

KT AIVLE School



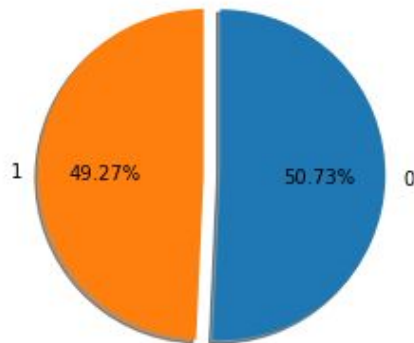
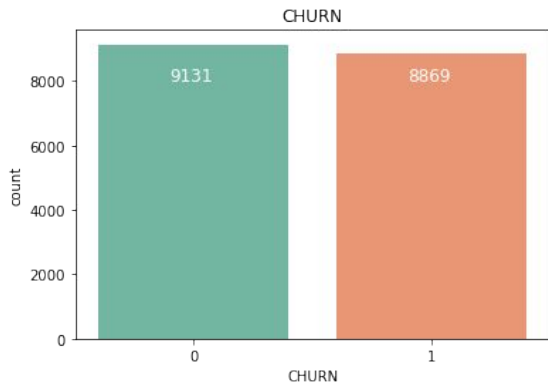
[미니 프로젝트] 이동통신 고객 이탈 분석

수도권 2반 3조(이세빈, 조혜빈, 차선홍, 차혜돈, 최기호)

01. Target

01. Target

CHURN(이탈여부)



- ✓ 잔존과 이탈 비율 차이가 크지 않다.
- ✓ 절반에 가까운 고객들이 이탈하는 것은 회사에 큰 손해로 작용할 수 있으므로 대책 검토가 시급하다.

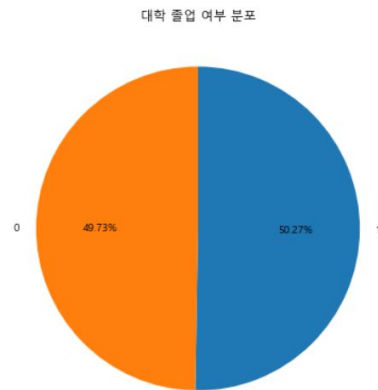
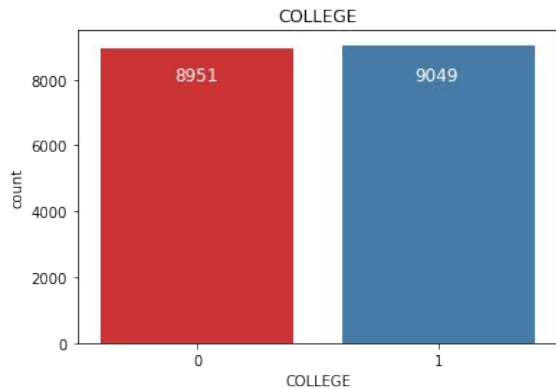
02. COLLEGE

(대학졸업 여부)

가설 : 대학졸업 여부가 고객 이탈 여부에 영향을 미칠 것이다.

02. COLLEGE

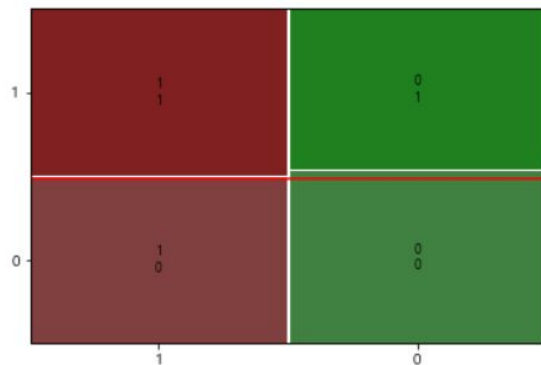
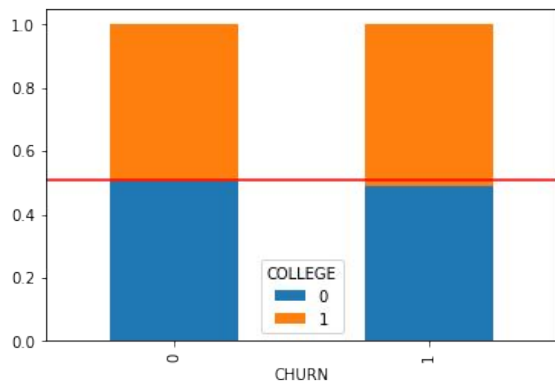
단변량 분석



- ✓ 대학을 졸업한 고객과 졸업하지 않은 고객의 비율이 거의 비슷하다.
- ✓ 졸업 유무에 따라 데이터를 분리하여 분석해 볼 필요가 있다.

02. COLLEGE

이변량 분석 (COLLEGE > CHURN)



CHURN	0	1
COLLEGE		
0	4616	4335
1	4515	4534

카이제곱 통계량 : 4.982102934693502
P-value : 0.02561083347822053
기대빈도 : [[4540.64338889 4410.35661111]
[4590.35661111 4458.64338889]]

✓ 그래프 상으로는 평균과 가까워 대학 졸업 여부와 고객 이탈과는 관계가 없어 보인다.

