

고객 이탈 분석

전남 전북 1반 1조

팀원 소개



김경준



박민지



안지예



임채걸



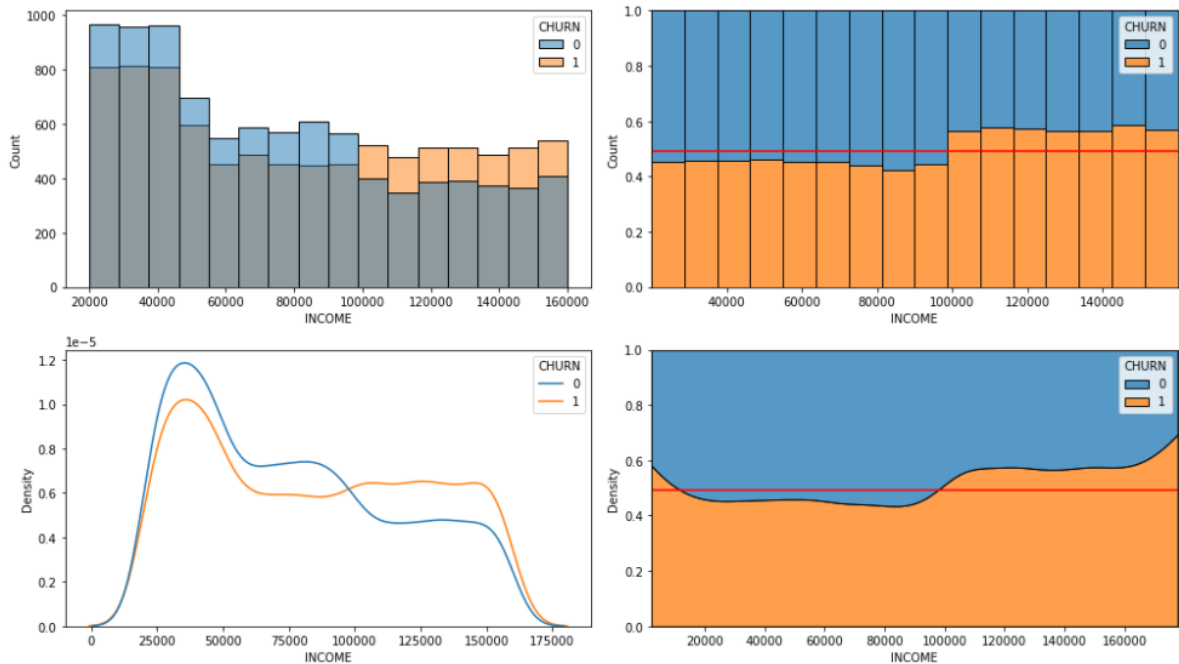
조훈근

Part 1

이번량 분석 - 숫자 vs 범주



2.2 연 수입액 - 이탈 여부



histogram, densityplot

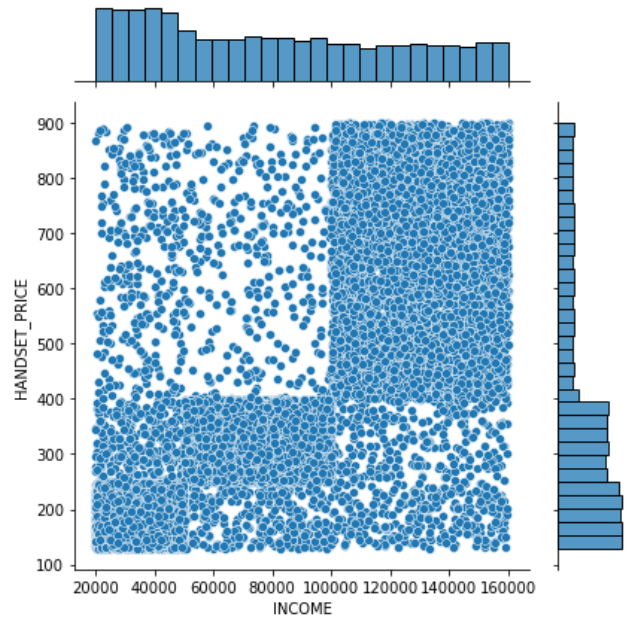
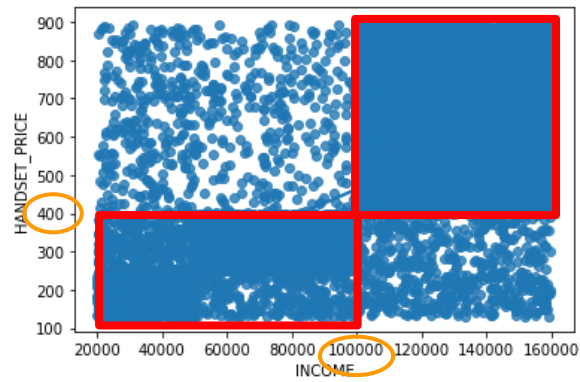
```
Optimization terminated successfully.  
Current function value: 0.692683  
Iterations 2  
INCOME 0.000044  
dtype: float64
```

로지스틱회귀

- 연 수입 100,000 달러 이상부터 이탈자가 더 많음
- 이탈하는 이유?
 - 더 좋은 통신사를 찾아 이탈
 - 핸드폰 가격과 관계 추가 분석
- 로지스틱회귀의 p 값이 0.05보다 작음 \Rightarrow 중간 관계

2.2 핸드폰 가격 + 연 수입액 - 이탈 여부

1. 핸드폰 가격과 연 수입액의 관계



- 앞서 이변량 분석에서 봤던 것과 같이 구간별로 나뉜 것을 볼 수 있음

2. 숫자 → 범주

