
통신회사에서 고객의 이탈을 어떻게 줄일 수 있을까?

AI_05_김민주_Section2

목차

1. 데이터 선정이유 및 문제정의
2. 데이터를 이용한 가설 및 평가지표, 베이스라인 선택
3. EDA와 데이터 전처리
4. 머신러닝 방식 적용 및 교차검증
5. 머신러닝 모델 해석
6. 결론

1. 데이터 선정이유 및 문제정의

(1) 데이터 선정 이유 : 모든 기업의 핵심은 "고객"

모든 기업에서 가장 중요한 것은 회사의 이익을 위해 "고객을 모으는 것"입니다. 하지만 새로운 고객을 얻는 것은 많은 비용과 시간이 필요합니다. 그와 비교해서 "고객 이탈을 막고 현재 고객을 잘 유지하는 것"은 장기적으로 더 경제적이고 효과적인 전략입니다.

따라서 많은 기업들의 가장 큰 이슈인 고객 이탈에 대한 데이터를 가지고 분석을 하였습니다.

(2) 문제 정의 : "통신회사에서 고객의 이탈을 어떻게 줄일 수 있을까?"

(3) 문제 접근 : 분류

2. 데이터를 이용한 가설 및 평가지표, 베이스라인 선택

(1) 가설 설정

- 성별에 따라서 고객의 이탈하는 정도에 차이가 있을까?
- 사용기간에 따라서 고객이 이탈하는 정도에 차이가 있을까?
- TotalCharges'가 'tenure'과 'MonthlyCharges'와 관련이 있을까?

