

2025

고객 이탈 리스크 기반 세그먼트 분석 및 리텐션 전략 제안

서상우

프로젝트 개요

본 프로젝트는 Kaggle에서 제공하는 Telco 고객 데이터를 활용하여, 고객의 이탈(Churn)에 영향을 주는 요인을 분석하고, 이탈 위험이 높은 고객군을 조기에 식별하여 맞춤형 리텐션 전략 수립에 기여하는 것을 목표로 한다.

단순한 이탈 여부 예측을 넘어, 고객의 계약 형태, 요금 수준, 서비스 이용 여부 등 다양한 속성을 바탕으로 시간 흐름에 따른 이탈 위험 요인을 정량적으로 도출하였다.

이를 바탕으로, 주요 변수를 이용해 고객을 클러스터링한 뒤, 각 고객군별 생존 패턴을 비교 분석함으로써 현실적인 고객 유형별 유지 전략 수립에 활용 가능한 인사이트를 도출하였다.

분석 결과는 고객 관리 우선순위 설정, 요금제 설계 및 서비스 구성 등 실질적인 전략 수립에 기여할 수 있을 것으로 기대된다.

(데이터 출처: [Telco Customer Churn](#))

● 분석 흐름 요약

1. 고객군 특성 분석

- 계약 유형, 요금, 서비스 사용 여부 등 주요 항목별 이탈률 비교

2. 고객 이탈 요인 분석

- CoxPH 생존 분석을 통한 시간 기반 이탈 위험 요인 해석

3. 고객 세그먼트별 생존 패턴 비교

- 앞서 도출한 주요 위험요인 기반으로 고객군 클러스터링(K-Means 등) 수행
- 각 고객군별 Kaplan-Meier 생존곡선 비교 → 고객 유형별 유지 패턴 시각화
- 분석 결과를 바탕으로 고객 유형별 유지 전략 수립

데이터 개요 및 전처리

■ 데이터는 총 7,043명의 고객 정보로 구성되어 있으며, 주요 변수는 아래와 같음:

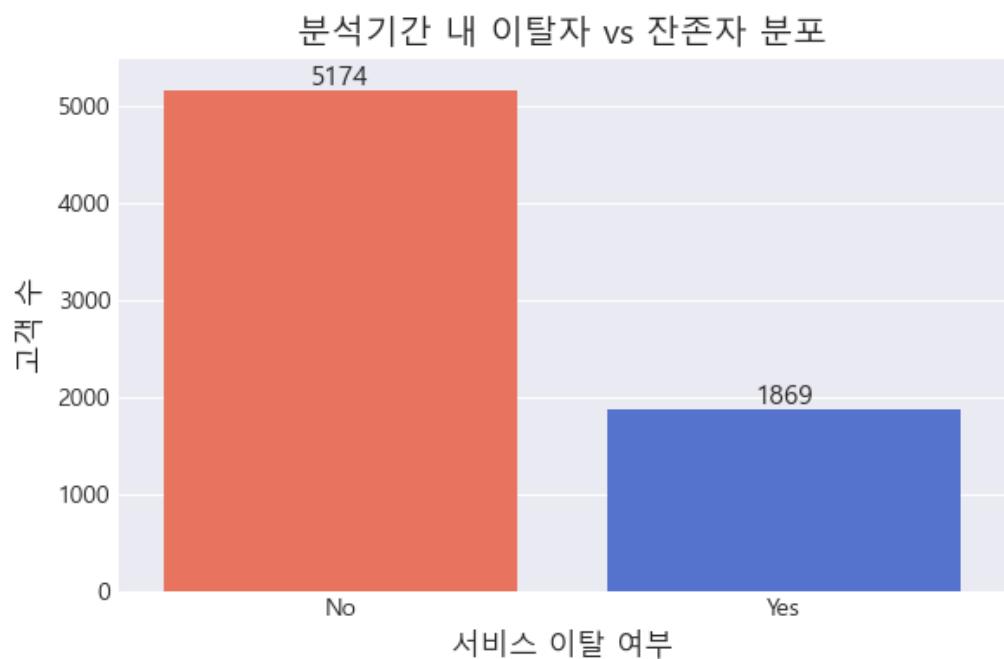
- 고객 ID
- 이탈 여부 ('Churn')
- 서비스 지속 기간 ('Tenure')
- 계약 형태 ('Contract': Month-to-month, One year, Two year)
- 부가 서비스 이용 여부 ('InternetService', 'OnlineSecurity', 등)
- 요금 관련 정보 ('MonthlyCharges', 'TotalCharges')
- 기타 항목 (성별, 부양가족 여부, 결제 방법 등)

