



AI 01 김세진

Telco Customer Churn Project

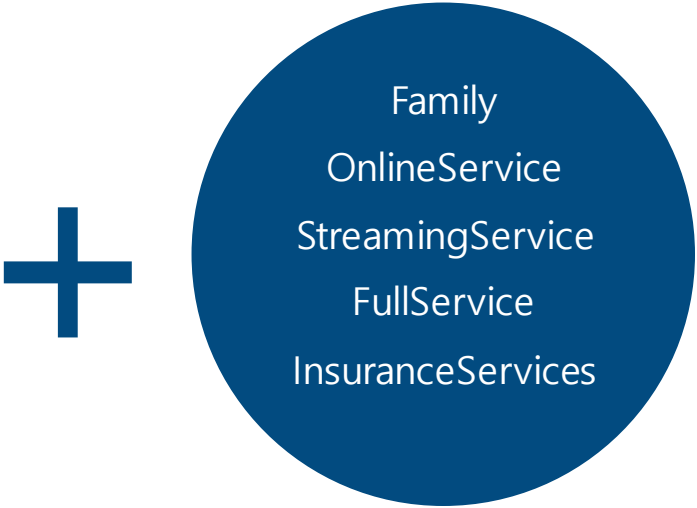
“

나는 오랜 취준 끝에 국내 3대 통신사 중 한 곳에 신입사원으로 입사했다.
입사한지 이제 1주일이 지난 나에게, 팀장님이 중요한 업무를 주셨다.
우리 회사가 내년 하반기 미국 진출을 목표로 고객 유치 전략을 짜고 있는데,
회사에서 입수한 미국 시장 데이터를 통해 **중요한 고객 특성**을 찾아보라는 것!!
팀장님께서 주신 데이터는 **미국 통신사의 고객 이탈 여부** 데이터였다.
심지어 팀장님은 나에게 팀 회의에서 분석 결과를 발표하라고 하셨다 π.π
그래서 머신러닝을 통해 **분류 모델**을 만들고 **어떤 특성이 고객 이탈 여부 예측에
가장 큰 영향을 미쳤는지 알아보는 프로젝트**를 진행하기로 했다. 화이팅!

”

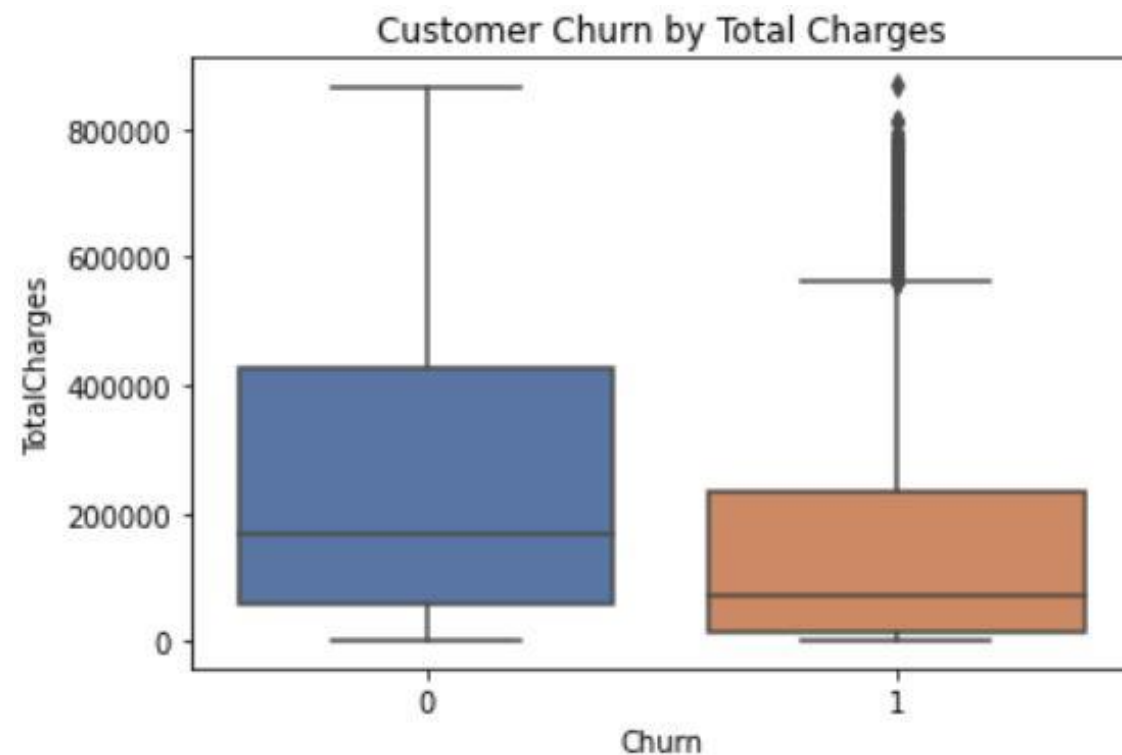
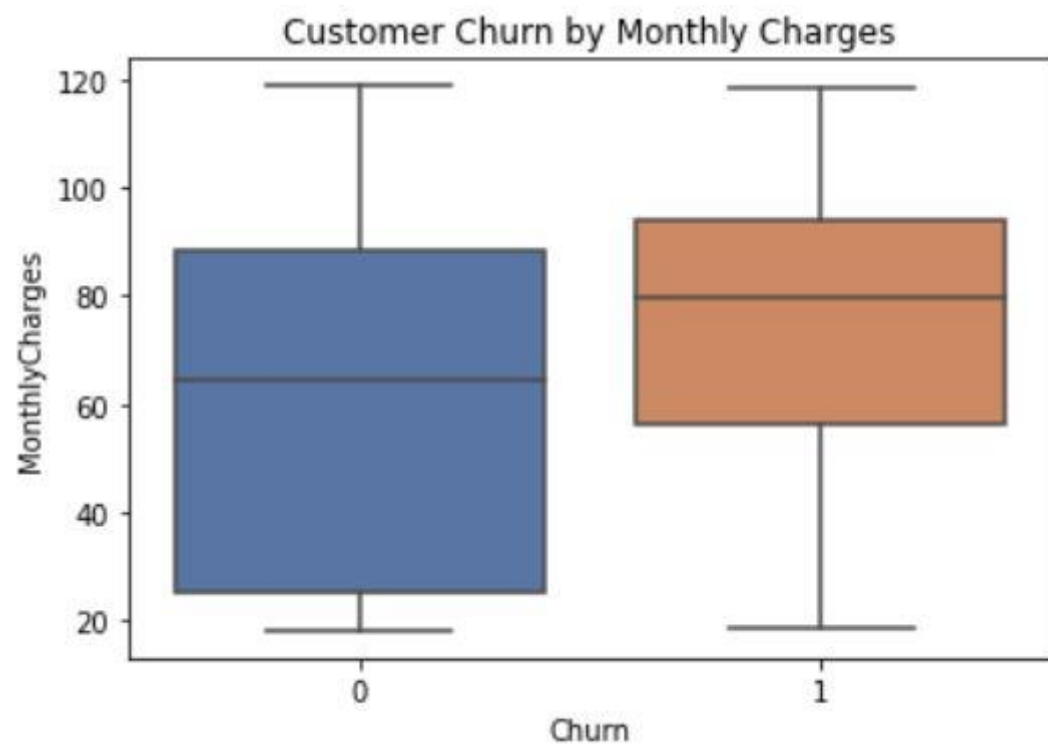
특성 설명

