

# 통신사 고객 이탈 예측을 위한 탐색적 데이터 분석 보고서

# 1. 분석 목적

본 프로젝트는 통신사 고객 데이터를 바탕으로 고객 이탈(Churn) 여부를 예측하고, 이탈에 영향을 미치는 주요 요인을 도출하기 위해 탐색적 데이터 분석을 수행하는 것을 목적으로 합니다. 고객 이탈은 기업 수익성과 직결되는 중요한 경영 지표로, 신규 고객 확보보다 기존 고객 유지가 평균 5배 이상 비용 효율적이라는 점에서 이탈 방지 전략은 기업 경쟁력 강화의 핵심입니다. 이에 따라 다음과 같은 세부 목적을 설정하였습니다:

1. 고객 이탈 여부(Churn)를 기준으로 주요 변수(계약 유형, 요금, 가입 기간 등) 간의 분포 및 차이 탐색
2. 이탈 고객의 특성 및 패턴을 시각화하여 이탈 고위험군을 조기에 식별할 수 있는 근거 확보
3. 향후 예측 모델링에 활용 가능한 변수 구조 및 데이터 품질을 사전 점검
4. 마케팅 전략에 활용 가능한 인사이트 도출

## 2. 데이터셋 개요

이번 분석에 사용된 데이터는 Kaggle에서 공개된 Telco Customer Churn Dataset으로, 통신사 고객의 서비스 이용 이력 및 이탈 여부(Churn)를 포함한 7,000건 이상의 고객 데이터로 구성되어 있습니다.

이 데이터는 고객의 계약 방식, 요금제, 서비스 이용 항목, 결제 방식 등 다양한 특성과 함께 이탈 여부를 포함하고 있어, 고객 이탈 예측 모델 구축에 적합합니다.

항목	내용
데이터 수	7,032건
변수 수	총 20개 변수 (설명변수 19개 + 타겟변수 1개)
타겟 변수	Churn (Yes / No) – 고객 이탈 여부
데이터 유형	수치형(연속형) + 범주형 혼합 데이터
출처	Kaggle: Telco Customer Churn Dataset

