

이동통신 LTV 모델 개발로 고객 관리와 이탈률 저감을 통한 매출 증대



CONTENTS

01

추진배경

02

개선기회

03

분석계획

04

분석결과

05

개선안

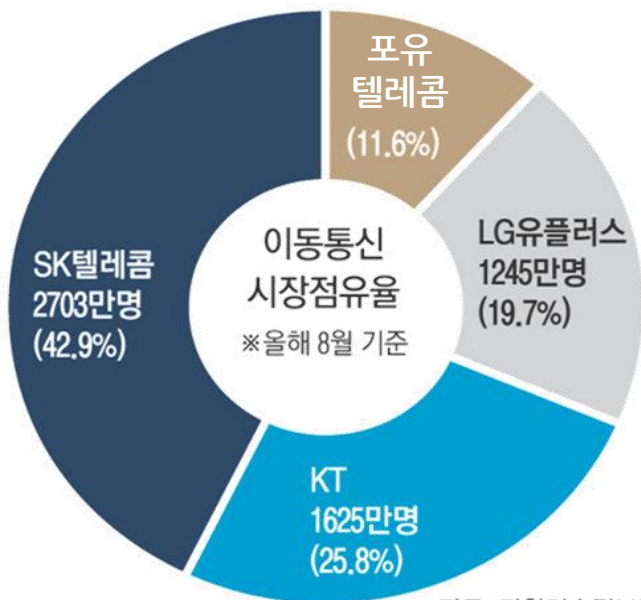
06

향후계획

01. 추진배경

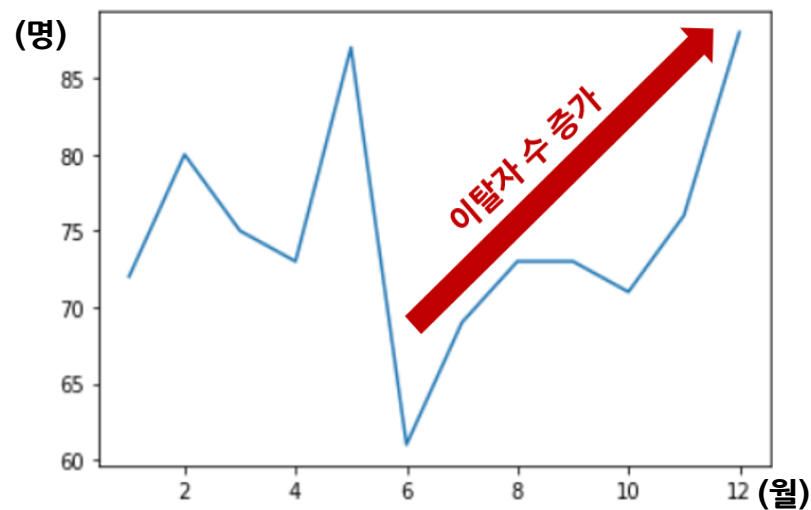
통신시장의 출혈경쟁으로 인한 자사의 고객 이탈률 증가 및 매출 감소

출혈 경쟁 지속



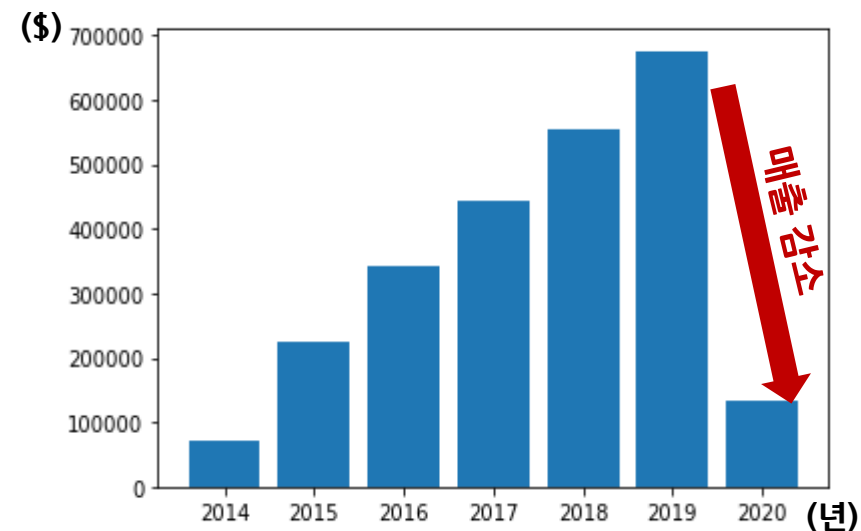
자료: 과학기술정보통신부

이탈자 수 증가



[월별 이탈자 수]

총 매출 감소



[연도별 자사 매출]

핵심 고객 유지 및 신규 고객 유치를 통한 매출 증대 필요

02. 개선기회 KPI

고객 특성에 따른 분류, 서비스의 낮은 만족도, 차별화된 마케팅 부족 개선을 통한 매출 증대 목표

목표 : 매출 증대

LTV 모델 개선으로 고객 이탈률을 저감하여 매출 증대, 약 7억 8천만원 매출 회복



KPI : 고객 이탈률 감소

현재 이탈률 17.5% 에서 10.3%까지 저감 목표 (매년 약 3.6%)



중점 개선 영역

- **핵심고객 분류 부족** : 연령, 결혼 여부, 부양자 수, 만족도와 같은 고객 특성에 따른 분류 부족
- **서비스의 낮은 만족도** : 자사가 제공하는 서비스에 대한 고객 불만족
- **경쟁사 대비 차별적 마케팅 부족** : 경쟁사의 마케팅에 대응할 차별화된 전략 수립

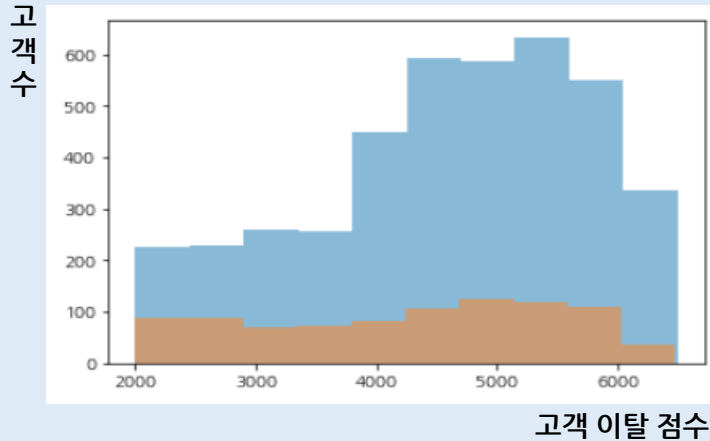
02. 개선기회 중점 개선 영역

이탈률 개선을 위한 중점 개선 영역

핵심고객 분류 미흡

고부가가치의 높은 LTV* 고객을 잡는 것이 목표

*LTV(Lifetime Value) : 고객의 생애가치.



LTV 점수에 상관없이 이탈률이 일정

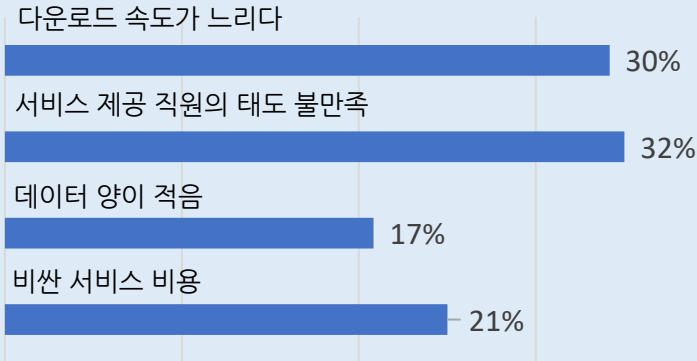
→ 현재 모델은 핵심고객 분류 X

고객의 특성에 맞춘

LTV 모델 개선으로 핵심고객 분류

서비스의 낮은 만족도

서비스 관련 고객의 이탈이유

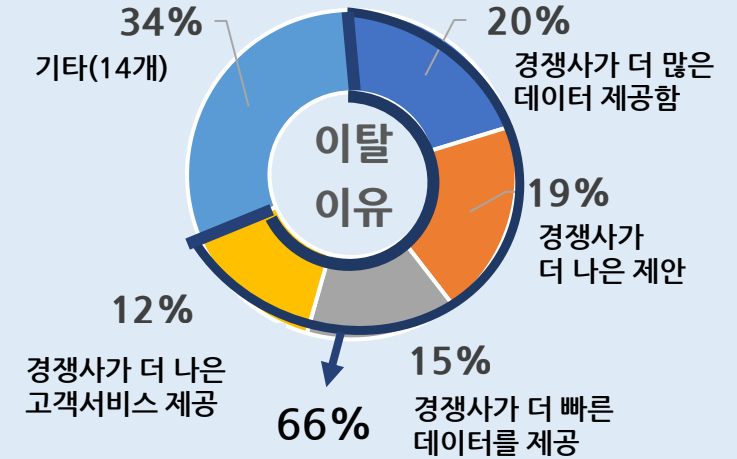


- 무제한 데이터 서비스 사용률(67.6%)으로 가입자가 많지만, 데이터 품질 불만족
- 애프터서비스 담당 직원 태도 불만족

부가서비스 품질 개선,

직원 CS 교육 필요

경쟁사 대비 차별화된 마케팅 미흡



- 경쟁사에 의한 이탈 이유 중 66%가 자사의 서비스 불만족과 연관 有
- 사용 기간동안 요금제, 부가 서비스에 대한 주기적인 모니터링과 제안 필요

요금제와 부가서비스 추천서비스,

경쟁사와 차별화된 마케팅 필요

핵심고객 분류, 서비스 품질 향상, 차별화 마케팅으로 이탈률 저감 → 7억 8천만원 이상의 매출 증대 목표

03. 분석계획

이탈률 예측 모델링을 위한 파생변수 생성 및 분석계획 수립

[Data set & 파생변수]

데이터	데이터 특성	파생 변수
telco_charge .csv	행 214,671개, 열 4개	LTV_category, TenureMonths, TotalCharge 외 11개
telco_customer.csv	행 5,023개 , 열31개	

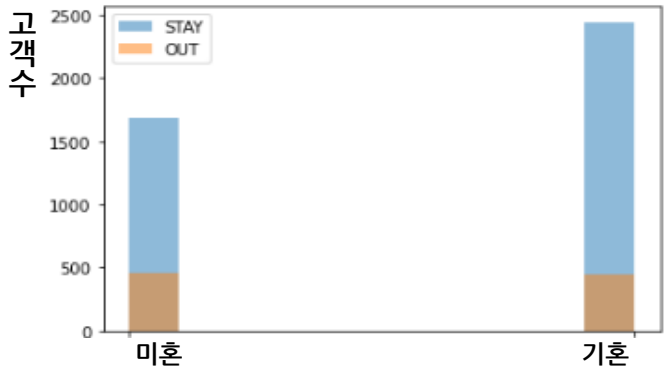
[분석 계획]

목적	분석 방법	주요 내용
각 변수 간의 관계와 변수 별 분포 파악	Box plot	이상치 확인
	Histogram, Bar plot, Cross tab	결혼여부, 추천 횟수 등에 따른 변수 간의 관계 확인
	Line Plot, Heatmap	결혼여부, 성별에 따른 churnscore평균
변수 중요도 파악	Gradient Boosting, Decision Tree, Random Forest	변수 특성에 따른 중요 변수 추출 및 파라미터 조정
변수 수준간 유의미한 특성 파악	연관분석	부가 서비스 간의 관계 파악
	군집분석	만족도에 따른 군집분석 후 특성 확인
모델링	XG Boosting, Random Forest, Decision Tree, SVM, Neural Network	이탈률 분류 모델 개발 및 평가, 정확도 증가

04. 분석결과 이탈률 개선을 위한 분석 1

고객 특성 : 결혼여부, 부양자 수, 추천 횟수, 만족도 인사이트 도출

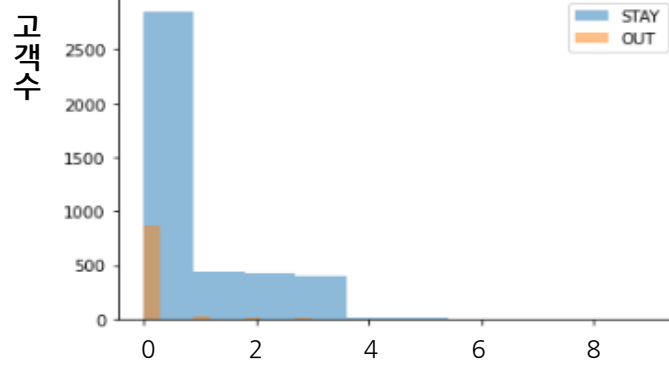
결혼 여부



기혼인 경우 이탈률이 낮음 → 신혼부부 서비스 추천

미혼인 경우 이탈률이 높음 → 미혼 타겟팅 필요

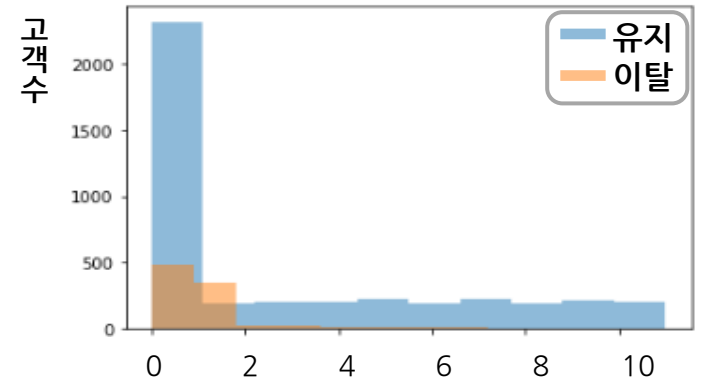
부양자 수



부양자가 많으면 이탈률이 적음 → 가족 결합 서비스 추천

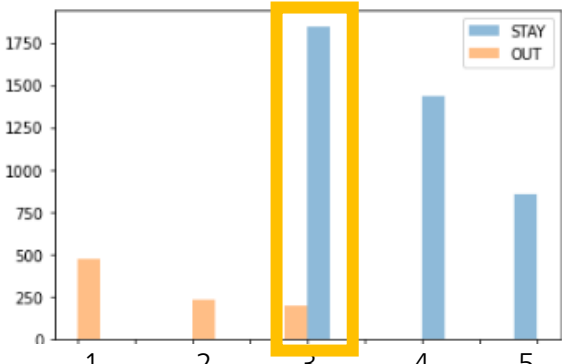
미혼인 경우 부양자수가 없음 → 반려견 서비스 추천

추천 횟수



추천횟수가 많을 수록 이탈률이 적음 → 지인 추천 이벤트

만족도



[만족도와 이탈 여부 간의 분포]

