0.489147	0.486696	0.499726	0.524008

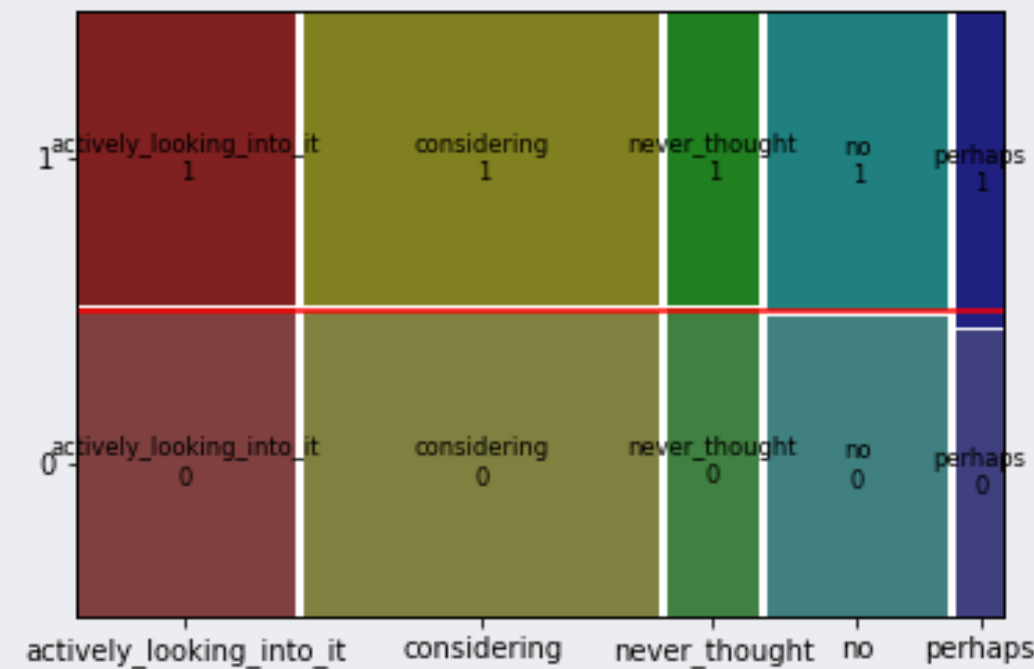
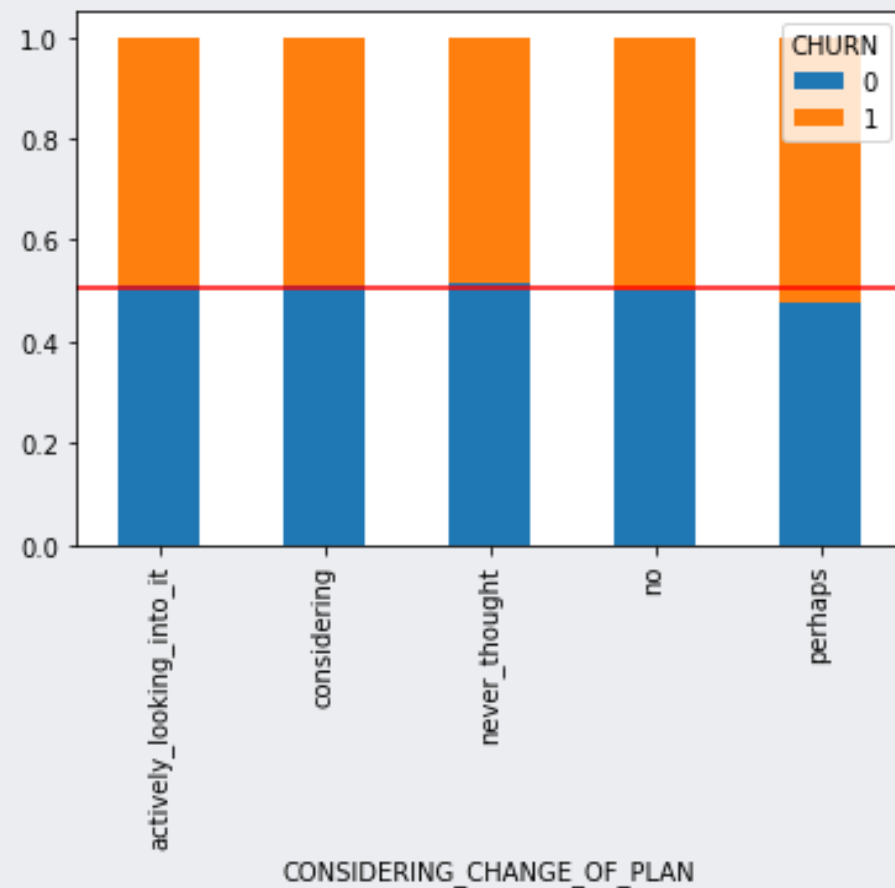
카이제곱통계량 : 5.421434419620689  
p-value : 0.2467222586679969

#### 분석 결과

- 변경 계획 설문 결과와 이탈 여부는 큰 관련이 없다.