(2) 평가지표 : Recall, F1 score

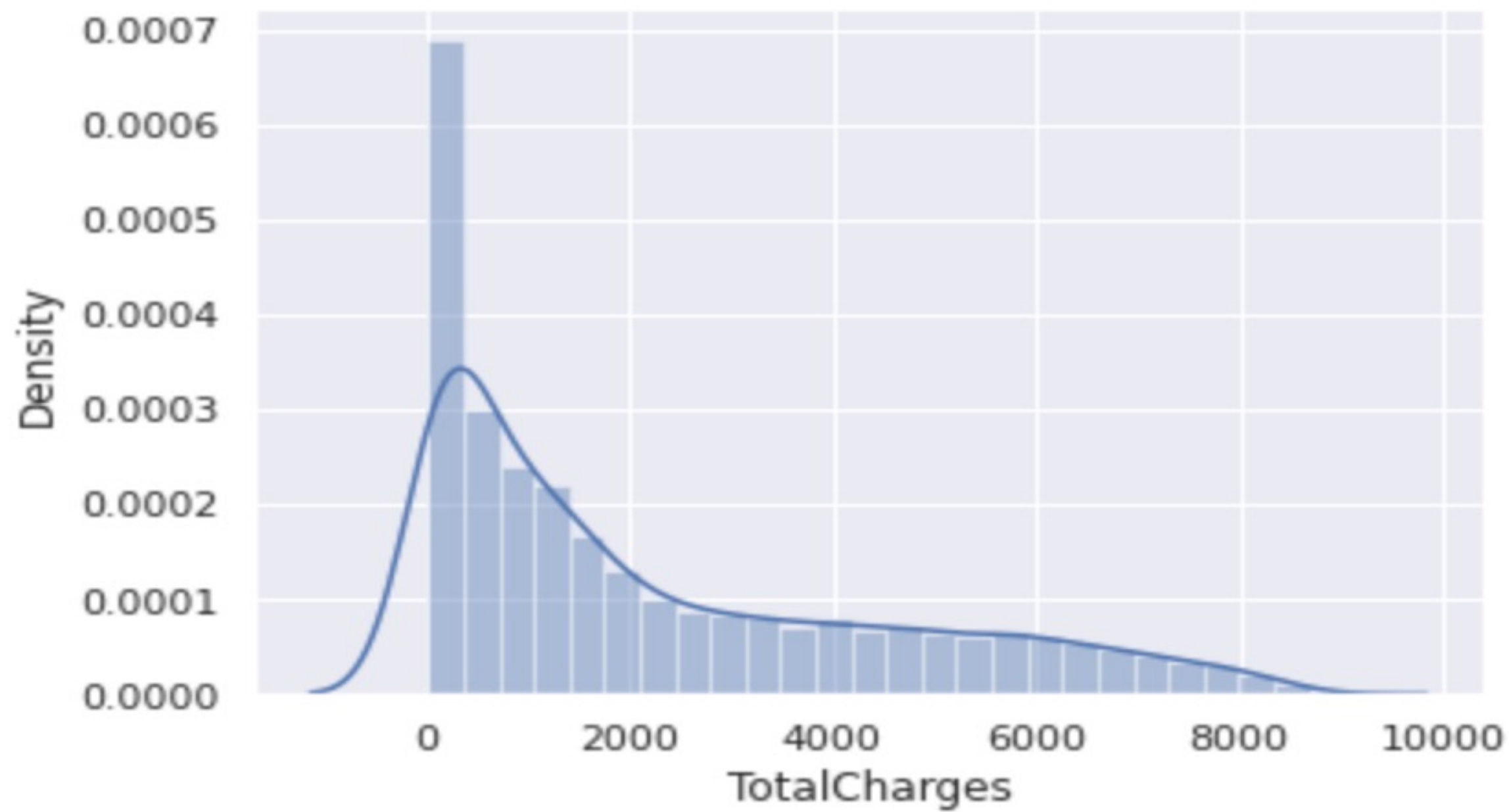
(3) 베이스라인 : 타겟의 최빈값

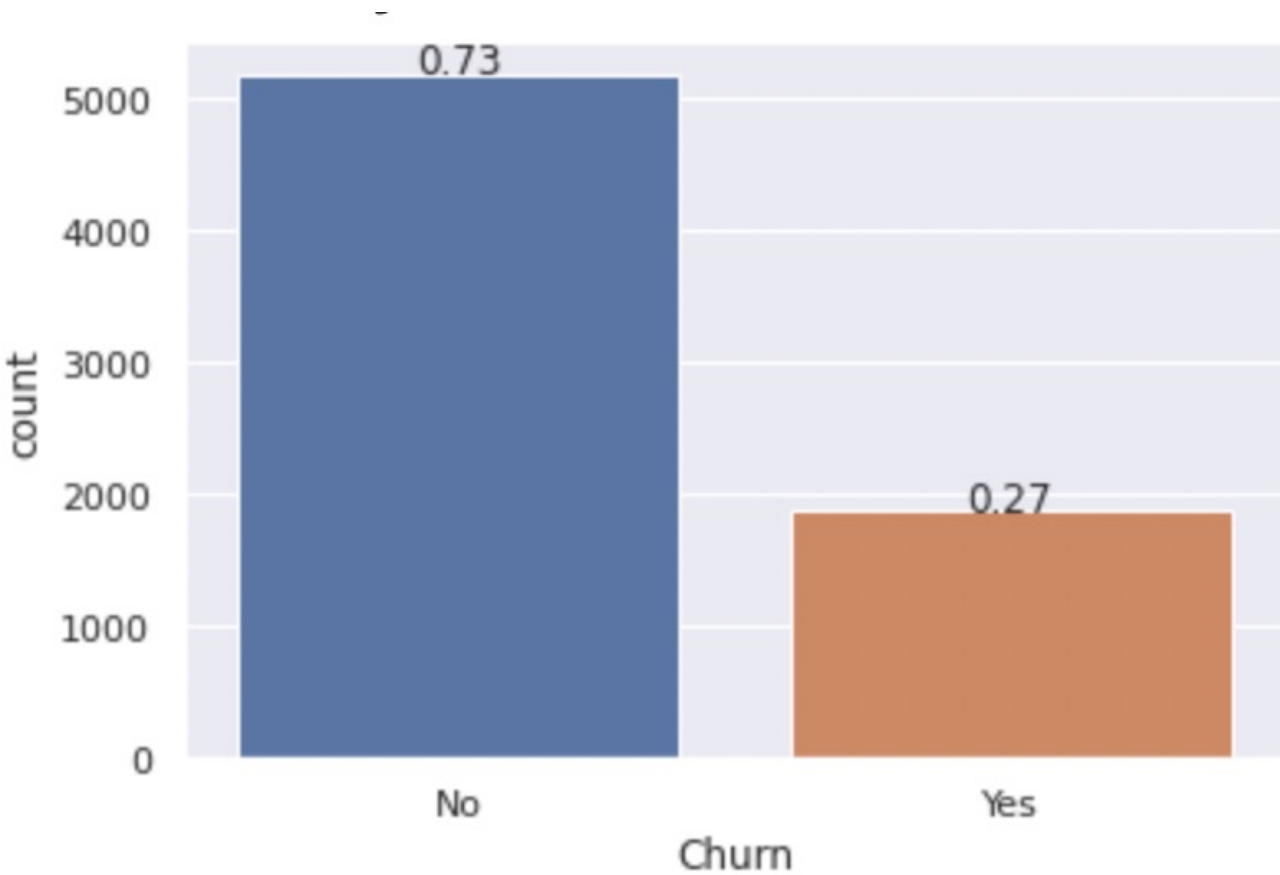
3. EDA와 데이터 전처리

(1) 데이터 분석

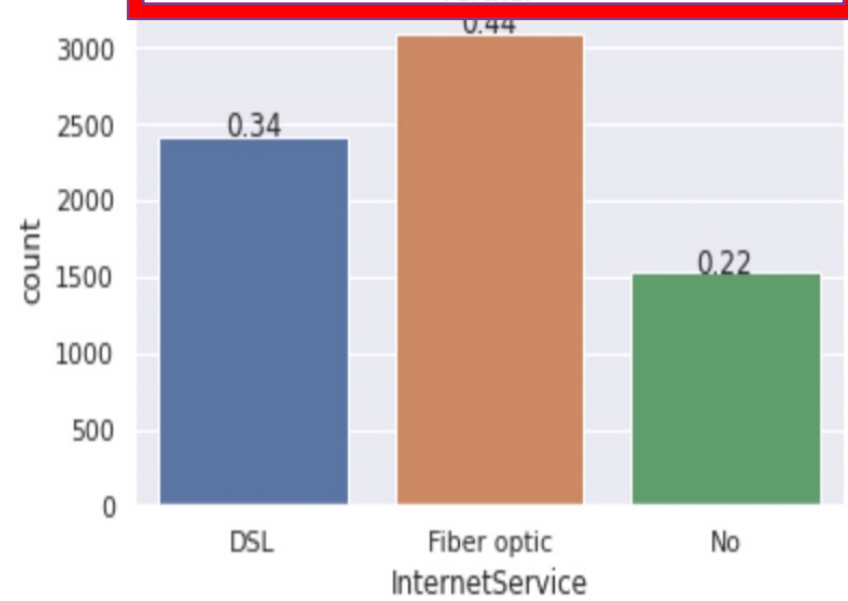
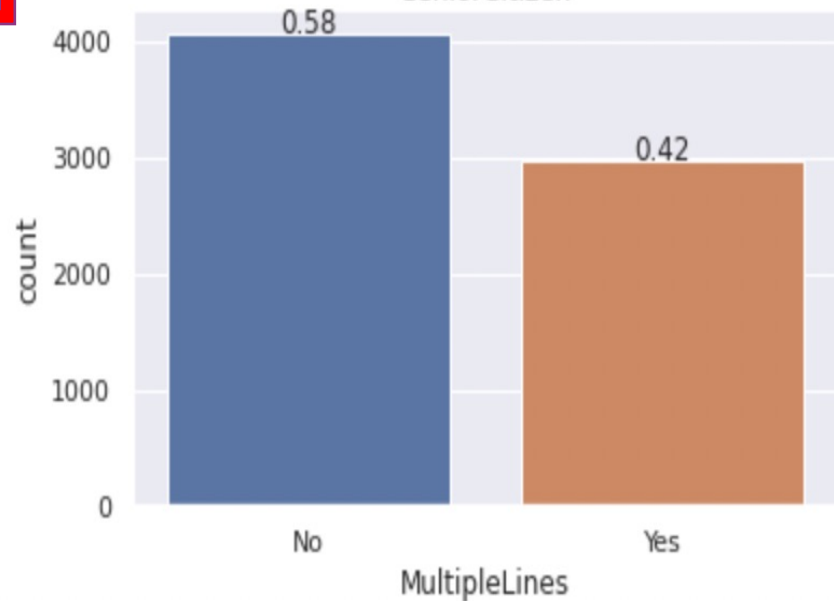
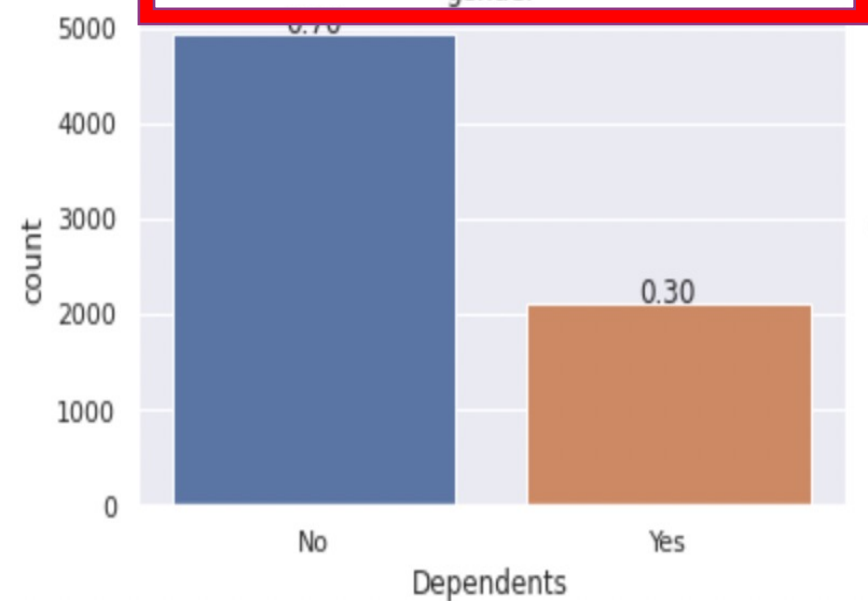
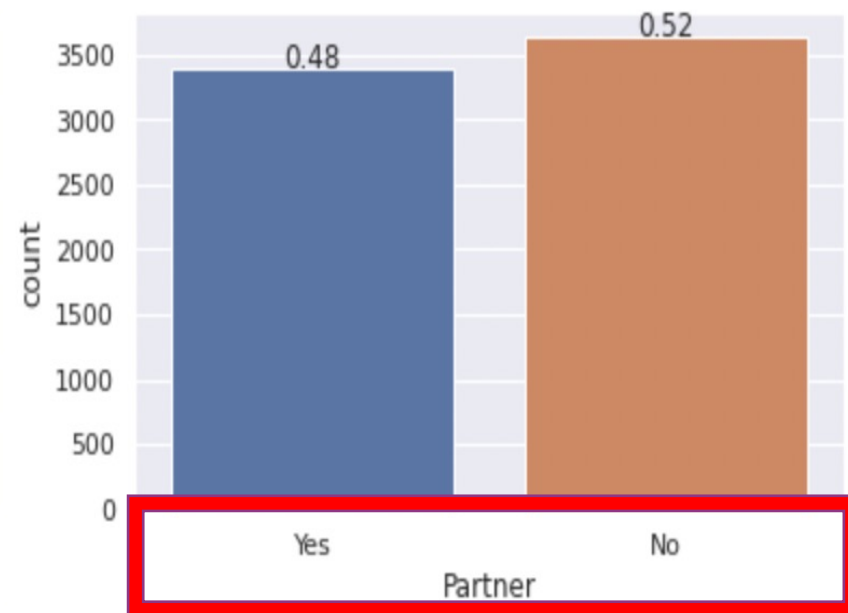
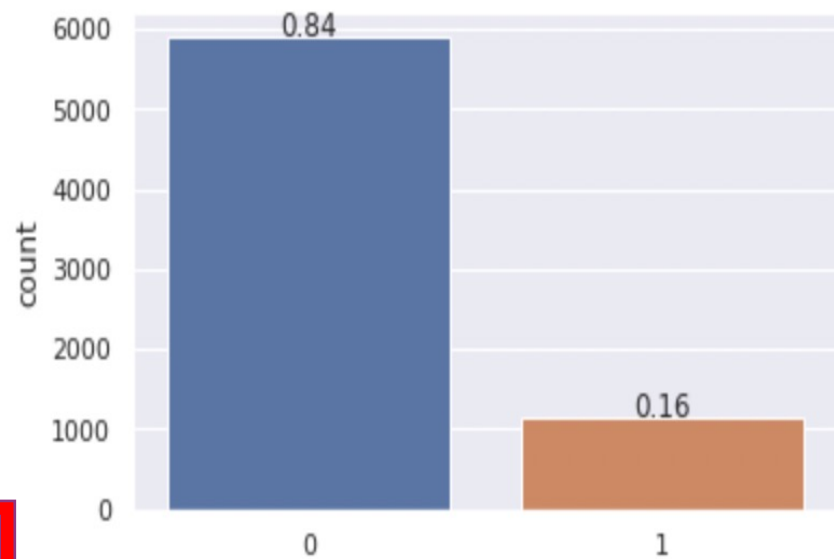
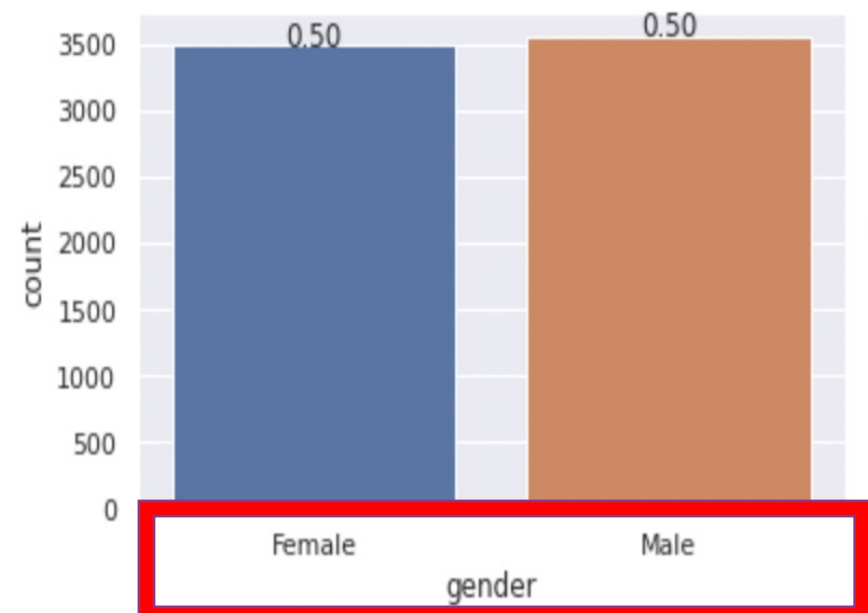
1. 성별, 결혼 여부를 포함한 고객에 대한 인구 통계 정보
2. 회사를 이용한 개월수, 페이퍼리스 과금, 결제수단, 월 이용요금, 총 이용요금 등 고객 계정 정보
3. TV 스트리밍, 영화 스트리밍 등 고객 이용 행태
4. 고객이 가입한 서비스: 전화 서비스, 다중 서비스, 인터넷 서비스, 온라인 보안, 온라인 백업, 장치 보호 및 기술 지원
5. 각 고객별로 이탈여부