만족도가 3점에서는 이탈/유지 공존

[이탈에 영향을 주는 중요 변수]

Feature	importance
만족도	0.896
평균 과금액	0.028
추천횟수	0.024

DT, Accuracy : 0.972

만족도에 영향을
미치는 변수 파악

[만족도에 영향을 주는 중요 변수]

Feature	importance
평균 과금액	0.374
추천횟수	0.203
사용기간	0.118

RF

평균 과금액이 낮거나,
추천횟수가 많거나,
사용기간이
긴 고객일수록
만족도 ↑ → 이탈률 ↓

04. 분석결과 이탈률 개선을 위한 분석 2

부가서비스 : 5가지 서비스에 따른 인사이트 도출

서비스 간의 연관분석

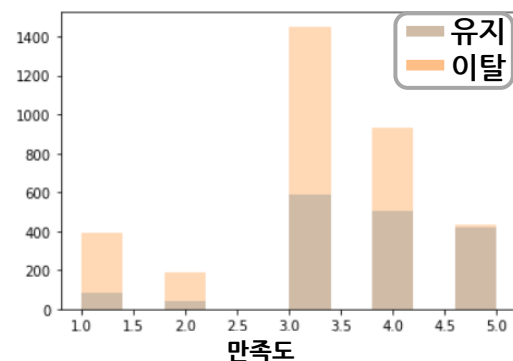
Support	Itemsets
0.676	(무제한데이터)
0.421	(기기보호)
0.354	(기술지원)
0.350	(보안)
0.365	(기기보호, 무제한데이터)
0.357	(백업, 무제한데이터)
0.304	(보안, 무제한데이터)
0.304	(무제한데이터, 기술지원)
0.222	(기기보호, 기술지원)
0.211	(백업, 기기보호, 무제한데이터)

(예) $P(\text{기기보호}) = 0.421$, $P(\text{기기보호} \cap \text{무제한데이터}) = 0.365$
 $P(\text{기술지원}) = 0.354$, $P(\text{기술지원} \cap \text{무제한데이터}) = 0.304$

➡ 부가 서비스를 하나라도 가입한 사람은
대부분 무제한데이터를 함께 사용함

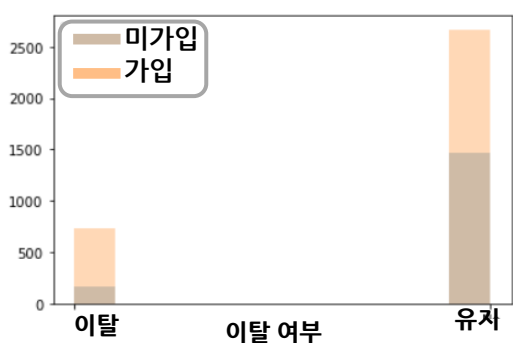
무제한 데이터

[만족도 - 무제한 데이터]



➡ 무제한 데이터에 가입한 고객의
만족도가 낮음

[이탈여부 - 무제한데이터]



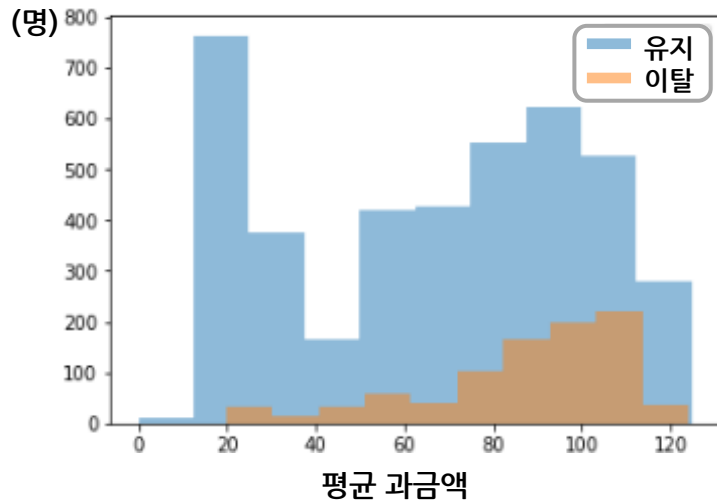
➡ 무제한 데이터 고객 중
이탈률 22%
(전체 이탈률 : 17.5%)

서비스 가입률은 높은 편(67.6%)이지만 만족도가 낮아 이탈자 발생
→ 서비스 품질에 대한 문제 파악

04. 분석결과 이탈률 개선을 위한 분석 3

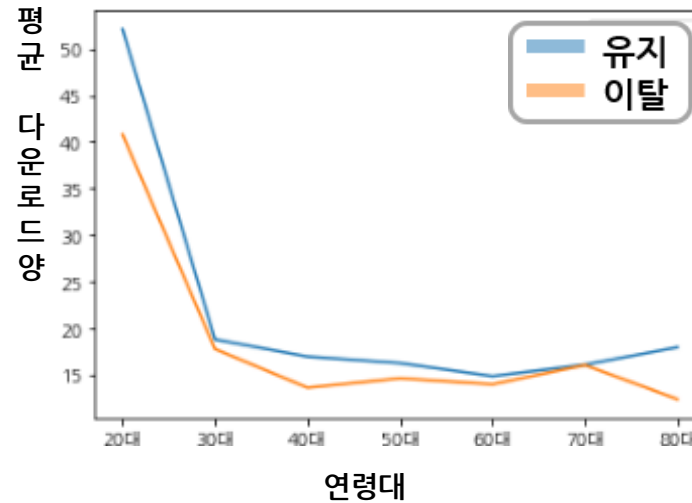
사용패턴 : 평균 과금액, 다운로드양, 계약기간에 따른 인사이트 도출

평균 과금액



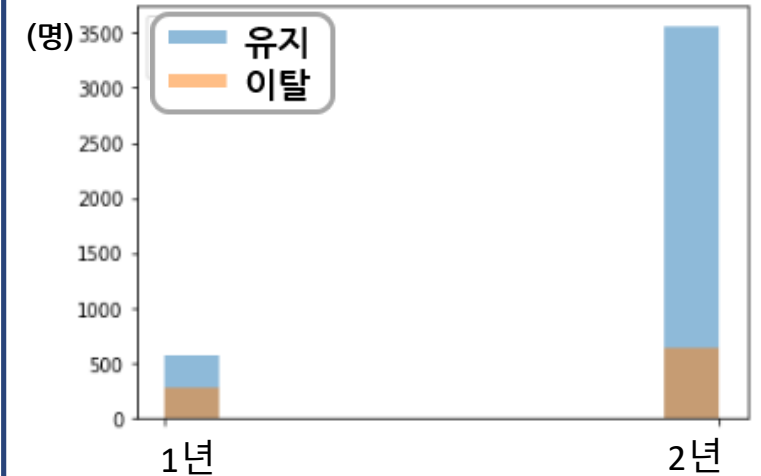
- 평균 과금액이 높을수록 이탈자 수가 많음
- 서비스 품질에 비해 비싼 비용에 부담을 느낀다는 것을 알 수 있음
 - ➔ 서비스 품질 관리와 요금제에 대한 개선이 필요함

다운로드양



- 2030대의 다운로드 양이 가장 많음
- ➔ 이탈 이유에서 데이터 품질 문제가 다수 나타났기때문에 개선이 필요
 - ➔ 다운로드 양이 많기 때문에 저장용량을 넉넉히 할 수 있는 클라우드 서비스 추천

약정기간



- 약정기간 2년일 때 이탈자 수가 적음
- ➔ 마케팅 시 1년보다는 2년 약정이 더 나을 것임

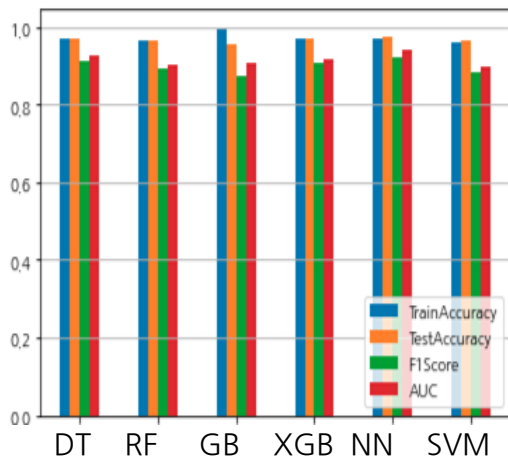
04. 분석결과 모델링 및 평가

고객 이탈 여부 예측 모델링 및 평가와 LTV 모델 개선 필요성 확인

고객 이탈여부 모델

[고객 이탈률 정의] $\frac{(\text{이탈하는 고객 수})}{(\text{현재 고객 수})} * 100$ [현재 고객 이탈률] 17.5 %

[고객 이탈률 분류 모델 성능 비교]



model	F1 score	AUC
Decision Tree	0.971	0.924
Random Forest	0.965	0.901
Gradient Boosting	0.956	0.906
XGB	0.969	0.916
NeuraNet	0.972	0.939
SVM	0.884	0.896

- F1score이 가장 높은 Neural net 모델 최종 선정
- 중요변수 : 만족도, 평균 금액, 추천 횟수, 사용 기간
- 고객 이탈률을 잘 예측한다고 할 수 있음

LTV 모델 개선 필요성

LTV(Lifetime Value)

: 고객의 평생가치. 거래 기간동안 고객에서 얻는 수익 또는 손실

[LTV를 왜 개선해야 하는가]

- 이탈 예상 고객 중 기업에 이윤을 주는 고객을 우선 타겟팅
- 데이터상, 인원과 데이터 사용량 가장 많은 **20~30대** 가 주요 타겟



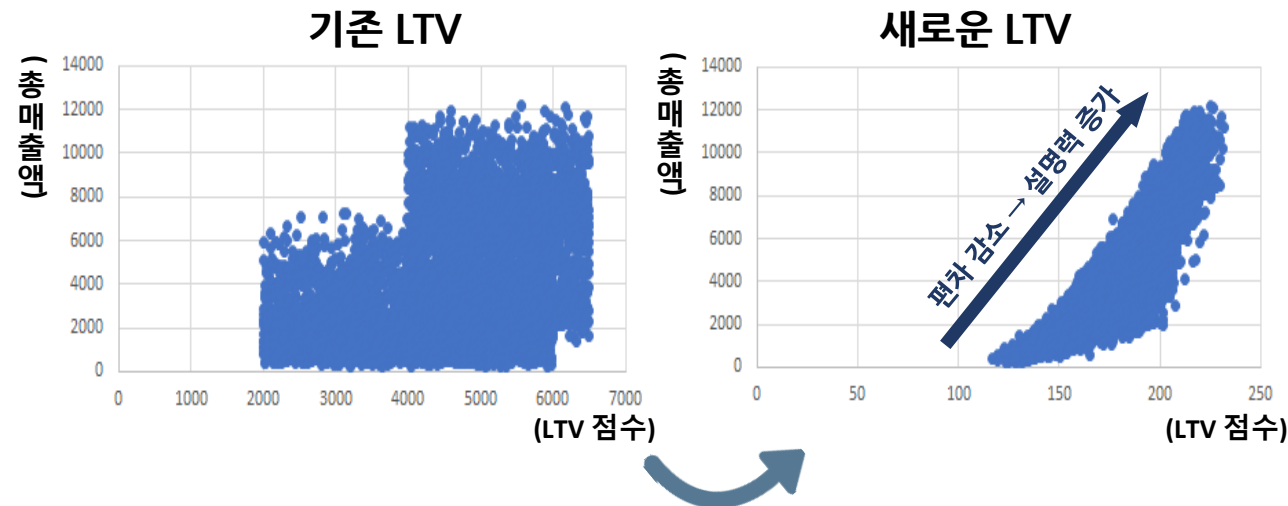
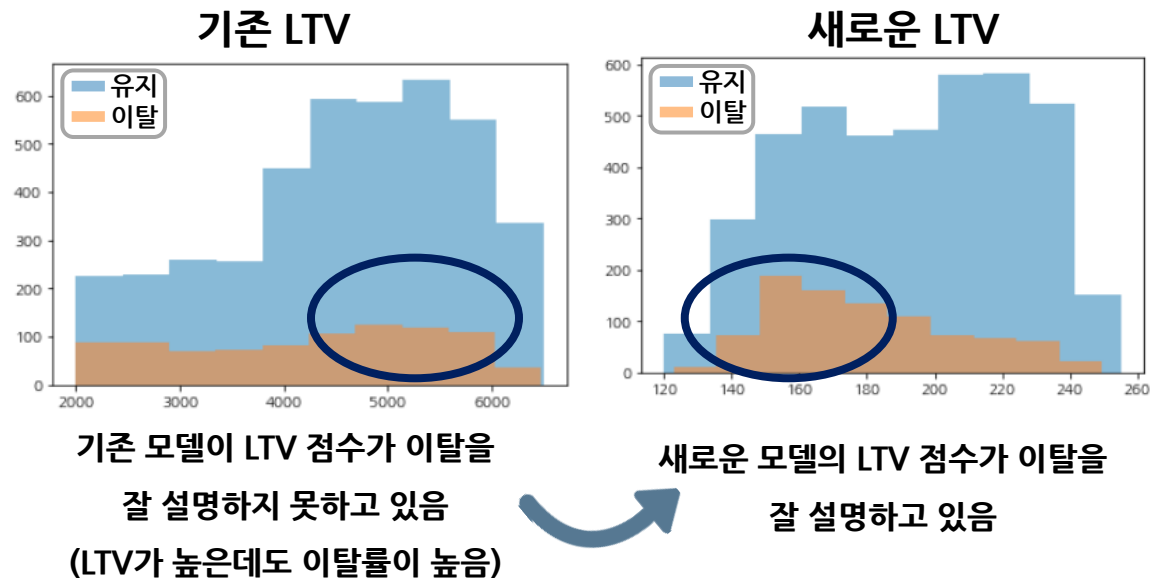
신규고객 유치하는 비용보다 장기고객 유지하는 것이 더 효율적!