✓ 검정 결과 유의수준이 0.05보다 작지만 보수적으로 접근하였을 때 0.01보다 크기 때문에 약한 관계라고 할 수 있다.

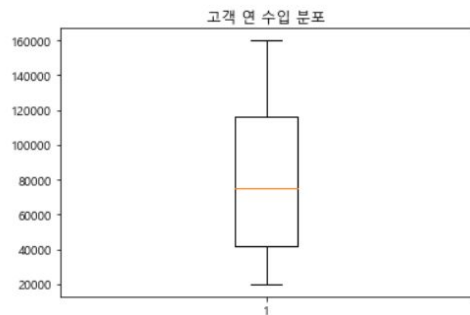
03. INCOME

(연 수입액)

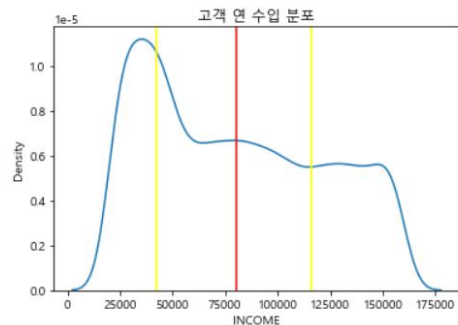
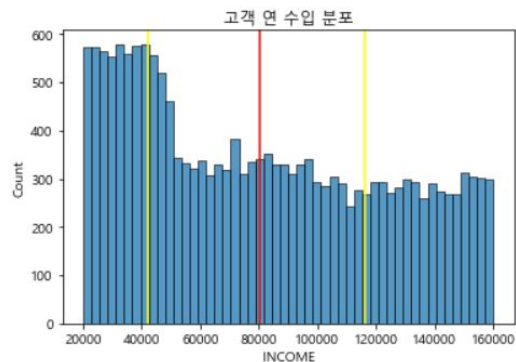
가설 : 연 수입액이 고객 이탈여부에 영향을 미칠 것이다.

03. INCOME

단변량 분석



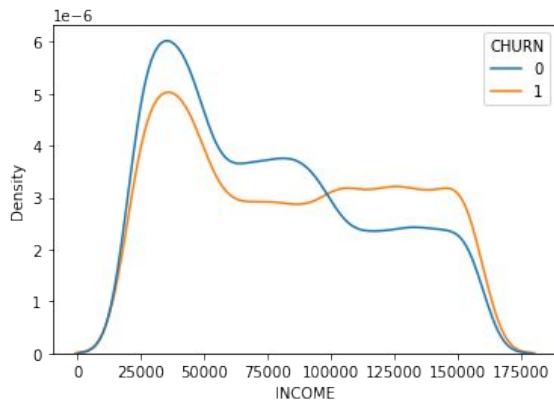
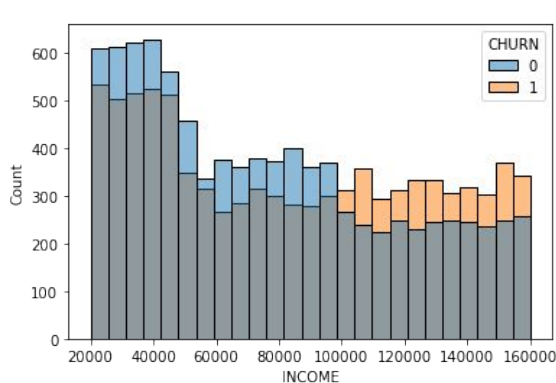
[42151.5 20007.]
[116082.25 159983.]



- ✓ 연 수입액은 50,000달러 기준으로 빈도수가 급격하게 변하는 것으로 볼 때 데이터를 나누어 비교 분석해 볼 필요가 있다.
- ✓ 데이터 분포가 약간 왼쪽으로 치우쳐있으나 대체로 고르게 분포되어 있다.

03. INCOME

이변량 분석 (INCOME > CHURN)



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

- ✓ 약 100,000달러 이상의 수입을 버는 사람들의 이탈이 더 많다.
- ✓ 이탈이 더 많아지는 부분을 기점으로 데이터를 분리하여 추가 분석할 필요가 있다
- ✓ 검정 결과 **p-value**의 값이 충분히 작기 때문에 연 수입액이 고객의 이탈 여부에 영향을 미친다고 할 수 있다.