#### 추가 분석

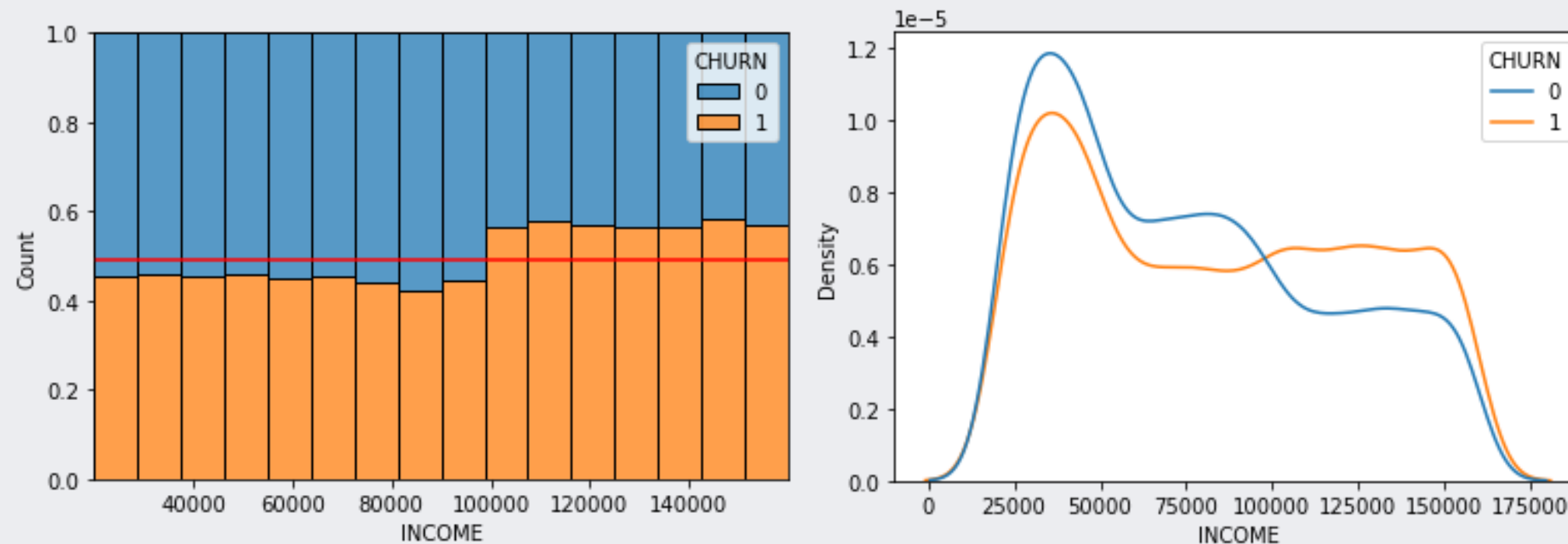
- 설문조사는 그 방식과 질문에 따라 결과가 달라질 수 있다.  
따라서 설문조사를 진행한 방식을 더 살펴볼 필요가 있다.





### 4. INCOME(연 수입, 달러) --> CHURN

```
Optimization terminated successfully.  
Current function value: 0.692683  
Iterations 2  
INCOME    0.000044  
dtype: float64
```



#### 분석 결과

- 연 수입이 일정 금액을 넘어서면, 이탈여부가 증가한다.
- 연 수입과 이탈 여부 사이에 관련이 있다.

#### 추가 분석

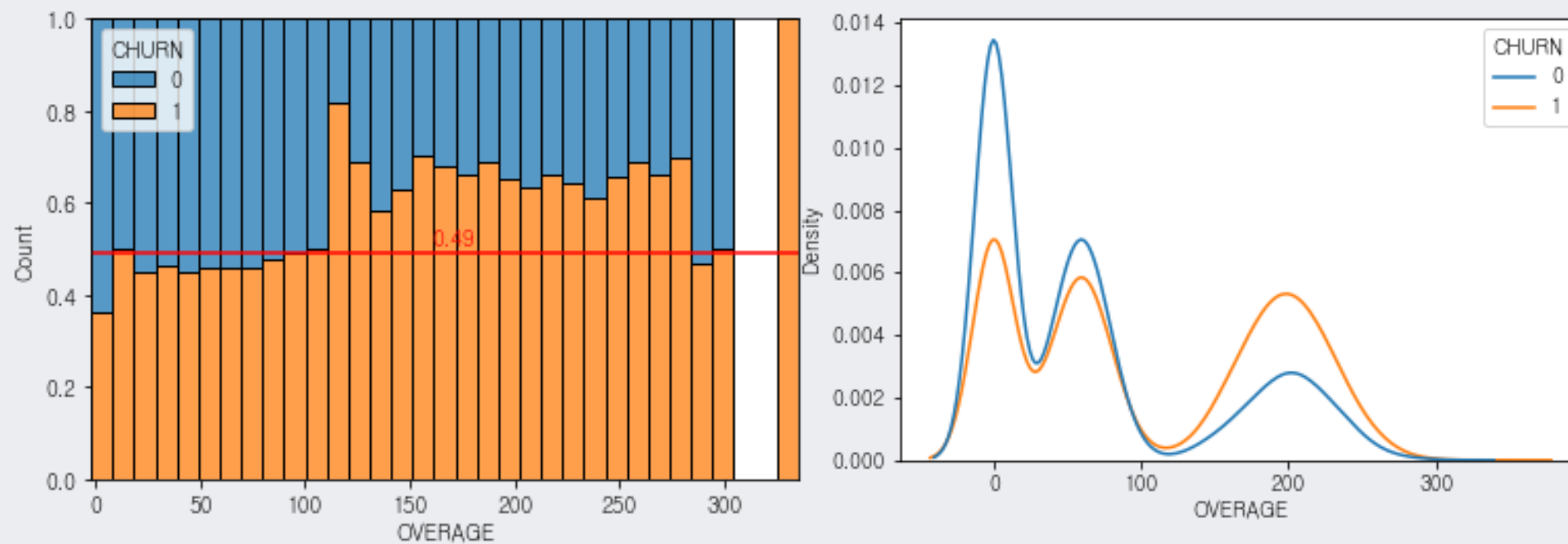
- 고객 연봉자에게는 금액을 지불할 능력이 있기 때문에, 기본적인 통화량 제공 뿐 아니라, 고객에게 별도로 제공될 프리미엄 상품을 준비해야 한다.  
(고객 연봉자에게는 심민적인 요소가 더 크게 작용할 수 있다.)

### 5. OVERAGE --> CHRUN

OVERAGE: 월 초과사용 시간(분)

```
count    18000.000000
mean      86.014833
std       85.963981
min       -2.000000
25%        0.000000
50%       59.000000
75%      179.000000
max      335.000000
Name: OVERAGE, dtype: float64
```

```
Optimization terminated successfully.
Current function value: 0.680553
Iterations 4
OVERAGE    3.521095e-97
dtype: float64
```



#### 분석 결과

- 그래프 상에서,
  - 약 110분을 기준으로 대부분의 구간에서는 전체 평균보다 이탈 비율이 높음을 알 수 있다.
- 로지스틱회귀로부터,
  - p-value가 거의 0에 수렴함으로 유의미한 값으로 볼 수 있다.
- 결론: 월 초과사용 시간에 따라 이탈여부 차이가 있다고 볼 수 있다.
  - 강한 관계