Gender	성별	DeviceProtection	기기보호 서비스 이용 여부
SeniorCitizen	고령 고객 여부	TechSupport	기술지원 서비스 이용 여부
Partner	배우자 유무	StreamingTV	TV채널 스트리밍 이용 여부
Dependents	부양가족 유무	StreamingMovies	영화 스트리밍 이용 여부
Tenure	직장 근속 개월수	Contract	계약 기간
PhoneService	전화 서비스 여부	PaperlessBilling	종이 고지서 신청 여부
MultipleLines	다중 회선 여부	PaymentMethod	요금 지불 방법
InternetService	인터넷 공급자	MonthlyCharges	월별 요금
OnlineSecurity	온라인 보안 서비스 이용 여부	TotalCharges	총 요금
OnlineBackup	온라인 백업 서비스 이용 여부	Churn 	통신사 이탈 여부

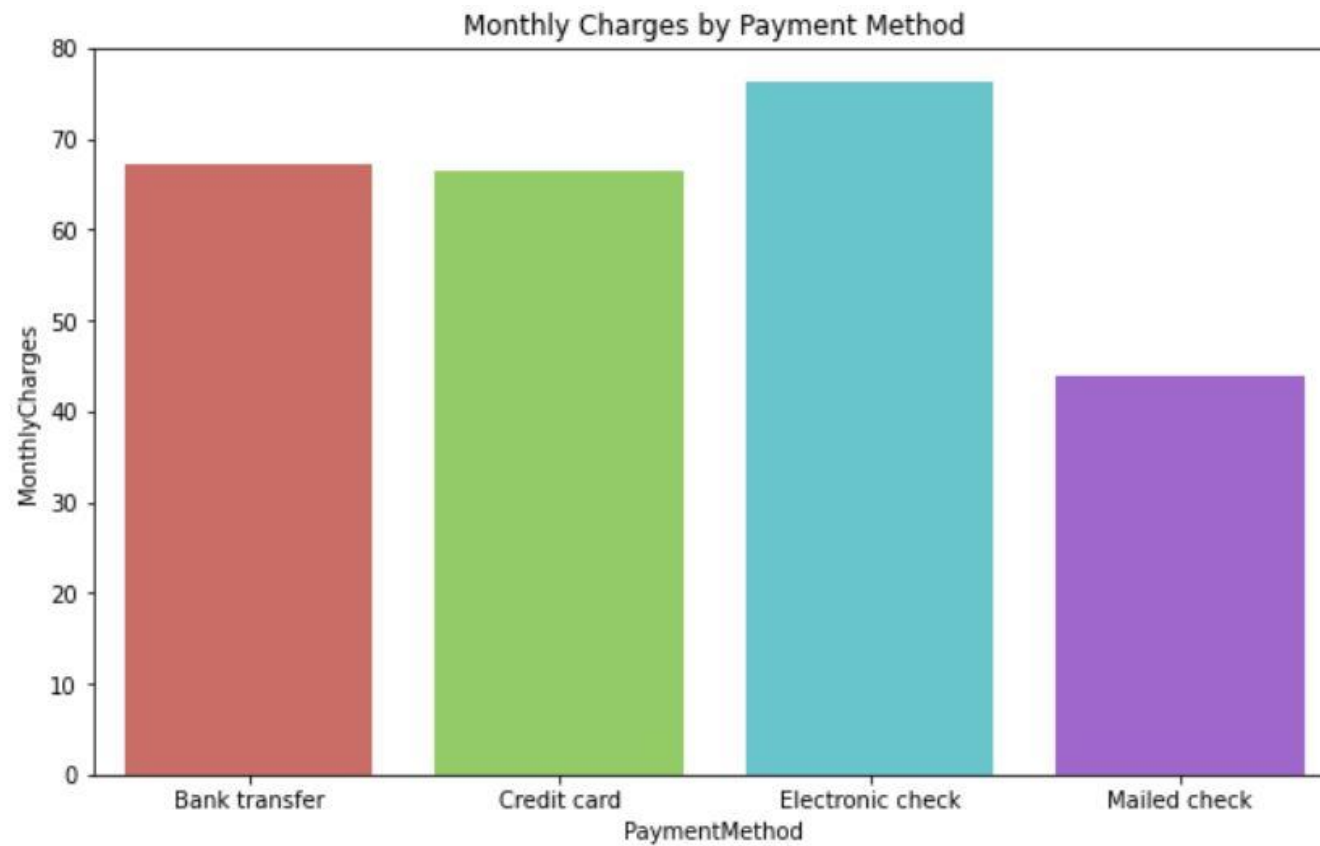


EDA

요금에 따른 이탈 여부 : 월별 요금이 높은 고객들의 이탈 비율이 높고, 총 요금이 낮은 고객들의 이탈 비율이 높다.