- 핸드폰 가격은 400 달러 전후로 cut
- 연 수입액은 100,000 달러 전후로 cut

```
var = 'Handset_Group'

bins = [130, 400, 900]
labels = ['HL', 'HH']
mobile[var] = pd.cut(mobile['HANDSET_PRICE'], bins=bins, labels=labels, right=False)

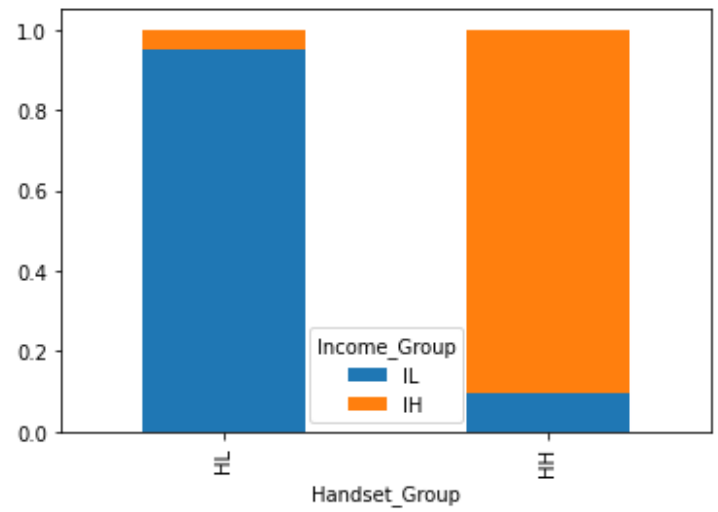
mobile.head()

var = 'Income_Group'

bins = [-np.inf, 100000, np.inf]
labels = ['IL', 'IH']
mobile[var] = pd.cut(mobile['INCOME'], bins=bins, labels=labels, right=False)

mobile.head()
```

3. 시각화



4. 카이제곱검정

카이제곱통계량 13139.33369248936
p-value 0.0
자유도 1
기대빈도
[[7860.60894444 4042.39105556]
[4026.39105556 2070.60894444]]

- 연 수입액이 낮은 그룹은 핸드폰 가격 또한 낮고, 연 수입액이 높은 그룹은 핸드폰 가격 또한 높은 것을 볼 수 있다.