#### 추가 분석

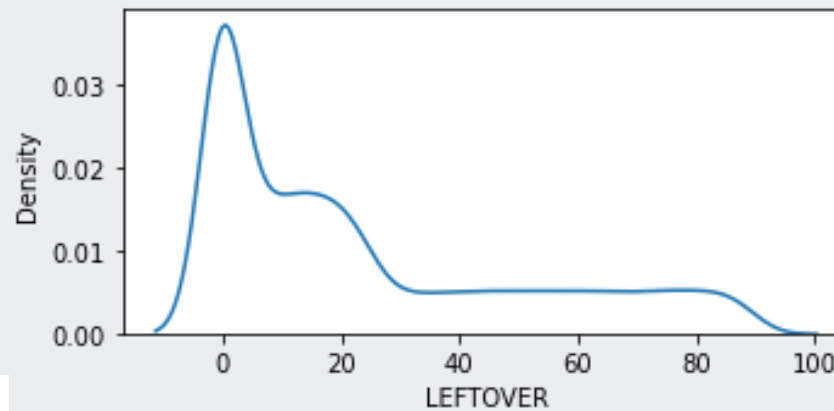
- 약 280분 구간에서 다시 평균 밑으로 내려온 이유에 대해 확인해볼 필요가 있다.

### 6. LEFTOVER--> CHURN

LEFTOVER: 월 사용 잔여시간(%)

```
count    18000.000000
mean      23.995500
std       26.860363
min        0.000000
25%        0.000000
50%      14.500000
75%      42.000000
max       89.000000
Name: LEFTOVER, dtype: float64
```

```
Optimization terminated successfully.
Current function value: 0.692532
Iterations: 3
LEFTOVER    0.000003
dtype: float64
```

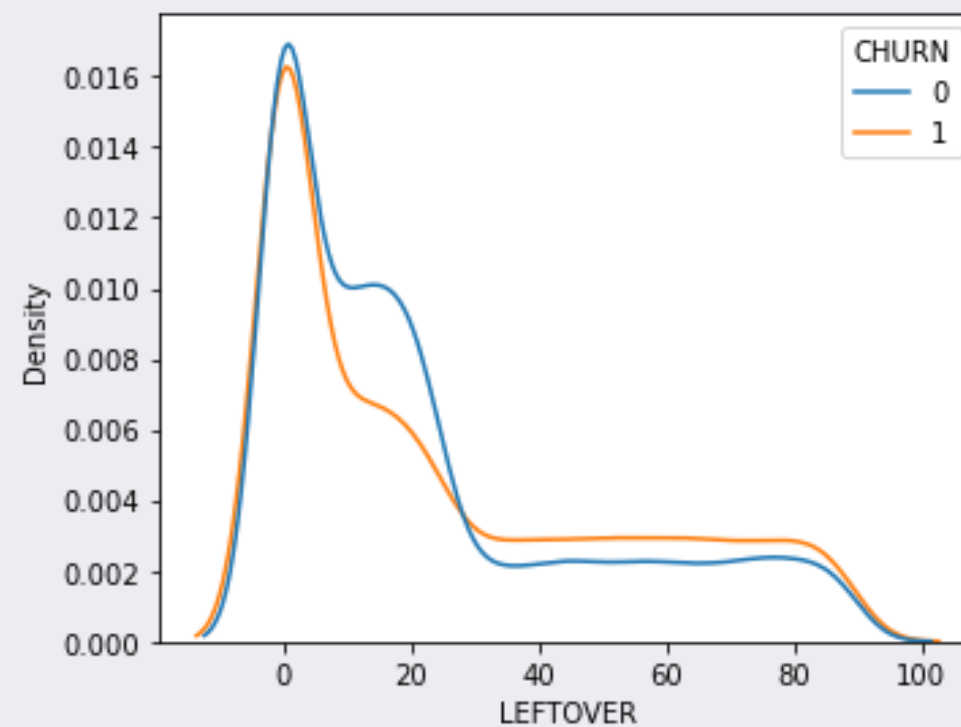
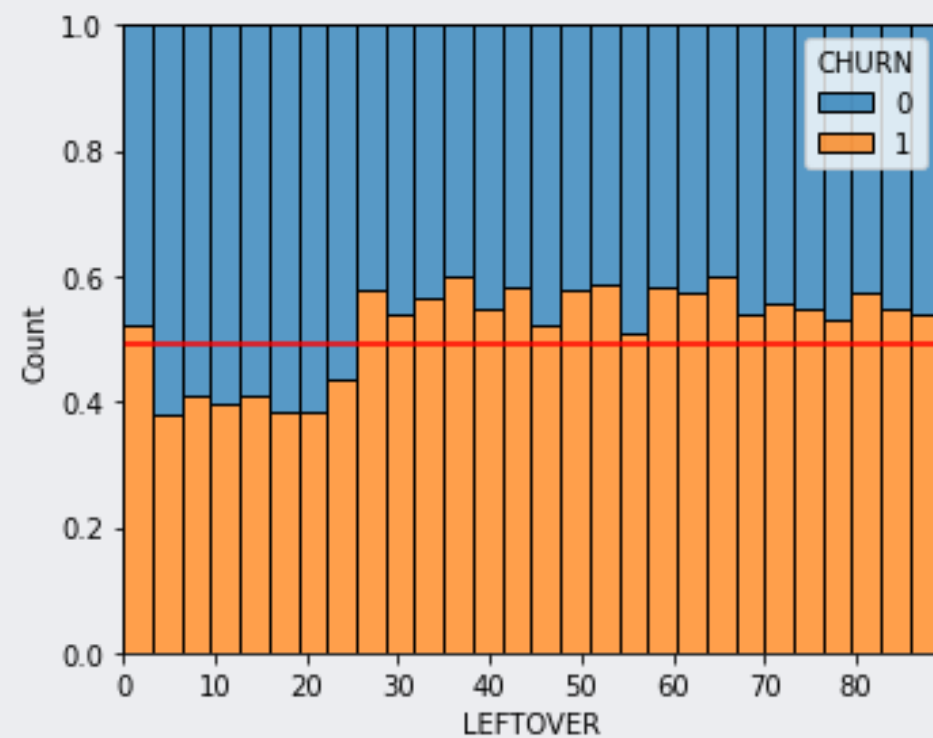


#### 분석 결과

- 월 사용 잔여시간(%)은 10%, 이내 30% 이내 그리고 그 이상을 기준으로 3개의 그룹으로 나뉜다. 보통 30% 이내의 고객의 이탈율이 낮은 편이었다.
- 로지스틱 회귀 결과 월 사용 잔여시간과 통신사 이탈 여부 간의 관계가 어느 정도는 있어 보인다.