고객 특성에 맞는 LTV 모델 개선으로

정확한 고객 분류의 필요

04. 분석결과 LTV 점수 분석

LTV 점수가 높은 고객의 이탈 多 → 이탈률을 정확하게 예측할 수 있는 LTV 모델 필요



$$LTV \text{ 점수} = 120 - 0.236 \times \text{연령} - 0.08 \times \text{총환불액} + 2.28 \times \text{부양자수} + 0.3 \times \text{평균 과금액} + 1.386 \times \text{사용기간}$$

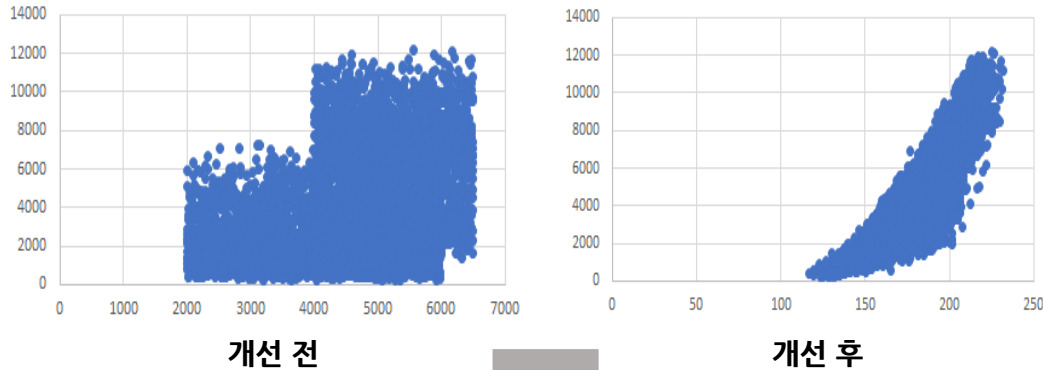
- LTV는 고객 가치를 우선해 핵심 고객에게 마케팅 하는데 사용하는 지표이다.
- 개선된 LTV 점수로 핵심고객을 개발하고 유지하는 마케팅에 활용한다.
- 위의 식에 다른 요소를 추가하더라도 위와 같은 모델과 성능이 크게 다르지 않다.

(출처 : 이동통신회사에서의 Customer Value 측정을 통한 고객 세분화)

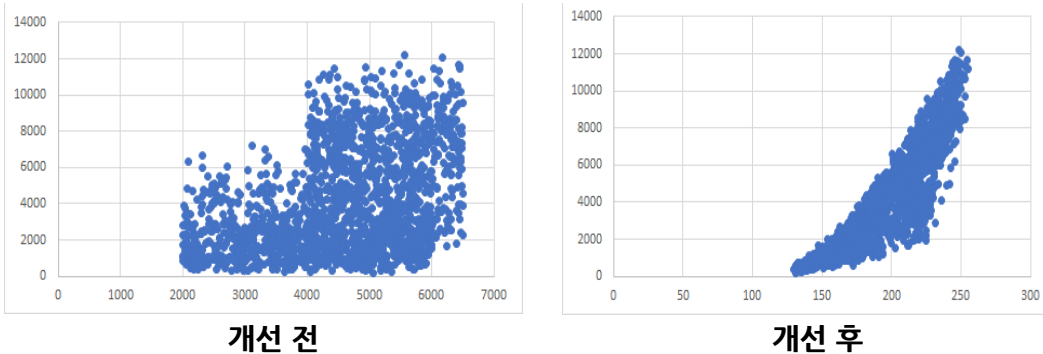
04. 분석결과

새로운 LTV 점수 타겟층(2030대) 적용

전체 연령대상_ 실제 LTV 모델과 개선 모델



타겟층인 20-30대_ 실제 LTV 모델과 개선 모델



“ 타겟층에서도 정확하게 모델이 맞음 ”

[개선된 LTV 모델을 기준으로 핵심 고객에 대한 총별화] → 멤버십 적용

민간 납부 금액 (\$ 기준)	가입기간		
	2년미만	2년이상 - 5년미만	5년이상
100 이상	Gold	VIP	VIP
80 - 100	Gold	Gold	VIP
60 - 79	Sliver	Gold	Gold
40 - 59	Sliver	Sliver	Gold
39 이하	Sliver	Sliver	Sliver

- 개선된 LTV 모델을 바탕으로 고객 총별화
- 고객의 LTV 점수는 대외비이므로 고객별 납부액 기준으로 멤버십 구성
- 이를 통해 핵심 고객 유치 및 장기 고객화

05. 개선안 1 고객 특성에 맞춘 서비스 제공

다양한 결합 상품과 추천제를 통해 신규고객 유치 및 기존 고객 유지

신혼 결합 할인 상품

제안 배경 : 기혼 고객이 추천횟수 많고, 이탈률도 낮음



제안 내용 : 신혼부부가 휴대폰을 결합하면, 추가 혜택 제공
(넷플릭스 무료 제공 , 동반 가입시 요금 할인)
이탈률 : 기혼 15.4% , 미혼 21.1%
기대 효과 : 신규 고객 유치 및 매출 증대

지인 추천 할인 상품

제안 배경 : 추천을 많이 한 고객이 이탈률 낮음



제안 내용 :
- 추천을 많이 할수록 , 추가 혜택 제공
(데이터 쿠폰, 포인트 제공, 넷플릭스)
- 3회까지는 할인 제공, 그 이후에는 포인트 제공
이탈률 : 추천多 2.5% , 추천少 24.1%
기대 효과 : 신규 고객 유치 및 이탈률 저감으로 매출 증대

05. 개선안 2 고객 니즈에 맞춘 서비스 제공

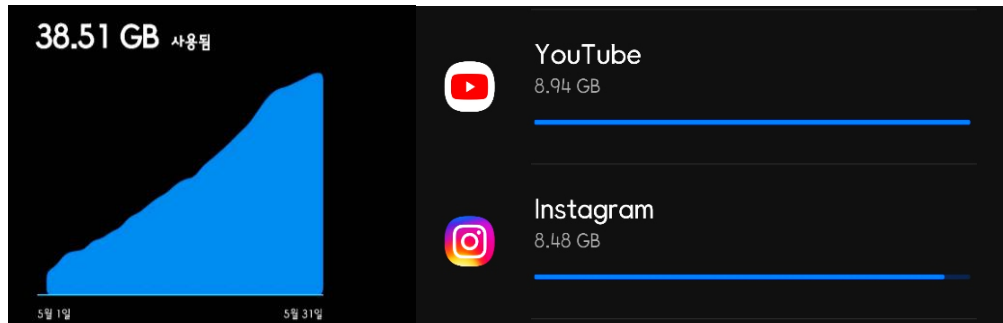
고객이 불만족했던 데이터 및 부가서비스 문제 개선

고객 맞춤형 추천 서비스

고객의 데이터 사용량 및 활동, 패턴 분석

➡ 고객에게 맞는 요금제 및 부가서비스 추천

(예)



월 38GB 사용, 주요 콘텐츠는 Youtube



IPTV/유튜브 연계 데이터 요금제 추천
(스트리밍 데이터 무료 제공)

클라우드 서비스

20~ 30대의 특징 : 다운로드양과 데이터 사용량이 많음

➡ 2년 약정 고객에게 클라우드 서비스제공
(1년간/ 온라인 백업 서비스 등 포함)

AI 고객 대응 서비스

이탈자 중 약 **18%**가 자사고객 대응 서비스에 불만

➡

- 감정분석 AI 서비스 & STT*를 통해 고객 대응
- 챗봇 서비스
- 직원 CS 교육 실시

*STT : Speech to text

통신망 개선

데이터 서비스와 이탈률이 큰 상관관계를 가짐

➡ 통신망의 유동적인 배정을 통해서 통신 속도와
데이터 불편사항을 개선

05. 개선안 3 경쟁사 대비 차별화 마케팅

펫팜족을 위한 차별화된 마케팅

반려견 위치 파악 서비스



국내 반려동물 인구 1,000만 돌파

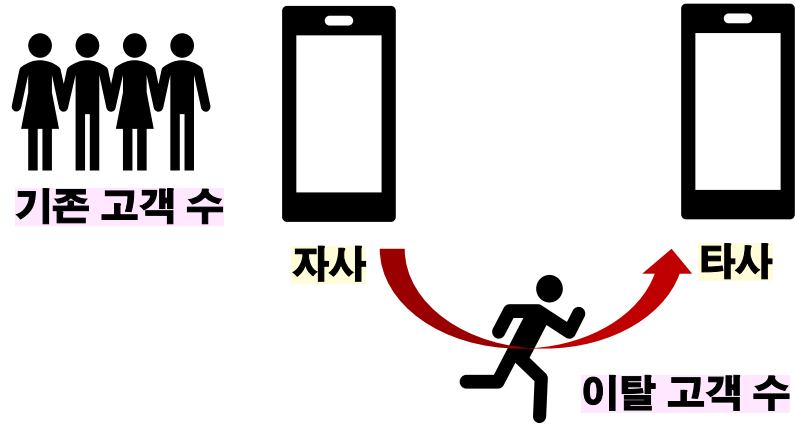
반려견을 키우고 계신가요?

소중한 댕댕이를 저희가 지켜드려요!

2030 미혼 가정을 위한 타 통신사와 차별화된 마케팅 출시

- 스마트 태그를 활용한 반려동물 위치 파악
- 반려동물의 시야를 확인할 수 있는 소형 카메라 장착
- 핸드폰에서 바로 확인 가능

“고객 이탈률” = 이탈하는 고객 수 대비 신규 고객 수를 의미



[Churn Rate (이탈률 수치) 계산식]

$$\frac{(\text{해당 통신사를 이탈하는 고객의 수})}{(\text{해당 통신사의 유지 고객의 수})} \times 100$$

이탈 가능성이 높은 고객을 미리 예측하여 막는다면 더 많은 새로운 고객을 확보하는 것보다 더 좋은 효과

새로운 고객을 확보하는데 드는 비용 > 기존 고객을 유지하는 비용

기존 고객을 유지하는 것이
더욱 효과적임!!

고객 이탈률을 정확히 예측할 수 있다면 고객 생애가치(LTV), 마케팅 투자 대비 수익(Return on Marketing Investing) 등
더 정확하게 계산 & 올바른 의사결정에 도움

Appendix p.3 추진배경_총 매출과 이탈자 수

1. 총 매출 매트릭스

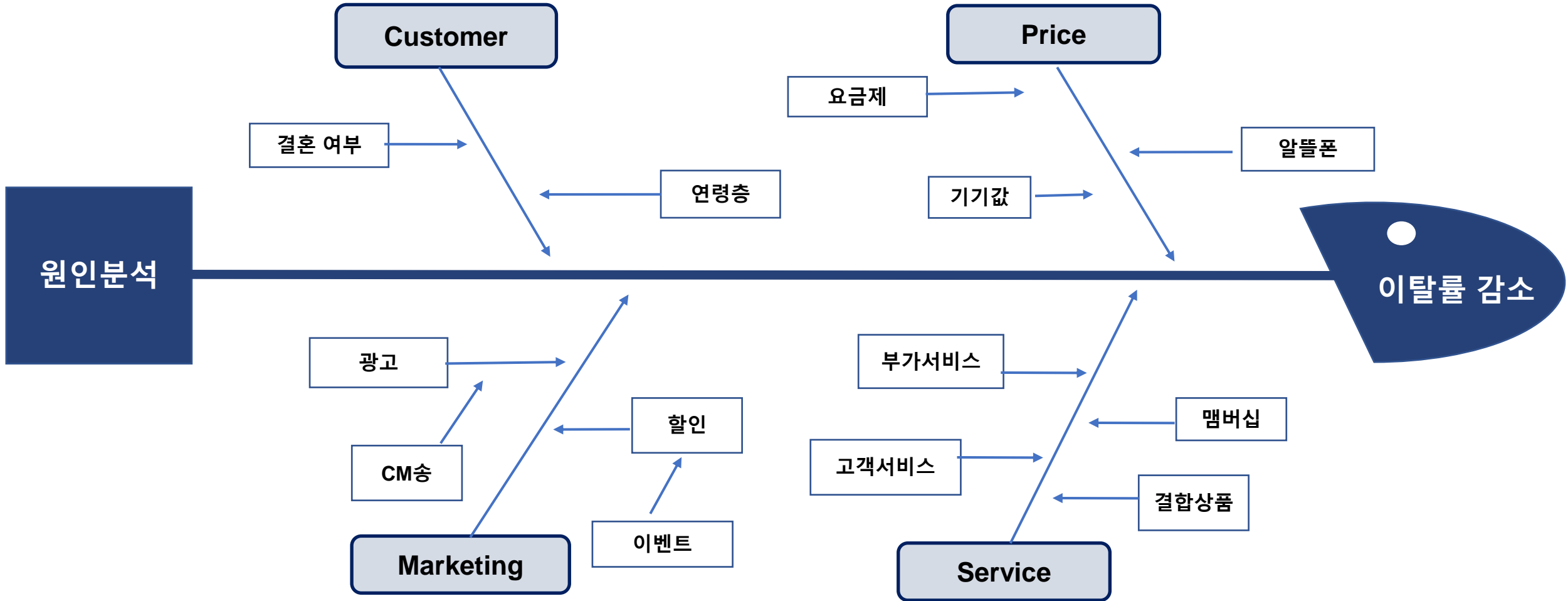
	ChargeAmt	chargemonth
chargeyear		
2014	848436.25	81847
2015	1651915.10	148017
2016	2310761.70	206993
2017	2931181.45	266746
2018	3623706.65	337090
2019	4080410.70	369479
2020	419714.20	20811

총 매출은 증가하다가 2020년에 급격하게 하락

2. 2020년 월별 이탈자 수

1월	2월	3월	4월	5월	6월
72	80	75	73	87	61
7월	8월	9월	10월	11월	12월
69	73	73	71	76	88

2020년 월 평균 이탈자는 약 74명
하지만 신규 가입자가 2월 이후 없음
→ 이탈자가 총 매출에 큰 영향을 끼침



1. LTV란?

: Life Time Value, 고객 생애 가치로 사용자 1명당 서비스에서 이탈하기 전까지 지불하는 금액을 의미한다.