2.2 핸드폰 가격 + 연 수입액 - 이탈 여부

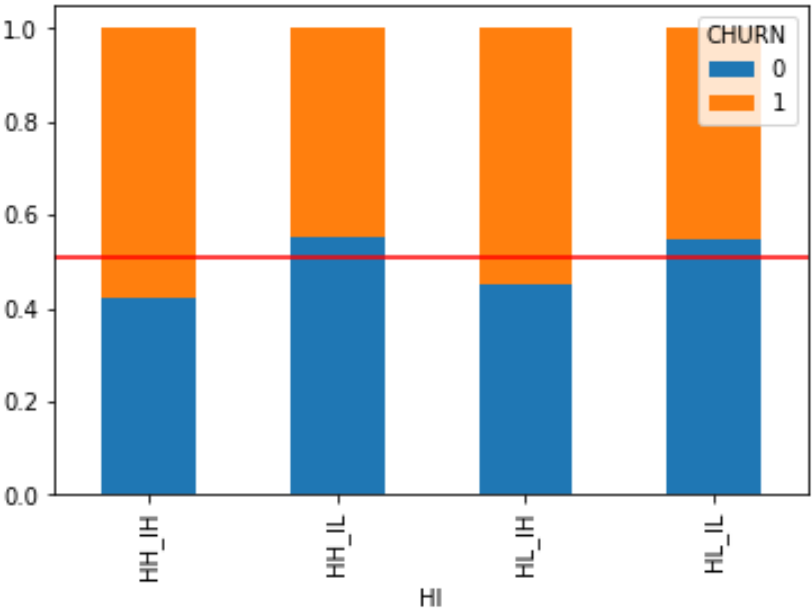
5. 핸드폰 가격과 연 수입액을 하나의 그룹으로 묶기

```
mobile['HI'] = 'HL_IL'
mobile.loc[(mobile['Handset_Group'] == 'HL') & (mobile['Income_Group'] == 'IH'), 'HI'] = 'HL_IH'
mobile.loc[(mobile['Handset_Group'] == 'HH') & (mobile['Income_Group'] == 'IL'), 'HI'] = 'HH_IL'
mobile.loc[(mobile['Handset_Group'] == 'HH') & (mobile['Income_Group'] == 'IH'), 'HI'] = 'HH_IH'
mobile.head()
```