#### 추가 분석

- 월 사용 잔여시간을 10% 미만, 30% 미만, 그리고 그 이상으로 범주로 바꾸어 분석할 필요가 있다. 왜냐하면 수치화에서 사용된 로지스틱 회귀는 미봉책이기 때문이다.



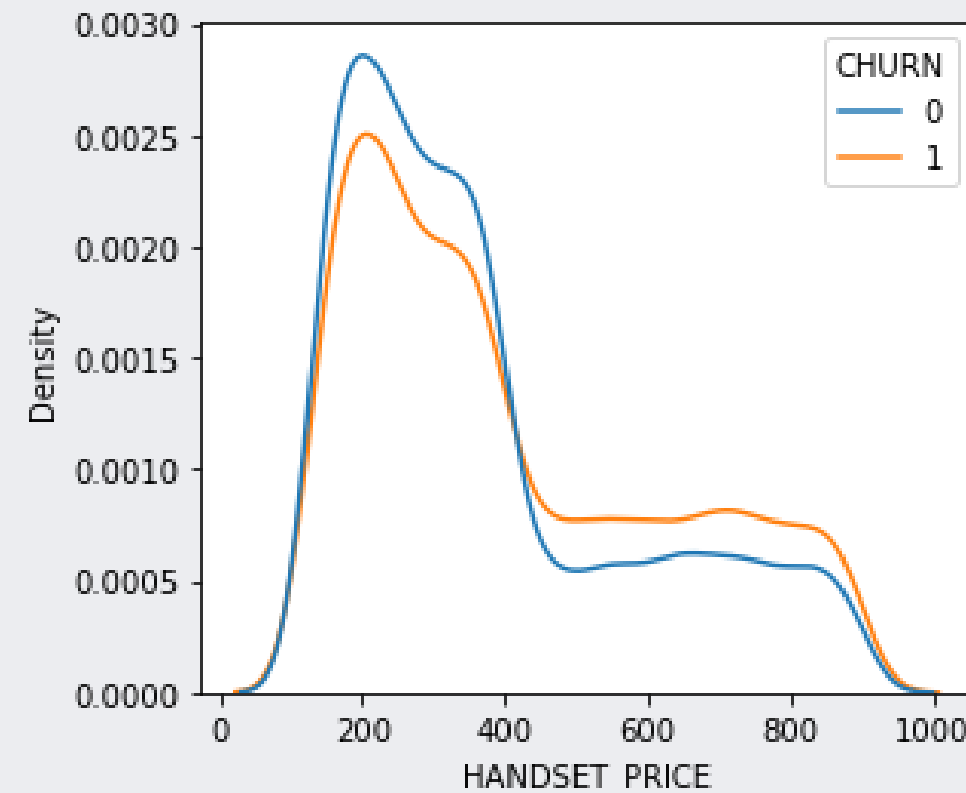
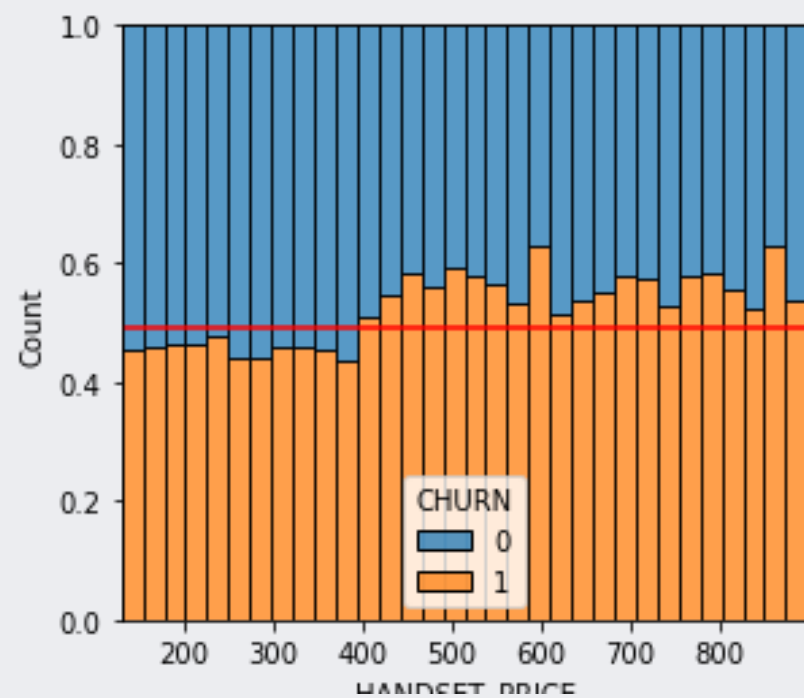
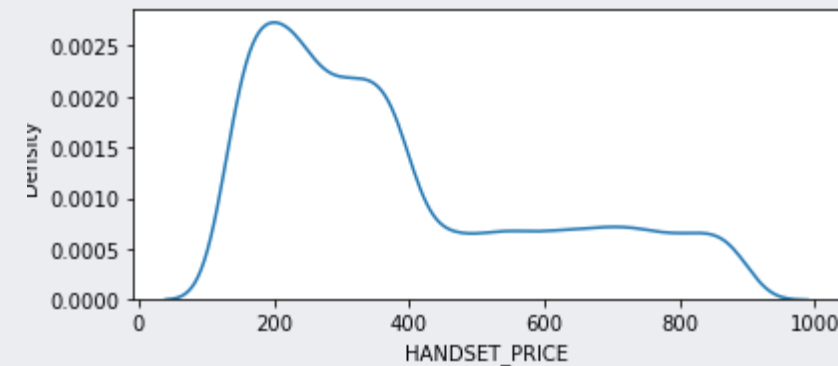


### 7. HANDSET\_PRICE --> CHURN

HANDSET\_PRICE: 휴대폰 가격

```
count    18000.000000
mean      390.096667
std       214.239639
min        130.000000
25%       219.000000
50%       326.000000
75%       536.000000
max        899.000000
Name: HANDSET_PRICE, dtype: float64
```

```
Optimization terminated successfully.
Current function value: 0.692730
Iterations: 3
HANDSET_PRICE    0.000107
dtype: float64
```