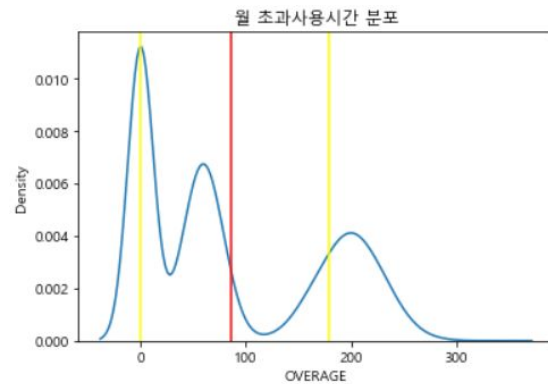
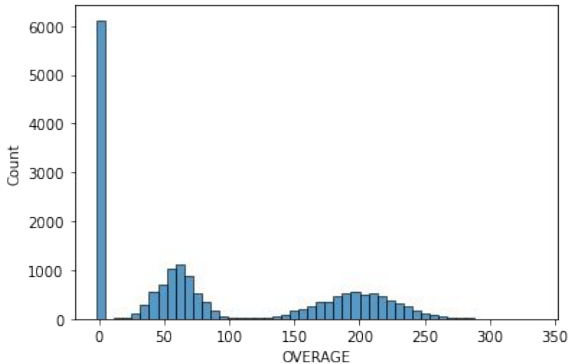
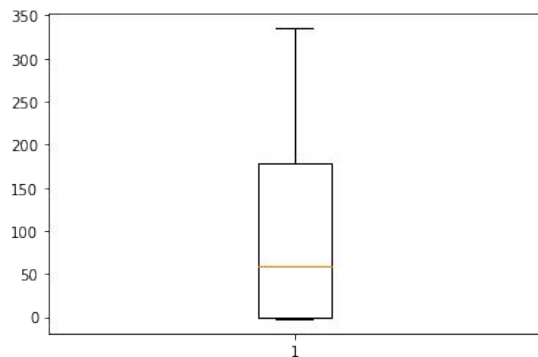
04. OVERAGE

(월 초과사용 시간)

가설 : 월 초과사용 시간이 고객 이탈여부에 영향을 미칠 것이다.

04. OVERAGE

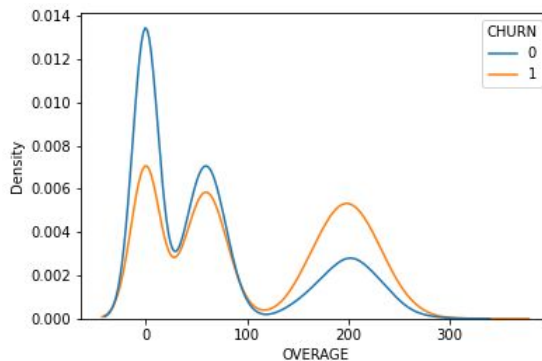
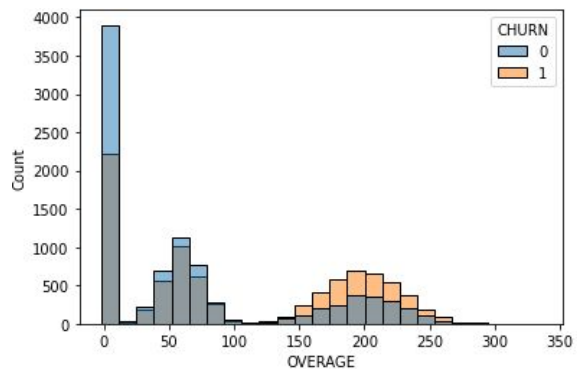
단변량 분석



- ✓ -2라는 이상치의 사용시간이 나타났기 때문에 이 데이터에 대한 처리 방안이 필요하다.
- ✓ 고객의 25%가량이 초과 사용시간이 0으로 나타났기 때문에 이 데이터를 제외하고 분석해 볼 필요가 있다.

04. OVERAGE

이변량 분석 (OVERAGE > CHURN)



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

- ✓ 초과시간 대략 100분 정도를 기준으로 이탈 여부에 큰 차이를 보이고 있다.
- ✓ 검정 결과 p-value의 값이 0에 가깝기 때문에 월 초과 사용시간이 이탈 여부에 영향을 미친다고 할 수 있다.

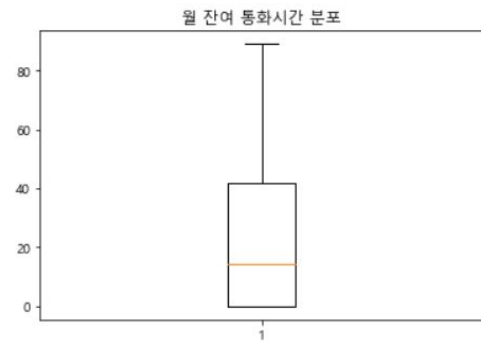
05. LEFTOVER

(월 사용 잔여시간)