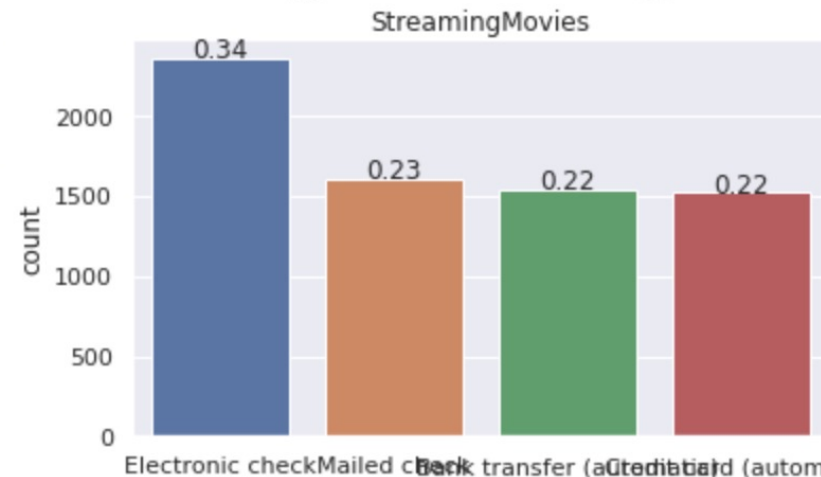
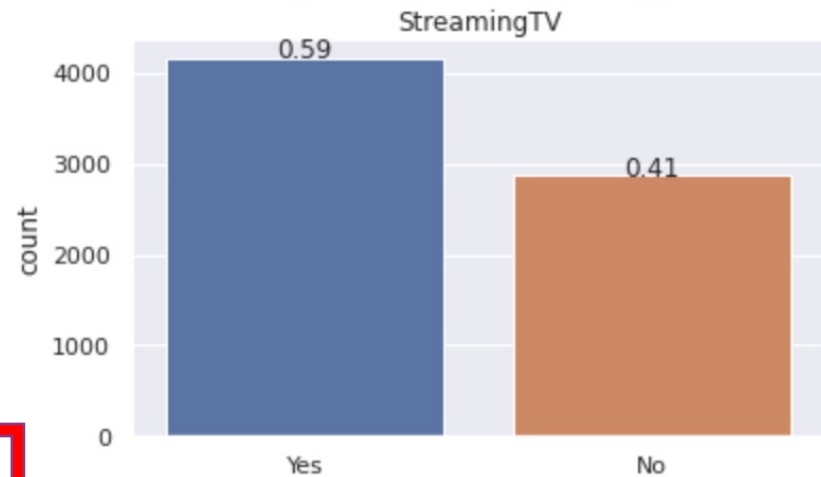
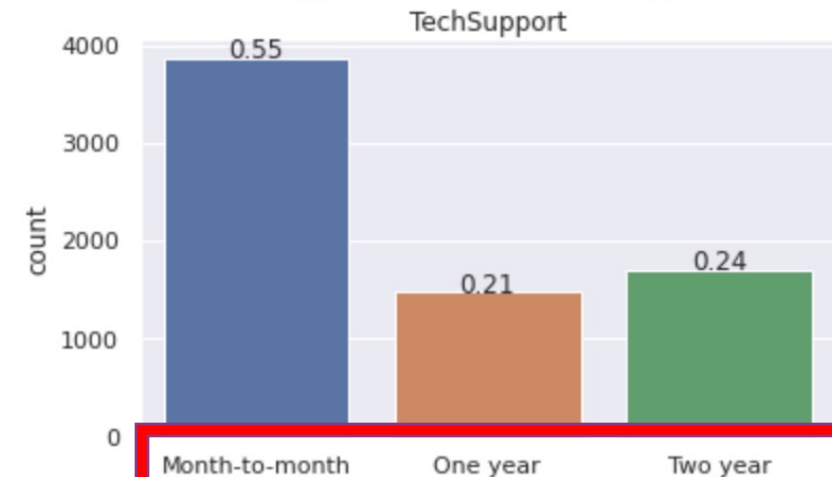
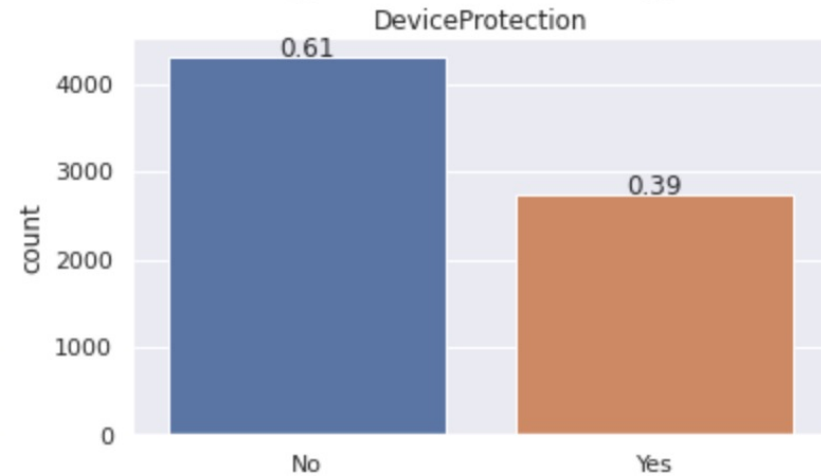
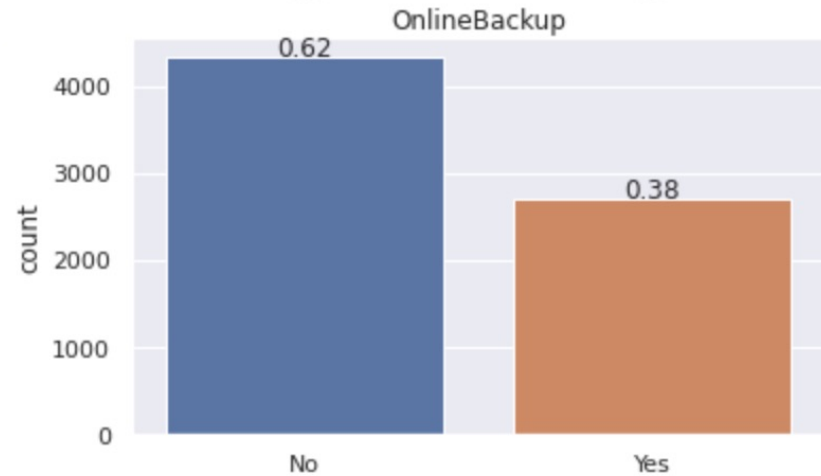
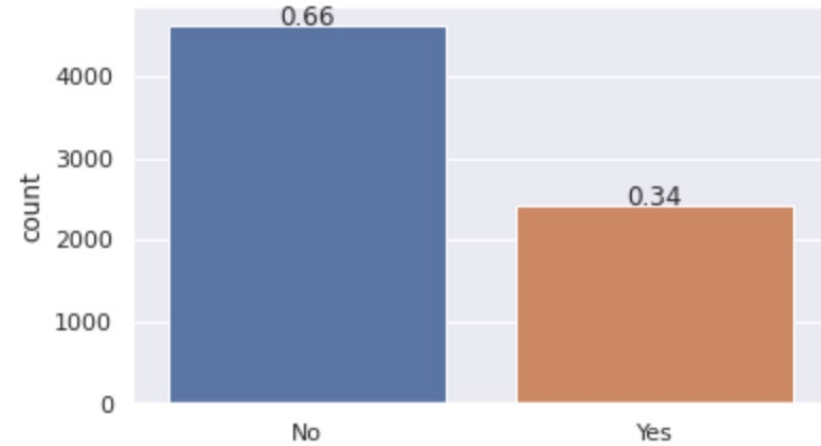
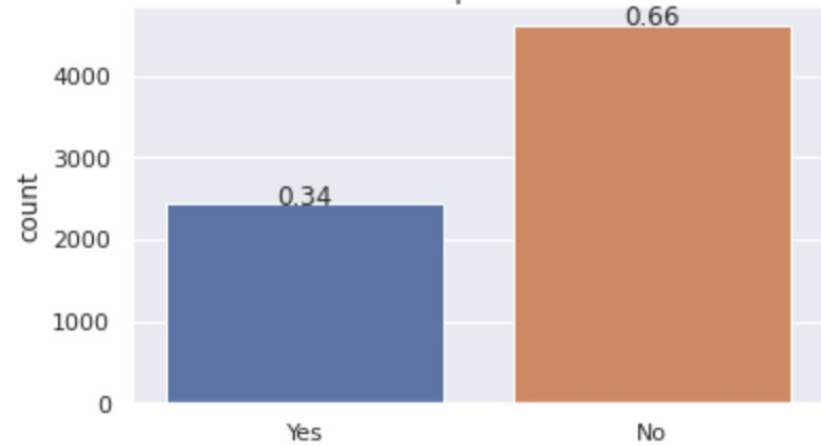
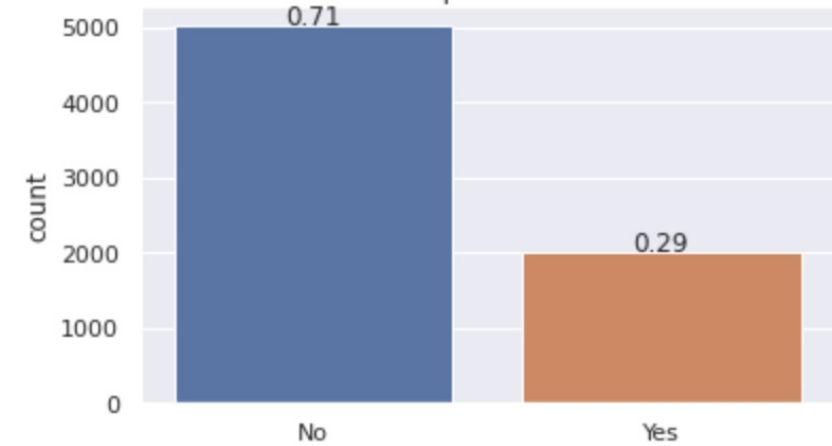
customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	DeviceProtection	TechSupport	StreamingTV	StreamingMovies	Contract	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges	Churn
7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	Yes	No	No	No	No	month-to-month	Yes	Electronic check	29.85	29.85	No
5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	No	Yes	No	No	No	one year	No	Mailed check	56.95	1889.0	No
3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	Yes	No	No	No	No	month-to-month	Yes	Mailed check	53.85	108.1	Yes
7795-CFOCV	Male	0	No	No	45	No	No phone service	DSL	Yes	No	Yes	Yes	No	No	one year	No	Bank transfer (automatic)	42.3	1840.7	No
9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	No	No	No	No	No	month-to-month	Yes	Electronic check	70.7	151.6	Yes
9305-CDSKC	Female	0	No	No	8	Yes	Yes	Fiber optic	No	No	Yes	No	Yes	Yes	month-to-month	Yes	Electronic check	99.65	820.0	Yes
1452-KIOVK	Male	0	No	Yes	22	Yes	Yes	Fiber optic	No	Yes	No	No	Yes	No	month-to-month	Yes	Credit card (automatic)	89.1	1949.0	No
6713-OKOM	Female	0	No	No	10	No	No phone service	DSL	Yes	No	No	No	No	No	month-to-month	No	Mailed check	29.75	301.0	No
7892-POOKF	Female	0	Yes	No	28	Yes	Yes	Fiber optic	No	No	Yes	Yes	Yes	Yes	month-to-month	Yes	Electronic check	104.8	3046.0	Yes
6388-TABGU	Male	0	No	Yes	62	Yes	No	DSL	Yes	Yes	No	No	No	No	one year	No	Bank transfer (automatic)	56.15	3487.9	No
9763-GRSKD	Male	0	Yes	Yes	13	Yes	No	DSL	Yes	No	No	No	No	No	month-to-month	Yes	Mailed check	49.95	587.4	No
7469-LKBCI	Male	0	No	No	16	Yes	No	No	No internet service	No internet service	No internet service	No internet service	No internet service	No internet service	two year	No	Credit card (automatic)	18.95	326.0	No
8091-TTVAX	Male	0	Yes	No	58	Yes	Yes	Fiber optic	No	No	Yes	No	Yes	Yes	one year	No	Credit card (automatic)	100.35	5681.0	No
0280-XJGEX	Male	0	No	No	49	Yes	Yes	Fiber optic	No	Yes	Yes	No	Yes	Yes	month-to-month	Yes	Bank transfer (automatic)	103.7	5036.0	Yes
5129-JLPIS	Male	0	No	No	25	Yes	No	Fiber optic	Yes	No	Yes	Yes	Yes	Yes	month-to-month	Yes	Electronic check	105.5	2686.0	No
3655-SNQYZ	Female	0	Yes	Yes	69	Yes	Yes	Fiber optic	Yes	Yes	Yes	Yes	Yes	Yes	two year	No	Credit card (automatic)	113.25	7895.1	No
8191-XWSZO	Female	0	No	No	52	Yes	No	No	No internet service	No internet service	No internet service	No internet service	No internet service	No internet service	one year	No	Mailed check	20.65	1022.9	No
9959-WOFK	Male	0	No	Yes	71	Yes	Yes	Fiber optic	Yes	No	Yes	No	Yes	Yes	two year	No	Bank transfer (automatic)	106.7	7382.2	No
4190-MFLUV	Female	0	Yes	Yes	10	Yes	No	DSL	No	No	Yes	Yes	No	No	month-to-month	No	Credit card (automatic)	55.2	528.3	Yes
4183-MYFRE	Female	0	No	No	21	Yes	No	Fiber optic	No	Yes	Yes	No	No	Yes	month-to-month	Yes	Electronic check	90.05	1862.0	No
8779-QRDM	Male	1	No	No	1	No	No phone service	DSL	No	No	Yes	No	No	Yes	month-to-month	Yes	Electronic check	39.65	39.65	Yes
1680-VDCW	Male	0	Yes	No	12	Yes	No	No	No internet service	No internet service	No internet service	No internet service	No internet service	No internet service	one year	No	Bank transfer (automatic)	19.8	202.2	No
1066-JKSGK	Male	0	No	No	1	Yes	No	No	No internet service	No internet service	No internet service	No internet service	No internet service	No internet service	month-to-month	No	Mailed check	20.15	20.15	Yes
3638-WEABV	Female	0	Yes	No	58	Yes	Yes	DSL	No	Yes	No	Yes	No	No	two year	Yes	Credit card (automatic)	59.9	3505.0	No
6322-HRPFA	Male	0	Yes	Yes	49	Yes	No	DSL	Yes	Yes	No	Yes	No	No	month-to-month	No	Credit card (automatic)	59.6	2970.0	No
6865-JZNKD	Female	0	No	No	30	Yes	No	DSL	Yes	Yes	No	No	No	No	month-to-month	Yes	Bank transfer (automatic)	55.3	1530.0	No
6467-CHFZV	Male	0	Yes	Yes	47	Yes	Yes	Fiber optic	No	Yes	No	No	Yes	Yes	month-to-month	Yes	Electronic check	99.35	4749.1	Yes
8665-UTDHZ	Male	0	Yes	Yes	1	No	No phone service	DSL	No	Yes	No	No	No	No	month-to-month	No	Electronic check	30.2	30.2	Yes
5248-YGIJN	Male	0	Yes	No	72	Yes	Yes	DSL	Yes	Yes	Yes	Yes	Yes	Yes	two year	Yes	Credit card (automatic)	90.25	6369.4	No
8773-HHUOZ	Female	0	No	Yes	17	Yes	No	DSL	No	No	No	No	Yes	Yes	month-to-month	Yes	Mailed check	64.7	1093.0	Yes
3841-NFECX	Female	1	Yes	No	71	Yes	Yes	Fiber optic	Yes	Yes	Yes	Yes	No	No	two year	Yes	Credit card (automatic)	96.35	6766.9	No
4929-XIHVV	Male	1	Yes	No	2	Yes	No	Fiber optic	No	No	Yes	No	Yes	Yes	month-to-month	Yes	Credit card (automatic)	95.5	181.6	No
6827-IEAUQ	Female	0	Yes	Yes	27	Yes	No	DSL	Yes	Yes	Yes	Yes	No	No	one year	No	Mailed check	66.15	1874.4	No