■ 전처리 수행 내용:

- 범주형 변수(Label 컬럼) → 숫자로 변환 (Label Encoding 또는 One-hot Encoding)
- 'TotalCharges' 컬럼의 문자형 → 숫자형 변환 및 결측값 제거
- 'Churn' 값을 1(이탈), 0(잔존)으로 이진화
- 불필요한 ID 컬럼 제거

1. 고객군 특성 분석 (탐색적 분석)

1) 전체 고객 및 이탈 고객 현황



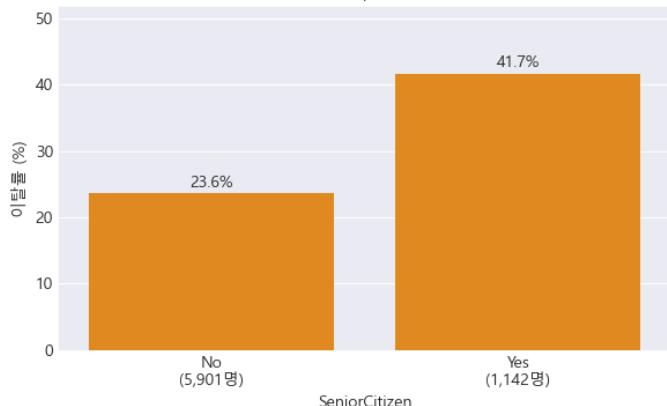
- 분석 대상 고객 수: 총 7,043명
- 이탈 고객 수: 1,869명 (전체의 약 26.5%)

2) 주요 변수별 이탈률 비교

변수	이탈률												
<p>계약 형태</p> <p>계약 형태별 이탈률</p> <table border="1"> <thead> <tr> <th>계약 형태</th> <th>이탈률 (%)</th> <th>고객 수</th> </tr> </thead> <tbody> <tr> <td>월 단위</td> <td>42.7%</td> <td>(3,875명)</td> </tr> <tr> <td>1년 계약</td> <td>11.3%</td> <td>(1,695명)</td> </tr> <tr> <td>2년 계약</td> <td>2.8%</td> <td>(1,473명)</td> </tr> </tbody> </table>	계약 형태	이탈률 (%)	고객 수	월 단위	42.7%	(3,875명)	1년 계약	11.3%	(1,695명)	2년 계약	2.8%	(1,473명)	<p>월간 계약 고객의 이탈률이 월등히 높음 → 장기 계약 유도가 핵심</p>
계약 형태	이탈률 (%)	고객 수											
월 단위	42.7%	(3,875명)											
1년 계약	11.3%	(1,695명)											
2년 계약	2.8%	(1,473명)											
<p>자동 이체 여부</p> <p>자동이체 여부별 이탈률</p> <table border="1"> <thead> <tr> <th>자동이체 여부</th> <th>이탈률 (%)</th> <th>고객 수</th> </tr> </thead> <tbody> <tr> <td>No</td> <td>16.3%</td> <td>(2,872명)</td> </tr> <tr> <td>Yes</td> <td>33.6%</td> <td>(4,171명)</td> </tr> </tbody> </table>	자동이체 여부	이탈률 (%)	고객 수	No	16.3%	(2,872명)	Yes	33.6%	(4,171명)	<p>자동이체 이용 고객의 이탈률 상대적으로 높음</p>			
자동이체 여부	이탈률 (%)	고객 수											
No	16.3%	(2,872명)											
Yes	33.6%	(4,171명)											
<p>부양가족 유무</p> <p>부양가족 유무별 이탈률</p> <table border="1"> <thead> <tr> <th>부양가족 유무</th> <th>이탈률 (%)</th> <th>고객 수</th> </tr> </thead> <tbody> <tr> <td>No</td> <td>31.3%</td> <td>(4,933명)</td> </tr> <tr> <td>Yes</td> <td>15.5%</td> <td>(2,110명)</td> </tr> </tbody> </table>	부양가족 유무	이탈률 (%)	고객 수	No	31.3%	(4,933명)	Yes	15.5%	(2,110명)	<p>부양 가족이 있는 고객은 이탈률이 낮음</p>			
부양가족 유무	이탈률 (%)	고객 수											
No	31.3%	(4,933명)											
Yes	15.5%	(2,110명)											