2. LTV 계산식

ARPA(Average Revenue per Account)란?

:결제계좌당 평균 매출을 뜻하는 용어로 통신사의 수익성을 측정할 때 활용하는 ARPU(가입자당 평균 매출)을 대체하는 지표

ARPU는 가입자들이 한 달에 납부하는 평균 금액과 유사하고 직관적이란 점에서 통신업계의 대표적인 수익성 지표로 활용되어 왔다.

why?

'1인 1회선'이란 공식이 더 이상 성립되지 않는 상황에서 ARPU보다 다회선을 반영한 ARPA로 통신사의 수익성을 평가하는 것이 합당하다는 지적이 제기된다.

EQUATION:

$$LTV = \frac{ARPA}{\text{Customer Churn Rate}}$$

SIMPLIFIED:

$$LTV = \frac{\left[\text{Sum of all customer MRR} / \text{Total \# of customers} \right]}{\left[\# \text{ of customers who churned} / \text{Total \# of customers} \right]}$$

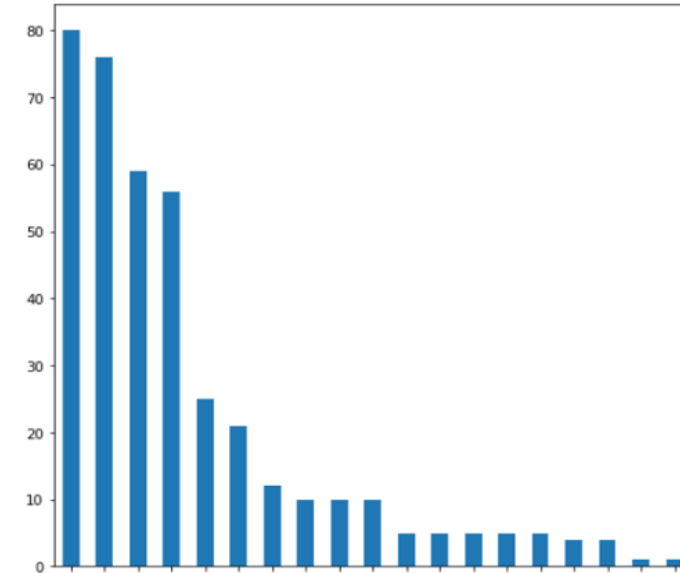
EXAMPLE:

$$LTV = \left(\frac{\$100 \text{ ARPA}}{0.1 \text{ Churn Rate}} = \$1000 \text{ LTV} \right)$$

Appendix p.5 개선기회_이탈 이유 데이터 분석

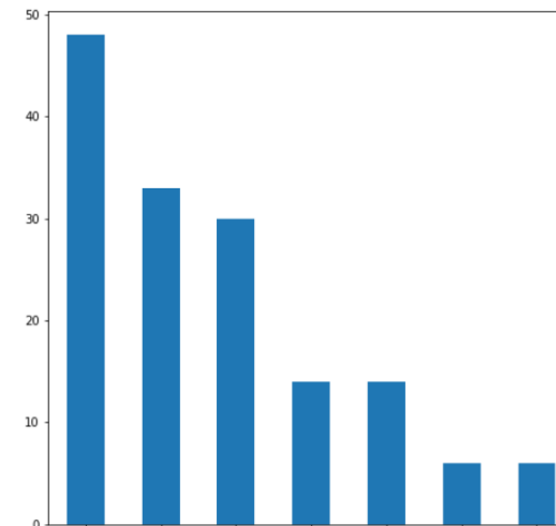
1. 이탈 이유 - 경쟁사

이유	인원(명)
경쟁사가 더많은 데이터 제공	80
경쟁사의 다운로드 속도가 더 빠르다	76
경쟁사가 더 좋다.	59
경쟁사의 기기가 더 좋다.	56
고객서비스 직원의 태도	25
A/S서비스 직원의 태도	21
자사 서비스의 불만족	12
추가 데이터 요금	10



2. 이탈 이유 - 자사에 대한 불만족

이유	인원(명)
제품 불만족	48
서비스 불만족	33
네트워크 지연율	30
부족한 인터넷 서비스	14
한정적인 서비스	14



NaN가 있는 변수는 3개, 이상치는 없음

```
In [455]: df.isnull().sum()
```

```
Out [455]: Unnamed: 0      0
CustomerId      0
ContractID      0
CustomerLTV      0
ChurnScore      0
SatisScore      0
ChurnLabel      0
Gender          0
Age            0
AgeGroup        0
Married         0
Dependents      0
noDependents    0
Referrals        0
noReferrals     0
Contract        0
PaperlessBilling 0
PaymentMethod   0
OnlineSecurity  0
OnlineBackup    0
DeviceProtection 0
TechSupport     0
UnlimitedData    0
AvgDownloadGB   0
TotalRefund     0
TotalExtraDataCharge 0
AvgLongDistanceCharge 0
TotalLongDistanceCharge 0
StartDate       0
EndDate         4136
ChurnCategory   4136
ChurnReason     4136
LTV_category    0
AvgCharge       0
```

```
In [471]: df["EndDate"]
```

```
Out [471]: 0      01-01-2020
1      04-01-2020
2      06-01-2020
3      02-01-2020
4      02-01-2020
...
5029      NaN
5030      NaN
5031      NaN
5032      NaN
5033      NaN
```

```
df["ChurnCategory"]
```

```
0      Competitor
1      Dissatisfaction
2      Competitor
3      Competitor
4      Competitor
...
5029      NaN
5030      NaN
5031      NaN
5032      NaN
5033      NaN
```

```
df["ChurnReason"]
```

```
0      Competitor offered higher download speeds
1      Product dissatisfaction
2      Attitude of support person
3      Competitor had better devices
4      Competitor offered more data
...
5029      NaN
5030      NaN
5031      NaN
5032      NaN
5033      NaN
```

NaN을 결측치라고 하기에는 계약 지속 고객은 이탈 날짜, 이탈 이유 無
따라서 결측치가 아닌 정상

Appendix p.6 분석계획_Data Set과 파생변수

[Data Set]

Variable	Description	Type
CustomerID	고객ID	문자
ChargeAmt	과금액(월)	연속
CustomerLTV	고객LTV	연속
ChurnScore	Churn점수	연속
ChurnLabel	Churn여부	범주
SatisScore	만족도 점수	이산
AgeGroup	연령대	이산
Married	결혼여부	범주
Dependents	부양자 유무	범주
noDependents	부양자 수	이산
noReferrals	추천 횟수	이산
Contract	계약/약정방식	범주
PaperlessBilling	종이영수증 발급여부	범주
OnlineSecurity	서비스 가입(보안)	범주
OnlineBackup	서비스 가입(백업)	범주
DeviceProtection	서비스 가입(기기보호)	범주
TechSupport	서비스 가입(기술지원)	범주
UnlimitedData	서비스 가입(데이터 무제한)	범주
AvgDownloadGB	Download 양	연속
AvgCharge	평균 과금액(월)	연속
TotalExtraDataCharge	요금(추가지불 총액)	연속
StartDate	가입일자	날짜
EndDate	해지일자	날짜
ChurnCategory	Churn카테고리	범주
ChurnReason	Churn사유	범주

[파생 변수]

Variables		Description
TotalCharge	총 과금액(전 기간)	Telco_charge의 과금액(월)의 총합
AvgCharge	평균 과금액(월)	TotalCharge / 개월 수
TotalRevenue	총 매출	=총 과금액 + 추가지불 총액 + 장거리 전화총액 - 환불총액
TenureMonths	사용기간(월)	과금납부일(월) 기준으로 산출
LTV_category	고객 LTV 지수 범주화	(2000 ~ 5500) 500단위로 범주화
DownloadGB_category	다운로드 양 범주화	(15 ~ 90) 15 단위로 범주화
Refund_category	요금(환불 총액) 범주화	환불하면 1, 아니면 0으로 범주화
AVG_Distance_category	요금(장거리 전화 평균) 범주화	(10 ~ 50) 10단위로 범주화
Total_Distance_category	요금(장거리 전화 총액) 범주화	(500 ~ 3000) 500단위로 범주화
ExtraData_category	요금(추가지불 총액) 범주화	추가지불 금액이 없으면 0, 있으면 1로 범주화
ChurnS_category	Churn Score 범주화	(10 ~ 100) 10단위로 범주화
TotalRevenue_category	총 매출 범주화	(1000 ~ 12000) 1000단위로 범주화
TotalCharge_category	총 과금액(전 기간) 범주화	(2000 ~ 8000) 2000단위로 범주화
AvgCharge_category	평균 과금액(월) 범주화	(30 ~ 120) 30단위로 범주화

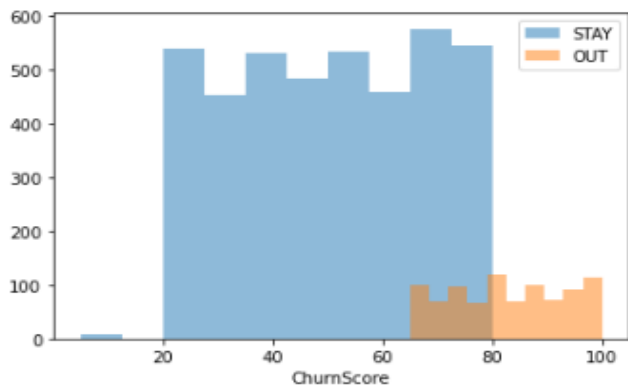
1. 전반적인 데이터 분포 확인 : Churn Label 여부로 데이터 나누어서 분석

- 이탈 고객을 잔존한다고 잘못 예상하고 있기에
나간 고객과 남은 고객의 특성을 비교, 분석하기 위해 관련 변수인 ChurnLabel을 사용

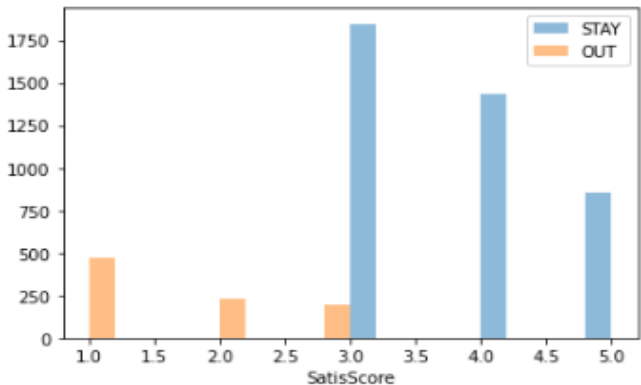
Churn Label 기준_유의미한 변수

- 천스코어 / 만족도 / 부양자여부 / 부양자수(확인필요) / 추천횟수 /
계약기간 / 종이영수증(확인필요) / 다운로드양(0~15) / 평균 과금액
/장거리평균요금(0~5) / 결혼 / 추천여부
- 0 빼고 : Refund , 추가 데이터 요금

1. 전반적인 데이터 분포 확인 : 유의한 영향을 끼치는 변수 파악

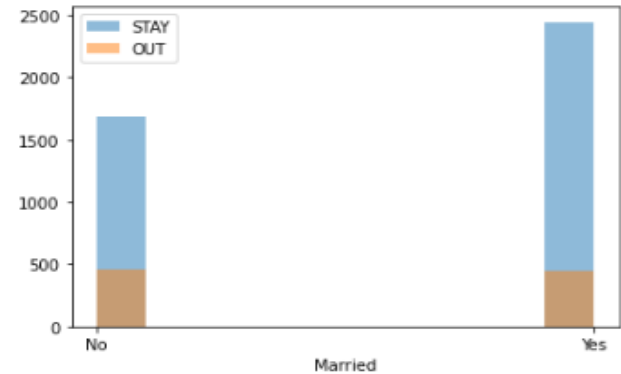


[Churn Score]



[Satis Score]