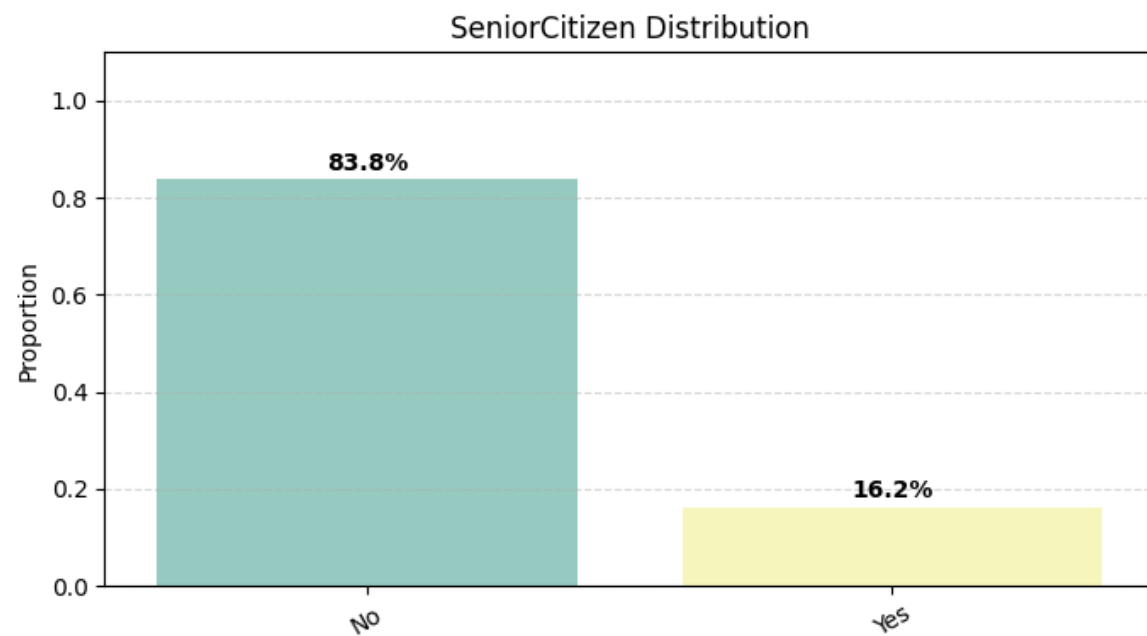
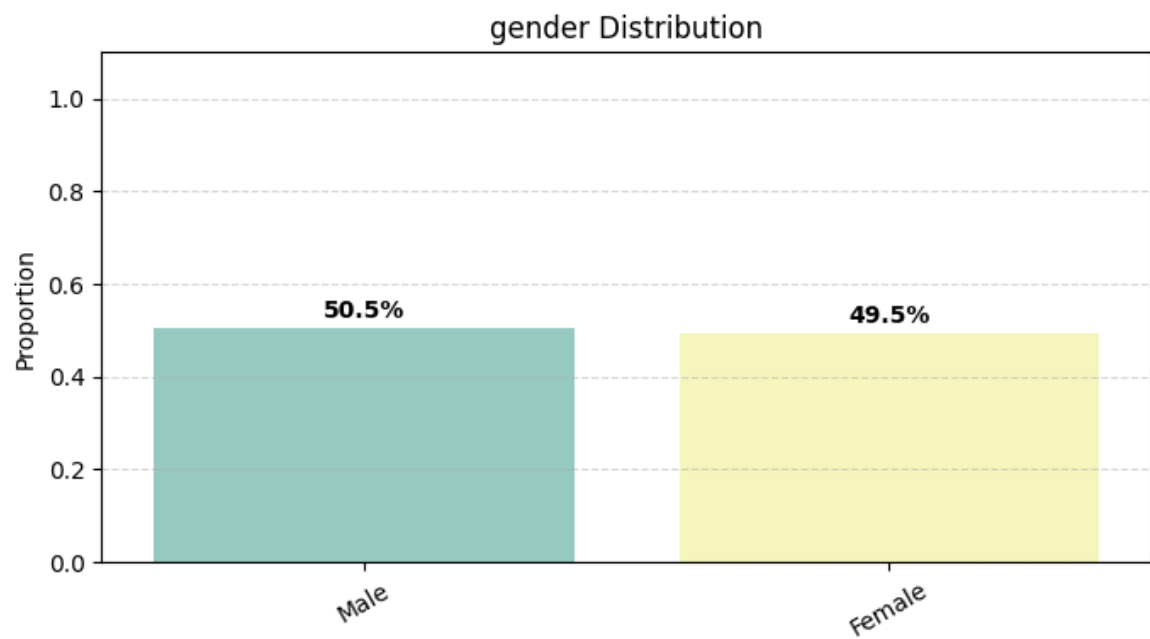
### 3. 기술통계요약

1) 수치형 기술 통계

	Tenure	Monthly Charges	Total Charges
count	7032	7032	7032
mean	32.4218	64.7982	2283.3004
Std	24.5453	30.0860	2266.7714
min	1.0000	18.2500	18.8000
25%	9.0000	35.5875	401.4500
50%	29.0000	70.3500	1397.4750
75%	55.0000	89.8625	3794.7375
Max	72.0000	118.7500	8684.8000