노년층 여부

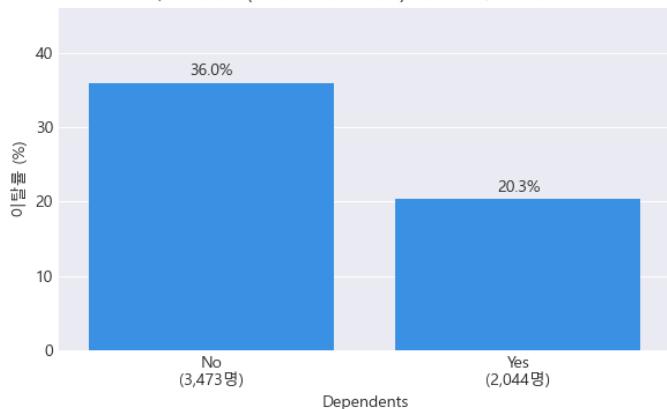
노년층 여부별 이탈률



노년층의 이탈률이 다소 높음

부가서비스 이용 여부

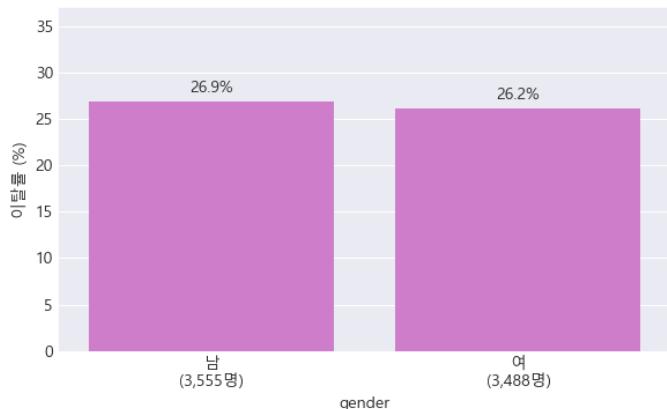
부가서비스(인터넷 기술 지원) 이용 유무별 이탈률



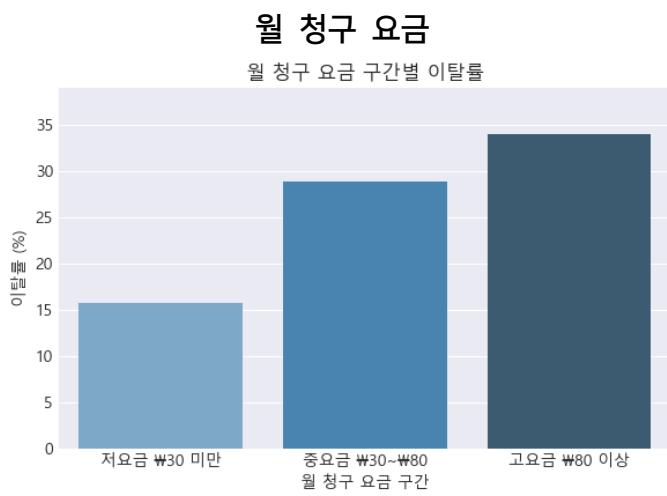
부가서비스 이용 고객의 이탈률이 낮음

성별

성별 이탈률



성별에 따른 이탈률의 큰 차이는 없음



월 요금 부담이 낮을수록 이탈률이 낮음

3) 소결 및 요약

- 계약 형태에 따라, 부가서비스 이용 여부, 고령 등의 특성, 요금제 등 특성이 이탈률과 유관하다는 부분을 간략히 살펴봄
- 고객군 분류, 생존분석 및 리텐션 전략 설계 시 설명 변수 후보군으로 유의미
- 해당 분석 결과는 향후 이탈 예측 모델 및 세그먼트 전략 수립을 위한 변수 탐색 목적이며, 실제 유의성 검정은 후속 분석에서 진행