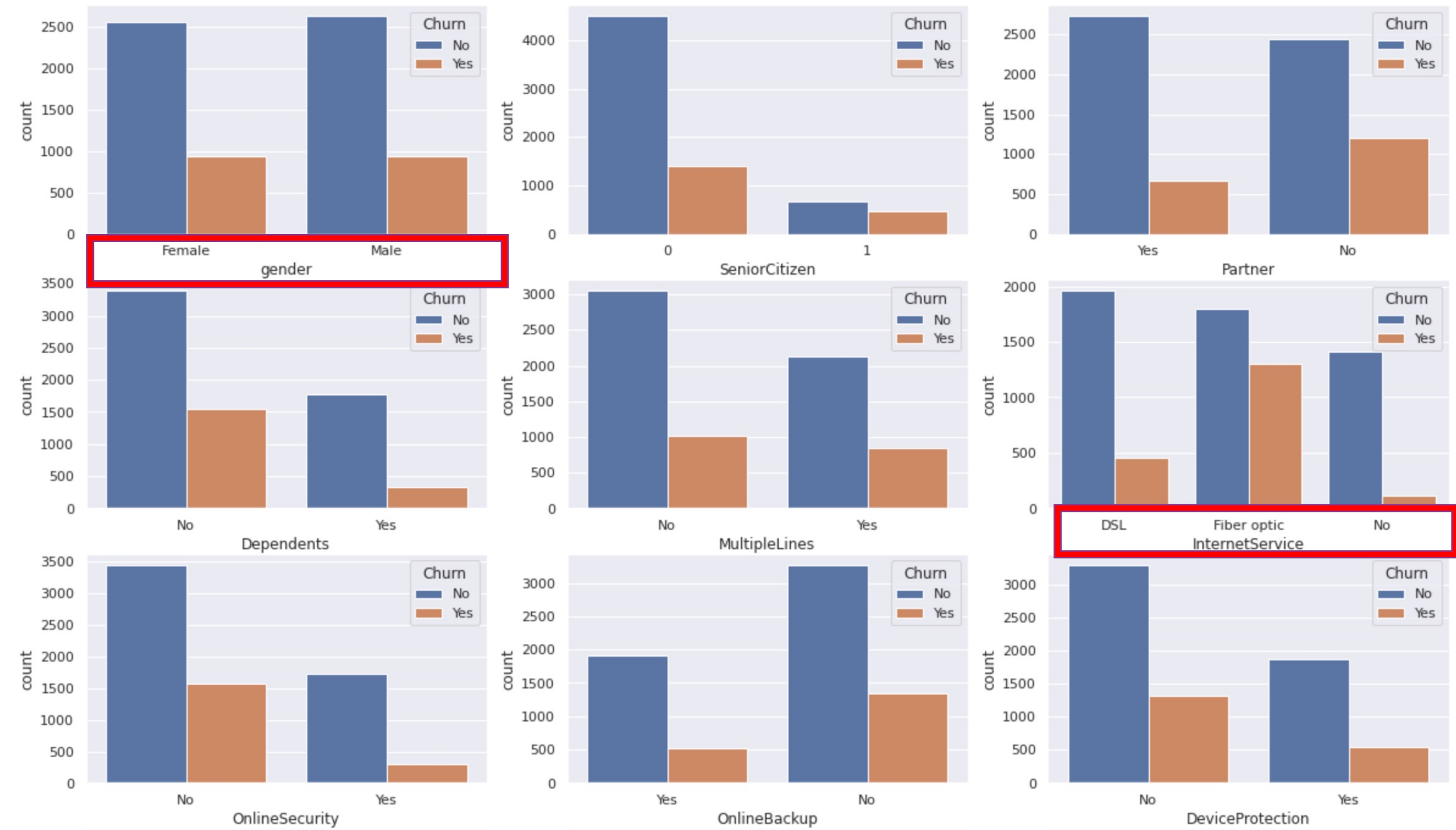


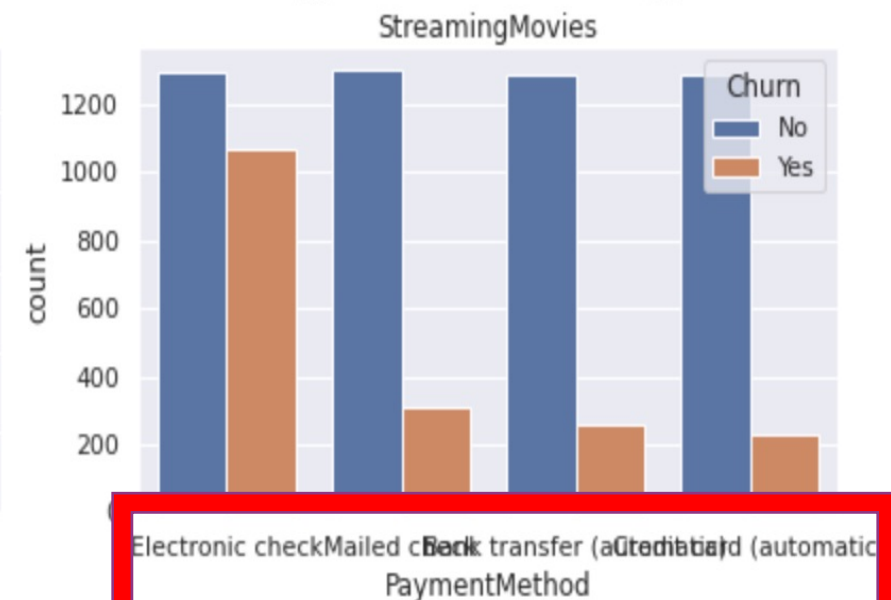
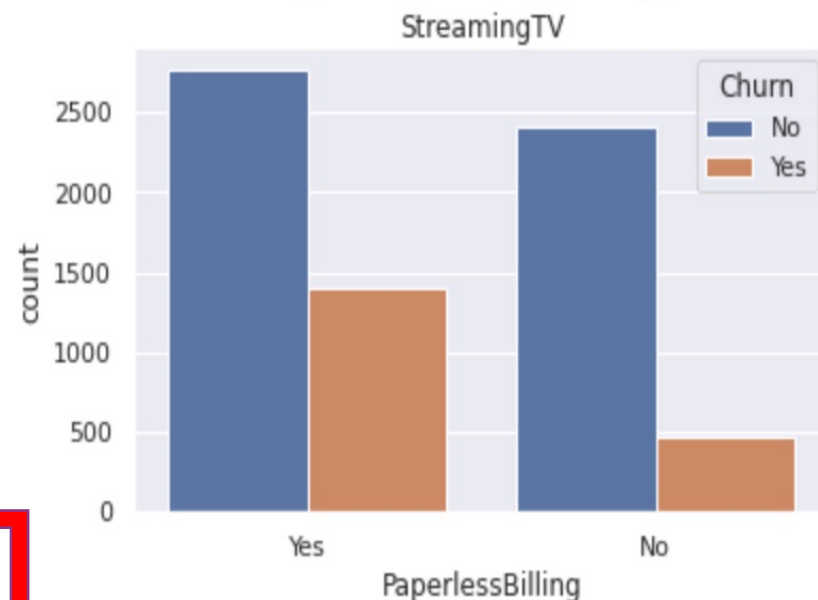
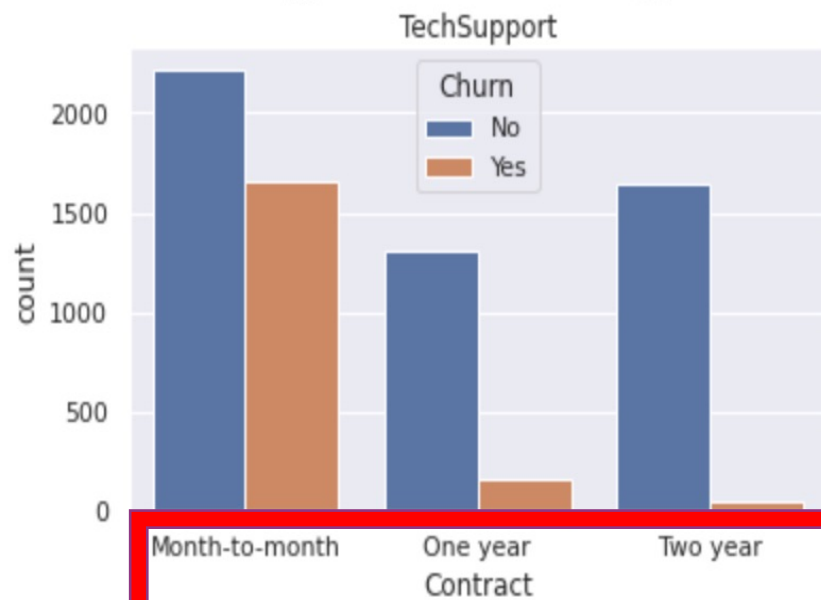
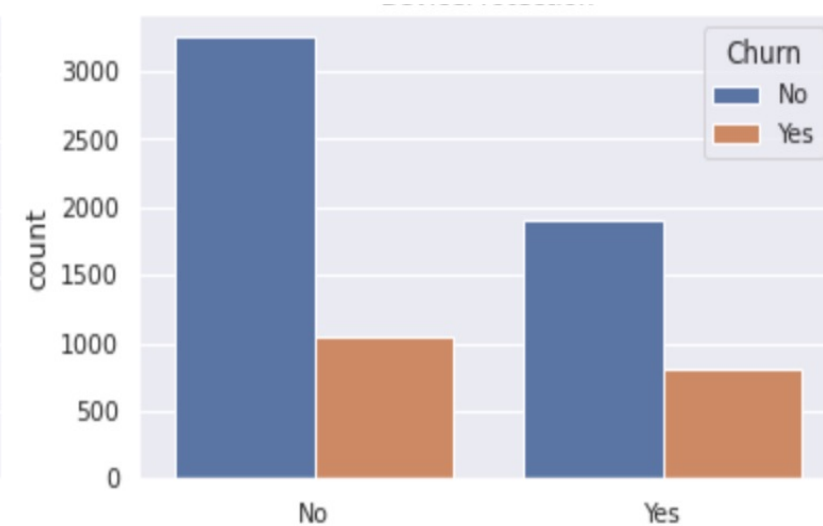
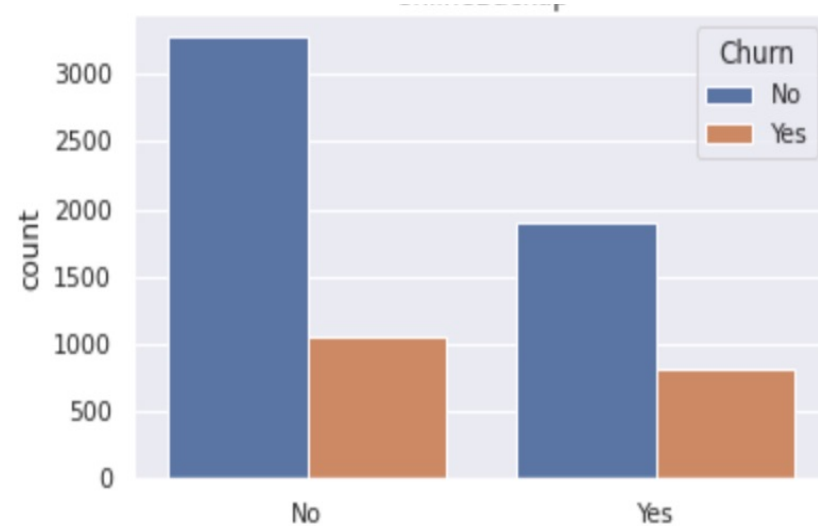
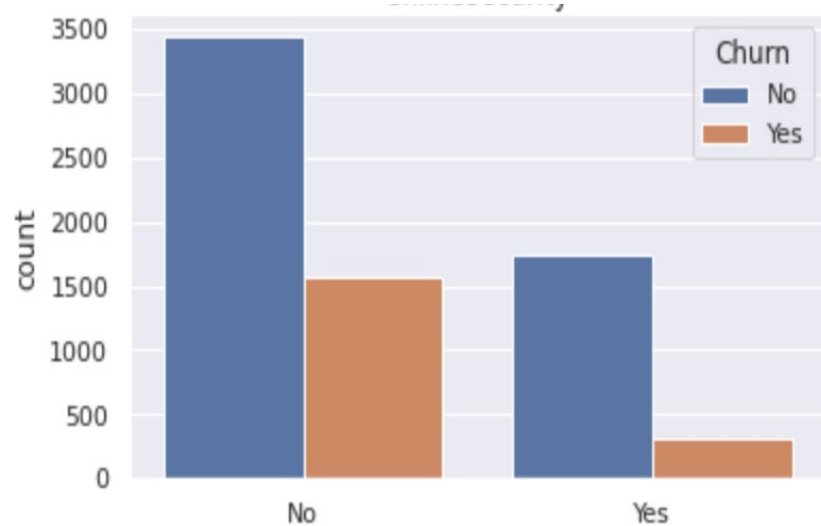