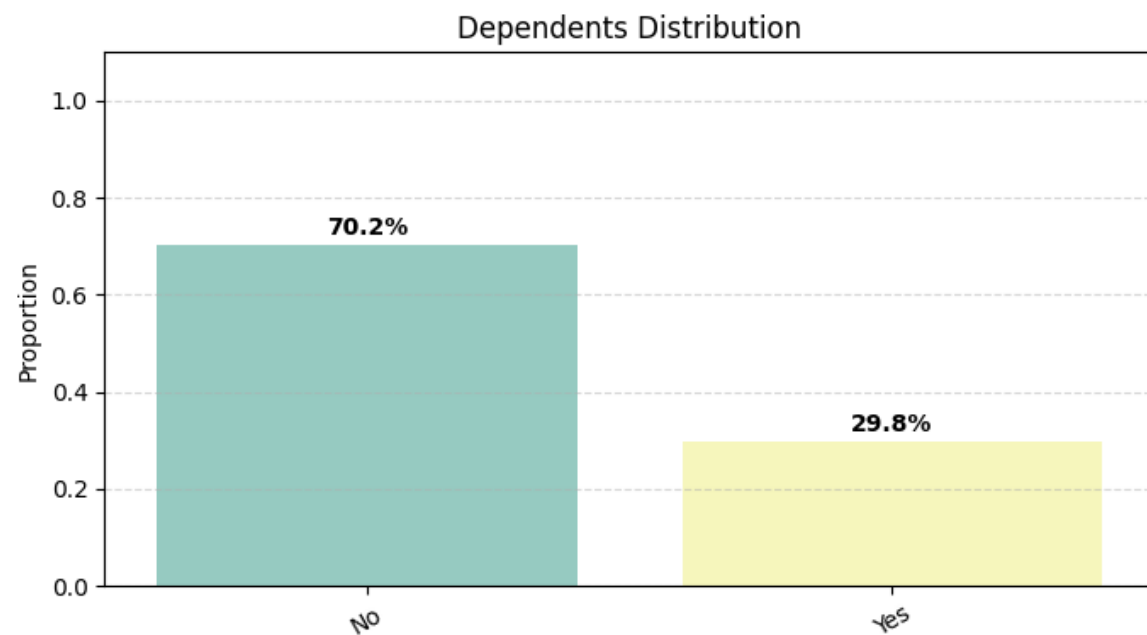
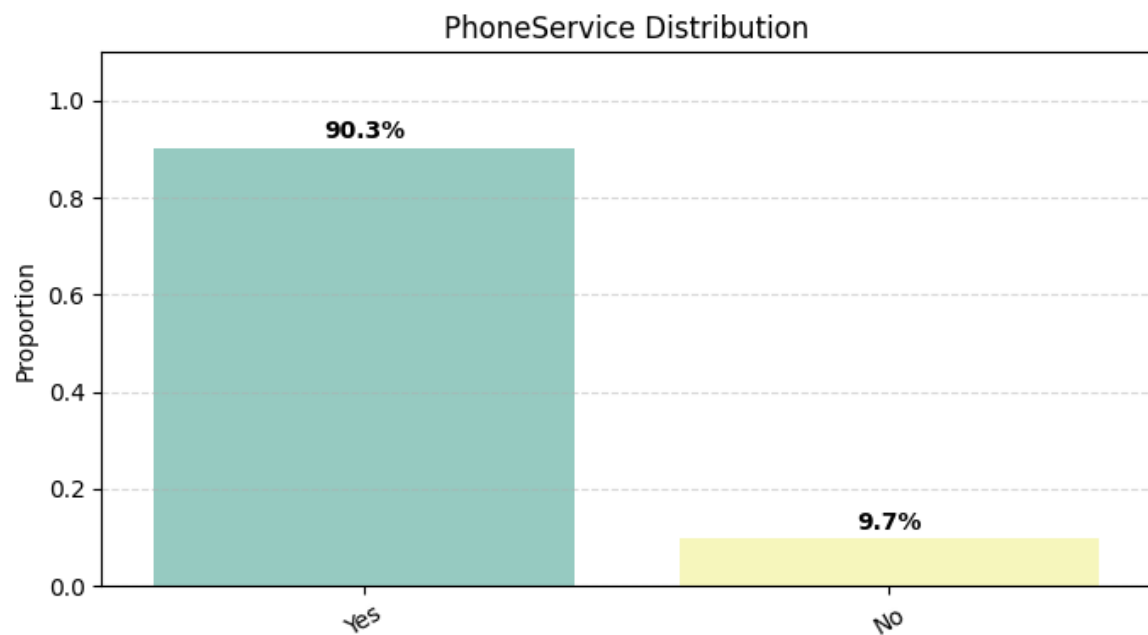
### 3. 기술통계요약