2. 고객 이탈 요인 분석

1) 분석 목적

- 고객의 이탈 여부(Churn) 뿐만 아니라, 이탈까지 걸리는 시간을 함께 고려하여, 어떤 요인이 시간에 따라 고객 이탈 위험을 높이지는지를 파악하기 위해 Cox 비례위험 모형을 활용
- 이를 통해 이탈에 영향을 미치는 핵심 위험 요인(Hazard Ratio)를 도출하고자 함

2) 변수 선택 과정

- 분석 초기에는 고객 특성과 요금·계약·서비스 항목을 포괄하는 다수의 변수들이 고려되었으나, 이후 다음과 같은 기준을 바탕으로 변수 정제를 진행

기준	설명
해석 가능성	실무적 메시지가 명확한 변수 우선 채택
다중공선성 제거	중복되거나 상관관계가 높은 변수 제거
비선형성 보완	요금 변수는 구간화하여 범주형 변수로 재구성
희소범주 제거	빈도 낮은 서비스 항목은 통합 또는 제외

- 특히 월 청구 요금('MonthlyCharges')은 비선형성 보완 및 해석의 편의성을 위해 일정 구간(저·중·고 요금)으로 구분하여 변수를 재가공 하였음.

- 최종 변수 목록 (8개 내외)

변수명	설명
-----	----

Contract	계약 유형 (단기, 1년, 2년)
MonthlyChargesGroup	요금 수준 그룹 (저·중·고)
PaperlessBilling	무지류 자동이체 여부
TechSupport, OnlineSecurity	핵심 부가 서비스 이용 여부
SeniorCitizen, Dependents	고객 특성 관련 변수

3) 분석 방법 및 모델

- 모형 종류: Cox Proportional Hazards Model

- 종속 변수

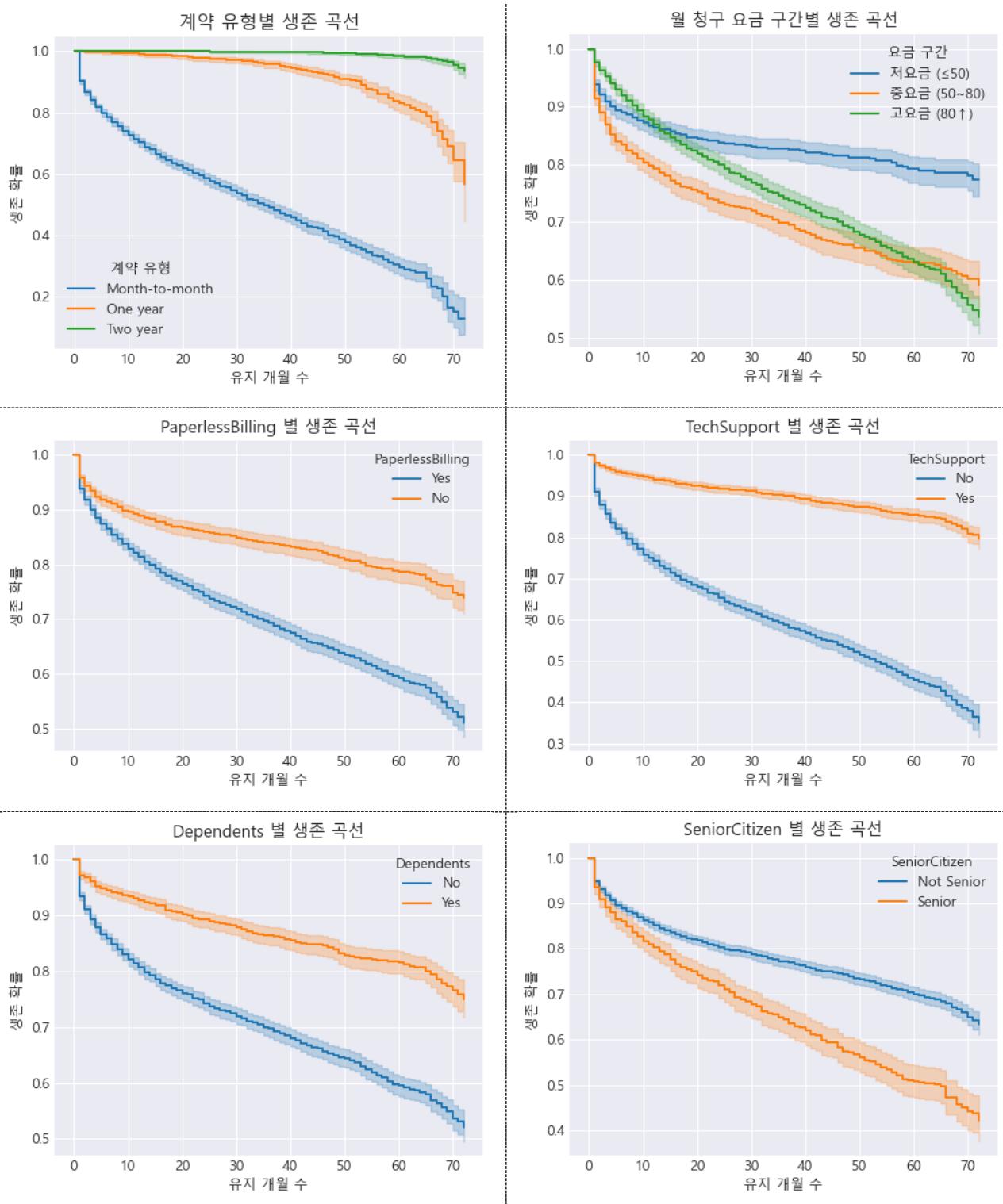
- Tenure (서비스 유지 개월 수)
 - Churn (이탈 여부: 1 = 이탈, 0 = 유지)