#### 분석 결과

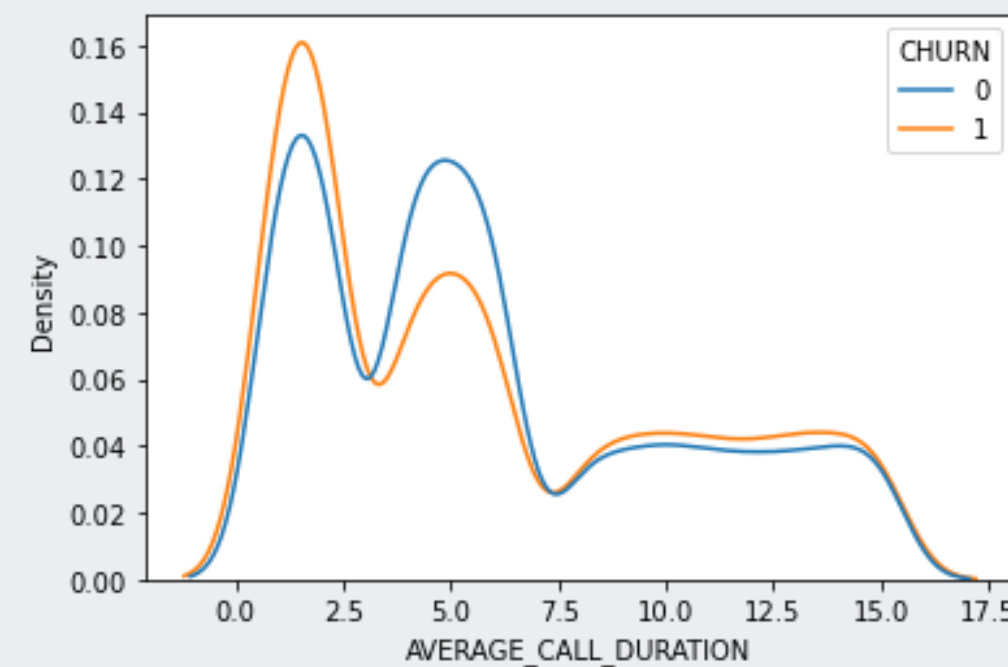
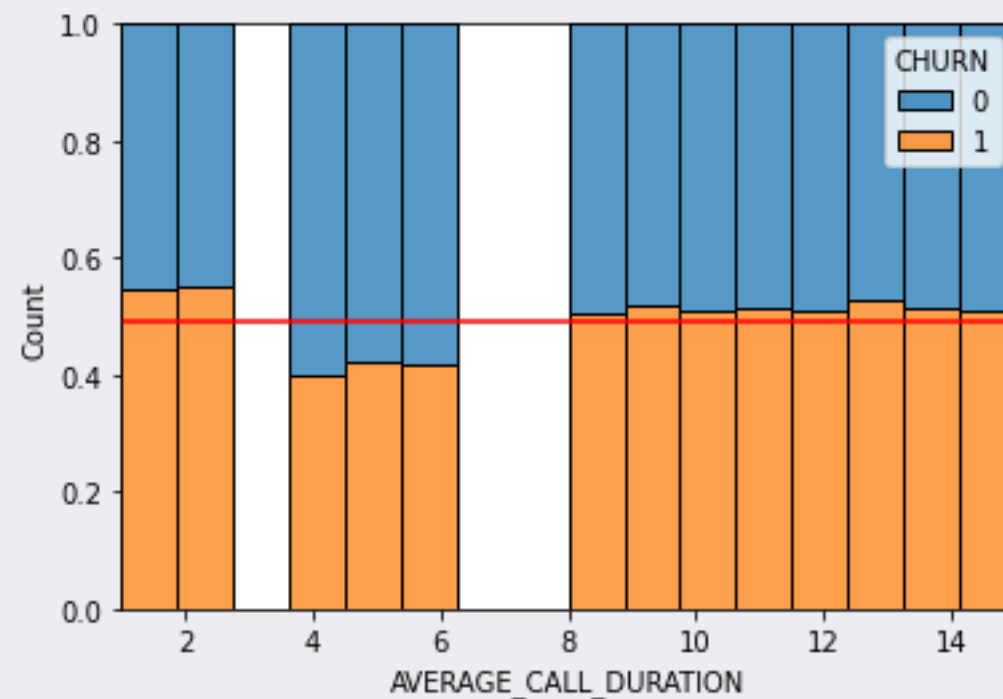
- 휴대폰 가격의 평균은 390달러이며, 대부분의 분포가 평균 이내에 분포되어 있다. 핸드폰 가격이 높은 사람일수록 번호이동을 할 확률이 높을 것으로 예상된다.
- 휴대폰 가격 400달러를 기준으로 이탈 여부율도 달라진다. 휴대폰 가격이 비쌀 수록 이탈율이 높다. p-value도 0.05 미만으로 휴대폰 가격과 이탈 여부 간 관계는 어느 정도 있어 보이지만, 그렇게 큰 관계는 아닌 것으로 보인다.

#### 추가 분석

- 휴대폰 가격을 400달러 미만, 400달러 이상으로 범주로 바꾸어 분석할 필요가 있다. 왜냐하면 수치화에서 사용된 로지스틱 회귀는 미분책이기 때문이다.

### 8. AVERAGE\_CALL\_DURATION(평균 통화시간) --> CHURN

```
Optimization terminated successfully.  
    Current function value: 0.693027  
    Iterations 3  
AVERAGE_CALL_DURATION    0.03783  
dtype: float64
```



#### 분석 결과

- 평균 통화 시간이 약 7분 아래인 고객들의 이탈 여부는 차이를 보인다.
- 평균 통화 시간과 이탈 여부는 관련이 있으나, 크지 않다.

#### 추가 분석

- 평균 통화시간이 7분 아래인 고객들의 정보만을 따로 모아 분석한 결과, p-value가 0에 수렴하기 때문에 이 고객들의 데이터만을 모아 볼 필요가 있다.