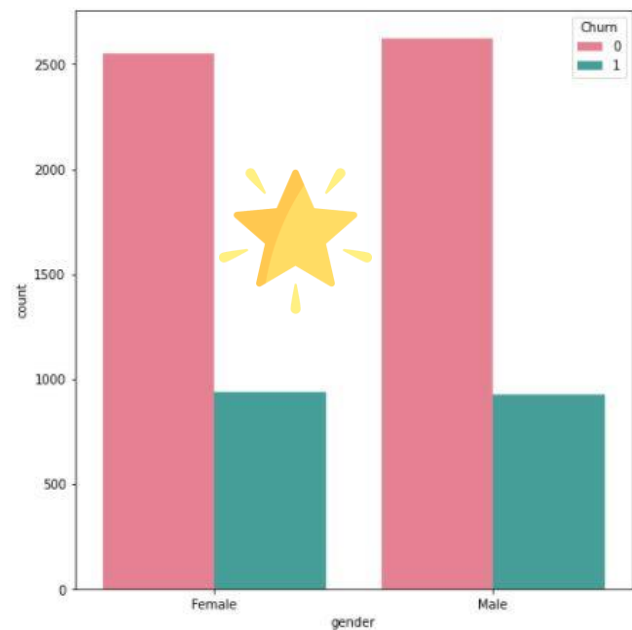


월별 요금이 가장 높은 지불 방법

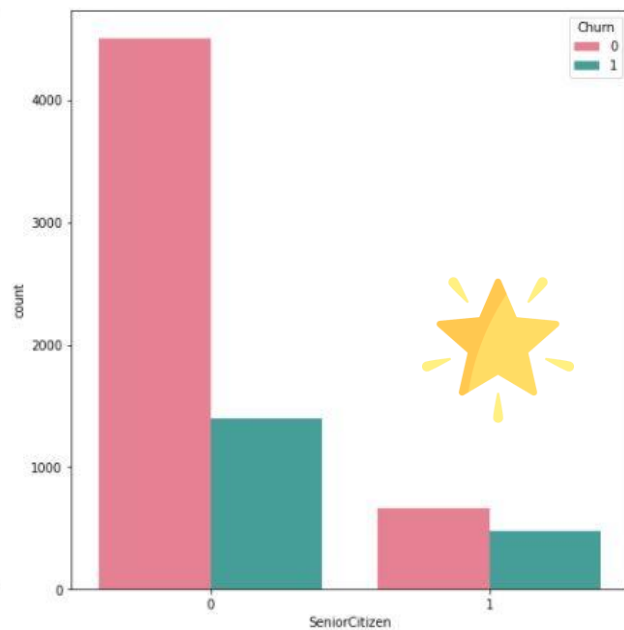


EDA

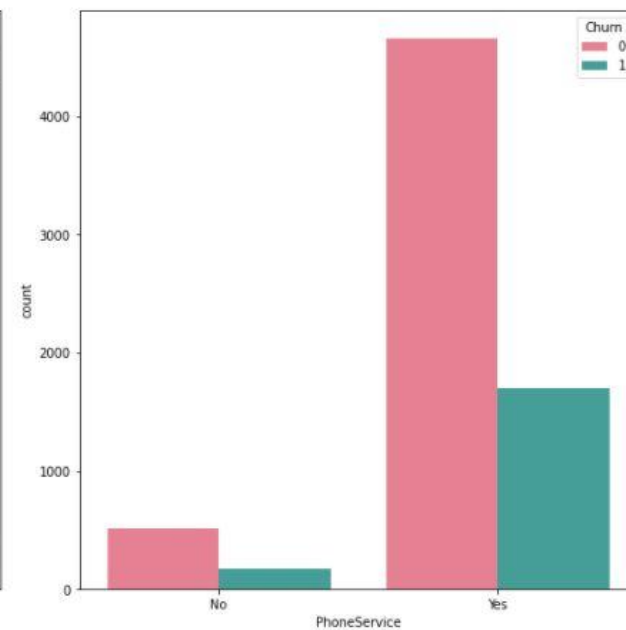
각 특성별 고객 이탈 비율 : 성별과는 관계 X, 고객의 연령대가 높을수록 이탈률이 낮아진다.



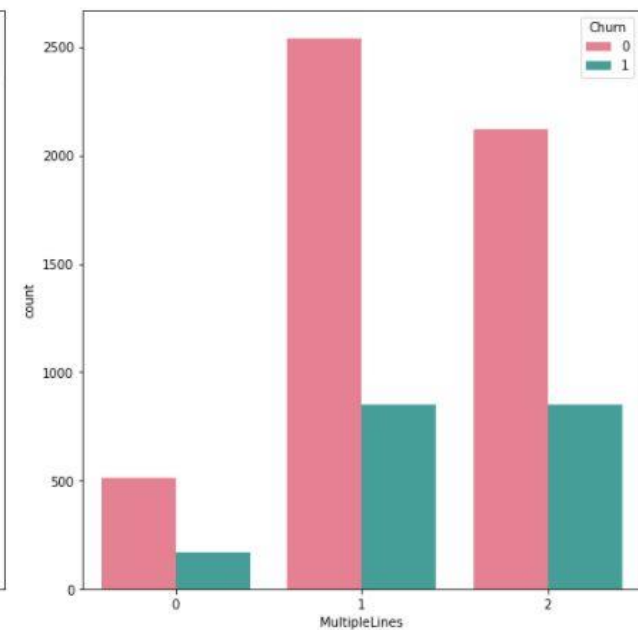
성별



고령 고객



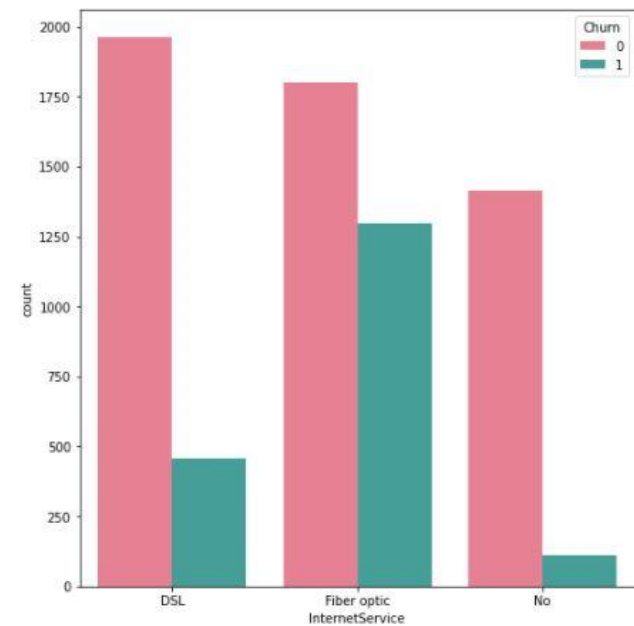
전화 서비스



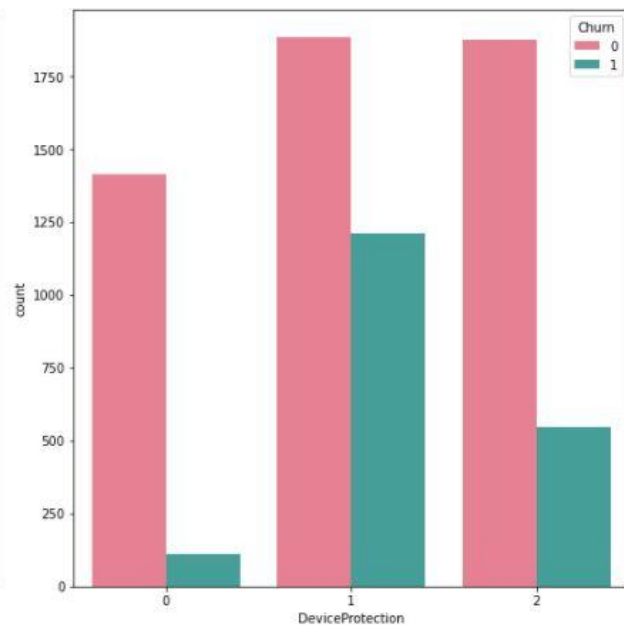
다중 회선 사용

EDA

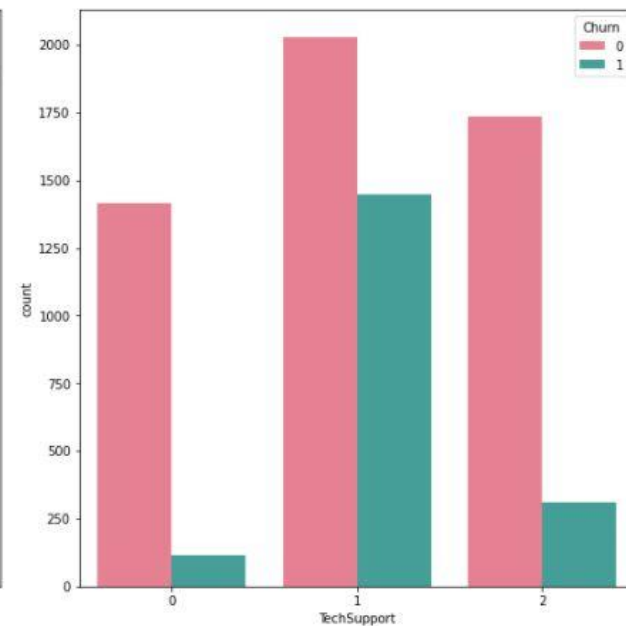
각 특성별 고객 이탈 비율 : 계약 기간이 길어질 수록 이탈률이 낮아진다.