6. 이탈여부와 관계 수치화

CHURN	0	1
HI		
HH_IH	0.423523	0.576477
HH_IL	0.552941	0.447059
HL_IH	0.449050	0.550950
HL_IL	0.548727	0.451273

7. 시각화

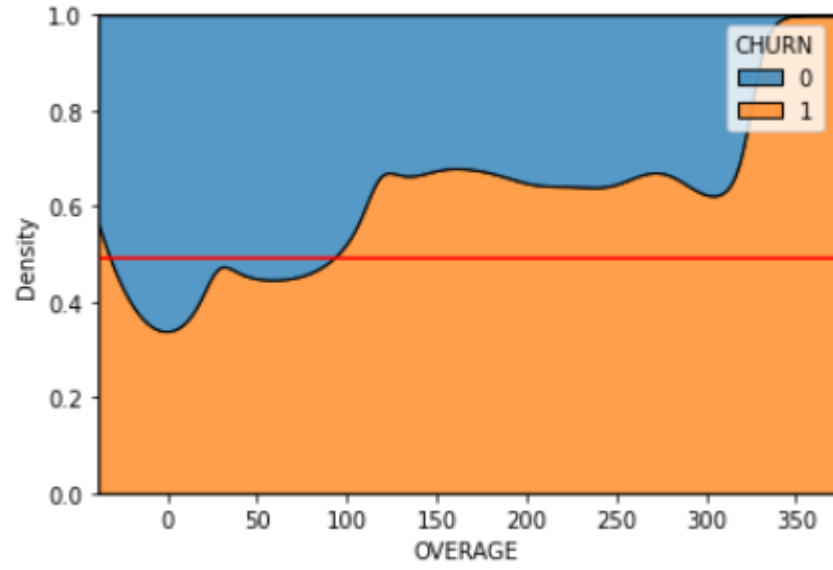
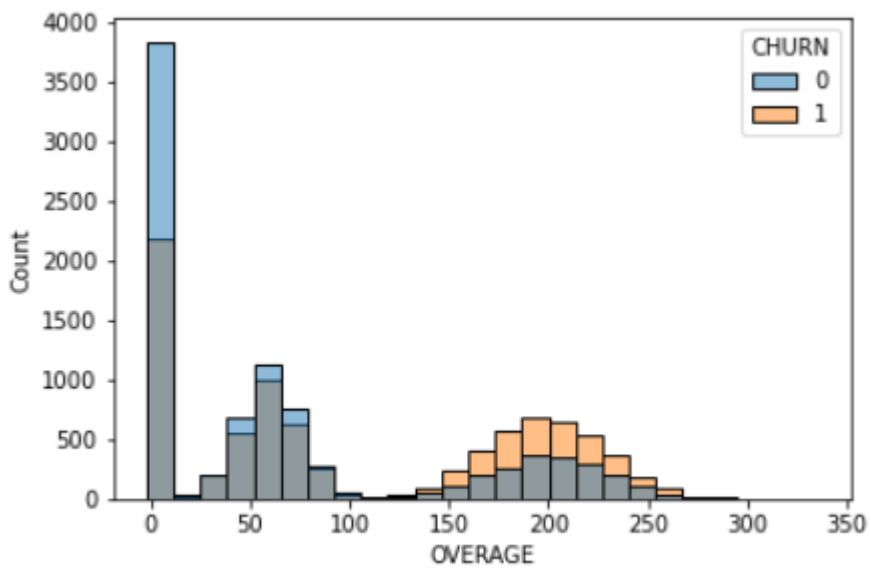


8. 카이제곱검정

카이제곱통계량 245.40761510993707
p-value 6.442531397969708e-53
자유도 3
기대빈도
[[2799.15877778 301.83027778 293.71383333 5736.29711111]
[2718.84122222 293.16972222 285.28616667 5571.70288889]]

- 핸드폰 가격, 연 수입액 따로 봤을 때보다 더 강한 관계성을 드러냄.
- 핸드폰 가격과 상관 없이 연 수입액이 높은 사람들이 이탈자가 많고, 연 수입액이 상대적으로 낮은 사람들이 이탈자가 더 적음

2.3 월 초과사용 시간 - 이탈여부



Optimization terminated successfully.
Current function value: 0.680632
Iterations 4
OVERAGE 2.805475e-95
dtype: float64

- 그래프상

- 월 초과사용 시간이 많을수록 이탈자가 많다.
- 월 초과사용 시간이 최대치이면 무조건 이탈인거 같다.
- 월 초과사용 시간이 적을수록 이탈자가 적은편이다.