- 해석 기준

- 계수(β)의 지수화 값 $\exp(\beta)$ = Hazard Ratio (HR)
 - $HR > 1$: 이탈 위험 증가 / $HR < 1$: 이탈 위험 감소

4) 분석 결과 요약

- Kaplan-Meier 생존 분석 결과 시각화



■ CoxPH 모델 결과 요약

변수	계수 (β)	HR ($\exp(\beta)$)	p-value	해석 요약
Contract_One year	-2.04	0.13	<0.005	1년 계약 고객은 단기 계약 대비 이탈 위험 87% 감소
Contract_Two year	-3.95	0.02	<0.005	2년 계약 고객은 단기 계약 대비 이탈 위험 98% 감소
TechSupport_Yes	-0.51	0.6	<0.005	기술지원 이용 시 이탈 위험 40% 감소
Dependents_Yes	-0.42	0.66	<0.005	부양가족 있으면 이탈 위험 34% 감소
PaperlessBilling_Yes	0.24	1.27	<0.005	무지류 자동결제 사용 고객은 이탈 위험 27% 증가
MonthlyChargesGroup_중요금	0.2	1.22	<0.005	중간 요금대 고객은 저요금 대비 이탈 위험 22% 증가
MonthlyChargesGroup_고요금	-0.05	0.96	0.49	고요금은 통계적으로 유의하지 않음
SeniorCitizen	-0.1	0.9	0.06	노년층은 약간의 위험 감소 경향 있으나 유의하지 않음

5) 핵심 인사이트 요약

■ 계약 기간이 가장 강력한 이탈 방지 요인

- 특히 2년 계약 고객의 HR=0.02로 이탈 위험 매우 감소

■ 부가서비스(기술 지원 서비스 등) 이용 시 서비스 이탈 위험 감소

■ 부양 가족 여부는 생존 기간에 긍정적 영향

- 부양 가족이 있는 고객이 더 오래 유지되는 경향

■ 자동 이체 고객(Paperless Billing) 은 이탈 경향이 강함

- 해지가 간편해서?

■ 요금 수준은 혼재된 결과

- KM 곡선상 고요금 고객은 초기 생존률은 높지만, 시간이 지나며 하락 폭이 크고 장기 생존률은 오히려 낮아지는 경향
- CoxPH 분석에서는 중요금 고객만 이탈 위험 유의 증가 ($HR=1.22$)
→ 요금만으로는 이탈 위험을 판단하긴 어렵고, 다른 변수와의 복합 영향 고려 필요



본 분석을 통해 이탈에 영향을 미치는 핵심 위험 요인을 도출할 수 있었으며, 다음 장에서는 이들 요인 기반으로 고객군을 구분하고, 각 고객 세그먼트의 생존 패턴을 비교함으로써 보다 정밀한 타겟 리텐션 전략 도출 가능성을 확인할 예정