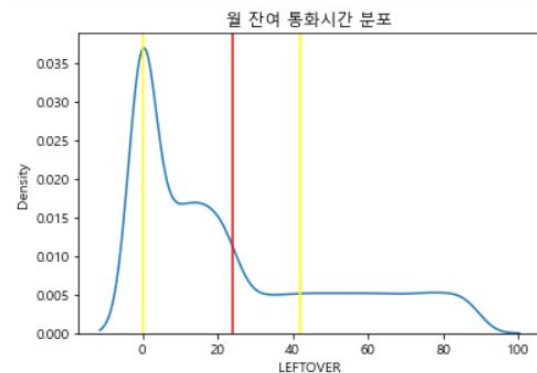
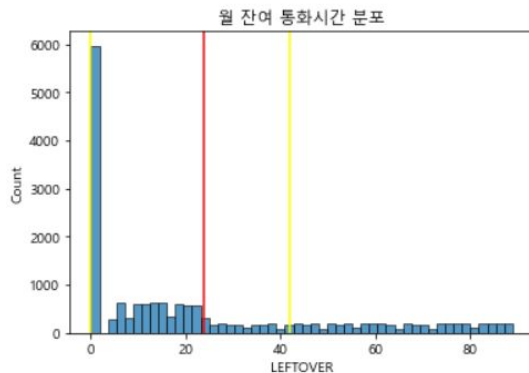
가설 : 월 잔여사용시간이 고객 이탈여부에 영향을 미칠 것이다.

05. LEFTOVER

단변량 분석



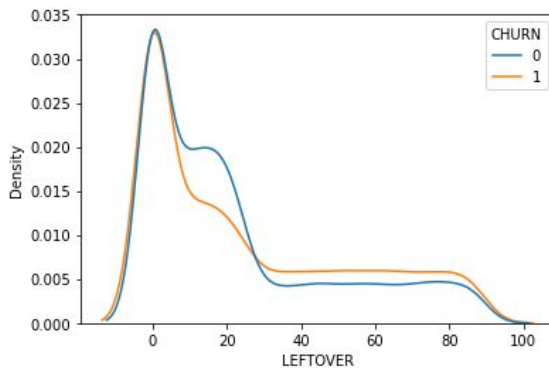
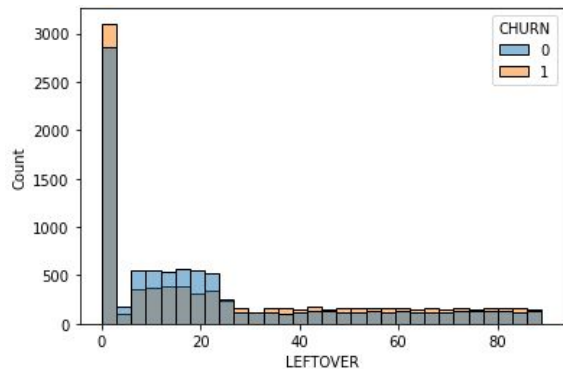
[0, 0.]
[42, 89.]



- ✓ 잔여시간이 전혀 남아있지 않은 고객이 전체의 25%를 차지하며 이 부분이 이탈 여부와 관련이 있는지 분석해 보아야 한다.
- ✓ 그래프가 전체적으로 왼쪽으로 치우친 경향이 있어 고객들에게 제공되는 통화시간이 부족한 것은 아닌지 추가 조사가 필요하다.
- ✓ 잔여시간이 0인 사람들과 아닌 사람들을 나누어 비교 분석할 필요가 있다.

05. LEFTOVER

이변량 분석 (LEFTOVER > CHURN)



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

- ✓ 잔여시간이 없는 고객 제외, 대략 25분 이하 정도 잔여시간이 있는 고객들의 경우, 잔류하는 고객이 이탈하는 고객보다 많다.
- ✓ 검정 결과 p-value의 값이 0에 가깝기 때문에 잔여시간이 고객의 이탈 여부에 영향을 미친다고 할 수 있다.

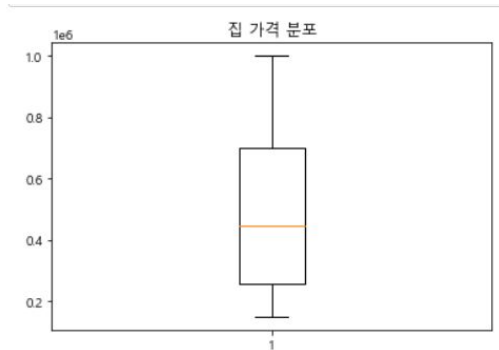
06. HOUSE

(집 가격)

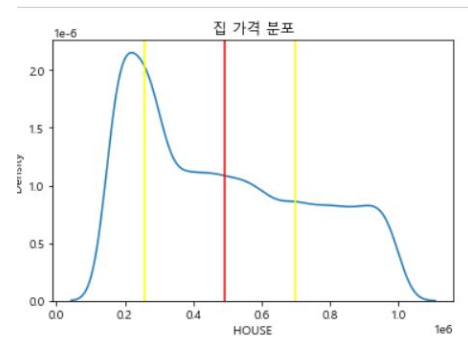
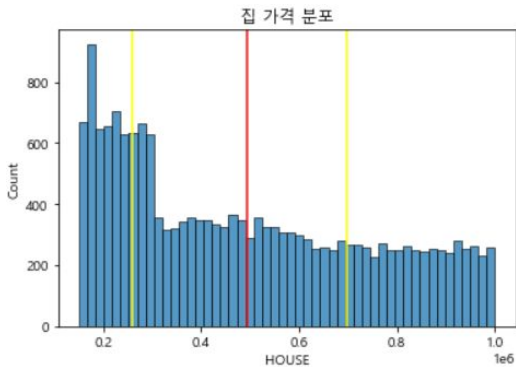
가설 : 집 가격이 고객 이탈여부에 영향을 미칠 것이다.