## 03 추가 가설수립

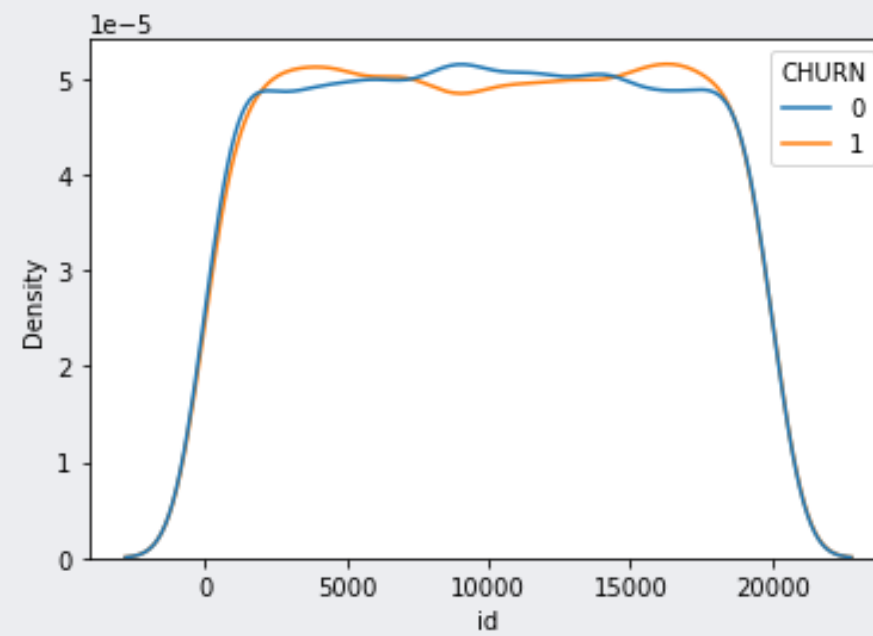
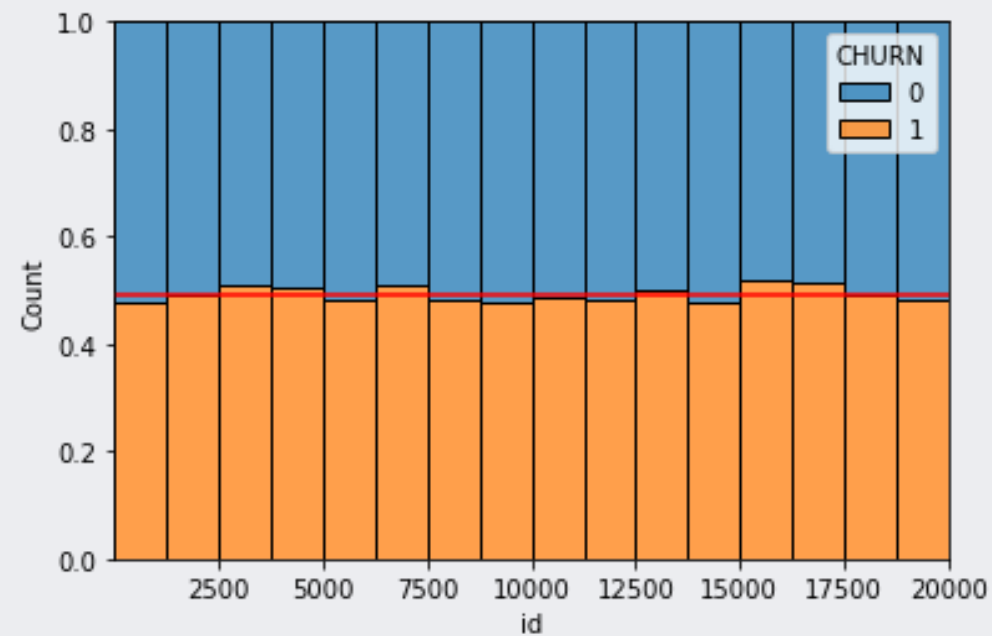


1. id에 따라 CHURN에 차이가 있다.
2. MOBILE\_PLAN에 따라 CHURN에 차이가 있다.
3. LEFTOVER에 따라 CHURN에 차이가 있다.
4. HANDSET\_PRICE에 따라 CHURN에 차이가 있다.

## 03 추가 가설수립

### 1. id --> CHURN

```
Optimization terminated successfully.  
Current function value: 0.693088  
Iterations 2  
id 0.144972  
dtype: float64
```