3. 고객 세그먼트별 생존 패턴 비교

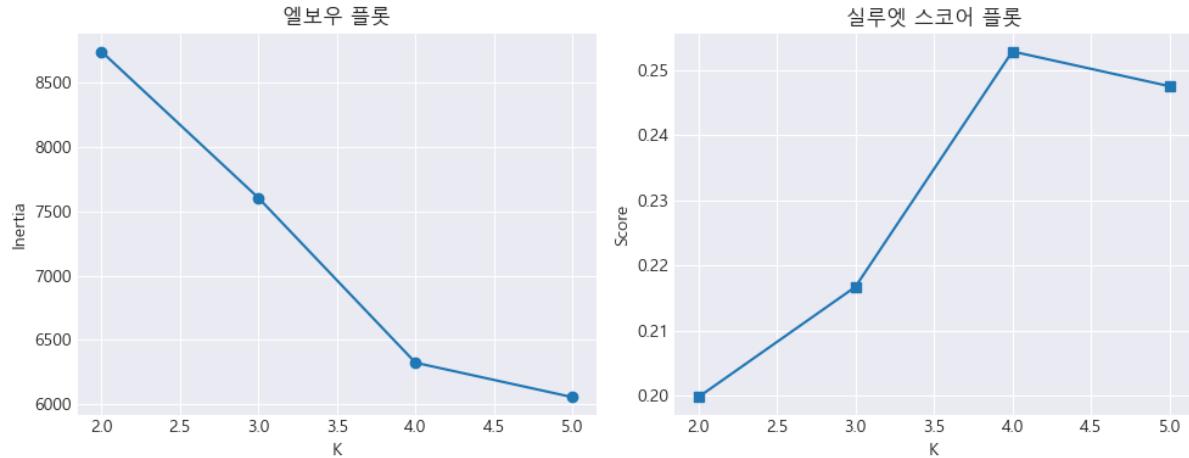
1) 분석 개요 및 목적

- 고객 이탈의 위험 요인이 단일 변수로 설명되지 않고, 다양한 속성의 조합으로 나타나는 경우가 많다는 점에서, 단일 변수 기반 비교는 설명력에 한계가 있음
- 이에 본 장에서는 앞서 생존 분석 결과를 통해 도출된 주요 변수들을 바탕으로 고객을 다차원적으로 군집화하고, 각 세그먼트별 생존 행태의 구조적 차이를 분석하고자 함
- 이를 통해 다음과 같은 목적을 달성하고자 함:
 - 고객 이탈 행태에 영향을 주는 변수들의 조합 기반 세분화 시도
 - 각 고객군의 생존 패턴(이탈률, 유지 기간, 이탈 시점 등) 비교
 - 각 군집을 대상으로 하는 맞춤형 리텐션 전략 도출 가능성 확보

2) 분석 방법론

- 변수 선택
 - Contract: 계약 형태 (Month-to-month, One year, Two year)
 - TechSupport: 부가서비스(기술 지원 서비스) 이용 여부
 - PaperlessBilling: 자동 이체 여부
 - MonthlyCharges Group: 월 청구 요금 (저/중/고 요금)
 - Dependents: 부양 가족 여부
- 클러스터링 기법
 - 모델: K-Means 클러스터링

- 클러스터 수 결정 기준: 엘보우 메서드 분석 기반으로 적정 K 검토 (*K=4)