06. HOUSE

단변량 분석



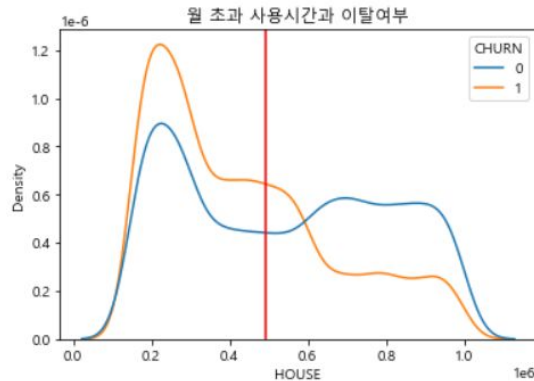
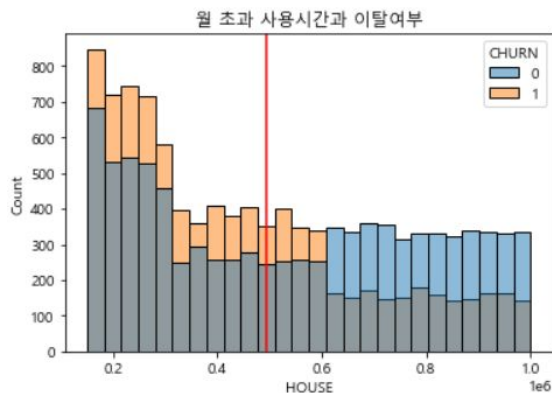
```
[259092.5 150015. ]  
[698473.5 999970. ]
```



✓ 집값은 대체로 고르게 분포되어 있으나 300,000달러 이하에서 빈도수가 높게 나타난다.

06. HOUSE

이변량 분석 (HOUSE > CHURN)



```
Optimization terminated successfully.  
Current function value: 0.687308  
Iterations: 2  
HOUSE      3.850174e-47  
dtype: float64
```

- ✓ 집값 600,000 달러 구간을 기준으로 이탈 여부에 두드러지는 차이가 보인다.
- ✓ 이탈 여부가 바뀌는 구간을 기준으로 나누어 분석을 진행해 볼 필요가 있다.
- ✓ 검정 결과 p-value값이 0에 가깝기 때문에 집값이 고객 이탈 여부에 영향을 미친다고 할 수 있다.

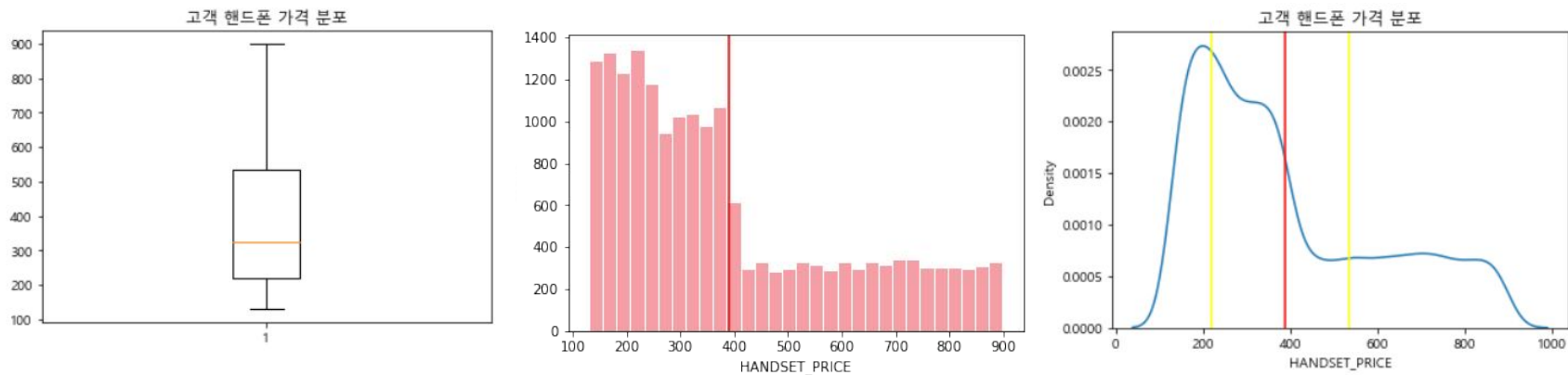
07. HANDSET_PRICE

(핸드폰 가격)

가설 : 핸드폰 가격이 고객 이탈여부에 영향을 미칠 것이다.