#### 1) 범주형 기술 통계 (주요 변수 시각화 결과)



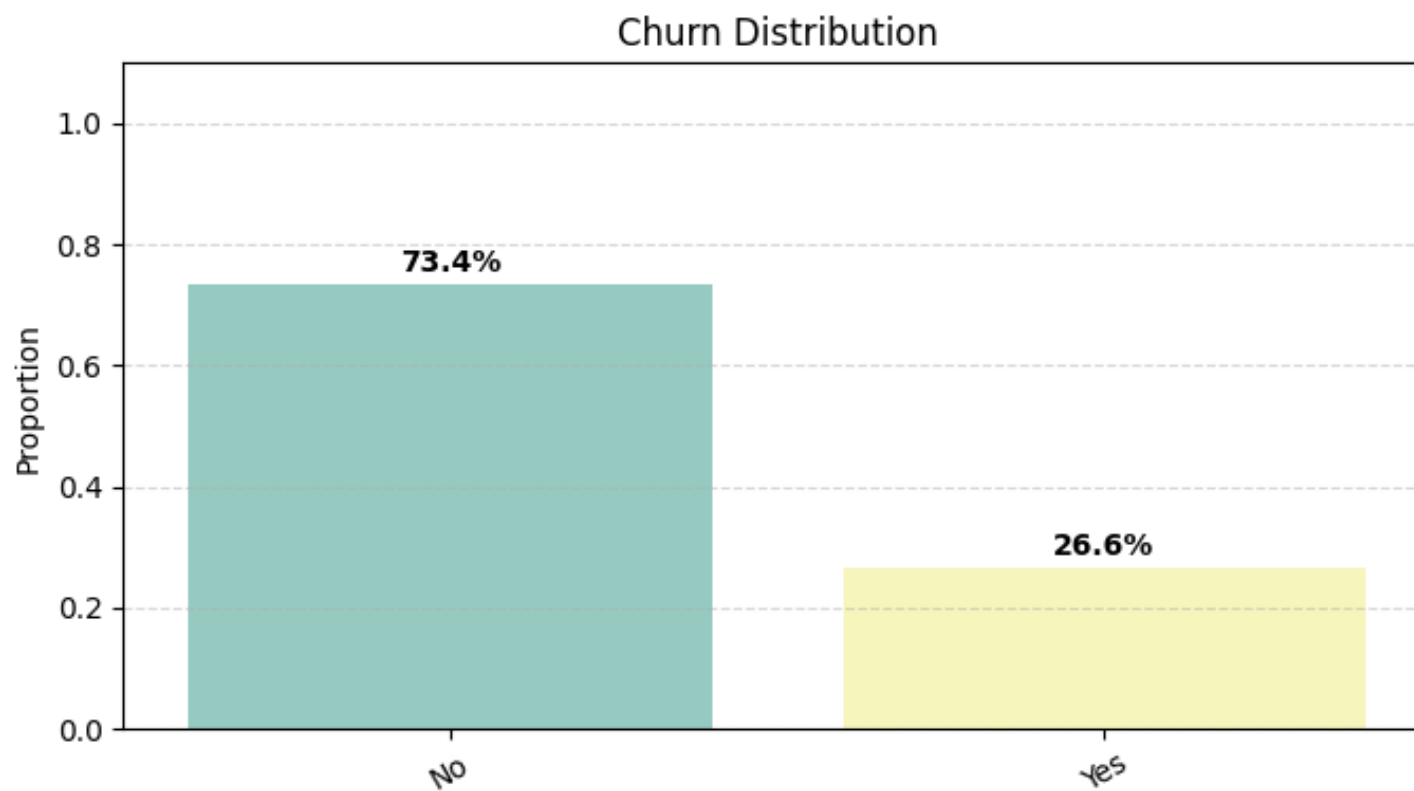
### 3. 기술통계요약

#### 1) 범주형 기술 통계 (주요 변수 시각화 결과)

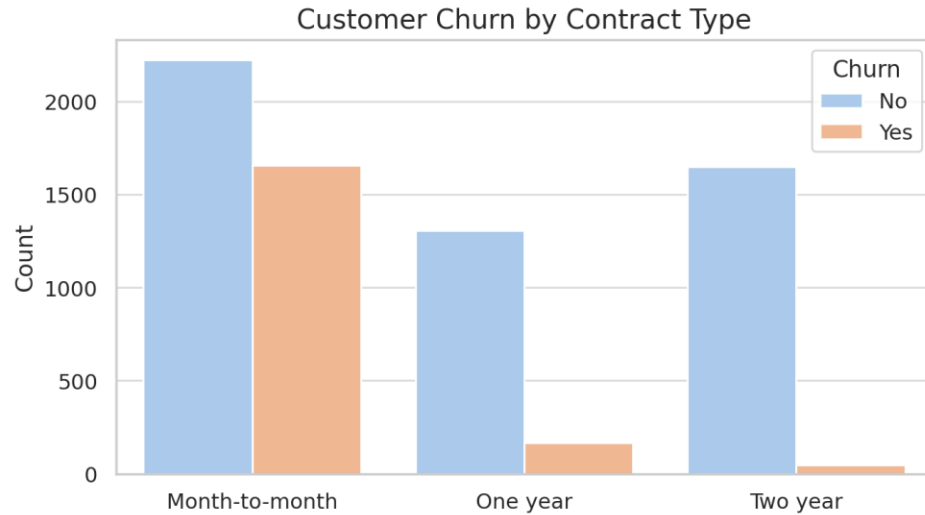


### 3. 기술통계요약

#### 1) 범주형 기술 통계 (주요 변수 시각화 결과)

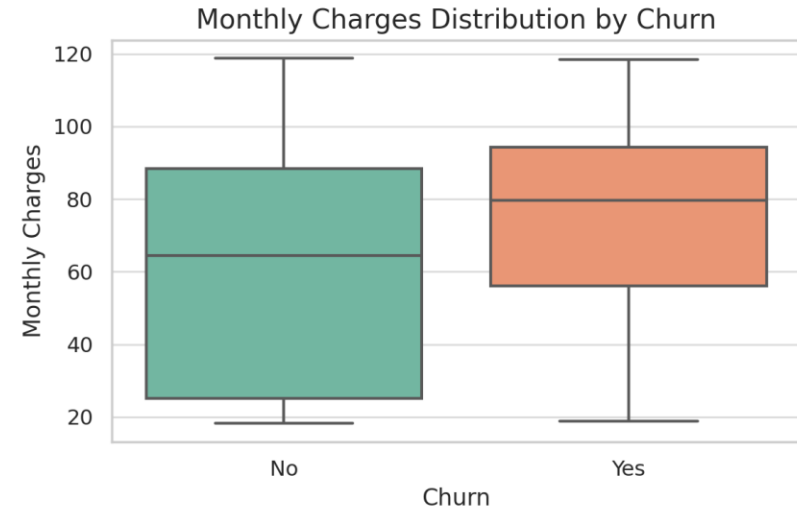


## 4. 시각화 및 주요 인사이트



### 1) 계약유형에 따른 이탈 분석

고객의 계약 유형에 따라 이탈 여부를 시각화한 결과, Month-to-month(월 단위 계약)을 이용하는 고객군에서 이탈 비율이 가장 높게 나타났습니다. 반면, one year, two year와 같은 장기 계약 고객군은 이탈 비율이 현저히 낮았습니다. 