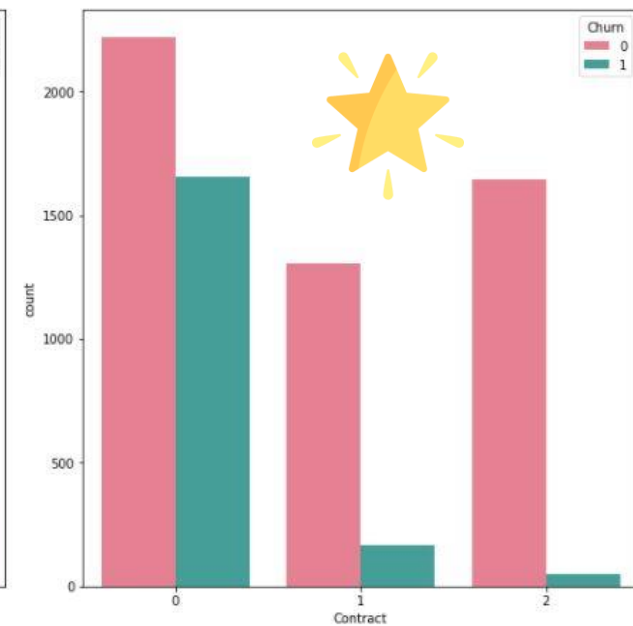
인터넷 공급자



기기 보호 서비스

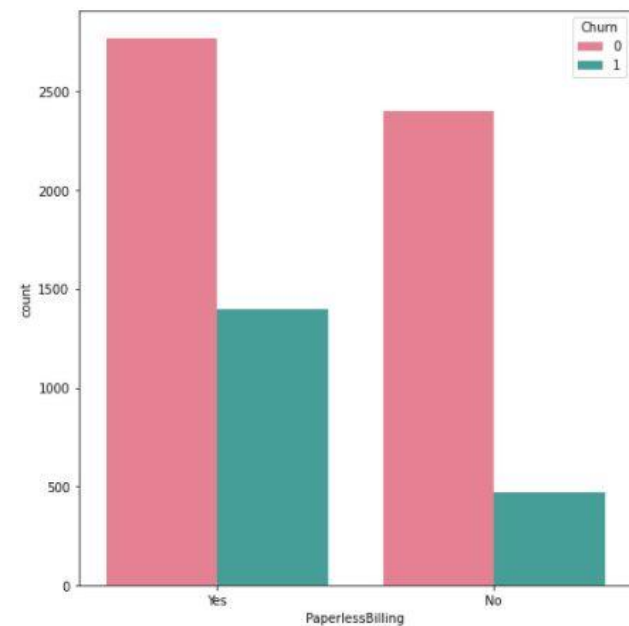


기술 지원 서비스

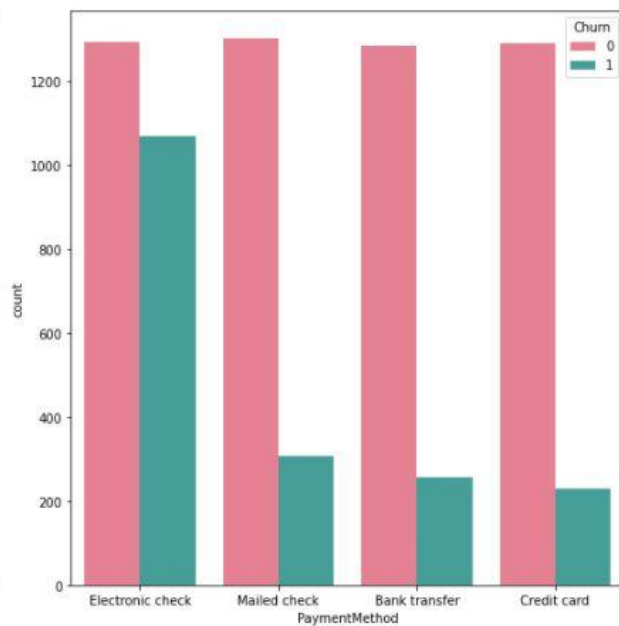


계약 기간

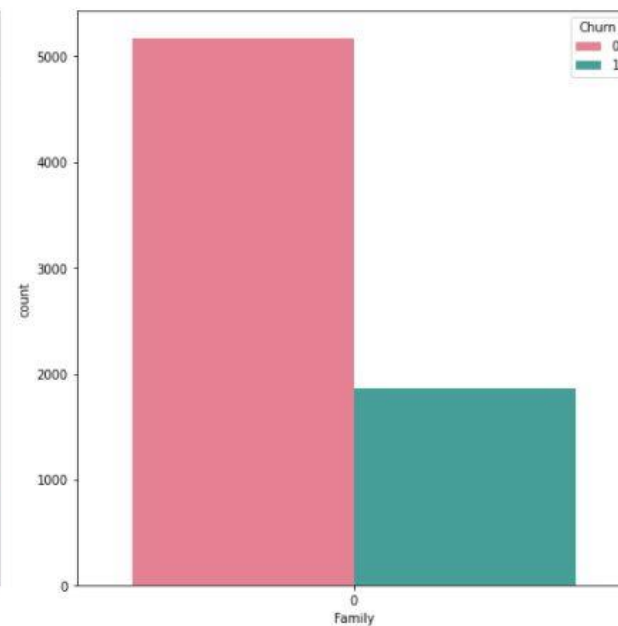
각 특성별 고객 이탈 비율



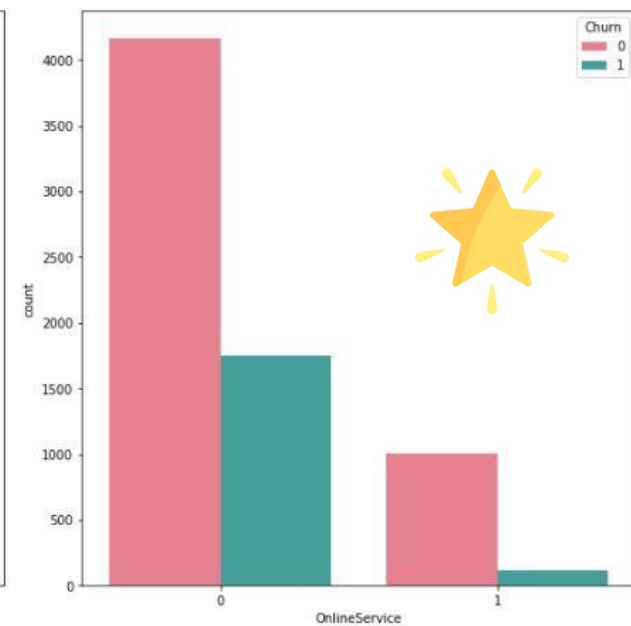
종이 고지서 신청 여부



요금 지불 방법



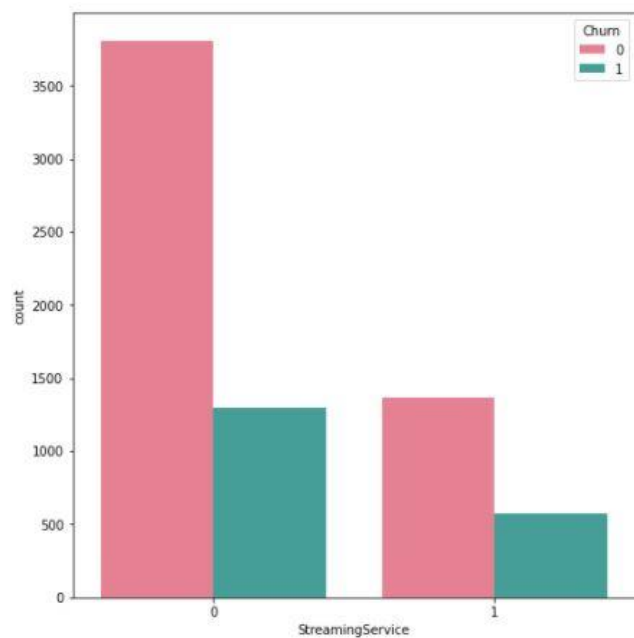
가족



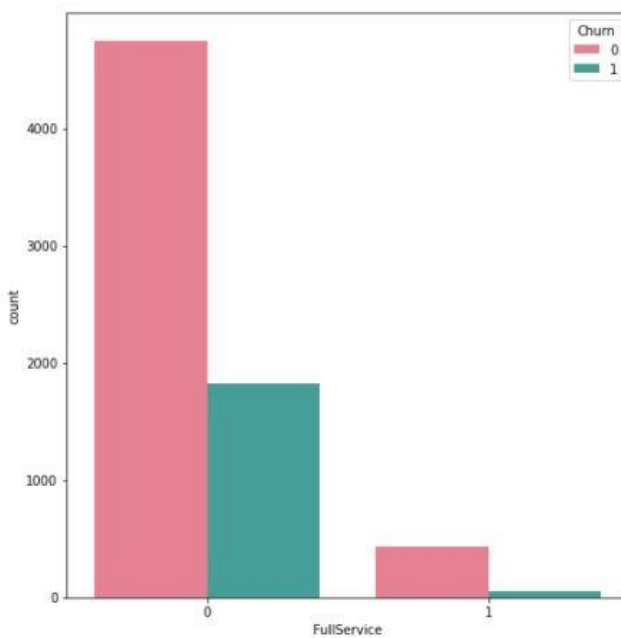
온라인 서비스

EDA

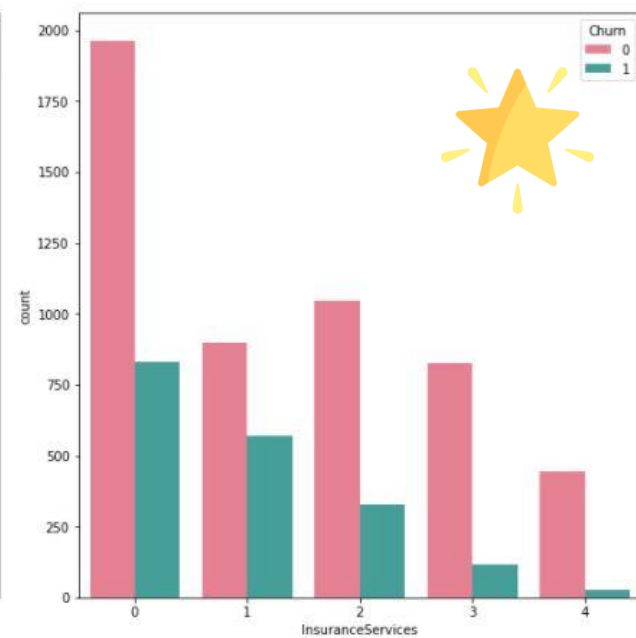
각 특성별 고객 이탈 비율 : 서비스를 많이 이용할수록 이탈률이 낮아진다.



스트리밍 서비스



모든 서비스 이용



보험 서비스

특성 선택

Weight	Feature
0.0165 ± 0.0114	tenure
0.0096 ± 0.0155	Contract
0.0085 ± 0.0060	MultipleLines
0.0062 ± 0.0135	PaymentMethod
0.0043 ± 0.0119	PaperlessBilling
0.0026 ± 0.0085	OnlineService
0.0017 ± 0.0105	TechSupport
0.0014 ± 0.0090	TotalCharges
0.0006 ± 0.0079	gender
0 ± 0.0000	Family
-0.0003 ± 0.0113	DeviceProtection
-0.0006 ± 0.0117	SeniorCitizen
-0.0009 ± 0.0034	FullService
-0.0011 ± 0.0011	PhoneService
-0.0020 ± 0.0038	StreamingService
-0.0057 ± 0.0084	InsuranceServices
-0.0085 ± 0.0173	InternetService