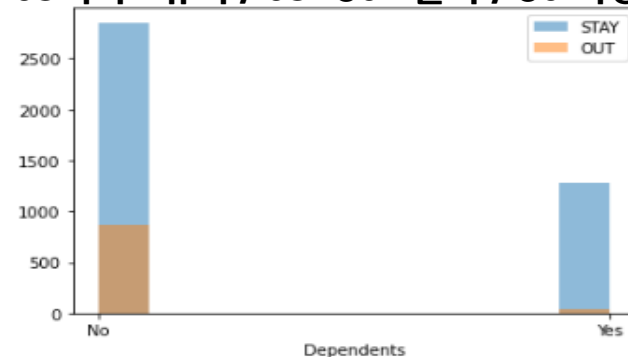
3 미만 : 이탈 / 3점 : 둘 다 / 3 초과 : 유지



[Married]

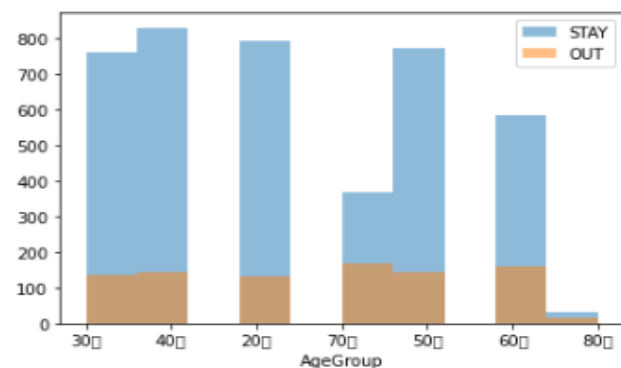
결혼할수록 남아있는 확률 높음

65이하 : 유지 / 65~80 : 둘다 / 80 이상 : 이탈



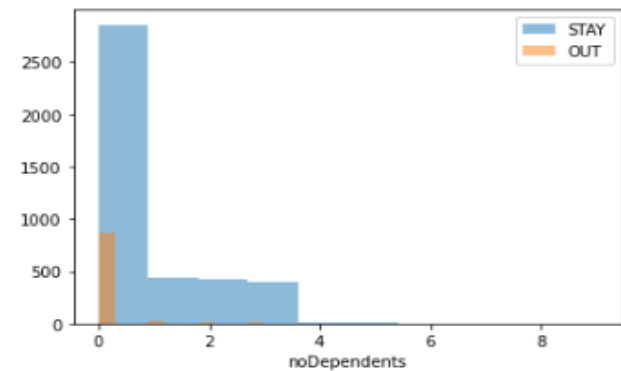
[Dependents]

부양자가 없을 때 더 많이 나감



[Age Group]

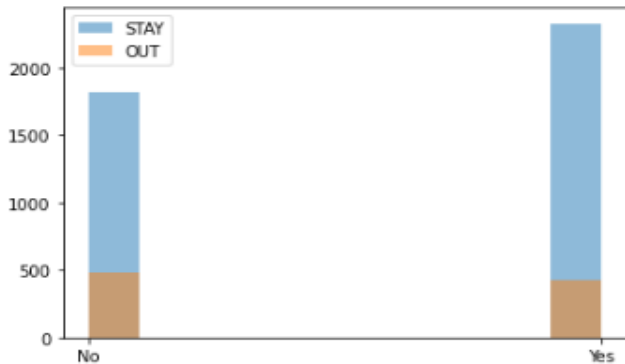
70-80대에 많이 이탈함



[noDependents]

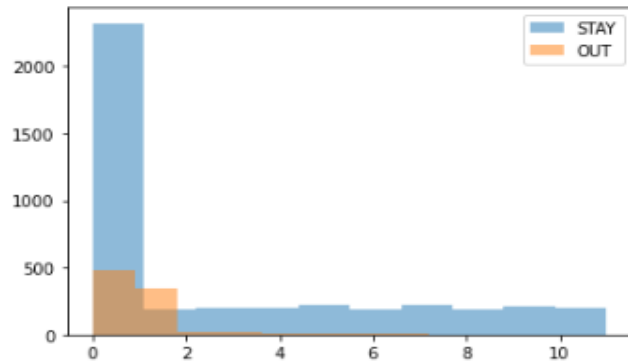
결혼할수록 남아있는 확률 높음

1. 전반적인 데이터 분포 확인 : 유의한 영향을 끼치는 변수 파악



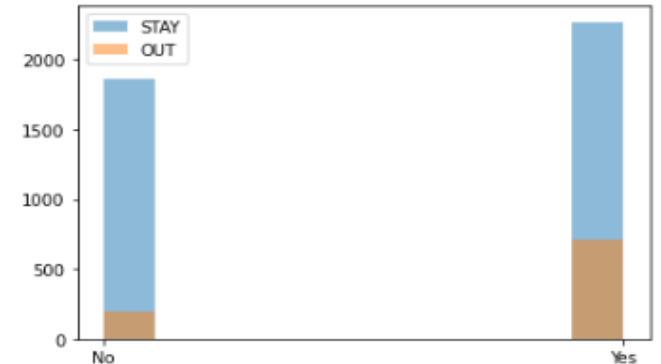
[Referrals]

추천을 안할수록 이탈확률 높음



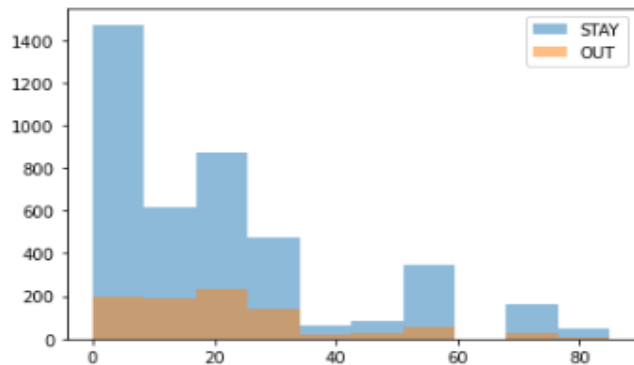
[noReferrals]

추천수가 낮을 경우에만 이탈 확률 존재



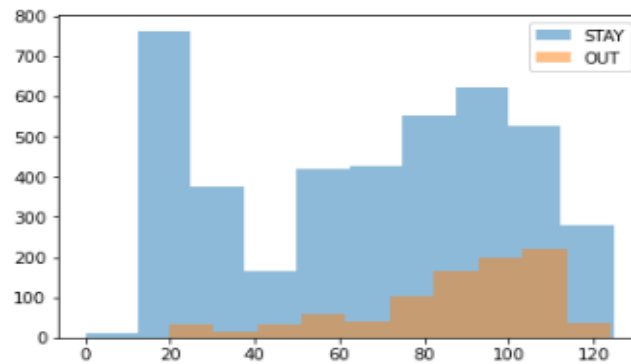
[PaperlessBuilding]

종이영수증에는 차이가 없다.



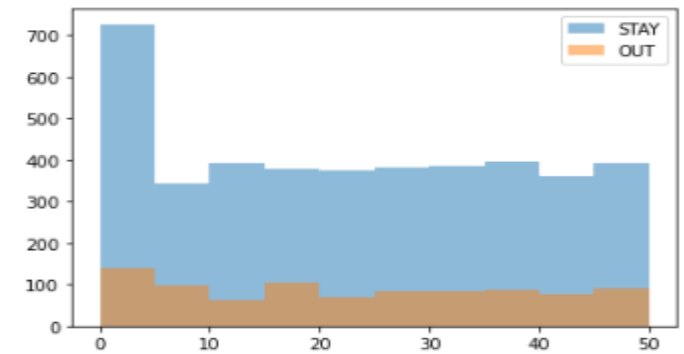
[AvgDownloadGB]

평균 다운로드양이 적을수록 이탈확률 높음



[AvgCharge]

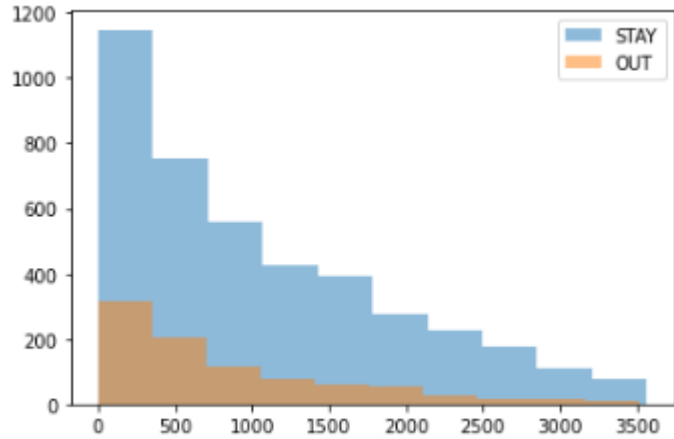
평균요금이 높아질수록 이탈확률 높음



[AvgLongDistanceChange]

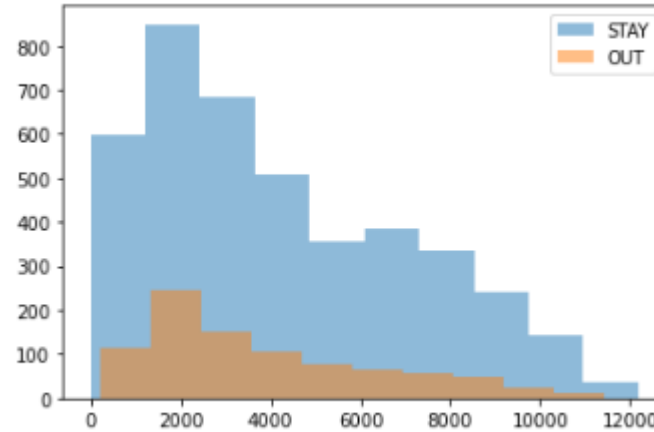
장거리 요금제에는 차이가 없다.

1. 전반적인 데이터 분포 확인 : 유의한 영향을 끼치는 변수 파악



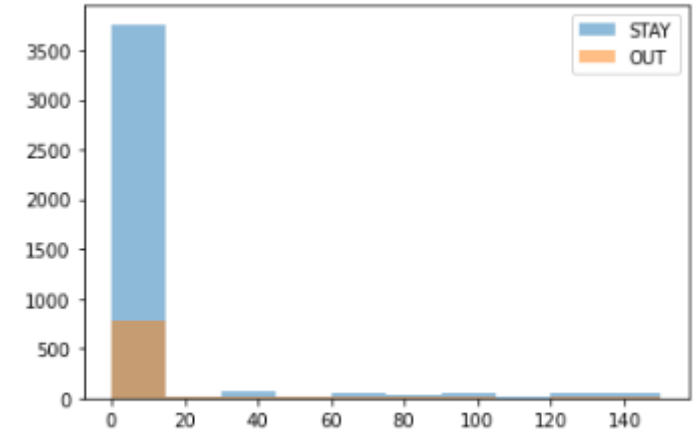
[TotalLongDistanceChange]

총 거리 추가 요금에는 차이 없음



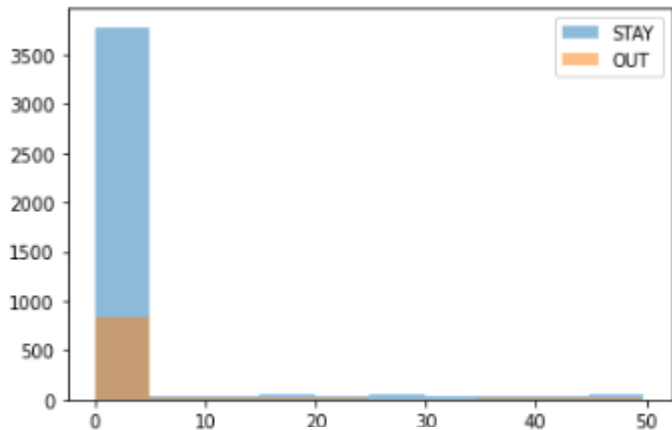
[TotalRevenue]

총 매출에는 차이 없음



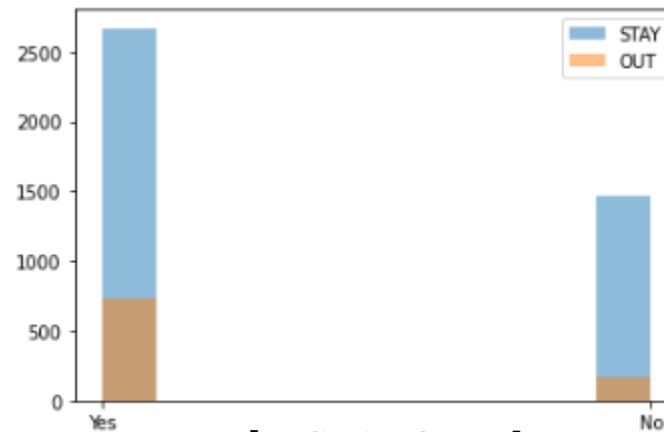
[TotalExtraDataChange]

총 추가데이터 요금에는 차이없음



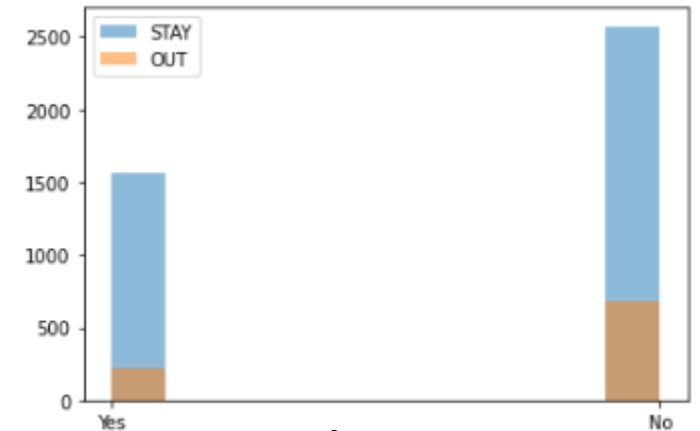
[TotalRefund]