07. HANDSET_PRICE

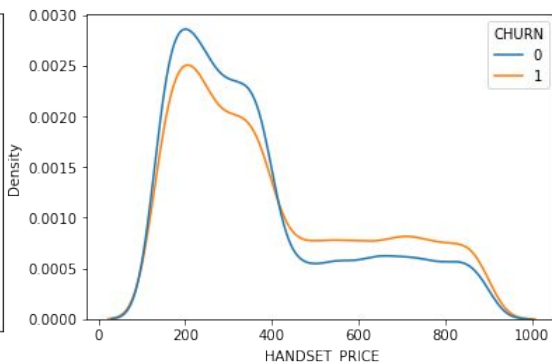
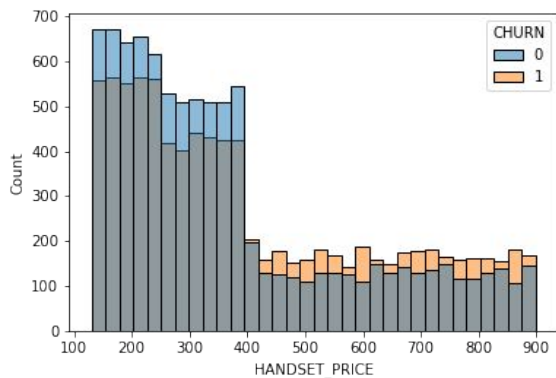
단변량 분석



- ✓ 핸드폰 가격은 약 400달러 기준으로 값이 극명하게 나뉘기 때문에 고객 데이터를 두 가지로 나누어 분석해 볼 필요가 있다.
- ✓ 고객의 50%가 326달러 이하의 핸드폰을 사용하고 있고 이것은 평균보다 낮은 수치로 고객들이 대체로 저가 핸드폰을 사용하는 것으로 보인다.

07. HANDSET_PRICE

이변량 분석 (HANDSET_PRICE > CHURN)



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

- ✓ 핸드폰 가격이 약 400달러 근처의 평균선을 기준으로 이탈 여부에도 차이를 보이고 있다.
- ✓ 검정 결과 p-value가 0에 가까워 핸드폰의 가격이 이탈 여부에 영향을 미친다고 할 수 있다.

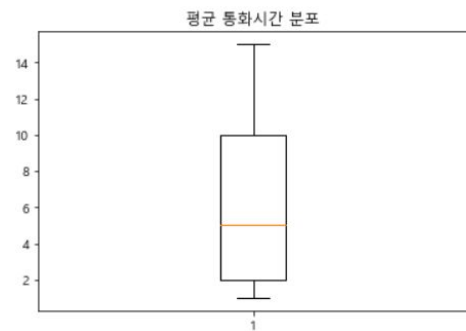
08. AVERAGE_CALL_DURATION

(평균 통화시간)

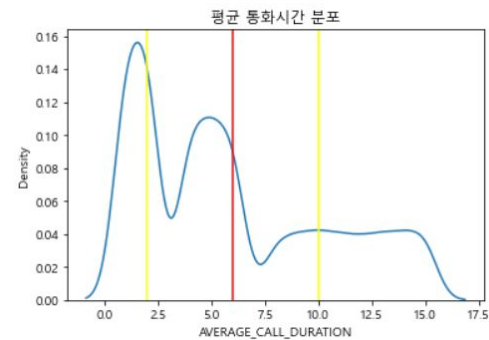
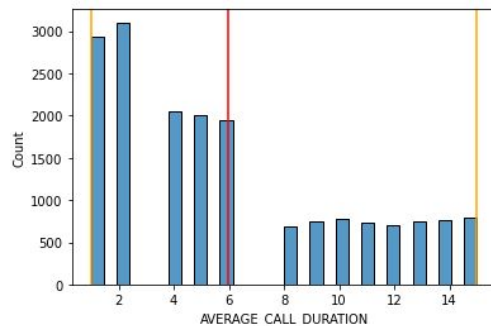
가설 : 평균 통화시간이 고객이탈여부에 영향을 미칠 것이다.

08. AVERAGE CALL DURATION

단변량 분석



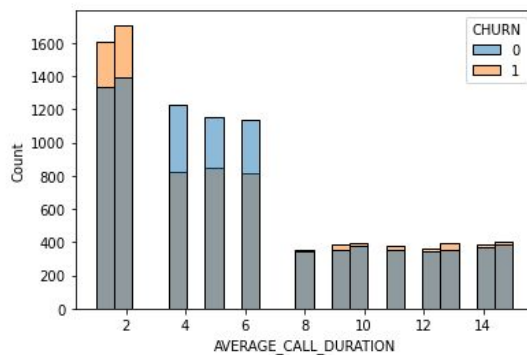
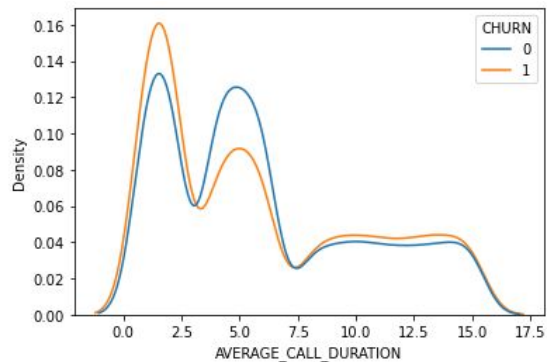
[2. 1.]
[10. 15.]