3) 클러스터링 결과 요약

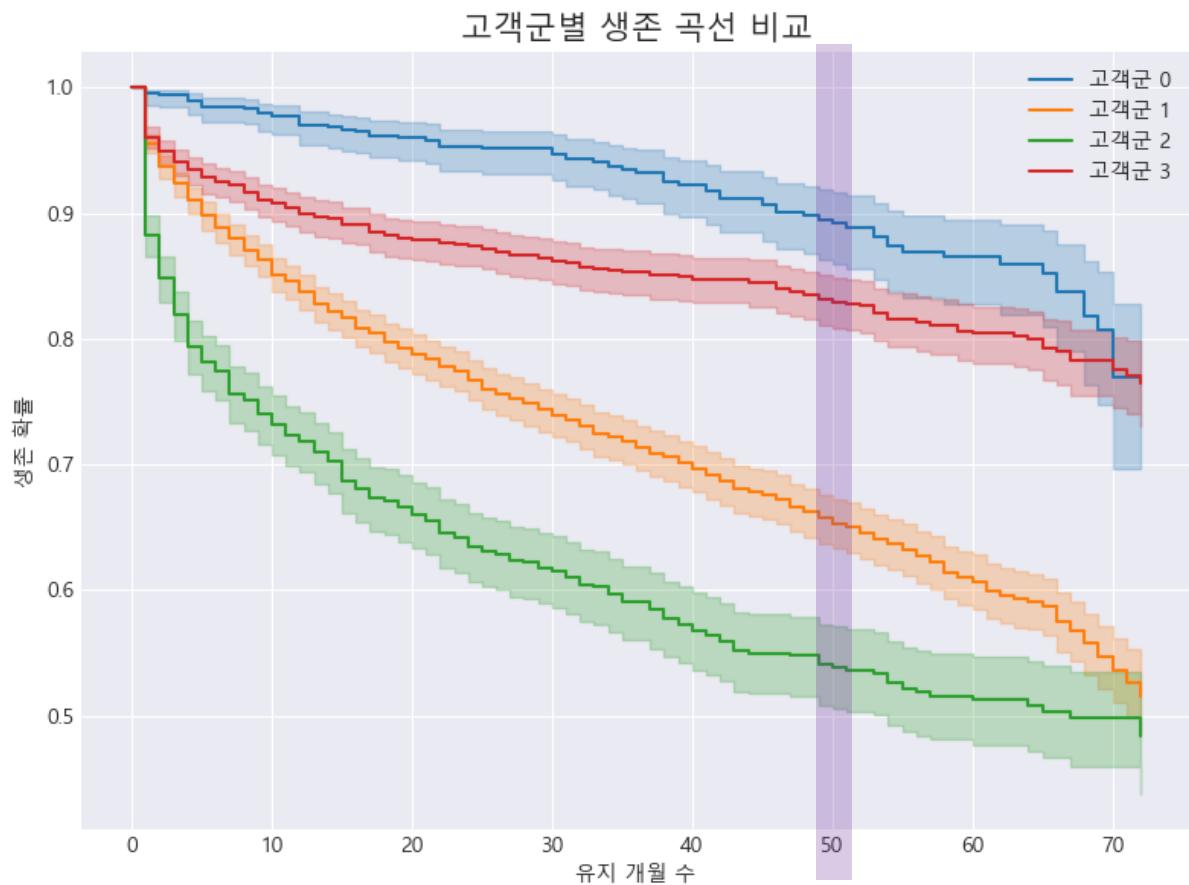
- 선택된 클러스터 수: K=4

- 각 클러스터의 특성 요약

	계약기간 1년	계약기간 2년	부가서비스 (기술지원) 사용 O	자동이체 사용 O	월 청구 요금 중간 수준	월 청구 요금 높은 수준	부양가족 O	고객수
고객군 0	65%	20%	87%	22%	84%	10%	39%	692
고객군 1	19%	21%	29%	100%	0%	70%	26%	2,872
고객군 2	7%	13%	19%	78%	100%	0%	22%	1,475
고객군 3	19%	38%	16%	0%	1%	29%	38%	2,004

고객군	주요 특성 요약
고객군 0	<p>중요금+1년 계약+기술지원 이용</p> <ul style="list-style-type: none"> 계약기간 1년 비율이 가장 높음(65%) 기술지원 서비스 이용률 87%로 매우 높음 월 청구요금은 중간 수준(84%)이 주를 이룸 부양가족 있음: 39%→ 안정적 이용 성향, 서비스 충성도 높음
고객군 1	<p>고요금+자동이체+기술지원 적음</p> <ul style="list-style-type: none"> 자동이체 100%로 결제 효율적 고요금 이용자 70%로 수익성 높은 군 기술지원 이용률 낮음(29%)→ 디지털 친화적 고가 요금제 이용자, 서비스 확대 여지
고객군 2	<p>저요금+기술지원 적음+가족 없음</p> <ul style="list-style-type: none"> 중간 요금(100%), 고요금 없음 기술지원 이용률 낮음(19%)• 부양가족 거의 없음(22%) 계약기간도 짧은 편(1년: 7%, 2년: 13%)→ 이탈 가능성 높은 단기·저가 사용자
고객군 3	<p>장기계약+고요금+가족 보유</p> <ul style="list-style-type: none"> 2년 계약 비중 38%로 가장 높음• 고요금(29%), 부양가족 비율 높음(38%) 기술지원, 자동이체 이용 낮음→ 전통적 충성고객군, 다소 수동적 이용 패턴