환불 금액에는 비교 불가능



[UnlimitedData]

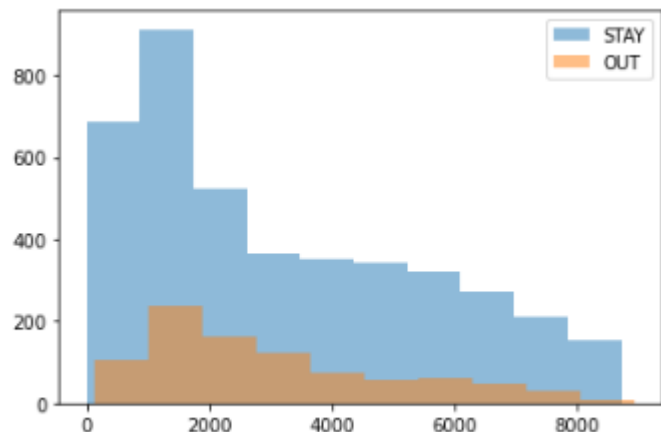
무제한 데이터에는 차이 없음



TechSupport

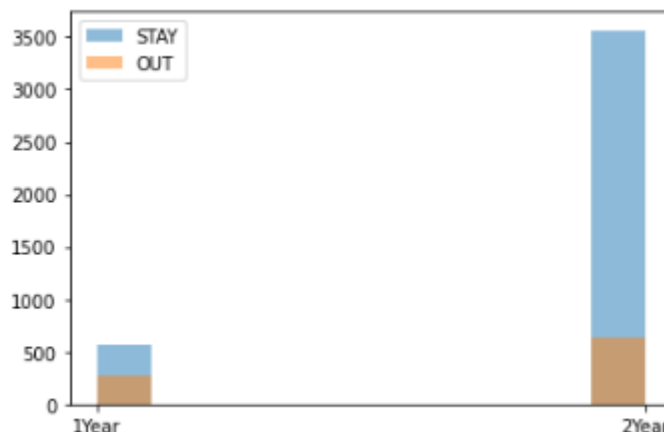
테크 서비스에는 차이 없음

1. 전반적인 데이터 분포 확인 : 유의한 영향을 끼치는 변수 파악



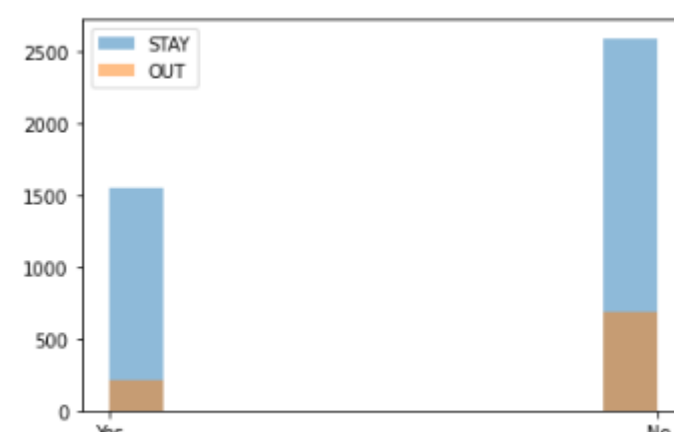
[TotalChange]

총 요금에는 차이 없음



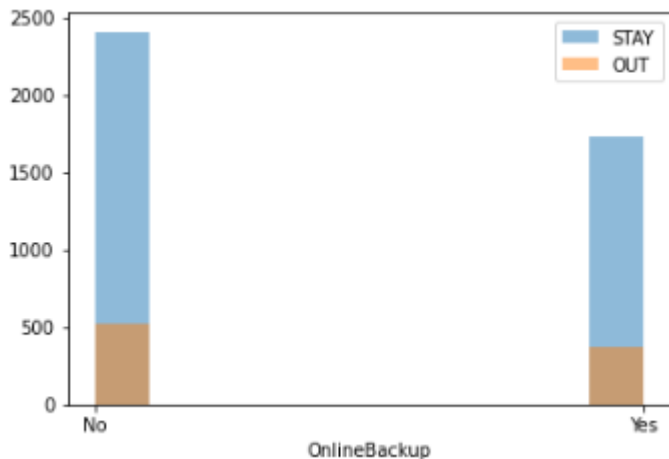
[Contract]

1년 계약의 이탈이 높음

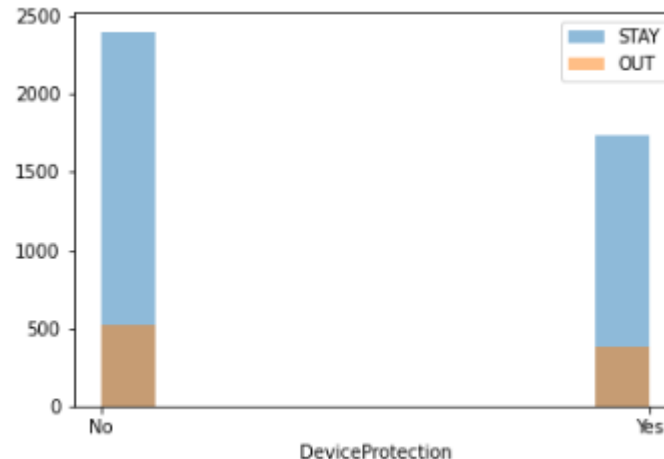


[OnlineSecurity]

온라인 보안 서비스는 큰 차이 없음



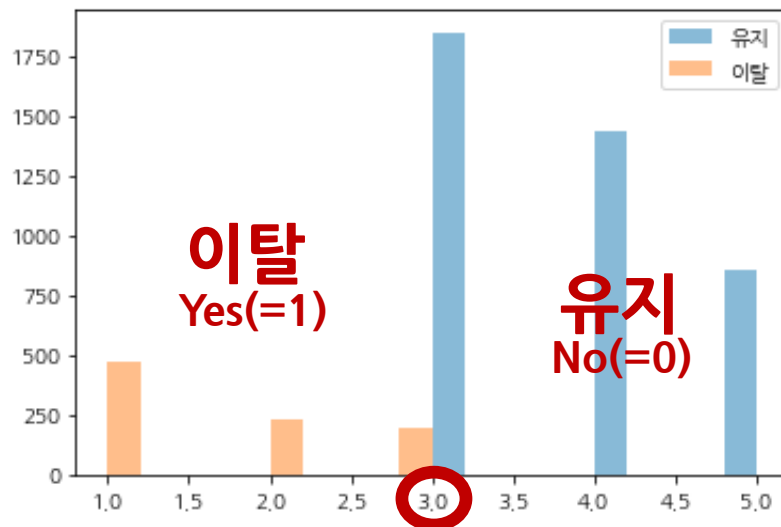
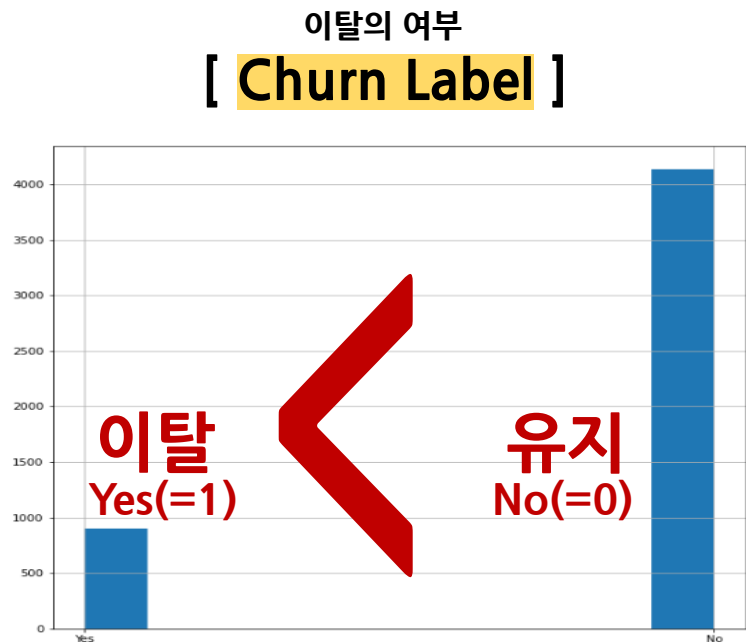
[OnlineBackup]



[DeviceProtection]

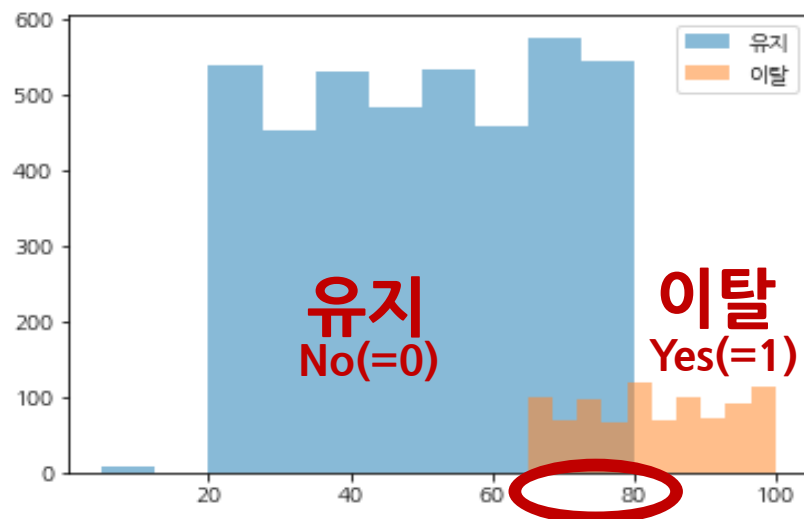
온라인 백업 서비스와 장치 보호
서비스도 큰 차이가 없음

2. Churn Label과 큰 연관성을 가진 변수 : 만족도, 이탈률 점수



만족도 점수
[Satis Score]

- 2점 이하는 모두 이탈
- 3점에서 이탈/유지 모두 발생
- 4점부터는 모두 유지



이탈률 점수
[Churn Score]

- 65점 이하는 모두 유지
- 80점 이상은 모두 이탈
- 65점에서 80점 사이에서 이탈/유지 모두 발생

3. Random Forest, Decision Tree, Gradient Boosting 을 통한 중요변수 도출

Y = Satis Score

GB 중요변수

	Feature	Importance
8	AvgCharge	0.234
2	noReferrals	0.110
9	TotalCharge	0.105
10	TotalRevenue	0.081
3	AvgDownloadGB	0.069
7	TotalLongDistanceCharge	0.065
6	AvgLongDistanceCharge	0.060
0	Age	0.057

RF 중요변수

	Feature	Importance
7	AvgCharge	0.302
1	noReferrals	0.208
2	AvgDownloadGB	0.111
8	TotalCharge	0.079
33	OnlineSecurity_res	0.034

DT 중요변수

[70]:	Feature	Importance
9	OnlineSecurity	0.367
2	AvgCharge	0.147
7	noReferrals	0.127
23	TenureMonths	0.072
18
22	TotalRevenue	0.049
5	Referrals	0.045
3	Dependents	0.039
0	CustomerLTV	0.038
14	AvgDownloadGB	0.019
8	PaperlessBilling	0.013

Statis Score에 대해 DT와 GB, RF로 중요 변수를 확인해본 결과,
평균 과금액, 추천 횟수, 평균다운로드양, 총 과금액 가 도출됨

3. PCA 를 통한 중요변수 도출

	Feature	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6
0	noDependents	-0.026	-0.272	-0.088	-0.228	-0.378	-0.116
1	noReferrals	-0.126	-0.274	-0.093	0.073	0.151	0.041
2	AvgDownloadGB	-0.154	0.059	-0.188	-0.168	-0.169	-0.038
3	TotalRefund	0.006	0.001	-0.021	0.001	0.000	0.028
4	TotalExtraDataCharge	-0.005	-0.004	0.115	0.029	-0.056	-0.081
5	AvgLongDistanceCharge	-0.034	-0.020	0.182	0.167	-0.148	-0.041
6	TotalLongDistanceCharge	-0.160	-0.053	0.321	0.161	-0.130	-0.050
7	AvgCharge	-0.262	0.179	-0.113	0.079	-0.074	-0.016
8	TotalCharge	-0.323	0.066	0.130	0.056	-0.048	-0.024
9	TotalRevenue	-0.313	0.038	0.207	0.096	-0.080	-0.036
10	TenureMonths	-0.240	-0.071	0.312	0.046	-0.011	-0.020
11	Gender_Female	0.000	0.003	-0.031	-0.025	0.244	-0.654
12	Gender_Male	-0.000	-0.003	0.031	0.025	-0.244	0.654
13	AgeGroup_20대	-0.017	-0.016	-0.050	-0.121	-0.088	-0.032
14	AgeGroup_30대	0.000	-0.020	0.009	-0.033	-0.017	-0.006
15	AgeGroup_40대	0.022	-0.025	0.041	-0.041	0.014	-0.042
16	AgeGroup_50대	0.012	-0.016	0.023	-0.032	0.043	0.034
17	AgeGroup_60대	0.008	0.027	0.019	0.077	0.021	0.001
18	AgeGroup_70대	-0.028	0.058	-0.049	0.186	0.045	0.065