- ✓ 0~3분, 4~7분, 8분 이상으로 세 집단으로 나누어 분석해 볼 필요가 있다.
- ✓ 전체의 75%가 평균적으로 10분 이내로 통화한다.

08. AVERAGE CALL DURATION

이변량 분석 (AVERAGE CALL DURATION > CHURN)



```
Optimization terminated successfully.  
Current function value: 0.693031  
Iterations 3  
AVERAGE_CALL_DURATION    0.042621  
dtype: float64
```

✓ 그래프상 차이가 있어 보이며, 약 3, 7.5 에서 교차한다.

✓ 세 구간 중 3~7분 제외, 모두 이탈률이 높으므로 구간별 분석이 필요하다.

✓검정 결과 p-value 값이 0.04262로 0.05보다는 작으므로 평균 통화시간이 고객 이탈 여부에 약한 영향을 미칠 수 있다.

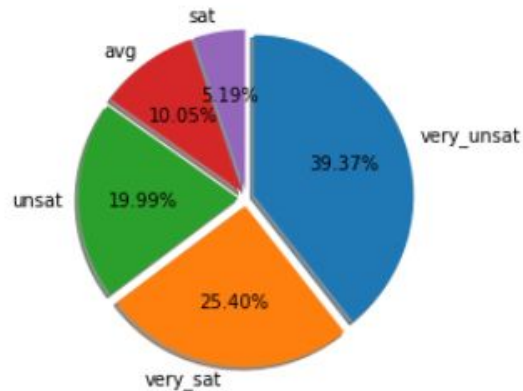
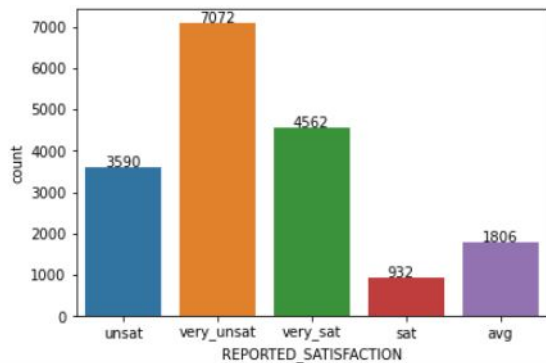
09. REPORTED_SATISFACTION

(만족도 설문)

가설 : 만족도 설문이 고객 이탈여부에 영향을 미칠 것이다.

09. REPORTED SATISFACTION

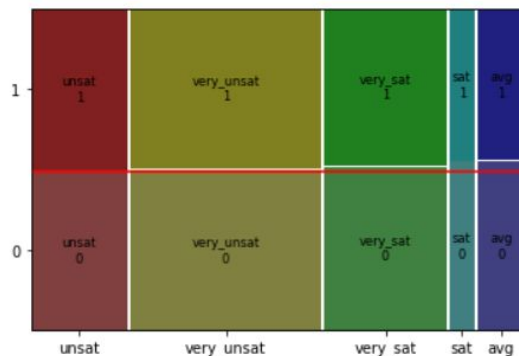
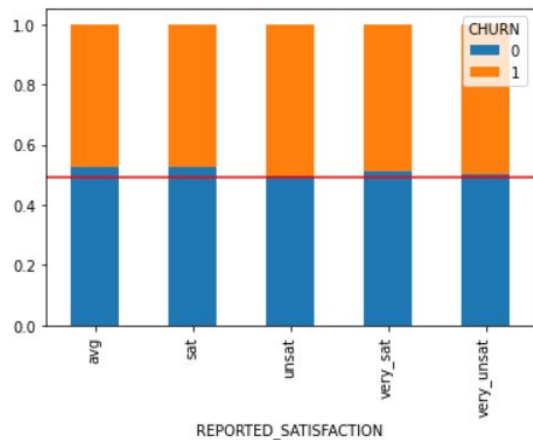
단변량 분석



- ✓ 부정적인 평가(매우 불만족, 불만족)를 한 고객이 전체 59.36%로 과반 이상을 차지한다.
- ✓ 부정적인 평가를 한 고객들의 이유를 조사해 볼 필요가 있다.
- ✓ 부정적인 평가를 한 고객들의 실제 이탈 여부를 분석해 볼 필요가 있다.
- ✓ 결측치가 존재하므로 결측치를 고려한 분석이 필요하다.

09. REPORTED SATISFACTION

이변량 분석 (REPORTED SATISFACTION > CHURN)



카이제곱 값 : 6.841100957205645

p-value : 0.14452738878905058

자유도 : 4

기대도수

```
[[ 916.27201871  889.72798129]
 [ 472.84912593  459.15087407]
 [1821.38236277 1768.61763723]
 [2314.5254426  2247.4745574 ]
 [3587.97104999 3484.02895001]]
```

✓ 그래프를 보면 대부분 평균에 걸쳐있다.