-0.0128 ± 0.0145	MonthlyCharges

eli5 Permutation Importance

선택한 특성

tenure, Contract, MultipleLines, PaymentMethod, PaperlessBilling, OnlineService, TechSupport, TotalCharges

하이퍼 파라미터 튜닝 & 모델링

Random Forest Model

파라미터 튜닝	Randomized Search CV n_estimators : 810 max_depth : 8 max_features : 3 min_samples_split : 5
파이프라인	Ordinal Encoder Simple Imputer Random Forest Classifier

Running time : 3.0 min
Accuracy : 0.8212
Recall : 0.4451
F1 Score : 0.5531

XGBoost Model



파라미터 튜닝	Randomized Search CV n_estimators : 910 max_depth : 8 max_features : 3 min_samples_split : 7 num_iterations : 800 reg_alpha : 1 reg_lambda : 10
파이프라인	Ordinal Encoder Simple Imputer XGB Classifier

Running time : 2.1 min
Accuracy : 0.8141
Recall : 0.2743
F1 Score : 0.4072

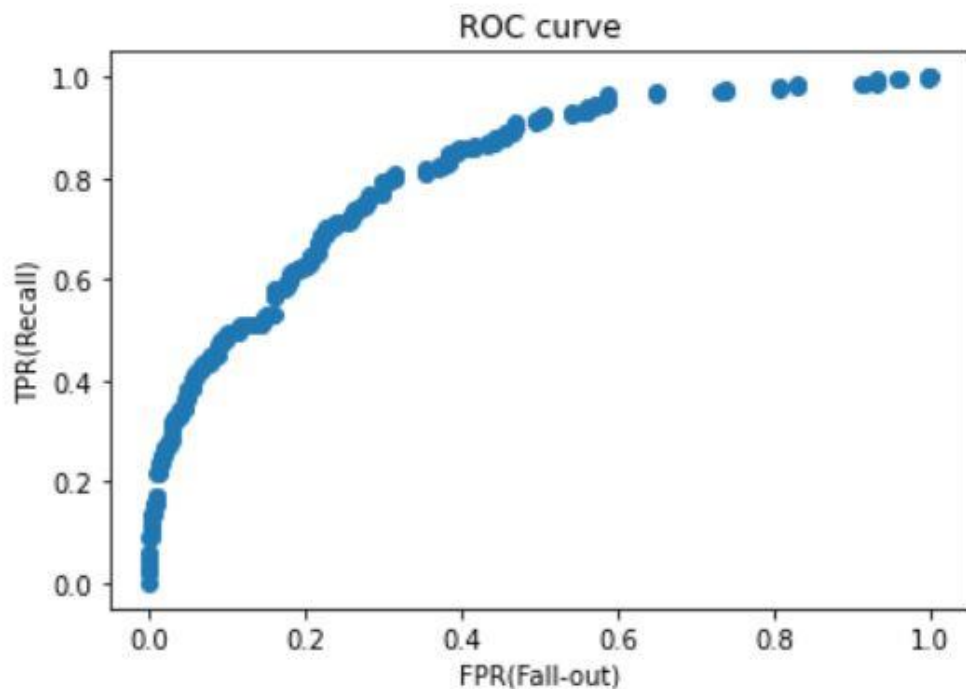
Light GBM Model

파라미터 튜닝	Randomized Search CV n_estimators : 510 max_depth : 6 max_features : 6 min_samples_split : 7 num_iterations : 800 reg_alpha : 0 reg_lambda : 10
파이프라인	Ordinal Encoder Simple Imputer LGBM Classifier