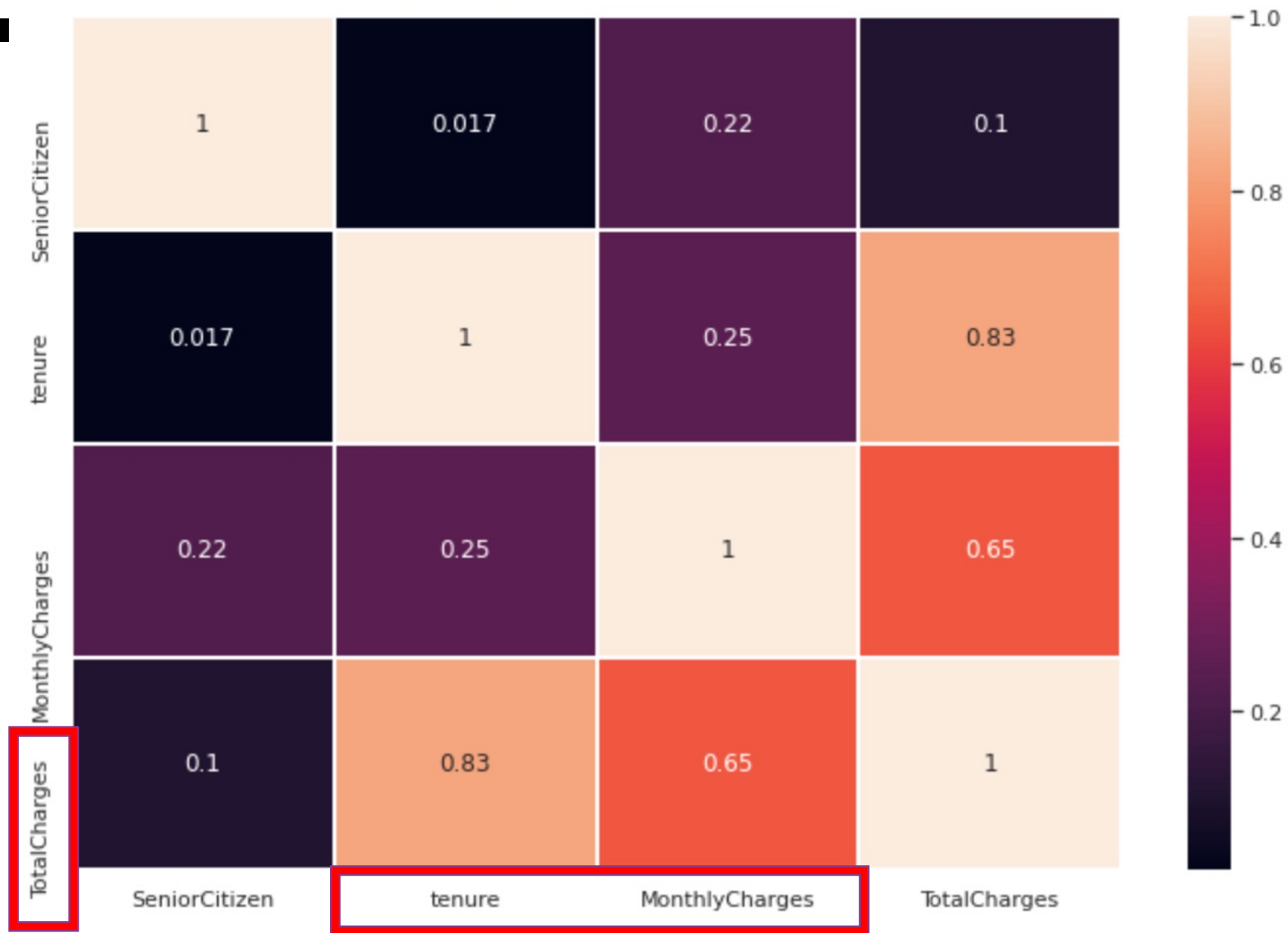
1. 불균형한 분포의 타겟
2. 평가지표 : recall, f1 score
- 3 .Baseline 모델 정하기
타겟의 최빈값인 no churn(약 0.73)
4. SMOTE 를 사용한 오버샘플링