- 로지스틱회귀

- p 값은 0에 수렴

종합의견

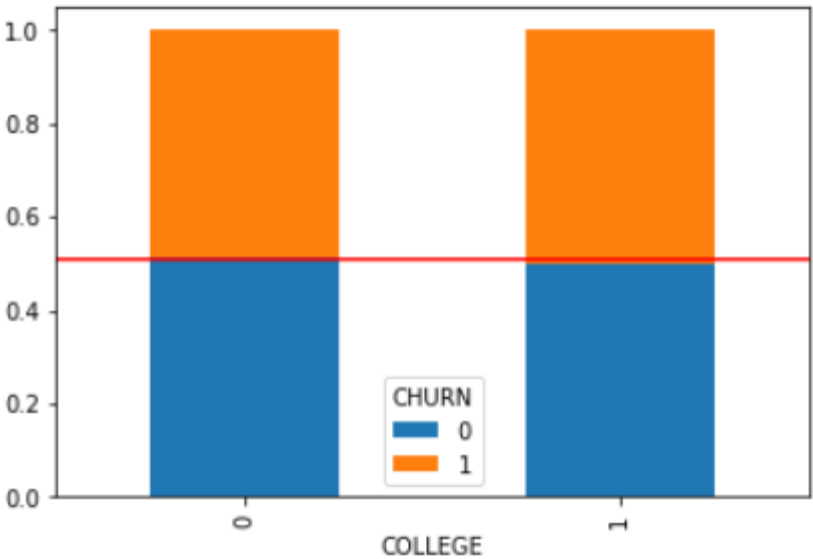
- 초과 사용 시간이 많아 질수록 이탈율이 높아진다.
- 최대치이면 무조건 이탈(350~)
- 초과사용 시간이 적을수록 이탈자가 적은편
- 평균통화시간과 비교를 해보는건 어떨지

Part 2

이번량 분석 - 범주 vs 범주



대학 졸업 여부 - 이탈 여부



카이제곱통계량 5.399292741405257
p-value 0.020144913376353327
자유도 1
기대빈도
[[4532.65460416 4401.34539584]
[4580.34539584 4447.65460416]]

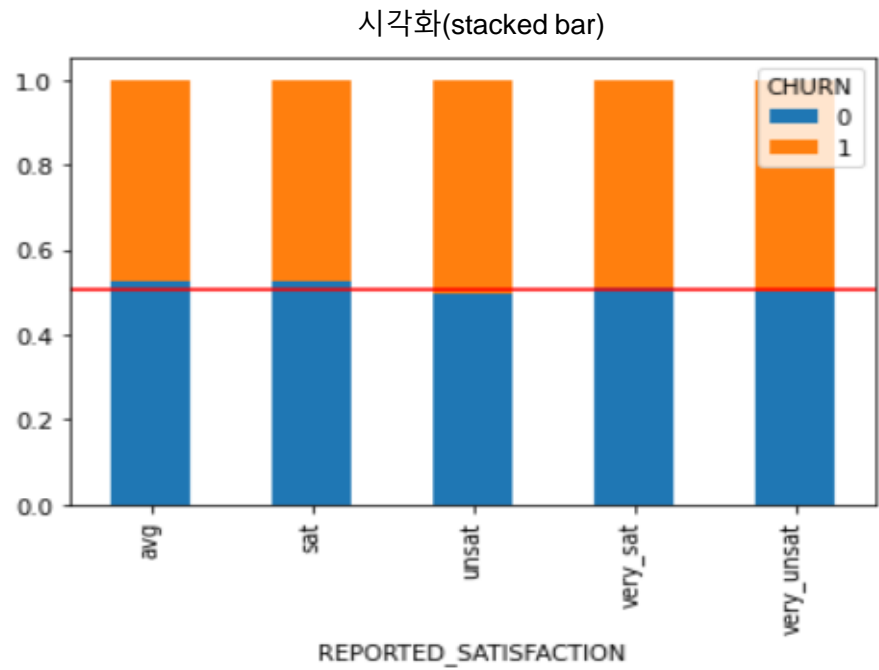
- 대학 졸업 여부와 이탈여부는 중간정도의 상관관계가 있다.
- 시각화 차트와 카이제곱검정 결과 p-value가 0.02로 0.05보다 낮기 때문에 귀무가설이 기각 된다.
- 대학 졸업 후 이탈율이 재학중 이탈율 보다 조금 높다.
- 왜 대학 졸업 후 이탈여부가 높을까?
 - 대학 졸업 후 취업으로 인해 돈을 벌기 때문에 더 좋은 통신사로 이탈할 확률이 높다. -> 연 수입 액과 비교
 - 재학 중엔 부모님의 의사로 통신사를 결정 하였지만, 졸업 후엔 자신이 원하는 통신사로 이동했을 가능성이 있다.
 - 졸업 후 새로운 마음가짐으로 통신사를 이동할 가능성이 있다.