OLS Regression Results						
=====						
Dep. Variable:	SatisScore	R-squared:	0.076			
Model:	OLS	Adj. R-squared:	0.076			
Method:	Least Squares	F-statistic:	138.1			
Date:	Mon, 31 May 2021	Prob (F-statistic):	5.51e-86			
Time:	23:47:52	Log-Likelihood:	-7459.1			
No. Observations:	5023	AIC:	1.493e+04			
Df Residuals:	5019	BIC:	1.495e+04			
Df Model:	3					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	3.3888	0.015	224.734	0.000	3.359	3.418
Prin1	-0.0345	0.006	-6.207	0.000	-0.045	-0.024
Prin2	0.1000	0.007	15.011	0.000	0.087	0.113
Prin3	-0.1092	0.009	-12.267	0.000	-0.127	-0.092
=====						
Omnibus:	102.359	Durbin-Watson:	0.140			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	105.579			
Skew:	-0.341	Prob(JB):	1.19e-23			
Kurtosis:	2.801	Cond. No.	2.71			
=====						
Notes:						
[1] Standard Errors assume that the covariance matrix of the errors is correctly						

PCA분석과 회귀분석을 통해서

총과금액, 총매출로 비용이 중요함을 확인

3. Random Forest, Decision Tree, Gradient Boosting 을 통한 중요변수 도출

Y = ChurnLabel

GB의 중요변수

	Feature	Importance
6	noReferrals	0.337
20	AvgCharge	0.227
23	TenureMonths	0.125
4	noDependents	0.122
2	Married	0.043
12	TechSupport	0.036
22	TotalRevenue	0.030
18	TotalLongDistanceCharge	0.027
17	AvgLongDistanceCharge	0.020
15	TotalRefund	0.011

RF 중요변수

	Feature	Importance
4	noReferrals	0.331
17	AvgCharge	0.323
20	TenureMonths	0.145
18	TotalCharge	0.048
19	TotalRevenue	0.034
16	TotalLongDistanceCharge	0.033
3	Referrals	0.023
10	TechSupport	0.020
7	OnlineSecurity	0.011
12	AvgDownloadGB	0.010
0	Married	0.007

DT 중요변수

	Feature	Importance
17	AvgCharge	0.201
4	noReferrals	0.193
20	TenureMonths	0.169
18	TotalCharge	0.054
2	noDependents	0.053
1	Dependents	0.051
3	Referrals	0.038
16	TotalLongDistanceCharge	0.037
12	AvgDownloadGB	0.033
6	PaperlessBilling	0.029
10	TechSupport	0.028
15	AvgLongDistanceCharge	0.028
19	TotalRevenue	0.026

ChurnLabel에 대해 DT와 GB, RF로 중요 변수를 확인해본 결과,
평균 과금액, 추천 횟수, 사용 기간이 도출됨

3. PCA 를 통한 중요변수 도출

Y = Churn Label _ Satis Score/ ChurnScore 포함

[528]:

1	SatisScore	0.893
7	noReferrals	0.038
21	AvgCharge	0.025
10	OnlineSecurity	0.022

DT의 중요변수 :

만족도, 추천횟수,
평균 과금액

Y = LTV

	Feature	Importance
8	TotalCharge	0.381
9	TotalRevenue	0.379
7	AvgCharge	0.182
5	AvgLongDistanceCharge	0.030
6	TotalLongDistanceCharge	0.014

GF의 중요변수 :

총 과금액, 총 매출액,
평균 과금액

3. Random Forest, Decision Tree, Gradient Boosting 을 통한 중요변수 도출

최종 GB

```
[156]: 1 y_pred = rf_final.predict(X_test)
      2 print('Accuracy: {:.3f}'.format(accuracy_score(y_test, y_pred)))
      3
      4 print('Confusion matrix: \n', confusion_matrix(y_test, y_pred))
```

```
Accuracy: 0.967
Confusion matrix:
[[1672   0]
 [  66 272]]
```

```
[157]: 1 a=272/(66+272) #recall
      2 b=272/(272) #precision
      3 print("recall : {:.3f}".format(a))
      4 print("precision : {:.3f}".format(b))
      5
      6 c=2*(a*b)/(a+b) #F1 score
      7 print("F1 score : {:.3f}".format(c))
```

```
recall : 0.805
precision : 1.000
F1 score : 0.892
```

최종 RF

```
[128]: y_pred = tree_final.predict(X_test)
      print('Accuracy: {:.3f}'.format(accuracy_score(y_test, y_pred)))
      print('Confusion matrix: \n', confusion_matrix(y_test, y_pred))
```

```
Accuracy: 0.972
Confusion matrix:
[[1230   8]
 [  34 235]]
```

```
[151]: a=235/(34+235) #recall
      b=235/(8+235) #precision
      print("recall : {:.3f}".format(a))
      print("precision : {:.3f}".format(b))
      c=2*(a*b)/(a+b) #F1 score
      print("F1 score : {:.3f}".format(c))
```

```
recall : 0.874
precision : 0.967
F1 score : 0.918
```

최종 DT

```
y_pred = gb_final.predict(X_test)
print('Accuracy: {:.3f}'.format(accuracy_score(y_test, y_pred)))
print('Confusion matrix: \n', confusion_matrix(y_test, y_pred))
```

```
Accuracy: 0.968
Confusion matrix:
[[1233   5]
 [  43 226]]
```

```
a=226/(43+226) #recall
b=226/(5+226) #precision
print("recall : {:.3f}".format(a))
print("precision : {:.3f}".format(b))
c=2*(a*b)/(a+b) #F1 score
print("F1 score : {:.3f}".format(c))
```

```
recall : 0.840
precision : 0.978
F1 score : 0.904
```

최종 Neural Network

```
print('Accuracy on training set: {:.3f}'.format(accuracy_score(y_train, y_train_pred)))
print('Accuracy on test set: {:.3f}'.format(accuracy_score(y_test, y_test_pred)))
warnings.warn('')
print('Confusion matrix: \n', confusion_matrix(y_test, y_test_pred))
```

```
Confusion matrix:
[[1641   31]
 [  47 291]]
```

```
recall = 291/(291+47)
recall
```

```
0.8609467455621301
```

```
precision = 291/(291+31)
precision
```

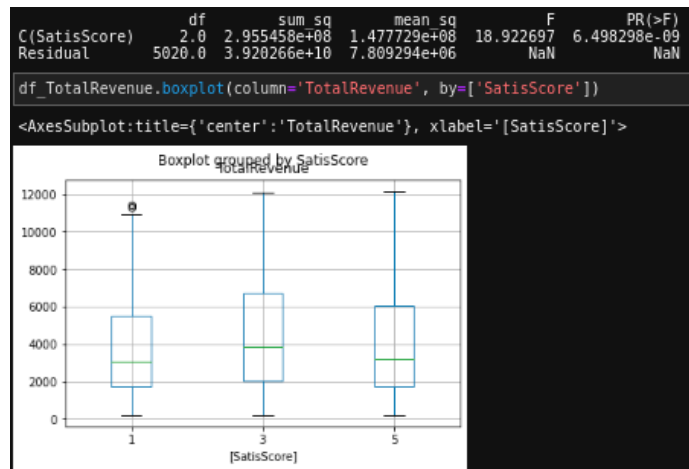
```
0.9037267080745341
```

```
f1score = 2*((recall*precision)/(recall+precision))
f1score
```

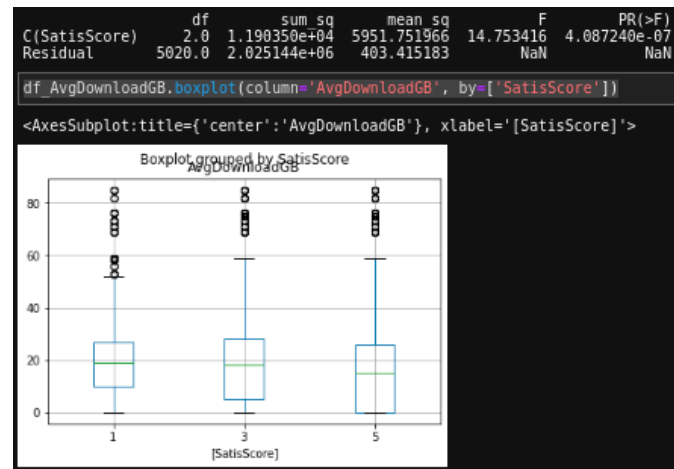
```
0.8818181818181817
```

만족도(SatisScore)에 대한 깊은 분석

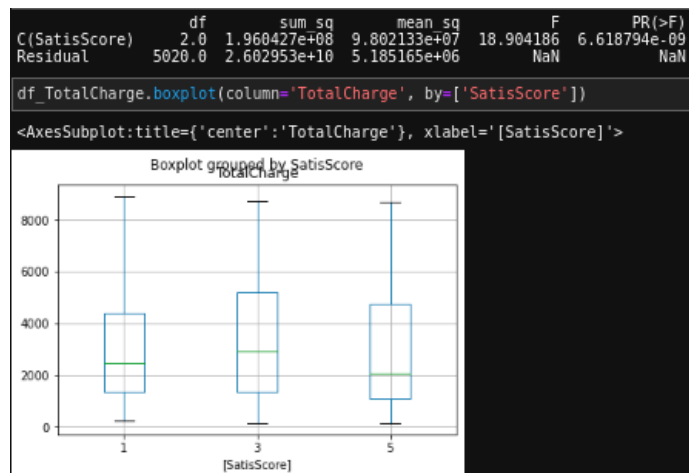
Variables for each SatisScore



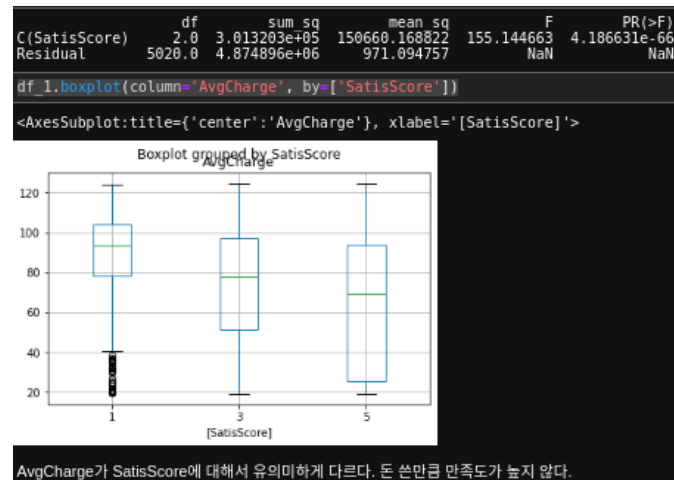
만족도에 따라 총 매출은
유의미하게 다르다



만족도에 따라 평균 데이터
사용량은 유의미하게 다르다
➢ 만족도가 낮을 수록 데이터
사용량이 크다.
➢ 데이터에 불만족



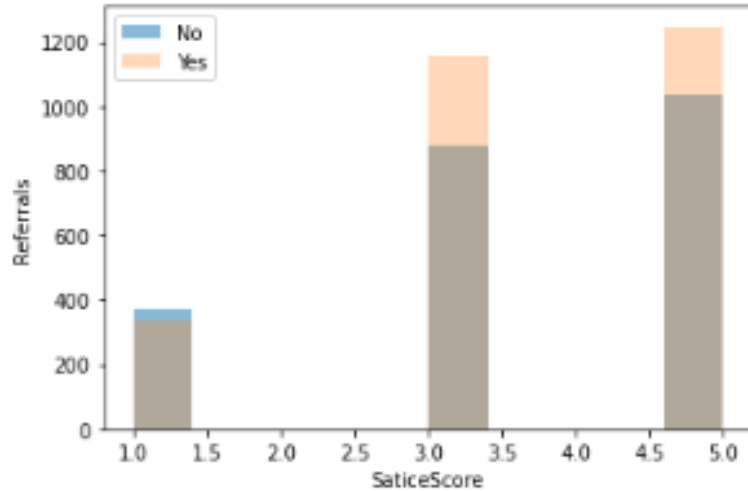
만족도에 따라 총 요금은
유의미하게 다르다



만족도에 따라 평균 요금은 다르다.
➢ 만족도가 낮을수록 평균요금
비싸다
➢ 비용값을 못함
➢ 데이터나 부가서비스 혹은
통화품질이 안 좋다

AvgCharge가 SatisScore에 대해서 유의미하게 다르다. 돈 쓴만큼 만족도가 높지 않다.

만족도(SatiscScore)에 대한 깊은 분석

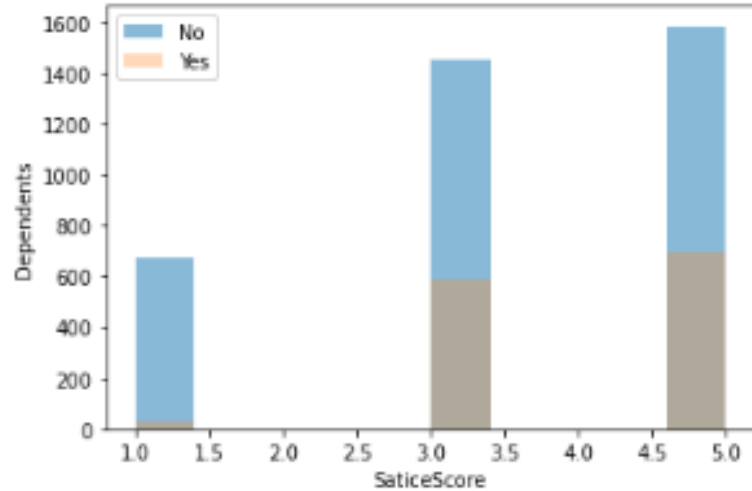


결혼 안한 사람의 만족도가 대부분 낮다.

- > 결혼 안한 사람 데이터 사용량이 많다
- > 데이터 속도가 느리거나 비용이 비싸다

결혼 안한 사람만 하고 있는 것은 무엇인가?

- 추천을 안함.
- 계약 년수가 1년임.