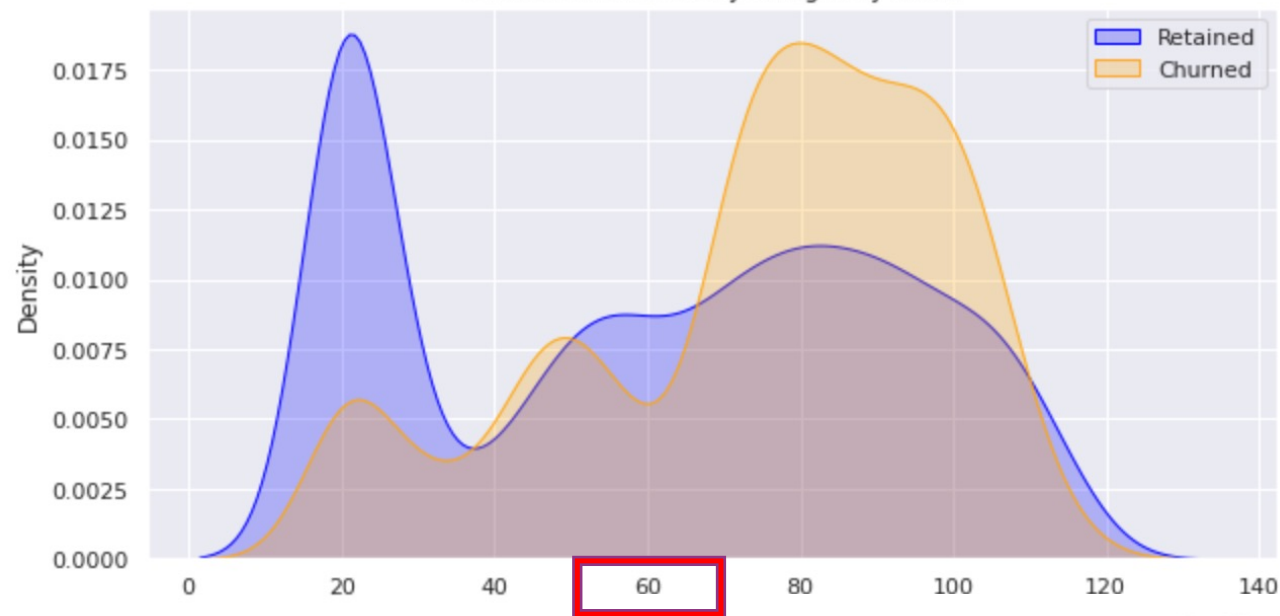




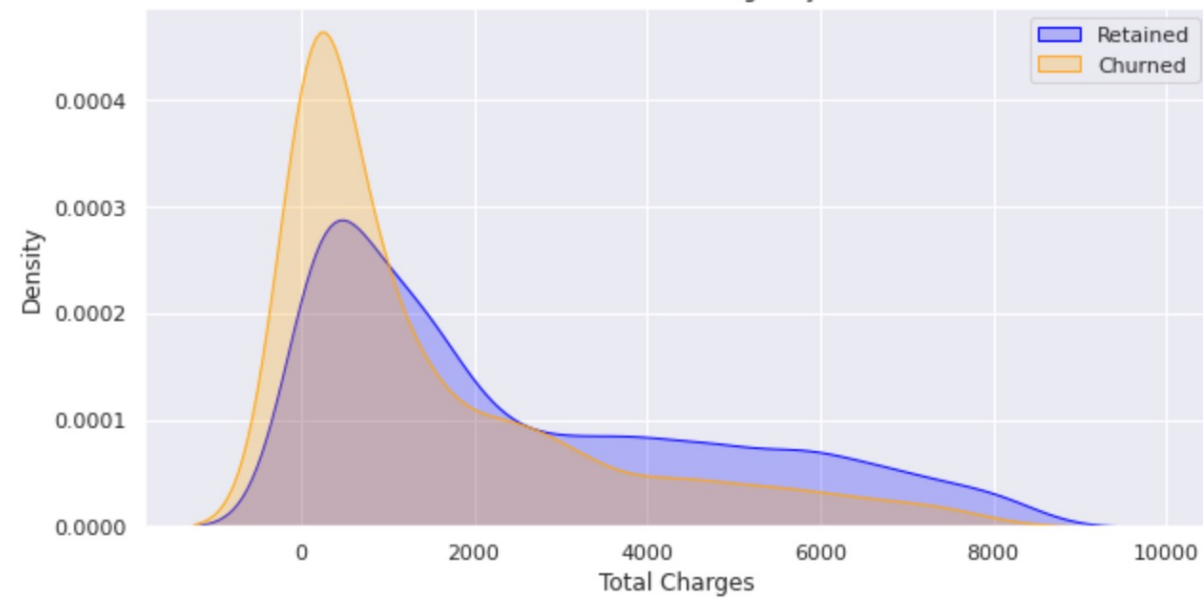




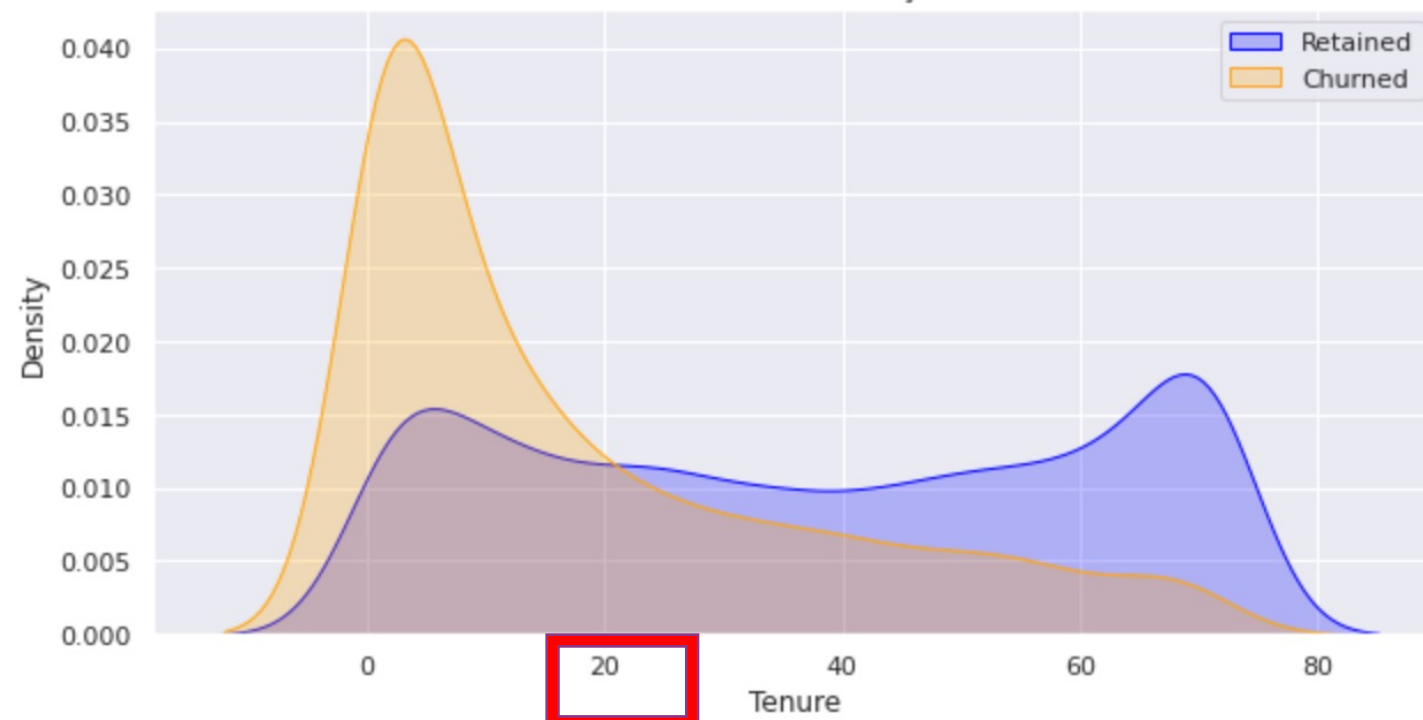
Distribution of monthly charges by churn



Distribution of total charges by churn



Distribution of tenure by churn



4. 머신러닝 방식 적용 및 교차검증

Model Name	Score
Logistic Regression	0.7985803016858918
Decision Tree	0.7187222715173026
Random Forest	0.7870452528837621
Xgboost	0.7293700088731144
Cat Boost	0.7577639751552795
KNN(k-nearest neighbor)	0.7284826974267968

4. 머신러닝 방식 적용 및 교차검증

Metrics	Random Forest
Accuracy	0.8105
Recall	0.8385
F1 score	0.7069

Metrics	Smote Random Forest
Accuracy	0.8269
Recall	0.8885
F1 score	0.8381

Metrics	Logistic
Accuracy	0.7935
Recall	0.5156
F1 score	0.5764

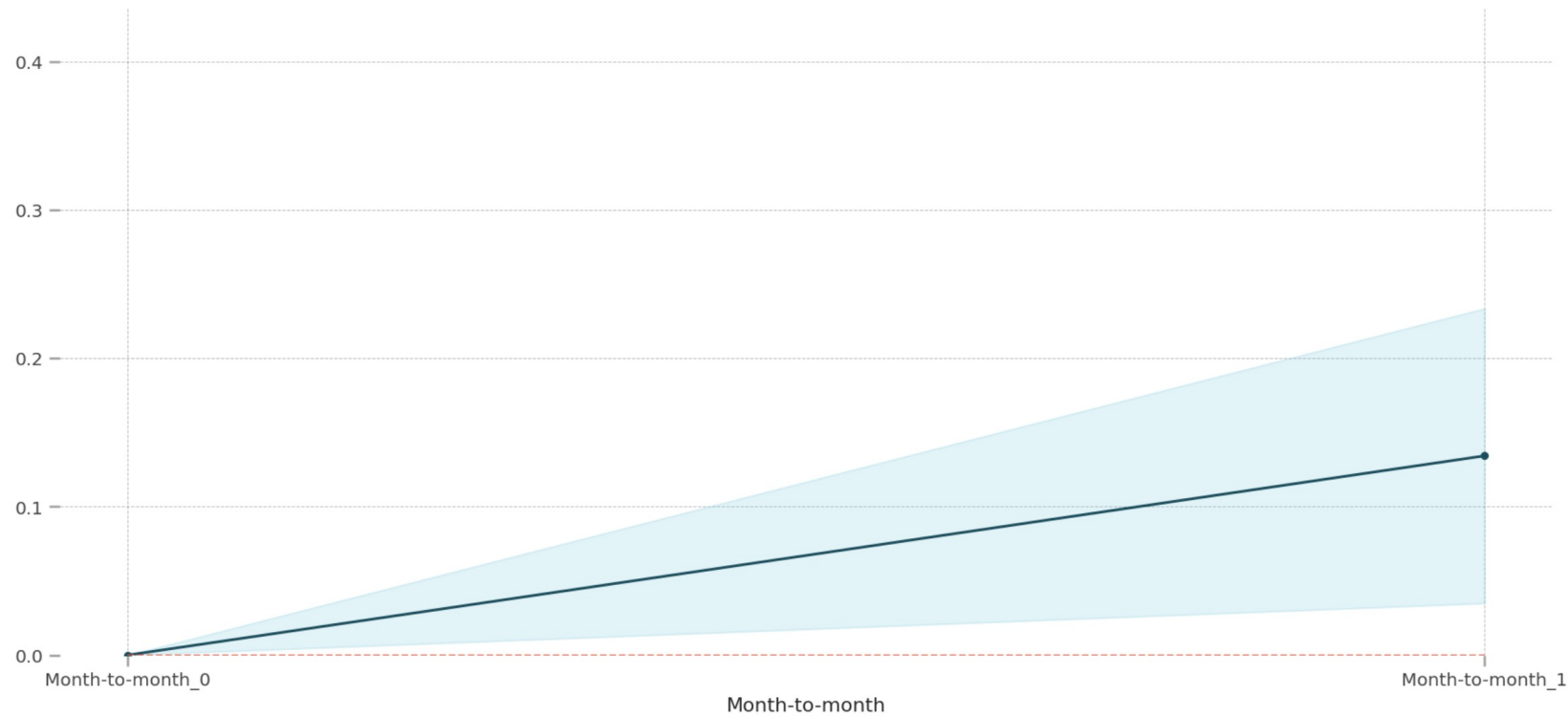
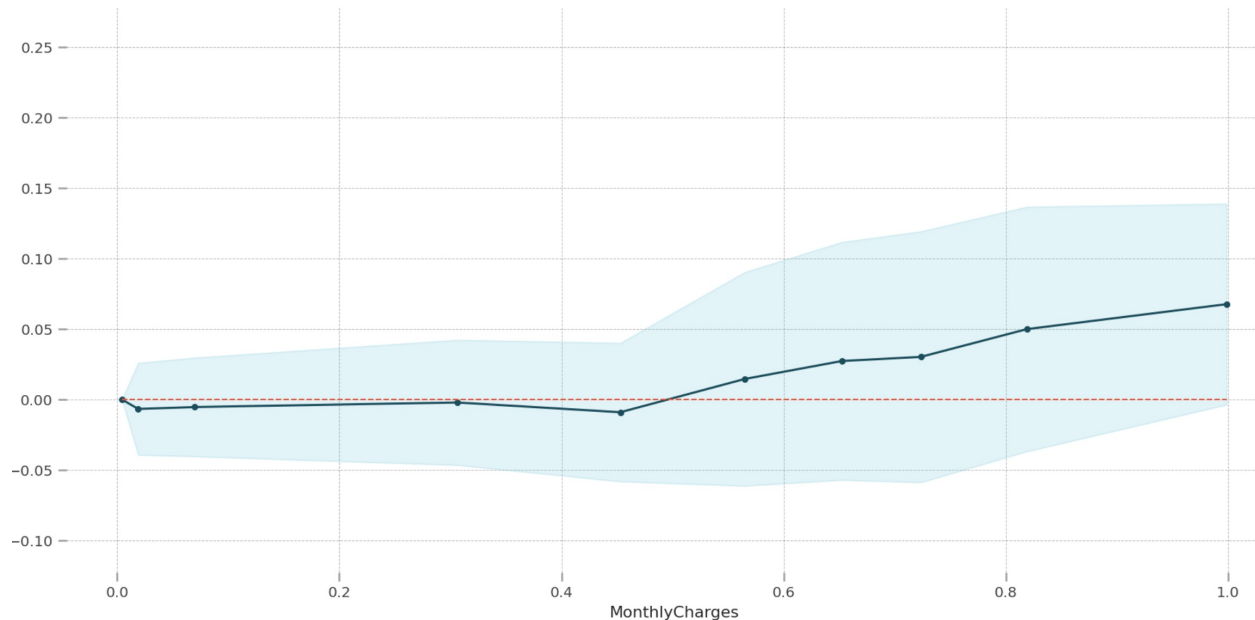
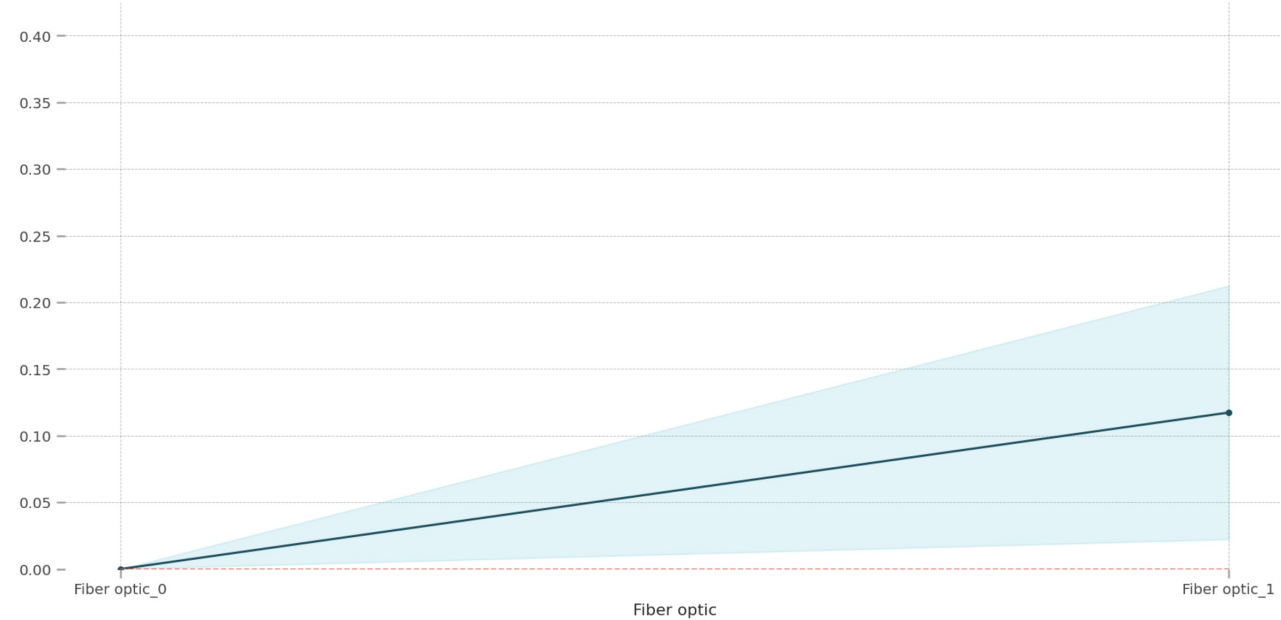
5. 머신러닝 모델 해석