Running time : 44.8s
Accuracy : 0.8099
Recall : 0.2682
F1 Score : 0.3963

ROC Curve, AUC Score

AUC Score : 0.8171746088995085

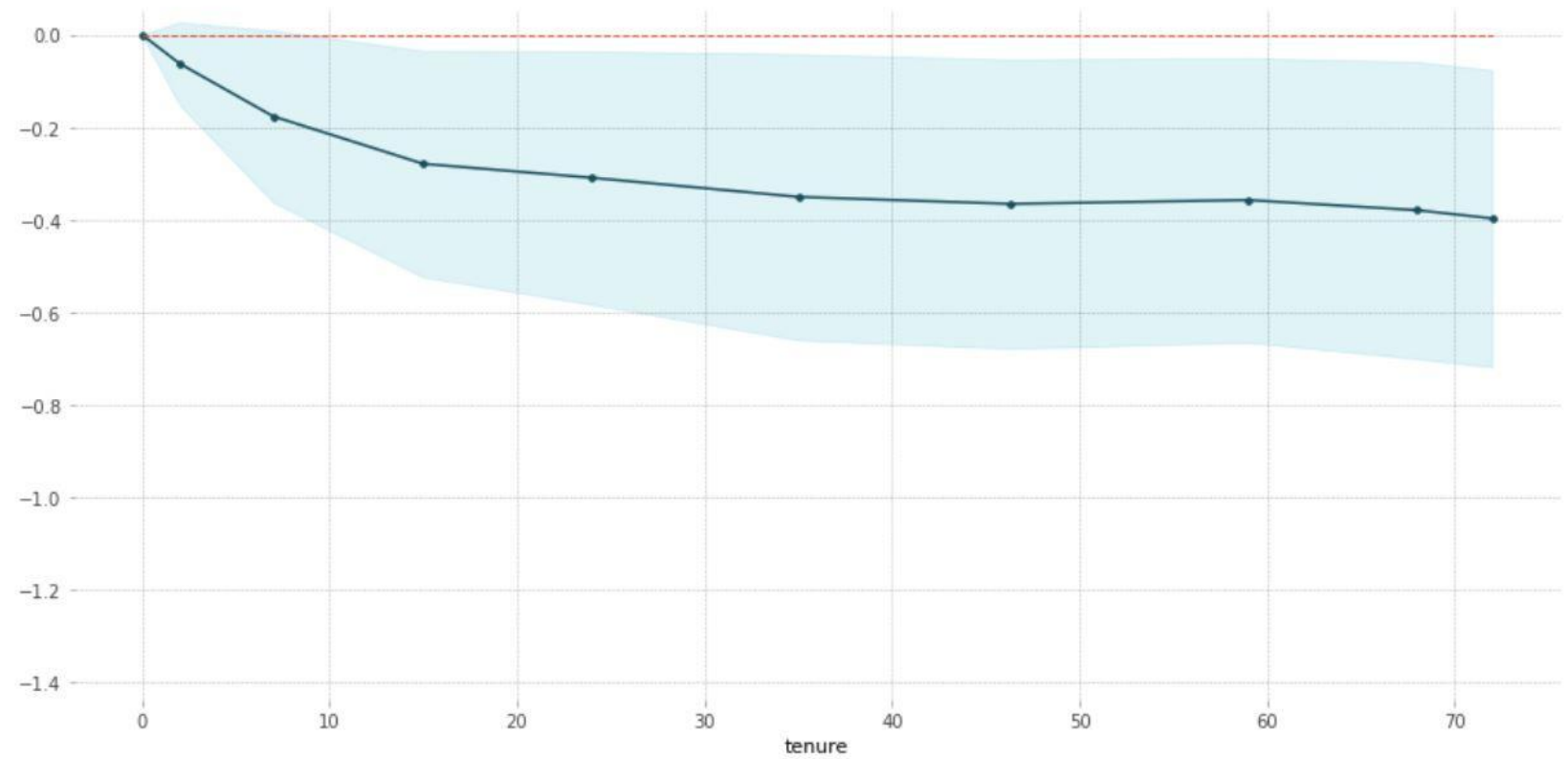


X축은 이탈하지 않은 고객을 이탈했다고 예측한 비율, Y축은 이탈한 고객을 이탈했다고 예측한 비율이므로 곡선이 굽어질수록 고객의 이탈률을 정확히 예측했다는 것을 알 수 있다. 또한 AUC Score가 1에 가까울수록 정확한 모델이다. 이 모델은 AUC Score가 0.8 이상이므로 비교적 정확한 모델이라고 볼 수 있다.