2.5 만족도 설문 - 이탈 여부

```
pd.crosstab(mobile2[target], mobile2[feature], normalize = 'index')
```

REPORTED_SATISFACTION	avg	sat	unsat	very_sat	very_unsat
CHURN					
0	0.10426	0.055854	0.196255	0.255613	0.388019
1	0.09629	0.051866	0.202729	0.251212	0.397903

교차표(crosstab)



- barplot으로 봤을 때 평균과 만족을 제외하면 전체 평균비율이고 거의 차이가 없음
- 만족도가 불만족인 경우 이탈율이 높다. 하지만 만족도가 매우 높아도 이탈율이 꽤 높다.
- 만족도가 낮지만 왜 이탈하지 못하는가?
 - 경제상황, 대학 졸업
- 만족도가 높지만 왜 이탈하는가?
 - 잔여시간, 월초과 사용량

③ 가설검정(카이제곱검정) - 결측치는 sat로 처리함

```
table = pd.crosstab(mobile2[feature], mobile2[target])

result = spst.chi2_contingency(table)
print('카이제곱통계량', result[0])
print('p-value', result[1])
print('자유도', result[2])
print('기대빈도\n', result[3])
```

```
카이제곱통계량 6.58368596842852
p-value 0.15959327308003016
자유도 4
기대빈도
[[ 916.14366667  889.85633333]
 [ 492.05944444  477.94055556]
 [1821.12722222 1768.87277778]
 [2314.20122222 2247.79877778]
 [3587.46844444 3484.53155556]]
```

• 결측치 있는 상태로 가설검정

```
table = pd.crosstab(mobile[feature], mobile[target])

result = spst.chi2_contingency(table)
print('카이제곱통계량', result[0])
print('p-value', result[1])
print('자유도', result[2])
print('기대빈도\n', result[3])
```

```
카이제곱통계량 6.58368596842852
p-value 0.15959327308003016
자유도 4
기대빈도
[[ 916.14366667  889.85633333]
 [ 492.05944444  477.94055556]
 [1821.12722222 1768.87277778]
 [2314.20122222 2247.79877778]
 [3587.46844444 3484.53155556]]
```

- 자유도는 4이고, 카이제곱통계량은 자유도의 2~3배 이상이 되지 못한다.
- p값이 0.05보다 크다.
- 만족도 설문에 따라 고객 이탈여부는 관계가 없다.(관계 없음)
- 결측치가 있는 상태로도 그냥 돌려보아도 크게 p값의 변화가 없었다.

```
from sklearn.ensemble import GradientBoostingRegressor
from sklearn.metrics import *
model_gb = GradientBoostingRegressor(n_estimators=100,
                                     learning_rate=0.1,
                                     max_depth=5,
                                     min_samples_split=30,
                                     min_samples_leaf=15,
                                     random_state=42)

model_gb.fit(x_train, y_train)
y_pred = model_gb.predict(x_test)
print('r2:', r2_score(y_test, y_pred))
```

r2: 0.3178305509267876

2.6 변경 계획 설문 - 이탈 여부

```
[43]: from xgboost import XGBRegressor
      from sklearn.metrics import r2_score

      xgb_model=XGBRegressor(random_state=42,
                             n_estimators=300,
                             eta=0.1,
                             gamma=3,
                             max_depth=5,
                             reg_lambda=5,
                             reg_alpha=5,
                             |colsample_bytree = 0.5)

      xgb_model.fit(train_x,train_y)
      y_pred=xgb_model.predict(test_x)

      print(r2_score(test_y,y_pred))

0.37384437093926304
```