(1) Permutation Importance

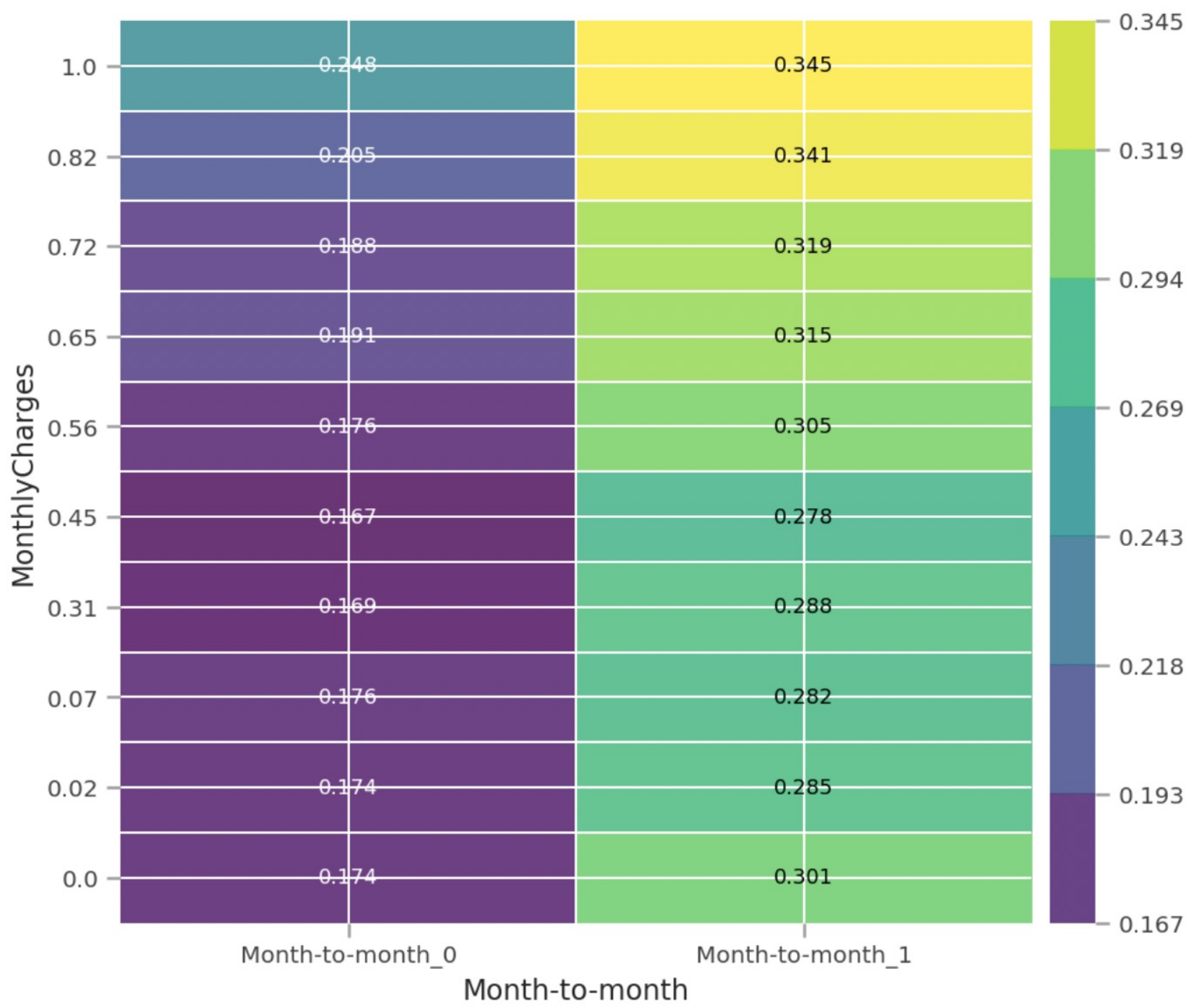
(2) PDP plot

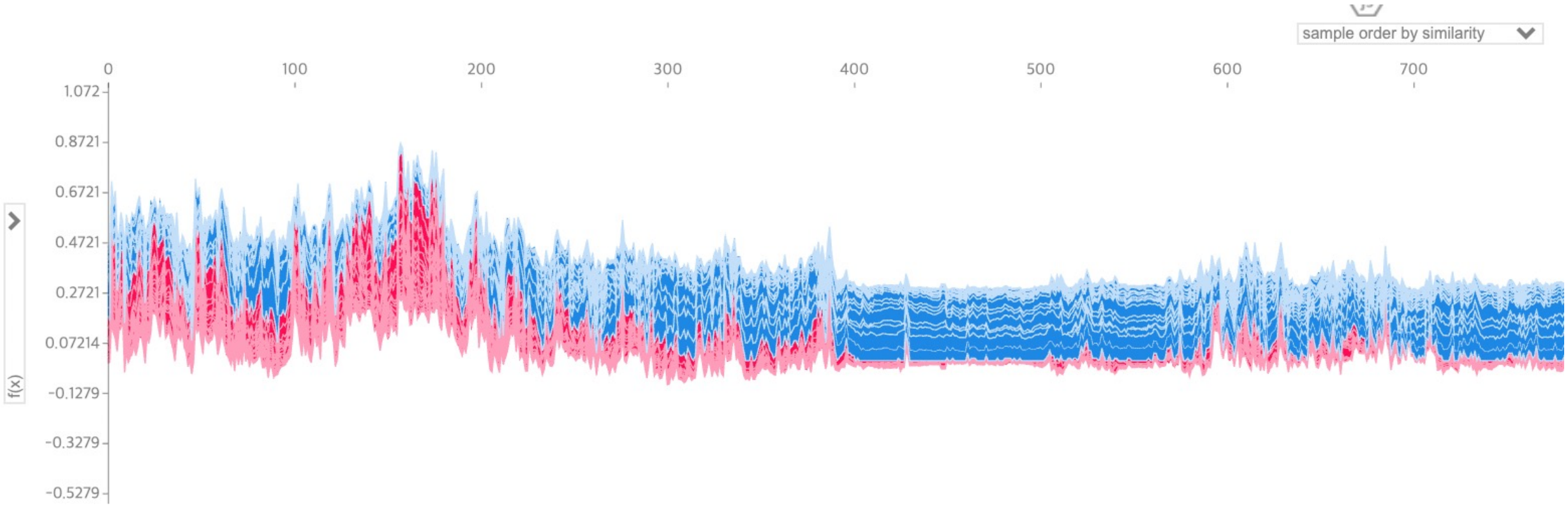
(3) Shap

Weight	Feature
0.1851 ± 0.0273	Month-to-month
0.1846 ± 0.0340	tenure
0.1592 ± 0.0192	Fiber optic
0.1295 ± 0.0174	MonthlyCharges
0.0518 ± 0.0088	TotalCharges
0.0386 ± 0.0126	DSL
0.0353 ± 0.0128	Two year
0.0259 ± 0.0154	Electronic check
0.0253 ± 0.0095	No
0.0226 ± 0.0081	PaperlessBilling
0.0204 ± 0.0166	One year
0.0143 ± 0.0172	OnlineSecurity
0.0072 ± 0.0124	SeniorCitizen
0.0072 ± 0.0027	PhoneService
0.0066 ± 0.0096	StreamingTV
0.0061 ± 0.0088	MultipleLines
0.0061 ± 0.0054	OnlineBackup
0.0033 ± 0.0081	Dependents
0.0028 ± 0.0078	StreamingMovies
0.0011 ± 0.0027	gender
0.0006 ± 0.0054	Credit card (automatic)
0.0006 ± 0.0064	TechSupport
0.0006 ± 0.0054	Partner
-0.0017 ± 0.0027	Bank transfer (automatic)
-0.0022 ± 0.0064	DeviceProtection
-0.0022 ± 0.0088	Mailed check