PDP

Tenure 특성

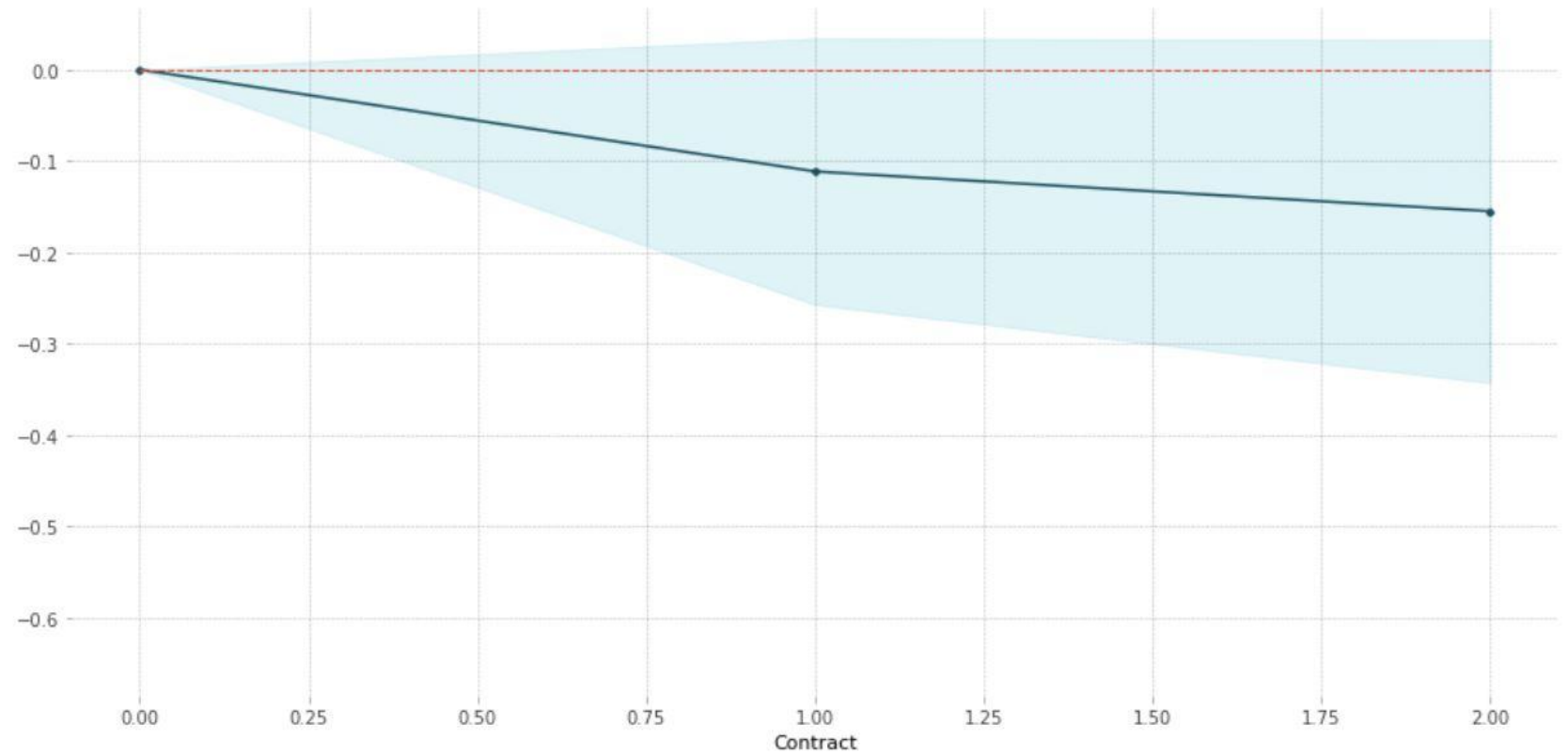
PDP for feature "tenure"
Number of unique grid points: 10



Contract 특성

PDP for feature "Contract"

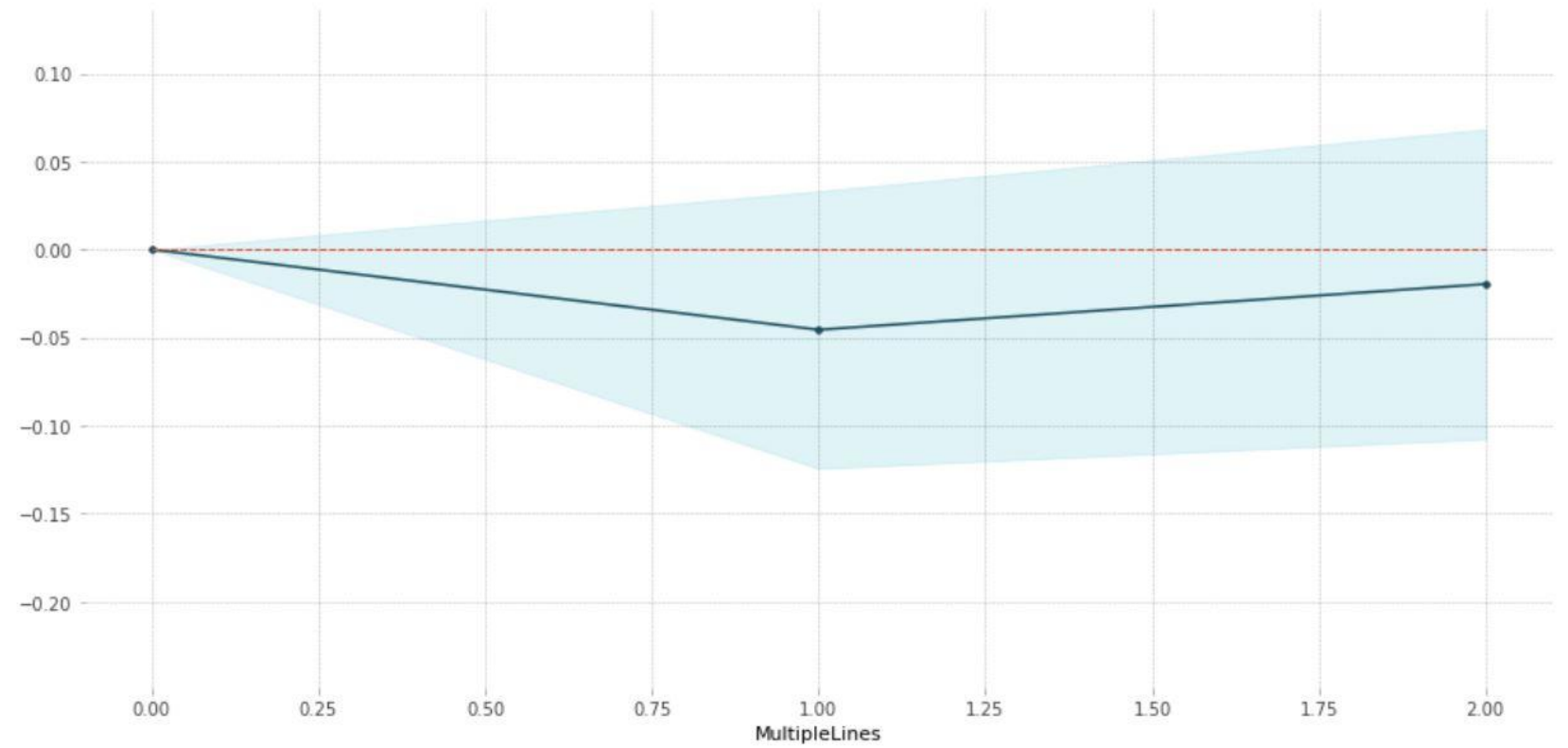
Number of unique grid points: 3



MultipleLines 특성

PDP for feature "MultipleLines"

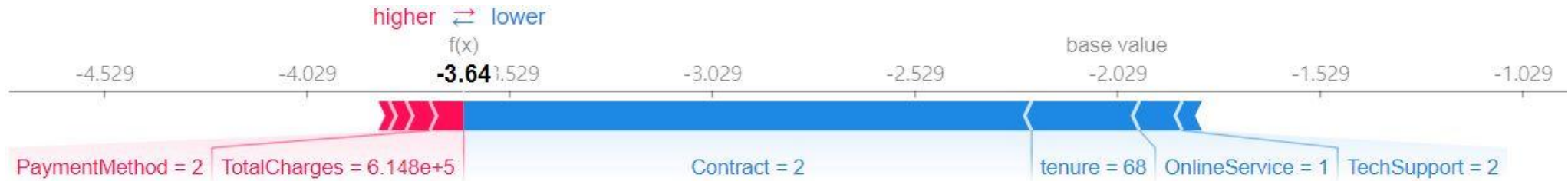
Number of unique grid points: 3



SHAP

Case 1

	tenure	MultipleLines	TechSupport	Contract	PaperlessBilling	PaymentMethod	TotalCharges	OnlineService
	337	68	2	2	2	2	614845.0	1