4) 고객군별 생존 곡선 비교



■ 고객군 0

- 전체 기간 동안 가장 높은 생존율 유지
- 50개월 시점 기준 잔존율 90% 이상 고객이 유지되는 것으로 보임
→ 장기 유지 가능성이 높은 안정적 고객군

■ 고객군 1

- 중장기적으로 이탈률이 높은 집단
- 50개월 시점 기준 이탈률 30% 대로 이탈 위험이 높은 고객군

■ 고객군 2

- 초기부터 급격한 이탈이 발생하며, 전체 기간 중 가장 낮은 생존 곡선
- 50개월 시점 기준 절반 가량의 고객이 이탈

■ 고객군 3

- 상대적으로 느린 이탈 속도를 보임
- 50개월 시점 기준 10% 대 중후반의 고객 이탈 발생
→ 중간 수준의 유지력을 가진 고객군

5) 전략적 시사점 및 활용 방안

■ 이탈 취약 고객군 사전 식별 및 리스크 기반 관리

- 단기 계약, 기술지원 미이용, 부양 가족 없음, 저요금 이용 등의 특성을 가진 고객군 2는 이탈 위험이 가장 높음
- 이들은 장기적인 수익 기여도가 낮고 생존률 감소 속도도 가파르므로, 사전 리스크 알림 시스템을 통해 조기 개입이 필요함

■ 고수익·안정 고객군의 충성도 유지 전략 강화

- 고객군 0, 3은 계약 기간이 길고 기술지원·자동이체 이용률이 낮지 않아 생존률이 높음
- 이들의 충성도를 유지하기 위해 맞춤형 리워드 프로그램, 업셀링 전략 도입 가능

■ 서비스 적응률이 낮은 고객에 대한 기능성 확대

- 고객군 1은 자동이체는 이용하나 기술지원 이용률이 낮고 생존률도 중간 수준
- 디지털 접근성 또는 서비스 이해도 부족으로 인한 이탈 가능성이 있는 만큼, 튜토리얼·상담 지원 확대 전략 고려

■ 클러스터 기반 타깃 마케팅 및 맞춤 전략 수립

- 단일 변수 분석이 아닌 다변량 기반 고객군 분류를 통해 보다 정밀한 마케팅 및 요금제 설계 가능
- 각 고객군별 특성에 기반해 할인 혜택, 리텐션 캠페인, 서비스 번들링 전략 차등화 필요

마무리: 종합 결론 및 제언

주요 결론 요약

- 이탈 위험 요인의 정량화: CoxPH 생존 분석을 통해 계약 기간, 부가 서비스 이용, 요금 수준, 고객 특성(부양가족 등)이 이탈 위험에 유의미한 영향을 미치는 요인임을 확인하였다.
- 고객군별 생존 패턴 차별화: 클러스터링 기반 고객 세그먼트를 구성하고 Kaplan-Meier 곡선을 활용해 생존 패턴을 비교한 결과, 각 고객군의 특성과 이탈 양상에 명확한 차이가 존재함을 확인하였다.

전략적 제언

1. 이탈 위험군 조기 식별 시스템 도입

고객군 2와 같이 이탈 위험이 높은 특성을 가진 집단에 대해 리스크 기반 모니터링을 적용하고, 조기 경고 시스템을 통해 선제적 개입이 필요

2. 고충성 고객 대상 차별화된 리워드 프로그램 운영

장기 계약 및 부가 서비스 이용률이 높은 고객군(예: 군집 0, 3)은 충성 고객으로서 장기적인 수익 기여도가 높으므로, 포인트 적립, 업셀링 캠페인 등을 통한 유인 전략이 유효

3. 서비스 이해도 및 접근성 지원 확대

기술 지원 서비스 이용이 저조한 고객군은 서비스 미숙지 또는 디지털 접근성 문제일 가능성이 높으므로, 고객 맞춤형 튜토리얼, 채팅 상담 확대 등 기능성 강화가 필요

4. 클러스터 기반 맞춤 마케팅 체계 구축

단일 항목 분석을 넘어 다변량 기반으로 도출된 고객군에 맞춰 요금제, 프로모션, 유지 전략 등을 세분화하여 운영하는 것이 효과적