[REDACTED]





Month-to-month

Fiber optic

tenure

TotalCharges

Two year

MonthlyCharges

Electronic check

No

PaperlessBilling

OnlineSecurity

MultipleLines

TechSupport

StreamingTV

StreamingMovies

SeniorCitizen

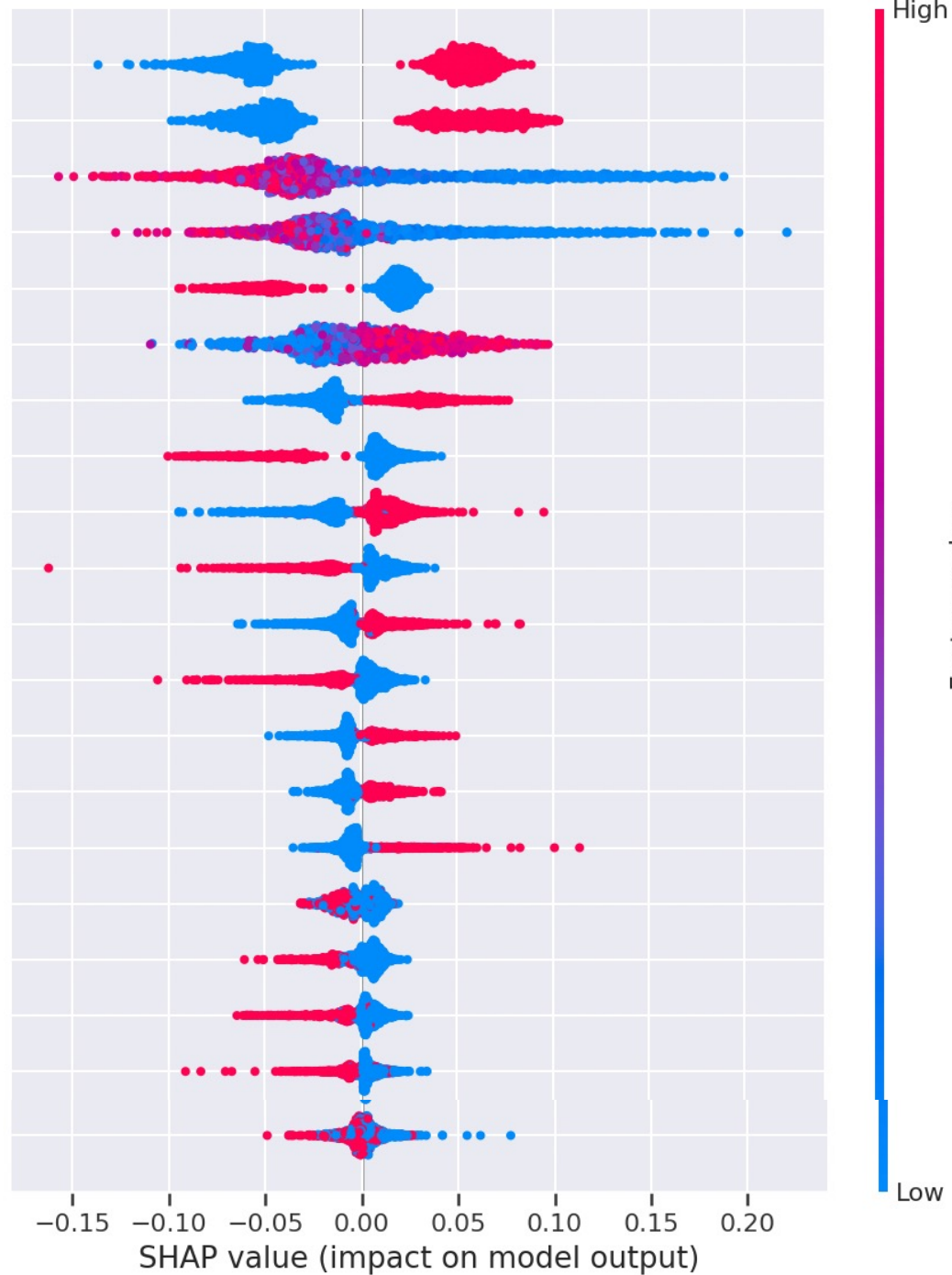
DSL

One year

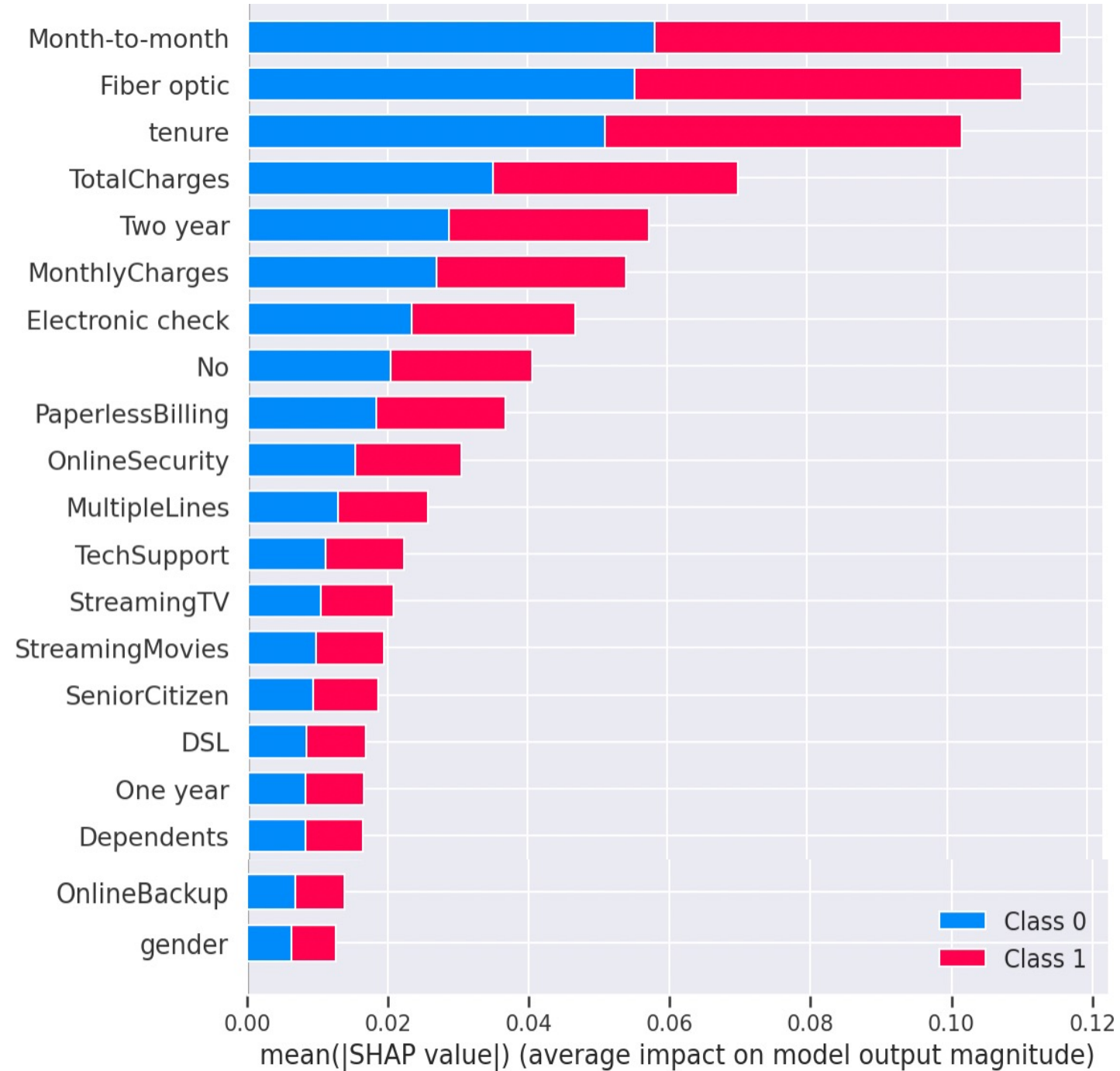
Dependents

OnlineBackup

gender



[Redacted]



6. 결론

(1) 가설 검증 :

- 성별에 따라서 고객의 이탈하는 정도에 차이가 있을까?

=> 성별에 따라서 고객의 이탈하는 정도에는 차이가 없었습니다.

- 사용기간에 따라서 고객이 이탈하는 정도에 차이가 있을까?

=> 약 10개월 미만의 tenure를 갖는 고객들이 훨씬 더 많이 이탈합니다. tenure가 60개월 이상인 사람들은 더 많이 유지됩니다.

따라서 사용기간에 따라서 고객이 이탈하는 정도에 차이가 있습니다.

- TotalCharges'가 'tenure'과 'MonthlyCharges'와 관련이 있을까?

=> TotalCharges는 tenure, MonthlyCharges와 높은 상관관계를 보였습니다. 따라서 셋은 서로 관련이 있습니다. (히트맵)

6. 결론

(2) 문제 정의 : "통신회사에서 고객의 이탈을 어떻게 줄일 수 있을까?"

(3) 해결책 :

1. 계약(Contract)을 Month-to-month로 한 고객들에 MonthlyCharges가 낮은 상품을 마케팅해서 이탈하지 않도록 합니다. Month-to-month 고객들이 이탈을 하는 빈도가 높으면서 동시에 이탈(churn)이라는 타겟에 유의미한 영향을 끼치는 중요한 고객입니다. 그들은 MonthlyCharges가 높으면 이탈하는 경향이 있기에 월별 요금이 적은 상품을 마케팅하면 장기고객으로 만들어낼 수 있습니다.

2. InternetService 중 하나인 Fiber optic 서비스를 점검할 필요가 있습니다. Fiber optic은 타겟에 큰 영향을 주는 특성 중 하나입니다. countplot을 통해 대부분의 이탈고객들이 fiber optic을 사용하고 있었기때문입니다. 따라서 해당 서비스에 문제가 있어 고객에게 부정적인 경험을 제공한건 아닌지 설문조사를 해서 해당 부분을 고친다면 고객의 이탈을 막을 수 있습니다.

6. 결론

(4) 데이터의 한계

- 시간이나 분기와 관련된 데이터가 전무
- 따라서 그때 그때의 흐름과 트렌드를 읽어야하는 고객 이탈 이슈에서 이 데이터로는 완벽한 전략을 구사하는데 제한적
- 따라서 각 고객별로 분기나 시간 등 언제 어떤 기간에 이탈(churn)을 했는지 알 수 있었다면 좀 더 정확하게 다음 분기 전략을 짤 수 있었을 것이다.



감사합니다