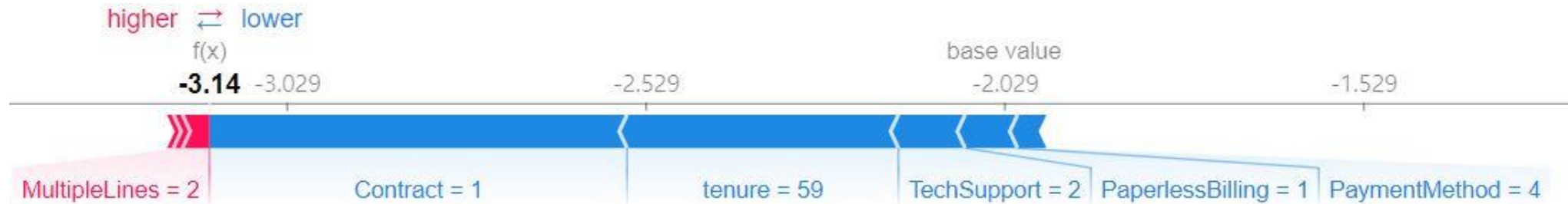


직장 근속 68개월(5년 8개월), 다중 회선 사용, 기술 지원 서비스 사용, 계약 기간 2년 이상, 전자 고지서 신청, 은행 자동이체로 요금 지불, 총 요금 61만 달러, 온라인 서비스 모두 사용
→ 지불 방법과 총 요금이 예측에 높은 영향을 미쳤다.

SHAP

Case 2

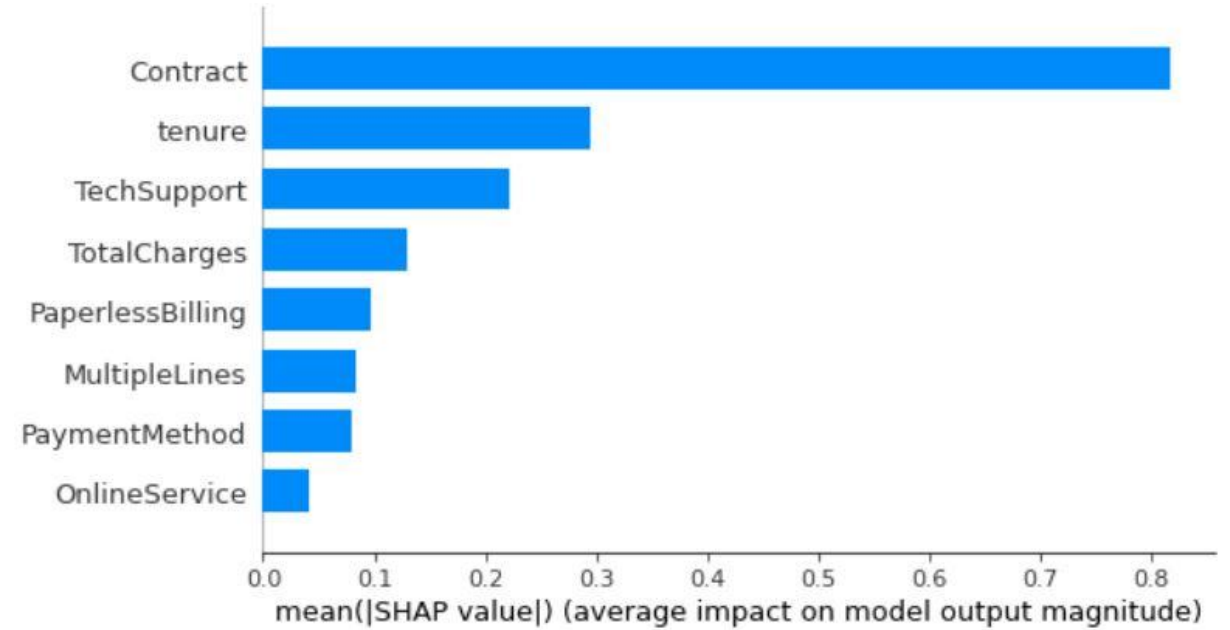
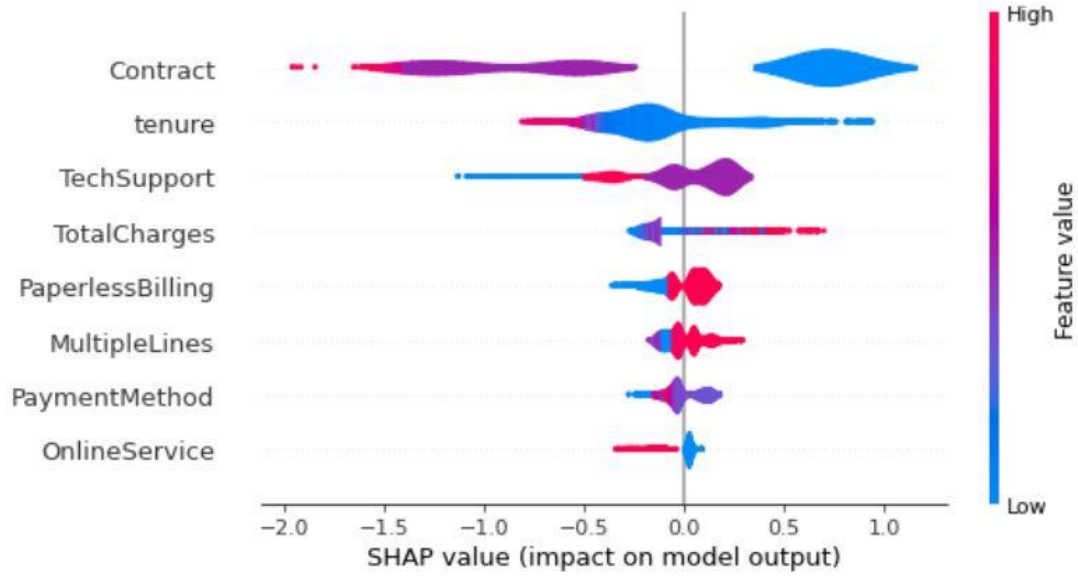
	tenure	MultipleLines	TechSupport	Contract	PaperlessBilling	PaymentMethod	TotalCharges	OnlineService
	5262	59	2	2	1	4	407095.0	0



직장 근속 59개월(4년 11개월), 다중 회선 사용, 기술 지원 서비스 사용, 계약 기간 1년 이상, 종이 고지서 신청, 이메일 결제로 요금 지불, 총 요금 40만 달러, 온라인 서비스 모두 사용하지 않음

→ 다중 회선을 사용하는 것이 예측에 높은 영향을 미쳤다.

SHAP



Contract 계약 기간, **tenure** 근속 개월수, **TechSupport** 기술지원 서비스 이용 여부

→ 이 특성들에 주목하면 고객 이탈률을 예측하는 데에 도움이 된다

가장 중요한 특성

: **Contract** 계약 기간 / **tenure** 근속 개월수 / **TechSupport** 기술지원 서비스 이용 여부

특성 활용 전략

1. 기존 고객을 위한 이벤트나 서비스 제공
2. 기존 고객 추천으로 가입 시 추천자와 가입자 모두에게 서비스 제공
3. 기술지원 서비스에 대한 홍보를 늘리고 기존 서비스를 보완



Thank You