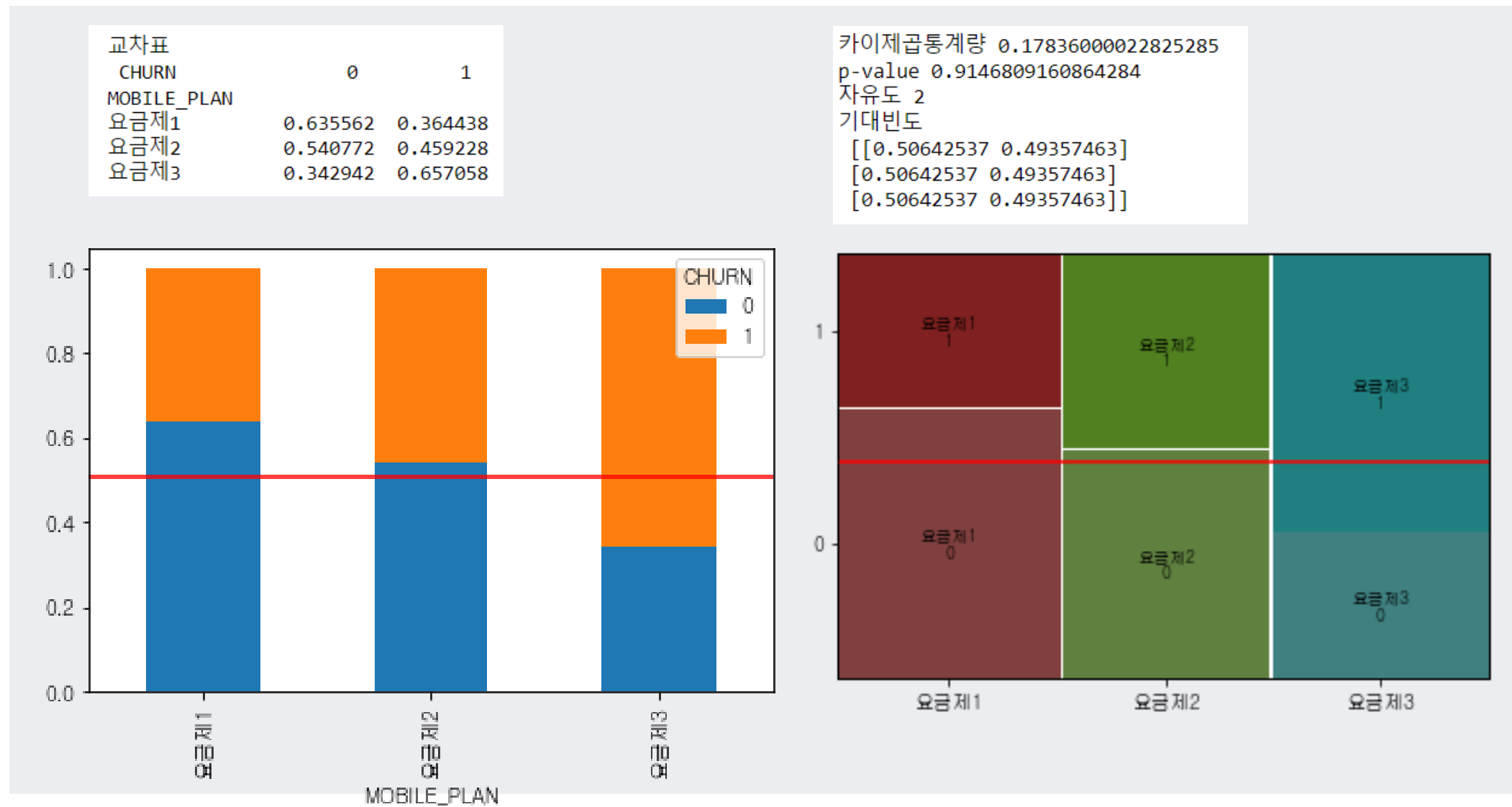


#### 분석 결과

- id를 고객의 가입 순서라 가정하고 데이터를 분석한 결과, 이탈 여부와는 큰 관계가 없는 것을 확인했다.
- id가 고객의 가입 순서라는 가정의 오류인지, 혹은 가입 순서가 맞으나 이와 이탈 여부가 관련이 없는 데이터인지 확인할 필요가 있다.

## 03 추가 가설수립

### 2. MOBILE\_PLAN --> CHURN

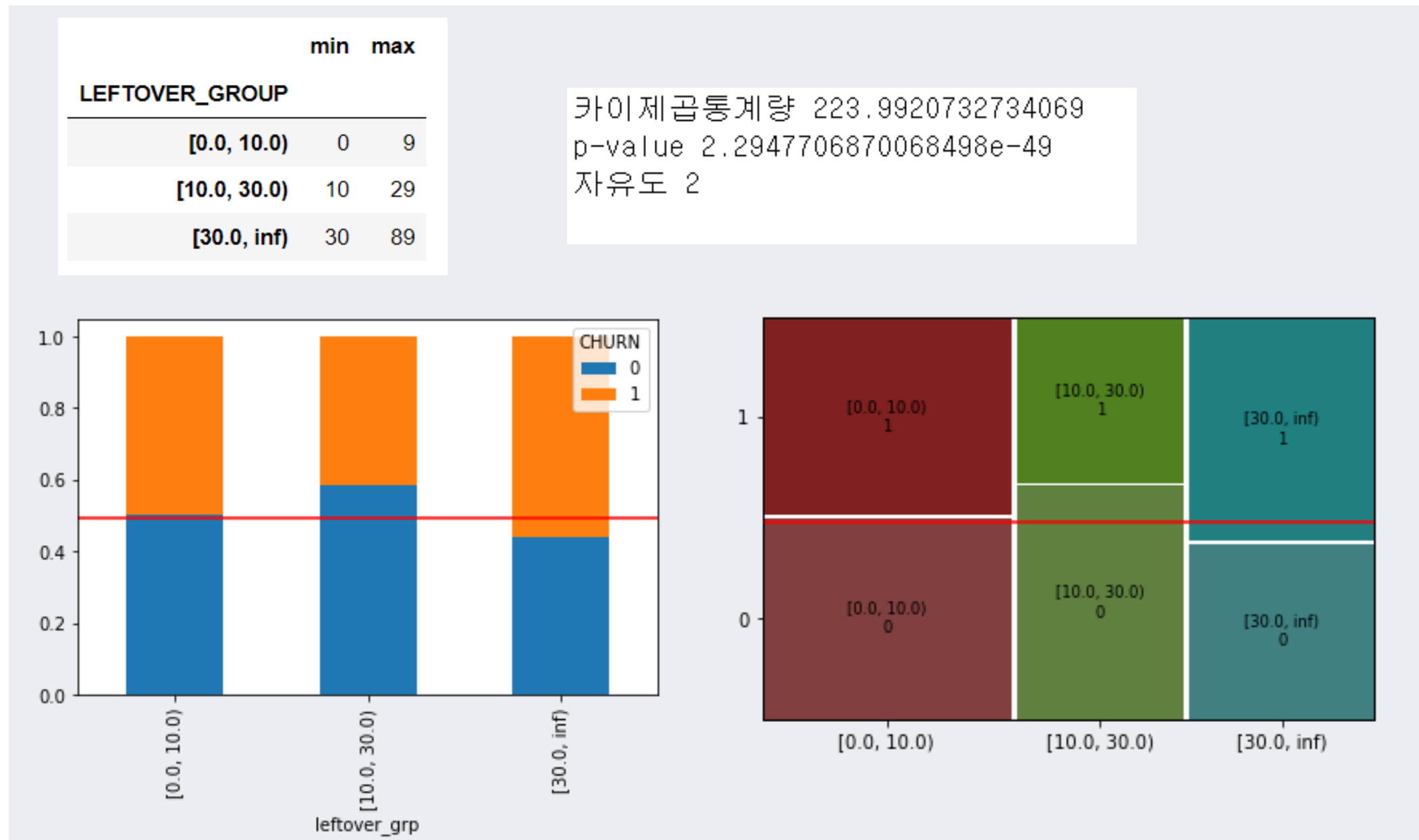


#### 분석 결과

- OVERAGE 단변량 분석 결과, 요금제를 ~ 30, 30~100, 100 ~ 구간으로 나눠서 분석해볼 필요가 있어 보였다.
  - 요금제에 따라 구간이 구분된 것으로 보였다.
- 예상과 다르게 카이제곱검정에서 카이제곱통계량, p-value, 자유도 등을 살펴보면 요금제와 이탈여부는 상관 없는 것으로 나타났다.
  - 따라서 MOBILE\_PLAN에 따른 CHURN은 차이가 없다(귀무가설 채택)