교차표(corsstab)


```
from sklearn.ensemble import GradientBoostingRegressor
from sklearn.metrics import *
model_gb = GradientBoostingRegressor(n_estimators=100,
                                     learning_rate=0.1,
                                     max_depth=5,
                                     min_samples_split=30,
                                     min_samples_leaf=15,
                                     random_state=42)

model_gb.fit(x_train, y_train)
y_pred = model_gb.predict(x_test)
print('r2:', r2_score(y_test, y_pred))
```

r2: 0.3178305509267876

```
from sklearn.ensemble import GradientBoostingRegressor
from sklearn.metrics import *
model_gb = GradientBoostingRegressor(n_estimators=100,
                                     learning_rate=0.1,
                                     max_depth=5,
                                     min_samples_split=30,
                                     min_samples_leaf=15,
                                     random_state=42)

model_gb.fit(x_train, y_train)
y_pred = model_gb.predict(x_test)
print('r2:', r2_score(y_test, y_pred))
```

r2: 0.3178305509267876

```
from sklearn.ensemble import GradientBoostingRegressor
from sklearn.metrics import *
model_gb = GradientBoostingRegressor(n_estimators=100,
                                     learning_rate=0.1,
                                     max_depth=5,
                                     min_samples_split=30,
                                     min_samples_leaf=15,
                                     random_state=42)

model_gb.fit(x_train, y_train)
y_pred = model_gb.predict(x_test)
print('r2:', r2_score(y_test, y_pred))
```

r2: 0.3178305509267876

관계 정리

	강한 관계의 변수	약한 관계의 변수	관계 없음
강가영	HOUSE	COLLEGE, INCOME, HANDSET_PRICE	REPORTED_SATISFACTION, CONSIDERING_CHANGE_OF_PLAN
김경준	HOUSE, OVERAGE	HANDSET_PRICE, COLLEGE	REPORTED_SATISFACTION, CONSIDERING_CHANGE_OF_PLAN
박민지	OVERAGE	INCOME, HANDSET_PRICE, COLLEGE	REPORTED_SATISFACTION, CONSIDERING_CHANGE_OF_PLAN
안지예	HOUSE, Call_Group, HI(핸드폰 가격과 연 수입액의 조합)	HANDSET_PRICE, INCOME	REPORTED_SATISFACTION, CONSIDERING_CHANGE_OF_PLAN
조훈근	LEFTOVER, HANDSET_PRICE	REPORTED_SATISFACTION , COLLEGE, AVERAGE_CALL_DURATION	CONSIDERING_CHANGE_OF_PLAN
최종 결론	HOUSE, OVERAGE	COLLEGE, INCOME, HANDSET_PRICE	REPORTED_SATISFACTION, CONSIDERING_CHANGE_OF_PLAN

Part 3

결론

- 고객 이탈 방지 대책



- 월 초과시간이 많을 수록 이탈율이 높음
-> 기본 제공 통화량이 많은 요금제를 조금 저렴하게 만들어 통화량이 많이 필요한 고객 확보 및 이탈 방지
- 대학 졸업 후 이탈율 대학 졸업 전 이탈율보다 미세하게 높음
-> 대학 재학중인 고객을 위해 같은 대학생끼리 이용할 수 있는 같은 대학 결합 패키지 요금제를 만듦
(졸업 후 n년까지 가능)
-> 졸업생을 위해 수능 할인 혜택과 비슷하게 졸업 할인 혜택을 만들어서 졸업생 고객의 이탈을 방지
- 핸드폰 가격이 높을 수록 이탈율이 높음
-> 핸드폰 가격에 따른 추가 혜택을 제공하여 이탈 방지
- 약정을 3년으로 늘려 고객 이탈 방지

감사합니다.