✓ 검정 결과 p-value가 0.14453 으로 큰 편으로 만족도 설문이 이탈 여부에 영향을 준다 보기 어렵다.

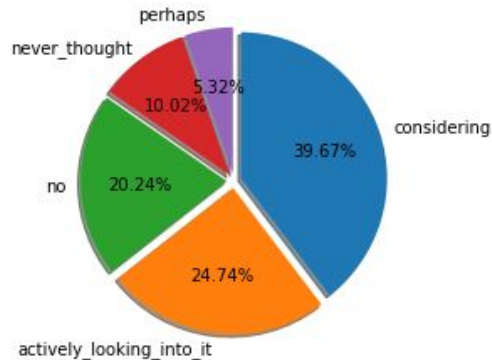
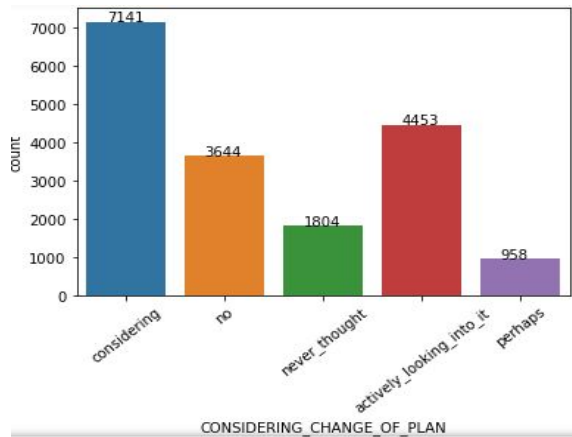
10. CONSIDERING_CHANGE_OF_PLAN

(변경 계획 설문)

가설 : 변경 계획 설문이 고객 이탈여부에 영향을 미칠 것이다.

10. CONSIDERING CHANGE OF PLAN

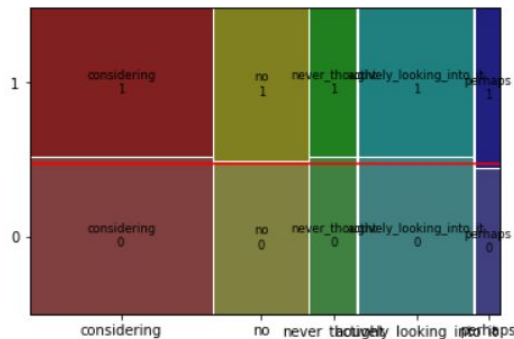
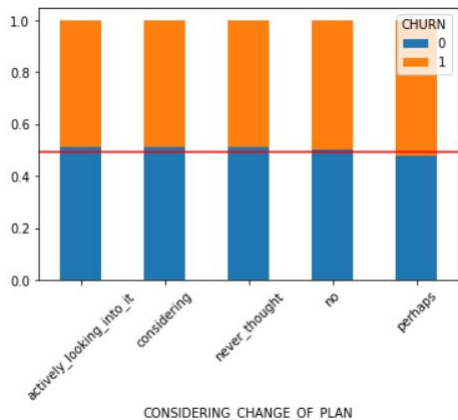
단변량 분석



- ✓ '절대 아니다', '아니다' 제외, 변경 가능성이 있는 고객의 비율이 69.73%로 높은 편이다.
- ✓ 회사는 변경할 가능성이 있는 고객들을 유지시킬 수 있는 방안을 모색해야 한다.
- ✓ 변경계획 조사와 실제 변경 조사를 비교하여 변경계획 조사의 신빙성을 검토해 볼 필요가 있다

10. CONSIDERING CHANGE OF PLAN

이변량 분석 (CONSIDERING CHANGE OF PLAN > CHURN)



카이제곱 값 : 5.421434419620688
p-value : 0.24672225866799702
자유도 : 4
기대도수
[[2258.90794444 2194.09205556]
[3622.47061111 3518.52938889]
[915.12911111 888.87088889]
[1848.52022222 1795.47977778]
[485.97211111 472.02788889]]

✓ 그래프를 보면 대부분 평균에 걸쳐있다.

✓ p-value 가 0.24672 으로 큰 편이므로 변경계획 설문이 고객 이탈여부에 영향을 준다 보기 어렵다.

11. 관계 정리

11. 관계 정리

✓ 강한 관계의 변수

- INCOME(연 수입액), LEFTOVER(월 사용 잔여시간(분)),
HOUSE(집 가격), HANDSET_PRICE(핸드폰 가격) OVERAGE(월 초과사용 시간(분))

✓ 약한 관계의 변수

- COLLEGE(대학졸업 여부), AVERAGE CALL DURATION(평균 통화시간 (분))

✓ 관계 없음

- REPORTED SATISFACTION(만족도 설문), CONSIDERING CHANCE OF PLAN(변경 계획 설문)