이는 장기 계약이 고객 충성도에 긍정적인 영향을 미칠 수 있음을 나타냅니다. 따라서 통신사는 단기 계약 고객을 대상으로 장기계약으로의 전환을 유도하는 전략(예: 장기 할인, 추가 혜택 제공 등)을 검토할 필요가 있습니다.

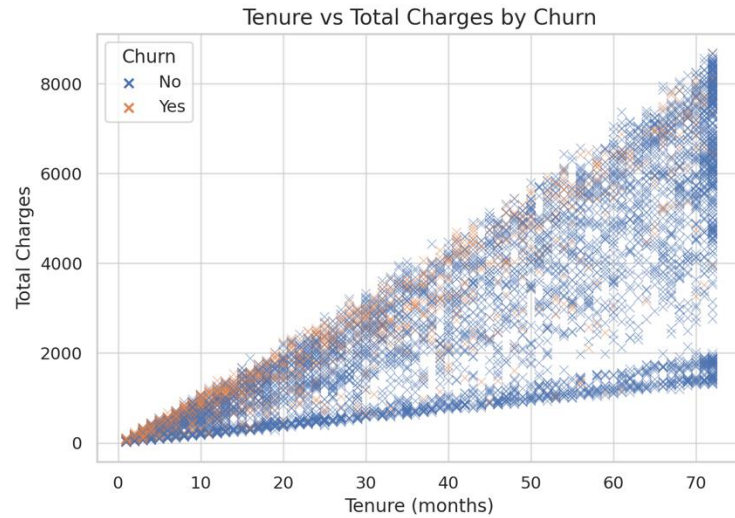


### 2) 월 요금에 따른 이탈 분석

월 요금에 대한 박스플롯을 통해, 이탈 고객들의 월 납부 요금이 평균적으로 더 높은 경향을 보였습니다. 특히 상위 25%에 해당하는 고요금 고객군에서 이탈률이 두드러졌습니다. 이는 고요금 고객이 가격에 대한 부담을 느끼거나, 요금 대비 서비스 품질에 불만을 가질 가능성이 있음을 나타냅니다. 따라서 고요금 사용자군을 대상으로 요금제 구성의 투명성 확보, 가격 대비 혜택 설명 강화, 맞춤형 할인 정책 등을 마련하는 것이 이탈률 감소에 기여할 수 있습니다.