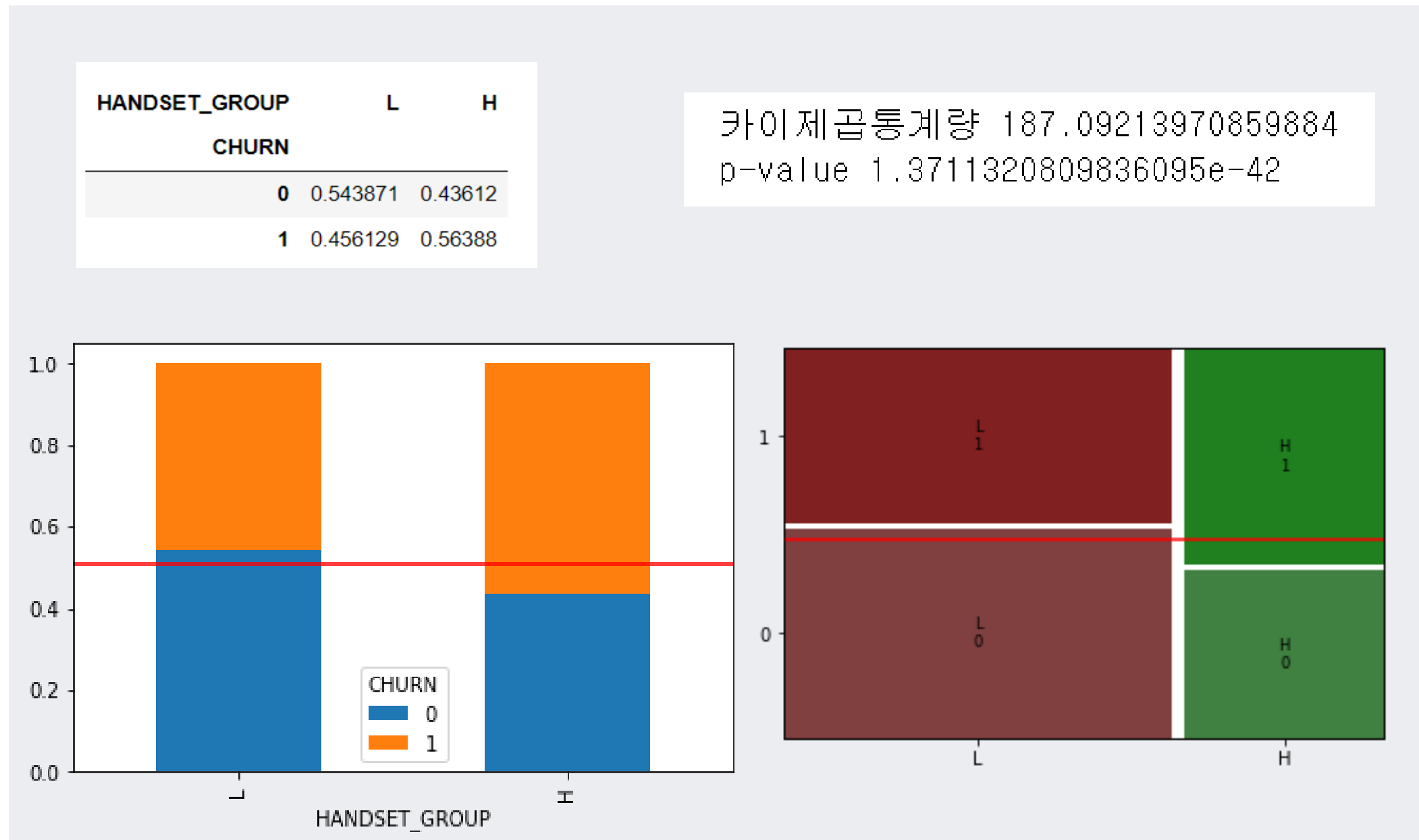
## 3. LEFTOVER\_GROUP--&gt; CHURN



## 분석 결과

- 기존의 핸드폰 가격은 로지스틱 회귀를 사용하였고, 또 p-value가 그렇게 낮지 않았다.
- 그래서 추가분석으로 월 사용 잔여시간을 10%이내, 30% 이내, 30% 이상의 그룹으로 나누어 비교해 본 결과, 매우 유의미한 차이가 있다는 것을 알 수 있다.

## 4. HANDSET\_GROUP--> CHURN



### 분석 결과

- 기존의 핸드폰 가격은 로지스틱 회귀를 사용하였고, 또 p-value가 그렇게 낮지 않았다.
- 그래서 추가분석으로 핸드폰의 가격을 400달러를 기준으로 낮은 그룹, 높은 그룹으로 나누어 비교해 본 결과, 매우 유의미한 차이가 있다는 것을 알 수 있다.

### 추가 분석

- 400달러 이내의 휴대폰은 보통 저렴한 보급형의 휴대폰이거나 알뜰폰일 확률이 높다. 400달러 이내의 휴대폰을 구매하는 고객의 이탈율보다 더 비싼 휴대폰을 구매하는 고객의 이탈율이 높은 만큼 고가의 휴대폰에 알맞은 요금제를 제공하거나 혜택을 제공하는 등의 노력이 필요해 보인다.



## 04 관계정리



관계 없음

COLLEGE  
REPORTED\_SATISFACTION  
CONSIDERING\_CHANGE\_OF\_PLAN  
Id  
MOBILE\_PLAN

중간 관계

INCOME  
LEFTOVER  
HANDSET\_PRICE  
AVERAGE\_CALL\_DURATION

강한 관계

OVERAGE  
LEFTOVER\_GROUP  
HANDSET\_GROUP



## 05 느낀점

- 기존 요인들을 분석하는 것 외에도 FE(Feature Engineering)를 하기 위한 노력이 더 필요하다고 생각한다.
- 설문조사의 결과는 그 방식에 따라 큰 차이를 보일 수 있기 때문에, 그 세부적인 방식을 정함에 있어 신중해야 한다.
- 예상과 다른 결과가 나올 때 흥미로웠고 왜 그런 지에 대해 생각해볼 때 재밌었다.



Thank you