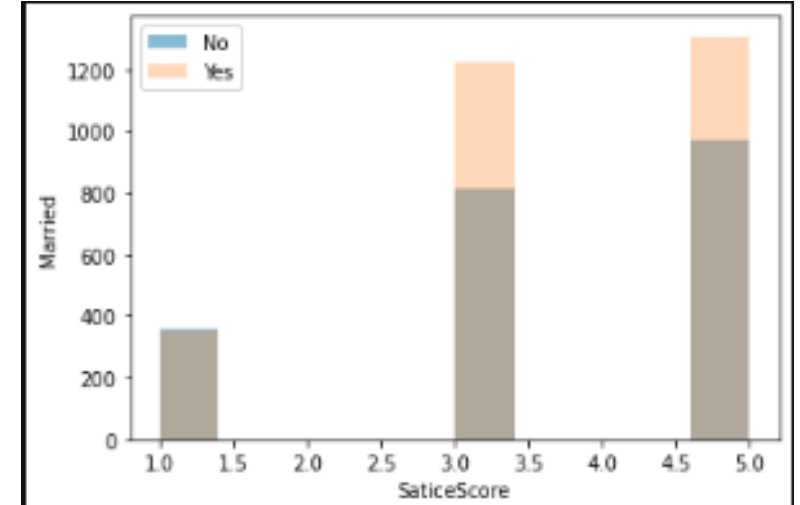


부양자가 없는 사람의 만족도가 낮다

- > 부양자 없는 사람 데이터 사용량 높다
- > 데이터 속도가 느리거나 비용이 비싸다.

부양자가 없는 사람만 하고 있는 것은?

- 추천을 잘 안함.
- 계약 년수가 1년
- 종이 영수증을 많이 받음
- 무제한 데이터를 많이 씀
- 평균 과금액이 많음



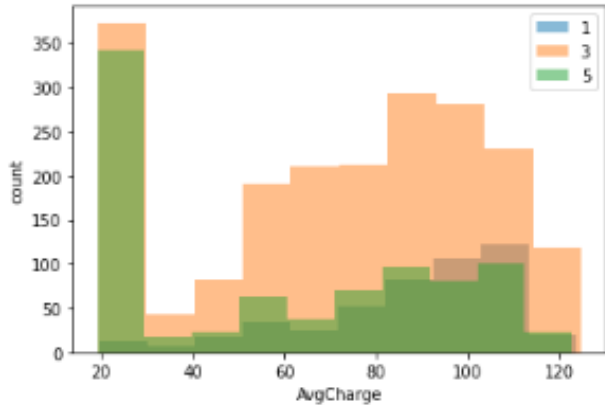
만족도가 낮은 사람의 대부분은 추천이 없다.

추천이 없는 사람이 더 많이 한 것은?

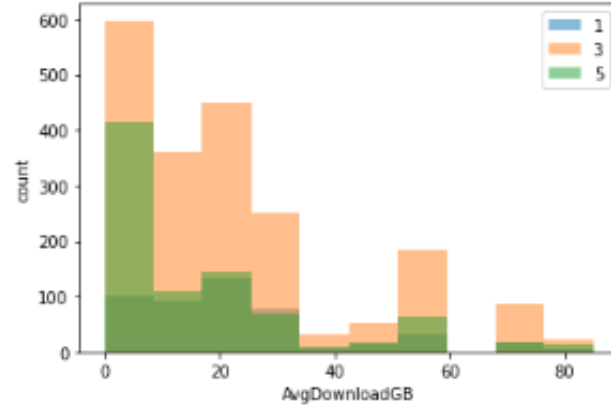
- 계약 년수가 1년임

만족도(SatisScore)에 대한 깊은 분석

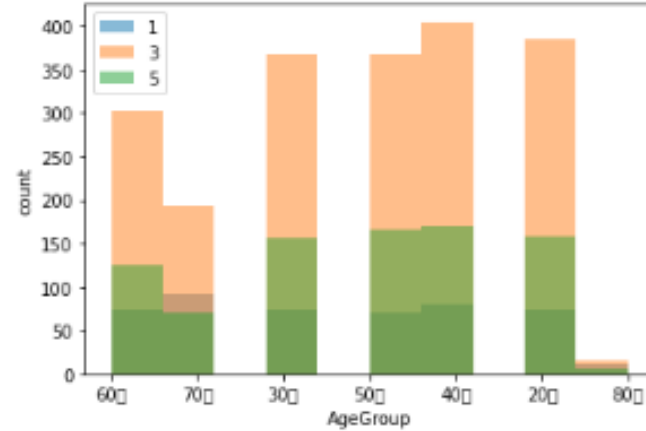
1,2점은 1점으로, 4,5점은 4점으로 통일시켜 만족도간 차이를 뚜렷하게 확인해보려 하였다.



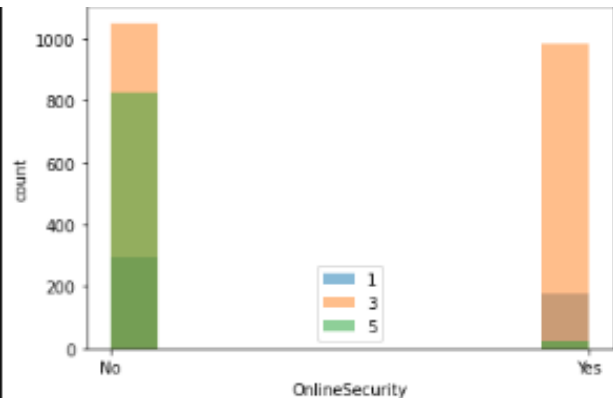
데이터를 안 쓰는 사람 만족도가 높다.
이것으로 보아 데이터 속도가 느리거나
비용이 비싸다.



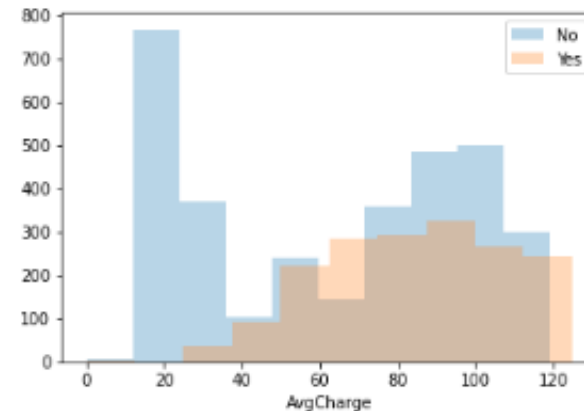
평균 요금이 낮은 사람의 만족도가 높다.
이것으로 보아 돈값을 못한다고 추측할 수 있다.
예로 데이터 비용이 비정상적으로 비싸거나
통화요금이 비정상적으로 비싸다.



70대에서 만족도가 낮다.
70대를 집중적으로 조사해볼 필요가 있다.
그런데 이 데이터는 해외 데이터로 70대가 장거리 전화요금(해외 통신특성)에
의해 요금은 높는데 만족도가 낮은 것으로 밝혀져서 7,80대는 제외하기로 한다.



온라인 보안 기능을 쓰는 사람 만족도가
매우 낮은걸로 보아 온라인 보안 기능의
문제점 파악이 필요하다



평균 과금액이 높을수록
온라인 보안 서비스를 많이 받는다.

연관분석을 통해 서비스간의 연관성 분석

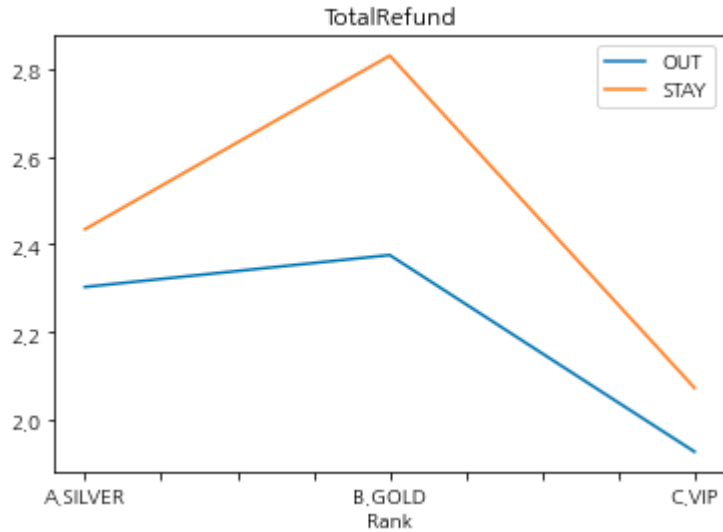
연관분석_서비스 4개 + 무제한 데이터

support	itemsets
0 0.350	(OnlineSecurity)
1 0.417	(OnlineBackup)
2 0.421	(DeviceProtection)
3 0.354	(TechSupport)
4 0.676	(UnlimitedData)
5 0.208	(OnlineSecurity, OnlineBackup)
6 0.208	(OnlineSecurity, DeviceProtection)
7 0.202	(OnlineSecurity, TechSupport)
8 0.304	(OnlineSecurity, UnlimitedData)
9 0.246	(OnlineBackup, DeviceProtection)
10 0.211	(OnlineBackup, TechSupport)
11 0.357	(OnlineBackup, UnlimitedData)
12 0.222	(DeviceProtection, TechSupport)
13 0.365	(DeviceProtection, UnlimitedData)
14 0.304	(UnlimitedData, TechSupport)
15 0.133	(OnlineSecurity, DeviceProtection, OnlineBackup)

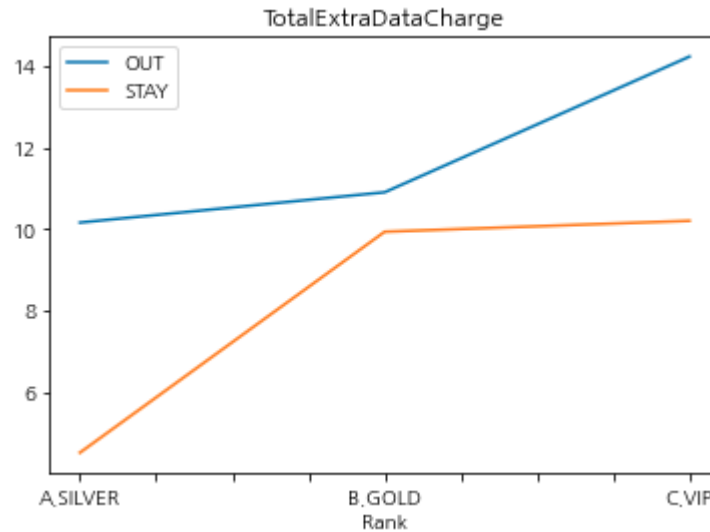
16	0.130	(OnlineSecurity, OnlineBackup, TechSupport)
17	0.180	(OnlineSecurity, OnlineBackup, UnlimitedData)
18	0.135	(OnlineSecurity, DeviceProtection, TechSupport)
19	0.183	(OnlineSecurity, DeviceProtection, UnlimitedData)
20	0.174	(OnlineSecurity, UnlimitedData, TechSupport)
21	0.138	(OnlineBackup, DeviceProtection, TechSupport)
22	0.211	(OnlineBackup, DeviceProtection, UnlimitedData)
23	0.180	(OnlineBackup, UnlimitedData, TechSupport)
24	0.191	(DeviceProtection, UnlimitedData, TechSupport)
25	0.091	(OnlineSecurity, DeviceProtection, OnlineBacku...
26	0.116	(OnlineSecurity, DeviceProtection, OnlineBacku...
27	0.112	(OnlineSecurity, UnlimitedData, OnlineBackup, ...
28	0.118	(OnlineSecurity, UnlimitedData, DeviceProtecti...
29	0.118	(OnlineBackup, UnlimitedData, DeviceProtection...
30	0.079	(DeviceProtection, TechSupport, OnlineBackup, ...

서비스 가입자 중
무제한 데이터 가입자가
가장 많으며, 연관성이 가장 크다.

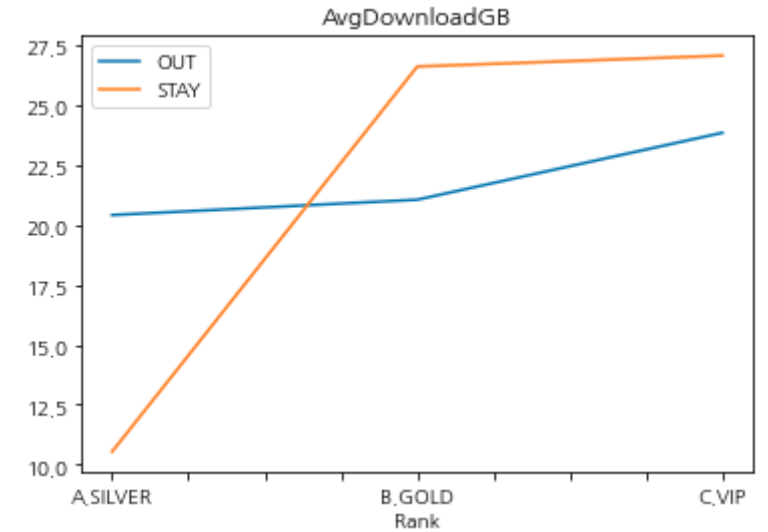
멤버십별 유의미한 그래프



전체 환불에 있어서 확실히 VIP 고객이
가장 적은 환불 수를 보임



추가 비용에 있어서 VIP가 제일 많이 내는데,
이탈한 고객 말고 이탈하지 않은 고객이 덜
내는 것을 확인할 수 있음
→ 결국 이탈한 고객은 돈이 많이 들어서
이탈했다고 추측가능



VIP 고객이 가장 많은 데이터 다운로드 양을
보임을 확인