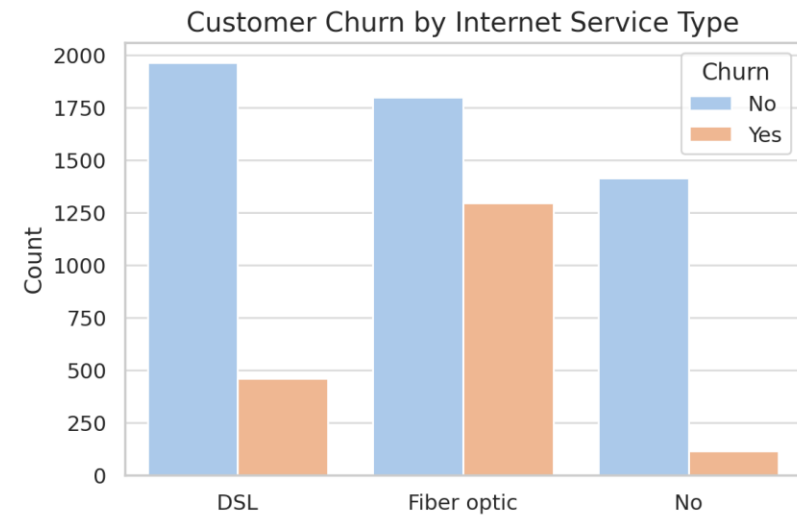


## 4. 시각화 및 주요 인사이트



### 3) 가입 기간과 누적 요금에 따른 이탈 분석

가입 기간과 누적 지불 요금 간의 관계를 이탈 여부에 따라 시각화한 결과, 가입 기간이 짧고 누적 요금이 낮은 고객군에서 이탈 비율이 뚜렷하게 높게 나타났습니다. 이는 고객들이 가입 초기(특히 3~6개월 내)에 서비스에 만족하지 못하고 이탈하는 경우가 많다는 점을 보여줍니다. 따라서 초기 고객 Onboarding 및 유지 프로그램(예: 초기 3개월간 요금 할인, 정기적인 만족도 조사 및 대응 등)을 통해 초기 이탈을 줄이는 노력이 필요합니다.



### 4) 인터넷 서비스 유형에 따른 이탈 분석

인터넷 서비스 유형별 고객 이탈률을 살펴본 결과, 'Fiber optic' 서비스를 사용하는 고객군에서 이탈률이 가장 높게 나타났습니다. 이는 고속 인터넷을 제공하지만 상대적으로 요금이 높거나 품질에 대한 불만이 발생할 가능성을 가지고 있습니다. 반면, DSL 또는 인터넷을 사용하지 않는 고객군에서는 이탈률이 상대적으로 낮았습니다. 이는 통신사가 'Fiber optic' 사용자들의 고객만족도를 집중적으로 수집, 분석하고, 불만족 요인을 사전에 제거하거나 고객 맞춤형 보완책을 제시해야 할 필요성을 보여줍니다.

## 6. 결론 및 향후 분석 방향

### 1) 결론

이번 탐색적 데이터 분석을 통해 고객 이탈과 밀접한 연관이 있는 주요 특성을 도출할 수 있었습니다.

특히, 단기 계약(Month-to-month) 고객, 월 요금이 높은 고객, 가입 초기(tenure 짧음) 고객, 그리고 Fiber optic 서비스를 사용하는 고객이 이탈할 가능성이 높은 것으로 나타났습니다.

또한, 시니어 고객과 전자결제 방식 고객 등 일부 특성군에서도 이탈률이 다소 높게 나타났으며,

이는 고객군별 맞춤 전략의 필요성을 보여줍니다. 이러한 결과는 단순히 이탈을 예측하는 데 그치지 않고, 실제 마케팅 전략 및 고객 관리 정책 수립에 활용될 수 있는 실질적인 인사이트를 제공합니다.

예를 들어, 장기 계약 전환 유도, 고요금 고객 케어, 초기 3개월 집중관리 등은 이탈률을 줄이는 데 효과적인 방향이 될 수 있습니다.

### 2) 향후 분석 방향

#### (1) 예측 모델 구축 (지도학습)

로지스틱 회귀, 의사결정나무, 랜덤포레스트, XGBoost 등의 분류 모델을 적용하여 이탈 예측 정확도를 비교

#### (2) 변수 중요도 분석

모델 기반 feature importance 분석을 통해 이탈에 가장 영향을 많이 미치는 요인 정량화

#### (3) 고객 세분화 및 타겟 마케팅

군집화(K-Means 등) 또는 프로파일링을 통해 고객군 세분화, 각 군집별 맞춤 이탈 방지 전략 제시

#### (4) 이탈 시점 분석

고객의 서비스 사용 패턴을 시간 흐름에 따라 분석하여 이탈 조짐 조기 탐지

#### (5) 비즈니스 효과 정량화

이탈 예측 모델을 기반으로 고객 유지 시 수익 증대 효과 추정